**An Attempt at Multimodal Public Transit Accessibility Modeling for Westchester County, New York**

Deana Baron

GTECH733

5/22/19

**Introduction**

As urbanization increases and much of the world’s population begins to move, work, and live in urban areas, there needs to become a better way to move people from one destination to another. The purpose of transportation is to effectively move through space and time. Transit accessibility is one of the major overarching goals of any public transit authority, however it is a concept that is relatively difficult to identify.

Transportation accessibility is becoming a particular topic of interest in the transportation geography community as urban areas shift toward public transportation as the main method of transportation. To begin with defining transportation accessibility and how it manifests and can be measured in an urban area, we start with a dissertation by Hunter College graduate Maxwell Siegel (2016), who attempts to quantify transit accessibility in New York City and develop an index to identify spatial and social patterns in public transportation. Siegel discusses how there is only a finite number of resources that planners and municipalities can put into transportation, and there are a multitude of factors that go into establishing the best public transportation network for an area. Accessibility and mobility are two concepts that are commonly brought up for the overall goals of a government agency’s transportation system. There are a myriad of definitions for both. Mobility is generally identified as an efficiency measure of movement to, from, and within a system. Accessibility is identified as more flexible depending on its application, but generally is the idea of availability of goods and services. These concepts are different, however cannot be separated in such a study (Siegel 2016).

Establishing a measure for public transportation accessibility is important and there are multiple ways to do so. Many of the models set out to measure accessibility use a different measure of travel time, like Tribby and Zandbergen (2012), who set out to design a high resolution spatio-temporal model for public transit accessibility. While accessibility can be thought of as a social measure, implementing this into a network requires some other factors. Tribby and Zandbergen use travel time from origins to destinations using walking times, waiting times, and bus travel times to measure transportation accessibility in this network (Tribby and Zandbergen 2012).

Many transit accessibility models are created and conducted using smaller scale areas, usually large and robust cities. It makes sense, since public transit data from these cities tends to be more readily available than in other areas. However, what about on a larger scale? There are not many accessibility models run on a county scale, in a network composed of a lot of routes and stops. The first question is how does a transit accessibility model work on a county scale?

The study area chosen for this is Westchester County, New York, a county in lower New York State, located just above the Bronx.  According to the 2010 Census, the county had a population of 949,113 and was estimated to have increased to 967,612 in 2017 (U.S. Census Bureau 2018). Westchester is centrally located in the Hudson Valley and within the New York City metropolitan area and due to the upper middle class development of some of its communities in the late 19th century along with rapid population growth, it was the first suburban area of its scale to develop. Much of the county, and the southern portion in particular, are as densely developed as New York City due to its many short connections with mass transit (Panetta & Jackson 2006).

The goal of this project is to create a multimodal network using walking routes and bus routes in Westchester and then to measure transit accessibility along the network using multiple measures. It would be interesting to see how and if the network benefits the county equally.

**Data and Methods**

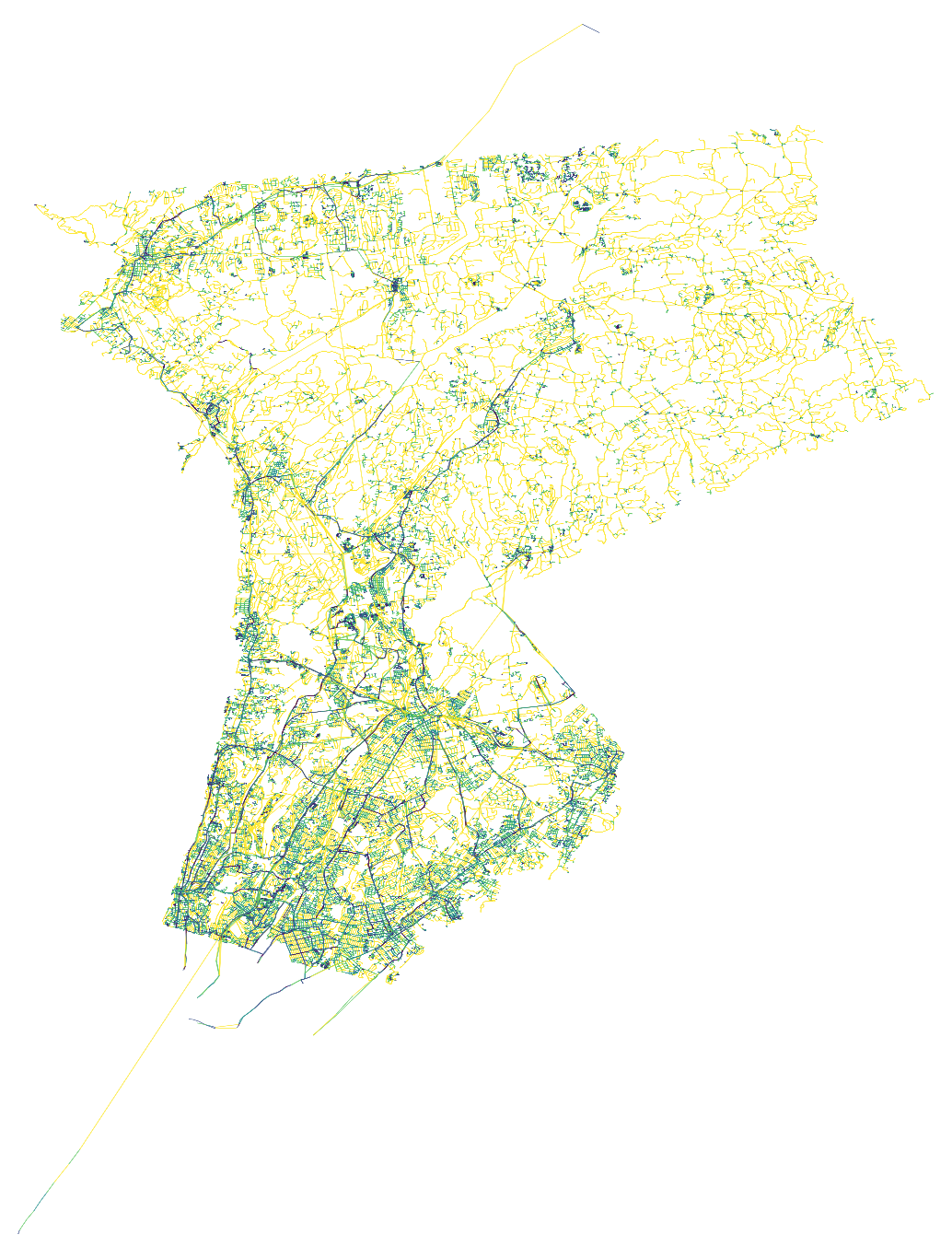
This project was conducted entirely with Python 3.6 and libraries available online. All of the data is also open source and publicly available. The Bee-Line Bus system is Westchester County’s sole bus operator. The Beeline Bus follows the GTFS (General Transit Feed Specification) standard in which multiple public transit agencies share their data in a way that can be used by a multitude of different software. The data is composed of different comma delimited text files representing various parts of the transit network, like trips, stops, stop times, routes, and differences per date. The latest Beeline Bus data was dated March 28, 2019 and is in effect from September 4, 2018 to present. The transit network contains 2,864 stops spread across 114 routes.

The pedestrian network and points of interest used in this project was downloaded from OpenStreetMap using OSMnx, a Python library designed to use street network data queried directly from OpenStreetMap (Boeing 2017). This included a pedestrian network and points of interest based on different amenities tagged in OpenStreetMap from Westchester County. The last data used in this project were 2010 U.S. Census blocks for Westchester County, and Longitudinal Employer-Household Dynamics (LEHD) LEHD Origin-Destination Employment Statistics (LODES) information, which counts the total number of jobs per census block for every block in the country. Both were obtained from the U.S. Census Bureau (U.S. Census Bureau 2019).

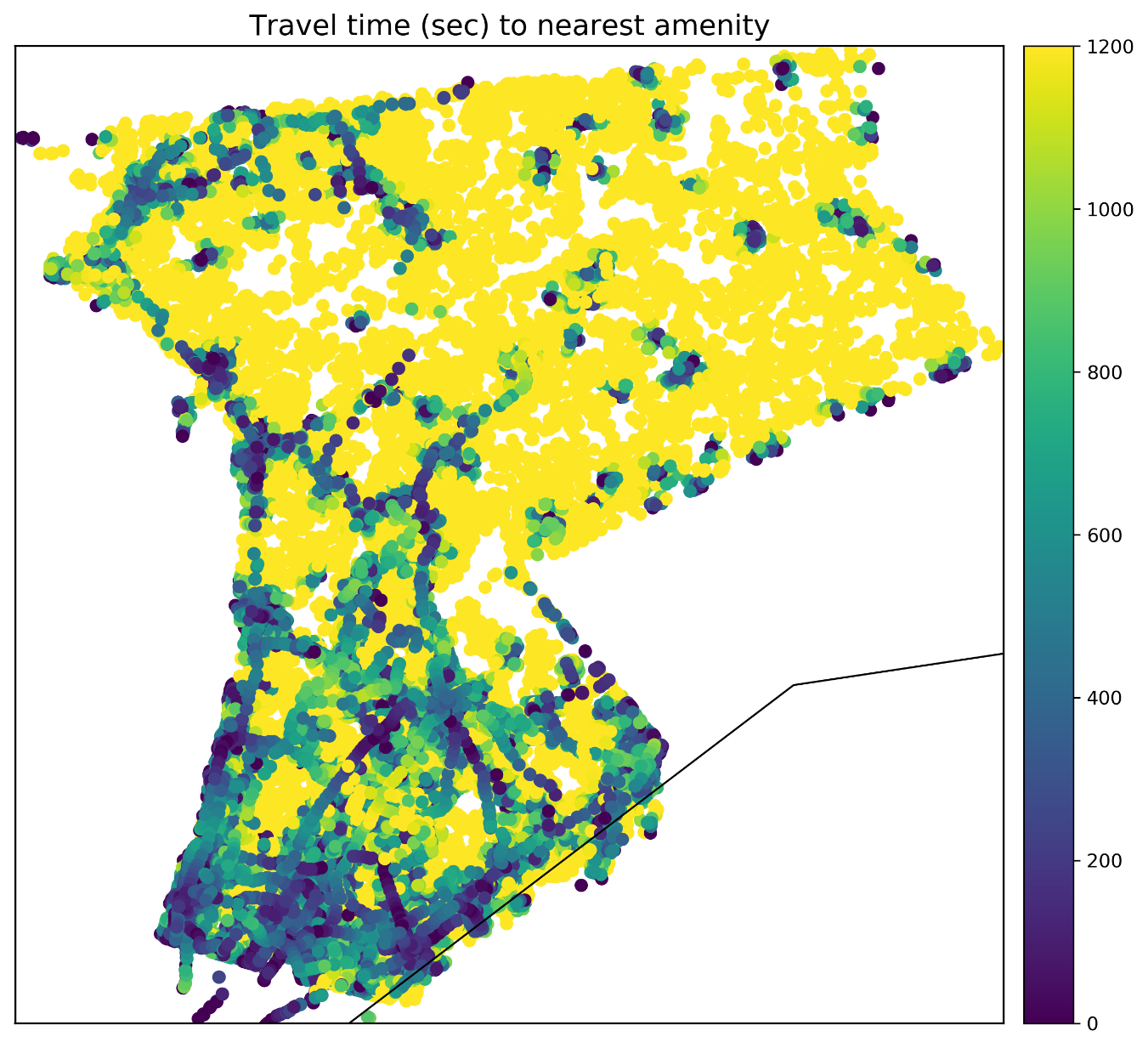
The first step for building the multimodal network for Westchester involved loading the transit network and the walking network separately and then combining them together. The GTFS data was handled using a library called peartree. Peartree reads the busiest day in the data and takes time inputs to calculate impedance along the network. For this project I used the morning rush hour time for impedance, which would be 7:00 AM to 9:00 AM. Peartree returns a network graph with edge weights represented as travel time. The pedestrian network was loaded using OSMnx, which has an edge weight of actual distance in meters. To combine these graphs together, the edge lengths in the pedestrian network were converted from distance in meters to travel time in seconds. Then they were combined by overlaying the graphs in peartree (Figure 1).

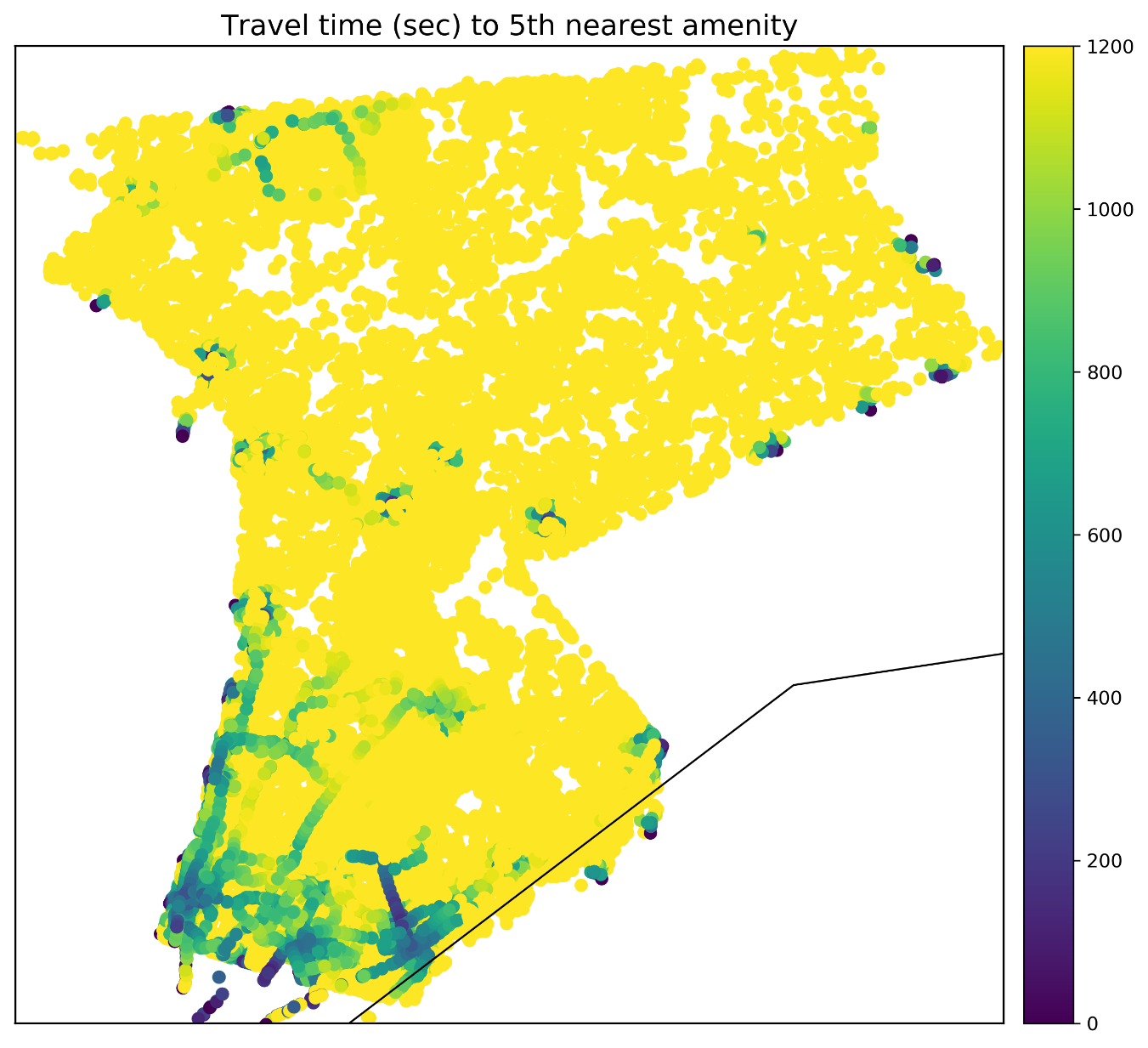
After creating the multimodal network, some analyses were run on the network using pandana (pandas network analysis), a Python library specifically designed for accessibility calculations on street networks. Pandana works by loading networks as nodes and edges pandas dataframes and then constructing a network. Impedance was travel time in seconds, recorded as the “length” attribute in the network edges (Foti et al. 2012). The first accessibility analysis run was on Westchester points of interest, pulled from OpenStreetMap using OSMnx (Figures 2, 3). The amenities chosen were: 'doctors', 'clinic', 'restaurant', 'cafe', 'school', 'bank', 'pharmacy', 'library', and 'cinema'. The second accessibility analysis was run on U.S. Census and LEHD data from 2015, which were combined together using CSV files and pandas dataframes (Figures 4a, 4b, 4c).

**Results and Discussion**

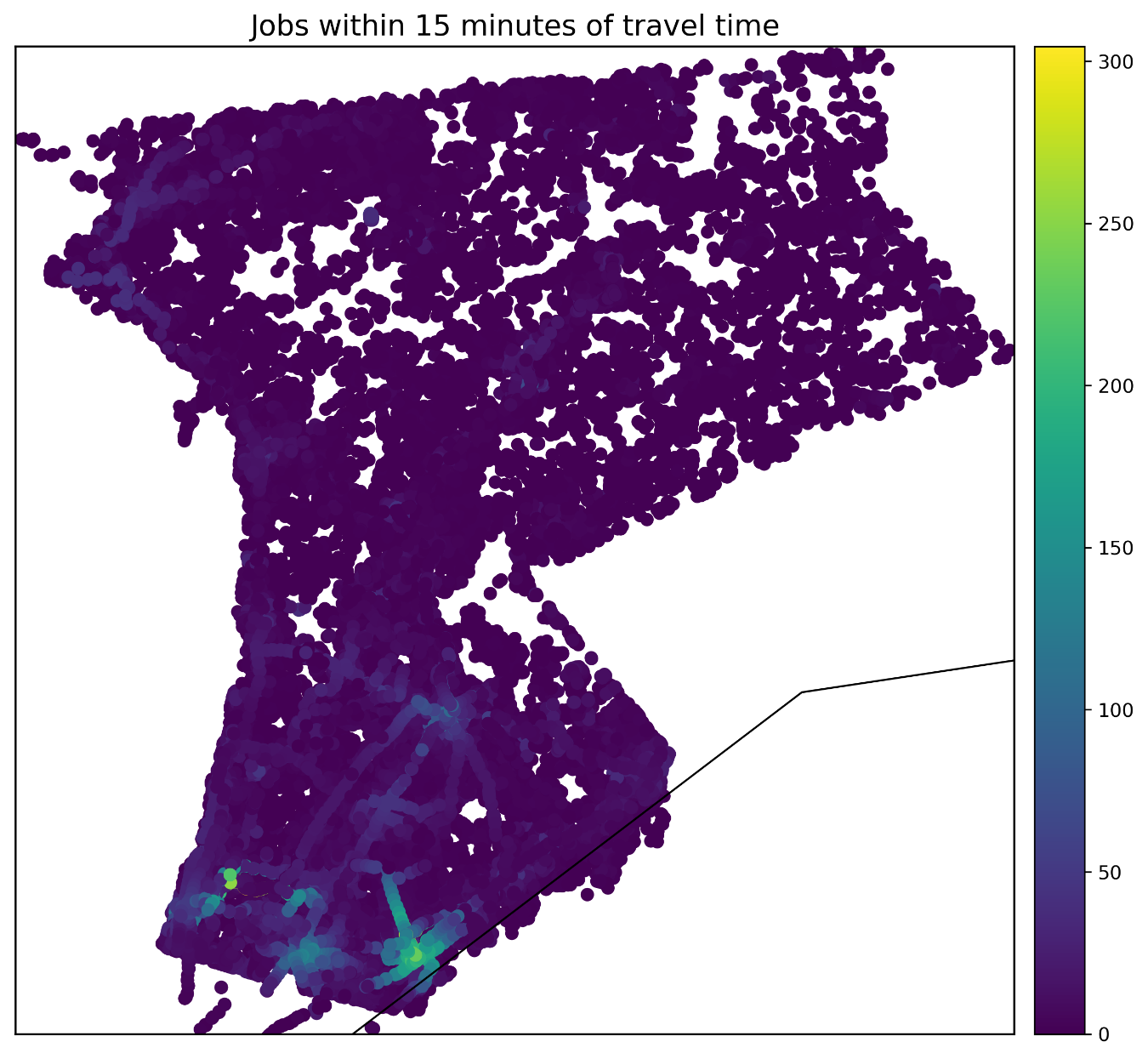
  
Figure 1: Multimodal public transit network with pedestrian and bus routes of Westchester County, New York.

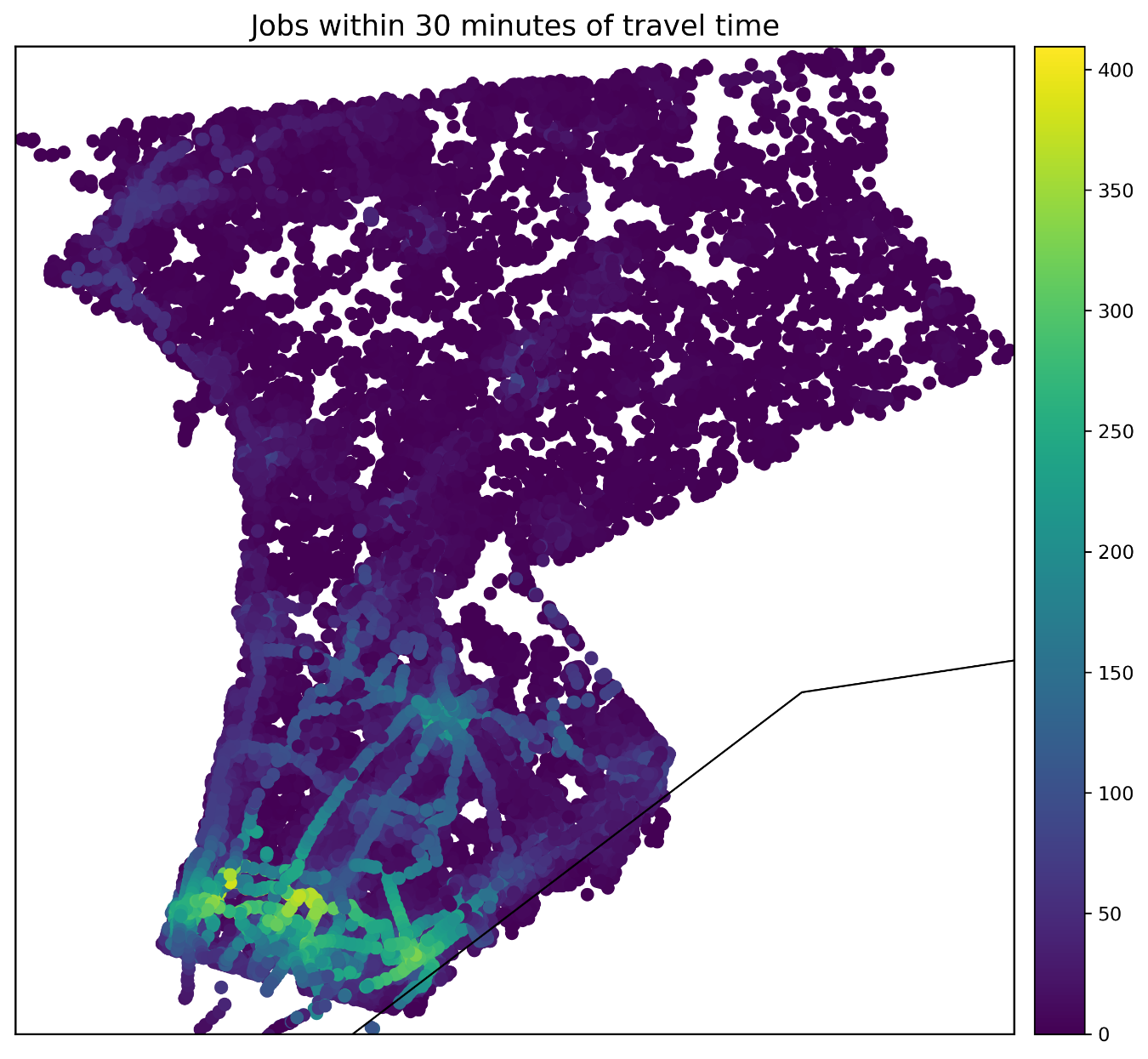
The first result was the multimodal network for Westchester County, as shown in Figure 1. This combined the pedestrian network pulled from OpenStreetMap and the bus route network created from GTFS data. There are 43,412 nodes and 113,348 edges in the network. The edge colors represent travel time in seconds, with the darker colors meaning faster travel times through the network. This was the network with which the accessibility calculations were run on, as shown in the below figures.

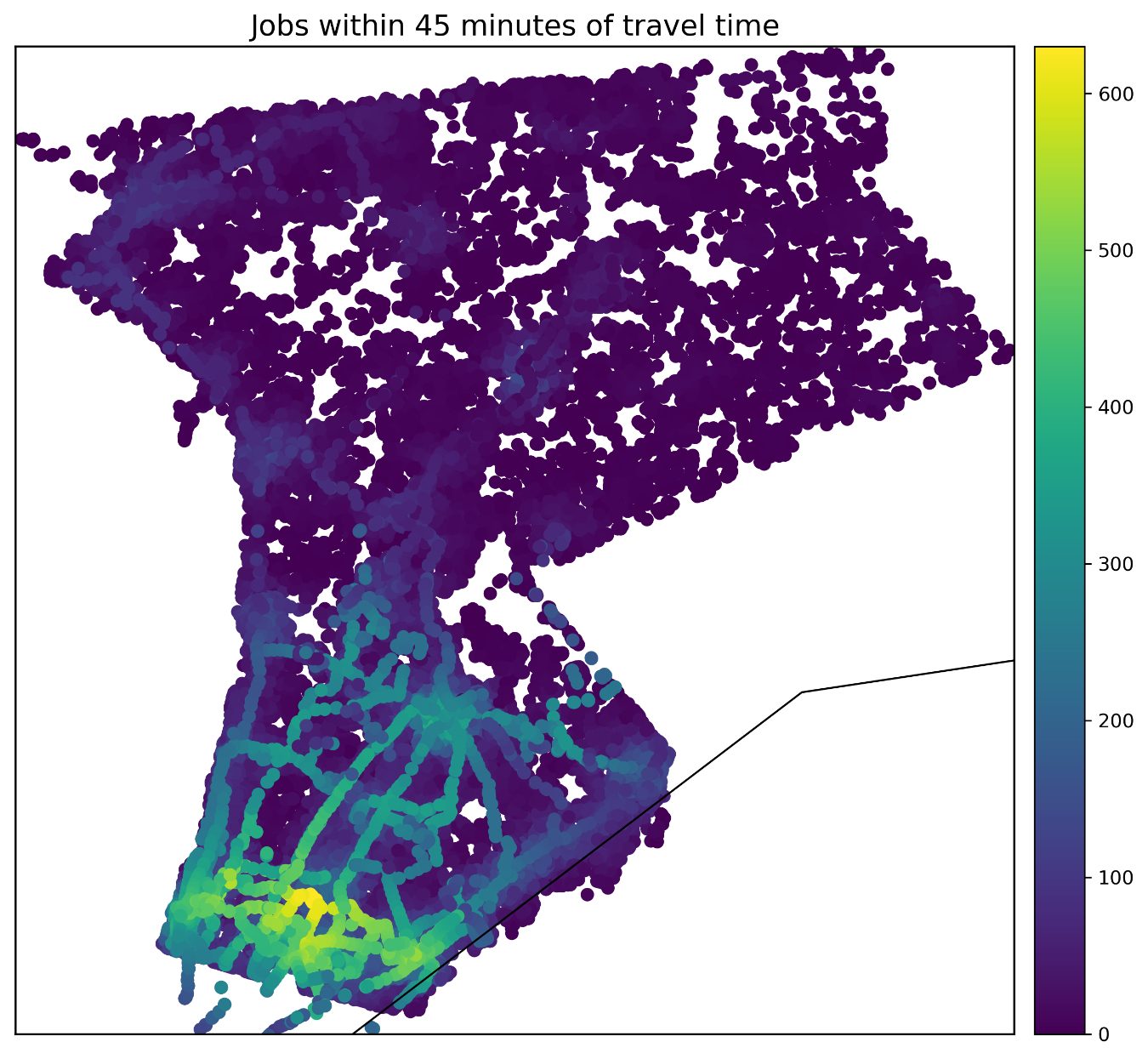
  
Figure 2: Travel time in seconds along the network to the nearest amenity.

  
Figure 3: Travel time along the network to the 5th nearest amenity.

Figures 2 shows the travel time from every node in the network to its nearest point of interest or amenity, regardless of its category. There were 1,238 total points of interests found. Figure 3 shows the same, except for every node’s fifth nearest amenity. The darker the colors, the shorter the travel time is to the nearest amenity. This measurement is up to 20 minutes of travel time along the network. Figures 4a, 4b, and 4c show travel time to jobs at the census block level as 15 minutes, 30 minutes, and 45 minutes respectively. These figures are reverse from the others; the lighter the color, the shorter the travel time is to the jobs found in the area.

  
Figure 4a: Jobs that are accessible within 15 minutes of travel time along the network.

  
Figure 4b: Jobs that are accessible within 30 minutes of travel time along the network.

  
Figure 4c: Jobs that are accessible within 45 minutes of travel time along the network.

It is clear from the multimodal network constructed in Figure 1 that the bus routes are obvious within the network. The addition of bus routes definitely improves a transit network. It’s also clear that southern Westchester has a greater concentration of bus routes and stops, so travel times in this area are overall much faster than in northern Westchester. Figure 2 also shows that much of southern Westchester has the greatest accessibility to amenities, and this is just proven further with Figure 3. Only a small portion of nodes in southern Westchester are within 20 minutes of travel time, as well as some clusters in the larger urban areas in northern Westchester.

Figures 4a through 4c also show a similar picture with access to jobs. Southern Westchester is much more densely populated with urban areas so it makes the most sense that more jobs would be located in these regions as opposed to in the much more suburban areas in northern Westchester. The jobs accessible within 15 minutes of travel time along the network are concentrated in the urban region of New Rochelle. Jobs accessible within 30 minutes spread across more of central southern Westchester. Job accessible within 45 minutes are similar, with some spread up north to the White Plains area, but concentrated in the southern region and along the bus route network. From all of these images, there is a clear bias in this network for southern Westchester as opposed to northern Westchester.

**Conclusions**

From this multimodal network and the resulting analyses, it’s apparent that southern Westchester is much further developed in terms of bus routes and walking than in northern Westchester. From the author’s personal experience, this makes the most sense, since the urban development in southern Westchester is much greater than in northern Westchester, especially due to its proximity to New York City. Therefore, not having a car and commuting in Westchester only with walking and buses makes sense in Southern Westchester but makes less sense in northern Westchester.

There is a clear disparity in accessibility between geographic regions in Westchester, but is there a disparity among other socioeconomic factors? There is a lot of future work that could go into this since the author’s main goal was to compute accessibility based on an index of transit need, which would be calculated from socioeconomic factors. This is similar to Siegel’s index of accessibility, while also using an origin-destination analysis to compute accessibility. There was also a final step in computing how accessibility could be approved in the bus system. It’s clear that moving some stops further north would increase travel speeds and decrease travel time from one node to another, but rearranging the network requires some other sort of measurement of the population like population density and socioeconomic need for buses.

Overall, there was success in building the multimodal network, however a lot of assumptions were made. This analysis was only conducted on a network with morning rush hour travel times, and it would be interesting to also do afternoon rush hours, off peak times, and weekend or holiday times. The resulting graphs are also multidigraphs, which means that there is direction along the edges and two nodes could have more than one edge connecting them together, which was not accounted for in normalizing the network. In the future, it would make sense to do multiple impedances along the network and use more census data to account for exactly what the accessibility is in a walking and bus route only network for Westchester County.

**References**

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, *65*, 126-139.

Foti, F., Waddell, P., & Luxen, D. (2012, April). A generalized computational framework for accessibility: from the pedestrian to the metropolitan scale. In *Proceedings of the 4th TRB Conference on Innovations in Travel Modeling. Transportation Research Board*.

Panetta, R., & Jackson, K. T. (2006). Westchester: The American Suburb. *New York State History*. 1. https://fordham.bepress.com/nysh/1

Siegel, Maxwell S. (2016). *Quantifying Transit Access In New York City: Formulating an Accessibility Index for Analyzing Spatial and Social Patterns of Public Transportation,* Hunter College, School of Arts & Sciences Theses.

Tribby, & Zandbergen. (2012). High-resolution spatio-temporal modeling of public transit accessibility. *Applied Geography,34*(C), 345-355.

U.S. Census Bureau. (2019). LEHD Origin-Destination Employment Statistics Data (2002-2015) [computer file]. Washington, DC: U.S. Census Bureau, Longitudinal-Employer Household Dynamics Program [distributor], accessed on 5/12/19. at https://lehd.ces.census.gov/data/#lodes. LODES 7.3.

U.S. Census Bureau (2018). QuickFacts Westchester County New York. Retrieved from https://www.census.gov/quickfacts/fact/table/westchestercountynewyork/PST045218.