FBI Crime Capstone Project Milestone Report

Ankita Hanmiah

February 7, 2019

Introduction

The Federal Bureau of Investigation's Uniform Crime Reporting program contains a plethora of data intended to provide information to law enforcement administration and the public on the prevalence of different crimes throughout the United States. Crimes that come to their attention are compiled into four collections of data: national incidents, crimes against law enforcement officials, hate crimes, and general criminal offenses. In this project, the focus is on general crime: violent and property crime. The intended questions to be answered are: which states are these crimes most prevalent in? Which are the "safest"" states and the most "dangerous"? Is there a relationship between violent and property crime? And does poverty share a relationship with crime?

This information would be beneficial for people, probably starting a family, who are looking for a place to put down roots that is safe for their family, or at the very least so that they may be informed and aware of the kind of crime that is most prevalent where they live. This would also be useful for law enforcement and state governments, so that they may be able to more thoroughly understand the current most widespread crime-related, and possibly poverty-related, issues within their jurisdiction. This information could help them build new laws, programs, or make reforms that move towards preventing these kinds of crimes within their state.

Data and Data Wrangling

The original crime data used for this project was collected from Table 4 of the 2017 FBI UCR (FBI Crime Data - State and Region (Original).csv). The FBI compiled 24 tables of data for their offenses, with different breakdowns of regions, states, counties, cities, metropolitan versus non-metropolitan areas, suburban versus non-suburban areas, generalized location of the offenses, weapons used in violent crime, time of day for property crime, and other circumstances. There is additional expanded data on specific crimes as well as clearance data, providing information on how the case was closed: whether or not an arrest was made or if there was another circumstantial close to the case. The table used for this project contained data on the counts and rates of different criminal offenses, as well as counts and rates for violent and property crime, for each state plus the District of Columbia. Violent crime is defined by the FBI as "offenses that involve force or threat of force" and encompasses murder and nonnegligent manslaughter, aggravated assault, rape, and robbery. Property crime is defined as "offenses in which there is no force or threat of force against the victims" and includes burglary, larceny, and motor vehicle theft. The crime dataset being used contains data for 2016 and 2017, though only 2017 will be used for this project to determine what the most current answers are, as this is most beneficial to families and law enforcement looking to make current decisions.

The poverty data (Poverty_State2017 (Original).csv) is compiled by Census using their measures of determining which families fall within the poverty limit. This is measured by using sets of income thresholds that differs based on the size of the family and their composition. The family's total income, before taxes and not including benefits, must be under the determined threshold for the family to be considered in poverty.

The data collected from the FBI UCR and Census was reorganized into three separate data sets: "FBICrime"(CrimeByState&Offenses.csv), "FBICrime2"(Violent&PropertyCrime.csv), and "Crime-Poverty"(Crime&PovertyByState.csv):

```
library(tidyverse)

FBICrime <- read_csv("FBI Crime Data - State and Region (Original).csv")
Poverty <- read_csv("Poverty_State2017 (Original).csv")

#With the poverty data, the columns to be used in this project were subset and renamed
```

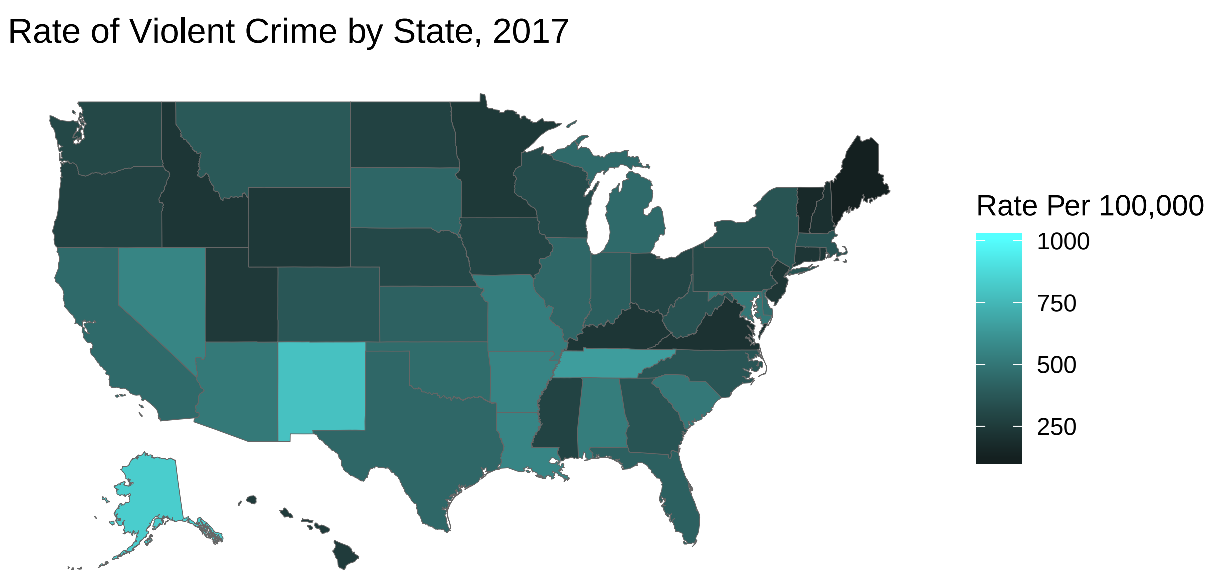
```r
poverty_names <- make.names(c("State", "Poverty Count", "Poverty Rate"))
Poverty <-
  Poverty[c("State / County Name", "All Ages in Poverty Count", "All Ages in Poverty Percent")] %>%
  setNames(poverty_names)
Poverty$State <- tolower(Poverty$State)

#The unnecessary data, extra rows, and empty columns were removed
#The states, population, and crime data for 2017 was isolated
#Column names and certain state names were adjusted
FBICrime <- FBICrime %>%
  fill(Area) %>%
  filter(Year != "Percent change", Year != "2016", !is.na(Year)) %>%
  filter(!(Area %in% c("United States Total5, 6", "Puerto Rico", "Pacific",
                       "Mountain", "West", "West South Central", "East South Central",
                       "South Atlantic5,6", "South5,6", "West North Central", "East North Central",
                       "Midwest", "Middle Atlantic", "Northeast", "New England"))) %>%
  arrange(Area)

column_names <- make.names(c("State", "Population", "Violent Crime", "Murder", "Rape",
                             "Robbery", "Aggravated Assault", "Property Crime", "Burglary",
                             "Larceny", "Motor Vehicle Theft"))
FBICrime <- FBICrime %>%
  select(Area, Population2, starts_with("X"), -X22, -X23, -X24) %>%
  setNames(column_names)

FBICrime$State[FBICrime$State %in% "District of Columbia5"] <- c("District of Columbia")
FBICrime$State[FBICrime$State %in% "North Carolina6"] <- c("North Carolina")

#Gathered the crimes into an offense column and separated rate to use for further wrangling
#Converted rates and population to numeric values after removing commas
FBICrime <- FBICrime %>%
  gather("Offense", "Rate", Violent.Crime:Motor.Vehicle.Theft)
FBICrime$Rate <- gsub(",", "", FBICrime$Rate) %>% as.numeric(FBICrime$Rate)
FBICrime$Population <- gsub(",", "", FBICrime$Population) %>%
  as.numeric(FBICrime$Population)
FBICrime$State <- tolower(FBICrime$State)

#Filtered violent crime and property crime into a separate data frame
#Removed violent crime and property crime from the original dataset
#Spread out the offenses and rates in FBICrime to prepare the data for plotting and statistics
Violent_Crime <- filter(FBICrime, Offense == c("Violent.Crime"))
Property_Crime <- filter(FBICrime, Offense == c("Property.Crime"))
FBICrime2 <- bind_rows(Violent_Crime, Property_Crime) %>%
  arrange(State)

FBICrime <- FBICrime %>%
  spread(Offense, Rate) %>%
  select(-Violent.Crime, -Property.Crime)

#Created a separate dataset containing average total crime rates per state
#Spread out the offenses and rates in FBICrime2 to prepare the data for plotting and statistics
AvgStateRate <- FBICrime2 %>%
  group_by(State) %>%
```

```r
  summarise(Average.Crime.Rate = mean(Rate))
FBICrime2 <- FBICrime2 %>% spread(Offense, Rate)

#Appended the average crime rates with the poverty data and adjusted decimal places
CrimePoverty <- left_join(AvgStateRate, Poverty, by = "State")
is.num <- sapply(CrimePoverty, is.numeric)
CrimePoverty[is.num] <- lapply(CrimePoverty[is.num], round, 1)

#Saved the three datasets to csv files
write.csv(FBICrime, file = "CrimeByState&Offenses.csv")
write.csv(FBICrime2, file = "Violent&PropertyCrime.csv")
write.csv(CrimePoverty, file = "Crime&PovertyByState.csv")
```

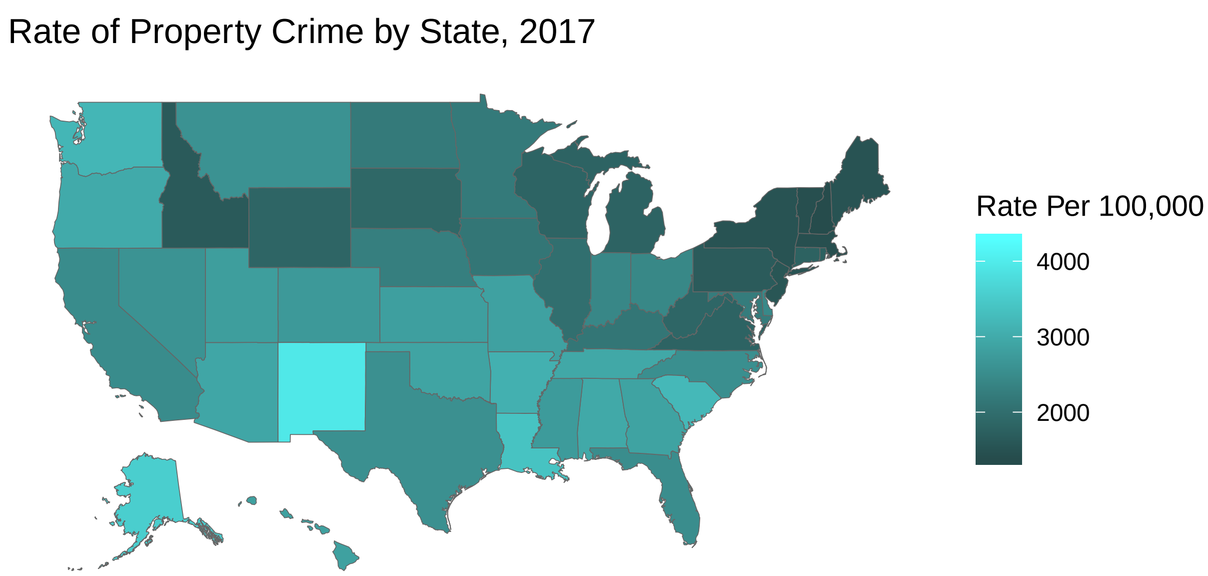
Exploratory Data Analysis and Statistics

Looking at these data sets, some of the things that can be explored are comparisons between the rates of different categories of crime (violent crime and property crime), as well as the rate of crime between states for each of the different criminal offenses. To be able to visually see which states are most affected, the rates were mapped onto the fifty states: the higher rates are in brighter shades of teal whereas the lower rates are in shades of green leading to black.

```r
library(fiftystater)

ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Violent.Crime, map_id = State),
           map = fifty_states, data = FBICrime2, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Violent Crime by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```




```r
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Property.Crime, map_id = State),
           map = fifty_states, data = FBICrime2, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Property Crime by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



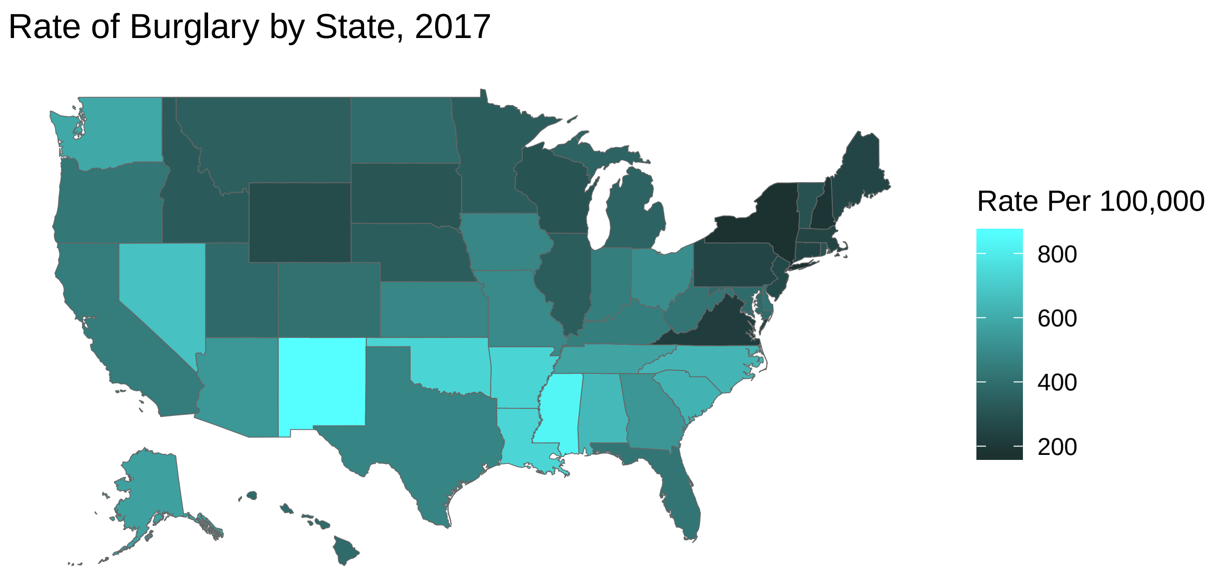
One of the trends that can be seen by looking at these maps is that there is a higher prevalence of property crime throughout the United States as compared to violent crime. While the scale of violent crime rates per 100,000 is from below 250 to over 1000, the scale for property crime rates per 100,000 is from below 2000 to over 4000.

These maps of generalized violent and property crime show that New Mexico and Alaska seem to have the highest rates of both violent and property crime. It can also be seen by the difference in scale and in the coloring that there are higher rates of property crime than violent and property crime is more highly widespread across the states than violent crime. Property crime seems to be most prevalent in the southern and western states, with the least amount in the northeast and parts of the midwest and mountain region, whereas violent crime is most prevalent in New Mexico, Alaska, and Tennessee, with mid to low rates across the rest of the states.

The following maps depict the distribution of the individual criminal offenses throughout the states. Glancing at the rate scales, it can be seen that larceny seems to have a highest prevalence with rates in the thousands.

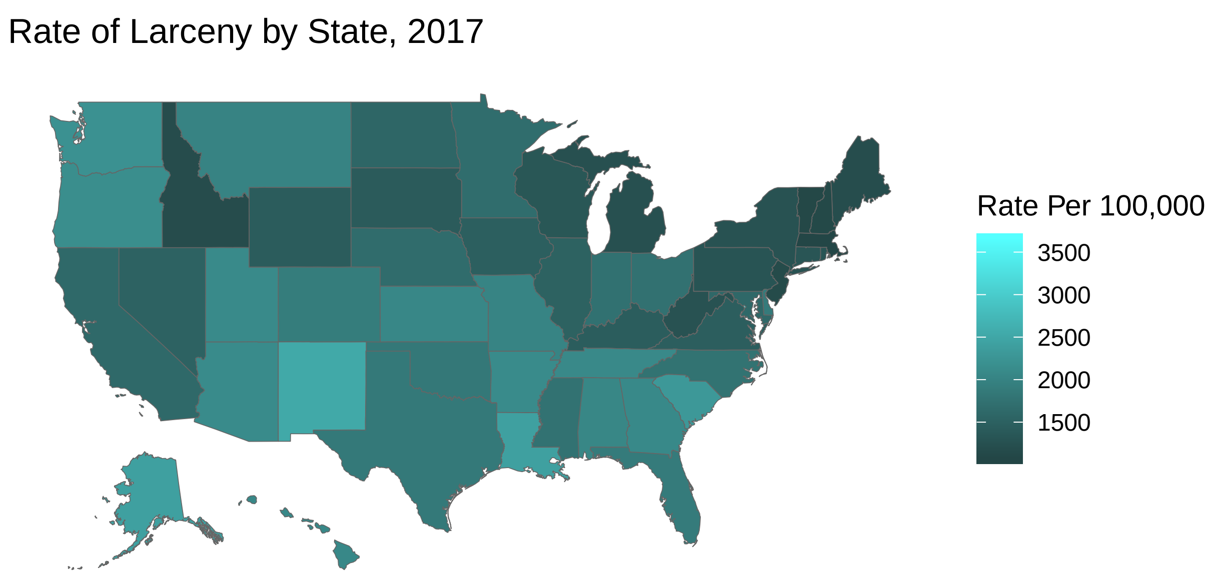
For property crimes:

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Burglary, map_id = State),
           map = fifty_states, data = FBICrime, size = 0.15, color = "grey40") +
 scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Burglary by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



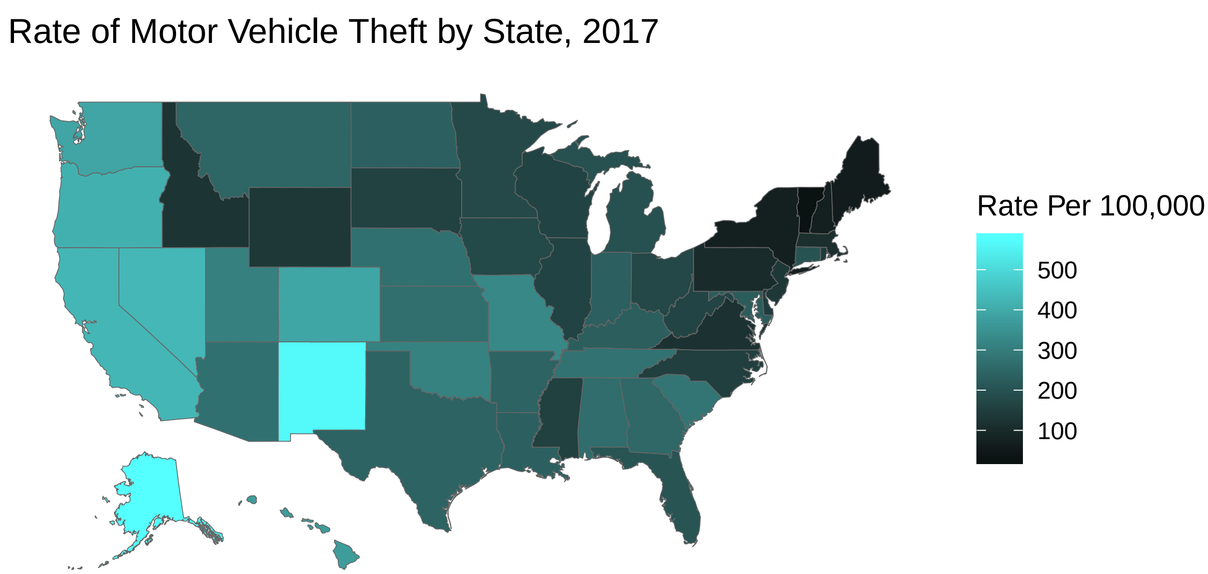
Burglary shows high rates throughout the south, ranging from 500 to over 800 cases per 100,000, with the most highlighted states being New Mexico and Mississippi, followed by states like Oklahoma, Arkansas, and Louisiana. The lowest rates of burglary (ranging from below 200 to 250 per 100,000) seem to be in New York, New Hampshire, and Virginia.

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Larceny, map_id = State),
           map = fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Larceny by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



Larceny seems to show highest rates (2500 to 3500 per 100,000) in New Mexico, Louisiana, South Dakota, and Alaska but is fairly widespread throughout the United States and seems to be the most prevalent of the individual criminal offenses reported to the FBI. The lowest rates (below 1500 per 100,000) seem to be found in parts of New England/northeast regions, parts of the midwest (such as Wisconsin and Michigan), and Idaho.

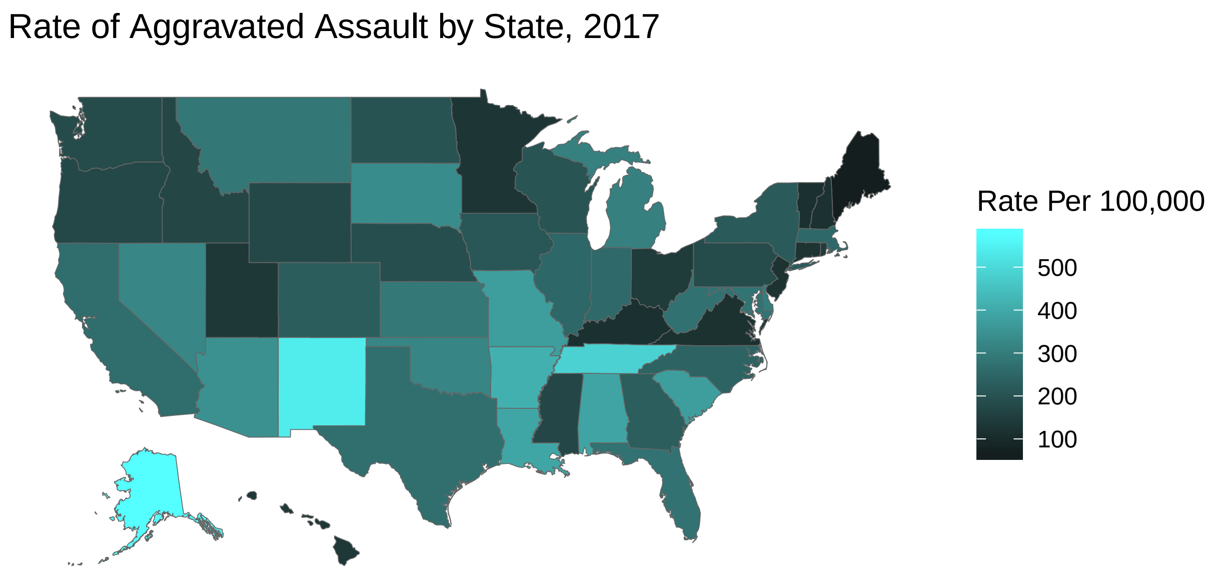
```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Motor.Vehicle.Theft, map_id = State),
           map =fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Motor Vehicle Theft by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



Motor vehicle theft seems to be most prevalent in New Mexico and Alaska (with rates over 500 per 100,000), followed by the west coast states (Washington, Oregon, California) in the 400 to 500 range, as well as Nevada and Colorado. The lowest rates (below 100) look to be in the New England/northeast states (Maine, New Hampshire, Vermont, New York, and Pennsylvania.

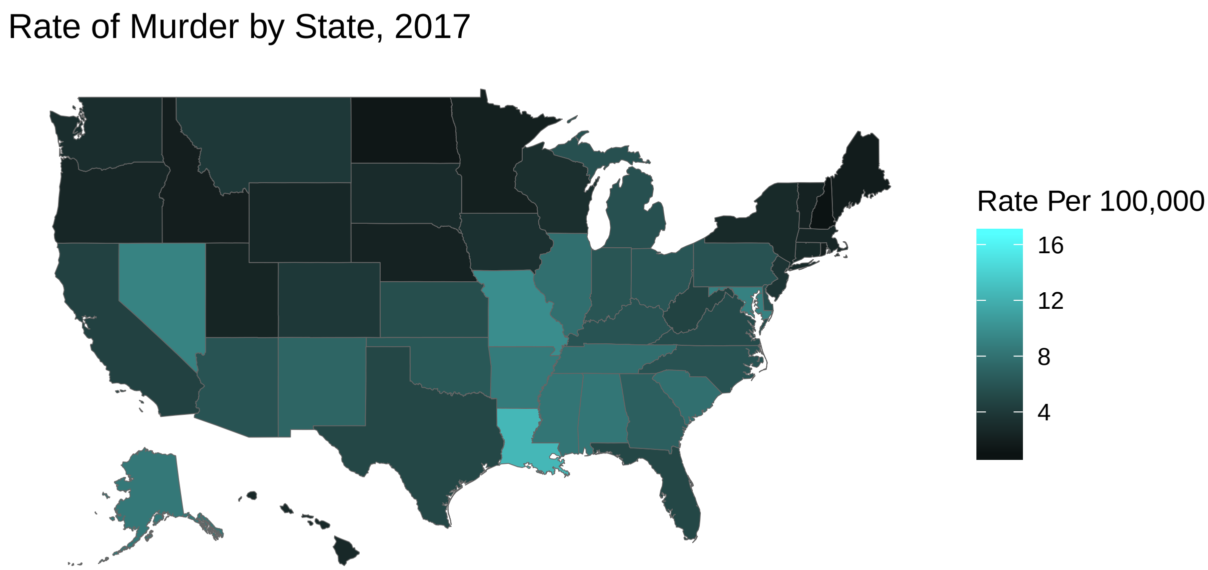
For violent crimes:

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Aggravated.Assault, map_id = State),
           map = fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Aggravated Assault by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



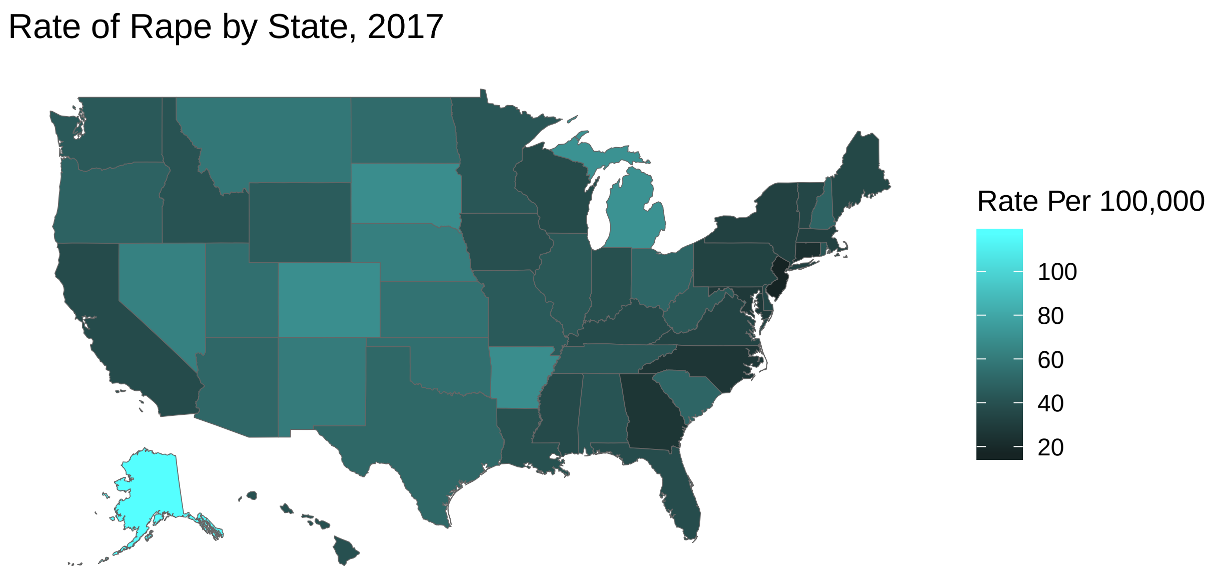
Aggravated assault seems to show its highest rates of over 500 cases per 100,000 in New Mexico, Alaska, and Tennessee, followed by 400 to 500 rates in Arkansas, Louisiana, Missouri, Alabama, South Carolina, and South Dakota. The lowest prevalence (below 100) seems to be in Maine, followed by 100 to 200 per 100,000 rates in New Hampshire, Vermont, Connecticut, New Jersey, Virginia, Kentucky, Ohio, Minnesota, and Utah.

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Murder, map_id = State),
           map = fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Murder by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



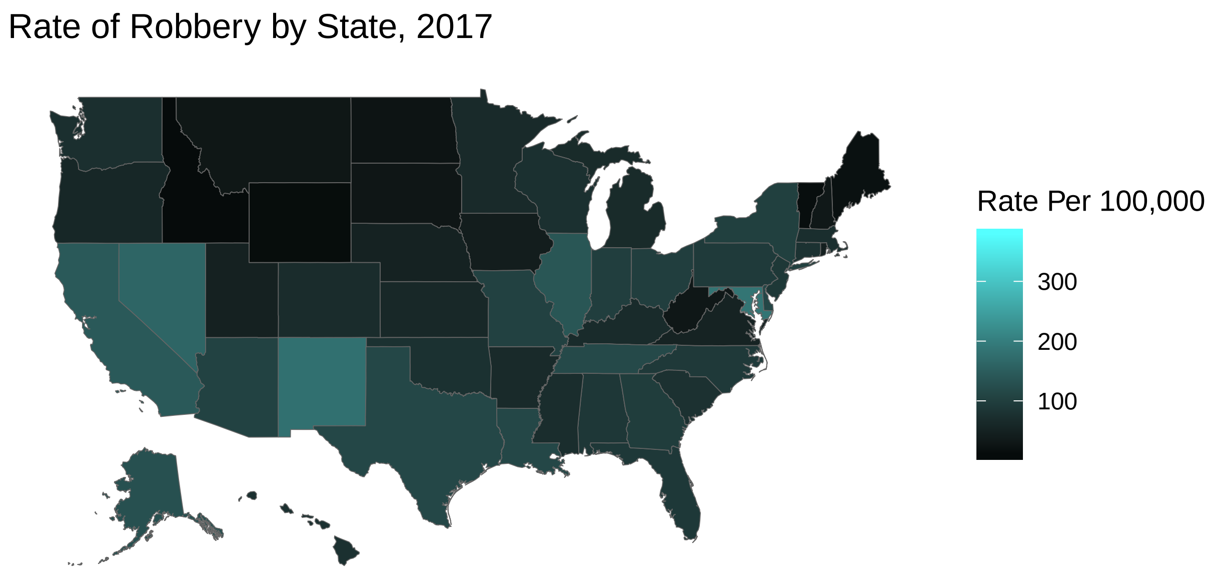
Murder seemes to be the least prevalent crime reported to the FBI with the lowest of crime rates at a range of 0 to over 16 cases per 100,000 people, showing mostly mid to low rates (from 0 to 12) throughout the United States. The higher murder rates are found in Louisiana, Nevada, Missouri, and Maryland. The lowest rates below 4 are seen in the New England/northeast region along with Hawaii, and states in the northern midwest and northwest such as Minnesota, North Dakota, Nebraska, Idaho, and Utah.

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Rape, map_id = State),
           map = fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
 scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Rape by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



Rape shows highest prevalence rating over 100 per 100,000 in Alaska. Most other states show middle range rates (between 40 to 90 per 100,000) except for Connecticut and New Jersey, which stand out as the lowest rated states below 20, followed by Georgia and North Carolina in the 20 to 40 range.

```
ggplot() +
  geom_map(aes(x = long, y = lat, map_id = id),
           data = fifty_states, map = fifty_states, size = 0.15) +
  geom_map(aes(fill = Robbery, map_id = State),
           map = fifty_states, data = FBICrime, color = "grey40", size = 0.15) +
  scale_fill_gradient2(name = "Rate Per 100,000", low = "black", mid = "black", high = "#55FFFF") +
  coord_map() +
  labs(x = NULL, y = NULL, title = "Rate of Robbery by State, 2017") +
  theme(axis.ticks = element_blank(), axis.text = element_blank(),
        panel.background = element_blank(), panel.border = element_blank())
```



Robbery is fairly low on its scale throughout the United States with rates below 150 except for states like Nevada, New Mexico, Maryland, and Illinois where you can probably find rates over 200.

Minimums for each of the criminal offenses:

```
FBICrime <- FBICrime %>%
  gather(Offense, Rate, Aggravated.Assault:Robbery) %>%
  group_by(Offense)

min.crime <- FBICrime %>%
  filter(Rate == min(Rate))
min.crime
```

```
## # A tibble: 7 x 4
## # Groups:   Offense [7]
##   State         Population Offense              Rate
##   <chr>              <dbl> <chr>               <dbl>
## 1 maine            1335907 Aggravated.Assault   65.3
## 2 new york        19849399 Burglary            176.
## 3 massachusetts    6859819 Larceny            1078
## 4 vermont           623657 Motor.Vehicle.Theft  31.1
## 5 new hampshire    1342795 Murder                1
## 6 new jersey       9005644 Rape                 16.7
## 7 idaho            1716943 Robbery              11.4
```

Maximums for each of the criminal offenses:

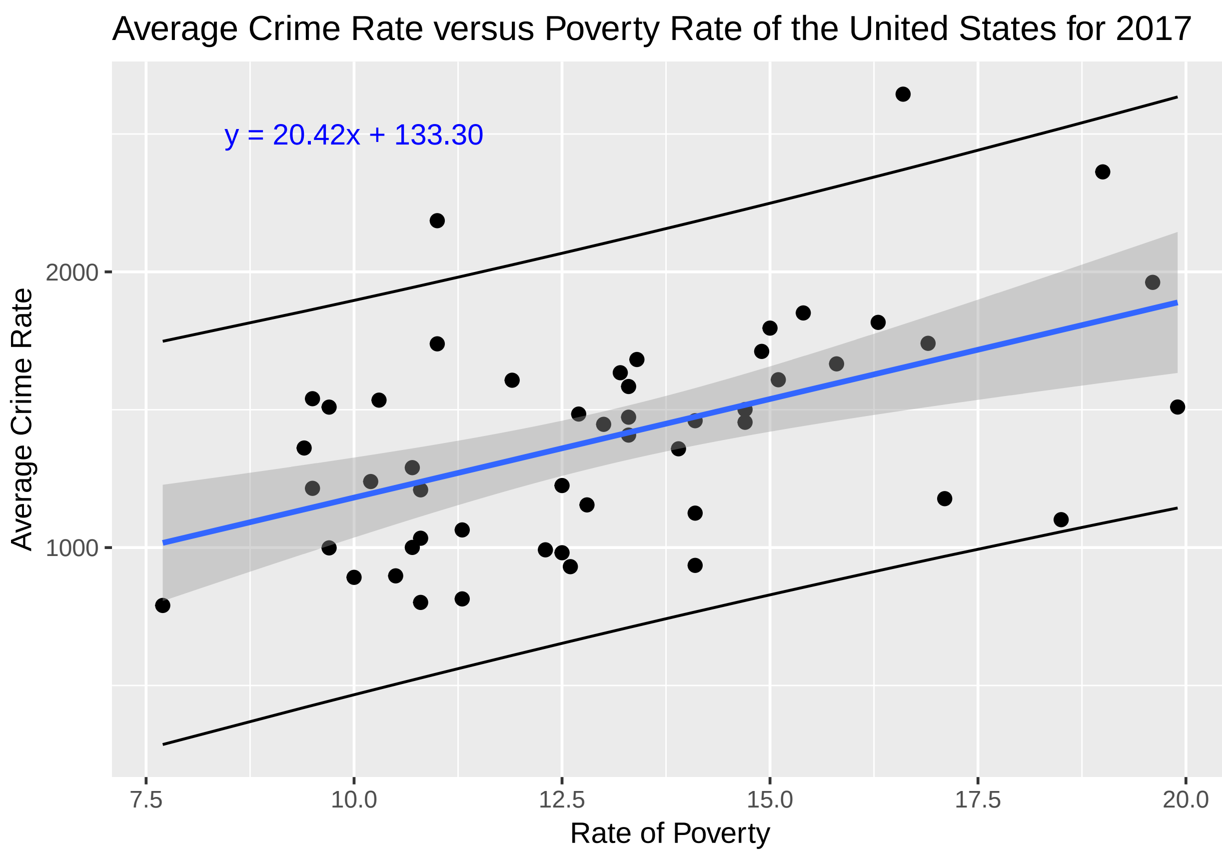
```r
max.crime <- FBICrime %>%
  filter(Rate == max(Rate))
max.crime
```

```
## # A tibble: 7 x 4
## # Groups:   Offense [7]
##   State                Population Offense                Rate
##   <chr>                     <dbl> <chr>                 <dbl>
## 1 alaska                   739795 Aggravated.Assault   575.
## 2 new mexico              2088070 Burglary             858.
## 3 district of columbia     693972 Larceny             3650.
## 4 alaska                   739795 Motor.Vehicle.Theft  576.
## 5 district of columbia     693972 Murder                16.7
## 6 alaska                   739795 Rape                 117.
## 7 district of columbia     693972 Robbery              378
```

Here we have the exact minimum and maximum rated states for each criminal offense. The minimums–as we were able to observe from looking at the maps–are mostly in the New England/northeast regions, except for Idaho having the lowest rate of robberies. With maximums–again as seen on the maps–Alaska and New Mexico showed the highest rates for many of the crimes, specifically aggravated assault, burglary, motor vehicle theft, and rape. District of Columbia was not really visible on the map so it is interesting to observe that it shows the highest rates of larceny, murder, and robbery.

The relationship between crime and poverty:

```r
dat <- data.frame(Poverty.Rate = CrimePoverty$Poverty.Rate,
                  Average.Crime.Rate = CrimePoverty$Average.Crime.Rate)
CrimePov.lm <- lm(data = dat, formula = Average.Crime.Rate ~ Poverty.Rate)
cppi <- cbind(CrimePoverty, predict(CrimePov.lm, interval = "prediction"))

p <- ggplot(cppi, aes(x = Poverty.Rate)) +
  geom_point(aes(y = Average.Crime.Rate), size = 2) +
  geom_smooth(aes(y = Average.Crime.Rate), method = lm, se = TRUE, na.rm = TRUE) +
  geom_line(data = cppi, aes(x = Poverty.Rate, y = upr)) +
  geom_line(aes(x = Poverty.Rate, y = lwr))

p + annotate("text", label = "y = 20.42x + 133.30", x = 10.0, y = 2500, size = 4, color = "blue") +
  labs(x = "Rate of Poverty", y = "Average Crime Rate",
       title = "Average Crime Rate versus Poverty Rate of the United States for 2017")
```



Correlation coefficient for crime vs. poverty:

```
cor(x = CrimePoverty$Poverty.Rate, y = CrimePoverty$Average.Crime.Rate)
```

```
## [1] 0.5095537
```

Linear regression model for crime vs. poverty:

```
summary(CrimePov.lm)
```

```
##
## Call:
## lm(formula = Average.Crime.Rate ~ Poverty.Rate, data = dat)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -687.70 -228.84   43.88  203.86  991.21
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    466.64     230.86   2.021 0.048729 *
## Poverty.Rate    71.48      17.24   4.145 0.000134 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 348.4 on 49 degrees of freedom
```

```
## Multiple R-squared:  0.2596, Adjusted R-squared:  0.2445
## F-statistic: 17.18 on 1 and 49 DF,  p-value: 0.000134
```

```
CrimePov.lm$residuals
```

```
##          1          2          3          4          5          6
##   66.167173  932.693536  179.825262  185.054600   55.691734  331.928868
##          7          8          9         10         11         12
## -160.983706   51.135447  991.210886  -14.291502   62.729453  394.312103
##         13         14         15         16         17         18
## -436.372935 -135.025031   -9.308266  -29.210655  289.762396 -511.428636
##         19         20         21         22         23         24
##   94.273752 -460.350177  222.660008 -319.666941 -349.491502   69.212103
##         25         26         27         28         29         30
## -379.269961  257.643829  109.879160   58.537250  166.891734 -226.825616
##         31         32         33         34         35         36
## -289.327419  537.861179 -538.991502  -62.878929   43.876772 -101.895693
##         37         38         39         40         41         42
##   70.294122  224.039638 -378.725031 -353.929222  283.585740 -226.668744
##         43         44         45         46         47         48
##  257.177358  -16.578929  349.516294 -437.410655 -230.862750  486.093536
##         49         50         51
## -687.699299 -210.250177 -204.610655
```

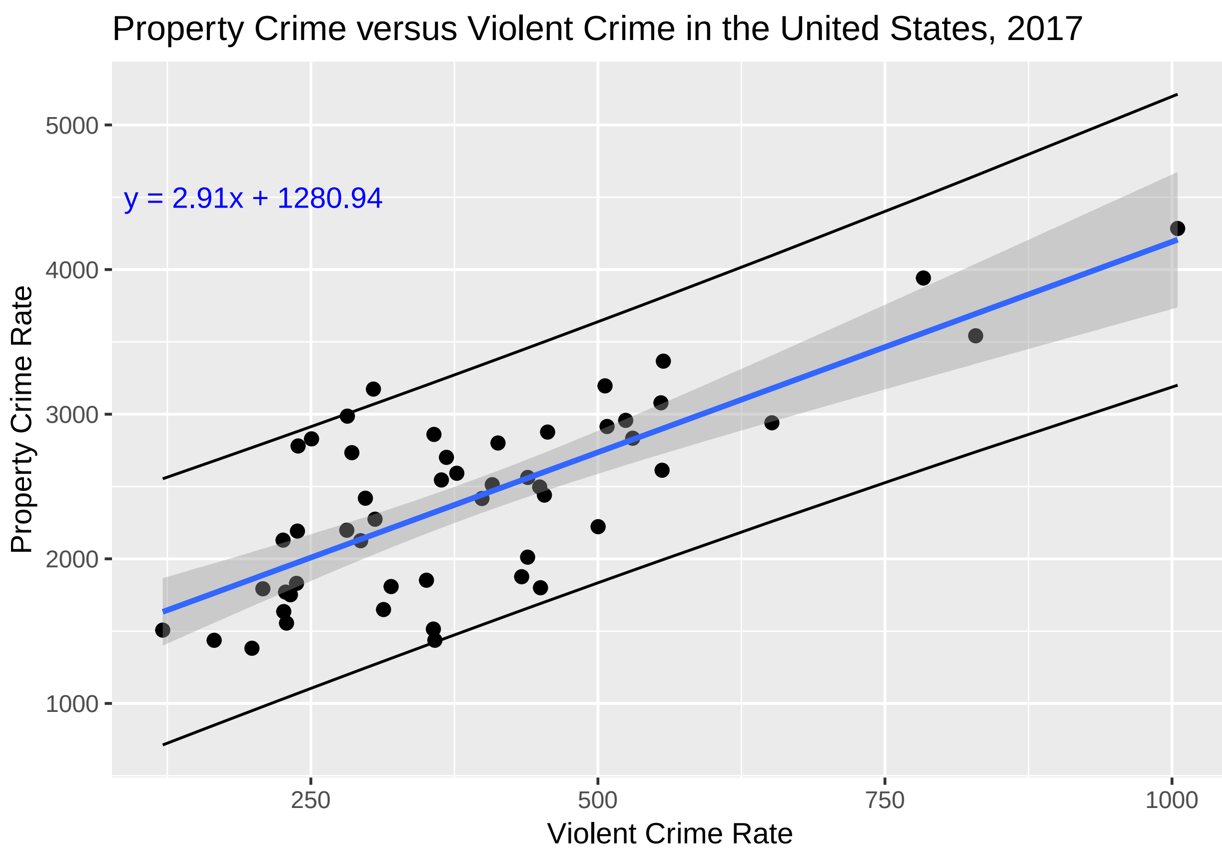
```
CrimePov.SSE <- sum(CrimePov.lm$residuals^2)
sqrt(CrimePov.SSE/51)
```

```
## [1] 341.518
```

The graph shows a positive relationship between the rate of poverty and average crime rate in the United States, so as the rate of poverty increases, the average crime rate appears to increase as well. On the graph, a 95% confidence interval and prediction interval are shown. The confidence interval shows the range within which there is a 95% probability that the actual observed average crime rate will be found. The prediction interval shows the estimated range within which it is probable that future predicted average crime rates will be found, based on the actual data observed. The correlation coefficient is a positive 0.510, which indicates that poverty rate and average crime rate have a moderately positive correlation, but it is not close to a perfect positive relationship. The R squared value is 0.2596, which is fairly low, and the root mean squared errors value is 341.518.

The relationship between violent and property crime:

```
dat1 <- data.frame(Violent.Crime = FBICrime2$Violent.Crime,
                   Property.Crime = FBICrime2$Property.Crime)
VPCrime.lm <- lm(data = dat1, formula = Property.Crime ~ Violent.Crime)
cppi1 <- cbind(FBICrime2, predict(VPCrime.lm, interval = "prediction"))


vp <- ggplot(cppi1, aes(x = Violent.Crime)) +
  geom_point(aes(y = Property.Crime), size = 2) +
  geom_smooth(aes(y = Property.Crime), method = lm, se = TRUE, na.rm = TRUE) +
  geom_line(data = cppi1, aes(x = Violent.Crime, y = upr)) +
  geom_line(aes(x = Violent.Crime, y = lwr))

vp + annotate("text", label = "y = 2.91x + 1280.94", x = 200, y = 4500, size = 4, color = "blue") +
  labs(x = "Violent Crime Rate", y = "Property Crime Rate",
       title = "Property Crime versus Violent Crime in the United States, 2017")
```



Correlation coefficient for property crime vs. violent crime:

```
cor(x = FBICrime2$Violent.Crime, y = FBICrime2$Property.Crime)
```

```
## [1] 0.7514964
```

Linear regression model for property crime vs. violent crime:

```
summary(VPCrime.lm)
```

```
##
## Call:
## lm(formula = Property.Crime ~ Violent.Crime, data = dat1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -886.13 -295.74    3.95  242.11 1006.21
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1280.9363   155.1236   8.258 7.74e-11 ***
## Violent.Crime    2.9112     0.3651   7.974 2.10e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 443 on 49 degrees of freedom
```

```
## Multiple R-squared:  0.5647, Adjusted R-squared:  0.5559
## F-statistic: 63.58 on 1 and 49 DF,  p-value: 2.097e-10
```

```
VPCrime.lm$residuals
```

```
##           1           2           3           4           5           6
##  150.328806 -152.195590  155.089748  182.255909  -92.224616  349.062330
##           7           8           9          10          11          12
## -174.782854 -160.260410   77.529734   43.706676  539.394075  819.024720
##          13          14          15          16          17          18
## -304.624983 -546.957338  -25.592801   -9.773325  317.650829  190.821718
##          19          20          21          22          23          24
##  464.342453 -126.087741 -514.803132 -886.134860 -790.962434  216.832102
##          25          26          27          28          29          30
##  621.242678    9.170673  212.861806  102.537059 -286.855261 -477.585594
##          31          32          33          34          35          36
## -391.511790  379.862612 -805.150340  205.571475   97.951823  272.090881
##          37          38          39          40          41          42
##  267.388316  885.196239 -543.605594 -205.309765  441.329853 -667.019258
##          43          44          45          46          47          48
## -236.963044    3.951545  803.785401 -326.908125  -94.141703 1006.212696
##          49          50          51
## -449.883325 -403.919311 -141.938962
```

```
VPCrime.SSE <- sum(VPCrime.lm$residuals^2)
sqrt(VPCrime.SSE/51)
```

```
## [1] 434.2271
```

Again, with violent crime and property crime, we see a positive relationship showing that there does tend to be higher property crime wherever there is higher violent crime. Here the correlation coefficient is 0.7515, which indicates a fairly strong correlation. We see that the relationship is stronger than the one we saw between poverty and average crime. The R squared value is 0.5647 and the root mean squared errors value is 434.227.

Conclusion

Through this project, we were able to look at the distributions of crime throughout the United States on the state level and determine which states could be considered "safer" and which "more dangerous", based on these distributions. For cautious families looking for a safer state to settle down in for the growing years of their children, they may want to veer away from states that currently have higher rates of violent crime. As we saw, there is a high prevalence of both violent and property crime in New Mexico, Alaska, and Washington D.C. For people owning homes and vehicles, it may be of value for them to know that these states, as well as the west coast and south, have a higher spread of property crimes–being aware would allow people to invest in cautious measures of protection, like a home security system.

For law enforcement branches and state governments, this project provides information on how their state compares to others, and what their problem areas are. New Mexico showed the highest rate of burglary but also one of the highest for motor vehicle theft, aggravated assault, and larceny. Knowing where the problem specifically lies allows for programs and measures to be put in place. Knowing the relationship between violent and property crime and also crime and poverty allows for an understanding of how best to approach the problems on a larger scale looking forward. Figuring out what drives people to commit crimes is an important part of preventing crimes by not allowing others in similar situations to follow the same path.

Possibilities for Further Investigation

It would be interesting to delve deeper into the data provided by the FBI and look further into the crime rates within the states to observe metropolitan versus non-metropolitan crime and poverty. Since there is some positive correlation on an overall state scale between crime rates and poverty, it would be interesting to look at whether or not this continues to show at a county or city level. Are the poverty rates higher in urban areas or rural? And do higher rates of crime follow wherever higher poverty rates are found within these specifically defined areas?

For further study, it would also be interesting to include other years of crime data and build a time series plot to look at how levels of crime has changed over the years at a national level but also at a state level, observing the increase or decrease of crime over time for all the states. This would also help to observe the trend of crime over the years to allow for making predictions of where crime rates are heading in the future and whether or not certain actions being taken are making a difference in bringing rates down.