INFO 3300 Project 2 Report

Sunil Sabnis, Rishik Zaparde, Esther Lee

a) Description of the Data

Criteria:

Through our visualizations, we wanted to explore 1) police violence and the victims in different regions of the US and 2) the relationship between number of victims and police deaths. For 1), we set the criteria for data selection to have the location of the incident, since our visualization will include the US map. Also, we wanted the dataset to include victims' personal information, especially age, gender, and race. For 2), we set the criteria to have both the number of police deaths (or a list of police officers who died) and location (city, county, etc.).

Sources/Filtering Process:

We got our first dataset from https://mappingpoliceviolence.org/. The dataset was gathered by researchers interested in mapping deaths caused by police officers at the individual level with a high granularity of data. Initially, we tried using datasets generated by government agencies, but the granularity was not specific enough. We did the data cleaning using the python pandas and the geopy package and a simple JavaScript filter function. We got rid of data that had null values in d.position, city, and state (but not the age and race). Since we were considering plotting the US map, we used a python3 library geopy, (https://geopy.readthedocs.io/en/stable/) to extract the longitude and latitude based on a US city and state. Specifically we used the Nominatim geocoder (https://nominatim.org/release-docs/develop/api/Search/) to translate city, state data for each row of the police_revised.csv into longitude and latitude coordinates. This required creating a state dictionary to convert state abbreviations and a considerable amount of string processing. The API was unable to convert about ~15 cities out of ~9000, which might be due to spelling errors of the cities in the dataset. We noticed some cities were miscalculated, and we had to hardcode the coordinates, but for the majority of cities and towns the geocoder was able to successfully locate them. We got our second dataset from https://github.com/fivethirtyeight/data/tree/master/police-deaths , which was created by 538. This dataset includes a list of police deaths, along with the officer's department, state, and year. We tried to use the geocoder to generate coordinates from department names, but the geocoder was only able to locate ~50% of the police departments. Thus, we decided it would be best to show police death data by state. This also makes more sense because there are considerably less police deaths than deaths caused by police officers. For the police_deaths_cleaned.csv dataset, we filtered out year values < 2000. We wanted the years from each dataset to be relatively comparable. See the jupyter notebook for the specific technical details on how we cleaned and processed the datasets.

To display a victim in our visualization, we used red and black standing man figure svgs, which we got from https://www.svgrepo.com/svg/6998/standing-up-man. We incorporated a multi touch slider using a simple slider d3-esque library: https://github.com/johnwalley/d3-simple-slider. Initially, we tried to design a multi-touch slider from scratch, but ran into many issues.

Integration:

Rather than integrating two different datasets, we imported two different datasets to the code and worked on it separately. We were planning on having a choropleth and a map with points on top of each other, so we did not find it necessary to combine the datasets.

b) An overview of the visual design rationale.

1) US Map

   Since our major goal is to show police violence throughout the US, we wanted to include the US map as part of our visualization. We made the US map to take the majority of the space, but not the entire website. This is so that the users can easily navigate the US map, while having enough space on the side to show the details of the data, such as the number of victims and information (name, age, etc.) about a victim.

2) Marks + Design Decisions

   Our marks include circles on the US map and human figures that show up on the side of the website. We thought using circles and human figures would be the most neutral/appropriate shape to show the victims. The opacity of the circles was set to 0.4, since some larger circles were covering smaller circles. In terms of colors, we used a neutral color (gray) for background, black for non-selected victim, and red — which we thought was the best color to represent death — for selected victim and the circles on the map. We avoided flashy colors and limited the number of colors used for the visualization because of the seriousness of the topic.

3) Channels

   Our channels include positions and areas of the circles and color saturation (of the states). The positions of the circles represent a location in the US, and the size of the circles represent the number of victims. We chose to make the size of the circles on a linear scale, since there were no outliers. The color saturation represents the number of police deaths, with high saturation indicating more police deaths (darker blue). The color saturation was based on how many police deaths occurred in each state using a quantile scale. Because of this, we were able to account for the whole death distribution in coloring each state.

c) An overview of the interactive elements and their design rationale

We implemented multiple interactive elements in our visualization, and designed it so that the interactive elements are usable, discoverable, and intuitive.

1) We made a slider that can be used to filter the circles on the map based on the number of victims. We were considering multiple options, such as checkbox, buttons, and textbox, but we thought that slider would provide users with more detailed inputs than checkbox and buttons (which would only allow a range of values as input). Also, we thought that textbox will not make the visualization interesting/tangible, and users would have to take more steps to filter the data if we end up using a textbox. The slider allows the user to dynamically filter the data in a very intuitive manner.
2) The map in visualization can be zoomed, so that users can click on smaller circles easily. The sizes of the circles vary, and there were a lot of small circles that were hard to click when the map was

not zoomed in. So, we thought including the zoom-in feature was essential and would ease the navigation of the US map. The map is usable and intuitive, since many users associate scrolling in (mouse/trackpad) as zooming in on a map. When the map is zoomed in, the users can see the county boundaries so that they can see the specific city/county, which is animated in smoothly.

3) The map can also be dragged to ease the exploration of the US map.
4) We made the circles and human figures clickable. We wanted to let the users navigate the details of police violence in the US. We thought allowing users to click on different locations on the map was the most intuitive way to do so.
   When a circle is clicked, it displays human figures on the side, and each of them represents a victim. The circle also is modified with a grey border to signify that it is selected. Clicking on the victim reveals specific information of the death such as the reason for death, age, name, race, and context of the killing. Instructions are shown initially when the web page is first loaded. The person is changed to a red color when clicked to signify that it is selected.
5) We added a button that displays a choropleth of police deaths. We thought using a button was the simplest way for users to toggle on/off the choropleth. It allows users to turn off the choropleth if it is distracting them from viewing the points. The choropleth button and map are smoothly animated simultaneously when clicked.

d) The story

Our visualization shows that there tends to be a higher number of victims of police brutality in states where there are many police deaths. Also, it tells us that there was at least one victim of police brutality in each state from 2013 to 2021. Moreover, over the past eight years, there were a total of ~9000 — a surprisingly high number — deaths from police violence in the US. While displaying the overall trend of police violence in the US, we ensured that our visualization allows users to learn about every victim at a personal level. We made this visualization so that users can learn about police violence in the US and recognize the relationship between the number of victims and police deaths.

Team Contributions

Esther: I helped come up with visualization ideas and wrote the final report. I also helped implement choropleth and zoom-in features. I spent most of the time writing the report.
Sunil: I found the datasets from 538 and https://mappingpoliceviolence.org/. I cleaned the dataset in a jupyter notebook (using pandas and geopy). I implemented the button to turn the choropleth on/off, the multi touch slider, the function to draw the circles appropriately, and zoom functionality. I also helped implement other features of the map visualization such as map initialization, css, and clicked circle/person events.
Rishik: I created the map projection and the choropleth, which displayed the police death data. I also wrote functions, which helped display individual victim data and created the color scale and legend for the choropleth. I also helped to style the components into divs arranged on the page.

Sources:

https://mappingpoliceviolence.org/
https://github.com/fivethirtyeight/data/tree/master/police-deaths

https://geopy.readthedocs.io/en/stable/
https://nominatim.org/release-docs/develop/api/Search/
https://www.svgrepo.com/svg/6998/standing-up-man
https://github.com/johnwalley/d3-simple-slider