Violent Crime in America

Andrew Wang

10/18/2020

Introduction

In this report, I will be going over some violent crime data recorded in the United States. First, I will go over some background information regarding crime in the United States, such as the types of criminal activity seen in the dataset and a brief overview of crime trends in the United States' history. After this, I will use R's ggplot 2 package to conduct graphical analysis of my crime dataset, and draw some conclusions based on the patterns in the data.

Criminal activity has been throughout the entirety of the United States' history, and the statistics on criminal activity are recorded in the annual Uniform Crime Reports (UCR), created by the Federal Bureau of Investigation (FBI). The UCR splits crime into two categories, Part I offenses and Part II offenses. For the purpose of this report, I am focusing on Part I offenses. Part I offenses are also known as index crimes, and are broken into two groups: violent crimes and property crimes. Assault, rape, murder, and robbery are all violent crimes, while arson (not in our dataset), burglary, larceny, and vehicle theft are property crimes. What follows is a description of each kind of crime we are looking at .

In reading the follow crime explanations, it is important to recognize the width and large variety that can be counted under each of these crimes. Each of these crimes has extreme variation in specific cases, and the explanations are simply overviews of the criminal activity. For example, robbery may range from a gas station robber to a bank robbery, the bank robbery being much more serious than the gas station robbery.

Assault The act of inflicting physical and unwanted harm or contact on another person, or the threat of committing such an action.

Rape Rape is a form of sexual assault usually involving sexual intercourse carried out against someone's consent.

Murder Murder is the illegal killing of another human without justification. Murder is also separated into manslaughter, which is a killing without malice, such as in the case of self-defense. .

Robbery Robbery is the taking of an object of value through force or the threat of force. Robbery is considered an inherently violent crime, and that is how it can be distinguished from other kinds of theft.

Burglary Burglary is the illegal entering into of a building or other area with the intent of theft. Burglary is different from robbery becaues it specifically must occur on someone's property.

Larceny Larceny is a crime based around the illegal theft of personal property. Larceny is a very broad term, and usually considered a less substaintile crime.

Vehicle theft Vehicle theft is very simply the criminal act of stealing a motor vehicle, such as a car or bus.

Criminal patterns in America

Regarding the overall picture, the crime rate has been decreasing with time in the United States. While there was a brief period of crime increase between the 1940s and the 1990s, that has since gone down. For example, total property crime (out of 100,000) has decreased from around 5,000 in 1990 to around 2,500 in 2015. Scientists point to a variety of factors which many have influenced the crime decrease in the United States, such as larger police forces and larger prison populations, but there has yet to be a definitive factor which has decreased the crime rate.

Data setup and cleanup

In this section, I am going to review basic environment clean-up and imported packages, along with cleaning the data and making sure everything is ready for analysis. First, I clean up and set up the R environment with the working directory. Keep in mind that your directory structure is likely different from mine, and adjust accordingly.

```r
# Andrew Wang
# October 18
# 2016 Violent Crime in America script
#
# clean up and setup
rm(list=ls()) # clean up any old stuff in R
setwd("C:/Users/hyper/OneDrive/Desktop/Desktop Folders/Programming/R/Assignments/Week 7") # go to this folder
#load up myfunctions.R
source("C:/Users/hyper/OneDrive/Desktop/Desktop Folders/Programming/R/myfunctions.R")
```

After this is done, I import the packages I'm using. Keep in mind that you will likely need to install these packages first before using them.

```r
#library import
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(maps)
```

```
## Warning: package 'maps' was built under R version 4.0.3
```

```
##
## Attaching package: 'maps'
```

```
## The following object is masked from 'package:purrr':
##
##     map
```

```r
library(ggcorrplot)
```

Finally, we can import our .csv file, and assign it to a variable. I would normally run some basic observations on the dataset as a starting point, but in the effort of saving paper I've commented out those lines.

```
crime <- read.csv("violentcrimes.csv")
#str(crime)
```

Furthermore, I need to check if there is any missing data in the dataset, and because the return is 0 I know that my data is filled.

```
print(sum(is.na(crime))) #no missing data
```

```
## [1] 0
```

Looking at a table of the data, I noticed that the property column was just a sum of the burglary, larceny, and auto columns, meaning it would be redundant in my summation graphs. Therefore, I've made the decision to remove it.

```
#remove the "property" column - its just a sum of some other columns
crime$Property <- NULL
```

Because I know that I am going to use ggplot2, I need to convert certain columns of currently character data to numeric. First, I need to remove the commas (,) from the numeric rows. I used the `gsub()` function to replace all the commas in each value with nothing ("").

```
#takes out commas in columns we want to change
crime$Population <- gsub(",","",crime$Population)
crime$Annual <- gsub(",","",crime$Annual)
crime$Murder <- gsub(",","",crime$Murder)
crime$Rape <- gsub(",","",crime$Rape)
crime$Robbery <- gsub(",","",crime$Robbery)
crime$Assault <- gsub(",","",crime$Assault)
crime$Burglary <- gsub(",","",crime$Burglary)
crime$Larceny <- gsub(",","",crime$Larceny)
crime$Auto <- gsub(",","",crime$Auto)
```

With the commas removed from the columns, I can now simply convert the data type of the column from character to string.

```
#changes the columns to numeric
crime$Population <- as.numeric(crime$Population)
crime$Annual <- as.numeric(crime$Annual)
crime$Murder <- as.numeric(crime$Murder)
crime$Rape <- as.numeric(crime$Rape)
crime$Robbery <- as.numeric(crime$Robbery)
crime$Assault <- as.numeric(crime$Assault)
crime$Burglary <- as.numeric(crime$Burglary)
crime$Larceny <- as.numeric(crime$Larceny)
crime$Auto <- as.numeric(crime$Auto)
```

Graphical analysis

Correlogram of crime types

As a baseline comparison, I wanted to get a correlogram of my crime types, so I could see how related they were. I hypothesized that most crimes would be pretty closely correlated and I was proven right. It seems reasonable that criminal activity would be correlated, as an increase in criminal activity would lead to more criminal activity.

```
# correlogram of crimes
crimeSub <- crime[2:11]
crimeCor <- round(cor(crimeSub), 1)
corPlot <- ggcorrplot(crimeCor,
                hc.order = TRUE,
                type = "upper",
                lab = TRUE,
                title="Correlogram of crime types")
corPlot
```
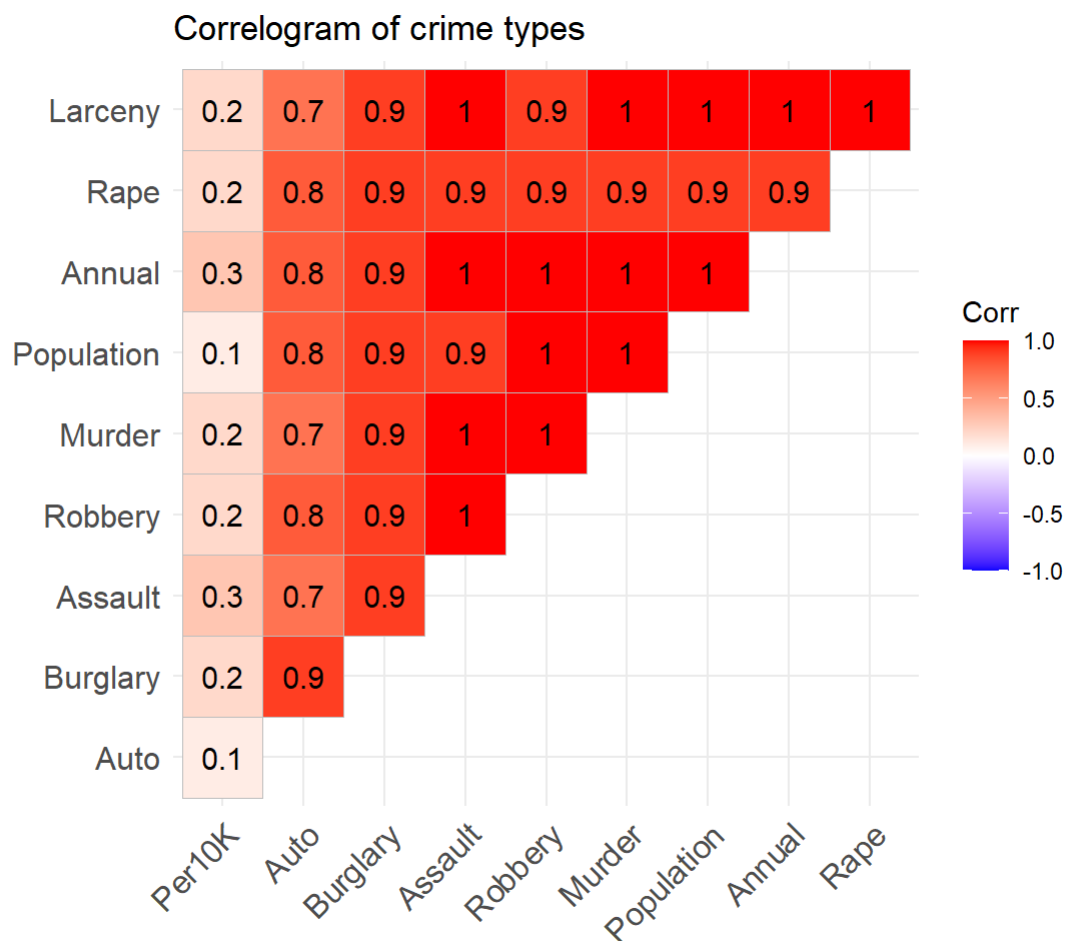

Merging data

Before we can begin plotting maps, we need to merge some data from another dataset so we can have accurate longitude and latitude for our map. I take the state data from the maps package, and merge it with my current crime data to produce a new dataset capable of being turned into a map.

```
#import data
states <- map_data("state")
#new "region" column for the two datasets to be joined with
crime$region <- tolower(crime$State)
#join the state data into the crime data
stateCrime <- left_join(states,crime)
```

```
## Joining, by = "region"
```
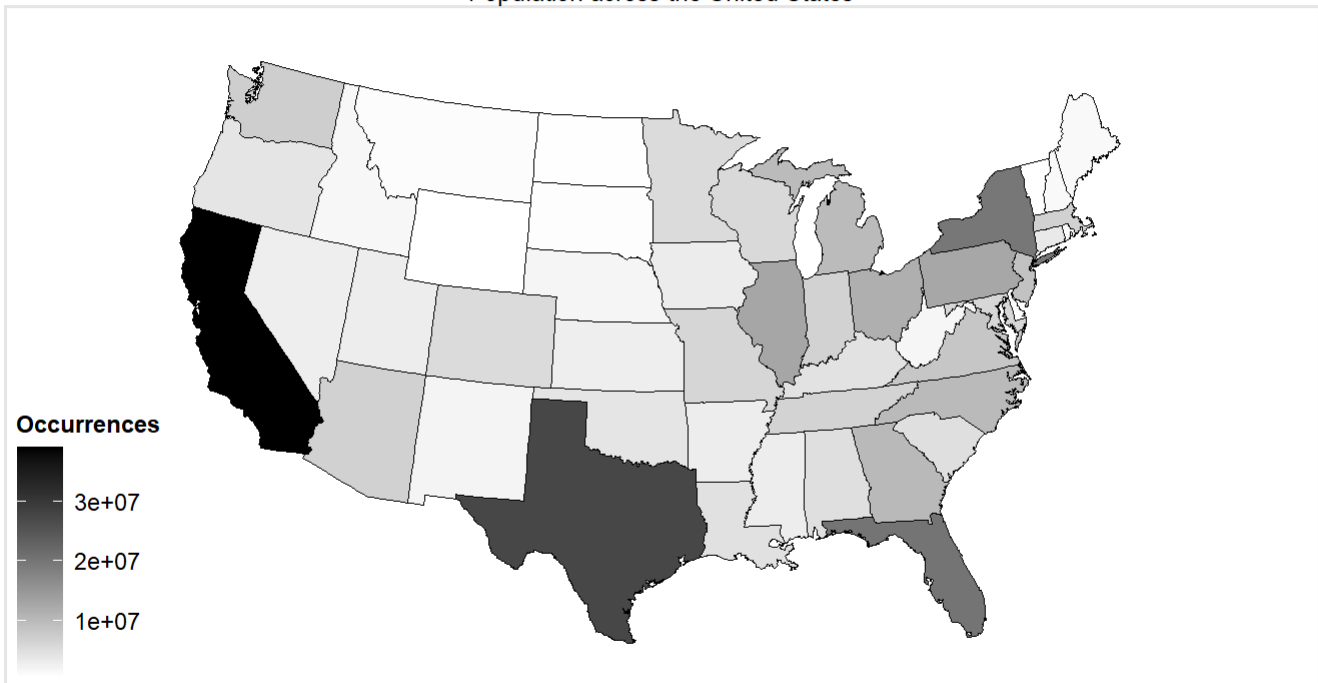
Population map

As an overview, I wanted to see the population distribution across the United States, so that I could have the high
population centers in mind when analyzing crime patterns. Based on the map, it seems California, Texas, New
York, and Florida have the largest populations.

```
#population map
popPlot <- ggplot(stateCrime, aes(x=long, y=lat, group=group, fill=Population)) +
  geom_polygon(color = "black", size = 0.1) +
  coord_map(projection = "albers", lat0 =30, lat1 = 45) +
  theme_map() +
  labs(title = "Population across the United States", fill = "Occurrences") +
  scale_fill_gradient(low =  "#ffffff", high = "#000000")
```

```
## Loading required package: grid
```

```
popPlot #output point
```
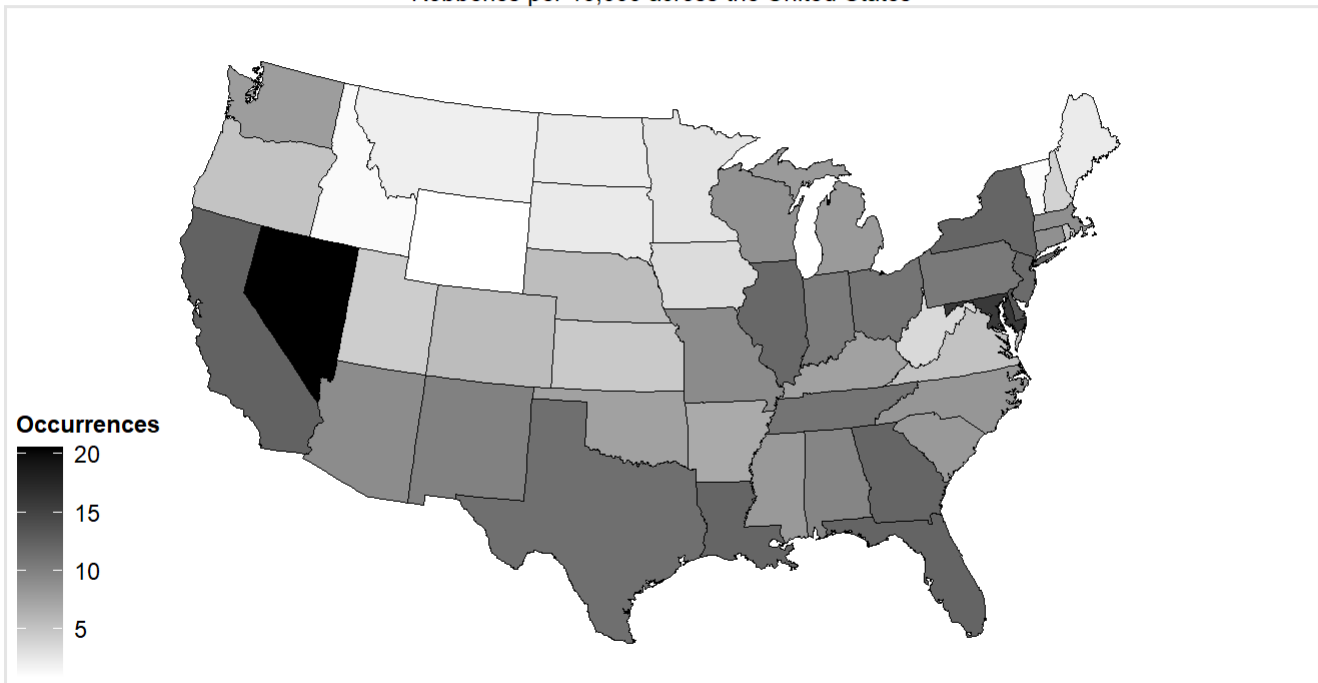
Population across the United States

Robbery map

Next, I decided to look at the robberies across the United States. However, because the data does not give the occurrences per 10,000, I had to calculate that. I did so by dividing the robbery by the population, giving an occurrence per person, and then multiplying by 10,000, to represent the occurrences per 10,000 people. It seems like the areas with high population had a relatively high count of robberies per 10,000 people, but I was more surprised at the high concentration in the East, particularly states like Idaho, Ohio, and Tennessee. Furthermore, it seems there are not many robberies in the North west, which makes sense as the population density there is less.

```
#robbery map
robPlot <- ggplot(stateCrime, aes(x=long, y=lat, group=group, fill=Robbery/Population*10000)) +
  geom_polygon(color = "black", size = 0.1) +
  coord_map(projection = "albers", lat0 =30, lat1 = 45) +
  theme_map() +
  labs(title = "Robberies per 10,000 across the United States", fill = "Occurrences") +
  scale_fill_gradient(low =  "#ffffff", high = "#000000")
robPlot #output point
```

Robberies per 10,000 across the United States

Larceny map

For the next map, I wanted to look at the distribution of larceny per 10,000 across the United States. In looking at the table of the dataset, I recognized that Larceny was one of the highest occurring crimes, so I wanted to get a closer look at it. I was not particularly shocked to see that there was much more larceny across the whole United States, although I was surprised to see a high larceny rate in the Northwest. I was particularly surprised to see Minnesota in white, as its larceny rate was the clear lowest in the whole nation.

```
#Larceny map
larPlot <- ggplot(stateCrime, aes(x=long, y=lat, group=group, fill=Larceny/Population*10000)) +
  geom_polygon(color = "black", size = 0.1) +
  coord_map(projection = "albers", lat0 =30, lat1 = 45) +
  theme_map() +
  labs(title = "Larceny per 10,000 across the United States", fill = "Occurrences") +
  scale_fill_gradient(low =  "#ffffff", high = "#000000")
larPlot #output point
```

Larceny per 10,000 across the United States
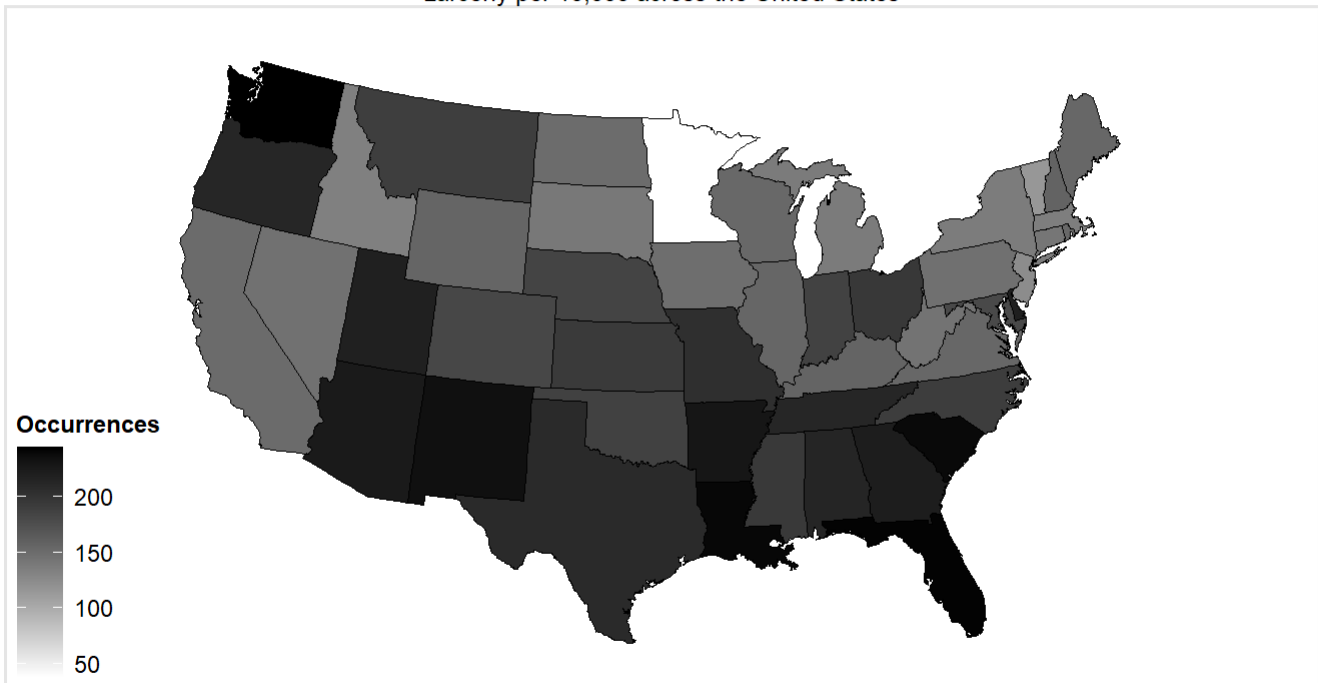
Murder map

After the larceny map, I wanted to look at one of the more serious violent crimes, murder. I felt like looking at murder next to larceny would help me draw some contrast and relationship between the two. The thing that immiedatly caught my attention was the high murder rate in Lousiana, with around 1 murder per 10,000 people. Louisiana easily had the highest murder rate in the nation, however it seemed strange that while it would have a high murder and larceny rate, its robbery rate was relatively low. Perhaps it could be attributed to the higher income inequality in the state, as this would drive people to commit more small crimes, like larceny, but not breaking-and-entering, because of the income inequality. Furthermore, Lousiana's ranking as the poorest state in the nation would also explain its higher murder rate.

```
#murder map
murPlot <- ggplot(stateCrime, aes(x=long, y=lat, group=group, fill=Murder/Population*10000)) +
  geom_polygon(color = "black", size = 0.1) +
  coord_map(projection = "albers", lat0 =30, lat1 = 45) +
  theme_map() +
  labs(title = "Murders per 10,000 across the United States", fill = "Occurrences") +
  scale_fill_gradient(low =  "#ffffff", high = "#000000")
murPlot #output point
```

Murders per 10,000 across the United States
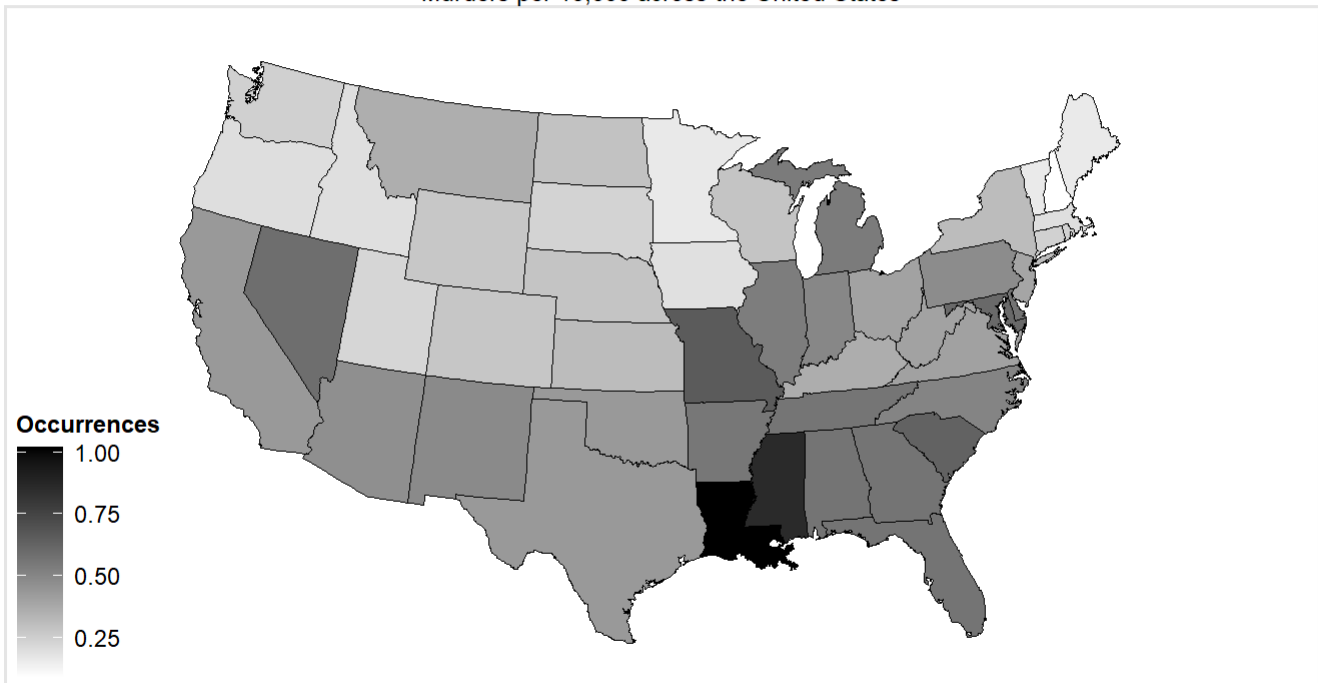
Total map

For my final map, I wanted to look at a sum of all the crime data cross the nation. However, I realized that the annual problem in the existing dataset was not accurate, so I needed to recalculate the column with accurate values. I rewrote the annual column to a sum of all crime types (this is one of the reasons why I took out the property column).

```
#recalculate annual crime total because it is inaccurate
stateCrime$Annual <- stateCrime$Murder + stateCrime$Rape + stateCrime$Robbery + stateCrime$Assau
lt + stateCrime$Burglary + stateCrime$Larceny + stateCrime$Auto
```

With an accurate annual column, I can now map out total criminal activity across the United States. I toned down the rate to 100 people here, because keeping it at 10,000 was giving numbers that were too large. What I immiedatly noticed was the high crime rate in Minnesota, possibly one of the highest in the nation. I was very surprised, as Minnesota had been a very safe state in previous maps. However, looking back at the data I can see that Minnesota has extremely high occurrences of burglary and vehicle-theft, which contributed to make it one of the states with the highest crime rate. I was not suprised to see states like Washington and Lousiana to have high crime rates, as both of those states had had high crime rates in previous maps.

```
#total map (computed column)
totPlot <- ggplot(stateCrime, aes(x=long, y=lat, group=group, fill=Annual/Population*100)) +
  geom_polygon(color = "black", size = 0.1) +
  coord_map(projection = "albers", lat0 =30, lat1 = 45) +
  theme_map() +
  labs(title = "Crime per 100 across the United States", fill = "Occurrences") +
  scale_fill_gradient(low =  "#ffffff", high = "#000000")
totPlot #output point
```


Crime per 100 across the United States
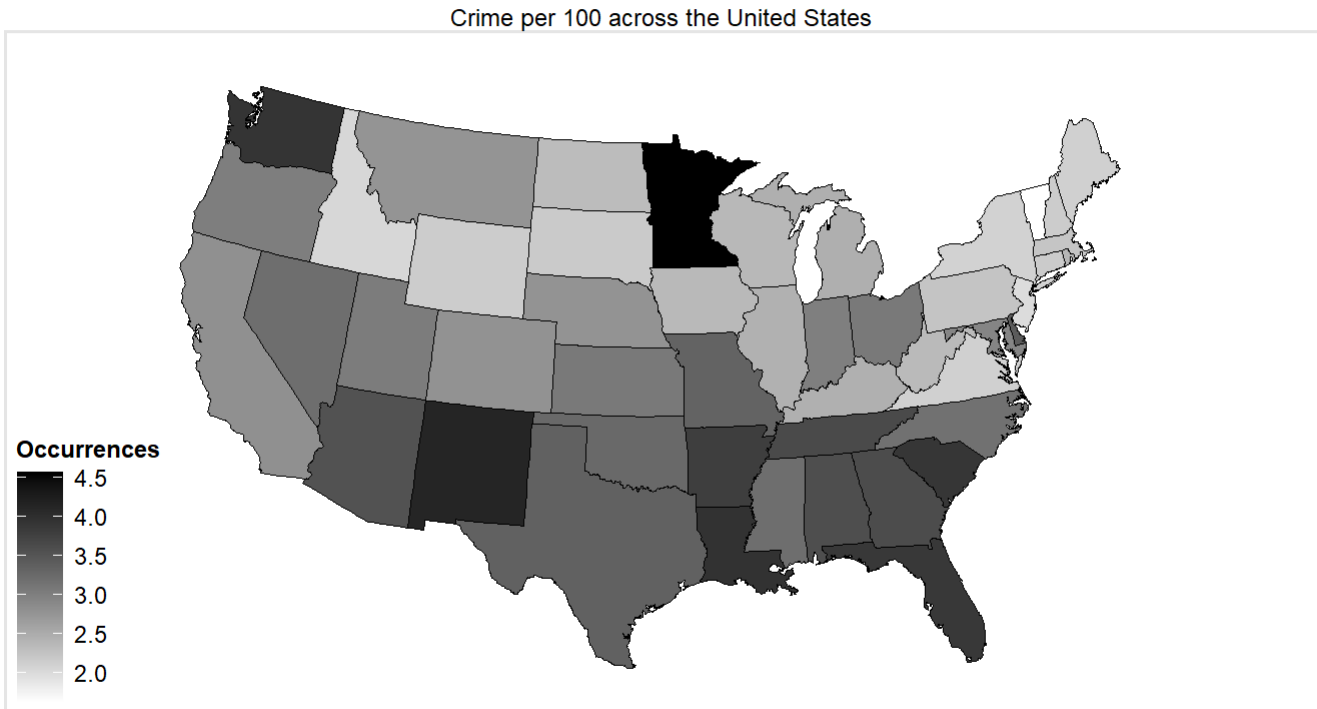
Conclusions

From my graphical analysis of the crime data I was able to draw a few conclusions:

1. The Northwestern states generally have a low crime rate. States like Idaho, Montana, North Dakota, South Dakota, and Wyoming showed a consistently low crime rate across all maps. I would attribute this lower crime rate to their lower populations, coupled with their large areas.

2. Louisiana is one of the states with the highest criminal activity. Louisiana had consistently high crime rates across all maps, therefore it is one of the states with the highest criminal activity. I would attribute this high crime rate to their large income inequality and rampant poverty, as both are factors for high crime.

3. High population does not imply high crime. While one would assume that a high population would mean there is a higher likelihood of crime, but the graphics in this report would suggest otherwise. The states with a higher population, while they did have higher crime rates, did not have crime rates proportional to their

large populations. For example, while Texas and California both had large crime rates than states like Idahao and Wyoming, their crime rates were no where near states like New Mexico and Lousiana, therefore disproving the correlation between crime rate and population.