Project: Investigate FBI Guns Dataset.

Table of Contents

Introduction

The data comes from the FBI's National Instant Criminal Background Check System. The NICS is used by to determine whether a prospective buyer is eligible to buy firearms or explosives. Gun shops call into this system to ensure that each customer does not have a criminal record or isn't otherwise ineligible to make a purchase. The data has been supplemented with state level data from census.gov.

In [2]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

Data Wrangling

> **Tip**: In this section of the report, I will load in the data, check for cleanliness, and then trim and clean your dataset for analysis.

General Properties

In [14]:
```
guns=pd.read_csv('gun_data.csv')
guns.head()
```

Out[14]:

	month	state	permit	permit_recheck	handgun	long_gun	other	multiple	admin	prepaw
0	2017-09	Alabama	16717.0	0.0	5734.0	6320.0	221.0	317	0.0	
1	2017-09	Alaska	209.0	2.0	2320.0	2930.0	219.0	160	0.0	
2	2017-09	Arizona	5069.0	382.0	11063.0	7946.0	920.0	631	0.0	
3	2017-09	Arkansas	2935.0	632.0	4347.0	6063.0	165.0	366	51.0	
4	2017-09	California	57839.0	0.0	37165.0	24581.0	2984.0	0	0.0	

5 rows × 27 columns

In [15]:
```
guns.shape
```

Out[15]:
```
(12485, 27)
```

In [16]:
```
guns.describe()
```

Out[16]:

	permit	permit_recheck	handgun	long_gun	other	multiple
count	12461.000000	1100.000000	12465.000000	12466.000000	5500.000000	12485.000000
mean	6413.629404	1165.956364	5940.881107	7810.847585	360.471636	268.603364
std	23752.338269	9224.200609	8618.584060	9309.846140	1349.478273	783.185073
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.000000	865.000000	2078.250000	17.000000	15.000000
50%	518.000000	0.000000	3059.000000	5122.000000	121.000000	125.000000
75%	4272.000000	0.000000	7280.000000	10380.750000	354.000000	301.000000
max	522188.000000	116681.000000	107224.000000	108058.000000	77929.000000	38907.000000

8 rows × 25 columns

In [17]:
```
census_data=pd.read_csv('census data.csv')
census_data.head()
```

rizona	Arkansas	California	Colorado	Connecticut	Delaware	...	South Dakota	Tennessee	Texas	Utah	Ver
31,071	2,988,248	39,250,017	5,540,545	3,576,452	952,065	...	865454	6651194	27,862,596	3,051,217	624
92,301	2,916,025	37,254,522	5,029,324	3,574,114	897,936	...	814195	6346298	25,146,100	2,763,888	625
8.40%	2.50%	5.40%	10.20%	0.10%	6.00%	...	0.063	0.048	10.80%	10.40%	-0

In [18]:
```
census_data.shape
```

Out[18]: (85, 52)

In [19]:
```
census_data.describe()
```

Out[19]:

	Fact	Fact Note	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	De
count	80	28	65	65	65	65	65	65	65	
unique	80	15	65	64	64	64	63	64	63	
top	Persons in poverty, percent	(c)	53.00%	7.30%	50.30%	50.90%	6.80%	3.30%	5.70%	
freq	1	6	1	2	2	2	2	2	2	

4 rows × 52 columns

In [20]:
```
print("GUNS DATAFRAME COLUMNS:")
print(guns.columns)
print("CENSUS DATAFRAME COLUMNS:")
print(census_data.columns)
```

```
GUNS DATAFRAME COLUMNS:
Index(['month', 'state', 'permit', 'permit_recheck', 'handgun', 'long_gun',
       'other', 'multiple', 'admin', 'prepawn_handgun', 'prepawn_long_gun',
       'prepawn_other', 'redemption_handgun', 'redemption_long_gun',
       'redemption_other', 'returned_handgun', 'returned_long_gun',
       'returned_other', 'rentals_handgun', 'rentals_long_gun',
       'private_sale_handgun', 'private_sale_long_gun', 'private_sale_other',
       'return_to_seller_handgun', 'return_to_seller_long_gun',
       'return_to_seller_other', 'totals'],
      dtype='object')
CENSUS DATAFRAME COLUMNS:
Index(['Fact', 'Fact Note', 'Alabama', 'Alaska', 'Arizona', 'Arkansas',
       'California', 'Colorado', 'Connecticut', 'Delaware', 'Florida',
       'Georgia', 'Hawaii', 'Idaho', 'Illinois', 'Indiana', 'Iowa', 'Kansas',
       'Kentucky', 'Louisiana', 'Maine', 'Maryland', 'Massachusetts',
       'Michigan', 'Minnesota', 'Mississippi', 'Missouri', 'Montana',
       'Nebraska', 'Nevada', 'New Hampshire', 'New Jersey', 'New Mexico',
       'New York', 'North Carolina', 'North Dakota', 'Ohio', 'Oklahoma',
       'Oregon', 'Pennsylvania', 'Rhode Island', 'South Carolina',
       'South Dakota', 'Tennessee', 'Texas', 'Utah', 'Vermont', 'Virginia',
       'Washington', 'West Virginia', 'Wisconsin', 'Wyoming'],
      dtype='object')
```

We Are Gonna drop some useable columns and clean our DataFrame.

In [48]:
```
guns.drop()
```

Out[48]:

	Fact	Alabama	Alaska	Arizona	Arkansas	California	Colorado	Connecticut	Delawar
0	Population estimates, July 1, 2016, (V2016)	4,863,300	741,894	6,931,071	2,988,248	39,250,017	5,540,545	3,576,452	952,06
1	Population estimates base, April 1, 2010, (V2...	4,780,131	710,249	6,392,301	2,916,025	37,254,522	5,029,324	3,574,114	897,93
2	Population, percent change - April 1, 2010 (es...	1.70%	4.50%	8.40%	2.50%	5.40%	10.20%	0.10%	6.00%
3	Population, Census, April 1, 2010	4,779,736	710,231	6,392,017	2,915,918	37,253,956	5,029,196	3,574,097	897,93
4	Persons under 5 years, percent, July 1, 2016, ...	6.00%	7.30%	6.30%	6.40%	6.30%	6.10%	5.20%	5.80%

5 rows × 51 columns

Let's check which capita has the highest gun trend(Mix some columns up)

In [ ]:
```
guns['']
```

> **Tip**: You should *not* perform too many operations in each cell. Create cells freely to explore your data. One option that you can take with this project is to do a lot of explorations in an initial notebook. These don't have to be organized, but make sure you use enough comments to understand the purpose of each code cell. Then, after you're done with your analysis, create a duplicate notebook where you will trim the excess and organize your steps so that you have a flowing, cohesive report.

> **Tip**: Make sure that you keep your reader informed on the steps that you are taking in your investigation. Follow every code cell, or every set of related code cells, with a markdown cell to describe to the reader what was found in the preceding cell(s). Try to make it so that the reader can then understand what they will be seeing in the following cell(s).

Data Cleaning (Replace this with more specific notes!)

In [35]:
```
# After discussing the structure of the data and any problems that need to be
#   cleaned, perform those cleaning steps in the second part of this section.
```

Exploratory Data Analysis

> **Tip**: Now that you've trimmed and cleaned your data, you're ready to move on to exploration. Compute statistics and create visualizations with the goal of addressing the research questions that you posed in the Introduction section. It is recommended that you be systematic with your approach. Look at one variable at a time, and then follow it up by looking at relationships between variables.

Research Question 1 (Replace this header name!)

In [ ]:
```
# Use this, and more code cells, to explore your data. Don't forget to add
#   Markdown cells to document your observations and findings.
```

Research Question 2 (Replace this header name!)

In [ ]:
```
# Continue to explore the data to address your additional research
#   questions. Add more headers as needed if you have more questions to
#   investigate.
```

Conclusions

> **Tip**: Finally, summarize your findings and the results that have been performed. Make sure that you are clear with regards to the limitations of your exploration. If you haven't done any statistical tests, do not imply any statistical conclusions. And make sure you avoid implying causation from correlation!

> **Tip**: Once you are satisfied with your work, you should save a copy of the report in HTML or PDF form via the **File** > **Download as** submenu. Before exporting your report, check over it to make sure that the flow of the report is complete. You should probably remove all of the "Tip" quotes like this one so that the presentation is as tidy as possible. Congratulations!

In [ ]: