Analisis Espacial Hurto Celulares

Alejandro Rodas Vásquez

Introduction

Usually, most of the published papers don't show the technical details releated with the investigation (like the source code). Therefore, the purpose of this document is share some of the main source code that was the final product of the present investigation.

Also, since the context of this investigation was in Pereria (Colombia) all the variables names are in Spanish.

You can consult the following github repository if you want to know more. http://rmarkdown.rstudio.com.

Read of Data Set

```
library(dplyr)

path<-file.path(getwd(),"DataSet","Hurto.tsv")
column_names <- c("fecha","departamento","municipio","dia","hora","barrio","zona",
                  "claseDeSitio","armaEmpleada","movilAgresor","movilVictima","edad",
                  "sexo","estadoCivil","paisNacimiento","claseEmpleo","profesion",
                  "escolaridad","codigoDANE","clase","marca","linea","cantidad")

hurto.tsv <- read.delim(path, col.names = column_names,stringsAsFactors = TRUE)


#Save the new data set
save(hurto_colombia_data_set, ascii = TRUE,
     file = file.path(getwd(),"DataSet","hurto_colombia_data_set.Rda"))
```

Change Some Data

One of the aims on this project was to analyze the Cell Phone Thefs from the Temporal perpective. To do this, it most be imperative transform some data. In the previous analysis of the data set, days like Miercoles (Wednesday) and Sabado (Saturday) are represent with the diferent encoding.

For that reason, the function called dias_transformar was created.

```
library(dplyr)
library(tidyverse)

load(file =file.path(getwd(),"DataSet","hurto_colombia_data_set_tbl.Rda"))

dias_transformar <- function(dia) {
  if(dia == "Mi\303\251rcoles") {
    return("Miercoles")
  }
  if(dia == "S\303\241bado") {
    return("Sabado")
  }
```

```
  return(as.character(dia))
}

hurto_colombia_data_set_tbl$dia_formato <-
  sapply(hurto_colombia_data_set_tbl$dia,dias_transformar)
```

Making Intervals

To make a successfull Temporal Analysis the column that containts time information has to be transform to an POSIXct object. This transformation is required to begin to manipulate Time Series.

```
# Transform days to factor
hurto_colombia_data_set_tbl$dia_formato <-
  factor(hurto_colombia_data_set_tbl$dia_formato,
        order=TRUE,levels = c("Lunes","Martes","Miercoles",
                              "Jueves","Viernes","Sabado","Domingo"))


# Just simple do some format date field
hora_format <- as.POSIXct(hurto_colombia_data_set_tbl$hora, format="%m/%d/%Y %I:%M:%S %p")
hurto_colombia_data_set_tbl$hora_hurto <- format(hora_format, "%H:%M:%S")

# This column is created for store time fields and help to create
# the intervals to group the data
hurto_colombia_data_set_tbl$hora_hurto_time <- times(format(hora_format, "%H:%M:%S"))

# Create intervals
time.tag <- chron(times=c('00:00:00','06:00:00','12:00:00','18:00:00','23:59:00'))
hurto_colombia_data_set$hurto_interval <-
        cut(hurto_colombia_data_set_tbl$hora_hurto_time,breaks = time.tag,
         labels = c("00-06","06-12", "12-18", "18-00"),
         include.lowest = TRUE)

# Create data set for only Pereira
hurto_pereira_data_set <- hurto_colombia_data_set_tbl %>%
  select("dia_formato","marca","sexo","departamento","municipio",
        "barrio","hora_hurto","hora_hurto_time","hurto_interval") %>%
  filter(municipio == "PEREIRA (CT)")


save(hurto_pereira_data_set, ascii = TRUE,
     file = file.path(getwd(),"DataSet","hurto_pereira_data_set_tbl.Rda"))
```

Some Temporal Analysis

```
# Generate a time series line geometry diagram to show the Afertoon and Night

library(ggplot2)
library(dplyr)

load(file =file.path(getwd(),"DataSet","hurto_pereira_data_set_tbl.Rda"))
```

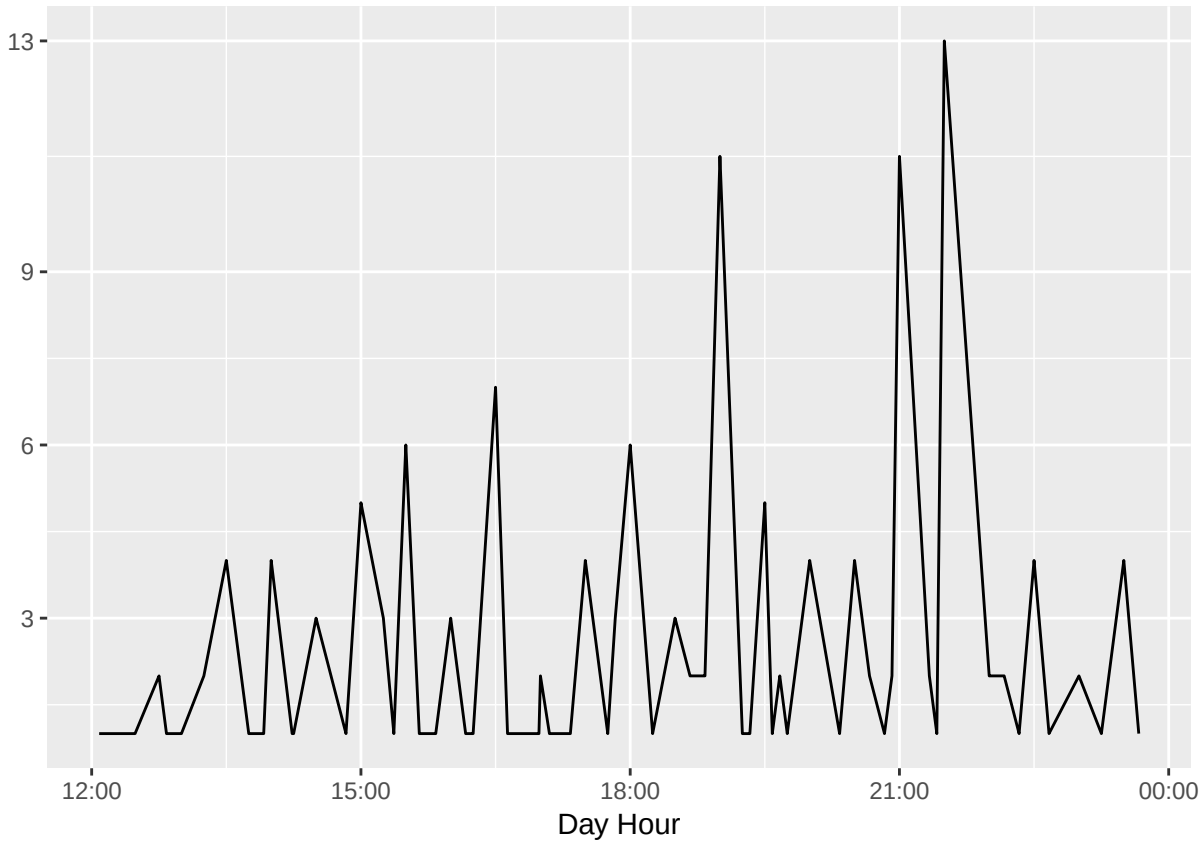
```
hurtos_tarde_sabado <- hurto_pereira_data_set %>%
  select(dia_formato,hora_hurto,hurto_interval,hora_format_POSIXct) %>%
  filter(dia_formato=="Sabado", hurto_interval=="12-18"| hurto_interval=="18-00") %>%
  group_by(hora_format_POSIXct,hurto_interval) %>%
  summarise(num_hurtos=n()) %>%
  arrange(hora_format_POSIXct)

hurtos_tarde_sabado
```

Pueden ver la tabla resultante

```
## # A tibble: 72 x 3
## # Groups:   hora_format_POSIXct [72]
##    hora_format_POSIXct hurto_interval num_hurtos
##    <dttm>              <ord>               <int>
##  1 2018-12-26 12:05:00 12-18                   1
##  2 2018-12-26 12:12:00 12-18                   1
##  3 2018-12-26 12:27:00 12-18                   1
##  4 2018-12-26 12:29:00 12-18                   1
##  5 2018-12-26 12:45:00 12-18                   2
##  6 2018-12-26 12:50:00 12-18                   1
##  7 2018-12-26 13:00:00 12-18                   1
##  8 2018-12-26 13:15:00 12-18                   2
##  9 2018-12-26 13:30:00 12-18                   4
## 10 2018-12-26 13:45:00 12-18                   1
## # ... with 62 more rows
```

Linea temporal

```
ggplot(hurtos_tarde_sabado,aes(hora_format_POSIXct,num_hurtos)) +
  geom_line() +
  xlab("Day Hour")+
  scale_y_continuous("",breaks = c(3,6,9,13))
```



Spatial Analysis

Doing some mergeing

```
library(tidyverse)
library(rgdal)
library(tmap)
library(sp)
library(RColorBrewer)

load(file =file.path(getwd(),"DataSet","hurto_pereira_barrio_comuna_data_set_tbl.Rda"))
barrios.comunas <- read_csv2(file.path(getwd(),"DataSet","comunas_barrios_pereira.csv"),
                             col_types = cols(
                               nombre_comuna = col_character(),
                               barrio = col_character()
                             ))

glimpse(barrios.comunas)
```

Data set barrios y comunas

```
## Using ',' as decimal and '.' as grouping mark. Use read_delim() for more control.
```

```
## Observations: 441
## Variables: 2
```

```
## $ nombre_comuna <chr> "COMUNA CONSOTA", "COMUNA CONSOTA", "COMUNA RIO ...
## $ barrio        <chr> "EL DORADO I", "EL DORADO II", "PRIMERO DE MAYO"...
```

Ahora mostramos la lectura del archivo shape file

```
# Read the shapefile that contains the map of city of Pereira
Mapa.pereira <- readOGR("/Users/alejandrorodas/Documents/Libros/R/R_ejercicios/R_in_action/SpatialAnaly
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/alejandrorodas/Documents/Libros/R/R_ejercicios/R_in_action/SpatialAnalysis/comunas/C
## with 21 features
## It has 10 fields
## Integer64 fields read as strings:  FID id_comuna
```

Estas son las columnas del DataFrame del objeto SpatialDataFrame

```
glimpse(Mapa.pereira@data)
```

```
## Observations: 21
## Variables: 10
## $ FID        <fct> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ...
## $ id_comuna  <fct> 3, 14, 15, 10, 4, 1, 16, 17, 5, 6, 19, 11, 12, 13, ...
## $ cod_suelo  <fct> 6600101, 6600101, 6600101, 6600101, 6600101, 660010...
## $ codcom     <fct> 12, 07, 09, 11, 03, 18, 01, 19, NA, 08, 13, 06, 16,...
## $ nombre     <fct> Comuna Ferrocarril, Comuna Boston, Comuna Cuba, Com...
## $ norma      <fct> Ac 63/1992, Ac 2/1999, Ac 63/1992, Ac 63/1992, Ac 6...
## $ act_geogra <fct> 2006-07-01T00:00:00.000Z, 2006-12-27T00:00:00.000Z,...
## $ area       <dbl> 1850916.1, 1773483.8, 758100.2, 3069008.9, 2122294....
## $ Shape__Are <dbl> 1.508069e-04, 1.444893e-04, 6.176647e-05, 2.500523e...
## $ Shape__Len <dbl> 0.13661251, 0.07851261, 0.04367972, 0.11276787, 0.0...
```

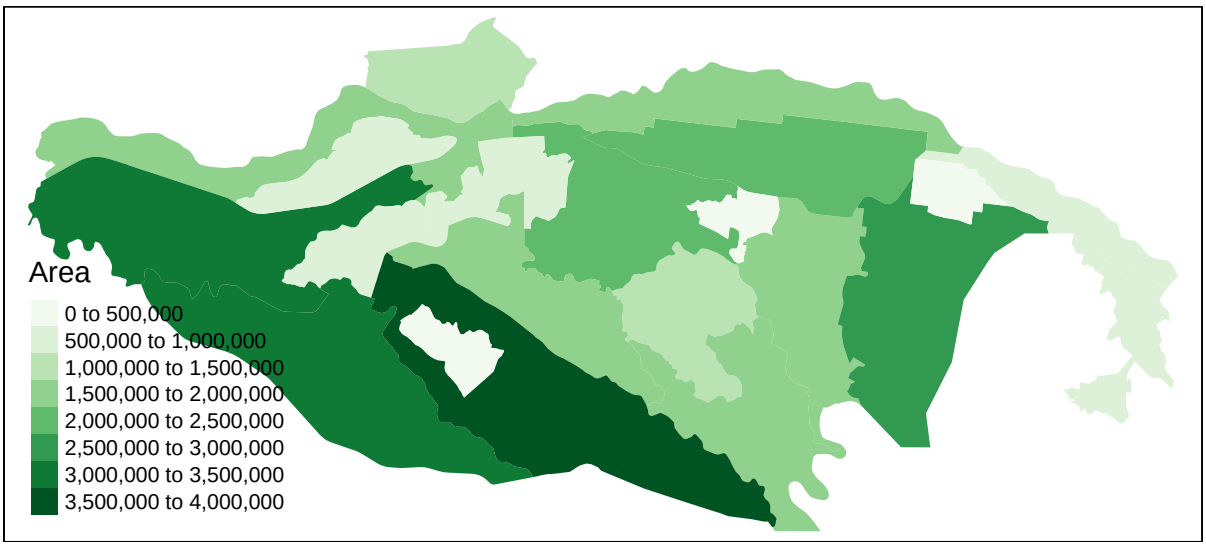
Combina el ShapeFile de las Comunas en Pereira con la lista de Barrios/Comunas

```
mapa_pereira_comuna_barrio <- merge(Mapa.pereira, barrios.comunas, by.x="nombre", by.y="nombre_comuna",

# Guarda shape file Mapa de Comunas de Pereira asociado con los hurtos
writeOGR(mapa_pereira_comuna_barrio,
         file.path(getwd(),"DataSet","MapaPereiraComunaBarrioHurto"),
         'mapa_pereira_comuna_barrio',
         "ESRI Shapefile")
```

```
library(rgdal)
library(tmap)
library(sp)
library(RColorBrewer)

mapa_pereira_comuna_barrio <-
  readOGR(file.path(getwd(),"DataSet","MapaPereiraComunaBarrioHurto"),'mapa_pereira_comuna_barrio')
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/alejandrorodas/Documents/Libros/R/R_ejercicios/R_in_action/ArticuloAnalisisEspacial/
## with 21 features
## It has 6 fields
```

```
tm_shape(mapa_pereira_comuna_barrio) +
  tm_fill("area",title = "Area",palette ="Greens")
```


Area
0 to 500,000
500,000 to 1,000,000
1,000,000 to 1,500,000
1,500,000 to 2,000,000
2,000,000 to 2,500,000
2,500,000 to 3,000,000
3,000,000 to 3,500,000
3,500,000 to 4,000,000