GR5063 - Data Visualization Final Report Proposal

Team Members:
Nicholas Ding Yang Choong
Rei Hrn Lim
Jaydon Jun Yu Chin
Jason Chien

Title:
New York City (NYC) 311 Non-Emergency Complaints

Abstract:
We hope to explore patterns in urban complaints, the efficacy of municipal responses, and the intricate web of socio-economic and geographical factors influencing these dynamics. The primary motivation for this is to provide valuable insights into the city's operational strengths and areas for improvement, offering a comprehensive understanding that could inform policy decisions, resource allocation, and urban planning strategies to enhance the quality of urban life.

Preliminarily, we hope to explore the following:

- Basic data comparisons:
  How different is the average response time across different complaint types and agencies?
  Investigating the time taken to close complaints after they have been reported could be one measure of the efficiency of city agencies. We hope to investigate this further and visualize this comparison of response times by complaint type or agency through bar charts or box plots.

- Temporal Patterns of Complaints:
  What are the trends and seasonal patterns in complaint types over time?
  This can be explored through time-series analysis. Line graphs and heatmaps would help with the identification of peak times for specific complaints and assess changes in complaint volume over the years.

- Geographical Distribution of Complaints:
  How are different types of complaints distributed across various neighborhoods in New York City?
  This question invites a geographical analysis that maps the concentration of complaints, potentially uncovering areas with higher incidences of specific issues. Geospatial heat maps and choropleth maps would be ideal for visualizing these distributions, facilitating the identification of hotspots for targeted urban interventions.

- Text Analysis of Descriptors:
  What common themes emerge from the descriptions of complaints?
  We hope to utilize natural language processing (NLP) techniques to analyze the text data in complaint descriptors. This can potentially reveal prevalent issues that might not be explicitly categorized. Word clouds, topic modeling, and sentiment analysis could provide insights into the nature of complaints and public sentiment.

- Geographical Network of Complaint Types: By treating different geographical areas (e.g., ZIP codes, neighborhoods) as nodes and establishing edges based on the similarity of complaint profiles (types and volumes), one can visualize the similarity in complaint patterns across the city. Such a network could help identify clusters of areas with similar issues, guiding localized policy interventions. Techniques like community detection could be particularly insightful here, uncovering regions of the city with distinct complaint characteristics.

- Correlation Between Complaint Types and Socioeconomic Indicators:
  Is there a correlation between the type or volume of complaints and socioeconomic indicators of neighborhoods (e.g., income levels, population density)?
  This analysis could involve combining the 311 data with external datasets on socioeconomic factors and employing statistical methods to identify correlations. Scatter plots and correlation matrices could help visualize these relationships.

Techniques:
ggplot2, ggmap, Shiny, NLP text analysis (To be covered), Network Analysis (To be covered)

Dataset:

- The primary dataset we hope to use originates from the NYC 311 service. (https://data.cityofnewyork.us/Social-Services/NYC-311-Data/jrb2-thup/about_data)

  This crucial municipal service offers a public interface for city residents to report non-emergency issues ranging from noise complaints to road damage. Spanning from 2016, it encapsulates a wide array of data points such as the unique identification of complaints, the dates of complaint initiation and closure, the responsible agency, complaint categories, detailed descriptions, location specifics, and additional related information.

- One possible avenue of extension would be to merge this dataset with census data. (https://www.census.gov/data.html)

  Census data likely contains relevant socioeconomic indicators that might influence the type or volume of complaints, such as median household income, percentage of residents with a college degree, unemployment rate, and housing quality metrics. This would help answer our last research question: *Is there a correlation between the type or*

volume of complaints and socioeconomic indicators of neighborhoods (e.g., income levels, population density)?