ORF467F16 Final Report: Nation-wide Ride-sharing:

A first look at the potential and its implications on Fleet Size, Fleet Operations, Integration with Existing Urban Rail, Intercity Rail, Intercity air system, and Societal benefits

Professor Alain Kornhauser
Richard Bush
Alex Manoloff
Houston Martinez
Wyatt Nevarro

Sam Button
Jessica Deng
Hassan Ejaz Chaudhry
Adam Hare
William Haynes
Jeb Helmers
Ifeyinwa Ikpeazu
August Kiles
Che-yu Liu
Jarret Lowell
Julia Ni
Bill Van Cleve
Evan Wood
Alex Yablonski

Table of Contents

Chapter 1:
Executive Summary

A first look at Nationwide data sets:
Comparing states and distinguishing trips by length and type.

Richard Bush, Alex Manoloff, Wyatt Navarro,
Houston Martinez

1.1 Description of data sets and Summaries of Synthesized Data Sets

1.1.1 Population and PersonTrips in the United States

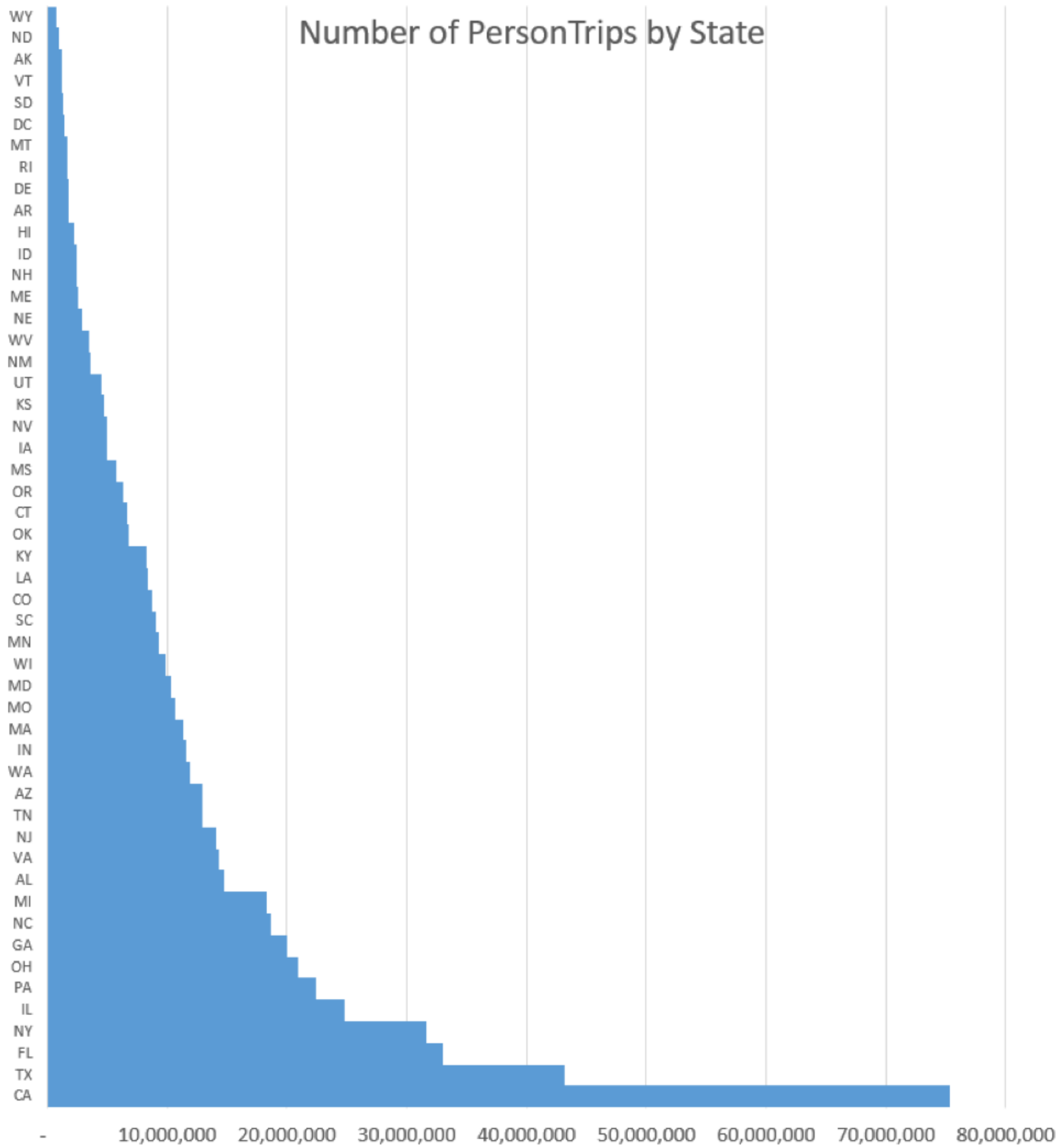
Our process began by creating 308,745,538 "US Residents" in the form of Excel sheet rows. Each row, one for every 'resident' of the US, held all of the information necessary to identify a person, including their age, income, sex, place of residence, amongst others.

From these records, we generated a daily 'trip tour' that a person would take on a given day based on their age, income, etc. From these trip tours, we were able to quantify the number of PersonTrips (trips taken by a person, as opposed to a VehicleTrip) by each state. Find below the state-by-state breakdown of the total number of PersonTrips that take place each day in the US.

Figure 1.1 - PersonTrips count by state

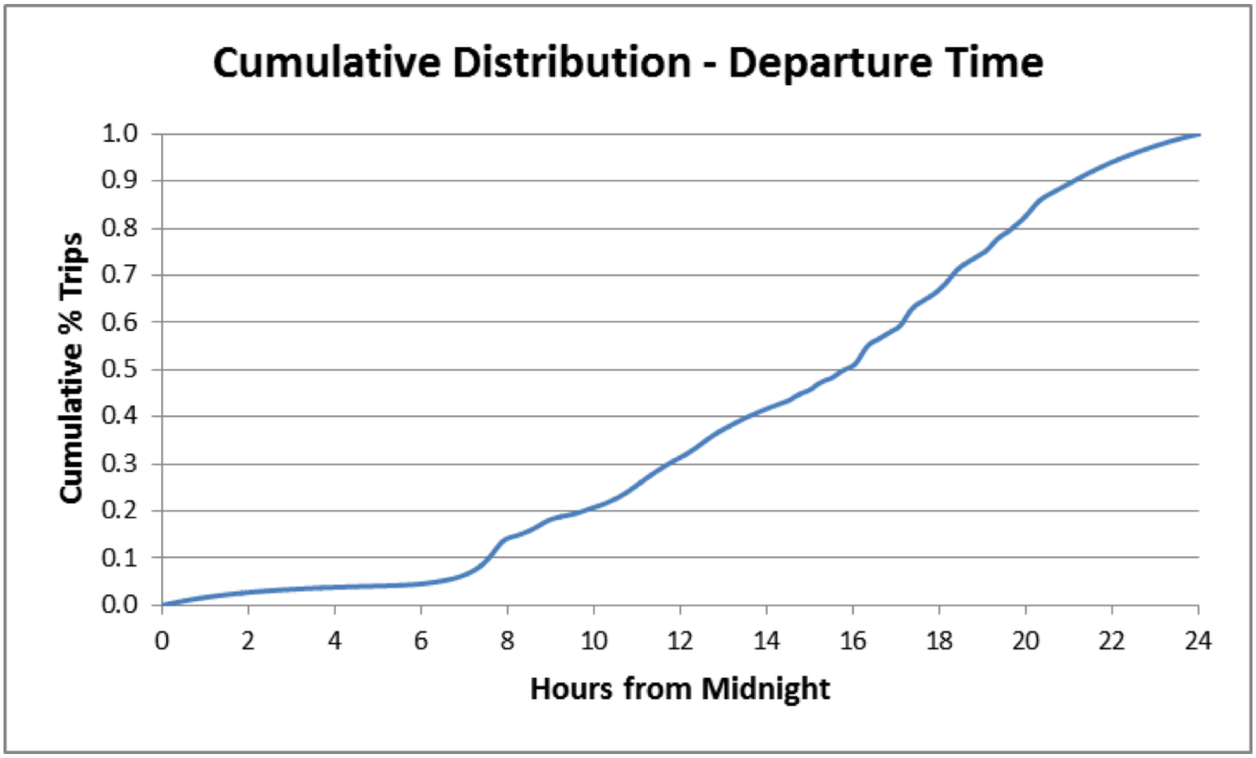
State	Trip Count	State	Trip Count	State	Trip Count
Alabama	15,825,280	Kentucky	14,349,637	North Dakota	2,196,072
Alaska	2,397,128	Louisiana	8,148,094	Ohio	38,087,870
Arizona	20,903,632	Maine	4,429,255	Oklahoma	12,262,687
Arkansas	9,554,065	Maryland	19,316,346	Oregon	12,730,747
California	123,852,078	Massachusetts	21,866,204	Pennsylvania	41,709,485
Colorado	16,839,860	Michigan	32,833,666	Rhode Island	3,489,284
Connecticut	11,850,814	Minnesota	11,167,667	South Carolina	15,262,810
DC	2,040,597	Mississippi	9,717,958	South Dakota	2,627,648
Delaware	2,970,506	Missouri	19,699,659	Tennessee	21,096,931
Florida	61,275,215	Montana	3,261,369	Texas	83,584,971
Georgia	32,302,424	Nebraska	5,979,671	Utah	9,047,267
Hawaii	4,437,926	Nevada	9,048,868	Vermont	2,104,664
Idaho	5,141,420	New Hampshire	4,444,770	Virginia	26,646,786
Illinois	42,657,513	New Jersey	29,237,285	Washington	22,475,312
Indiana	21,431,504	New Mexico	6,779,308	West Virginia	6,106,765
Iowa	9,943,451	New York	64,529,719	Wisconsin	18,824,313
Kansas	9,327,336	North Carolina	31,638,807	Wyoming	1,870,191
Total	1,009,322,835				

1.1.2 PersonTrips by State

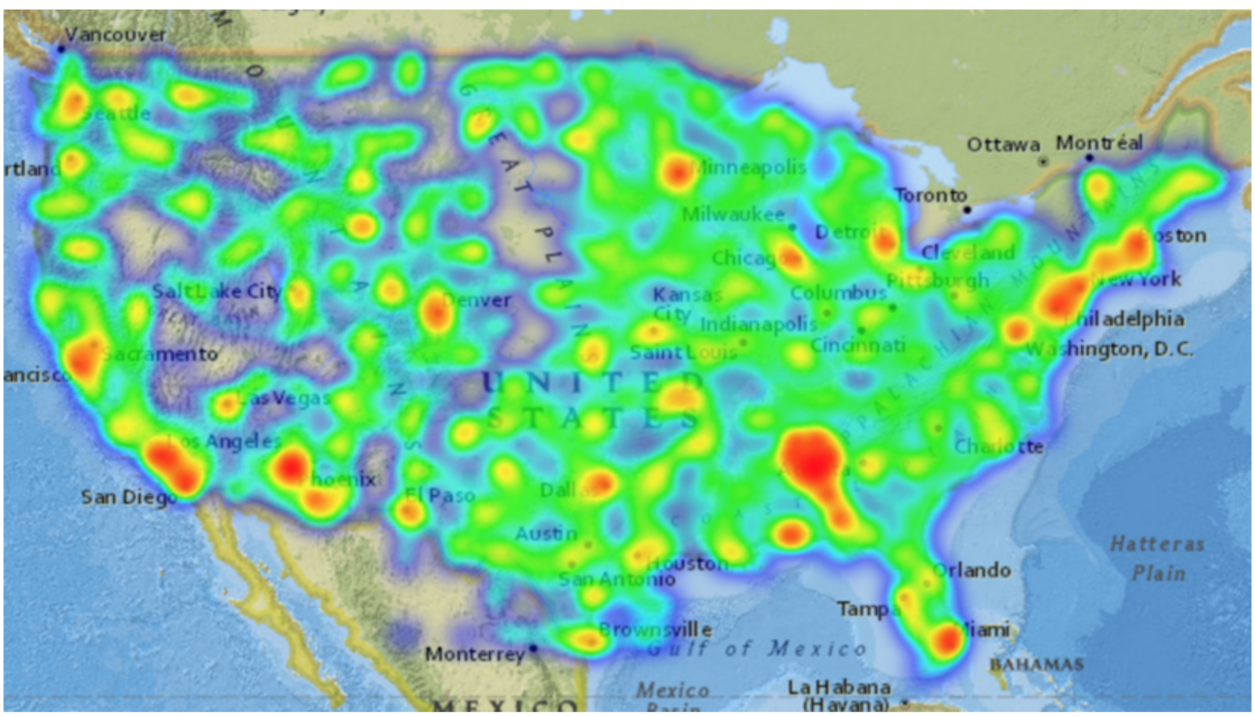


State	Population	PersonTrips	Average Trips/Person	% of all Walking Trips	% of all Short Trips	% of all Regular Trips
Alabama	4,833,722	15,825,280	3.27	0.7	1.3	2.6
Alaska	735,132	2,387,128	3.25	0.3	0.2	0.2
Arizona	6,626,624	20,903,632	3.15	2.6	1.8	2.3
Arkansas	2,959,373	9,554,065	3.23	0.4	0.3	0.3
California	38,332,521	123,852,078	3.23	19.9	13.9	13.3
Colorado	5,268,367	16,839,860	3.20	2	1.7	1.5
Connecticut	3,596,080	11,850,814	3.30	1	1.2	1.2
Delaware	925,749	2,040,597	2.20	0.2	0.3	0.3
District of Columbia	646,449	2,970,506	4.60	0.5	0.3	0.3
Florida	19,552,860	61,275,215	3.13	3.5	5.5	5.8
Georgia	9,992,167	32,302,424	3.23	1.5	2.9	3.5
Hawaii	1,404,054	4,437,926	3.16	0.5	0.5	0.4
Idaho	1,612,136	5,141,420	3.19	0.6	0.6	0.4
Illinois	12,882,135	42,657,513	3.31	6.4	4.2	4.4
Indiana	6,570,902	21,431,504	3.26	1.5	2.1	2
Iowa	3,090,416	9,943,451	3.22	1.2	1.1	0.9
Kansas	2,893,957	9,327,336	3.22	1.2	1	0.8
Kentucky	4,395,295	14,349,637	3.26	0.8	1.3	1.5
Louisiana	4,625,470	8,148,094	1.76	0.9	1.5	1.5
Maine	1,328,302	4,429,255	3.33	0.2	0.4	0.5
Maryland	5,928,814	19,316,346	3.26	1.3	1.9	1.8
Massachusetts	6,692,824	21,866,204	3.27	2.1	2.3	2
Michigan	9,895,622	32,833,666	3.32	2.5	3	3.2
Minnesota	5,420,380	11,167,667	2.06	1.6	1.8	1.6
Mississippi	2,991,207	9,717,958	3.25	0.5	0.8	1
Missouri	6,044,171	19,699,649	3.26	1.5	1.9	1.9
Montana	1,015,165	3,261,369	3.21	0.4	0.4	0.3
Nebraska	1,868,516	5,979,671	3.20	0.9	0.7	0.5
Nevada	2,790,136	9,048,868	3.24	0.9	0.8	0.9
New Hampshire	1,323,459	4,444,770	3.36	0.2	0.4	0.4
New Jersey	8,899,339	29,237,285	3.29	3.3	3.1	2.5
New Mexico	2,085,287	6,779,309	3.25	0.6	0.7	0.6
New York	19,651,127	64,529,719	3.28	8.8	6.9	5.6
North Carolina	9,848,060	31,638,807	3.21	1.2	2.6	3.3
North Dakota	723,393	2,196,072	3.04	0.3	0.3	0.2
Ohio	11,570,808	38,087,870	3.29	2.7	3.6	3.7
Oklahoma	3,850,568	12,262,687	3.18	1.2	1.2	1.2
Oregon	3,930,065	12,730,747	3.24	1.4	1.4	1.1
Pennsylvania	12,773,801	41,709,485	3.27	3.2	4	4
Rhode Island	1,051,511	3,489,284	3.32	0.3	0.4	0.3
South Carolina	4,774,839	15,262,810	3.20	0.5	1.2	1.6
South Dakota	844,877	2,627,648	3.11	0.4	0.3	0.2
Tennessee	6,495,978	21,096,931	3.25	1	1.8	2.3
Texas	26,448,193	83,584,971	3.16	8.3	7.3	7.6
Utah	2,900,872	9,047,267	3.12	1.2	1	0.8
Vermont	626,630	2,104,664	3.36	0.1	0.2	0.2
Virginia	8,260,405	26,646,786	3.23	2.9	3.1	2.5
Washington	6,971,406	22,475,312	3.22	2.2	2.2	2.1
West Virginia	1,854,304	6,106,765	3.29	0.4	0.5	0.6
Wisconsin	5,742,713	18,824,313	3.28	1.7	1.9	1.7
Wyoming	582,658	1,870,191	3.21	0.3	0.2	0.1
Nationwide	**308,745,538**	**1,009,322,835**	**3.27**	**1**	1	1

1.1.2.1 National Cumulative Distributions



1.1.2.2 Heat map of PersonTrip/pixel of whole country



0.2 Modal Share Nationwide Summary results

1.2.1 Walking Trips

Walking trips for this analysis were defined as being less than 0.5 miles in Great Circle distance from their person trip records. Due to the cumbersome nature of having an aTaxi pick up and drop off a passenger within 0.5 miles, this segment of the analysis is used as somewhat of a filter for the person trip files since walking trips would not be serviced by aTaxis and thus are being 'filtered' out of the aTaxi ridership feasibility analysis data.

# Trips	% Total	personTripMiles	% Total	AveragePersonTripLength
22,803,288	.026%	6,868,765	.067%	.3 Miles

*totals include driving trips (less than 200 miles)

1.2.2 Short aTaxi Trips

Short trips were defined between 0.5 and 2 miles in Great Circle distance from their person trip records.

1.2.2.1 personTrips

# Trips	% Total	personTripMiles	% Total	AveragePersonTripLength
282,346,209	32.5%	307,524,675	3%	1.09 Miles

*totals include driving trips (less than 200 miles)

1.2.2.2 vehicleTrips

# Trips	PersonTripMiles	vehicleMiles	AVO
72,295,574	2.82E+08	2.03E+08	1.39

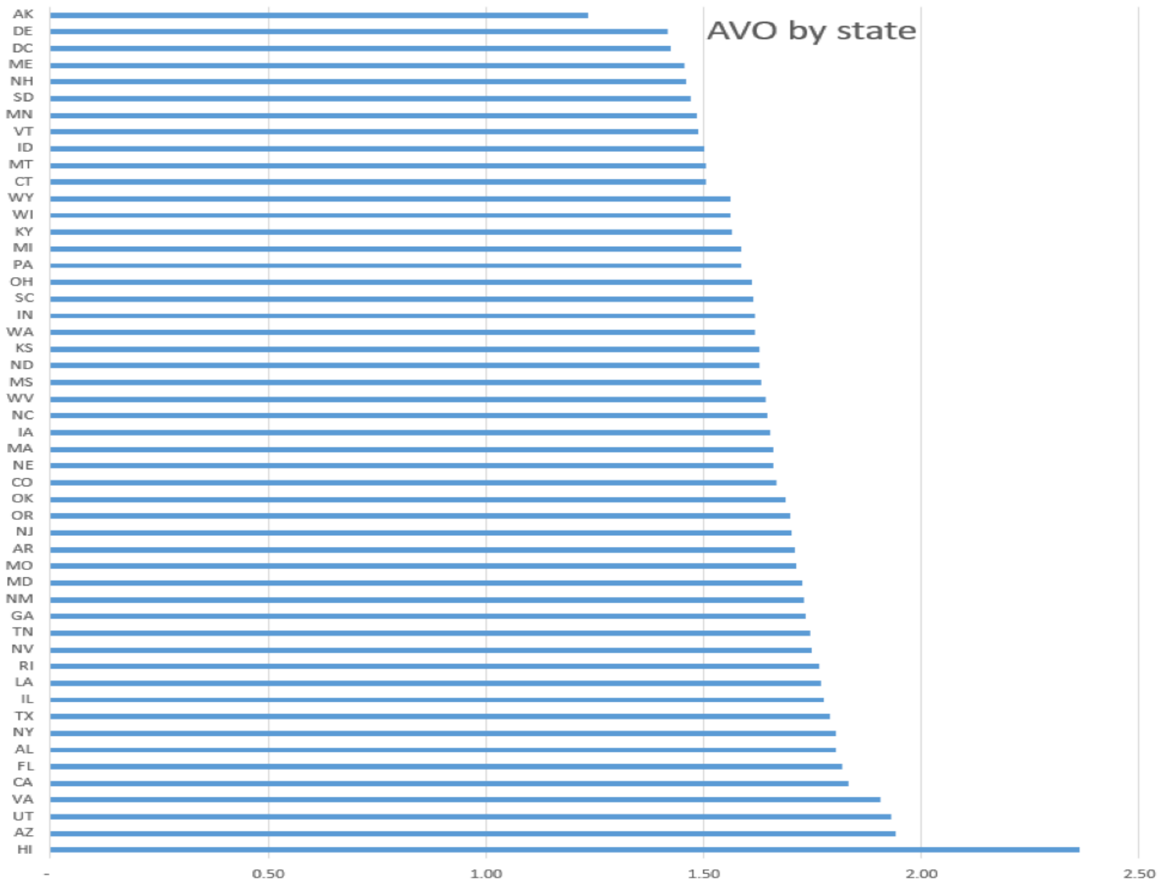
1.2.3 Normal aTaxi Trips (2.0 < X < 200miles)

1.2.3.1 personTrips

State	State Code	regTrips	aTaxi TripCount	regTrips/ aTaxiTripCount	regTrips Miles	aTaxi Miles	AVO (regTripsMiles / aTaxiMiles)
Alabama	AL	14,817,309	6,376,659	2.32	632,158,317	350,381,323	1.80
Alaska	AK	1,208,987	521,815	2.32	26,418,510	21,419,061	1.23
Arizona	AZ	12,927,641	4,841,858	2.67	305,147,214	157,255,392	1.94
Arkansas	AR	1,759,568	766,139	2.30	90,314,799	52,843,918	1.71
California	CA	75,274,095	25,365,328	2.97	1,242,516,857	678,132,130	1.83
Colorado	CO	8,706,925	3,498,241	2.49	126,435,035	75,837,941	1.67
Connecticut	CT	6,662,969	2,824,128	2.36	122,783,818	81,537,716	1.51
Delaware	DE	1,738,643	723,154	2.40	34,042,765	24,028,432	1.42
District of Columbia	DC	1,468,686	401,899	3.65	35,545,392	24,938,371	1.43
Florida	FL	32,998,488	12,969,351	2.54	523,231,835	287,869,332	1.82
Georgia	GA	19,976,818	8,756,810	2.28	442,483,195	255,476,928	1.73
Hawaii	HI	2,267,887	693,005	3.27	30,742,076	13,004,381	2.36
Idaho	ID	2,417,771	1,154,965	2.09	34,936,467	23,250,730	1.50
Illinois	IL	24,782,640	9,517,125	2.60	418,412,917	235,816,448	1.77
Indiana	IN	11,564,843	5,369,402	2.15	198,413,872	122,712,820	1.62
Iowa	IA	4,977,361	2,459,759	2.02	99,185,484	60,060,919	1.65
Kansas	KS	4,707,145	2,201,806	2.14	94,924,363	58,297,679	1.63
Kentucky	KY	8,278,433	3,882,177	2.13	145,335,090	92,779,380	1.57
Louisiana	LA	8,391,554	3,679,703	2.28	182,573,621	103,311,587	1.77
Maine	ME	2,546,828	1,382,888	1.84	46,785,614	32,169,170	1.45
Maryland	MD	10,380,493	4,010,447	2.59	127,987,692	74,163,606	1.73
Massachusetts	MA	11,358,238	4,668,683	2.43	156,491,863	94,287,883	1.66
Michigan	MI	18,323,460	8,635,735	2.12	310,166,620	195,531,912	1.59
Minnesota	MN	9,303,861	4,327,872	2.15	146,621,109	98,836,312	1.48
Mississippi	MS	5,767,286	2,872,499	2.01	136,004,338	83,408,320	1.63
Missouri	MO	10,713,995	4,977,922	2.15	202,358,872	118,209,384	1.71
Montana	MT	1,612,789	816,008	1.98	34,484,275	22,918,115	1.50
Nebraska	NE	2,903,605	1,321,666	2.20	44,873,601	27,022,196	1.66
Nevada	NV	4,924,667	1,787,517	2.76	52,185,662	29,838,669	1.75
New Hampshire	NH	2,490,348	1,278,855	1.95	41,026,938	28,108,055	1.46
New Jersey	NJ	14,073,301	5,552,053	2.53	179,902,504	105,708,554	1.70
New Mexico	NM	3,609,809	1,506,586	2.40	77,465,279	44,728,119	1.73
New York	NY	31,693,655	11,426,112	2.77	480,684,713	266,469,297	1.80
North Carolina	NC	18,600,152	8,725,553	2.13	290,286,662	176,474,069	1.64
North Dakota	ND	1,023,363	527,031	1.94	25,368,319	15,564,811	1.63
Ohio	OH	20,892,719	9,481,675	2.20	352,905,195	219,239,693	1.61
Oklahoma	OK	6,776,061	3,130,132	2.16	128,165,956	75,950,637	1.69
Oregon	OR	6,304,511	2,647,637	2.38	96,007,576	56,579,822	1.70
Pennsylvania	PA	22,407,244	10,000,119	2.24	343,308,031	216,421,579	1.59
Rhode Island	RI	1,671,057	632,978	2.64	16,060,307	9,096,156	1.77
South Carolina	SC	9,019,553	4,315,832	2.09	162,688,604	100,767,894	1.61
South Dakota	SD	1,276,809	664,371	1.92	22,871,003	15,539,624	1.47
Tennessee	TN	12,943,835	5,974,292	2.17	277,827,939	159,360,395	1.74
Texas	TX	43,164,348	18,805,385	2.30	821,456,690	459,059,985	1.79
Utah	UT	4,521,535	1,683,899	2.69	58,780,268	30,433,413	1.93
Vermont	VT	1,229,261	658,356	1.87	22,970,887	15,456,947	1.49
Virginia	VA	14,289,469	5,432,099	2.63	238,670,438	125,151,714	1.91
Washington	WA	11,952,819	4,976,597	2.40	173,231,921	107,043,683	1.62
West Virginia	WV	3,513,089	1,698,465	2.07	61,324,309	37,369,639	1.64
Wisconsin	WI	9,881,693	4,816,275	2.05	167,141,424	106,939,645	1.56
Wyoming	WY	784,251	393,036	2.00	16,502,851	10,568,042	1.56

1.2.3.2 vehicleTrips LoS

Our vehicle trips used a LoS (Level of Service) with three defining characteristics. We used a departure delay (DD) of 300 seconds, a maximum occupancy of 3 passengers, and a max circuity of 0.25.



1.2.4 Transit Trips

For this analysis, we allowed for 3 different trip components: a transit trip, a aTaxi trip to the transit station, and a aTaxi trip from the transit station. We also allow for a walk to/from a transit station if the destination or origin is within the same pixel as the transit station. Our analysis assumes that everyone who can use transit transportation does so, and thus provides an upper bound on transit ridership potential. Across the 13 cities we analyzed, there was an average potential increase factor of 3.18 times the current transit ridership levels with Los Angeles having the greatest at 5.31 and St. Paul having the lowest at 1.87. When analyzing the modal split, it is apparent that one reason the current transit ridership levels are so low is that a great majority of transit trips begin or end with an aTaxi trip (85% or greater).

1.2.5 Long Trips (> 200miles)

1.2.5.1 Amtrak Trips

The mode-split model utilized to analyze Amtrak trips contained 3 sections: 1) an origin close to an Amtrak station which would be serviced by an aTaxi station; 2) an Amtrak ride to an Amtrak station closer to the passenger's destination; and 3) Amtrak station to passenger's final destination serviced also by an aTaxi station. Due to the fact that only 70% of data from the Amtrak Trips files were able to be run through the model, no reasonable comparison with current Amtrak ridership exists unfortunately.

1.2.5.2 Airline Trips

Generated Plane Data:

Number of plane trips per day	107.8 million
Plane miles travelled (Plane Miles = sum of miles travelled by each passenger)	162.23 billion miles
Busiest time	5:30 pm - 6:30 pm
Busiest Airport	Huntsville International Airport-Carl T Jones Field
Number of Busiest airports	377

Comparison with Historical Data:

	Our Long Trip Data	Bureau of Transportation Statistics
Number of Trips	107.8 million trips per day	2.57 million trips per day
Plane Miles Travelled	162.23 billion miles per day	3.306 billion miles per day

Key Differences with Historical Data:

- There are simply far too many long trips in the datasets compared to the historical data.
- The top 30 busiest airports do not match with actual data.
- We are missing important airports like JFK and ATL which are absent from our list of top 30 busiest airports.
- The states of Alabama, Alaska, Arizona, and California seem to be overrepresented in the long trip data in comparison with historical data.
- Alabama in particularly was anomalous. Birmingham, AL with a population of about 200,000 produced around 140,000 long trips out of the area per day.

Chapter 2:
Short Trips

The hundreds of millions of trips that people in the United States take every day vary wildly in terms of distance traveled; it is imperative to understand the stark difference between how autonomous taxi service should assist someone traveling hundreds of miles versus someone traveling only hundreds of feet. This chapter delves deeper into those trips which are the shortest, detailing how these short trips are distributed across the United States and ultimately looking into the potential benefits of specialized autonomous taxis servicing short trips.

Jeb Helmers and Gavin Zhu

Section 2.1: Overview of Short Trips

Once all of the person trips are simulated and the individual trip data is compiled, these person trips can then be subsequently categorized based on the Great Circle Distance of each trip. This chapter will focus on the shortest two categories of these trips: walking trips and short haul trips.

Individual person trips which are a distance of less than 0.5 miles (exclusive) are characterized as walking trips, based on the Great Circle Distance from their person trip records. Ultimately, as the name suggests, these trips are assumed to not be worth the costs of aTaxi service; people would prefer to simply walk the extremely short distance rather than wait for an aTaxi to pick them up. This distance is justified by the assumption that people walking 0.5 miles (which is the distance from one pixel to the next) would find the process of walking from the centroid of one pixel to be picked up, waiting for an aTaxi to depart, and then walking to their final destination (all to move less than one pixel away) too cumbersome.

Next, individual person trips which are a distance of greater than 0.5 miles (inclusive) and less than 2 miles (exclusive) are characterized as short haul Trips. These trips will still be serviced by aTaxis, but these aTaxis will specialize in short haul trips which are centered close by the originating pixel, meaning that all potential destinations are relatively nearby. This allows for more freedom when picking passengers, so that potential passengers will not be denied for being too far away (or violating the Max Circuity condition). The upper bound of 2 miles is a bit more arbitrary; it was assumed that trips longer than 2 miles had too much potential to substantially increase the distance aTaxis had to travel, though this can be adjusted in future analysis.

Trips longer than 2 miles will be further broken down and discussed in the forthcoming chapters.

Section 2.2: The Distribution of Short Trips in the U.S.

As previously stated, there are two types of short trips, walking trips and short haul trips. In this section, the distribution of both types of short trips across the different states in the U.S. (including Washington D.C.) is analyzed. This was done by iterating through all of the person trip files and categorizing each person trip as either a walking trip, short haul trip, or normal aTaxi trip (>2 miles), and storing which state the trip originated from in a 51 X 1 array. This process was performed on all trips less than 200 miles long. In total, there were 3,596 person trip files (about 121 GB) analyzed, and there were about 870 million trips analyzed totaling about 10.4 billion trip miles.

In America, there were a little over 22.8 million walking trips, and roughly 6.9 million total walking trip miles. This means that about 2.6% of all trips less than 200 miles were walking trips, and 0.07% of all trip miles (for trips less than 200 miles) were walking trip miles. This difference in proportions is explained by the fact that walking trips are significantly shorter than the longer trips, especially the normal aTaxi trips which can reach up to 200 miles long. In terms of a statewide breakdown, California accounted for 19.9% of total walking trips and 18.7% of total walking trip miles. California had by far the largest proportion of the total walking trips and walking trip miles, with New York and Texas accounting for the next most. In contrast, Vermont had very few walking trips, only about 27,000. A breakdown of the walking trips by the state in which they originate is shown in Figure 2.1:

State	State Code	FIPS code	walkingTrips	walkingTripMiles	walkingTrips%	walkingTripsMiles%
Alabama	AL	1	157,223	48,334	0.7%	0.7%
Alaska	AK	2	62,058	18,530	0.3%	0.3%
Arizona	AZ	4	600,557	190,543	2.6%	2.8%
Arkansas	AR	5	101,878	32,493	0.4%	0.5%
California	CA	6	4,544,965	1,287,474	19.9%	18.7%
Colorado	CO	8	446,334	138,826	2.0%	2.0%
Connecticut	CT	9	216,972	66,785	1.0%	1.0%
Delaware	DE	10	45,616	14,486	0.2%	0.2%
District of Columbia	DC	11	118,825	27,456	0.5%	0.4%
Florida	FL	12	801,995	262,336	3.5%	3.8%
Georgia	GA	13	344,054	114,264	1.5%	1.7%
Hawaii	HI	15	104,329	32,206	0.5%	0.5%
Idaho	ID	16	137,226	43,395	0.6%	0.6%
Illinois	IL	17	1,461,827	395,519	6.4%	5.8%
Indiana	IN	18	344,978	110,892	1.5%	1.6%
Iowa	IA	19	283,437	88,137	1.2%	1.3%
Kansas	KS	20	281,502	87,836	1.2%	1.3%
Kentucky	KY	21	188,705	59,877	0.8%	0.9%
Louisiana	LA	22	215,825	68,976	0.9%	1.0%
Maine	ME	23	49,955	15,666	0.2%	0.2%
Maryland	MD	24	302,254	97,402	1.3%	1.4%
Massachusetts	MA	25	467,957	141,879	2.1%	2.1%
Michigan	MI	26	568,778	181,147	2.5%	2.6%
Minnesota	MN	27	365,353	114,614	1.6%	1.7%
Mississippi	MS	28	107,059	34,617	0.5%	0.5%
Missouri	MO	29	350,405	111,975	1.5%	1.6%
Montana	MT	30	90,660	27,259	0.4%	0.4%
Nebraska	NE	31	215,771	65,536	0.9%	1.0%
Nevada	NV	32	204,141	64,474	0.9%	0.9%
New Hampshire	NH	33	54,291	16,982	0.2%	0.2%
New Jersey	NJ	34	751,112	218,173	3.3%	3.2%
New Mexico	NM	35	141,879	45,716	0.6%	0.7%
New York	NY	36	2,006,210	539,802	8.8%	7.9%
North Carolina	NC	37	264,245	87,453	1.2%	1.3%
North Dakota	ND	38	70,500	21,119	0.3%	0.3%
Ohio	OH	39	619,735	195,992	2.7%	2.9%
Oklahoma	OK	40	262,250	82,697	1.2%	1.2%
Oregon	OR	41	313,002	99,348	1.4%	1.4%

Pennsylvania	PA	42	736,008	221,484	3.2%	3.2%
Rhode Island	RI	44	75,203	22,483	0.3%	0.3%
South Carolina	SC	45	122,594	40,609	0.5%	0.6%
South Dakota	SD	46	86,458	26,150	0.4%	0.4%
Tennessee	TN	47	221,916	72,591	1.0%	1.1%
Texas	TX	48	1,881,888	601,305	8.3%	8.8%
Utah	UT	49	277,527	88,692	1.2%	1.3%
Vermont	VT	50	27,465	8,402	0.1%	0.1%
Virginia	VA	51	664,959	214,919	2.9%	3.1%
Washington	WA	53	509,253	157,519	2.2%	2.3%
West Virginia	WV	54	86,637	26,373	0.4%	0.4%
Wisconsin	WI	55	391,115	119,982	1.7%	1.7%
Wyoming	WY	56	58,402	18,036	0.3%	0.3%
Total			22,803,288	6,868,765		

Figure 2.1

Below is Figure 2.2, a heat map representing each state's proportion of walking trips, where once again it is clear just how concentrated walking trips are in the few aforementioned states.

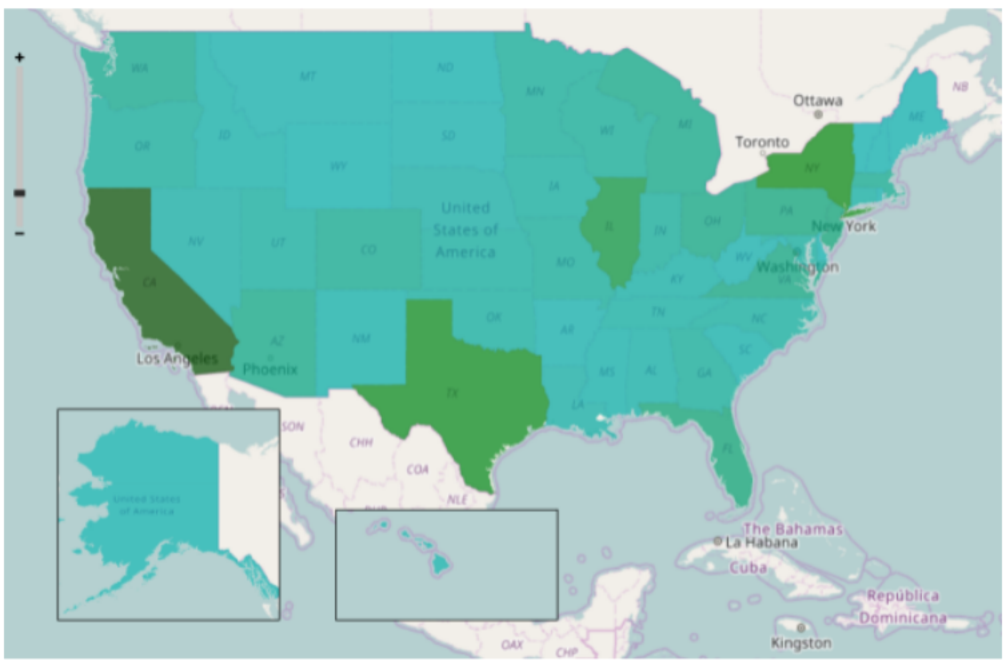


Figure 2.2

WalkingTrips%	0.100	9.05	18.0

Next, analysis was done on the proportion of each state's total trips that were walking trips, as opposed to other types of trips (less than 200 miles). Interestingly, of all regions in the United States, Washington DC had clearly the highest proportion of walking trips out of the total trips in its region. Not surprisingly, New York was in the top ten, with the 9th highest ratio of walk trips to total trips. The 5 regions with the highest ratios were DC, NE, SD, ND, and WY. A graph of each region's ratio is shown below in Figure 2.3, where the y-axis shows what percent of the state's trips are walking trips:

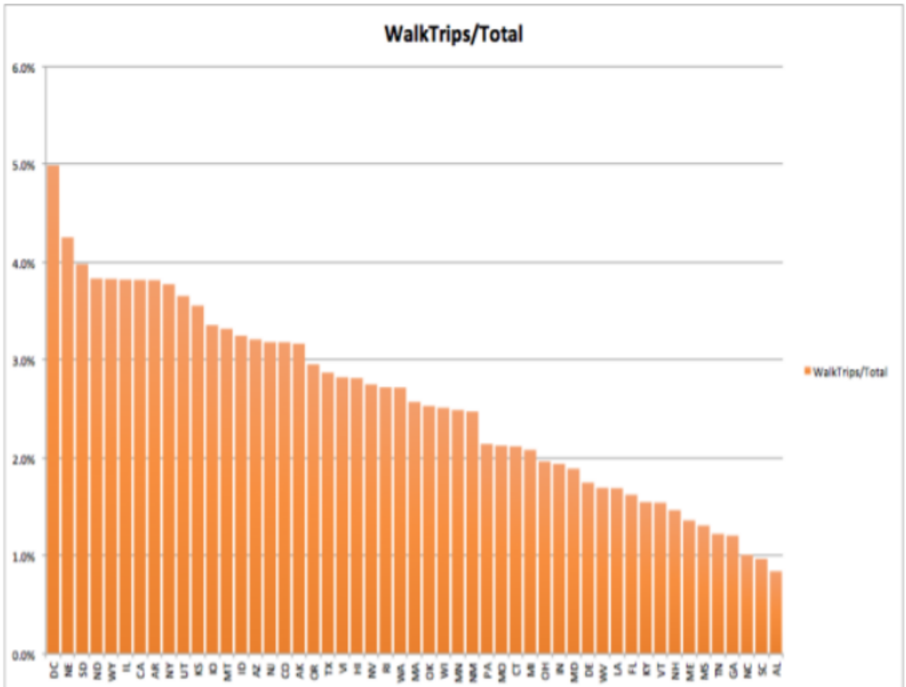


Figure 2.3

Similar analysis was performed for short haul trips. In total, there were about 282.3 million short haul trips and 305.5 million short haul trip miles, meaning that about 32.5% of all trips less than 200 miles were short haul trips, and 3% of all trip miles (for trips less than 200 miles) were short haul trip miles. Again, the disparity is explained by how much longer normal aTaxi trips are.

Not surprisingly, California also had the largest proportion of short haul trips originating there. However, it did not have as large a proportion as it did for walk trips. Only 13.9% of short haul trips and 13.7% of short haul trip miles originated from California. Once again, the fewest number of trips (about 550,000) originated from Vermont. An overall statewide distribution is displayed in Figure 2.4:

State	State Code	FIPS code	shortHaulTrips	shortHaulTripMiles	shortHaulTrips%	shortHaulTripMiles%
Alabama	AL	1	3,710,220	4,167,435	1.3%	1.4%
Alaska	AK	2	689,936	755,347	0.2%	0.2%
Arizona	AZ	4	5,185,789	5,652,510	1.8%	1.8%
Arkansas	AR	5	809,987	886,384	0.3%	0.3%
California	CA	6	39,281,485	42,283,737	13.9%	13.7%
Colorado	CO	8	4,885,357	5,345,514	1.7%	1.7%
Connecticut	CT	9	3,370,813	3,695,673	1.2%	1.2%
Delaware	DE	10	827,437	907,373	0.3%	0.3%
District of Columbia	DC	11	795,471	882,976	0.3%	0.3%
Florida	FL	12	15,640,857	17,338,743	5.5%	5.6%
Georgia	GA	13	8,229,801	9,299,847	2.9%	3.0%
Hawaii	HI	15	1,335,708	1,426,896	0.5%	0.5%
Idaho	ID	16	1,672,255	1,797,996	0.6%	0.6%
Illinois	IL	17	12,033,913	12,874,254	4.3%	4.2%
Indiana	IN	18	5,895,699	6,460,920	2.1%	2.1%
Iowa	IA	19	3,188,203	3,315,713	1.1%	1.1%
Kansas	KS	20	2,933,151	3,095,559	1.0%	1.0%
Kentucky	KY	21	3,729,900	4,162,909	1.3%	1.4%
Louisiana	LA	22	4,171,539	4,603,869	1.5%	1.5%
Maine	ME	23	1,076,936	1,159,863	0.4%	0.4%
Maryland	MD	24	5,318,295	5,876,001	1.9%	1.9%
Massachusetts	MA	25	6,381,237	6,913,485	2.3%	2.2%
Michigan	MI	26	8,448,057	9,246,561	3.0%	3.0%
Minnesota	MN	27	5,018,389	5,389,303	1.8%	1.8%
Mississippi	MS	28	2,307,052	2,541,271	0.8%	0.8%
Missouri	MO	29	5,432,312	5,896,730	1.9%	1.9%
Montana	MT	30	1,029,917	1,089,954	0.4%	0.4%
Nebraska	NE	31	1,956,069	2,059,786	0.7%	0.7%
Nevada	NV	32	2,298,009	2,538,547	0.8%	0.8%
New Hampshire	NH	33	1,161,782	1,269,753	0.4%	0.4%
New Jersey	NJ	34	8,789,460	9,442,935	3.1%	3.1%
New Mexico	NM	35	1,986,262	2,168,553	0.7%	0.7%
New York	NY	36	19,479,038	20,921,909	6.9%	6.8%
North Carolina	NC	37	7,339,717	8,295,141	2.6%	2.7%
North Dakota	ND	38	746,703	781,732	0.3%	0.3%
Ohio	OH	39	10,064,119	11,040,558	3.6%	3.6%
Oklahoma	OK	40	3,330,283	3,563,202	1.2%	1.2%
Oregon	OR	41	3,977,493	4,282,214	1.4%	1.4%
Pennsylvania	PA	42	11,236,259	12,194,353	4.0%	4.0%
Rhode Island	RI	44	1,018,381	1,123,195	0.4%	0.4%
South Carolina	SC	45	3,519,534	3,948,779	1.2%	1.3%
South Dakota	SD	46	810,400	844,791	0.3%	0.3%
Tennessee	TN	47	4,980,583	5,616,969	1.8%	1.8%
Texas	TX	48	20,520,154	22,332,727	7.3%	7.3%
Utah	UT	49	2,799,941	3,045,076	1.0%	1.0%
Vermont	VT	50	525,520	557,369	0.2%	0.2%
Virginia	VA	51	8,614,073	9,565,842	3.1%	3.1%
Washington	WA	53	6,279,229	6,785,901	2.2%	2.2%
West Virginia	WV	54	1,519,871	1,657,228	0.5%	0.5%
Wisconsin	WI	55	5,309,838	5,691,935	1.9%	1.9%
Wyoming	WY	56	683,775	729,359	0.2%	0.2%
			282,346,209	307,524,675		

Figure 2.4

Below in Figure 2.5 is a heat map of each state's proportion of total short haul trips, demonstrating how short haul trips are similarly distributed to walking trips in the U.S.

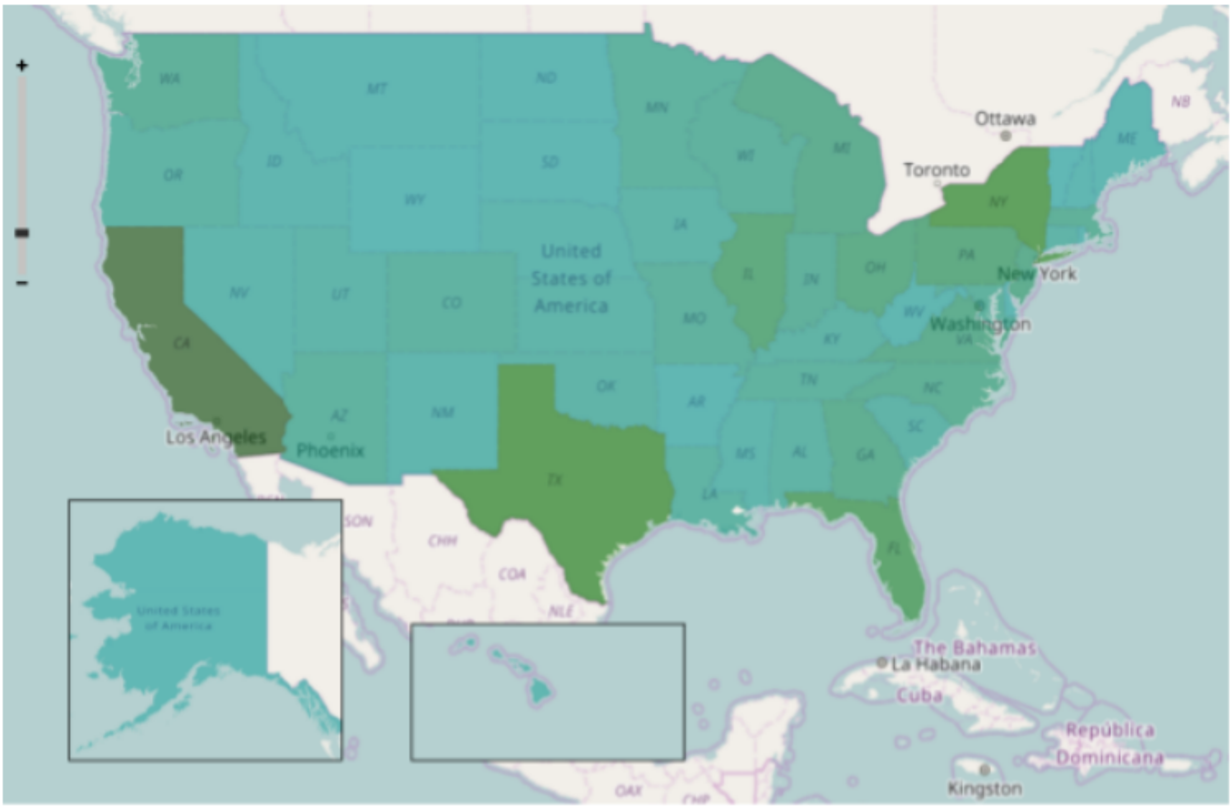


Figure 2.5

ShortHaulTrips% 0.200 7.10 14.0

Again for short haul trips, analysis was performed on which states/regions had the highest ratio of short haul trips to total trips for their respective regions. For short haul trips, Wyoming had the highest ratio, which makes sense because it had very few walk trips. A breakdown of each state's ratio is below in Figure 2.6, where the y-axis shows what percentage of the total trips are short haul trips:

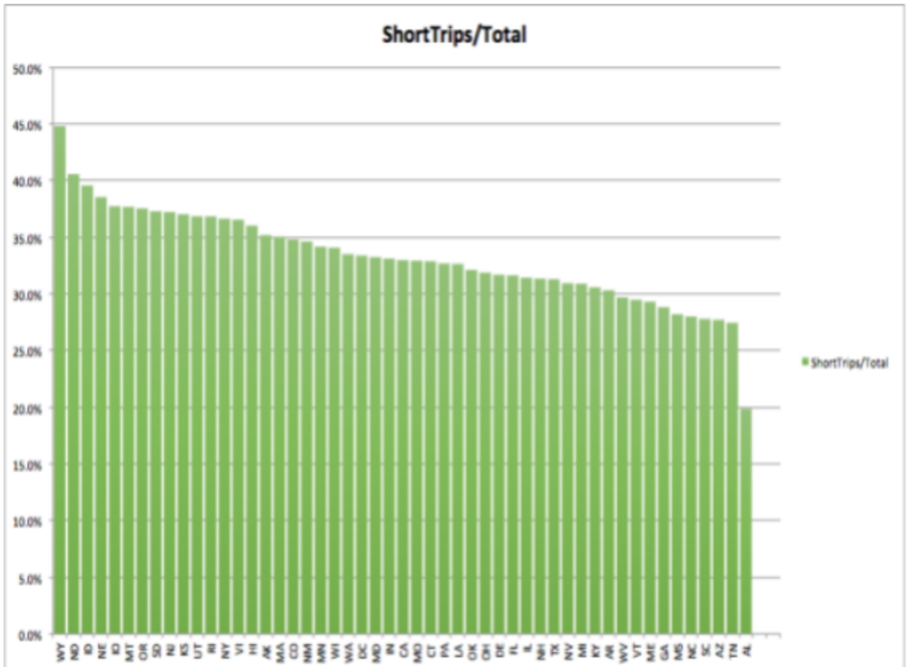


Figure 2.6

Section 2.3: Superpixel Analysis of Short Trips

Initially, the Person Trips are split up based on a pixilation of the United States into 0.5 mile by 0.5 mile pixels. aTaxis depart from and arrive at the center of these pixels, but in order to make the data more manageable, we split the United States into "superpixels" that are 10 times the size of the of the original pixels (5 miles by 5 miles). With help from Kyle Marocchini, the range of pixel values for the United States was determined to be:

MinX=-11018, MaxX=941
MinY=-2756, MaxY=4481

Once the max and min values of the pixels, two-dimensional arrays were created with each entry in the arrays representing every single possible superpixel location, and the number of trips originating from a pixel in that superpixel. These superpixels would have 1/10th of the range that the pixels have (ranging from -1102 to 94 for X and -276 to 448 for Y).

After this was determined, the trips were analyzed based on where they originate from. Using MATLAB code, all the trip files were iterated through and incremented the value of whichever superpixel they were contained within, which was performed separately for both walking trips and short haul trips.

33,904 superpixels had at least one walking trip originating in it, with many more superpixels being completely empty, representing bodies of water and other unpopulated or sparsely populated areas. The maximum number of walking trips in any single superpixel is about 247,000, located in

Manhattan (New York City). However, the vast majority of these superpixels did not have many walking trips originating in it, leading to the very skewed distribution shown in Figure 2.7:

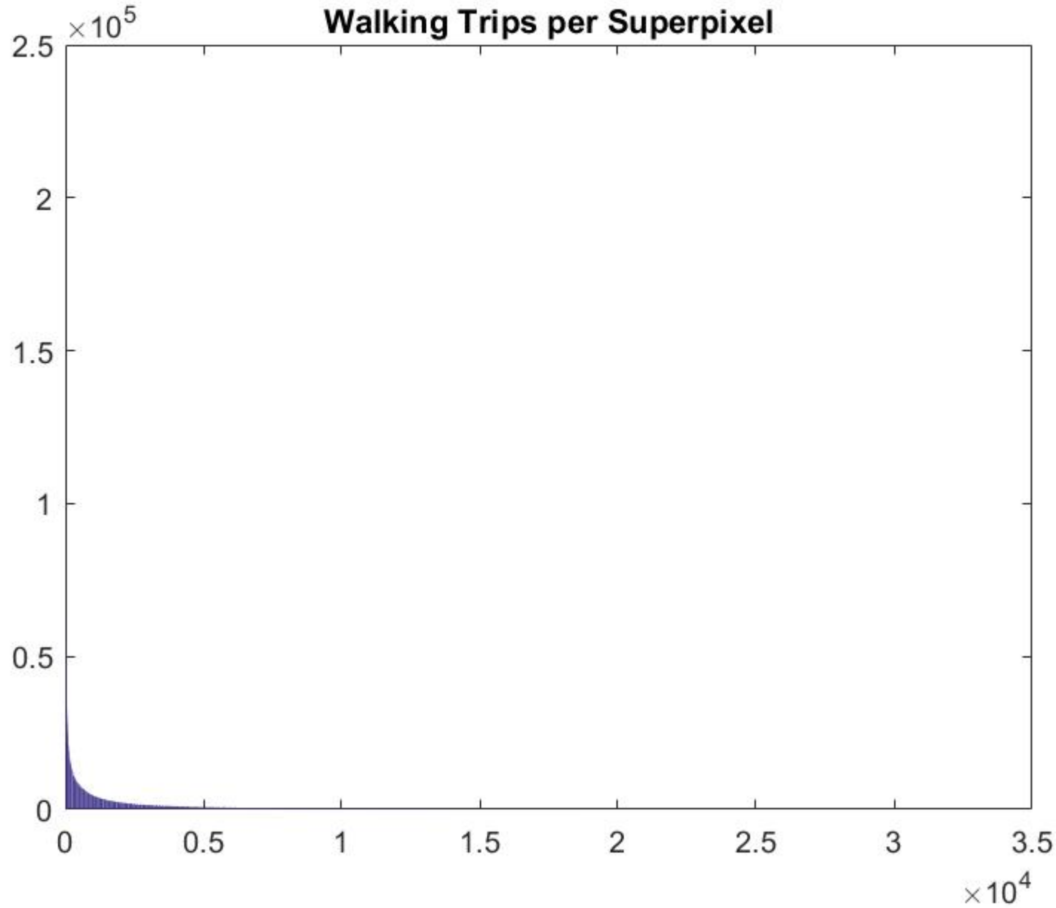


Figure 2.7

By zooming on the section of the graph with the 2000 most active superpixels, it is clear that there are still few superpixels with a significant amount of walking trips. This could be explained by the assertion that walking trips are concentrated primarily in the most urban areas. Figure 2.8 below zooms in on the 2000 most active superpixels.

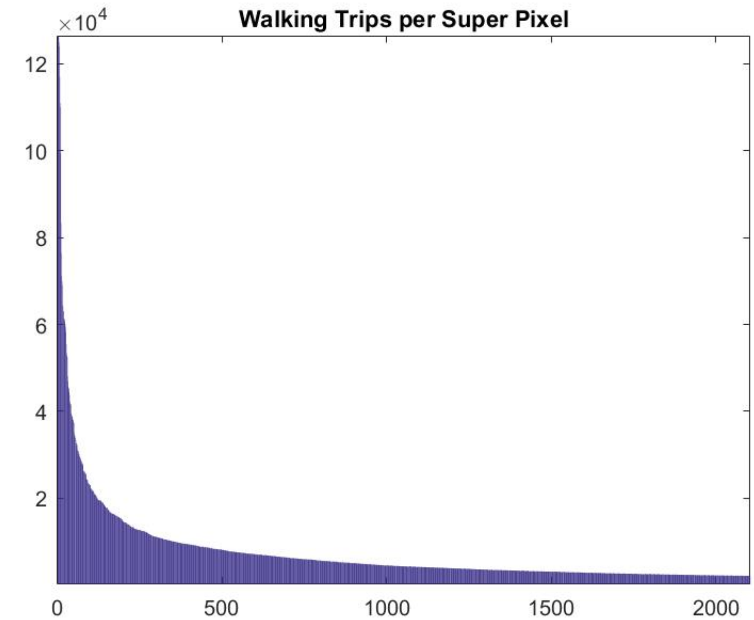


Figure 2.8

Next, a cumulative distribution (cdf) of the walking trips in superpixels is constructed to better see the distribution amongst superpixels. The first two thousand superpixels constitute about 70% of all walking trips, and the first 10,000 superpixels constitute about 90% of all walking trips. The cdf of walking trips is shown below in Figure 2.9.

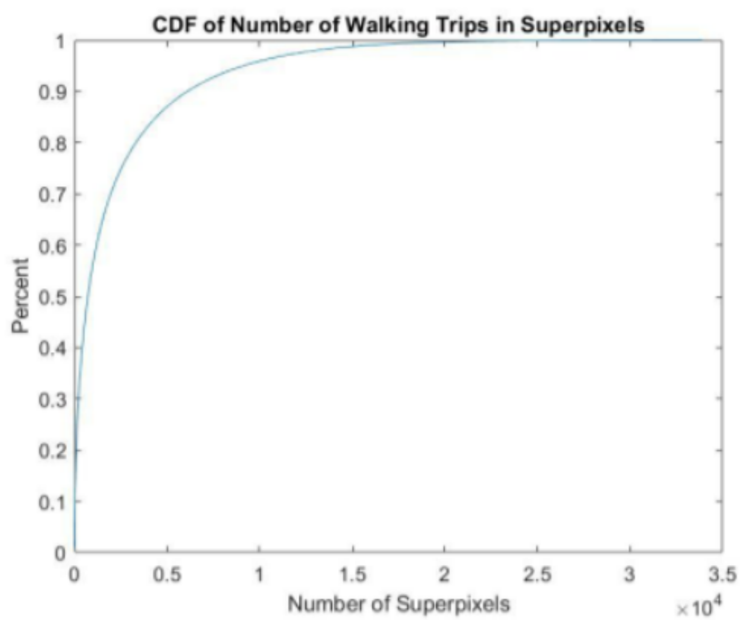


Figure 2.9

Finally, the superpixel data can be used to create heat maps of how walking trips are distributed across the United States. Since the data is so skewed, two different heat maps are created, one with a linear scale and one with a logarithmic scale.

The heat map with a linear scale is presented to show the uneven distribution of walking trips in the United States. Though the range of superpixels is split into thirds, from red to yellow to green, very few of the superpixels are within the upper third of the range as demonstrated by the little amount of green within the heat map (Figure 2.10).

Figure 2.10

The abundance of yellow makes it a bit difficult to understand the distribution, so in order to better visualize the walking trips, a logarithmic heat map is presented, first with Alaska and Hawaii (Figure 2.11), and then zoomed in on the continental United States (Figure 2.12). Once again it is clear that the majority of trips are concentrated in urban areas, but now it easier to observe where other pockets of walking trips are located in the United States. Note that the scale on the axes is purely positive numbers, the actual super pixel values should range from -1200 to 100 on the X axis, and -300 to 500 on the Y axis.

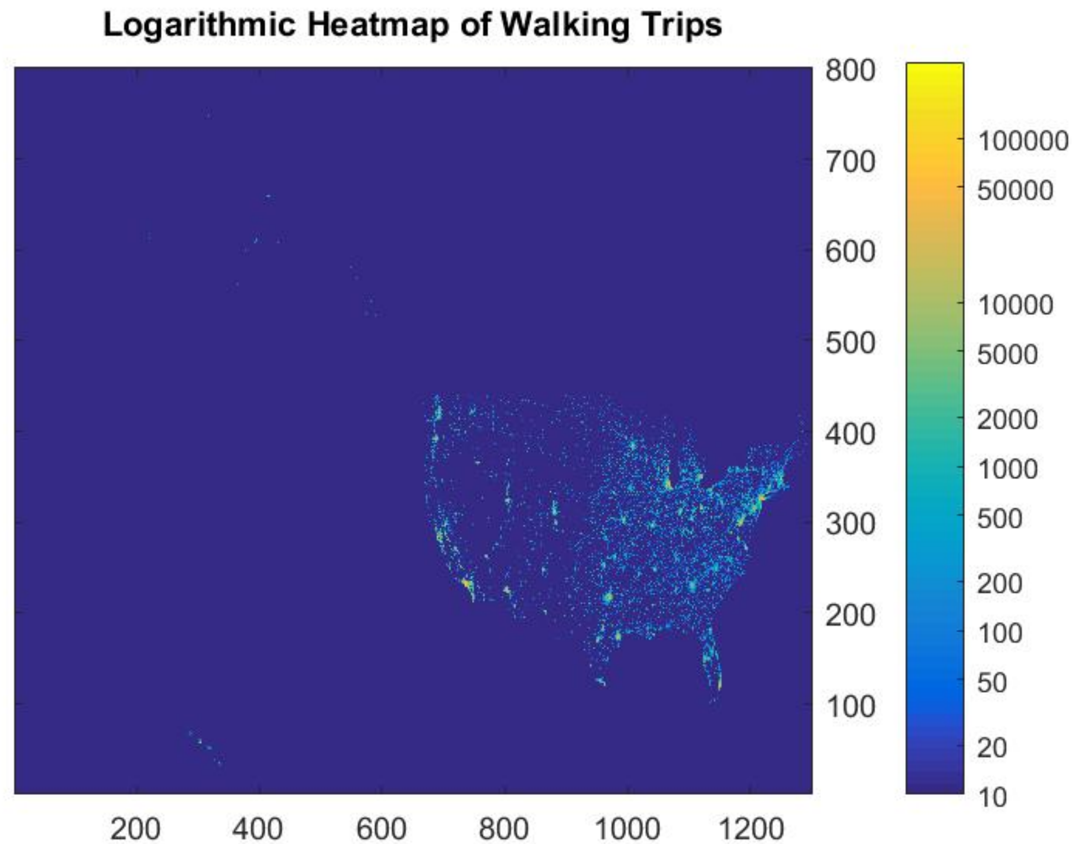


Figure 2.11

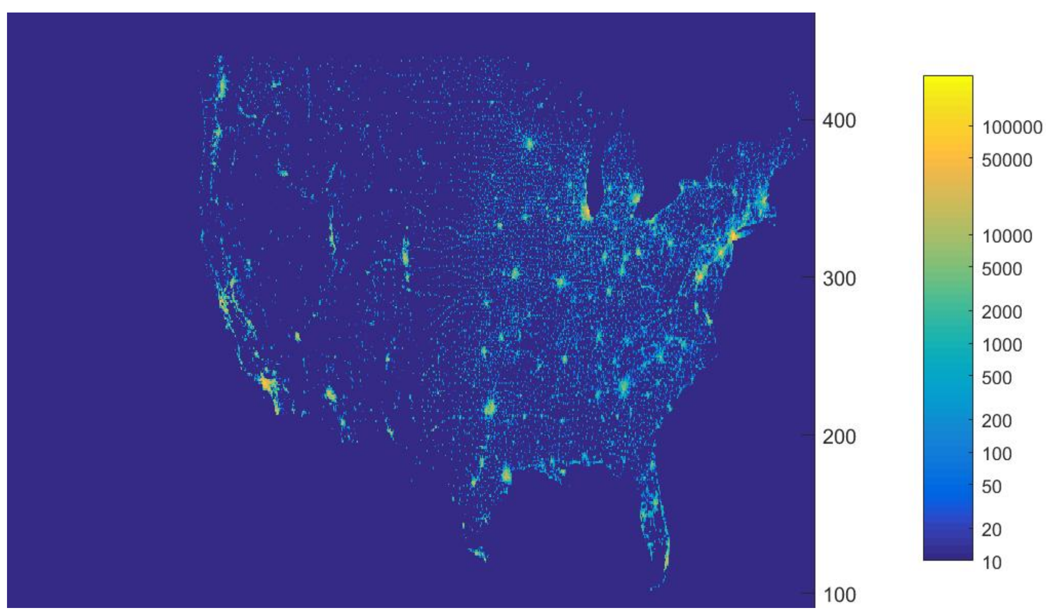


Figure 2.12

For the short haul trips, 67,592 superpixels had at least one short haul trip originating in it, again with many more superpixels being completely empty, representing bodies of water and other unpopulated or sparsely populated areas. The maximum number of trips from any single superpixel being about 1.9 million, also located in New York City. Once again, the vast majority of relevant superpixels had relatively few trips.

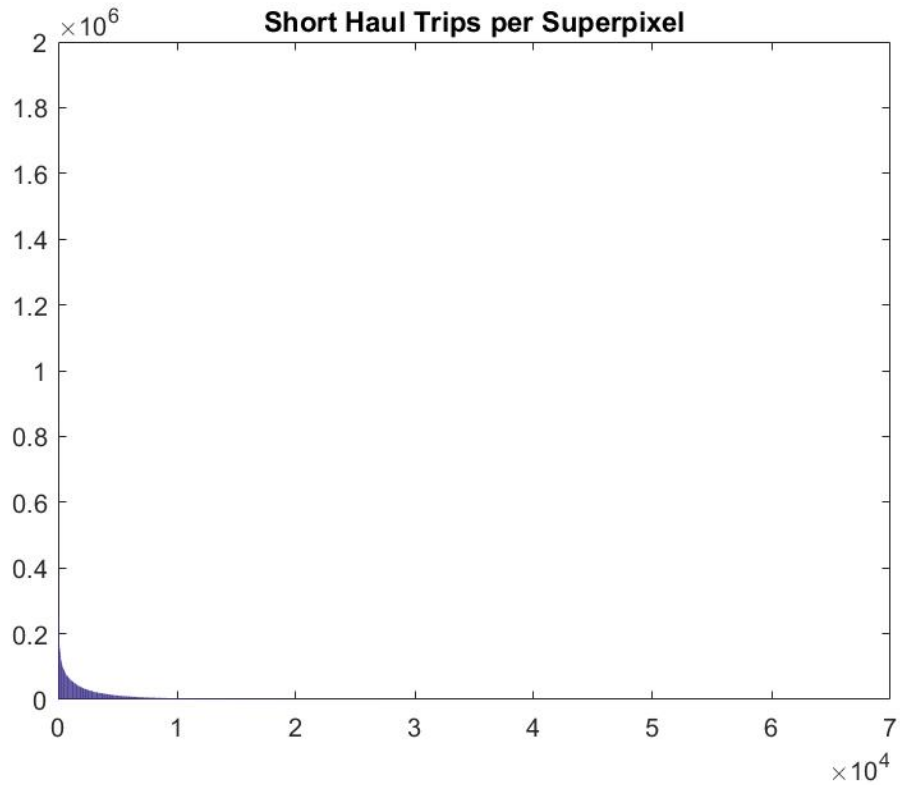


Figure 2.13

In contrast to the walking trips, when zooming in on the 2000 superpixels with the most trips (Figure 2.14), it is clear that short haul trips are more evenly distributed amongst the most active superpixels than for walking trips. This implies that short haul trips are less concentrated in the most urban areas than walking trips.

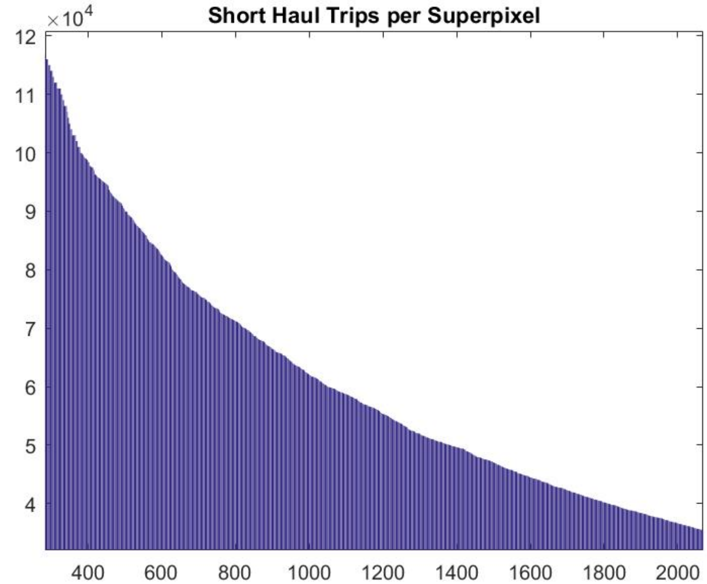


Figure 2.14

A cumulative distribution of the short haul trips was also constructed and is presented below in Figure 2.15. Once again, the majority of the short haul trips in the U.S. are concentrated in a relatively small proportion of the superpixels, with only about 10,000 superpixels containing 90% of the short haul trips.

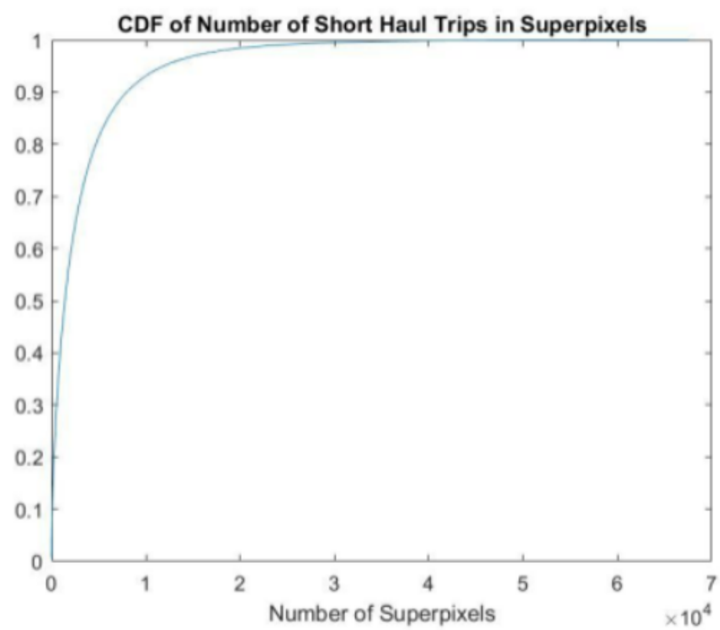


Figure 2.15

Similar to walking trips, heat maps are constructed for linear and logarithmic scales. From the linear heat map in Figure 2.16, it is clear that very few of the trips fall within the upper third of the range, colored green. It is also clear that the distribution of the middle range of superpixels is much more widespread than for walking trips.

Figure 2.16

From the logarithmic heat maps in Figures 2.17 and 2.18, it is once again clear that there is a stark difference between the amount of walking trips and short haul trips, but the two trip types are still concentrated in similar areas. There are much more areas of highly active superpixels for short haul trips, including smaller cities than the walking trips. Again, the actual super pixel values should range from -1200 to 100 on the X axis, and -300 to 500 on the Y axis.

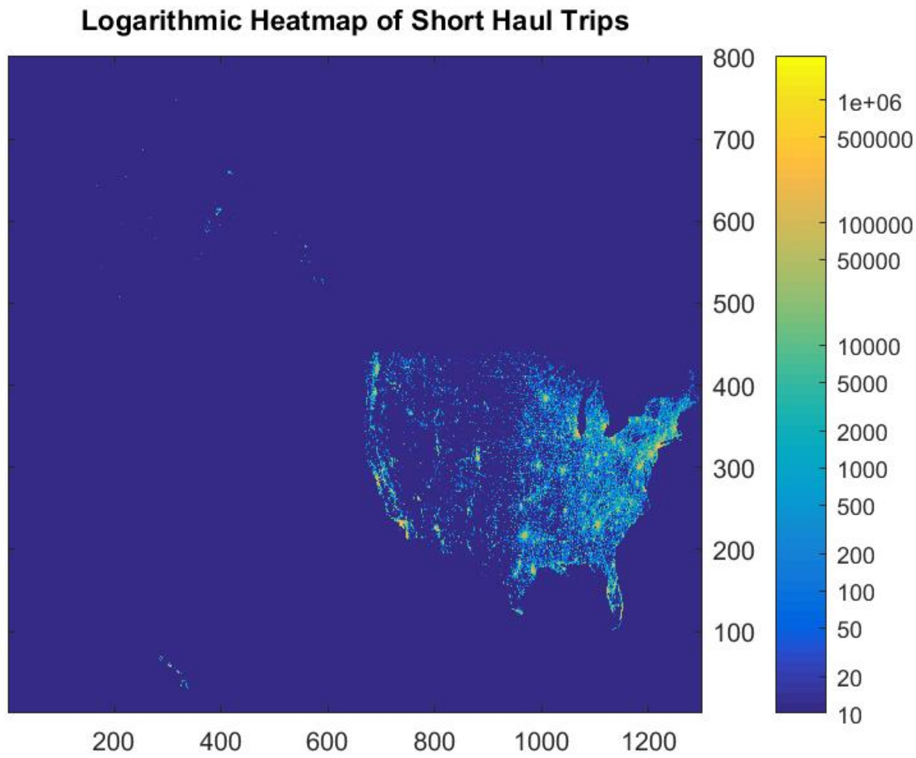


Figure 2.17

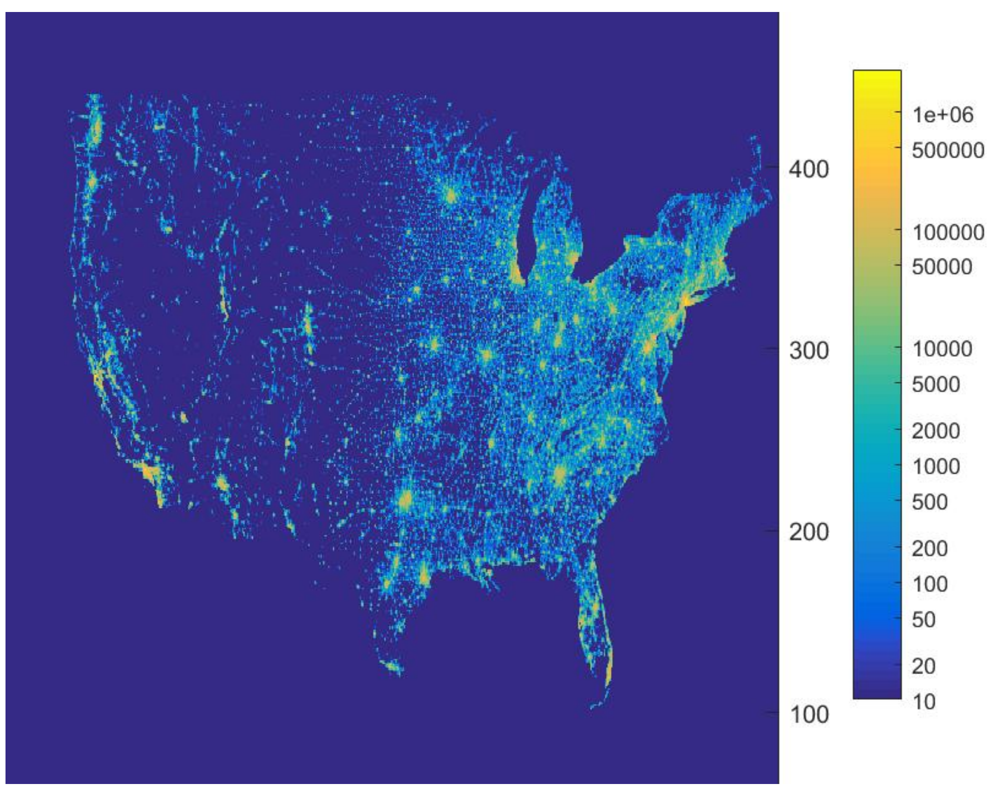


Figure 2.18

Section 2.4: AVO Analysis of Short Haul Trips

After aggregating the data on short haul trips, it is now possible to perform ride-sharing analysis on each state and simulate how aTaxis would pick up and drop off passengers looking to travel between 0.5 and 2 miles. The difference between this ride share analysis and those performed previously is that it is assumed that these aTaxis will have an infinite capacity of passengers and will not reject passengers based on potential added travel distance (Max Circuity). With this set of data and assumptions, each set of person trips that originate from each state is analyzed to determine the number of Vehicle Miles Traveled (distance the aTaxis travel, VMT), Person Trip Miles (distance that the passengers desired to travel), and Average Vehicle Occupancy (Person Trip Miles / Vehicle Miles Traveled) for each state and D.C.

The results from the analysis are compiled below in Table 2.19, which includes the number of aTaxi trips were made, as well as the AVO of the short haul trips in each state. The AVOs ranged from close to 1 (Alaska) to a bit over 2 (Hawaii), with most falling within the 1.2 to 1.5 range. Hawaii stands out as an outlier, which might be due to the innate clustering of its islands, leading to access short haul trips with greater efficiency and demand. Other states with high AVOs include Arkansas, Washington D.C., and Texas.

The fact that the AVOs for short haul aTaxi service is rather high in every state implies that short haul aTaxi service would be beneficial for society, as the specialized ridesharing service

However, it is important to remember that the assumptions that we have made during this ride share analysis, such as the infinite capacity aTaxi fleet, may substantially influence the AVOs that are obtained. To further determine how beneficial short haul aTaxi service truly is, different sets of assumptions should be tested.

State	Trip Count	Vehicle Miles	Person Trip Miles	AVO
Alabama	1,311,388	2.98E+06	3.71E+06	1.244
Alaska	220,156	6.80E+05	6.90E+05	1.015
Arizona	1,303,476	3.59E+06	5.19E+06	1.446
Arkansas	192,889	4.23E+05	8.10E+05	1.916
California	5,466,289	2.45E+07	3.93E+07	1.603
Colorado	1,319,764	3.64E+06	4.89E+06	1.341
Connecticut	981,573	2.62E+06	3.37E+06	1.285
Delaware	236,666	6.23E+05	8.27E+05	1.329
D.C.	120,621	4.12E+05	7.95E+05	1.933
Florida	4,214,287	1.12E+07	1.56E+07	1.394
Georgia	2,612,384	6.33E+06	8.23E+06	1.300
Hawaii	247,718	6.12E+05	1.34E+06	2.181
Idaho	491,286	1.27E+06	1.67E+06	1.315
Illinois	3,092,814	8.72E+06	1.20E+07	1.379
Indiana	1,784,803	4.53E+06	5.90E+06	1.301
Iowa	945,635	2.27E+06	3.19E+06	1.407
Kansas	846,751	2.10E+06	2.93E+06	1.398
Kentucky	1,215,203	2.92E+06	3.73E+06	1.276
Louisiana	1,261,046	3.04E+06	4.17E+06	1.371
Maine	413,469	8.90E+05	1.08E+06	1.210
Maryland	1,340,227	3.77E+06	5.32E+06	1.411
Massachusetts	1,694,822	4.68E+06	6.38E+06	1.363
Michigan	2,716,000	7.04E+06	8.45E+06	1.199
Minnesota	1,503,189	3.88E+06	5.02E+06	1.294
Mississippi	791,733	1.72E+06	2.31E+06	1.344
Missouri	1,651,709	4.11E+06	5.43E+06	1.323
Montana	330,823	7.97E+05	1.03E+06	1.292
Nebraska	534,264	1.38E+06	1.96E+06	1.419
Nevada	582,548	1.69E+06	2.30E+06	1.360
New Hampshire	414,741	9.50E+05	1.16E+06	1.223
New Jersey	2,125,269	6.19E+06	8.79E+06	1.420
New Mexico	565,451	1.43E+06	1.99E+06	1.385
New York	4,081,506	1.18E+07	1.95E+07	1.650
North Carolina	2,517,417	5.98E+06	7.34E+06	1.227
North Dakota	222,084	5.54E+05	7.47E+05	1.347
Ohio	3,083,748	8.02E+06	1.01E+07	1.254
Oklahoma	981,930	2.39E+06	3.33E+06	1.396
Oregon	1,092,596	3.03E+06	3.98E+06	1.311
Pennsylvania	3,261,535	8.45E+06	1.12E+07	1.330
Rhode Island	254,757	7.47E+05	1.02E+06	1.364
South Carolina	1,221,386	2.82E+06	3.52E+06	1.250
South Dakota	255,742	5.96E+05	8.10E+05	1.361
Tennessee	1,733,363	4.06E+06	4.98E+06	1.226
Texas	4,122,999	1.07E+07	2.05E+07	1.918
Utah	689,928	1.96E+06	2.80E+06	1.426
Vermont	205,135	4.21E+05	5.26E+05	1.250
Virginia	1,909,818	5.11E+06	8.61E+06	1.685
Washington	1,752,690	4.93E+06	6.28E+06	1.274
West Virginia	488,271	1.11E+06	1.52E+06	1.371
Wisconsin	1,685,563	4.25E+06	5.31E+06	1.250
Wyoming	206,112	4.97E+05	6.84E+05	1.376
Total	72,295,574	1.98E+08	2.82E+08	1.423

Figure 2.19 (Total below)

Below, in Figure 2.20, is a color map by state of AVO, giving a visual representation of how the short haul AVOs look across the United States.

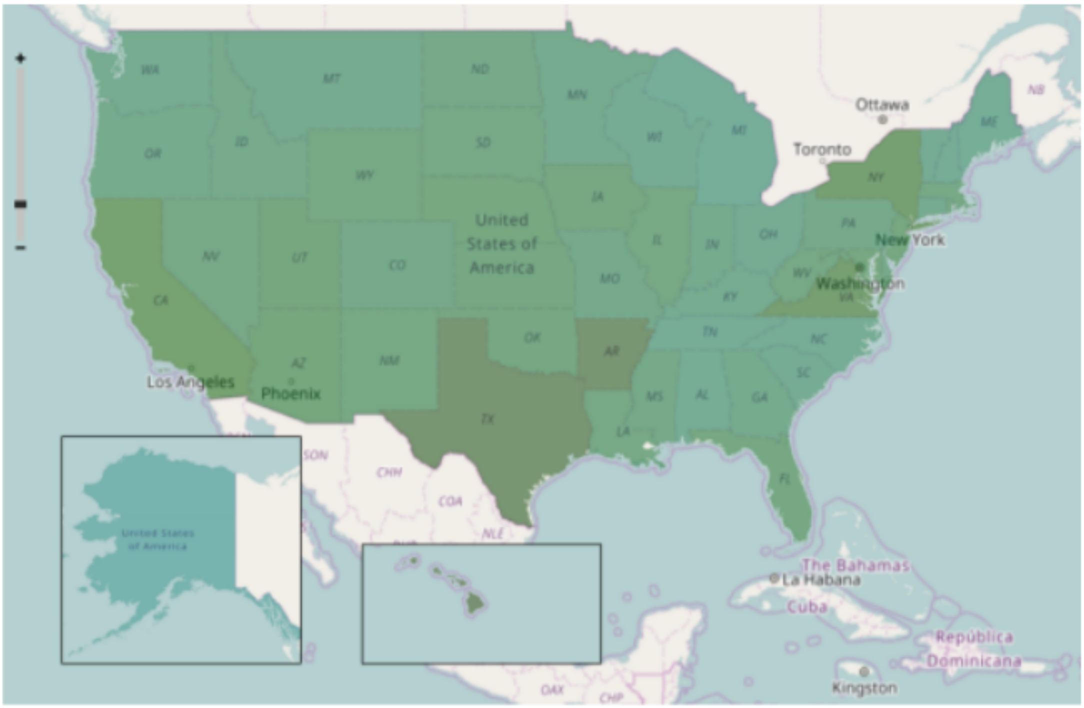


Figure 2.20

AVO	1.00	1.55	2.10

Finally, in Table 2.21, included is the breakdown of how many passengers were in each aTaxi amongst the states, including a category for aTaxis with 15 or more passengers. The majority of aTaxis tend to contain between 1 to 5 passengers, with a steady drop off after that. There is a general increase from the number of aTaxis with 2 passengers to those with 3 passengers, implying either that 3 passenger aTaxis might be an ideal capacity (or close to it), or perhaps that the analysis is skewed in some fashion. Either way, it is clear that aTaxis do not need to be infinitely large to be beneficial, many beneficial trips can be serviced with smaller aTaxis.

# of Passengers:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15+
Alabama	490,004	217,628	322,123	143,863	60,811	27,374	13,652	7,596	4,476	3,088	2,329	1,815	1,480	1,205	13,944
Alaska	71,418	33,807	57,526	26,593	12,339	6,200	3,400	2,002	1,315	879	666	518	361	322	2,810
Arizona	256,784	160,099	414,140	223,821	106,814	51,746	26,227	14,372	8,580	5,680	3,898	2,940	2,238	1,887	24,250
Arkansas	71,708	29,246	38,296	18,841	9,392	5,297	3,283	2,141	1,692	1,192	1,001	803	770	635	8,592
California	1,337,450	753,462	1,743,157	815,835	355,395	163,421	80,715	44,416	25,987	17,672	12,467	9,795	7,747	6,727	92,044
Colorado	284,812	169,391	400,617	212,654	104,648	52,529	28,495	16,177	10,130	6,547	4,553	3,418	2,622	2,068	21,103
Connecticut	265,293	145,226	270,269	133,144	65,201	34,218	19,133	11,311	7,283	5,095	3,553	2,735	2,039	1,717	15,356
Delaware	63,928	33,519	66,912	33,588	15,739	7,775	4,142	2,320	1,482	1,112	765	602	475	419	3,888
D.C.	5,640	4,891	27,670	21,083	14,184	9,711	6,898	5,067	3,786	2,946	2,394	1,969	1,572	1,330	11,480
Florida	921,647	559,765	1,251,857	676,052	336,241	171,769	91,550	52,051	31,226	20,003	13,584	9,905	7,647	6,066	64,924
Georgia	833,930	435,059	697,632	315,192	136,026	68,634	32,218	21,193	11,066	8,865	5,477	5,012	3,379	3,311	35,390
Hawaii	49,221	28,986	44,530	32,720	22,841	16,145	11,577	8,326	6,114	4,699	3,499	3,066	2,123	1,944	11,927
Idaho	139,615	66,424	132,181	68,688	33,636	17,588	9,529	5,828	3,563	2,301	1,727	1,264	941	732	7,269
Illinois	626,662	372,351	969,875	521,126	252,024	124,465	65,924	37,125	22,694	14,902	10,628	7,970	6,351	5,137	55,580
Indiana	515,827	258,083	491,738	246,953	116,131	56,780	29,677	16,744	10,052	6,550	4,614	3,308	2,726	2,202	23,418
Iowa	273,001	132,799	235,670	132,558	69,681	37,242	20,411	11,807	7,293	4,574	3,209	2,323	1,696	1,391	11,980
Kansas	214,300	117,961	237,361	127,009	62,491	31,958	17,025	9,727	5,895	3,900	2,627	1,961	1,496	1,236	11,804
Kentucky	409,572	184,486	300,555	150,476	71,587	35,499	18,810	10,840	6,603	4,407	3,046	2,198	1,726	1,412	13,986
Louisiana	368,991	187,824	333,909	171,957	83,476	41,914	22,099	12,396	7,537	4,912	3,467	2,590	1,960	1,610	16,404
Maine	187,728	67,274	76,988	36,724	17,733	9,295	5,216	3,133	2,037	1,340	905	704	542	422	3,428
Maryland	289,446	156,344	361,699	218,132	121,075	67,672	39,078	23,126	14,381	9,424	6,360	4,604	3,395	2,682	22,809
Massachusetts	401,381	232,577	466,186	244,779	126,562	69,388	40,956	25,654	17,256	12,092	8,619	6,738	5,214	4,072	33,348
Michigan	840,449	407,036	791,907	347,582	144,793	64,669	32,049	17,731	11,050	7,470	5,430	4,286	3,584	2,967	34,997
Minnesota	428,810	220,860	415,682	204,846	95,962	47,654	25,198	14,574	8,960	6,095	4,271	3,406	2,635	2,143	22,093
Mississippi	299,488	130,464	178,770	86,211	40,384	20,265	10,078	6,310	3,530	2,729	1,703	1,454	1,005	872	8,470
Missouri	473,529	242,742	458,113	229,566	106,649	51,805	26,656	14,778	8,874	5,829	4,029	3,185	2,368	2,041	21,545
Montana	109,881	48,816	77,167	41,799	21,530	11,575	6,448	3,696	2,250	1,501	1,082	728	576	417	3,357
Nebraska	126,815	66,789	149,470	83,953	43,493	23,058	12,397	7,092	4,525	2,900	2,085	1,463	1,120	898	8,206
Nevada	103,342	71,619	196,937	101,126	47,511	22,841	11,750	6,612	3,790	2,394	1,633	1,221	906	741	10,125
New Hampshire	167,696	68,753	86,637	41,777	20,041	10,310	5,620	3,372	2,158	1,454	1,018	762	588	500	4,055
New Jersey	380,671	248,775	632,565	350,158	184,237	101,155	59,857	37,981	25,582	18,227	13,318	10,304	7,925	6,337	48,177
New Mexico	143,750	76,165	156,031	85,323	42,920	22,330	12,106	6,941	4,152	2,657	1,954	1,404	1,001	848	7,869
New York	866,653	435,050	984,795	576,995	337,427	211,826	142,416	101,639	74,880	56,978	43,780	34,828	28,139	23,045	163,055
North Carolina	895,249	410,876	636,372	288,377	125,514	57,444	29,048	15,991	9,801	6,705	4,730	3,729	3,039	2,587	27,955
North Dakota	62,042	31,011	57,504	31,690	16,193	8,722	4,727	2,762	1,682	1,099	776	560	407	316	2,593
Ohio	860,341	452,654	901,024	429,827	193,685	91,013	46,491	25,483	15,449	9,966	7,026	5,299	4,165	3,354	37,971
Oklahoma	270,604	137,670	262,355	143,323	71,152	35,804	18,919	10,824	6,292	4,133	2,973	2,125	1,621	1,267	12,868
Oregon	250,503	132,817	323,432	175,517	87,166	44,957	24,074	13,765	8,476	5,463	3,825	2,739	2,126	1,614	16,122
Pennsylvania	935,262	449,547	845,277	451,505	230,287	121,587	67,594	40,230	25,186	17,004	11,916	8,553	6,704	5,142	45,741
Rhode Island	50,396	29,934	74,791	41,344	21,969	12,292	6,958	4,366	2,746	1,882	1,380	952	753	555	4,439
South Carolina	437,262	208,329	306,810	135,003	57,909	26,876	13,166	7,582	4,712	3,223	2,399	1,841	1,556	1,288	13,430
South Dakota	80,271	39,497	62,493	33,272	16,623	8,632	4,714	2,570	1,545	1,051	759	588	417	351	2,959
Tennessee	625,639	292,734	439,938	190,913	79,821	36,931	17,852	10,599	6,178	4,539	2,996	2,612	1,938	1,741	18,932
Texas	1,008,784	568,306	1,314,792	615,351	268,060	123,262	60,880	33,501	19,601	13,329	9,403	7,388	5,843	5,074	69,425
Utah	129,377	82,336	209,698	118,806	59,981	31,487	16,852	9,866	6,188	3,991	2,678	1,979	1,541	1,214	13,934
Vermont	98,358	33,400	34,655	16,799	8,612	4,458	2,458	1,539	1,006	698	514	384	308	223	1,723
Virginia	473,798	234,133	466,858	280,542	158,390	91,205	54,414	34,670	22,168	15,269	10,768	7,893	5,972	4,623	49,115
Washington	437,257	232,493	508,534	261,701	127,844	64,739	34,600	19,923	12,226	8,218	5,756	4,329	3,362	2,785	28,923
West Virginia	174,715	78,040	99,919	55,836	30,110	16,847	9,890	6,103	3,940	2,523	1,860	1,330	1,024	823	5,311
Wisconsin	517,240	239,064	462,633	230,435	105,311	49,801	25,306	13,686	8,003	5,280	3,675	2,772	2,249	1,821	18,287
Wyoming	60,192	30,813	49,699	28,112	14,902	7,936	4,495	2,812	1,633	1,047	731	560	408	301	2,471

Figure 2.21(total below)

Total	19,386,732	10,047,925	20,125,749	10,249,477	4,952,503	2,528,101	1,377,028	820,350	517,031	355,834	253,856	194,912	151,780	124,415	1,209,882

Section 2.5: Acknowledgements

Thank you to Kyle Marocchini for helping compile all of the person trip files, which were the basis of all of aforementioned analysis. Thank you to Bill Van Cleve and Tianay Zeigler, who assisted us with the AVO analysis and code, and were a huge help with accessing the Della supercomputer to perform simultaneous analysis.

Chapter 3:
Transit Trips West of Chicago

Our assignment was to analyze opportunities to increase ridership of urban rail systems by solving the last mile problem using autonomous taxis. We analyzed transit trips west of Chicago, focusing on cities that have subway or light rail systems. Despite the large area of this portion of the United States, only 14 of these cities have significant urban rail systems.

Elizabeth Haile and Evan Wood

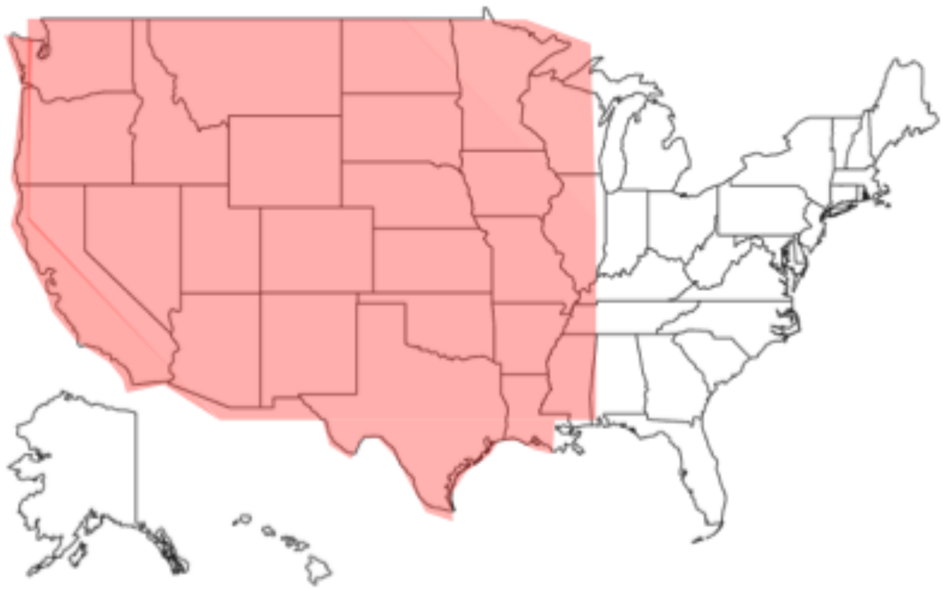


Figure 3.1

Section 1.1: Introduction-Type of Trips Analyzed

For all person trips with great circle distances greater than 0.5 miles and less than 200 miles (any trip greater than 200 miles were considered for Amtrak trips or flights and any trip less than 0.5 miles were designated as walking trips), we considered three different types of trips for transit use.

1. Walk → Transit → Walk
   Both the origin and destination of the person trip are in the same 0.5 mile x 0.5 mile pixel as a transit station.
2. aTaxi → Transit → Walk
   The origin of the person trip is within 5 miles of a transit station. The destination of the person trip is in the same pixel as a transit station.
3. Walk → Transit → aTaxi
   The origin of the person trip is in the same pixel as a transit station. The destination of the person trip is within 5 miles of a transit station.

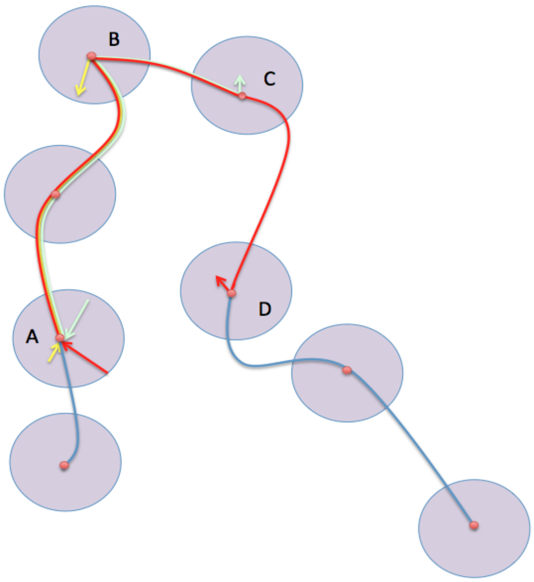


Figure 3.2: Transit Trip Simplification

We used this level of service because we assume that a person would not want to take more than 2 modes of transportation (other than walking) to get to their destination. A person would not be expected to take a taxi to *and* from a train station. We also made some assumptions about the connectivity of the transit lines; we assume that passengers are able to transfer to other rail lines at various stations along the way, so that they can easily get from the station closest to their origin point to the station closest to their destination point. The image to the right depicts the theoretical movement of 3 people who start within the vicinity of station A, and travel to stations B, C, and D.

Section 3.2: Generating Transit Trip Files

Drawing from publicly-available geospatial data for the stations of major rail systems and using an efficient algorithm to find the nearest station to an individual's origin and destination (kdTree), we generated a file with transit trips and associated aTaxi trips under the following constraints.

1. Walking trips travel at a uniform 3 mph.
2. aTaxi trips travel at a uniform 30 mph.
3. Transit trips travel at a uniform 40 mph.
4. There is a maximum circuity of 30% on trip distance.

The transit trip file generated had the following format.

0	1	2	3	4	5
X-Coord(Origin)	Y-Coord(Origin)	Long(Origin)	Lat(Origin)	X-Coord(Dest)	Y-Coord(Dest)
-4687	-649	-118.62	34.21	-4678	-646
-4678	-646	-118.54	34.23	-4646	-671
-4687	-649	-118.62	34.21	-4678	-646
...	...	...	...	...	...

6	7	8	9	10	11	12
Long(Dest)	Lat(Dest)	Departure Time (sec)	Trip Segment Dist	Arrival Time (sec)	Type of Trip	Trip Count
-118.54	34.23	27231	4.95	27825	2	1
-118.25	34.04	28125	20.64	30283	0	1
-118.54	34.23	27451	4.95	28045	2	2
...	...	...	...	...	...	...

Figure 3.3: Sample Transit Trip File

Indexes 0-7: Keep track of the segment origin and destination (lat/long & pixel coordinates)
Index 8, 10: Stores departure time and arrival time
Index 9: Stores the segment distance
Index 11: Specifies the type of trip

- Transit trip
- aTaxi trip to station
- aTaxi trip from station

Index 12: Has a counter to synthesize segments of individual trips together

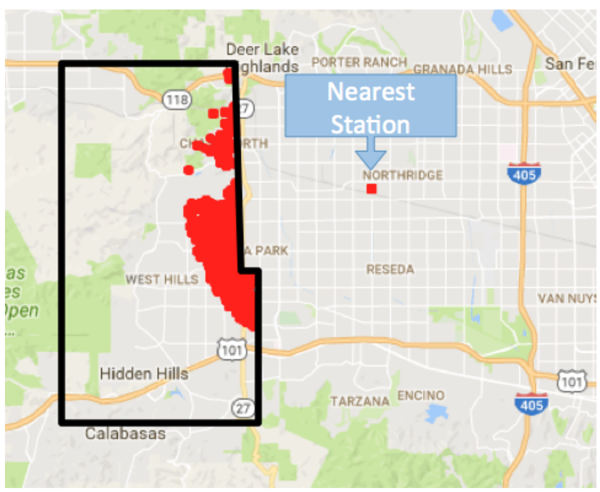


Figure 3.4: Northridge, Los Angeles

When considering trip files, we made sure that the only people who were considered in this analysis were people who lived within 5 miles of a train station. The map above is a sample of a district in the trip file in Los Angeles. Northridge station is the west-most station in the Los Angeles transit system, so it makes sense that the only people living in West Hills who will take the Metro are people who live in this 5 mile radius. After searching through all relevant trips in the Los Angeles area, we were able to compile a map of all trips that include a transit trip.

Figure 3.5: Los Angeles **Figure 3.6 Los Angeles aTaxi Trips**

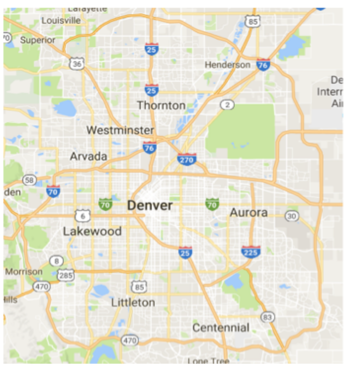
Figure 3.5 is a map of the city of Los Angeles, while Figure 3.6 is a depiction of all of the trips to and from the nearest transit station. The dark black regions are the epicenters of the train stations, and the black lines that extend outwards represent the individual trips made by aTaxis. This web visually shows all of the areas of Los Angeles that are within 5 miles of a station. The large white area in the middle of the map is the Angeles National Forest, so there are obviously no transit trips to this area. The maps of all 14 cities that we analyzed are shown below:

SEATTLE

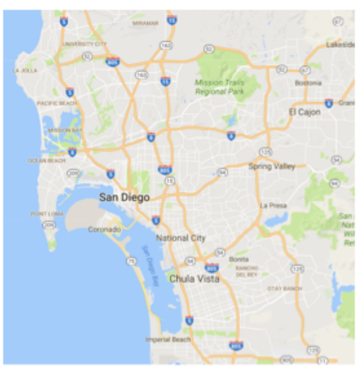
BAY AREA

LOS ANGELES

DENVER

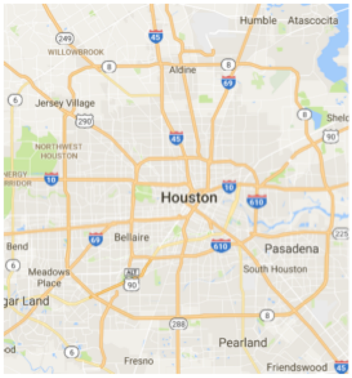


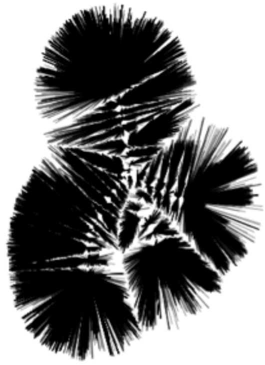
San Diego



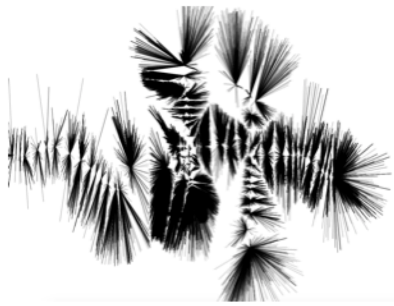
Salt Lake City

Houston



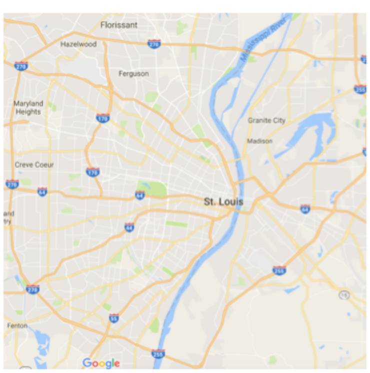


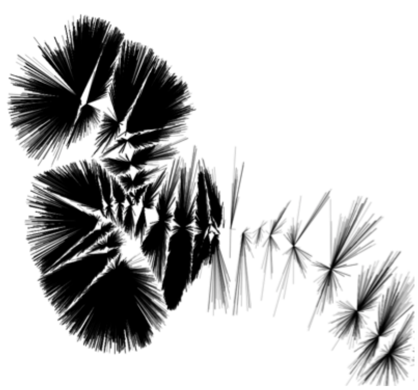
Portland

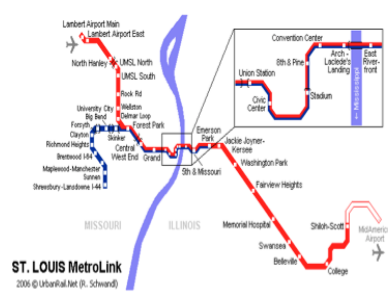


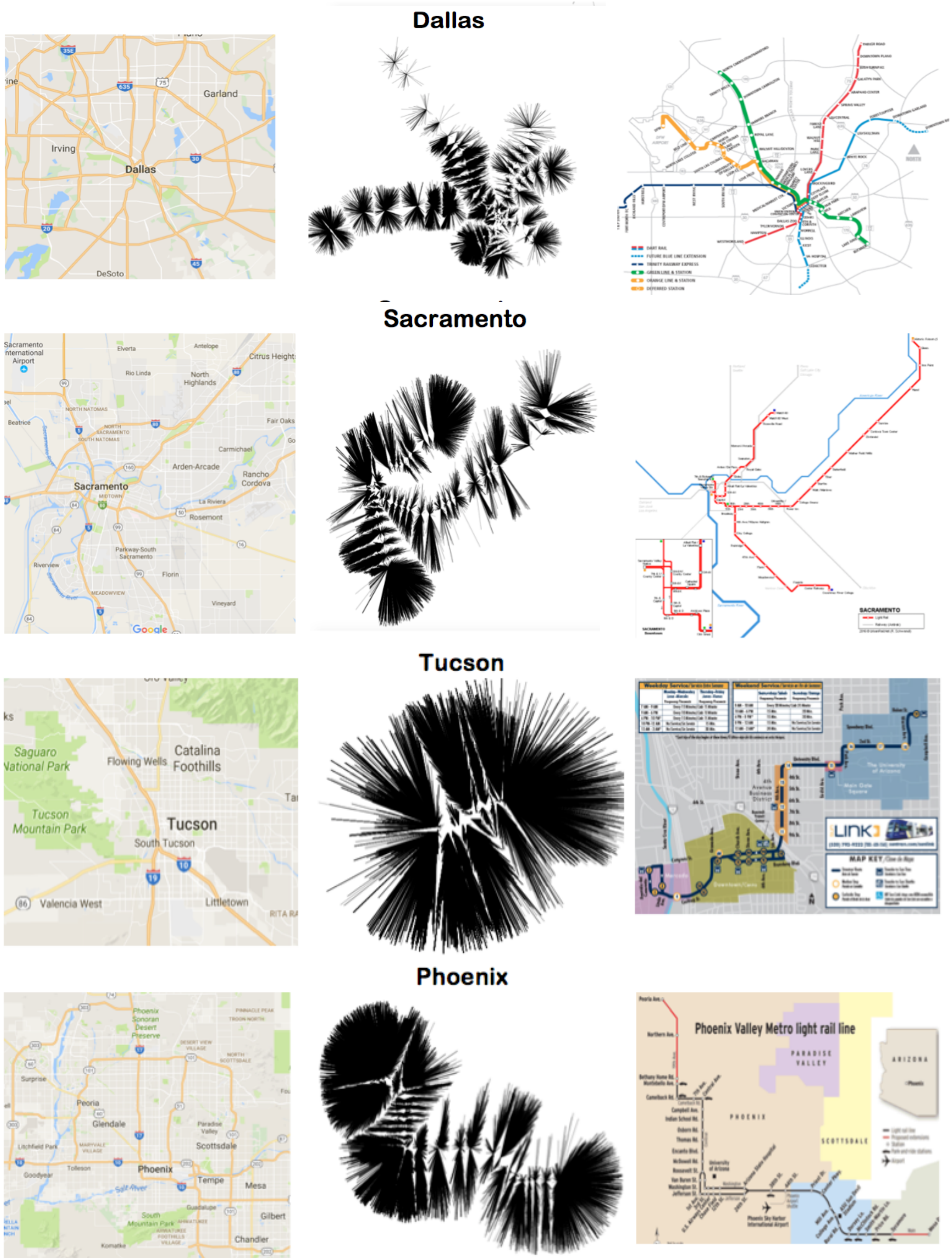
St. Paul

St. Louis









Dallas
Garland
Irving
Dallas
DeSoto
Sacramento
Sacramento International Airport
Elverta
Antelope
Citrus Heights
Rio Linda
North Highlands
Fair Oaks
Carmichael
Arden-Arcade
Rancho Cordova
Sacramento
La Riviera
Rosemont
Parkway-South Sacramento
Riverview
Florin
Vineyard
Google
Tucson
Saguaro National Park
Flowing Wells
Catalina Foothills
Tucson Mountain Park
Tucson
South Tucson
Valencia West
Littletown
The University of Arizona
Main Gate Square
LINK
MAP KEY
Phoenix
Phoenix Sonoran Desert Preserve
Surprise
Peoria
Glendale
Paradise Valley
Scottsdale
Litchfield Park
Goodyear
Tolleson
Phoenix
Tempe
Mesa
Salt River
South Mountain Park
Guadalupe
Gilbert
Chandler
Komatke
Phoenix Valley Metro light rail line
PARADISE VALLEY
PHOENIX
SCOTTSDALE
ARIZONA

Figure 3.7

It is worth noting that when comparing our analysis (center column) to the urban transit maps (the rightmost column), not all of the lines appear to match up perfectly. This is because many map designers choose to draw subway maps with simpler, more elegant paths, rather than match the actual geography of the region. For this reason, both the transit maps and the regular maps (leftmost column) are provided. One interesting observation from looking at these maps are the white "stripes" that exist; these visually show a separation line that represents the areas of the city that are most inaccessible by public transit. These stripes are the result of being equally far from two stations. These 14 sets of maps show not only the detail of each city, but also the scalability of the model that we developed. Given station and trip coordinates, this analysis can be done on any region of the world!

Section 3.3: Implications on Ridership

Now that we have mapped the movement of people using transit systems, we can begin to dig deeper by looking at the number of people who use the system, and how they get to and from the stations. As described earlier, our transit trip file breaks down each trip into 3 components: (1) transit trip, (2) aTaxi trip to station, and (3) aTaxi trip from station. By adding up the total number of transit trips made, we can calculate a ridership estimate for the number of people who use public rail transportation in each city. Because we assume that everyone who is eligible to ride the train does so, our ridership estimates are much larger than the current numbers. Our analysis provides an upper bound for the volume of passengers who could utilize public transit once autonomous taxis and more efficient ride-sharing becomes available. Figure 3.1 shows our estimated increase in ridership for each of the cities that we analyzed. The cities are sorted from left to right by the increase in improvement.

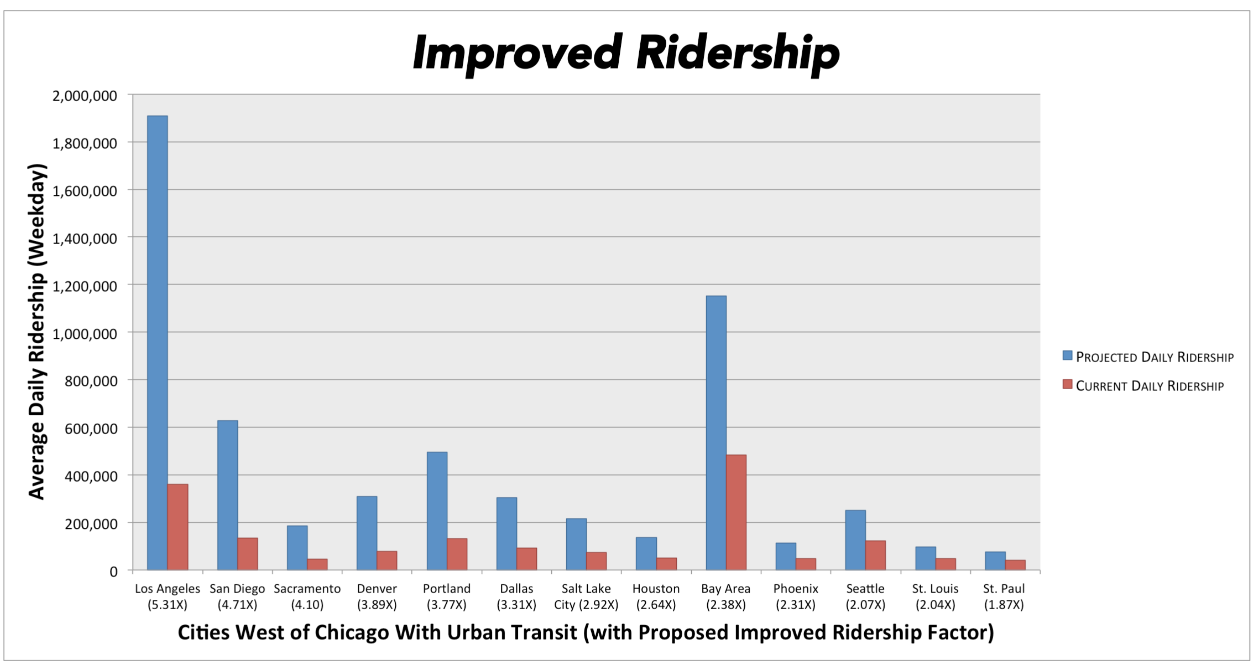


Figure 3.8

Figure 3.9 summarizes the data presented in Figure 3.8

City	Projected Daily Ridership	Current Daily Ridership	Increase Factor
Los Angeles (5.31X)	1,909,829	359,861	5.31
San Diego (4.71X)	626815	133,200	4.71
Sacramento (4.10)	185800	45,300	4.1
Denver (3.89X)	307,861	79,200	3.89
Portland (3.77X)	493,983	131,100	3.77
Dallas (3.31X)	304,303	92,000	3.31
Salt Lake City (2.92X)	215,464	73,800	2.92
Houston (2.64X)	135,425	51,332	2.64
Bay Area (2.38X)	1,151,347	483,339	2.38
Phoenix (2.31X)	113051	48,900	2.31
Seattle (2.07X)	250,805	121,382	2.07
St. Louis (2.04X)	97175	47,600	2.04
St. Paul (1.87X)	76079	40,502	1.87

Figure 3.9

The information about current ridership statistics can be found here.

Because we also kept track of the number of autonomous taxi trips to and from the station, we were able to calculate the modal split of passengers. Using the fact that there is a complementary distribution of passengers who get picked up by aTaxis and those who get dropped off by aTaxis, it was possible to find the number of people who walked both to and from the station.

Figure 3.10 shows the breakdown of the modes of transportation for each trip in each city. Note that a Walking trip indicates that the person walked on *both* ends of the trip, while an aTaxi Drop-off indicates that a person took a taxi to the station, and walked from the station closest to their final destination.

City	aTaxi Drop-off	aTaxi Pick-up	Walking
Los Angeles (5.31X)	873,470	865,007	171,352
Bay Area (2.38X)	555,956	476,197	119,194
San Diego (4.71X)	291,957	266,114	68,744
Portland (3.77X)	225,841	192,514	75,628
Denver (3.89X)	158,627	121,008	28,226
Dallas (3.31X)	140,221	136,625	27,457
Seattle (2.07X)	121,231	105,836	23,738
Salt Lake City (2.92X)	105,811	90,263	19,390
Sacramento (4.10)	93,479	75,927	16,394
Houston (2.64X)	61,147	57,154	17,124
Phoenix (2.31X)	55,294	48,639	9,118
St. Louis (2.04X)	46,974	39,589	10,612
St. Paul (1.87X)	37,896	32,662	5,521

Figure 3.10: Transit Trip Mode Split in Western U.S. Cities

For the sake of simplicity, the next few pages will focus specifically on just 4 of the 14 cities that we analyzed. Figure 3.11 looks in depth at the breakdown of transportation modes for the cities of Los Angeles Portland, San Francisco, and Dallas. It shows the relative percentages of people who utilize each mode. The size of the pie charts are indicative of the ridership numbers in each city.

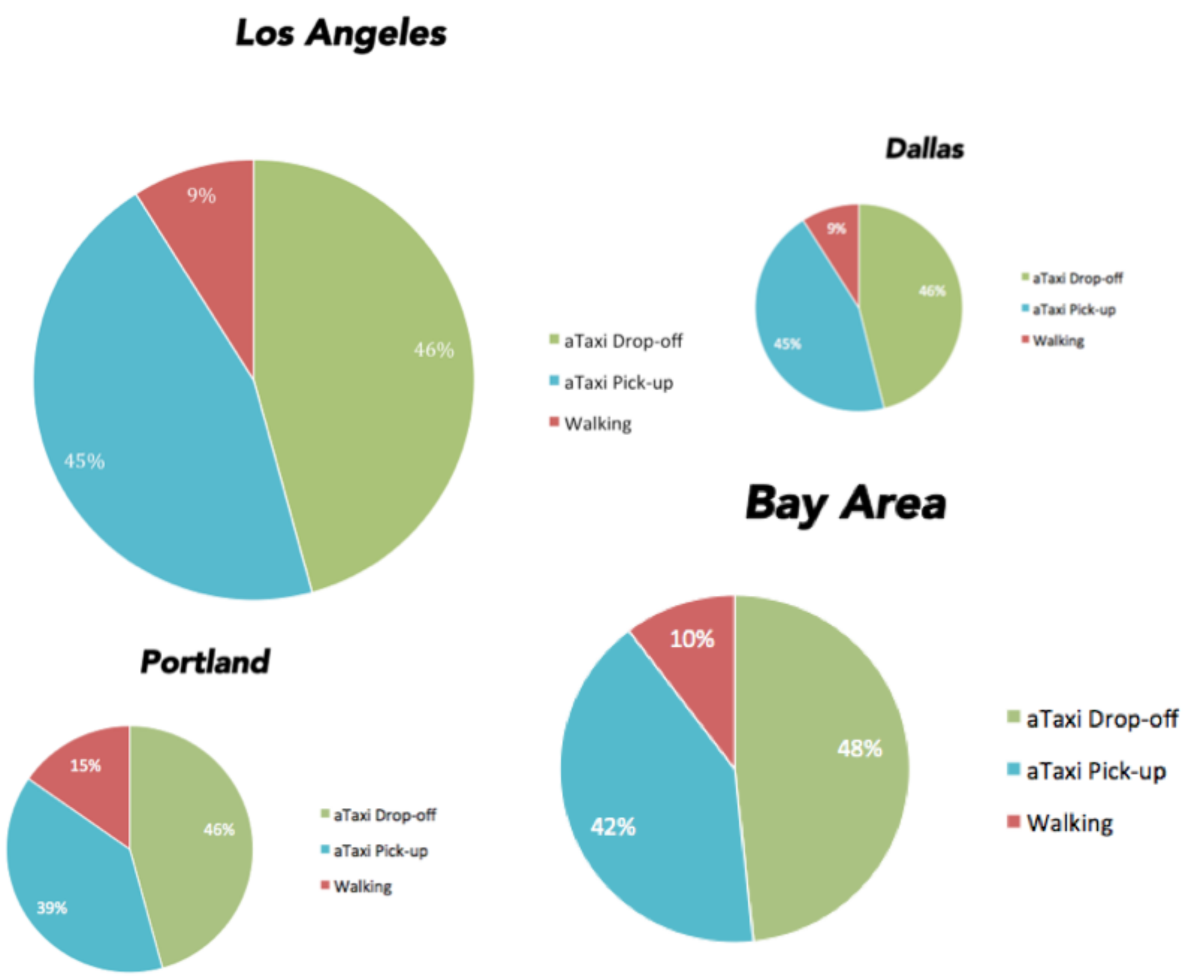


Figure 3.11: Mode Split

From the pie charts, Portland has the highest percentage of pure walking trips to and from each transit station. Dallas and Los Angeles have nearly identical mode distributions. The relatively low number of pure walking trips to and from transit stations in cities in the Western United States as compared to cities like New York with much more extensive public rail transit systems indicates some of the shortfalls of existing transit systems and urban design of Western cities.

Next we look closely at the individual stations of each city, to see the relative concentrations of passengers. Plots of the stations of these four cities are in Figures 3.12-3.15.

Daily Volume of Passengers at Each Station in Four Cities

Dallas

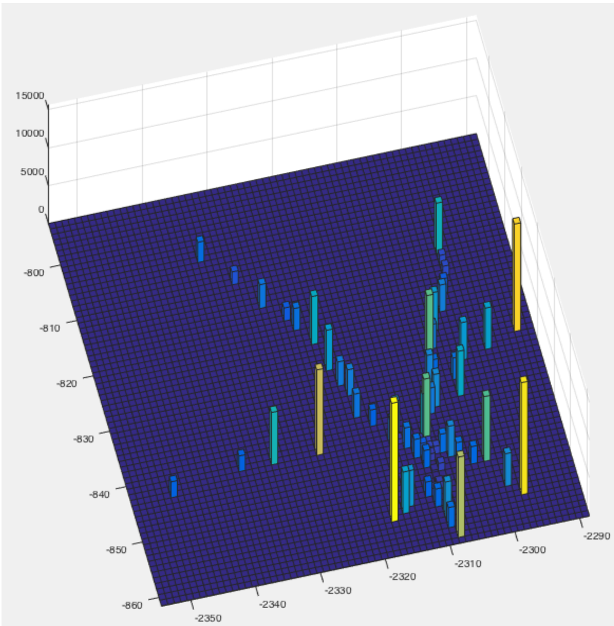


Figure 3.12

Los Angeles

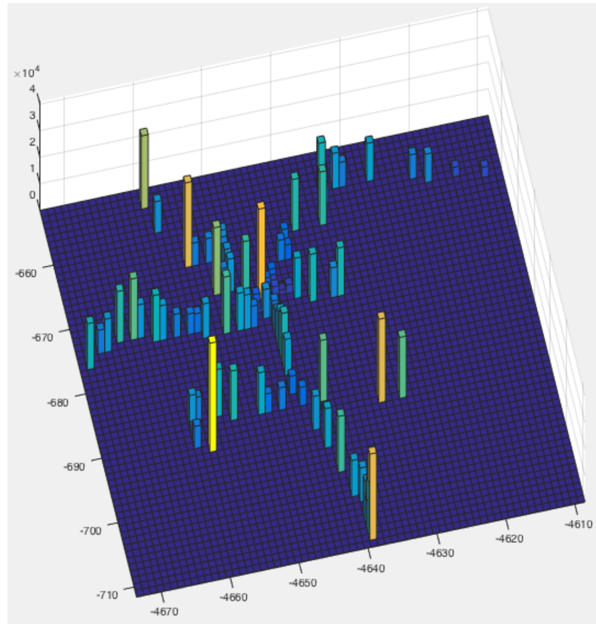


Figure 3.13

Portland

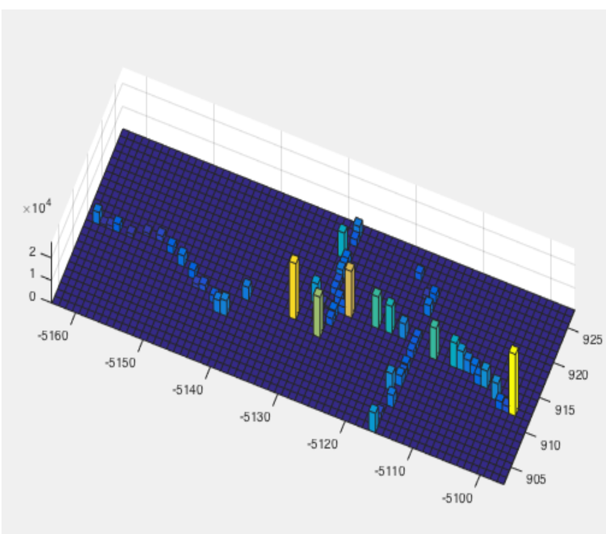


Figure 3.14

Bay Area

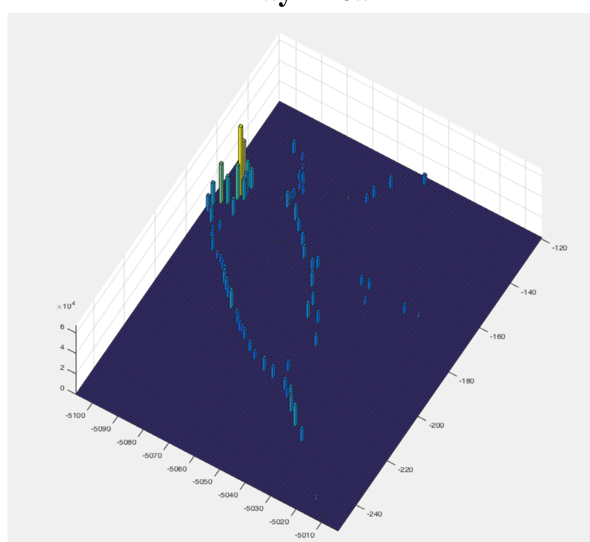


Figure 3.15

The 3D histograms above indicate the relative activity of aTaxi transit trips to and from transit stations in Los Angeles, the Bay Area, Dallas, and Portland. Each colored cell represents the location of a train station, and the height of the cell corresponds to the volume of activity at each station. As expected, there is a high level of activity in city centers. Interestingly, there is also a high level of activity at the end of lines, indicating that many people ride the trains to the end of the line. This is particularly evident in Figures 3.12 and 3.13. This is an indication that the transit line could be expanded in order to more effectively meet the travel demands of the passengers who use public transportation. Not surprisingly, Dallas and Los Angeles, the two cities with the highest concentration of station activity at the ends of the train lines, have some of the least developed transportation systems in the county. Expanding the transit lines of these cities could vastly improve the ride-sharing possibilities.

Section 3.4: Simulated Scheduling of Transit System

As part of our analysis of transit trips, we decided to simulate the actual transit line by running trains every 10 minutes to each station, and picking up the passengers in each 10 minute window. We assume that all people board the first available train that arrives at the station, and that from that station you can get to any other station. Once we determined the start time of all passengers boarding the trains to their respective destinations, we were able to calculate the expected arrival time of each person. To do so, we assumed that trains traveled at around 40 miles per hour, and multiplied the distance traveled by the speed to find the duration of each trip.

At each station, we kept track of two different timelines: one of all the passengers that arrive at this given station, and one of all the passengers that depart the given station. The figure below shows the modeling of people who are entering and leaving the train stations (this diagram shows the movement of passengers *before* the simulation of the transit line). The vertical dotted lines represent ten minute intervals at which trains might arrive at a station.

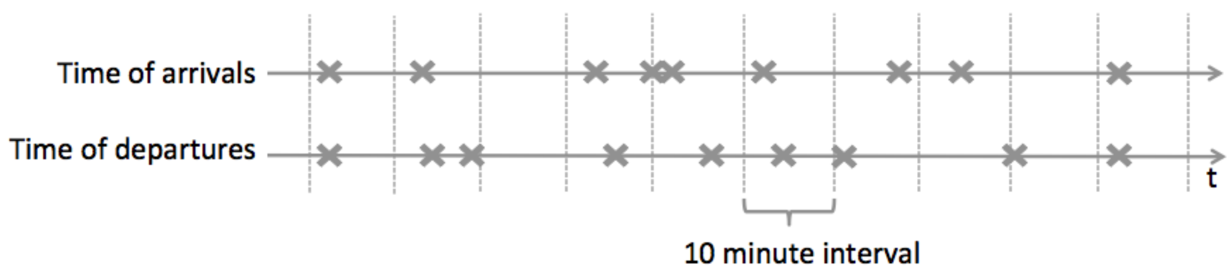


Figure 3.16

Once we have created this data structure for each station, we are able to simulate the actual transit line, as discussed above. We simply transform the two timelines by shifting the passengers to the next 10 minute interval. Thus, for each station, we can simulate the number of passengers that embark and disembark a single train.

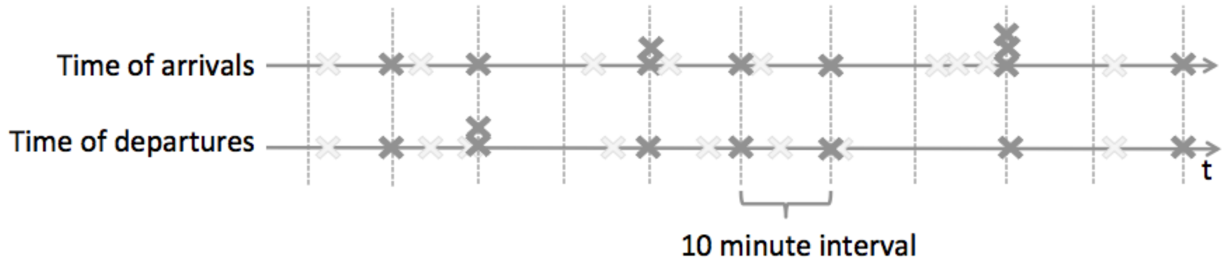


Figure 3.17

What we can learn from this simulation is even more interesting; once we know how many people board and exit the trains at any given time and at every station, we can begin to quantify the capacities of not only the trains, but also the autonomous taxis that pick people up from each station. Because of the rudimentary model that we use for simulating the movement of trains, we will not analyze the minimum number cars per train needed to pick up all of the waiting passengers in this report; such an analysis could easily be done in a later report. We will focus on the movement of autonomous taxis that pick up passengers from each station instead. This is an ideal scenario to consider, as it provides the most efficient level of service for both the taxis and the people riding them. Because a train arrives at each station every 10 minutes, this is an ideal ride-sharing situation. The only people who get off at a particular station in our model must be within 5 miles of their destination. This means that passengers who disembark a train at the same station have a high probability of having a similar destination, which indicates that an efficient autonomous taxi ride-sharing program can be implemented.

Section 3.5: Pico Station, Los Angeles: aTaxi Analysis Case Study

In order to look at the implications this simulated scheduling of the transit system, let us focus on one station in particular as a case study. It is located right next to Downtown Los Angeles, the financial center of the city, and other commercial centers like the jewelry, garment, and fashion districts. Pico Station is also located next to the L.A. Live, a restaurant and recreation complex, the Staples Center, and the Los Angeles Convention Center.

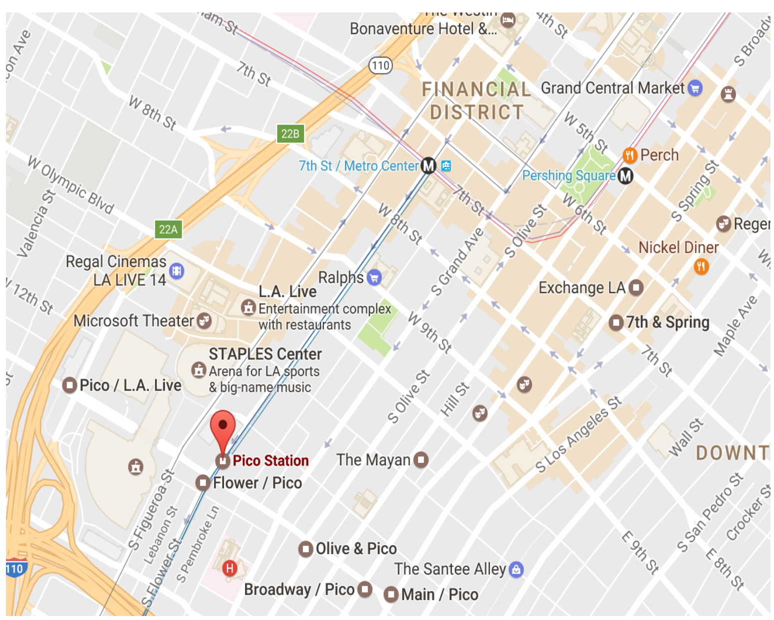
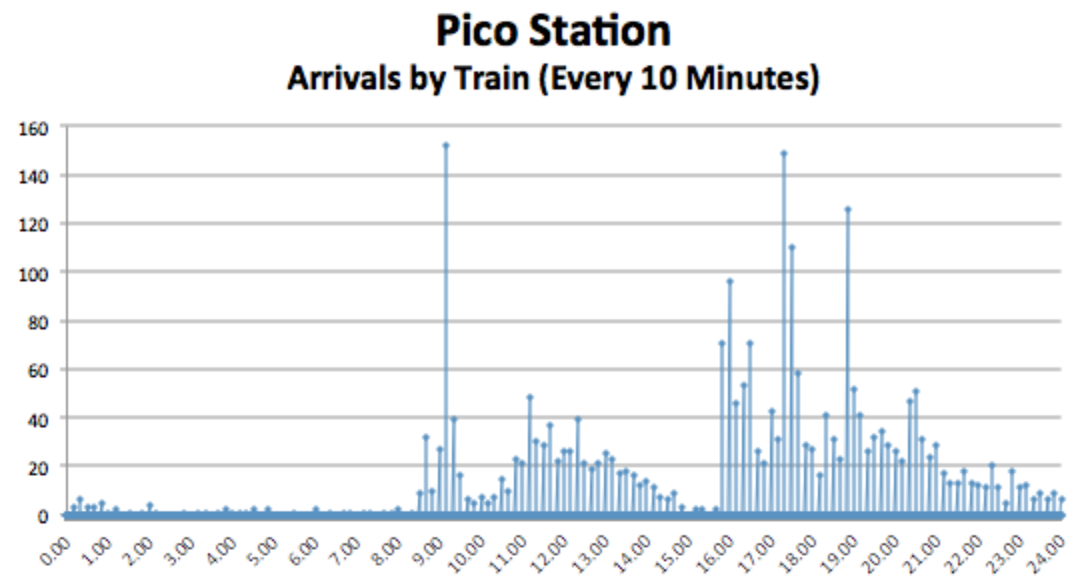


Figure 3.18: Map of Pico Station

Using Pico Station on the L.A. Metro as a case study, we are able to simulate the arrivals and departures (Figure 3.19) by train for every 10 minute interval of the day. Not surprisingly, the peak arrivals at Pico Station occurs at 9AM, corresponding to the morning commutes of Angelinos heading downtown. Similarly, the peak departures from Pico Station shortly after 5PM matches the corresponding rush hour commute out of the city at the end of the workday. Notably, there continues to be a relatively high number of arrivals into Pico Station after the end of the work day, perhaps corresponding to individuals coming to the L.A. Live area for restaurants and evening recreation.



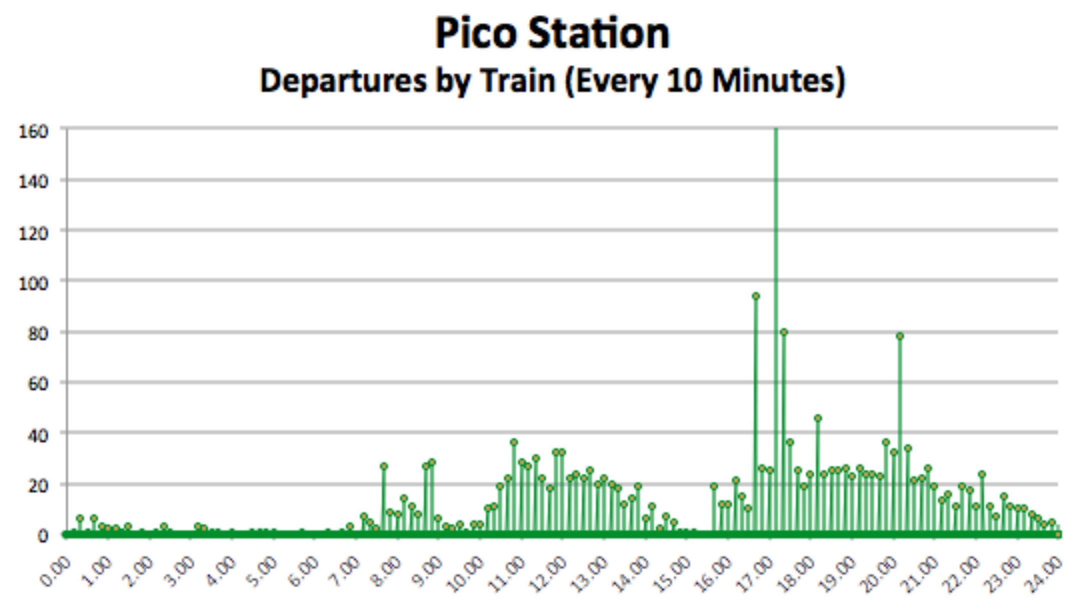


Figure 3.19

These peak hours of arrivals into Pico Station provide ideal ride-sharing opportunities. Using the following constraints on the level of service of aTaxi trips in Figure 3.20, we were able to simulate the ride-sharing capacity of person trips leaving Pico Station.

Level-of-Service Constraints on aTaxi Trips
1. Maximum circuity of 1.25% 2. Maximum capacity of 4 passengers 3. Maximum of 3 destinations for a single aTaxi trip

Figure 3.20

We are able to simulate the number of aTaxis needed at Pico Station for every 10 minute interval of the day, information that can better inform the repositioning of aTaxis and better serve the needs of commuters by decreasing waiting time. It follows that the time at which the greatest number of aTaxis are needed at Pico Station corresponds to the time at which there are the greatest number of arrivals by train to the station. Thus, the plot in Figure 3.21 looks nearly identical to the plot of arrivals by train in Figure 3.19, except at a different scale.

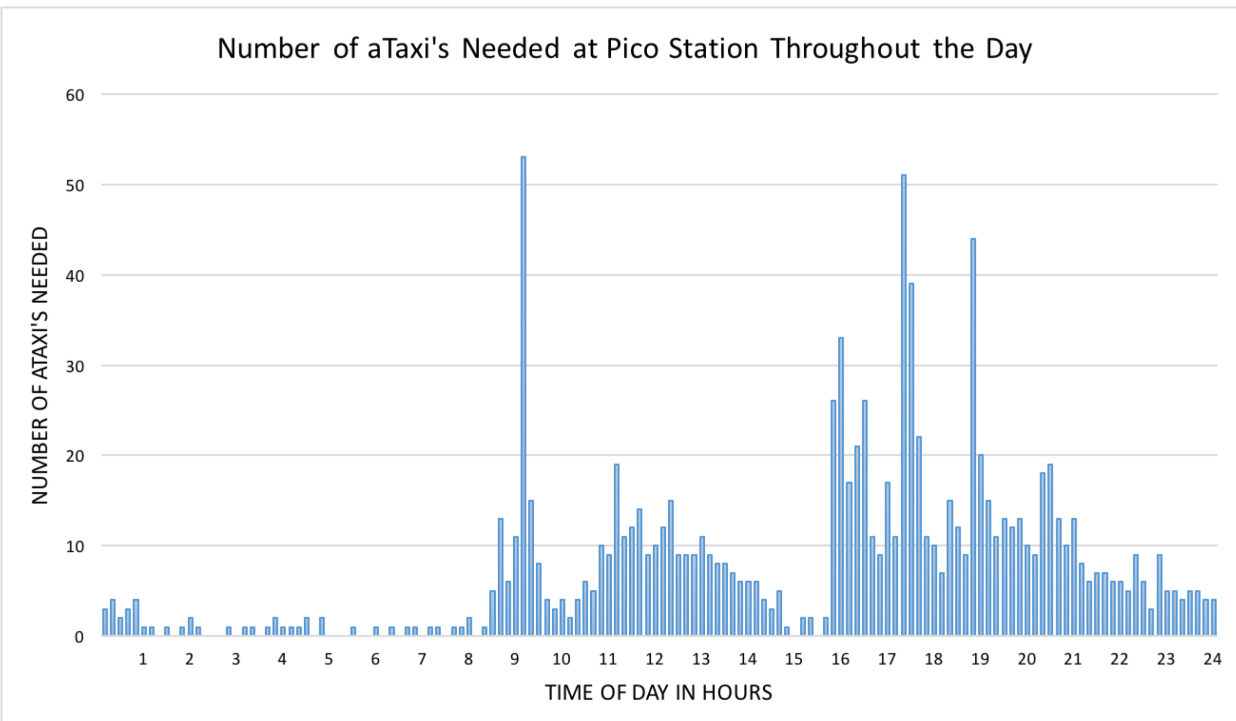


Figure 3.21

Figure 3.22 is a summary of the maximum number of aTaxis needed at the peak time for each L.A. Metro station on an average workday. The table also contains the corresponding number of passenger arrivals, average vehicle occupancy at departure, and time of day when this maximum number of aTaxis is required. As expected, the busiest time at most stations in L.A. is around rush hour at the end of the workday, with a few notable exceptions corresponding to the morning rush hour. Atlantic station is located next to a number of schools, which could explain this peak activity time of 3:48 pm.

Station	Passengers Arriving	aTaxi's Needed	AVO at Departure	Time
Sorrento Valley	720	193	3.7	6:12 pm
Crenshaw Station	516	138	3.7	5:30 pm
Indiana	427	122	3.5	5:18 pm
Florence	441	121	3.6	5:30 pm
Pico	152	53	2.9	***9:00 am***
Vincent Grade	67	22	3.0	6:00 pm
Memorial Park	430	136	3.2	***8:30 am***
Expo Park/USC	276	75	3.7	5:18 pm
Atlantic	48	17	2.8	3:48 pm
Harbor Freeway	170	47	3.6	5:30 pm
Culver City	325	89	3.7	5:30 pm
…	…	…	…	…

Figure 3.22

This study can be extrapolated to every transit station in Los Angeles and every city in the United States with a rail transit system. These findings present an opportunity to greatly improve the utility of transportation services. As soon as an individual begins a transit trip, it is possible to match that individual with others going to similar locations and accurately determine the number of aTaxis needed at the destination transit station. Issues of long waiting times and inefficient distribution of individuals to the appropriate vehicles become virtually nonexistent.

Section 3.6: Conclusion

Analysis of the use of aTaxis to solve the last mile problem presents an opportunity to greatly increase ridership of existing rail transit systems in U.S. cities west of Chicago. Our findings in Los Angeles, for example, suggest a 5.31 times increase in L.A. Metro daily ridership. In places like Los Angeles, a city with one of the worst traffic problems in the country, aTaxi development and integration into the transport paradigm has the potential to get more cars off the road, which will decrease congestion and reduce the environmental impacts associated with heavy personal car usage. Autonomous taxis have the potential to greatly improve the time-cost utility of existing public transit systems and may very well save the future of urban rail systems in the United States.

Chapter 4:
Transit Trips In the New York Area (NY, NJ, CT)

We found and analyzed all trips in counties in the New York Metropolitan area that are affected by major transit systems.

Jessica Deng and Julia Ni

Section 4.1: Description of Data and Problem Set-up

We used the definition of a "Combined Statistical Area" (CSA), which groups together adjacent core-based statistical areas with a high degree of economic interconnection, to find counties in the New York Metropolitan area. This includes the following counties:

- New York City: The five boroughs of the city are serviced by the New York subway system, with outgoing trips on Metro North, LIRR, and New Jersey Transit. Staten Island is the only borough without direct subway access, it is connected to Manhattan through a ferry system.
    - Manhattan (36061)
    - Bronx (36005)
    - Brooklyn (36047)
    - Queens (36081)
    - Staten Island (36085)
- New York Commuters: We analyzed the following counties in New York State minus the five boroughs of New York City as one group. These counties are serviced by the Port Jervis, Pascack, Hudson, and Harlem lines of Metro North. There is a significant population that commutes from these counties in New York State to New York City for work every day so it is common to have a New York subway station as the transit destination.
    - Dutchess County (36027)
    - Kings County (36047)
    - Nassau County (36059)
    - Orange County (36071)
    - Putnam County (36079)
    - Rockland County (36087)
    - Suffolk County (36103)
    - Ulster County (36111)
    - Westchester County (36119)

- New Jersey Commuters: We analyzed the following counties in New Jersey State as one group. The counties in this group are serviced by New Jersey Transit. There is a significant population that commutes from these counties in New Jersey to New York City for work every day so it is common to have a New York subway station as the transit destination.
    - Bergen County (34003)
    - Essex County (34013)
    - Hudson County (34017)
    - Hunterdon County (34019)
    - Mercer County (34021)
    - Middlesex County (34023)
    - Monmouth County (34025)
    - Morris County (34027)
    - Ocean County (34029)
    - Passaic County (34031)
    - Somerset County (34035)
    - Sussex County (34037)
    - Union County (34039)
    - Warren County (34041)

- Connecticut Commuters: We analyzed all three counties in Connecticut as a group. These counties are serviced by the New Haven line of Metro North. There is a significant population that commutes from Connecticut to New York City for work every day so it is common to have a New York subway station as the transit destination.

- Fairfield County (09001)
- New Haven County (09009)
- Litchfield County (09005)

The goal of the project is to find and run a mode-split model on all transit trips in these areas. A transit trip is defined as all trips generated between 0.5 miles and 200 miles in distance that start or terminate within 0.5 miles (10 minute walk) of a transit station. The mode split was performed such that for the segments of each transit trip to and from transit stations, the passenger was assigned to an aTaxi if the trip segment is more than 0.5 miles (10 minute walk) and assigned to walk otherwise.

We found the GPS locations of transit stations for the following transit systems that affect the areas we are interested in:

- New York City Subway System
- LIRR
- New Jersey Transit
- Metro-North

There are 952 stations in total, plotted in the figure below. We pixelated the location of each station using the formula given in class to facilitate our analysis:

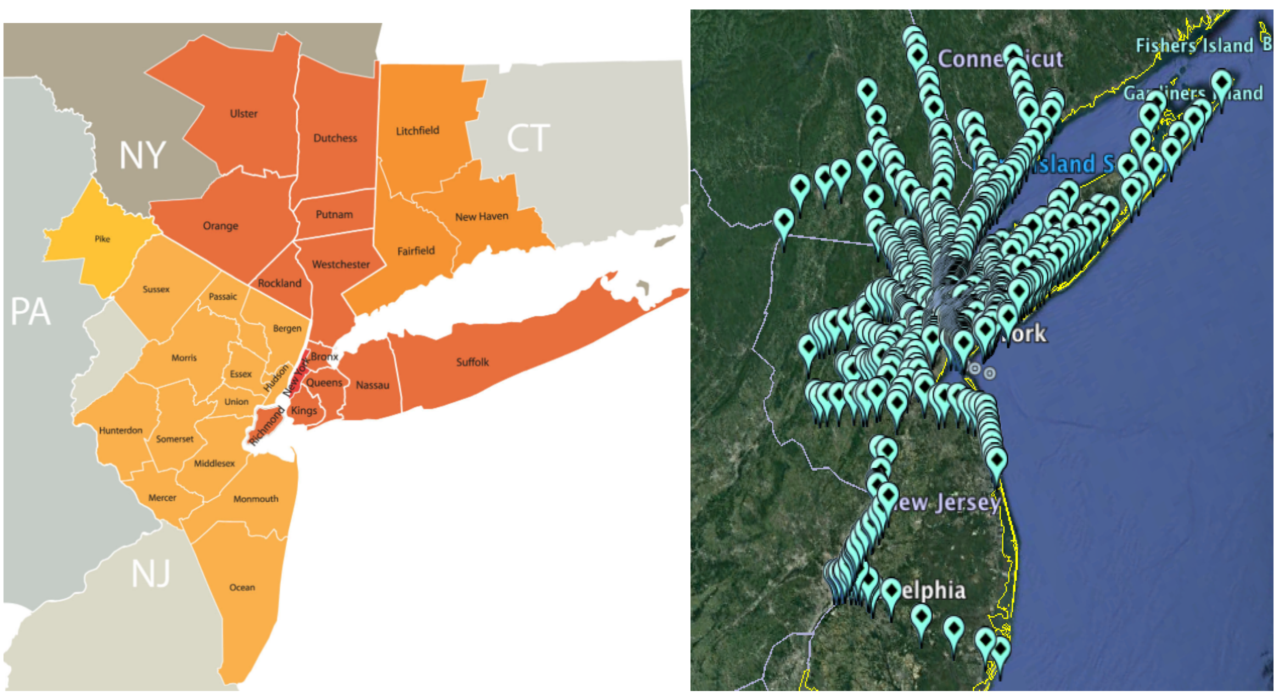


Figure 4.1

Section 4.2: Methodology

Step 1: For each area, compile raw data for trips that originate within all counties in the area. Find all trips that satisfy 0.5 miles < GCDistance < 200 miles. These are the trips we are interested in since all other trips are either walking trips or long distance trips.

Step 2: Of these trips, find all trips that have a transit station within the same pixel as the origin or destination by cross referencing a master list of all transit stations.

Step 3: If the origin is in the same pixel as a transit station, find the closest transit station to the destination and vice versa. We used a KD tree data structure to efficiently implement this optimization process.

Step 4: Check to see if the trips found so far satisfy the max circuity constraint that using transit would not results in more than a 30% increase in total distance traveled from origin to destination compared to using a direct aTaxi.

Step 5: Trips found following the steps above are all the Transit Trips that originate within the area in question. For each Transit Trip we now perform a mode split using the following principles: if both the origin and destination are within the same pixel as a transit station the transit trip is recorded as transit only, if the origin is within the same pixel as a transit station but the destination is not the transit trip is recorded as aTaxi pick up, if the destination is within the same pixel as a transit station but the origin is not the transit trip is recorded as aTaxi drop off.

Step 6: We add each Transit Trip to a master list. Segments of each Transit Trip is recorded as separate trips adjacent to each other. For example, given aTaxi drop off trip, we record an aTaxi trip and a pure Transit Trip from station to station. For each trip segment, we record the following information: X coordinate of origin, Y coordinate of origin, Longitude of origin, Latitude of origin, X coordinate of destination, Y coordinate of destination, Longitude of destination, Latitude of destination, departure time, distance of trip segment, arrival time and type of trip. We assumed that the average walking speed is 3 mi/hr, average aTaxi speed is 30mi/hr and average transit speed is 40 mi/hr to calculate arrival times. We also included a five minute buffer time to accommodate transit wait times.

Step 7: We now have a complete list of all Transit Trips and details on their mode of travel. For each area we performed the following analysis: the percentage of trips of each mode type in the area, a “whisker” plot that visualizes trips by connecting their origin with station or destination with station, and a 3D histogram that shows the spatial distribution of Transit Trips.

Section 4.3: Results

Results by Area

Manhattan (36061, Total Trips = 3,490,832):

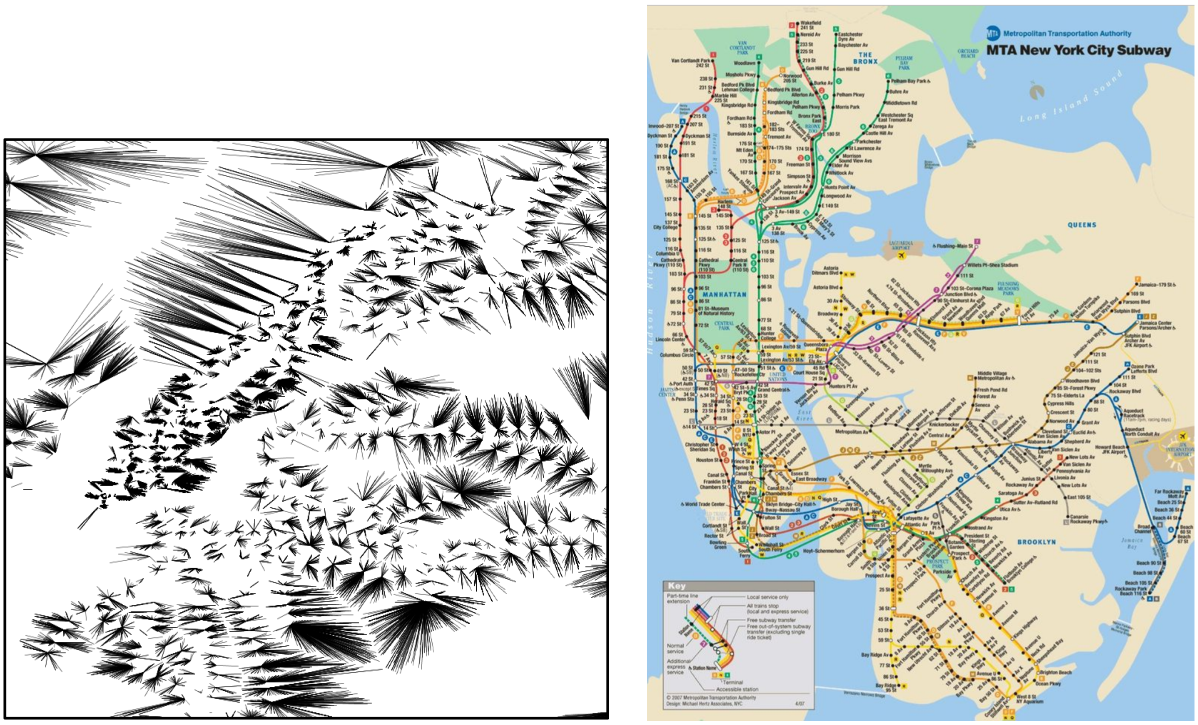


Figure 4.2

We can see Central Park in the middle of the island!

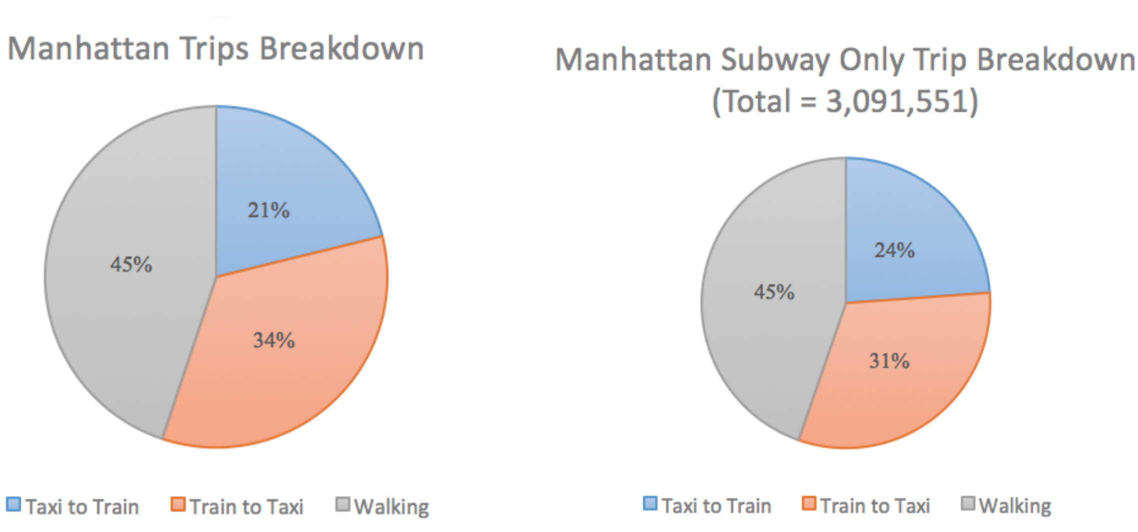


Figure 4.3

It is interesting to note that when we ran Manhattan with only Subway stations, there were still 3,091,551 trips -- indicates that most trips in Manhattan can be served with the city subway.

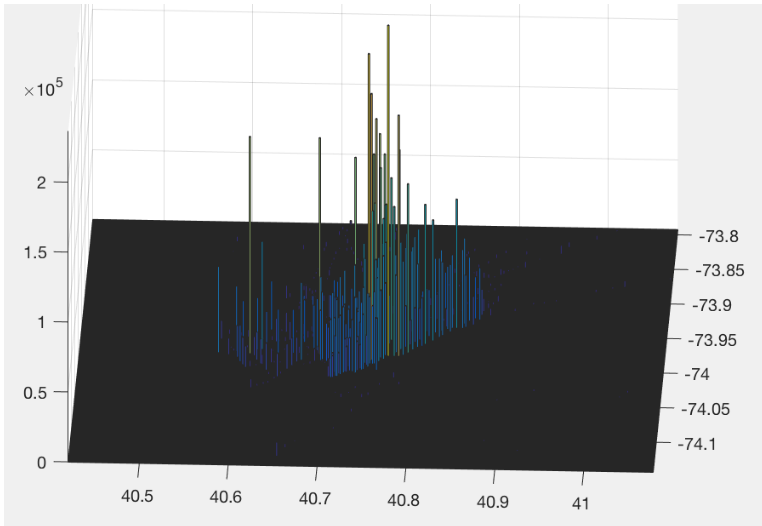


Figure 4.4: Manhattan 3D Histogram

Bronx (36005, Total Trips = 1,558,771):

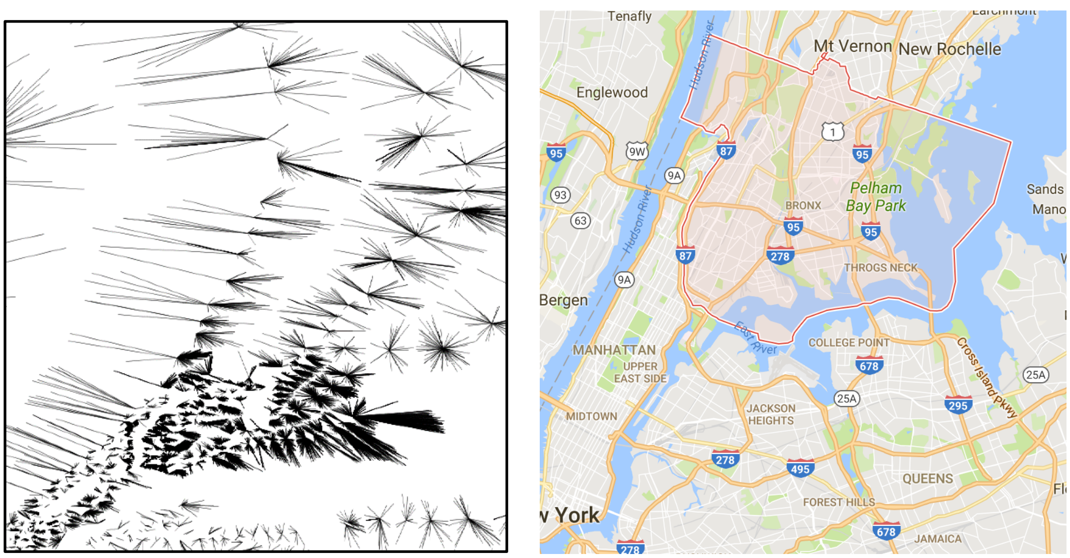
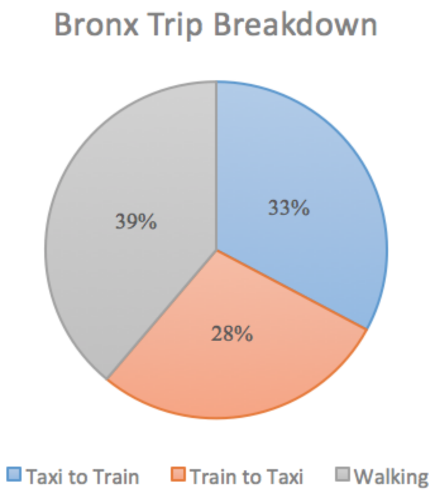


Figure 4.5



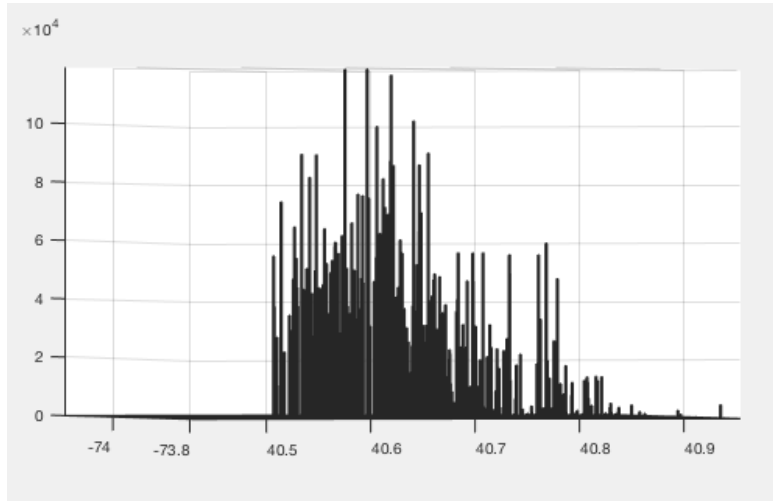
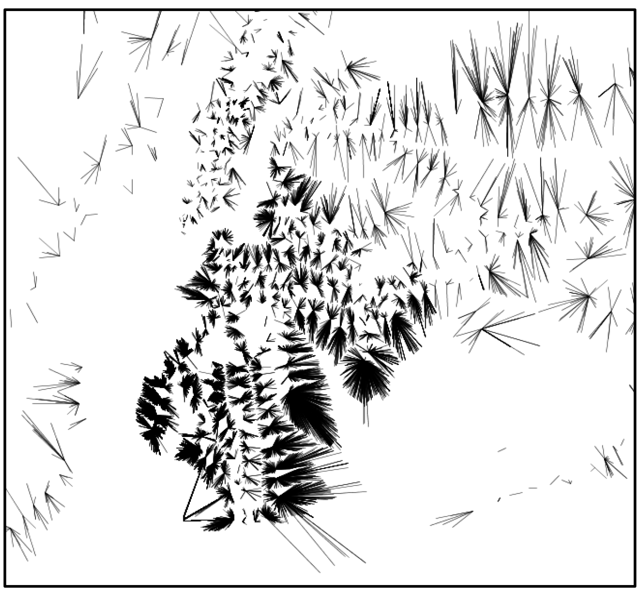


Figure 4.6: Bronx Trip Breakdown and 3D Histogram

Brooklyn (36047, Total Trips = 3,056,045):



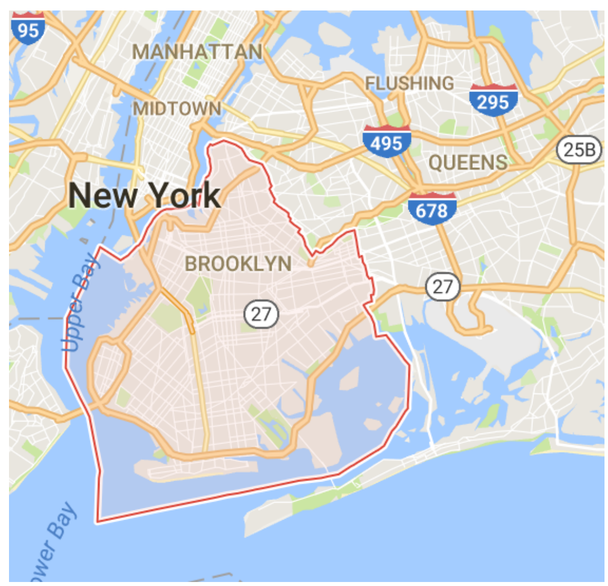
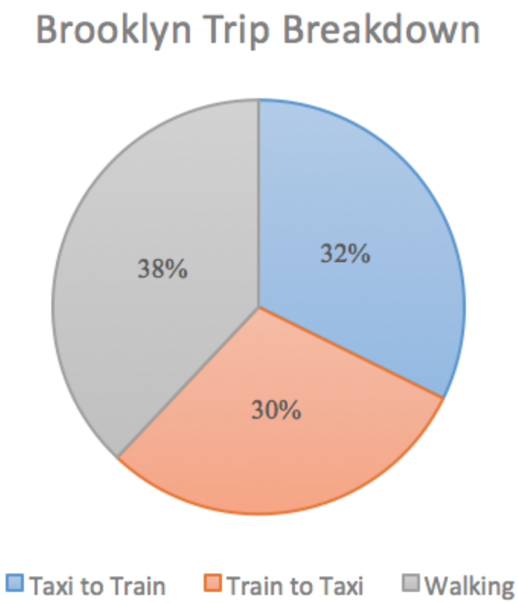


Figure 4.7



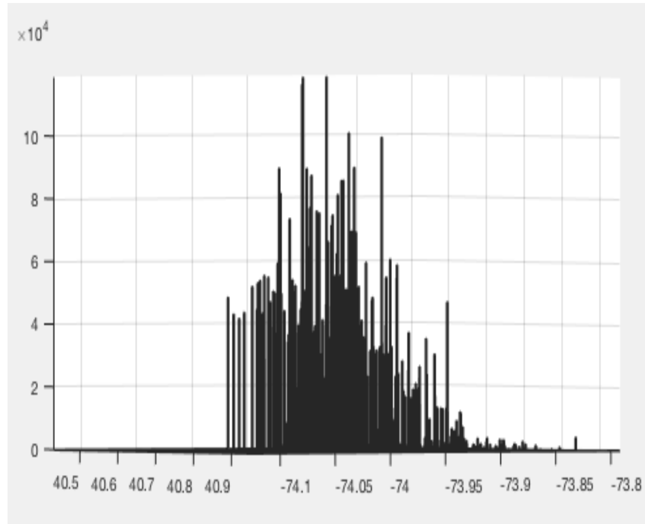


Figure 4.8: Brooklyn Trip Breakdown and 3D Histogram

Queens (36081, Total Trips = 1,710,629):

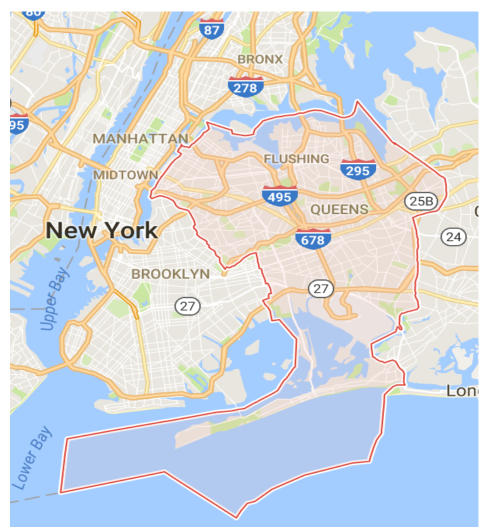


Figure 4.9

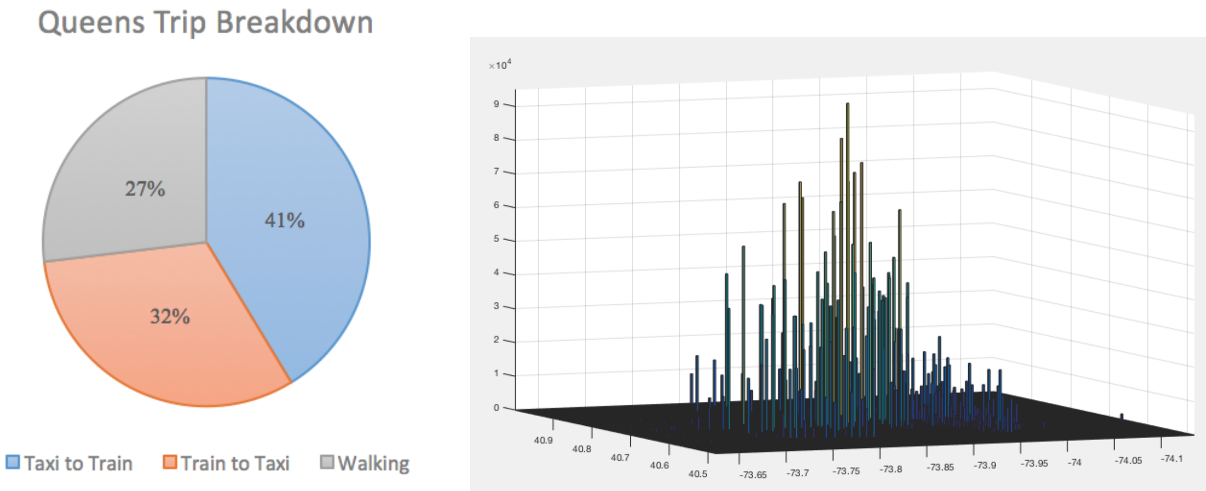


Figure 4.10: Queens Trip Breakdown and 3D Histogra

Staten Island (36085, Total Trips = 141,607):

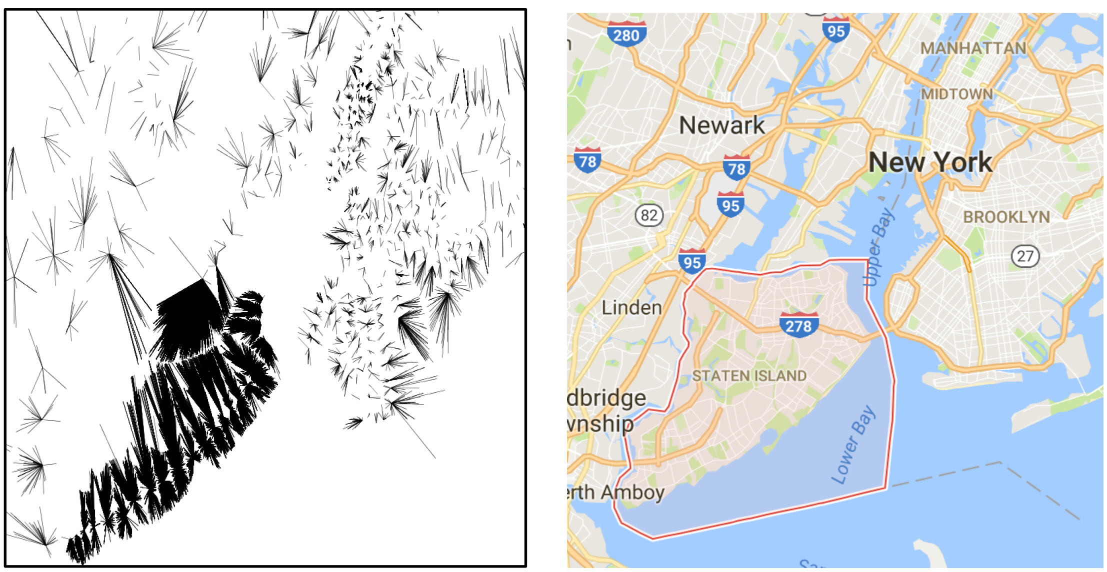


Figure 4.11

I included a close-up of the subway line that goes to Staten Island to show why there is such a cluster there.

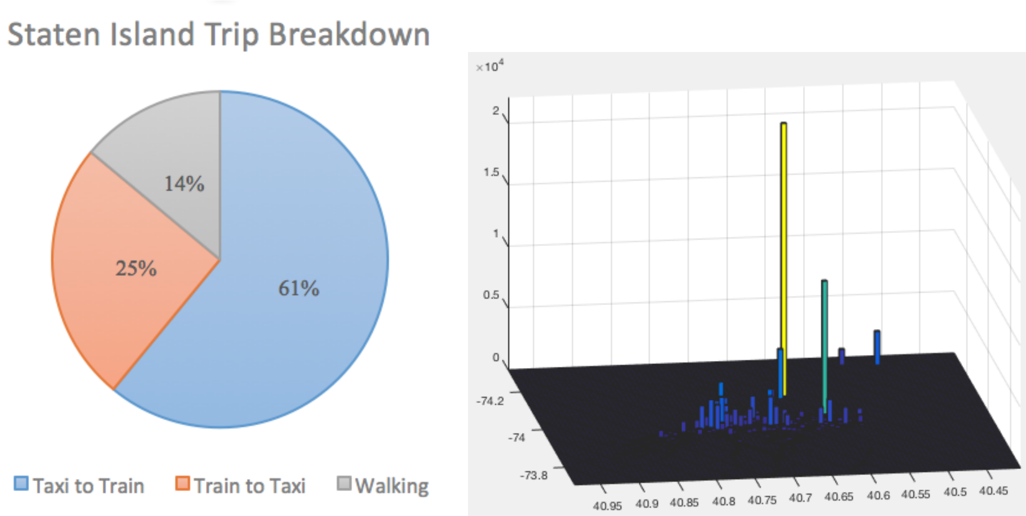


Figure 4.12: State Island Trip Breakdown and 3D Histogram

New York State Commuters (Total Trips = 932,012):

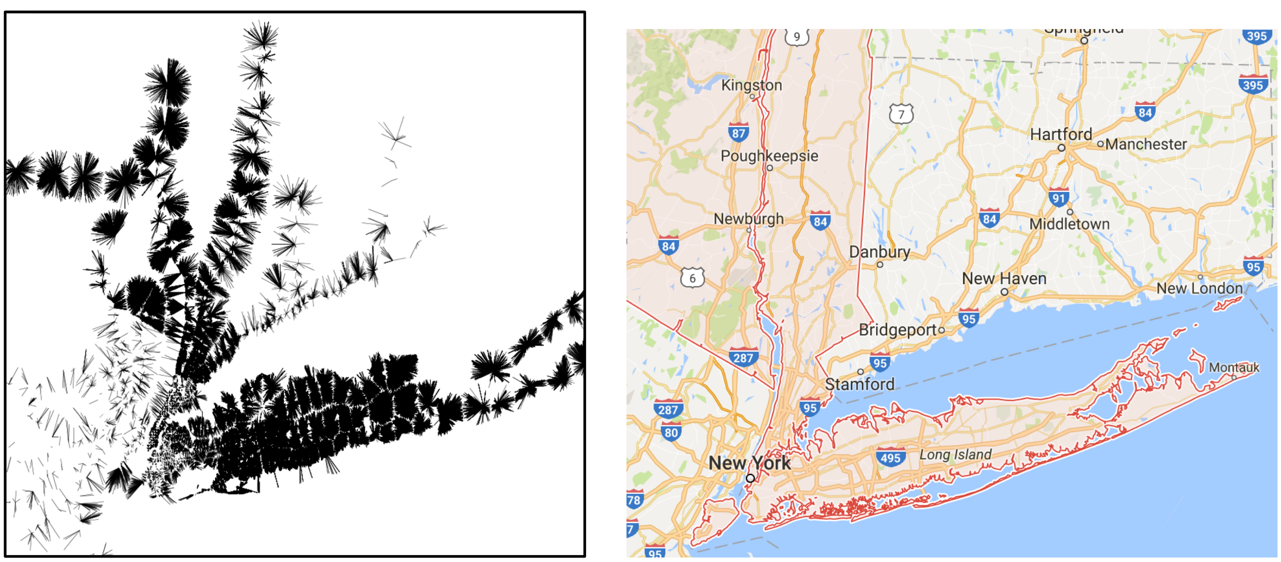
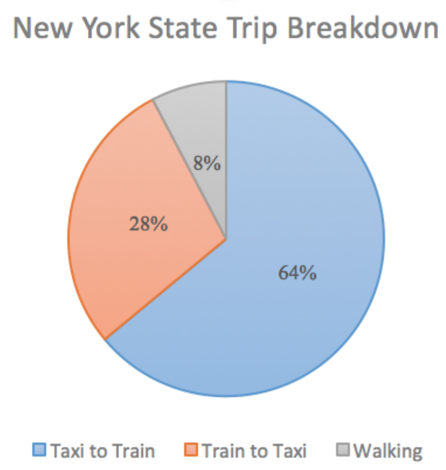


Figure 4.13



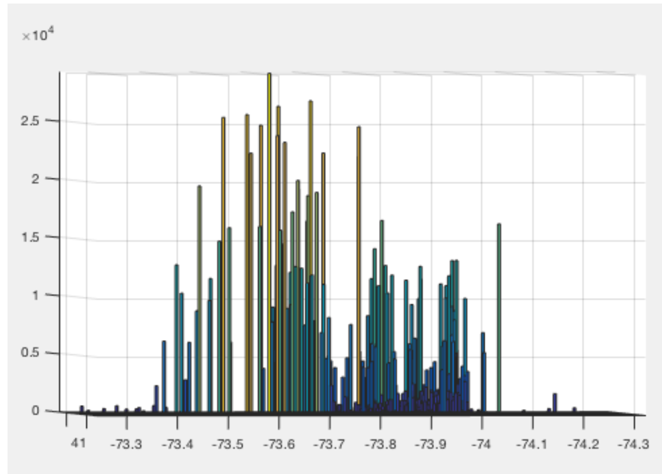
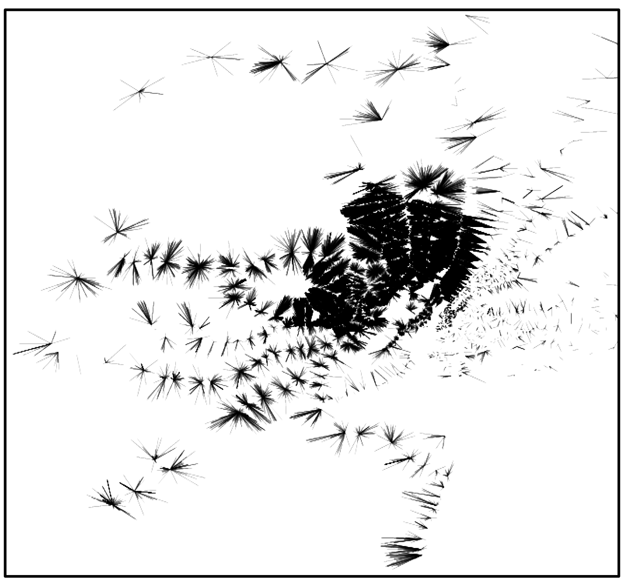
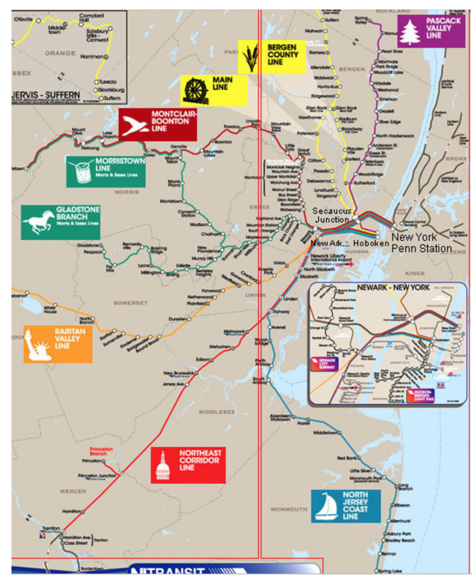
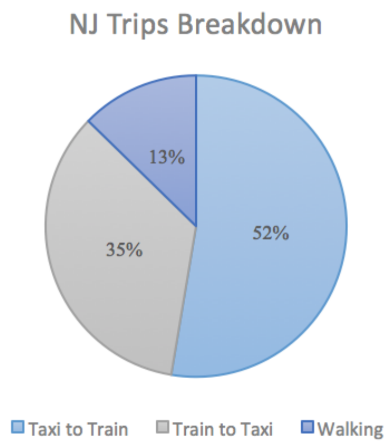


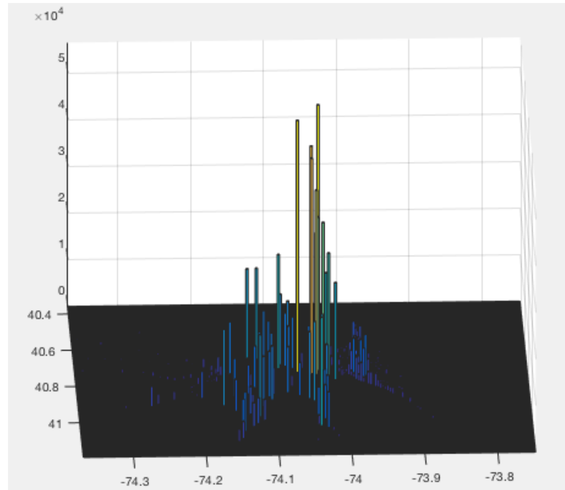
Figure 4.14: New York State trip Breakdown and 3D Histogram

New Jersey Commuters (Total Trips = 1,144,614)



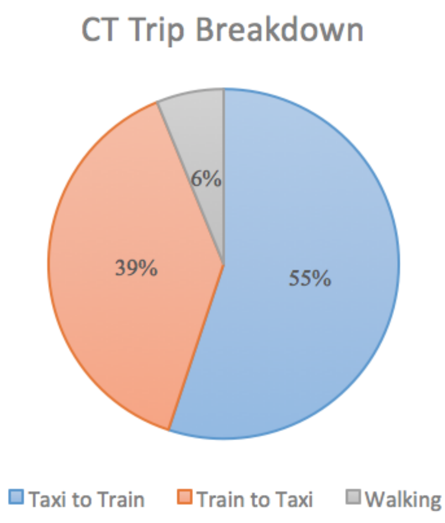


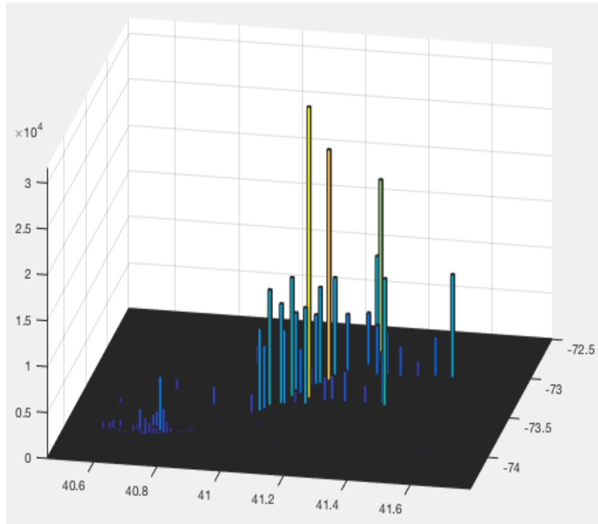




New Jersey 3D Histogram

h. Connecticut Commuters (Total Trips = 180,324)





CT 3D Histogram

Section 2.2 Overview of Results

Though the rail systems in the New York City metropolitan area are arguably some of the most heavily used and well-designed systems in the country, our analysis showed that they still have more potential to serve trips, with a 62.4% overall increase in ridership across all 8 geographical regions and 5 major train lines.

Train Line	Daily Ridership (Weekday)
New York City Subway	5,650,610
NJTransit	940,877
MetroNorth	298,900
LIRR	337,800
PATH	276,417
Total	7,504,604

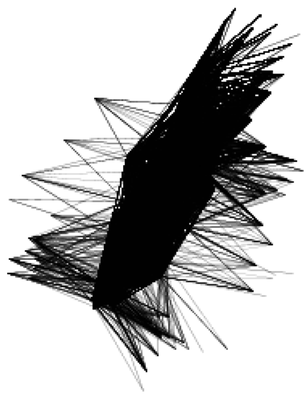
Geographical Region	Daily Number of Transit Trips (Weekday, not counting how many train lines the rider would use to complete the trip)
Manhattan (36061)	3,490,832
Bronx (36005)	1,558,771
Brooklyn (36047)	3,056,045
Queens (36081)	1,710,629
Staten Island (36085)	141,607
New York State (many)	932,012
New Jersey (many)	1,114,614
Connecticut (09001, 09005, 09009)	180,324
Total	12,184,832

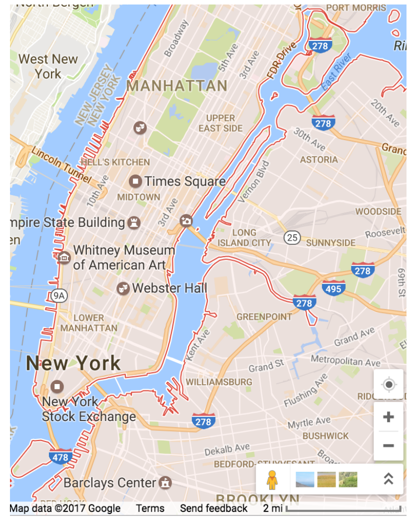
	Transit Only	aTaxi Pickup	aTaxi Drop off	% of all trips originating in this area that is a Transit Trip
Manhattan	45%	34%	21%	66.70%
Manhattan(Subway only)	45%	31%	24%	66.70%
Bronx	39%	28%	33%	80.30%
Brooklyn	38%	30%	32%	45.40%
Queens	27%	32%	41%	61.00%
Staten Island	14%	25%	61%	3.30%
New York Commuters	8%	28%	64%	5.20%
New Jersey Commuters	13%	35%	52%	11.22%
Connecticut Commuters	6%	39%	55%	3.30%

Finally, we looked at special "short-rail" trips in the 5 boroughs of New York City, which had to satisfy the following constraints: 1) between 0.5 and 2 miles, and 2) Both the origin and destination pixel must be in the same pixel as the origin and destination train station (i.e. the rider must walk to the train station on both ends). We found 1,917,852 of these trips (out of 9,957,884 transit trips in all 5 boroughs). This number may appear low, but then we realized how many subway trips in Manhattan actually travel more than 2 miles, and how very few trips in Queens/Staten Island (which are more suburb-like) would be under 2 miles.

Borough	Short-Rail Trips	Total Transit Trips	% Short-Rail Trips
Manhattan	768,872	3,490,832	22.03%
Bronx	351,924	1,558,771	22.58%
Brooklyn	570,817	3,056,045	18.68%
Queens	216,017	1,710,629	12.63%
Staten Island	10,222	141,607	5.22%
Total	1,917,852	9,957,884	19.26%

We attempted to whisker plot these short trips, but because many of them are overlaid over the same small area, the graph does not look very much like New York City. In the lower right hand corner, you can see the scale on Google Maps that shows how far 2 miles is.





Section 3. Conclusions

The four boroughs of New York City besides Staten Island (Manhattan, the Bronx, Queens and Brooklyn) exhibited similar patterns while the commuting trips from each state and Staten Island were similar. Though a borough of New York, Staten Island is more similar to the suburbs in traffic pattern and trip breakdown. Staten Island is not directly incorporated into the New York subway station. Instead, passengers must use a Ferry at St.George's Terminal at the North end of the island to reach Manhattan. There is a light rail line that runs through the North-South axis of the island but trains on this line only run from the South end to the North end in the morning and vice versa in the afternoon to accommodate work commutes.

For the four boroughs the most popular types of mode split are in order: Transit only, aTaxi Pickup and aTaxi Drop off. Because we only looked at trips *originating* in each borough, this makes sense because most origins and destinations within the city are within a walking distance from a subway station. There is more aTaxi Pickup than aTaxi Drop off trips because trips originating in the city are more likely to have an origin within walking distance from a station while their destination could be located outside of the city, where an aTaxi is needed to take them from their terminating station to final destination.

Trips originating in the three commuting areas and Staten Island present opposite trends. The most popular types of mode split are in order: aTaxi Drop off, aTaxi Pickup and Transit only. Following the same logic as above, it makes sense that there is only a small portion of trips originating in these suburban areas that have origins within a walking distance from a transit station. The typical trip would be of the morning commute type where the passenger rides aTaxi from his or her home to the nearest transit station from which he or she takes the train into the city for work. The passenger would then walk from his or her destination station (probably a subway station) to his or her workplace.

The four boroughs with subway access have a high proportion of trips affected by the inclusion of transit system. Notably in the Bronx, 80.3% of all trips can be classified as Transit Trips. Although Manhattan has denser transit coverage it also has a much higher proportion of short trips (those less than 0.5 miles) that we relegated to walking trips. It would make sense that people rely more on the subway system than on walking in the Bronx than in Manhattan. In the suburban regions, the percentage of trips affected by transit are much lower. Even in the New York City borough of Staten Island, this percentage is as low as 3.3%. The percentage of trips in New Jersey affected by transit is relatively high compared to New York State and Connecticut because the counties we analyzed in New Jersey is much more densely covered by New Jersey Transit than New York or Connecticut is by their respective transit systems.

The total number of transit trips originating in Manhattan was 3,091,551 when we only included subway stations while there were 3,490,832 trips when we included all four transit systems in our analysis. Including the mid distance transit systems only increased the number of transit trips by 11%. This means that the vast majority of transit trips in Manhattan are intracity. The difference in mode split between the two different analysis for Manhattan did not produce significantly different results. There was proportionally a slight increase in the number of aTaxi drop off trips with a corresponding decrease in aTaxi pickup trips when we only included subway stations. This could be because the mid distance transit systems involve more destinations that require aTaxi pickups.

Section 4. Challenges and Lessons Learned

The biggest challenge we faced was the scale of the problem at hand. The raw data in the trip files for relevant counties we looked at totaled 8.72GB. When we tried to run more trips at once to get more symmetric Train-to-Taxi and Taxi-to-Train breakdown in the pie charts, our computers both ran out of “Java heap space” for trying to run just 2 boroughs of New York City at once -- this is the reason why our analysis is broken down by geographical region, rather than by train line (there were simply too many trips to combine them into one file, and run them all at once). Besides the pure size of the data, the problem requires us to find the nearest transit station to the origin and destination out of 952 possible candidates. The calculations involved in this optimization created a bottleneck. Given the size of the problem, our original plan of implementing Trip boarding and Trip departure matrices in MATLAB was unfeasible.

The problem was resolved with the help of Evan Wood and Elizabeth Haile who were working on the same problem for the West Coast data set. Using their method of implementing KD trees in Java the runtime became manageable. We also considered using Princeton’s supercomputer Della for computation but this was unnecessary with the new method. KD trees are a more efficient data structure and Java is more efficient than MATLAB because of less built in overhead.

We also found the pixelation method we’ve used in class that relegates each 0.25 mile square area to one pixel to be unreasonable for the densely populated area of New York City. For perspective, Manhattan is 22.87 square miles with 151 unique subway stations. About 120 of these stations are clustered in the lower part of the island, which results in more than one station in each pixel on average. We resolved this issue using exact GPS coordinates instead of pixelation. We recommend that more precise pixelation be used in future studies.

Section 5. Acknowledgements

We would like to give thanks to Elizabeth Haile, Evan Wood, Jamie Cuff, Jarret Lowell, and Alex Yablonski for collaborating with us on this project and allowing us to use their code. We would also like to acknowledge Professor Kornhauser for the excellent food provided during the symposium as well as the support and feedback throughout the process.

Chapter 5:
Transit Trips East of Chicago (excluding NY, NJ, CT)

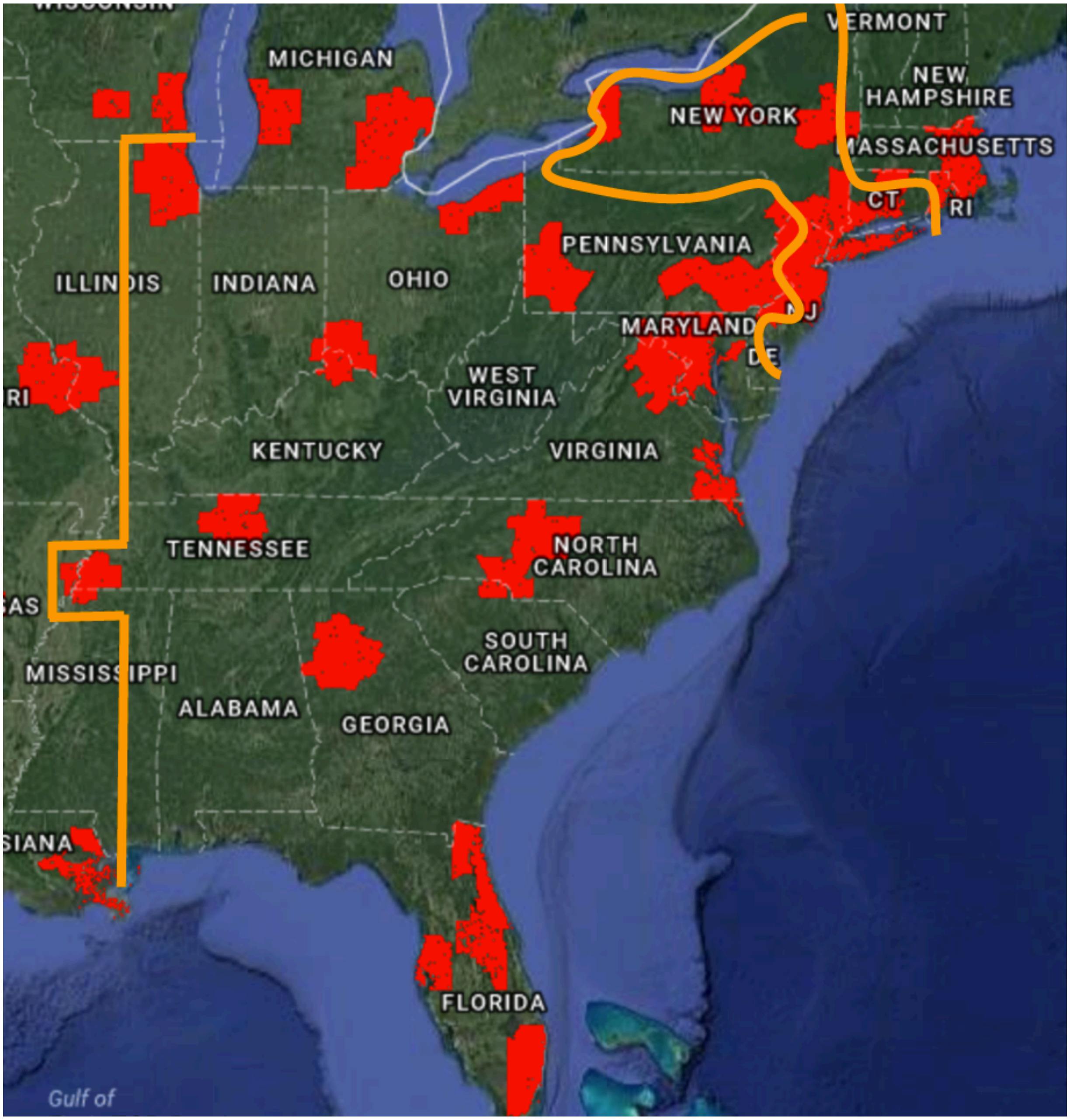
This chapter analyzes ridership potential on existing transit, passenger, and light rail systems in the Eastern United States using a modal split model to compare to current transit ridership. To begin, the region we analyzed is everything including Chicago going East in the contiguous United States, excluding NY, NJ, and CT. A line was drawn down the middle of the country through Chicago, allowing our region to include Michigan, eastern Illinois, Kentucky, Tennessee, Mississippi, and everything East of that, excluding the New York tri-state region. Figure 5.1 below, courtesy of the TOD online database, gives a visual of the region analyzed, with the "red" areas indicating the transit systems in the metro-city areas.

James Cuffe, Alex Yablonski, Jarret Lowell

Abstract

This chapter analyses the impact of aTaxis on transit trips in all 20 cities in the region from Chicago and east. We analysed Professor Alain Kornhauser's and Kyle Marocchini's database of the billion daily trips in conjunction with the Transit Oriented Development Database of the rail network in the US using multiple mode split, ride sharing and visualisation scripts. As a result, we were able to produce mode split, purpose split, time of day, trip length distributions, and ride-sharing analysis for transit trips in the East. Our final deductions showed that aTaxis would increase transit utilisation by on average 2.43 times in terms of passenger miles travelled and 3.12 times in terms of unlinked trips. The financial effects of this are an increase in operating expenses by 2.43 times and in fare revenues by 3.12 times leading to a 112% decrease in profits from transit systems on average across the East. When aTaxis are implemented as an efficient means for transporting individuals to and from transit stations, a change in the current allocation of funding to existing transit systems will be needed to address this larger loss resulting from increased transit utilisation.

Figure 5.1: A Map of the Chicago and East Area (Excluding NY, NJ and CT)



From TOD Database

Within our region, we focused on the transit systems of the following 20 metro-city areas: Atlanta, Baltimore, Boston, Charlotte, Chicago, Cleveland, Cincinnati, Detroit, Grand Rapids, Greensboro, Harrisburg, Jacksonville, Memphis, Miami, Nashville, Orlando, Philadelphia, Pittsburgh, Tampa, Washington D.C.

The Data Set

Within the 20 cities mentioned in the previous section, we determined all the counties within each city region that are served by the transit system in the metro area. This amounted to an analysis of 137 counties across the 20 major cities in the analysis. The rest of the counties in the Eastern United States were either in too remote of a location to be served by transit and light systems (greater than 5 miles from any transit station in our network) or were greater than 200 miles outside of a major city region. That is, if the centroid of the city was greater than 200 miles from the nearest transit station, we did not deal with those trips. It is assumed that person trips originating in counties not having a transit station within 200 miles have no opportunity to be served by existing rail transit. The analysis of these cities is left for the aTaxi ride-sharing and Amtrak (long-trip) sections of this report. We analyzed the trips with great circle distance between 2.0 and 200 miles generated by Marocchini and Kornhauser. Within each region we analyzed, there was one major city and at least one county. Each county had at least one file, which was pixelated into ½ mile by ½ mile squares, based on Latitude and Longitude calculations. We compiled two files to read data into our program. First, a city file listed all files from the Kornhauser database pertaining to the counties within the city. Then, a transit file listed all train stations from the TOD database within 200 miles of the city. Please refer to the following link to see what the transit file from Atlanta looked like: https://drive.google.com/file/d/0B-zTKW2_9Lnma1BKdEh4TTNVS00/view?usp=sharing. These were then pixelated to give a vector set of all of the pixels containing transit stations. For example, this is the Atlanta vector set: https://drive.google.com/file/d/0B-zTKW2_9Lnmek1NcTJueE40STA/view?usp=sharing.

How the Code Worked

Using Matlab as the coding environment, we looped through all the county files for one of our 20 cities and outputted a Matlab workspace for each county file, containing the following analysis:

1. Sorting of all trips by trip type (use transit, transit to aTaxi, aTaxi to transit, and pure aTaxi)
2. Recording of all relevant information for each trip, including tour type, origin and destination identity, and information about train stations visited, with their location and information about the train lines

Use transit trips recorded all trips that originated within a 10 minute walk the nearest transit stations to their origins and destinations. To simulate a 10 minute walk between the origins, destinations, and train stations, we constructed a super pixel with the train station in the center pixel and the origin/destination in one of the 9 squares in the super pixel (refer to Figure 5.2 below). This type of trip assumes that the traveler will walk up to 10 minutes to the nearest transit station, take transit to the station nearest to their destination, and walk up to 10 minutes to their final destination.

aTaxi to transit trips recorded all trips in which the origin of the trip was within 10 miles (to be services by an aTaxi) of the nearest train station and the destination was within a 10 minute walk of the nearest transit station (refer to Figure 5.2 below).

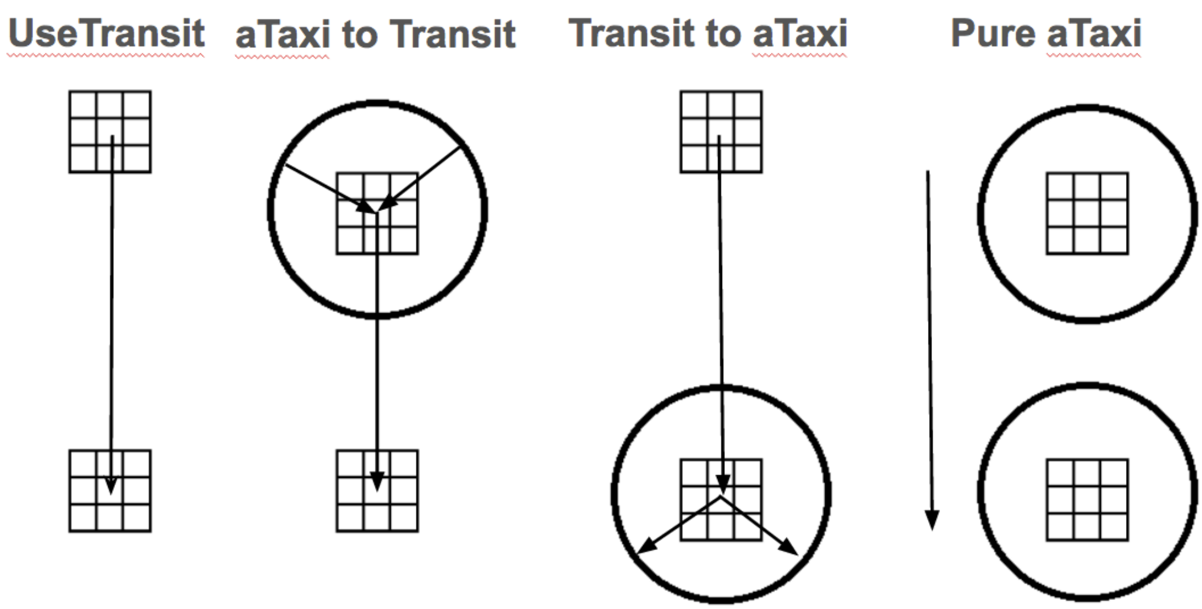
Transit to aTaxi trips recorded all trips in which the origin of the trip was within a 10 minute walk of the nearest transit station and the destination was within 10 miles of the nearest transit station (refer to Figure 5.2 below).

Pure aTaxi trips recorded all the rest of the trips: that is, trips whose origins and destinations were both greater than 10 miles away from the nearest transit stations (refer to Figure 5.2).

The Pixelization process was one of the many keys to success throughout this project. The dataset we examined was simply too large to be run all at once, so we split up the data into .25 mi^2 regions (.5 mi on a side). Based on an origin in Eastern New Jersey, we split the nation into these pixels, and then included pixels uniquely into counties. These counties, together with the PersonTrip files, composed a region about one city. We then had a structure that contained a whole city's worth of trips, broken down into tiny, discreet, and unique pixels. This was matched by the train network file we built, which contained every train station within a 200mi radius of the city center. This idea was nothing short of necessary, as it would be computationally impossible to run through the entire dataset of about 1 billion trips for each of 2555 train stations, for each of 20 cities. Many thanks to Prof. Kornhauser for providing the formula needed to convert latitude and longitude coordinates into pixels.

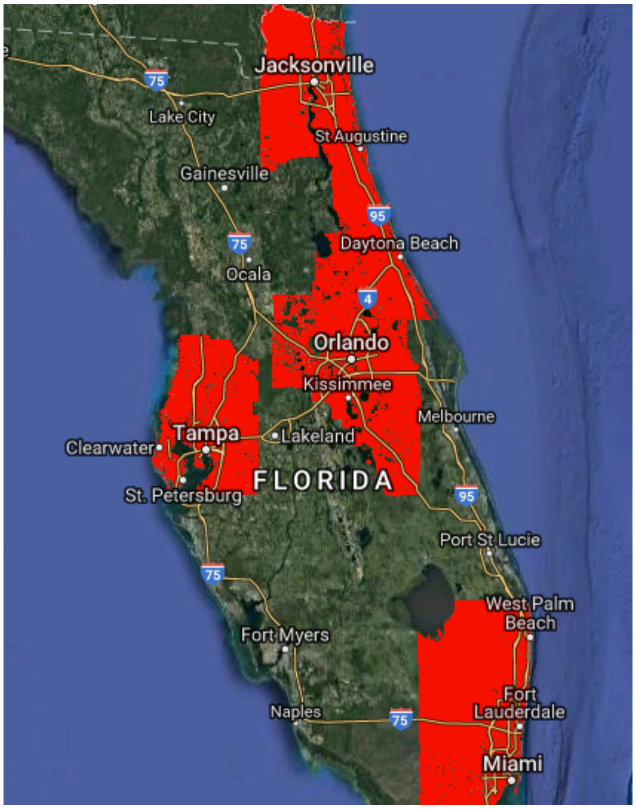
e

Figure 5.2: The Modes of Travel Analysed and The Super Pixel Method



A Look at an Example County: Tampa County 12057

Figure 5.3a: A Map of The Florida Transit System



From TOD Database

To better illustrate the output generated by our program, we include several figures that summarize the activity of transit trips throughout a typical day. The city of Tampa is composed of a single county, 12057, and we analyzed part two of the data (there were four parts).

To begin, we look at the trips that use transit only. Figure 5.3 displays the output from our Matlab program. The figure shows a table of output. The row of data under each of the headings represents a single trip and all the information relevant to that trip. This trip represented an other to home trip from a local Wendy's restaurant to Home. Our data recorded the origin x and y pixels and destination x and y pixels, along with the pixelation and identity data of the two train stations visited during the trip. The use transit trip in Figure 3 is interpreted in the following way: the traveler walks from Wendy's, pixel (-748,-1516), to the Dick Greco Plaza/Transportation Center, pixel (-747,-1515), takes transit to the Cadrecha Plaza Station, pixel (-746,-1512), and then walks to his final destination, home, at pixel (-747,-1511).

Figure 5.4 - Example Use Transit Trip in Tampa County 12057 part 2

OFIPS	OLon	OLat	OXCoord	OYCoord	ODepartureTime	DFIPS	DLon	DLat	DXCoord	DYCoord	GCDistance
12057	-82.459572	27.936865	-748	-1516	5305	12057	-82.454606	27.971643	-747	-1511	2.424665
OType	OName	DType	DName								
'O'	Wendy's'	'H'	'Home'								
TransitOLat	TransitOLon	TransitOYCoord	TransitOXCoord								
27.94155717	-82.454825	-1515	-747								
Buffer	Agency	Lines	StationName	YearOpened							
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'							
TransitDLat	TransitDLon	TransitDYCoord	TransitDXCoord								
27.96067524	-82.445515	-1512	-746								
Buffer	Agency	Lines	StationName	YearOpened							
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'							

In a similar way, we output an example of aTaxi to transit, transit to aTaxi, and pure aTaxi trips, interpreted in the same manner as described in the previous paragraph. These next several figures include several example trips. Each row corresponds to a distinct trip.

Figure 5.5a - Example of 7 aTaxi to Transit Trips in Tampa County 12057 part 2

OFIPS	OLon	OLat	OXCoord	OYCoord	ODepartureTime	DFIPS	DLon	DLat	DXCoord	DYCoord	GCDistance
12057	-82.504954	28.015283	-752	-1505	2020	12057	-82.467417	27.948155	-748	-1514	5.17854843
12057	-82.498355	28.011888	-752	-1505	9918	12031	-81.662187	30.334361	-661	-1184	168.395045
12057	-82.498355	28.011888	-752	-1505	10060	12031	-81.662187	30.334361	-661	-1184	168.395045
12057	-82.498449	28.013967	-752	-1505	14690	12057	-82.463067	27.951329	-748	-1514	4.84182256
12057	-82.498449	28.013967	-752	-1505	20639	12095	-81.369118	28.576753	-629	-1427	79.0337014
12057	-82.496616	28.012812	-752	-1505	22488	12057	-82.42495	27.958398	-744	-1513	5.77312016
12057	-82.498355	28.011888	-752	-1505	25846	12057	-82.442306	27.956318	-746	-1513	5.14742481
OType	OName	DType	DName								
'O'	'UNIVERSITY	'H'	'Home'								
'H'	'Home'	'S'	'FLORIDA COMMUNITY CLG-JCKSNVLL'								
'H'	'Home'	'S'	'FLORIDA COMMUNITY CLG-JCKSNVLL'								
'H'	'Home'	'O'	'TAMPA BAY PERFORMING ARTS CTR'								
'H'	'Home'	'S'	'FLORIDA HOSPITAL'								
'H'	'Home'	'O'	'PCL CIVIL CONSTRUCTORS INC'								
'H'	'Home'	'W'	'KIMMINS CONTRACTING CORP'								

Figure 5.5b - Information about the Train Station where the aTaxis Drop off their Passengers, each row corresponding to the ones in Figure 5.5a

TransitOLat	TransitOLon	TransitOYCoord	TransitOXCoord	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
27.96067524	-82.445515	-1512	-746	
Buffer	Agency	Lines	StationName	YearOpened
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'

Figure 5.5c - Information about the Train Station where the Passengers get off Transit and Walk to their Destinations, each row corresponding to the ones in Figures 5.5a and 5.5b

TransitDLat	TransitDLon	TransitDYCoord	TransitDXCoord	
27.94155717	-82.454825	-1515	-747	
30.327	-81.6623	-1185	-661	
30.327	-81.6623	-1185	-661	
27.94155717	-82.454825	-1515	-747	
28.5741	-81.3731	-1428	-629	
27.96103535	-82.437268	-1512	-745	
27.96067524	-82.445515	-1512	-746	
Buffer	Agency	Lines	StationName	YearOpened
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Existing Transit'	'JTA '	'Skyway'	'Central'	'Pre-2000'
'Existing Transit'	'JTA '	'Skyway'	'Central'	'Pre-2000'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Planned Transit'	'Central Flori	'Final Design'	'Florida Hospital Station'	2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Centennial Park Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'

Figure 5.6a - Example of 5 Transit to aTaxi Trips in Tampa County 12057 part 2, includes the location data of the Transit Station Visited first by each passenger

OFIPS	OLon	OLat	OXCoord	OYCoord	ODepartureTime	DFIPS	DLon	DLat	DXCoord	DYCoord	GCDistance
12057	-82.459854	27.936708	-748	-1516	242	12057	-82.482695	27.946021	-750	-1514	1.53722918
12057	-82.459068	27.937181	-748	-1516	337	12057	-82.44906	27.933445	-746	-1516	0.66394783
12057	-82.459854	27.936708	-748	-1516	395	12057	-82.490893	27.895157	-751	-1521	3.44377498
12057	-82.459068	27.937181	-748	-1516	507	12057	-82.452289	27.927404	-747	-1517	0.79308204
12057	-82.459854	27.936708	-748	-1516	552	12057	-82.464658	27.970999	-748	-1511	2.39001192

OType	OName	DType	DName
'O'	'TAMPA GENI	'O'	'HARDEMAN LANDSCAPE NURSERY'
'O'	'USFIVF'	'H'	'Home'
'O'	'TAMPA GENI	'H'	'Home'
'O'	'USF REGION/	'H'	'Home'
'O'	'TAMPA GENI	'H'	'Home'

TransitOLat	TransitOLon	TransitOYCoord	TransitOXCoord
27.94155717	-82.454825	-1515	-747
27.94155717	-82.454825	-1515	-747
27.94155717	-82.454825	-1515	-747
27.94155717	-82.454825	-1515	-747
27.94155717	-82.454825	-1515	-747

Figure 5.6b - Information about the Identities of the Transit Origination stations from Figure 5.6a and about the Transit Destination stations, each row corresponding to the trips in Figure 5.6a

Buffer	Agency	Lines	StationName	YearOpened
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Dick Greco Plaza/ Transportation Center'	'2002'

TransitDLat	TransitDLon	TransitDYCoord	TransitDXCoord
27.96067524	-82.445515	-1512	-746
27.96067524	-82.445515	-1512	-746
27.94573695	-82.445585	-1514	-746
27.96067524	-82.445515	-1512	-746
27.96067524	-82.445515	-1512	-746

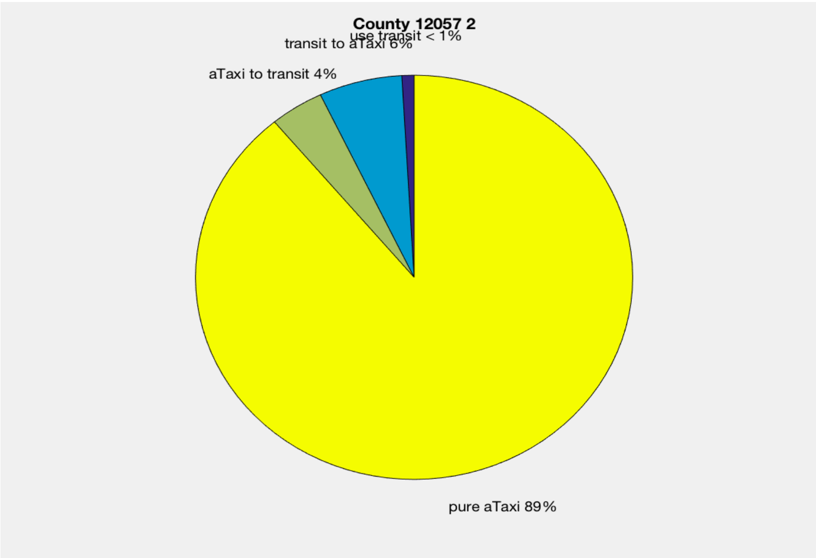
Buffer	Agency	Lines	StationName	YearOpened
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cumberland Avenue Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'
'Existing Transit'	'HART'	'Teco Line Streetcar System'	'Cadrecha Plaza Station'	'2002'

Figure 5.7 - Example of 5 Pure aTaxi Trips in Tampa County 12057 part 2

OFIPS	OLon	OLat	OXCoord	OYCoord	ODepartureTime	DFIPS	DLon	DLat	DXCoord	DYCoord	GCDistance
12057	-82.459068	27.937181	-748	-1516	211	12057	-82.379225	27.851916	-739	-1527	7.65565271
12057	-82.459068	27.937181	-748	-1516	287	12103	-82.636771	27.847183	-767	-1528	12.5210117
12057	-82.459854	27.936708	-748	-1516	293	12057	-82.605148	28.001711	-763	-1507	9.95003296
12057	-82.459854	27.936708	-748	-1516	566	12057	-82.320025	28.146413	-732	-1487	16.8308836
12057	-82.459068	27.937181	-748	-1516	799	12057	-82.410998	28.033718	-742	-1502	7.294537

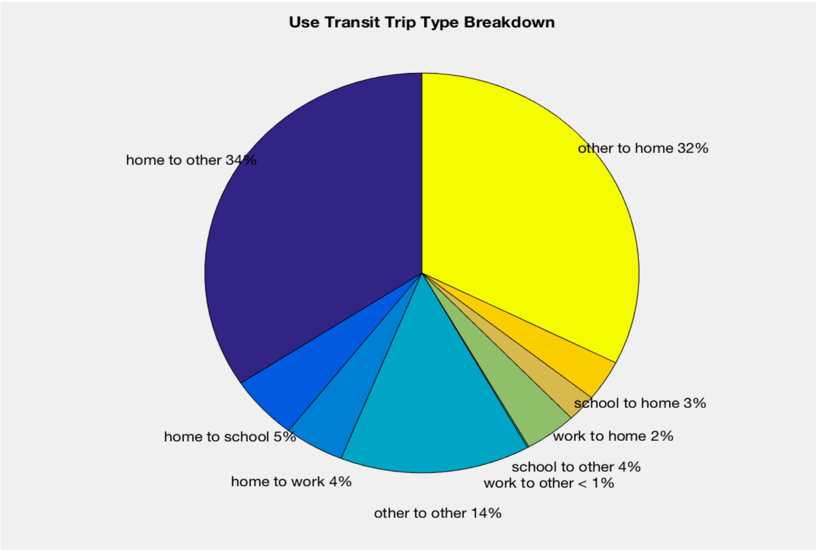
OType	OName	DType	DName
'O'	'USF PHYSICI/	'H'	'Home'
'O'	'US F PHYSICI	'H'	'Home'
'O'	'TAMPA GENI	'H'	'Home'
'O'	'TAMPA GENI	'H'	'Home'
'O'	'USF PHYSICI/	'H'	'Home'

Figure 5.8 - Pie Chart Summarizing Breakdown of Trip Types in Tampa County 12057 part 2



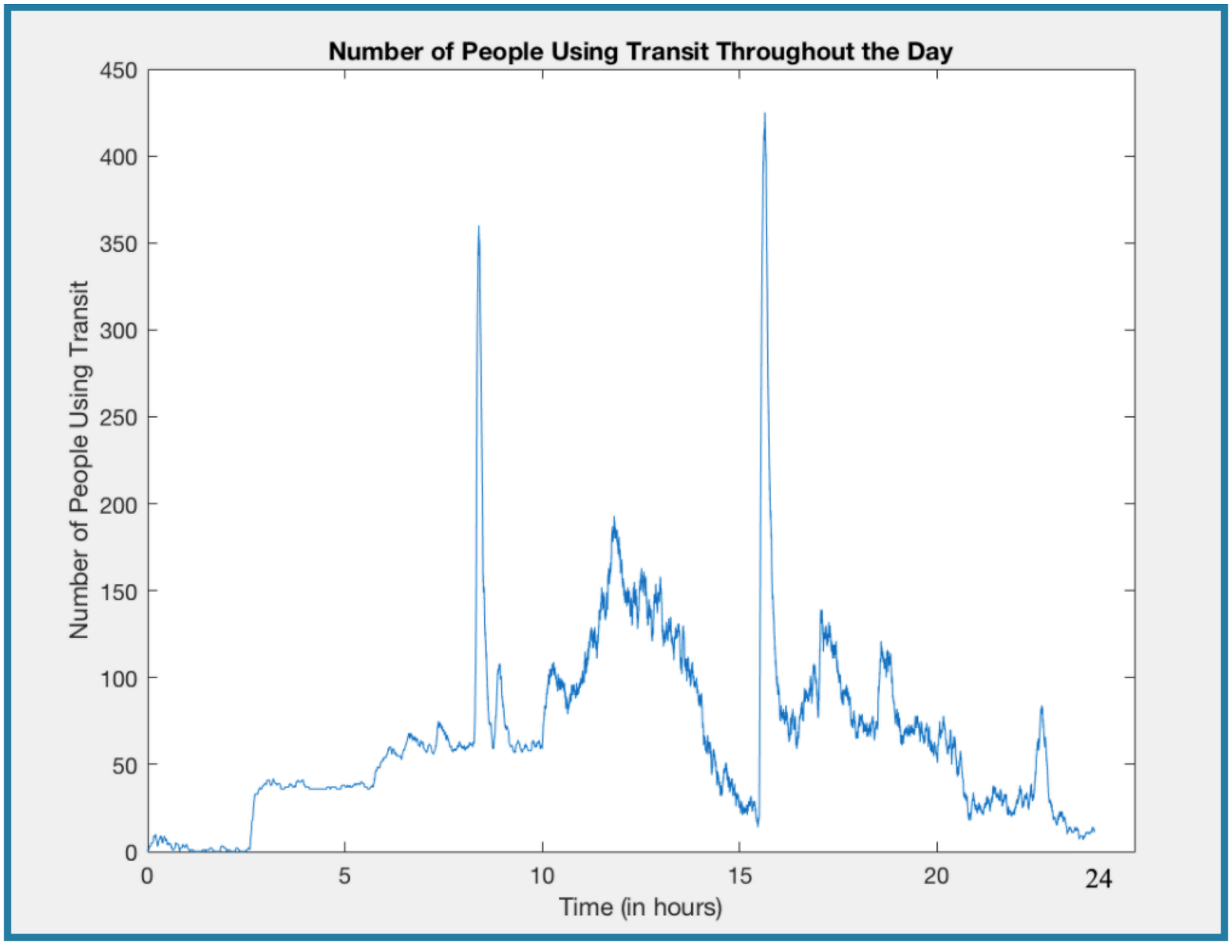
Most trips in Tampa go via the pure aTaxi mode. This is because there is a only a small subway in the center of Tampa as can be seen in Figure 5.3b. However, a significant proportion of transit trips use the transit to aTaxi and aTaxi to transit modes when compared to <1% using transit only showing that aTaxis do have an effect in increasing the number of trips that will utilize transit.

Figure 5.9a - Pie Chart Summarizing Breakdown of Tour Types in Use Transit Trips in Tampa County 12057 part 2



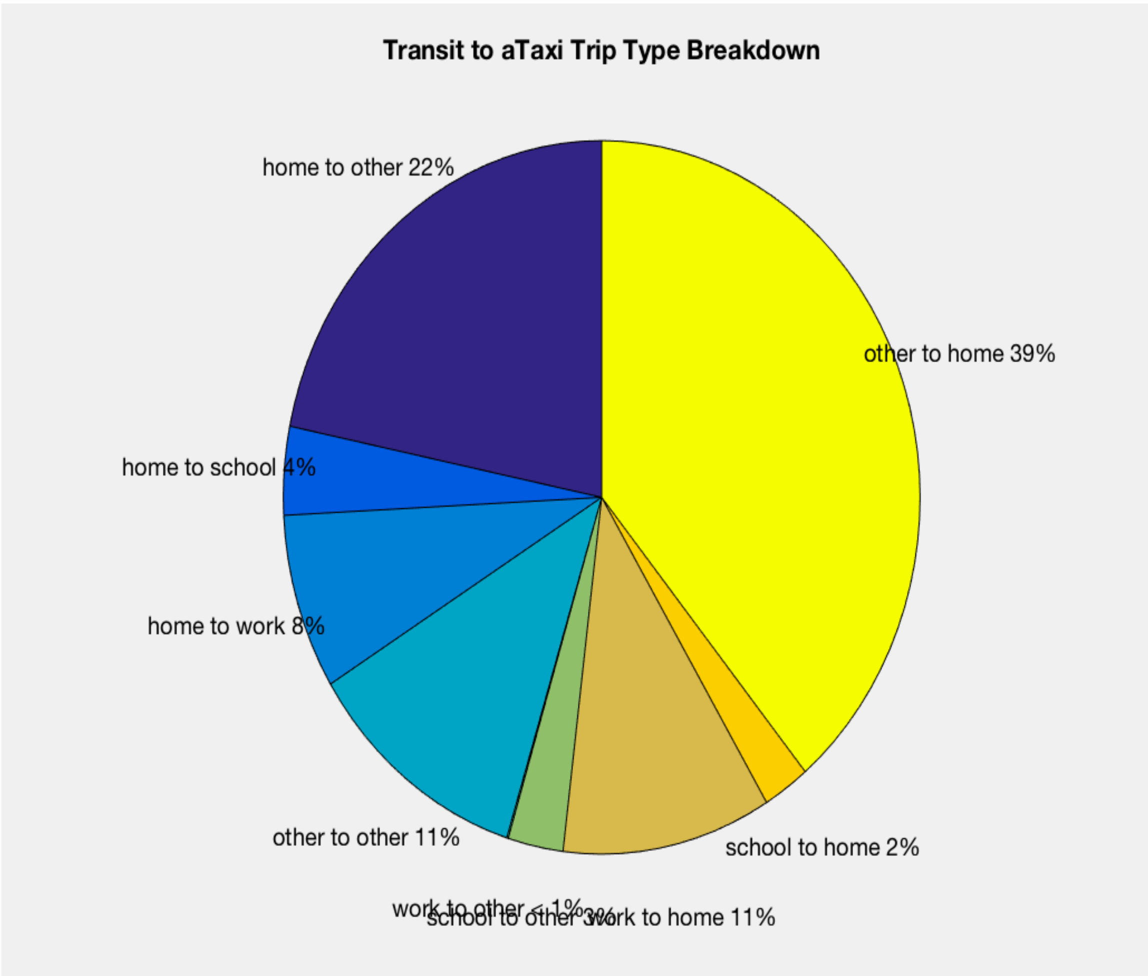
The majority of the transit trips go from home to other (34%) or other to home (32%). This means that the transit stations in Tampa are within walking distance of more residential and entertainment locations than schools or workplaces.

Figure 5.9b - Time Distribution of Use Transit Riders in Tampa County 12057 part 2



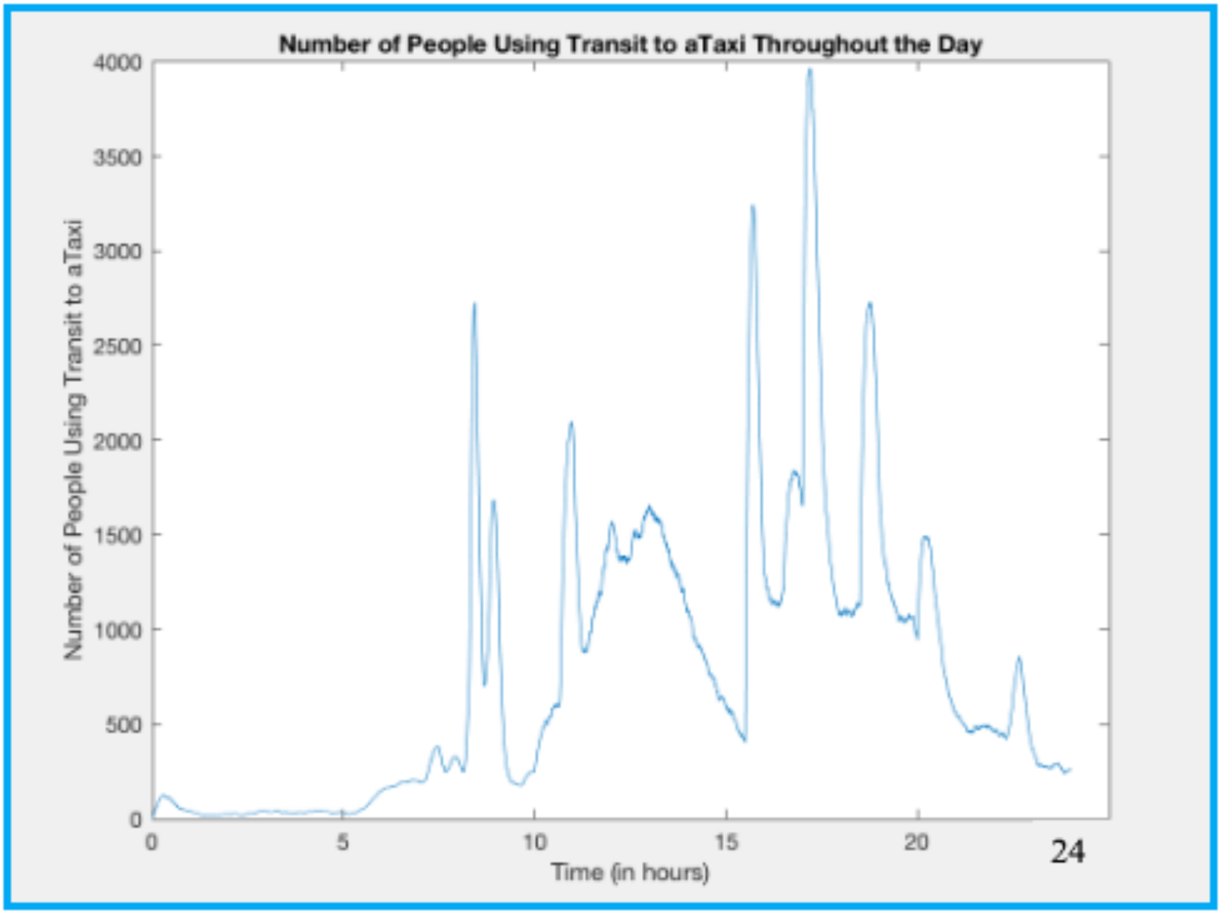
There is a distinctive morning peak around 8:30am for the start of work and an afternoon peak around 3:30pm when schools let out.

Figure 5.10a - Pie Chart Summarizing Breakdown of Tour Types in Transit to aTaxi Trips in Tampa County 12057 part 2



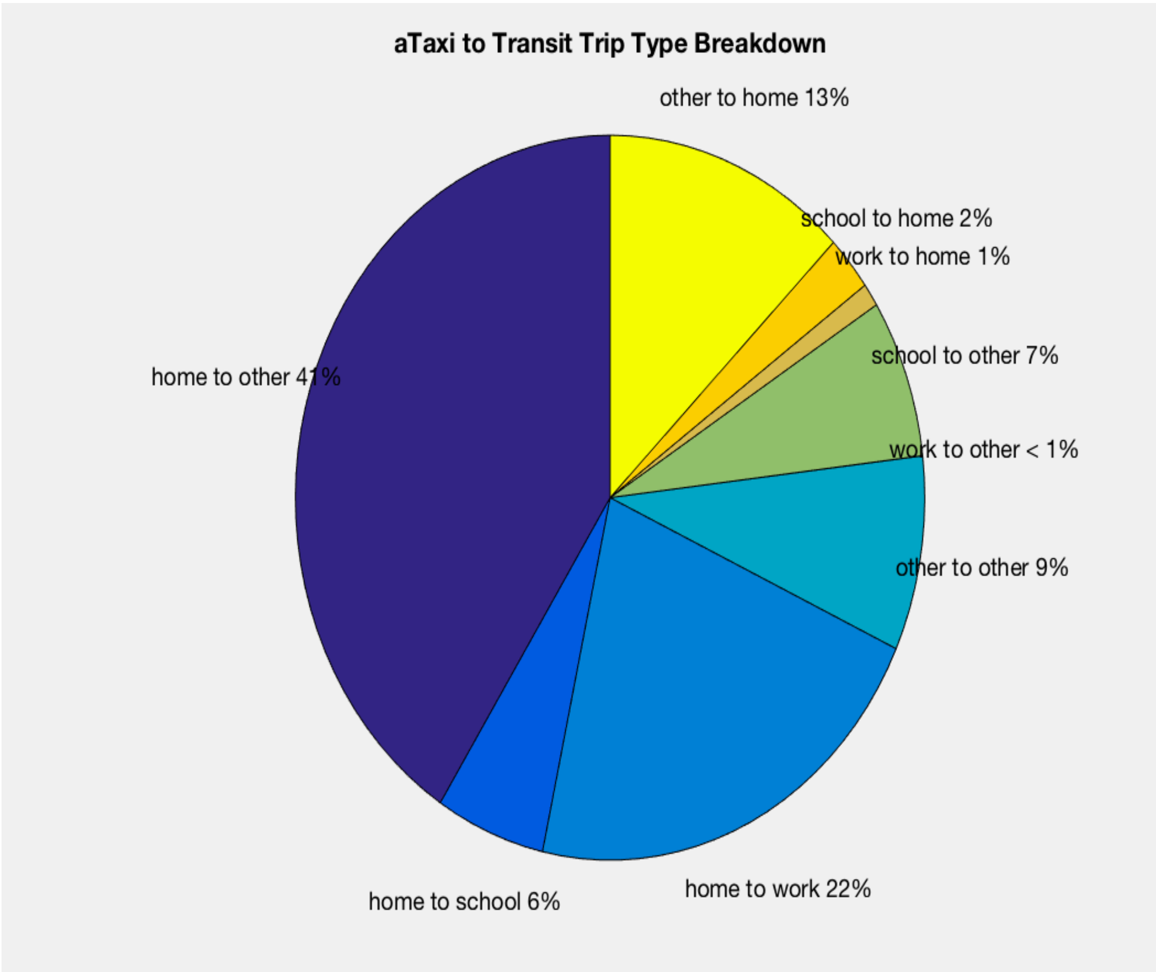
The majority of the trips for transit to aTaxi are other to home (39%). This means a large number of passengers are coming from entertainment areas where there are stations within walking distance and then travelling less than 5 miles in an aTaxi from the end station back to their home.

Figure 5.10b - Time Distribution of Transit to aTaxi Riders in Tampa County 12057 part 2



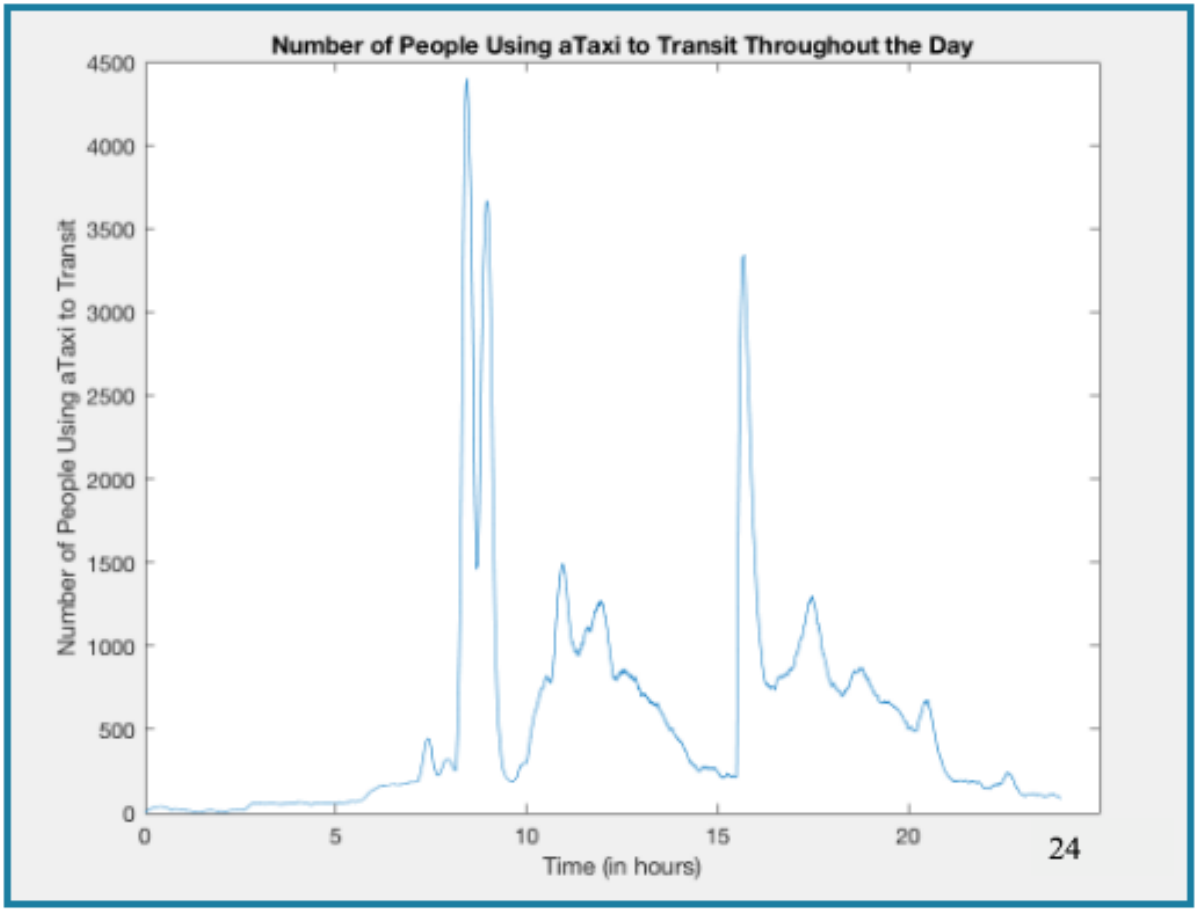
For transit to aTaxi trips, there is a smaller morning peak but a large set of peaks from 3:30pm to 8:00pm as people return home from school, work and dinner out.

Figure 5.11a - Pie Chart Summarizing Breakdown of Tour Types in aTaxi to Transit Trips in Tampa County 12057 part 2



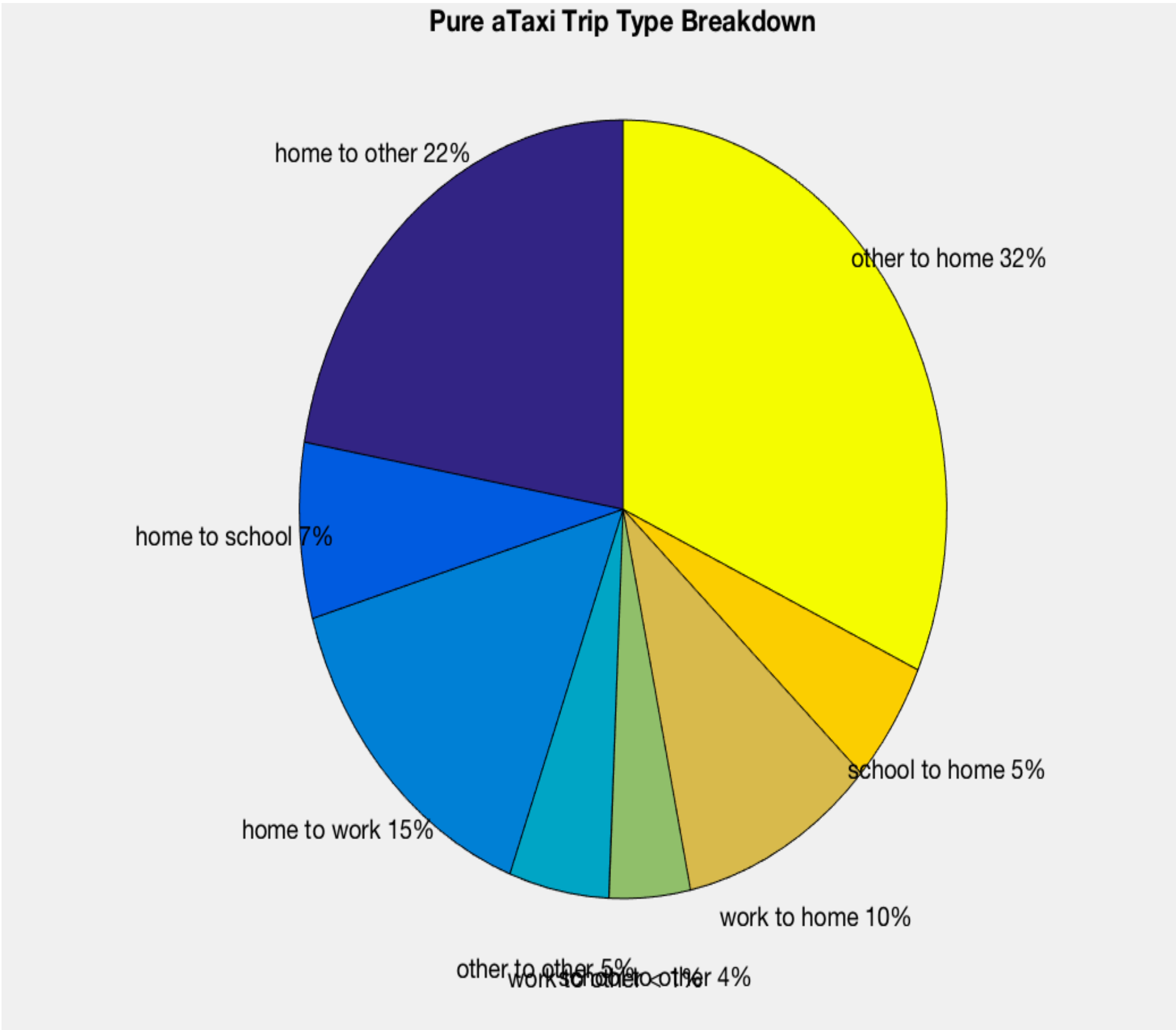
For aTaxi to transit trips, the majority of the trips are from home to other (41%) corresponding to the other to home trips that were the majority for the transit to aTaxi trips.

Figure 5.11b - Time Distribution of aTaxi to Transit Riders in Tampa County 12057 part 2



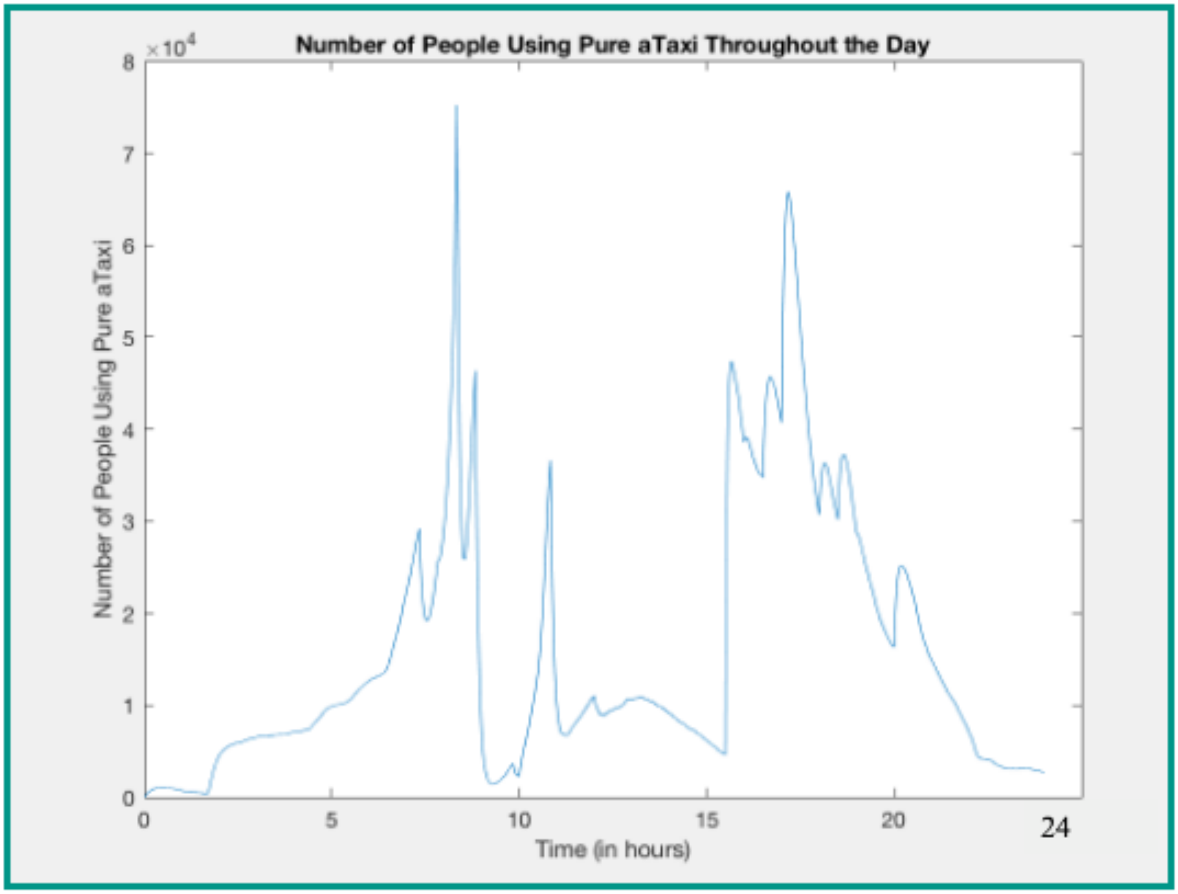
The largest peak is in the morning between 8:00am and 9:30am rather than the evening as was the case for the aTaxi to transit trips. This suggests that many of these aTaxis used to take people from their homes to the stations in the morning could be utilised again in the evening to take people back from the stations to their homes. However, there would need to be enough other trips during the day to keep them occupied during the rest of the day.

Figure 5.12a - Pie Chart Summarizing Breakdown of Tour Types in Pure aTaxi Trips in Tampa County 12057 part 2



The pure aTaxi trips actually mirror the transit to aTaxi trips in terms of their tour type break down with the majority being other to home. However, these trips either do not start nor end within 10 minutes walk of a station or one or both ends of the trip are more than a 10 mile aTaxi ride from a station.

Figure 5.12b - Time Distribution of Pure aTaxi Riders in Tampa County 12057 part 2



The number of aTaxis at each time throughout the day is almost a factor of 10 higher which corresponds to the 89% of trips going via the pure aTaxi mode. There is a very large and sharp peak in the morning at 8:00am for the work commute and then a smaller peak around 3:30pm when school let out and around 5:00pm for the end of the working day.

Takeaways from Tampa County 12057 Output

Our code provided a very rich dataset separating the trips in Tampa County 12057 into use transit only, transit to aTaxi, aTaxi to transit, and pure aTaxi trips, allowing us to break down the data into tour type and to take a look at the time distribution of each type of trip at every second of each day. Yet, because we constructed our code in Matlab, the run time was exceeding an hour for one file. Given more computing resources and faster run time, our Matlab code could output a very rich dataset on trips in the entire country, allowing us to break down the trips by tour type and potentially by most popular origins and destinations. However, due to the time constraints of this assignment and after some consultation with Professor Kornhauser, Evan Wood, and Elizabeth Haile, who analyzed potential transit trips in the Western United States, we were granted access to more efficient code from Wood and Haile, written in Java and were able to run it on each of the 20 cities we analyzed.

The output of the code of Wood and Haile run on our data sets differed from the output of our Matlab code in several key ways:

1. Output assumed 5 minute walking distance to transit stations and 5 miles within transit stations for aTaxi trips (ours assumed travelers would be willing to walk up to 10 minutes and that aTaxi trips would occur within 10 miles of the nearest transit station)
2. Output did not include pure aTaxi trips
3. Output did not include qualitative information about the trip (i.e, the identity of the origin and destination)
4. Output did not keep track of the qualitative information concerning the transit stations visited during the course of the trip. The output did however keep track of the tour taken by an individual, similar to ours (i.e, kept track of the latitude-longitude of the transit stations visited)

Nevertheless, the output included the number of unlinked trips, along with the number of people who made trips, and what type of trip each segment was (0 signified purely transit, 1 signified transit to aTaxi, and 2 signified aTaxi to transit)

<u>Legend for interpreting the output of Wood and Haile's Java script</u>
Column 1 X-Coord (Origin)
Column 2 Y-Coord (Origin)
Column 3 Long (Origin)
Column 4 Lat (Origin)
Column 5 X-Coord (Destination)
Column 6 Y-Coord (Destination)
Column 7 Long (Destination)
Column 8 Lat (Destination)
Column 9 Departure Time
Column 10 Trip Distance
Column 11 Arrival Time
Column 12 Type of Trip {0 = transit, 1 = transit to aTaxi, 2 = aTaxi to transit
Column 13 Trip Count

Supercomputers and Run Time Efficiency

To get our output run in time we had set up directories on the Della research computing cluster. Using a bash file that Mr. Michael Bino created, we were able to download files more efficiently, which streamlined the run process for both the Matlab and Java implementations.

Although our code followed a similar logic methodology, the more lightweight Java environment as well as unlinking the trips and decreasing the information output meant that the new code ran in linearithmic rather than polynomial time. One other major improvement to the efficiency came from using my KD-Tree implementation from COS 226. The KD-Tree is a data structure implementation that uses a symbol table implementation with 2-D points as the key to enable efficient nearest neighbour search on a two dimensional map. These efficiency gains meant we could run the code locally.

Velocity Assumptions

We made the following assumptions about the velocity of each mode to calculate travel and arrival times:

- 3 mph for Walks
- 30 mph for aTaxis
- 40 mph for Transit

A Look at an Example City: Atlanta

To illustrate the output generated by our visualization program on the data run on the code of Wood and Haile, we offer a case study of the transit system analysis in Atlanta. Below the following maps, which offer a detailed view of the transit stations in the metro-Atlanta area, we include a pie chart summarizing the breakdown of transit trips by type in Atlanta and display both the time distribution graphs of trip type as well as the cumulative distance distribution of trips by trip type.

Figure 5.13a: A Map of The Transit Systems in The Atlanta Area

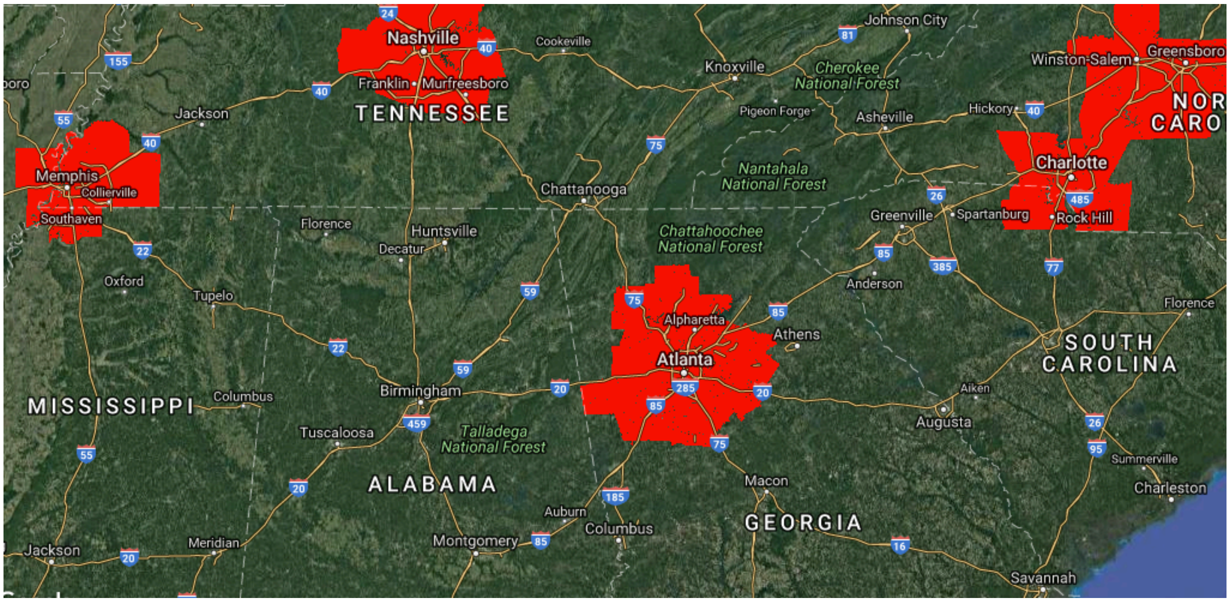


Figure 5.13b: A Map of The Transit Stations in Atlanta

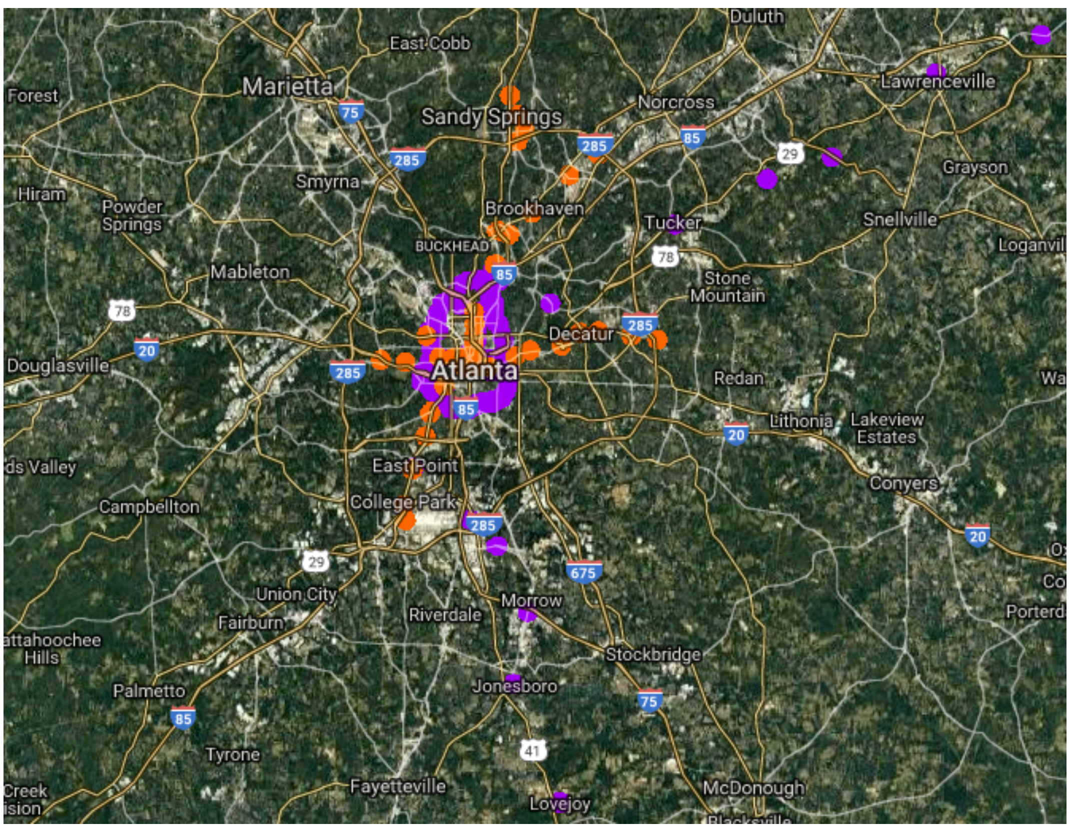
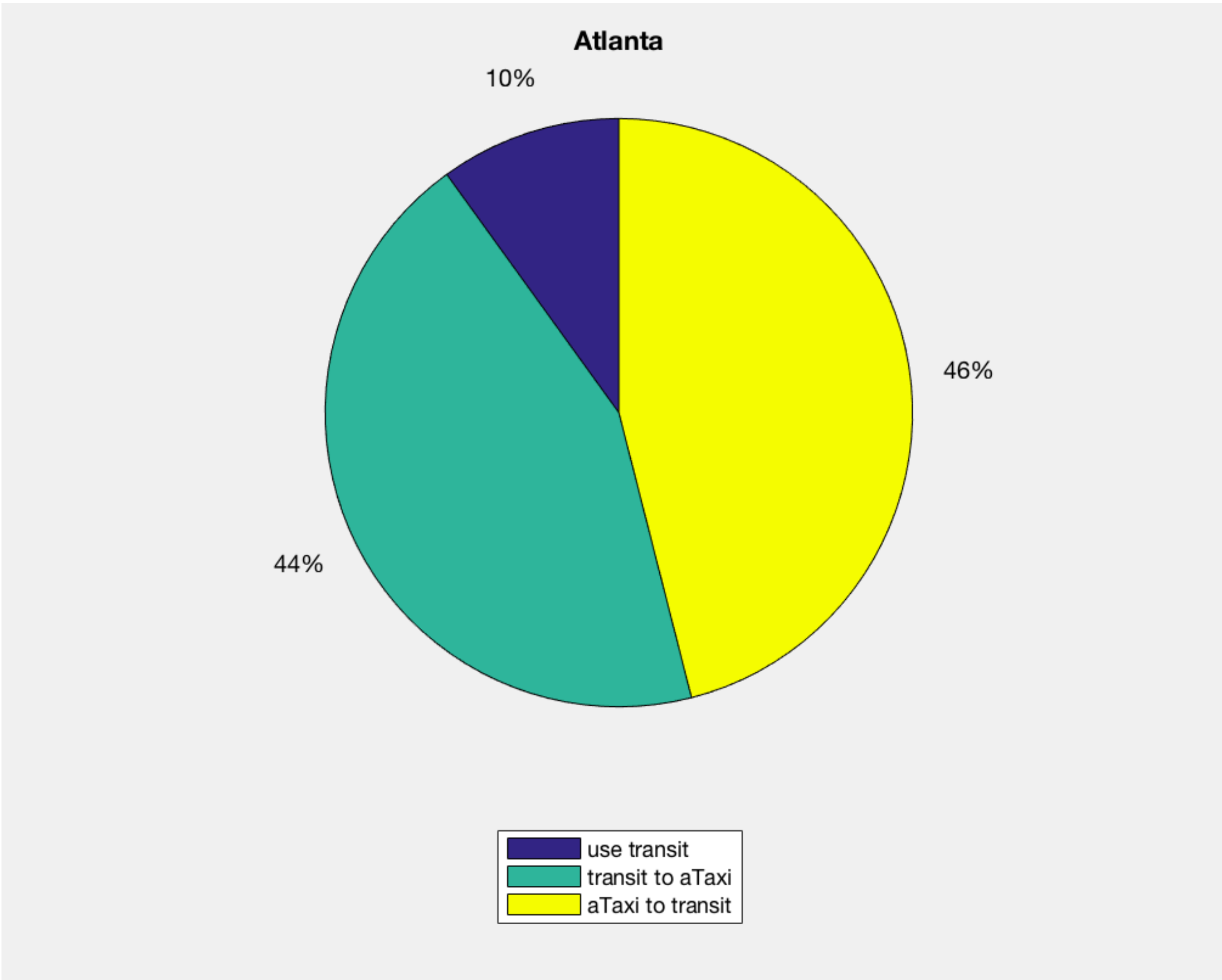
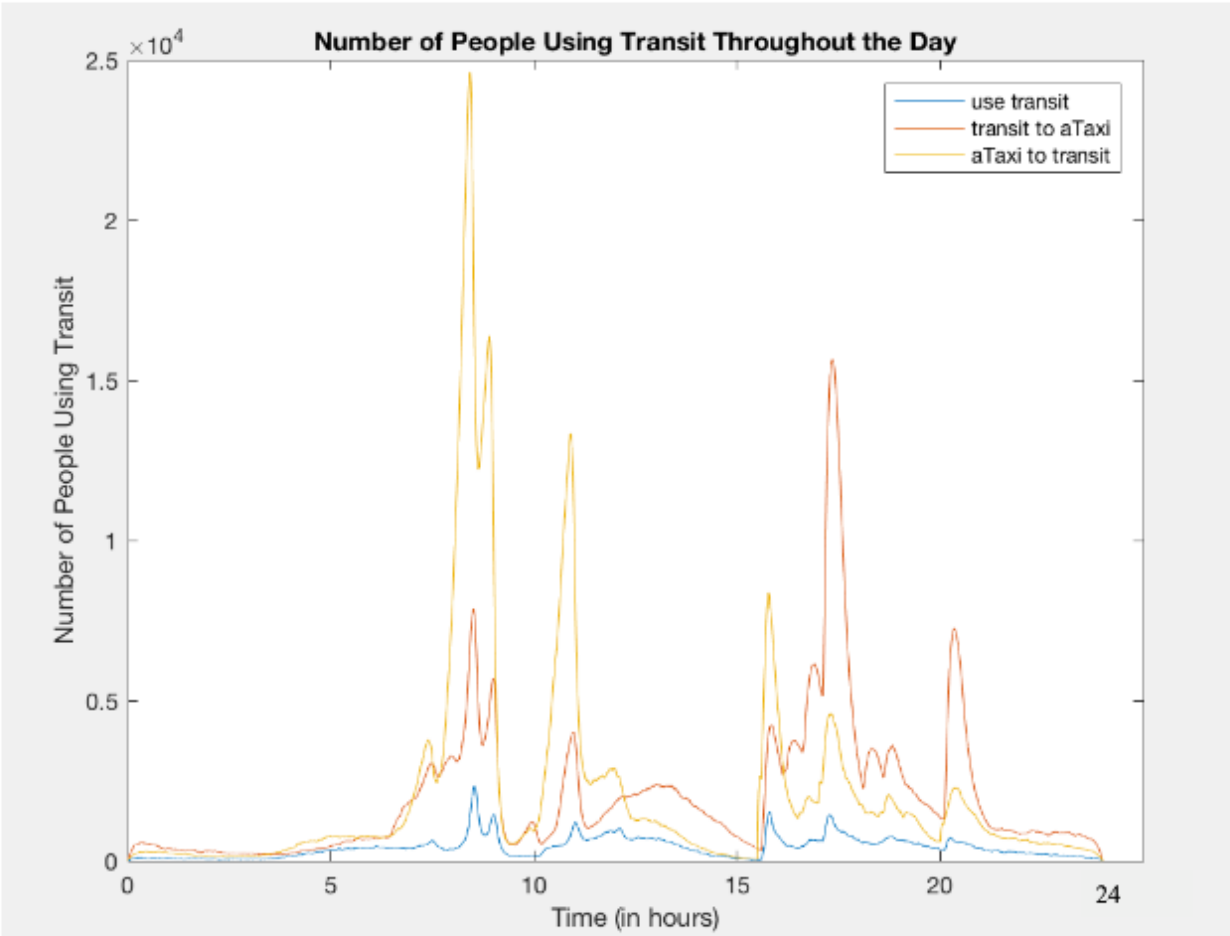


Figure 5.14 - Pie Chart Summarizing The Mode Split in Atlanta



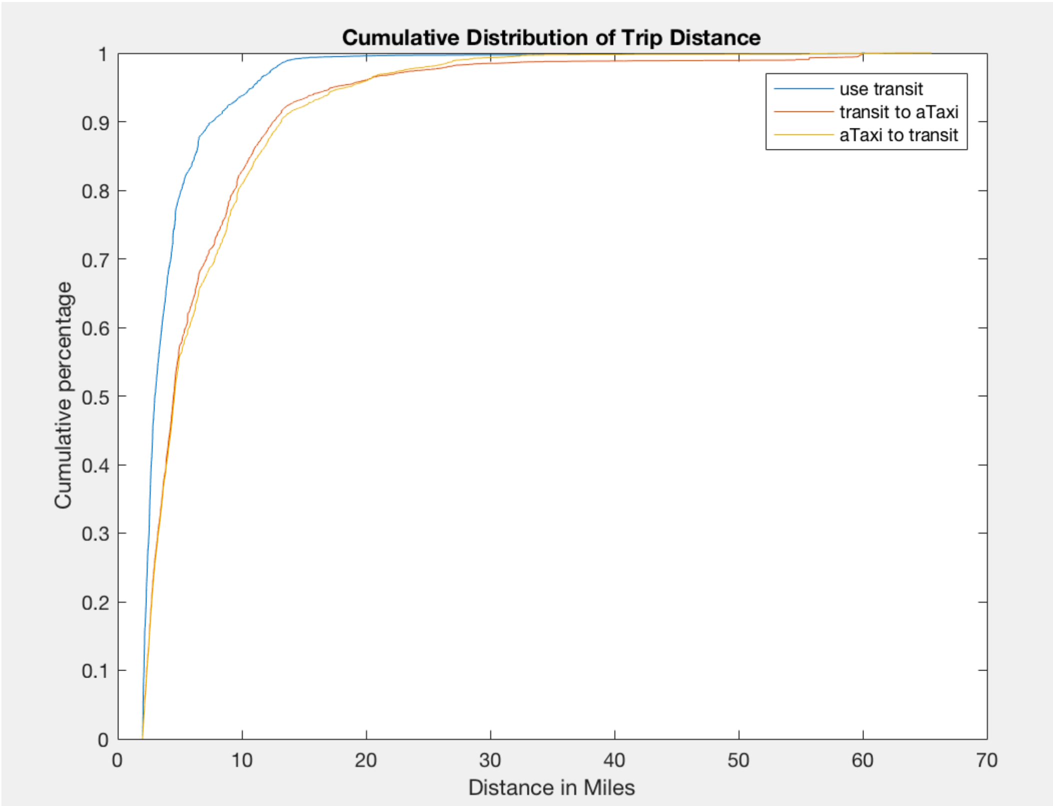
As we can see from Figure 5.14, the impact of aTaxis is highly significant in causing more trips to take transit. The transit to aTaxi and aTaxi to transit segments should be roughly equal as they are here. This is because we considered every county that was within 5 miles of a station in the Atlanta network thus all trips were considered.

Figure 5.15 - Time Distribution of Transit Trips in Atlanta



There is a large morning peak in aTaxi to transit from 8:00am to 9:30am. There is a smaller peak in transit to aTaxi suggesting that more work commuters and school trips are coming from more suburban areas further from transit stations to work/schools in more urban locations within walking distance of a station. There is a corresponding large afternoon peak in transit to aTaxi when works gets out around 5:00pm and a smaller peak in aTaxi to transit when school gets out around 3:30pm. The peak is higher at 9am as this start time for work and school is very common but leaving times tend to be more staggered so the peak is lower but broader in the tales meaning, as the pie chart showed, the area under the line transit to aTaxi and aTaxi to transit lines would be roughly similar. There is also a lunch time peak around 12:00pm when people predominantly travel via the aTaxi to transit mode suggesting they are heading to restaurants in more urban areas for lunch. There is also a peak around 8:30-9:00pm as head back via the transit to aTaxi mode from urban areas after a dinner out or entertainment to their suburban homes.

Figure 5.16 - Cumulative Distribution of Trip Distance in Atlanta



Minimum trip length was 2.0 miles so the cumulative distribution of the trip distance (Figure 5.16) for all three modes starts at 2 miles. Although the aTaxi portion of the trip was a maximum of 5 miles, the transit portion of the trip could be any length hence the tail with a small number of transit to aTaxi and aTaxi to transit trips at greater than 50 miles. However, 90% of the transit only trips were under 5.52 miles, 90% of the transit to aTaxi were under 13.31 miles and 90% of the aTaxi to transit trips were under 14.17 miles. The average distance and travel times for each mode are shown in Figure 5.15.

Figure 5.17 - Average Distance and Travel Times For Each Mode Split in Atlanta

Trip Type	**Metric**	**Value**
Transit	Average Distance	1.49 miles
Transit	Average Travel Time	6.23 min
Transit to aTaxi	Average Distance	3.67 miles
Transit to aTaxi	Average Travel Time	6.08 min
aTaxi to Transit	Average Distance	3.77 miles
aTaxi to Transit	Average Travel Time	5.83 min

From Figure 5.17, the transit trips are on average more than half the distance of the transit to aTaxi or aTaxi to transit trips. The existence of aTaxi's makes these longer trips possible by solving the final mile problem and hence have a longer average distance. However, the travel time of these transit to aTaxi and aTaxi to transit trips is roughly comparable as it omits the slow (2mph) walking segment at one end and replaces it with a fast (30 mph) aTaxi segment.

Transit Utilization

From our analysis of the transit trips across the 20 cities, we have been able to predict the number of trips and person miles travelled after the implementation of aTaxis. Thus, we have modeled the change in transit utilization from current levels (data in the "Now" columns) aggregated from the National Transit Database to the predicted levels with aTaxis (data in the "New" columns) for each major city.

Figure 5.18a - Transit Utilisation Tables: Passenger Miles Travelled

City	PMT Now*	PMT New	Multiplier
Atlanta	765,469,408	1,030,459,787	1.35
Baltimore	794,420,383	1,489,399,653	1.87
Boston	1,847,714,947	5,382,869,748	2.91
Charlotte	197,566,308	618,673,687	3.13
Chicago	3,799,121,983	17,446,181,136	4.59
Cleveland	223,146,222	615,225,860	2.76
Cincinnati	91,524,056	117,634,310	1.29
Detroit	197,566,308	639,241,306	3.24
Grand Rapids	48,755,929	336,969,622	6.91
Greensboro	38,016,971	69,862,146	1.84
Harrisburg	117,336,054	402,013,174	3.43
Jacksonville	80,165,368	47,346,019	0.59
Miami	655,965,657	613,231,874	0.93
Memphis	46,745,902	131,759,095	2.82
Nashville	65,725,205	59,379,483	0.90
Orlando	178,129,638	324,969,503	1.82
Philadelphia	1,546,679,224	6,104,306,038	3.95
Pittsburgh	292,806,101	459,782,009	1.57
Tampa	82,678,376	41,067,633	0.50
Washington D.C.	1,968,724,491	4,450,069,759	2.26
Average	651,912,927	2,019,022,092	2.43
Standard Deviation	961,196,038	4,067,476,874	1.55
Total	13,038,258,531	40,380,441,841	3.10

* The numbers in the "PMT Now" column are summations of PMT data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to PMT data.

Person miles travelled increased in each city by, on average, 2.43 times with a standard deviation of 1.55 leading to a total increase of 3.10 times for the whole region. The biggest increase was in Grand Rapids (6.91x) but large increases were also found in the major cities of Chicago (4.59x) and Philadelphia (3.95x). PMT increased in every city except those in Florida (Jacksonville, Miami, and Tampa) as well as Nashville. These discrepancies were double checked to verify this decrease and are definitely correct based on the data used. This could point to the need to incorporate more vacation trips, particularly in the case of Florida, into the billion trip simulation.

Figure 5.18b - Transit Utilisation Tables: Unlinked Passenger Trips

City	Unlinked Trips Now*	Unlinked Trips New	Multiplier
Atlanta	131,756,876	263,275,595	2.00
Baltimore	113,995,672	298,480,575	2.62
Boston	409,248,438	1,114,070,520	2.72
Charlotte	37,461,838	154,778,980	4.13
Chicago	588,598,934	2,712,435,085	4.61
Cleveland	49,245,884	187,037,680	3.80
Cincinnati	16,624,349	65,147,755	3.92
Detroit	37,461,838	102,039,765	2.72
Grand Rapids	12,524,771	48,489,520	3.87
Greensboro	5,400,899	30,786,290	5.70
Harrisburg	4,032,961	25,080,975	6.22
Jacksonville	12,596,111	23,635,575	1.88
Miami	111,354,011	84,413,915	0.76
Memphis	9,354,609	43,643,050	4.67
Nashville	10,238,898	1,759,300	0.17
Orlando	30,141,247	73,454,425	2.44
Philadelphia	347,177,503	1,438,480,330	4.14
Pittsburgh	63,919,450	148,850,650	2.33
Tampa	15,687,946	18,705,885	1.19
Washington D.C.	411,323,792	1,001,410,715	2.43
Average	120,907,301	391,798,829	3.12
Standard Deviation	172,495,230	683,404,894	1.58
Total	2,418,146,027	7,835,976,585	3.24

* The numbers in the "Unlinked Trips Now" column are summations of unlinked trips data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to unlinked trips data.

The number of unlinked trips also increased in each city by, on average, 3.12 times with a standard deviation of 1.58 leading to a total increase of 3.24 times for the whole region. The biggest increase was in Harrisburg (6.22x) but large increases were also found in the major cities of Chicago (4.61x) and Philadelphia (4.14x). Number of trips increased in every city except Miami and Nashville. These discrepancies were double checked to verify this decrease and are definitely correct based on the data used. This could point to the need to incorporate more vacation trips, particularly in the case of Miami, into the billion trip simulation. These two cities showed a decrease in both metrics and Nashville's was particularly significant for PMT suggesting perhaps there is a more long distance transit line that may have been omitted from the TOD database. More research was conducted into the rail links in Nashville but no additional lines were found.

Financial Analysis

While the previous section of our chapter focused on the ambitious parts of calculating and synthesizing all of this data, these last sections examine the bottom line impacts this type of model yields. Below we examine the change in operating expenses, fare revenue, and profits for each transit system broken down by region.

Our first step was to go to the National Transit Database, a service offered by the Federal Transit Administration, a department of the US Department of Transportation. From there we collected two pieces of data for each city. The first was Operating expenses per vehicle mile (Train miles, not person miles). The second was Fare Revenue, and this was a total measure for the agency. We retrieved these data from the first available heavy rail system in a city, opting for light rail if no heavy rail measures were reported. From those two data points, and our own mileage data, we were able to create comprehensive data tables for each city outlining the financial impact of our model (essentially increasing ridership).

This data has a very poignant effect on a government agency looking for a financial forecast in the next 10 years. More funds and creative pricing structures will be needed, and based on the industry’s ability to deliver those services, it will either benefit financially from the prevalence of aTaxis, or be stunted by it.

This initial research is a start to allow us to at least ask some of these questions. With access to more data from the FTA, we might be able to test the assumption of whether operating expenses (which do only include the variable costs not the sunk capital costs) do vary linearly with such a large increase in ridership and a decrease in average trip length as shown in Figure 5.19a. However, this may be the first time that we have even been able to see estimates of the budgetary effects of aTaxis on transit and will hopefully be a starting point for future research.

Figure 5.19a - Average Trip Length Analysis

	Now*	New	Multiplier
Average Trip Length	5.39	5.15	0.96

* The number in the "Now" column is an average of the trip length data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to trip length data.

Figure 5.19b - Current Average Unit Economic Effects In Per PMT and Per Trip

	Operating Expenses Now*	Fare Revenue Now	Profit Now
Per PMT	1.01	0.17	-0.83
Per Trip	5.42	0.94	-4.48

* The numbers in the "Operating Expenses Now" column are averages of the operating expenses data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to operating expenses data.

Figure 5.19c - Change In Average Unit Economic Effects

	Now*	New	Multiplier
Operating Expenses Per Trip	5.42	5.18	0.96
Fare Revenue Per PMT	0.17	0.18	1.05

* The numbers in the "Now" column are averages of the operating expenses data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to operating expenses and fare revenue data.

Figure 5.19a shows a 4% decrease in average trip length. This seems to conflict with the analysis from Atlanta showing that aTaxis are bringing in longer trips by solving the final mile problem. However, we did not consider any trips over 200 miles so maybe these would increase the average distance up to the current level in the FTA database. However, our the TOD nor the FTA figures included amtrak stations so this is unlikely to be the issue. It could mean that the trip generation need to be adjusted to include more long medium length trips.

Assuming that the operating expenses per PMT and revenue per trip remains the same as in the FTA database. In Figure 5.19b, we have worked back to give the current revenue per PMT and operating expenses per trip for reference.

In Figure 5.19c, the decrease in average trip length means the operating expenses per trip decreases by 4% while the revenue per PMT increases 5%. However, currently transit authorities are losing $0.83 per mile and $4.48 per trip so, despite this improvement in unit economics, the increased total number of trips and PMT means that losses on transit increase. Figures 5.20 and 5.22 show the aggregated financial effects on an individual city basis.

Figure 5.20a - Financial Impact Tables: Operating Expenses

City	Operating Expenses Per PMT Now*	Operating Expenses Now*	Operating Expenses New	Multiplier
Atlanta	0.49	375,080,010	504,925,296	1.35
Baltimore	0.74	587,871,083	1,102,155,743	1.87
Boston	0.54	997,766,071	2,906,749,664	2.91
Charlotte	0.53	104,710,143	327,897,054	3.13
Chicago	0.38	1,443,666,354	6,629,548,832	4.59
Cleveland	0.75	167,359,667	461,419,395	2.76
Cincinnati	0.94	86,032,613	110,576,252	1.29
Detroit	1.05	207,444,623	671,203,371	3.24
Grand Rapids	0.83	40,467,421	279,684,786	6.91
Greensboro	0.92	34,975,613	64,273,175	1.84
Harrisburg	0.45	52,801,224	180,905,928	3.43
Jacksonville	0.88	70,545,524	41,664,497	0.59
Miami	0.57	373,900,424	349,542,168	0.93
Memphis	3.71	173,427,296	488,826,241	2.82
Nashville	1.00	65,725,205	59,379,483	0.90
Orlando	0.59	105,096,486	191,732,007	1.82
Philadelphia	0.43	665,072,066	2,624,851,596	3.95
Pittsburgh	1.57	459,705,579	721,857,754	1.57
Tampa	3.10	256,302,966	127,309,663	0.50
Washington D.C.	0.63	1,240,296,429	2,803,543,948	2.26
Average	1.01	375,412,340	1,032,402,343	2.43
Standard Deviation	0.87	416,138,428	1,608,220,788	1.55
Total	20.10	7,508,246,798	20,648,046,852	2.75

* The numbers in the "Operating Expenses Per PMT Now" and "Operating Expenses Now" columns are the operating expenses data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to operating expenses data.

The operating costs increased in each city by, on average, 2.43 times with a standard deviation of 1.55 leading to a total increase of 2.75 times for the whole region. The biggest increase was in Grand Rapids (6.91x) but large increases were also found in the major cities of Chicago (4.59x) and Philadelphia (3.95x). As operating costs per PMT scale linearly with PMT, the cities where PMT fell, namely Jacksonville, Miami, Tampa and Nashville, also experienced a decrease in operating expenses.

Figure 5.20b - Financial Impact Tables: Fare Revenue

City	Average Fare Revenue Now*	Fare Revenue Now*	Fare Revenue New	Multiplier
Atlanta	0.57	74,914,218	149,693,025	2.00
Baltimore	0.11	13,020,548	34,092,353	2.62
Boston	0.48	197,899,125	538,727,972	2.72
Charlotte	0.12	4,553,044	18,811,557	4.13
Chicago	0.49	290,337,682	1,337,960,485	4.61
Cleveland	0.13	6,380,158	24,232,075	3.80
Cincinnati	1.78	29,521,729	115,690,206	3.92
Detroit	0.50	18,596,179	50,652,873	2.72
Grand Rapids	0.72	9,050,678	35,039,605	3.87
Greensboro	0.62	3,343,986	19,061,442	5.70
Harrisburg	8.19	33,009,837	205,288,099	6.22
Jacksonville	0.88	11,081,896	20,794,274	1.88
Miami	0.23	25,654,430	19,447,803	0.76
Memphis	0.08	706,781	3,297,420	4.67
Nashville	1.06	10,847,629	1,863,895	0.17
Orlando	0.91	27,296,885	66,522,695	2.44
Philadelphia	0.30	103,850,277	430,288,770	4.14
Pittsburgh	0.18	11,385,832	26,514,441	2.33
Tampa	0.03	465,012	554,468	1.19
Washington D.C.	1.44	593,323,968	1,444,509,145	2.43
Average	0.94	73,261,995	227,152,130	3.12
Standard Deviation	1.77	143,076,931	423,662,511	1.58
Total	18.81	1,465,239,894	4,543,042,604	3.10

* The numbers in the "Average Fare Revenue Now" and "Revenue Now" columns are the fare revenue data found by visiting https://www.transit.dot.gov/ntd/transit-agency-profiles and entering the city of interest into the search field. After entering the city name, at least one option appears for each of the transit commissions in the city, each of which leads to a 2014 report of the data collected from the previous year for that commission. Refer to fare revenue data.

The operating costs increased in each city by, on average, 3.12 times with a standard deviation of 1.58 leading to a total increase of 3.10 times for the whole region. The biggest increase was in Harrisburg (6.22x) but large increases were also found in the major cities of Chicago (4.61x) and Philadelphia (4.14x). As fare revenue scales linearly with the number of trips, the cities where the number of trips fell, namely Miami and Nashville, also experienced a decrease in fare revenues.

Figure 5.20c - Financial Impact Tables: Profits

City	Profits Now*	Profits New	Profits Percentage Change
Atlanta	-300,165,792	-355,232,271	-18.35
Baltimore	-574,850,535	-1,068,063,391	-85.80
Boston	-799,866,946	-2,368,021,692	-196.05
Charlotte	-100,157,099	-309,085,497	-208.60
Chicago	-1,153,328,672	-5,291,588,347	-358.81
Cleveland	-160,979,509	-437,187,320	-171.58
Cincinnati	-56,510,884	5,113,955	109.05
Detroit	-188,848,444	-620,550,497	-228.60
Grand Rapids	-31,416,743	-244,645,181	-678.71
Greensboro	-31,631,627	-45,211,733	-42.93
Harrisburg	-19,791,387	24,382,171	223.20
Jacksonville	-59,463,628	-20,870,223	64.90
Miami	-348,245,994	-330,094,365	5.21
Memphis	-172,720,515	-485,528,821	-181.11
Nashville	-54,877,576	-57,515,588	-4.81
Orlando	-77,799,601	-125,209,311	-60.94
Philadelphia	-561,221,789	-2,194,562,826	-291.03
Pittsburgh	-448,319,747	-695,343,313	-55.10
Tampa	-255,837,954	-126,755,195	50.45
Washington D.C.	-646,972,461	-1,359,034,803	-110.06
Average	-302,150,345	-805,250,212	-111.98
Standard Deviation	308,262,305	1,260,072,326	194.21
Total	-6,043,006,905	-16,105,004,248	-166.51

* The numbers in the “Profits Now” column are the profits in each city found by summing the “Operating Expenses Now” column and the “Revenue Now” column.

Profits decreased in each city by, on average, 112% with a standard deviation of 194 leading to a total decrease in profits (or actually an increase in losses) of 167% for the whole region. The largest change in losses was in Grand Rapids (679%) but large decreases in profits were also found in the major cities of Chicago (359%) and Philadelphia (291%). There were 5 cities were profits actually increase. Excluding the three cities where operating costs also fell, namely Jacksonville, Miami and Tampa, the two remaining cities are Cincinnati and Harrisburg. As they both went from a loss to a profit, their percentage increase is basically meaningless but their profits did increase. In both cases, the number of unlinked trips and so fare revenue (Harrisburg both 6.22x and Cincinnati both 3.92x) increases by a significant amount more than PMT and so operating expenses (Harrisburg both 3.43x and Cincinnati both 1.29x). This suggests that an increase in the number of short trips, evidenced by the decreased average trip length in Figure 19a, causes an increase in profits for both of these cities.

Annualisation Factor

We began our analysis by assuming the trip data represented an average day and so multiplied by 365 to get the annualised trips and PMT with which to compare to the annualised data in the National Transit Database.

However, as vacation trips were not necessarily fully incorporated into the database, we decided to offer a refinement to our analysis by only multiplying by roughly the number of working days in a year in the US. As Professor Kornhauser suggested, this is roughly 300 days. This is derived from the fact that adults in the US work, on average, 6 days a week (Isidoer and Luhby, 2015) and have 10 days of paid vacation (Roper, 2016). This equates to roughly 300 working days per year, which we used to annualise the statistics in Figures 20 and 21.

This data still suggests the broad and impactful conclusions as stated above, but perhaps more muted. Similar changes occur in the later financial tables, as we simply change the scale of operation on which the transit agencies would operate.

All numbers in the "Now" columns in this section come from the same sources footnoted in the charts above in the "Transit Utilization" section of the report. Please refer to those footnotes for references to the data sources.

Figure 5.21a - Transit Utilisation Tables (300 day assumption): Passenger Miles Travelled

City	PMT Now	PMT New	Multiplier
Atlanta	765,469,408	846,953,250	1.11
Baltimore	794,420,383	1,224,164,098	1.54
Boston	1,847,714,947	4,424,276,505	2.39
Charlotte	197,566,308	508,498,921	2.57
Chicago	3,799,121,983	14,339,326,961	3.77
Cleveland	223,146,222	505,665,090	2.27
Cincinnati	91,524,056	96,685,735	1.06
Detroit	197,566,308	525,403,813	2.66
Grand Rapids	48,755,929	276,961,333	5.68
Greensboro	38,016,971	57,420,942	1.51
Harrisburg	117,336,054	330,421,786	2.82
Jacksonville	80,165,368	38,914,536	0.49
Miami	655,965,657	504,026,198	0.77
Memphis	46,745,902	108,295,146	2.32
Nashville	65,725,205	48,805,055	0.74
Orlando	178,129,638	267,098,221	1.50
Philadelphia	1,546,679,224	5,017,237,839	3.24

Pittsburgh	292,806,101	377,903,021	1.29
Tampa	82,678,376	33,754,219	0.41
Washington D.C.	1,968,724,491	3,657,591,583	1.86
Average	651,912,927	1,659,470,213	2.00
Standard Deviation	961,196,038	3,343,131,677	1.27
Total	13,038,258,531	33,189,404,253	2.55

Previously the 365 day analysis had a bottom line of roughly 40 billion PMT and a multiplier of 3.10. We see about a 18% reduction in PMT to 33 billion (this is just scaled by 300/365) and a new multiplier of 2.55.

Figure 5.21b - Transit Utilisation Tables (300 day assumption): Unlinked Passenger Trips

City	Unlinked Trips Now	Unlinked Trips New	Multiplier
Atlanta	131,756,876	216,390,900	1.64
Baltimore	113,995,672	245,326,500	2.15
Boston	409,248,438	915,674,400	2.24
Charlotte	37,461,838	127,215,600	3.40
Chicago	588,598,934	2,229,398,700	3.79
Cleveland	49,245,884	153,729,600	3.12
Cincinnati	16,624,349	53,546,100	3.22
Detroit	37,461,838	83,868,300	2.24
Grand Rapids	12,524,771	39,854,400	3.18
Greensboro	5,400,899	25,303,800	4.69
Harrisburg	4,032,961	20,614,500	5.11
Jacksonville	12,596,111	19,426,500	1.54
Miami	111,354,011	69,381,300	0.62
Memphis	9,354,609	35,871,000	3.83
Nashville	10,238,898	1,446,000	0.14
Orlando	30,141,247	60,373,500	2.00
Philadelphia	347,177,503	1,182,312,600	3.41
Pittsburgh	63,919,450	122,343,000	1.91
Tampa	15,687,946	15,374,700	0.98
Washington D.C.	411,323,792	823,077,300	2.00
Average	120,907,301	322,026,435	2.56
Standard Deviation	172,495,230	561,702,652	1.30
Total	2,418,146,027	6,440,528,700	2.66

Previously the 365 day analysis had a bottom line of roughly 5.8 billion trips and a multiplier of 3.24. We see the same 18% reduction to 6.4 billion and a new multiplier of 2.66.

Figure 5.22a - Financial Impact Tables: Operating Expenses

City	Operating Expenses Per PMT Now	Operating Expenses Now	Operating Expenses New	Multiplier
Atlanta	0.49	375,080,010	415,007,092	1.11
Baltimore	0.74	587,871,083	905,881,433	1.54
Boston	0.54	997,766,071	2,389,109,313	2.39
Charlotte	0.53	104,710,143	269,504,428	2.57
Chicago	0.38	1,443,666,354	5,448,944,245	3.77
Cleveland	0.75	167,359,667	379,248,818	2.27
Cincinnati	0.94	86,032,613	90,884,591	1.06
Detroit	1.05	207,444,623	551,674,003	2.66
Grand Rapids	0.83	40,467,421	229,877,906	5.68
Greensboro	0.92	34,975,613	52,827,267	1.51
Harrisburg	0.45	52,801,224	148,689,804	2.82
Jacksonville	0.88	70,545,524	34,244,792	0.49
Miami	0.57	373,900,424	287,294,933	0.77
Memphis	3.71	173,427,296	401,774,993	2.32
Nashville	1.00	65,725,205	48,805,055	0.74
Orlando	0.59	105,096,486	157,587,951	1.50
Philadelphia	0.43	665,072,066	2,157,412,271	3.24
Pittsburgh	1.57	459,705,579	593,307,743	1.29
Tampa	3.10	256,302,966	104,638,079	0.41
Washington D.C.	0.63	1,240,296,429	2,304,282,697	1.86
Average	1.01	375,412,340	848,549,871	2.00
Standard Deviation	0.87	416,138,428.55	1,321,825,305.46	1.27
Total	20.10	7,508,246,798.72	16,970,997,412.94	2.26

Previously the 365 day analysis had a bottom line of roughly $20.6 billion in operating expenses and a multiplier of 2.75. We see the same 18% reduction to $17 billion and a new multiplier of 2.26.

Figure 5.22b - Financial Impact Tables: Fare Revenue

City	Average Fare Revenue Now	Fare Revenue Now	Fare Revenue New	Multiplier
Atlanta	0.57	74,914,218	123,035,363	1.64
Baltimore	0.11	13,020,548	28,021,112	2.15
Boston	0.48	197,899,125	442,790,114	2.24
Charlotte	0.12	4,553,044	15,461,554	3.40
Chicago	0.49	290,337,682	1,099,693,549	3.79
Cleveland	0.13	6,380,158	19,916,774	3.12
Cincinnati	1.78	29,521,729	95,087,841	3.22
Detroit	0.50	18,596,179	41,632,499	2.24
Grand Rapids	0.72	9,050,678	28,799,676	3.18
Greensboro	0.62	3,343,986	15,666,939	4.69
Harrisburg	8.19	33,009,837	168,729,944	5.11
Jacksonville	0.88	11,081,896	17,091,184	1.54
Miami	0.23	25,654,430	15,984,496	0.62
Memphis	0.08	706,781	2,710,209	3.83
Nashville	1.06	10,847,629	1,531,969	0.14
Orlando	0.91	27,296,885	54,676,188	2.00
Philadelphia	0.30	103,850,277	353,662,003	3.41
Pittsburgh	0.18	11,385,832	21,792,691	1.91
Tampa	0.03	465,012	455,727	0.98
Washington D.C.	1.44	593,323,968	1,187,267,790	2.00
Average	0.94	73,261,995	186,700,381	2.56
Standard Deviation	1.77	143,076,931	348,215,762	1.30
Total	18.81	1,465,239,894	3,734,007,620	2.55

Previously the 365 day analysis had a bottom line of roughly $4.5 billion in fare revenues and a multiplier of 3.10. We see the same 18% reduction to $3.7 billion and a new multiplier of 2.55.

Figure 5.22c - Financial Impact Tables: Profits

City	Profits Now	Profits New	Profits Percentage Change
Atlanta	-300,165,792	-291,971,729	2.73
Baltimore	-574,850,535	-877,860,321	-52.71
Boston	-799,866,946	-1,946,319,199	-143.33
Charlotte	-100,157,099	-254,042,874	-153.64
Chicago	-1,153,328,672	-4,349,250,696	-275.10
Cleveland	-160,979,509	-359,332,044	-123.22
Cincinnati	-56,510,884	4,203,250	105.44
Detroit	-188,848,444	-510,041,505	-170.08
Grand Rapids	-31,416,743	-201,078,231	-540.04
Greensboro	-31,631,627	-37,160,328	-15.48
Harrisburg	-19,791,387	20,040,140	201.26
Jacksonville	-59,463,628	-17,153,608	71.15
Miami	-348,245,994	-271,310,437	22.09
Memphis	-172,720,515	-399,064,784	-131.05
Nashville	-54,877,576	-47,273,086	13.86
Orlando	-77,799,601	-102,911,763	-32.28
Philadelphia	-561,221,789	-1,803,750,268	-221.40
Pittsburgh	-448,319,747	-571,515,052	-25.48
Tampa	-255,837,954	-104,182,352	59.28
Washington D.C.	-646,972,461	-1,117,014,907	-72.65
Average	-302,150,345	-661,849,490	-74.23
Standard Deviation	308,262,305	1,035,675,884	159.63
Total	-6,043,006,905	-13,236,989,793	-119.05

Previously the 365 day analysis had a bottom line of roughly $16 billion in losses and a multiplier of -1.65. We see the same 18% reduction to $17 billion and a new multiplier of 2.26.

Our results are displayed and discussed in the previous figures but the fully linked spreadsheet of transit utilisation and financial analysis, which can be easily updated with new assumptions for future research, can be found here:
https://docs.google.com/spreadsheets/d/1KAy7Lxv6o2wz3ids0kccw8EjxHn58cqPWObKAlPlIG0/edit?usp=sharing

Ride-Sharing Analysis

The aTaxi end of the transit to aTaxi and aTaxi to transit mode splits are an ideal opportunity for ridesharing. This is because, as train unload numerous passengers at the same time, you can load an aTaxi that satisfies common destination and circuity constraints with a minimal departure delay.

We ran some proof of concept tests on one of the most populated pixels in Atlanta (-710, -940). The ride-sharing analysis was done on the aTaxis serving the transit to aTaxi mode. The output was as follows: https://drive.google.com/file/d/0B-zTKW2_9LnmeGFOdTZsbXA5WFk/view?usp=sharing

Figure 5.23 - Cumulative Departure of aTaxis Serving Transit to aTaxi Trips Throughout the Day

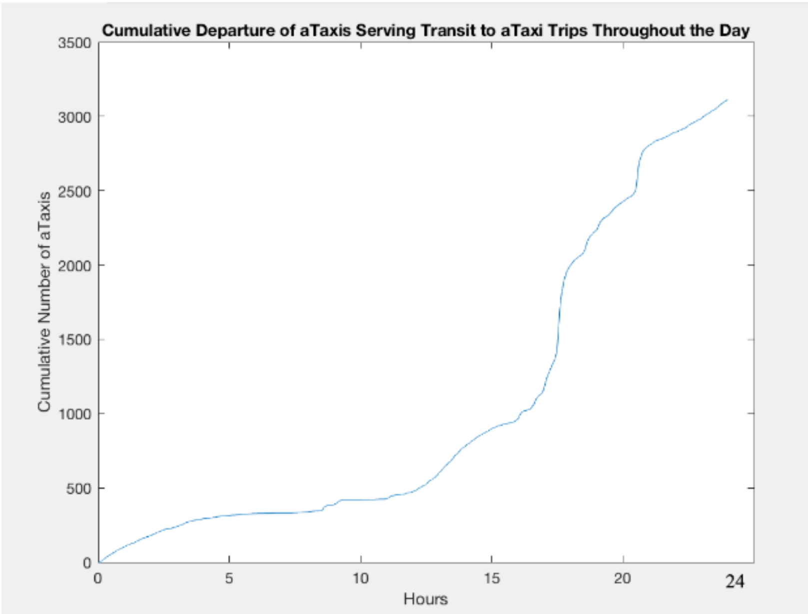
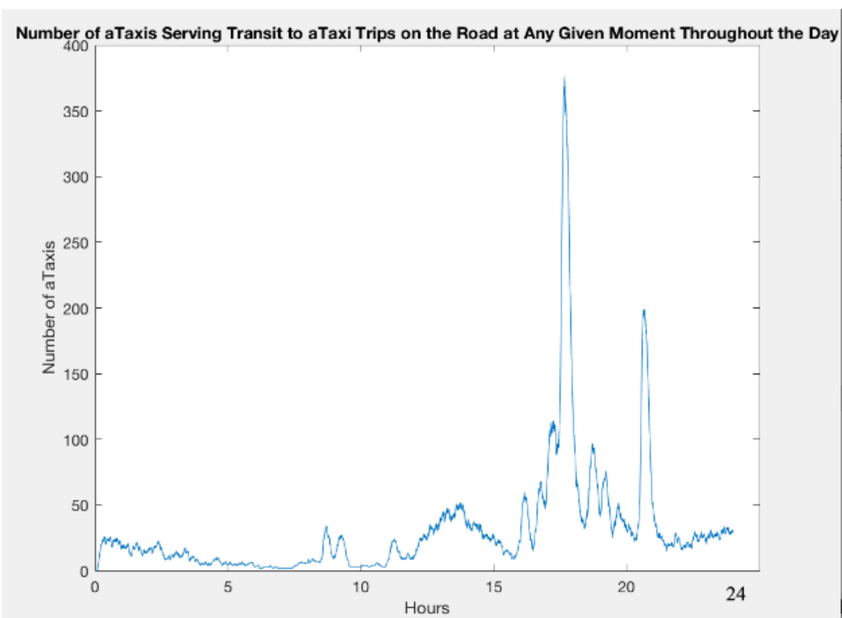


Figure 5.24 - Number of aTaxis Serving Transit to aTaxi Trips on the Road at Any Given Moment Throughout the Day



There is a major rise in Figures 5.23 and 5.24 around 5:30pm after work and then again at 8:30pm after dinner out as was shown in the analysis of transit to aTaxi trips across Atlanta in Figure 5.15. From Figure 5.23, the minimum fleet size assuming there is no empty vehicle repositioning is 3,119. From Figure 5.24, we can see that the maximum number of aTaxis required is 379 at 5:30pm. This corresponds to the minimum fleet size if empty vehicle repositioning was fully optimized.

Figure 5.25 - Number of aTaxis Serving Transit to aTaxi Trips Vs Number of Occupants per Departing Vehicle

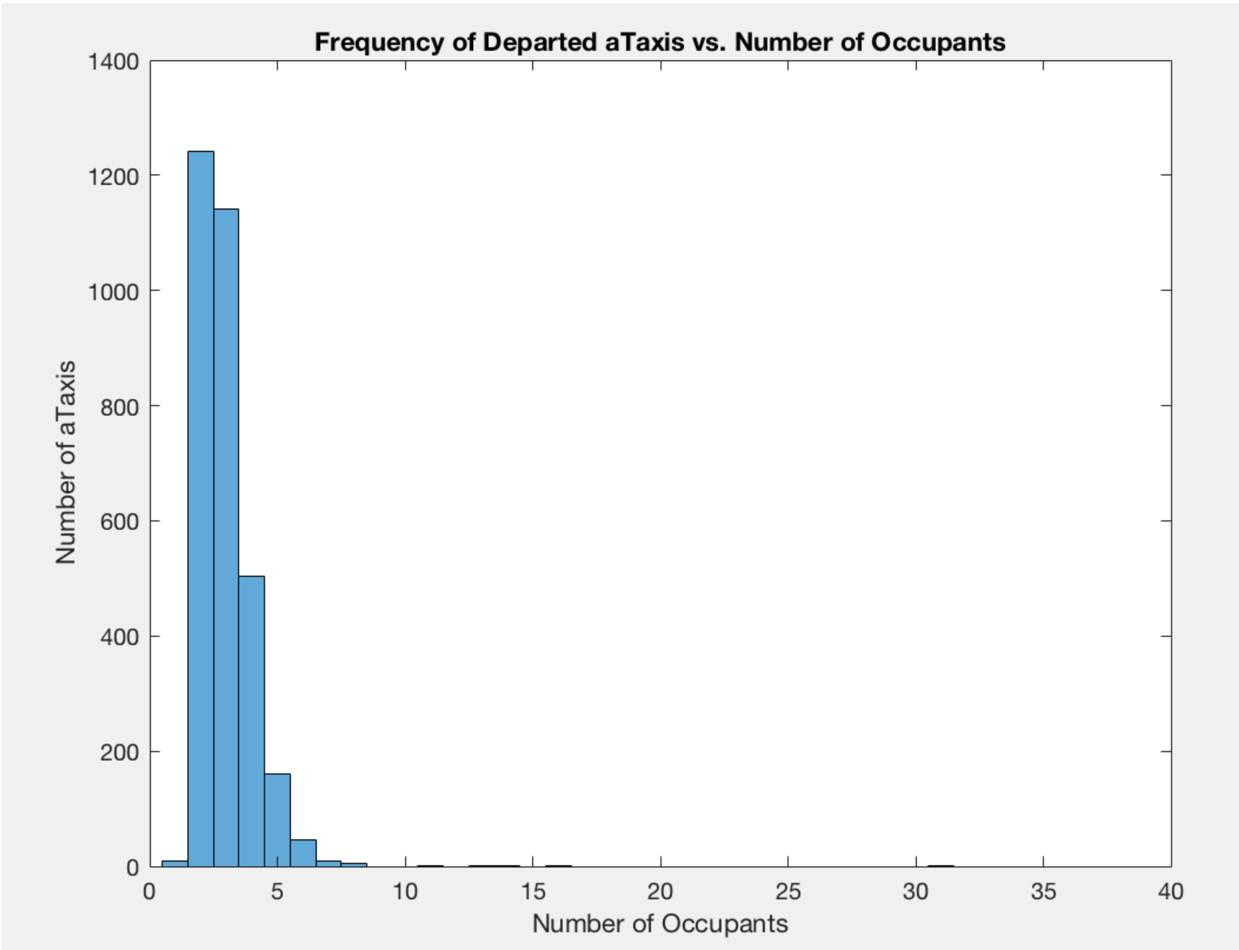


Figure 5.25 shows the number of aTaxis against the number of occupants when the vehicle departs. It seems odd that the number of aTaxi departing with 2 occupants is so much greater than the number departing with 1 occupant. However, we double checked the values to verify. One potential explanation is that, as many transit passengers are arriving at exactly the same location, it is rare for a single passenger to not have at least one other passenger present that meets the circuity and common destination constraints. What is clear is that, with an AVO of 2.95, there is significant ride-sharing potential.

The ride-sharing analysis took too long to run on a whole city but, in future research, this ride-sharing code (or even other more data rich version we have developed, see Code Appendix) could be run using our della cluster setup for all of the aTaxis serving transit to aTaxi and aTaxi to transit modes in each of our 20 cities.

Conclusions

This project focused on creating mode split, purpose split, time of day, trip length distributions and ride-sharing analysis for transit trips in the East. We showed that Pure aTaxi trips are the most used mode of travel. We found that the most frequent travel purposes are "Home to other" and "Other to home". We investigated the time of day and distance distributions for transit, transit to aTaxi and aTaxi to transit modes with impacts on fleet size and repositioning strategies for aTaxis and trains alike. We conducted an initial investigation, which provides a clear starting point for future research, into the ride-sharing potential for the transit to aTaxi trips in a pixel in Atlanta isolating its minimum fleet size and determining, with an AVO of 2.95, there is significant ride-sharing potential.

Our final deductions showed that aTaxis would increase transit utilisation by 3.10 (2.55 for 300 days) times in terms of passenger miles travelled and 3.24 (2.66 for 300 days) times in terms of unlinked trips across our whole Eastern region. The financial effects of this are an increase in operating expenses by 2.75 (2.26 for 300 days) times and in fare revenues by 3.10 (2.55 for 300 days) times leading to a 167% (120% for 300 days) decrease in profits from transit systems across the East.

References

Center for Transit-Oriented Development,. "Transit Oriented Development Database". *TOD Database*. N.p., 2015. Web. 20 Dec. 2015.

Haile, Elizabeth and Evan Wood. *Mode Split Java Script*. Princeton University: N.p., 2015. Print.

Isidore, Chris and Tami Luhby. "Turns Out Americans Work Really Hard... But Some Want To Work Harder". *CNN*. N.p., 2015. Web. 16 Jan. 2015.

Kornhauser, Alain and Kyle Marocchini. "Nationwidetrips'16". *ORF 467 - Princeton University*. N.p., 2015. Web. 16 Dec. 2015.

Kornhauser, Alain. "ORF 467: Transportation System Analysis". 2016. Lecture.

Princeton University,. "Della Research Computing". *Princeton University*. N.p., 2016. Web. 10 Jan. 2015. Note: special thanks to Michael Bino to his help with configuring this and his bash script to download the csvs from the Nation Wide Trips database more efficiently.

Roper, Colin. "Paid Vacation Time: How Do You Stack Up?". *Gusto*. N.p., 2016. Web. 16 Jan. 2015.

United States Department of Transportation,. "National Transit Database". *Federal Transit Adminitration*. N.p., 2014. Web. 13 Jan. 2015.

References were created using: *Cite This For Me*. N.p., 2016. Web. 16 Jan. 2016.

Code Appendix

Mode Split Codebase:
As discussed earlier, the following script is our original data rich mode split program:
https://drive.google.com/open?id=0B-zTKW2_9LnmSU9uVUFJSG1iUkE
As explained, the code used to get output for the 20 cities was produced by Evan Wood and Elizabeth Haile and is linked here:
.java file: https://drive.google.com/file/d/0B-zTKW2_9LnmZ0lkV3BQbWxteTA/view?usp=sharing
Compiled .class file: https://drive.google.com/file/d/0B-zTKW2_9LnmN2pqQ2t2Um9iZEE/view?usp=sharing
This was used in conjunction with my KD-Tree implementation code which cannot be linked due to COS 226 Honor Code restrictions.

Visualisation Codebase:
This is the visualisation script that was written by our team and given to all of the transit teams to use to visualise the Java code output: https://drive.google.com/file/d/0B-zTKW2_9LnmMklONktTLWlhaGc/view?usp=sharing
This is the visualisation script that we wrote for the data rich output of our original Matlab code:
https://drive.google.com/file/d/0B-zTKW2_9LnmQ24xQVVDQVJnSU0/view?usp=sharing

Ride-Sharing Codebase:
For a more data rich analysis, as in our Empty Vehicle Repositioning report (https://docs.google.com/document/d/15GdT4zhRSJ2GeGS1jwJMXSfp-e4WaWJ1rQdhmlZFpSY/edit?usp=sharing), we wrote the following scripts that were created and implemented in Matlab. They are linked here for use in case they are of use in future research:
https://docs.google.com/document/d/14D5mf41H3_UiSApzOtcu3LvUfTBomxNv8V8ZhRRlcA8/edit?usp=sharing

1. updated_pixelate : created our own pixelation scheme for a coordinate system starting at zero to make computations easier; essentially a pixelation procedure that scaled the latitude-longitude data in a slightly different way than the data provided by Professor Kornhauser on Blackboard
2. final_county_origin : generated all the aTaxi trips within our contiguous area under the assumption of unlimited capacity within the aTaxis; outputted an aggregate total trip array with all the aTaxi trips from every pixel within our contiguous area
3. final_county_max3 : generated all the aTaxi trips within our contiguous area under the assumption that each aTaxi could hold a maximum of 3 passengers
4. walk_trips : generated the walking trips within our contiguous area
5. short_haul_trips : created the array for all the short haul trips within the contiguous area
6. short_haul_ataxi : generated the short-haul aTaxi trips within the contiguous area; a simple modification was added to change the departure delay between 180 and 300 seconds and to restrict the number of passengers that could ride in each aTaxi
7. normal_ataxi_greater_than_2 : generated the normal aTaxi trips within the contiguous area; a simple modification was added to restrict the number of passengers that could ride in each aTaxi to further break down the analysis

Chapter 6:
Potential aTaxi and Amtrak Ridership Opportunities

For this section of the analysis on nation-wide ride sharing, we sought to analyze all personTrips that were greater than 200 miles but less than 500 miles long.

Franco D'Agostino and William Chance

Our main method of studying and aggregating these trips was via the use of a mode-split model; more specifically, we broke up all trips into the following segments:

- **Segment 1:** Origin to closest Amtrak station, which could be theoretically serviced by an aTaxi service.
- **Segment 2:** Amtrak train ride to Amtrak station closest to destination.
- **Segment 3:** Amtrak station closest to destination to original intended destination, which could also be theoretically by an aTaxi service.

Figure 6a: Simple Illustration of Broken Up Trips

Broken-up Trips Algorithm Description

In order to actually break up all the trips included in the nation-wide trip file database that are greater than 200 miles but less than 500 miles, we implemented the following algorithm:

1. Locate nearest Amtrak train station to the trip origin's latitude and longitude.
   a. Record distance from Origin to nearest Amtrak station (let this distance be **A**).
2. Locate nearest Amtrak train station to the trip destination's latitude and longitude.
   a. Record distance from Destination to nearest Amtrak station (**B**)**.**
3. Using a master Amtrak database, compute the time and distance that it takes to ride from any Amtrak station to any other Amtrak station in the United States.
   a. Record distance from Origin Amtrak station to Destination Amtrak station (**C**).
4. Sum distances **A** + **B** + **C** to obtain total trip distance
5. Compare total broken-up trip distance to Great Circle distance of original trip.
   a. If Total Trip Distance <= 1.3 * Great Circle Distance, then break up trip and write data to output file.

Remark: The 1.3 multiplying the Great Circle Distance is meant to illustrate the 30% circuity constraint to fit our analysis; this number can vary depending on the specifications of the user.

Sample Data for Broken Up Trips

Given the sheer size of the data that we are analyzing, it is impossible and impractical to include the entire output file in this report. We therefore include below roughly 20 broken-up trips in Figure 6a using the algorithm described above:

Origin Lat	Origin Lon	oStation	Dest Lat	Dest Lon	dStation	Train Trip Distance	Train Trip Distance with Train Trip	Car Only Trip Distance
32.63	-85.91	725	39.57	-86.10	5	460.22	528.40	481.19
32.63	-85.91	725	39.62	-86.11	5	460.22	524.75	484.69
32.63	-85.91	725	39.41	-86.04	5	460.22	540.10	470.44
32.63	-85.91	725	39.61	-86.17	5	460.22	524.88	484.00
32.63	-85.91	725	39.36	-85.95	5	460.22	544.75	467.50
32.63	-85.91	725	39.62	-86.14	5	460.22	524.49	484.59
32.63	-85.91	725	39.50	-86.06	5	460.22	533.38	476.79
32.63	-85.91	725	39.63	-86.18	5	460.22	523.93	484.95
32.63	-85.91	725	39.54	-86.20	5	460.22	529.92	478.99
32.63	-85.91	725	39.35	-85.96	5	460.22	545.05	467.10
32.63	-85.91	725	39.60	-86.11	5	460.22	526.38	483.06
32.63	-85.91	725	39.50	-86.13	5	460.22	532.73	476.42
32.63	-85.91	725	39.58	-86.20	5	460.22	527.12	481.81
32.63	-85.91	725	39.58	-86.19	5	460.22	527.21	481.68
32.63	-85.91	725	39.59	-86.10	5	460.22	526.99	482.57
32.63	-85.91	725	39.57	-86.10	5	460.22	528.40	481.19
32.63	-85.91	725	39.63	-86.18	5	460.22	523.84	485.04
32.58	-86.89	725	43.32	-83.74	133	710.33	789.91	761.81
32.58	-86.89	725	43.45	-84.02	133	710.33	802.85	766.38
32.58	-86.89	725	43.48	-83.99	133	710.33	804.03	769.15
32.58	-86.89	725	43.25	-83.78	133	710.33	785.31	756.16

Figure 6b: Sample Data for Broken Up Trips

All the column headers in Figure 6b should be self-explanatory, except for "oStation" and "dStation"; these are merely indices that map to a specific Amtrak station in the United States.

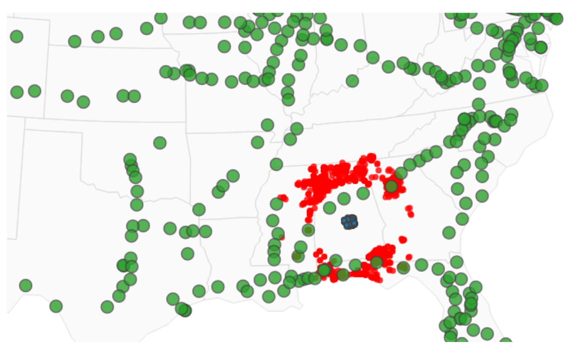


Figure 6c: Long Trips Plot

Figure 6c also illustrates the inner workings of our algorithm; the blue dots in the middle represent the origins of trips for a single county in Alabama, while the red dots indicate their destinations. The green dots depict all the Amtrak stations in the United States.

Statistics for Nationwide Trips

Remark: Due to the large quantity of files that our code was processing, we were only able to process roughly 70% of the data before our computer crashed due to a segmentation fault. Therefore, all statistics and analyses provided from this point forward in this chapter only encompass the 70% of the data that our code managed to cover. For future reference, our code is annexed.

Total Miles Statistics

Origin -> Origin Station	359,902,939.27
Destination Station -> Destination	349,124,658.23
Train Trips	7,542,070,061.06

Average Miles Statistics

Origin -> Origin Station	16.89
Destination Station -> Destination	16.37
Train Trips	353.85

Figure 6d: Statistics for Nationwide Trips

Even though our code only covered 70% of the data, it is still remarkable to observe the sheer distance covered by all the broken up trips in Figure 6d; not to mention that it is also a healthy sanity check to see that the distances from the origin/destination to their closest Amtrak stations are roughly equidistant.

Cumulative Distribution Functions

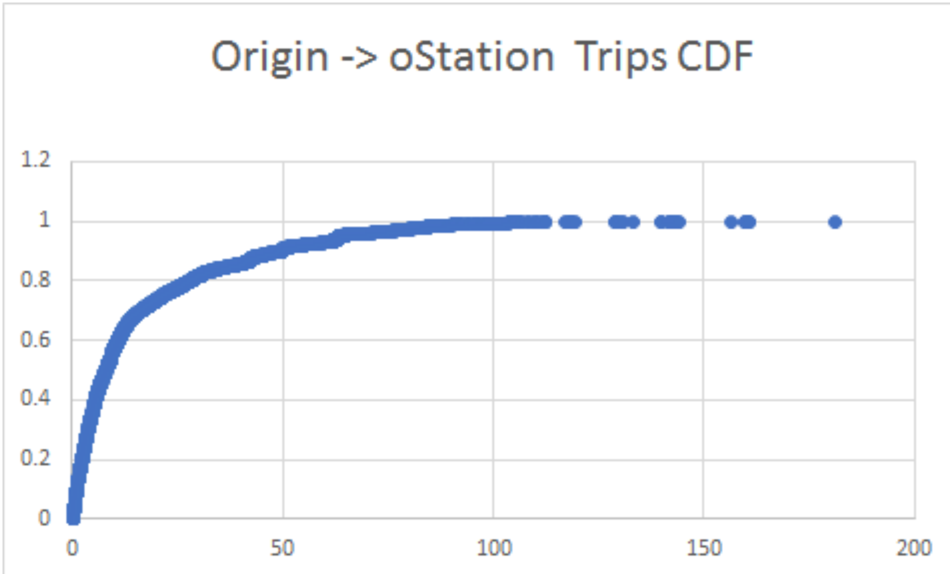


Figure 6e: CDF of trips from Origin to closest Amtrak station to origin (oStation)

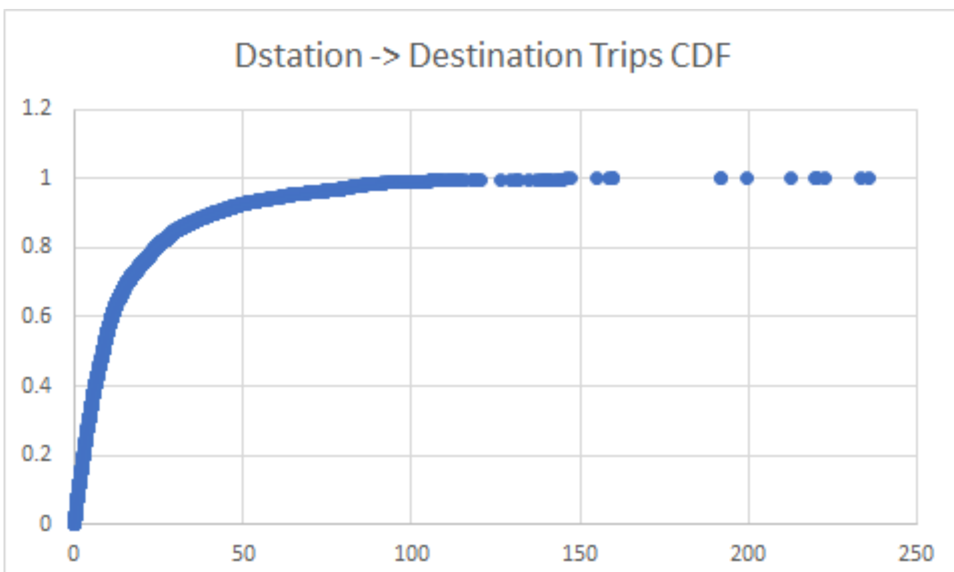


Figure 6f: CDF of trips from closest Amtrak station to destination (dStation) to intended destination

It is interesting to note that the total distance travelled from each Amtrak station roughly equaled each other, as evidenced in Figures 6e and 6f; that is, the distance from the origin to its closest Amtrak station and the distance from the destination to its closest Amtrak station seemed roughly equidistant, based on the two cumulative distribution functions above. Moreover, it is fascinating to note that roughly 80% of trips between stations fall below the 30 mile mark, which is a relatively short distance to cover for trips that are intended to be greater than 200 miles.

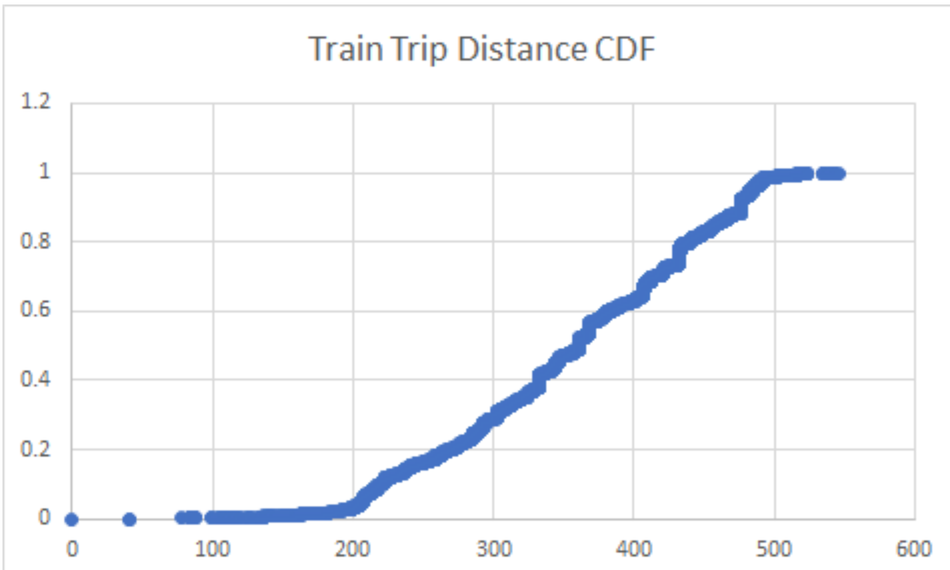


Figure 6g: CDF of total train trip distance

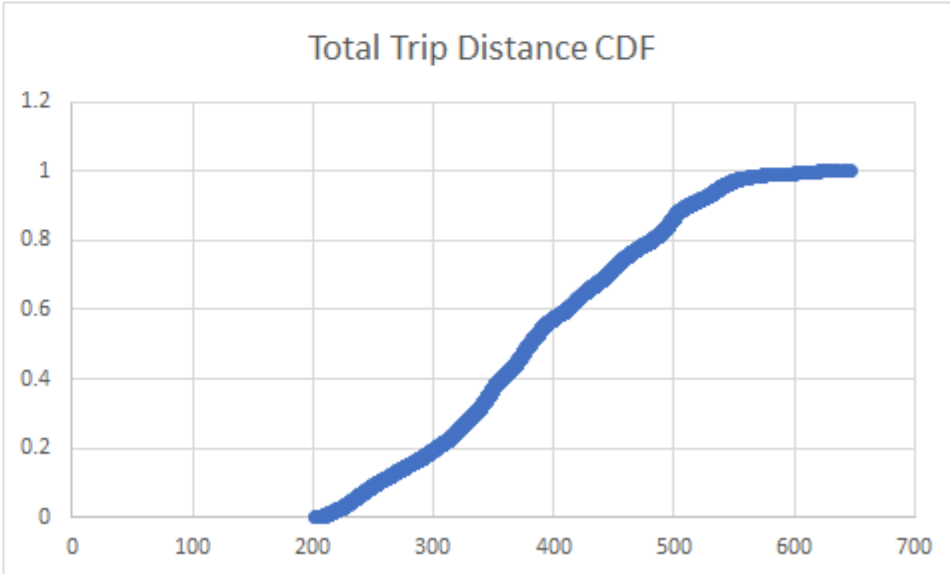


Figure 6h: CDF of total trip distance of broken up trips

Even though we are only studying trips that are no greater than 500 miles, it may be puzzling as to how come a small percentage trips are greater than 500 miles in Figure 6h. This is no computing error; since there might not be a high density of Amtrak stations in some areas of the country, some trips became longer than the original "straight line" trip, since the passengers have to theoretically travel large distances to reach an Amtrak station.

Chapter 7:
Long Trips

The goal of this project was to analyze the plane trips from all over the United States. We performed analysis on all long trips greater than 500 miles and then compared the results with historical airplane trip data.

Hassan Chaudhry, Daniel Liu, Adam Hare

Section 5.1: The Data

In order to procure the data for such trips, we went about the following process:

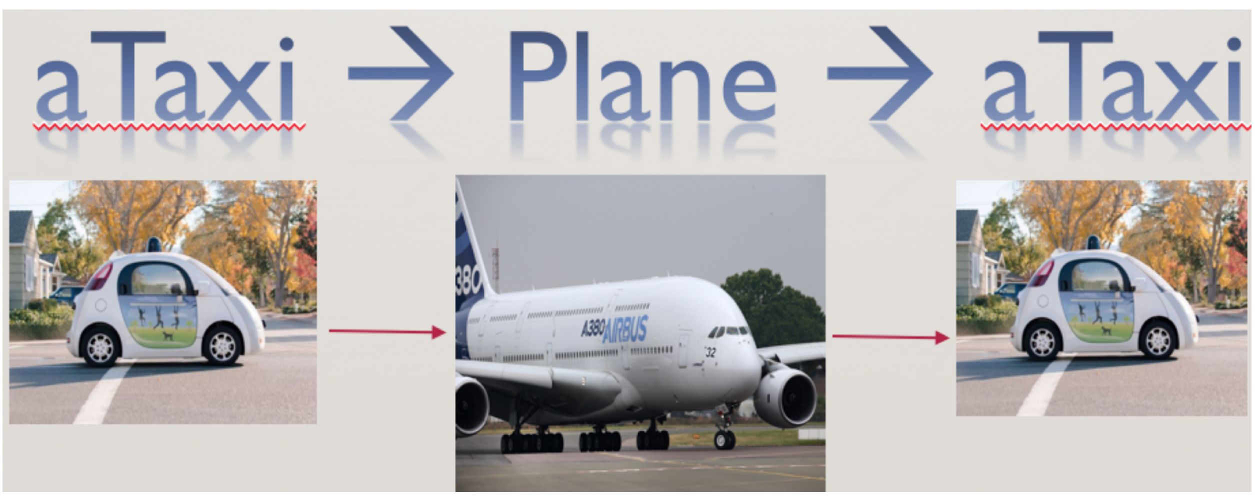
- First we downloaded the 'Long Trip' data (~6200) excel files from the ORF directory.
- This data contained trips which were greater than 200 miles from all over the United States and which amounted to about 40 gigabytes of data.
- We filtered the data to trips greater than 500 miles and eliminated unnecessary columns in order to cut the file size so we could process the complete data more efficiently.

Section 5.2: Assumptions

Before we describe the algorithm that we employed there were three main assumptions that we made.

- First, we assumed that the passengers were all taking commercial flights and did not have their own chartered planes.
- Second, we assumed that the passengers would always take planes for trips greater than 500 miles over other modes of transport namely trains and cars.
- Lastly, for all the trips that did not originate at the airports, we assumed that the passenger utilized an aTaxi from the origin to get to the airport and similarly from the destination airport to the destination.

The general route for a long trip can therefore be presented as follows:



Section 5.3: Steps Taken

To find the nearest origin and destination airports, we took the following steps:

- First we downloaded airport data from OpenFlights Airports Database (http://openflights.org/data.html)
- Next, we filtered the data to include 377 of the busiest airports in the United States. We were able to retrieve the busiest airport data from the FAA (http://tinyurl.com/zyatba8). Our file that contains the busiest airports is called mergedairports.csv and can be accessed at

https://www.dropbox.com/sh/uwrovzswl7x73d9/AAB9MpgYGmeEaJP22DkRgAz1a?dl=0&preview=mergedairports.csv

- For each long trip, we found the nearest origin airport and nearest destination airport using a Kd-tree nearest neighbor algorithm in Python
- Next, we appended the nearest airport data for both origin and destination points
- We were then able to calculate the plane trip miles by taking the great circle distance between the two airport coordinates
- Finally we did a mode split analysis by dividing the trips into plane and aTaxi trips
- Lastly, we analyzed all the trips according to their temporal and spatial properties as well as the trip distances

Section 5.4: Sampling the Dataset

We ran into issues calculating the nearest airport and appending the airport information for the entirety of the data because it was simply taking too much time to run. In order to further our progress, we decided to perform analysis on a random sample of the data. We sampled 0.5% of the data from each FIPS code and combined them into a new file which was representative of the data from all over the United States.

In our result.csv file we have appended the sample data for each FIPS code and included all nearest origin and destination airports. This file can be accessed at:

https://www.dropbox.com/sh/uwrovzswl7x73d9/AAB9MpgYGmeEaJP22DkRgAz1a?dl=0&preview=result.csv

Note that the file is too big to be previewed on Dropbox and needs to be downloaded and opened in Matlab, Python, or R. One row of the results.csv file looks like:

County Code	OType	OLon	OLat	ODeparture	DType	DLon	DLat	GCDistance	origin_airpo	origin_airpo	origin_airpo	origin_airpo	origin_dist	dest_airport	dest_airport	dest_airport	dest_airport	dest_dist
26145	O	-86.412096	32.488821	68599	H	-84.054875	43.407668	766.033248	4115	Montgomery	32.300639	-86.393972	13.053493	4128	Mbs Intl	43.532913	-84.079647	8.815295

After we had performed this analysis, we then were able to extract the aTaxi trips from this file as well. As highlighted before, every plane trip was assumed to include two car trips to and from the airport. Therefore, we were able to generate two aTaxi trips for every plane trip:

1) Trip from origin coordinates to nearest origin airport
2) Trip from nearest destination airport to the destination coordinates

Section 5.5: Findings

After this process of finding the nearest airport, we were finally able to perform analysis on the trip files. The biggest issue that we found with the long trip data was that it contained way more trips than were actually taken by planes historically. In addition to this, certain areas and states had a disproportionate amount of trips originating from them such as Alabama.
We will first summarize the findings of the plane trips and then follow up by comparing it with historical data. We will also try to make a conjecture on which states have the most disproportionate number of trips.

Number of plane trips per day	105.8 million
Plane miles travelled (Plane Miles = sum of miles travelled by each passenger)	162.23 billion miles
Busiest time	5:30 pm - 6:30 pm
Busiest Airport	Huntsville International Airport-Carl T Jones Field
Number of Busiest airports	377

Figure 7a

Note: The data has been scaled to represent the entire United States (i.e. we multiplied our metrics by 200 in order to scale)

Comparing Generated Data with Historical Data

In order to better understand the validity of our results, we decided to compare the data that we had generated from the Long Trip files with historical data from the Bureau of Transportation Statistics.

According to the Bureau of Transportation Statistics there were about 895 million plane trips undertaken in 2015 which results in 2.57 million trips per day. In addition to this there were about 1205.7 billion plane miles in a year which results in about 3.306 billion miles per day. The data from the Bureau of Transportation Statistics assumes each passenger is a traveler who takes a plane ride. Therefore, if the same passenger were to take a plane five times, that would count as five different passengers in this dataset.
The comparison on a per day basis is given below:

	Our Long Trip Data	Bureau of Transportation Statistics
Number of Trips	105.8 million trips per day	2.57 million trips per day
Plane Miles Travelled	162.23 billion miles per day	3.306 billion miles per day

Figure 7b

Temporal distribution of Trips Across the United States

Here is the temporal distribution of the trips across the United States for. These numbers have **not** been scaled and represent the sample data. We believe the distribution is very similar to the entirety of the data because our samples were completely random. As you can see the distribution appears to be skewed towards the right and the highest amount of plane trips are from 5:30 pm to 6:30 pm. This is understandable since most people travel in the evening.

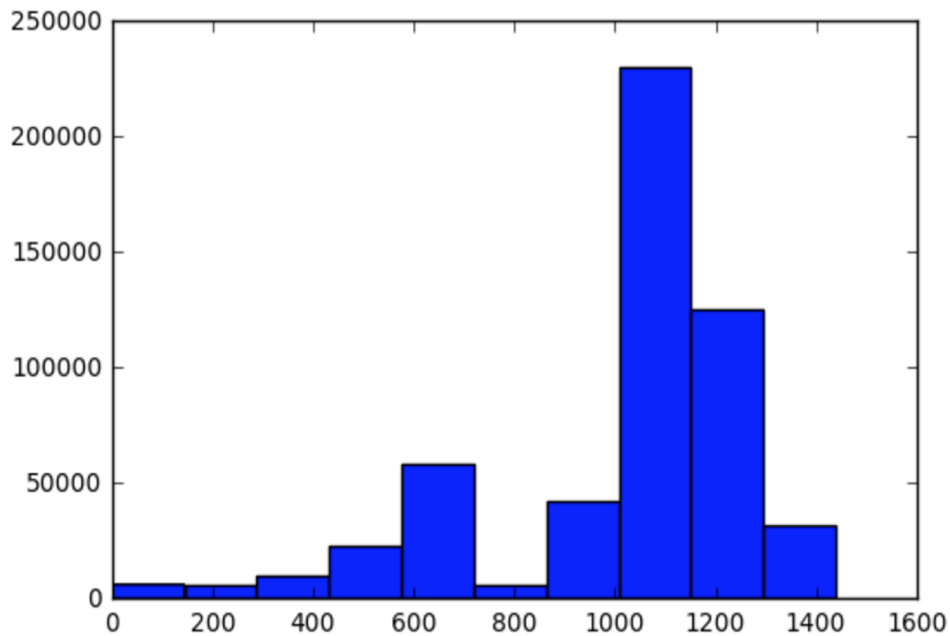


Figure 12c: Minutes After Midnight

Distribution of Trips according to Trip Type

We also analyzed the plane and by extension, the aTaxi trips according to trip type and got the following graph. These graphs are true for both aTaxis and Plane trips because the first type of trips are derived from the latter. As expected, almost all the trips are originating from either home, work or other. While other is a broad category, we believe that there is a reasonable chance that most of the other trips also originate from a place that a person might be at because of work. As you can see in the pie chart below, there are hardly any trips originating from schools.

\

Origin Trip Type

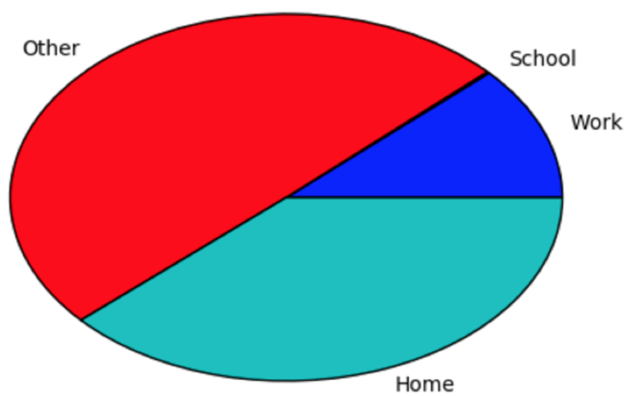


Figure 12d

Destination Trip Type

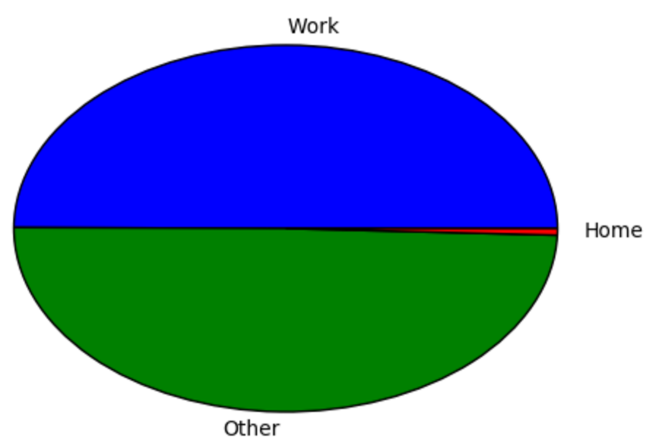


Figure 12e

Distribution of Plane Trip Distances

Let us now analyze the distribution of the distances of the plane trips. As it was a large dataset, we decided to go ahead with a histogram with bins of distance ranges rather than having a line graph as we did before for Assignment 6. It makes sense that a majority of the trips are between 500 miles and 1000 miles. The number of trips decreases with an increase in distance.

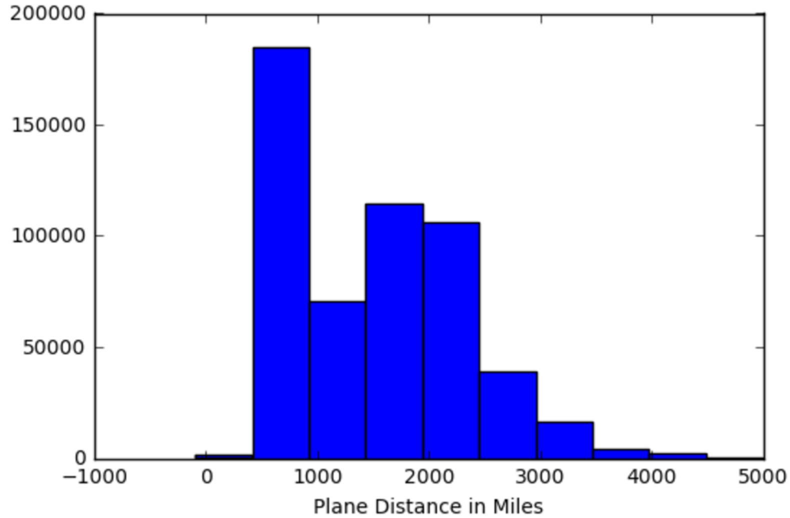


Figure 12f

General Statistics:

	Distance (miles)
Mean	1491.3
25%	700.3
50%	1540.7
75%	2011.1

Spatial Distribution

In order to better understand the deficiencies and the issues in the data we were able to plot an interactive heatmap of the origination and destinations in the entire United States using our sample dataset result.csv. Please find below screenshots of the heatmap. Note that the scale is the same for both the origin and destination heatmaps. To create these maps, we used a python package called folium which allows you to set the magnitude (which we set to be the same value in both maps).

Heat Map for Origins

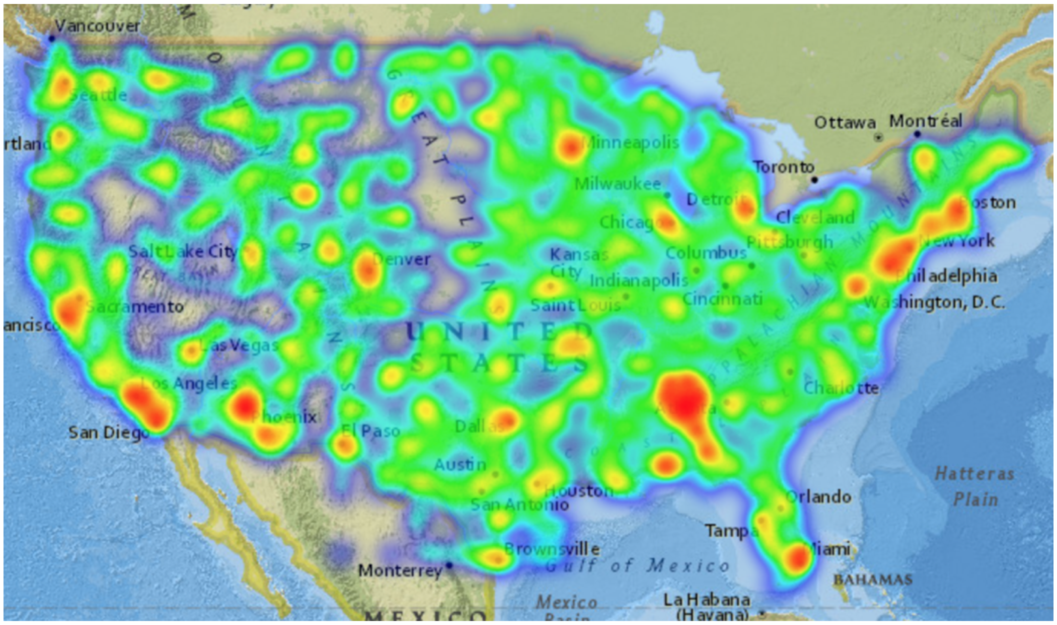


Figure 12g

Heat Map for Destinations for the entire US

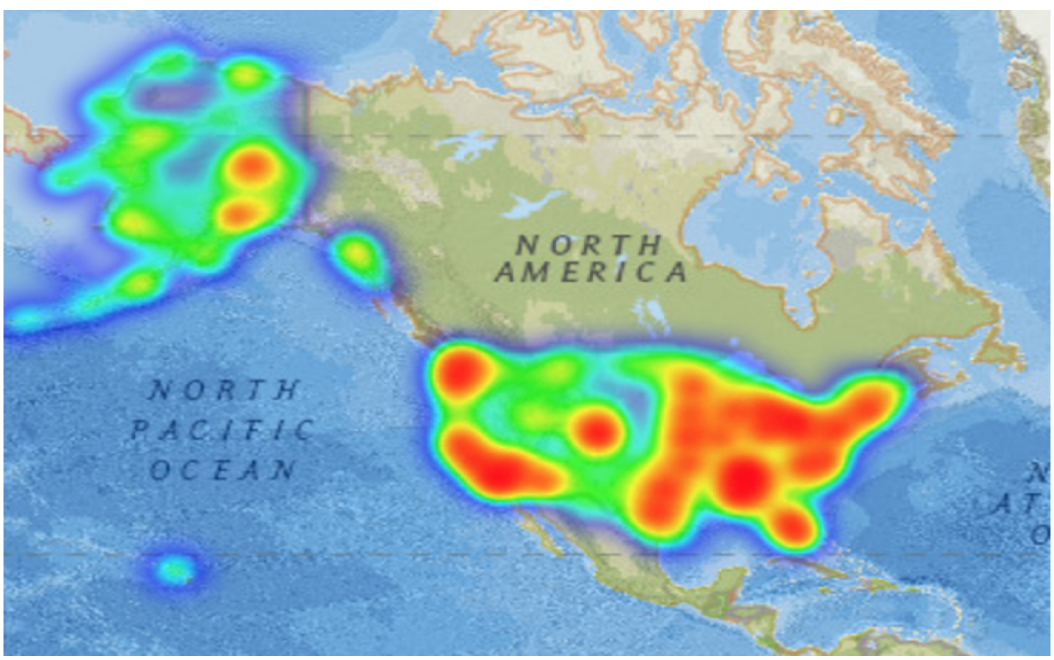


Figure 12h

Heat Map for Destinations (Zoomed In)

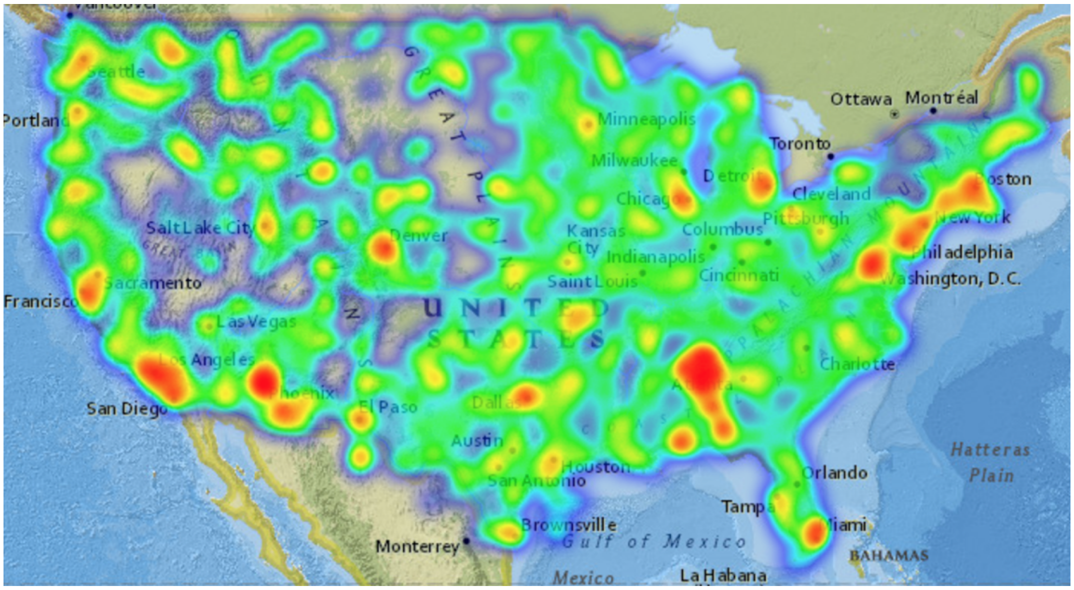


Figure 12i

Spatial Comparison of Busiest Origin Airports to Historical Data

We have included circle maps of historical busiest airports and our long trip data busiest airports. The larger the circle radius, the more trips there are from the airport. The historical data is from the Bureau of Transportation Statistics (BTS). Each airport has its IATA code next to it. Note that for both graphs, we only included the top 30 airports so the maps are bit clearer. We used the same scale for both maps.

Historical Data from Bureau of Transportation Statistics

Figure 12j

Our Long Trips Data

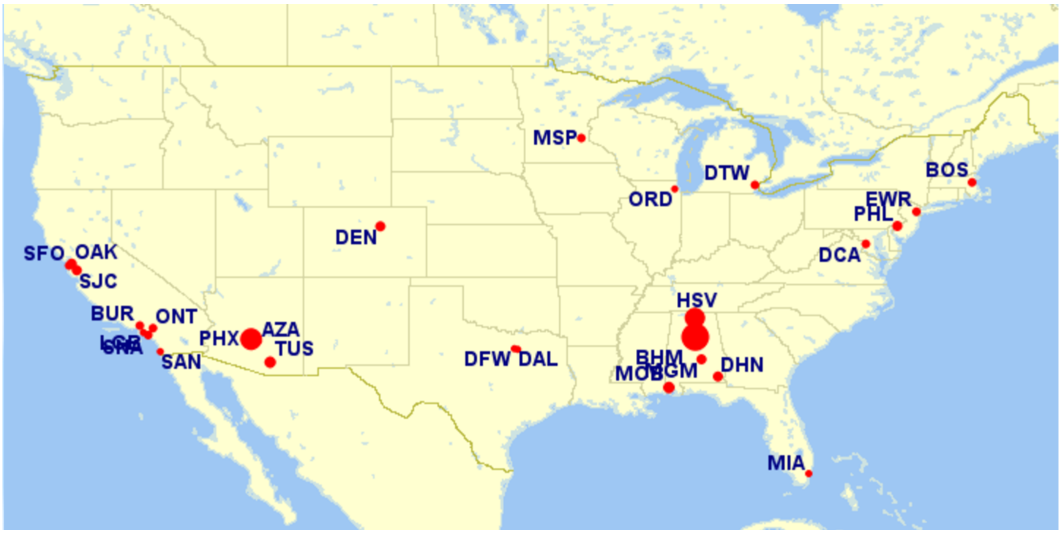


Figure 12k

Similarities:

- In both cases we see that many of the most popular airports are in the northeast and the west coast

Differences:

- It is clear that we have a lot more trips than we should in our long trip data by the size of the circles. In the historical data, the circles are so small because compared to our long trip data there are far fewer trips
- In addition to this, the top 30 busiest airports do not match
- We are missing important airports like JFK and ATL which are absent from our list of top 30 airports by enplanements.
- There are a lot more trips from the states of Alabama, Alaska, Arizona, and California in the long trip data than in the historical data.
- It is clear that there are some FIPS codes whose number of trips have not been properly generated. The biggest issue was in Alabama. Reasonably, given the population of cities like Birmingham in Alabama(~200,000), it is surprising that such a huge number of trips are originating from the region(~140,000 per day).

Similarly, small towns like Montgomery, which has a population of ~200,000 people, and Tuscaloosa, which has a population of ~90,000 people, also look more dense than Atlanta, which has a higher population and a historically busier airport. The zoomed in heat map below shows the issues with Alabama long trips in Figure 12

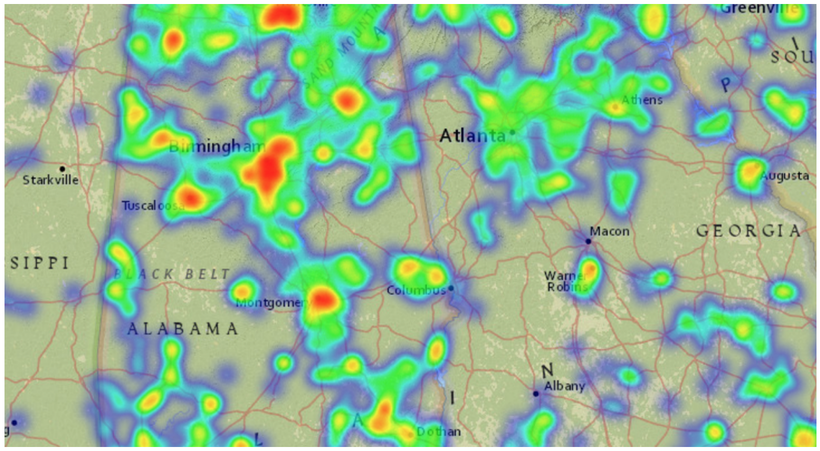


Figure 12l

Closeup Circle Maps

California:

Enplanements in our data / Enplanements in Historical Data

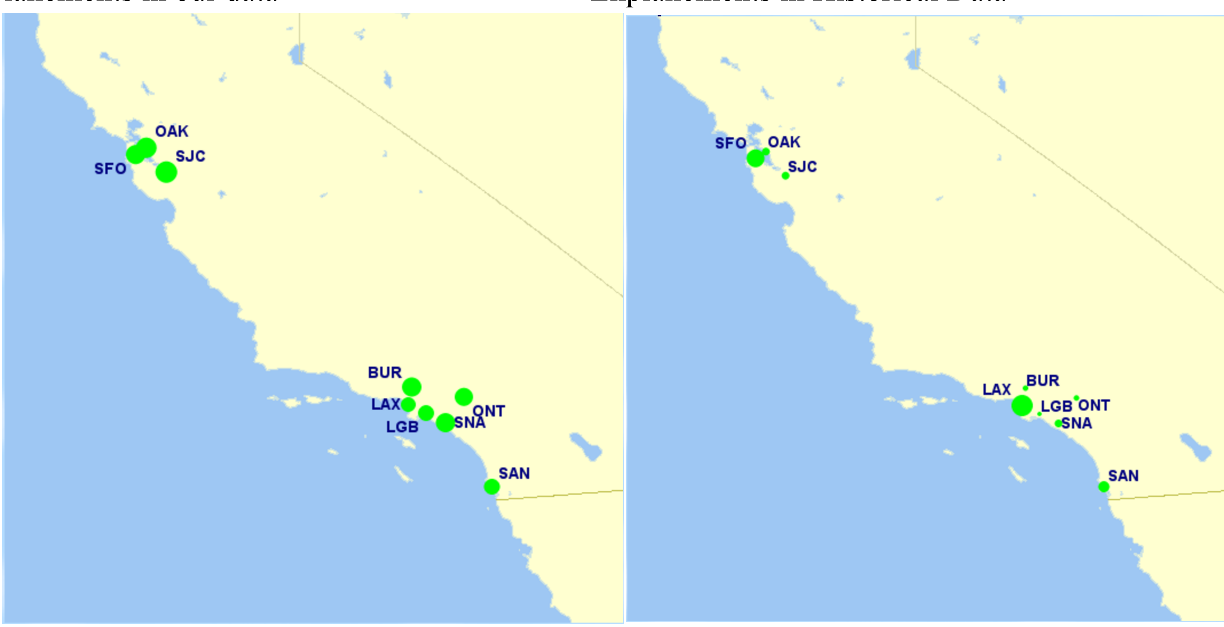


Figure 12m

California appears to be fine apart from the inflated number of trips in our data as compared to historical data.

New York:

Enplanements in our data — Enplanements in Historical Data

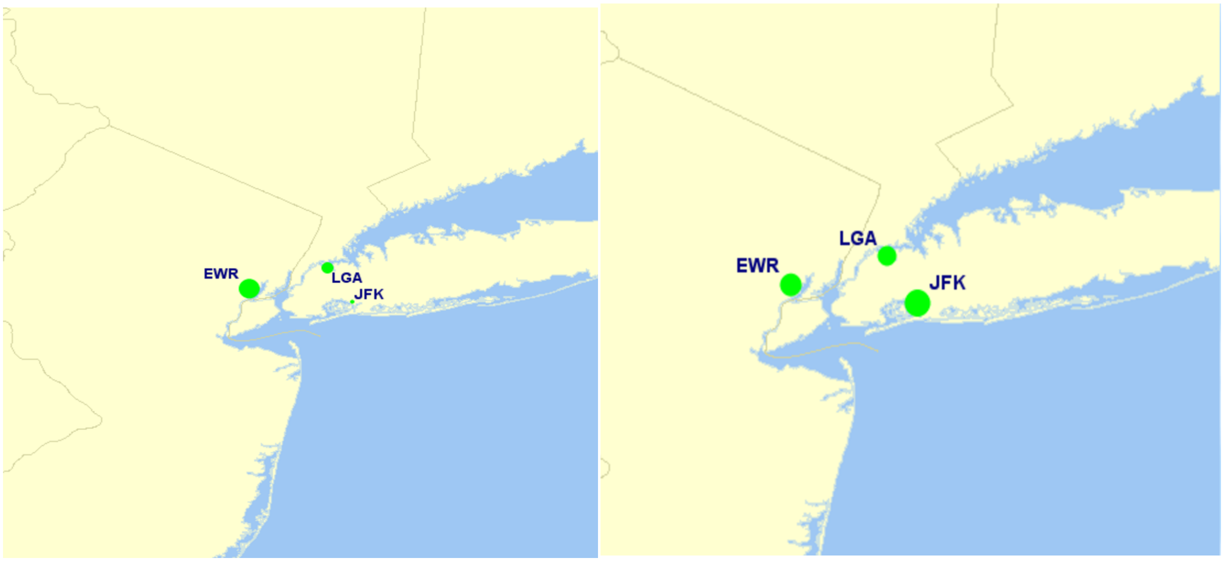


Figure 12n

This comparison shows that the long trip data is clearly wrong because the number of trips coming from LGA and JFK are far too small. Generally, the long trip data included too many trips for most of the airports but in this case, there are too little trips coming from historically busy airports such as LGA and JFK.

Alabama/Atlanta:

Enplanements in our data — Enplanements in Historical Data

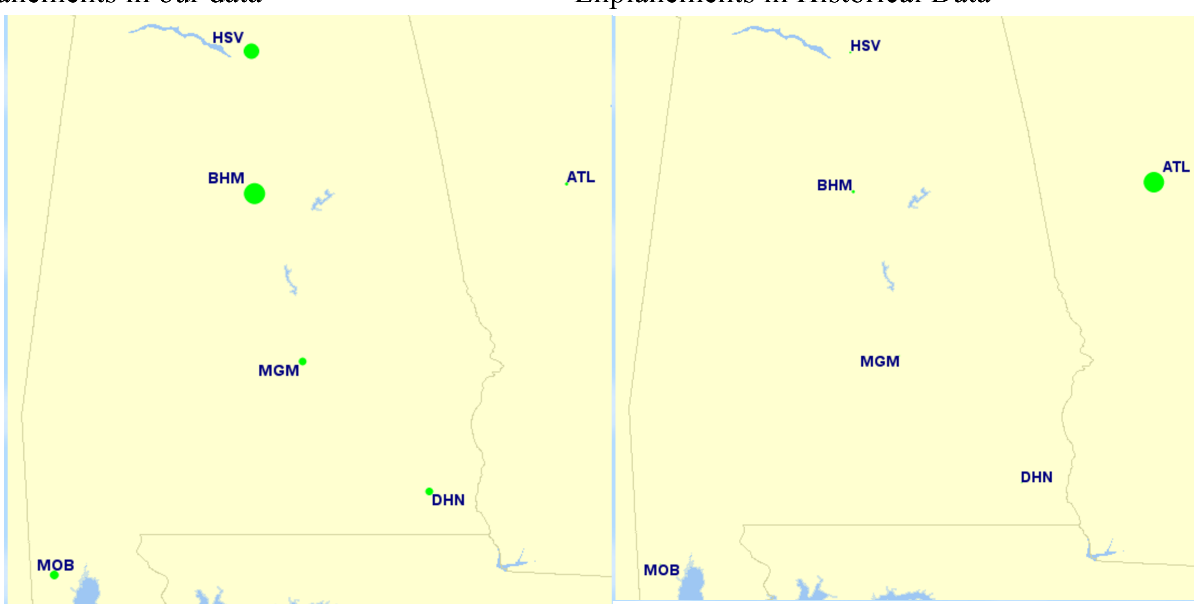


Figure 12o

It is easy to see that our data for BHM (Birmingham), MGM(Montgomery) and ATL(Atlanta) is way off compared to historical estimates.

Busiest Airports
Comparing our busiest origin airports to historical busiest airports we have:

Rank	Origin Airport	Historical Busiest Airports
1	Birmingham Intl	Hartsfield - Jackson Atlanta International
2	Phoenix Sky Harbor Intl	Los Angeles International
3	Huntsville International Airport-Carl T Jones Field	Chicago O'Hare International
4	Merrill Fld	Dallas/Fort Worth International
5	Mobile Rgnl	John F Kennedy International
6	Tucson Intl	Denver International
7	Dothan Rgnl	San Francisco International
8	Norman Y Mineta San Jose Intl	Charlotte/Douglas International
9	Montgomery Regional Airport	McCarran International
10	Metropolitan Oakland Intl	Phoenix Sky Harbor International

Figure 12p

Busiest Airport by Passenger Boarding was retrieved from the FAA at the following link: *https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/media/cy14-commercial-service-enplanements.pdf*

Top 10 Origin Airports with the Most Trips

Full data can be accessed in the originairportscount.csv file
https://www.dropbox.com/sh/uwrovzswl7x73d9/AAB9MpgYGmeEaJP22DkRgAz1a?dl=0&preview=originairportscount.csv
Note that this analysis was performed on the sample data so therefore the actual number of trips of much higher.

Origin Airport	# of Trips
Birmingham Intl	56,073
Phoenix Sky Harbor Intl	34,997

Huntsville International Airport-Carl T Jones Field	29,814
Merrill Fld	21,717
Mobile Rgnl	10,251
Tucson Intl	10,082
Dothan Rgnl	8,338
Norman Y Mineta San Jose Intl	7,641
Montgomery Regional Airport	7,538
Metropolitan Oakland Intl	6,886

Figure 12q

Top 10 Destination Airports with the Most Trips

Full data can be accessed in the destairportscount.csv file
https://www.dropbox.com/sh/uwrovzswl7x73d9/AAB9MpgYGmeEaJP22DkRgAz1a?dl=0&preview=destairportscount.csv
Note that this analysis was performed on the sample data so therefore the actual number of trips of much higher.

Destination Airport	# of Trips
Birmingham Intl	55,596
Phoenix Sky Harbor Intl	35,169
Huntsville International Airport-Carl T Jones Field	29,710
Merrill Fld	21,753
Mobile Rgnl	10,675
Tucson Intl	10,012
Dothan Rgnl	8,506
Montgomery Regional Airport	7,719
Norman Y Mineta San Jose Intl	7,423
Phoenix-Mesa Gateway	6,967

Origin Airport to Destination Airport Highest Pair Counts

The top 10 pairs of origin and destination airports is shown below and the full data is located in airportsgrouped.csv
https://www.dropbox.com/sh/uwrovzswl7x73d9/AAB9MpgYGmeEaJP22DkRgAz1a?dl=0&preview=airportsgrouped.csv
Note that this analysis was performed on the sample data so therefore the actual number of trips of much higher.

Origin Airport	Destination Airport	# of Trips
Huntsville International Airport-Carl T Jones Field	Norman Y Mineta San Jose Intl	4,902
Norman Y Mineta San Jose Intl	Huntsville International Airport-Carl T Jones Field	4,737
Birmingham Intl	Dallas Love Fld	3,954
Phoenix Sky Harbor Intl	Huntsville International Airport-Carl T Jones Field	3,766
Huntsville International Airport-Carl T Jones Field	Phoenix Sky Harbor Intl	3,520
Dallas Love Fld	Birmingham Intl	3,486
Birmingham Intl	Philadelphia Intl	3,431
Miami Intl	Birmingham Intl	3,320
Birmingham Intl	John Wayne Arpt Orange Co	3,244
Birmingham Intl	Miami Intl	3,200

Figure 12s

Below is the historical data for busiest pairs of airports (http://www.transtats.bts.gov/) from the Bureau of Transportation Statistics:

United States (September 2014 - August 2015) [edit]

Busiest air routes by city pairs within the United States [11]

Rank	City 1	City 2	Passengers (Sep 2014-Aug 2015)
1	Chicago, IL (Metro Area)	New York City, NY (Metro Area)	4,020,000
2	Los Angeles, CA (Metro Area)	San Francisco, CA (Metro Area)	3,660,000
3	Los Angeles, CA (Metro Area)	New York City, NY (Metro Area)	3,420,000
4	Chicago, IL (Metro Area)	Los Angeles, CA (Metro Area)	3,010,000
5	Miami, FL (Metro Area)	New York City, NY (Metro Area)	2,750,000
6	Atlanta, GA (Metro Area)	Chicago, IL (Metro Area)	2,720,000
7	Chicago, IL (Metro Area)	Minneapolis, MN (Metro Area)	2,720,000
8	Atlanta, GA (Metro Area)	New York City, NY (Metro Area)	2,600,000
9	Atlanta, GA (Metro Area)	Orlando, FL (Metro Area)	2,620,000
10	Chicago, IL (Metro Area)	Washington, DC (Metro Area)	2,620,000

Figure 12t

This comparison of our data with historical data shows us that we have **way too many trips from Alabama, Arizona, and Alaska.** In particular, Alabama has highly disproportionate number of trips. Many of our top 10 busiest airports were from Alabama. The long trip data also does not have enough trips from some of the top busiest airports in Atlanta, Chicago, Dallas, Los Angeles, New York, and San Francisco. It seems like the long trip data that we received had many more data points coming from the first section of the data, namely from states that start with A.

aTaxis

There are two aTaxi trips for each plane trip. The first trip is from the origin to the nearest origin airport and the second trip is from the nearest destination airport to the destination. The aTaxi analysis was performed on the sample data because the complete data was simply too large.

Important Findings for the sample of the data

aTaxi miles: 16,672,925
Number of aTaxi trips: 215.6 Million
AVO: 1
Minimum Fleet size: 531,443 aTaxis
<u>Important Findings Scaled for the entirety of the data:</u>
aTaxi miles: 3,334,585,000 miles
Number of aTaxi trips: 43.1 Billion
AVO: 1
Minimum Fleet size: 106,288,600 aTaxis

aTaxi Distances from Origin to nearest Origin Airport

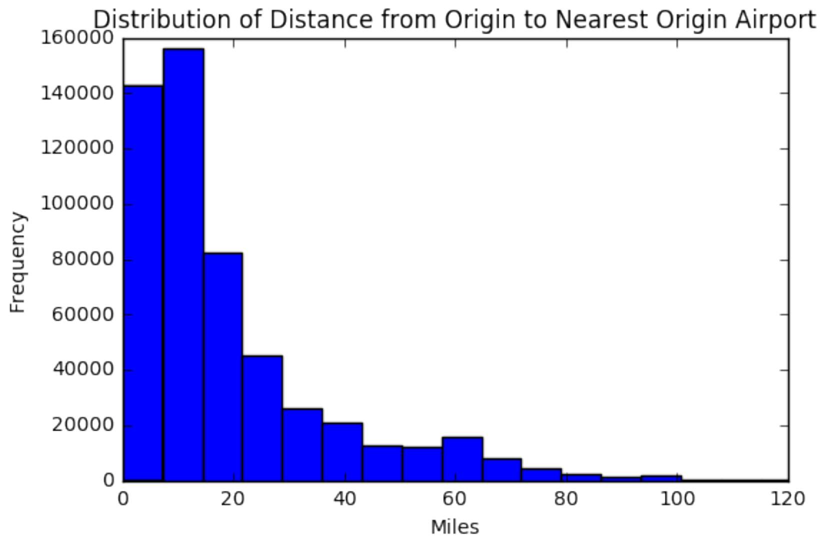


Figure 12u

General Statistics:

	Distance (miles)
Mean	21.8
25%	6.9
50%	12.8
75%	24.8

aTaxi Distances from Nearest Destination Airport to Destination

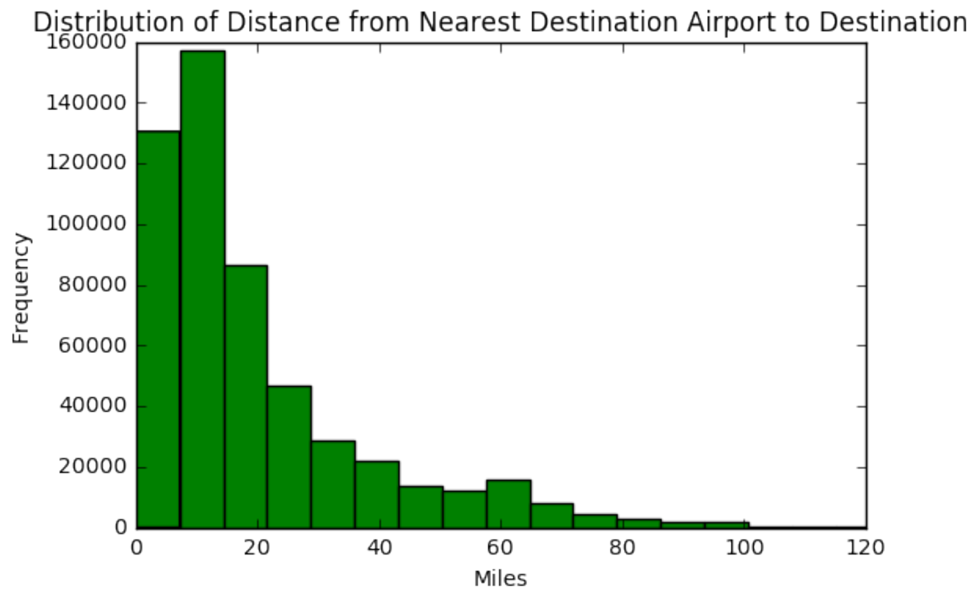
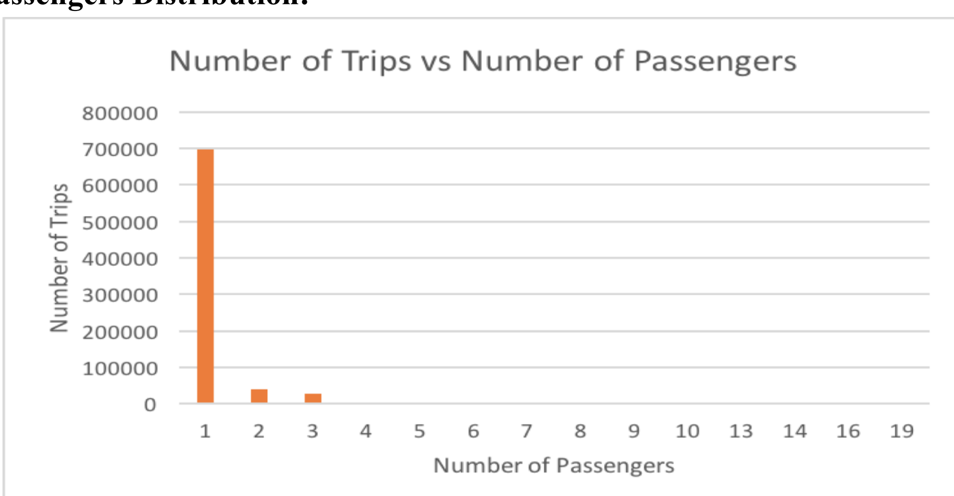


Figure 12v

General Statistics:

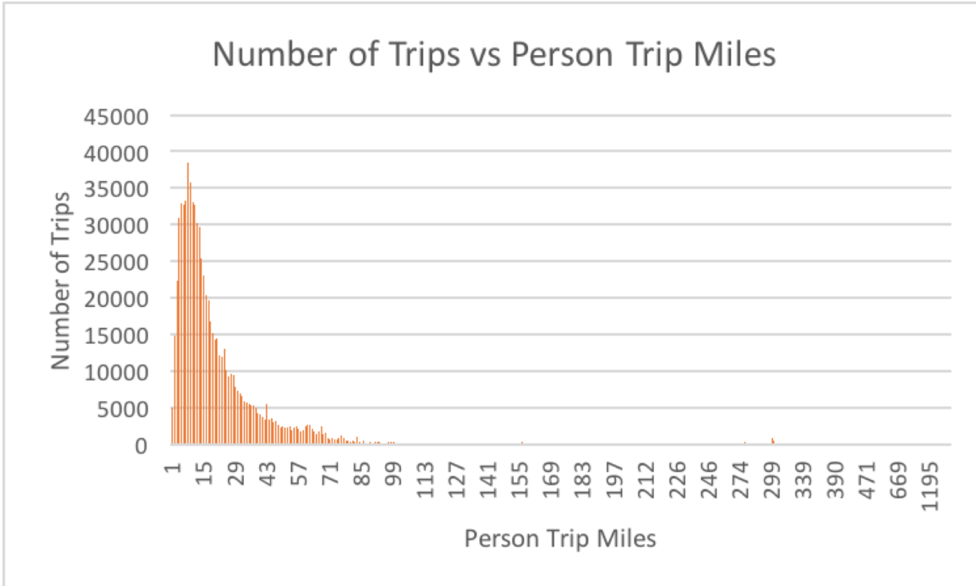
	Distance (miles)
Mean	22.3
25%	5.3
50%	13.4
75%	26.0

Number of Passengers Distribution:

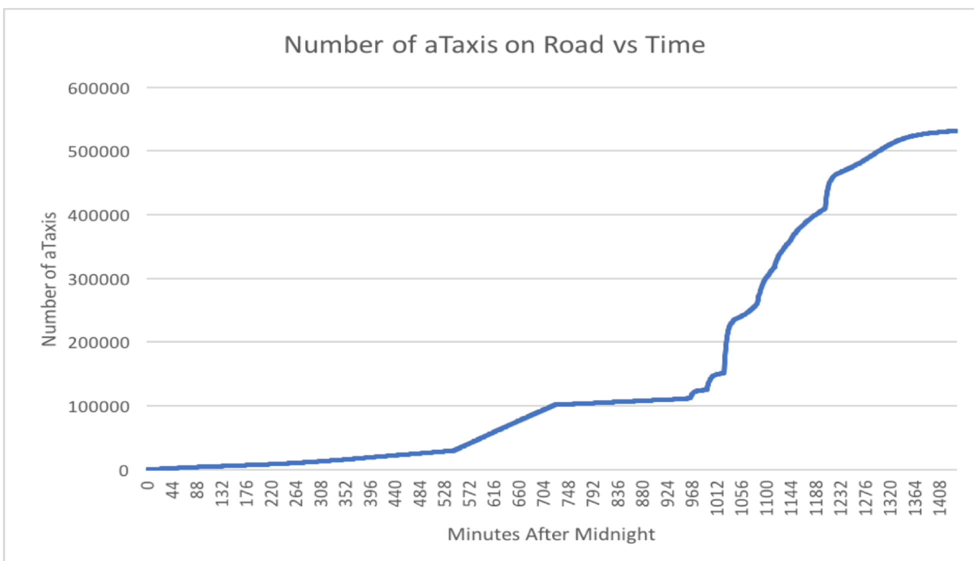


This graph tells us that there is not much ridesharing going on to the airports. It also explains why we got our aTaxi AVO (Average Vehicle Occupancy) to be 1(up to 4 significant figures).

Person Trip Miles Distribution:



Temporal Distribution:



Chapter 8:
Nationwide AVO

The task was to take the nationwide Person Trip files and generate the set of nationwide aTaxi Vehicle Trips assuming infinite size aTaxis. Section 8.1 provides an overview of the input data, Section 8.2 analyzes the Person Trips, Section 8.3 walks through the aTaxi Vehicle Trip File generation process, and Section 8.4 analyzes the aTaxi data and evaluates implications for Average Vehicle Occupancy.

August Kiles, Bill Van Cleve, Tianay Zeigler

Section 8.1 Overview of Input Data

This section will describe the input data, the nationwide Person Trips. Thank you to Kyle Marocchini '18 for producing these files. We were given [CSV files] labeled by FIPS code corresponding to trips originating in all 50 states plus the District of Columbia. The files were about 115gb in all. Each file contained roughly a million or fewer Person Trips. For many counties, one file was enough to contain all Person Trips originating in that county. For the larger counties, though, the trips were split into multiple files for a single FIPS code. Each row corresponds to a single trip taken by one person. As an example, Figure 8a shows the beginning of file "NewRecoveredOriginPixel40001_1.csv", which contains all trips originating in the county of Adair, Oklahoma.

Figure 8a: "NewRecoveredOriginPixel40001_1.csv"

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	County Co	Person ID	Trip Type	OType	OName	OFIPS	OLon	OLat	OXCoord	OYCoord	ODepartu	DType	DName	DFIPS	DLon	DLat	DXCoord	DYCoord	GCDistance
2	40001	19149	15	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	2660	H	Home	40001	-94.6471	35.75373	-2075	-435	14.89984
3	40021	353079	14	H	Home	40021	-94.825	35.91464	-2094	-413	34912	W	WENDY'S	40101	-95.4027	35.75302	-2157	-435	34.27019
4	40001	9919	13	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	44480	H	Home	40001	-94.6958	35.65368	-2080	-449	19.45101
5	40001	5409	8	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	46518	O	BALDOR E	40001	-94.5676	35.9962	-2066	-402	15.33978
6	40001	11291	12	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	60659	H	Home	40001	-94.7994	35.8936	-2091	-416	1.993548
7	40001	5598	12	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	61237	H	Home	40001	-94.7825	35.9627	-2090	-406	3.976178
8	40001	22452	12	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	62343	H	Home	40001	-94.5551	35.73741	-2065	-438	19.40886
9	40001	7580	18	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	68757	H	Home	40001	-94.5792	35.98423	-2067	-403	14.44481
10	40001	11289	15	O	COLLINS S	40021	-94.8227	35.91536	-2094	-413	71872	H	Home	40001	-94.7994	35.8936	-2091	-416	1.993548
11	40001	1	14	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	37491	O	AMERICA	40001	-94.637	35.80859	-2074	-428	12.29363
12	40001	4601	14	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	42464	O	NORTHWE	5007	-94.5126	36.18318	-2060	-376	24.95361
13	40001	5077	3	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	42555	H	Home	40001	-94.7851	36.07039	-2090	-392	10.80711
14	40001	10870	3	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	42791	H	Home	40001	-94.7897	35.90683	-2090	-414	1.39036
15	40001	5294	3	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	43925	H	Home	40001	-94.6775	36.05021	-2078	-394	11.96875
16	40001	10165	13	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	49616	H	Home	40001	-94.774	35.95026	-2089	-408	3.202264
17	40001	20670	8	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	50156	H	Home	40001	-94.5552	35.85692	-2065	-421	14.95474
18	40001	10885	13	O	ON-SITE C	40021	-94.812	35.91568	-2093	-413	54557	H	Home	40001	-94.7913	35.90114	-2091	-415	1.536073
19	40001	2925	15	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61233	O	STEPPING	5007	-94.1769	36.38367	-2024	-348	48.02362
20	40001	3027	17	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61257	H	Home	40001	-94.6296	36.05387	-2073	-394	13.98285
21	40001	11172	19	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61279	O	TRANSPLA	5007	-94.1213	36.25844	-2018	-366	45.305
22	40001	16656	6	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61369	O	M CLARK	5007	-94.165	36.36675	-2022	-351	47.74436
23	40001	1797	1	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61403	H	Home	40001	-94.6379	36.10782	-2074	-386	16.47626
24	40001	6174	19	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61426	O	MID-AME	5007	-94.5315	36.24685	-2062	-367	27.75945
25	40001	17911	19	W	ON-SITE C	40021	-94.812	35.91568	-2093	-413	61592	O	DR GREG	5007	-94.5594	36.18204	-2065	-376	23.21702

Columns G and H give the specific latitude and longitude coordinates where a trip originates. Column K gives the exact time of day, in seconds, when the person departs. OName and DName, columns E and L, describe where the individual began and ended the trip. DLon and Dlat give the ending coordinates. The final column, GCDistance, gives the distance of the trip, calculated with GreatCircleDistance.

Kyle's files contain all trips with GCDistance under 200 miles. For the purpose of analysis, we only examined trips with GCDistance greater than 2 miles, but less than 200, leaving the analysis of very short trips and very long trips to other groups.

Section 8.2 Analysis of Person Trips and Person Trip Miles

Before generating the aTaxi Vehicle Trip files, the first step was to analyze the existing Person Trips (between 2 and 200 miles). the program looked through the Person Trip files and summed up the number of Person Trips originating in that state by state as well as the total number of Person Trip miles by state. Figure 8b lists the total number of "regTrips," or trips between 2 and 200 miles, by state.

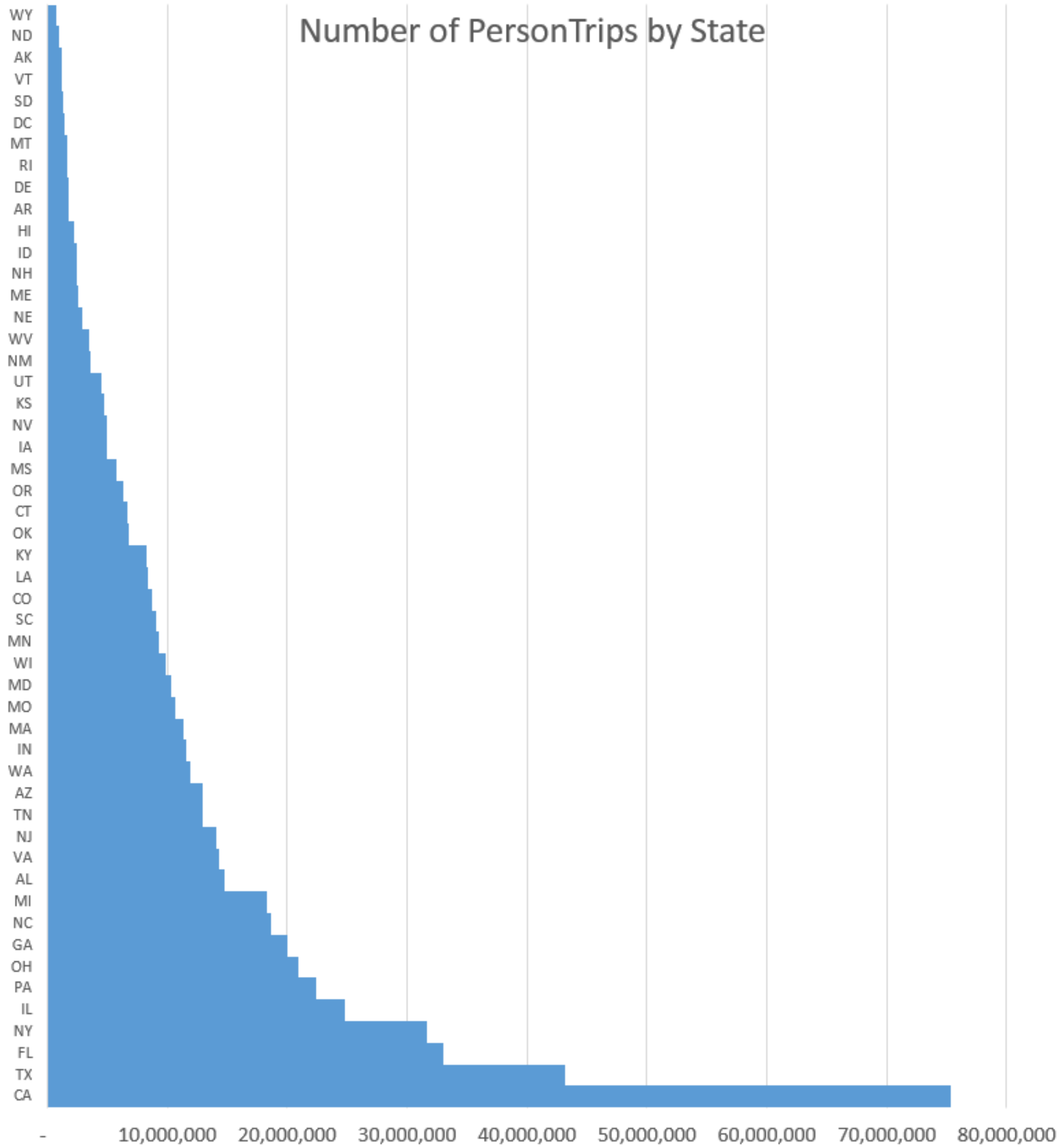
Figure 8b: Trips by state

State	State Code	FIPS code	regTrips	regTrips %
California	CA	6	75,274,095	13.3%
Texas	TX	48	43,164,348	7.6%
Florida	FL	12	32,998,488	5.8%
New York	NY	36	31,693,655	5.6%
Illinois	IL	17	24,782,640	4.4%
Pennsylvania	PA	42	22,407,244	4.0%
Ohio	OH	39	20,892,719	3.7%
Georgia	GA	13	19,976,818	3.5%
North Carolina	NC	37	18,600,152	3.3%
Michigan	MI	26	18,323,460	3.2%
Alabama	AL	1	14,817,309	2.6%
Virginia	VA	51	14,289,469	2.5%
New Jersey	NJ	34	14,073,301	2.5%
Tennessee	TN	47	12,943,835	2.3%
Arizona	AZ	4	12,927,641	2.3%
Washington	WA	53	11,952,819	2.1%
Indiana	IN	18	11,564,843	2.0%
Massachusetts	MA	25	11,358,238	2.0%
Missouri	MO	29	10,713,995	1.9%
Maryland	MD	24	10,380,493	1.8%
Wisconsin	WI	55	9,881,693	1.7%
Minnesota	MN	27	9,303,861	1.6%
South Carolina	SC	45	9,019,553	1.6%
Colorado	CO	8	8,706,925	1.5%
Louisiana	LA	22	8,391,554	1.5%
Kentucky	KY	21	8,278,433	1.5%
Oklahoma	OK	40	6,776,061	1.2%
Connecticut	CT	9	6,662,969	1.2%
Oregon	OR	41	6,304,511	1.1%
Mississippi	MS	28	5,767,286	1.0%
Iowa	IA	19	4,977,361	0.9%

Nevada	NV	32	4,924,667	0.9%
Kansas	KS	20	4,707,145	0.8%
Utah	UT	49	4,521,535	0.8%
New Mexico	NM	35	3,609,809	0.6%
West Virginia	WV	54	3,513,089	0.6%
Nebraska	NE	31	2,903,605	0.5%
Maine	ME	23	2,546,828	0.5%
New Hampshire	NH	33	2,490,348	0.4%
Idaho	ID	16	2,417,771	0.4%
Hawaii	HI	15	2,267,887	0.4%
Arkansas	AR	5	1,759,568	0.3%
Delaware	DE	10	1,738,643	0.3%
Rhode Island	RI	44	1,671,057	0.3%
Montana	MT	30	1,612,789	0.3%
District of Columbia	DC	11	1,468,686	0.3%
South Dakota	SD	46	1,276,809	0.2%
Vermont	VT	50	1,229,261	0.2%
Alaska	AK	2	1,208,987	0.2%
North Dakota	ND	38	1,023,363	0.2%
Wyoming	WY	56	784,251	0.1%
Sum			564,881,867	100%

The table presents no big surprises, with California, Texas, Florida and New York coming out on top, as they are the more highly populated states. In total, we were working with just over 550 million Person Trips. (After walking trips, short haul trips, and long trips are taken into account, this number will be roughly 1 billion trips in total).

Figure 8c presents the results from Figure 8b graphically:



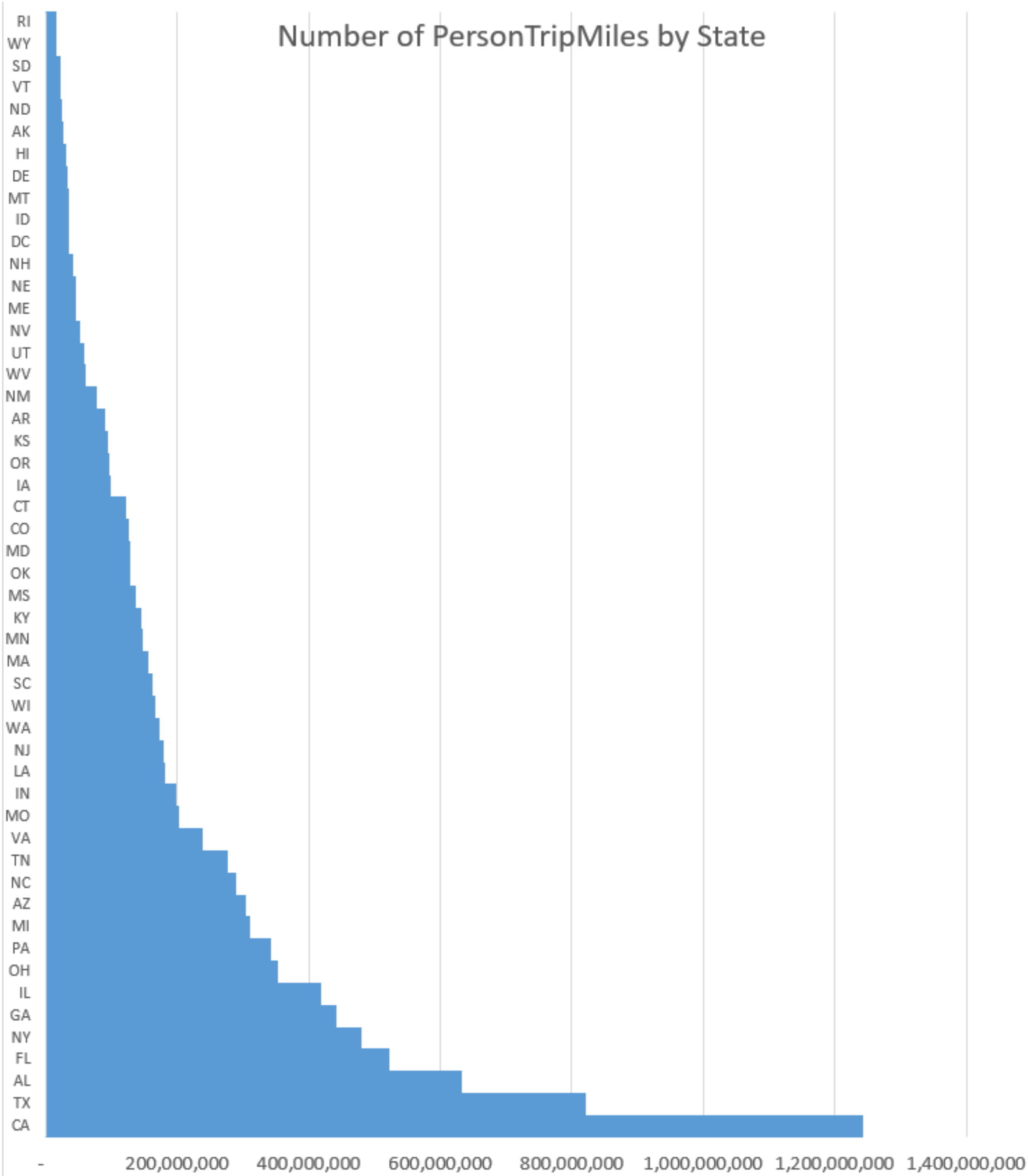
We also examined the states by number of Person Trip Miles, or the total distance of all trips originating in that state. The results are presented in Figure 8d.

Figure 8d: Trip miles by state

State	State Code	FIPS code	regTripsMiles	regTripsMiles%
California	CA	6	1,242,516,857	12.3%
Texas	TX	48	821,456,690	8.1%
Alabama	AL	1	632,158,317	6.3%
Florida	FL	12	523,231,835	5.2%
New York	NY	36	480,684,713	4.8%
Georgia	GA	13	442,483,195	4.4%
Illinois	IL	17	418,412,917	4.1%
Ohio	OH	39	352,905,195	3.5%
Pennsylvania	PA	42	343,308,031	3.4%
Michigan	MI	26	310,166,620	3.1%
Arizona	AZ	4	305,147,214	3.0%
North Carolina	NC	37	290,286,662	2.9%
Tennessee	TN	47	277,827,939	2.8%
Virginia	VA	51	238,670,438	2.4%
Missouri	MO	29	202,358,872	2.0%
Indiana	IN	18	198,413,872	2.0%
Louisiana	LA	22	182,573,621	1.8%
New Jersey	NJ	34	179,902,504	1.8%
Washington	WA	53	173,231,921	1.7%
Wisconsin	WI	55	167,141,424	1.7%
South Carolina	SC	45	162,688,604	1.6%
Massachusetts	MA	25	156,491,863	1.5%
Minnesota	MN	27	146,621,109	1.5%
Kentucky	KY	21	145,335,090	1.4%
Mississippi	MS	28	136,004,338	1.3%
Oklahoma	OK	40	128,165,956	1.3%
Maryland	MD	24	127,987,692	1.3%
Colorado	CO	8	126,435,035	1.3%
Connecticut	CT	9	122,783,818	1.2%
Iowa	IA	19	99,185,484	1.0%
Oregon	OR	41	96,007,576	1.0%

Kansas	KS	20	94,924,363	0.9%
Arkansas	AR	5	90,314,799	0.9%
New Mexico	NM	35	77,465,279	0.8%
West Virginia	WV	54	61,324,309	0.6%
Utah	UT	49	58,780,268	0.6%
Nevada	NV	32	52,185,662	0.5%
Maine	ME	23	46,785,614	0.5%
Nebraska	NE	31	44,873,601	0.4%
New Hampshire	NH	33	41,026,938	0.4%
District of Columbia	DC	11	35,545,392	0.4%
Idaho	ID	16	34,936,467	0.3%
Montana	MT	30	34,484,275	0.3%
Delaware	DE	10	34,042,765	0.3%
Hawaii	HI	15	30,742,076	0.3%
Alaska	AK	2	26,418,510	0.3%
North Dakota	ND	38	25,368,319	0.3%
Vermont	VT	50	22,970,887	0.2%
South Dakota	SD	46	22,871,003	0.2%
Wyoming	WY	56	16,502,851	0.2%
Rhode Island	RI	44	16,060,307	0.2%
Sum			10,098,209,086	100%

Figure 8e presents this information graphically:



Interestingly, several states move up in the rankings when sorted by Person Trip Miles. Alabama, for example, goes from 11th to 3rd, and Georgia goes from 8th to 6th. This indicates that the trips in those states are relatively longer compared to other more dense parts of the US. With the data from Person Trips and Person Trip Miles largely confirming what we expected to see, we moved on to generating the aTaxi Vehicle Trips.

Figure 8f: Person Trip count by state

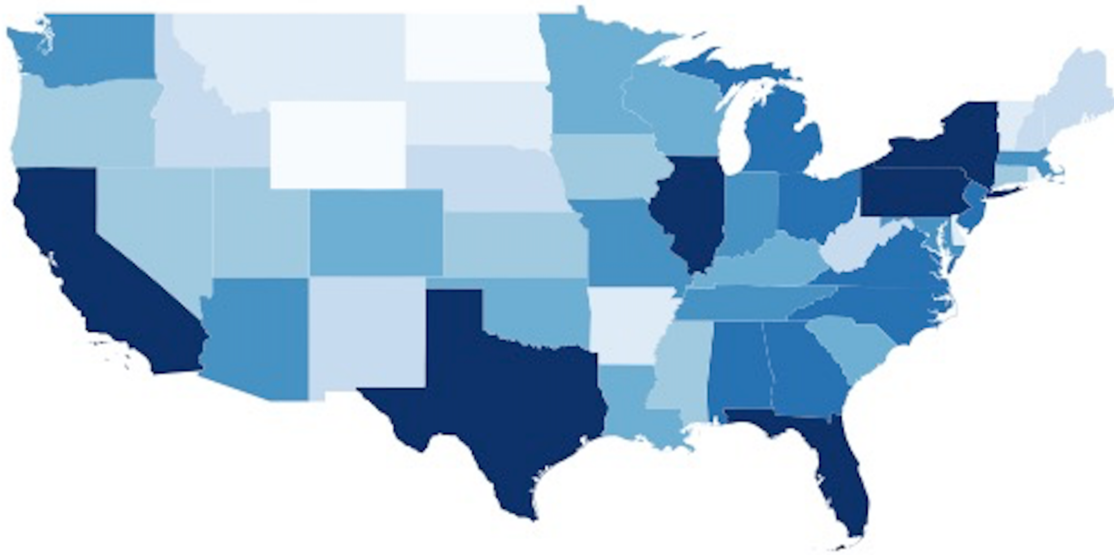
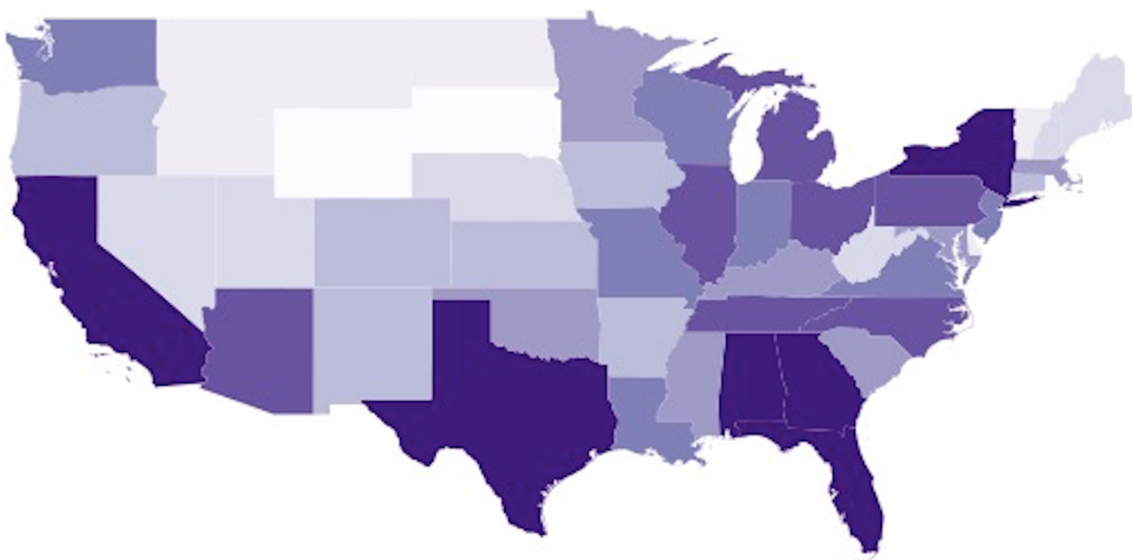
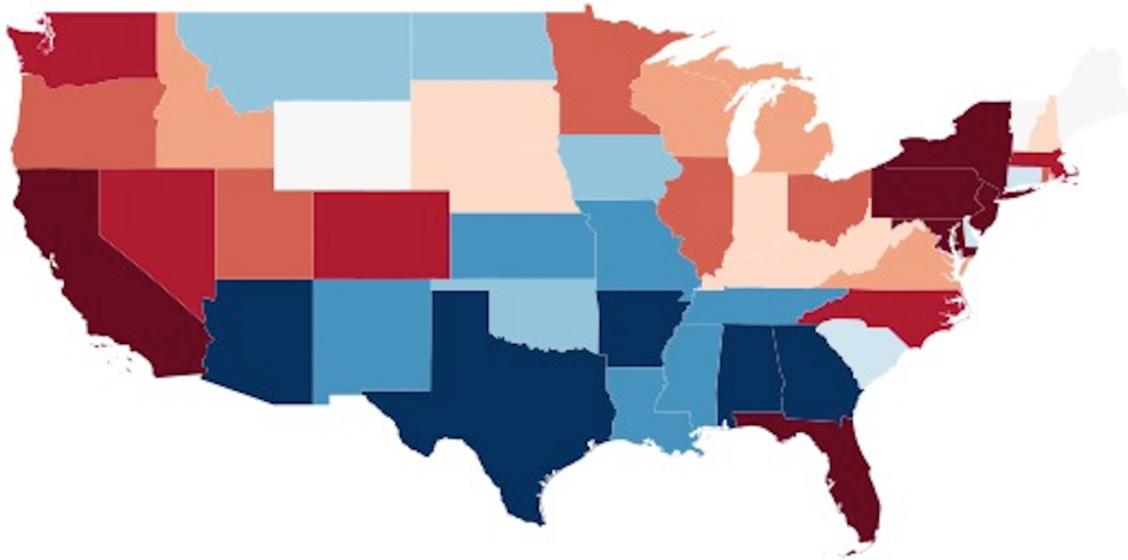


Figure 8g: Person Trip Miles by state



By calculating the quantile of both the number of person trips and total trip miles for each state, it was easy to see which states showed the most significant difference. In the map below, blue states showed an increase in quantile when the analysis shifts from number of trips to total miles, while red states showed a decrease. In other words, blue states are home to the longest average trip length while red states are home to the shortest. Darker colored states demonstrate a greater discrepancy between the two.

Figure 8h: Trip Length heat map



Section 8.3 Description of aTaxi Vehicle Trip File generation

Level of Service

Before writing the code to generate nationwide aTaxi Vehicle Trips, we needed to make some assumptions about the Level of Service (LOS) we would provide. To start, we set Common Destinations (CD) = 3. This means that no aTaxi will stop more than 3 times in total after picking up passengers. The rationale here is simply that customers will tolerate a couple stops along their route, but not more than 2. We then assumed a Departure Delay (DD) of 300 seconds, or 5 minutes. This means that as soon as the first passenger gets in the aTaxi, it will wait no more than 5 minutes for other passengers to come along and rideshare before departing. Again, this constraint has to do with how much waiting customers will be able to tolerate. The next LOS term we set was a Maximum Circuity of 25%. With this constraint in place, no passenger will travel more than 125% of the distance they would have if they were going in a straight line to their destination. Finally, we assumed an infinite aTaxi size, leaving the analysis of aTaxi size to other groups. These constraints are simply reasonable starting points to get an idea of what the nationwide ridesharing potential might be; they could be altered to deliver a better passenger experience or to try to increase the average vehicle occupancy (AVO).

Description of aTaxi Vehicle Trips generation

Since the ridesharing plan relies entirely on people leaving from the same pixel and all departures from a single pixel are contained in one of the PersonTrips files, we were able to examine each PersonTrip file only once. Further, since the PersonTrip files were ordered first by pixel, and then by time of departure, we were able to read each row of each PersonTrip file in order. This meant that we never had to read a Person Trip more than once or jump around within the PersonTrip file.

This paragraph describes the basic steps of the algorithm. the outer loop simply iterated through every PersonTrip file. Within each file, we iterate through all Person Trips in that file. Once the program reads in a departure from a given pixel, that departure is automatically entered into the aTaxiBoarding variable. This variable essentially keeps track of all the aTaxis that have not departed their arrival point yet. The first person trip is loaded into the variable, and the first aTaxi is created to accommodate this trip. Each sequential person trip that is entered into the loop is first compared to the trips in each aTaxi that has not yet been boarded. If the destination, and time of departure are similar enough that a ridesharing trip would meet the criteria described above, then the person trip is added to this preexisting aTaxi. (If that aTaxi has multiple destinations, the route is reconfigured to be optimal.) If, on the other hand, no yet-to-be-departed aTaxi works with the new person trip, then the function creates a new aTaxi. Each of these aTaxis are departed after a designated period of five minutes. Once we get to a new departure pixel, all of the aTaxis still waiting at the old pixel are departed. This process was continued until the program had looped through every person trip in the file, and ultimately every file in the directory.

Format of output files

This section describes the files that we produced as output. the program outputted one aTaxi Trip file for every state plus D.C., for a total of 51 files. In total, the files took up 27.2 gb. Each row in a file corresponds to one aTaxi trip, and the columns are defined in the table below.

1	ith aTaxi
2	oXpixel
3	oYpixel
4	DepartureTimeFromThisPixel (Person Trip oTime +300)
5	Riders@Departure
6	CD1XPixel
7	CD1YPixel
8	RidersGettingOff@CD1
9	CD2XPixel
10	CD2YPixel
11	RidersGettingOff@CD2
12	CD3XPixel
13	CD3YPixel
14	RidersGettingOff@CD3
15	PersonTripMilesServed
16	DXPixel
17	DYPixel
18	ArrivalTime
19	ATaxiVehichleMiles
20	AVO
21	FIPS Code

Most of the columns are self-explanatory, but a few require explanation. Columns 6, 7, and 8 give the location of the aTaxi's first destination and the number of passengers disembarking there. Columns 9-11 and 12-14 do the same for CD2 and CD3, but are not always defined, as many aTaxis only go to one or two destinations. In that case, the values in columns 9-11 or 12-14 are set to matlab's "realmax," which is just the maximum value matlab can hold. AVO, column 20, is calculated for each trip as the ratio of PersonTripMilesServed to ATaxiVehicleMiles.

Computational challenges and Della

By far the biggest challenge we faced during this analysis was the immense amount of data we were working with. As noted above, the input files alone amounted to 115gb. There was no way anyone in the group could fit that much data on a personal computer, so the first thought was to store it on two Bloomberg terminals in the Sherrerd basement. As it turned out, these machines did not even have enough space for all this data, so we were forced to split the data between two different terminals.

To process the data we used matlab, and as it turned out simply loading the input data into the matlab software proved to be quite a challenge. To begin, it took hours just to loop through all ~3500 trip files (one for each county). On top of that, about one in every one hundred files was at first unreadable by the computer for reasons unknown. Every time the computer encountered on of these files, the program would crash and we would have to start from scratch after fixing the unreadable file. Soon we figured out that it would be smarter to have the loop start where it left off, as opposed to at the beginning again. So if the program crashed on the 500th file, we would fix that file and then have the program start up again on the 500th file.

Once all the data was read into matlab as a single dataset, we realized that we simply did not have the processing power to analyze all of it for ride sharing potential in any reasonable amount of time. So we went to Michael Bino, the IT specialist for the ORFE department and had him set us up with a Della account. Using Della's massive parallel capabilities, we were able to analyze the smallest forty or so states very easily in a matter of hours. We ran into problems with some of the bigger states, for two reasons. First, Della automatically limits the amount of memory allocated per job, and some of the bigger states required more than this maximum amount. Michael Bino was immensely helpful once again, showing us the command we could use in order to tell Della to allocate more memory for each job. Additionally, we had set a cap on the number of aTaxis the program could use at 10 million, for the sake of computational efficiency. However, many of the bigger states required more than 10 million aTaxis to cover all of their person trips. In these cases, we were forced to opt for greater run time and raise this maximum value. At this point, Texas was still giving us problems so we divided it up into five pieces and ran each of them separately without issue. They were divided into five portions so that the information that was written out was still fairly consolidated and easy to append and analyze.

Below are a few screenshots of different steps along the way with Della.

Figure 8i: the uploaded Person Trip files in the Della account

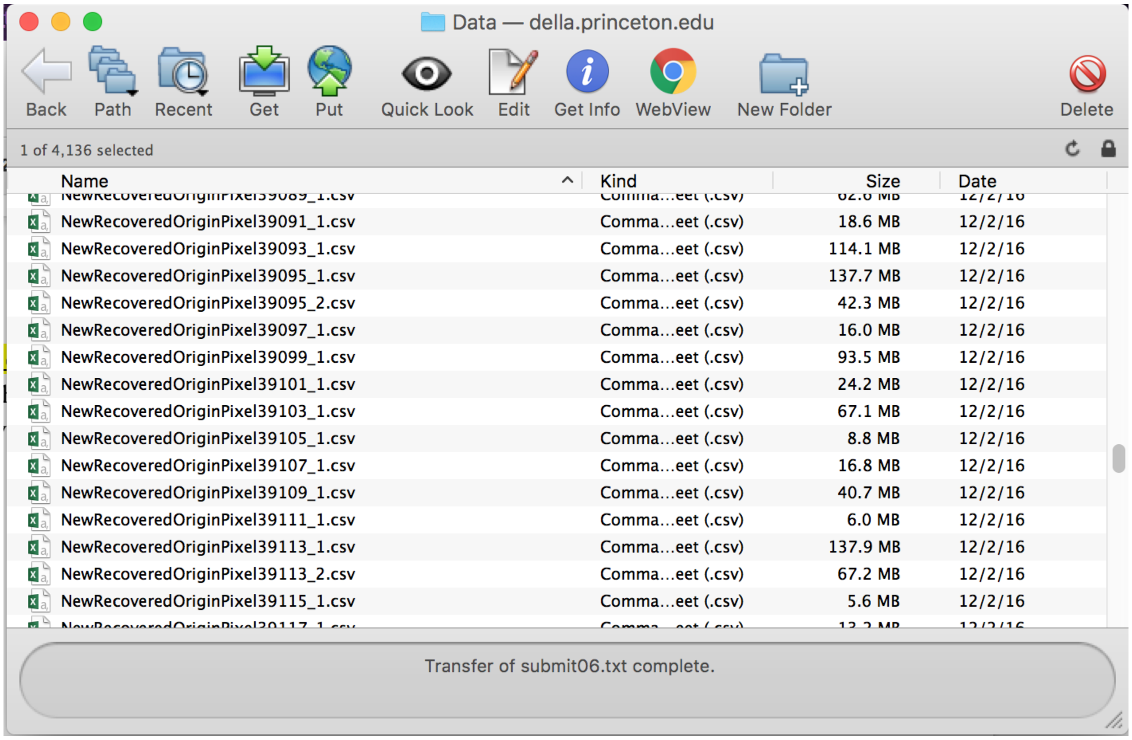
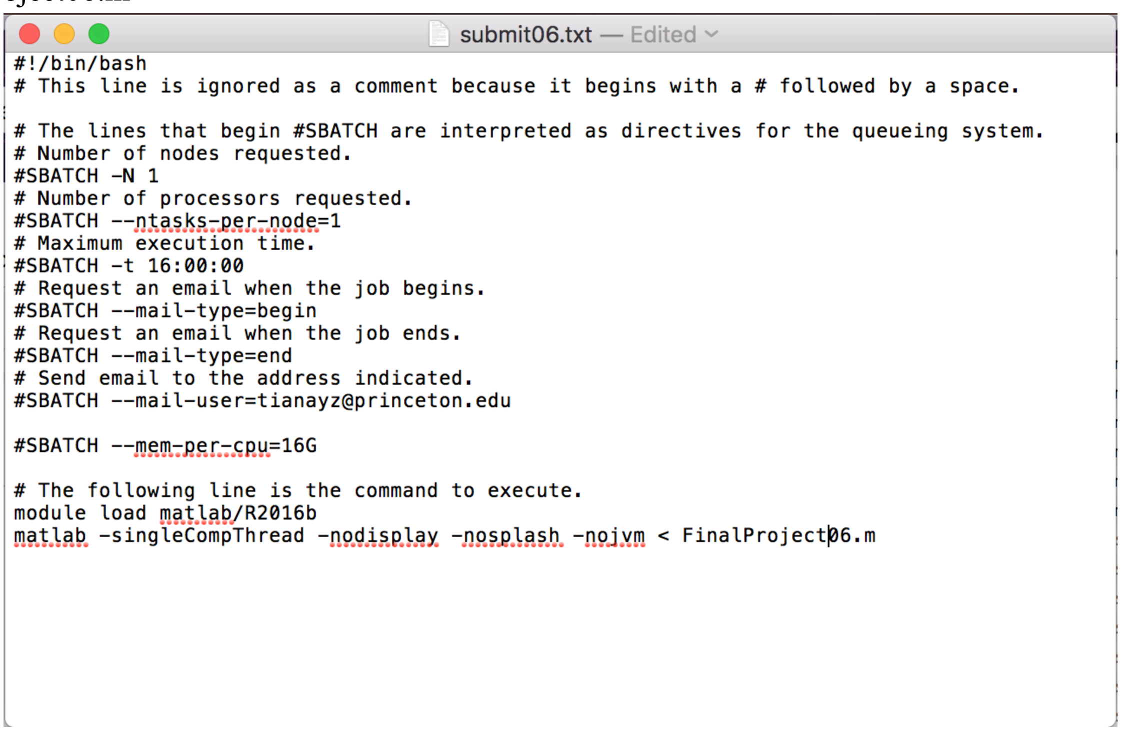


Figure 8j: a sample submission text that tells Della to run the Matlab code (in this case, “FinalProject06.m”


submit06.txt — Edited

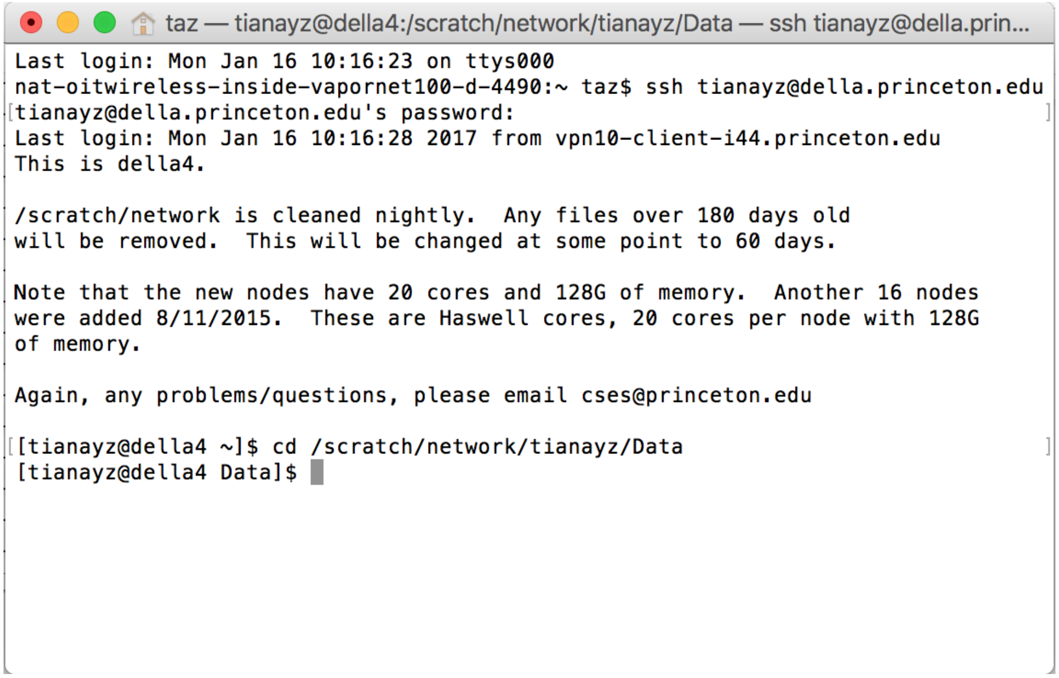
```
#!/bin/bash
# This line is ignored as a comment because it begins with a # followed by a space.

# The lines that begin #SBATCH are interpreted as directives for the queueing system.
# Number of nodes requested.
#SBATCH -N 1
# Number of processors requested.
#SBATCH --ntasks-per-node=1
# Maximum execution time.
#SBATCH -t 16:00:00
# Request an email when the job begins.
#SBATCH --mail-type=begin
# Request an email when the job ends.
#SBATCH --mail-type=end
# Send email to the address indicated.
#SBATCH --mail-user=tianayz@princeton.edu

#SBATCH --mem-per-cpu=16G

# The following line is the command to execute.
module load matlab/R2016b
matlab -singleCompThread -nodisplay -nosplash -nojvm < FinalProject06.m
```

Figure 8k: Accessing Della via the terminal



```
taz — tianayz@della4:/scratch/network/tianayz/Data — ssh tianayz@della.prin...
Last login: Mon Jan 16 10:16:23 on ttys000
nat-oitwireless-inside-vapornet100-d-4490:~ taz$ ssh tianayz@della.princeton.edu
tianayz@della.princeton.edu's password:
Last login: Mon Jan 16 10:16:28 2017 from vpn10-client-i44.princeton.edu
This is della4.

/scratch/network is cleaned nightly.  Any files over 180 days old
will be removed.  This will be changed at some point to 60 days.

Note that the new nodes have 20 cores and 128G of memory.  Another 16 nodes
were added 8/11/2015.  These are Haswell cores, 20 cores per node with 128G
of memory.

Again, any problems/questions, please email cses@princeton.edu

[tianayz@della4 ~]$ cd /scratch/network/tianayz/Data
[tianayz@della4 Data]$
```

Section 8.4 Analysis of aTaxi Vehicle Trip Files

After producing an aTaxi Vehicle Trip File for the entire nation, the next step was to analyze the reduction in trips and in trip miles. Figure 8l presents the number of Person Trips per state, the number of aTaxi Trips per state given the LOS, the number of Person Trip Miles per state, and the number of aTaxi Trip Miles per state with the LOS. It also displays ratios for the reduction in number of trips as well as for the reduction in miles. This second fraction, regTripsMiles/aTaxiMiles, is how we define AVO.

Note that this definition is much more strict than a simple calculation of the average number of people in an aTaxi. AVO as we define it is a ratio of the benefits of service over the cost (without taking repositioning into account). The numerator, regTripsMiles, is not how many miles each customer rode in the aTaxi. Rather, it is the straight-line distance between their origin and destination, i.e. the actual mobility that was provided to that rider. With this definition, AVO under the conventional car-ownership model would also be quite a bit lower. For example, consider a parent dropping off a child at school and then returning home. Under the simple average-number-of-passengers definition, AVO would appear to be 1.5. But how much mobility was actually provided? Only the child's place-time utility was improved by the trip, so AVO would actually be 0.5. It is important to keep this in mind when analyzing the AVO produced by the LOS.

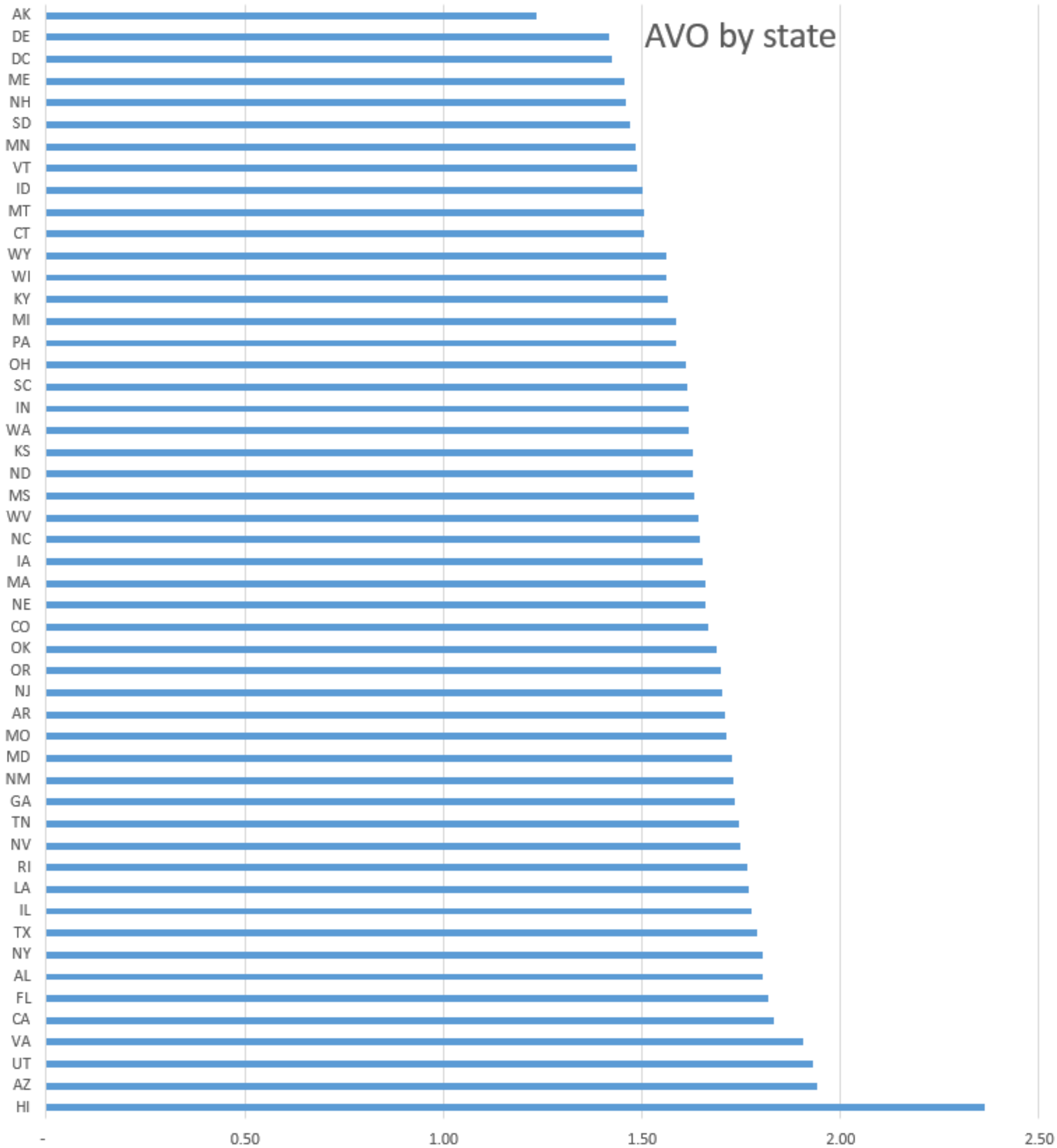
Figure 81: Impact of ridesharing on number of trips and trip miles (2 <GCD<200)

State	State Code	regTrips	aTaxi TripCount	regTrips/ aTaxiTripCount	regTrips Miles	aTaxi Miles	AVO (regTripsMiles / aTaxiMiles)
Alabama	AL	14,817,309	6,376,659	2.32	632,158,317	350,381,323	1.80
Alaska	AK	1,208,987	521,815	2.32	26,418,510	21,419,061	1.23
Arizona	AZ	12,927,641	4,841,858	2.67	305,147,214	157,255,392	1.94
Arkansas	AR	1,759,568	766,139	2.30	90,314,799	52,843,918	1.71
California	CA	75,274,095	25,365,328	2.97	1,242,516,857	678,132,130	1.83
Colorado	CO	8,706,925	3,498,241	2.49	126,435,035	75,837,941	1.67
Connecticut	CT	6,662,969	2,824,128	2.36	122,783,818	81,537,716	1.51
Delaware	DE	1,738,643	723,154	2.40	34,042,765	24,028,432	1.42
District of Columbia	DC	1,468,686	401,899	3.65	35,545,392	24,938,371	1.43
Florida	FL	32,998,488	12,969,351	2.54	523,231,835	287,869,332	1.82
Georgia	GA	19,976,818	8,756,810	2.28	442,483,195	255,476,928	1.73
Hawaii	HI	2,267,887	693,005	3.27	30,742,076	13,004,381	2.36
Idaho	ID	2,417,771	1,154,965	2.09	34,936,467	23,250,730	1.50
Illinois	IL	24,782,640	9,517,125	2.60	418,412,917	235,816,448	1.77
Indiana	IN	11,564,843	5,369,402	2.15	198,413,872	122,712,820	1.62
Iowa	IA	4,977,361	2,459,759	2.02	99,185,484	60,060,919	1.65
Kansas	KS	4,707,145	2,201,806	2.14	94,924,363	58,297,679	1.63
Kentucky	KY	8,278,433	3,882,177	2.13	145,335,090	92,779,380	1.57
Louisiana	LA	8,391,554	3,679,703	2.28	182,573,621	103,311,587	1.77
Maine	ME	2,546,828	1,382,888	1.84	46,785,614	32,169,170	1.45
Maryland	MD	10,380,493	4,010,447	2.59	127,987,692	74,163,606	1.73
Massachusetts	MA	11,358,238	4,668,683	2.43	156,491,863	94,287,883	1.66
Michigan	MI	18,323,460	8,635,735	2.12	310,166,620	195,531,912	1.59
Minnesota	MN	9,303,861	4,327,872	2.15	146,621,109	98,836,312	1.48
Mississippi	MS	5,767,286	2,872,499	2.01	136,004,338	83,408,320	1.63
Missouri	MO	10,713,995	4,977,922	2.15	202,358,872	118,209,384	1.71
Montana	MT	1,612,789	816,008	1.98	34,484,275	22,918,115	1.50
Nebraska	NE	2,903,605	1,321,666	2.20	44,873,601	27,022,196	1.66
Nevada	NV	4,924,667	1,787,517	2.76	52,185,662	29,838,669	1.75
New Hampshire	NH	2,490,348	1,278,855	1.95	41,026,938	28,108,055	1.46
New Jersey	NJ	14,073,301	5,552,053	2.53	179,902,504	105,708,554	1.70
New Mexico	NM	3,609,809	1,506,586	2.40	77,465,279	44,728,119	1.73
New York	NY	31,693,655	11,426,112	2.77	480,684,713	266,469,297	1.80
North Carolina	NC	18,600,152	8,725,553	2.13	290,286,662	176,474,069	1.64
North Dakota	ND	1,023,363	527,031	1.94	25,368,319	15,564,811	1.63
Ohio	OH	20,892,719	9,481,675	2.20	352,905,195	219,239,693	1.61
Oklahoma	OK	6,776,061	3,130,132	2.16	128,165,956	75,950,637	1.69
Oregon	OR	6,304,511	2,647,637	2.38	96,007,576	56,579,822	1.70
Pennsylvania	PA	22,407,244	10,000,119	2.24	343,308,031	216,421,579	1.59
Rhode Island	RI	1,671,057	632,978	2.64	16,060,307	9,096,156	1.77
South Carolina	SC	9,019,553	4,315,832	2.09	162,688,604	100,767,894	1.61
South Dakota	SD	1,276,809	664,371	1.92	22,871,003	15,539,624	1.47
Tennessee	TN	12,943,835	5,974,292	2.17	277,827,939	159,360,395	1.74
Texas	TX	43,164,348	18,805,385	2.30	821,456,690	459,059,985	1.79
Utah	UT	4,521,535	1,683,899	2.69	58,780,268	30,433,413	1.93
Vermont	VT	1,229,261	658,356	1.87	22,970,887	15,456,947	1.49
Virginia	VA	14,289,469	5,432,099	2.63	238,670,438	125,151,714	1.91
Washington	WA	11,952,819	4,976,597	2.40	173,231,921	107,043,683	1.62
West Virginia	WV	3,513,089	1,698,465	2.07	61,324,309	37,369,639	1.64
Wisconsin	WI	9,881,693	4,816,275	2.05	167,141,424	106,939,645	1.56
Wyoming	WY	784,251	393,036	2.00	16,502,851	10,568,042	1.56

A brief look at this table demonstrates that ridesharing at this LOS has great potential. First looking solely at the number of trips taken on a typical day, one can see that the number of aTaxi trips is significantly less than the number of Person Trips. California, for example, goes from about 75 million Person Trips to 25 million aTaxi Trips. New York goes from 32 million to 11 million. Even states where you might expect there to be little to no ridesharing see a significant drop in trips. Wyoming, which has only 784,000 Person Trips, can get by with just 393,000 aTaxi Trips. South Dakota goes from 1.28 million to 660,000. Clearly, there is a lot of potential for ridesharing to reduce the number of trips across the entire nation.

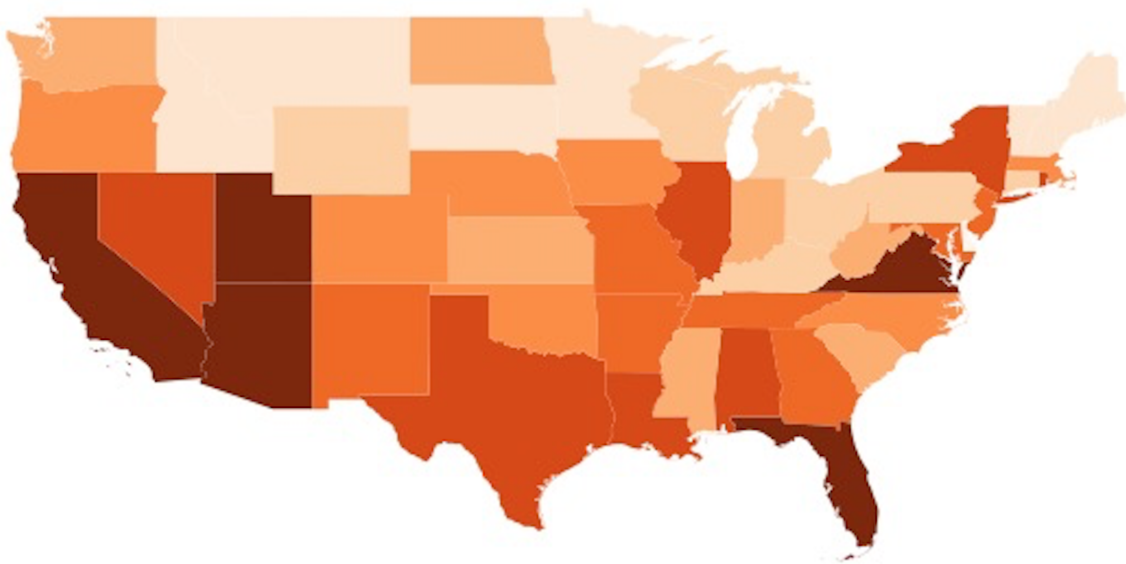
More important than the reduction in the number of trips, though, is AVO: the reduction from Person Trip Miles to aTaxi Vehicle Miles. The results in Figure 8f are encouraging while still being believable. Figure 8m presents AVO by state graphically.

Figure 8m: AVO by state



As Figure 8m shows, AVO is fairly consistent across states, with a few notable outliers. Hawaii's AVO is well above every other state, coming in at 2.36. Surely this has to do with the fact that all trips less than 200 miles leaving from Hawaii are staying in Hawaii, so there are simply fewer destinations to go to and thus more opportunities for ridesharing. It is also possible that trips from one island to another are being assigned to aTaxis by the program (not possible in real life), and thus Hawaii's AVO may be somewhat artificially inflated. Alaska is the other large outlier, with an AVO of only 1.23. This makes sense, as it is the least densely populated state in the US. Washington, D.C. is also surprisingly low, with AVO = 1.43. At first, this might seem wrong: surely ridesharing opportunities are greater in cities, so why would DC be third to last? But consider that the analysis only examines trips with GCDistance greater than 2. Many of the trips inside DC are probably walkable, with GCDistance less than 0.5, and even more can be taken care of by the short-haul aTaxi service, providing mobility for trips longer than half a mile but less than 2 miles. Many of the trips greater than 2 miles then, are probably people returning home from work, which would mean they are fanning out in all directions from the district, reducing ridesharing opportunities.

Figure 8n: AVO by state



The nationwide figures are as follows. Nationwide Person Trips, as mentioned before, sum to 564,881,867. With ridesharing at the given LOS, the number of aTaxi trips is 235,131,867. That represents a reduction by a factor of 2.4. Nationwide Person Trip Miles sum to 10,098,209,086, while total nationwide aTaxi Vehicle Miles are 5,877,371,824. So if riders were willing to wait a maximum of 5 minutes for their aTaxi to depart and tolerate at most 2 additional stops with a max circuity of 25%, we would see a reduction from just over 10 billion Person Trip Miles to under 6 billion aTaxi Vehicle Miles.

The ratio of those two figures gives a nationwide AVO of 1.72 (recall that is under the strict definition of AVO). Further groups will analyze the implications of fleet size and empty vehicle repositioning strategies on this number. As a baseline, though, this is an encouraging result that demonstrates the capacity of a nationwide aTaxi fleet to enable ridesharing, reduce Vehicle Miles Traveled, and more efficiently use the roads and vehicles.

Acknowledgments

Thanks to Michael Bino in the ORFE department for his help in getting us access to Della, explaining how the system works, and resolving issues we encountered with Della.

Thanks to Kyle Marocchini '18 for generating the Person Trip files we used as input to our aTaxi Vehicle Trip file generation.

And finally thanks to Professor Kornhauser for the whole sem

Chapter 9:
Single Sized aTaxi Fleet

This chapter's analysis involves looking at nationwide five-passenger aTaxi trips between two and two hundred miles. It analyzes supply and demand for nationwide trips and tried to best serve the need with aTaxis having a finite seating capacity of five passengers. It discusses different repositioning strategies to find the optimal method, which creates the most efficient fleet operation. Attached to the submission is all of the excel and MATLAB files used to generate these results.

Sam Button and Antigone Valen

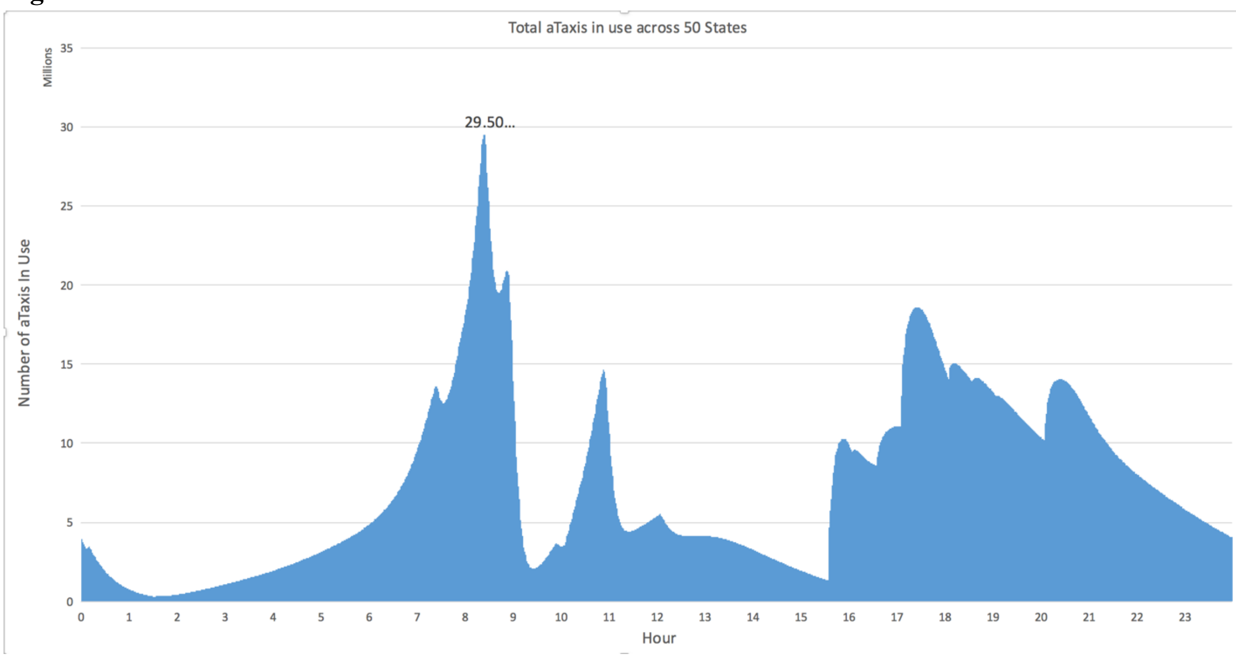
1. aTaxis in Motion

The goal was to find the necessary fleet size to satisfy nationwide demand for trips between two and two-hundred miles.

First the vehicleTrip files generated by Bill Van Cleve and Tianay Zeigler were read in using the aTaxicounter.m file. The code initializes a function called 'countThoseTaxis()' that then counts the number of taxis leaving and arriving at each pixel at any given minute of the day. At any given minute if a taxi leaves a pixel, it is added to the number of aTaxis in motion. If it arrives at a pixel then it is taken out of the number of moving aTaxis. This shows how many aTaxis are in motion per minute.

Below is the graph of aggregate aTaxi motion across the nation.

Figure 9a. Total aTaxis in use Across the 50 States



The peak here is at the same time, around 8:30am in the morning with 29,520,517 aTaxis in motion. There are other peaks throughout the day a little before 11:00am, before 5:30pm, and around 8:15pm. There's are national low points at about 9:30am and around 3:30pm, right before trips shoot back up again. aTaxis in motion trails off after 8:30pm.

For 5-passenger capacity aTaxis, the trip data gives the following.

Table 9b. Minimum Fleet Size Required and AVO for Nationwide Trips

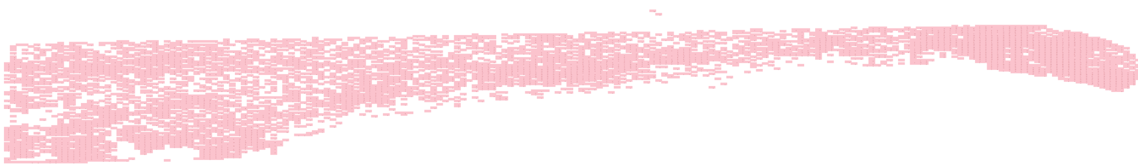
Minimum Fleet Size	29,520,517
Average Vehicle Occupancy (AVO)	1.63

2. aTaxi Supply and Demand

In addition to keeping track of the aTaxi motion throughout the day, the algorithm developed a snapshot of the total aTaxi demand for each pixel of each state. This was the method to establish a minimum fleet size, which would allow for no taxi repositioning during the day.

For example, a summary of the total aTaxi demand for the state of Delaware is shown below.

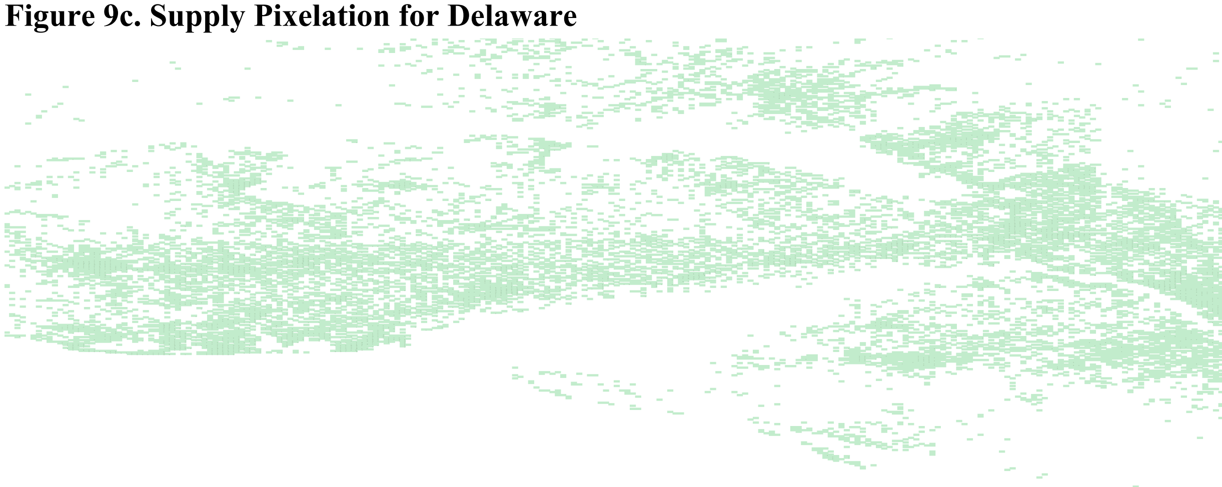
Figure 9c. Demand Pixelation for Delaware



Every colored pixel in the above array represents a pixel that requires a certain number of taxis to be located there at the beginning of the day. For the state of Delaware, the total number of taxis required (the maximum fleet size) is 216,105.

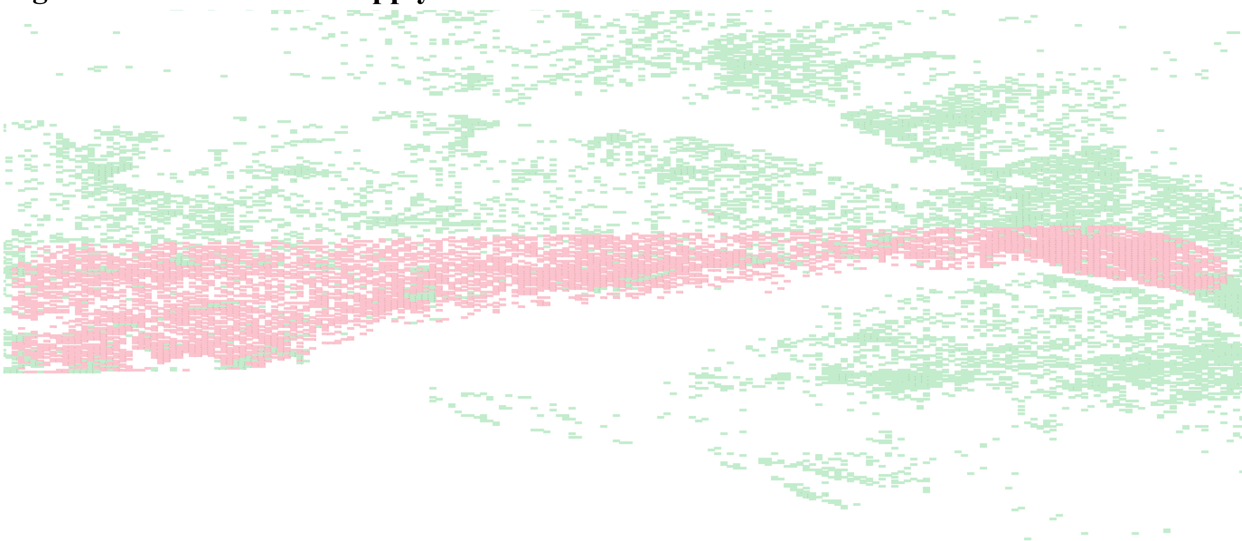
The algorithm also developed a picture of the locations of all aTaxis (supply) at the end of the day. This was useful because our initial attempt at aTaxi management consisted of a one-time overnight mass reposition for each state. This approach proved to be inefficient, partly because of the obstacles that we ran into dealing with the data from each state. Below is an example of the supply of aTaxis at the end of the day, for Delaware.

Figure 9c. Supply Pixelation for Delaware



As one can see, the supply of aTaxis at the end of the 24 hour period is scattered across a greater area than the demand covered. This scattering happens because the destinations of the aTaxi trips were not restricted to the boundaries of Delaware. Therefore, the overnight mass reposition would consist of pixels with a surplus of taxis moving vehicles to pixels with a deficit. The difference between aTaxi supply and demand for Delaware trips is shown below.

Figure 9d. Difference in Supply and Demand Pixelation for Delaware



The repositioning method called for the taxis located at green pixels to move to red pixels, setting up the next day of aTaxi use at midnight. We decided that this method of aTaxi management did not fit our needs, mainly because of our inability to deal with more than one state at a time.

In the example above, one was only able to analyze the taxi movement of Delaware trips for the 24-hour period. This proves, however, to be inadequate, because it is clear to us that aTaxi supply from trips from New Jersey is relevant to aTaxi supply in Delaware at midnight.

An improvement that could be made to this analysis in the future would be to enable the code to capture the dynamics of interstate taxi movement, in an array that is populated across a single 24-hour period, but over 50 states. This would allow to more realistically consider a mass overnight repositioning system.

After running the same algorithm for all the nation's trip files it was able to generate the maximum number of aTaxis needed to serve the whole country's demand. A table is shown below of each state's demand, totaling to 73,500,704 for the whole country.

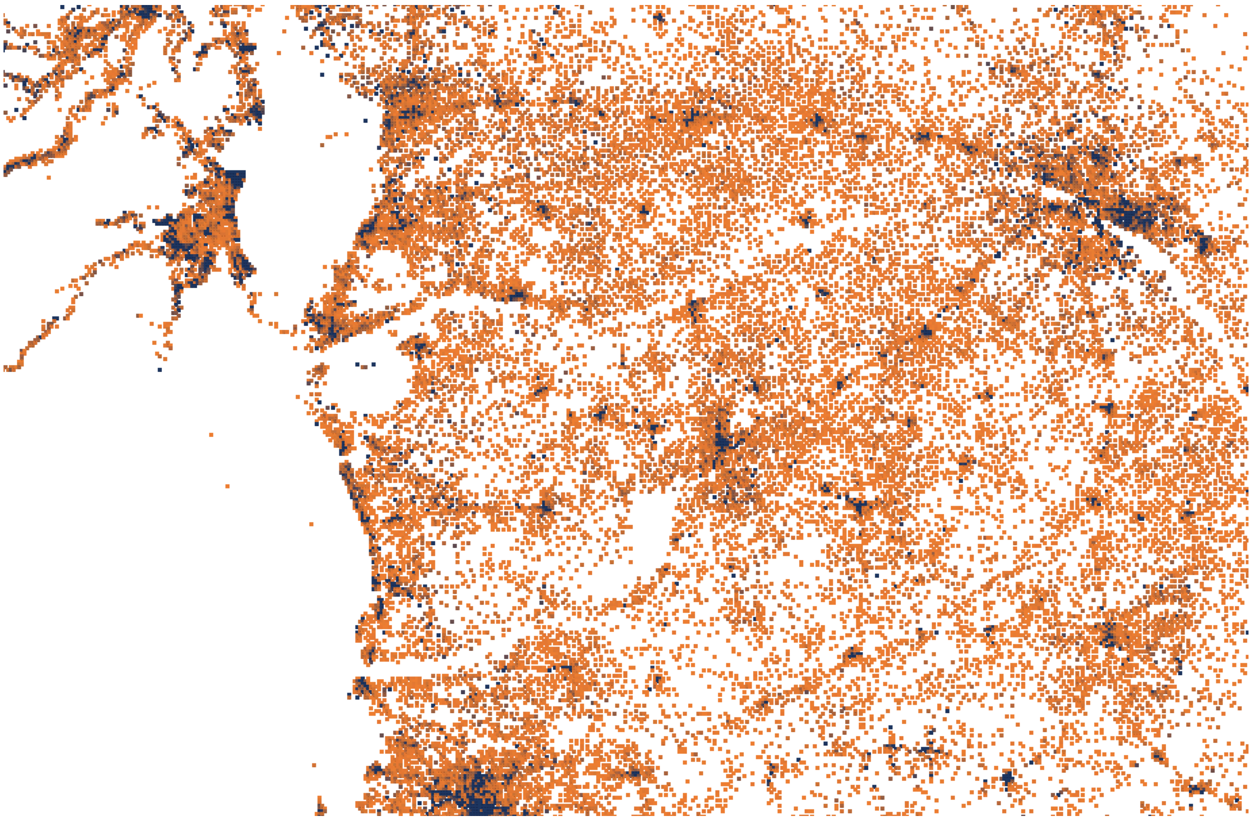
Table 9e. aTaxi Demnd by State

State	aTaxi Demand
California	8,962,468
Texas	6,226,425
Florida	4,594,885
New York	4,560,916
Illinois	2,975,112
Pennsylvania	2,953,546
Ohio	2,677,980
Georgia	2,332,250
North Carolina	2,296,823
Michigan	2,288,949
New Jersey	2,064,509
Virginia	1,923,178
Washington	1,631,048
Massachusetts	1,558,031
Arizona	1,554,815
Indiana	1,523,719
Tennessee	1,512,747
Missouri	1,400,547
Maryland	1,380,410
Wisconsin	1,329,862
Minnesota	1,260,479
Colorado	1,237,077
Alabama	1,120,092
South Carolina	1,116,190
Louisiana	1,073,966
Kentucky	1,019,405
Oregon	917,032
Oklahoma	895,739
Connecticut	830,748
Iowa	717,673
Arkansas	691,562
Mississippi	689,448
Utah	679,741
Kansas	670,761
Nevada	655,765
New Mexico	481,718
Nebraska	434,583
West Virginia	427,382
Idaho	377,523
Hawaii	327,885
Maine	307,219
New Hampshire	306,463
Rhode Island	243,720
Montana	236,423
Delaware	216,105
South Dakota	197,063
North Dakota	170,803
Alaska	170,399
Vermont	144,594
Wyoming	134,926
Total	**73,500,704**

Attached to the report is a zip file containing files showing the pixelations for supply, demand, and difference between the two for nationwide trips. The files have color gradients, showing locations of high demand.

Below is a snippet of the demand pixelation for the whole country. The full picture can be seen in the attached "DemandFinal" file.

Figure 9f. Demand Pixelation for Nation

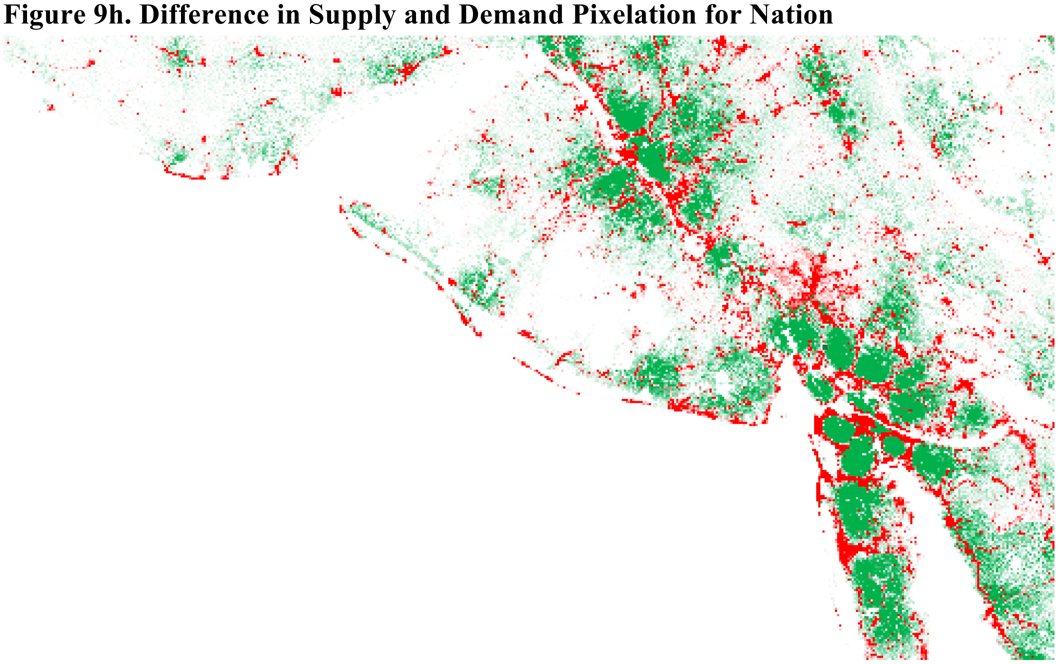


Below is a snippet of the supply pixelation for the whole country. The full picture can be seen in the attached "SupplyFinal" file.

Figure 9g. Supply Pixelation for Nation

Below is a snippet of the difference pixelation for the whole country between supply and demand. The full picture can be seen in the attached "DifferenceFinal" file.

Figure 9h. Difference in Supply and Demand Pixelation for Nation



3. Vehicle Repositioning

There are two different strategies for vehicle repositioning: the system-wide reposition at one time during the day or the individual repositioning of unoccupied aTaxis throughout the day.

System-Wide Repositioning

The system-wide repositioning of taxis at one time during the day seeks to fill the expected demand at each pixel. First, it's important to find the hour of day where aTaxis have the least expected demand. For many states this is around midnight. This is the same for the aggregated national demand.

To perform this repositioning strategy, one can model it as a network flow problem. After choosing an hour for repositioning (where demand is the lowest), we must choose an hour of expected demand we want to complete the repositioning for (where demand starts to pick up, say, at the beginning of the day before school and work). Say our state repositions at midnight each night. We may want our repositioning then to fill our expected demand at 8:00am, where demand starts to pick up. For each pixel, we subtract the number of aTaxis from the pixel at the repositioning hour from the expected demand at the hour we want to satisfy demand. If positive, this gives us the number of aTaxis we need to have repositioned to that pixel. If negative, we have excess aTaxis and can reposition those elsewhere. If zero, our expected demand for that pixel is satisfied and no repositioning occurs.
This gives supply and demand nodes. The supply nodes are the pixels with negative values, i.e. excess vehicles. The demand nodes are the ones that need repositioning to satisfy the morning's demand. Each supply node is connected to every demand node to determine where the supply node should reposition its aTaxis amongst the demand nodes. Each arc then has a cost, which is 1.2 times the Cartesian distance from supply to demand and we minimize so that supply nodes are repositioning to the closest demand nodes.

Hourly Repositioning

Similar to the prior strategy but repositions at every hour of the day. The goal of this strategy is to deal with imbalance throughout the day instead of it adding up at the end of the day as in the system-wide one time repostioning strategy. At each hour, we calculate the expected demand imbalance for each pixel (as in the system-wide repositioning). This gives us a higher percentage of aTaxis in motion at any given time, a higher number of trips per aTaxi per day, and a smaller fleet size.

Individual Repositioning

Another potential strategy is that of individual repositioning throughout the day. In this case, unoccupied aTaxis reposition themselves if their drop off location has low expected demand. One way to calculate this is to assign a value to each pixel, computing the number of expected arrivals in an hour minus the existing number of aTaxis currently there. We then subtract the distance between pixels so as to again minimize cost and travel time so that aTaxis are making smart repositioning decisions.

This strategy gives a much lower recommended fleet size because aTaxis can be reused many more times throughout the day.

Demand-centric Close Radius Pickup

Rather than generating a new aTaxi when demand is not being serviced, this strategy searches for available taxis within a close radius (say, within 30 closest pixels). This is demand-centric in that it looks at individual demand instead of looking at the system as a whole. This will reduce the fleet size but increase passenger wait time.

4. Conclusion

Our results show that the nationwide aTaxi fleet size must be somewhere between:

Table 9i. Minimum and Maximum aTaxi Fleet Size for the Nation

Minimum Fleet Size	**29,520,517**
Maximum Fleet Size	**73,500,704**

Chapter 10:
Optimal Vehicle Capacity

The goal of this chapter is to find the single best vehicle capacity for nationwide ride sharing. By modifying the vehicle trip file for infinite capacity (created by Van Cleve and Ziegler in Chapter 8) we study the effects of enforcing a finite capacity on ride sharing. Specifically, we look at total number of trips, average vehicle occupancy (AVO), PersonTripMiles, VehicleTripMiles, and Fleet Size for finite capacities of 3, 4, 5, 15, and 60. With these calculations, we are able to estimate capital and energy operating costs for each type of fleet, which we then use to suggest an "optimal" vehicle capacity for nationwide ride sharing. Finally, we will isolate individual states in order to answer whether or not a single optimal vehicle capacity exists, or if, rather, different regions require different capacities.

Ify Ikpeazu and William Haynes

Method:

We started with the 51 vehicleTripFiles created by Bill Van Cleve and Tianay Ziegler in chapter 8, one for each state plus one for Washington, D.C. Together, these 51 files contain every daily vehicle trip in the nation. The algorithm we perform is as follows:

Algorithm 10.1

```
1    For each vehicleTripFile
2        For each capacity (3, 4, 5, 15, 60)
3            Total Trips Required = ceiling(Riders at Departure / capacity)
4            New PersonTripMiles = PersonTripMiles
5            New VehicleTripMiles = VehicleTripMiles * Total Trips Required
6            New AVO = New PersonTripMiles / New VehicleTripMiles
```

Notice the statements on Lines 3 and 4 of Algorithm 10.1. All of the calculations in the analysis are based on the logic we use here. Enforcing a vehicle capacity does not affect the time at which people travel nor does it change their starting and ending points. Therefore, we can think of enforcing a vehicle capacity as requiring more vehicle to complete the same trips. In the infinite capacity model, 10 (or more) people could share one vehicle. If we enforce a capacity of 3, we assume that the same people will make the same journeys as in the other case, but will have to do so in ceiling(10/3) = 4 aTaxis. As a result, PersonTripMiles do not change.

Our model has advantages and disadvantages. On one hand, it is simple to understand, easy to implement, and fast to run. Unfortunately, though, it does not handle trips with multiple destinations quite correctly. If 2 groups of 3 people with nearby destinations (under max circuity limit, see chapter 8) arrive at an aTaxi stand at the same time, infinite capacity ride sharing analysis would put them in the same vehicle. This vehicle visits both destinations. Our model, in the capacity of 3 situation, puts them into 2 taxis, but still forces them to go to both destinations. In a real-world scenario, aTaxis would likely departure as soon as they were at capacity. There is a chance that the aTaxis would visit both locations, but it would depend on the order of arrival. If the 3 passengers headed for destination A all arrived before the three headed for destination B, though, the two aTaxis would only visit one location. Our model makes the taxis visit both (or all 3) locations every time. As a result, our model is slightly inaccurate for vehicle trips that are a) bound for multiple locations and b) resulted from an arrival order like the one I mention above, when a real aTaxi would depart upon being filled. For us to have eliminated this inaccuracy, we would have had to run the entire ride sharing analysis from the start, starting with personTrip Files and departing aTaxis when they fill up. Because the inaccuracy occurs in such a specific scenario, we believe it affects very few of the vehicleTrips, so its effect on our results is minute. Still, its effect would be to make both personTripMiles and vehicleTripMiles to be larger than they should be. Since the situation only occurs when there is ride sharing, this would increase personTripMiles by more than vehicleTripMiles, so our AVOs may also be larger than they should be. As we will explain later, the existence of this slight inaccuracy does not affect any of our conclusions.

Performing algorithm 10.1 for each of the 5 capacites gives us 5 sets of new values for Total Trips, AVO, PersonTripMiles, and VehicleTripMiles. For each vehicleTripFile we summed up these values and recorded our state totals (for statewide AVO we recorded the sum of PersonTripMiles divided by the sum of VehicleTripMiles).

Fleet Size Calculation

To calculate statewide fleet size for the different capacities, we find the maximum demand at any point in the day for each pixel, then sum up across the pixels in a state. We use Algorithm 10.2 (below), modeled after Hana Ku and Shirley Zhu's Minimum Fleet Sizing Method in Chapter 1 of the Orf467 Fall 2015 Final Report.

Algorithm 10.2

```
For each vehicleTripFile
    For each capacity (3, 4, 5, 15, 60)
        For Each Pixel
            Extract all departures and arrivals at that pixel
            Set Current Demand for Pixel to 0.
            Set Max Demand equal to 0
            Sort departures and arrivals by time
            For each departure or arrival:
                If arrival
                    Increment current demand by Total Trips
                If departure
                    Decrement current demand by Total Trips
                If current demand < max demand
                    Max demand = current demand
            If Max Demand < 0
                Demand at pixel = -1 * Max demand
            If Max Demand >= 0
                Demand at pixel = 0
        Fleet size equals sum of demand across all pixels in file
```

Results

After the completion of algorithms 10.1 and 10.2, we have Total Trips, AVO, PersonTripMiles, VehicleTrip Miles, and Fleet size for each state for the 5 different capacities. Tables 10.1 (next page) shows the statewide statistics for capacity 3. Tables 10.2, 10.3, 10.4, and 10.5 are attached at the end of the chapter and contain these statistics for capacities 4, 5, 15, and 60, respectively.

Table 10.1: Statewide Ridesharing Statistics for Capacity 3

State	Total Trips	AVO	PersonTripMiles	VehicleTripMiles	Fleet Size
Alabama	7,155,363	1.58	626,102,847	396,921,645	2,795,115
Alaska	614,289	1.44	37,459,606	26,081,969	215,087
Arizona	5,710,540	1.53	300,590,679	196,134,270	1,871,345
Arkansas	881,635	1.55	87,929,637	56,802,447	525,836
California	31,816,094	1.43	1,233,534,989	865,331,983	8,804,063
Colorado	4,154,439	1.38	129,052,522	93,668,051	1,341,672
Connecticut	3,225,871	1.37	128,603,686	94,067,596	1,004,031
Delaware	840,565	1.29	35,519,946	27,469,410	303,530
DC	657,632	1.00	37,900,605	37,900,539	478,809
Florida	15,178,126	1.44	504,623,994	350,185,555	4,390,710
Georgia	10,079,260	1.39	437,590,982	313,835,747	3,243,064
Hawaii	991,642	1.47	28,057,815	19,061,913	316,446
Idaho	1,306,638	1.34	36,962,916	27,571,601	410,977
Illinois	10,526,191	1.44	405,915,283	282,443,290	3,094,260
Indiana	6,075,335	1.38	204,655,921	148,162,102	1,928,464
Iowa	2,781,736	1.40	103,767,050	74,341,188	884,822
Kansas	2,492,303	1.37	95,990,327	69,970,314	795,280
Kentucky	4,422,864	1.33	147,700,006	110,699,773	1,439,661
Louisiana	4,265,494	1.37	175,410,171	128,104,456	1,354,135
Maine	1,497,885	1.35	49,617,670	36,738,582	508,168
Maryland	4,839,363	1.39	129,738,975	93,336,327	1,591,494
Massachusetts	5,459,849	1.41	162,957,686	115,911,672	1,586,157
Michigan	9,556,643	1.42	327,209,044	231,096,864	2,882,778
Minnesota	4,913,386	1.34	157,727,822	117,731,968	1,574,937
Mississippi	3,211,157	1.36	134,072,982	98,874,981	1,093,041
Missouri	5,633,229	1.41	203,029,143	144,214,013	1,846,176
Montana	917,633	1.37	37,164,882	27,225,400	289,621
Nebraska	1,533,297	1.37	46,165,634	33,709,592	484,823
Nevada	2,184,011	1.41	52,483,186	37,148,045	640,025
New hampshire	1,391,569	1.35	42,574,082	31,648,660	459,930
New jersey	6,540,073	1.41	184,621,675	130,486,453	2,017,123
New mexico	1,779,907	1.37	77,450,670	56,547,610	558,592
New york	14,675,339	1.38	495,006,193	358,475,782	4,255,741
North carolina	9,796,645	1.38	289,062,419	209,490,790	3,124,203
North dakota	595,674	1.43	28,353,985	19,828,249	192,587
Ohio	10,725,240	1.39	363,672,072	261,430,632	3,308,958
Oklahoma	3,556,161	1.38	127,394,207	92,283,076	1,140,702
Oregon	3,109,642	1.41	100,915,495	71,381,058	923,164
Pennsylvania	11,424,049	1.37	354,153,462	258,938,292	3,497,440
Rhode island	771,892	1.42	16,337,099	11,525,173	259,349
South carolina	4,807,026	1.39	162,752,948	117,354,776	1,548,082
South dakota	743,137	1.31	23,930,795	18,239,371	235,878
Tennessee	6,786,150	1.41	279,847,280	199,025,886	2,165,046
Texas	21,215,137	1.45	794,293,785	547,460,984	8,995,955
Utah	2,071,668	1.46	59,389,102	40,584,521	608,531
Vermont	719,989	1.35	23,933,950	17,711,639	240,214
Virginia	6,841,929	1.43	237,622,033	166,365,393	2,326,317
Washington	5,781,970	1.42	186,484,344	131,629,947	1,715,265
West virginia	1,917,670	1.36	61,887,765	45,522,992	692,939
Wisconsin	5,381,784	1.37	174,515,750	127,015,194	1,683,953
Wyoming	446,769	1.33	16,995,216	12,749,650	142,595

Analysis

Summing up the results of the previous section gives the nationwide ridesharing statistics for the 5 capacities. These are displayed in Table 10.6 below:

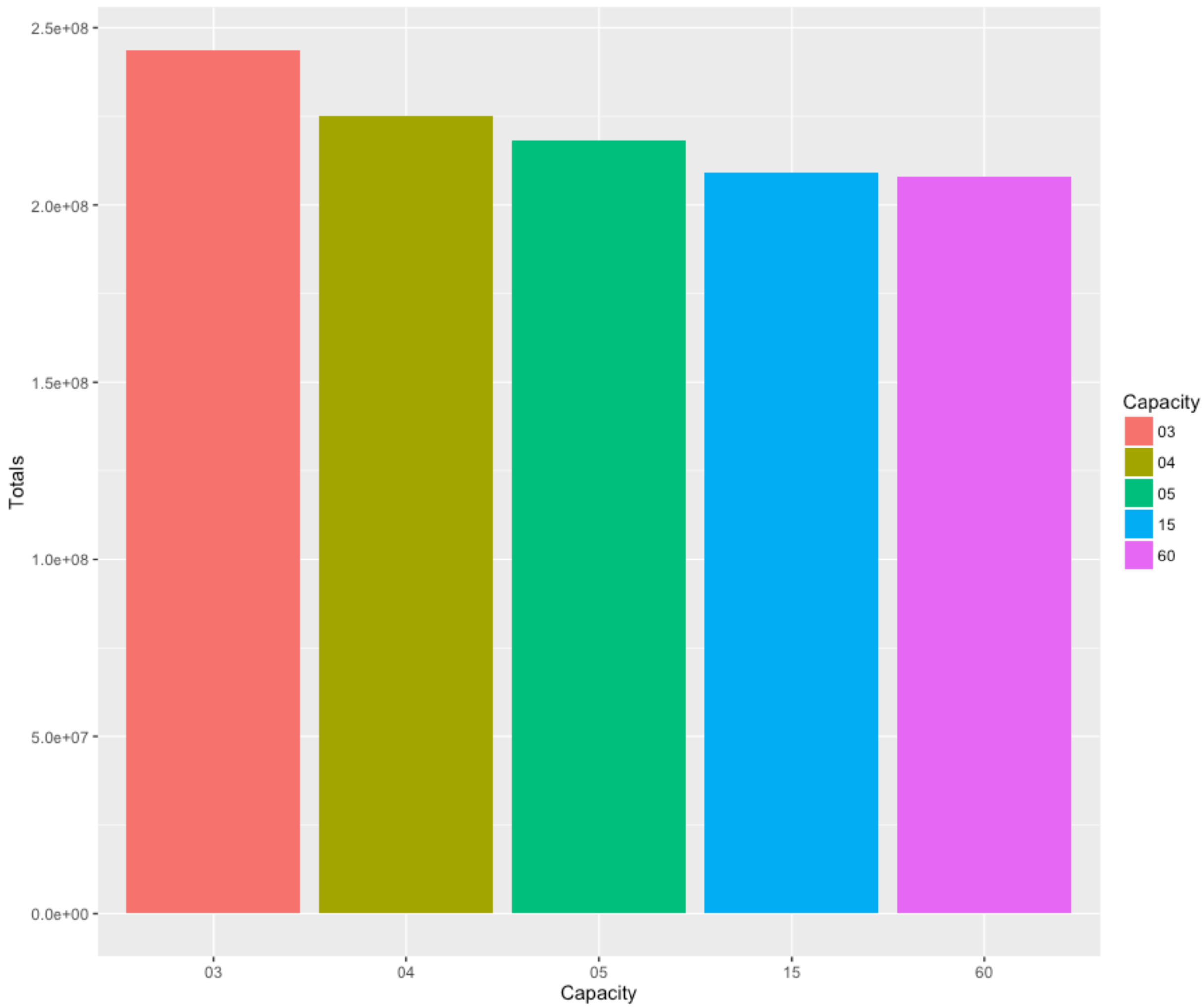
Table 10.6: Nationwide Ridesharing Statistics

Capacity	Trips	AVO	PersonMiles	VehicleMiles	Fleet Size
3	274,001,890	1.41	10,158,730,343	7,180,437,418	87,787,091
4	253,521,378	1.56	10,158,730,343	6,513,264,346	80,290,441
5	245,921,322	1.62	10,158,730,343	6,261,043,123	75,410,333
15	235,527,205	1.72	10,158,730,343	5,904,686,695	72,884,021
60	234,430,147	1.73	10,158,730,343	5,864,232,185	72,466,036

As we increase capacity, AVO increases while Total Trips, VehicleMiles, and Fleet Size decrease. PersonMiles remains the same due to our way of modeling the different capacities. These results are as expected. Obviously, it will take more vehicles of smaller capacity to move the same amount of people around. More vehicles performing the same trips means more vehicle miles and more trips. Fewer people in a vehicle means a smaller AVO. In order to suggest an optimal capacity, we need to know just how much these statistics change across the capacities and how these changes affect operating and capital costs. We'll start by looking at each statistic individually. For each we show a bar chart portraying a simple comparison between capacities. In addition, we include 3-by-3 matrices that help whose entries give the ratios of these statistics for capacities 5, 15, and 60. The purpose of these matrices is to show the degree of change in our statistics when we moving from capacity 5 to 15, 5 to 60, 15 to 60, etc. Each entry M_{ij} corresponds to a capacity i and a capacity j. The ratio shown in the entry is j's value of the particular statistic divided by i's value of the statistic. Thus, for our matrices, $M_{ij} = 1/M_{ij}$ and $M_{ij} = 1$ when i = j. We first focus on Total Trips.

Total Trips

Figure 10.1: Total Trips at Different Capacities



The number of trips decreases as the capacity of a vehicle increases. Naturally, a smaller vehicle will need to take more trips than a large vehicle to accommodate the same amount of passengers. In this case, the ratio of the five passenger car to the 60 passenger car is 1.05. This ratio shows the decreasing difference between the number of total trips as the capacity approaches 60 and approximates infinite capacity. Table 10.7 is the matrix of ratios for Total Trips.

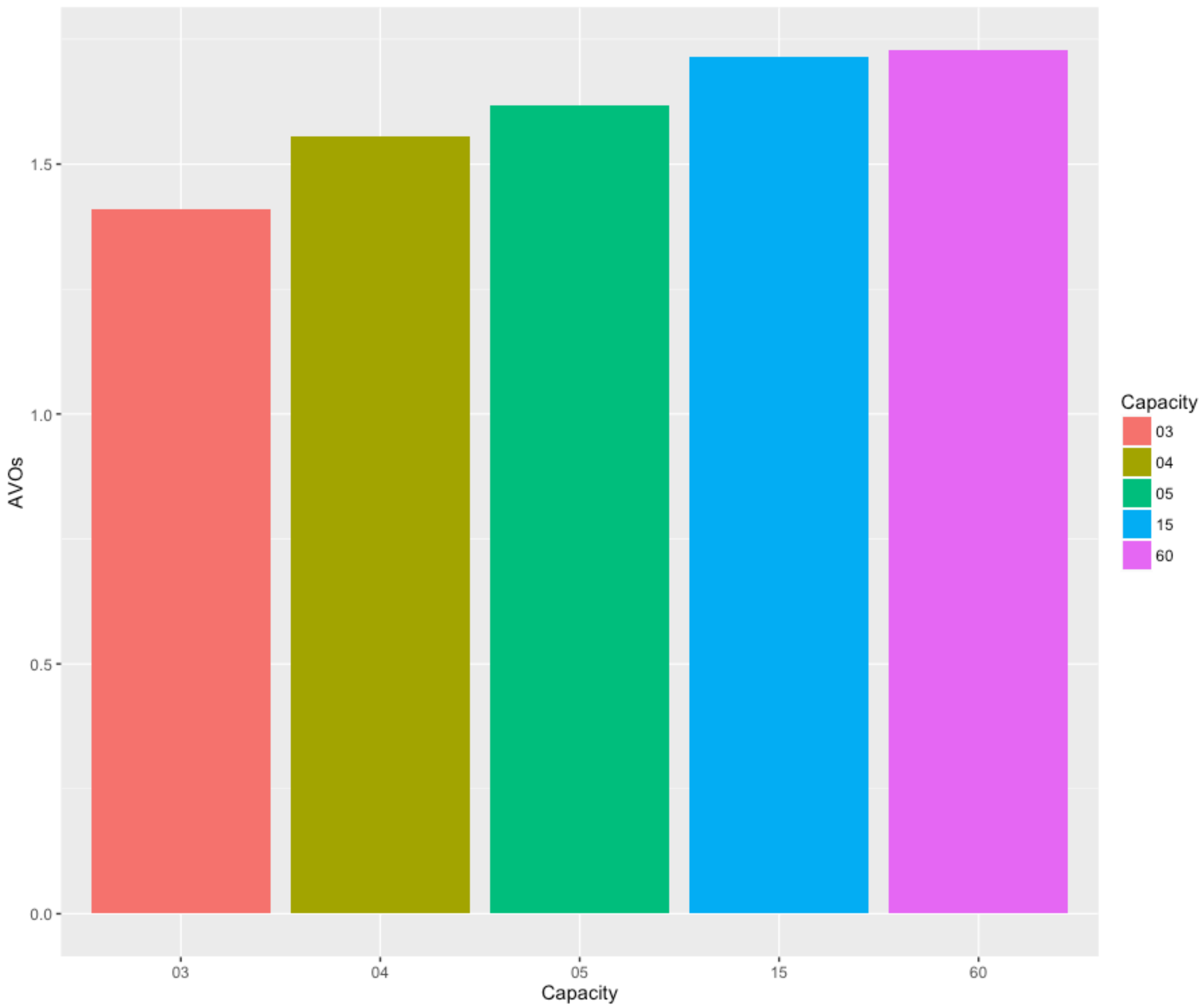
Table 10.7: Total Trips Ratios

	5	15	60
5	1.00	1.04	1.05
15	0.96	1.00	1.01
60	0.95	0.99	1.00

As we can see from the table, moving between 15 passenger vehicles and 60 passenger vehicles has almost no effect on total number of trips, indicating that hardly any trips would carry more than 15 passengers. Even moving from 5 passenger vehicles to 60 passenger vehicles only results in a 5% decrease. Almost that entire decrease comes as the vehicle is increased from 5 to 15 passengers.

AVO

Figure 10.2: AVO at Different Capacities



The average vehicle occupancy increases as the passenger size increases. Because a large capacity vehicle can receive more passengers than a small capacity vehicle, the average number of passengers across all trips is higher. The AVO ratio of a five passenger car to the 60 passenger car is .94. The ratio represents the decreasingly small difference in AVO for larger vehicles and also indicates that a larger vehicle, like the 60-sized, does not take many more passengers than the smaller 5 passenger vehicle, which fits 12 times as few passengers. We can infer that there would be many empty seats in a 60 person vehicle scenario. Table 10.8 contains the ratios between 5, 15, and 60.

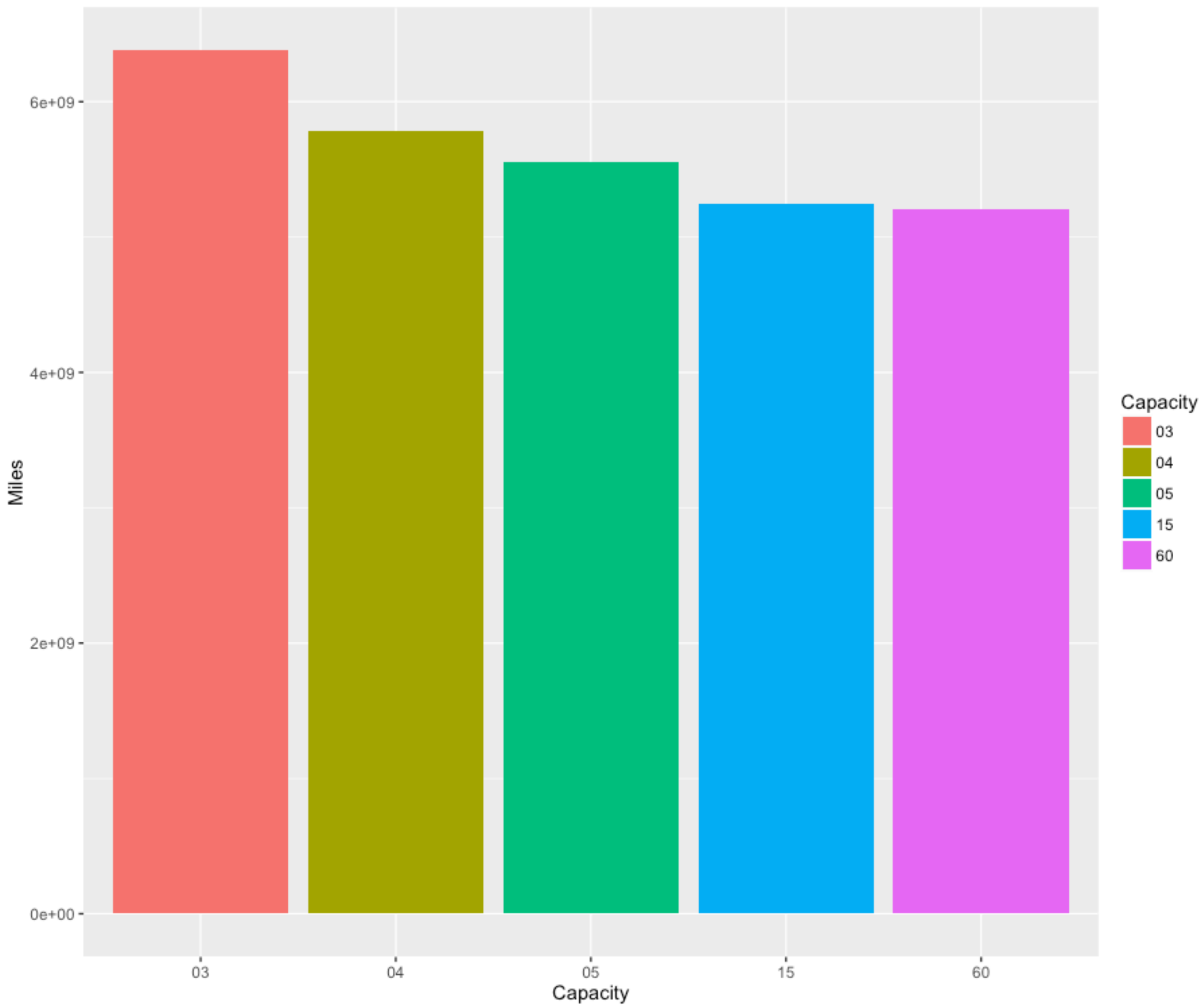
Table 10.8: Total Trips Ratios

	5	15	60
5	1.00	0.94	0.94
15	1.06	1.00	0.99
60	1.07	1.01	1.00

Table 10.8 shows that increases vehicle capacity has a slightly larger effect on AVO than it does on Total Trips. For AVO, there is almost a 7% increase when moving from capacity 5 to capacity 60. Again, the statistics for capacity 15 and capacity are very similar, almost identical. As we explain in the introduction, any extra increase in AVO could be the result of our own method of enforcing a capacity. In a real world scenario, AVO may not be quite as larger for larger vehicles.

Vehicle Miles

Figure 10.3: VehicleMiles at Different Capacities



Vehicle miles decreases as capacity size increases. A small capacity vehicle will travel more total trips, which increases the number of vehicle miles. The 60/5 VehicleMiles ratio is 1.05. Although this represents a rather decrease from 5 passenger to 60 passenger vehicles, the difference between the three and five passenger vehicles is sizeable; the ratio between those sized vehicles is 1.15 and three passenger vehicles travel 919394295 more miles than five passenger ones. We see from the bar chart that the VehicleMiles varies much more between small size vehicles than between large ones again due to smaller cars requiring more vehicles and also the average number of people departing from each pixel position. We can further conclude that most ridesharing occurs with 4 or 5 people in a vehicle, not 15 and above. Table 10.9 shows the VehicleMiles ratios between 5, 15, and 60 passenger vehicles.

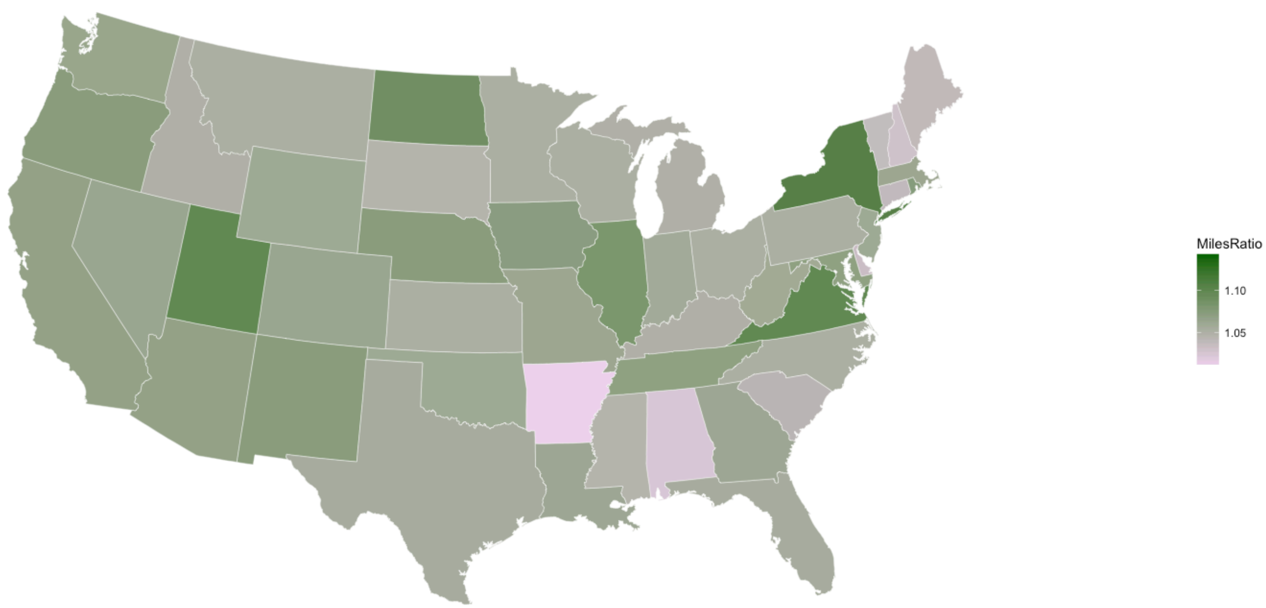
Table 10.9: Vehicle Miles Ratios

	5	15	60
5	1.00	1.06	1.07
15	0.94	1.00	1.01
60	0.94	0.99	1.00

As we have seen before, we see most of the change when moving from 5 passenger vehicles to 15 passenger vehicles. Still this change is not as great as the decrease in Vehicle Miles we see when moving from 3 passenger vehicles to 5 passenger vehicles.

In order to get a feel for the relative effect of changing vehicle capacities in different parts of the country, we introduce the following graphic, Figure 10.4. This map shows the ratio of Vehicle Miles with capacity 15 to Vehicle Miles with capacity 5 in each state. The darker the shade of green, the higher the Vehicle Miles(5) to Vehicle Miles (15) ratio.

Figure 10.4: Vehicle Miles Ratios Map

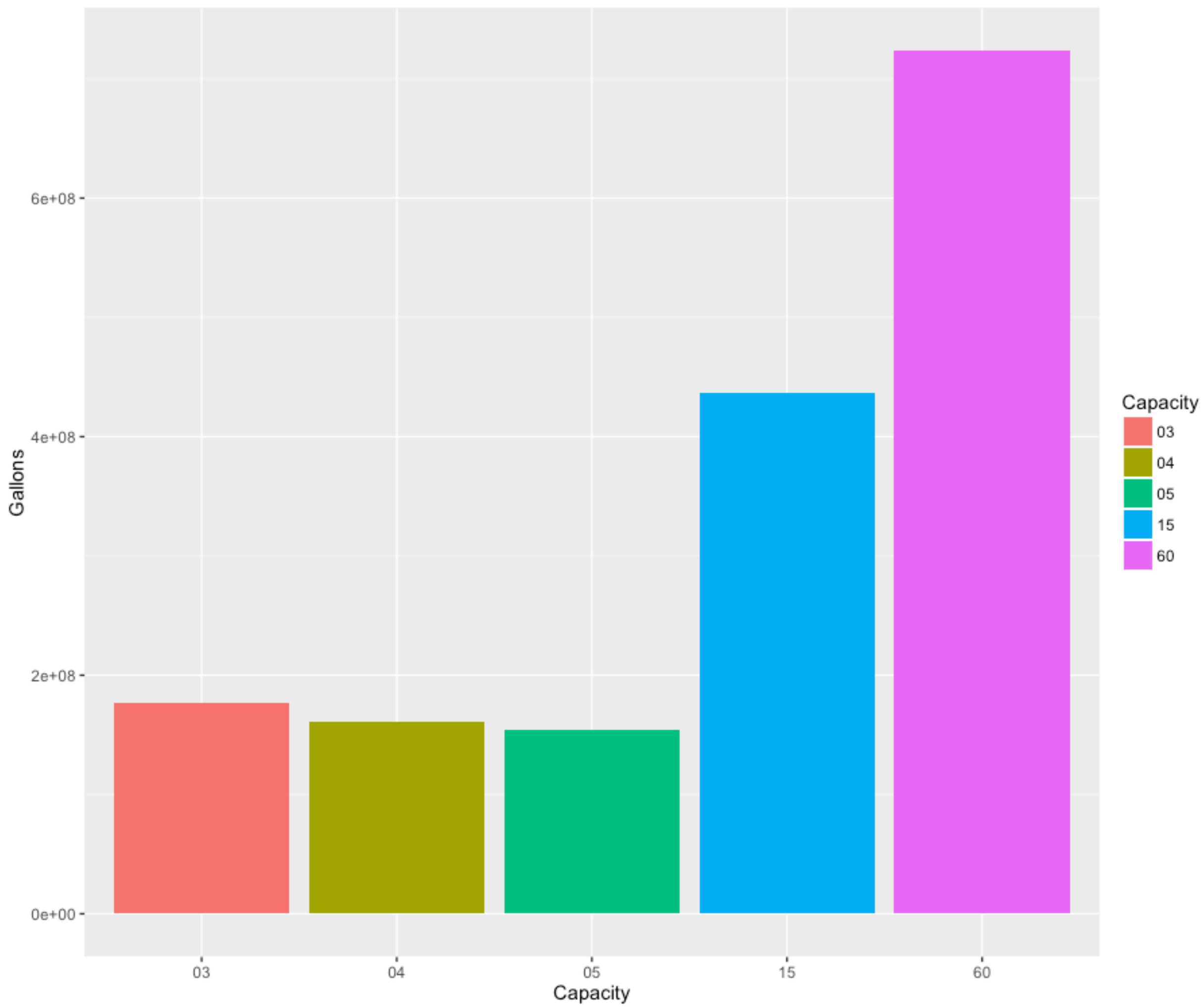


As we may have guessed, the greatest increases in Vehicle Miles when moving from capacity 5 to capacity 15 occur in some of the more urbanized states, like New York and Virginia. Rural states like Alabama and Arkansas see little change, indicating relatively few ridesharing opportunities. One outlier appears to be North Dakota, which appears a darkish shade of green. One reason for its large ratio may be that North Dakota has the 6th fewest Vehicle Miles. Any change in Vehicle Mile due to ridesharing will result in a relatively high ratio.

Before looking at our results for fleet capacity, we now look to get a feel for energy costs at the different capacities.

Fuel Economy

Figure 10.5: Gallons Consumed at Different Capacities



We determined our energy costs for the vehicles using an average of the highway and city fuel economies of each type of car. We divided the total VehicleMiles for each size car by the average fuel economy basing the 3, 4, and 5 passenger cars on a sedan (36 mpg), the 15 passenger car on a large van (12 mpg), and the 60 passenger on a bus (5.2 mpg). Larger vehicles use more fuel per mile because of the weight of the car so we expected these variances. As a result, the ratio between 60 and 5 passenger is 4.69. We conclude that fuel economy is a best indicator of which size vehicle is the optimal one. In our graph, five has the lowest cost followed by four, three, fifteen, and sixty is last. This is expected because we used the same size car (a five passenger) for 3,4, and 5 capacity vehicles. To have more accurate data on fuel size, we would have extrapolated fuel economy for 3 and 4 capacity cars. However, we wanted to use realistic data and because the capacities differ by at most two and the smaller capacities have much larger VehicleMiles it's not likely that a more exact fuel economy would make a reasonable difference in the cost.

Fleet Capacity

We now will look at capital costs across the different capacities. Our primary indicator for these costs will be fleet capacity. We arrived at estimates for fleet capacity using algorithm 10.2 included at the beginning of this report. Figure 10.6 shows a bar plot of the different capacities, and Table 10.10 shows the ratios of fleet size of capacities 5, 15, and 60.

Figure 10.6: Fleet Size at Different Capacities

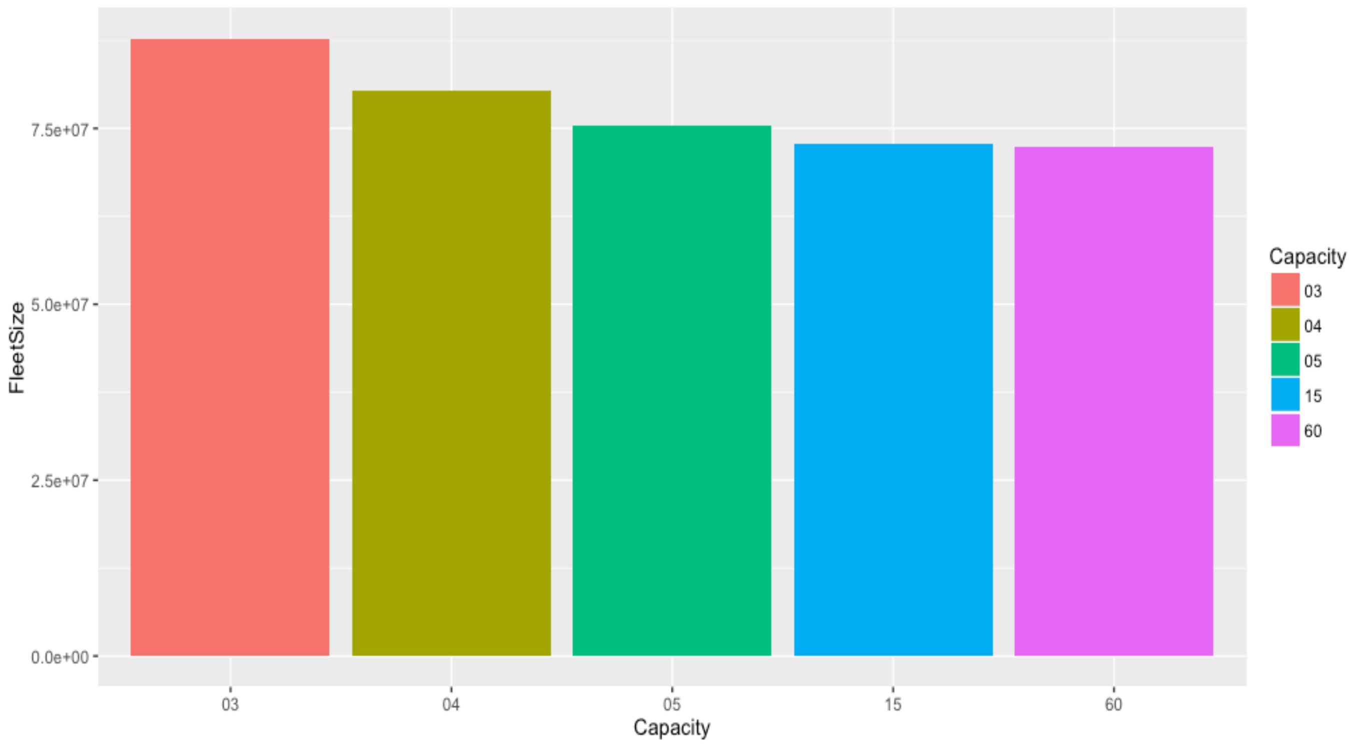


Table 10.10: Fleet Size at Different Capacities

	5	15	60
5	1.00	1.04	1.04
15	0.97	1.00	1.01
60	0.96	0.99	1.00

Fleet Size changes the least of all statistics as we increase vehicle capacity. Our ridesharing analysis indicates that it would not take very many more 5 person vehicles to serve the nation than it would 15 or 60 person vehicles.

Total Cost Analysis

Finally, we incorporate energy and capital costs in order to get an idea each capacity choice's cost implications. To calculate fuel cost, we use AAA's national average for gas prices. We use $2.35/gallon for capacities 3, 4, and 5, and $2.53/gallon (diesel price) for capacities 15 and 60.

To estimate daily capital costs, we researched average prices for each capacity. We use $23,000 as our estimate for capacities 3, 4, and 5 (new Toyota Camry). We use $33,000 as the cost of a 15 passenger van (cost of a Chevy van). We price a 60 passenger bus at $300,000. We multiplied vehicle cost by fleet size to get fleet cost. Assuming a 250,000 mile life for each vehicle, we can multiply this number by fleet size to get total fleet mileage expectancy. Next we divide fleet mileage expectancy by daily vehicle miles to get an estimate for each fleet's days in operation. Finally, dividing fleet cost by days in operation, we arrive at daily capital costs. Table 10.11 contains our energy and capital cost calculations and results:

Table 10.11: Fleet Size at Different Capacities

Capacity	3	4	5	15	60
Daily Gallons of Fuel	199,500,000	180,900,000	173,900,000	492,100,000	814,476,692
Price Per Gallon	$2.35	$2.35	$2.35	$2.53	$2.53
Daily Energy Operating Costs	**$468,825,000**	**$425,115,000**	**$408,665,000**	**$1,245,013,000**	**$2,060,626,031**
Fleet Size	87,790,000	80,290,000	75,410,000	72,880,000	72,470,000
Cost Per Vehicle	$20,000	$20,000	$20,000	$33,000	$300,000
Fleet Cost	1,755,800,000,000	1,605,800,000,000	1,508,200,000,000	2,405,040,000,000	21,741,000,000,000
Mileage Expectancy/Vehicle	250,000	250,000	250,000	250,000	250,000
Fleet Mileage Expectancy	21,947,500,000,000	20,072,500,000,000	18,852,500,000,000	18,220,000,000,000	18,117,500,000,000
Daily Miles Travelled	7,180,000,000	6,513,000,000	6,261,000,000	5,905,000,000	5,864,000,000
Fleet Life Expectancy (days)	3057	3082	3011	3086	3090
Daily Capital Cost	**$574,400,000**	**$521,040,000**	**$500,880,000**	**$779,460,000**	**$7,036,800,000**
Total Daily Cost	**$1,043,225,000**	**$946,155,000**	**$909,545,000**	**$2,024,473,000**	**$9,097,426,031**

Energy and capital costs are quite similar for capacities 3, 4, and 5. Energy costs dominate for capacity 15 and capital costs dominate for capacity 60. As our analysis up to this point has shown, changing capacities does not have a huge effect on total trips, vehicle miles, AVO, or fleet size. Thus, operating costs per mile will play the biggest role in determining the optimal capacity. Total cost is minimized at capacity 5. We would expect our analysis to produce this result because we assume the same operating costs per mile for capacities 3, 4, and 5 (and the higher the capacity, the lower the fleet size and vehicle miles). These costs are dwarfed those of 15 and 60 passenger vehicles. Smaller vehicles are more fuel efficient and cheaper to build.

Our ridesharing model shows that optimal vehicle capacity is attained when we cram as many people as possible into the vehicle that is cheapest to build and run. Nationwide, at least, there is simply not enough ridesharing potential to warrant the use of large-capacity automobiles like vans and buses.

Appendix

Table 10.2: Statewide Ridesharing Statistics for Capacity 4

State	Total Trips	AVO	PersonTripMiles	VehicleTripMiles	Fleet Size
Alabama	6,712,634	1.70	626,102,847	367,650,004	2,616,870
Alaska	567,842	1.57	37,459,606	23,786,660	195,935
Arizona	5,243,776	1.70	300,590,679	176,528,714	1,668,409
Arkansas	827,257	1.63	87,929,637	54,055,478	498,321
California	28,343,141	1.61	1,233,534,989	765,355,298	7,618,738
Colorado	3,812,288	1.52	129,052,522	84,681,744	1,203,728
Connecticut	3,018,193	1.47	128,603,686	87,257,546	930,254
Delaware	780,208	1.39	35,519,946	25,470,438	278,760
DC	544,674	1.20	37,900,605	31,474,607	397,872
Florida	13,946,608	1.59	504,623,994	317,216,439	3,958,951
Georgia	9,397,854	1.54	437,590,982	283,262,765	2,990,256
Hawaii	858,274	1.72	28,057,815	16,355,846	265,764
Idaho	1,226,653	1.46	36,962,916	25,373,715	383,408
Illinois	9,652,293	1.60	405,915,283	253,357,218	2,786,625
Indiana	5,709,455	1.51	204,655,921	135,648,848	1,887,994
Iowa	2,621,918	1.53	103,767,050	67,833,952	836,467
Kansas	2,339,342	1.50	95,990,327	63,979,532	742,704
Kentucky	4,145,936	1.46	147,700,006	101,203,760	1,343,111
Louisiana	3,962,242	1.52	175,410,171	115,114,586	1,239,992
Maine	1,437,930	1.44	49,617,670	34,494,430	486,978
Maryland	4,414,407	1.55	129,738,975	83,706,418	1,420,382
Massachusetts	5,055,594	1.55	162,957,686	105,099,868	1,440,743
Michigan	9,060,747	1.53	327,209,044	213,471,984	2,704,328
Minnesota	4,616,422	1.46	157,727,822	108,373,522	1,468,883
Mississippi	3,037,431	1.48	134,072,982	90,587,665	1,035,304
Missouri	5,291,148	1.54	203,029,143	131,492,146	1,729,022
Montana	866,565	1.48	37,164,882	25,111,191	273,857
Nebraska	1,428,496	1.51	46,165,634	30,582,133	448,171
Nevada	1,979,306	1.57	52,483,186	33,343,553	558,227
New hampshire	1,331,152	1.43	42,574,082	29,731,684	439,197
New jersey	6,002,672	1.57	184,621,675	117,401,272	1,815,027
New mexico	1,637,889	1.53	77,450,670	50,675,600	509,431
New york	13,183,704	1.56	495,006,193	316,822,506	3,766,373
North carolina	9,239,180	1.50	289,062,419	193,266,556	2,919,156
North dakota	562,846	1.57	28,353,985	18,020,990	182,226
Ohio	10,079,846	1.51	363,672,072	240,148,886	3,078,035
Oklahoma	3,334,902	1.51	127,394,207	84,174,496	1,070,352
Oregon	2,866,737	1.57	100,915,495	64,233,710	839,536
Pennsylvania	10,687,813	1.49	354,153,462	237,455,092	3,251,256
Rhode island	703,233	1.58	16,337,099	10,347,484	229,805
South carolina	4,546,133	1.50	162,752,948	108,631,500	1,453,042
South dakota	703,222	1.42	23,930,795	16,862,683	223,375
Tennessee	6,376,234	1.56	279,847,280	179,559,419	2,010,583
Texas	19,933,466	1.58	794,293,785	503,959,926	8,317,275
Utah	1,876,289	1.66	59,389,102	35,874,654	538,118
Vermont	688,518	1.45	23,933,950	16,532,811	230,073
Virginia	6,196,469	1.61	237,622,033	147,756,899	2,067,289
Washington	5,359,768	1.56	186,484,344	119,536,067	1,564,015
West virginia	1,802,112	1.49	61,887,765	41,480,316	661,182
Wisconsin	5,090,098	1.49	174,515,750	117,211,799	1,581,994
Wyoming	420,461	1.45	16,995,216	11,709,926	133,047

Table 10.3: Statewide Ridesharing Statistics for Capacity 5

State	Total Trips	AVO	PersonTripMiles	VehicleTripMiles	Fleet Size
Alabama	6,571,186	1.74	626,102,847	359,827,671	2,570,001
Alaska	549,413	1.64	37,459,606	22,860,642	187,674
Arizona	5,081,918	1.78	300,590,679	169,099,081	1,590,263
Arkansas	806,182	1.64	87,929,637	53,496,237	491,123
California	27,068,877	1.69	1,233,534,989	728,063,819	7,144,191
Colorado	3,684,922	1.59	129,052,522	81,257,144	1,147,482
Connecticut	2,943,120	1.51	128,603,686	85,032,999	903,301
Delaware	758,220	1.43	35,519,946	24,886,303	270,080
DC	491,836	1.32	37,900,605	28,695,730	361,607
Florida	13,536,902	1.65	504,623,994	305,395,221	3,795,508
Georgia	9,152,128	1.61	437,590,982	272,422,735	2,893,394
Hawaii	798,735	1.85	28,057,815	15,161,485	241,614
Idaho	1,196,787	1.51	36,962,916	24,542,653	371,717
Illinois	9,323,415	1.68	405,915,283	241,719,882	2,661,220
Indiana	5,574,959	1.56	204,655,921	130,872,639	7,819
Iowa	2,559,095	1.59	103,767,050	65,130,755	816,256
Kansas	2,282,881	1.56	95,990,327	61,684,796	722,798
Kentucky	4,040,746	1.51	147,700,006	97,720,416	1,302,013
Louisiana	3,847,947	1.59	175,410,171	110,280,755	1,194,468
Maine	1,415,696	1.48	49,617,670	33,631,201	478,104
Maryland	4,255,552	1.62	129,738,975	80,061,726	1,349,006
Massachusetts	4,904,020	1.61	162,957,686	100,960,042	1,381,276
Michigan	8,887,652	1.58	327,209,044	206,724,260	2,634,429
Minnesota	4,502,717	1.51	157,727,822	104,688,725	1,425,670
Mississippi	2,973,493	1.53	134,072,982	87,671,576	1,011,955
Missouri	5,164,034	1.60	203,029,143	126,512,186	1,680,101
Montana	846,320	1.53	37,164,882	24,277,930	266,778
Nebraska	1,387,128	1.58	46,165,634	29,301,806	433,194
Nevada	1,903,328	1.64	52,483,186	31,921,765	524,804
New hampshire	1,310,185	1.46	42,574,082	29,086,926	430,784
New jersey	5,812,634	1.64	184,621,675	112,581,083	1,737,409
New mexico	1,584,721	1.60	77,450,670	48,372,552	490,510
New york	12,542,497	1.66	495,006,193	298,839,021	3,538,000
North carolina	9,037,051	1.55	289,062,419	187,021,145	2,831,487
North dakota	549,387	1.65	28,353,985	17,230,723	177,963
Ohio	9,844,798	1.57	363,672,072	232,113,156	2,984,384
Oklahoma	3,253,214	1.57	127,394,207	81,075,460	1,041,333
Oregon	2,776,518	1.64	100,915,495	61,377,149	806,504
Pennsylvania	10,414,359	1.54	354,153,462	229,266,599	3,145,743
Rhode island	676,812	1.65	16,337,099	9,888,449	217,714
South carolina	4,452,957	1.54	162,752,948	105,463,419	1,413,468
South dakota	687,387	1.47	23,930,795	16,333,161	218,158
Tennessee	6,223,109	1.63	279,847,280	172,113,170	1,947,264
Texas	19,485,239	1.63	794,293,785	487,053,329	8,065,328
Utah	1,801,867	1.75	59,389,102	33,927,565	509,708
Vermont	676,531	1.49	23,933,950	16,108,637	225,879
Virginia	5,926,144	1.70	237,622,033	139,897,102	1,951,564
Washington	5,207,810	1.62	186,484,344	114,857,175	1,503,878
West virginia	1,758,740	1.55	61,887,765	39,891,265	647,685
Wisconsin	4,980,601	1.54	174,515,750	113,361,426	1,538,615
Wyoming	409,552	1.51	16,995,216	11,282,415	129,109

Table 10.4: Statewide Ridesharing Statistics for Capacity 15

State	Total Trips	AVO	PersonTripMiles	VehicleTripMiles	Fleet Size
Alabama	6,394,729	1.78	626,102,847	351,338,974	2,525,093
Alaska	524,183	1.74	37,459,606	21,571,751	175,436
Arizona	4,864,660	1.90	300,590,679	158,455,972	1,485,452
Arkansas	770,725	1.66	87,929,637	52,911,216	482,638
California	25,508,866	1.81	1,233,534,989	682,216,693	6,543,960
Colorado	3,515,443	1.69	129,052,522	76,417,595	1,070,389
Connecticut	2,837,640	1.57	128,603,686	81,965,360	867,591
Delaware	727,004	1.47	35,519,947	24,121,795	259,290
DC	410,420	1.51	37,900,605	25,116,091	311,163
Florida	13,024,769	1.74	504,623,994	289,528,981	3,573,463
Georgia	8,793,377	1.70	437,590,982	256,955,795	2,740,250
Hawaii	705,737	2.11	28,057,815	13,273,300	202,049
Idaho	1,157,732	1.58	36,962,916	23,369,691	355,662
Illinois	8,835,632	1.82	405,915,283	223,246,732	2,486,310
Indiana	5,390,475	1.65	204,655,921	123,745,593	1,687,589
Iowa	2,470,258	1.71	103,767,050	60,801,751	787,155
Kansas	2,208,479	1.64	95,990,327	58,584,060	696,210
Kentucky	3,897,088	1.58	147,700,006	93,226,484	1,243,519
Louisiana	3,695,258	1.69	175,410,171	103,958,405	1,134,341
Maine	1,385,962	1.53	49,617,670	32,366,726	465,928
Maryland	4,036,718	1.73	129,738,975	74,850,182	1,250,792
Massachusetts	4,693,579	1.71	162,957,686	95,059,087	1,294,651
Michigan	8,658,536	1.66	327,209,044	196,921,189	2,539,667
Minnesota	4,343,786	1.59	157,727,822	99,450,318	1,363,612
Mississippi	2,881,256	1.60	134,072,982	83,772,169	977,832
Missouri	4,994,369	1.70	203,029,143	119,202,864	1,610,960
Montana	818,582	1.61	37,164,882	23,069,836	256,194
Nebraska	1,328,455	1.69	46,165,634	27,324,930	412,381
Nevada	1,800,098	1.75	52,483,186	30,044,116	479,295
New hampshire	1,281,999	1.51	42,574,082	28,227,154	418,534
New jersey	5,573,254	1.74	184,621,675	106,269,358	1,636,604
New mexico	1,513,877	1.72	77,450,670	45,108,936	465,687
New york	11,556,061	1.83	495,006,193	270,288,999	3,170,006
North carolina	8,755,459	1.63	289,062,419	177,721,379	2,702,538
North dakota	529,400	1.79	28,353,985	15,835,770	171,861
Ohio	9,517,875	1.65	363,672,072	220,611,722	2,847,729
Oklahoma	3,141,615	1.66	127,394,207	76,552,161	999,336
Oregon	2,658,953	1.77	100,915,495	57,138,494	764,480
Pennsylvania	10,040,750	1.63	354,153,462	217,848,705	2,996,703
Rhode island	638,410	1.78	16,337,099	9,200,424	200,159
South carolina	4,328,643	1.61	162,752,948	101,212,878	1,356,039
South dakota	666,452	1.53	23,930,796	15,618,505	211,231
Tennessee	5,998,355	1.74	279,847,280	160,783,598	1,855,936
Texas	18,866,137	1.72	794,293,785	461,782,759	7,705,067
Utah	1,696,384	1.92	59,389,102	30,870,983	468,151
Vermont	660,273	1.54	23,933,950	15,540,475	220,112
Virginia	5,500,774	1.87	237,622,033	127,273,936	1,763,153
Washington	4,999,455	1.73	186,484,344	107,891,336	1,420,844
West virginia	1,703,839	1.64	61,887,765	37,695,743	627,795
Wisconsin	4,830,983	1.62	174,515,750	107,693,232	1,479,552
Wyoming	394,441	1.60	16,995,216	10,652,478	123,632

Table 10.5: Statewide Ridesharing Statistics for Capacity 60

State	Total Trips	AVO	PersonTripMiles	VehicleTripMiles	Fleet Size
Alabama	6,377,443	1.79	626,102,847	350,440,621	2,523,470
Alaska	521,888	1.75	37,459,606	21,428,719	174,313
Arizona	4,842,836	1.91	300,590,679	157,310,844	1,476,432
Arkansas	766,254	1.66	87,929,637	52,847,696	481,890
California	25,371,379	1.82	1,233,534,989	678,298,836	6,497,232
Colorado	3,498,914	1.70	129,052,522	75,871,322	1,064,144
Connecticut	2,825,008	1.58	128,603,686	81,569,190	864,396
Delaware	723,291	1.48	35,519,947	24,032,331	258,449
DC	402,409	1.52	37,900,605	24,945,004	307,800
Florida	12,970,963	1.75	504,623,994	287,951,231	3,549,968
Georgia	8,758,153	1.71	437,590,982	255,545,903	2,726,288
Hawaii	693,705	2.15	28,057,815	13,021,665	197,365
Idaho	1,155,023	1.59	36,962,916	23,256,097	354,654
Illinois	8,768,989	1.84	405,915,283	220,386,997	2,469,075
Indiana	5,370,437	1.67	204,655,921	122,791,371	1,657,577
Iowa	2,460,332	1.73	103,767,050	60,129,221	784,573
Kansas	2,201,921	1.65	95,990,327	58,304,773	693,985
Kentucky	3,882,664	1.59	147,700,006	92,799,925	1,238,219
Louisiana	3,680,212	1.70	175,410,171	103,341,821	1,129,412
Maine	1,383,098	1.54	49,617,670	32,188,515	464,991
Maryland	4,011,691	1.75	129,738,975	74,206,621	1,242,437
Massachusetts	4,670,039	1.73	162,957,686	94,339,735	1,285,801
Michigan	8,636,801	1.67	327,209,044	195,643,957	2,532,050
Minnesota	4,328,507	1.60	157,727,822	98,878,266	1,358,179
Mississippi	2,872,778	1.61	134,072,982	83,425,576	975,081
Missouri	4,978,607	1.72	203,029,143	118,287,002	1,605,739
Montana	816,076	1.62	37,164,882	22,927,040	255,346
Nebraska	1,321,992	1.71	46,165,634	27,043,518	410,914
Nevada	1,787,855	1.76	52,483,186	29,843,285	474,619
New hampshire	1,279,019	1.51	42,574,082	28,117,514	417,321
New jersey	5,552,588	1.75	184,621,675	105,724,446	1,628,044
New mexico	1,506,941	1.73	77,450,670	44,752,982	464,197
New york	11,432,949	1.86	495,006,193	266,660,781	3,126,942
North carolina	8,726,659	1.64	289,062,419	176,556,692	2,690,980
North dakota	527,152	1.82	28,353,985	15,583,511	171,438
Ohio	9,482,844	1.66	363,672,072	219,311,540	2,833,587
Oklahoma	3,130,697	1.68	127,394,207	75,992,089	995,838
Oregon	2,648,126	1.78	100,915,495	56,617,955	761,810
Pennsylvania	10,002,065	1.64	354,153,462	216,522,423	2,983,140
Rhode island	633,312	1.79	16,337,099	9,103,579	198,150
South carolina	4,316,132	1.61	162,752,948	100,781,073	1,350,067
South dakota	664,458	1.54	23,930,796	15,543,902	210,621
Tennessee	5,975,187	1.76	279,847,280	159,450,524	1,849,395
Texas	18,807,281	1.73	794,293,785	459,171,843	7,670,562
Utah	1,684,491	1.95	59,389,102	30,462,462	464,267
Vermont	658,479	1.55	23,933,950	15,462,253	219,610
Virginia	5,436,358	1.90	237,622,033	125,311,824	1,737,281
Washington	4,977,485	1.74	186,484,344	107,086,735	1,414,247
West virginia	1,698,753	1.65	61,887,765	37,397,250	626,175
Wisconsin	4,816,808	1.63	174,515,750	106,990,361	1,474,602
Wyoming	393,098	1.61	16,995,216	10,573,361	123,363

Chapter 11:
Multi-Sized aTaxi Fleet

In order to consider a complete transition away from traditional transportation and towards fully autonomous vehicle mobility, a thorough economic analysis of such transition is compulsory. Economic feasibility is a driving force in policy-making, technological research & development, and consumer adaptation. This chapter explores the economic cost of a fully autonomous aTaxi fleet in select sates of the United States, outlining what factors into the cost of every vehicle mile traveled by an autonomous taxi. In addition, this chapter presents a breakdown of the types of vehicles by passenger capacity that minimizes the total cost of fully meeting the mobility demand in the selected states.

Isabelle Uhl and Emilio Moreno

Section 11.2: Cost of every vehicle mile

The types of vehicles that have been considered for this aTaxi fleet are:

- 3-passenger aTaxi
- 6-passenger aTaxi
- 15-passenger aTaxi
- 50-passenger aTaxi

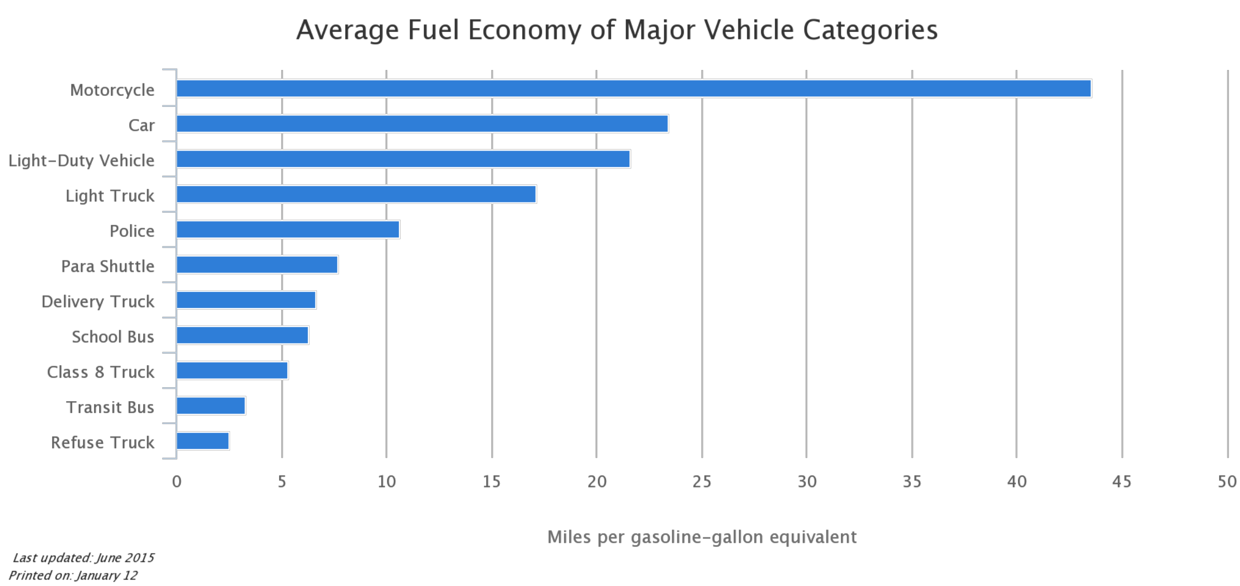
Currently, autonomous vehicle manufacturers have focused completely on small and sedan cars that fit four to six passengers. There is little public information about the manufacturing costs for these vehicles, as they have not been yet fully commercialized and they are still in a development stage; as such, it would be misleading to look into the capital costs of autonomous vehicle developers to estimate the cost of this aTaxi fleet.

Instead, the approach taken in this analysis has been under the assumption that autonomous driving technology has been fully developed, and is equally expensive to apply this technology to any sized vehicle. This cost analysis takes data from the vehicles currently serving mobility in the United States, and compares them to each other to determine the optimal fleet composition. Major vehicle costs have been broken down into Operating, Capital, and Environmental costs.

<u>11.2.1: Operating Costs:</u>

11.2.1.1: Miles Per Gallon:

The following data on Miles Per Gallon was collected from the United States Department of Energy (http://www.afdc.energy.gov/data/10310):



From the vehicle sizes presented by the US Department of Energy, it was determined that ‘Motorcycles’ were the most suitable comparable for a 3-passenger vehicle, performing at 43.54 MPG. Mileage for the category ‘Car’ was used for 6-person aTaxis, standing at 23.41 MPG. ‘Para Shuttles’ were deemed comparable to 15-person aTaxis, with a mileage of 11.69 MPG. Mileage for a 50-person aTaxi was taken from ‘Transit Buses’ with 3.26 MPG.

11.2.1.2: Lifespan:

According to the New York City Taxi & Limousine Commission, a taxi operating full time in New York City drives around 70,000 miles per year, and the average taxi operates for 3.3 years (http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf). Assuming the aTaxis in question will perform similarly to the current fleet of taxis in the City, an estimated lifetime of 231,000 miles has been assigned to 3-person and 6-person aTaxis:

70,000 miles per year x 3.3 years = 231,000 miles

A similar lifetime of 250,000 miles was considered appropriate for a 15-person vehicle.

Given the current the current usage of transit buses, a lifetime of 500,000 miles was considered appropriate for 50-person aTaxis, as suggested by a representative of UITP, *Union Internationale des Transports Publics*, during an in-class presentation with Professor Kornhauser.

11.2.1.3: Variable costs:

To determine variable costs, the report “Transportation Cost and Benefit Analysis II” by the Victoria Transport Policy Institute (VTPI) was consulted. This cost analysis defines vehicle costs into Fixed Costs and Variable Costs to the passenger. Fixed Costs include vehicle purchase, insurance and taxes – these costs have been disregarded, as the report assumes a fixed cost of zero for larger vehicles (since the fixed cost to a passenger of a bus is 0, given that they don’t own the vehicle themselves). In the aTaxi system, it would be the case that it’s a complete ride-sharing system, and so no passengers are vehicles owners – hence, fixed costs are zero.

Variable Costs as reported by VTPI include maintenance and repair costs, fuel, fuel taxes and oil, and paid parking and tolls. The report includes data for compact cars (proxy for 3-person aTaxi), average cars (proxy for 6-person aTaxi), Van/Light Truck (proxy for 15-person aTaxi) and Diesel Bus (proxy for 50-person aTaxi). Respectively, their costs are \$0.118, \$0.1637, \$0.2310 and \$2.4948 per vehicle mile (http://www.vtpi.org/tca/tca0501.pdf).

11.2.2: Capital Costs:

11.2.2.1: Purchase Price:

After consulting with Professor Alain Kornhauser, Director of Princeton University Transportation Program, it was determined that purchase prices for the 3, 6, 15, and 50 person aTaxis should be approximately $30,000, $60,000, $100,000 and $175,000. These prices carry a substantial premium over the current market comparable vehicles because it was assumed that the quality of vehicles needed had to be substantially higher to withstand the additional stress of constant usage in an autonomous vehicle fleet.

11.2.3: Environmental Costs:

11.2.3.1: Cost per ton of CO2:

According to the Environmental Defense Fund, the Institute for Policy Integrity and the National Resources Defense Council, the current social cost of every ton of CO2 emissions is $40 (http://costofcarbon.org/faq).

11.2.3.2: CO2 emissions:

According to the United States Environmental Protection Agency, every gallon of gasoline emits 8,887 grams of CO2, and every gallon of diesel emits 10,180 grams of CO2. It was assumed that the smaller vehicles (3 and 6 passengers) would operate with gasoline, and that the larger vehicles would operate with diesel. Vehicle mileage and environmental cost of carbon emissions were used to determine the economic cost of each of these vehicles due to their environmental impact.
(https://www.epa.gov/sites/production/files/2016-02/documents/420f14040a.pdf)

Below is a table that summarizes the cost per vehicle mile traveled by each vehicle type. Values in blue are all the assumptions discussed above, while values in black, under 'Total Costs', are computations derived from these assumptions.

Cost For Vehicle Miles (in 2015 dollars)				
# passengers	3	6	15	50
Operating				
MPG	43.54	23.41	11.69	3.26
Lifespan (miles)	231,000	231,000	250,000	500,000
Variable Cost (per mile)	0.119	0.164	0.231	2.495
Capital				
Purchase price	30,000	60,000	100,000	175,000
Environmental				
Environmental cost (per CO2 ton)	40	40	40	40
CO2 emissions (grams per gal)	8,887	8,887	10,180	10,180
Total Costs				
Purchase price over lifetime	0.130	0.260	0.400	0.350
Environmental	0.008	0.015	0.053	0.125
Operating	0.119	0.164	0.231	2.495
Total (per Vehicle Mile Traveled)	**0.257**	**0.439**	**0.684**	**2.970**

From the cost per vehicle mile traveled, it is clear that vehicles of 50 passengers are never an efficient choice. It is always cheaper to send four 15-person vehicles to serve 50 people instead of sending a single 50-person vehicle:

$$\$0.684 * 4 = \$2.736 < \$2.970$$

Given this inefficiency of 50-person aTaxis, this type of vehicle was excluded from the analysis, resulting in an optimal fleet containing only 3, 6 and 15-person vehicles.

Section 11.3: Methodology to optimize fleet size

In order to determine the optimal vehicle assignment in order to minimize the cost of operation, every possible combination of trip assignments must be assessed. All aTaxi trips have 1, 2 or 3 stops, originating from a single location and without the ability to pick up passengers in any of the stops. To compute the optimal vehicle assignment, it is necessary to check every combination of vehicle for every possible route the vehicle can take, in such a way that cost is minimized.

For the case when all passengers share 1 destination, it is easy to see (given the costs outlined above) that a 3-person vehicle will serve 0-3 passengers, a 6-person vehicle will serve 4-6 passengers, and a 15-person vehicle will serve 7-15 passengers. When the number of passengers 'N' exceeds 15, the modulus operator is applied: N % 15. This determines the number of 15-person vehicles needed, and the remainder of N will always fit into one of the three categories just described.

For the cases with 2 or 3 destinations, it is sometimes possible that it is more efficient to send two vehicles in a trip that could have in theory be served by a single vehicle. Because it is impossible to know which trips will fit into this category, it is necessary to check all possibilities, making the algorithm extremely inefficient. Due to this inefficiency in the system, the analysis for a nationwide optimization was infeasible (system memory runs out with files larger than 500 MB, which is lower than the average file size for every state's aTaxi trips. With files this large, the computer Terminal failed to process all the necessary information). For the purpose of this analysis, the following states were selected for the optimization of their aTaxi fleet:

- Alaska
- Washington D.C.

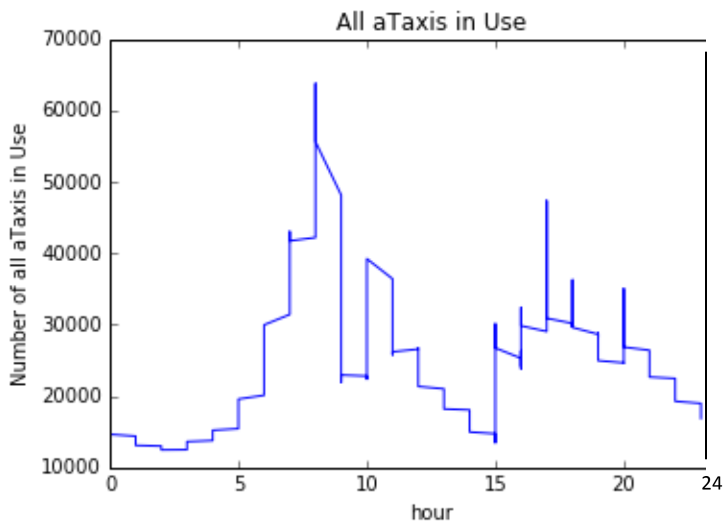
Section 11.4: Alaska

Alaska is the largest state of the United States, encompassing 663,268 square miles of area. The analysis in this project is for aTaxi trips with trip length between 0.5 miles and 200 miles; because Alaska is not part of the continental United States (and hence is over 200 miles away from the nearest state), all the trips contained in this analysis have an origin and/or destination in Alaska.

The population of Alaska is approximately 738,432 people. The largest cities by population are Anchorage (260,283 people), Fairbanks (30,224 people) and Juneau (30,711 people).

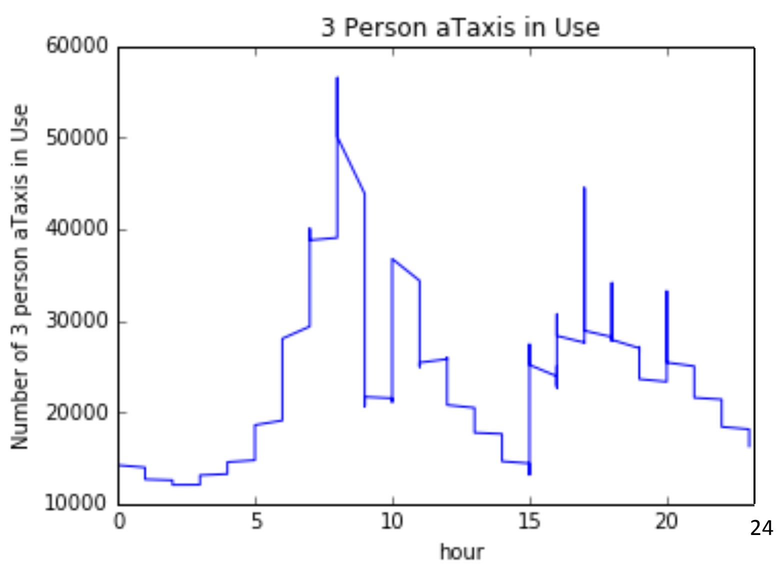
Section 11.4.1.1: aTaxis in use

This section explores the potential usage of autonomous taxis in Alaska throughout a typical day, followed by a concluded minimum fleet size for the state, along with an estimated economic cost of a fully autonomous fleet meeting 100% of the mobility demand in the state.



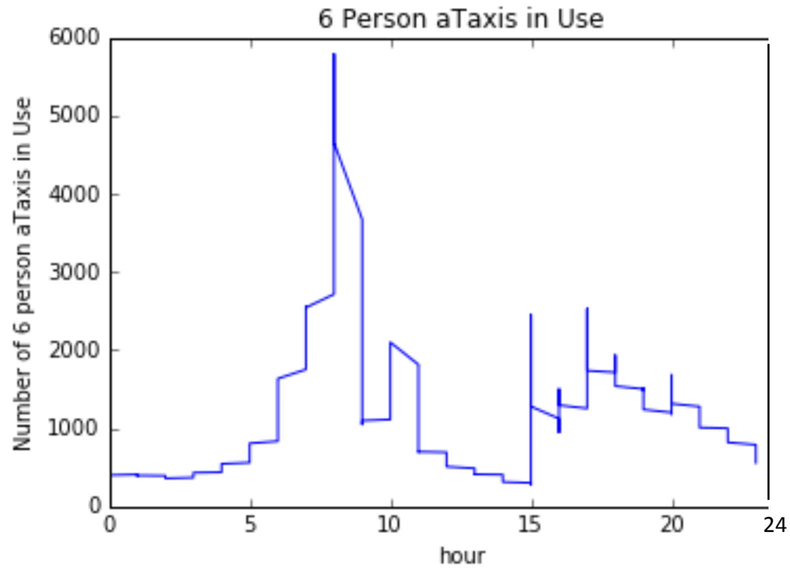
Max: 63,760 aTaxis in use at 8:00 am.
Min: 12,460 aTaxis in use at 2:00 am.

The distribution of taxi demand in Alaska seems normal, with sharp peaks around 8:00am and 6:00pm. The peak that occurs right before 11:00am seems deviant from the norm throughout the United States (such peak usually occurs around noon, when workers go out for lunch). This may reflect different behavioral patterns in Alaska.



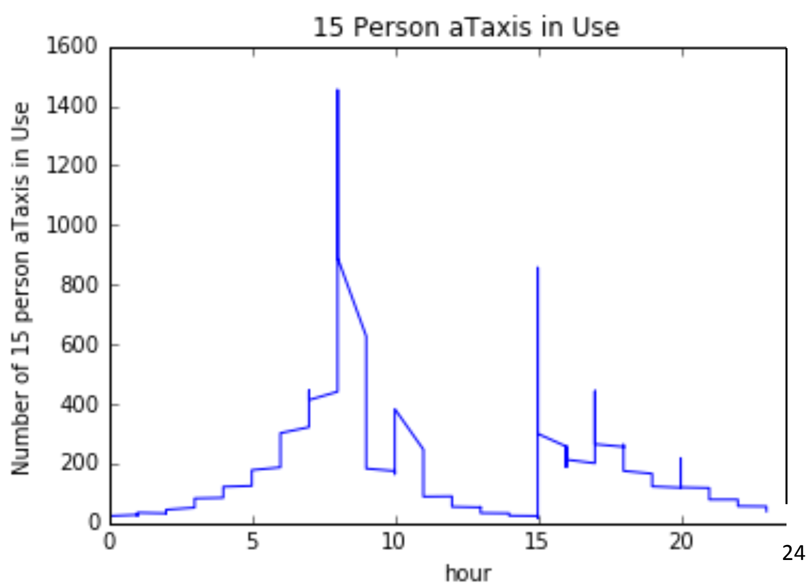
Max: 56,522 3-person aTaxis in use at 8:00 am
Min: 12,056 3-person aTaxis in use at 2:00 am

The lower bound on the number of 3-person aTaxis in Alaska is 56,522 vehicles. This amount would allow for all 3-person vehicle trips to be serviced, assuming a fully efficient relocation system is in place.



Max: 5,785 6-person aTaxis in use at 8:00 am
Min: 279 6- person aTaxis in use at 3:00 pm

The lower bound on 6-person aTaxis is 5,785 vehicles, to service the peak demand for this type of vehicle during the morning rush hour.



Max: 1,453 15-person aTaxis in use at 8:00 am
Min: 16 15-person aTaxis in use at 3:00 pm

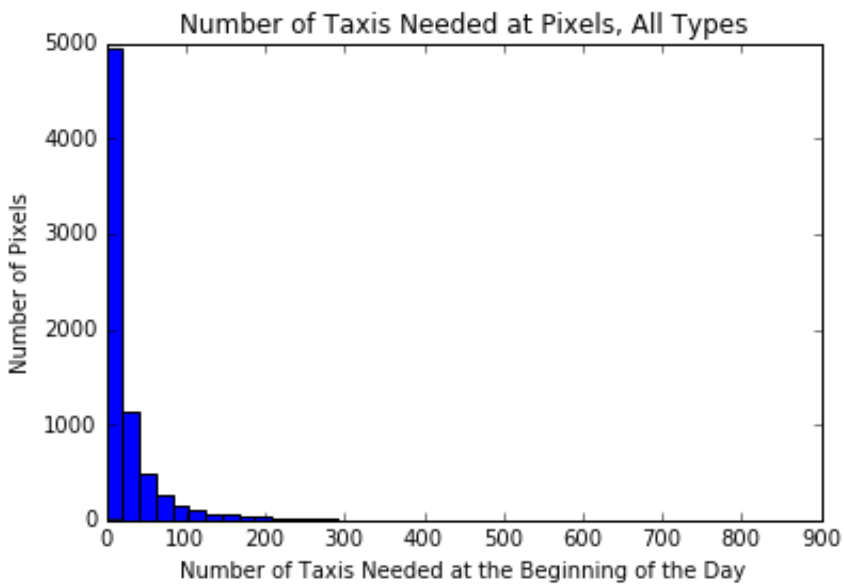
The minimum number of 15-person vehicles is 1,453.

11.4.1.2: Conclusions

The minimum fleet size to service all aTaxi trips in Alaska is 63,760 taxis, 88.6% three-person vehicles, 9.1% six-person vehicles, and 2.3% fifteen-person vehicles.

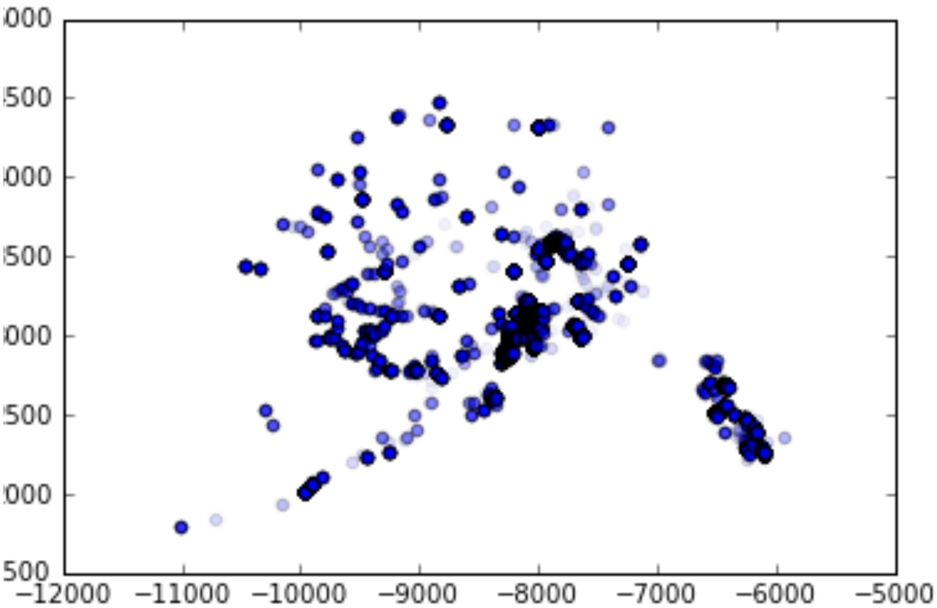
The total cost of operating this fleet is $2,658,949, servicing 28,360,568 person trip miles and 17,443,749 vehicle trip miles. The average vehicle occupancy (AVO) in a fully autonomous taxi fleet in Alaska is 1.626

Section 11.4.2: Vehicle Management

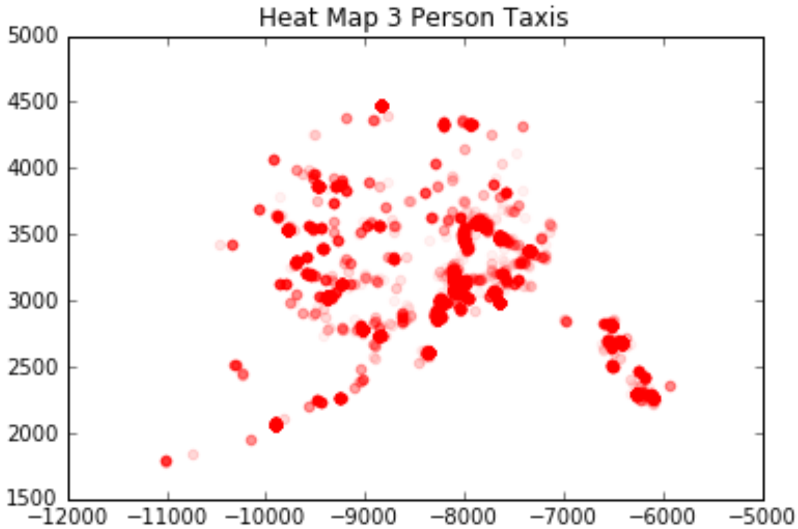


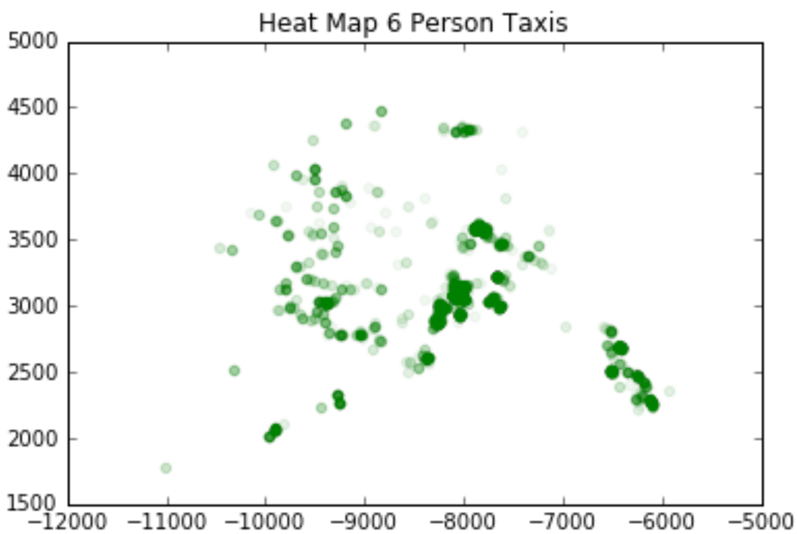
The histogram above displays the number of pixels that need a certain number of aTaxis in the beginning of the day to fully service their trips throughout the day. Most pixels need very few aTaxis in the beginning of the day, as expected, because most trips being generated have very few people on them.

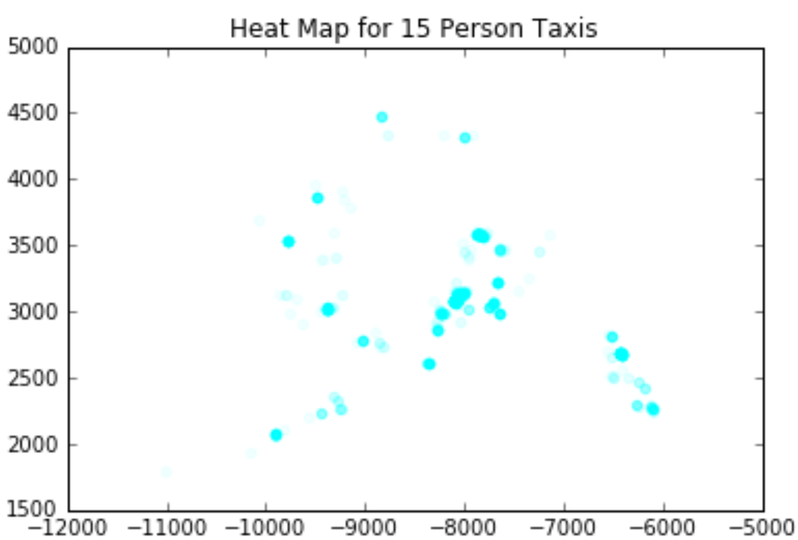
The map below displays a heat map for the vehicle demand at the beginning of the day throughout the state. Clearly, the regions with the highest demand occur near the cities of Anchorage, Fairbanks and Juneau, which are the three largest cities of the state. In addition, there is significant demand for mobility near the city of Bethel along the Southwest coast; Bethel is the main port on the Kuskokwim River, and Bethel Airport is a transportation hub for the region.



The maximum demand for 3-person taxis in a single pixel is 1,083. For a 6-perosn vehicle it is 134 and for a 15-person vehicle it is 54. Below are the maps for the demand for each type of vehicle individually, by trip origination (where x-person taxi represents a maximum capacity of ‘x’ for every taxi in transit particular taxi:


Heat Map 3 Person Taxis
5000
4500
4000
3500
3000
2500
2000
1500
−12000
−11000
−10000
−9000
−8000
−7000
−6000
−5000

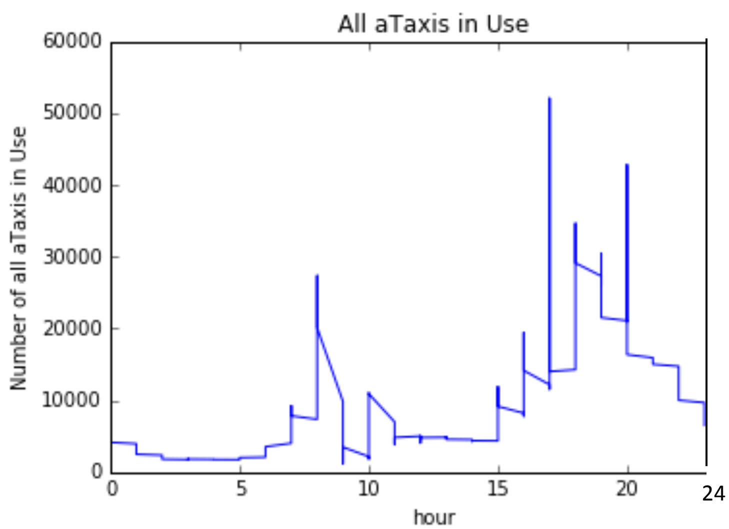

Heat Map 6 Person Taxis
5000
4500
4000
3500
3000
2500
2000
1500
−12000
−11000
−10000
−9000
−8000
−7000
−6000
−5000


Heat Map for 15 Person Taxis
5000
4500
4000
3500
3000
2500
2000
1500
−12000
−11000
−10000
−9000
−8000
−7000
−6000
−5000

Section 11.5: Washington D.C.

In contrast to the extensive area of Alaska, Washington DC has around 68.34 square miles in area, roughly 10% of that of Alaska. The population in the District is 681,170, only 100,000 less than in Alaska. The small area and high population density of Washington DC provide an interesting contrast to the large area and minimal population density of Alaska. The following section explores the feasibility and cost of an autonomous taxi fleet in the District of Columbia.

Section 11.5.1.1: aTaxis in use

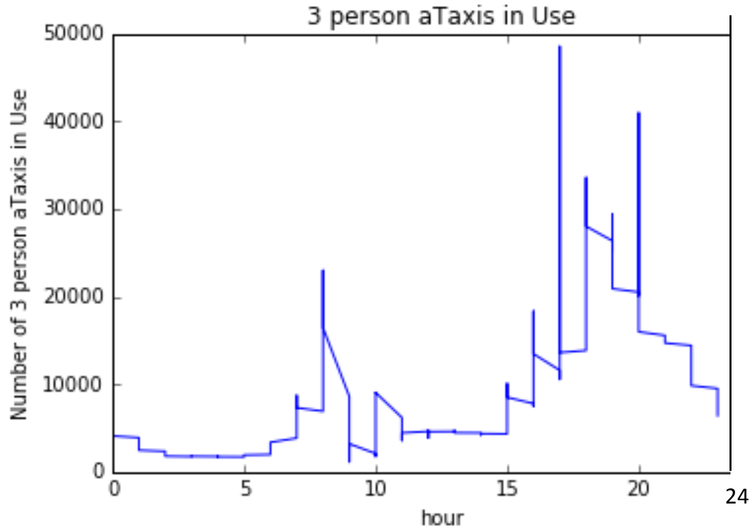


Max: 52,065 aTaxis in use at 5:00 pm
Min: 1,150 aTaxis in use at 9:00 am

The time-distribution of vehicle demand in Washington DC is markedly different from Alaska and from the general trends of mobility across the country. The morning peak in the District is roughly half the size of the afternoon peak; this is reasonable given that Washington DC is a primarily governmental employment region, with a relatively older population. The reduced number of children that must go to school in the morning, and the more flexible nature of diplomatic schedules, make the morning peak abnormally small.

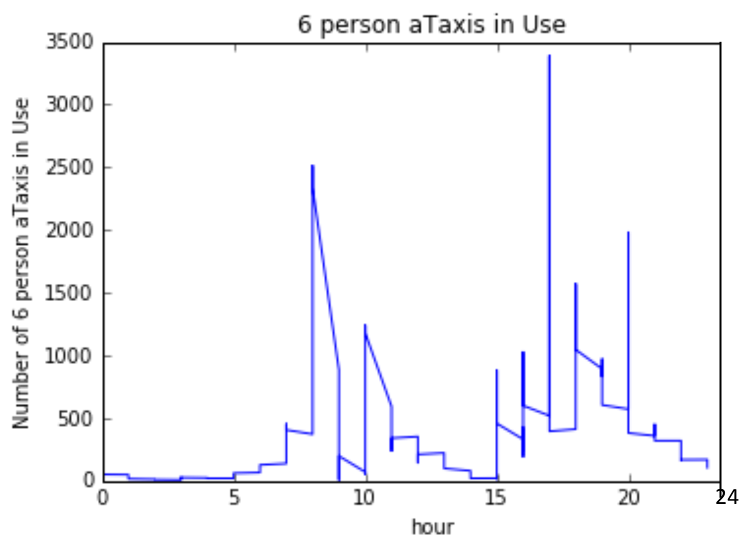
Likewise, the second peak that occurs around 8:00pm is most likely explained by the later schedules kept by many diplomatic workers that don't get out of their job at a regular 5:00pm schedule.

Below is a breakdown of mobility demand by each type of aTaxi.



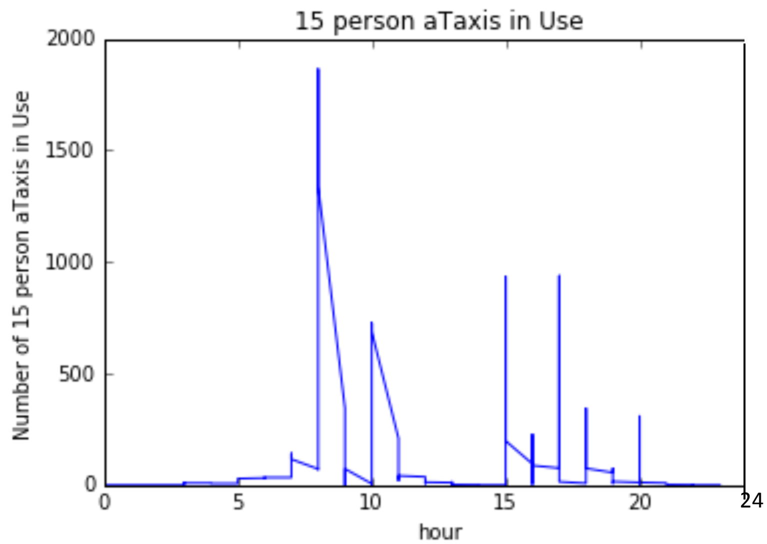
Max: 48,522 3-person aTaxis in use at 5:00 pm
Min: 1,138 6-person aTaxis in use at 9:00 am

The distribution of 3-person aTaxis in use throughout the day is very similar to the distribution of all taxis because smaller vehicles service the majority of mobility in DC.



Max: 3,391 6-person aTaxis in use at 5:00 pm
Min: 8 6-person aTaxis in use at 3:00 am

The morning peak for 6-person aTaxis is more significant that the morning peak for the smaller vehicles; likewise, the 8:00pm peak for 6-person taxis is greatly reduced. A plausible explanation for this behavior is that morning trips are easier to share among more people: most people live in very densely packed pockets in the region, and work in similar areas, so the commute to work/school is optimized by ride-sharing. However, as workers disperse throughout the day, it is harder for them to share rides back home towards the end of the working day.



Max: 1,864 15-person aTaxis at 8:00 am
Min: 0 15-person aTaxis at 12:00 am

11.5.1.2: Conclusions

The minimum total fleet size to fully service Washington DC is 53,777 vehicles, servicing 90.2% 3-person taxis, 6.3% 6-eperson taxis and 3.5% 15-person taxis:

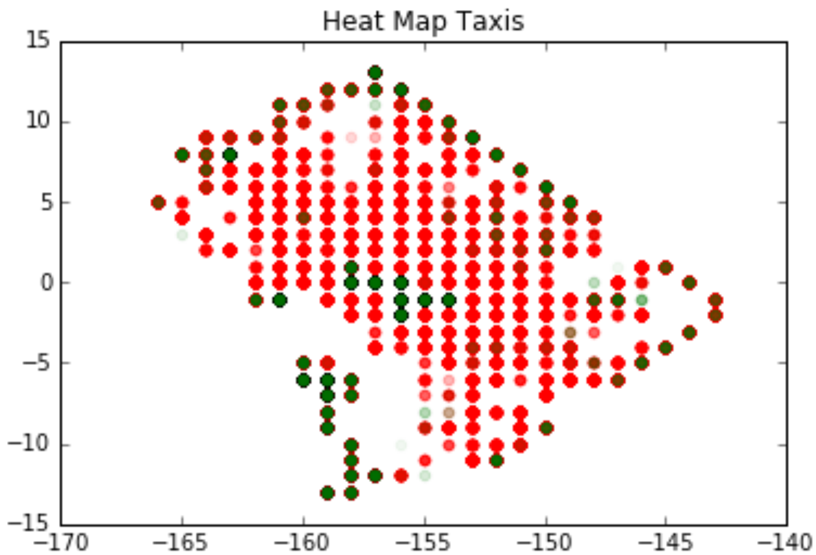
48,522 three-person taxis + 3,391 six-person taxis + 1,864 fifteen-person taxis = 53,777 taxis

This number is slightly larger than the minimum number of taxis in use at any point in the day (52,065). This happens because the peak demand for individual types of vehicles does not occur at the exact same time in the day (for Alaska, it happened that the peaks overlapped, and so the two figures matched exactly).

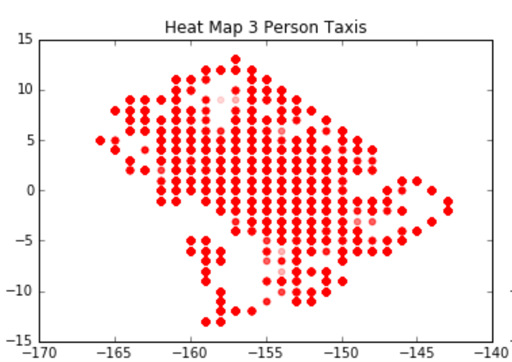
The total cost of operating this fleet is $1,723,907, servicing 11,970,176 person trip miles and 6,869,580 vehicle trip miles. The average vehicle occupancy (AVO) in a fully autonomous taxi fleet in Washington DC is 1.742, slightly higher than in Alaska. A higher average vehicle occupancy is expected in more densely packed states, because the potential of ride sharing is maximized by people working and living near each other; the above results are consistent with the expectations.

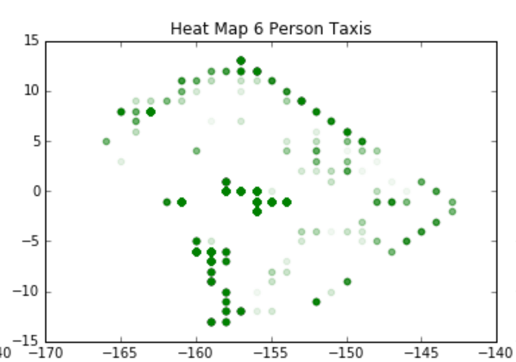
Section 11.5.2: Vehicle management

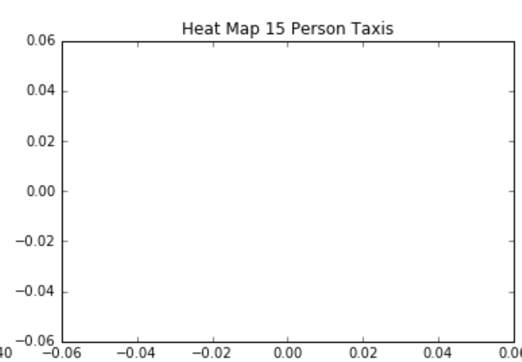
The following maps display the vehicle demand by pixel at the beginning of the day throughout Washington DC. Pixels in red represent demand for 3-person Vehicles, pixels in green represent demand for 6-person vehicles and pixels in cyan represent the demand for 15-person taxis.



The maximum demand for 3-person vehicles in a single pixel is 38,461 vehicles. The maximum for 6-person vehicles is 3,889 and the maximum for 15-person vehicles is 285. However, by looking at the individual heat map for the need of 15-person aTaxis, it becomes clear that demand at the beginning of the day is largely met by the same vehicles that end their trips at the given pixel. In other words, the demand at the beginning of the day is often met by the supply at the end of the day, minimizing the need for empty vehicle repositioning, which is why the demand heat map for the larger vehicles is completely empty. The maps below are a breakdown by vehicle type of all vehicle demand that occurs in Washington DC at the beginning of the day.







Chapter 12:
Vacation Trips & Critiques

Much of the previous data and research has focused on modelling an aTaxi service based on a typical day's travel needs for all people. But what about atypical trips? How do we model the rare vacation trip? We use a simple approach in this chapter to explore the characteristics of trips that are greater than 200 miles that originate from Orlando, Florida as a representative city. Through this process, we realized that the data used to conduct all of the previous research is inadequate and incorrect to build a model of vacation trips.

Houston Martinez & Wyatt Navarro

Section 12.1: Overview of Vacation Trips

For this analysis, we chose Orlando, FIPS code 12095, as our model city since there is a lot of vacation activity there. We restricted our focus to trips greater than 200 miles, so we used the LongOriginPixel_12095_1.csv file which only contained these trips. This file contains trips that originate in FIPS code 12095 and have a distance greater than or equal to 200 miles. However, we realized that it had many trips that did not actually originate in the vicinity of Orlando. Some were originating in places like South Carolina. So we filtered this file to only include trips the originated in the counties in the immediate vicinity of Orlando. While we hard coded this filter, we imagine that it would be useful to use a clustering algorithm or Kd-tree to assign each person to their nearest airport or major city. We chose to use the trips that originated in Orlando versus the trips that terminated in Orlando for two reasons. One, we assumed that modeling the vacation trips leaving Orlando trips would be just as useful as modeling the trips terminating in Orlando since vacation trips are usually roundtrip. Second, the originating trip data was simply more accessible and required little work to acquire. Once we had these trips correctly filtered we were ready to start the analysis.

Section 12.1: Level of Service Formulation

We decided to analyze vacation trips on a macroscopic scale. In anticipation of a large number of people heading back to the same destinations, we modeled the parameters of our aTaxi trips on an airport system. We envisioned one major aTaxi hub in a central location near Orlando where all trips over 200 miles would depart from on a type of vehicle that would be determined by the results of our analysis. For this location we chose coordinate (28.531100, -82.316506). Then, we chose to assign an automatic delay departure of 2 hours to all departing trips in order to produce a more significant probability of travelers having a common location. As soon as a new aTaxi trip was created, anyone in the whole Orlando area that also was traveling to the same destination as that aTaxi within the next two hours would be assigned to that aTaxi. This is much like how an airport works. Anyone in all of Orlando that is heading to New York City will go to the same airport and get on a new flight to New York City that probably departs at least every two hours.

This scenario ignores traveler's commutes to the departure hub because we assume that it is not that important since the primary length of the trip will be over 200 miles. We are also aware that another group did the analysis for trips to the nearest airport so we felt that this was unnecessary. We also purposefully chose to ignore existing airports in order to be unconstrained by existing infrastructure. By doing this, we are able to save the actual destinations that people's trips terminate at rather than the airport that they are forced to arrive at. This would allow future transportation systems, whether that be some sort of hyperloop system or high-speed bus system, to optimally place their long distance hubs near people's destinations or to optimally link these systems between cities together. Finally, we did not set a maximum capacity to the generated trips because we wanted to get a full picture of the demand for this service, especially if it has similar capacity to airplanes as we hypothesized.

Section 12.2: Methodology

Our analysis of vacation trips relied heavily on our Homework 6 assignment's generation of aTaxi trips for various Level of Service parameters. But it was a much simpler task. The entire file of person trips was loaded into a Matlab buffer after being sorted by OYPixel, OXPixel, and departure time. To start the process, our program reads in the first person trip line from the OriginPixel file for FIPS code 12095, and uses its destination as the aTaxi vehicle's destination (ignoring any preexisting airports or transportation systems). The aTaxi will also have a departure time that exactly two hours after this first passenger is assigned to the aTaxi. Since the person trip file is sorted by departure time and the trips are filtered to originate from the Orlando area already, we use a binary search to find all the trips that arrive within the next two hours. Once we have those trips, determining which ones get assigned to the aTaxi is a simple task. We chose to assign trips to an aTaxi based on a destination radius. We used two different radii in our analysis: 20 miles and 30 miles. If a person trip's destination is within the chosen radius of the aTaxi then it is assigned to the aTaxi, much how like people take airports that take them within a reasonable distance of their destination. After all the trips that depart within the delay departure of the aTaxi are analyzed, the aTaxi would depart and we print out the trip to a file. Then the next person trip is read in from the OriginPixel file and a new aTaxi trip is created.

The output of our program was [County_Code, Capacity,TripLength, OLat, OLon, DLon, DLat, DTime, Delay]. By only requiring that a person's destination be within a 20 or 30 mile radius and leave within a delay departure of two hours, we hoped to aggregate a greater number of people onto the aTaxi trips and generate a larger average vehicle occupancy (AVO).

Section 12.3: Results

As was probably apparent in the Transportation Symposium where we presented our results, we had some confusing results that were inconsistent with what we know based on our general knowledge about travel patterns in the United States. I will recap those results here.

Level of Service 1: Delay Departure=2 hours, Destination Radius = 30 miles

Using a delay departure of 2 hours, a destination radius of 30 miles and an uncapped capacity, we produced aTaxi trips with an AVO of 1130 and a median trip length of 880 miles which is about the distance to Washington, DC. Needless to say, this extremely high AVO was an immediate signal that something might have been wrong. When we went through the generated trips we also found that the maximum capacity on a trip was one to a suburb of Montgomery, Alabama with a capacity of 104,009. This clearly had to be incorrect.

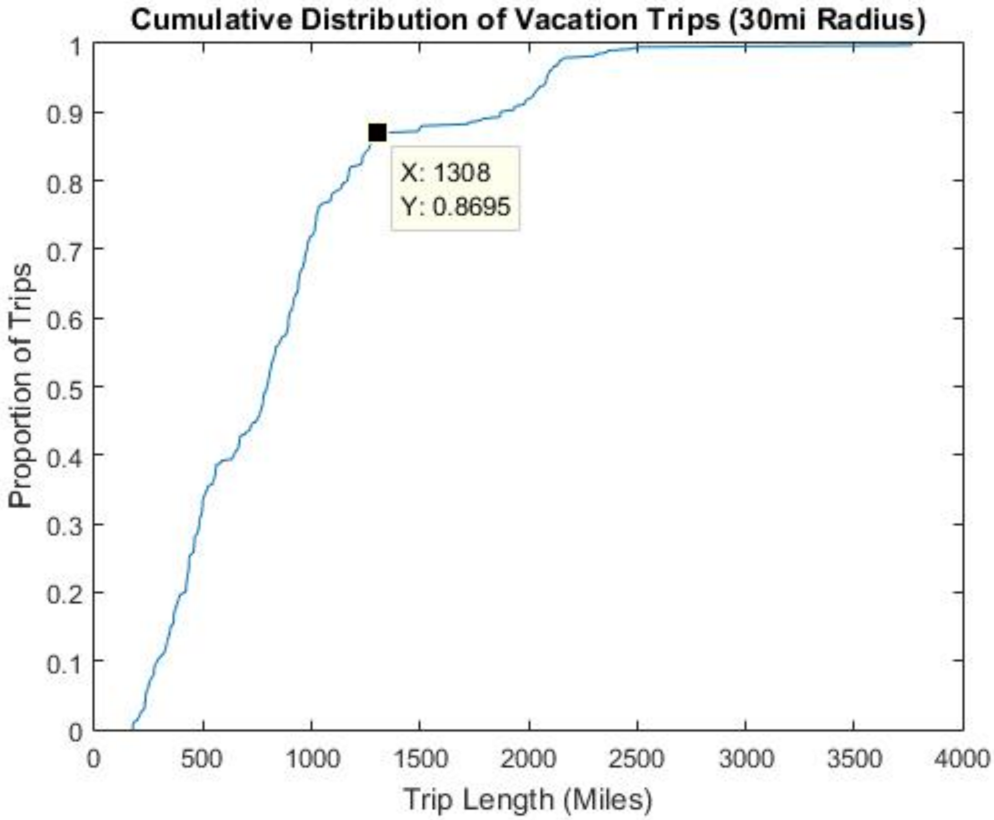


Figure 12b

In figure 12b above, we can see that the about 95% of trips are shorter than 2100 miles and at least 87% of trips are shorter than 1300 miles. A service for vacation trips that does not involve air travel would likely ignore the trips above 1750 miles and would definitely ignore trips longer than 2500 miles. For reference, 2500 miles would get you all the way to Seattle from Orlando in great circle distance and 1500 miles would get you to Denver. This distribution of long trip lengths bodes well for a high-speed autonomous vehicle transportation system that wants to serve a large majority of vacation trips within a manageable distance. 1300 miles gets you to New York City, Dallas, and Kansas City.

In figure 12c below, we have made a bubble chart that aggregates the autonomous vehicle generated trips by their destinations. It is important to remember that this chart does not take into account the number of people on each of these trips and that the capacity is unlimited on these trips. If these anomalous trips are excluded from the generated trips data, then the AVO for vacation trips drops down to a more reasonable 364 and the trip with the highest capacity is to Tallahassee, FL which is more reasonable.

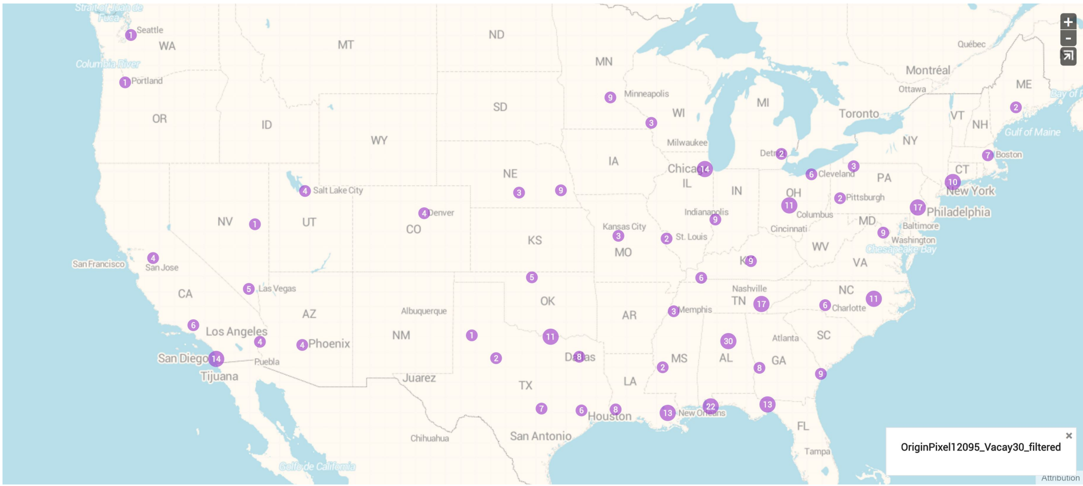


Figure 12c

As you can see, Alabama has the most trips terminating within its borders. It has two aggregated bubbles with a total of 52 trips terminating there. In contrast, the bubble over Chicago's and New York City's airports have only 14 and 10 trips terminating there respectively. This does not adhere to logic since those cities have some of the busiest airports in the nation and would be expected to be some of the most popular destinations for people taking long trips out of Orlando. Moreover, the top five generated trips by capacity all had destinations in Alabama, and they all had capacities over 20,000.

Level of Service 2: Delay Departure=2 hours, Destination Radius=30 miles, Max Capacity=300 passengers

Once we knew that there was some pretty significant demand for vacation trip aTaxis, we imposed a max capacity of 300 passengers on the trips. This generated trips that had an AVO of 239 passengers, a mean trip length of 545 miles, and a median trip length of 368 miles. The cumulative distribution of these trips' lengths is shown below in figure 12d.

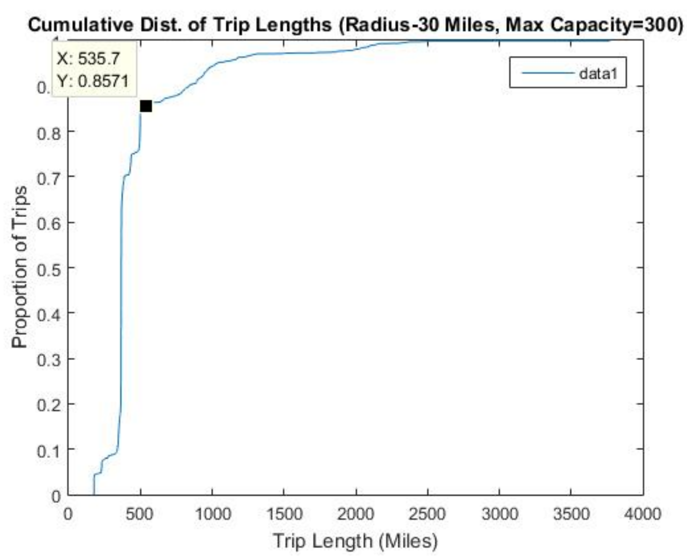


Figure 12d

This distribution of trips is even more amenable to an autonomous taxi system. 85% of trips are less than 535 miles. This means that it would not be too difficult to create a network or transportation system to serve most of these trips.

Level of Service 3: Delay Departure=2 hours, Destination Radius=20 miles, Infinite Capacity

We also generated aTaxi trips using all the same parameters as before but instead using a 20 mile destination radius. The AVO generated in this scenario was 325. The cumulative distribution of the trip lengths is shown below in Figure 12e. The median trip length is 816 miles for this scenario.

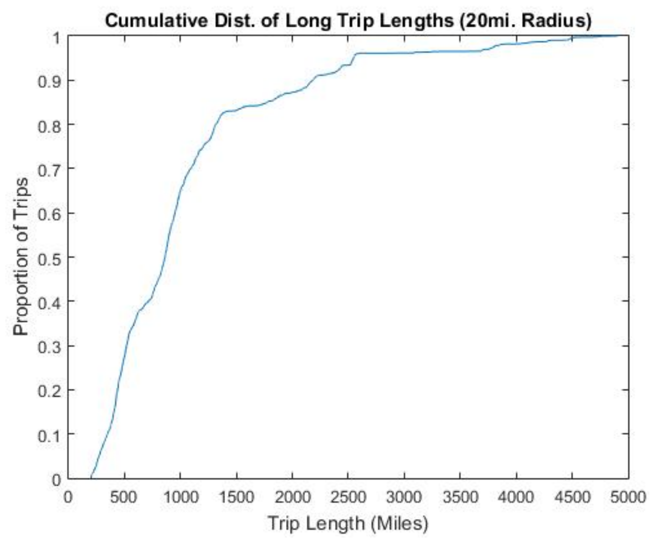


Figure 12e

Since these trips were also generated from the same underlying data from the OriginPixel12095 file, it has the same problems as the 30 miles radius data. There were far too many people taking trips to Alabama. This can clearly be seen in figure 12f below.

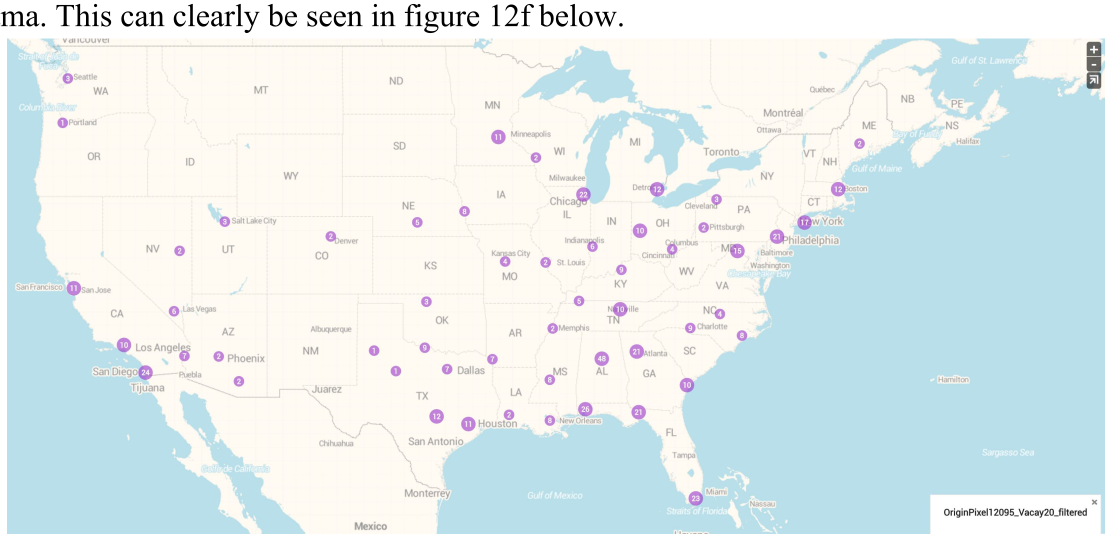


Figure 12f

Alabama clearly has a disproportionate number of trips terminating in it. And this graphic does not consider the capacity of these trips or Alabama would stand out even more because of the huge capacity of trips terminating there.

We also generated trips for one more scenario in which we cap the trip capacity at 300 passengers, but at this point it is fruitless to present that data here. It follows a similar pattern as the 30 mile radius scenario seen above. Instead, we offer a few solutions to address the vacation trip data in the future.

Section 12.4: Vacation Trip Critique

We conferred with Professor Kornhauser, and he agreed with us that the underlying data in the LongOriginPixel12095_1.csv file, and likely all other long trip files, was incorrectly generated. So thankfully, our code was not incorrect when said at the symposium that we had over 100,000 people taking a trip to Alabama. At this time, we have not addressed the problem with the long trip generation process but we have a few suggestions.

Incorrect Assumptions from the Top Down

The process for generating all trips involved sampling from census data and various other sources to give each simulated individual a ‘home’ and a ‘work’ and many other necessary characteristics that help determine their ‘average’ daily tour. Theoretically, the only trips that should be taken that are over 200 miles are those few people that had their work location assigned to them over 200 miles away. But while going through the LongTrip file even that does not seem to be the case because the majority of the trip destinations are of type ‘Other’. Some of the trips just do not make any sense at all. For example, I love Chick-Fil-A chicken sandwiches, but one person in Orlando, Florida traveled all the way from his workplace at Publix Super Market to Alabama to eat some Chick-Fil-A. I just cannot imagine that that is in any way realistic. So instead of trying to simulate vacation trips in the same disaggregate manor as the normal trips are simulated, we recommend a different approach.

Disaggregate for the Typical Trip, Aggregate for Vacation Trips

Depending on your goals, it is likely that performing a disaggregate generation of person trips for vacation trips is unnecessary and inaccurate. Exactly what details about a person's characteristics would tell you that he or she is more likely to take a trips on a given day? Using a person's income to scale the probability that he or she decides to take a vacation on a given day may be helpful. But even then, a wealthy person still probably only has a 2% chance of going on a vacation on an average day. Doing that may provide a more realistic number of people that decide to take a trip greater than 200 miles on a given day, but how do you then chose that person's destination? Surely nothing about a person could ever tell us where he or she is likely to take a vacation. What would be more useful is to abandon the disaggregate approach for long trips all together. The normal, non-vacation trips should still be produced by the disaggregate approach because that detail is essential to aTaxi system planning. But that detail is not necessary for the long stretch of distance between long trip origins and destinations. The disaggregate part of aTaxi trip generation should be certain that there are no trips greater than 200 miles that are of type 'Other'. It should also be careful to have a very small number of people whose work is assigned to them greater than 200 miles from their homes.

The aggregate trip generation method should produce the great majority of the long distance, vacation trips. Almost all of them should have a destination of type 'Other'. In order to determine the originations of these vacation trips, we could again sample the census data but mainly to determine the income distribution in a given FIPS code. Using that data, a probability can be assigned to each individual in the disaggregate person trip file that determines their likelihood of taking a vacation that day and increases with their income. Using those statistics, we would resample from the disaggregate person trip file to determine which individuals would be taking a vacation on an average day in each FIPS code. This is the end of the disaggregate approach and where the aggregate approach begins.

After sampling from the person trip file to generate the individuals who are taking vacations, the only information that is useful is the coordinate at the center of the FIPS code to which every long distance vacation trip will originate from, and the total number of vacation trips being taken from a FIPS code. Then, we could use hotel data scraped from the internet, rental car data from companies like Hertz, credit card data of customers buying things from cities that they do not live in, data on concert locations, and other people-attracting events to create an attractiveness score for each city in America (or maybe each city in America with a population greater than 50,000 or 100,000 people). Once this matrix is created, every person taking a trip greater than 200 miles can be assigned a destination city using the gravity model aggregate approach. Using these assigned destinations, the vacation trips could be reinserted into the original person trip file (the OriginPixel##### file) or kept separate. At this point, the vacation trip data should be much more realistic. We will only know once someone has tried it.

Conclusion

We used the case study FIPS Code location of Orlando to begin testing the feasibility of an autonomous taxi service to operate as a mode of transportation for riders wishing to travel long distances over short time spans for vacations has yielded particular insights. Firstly, the methodology utilized to model vacation trips has shown that vacation trips using aTaxi's will remain to be somewhat multimodal due to the fact that vacation trips are clustered into a common origin to service common destinations between similar passengers. This means that the use of an aTaxi service will not eliminate the need for public spaces used for aggregation of ridership much in the way that airports, greyhound bus terminals, and amtrak stations function in the present. Moreover, vacation trips will still require a trip to the common origin place for vacation aTaxi trips thereby maintaining the multi modal style of transportation. The simulated aTaxi vacation trip model also contains a 2 hour delay of departure constant which is again very similar to the average wait at an airport as a result of the need for passengers to wait until all similar riders are grouped into their respective vehicles. These designed realities of the vacation trip aTaxi model show that the efficiency increases associated with aTaxi vacation trips are not altogether immediately obvious in comparison to the current system of vacation travel. At the current moment, vacationers already need to find a method of transportation to get to a central hub to aggregate and wait until other passengers traveling with them to roughly the same destination can embark on the journey together.

That being said, the model we have created has signaled that there is considerable demand for vacation trips undertaken by autonomous taxis. This demand is exemplified by the calculated Average Vehicle Occupancy from the model in which there was no maximum occupancy constraint imposed upon the aTaxi's. In fact, even after imposing maximum occupancy constraints on these aTaxi trips traveling over 200 miles, it is clear that a consistent demand for ride sharing for vacation trips or long trips remains. In the face of these positive findings of the vacation trip aTaxi simulation, it is quite obvious that the dataset used to model such an implementation of autonomous ridesharing is very incomplete and/or inaccurate. Thus, while this chapter provides crucial information regarding a model that could be used for aTaxi vacation trips, this data and analysis alone cannot determine the feasibility of implementing such a vacation trip service. Only with more data and, specifically, more holistic data on vacation trips taken in the US will the aTaxi vacation trip service be able to be analyzed in a more realistic manner.