Visualization of Misconduct in New York City Police Department

Group number: G01
Akshay Deshpande
Chaitanya Mundle
Pranav Kalbhor
CPCS 4030/6030, Fall 2020

Introduction

Racial descrimination is one of the major social problems that we are facing across the globe. According to a report by the Brooklyn Daily Eagle, complaints against New York City police officers have spiked to 20% as compared to last year. The complaints against uniformed police officers mainly involve offensive language, discourtesy, abuse of authority, and excessive use of force. The job of the police is difficult but they have to perform their duties within the limits specified by the law and our free society. The recent incident involving George Floyd displayed similar offensive behaviour by the cops which led to nationwide protests. Statistics on several websites show an exponential increase in such offenses, highlighting an increase in police brutality. In this project we aim to visualize the racial discrimination taking place in New York City Police Department.

Dataset Description

In order to analyze how often racial discrimination is happeinig and how is it being tackled, we decided to analyze visually the NYPD complaints dataset. This dataset mainly flashes out broader information about the number of complaints filed against police New York City police officers by the civilians by considering different parameters like an officer's rank, complaint date, outcome, allegations, etc. An important factor that the dataset comprises is the span for which the complaints have been registered, which would be from years 1985 to 2020. The dataset constitutes the name of each officer, the designation of the officer, the details of the complainant and the officer, and the decision of CCRB against the officer. The overall size of the dataset consists of 33,359 tuples with 27 attributes wherein some data cleaning was required to refine the data and then the required data analysis is performed.

The source for this dataset:

Civilian Complaints Against New York City Police Officers

Design Solution

Story Description

Considering the problem statement, we decided to analyze the trend of crimes followed by the precinct wise distribution of allegations with respect to ethnicity and finally determining if there exists any racism by visualizing all three graphs. The first level of our analysis can be give by the trendline graph were we only tend to show the variation in the number of allegations over the years, In the second level we go in depth of the number of allegations by representing the distribution of allegation across different precincts and finally it shows followed by analysis of complaints with respect to ethnicity. The in-depth description of each visualization is given as follows:

Visualization 1: Interactive Bar Chart

The First Visualization on our website describes the trend of complaints from **1985 to 2019.** The trend of complaint is an essential visualization to get a perspective of the number of complaints being launched against the officers of New York Police Department over the years. The Interactive Bar graph is used to show this trend of complaints that indicate 2016 to be the year with a maximum of 2345 complaints. While deciding on the visualization for displaying the trend, the line graph seems an obvious choice, but on analysis using tableau, we found bar charts to be more effective since the X axis had a large range of years to display and bar charts help to clearly distinguish data of one year with another. The visual Encoding for this graph is as below:

- Dataset Type: Bar chart
- Data Type:
    - Item: Number of complaints represented with the mark line
- Attribute:
    - Number of complaints represented by visual channel length
    - Year received encoded by visual channel position

Visualization 2: Map Graph for New York City

Once we have the trendline displaying the number of crimes each year, we explore the dataset further to analyze the number of crimes occurring at each precinct. Considering visualization options with precinct data, a map describing the data with a channel of hue seems to be an obvious choice. This Visualization complements the first visualization of the trend of number of allegations by describing the number of allegations across different precincts. We can identify which precinct has contributed towards the maximum and minimum number of allegations. The map of New York city has radio buttons below to

select the ethnicity of the complainant, that categorizes the number of allegations on the police officers by citizens of various ethnicities. Map graph was chosen since it makes it really easy to view the geographical distribution of allegations across New York. The Visual Encoding for this visualization are:

- Dataset Type: Graph
- Data Type:
  - Item: Precinct represented with the mark area
  - Item: Number of complaints from people of different ethnicity represented by color
- Attributes:
  - Number of complaints encoded using the visual channel hues (color) and position
  - Precinct of New York City using the visual Channel area
  - Ethnicity of complainant (Radio Button) not encoded

Visualization 3: Interactive Sankey Graph

The third Visualization dives deep into the second visualization and describes the distribution of complaints with respect to the ethnicity of the complainant and ethnicity of the police officer. The Sankey graph has the officer's ethnicity on one end and complainant's ethnicity on the other end and the distribution clearly shows that the officers with white ethnicity ethnicity have the maximum number of allegations against them and the maximum allegations are from the citizens of black ethnicity. This visualization mainly answers our problem statement of determining racial discrimination since we can observe a clear majority of black complainant;s against white officers. Hue is used as the channel in this graph to distinguish between different ethnicities.

The Visual Encoding for this visualization are:

- Dataset type: Graph
- Data Type:
  - Item: Complainant Ethnicity represented with the mark Line
  - Item: Mos Ethnicity (Police Officer's ethnicity) represented with mark Line
  - Link: Number of complainants for an officer of a particular ethnicity represented by the mark area
- Attribute:
  - Number of complaints encoded using the visual channel hues (color) and position
  - Number of complaints lodged by the people of a particular ethnicity using the visual channel length
  - Number of complaints lodged for an officer of a particular ethnicity using the visual channel length
  - Number of complaints encoded using visual channel area
  - Police and complainant ethnicity encoded using the visual channel position

Literature Review

The Data Visualization experts Edward Tufte and Stephan Few insist on visualization to be free from chart junk, clearly show all the data and are easy to comprehend without any distractions [I] [II] [III].

The research [IV] shows the emotions of India based on live twitter data using an Interactive bar graph and Map. In this research the x-axis of the bar graph represents the mood or emotions of the people and the y axis represents the count. The visualizations also include a map graph to display state wise emotions and is very similar to what we have done in our project.

Huge capital is spent every year by the U.S. Department of Health and Human Services to keep track of various health parameters across the United State that result in a large geographically varying dataset. In this research [V] paper the authors describe a web application called Community Health Map that allows users to view the health care data geospatially. The Visualization represents an interactive map of the US with all the essential parameters and this visualization can be used by journalists, policy makers, academic researchers and Consumer groups to draw key insights. The visualization has dynamic filters which helps to narrow down the data on the map to specific regions and ranges. The visualization targets to address the following:

1. Visualize Health Parameters across the United States or Hospital Referral Region in the country
2. Include essential filters to categorize the data based on demographic factors such as age, poverty, income, and education
3. Produce Visual Conclusions to assist policy makers to take relevant actions to improve the Health care in their regions.

The Sankey Graph is used to represent complex flows and, in the research [VI], the energy flow in a city is visualized using a sankey graph. The various forms of energy are distributed within the city and they are transformed into many forms like heat, electricity. These processes have been visualized effectively by having interactions on the designed Sankey graph. This graph is used to answer the following questions in the research paper:

1. How is the energy distributed?
2. Where are the biggest losses?
3. What kind of energy is used for which purpose?

References:

[I] S. Few. Benefitting infovis with visual difficulties? Provocation without a cause. Visual Business Intelligence Newsletter, 2011

https://perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

[II] S. Few. The chartjunk debate: A close examination of recent findings. Visual Business Intelligence Newsletter, 2011.

https://perceptualedge.com/articles/visual_business_intelligence/the_chartjunk_debate.pdf

[III] E. Tufte. The Visual Display of Quantitative Information. Cheshire (Conn.), 2001.

https://webspace.science.uu.nl/~telea001/uploads/VACourse/Tufte01-2.pdf

[IV] Venigalla, Akhila Sri Manasa, Sridhar Chimalakonda, and Dheeraj Vagavolu. "Mood of India During Covid-19-An Interactive Web Portal Based on Emotion Analysis of Twitter Data." *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*. 2020.

https://arxiv.org/pdf/2005.02955.pdf

[V] Sopan, Awalin, et al. "Community Health Map: A geospatial and multivariate data visualization tool for public health datasets." *Government Information Quarterly* 29.2 (2012): 223-234.

https://d1wqtxts1xzle7.cloudfront.net/42167908/Sopan2012Community.pdf?1454711040=&response-content-disposition=inline%3B+filename%3DCommunity_Health_Map_A_geospatial_and_mu.pdf&Expires=1607289338&Signature=JAvrHKeYSfKmL21eGqu5LCbzGLut1uNwut43yz20oVCTaY2B46EYtpPX4lSD0Hc~3R1ImEvEK0Jt1iP1EjMJCB~0xNq9PX3RjuYc4daEPJWACTULxn2sElJ~MEswaIXD3FWgL8NHFuNdnFsIwGN4yF-Vn3C9VcZRLINIc7AgVAIZQq3R4vKIeiyIn6ouhJ4Hc53CIsqx-W50SstK7vrdtOTe6KDBZ7lK4EeArgt4hXsAB2WTTVzydcfXnPzk5tWESpI4a5MSsZ9mc8KeWnbXH8HOBs7XybyNYxR7jCYu5TvjEZksc~ZGH26334Vd7j1xo-x9jiB1eoNTdxXt9GXAdw__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[VI] Froehlich, P. "Interactive Sankey Diagrams." *IEEE Symp. on Information Visualization*. 2005.

https://ieeexplore.ieee.org/abstract/document/1532152?casa_token=NwBWTO1KjZgAAAAA:Jf_S7Jd7MngzKzCxx_8RiJGkZbzqedsPw7E0CF1uRVAcCCe0k9Ae-HeBtF2hZOoRfjH3jjEy