**Final Project – Working Journal and decision-making notes/ramblings.**

**3/25/2023** – (1 hour) Started to find and evaluate data sources. This was cut very short, a little over an hour in, due to power outage and network outage that lasted well into the evening hours. I was still working on narrowing things at this time, but have a few target focal points for the data.

1. Finding the data around Rifle (typically “long guns”) uses in assaults/homicides vs other firearms to understand why there is outrage.
2. Pivot this to evaluate a rise/fall in the past 5 or so years in gun violence and it’s relationship to Law Enforcement Officers (LEO’s) employment counts. This appears at surface level to be on the decline.
3. Finding both firearm related deaths/assault numbers and comparing them to firearm sales statistics to validate or discredit the idea of “more” guns equates to more violent crime.

Again, seeing the target being law makers/government officials to push for changes in accordance with what the data reveals.

**3/27/2023** – (2.25 hours) Finally back online and looking to nail down data sources. This seems to be a bit of a quagmire when looking to find this data. I have found some areas where I might be able to pull this down and will be working to identify if I can get the data I’m looking for. As forecast, by you Dr. Silva, this is proving to be a bit daunting as finding this data isn’t as easy as you would think and has started to become a bit overwhelming.

Working to collect data from here: [https://cde.ucr.cjis.gov/LATEST/webapp/#](https://cde.ucr.cjis.gov/LATEST/webapp/) gathering annual data files for “weapon” use. Will look to get data for the past 5 years that I can get starting with 2021. This will require some more data work to get these “together” as they’re individual files per year. Correction, this is limited to only 2020 and 2021. Not ideal. (

**3/28/2023** – (3.5 hours) Back looking again to find any good data source that will help to tell ANY relevant 2nd Amendment story at this point. Getting a bit discouraged with the results from the CDC and FBI resources at those are basic reports rather than raw data. I will need to continue to dig in order to better identify if the data can tell the story I am trying to tell?

Again, looking for another 2 hours with no reasonable data set that isn’t in a report to work with. I have tried cdc data, fbi data, atf data, and anything else I can find by searching. Ready to give up on this topic all together and just to get something to work with at this point. I don’t have an additional 5-6 hours to keep looking for this same data that I’m no able/capable of sourcing.

Found some FBI data that indicates weapons used in homicides from 2015-2019 for all 50 states, DC and US Virgin Islands. These files are only available per year and will need to combine them. This data doesn’t really tell that great of a story as it relates to supporting the 2nd Amendment. This is the most common weapon used, but is broken down into subcategories like handgun, rifle, shotgun, and Firearm “other”. I’m going to need to look deeper into this data and see if it’s helpful to tell a story here. Data was gathered here by going to the individual years 🡪 Crim in the U.S. “YEAR” 🡪 “Offenses Known to Law Enforcement section 🡪 “Weapons” 🡪 Murder – Weapons “Table 20”. <https://ucr.fbi.gov/crime-in-the-u.s>   
  
I’m hoping that this is an acceptable step and in the interest of time, I’m going to use Microsoft Access to mash together these spreadsheets quickly and allow me to add in a field for the year that is not currently present in the source data. I’m still not too confident that I can work with this data set to really tell a compelling story, but I have a start to at least begin to evaluate.

I have now created a new .xls file that I’ll likely save as a CSV and clean up the field names for import. Currently they have spaces in them which seems to be problematic at times. Now after about 3 hours I think I have some data to examine to see what it provides. I’ll be looking to work with this csv in a jupyter notebook and do some data manipulation there for getting better visuals.

**4/1/2023 –** (1 Hour) Ok, worked with the data to get a decently clean file for use to try to see what visualizations in Plotly may be useful. I can’t say I’m happy with the limited extent of the data I have collected, I might look to gather more historical years and potentially have 10 years of history (only considering this if what I have doesn’t show a story).

I began my journey just plugging my data into the charts that I had already put together for the Plotly session that we had led. I just wanted to get a baseline of how the data was shaping up and what the charts could look like based of our in-class examples. Clearly these aren’t going to be stellar, but interesting either way. A couple in, not sure this is the right way to use this data. Going to try a few things later today and see what I get. I think I can just leverage the source data without the “describe” as it provides zero value to the visualization. Will regroup and attempt some visuals later, I think they’re likely to be easier to pull potentially. Will see.   
  
Lost power from 4/1/2023 at 1pm until 4/2/2023 at roughly 5pm. The household requirements took over my time and was not able to revisit this again.

**4/4/2023** –

This might be helpful to get some “scores” up for strictest laws and least restrictive laws: <https://worldpopulationreview.com/state-rankings/strictest-gun-laws-by-state>. I might look to add this to my tables to have an additional

**4/7/2023**

Spent some time reviewing the Dr. Silva’s findings via the API calls into the FBI data. This doesn’t look to be all that dissimilar to what I’ve gathered from here [FBI — Crime in the U.S.](https://ucr.fbi.gov/crime-in-the-u.s) on 3/28. While helpful, as Dr. Silva added, not sure if it helps me much.  
  
I did do some additional digging into that site and did consider another angle as it pertains to the number of police officers per state. I’m going to gather that data for the years I have 2015-2019 and perhaps attempt to see any correlation to number of officers to gun crime or just homicides in general.

Started to gather the Police officer tables, wonderfully 2016 data does not exisit as it does for all other years I’ve gathered. So, I may continue this path or I may scrap doing this route too. Perhaps this is a perfect fit to see that we can identify 2016 data as “NONE” or 0 for graphing purposes as we recently learned in class.   
  
I’m going to work tomorrow on getting these new tables to combine with my existing table and look to build this source data further.

**4/8/2023**

(Morning 2 hours) - So, after doing some thinking last night, perhaps for 2016 police numbers I can take 2015 and 2017 and simply get the average between the two and leverage that number for 2016 counts. While a little goofy and graphically would be linear, it would likely be a close number to actual and provide me one additional data point for visual reference. At this point, I would like to have data that may be “reasonably” wrong there rather than nothing for visual representation for this project.

\*\*\* I would like to go on record stating that I DO NOT like the idea of “fudging” the data, but in this case it’s a calculated assumption that I’ll be making in the interest of having a quasi-complete dataset that should have a minimal margin of error here.

I’ll be working on getting the csv expanded to include this new police data for each year today. I hope to get some time to work on starting some of the Plotly visuals in hopes of finding a telling story.

Been working via Microsoft Access to mash the files together. I’m working from 5 original Homicide weapon .xls and now 4 LEO .xls files via access to get them all together into one table. I have been able to get the calculated average for 2016 LEO information as indicated above. I’m not sure I really like the way that’s going to flow in between years….I might toss in some minor random factoring to give is a little variation rather than the linear nature of the resultant numbers. This did result in some previously unobserved mission data. I’m thinking about how I want to populate that data, once again, as discussed in class this week.   
  
Funny enough to me, Florida information is what’s missing from the Homicides table. This is somewhat funny, as I could have focused on the “Florida Man” angle/story here…very disappointing. HAHAHAHA 😊

(Afternoon plan) I’m going to get this into a jupyter notebook to start working with visualizations at this point. I’m going to attempt to find something of value in the data that I’ve been able to compile. I think the original data may have some stories as it pertains to protecting rifles considering their limited use vs handguns. I’m also going to investigate how to potentially leverage the use of MAPS via Plotly to see if that will allow me to give a nice representation of data considering the states are a “sorting” category for me presently.

Sadly, my afternoon got re-arranged with household demands. I’ll be working on getting those initial visualizations together throughout the week this week.

-------------- Project Milestone 1 --------------------------------------------------------------------------------------------

**4/11/2023** – Class night discussion and findings. Desire to use maps to represent the data for each state and will require the use of the following file: <https://www.kaggle.com/datasets/pompelmo/usa-states-geojson/code> along with this <https://plotly.com/python/mapbox-county-choropleth>

Leverage this mathematical path to calculate ratio of per “100,000 residents”

A : B = C : D where D will equal 100,000 as it commonly does.   
Thus : C = D \* (A/B)

Post class time conversation with you Dr. Silva, you had indicated that the FBI information is limited due to privacy restrictions. This provides an interesting variable to throw into the mix of the data that I did collect.

**4/12/2023** – Ok, tonight I’m going work on cleaning my data and adding some additional data points worth examining. I’m going to attempt to accomplish this via Python and create new rows for these calculated fields as indicated in the ratio notes from above. These calculations will go against the estimated population data to calculate the total gun homicide ratio per 100,000 residents and the policing ratio per 100,000 residents.   
  
I hope to move to some visualizations before the night is out, I’ll be looking to see how to deal with some of the data that is blank. Specifically, data for Florida, which is mostly missing for the years I’ve collected as it pertains to homicide information. I also have some states that look to have data that doesn’t seem to be right. I’ll have to look at the visualizations to validate some things, but I am not confident that I have quality data or even complete data. My current assumption at this point is that this data is erroneously reported for a handful of states.

Ok, I have added my ratio calculated fields into my data frame. Took a few trial-and-error attempts but was able to successfully get these added into the jupyter notebook. Now I’m going to try and get some basic visualizations to attempt to see correlations here. “Fast forward a good hour +” I’m starting to play around a little with the “Plotly GO” options and needed to add that to the initial load in. Visuals are progressing and just playing around now to find ANYTHING that looks promising and seems to stand out at present. I may consider top 5 and bottom 5 states in some of the calculated fields to pick out for that visualization as all 50 states + DC gets muddy in many cases. I’m also finding that the “NAN” values in some areas are affecting my ability to represent things like “size” for a point in a scatter plot due to NAN values. I will need to evaluate the impact of substituting data there with something like a 0 or something else. I still think the “MAP” look will really be the best option, so may also look to spend time getting that happening.   
  
Using the following as a guide to try and pull in my geoJSON file for the mapping: <https://datatofish.com/load-json-pandas-dataframe/> . My attempts to get the mapping working is not going anywhere. I believe I’m having an issue with how I’m bringing in the geoJSON file here to make the connection to my source data in the data frame that I’m working with. I’m not at about 3.5 hours this evening and I need to call it quits.   
  
Dr. Silva, this is the night prior to our Teams meeting (4/13/2023). Hopefully when we connect I can bring up these issues and see if you might have some insight here to get the “map” charts to work.

**4/13/2023** – Considerations for a PowerPoint presentation? [https://pypi.org/project/plotlyPowerpoint](https://pypi.org/project/plotlyPowerpoint/)

Had an excellent meeting with Dr. Silva to get the map box/choropleth functionality working. The issue was around the way I was reading in the geojson file and placing it into a data frame which does NOT work with the map functionality. Also, check your zoom level and centering coordinates. Once corrected, the visualization mapping worked perfectly. This should prove useful in my attempts to tell a compelling story.

Observationally, “nan” values were being interpreted as zeros when heat mapping. This is going to give some misleading visualizations for states like Florida and Alabama. This continues to indicate that the quality of the data obtained from the FBI website isn’t likely to be the most reliable. I will need to create additional visualizations where I change up what value is in place of “nan” to try to address these clear inconsistencies.

**4/15/2023** – Getting back into jupyter notebook today to work on the mapping visualizations. I’m going to attempt to work on the “mapbox\_style” to see if another version will better represent the United states to include Alaska and Hawaii easily.

I’ve started to look into changing the NAN values to “0”. This has resulted in some of my calculated values to return an inf value. It would appear that I can make a change to these values yet again by doing a replace statement found here: <https://www.statology.org/pandas-replace-inf-with-0> I’m going to attempt this in the jupyter notebook and see if this will clear up those values. BINGO, this worked as hoped. Will continue down the path of zero values.   
  
Playing around some with different options within the visualization output. Finding that the opacity is a nice touch and will look to incorporate that further. Lots of progress here with different data getting “mapped” out and finding some mildly interesting directions to take this. The raw number data shows some compelling things, but the per 100000 residents shows different vantage points. I’m still working on narrowing my focus. This has me working in a good direction, and I’m considering using -1 or lower to indicate states that don’t have good data and find a color scheme that will color those outliers accordingly.

-Took a break-

Back at it and looking at changing zero values to -5 or -1 or similar and allowing that to determine how some of the missing information is shown in the visualizations. I’ll go back up into the notebook to work with the original data and create another data frame that has -1 for the nan values. I might have to play with this to be “more” negative to further separate it from the rest of the actual data that I do have. Now that I’ve implemented the -1 data, I think this is messing with some of the calculations, and things aren’t coming out right from what I can see. I may be sticking with the zero value and simply indicate the states that don’t have complete data.

The maps have been able to provide a few options for the focus of my “story” I think I’m full circle back to the unjustified focus on rifles. Nearly all other “weapons” out pace that of those homicides that use a rifle. I did comparative calculations for knives and blades, Pistols, and Hands feet etc (bare hands) and all are used more for homicides than that of a rifle.

-------------- Project Milestone 2 --------------------------------------------------------------------------------------------

**4/22/2023** – Today I am working on getting some different visuals around the comparative calculations discussed in my prior working session and looking at them based on different variables. I’m likely going to try to see what stands out vs. LEO Ratio and maybe even Population? Not sure what will be exposed here but will be working to see what could stick out to support the outputs that the mappings initially highlight.

Through some quick additional modeling, I think I’ve found some direction on the narrowing of the funnel for additional visualizations. I think there may be some additional scatter plots and perhaps even line graphs that will allow some of the comparative insights that will prove beneficial to the argument. And while I’m still circling around what that argument is, I believe it ultimately will focus on the unjustified attack and regulation on rifles.

I’m having an issue now where I’d like use df.query to select 5 of the StateID’s and I’m not sure how to accomplish this. I did find this that might be helpful: df['column'].nlargest(n=5). With some fiddling around I was able to figure out how to call out multiple values with the df.query and here is the example I was able to get to work: datazero.query("StateID==['NY','KS', 'AK', 'NJ', 'CT']"). This allowed me to select the top five, based on simply visually looking through the data, for a specific “Ratio” that I had calculated. I’ll look to get some visuals that help show this…I’m actually thinking some lines might work well for this purpose.

**4/23/2023**

Just going to spend a little time working seeing if some line charts will help make some of the comparative visuals to support the argument. I’m wanting something that doesn’t have “year” as the x-axis because I don’t like how this data looks in some forms, perhaps a bar chart might be the way to go here?   
  
Found some decent charting options with scatter and bar charts. I’m thinking comparative line charts would be nice, but I’m going to need to tap into the “Plotly go” library. I’m hesitant to open up that can of worms, but that might be my Monday Night adventure during our working class period. This is the link to the line chart documentation for future reference: [Line charts in Python (plotly.com)](https://plotly.com/python/line-charts/) ¾ of the way down the page is where the fig.add\_trace() charts begin. These will be valuable references for these visualizations.

I’m still trying to determine what to do with the visuals once I get them where I want them. When I say that, am I grabbing a screen scrape of the data for use in power point, or flyer, or another type of static publication? This all but eliminates any options for interactivity out of Plotly.

**4/25/2023 – Class night working session.**

Much of my evening in class was determining the story “path” of my visuals along with trying to fully land on how I plan to deliver the final project. I’m working in between a document/pamphlet or a PowerPoint presentation that would again be geared towards legislators in hopes of educating them on the lack of real evidence that Rifles are utilized at greater rates for homicides. My raw numbers from the FBI support this assessment as some of the calculated fields.

I’m will likely work on getting all of my graphics organized with the same colors schemes for the same information in order to carry a connection between the mapping and supporting graphics. I’m working on the flow of information presented…I have 3 main area’s that I’m going to be drilling down into and those are Rifle percentage of Gun related homicides, Blade to Rifle Ratio, and Hands Feet ect to Rifle Ratio. I do have a few over arching numbers that will likely target Handgun to Rifle ratio where all of these ratio numbers will show each of the alternative means is more likely to be used in a homicide than a rifle.

I will note that I think the course time to spend working on things and to ask questions and share ideas was very helpful. I think this could be better led by a short window of quick sharing for each person. They can 2-4 minute pitch what they have, what they did and where they’re at so others that may have questions about this or that may ask their fellow classmates. This will likely trigger more conversations between students like it did as we ended the class session.

**4/26/2023**

I have concluded that I might need to gather some factoring where I have instances in my ratios that are not greater than one. This would happen when rifles were used more frequently and worth noting on these numbers. This was found as we looked at the information from NV and the Las Vegas shooting that happened in 2017. The information about rifles was far more in favor than other means for that specific data and would like to highlight X number of instances where the rifle was more likely.

I’m going to work some into a new notebook where I do copy over most of my code to work on the very specific visuals that I want to get lined up and get them better organized. I have decided that I’ll be working with the nan values being set to 0 and working from there.

One observation from my visualizations that I’m seeing now, is that the 2019 data “appears” to be the most complete. There are little not “zero” values that seem present here. This begs the question that did something change with the states or how the FBI collects that data that may have happened in 2019. This might warrant a quick google search to see if there was an emphasis on accurate homicide statistics being reported by each state. That said, it does lead me to even further question the validity of the data collected in those years prior to 2019.

I believe I have a good number of the graphs ready as far as the colors and scheme and visual information. There will need to be some polishing now on labels and getting the legend to look nice with an appropriate label as well. This will be the next task I look to tackle when I jump back into these visuals. I have the makings of these labels further down in my code examples, and will work with that to polish the visuals I intend to use.

**\*\*\* NOTE: New FILE with condensed images and focused color themes: FinalProjectSmall.ipynb**

4/30/2023

No new work in Python today, just gathering up my files for milestone 3. I have put this into my report, but I’ll be looking to generate a informational pamphlet to display my visualizations. This was decided due to the fact that my targeted audience, legislators, will not likely be a captive audience for me to provide as presentation for. I’m also not looking to generate a web page, while modern, doesn’t fell “tangible” for these legislators to be drawn into looking at it.

This decision will prompt some additional conversation around supporting text and graphics as it relates to the final “product” that will be submitted. I also have some curiosity about identifying where our source data has been gathered from once in this final form.

There are some cleanup items that I need to finalize with my visuals as it pertains to labels which I was not able to get to with other course demands. I’ll have a full plate of work in this coming week and weekend to generate a final product that I can be happy with. I hope to be proud of it, we will see where I’m at when this concludes.

I will also try to revisit the numbers where ratios where more in favor of rifles as previously mentioned. Listing this count may be important to add credibility to the arguments and indicate a more impartial analysis of the data. I have yet to do this, and would be good a exercise, I just don’t know if I’ll be able to make the time to generate these values. I think it should be easy, but need to look further at it.

-------------- Project Milestone 3 --------------------------------------------------------------------------------------------