Chicago Crime: Data Analysis and Visualisations using R

Student ID: 201081646

November 3, 2018

1 Introduction

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

- How crime has changed in the city of Chicago over the years?
- What time of day do most types of crime occur?
- In which locations are specific types of crime more likely to happen?
- Which districts of the city are potentially more dangerous per type of crime?

2 Data and methods

The dataset analysed is a sample of the original data of crimes extracted from the Chicago Police Department which content the crimes 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 basic univariable analysis of the dataset, and then, with a multianalysys based on heatmaps we answer our research questions. Finally, we sumarise the findings.

This report has been done with `Rmarkdown`, although due to space limitations only the `R` code cells that have been considered most significant are shown. To see the complete code see https://github.com/eugenividal/Chicago-Crime-Data-Analysis.

3 Results

3.1 Data preparation

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

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

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

```
# Create a variable count with value 1
dd$Count <- 1
# Extract hour from Date
dd$Hour <- substring(dd$Date, 12,13)
# Drop time from Date
```

```r
dd$Date <- as.Date(dd$Date, format="%m/%d/%Y")
# Drop days from Date
dd$Month_year <- as.Date(as.yearmon(dd$Date, "%Y-%m"))
# Change format of variables
dd$Hour <- as.factor(dd$Hour)
dd$Primary.Type <- as.factor(dd$Primary.Type)
dd$Location.Description <- as.factor(dd$Location.Description)
dd$District<- as.factor(dd$District)
```

Third, we simplify the variables `Primary.Type` and `Location.Description` grouping their categories and call them `Type_grouped` and `Location_grouped` respectivelly [1].

Next, we drop all variables we do not need to perform the general anasylis or answer our research questions.

```r
# Drop all variables we are not interested in
dd <- dd[, -c(1:2, 4:11, 13:18)]
```

Then, we clean the dataset of missing values.

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

Finally, the dataset is ready for the explotation.

```r
# Show first 5 records
head(dd)
```

```
##         Date District Count Hour Month_year Type_grouped Location_grouped
## 1 2013-07-20       19     1   12 2013-07-01       Batery           Street
## 2 2013-07-20       19     1   01 2013-07-01       Others           Street
## 3 2013-07-19        2     1   09 2013-07-01      Assault        Apartment
## 4 2013-07-20        9     1   02 2013-07-01    Narcotics           Street
## 5 2013-07-12        3     1   05 2013-07-01        Theft           Street
## 6 2013-07-20        9     1   01 2013-07-01       Batery        Apartment
```

3.2 Data exploration

3.2.1 How crime has changed in the city of Chicago over the years?

Figure 1[2] shows the evolution by month of the crimes from 2001 to the present. There is an obvious periodic pattern and a clear downward trend.

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

[1] This code is not showed here due to space limitations
[2] The code to create the graphs is not showed here either

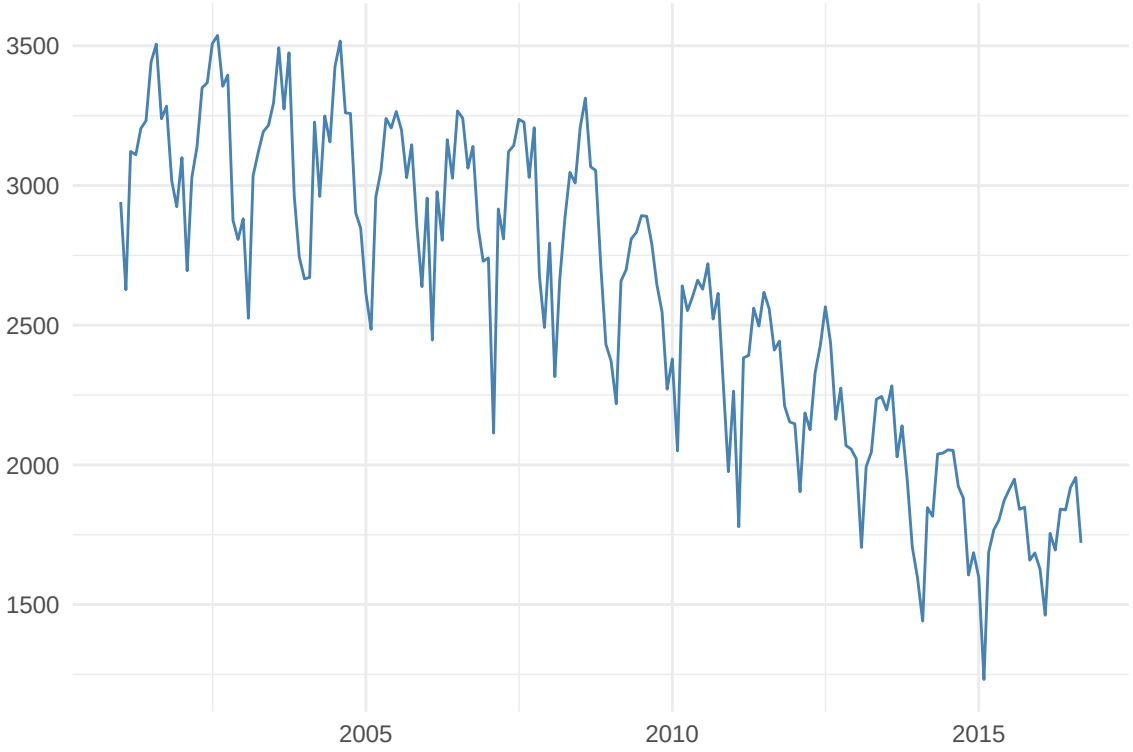

Figure 1: Crimes evolution

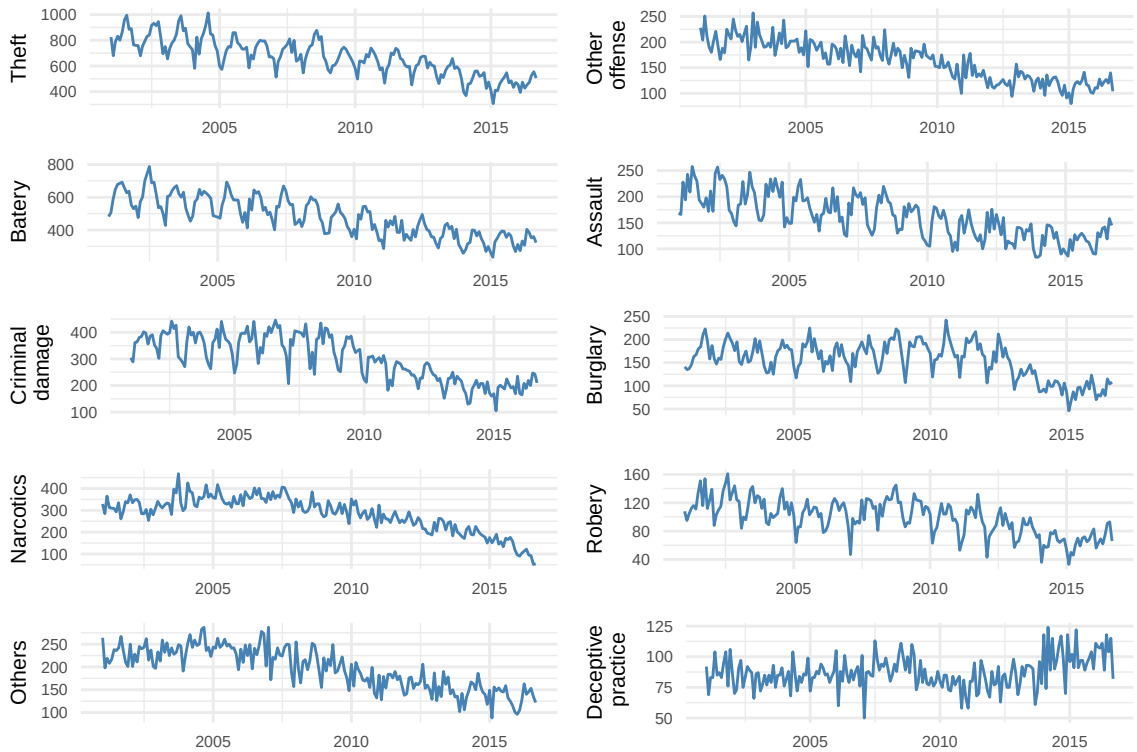

Figure 2: Evolution per type of crime

3.2.2 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?

The most dangerous hours per Thefth are 00 and 12.

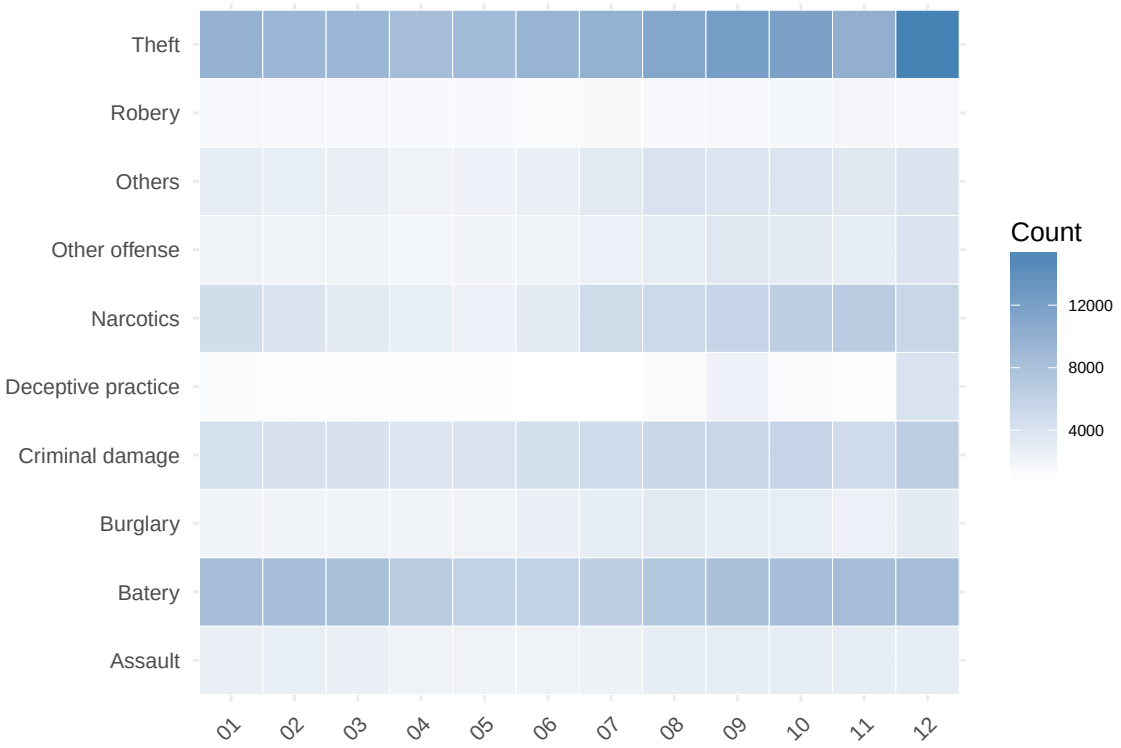

Figure 3: Type of crime vs hour

3.2.3 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.

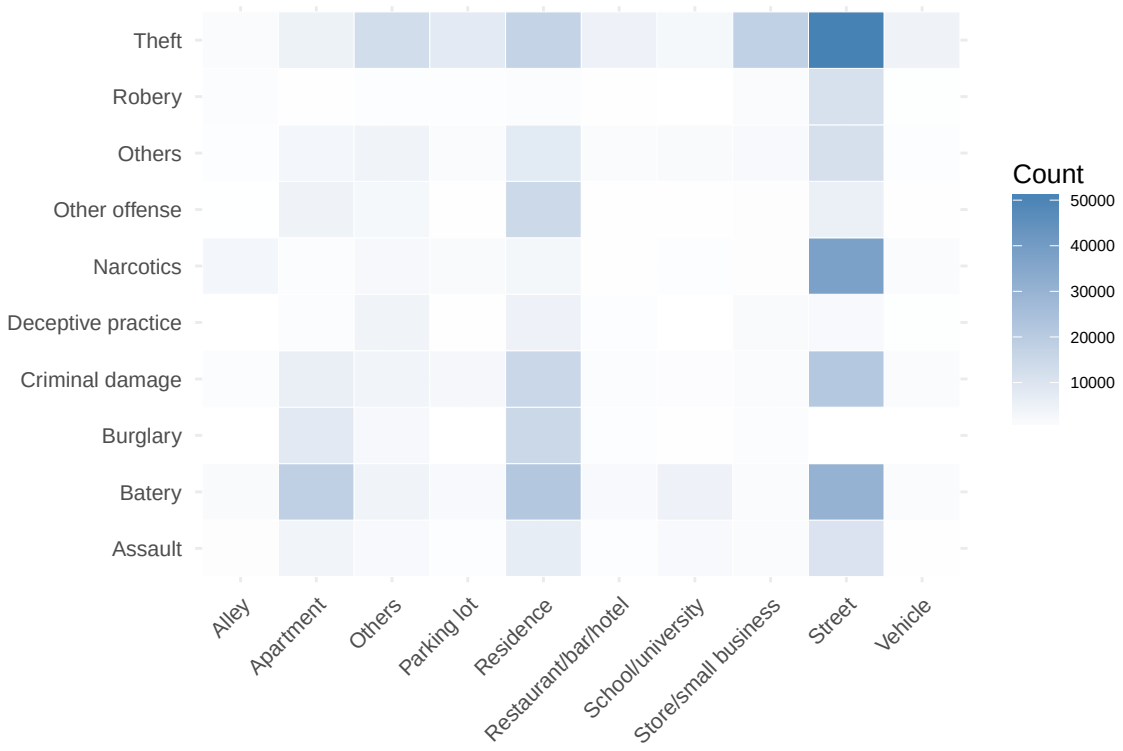

Figure 4: Type of crime vs location

3.2.4 Which districts of the city 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.

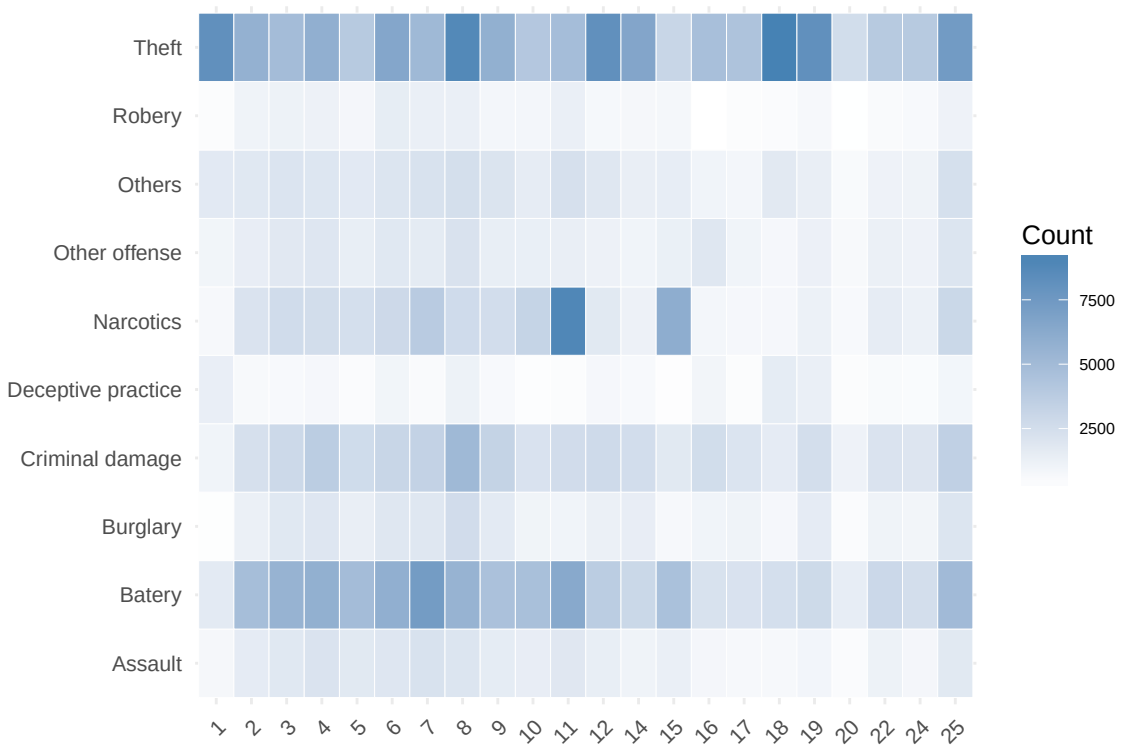

Figure 5: Type of crime vs district

4 Conclusions