---

Title of Study: Case Study on Gun Related Deaths

Author: Amber P. Giza

File Format and Title: Updated Case Study on Gun Violence and Deaths.docx

---

---------------------------

CONTENTS:

1. ABSTRACT

2. INTRODUCTION

3. KAGGLE

4. JUPYTER NOTEBOOKS USAGE

5. CODE

6. RESULTS

7. COMPARING TO PAST RESULTS

8. CONCLUSION

9. REFERENCES AND CITATIONS

---------------------------

# PART 1: Abstract

About a few months ago, I was tasked to research a module for Google Coursera, which is the capstone module (Final), a dataset from Kaggle was picked, to demonstrate a need for stricter gun laws. Gun violence within the United States

has been a prevalent subject to bring up, as thousands of people across the country have either access to a gun or know someone who owns one. What this research will entail is that this will examine a dataset that spans a few years, from 2006 to 2020, which is an expansion of the previous research that I conducted. This project aims to identify key factors that contribute to gun violence, such as socioeconomic conditions, urbanization, gun ownership laws, and the mental state of the shooter. Next, data visualizations will be included with the ggplot2 and plotly packages, in R, to visualize the amount of gun violence. Lastly, this study will underscore the importance of sociopolitical considerations, to address the challenges posed by a want or desire to change gun laws. For this project, Jupyter Notebooks with the use of R and Python there will help visualize the data.

# PART 2: Introduction

"A well regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed" (The Declaration of Independence, 1776). This is known as the 2nd Amendment inside The Bill of Rights, which are the first ten amendments listed inside the Constitution of the United States. Basically, this law in part helps protect the right for Americans to possess guns and similar weapons for protecting themselves. However, in recent years, this amendment has sparked many debates as to how it should be regulated, as gun-related homicides have increased.

Homicides are not the only form of gun-related incidents that are common. Sadly, some of these incidents are caused by accidents, where the individual in question had no knowledge of how to treat or use a firearm, properly. Therefore, said individual may've wrongfully ended the life of another individual, because of improper firearm use.

Lastly, the other form of gun-related incidents that are common are forms of suicide. Guns have been used in suicide attempts, as they are a weapon that utilizes speed and precision. Suicide via firearm is sadly one of the most common ways people take their own lives, when compared to other methods (strangulation/hanging, pesticide/poison ingestion, suffocation, etc.). Since this is a dark subject, a list of suicide hotlines has been included in the file title "WARNINGS.txt," in case anyone reading this study needs someone to reach out to. Race is something that some people argue is a contributing factor, as well as age and gender. Statista graphs show that in 2021, "gun homicide rates were highest among Black people aged between 15 and 24 years, at 70.65 gun homicides per 100,000 of the population. In comparison, there were only 2.71 gun homicides per 100,000 of the White population within this age range. However, the risk for gun homicide was greatest among all adolescents and young adults between the ages of 15 to 34 in that year" (Korhonen, 2024). Plus, according to Brady United citing a survey that surveyed gun violence, "Over 45% of Black and Hispanic Latino Americans believe gun violence is an important issue impacting their neighborhood, compared to only 27% of white Americans" and inside the same article, Black and Hispanic/Latino youth are more than 11 times and 2.5 times more likely to die from firearm homicide than their white peers, respectively" (Brady United, n.d.).

# PART 3: Kaggle Integration

The datasets were pulled from Kaggle, with the user name "Konrad Banachewicz." The original data is from collaborating with the websites FiveThirghtyEight, CDC (Centers for Disease Control), and NCHS (National Center for Health Statistics). The repository from Kaggle has 15 csv files, one for each year from 2006-2020. For each attribute in the file, after cleaning:

* Year: Year that the incident took place. In each folder, each year is partitioned to hold data and files related to said year. For
* example, in the "Year 2006" folder, there is data and analyses for that specific year.
* Month: Month that the incident took place. In this dataset, the months are numbered (01 for January, 02 for February, etc.)
* Intent: The intentions of the victim.
* Police: Were the police contacted? 0 means no or unknown, and 1 means yes.
* Gender: The gender of the victim.
* Age: The age of the victim.
* Race: Race of the victim.
* Hispanic: If the victim is Hispanic, yes. If not, no.
* Place: Where the incident took place
* Education: The highest amount of education that the victim has.

NA values in the age column were replaced with the average age of the victim, for each csv file. NA values in the place column were replaced with the value "Other unspecified" as since that value already exists, it also functions as an "unknown" location holder of sorts. NA values were replaced in the education column with "Unknown." NA values were replaced in the intention column with “Undetermined.” For the Hispanic column, this was one of the more confusing columns, because instead of basic "Yes" or "No" values, there were numerical values. For the most part, anyone not Hispanic had a value either 100 or 998, and Hispanic victims have a value between 200-299. I have no idea what these values or possible codes mean, so I had to painstakingly replace each value in Excel with "Yes" or "No" to make things less confusing.

# PART 4: Jupyter Notebook Usage

Jupyter Notebooks are the result of following Project Jupyter, which "is a non-profit, open-source project, born out of the IPython Project in 2014 as it evolved to support interactive data science and scientific computing across all programming languages" (Jupyter, n.d.). Because of the fact that Jupyter is open-source, many companies and applications such as Google Colab, Amazon Web Services (AWS), and Microsoft Azure, utilize either as a free or paid service. Jupyter's source code and interfaces are free to use, regardless if you choose to go through a company's service. For this project, I utilized Google Colab as this is something I have prior experiences of using. Google Colab is a free to use service for writing and executing Python, R, and Julia scripts as well as sharing them through the use of Google Drive. Because of the shareability as well as being able to easily change the Runtime language, Google Colab will be the main source for hosting the Python and R code.

# PART 5: Code

+++++++++++++++++++++++++++

Sub-part 1: Python

+++++++++++++++++++++++++++

Most of the code is recycled, because there are several csv files to filter through. First off, the packages used for the Jupyter Notebooks with using the Python Runtime and their functions are:

* Pandas (as pd): Data manipulation and analytical tool for creating DataFrames. "pandas is a fast, powerful, flexible
* and easy to use open source data analysis and manipulation tool,
* built on top of the Python programming language" (OVHCloud, n.d.).
* Matplotlib.pyplot (as plt): Plotting data and visualizations. "Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible" (Matplotlib Development Team, n.d.).
* Seaborn (as sns): Data visualization. "Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics" (Waskom, n.d.).
* NumPy (as np): Data manipulation for making arrays. "The fundamental package for scientific computing with Python... NumPy brings the computational power of languages like C and Fortran to Python, a language much easier to learn and use. With this power comes simplicity: a solution in NumPy is often clear and elegant" (NumPy Team, n.d.).
* Statistics: Determining the mode and other statistical measures. "This module provides functions for calculating mathematical statistics of numeric (Real-valued) data" (Python Software Foundation, n.d.).
* SciPy: Scientific/Statistical computing, as an extension of Numpy. "Fundamental algorithms for scientific computing in Python... SciPy provides algorithms for optimization, integration, interpolation, eigenvalue problems, algebraic equations, differential equations, statistics and many other classes of problems" ("SciPy", n.d.).

Second, as casting the csv files to a DataFrame, there was only one numerical column to determine the statistics of: the victim's age. Third, data visualizations for the boxplots and histograms helped us visualize:

* Skewness of the age column
* Visualizing the data of the non-numerical columns

+++++++++++++++++++++++++++

Sub-part 2: R

+++++++++++++++++++++++++++

The code here is also recycled for the R and RMD file, because there are several csv files to filter through. First off, the packages used for the Jupyter Notebooks with using the R Runtime and their functions are:

* ggplot2: Based on The Grammar of Graphics by Leland Wilkinson et al., "ggplot2 is a system for declaratively creating graphics... You provide the data, tell ggplot2 how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details" (Posit Software, n.d.).
* tinytex: "TinyTeX is a custom LaTeX distribution based on TeX Live that is relatively small in size, but functions well in most cases, especially for R users" (Xie, Y., et al., 2021).
* rmarkdown: "R Markdown is a file format for making dynamic documents with R. An R Markdown document is written in markdown (an easy-to-write plain text format) and contains chunks of embedded R code, like the document below" (Grolemund, 2014).
* plotly: "Plotly's R graphing library makes interactive, publication-quality graphs. Examples of how to make line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heatmaps, subplots, multiple-axes, and 3D (WebGL based) charts" (Plotly, n.d.). In conjunction with ggplot2, this is one excellent resource to use for data visualization.
* DescTools: "The DescTools package in R programming is a collection of functions that are used in various scenarios where data description, summary, and exploration are needed. It is a widely used package that was designed to help data scientists, researchers, and data analyst to understand their data and identify their findings" (GeeksForGeeks,2023).
* The dplyr package is "a grammar of data manipulation, providing a consistent set of verbs that help you solve the most common data manipulation challenges" (Wickham, et al., 2023).

First off, DescTools is used for getting the main basic statistical measurements (Mean, Median, Mode, and Standard Deviation). This should give us an idea of how the data is skewed. Second, with the ggplot2 and plotly packages, these are used to make graphs (boxplots and a histogram) to further visualize the data. DescTools is necessary to display the mode of a dataset in R, which is the data value that occurs the most inside a column of data. Third, dplyr is used for counting up the non-numerical columns to help determine which victims died the most, and the most common method/intent utilized. Lastly, two more histograms were constructed to visualize when the most deaths occurred (by month and by year).

# PART 6: Results

Both Python and R showed that the data from 2006-2020 in the csv files had similar trends to one another:

* Mean: The mean across all csv files ranged from 41 - 44 years old.
* Median: The median across all csv files was around 37 - 41 years old
* Mode: Most of the victims killed were between the ages of 21 - 25 years old
* Standard deviation: The Standard Deviation was between 19 - 20 years old, only changing slightly across each csv file.

R was better for visualizing the boxplots with the age and categorical attributes of the victim. First off, the boxplot with comparing the victims ages to the victims races, victims that were White or Caucasian had no outliers. Not only that, but across all csv files, the boxplots had a normalized central tendency, or were the most "normal." In statistics, that means that the mean and median for this boxplot are nearly the same value. The second interesting fact from the data column is that African-American/Black victims had the most outliers when comparing the ages of Black victims to other victims' ages and race. For one of the next groups of the boxplots, comparing males to females based on their ages had the least amount of outliers, which is interesting. Lastly, for the final group of boxplots, this regarded the intentions of the use behind the weapon/gun. Sadly, the boxplot on suicide had the most normalized central tendency, across all csv files. The most common time for people to die by gun usage was in July, which may attribute to seasonal depression.

# PART 7: Comparisons

From the past data analysis, I determined that most of the deaths that were cited were caused by successful suicide attempts, at almost 63%. From the past analysis, most of the victims were White/Caucasian, most of the victims were Males, the year that had the most gun-related deaths was 2013, and the month that had the most gun-related deaths was July. For the deaths located in these csv files, suicides were also the most gun-related deaths across all 15 csv files. From this study, the year that had the most gun-related deaths was 2020, the month that had the most gun-related deaths from 13 csv files was July (1 csv file had August and another had January as their most gun-related deaths per month), and most of the suicide attempts range between 53 - 63%. If we look:

* Suicide attempts (2006): 54.64%
* Suicide attempts (2007): 55.57%
* Suicide attempts (2008): 57.68%
* Suicide attempts (2009): 59.77%
* Suicide attempts (2010): 61.23%
* Suicide attempts (2011): 61.79%
* Suicide attempts (2012): 61.57%
* Suicide attempts (2013): 62.95%
* Suicide attempts (2014): 63.496%
* Suicide attempts (2015): 60.74%
* Suicide attempts (2016): 59.34%
* Suicide attempts (2017): 59.98%
* Suicide attempts (2018): 61.48%
* Suicide attempts (2019): 60.29%
* Suicide attempts (2020): 53.72%

The mean from adding these up is: 55.497%.

Something I noticed from looking back at the past data that I noticed was:

* 2012 had 33563 deaths occurred in total
* 2013 had 33636 deaths occurred in total
* 2014 had 33599 deaths occurred in total

The csv files analyzed for the new analysis for the years 2012, 2013, and 2014 have the same number of deaths and columns. It could be possible that those years hold the same data from this analysis, though I decided to analyze the files with the use of Python, as well as R. Plus, this analysis has 12 other files from different years.

# PART 8: Conclusion

The analysis done shows that most of the victims were males, most of the gun related deaths occurred in 2020, the month that had the most gun-related deaths was July, and the most deaths that were cited were caused by suicide attempts. Over the years, suicide attempts via gun usage tended to fluctuate, but the percentages are still pretty high, in my opinion.

For the next steps to help prevent any future gun-related deaths and suicide attempts, this requires a multi-faceted approach that will address not only societal influences, root causes, gun safety laws, and addressing any root causes that are overlooked. First off, re-addressing gun safety laws and regulations should be considered as the United States of America doesn't have fairly lax laws and regulations when a person is considering to buy a firearm. For most states, the person must be older than the age 18 to buy firearms. In Washington State, the minimum age to buy a firearm is 21 years old. Some states don't regulate a minimum age for buying a firearm, though the United States Government has put a federal minimum age of 18 years old for purchasing a firearm. Guns should be kept away from children and teens by any means necessary, as this group of people either:

* Doesn't know how to handle a gun and misfires
* Knows what a gun is and they want to show it off to friends, but not hurt anyone
* Knows what a gun is and they want to harm others

For other laws regarding the possession of guns, that comes down to how the gun operates. Guns are not all built the same way. Some have larger bullets than others, and some guns require the user to reload them manually, or the gun can automatically or semi-automatically reload bullets. Some semi-automatic weapons and automatic weapons are used by professionals like the military and police officers, but civilians have no need or use for guns like the AR-15, outside of gun collectors or prominent hunters. These guns were once more available for ordinary civilians to purchase, but nowadays, are being more regulated, which is a good thing. However, these guns can still slip into the wrong hands. Only a handful of guns are banned inside the United States of America, which include machine guns/automatic firearms, short-barreled shotguns, short-barreled rifles, and most states even ban AK-47s (Avtomat Kalashnikova-47s). Lastly, storing the guns at home should be carefully considered. Guns should be stored away from children, teens, adolescents, and individuals that suffer from mental illnesses.

Second, selling firearms to an individual should be heavily regulated. Background checks aren't done that often for people who are willing to buy a firearm, and if a background check is done on said individual, the process isn't thorough. Firearms should not be sold to anyone of the following, in my opinion:

* Individuals who are younger than the minimum age for buying a firearm in their respective state (or younger than 21).
* Individuals who have severe mental illnesses, cognitive impairments, or behavioral disorders
* Individuals who have been convicted for a felony and have served prison or jail time
* Individuals who live or share a living space with someone mentioned above

Third, society needs to address socio-economic issues that are associated with gun violence. Most of the gun violence that occurs in the United States of America occurs within lower-income communities. While most sources and professionals can see a rise in gun violence in lower-income neighborhoods and communities, these professionals and analysts don't understand why there's an increase. The reason why there is an increase in poorer communities is because of political unrest and stress that takes over an individual's life.

This leads into my fourth and last point in regulating guns. For suicide methods, the use of a gun is a common method as the victim points the barrel of the gun towards the roof of their mouth, with the intent of having the bullet penetrate through their skull and affecting their brain. Forming a stronger social network for people who are experiencing suicidal thoughts as well as making counseling more available for those who can't afford aid for their mental health. As someone who has attempted suicide (strangulation) and is a survivor, I vouch for having counseling, therapists, and similar methods to help lower stress and symptoms that come with depression and anxiety to become more available. This is why a list of resources are included within this file for anyone reading this to reach out to, should the individual be dealing with suicidal thoughts.

# PART 9: References

Adams, D., Martin, M., Perez, A., & Mai, H. J. (2024, April 16). Suicides make up majority of gun deaths, but remain overlooked in gun violence debate. NPR. https://www.npr.org/2024/04/16/1243924108/gun-violence-suicide-united-states

Banachewicz, K. (2023, June 20). Gun deaths in the US: 2006 - 2020. Kaggle. https://www.kaggle.com/datasets/konradb/gun-deaths-in-the-us-1968-2021

Bencasselman. (2016, July 13). Gun deaths in America. FiveThirtyEight. https://fivethirtyeight.com/features/gun-deaths/

Brady United. (n.d.). Gun violence is a racial justice issue. Gun violence is a Racial Justice Issue | Brady. https://www.bradyunited.org/resources/issues/gun-violence-is-a-racial-justice-issue

Centers for Disease Control and Prevention. (2024, December 19). NVSS - Public Use Data File Documentation. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nvss/mortality\_public\_use\_data.htm

GeeksforGeeks. (2023, June 8). Desctools package in R. https://www.geeksforgeeks.org/desctools-package-in-r/

Google. (n.d.). Google colab. https://colab.research.google.com/#scrollTo=Wf5KrEb6vrkR

Grolemund, G. (2014, July 16). Introduction to R Markdown. Introduction to R markdown. https://rmarkdown.rstudio.com/articles\_intro.html

Jupyter. (n.d.). Project Jupyter. Project Jupyter | About Us. https://jupyter.org/about

Korhonen, V. (2024, November 21). Gun homicide rate by race and age U.S. 2021. Statista. https://www.statista.com/statistics/1466060/gun-homicide-rate-by-race-and-age-us/

Kravitz-Wirtz, N., Bruns, A., Aubel, A. J., Zhang, X., & Buggs, S. A. (2022). Inequities in Community Exposure to Deadly Gun Violence by Race/Ethnicity, Poverty, and Neighborhood Disadvantage among Youth in Large US Cities. Journal of urban health : bulletin of the New York Academy of Medicine, 99(4), 610–625. https://doi.org/10.1007/s11524-022-00656-0

Matplotlib Development Team. (n.d.). Matplotlib: Visualization with Python. Matplotlib -- Visualization With Python. https://matplotlib.org/

Nra-Ila, & Association, N. R. (2025). Ila: State gun laws. NRA. https://www.nraila.org/gun-laws/state-gun-laws/

NumPy Team. (n.d.). NumPy. https://numpy.org/

OVHCloud. (n.d.). Pandas. pandas | Python Data Analysis Library. https://pandas.pydata.org/

Plotly. (n.d.). Plotly. Plotly r graphing library in R. https://plotly.com/r/

Posit Software. (2023, September 12). Open source for a Better World. Posit. https://posit.co/about/

Posit Software. (n.d.). Create elegant data visualisations using the grammar of graphics. Create Elegant Data Visualisations Using the Grammar of Graphics •. https://ggplot2.tidyverse.org/

Python Software Foundation. (n.d.). Statistics - mathematical statistics functions. statistics — Mathematical statistics functions | Python 3.13.2. https://docs.python.org/3/library/statistics.html

*Scipy.* SciPy. (n.d.). https://scipy.org/

Waskom, M. (n.d.). seaborn: statistical data visualization #. seaborn. https://seaborn.pydata.org/

Wickham, H. (2016). GGPLOT2: Elegant graphics for data analysis. Springer.

Wickham, H., Francois, R., Henry, L., Muller, K., & Vaughan, D. (2023). A grammar of data manipulation. A Grammar of Data Manipulation •. https://dplyr.tidyverse.org/

Xie, Y., Dervieux, C., & Riederer, E. (2021). R markdown cookbook. CRC Press.

Zurda. (2017, January 25). Gun deaths in the US: 2012-2014. Kaggle. https://www.kaggle.com/datasets/hakabuk/gun-deaths-in-the-us