

Preliminary Report

Group 13
Jin Chen
Huiyi Shen
Shuo Qiang

Introduction

Nowadays, society and the economy are rapidly developing, which brings convenient, comfortable, and diversification life to human beings. However, it seems to fuel conflict between people from different classes, various cultural backgrounds, or beliefs. In that case, although today's crime rate is much lower than the historical crime rate, we still always can see the news about rob, shooting, and other crimes. In the United States, most of the shops will be closed around 7pm, and people usually don't go to the street after 10pm. People still feel unsafe currently. Under this situation, our group wants to create a crime report to understand the crimes. We also hope to generate some useful advices to people to prevent them from crimes. Since it is difficult to analyze the crimes across the United States, we picked New York City as our analyzing object.

Data description

The dataset we want to use is called "NYPD Complaint Data Current (Year To Date)." This dataset includes all valid felony, misdemeanor, and violation crimes reported to the New York City Police Department for all complete quarters so far in 2019. Which includes the precinct in which the incident occurred, time, level of offense, suspect's age group, race, sex, and so on.

Proposed analysis

- The demographic pattern of the crime such as gender, race, and age
- The region has the highest crime rate
- What time during the day happens the most frequently
- Which type of crime happens most frequently
- Which type of crime happens most frequently in a certain borough

Analysis methods

First, we pick the columns we need and put them into several tables to from a database. For example, we put 'SUSP_AGE_GROUP', 'SUSP_RACE', and 'SUSP_SEX' plus 'SUSP_ID' as the primary key into one table named 'Suspects'. Then, we manipulate the data in the database, like fitting the various models to see the relationship between the crime rate and age. Finally, we will use the library like Matplotlib to visualize our findings.

Milestones

Now, we had the columns and the tables. But, we find the data needs to be cleaned. So we are trying to figure out how to clean the data. We may choose to get rid of the rows which contain the null value or ignore the null value; just use the existing data. We are also thinking about the models that we can use to fit the data. We plan to get the database done this week and finish everything in the next two weeks.

References

NYPD Complaint Data Current (Year To Date):

<https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Current-Year-To-Date-/5uac-w243>