DTSA-5304 Final Project

**A brief recap of your data, goals, and tasks, focusing on those that most directly influence your design**

I would like to explore school shootings in the US since 1999. The data set can be found here (school-shootings-data.csv)

The key attributes that I would like to explore are school district, city, state, type of school, number of students enrolled, killed, injured as well as shooter information such as age, gender and his relationship to the incident.

The key questions I would like to learn are:

1. How many people were killed or injured over time?
2. What is the distribution of school shootings by city and state?
3. Is there a relationship between the shooters age number of shootings?
4. Is there a relationship between shooting type, targeted or non targeted and the number of people injured or killed?
5. Over time, are there any times of the year where shootings are more common?

**Screenshots of and/or a link to your visualization implementation (see below for additional guidance)**

Link to visualization:Please open the Jupyter file associated with this submission. Note, to see the visualizations, you MUST download the data into the same folder as the Jupyter file.

Screenshots of the five-design sheet approach:

Text, letter

Description automatically generated

Diagram

Description automatically generated

Diagram

Description automatically generated with low confidence

**A summary of the key elements of your design and accompanying justification**

Task 1 - Analyze people impacted by school shootings. Explore data to assess similarities between shootings. Navigate between overview levels to find large scale patterns and individual occurrences to find small scale changes in people impacted. Use the outcomes to examine relationship between shooting type and people impacted. A breakdown of this task into its individual components can be found here:

* Purpose - (Explore) Consume data on people impacted by school shootings to establish new relationships between variables, if any exist.
* Means - Organize data by grouping people impacted when shooting was targeted vs non targeted. Allowing the end user to filter between total people impacted and subsets such as killed or injured to determine if a relationship exists.
* Data Characteristics
  + Low level characteristics - compare maximum people impacted when the shooting was targeted or non-targeted
  + High level characteristics - is there a trend in targeted or non-targeted school shootings over time
* Target data - (Absolute reference frame) number of people impacted by shootings observed over time
* Workflow
  + Information we have going in: number of people injured or killed in each shooting, shooters motive (targeted or non targeted)
  + Information we want to take out: people injured in shootings over time, whether targeted or non targeted shooters result in greater people impacted
* Roles
  + Who is looking at this? Principle
  + Why are they looking at this? Making a decision on if school should invest money into programs that foster team building, decrease bullying, and increase mental health awareness
* Thought Process for iterations
  + I combined idea 1 and 2 into one graph by introducing a drop down menu that lets the user switch between total casualties, killed and injured individuals.
  + From there, I used colors to identify the different motives of the shooter.

Task 2: Analyze geographic occurrence of school shootings. Explore data to assess distribution of school shootings. Navigate between overview levels to find large scale patterns and individual cities to find small scale changes in people impacted. Use the outcome to examine difference in school shootings by geographic occurrence. A breakdown of this task into its individual components can be found here:

* Purpose - (Explore) Consume data on geographic location of shootings to determine if there’s a relationship between geographic location and shootings
* Means - Organize data by city and state of shootings. Allow the end user to filter by state to see if there’s a pattern in shootings in a geographic location over time.
* Data Characteristics
  + Low level characteristics - which state had the least number of people impacted by a school shooting
  + High level characteristics - is there a trend in shootings by state
* Target data - (absolute reference frame) data is compared against geographic location
* Workflow
  + Information we have going in: states and cities where shootings occurred, number of people injured or killed in shootings
  + Information we want to take out: death or injury rate from school shootings in each state
* Roles
  + Who is looking at this? Analyst
  + Why are they looking at this? Lobbying for stricter gun control in their state by looking at incidents in states with stronger or weaker gun control laws
* Thought process for iterations
  + I began by looking at casualties across states however wanted to get a closer look at the geographic distribution within a state
  + As a result, I build a overview + detail visualization. Now, the user can look at the state level trends, and then pick a state to see casualties for cities within the state.
  + I hadn’t planned to include public vs. private school information on this visualization however once it was built in Jupyter, I was curious if there’s a trend in casualties when the shooting occurs at the public or private school and whether that varies by city or state.

**A discussion of your final evaluation approach, including the procedure, people recruited, and results. Note that, due to the difficulty of recruiting experts, you can use colleagues, friends, classmates, or family to evaluate your designs if experts or others from your target population are unavailable.**

* The target question you want to answer
  + Does the visualization provide insight into US school shooting casualties by state?
* The people you would recruit to answer that question
  + Data analysts, policy makers of gun control laws (ideally)
  + For this assignment, I used my family (software engineers)
* The kinds of measures you would use to answer your data and what these measures would tell you about the core question
  + Time to determine casualties between various types of shooters. Time to determine how school shootings occur geographically. Qualitative feedback on ease of use of the visualization.
* The approach you will use to answer that question
  + Thank aloud studies for qualitative feedback and an experimental design for quantitative feedback
* How would you instantiate those methods?

1. I would show my participants the tool and explain how it works
   1. For example, I’d highlight the dependent and independent variable as well as identify the different ways that they can explore the data.
2. Then, I’d get the participant to use the tool to explore the data.
   1. The goal here is to see how finding one pattern leads to additional questions and if they’re able to find those answers within the data set. I’d also like to see how easily they’re able to use the tool once it has been presented to them.

* What criteria would you use to indicate that your visualization was successful
  + Abie to quickly identify trends across time and geography (time to insight)
  + Qualitative feedback from users about the system adoption
* Results
  + Time to determine trend in casualties between various types of shooters was on average 65 seconds.
    - Note: this is an average from 3 family members
    - All participants noted that non targeted shootings were the most common
  + Time to observe trends in school shooting between states with strict and non-strict gun control laws was on average 50 seconds
    - Note: this is an average from 3 family questions
    - Additional questions that arose from this observation was if there was a significant difference between number of students impacted when the shooting occurred at a private or public school

**A synthesis of your findings, including what elements of your approach worked well and what elements you would refine in future iterations.**

* The elements of my approach that worked well was the ability to see in real-time how my visualization was being used. Showing the participants how the visualization tool worked helped them jump right into what they wanted to look at. The participants shared that without the introduction, they wouldn’t know where to start or what the tool could even do (ex: clicking one bar in the graph cause the second graph to filter).
* For the future, I would like to measure the number of interactions and explore that against time to interaction. This would show me if and where my visualization slows or confuses the user. While thinkalounds may catch this, the best way to confirm is by exploring the data.