

Proyecto por Modalidad 2018

Caro Vivianita

6/6/2019

```
library(tidyverse)

## -- Attaching packages -----
tidyverse 1.2.1 --

## v ggplot2 3.1.1      v purrr 0.3.2
## v tibble 2.1.1      v dplyr 0.8.0.1
## v tidyr 0.8.3       v stringr 1.4.0
## v readr 1.3.1       v forcats 0.4.0

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()

library(sf)

## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3

Robo_2018 <- read.csv2("C:/Users/DEGIYEC/Desktop/CIENCIA DE DATOS/Robo
2018 - para proyecto final.csv")
```

Queremos ver si hay una correlación entre ROBO A Mano Armada y los días de la semana

`head(Robo_2018)` # para que me muestre lo que subi

```
##   INTEG_CAR_CODIGO..USAR.Y.NO.PUBLICAR.  A.O    MES DIA.DE.LA.SEMANA
## 1                P01541-2018-0249989 2018    Mayo           MARTES
## 2                D02207-2018-0546980 2018 Octubre          LUNES
## 3                P02184-2018-0512127 2018 Octubre          JUEVES
## 4                P02167-2018-0512051 2018 Octubre          JUEVES
## 5                K01955-2018-0534481 2018 Octubre          JUEVES
## 6                D00609-2018-0319863 2018    Julio          DOMINGO
##   FECHA.DEL.HECHO DIA FRANJA.24.HS HORA.DEL.HECHO    BARRIO
Comuna
## 1      22/5/2018  22              22          22:30    Flores
Comuna 7
## 2      22/10/2018  22              19          19:46    Barracas
Comuna 4
## 3       4/10/2018   4              12          12:35    Belgrano Comuna
13
```

## 4	4/10/2018	4	20	20:37	Nueva Pompeya
Comuna 4					
## 5	11/10/2018	11	17	17:10	Floresta Comuna
10					
## 6	1/7/2018	1	22	22:30	Villa Crespo Comuna
15					
##	Longitud	Latitud	ROBO	ROBO.de.MOTO	ROBO.de.MOTO.AMA AUTO
AUTO.AMA					
## 1	-58,7030326	-34,479665	1	0	0 0
0					
## 2	-58,6753978	-34,6570243	1	0	0 1
1					
## 3	-58,6570859	-34,8495826	1	0	0 0
0					
## 4	-58,6229664	-34,7862476	1	0	0 1
1					
## 5	-58,6167212	-34,6558611	1	0	0 0
0					
## 6	-58,5894662	-34,6128487	1	0	0 0
0					
##	En.MOTO	En.MOTO.AMA	AMA	AUSENCIA	
## 1	0	0	0	0	
## 2	0	0	1	0	
## 3	0	0	0	0	
## 4	0	0	1	0	
## 5	0	0	0	0	
## 6	0	0	0	0	

LIMPIEZA GENERAL

Arreglamos coordenadas, que a veces aparecen con comas, y otras con puntos, por lo cual R las leyo como texto. También pasamos el campo comuna a todo mayuscula (unificamos)

```
Robo_2018 <- Robo_2018 %>% # piso el archivo con mis correcciones (que voy a hacer ahora)
  mutate(Longitud = sub(",", ".", Longitud), # substituyo "," por "."
         Latitud = sub(",", ".", Latitud),
         Longitud = as.numeric(Longitud), # convierto el campo de texto a número "as"
         Latitud = as.numeric(Latitud))

## Warning: NAs introducidos por coerción

## Warning: NAs introducidos por coerción
```

GRAFICO AGRUPADO POR MODALIDAD

Convertimos tanto el mes como el dia de la semana en factores (data categorica), ordenados

```
Robo_2018 <- Robo_2018 %>%
  mutate(MES = factor(MES, levels = c("Enero", "Febrero", "Marzo",
    "Abril", "Mayo", "Junio", "Julio",
    "Agosto", "Septiembre", "Octubre",
    "Noviembre", "Diciembre"),
    ordered = TRUE),
  DIA.DE.LA.SEMANA = factor(DIA.DE.LA.SEMANA, levels = c("LUNES",
    "MARTES", "MIERCOLES", "JUEVES",
    "VIERNES", "SABADO", "DOMINGO"),
    ordered = TRUE))

Robo_AMA_2018 <- Robo_2018 %>%
  group_by(MES, DIA.DE.LA.SEMANA) %>%
  summarise(TOTAL_AMA = sum(AMA))
```

```
Robo_AMA_2018
```

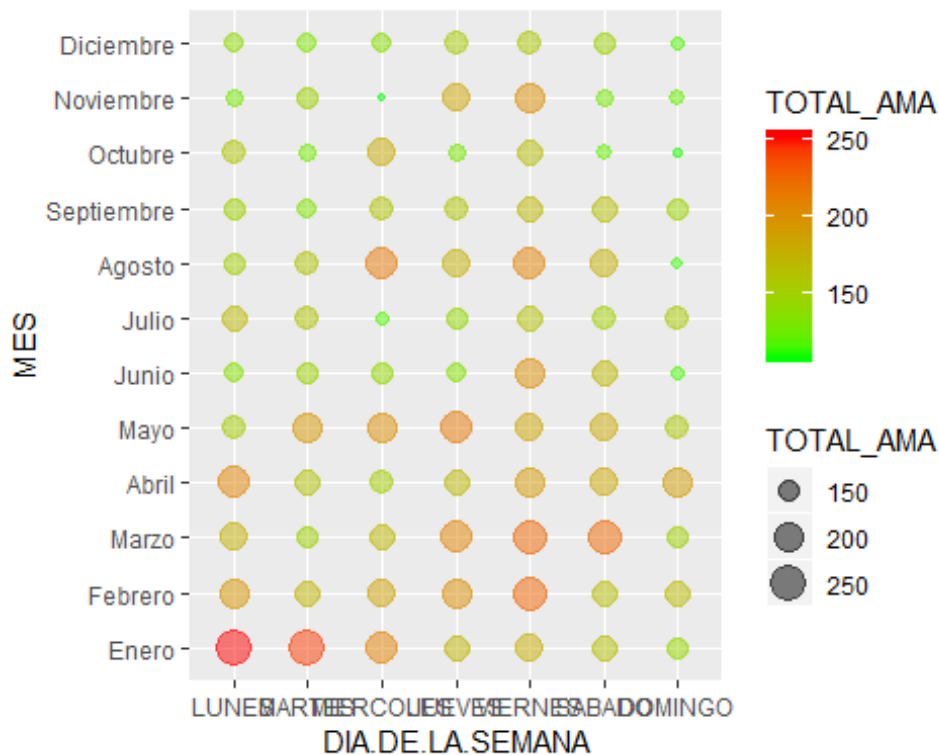
```
## # A tibble: 84 x 3
## # Groups:   MES [12]
##   MES      DIA.DE.LA.SEMANA TOTAL_AMA
##   <ord>    <ord>              <int>
## 1 Enero   LUNES                252
## 2 Enero   MARTES                243
## 3 Enero   MIERCOLES             210
## 4 Enero   JUEVES                172
## 5 Enero   VIERNES                177
## 6 Enero   SABADO                164
## 7 Enero   DOMINGO               144
## 8 Febrero LUNES                196
## 9 Febrero MARTES                172
## 10 Febrero MIERCOLES             188
## # ... with 74 more rows
```

Gráfico por mes y días de los Robos AMA:

```
options(scipen = 20) # para que la escala de los X no me aparezca en
notación científica
ggplot() +
  geom_point(data = Robo_AMA_2018, aes (x=DIA.DE.LA.SEMANA, y = MES,
    color = TOTAL_AMA, size=TOTAL_AMA), #Gráfico día de la semana vs robo
    auto distinguido por colores y por tamaño de población
    alpha = 0.5) + # con 0.5 de transparencia para que me
    permita ver todos los puntos
  scale_color_gradient(low = "green", high = "red") + # para agregar los
    colores a mi escala como yo quiera (pongo el color de los extremos --> el
    mas alto y el mas bajo) Sin esta aclaración, el color (puesto como Color
    = TOTAL_AMA) lo elige solo
  geom_smooth(data= Robo_AMA_2018, aes(x = DIA.DE.LA.SEMANA, y = MES), #
```

hasta acá, mi gráfico me muestra la tendencia apegada a segmentos, PERO MUY sensible a los datos... deben ser muy apegados a la realidad y no me ayuda con los nuevos (o "a futuro")

```
method = "lm") # para que mi regresión sea lineal
```



Correlación

```
modelo_Robo_AMA_2018 <- lm(Robo_AMA_2018$TOTAL_AMA ~
as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)) # usamos un modelo global
para ver cuánto hay de correlación entre Robo AMA y los días de la
semana. Como los ordené como YO QUERIA (y eso marea a mi "modelo lineal",
le digo que tome de mi archivo "Robo_AMA_2018" la columna "TOTAL_AMA" con
el signo $ ; y lo mismo con DIA.DE.LA.SEMANA (pero le aclaro que es un
character para que no tome el orden que le di antes: Lunes no vale mas
que martes, miércoles no vale mas que jueves, etc.))
```

```
modelo_Robo_AMA_2018
```

```
##
```

```
## Call:
```

```
## lm(formula = Robo_AMA_2018$TOTAL_AMA ~
as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA))
```

```
##
```

```
## Coefficients:
```

```
##
```

```
(Intercept)
```

```
##
```

```
140.42
```

```

##      as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)JUEVES
##                                     30.08
##      as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)LUNES
##                                     27.33
##      as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MARTES
##                                     20.25
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MIERCOLES
##                                     24.67
##      as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)SABADO
##                                     24.25
##      as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)VIERNES
##                                     49.08

summary(modelo_Robo_AMA_2018) # para ver las estrellas "*"

##
## Call:
## lm(formula = Robo_AMA_2018$TOTAL_AMA ~
as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -57.083 -23.417  -0.208   15.771   84.250
##
## Coefficients:
##                                     Estimate Std.
Error
## (Intercept)                                     140.417
8.696
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)JUEVES      30.083
12.298
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)LUNES      27.333
12.298
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MARTES      20.250
12.298
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MIERCOLES    24.667
12.298
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)SABADO      24.250
12.298
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)VIERNES     49.083
12.298
##                                     t value
## (Intercept)                                     16.147
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)JUEVES      2.446
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)LUNES      2.223
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MARTES      1.647
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MIERCOLES    2.006
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)SABADO      1.972
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)VIERNES     3.991
##

```

```

Pr(>|t|)
## (Intercept) <
0.0000000000000002
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)JUEVES
0.016720
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)LUNES
0.029181
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MARTES
0.103719
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MIERCOLES
0.048400
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)SABADO
0.052220
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)VIERNES
0.000149
##
## (Intercept) ***
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)JUEVES *
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)LUNES *
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MARTES
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)MIERCOLES *
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)SABADO .
## as.character(Robo_AMA_2018$DIA.DE.LA.SEMANA)VIERNES ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 30.12 on 77 degrees of freedom
## Multiple R-squared:  0.1778, Adjusted R-squared:  0.1138
## F-statistic: 2.776 on 6 and 77 DF,  p-value: 0.01697

```

Llegamos a la conclusión, que hay una correlación positiva entre el día Viernes y los Robos con Arma.

Dato que resulta fundamental para la prevención.

ABRIR EL ARCHIVO EN R

```
Robo_2018 <- read.csv2("C:/Users/DEGIYEC/Desktop/CIENCIA DE DATOS/Robo  
2018 - para proyecto final.csv")
```

```
head(Robo_2018) # para que me muestre lo que subi
```

```
##   INTEG_CAR_CODIGO..USAR.Y.NO.PUBLICAR.  A.O      MES DIA.DE.LA.SEMANA  
## 1                P01541-2018-0249989 2018    Mayo           MARTES  
## 2                D02207-2018-0546980 2018 Octubre          LUNES  
## 3                P02184-2018-0512127 2018 Octubre          JUEVES  
## 4                P02167-2018-0512051 2018 Octubre          JUEVES  
## 5                K01955-2018-0534481 2018 Octubre          JUEVES  
## 6                D00609-2018-0319863 2018    Julio           DOMINGO  
##   FECHA.DEL.HECHO DIA FRANJA.24.HS HORA.DEL.HECHO      BARRIO  
COMUNA  
## 1      22/5/2018  22          22          22:30      Flores  
Comuna 7  
## 2      22/10/2018 22          19          19:46      Barracas  
Comuna 4  
## 3      4/10/2018  4          12          12:35      Belgrano Comuna  
13  
## 4      4/10/2018  4          20          20:37 Nueva Pompeya  
Comuna 4  
## 5      11/10/2018 11          17          17:10      Floresta Comuna  
10  
## 6      1/7/2018   1          22          22:30      Villa Crespo Comuna  
15  
##      Longitud      Latitud ROBO ROBO.de.MOTO ROBO.de.MOTO.AMA AUTO  
AUTO.AMA  
## 1 -58,7030326  -34,479665    1          0          0      0  
0  
## 2 -58,6753978  -34,6570243    1          0          0      1  
1  
## 3 -58,6570859  -34,8495826    1          0          0      0  
0  
## 4 -58,6229664  -34,7862476    1          0          0      1  
1  
## 5 -58,6167212  -34,6558611    1          0          0      0  
0  
## 6 -58,5894662  -34,6128487    1          0          0      0  
0  
##   En.MOTO En.MOTO.AMA AMA AUSENCIA  
## 1      0          0      0      0  
## 2      0          0      1      0  
## 3      0          0      0      0  
## 4      0          0      1      0  
## 5      0          0      0      0  
## 6      0          0      0      0
```

LIMPIEZA GENERAL

Arreglamos coordenadas, que a veces aparecen con comas, y otras con puntos, por lo cual R las leyo como texto. También pasamos el campo comuna a todo mayuscula (unificamos)

```
Robo_2018 <- Robo_2018 %>% # piso el archivo con mis correcciones (que voy a hacer ahora)
  mutate(Longitud = sub(",", ".", Longitud), # substituyo "," por "."
         Latitud = sub(",", ".", Latitud),
         Longitud = as.numeric(Longitud), # convierto el campo de texto a número "as"
         Latitud = as.numeric(Latitud),
         COMUNA = toupper(COMUNA), # que me ponga comuna todo en mayuscula
         COMUNA = factor(COMUNA, levels = c("COMUNA 1",
                                             "COMUNA 2",
                                             "COMUNA 3",
                                             "COMUNA 4",
                                             "COMUNA 5",
                                             "COMUNA 6",
                                             "COMUNA 7",
                                             "COMUNA 8",
                                             "COMUNA 9",
                                             "COMUNA 10",
                                             "COMUNA 11",
                                             "COMUNA 12",
                                             "COMUNA 13",
                                             "COMUNA 14",
                                             "COMUNA 15"),
                        ordered = TRUE)) # que me ordene como quiero

## Warning: NAs introducidos por coerción
## Warning: NAs introducidos por coerción
```

GRAFICO AGRUPADO POR COMUNA

n(): es para que cuente filas

```
Robos_por_Comuna <- Robo_2018 %>%
  group_by(COMUNA) %>%
  summarise(TOTAL = n())

## Warning: Factor `COMUNA` contains implicit NA, consider using
## `forcats::fct_explicit_na`

Robos_por_Comuna

## # A tibble: 16 x 2
##   COMUNA     TOTAL
```



```
##      <ord>      <int>
##  1 COMUNA 1    9262
##  2 COMUNA 2    3587
##  3 COMUNA 3    5862
##  4 COMUNA 4    6688
##  5 COMUNA 5    4484
##  6 COMUNA 6    3375
##  7 COMUNA 7    5754
##  8 COMUNA 8    3586
##  9 COMUNA 9    3857
## 10 COMUNA 10   3213
## 11 COMUNA 11   3259
## 12 COMUNA 12   3284
## 13 COMUNA 13   4163
## 14 COMUNA 14   4983
## 15 COMUNA 15   4052
## 16 <NA>       1712
```

Leo la población por comuna

```
habitantes_x_comuna <- read.csv("C:/Users/DEGIYEC/Desktop/CIENCIA DE
DATOS/Proyecto en R/gcba_pob_comunas_17.csv")
```

Le agrego la palabra comuna para poder hacer un match

```
habitantes_x_comuna <- habitantes_x_comuna %>% # piso el archivo con mis
correcciones (que voy a hacer ahora)
```

```
  mutate(COMUNA = case_when(COMUNA == 1 ~ "COMUNA 1", # que reemplace si
encuentra la variable que le pido. Tipo la función "if"
```

```
    COMUNA == 2 ~ "COMUNA 2",
    COMUNA == 3 ~ "COMUNA 3",
    COMUNA == 4 ~ "COMUNA 4",
    COMUNA == 5 ~ "COMUNA 5",
    COMUNA == 6 ~ "COMUNA 6",
    COMUNA == 7 ~ "COMUNA 7",
    COMUNA == 8 ~ "COMUNA 8",
    COMUNA == 9 ~ "COMUNA 9",
    COMUNA == 10 ~ "COMUNA 10",
    COMUNA == 11 ~ "COMUNA 11",
    COMUNA == 12 ~ "COMUNA 12",
    COMUNA == 13 ~ "COMUNA 13",
    COMUNA == 14 ~ "COMUNA 14",
    COMUNA == 15 ~ "COMUNA 15",))
```

Agregamos los habitantes a cada Comuna

```
Robos_por_Comuna <- Robos_por_Comuna %>% # piso el archivo con mis
correcciones (que voy a hacer ahora)
```

```
  left_join(habitantes_x_comuna, by = "COMUNA")
```

```
## Warning: Column `COMUNA` joining factor and character vector, coercing
into
## character vector

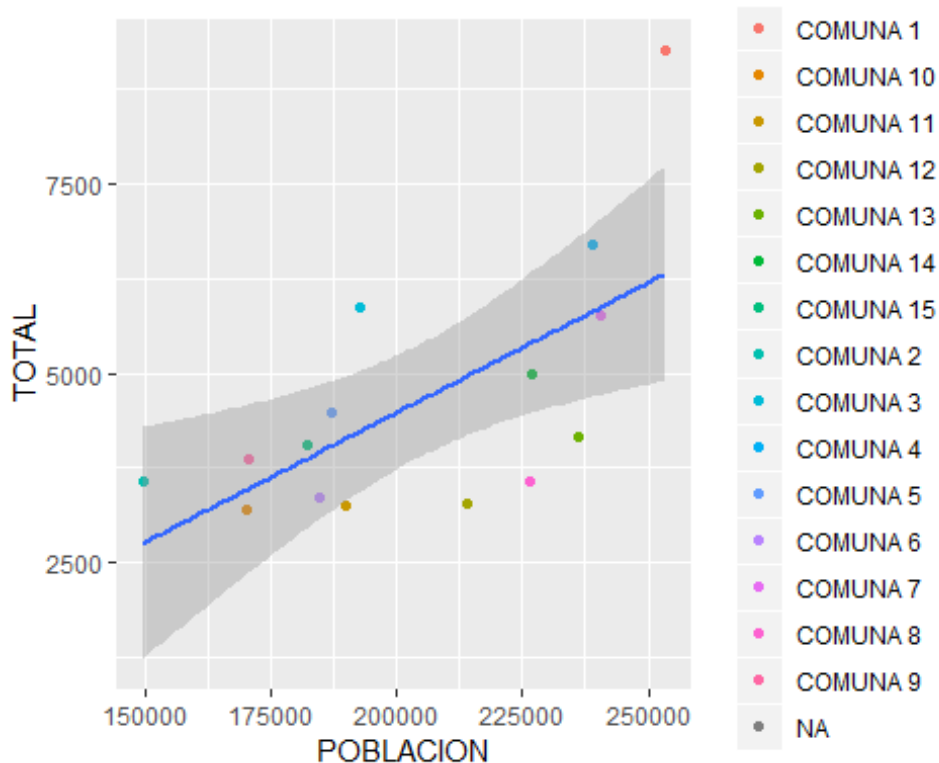
Robos_por_Comuna

## # A tibble: 16 x 3
##   COMUNA   TOTAL POBLACION
##   <chr>   <int>   <int>
## 1 COMUNA 1    9262    253271
## 2 COMUNA 2    3587    149720
## 3 COMUNA 3    5862    192763
## 4 COMUNA 4    6688    238809
## 5 COMUNA 5    4484    186956
## 6 COMUNA 6    3375    184846
## 7 COMUNA 7    5754    240607
## 8 COMUNA 8    3586    226649
## 9 COMUNA 9    3857    170605
## 10 COMUNA 10  3213    170282
## 11 COMUNA 11  3259    189986
## 12 COMUNA 12  3284    213914
## 13 COMUNA 13  4163    235967
## 14 COMUNA 14  4983    226944
## 15 COMUNA 15  4052    182409
## 16 <NA>      1712         NA
```

Ahora deberíamos graficar para comparar densidades

```
ggplot() +
  geom_point(data = Robos_por_Comuna, aes (x=POBLACION, y = TOTAL, color
= COMUNA)) +
  geom_smooth(data = Robos_por_Comuna, aes(x = POBLACION, y
= TOTAL),
              method = "lm")

## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
```



```
library(sf) # para que incluya un lector de datos espaciales

## Linking to GEOS 3.6.1, GDAL 2.2.3, PROJ 4.9.3

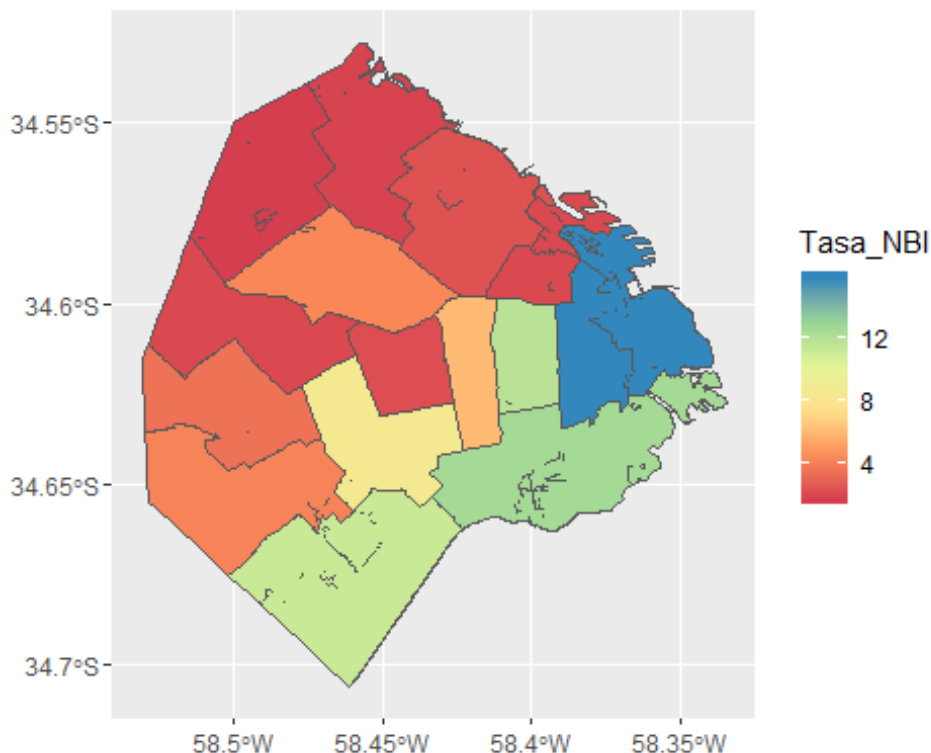
radios <-
st_read("http://cdn.buenosaires.gob.ar/datosabiertos/datasets/informacion-
-censal-por-radio/CABA_rc.geojson") #traje los datos censales

## Reading layer `CABA_rc' from data source
`http://cdn.buenosaires.gob.ar/datosabiertos/datasets/informacion-censal-
por-radio/CABA_rc.geojson' using driver `GeoJSON'
## Simple feature collection with 3554 features and 8 fields
## geometry type:  MULTIPOLYGON
## dimension:      XY
## bbox:           xmin: -58.53092 ymin: -34.70574 xmax: -58.33455 ymax:
-34.528
## epsg (SRID):    4326
## proj4string:     +proj=longlat +datum=WGS84 +no_defs

Tasa_Hogares_NBI <- radios %>% #Nombre como se va a llamar mi archivo que
estoy creando con las nuevas variables
  group_by(COMUNA) %>%
  summarize(HOGARES = sum(HOGARES, na.rm = TRUE),
            HOGARES_NBI = sum(HOGARES_NBI, na.rm = TRUE)) %>% # con el
na.rm = TRUE hago que me saque los valores invalidos/faltantes
  mutate(Tasa_NBI = HOGARES_NBI/HOGARES*100)
```

Graficamos la tasa de Hogares NBI por comuna (que es lo que me interesa comparar con los Robo por comuna)

```
ggplot() +  
  geom_sf(data = Tasa_Hogares_NBI, aes(fill = Tasa_NBI)) + # Graficar en  
  un mapa las comunas y la tasa de hogares NBI  
  scale_fill_distiller(palette = "Spectral", direction = 1)
```



Si quiero renombrar las comunas de mi archivo nuevo (Tasa_Hogares_NBI) para que me coincida con mis comunas

```
Tasa_Hogares_NBI <- Tasa_Hogares_NBI %>% # piso el archivo con mis  
  correcciones (que voy a hacer ahora)  
  mutate(COMUNA = case_when(COMUNA == 1 ~ "COMUNA 1", # que reemplace si  
    encuentra la variable que le pido. Tipo la función "if"  
    COMUNA == 2 ~ "COMUNA 2",  
    COMUNA == 3 ~ "COMUNA 3",  
    COMUNA == 4 ~ "COMUNA 4",  
    COMUNA == 5 ~ "COMUNA 5",  
    COMUNA == 6 ~ "COMUNA 6",  
    COMUNA == 7 ~ "COMUNA 7",  
    COMUNA == 8 ~ "COMUNA 8",  
    COMUNA == 9 ~ "COMUNA 9",  
    COMUNA == 10 ~ "COMUNA 10",  
    COMUNA == 11 ~ "COMUNA 11",  
    COMUNA == 12 ~ "COMUNA 12",  
    COMUNA == 13 ~ "COMUNA 13",
```

```
COMUNA == 14 ~ "COMUNA 14",
COMUNA == 15 ~ "COMUNA 15",))
```

Agregamos la tasa de NBI a cada Comuna

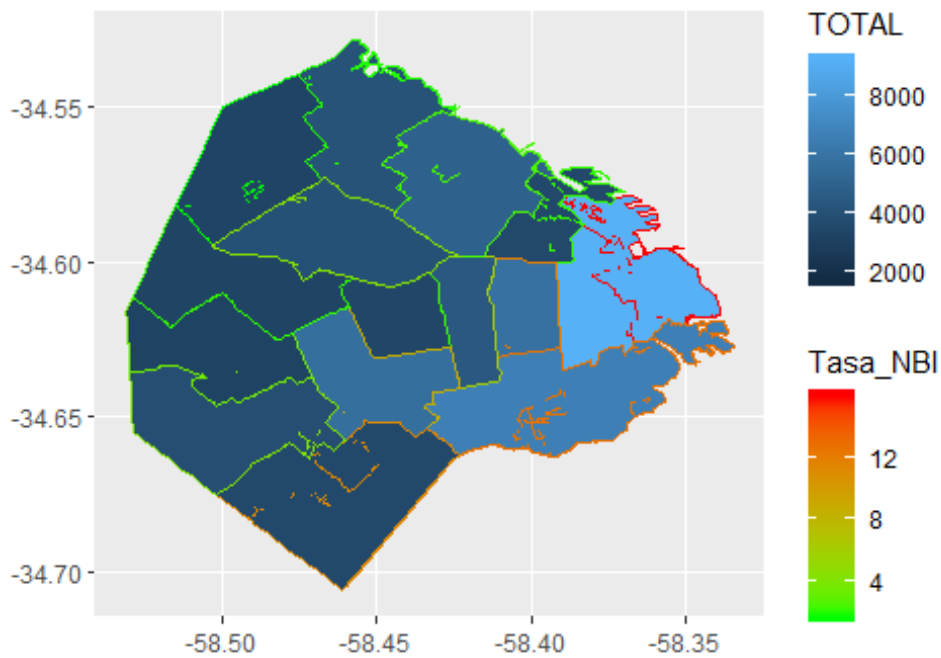
```
Robos_por_Comuna <- Robos_por_Comuna %>% # piso el archivo con mis
correcciones (que voy a hacer ahora)
  left_join(Tasa_Hogares_NBI, by = "COMUNA")
```

Robos_por_Comuna

```
## # A tibble: 16 x 7
##   COMUNA TOTAL POBLACION HOGARES HOGARES_NBI Tasa_NBI
##   <chr> <int> <int> <dbl> <dbl> <dbl>
## 1 COMUN~ 9262 253271 84468 13429 15.9
## 2 COMUN~ 3587 149720 73156 1497 2.05
## 3 COMUN~ 5862 192763 80489 9560 11.9
## 4 COMUN~ 6688 238809 76455 9678 12.7
## 5 COMUN~ 4484 186956 76846 4652 6.05
## 6 COMUN~ 3375 184846 75189 1656 2.20
## 7 COMUN~ 5754 240607 81483 7040 8.64
## 8 COMUN~ 3586 226649 58204 6582 11.3
## 9 COMUN~ 3857 170605 56495 2345 4.15
## 10 COMUN~ 3213 170282 61453 2149 3.50
## 11 COMUN~ 3259 189986 71460 1444 2.02
## 12 COMUN~ 3284 213914 78547 1335 1.70
## 13 COMUN~ 4163 235967 100257 1879 1.87
## 14 COMUN~ 4983 226944 103167 2425 2.35
## 15 COMUN~ 4052 182409 72465 3105 4.28
## 16 <NA> 1712 NA NA NA NA
## # ... with 1 more variable: geometry <GEOMETRY [Â°]>
```

Finalmente grafico la tasa de Hogares NBI y la cantidad de robos, ambas por comuna para ver si hay alguna relación entre ambas.

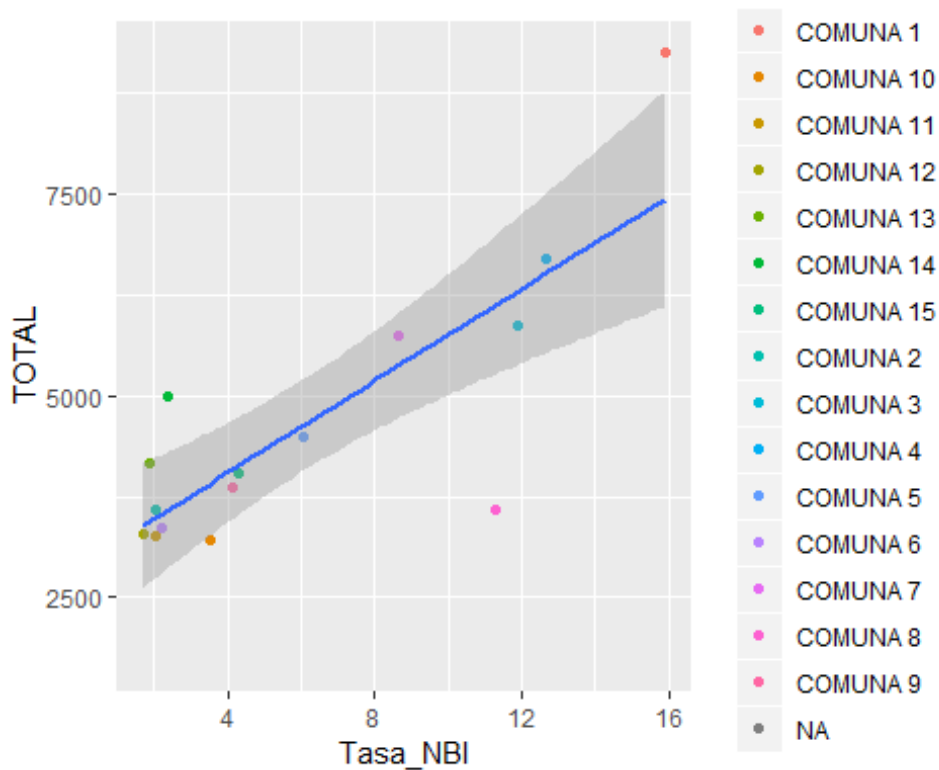
```
ggplot() +
  geom_sf(data = Robos_por_Comuna, aes(fill = TOTAL, color = Tasa_NBI)) +
  # Graficar en un mapa las comunas y la tasa de hogares NBI
  scale_color_gradient(low = "green", high = "red")
```



Otra forma de graficar

```
ggplot() +
  geom_point(data = Robos_por_Comuna, aes (x=Tasa_NBI, y = TOTAL, color =
COMUNA)) +
  geom_smooth(data = Robos_por_Comuna, aes(x = Tasa_NBI, y =
TOTAL),
              method = "lm")

## Warning: Removed 1 rows containing non-finite values (stat_smooth).
## Warning: Removed 1 rows containing missing values (geom_point).
```



Por último, veamos si hay alguna correlación estadística significativa entre estos datos

```
Regresion_lineal <- lm(data = Robos_por_Comuna,
  TOTAL ~ COMUNA + Tasa_NBI + POBLACION)
```

```
summary (Regresion_lineal)
```

```
##
## Call:
## lm(formula = TOTAL ~ COMUNA + Tasa_NBI + POBLACION, data =
Robos_por_Comuna)
##
## Residuals:
## ALL 15 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (2 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9262         NA      NA      NA
## COMUNACOMUNA 10  -6049         NA      NA      NA
## COMUNACOMUNA 11  -6003         NA      NA      NA
## COMUNACOMUNA 12  -5978         NA      NA      NA
## COMUNACOMUNA 13  -5099         NA      NA      NA
## COMUNACOMUNA 14  -4279         NA      NA      NA
## COMUNACOMUNA 15  -5210         NA      NA      NA
## COMUNACOMUNA 2   -5675         NA      NA      NA
## COMUNACOMUNA 3   -3400         NA      NA      NA
```

```
## COMUNACOMUNA 4      -2574      NA      NA      NA
## COMUNACOMUNA 5      -4778      NA      NA      NA
## COMUNACOMUNA 6      -5887      NA      NA      NA
## COMUNACOMUNA 7      -3508      NA      NA      NA
## COMUNACOMUNA 8      -5676      NA      NA      NA
## COMUNACOMUNA 9      -5405      NA      NA      NA
## Tasa_NBI              NA      NA      NA      NA
## POBLACION            NA      NA      NA      NA
##
## Residual standard error: NaN on 0 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:      1, Adjusted R-squared:      NaN
## F-statistic:      NaN on 14 and 0 DF, p-value: NA
```

Conclusión: Solo hay correlación entre Robos y Tasa_NBI, pero no es estadísticamente significativa en ninguna comuna en particular.

ABRIR EL ARCHIVO EN R

```
Robo_2018 <- read.csv2("C:/Users/DEGIYEC/Desktop/CIENCIA DE DATOS/Robo  
2018 - para proyecto final.csv")
```

```
head(Robo_2018) # para que me muestre lo que subi
```

```
##   INTEG_CAR_CODIGO..USAR.Y.NO.PUBLICAR.  A.O      MES DIA.DE.LA.SEMANA  
## 1                P01541-2018-0249989 2018    Mayo           MARTES  
## 2                D02207-2018-0546980 2018 Octubre          LUNES  
## 3                P02184-2018-0512127 2018 Octubre          JUEVES  
## 4                P02167-2018-0512051 2018 Octubre          JUEVES  
## 5                K01955-2018-0534481 2018 Octubre          JUEVES  
## 6                D00609-2018-0319863 2018    Julio           DOMINGO  
##   FECHA.DEL.HECHO DIA FRANJA.24.HS HORA.DEL.HECHO      BARRIO  
COMUNA  
## 1      22/5/2018  22          22          22:30      Flores  
Comuna 7  
## 2      22/10/2018 22          19          19:46      Barracas  
Comuna 4  
## 3      4/10/2018  4          12          12:35      Belgrano Comuna  
13  
## 4      4/10/2018  4          20          20:37      Nueva Pompeya  
Comuna 4  
## 5      11/10/2018 11          17          17:10      Floresta Comuna  
10  
## 6      1/7/2018   1          22          22:30      Villa Crespo Comuna  
15  
##      Longitud      Latitud ROBO ROBO.de.MOTO ROBO.de.MOTO.AMA AUTO  
AUTO.AMA  
## 1 -58,7030326  -34,479665    1          0          0      0  
0  
## 2 -58,6753978  -34,6570243    1          0          0      1  
1  
## 3 -58,6570859  -34,8495826    1          0          0      0  
0  
## 4 -58,6229664  -34,7862476    1          0          0      1  
1  
## 5 -58,6167212  -34,6558611    1          0          0      0  
0  
## 6 -58,5894662  -34,6128487    1          0          0      0  
0  
##   En.MOTO En.MOTO.AMA AMA AUSENCIA  
## 1      0          0      0      0  
## 2      0          0      1      0  
## 3      0          0      0      0  
## 4      0          0      1      0  
## 5      0          0      0      0  
## 6      0          0      0      0
```

LIMPIEZA GENERAL

Arreglamos coordenadas, que a veces aparecen con comas, y otras con puntos, por lo cual R las leyó como texto. También pasamos el campo comuna a todo mayúscula (unificamos)

```
Robo_2018 <- Robo_2018 %>% # piso el archivo con mis correcciones (que voy a hacer ahora)
  mutate(Longitud = sub(",", ".", Longitud), # substituyo "," por "."
         Latitud = sub(",", ".", Latitud),
         Longitud = as.numeric(Longitud), # convierto el campo de texto a número "as"
         Latitud = as.numeric(Latitud))

## Warning: NAs introducidos por coerción

## Warning: NAs introducidos por coerción
```

Leo los meses con lluvia

```
Lluvia_x_mes <- read.csv2("C:/Users/DEGIYEC/Desktop/CIENCIA DE DATOS/Lluvia por mes 2018.csv")
```

Lluvia_x_mes

##	Año	MES	mm	DIAS
## 1	2018	Enero	49.3	7
## 2	2018	Febrero	39.8	6
## 3	2018	Marzo	70.1	6
## 4	2018	Abril	263.5	11
## 5	2018	Mayo	216.6	14
## 6	2018	Junio	6.6	4
## 7	2018	Julio	125.3	12
## 8	2018	Agosto	54.3	8
## 9	2018	Septiembre	111.7	8
## 10	2018	Octubre	32.0	10
## 11	2018	Noviembre	189.5	8
## 12	2018	Diciembre	314.2	10

Filtramos la cantidad de Robo con Moto

```
Robo_2018_Moto_lluvia <- Robo_2018 %>%
  group_by(MES)%>%
  summarise(Total_Robo_en_Moto = sum(En.MOTO))
```

Robo_2018_Moto_lluvia

```
## # A tibble: 12 x 2
##   MES      Total_Robo_en_Moto
```

```
##      <fct>                <int>
##  1 Abril                    787
##  2 Agosto                   890
##  3 Diciembre               1271
##  4 Enero                   1030
##  5 Febrero                  953
##  6 Julio                    800
##  7 Junio                   639
##  8 Marzo                    992
##  9 Mayo                     812
## 10 Noviembre               1131
## 11 Octubre                 1050
## 12 Septiembre              916
```

Hacemos un Join de Robo con Moto y días de lluvia

```
Robo_2018_Moto_lluvia <- Robo_2018_Moto_lluvia %>%
  left_join(Lluvia_x_mes, by = "MES")
```

```
Robo_2018_Moto_lluvia
```

```
## # A tibble: 12 x 5
##   MES      Total_Robo_en_Moto  Año    mm  DIAS
##   <fct>          <int> <int> <dbl> <int>
##  1 Abril            787  2018  264.    11
##  2 Agosto           890  2018   54.3    8
##  3 Diciembre       1271  2018  314.   10
##  4 Enero          1030  2018   49.3    7
##  5 Febrero          953  2018   39.8    6
##  6 Julio            800  2018  125.   12
##  7 Junio           639  2018    6.6    4
##  8 Marzo           992  2018   70.1    6
##  9 Mayo            812  2018  217.   14
## 10 Noviembre       1131  2018  190.    8
## 11 Octubre        1050  2018   32    10
## 12 Septiembre      916  2018  112.    8
```

Veo si hay una correlación entre la cantidad de Robos y la cantidad de Días que llovió

```
Robo_2018_Moto_lluvia <- lm(data = Robo_2018_Moto_lluvia,
  Total_Robo_en_Moto ~ DIAS) # usamos un modelo global para ver cuánto hay
  de correlación entre Robo con moto y días de lluvia
```

```
Robo_2018_Moto_lluvia
```

```
##
## Call:
## lm(formula = Total_Robo_en_Moto ~ DIAS, data = Robo_2018_Moto_lluvia)
##
## Coefficients:
```

```
## (Intercept)      DIAS
##    930.9417      0.9586

summary(Robo_2018_Moto_lluvia) # para ver las estrellas "*"

##
## Call:
## lm(formula = Total_Robo_en_Moto ~ DIAS, data = Robo_2018_Moto_lluvia)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -295.78 -134.88   -3.15    96.63   330.47
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  930.9417   173.1171   5.378 0.000311 ***
## DIAS          0.9586    19.0596   0.050 0.960876
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 179.5 on 10 degrees of freedom
## Multiple R-squared:  0.0002529, Adjusted R-squared:  -0.09972
## F-statistic: 0.00253 on 1 and 10 DF, p-value: 0.9609
```

No hay correlación entre estos datos, pero consideramos que la cantidad de datos no es suficiente para estar seguros de que no hay relación estrecha entre ambos factores.