Fatal Encounters

Submitted by Jacob Martin and Robert Bintu – 07/30/2020

Introduction

Our final project is on a dataset from the Fatalencounters.org website. Fatal Encounters is a website that was founded by D. Brian Burghart, an award-winning journalist and part-time researcher for the University of Southern California. The objective of the site is to provide "a step toward creating an impartial, comprehensive national database of people killed during interactions with law enforcement". Fatal Encounters documents non-police deaths that occur when police are present or are precipitated by police action or presence. Officer deaths are included when caused by another officer, including friendly fire incidents, and criminal actions e.g. domestic violence, and suicides that occur when other officers are present. Officer vehicle-related deaths are included when they are caused by another officer. Homicides of officers by felons or deaths in the regular course of duties are not generally documented in the database. The data is believed to include all the available records for all 50 states and DC back to 2000, but includes the following caveats

- Several data points that are thought to be too poorly reported in the news media to result in accurate results for analysis.
- The racial data is the best that exists, but it's pretty spotty and gets worse prior to 2013.
- Government data also suggests that police chase deaths are often not reported in news media, so our data almost certainly understates those totals
- The vast majority of these records come from media sources and police records meaning only one version of the story (i.e. the police story) is generally told in the descriptions.
- The site is not a finished product but rather a first step toward creating an impartial, comprehensive and searchable national database of people killed during interactions with law enforcement.

Domain problem characterization

This site is presented as impartial and data-driven as possible. It is driven by the theory that Americans should be able to answer some simple questions about the use of deadly force by police: How many people are killed in interactions with law enforcement in the United States of America? Are they increasing? What do those people look like? How can this data be easily visualized so that the general public can benefit?

It is these questions that the project aims to provide an added layer of transparency through data visualization on how many people are killed during interactions with law enforcement. We also wanted to shed light on why they are killed and if there are common threads in the data on state, gender, and race levels.

Using data visualization, our project aims to:

- Create easy to understand visualizations so that others, who do not have data visualization skills, can benefit from the dataset
- Provide interactive ability to filter out the deaths of specific interest to the user including gender, race, and cause of death
- Show trends across a state level through geographical mapping
- Allow visualization of the description text of each incident

Data/operation abstraction design

We felt like the data was structured in a way in which it would be useful to allow users to visualize based on geographical region. The extent of our project was only to make a user interface so that others can take it a step further to explore if there are deeper trends. They would then be able to explore if state level laws fuel these events or see how population density affects the number of cases.

The dataset is provided via a google spreadsheet and comes from the website Fatalencounters.org. They describe the dataset by saying "We try to document all deaths that happen when police are present or that are caused by police: on-duty, off-duty, criminal, line-of-duty, local, federal, intentional, accidental–all of them." The full set of column names is detailed in the appendix information, but more generally, it is composed of information about the subject (gender, demographics, name, etc.), geographical information about the incident, and information about the incident itself (dates, description, departments involved, links to news articles, etc.). There were 28,434 observations included in the dataset at the time when we pulled it from the Fatal Encounters website on 7/14/2020. The data has been collected by their team since the year 2000 and includes to the best of their ability, all deaths as noted above. It is based on the data collected across 50 states for 29 unique parameters.

We also included a data file in our project of a geojson file collected from the leaflet.js tutorial page on creating a choropleth. It includes a geometry polygon which gives a shape object for each state. Then we attached added data from our Fatal dataset.

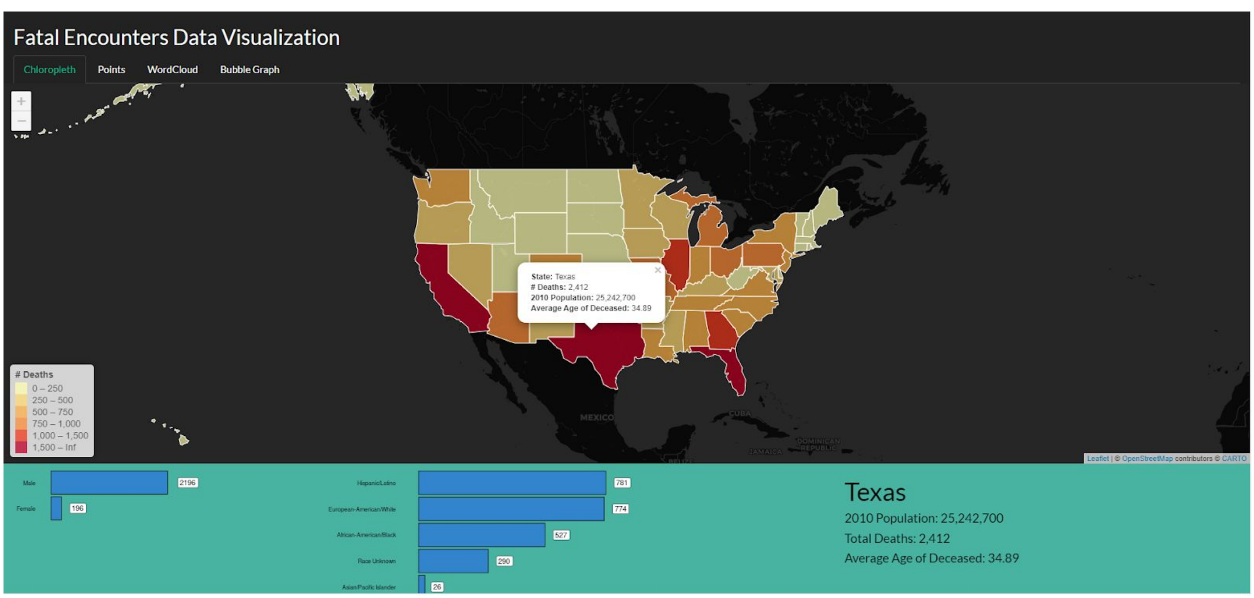
Lastly, we pulled from the web on population data for each state. We used the population data from 2010 since that was the most recent census year and because it was in the chronological middle of our dataset. We also added the state abbreviations to it for ease of performing a join operation to the geojson file.

To start the cleaning process of our data we began looking for data points that had possible data errors and removing data that may have integrity issues. Some of the data columns from the Fatal Encounters dataset were labeled as redundant or earmarked as "Not for analysis" so those columns were removed. There was one observation with the year 2100 so that observation was removed due to error. During the import of the data we needed to adjust some of the column names and amend data types to allow further analysis later on. The data fields were converted to a usable format as well.

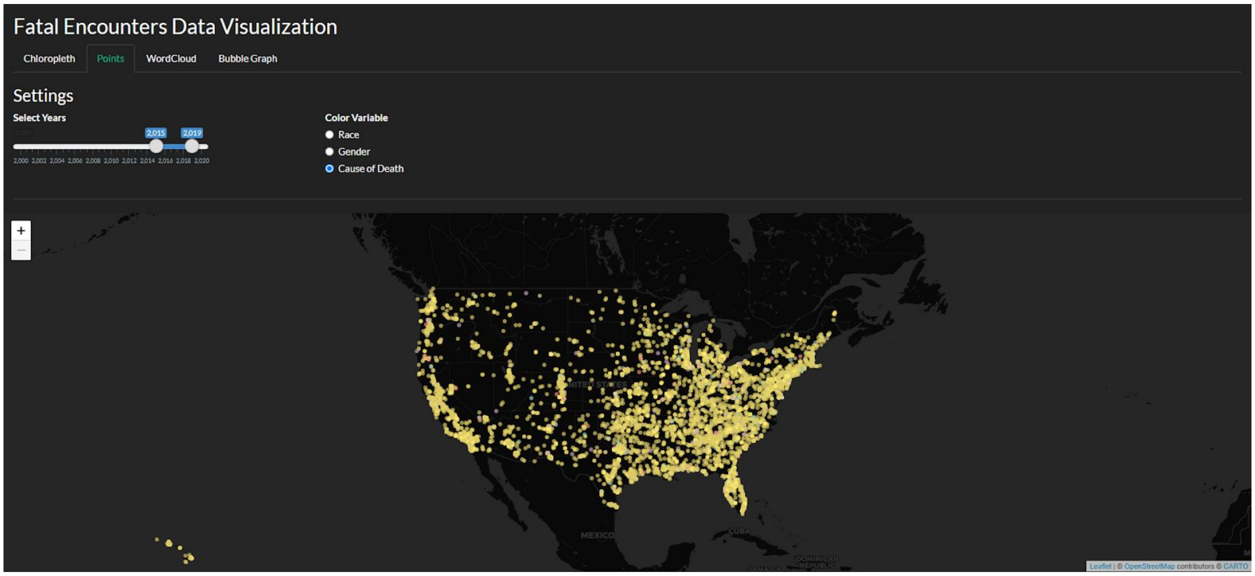
After the initial cleaning steps, the state geojson file was imported and we aggregated our Fatal Encounters dataset to a state level. Columns for the subject's average age, number of total deaths, and totals for race, gender, and cause of death classes were combined to the spatial polygon data. The spatial dataset will allow us to create a choropleth of the United States later on. The final metrics we added to this new geojson file were from the state population figures.

Encoding/Interaction technique design:

The dataset contains information linked to physical locations where the incidents occurred in the form of the nearest address and longitude/latitude. In this case, we found it helpful to visualize the data in their proper geospatial context i.e. to show the data on a map or alternatively as a map-like diagram (*Fundamentals of Data Visualization, Clause Wilke*). We employed the use of a common mapping technique, the choropleth map. It consists of data values represented as differently colored spatial areas. Each state is a polygon shape with data connected to it and as the user selects a state from the map they are able to see bar chart breakdowns of two categories. These show the breakdown of gender and race. We decided that the most simple way for users to interact with the data was to select the state directly on the plot by clicking the shape object. The main purpose of this plot is to allow users to compare on a state by state level. There is also a way to standardize the dataset to show the choropleth colors represented by deaths as a measure of the total population. This is helpful so that population metrics are included and the larger population states aren't more heavily skewed.



Our next visualization is a map that plots the points of every death included in the dataset. It enables the ability to show points on the map with the color based on race, gender, and cause of death. We felt that out of the dataset these were the most interesting metrics to display. There is also a slider tool to filter by the year the user wants to view. It is set up as a range selector so multiple years can be set at the same time. There are over 28,000 data points to display though so it is often best to limit the range to 5 or less years to aid loading times. This is an interesting graph because it shows where deaths occur in geographic position. At a glance, it seems that density of points occurs near large population centers. This logically makes sense though since there would be more police presence in those areas and there would be more people as well. A display of this type is helpful because it makes the data more personable. When the user is looking at the individuals affected in each incident they are able to select the related link and see the victim's photo.



There was also a column of text information in the Fatal Encounters dataset. The third tab on our shiny app is a wordcloud. These are easy to understand and interpret since the words most commonly used will be at the center in larger font. The color differentiation is based on the number of times a word was used. This feature makes it easy to interpret how points relate to one another. We included the ability for individuals to base the word cloud on the cause of death. The field can either be selected to represent all causes of death or a single one. Then we allowed for some flexibility in the custom settings for the number of words to include and size of the words. Overall, this plot is a simple one but is helpful in understanding how the text in the "A brief description of the circumstances surrounding the death" column is organized.

Algorithmic design:

Our system is designed to give several different granular levels of the data. The choropleth shows state by state comparison, the individual points graphing shows the most granular with individual cases, and the word cloud shows different levels of the cause of death in the text. At each stage the main characteristics we wanted individuals to be able to sort by were race, gender, and cause of death. We accomplished this through different selection and slider mechanisms which provide interaction with the user. Interaction in our system helps to achieve our goal of visualizing the dataset. Another aspect we were shooting for with the data was to make simple interactions so that users of all levels are able to operate the system. This was especially achieved in the maps as it requires the user to select the object of interest.

The system was also tested to ensure it isn't sluggish. There is a slight start up lag where the shapes are being added to the choropleth. After the initial loading the system performs well and isn't sluggish with casual use. If the user selected many items at 1 time on the choropleth it will cause some lag though. The individual points map also has some sluggishness when the user displays all points on the map. It is advised that when using this graph that only a smaller portion of the points should be plotted. The issue with both of these has to do with the amount of data that is being displayed on the map and isn't something we were able to fix.

One issue to note with the word cloud plot is that some causes of death do not have enough values to generate a high number of words. Since there were so few of some causes, when the user selects them, the plot ends up showing many words that were only used 1 time in the dataset. This may cause some confusion to the user if they think certain words were very common throughout the dataset as a result. The solution to this will be future collection of data and the addition to the word cloud. The most common cause of loss was gunshot with about 70% of the cases.

User evaluation

The evaluation of the designed system was performed during the design phase of the application. A formative evaluation i.e. directed at identifying potential problems and indicating how to possibly improve the system design (*Introduction to Information Visualization, Riccardo Mazza*), was employed. Determining the specification of the requirements was key to evaluating the interface system. The objectives of the project were clearly defined as presented in the "Domain Problem Characterization" section, and were identified prior to its commencement that were utilized to create an evaluation strategy. An analytical method was used to evaluate whether the system is compliant with heuristics defined on the principles of Usability and Accessibility of the interface. The main goal of these heuristic evaluations is to identify any problems associated with the design of user interfaces before being adopted by the final user. Five of the usability heuristics for user interface design (*Usability Engineering, Jakob Nielsen*) were applied to the evaluation of the created system. The following analysis defines the criteria for the evaluation strategy:

Criteria	**Description of Criteria**	**System Evaluation**
Visibility of system status	The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.	Interactive design provides user the ability to look up specific information by state and breakout of demographics by Gender and Race
Match between system and the real world	The system should speak the user's language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.	Language used mimics the information in the data set that relates to the objective of the project
User control and freedom	Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.	Basic Point and display system that delivers the information instantly with no ability to select unwanted information
Consistency and standards	Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.	System uses consistent terminology aligned with the subject of the project
Error prevention	Even better than good error	App is designed to use data

	messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.	directly from the data source and provide visualization graphic with no data distortion

Future work

The goal of our system was to create an easy to understand visualization of the dataset provided by the Fatal Encounters team. We accomplished that through mapping the data geographically, but it may be interesting to explain further questions as to why certain locations are hot spots. This type of work could be done by overlaying population centers, dangerous intersections and roads, or possibly looking at areas of poverty to see how those match up to deaths in our data. Further applications could be created using modeling to predict future hot spots of deaths involving a police presence. Models may also be useful for determining what factors play the largest role in why an area is prone to have events.

One area of interest that we weren't able to broach is looking at trends over time. Our dataset has metrics for dates that events occurred. With that we would like to look at forecasting and trends over time to help answer some questions we had about seasonality or annual changes. It may also be interesting to look at legislation and state level policies to see how those affect the number of data points after enactment. This is one area we were not able to include due to time constraints, but would like to revisit it in the future.

Our methods of visualization were designed for the common user so that they could see the data in an easy to understand method. The next stage of visualization would be to create more complex interactions that could be used for policy changes. The creators of the dataset aim to provide the data so that future policy changes can be enacted to prevent high numbers of deaths with law enforcement interaction. Fatal encounters aims to provide data "that can be used to facilitate improvements in training and policies to decrease the number of officer-involved deaths"

Appendix

Fatal Encounters Dataset columns

- Unique ID
- Subject's name
- Subject's age
- Subject's gender
- Subject's race
- Subject's race with imputations
- Imputation probability
- URL of image of deceased
- Date of injury resulting in death (month/day/year)
- Location of injury (address)
- Location of death (city)
- Location of death (state)
- Location of death (zip code)
- Location of death (county)
- Full Address
- Latitude
- Longitude
- Agency responsible for death
- Cause of death
- A brief description of the circumstances surrounding the death
- Dispositions/Exclusions INTERNAL USE, NOT FOR ANALYSIS
- Intentional Use of Force (Developing)
- Link to news article or photo of official document
- Symptoms of mental illness? INTERNAL USE, NOT FOR ANALYSIS
- Video
- Date&Description
- Unique ID formula
- Unique identifier (redundant)
- Date (Year)

References

Fatal Encounters Dataset by: D. Brian Burghart and the entire Fatal Encounters team. They can be found at https://fatalencounters.org/ and the dataset can directly be pulled from their google sheets link at: FATAL ENCOUNTERS DOT ORG SPREADSHEET

State Geometry file provided by Mike Bostock (https://bost.ocks.org/mike/) and pulled from the Leaflet.js team's example (https://leafletjs.com/examples/choropleth/)

Inspiration for the Point's tab visualization from the Fatal encounters visualization team at https://github.com/adv-datasci/fatalencounters and Visualization by Kenneth Morales

Introduction to Information Visualization, Riccardo Mazza

Usability Engineering, Jakob Nielsen