

Project 2 Proposal

Graffiti Analysis in Saint Paul, Minnesota

Team Member: Mia Yin, Jude Wentian Zhu

Github repository: https://github.com/UC-Berkeley-I-School/Project2_Yin_Zhu.git

Dataset: Parks Graffiti Report - Dataset

This dataset contains graffiti related issues responded to by the Parks and Recreation Department for the year 2015. It includes work done in the City's parks and non-park locations. It also includes cleanup costs.

Source: <https://information.stpaul.gov/City-Infrastructure/Parks-Graffiti-Report-Dataset/gcu2-spkd>

Supplemental datasets: Crime Incident Report - Dataset

The report contains incidents from Aug 14 2014 through the most recent available, as released by the Saint Paul Police Department, for the following categories: Homicide, Rape, Robbery, Agg. Assault, Burglary, Theft, Auto Theft, Arson, Domestic Assaults, Vandalism, Narcotics, and Firearm Discharges.

Source: <https://information.stpaul.gov/Public-Safety/Crime-Incident-Report-Dataset/gppb-g9cg>

Background:

Graffiti, which is often seen as sign of disorder in communities, can open the door to individuals breaking other social norms and are associated with increased crime rates in many cities across the US. In addition to being a sign of destruction, graffiti vandalism in national parks destroys the natural beauty and the cleanup work becomes a financial burden to municipality.

Project Plan

In this project, we would like to conduct a cost analysis using the dataset of graffiti related issues responded to by the Parks and Recreation Department for the year 2015. The goal is to explore the following:

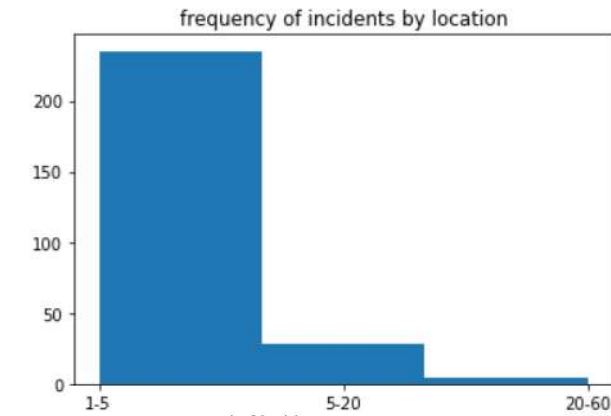
- General trend of graffiti incidents from 2013 to 2015, any seasonal trend
- Where does graffiti issue occur most frequently?
 - City vs. parks
 - Distribution by location on a map
 - Whether there is presence of vandalism clustering
- Is there any visible trend of what surface do the vandals prefer
- Comparison of material vs. labor cost
- Is there any difference in terms of cost of cleanup by graffiti surface
- Do labor cost and material cost correspond to the size of graffiti

Initial data cleaning and manipulation includes but not limited to:

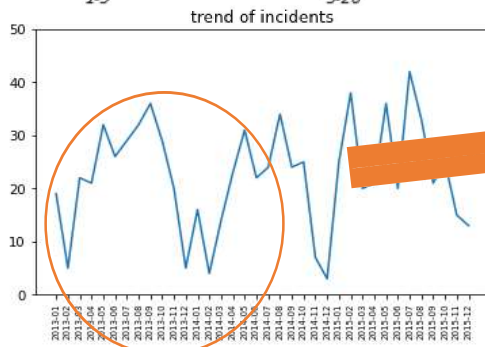
- Remove missing values
- Look for duplication, check for recording errors
- Create a new variable "work_date_y_m" that buckets the time of work by month, to simplify the analysis and display seasonal trends
- Create a new variable "surface" which extracts the keyword of where the graffiti is removed from, such as "wall", "trash can", "sign", from the description column
- Re-arrange the location variable to split latitude and longitude to show

- Clustering of incidents on a map
- Match crime rate distribution in the same area

Some exploratory analysis to help us draw some initial conclusions:

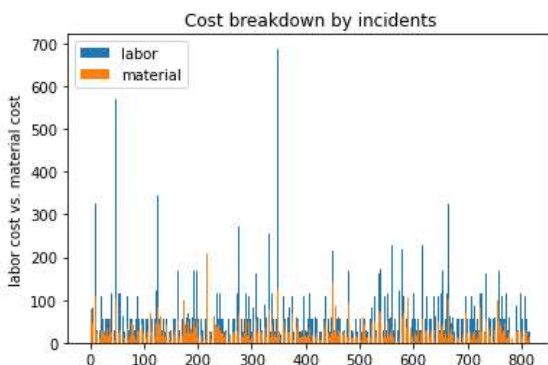


During this period, each location incurred between 1 to 60 cleanups of graffiti. Most location incurred under 5 cleanups, some locations have extreme cases of cleanups over 20.



Some conclusions we can draw from the above trend analysis is 1. Vandalism exhibits seasonality (assuming the cleanup time is shortly after the occurrence of incidents) 2. The peak remains relatively stable, 2015 slightly higher than the other two years and 2014 slightly lower than the other two years.

The seasonality of cleanup cost is in line with the frequency of incidents. However, the cost incurred is highest in 2013, where the total number of cleanup is the median of all three years – this shows the total cost per cleanup might have reduced over the years, whether the labor is lower or material is cheaper.



In many cases, labor cost is significantly higher than material cost and we might want to explore if that's related to the kind of surface.

Final Report:

Summary Findings about Graffiti trend, cost analysis, surface preference, and relationship with crime rates thus the importance of vandalism prevention.