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)

In the body report we will explain the most essential code. The whole code can be checked in an appendix at the end of the document.

This report has been done with Rmarkdown and it can be reproducible in github.

# 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 explotation.

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 explotation.

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.

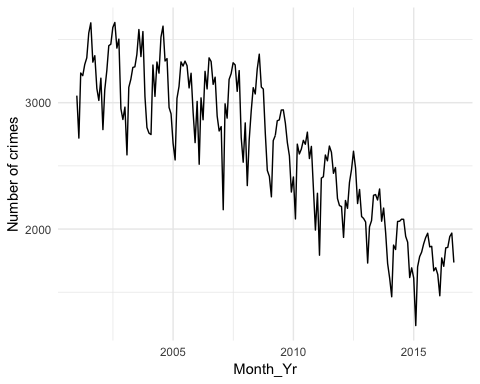


Figure 1. Vehicles, casualties and secs pairwise

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

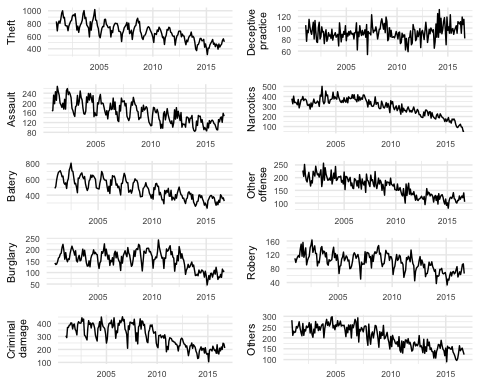


Figure 2. Vehicles, casualties and secs pairwise

#### Records by time

##### Hour

The crimes are concentrated in hours

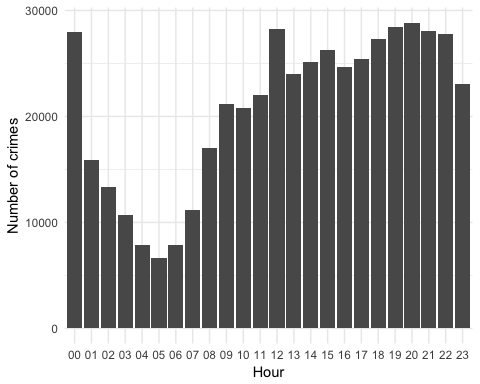


Figure 3. Crimes per hour

##### Weekday - Order days

Friday concentrated most of the crimes, percentage?

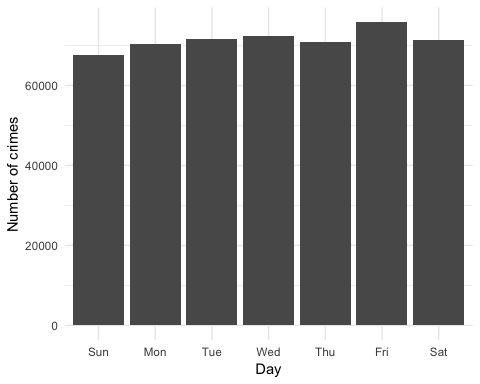


Figure 4. Crimes per week day

##### Month - Order months

Summer is in difference the period with more crimes recorded.

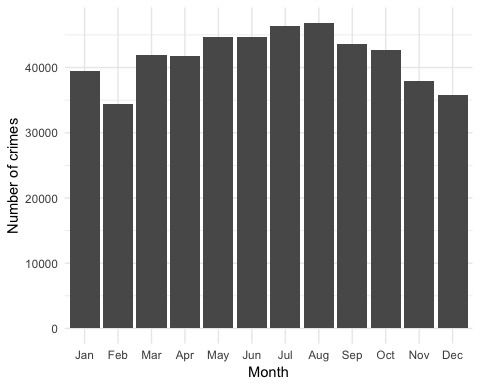


Figure 5. Crimes per month

#### Type of crimes

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

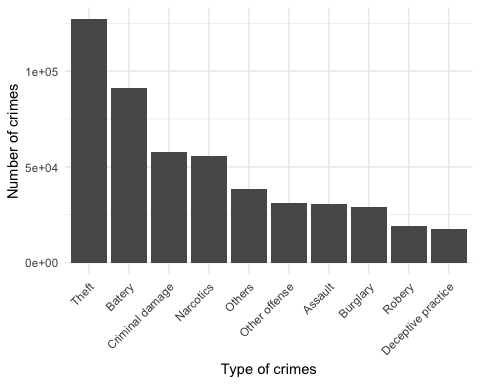


Figure 6. Crimes per type

#### Location of crimes

These crimes are concentrated in Streets, give percentage.

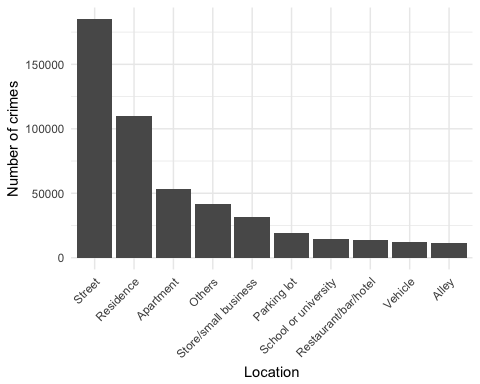


Figure 7. Crimes per location

#### Crimes per districts (map?)

Per districts the most dangerous are 8.

### Two variables analysis

The following part analyse the data in two variables. Through eat maps.

#### Type of crime vs hour

The most dangerous hours per Thefth are 00 and 12.



Figure 9. Hour vs type of crime heatmap

#### Type of crime vs location

Street is particularly important for Theft.

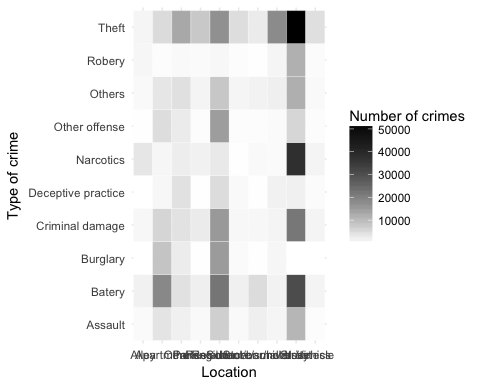


Figure 11. Location vs type of crime heatmap

#### Type of crime vs district

Narcotics in district 11 is crealy a problem.

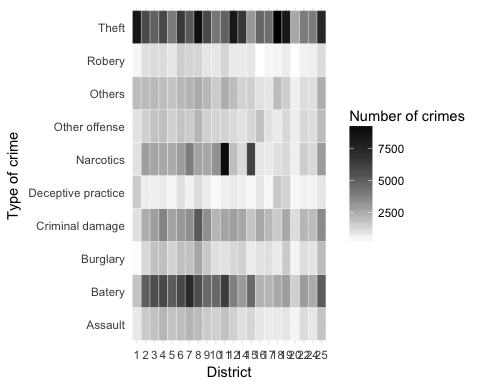


Figure 10. District vs type of crime heatmap

# Conclusions