Data Viz Individual Project – First Redesign

Kailin Hu

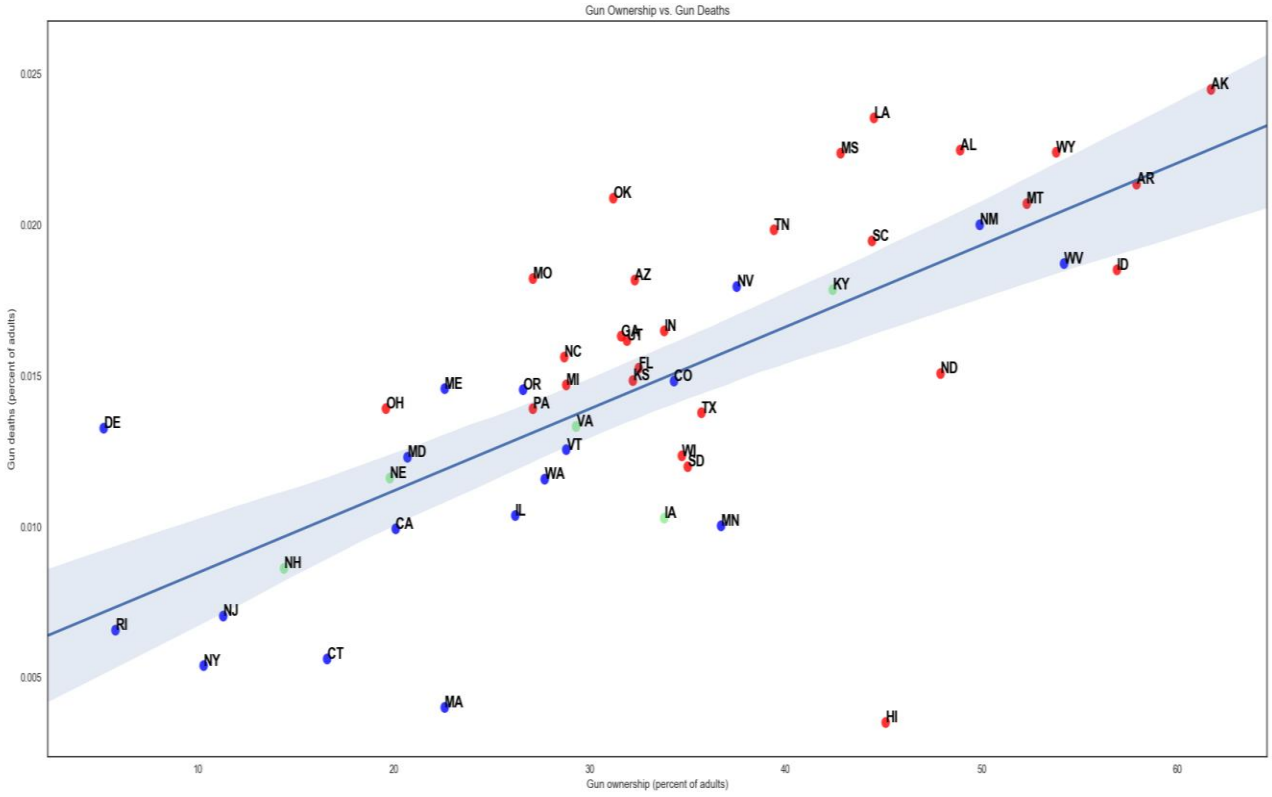
Introduction

The following graphs are modeled after graphs presented in this Vox article:

https://www.vox.com/policy-and-politics/2017/10/2/16399418/us-gun-violence-statistics-maps-charts

The data used to recreate the graphs was obtained from the same sources as those in the article as much as possible.

1. GRAPH 5



Differences from original graph: The redesign looks quite similar to the original graph, even though now both variables (gun ownership and gun deaths) are expressed in percentages. A new variable is also visualized here: the dominant political party of each state, represented by the color of each point. Blue means Democratic, red means Republican, and green means 'Other' (unicameral or split). This provides more context and a potential explanation for differences in gun ownership rates across different states.

Findings: The graph does show that as gun ownership rates increase, the gun death rates increase as well. It also shows that most of the Democratic states have lower gun ownership and death rates, while

most of the Republican states have higher rates. This suggests that a state's political party may have an impact on gun ownership and death rates.

Workflow:

1. *Obtaining Data*

To recreate this graph, we need data on number of gun deaths by state and also number of adult gun ownerships by state. The data sources used are listed below.

> Data for gun ownership in 2013 was obtained from this published literature, also used in the original graph:
>
> http://injuryprevention.bmj.com/content/injuryprev/early/2015/06/09/injuryprev-2015-041586.full.pdf?keytype=ref&ijkey=doj6vx0laFZMsQ2T
>
> The following is a description from the article of the data collection method: "*This study was performed using the 4000 respondents who were identified as a nationally representative cohort.*" All respondents surveyed were over age 18. The data was presented as rates (number of people owning guns out of number of total people sampled per state). There is no spreadsheet available for the data presented in this study, so the data for each state was manually entered into Excel.
>
> Data for gun deaths from 2013 was obtained from the CDC website:
>
> https://webappa.cdc.gov/cgi-bin/broker.exe
>
> The following filters were used on the WISQARS site using the "Fatal Injury Reports, National, Regional and State (RESTRICTED), 1999-2016": cause of injury = Firearm; census region/state = [selected each state separately]; years of report = 2013; age groups = custom age range: 19 to 85+. The data outputs included both counts and rates, so I just used the rates.

Fortunately, the data required was clean and simple already so no data wrangling or conversion was required. The gun ownership rates and gun death rates were combined together into one Excel spreadsheet.

2. *Creating the Visual*

The scatterplot was created in Python with the **seaborn** package. The Python code and data used to generate the plot can be found in the Github repository: https://github.com/wooohu/Data-Viz-Project

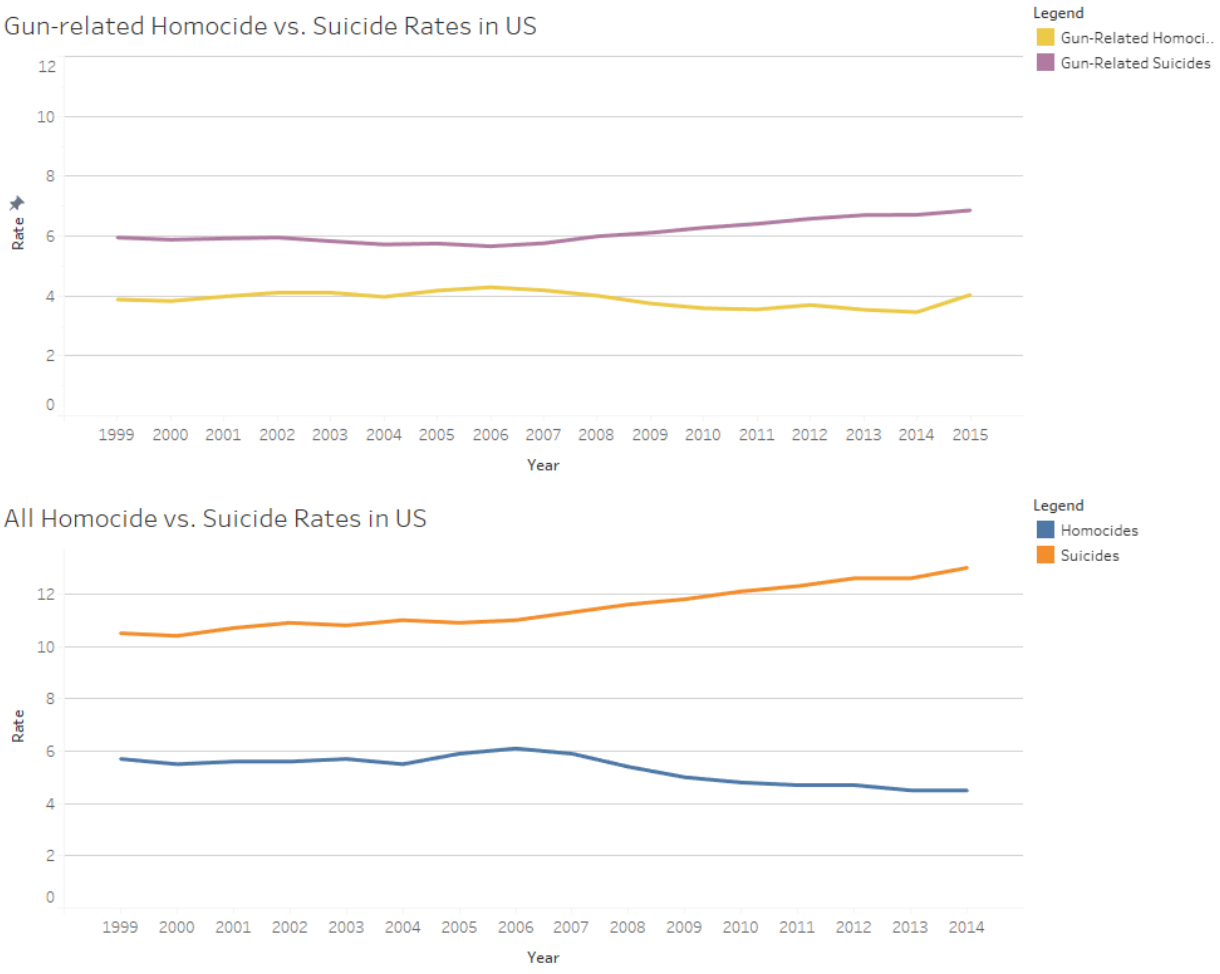
Future features/enhancements:

It is not exactly clear from the *Injury Prevention* article how the respondents were selected. This is a concern because if the number of respondents selected per state is not proportional to the total population of that state relative to others, then this data is flawed. Population varies greatly among different states, so each state should not be given the same weight in the sample especially if the results are presented as percentages of the population.

For the next version, I would like to obtain data for 2013 gun ownership by state from a different source that does account for the difference in population among different states. The same type of plot will be used to represent the data.

Some aesthetic aspects also need enhancing. For example, the axis titles should be in larger font and there should be a legend in the bottom-right corner to explain what each color represents.

2. GRAPH 10



Differences from original graph: The redesign includes line graphs of both the total homocide and suicide rates in the US as well as firearm-specific homocide and suicide rates, from 1999 to 2014. The total rates were included as a baseline so we can see whether the firearm-specific trends follow the overall trends. The dots from the original line graph were also removed, presenting smoother and more easily interpretable lines. One important thing to note is that while the original graph used counts, the

redesign uses rates because they were more easily accessible online, especially since WISQARS were temporarily down.

Findings: It turns out that the firearms lines follow the overall lines quite well over the years, weakening the original graph's claim that more access to guns lead to more suicides. If this were true, the trends of the lines would be different between the two graphs. Additionally, in both graphs here, the suicide line is noticeably above the homocide line.

Workflow:

1. *Obtaining Data*

Data on total homocide and suicide rates in the US were found in this CDC report:

https://www.cdc.gov/nchs/data/databriefs/db241_table.pdf#1

The report only gives data up through 2014 so the same year range was used for gun-specific rates. The rates were provided so no data transformation was needed.

The gun-specific data was obtained from the WISQARS website: https://webappa.cdc.gov/sasweb/ncipc/mortrate10_us.html

On the site, I went to "Fatal Injury Reports, National and Regional, 1999-2015" and used the following filters: intent/manner of the injury = "Homocide", cause or mechanism = "All injury", years of report = 1999 to 2014. Again, the rates were provided so no data transformation was needed.

These two datasets were then combined into one Excel spreadsheet and read into Tableau.

2. *Creating the Visual*

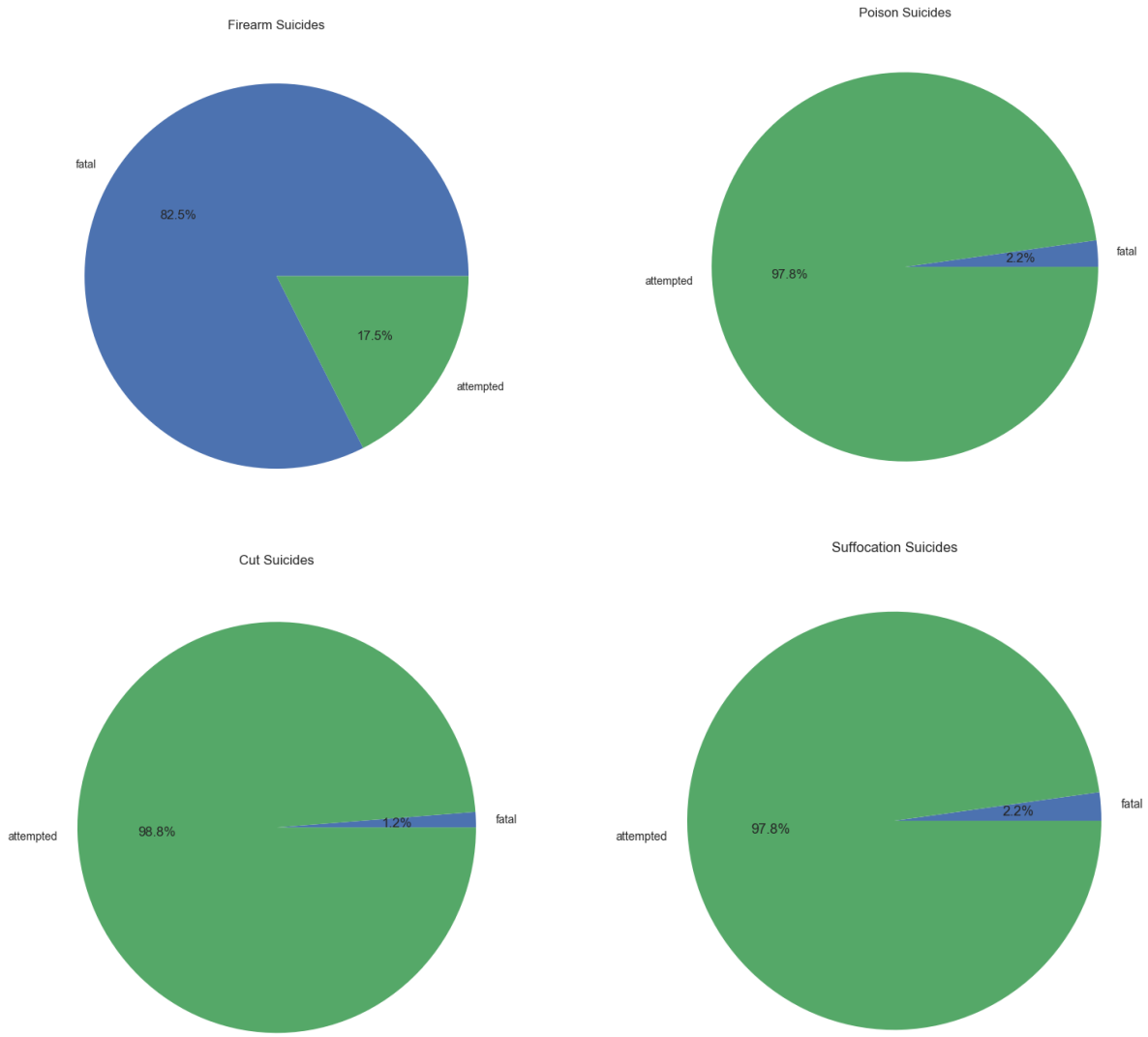
The spreadsheet was read into Tableau and used to generate line graphs. Creating the first version of the graphs was relatively easy. One sheet was made for each of the two graphs. The 'Year' dimension was added to Columns, and the 'Homocide Rate' and 'Suicide Rate' were added to Rows.

One challenge was getting the two lines (homocides and suicides) to be plotted on the same axes for easy comparison rather than stacked on top of each other, so I dragged one of measures onto the y-axis of the measure already plotted. Then I selected "Synchronize axes".

Future features/enhancements:

For the next version, if the WISQARS site is back up, I would like to get the actual number of homocides and suicides, plot them, and compare to the current version I have, which is based on the rates. I would also like to display both the homocide and suicide numbers when hovering over any given point along the lines. Right now, those numbers are displayed separately. Additionally, maybe it would be better to represent both sets of line graphs with the same two colors so they will be more consistent and easier to compare.

3. Graph #12



Differences from original graph:

The original graphs only included data from Indiana between 1990 to 1997. This data is from a collection of hospital records from 8 different states during the same time period. I also added a Suffocation pie chart because this was actually the second most common method of suicide from the study. Unlike the original pie charts, the redesigns also have labels for attempted versus fatal, which makes them much easier to interpret.

Findings: Even though the data was from a different source, these recreated pie charts look quite similar to the original, with the percentage of deaths from firearms being much higher than those for other suicide methods.

Workflow:

1. *Obtaining Data:*

> Because WISQARS was down, data was instead obtained from an article in the American Journal of Public Health titled "Previous Suicide Attempt and Its Association With method Used in a Suicide Death", by R. S. Spicer and T. R. Miller. The article can be found here:
>
> *https://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.90.12.1885*
>
> The study includes suicide fatalities and attempts pooled from 8 different states' hospital records from 1989 to 1997, which is a very similar timeframe to the one used in the original graph. The relevant data was manually entered into Excel and then read into Python. While finding the data source was difficult for this redesign, no data wrangling was required because the article already had the suicide numbers broken down by "attempted" and "fatal" for each suicide method. The data also did not need to be converted to rates for this one due to the nature of pie charts.

3. *Creating the Visual*

The pie charts were all made in Python using the **matplotlib** package. The Python code and data used to generate the plot can be found in the Github repository: https://github.com/wooohu/Data-Viz-Project

Future features/enhancements: I am curious to see how the data looks for a more recent time period, for instance 2005 to 2015. For the next version, I would like to obtain data from those years and compare them to the current version to see if there are any big differences.

In terms of aesthetics, the next version should have larger font for all the labels and titles.