Chicago Crime Data Analysis

Student ID:/ 201081646

Introduction

This is the first assessment for the **MATH5741M Statistical Theory and Methods** module. Its objective is to summarise statistically a crime dataset sample from the city of Chicago and answer the following research questions:

- What time of day do most types of crime occur?
- In which locations are specific types of crime more likely to happen?
- Which districts are potentially more dangerous per type of crime?

Data and methods

The dataset analysed is a sample of the original data of crimes extracted from the Chicago Police Department that occurred in the city of Chicago from 2001 to present.

For the analysis, first, we prepare the data creating, transforming and simplifying variables, as well as cleaning the dataset keeping the variables we are interested in. Secondly, we make a general univariable analysis of the data set, and then, with a multianalysys based on heatmaps we answer our questions. Finally, we sumarise the findings.

The report describes not only the statistical process followed but also incorporates the most important R code used to carry it out. Unfortunately, space constraints did not allow to include all the code use, however it is available in https://github.com/eugenividal/Chicago-Crime-Data-Analysis.

Results

Data preparation

First, we load the data into the `R` environment.

```r
# read csv in R
dd=read.csv("http://www1.maths.leeds.ac.uk/~charles/math5741/crime.csv",header=T)
```

Second, we create new variables (`count`, `hour`) based on the existing ones, and give them the right format for later explotation.

Third, we simplify the variables `Primary.Type` and `Location.Description` grouping their categories and call them `Type_grouped` and `Location_grouped` respectivelly.

Next, we keep only those variables which will help us to answer our questions. So, we drop all the variable we do not need.

```r
# drop all variables we are not interested in
dd <- dd[, -c(1:8, 10:11, 13:18)]
```

Then, we clean the dataset of missing values.

```
# remove NAs
dd <- dd[complete.cases(dd),]
```

Finally, the data is ready for the explotation.

```
head(dd)
```

```
##   Arrest District count hour   Month_Yr Type_grouped Location_grouped
## 1   true       19     1   00 2013-07-01        Batery           Street
## 2   true       19     1   01 2013-07-01        Others           Street
## 3  false        2     1   21 2013-07-01       Assault        Apartment
## 4   true        9     1   02 2013-07-01     Narcotics           Street
## 5  false        3     1   17 2013-07-01         Theft           Street
## 6   true        9     1   01 2013-07-01        Batery        Apartment
```

Data exploration

General analysis

Crime evolution

The number of crimes in Chicago has decrease dramatically per year from 2001 to present.

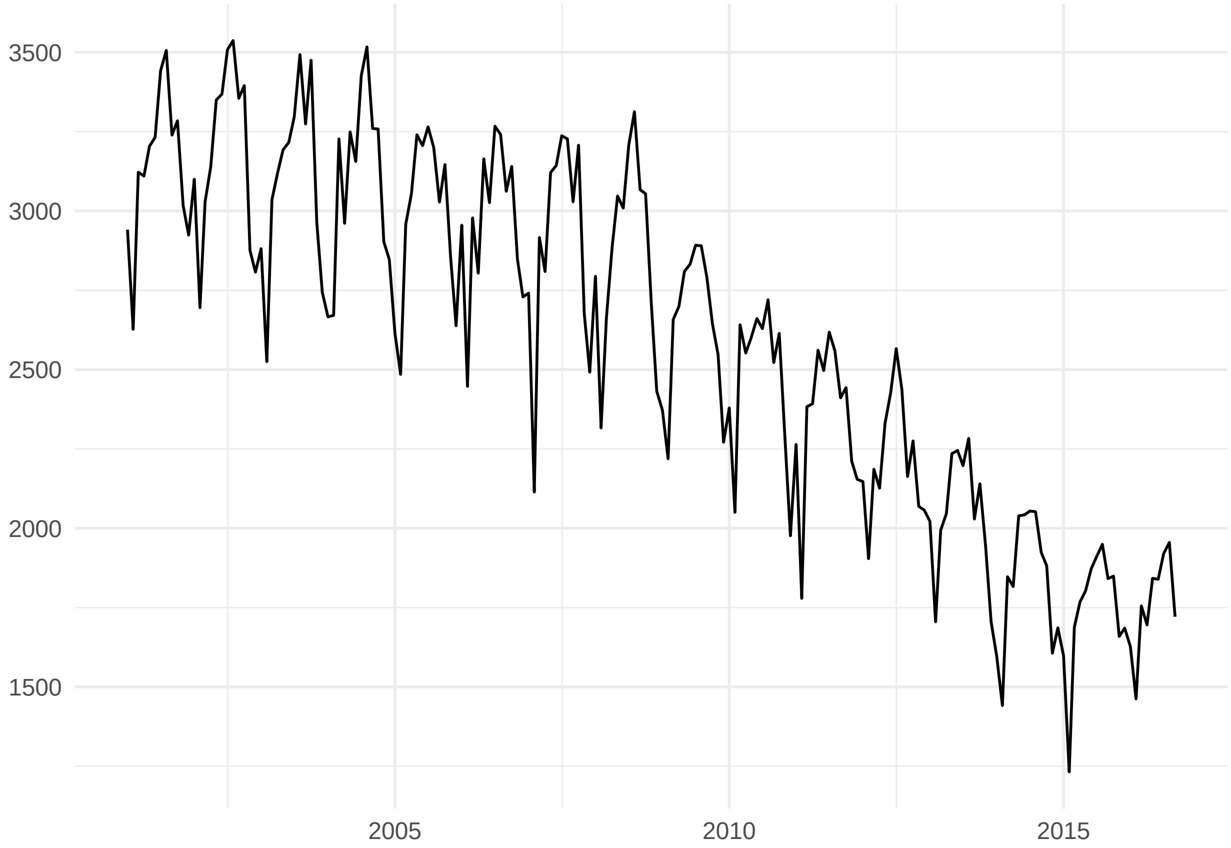

Figure 1: Crimes evolution

Except for the deceptive practice, all the crimes have decresead in more or less grade.

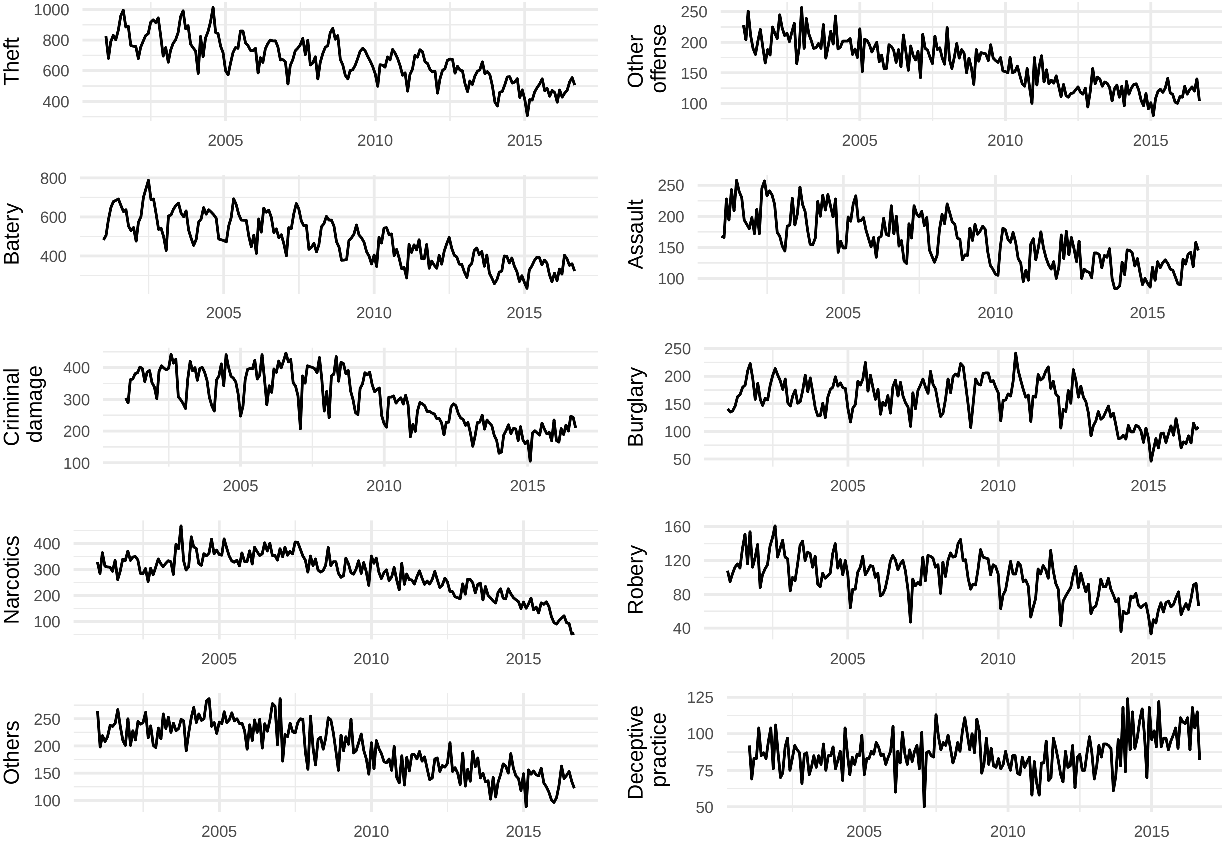

Figure 2: Evolution per type of crime

Crime per hour

The crimes are concentrated in hours

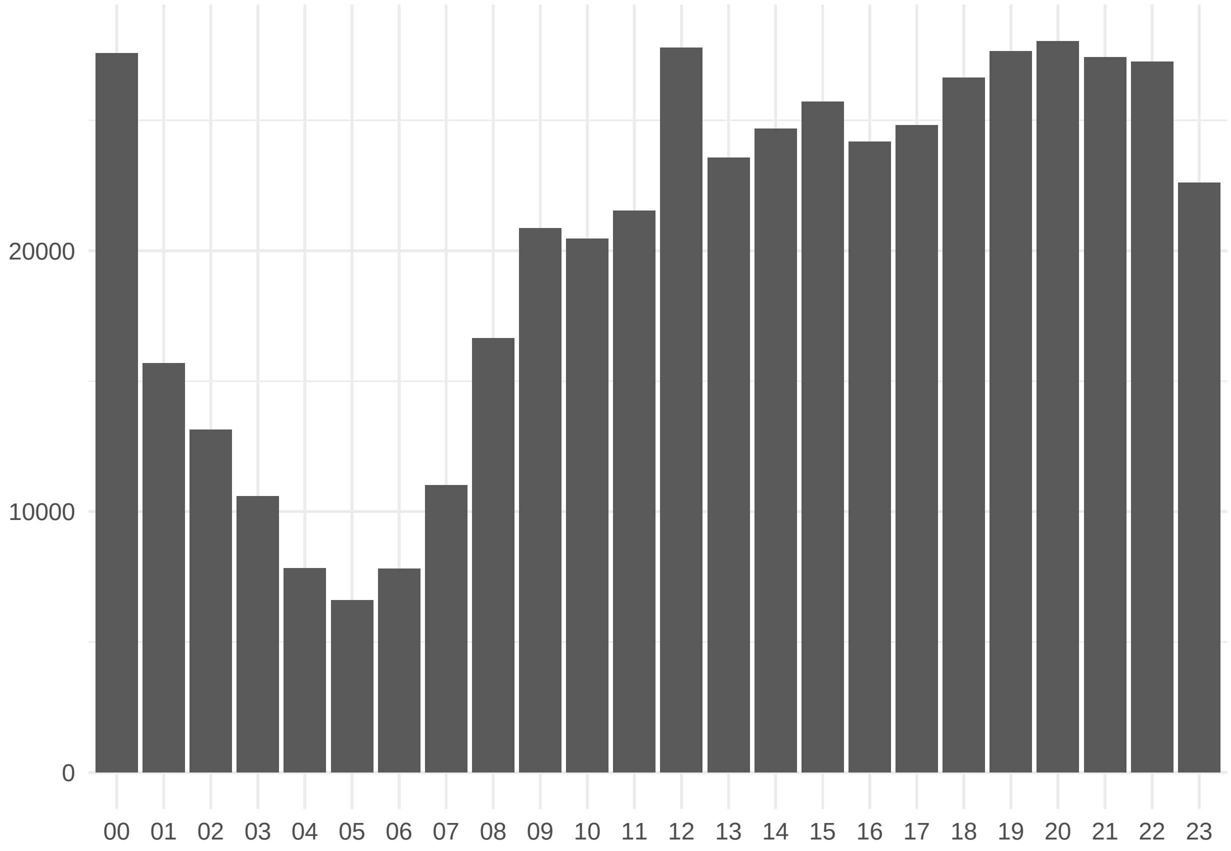

Figure 3: Crimes per hour

Type of crimes

Per type of crime Theft is in difference the biggest number. Change the scientifyc number.

Location of crimes

These crimes are concentrated in Streets, give percentage.

Crime per districts

Per districts the most dangerous are 8.

Answers to our questions

The multiple analysis focuses on type of crime crossed with hour, location and district.

What time of day do most types of crime occur?

Are some types of crimes more likely to happen in specific time of the day?

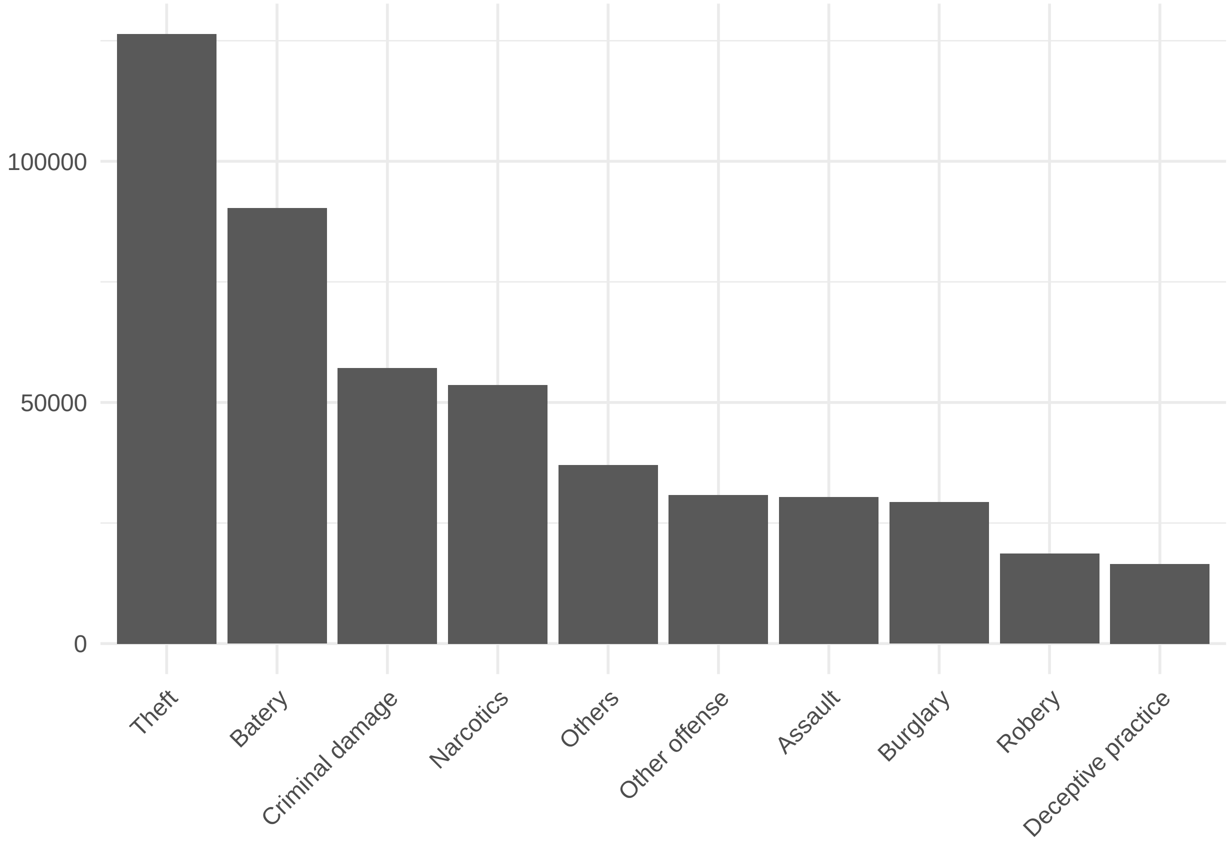

Figure 4: Crimes per type

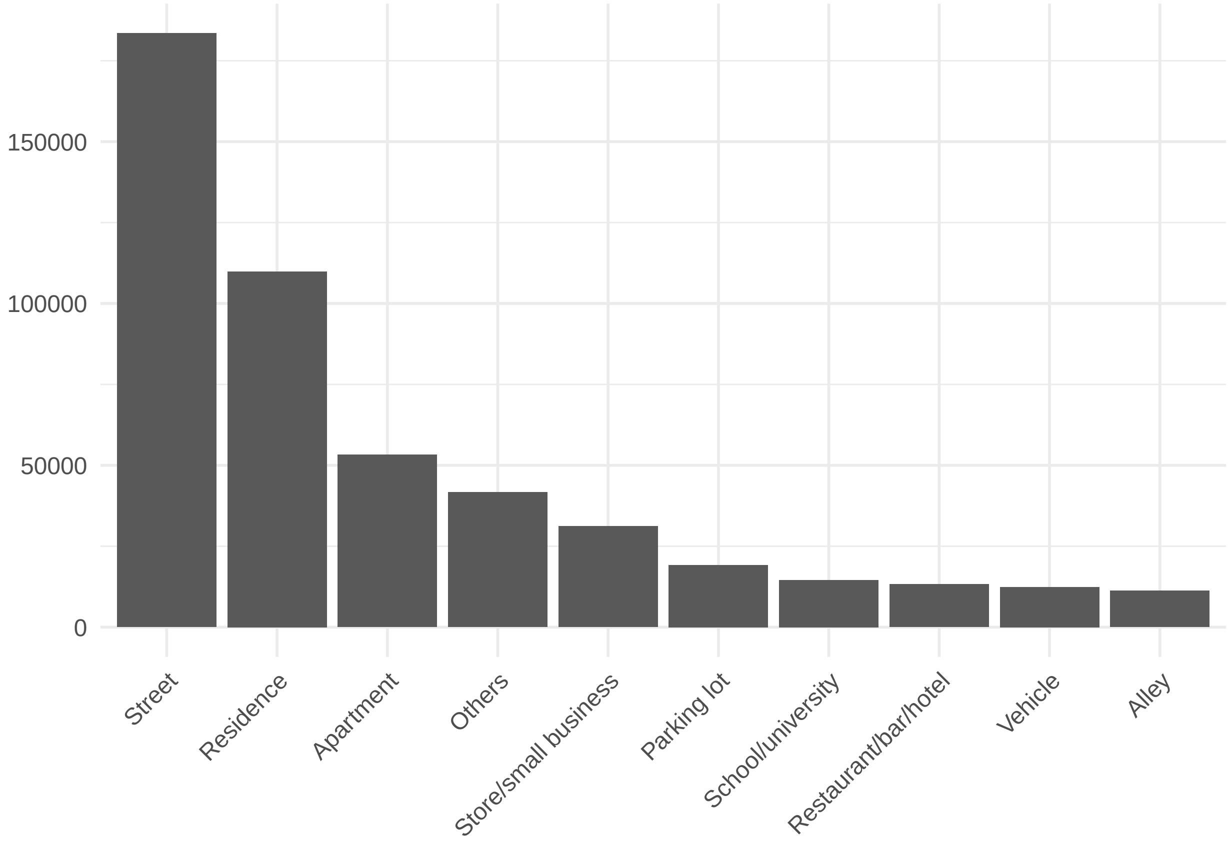

Figure 5: Crimes per location

The most dangerous hours per Thefth are 00 and 12.

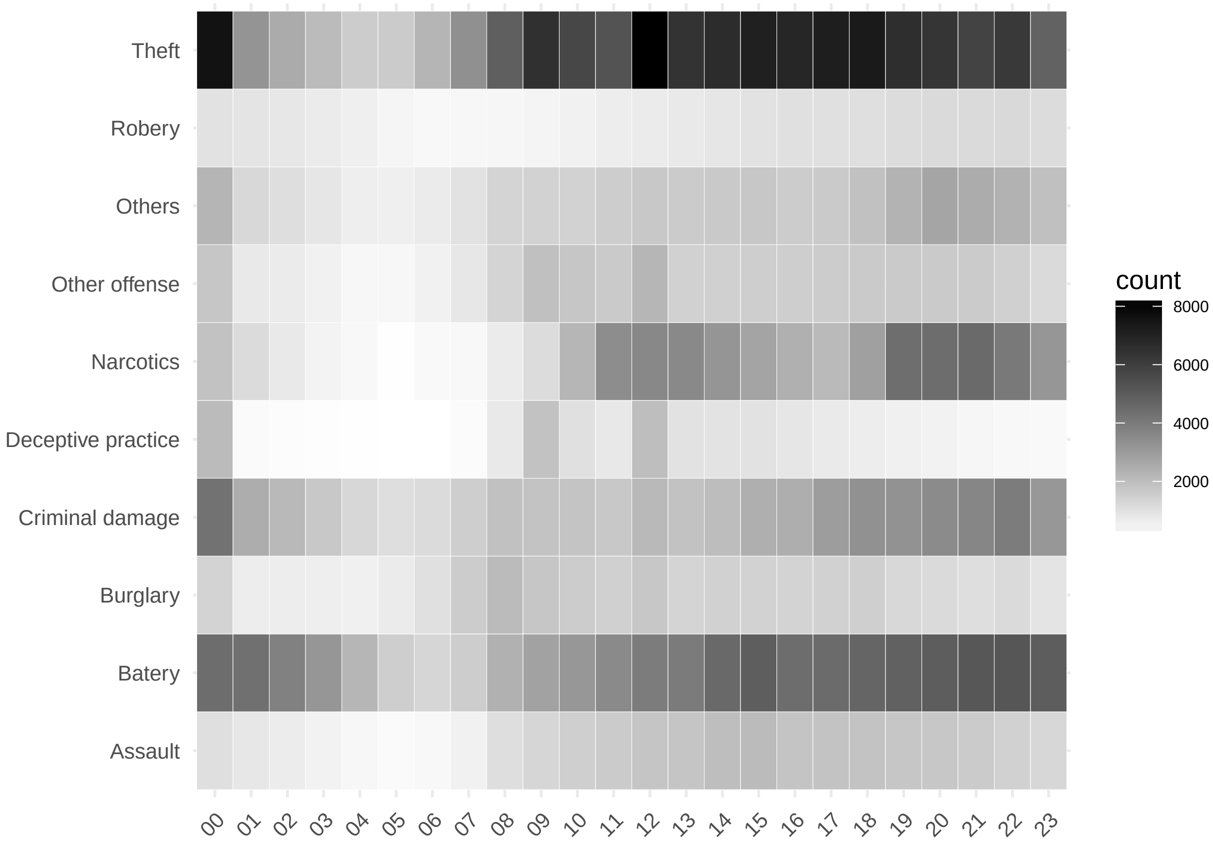

Figure 6: Type of crime vs hour

In which locations are specific types of crime more likely to happen?

Are some types of crimes more likely to happen in specific locations?

Street is particularly important for Theft.

Which districts are potentially more dangerous per type of crime?

Are some types of crimes more likely to happen in specific distrits?

Narcotics in district 11 is crealy a problem.

Conclusions

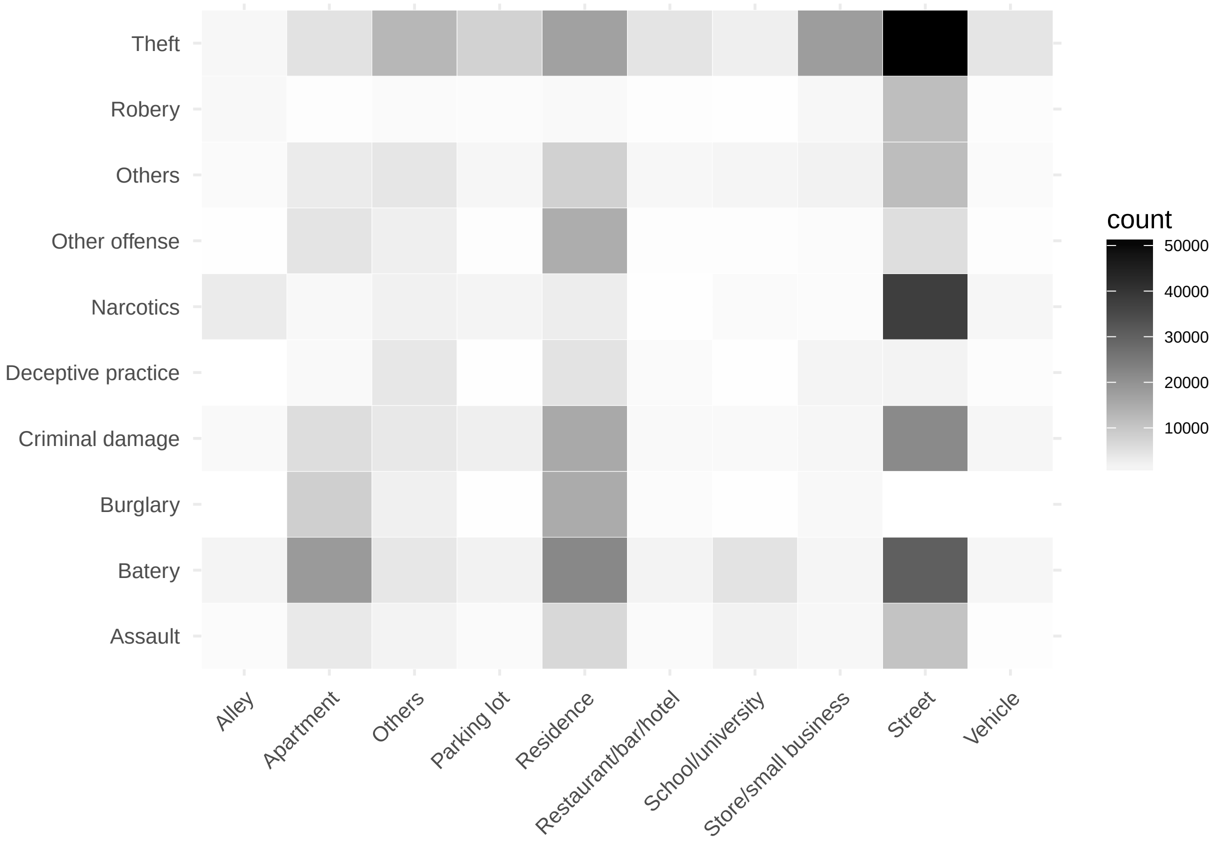

Figure 7: Type of crime vs location

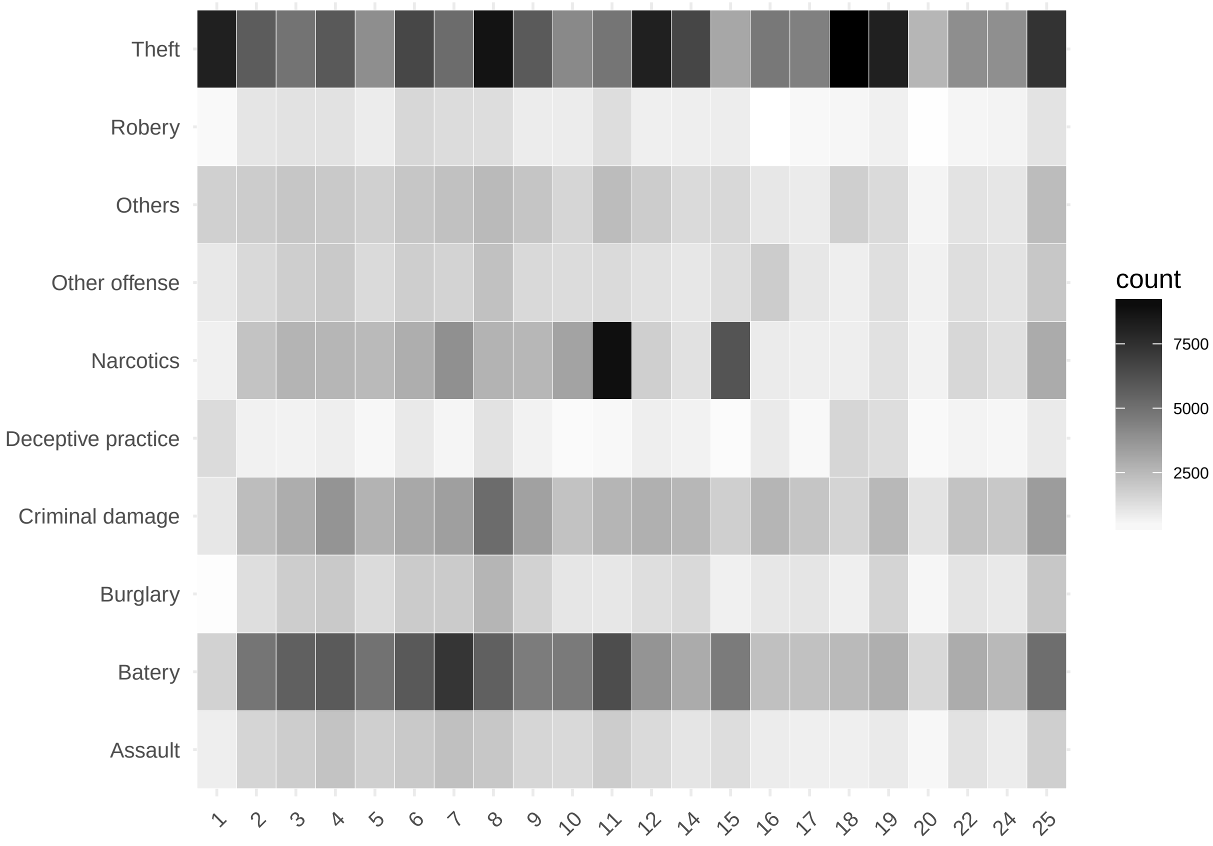

Figure 8: Type of crime vs district