Marvelous Starmie

Report

Introduction

Our team is investigating the relationship between crime data and state allocation of spending. Our research question is analyzing whether state allocation of funds has a correlation with violent crime rates. The hypothesis is that states with more state funding for prisons, education, and the police tend to have lower violent crime rates.

Data description

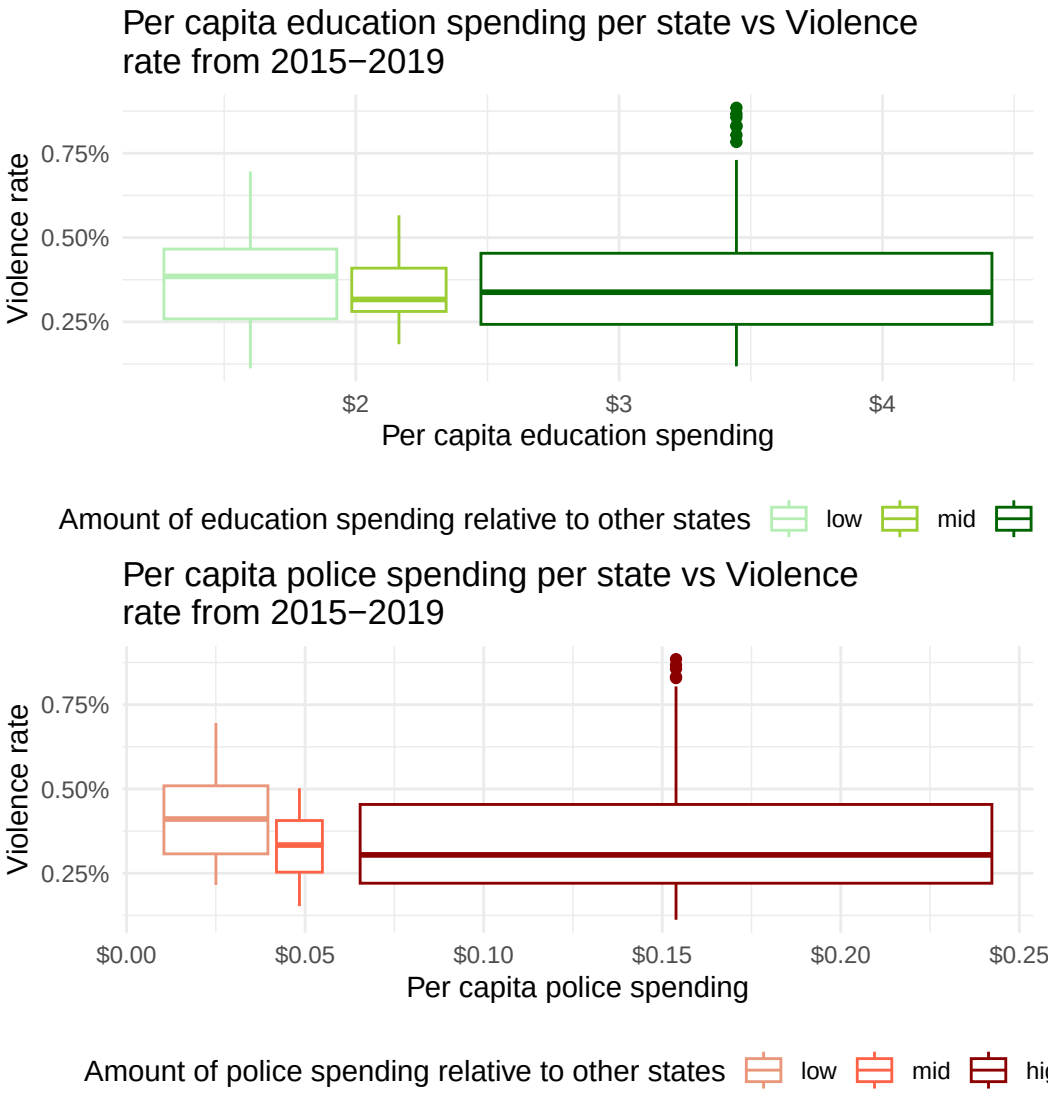
Our data set (finance_crime) combined columns of violent crime counts for each state between 2015-2019 with proportions of state spending for fields including prisons, education, and the police. Each row is the data for one state for a given year. Each state has 5 years, so there are 5 rows per state. We merged two different data sources, one from the National Incident-Based Reporting System's crime reports and the other from state spending reports.The crime data was published by the Uniform Crime Reporting Program from the FBI to help provide comprehensive information surrounding police forces for research, analysis, and public use. Funded by the FBI, the data was obtained from official crime arrest records through a partnership with the police, so a reasonable factor influencing data collection is underlying biases leading to the arrests. Since each year of the data was separated, we had to merge each year and select the necessary columns and rows from each data frame, dropping NA values, and calculating the violence rates from the provided frequencies. The state spending reports come from the United States Census Bureau, which is funded by the government. The state government representatives were aware of the data collection and its intent. This data was created to aid policy research, GDP estimates, and education. Our group cleaned the the spending data before merging it with the crime data.

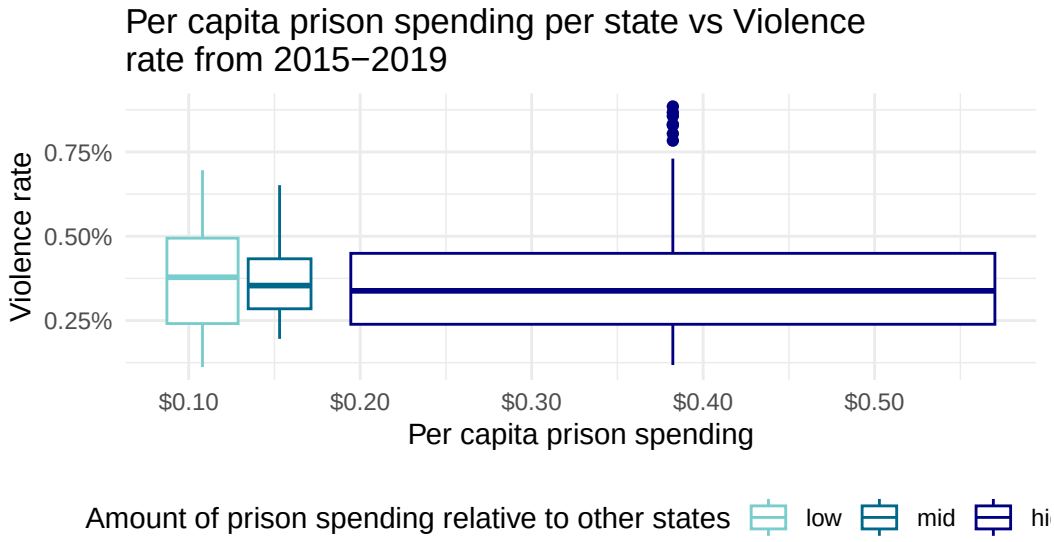
Data analysis

```
[1] 0.1565091
```

```
[1] 0.002084022
```

```
[1] 0.3096611
```





The correlation value between per capita education spending and violence rate is 0.002084022, indicating a weak correlation. This is evident in the graph, as states with a 'mid' quantity of education spending have a significantly lower median violence rate (0.27%) than states with a 'high' quantity of education spending (0.30%). This contradicts our hypothesis because we expected the violence rate to decrease as education spending increased.

The correlation value between per capita police spending and violence rate is 0.1565091, indicating a weak positive correlation, although significantly stronger than police spending's correlation. As shown, states with a 'low' quantity of police spending (0.4%) have a higher median violence rate than median violence rate of states with a 'mid' quantity of police spending (0.3%), which have a significantly lower median violence rate than states with a 'high' quantity of police spending (0.27%). This data is consistent with our hypothesis, because the violence rate sequentially decreases as police spending increases. However, correlation is still weakly positive so there may be other factors influencing the trend.

The correlation value between per capita prison spending and violence rate is 0.3096611, indicating a small positive correlation but is significantly stronger than both police and education spending's correlation. As shown, states with a 'low' quantity of prison spending (0.375%) have a higher median violence rate than median violence rate of states with a 'mid' quantity of prison spending (0.35%), which have a significantly lower median violence rate than states with a 'high' quantity of prison spending (0.27%). This data supports our hypothesis because the violence rate decreases as prison spending increases. Although this correlation is considered the strongest of the three sectors, it is still not significant enough, so there must be other factors contributing to the decreasing median violence rates.

##Linear Model

```
Call:
lm(formula = Violent_Rate ~ percap_prison, data = finance_crime)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0029664 -0.0009922 -0.0001928  0.0007387  0.0044274

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0025574  0.0002393  10.688  < 2e-16 ***
percap_prison 0.0068563  0.0013369   5.129 5.89e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001406 on 248 degrees of freedom
Multiple R-squared:  0.09589,   Adjusted R-squared:  0.09224
F-statistic:  26.3 on 1 and 248 DF,  p-value: 5.885e-07



Call:
lm(formula = Violent_Rate ~ percap_edu, data = finance_crime)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0025719 -0.0011950 -0.0001194  0.0007996  0.0051466

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.685e-03  3.682e-04  10.007   <2e-16 ***
percap_edu  5.158e-06  1.572e-04   0.033    0.974
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001478 on 248 degrees of freedom
Multiple R-squared:  4.343e-06, Adjusted R-squared:  -0.004028
F-statistic: 0.001077 on 1 and 248 DF,  p-value: 0.9738



Call:
lm(formula = Violent_Rate ~ percap_police, data = finance_crime)

Residuals:
```

```
       Min         1Q     Median         3Q        Max
-0.0031688 -0.0010605 -0.0000954  0.0008181  0.0047902

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.0033107  0.0001801  18.378   <2e-16 ***
percap_police 0.0065531  0.0026260   2.495   0.0132 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.00146 on 248 degrees of freedom
Multiple R-squared:  0.0245,    Adjusted R-squared:  0.02056
F-statistic: 6.227 on 1 and 248 DF,  p-value: 0.01323



Call:
lm(formula = Violent_Rate ~ percap_prison + percap_edu + percap_police,
    data = finance_crime)

Residuals:
       Min         1Q     Median         3Q        Max
-0.0022304 -0.0011045 -0.0001721  0.0006560  0.0044675

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.0031154  0.0003759   8.288 7.52e-15 ***
percap_prison  0.0116561  0.0021776   5.353 1.99e-07 ***
percap_edu    -0.0004193  0.0001809  -2.318   0.0213 *
percap_police -0.0068776  0.0043622  -1.577   0.1162
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001379 on 246 degrees of freedom
Multiple R-squared:  0.137, Adjusted R-squared:  0.1264
F-statistic: 13.01 on 3 and 246 DF,  p-value: 6.451e-08



Call:
lm(formula = Violent_Rate ~ percap_prison * percap_edu * percap_police,
    data = finance_crime)

Residuals:
```

```
       Min         1Q     Median         3Q        Max
-0.0023054 -0.0008891 -0.0000250  0.0006738  0.0047092

Coefficients:
                                        Estimate Std. Error t value Pr(>|t|)
(Intercept)                            0.0148470  0.0021943   6.766 9.88e-11
percap_prison                         -0.0614422  0.0141342  -4.347 2.03e-05
percap_edu                            -0.0034030  0.0007952  -4.279 2.70e-05
percap_police                         -0.1797356  0.0376931  -4.768 3.21e-06
percap_prison:percap_edu               0.0198603  0.0051250   3.875 0.000137
percap_prison:percap_police            0.9520338  0.1897377   5.018 1.01e-06
percap_edu:percap_police               0.0432150  0.0116936   3.696 0.000271
percap_prison:percap_edu:percap_police -0.2518722  0.0558739  -4.508 1.02e-05

(Intercept)                           ***
percap_prison                         ***
percap_edu                            ***
percap_police                         ***
percap_prison:percap_edu              ***
percap_prison:percap_police           ***
percap_edu:percap_police              ***
percap_prison:percap_edu:percap_police ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.001217 on 242 degrees of freedom
Multiple R-squared:  0.3385,    Adjusted R-squared:  0.3193
F-statistic: 17.69 on 7 and 242 DF,  p-value: < 2.2e-16
```

In this analysis, we compared five different linear models.

1. Prison Model: The prison_model shows that about 9.22% of the variation in the violent crime rate can be explained by per capita spending on prisons. The percap_prison variable is statistically significant ($p < 0.001$).
2. Education Model: The edu_model suggests that the percap_edu variable does not explain any variation in the violent crime rate, and it is not statistically significant ($p = 0.974$).
3. Police Model: In the police_model, approximately 2.056% of the variation in the violent crime rate can be explained by per capita spending on police. The percap_police variable is statistically significant ($p = 0.0132$).

4. Additive Model (pep_add_model): This model accounts for 12.64% of the variation in the violent crime rate by combining per capita prison, education, and police expenditure. It shows a significant positive relationship between violent crime rate and per capita prison expenditure ($p < 0.001$), a significant negative relationship between violent crime rate and per capita education expenditure ($p = 0.0213$), and no statistically significant relationship between violent crime rate and per capita police expenditure ($p = 0.1162$).
5. Interaction Model (pep_int_model): The interaction model explains approximately 31.93% of the variation in the violent crime rate by considering the interactions between the three variables. Since $p < 0.001$, the model indicates that their combined effect is significant and not independent of each other.

The interaction model (pep_int_model) has the highest adjusted R-squared value, suggesting that it is crucial to consider the interactions of police, education, and prison spending on violent crime.

Evaluation of significance

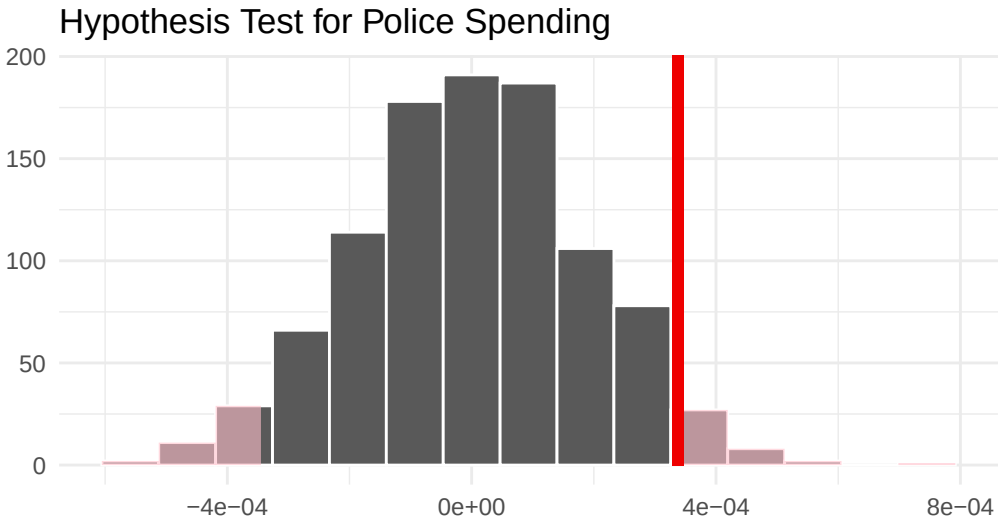
Hypothesis Test - Police

$$H_0 : p_c - p_p = 0$$

$$H_A : p_c - p_p \neq 0$$

Null Hypothesis: There is no significant difference in violence rate in states with low per cap police spending and high per cap police spending.

Alternative Hypothesis: There is a significant difference in violence rate in states with low per cap police spending and high per cap police spending.




```
# A tibble: 1 x 1
  p_value
    <dbl>
1   0.068
```

Since the p-value is 0.068 which is greater than 0.05(95% significance level), we fail to reject the null hypothesis. The data does not provide convincing evidence that the average violence crime rates of states with low per cap police spending and high per cap police spending is different.

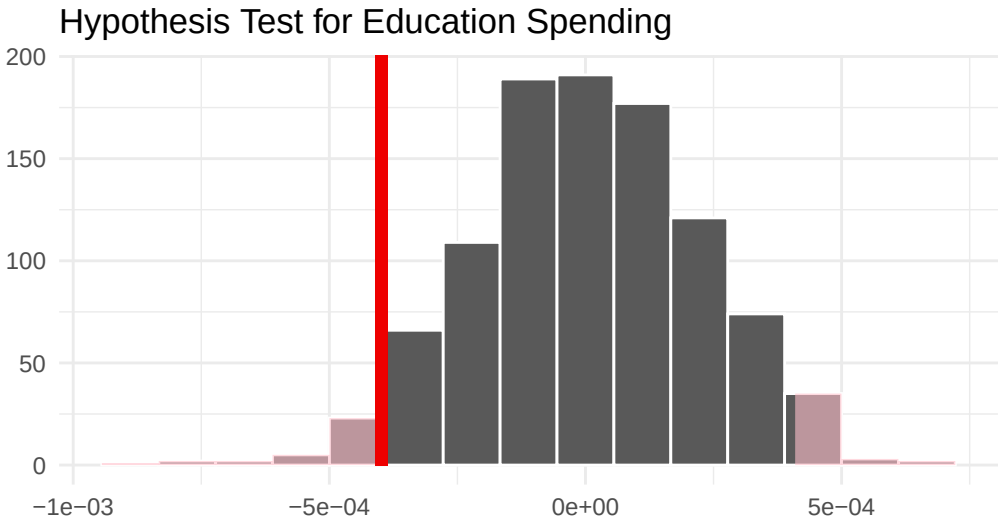
Hypothesis Test - Education

$$H_0 : p_c - p_e = 0$$

$$H_A : p_c - p_e \neq 0$$

Null Hypothesis: There is no significant difference in violence rate in states with low per cap education spending and high per cap education spending.

Alternative Hypothesis: There is a significant difference in violence rate in states with low per cap education spending and high per cap education spending.




```
# A tibble: 1 x 1
  p_value
    <dbl>
1    0.06
```

Since the p-value is 0.06 which is greater than 0.05(95% significance level), we fail to reject the null hypothesis. The data does not provide convincing evidence that the average violence crime rates of states with low per cap education spending and high per cap education spending is different.

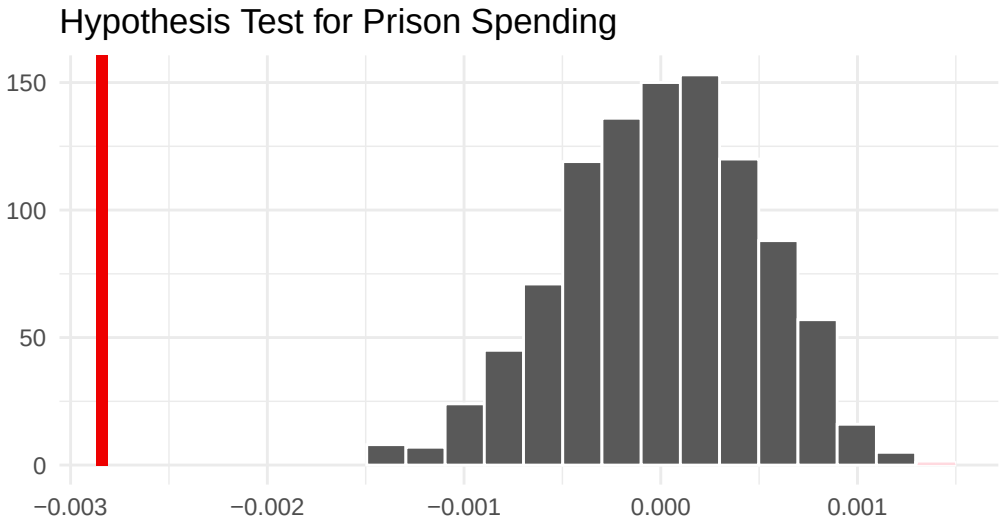
Hypothesis Test - Prison

$$H_0 : p_c - p_r = 0$$

$$H_A : p_c - p_r \neq 0$$

Null Hypothesis: There is no significant difference in violence rate in states with low per cap prison spending and high per cap prison spending.

Alternative Hypothesis: There is a significant difference in violence rate in states with low per cap prison spending and high per cap prison spending.




```
# A tibble: 1 x 1
  p_value
    <dbl>
1       0
```

Since the p-value is 0 which is less than any significance level, we can reject the null hypothesis in favor of the alternative hypothesis. The data provides convincing evidence that the average violence crime rates of states with low per cap prison spending and high per cap prison spending is different.

Interpretation, Conclusions, and Limitations

Across the country, states spend more on education than on police or prisons per capita. There is a relatively low correlation between types of spending and violence rates, which could be because prison and police spending is similar for each state. New Mexico and Alaska are clear outliers, with higher violence rates than other states, and Nevada's crime rates dropped from 2016 to 2019, although it isn't clear whether it was due to a spending difference. For the hypothesis tests, we failed to reject on a 95% significance level the difference in violence rates between states with high or low police and education spending, but were able to reject the difference in violence rates between states with high or low prison spending, indicating a closer correlation between crime and prison reform.

A limitation of our project is that the crime rate data only had data up to 2019, so obtaining data until 2023 would further improve our analysis. Because there was a lot of state data, it was difficult to visualize all of it clearly. Although we analyzed data related to spending on police, education, and prisons, more analysis relating recent policies to crime and spending would offer additional insight.