Part I: Introduction

The bus. We all know it as this behemoth of a vehicle that roams around the city streets and highways picking up people and dropping them off elsewhere. You can generally know where it is by the roar of its large engine, or perhaps the lights showing the route number and destination, or even the large flashy advertisements on the sides of the thing. We all know what the bus is, and chances are we have made some use of it. Public transportation is an integral part of any city's infrastructure. For many, it as seen as a convenient way to get from one place to another without having to own a car or walk everywhere. It is also seen as a cleaner alternative to driving as more people using public transit means fewer cars on the road, resulting in fewer emissions. As such, many rely on the bus to travel to work, or even to school. No doubt living near a bus stop will make things more convenient. One could expect more people to live near public transit out of convenience. But of course, with higher convenience comes higher rent as well, which may lead one to suspect that those with more money would have a better chance at living in higher rent areas. This brings us to our research question: Do areas near more bus stops have a higher population density, and do those in areas with more easier access to buses have a higher income? To answer that, we will be using data to look at public transportation in the city of Seattle.

Part II: Data Sources and Preparation

For this project, I ended up looking to King County's official GIS department to look for data that I needed. I managed to find a geodatabase file containing all sorts of transportation information about king county; the layers that were of particular interest to me was a layer that mapped out bus routes as well as another layer that mapped out bus stations all over King

County. These have all sorts of attributes, such as route numbers, although I did not make use of them here. I also managed to find a geodatabase file containing all sorts of census data about King County, as well as a layer that mapped out the most recent census tracts in the county as blank polygons that I could attach additional data too, such as from various tables regarding statistics such as income that were also included in the file. However, it did not seem to include data for the total population in each census tract, so instead I had to rely on excel spreadsheet from the U.S. Census Bureau for the total population as well as population density and land area.

Each polygon in the base census tract layer had an attribute called a GEO ID, which was essentially an eleven-digit identification number assigned to each census tract. It was present as a common attribute amongst all the census data tables that I used, for joining additional data tables to the base layer using the GEO ID. I used this to join the excel data onto the base tract layer, as well as additional data tables from the King County census geodatabase that contained statistics on income per capita. This income per capita layer was based on estimates from 2011-2015, taken over the course of twelve months until the data file was released. The totals were all adjusted for inflation in 2015. I also linked a table for median family income, but I didn't end up using it.

Since I only wished to study Seattle rather than all of King County, I had to prepare my base layer to only show Seattle. For that, I used a layer from an administrative geodatabase file from King County that represented city borders as polygons. I selected the polygon named "Seattle" and gave it its own layer. I then did a spatial-selection query to select all the census tracts that had their centroid ("center of mass," so to speak) located within the Seattle polygon, giving me a selection of census tracts that approximated the borders of Seattle, although not entirely perfectly, but still close. I made this selection of tracts into its own layer, which became the new “base” tract layer for my analysis. I then performed a clip on the bus stop and route layer, using the new Seattle tract layer as a cookie-cutter to clear away stops and routes that weren’t in Seattle proper. After clearing away the excess layers, and having joined data to the Seattle tract layer, I was ready to start performing my analysis.

Part III: Analysis and Mapping

I realized I would somehow need to come up with a way of gauging how much land in a tract in considered close to a bus stop, so I decided to create a buffer around all the bus stop points. I decided close would be defined as “within reasonable walking distance,” which I discovered is accepted to be about a quarter-mile, which is approximately a five-minute walk for most people. So, I ended up creating a 0.25-mile radius buffer around the bus stop points; this buffer painted out the areas of Seattle that would be within reasonable walking distance of bus stops. I then clipped the buffer using the tracts layer to get rid of portions of the buffer that extended outside the city or into bodies of water, as those would be “out of bounds.” (Plus, I doubt one is going to crawl out of the Puget Sound to try and catch a bus.)

I decided it might be effective to measure the percentage of land area of each tract that was within the buffer, so I ended up performing a union between the now-clipped buffer and the tracts layer, but not before creating a copy of the “shape-size” column in the tracts layer, so that I could use the total area of the whole tract as the denominator for calculating the percentage of land within the buffer. The new layer from the union dived the many of the census tracts into two features each, one that overlapped with the old buffer layer, and a part that did not; these pairs actually shared some common attributes, including the GEO ID of the tracts they were originally both part of. Other tracts remained whole because they were entirely overlapped by the buffer.

Each new feature had an FID attribute that indicated whether they were a part of the original buffer or not; those that were had a value of 1, while those that did not had a value of -1. I used that attribute to select all the features that had a value of 1, and created a new layer from the selection. This layer covered the same area as the buffer, but was now divided into multiple polygons based on the census tract boundaries; each feature also now had data for their respective tracts as well, including the duplicate column I had added for the total area of each tract. The regular shape area attribute had changed to reflect the new shapes' area, hence why I duplicated it in the tracts layer before performing the union.

I added a new field to this new buffer layer for the percentage of the total tract area each buffer polygon took up, which I calculated by taking the shape area attribute of the new buffer layer and dividing by the whole tract area attribute and multiplying by 100, giving me the percentage or each census tract that was within 0.25 miles of a bus stop. I then proceeded to add that data to the old Seattle tract layer, but for some reason I had to duplicate the new buffer layer in order to join the data properly.

The final step of this analysis, which I probably could have done much earlier, was figuring out how many bus stops were in each tract. To figure this out, I decided to perform a spatial join between the tracts layer and the bus stops point; as it was a one-to-many join, I had to do a summarized join to link the data, although the only summarized data I needed was the count of how many bus stops were within each tract. Due to weird issues I was encountering with joining, I had to create a duplicate of the tracts layer to produce a successfully joined layer with bus stop count, and this new layer actually got disconnected from the data I had connected the old layer, so I had to exclusively use this new tract layer for displaying the results of the stop count. But, after I cleared up all the extra layers, I was ready to begin mapping. I decided to create multiple choropleth maps that would be compared to each other, each showing different types of data. Each map used largely the same basic layout and colors, as I wanted to keep things simple and easy to compare. I also decided to classify my data using the equal interval method, as that method is supposed to be ideal for comparing maps. The first few maps did not include the buffer layer, but it was implemented in the later, more complex maps.

Part IV: Interpretation of Results

Note: All maps are included at the end of this paper due to their size.

The first map, Figure 1, displays the number of bus stops per square mile of each tract. This resulted in a simple map showing the density of bus stops throughout each tract in Seattle, and which tract had more relative to their size. Unsurprisingly, the most stop-dense areas appeared to be downtown, which makes sense since it's the center of the city and many commute there for work.

Figure 2 is similar to figure 1, but displays population density instead. Again, the densest portions of the city seem to be around downtown, although more inland this time, more towards Capitol Hill or South Lake Union. It makes sense, since the center of downtown where the previous map showed dense bus stops is mainly commercial rather than residential; more residents would be seen at the edge of downtown and in the surrounding neighborhoods.

Figure 3 depicts the average number of people for each bus stop in each census tract. The densest area of this map corresponds with the most densely populated areas shown in the last map, which makes sense as a smaller, more population dense area with fewer bus stops would mean more people for each stop. Interestingly, there is another dense spot near the University of Washington campus. Another similar spot can be seen in the previous map. The denseness around here is probably due to the amount of students living near campus, and the fact that many of the stops here serve multiple routes, requiring fewer stops.

Figure 4 is the first map that includes the area within walking distance of the stop, and also maps out the per capita income of each tract. I chose diagonal lines for the buffer zone so that the colors of the tract underneath would still be visible without distorting the original color. The buffer is shaded based on the percentage of area in each tract that it occupies. Not surprisingly, the buffer is largely situated downtown where most buses go, with the buffer completely covering a lot of the tracts in that area, meaning you shouldn't have to walk further than a quarter mile in that region to catch a bus. Interestingly, the highest per capita income is not located in the area with most stop coverage, but instead is situated out in the East. I suspect this might be where the particularly well-off folks live, or maybe I'm just not looking at it right. But from this map, we cannot say that increased bus coverage results in higher income people living nearby.

The final map, Figure 5, once again displays population density like Figure 2, but also includes the bus coverage buffer to compare areas within walking distance to stops with areas with more people per square mile. Not surprisingly, a lot of the dense areas also have more bus coverage, while the areas with less bus coverage tend to be less densely populated.

Part V: Conclusion

From this analysis, and by comparing the maps, we can see that populations does seem to be denser in areas with more convenient access to public transit. However, some areas like downtown may still be less populated while having more stops simply because they are commercial zones where people go to work rather than residential. It would seem, however, that higher income people are not necessarily more likely to live near area with more bus coverage than others; further testing is needed.

Personally, I feel like I could have done a lot more on this project, like analyzing population, transit, and schools, but due to time constraints I had to simplify it. Were I to do this project again, I would spend more time coming up with a solid plan for analysis, as I sort of went with the flow here. But overall, I'm glad I got to show off my GIS skills, and I feel like I showed how I can use spatial analysis to compare relations between different phenomena. I guess I see on the bus trip to future improvement.

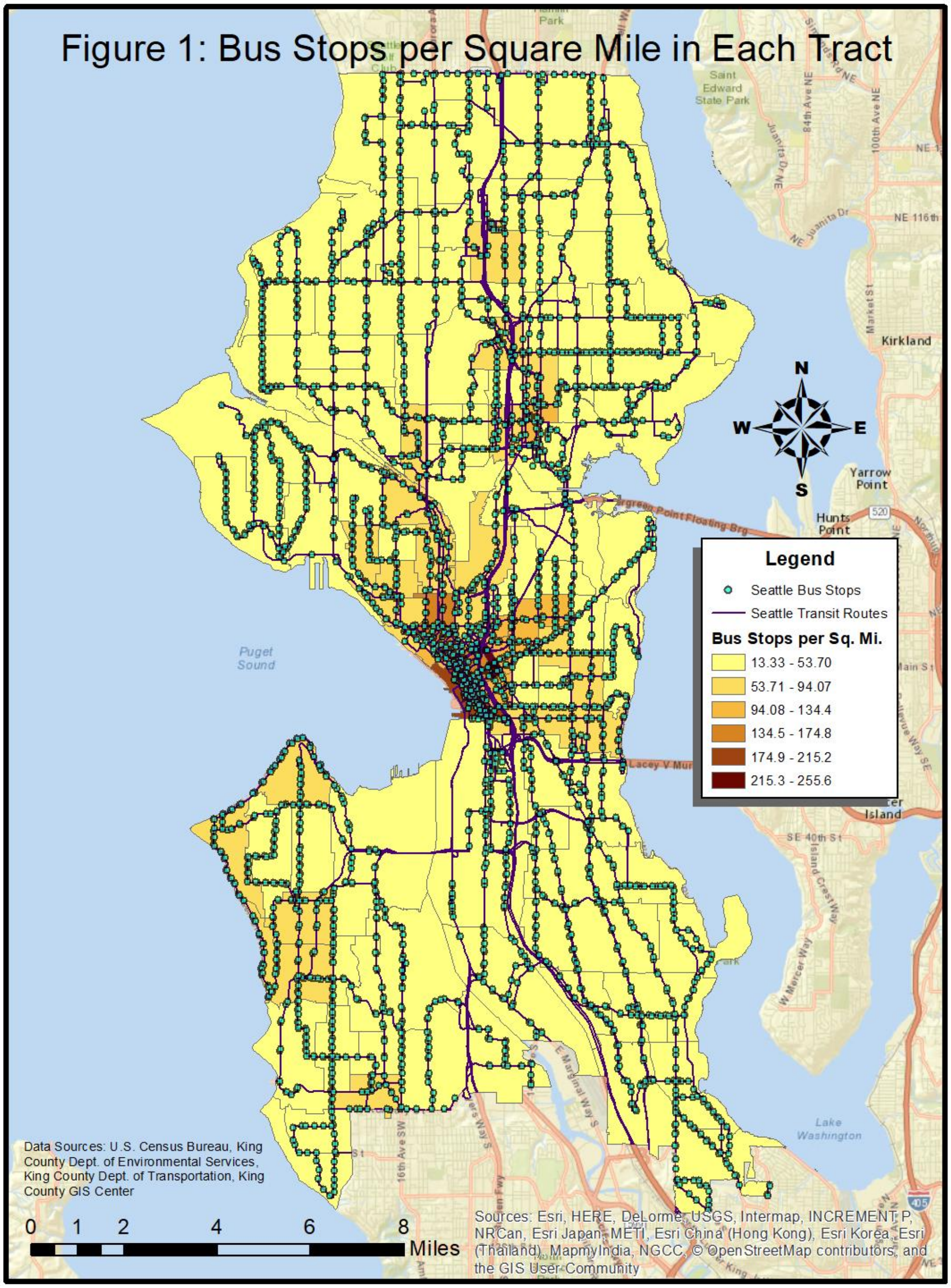


Figure 1: Bus Stops per Square Mile in Each Tract

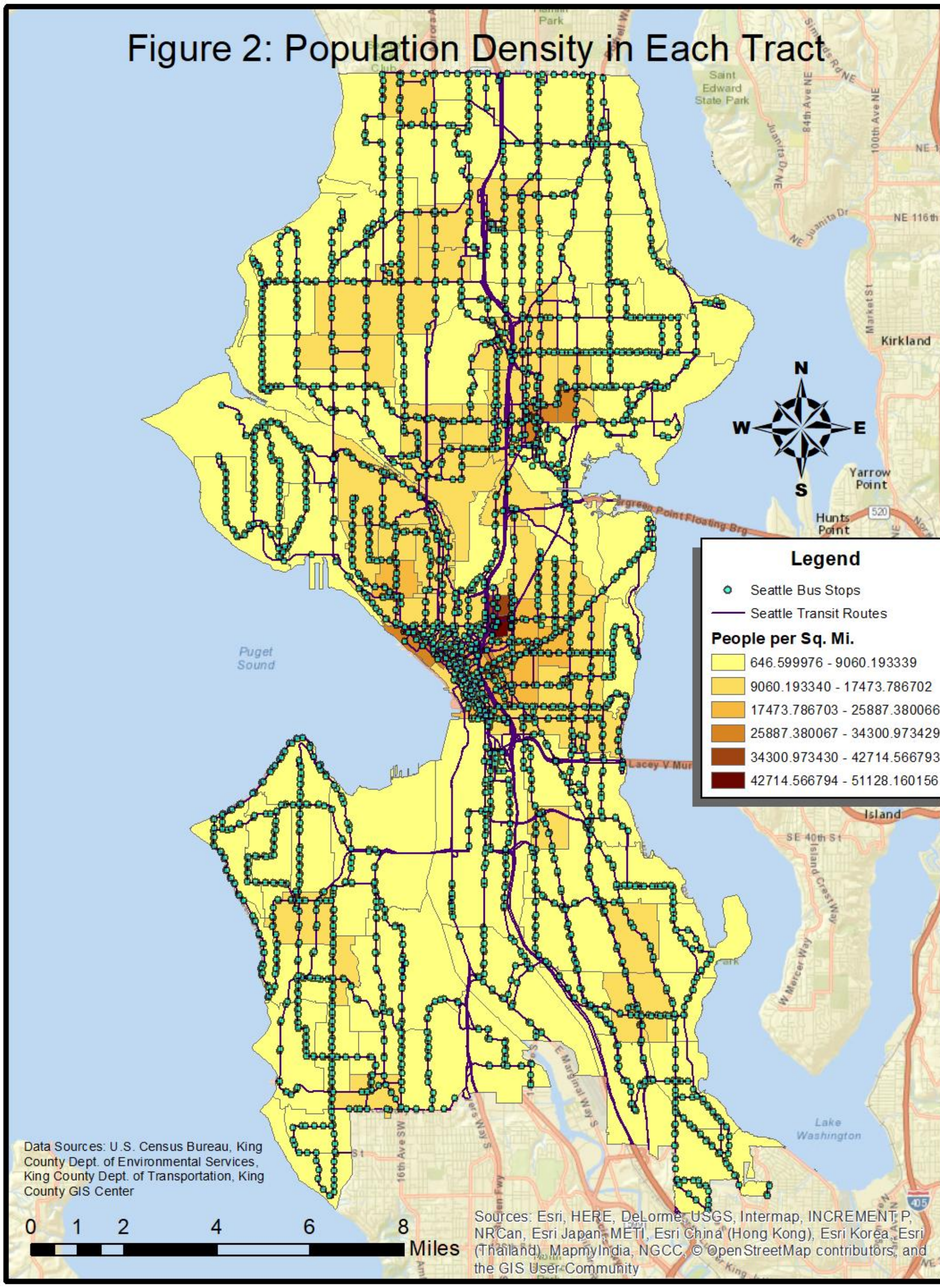


Figure 2: Population Density in Each Tract

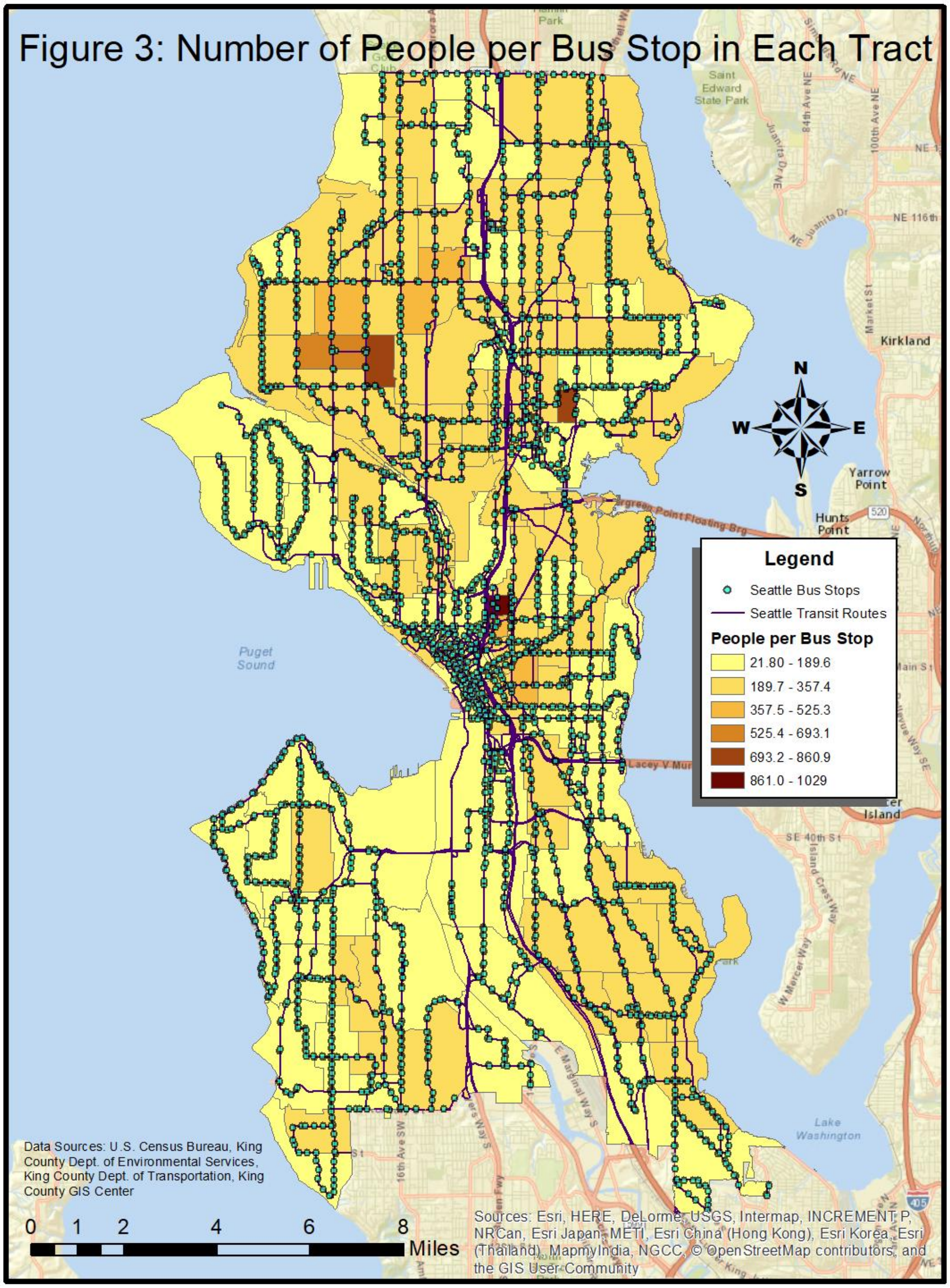
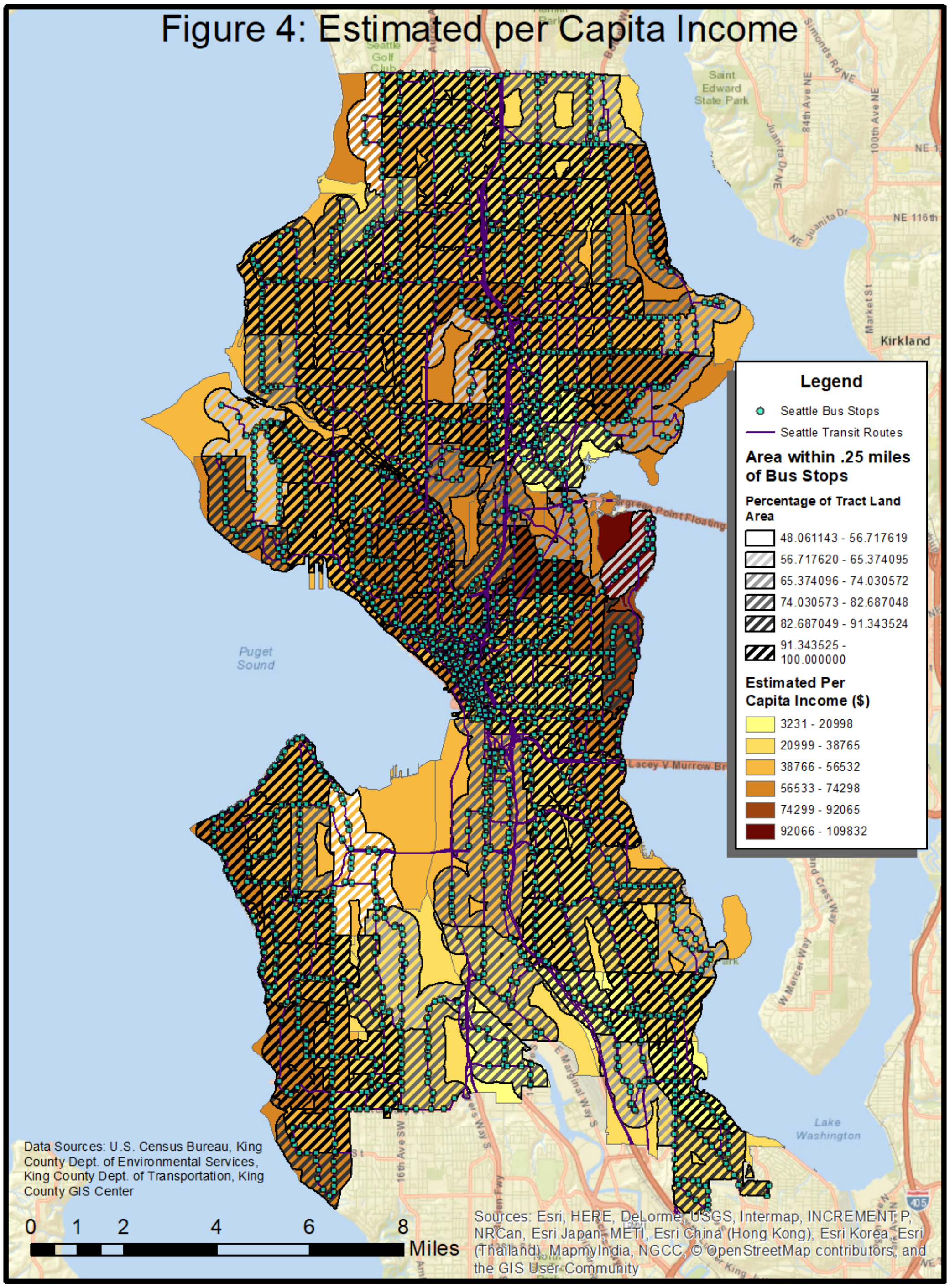


Figure 3: Number of People per Bus Stop in Each Tract
Legend
Seattle Bus Stops
Seattle Transit Routes
People per Bus Stop
21.80 - 189.6
189.7 - 357.4
357.5 - 525.3
525.4 - 693.1
693.2 - 860.9
861.0 - 1029
N
W
E
S
Puget Sound
Lake Washington
Kirkland
Yarrow Point
Hunts Point
Saint Edward State Park
Data Sources: U.S. Census Bureau, King County Dept. of Environmental Services, King County Dept. of Transportation, King County GIS Center
0 1 2 4 6 8
Miles
Sources: Esri, HERE, DeLorme, USGS, Intermap, INCREMENT P, NRCan, Esri Japan, METI, Esri China (Hong Kong), Esri Korea, Esri (Thailand), MapmyIndia, NGCC, © OpenStreetMap contributors, and the GIS User Community

Figure 4: Estimated per Capita Income



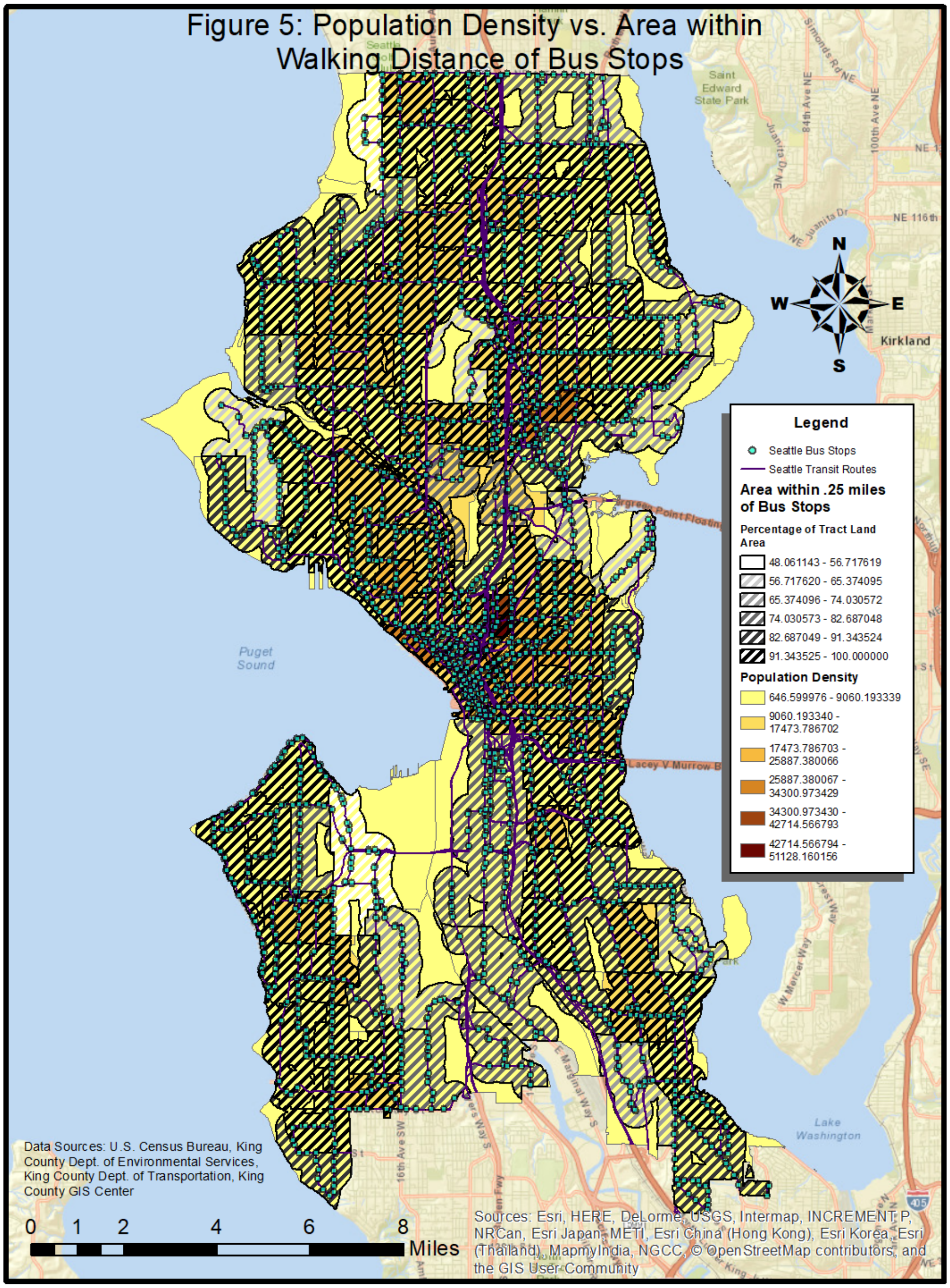


Figure 5: Population Density vs. Area within Walking Distance of Bus Stops