Friday edition
April 19th, 2019

Los Angeles Crimes

Do you want to interact with the data?
https://tabsoft.co/2ZgYcGV

Issue No. 315-719
Fall 2019

David Forteguerre
dfortegu@syr.edu

IST719 Information Visualization
Professor Hemsley

School of Information Studies
Syracuse University

Summary

Cities across the world are increasingly launching open data portals to make their data available to the public. Crime data helps us gain insight into a city's organization, population, economic situation, overall safety, and can undoubtely help individuals and official organizations make more informed decisions to ultimately prevent crime from happening and increase everyone's safety. This article seeks to shed light on crime data for the city of Los Angeles, CA.

About the Data

The dataset was retrieved from Los Angeles's open data portal and it is maintained/regularly updated by the LAPD. The raw dataset contained data from 2011-2019 for the greater Los Angeles area and had 1,914,826 rows and 26 columns. In this article, only data for the year of **2015** was kept and analyzed. After extensive cleaning (there were a great deal of inconsistencies, outliers, and missing values), the final dataset contained 214,698 rows & 20 columns.

Audience

This article may be of interest to Los Angeles residents, travelers, official organizations, and anyone who has an interest in crime data, information visualization, or data science for social good. The findings might help you to know where you should (or shouldn't) be booking your next Airbnb/renting your next apartment!

Questions

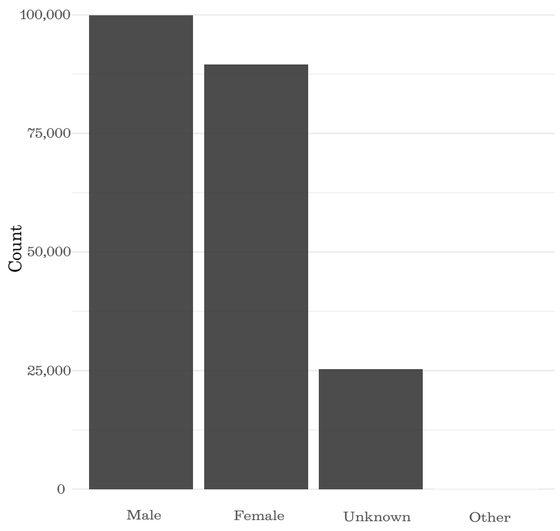
This article will answer the following questions:
(1) Who were the victims?
(2) What were the top 6 crimes, top crime premises, and where did crimes happen?
(3) When did crimes occur?

Demographics

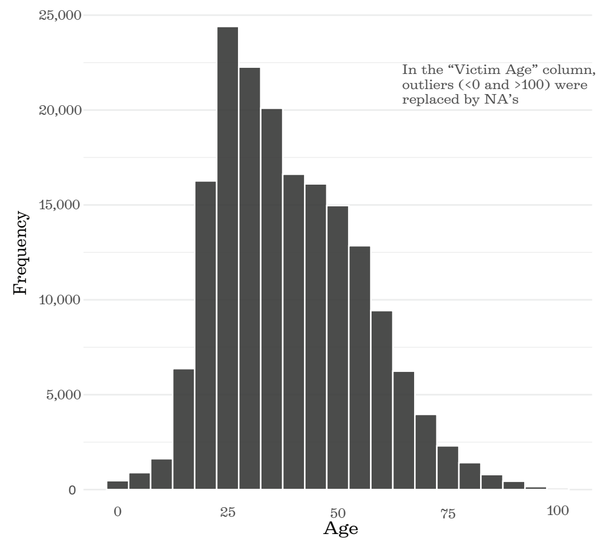
The plots on the right indicate that most of the victims were of Hispanic, Caucasian, and African-American descent. There were more male victims than female victims overall, and the age distribution shows that most victims were in their 20s-50s (female victims being slightly younger than male victims).

Data Source: Crime Data from 2010 to Present (https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-Present/y8tr-7khq)
R Packages used: dyplr, ggplot2, ggmap, treemap, stringr, lubridate, wordcloud

Number of Victims by Gender (2015)

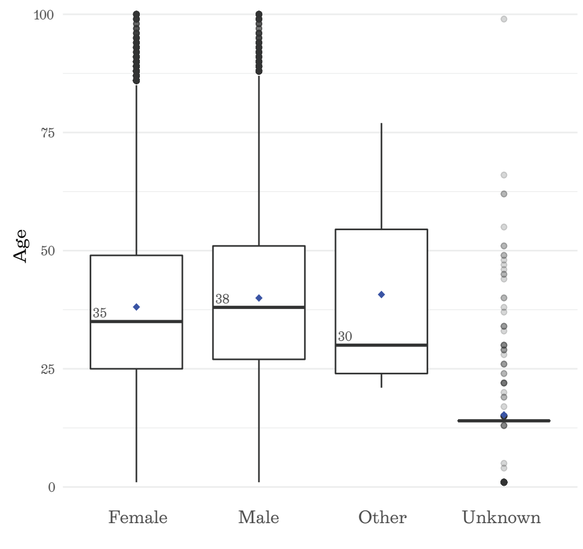


Victim Age Histogram (2015)

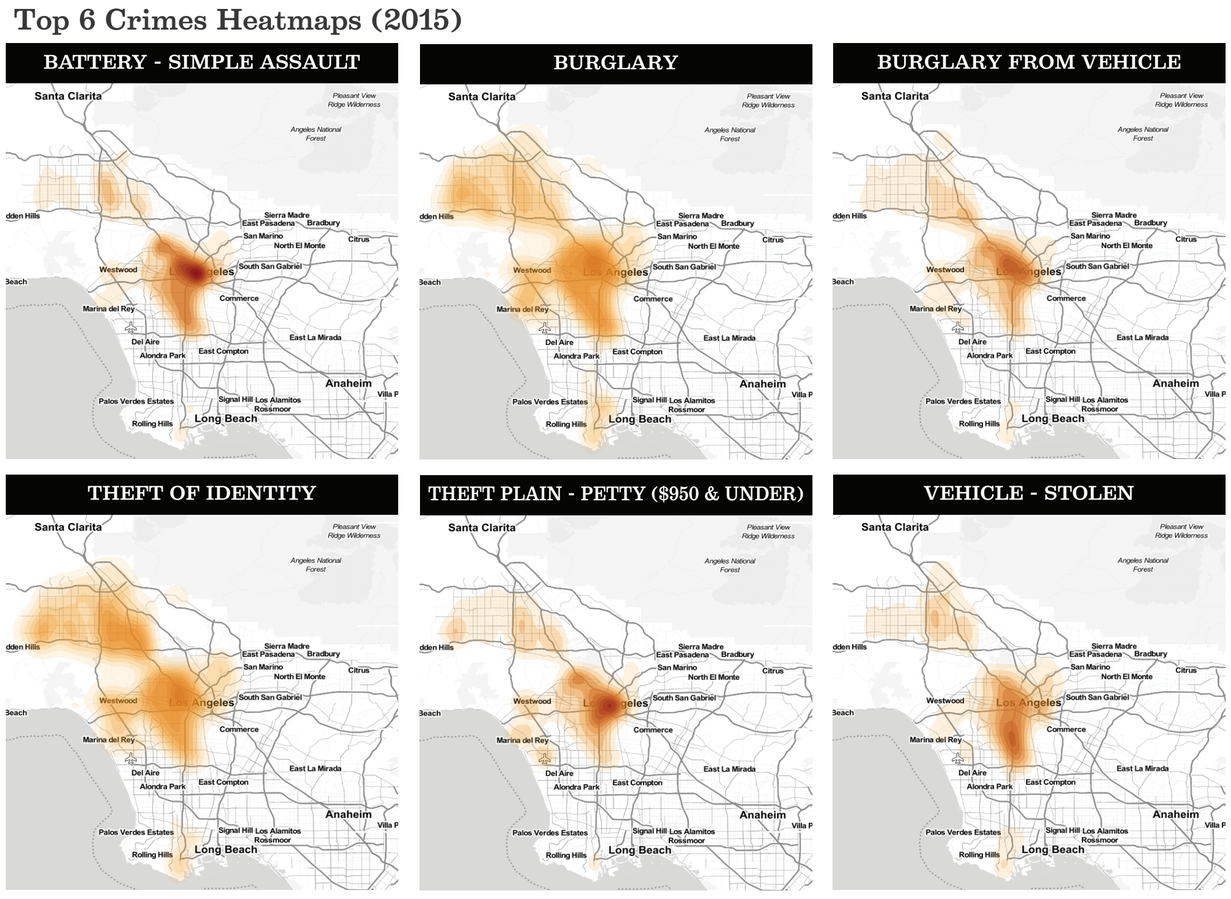


Victim Ethnicity Treemap (2015)

Victim Age Boxplot (2015)



Top 6 Crimes Heatmaps (2015)

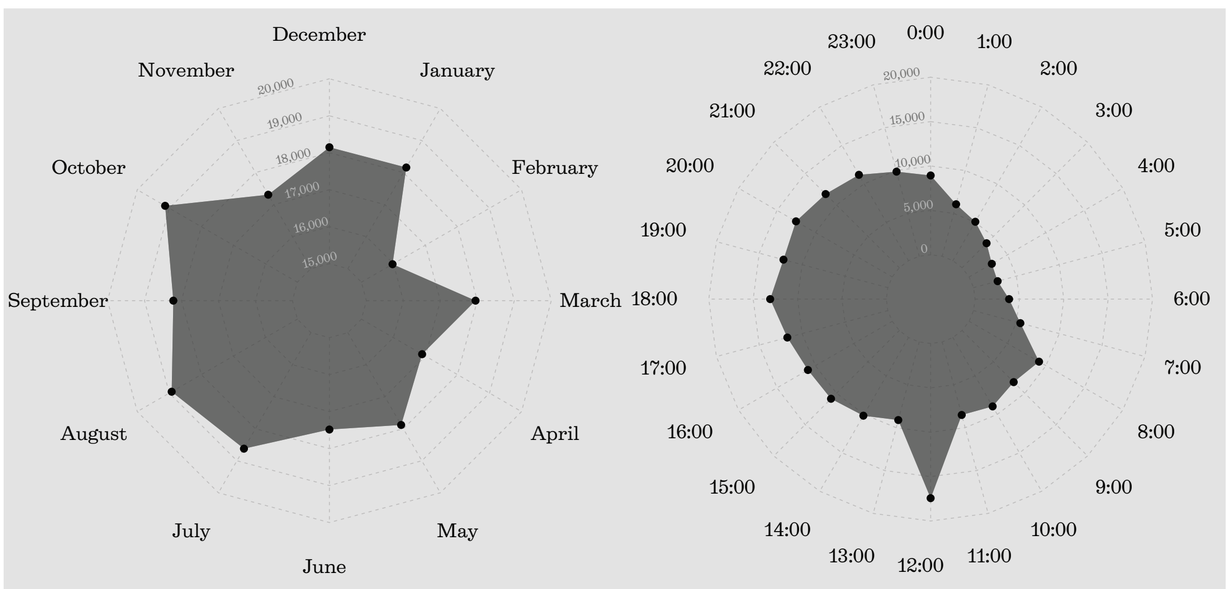


Location

The heatmaps on the left reveal that the top 6 crimes in the data were battery, vehicle (stolen), theft (petty), burglary, theft of identity, and burglary from vehicle. For the most part, crimes were concentrated in the DTLA and Hollywood areas (e.g., battery, theft (petty), burglary from vehicle). Some other crimes were more spread out (e.g., theft of identity, burglary) including the San Fernando Valley. For stolen vehicles, the epicenter was south of DTLA (South LA). The wordcloud on the right indicates that crimes most often occurred on streets, parking lots, sidewalks, and in single-family/multi-unit dwellings, vehicles, and driveways.

Top 30 Premises Wordcloud (2015)

Number of Victims by Month (2015)

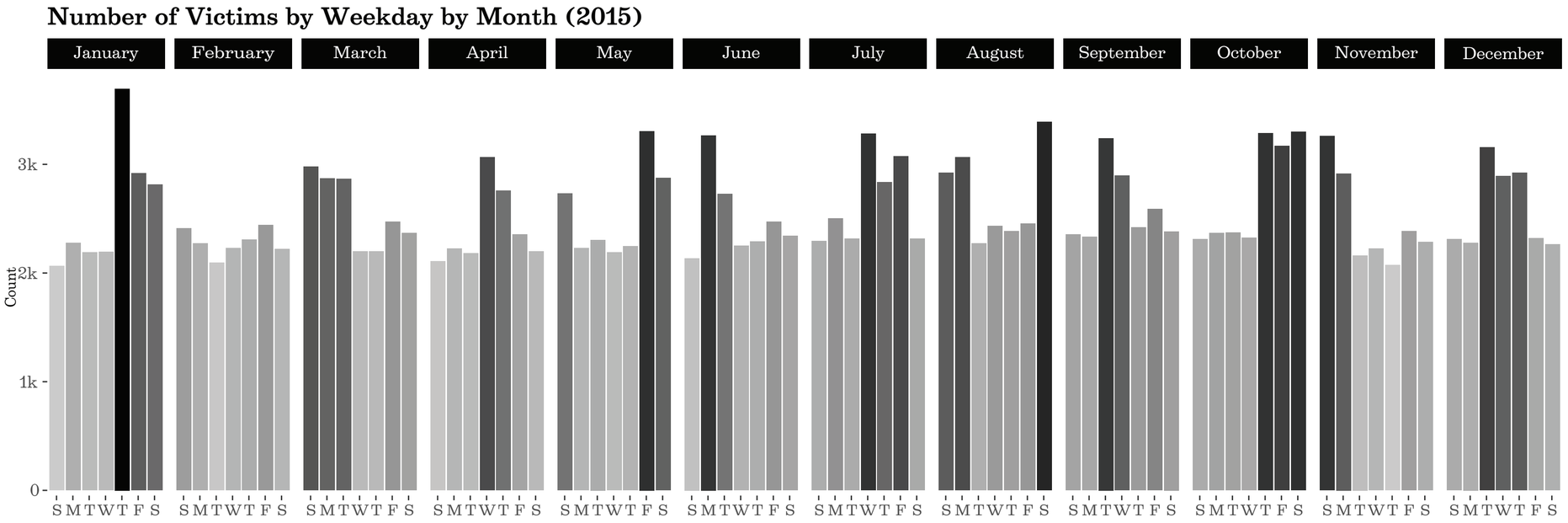


Number of Victims by Hour (2015)

Time

(1) MONTH: The radar chart on the left shows that there were more crimes overall in July, August, September, and October. On the other hand, February, April, and November were calmer months. **(2) HOUR:** The radar chart on the right indicates that there was a crime peak at noon (which may be a data entry matter; perhaps the LAPD used 12:00 pm as a reference point when the time reported was unclear or unknown.) Nonetheless, the plot does show that there were more crimes in the afternoon and evening overall (6:00 pm being the second peak). **(C) WEEKDAY BY MONTH:** The barplot below shows that, in 2015, different days of the week had crime peaks. However, this may be due to randomness as the first day of every month in the data always showed far more crimes (as a result, the weekday with most crimes in a given month would simply correspond to which weekday the first day of that month was.) Even so, filtering out the first day of every month (see online dashboard) still demonstrates that peaks were on different weekdays every month. To conclude, crimes are not necessarily higher on the weekend (Fridays and Saturdays) as many would expect.

Number of Victims by Weekday by Month (2015)



Correlation

A simple linear regression was performed in R between "Victim Age" (independent variable) and "Day Difference" (i.e., the time it took the victim to report the crime) (dependent variable). The R-squared of the model was close to 0, indicating that there is no relationship between the variables.

In 2015, the average number of days it took to report a crime was 16; the median was 1.