Data analysis using R and Quarto

Student 296356

June 22, 2025

Table of contents

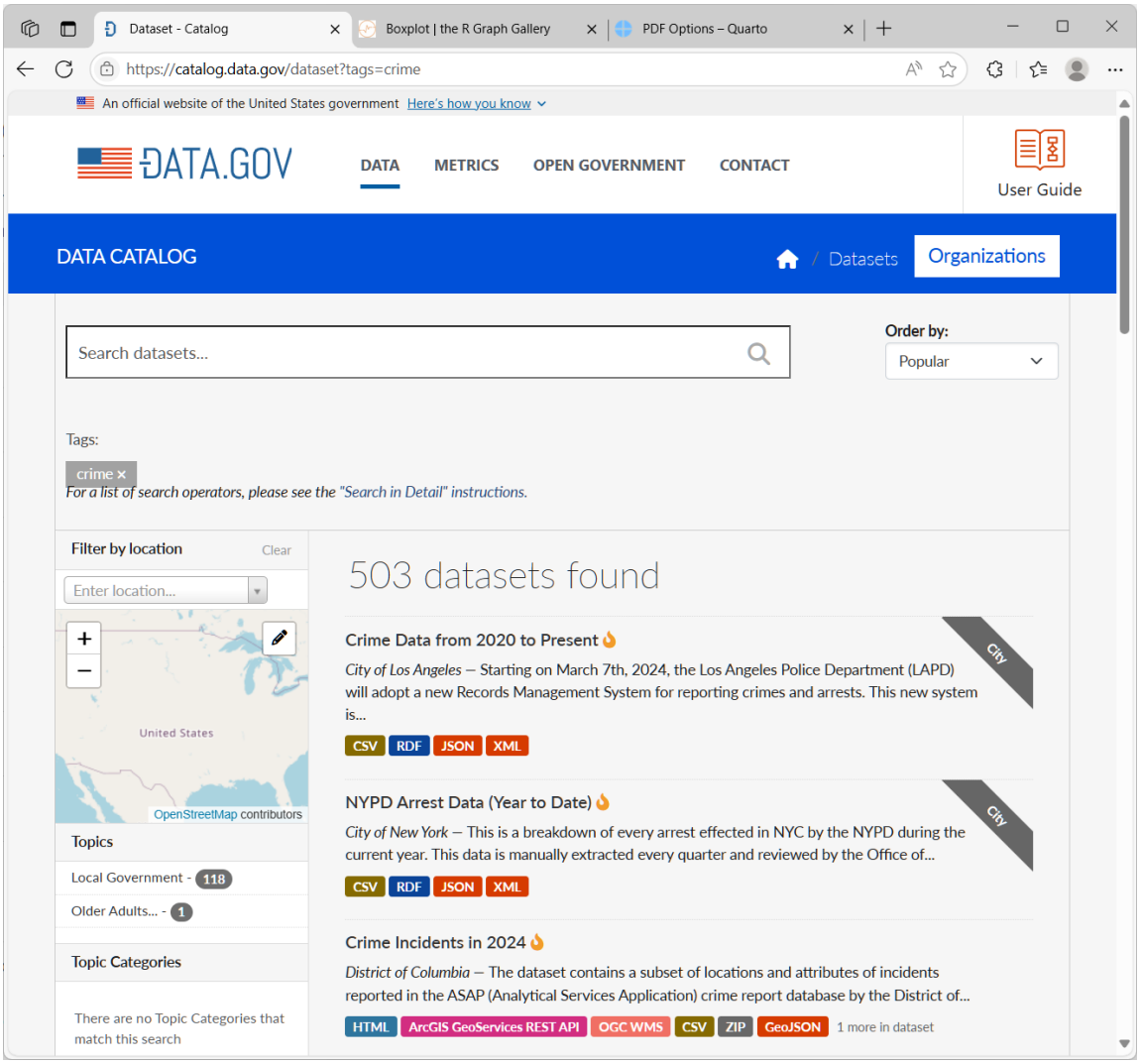
1 Information on the Quarto project

Project title: Data analysis using R and Quarto

Author: Student 296356

Subject: Statistical analysis of Los Angeles crime reports

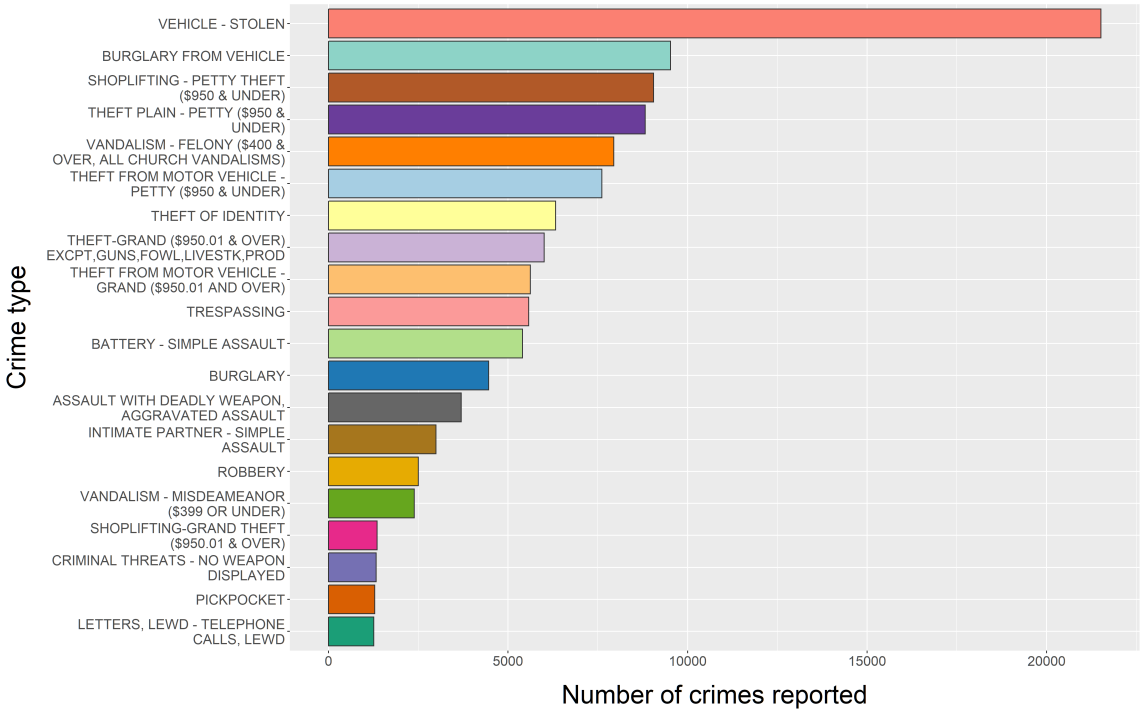
Data source: Los Angeles crime types (based on 2024 data) (https://catalog.data.gov/dataset?tags=crime)



2 Most common crime types

This chart shows an analysis of Los Angeles crime types in 2024 by frequency.

Twenty most common offences are presented.

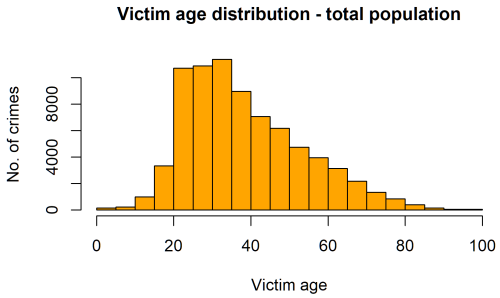


Conclusion:

The most common crime type in the LA area in 2004 was vehicle theft (an outstanding number of over 20,000 cases reported). The remaining types of offences are less variable in frequency, and none of them exceeded 10,000.

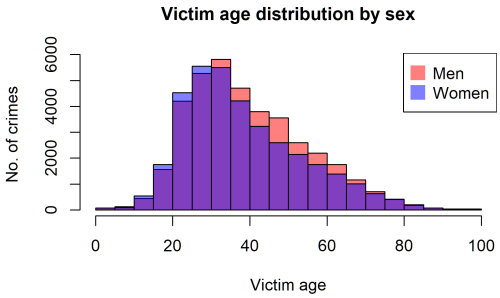
3 Victim age analysis

Victim age structure: entire sample



Conclusion: This chart presents age distribution in *the entire population*. Most victims are adults aged 20–40 years. The data seems to be distributed normally, as confirmed below in a statistical test.

Victim age structure: by age



Conclusion: This chart presents overimposed histograms of the age for *male* and *female* victims. The sex-specific histograms are very similar and nearly overlap, although *male* victims are slightly older.

Only *male* and *female* genders are included (categories *other*/*unknown* were omitted due to low sample size).

Summary stastistics

N: 76839

Min: 2

25th perc.: 27

Mean: 39

Median: 36

SD: 15

75th perc.: 49

Max: 99

Conclusion: The youngest victim was 2 years old, and the oldest 99 years old. Median victim age was 36 years.

Test for normal distribution:

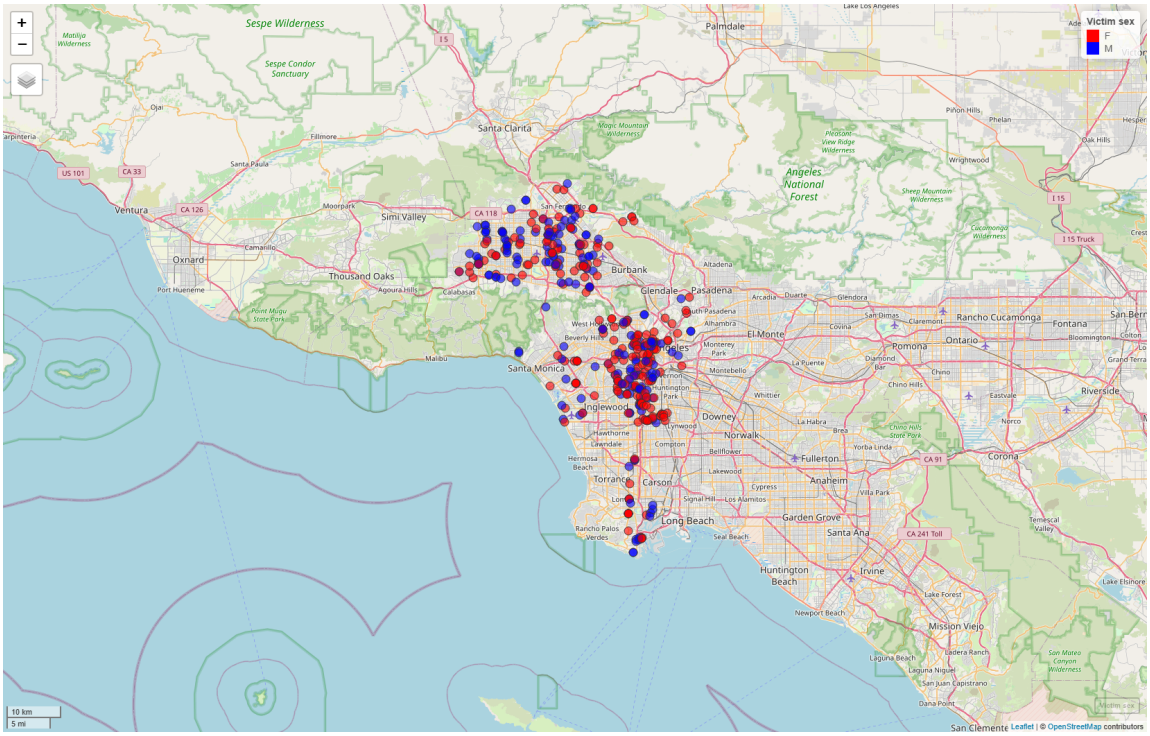
```
    Shapiro-Wilk normality test

data:  head(data_prep$Vict.Age, n = 5000)
W = 0.95593, p-value < 2.2e-16
```

Conclusion: The p value in Shapiro-Wilk test is very low, confirming normal statistical distribution of the data.

4 Geographical distribution of crime

Certain offences may exhibit a heterogeneous geographical distribution. To illustrate this, information on the geographical location of assaults on minors was extracted from the dataset, thereby yielding a smaller subset of data that can be clearly displayed on a map. Colors indicate victim sex.

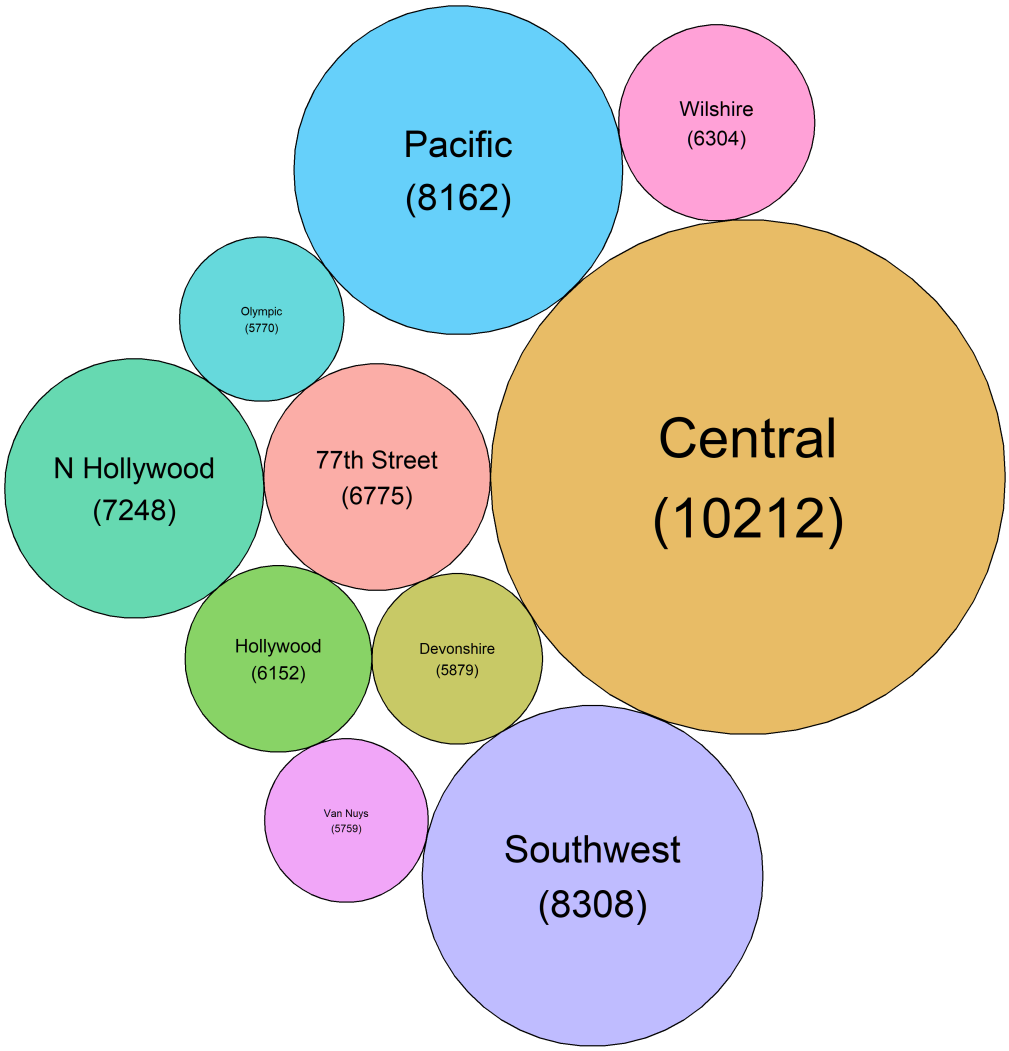


Conclusion:

The map clearly shows two large and two smaller clusters corresponding to areas where the majority of these offences (assaults on minors) occur, indicating where preventive measures in this regard should be intensified.
No clear pattern is evident regarding the victims' sex.

5 Areas most affected by crime

This chart shows 10 areas with the highest crime rate. The diameters of the circles are proportional to the number of reported crimes in the respective areas. Exponential transformation was applied to better highlight the differences.

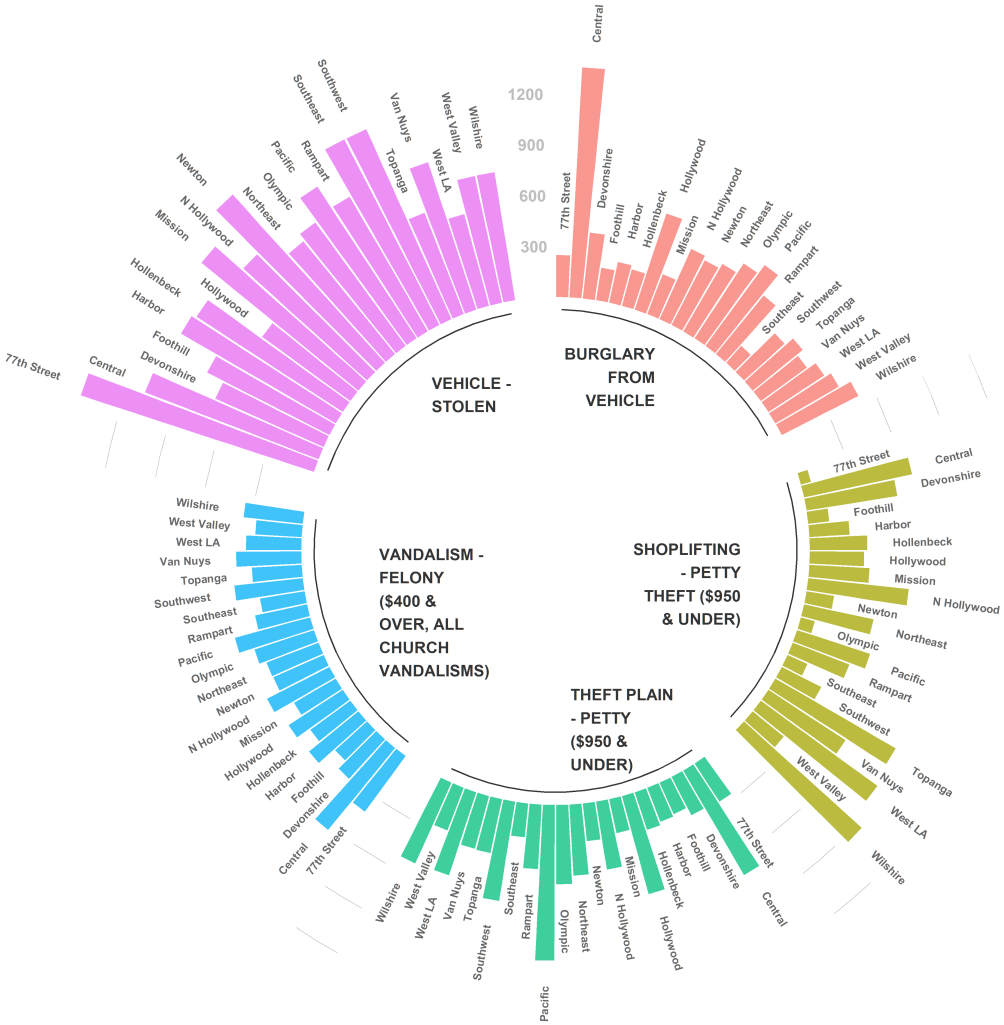


Conclusion:

Most reports are from the Central area, followed by Southwest and Pacific.

6 Crime distribution by area

The chart below shows crime distribution in the LA region by area. Five most common offences are presented.

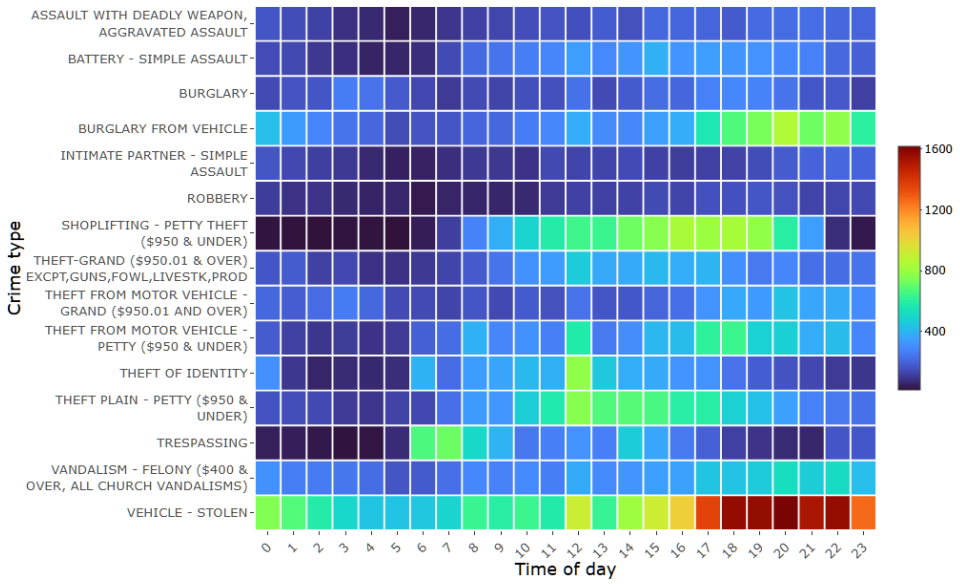


Conclusion:

For certain crimes there are marked differences in the number of reports depending on area. For instance, "BURGLARY FROM VEHICLE" occurred overwhelmingly in the "Central" area (with more than twice as many reports as in any other area). Other offences show more homogeneous distribution (e.g. "VANDALISM").

7 Crime types by time of day

The chart below shows the probability of specific crimes depending on time of day. Fifteen most common offences are presented.



Conclusions:

- Vehicle-related crimes (stealing or burglary from vehicle) are much more common in the evening hours, starting around 5:00 PM.
- Most cases of theft usually occur in the middle of the day.
- Trespassing is clearly limited to early morning hours.
- Other offences, such as burglary or robbery, are equally distributed throughout the day.