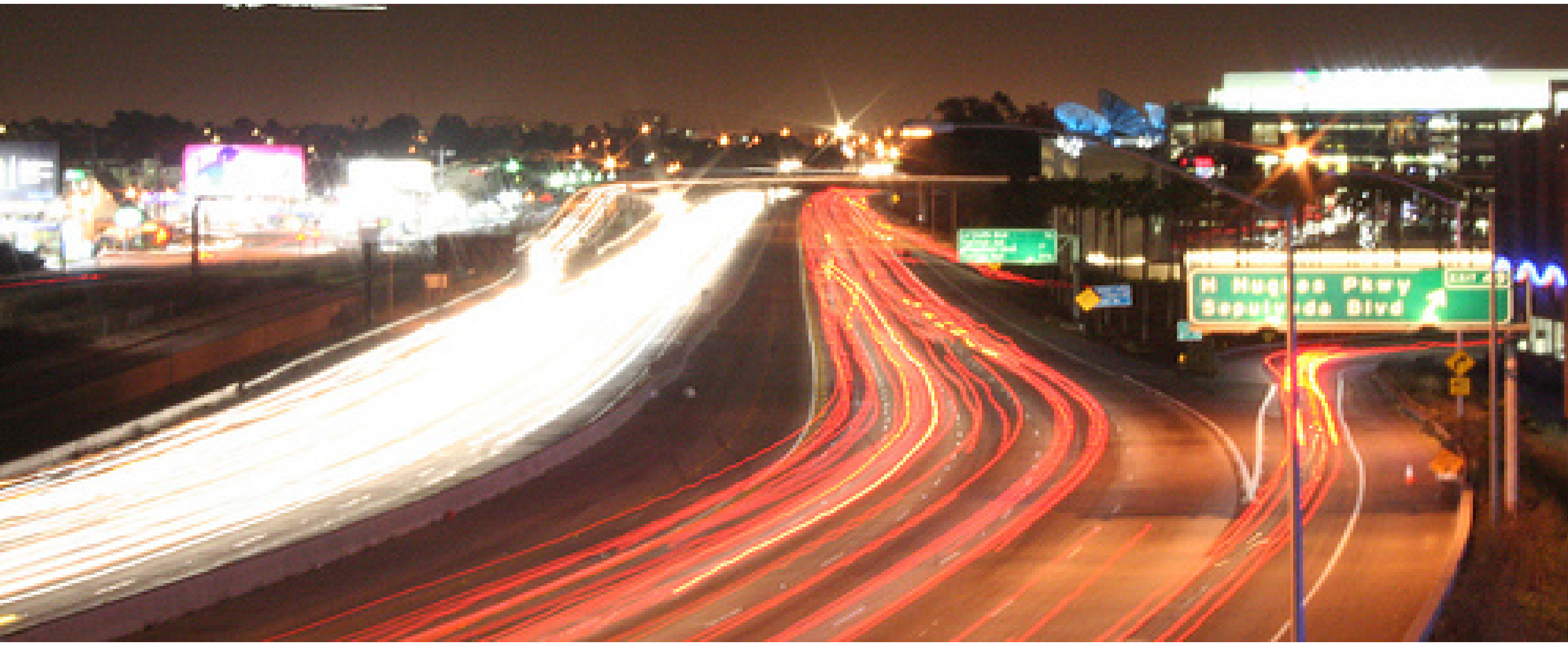
DRAFT COPY

Synthesizing Individual Travel Demand in New Jersey

Trips everyone in NJ wants and needs to make on a typical day



Philip Acciarito '12
Luis Quintero '12
Spencer Stroeble '12
Natalie Webb '12
Heber Delgado-Medrano GS '12
Talal Mufti GS '12
Bharath Alamanda '13
Christopher Brownell '13
Blake Clemens '13
Charles Fox '13
Sarah Germain '13
Akshay Kumar '13
Michael Markiewicz '13
Tim Wenzlau '13

Professor Alain L. Kornhauser

Department of Operations Research & Financial Engineering
Princeton University

January, 2012

DRAFT COPY

PROJECT CONTRIBUTORS

Task 1: Building A New Jersey Resident File	Natalie Webb Luis Quintero
Task 2: Assigning a Work County to each Worker	Akshay Kumar
Task 3: Assigning a Work Place to each Worker	Spencer Stroeble
Task 4: Assigning a School to each Child	Chris Brownell Blake Clemens
Task 5: Assigning a Trip Chain to each Person	Sarah Germain Tim Wenzlau
Task 6: Assigning The Other TripEnds	Charles Fox Michael Markiewicz
Task 7: Assigning a Departure Time to each Trip	Philip Acciarito Heber Delgado-Medrano
Generating Patronage and Employee Shift Time Distributions	Bharath Alamanda
Report Preparation	Bharath Alamanda Sarah Germain
Leadership Team	Talal Mufti

ABSTRACT

In the state of New Jersey, there is a growing need for accurate travel demand data for use in transportation systems analysis. Traditional travel survey techniques are often too expensive and fail to capture key segments of the population. Instead, using data from the US Census and other sources, a population was synthesized that is demographically largely identical to that of New Jersey and forecast the travel needs and desires for each resident in this population on an average weekday. Each resident was assigned key defining features including an age, gender, place of residence, demographic description (i.e. student, worker, retired, etc.), place of employment, and place of education. Using various distributional assumptions on trip chains and behavioral needs and choices, a NJ Trip File was generated that contains an individualized record for every trip each resident makes, detailing precisely where and when each trip originates and where each trip ends. The end result of our project is a data driven, spatial, and temporal process that characterizes the individual demand for travel in New Jersey that can be used for a variety of applications from designing PRT (Personal Rapid Transit) networks to anticipating infrastructure overloads.

DRAFT COPY

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

Everyday almost 9 million citizens of New Jersey and thousands of out of state workers travel through and within the 8,721 square miles that constitute the state of New Jersey. Currently, there is very little sense of the pattern of travel of individuals. Where are they coming from? Where do they want to go on a daily basis? When are they making their trips? By using GPS, tracking people's cell phones, and doing surveys, real life travel patterns can be measured. However, data collection is an expensive process that in the end produces less than comprehensive results. Further, there are limitations on our ability to extrapolate from these small surveys.

As a solution to this problem, our project seeks to synthesize via probabilistic selection a trip file that characterizes travel demand for the entire state of New Jersey. Establishing a framework for synthesizing daily travel demands of every individual in New Jersey is an imperative first step to designing or improving a transportation system that is capable of supporting these demands. In our project, profiles of 8.5 million New Jersey citizens as well as 0.5 million out-of-state workers were generated, providing each individual with a name, age, gender, place of residence, and demographic description (i.e. student, worker, retired, etc.). From there and given certain underlying assumptions about the population dynamics (which we will outline below), trip patterns for every individual in one typical weekday were synthesized, showing every trip each person makes, detailing precisely where and when each trip originates and where and when each trip ends.

Some key statistics from our simulated travel demand file include:
-30,564,582 trips were successfully assigned an origin, destination, departure time, and arrival time on a typical day in New Jersey
-the average New Jersey citizen makes 3.41 trips per day in our synthesis
-the average out-of-state worker makes 2.50 trips per day within the borders of New Jersey
-the average trip was 19.3 miles long
-the average commute to work was 19.1 miles long
-the number of children going to school was 1,605,929 in our simulation, closely matching the estimated 1.5 million children age 5-18 in New Jersey (based on census data)
-the average trip to school was 4.0 miles long

Given our substantial first step in the modeling of trip demand in New Jersey, there is definite room for improving upon our results and collecting more data to justify or modify our key assumptions in the future, making our work even more useful to designing and analyzing transportation systems based on our ability to generate comprehensive and realistic travel demands.

2. INTRODUCTION: OBJECTIVE

The main objective of this project is to obtain a spatial and temporal characterization of travel demand in New Jersey. Using the 2010 US census, data from other sources, and distributional assumptions, a NJ_TripFile that contains an individualized, probabilistic record of the each trip for each resident in New Jersey takes on an average weekday was generated.

3. INTRODUCTION: PURPOSE

The purpose of this project is to take steps toward building a more realistic demand model for use in transportation planning in New Jersey. Besides existing survey techniques, which are both cost and time intensive, our probabilistic approach is one of the leading alternatives to develop a better sense of travel patterns. As more real world data is incorporated into forming underlying assumptions, simulated data should prove increasingly useful in transportation systems analysis. Additionally, simulated data easily lends itself to what-if analysis of travel demand, allowing one to quantify the effects of changes to various parameters and assumptions. The data can also be particularly instrumental in designing new transportation networks since developers will have a detailed understanding of where and when trips are being taken.

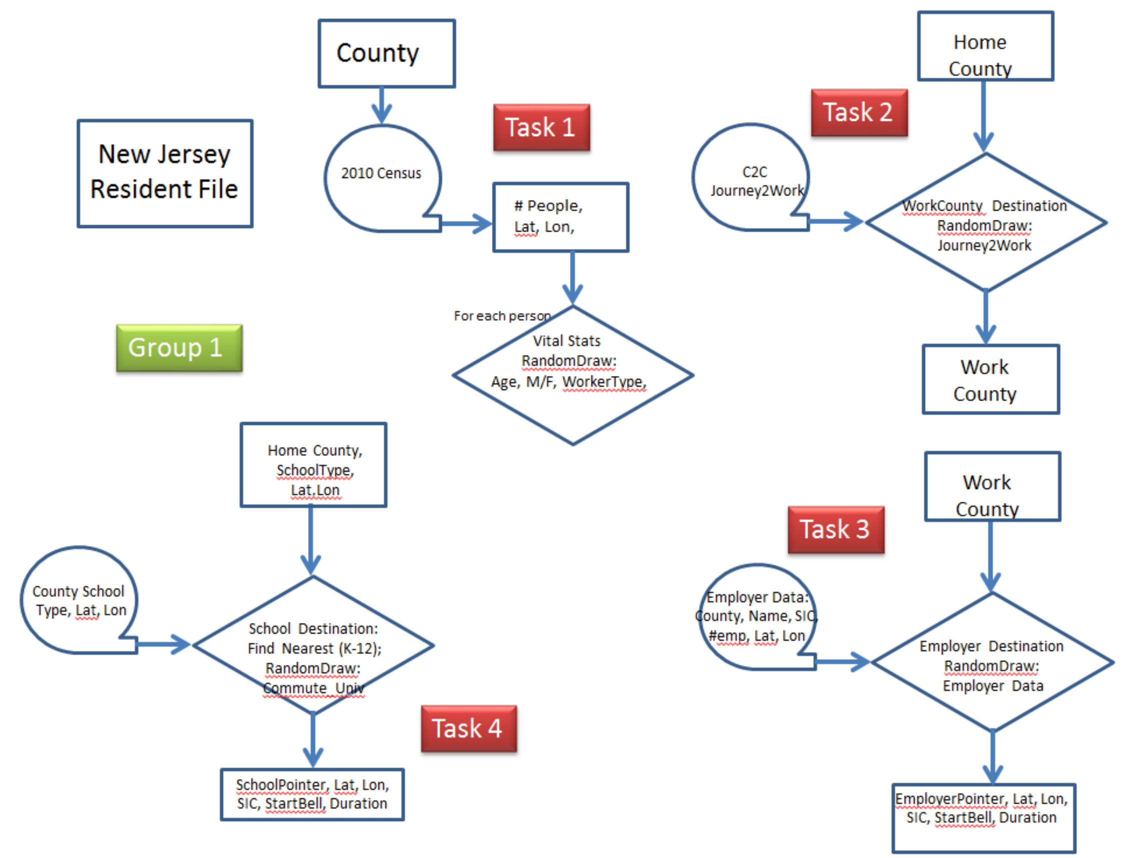
4. INTRODUCTION: PROCESS

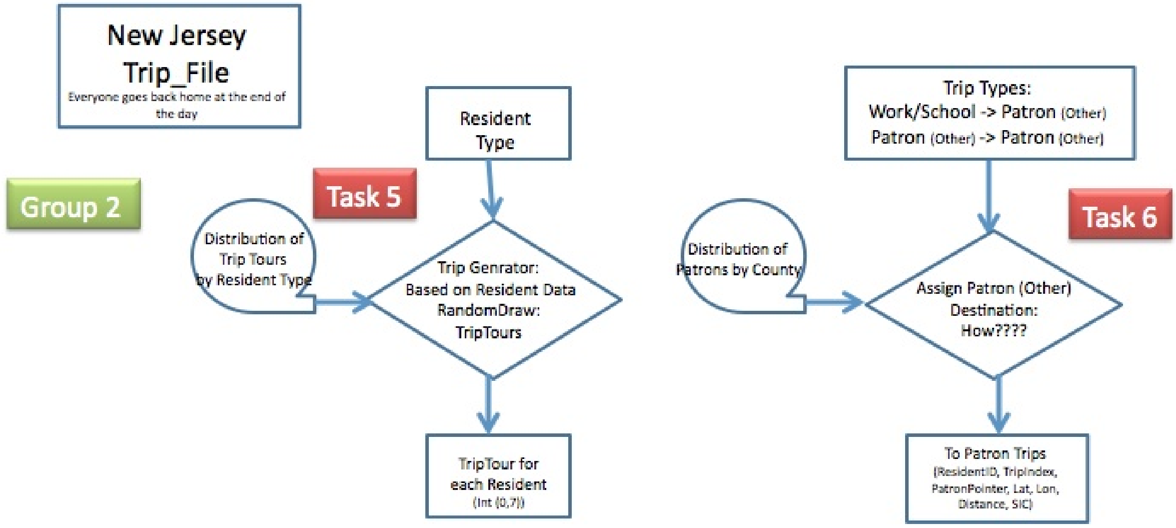
In order to generate a complete look at the trip demand of New Jersey, the building of the NJ_TripFile file was split into 7 sequential tasks. Tasks 1,2,3, and 4 were primarily responsible for recreating the population of New Jersey. Using demographic data on each census block, Task 1 created a NJ_Residents file that contains records for approximately each of the 8.5 million residents who reside in and/or work within the state. Using a random draw of the probability distributions acquired from the census, assigned to each resident were vital statistics such as name, age, gender, home location, and worker type. Worker type roughly corresponds with age and describes the general demographic description for the person with the available choices being 1) Under 5 child, 2)Elementary School Student, 3) Middle School Student, 4) College Commuter, 5) College Student on Campus, 6) Worker, 7)Out-of-State Worker, 8) At Home Worker (which includes stay at home spouses and retired workers), and 9) Nursing Home/Elderly Person. To determine places of employment for residents who were Workers, Task 2 first assigned a work county for them based on census data and Journey to Work data. Once a work county had been identified, Task 3 assigned a specific employer to each resident using the employee distribution for that particular work county. Task 4 assigned a specific school for each person who was a student.

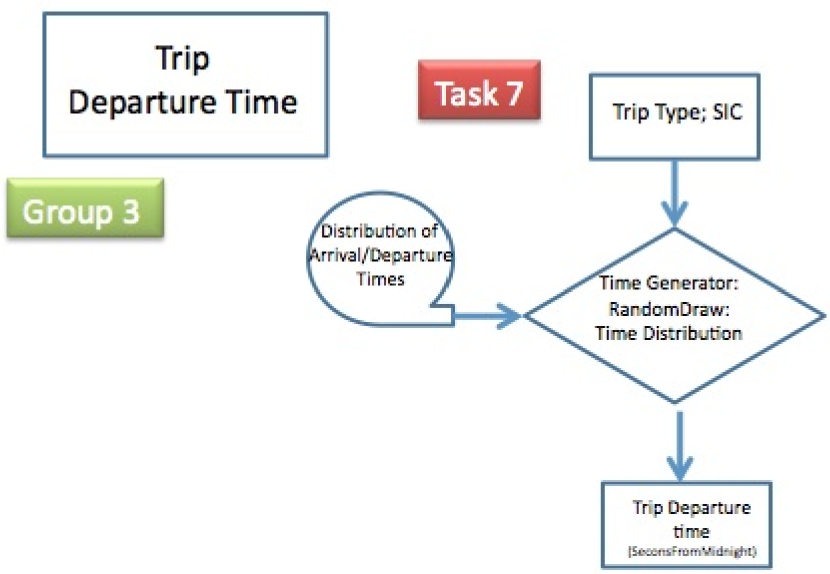
In the next stage of the synthesis, Tasks 5 and 6 were focused on consolidating the information regarding the number of trips taken and the origin and destination of each trip. Task 5 assigned each resident in our simulated population a certain trip chain. The trip chain describes the sequence and purpose of trips that a resident will take on a typical weekday. The trip chain was assigned using a random draw from distributions for each worker type based on assumptions about a reasonable number of trips that a certain type of worker would take in one day (stated in the Task 5 report section). Once each resident has been assigned a trip chain, Task 6 proceeded to append origin and destinations for each trip within a resident's trip chain. For home-to-work, home-to-school trips and their inverses (work-to-home, school-to-home), the locations were already assigned in previous tasks. Task 6, though, had to take particular care in assigning destinations for the (any location)-to-other trips since there were many locations to choose from for the other trips as they encompass attractions as varied as restaurants, shopping malls, and other recreational areas. Particular other location were chosen based on the patronage distribution (i.e. number of patrons visiting on a single day) of available options and the county of the origin location.

After each trip in the trip chains of all 9 million individuals had a origin and destination, the final stage of the project was completed by Task 8. Task 8 appended a departure time and roughly estimated an arrival time for each one of the trip records based on distributions of employee shift times, school start times, and other behavioral assumptions. For non-work, non-school trips (i.e. other trips), the arrival time was used to estimate a departure time for the subsequent trip.

The following flowchart below outlines our process including the inputs, outputs, and mechanism of each task:




New Jersey
Trip_File
Everyone goes back home at the end of the day
Group 2
Task 5
Resident Type
Distribution of Trip Tours by Resident Type
Trip Genrator: Based on Resident Data RandomDraw: TripTours
TripTour for each Resident
(Int (0,7))
Trip Types:
Work/School -> Patron (Other)
Patron (Other) -> Patron (Other)
Task 6
Distribution of Patrons by County
Assign Patron (Other) Destination: How????
To Patron Trips
(ResidentID, TripIndex, PatronPointer, Lat, Lon, Distance, SIC)

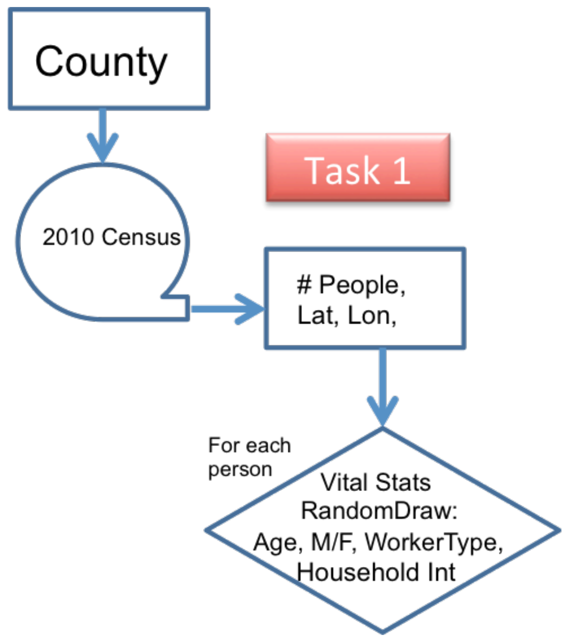

Trip
Departure Time
Group 3
Task 7
Trip Type; SIC
Distribution of Arrival/Departure Times
Time Generator: RandomDraw: Time Distribution
Trip Departure time
(SeconsFromMidnight)

5. TASK 1: BUILDING A NEW JERSEY RESIDENT FILE

1.1 Introduction

1.1.1 Objective

The objective of Task 1 is to generate a population of New Jersey and non-New Jersey residents who work in the state. Using population and location information from the 2010 census and set of input distributions, we generate, for each person in the state, a name, household integer, ID number, age, gender, WorkerType (elementary school, worker, at-home worker, etc.), and location of residence. The objective of the name generation is to generate names for the simulated NJ and out of state commuters that closely resembles the true names of the daily commuters.



1.1.2 Purpose

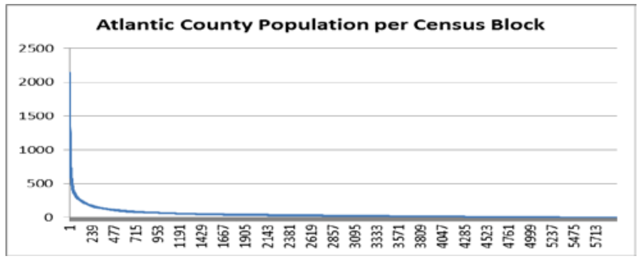
In creating this population for New Jersey, we want to generate information about each person that is necessary and sufficient for later tasks to append reasonably realistic work and school information and trip types. The purpose of generating names for the population is to make our Synthesis one degree more realistic by assigning the commuters individual names, as they have in reality that could be used in place of a simple ID number. Also, generating names allows one to identify the trips of a single person (or household) by referencing name rather than an ID number.

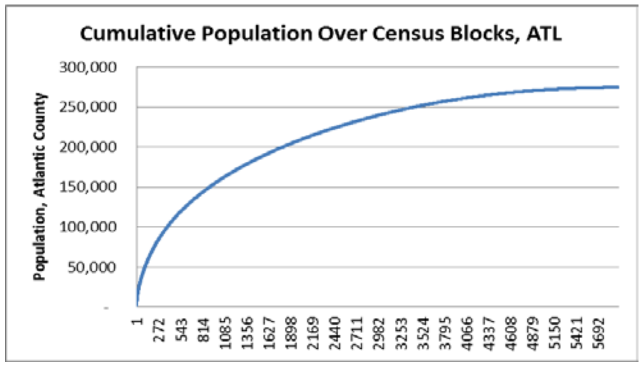
1.2 Process

1.2.1 Input data sets

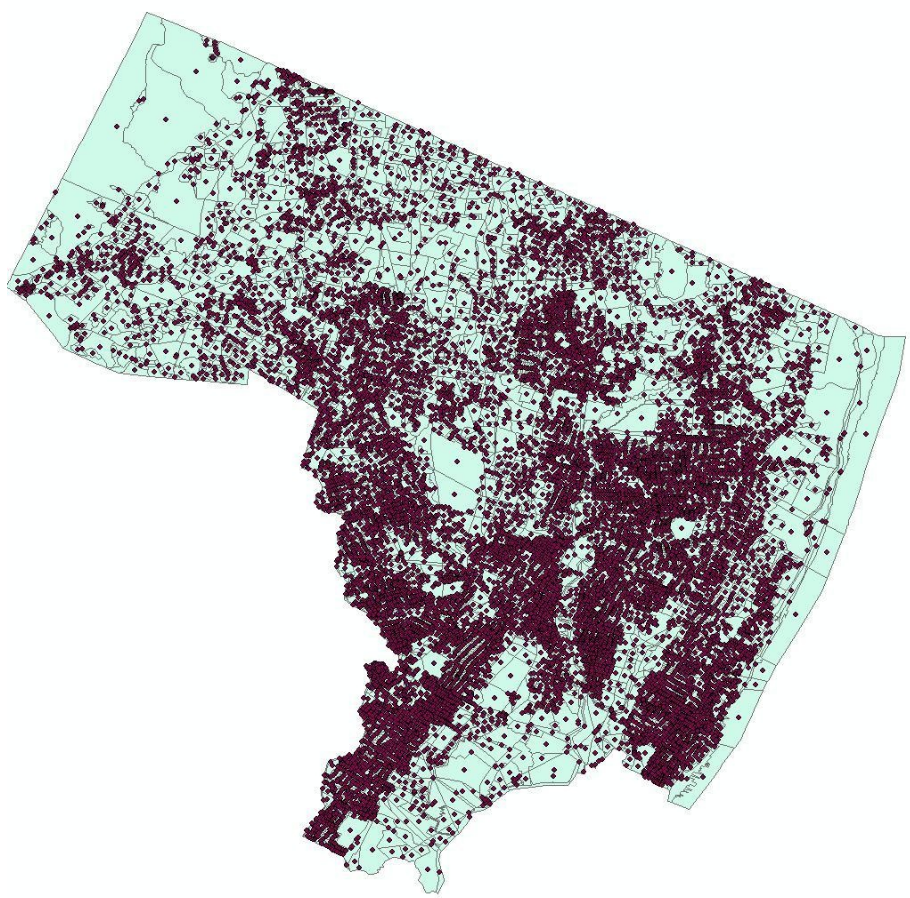
Data from the 2010 census provided the starting point.
http://www.genesys-sampling.com/pages/Template2/site2/61/default.aspx
It has, by county, the centroid and population of each census block - the smallest unit of geography defined by the U.S. Bureau of the Census and is used to report and collect Census Data. A Census Block is a geographic sub-division of a Census Tract and is typically the size of a city block in urban areas and slightly larger in rural areas. New Jersey's 2010 population of 8,791,894 individuals is distributed over 118,654 Census Blocks. The Table below documents New Jersey's population by county, the number of Census Blocks in each county and the median and average values of the distribution of population by Census Block for each county. Because the median values are so much lower than the average value, the distribution of population per block has a very long tail of high values. However, those high values tend to be blocks that are very small in size; thus, the assignment of the centroid of the block as their home location tends to be much more consistent to the location of their "front door" than for the blocks that comprise very few people but encompass a very much larger area.

County	Population	Census Blocks	Median Pop/ Block	Average Pop/Block
ATL	274,549	5,941	26	46
BER	905,116	11,171	58	81
BUR	448,734	7,097	41	63
CAM	513,657	7,707	47	67
CAP	97,265	3,610	15	27
CUM	156,898	2,733	34	57
ESS	783,969	6,820	77	115
GLO	288,288	4,567	40	63
HUD	634,266	3,031	176	209
HUN	128,349	2,277	31	56
MER	366,513	4,611	51	79
MID	809,858	9,845	50	82
MON	630,380	10,067	39	63
MOR	492,276	6,543	45	75
OCE	576,567	10,457	31	55
PAS	501,226	4,966	65	101
SAL	66,083	1,665	26	40
SOM	323,444	3,836	51	84
SUS	149,265	2,998	28	50
UNI	536,499	6,139	61	87
WAR	108,692	2,573	23	42
Total	**8,791,894**	**118,654**		**74.1**





Below is a display of the census block boundaries and their centroids for Atlantic County



The latitude and longitude of the block centroids specified the spatial location of the home of each person and demographic characteristics were assigned probabilistically from distributions assembled various state of New Jersey statistics sources.. (Note that the output information listed is from Atlantic county. Trying to find statistics on the entire nine million people generated was unwieldy and unnecessary for the purposes of this report - we did sanity checks on other counties as well, but did not include the results here.)

Gender:	**Input:**	**Output:**
female	51.3%	51.3%

Ages (varying linearly over interval):	**input:**	**output:**
[0,49]	67.5%	67.5%
[50,64]	18.0%	17.9%
[65,79]	12.0%	12.1%
[80,100]	2.5%	2.5%

The 2010 census gave information about the average household size and number of households consisting of a couple with no children, a couple with children, a single man with no children, a single man with children, a single woman with no children, a single woman with children. It did not include information, however, on how many children were in the household. Starting with this data from Atlantic County, we created the following distribution. Though we had information about the number of households with children present, we had to distribute that number over precise family size and otherwise tweak the distributions so that the probabilities would add to 1.

Household:	**Size:**	**Probability:**	**cdf:**	**Expectation:**
couple	2	0.30	0.300	0.6
couple + 1	3	0.08	0.380	0.24
couple + 2	4	0.06	0.440	0.24
couple + 3	5	0.04	0.480	0.2
couple + 4	6	0.04	0.520	0.24
couple + grandparent:	3	0.01	0.525	0.015
single woman	1	0.16	0.685	0.16
single mom + 1	2	0.07	0.755	0.14
single mom + 2	3	0.05	0.805	0.15
single mom + 3	4	0.03	0.835	0.12
single mom + 4	5	0.03	0.865	0.15
single man	1	0.12	0.985	0.12
single dad + 1	2	0.01	0.990	0.01
single dad + 2	3	0.005	0.995	0.015
single dad + 3	4	0.005	1.000	0.02
				2.42

The expected family size of this distribution is 2.42, lower than the Atlantic county average of 3.17 (and the NJ average of 2.64). In practice, the average family size generated for Atlantic county was 1.88. This discrepancy is partly due to the fact that, when small populations are generated, there is more likely to be difference between the number of men and women, resulting in more single-person households than there might be in reality (for algorithm, see section 1.2.2 below). Also, the algorithm does not account for households with non-family members living together – the bachelor pad of five 20-something guys, for example. And, as previously mentioned, the main problem here is a lack of more precise data.

WorkerType Int	WorkerType String:	Distribution:
0	grade school	100% ages [6,10]
1	middle school	100% ages [11,14]
2	high school	100% ages [15,18]
3	college: commute	distribution given below
4	college: on campus	distribution given below
5	worker	distribution given below
6	at-home worker and retired	at-home dist. given below, 100% ages [65,79]
7	nursing home and under 5	100% ages [0,5] and 100% ages [80,100]

The distribution for workers vs. at-home workers would be conditional on gender. Therefore, we used the following calculations:

$$P\{\text{at-home worker}|\text{female}\} = \frac{P\{\text{female}|\text{at-home worker}\}*P\{\text{at-home worker}\}}{P\{\text{female}\}}$$

$$= 0.97*0.33/.513 = 62.4\%$$

$$P\{\text{worker}|\text{female}\} = 1 - .624 = 37.6\%$$

Doing the corresponding calculations for males, yields the following distribution:

Female worker vs. at home, ages [24,64]	Input:	Output:
worker (5)	37.6%	37.6%
at-home worker (6)	62.4%	62.4%

Male worker vs. at home, ages [24,64]	Input:	Output:
worker (5)	79.7%	79.7%
at-home worker (6)	20.3%	20.3%

The number of at-home males seems high, but when we consider that this also includes unemployed, it might not be too bad. However, one of the improvements that could be made to this task is to find the distribution of worker vs. at-home worker by gender for each county.

The above numbers were used, together with the statistic that 51.3% of college-age students in NJ go to college, and that 86% of college students commute, to generate this distribution:

Female college-age students, ages [19,23]	Input:	Output:
college: commute (3)	44.1%	42.4%
college: on campus (4)	7.2%	7.0%
worker (5)	18.3%	18.4%
at-homeworker (6)	30.4	32.1%

Male college-age students, ages [19,23]	Input:	Output:
college: commute (3)	44.1%	44.6%
college: on campus (4)	7.2%	7.2%
worker (5)	38.8%	38.5%
at-homeworker (6)	9.9%	9.7%

1.2.1.1 Sample input data

From the 2010 census, from Atlantic county:

SUMLEV	STATE	COUNTY	TRACT	BLKGRP	BLOCK	HU100	POP100	INTPTLAT	INTPTLON
101	34	1	10200	1	1000	143	211	39.4394	-74.495
101	34	1	10200	1	1003	0	136	39.4428	-74.496
101	34	1	10200	1	1007	10	23	39.4354	-74.496
101	34	1	10200	1	1008	47	114	39.4365	-74.492
101	34	1	10200	1	1009	18	35	39.4365	-74.49

The column POP100 is the population of a census block, and the INTPTLAT and INTPTLON are the latitude and longitude, respectively, of the centroid of the census block.

The other input data was, as mentioned, various statistics used to create distributions for age, gender, WorkerType, etc.

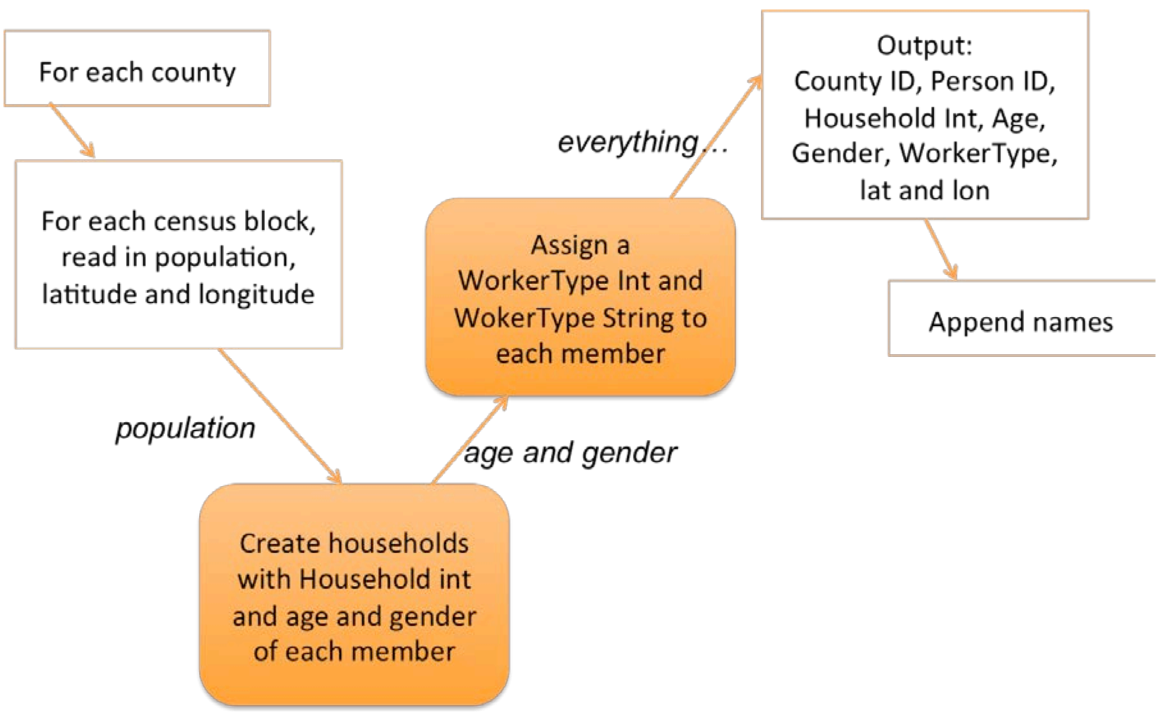
When generating Non-New Jersey counties (for non-residents that work in New Jersey), we only generated single workers between the ages of 22 and 64, and used the following counties and associated latitudes and longitudes:
NYC - New York City - Empire State Building: (40.748716,-73.986171)
PHL - Philadelphia - Ben Franklin statue: (39.952335,-75.163789)
BUC - Bucks County PA and east to CA - Newtown, PA: (40.229275,-74.936833)
SOU - South of Philadelphia - Wilmington DE: (39.745833,-75.546667)
NOR - North of Bucks County in PA - Allentown PA: (40.608431,-75.490183)
WES - Westchester County NY & East - White Plains: (41.033986,-73.76291)
ROC - Rockland and Orange & Rest of NY State - Rockland: (41.148946,-73.983003)

1.2.2 Process

Coding in Python, population, latitude and longitude associated with each census block was read in. We then called a function that generates households, taking in the population as an argument. For each person in the given census block population, we generated, with random number generators and the given input distributions, an age and gender. We separated these realizations into four vectors: children (ages 22 and under), men (ages 23-79), women (ages 23-79), and grandparents (80 and above). After sorting each vector according to age, we then sorted them into buckets and shuffled the entries in the buckets. The purpose for this shuffling was so that when we drew two children for one family, they would have slightly different ages and so that the parents would have slightly different ages from each other but about the right age difference between them and their kids. Using a random number generator, I then used the distribution given above to create families, couples, and single people, giving each household an ID number. If we cycled through all of the adults and there were children left over, if the children were over 18, they were treated as singles, and if under 18 their age was incremented by 10 and they were treated as singles. When there were still men and women left over, we formed couples (probability .75), single men (probability .1), and single women (probability .15). After that, if there were any other people left over, they formed single households. After generating households, we then generated a WorkerType for each person using a random number generator and based on age and gender.

Once we had finished generating the first portion of Task 1, we added the names by using the file from the first portion (without names) as input and allowing a MATLAB program to output the original file with names added to the fourth and fifth columns of the data.

1.2.2.1 Flow chart of complete process

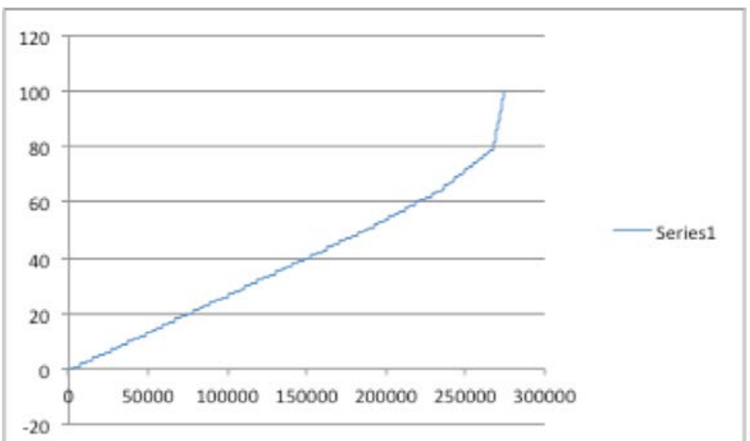


1.2.3 Output data sets

Output distributions have been indicated above (for Atlantic County).

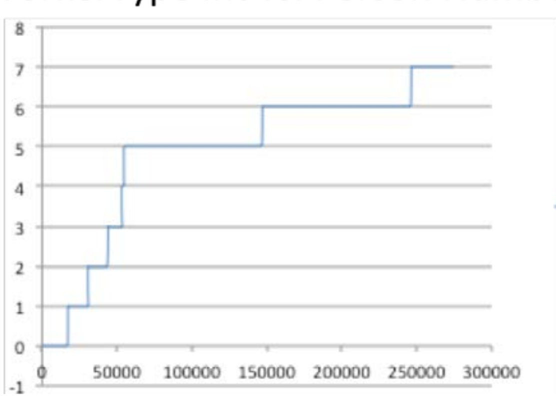
The County ID integer field has integers:
0-20 for NJ counties in alphabetical order
21 New York City (5 boroughs and Long Island) (NYC)
22 Philadelphia (PHL)
23 Bucks County PA and east to California (BUC)
24 South of Philadelphia (SOU)
25 North of Bucks County in PA (NOR)
26 Westchester County NY & East (WES)
27 Rockland and Orange and Rest of New York State (ROC)

Output:
Age vs. Person Number



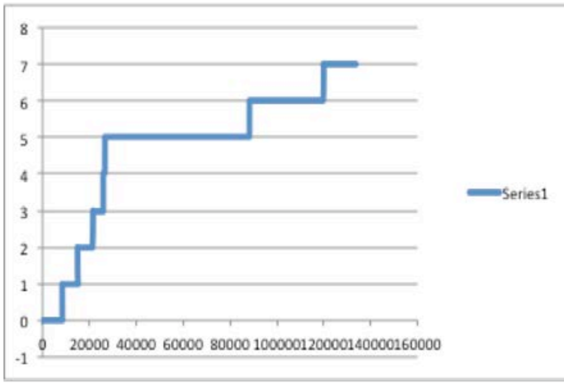
Here we see the expected linear pattern over the intervals [0-49], [50-64], [65-79], and [80-100] with a decrease for older ages.

Output:
WorkerType Int vs. Person Number

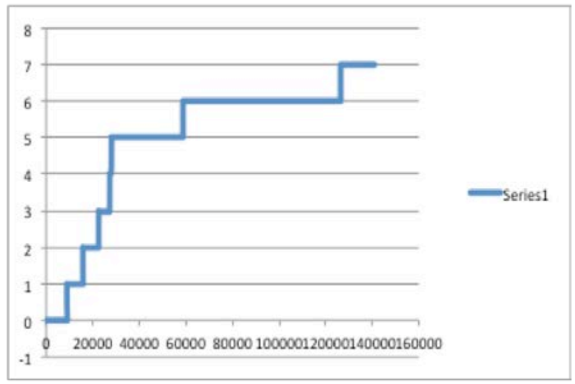


As expected, there are approximately equal numbers of grade school, middle school, and high school students, with a similar number of college-age students split between college: commute (3), college: on campus (4), working (5), and at-home workers (6). There are slightly more at-home workers than workers since the at-home category also includes retirees. There is a fairly small number of elderly/under 5 year-olds.

WorkerType Int vs. Person Number



Males

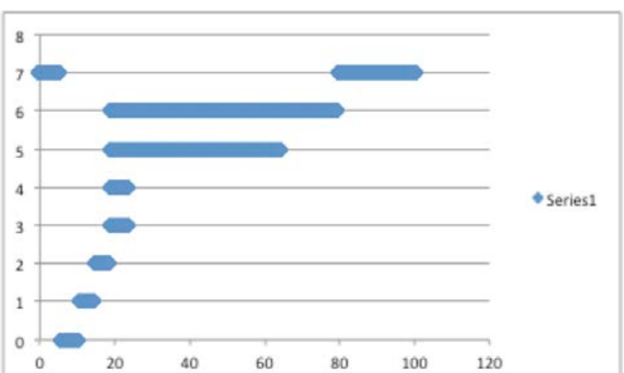


Females

Here we see the effects of this distribution being conditional on gender – far more women fill the at-home worker category.

Output:

WorkerType Int vs. Age



We see the expected relationships between age and WorkerType, especially:

- college-age student being split between college: commute, college: on campus, worker, and at-home worker
- category 6 including both at-home workers and retirees
- category 7 includes both the very young and the very old.

1.2.3.1 Format of output data set(s)

The output is given in csv files titled XXXTask1.csv, with XXX being the first three letters of the county.

Columns of csv files:	Datatype:
County ID	integer
Person ID	integer
Household Int	integer
Last Name	string
First Name	string
Age	integer
Gender	Boolean
WorkerType integer	integer
WorkerType string	string
Latitude of residence	float
Longitude of residence	float

1.2.3.2 Sample output data

County ID	Person ID	Household ID	Last Name	First / MI	Age	Gender	Worker Int	Worker Str	Lat	Long
0	1	1	PREVILLE	RICHARD G.	24	FALSE	5	worker	39.439369	-74.495087
0	2	1	PREVILLE	JACK J.	7	FALSE	0	grade school	39.439369	-74.495087
0	3	1	PREVILLE	CHARLES X.	1	FALSE	7	under 5	39.439369	-74.495087
0	4	2	DEVEREUX	SUE B.	24	TRUE	6	at-home worker	39.439369	-74.495087
0	5	2	DEVEREUX	ANTON P.	2	FALSE	7	under 5	39.439369	-74.495087
0	6	2	DEVEREUX	KATIE S.	6	TRUE	0	grade school	39.439369	-74.495087
0	7	3	WHEDBEE	LINDA C.	26	TRUE	6	at-home worker	39.439369	-74.495087
0	8	4	CARVER	ROBERT Z.	24	FALSE	5	worker	39.439369	-74.495087
0	9	4	CARVER	JENNIFER P.	25	TRUE	6	at-home worker	39.439369	-74.495087
…	…	…	…	…	…	…	…	…	…	…

1.3 Characteristics of one realization of complete output

Run time for the first portion of Task 1 (i.e., not including name generation):
NJ counties: approximately 3 minutes, 45 seconds.
NonNJ counties: approximately 4 seconds

File Lengths:

ATL	274,549
BER	905,116
BUR	448,734
CAM	513,657
CAP	97,265
CUM	156,898
ESS	783,969
GLO	288,288
HUD	634,266
HUN	128,349
MER	366,513
MID	809,858
MON	630,380
MOR	492,276
OCE	576,567
PAS	501,226
SAL	66,083
SOM	323,444
SUS	149,265
UNI	536,499
WAR	108,692
NYC	86,418
PHL	18,586
BUC	99,865
SOU	13,772
NOR	5,046
WES	6,531
ROC	32,737
Total:	9,054,849

1.4 Limitations of Current Results

One of the primary limitations of the current results is in the household algorithm, described in some detail in section 1.2.2. As mentioned there, it results in a lower average household size than expected. Part of this issue is due to limited data on precise household size, but part of it is also the function of the algorithm itself. It fails to account for unrelated persons living together, and, as mentioned, is sensitive to small discrepancies in the numbers of adult men vs. adult women.

The largest obstacle for the name generation process was efficiency. Due to the large number of residents being generated and the large sizes of the name distribution files, algorithm choice is a major factor in making a generator that will work in a reasonable amount of time. With regards to the actual data, there are two main limitations: the independence of first and last names and the choice of using New Jersey-specific names. The first limitation is a result of the data sets available. Since first names and last names are in three different files (male and female first names are separate) with no reference to joint distributions, there was no way of using the Census data files and creating correlated first-last name choices. Secondly, the name files used were of all of the United States and not specifically New Jersey. As a result the names generated would probably resemble a sample of United States citizens and less of New Jersey commutes (although, there will be some interplay since New Jersey is a state of the United States). Some ways to better these methods are explained in the next section.

1.5 Suggestions for Future Efforts

As mentioned above, refining the household selection would be a significant improvement to this task. Getting more precise data (and by county), and then rewriting the algorithm to account for non-family members living together, etc. One option would be to not generate any singles while there are still children available, but that wouldn't be very accurate since not all singles are late-middle age, which is what that would generate. Another option would be to call the household algorithm with larger populations (perhaps the entire county) so there is a smaller chance of a discrepancy between the number of men and women, then grab enough households to fill a given census block.

The input distributions assume that each county in the NJ has the same characteristics. Finding distribution information by county and using it as input could improve the precision of this project. Also, while we found the worker vs. at-home worker distribution by gender using Bayes' Theorem, more precise data can be found by county.

Another project would be to get more specific location information for residences. Once the housing algorithm is refined, one could also find the area of a census block and distribute the houses over that area. One could assume that the census block is circular and locate the houses on its perimeter.

To better the name generation process, there are two main changes in methodology that could be used in the future to better simulate names for the purposes of this project. The first change addresses the generation of New Jersey-specific names. This can be accomplished by "scraping" an online phonebook website of New Jersey to gather names of real New Jersey residents. Then one could use these as the population of names for simulated commuters. This method could eliminate the need of separating last and first name generation if one uses the first and last name pairs. If one does not want to eliminate that separation, one just has to separate last and first names in the "scraping" *and* separate first names by gender (which could prove to be more difficult). The second change addresses the independence of first and last names. To better the process, one can use a relationship variable, for example a statistically common race associated with a last name, to correlate first and last names given that relationship variable. Having done some searching, I do know there are lists of baby first names available separated by gender for a specific race. Obviously there may be better relationship variables that can be used to correlate first and last names as well.

6. TASK 2: ASSIGNING WORK COUNTY TO WORKERS

2.1 Introduction

2.1.1 Objective

The objective of this task was to generate work counties for each entry of the trip file.

2.1.2 Purpose

The purpose of this task was to set the groundwork for the other tasks to generate data (such as workplace, trip-type, etc.) based on where each NJ resident worked.

2.2 Process

2.2.1 Input data sets

The program takes three inputs: Home-base Journey to Work (HJ2W) census data, Work-based Journey to Work (WJ2W) census data, and the NJ_Resident file. Here is a sample of the HJ2W:
http://www.census.gov/population/www/cen2000/commuting/files/2KRESCO_NJ.xls

34,1,6162,560,Atlantic Co. NJ,6,59,4472,5945,Orange Co. CA,12
34,1,6162,560,Atlantic Co. NJ,6,85,7362,7400,Santa Clara Co. CA,9
34,1,6162,560,Atlantic Co. NJ,10,3,6162,9160,New Castle Co. DE,175
34,1,6162,560,Atlantic Co. NJ,10,5,9999,9999,Sussex Co. DE,9

Using this data, the Task 2 program is able to compute conditional probabilities for each work county for all NJ residents (for more details, see section 2.2.2). However, the HJ2W census data does not include non-NJ residents who work in NJ, so these data had to be supplemented by the WJ2W. An example of the WJ2W census data is shown below:
http://www.census.gov/population/www/cen2000/commuting/files/2KWRKCO_NJ.xls

6,37,4472,4480,Los Angeles Co. CA,34,1,6162,560,Atlantic Co. NJ,33
6,65,4472,6780,Riverside Co. CA,34,1,6162,560,Atlantic Co. NJ,7
9,3,*,*,Hartford Co. CT,34,1,6162,560,Atlantic Co. NJ,5
9,5,*,*,Litchfield Co. CT,34,1,6162,560,Atlantic Co. NJ,4

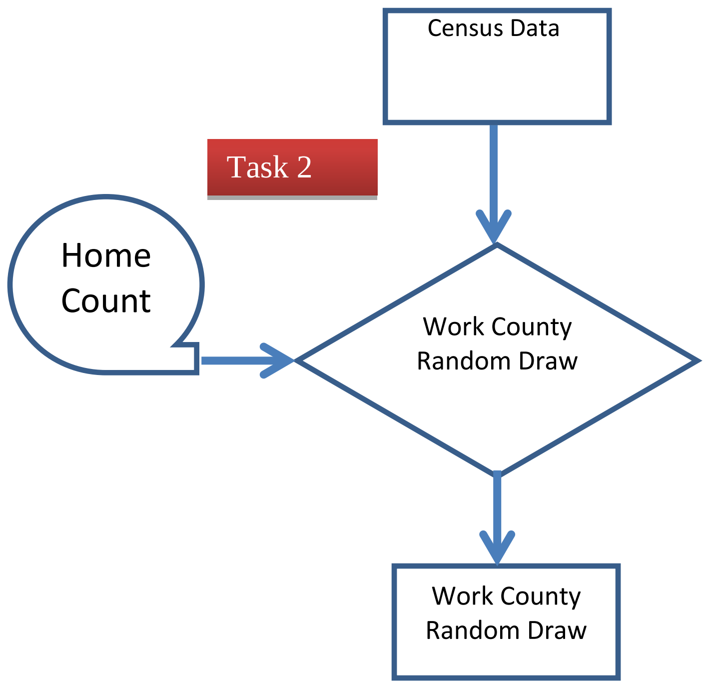
Notice here that the first values (the state codes) are numbers other than 34, signifying non-NJ states. Thus, both census data files provide the program with the information required to generate the underlying probability distribution of the counties.

The final input data file is the output of Task 1, which contains all the residential information for each person in the trip file. The Task 2 program appends a work county to each entry of this input file. A sample of the Task 1 output file is shown below:

1,1,14,f,1,middle school,41.0384561,-74.125712
1,2,48,m,5,worker,41.0384561,-74.125712
1,3,35,m,5,worker,41.0384561,-74.125712
1,4,74,m,6,retired,41.0384561,-74.125712
1,5,44,m,5,worker,41.0384561,-74.125712

The program takes the first value (the numerical representation of the home county) and uses it to generate the work county.

2.2.2 Flow chart of complete process



Briefly, the program has three main steps: data collection and standardization, probability distribution calculation, and work county generation.

Data Collection and Standardization

The program first reads in both census data files, and stores the number of people in each home-county/work-county pair in a matrix where the row number represents the home county and the column number represents the work county. The census state and county numbers are parsed into a uniform set of numbers from 0-27, which make up the indices of the matrix. NJ counties are numbered 0-20, and all other locations are sorted into arbitrary buckets (i.e. "virtual counties") numbered 21-27.

Probability Distribution Calculation

Each row of this "count matrix" is then divided by the row sum. This produces the probability distribution of the work county conditioned on the home county. Adding all of the numbers in a row behind a given entry yields the conditional cumulative distribution.

Work County Generation

Now, the program turns to the input file. It first reads a line of the input file and gets the integer representation of the home county. Then, it goes to the corresponding row in the conditional cumulative distribution matrix and generates a uniform random variable from 0 to 1. Finally, it chooses the work county that from the cumulative distribution matrix that matches the uniform random variable. It then appends this work county integer to the input file and moves on to the next line, repeating the process.

2.2.3 Output data sets

The format of the data set is the same as the input, with one additional entry (the work county).

Sample Output Data

This is a sample line of output from the program. The underlined part is generated by the Task 2 program.

1,1,14,f,1,middle school,41.0384561,-74.125712, 1

2.3 Characteristics of one realization of complete output

The aggregated output of the program matches the underlying census distribution very well. The matrix below is the absolute value of the difference between the input distribution and the output distribution (obtained by running each home county 100,000 times and storing the results). No difference is greater than 3% from the original distribution, which suggests that the program will generate work counties in a way that will closely reflect the NJ census data.

0.0064	0.0000	0.0018	0.0026	0.0006	0.0007	0.0002	0.0021	0.0001	0.0000	0.0009	0.0003	0.0004	0.0002	0.0017	0.0000	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0.0023	0.0012	0.0002	0.0000	0.0000	0.0001
0.0000	0.0347	0.0000	0.0000	0.0000	0.0000	0.0012	0.0000	0.0000	0.0000	0.0003	0.0001	0.0005	0.0023	0.0002	0.0042	0.0000	0.0009	0.0004	0.0007	0.0001	0.0210	0.0001	0.0002	0.0003	0.0016	0.0014	0.0012
0.0006	0.0001	0.0207	0.0030	0.0000	0.0008	0.0004	0.0008	0.0001	0.0000	0.0040	0.0001	0.0008	0.0002	0.0008	0.0000	0.0001	0.0007	0.0000	0.0002	0.0000	0.0009	0.0104	0.0056	0.0013	0.0001	0.0000	0.0001
0.0023	0.0006	0.0052	0.0198	0.0000	0.0009	0.0001	0.0007	0.0001	0.0000	0.0012	0.0003	0.0002	0.0005	0.0003	0.0000	0.0005	0.0002	0.0000	0.0004	0.0001	0.0009	0.0170	0.0050	0.0012	0.0001	0.0000	0.0001
0.0043	0.0002	0.0007	0.0014	0.0045	0.0006	0.0000	0.0003	0.0000	0.0000	0.0001	0.0005	0.0001	0.0001	0.0005	0.0000	0.0002	0.0001	0.0001	0.0001	0.0000	0.0002	0.0031	0.0035	0.0005	0.0002	0.0000	0.0000
0.0039	0.0000	0.0006	0.0005	0.0001	0.0068	0.0002	0.0016	0.0001	0.0000	0.0004	0.0004	0.0001	0.0001	0.0002	0.0001	0.0040	0.0001	0.0001	0.0006	0.0001	0.0001	0.0011	0.0004	0.0014	0.0000	0.0000	0.0000
0.0000	0.0020	0.0001	0.0000	0.0000	0.0000	0.0196	0.0001	0.0000	0.0002	0.0003	0.0018	0.0002	0.0024	0.0001	0.0024	0.0000	0.0011	0.0001	0.0052	0.0002	0.0103	0.0001	0.0004	0.0007	0.0003	0.0003	0.0003
0.0004	0.0006	0.0025	0.0047	0.0002	0.0006	0.0001	0.0109	0.0002	0.0000	0.0006	0.0000	0.0003	0.0001	0.0002	0.0000	0.0003	0.0000	0.0000	0.0000	0.0001	0.0004	0.0114	0.0056	0.0014	0.0002	0.0000	0.0000
0.0000	0.0102	0.0000	0.0000	0.0000	0.0001	0.0007	0.0000	0.0170	0.0000	0.0005	0.0017	0.0006	0.0016	0.0003	0.0017	0.0000	0.0000	0.0001	0.0022	0.0001	0.0220	0.0000	0.0003	0.0005	0.0004	0.0008	0.0001
0.0001	0.0013	0.0002	0.0001	0.0000	0.0000	0.0023	0.0000	0.0002	0.0134	0.0003	0.0001	0.0000	0.0014	0.0001	0.0007	0.0000	0.0062	0.0002	0.0033	0.0019	0.0005	0.0005	0.0036	0.0006	0.0001	0.0001	0.0001
0.0000	0.0003	0.0010	0.0001	0.0000	0.0000	0.0000	0.0001	0.0001	0.0003	0.0081	0.0005	0.0001	0.0004	0.0001	0.0003	0.0000	0.0014	0.0002	0.0009	0.0001	0.0040	0.0001	0.0030	0.0003	0.0003	0.0000	0.0001
0.0002	0.0014	0.0000	0.0001	0.0000	0.0000	0.0016	0.0001	0.0003	0.0000	0.0003	0.0125	0.0020	0.0027	0.0007	0.0000	0.0000	0.0009	0.0001	0.0055	0.0000	0.0061	0.0001	0.0007	0.0005	0.0001	0.0000	0.0001
0.0001	0.0007	0.0005	0.0000	0.0000	0.0000	0.0012	0.0001	0.0017	0.0005	0.0018	0.0003	0.0064	0.0007	0.0012	0.0000	0.0000	0.0005	0.0001	0.0003	0.0003	0.0063	0.0003	0.0007	0.0001	0.0002	0.0000	0.0001
0.0000	0.0032	0.0000	0.0000	0.0000	0.0000	0.0042	0.0000	0.0010	0.0001	0.0001	0.0002	0.0002	0.0012	0.0002	0.0007	0.0000	0.0029	0.0001	0.0004	0.0008	0.0038	0.0000	0.0004	0.0003	0.0006	0.0002	0.0003
0.0019	0.0003	0.0003	0.0004	0.0003	0.0001	0.0010	0.0001	0.0002	0.0000	0.0003	0.0010	0.0002	0.0001	0.0003	0.0002	0.0000	0.0002	0.0001	0.0018	0.0004	0.0014	0.0000	0.0002	0.0003	0.0002	0.0001	0.0003
0.0000	0.0211	0.0001	0.0001	0.0000	0.0000	0.0022	0.0000	0.0002	0.0000	0.0001	0.0009	0.0002	0.0021	0.0002	0.0133	0.0000	0.0006	0.0002	0.0000	0.0004	0.0055	0.0001	0.0004	0.0002	0.0000	0.0002	0.0005
0.0008	0.0000	0.0014	0.0021	0.0004	0.0040	0.0003	0.0040	0.0001	0.0003	0.0003	0.0006	0.0000	0.0000	0.0001	0.0000	0.0018	0.0000	0.0000	0.0001	0.0002	0.0000	0.0008	0.0024	0.0074	0.0001	0.0001	0.0000
0.0000	0.0004	0.0003	0.0002	0.0000	0.0000	0.0030	0.0001	0.0017	0.0003	0.0024	0.0049	0.0003	0.0005	0.0001	0.0000	0.0000	0.0059	0.0002	0.0042	0.0000	0.0018	0.0005	0.0002	0.0005	0.0003	0.0001	0.0004
0.0000	0.0067	0.0001	0.0001	0.0000	0.0000	0.0028	0.0000	0.0004	0.0002	0.0002	0.0010	0.0010	0.0026	0.0005	0.0018	0.0000	0.0003	0.0063	0.0004	0.0008	0.0024	0.0001	0.0005	0.0006	0.0004	0.0002	0.0010
0.0000	0.0031	0.0004	0.0001	0.0001	0.0001	0.0071	0.0001	0.0005	0.0001	0.0002	0.0010	0.0007	0.0001	0.0002	0.0002	0.0000	0.0019	0.0001	0.0067	0.0003	0.0028	0.0000	0.0006	0.0006	0.0001	0.0002	0.0001
0.0002	0.0024	0.0001	0.0002	0.0000	0.0000	0.0001	0.0001	0.0002	0.0068	0.0009	0.0015	0.0001	0.0047	0.0003	0.0001	0.0000	0.0004	0.0003	0.0010	0.0107	0.0007	0.0000	0.0039	0.0001	0.0001	0.0001	0.0002
0.0004	0.0138	0.0008	0.0002	0.0001	0.0002	0.0032	0.0004	0.0002	0.0005	0.0007	0.0023	0.0007	0.0030	0.0001	0.0029	0.0000	0.0025	0.0002	0.0048	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0019	0.0006	0.0058	0.0069	0.0011	0.0008	0.0009	0.0018	0.0010	0.0005	0.0056	0.0010	0.0008	0.0002	0.0002	0.0004	0.0003	0.0026	0.0001	0.0002	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0011	0.0015	0.0031	0.0044	0.0002	0.0000	0.0001	0.0027	0.0005	0.0017	0.0009	0.0013	0.0005	0.0018	0.0003	0.0003	0.0003	0.0030	0.0024	0.0006	0.0064	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0016	0.0044	0.0033	0.0036	0.0007	0.0010	0.0033	0.0026	0.0010	0.0012	0.0001	0.0018	0.0010	0.0019	0.0008	0.0013	0.0073	0.0026	0.0002	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0006	0.0186	0.0001	0.0008	0.0005	0.0008	0.0033	0.0002	0.0001	0.0005	0.0018	0.0046	0.0022	0.0083	0.0011	0.0035	0.0004	0.0017	0.0005	0.0023	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0000	0.0216	0.0002	0.0001	0.0000	0.0004	0.0002	0.0000	0.0044	0.0002	0.0004	0.0028	0.0004	0.0023	0.0000	0.0053	0.0000	0.0003	0.0005	0.0050	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.0004	0.0247	0.0002	0.0001	0.0000	0.0001	0.0030	0.0001	0.0023	0.0003	0.0000	0.0016	0.0003	0.0044	0.0010	0.0054	0.0000	0.0007	0.0037	0.0020	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

7. TASK 3: ASSIGNING A WORKPLACE TO EACH WORKER

3.1 Introduction

3.1.1 Objective

The objective of task 3 is to assign each worker a place of employment. I attempt to estimate a probability distribution of employers in each county. Then, using the percentage of total employees in the county, I will use my probability distribution to synthesize employers for each worker.

3.1.2 Purpose

One of the most essential times of the day for urban planners and transportation experts is the morning rush hour. Every worker is coming in to work at roughly the same time, putting heavy stress on the system. Before we can begin to get to rush hour congestion, we must first understand the demand for trips of the workers of New Jersey. Once we get a better idea of the demand for trips, we can begin to formulate solutions that increase the utility of transportation for all.

3.2 Process

3.2.1 Input data sets

In order to formulate our assignment of employers, we must have two files as input: the resident file with work county appended and the data including all businesses located in New Jersey. The resident file is produced in Task 1 and added to in task 2. Task 2 is the essential step in the chain, as once I know the work county of a given worker, I can then sample my distribution of employers to assign his place of work. The other essential input is our business data. We have as input a file listing all the businesses for each county in New Jersey, including information like id, name, latitude, longitude, number of employees, SIC and NAICS codes.

3.2.1.1 Sample input data

Some sample input data for the residents file appears as follows:

0 1 1 PREVILLE RICHARD G. 24 FALSE 5 worker 39.439369 -74.495087 22
0 2 1 PREVILLE JACK J. 7 FALSE 0 grade school 39.439369 -74.495087 7
0 3 1 PREVILLE CHARLES X. 1 FALSE 7 under 5 39.439369 -74.495087 0
0 4 2 DEVEREUX SUE B. 24 TRUE 6 at-home worker 39.439369 -74.495087 0
0 5 2 DEVEREUX ANTON P. 2 FALSE 7 under 5 39.439369 -74.495087 0
0 6 2 DEVEREUX KATIE S. 6 TRUE 0 grade school 39.439369 -74.495087 0
0 7 3 WHEDBEE LINDA C. 26 TRUE 6 at-home worker 39.439369 -74.495087 0
0 8 4 CARVER ROBERT Z. 24 FALSE 5 worker 39.439369 -74.495087 0
0 9 4 CARVER JENNIFER P. 25 TRUE 6 at-home worker 39.439369 -74.495087 9
0 10 5 TINSLEY ELLEN U. 23 TRUE 4 college: on campus 40.856461 -74.197833 0

The column headings for this input file are {Home county, ID, Household, Last Name, First Name and Middle Initial, Age, Gender, Worker Type, Home Latitude, Home Longitude, and Work County}.

The input file for businesses in a county appears as follows (several other data characteristics are available, but these are not necessary for further tasks):
Name County SIC Code SIC Description
1 VIP SKINDEEP Atlantic 729963 Massage
10 Acres Motel Atlantic 701101 Hotels & Motels
1001 Grand Street Investors Atlantic 679999 Investors NEC
1006 S Main St LLC Atlantic 651301 Condominiums
11th Floor Creative Group Atlantic 781205 Motion Picture Producers & Studios
123 Cab Co Atlantic 412101 Taxicabs & Transportation Service
123 Junk Car Removal Atlantic 593215 Junk-Dealers
1400 Bar Atlantic 581301 Bars
1-800-Got-Junk? Atlantic 495326 Junk Removal

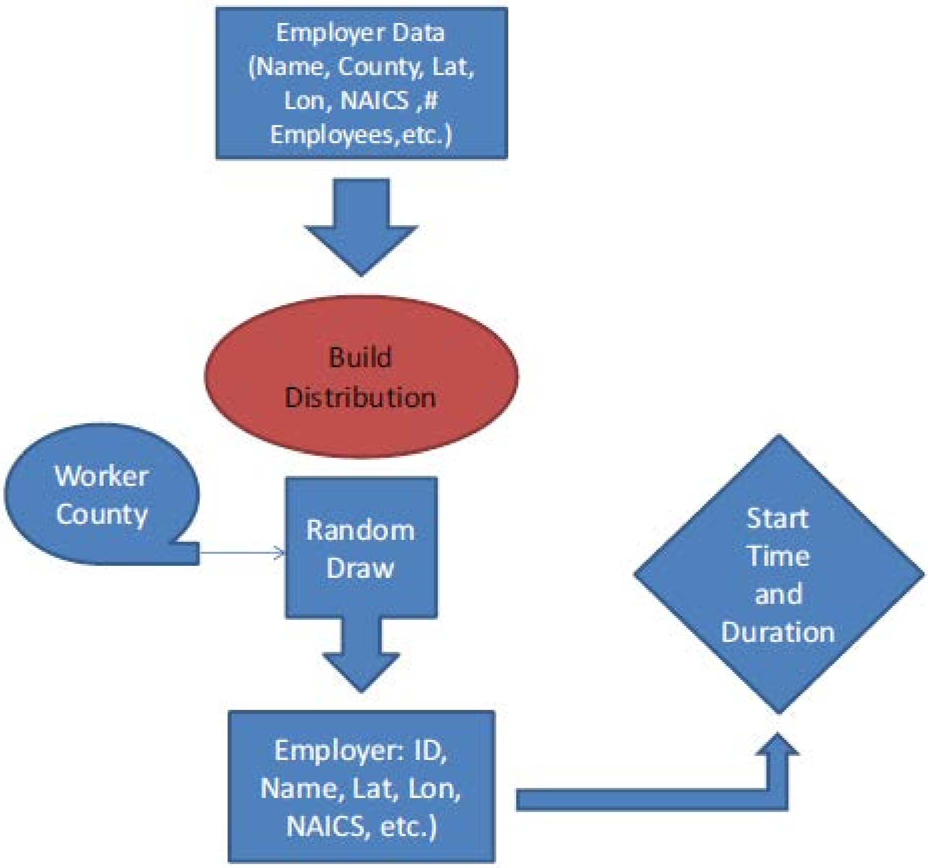
NAICS Code NAICS Description Employment Latitude Longitude
81219915 Other Personal Care Svcs 2 39.401104z -74.514228
72111002 Hotels & Motels Except Casino Hotels 2 39.437305 -74.485488
52399903 Misc Financial Investment Activities 3 39.619732 -74.786654
53111004 Lessors Of Residential Buildings 5 39.382399 -74.530785
51211008 Motion Picture & Video Production 2 39.359014 -74.430151
48531002 Taxi Svc 2 39.3916 -74.521715
45331021 Used Merchandise Stores 2 39.361705 -74.435779
72241001 Drinking Places Alcoholic Beverages 4 39.411266 -74.570083
56221910 Other Non-Hazardous Waste Disposal 4 39.423954 -74.557892

3.2.2 Process

The process of assigning work places is a multi-step process. My process takes the following steps:

1) Read in the file containing business information for each county. Create a new _le including only necessary information (ID#, Name, Latitude, Longitude, SIC Code, SIC Description, NAICS Code, NAICS Description, # of Employees) and append the nearest NJ Transit station along with its coordinates.
2) Create a file for the distribution of the employees for each county. For each business with n employees, write the ID n times.
3) Read through the residential files. Use the work county of each worker to pick the distribution from which to select the employer. For each worker, append necessary employer information, including distance from home to work. Assign each worker a start time and duration from the distribution specified by the employer NAICS code.

3.2.2 Flowchart of Complete Process



3.2.3 Output data sets

3.2.3.1 Format of output data set(s)

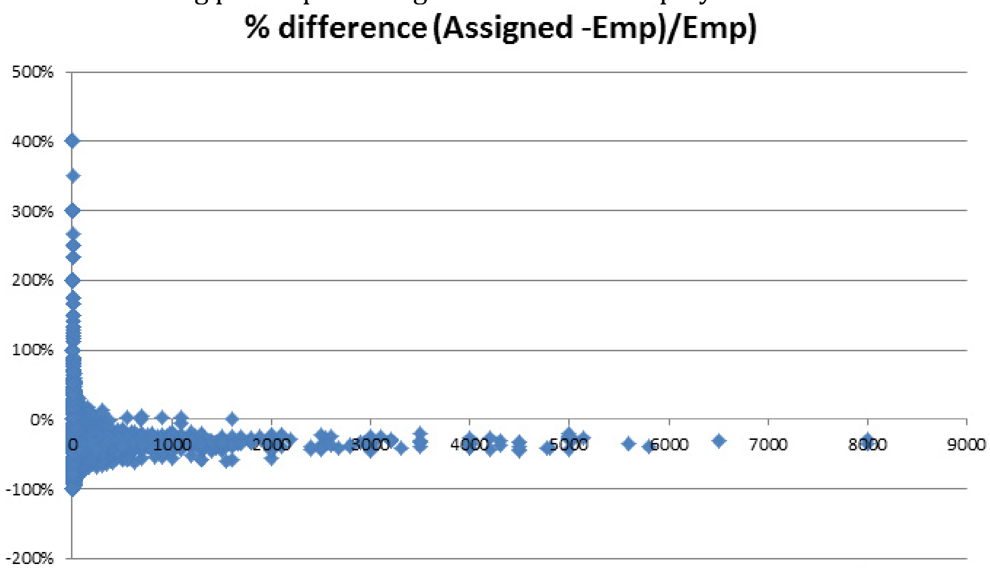
The output takes the form of employer information appended onto the residential files for all workers. We include a pointer to the employer on a list of all businesses in the state, the name of the employer as well as its coordinates, SIC and NAICS codes and descriptions, the distance from home to work, and a start and end time for work.

3.2.3.2 Sample output data

Res Info	County	ID	Name	Lat	Lon	SIC	SIC Des	NAICS	NAICS Des	Dist	Start	End
*	22	0		39.95	-75.16	0	0	0	0	50.3	31162	59657
*	0	10101	Seaview Golf	39.45	-74.47	701	Resorts	721	All Other	20.1	31392	63101
*	0	11777	Univ Chimn	39.37	-74.42	174	Chimney	238	Masonry	15.2	30903	59765
*	12	6823	Shell	40.29	-74.00	435	Gasoline	567	Gasoline	48.9	30222	61295
*	14	1107	Target	39.17	-74.78	365	Retail	456	Retail	39.8	28626	57389

3.3 Characteristics of one realization of a complete output

The larger the number of employees an employer has, the closer the synthesized employment matches the actual employment figures. This makes sense, as a small employment number can vary by a large percentage even if employment differs by only a few employees. This effect can be seen on the following plot of percentage difference vs. employment:



Also, the number of workers in our residential differs significantly from the employment statistics offered in the employer file. The employer file indicates a total of 4,254,762 employees, while we have 2,840,611 workers in our residential file. This is a difference of about 67%, a large deviation. This indicates that we may have made mistakes in determining our distribution of worker types in our residential files.

3.4 Limitations of Current Results

Our current results are primarily oriented towards full time workers. It does not include part time workers who may also be attending school. Also, the deviation of employment figures from our number of workers in the residents file indicates that there may be mistakes in our distribution of worker types. There are also issues with our database of employers in New Jersey. Duplicate records abound, as well as some employee statistics that do not seem correct. Given our data resources, we have done an effective job of allocating workers to employers. However, a more reliable set of data would produce much more realistic results.

3.5 Suggestions for Future Efforts

In the future, we could search for a more reliable database of employer information from which to create our employment distributions. Furthermore, adding the capability to allow individuals to be both students and workers would bring our Synthesis closer to that of the real world. I am currently working on more analysis that will be useful in judging the effectiveness of our worker allocation. I will be mapping the home locations of synthesized employees of familiar businesses to gauge the characteristics of workers that we assign to businesses. This analysis will allow us to better understand our results and identify areas for improvement.

8. TASK 4: ASSIGNING A SCHOOL TO EACH CHILD

4.1 Introduction

4.1.1 Objective

The third task in creating the Garden State's daily trip file is the important job of making sure that each young person has a school to attend. Selecting a school for one's children is often a matter of great importance to families, and the options that each student has reflect that. Of course there are the New Jersey public schools, which are consistently rated near the top of the nation, but there are also hundreds of private schools throughout the state, and the option of homeschooling is growing in popularity in a state that was once vehemently opposed to the idea.

The objective of this task is to assign a school to every student, including those at university. In so doing, it is imperative that we adequately mirror the real-life distributions of students at public and private schools, and the recorded enrollments of the schools in the state.

4.1.1 Purpose

The purpose of this task is to add more specialized attributes to the data generated in Tasks 1 and 2. The school decision is more specialized because it depends upon the data generated in Task 1, as well as upon real-life distributions. The school-specific data generated in Task 4 will play a major role in the final trip file, as more than ninety percent of students travel to their school each day before they go anywhere else.

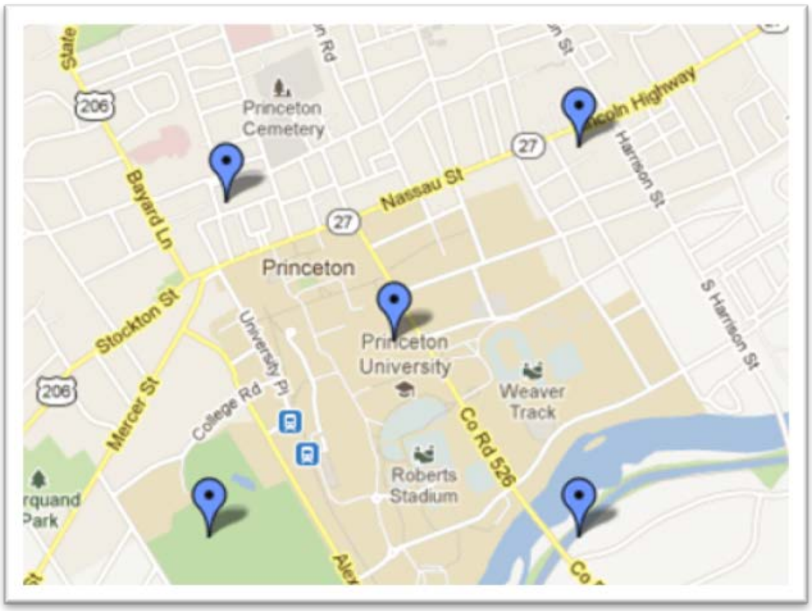
4.2 Process

4.2.1 Input data sets

The program takes two inputs: a School Data file and the PersonFile generated in Task 1. Below is a sample of the School Data file. The selected cells refer to elementary schools in Atlantic and Bergen counties. Overall, the School Data file lists 4918 schools. To expedite Task 4 program run time, we have broken up the file by school type. The result is nine independent School Data files, named Elem, Mid, High, PElem, PMid, PHigh, Special, CommUniv, and NonCommUniv. By separating the data into these nine files, we allow the Task 4 program to through only the relevant schools for the student at hand.

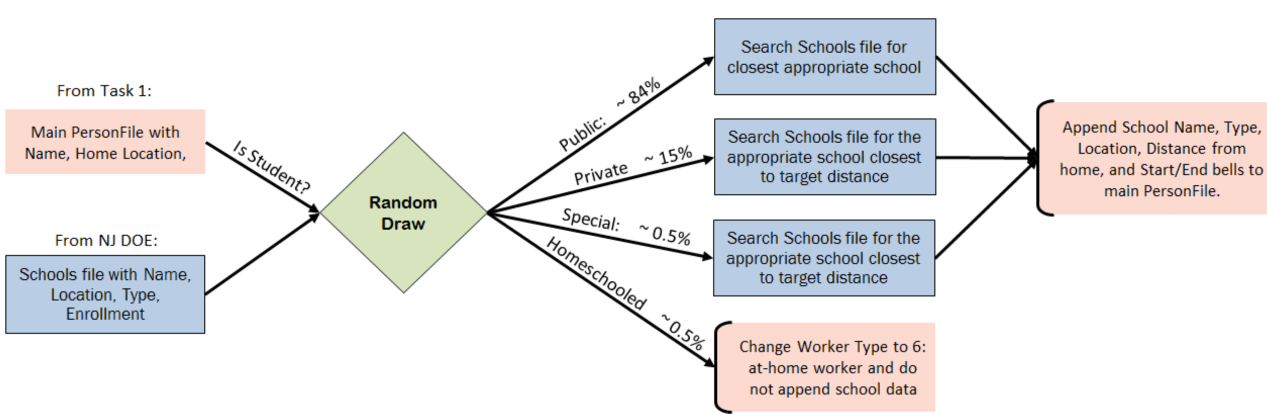
COUNTY	ID	SCHOOL	SOURCE	ENR	LAT	LON
Atlantic	0	Warren E. Sooy Jr. Elementary School	PUBLIC	731	39.64779	-74.8046
Atlantic	0	Washington Avenue Elementary School	PUBLIC	397	39.39315	-74.5259
Atlantic	0	Weymouth Township Elementary School	PUBLIC	212	39.28806	-74.7558
Atlantic	0	William B. Donini Elementary School	PUBLIC	162	39.5261	-74.9421
Atlantic	0	William H. Ross III Intermediate School	PUBLIC	267	39.43	-74.58
Bergen	1	5-6 School	PUBLIC	618	40.88856	-74.0422
Bergen	1	Abraham Lincoln Elementary School	PUBLIC	364	41.01371	-74.1598
Bergen	1	Alexander Hamilton School	PUBLIC	268	40.95316	-74.1309
Bergen	1	Alfred S. Faust Intermediate Elementary School	PUBLIC	247	40.83492	-74.0996
Bergen	1	Alpine Public Elementary School	PUBLIC	154	40.94258	-73.9283
Bergen	1	Ann Blanche Smith Elementary School	PUBLIC	406	41.00433	-74.0216
Bergen	1	Anna C. Scott Elementary School	PUBLIC	639	40.86045	-73.9872

The files for all primary schools, secondary schools, and commuter universities resemble this sample from Elem, and contain sufficient information to assign commuting students to the school they will arrive at each weekday morning. Non-commuter universities such as Princeton and Rutgers, however, offer a unique challenge because of the multiple purposes they serve. A Princeton or Rutgers is not just a destination for its students, but also a home to the vast majority of them, even if their "listed" household address is in Paramus or Trenton. To handle these boarding universities, we created for each a bounding box around the campus's centroid, using an online maps tool[1]. Princeton University's bounding box is shown below. Students who are assigned to Princeton in our program are also assigned an approximate dorm location, in a random spot uniformly distributed across the bounding box. This replaces their home latitude and longitude, and acts as their home for the remainder of the trip file generation. They are also assigned a "classroom" location within campus, which serves the same purpose that school latitudes and longitudes serve for other students.



The second file input data file is the PersonFile from Task 1, which contains all the residential information for each person. The Task 4 program appends to each student's row a School Name, Type, Latitude, Longitude, Distance from Home, Start Bell, and End Bell. For non-commuter university students, as discussed above, the Task 4 program also updates the student's latitude and longitude to his or her "dorm address," while keeping household and home county data unchanged.

4.2.2 Flow Chart of Complete Processes



The complete process of Task 4 is illustrated in the Flow Chart above. The program reads in data from the Main PersonFile, and if a person is identified as a student (a person with Worker Type 0, 1, 2, 3, or 4), it sends that person through random draw and based on the outcome, designates him or her as a public-schooler, private-schooler, pupil at a special school for the handicapped, or homeschooled student. The program's type-specific actions are explained below:

Handling Homeschoolers

New Jersey has historically been one of the least friendly states to homeschooling.[2] While it is permitted today, the state does not keep an annual count of homeschooled students. Estimates range from 3,000 to 30,000. We chose a reasonable estimate of 10,000 homeschooled children when constructing our program. This works out to 0.618% of New Jersey's 1.6 million primary and secondary school students.

When the Task 3 code encounters a student that has been identified by the Random Draw as homeschooled, no data is appended to that child's entry in the PersonFile, but his or her Worker Type is changed to "6: at-home worker." The rationale behind this choice is that a homeschooled student makes trips in much the same way a stay-at-home parent would, without the time-restrictions of a rigid schedule.

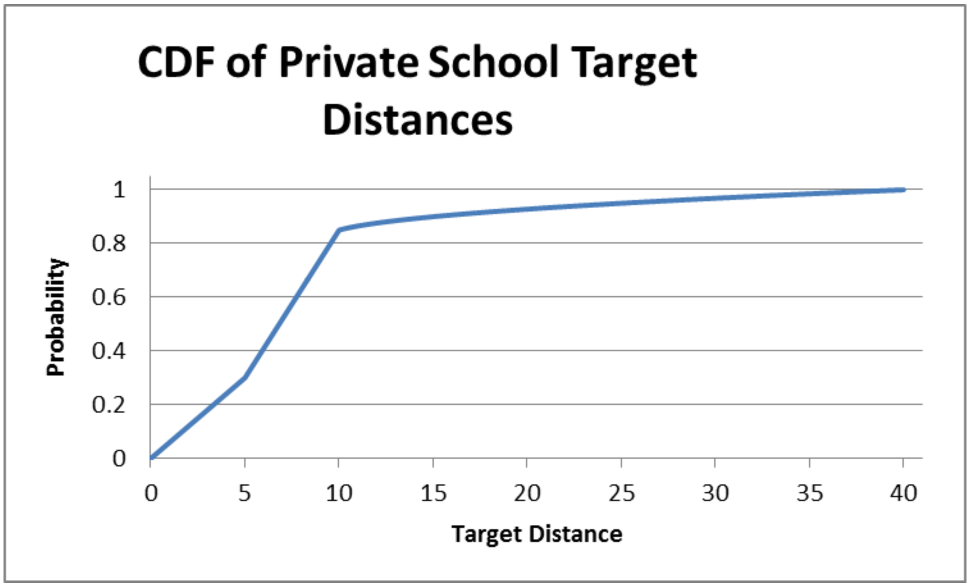
Handling Students at Special Schools

In New Jersey, 204,949 public school students qualify as "Special Needs," but only a small fraction attend a school that solely serves the handicapped. That number across public and private schools is estimated by the NJ Department of Education at 10,660. This works out to 0.659% of all primary- and secondary-schoolers. The students who attend these schools are often the most severely handicapped, for whom age is not a good indicator of grade-level as it is with most children. For this reason our program does not parse up the students of handicapped schools into Elementary, Middle, and High school like their public- and private-school attending contemporaries. Instead, it simply assigns the student to the closest special needs school to their home, regardless of county or any other factor.

Handling Private School Students

Some 240,555 students are reported to attend private schools in New Jersey. This comes out to 14.86% of all students. One potential difficulty in assigning private school students to schools is the inability to model the complex decision-making that goes into choosing a school for one's child. While the vast majority of private school students do attend a school nearby, some parents are willing to drive their children dozens of miles each day to a school they think is best. In an attempt to model distance preference for private school assignment, we developed a piecewise cumulative distribution function.

[2] New Jersey Homeschool Association http://jerseyhomeschool.net/



For each private schooler we randomly generated a target distance for their parents' private school of choice. That way, instead of choosing the nearest Lutheran school, a family instead walks through the program until it finds the school closest to a target distance t, which is a very basic way to model personal choice in a pseudo-random way. T is distributed as shown above, and 30% of all private school students attend school within 5 miles of their home. Some 85% go to a school no more than 10 miles away. The remaining 15% are the children of diehard parents who drive them between 10 and 40 miles to school each day.

For private school students, our program pays no mind to the county in which a school is located, but rather it searches for the one closest to that student's desired travel distance t. Much like a helicopter parent, our program allows no consideration to come between it and what it wants for the child at hand.

Handling Public School Students

Public school students are assigned to schools in a very straightforward way. The program searches within the student's home county, and chooses the nearest school that is not already at capacity. A bit of a hang-up does arise when reconciling the NJ Department of Education's 2010 enrollment numbers and the data generated in Task 1, however. In some counties such as Hudson and Bergen, the state's enrollment numbers for middle school students are substantially lower than the number of students who need to be assigned a school in Task 4. While our program caps the majority of public schools at 110% of 2010 enrollment data, it rectifies this capacity discrepancy by capping the state's middle schools at 200% of 2010 enrollment. This figure still strands some middle school students in Hudson and Passaic counties, so the code was updated to 300% and 210% respectively, for these two special cases.

Handling Commuter College Students

College students who commute each day are often looking for the most convenient way to attend classes and work toward a degree, while staying close to home and often holding down a job. For this reason our program assigns every commuter college student to the nearest commuter college to their home.

Handling Non-Commuter College Students

Assigning students to non-commuter colleges is not nearly as straightforward. Like private schools, boarding colleges are often chosen in a very complex way. Also, the assignment of students to boarding colleges becomes an inter-state endeavor, as a large proportion of New Jersey college students come from outside of New Jersey, and a large proportion of New Jersey high school students venture elsewhere for college. In an attempt to keep the student portion of our trip file an intra-state endeavor, we made the assumption that the number of non-commuter university students who leave New Jersey for school is roughly equal to the number who come to New Jersey from other states and nations. This assumption proved to be incorrect, as the enrollments of New Jersey's non-commuter colleges that are generated in Task 3 are substantially lower than the colleges' known enrollments. Princeton University's generated enrollment of 1298, for example, is well below the known 2011 enrollment of 7806. This is the case across the board however, as Task 1 only generates approximately 46,000 boarding college students, while the recorded enrollment at such schools in 2010 was 258,015. The non-commuter under-enrollment is coupled with commuter school over-enrollment (281,735 generated vs. 183,889 actual), so tweaking the Task 1 distributions might mitigate some, but not all, of the error.

In an attempt to account for at least one of the many attributes students look for in their college, we divided the 37 non-commuter colleges up into 5 groups, based on enrollment size. The smallest group consists of colleges with up to 1000 students, and the largest group contains any schools with more than 17,000. In New Jersey, this means that the largest group contains only Rutgers, which serves over 58,000 students. The Task 3 program assigns a size preference to each non-commuter college student based on a random draw, and then randomly selects a school from that size grouping.

4.2.3 Output data sets

4.2.3.1 Format of output data set(s)

Task 4's output data is formatted like its input data, the PersonFile, but includes the added fields "School," "Latitude," "Longitude," "Distance," "Start Bell," and "End Bell" in the rows that contain students.

4.2.3.2 Sample Output Data,

FIRST	AGE	GENDER	WORKER TYPE		LAT	LON	SCHOOL	TYPE	LAT	LON	DIST	START	END
MELISSA Y.	33	TRUE	6	at-home worker	39.28709	-74.5715							
LYDIA Y.	7	TRUE	0	grade school	39.28709	-74.5715	Upper Township Primary	PUBLIC ELEMENTARY	39.26946	-74.6442	4.08 miles	32100	55800
JOHN V.	14	FALSE	1	middle school	39.28709	-74.5715	Upper Township Middle School	PUBLIC MIDDLE	39.26137	-74.7333	8.85 miles	29700	54300
ANGELA L.	50	TRUE	6	at-home worker	39.28709	-74.5715							
CAROL C.	52	TRUE	6	at-home worker	39.28709	-74.5715							
FRANKLIN X	32	FALSE	6	at-home worker	39.28709	-74.5715							
JEFFREY D.	47	FALSE	5	worker	39.28709	-74.5715							
RUBY F.	65	TRUE	6	retired	39.28709	-74.5715							
ROBERTO F.	49	FALSE	5	worker	39.28709	-74.5715							
YVONNE E.	67	TRUE	6	retired	39.28709	-74.5715							
ANGELA D.	17	TRUE	2	high school	39.28709	-74.5715	Cape May County Tech. High School	PUBLIC HIGH	39.10147	-74.7969	17.63 miles	27600	52200
TONY G.	11	FALSE	1	middle school	39.28709	-74.5715	Upper Township Middle School	PUBLIC MIDDLE	39.26137	-74.7333	8.85 miles	29700	54300
RICHARD I.	52	FALSE	6	at-home worker	39.28709	-74.5715							
BETTY P.	68	TRUE	6	retired	39.28709	-74.5715							
WILLIAM S.	19	FALSE	4	college: on campu	40.34409	-74.6525	Princeton University	NON_COMMUTER UNIV	40.34651	-74.6595	Within Cam	36000	59400
DEANNA M.	18	TRUE	2	high school	39.28709	-74.5715	Cape May County Tech. High School	PUBLIC HIGH	39.10147	-74.7969	17.63 miles	27600	52200
DONALD Z.	57	FALSE	5	worker	39.28709	-74.5715							

Notice that Princeton University student William does not reside inside his census block like the rest of the people in this portion of the Cape May county PersonFile. His new latitude and longitude identifies his “home” location within campus to be just outside of Ichan Lab. His “class” location is inside Little Hall.

Task 3 also outputs the generated enrollment data of all the schools in the School Data file. A sample of such data, from elementary schools in Atlantic and Bergen counties, is included below. The fifth column from the left is the 2010 recorded enrollment, while the last column contains the enrollment numbers generated by the task three code.

48	Atlantic	0	Warren E. Sooy Jr. Elementary School	PUBLIC	731	39.64779	-74.8046	751
49	Atlantic	0	Washington Avenue Elementary School	PUBLIC	397	39.39315	-74.5259	100
50	Atlantic	0	Weymouth Township Elementary School	PUBLIC	212	39.28806	-74.7558	29
51	Atlantic	0	William B. Donini Elementary School	PUBLIC	162	39.5261	-74.9421	179
52	Atlantic	0	William H. Ross III Intermediate School	PUBLIC	267	39.43	-74.58	217
53	Bergen	1	5-6 School	PUBLIC	618	40.88856	-74.0422	680
54	Bergen	1	Abraham Lincoln Elementary School	PUBLIC	364	41.01371	-74.1598	198
55	Bergen	1	Alexander Hamilton School	PUBLIC	268	40.95316	-74.1309	235
56	Bergen	1	Alfred S. Faust Intermediate Elementary S	PUBLIC	247	40.83492	-74.0996	272
57	Bergen	1	Alpine Public Elementary School	PUBLIC	154	40.94258	-73.9283	64
58	Bergen	1	Ann Blanche Smith Elementary School	PUBLIC	406	41.00433	-74.0216	170
59	Bergen	1	Anna C. Scott Elementary School	PUBLIC	639	40.86045	-73.9872	703

4.3 Characteristics of one realization of a complete output

One realization of complete output for all 21 counties of New Jersey assigns 1,754,516 students to all schools. Of this number, 475,826 are in a public elementary school, 368,609 attend a public middle school, and 360,468 are public high school students. In terms of private school students, there are 84,265 in elementary school, 65,327 in middle school, and 63,580 in high school. 8,666 students attend special school, 281,735 are commuting college students, and 46,040 go to boarding colleges.

The probability distributions resemble reality quite well. Public school students make up 83.97% of school-age children, and private schoolers comprise 14.86%. The real-life percentages come out to 83.86% and 14.86% respectively.

4.4 Limitations of Current Results

One major area in which the results are currently limited is the accuracy of the NJ Department of Education data. In the Schools Data, many schools are listed with incorrect addresses, or at P.O. Boxes, or more than one time by a slightly different name. We did considerable work to clean up the data, but could have doubled our efforts for even truer results.

A second limitation, as discussed above, is the lack of an inter-state college student make-up. The assumption that the number of New Jersey residents who leave the state for college equals those who come in was a very lofty, and as it turns out, not quite accurate one. Removing the intra-state constraint from the college student assignment process would help our results be a more accurate representation of daily trips in New Jersey.

4.5 Suggestions for Future Efforts

The most important suggestions for future endeavors to assign a school to each student in New Jersey would be to deal with the two major limitations that we found, and discussed in section 3.4.

Beyond that, a further suggestion is to seek out accurate private school enrollment data. We have done quite a bit of extrapolation to impose enrollment limits on private schools, which has contributed to a bit of inaccuracy in that arena.

9. TASK 5: ASSIGNING A DAILY TRIP TOUR TO EACH PERSON

5.1 Introduction

5.1.1 Objective

Task 5 assigns each resident in our simulated population exactly one trip chain type, which represents the pattern of trips taken on a typical weekday. The trip chain is randomly drawn from a probability distributions made for each worker type (shown in section 5.2.2 Process below).

The objective of Task 5 is to begin taking the demographic information about our population in our New Jersey Resident file and using it to generate information about the number and pattern of trips for each individual. We append a trip chain type to each resident in the file. Once a designated trip chain type is associated with each resident, following tasks will associate locations with the various trip ends for each resident, which will accurately depict the travel demand for the resident.

5.1.2 Purpose

The purpose of Task 5 is to limit the number of possible trip chain patterns a resident can take to a finite number of possibilities so that we are able to later assign actual origins, destinations, start and end times to each trip on a manageable scale. In our task, we have limited the number of possible trips to 7 per person per day (in 1 of 8 possible trip chain patterns) so that later tasks are feasible.

5.2 Process

5.2.1 Input data sets

We use the NJ_Resident file as out input data set. The only piece of data needed to make our trip chain determination is the worker type.

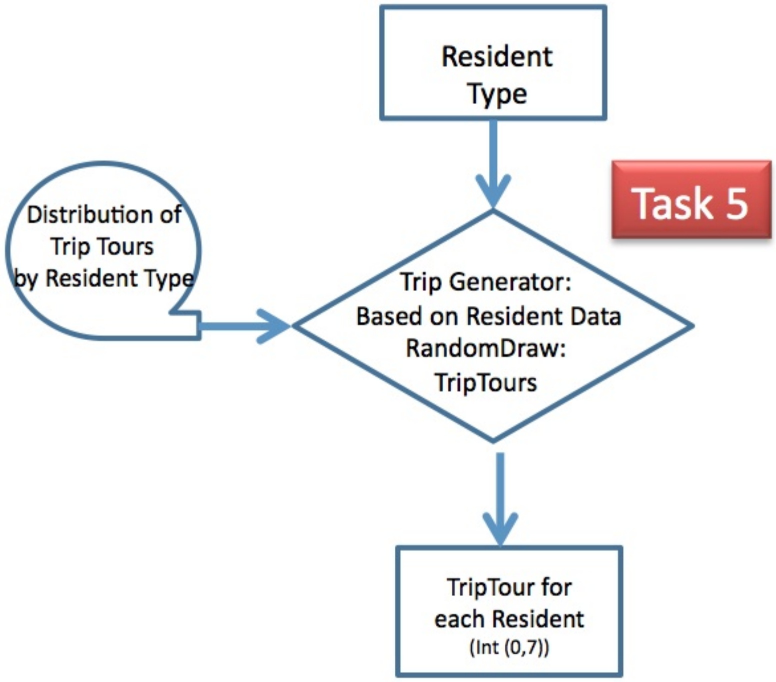
5.2.1.1 Sample input data

As in previous tasks, the NJ_Resident files is formatted as follows:

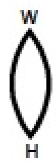
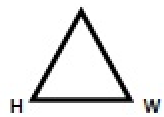
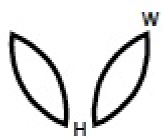
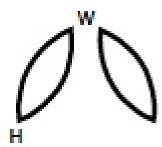
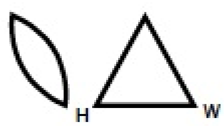
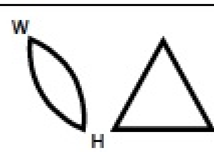
```
ATLTask1.csv (Atlantic County):
0,1,1,PREVILLE,RICHARD G.,24,False,5,worker,39.4393697,-74.4950892
0,2,1,PREVILLE,JACK J.,7,False,0,grade school,39.4393697,-74.4950892
0,3,1,PREVILLE,CHARLES X.,1,False,7,under 5,39.4393697,-74.4950892
0,4,2,DEVEREUX,SUE B.,24,True,6,at-home worker,39.4393697,-74.4950892
0,5,2,DEVEREUX,ANTON P.,2,False,7,under 5,39.4393697,-74.4950892
        …
```

5.2.2 Process

5.2.2.1 Flow chart of complete process



Before we could assign any trip chains, we had to decide exactly which patterns of trips we would consider in our Synthesis. As stated previously, we had to pick a subset of possible trip chains that both represented the most common trip patterns but was small enough to be manageable for later tasks. We chose to limit ourselves to the following trip chains:

TripChainType Number	What it looks like	Number of trip ends
0	H	0
1		2
2		3
3		4
4		4
5		5
6		5
7		7

We assumed that each resident begins and ends their day at home, so the trip chain will start and end at the resident's home address given in the NJ_Residents file in all cases, indicated by the 'H' in the chart above. If the resident falls into the category of Worker (or Out of State Worker), we assumed that the first trip made was directly to work, indicated by the 'W' in the chart above. If the resident falls into the category of student (Grade School, Middle School, High School, College Commuter, or College on Campus), we assumed that the first trip made will be to school, also represented by the 'W' above. All other trips, which occur if the second trip is not immediately back to home for workers and students and occur every time for people who do not travel to work or school (i.e. college on campus or at home workers), are made to 'Other' locations. 'Other' destinations include recreation, shopping, and all other non-work or non-school locations.

In order to assign a trip chain type to each resident, we created probability distributions for each worker type, based on certain assumptions about the travel demands of a particular type of worker. All of our probability distributions are based on the assumption that the average number of trips an active resident makes in a day varies between 3.5 and 4.5 trips/day, with exceptions for college students who live on campus, children under 5, and people living in retirement homes. Each trip is forecasted to represent the demands for the individual. In our task, we do not account for trips where one resident is chaperoning another. For example, we would not account for a trip in which a parent drives their child to school, but rather we are only account for the child's individual trip demand.

5.2.2.2 Probability Distribution Assumptions:

The numbers in the column 'Trip Chain Type' correspond with the trip chains shown in the table above. We assumed that every resident follows exactly one of the designated trip chains based on a random draw from a particular probability distribution that we determined for each worker type.

	Probabilities								
Trip Chain Type	**Grade School**	**Middle School**	**High School**	**College Commuter**	**College on Campus**	**Worker**	**Out of State Worker**	**At Home Worker**	**Nursing Home & Under 5**
0	0.050	0.025	0.025	0.050	0.300	0.010	0.000	0.100	1.000
1	0.175	0.150	0.050	0.075	0.300	0.050	0.600	0.300	0.000
2	0.250	0.200	0.200	0.250	0.200	0.100	0.300	0.200	0.000
3	0.200	0.275	0.225	0.225	0.100	0.150	0.000	0.150	0.000
4	0.000	0.000	0.050	0.000	0.000	0.150	0.100	0.000	0.000
5	0.200	0.200	0.250	0.150	0.040	0.250	0.000	0.100	0.000
6	0.075	0.100	0.150	0.150	0.040	0.200	0.000	0.100	0.000
7	0.050	0.050	0.050	0.100	0.020	0.090	0.000	0.050	0.000
AVG	**3.625**	**3.850**	**4.150**	**4.000**	**2.140**	**4.480**	**2.500**	**3.150**	**0.000**

Grade School Average Trip Ends Per Day: **3.625**
The average number of daily trips a person in grade school makes is **3.625.** The probabilities for the different trip chain types can be found in the above table. We assumed that about 5% of kids stay home from school each day. This number is relatively high compared to students in middle and high school because younger kids generally are more susceptible to becoming sick. Trip Chain Type 2 is comparatively larger because we assumed that a significant percentage of kids this age are involved in after-school activities that take place after school. We assumed that most kids left to these participate activities directly from school, which explains why the probability of Trip Chain Type 2 is larger than Trip Chain Type 3.

Middle School Average Trip Ends Per Day: **3.85**
The average number of daily trips a person in Middle School makes is **3.85**, which is larger than the Grade School demographic. We assumed that a small number of students stay home from school, 2.5%. A smaller percentage of kids return directly home after school, taking into account the assumption that middle school kids are more involved in after-school activities. The largest percentages, Trip Chain Types 2 and 3, characterize the students' after-school activities, with a larger percentage of students returning home before going to another location. A significant number of

students make an average of 5 trips daily, which explains other random trips made throughout the day. The probability of a student making 7 trips in a day is relatively unlikely but still probable.

High School Average Trip Ends Per Day: **4.15**
High School students make significantly more trips per day on average than Grade School students and Middle School students, **4.15** trips. We assumed that most students are involved in some sort of activity after school, so Trip Chain Type 0 and 1 are comparatively low. With much larger probabilities, students make an average of 5 trips per day. This could be credited to high school students becoming licensed drivers and gaining more freedom from their parents or guardians.

College Commute Average Trip Ends Per Day: **4**
The college commuter makes an average of 4 trips per day. They make fewer trips per day than the average High School student. This could be credited to the longer school days typical of college and that most after school activities such as recreation and dining are located close to the campus. College commuters return home after school with a relatively low probability, .075, because we assumed that most students make at least one extra trip throughout the day.

College on Campus Average Trip Ends Per Day: **2.14**
The most interesting thing about college students who live on campus is that 30% of the time they will stay on campus and make zero trips during the day. This is significantly higher than any other demographic and they make an average of **2.14** trips per day. Many universities, including Princeton University, are self-sufficient. Few students will make an average of 5 or more trips per day, with relatively zero students making 7 trips per day.

Worker Average Trip Ends Per Day: **4.48**
This demographic characterizes the bulk of the population, people who are in the work force between ages 18 and 65. Workers make an average of **4.48** trips per day. We assumed that workers rarely skip work, 1% of the time workers will stay at home. Similarly, we assumed that workers would return directly home and stay home for the remainder of the day with a probability of 0.05. The 'Worker' demographic has a special Trip Chain Type 4 which explains trips made during the lunch break, we assumed with a probability of 0.15 workers will go out to lunch. With a probability of 0.45, workers will make an average of 5 trips, which is significantly higher than other categories.

Out-of-State Worker Average Trip Ends Per Day: **2.50**
We assumed that out-of-state workers only display trip chain types 1, 2, and 4. That is, workers in this category will return home with a probability of 0.6, workers will return home after going to another destination with a probability of 0.3, and workers will make a trip leaving from work (maybe to go out to lunch) with a probability of 0.1 before returning home. Out-of-state workers will only follow one of these trip chains because we assumed that they would conduct all 'other' trips outside of New Jersey. In this file, we are only accounting for trips made in New Jersey.

At Home Worker Average Trip Ends Per Day: **3.15**
This category takes into account people who work from home, stay at home parents, unemployed, and retired people. With a probability of 0.1, these people will stay at home for the entire day. This probability distribution is slightly skewed towards making fewer trips in a given day. However, it is still probable that someone of this category might make between 5 and 7 trips per day, thus bringing the average trips per day to be **3.15**.

Nursing Home and Under 5 Average Trip Ends Per Day: **0**
We assume that people under the age of 5 and people in assisted living make zero trips per day.

People in nursing homes will likely not leave their residence, while children under the age of 5 do not have their own trip demand but are simply along for the ride wherever their parents go.

5.2.3 Output data sets

5.2.3.1 Format of output data set(s)

We append the trip chain type and trip chain type index (0-7) to the NJ_Residents file.

5.2.3.2 Sample output data

ATLTask5.csv (Atlantic County):
Worker ID,Trip Chain Type Name,Trip Chain Type Index
1,FIVE_B,6
2,FOUR_A,3
3,ZERO,0
4,TWO,1
5,ZERO,0
...

5.3 Characteristics of one realization of a complete output

We successfully randomly drew from our probability distribution types such that one realization of the population we generated followed our distribution of trip chain types (based on worker type). One indication of this is part of our sanity check (shown below) in which we calculated average trip ends for all of the residents generated in our NJ_Residents file. The overall average number of trip ends for New Jersey residents was 3.41 trips per person per day. The overall average number of trip ends for out of state workers was 2.5 trips per person per day. The realized averages for every worker type matched closely with our intended average trip ends that we established while choosing our probability distributions.

One Realization of Average Trip Ends Based on Worker Type:

Worker Type	Realized Average Trip Ends (based on the output of our Task 5 code)	Intended Average Trip Ends (based on our probability distribution assumptions)
Grade School	3.626283129385837	3.625
Middle School	3.830512282804074	3.85
High School	4.153204172876304	4.15
College Commuter	3.998464407151475	4.00
College On Campus	2.228152101400934	2.14
Worker	4.476620633631194	4.48
At Home Worker (and Retired)	3.141953074694140	3.15
Nursing Home and Under 5	0.0	0.00

5.4 Limitations of Current Results

We created the probability distributions of trip chain types for each of the given resident types based completely on our assumptions about how one type of resident would behave in relation to another (e.g. a high-schooler would be more active than a middle-schooler). There was no readily accessible data source available to justify these assumptions, which could limit the accuracy of our results if we are not as close to reality in our assumptions as we attempted to be.

There is also a lack of real world data to support our assumption of ~4 trip ends per person per day. This lack of data most likely arose because there is no way to concretely classify and define a trip. We did no include short walking trips in our definition of a trip, but there is room to debate when a trip is insignificant and when it should be counted.

Another limiting factor of our results is the number of trip chain types we made available to each resident type. The assumption that all residents make one of eight trip tours in a given day is almost certainly false, as in reality some residents make trips throughout the day that look nothing like our set of trip chains and include more than the 7 trips that we limited our scope to. We justify this by thinking that, on average, we have covered the majority of trip chains that a typical person will make. Trip chains differing from those we included in our project are probably make up a relatively small percentage of trip chains, but affect our accuracy nonetheless.

5.5 Suggestions for Future Efforts

To create more accurate probability distributions for each resident, we suggest finding and incorporating real world data about the actual travel habits of each resident type into our project. The average number of trip ends for each resident type would be more accurate and credible if the probability distributions reflected more real world data to justify our assumptions.

Further research and effort could also be put into creating trip tours specific to different careers and counties. For example, some counties are less densely populated than others, which could influence the number of trips that a resident is willing to make in a given day. Other counties might have higher, more compact populations which influence the amount of traffic and thus inducing people to make fewer trips. We assumed that all residents in the work force have the same probability distribution over the available trip tours, which is a large generalization. We suggest making career specific trip tours, to account for the fact some jobs require the employees to travel more than others.

Additionally, we suggest creating different trip chain probability distributions depending on the day of the week, as a typical work day on a Monday probably varies significantly from a typical work day on a Friday. If we can get real world data to get an idea of the difference between traffic patterns on a specific day (and in the future even the weekends), we could more accurately generate a NJ trip file, perhaps even expanding it to show a realization of a typical work work rather than one day.

10. TASK 6: ASSIGNING THE "OTHER" TRIP ENDS

6.1 Introduction

6.1.1 Objective

We were tasked with the responsibility of building the trip file for all New Jersey residents. Using each resident's trip chain (a description of how each resident travels throughout the day), we assigned origin and destinations to all trips in a resident's chain.

6.1.2 Purpose

The trip file is a critical piece of the trip generator file, as it ultimately determines how traffic flows across the state. A thorough understanding of traffic flows will be very important in the future to better meet transportation demands.

6.2 Process

First, we needed to figure out how many people would visit each "other" destination. We were provided with a comprehensive list of all businesses in New Jersey arranged alphabetically and by county. The business data included name, address, NAICS code, employees, county, and latitude and longitude. In order to create a patron number for each location, the NAICS code of each business was truncated to 3 digits, allowing us to assign a ratio of patrons per employee to each type of business. These ratios were generally the result of educated guesses and checks. Unique errors and outliers were then identified and fixed. We used these patron numbers as weights to assign trips given a trip chain for each resident. A gravity type model was used to assign residents to "other" destinations, factoring in the patron number weight and distance from their current location. The destination county for each trip was chosen using the distance from the person's home and from the trip's origin. Next, the business within the county was chosen randomly using the percentage of that county's patrons that were in each business. Distance between counties was calculated from a centroid of the places of patronage. The measure of distance chosen was the Manhattan Walk Distance as this helped reduce run time.

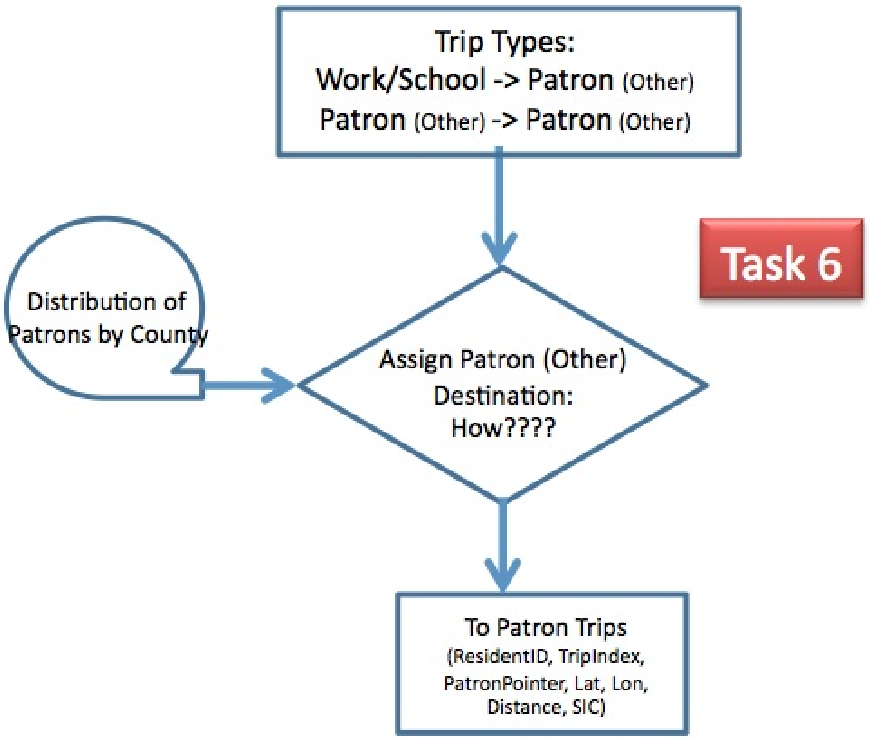
6.2.1 Input data sets

The resident file which included home county, person ID, household number, name, age, gender, worker type by number, worker type by string, home location, School location, work county, pointer, work location, NAICS code, and work start/end time. The trip chain integer was then appended in the final column.

6.2.1.1 Sample input data

0	1	1	PREVILLE	RICHARD	24	FALSE	5	worker	39.43937	-74.4951								22	0	39.95234	-75.1638	0	31162.18	28494.57	5
0	2	1	PREVILLE	JACK J.	7	FALSE	0	grade sch	39.43937	-74.4951	H. Ashton	PUBLIC E	39.43596	-74.4953	0.24 miles	32100	55800	7	0	0	0	0	0	0	2
0	3	1	PREVILLE	CHARLES	1	FALSE	7	under 5	39.43937	-74.4951								0	0	0	0	0	0	0	0
0	4	2	DEVEREU	SUE B.	24	TRUE	6	at-home w	39.43937	-74.4951								0	0	0	0	0	0	0	2
0	5	2	DEVEREU	ANTON F	2	FALSE	7	under 5	39.43937	-74.4951								0	0	0	0	0	0	0	0
0	6	2	DEVEREU	KATIE S.	6	TRUE	0	grade sch	39.43937	-74.4951	H. Ashton	PUBLIC E	39.43596	-74.4953	0.24 miles	32100	55800	0	0	0	0	0	0	0	3
0	7	3	WHEDBE	LINDA C.	26	TRUE	6	at-home w	39.43937	-74.4951								0	0	0	0	0	0	0	2
0	8	4	CARVER	ROBERT	24	FALSE	5	worker	39.43937	-74.4951								0	10101	39.44931	-74.4742	721	31392.93	31707.57	7
0	9	4	CARVER	JENNIFER	25	TRUE	6	at-home w	39.43937	-74.4951								9	0	0	0	0	0	0	0
0	10	5	TINSLEY	ELLEN J.	23	TRUE	4	college: o	40.85646	-74.1978	Montclair	NON_COL	40.8538	-74.2003	Within Ca	36000	59400	0	0	0	0	0	0	0	3

6.2.2 Flow chart of complete process



6.2.3 Output data sets

6.2.3.1 Format of output data set(s)

Person ID, Number of Trips, Trip Chain, Home Latitude, Home Longitude, Home County, Trip Type, Distance, Trip # of day, Destination (Latitude), Destination (Longitude), NAICS code or 0 if to home or school type, Destination County, Pointer, Start Time, End Time

6.2.3.2 Sample output data

Person Id	# of Trips	Trip Chain	Home Lat and Long														
1	5	5	39.4394	-74.4951	0	3	60.2034	1	39.9523	-75.1638	0	22	0	31162	59657	7	27.5374
2	3	2	39.4394	-74.4951	0	1	0.28294	1	39.436	-74.4953	0	0	0	32100	55800	6	6.18015
3	0	0	39.4394	-74.4951	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	2	39.4394	-74.4951	0	5	7.7804	1	39.3614	-74.4274	812	0	10433	0	0	8	8.04916
5	0	0	39.4394	-74.4951	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	3	39.4394	-74.4951	0	1	0.28294	1	39.436	-74.4953	0	0	0	32100	55800	2	0.28294
7	3	2	39.4394	-74.4951	0	5	11.0077	1	39.331	-74.5945	722	0	817	0	0	8	4.70985
8	7	7	39.4394	-74.4951	0	3	1.56918	1	39.4493	-74.4742	721	0	10101	31393	63101	7	45.7439
9	0	0	39.4394	-74.4951	0	0	0	0	0	0	0	0	0	0	0	0	0
10	4	3	40.8565	-74.1078	0	1	0.26785	1	40.8538	-74.2003	4	0	0	36000	50400	2	0.26785

6.3 Characteristics of one realization of a complete output

While looking at the data set above, the resident with Person ID “1”, makes 5 trips during that day, following the chain of home to work, work to other, other to home, home to other, and finally other to home. The home latitude and longitude can be seen next along with the county code of 0 (Atlantic County). The first trip of this person’s day is home to work, represented by the “3” trip type and they travel 60.2 miles. This resident works in County 22 (Philadelphia).

6.4 Limitations of Current Results

1. The Patrons/Employee ratios were largely made up, yet determined the results for the “to other” trips.
2. Truncating NAICS codes to three digits to determine patrons resulted in the grouping of very different types of businesses with widely varying patron numbers.
3. County to county travel was based on centroids in each county. The centroid was found by using the patron numbers at businesses within a county. This caused unlikely travel patterns. For example, a resident in western Monmouth was more likely to travel to a mall closer to the coast than a mall in Mercer County.
4. Within the business data file there was an issue with duplicate entries. Some businesses had the same address and coordinate location, yet different employee counts. We chose to delete duplicates where the employee count was the same, leaving one business location, but to delete rows which were identical duplicates.

6.5 Suggestions for Future Efforts

1. The Patrons/Employee ratios could be greatly improved with some research.
2. NAICS codes could be truncated to 4 or 5 digits rather than 3 digits to more accurately assign patrons to businesses.
3. Counties could be broken down into partial counties using blocks of latitudes and longitudes to deal more efficiently with the centroid problem.
4. Regarding the duplicates in the business data file, it may have been better in some cases to add the employees of each duplicate and then to combine the locations.

11. TASK 7: ASSIGNING A DEPARTURE TIME TO EACH TRIP

7.1 Introduction

7.1.1 Objective

The objective of Task 8 is to assign (by assumption, calculation or estimation) Departure/Arrival Times to all trips made in New Jersey by 8.5 million person-observations on a given weekday.

7.1.2 Purpose

The purpose of assigning Departure/Arrival Times is to obtain a temporal distribution of trips throughout a given weekday in New Jersey in order to observe what are the peak and valley times throughout the day. With this information transportation system planners could determine what are the peak times that need to be serviced by public transit on a given weekday, for instance.

7.2 Process

7.2.1 Input data sets

The main input for Task 8 is a dataset containing 8.5 million person-observations with the following data:

- Trip chain: includes number of trips in given day per person-observation, trip purpose type for each trip and origin/destination information for each trip
- Origin and destination SIC codes and school classification codes for all businesses and schools
- Distance information: home location (latitude/longitude) for every person-observation; latitude/longitude for every origin/destination (businesses/schools)
- Average speeds
- Start/end times for schools
- Worker shifts for all workers based on business start/end times
- Shopping/service acquisition peak times for non-workers per business category
- Other task groups used a different set of SIC codes for businesses. For this task, we wanted to reduce the number of SIC codes as much as possible in order to limit the number of trip combinations. To match up the SIC codes from different groups, we used a key that translates SICs and school classifications obtained from other Task Groups into 14 business codes and 4 school codes used in Task 8
- For every trip combination (161 in total):
    - Modality of distribution (i.e. unimodal or bimodal)
    - Distributed elements: an indication of what elements will be subjected to a distribution model (e.g. departure time, arrival time, etc.) and whether earliness/lateness is also estimated
    - Assumed/calculated elements: an indication of what elements will be assumed or calculated using other inputs, as opposed to being estimated through distribution
    - For Home-School, Home-Work, Home-Other, Other-Other and Other-Home trip combinations, for which a fixed arrival (or departure) time is assumed/calculated for a given person-observation, we use as inputs the width of a 95% confidence interval around a calculated average departure (or arrival) time (using distance and average speed) and a given standard deviation for the distribution associated with that trip combination
    - For School-Home, Work-Home, School-Other and Work-Other trip combinations, for which a nominal departure time is assumed, but for which early/late

departures are required, we use expected earliness and lateness values, 95% confidence intervals around those values, a given standard deviation for each distribution around those values and an estimate of the number of people that depart early and late for every trip combination

- For Other-Other and Other-Home trip combinations, estimated duration time in Other, per trip combination

- Sample Trip Combination: (1) “Home-School: Elementary Trip”

	Trip Combinations				
	Trip Type	SIC Origin	SIC Destination	Distribution Modality	Distributed Elements
1	1	0	18	Unimodal	Departure Time

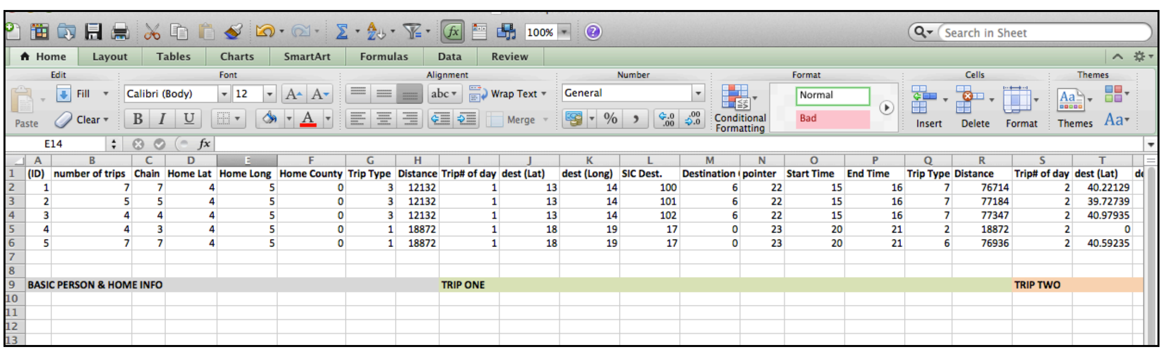
					Ordinary Distribution Function	
Assumed/Calculated Elements	**Distribution Type**	**Skew 1**	**Skew 2**	**Estimated Duration (mins.)**	**95% Time Window (mins.)**	**Standard Deviation (mins.)**
Arrival Time	Normal	0	N/A	N/A	30	7.5

	Earliness Distribution Function	
Expected Earliness Value (mins.)	**95% Time Window (mins.)**	**Standard Deviation (mins.)**
N/A	N/A	N/A

	Lateness Distribution Function			
Expected Lateness Value (mins.)	**95% Time Window (mins.)**	**Standard Deviation (mins.)**	**Early %**	**Late %**
N/A	N/A	N/A	N/A	N/A

7.2.1.1 Sample input data

We are given a Total Trip File input for each county that is organized by Person ID, such that each row represents a different person’s entire trip chain for the day, beginning with trip one, and ending with up to a possibility of trip seven. Each trip has the following 10 columns of useful input information that allows us to calculate departure times for these individual trips:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	(ID)	number of trips	Chain	Home Lat	Home Long	Home County	Trip Type	Distance	Trip# of day	dest (Lat)	dest (Long)	SIC Dest.	Destination (	pointer	Start Time	End Time	Trip Type	Distance	Trip# of day	dest (Lat)
2	1	7	7	4	5	0	3	12132	1	13	14	100	6	22	15	16	7	76714	2	40.22129
3	2	5	5	4	5	0	3	12132	1	13	14	101	6	22	15	16	7	77184	2	39.72739
4	3	4	4	4	5	0	3	12132	1	13	14	102	6	22	15	16	7	77347	2	40.97935
5	4	4	3	4	5	0	1	18872	1	18	19	17	0	23	20	21	2	18872	2	0
6	5	7	7	4	5	0	1	18872	1	18	19	17	0	23	20	21	6	76936	2	40.59235
9	BASIC PERSON & HOME INFO								TRIP ONE										TRIP TWO	

7.2.1.2 Key Assumptions

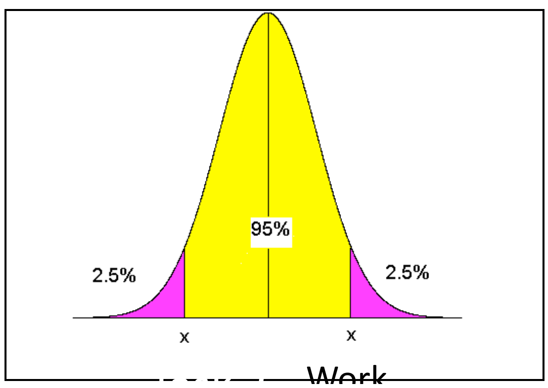
- All Home-School, School-Home and School-Other trip combinations are assumed unimodal, except those that involve Colleges, which are assumed bimodal
- All Home-Work and Work-Home trip combinations are assumed unimodal, except those that involve Food Services, General Services, Health Services, Restaurants/Bars and Retail & Wholesale, all of which are assumed bimodal (to account for multiple work shifts in those industries)
- All Work-Other trip combinations are assumed unimodal; individuals in late shifts in industries like Food Services or General Services, for instance, are assumed to go home after work instead of going shopping, etc.
- All Home-Other, Other-Other and Other-Home trip combinations are assumed bimodal (to account

for two peaks during the day); with regards to Restaurants/Bars, for instance, we assume that they only serve lunch and dinner, no breakfast

- All distributions used in this exercise are normal, with varying standard deviations depending on the context
- Absolute expected earliness values are smaller than absolute expected lateness values
- Normal distributions around expected earliness values are tighter than those around expected lateness values
- The percentage of workers that leave early is extremely small for most trip combinations with few exceptions, such as those involving government workers

Upon receipt of the 8.5 million person-observations and the concomitant descriptive data, all trips carried out by the person-observations are matched with the corresponding trip combination using the trip chain information for each person-observation. Given the trip combination, the code produces assumed, calculated or estimated departure and arrival times.

Speed is assumed to be normally distributed about 20 miles per hour. Considering the mode split among commuters varies greatly, the standard deviation is relatively large, allowing for 95% individuals to fall into a (-10, +10) miles per hour window such that the standard deviation is 5 miles per hour.



2.2.1. Flow Chart of Complete Process

A) Trips that have an expected arrival time (i.e. **to** school or work):

Read and input E(Arrival Time)→read and input probability distribution + noise→Actual Arrival Time→input speed function (+noise) to compute time of travel (dist/speed)→Subtract t.o.travel→Actual Departure Time->**Output departure time** to corresponding cell (PersonID, Trip #)

B) Trips that have an expected departure time (i.e. **from** school or work):

Read and input E(Departure Time)→read and input probability distribution + noise→**Actual Departure Time***((→input speed function (+noise) to compute time of travel (dist/speed)→Add t.o.travel→Actual Arrival Time))->***Output departure time** to corresponding cell (PersonID, Trip #)

C) Trips that depend significantly upon the previous trip in chain (i.e. other to other)

Read and input Actual Arrival Time of trip-1, (person's previous trip in chain)→ read and input E(duration at new origin location) + noise→Actual Departure Time of new trip*((→input speed function (+noise) to compute time of travel (dist/speed)→Add t.o.travel→Actual Arrival Time for new trip))-*

>**Output departure time** to corresponding cell (PersonID, Trip #

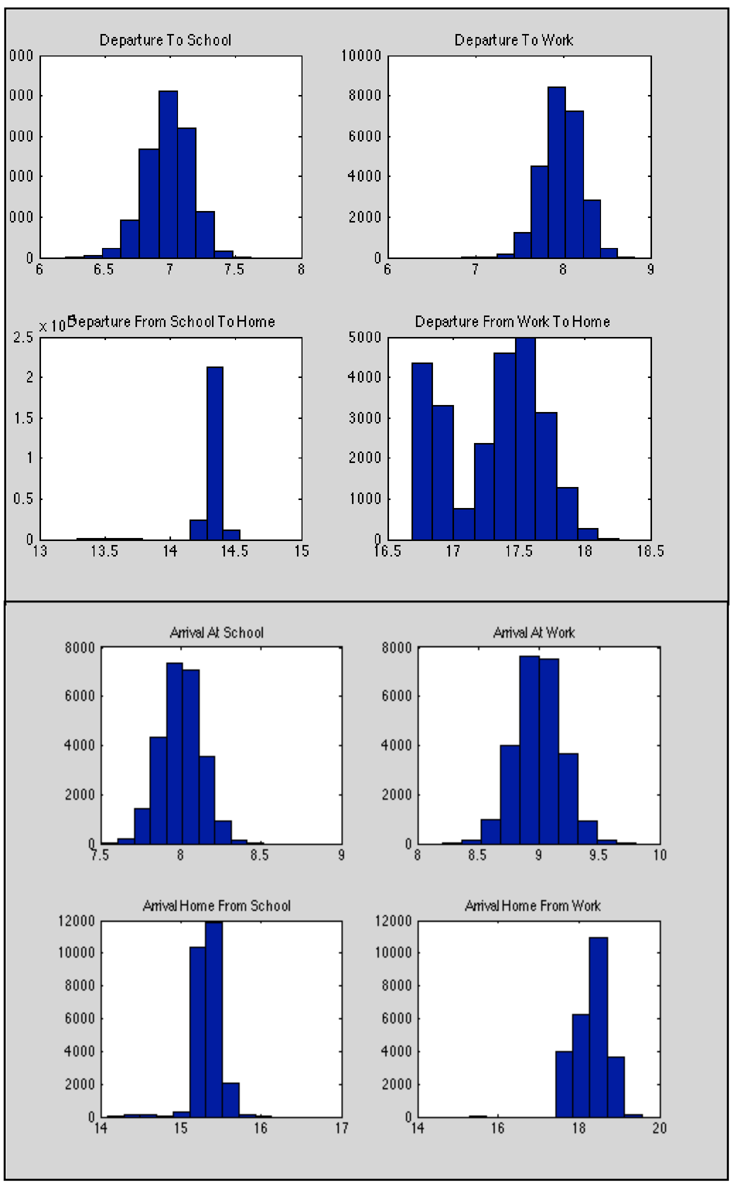
7.2.3 Output Data Sets

7.2.3.1 Format of output data set(s)

For each person-observation that is inputted, the code will output a departure time and arrival time for every single trip that is taken by all of the person-observations.

7.2.3.2 Sample output data and Sanity Checks

Taking a random sample of 25,000 Home to Elementary School trips, 25,000 Home to Administrative/Office job trips, 25,000 Back to Home from Elementary School trips, and 25,000 Back to Home from Administrative/Office job trips, we arrive at the following frequency charts below. Accurately so, school and work arrivals (and departures from home to here) are centered around early mornings, with some leeway, and school and work departures (and arrivals to home) are centered around mid-afternoon to evening.



7.3 Characteristics of one realization of a complete output

Once the departure times for each trip are computed, they are organized, in seconds to midnight, in an array, such that each PersonID for the county has his or her own row in the county trip file, where the first output column within the row refers to the person's first trip of the day, all the way to possibly a 7th trip, representing his or her final trip of the day. For example, person 124 in the GLO county has the following trip departure chain:

124	35911	40457	60305	65700	0	0	0	0

Looking at this individual's trip file, this person first takes a trip to a museum, and arrives at approximately 10:00 a.m. (or 35911 seconds from midnight). The person then returns back home, and departs for home at 11:15 a.m. after spending some time at the museum. The individual then leaves their house again at around 4:45 p.m. (60305 seconds from midnight) to go to a motor vehicles or car dealer shop, and finally departs to come back home from the dealership at around 6:15 p.m. (65700 seconds from midnight). The last 3 trips of the day are filled with zeros, as this individual only takes four trips on the day.

7.4 Limitations of Current Results

The results of this process are *estimates* of the actual departure/arrival times of every trip in New Jersey on a given weekday and are thus subject to inaccuracies; they cannot, thus, be used for concrete predictive purposes. Also, the results of this process are for an average weekday in New Jersey and are not reflective of travel patterns on weekends, holidays or days that are otherwise extraordinary within the year. The procedures carried out in Task 8 employ various assumptions and data that are either assumed as best guesses or that are estimates from other procedures carried out in Tasks 1-7; in an ideal situation, all of this data would be factual, rather than assumed. For simplicity all distributions used to estimate departure or arrival times are normal, which in some cases may not be the best assumption. All of these limitations are due to data and time constraints. In addition, the runtime of this program is quite slow, and may take anywhere from 4-6 hours to assign departure times to all trips of all counties.

7.5 Suggestions for Future Efforts

In future efforts, more research can be done to reduce the number of assumptions made on critical variables (e.g. percentage of workers that leave early/late in a given sector) and replace those assumptions with concrete data. Also, with more time and better data, more sophisticated distributions could be used to calculate more accurate estimates of departure/arrival times. We can also focus on a mode split for more accurate estimations of speed distributions (and therefore times of travels and corresponding arrival/departure times for trips), and open up the door for analyses on specific modes of transportations. The code can also take into account more robust and speedy algorithms to delegate departure times quicker and more efficiently. In addition, one crucial element that would greatly enhance the project for future efforts would be a uniform system for input data. In particular, having one unified pointer system would empower us to make much more precise analyses regarding specific businesses. For schools, no pointers were given to us, for "other" trips, one system was given (with truncated/rounded pointers, thereby defeating the individuality of businesses and locations), and for work, another system was used. Essentially 3 different pointers can refer to the same location, leading to confusion. In the absence of, or rounding of, these pointers, no analyses on actual business/school/"other" locations could be done. Finally, this model could be calibrated with GPS data collected in NJ to determine its predictive accuracy.

12. CHARACTERISTICS OF OUTPUT FILES: A TYPICAL WEEKDAY'S NEW JERSEY TRAVEL DEMAND

12.1 Summary Statistics

As stated previously, key take-away statistics about our NJ Trip files are:

-we successfully assigned an origin, destination, departure time, and arrival time to 30,564,582 trips on a typical day in New Jersey
-the average New Jersey citizen makes 3.41 trips per day
-the average out-of-state worker makes 2.50 trips per day within the borders of New Jersey
-the average trip was 19.3 miles long
-the number of people traveling to work was 3,238,548 in our Synthesis, shy of the estimated 4.4 million people employed in New Jersey
-the average commute to work was 19.1 miles long
-the number of children going to school was 1,605,929
-the average trip to school was 4.0 miles long

It should be noted that we tried to model travel demand for a typical weekday in New Jersey. This means that we following considerations have to be taken into account:

1) All our data is current or historical. We did not try to forecast how population distributions would evolve in the future or how travel behaviors might change. It would be possible to generate travel demand out in the future if careful and sound assumptions are made about where population growth might arise or how the labor force might change within the state among other variables.
2) We did not account for day-specific travel demands, or special cases such as holidays or large events.

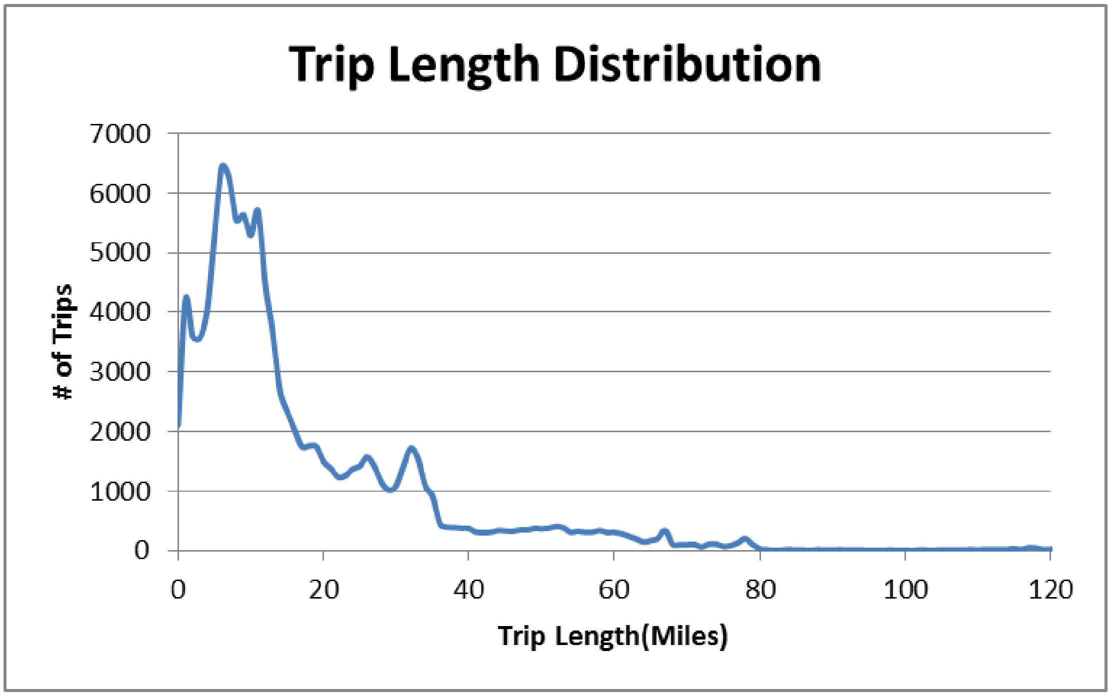
On the following pages are tables that summarize trips by count, length, average length and by trip purpose. Following each table, there are also charts that show the trip length distribution for trips taken in Atlantic county by trip purpose. Due to Excel's limitations in charting data that contains over 250,000 records, we had to constrain ourselves to only plotting the trip length distributions for Atlantic county.

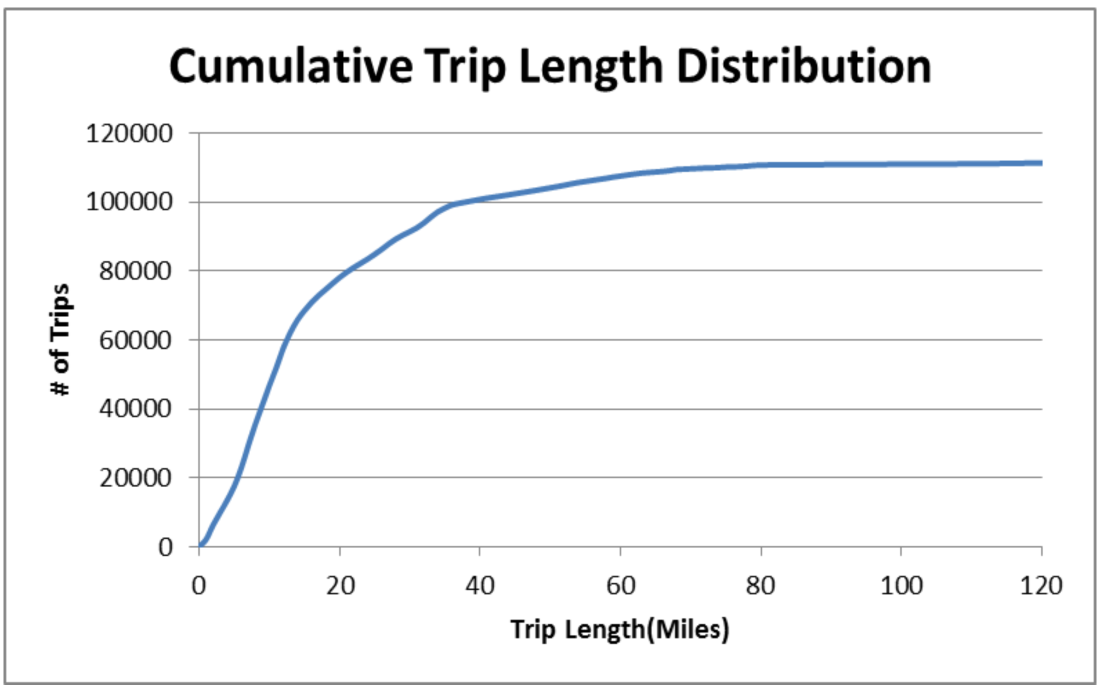
	All Trips		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	936,585	27,723,931	29.6
BER	3,075,434	40,006,145	13
BUC	250,006	9,725,080	38.9
BUR	1,525,713	37,274,682	24.4
CAM	1,746,906	27,523,679	15.8
CAP	333,690	11,026,874	33
CUM	532,897	18,766,986	35.2
ESS	2,663,517	29,307,439	11
GLO	980,302	23,790,798	24.3
HUD	2,153,677	18,580,585	8.6
HUN	437,598	13,044,440	29.8
MER	1,248,183	22,410,297	18
MID	2,753,142	47,579,551	17.3
MON	2,144,477	50,862,651	23.7
MOR	1,677,161	33,746,360	20.1
NOR	12,534	900,434	71.8
NYC	215,915	4,131,764	19.1
OCE	1,964,014	63,174,466	32.2
PAS	1,704,184	22,641,201	13.3
PHL	46,468	1,367,405	29.4
ROC	81,740	2,163,311	26.5
SAL	225,725	8,239,593	36.5
SOM	1,099,927	21,799,647	19.8
SOU	34,493	2,468,016	71.6
SUS	508,674	16,572,792	32.6
UNI	1,824,093	21,860,031	12
WAR	371,169	13,012,489	35.1
WES	16,304	477,950	29.3
Total	**30,564,528**	**590,178,597**	**19.3**

Trip Purpose: Home to Work

	Home2Work		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	93,643	1,724,814	18.4
BER	306,461	4,166,295	13.6
BUC	99,865	3,925,228	39.3
BUR	152,414	3,024,331	19.8
CAM	174,217	2,921,137	16.8
CAP	33,651	809,268	24
CUM	53,057	976,007	18.4
ESS	264,128	3,068,551	11.6
GLO	97,728	3,522,562	36
HUD	212,065	2,017,156	9.5
HUN	43,597	1,040,547	23.9
MER	123,537	2,085,829	16.9
MID	273,830	4,936,443	18
MON	213,592	4,831,167	22.6
MOR	167,281	3,068,346	18.3
NOR	5,046	406,030	80.5
NYC	86,418	1,701,041	19.7
OCE	196,172	5,384,085	27.4
PAS	168,799	2,186,221	13
PHL	18,586	539,041	29
ROC	32,737	895,123	27.3
SAL	22,555	600,869	26.6
SOM	109,273	2,045,880	18.7
SOU	13,772	1,085,350	78.8
SUS	50,702	1,309,213	25.8
UNI	181,743	2,384,690	13.1
WAR	37,148	1,058,270	28.5
WES	6,531	200,474	30.7
Total	**3,238,548**	**61,913,969**	**19.1**

Home to Work Trip Length Distribution for Atlantic County:

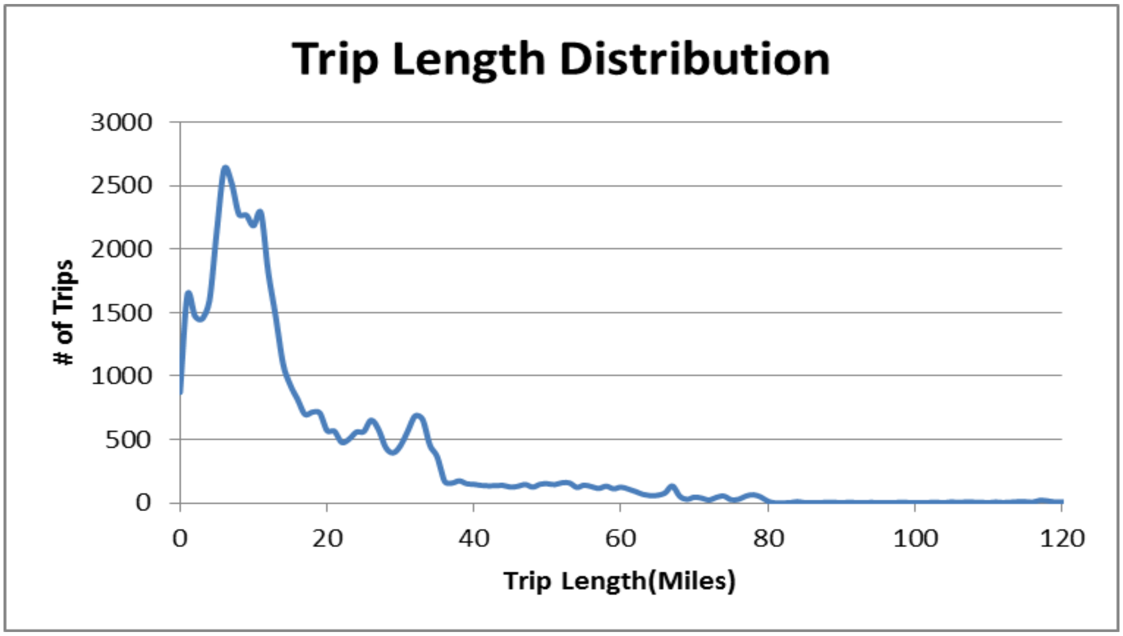


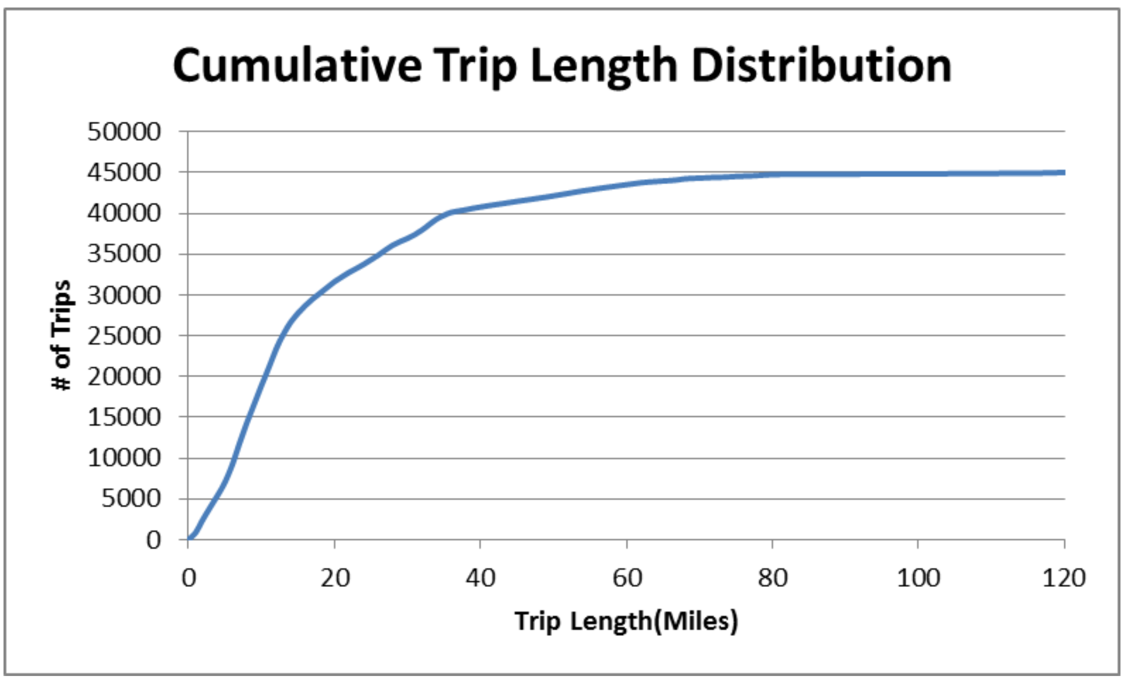


Trip Purpose: Work to Home

	Work2Home		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	52,915	984,156	18.6
BER	173,598	2,371,782	13.7
BUC	69,605	2,736,588	39.3
BUR	86,257	1,720,275	19.9
CAM	98,743	1,666,533	16.9
CAP	19,020	460,518	24.2
CUM	29,999	556,068	18.5
ESS	149,904	1,749,433	11.7
GLO	55,550	2,007,054	36.1
HUD	120,313	1,151,451	9.6
HUN	24,598	586,236	23.8
MER	69,516	1,179,898	17
MID	154,681	2,788,302	18
MON	120,983	2,739,547	22.6
MOR	94,806	1,739,121	18.3
NOR	3,652	294,014	80.5
NYC	60,373	1,185,471	19.6
OCE	111,004	3,050,984	27.5
PAS	95,605	1,247,250	13
PHL	13,054	381,059	29.2
ROC	22,893	626,062	27.3
SAL	12,759	339,546	26.6
SOM	61,819	1,159,204	18.8
SOU	9,623	757,079	78.7
SUS	28,893	747,932	25.9
UNI	102,738	1,350,641	13.1
WAR	21,103	600,186	28.4
WES	4,555	139,228	30.6
Total	**1,868,559**	**36,315,620**	**19.4**

Work to Home Trip Length Distribution for Atlantic County:

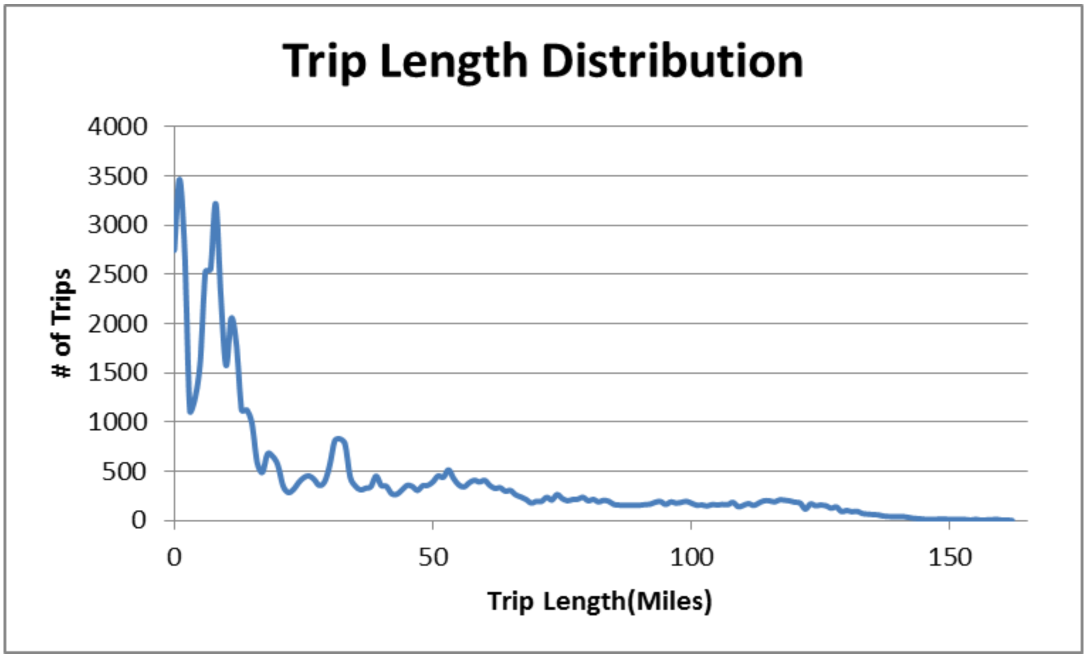


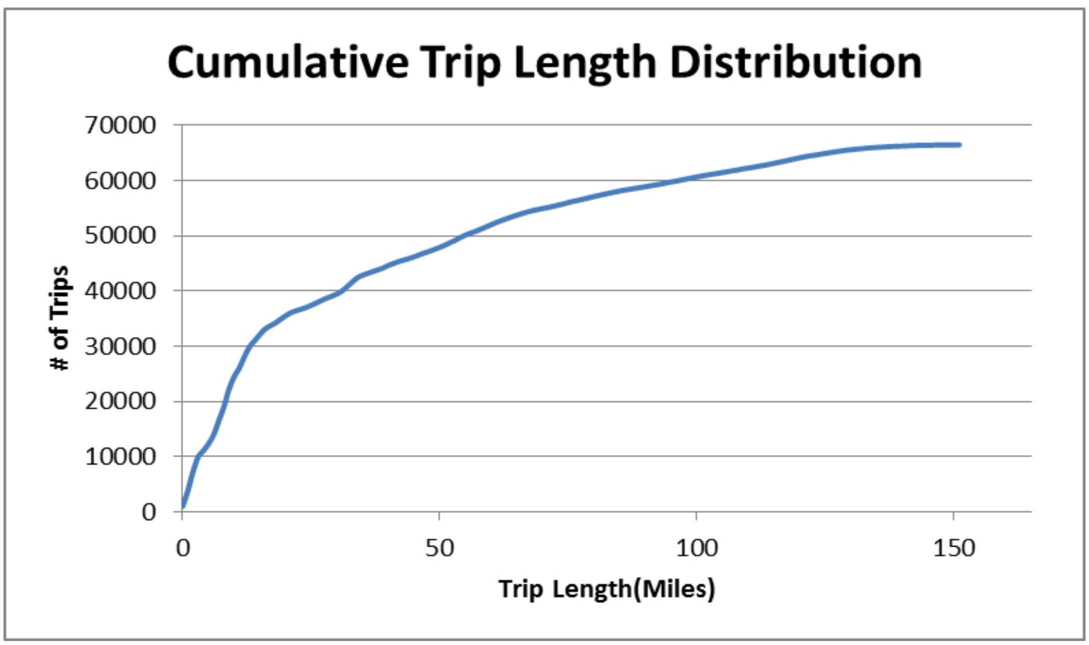


Trip Purpose: Work to Other

	Work2Other		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	54,559	1,890,623	34.7
BER	177,817	2,726,683	15.3
BUC	40,268	1,470,359	36.5
BUR	88,742	2,485,030	28
CAM	101,505	2,270,969	22.4
CAP	19,665	761,388	38.7
CUM	30,969	1,280,128	41.3
ESS	153,536	2,281,238	14.9
GLO	56,753	2,182,113	38.4
HUD	123,339	1,426,163	11.6
HUN	25,514	782,312	30.7
MER	72,096	1,604,118	22.2
MID	159,863	3,354,326	21
MON	124,624	3,312,717	26.6
MOR	97,053	2,068,905	21.3
NOR	1,918	75,923	39.6
NYC	34,562	641,954	18.6
OCE	113,809	3,993,928	35.1
PAS	97,903	1,540,416	15.7
PHL	7,414	223,398	30.1
ROC	13,055	259,537	19.9
SAL	13,141	578,438	44
SOM	63,658	1,467,693	23.1
SOU	5,549	258,367	46.6
SUS	29,302	924,169	31.5
UNI	105,901	1,721,841	16.3
WAR	21,500	752,789	35
WES	2,609	54,476	20.9
Total	**1,836,624**	**42,390,000**	**23.1**

Work to Other Trip Length Distribution for Atlantic County:

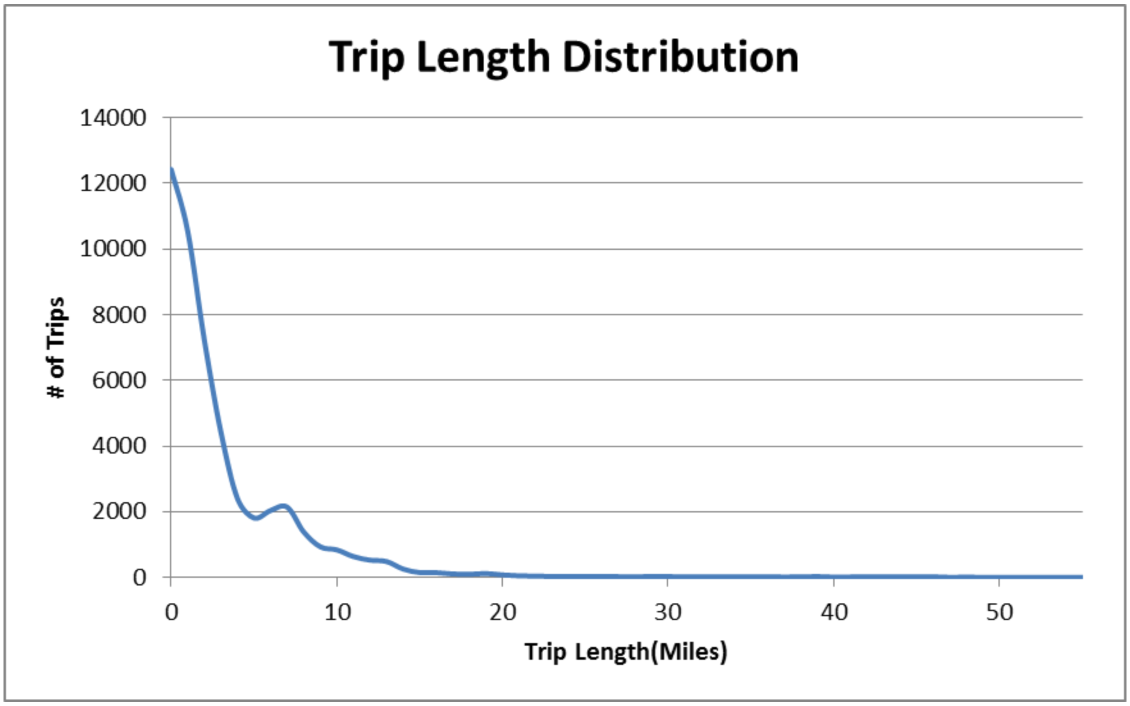


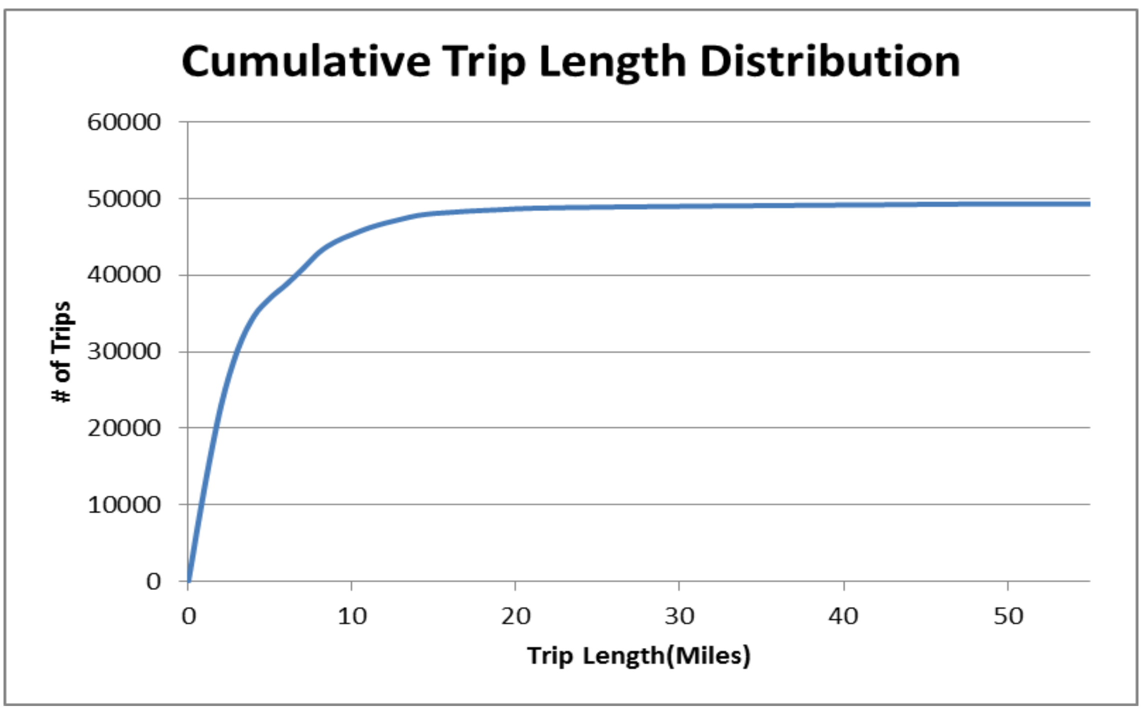


Trip Purpose: Home to School, the not applicable refers to locations that don't have school children.

	Home2School		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	49,782	269,330	3.8
BER	165,065	577,658	3.5
BUC	N/A	N/A	N/A
BUR	80,872	424,834	5.3
CAM	92,887	358,912	3.9
CAP	17,193	198,169	11.5
CUM	28,595	149,412	5.2
ESS	145,581	402,905	2.8
GLO	52,214	220,591	4.2
HUD	119,709	314,259	2.6
HUN	23,262	148,650	6.4
MER	67,326	243,428	3.6
MID	147,919	559,296	3.8
MON	113,975	481,826	4.2
MOR	89,671	365,912	4.1
NOR	N/A	N/A	N/A
NYC	N/A	N/A	N/A
OCE	103,882	594,404	5.7
PAS	92,764	299,433	3.2
PHL	N/A	N/A	N/A
ROC	N/A	N/A	N/A
SAL	11,763	68,822	5.9
SOM	59,572	236,603	4
SOU	N/A	N/A	N/A
SUS	26,929	160,082	5.9
UNI	97,441	266,672	2.7
WAR	19,527	103,055	5.3
WES	N/A	N/A	N/A
Total	**1,605,929**	**6,444,255**	**4**

Home to School Trip Length Distribution for Atlantic County:

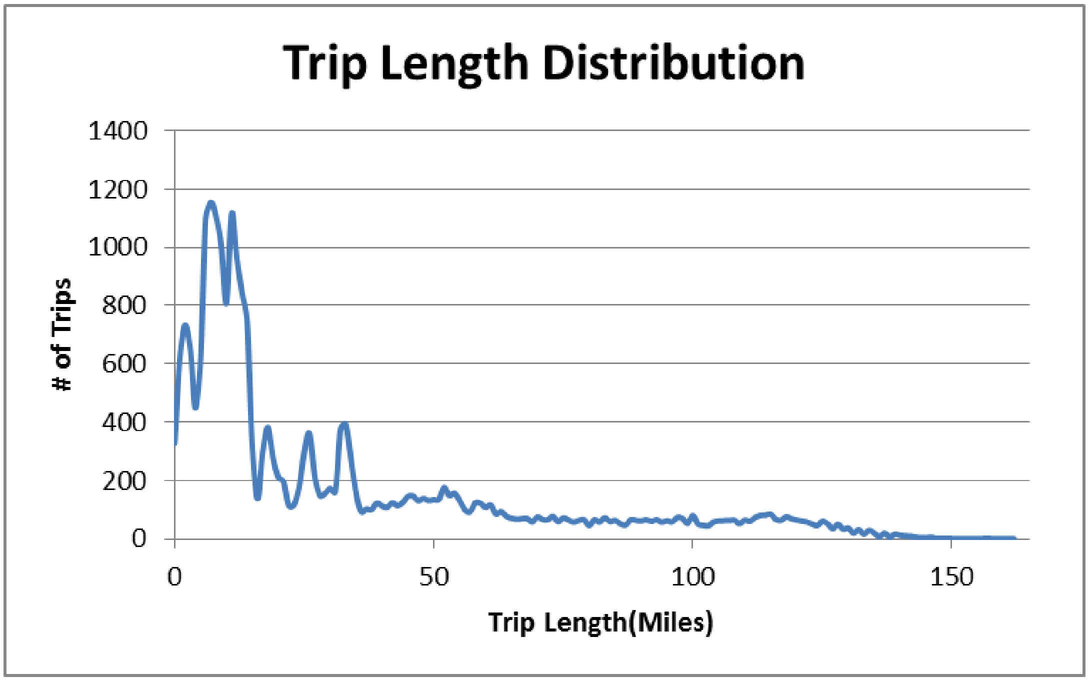


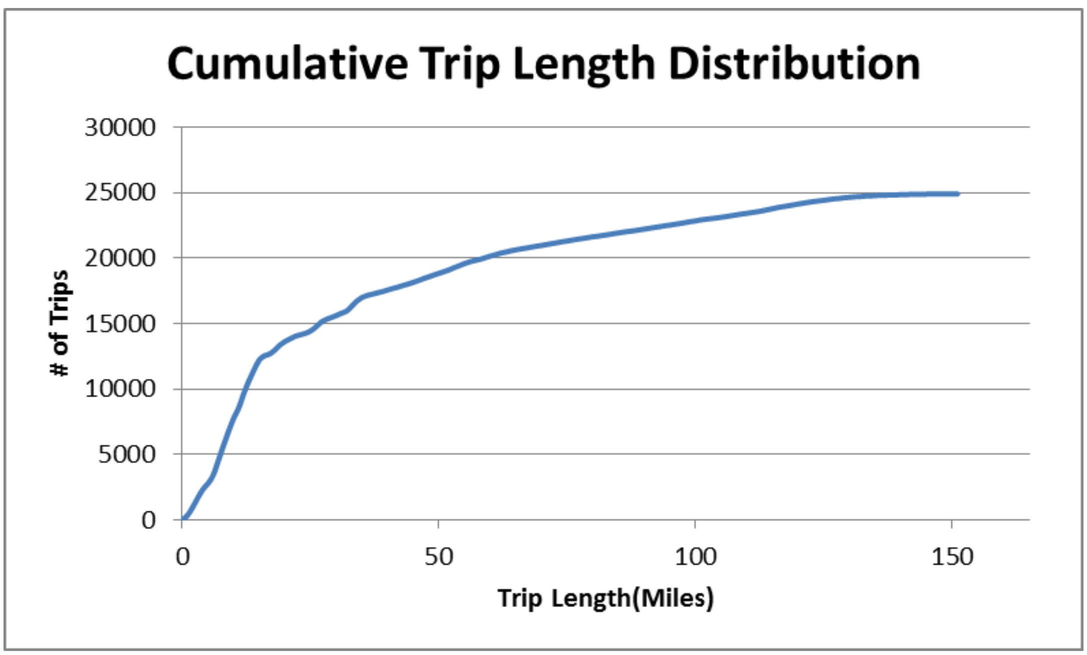


Trip Purpose: School to Other

	School2Other		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	26,945	884,188	32.8
BER	89,077	1,208,350	13.6
BUC	N/A	N/A	N/A
BUR	43,777	1,191,094	27.2
CAM	50,402	806,246	16
CAP	9,342	343,772	36.8
CUM	15,326	603,608	39.4
ESS	78,630	912,275	11.6
GLO	28,288	608,632	21.5
HUD	64,220	581,852	9.1
HUN	12,625	416,004	33
MER	36,127	657,159	18.2
MID	80,020	1,412,052	17.6
MON	61,366	1,532,125	25
MOR	48,413	1,072,024	22.1
NOR	N/A	N/A	N/A
NYC	N/A	N/A	N/A
OCE	56,284	1,938,637	34.4
PAS	50,159	708,956	14.1
PHL	N/A	N/A	N/A
ROC	N/A	N/A	N/A
SAL	6,369	252,917	39.7
SOM	32,231	665,883	20.7
SOU	N/A	N/A	N/A
SUS	14,580	539,227	37
UNI	52,383	624,618	11.9
WAR	10,661	413,239	38.8
WES	N/A	N/A	N/A
Total	**867,225**	**17,372,862**	**20**

School to Other Trip Length Distribution for Atlantic County:

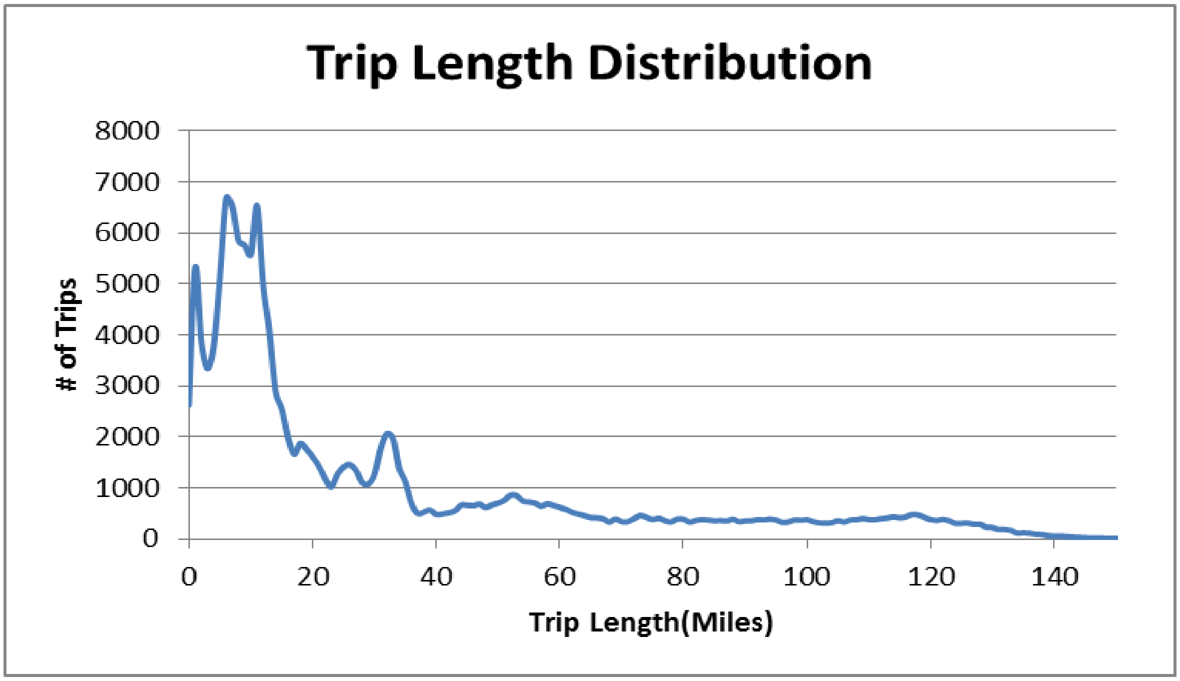


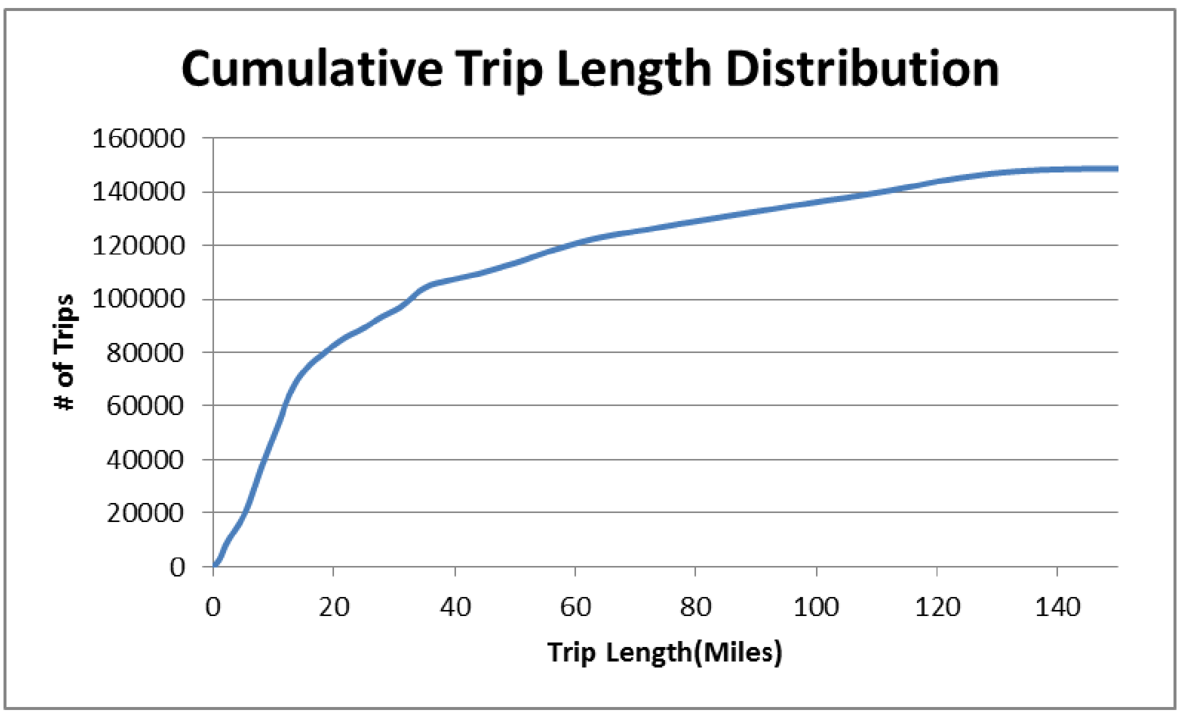


Trip Purpose: Home to Other

	Home2Other		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	242,103	7,743,632	32
BER	795,172	10,496,093	13.2
BUC	N/A	N/A	N/A
BUR	395,040	10,293,821	26.1
CAM	451,776	6,697,353	14.8
CAP	86,500	2,916,664	33.7
CUM	137,666	5,527,264	40.1
ESS	687,012	7,302,337	10.6
GLO	253,621	5,334,855	21
HUD	555,177	4,532,973	8.2
HUN	113,331	3,719,854	32.8
MER	323,389	5,755,052	17.8
MID	711,777	12,320,783	17.3
MON	555,388	13,712,887	24.7
MOR	433,832	9,379,604	21.6
NOR	N/A	N/A	N/A
NYC	N/A	-	N/A
OCE	508,961	17,601,991	34.6
PAS	440,515	6,051,914	13.7
PHL	N/A	N/A	N/A
ROC	N/A	N/A	N/A
SAL	58,583	2,309,688	39.4
SOM	283,987	5,824,078	20.5
SOU	N/A	N/A	N/A
SUS	131,921	4,851,367	36.8
UNI	471,982	5,402,996	11.4
WAR	96,160	3,741,981	38.9
WES	N/A	N/A	N/A
Total	**7,733,893**	**151,517,190**	**19.6**

Home to Other Trip Length Distribution for Atlantic County:

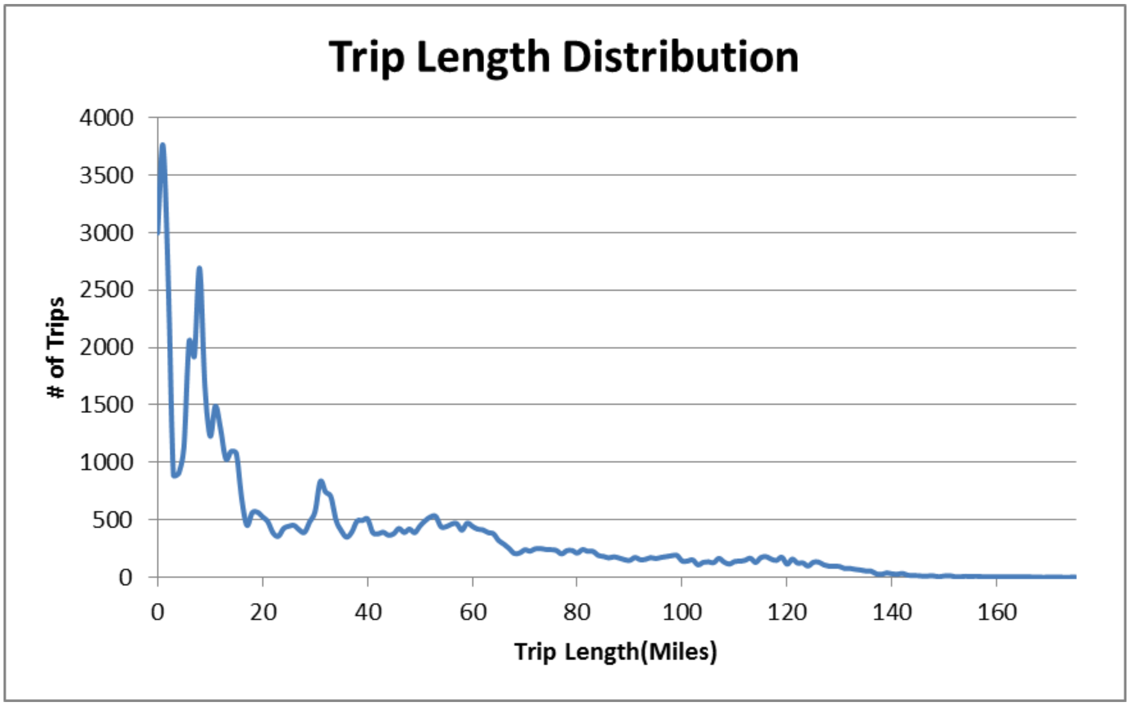


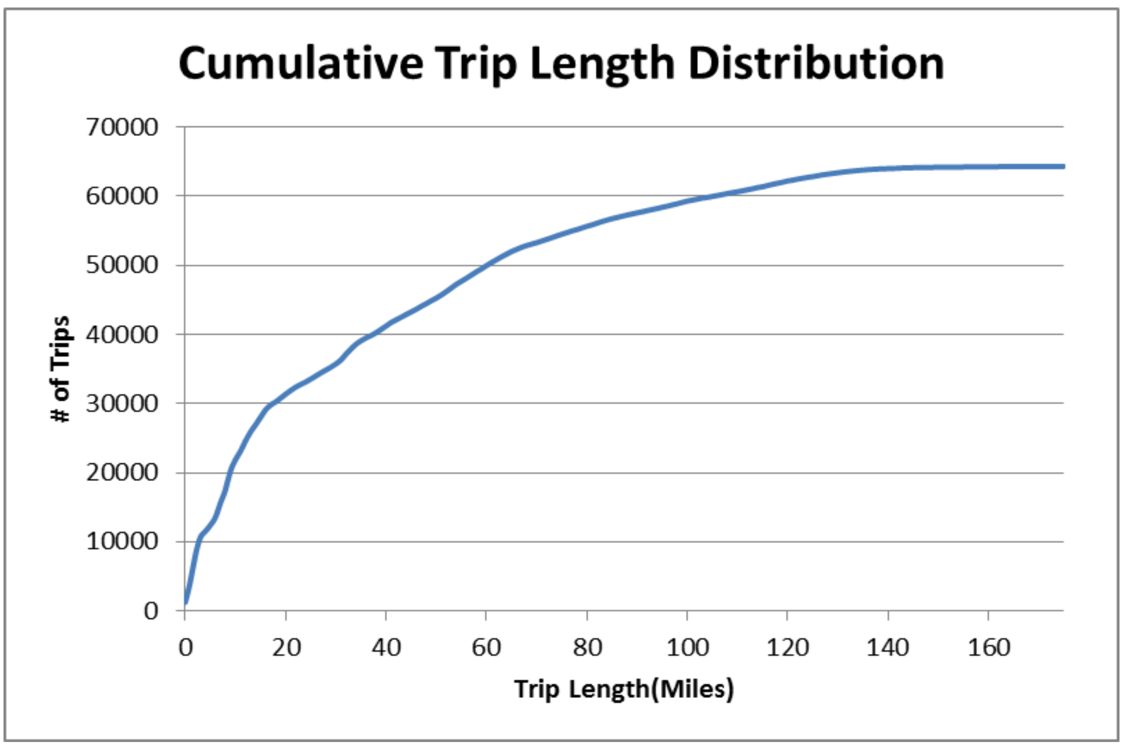


Trip Purpose: Other to Other

	Other2Other		
	Trips	TripMiles	AverageTM
Home County	#	Miles	Miles
ATL	69,464	2,538,600	36.5
BER	227,944	3,368,265	14.8
BUC	N/A	N/A	N/A
BUR	112,905	3,292,505	29.2
CAM	129,984	2,468,093	19
CAP	24,735	942,999	38.1
CUM	39,677	1,581,443	39.9
ESS	196,627	2,782,157	14.1
GLO	72,859	1,731,954	23.8
HUD	159,045	1,663,444	10.5
HUN	32,228	1,056,921	32.8
MER	92,482	2,136,580	23.1
MID	203,583	4,239,167	20.8
MON	158,977	4,431,729	27.9
MOR	124,373	2,793,651	22.5
NOR	N/A	N/A	N/A
NYC	N/A	N/A	N/A
OCE	145,803	5,284,086	36.2
PAS	126,066	1,958,041	15.5
PHL	N/A	N/A	N/A
ROC	N/A	N/A	N/A
SAL	16,922	676,171	40
SOM	81,345	1,977,688	24.3
SOU	N/A	N/A	N/A
SUS	37,840	1,213,308	32.1
UNI	135,262	2,008,660	14.9
WAR	27,617	991,054	35.9
WES	N/A	N/A	N/A
Total	**2,215,738**	**49,136,514**	**22.2**

Other to Other Trip Length Distribution for Atlantic County:





12.2 Trip Length Distributions

The trip length distributions largely conform to what we would expect given the methodology used to generate them. Some interesting aspects of the distributions to note are:

- Most of the distributions are extremely right skewed except the trip purposes that end in an "other" location. This is probably due to the fact that those trip purposes are the ones that weigh the importance of distance the least when assigning destinations of a trip. Our process for choosing X-other trips did not exactly follow the gravity model, which does take into account the relative distances of potential destination locations. Given a starting location, there was a preference for the home county, and once a county was picked, the selection of a destination within the county was done at random and didn't factor in the distance from the starting point. So, the preference for a home county introduces some skewness into the distribution but the random selection afterwards makes it much less skewed compared to the other trip purpose distributions.

- The home to work and work to home trip length distributions are identical except that the work to home distribution has fewer trips. This is expected since a proportion of workers make work to other trips.

- The home to school trip length distribution has the lowest average and is the most skewed distribution. This is because the school assignment weighs distance from the home the most.

13. LIMITATIONS OF CURRENT RESULTS AND SUGGESTIONS FOR FUTURE EFFORTS

13.1 Limitations of Current Results

The ability to create a realistic NJ Trips file was limited by two main factors: available data and time.

Limitations on available data caused assumptions to be made based on intuition in many instances. As discussed before, the household algorithm in Task 1, the input data for names in New Jersey and first and last name independence in Task 1, the probability distribution of trip chain types based on worker types in Task 5, the assumption of approximately 4 trip ends per person per day in Task 5, the Patron/Employee ratios in Task 6, and the assumption of normality for departure and arrival time distributions in Task 7 are all aspects of our project that suffered from a lack of available data. Because of the use of Patron/Employee ratios to construct Other-to-Other trips, distance between Other locations was not sufficiently accounted for and subsequently Other-to-Other trips were the longest on average, 22.2 miles. This does not follow the intuition that people will not travel very far to go to a restaurant, shop, or recreational area, but would prefer one close to their house or on their way home from work or school. Further, because of assumptions and procedure (notably that the number of workers was calculated based on census data about the population distribution and an estimate of the employment level in New Jersey), the number of people traveling to work in a realization of our simulation was 3,238,548 worker, a good bit shy of the estimated 4.4 million people employed in New Jersey. In all the aforementioned cases there was a limited amount of data that was readily available to give a general sense of demographics, distributions, and travel demand. Any missing data was supplemented with intuition, personal experience, and consultations with industry insiders with more experience in transportation.

The other key limitation in our project was time. The code used to generate the population had to be reasonably efficient in order to deal with the time constraints that come with simulating the travel demands of over 9 million people. As mentioned previously, parts of the process that were limited by time include: name generation (because the input files were so large) in Task 1 and the number of trip chain types being limited to 8 types in which one person can make a maximum of 7 trips/day in Task 5. In general, the run time of our program as a whole amounted to several hours, approaching one day. In order to keep this number reasonable, the depth of our assumptions (i.e. the number of factors considered) had to be sacrificed in order to achieve the breadth of covering all of New Jersey that was our stated goal.

Overall, the procedures carried out rely heavily on various assumptions, estimates, and simplifications; in an ideal situation, all of the required data and input probability distributions would be available and factual. Factual data included in the underlying input data of the synthesis includes: population spatial distribution; work, school, and patronage distribution; and Home-to-Work county distributions. Otherwise, input distributions were constructed using educated guesses.

13.2 Suggestions for Future Efforts

The main improvement to our project would come with finding, collecting, and incorporating more precise real world data to reduce the number of assumptions made on critical variables.

In the future, the difference in each county's features could be accounted for by using county demographics and Census data rather than applying state demographics to each county. Currently, our input distributions assume that each county in the NJ has the same characteristics, but it is likely, if not certain, that different counties have different demographic breakdowns, different common trip patterns, and overall different travel demands depending on their population, layout, proximity to New York City, etc.

Finding (or creating) and utilizing more specific location information for residences rather than centroids of counties could also be more illuminating. If counties can be broken down into smaller units or incorporate actual residential location data, it would be easier to get a more realistic picture of where people would want to travel within New Jersey.

To create more accurate probability distributions for each resident, it would also be advisable to find and incorporate real world data about the actual travel habits of each resident type into our project. With more time and better data, our assumed input probability distributions could be justified and/or more sophisticated distributions could be used to find more accurate travel demands. To improve efficiency, our programming code can also take into account more robust and speedy algorithms in generating our NJ Trip File.

Finally, different input probability distributions that vary depending on the day of the week could be used. Our current results are for an average weekday in New Jersey and are not reflective of travel patterns on weekends, holidays or days that are otherwise extraordinary within the year. If real world data that reveals the difference in traffic patterns on specifics days could be obtained and utilized, it might be possible to accurately generate a NJ Trip File that even shows a realization of a typical work week rather than one single day.