Discover St. Louis (Crime/Income)  
Project 1 – REbecca Kalhorn

August 20, 2018

# Overview

## Project Background and Description

**Group Members** – Rori Cooper, Rebecca Kalhorn, and Raj Sikka

**Background** – Our group initially thought about probing into ATM and banking data; however, after having difficulties finding public banking data we decided to try a different approach. It turns out Rebecca and Rori both reside in St. Louis City (Raj lives in St. Charles) and the group collectively came up with the idea to go on a hunt for public data to explore what is unique about St. Louis neighborhoods. We eventually found out that identifying datasets was trickier than we thought, but that’s what made our project adventurous and fun. The sole purpose was a fact-finding mission using descriptive statics.

**Goals** - Our project is to describe each zip code in Saint Louis and provide an overview of their economies and demographics. We will compile and analyze multiple data sources and let the data show what is unique about each zip code.

**Questions** –

•What is unique about each zip code located within the Saint Louis city limits?

•How do the zip codes (neighborhoods) compare by a number of demographic variables using descriptive statistics?

•If you were moving to Saint Louis and were new to the area, which is the most attractive zip code to live and why?

**Description** - Use public data to show the distinctive neighborhood characteristics of St. Louis City.

## Project Scope

We delegated data-gathering duties amongst each teammate respectively:

Rori – Education

Rebecca – Crime and Income

Raj – Business

## High-Level Requirements

|  |  |
| --- | --- |
|  |  |

## Crime/Income Data Sources and Method

|  |  |
| --- | --- |
| St. Louis – gov has a lot of information including detailed monthly crime reports. The data is given with an address associated with each crime, but without a zip code. The neighborhood code is given for each crime listed. Neighborhood maps and zip code maps could be found and cross referenced to then approximate the zip code where the crime occurred. To make the data even better, the address could be used in conjunction with the google api to get the exact zip code for each crime. Since counts can be misleading on their own, population was also explored by zip code as well as income. Crime relative to population was investigated. Population for each zip code was found using census data. This data was given via tracts necessitating cross reference with a tract map.  St. Louis – gov, used for the monthly crime report – <https://www.stlouis-mo.gov/data/>  Neighborhood map used - <https://nextstl.com/2011/01/groth-guides-to-the-neighborhoods-of-the-city-of-st-louis/>  Tract map and census information - Home Mortgage Disclosure Act from [www.stlouis-mo.gov](http://www.stlouis-mo.gov)  List of crime codes - https://www.courts.mo.gov/page.jsp?id=270  Extracted crime data from 2014 to present as well as population data, all from csv files. Concatenated the crime data and merged this with the population data. See Jupyter Notebook for more detail. For income, the median income for each tract was given. So, it was decided that the average of the median income of tracts in each zip code would be looked at for the income in each tract. A tract was only put in one zip code even though many of them span more than one. A tract was assigned to the zip code where the majority of the tract’s land mass was. |  |
| |  |  |  | | --- | --- | --- | | Census Tracts | Neighborhoods | Zip Codes | |  |

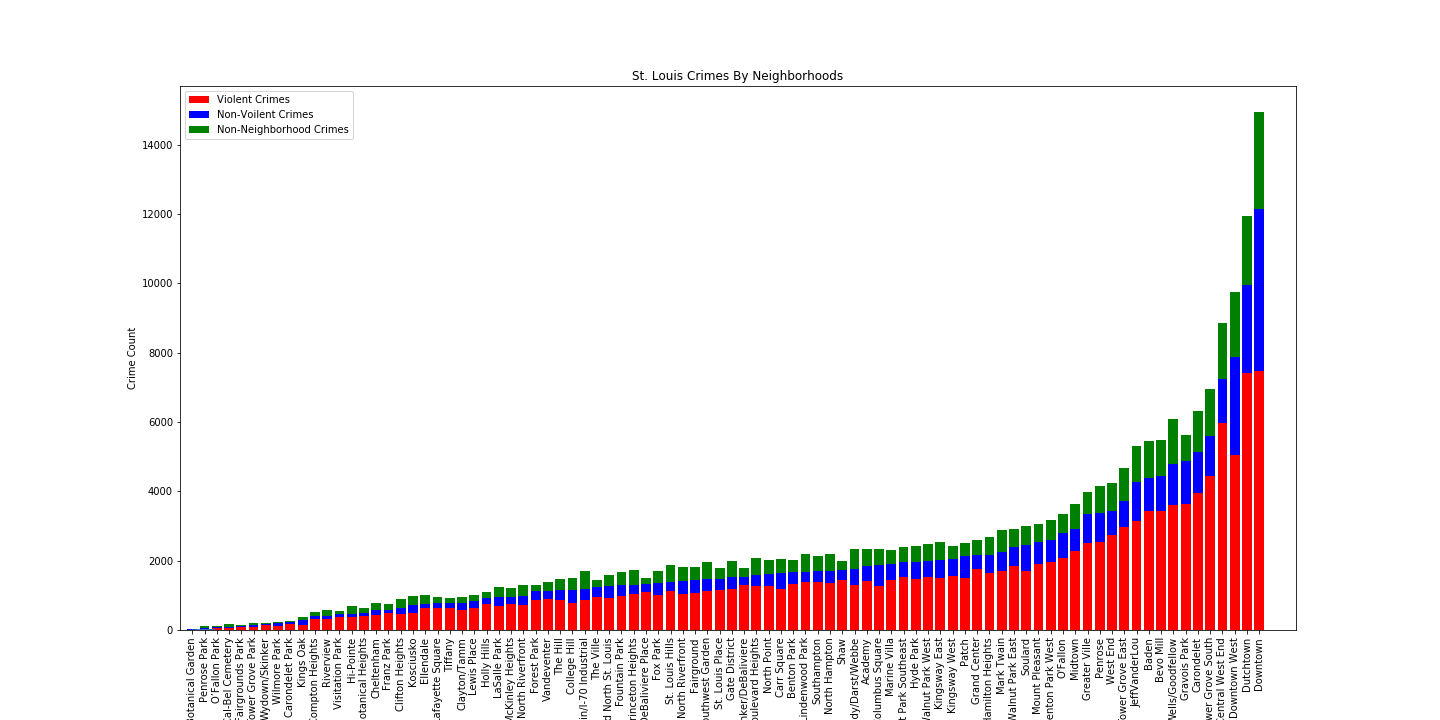
## Analysis/Results

Saved the resulting dataframes as well as the output bar charts:

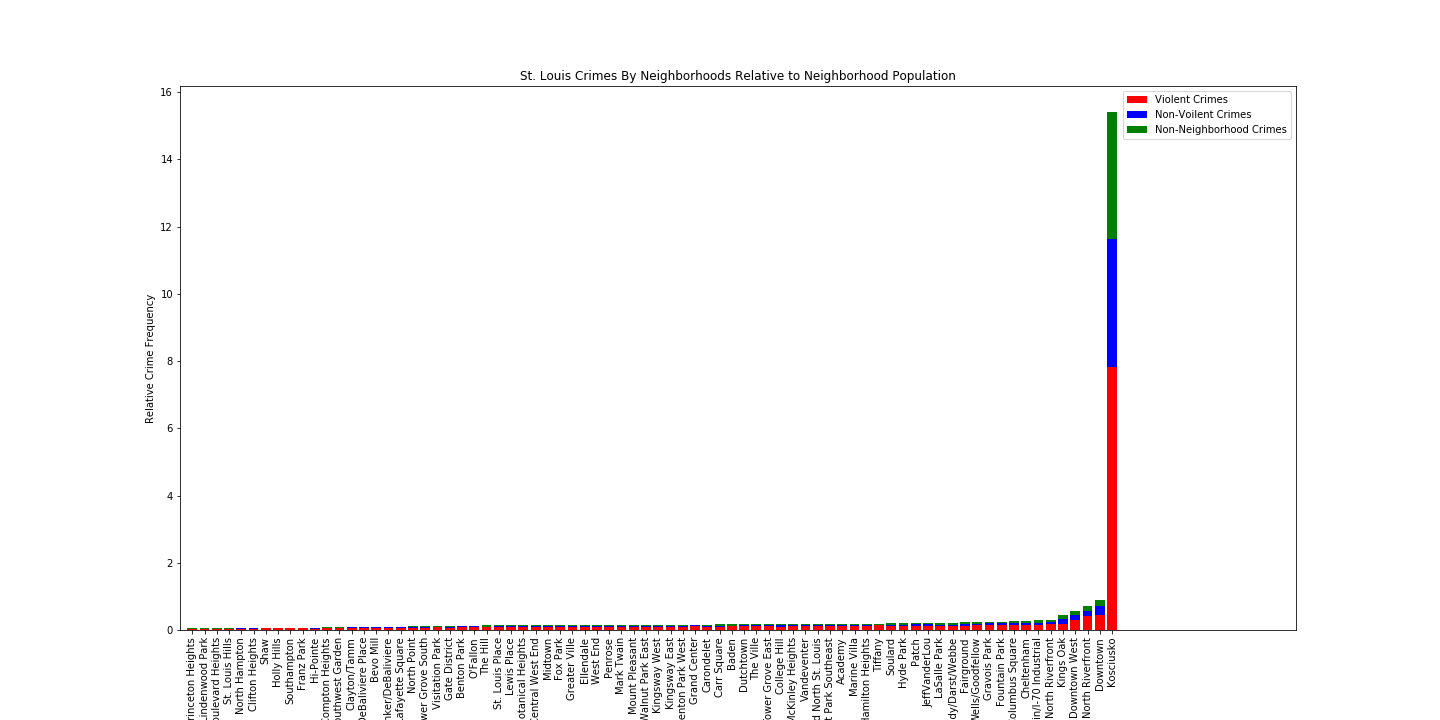
stl\_crime\_data\_2014\_to\_June\_2018 – combined data frame

Stacked Bar Charts:

Crime count by neighborhood: crimeCount.png

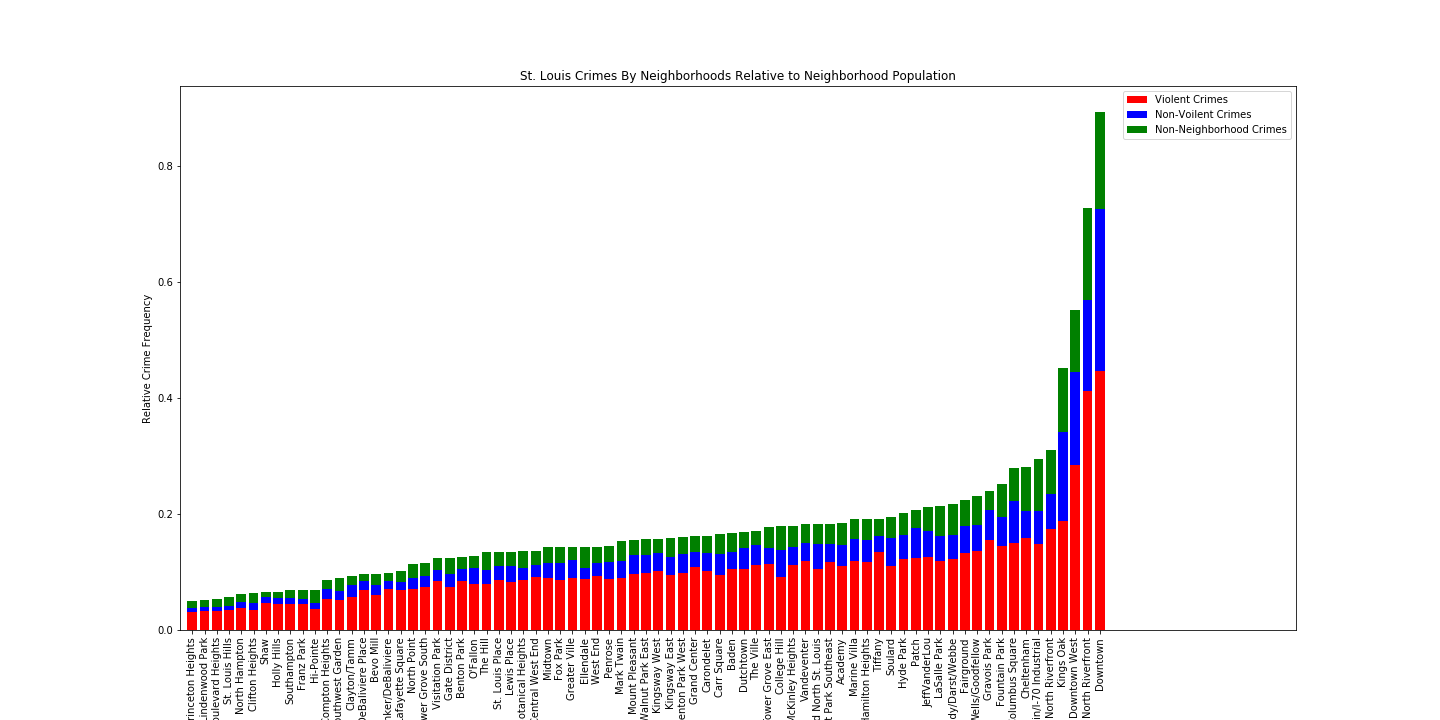


Crime relative to population by neighborhood: relCrime.png

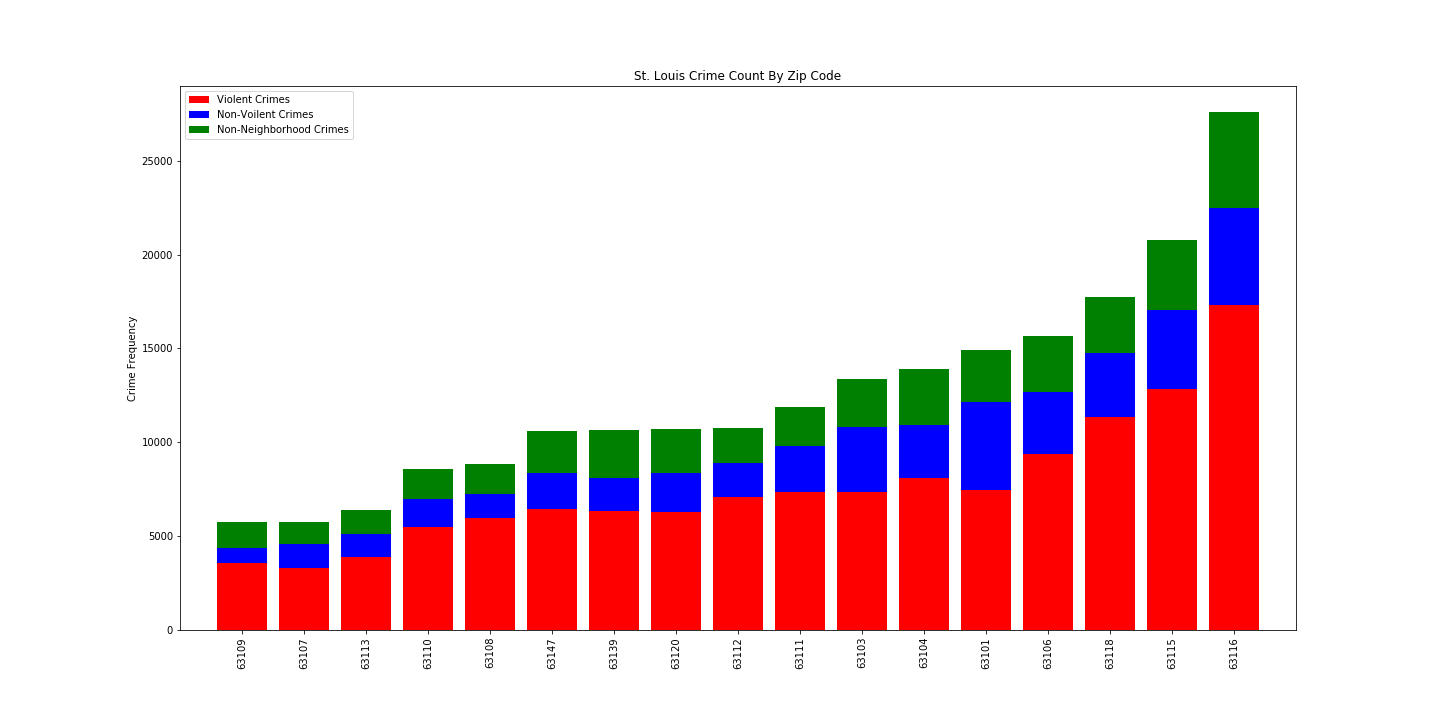


The 4 nieghborhoods with the most crime relative to population is somewhat expected since they are all in the downtown area. This area has more venues for events as well as businesses and government buildings which translates into a lot of people going in and out of the area on a daily basis. Therefore, it is expected that they would have more crime compared to the number of people who live there.

Crime relative to population by neighborhood w/o relative outlier neighborhood Kosciusko: relCrimewoKosciusko,png

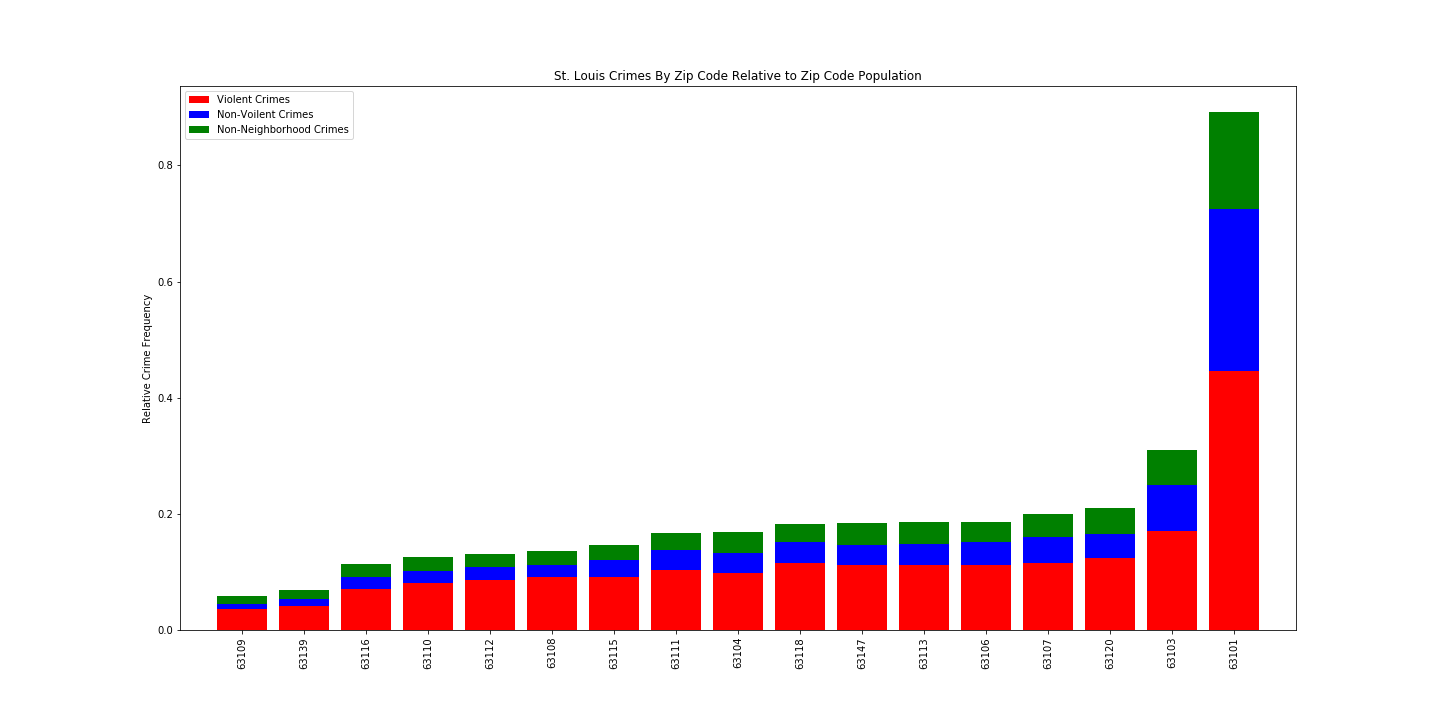


Crime count by neighborhood: zipCrimeCount.png



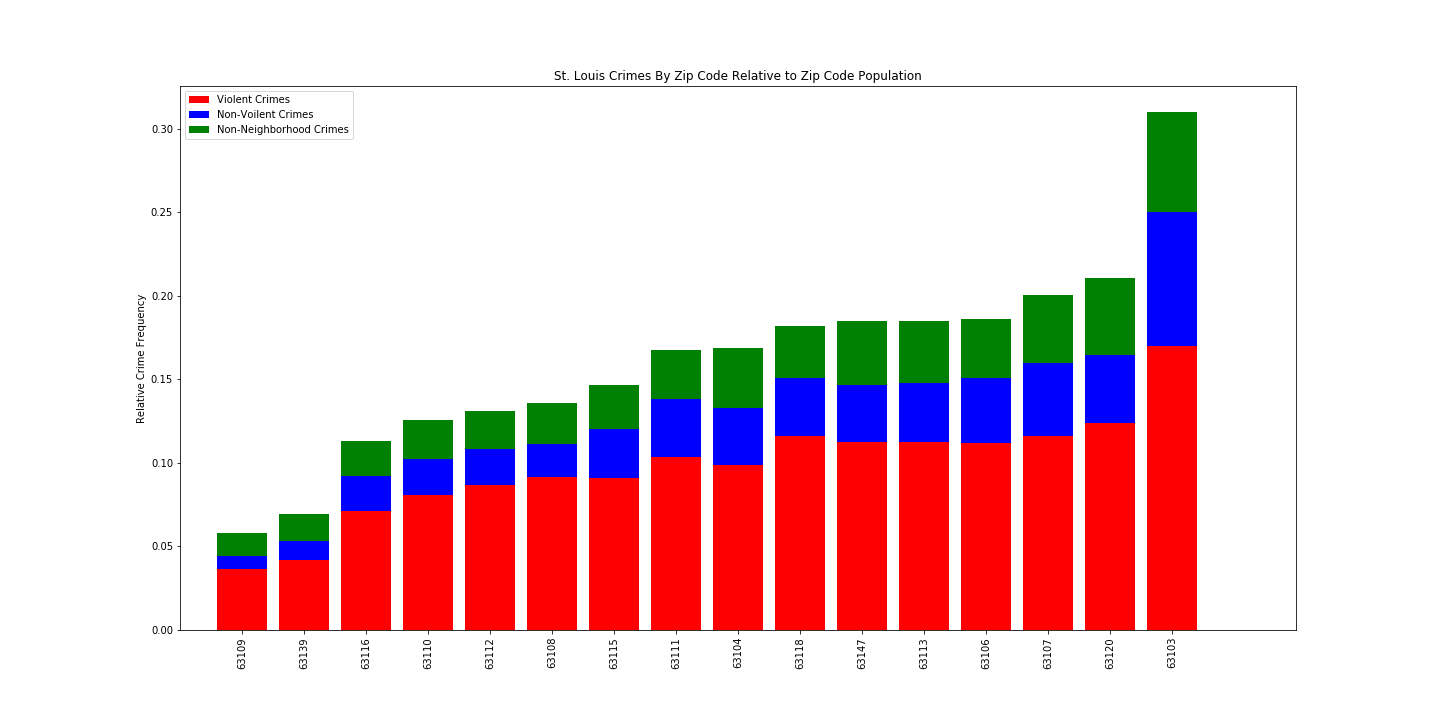
63116 also has the most population. Therefore, it may be expected for it to have the higest crime count.

Crime relative to population by zip: relZipCrime.png

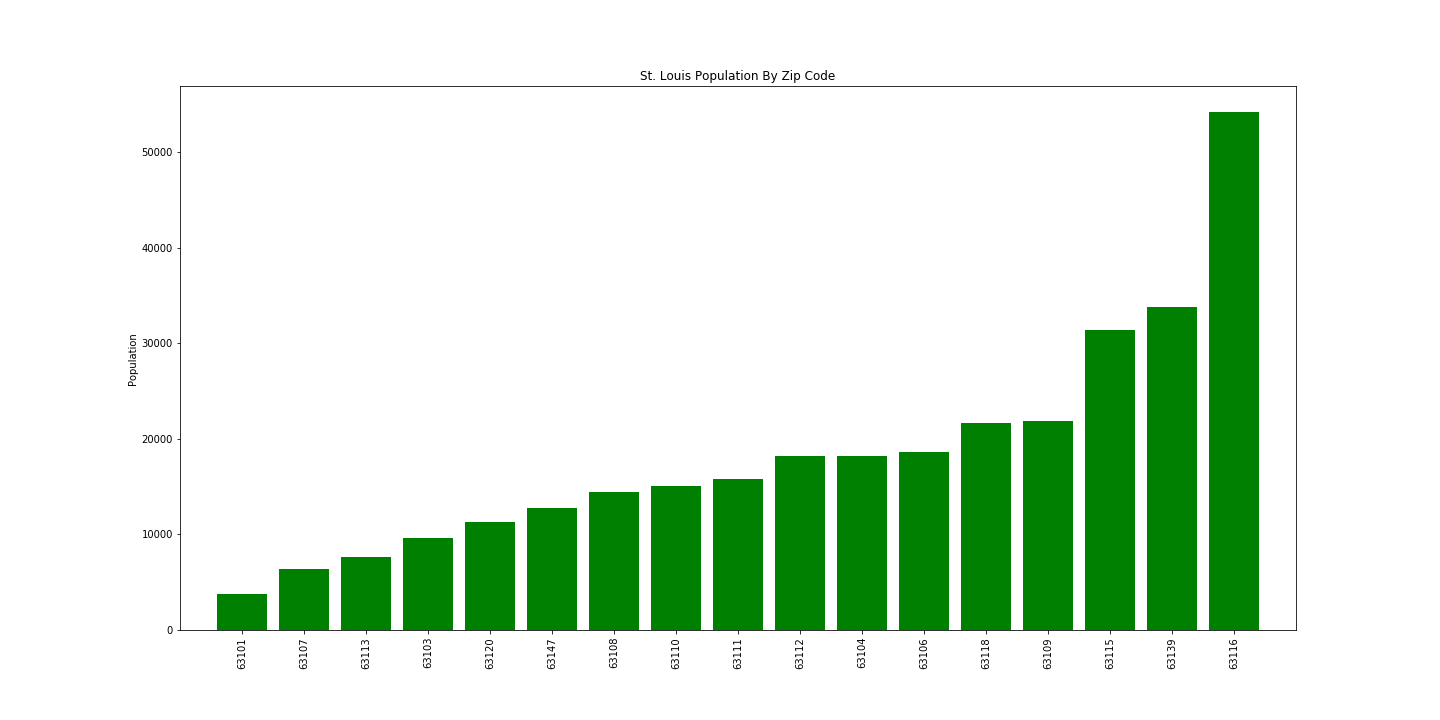


63101 contains downtown. Therefore, it is expected that it will have more crime relative to the population of the same zip for the same reasons metioned in relative neighborhood crime. 63103 is adjacent to downtown, and therefore probably has higher relative crime for the same reason. The rest of the zip codes, until you get to the lowest 2, have fairly similar relative crime rates. Additional analysis could be done to see if the difference in these middle 13 zip codes is statistically significantly.

Crime relative to population by zip without relative outier zip of 63101: relZipCrimewo63101.png



Population by zip: pop.png



Opened CSV in excel to take advantage of Excel’s 3D Map feature. This was helpful in getting maps that showed the crime by zip code in a way that made the location stand out. This also made side by side comparisons of population, income and crime by zip easier to visualize.

|  |  |  |
| --- | --- | --- |
| Population: | Income: | Crime: |

## Next Steps

As stated before, google maps api could be used to get better location of the crimes by zip code. Crime could also be looked at relative to factors other than population. Ideally this could be broken down to neighborhoods using google maps.