

# Crime Rates in the US

Jade Towgood

2023-11-19

# Data Set USArrests

This data set displays the violent crime rates per 100,000 residents by US State in 1973.

There are 4 variables with 50 states in this data set:

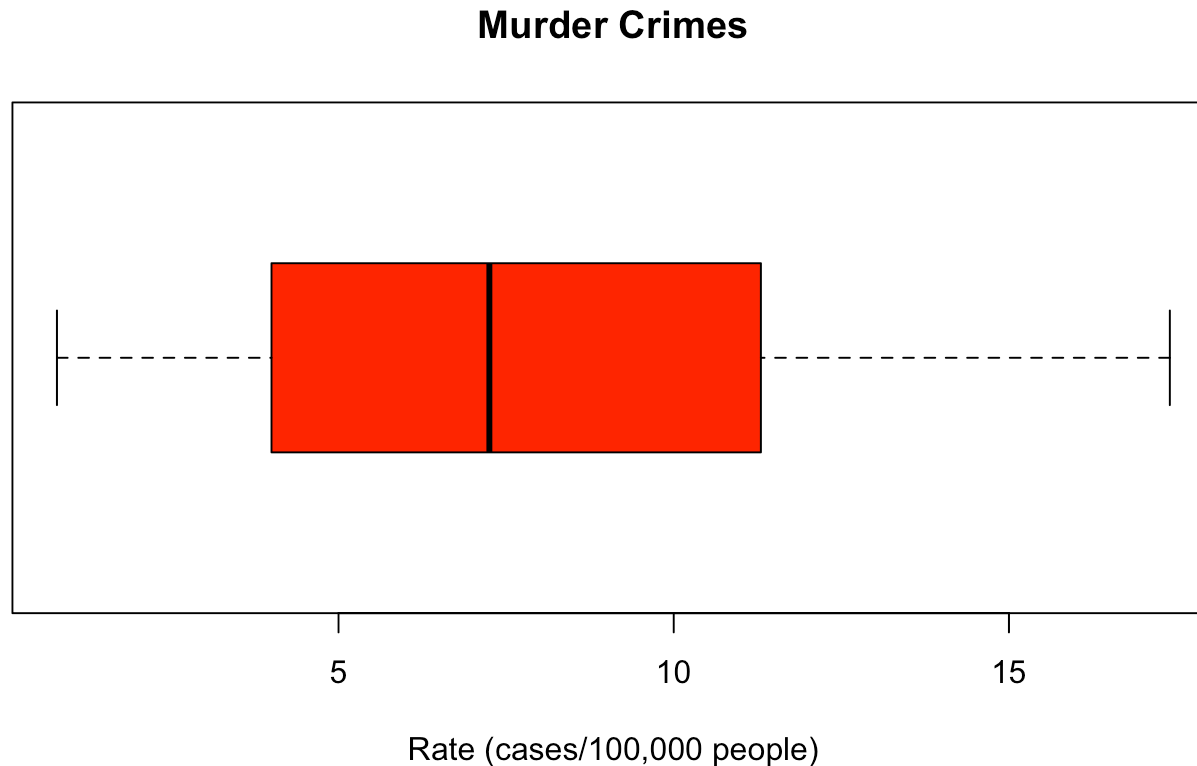
- murder
- assault
- rape
- urban population

# The Question

Is there a correlation between the violent crime rates and each other?

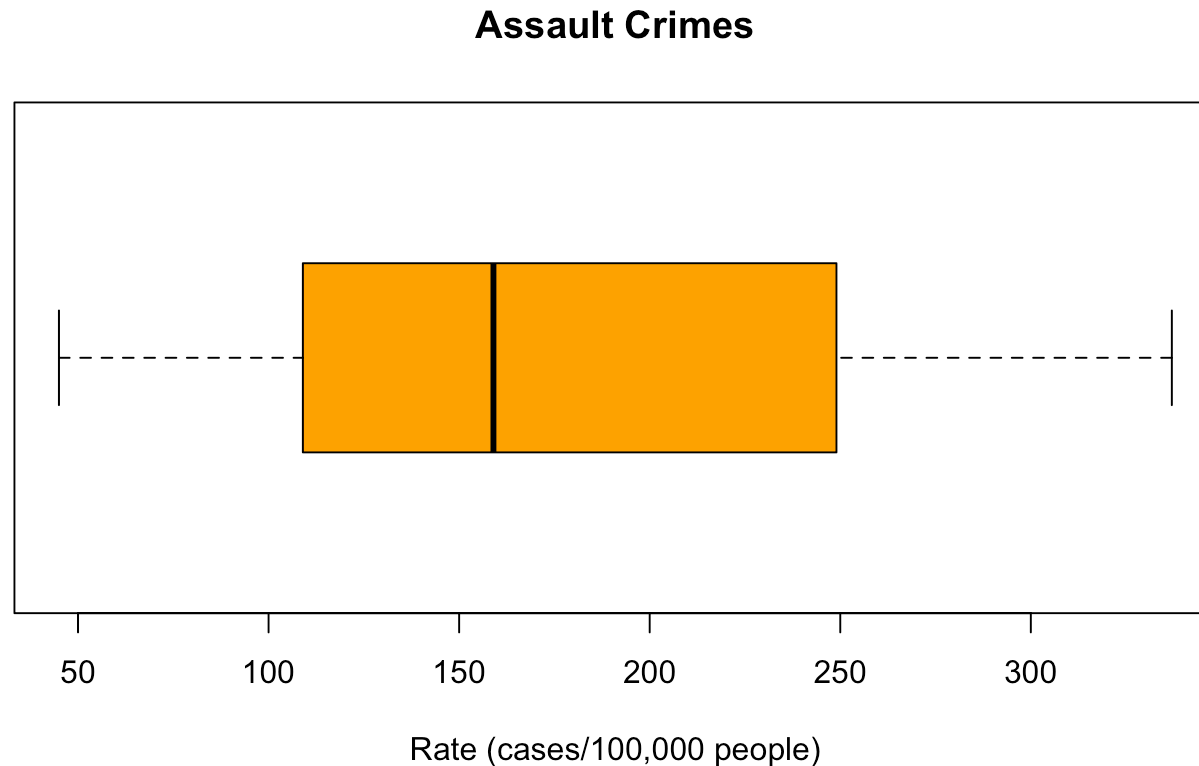
- rape vs murder
- murder vs assault
- assault vs rape

# Summary Violent Crimes: Murder



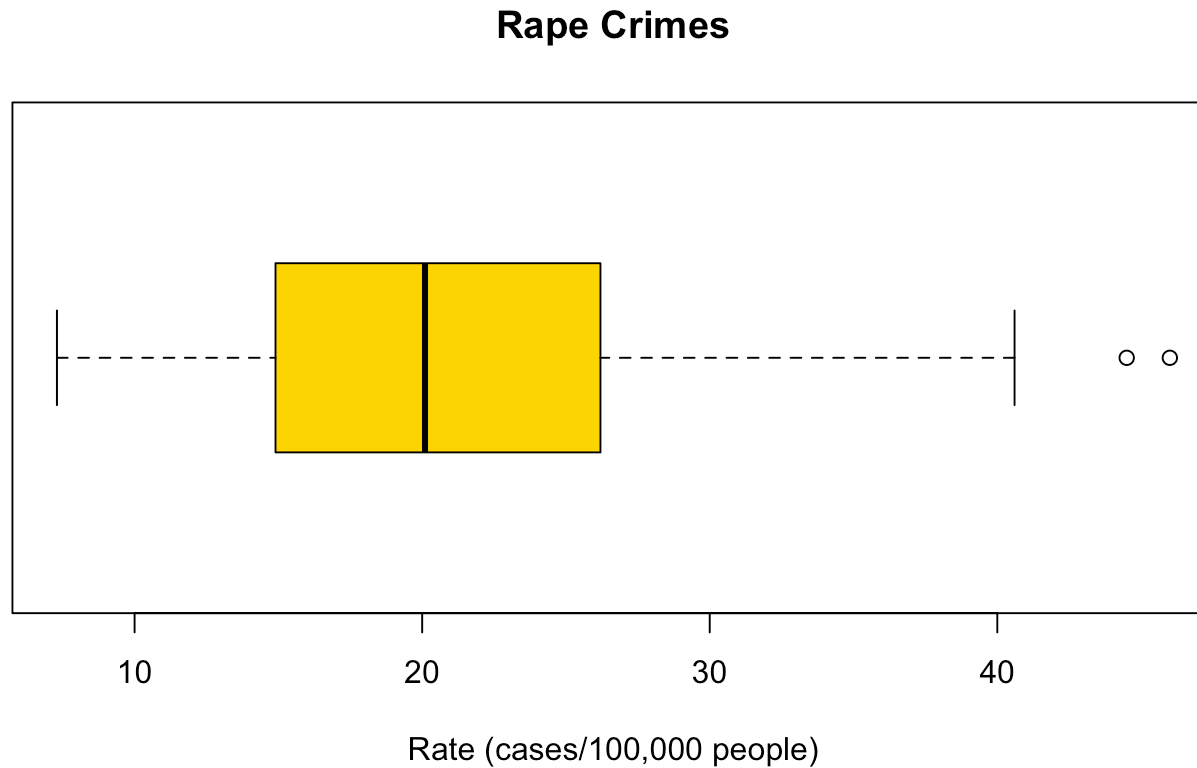
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	0.800	4.075	7.250	7.788	11.250	17.400

# Summary Violent Crimes: Assault



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	45.0	109.0	159.0	170.8	249.0	337.0

# Summary Violent Crimes: Rape



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	7.30	15.07	20.10	21.23	26.18	46.00

# Explanation of Code: Boxplots

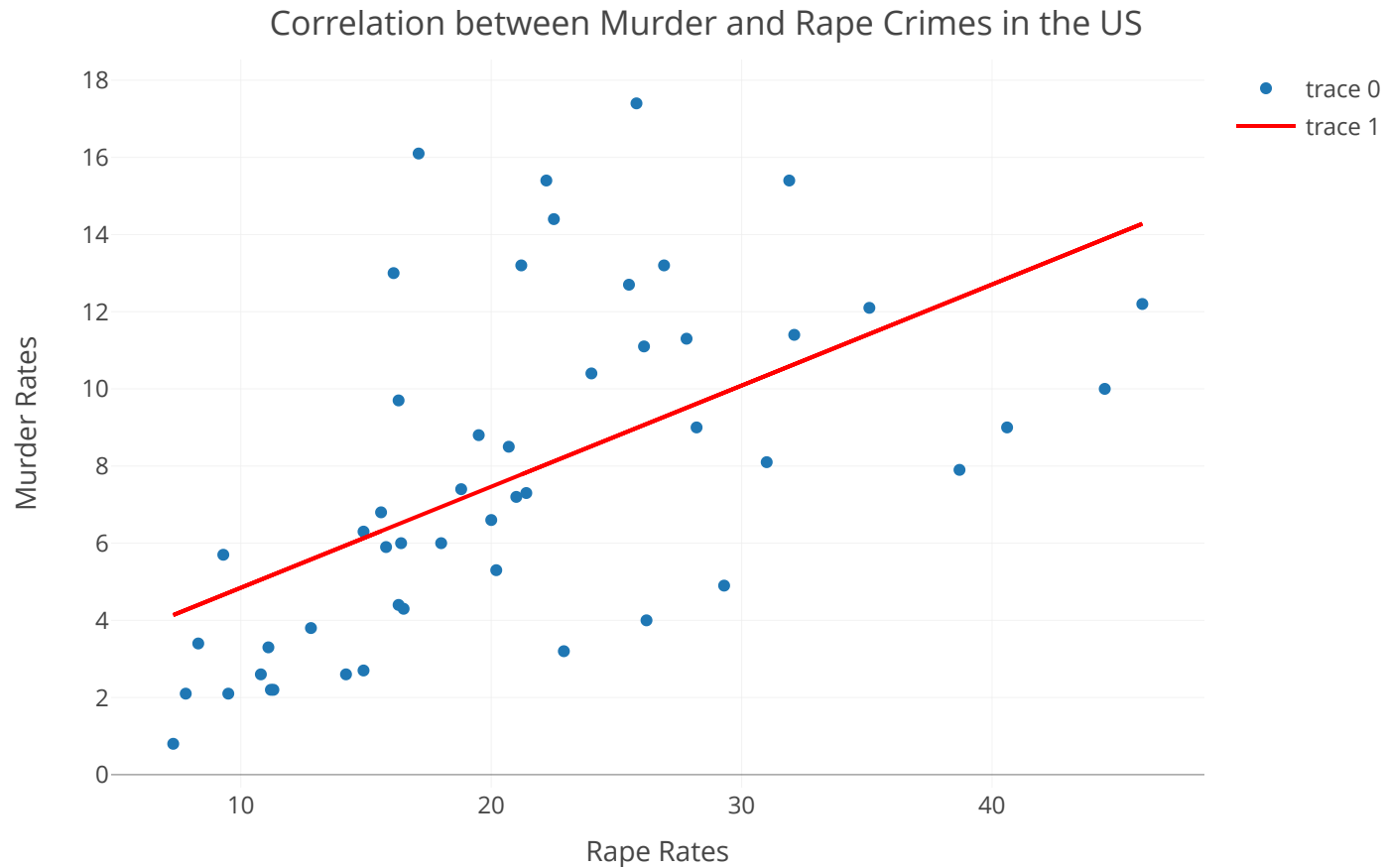
```
boxplot(USArrests$Murder, horizontal = TRUE, xlab = "Rate  
  (cases/100,000 people)", main = "Murder Crimes", col = "red")  
summary(USArrests$Murder)
```

```
boxplot(USArrests$Assault, horizontal = TRUE, xlab = "Rate  
  (cases/100,000 people)", main = "Assault Crimes", col =  
  "orange")  
summary(USArrests$Assault)
```

```
boxplot(USArrests$Rape, horizontal = TRUE, xlab = "Rate  
  (cases/100,000 people)", main = "Rape Crimes", col = "gold")  
summary(USArrests$Rape)
```

I decided to do 3 separate boxplots and summaries to make the data easier to read and understand it instead of putting it all together. I excluded the urban population data because it was not a variable I was discussing.

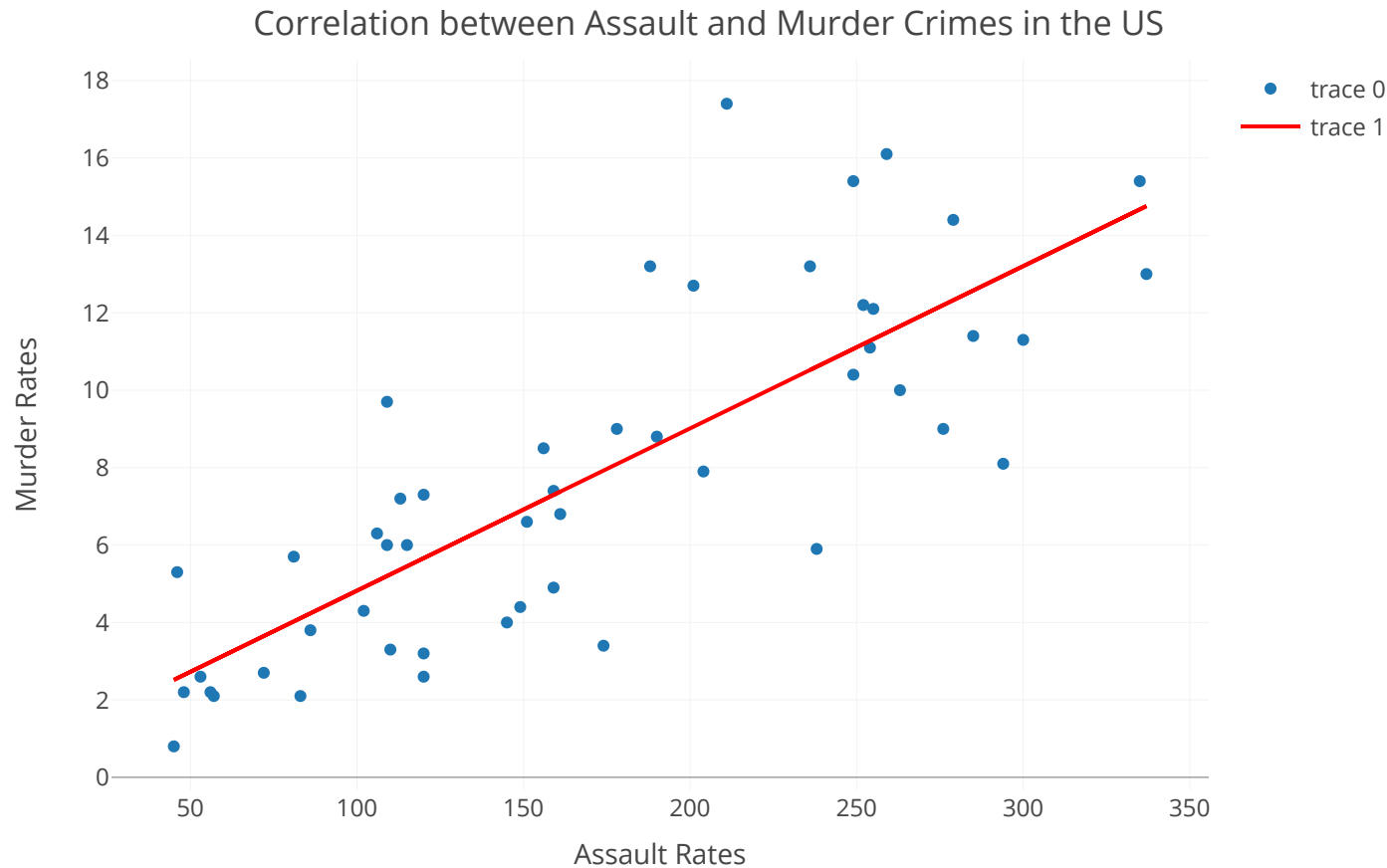
# Scatter Plots: Rape vs Murder



This scatter plot displays a positive correlation between the rate of rape and murder crimes.

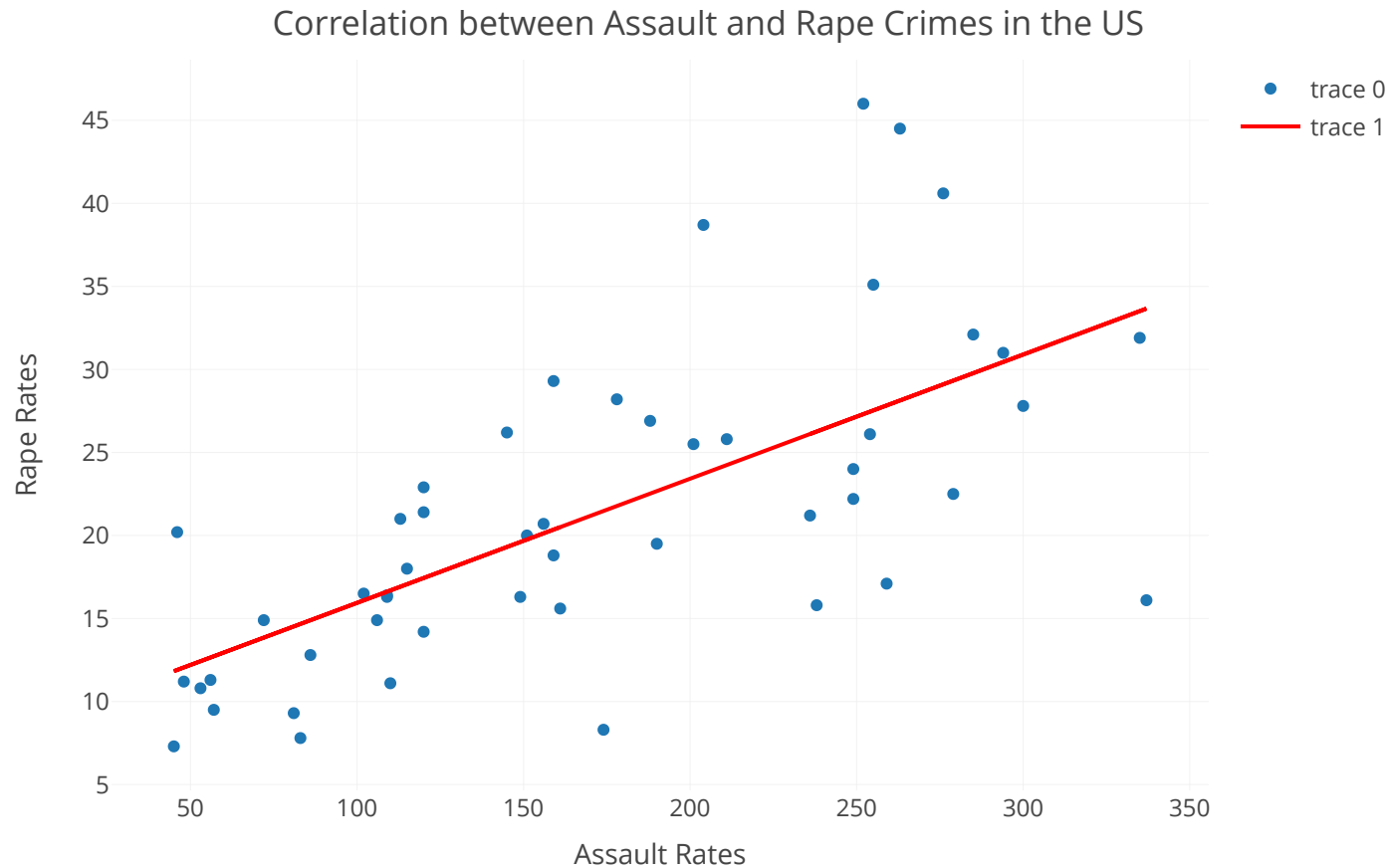


# Scatter Plots: Assault vs Murder



This scatter plot displays a positive correlation between the rate of assault and murder crimes.

# Scatter Plots: Assault vs Rape



This scatter plot displays a positive correlation between the rate of assault and rape crimes.

# Explanation of Code: Scatterplots

```
murderRates <- USArrests$Murder  
rapeRates <- USArrests$Rape
```

```
x = rapeRates  
y = murderRates
```

```
plot_ly(x = x, y = y, type = "scatter", mode = "markers") %>%  
  add_trace(x = rapeRates, y = predict(lm(murderRates~rapeRates)),  
    mode = "lines", type = "scatter", line = list(color =  
      "Red")) %>%  
  layout(title = "Correlation between Murder and Rape Crimes in  
the US", xaxis = list(title = "Rape Rates"), yaxis = list  
    (title = "Murder Rates"))
```

# Cont.

```
assultRates <- USArrests$Assault  
murderRates <- USArrests$Murder
```

```
x = assultRates  
y = murderRates
```

```
plot_ly(x = x, y = y, type = "scatter", mode = "markers") %>%  
  add_trace(x = assultRates, y = predict(lm(murderRates~assultRates)),  
            mode = "lines", type = "scatter", line = list(color =  
              "Red")) %>%  
  layout(title = "Correlation between Assault and Murder Crimes  
in the US", xaxis = list(title = "Assault Rates"), yaxis =  
    list(title = "Murder Rates"))
```

## Cont.

```
assultRates <- USArrests$Assault
rapeRates <- USArrests$Rape

x = assultRates
y = rapeRates

plot_ly(x = x, y = y, type = "scatter", mode = "markers") %>%
  add_trace(x = assultRates, y = predict(lm(rapeRates~assultRates)),
    mode = "lines", type = "scatter", line = list(color =
      "Red")) %>%
  layout(title = "Correlation between Assault and Rape Crimes
in the US", xaxis = list(title = "Assault Rates"), yaxis =
  list(title = "Rape Rates"))
```

I decided that scatterplots would be the best way to analyze whether there was a correlation with a linear regression model included.

# Linear Regression Model: $R^2$ Value

Rape vs Murder

## R-squared value: 0.3176211

Assault vs Murder

## R-squared value: 0.6430008

Assault vs Rape

## R-squared value: 0.4425459

# Explanation of Code: Linear Regression Model

```
assultRates <- USArrests$Assault
rapeRates <- USArrests$Rape
murderRates <- USArrests$Murder

x1 = assultRates
y1 = rapeRates
y2 = murderRates

lm_model <- lm(y2~y1, data = USArrests)
r_squared <- summary(lm_model)$r.squared
cat("R-squared value:", r_squared, "\n")
```

# Conclusion

Assuming:

- The size of error doesn't change significantly across the values.
- There are no hidden relationships among the variables.
- The data has a normal distribution.
- The Relationship between the variables are linear.

There seems to be a positive correlation between all crimes in this data set, which answers my question of whether each violent crime has some type of correlation with one another.