State Incarceration Rates and Political Affiliations
Olivia, Benny, Anna, Macy, and Jacob

1. Introduction

Crime and incarceration rates in the United States are a hot topic in political conversation. With that being said, knowing if there is a correlation between state political affiliation and crime rates in the United States, across all 50 states, could affect the way in which people decide to vote, or where they decide to live.

In this project, we will use the incarceration rates of each state over 2015-2017 and compare it with the political makeup of each state to examine how much politics affect crime rates, if any effect at all.

2. Data[1]

In this project, we are using 2 different sources of data. Our first source is from Gallup to collect the years, states, and which way the state leans in terms of democratic or republican for that given year. The second source is from Kaggle and includes crime rates of different types of crimes of each state over the given year.

2.1 Political Makeup of State by Year

In order to find the political makeup of each state over 3 three years, we had to scrape data from the Gallup website which provided tables that included the state, year, and different numbers to show which way a state leaned politically. Each table was on a different webpage within Gallup, so we needed to do two different scrapes for each table. Additionally, one of the tables (year 2015) we had to import into excel and read into R, because it was a picture that we were unable to scrape.

For the scraping portion, we had to scrape each of the 6 columns as a separate line of code to grab from the table. The first table of the year 2017, we had to add a year column to indicate this was the 2017 column and then add the 7 columns from the table we scraped to a data frame. We repeated these steps for the table with the year 2016 and made a different data frame. We renamed the columns in the 2016 scrape to match those in 2017 exactly, in order to bind them together. We then created one overall data frame of the two separate scrapes.

Lastly, we read the data from excel from the 2015 table into r and vertically integrated it with the data frame of the two scrapes. The scraped data along with the table imported from excel created a data frame with 152 observations and 7 variables. This can be found in our r script: scraping_finar.r.

---

[1]https://www.kaggle.com/christophercorrea/prisoners-and-crime-in-united-states?select=ucr_by_state.csv
https://news.gallup.com/poll/226643/2016-party-affiliation-state.aspx
https://news.gallup.com/poll/203117/gop-maintains-edge-state-party-affiliation-2016.aspx
https://news.gallup.com/poll/188969/red-states-outnumber-blue-first-time-gallup-tracking.aspx

2.2 Crime and Incarceration Rates in the United States, by State and Year

Next, we had to read our Kaggle dataset into r for our incarceration rate data. We made sure that the headers matched up to those of the scraped data. We made the states with two names into a one name feature because it would not match up otherwise.

We then horizontally merged the scraped data frame and the Kaggle data frame into one data frame called *all* and made sure "all.X = True" was included in the merge. Our final data frame has 149 observations (NY is not included in 2015) of 20 variables (becomes 26 variables after the analysis) and can be found in the r script: Integration_final.r.

Table 1 - Data Dictionary

Column	Type	Description
Jurisdiction1	Text	state
Year1	timestamp	year
crime_reporting_change	binary	Indicates whether state changed reporting systems affecting comparisons with previous years
crimes_estimated	binary	Indicates whether some portion of state's crimes were estimated
State_population	numeric	Population of each state
violent_crime_total	numeric	Total violent crime
murder_manslaughter	numeric	Total number of murders or manslaughter
rape_legacy	numeric	total rape (using legacy definition of rape)
rape_revised	numeric	total rape (using revised definition of rape)
robbery	numeric	total robbery
agg_assult	numeric	Total aggravated assault
property_crime_total	numeric	Total number of property crimes
burglary	numeric	Total number of burglaries
larceny	numeric	Total number of larceny
vehicle_theft	numeric	Total number of vehicle theft
Democrat_Lean1	numeric	How much the state leans democratically

Republican_Lean1	numeric	How much the state leans republican
Democratic_Advantage1	numeric	By how much more does the state lean democratic
N1	numeric	Sample size used to calculate the leaning
Classification1	text	What the state is officially classified as for democrat/republican

3. Analysis

The goal of this project is to look at the correlation between incarceration and crime rates in the United States, by the state and year, and the classification of the state, whether republican or democratic.

3.1.1 Highest Crime/Incarceration Rates in States

Which States have the highest murders, violent crimes, and property crimes in the United States?

To answer this question, we started by creating separate columns in the data frame, that included the per capita calculations of murders, violent crimes, and property crimes to show an accurate representation of the states, otherwise the states with the highest population would most often have the highest percentages.

We then created a data frame for each of the 3 different crime categories and sorted them from least to greatest. We pulled the top and bottom of each data frame to get the highest and lowest per capita calculations in the United States of each crime category.

Tables 1, 2, and 3 show the percentages of each type of crime per capita, with the highest state, and lowest state as well as the corresponding year.

Table 2 - Murders

Jurisdiction	Murder Per 100,000 People	Year
Louisiana - Highest	12.424394	2017
New Hampshire - Lowest	1.423205	2017

Table 3 - Violent Crimes

Jurisdiction	Violent Crime Per 100,000 People	Year
Alaska - Highest	829.0134	2017
Vermont - Lowest	118.0432	2015

Table 4 - Property Crimes

Jurisdiction	Property Crime per 100,000 People	Year
New Mexico - Highest	3941.726	2017
New Hampshire - Lowest	1381.819	2017

After creating 3 different data frames that relate to murders, property crimes, and violent crimes, we found that when it came to violent crimes in the United States (Table 3), that Alaska has the highest rate at 829 people per 100,000, whereas Vermont has the lowest rate at 118 people per 100,000 committing violent crimes. Alaska came as a surprise as it is a very low populated state and doesn't strike you as a violent area. Vermont did not shock us as it seems like a very safe place to live and there doesn't seem to be many crimes there in general.

In Table 4, we found that the highest rate of property crime comes from New Mexico at 3941 people per 100,000. These property crimes include burglary, larceny, and vehicle theft. We found that the lowest per capita rate was at 1382 per 100,000 people in New Hampshire and is almost ⅓ the rate of the top state, New Mexico. We were not surprised that property crimes had a much higher number of people committing these, as property crime includes theft, burglary, larceny, and car damage as well.

Finally, looking deeper into violent crimes, at murders in particular, we found that Louisiana has the highest rate at 12.42 per 100,000 people and New Hampshire has the lowest at 1.42 people per 100,000. We thought that the violent crime and murders would have corresponding highest and lowest per capita in regards to the states, but they did not.

3.1.2 Summary table of changes in political affiliation each year of those corresponding states

Do the highest and lowest states in their respective crime categories show any patterns in their classification each year? AND **Out of the states with the highest and lowest crime rates, do any change classifications each year?**

We created a dplyr summary table of the 5 states in tables 2, 3, and 4 to analyze changes in politics over the years with high and low incarceration rates of different crime categories.

Table 5 - Political Affiliations by State

Jurisdiction	Classification	Year
Louisiana	Competitive	2015
Louisiana	Competitive	2016
Louisiana	Competitive	2017

New Hampshire	Lean Rep	2015
New Hampshire	Competitive	2016
New Hampshire	Competitive	2017
Alaska	Solid Rep	2015
Alaska	Lean Rep	2016
Alaska	Solid Rep	2017
Vermont	Solid Dem	2015
Vermont	Solid Dem	2016
Vermont	Solid Dem	2017
New Mexico	Solid Dem	2015
New Mexico	Lean Dem	2016
New Mexico	Solid Dem	2017

When looking at tables 2, 3, and 4, the top murder, property crime, and violent crime rates come from Louisiana, which has been competitive all 3 years, Alaska, which has been republican all 3 years,and New Mexico which has been democratic all 3 years. So, there does not look to be a pattern across the highest crimes per capita and the state political affiliation.

On the low side, New Hampshire has leaned republican and been competitive and Vermont has been a solid democratic state all 3 years. Again, there looks to be no pattern from these tables off the bat.

We will go into more depth with regression analysis and other hypothesis testing to analyze patterns across the data and see where there actually may be some correlations.

3.2 Incarceration rates and political affiliation independence

What is the effect, if any, of political affiliation on incarceration rate?

To answer this question, we first calculated the overall incarceration rate percentage for each state by adding total violent crimes and total property crimes, and created a new column to store this information called "incarceration_rate".

To test the independence of state political affiliation and the incarceration rate, we conducted a chi-square test using the democratic advantage column to indicate the state's political leaning. Our null hypothesis (H_o) for this test was that a state's incarceration rate is independent of its democratic leaning. The

p-value of our chi-sq test was .2589. Our p-value is greater than .05, therefore, we fail to reject the null hypothesis that incarceration rate is independent of the state's democratic leaning.

We also conducted a correlation test to answer this question. Our null hypothesis (H_o) for this test was that there is no correlation between a state's incarceration rate and its democratic leaning. The results showed a -.196 correlation. The p-value for this test was .0165. Our p-value is less than .05, therefore, we reject the null hypothesis that there is no correlation between incarceration rate and a state's democratic leaning. This conclusion means that there is a weak negative correlation between the democratic advantage and the incarceration rates. As democratic advantage increases, incarceration rate decreases slightly.

We also created a linear regression model to help answer this question. We found that the incarceration rate is estimated to be 2.8 when a state has a democratic advantage of 0. Each additional point increase of democratic advantage is associated with a decrease of .0097 in the state's incarceration rate. From the r-squared value, we learn that a state's political leaning explains 3.85% of the variation in incarceration rate. Because of these low values, we can determine that democratic advantage is a very weak predictor of incarceration rate.

	Estimate	Std. Error	T-Value	Pr
Intercept	2.8055	0.05649	49.66	<2e-16
Democratic_Advantage	-.009667	0.00399	-2.426	0.0165
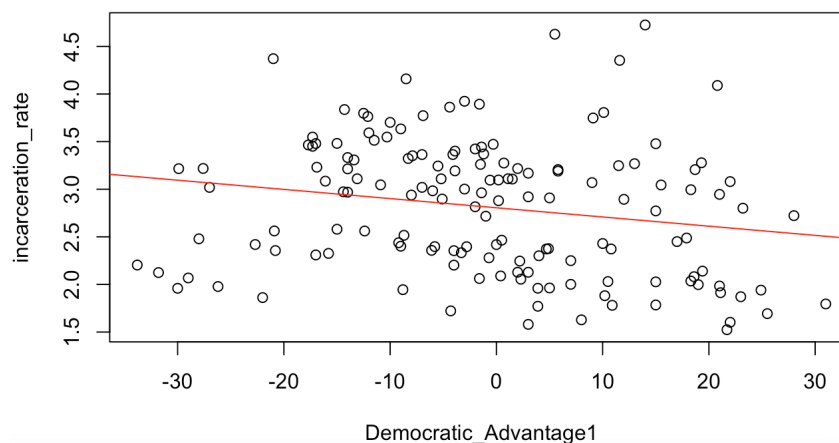
3.3 Difference between the means of Republican, Democrat, and Competitive states

Is there a difference in incarceration rates between Republican, Democrat, and Competitive states?

The goal of our analysis is to determine whether Republican and Democrat states have differences in incarceration rates. To answer this question, we used the "incarceration_rate" column discussed in 3.2 to run an ANOVA test. This test also requires categorical variables to test the differences between the categories.

We decided that the best way to break down states was by Republican, Democrat, and Competitive. To do this, we used the variable Classification1 to create a new variable called Classification2. Classification 2 took the categories "Solid Rep" and "Leaning Rep" and classified those both as "Rep" for Republican states. The same was done with Democratic states. We kept "Competitive" states as Competitive.

Before running the ANOVA test, we established our null hypothesis (H_o) as being all incarceration_rate means were equal between Republican, Democratic, and Competitive states. Now that all the elements of an ANOVA test were set, we ran the code to get the results.

After getting the summary of the ANOVA test, we found that the p-value was 0.012. This means that we will reject the null hypothesis that Incarceration Rate means are equal across the Republican, Democratic, and Competitive states. To further investigate which categories were causing means to not be equal, we ran Tukey's range test. Within this test, we got the following results:

Classification2	diff	lwr	upr	p adj
Dem-Competitive	-0.0848	-0.425	0.256	0.825
Rep-Competitive	0.308	-0.017	0.633	0.067
Rep-Dem	0.393	0.064	0.722	0.015

After reviewing the table above, we see that the Rep-Dem p-value is less than 0.05 and the other comparisons are greater than 0.05. The Rep-Dem p-value indicates that their means are unequal, whereas the other comparison p-values indicate that we fail to reject the null hypothesis that means are equal.

In conclusion, there is a statistically significant difference in Incarceration Rates between Republican and Democratic states. When comparing Republican and Democratic states to competitive states, there is not a statistically significant difference between Incarceration Rates.


3.4 Number of crimes vs. political affiliations

**Does crime increase or decrease based on how political affiliations flip year by year?**

To answer this question, we created histogram plots that counted the number of states that affiliated with a political party. In total, we created three histogram plots, one each for years 2015, 2016, and 2017. Then we created a line graph chart with two lines showing how property crimes and violent crimes changed

over the years. Then by comparing the two different types of graphs, we determined whether crimes increased or decreased based on political affiliation.

When looking at our histograms, we determined that in 2015, 13 states either leaned republican or were solid republican. 21 states were either lean democrat or solid democrat. In 2016, Republican states increased by +8 while democratic states decreased by -7. Then in 2017, republican states decreased by -5 to a total 16 and democratic states increased by +4 to a total of 18.

2015

2016

2017
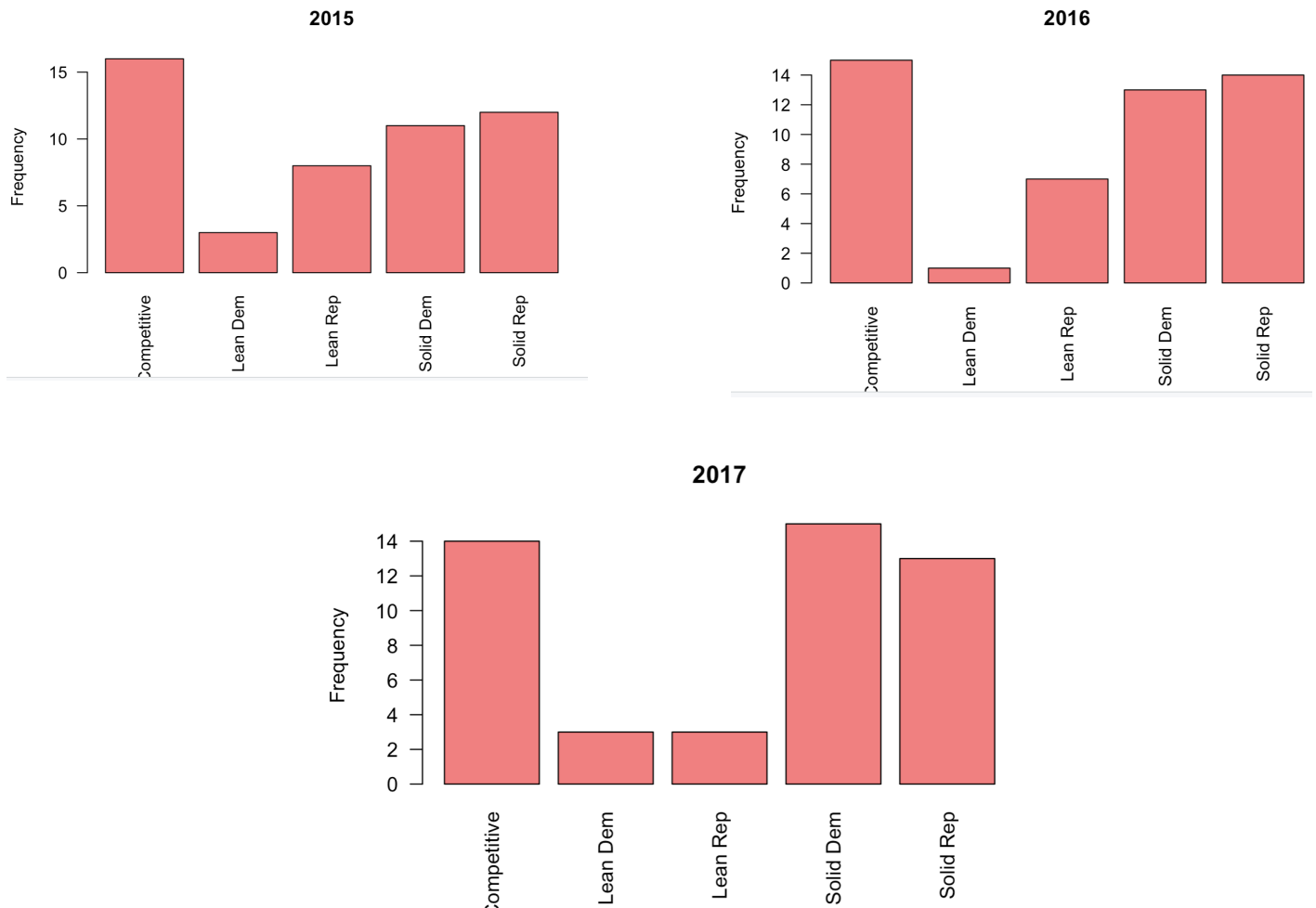
When looking at the line graph, as expected, property crimes largely outnumber violent crimes in each and every year. In fact, property crimes outnumber violent crimes by nearly 7 times in all three years of our dataset. Taking a closer look at our dataset, we found that both violent and property crimes increased in 2016 compared to 2015. In 2016, violent crimes increased by a staggering 11.3% while property crimes increased by 3.3%. Then in 2017, violent crimes decreased by ~1% while property crimes decreased by 3%.

Property Crimes vs. Violent Crimes by Year
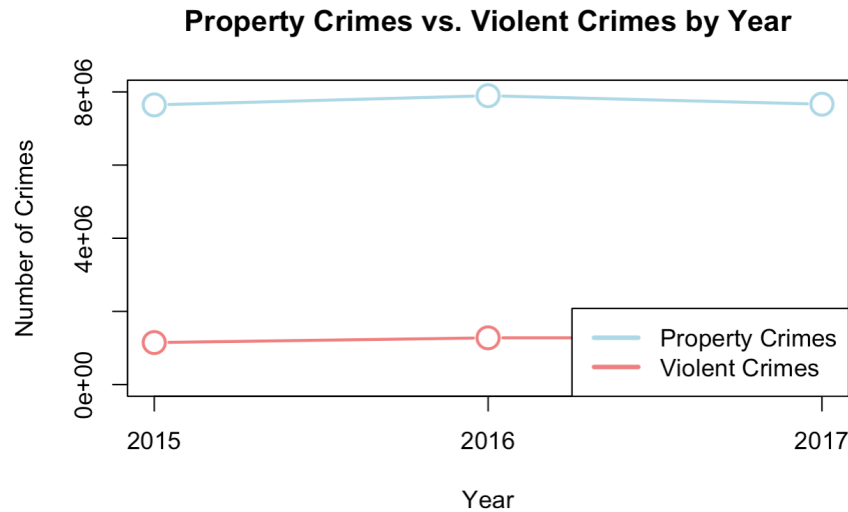
In conclusion, we found that from 2015 to 2016, violent crime increased by 11.3% and property crimes increased by 3.3%. In that same time frame, our data set went from 13 republican states to 21, leading to a 65% increase in republican leaning states. Then from 2016 to 2017, we saw crime decrease, albeit slightly. In that same time frame, we found that republican states decreased by 23% while democratic states increased by 28%. All in all, we can conclude from these findings that an increase in republican states leads to a higher number of crimes while increases in democratic states leads to a decrease in overall crime.

4. Conclusion

This project utilizes United States government incarceration rate data joined with political affiliation data scraped from Gallup to examine the effect of political affiliation on state incarceration rates. Through the use of summary tables, correlation testing, linear regression, ANOVA, and the Tukey's range test, we concluded that a minor relationship does exist between political affiliation and incarceration rates. States with a greater democratic advantage, meaning they lean more democratically, have slightly lower incarceration rates, and solid democratic states have statistically lower incarceration states than solid republican states.

The most limiting aspect of our analysis was the time frame of data collection. We only had three years of data, which does not fully encompass the changes in political affiliation that occur over time. While we did have an election year (2016) present in our data, a more comprehensive dataset covering multiple election years would have allowed us to more accurately assess changes in the political climate of the United States. In the same way, a larger time frame would show a better assessment of the changes in incarceration rates over time.