

Linear Regression Project Part 1: Violent Crime

4th Side

Pledge

Please type your names in the appropriate space below. Failing to do so will result in a 0 on this assignment.

“We have neither given nor received unauthorized help on this assignment”

- Member 1: Cooper Carlson
- Member 2: Ryan Erhardt
- Member 3: Nick Steinhilber
- Member 4: Grant Zhou

Introduction

This semester, our team has decided to conduct a research project to find any variables that seem to have a relationship with crime rates in the United States. According to Statista, the US has seen an unmatched violent crime rate over the past coupled decades, specifically with respect to homicides, when compared with rates in other countries. The United States has the highest crime rate of high-income countries by a wide margin. According to the Institute for Health Metrics and Evaluation, “Age-adjusted firearm homicide rates in the US are 19 times greater than they are in France, and 77 times greater than in Germany. The US has 33 times the rate of firearm homicide seen in Australia” (Sirull, 2023). Most Americans would agree that violent crime is a prevalent issue throughout the country, and the United States seems to be a clear outlier when compared with other developed countries on this basis. We believe that the first step to solving this problem is identifying some possible causes, and that is what we seek to outline throughout our project. Before our team chose its explanatory variables, we found a few articles that suggested potential relationships between certain variables and violent crime rates that guided the direction of our project. One of the most captivating articles was research conducted by the Harvard School of Public Health into the relationship between gun ownership and homicide. Researcher David Hemenway and his team found that “where there are more guns there is more homicide” across both high-income nations and US states (Hepburn & Hemenway, 2004, 417). Our project will outline the relationship with more detail and provide the significance of gun rates in violent crime rates when compared to other possible predictors of violent crime rates in the United States. Here are 3 of the key questions that we will answer throughout this project:

- Do states that legally allow capital punishment have a lower rate of violent crime per capita?
- Do states with more police funding (in dollars per capita) have a lower rate of violent crime per capita?
- Do states with higher gun ownership rates have a higher rate of violent crime per capita?

In addition to these questions, we will also be exploring the relationship between our response variable of violent crime per capita and numerous other potential explanatory variables including: median income, poverty rate, state political affiliation, the legality of abortion, and the decriminalization of marijuana

In a recent poll, nearly 2/3 of Americans said they believe crime is a serious problem (Smith, 2024), and by identifying variables we want to create meaningful inferences that can be put

to action. It is important to note that this is an observational study and not an experiment, so we cannot draw any causation from the relationships we discover in this project. Still, noting strong relationships could be helpful to identify areas for future research.

Data Summary

Our data has been compiled in a table with each state observation and its respective data being captured in a single row. Each row has the following values in its columns in the following order: State name, presence of the death penalty (Y, N, or M), state political affiliation (R or D), median income (in USD), decriminalization of marijuana (Y or N), poverty rate (as a % of people in the state), gun ownership rate (percentage of population), average police funding (in USD per capita), and legality of abortion (Y or N). We are going to give a short background summary for each variable.

- Our response variable is violent crime, and is considered to be quantitative. This data was gathered from the Federal Bureau of Investigation in 2022. Taken from 15,726 of 18,888 participating law enforcement agencies, the data shows the number of reported offenses per 100,000 of the population from every state. Because not every agency submitted information, the data is not entirely accurate to what actually occurred, but it gives us a more than reasonable estimate.
- Our first explanatory variable is the presence of the death penalty, and is considered to be qualitative. This data was gathered from the Death Penalty Information Center and is current as of 2023. Information on the website was taken from each state's current legal standing on the death penalty.
- Our second explanatory variable is average police funding, and is considered to be quantitative. This data was gathered from MoneyGeek, where they analyzed police and corrections spending data for each state and is current as of 2024.
- Our third explanatory variable is median household income, and is considered to be quantitative. This data was gathered from the US Census Bureau in 2022. Coming from the United States Census, we can say that the data is fairly accurate and representative of the population.
- Our fourth explanatory variable is poverty rate, and is considered to be quantitative. This data was gathered from the U.S. Department of Agriculture under the Economic Research Service, and is current as of 2021.

- Our fifth explanatory variable is gun ownership, and is considered to be quantitative. This data was gathered from World Population Review and is current as of 2024. Because of variation in regulations throughout the nation, it is impossible to get exact numbers for the number of guns in each state.
- Our sixth explanatory variable is state political party affiliation, and is considered to be qualitative. We chose to base each state's party affiliation on the party of the governor. This data was gathered from UVA's Center For Politics and is current as of 2023, taking the results from the most recent gubernatorial races.
- Our seventh explanatory variable is the legality of abortion, and is considered to be qualitative. This data was gathered from the Center for Reproductive Rights, and is current as of 2022. Information was gathered from each state's current legal standing on abortion.
- Our eighth and last explanatory variable is the legality of marijuana and is considered to be qualitative. This data was gathered from DISA Global Solutions, and is current as of 2024. Information was gathered from each state's standing on its legality; we are considering marijuana legal in a state if it is decriminalized and allowed to be sold.

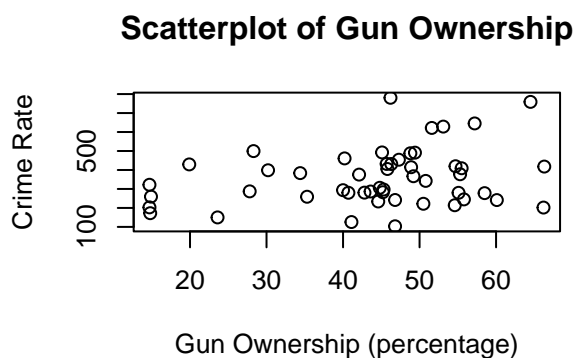
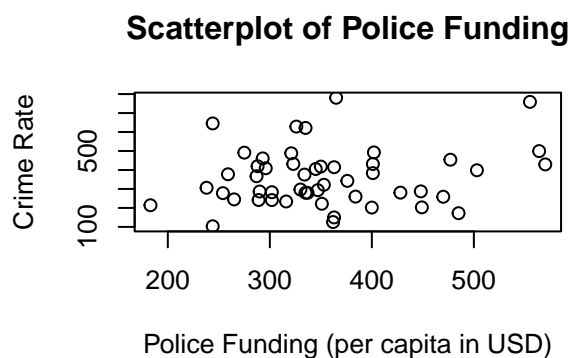
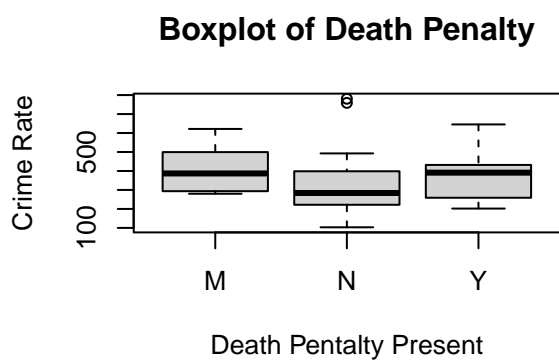
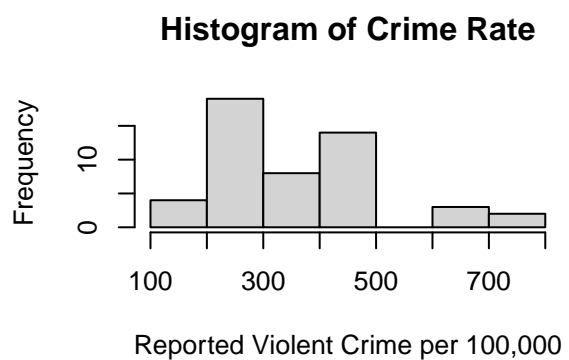
Throughout our data, we made some manipulations in order to facilitate the process. The main manipulation we made was excluding the District of Columbia from some of the gathered data sets, in order to keep our data consistent with 50 observations for every explanatory variable. Additionally, some of the data sets came in different order, such as being alphabetical or in terms of highest value to lowest value. We reordered every data set so that they the states are in order alphabetically.

One general potential issue we have found within the data is the variation in the times of which it was gathered. All of our data was gathered within the last three years, so it is very recent. However, there is variation between the years they were gathered and not all of the data correlates to the same point in time. Due to the contemporary nature of some of these issues (such as legalization of weed or abortion), drastic changes can occur from year to year. As all of the data is not from the exact same year, there might be some discrepancy and error present.

Overall, our sources are generally very trustworthy. Most of our data came from governmental institutions such as the FBI, Consensus Bureau, or Department of Agriculture. The rest of our data came from third party organizations, but they just compiled information from each state's standing or used verifiable sources like the Pew Research Center, so the validity of the data can be trusted.

EDA

```
Crime_Compiled<-read.csv("https://raw.githubusercontent.com/RyanErhardt/Crime_Compiled/main/crime_data.csv")
par(mfrow=c(2,2))
hist(Crime_Compiled$Crime_Rate, xlab="Reported Violent Crime per 100,000", main="Histogram of Crime Rate")
boxplot(Crime_Rate~Death_Penalty_Present, data=Crime_Compiled, main="Boxplot of Death Penalty")
plot(Crime_Rate~Police_Funding,data=Crime_Compiled, main="Scatterplot of Police Funding")
plot(Crime_Rate~Gun_Ownership, data=Crime_Compiled, main="Scatterplot of Gun Ownership",
```



```
round(cor(Crime_Compiled[,c(5,6)],Crime_Compiled$Crime_Rate),3)
```

```
##           [,1]
## Police_Funding 0.177
## Gun_Ownership  0.273
```

Conclusions

From the box plot of death penalty presence it does seem like there is a difference in average crime rate between states that have the death penalty, that don't have it, and ones where the death penalty is paused. The average crime rate for paused and yes is around 400 while the average for no is 300. There are two outliers present in the data for no death penalty which might be causing the difference in means and this will be analyzed further during regression. Answering the first research question on the death penalty based on our preliminary data analysis it appears that no, states with the death penalty do not have lower rates of crime, in fact it appears to be the opposite.

From the scatter plot of crime rate and police funding there does not immediately appear to be a relationship present and if there is, it is most likely fairly weak. The correlation value of .177 is quite low, suggesting the relationship is very weak and positive, may not be linear, or may not be present at all. Answering the second research question on if states with higher police spending have lower rates of crime, no it does appear so. In fact, the positive correlation value suggests that higher police spending per capita might be associated with higher crime rates. There is most likely no relationship since the correlation is very low, but this will be analyzed further during regression.

From the scatter plot of gun ownership rates and crime rate it appears that there is not a very strong relationship present. The correlation value of .273 is low, suggesting the relationship is fairly weak and positive, may not be linear, or may not exist. Answering the third research question of whether states with higher gun ownership rates have higher crime rates, no it does not seem to be the case. It seems like there is most likely no relationship since the correlation value is very low.

Crime rate is currently not very suitable to be the response variable in regression. There is a gap in the data from 500-600 that creates an island of higher value data points. The histogram is also not unimodal or very symmetric which also creates an obstacle to regression. This will likely need to be fixed by transforming the crime rate response variable.

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Violent Crime	Crime rate	Violent crime is composed of four offenses: murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault. These are all crimes that involve force or threat of force. The rate is the number of reported offenses per 100,000 people in the population of the state. For example, a crime rate of 400 in a state with a population of 1,000,000 means that 4,000 violent crimes were reported.
Presence of the Death Penalty	Death penalty	The death penalty is the highest punishment administratable by courts and means that the convicted will be put to death. Not all states allow this punishment, so there are two levels of this binary variable: 0 for no death penalty, 1 for death penalty.
Average police funding	Police funding	Police funding measures the amount of money a states spends on police departments per capita(person) in dollars.

Variable Name	Abbreviated Name	Description
Median household income	Household income	The median household income is the 50th percentile of the household incomes for households in the United States, calculated using \$2,500 intervals. Household income measures the amount of money that the providers of a household bring in.
Poverty rate	Poverty rate	The percentage of people living below the poverty line in each state. The poverty line depends on the number of individuals in a household, beginning at \$14,580/year for individuals
Gun ownership rate	Gun ownership rate	The percentage of people who own guns in each state.
State political affiliation	State political affiliation	A state's primary political affiliation, based on the party of the governor. The levels of the qualitative variable are 0 for democrat and 1 for republican
The legality of abortion	Legality of abortion	Whether abortion is legal in a state or not. Abortion is illegal in a state if it is completely banned. The levels of the qualitative variable are 0 for not legal and 1 for legal

Variable Name	Abbreviated Name	Description
The legality of marijuana	Legality of marijuana	Whether marijuana is legal in a state or not. Marijuana is considered legal in a state if it is decriminalized and allowed to be sold. The levels of the qualitative variable are 0 for not legal and 1 for legal

Appendix B: Data Rows

```
head(Crime_Compiled)
```

##	State	Crime_Rate	Governor_Party	Death_Penalty_Present	Police_Funding
## 1	AL	409.1	R	Y	296
## 2	AK	758.9	R	N	555
## 3	AZ	431.5	R	M	401
## 4	AR	645.3	R	Y	244
## 5	CA	499.5	D	M	564
## 6	CO	492.5	D	N	402

##	Gun_Ownership	Abortion_Legality	Marijuana_Legality	Median_Income
## 1	55.5	No	No	59910
## 2	64.5	Yes	Yes	89740
## 3	46.3	Yes	Yes	73450
## 4	57.2	No	No	53980
## 5	28.3	Yes	Yes	85300
## 6	45.1	Yes	Yes	89930

Appendix C: References

Background

References Hepburn, L., & Hemenway, D. (2004). Firearm availability and homicide: A review of the literature. *Aggression and Violent Behavior: A Review Journal*, 9(417), 40. <https://www.hsph.harvard.edu/hicrc/firearms-research/guns-and-death/>

Sirull, R. (2023, October 31). On gun violence, the United States is an outlier. Institute for Health Metrics and Evaluation. Retrieved March 28, 2024, from <https://www.healthdata.org/news-events/insights-blog/acting-data/gun-violence-united-states-outlier>

Smith, C. (2024, January 11). Why It's Confusing to Know Whether Crime's Really Up or Down. Why It's Confusing to Know Whether Crime's Really Up or Down. Retrieved March 28, 2024, from <https://www.governing.com/urban/why-its-confusing-to-know-whether-crimes-really-up-or-down>

Data

CBS News. (April 14, 2022). Gun Ownership by State 2022 [Graph]. In *World Population Review*. Retrieved March 28, 2024, from <https://worldpopulationreview.com/state-rankings/gun-ownership-by-state>

Center for Politics. (November 9, 2023). Party control of governorships following 2023 elections [Graph]. Retrieved March 28, 2024, from <https://centerforpolitics.org/crystalball/2024-governor/>

Center for Reproductive Rights. After Roe Fell: Abortion Laws by State [Graph]. Retrieved March 28, 2024, from <https://reproductiverights.org/maps/abortion-laws-by-state/>

Death Penalty Information Center. States with and without the death penalty – 2023 [Graph]. Retrieved March 28, 2024, from <https://deathpenaltyinfo.org/states-landing>

Defense Information System Agency. (March 1, 2024). MARIJUANA LEGALITY BY STATE - Updated March 1, 2024 [Graph]. Retrieved March 28, 2024 from <https://disa.com/marijuana-legality-by-state>

FBI. (October 16, 2023). Reported violent crime rate in the United States in 2022, by state (per 100,000 of the population) [Graph]. In *Statista*. Retrieved March 28, 2024, from <https://www.statista.com/statistics/200445/reported-violent-crime-rate-in-the-us-states/>

US Census Bureau. (September 13, 2023). Median household income in the United States in 2022, by state (in 2022 U.S. dollars) [Graph]. In Statista. Retrieved March 28, 2024, from <https://www.statista.com/statistics/233170/median-household-income-in-the-united-states-by-state/>

US Census Bureau. (October 18, 2022). Police and Corrections Spending by State [Graph]. In MoneyGeek. Retrieved March 28, 2024, from <https://www.moneygeek.com/living/state-policing-corrections-spending/#detailed-findings>

US Department of Agriculture. (June 16, 2023). Percent of total population in poverty, 2021 [Graph]. Retrieved March 28, 2024, from <https://data.ers.usda.gov/reports.aspx?ID=17826>