

Process book of Group_U

Group Member: shiyun li, xia shan, junjie ma, peiyuan huo

Idea development

We first wanted to work on analyzing people's thoughts on the recent Ukraine war. However, due to data complexities, we would need to look at multiple languages of data. It would be difficult to analyze these languages. If we only looked at English-language posts, we would exclude a lot of important opinions. We then turned to some other trending topics such as COVID-19, the past elections, housing prices, and increasing gas prices, but these were dismissed due to duplicate topics, data complexities, etc. Finally, we have decided our topic on analyzing the crime activities in Los Angeles since crime has become one of the most striking problems in the United States.

Process timeline

Group meeting 1: the first discussion about the project topic. Going through all the hot issues recently, we decided to conduct sentiment analysis about tweets regarding the conflicts between Ukraine and Russia.

Group meeting 2: conducted project agenda and divided the contents into three parts, that are abstract, dataset and visualization.

Group meeting 3: received TA's feedback and discussed certain problems in the prior topic. We decided to change our topic.

Group meeting 4 : chose crime analysis of LA as our final topic, constructed the agenda and timeline of final project

Group meeting 5 : completed the draft version and checked the process

Group meeting 6 : finished the final version

Group meeting 7: revised the project according to classmates' comments

Why the topic

In recent years, crime has become one of the most striking problems in the United States.

However, methods for analyzing patterns of association with place and time in crime research remain to be further developed. This project aims to help us better visualize and understand spatial and temporal trends in criminal activity.

Data source

<https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8>

The dataset is from LA open data, reflecting incidents of crime in the City of Los Angeles dating back to 2020. This data is transcribed from original crime reports from LAPD and is updated weekly.

We first targeted this dataset but the original file was too large to analyze, and there were too many small categories which were difficult to visualize. So we cleaned the data and narrowed it down to six types of crime which are all over 500 cases.

Visualize Data

- Crime across space
- Crime over time
 - Overall Trend
 - Divided into Category

Aggregate data

- the total numbers of each crime
- percentage of each
- histogram?pie chart? statistics?

Temporal trends

- pick one or two crime and think about the underlying influence factors
- overall trends over the past two years?
or
- weekly trends?
- line charts
- heatmap

Correlation analysis

- Want to see if certain types of crime are time dependent
- Fill crime types into 6 categories; time stamp data
 - Horizontal: day of week; Vertical: hour of day
 - Normalization to get a better view
- Factor By Police District to see if trends appear
- Factor By Month
- Factor By Year