U.S. Crimes Data Visualization

Advaith Cheruvu

10/5/2021

Importance of Study

Criminal offenses have been increasing as of late, causing damage to public security and social order. Often, this damage is thought of as the result of the crime, including personal property theft, victim injury, and others. But also, these crimes often lead to uncertainty in the local and national government being able to keep their people safe and secure. To combat these problems, the government must focus funding on law enforcement, emergency response, and other related public safety measures. This way, crime decreases and the people are better protected. However, With government's limited resources, it begs the question, which areas need the most support? Knowing the states that have the highest crime rates will help the government focus funding. Also, knowing the types of crimes that these states are most often faced with can help the government focus funding. For example, property crimes and larceny yield less severe consequences than murder and and rape, allowing the national government to focus funding on the states with more severe crimes. Finally, understanding relationships that result in crimes will help the government focus funding on the states that most require crime prevention measures.

Crime in relation to COVID-19

The COVID-19 pandemic has caused economic and mental distress for many individuals, causing some of them to turn to crime. Even before the pandemic, crimes were mainly focused in urban centers. As the pandemic rolled through, these urban areas were more deeply affected than their rural and suburban counterparts. This means that the already crime-ridden urban regions had even worse crime rates. These changes have further exacerbated the issue and brings more importance to the analysis of crime data.

Setting up Workspace and Installing Packages

The code below shows the preparation of the workspace and cleaning up the environment, as well as installing and loading packages.

```
#
# clean up and set up
rm(list=ls())
setwd("/Users/advai/Documents/DSFS")
#
# install and load libraries
library(tidyverse)
```

Loading and Looking at Crimes and US States Data

Eventually these datasets will be combined, but the crimes data requires some minor cleaning.

```
#
# load crimes data
crimes <- read.csv(file = "C:\\Users\\advai\\Documents\\DSFS\\newCrimes.csv",header=T)
#
# looking at crime data
names(crimes)
```

```
##  [1] "State"      "Population" "Annual"     "Per10K"     "Murder"
##  [6] "Rape"       "Robbery"    "Assault"    "Property"   "Burglary"
## [11] "Larceny"    "Auto"
```

```
summary(crimes)
```

```
##     State              Population           Annual           Per10K
##  Length:50          Min.   :  586107   Min.   :   739   Min.   :12.00
##  Class :character   1st Qu.: 1857144   1st Qu.:  5602   1st Qu.:26.00
##  Mode  :character   Median : 4547908   Median : 15962   Median :35.00
##                     Mean   : 6417926   Mean   : 24364   Mean   :36.30
##                     3rd Qu.: 7084780   3rd Qu.: 27875   3rd Qu.:42.75
##                     Max.   :39144818   Max.   :166883   Max.   :73.00
##      Murder            Rape            Robbery         Assault
##  Min.   :  10.0   Min.   :  110.0   Min.   :   53   Min.   :  432
##  1st Qu.:  55.5   1st Qu.:  773.8   1st Qu.: 1060   1st Qu.: 3365
##  Median : 165.0   Median : 1547.0   Median : 3248   Median :10037
##  Mean   : 282.9   Mean   : 2323.5   Mean   : 6401   Mean   :14682
##  3rd Qu.: 369.5   3rd Qu.: 2503.2   3rd Qu.: 7173   3rd Qu.:17500
##  Max.   :1699.0   Max.   :11527.0   Max.   :48680   Max.   :91803
##     Property         Burglary         Larceny            Auto
##  Min.   :  6729   Min.   :  1689   Min.   :  7273   Min.   :   244
##  1st Qu.: 39060   1st Qu.:  8206   1st Qu.: 27994   1st Qu.:  3947
##  Median :115144   Median : 23912   Median : 81709   Median :  9720
##  Mean   :162501   Mean   : 36619   Mean   :115123   Mean   : 15472
##  3rd Qu.:194395   3rd Qu.: 44029   3rd Qu.:135388   3rd Qu.: 13803
##  Max.   :947192   Max.   :202670   Max.   :592670   Max.   :151852
```

```
dim(crimes)
```

```
## [1] 50 12
```

```
class(crimes)
```

```
## [1] "data.frame"
```

```
head(crimes)
```

```
##        State Population Annual Per10K Murder  Rape Robbery Assault Property
## 1    Alabama    4858979  22952     47    276  2005    4701   13745   154094
## 2     Alaska     738432   5392     73     41   771     629    3243    20334
## 3    Arizona    6828065  28012     41    319  3378    6249   16970   215240
## 4   Arkansas    2978204  15526     52    165  1763    2050   10265    99018
## 5 California   39144818 166883     43   1699 11527   48680   91803   947192
## 6   Colorado    5456574  17515     32    151  3039    3039   10325   135510
##   Burglary Larceny   Auto
## 1    39715  104238  10141
## 2     3150   15445   1739
## 3    43562  154091  17587
## 4    24790   68627   5601
## 5   202670  592670 151852
## 6    23472   99464  12574
```

```
#
# Looking at US States data
us_states<-map_data("state")
head(us_states)
```

```
##        long      lat group order  region subregion
## 1 -87.46201 30.38968     1     1 alabama      <NA>
## 2 -87.48493 30.37249     1     2 alabama      <NA>
## 3 -87.52503 30.37249     1     3 alabama      <NA>
## 4 -87.53076 30.33239     1     4 alabama      <NA>
## 5 -87.57087 30.32665     1     5 alabama      <NA>
## 6 -87.58806 30.32665     1     6 alabama      <NA>
```

```
dim(us_states)
```

```
## [1] 15537     6
```

```
summary(us_states)
```

```
##       long              lat            group          order       
##  Min.   :-124.68   Min.   :25.13   Min.   : 1.00   Min.   :    1  
##  1st Qu.: -96.22   1st Qu.:33.91   1st Qu.:15.00   1st Qu.: 3899  
##  Median : -87.61   Median :38.18   Median :26.00   Median : 7794  
##  Mean   : -89.67   Mean   :38.18   Mean   :30.15   Mean   : 7798  
##  3rd Qu.: -79.13   3rd Qu.:42.80   3rd Qu.:47.00   3rd Qu.:11699  
##  Max.   : -67.01   Max.   :49.38   Max.   :63.00   Max.   :15599  
##     region           subregion        
##  Length:15537       Length:15537      
##  Class :character   Class :character  
##  Mode  :character   Mode  :character  
##                                       
##                                       
##
```

Creating New Variables

The “Property_Per10k” variable was created to compare the property crime rate for various states. The “Murder_Per10k” variable was created to compare the murder rate for various states. Property crime can be considered less impactful than murder, meaning that I can analyze which states require more government support, based more heavily on the murder rate than property crime rate. Finally, the “PopInMil” variable simply converts the population number to the population in millions to allow the graphs to be easier to read.

```
#
# Creating a new variable for property per 10k
crimes$Property_Per10k <- (crimes$Property/crimes$Population) * 10000
#
# Creating a new variable for murders per 10k
crimes$Murder_Per10k <- (crimes$Murder/crimes$Population) * 10000
#
# Creating a new variable for population in millions
crimes$PopInMil <-(crimes$Population/1000000)
```

Merging Datasets

Merging the datasets allows R to associate the crime data with the states, thus allowing me to create visualizations.

```
#
# Merging us_states and crimes
crimes$region <- tolower(crimes$State)
us_crimes<- left_join(us_states, crimes)
```

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

```
summary(us_crimes)
```

```
##       long              lat            group           order       
##  Min.   :-124.68   Min.   :25.13   Min.   : 1.00   Min.   :    1  
##  1st Qu.: -96.22   1st Qu.:33.91   1st Qu.:15.00   1st Qu.: 3899  
##  Median : -87.61   Median :38.18   Median :26.00   Median : 7794  
##  Mean   : -89.67   Mean   :38.18   Mean   :30.15   Mean   : 7798  
##  3rd Qu.: -79.13   3rd Qu.:42.80   3rd Qu.:47.00   3rd Qu.:11699  
##  Max.   : -67.01   Max.   :49.38   Max.   :63.00   Max.   :15599  
##                                                                   
##     region           subregion            State              Population      
##  Length:15537       Length:15537       Length:15537       Min.   :  586107  
##  Class :character   Class :character   Class :character   1st Qu.: 4093465  
##  Mode  :character   Mode  :character   Mode  :character   Median : 6619680  
##                                                           Mean   : 9762434  
##                                                           3rd Qu.:10214860  
##                                                           Max.   :39144818  
##                                                           NA's   :10        
##      Annual            Per10K          Murder            Rape       
##  Min.   :   739   Min.   :12.00   Min.   :  10.0   Min.   :  110  
##  1st Qu.: 12475   1st Qu.:28.00   1st Qu.: 160.0   1st Qu.: 1411  
##  Median : 25208   Median :38.00   Median : 338.0   Median : 2310  
##  Mean   : 38163   Mean   :36.67   Mean   : 450.3   Mean   : 3565  
##  3rd Qu.: 41231   3rd Qu.:43.00   3rd Qu.: 535.0   3rd Qu.: 5042  
##  Max.   :166883   Max.   :70.00   Max.   :1699.0   Max.   :11527  
##  NA's   :10       NA's   :10      NA's   :10       NA's   :10     
##     Robbery          Assault          Property         Burglary      
##  Min.   :   53   Min.   :  432   Min.   :  6729   Min.   :  1689  
##  1st Qu.: 2430   1st Qu.: 6264   1st Qu.: 99018   1st Qu.: 23122  
##  Median : 5695   Median :15133   Median :160824   Median : 38337  
##  Mean   :10195   Mean   :23477   Mean   :256502   Mean   : 57229  
##  3rd Qu.:12417   3rd Qu.:27519   3rd Qu.:285697   3rd Qu.: 76428  
##  Max.   :48680   Max.   :91803   Max.   :947192   Max.   :202670  
##  NA's   :10      NA's   :10      NA's   :10       NA's   :10      
##     Larceny            Auto         Property_Per10k  Murder_Per10k    
##  Min.   :  7273   Min.   :   244   Min.   : 12.33   Min.   :0.09018  
##  1st Qu.: 68627   1st Qu.:  7298   1st Qu.:204.13   1st Qu.:0.31168  
##  Median :124022   Median : 13040   Median :250.39   Median :0.43403  
##  Mean   :181102   Mean   : 23832   Mean   :254.01   Mean   :0.45093  
##  3rd Qu.:199926   3rd Qu.: 21157   3rd Qu.:296.31   3rd Qu.:0.56210  
##  Max.   :592670   Max.   :151852   Max.   :364.98   Max.   :1.02126  
##  NA's   :10       NA's   :10       NA's   :10       NA's   :10       
##     PopInMil      
##  Min.   : 0.5861  
##  1st Qu.: 4.0935  
##  Median : 6.6197  
##  Mean   : 9.7624  
##  3rd Qu.:10.2149  
##  Max.   :39.1448  
##  NA's   :10
```

```
us_crimes$Per10K <- as.numeric(us_crimes$Per10K)
```

Visualizations

I will be mainly creating maps since this main point of this data analysis is to determine which states need government support. This means that spatial organization found in a map will be helpful.

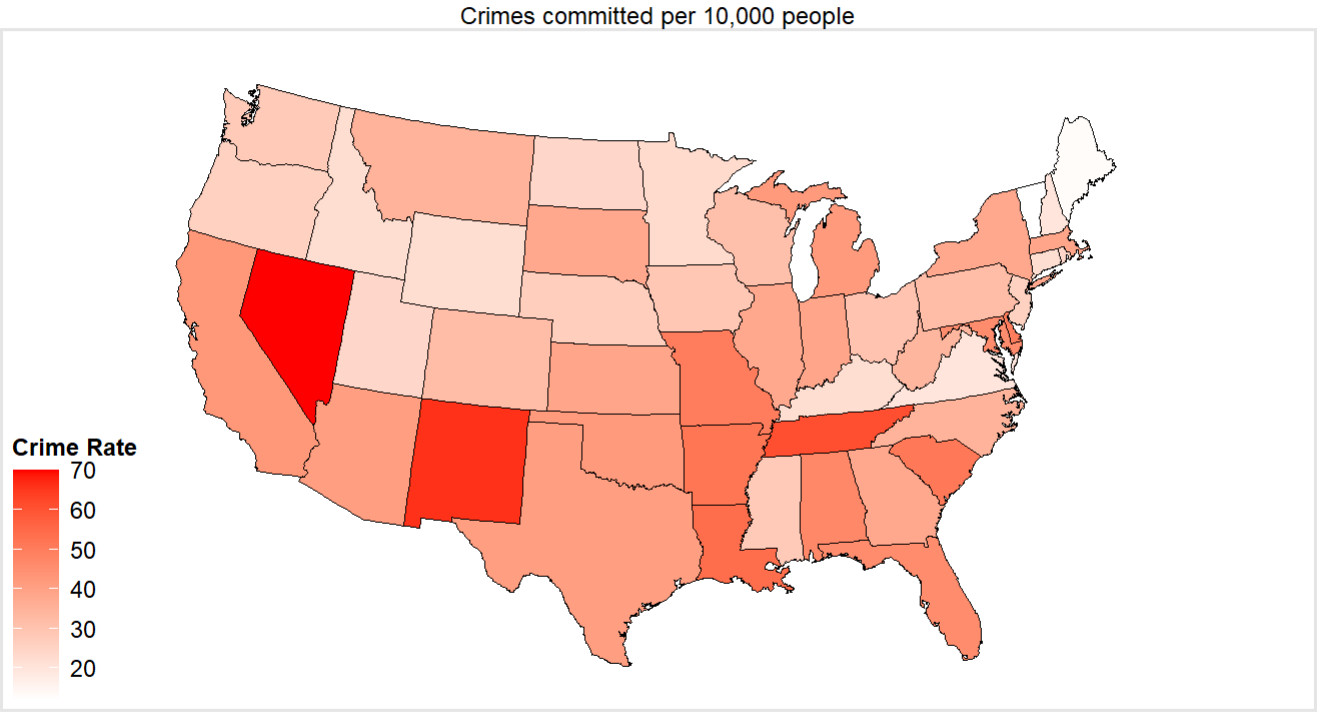
Map of Crime Rate per 10,000 People

I will first create a map of crime committed per 10,000 people. This will give a general idea of the crime rates, giving us a basic answer to which states require the most government support.

```
##
# Creating map of crime rates
per10kMap <-
  ggplot(data=us_crimes,
  mapping=aes(x=long, y=lat, group=group, fill=Per10K)) + # maps aesthetics
  scale_fill_continuous(low = "white", high="red") + # color scale
  geom_polygon(color="black", size=0.08) + # creates border aesthetics
  coord_map(projection="albers", lat0=30, lat1=45) + # albers projection
  labs(title="Crimes committed per 10,000 people", fill="Crime Rate") + # title and legend
  theme_map()
```

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

```
#
# Print the per10k map
per10kMap
```

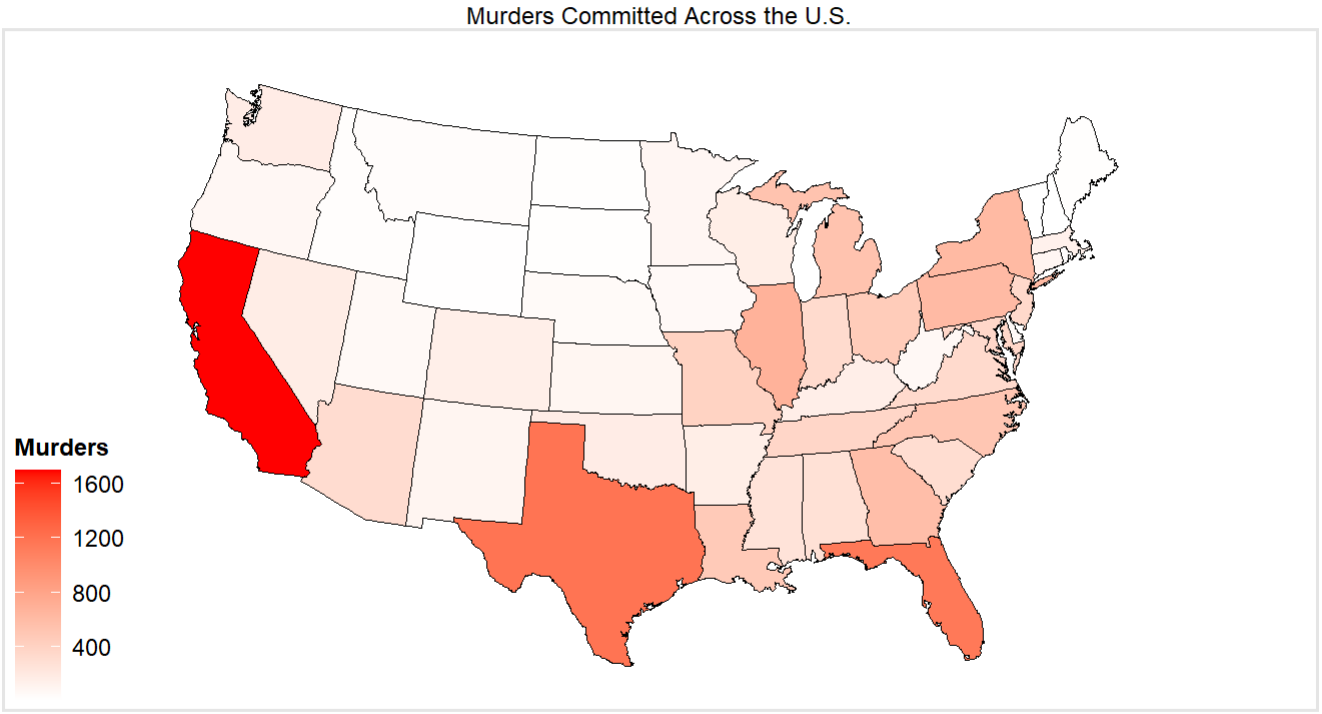


This map shows that Nevada and New Mexico stand out. This means that more people per 10,000 commit crimes in Nevada and New Mexico than in other states. These states would need general support from the government, but we do not know which crimes occur in these states, only that there are more crimes in these states.

Map of Murders

Next, I'll create a map showing where murders happen in the U.S. This will informs us which states have the most murders. This violent crime provides a comparison between murder and larceny (the next map), allowing us to cross reference the states that are high in these crimes.

```
#
# Creating a map of murders
murdersMap <-
  ggplot(data=us_crimes,
         mapping=aes(x=long, y=lat, group=group, fill=Murder)) + # maps aesthetics
  scale_fill_continuous(low = "white", high="red") + # color scale
  geom_polygon(color="black", size=0.08) + # creates border aesthetics
  coord_map(projection="albers", lat0=30, lat1=45) + # albers projection
  labs(title="Murders Committed Across the U.S.", fill="Murders") + # title and legend
  theme_map()
#
# Print the murdersMap
murdersMap
```

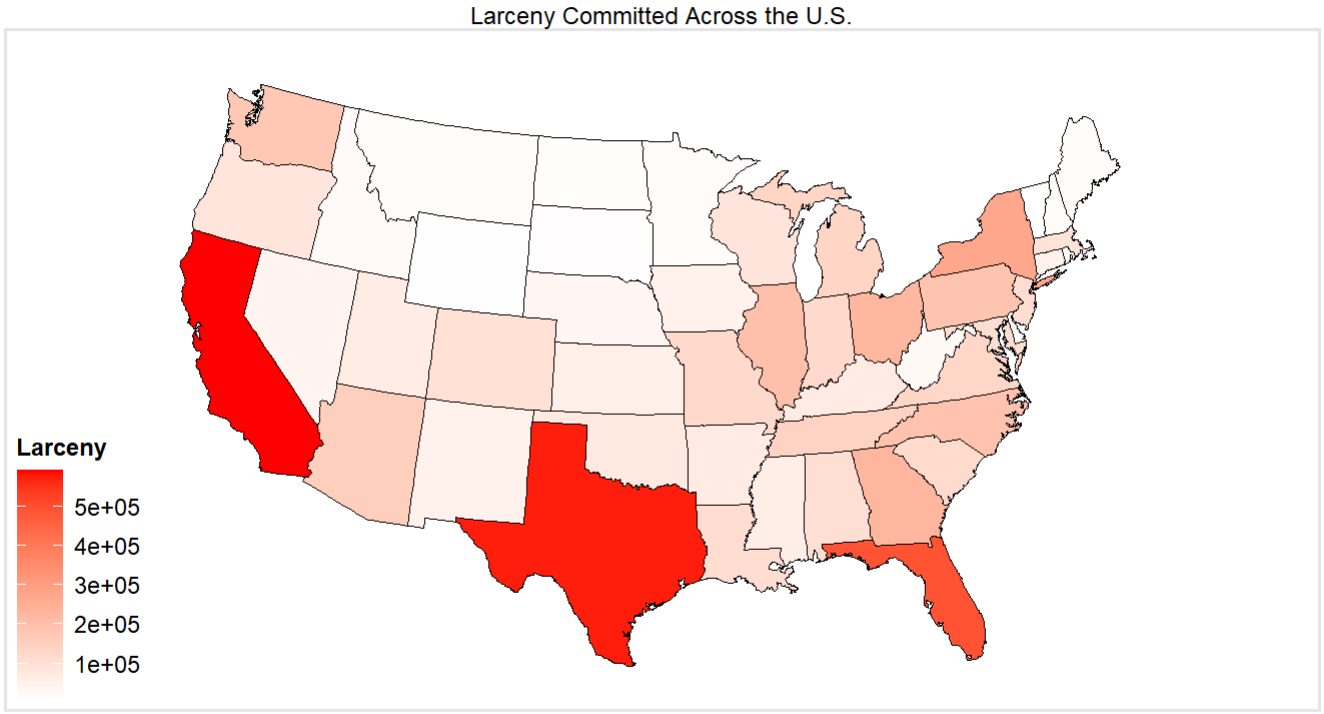


According to the map, California, Texas, and Florida seem to have the highest amount of murders. This indicates that the government can focus police investigation and organized crime prevention in these states. Also, it should come as no coincidence that these states have the highest population due to number of cities. Therefore, this graph gives us an idea of the relationship between population and murders.

Map of Larceny

To contrast the murder map, This range from murder to larceny gives the perspective of scale. A smaller scale crime, such as larceny, may provide different patterns from a larger scale crime, such as murder.

```
#
# Creating a map of Larceny
larcenyMap <-
  ggplot(data=us_crimes,
         mapping=aes(x=long, y=lat, group=group, fill=Larceny)) + # maps aesthetics
  scale_fill_continuous(low = "white", high="red") + # color scale
  geom_polygon(color="black", size=0.08) + # creates border aesthetics
  coord_map(projection="albers", lat0=30, lat1=45) + # albers projection
  labs(title="Larceny Committed Across the U.S.", fill="Larceny") + # title and legend
  theme_map()
#
# Print the larcenyMap
larcenyMap
```

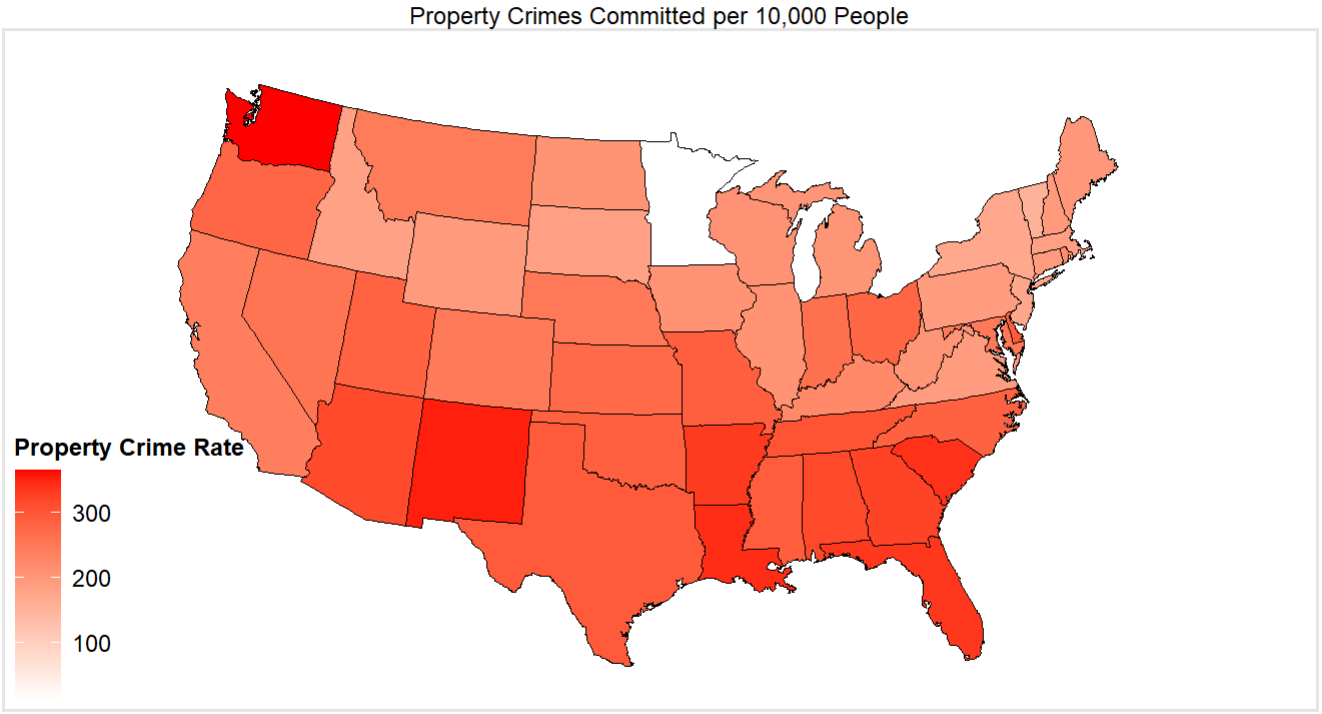


This map shows that larceny was most often committed in California, Texas, and Florida, which supports idea that more urban centers relates to more crime. Except for some minor changes most states have similar relative murder rates and larceny rates, meaning that the population and urban centers could affect the crime rate across the scale of the crime.

Property Crimes Rate Map

Property crimes are another small scale crime which gives us the information about how much small scale crime each state has. This will provide a comparison between the states, allowing the government to focus funding on basic security measures that are tailored towards property crimes.

```
#
# Creating a map of Property_Per10k
propertyPer10KMap <-
  ggplot(data=us_crimes,
         mapping=aes(x=long, y=lat, group=group, fill=Property_Per10k)) + # maps aesthetics
  scale_fill_continuous(low = "white", high="red") + # color scale
  geom_polygon(color="black", size=0.08) + # creates border aesthetics
  coord_map(projection="albers", lat0=30, lat1=45) + # albers projection
  labs(title="Property Crimes Committed per 10,000 People", fill="Property Crime Rate") + # titl
e and legend
  theme_map()
#
# Print the Property_Per10k
propertyPer10KMap
```

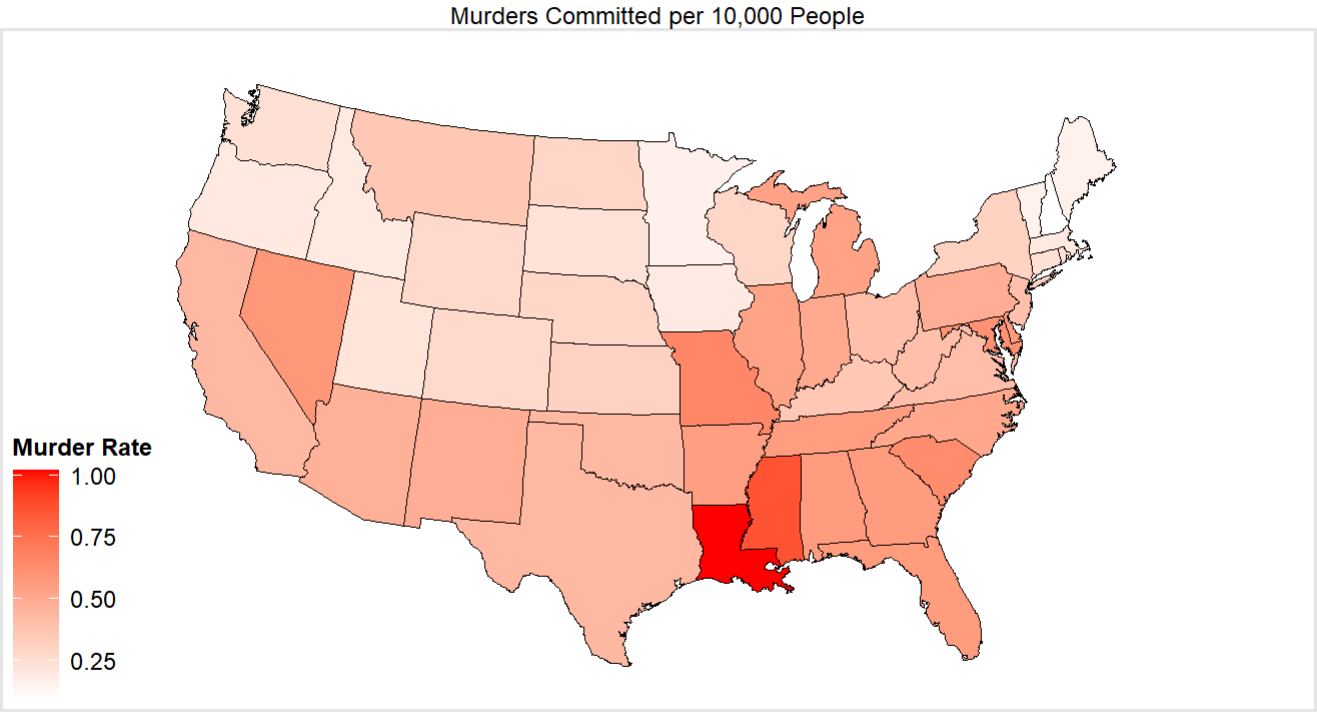


This map shows that most U.S. states have around the same property crime rate, though the southeast coast has considerably higher property crime rates. Also, it should be noted that Washington State has the highest property crime rate. More analysis is needed to determine exactly why this is the case, though I can speculate that it may be due to the fact that this state has a higher share of the population in cities vs rural and suburban areas. Also, Minnesota has an extremely low property crime rate. Again, more analysis is needed to discover exactly why, but the tactics Minnesota is using to prevent property crime may be applied elsewhere for similar success.

Murder Rate Map

The murder rate map can be used to compare to the property crime map to determine how a change in the scale of the crime can affect its prevalence and could perhaps indicate different patters across the scale.

```
#
# Create a map of Murders_per10k
murders_per10KMap <-
  ggplot(data=us_crimes,
         mapping=aes(x=long, y=lat, group=group, fill=Murder_Per10k)) + # maps aesthetics
  scale_fill_continuous(low = "white", high="red") + # color scale
  geom_polygon(color="black", size=0.08) + # creates border aesthetics
  coord_map(projection="albers", lat0=30, lat1=45) + # albers projection
  labs(title="Murders Committed per 10,000 People", fill="Murder Rate") + # title and legend
  theme_map()
#
# Print the Property_Per10k
murders_per10KMap
```



This map yields some particularly interesting results. Firstly, the murder rate varies much more than the property crime rate. I suspect this may be due to the fact that number of murders was much less than the number of property crimes, meaning that a single murder would affect the murder rate much more than the property crime affecting the property rate. Also, where Washington state had the highest property crime rate, it now has one of the lowest murder rates, meaning that crimes of different scales can be committed with different rates. Also, Louisiana, Mississippi, and Missouri seem to have the highest murder rates. This, along with the fact that the south in general has a higher murder rate could indicate that murder rates are not only affected by city amount and distribution. More analysis is required to determine exactly why these states exhibit these patterns. The national government can focus more on funding investigative work in the south and specifically Missouri, Louisiana, and Mississippi.

Population and Annual Crimes Scatter Plot

Finally, this last plot shows the relationship between annual crimes and population. Though it seems obvious that a higher population would mean that more criminals would be part of the population, this plot can provide valuable information about states that are above or below the average.

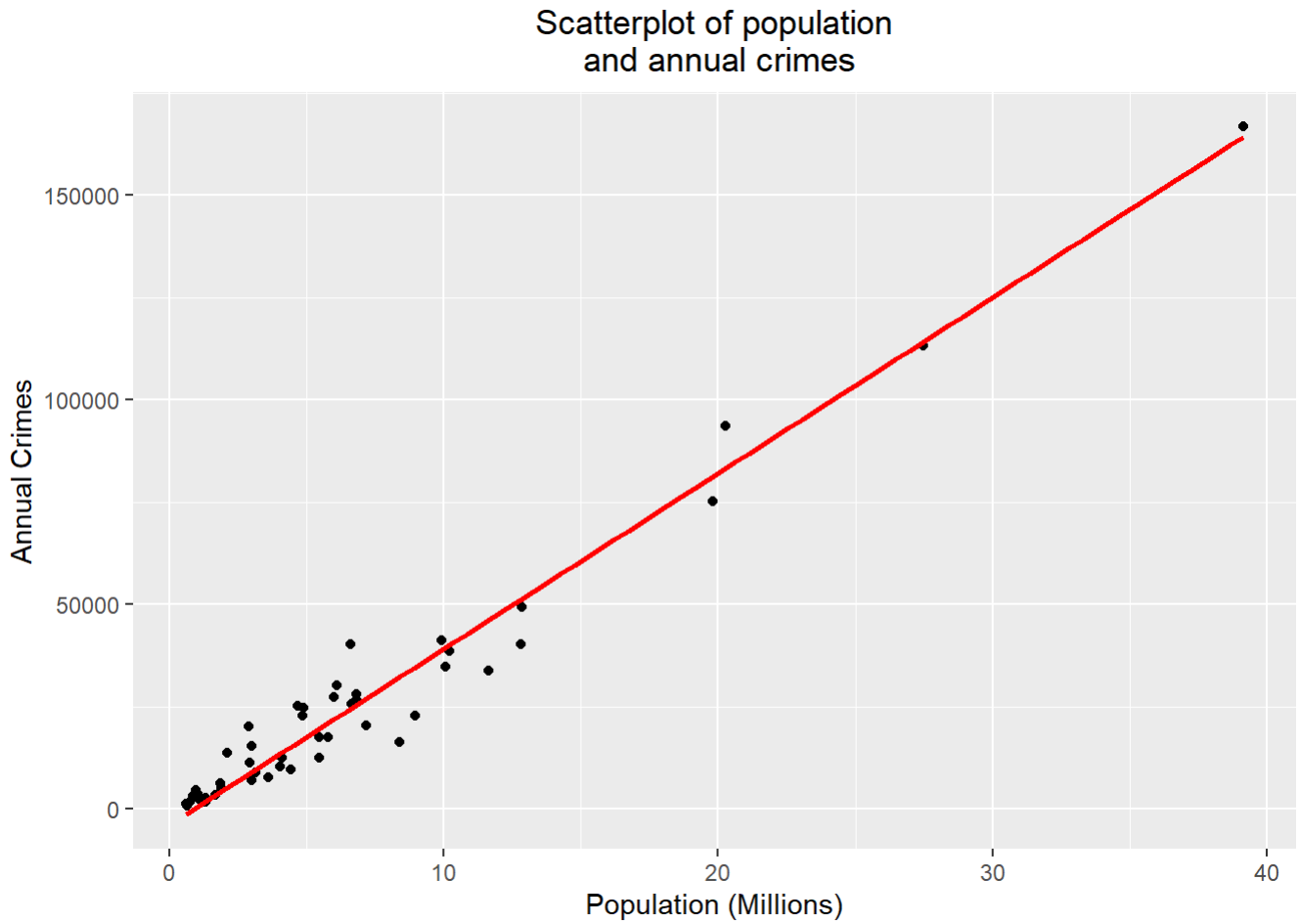
```
#
# Scatterplot between population and annual crimes
popCrimesPlot <- ggplot(us_crimes, aes(x=PopInMil,                    # sets the x to populatio
n
                        y=Annual))+
  geom_point() +                                      # tells R to create a scatter plot
  xlab("Population (Millions)") +                     # labels the x axis
  ylab("Annual Crimes") +                                    # labels the y axis
  ggtitle("Scatterplot of population\n and annual crimes") +  # labels the plot
  theme(plot.title = element_text(hjust = 0.5)) +            # centers the title
  geom_smooth(method=lm, color="red")                        # colors the linear regression line
red

# Prints the plot
popCrimesPlot
```

```
## `geom_smooth()` using formula 'y ~ x'
```

```
## Warning: Removed 10 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 10 rows containing missing values (geom_point).
```



The states below the linear regression line are the states that have the lowest amount of annual crimes for their population. These are the states that can be used as exemplars for other states to reduce their annual crimes. Similar policies from these states can be used in the states that aren't using the policies and other similar traits can be extracted with more analysis.

Work Cited

"Crime." Legal Information Institute, Legal Information Institute, 2021, www.law.cornell.edu/wex/crime.

"The Impact of Crime." From Pain to Power: The Impact of Crime, Office for Victim Clinic, 2021, www.ncjrs.gov/ovc_archives/reports/fptp/impactcrm.htm.

Justice., Council on Criminal. "NCCCJ - Impact Report: Covid-19 and Crime." National Commission on COVID-19 and Criminal Justice, National Commission on COVID-19 and Criminal Justice, 21 May 2021, covid19.counciloncj.org/2021/05/21/impact-report-covid-19-and-crime-4/.

Walden University. "Why-National-Crime-Statistics-Are-Important." Walden University, Walden University, 25 Mar. 2021, www.waldenu.edu/online-bachelors-programs/bs-in-criminal-justice/resource/why-national-crime-statistics-are-important.