Dane Estok
GEOG 586 – Spring 2015
March 9, 2015
Project 9

Elevation, Slope, and Gender – July 2013 Citi Bike Program

For my Final Project, I decided to investigate the Citi Bike Program in New York City, operated by the NYC Bike Share. Per the Citi Bike website, their services offer "a convenient solution for quick trips around New York City" (www.citibikenyc.com). Because this program was brought to my attention earlier in the class, and because I live in a city (Pittsburgh, PA) that is slowly building their bike share program, I figured it would be fruitful to investigate some aspect of the NYC Bike Share Program. Since Pittsburgh is extremely hilly, my initial interest regarding a final project with the NYC Bike Share Program revolved around the elevation and slope of the study area in New York City. Further exploration made me curious about the relationship between slope and the gender of riders. Because the Citi Bike Program data is enormous and because there are 332 bike stations, I decided to focus on two stations and their routes for a particular month. I arbitrarily picked to work with the July 2013 data from these two stations. The two stations chosen were Station 327 in Battery Park, and Station 420 in Fort Green. Station 327 has an elevation of 0 feet, and Station 420 has an elevation of 28.4 ft. Since these elevations represent the lowest and highest of the stations in the study area, I thought it would be interesting to see what story the data told between the two in regards to slope and gender. In particular, I became curious as to which routes to both stations were the steepest, and if they had an effect on gender. There were four stations with an elevation of 0 feet, but Station 327 was chosen since it is the station with zero elevation furthest away from Station 420.

The data used for the project were taken from the Citi Bike Share Program website at https://www.citibikenyc.com/system-data. Dr. Blanford and past students have cleaned up an enormous amount of data, which I explored, but I wanted to take the data straight from the website for this project. I downloaded a 25.8MB Excel file (.CSV) for the July 2013 dataset that contains 843,416 rows of individual trips. The dataset shows information such as trip duration, start and stop time, station information (start and end with lat/long), type of customer, gender, and birth year. For this project, I was only interested in the station start and end (lat/long), and the rider's gender. I downloaded three DEMs from the Cornell University Geospatial Information Repository that covered the areas of Central Park, Brooklyn, and Jersey City. Four stations fell out of DEM reach since I could not find DEM coverage. The following images show the study area and stations:

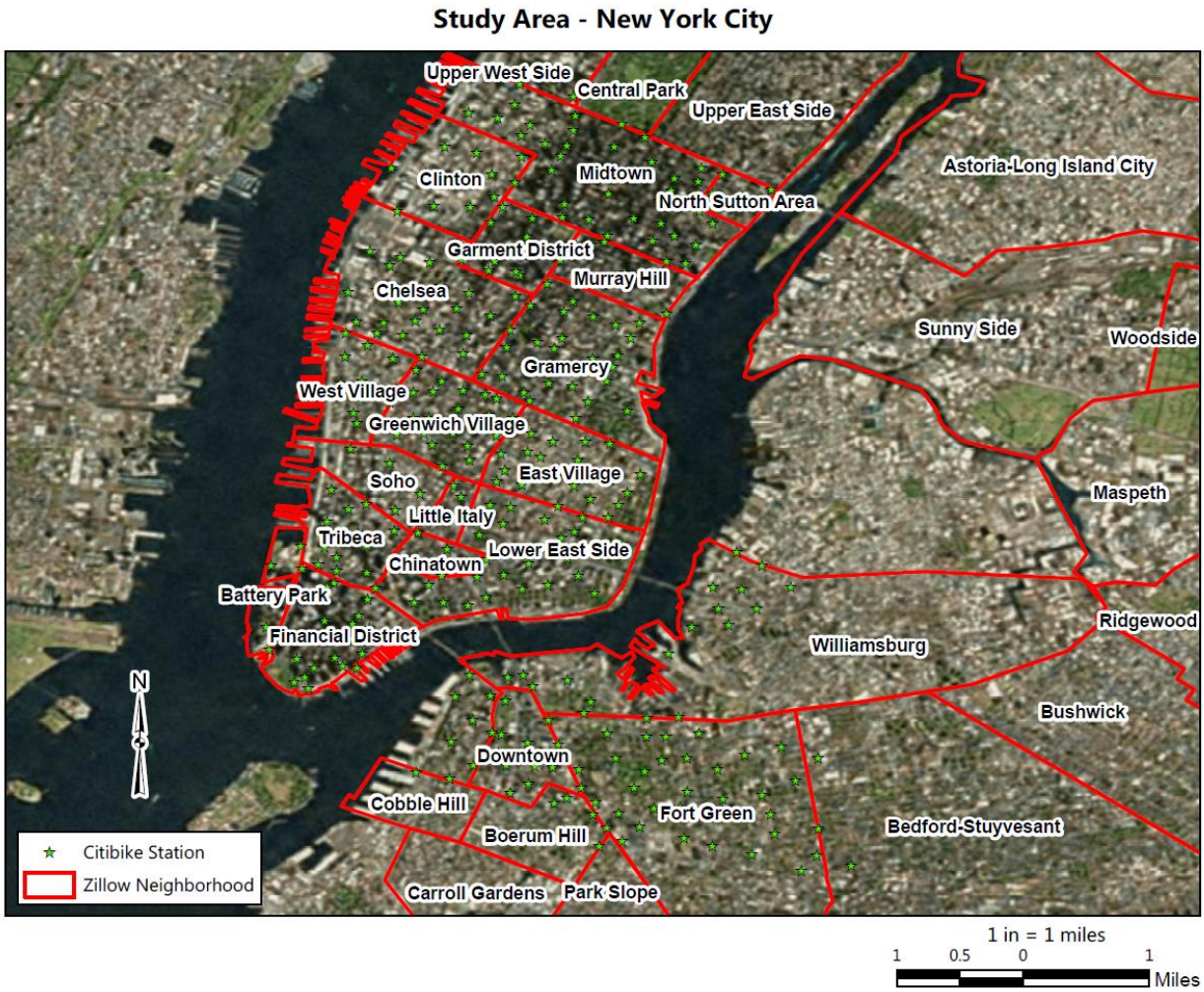

Figure 1 The study area for my final project. The red outlines represent neighborhood boundaries, and the green stars represent each CitiBike stations.

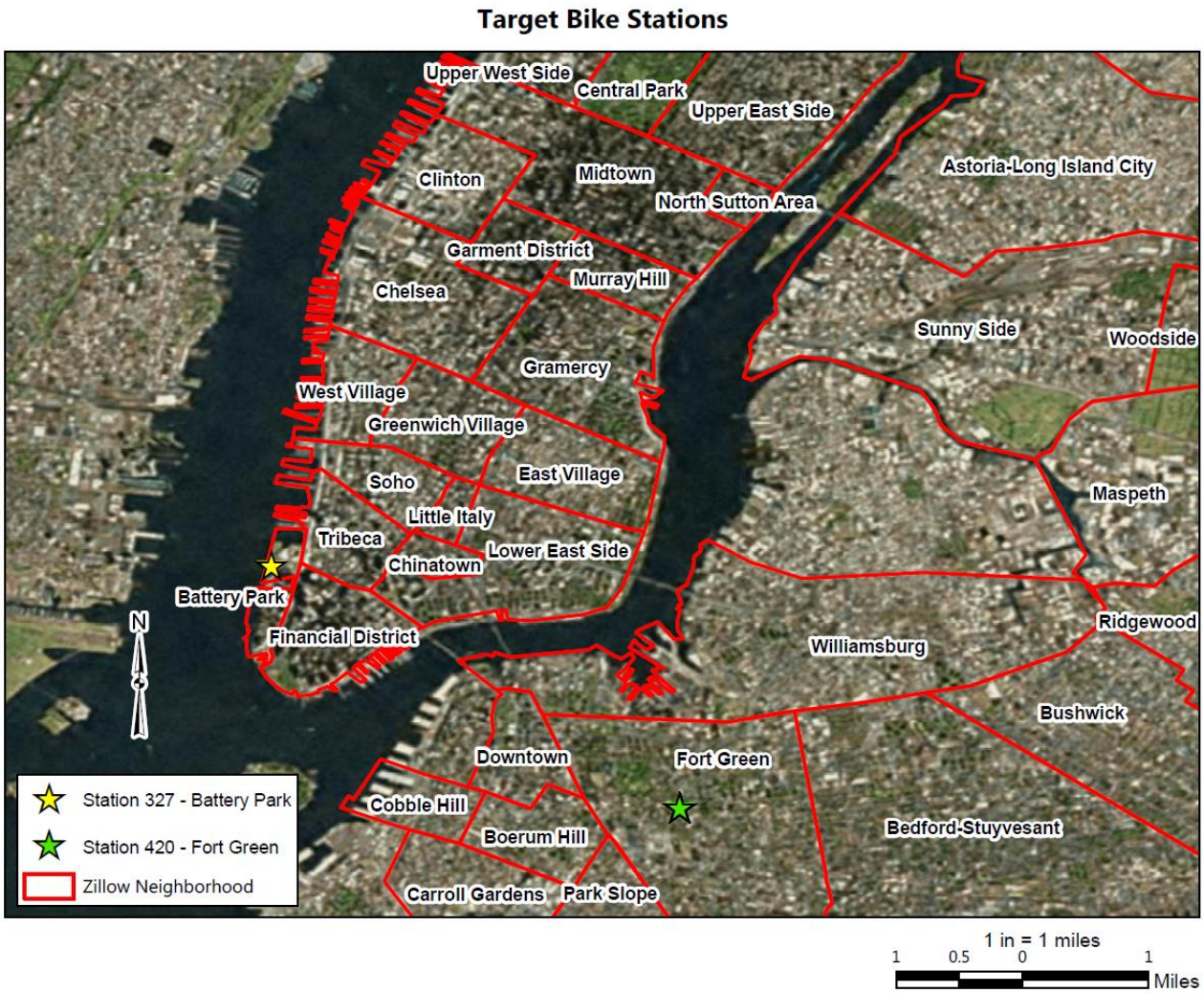

Figure 2 This map shows the two stations I chose for further analysis. The yellow star represents Station 327 in Battery Park, and the green star represents Station 420 in Fort Green.

After downloading the July 2013 dataset, I created two separate files based on the Ending Station ID. I filtered the data to show all the routes ending at Station 327, and Station 420. I changed the Lat/Long format to Numbers in Excel, and saved each file as a .csv. I imported each file into ArcMap, and added the station points by adding the XY Data using the station's Lat/Long. Once both sets of points were created, I converted the DEMs to raster so I could use the ArcMap tool "Extract by Points" to calculate the elevation of each station. Both station sets were then symbolized by surface elevation. Next, I wanted to create a general route from the starting station to the ending station. Because I don't have any data that shows the exact route each rider took, I had to generalize the trip by using the "XY to Line" tool in ArcMap. Using each trip's start and end Lat/Long coordinates, this tool created a straight line route representing the rider's journey. The following images show the results of these steps.

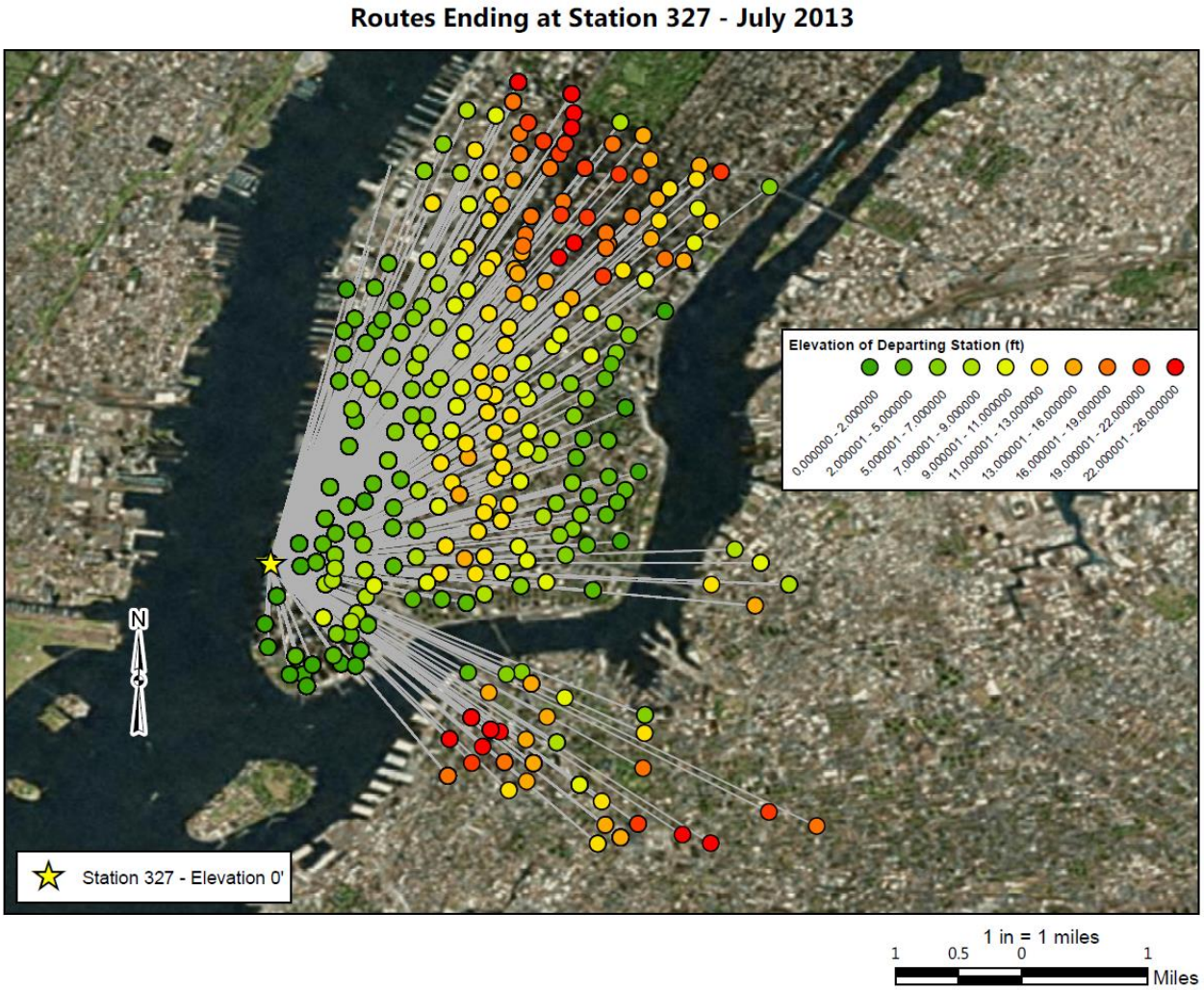

Figure 3 All routes traveled in July 2013 that ended at Station 327 (elevation 0 feet). Red represents stations with higher elevation, while green represents stations at lower elevations closer to sea level/zero.

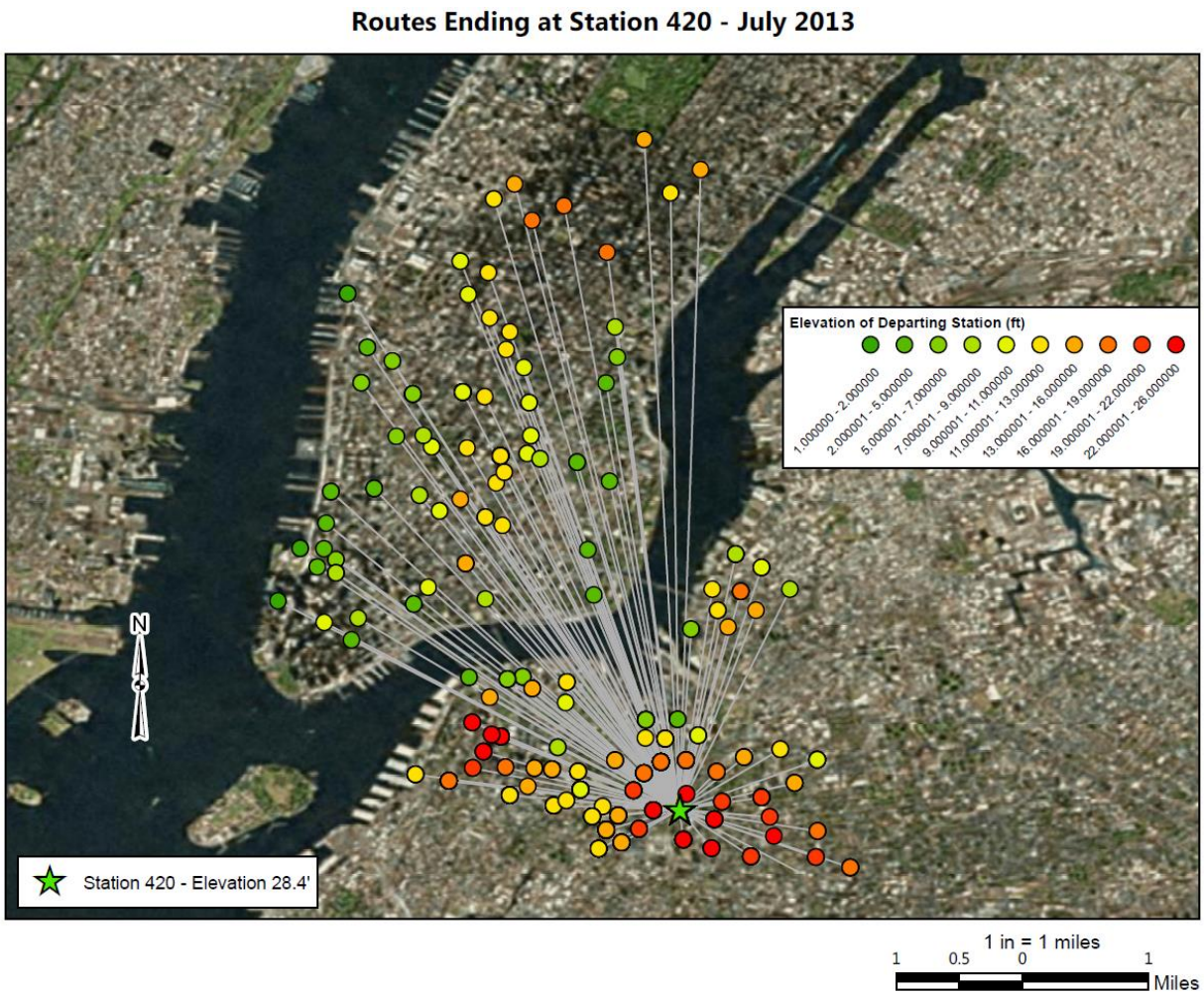

Figure 4 All routes traveled in July 2013 that ended at Station 420 (elevation 28.4 feet). Red represents stations with higher elevation, while green represents stations at lower elevations closer to sea level/zero.

It is interesting to see the maps showing the start station elevations with their routes to the ending stations. It provides an interesting perspective on the stations, and on the study area as a whole. Higher elevations can be seen towards the south of the study area in neighborhoods such as Fort Green and Boerum Hill. Lower elevations are apparent in West Village, Soho, Tribeca, and Battery Park.

After seeing a general trend in the station elevation data, it was time to investigate the general slope of the routes. Using the "Slope" tool in ArcMap, I was able to create a raster showing the slope of the study area. Next, I used the "Zonal Statistics" tool to create the "mean" slope for the routes. This tool analyzed the slope of each route's polyline, and created a raster with each route coded with the average slope of the whole route.

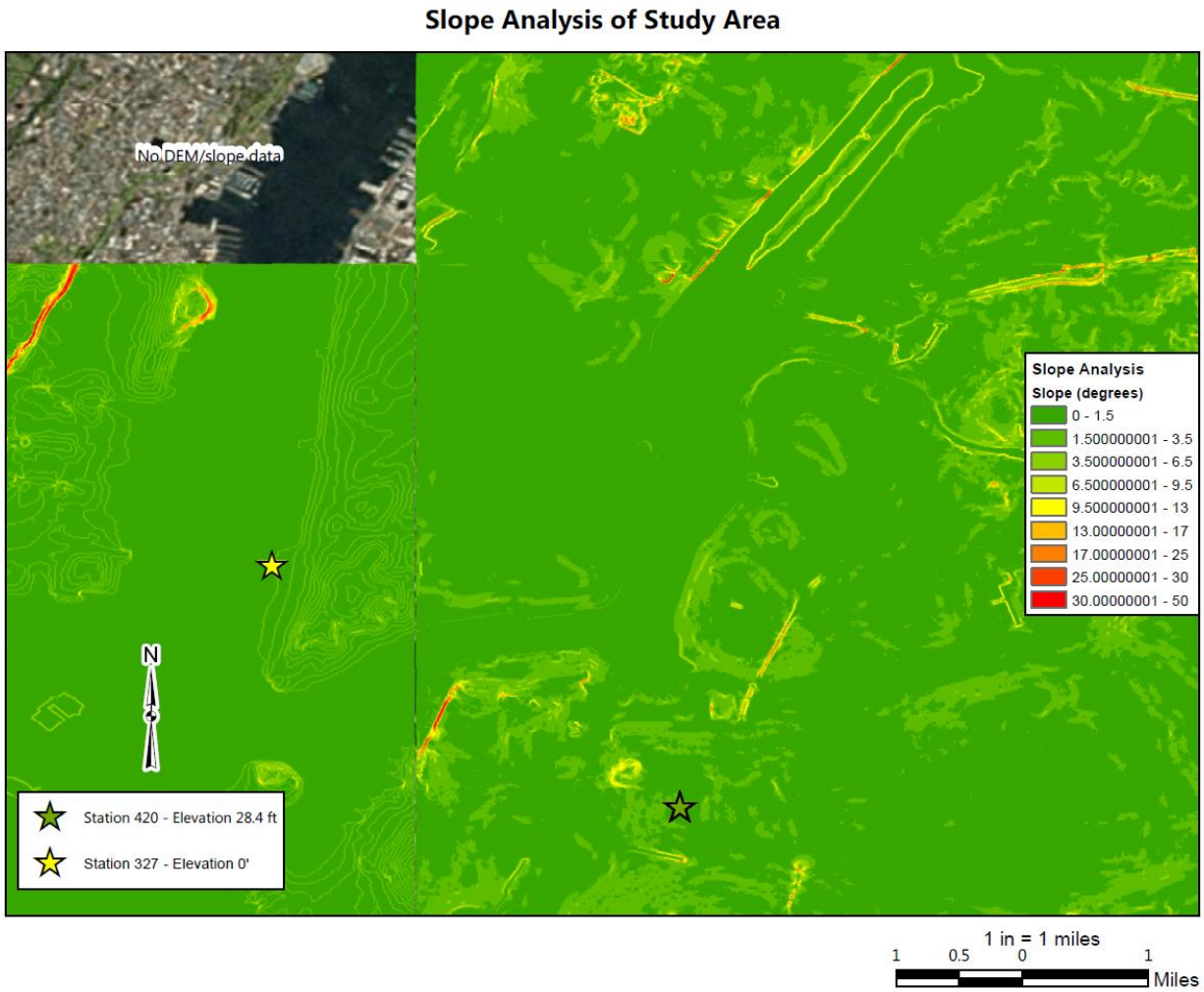

Figure 5 Map of the study area showing slope. Areas of green represent low slope and areas of red represent steep slopes. The upper-left corner of the map doesn't have data due to the lack of a DEM.

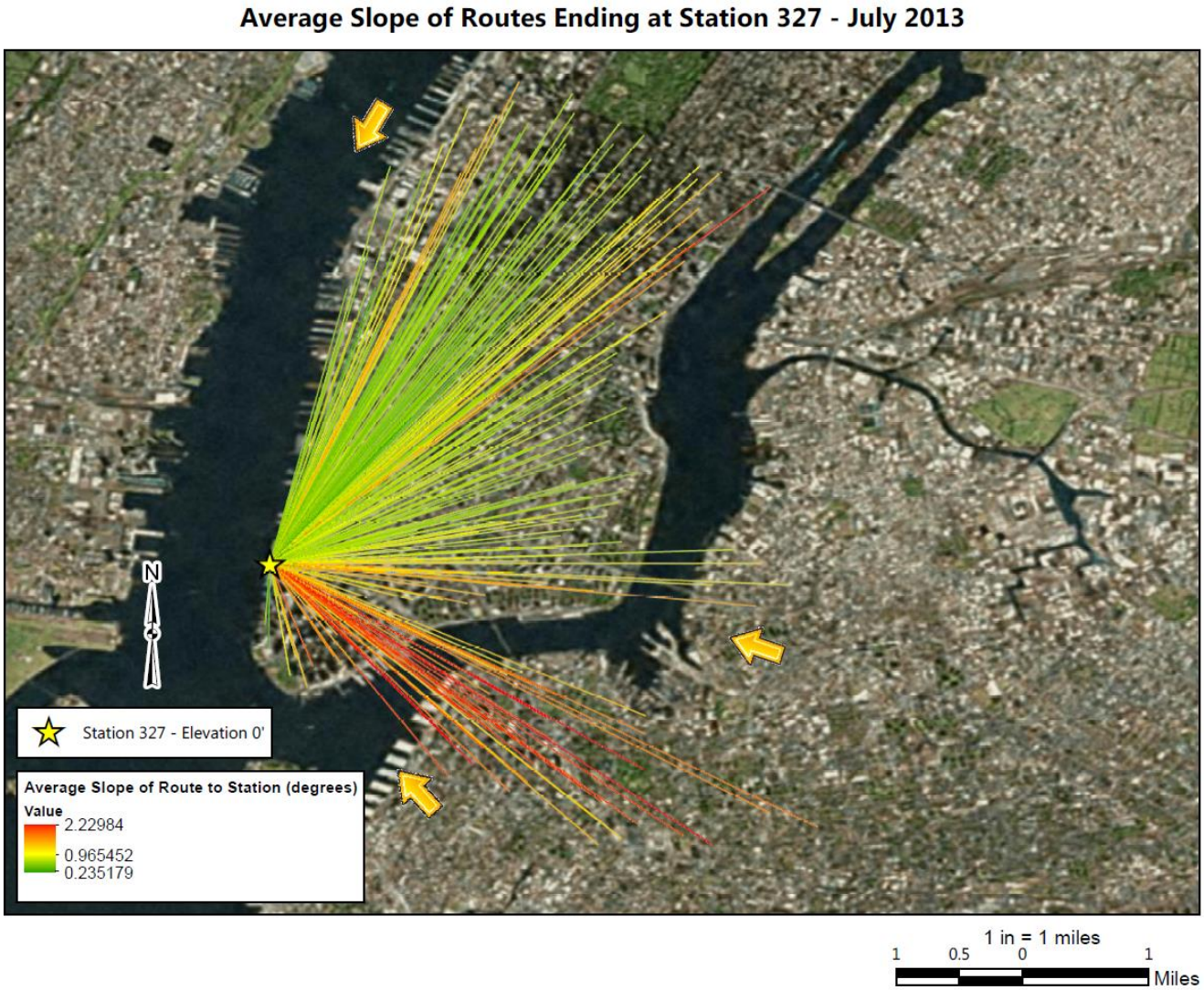

Figure 6 The average slope of the routes ending at Station 327. Green lines represent routes with a generally low average of slope, while red line represent areas of generally high average slope. The yellow arrows point towards the direction of travel.

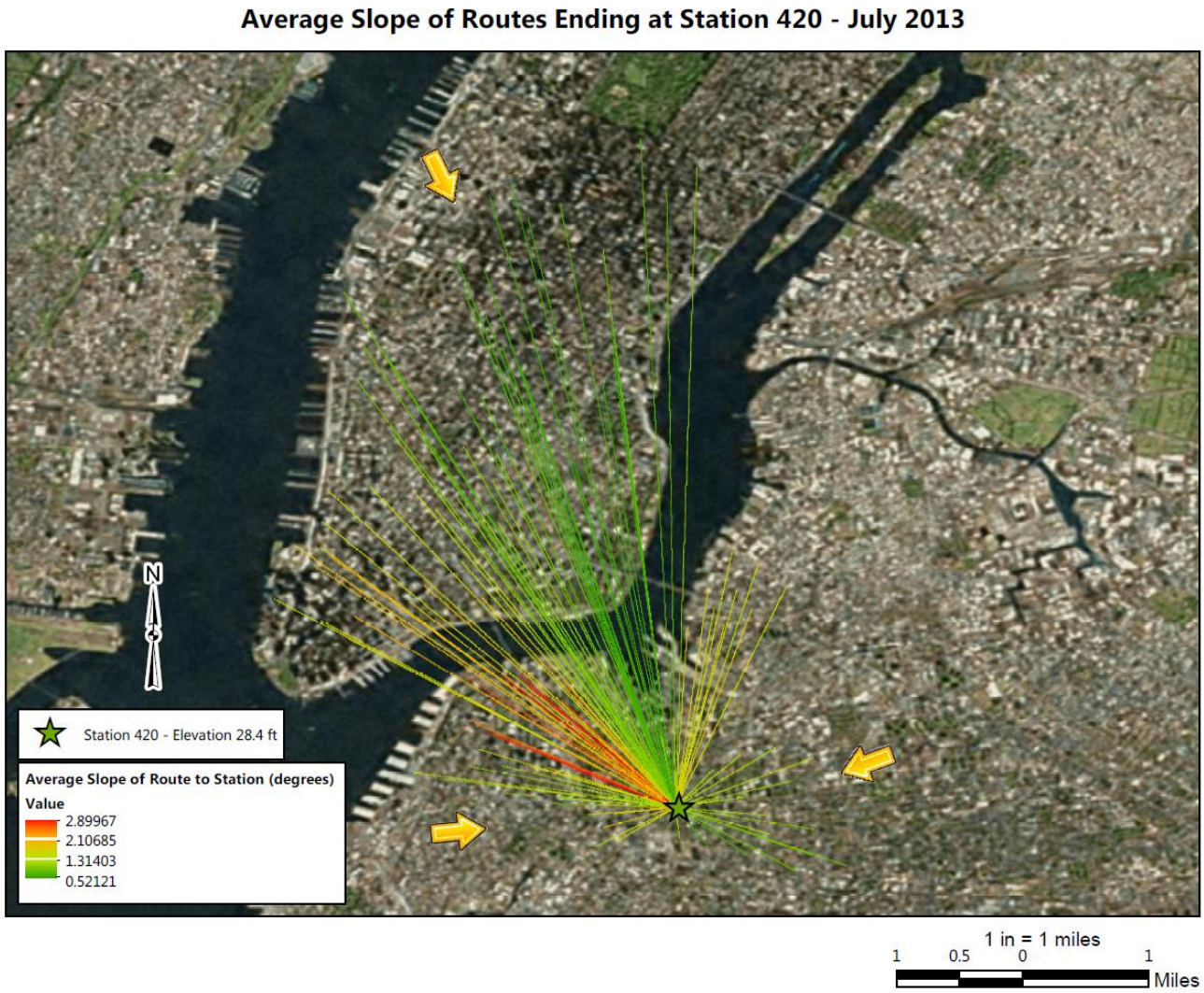

Figure 7 The average slope of the routes ending at Station 420. Green lines represent routes with a generally low average of slope, while red line represent areas of generally high average slope. The yellow arrows point towards the direction of travel.

Next, I wanted to find out the route with the greatest slope to each station. In order to do this, I looked at the highest slope value from the maps in Figure 6 and 7, and used the raster calculator to produce a raster showing the specified equation in the calculator. For example, for the routes to Station 327 (Figure 6), I used an equation to find routes GREATER THAN OR

EQUAL to 2, and for the routes to Station 420, I used the equation GREATER THAN OR EQUAL to 2.8. The following images describe the results:

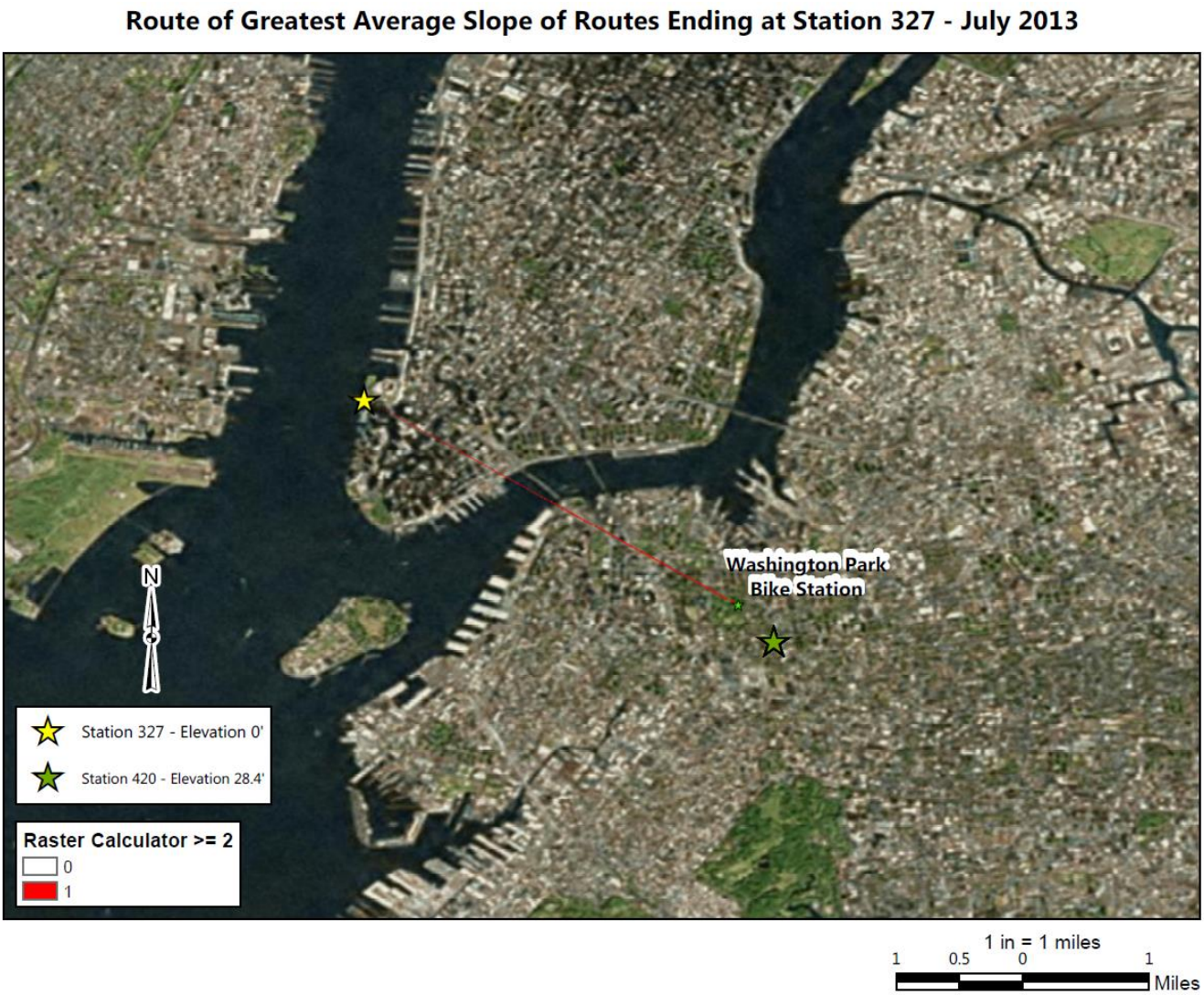

Figure 8 The red line is a product of the raster calculator, and represents the greatest average slope of all routes ending at Station 327 in July 2013. The red line equals an average slope of 2.22 degrees at the origin station of Washington Park.

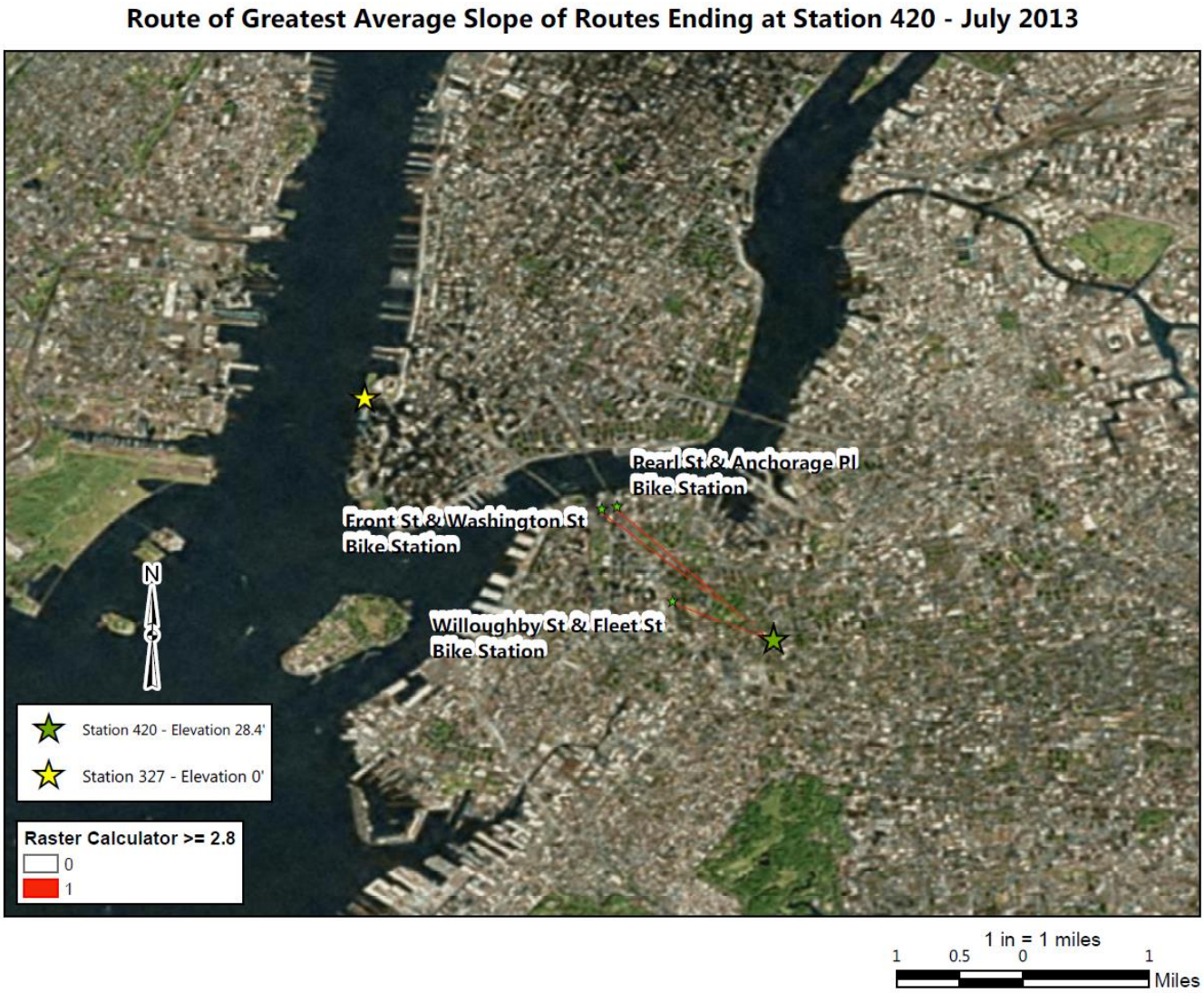

Figure 9 The red line is a product of the raster calculator, and represents the greatest average slope of all routes ending at Station 420 in July 2013. The red lines equal an average slope of 2.8 degrees or greater. The route originating at Willoughby & Fleet had the greatest average slope of 2.89 degrees.

After discovering the routes with the highest average slope, I decided to look at the gender statistics for all trips ending at Stations 327 and 420 in July 2013. The results were interesting. Since Station 420 was the highest elevation, the trips ending here were the most difficult. Station 327 had the lowest elevation, so trips ending here were most likely, on average, the easiest. I wanted to see if this affected the gender of the riders. In both cases, males rode more trips than females, and the following graphs and charts help explain this:

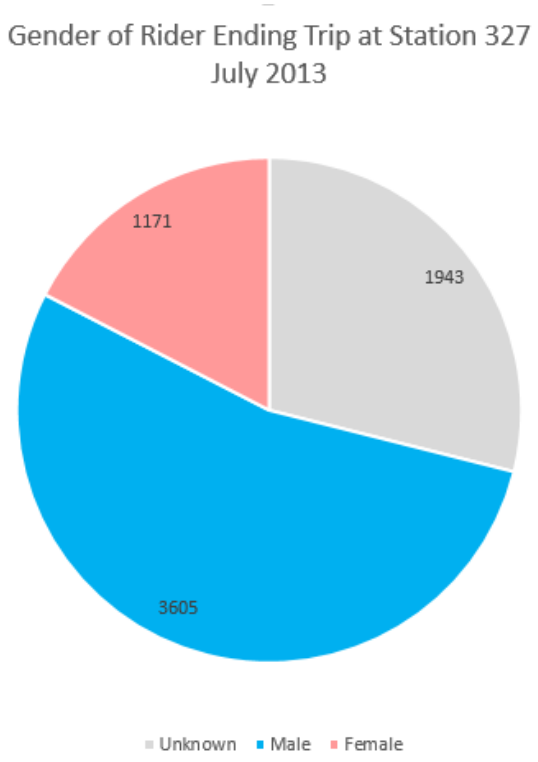

Figure 10 This pie chart shows the number of riders, per gender, that took trips that ended at Station 327. There were 278 unique routes ending at Station 327, and 6,719 total trips. 54% = male, 17%= female, 29% unknown.

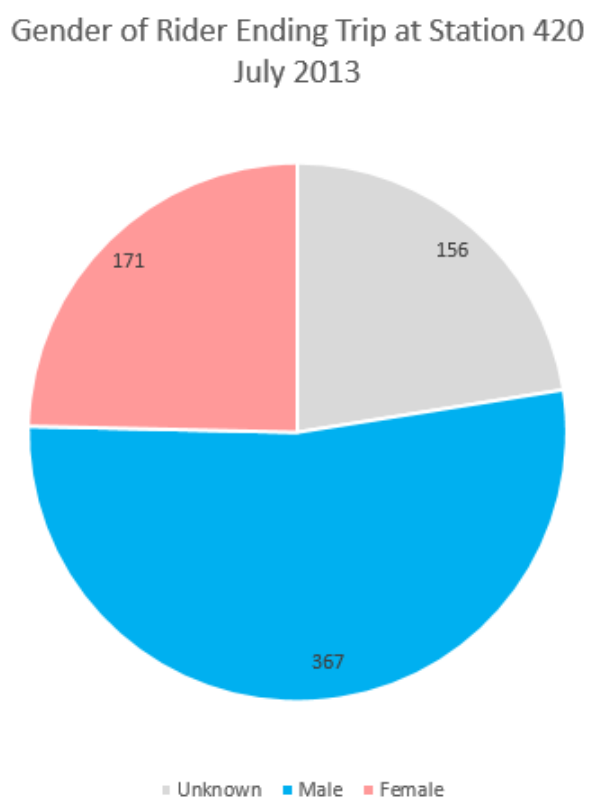

Figure 11 This pie chart shows the number of riders, per gender, that took trips that ended at Station 420. There were 131 unique routes ending at Station 420, and 694 total trips. 53% male, 25% female, 22% unknown.

It is interesting to note that although males dominated the trips ending at both stations, there was an increase in the total percentage of female riders ending at Station 420, which was the station with the highest elevation. Another question I had was the relationship between gender and the elevation of the origin stations. Using the Summarize tool, I found the following information:

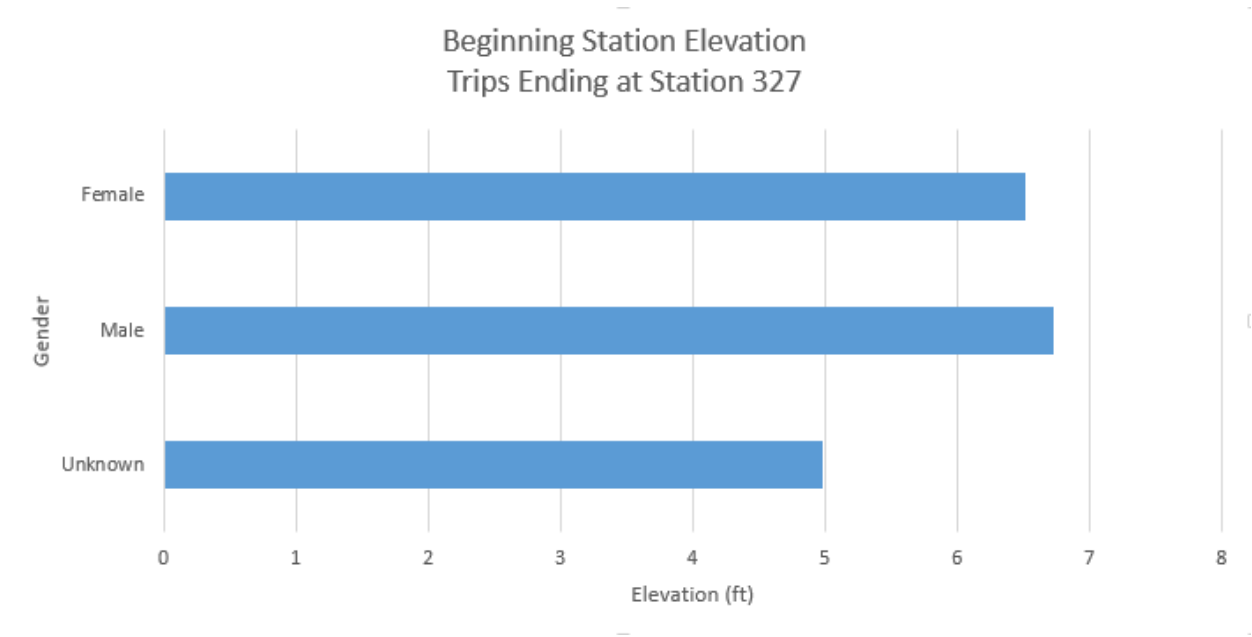

Figure 12 This bar chart shows the relationship between gender and the average elevation of the origin station that ended at Station 327 (the lowest station). Males tended to start at higher elevations compared to female riders, but not by a large amount at all.

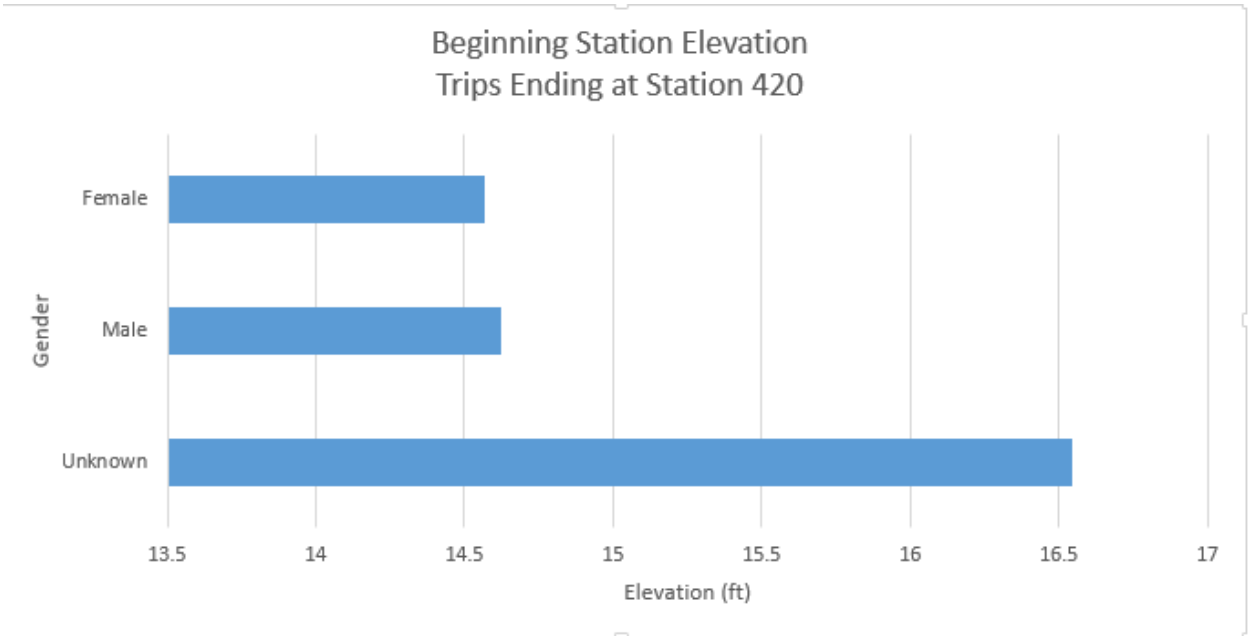

Figure 13 This bar chart shows the relationship between gender and the average elevation of the origin station that ended at Station 420 (the highest station). Males tended to start at higher elevations compared to female riders, but not by a large amount at all. It is interesting to see the higher elevations of origin stations by riders with an unknown gender.

This project was a lot of fun. Initially, I had intended to look at more months of data, but I soon realized the enormous amount of time it would take to work through multiple months of data. I think it was valuable to find the elevations of the stations, especially the ones I was taking a closer look at, in order to understand the context of the study area better in terms of elevation and slope. The downside of a project like this is that, as mentioned earlier, we don't know the exact route taken by the rider. So, in order to get something out of the route data, we needed to generalize the routes as straight lines from the origin XY to the destination XY. It is possible that the rider took a route that didn't involve much of the extreme slopes factored in the average slope of the routes I investigated. Before this project, I had an assumption that more males would travel routes with the greater slope than female. The charts in Figures 10 and 11 show that yes, there were more male riders taking trips that ended at both stations, but there was an increase in the percentage of females that took the trips that ended at the highest station – Station 420. I don't know the reason for this. Much more time would have to be spent investigating cultural influences in the different neighborhoods in NYC. It would be interesting to note where the schools and colleges are located in relation to the bike stations.

If I had to do this project again, I would figure out a way to use the Cost Allocation tool in ArcMap. I think it would be interesting to weigh the routes based on their relation or distance to some point of interest, such as nightlife venues/bars or locations of education. This would take more time than this class would allow, but it might be fruitful to see if further analysis could somehow predict the gender use for each route. However, I accomplished the goals I set before this project began. I wanted to see the overall pattern of station elevation, create routes that factored in an average slope, and produce some basic numbers regarding the number of riders and their genders for the routes ending at the lowest and highest station elevations during July of 2013.