Matthew Timmermann

Title: Income, Unemployment, and Population Around Metra Train Lines in Illinois

Abstract:

The goal in this research project was to test the difference in median income, population density, and unemployment rates between populations that lived on or near Metra train lines, and populations that did not live on or near Metra train lines. I broke down the Illinois population into census tracts, and only focused on tracts near the Chicagoland area, which I considered the grand scope of how far the tracks spread out from Chicago. I created three choropleth maps depicting each of my three factors with the Metra train lines overlaying these color maps. I was able to split the tracts into the rail line/non rail line groups and then analyzed the averages of my three factors based on this split. My original hypothesis of believing median income, population density, and unemployment would all be higher in rail line tracts was proven wrong after the analysis was done. Out of the three, only unemployment showed to be higher in the rail line tracts, while median income and population density both were higher in non-rail tracts.

Introduction:

Public transportation, and more specifically train lines, have been a very hot topic as of late. Both regarding convenience for tax payers, reduction of costs and waste for the city, overpopulation of vehicles on the roads, and large emissions of greenhouse gasses from vehicles. There are many theoretical benefits to a switch to all public transportation, and I think train lines are the most important part of that. I noticed that a lot of studies looked at costs to overhaul a larger public transportation system, current costs of operating these systems compared to individuals driving, and even carbon emissions of public transportation versus individuals driving. I didn’t notice very many studies regarding the actual people who lived around public transportations systems, compared to those who did not have such a system around them. That is why for my project, my aim is to compare just that. I am going to focus on just one part of the public transportation system, and that is the Metra Rail line that goes throughout the Chicagoland area. I would like to look at three factors: Population Density, Median Income, and Unemployment Rates in the Chicagoland Area. I will be looking at two very specific groups to contrast them: People who live within a Metra Rail Line, and People who do not live within a Metra Rail Line. My hypothesis is that all three of these factors will be higher for the group of people who live within the rail line, rather than those who do not.

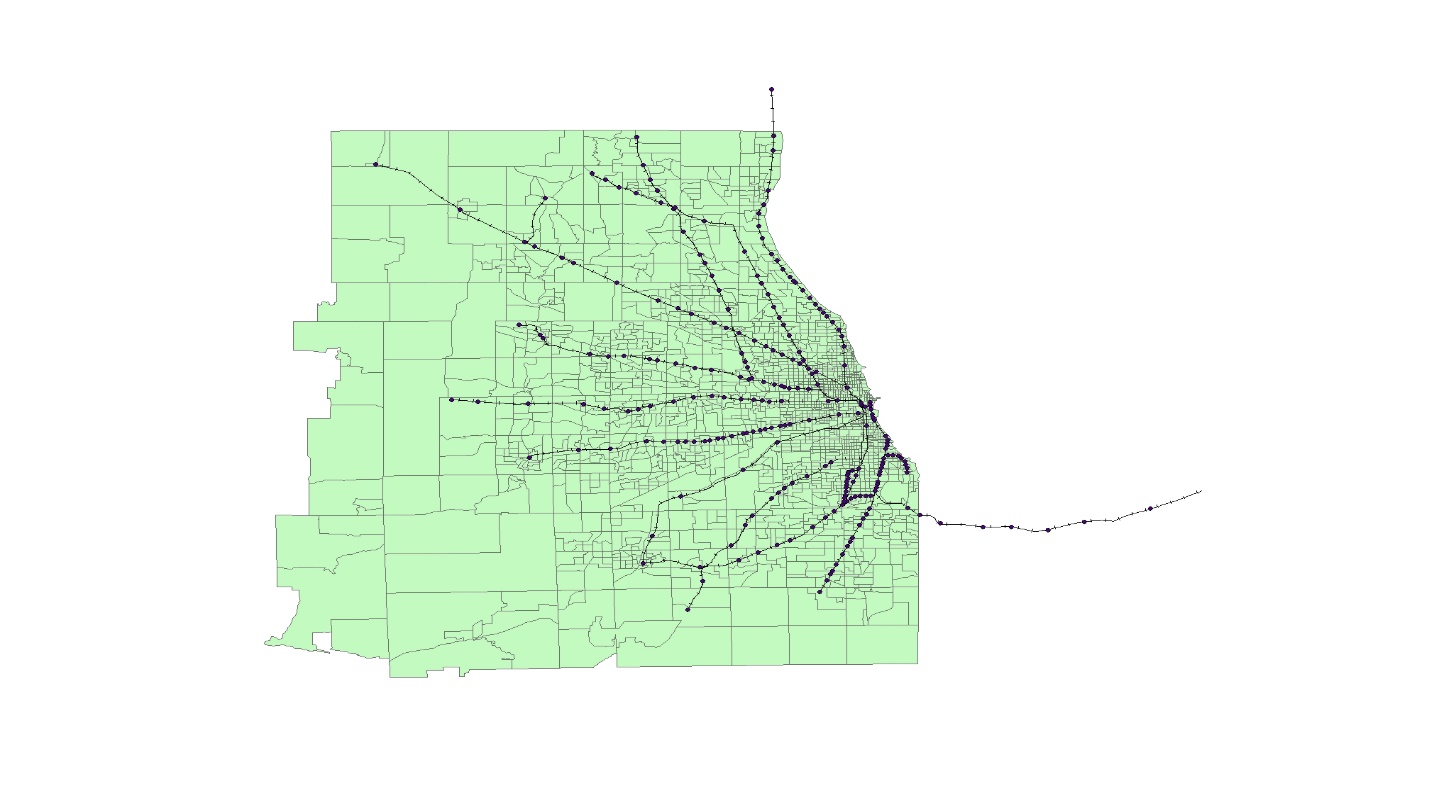
Data and methods:

To start out, I initially thought that I would focus my study around the different counties of the Chicagoland area, and compare the counties that a rail line runs through versus the counties that do not. After using the shapefile of counties that was provided to us in a prior lab, I very quickly realized that essentially every county, since they are so large, had at least one Metra rail line run through it. This would not work if I was trying to find a relatively even sample of counties that were in and out of the rail line. I could have taken the entirety of Illinois, which heavily favors toward counties not being in the vicinity of a rail line, however I wanted to focus mainly just in the surrounding area of the Metra lines. I then decided to do the break down by Census tracts instead of counties. I found the shapefile for these counties from the United States Census Bureau. I now had a shapefile, with each census tract feature having a specific tract code associated with it which was called AFFGEOID. I also needed my shapefiles for the train stops and rail line, which are easily accessible from the City of Chicago website. My next step was to find the median income, population, and unemployment rates for each of the census tracts I was looking at. I found information for all three of these in two different reports from American FactFinder. I decided to focus my data on 18-64-year-old people living in the tracts as those are the primary ages of working people. My study is looking at income and unemployment, so it was a good fit to look at this slice of the population.

All of my original data from these two reports is in the Data Tables folder, in addition to two edits of that data in excel, to condense and simplify the data for import into ArcMap. There were about 25 columns of data that I did not need, so I found it easier to unclutter that in excel, and then import only what was needed into ArcMap. Those final two data files, median income.csv and unemployment and population.csv are in the main folder for my project.

The next part of the process was to join the two data tables just imported, with my shapefile of the census tracts. This was extremely simple as my data files both came with that census tract ID number that I pointed out earlier. So, both of them joined based on that ID number. The last part of table data manipulation I had to do was create a normalized population, or population density number. My census tract shapefile data table came with a field called ALAND, which is the total area of the tract in square meters. So, my population density ended up being persons per square meter (since my calculation for this was Total Population / ALAND). I know I also could have normalized the data when making a graduated color map, however I wanted to have the population density statistic later for export. My tract shapefile did already come with a population density field called Population, however like state earlier, I wanted to use just the population of 18-64-year-old people that I acquired from the unemployment data set. For my median income data, I used nonhousehold income as I thought that would flow best with my data focused on 18-64 working class.

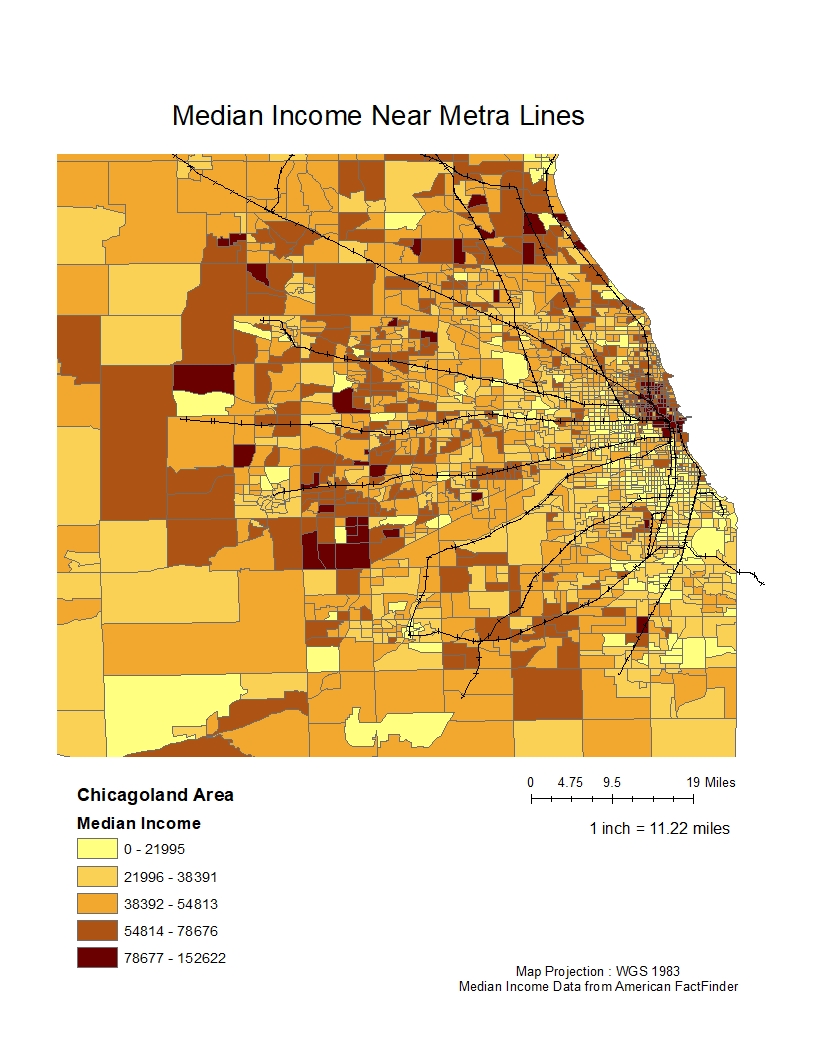
Now that I had all my tract data joined with my external data, and all calculations done, I needed to remove a good chunk of the census tract features from my map. This shapefile came with every census tract for Illinois, and like stated earlier, I wanted to only look at the Chicagoland area, or the area around where the Metra Lines run. This next part was completely subjective on my part as I did not find any surefire way to break down these tracts for my goal. I simply selected, by eye, the group of tracts that I thought would best represent my data, and this picture shows a good idea of the total tracts I used for this report.

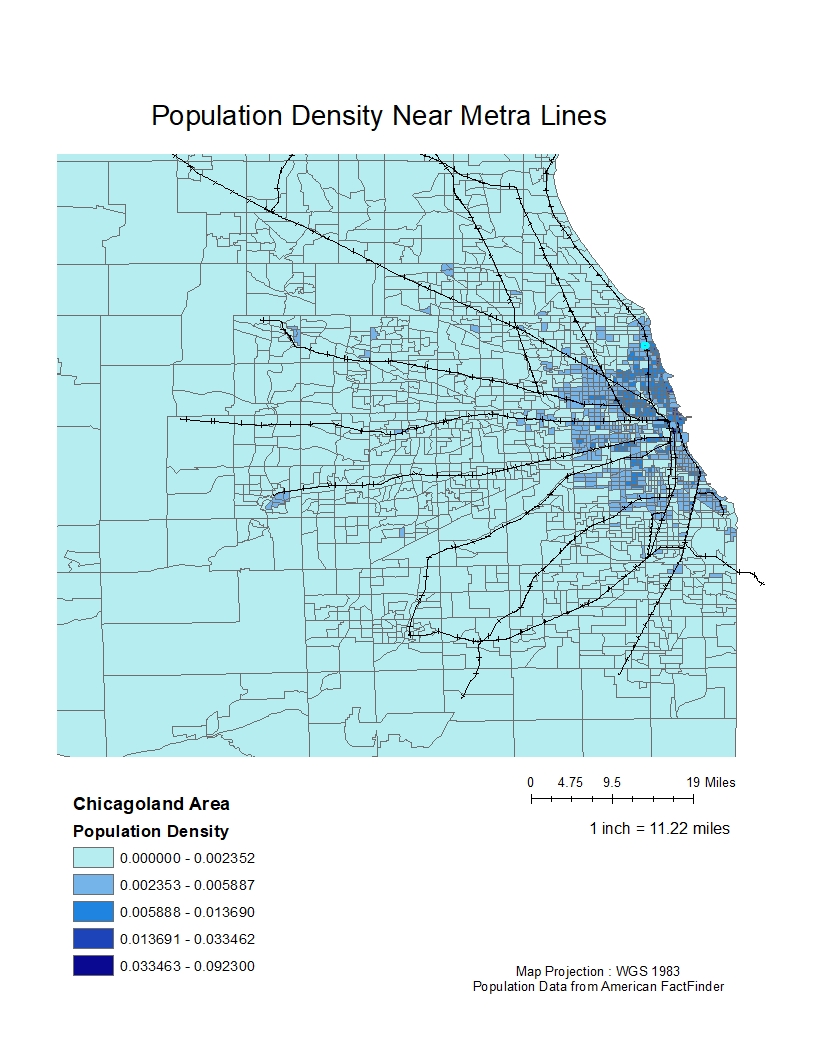


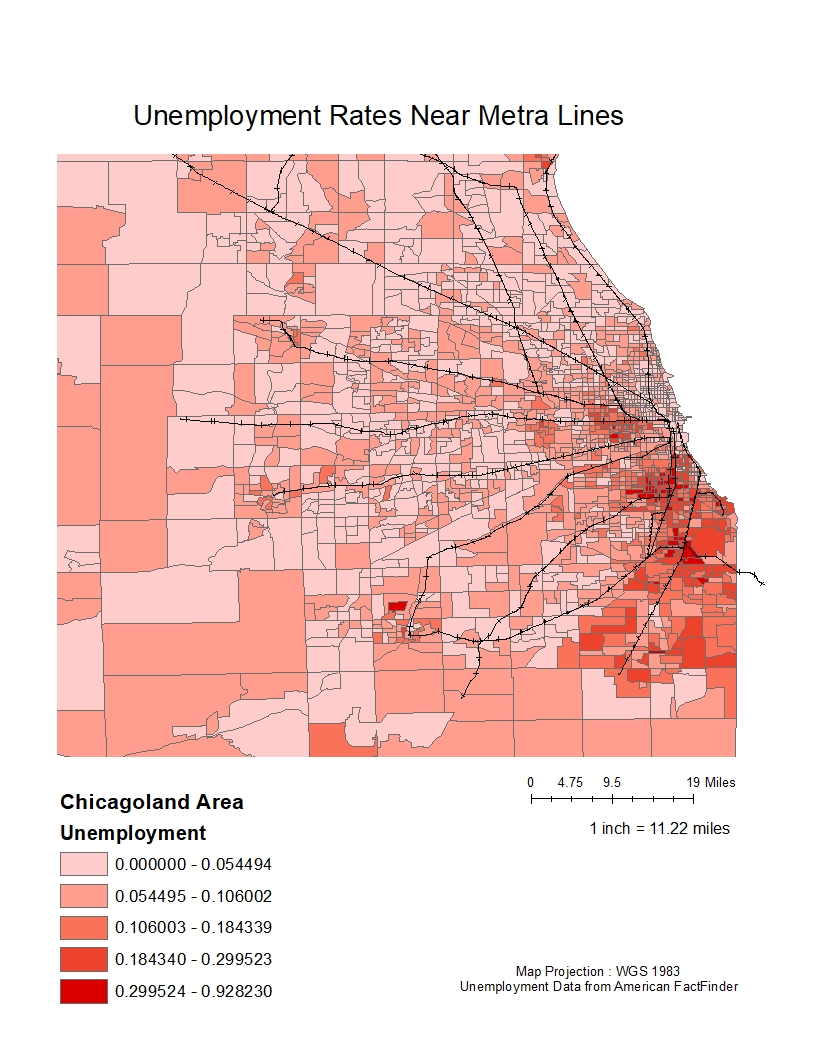
Now with this chunk of tracts of the Chicagoland area, I could easily make 3 choropleth maps describing my three main factors using graduated colors. Since I already had all the data linked to the tracts, I was able to make the three maps instantly. The last thing I needed to do, was to separate my tract data into two categories, tracts that have a rail line going through them (or extremely close to, such as a tract having a rail line be one of the borders of that tract) and tracts that do not have a rail line going through them. I was able to separate these by creating two new shapefiles. I did a select by location and selected all tracts that intersected with the rail lines and had a buffer of 0.001 to ensure that I got tracts that were extremely close to a rail line, but not quite intersecting it. I then was able to get the opposite of that data by simply hitting the switch selection key from the layer data. I exported both of these selections to new shapefiles labeled In\_Rail\_Line and Not\_In\_Rail\_Line. Now that ArcMap successfully split my data for me, all I had to do was export this data to two excel sheets, and I was able to sort the data into the top 10 tracts for each of my 3 factors, and find the averages of income, population density, and unemployment for each of the two data sets.

Results:

From my data, I was able to create three choropleth maps based on the distribution of my three factors over the Chicagoland Area census tracts. They are as follows:







Looking at these three maps, there is definitely a trend along the track lines that is generally a darker color (higher number) than the areas slightly further out. The main city of Chicago, which all the Metra lines converge to, consistently has the darkest shade for all 3. The biggest thing I notice is that, despite not having a train line, it seems like the lowest unemployment rates fall into the category of not being near a line. The next thing that steps out is the median income stat. It does follow that median income seems to be pretty good around the train lines, but there are a lot of certain outlier tracts, particularly in the western suburbs that have an extremely high median income, despite not being by the tracks. The population density to no surprise is the darkest in Chicago, however the other few dark spots on the map do seem to fall along the track lines.

Next, I was able to produce 6 tables, showing the top 10 values of unemployment, median income, and population density for inside the rail lines, and not inside the rail lines. In addition I have produced 3 bar charts showing the comparison of the in rail versus not in rail of the averages of unemployment, median income, and population density. They are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| Top 10 Median Incomes for Tracts With Rail Lines | | Top 10 Median Incomes for Tracts Without Rail Lines | |
| Number | Median Income | Number | Median Income |
| 1 | $119,500.00 | 1 | $152,622.00 |
| 2 | $116,914.00 | 2 | $139,500.00 |
| 3 | $114,625.00 | 3 | $132,875.00 |
| 4 | $110,458.00 | 4 | $126,250.00 |
| 5 | $104,542.00 | 5 | $123,587.00 |
| 6 | $103,665.00 | 6 | $123,571.00 |
| 7 | $103,125.00 | 7 | $113,125.00 |
| 8 | $100,045.00 | 8 | $112,277.00 |
| 9 | $96,875.00 | 9 | $110,764.00 |
| 10 | $96,250.00 | 10 | $106,953.00 |

|  |  |  |  |
| --- | --- | --- | --- |
| Top 10 Unemployment Rates for Tracts With Rail Lines | | Top 10 Unemployment Rates for Tracts Without Rail Lines | |
| Number | Unemployment Rate | Number | Unemployment Rate |
| 1 | 92.82% | 1 | 49.34% |
| 2 | 42.81% | 2 | 48.51% |
| 3 | 41.17% | 3 | 47.40% |
| 4 | 38.62% | 4 | 45.43% |
| 5 | 38.61% | 5 | 42.91% |
| 6 | 38.55% | 6 | 42.75% |
| 7 | 38.40% | 7 | 40.90% |
| 8 | 37.14% | 8 | 40.38% |
| 9 | 37.02% | 9 | 38.25% |
| 10 | 36.91% | 10 | 37.93% |

|  |  |  |  |
| --- | --- | --- | --- |
| Top 10 Population Densities for Tracts With Rail Lines | | Top 10 Population Densities for Tracts Without Rail Lines | |
| Number | Population Density | Number | Population Density |
| 1 | 0.0146 | 1 | 0.0923 |
| 2 | 0.0124 | 2 | 0.0335 |
| 3 | 0.0112 | 3 | 0.0286 |
| 4 | 0.0107 | 4 | 0.0267 |
| 5 | 0.0107 | 5 | 0.0258 |
| 6 | 0.0105 | 6 | 0.0230 |
| 7 | 0.0101 | 7 | 0.0222 |
| 8 | 0.0100 | 8 | 0.0213 |
| 9 | 0.0090 | 9 | 0.0213 |
| 10 | 0.0087 | 10 | 0.0211 |

These results show that the average population density and the average median income are higher for in rail tracts, and the average unemployment rate is higher for in rail. You can see all of the top ten data for each factor further reflects these averages.

Conclusion:

This study ended up proving the exact opposite of what I thought would potentially happen regarding median income, population density, but held firm in my hypothesis for unemployment rates. I believe that this study was severely skewed from the simple fact that most of the tracts that contain railroad lines were either in the city of Chicago, or very close to the city of Chicago, since all the Metra lines start/end at Union station. The maps I got from my data did show a promising trend of high-income high population density along the lines further away from Chicago, but then again were overshadowed by the outlier tracts that were in the suburbs. I believe that this study could be used and improved upon in the future, looking more closely at multiple forms of public education, and finding a way to separate suburban data from city data, as I think that was the main cause of any error that I got.

References:

American FactFinder - Results. (2019). Retrieved from <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_S1901&prodType=table>

American FactFinder - Results. (2019). Retrieved from <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_17_5YR_B21005&prodType=table>

(2019). Retrieved from <https://data.cityofchicago.org/widgets/q8wx-dznq>

(2019). Retrieved from <https://data.cityofchicago.org/widgets/q8wx-dznq>

Branch, G. (2019). Cartographic Boundary Shapefiles - Census Tracts - Geography - U.S. Census Bureau. Retrieved from <https://www.census.gov/geo/maps-data/data/cbf/cbf_tracts.html>