DRAFT User's Manual

NJDOT 2012-13: *Integration of Bus Stop Count Data with Census data for Improving Bus Service and Efficiency*

Prepared for New Jersey Department of Transportation

Albany Visualization And Informatics Lab (AVAIL)

Lewis Mumford Center, University at Albany, SUNY

8/13/2014

Catherine T. Lawson, Ph.D.

Chair, Geography and Planning Department

Director, Lewis Mumford Center

AS 218 1400 Washington Ave

Albany, New York 12222

TABLE OF CONTENTS

1 TABLES AND FIGURES

2 Overview

2.1 Transit Demand Modeling

This collection of tools and methodologies are intended to allow planners to assess changing transit demand in customizable market areas defined simply by GTFS routes and census geographies. The web-tool designed by AVAIL aggregates a number of data sets which are universally available in the US, such as the American Community Survey (ACS), The Census Transportation Planning Products (CTPP) and The Longitudinal Employment and Household Dynamics (LEHD) Data, with data generated by transit agencies like GTFS and ridership surveys. These data sets are then run through an algorithm to approximate bus ridership. A software called Open Trip Planner is then used to microsimulate bus ridership in a given market area. The collection of tools and methodologies together, illuminate dynamics of bus ridership in a given area.

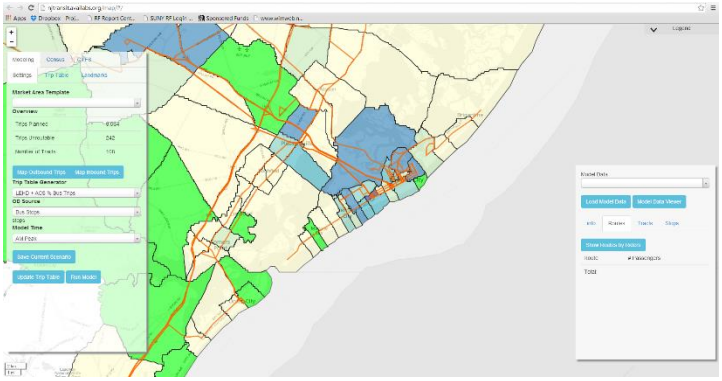


Figure 1 Map Viewer

2.2 User's Guide

This User's Guide is broken into two sections:

1. Admin Tool: This section discusses every functionality of the web tool guiding a user through the process of uploading data, creating market areas, running models, and reviewing model outputs.

Figure 2 Admin Tool

2. Modeling: This section contains a summary of data elements and the regression model specifications used in the microsimulation process, including the sources of data, descriptive statistics, the correlations coefficients, regression model development and latest regression models (as of 8.13.14) for the three market areas; Atlantic City, Princeton/Trenton, and Paterson.

Atlantic City				
	Atlantic City1	Atlantic City2	Atlantic City3	**Atlantic City4**
Dependent Variable	bus_to_wor	bus_to_wor	bus_to_wor	**bus_to_wor**
Constant	4.47	-25.17	-13.85	**-19.36**
CAR_0	.77**	0.69**	.63**	**.54****
CAR_1		0.10**	0.13**	**.13****
INFORMATIO			-0.79**	**-.71***
EMP_DEN				**.00***
R Sq.	0.57	0.59	0.62	**0.63**
N	108	108	108	**108**

Table 1 Regression Modelling

3 ADMIN TOOL

The Admin Tool contains three sections: a navigation panel on the left; a content section in the middle; and a user's guide at right. As a user navigates throughout the website, the User's Guide at right changes to reflect the functionalities located in the center pane. The right panel User's Guide provides information about how to use each of the tools located in the center panel while never navigating away from those tools.

The Admin Tool (see Figure 1, below) is the landing page or "**Dashboard**" which users will see when they first sign in. This figure illustrates clearly the Left Navigation Pane, the Center Pane, and the User's Guide at Right. A user can access a PDF of the entire User's Guide by clicking on the "**User's Guide**" in the Left Navigation Pane.

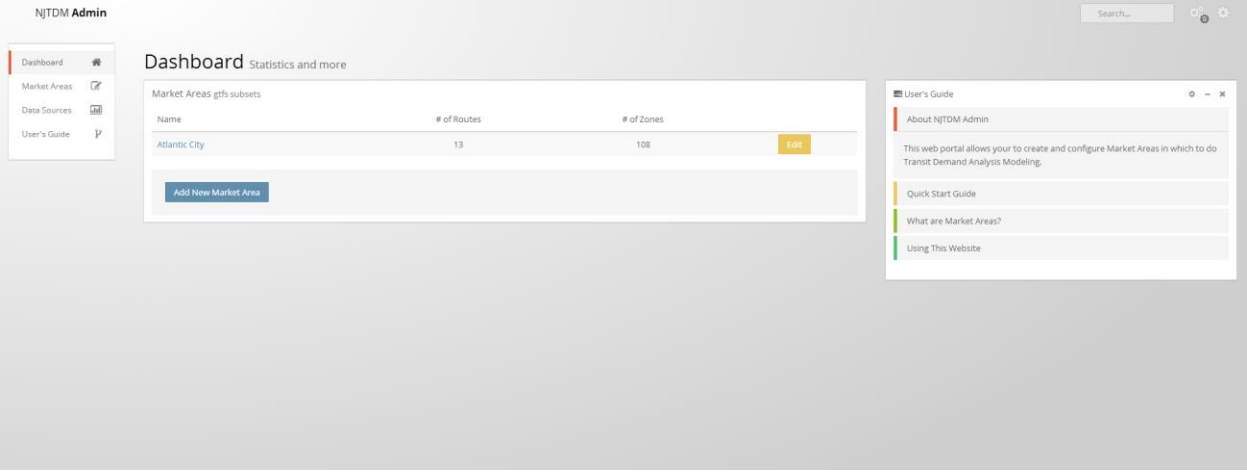


Figure 3 Admin Tool: Dashboard

3.1 USER'S GUIDE - RIGHT PANE

The Admin Tool contains a User's Guide in the Right Side Panel (see Figure 4 at right). This User's Guide contains information and instructions related to the functionalities of the Admin Tool's Center Pane. The category titles are named for each of the functionalities and columns located in the Center Pane.

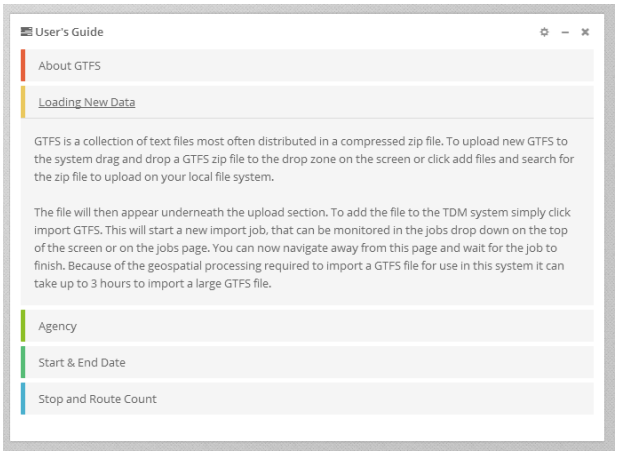


Figure 4 Admin Tool: Right Pane User's Guide

3.2 Left Navigation Pane

Before discussing the Center Pane functionality tools or their corresponding Right Pane User's Guide sections, the following section describes the full suite of functions in the Left Navigation Pane.

3.2.1 Left Navigation - Market Areas

By Clicking Market Areas you open a list of each of the Cities in your web tool. Below each city is an Overview, Model Runs, and access to the Map Tool.

3.2.2 Left Navigation - Data Sources

By Clicking on **Data Sources**, you open a list of each of the data sources used in building the web tool and it allows you to upload new data.

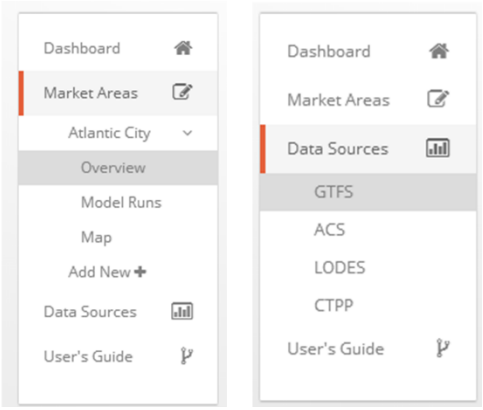


Figure 5 Admin Tool: Left Navigation Pane

3.3 Dashboard

Upon opening the Admin Tool, users will find the Dashboard page with the three navigation panes discussed above. The Center Pane is currently set to land on "**Market Areas**." At this time, the only market area in the web tool is Atlantic City.

3.3.1 Dashboard - Center Pane

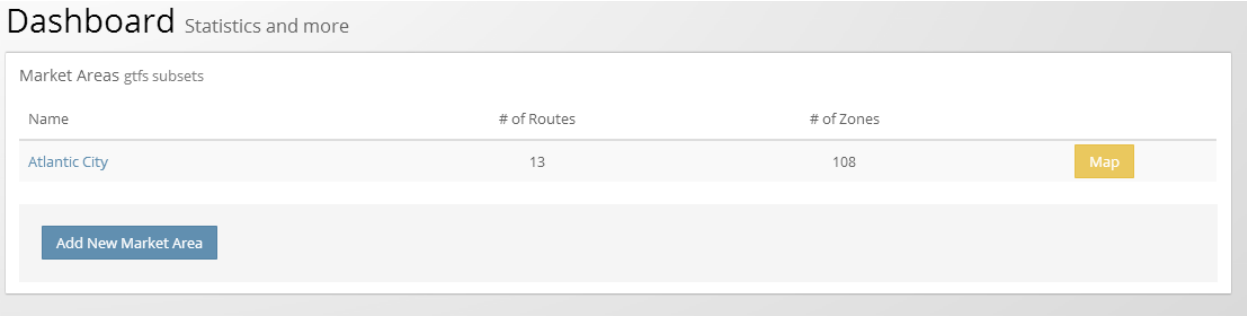


Figure 6 Admin Tool: Dashboard Center Pane

The Center Pane has three functionalities:

1. By Clicking on the name of a city, a user can access the Overview for that city
2. The "**Add New Market Area**" Button is for uploading new Market area
3. The "**Map**" Button navigates to the Map Tool

3.4 Market Areas

When you click on "**Market Areas**" in the Left Navigation Panel you'll get a dropdown list of all of the cities listed in your tool and an "**Add New**" option. At this time, the only city in this tool is Atlantic City.

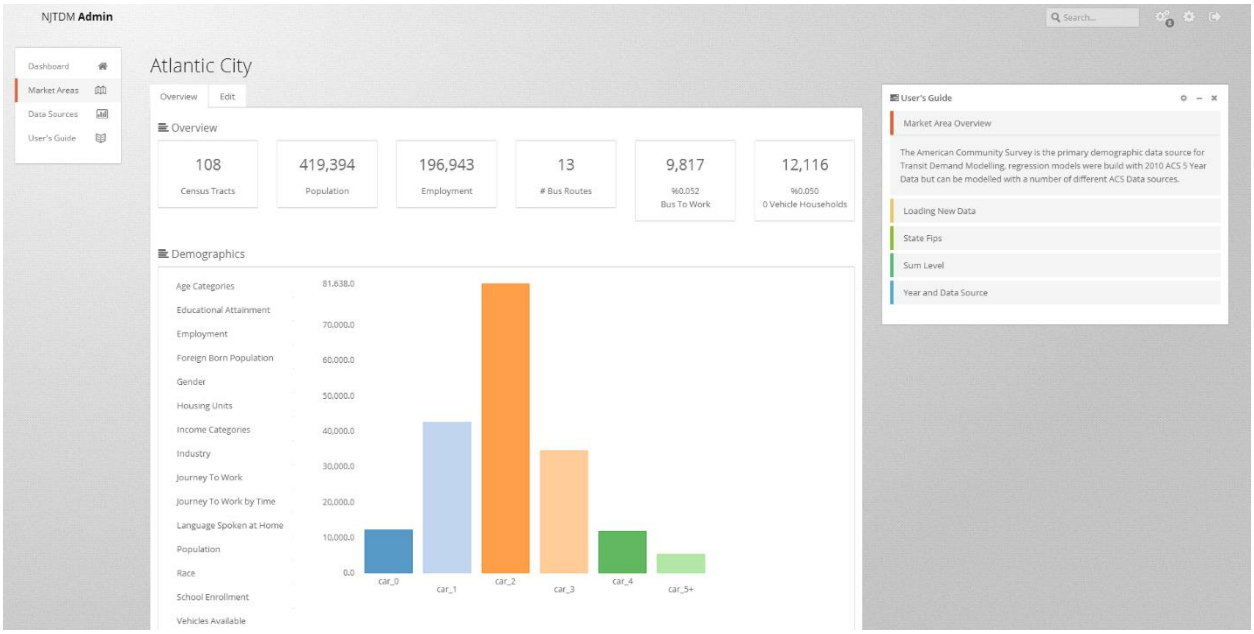
By clicking on Atlantic City, the user opens a dropdown menu that includes three sections:

1. Overview
2. Market Runs
3. Map View

3.4.1 Overview (Market Areas>Atlantic City)

The Overview Section gives a graphic overview of a Market Area's demographics.

Figure 7 Admin Tool: Market Area Overview



Across the top of the Overview Section, you will find a left to right tab function. The two tabs in the overview section are titled "**Overview**" and "**Edit.**" The tab titled "Overview" is the default landing page for this section.

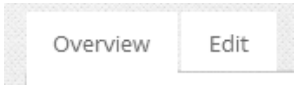


Figure 8 Admin Tool: Market Area Overview Tabs

3.4.1.1 Overview Tab (Market Areas>Atlantic City>Overview)

In the Center Pane of the Overview Section, there is a left to right feature. This feature provides broad market area statistics relevant to transit modeling.

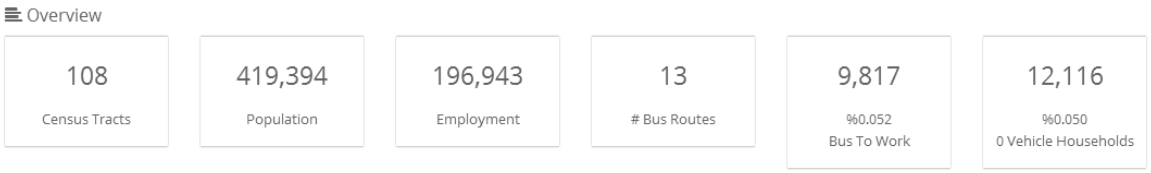


Figure 9 Admin Tool: Market Area Overview Census Data Table

The data listed left to right above the graph includes: Census Tracts, Population, Number of Bus Routes, Number of people indicating "Bus to Work" on the Census, and Number of Zero Car Households.

On the left side of the Overview Center Pane, there is a column of data. The categories listed here can be clicked on and they will change the graph in the center. The data in this column includes:

Age Categories, Education Attainment, Employment, Foreign Born Population, Gender, Housing Units, Income Categories, Industry, Journey to Work, Journey to Work by Time, Language Spoken at Home, Population, Race, School Enrollment and Vehicles Available.

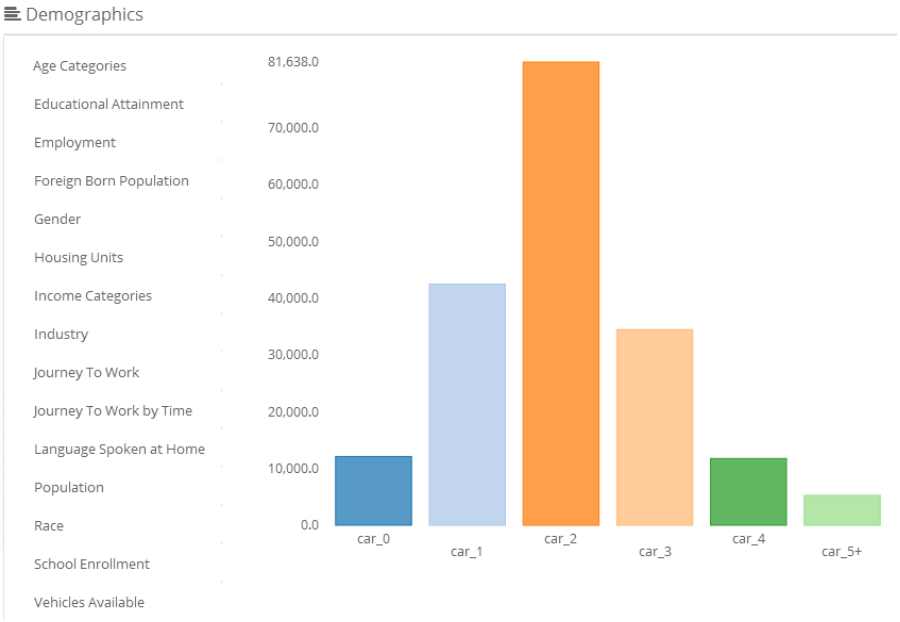


Figure 10 Admin Tool: Market Area Overview Census Data Graph Section

3.4.1.2 The Edit Tab (Market Areas>Atlantic City>Overview>Edit)

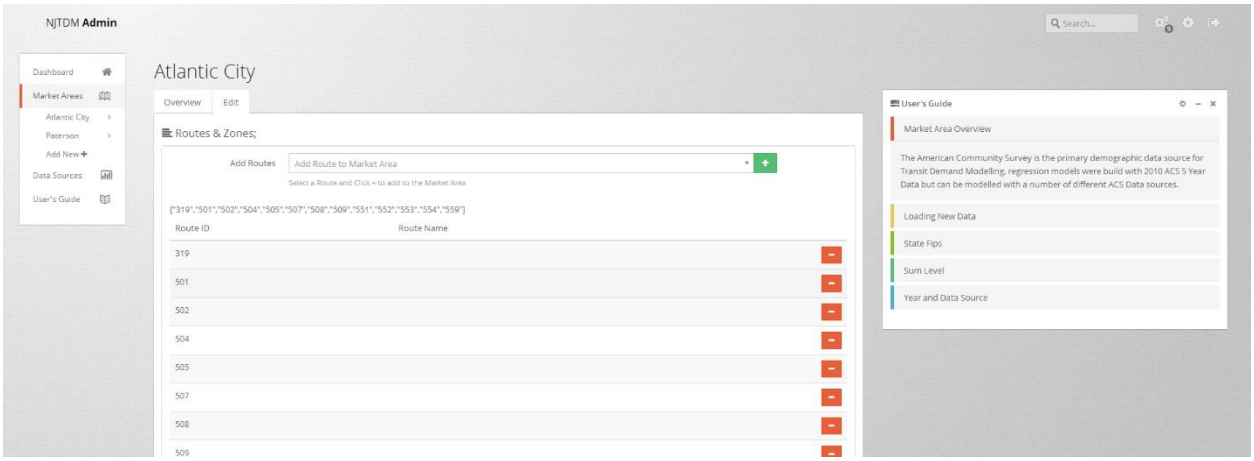


Figure 11 Admin Tool: Edit Market Area Tab

The Routes Tab section allows you to add (Green + Button) or remove (Red – Button) a bus route.

The Routes Tab also allows you to search all of the bus routes in a given Market Area in a smart dropdown menu. You can either search all bus routes or narrow your search by typing one of the numbers in the bus route you would like to use.

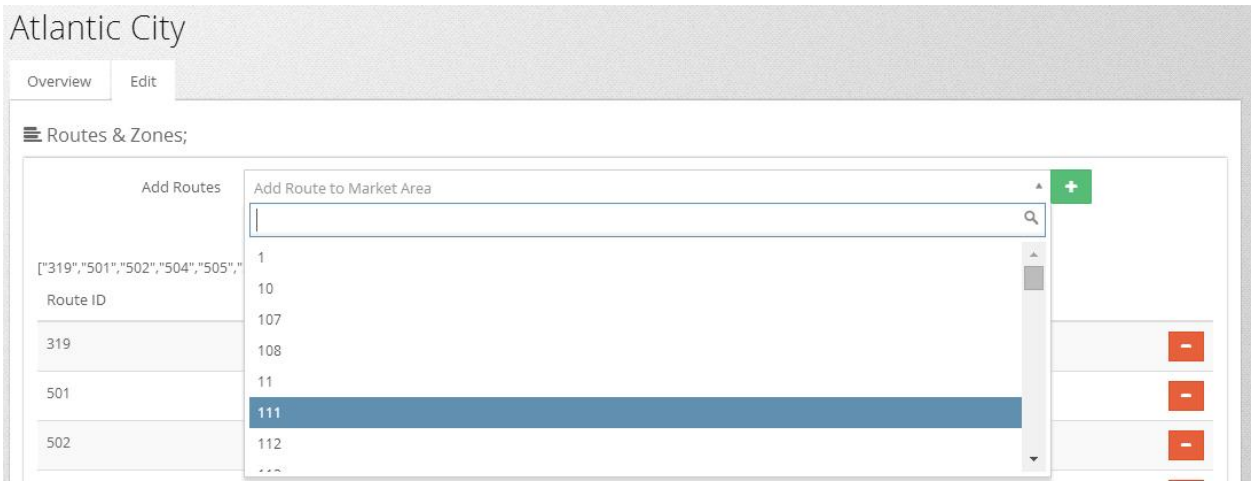


Figure 12 Admin Tool: Edit Bus Routes Dropdown Menu

3.4.2 Add Market Area

You can add new market areas by following the step by step process of the **Add Market Area Wizard**.

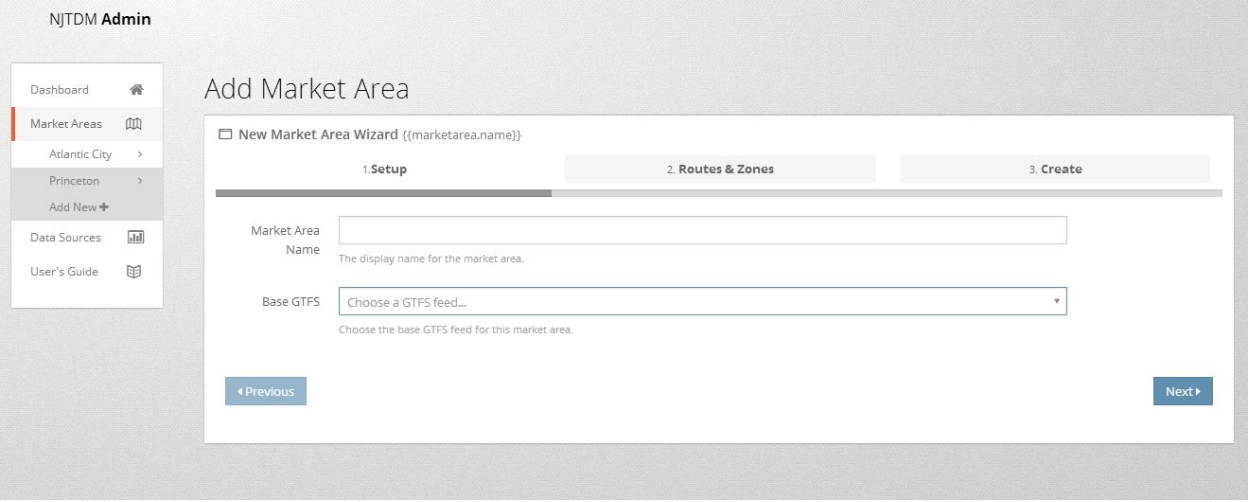


Figure 13 Admin Tool: Add Market Area

3.4.2.1 Step 1 - Name the Market Area

The first step is to give your market area a name and choose a base GTFS code. Then click next.

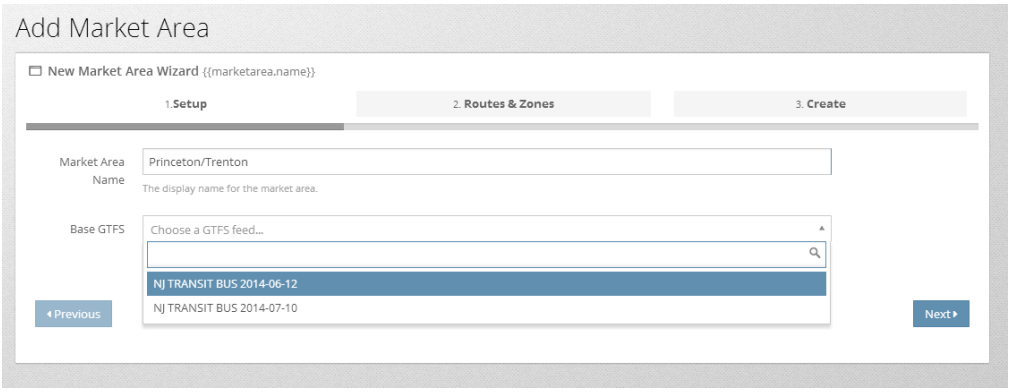


Figure 14 Admin Tool: Add Market Area Wizard

3.4.2.2 Step 2 - Routes & Zones.

Here you'll need some knowledge of your market area's bus routes. You'll need to type in the route numbers then click on the green plus symbol to add it to your market area.

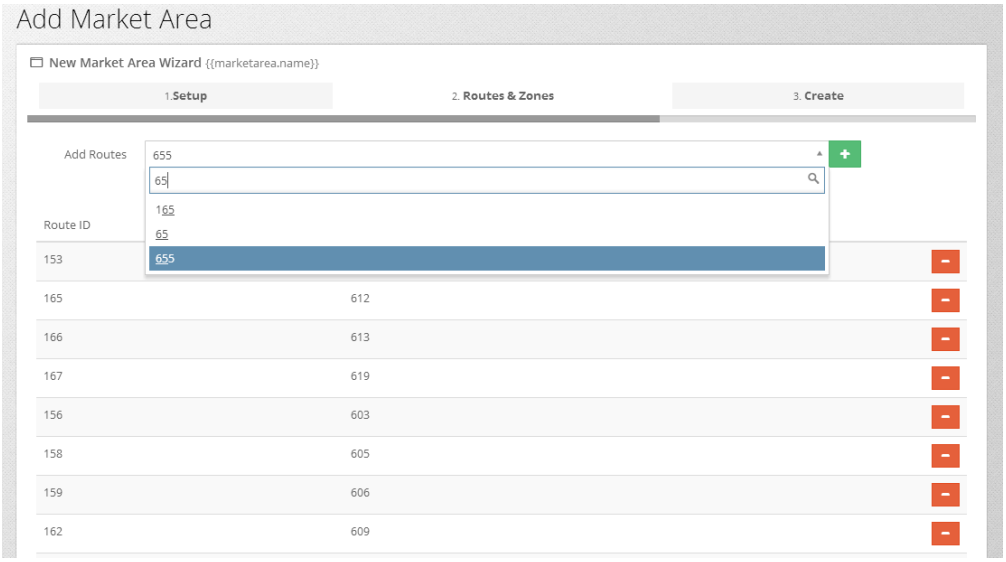


Figure 15 Admin Tool: Add Market Area Wizard Step 2

Each Route you add to the Market Area will be included as a red line in the map below. Each route you add will include its corresponding census tracts as seen in Figure 16 as black lines. After viewing the map of census tracts and bus routes click **Next** to go to step 3.

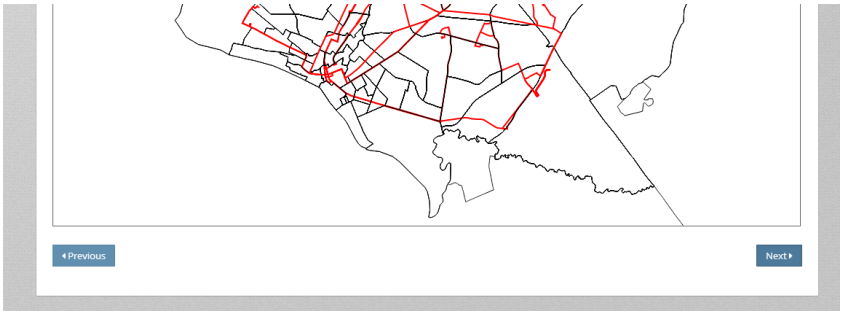


Figure 16 Admin Tool: Add Market Area Wizard Step 2 Routes and Zones Map

3.4.2.3 Step 3 - Review Market Area and Create

Here you'll view the number of routes and census tracts in your new Market Area. When you click Finish you will be brought to the Market Area > Overview page for your new Market Area.

Figure 17 Admin Tool: Add Market Area Wizard Step 3 Finish

3.4.3 Model Analysis

The Model Analysis section of the Admin Tool allows you to view graphs and data about model you've run in the Map View (See section 3.6). To analyze a model click on the dropdown menu, locate the model you like to review and click **Load Model**. After loading your model a graph will appear.

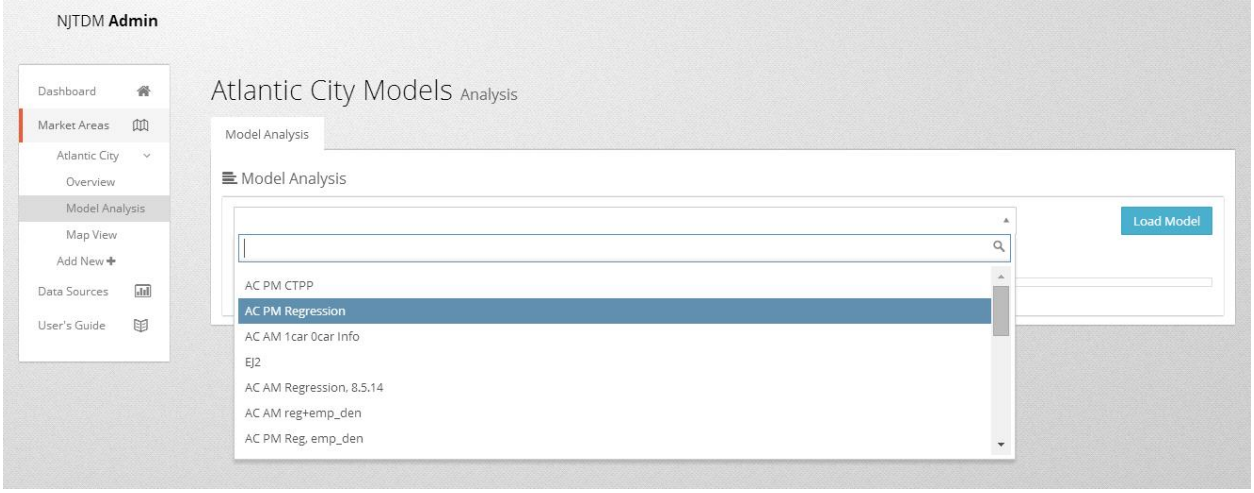
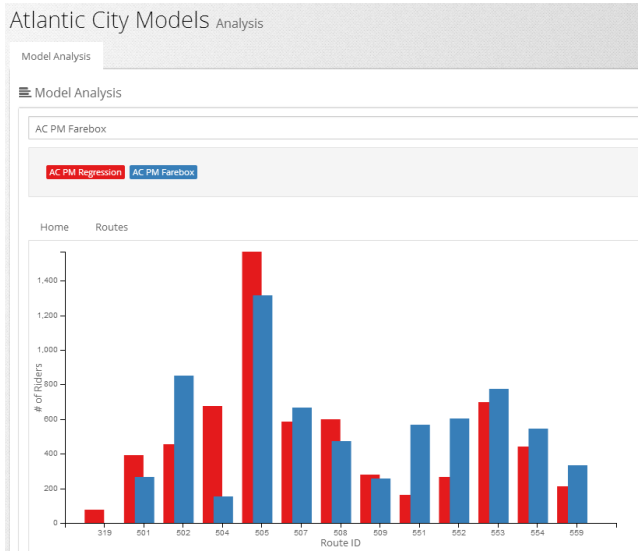


Figure 18 Admin Tool: Model Analysis

Figure 19 shows a second data set (AC PM Farebox). This was added to the Model Analysis view in order to compare the regression model to the actual Farebox totals.

Below the graph are a series of graphs and charts for analyzing the accuracy of the model run.

The Admin Tool does not currently have an **Export to CSV** function but this is a planned addition. Export to CSV will be added to this section to allow for easy export of graphs and data.



Route ID	AC PM Regression	AC PM Farebox
319	76	undefined
501	394	264
502	453	852
504	675	153
505	1568	1314
507	587	667
508	600	475
509	279	256
551	162	569
552	266	602
553	698	775
554	441	543
559	211	332

Figure 19 Admin Tool: Model Analysis Graphs and Tables

3.5 DATA SOURCES

The Data Sources section serves two functions:

1. Data Upload and Management: This functionality is located in the Admin Tool Center Pane.

2. Information about Data Sources and User's Guide: This functionality is located in the Right Pane.

3.5.1 GTFS

3.5.1.1 About GTFS

The **General Transit Feed Specification (GTFS)** is a data standard for encoding transit schedule and operations information. GTFS is a collection of text files most often distributed in a compressed zip file.

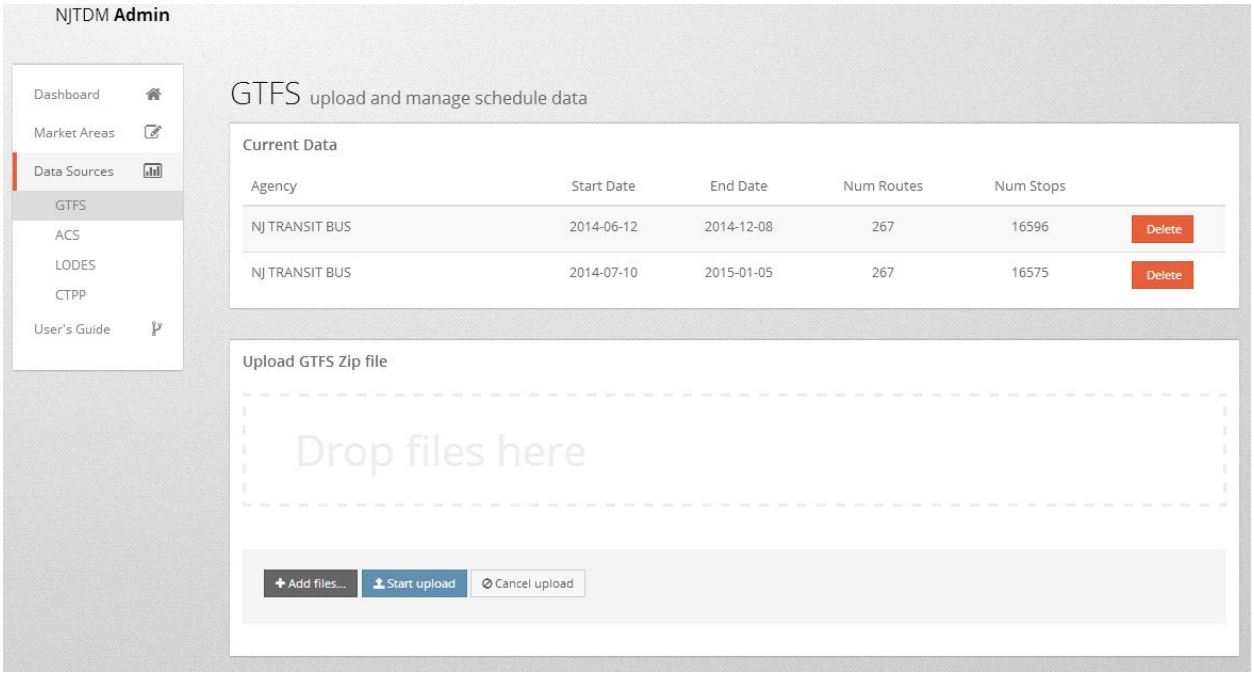
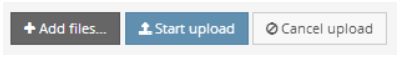


Figure 20 Admin Tool: Data Sources Section GTFS Upload and Manage

3.5.1.2 Loading New Data

To upload new GTFS to the system drag and drop a GTFS zip file to the drop zone on the screen or click **add files** and search for the zip file to upload on your local file system.



The file will then appear underneath the upload section. To add the file to the TDM system simply click **Start upload** to import GTFS. This will start a new import job that can be monitored in the jobs drop down on the top of the screen or on the jobs page. You can now navigate away from this page and wait for the job to finish. Because of the geospatial processing required to import a GTFS file for use in this system it can take up to 3 hours to import a large GTFS file.

3.5.1.3 Agency

Every GTFS file is required to have an agency file that lists the agency that created the GTFS file. This column lists the agency of loaded GTFS Data.

3.5.1.4 Start and End Date

A GTFS file is only valid for a specific amount of time as specified by the Calendar and/or Calendar dates files of the GTFS feed. These listings let you know when the file starts and ends its valid use.

3.5.1.5 Stop and Route Count

Lists the number of stops and routes present in each GTFS file.

3.5.2 ACS

The American Community Survey is the primary demographic data source for Transit Demand Modelling, regression models were built with 2010 ACS 5 Year Data but can be modelled with a number of different ACS Data sources.

3.5.2.1 Add New Data

ACS Data is loaded directly from the census ACS Application Programming Interface (API) into our Transit Demand Modeling database. To add data from neighboring states use the state dropdown menu to locate the desired state, enter Data Source, Base Year and Sum Level then click **Add Dataset.**

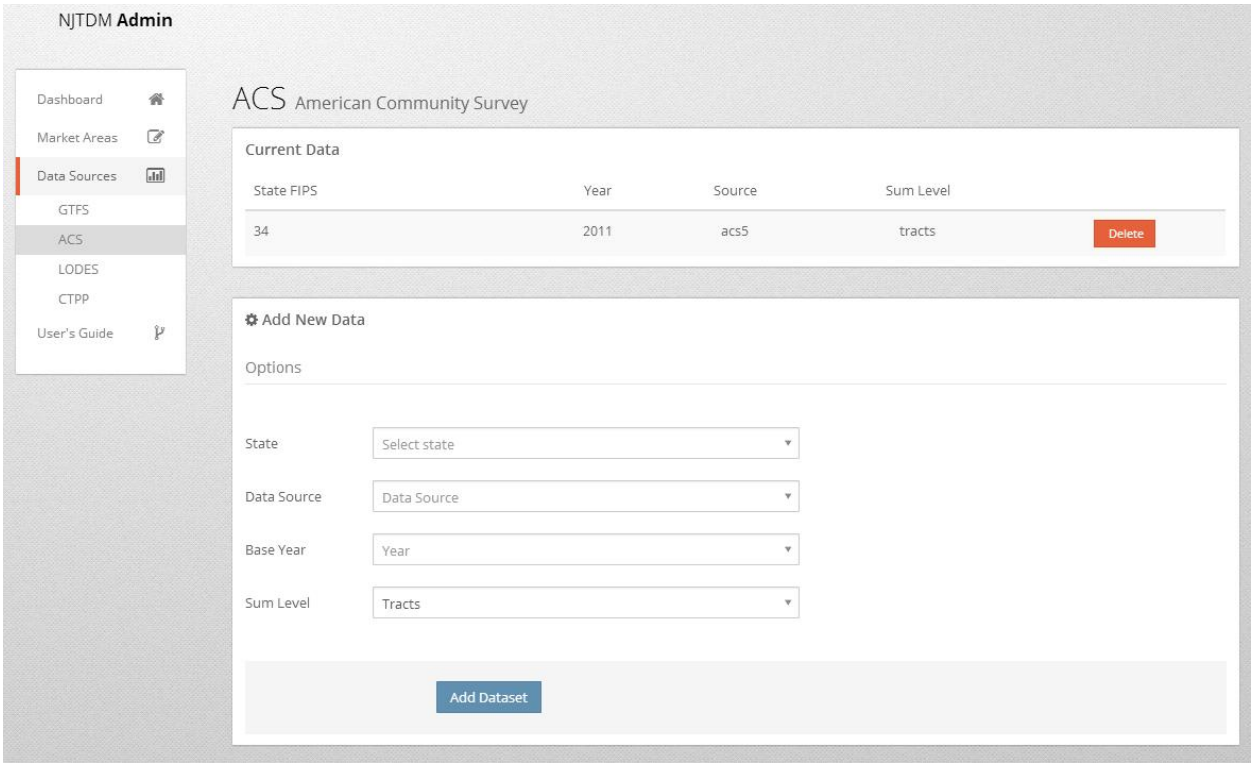


Figure 21 Admin Tool: Data Sources Section ACS Upload and Manage

3.5.2.2 State FIPS

State FIPS is the code used by the census to identify a state. New Jersey State FIPS Code is 34.

3.5.2.3 Sum Level

In this case sum level is the geographic level to which the data is summed. For this tool we can use either census tracts or block groups.

3.5.2.4 Year and Data Source

ACS Data is available in 1, 3 & 5 year summary groups, the data source denotes which of these groups is being used. The year always denotes the last year of the summary group.

3.5.3 LODES

LEHD Origin-Destination Employment Statistics is an annual data set provided by the US Census Bureau. This data set provides mode-less commuting data at the census tract level.

This Section of the Admin Tool is not yet complete.

3.5.4 CTPP

This section of the Admin Tool is not yet complete

3.5.5 Farebox

Farebox and Custom Data uploads will need to standardized.

3.5.6 Custom Data

3.6 Map View

The Map View is accessed through the Left Navigation Panel (Market Areas>City>Map View). The Map View will open in another tab.

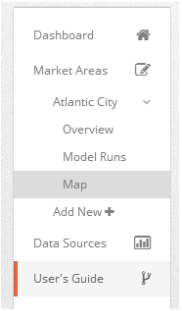


Figure 22 Admin Tool: Left Navigation Panel Access to Map View

There are four Map View Tools to help in building and analyzing your model runs. Three of them are located on the left navigation panel. The fourth is the Model Data Viewer button on the right side of the page. The three tabs in the left navigation panel are as follows:

1. The Modeling Tab – This is the default tab where models are run.
2. The Census Tab – For viewing data about a given Market Area.
3. The GTFS Tab – For viewing routes and adding or removing routes.

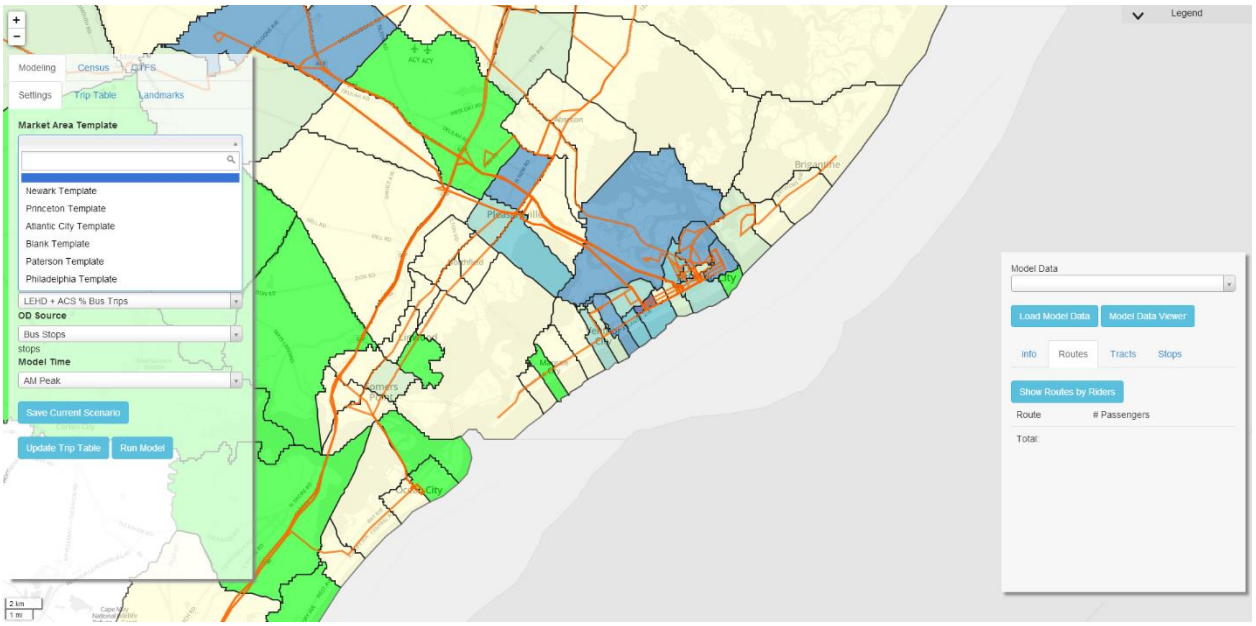


Figure 23 Map View

3.6.1 Modeling View – How to use Map View to Run Models

3.6.1.1 Step 1 - Load a Template

The first step in running a model involves loading your template.

Atlantic City is the default template. To choose another template, click on the dropdown menu titled "**Market Area Template**" located on the left side of the page.

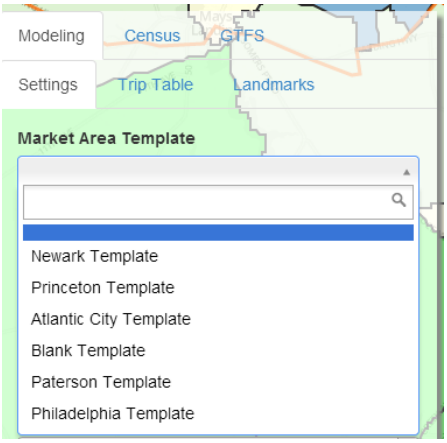


Figure 24 Map View: Load Market Area Template

3.6.1.2 Step 1.2 Editing the Scenario

You can add or remove census tracts from your market area simply by holding the ctrl button and clicking on the census tract you want to add or remove. When a census tract is changed the census data and trip table for the market area are automatically recalculated. When you remove a census tract that tract is dimmed as can be seen on the right side of Figure 25.

Figure 25 Map View: Illustration of Census Tract Removal

3.6.1.3 Step 2 - Choose Your Trip Table Generation Algorithm (Model)

We have designed a number of algorithms to generate trip tables for a market area, they are based on different data sources and methodologies. These different algorithms perform differently in each different market area and respond to different factors. Specific regression models calibrated with ACS 5 Year 2010 data have been designed for Atlantic City, Princeton Trenton and Paterson.

Choose your algorithm by clicking on the dropdown menu titled **Trip Table Generator** located on the left navigation panel.

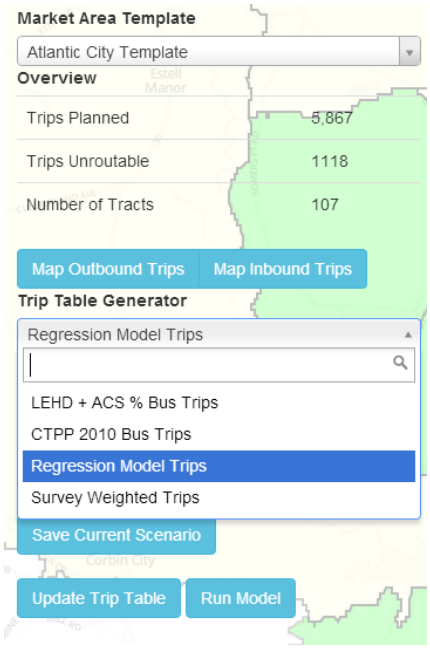


Figure 27 Map View: Trip Table Generation Algorithm Selection

3.6.1.4 Step 3 - Update Trip Tables

Once you have selected the trip table algorithm you want to use, click on the button below titled "**Update Trip Table**." This will update the number of trips planned by the model you've chosen. You'll notice a change in the number of **Trips Planned, Trips Unroutable** and **Number of Tracts**. These numbers are generated by a set of algorithms based on the model you've chosen combined with the census tracts and bus routes chosen and the time of day.

3.6.1.5 Step 4 - Choose Your OD Points Source

Trip table generation algorithms work at the level of a geographic zone, like a census tract. To create a trip table that can be simulated by Open Trip Planner, we need specific points of latitude and longitude inside each census tract to map those individual trips to. **Bus Stops** - Uses bus stops from GTFS as origin destination points. **Parcel Data (experimental)** - Uses centroids of

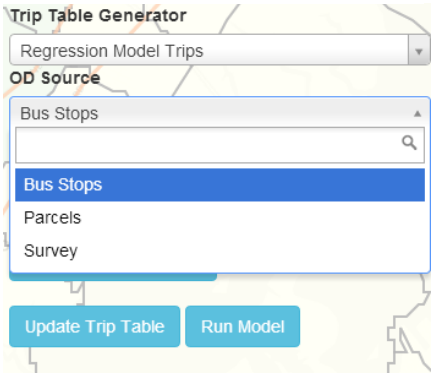


Figure 26 Map View: OD Source Selection

random parcels within 1/2 mile of bus stops. **Survey Data** - Uses survey origins and destinations as origin and destination points.

3.6.1.6 Step 4 - Choose Your Model Time Period

Data in CTPP, LODES and Regression Models use time of travel to work to understand and schedule ridership in the trip table during peak times. Survey data is available for AM only, pm is generated by reversing the direction of AM work traffic.

AM Peak - AM peak includes work trips from 6am to 10am.
PM Peak - PM Peak includes trips from 3pm to 7pm.
Full Day - (Coming soon) All work trips from peaks, plus non work trip estimate.

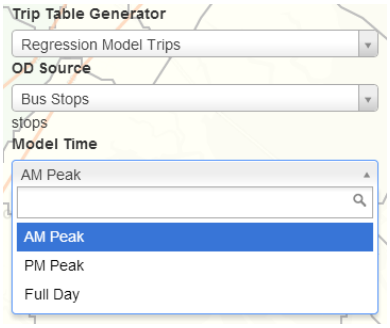


Figure 28 Map View: Model Time Selection

3.6.1.7 Run Model

Once you have chosen all your settings you are ready to run your model. It's worth noting that in the Trip Table tab of the model pane you can hide or show your trip table origins and destinations, you may want to investigate your trip table closely before running your model.

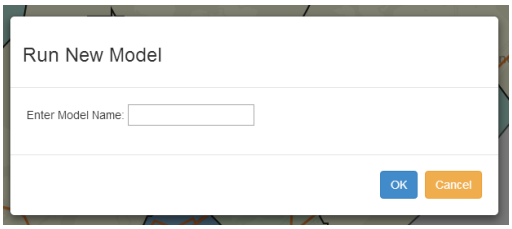


Figure 29 Map View, Model Name Window

The final step in developing a model is to run the model. This step takes all of the Trips Planned and runs those trips through Open Trip Planner to simulate the trips. To initiate this final step, click the button titled "**Run Model**." You'll be asked to name the model. AVAIL has been naming the models based on city name, followed by time of day, followed by algorithm. For instance, AVAIL has chosen titles like AC AM Regression or AC PM Regression.

Once you've named the model click OK. The model should start running. Check the bottom of the left navigation panel to see if the model is running properly. A bar will appear at the bottom of the left navigation pane as illustrated in Figure 30. It will contain a set of numbers, in Figure 30 you see 1/3064. This means that Open Trip Planner has simulated 1 out of a total of 3064 trips. It may take anywhere from 10 minutes to over an hour to complete the model run depending on the number of planned trips and your internet bandwidth. However even if you navigate away from the page the model will continue to run on the server.

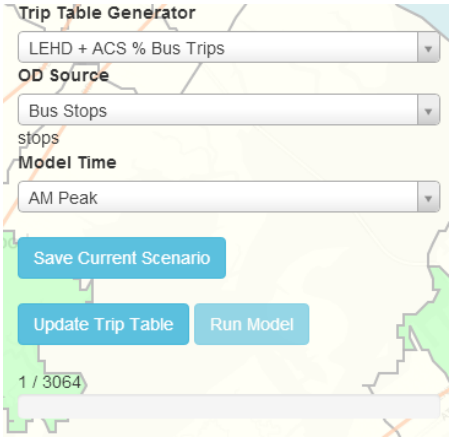


Figure 30 Map View: Run Model Button, Trips Run Bar with ratio of trips run to total trips planned

3.6.2 Model Analysis

The Model Analysis button on the Left Navigation Pane of the Map Viewer will take you back to the Admin Tool>Market Area>Model Analysis (see section 3.4.3 of this document).

3.6.3 GTFS View

The GTFS tab allows you to turn Routes and Stops on and off. This can help in analyzing model runs.

To filter down to a single bus stop, start clicking on the Show/Hide Stops button. Routes are shown by default, Stops are hidden. Next, mouse over the map to find the "Stop Code" for the bus stop you want. In the example at right we chose Arctic Ave at Christopher Columbus. We then entered the Stop Code for this bus stop, 10054, into the filter bar. You'll notice on the second map that Arctic Ave at Christopher Columbus is now the only bus stop showing.

This very same process can be used to turn Routes on or off.

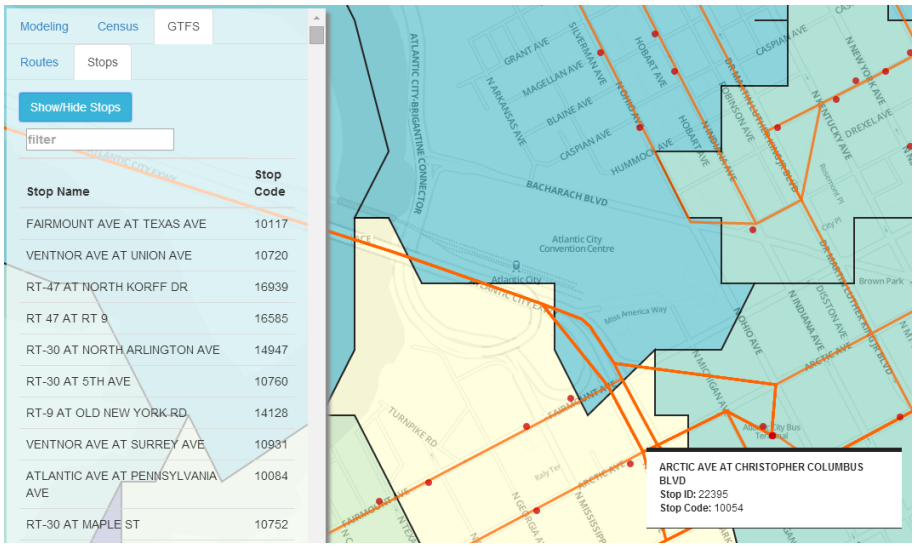


Figure 31 Map View: with Stops and Routes Shown

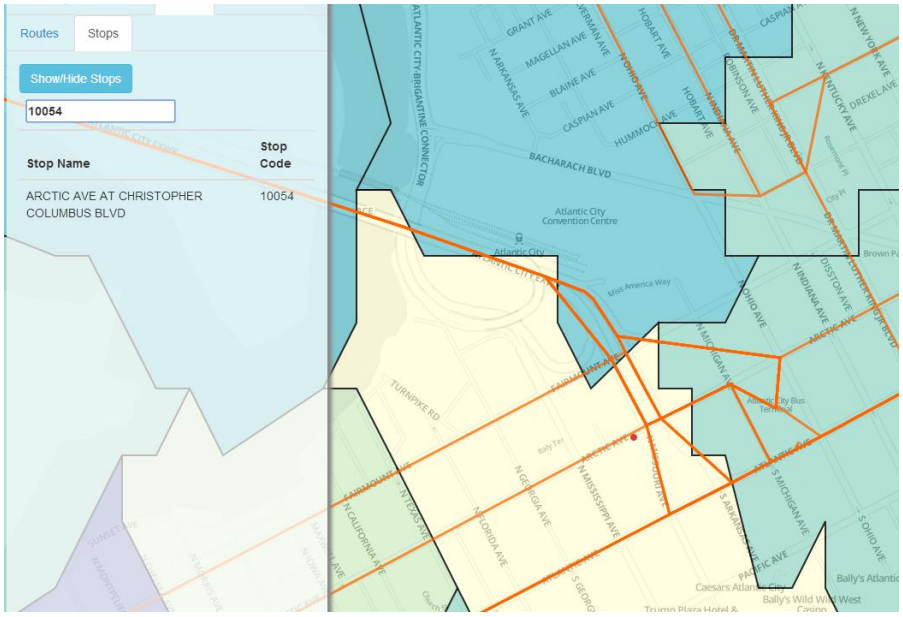
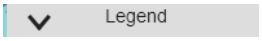


Figure 32 Map View: Filtered to show only one stop

3.6.4 Census View

The Census Tab allows you to view census data in choropleth shading on the map. To view census data, click on the box below the population number that looks like this: 

To see what the shadings mean, click on the "Legend" tab at the top right: 

For Figure 33 we clicked on Income Categories followed by the category titled "$25,000 to $29,999." For Figure 34 we clicked on "$30,000 to $34,999." Notice how the map coloring changed to reflect population distribution.

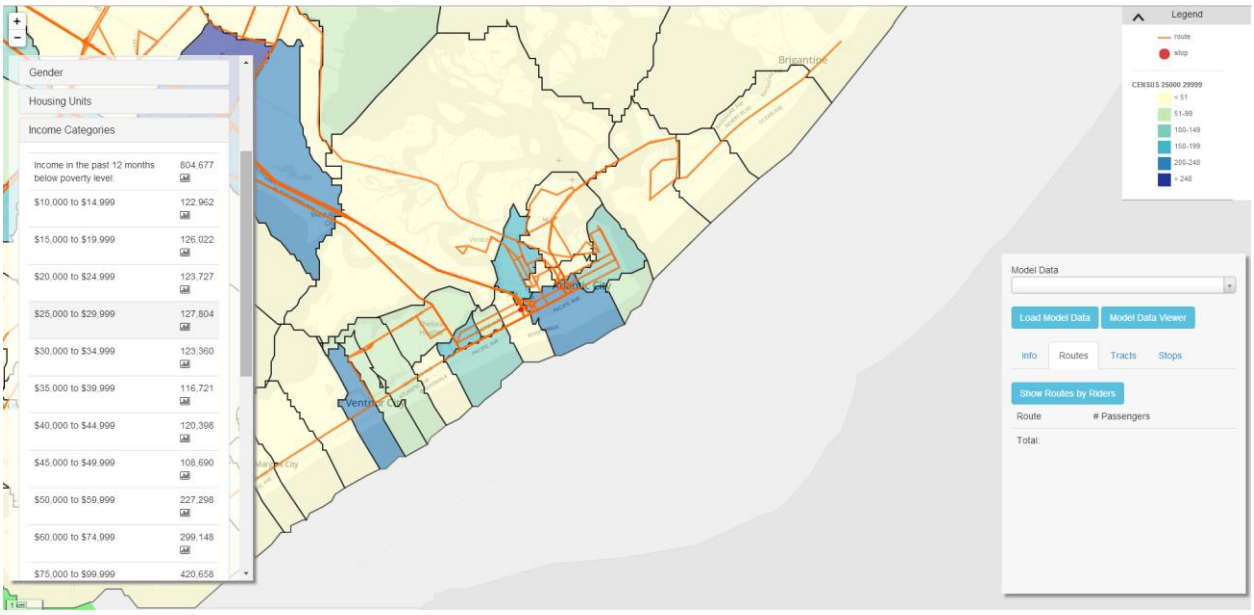


Figure 33 Map View: With census data shown in chloropleth, income $25,000 to $29,999

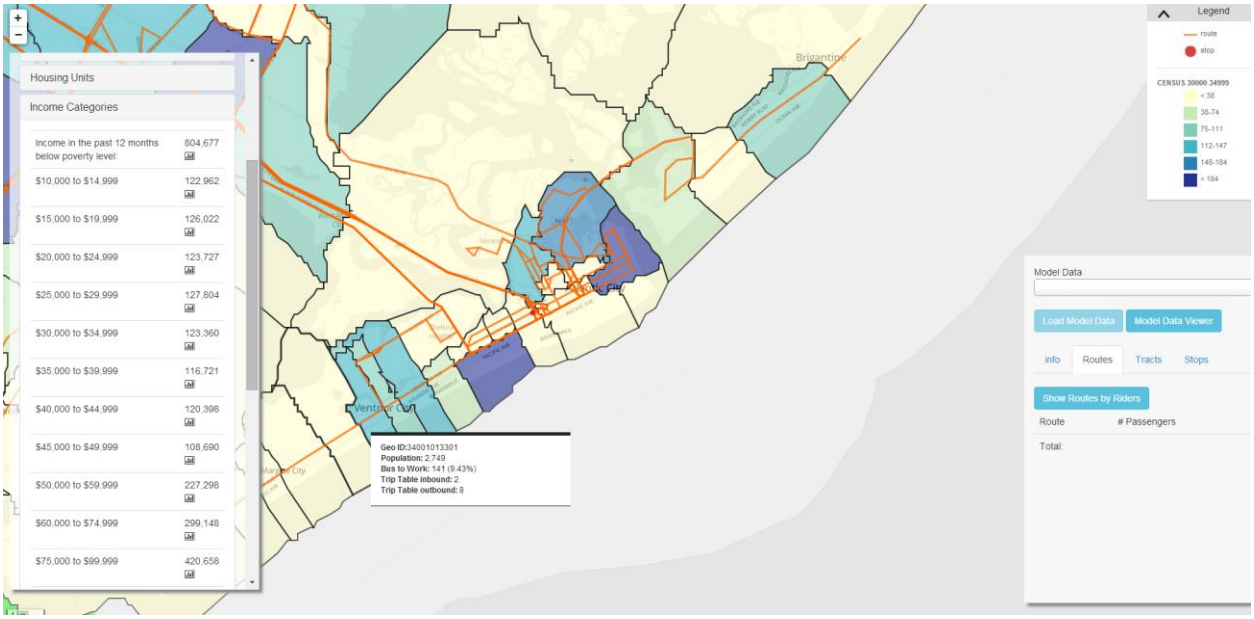


Figure 34 Map View: With census data shown in chloropleth, income $30,000 to $34,999

4 MODELING

4.1 INTRODUCTION

This research uses data science techniques to produce new forms of transit forecasting tools and algorithmic modeling. Forecasting is performed using a number of different underlying data sources, including the American Community Survey five-year summary dataset. The ACS is extracted at the tract level and used in a linear regression model and further integrated within a web environment microsimulation using CTPP and Open Trip Planner to produce route level forecasts.

This report contains a summary of data elements and the regression model specifications used in the microsimulation process, including the source of the data, descriptive statistics, the correlations coefficients, regression model development, and microsimulation model outputs.

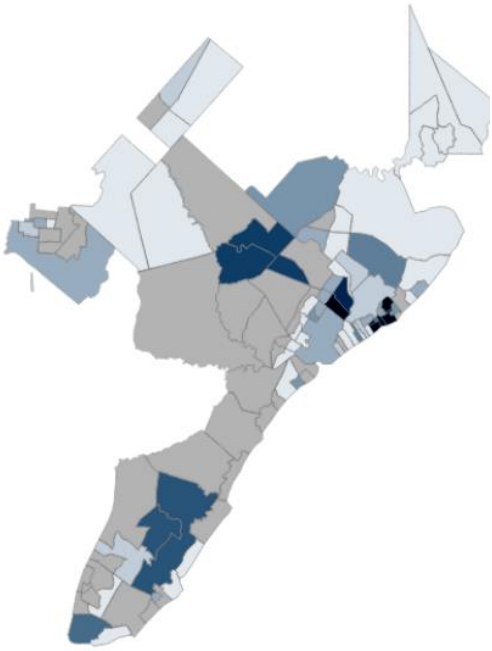


Figure 35 Map of Atlantic City by Census Tract

4.2 META DATA

The data used in this study was obtained from the US Census Application Programming Interface (API)[1]. The data set is called the American Community Survey Five-Year Data 2006-2010 (ACS). The ACS is an ongoing survey that provides data every year. The ACS provides communities with the current information they need to plan investments and services. It covers a broad range of topics about social, economic, demographic, and housing characteristics of the U.S. population. For this research, we developed two new variables: Employment Density (EMP_DEN) and Population Density (POP_DEN). These variables were derived by dividing the employment at tract level by polygon tract area and population at tract level by polygon tract area.

[1]United States Census Bureau, American Community Survey, http://www.census.gov/data/developers/data-sets/acs-survey-5-year-data.html

4.3 DESCRIPTIVE STATISTICS

The first task in developing the regression models was to review the descriptive statistics for the most relevant variables. Table 1 includes the variable name, variable description, mean, median, and standard deviation of each of the variables. The mean and median are measures of central tendency, where the mean is the numerical value found by summing the values and dividing by the number of cases, while the median is the numerical value separating the higher half of a data sample from the lower half. The standard deviation measures the amount of variation or dispersion from the average and is equal to the square root of the sample variance. These statistics provide a better understanding of the characteristics of the variables at the tract level.

Table 2 Atlantic City Descriptive Census Statistics

Variable	Description	Mean	Median	Std. Deviation
unemployme	Unemployed Population	207.99	187.00	142.067
public_tra	Journey to Work by Public Transportation Total	96.95	42.50	135.545
bus_to_wor	Journey to Work by Public Transportation by Bus or Trolley Bus	90.90	37.00	133.890
agricultur	Employment in Agriculture Sector	14.55	0.00	30.605
constructi	Employment in Constuction Sector	110.88	93.50	84.463
manufactur	Employment in Manufacturing Sector	84.58	45.50	128.333
informatio	Employment in Information	23.66	17.00	28.786
finance	Employment fi Finance	96.17	82.50	71.379
arts	Employment in Arts	409.96	320.00	358.514
poverty_st	Poverty Status	449.63	340.00	411.014
no_high_sc	No High School Education	255.67	168.50	247.300
high_schoo	High School Education	685.81	584.00	452.165
foreign_bo	Foreign Born	488.94	307.00	520.705
spanish_sp	Spanish Speaking	217.90	92.00	323.253
other_lang	Other Language Speaking	135.14	68.00	193.328
male_pop	Male Population	1879.50	1619.50	1102.515
age22_24	Age 22 to 24 Total	137.34	106.00	118.432
age25_29	Age 25 to 29 Total	216.15	194.00	160.436
age30_34	Age 30 to 34 Total	212.42	159.50	176.693
age35_39	Age 35 to 39 Total	229.38	189.00	192.564
race_white	Race White	2815.49	2449.00	1758.957
race_black	Race Black	502.25	197.50	711.065
race_asian	Race Asian	206.15	65.50	351.308
race_other	Race Other	242.53	95.00	355.714
race_two	Bi-racial	99.40	52.00	143.735
total_hous	Households, Occupancy Status, Total	2261.16	2093.00	1338.495
occupancy_	Tenure, Occupancy Status, Owner Occupied	1080.65	899.00	713.725
occupanc_1	Tenure, Occupancy Status, Renter Occupied	435.08	360.00	316.667
car_0	Households, Zero Vehicles Available	112.19	64.00	130.792
car_1	Households, One Car Available	393.51	345.50	246.567
car_2	Households, Two Cars Available	755.91	586.50	585.424
car_3	Households, Three Cars Available	319.32	236.50	327.875
car_4	Households, Four Cars Available	108.89	84.50	116.780
emp_den	Employment/Area	16915463544.87	9650409240.25	26504965533.74
pop_den	Population/Area	38184941855.88	20868714033.20	62285032124.07

4.4 CORRELATIONS

To better understand how the variables covary with our dependent variable, Bus_to_Wor, we used SPSS[2] to a produce a correlation matrix. Table 2 displays the statistically significant correlation coefficients, indicating the measure of strength of the linear association between two variables (-1 to +1).

Table 3 Atlantic City Bus to Work Correlation Coefficients

Variable	Description	Correlation Coefficient
Bus_to_Wor	Journey to Work by Public Transportation by Bus or Trolley Bus	**1**
unemployme	Unemployed Population	.283**
public_tra	Journey to Work by Public Transportation Total	.997**
constructi	Employment in Construction Sector	-.220*
informatio	Employment in Information Sector	-.225*
arts	Employment in Arts	.503**
poverty_st	Poverty Status	.444**
no_high_sc	No High School Education	.500**
foreign_bo	Foreign Born	.566**
spanish_sp	Spanish Speaking	.401**
other_lang	Other Language Speaking	.424**
age25_29	Age 25 to 29 Total	.332**
age30_34	Age 30 to 34 Total	.227*
race_white	Race White	-.283**
race_black	Race Black	.500**
race_asian	Race Asian	.344**
race_other	Race Other	.387**
race_two	Bi-racial	.215*
occupancy_	Tenure, Occupancy Status, Owner Occupied	-.202*
occupanc_1	Tenure, Occupancy Status, Renter Occupied	.526**
car_0	Households, Zero Vehicles Available	.753**
car_1	Households, One Car Available	.471**
car_3	Households, Three Cars Available	-.226*
car_4	Households, Four Cars Available	-.223*
emp_den	Employment/Area	.532**
pop_den	Population/Area	.514**

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

[2] http://www-01.ibm.com/software/analytics/spss/

4.5 Regression Methodology

Statistically significant correlations provides guidance for the specification of a linear regression model. Those variables that are too highly correlated are most likely illustrating multicollinearity problems and are not useful. A linear regression model assumes a linear relationship between our dependent variable (bus_to_wor), and a set of independent variables. A well-specified regression model includes all the necessary relevant variables and excludes any irrelevant variables, but remaining as sparse as possible to make them "practice-ready" for applications.

The regression model will produce a number of parameters and model fitting indicators, such as the coefficient of determination (R Squared). The R Squared is defined as the percent of the variation of the dependent variable (bus_to_wor) explained by the set of independent variables included in the regression model. Therefore, the percent of variation in the number of bus riders from each census tract will be explained by the set of independent variables. Therefore, the higher the R Squared, the more explanatory power the model provides.

The regression model output also provides a constant (intercept), which is the average value of the dependent variable when the independent variables equal zero.[3] In addition, the regression model output will provide slope coefficients for each independent variable, indicating the average change in the dependent variable with a one unit change in the independent variable, holding all else contact.

For the purposes of this modelling effort, statistical significance is defined as a p-value of <.05, or a t-value >2.5.

4.6 Overview on Building Regression Models

The following is an overview of processes followed in building the regression models derived from the census data. At this time, the regression models show high sensitivity to changes in geographies.

4.7 Extract Census Tracts from Web Tool

Our first step is to extract the GeoID's of the Census Tracts for a given market area from the database of our web-tool. For Atlantic City, the GeoID's are shown below:

"34001010102","34001010104","34001010105","34001001400","34001002500","34001002400","34001001900","34001001200","34001001500","34001001100","34001000300","34001002300","34001000500","34001000400","34001000100","34001000200","34001013201","34001013202","34001013302","34001013301","34001013000","34001013101","34001013102","34001001300","34001010101","34001010505","34001010200","34001011900","34001010300","34001012000","34001012100","34001012200","34001012401","34001012302","34001012402","34001012501","34001013500","34001012602","34001012701","34001012801","34001011803","34001012502","34001011702","34001011701","34001010506","34001010503","34001983400","34001010403","34001011802","34001011804","34001011404","34001012702","34001011805","34001011403","34009020201","34009020101","34009020102","34009020203","34001012802","34001011600","34001011500","34001011401","34011030100","34009020301","34009020205","34009020800","34009020206","34001010401","34001010501","34009021001","34009020902","34009020901","34009021100","34009022102","34009022101","34009021804","34009021803","34009021701","34009021702","34009021805","34009021806","34009021900","34009022000","34009021600","34009021500","34009021400","34009021300","34009021002","34009020302","34009020700","34011041000","34011040800","34011040700","34011040600","34011040500","34011041100","34011040300","34011040400","34001011202","34001011100","34001011000","34001010900","34001010800","34029736101","34029736102","34029737000","34029736002","34029736001"

[3] Lewis-Beck, Michael S., 1980, *Applied Regression, An Introduction,* Newbury Park: Sage Publications

4.8 Creating a Data Rich GIS File

Now that we have a subset of Census Tracts, we slice this group out of our database that contains New Jersey Census Tracts, and query the Census API to acquire a selection of ACS data for these tracts. Once completed, this data is exported from our database as a JavaScript Object Notation (JSON) object that contains the census geographies and with the ACS data as attributes, this JSON object is then converted to a shapefile for use in GIS applications.

geoid	total_popu	employment	unemployme	travel_to_	car_to_wor	public_tra	bus_to_wor
34009020901	828	318	27	308	244	0	0
34009021002	3623	1930	141	1897	1735	16	16
34009022102	5533	2367	356	2299	2140	37	37
34009021001	2600	848	89	835	738	20	20
34009022101	1887	649	57	649	545	0	0
34009020301	4202	2136	185	2070	1969	10	0
34009021500	2150	1116	206	1084	944	43	24
34009020102	2413	1220	47	1210	867	19	9
34009021400	3650	1687	202	1657	791	104	101
34009020101	3307	1703	172	1703	1190	48	38
34009021804	5603	2789	343	2633	2448	79	79
34009021300	4111	1670	196	1580	1411	25	12

4.9 GeoDA

Next, we take advantage of the Open Source program GeoDA[4], a statistical tool for analyzing sets of spatial data. Using GeoDA, we developed a regression model for Atlantic City that provides significant ridership determination values using ACS variables.

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  atlantic_city
Dependent Variable  :  bus_to_wor  Number of Observations:  108
Mean dependent var  :     90.8981  Number of Variables   :    5
S.D. dependent var  :     133.269  Degrees of Freedom    :  103

R-squared           :    0.635076  F-statistic           :     44.8126
Adjusted R-squared  :    0.620904  Prob(F-statistic)     : 9.5755e-022
Sum squared residual:      699980  Log likelihood        :    -627.186
Sigma-square        :     6795.93  Akaike info criterion :     1264.37
S.E. of regression  :     82.4374  Schwarz criterion     :     1277.78
Sigma-square ML     :      6481.3
S.E of regression ML:     80.5065

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------
    CONSTANT     -19.35818       15.86841      -1.219919      0.22528
  informatio    -0.7130557      0.3039812      -2.345723      0.02091
       car_0     0.5432931     0.08102835       6.704975      0.00000
       car_1     0.1331132     0.03839082       3.467318      0.00077
     emp_den  8.155064e-010   3.607904e-010       2.260333      0.02590
------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER    5.011276
TEST ON NORMALITY OF ERRORS
TEST                    DF           VALUE             PROB
Jarque-Bera              2            27.4101          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE             PROB
Breusch-Pagan test     4            87.5447          0.00000
Koenker-Bassett test   4            39.3484          0.00000
========================= END OF REPORT =============================
```

Figure 36 GeoDA Regression Model Output for Atlantic City, 08.04.2014

[4] http://geodacenter.asu.edu/software/downloads

4.10 Regression Models for Atlantic City, Princeton/Trenton, and Paterson

4.10.1 Atlantic City

4.10.1.1 Atlantic City Model Development

We developed a series of regression models for Atlantic City (see Table 3), with a final specification that produced an R-Squared of .629. The independent variables tested in these models include CAR_0, CAR_1, INFORMATIO and EMP_DEN. The constant in all the models was not statistically significant.

Table 4 Atlantic City Regression Model Development

Atlantic City				
	Atlantic City1	Atlantic City2	Atlantic City3	**Atlantic City4**
Dependent Variable	bus_to_wor	bus_to_wor	bus_to_wor	**bus_to_wor**
Constant	4.47	-25.17	-13.85	**-19.36**
CAR_0	.77**	0.69**	.63**	**.54****
CAR_1		0.10**	0.13**	**.13****
INFORMATIO			-0.79**	**-.71***
EMP_DEN				**.00***
R Sq.	0.57	0.59	0.62	**0.63**
N	108	108	108	**108**

**. T-value >2.5 and P-value <.05.

*. T-value >2.5 or P-value <.05.

4.10.1.2 Atlantic City Regression Equation

The final regression model equation used in the microsimulation process, with an R Squared of 63% is specified as follows:

BUS_TO_WOR = 0.543293 *(CAR_0) + 0.1331132*(CAR_1) - 0.713056*(INFORMATIO) + 8.155063e-010*(EMP_DEN)

4.10.1.3 Implementing Regression Models For Microsimulation

To use the model equation to predict ridership in our microsimulation, we use the following formula for each tract in the market area.

a. Find the number of riders predicted by the regression model for the tract.

b. We take the number of riders predicted by the regression, and divide that by ACS variable Journey to Work: Public Transportation – Bus for the tract, giving us the ratio of predicted riders to census counted riders.

c. Then for each home tract to work tract bus travel count in the CTPP, we multiply the count by the regression ratio.

d. All resulting trips are added to our trip table to be simulated by the modeling software.

e. The trip is microsimulated using Open Trip Planner.

4.10.2 Princeton/Trenton Regression Model, 8.5.14

Table 5 Princeton/Trenton Regression Model Development

Princeton Regression Models					
	Princeton1	**Princeton2**	**Princeton3**	**Princeton4**	**Princeton5**
Dependent Variable	bus_to_wor	bus_to_wor	bus_to_wor	bus_to_wor	bus_to_wor
Independent Variable					
Constant	25.95**	16.69*	12.19	0.42	1.61
car_0	0.35**	.32**	0.33	0.3**	0.34**
poverty_st					-0.04
race_black		.01*		0.014**	0.02**
age_25-29			0.06	0.62*	0.07**
R Sq.	0.63	0.66	0.65	0.68	0.70
N	69	69	69	69	69

Table 6 GeoDA Regression Model Output for Princeton, 08.05.2014

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  princeton
Dependent Variable  :  bus_to_wor  Number of Observations:   69
Mean dependent var  :     70.4638  Number of Variables   :    5
S.D. dependent var  :     75.2699  Degrees of Freedom    :   64

R-squared           :    0.697629  F-statistic           :     36.9152
Adjusted R-squared  :    0.678731  Prob(F-statistic)     :5.55997e-016
Sum squared residual:      118204  Log likelihood        :    -354.796
Sigma-square        :     1846.93  Akaike info criterion :     719.592
S.E. of regression  :      42.976  Schwarz criterion     :     730.762
Sigma-square ML     :      1713.1
S.E of regression ML:     41.3896

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------
    CONSTANT      1.612588       10.39041      0.1551996     0.87715
       car_0     0.3400127     0.04336295       7.841088     0.00000
  race_black    0.02379176    0.008090476       2.940712     0.00455
    age25_29    0.07151607     0.02770711       2.581145     0.01215
  poverty_st    -0.0409672     0.02478712      -1.652762     0.10328
------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   7.302327
TEST ON NORMALITY OF ERRORS
TEST                    DF           VALUE             PROB
Jarque-Bera              2             2.2319          0.32760

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                    DF           VALUE             PROB
Breusch-Pagan test       4             8.8273          0.06557
Koenker-Bassett test     4             6.7490          0.14976
========================== END OF REPORT ==============================
```

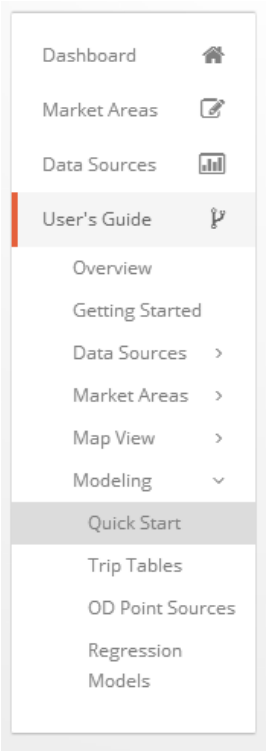
4.10.3 Paterson Regression Model, 8.5.14

Paterson Regression Models					
	Paterson1	Paterson2	Paterson3	Paterson4	Paterson5
Dependent Variable	bus_to_wor	bus_to_wor	bus_to_wor	bus_to_wor	bus_to_wor
Constant	133.44**	50.20*	32.89	59.55**	44.07*
poverty_st				-.05*	.06**
50+_Units				.15**	.13**
car_0	.27**	.23**	.13**	.31**	.023**
car_1		.16**	.11**	.14**	.09**
Employment Density			.00**		.00**
population					
R Sq.	0.30	.41	.47	51	.56
N	128	128	128	128	128

Table 7 GeoDA Regression Model Output for Paterson, 08.05.2014

```
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  paterson
Dependent Variable  :  bus_to_wor  Number of Observations:  128
Mean dependent var  :     192.938  Number of Variables   :    6
S.D. dependent var  :      136.32  Degrees of Freedom    :  122

R-squared           :    0.561941  F-statistic           :     31.3003
Adjusted R-squared  :    0.543988  Prob(F-statistic)     :2.15646e-020
Sum squared residual:1.04198e+006  Log likelihood        :    -757.919
Sigma-square        :     8540.82  Akaike info criterion :     1527.84
S.E. of regression  :     92.4165  Schwarz criterion     :     1544.95
Sigma-square ML     :     8140.47
S.E of regression ML:     90.2245

------------------------------------------------------------------------
    Variable   Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------
    CONSTANT       44.0737       18.68106       2.359271     0.01990
       car_0     0.2265559     0.04817075       4.703184     0.00001
       car_1    0.09415446     0.03179322       2.961463     0.00368
   50+_units      0.131551     0.03525934       3.730956     0.00029
  poverty_st    -0.05523987     0.01947491      -2.836464     0.00534
     emp_den  7.776513e-010   2.019496e-010       3.850719     0.00019
------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   6.706108
TEST ON NORMALITY OF ERRORS
TEST                 DF          VALUE            PROB
Jarque-Bera           2            2.3264         0.31249

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                 DF          VALUE            PROB
Breusch-Pagan test    5            7.7766         0.16898
Koenker-Bassett test  5           11.4090         0.04385
========================== END OF REPORT ==============================
```

4.10.4 Left Navigation - User's Guide



By Clicking on the User's Guide a user can drill into a series of explanatory sections. The outline for the User's Guide is as follows:

1. Navigating the Admin Tool and Map View (this PDF)
2. Admin Tool Modeling
   - 2.1. Running a model
   - 2.2. Viewing Model Outputs
   - 2.3. Trip Tables
   - 2.4. OD Point Sources
   - 2.5. Regression Modeling (the Final Regression Model Report)
   - 2.6. How the Modeling Tool Works
3. Map View Tools
   - 3.1. How to use Map View to Run Models
     - 3.1.1. Load a scenario
     - 3.1.2. Update Trip Tables
     - 3.1.3. Turning Census Tracts on and off
     - 3.1.4. Turning Routes on and off
4. Market Areas (Can we put the Right Nav here?)
   - 4.1. Add New Market Area (Link to Right Nav Guide for Users Guide)
   - 4.2. Edit Existing Market Areas (link to Right Nav Guide for User's Guide)
5. Data Sources
   - 5.1. GTFS (Put Right Nav User's Guide here)
   - 5.2. ACS(Put Right Nav User's Guide here)
   - 5.3. LODES (Put Right Nav User's Guide here)
   - 5.4. CTPP (Put Right Nav User's Guide here)
   - 5.5. Farebox (Put Right Nav User's Guide here)
   - 5.6. Custom Data

Running a Model Introduction

Running a model is done through the map explorer view, which allows your to dynamically edit data sources and visualize the impact on your model run. It also allows you to closely monitor the trip origins and destinations of your trip table down to the geographical unit of your market area. Once you have your market area scenario selected you start your model run which runs a microsimulation of your trip table through Open Trip Planner to understand ridership from the whole market area level down to individual stops. Step 1 - Choose Your Market Area Template

Market Area templates have the basic information needed to run a model for a particular market area, which are the routes included and the census tracts of the market area. Simply selecting a new scenario from the selector will load that scenario onto the map. This will zoom you to that scenario and show all the routes from the map and load a trip table. The census tracts will be visualized based on out-bound trips from that census tracts.

Editing the Scenario

You can add or remove census tracts from your market area simple by holding the ctrl button and clicking on the census tract you want to add or remove. When a census tract is changed the census data and trip table for the market area are automatically recalculated.

Changing Data Sources

The ability to switch data sources like ACS and LODES vintages will be added shortly.

Step 2 - Choose Your Trip Table Generator

We have designed a number of algorithms to generate trip tables for a market area, they are based on different data sources and methodologies. These different algorithms perform differently in each different market area and respond to different factors. Specific regression models calibrated with ACS 5 Year 2010 data have been designed for Atlantic City, Princeton Trenton and Paterson.

LODES + ACS % - This model looks at the number of work trips from each tract in the lodes data and multiplies that data by the rate of bus ridership to work in that tract to get the number of trips generated for each destination tract.

CTPP Model - This model uses the public transporation to work value from CTPP 2010 to generate a trip table.

Regression Model - Regression models take a number of ACS variables to estimate a number of transit riders from a specific tract. This number is then used with the total bus riders from the ACS to create a regression ratio

Survey Based Model - Uses survey origin and destination and weights to generate trip table. Step 2 A - Update Trip Table

Once you have selected the trip table algorithm you want to use you click the update

Step 3 - Choose Your OD Points Source

Trip table generation algorithms work at the level of a geographic zone, like a census tract, to create a trip table that can be simulated by Open Trip Planner we need specific points of latitude and longitude inside each census tract to map those individual trips to.

Bus Stops - Uses bus stops from GTFS as origin destination points.

Parcel Data (experimental) - Uses centroids of random parcels within 1/2 mile of bus stops

Survey Data - Uses survey origins and destinations as origin and destination points.

Step 4 - Choose Your Model Time Period

Data in CTPP, LODES and Regression Models use time of travel to work to understand and schedule ridership in the trip table during peak times. Survey data is available for AM only pm is generated by reversing the direction of AM work traffic.

AM Peak - AM peak includes work trips from 6am to 10am.

PM Peak - PM Peak includes trips from 3pm to 7pm.

Full Day - (Coming soon) All work trips from peaks, plus non work trip estimate.

Step 5 - Run the Model

Once you have chosen all your settings you are ready to run your model. Its worth noting that in the Trip Table tab of the model pane you can hide or show your trip table origins and destinations, you may want to investigate your trip table closely before running your model. Once you are satisfied with your trip table click run model. This will ask you to name your model, enter a name and click ok. You can track your model run progress on the bottom of the model run tab. However even if you navigate away from the page the model will continue to run on the server.