

# Crime in Chicago: Data Analysis and Visualizations

## using R

201081646

## Introduction

This is the first assessment for the **Statistical Theory and Methods module**.

Its objective is to:

- Summarise a sample of dataset.
- Highlight key findings.

## Data and methods

The dataset we use is sample of 500,000 rows of the original data which come from <https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2>.

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

First we prepare the data, then we explore it through a univariable analysis and a multianalysys based on heat maps. in Counclussions we sumarise the main findings.

This report has been done explaining the most essential code. However, due to space limitarion couldn't add all the code which can be analysed in the following github link.

## Results

### Data preparation

First, we have a look at the variables we got.

```
names(dd)
```

##	[1]	"X"	"ID"	"Date"
##	[4]	"Block"	"IUCR"	"Primary.Type"
##	[7]	"Description"	"Location.Description"	"Arrest"
##	[10]	"Domestic"	"Beat"	"District"
##	[13]	"Ward"	"Community.Area"	"FBI.Code"
##	[16]	"Year"	"Latitude"	"Longitude"

Due to time limitation and lenght constraint. We will analyse 8 of them: `Date`, `Primary.Type`, `Location.Description`, `Arrest`, `Domestic`, `District`.

We drop the rest of them.

Secondly, we clean the dataser of missing values - drop the NAs.

Third, we create new variables: `count`, `hour`, `Month_Yr`, `Month`, `weekday`.

And we give them the right format for later exploitation.

Next, we group labels of variables 'Type\_grouped' and 'Location.Description' of accidents in bigger categories.

We do the same with Location.Description.

Something is wrong here because there are NAs later.

Finally, the data is ready for exploitation.

```
head(dd[dd$VAR1==4,],6)
```

```
## [1] Date           Primary.Type      Location.Description
## [4] Arrest          Domestic         District
## [7] count           hour            Month_Yr
## [10] mon             weekday          Type_grouped
## [13] Location_grouped
## <0 rows> (or 0-length row.names)
```

## Data exploration

This section explores the data visually to identify the key patterns.

### Univariable analysis

#### Evolution crime

- Check how to describe graphs in English.

The number of crimes in Chicago has decrease dramatically per year from 200x un til 2015.

In more or less grade all the crimes have decresead, except for the grey ones.

#### Crime per Hour

The crimes are concentrated in hours

#### Crime per weekday

Friday concentrated most of the crimes, percentage?

#### Crime per month

Summer is in difference the period with more crimes recorded.

#### Type of crimes

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

#### Location of crimes

These crimes are concentrated in Streets, give percentage.

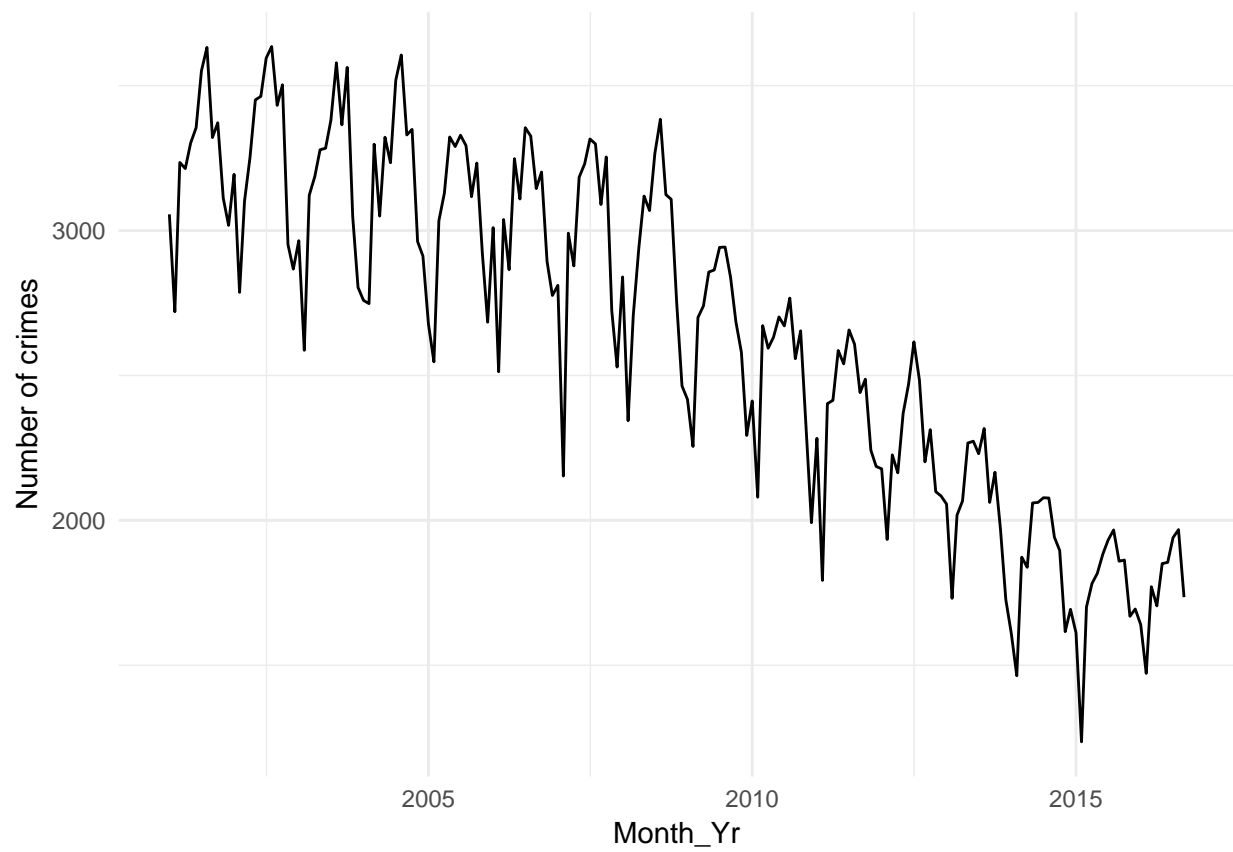


Figure 1: Figure 1. Vehicles, casualties and secs pairwise

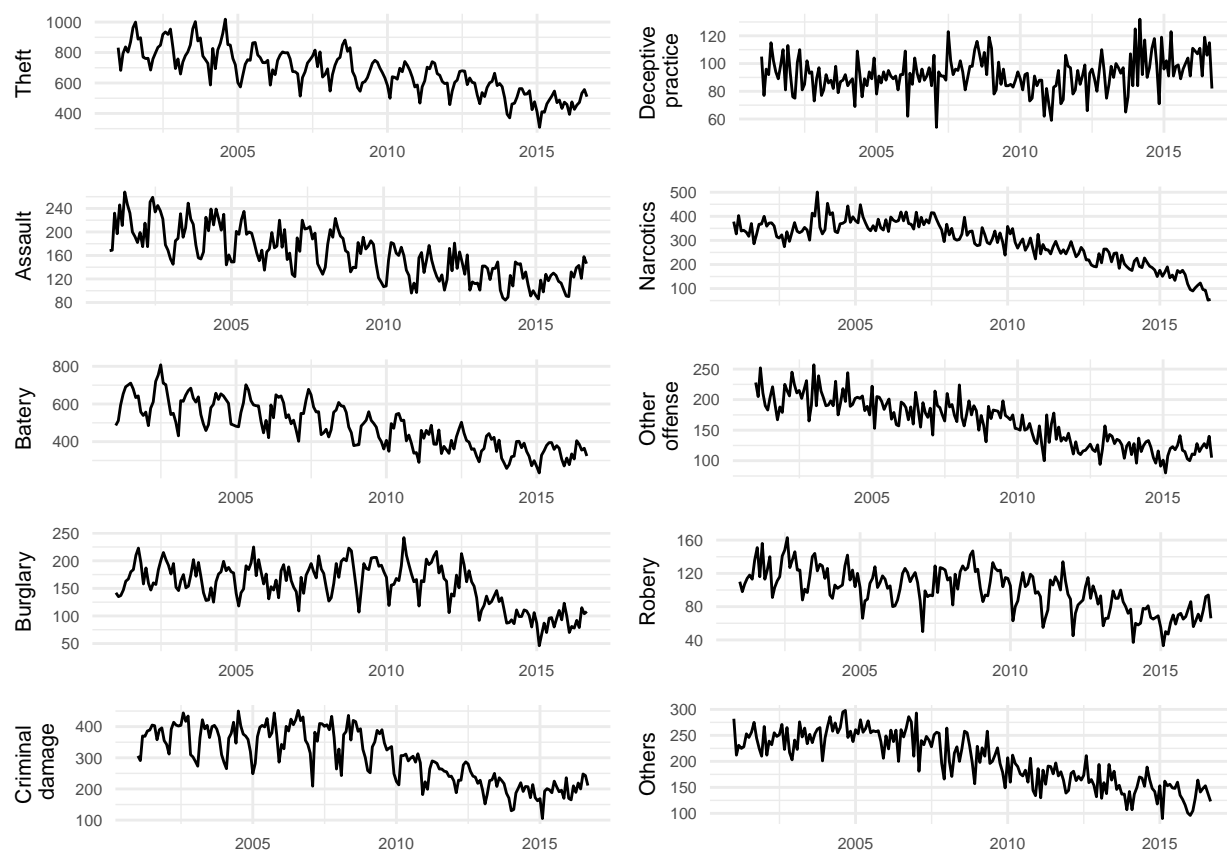


Figure 2: Figure 2. Vehicles, casualties and secs pairwise

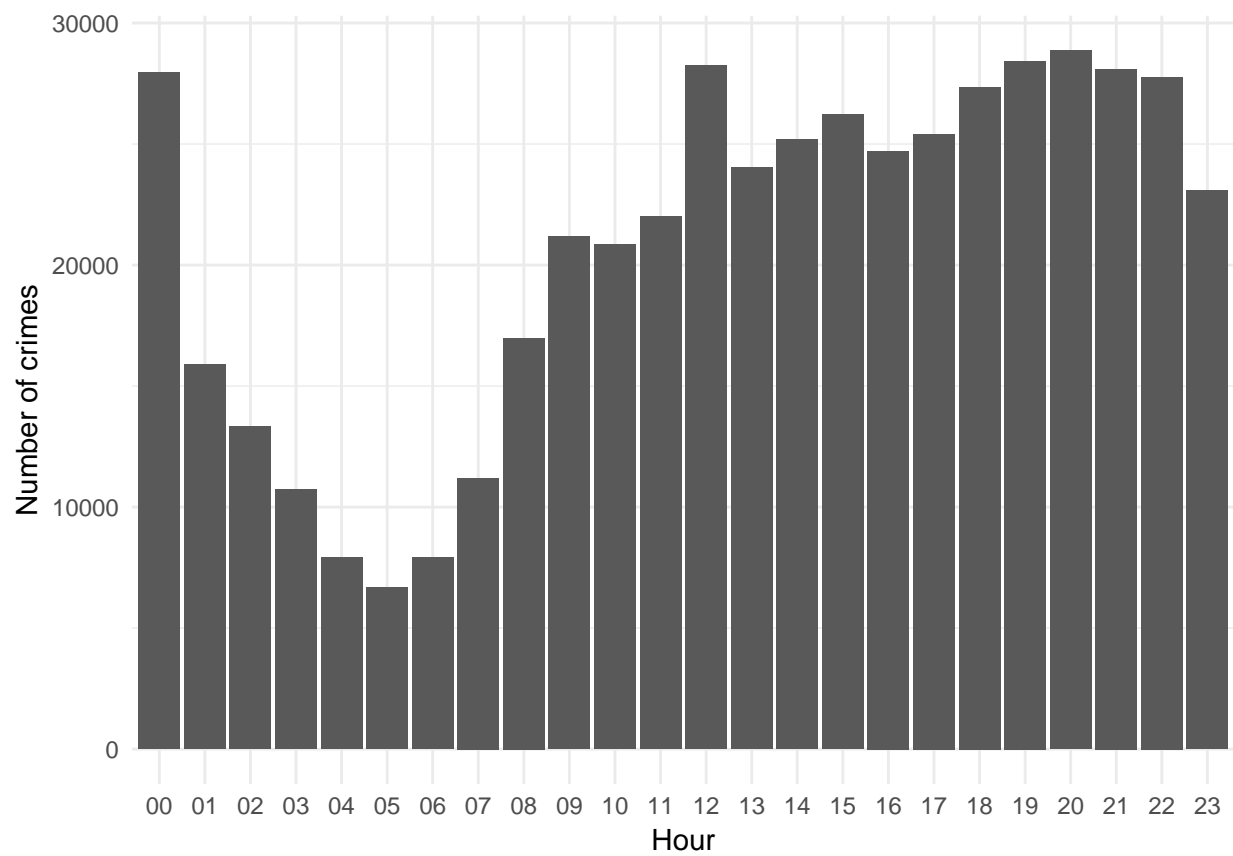


Figure 3: Figure 3. Crimes per hour

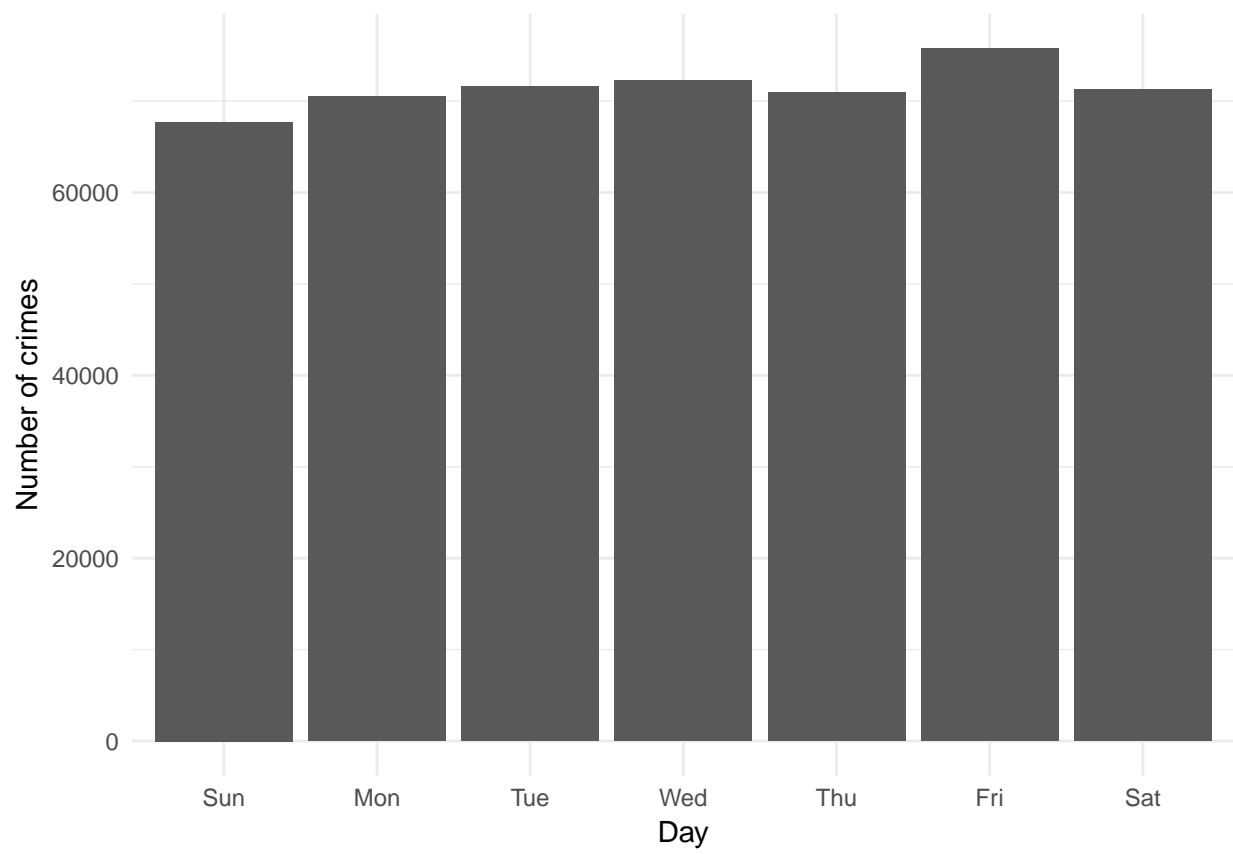


Figure 4: Figure 4. Crimes per week day

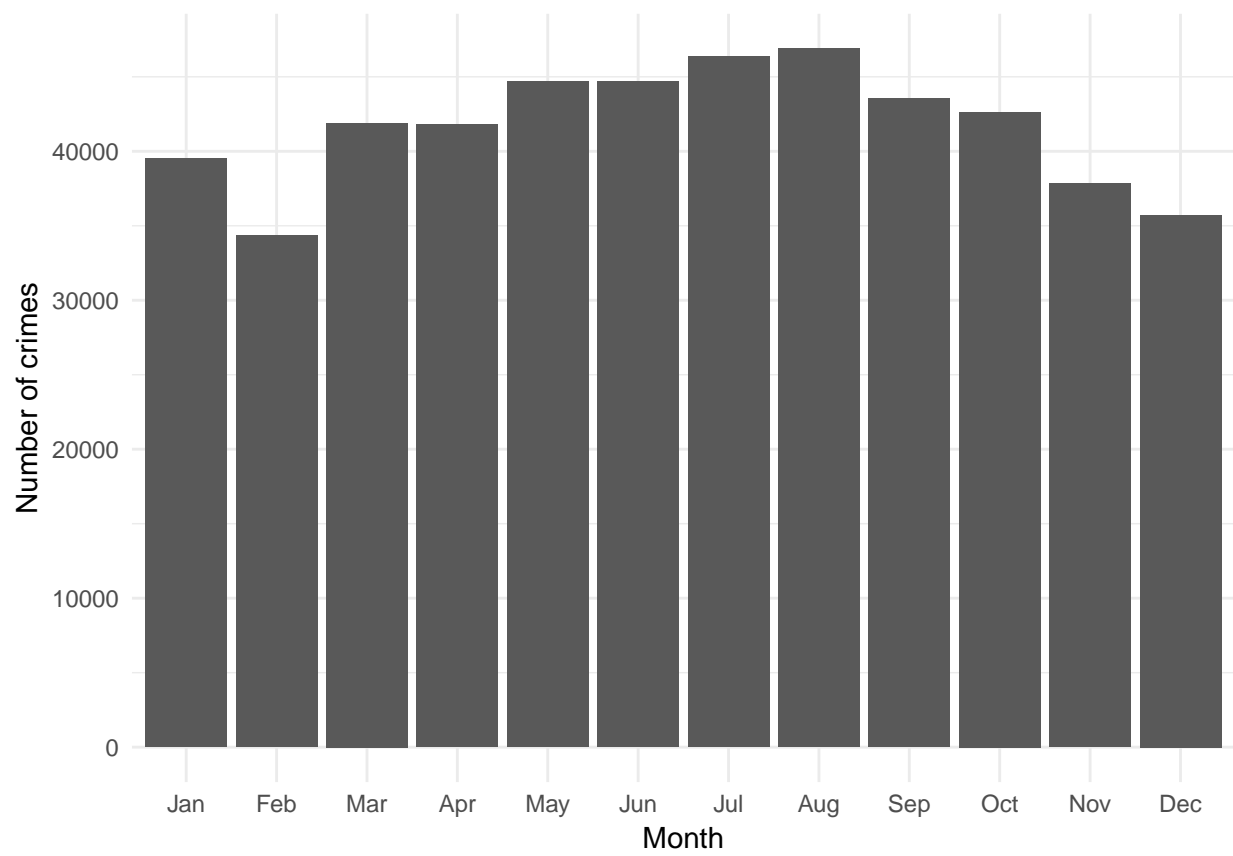


Figure 5: Figure 5. Crimes per month

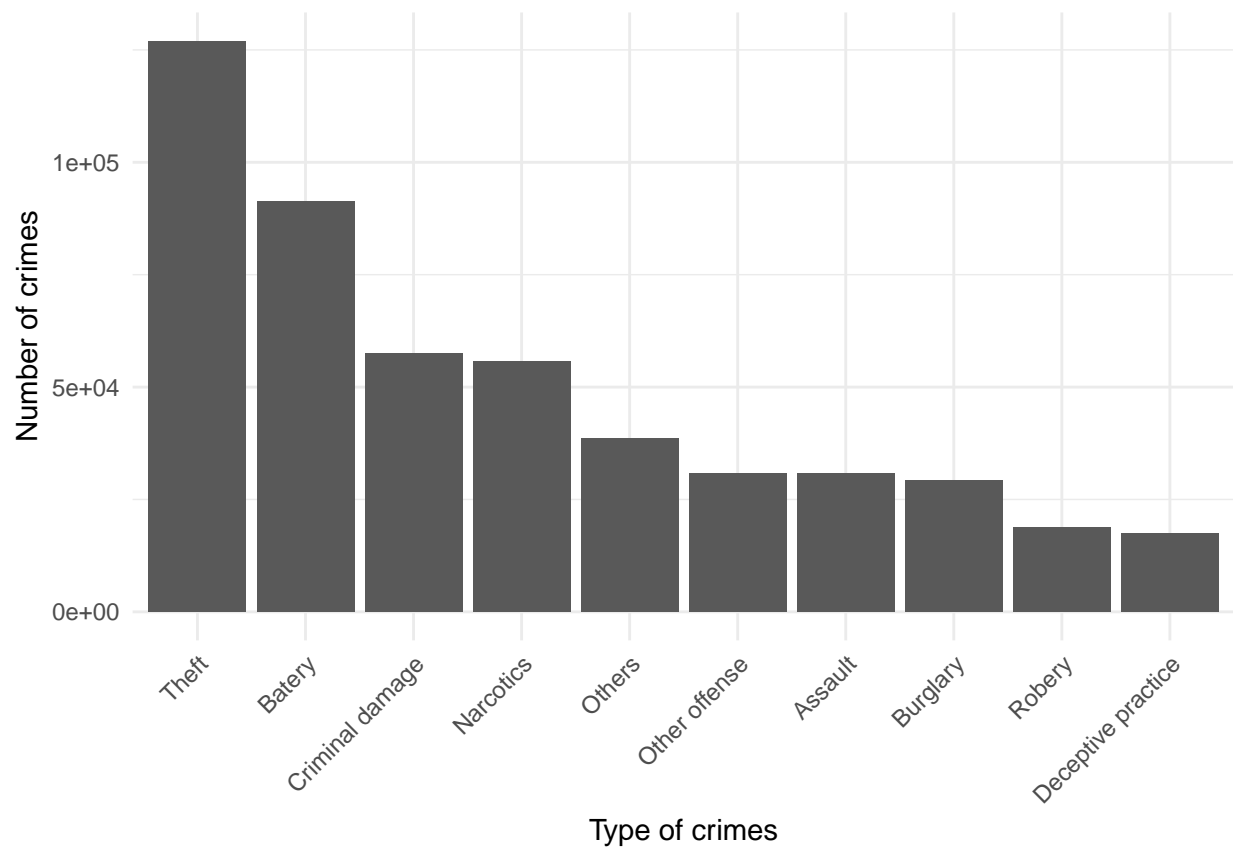


Figure 6: Figure 6. Crimes per type



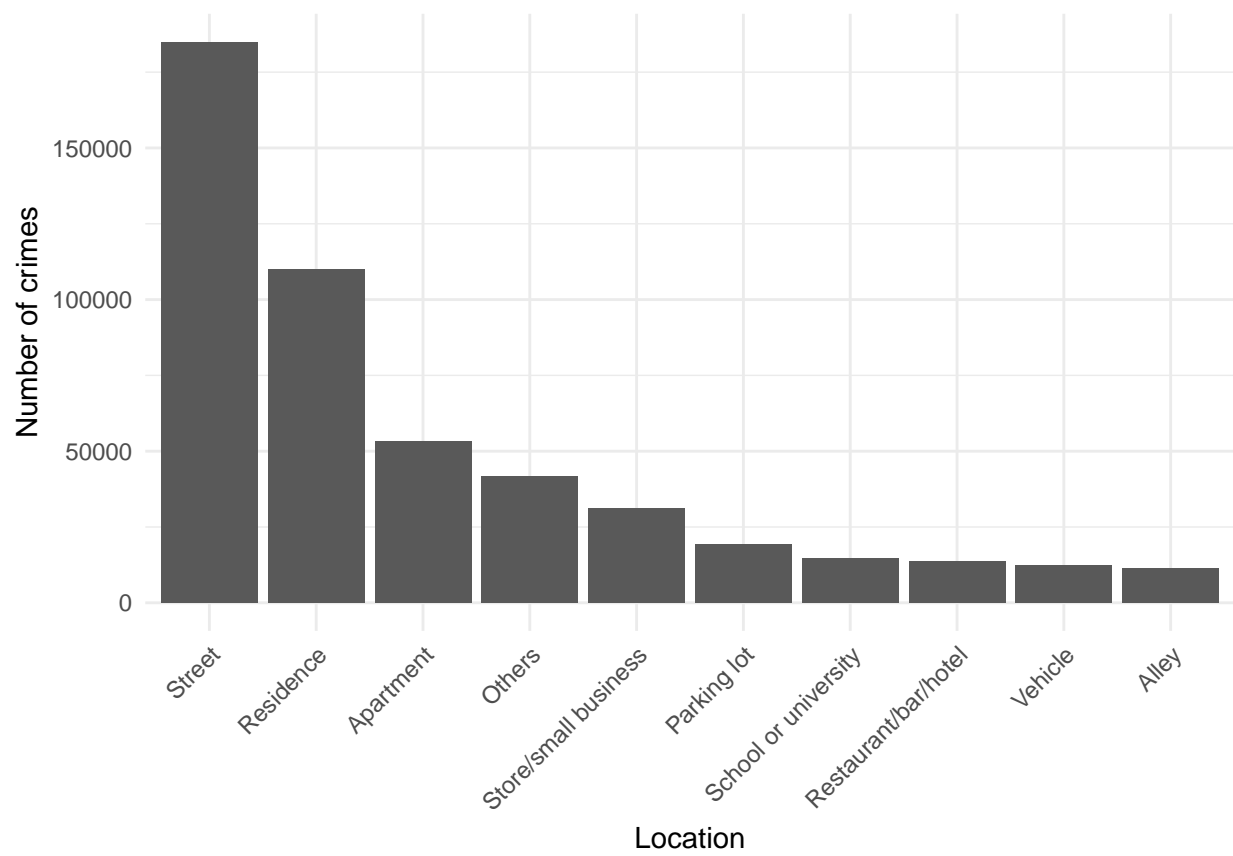


Figure 7: Figure 7. Crimes per location

### Crimes per districts

Per districts the most dangerous are 8.

### Multivariable analysis

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

### Type of crime vs hour

The most dangerous hours per Theft are 00 and 12.

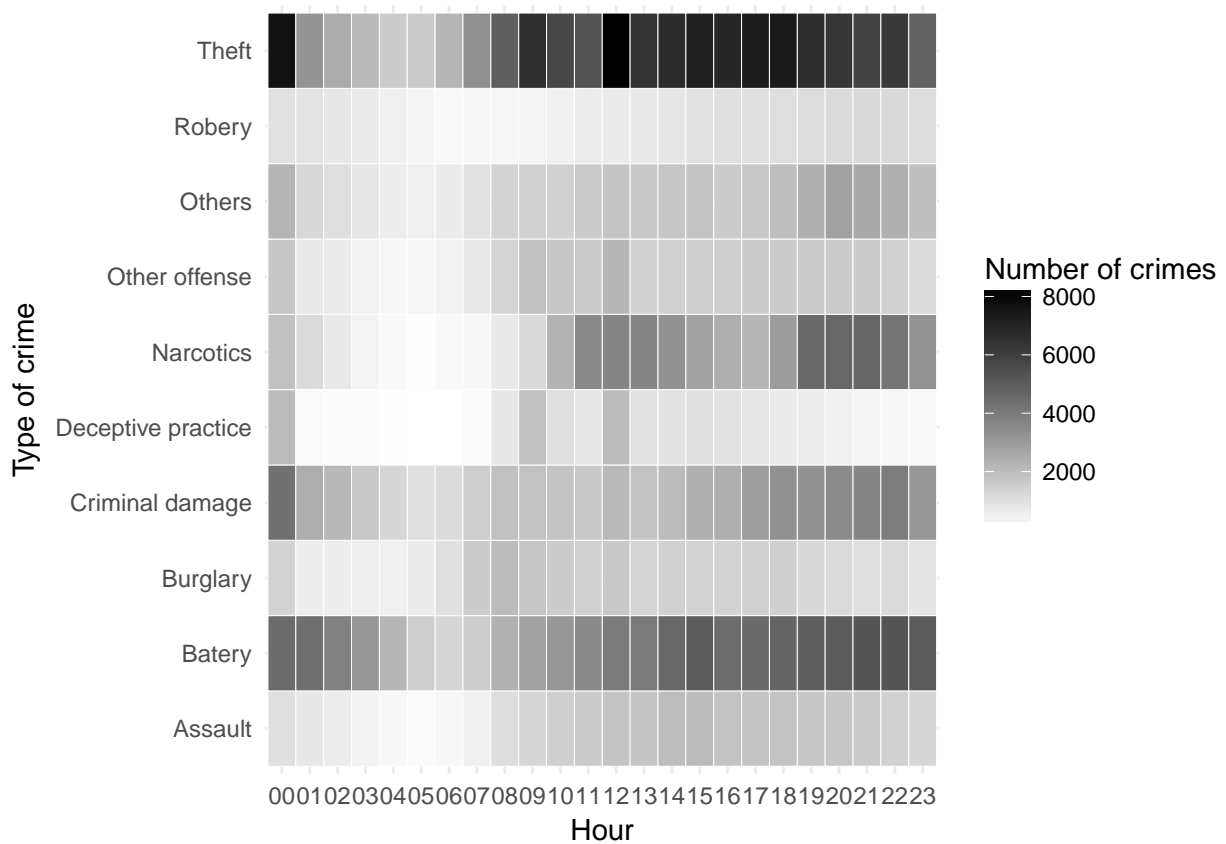


Figure 8: Figure 9. Hour vs type of crime heatmap

### Type of crime vs location

Street is particularly important for Theft.

### Type of crime vs district

Narcotics in district 11 is crealy a problem.

## Conclusions

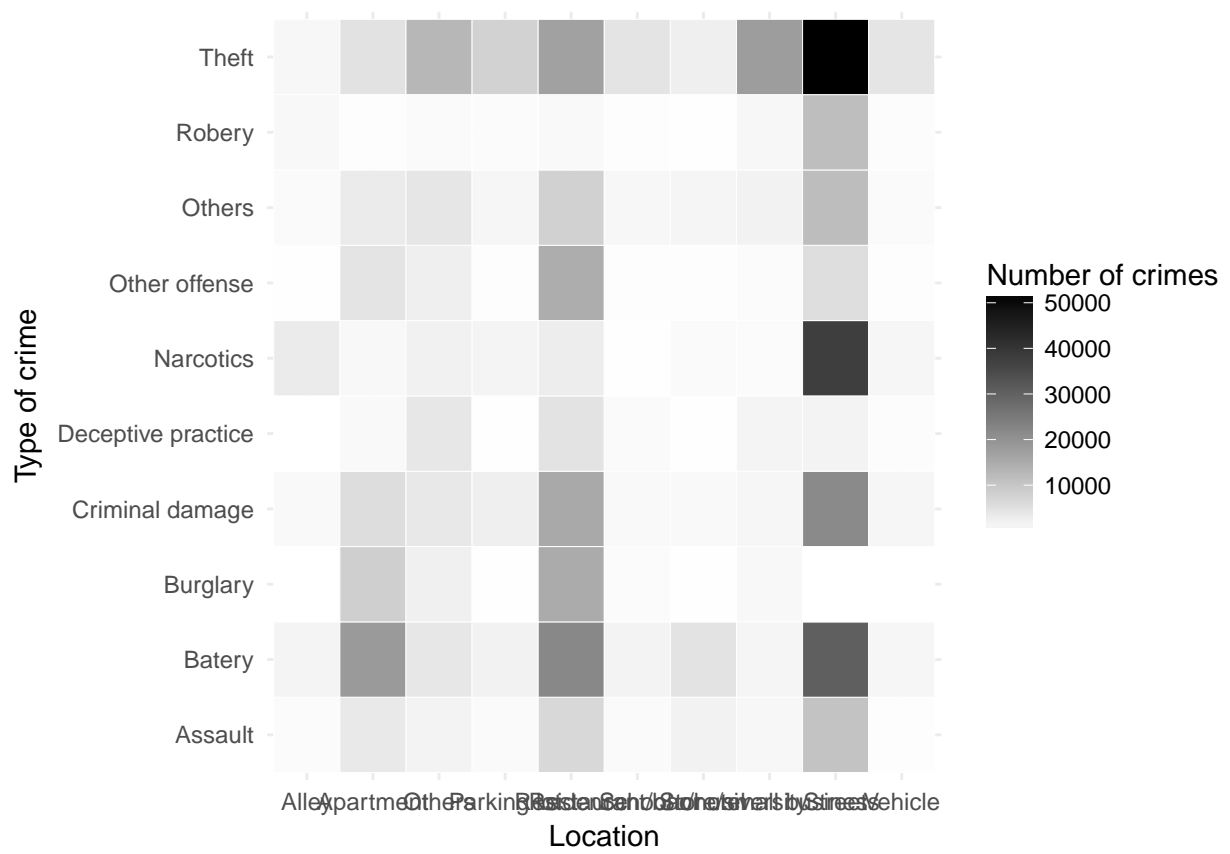


Figure 9: Figure 11. Location vs type of crime heatmap

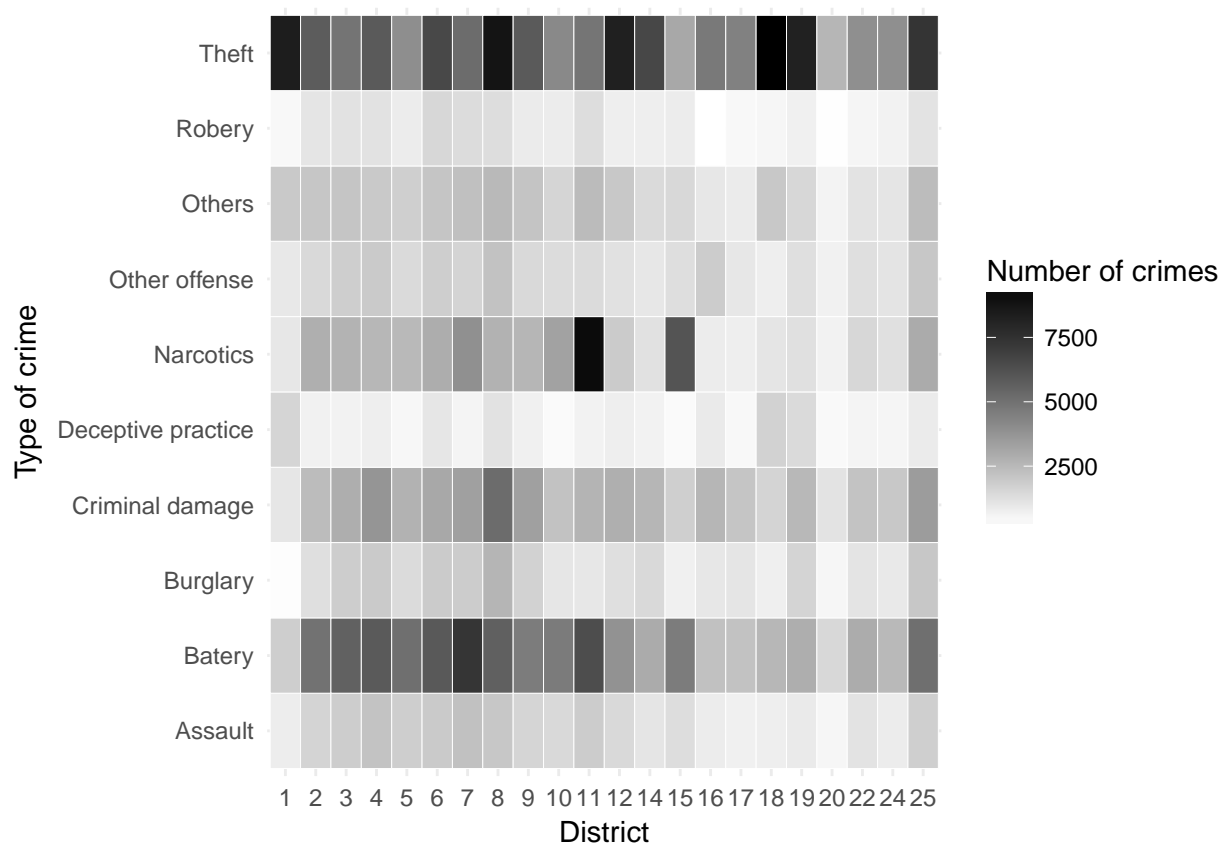


Figure 10: Figure 10. District vs type of crime heatmap