

RYAN XIA - CLASS OF 2022 - IW WRITE UP

CONTEXTUALIZING GUN VIOLENCE: VISUALIZING THE ECONOMIC AND POLITICAL CONTEXTS OF GUN VIOLENCE

ABSTRACT

The study of gun violence remains difficult despite recent progressions in legislation allowing research into the field again. This report will present the results of an effort to process and make publicly available a collection of gun violence datasets spanning several data sources in hopes of aiding public research and legislative change in the area. The process revealed several flaws in the current state of gun violence data including inefficient indexing, incomplete shapefiles, and unreliable, non-standardized data, which the final product attempts to address. The final product includes a github repository with all relevant code, a website providing insight into the project and guidance on use, and a series of example studies on severely affected cities using the processed datasets.

INTRODUCTION

Over the past five decades, the United States has been reliving the same nightmare, albeit with a case of severe amnesia. In 1966, the “Texas Tower Sniper” killed 16 and wounded 31 in 90 minutes. He was armed with a bolt-action rifle and semi-automatic weapons. In 2017, the Las Vegas music festival shooter killed 58 and wounded 546 in 10 minutes. He was armed with an automatic weapon. Over the past fifty years, it would seem that the issue of domestic gun violence has only worsened.

It is commonly assumed that the government is “working on it”, that society as a whole is “working on it”. However, this wasn’t the case for a significant period of time as a result of a series of regulations put in place by Congress with the lobbying of the NRA. Starting around the 1990s, the CDC put out a series of studies, one of which linked higher homicide rates with the presence of a firearm in the household, that shook the public and the media. The studies came from a specific branch within the CDC dedicated towards lowering casualties resulting from violence, as opposed to illness or poor health habits.¹

Predictably, the NRA was irritated by the CDC’s work and advocated for the elimination of the violence research branch entirely. The result was

what is referred to as the Dickey Amendment, which declared that “none of the funds made available in this title may be used, in whole or in part, to advocate or promote gun control.” While the wording of the legislation was muddled and ambiguous, the effects were clear—funding for gun-related research at the CDC, and later the NIH, became nonexistent.²

As the Dickey Amendment persisted despite democratic opposition, a series of high-profile shootings and police brutality incidents continued to put gun violence and racism in the spotlight of the public. As a result, the subject became similar to some topic in the field of pop science, welcoming several simple data visualizations, explanatory videos, and articles trying to tackle the problem of gun violence. Unfortunately, it also happens to be the case that many of these graphs, charts, videos, and articles succeed only in reiterating the issue, rather than providing any further insight into the context of gun violence. This is the problem that this project is trying to help solve, and we will analyze what specific issues much of the previous work done in the field suffers from in the next section.

PAST WORK

We will analyze three particular types of data visualization projects done in the field, providing multiple examples of each. Broadly speaking, we can categorize the three as Demographic Based Visualizations, Broad Geographic Visualizations, and Narrow Geographic Visualizations.

Demographic Based Visualizations refer to projects that sought to provide insight into who was affected by gun violence, categorizing by demographic information such as age, cultural heritage, income, type of violence, etc. A good example of this is a project done by FiveThirtyEight, which provides terrific information on what types of people are affected by gun violence.³ FiveThirtyEight’s project does an excellent job of providing a comprehensive picture demonstrating the complexity and breadth of gun violence as a social issue. While it does well in promoting

¹ Rostron, A. (2018). The Dickey Amendment on Federal Funding for Research on Gun Violence: A Legal Dissection. *American Journal of Public Health*, 108(7), 865-867. doi:10.2105/ajph.2018.304450

² Ibid.

³ Bencasselman. (2016, July 13). Gun Deaths In America.

gun violence as an issue, it performs rather poorly at informing policy decisions at a local level, which is critical to driving legislative change.

A similar project was done by Periscopic, whose visualization provided a powerful perspective on the years lost to gun violence.⁴ Periscopic's project is likely more representative of the visualization much of the public is exposed to—provocative visualizations intended to highlight and promote gun violence as a critical issue. Similar to FiveThirtyEight's project, Periscopic's visualization also fails to contextualize gun violence on a local level and doesn't effectively inform legislators or influencers about what to specifically focus on.

From a dataset perspective, both studies actually only use one dataset, which turns out to be more the case than not for most visualizations done in the field. FiveThirtyEight uses the CDC's multiple cause of death database.⁵ Periscopic uses the Federal Bureau of Investigation's Supplementary Homicide Report.⁶ It is strange, however, that two visualizations using overlapping data would choose to use two different datasets from two government agencies.

The reasons for this can be attributed to both the existence of the Dickey Amendment and the lack of a single reliable database tracking gun deaths in the United States.⁷ For example, while many continue to trust gun injury figures from the CDC, an analysis from FiveThirtyEight has demonstrated significant inconsistencies between CDC data and that of other large sources in the field. Furthermore, the CDC itself has released statements warning that its own gun injury database may be “unstable and potentially unreliable.”⁸

Broad Geographic Visualizations, on the other hand, are likely the least helpful type of study we will analyze. A good example would be an interactive visualization done by the Rockefeller Institute of the State University of New York.⁹ While it does correlate several databases, including shapefiles, gun death databases, and temporal data, what it fails to do is provide meaningful information beyond “gun deaths happen here.” Furthermore, visualizations such as these can be dangerous by playing into theory maintenance and confirmation bias behaviors that may further polarize perception of the field.^{10 11} While this warrants further discussion, it would require a separate paper, so I digress. The critical point is recog-

nizing that Broad Geographic Visualizations, such as the one above, are useless in providing useful information to drive legislative change.

Another detail to notice is that the Rockefeller Institute's visualization draws from the CDC WISQARS database which, as mentioned before, has been demonstrated to be potentially unreliable and unstable. This serves as another driving factor for the final product—a minimally reliable collection of databases (the meaning of which will be elaborated on later).

Finally, there are **Narrow Geographic Visualizations**. While there are quite a few of these, it was one study done by The Guardian that we will analyze. This study, in particular, did a beautiful job of describing the issue of Gun Violence and providing actionable information and context for its data.¹³ First, it presents data demonstrating that gun shootings tend to be geographically heavily clustered, which also serves to be one of the foundational reasons why geographic visualization was chosen for my project. Then it identifies and presents the cities in which shooting clusters exist and overlays demographic data providing context into each instance. Specifically, it clearly demonstrates a correlation between race/segregation and gun violence; in areas with low income, historical segregation, and high African American or Hispanic population percentages, there were more gun violence incidents. In clarifying this relationship, the study provides a clear challenge for legislators to tackle and address, as well as for researchers to delve into and understand. In short, it does a good job and is something that my project will seek to expand on.

However, it is important to note that The Guardian's study, while using many databases and sources, is difficult to replicate and doesn't attempt to aid public research or explicitly provide tools for interactivity or advocate legislative change. The process for producing the figures in the article is arduous and difficult to automate and it is this process that my project will seek to simplify and enable.

APPROACH

The primary goal of the project is to provide a set of tools and resources to enable both informed legislative decision making and efficient public

⁴ United States gun death data visualization by Periscopic. (2018).

⁵ Bencasselman. (2016, July 13). Gun Deaths In America.

⁶ United States gun death data visualization by Periscopic. (2018).

⁷ Campbell, S. (2018, October 04). The CDC Is Publishing Unreliable Data On Gun Injuries. People Are Using It Anyway.

⁸ Ibid.

⁹ Rockefeller Institute, Gun Violence Dashboard. (2020, October 22).

¹⁰ Gilovich, T. (1993). *How We Know What Isn't So*. New York: Free Press.

¹¹ Shafir, E. (2013). *The behavioral foundations of public policy*.

¹² Rockefeller Institute, Gun Violence Dashboard. (2020, October 22).

¹³ Aufrichtig, A., Beckett, L., Lartey, J., Diehm, J., & Team, G. (2017, January). Want to fix gun violence in America? Go local.

research. Using the Guardian's study as a template of sorts, the project provides the datasets used in a preprocessed, organized manner, and guides and examples demonstrating how to use the resources provided. The goal is to take the Guardian's study and make it more accessible, interactive, and usable.

For databases, the primary sources will be the Gun Violence Archive, Census Tract Shapefiles, the Rand Gun Law Database, American Community Survey, and the Census Bureau Databases. Practically speaking, the user flow will be identifying a county of interest via the Gun Violence Archive, then provide visualizations and context via the shapefiles and supplementary databases.

An important decision made was using the Gun Violence Archive, as opposed to the Federal Bureau of Investigation Supplementary Database or the CDC WISQARS database, as the primary source of gun violence data. The reasons were straightforward: 1) Each incident has an option for a linked, publicly available report, making verification easier. 2) Each incident has geolocal data that helps with visualization. 3) Each incident has minimal but helpful demographic data, making it easier to sort through.¹⁴

IMPLEMENTATION

The primary goal of the project is to provide a set of tools and resources to enable both easier research and informed decision making. The primary steps in the implementation phase were: data gathering and preparation, data analysis, and data visualization. Each step will be described in detail.

For the data gathering and preparation phase, the first hurdle was compiling a list of usable databases and their corresponding shapefiles and storage methods. It is important to first note that shapefiles are readily available at the census online database and should be the first place to check.¹⁵ Census shapefiles are split into several different categories, but the two being used are the census and tract based shapefiles / boundary files. The Census based files are meant for country scale visualizations, while the tract based files are meant for county level visualizations.

The next step was processing and reducing the Gun Violence Archive, Rand Gun Law Database, and the American Community Survey. Unfortunately, many of these databases do not follow one standardized indexing, which is why

it is very involved to produce visualizations and studies using them.

Packages used include MapClassify, Pandas, GeoPandas, Numpy, Matplotlib, Shapely, requests, urllib, etc. Details are included in the code as well as on the website. In each case, unnecessary columns were merged and removed, the details of which are contained below:

Census TIGER Shapefiles: removed columns included: ['COUNTYNS', 'NAMELSAD', 'LSAD', 'CLASSFP', 'MTFCC', 'CSAFP', 'CBSAFP', 'METDIVFP', 'FUNCSTAT', 'INTPTLAT', 'INTPTLON', 'ALAND', 'AWATER']. The aim was to make the Census shapefile the base dataframe for merging and manipulation. Thus, the only remaining features were the county and state FIPS codes, which are combined to form the index 'GEOID'. Each one is unique and serves to identify the county.

Gun Violence Archive: remaining columns after processing include: ['date', 'state', 'city_or_county', 'n_killed', 'n_injured', 'latitude', 'longitude', 'n_guns_involved', 'participant_age_group', 'participant_gender']. The goal here is to cut down on unnecessary information so that the preprocessing loop is more efficient. Furthermore, during the preprocessing loop, each incident is stored into a list of incidents belonging to each county. Thus, cutting down on row size allows for quicker processing and storage.

As the top two databases are the most pertinent and important, the remaining descriptions will be more brief.

Rand Gun Law Database: the database contains 3,000 gun laws across the 50 states, providing context for narrowed down searches. The processing limits the dataset to FIPS identification, law class, and law content.¹⁶

American Community Survey: this database, which is updated annually by the Census Bureau, is thankfully able to be sorted via an online interface that the Census Bureau provides. In it's current state, the dataset being used contains all census tracts and their respective data on racial composition and income.¹⁷

¹⁴ Gun Violence Archive. (2018).

¹⁵ Bureau, U. (2020, March 12). TIGER/Line Shapefiles.

¹⁶ Cherney, S., Morral, A., Schell, T., & Smucker, S. (2020, September 17). The RAND State Firearm Law Database.

¹⁷ Bureau, C. (2021). American Community Survey

The merging and processing phase was completed via a loop that queried the Census Geocoding API for each entry within the Gun Violence Archive. Unfortunately, the actual implementation of this was very problematic and many revisions were made in order to account for imperfect data from both the Gun Violence Archive and the Geocoding service.

In particular, two of the most common errors were `KeyErrors` and `TypeErrors`. On occasion the Gun Violence Archive data entry will have a faulty longitude latitude pair that points a non-existent county, thus causing a `TypeError`. At other times, the Census Geocoding service would return `NaN` for its GEOID, which in turn causes a `KeyError`.

Another hurdle was the issue of long runtimes and incremental saving. In the initial implementation of the data processing loop, there were no incremental saves, no error-catching, and inefficient assertions and if-statements. Over multiple iterations, the current product contains error-catching for `KeyErrors` and `TypeErrors`, as well as incremental saving separated by the year of each gun incident. For the 2016 incremental save, there were around 200 `KeyErrors`, which results in a .3% data loss, which was decided to be within acceptable given the 60,000 entries for that year. Unfortunately, the issue of runtime and efficiency is largely limited by the efficiency of the Census Geocoding API, which is understandably slow.

Further data processing includes primarily that of the American Community Survey and parsing through it via the API. Since the final goal is to make an interactive visualization spanning the entire United States, there are two ways to achieve this.

One is to download the entire dataset and access the parts that the user wants to use in the visualization. This approach is, however, unsustainable. Since the ACS database is enormous, the approach wouldn't be viable due to both not having a server capable of handling querying such a large database and it would require constant maintenance due to the ACS updating frequently.

The other is to utilize the Census Data API to query the ACS for up-to-date information on each census tract during the visualization. This is the approach that the product currently uses. The downside is that the load times for each in-

dividual visualization is longer compared to the approach above. However, the upsides include up-to-date information, user flexibility, and lower server load.

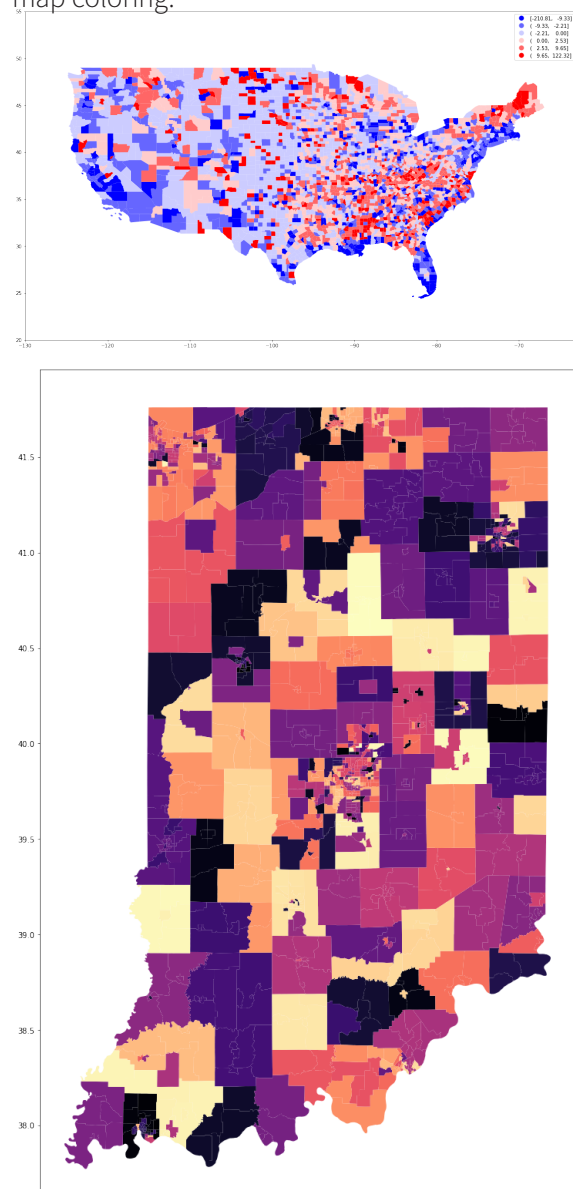
The data analysis phase is less involved than the data processing phase. The goal is straightforward: to identify highly affected counties and use them for visualization. This is done in a few ways. First, counties are filtered by unadjusted casualty or incident numbers, putting highest incident counts at the top. Second, counties are filtered by adjusted casualties. Third, counties are filtered by both rural percent and adjusted casualties, which aims to include both rural areas and urban areas. This is included specifically to avoid suggesting that only cities or high population density areas are affected by Gun Violence. A few results of county identification:

Louisiana	Missouri
Orleans Parish	St. Louis city
Rural Pct: -49.408136	Rural Pct: -49.999687
Incident Count: 1467	Incident Count: 1016
Casualty Count: 1204	Casualty Count: 1114
Total Population: 343829	Total Population: 319294
Density: 0.078351	Density: 0.199665
Median Income: 38855	Median Income: 43263
-----	-----

Some interesting findings during the data analysis phase include: 1) As expected, urban areas have far higher numbers of incidents. 2) Filtering by incidents per 1,000 population skewed results to low population rural areas as urban areas tend to have much higher population densities that drown out the number of incidents in that county. 3) 100 percent rural areas tend to be excluded from most filters used, which is expected given much lower population density.

Finally, the data visualization phase, which relies heavily on matplotlib, was implemented so that users could choose the colormap, scheme, and county they were interested in. A folder containing the fifty different census tract shapefiles is queried depending on which county the user wishes to see and the boundaries are drawn only for that county. Following this, the program uses the Census Bureau Data API to query for a column on which matplotlib will color the visualization. All of this is performed according to user input for the function but hasn't yet been linked to the website. The website, as it is now, only houses the various data sources used and

serves as a storage site rather than an interactive visualization. A few examples of matplotlib and map coloring:



RESULTS

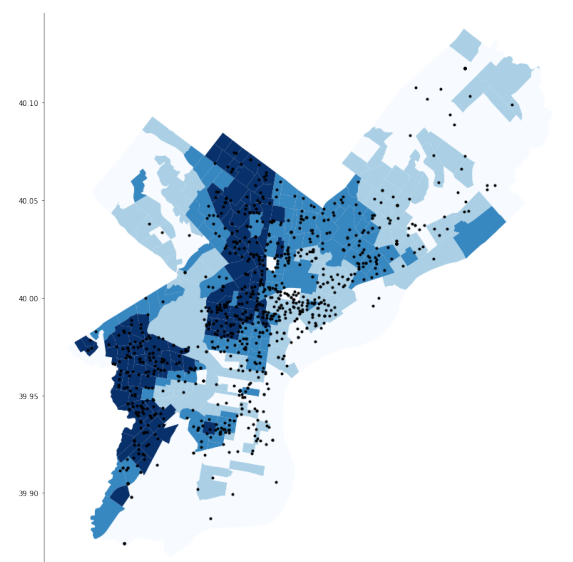
The results for a project such as this are rather ambiguous. First and foremost, everything works as intended. As explained in both the approach and the implementation, the functionality of the python scripts successfully enable flexible study of different counties in detail just as the Guardian had done. Some specific studies done on Gun Violence Archive Data from 2016 have been included below as examples of what the script enables.

PHILADELPHIA 2014

A quick incident analysis of the 2014 Gun Violence Archive yields Philadelphia as one of the most affected counties of that year. At 1,043 cases, Phil-

adelphia is close behind Chicago and followed by Washington D.C., Houston, and San Francisco. This will be the subject of the first example.

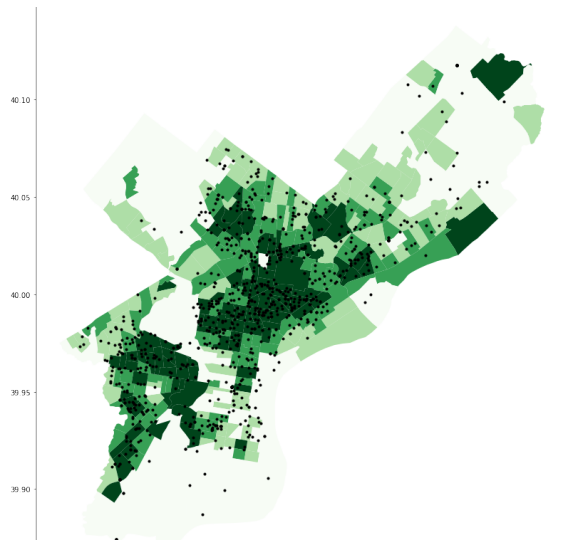
Now, among those concerned with political correctness, it can often be taboo to insinuate that African Americans are violent. However, behavioral psychology studies have constantly verified this implicit bias in most people. In particular, one can look to the study with regards to identifying weapons in the hands of Black people versus people of other races. There exists a clear inclination to misjudge Black people to have weapons and misjudge White people to not have weapons.¹⁸ Below is a geographic visualization where gun incidents are plotted onto a map of Philadelphia. Notice that the census tracts of Philadelphia are colored such that a darker blue indicates a higher percentage of Black people as part of the tract's overall population.



At first glance, this map does seem to confirm the implicit assumption that Black people are violent. This is, in part, true. However, it turns out that Black people turn out to be both the most affected by gun violence, with over 30 percent of homicide victims being young black males.¹⁹ So, how does this change the narrative? It doesn't really, not just yet. As far as the first visualization goes, it on its own serves only to confirm this unfortunate bias. However, this next map might provide better information. For this visualization, the census tracts are colored by percent of population determined, as opposed to undetermined by the Census Bureau, to be under the poverty level. Simply put, the Greener the census tract, the poorer it is.

¹⁸ Shafir, E. (2013). Ch.1 The behavioral foundations of public policy.

¹⁹ Bencasselman. (2016, July 13). Gun Deaths In America.



This visualization provides a clearer picture of the circumstances under which most of the gun incidents in Philadelphia take place—poverty. This has larger implications than it might seem at first glance. Behavioral Psychology studies have also linked lower cognitive ability with being in poverty.²⁰

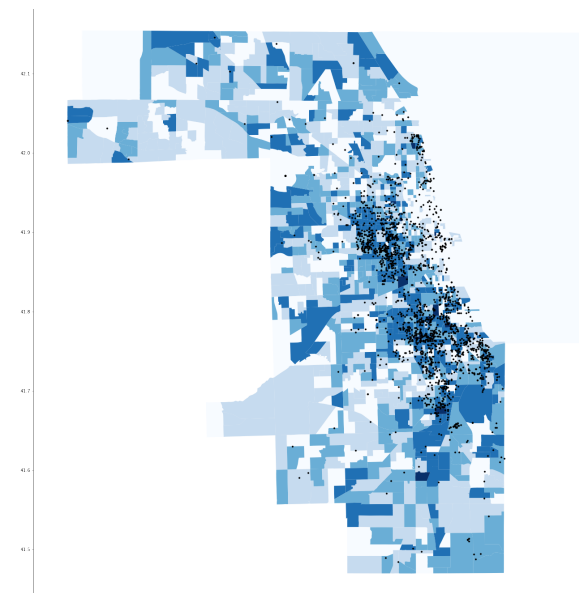
It is important to note the difference between being of low intelligence and having lower cognitive ability. Having lower cognitive ability can be both the effect of having low intelligence or of having lower cognitive capacity. In the case of the Behavioral Psychology study referred to above, it is a lowered cognitive capacity that leads to lower cognitive ability.

In essence, poverty makes decision making harder because there is too much going on. What's for dinner? My phone bill is due, do I leave my car unrepaired? Should I take a loan on my weekly check to pay it? Furthermore, poverty also leads to unhealthy childhoods that can lead to depression, anxiety, and even violent behavior, all of which can develop later in life or even during childhood.²¹

Thus, it would be unfair and incorrect to assume that all Black people are violent after looking at the second map. Rather, it would be more important to realize that those in poverty and the African American community suffer disproportionately to gun violence, at least in Philadelphia. Furthermore, this association is geographic and not just numeric. The visualization offers concrete evidence demonstrating which tracts are more affected by others and thus require legislative change or assistance.

CHICAGO 2014

In this next example, we will visualize incidents in Chicago, Illinois. As mentioned earlier, Chicago ranked first for number of incidents in 2014, far surpassing Philadelphia with almost double the number of incidents at 2,128 cases. We can start off a visualization. Below is a map of Cook county with incidents marked over census tracts colored by educational attainment.



It is rather surprising how closely gun incidents occur according to the boundaries of counties with low educational attainment. Or in the specific case of this visualization, the counties with the highest percentage of inhabitants without having graduated from highschool. This puts into context the popular concept of keeping kids in school to lower gun violence and crime. Unfortunately, we have seen both in action and in theory that this course of action doesn't address the root of the problem. Speaking just to the theory portion, it turns out that a large part of a successful education stems from interaction and engagement from parents in the household.²² So just keeping children in school doesn't necessarily keep them engaged or provide a good education. The visualization below colors census tracts by income and provides a fuller picture of the gun violence in Chicago in the context of education.

A quick glance shows that much of the incidents occur in low income, low educational attainment areas. This reinforces the notion that keeping kids in school is not the solution even though it seems that lower educational attainment is a cause of gun violence. As mentioned before, poverty has adverse effects for both par-

²⁰ Shafir, E. (2013). Ch.16 The behavioral foundations of public policy.

²¹ APA. (2018). Childhood poverty, living below the line.

²² Shafir, E. (2013). Ch.19 The behavioral foundations of public policy.

ents and children. Furthermore, in the context of education and schooling, parents dealing with keeping their family financially afloat are understandably and predictably less present in the household. As a result, providing a supportive environment for education becomes difficult and adversely affects the outcomes of students. This suggests that, rather than forcing children to stay in school, helping parents be more present and educationally supportive in the household might provide better results in terms of curbing gun violence. This concludes the visualization examples.

As you can see, with the set of scripts created via this project, it becomes possible to tell many different narratives about counties and people affected by gun violence. At the moment, it is difficult to provide census tract specific demographic data beyond the inventory of the Census Bureau since not many entities record information across the country at that level. However, for the moment, the ACS should provide ample information for the average user and more advanced users should be able to search for data sources that fit their needs.

CONCLUSION

While the script in the implementation phase functions properly and produces results, there remains a good deal to be done in order to make the work meaningful and far reaching. There are two tasks still to be completed, the details of which I will go through for the conclusion.

First, a github repository. This task is relatively straightforward. With the files processed, it is necessary to first make them openly accessible to the public. From here, a website can be created using the github websites functionality that both houses the data collection as well as provides interactive visualizations and online script usage for the project. In doing so, the scripts described in the implementation phase will actually see use and potentially make an impact.

The website will consist of two parts. One is, as mentioned above, the freeform interactive visualization. The other part will focus on storytelling. Looking back through the work of the Guardian, FiveThirtyEight, and Periscope, a common theme is the idea of telling stories through data rather than just having people interpret them. While this is likely an entire second project, it is something worth continuing beyond this final report.

Second, outreach. During the semester of searching, refining, and optimizing the scripts, a few things in particular stood out. For one, there in fact exist thousands of online APIs and interactive sites that allow users to geolocate, find census tract information, find gun law information, etc, that never reach the people that need to see it. Many of the tools are buried and require extremely specific search terms that are unlikely for a legislator or someone in their office to stumble upon. Thus, reaching out to legislators, news outlets, and other independent researchers is essential for the project to grow beyond a handful of python scripts.

Finally, an evaluation of the success of the project. While the website remains in construction and outreach hasn't been done, the capabilities of the project show promise. The initial goal was only to make a small step in the right direction, so the project in its current state can be considered a success. Furthermore, it's potential to be more than a handful of python scripts shows promise.

However, as part of researching over multiple months, it is also apparent that there does exist a lot of work in the field, albeit under the surface, that also accomplishes many of the things that this project does. The project does distinguish itself from the others, but does not necessarily outdo many of the other projects in the field. It is my hope that others will continue to build on this project via the public repository and the website.

ACKNOWLEDGEMENTS

I'd like to thank my Advisor, David Dobkin, as well as my Independent Work Seminar classmates for their feedback and advice on the project. I'd also like to thank the data librarians at the Firestone Library for their assistance in finding reliable data sources.

SOURCES

APA. (2018). Childhood poverty, living below the line. Retrieved January 07, 2021, from <https://www.apa.org/pis/res/resources/indicator/2014/06/childhood-poverty>

Aufrichtig, A., Beckett, L., Lartey, J., Diehm, J., & Team, G. (2017, January). Want to fix gun violence in America? Go local. Retrieved January 03, 2021, from <https://www.theguardian.com/us-news/ng-interactive/2017/jan/09/special-report-fixing-gun-violence-in-america>

Bencasselman. (2016, July 13). Gun Deaths In America. Retrieved January 03, 2021, from <https://fivethirtyeight.com/features/gun-deaths/>

Bureau, C. (2021). Retrieved January 06, 2021, from <https://data.census.gov/cedsci/>

Bureau, U. (2020, March 12). TIGER/Line Shapefiles. Retrieved January 05, 2021, from <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html>

Campbell, S. (2018, October 04). The CDC Is Publishing Unreliable Data On Gun Injuries. People Are Using It Anyway. Retrieved January 03, 2021, from <https://fivethirtyeight.com/features/the-cdc-is-publishing-unreliable-data-on-gun-injuries-people-are-using-it-anyway/>

Cherney, S., Morral, A., Schell, T., & Smucker, S. (2020, September 17). The RAND State Firearm Law Database. Retrieved January 06, 2021, from <https://www.rand.org/pubs/tools/TLA243-2.html>

Gilovich, T. (1993). *How We Know What Isn't So*. New York: Free Press.

Gun Violence Archive. (2018). Retrieved January 05, 2021, from <https://www.gunviolencearchive.org/>

Gun Violence Dashboard. (2020, October 22). Retrieved January 03, 2021, from <https://rockinst.org/gun-violence/gun-violence-dashboard/>

Rostron, A. (2018). The Dickey Amendment on Federal Funding for Research on Gun Violence: A Legal Dissection. *American Journal of Public Health*, 108(7), 865-867. doi:10.2105/ajph.2018.304450

Shafir, E. (2013). *The behavioral foundations of public policy*. Princeton, NJ: Princeton University Press.

United States gun death data visualization by Periscope. (2018). Retrieved January 03, 2021, from <https://guns.periscope.com/>