CS591 L1 Data Mechanics

Factors that influence the number of crime incidents

Joe Zhou & Ekaterina Prokopeva

Narrative

The five datasets are employee earnings in Boston, location of public schools in Boston, crime incidents in Boston, colleges and universities in Boston and street lights in Boston. These datasets can be combined on location to see various statistics relating to crime in different counties of Greater Boston area. Specifically, we want to see the presences of schools on the occurrences and different crimes. Also, we want to see if there is a correlation between wage of police officers and general crime in an area. These questions are interesting because politicians often point to increasing security and education as ways to improve living conditions.

Data Sources

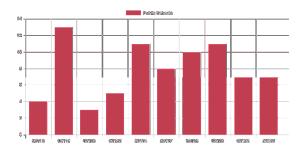## Crime Incidents In Boston

Techniques used:

Map Reduce to map only the report district and calculate the number of crimes occurred in every district.
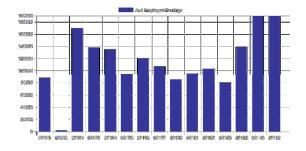We created our own dataset that maps the report district to a zip code.

Location of Public Schools

The given dataset contains the Location of Boston Public Schools for school year 2012-2013.

Employee Earnings

The given dataset contains the earnings of police officers for City of Boston employees for calendar year 2015. In order to retrieve useful results we had to find the zip code, the salaries within this zip code, and the average of these salaries. The technique we used was **Map Reduce**.

k-means

We used k-means algorithm in order to find the point with the highest density of crimes.
By traversing the dataset that contains crime incidents and their coordinates, we obtained the value equal to **(-71.08306194165799, 42.3237643132037)**, which is the value with the biggest number of crime.

p-value

Null Hypothehis: average earnings of police officers HAS NO effect on the number of crimes.
Alternative Hypothehis: average earnings of police officers HAS an effect on the number of crimes.
Results: 1. Therefore, we cannot conclude that a significant difference exists between null and alternative hypotheses.

Future Ideas

In the future we would like to explore more datasets of major cities in the states such as New York and Chicago to prove or disprove our hypothesis about the correlation between crimes and all other datasets that we mentioned before.