# Atlanta, GA, USA Transit Expansion

### Determining the best neighborhoods for transit expansion

**By Seth Rikard**

## Introduction:

### Background:

The Atlanta, GA region of the United States has experienced rapid growth over the past 20 years.

In the 2010 United States Census, the city had a population of 423,000 residents. In 2015, the US Census Bureau estimated the population had grown to 463,000. This represents a nearly 9% jump. In 2019, the US Census estimates that the population is nearly 500,000. This not only represents a 15% increase in the population over the past 10 years, it also means an exponentially increasing population rate.

In response, the citizens of the City of Atlanta voted on a passed a resolution to launch a $2.5 billion dollar planned expansion of the inner-city transit system. The primary focus of the money is to build out the light-rail system transit system with tie-ins to the current regional heavy-rail transit system.

### Problem:

Despite the investment, the city's transit expansion cannot touch every city neighborhood or community. As such, how can the city use data to prioritize spending the money? Where would invested dollars generate the greatest system usage and alleviate the most car traffic?

### Importance:

The city's department of transportation would be very interested in knowing how best to allocate the money. Collecting and analyzing data would allow the department to know how best to use these funds by identifying the areas of the city experiencing the greatest population growth.

# Data Analysis:

I used the follow data sources to access information on currently operating light-rail systems in North America, population and job growth trends those regions as well as Atlanta, and finally, population and job growth in the 24 Neighborhood Planning Units that comprise the city.

## Data Sources:

Wikipedia North America ligh-rail transit ridership:
https://en.wikipedia.org/wiki/List_of_North_American_light_rail_systems_by_ridership

*This open-source website allows us to pull information on currently operating light rail systems across North America.*

Us Census Data:
https://data.census.gov/cedsci/all?q=Atlanta,%20GA&hidePreview=false&table=DP05&tid=ACSDP1Y2017.DP05&g=1600000US1304000&vintage=2017&layer=place&cid=DP05_0001E&lastDisplayedRow=15

*Statistics and information provided by the United State's Department of Commerce. This information provides insight into population and job growth trends across the city.*

Atlanta Regional Commission Data: http://opendata.atlantaregional.com/

*Compiles the Census Data and provides information on a micro basis. This open source website allows me to scrub and glean insight into the city's demographic trends.*

City of Atlanta Data By NPU: https://neighborhoodnexus.org/maps-and-data/maps/city-of-atlanta-neighborhood-planning-units-npu/

*The City of Atlanta's website for their Neighborhood Planning Unit. Allows us to scrub NPU specific data.*

Atlanta NPU (Neighborhood Planning Unit) GeoJSON file:
https://raw.githubusercontent.com/jordanstreiff/atlanta-city-council-search/master/public/NPU.geojson

*Publicly available GeoJSON file used to generate the coordinates of the NPU district maps.*

## Data Prep:

The first step in my data prep process was wrangling the data from the various sources listed above. This required scrapping websites, downloading CSV files from government publications, combining tables, compiling results and loading this information back into m Jupyter Notebook.

There were two tables created from the multiple sources:

| | North American Light Rail Systems | Atlanta NPU District |
|---|---|---|
| **Name** | df_rs | df_atl |
| **Columns** | System Name, City, Annual Ridership, Daily Ridership, AVG Daily Boarding's Per Mile, Pop Density, Job Density | NPU District, Total Population, 2000 Population, Pop Change, Population Density, Jobs, Residential Permits, Commercial Building Permits, Total Permits |
| **Rows** | 28 | 24 |
| **Design Intent** | Correlation Modeling | Recommendation Modeling |

The Light Rail System data frame compiled information from the Wikipedia Page and the US Census Bureau. My first step was scraping the information from the Wikipedia page and formatting the values, which required removing annotations, hyperlink formatting, and dual-branded names on transit systems so that the 'System Name' column parsed correctly.

Similarly, the Atlanta NPU District data frame compiled information from the US Census Bureau, and the Atlanta Regional Commission. The downloaded CSV files were much easier to wrangle and format; however, it required the addition of a key-index column for this data to communicate with the GEOJSON file referenced above.

# Methodology:

### 1.1 Defining Success:

After wrangling and cleaning the data, I was prepared to conduct the first step in my methodology: develop a profile of the most successful light- rail systems in North America.

I decided to define success by those systems that have the highest average daily boarding's per mile.

I made this decision because I felt it was more reflective of system usage than Annual Ridership. A larger light-rail network might have a higher overall annual ridership, but that success could be accredited the size of the system, and not necessarily the usage.

### 1.2 Defining Outliers:

Secondly, I decided to remove the outliers from the data set. Three of the top four systems with the highest boarding's per passenger mile were located in the New York City area. New York City and the surrounding communities are the most densely populated in North America. For example, that region has a job density (jobs per square mile) over 100,000. The second closest city is San Francisco and Chicago with less than 45,000.

If I kept the New York Region in the data set, their job and population density would drastically screw the correlation and averages because it is one of the most densely populated regions in the world.

### 1.3 Exploring the Data:

### Light Rail Systems

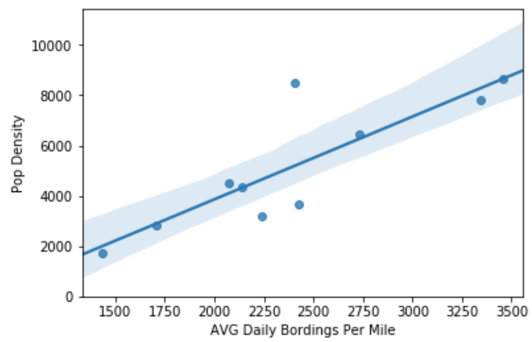
After cleaning and removing the outliers, I sorted the data set by the highest average daily boarding's per mile. This resulted in a new data set populated by the 10 most successful light-rail systems in North America.

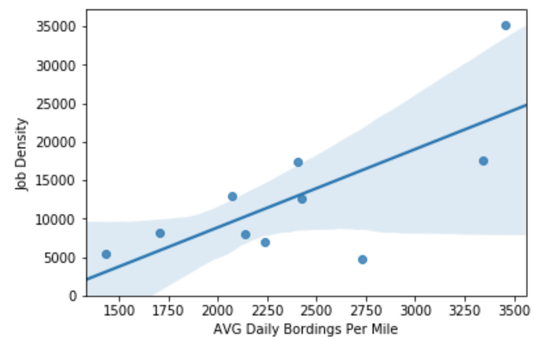
| | System Name | City | Annual Ridership | Daily Ridership | AVG Daily Bordings Per Mile | Pop Density | Job Density |
|---|---|---|---|---|---|---|---|
| 2 | Link | Seattle | 24155900 | 74400 | 3456 | 8642 | 35112 |
| 3 | METRO Light Rail | Minneapolis | 23811200 | 71900 | 3344 | 7820 | 17631 |
| 6 | Buffalo Metro Rail | Buffalo | 4560600 | 16900 | 2734 | 6436 | 4710 |
| 7 | METRORail | Houston | 18808000 | 61100 | 2427 | 3660 | 12556 |
| 8 | Metro Rail | Los Angeles | 67921600 | 219900 | 2403 | 8483 | 17386 |
| 9 | Valley Metro Rail | Phoenix | 16269000 | 48900 | 2240 | 3207 | 6984 |
| 10 | San Diego Trolley | San Diego | 37215800 | 112100 | 2140 | 4325 | 7937 |
| 11 | MAX Light Rail | Portland | 39173700 | 119700 | 2070 | 4504 | 12978 |
| 12 | Lynx | Charlotte | 5228500 | 16900 | 1708 | 2827 | 8092 |
| 13 | TRAX | Salt Lake City | 18823500 | 63000 | 1435 | 1709 | 5389 |

Using this new data set, I can a correlation function to determine which values correlated most strongly with AVG Daily Boarding's Per Mile.

| | Annual Ridership | Daily Ridership | AVG Daily Bordings Per Mile | Pop Density | Job Density |
|---|---|---|---|---|---|
| **Annual Ridership** | 1.000000 | 0.998189 | 0.039953 | 0.438033 | 0.308191 |
| **Daily Ridership** | 0.998189 | 1.000000 | 0.025802 | 0.436752 | 0.297633 |
| **AVG Daily Bordings Per Mile** | 0.039953 | 0.025802 | 1.000000 | 0.846095 | 0.721692 |
| **Pop Density** | 0.438033 | 0.436752 | 0.846095 | 1.000000 | 0.721716 |
| **Job Density** | 0.308191 | 0.297633 | 0.721692 | 0.721716 | 1.000000 |

It appeared that Population Density and Job Density had the strongest correlation to my definition of success. I ran two plotting functions to visualize the correlation.




I ran the Peason Correlation Coefficient function on these two values to verify (and crosscheck myself), and confirm the best metric. The Population Density was the best metric to use as a means to determine where the City of Atlanta should invest the transit money.

Job Density returned a value of 0.0184585 and Population Density returned a value of 0.0020297

Since the Peason Correlation Coefficient recommends selecting the correlation value closest to .001, I was able to confirm that Population Density was the better metric to use for recommending districts for transit expansion.

Finally, I performed a mean average on Population Density for the top-ten performing light rail systems. It was determined that the top performing systems operating in neighborhoods with an average of 5161 people per square mile.

**Atlanta NPU (Neighborhood Planning Districts)**

I performed a similar sort function on the Atlanta NPU data set. This allowed me to determine which districts were the most populated. I then created a new data set from the top-ten most densely populated NPU Districts.

| | NPU | District | TOTAL POPULATION | 2000 Pop | Pop Change | Sq Miles | POP SQ MILE | JOBS | Residential Permits | Commercial Building Permits | Single-Family | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **12** | M | NPU M | 30647 | 21343 | 43.6 | 3.78 | 8108 | 131381 | 46 | 6 | 19 | 71 |
| **4** | E | NPU E | 45375 | 34502 | 31.5 | 5.91 | 7678 | 99242 | 12 | 7 | 22 | 41 |
| **19** | T | NPU T | 17260 | 20095 | -14.1 | 2.74 | 6299 | 4965 | 2 | 2 | 0 | 4 |
| **21** | W | NPU W | 21502 | 20197 | 7.0 | 3.52 | 6108 | 6790 | 3 | 2 | 52 | 57 |
| **13** | N | NPU N | 19039 | 14730 | 29.3 | 3.44 | 5535 | 6426 | 41 | 3 | 6 | 50 |
| **11** | L | NPU L | 6970 | 7314 | -4.7 | 1.32 | 5280 | 1883 | 0 | 0 | 7 | 7 |
| **20** | V | NPU V | 15542 | 15840 | -1.9 | 3.17 | 4903 | 3790 | 1 | 0 | 38 | 39 |
| **1** | B | NPU B | 48709 | 38562 | 26.3 | 10.18 | 4785 | 91833 | 44 | 4 | 72 | 120 |
| **5** | F | NPU F | 25661 | 21378 | 20.0 | 6.00 | 4277 | 11200 | 21 | 3 | 55 | 79 |
| **10** | K | NPU K | 9973 | 11923 | -16.4 | 2.39 | 4173 | 2478 | 0 | 0 | 7 | 7 |

These districts most closely meet the mean average population density of the most successful light rail systems. As such, the characteristics of these districts are such that a future transit line in these districts would maximize the number of boarding's per transit system mile.

In addition to pulling this information, I felt it would be beneficial to analysis other population growth trends. I selected two other data points for consideration. First, I chose Population Growth. I selected this data point because it would allow planners to know which districts are experiencing the most growth in population.

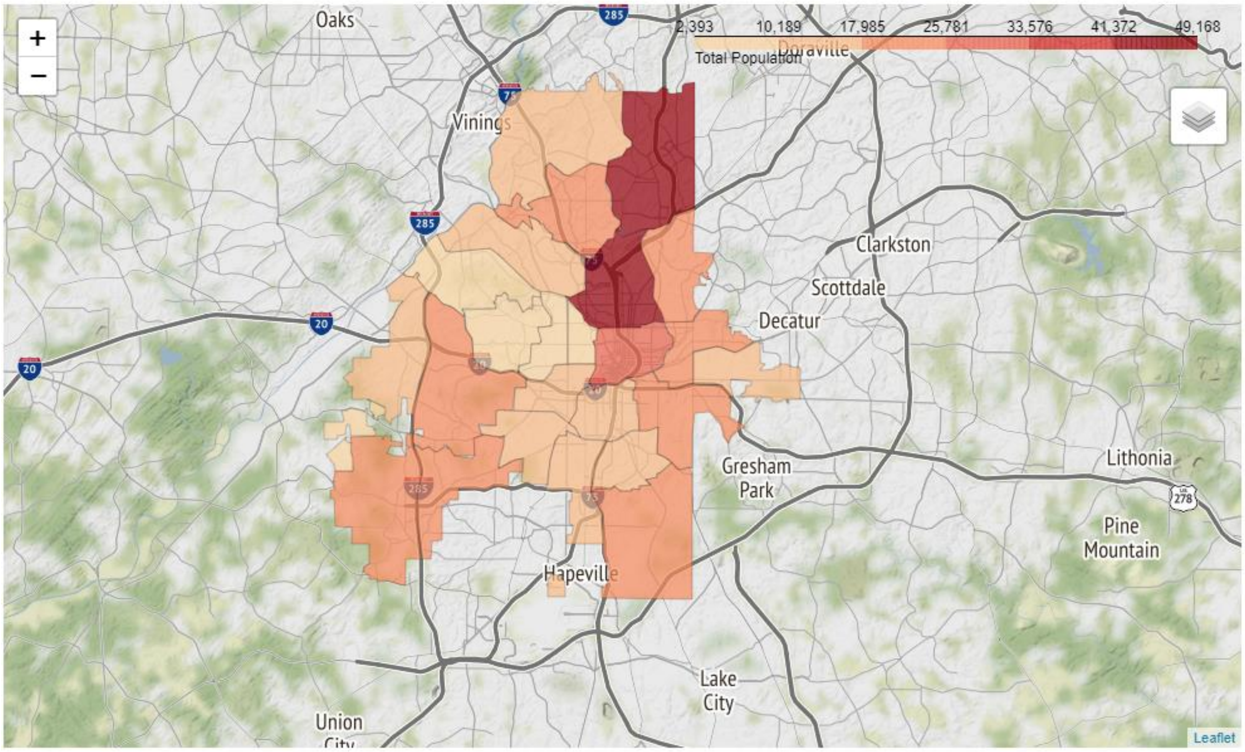
Planners would want this information to prevent shortsightedness in planning. A district may not be the most densely populated currently, but it could be the fastest growing and therefore could become one of the most densely populated in a matter of years.

Secondly, I selected the Residential Permits. Residential Permits are authorization from the city that allow a builder to construct new residential high-rise towers, condominiums, or apartment buildings. Planners would want to know where new residents are moving and therefore where population density will be increasing.

## Data Visualization:

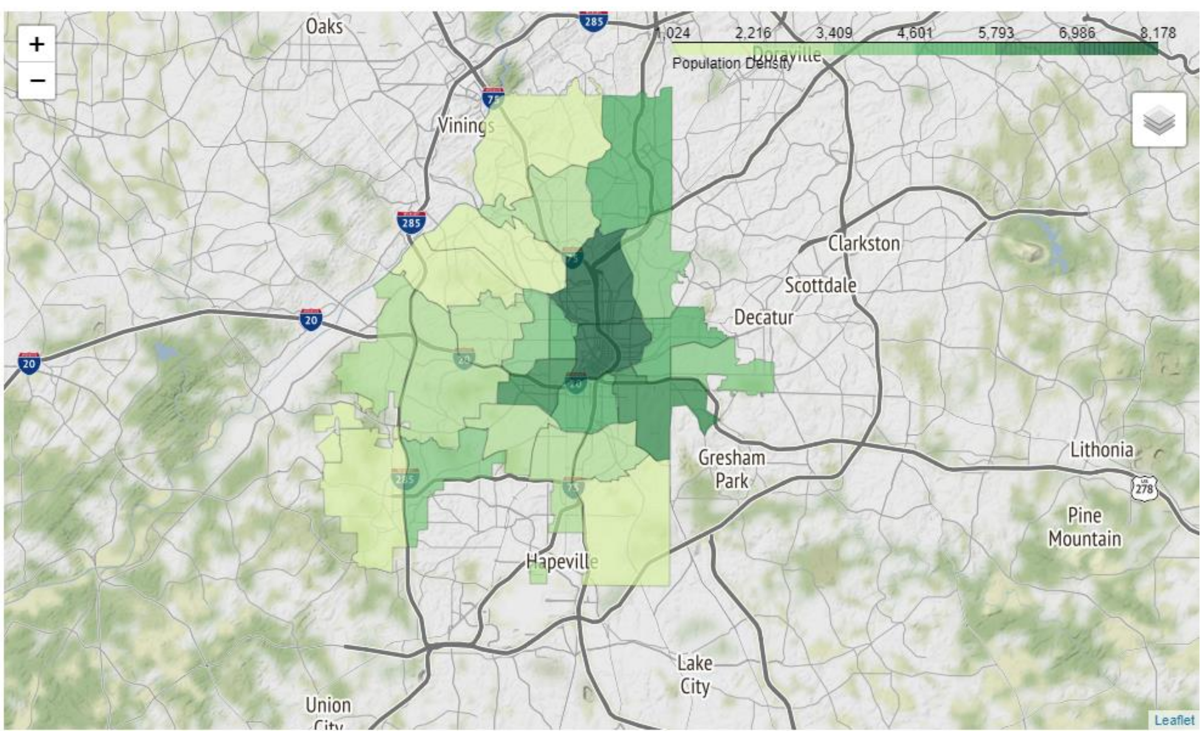
I began the data visualization by incorporating all data from the Atlanta NPU data set. I did this to determine which NPUS were the most populated, which were the most densely populated, which have experienced the greatest population change, and finally, which had the highest issuance of building permits.

2.1: Map of Atlanta NPU Total Population



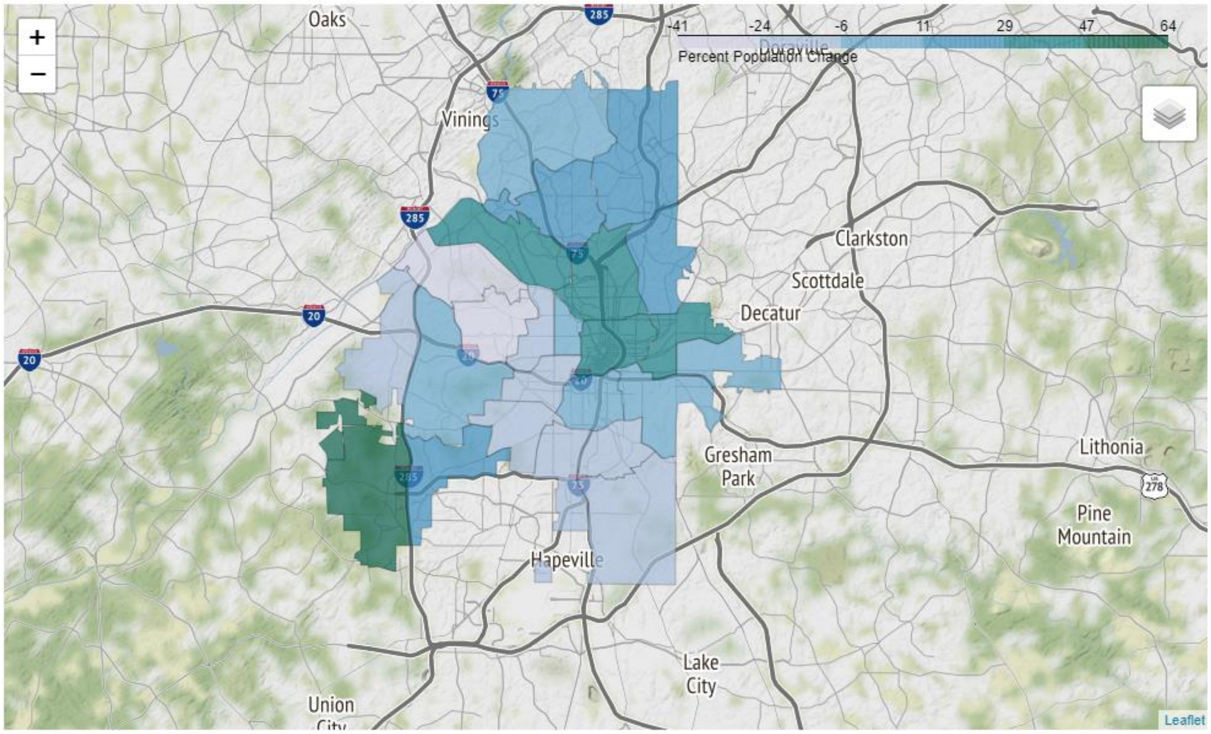

NPU B, E, and M are the most highly populated Neighborhood Planning Districts.

## 2.2 Map of Atlanta NPU's Population Density



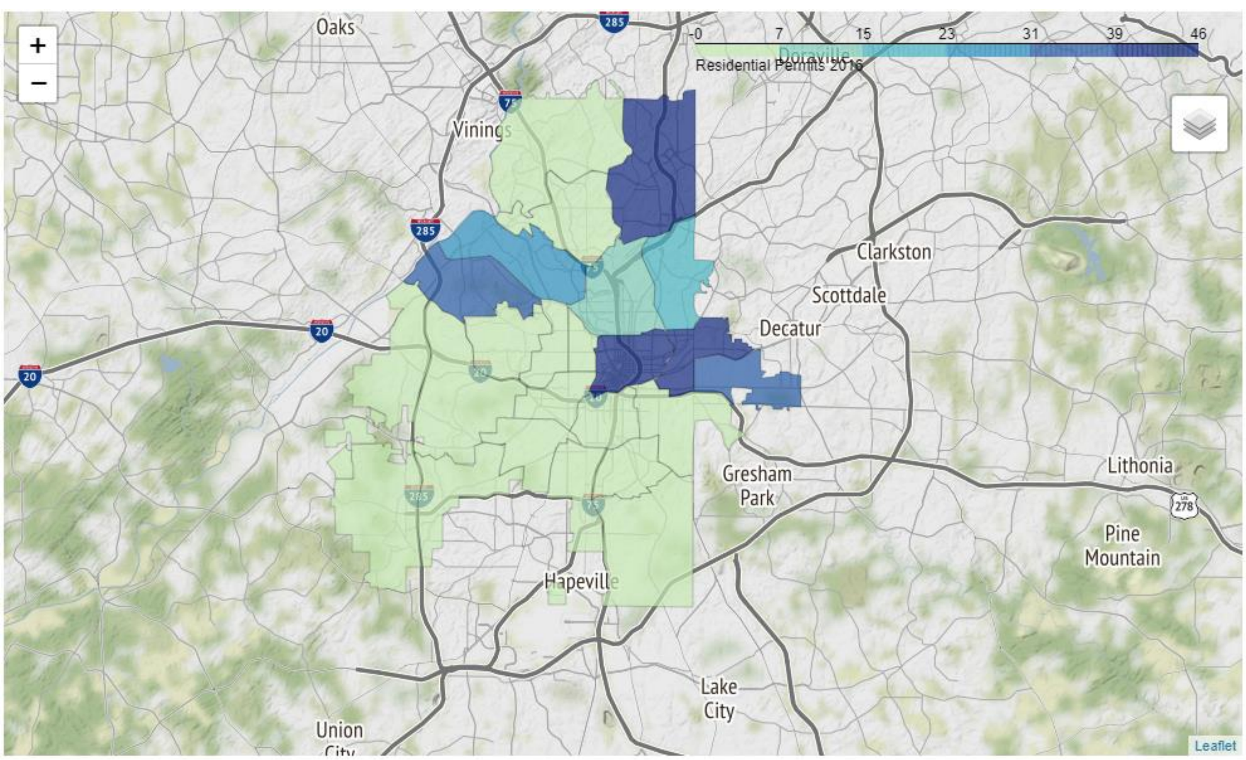

The greatest population density is located to the north of the city's Central Business District. This included NPU B, E, M, N, and W.

## 2.3 Map of Atlanta NPU Population Change



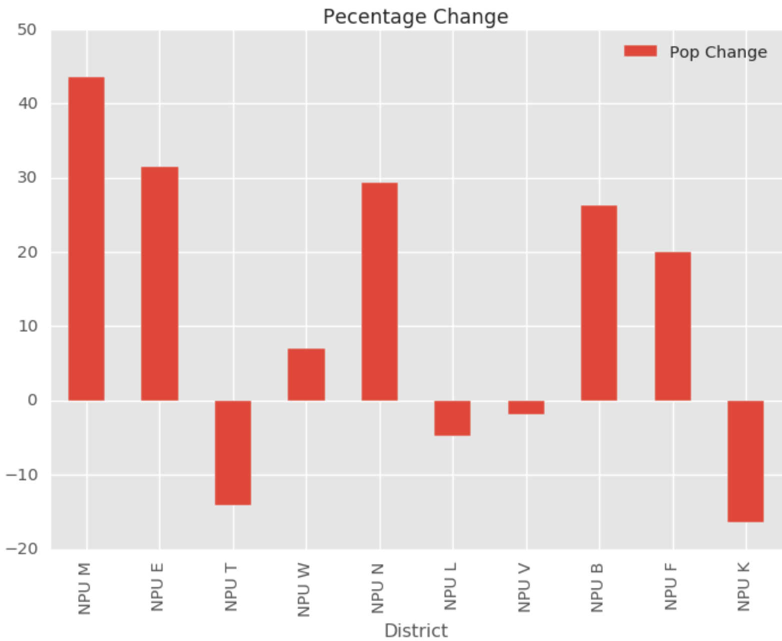

The greatest increase in population is occurring in NPU B, D, E, M, and P

## 2.4 Map of Residential Building Permits



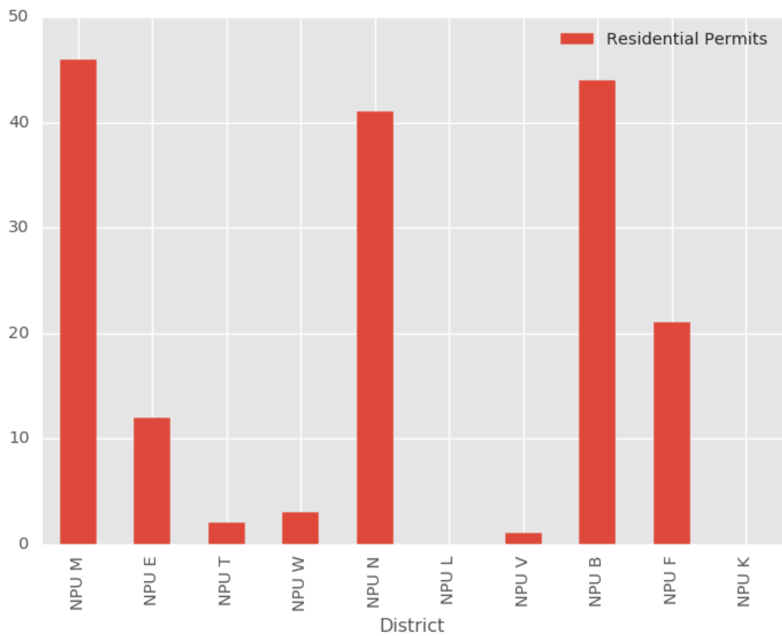
The highest number of permits issued in 2016: NPU B, G, D, E, M, N, O, and F.

After visualizing what the data was communicating, I decided to focus in on the current top-ten most densely populated NPU Districts. I felt this would allow planners to not only see what trends had been developing over time, but what current and trending information was available on the most densely populated neighborhoods.




This analysis added value because it allows planners to see holistically as well as within the subsection of those districts currently meeting the mean average of the most successful light rail systems.

The two bar charts of Top Ten Most Densely Populated NPU's outline the population change 2010 to 2016 and the number of building permits issued in 2016.




From these two charts, I inferred that the densely populated districts needed to be broken into two different and further subsections: growing and not growing.

This allows yet another layer of analysis for the city transportation planners.

**Extrapolation:**

There were districts that meet the threshold of minimal population density and were growing, those that met the minimal population density but were not growing, those that did not currently meet the minimal population density but were growing, and those that did not currently meet the minimal population density and were not growing.

I broke all three districts down into those four categories.

| | Populated | Not Populated |
|---|---|---|
| **Growing** | **NPU M, E, W, B, F** | **NPU G, D, N, O** |
| **Not Growing** | **NPU T, L, V, K** | **NPU A, C, H, I, J, P, R, S, X, Y, Z** |

## Recommendation:

Transportation planners should focus the money toward system expansion in the districts that exhibit both or one of the characteristics. Doing so will ensure the system maximizes efficiency by being placed in communities with the necessary density to generate the greatest ridership per system expansion mile.  This includes NPU B, E, M, F, and W.

As a secondary focus, the department should invest money in lines that transect the districts that meet one of the two characteristics. This ensures the expansion covers neighborhoods which do are trending towards meeting the mean average population density metric. This includes NPU G, D, N, and O.