Firearm Education Data Summary

Haley Bustle January 3, 2024

**Data Set Name:** Census Data 2009 to 2017. Download [[here](file:///C:\Users\boudl\Firearm%20Education%20Project\02%20Data\Original%20Data\census%20data%202009%20to%202017%20PDF.pdf)](https://docs.google.com/spreadsheets/d/1I4zNbY2lPxglzCL6bmKKppt8zoMOnaLM_D8JGyaXQGY/edit?usp=drive_link)

**Dataset source:** This is an external data source provided the US Census Bureau

**Dataset Summary:** As indicated from the US Census website, the information and data collected by them is collected both through surveys and administrative data ([Source](https://www.census.gov/about/what/admin-data.html#:~:text=The%20Census%20Bureau%20uses%20data%20from%20a%20variety%20of%20sources.&text=Some%20data%20are%20collected%20from,additional%20data%20from%20other%20sources.)). However, this collection only happens even ten years ([Source](https://www.census.gov/programs-surveys/censuses.html#:~:text=The%20U.S.%20census%20counts%20every,takes%20place%20every%2010%20years.&text=Learn%20about%20all%20the%20decennials%20from%201790%20to%20present.)), which can cause massive discrepancies in a data set. In addition to a time lag, this data has further room for error, as it is collected both through survey (where human error can occur) and administratively (where computer error can occur). Surveys are also at the mercy of people themselves, as not everyone will fill out or respond to them, which can also skew data collection for one or more populations. This data set includes population data from all fifty states, District of Columbia, and Puerto Rico for gender, age categories, and year data was collected. However, this data set is from 2009 – 2017 and the most recent census was in 2020, which will cause a data discrepancy

**Relevance & Limitations:** This data set will allow me to view how populations of various age groups, genders, and states have changed over time when in comparison to firearm related incidents. However, this dataset is now seven years old and the most recent census was completed in 2022. This will be kept in consideration when completing this analysis and allow room for interpretation of results due to this data gap.

**Data Set Name:** Firearm Deaths 2009 to 2017, download [here](https://docs.google.com/spreadsheets/d/19JtAzsNPKSNTKmDp4mLzjYQgu3vms9oh/edit?usp=drive_link&ouid=107389679865717903717&rtpof=true&sd=true)

**Dataset source:** this is an external dataset provided by the CDC Center for Disease Control and Prevention.

**Dataset Summary:** This dataset was a personal selection of factors as provided by data collected through the CDC Wonder Underlying Cause of Death 1999 to 2020 ([Source](https://wonder.cdc.gov/wonder/help/ucd.html)). The information collected on causes of deaths are based listed cause of death on an individual’s death certificate and thus broken down into several functions that can be used to shape a dataset specific to the needs of a data analysis. For this purpose, this dataset was created to study firearm-related incidents (assault, suicide, homicide, and accident) amongst all age groups, all states, and years 2009 to 2017. Please note this dataset does include zero values and suppressed values, where suppressed values are statistics representing one to nine deaths reported. These restrictions are also marked as unreliable, since the values reported are not only less than ten but can also be suppressed due to age, ethnicity, privacy, or security concerns. The original dataset was not immediately usable, and was converted by my hand into more legible rows and columns (please review dataset changes [here](https://docs.google.com/document/d/1dVi3kYPmLCJIPXCZe_KFI3hqMgXMKk7o/edit?usp=drive_link&ouid=107389679865717903717&rtpof=true&sd=true)).

**Relevance & Limitations:** While this dataset will allow me to study the relationship of types of firearm deaths in relation to population groups and states, I did have to create this dataset with the idea of my census dataset in mind. This dataset also does not specify if any of these incidents are self-defense, which could skew the data results. This dataset specifically studies 2009 to 2017, thus putting it at seven years old. There are several outside factors, socioeconomic factors, and legislation that could have potentially skewed these numbers since 2017. These factors will be kept in consideration for the final report and be left open to interpretation.

**Project Hypothesis**

The perception of firearms is greatly influenced by several factors, including educational materials and media perception, which can have a huge impact on firearm related incidents and deaths. Using data analysis of firearm related incidents and deaths will help visualize regions and demographics in need. This information can then be used to create custom education materials for each region and demographic with the long-term goal of reducing the number of firearm related incidents and deaths.

1. If there is a larger state population, there will be more firearm related deaths in all categories, with homicide being the leading cause of death.

**Data Exploration Questions**

**Clarify**

1. Which firearm incident is the most common amongst all age groups and genders?
2. Which state has the most incidents and deaths?
3. Will homicide be the leading cause of death in all age groups across all states, or only largely populated states?

**Funnel**

1. Which age groups and have a higher death count? Do males or females have higher deaths?
2. Is there a noticeable trend in population size when compared to the number of incidents and deaths? Is there a specific age group that is commonly affected in large population states versus small population states?

**Privacy & Ethics**

1. Does the dataset contain any information on individuals, or has it been properly removed?
2. Does this information accurately reflect all age groups regardless of race, minority, socioeconomic status, or level of education?
3. Does this information include all reports of incidents and deaths, or are there exclusions due to victim age, privacy, security, or non-reported incidents?
4. What are the limitations of my analysis and what is the best way to present a detailed report that focuses on informing the public on the facts relating to firearm incidents and deaths?