Project 1 Udacity

- **Dataset**
  The analyzed dataset is FBI Gun Data found in
  https://www.google.com/url?q=https://github.com/BuzzFeedNews/nics-firearm-background-checks/blob/master/README.md&sa=D&ust=1532469042126000

- **Research Questions**
  1. Which states perform the highest number of background checks in 2015?
  2. Which states have had the highest growth in gun registrations?
  3. What is the overall trend of gun purchases?

- **Description**
  - First Question
    1. Since records are every month of the year, in order to get the year, the value 2015 was specified. That gives the a dataframe for 2015 only.
    2. In order make an aggregation on each state, groupby state took place.
    3. Specifying which column to have the summation aggregation on which is totals, and then perform sum()
    4. Since the values are relatively big, and in order to pleasant in the graph, values are divided by thousand.
    5. Horizontal bar chart, on the x-axis total number of background checks in 2015, and on the y-axis, all states.

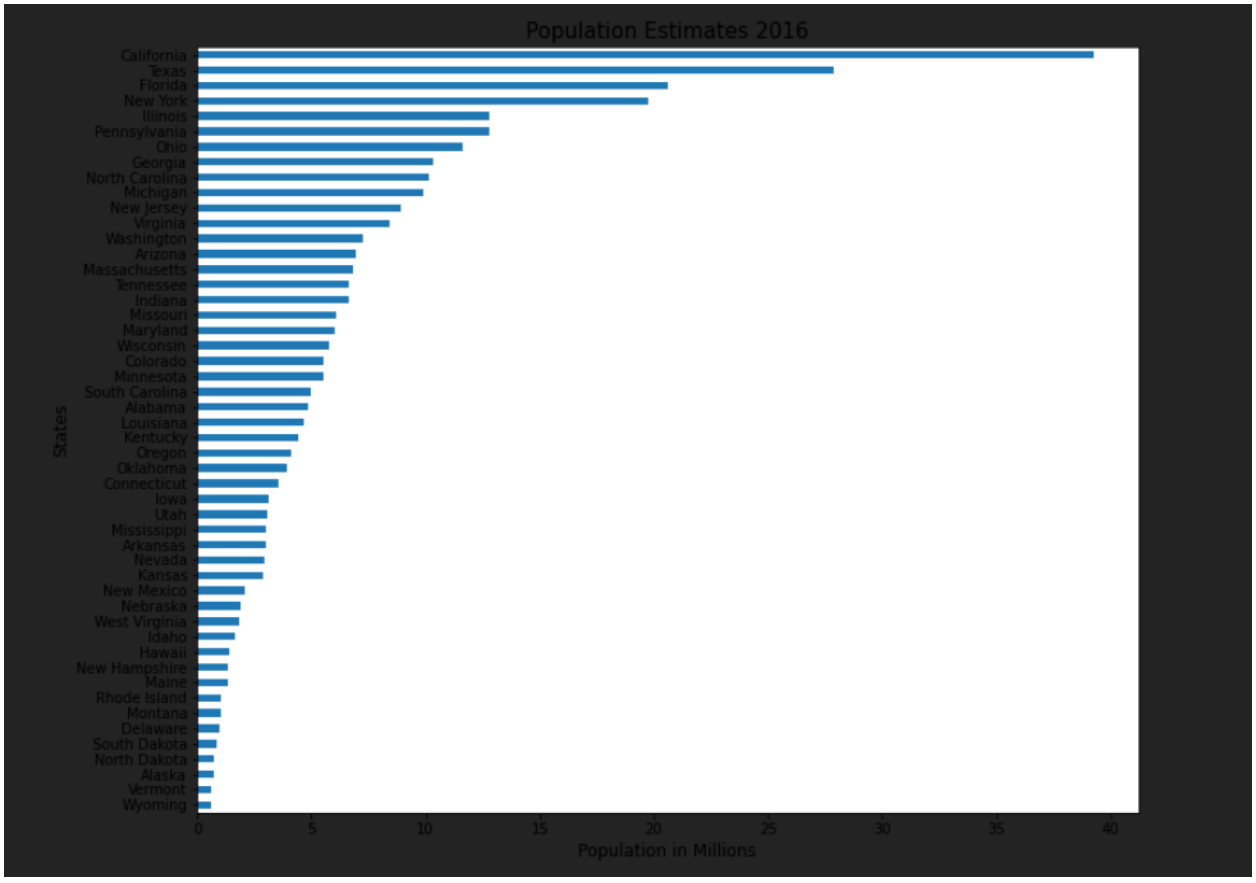
  - Second Question
    1. In order to take a look at the growth in gun registrations (which should be positively correlated with background checks or the same meaning), figure is declared firstly.
    2. A for loop with each state, and within the for loop one line of code specifying a state, and grouping by the year column, and again the totals is summed up, and finally plotted.
    3. The final plot contains all the states' lines of growth. Colors are not that distinguishable which is not convenient.
    4. For practicality, only the highest 5 states in totals will be specified by creating a dictionary with the state as the key and the mean of the totals of background checks as the value.
    5. After highlighting the highest 5 states, they are line plotted.

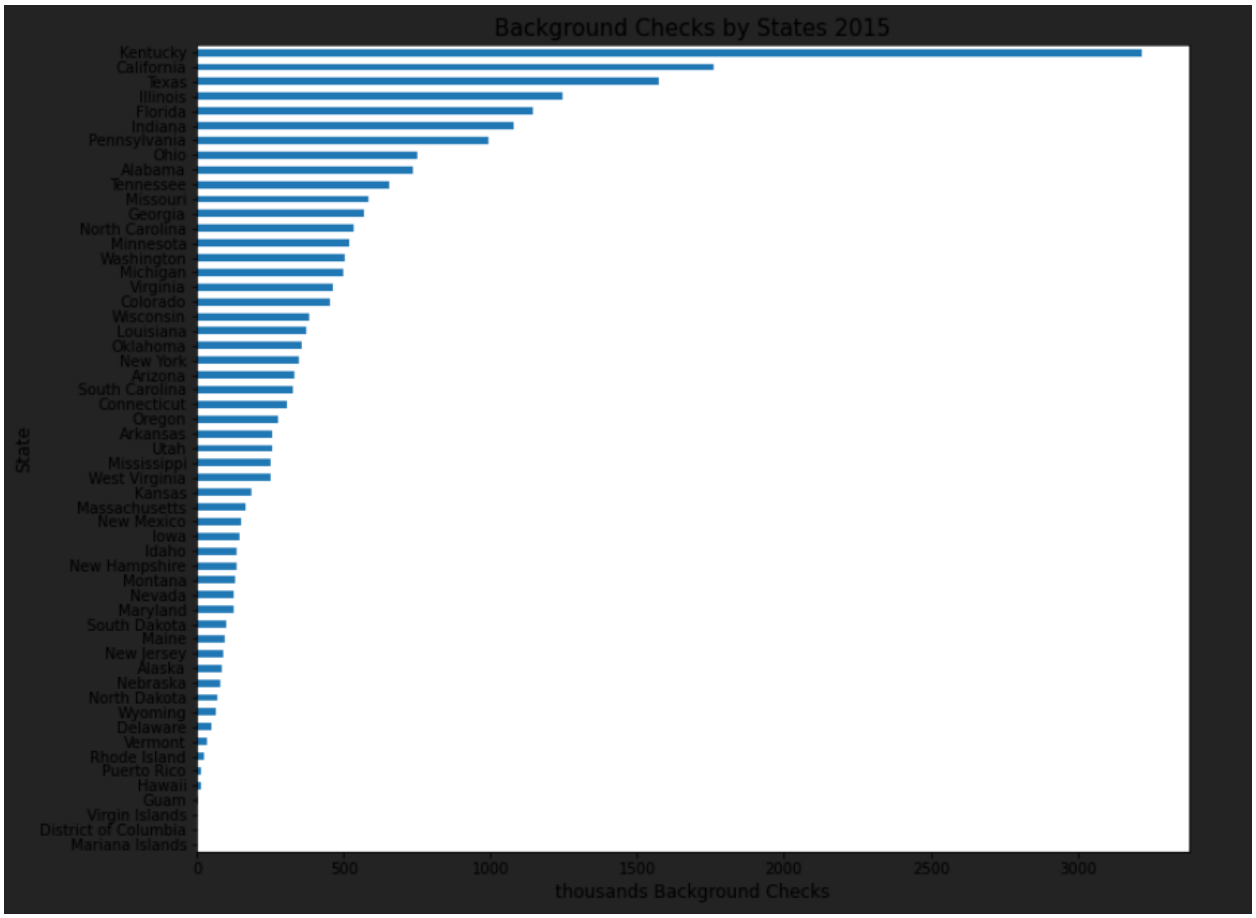
  - Third Question
    1. In order to get the overall trend, no states should be specified, but the groupby year is required.
    2. Column totals is specified for the summation aggregation.
    3. Plotted line with year on the x-axis and y is the number of background checks in millions.

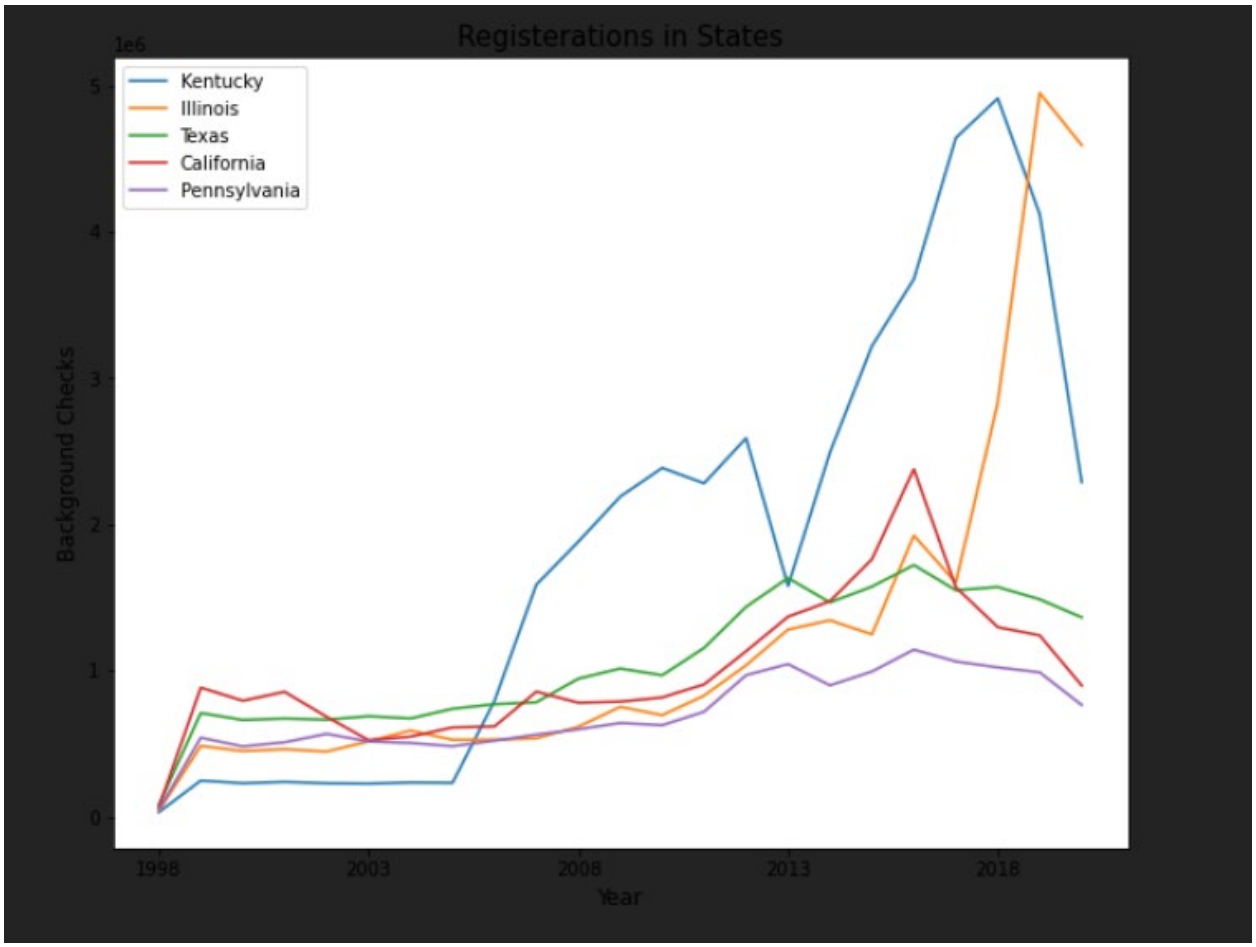
- **Data Wrangling**
  - gun_df
    1. Some columns have more than 1000 NaN values, these columns are not informative and will not be used, they were dropped.
    2. The column month has this format yyyy-mm, a new column called year was added with only yyyy, and the month column became with the format mm.

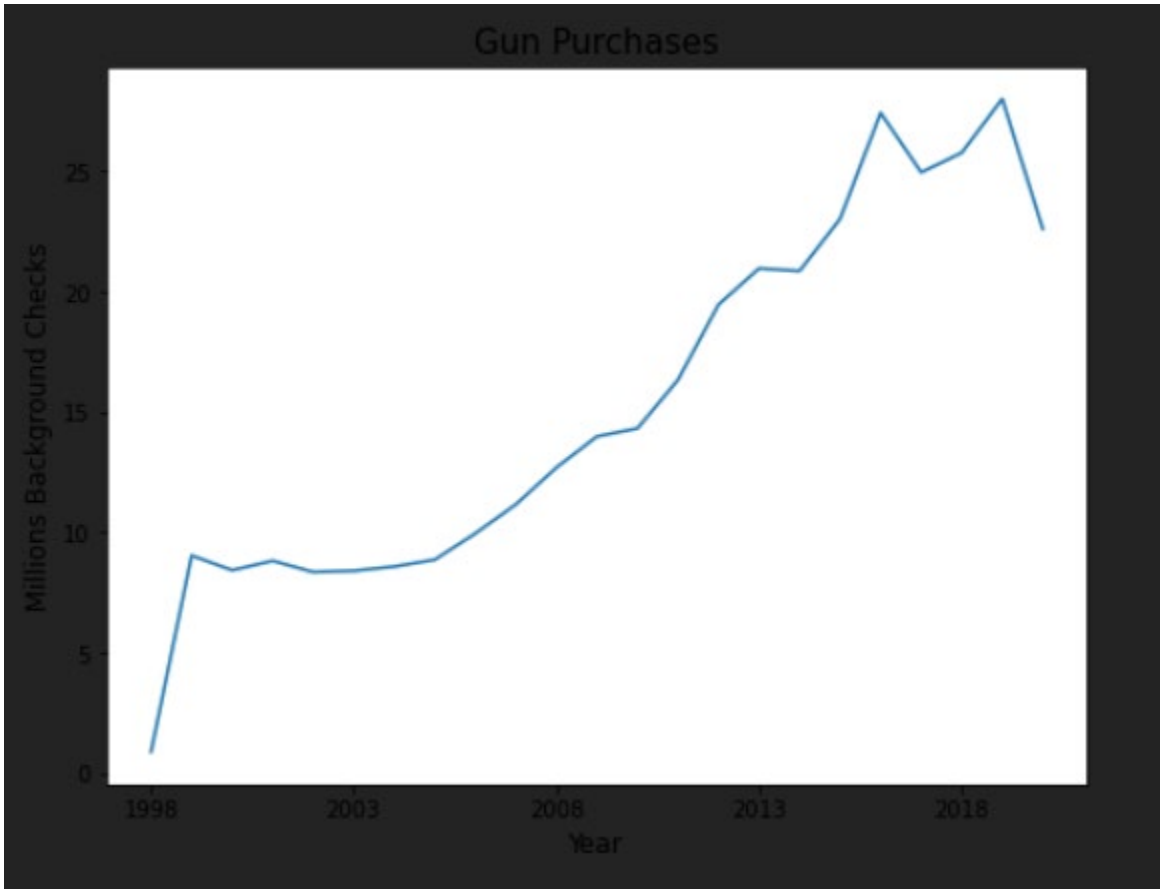
  - census_df
    1. 20 rows were all NaN values in all the states (columns), all were dropped.
    2. The column Fact Note was not informative, so it was dropped.
    3. The Fact column has description to what the corresponding fields in the row represent in the states columns, also the year it was recorded at, and if the fields are change from a year to another, and others does not have year value at all, that's why dataframe was divided into 3; census_0y_df, cesus_1_df, census_2y_df where the number represents the number of values of years in the Fact column.
    4. The data is object (string) with commas, dollar signs, percentage signs, and others. Cleaning all that by replacing them with '', and converting their type into float.

- **Plots**




Background Checks by States 2015
Kentucky
California
Texas
Illinois
Florida
Indiana
Pennsylvania
Ohio
Alabama
Tennessee
Missouri
Georgia
North Carolina
Minnesota
Washington
Michigan
Virginia
Colorado
Wisconsin
Louisiana
Oklahoma
New York
Arizona
South Carolina
Connecticut
Oregon
Arkansas
Utah
Mississippi
West Virginia
Kansas
Massachusetts
New Mexico
Iowa
Idaho
New Hampshire
Montana
Nevada
Maryland
South Dakota
Maine
New Jersey
Alaska
Nebraska
North Dakota
Wyoming
Delaware
Vermont
Rhode Island
Puerto Rico
Hawaii
Guam
Virgin Islands
District of Columbia
Mariana Islands
State
0
500
1000
1500
2000
2500
3000
thousands Background Checks


Registerations in States
1e6
Kentucky
Illinois
Texas
California
Pennsylvania
Background Checks
Year
0
1
2
3
4
5
1998
2003
2008
2013
2018


Gun Purchases
Millions Background Checks
0
5
10
15
20
25
1998
2003
2008
2013
2018
Year