

## Descobrimos regras de associação entre tipos de crimes realizados na cidade de Los Angeles - EUA

### 1 Introdução

Crimes que acontecem em uma cidade podem ser agrupados em transações. A partir destas transações é possível estabelecer associações entre tipos de crimes.

O objetivo deste trabalho é identificar regras de associação entre tipos de crimes que aconteceram na cidade de Los Angeles entre 1º de novembro e 31 de dezembro de 2014.

Neste trabalho as ocorrências de crimes serão agrupadas levando-se em consideração a região onde o crime aconteceu, a data e o horário do dia.

### 2 Dados

Várias cidades e países tem implementado iniciativas de *Open Data*. Los Angeles<sup>1</sup> talvez é uma das cidades no mundo com maior nível de maturidade no uso de *Open Data*.

Os registros sobre crimes que ocorreram na cidade de Los Angeles em 2014 podem ser encontrados em <sup>2</sup>. Para este trabalho foi aplicado um filtro sobre os dados acima restringindo a lista de ocorrências de crimes para somente aquelas que ocorreram entre 1º de novembro e 31 de dezembro de 2014. O resultado desta consulta pode ser encontrado em <sup>3</sup>.

Estes dados precisam ser pré-processados para serem representados no formato de transações. A primeira etapa consiste em ler os dados e converter os tipos:

```
ocorrencias <- read.csv("dataset/LAPD_Crime_and_Collision_Raw_Data_-_2014.csv")
ocorrencias <- ocorrencias[,c(3,4,6,9)]
ocorrencias$periodo <- ifelse(ocorrencias$TIME.OCC < 600, "Madrugada",
                             ifelse(ocorrencias$TIME.OCC < 1200, "Manha",
                                     ifelse(ocorrencias$TIME.OCC < 1800, "Tarde", "Noite")))
ocorrencias$TIME.OCC <- NULL
names(ocorrencias) <- c('data', 'area', 'descricao', 'periodo')
ocorrencias$data <- as.Date(ocorrencias$data, format = '%m/%d/%Y')
```

Depois, é possível fazer uma breve análise exploratória sobre os dados, por exemplo: calculando e imprimindo a quantidade de crimes por dia:

```
library(dplyr)
byData <- group_by(ocorrencias, data)
sumData <- summarize(byData, count=n())
plot(sumData$count ~ sumData$data, type='l',
     col='red', main="Quantidade de crimes por dia na cidade de Los Angeles",
     xlab="Período", ylab="Quantidade de crimes")
```

Calculando e imprimindo a quantidade de crimes por região:

<sup>1</sup><https://data.lacity.org/>

<sup>2</sup><https://data.lacity.org/A-Safe-City/LAPD-Crime-and-Collision-Raw-Data-2014/eta5-h8qx>

<sup>3</sup>[https://raw.githubusercontent.com/fbarth/posGraduacao/master/ExemplosRegrasAssociacao/dataset/LAPD\\_Crime\\_and\\_Collision\\_Raw\\_Data\\_-\\_2014.csv](https://raw.githubusercontent.com/fbarth/posGraduacao/master/ExemplosRegrasAssociacao/dataset/LAPD_Crime_and_Collision_Raw_Data_-_2014.csv)

```
par(las=2)
par(cex.axis=0.6)
barplot(table(ocorrencias$area), main="Quantidade de crimes por região",
        xlab="Região", ylab="Quantidade", col="cyan")
```

Calculando e imprimindo a quantidade de ocorrências por crime:

```
par(cex.axis=0.3)
barplot(sort(table(ocorrencias$descricao), decreasing = TRUE),
        main="Quantidade de ocorrências por crime",
        xlab="Crimes", ylab="Quantidade", col="cyan")
```

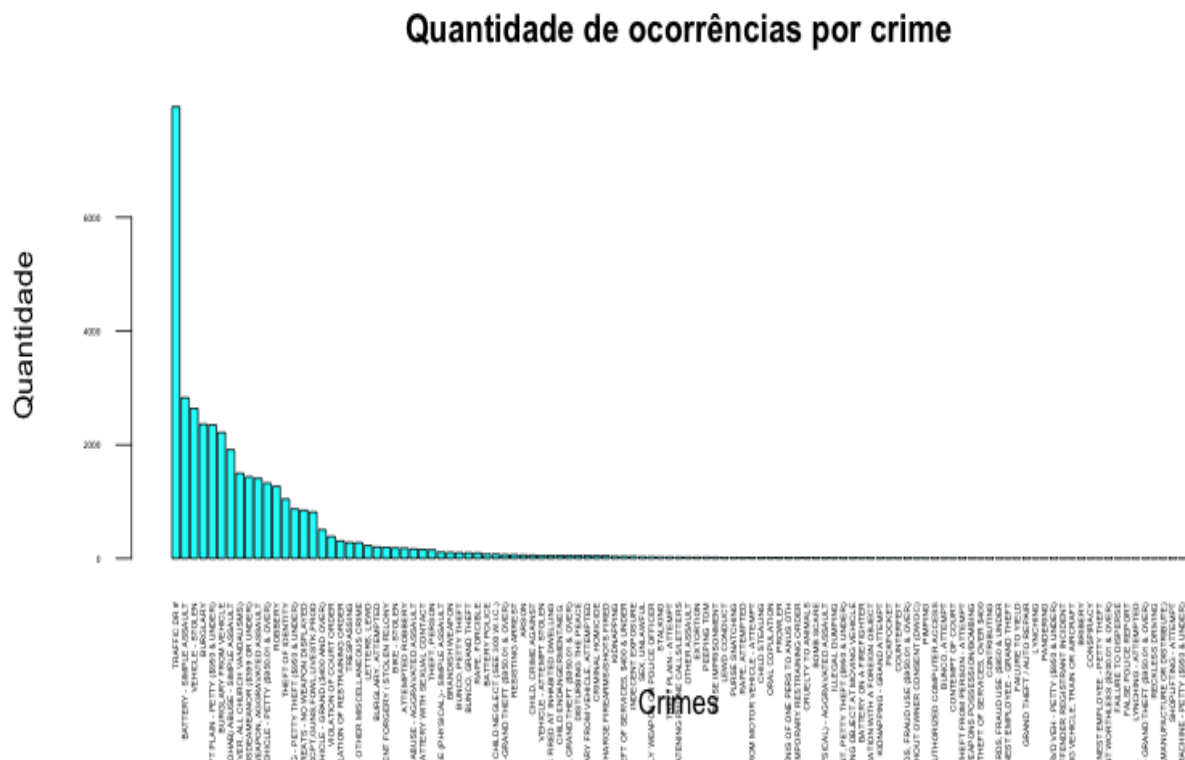


Figura 1: Quantidade de ocorrências por crime: este gráfico ilustra a quantidade de ocorrências por tipo de crime. No eixo x são apresentados os tipos de crimes com ocorrência no dataset e no eixo y são apresentadas as quantidades das ocorrências.

Na figura 1 é possível visualizar que o tipo de crime *TRAFFIC DR* ocorre com uma frequência bem superior aos demais tipos de crime. Portanto, este tipo de crime deve ser removido do dataset:

```
ocorrencias <- subset(ocorrencias, ocorrencias$descricao != 'TRAFFIC_DR.#')
```

Gerando a base de transações de tipo de crimes agrupadas por data, região e período do dia:

```
library(arules)
temp <- data.frame(id=paste(ocorrencias$periodo,
                           ocorrencias$data,
                           ocorrencias$area,
                           sep = "_",
                           collapse = NULL),
                  crime=ocorrencias$descricao)
write.csv(temp, file = "dataset/temp_trans.csv")

tr <- read.transactions("dataset/temp_trans.csv",
```

```
format="single", sep=',',  
cols = c('id', 'crime'),  
rm.duplicates = TRUE)  
tr
```

O script, na íntegra, que realiza este pré-processamento pode ser encontrado em <sup>4</sup>.

### 3 Método de trabalho

Sugere-se seguir as seguintes etapas para a realização desta análise:

- Aquisição, análise exploratório e pré-processamento dos dados com o script citado acima;
- Construção das regras de associação;
- Interpretação e análise das regras, e;
- Preparação do relatório e principais artefatos que devem ser entregues.

### 4 Itens que devem ser entregues

Os seguintes itens devem ser entregues:

- Um relatório apresentando os principais resultados da análise. Este relatório pode ser entregue na forma de slides, documento PDF ou documento HTML.
- Independente do formato do relatório, o conteúdo do mesmo deve possuir os seguintes itens:
  - introdução, com descrição do contexto e objetivo do trabalho;
  - descrição do método utilizado para alcançar o objetivo, e;
  - conclusão, ou seja, descrição dos resultados encontrados com a análise realizada.
- O relatório deve possuir duas figuras que ilustram e resumizam bem os achados. Cada uma das figuras deve ter um *caption* com uma descrição autocontida da figura.
- O script que realiza todas as transformações e análises no dataset. Este item geralmente não é entregue para o cliente final. No entanto, será relevante para a avaliação deste trabalho.

### 5 Equipes e prazos

Este trabalho poderá ser feito individualmente ou em dupla e deverá ser entregue até o final do curso.

---

<sup>4</sup><https://github.com/fbarth/posGraduacao/blob/master/ExemplosRegrasAssociacao/scripts/analiseLAPDCrime.R>