# Methodology:

Synthesizing Travel Demand across New Jersey

To restate the goals of this project in operational terms, the model creates a population of individuals whose characteristics, together, come to resemble the aggregate characteristics of people who live and/or work in New Jersey. Then for each of those individuals, the model assigns a ‘Traveler Type’ that is representative of individuals with such characteristics and a home that is representative of where people actually live in NJ. Next, it assigns them work, school and other activities as well as the timings for these functions that are representative of where and when people take part in those respective functions. This section reports and discusses the thought process and methods used to accomplish each of the tasks that are required for the project's high fidelity synthesis.

## Predictable Activities & Others Trips

The different tasks involved in the Synthesizer are of varying difficulties. Even if one were simply modeling his/her own travel patterns for just an average weekday, something as simple as where he/she might go for lunch or to relax after work can be surprisingly difficult to guess. On the other hand, that one will likely go to school and work, and eventually back home can be predicted with great certainty. The trip ends to the less difficult tasks mentioned, such as Home, Work, and School correlate with what are referred to in the literature as ‘more rigid activities,’ and as ‘anchors’ in travel survey documentation (NHTS, 2011). The time a person spends during such activities are considered ‘blocked periods’ in Kitamura and Fuji’s (1998) PCATS model, periods modeled before more variable ‘open periods’. Though this terminology is not used here, the principle remains that activities such as work and school are modeled first due to their greater feasibility of prediction when compared to ‘Other’ trips.

To illustrate, generating places of residence down to the Census Block level and then filling them with people of the right age, sex, and Traveler Type is somewhat easier than deciding where those people go to work and/or school, which is in turn easier than deciding where they choose to dine and recreate. Still this model does all this, in that order, and creates plausible, albeit synthetic, outcomes of trips in space and time. In addition to requiring a large amount of disaggregated location-specific data for such a model, many fundamental assumptions must be made.

## Fundamental Assumptions

A model of real world phenomena is only as good as the assumptions it is based on. The assumptions below cater mainly to the level of data available, as well as the issues of limited time and processing power. They are divided by the tasks to which they are relevant, and in doing so, they reveal the structure of the following section on building the complete New Jersey trip file, in which they are expounded. Some of these assumptions can be improved upon, and will be touched on later in the Conclusions, Limitations, and Next Steps section.

**Task 1** Generate the Populous

* Each household, and therefore each resident, is geographically located at the centroid of the block it is in, as provided by the census data fields INTPTLAT and INTPTLON.
* The number of people by age and sex is known down to the Census Block level, but ages are divided by the census into intervals, 0-4, 5-9, etc. Ages within these intervals are assumed to be distributed uniformly and are sampled as such[[1]](#footnote-1).
* The population is divided into households and group quarters such as dormitories and nursing homes. All are represented as households however and have a household type from 0 to 8. 0 and 1 refer to actual households and the rest refer to group quarters - a full list is shown in Table 1 Codes for Traveler Types, Household Types, and Income BracketsTable 1.
* Households are built by first choosing a household size and a female or male householder. The rest are filled based on household relations distributions as in table P29 in the Census SF1. All sampling used here (and later on) is done with replacement.
* Residents are assigned a Traveler Type from 0-7, which helps the Synthesizer categorize them and later specify their potential sequences of daily activities.
* Traveler Type is based on age and household type (particularly if the household is a group quarter).
* Incomes are assigned to each entire household to reflect in aggregate the income characteristics of each Census Tract. It is then divided among its residents that work to assign them individual incomes.

**Task 2** Assign Work Places

* Workers from out of state are generated deterministically from the 2000 Journey to Work Census data rather than sampled.
* Out-of-state workers are given Household and Traveler Types of 9 and 7 respectively and are immediately assigned a county to work in. Their records are saved in seven different files based on where they reside.
* Every resident worker is first assigned a working county where their employment is located to reflect in aggregate the county-to-county flow from the 2000 Journey to Work Census data.
* All non-workers like children and the elderly, as well as Homeworkers (Traveler Type 6)—including homemakers, the unemployed, or even workers on a sick day—are given a -1 instead of a working county.
* Workers who work outside the state are assigned a -2 instead of a working county.
* Workers who are in school, college, or university work in the same county that they live in by default.
* Workers are then assigned an industry, followed by an employer within that industry. Both are drawn from distributions built using attraction equations.

**Task 3** Assign Schools

* Despite the availability of data on preschools and kindergartens that have children under the age of 5, residents in this age range are of Traveler Type 0 and are not assigned a school, as their travel patterns are typically tied more to that of their parents.
* The data detailing the percent of students enrolled by level and age group used here is at the national level.
* The proportion of enrolled students in public and private institutions by age group, school level, and sex is available at the county level, though age group is used rather than school level.
* For simplicity, lists of schools, colleges, and universities drawn from, both public and private, are limited to those in the same county as the student.
* For public K-12 schools of any level, no sampling is done; rather the school nearest to the child’s resident Census Block is chosen.
* For private schools and higher education, sampling is done with replacement, as has been the case in previous modules.
* Private schools and colleges/universities are sampled from distributions built using an attraction equation, which is weighted by the size of the school over the squared distance between campus centroid and centroid of the Census Block the student lives in.

**Task 4** Assign Tours/Activity Patterns

* All tours begin and end at Home.
* Revised Traveler Type is assigned to deal with students (TT’s 1-4) who are assigned as “Not Enrolled” (Student Type 9). TT’s 1, 2, and 4 are changed to TT 1, Homeworkers. TT 3’s becomes 5’s as they simply work that day without attending college.
* For simplicity, there are exactly 17 different Activity Patterns (referred to in the code as Tour Types), with a different probability for every type of resident.
* If the resident is a Homeworker, all Work nodes in any of the Activity Patterns are considered Other nodes.

**Task 5** Assign Other Trips

* Other trips made from work during lunch hours must be within the work county (Type 11)
* The rest of the Other trips can be in the county itself or any county that is 1-adjacent to it, or neighboring.
* An O location (place of patronage) is drawn randomly with replacement from a distribution that is weighted by the daily patronage at the place divided by the L2 (Euclidean) distance from home to the place, even when it is an Other trip following another Other trip.
* Any trip less than the equivalent of a quarter-mile in distance is ignored, and for Other trips that are followed by a return to work (Type 11), they must be less than 5 miles away or the next nearest place of patronage.

**Task 6** Assign Arrival and Departure Times

* Arrival and Departure Times are all represented by asymmetrical triangular distributions for simplicity, such that few people arrive late or leave early.
* All times are in seconds after midnight.
* Only one average speed is used for all trips, 30 MPH.
* All distances here are calculated more precisely using Great Circle Distance (aka Haversine distance).
* Durations of stay at places of patronage are also drawn using a triangular distribution, the parameters of which are hardcoded to reflect times spent recreating. Minimum is set to 6 minutes, maximum to 2 hours and the mode to 20 minutes.

With the fundamental assumptions of each part of the simulation covered, the following sections proceed to explain more fully each task and how they come together to produce the final trip file. Each task is written up in python code as a module, links to which can be found in the appendix on page 77.

## Task 1: Generating the Populace

The first task operates primarily based on population and household demographics from the 2010 Decennial Census. The goal of Module 1—the programming counterpart to Task 1–is to output a complete resident file for each county in the state. This resident file can be seen as a synthetically generated database that includes rows/records for individual people and columns/fields for particular attributes. These attributes include county number, Household ID, Household Type, latitude and longitude, ID number, Age, Sex, Traveler Type and Income Bracket.

New Jersey counties are represented by an odd number between 1 and 41 following the FIPS County codes; though, within the modules’ coding a custom code from 0-20 is sometimes used for convenience. Out-of-state counties and their categorization into regions are also coded with numbers following 41 and 20 (FIPS and custom codes respectively) but are not dealt with until Task 2. Next, an integer household ID, tracks which household the resident is in. Residents in the same household are displayed in consecutive rows with the same household ID. Household Type uses an integer from 0 to 8 to describe the kind of household or group quarter as shown in Table 1 below.

The latitude and longitude of the center of population (2010 Census Centers of Population by County, 2010) of the Census Block which the resident is in are expressed to 7 decimal places. Every resident's ID starts with a three letter code for the county he/she lives in, followed by an 8 digit number. Then the age and sex of each resident are added, followed by an integer between 0 and 8 representing Traveler Type. And lastly, a code from 0 to 10 signifies which income bracket the resident falls under. All integer-represented attributes are detailed in the table below.

Table 1 Codes for Traveler Types, Household Types, and Income Brackets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Traveler Types** | | | **Household Types** | | **Income Brackets ($)** | |
| 0 | Do-Not-Travel | 0-5, 79 + those in HHT 2,3,4,5,7 | 0 | Family | 0 | < 10,000 |
| 1 | Non-Family | 1 | 10,000 - 14,999 |
| 1 | School-No-Work | 5-15, 16-18×99.81%\* | 2 | Correctional Facility | 2 | 15,000 - 24,999 |
| 2 | School-Work in County | 16-18×0.193% \* | 3 | Juvenile Detention | 3 | 25,000 - 34,999 |
| 3 | College-No-Commute | 18-22×90.34%\* + HHT 6 (Dorms) | 4 | Nursing Homes | 4 | 35,000 - 49,999 |
| 5 | Other institutionalized quarters | 5 | 50,000 - 74,999 |
| 4 | College-Work-in-County | 18-22×**9.66**%\* | 6 | Dormitories | 6 | 75,000 - 99,999 |
| 5 | Typical Traveler Type | 22-64×78% | 7 | Military Quarters | 7 | 100,000-149,999 |
| 6 | Home-Worker-Traveler | 22-64×22%\*\* + 65-79 | 8 | Other non- institutionalized quarters | 8 | 150,000-199,999 |
| 7 | Out-of-State-Worker | Out-of-State |  |  | 9 | > 200,000 |

\* Percentages based on Quarterly Workforce Indicator Q2 2012 data[[2]](#footnote-2)

\*\* Unemployment rounded up to 10%[[3]](#footnote-3) + work-at-home at about 8%[[4]](#footnote-4) + sick days at 4%[[5]](#footnote-5)

Module 1 begins by reading in comma-delimited text files prepared using the 2010 Census Summary File 1 (SF1) (US Census Bureau, 2011) and a VBA macro in MS Access (link in the appendix on page 77). Here, all census data drawn are from tables summarized to the block level. The particular tables drawn from are P12 (Population by Sex by Age), P16 (Population in Households by Age—the table differentiates by ages under/over 18), P29 (Household Type by Relationship), H13 (Household Size), and P43 (Group Quarter Population by Sex by Age by Group Quarter Type). There are likely many ways one could use these and other tables from SF1 to generate a synthetic population for a state. The method used in Module 1 is repeated for every Census Block in every county and is explained briefly below in the following paragraphs. In addition to data from SF1, income data is read in from the 2010 5-Year American Community Survey (US Census Bureau, 2011). This will be explained further below when describing assigning incomes to households and residents.

The census makes available exact block-level data stating the number of people for each sex in each age group (P12). These are iterated through, generating the appropriate number of residents for each group. Their exact age is then chosen randomly by uniformly sampling from within the particular age range. These are kept in four lists, male adults, female adults, male children, and female children, which are shuffled so that they do not remain in the original order of iteration, youngest to oldest age groups. The cut-off age for children in this model is 22 rather than 18 for simplicity that will become apparent in Task 3: Assigning Schools and other Educational Institutions where schools and universities are assigned.

Next, the module begins to form households of different sizes and types. It first iterates over a census data table (H13) which states exactly how many households of sizes 1 to 7+ exist in each block—in this model 7 is the maximum number of occupants generated for any Non-Group Quarter household. For each household in each of these household sizes, the program calls a function to create a single household of the appropriate size. This function works by first selecting whether or not the household is considered a family (Household Type 0) or non-family (Household Type 1), since this affects which distribution to use in determining household members. Next it chooses whether the main householder is a male or a female; again, the distribution sampled from to decide this differs based on family status. Afterwards the remaining members of the household are chosen where the main aspects differentiating them are sex and adult/child status. To illustrate this with an example, two of the fields in table P29 are "Male Biological Child" and "Male Adopted Child," however this level of detail is beyond the scope of this model and thus when either of these options is drawn, the household member created is simply considered a male child. Sampling this way, the appropriate number of times, creates an empty shell for the household. This is then represented by a list, which is filled by popping residents, as appropriate, from the male adults, female adults, male children, and female children lists (here used as stacks) mentioned earlier. Returning to our example, the male children list would be popped twice thus choosing two male children that were generated for this Census Block.

With households of types 0 and 1 generated for a Census Block, the model now generates residents living in other living spaces, which the Census calls Group Quarters. These include places such as military barracks and school dormitories among others detailed in Table 1 above. Table P43 includes a great level of detail, dividing the population into institutionalized quarters like correctional and juvenile facilities and noninstitutionalized quarters such as student housing and military quarters, with those all divided into three age categories: Under 18 years, 18 to 64 years, and 65 years and over. The model assumes only one of each type of quarter per Census Block. This follows the reasoning that most such quarters would be rather large in comparison to the area of a single Census Block. The presence of multiple ones is both unlikely and effectively the same for the purposes of this model. As such, the table is iterated through and group quarters, much like households are represented by lists which are populated by popping the appropriate types of residents from their respective lists. In the remainder of this thesis, unless otherwise mentioned, the term household will also include Group Quarters or Household Types 2 to 8. In populating the block's group quarters, certain other information can immediately be determined and assigned to their residents, namely, Traveler Type and Income Bracket, the final two attributes given to each resident in this model's resident file.

Now every resident is assigned a Traveler Type, numbered from 0 to 6 such as School-No-Work (1) and Homeworker-Traveler (6). These are based primarily on a resident's age and the type of household which they reside in. For example, people in adult correctional facilities and those over 65 in nursing facilities are all of Traveler Type, Do-Not-Travel (0). The rest are detailed in Table 1 above based on a distribution that is currently hard-coded to reflect the distribution for the whole state (see Conclusions, Limitations, and Next Steps for how this could be improved).



Figure 1 Process Chart of Task 1 Methods

School-aged children have two additional Travel Types in addition to the typical, School-non-work (1): School-Work-in-County ( 2), and College-Work-in-County( 4). Residents of these types, as well as Homeworker-Travelers are all assigned an Income Bracket coded between 1 and 10—0 indicates no income. For the first three, this is of consequence because it will be used in Module 2 to help choose where that resident works; not so for type 6 residents because they work at home by definition.

As mentioned earlier, before iterating over the Census Blocks in a county, Census data relevant to the county are read. Before this, however, household income data are read for the entire state. This is done because the data are available only at the Census Tract level, thus the file is not nearly as long. This file can be generated easily using the American FactFinder website (American FactFinder, 2012). It includes the estimated number of households of different types—family and non-family households are used here—in each income bracket. These estimates are used as distributions from which Non-Group Quarter household incomes are sampled. The file also includes margins of error as well as other estimates, however these are never used, and only relevant data are read by the module.

The data are first sampled for every household to generate a household income; a dollar amount is randomly drawn uniformly within the range of the income bracket. This is then distributed over all working members of the household. Once again there are many possible ways in which this could be done; for example, age and/or position in the household could be taken into consideration. In this instance, the module uses a simple function which randomly generates a coefficient for each worker (these coefficients sum to 1), which decides the portion of the household income that he/she makes annually. Each income is then aggregated to an Income Bracket (from 1 to 10) which it falls under.

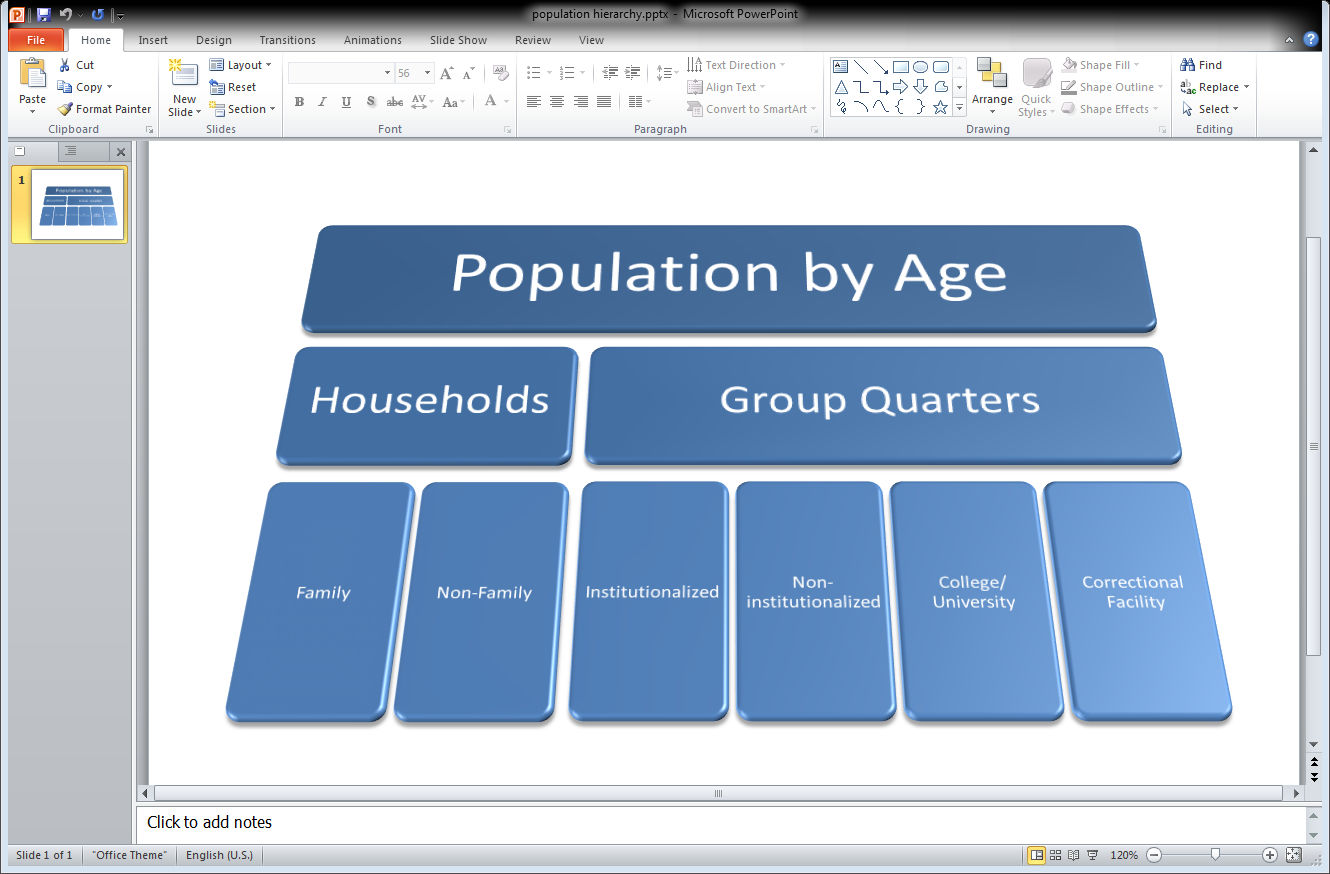


Figure 2 Population Hierarchy

Lastly, the module writes each person in every household to a row in a comma-delimited file. A snapshot of a sample output can be seen below in Figure 3.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Res County** | **HH ID** | **HH Type** | **Lat** | **Long** | **Person ID** | **Age** | **Sex** | **Traveler Type** | **Income Bracket** | **Income Amount ($)** |
| 21 | 1 | 1 | 40.2016752 | -74.7542921 | MER00000001 | 45 | 1 | 5 | 1 | 8410 |
| 21 | 1 | 1 | 40.2016752 | -74.7542921 | MER00000002 | 69 | 0 | 6 | 3 | 16367 |
| 21 | 2 | 1 | 40.2016752 | -74.7542921 | MER00000003 | 82 | 0 | 0 | 0 | 0 |
| 21 | 2 | 1 | 40.2016752 | -74.7542921 | MER00000004 | 97 | 0 | 0 | 0 | 0 |
| 21 | 2 | 1 | 40.2016752 | -74.7542921 | MER00000005 | 61 | 1 | 5 | 3 | 20608 |
| 21 | 3 | 0 | 40.2016752 | -74.7542921 | MER00000006 | 50 | 1 | 5 | 8 | 1173873 |
| 21 | 3 | 0 | 40.2016752 | -74.7542921 | MER00000007 | 9 | 0 | 1 | 0 | 0 |
| 21 | 4 | 1 | 40.2016752 | -74.7542921 | MER00000008 | 52 | 1 | 5 | 1 | 1859 |
| 21 | 4 | 1 | 40.2016752 | -74.7542921 | MER00000009 | 73 | 0 | 6 | 1 | 5649 |
| 21 | 5 | 1 | 40.2016752 | -74.7542921 | MER00000010 | 78 | 0 | 6 | 1 | 2212 |
| 21 | 5 | 1 | 40.2016752 | -74.7542921 | MER00000011 | 73 | 1 | 6 | 1 | 5549 |
| 21 | 6 | 1 | 40.2016752 | -74.7542921 | MER00000012 | 59 | 0 | 5 | 1 | 3594 |
| 21 | 6 | 1 | 40.2016752 | -74.7542921 | MER00000013 | 79 | 0 | 6 | 3 | 16336 |
| 21 | 7 | 0 | 40.2016752 | -74.7542921 | MER00000014 | 60 | 1 | 5 | 7 | 82731 |

Figure 3 Sample Output of Module 1

## Task 2: Assigning Work Places to Workers

The second task generates exact work places for every worker in New Jersey, including both working residents generated in Task 1 as well as out-of-state workers which commute to different counties in the state.

First, Module 2a, the first python script used in Task 2, creates seven resident files, identical in format to those made in Task 1, to account for people who work in New Jersey but reside outside the state. Those who reside outside the United States and Canada are ignored in our model due to their relatively low numbers. These workers are all assigned a Traveler Type of 7 and a Household Type of 9, which reflect that their households are not in the state and that their travel pattern reflects that only come to NJ for work. They are also given an age uniformly chosen between 22 and 65 and a sex drawn at random with a higher probability, 0.61, of being Male. In any case, these attributes play no role in choosing their work place or travel patterns within the scope of this model. In fact, the counties in which each worker lives and works is known deterministically from the 2000 Journey-to-Work Census data's County to County flows file sorted by work state and county. This data is only publicly available at the only county level for privacy reasons (US Census Bureau, 2000).

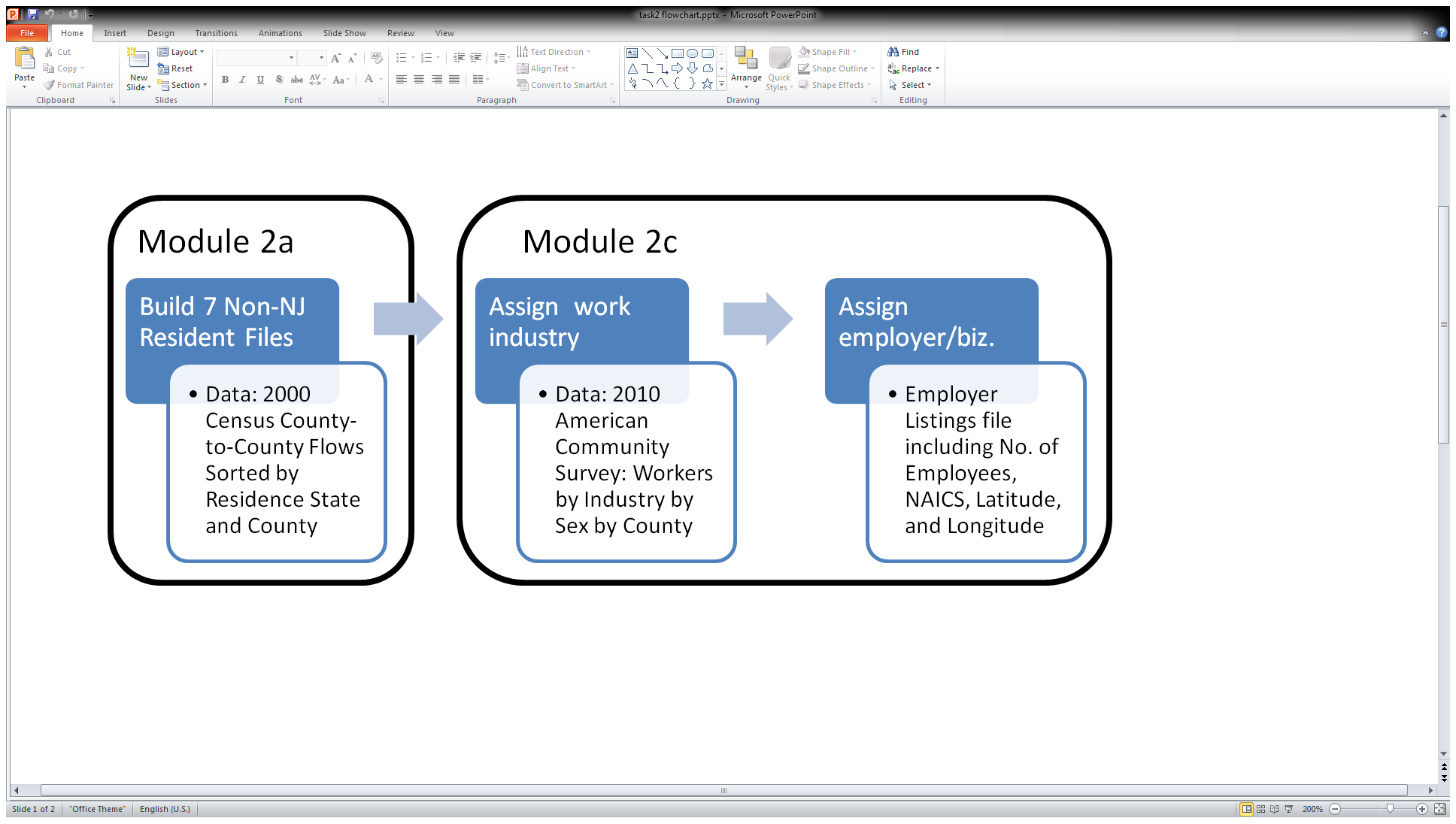


Figure 4 Process Chart of Task 2 Methods for non-NJ Counties

Nevertheless, since the county which they work in within New Jersey is given, determining the work county is trivial. As for their residence counties, all locations are categorized into 7 possible places for the scope of this project, outlined below (credit to N. Webb for its initial compilation).

Table 2 Out-of-State Locations and Categorizations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Custom Coding | Ext. FIPS | Region | Exact Location | Latitude, Longitude |
| NYC | 21 | 42 | New York City | Empire State Building | (40.748716,-73.986171) |
| PHL | 22 | 43 | Philadelphia | Ben Franklin statue | (39.952335,-75.163789) |
| BUC | 23 | 44 | Bucks County PA and West to CA | Newtown, PA | (40.229275,-74.936833) |
| SOU | 24 | 45 | South of Philadelphia | Wilmington DE | (39.745833,-75.546667) |
| NOR | 25 | 46 | North of Bucks County in PA | Allentown PA | (40.608431,-75.490183) |
| WES | 26 | 47 | Westchester County NY and East | White Plains | (41.033986,-73.76291) |
| ROC | 27 | 48 | Rockland, Orange and Rest of NY State | Rockland | (41.148946,-73.983003) |
| INTL | 28 | 49 | Outside the United States | NY Penn Station | (40.750580,-73.993580) |

A complete dictionary mapping each state and/or county to one of these locations, is used in all three parts of Module 2 and are based on work first done by A. Kumar for his part of the ORF467 Trip Synthesizer project Module 2a is essentially a simplified version of Task 1 for out-of-state workers. Module 2c assigns work related attributes in much the same as shall now described for the New Jersey residents and will be elaborated on at the end of this section to highlight noteworthy differences from 2b.

Module 2b, reads in the 21 New Jersey resident files generated in Task 1 so as to append to them the following fields, Work County, Simplified Industry Code, Company of Employment's Name, Employment Zip Code, 3-digit NAICS code, a pointer into the work file, Latitude and Longitude. Note that the pointer, currently, is a row number that refers directly into the Employer file with a header, as it would be viewed in a spreadsheet editor. Due to indices starting with 0 in the code—but 1 in say Excel—and the skipping of the header, should the pointer be used for later code, 1 or 2 may have to be subtracted.

First each resident is assigned an integer to indicate which county they work in, if they work at all. -1 indicates that they do not work, and odd numbers from 1 to 41 (FIPS county codes) represent the 21 counties in New Jersey, with the out-of-state locations represented by consecutive numbers following that, 42 – 49, where 49 is International and is not given an exact location. Rather, the coordinates for international workers are set to those of New York Penn Station. For Traveler Type 5 residents—workers—work counties are drawn from the 2000 Journey-to-Work Census data's County to County flows file sorted by residence state and county. When a county outside the state is drawn, one of the seven locations listed above is chosen based on the previously mentioned mapping (US Census Bureau, 2000).

Table 3 Industry Codes used in Module 2

|  |  |  |
| --- | --- | --- |
| Code | 2-digit Truncated NAICS | Name |
| -2 | - | Out-of-State; No Industry Assigned |
| 0 | 11 | Agriculture Forestry Fishing and Hunting |
| 1 | 21 | Mining |
| 1 | 22 | Utilities |
| 3 | 23 | Construction |
| 4 | 31 | Manufacturing |
| 4 | 32 | Manufacturing |
| 4 | 33 | Manufacturing |
| 5 | 42 | Wholesale Trade |
| 6 | 44 | Retail Trade |
| 6 | 45 | Retail Trade |
| 7 | 48 | Transportation and Warehousing |
| 7 | 49 | Transportation and Warehousing |
| 8 | 51 | Information |
| 9 | 52 | Finance and Insurance |
| 10 | 53 | Real Estate and Rental and Leasing |
| 11 | 54 | Professional Scientific and Technical Services |
| 12 | 55 | Management of Companies and Enterprises |
| 13 | 56 | Administrative and Support and Waste Management and Remediation Services |
| 14 | 61 | Education Services |
| 15 | 62 | Health Care and Social Assistance |
| 16 | 71 | Arts Entertainment and Recreation |
| 17 | 72 | Accommodation and Food Services |
| 18 | 81 | Other Services |
| 19 | 92 | Public Administration |

With the county of work chosen, now the module calls a function to select an industry sector for each resident to work in. To do so the module first creates a distribution from which to draw a sector for every different resident. Through the 2010 American Community Survey, exact numbers of workers in each county for each industry sector are publicly available, as are the median incomes in each of these sectors; furthermore, these are also broken up by sex. Combining these data with the worker's exact income, which was assigned in Task 1, the following equation is used for every industry to build a discrete distribution from which to draw a particular industry (indicated by 0-20) for each worker.

Equation 1 Industry Attraction[[6]](#footnote-6)

Here is the number of workers in a particular industry *i*. Rather than simply draw from such a list, we weight these frequencies by the squared inverse of the difference between the worker’s income and the median income of the particular industry. This heuristic is used to try to more accurately guess what industry a person might work in given their known work county and income without the availability of a detailed breakdown of workers in each industry by income bracket. To avoid errors, missing frequencies and median incomes in the data, which were represented by dashes, were replaced with 0.01 instead to avoid the possibility of a resulting NaN value due to dividing zero by itself. Furthermore, the assigned incomes to workers are to a greater decimal place than the medians so there is a small chance of a zero in the denominator. If the worker works outside of the state, then a -2 is placed instead of an industry code. See Table 3 above for the list of industry categories used for the purposes of this model, their NAICS 2-digit codes, and their simplified code.

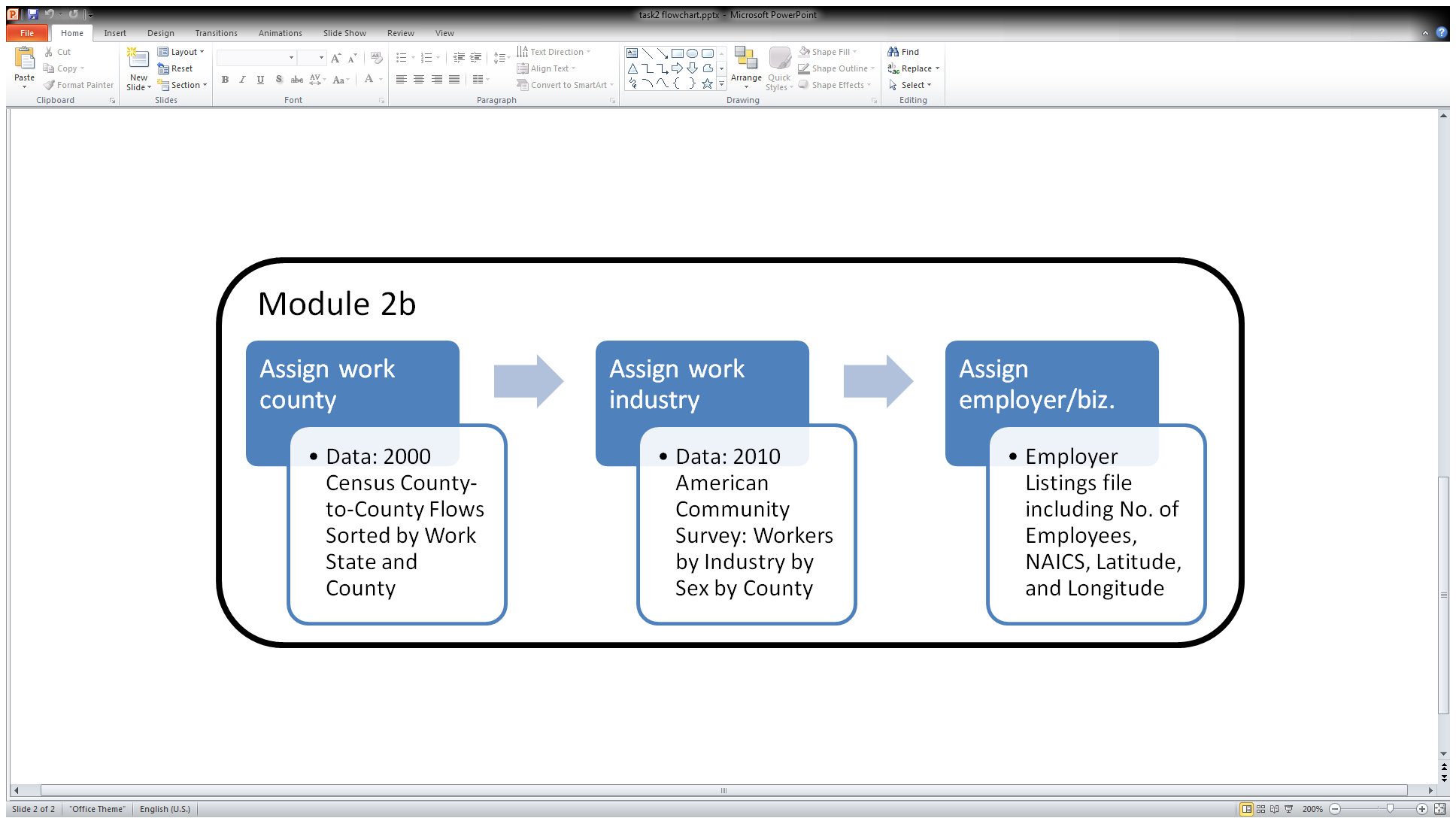


Figure 5 Process Chart of Task 2 Methods for NJ

Lastly, an exact employer is chosen from a dataset of “every” businesses for every county in the state. This list includes the name of the business, its zip code, its NAICS code, the number of employees there, as well as the business's latitude and longitude. A business is drawn by filtering this file by Work County and by the particular industry just assigned, and then drawing from the distribution whose values are a function of the number of employees in a particular business over the square distance to its location, see Equation 2 below.

Equation 2 Employer Attraction

Module 2c borrows all the same functions from 2b to add work attributes to the out-of-state workers generated in 2a. Module 2c differs only in that Work County is not drawn from any distribution but rather deterministically from the 2000 Journey-to-Work Census data's County to County flows as seen earlier in Module 2a.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Res County** | **HH ID** | **HH Type** | **Latitude** | **Longitude** | **Person ID** | **Age** | **Sex** | **Traveler Type** | **Income Bracket** | **Income Amount** | **Work County** |
| 43 | 4 | 9 | 39.95234 | -75.1638 | PHL00000004 | 23 | 0 | 7 | 8 | 105503 | 1 |
| 43 | 5 | 9 | 39.95234 | -75.1638 | PHL00000005 | 44 | 0 | 7 | 7 | 96781 | 1 |
| 43 | 6 | 9 | 39.95234 | -75.1638 | PHL00000006 | 24 | 0 | 7 | 8 | 115316 | 1 |
| 43 | 7 | 9 | 39.95234 | -75.1638 | PHL00000007 | 43 | 0 | 7 | 6 | 74728 | 1 |

Figure 6 Sample Output of Module 2a which generates out-of-state workers

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Work Industry** | **Company Name** | **Work Zip code** | **3 digit NAICS** | **Work Lat** | **Work Long** | **Work Pointer** |
| 13 | Atlantic City Convention | 8401 | 561 | 39.35577 | -74.4388 | 1176 |
| 16 | Caesars Atlantic City | 8401 | 713 | 0 | -74.4358 | 2520 |
| 16 | Hidden Creek Golf Club | 8234 | 713 | 39.37517 | -74.674 | 5321 |
| 2 | New Vistas Corp | 8225 | 237 | 39.38763 | -74.5573 | 7499 |

Figure 7 Sample Output of Module 2c which adds work attributes to out-of-state workers

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Res County** | **HH ID** | **……..**  **………**  **………** | **Traveler Type** | **Income Bracket** | **Income Amount** | **Work County** | **Work Industry** | **Company Name** | **Work Zip code** | **3 digit NAICS** | **Work Lat** | **Work Long** | **Work Pointer** |
| 1 | 1 | 5 | 1 | 3805.615 | 1 | 5 | Tilly's | 8330 | 452 | 39.45389 | -74.6433 | 10621 |
| 1 | 1 | 6 | 1 | 3589.231 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 1 | 5 | 1 | 4680.442 | 1 | 0 | Pleasantdale Farms Inc | 8037 | 111 | 39.63626 | -74.7496 | 8239 |
| 1 | 1 | 5 | 1 | 8279.122 | 1 | 0 | R F Demarco Nursery Inc | 8037 | 111 | 39.66168 | -74.7866 | 8637 |
| 1 | 1 | 2 | 1 | 9123.439 | 1 | 12 | Bask Holding LLC | 8221 | 551 | 39.34066 | -74.5702 | 1728 |
| 1 | 2 | 5 | 7 | 94953.5 | 1 | 6 | First Student Inc | 8213 | 485 | 39.4946 | -74.5985 | 4385 |
| 1 | 2 | 0 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 2 | 6 | 4 | 27796.23 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 2 | 1 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 3 | 5 | 4 | 26054.45 | 1 | 12 | Cape Bank | 8221 | 551 | 39.35674 | -74.5617 | 2632 |
| 1 | 3 | 0 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 3 | 1 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 3 | 0 | 0 | 0 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |
| 1 | 4 | 6 | 3 | 16409.12 | -1 | -1 | -1 | -1 | -1 | -1 | -1 | -1 |

Figure 8 Sample Output of Module 2b

## Task 3: Assigning Schools and other Educational Institutions

The third task deals with assigning a place of study to residents designated Traveler Types 1 through 4, namely students in K-12 schools, colleges, universities and other schools such ones for the severely handicapped.

To begin the module looks at each resident and assesses whether they are special needs students or not. Though a relatively large number of public school students qualify as "Special Needs," the number of students that attend a dedicated school for handicapped children is about l0,660, or 0.66 % of all K-12 students, according to the New Jersey Department of Education. A randomly drawn number between 0 and 1 is drawn and compared to this figure to decide whether the student in question should be assigned a Student Type of 6.

The module then decides whether or not they are enrolled in any non-special school and which level of education their schooling falls under, and whether that school is private or public. The datasets from which these attributes are drawn are detailed in the Data section under School Data Sets. A function performs these checks in that order and assigns one of the following Student Types to each resident:

* 0 – Public Elementary School
* 1 – Public Middle School
* 2 – Public High School
* 3 – Private Elementary School
* 4 – Private Middle School
* 5 – Private High School
* 6 – Special Needs
* 7 – Commuter College/University
* 8 – Non-commuter College/University
* 9 – Not Enrolled

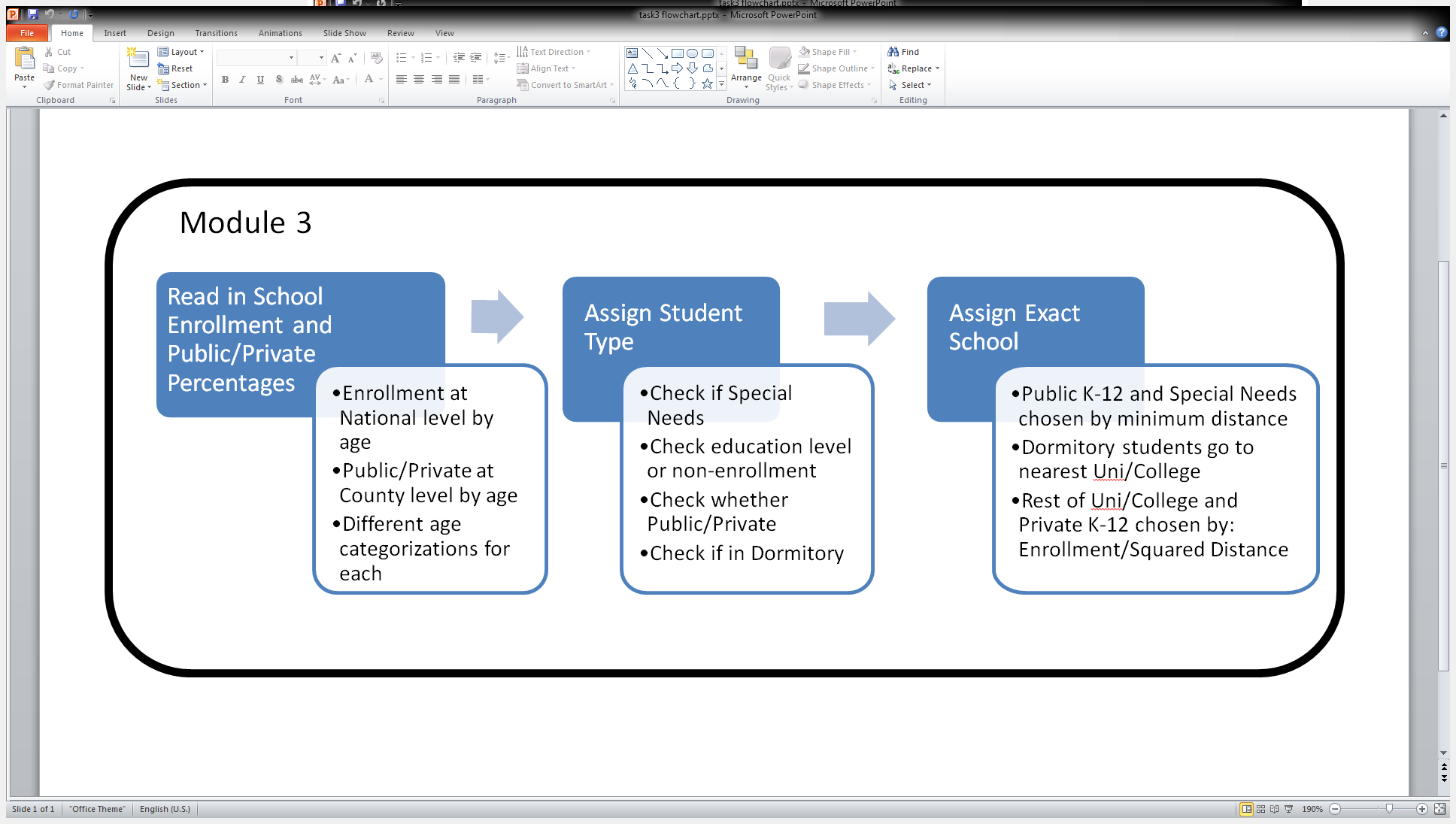


Figure 9 Process Chart of Task 3 Methods

Next, the resident's Household Type is checked. The assumption is made that residents in Group Quarters other than dormitories don't go to school or college for simplicity, though in reality a small number of juvenile detentions do allow children to go to school. Residents in dormitories (Household Type 6) are automatically assigned a School Type of 8, Non-commuter College/University. For Household Types of 0 or 1, more typical family or non-family residences, the enrollment distribution data and private versus public distribution data mentioned earlier are drawn from, based on age and county, to determine whether the student goes to, say, public middle school, private high school, or is perhaps not enrolled at all.

With a Student Type assigned to the resident in question, the module goes on to choose an exact school or educational institution which the resident goes to on a typical weekday morning. In its current build, the module reads from six files, each of which includes at least School Name, Position (latitude and longitude), and Total Student Enrollment. In the previous iteration of this project done by the class of ORF 467 Fall ’11, a separate file was used for each of the student types (other than 9) resulting in 8 files. While public school data and college/university data were readily accessible from the website of the New Jersey Department of Education, private school data were poorly fabricated due to difficulty locating enrollment data at the time. Further research led to finding these data at the National Center of Education Statistics Website which performs a yearly survey of private schools across the nation. As such all private school data are currently in one file, bringing the number of school enrollment files to 6.

Different methods were used to pick different types of institutions. Public K-12 schools and Special schools were picked purely by shortest great circle distance from schools to the household; both in the same county. This assumption isn’t unreasonable as School Districts never cross County boundaries. The mapping is not one-to-one, however. Multiple districts do sometimes exist in the same County. Ostensibly, a mapping of Census Blocks to School Districts should be simple to make and use. However, when attempted using the SF1 data, the districts simply did not match those in the school file (which are most likely the correct ones); so, distance was used instead. Furthermore, to save time on repeated calculations, these distances could have been calculated only once per Census Block, however due to the relatively small number of public schools of each level within a single county, they are calculated every time for different students. One advantage of this, however, is that in future renditions where each household may be given a location different from and more accurate than just the Census Block centroid, the code for choosing public schools would still work.

Private schools and all places of higher education are picked randomly from a distribution whose values are produced by a function of student enrollment (i.e. school size) and distance, much like the attraction equation used to select industries as a function of frequency and income difference. It is as follows in Equation 3 for every college/university and private K-12 school.

Equation 3 Private School Attraction

Here GCD refers to Great Circle distance, a method of calculating distance between two points on a sphere; in this case using latitude and longitude coordinates and assuming the Earth’s radius to be 3963.17 miles to cater to coordinates in the Northeast of America.

Once an institution is picked for every enrolled student, a new Task 3 file is created for every county relisting all information from the previous two tasks and appending school information, specifically Student Type, School Name, a pointer into the school file(i.e. a row number), and the school's Latitude and Longitude coordinates. Sample output can be found in Figure 10 below.

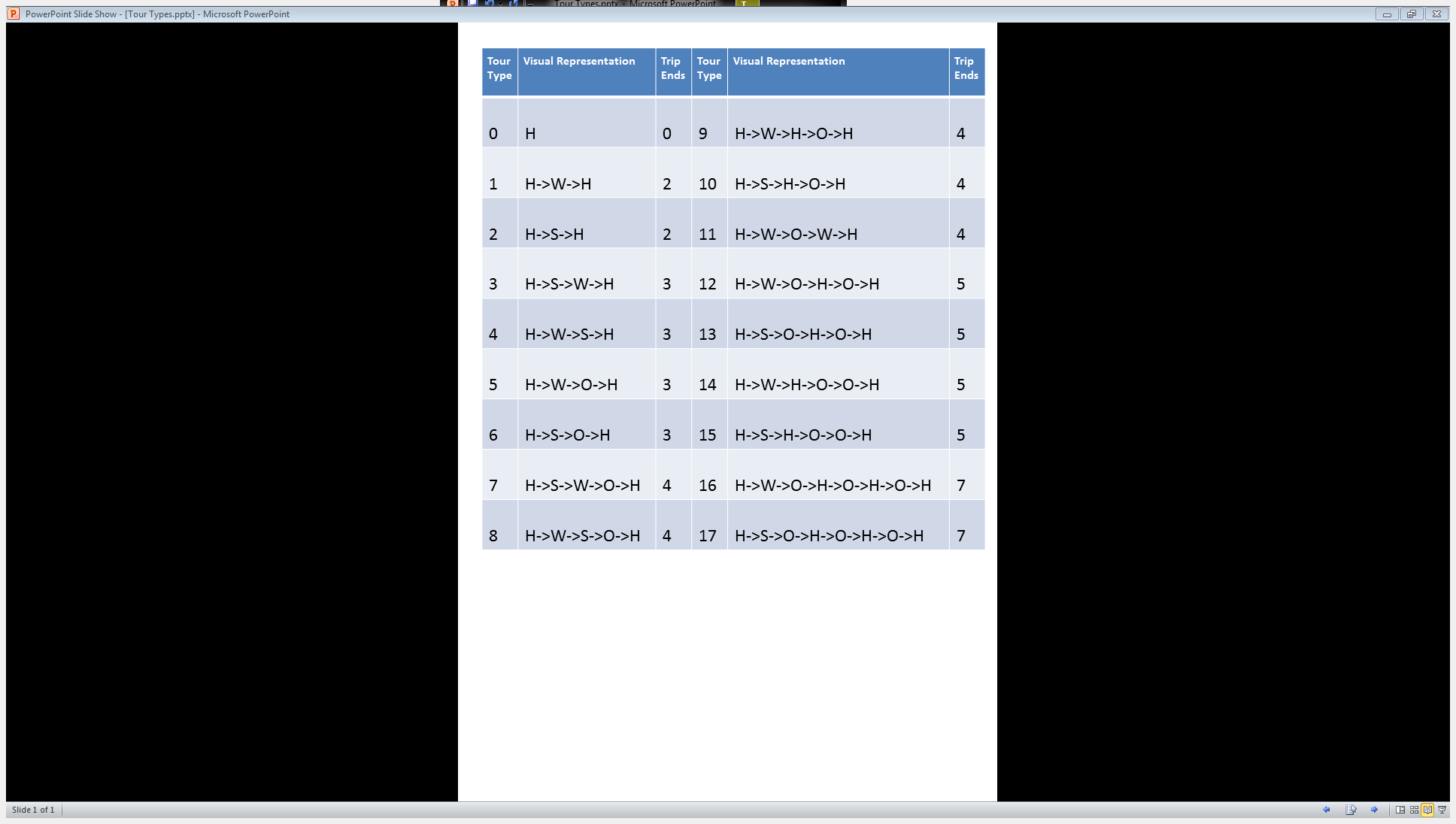
|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Res County** | **HH ID** | **HH Type** | **Res Lat** | **Res Long** | **Person ID** | **Age** | …..  …. | **School Type** | **School Pointer** | **School Name** | **School Lat** | **School Long** |
| 1 | 1 | 0 | 39.3578934 | -74.4607536 | ATL00000001 | 28 | 9 | -1 | -1 | -1 | -1 |
| 1 | 1 | 0 | 39.3578934 | -74.4607536 | ATL00000002 | 69 | 9 | -1 | -1 | -1 | -1 |
| 1 | 1 | 0 | 39.3578934 | -74.4607536 | ATL00000003 | 32 | 9 | -1 | -1 | -1 | -1 |
| 1 | 1 | 0 | 39.3578934 | -74.4607536 | ATL00000004 | 28 | 9 | -1 | -1 | -1 | -1 |
| 1 | 1 | 0 | 39.3578934 | -74.4607536 | ATL00000005 | 16 | 2 | 1 | Atlantic City High School | 39.369436 | -74.47558 |
| 1 | 2 | 0 | 39.3578934 | -74.4607536 | ATL00000006 | 44 | 9 | -1 | -1 | -1 | -1 |
| 1 | 2 | 0 | 39.3578934 | -74.4607536 | ATL00000007 | 4 | 9 | -1 | -1 | -1 | -1 |
| 1 | 2 | 0 | 39.3578934 | -74.4607536 | ATL00000008 | 65 | 9 | -1 | -1 | -1 | -1 |
| 1 | 2 | 0 | 39.3578934 | -74.4607536 | ATL00000009 | 8 | 0 | 5 | Chelsea Heights Elementary School | 39.355605 | -74.463162 |
| 1 | 3 | 0 | 39.3578934 | -74.4607536 | ATL00000010 | 44 | 9 | -1 | -1 | -1 | -1 |
| 1 | 3 | 0 | 39.3578934 | -74.4607536 | ATL00000011 | 2 | 9 | -1 | -1 | -1 | -1 |
| 1 | 3 | 0 | 39.3578934 | -74.4607536 | ATL00000012 | 7 | 0 | 5 | Chelsea Heights Elementary School | 39.355605 | -74.463162 |
| 1 | 3 | 0 | 39.3578934 | -74.4607536 | ATL00000013 | 2 | 9 | -1 | -1 | -1 | -1 |
| 1 | 4 | 0 | 39.3578934 | -74.4607536 | ATL00000014 | 39 | 9 | -1 | -1 | -1 | -1 |

Figure 10 Sample Output of Module 3

## Task 4: Assigning Activity Patterns

Task 4 assigns every resident created in Task 1 a travel activity pattern for the day, represented by a tour, or sequence of trips. The goal here is to use the demographics generated for each resident to try to approximate what their activities might look like on a typical day and hence determine what trips they would make. As discussed in the section on Activity-Based Models above, there exists much research on simulating human activity in varying degrees of complexity that continues today. For the purpose of this model, which focuses more on disaggregation than on pure complexity, tours as seen in Table 4 are drawn randomly from discrete probability distributions that are conditioned on the type of traveler. These distributions, seen in Table 5, are generated manually and exogenously, and are elaborated on below.

Table 4 Activity Patterns



Module 4 reads in every row of the output files of the previous module and collects each resident’s Traveler Type and Student Type. From this information it first creates a revised Traveler Type simply to account for Traveler Types 1 and 2 (i.e. K-12 students) who in Task 3 were assigned as Student Type 9 (Not Enrolled). They are changed to Traveler Type 6 (Home-Worker). With this minor change made, this revised type is used to select the appropriate probability distribution of trips and then draws an Activity Pattern. These trip patterns are numbered from 0 to 17 as seen in Table 4 above, where H refers to home, W to work, S to school, and O to others. For Home-Workers, Tours that include W’s but not S’s treat W’s as O’s. For example, Activity Pattern 1, H->W->H, is equivalent to H->O->H. Tours which include a school will have probability 0 for a Home-Worker, just as they will for a Worker who does not attend school.

Table 5 Probability Distributions of Activity Pattern by Traveler Type

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Traveler Type | Do-Not-Travel | School-No-Work | School-Work | College | College-Work | Typical Worker | Home-Worker | Out-of-State |  |
| Activity Pattern | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **Trip Ends** |
| 0 | 1 | 0.01 | 0.01 | 0.005 | 0.005 | 0.004 | 0.075 | 0 | 0 |
| 1 | 0 | 0 | 0 | 0.0075 | 0.0075 | 0.05 | 0.15 | 0.6 | 2 |
| 2 | 0 | 0.125 | 0.05 | 0.0075 | 0.0075 | 0 | 0 | 0 | 2 |
| 3 | 0 | 0 | 0.405 | 0.2 | 0.2 | 0 | 0 | 0 | 3 |
| 4 | 0 | 0 | 0 | 0.2 | 0.2 | 0 | 0 | 0 | 3 |
| 5 | 0 | 0 | 0 | 0.0075 | 0.0075 | 0.196 | 0.15 | 0.3 | 3 |
| 6 | 0 | 0.35 | 0.085 | 0.0075 | 0.0075 | 0 | 0 | 0 | 3 |
| 7 | 0 | 0 | 0.45 | 0.26 | 0.26 | 0 | 0 | 0 | 4 |
| 8 | 0 | 0 | 0 | 0.26 | 0.26 | 0 | 0 | 0 | 4 |
| 9 | 0 | 0 | 0 | 0.0075 | 0.0075 | 0.15 | 0.1 | 0 | 4 |
| 10 | 0 | 0.325 | 0 | 0.0075 | 0.0075 | 0 | 0 | 0 | 4 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0.15 | 0.125 | 0.1 | 4 |
| 12 | 0 | 0 | 0 | 0.0075 | 0.0075 | 0.15 | 0.125 | 0 | 5 |
| 13 | 0 | 0.15 | 0 | 0.0075 | 0.0075 | 0 | 0 | 0 | 5 |
| 14 | 0 | 0 | 0 | 0.005 | 0.005 | 0.15 | 0.125 | 0 | 5 |
| 15 | 0 | 0.025 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 16 | 0 | 0 | 0 | 0 | 0 | 0.15 | 0.15 | 0 | 7 |
| 17 | 0 | 0.015 | 0 | 0.01 | 0.01 | 0 | 0 | 0 | 7 |
| Total | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |
| Average Trips |  | 3.58 | 3.37 | 3.585 | 3.585 | 4.438 | 3.95 | 2.5 |  |

The distributions were made with two main points in mind. That the average New Jersey resident makes between 3.5 and 4.5 trips a day (note that this doesn’t apply to our Out-of-State workers), and that the probabilities of an Activity Pattern across different Traveler Types make sense based on our definitions of the categories.

Those who don’t travel (TT 0) indeed have only one possible tour, staying at home. School children who do not work (TT 1) are most likely to only take only one recreational (O) trip, and are more likely to take them right after school than later in the night. School aged kids who may work (TT 2) always go to school before work, and otherwise follow similar priorities to TT 1 children.

It was decided that both commuter and non-commuter college students (TT 3 and TT 4) should have the same probability distributions due to lack of clear distinction at the level of detail with which we’re working with. For example even when living in student housing near the university, there may still be a trip to school for students at certain universities. Indeed taking Household Type and particular schools into account could allow for a more nuanced distinction in travel patterns of commuter and non-commuter students in future renditions of this model – see the Limitations and Next Steps section for an elaboration on how this could be done.

Typical Travelers, or typical workers (TT 5), act quite similarly to students who do not work (TT 1) in that they are most likely to take one recreational trip right after Work. One unique tour of theirs is that they may make an Other trip and then return to work afterwards. This trip shall be limited to a certain radius, as shall be discussed in Task 5. This is to simulate a trip either for lunch or to run an errand. Lastly, Home-Workers (TT 6) are both more likely to stay at home all day, as well, as more likely to recreate overall, though the lack of work trips bring their average number of trips below that of workers.

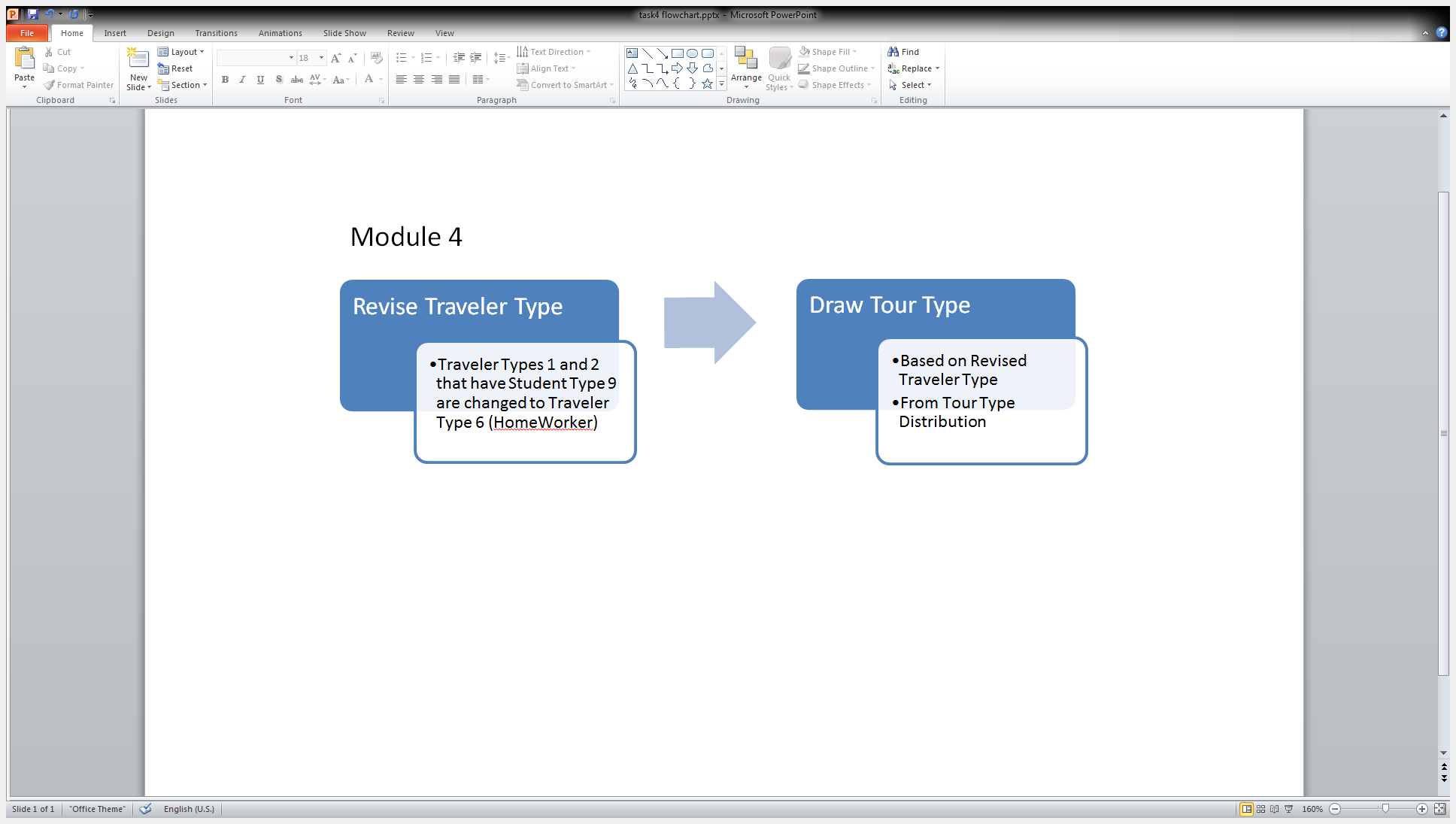


Figure 11 Process Chart of Module 4

With Activity Patterns chosen, the model then writes a new file that includes all past information from the previous tasks, in addition to appending the revised Traveler Type and the Activity Pattern from 0-17. These columns can be seen in the sample output in Figure 11 of Task 5 below.

## Task 5: Assigning Destinations for Other Trips

Task 5 brings the synthesizer one step closer to conclusion. In fact, it performs the final actions relevant to the spatial distribution of trips generated. To do so, it generates the Other trips that the resident makes, in addition to reading the locations of homes, schools and businesses as well as other information from the files generated in Task 4. It then appends to them 5 attributes for every trip the person makes, namely Trip Type, County Code, Pointer, Latitude, Longitude, and either School Type, Work Industry, or Patronage Industry, depending on the type of trip.

Trip Type is designated by one of H, W, S, and, O, or Home, Work, School, and Other. County Code (FIPS number) expresses the county relevant to the trip end. This is often the same as the residence county, since many people work and/or go to school within the county they live in. The Pointer field contains row numbers that each refer to the relevant place in the appropriate file of the appropriate county. For example, to use a school pointer (Output Column Index 21), one must first check the school type (Output Column Index 19) to know which school file to look in. Latitude and longitude are once again simply those of the places drawn in earlier modules. Similarly, School Type, Work Industry, and Patronage Industry are exactly as they are chosen in modules 3, 4 and 5, respectively. This repetition is to allow these attributes to be listed separately, creating a succinct but comprehensive Trip File. Before it can build this, however, the module must first assign all Other trips.

For any single resident, all Other trips—of which more than one are often chained in succession—are chosen based either on Residence County or Work County. Activity Patterns which do not include an Other trip between work trips base their Other trips on Residence County. In this case, places of patronage are chosen from counties that are at most 1 county away from the Residence County. These Neighboring Counties are listed in Table 6 below. The only exception to this is for Other trips that are in between work, which can be thought of as a lunch time break (Activity Pattern 11); these take place in the Work County instead. Such Other trips can sometimes turn up no venue within the set distance (between approximately .5 and 5 miles); in this instance, the closest place of patronage is chosen instead. Other trips made in tours other than type 11 are only constrained to having a minimum distance equivalent to half a mile. Trips shorter than this are beyond the scope of this project as most such trips would likely be taken without any motorized transport.

Table 6 Neighboring (1-adjacent) Counties

|  |  |  |  |
| --- | --- | --- | --- |
| County Name | FIPS Code | Custom Code | Neighboring Counties (FIPS) |
| Atlantic | 1 | 0 | 9, 11, 5, 7, 15, 29 |
| Bergen | 3 | 1 | 31, 17, 13 |
| Burlington | 5 | 2 | 1, 7, 25, 29 |
| Camden | 7 | 3 | 5, 15, 1 |
| Cape May | 9 | 4 | 1, 11 |
| Cumberland | 11 | 5 | 9, 1, 15, 33 |
| Essex | 13 | 6 | 31, 17, 3,27, 39 |
| Gloucester | 15 | 7 | 1, 33, 11, 7 |
| Hudson | 17 | 8 | 3, 13, 39 |
| Hunter | 19 | 9 | 41, 27, 35, 21 |
| Mercer | 21 | 10 | 19, 35, 23, 27, 5 |
| Middle | 23 | 11 | 25,21, 35,39 |
| Monmouth | 25 | 12 | 29, 5, 21, 23 |
| Morris | 27 | 13 | 35,19, 41, 37, 31, 13,39 |
| Ocean | 29 | 14 | 1, 5, 25 |
| Passaic | 31 | 15 | 3, 13, 27, 37 |
| Salem | 33 | 16 | 11, 15 |
| Somerset | 35 | 17 | 19, 21, 27, 39,23 |
| Sussex | 37 | 18 | 31, 27, 41 |
| Union | 39 | 19 | 17, 23, 13, 27, 35 |
| Warren | 41 | 20 | 19, 27,37 |

The module begins by reading in a data set of neighboring counties, as shown in Table 6, followed by each Patronage file by county. These files are in fact built from the same files used to select employer in Task 2, however they also have daily patron numbers for each business that has patrons such as banks or restaurants. For more information on these numbers, see the Employers and Patronage Data section. Then for each county, a larger list of all places of patronage is built by simply joining together all patronage lists that someone in that county may visit. So for example, a homework from Sussex County (37) might go to any of Morris(27), Passaic(31), Warren(41) or Sussex itself.

Next it runs through every resident in that county, for each of whom it reads the Activity Pattern. Each Activity Pattern has corresponding representation in the form of a pattern of H’s, W’s, S’s, O’s, and, possibly NO’s. Once this pattern is attained for the particular resident, the code runs through each one and appends the 5 attributes mentioned earlier. When the code reads an H, it appends ‘H’, County Code (Index 0), -1 (since there is no pointer to homes), and the Latitude (Index 3) and Longitude (Index 4) of the centroid of the Census Block. This forms a complete spatial representation of this point in the resident’s day of travel.

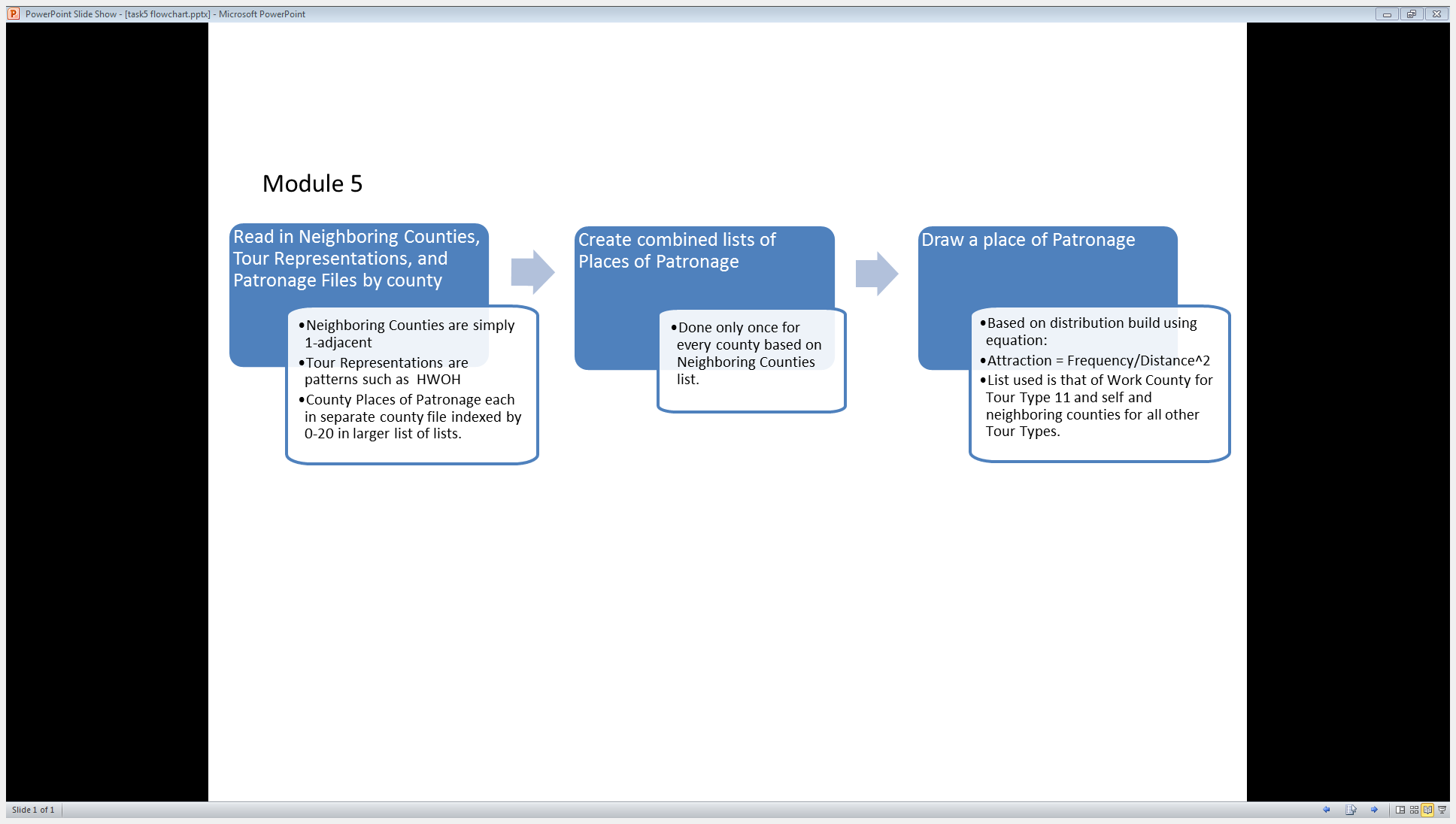


Figure 12 Process chart of Module 5

When the module reads an ‘O’, it then calls a function to draw an Other trip through the following method. The function uses a parameter to indicate whether the distance constraint for Other trips is set to above 0.5 miles as it is generally, or between 0.5 and 5 miles for the case of ‘lunch’ trips. In the latter case, the vector of patronage frequencies from which a place is randomly drawn is calculated for every resident of Activity Pattern 11. To do so, the module calculates the distance from the resident’s workplace to each possible place of patronage in the work county. The logic here is that it is unlikely to cross over county borders simply to get lunch or recreate in the duration of a lunch break. For all other cases, such distances are calculated once per Census Block since all households in that block have been given the same coordinates as those of the block in the 2010 Census Summary File. This is simply to save on processing time. It should also be noted that in order to

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Res County** | **HH ID** | **HH Type** | **Res Lat** | **Res Long** | **Person ID** | **Age** | **Sex** | **Traveler Type** | **…** | **Revised Traveler Type** | **Trip Type** | **Node 1: Node Type** | **Node 1: County** | **Node 1: Pointer** | **Node 1: Lat** | **Node 1: Long** | **Node 1: School Type/Work Ind/Patr Ind** | **Node 2: Node Type** | **Node 2: County** |
| 1 | 1 | 0 | 39.35789 | -74.4608 | ATL00000001 | 42 | 1 | 5 | … | 5 | 5 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | W | 1 |
| 1 | 1 | 0 | 39.35789 | -74.4608 | ATL00000002 | 13 | 0 | 1 | … | 1 | 2 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | S | 1 |
| 1 | 1 | 0 | 39.35789 | -74.4608 | ATL00000003 | 34 | 0 | 5 | … | 5 | 9 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | W | 1 |
| 1 | 1 | 0 | 39.35789 | -74.4608 | ATL00000004 | 3 | 0 | 0 | … | 0 | 0 | H | 1 | -1 | 39.35789 | -74.4608 | -1 |  |  |
| 1 | 1 | 0 | 39.35789 | -74.4608 | ATL00000005 | 7 | 1 | 1 | … | 1 | 10 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | S | 1 |
| 1 | 2 | 0 | 39.35789 | -74.4608 | ATL00000006 | 29 | 1 | 5 | … | 5 | 14 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | W | 1 |
| 1 | 2 | 0 | 39.35789 | -74.4608 | ATL00000007 | 16 | 1 | 2 | … | 6 | 16 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | W | 1 |
| 1 | 2 | 0 | 39.35789 | -74.4608 | ATL00000008 | 41 | 1 | 6 | … | 6 | 9 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | O | 29 |
| 1 | 2 | 0 | 39.35789 | -74.4608 | ATL00000009 | 27 | 1 | 5 | … | 5 | 9 | H | 1 | -1 | 39.35789 | -74.4608 | -1 | W | 1 |

Figure 13 Sample Output of Module 5.

reduce processing time, Euclidean L2 distance was used, rather than Great Circle Distance. For the purposes of using distance as a mere weight to create a distribution from, this is more than adequate, considering the relatively small range that the coordinates are spread over.

With distances and patronage frequencies available, the module draws from a distribution created in the following way.

Equation 4 Patronage Attraction

This is repeated until each letter in the resident’s trip chain is read and the row of the resident’s information is appended with the results of the module, Trip Type, County, Pointer, Latitude, Longitude, and Industry/School Type for each trip, for each tour, for each resident, for each county. A few row of abridged sample output are displayed above in Figure 11, which also shows Revised Traveler Type and Trip Type fields that were added in Module 4.

## Task 6: Adding the Temporal Dimension

The sixth and final task of the trip synthesizer assigns arrival and departure time for every node or location generated in Task 5, as well as distances between each. These three attributes are inserted, for every trip, after the five attributes appended in Task 5, bringing the number of attributes listed per trip to eight. These fields add a temporal aspect to what was, in-so-far, purely spatially distributed data. The purpose of this dimension within the scope of this project is simply to allow a basic analysis of the trips in time, as well as to aid visualization. With timestamps (in seconds after midnight) for every location, the trips can be visualized as filaments akin to GPS tracks, as demonstrated in Figure 14. The level at which the module generates timestamps, however, is not thoroughly deep or comprehensive but rather an alternative to leaving out the temporal aspect altogether.



Figure 14 Visualizing Trip Filaments Using a Google Earth Application[[7]](#footnote-7)

The module first reads in a file of the character representation of each Activity Pattern exactly as was done in the previous two modules. This adds a robustness to the modules should the Activity Patterns be changed in future versions. That said, individual exceptions may still have to be made through logic in the code. In addition, it reads in schedule files for school and work, which list arrival and departure parameters by school type and industry, respectively. Currently, the parameters are those of a triangular distribution, minimum, maximum, and mode for each arrival and departure time for every school type and work industry. (Maybe move this to data section, replace with: See the Data section for an elaboration on these datasets, and see Limitations and Next steps for ways to improve them).

As before, Module 6 iterates over every row—representing a person—in every task 5 output file—representing each county. A simple process chart gives an overview of the task in Figure 11. For every resident, the module first calculates the distance of every trip made. Since, however, every node in the trip chain is going to be given distance for consistency, the first node, Home, is given a -1; as such only nodes that are trip ends have an actual distance in miles. The same goes for the Arrival Time at the first node and the Departure Time at the last node; all set to -1.

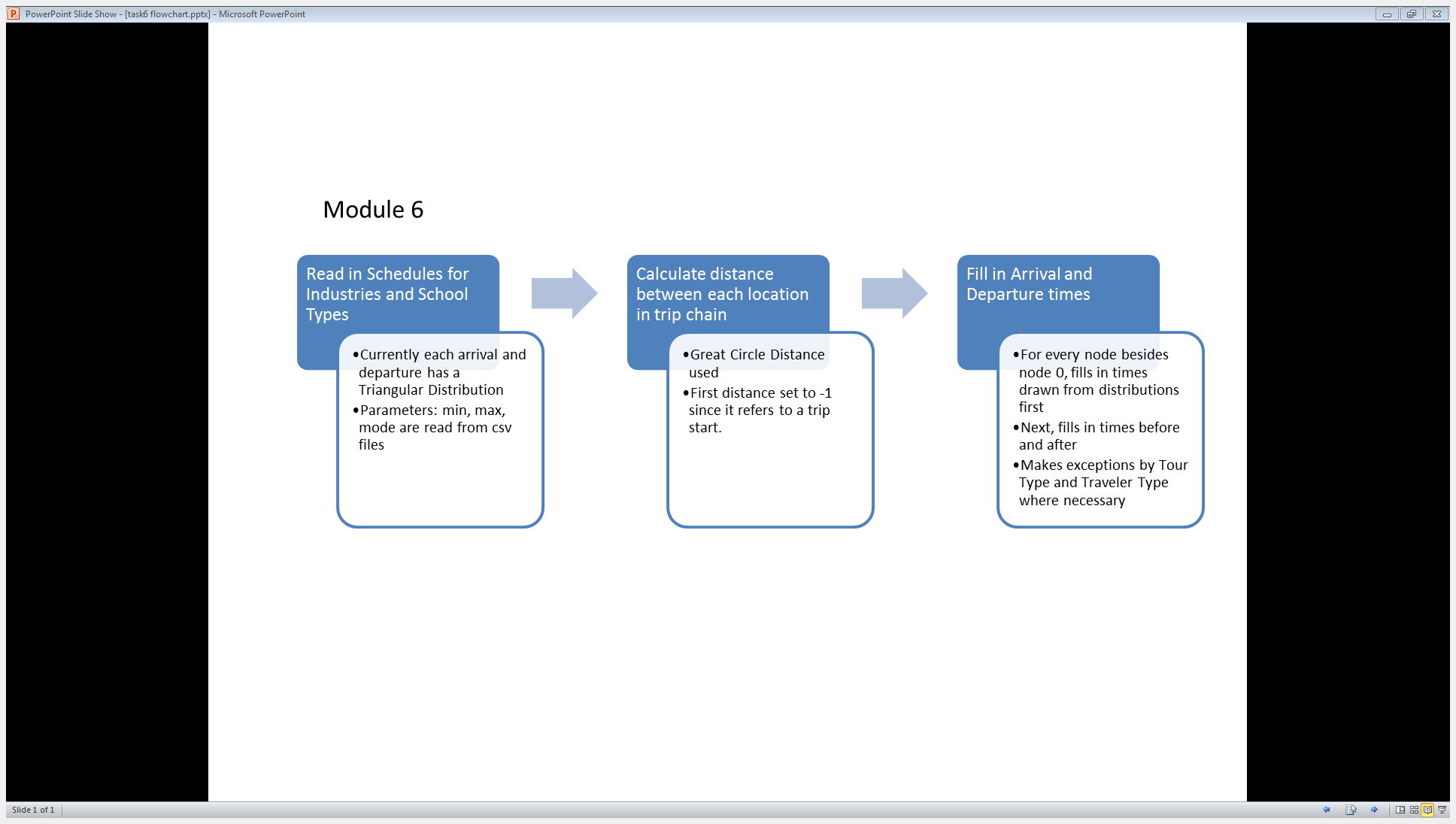


Figure 15 Process Chart of Module 6

Now for every resident, the Trip Representation is read and iterated over by a function that ultimately populates a list of Arrival Times and a list of Departure Times. The function essentially checks what type of location the current node is (H, W, or O) as well as attributes, like where in the chain it the node is, what Activity Pattern is being read, and the Traveler Type of the resident. In general, Arrival Times are either drawn from a distribution, as is the case for arriving at school or work, or they are calculated from the previous node’s Departure Time plus the time it takes to get there. This time is calculated simply by dividing the distance between the two nodes by an average speed. In this model only two such speeds are used, 15 mph for school trips and 30 mph for all other types. Departure Times are calculated very similarly in that they are either drawn or simply calculated either by subtracting from the Arrival Time of the next node (if known) or by adding a randomly drawn time spent at the location. The latter is used to decide how long a resident recreates at a certain O node. See the Limitations and Next steps section for an explanation of how this can be expanded to a whole distribution of patronage schedules by industry.

A few important exceptions in the function are worth noting as well. The function, *get\_times,* categorizes all its logic first by whether it is currently at second node (node/index 1) or whether it is at any node afterwards. In the former case, in between assigning the current node’s Arrival Time and its Departure Time, it calculates the previous node’s (always an H in this model) Departure Time by subtracting trip time, in the manner mentioned above. Arrival Time and Distance for the first node (node/index 0) are set to -1. Similarly, the Departure Time of the final node, also an H, is set to -1.

After node 1, another exception must be made for tours where the resident both works and goes to school, or vice versa. This is due to there not being a schedule for those who work and go to school. To make up for the current simplification, the function shortens both events and ensures that the second never begins before the first and so on.

With these Arrival Times, Departure Times, and Distances inserted in to each row in the appropriate trip columns, Module 6 outputs these rows to the final products of this trip synthesizer. Instead of a sample of this output a clear table of all output fields, their indices, and the module that generates them is listed below in Table 7.

Table 7 Output Fields by Module ordered by Field Index. Note that Module 6 inserts new fields rather than appending them.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Module | Field | Index | Module | Field | Index |
| Module 1 | Residence County | 0 | **Module 6** | Node 2: Arrival Time | 51 |
|  | HH ID | 1 |  | Node 2: Departure Time | 52 |
|  | HH Type | 2 |  | Node 2: Distance | 53 |
|  | Residence Latitude | 3 | **Module 5** | Node 3: Node Type | 54 |
|  | Residence Longitude | 4 |  | Node 3: County | 55 |
|  | Person ID | 5 |  | Node 3: Pointer | 56 |
|  | Age | 6 |  | Node 3: Latitude | 57 |
|  | Sex | 7 |  | Node 3: Longitude | 58 |
|  | Traveler Type | 8 |  | Node 3: School Type/Work Industry/Patronage Industry | 59 |
|  | Income Bracket | 9 | **Module 6** | Node 3: Arrival Time | 60 |
|  | Income | 10 |  | Node 3: Departure Time | 61 |
| Module 2 | Work County | 11 |  | Node 3: Distance | 62 |
|  | Work Industry | 12 | **Module 5** | Node 4: Node Type | 63 |
|  | Company Name | 13 |  | Node 4: County | 64 |
|  | Work Zip code | 14 |  | Node 4: Pointer | 65 |
|  | 3 digit NAICS | 15 |  | Node 4: Latitude | 66 |
|  | Work Latitude | 16 |  | Node 4: Longitude | 67 |
|  | Work Longitude | 17 |  | Node 4: School Type/Work Industry/Patronage Industry | 68 |
|  | Work Pointer | 18 | **Module 6** | Node 4: Arrival Time | 69 |
| Module 3 | Student Type | 19 |  | Node 4: Departure Time | 70 |
|  | School County | 20 |  | Node 4: Distance | 71 |
|  | School Pointer | 21 | **Module 5** | Node 5: Node Type | 72 |
|  | School Name | 22 |  | Node 5: County | 73 |
|  | School Lat | 23 |  | Node 5: Pointer | 74 |
|  | School Long | 24 |  | Node 5: Latitude | 75 |
| Module 4 | Revised Traveler Type | 25 |  | Node 5: Longitude | 76 |
|  | Trip Type | 26 |  | Node 5: School Type/Work Industry/Patronage Industry | 77 |
| Module 5 | Node 0: Node Type | 27 | **Module 6** | Node 5: Arrival Time | 78 |
|  | Node 0: County | 28 |  | Node 5: Departure Time | 79 |
|  | Node 0: Pointer | 29 |  | Node 5: Distance | 80 |
|  | Node 0: Latitude | 30 | **Module 5** | Node 6: Node Type | 81 |
|  | Node 0: Longitude | 31 |  | Node 6: County | 82 |
|  | Node 0: School Type/Work Ind/Patr Ind | 32 |  | Node 6: Pointer | 83 |
| Module 6 | Node 0: Arrival Time | 33 |  | Node 6: Latitude | 84 |
|  | Node 0: Departure Time | 34 |  | Node 6: Longitude | 85 |
|  | Node 0: Distance | 35 |  | Node 6: School Type/Work Ind/Patr Ind | 86 |
| Module 5 | Node 1: Node Type | 36 | **Module 6** | Node 6: Arrival Time | 87 |
|  | Node 1: County | 37 |  | Node 6: Departure Time | 88 |
|  | Node 1: Pointer | 38 |  | Node 6: Distance | 89 |
|  | Node 1: Latitude | 39 | **Module 5** | Node 7: Node Type | 90 |
|  | Node 1: Longitude | 40 |  | Node 7: County | 91 |
|  | Node 1: School Type/Work Ind/Patr Ind | 41 |  | Node 7: Pointer | 92 |
| Module 6 | Node 1: Arrival Time | 42 |  | Node 7: Latitude | 93 |
|  | Node 1: Departure Time | 43 |  | Node 7: Longitude | 94 |
|  | Node 1: Distance | 44 |  | Node 7: School Type/Work Ind/Patr Ind | 95 |
| Module 5 | Node 2: Node Type | 45 | **Module 6** | Node 7: Arrival Time | 96 |
|  | Node 2: County | 46 |  | Node 7: Departure Time | 97 |
|  | Node 2: Pointer | 47 |  | Node 7: Distance | 98 |
|  | Node 2: Latitude | 48 |
|  | Node 2: Longitude | 49 |
|  | Node 2: School Type/Work Ind/Patr Ind | 50 |

# Data

“Garbage in, garbage out…” - Anonymous

This section expounds upon the many data sets that are used by the synthesizer to model the characteristics and behaviors of New Jersey residents and out-of-state workers. Figure 16 contains a complete file tree displaying all files directly used and produced by the model. The section concludes with potentially useful data for future iterations or this or other synthesizers.

├───Code

├───Data

│ ├───ACS2010 5yr

│ ├───Census 2010

│ │ └───Census Matrices

│ ├───Employment

│ ├───Industry

│ ├───NHTS

│ ├───School

│ ├───JTW Flow

│ ├───Patron Employee

│ ├───School

│ │ └───Private

│ │ └───Public

│ └───Tours

│

├───Output

│ ├───Module 1

│ ├───Module 2

│ │ ├───2a

│ │ ├───2b

│ │ └───2c

│ ├───Module 3

│ ├───Module 4

│ ├───Module 5

│ └───Module 6

Figure 16 Tree Structure of Folders Relevant to Synthesizer

1. There exist a few blocks so lowly populated that this information is only available at tract level and not displayed at the block level, for privacy concerns. [↑](#footnote-ref-1)
2. (LED, 2012) [↑](#footnote-ref-2)
3. (LED, 2012) [↑](#footnote-ref-3)
4. (US Census Bureau, 2005) [↑](#footnote-ref-4)
5. The true average is closer to 2.5% (BLS, 2012) [↑](#footnote-ref-5)
6. Note that such attraction equations are used frequently throughout the model in this thesis and their results and limitations are discussed in their respective sections. [↑](#footnote-ref-6)
7. Image Source: M. Yaroshefsky [↑](#footnote-ref-7)