jjimenez_608_story3

February 28, 2024

CUNY SPS Data 608 Spring 2024

Jean Jimenez

1 Story 3 – Do stricter gun laws reduce firearm gun deaths? –

1.1 Introduction

The purpose of this Story is to answer the question, " Do stricter firearm control laws help reduce firearm mortality?"

To solve this question and tell this story, we were asked to create heatmap visualization of our data.

In this assignment, I compare firearm mortality data with data on Gun Law Strictness with the aim of answering the above question.

1.2 Data Handling and Processing

1.2.1 Importing CDC Firearm Mortality Data

I obtained Firearm Mortality Data from the CDC Database, specifically:

National Center for Health Statistics. NCHS - VSRR Quarterly provisional estimates for selected indicators of mortality. Date accessed 02/26/2024. Available from https://data.cdc.gov/d/489q-934x.

I imported the data using API request.

```python
import requests
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import plotly.express as px
from IPython.display import display, HTML
import statsmodels.api as sm
```

```python
url = "https://data.cdc.gov/resource/489q-934x.json"

response = requests.get(url)

if response.status_code == 200:
```

```
    df_mortality = pd.DataFrame(response.json())

    print(df_mortality.head())
else:
    print("Error: Unable to retrieve")
```

```
  year_and_quarter                  time_period  \
0          2021 Q1  12 months ending with quarter
1          2021 Q1  12 months ending with quarter
2          2021 Q1  12 months ending with quarter
3          2021 Q1  12 months ending with quarter
4          2021 Q1  12 months ending with quarter


                       cause_of_death      rate_type                   unit  \
0                          All causes  Age-adjusted  Deaths per 100,000
1                   Alzheimer disease  Age-adjusted  Deaths per 100,000
2                            COVID-19  Age-adjusted  Deaths per 100,000
3                              Cancer  Age-adjusted  Deaths per 100,000
4  Chronic liver disease and cirrhosis  Age-adjusted  Deaths per 100,000


  rate_overall rate_sex_female rate_sex_male rate_alaska rate_alabama  ...  \
0        866.3           716.3        1040.4       779.2       1123.4  ...
1         32.1            36.8          24.8        28.2         51.2  ...
2        120.7              94         153.9        44.4        160.2  ...
3          142           122.8         167.7         143        160.5  ...
4         13.9             9.8          18.3        23.6         17.2  ...


  rate_age_1_4 rate_age_5_14 rate_age_15_24 rate_age_25_34 rate_age_35_44  \
0          NaN           NaN            NaN            NaN            NaN
1          NaN           NaN            NaN            NaN            NaN
2          NaN           NaN            NaN            NaN            NaN
3          NaN           NaN            NaN            NaN            NaN
4          NaN           NaN            NaN            NaN            NaN


  rate_age_45_54 rate_age_55_64 rate_65_74 rate_age_75_84 rate_age_85_plus
0            NaN            NaN        NaN            NaN              NaN
1            NaN            NaN        NaN            NaN              NaN
2            NaN            NaN        NaN            NaN              NaN
3            NaN            NaN        NaN            NaN              NaN
4            NaN            NaN        NaN            NaN              NaN

[5 rows x 69 columns]
```

1.2.2 Cleaning CDC Firearm Mortality Data

After importing the data, I had to clean it up and process it.

I extracted only the columns that had data on Firearm Mortality by year.

The dataframe had alot of columns. Each state's mortality rate was in a separate column. Using the data dictionary from this dataset, GPT-4 generated a dictionary of each of the columns of interest from the dataframe as a key and the state abbreviation as a value.

I then process the text to extract year. Afterwards, I only extract firearm mortality rate for 2023 Q1 & Q2 (the last published data)

```python
#print(df_mortality.columns)


firearm_df = df_mortality[(df_mortality['cause_of_death'] == "Firearm-related␣
↪injury") &
                        (df_mortality['rate_type'] == "Crude") &
                        (df_mortality['time_period'] == "12 months ending with␣
↪quarter")]

firearm_df.reset_index(drop=True, inplace=True)

#GPT 4 used to generate the following dictrionary:

rename_cols = {
    "rate_alaska": "AK", "rate_alabama": "AL", "rate_arkansas": "AR",␣
↪"rate_arizona": "AZ",
    "rate_california": "CA", "rate_colorado": "CO", "rate_connecticut": "CT",␣
↪"rate_district_of_columbia": "DC",
    "rate_delaware": "DE", "rate_florida": "FL", "rate_georgia": "GA",␣
↪"rate_hawaii": "HI",
    "rate_iowa": "IA", "rate_idaho": "ID", "rate_illinois": "IL",␣
↪"rate_indiana": "IN",
    "rate_kansas": "KS", "rate_kentucky": "KY", "rate_louisiana": "LA",␣
↪"rate_massachusetts": "MA",
    "rate_maryland": "MD", "rate_maine": "ME", "rate_michigan": "MI",␣
↪"rate_minnesota": "MN",
    "rate_missouri": "MO", "rate_mississippi": "MS", "rate_montana": "MT",␣
↪"rate_north_carolina": "NC",
    "rate_north_dakota": "ND", "rate_nebraska": "NE", "rate_new_hampshire":␣
↪"NH", "rate_new_jersey": "NJ",
    "rate_new_mexico": "NM", "rate_nevada": "NV", "rate_new_york": "NY",␣
↪"rate_ohio": "OH",
    "rate_oklahoma": "OK", "rate_oregon": "OR", "rate_pennsylvania": "PA",␣
↪"rate_rhode_island": "RI",
    "rate_south_carolina": "SC", "rate_south_dakota": "SD", "rate_tennessee":␣
↪"TN", "rate_texas": "TX",
    "rate_utah": "UT", "rate_virginia": "VA", "rate_vermont": "VT",␣
↪"rate_washington": "WA",
    "rate_wisconsin": "WI", "rate_west_virginia": "WV", "rate_wyoming": "WY"
}
```

```python
firearm_df.rename(columns=rename_cols, inplace=True)

#print(firearm_df.head())

firearm_df.iloc[:, 5:69] = firearm_df.iloc[:, 5:69].apply(pd.to_numeric,␣
 ↪errors='coerce')
firearm_df['year'] = firearm_df['year_and_quarter'].str[:4]

yr_firearm_df = firearm_df.groupby('year').size()

#print(yr_firearm_df.head())

filtered_firearm_df = firearm_df[firearm_df['year_and_quarter'].str.
 ↪contains("2023 Q1|2023 Q2")]

#GPT 4 used to generate the value_vars list:

firearm_data_long = pd.melt(filtered_firearm_df, id_vars=['year',␣
 ↪'year_and_quarter'],
                            value_vars=['AK', 'AL', 'AR', 'AZ', 'CA', 'CO',␣
 ↪'CT', 'DC', 'DE', 'FL', 'GA',
                                        'HI', 'IA', 'ID', 'IL', 'IN', 'KS',␣
 ↪'KY', 'LA', 'MA', 'MD', 'ME',
                                        'MI', 'MN', 'MO', 'MS', 'MT', 'NC',␣
 ↪'ND', 'NE', 'NH', 'NJ', 'NM',
                                        'NV', 'NY', 'OH', 'OK', 'OR', 'PA',␣
 ↪'RI', 'SC', 'SD', 'TN', 'TX',
                                        'UT', 'VA', 'VT', 'WA', 'WI', 'WV',␣
 ↪'WY'],
                            var_name='state', value_name='rate')

final_df = firearm_data_long[['year', 'state', 'rate']]
final_df_cleaned = final_df.dropna(subset=['rate'])

final_df_aggregated = final_df_cleaned.groupby(['year', 'state'])['rate'].
 ↪mean().reset_index()
#print(final_df_aggregated)
```

C:\Users\bleac\AppData\Local\Temp\ipykernel_16504\3957919178.py:28:
SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
C:\Users\bleac\AppData\Local\Temp\ipykernel_16504\3957919178.py:32:
SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\bleac\AppData\Local\Temp\ipykernel_16504\3957919178.py:33:
SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

1.2.3 Importing and Processing Firearm Laws by State

Afterwards, I had to obtain data on Firearms Law by State.

I chose the following source:

State Firearm Law Database: State Firearm Laws, 1991-2019 Siegel, Michael 2020 doi:10.3886/ICPSR37363.v1 Inter-university Consortium for Political and Social Research https://www.icpsr.umich.edu/web/NACJD/studies/37363/versions/V1/summary

I chose this dataset because it essentially had many columns with each type of gun law and asks if each state has that type of law or no. The dataset has a column `LAWTOTAL` that is the total # of gun laws enacted in that state. I will use this as a proxy measure for Gun Law Strictness. I will make the assumption that More Gun Laws == More strict gun laws. I make this assumption because each law listed in the dataset restricted firearm use in some way.

I obtained the data and uploaded a copy on Github (which I call here to get the data).

I chose to filter only the year 2019 because I wanted to have the most updated gun laws.

Similar to the previous dataset, I had GPT-4 generate a dictionary of the State name and abbreviation. I used this dictionary to merge both datasets togeather by state.

```python
[ ]: gun_law_url = 'https://raw.githubusercontent.com/sleepysloth12/data608_story3/
      ↪main/37363-0001-Data.tsv'


     gun_law_dat = pd.read_csv(gun_law_url, sep='\t', usecols=['STATE', 'YEAR',␣
      ↪'LAWTOTAL'])
```

```python
gun_law_dat_latest = gun_law_dat[gun_law_dat['YEAR'] == 2019]


#print("Columns in final_df_aggregated:", final_df_aggregated.columns.tolist())
#print("Columns in gun_law_dat_latest:", gun_law_dat_latest.columns.tolist())

#GPT 4 used to generate the following dictrionary:

abbrev_to_state_names = {
    'AL': 'Alabama', 'AK': 'Alaska', 'AZ': 'Arizona', 'AR': 'Arkansas', 'CA':␣
↪'California',
    'CO': 'Colorado', 'CT': 'Connecticut', 'DE': 'Delaware', 'FL': 'Florida',␣
↪'GA': 'Georgia',
    'HI': 'Hawaii', 'ID': 'Idaho', 'IL': 'Illinois', 'IN': 'Indiana', 'IA':␣
↪'Iowa',
    'KS': 'Kansas', 'KY': 'Kentucky', 'LA': 'Louisiana', 'ME': 'Maine', 'MD':␣
↪'Maryland',
    'MA': 'Massachusetts', 'MI': 'Michigan', 'MN': 'Minnesota', 'MS':␣
↪'Mississippi', 'MO': 'Missouri',
    'MT': 'Montana', 'NE': 'Nebraska', 'NV': 'Nevada', 'NH': 'New Hampshire',␣
↪'NJ': 'New Jersey',
    'NM': 'New Mexico', 'NY': 'New York', 'NC': 'North Carolina', 'ND': 'North␣
↪Dakota', 'OH': 'Ohio',
    'OK': 'Oklahoma', 'OR': 'Oregon', 'PA': 'Pennsylvania', 'RI': 'Rhode␣
↪Island', 'SC': 'South Carolina',
    'SD': 'South Dakota', 'TN': 'Tennessee', 'TX': 'Texas', 'UT': 'Utah', 'VT':␣
↪'Vermont',
    'VA': 'Virginia', 'WA': 'Washington', 'WV': 'West Virginia', 'WI':␣
↪'Wisconsin', 'WY': 'Wyoming'
}
final_df_aggregated['STATE'] = final_df_aggregated['state'].
↪map(abbrev_to_state_names)


#Merging both datasets

final_df_aggregated_merged = pd.merge(final_df_aggregated, gun_law_dat_latest,
                                      left_on='STATE', right_on='STATE',␣
↪how='left',
                                      suffixes=('', '_from_gun_law'))


final_df_aggregated_merged.drop('state', axis=1, inplace=True)

final_df_aggregated_merged.rename(columns={'year': 'YEAR'}, inplace=True)
```

6

```
#print(final_df_aggregated_merged)

final_fa_df = final_df_aggregated_merged[[ 'STATE','rate', 'LAWTOTAL']].copy()

final_fa_df = final_fa_df.dropna(subset=['STATE'])

#print(final_fa_df)
```

1.2.4 Categorizing Gun Control Laws

As mentioned previously, the `LAWTOTAL` variable is the # of gun laws the state passed. We are using this as a proxy for gun law strictness.

Before assigning categories to my data, I wanted to analyze the spread of `LAWTOTAL`.

I created a plot of the distribution and got the summary statistics (commented out for cleanliness).

I then created 5 categories, from 0-4 (0 being lowest # of Gun Laws, 4 being heighest # of gun laws).

{ 0: 'Very Low # of gun laws', 1: 'Low # of gun laws', 2: 'Moderate # of gun laws', 3: 'High # of gun laws', 4: 'Very High # of gun laws' }

The cutoff points for each was detemined strategically based on the percentiles.

```
[ ]: lawtotal_stats = final_fa_df['LAWTOTAL'].describe()
     #print(lawtotal_stats)

     #sns.histplot(final_fa_df['LAWTOTAL'], kde=True)


     #plt.title('Distribution of LAWTOTAL')
     #plt.xlabel('LAWTOTAL')
     #plt.ylabel('Frequency')


     #plt.show()

     bins = [-float('inf'), 9, 20, 28, 40, float('inf')]
     labels = [0, 1, 2, 3, 4]

     final_fa_df['strict_cat'] = pd.cut(final_fa_df['LAWTOTAL'], bins=bins,
       ↪labels=labels, include_lowest=True)


     #print(final_fa_df.head())
```

1.3 Visualizations and Discussion

After cleaning the data, we now have a final dataframe with both mortality rate data and gun law data.

Now, I create visualizations to help answer the question.

1.3.1 Heatmap of Firearm Mortality per 100,000 Persons

First, I created a geographic heatmap showing the Firearm Mortality Rate (per 100,000 persons) by each state.

I used plotly package to create the graphs.

I had to convert rate to numeric to ensure plotly uses it as a continuous variable.

I used the Cividis color scheme to be accessible to colorblind individuals.

```
[ ]: final_fa_df['rate'] = pd.to_numeric(final_fa_df['rate'], errors='coerce')

     state_names_to_abbrev = {v: k for k, v in abbrev_to_state_names.items()}

     final_fa_df['STATE'] = final_fa_df['STATE'].map(state_names_to_abbrev)
     #print(final_fa_df['STATE'].unique())
     #print(final_fa_df['rate'].isna().sum())
     #print(final_fa_df['rate'].describe())




     rate_fig = px.choropleth(final_fa_df,
                     locations='STATE',
                     locationmode="USA-states",
                     color='rate',
                     scope="usa",
                     color_continuous_scale="Cividis",
                     labels={'rate':'Rate Value'})


     rate_fig.update_layout(title_text = 'Firearm Mortality Rate by State in the␣
       ↪USA')
     rate_fig.show()
```

As you can see, the states that stand out the most are LA, AL, NM, and MO. These states seem to have the highest Firearm Mortality Rate. More people seem to be dying in these states from gun violence.

The states with the lowest mortality rates are concentrated in the North East (NY, NJ, CT, MA, RI).

I wanted to display the top and bottom 5 states based on mortality rate in the following tables:

```
sorted_df2 = final_fa_df.sort_values(by='rate')

top_52 = sorted_df2.head(5)[['STATE', 'LAWTOTAL', 'rate']]
bottom_52 = sorted_df2.tail(5)[['STATE', 'LAWTOTAL','rate']]


styled_top_52 = top_52.style.set_table_styles(
    [{'selector': 'th', 'props': [('font-size', '12pt'), ('text-align',␣
 ↪'center')]}]
).set_properties(**{'text-align': 'center', 'background-color': '#f0f9e8',␣
 ↪'color': 'black'}).set_caption("Bottom 5 States by Firearm Mortality Rate")

styled_bottom_52 = bottom_52.style.set_table_styles(
    [{'selector': 'th', 'props': [('font-size', '12pt'), ('text-align',␣
 ↪'center')]}]
).set_properties(**{'text-align': 'center', 'background-color': '#0868ac',␣
 ↪'color': 'white'}).set_caption("Top 5 States by Firearm Mortality Rate")


display(styled_top_52)
display(styled_bottom_52)
```

<pandas.io.formats.style.Styler at 0x2b34fa1df00>

<pandas.io.formats.style.Styler at 0x2b35466e9b0>

The table confirms the findings of the previous heatmap.

The states with the Highest Firearm mortality rate are MS, LA, NM, AL, MO.

The states with the lowest mortality rate are RI, MA, HI, NY, NJ.

I missed HI in the map because it is separate from the continental US.

1.3.2   Heatmap of Gun Law Strictness Per State

Then I decided to graph a heatmap of the States by Gun Law Strictness level.

I used the categories that I previously defined.

Since this was a categorical variable, I couldnt use a continouous color scheme. Instead, I created
a custom color scale for each category based on the Cividis scheme to maintain its acceibility.

```
category_labels = {
    0: 'Very Low # of gun laws',
    1: 'Low # of gun laws',
    2: 'Moderate # of gun laws',
    3: 'High # of gun laws',
    4: 'Very High # of gun laws'
}
```

9

```python
final_fa_df['strict_cat_label'] = final_fa_df['strict_cat'].map(category_labels)


color_scale2_labels = {
    'Very Low # of gun laws': '#f8e23e',
    'Low # of gun laws': '#bfb06b',
    'Moderate # of gun laws': '#817f77',
    'High # of gun laws': '#4a536c',
    'Very High # of gun laws': '#032554'
}


law_fig = px.choropleth(final_fa_df,
                    locations='STATE',
                    locationmode="USA-states",
                    color='strict_cat_label',
                    color_discrete_map=color_scale2_labels,
                    scope="usa",
                    category_orders={"strict_cat_label": ['Very Low # of gun␣
 ↪laws', 'Low # of gun laws', 'Moderate # of gun laws', 'High # of gun laws',␣
 ↪'Very High # of gun laws']},
                    labels={'strict_cat_label': 'Gun Law Strictness Category'},
                    title='Strictness Category by State')

law_fig.show()
```

Alot of states have Low # of gun laws / less strict gun laws. Right off the bat some fo those states are GA, MS, KY, NM, AK, OK, KS, MO.

Some of the states with the most # of gun laws/ most strict gun laws are NY, MA, CT, NJ, MD, CA, WA, IL.

Again, I created tables to show the top and bottom 5 by # of gun laws.

```python
[ ]: sorted_df = final_fa_df.sort_values(by='strict_cat')

top_5 = sorted_df.head(5)[['STATE', 'LAWTOTAL', 'rate']]
bottom_5 = sorted_df.tail(5)[['STATE', 'LAWTOTAL','rate']]


styled_top_5 = top_5.style.set_table_styles(
    [{'selector': 'th', 'props': [('font-size', '12pt'), ('text-align',␣
 ↪'center')]}]
).set_properties(**{'text-align': 'center', 'background-color': '#f0f9e8',␣
 ↪'color': 'black'}).set_caption("5 States with Least Strict Gun Laws")
```

```
styled_bottom_5 = bottom_5.style.set_table_styles(
    [{'selector': 'th', 'props': [('font-size', '12pt'), ('text-align',␣
 ↪'center')]}]
).set_properties(**{'text-align': 'center', 'background-color': '#0868ac',␣
 ↪'color': 'white'}).set_caption("5 States with Most Strict Gun Laws")


display(styled_top_5)
display(styled_bottom_5)
```

<pandas.io.formats.style.Styler at 0x2b34fa291e0>

<pandas.io.formats.style.Styler at 0x2b34fa1dff0>

The 5 States with the Least Strict Gun Laws are AK, SD, OK, NH, MT.

The 5 states with the Most Strict gun laws are MA, CT, DE, RI, HI.

Right off the bat by looking at both map we can make a conclusion to our answer.

States that had a higher firearm mortality rate (represented as yellow in the map) were more likely to be on the less strict category (more yellow) when it came to gun law strictness map.

Applying the same logic, it seems that states with lower firearm mortality rate (blue on first map) were more likely to be more strict with gun laws (more blue in second map).

I conclude that yes, gun laws matter. It seems like the states that have enacted gun control legislation overall experience decreased firearm mortality when compared to states with less legislation.

1.3.3 Regression Model Trying to Predict Firearm Mortality Rate Using # of Gun Laws

I wanted to see if # of gun laws (the proxy for gun law strictness level) had a linear relationship with the firearm mortality rate.

I expect there to be a linear decreasing relationship, because we expect mortality to decrease as # of gun legislature / Strictness increases.

```
[ ]: plt.figure(figsize=(10, 6))
     sns.regplot(x='LAWTOTAL', y='rate', data=final_fa_df, scatter_kws={'alpha':0.
      ↪5}, line_kws={'color':'red'})

     plt.title('Regression of Firearm Mortality Rate on Total Gun Laws')
     plt.xlabel('Total Number of Gun Laws (LAWTOTAL)')
     plt.ylabel('Firearm Mortality Rate (rate)')

     plt.show()


     X = sm.add_constant(final_fa_df['LAWTOTAL'])
     y = final_fa_df['rate']
```
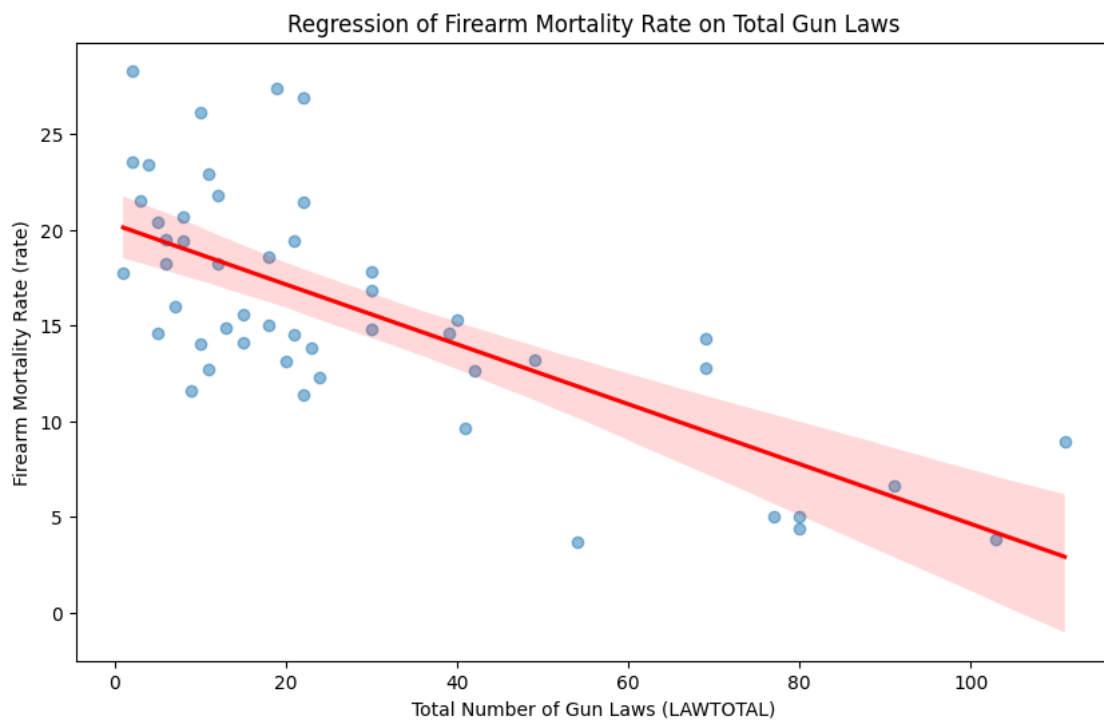
```
model = sm.OLS(y, X).fit()


model_summary = model.summary()


print(model_summary)


print(f"R-squared: {model.rsquared:.3f}")
print(f"p-value for LAWTOTAL: {model.pvalues['LAWTOTAL']:.3e}")
```


Regression of Firearm Mortality Rate on Total Gun Laws

OLS Regression Results

Dep. Variable:	rate	R-squared:	0.521
Model:	OLS	Adj. R-squared:	0.511
Method:	Least Squares	F-statistic:	52.21
Date:	Wed, 28 Feb 2024	Prob (F-statistic):	3.32e-09
Time:	21:01:45	Log-Likelihood:	-142.63
No. Observations:	50	AIC:	289.3
Df Residuals:	48	BIC:	293.1
Df Model:	1		
Covariance Type:	nonrobust		

```
==============================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const           20.2622      0.869     23.329      0.000      18.516      22.009
LAWTOTAL        -0.1563      0.022     -7.226      0.000      -0.200      -0.113
==============================================================================
Omnibus:                      2.190   Durbin-Watson:                   1.642
Prob(Omnibus):                0.335   Jarque-Bera (JB):                1.922
Skew:                         0.474   Prob(JB):                        0.382
Kurtosis:                     2.842   Cond. No.                        57.6
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
R-squared: 0.521
p-value for LAWTOTAL: 3.319e-09
```

There is in fact a linear, decreasing relationship between # of gun laws and the firearm mortality rate (as predicted).

52.1% (R-squared) of the variability in firearm mortality rates can be explained by the number of gun laws in different states. # of gun laws (or gun strictness) is a good predictor for firearm mortality.

The p-value was very small, suggesting that the number of gun laws has a statistically significant impact on firearm mortality rates.

1.4 Conclusion

In conclusion, stricter firearm control laws help reduce firearm mortality.

I have shown that you can use # of gun laws as a proxy for firearm control strictness. When using this as a proxy, states that enacted the most laws (AKA states with strict gun control) experienced a lower firearm mortality. States with less laws controlling firearms experienced higher firearm mortality.

Gun control legislation works and this data supports that.

For a future direction, I would like to compare Firearm mortality and gun strictness by nation. I am curious to see where the United States stands in comparision to other nations. That will put things more in context. After all, what if a low firearm mortality rate in the U.S. is high for Europe/ Asia?

1.5 Works Cited

Data Sources:

National Center for Health Statistics. NCHS - VSRR Quarterly provisional estimates for selected indicators of mortality. Date accessed 02/26/2024. Available from https://data.cdc.gov/d/489q-934x.

State Firearm Law Database: State Firearm Laws, 1991-2019 Siegel, Michael 2020, doi:10.3886/ICPSR37363.v1 Inter-university Consortium for Political and Social Research https://www.icpsr.umich.edu/web/NACJD/studies/37363/versions/V1/summary