

Title of the project:

Data Visualizations of Crimes in Chicago

Names of all project participants:

Camille Mancuso (clm2246@columbia.edu)

Cara Xu (jx2445@columbia.edu)

SooYeon Choi (sc4815@columbia.edu)

Herbert Dong (herbert.dong@columbia.edu)

Description of the proposed visualizations:

We are interested in exploring the crime dynamics of Chicago, a city known to have one of the highest crime rates in the US. For this project, we plan to focus on the year 2018, using data sourced from the Chicago open data portal, which offers public access to crime records in Chicago spanning from 2001 to the present. Our approach involves leveraging this dataset to create visually informative graphs and maps.

Through the project, we intend to shed light on how crime trends evolved throughout Chicago in 2018. Specifically, we aim to analyze:

1. The spatial distribution of crimes across Chicago, displaying frequency disparities among different locations.
2. The most frequent types of crimes occurring in each area.
3. Temporal changes in crime patterns over the course of 2018, exploring evolving trends over time.
4. Make some connections between the crimes and the location of police offices to visualize the potential links between the distribution of crimes and the local government's control.
5. The interrelation between the crimes and social factors related to local communities, such as household income and age.

Links to data sources:

Chicago crime data:

https://data.cityofchicago.org/Public-Safety/Crimes-2018/3i3m-jwuy/about_data

Chicago Map:

<https://data.cityofchicago.org/Public-Safety/Boundaries-Police-Beats-current-/aerh-rz74>

Chicago Data Portal:

<https://data.cityofchicago.org/>

Potential types of visualizations:

- Line chart (time series analysis): Using ggplot2, we plan to create a line chart that depicts the changes in crime frequency in Chicago across various crime types over time.
- Map: We would use ggplot2 and maps to visually represent the distribution of crime frequency and types across different locations in Chicago. Furthermore, by employing the leaflet function, we aim to generate an interactive map that displays regional crime trends through color differentiation. Besides, we will use the stadia map to visualize some figures to enhance the aesthetic properties of the maps.
- Bar charts: Since we know the location of each crime, we can create stacked bar charts that show the racial makeup of the most and least frequent areas for each particular crime type.
- Box plots: Since we know the location of each crime, we can create box plots that show the median household income of the most and least frequent areas for each particular crime type.
- Density Map: For visualizing which community had the largest crime rate and what the crime rate trend was around it. Combining it with OpenStreetMap to link it with other factors.
- Connection graph: Networks or Chords are likely to be used to visualize the connection between crimes and social factors.

Potential technical difficulties:

Thus far, we haven't encountered any substantial technical hurdles with data. However, it would be important to address the task of data cleaning and aggregation properly, given the substantial size of the original dataset and the multitude of variables it includes.