COVID dataset analyzing

Cristina Díaz García, Beatriz Huertas Calvillo, Miguel Lillo del Barrio, María Vázquez Pendón

2020-06-28

Contents

Description

This is an analysis of a COVID dataset using different approaches.

Chapter 1

Visualization

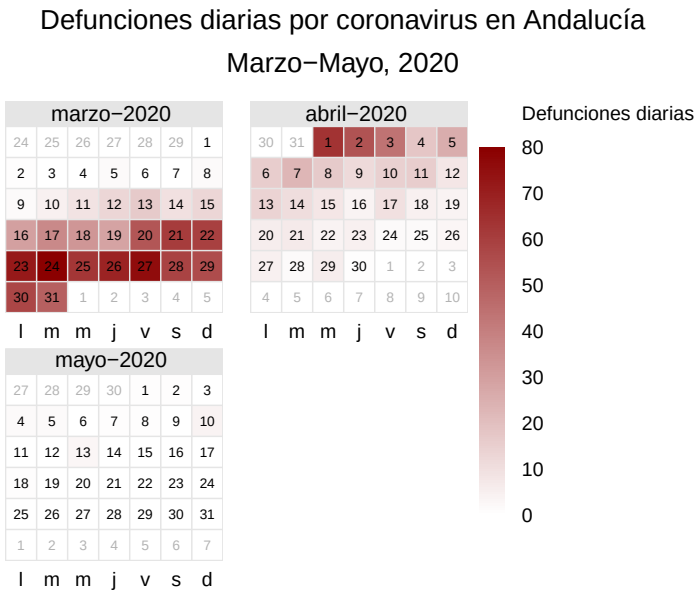
1.1 Death and Healed Calendar

```
library(readxl)
library(dplyr)
library(ggplot2)
library(lubridate)
library(openair)
library(lattice)
cs_export <- read_excel("cs_export.xls")
filasandalucia <- filter(cs_export, Territorio=="Andalucía" )
#View(filasandalucia)
colnames(filasandalucia)[1] <- "fecha"
##Calendario defunciones
filasandalucia$fecha <- as.factor(filasandalucia$fecha)
class(filasandalucia$fecha)
```

```
## [1] "factor"
```

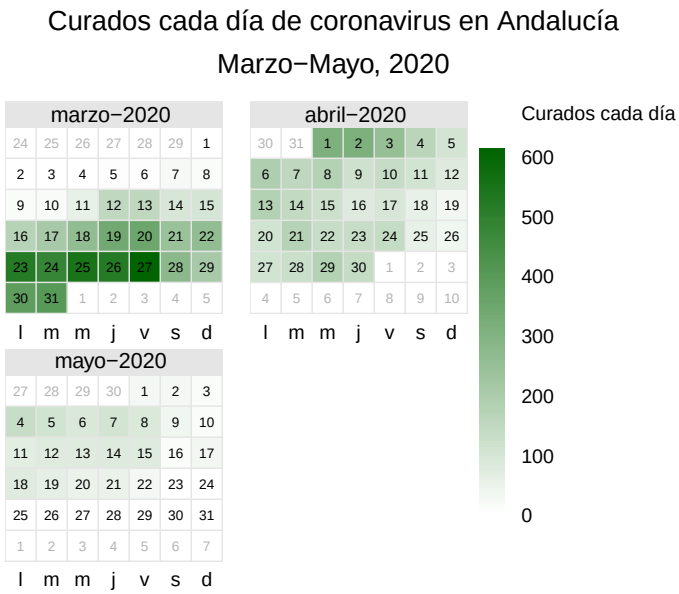
```
datos <- filasandalucia %>% dplyr::select(date = fecha, Defunciones) %>% mutate(date = as.Date(da
  # Quitar febrero
  filter(month(date) != 2)
datos <- as.data.frame(datos)
# Guardar calendario
#png("14.png", width = 10, height = 3.5, units = "in", res = 300)
calendarPlot(datos,
             pollutant = "Defunciones",
             # Título
```

```
            main = "Defunciones diarias por coronavirus en Andalucía \nMarzo-Mayo, 202
            # Para que el calendario empiece en lunes
            w.shift = 2,
            limits = c(0, max(datos$Defunciones)),
            # Colores para los eventos (del 0 al 3)
            cols = c("white", "darkred"),
            key.header = "Defunciones diarias")
```




```
#dev.off()
## Calendario curados
datos <- filasandalucia %>%
  dplyr::select(date = fecha, Curados) %>%
  mutate(date = as.Date(date,format="%d/%m/%Y")) %>%
  # Quitar febrero
  filter(month(date) != 2)
datos <- as.data.frame(datos)
#png("15.png", width = 10, height = 3.5, units = "in", res = 300)
calendarPlot(datos,
             pollutant = "Curados",
             # Título
             main = "Curados cada día de coronavirus en Andalucía \nMarzo-Mayo, 2020",
             # Para que el calendario empiece en lunes
             w.shift = 2,
             limits = c(0, max(datos$Curados)),
```

```
            # Colores para los eventos (del 0 al 3)
            cols = c("white", "darkgreen"),
            key.header = "Curados cada día")
```




```
#dev.off()
```

1.2 Confirmed cases bar graph

```
# Librerías a utilizar
library(tidyverse)
library(gganimate)
library(readxl)
# Cargo los datos a trabajar
cs_export <- read_excel("cs_export.xls") %>% print()
```

```
## # A tibble: 687 x 8
##    `Fecha declarac~ Territorio `Confirmados PC~ Hospitalizados   UCI Curados
##    <chr>            <chr>                  <dbl>          <dbl> <dbl>   <dbl>
##  1 22/05/2020       Andalucía                  9              6     0      26
##  2 22/05/2020       Almería                    1              2     0       4
```

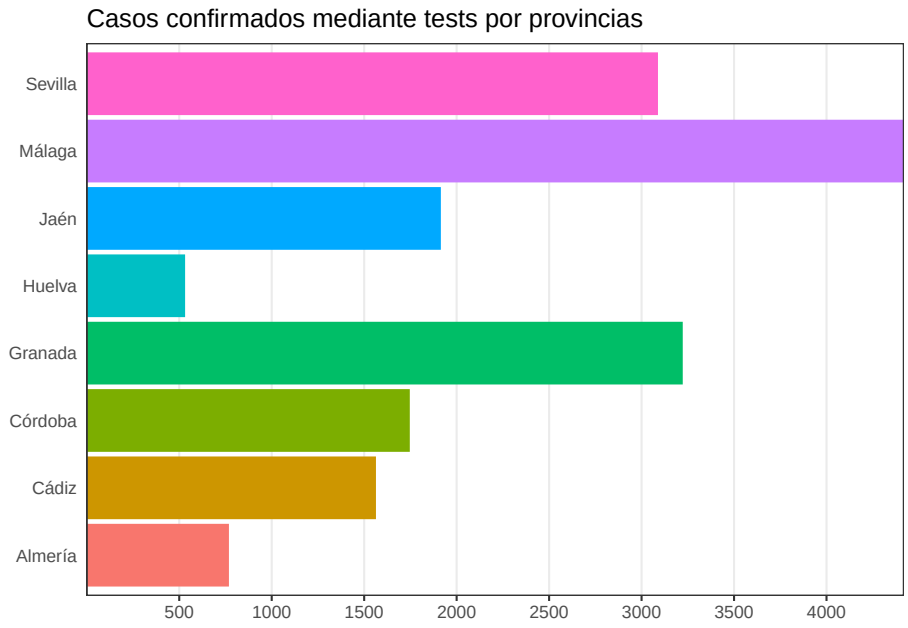
```
##  3 22/05/2020       Córdoba                   0              0     0       1
##  4 22/05/2020       Granada                   0              1     0      11
##  5 22/05/2020       Huelva                    1              1     0       0
##  6 22/05/2020       Jaén                      5              2     0       1
##  7 22/05/2020       Málaga                    2              0     0       5
##  8 22/05/2020       Sevilla                   0              0     0       4
##  9 21/05/2020       Andalucía                16              5     0      52
## 10 21/05/2020       Almería                   0              0     0       4
## # ... with 677 more rows, and 2 more variables: Defunciones <dbl>, `Total
## #   confirmados (PCR+test)` <dbl>
```

```
# proceso los datos a utlizar
confirmados <-
 cs_export %>%
  group_by(Territorio, `Fecha declaración`)%>%
  print()
```

```
## # A tibble: 687 x 8
## # Groups:   Territorio, Fecha declaración [686]
##    `Fecha declarac~ Territorio `Confirmados PC~ Hospitalizados   UCI Curados
##    <chr>            <chr>                 <dbl>          <dbl> <dbl>   <dbl>
##  1 22/05/2020       Andalucía                 9              6     0      26
##  2 22/05/2020       Almería                   1              2     0       4
##  3 22/05/2020       Córdoba                   0              0     0       1
##  4 22/05/2020       Granada                   0              1     0      11
##  5 22/05/2020       Huelva                    1              1     0       0
##  6 22/05/2020       Jaén                      5              2     0       1
##  7 22/05/2020       Málaga                    2              0     0       5
##  8 22/05/2020       Sevilla                   0              0     0       4
##  9 21/05/2020       Andalucía                16              5     0      52
## 10 21/05/2020       Almería                   0              0     0       4
## # ... with 677 more rows, and 2 more variables: Defunciones <dbl>, `Total
## #   confirmados (PCR+test)` <dbl>
```

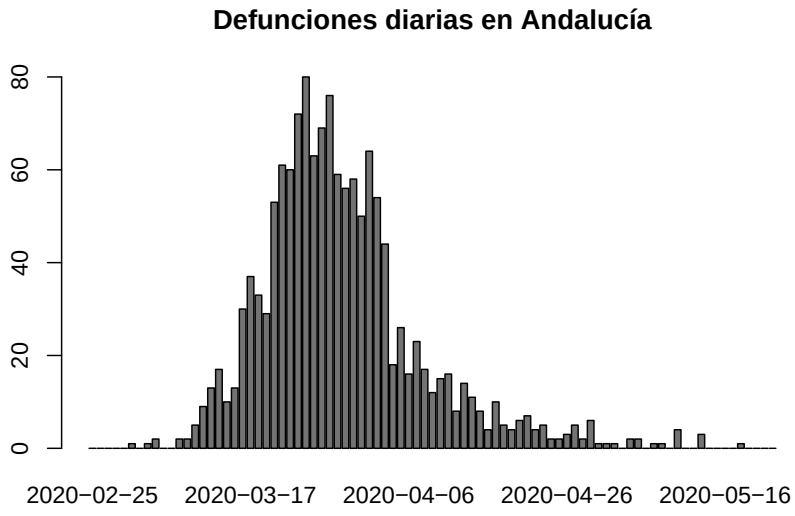
```
almeria <- confirmados %>% filter(Territorio=="Almería")
almeriaacum <- colSums(almeria[3:8])
cadiz <- confirmados %>% filter(Territorio=="Cádiz")
cadizacum <- colSums(cadiz[3:8])
cordoba <- confirmados %>% filter(Territorio=="Córdoba")
cordobaacum <- colSums(cordoba[3:8])
granada <- confirmados %>% filter(Territorio=="Granada")
granadaacum <- colSums(granada[3:8])
huelva <- confirmados %>% filter(Territorio=="Huelva")
huelvaacum <- colSums(huelva[3:8])
jaen <- confirmados %>% filter(Territorio=="Jaén")
```

```
jaenacum <- colSums(jaen[3:8])
malaga <- confirmados %>% filter(Territorio=="Málaga")
malagaacum <- colSums(malaga[3:8])
sevilla <- confirmados %>% filter(Territorio=="Sevilla")
sevillaacum <- colSums(sevilla[3:8])
dfaux <- data.frame("Provincia"=c("Sevilla","Málaga","Jaén","Huelva","Granada","Córdoba","Cádiz",
# genero el gráfico estático
plot_conf <-
  ggplot(dfaux,
         aes(x = dfaux$ConfirmadosAcum,
             y = dfaux$Provincia,
             colour = as.factor(dfaux$Provincia),
             fill = as.factor(dfaux$Provincia))) +
  geom_bar(stat = "identity",position="stack") +
  scale_x_continuous(breaks = seq(500, 5000, 500), expand = c(0,0)) +
  theme_bw() +
  theme(axis.title = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = "none",
        panel.grid.minor = element_blank(),
        panel.grid.major.y = element_blank())+ggtitle("Casos confirmados mediante tests por provi
plot_conf
```

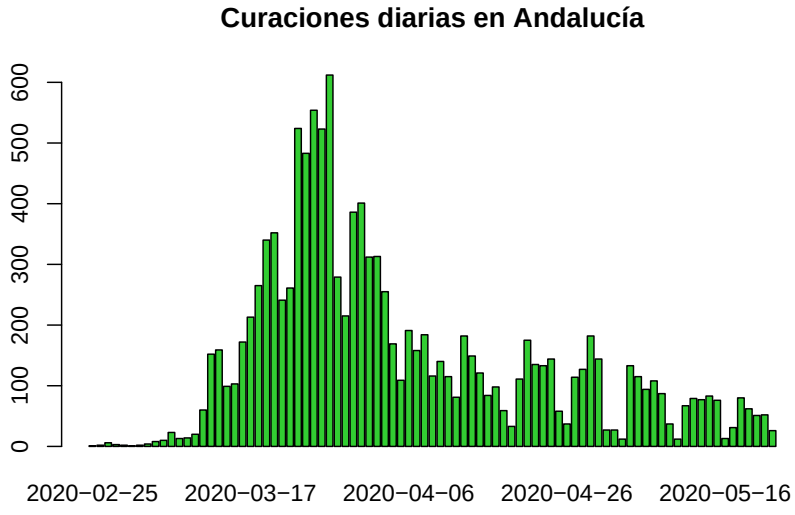


1.3 Andalusia bar graphs

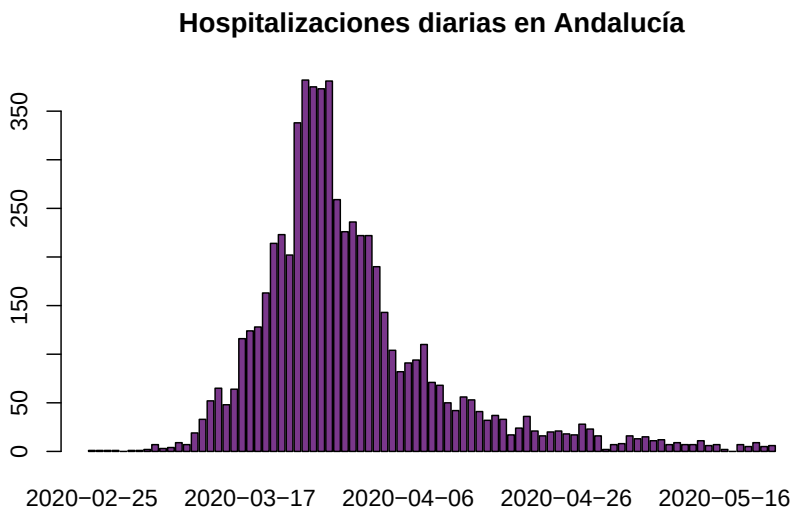
```
library(readxl)
library(dplyr)
library(ggplot2)
library(lubridate)
library(openair)
library(lattice)
cs_export <- read_excel("cs_export.xls")
filasandalucia <- filter(cs_export, Territorio=="Andalucía" )
aux <- filasandalucia
fechas <- as.Date(aux$`Fecha declaración`,"%d/%m/%Y")
aux$`Fecha declaración` <- sort(fechas)
c <- aux$Curados
h <- aux$Hospitalizados
d <- aux$Defunciones
uci <- aux$UCI
conf <- aux$`Confirmados PCR`
totalconf <- aux$`Total confirmados (PCR+test)`
salidac <- vector("numeric",length(c))
salidah <-vector("numeric",length(h))
salidad <- vector("numeric",length(d))
salidauci <- vector("numeric",length(uci))
salidaconf <- vector("numeric",length(conf))
salidatotalconf <- vector("numeric",length(totalconf))
for(i in seq_along(c)){

  salidac[length(c)+1-i] <- c[i]
  salidah[length(h)+1-i] <- h[i]
  salidad[length(d)+1-i] <- d[i]
  salidauci[length(uci)+1-i] <- uci[i]
  salidaconf[length(conf)+1-i] <- conf[i]
  salidatotalconf[length(totalconf)+1-i] <- totalconf[i]

}
aux$Curados <- salidac
aux$Hospitalizados <- salidah
aux$Defunciones <- salidad
aux$UCI <- salidauci
aux$`Confirmados PCR` <- salidaconf
aux$`Total confirmados (PCR+test)` <- salidatotalconf
barplot(names.arg=aux$`Fecha declaración`,aux$Defunciones,main="Defunciones diarias en
```



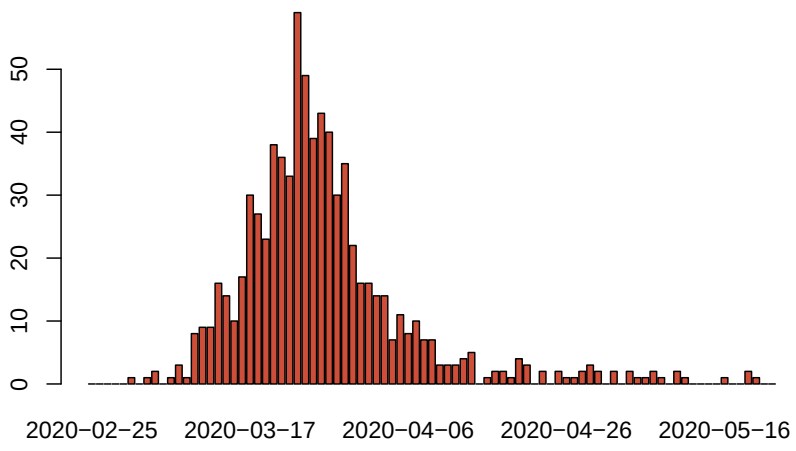

```
barplot(names.arg=aux$`Fecha declaración`,aux$Curados,main="Curaciones diarias en Andalucía",col=
```



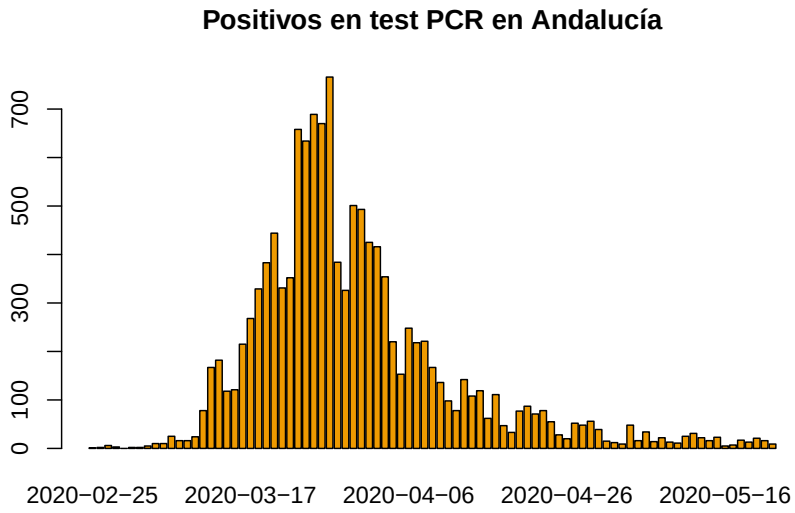
```
barplot(names.arg=aux$`Fecha declaración`,aux$Hospitalizados,main="Hospitalizaciones di
```



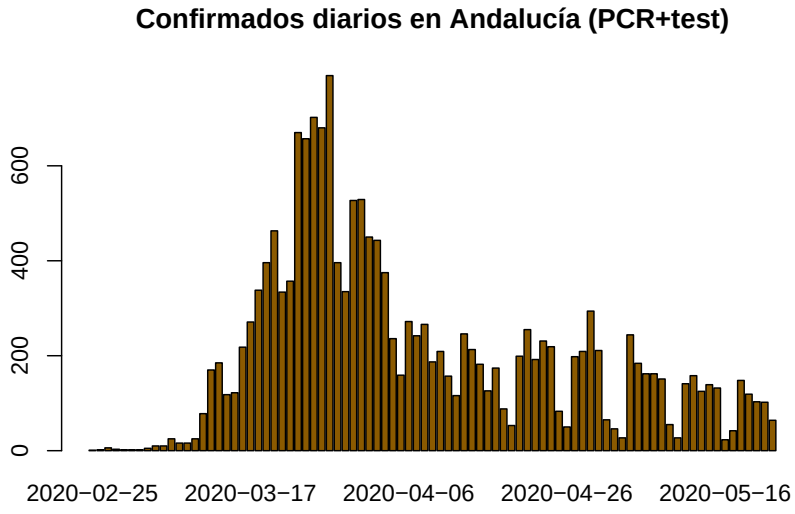

```
barplot(names.arg=aux$`Fecha declaración`,aux$UCI,main="Ingresos diarios en UCI en Anda
```



```
barplot(names.arg=aux$`Fecha declaración`,aux$`Confirmados PCR`,main="Positivos en test PCR en An
```




```
barplot(names.arg=aux$`Fecha declaración`,aux$`Total confirmados (PCR+test)`,ylim=c(0,max(aux$`To
```



1.4 Map of Andalusia

```
# para manipular dataframes
library(tidyverse)
# para importar archivos shapefiles y excel
library(rgdal)
library(readxl)
# Para transformar los archivos shapefiles
library(broom)
library(ggplot2)
library(plotly)
# Guardamos el archivo shapefile
shapefile_provincias <- rgdal::readOGR("Provincias_ETRS89_30N/Provincias_ETRS89_30N.sh
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "C:\Users\Beatriz Huertas\Desktop\3º Ingeniería informática\2º cuatrimestre
## with 52 features
## It has 5 fields
```

```
# Para convertir el archivo shapefile en un dataframe utilizamos la función tidy()
data_provincias <- tidy(shapefile_provincias)
nombres_provincias <- data.frame(shapefile_provincias$Texto)
nombres_provincias$id <- as.character(seq(0, nrow(nombres_provincias)-1))
head(nombres_provincias)
```

```
##   shapefile_provincias.Texto id
## 1                     Ã\201lava  0
## 2                    Albacete  1
## 3                    Alicante  2
## 4                    AlmerÃa  3
## 5                      Ã\201vila  4
## 6                     Badajoz  5
```

```
data_provincias_mapa <- left_join(data_provincias, nombres_provincias, by = "id")
reemplazos <-cbind(data_provincias,gsub("AlmerÃ a","Almeria", data_provincias_mapa$shap
colnames(reemplazos)[8] <- "Provincias"
reemplazos$Provincias <- as.character(reemplazos$Provincias)
reemplazos$Provincias[2443:3298] <- "Almeria"
reemplazos$Provincias <- as.factor(reemplazos$Provincias)
provinciasandalucia <- filter(reemplazos,reemplazos$Provincias %in% c("Sevilla","Huelv
cs_export <- read_excel("cs_export.xls") %>% print()
```

```
## # A tibble: 687 x 8
```

```
##    `Fecha declarac~ Territorio `Confirmados PC~ Hospitalizados   UCI Curados
##    <chr>            <chr>                 <dbl>          <dbl> <dbl>   <dbl>
##  1 22/05/2020       Andalucía                 9              6     0      26
##  2 22/05/2020       Almería                   1              2     0       4
##  3 22/05/2020       Córdoba                   0              0     0       1
##  4 22/05/2020       Granada                   0              1     0      11
##  5 22/05/2020       Huelva                    1              1     0       0
##  6 22/05/2020       Jaén                      5              2     0       1
##  7 22/05/2020       Málaga                    2              0     0       5
##  8 22/05/2020       Sevilla                   0              0     0       4
##  9 21/05/2020       Andalucía                16              5     0      52
## 10 21/05/2020       Almería                   0              0     0       4
## # ... with 677 more rows, and 2 more variables: Defunciones <dbl>, `Total
## #   confirmados (PCR+test)` <dbl>
```

```
# proceso los datos a utlizar
confirmados <-
  cs_export %>%
  group_by(Territorio, `Fecha declaración`)%>%
  print()
```

```
## # A tibble: 687 x 8
## # Groups:   Territorio, Fecha declaración [686]
##    `Fecha declarac~ Territorio `Confirmados PC~ Hospitalizados   UCI Curados
##    <chr>            <chr>                 <dbl>          <dbl> <dbl>   <dbl>
##  1 22/05/2020       Andalucía                 9              6     0      26
##  2 22/05/2020       Almería                   1              2     0       4
##  3 22/05/2020       Córdoba                   0              0     0       1
##  4 22/05/2020       Granada                   0              1     0      11
##  5 22/05/2020       Huelva                    1              1     0       0
##  6 22/05/2020       Jaén                      5              2     0       1
##  7 22/05/2020       Málaga                    2              0     0       5
##  8 22/05/2020       Sevilla                   0              0     0       4
##  9 21/05/2020       Andalucía                16              5     0      52
## 10 21/05/2020       Almería                   0              0     0       4
## # ... with 677 more rows, and 2 more variables: Defunciones <dbl>, `Total
## #   confirmados (PCR+test)` <dbl>
```

```
almeria <- confirmados %>% filter(Territorio=="Almería")
almeriaacum <- colSums(almeria[3:8])
cadiz <- confirmados %>% filter(Territorio=="Cádiz")
cadizacum <- colSums(cadiz[3:8])
cordoba <- confirmados %>% filter(Territorio=="Córdoba")
cordobaacum <- colSums(cordoba[3:8])
granada <- confirmados %>% filter(Territorio=="Granada")
```

```
granadaacum <- colSums(granada[3:8])
huelva <- confirmados %>% filter(Territorio=="Huelva")
huelvaacum <- colSums(huelva[3:8])
jaen <- confirmados %>% filter(Territorio=="Jaén")
jaenacum <- colSums(jaen[3:8])
malaga <- confirmados %>% filter(Territorio=="Málaga")
malagaacum <- colSums(malaga[3:8])
sevilla <- confirmados %>% filter(Territorio=="Sevilla")
sevillaacum <- colSums(sevilla[3:8])
dfaux <- data.frame("Provincia"=c("Sevilla","MÃ¡laga","JaÃ©n","Huelva","Granada","Córd
dfaux$id <- as.character(c(40,28,22,20,17,13,10,3))
confirmadosmapa <- provinciasandalucia %>%
  left_join(dfaux, by= "id")
mapa <- confirmadosmapa %>%
  ggplot(aes(x=long, y= lat, group = group)) +
  geom_polygon(aes(fill=Confirmados), color= "white", size = 0.2) +
  labs( title = "Tasa de Contagios por Provincia",
        fill = "") +
  theme_minimal() +
  theme(
    axis.line = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    plot.background = element_rect(fill = "snow", color = NA),
    panel.background = element_rect(fill= "snow", color = NA),
    plot.title = element_text(size = 16, hjust = 0),
    plot.subtitle = element_text(size = 12, hjust = 0),
    plot.caption = element_text(size = 8, hjust = 1),
    legend.title = element_text(color = "grey40", size = 8),
    legend.text = element_text(color = "grey40", size = 7, hjust = 0),
    legend.position = c(0.93, 0.3),
    plot.margin = unit(c(0.5,2,0.5,1), "cm")) +
  scale_fill_gradient(low = "yellow", high = "red")
ggplotly(mapa) %>%
  layout(title = 'Tasa de Contagios por Provincia')
```

```
## Hospitalizados
mapa <- confirmadosmapa %>%
  ggplot(aes(x=long, y= lat, group = group)) +
  geom_polygon(aes(fill=Hospitalizados), color= "white", size = 0.2) +
  labs( title = "Tasa de Hospitalizados por Provincia",
        fill = "") +
  theme_minimal() +
  theme(
```

```
    axis.line = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    plot.background = element_rect(fill = "snow", color = NA),
    panel.background = element_rect(fill= "snow", color = NA),
    plot.title = element_text(size = 16, hjust = 0),
    plot.subtitle = element_text(size = 12, hjust = 0),
    plot.caption = element_text(size = 8, hjust = 1),
    legend.title = element_text(color = "grey40", size = 8),
    legend.text = element_text(color = "grey40", size = 7, hjust = 0),
    legend.position = c(0.93, 0.3),
    plot.margin = unit(c(0.5,2,0.5,1), "cm")) +
  scale_fill_gradient(low = "green", high = "red")
ggplotly(mapa) %>%
  layout(title = 'Tasa de Hospitalizados por Provincia')
```

```
## Curados
mapa <- confirmadosmapa %>%
  ggplot(aes(x=long, y= lat, group = group)) +
  geom_polygon(aes(fill=Curados), color= "white", size = 0.2) +
  labs( title = "Tasa de Curados por Provincia",
        fill = "") +
  theme_minimal() +
  theme(
    axis.line = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    plot.background = element_rect(fill = "snow", color = NA),
    panel.background = element_rect(fill= "snow", color = NA),
    plot.title = element_text(size = 16, hjust = 0),
    plot.subtitle = element_text(size = 12, hjust = 0),
    plot.caption = element_text(size = 8, hjust = 1),
    legend.title = element_text(color = "grey40", size = 8),
    legend.text = element_text(color = "grey40", size = 7, hjust = 0),
    legend.position = c(0.93, 0.3),
    plot.margin = unit(c(0.5,2,0.5,1), "cm")) +
  scale_fill_gradient(low ="aquamarine", high = "darkblue")
ggplotly(mapa) %>%
  layout(title = 'Tasa de Curados por Provincia')
```

```
## Defunciones
mapa <- confirmadosmapa %>%
  ggplot(aes(x=long, y= lat, group = group)) +
```

```
  geom_polygon(aes(fill=Defunciones), color= "white", size = 0.2) +
  labs( title = "Tasa de Defunciones por Provincia",
        fill = "") +
  theme_minimal() +
  theme(
    axis.line = element_blank(),
    axis.text = element_blank(),
    axis.title = element_blank(),
    axis.ticks = element_blank(),
    plot.background = element_rect(fill = "snow", color = NA),
    panel.background = element_rect(fill= "snow", color = NA),
    plot.title = element_text(size = 16, hjust = 0),
    plot.subtitle = element_text(size = 12, hjust = 0),
    plot.caption = element_text(size = 8, hjust = 1),
    legend.title = element_text(color = "grey40", size = 8),
    legend.text = element_text(color = "grey40", size = 7, hjust = 0),
    legend.position = c(0.93, 0.3),
    plot.margin = unit(c(0.5,2,0.5,1), "cm")) +
  scale_fill_gradient(low ="gray46", high = "gray8")
ggplotly(mapa) %>%
  layout(title = 'Tasa de Defunciones por Provincia')
```

Chapter 2

Association Rules

```
library(readxl)
library(dplyr)
library(arules)
library(arulesViz)
datos <- read_excel("cs_export.xls")
#View(head(datos))
datos <- na.omit(datos)
```

We split the data of each province from those of Andalusia

```
datos$`Fecha declaración` <- as.Date(datos$`Fecha declaración`, "%d/%m/%Y")
datos$Territorio <- as.factor(datos$Territorio)
filasandalucia <- filter(datos, Territorio=="Andalucía" )
provincias <- setdiff(datos,filasandalucia)
nrow(provincias)
```

```
## [1] 687
```

```
nrow(filasandalucia)
```

```
## [1] 87
```

When working with numerical data the first step we have to carry out is to discretize them. The ICU and Death columns are very skewed so we use a different method.

```
summary(provincias$UCI)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00    2.23    2.00   59.00
```

```
summary(provincias$Defunciones)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   1.000   4.044   4.000  80.000
```

```
for(i in c(3,4,6,8)){
  provincias[[i]] <- discretize(provincias[[i]],breaks = 4, labels = c("Bajo","Normal"
}
provincias$UCI <- ordered(cut(provincias$UCI, c(-1,40,60),
  labels = c("Bajo", "Alto")))
provincias$Defunciones <-  ordered(cut(provincias$Defunciones, c(-1,5,80),
  labels = c("Bajo", "Alto")))
for(i in c(3,4,6,8)){
  filasandalucia[[i]] <- discretize(filasandalucia[[i]],breaks = 4, labels = c("Bajo",'
}
filasandalucia$UCI <- ordered(cut(filasandalucia$UCI, c(-1,40,60),
  labels = c("Bajo", "Alto")))
filasandalucia$Defunciones <-  ordered(cut(filasandalucia$Defunciones, c(-1,5,80),
  labels = c("Bajo", "Alto")))
```

We apply the *apriori* algorithm to the provinces dataset

```
reglas <- apriori(provincias[3:length(provincias)],parameter=list(supp=0.05,conf=0.5, m
```

```
## Apriori
##
## Parameter specification:
##  confidence minval smax arem  aval originalSupport maxtime support minlen
##         0.5    0.1    1 none FALSE            TRUE       5    0.05      2
##  maxlen target   ext
##      10  rules FALSE
##
## Algorithmic control:
##  filter tree heap memopt load sort verbose
##     0.1 TRUE TRUE  FALSE TRUE    2    TRUE
##
## Absolute minimum support count: 34
##
```

```
## set item appearances ...[0 item(s)] done [0.00s].
## set transactions ...[20 item(s), 687 transaction(s)] done [0.00s].
## sorting and recoding items ... [19 item(s)] done [0.00s].
## creating transaction tree ... done [0.00s].
## checking subsets of size 1 2 3 4 5 6 done [0.00s].
## writing ... [833 rule(s)] done [0.00s].
## creating S4 object  ... done [0.00s].
```

```
reglas <- sort(reglas,by="lift")
reglas <- reglas[!is.redundant(reglas)]
summary(reglas)
```

```
## set of 183 rules
##
## rule length distribution (lhs + rhs):sizes
##  2  3  4  5
## 74 79 26  4
##
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   2.000   3.000   2.781   3.000   5.000
##
## summary of quality measures:
##     support          confidence          lift            count
##  Min.   :0.05095   Min.   :0.5068   Min.   :0.9801   Min.   : 35.0
##  1st Qu.:0.10771   1st Qu.:0.7183   1st Qu.:2.0605   1st Qu.: 74.0
##  Median :0.15138   Median :0.8111   Median :2.9720   Median :104.0
##  Mean   :0.16187   Mean   :0.8128   Mean   :2.7265   Mean   :111.2
##  3rd Qu.:0.19505   3rd Qu.:0.9606   3rd Qu.:3.5112   3rd Qu.:134.0
##  Max.   :0.81951   Max.   :1.0000   Max.   :4.4593   Max.   :563.0
##
## mining info:
##                              data ntransactions support confidence
##  provincias[3:length(provincias)]           687    0.05        0.5
```

In principle, we will try to predict the behavior of the *Hospitalized* variable based on, for example, *Total confirmed (PCR + test)*

```
#reglas_3 <- reglas[which(size(reglas)==3)]
#inspect(head(reglas_3))
s1 <- subset(reglas,subset=lhs %pin% "Total confirmados")
inspect(head(s1))
```

```
##     lhs                                        rhs                  support confidence     li
## [1] {Confirmados PCR=Muy alto,
```

```
##      Hospitalizados=Muy alto,
##      Curados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.14410480  0.8048
## [2] {Confirmados PCR=Muy alto,
##      Hospitalizados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.15429403  0.8030
## [3] {Hospitalizados=Muy alto,
##      Curados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.14410480  0.7983
## [4] {Hospitalizados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.15429403  0.7910
## [5] {Confirmados PCR=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.15574964  0.7697
## [6] {Confirmados PCR=Bajo,
##      Hospitalizados=Normal,
##      Total confirmados (PCR+test)=Bajo}     => {Curados=Bajo}     0.06695779  0.9387
```

```
s1 <- subset(s1,subset=rhs %pin% "Hospitalizados")
inspect(head(s1))
```

```
##     lhs                                          rhs                           support
## [1] {Defunciones=Alto,
##      Total confirmados (PCR+test)=Muy alto} => {Hospitalizados=Muy alto} 0.15429403
## [2] {Confirmados PCR=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Hospitalizados=Muy alto} 0.19213974
## [3] {Confirmados PCR=Bajo,
##      Total confirmados (PCR+test)=Bajo}     => {Hospitalizados=Bajo}     0.08733624
## [4] {Curados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Hospitalizados=Muy alto} 0.18049491
## [5] {Total confirmados (PCR+test)=Muy alto} => {Hospitalizados=Muy alto} 0.19505095
## [6] {Confirmados PCR=Alto,
##      Curados=Alto,
##      Defunciones=Bajo,
##      Total confirmados (PCR+test)=Alto}     => {Hospitalizados=Alto}     0.08588064
```

We can see a clear relationship between the number of confirmed and the number of hospitalized.

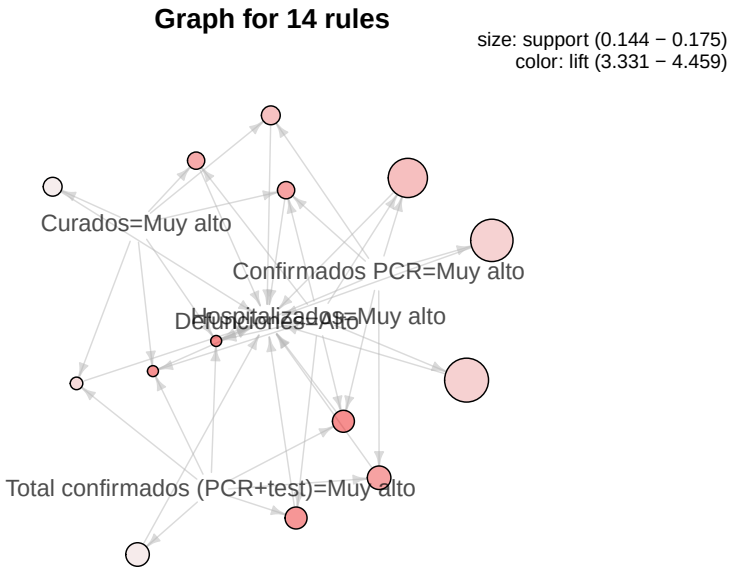
In the following way we look for implications that on the right side have the variable *Defunctions*

```
s2 <- subset(reglas,subset=rhs %pin% "Defunciones")
inspect(head(s2))
```

```
##     lhs                                          rhs                  support confide
```

```
## [1] {Confirmados PCR=Muy alto,
##      Hospitalizados=Muy alto,
##      Curados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.1441048  0.8048780 4.45928
## [2] {Confirmados PCR=Muy alto,
##      Hospitalizados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.1542940  0.8030303 4.44904
## [3] {Hospitalizados=Muy alto,
##      Curados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.1441048  0.7983871 4.42332
## [4] {Hospitalizados=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.1542940  0.7910448 4.38264
## [5] {Confirmados PCR=Muy alto,
##      Total confirmados (PCR+test)=Muy alto} => {Defunciones=Alto} 0.1557496  0.7697842 4.26485
## [6] {Confirmados PCR=Muy alto,
##      Hospitalizados=Muy alto,
##      Curados=Muy alto}                      => {Defunciones=Alto} 0.1499272  0.7686567 4.25860
```

```
s2Alto <- subset(reglas,subset=rhs %pin% "Defunciones=Alto")
plot(s2Alto, method="graph")
```



As we can see, most of the rules give us very obvious information such as ** {Hospitalized = Very high, Total confirmed (PCR + test) = Very high} => {Deaths = High} ** It can also be interpreted as that there is a strong correlation between the different columns, but it is more efficient to apply regression

methods in these cases in which we have quantitative variables.

Chapter 3

fcaR

```
library("fcaR")
library("arules")
covid <- read.csv("COVID.csv", header = TRUE, sep = ",")
head(covid)
```

```
##   ï..Fecha.declaraciÃ³n Territorio Confirmados.PCR Hospitalizados UCI Curados
## 1                   87          1               9              6   0      26
## 2                   87          2               1              2   0       4
## 3                   87          3               0              0   0       1
## 4                   87          4               0              1   0      11
## 5                   87          5               1              1   0       0
## 6                   87          6               5              2   0       1
##   Defunciones Total.confirmados..PCR.test.
## 1           0                           64
## 2           0                            7
## 3           0                            1
## 4           0                           14
## 5           0                            1
## 6           0                            9
```

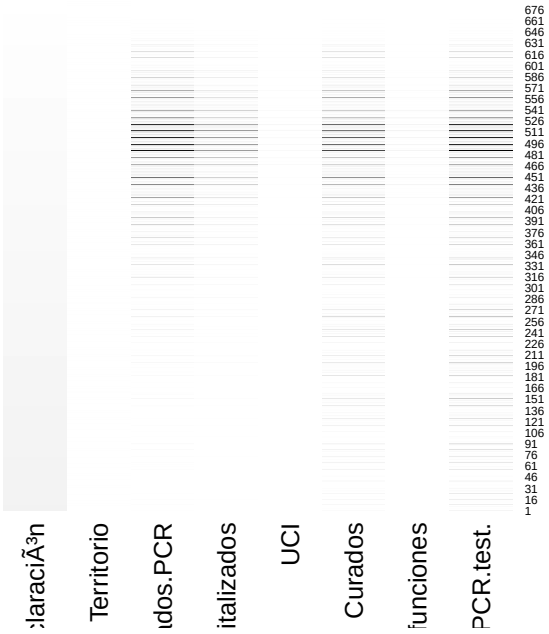
```
fc_covid <- FormalContext$new(covid)
print(fc_covid)
```

```
## Warning: Too many attributes, output will be truncated.
```

```
## FormalContext with 687 objects and 8 attributes.
## Attributes' names are: ï..Fecha.declaraciÃ³n, Territorio, Confirmados.PCR,
```

```
##   Hospitalizados, UCI, Curados, ...
## Matrix:
##      ï..Fecha.declaraciÃ³n Territorio Confirmados.PCR Hospitalizados UCI
## [1,]                    87          1               9               6   0
## [2,]                    87          2               1               2   0
## [3,]                    87          3               0               0   0
## [4,]                    87          4               0               1   0
## [5,]                    87          5               1               1   0
## [6,]                    87          6               5               2   0
##      Curados Defunciones
## [1,]      26           0
## [2,]       4           0
## [3,]       1           0
## [4,]      11           0
## [5,]       0           0
## [6,]       1           0
```

```
fc_covid$plot()
```




```
fc_covid$objects
```

```
##   [1] "1"   "2"   "3"   "4"   "5"   "6"   "7"   "8"   "9"   "10"  "11"  "12"
##  [13] "13"  "14"  "15"  "16"  "17"  "18"  "19"  "20"  "21"  "22"  "23"  "24"
##  [25] "25"  "26"  "27"  "28"  "29"  "30"  "31"  "32"  "33"  "34"  "35"  "36"
```

```
##  [37] "37"  "38"  "39"  "40"  "41"  "42"  "43"  "44"  "45"  "46"  "47"  "48"
##  [49] "49"  "50"  "51"  "52"  "53"  "54"  "55"  "56"  "57"  "58"  "59"  "60"
##  [61] "61"  "62"  "63"  "64"  "65"  "66"  "67"  "68"  "69"  "70"  "71"  "72"
##  [73] "73"  "74"  "75"  "76"  "77"  "78"  "79"  "80"  "81"  "82"  "83"  "84"
##  [85] "85"  "86"  "87"  "88"  "89"  "90"  "91"  "92"  "93"  "94"  "95"  "96"
##  [97] "97"  "98"  "99"  "100" "101" "102" "103" "104" "105" "106" "107" "108"
## [109] "109" "110" "111" "112" "113" "114" "115" "116" "117" "118" "119" "120"
## [121] "121" "122" "123" "124" "125" "126" "127" "128" "129" "130" "131" "132"
## [133] "133" "134" "135" "136" "137" "138" "139" "140" "141" "142" "143" "144"
## [145] "145" "146" "147" "148" "149" "150" "151" "152" "153" "154" "155" "156"
## [157] "157" "158" "159" "160" "161" "162" "163" "164" "165" "166" "167" "168"
## [169] "169" "170" "171" "172" "173" "174" "175" "176" "177" "178" "179" "180"
## [181] "181" "182" "183" "184" "185" "186" "187" "188" "189" "190" "191" "192"
## [193] "193" "194" "195" "196" "197" "198" "199" "200" "201" "202" "203" "204"
## [205] "205" "206" "207" "208" "209" "210" "211" "212" "213" "214" "215" "216"
## [217] "217" "218" "219" "220" "221" "222" "223" "224" "225" "226" "227" "228"
## [229] "229" "230" "231" "232" "233" "234" "235" "236" "237" "238" "239" "240"
## [241] "241" "242" "243" "244" "245" "246" "247" "248" "249" "250" "251" "252"
## [253] "253" "254" "255" "256" "257" "258" "259" "260" "261" "262" "263" "264"
## [265] "265" "266" "267" "268" "269" "270" "271" "272" "273" "274" "275" "276"
## [277] "277" "278" "279" "280" "281" "282" "283" "284" "285" "286" "287" "288"
## [289] "289" "290" "291" "292" "293" "294" "295" "296" "297" "298" "299" "300"
## [301] "301" "302" "303" "304" "305" "306" "307" "308" "309" "310" "311" "312"
## [313] "313" "314" "315" "316" "317" "318" "319" "320" "321" "322" "323" "324"
## [325] "325" "326" "327" "328" "329" "330" "331" "332" "333" "334" "335" "336"
## [337] "337" "338" "339" "340" "341" "342" "343" "344" "345" "346" "347" "348"
## [349] "349" "350" "351" "352" "353" "354" "355" "356" "357" "358" "359" "360"
## [361] "361" "362" "363" "364" "365" "366" "367" "368" "369" "370" "371" "372"
## [373] "373" "374" "375" "376" "377" "378" "379" "380" "381" "382" "383" "384"
## [385] "385" "386" "387" "388" "389" "390" "391" "392" "393" "394" "395" "396"
## [397] "397" "398" "399" "400" "401" "402" "403" "404" "405" "406" "407" "408"
## [409] "409" "410" "411" "412" "413" "414" "415" "416" "417" "418" "419" "420"
## [421] "421" "422" "423" "424" "425" "426" "427" "428" "429" "430" "431" "432"
## [433] "433" "434" "435" "436" "437" "438" "439" "440" "441" "442" "443" "444"
## [445] "445" "446" "447" "448" "449" "450" "451" "452" "453" "454" "455" "456"
## [457] "457" "458" "459" "460" "461" "462" "463" "464" "465" "466" "467" "468"
## [469] "469" "470" "471" "472" "473" "474" "475" "476" "477" "478" "479" "480"
## [481] "481" "482" "483" "484" "485" "486" "487" "488" "489" "490" "491" "492"
## [493] "493" "494" "495" "496" "497" "498" "499" "500" "501" "502" "503" "504"
## [505] "505" "506" "507" "508" "509" "510" "511" "512" "513" "514" "515" "516"
## [517] "517" "518" "519" "520" "521" "522" "523" "524" "525" "526" "527" "528"
## [529] "529" "530" "531" "532" "533" "534" "535" "536" "537" "538" "539" "540"
## [541] "541" "542" "543" "544" "545" "546" "547" "548" "549" "550" "551" "552"
## [553] "553" "554" "555" "556" "557" "558" "559" "560" "561" "562" "563" "564"
## [565] "565" "566" "567" "568" "569" "570" "571" "572" "573" "574" "575" "576"
## [577] "577" "578" "579" "580" "581" "582" "583" "584" "585" "586" "587" "588"
```

```
## [589] "589" "590" "591" "592" "593" "594" "595" "596" "597" "598" "599" "600"
## [601] "601" "602" "603" "604" "605" "606" "607" "608" "609" "610" "611" "612"
## [613] "613" "614" "615" "616" "617" "618" "619" "620" "621" "622" "623" "624"
## [625] "625" "626" "627" "628" "629" "630" "631" "632" "633" "634" "635" "636"
## [637] "637" "638" "639" "640" "641" "642" "643" "644" "645" "646" "647" "648"
## [649] "649" "650" "651" "652" "653" "654" "655" "656" "657" "658" "659" "660"
## [661] "661" "662" "663" "664" "665" "666" "667" "668" "669" "670" "671" "672"
## [673] "673" "674" "675" "676" "677" "678" "679" "680" "681" "682" "683" "684"
## [685] "685" "686" "687"
```

We find the implications and the concepts.

```
S <- SparseSet$new(attributes = fc_covid$objects)
fc_covid$find_implications()
fc_covid$implications
```

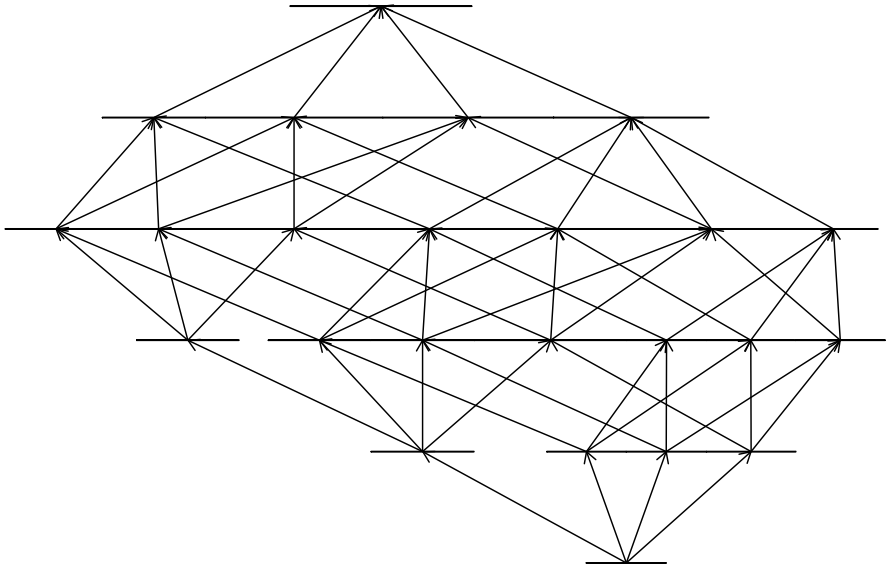
As we can see, there are no implications. Maybe this could be fixed discretizing the columns with data about the patients and healed.

```
fc_covid$find_concepts()
fc_covid$concepts
```

```
## A set of 24 concepts:
## 1: ({1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22,
## 2: ({33, 36, 74, 75, 78, 81, 101, 104, 105, 106, 107, 116, 123, 125, 129, 143, 144,
## 3: ({1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
## 4: ({33, 36, 74, 75, 78, 81, 101, 106, 107, 116, 123, 125, 129, 143, 144, 147, 152,
## 5: ({1, 2, 4, 5, 6, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 28, 30, 32, 3
## 6: ({33, 36, 74, 75, 78, 81, 101, 104, 105, 106, 107, 116, 123, 125, 129, 143, 144,
## 7: ({1, 2, 4, 6, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 28, 30, 32, 33,
## 8: ({33, 36, 74, 75, 78, 81, 101, 106, 107, 116, 123, 125, 129, 143, 144, 147, 152,
## 9: ({16, 21, 24, 26, 32, 49, 50, 92, 95, 101, 105, 106, 116, 124, 125, 128, 131, 134
## 10: ({101, 105, 106, 116, 125, 143, 144, 152, 158, 167, 171, 183, 189, 192, 196, 20
## 11: ({16, 21, 24, 26, 32, 49, 50, 92, 95, 101, 106, 116, 124, 125, 128, 131, 134, 14
## 12: ({101, 106, 116, 125, 143, 144, 152, 158, 167, 183, 189, 192, 196, 201, 205, 207
## 13: ({1, 2, 5, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 27,
## 14: ({33, 36, 74, 75, 78, 81, 101, 104, 105, 106, 107, 116, 123, 125, 129, 143, 147
## 15: ({1, 2, 6, 7, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 27, 28
## 16: ({33, 36, 74, 75, 78, 81, 101, 106, 107, 116, 123, 125, 129, 143, 147, 152, 156
## 17: ({1, 2, 5, 6, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 28, 30, 32, 33
## 18: ({33, 36, 74, 75, 78, 81, 101, 104, 105, 106, 107, 116, 123, 125, 129, 143, 147
## 19: ({1, 2, 6, 9, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 26, 28, 30, 32, 33, 36
## 20: ({33, 36, 74, 75, 78, 81, 101, 106, 107, 116, 123, 125, 129, 143, 147, 152, 156
## 21: ({16, 21, 24, 26, 32, 49, 92, 95, 101, 105, 106, 116, 124, 125, 128, 131, 134,
```

```
## 22: ({101, 105, 106, 116, 125, 143, 152, 158, 167, 171, 183, 189, 192, 196, 201, 205, 207, 210
## 23: ({16, 21, 24, 26, 32, 49, 92, 95, 101, 106, 116, 124, 125, 128, 131, 134, 140, 143, 152, 1
## 24: ({101, 106, 116, 125, 143, 152, 158, 167, 183, 189, 192, 196, 201, 205, 207, 210, 218, 219
```

```
fc_covid$concepts$plot()
```



In this last plot we can see it's a quite small lattice, and it would require a deeper analysis to extract some conclusions.

Chapter 4

Regression

```
library(readxl)
library(ggplot2)
library(dplyr)
library(magrittr)
library(ggplot2)
library(GGally)
```

We process the dataset to be able to work with it.

```
colnames(cs) <- c("Fecha", "Territorio", "Confirmados_PCR", "Hospitalizados", "UCI", "Curados", "
juntos <-
  filter(cs, cs$Territorio == "Andalucía")
juntos <- mutate(juntos, Total = cumsum(Total_confirmados))


c <- juntos$Curados
h <- juntos$Hospitalizados
d <- juntos$Defunciones
uci <- juntos$UCI
conf <- juntos$Confirmados_PCR
totalconf <- juntos$Total_confirmados
salidac <- vector("numeric",length(c))
salidah <-vector("numeric",length(h))
salidad <- vector("numeric",length(d))
salidauci <- vector("numeric",length(uci))
salidaconf <- vector("numeric",length(conf))
salidatotalconf <- vector("numeric",length(totalconf))
for(i in seq_along(c)){
```

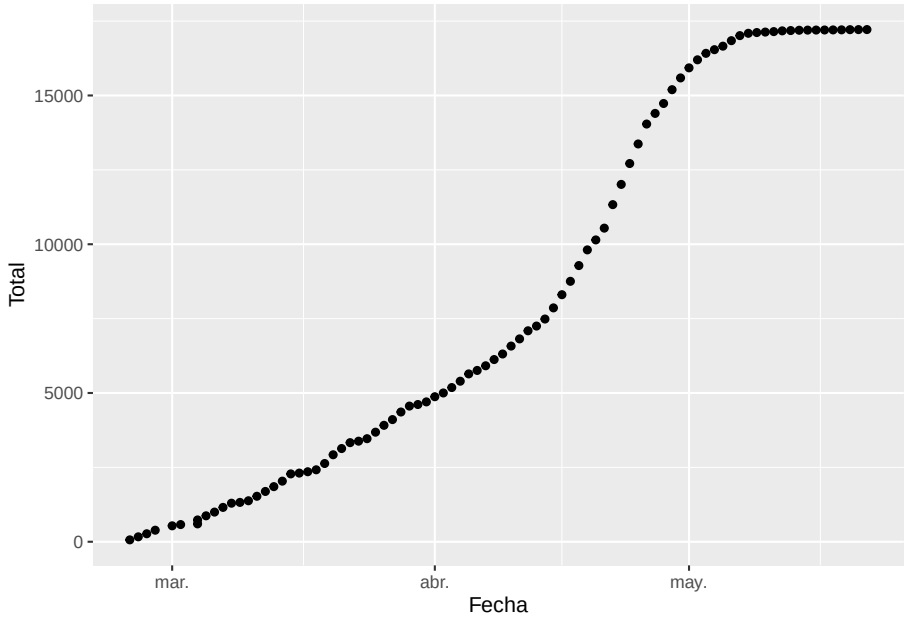
```
  salidac[length(c)+1-i] <- c[i]
  salidah[length(h)+1-i] <- h[i]
  salidad[length(d)+1-i] <- d[i]
  salidauci[length(uci)+1-i] <- uci[i]
  salidaconf[length(conf)+1-i] <- conf[i]
  salidatotalconf[length(totalconf)+1-i] <- totalconf[i]

}
juntos$Curados <- salidac
juntos$Hospitalizados <- salidah
juntos$Defunciones <- salidad
juntos$UCI <- salidauci
juntos$Confirmados_PCR <- salidaconf
juntos$Total_confirmados <- salidatotalconf

  fechas <- as.Date(juntos$Fecha ,"%d/%m/%Y")
juntos$Fecha <- sort(fechas)
```

We make the graph of how the cases have increased as a function of time.

```
attach(juntos)
```

```
## The following objects are masked from juntos (pos = 3):
##
##     Confirmados_PCR, Curados, Defunciones, Fecha, Hospitalizados,
##     Territorio, Total, Total_confirmados, UCI
```

```
grafica <- juntos %>% ggplot(aes(x= Fecha, y= Total))+geom_point()
grafica
```



In this graph we can see how the number of cases up to May increased exponentially. However, starting from this date, the number of cases per day begins to decrease, reaching the famous peak.

4.1 Model of the number of infections with respect to the date

```
modFC <- lm(formula = juntos$Total ~ juntos$Fecha, data = juntos)
modFC
```

```
##
## Call:
## lm(formula = juntos$Total ~ juntos$Fecha, data = juntos)
##
## Coefficients:
##  (Intercept)  juntos$Fecha
##   -4444803.2         242.5
```

```
summary(modFC)
```

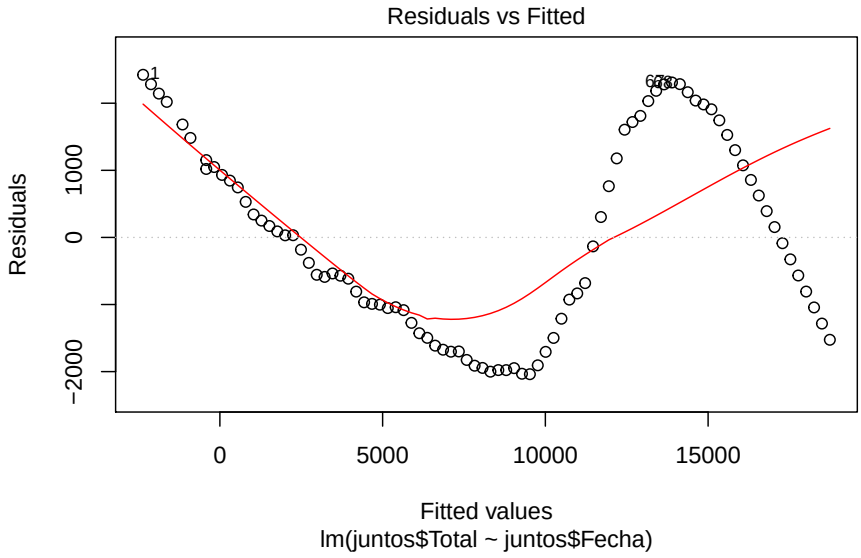
```
##
```

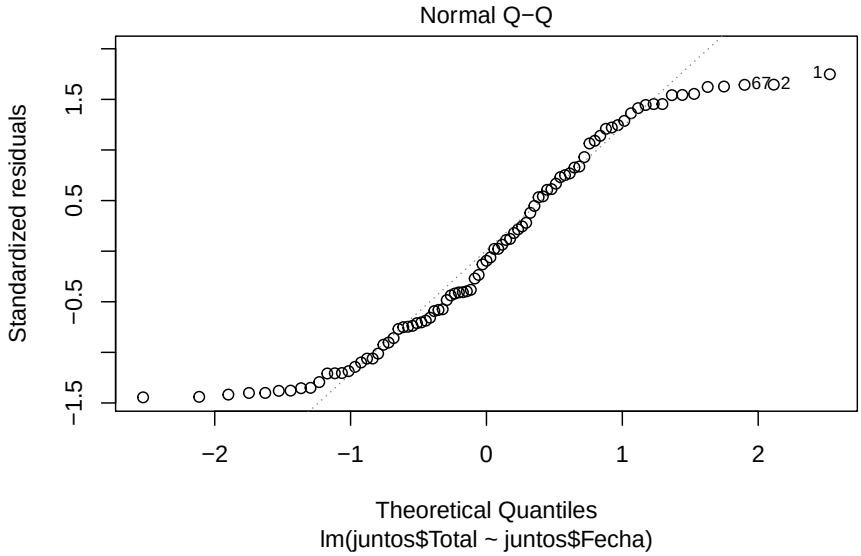
```
## Call:
## lm(formula = juntos$Total ~ juntos$Fecha, data = juntos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2038.5 -1147.8  -133.7  1165.0  2423.5
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -4.445e+06  1.110e+05  -40.03   <2e-16 ***
## juntos$Fecha  2.425e+02  6.047e+00   40.11   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1420 on 85 degrees of freedom
## Multiple R-squared:  0.9498, Adjusted R-squared:  0.9492
## F-statistic:  1609 on 1 and 85 DF,  p-value: < 2.2e-16
```

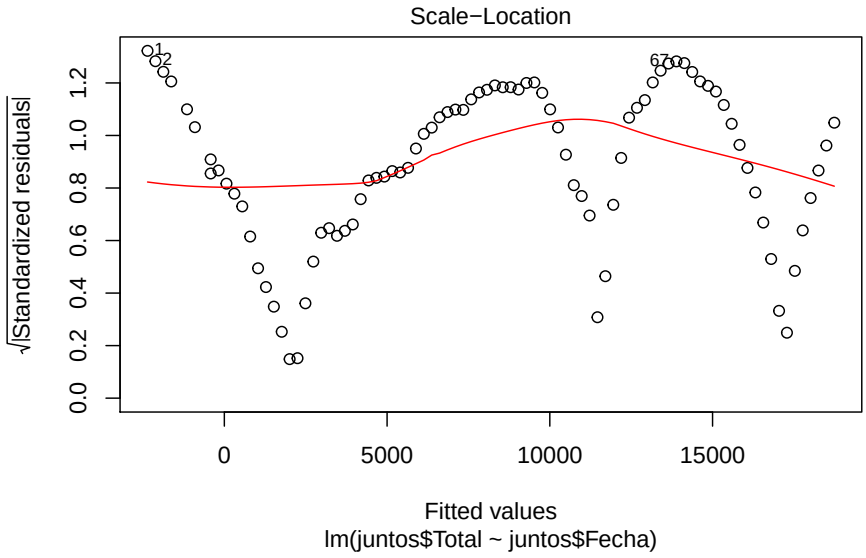
We start by checking if there is a model, for this we look at the values of F-statistic and p-value. ** F-statistic ** is quite far from 1 (1609), so it indicates that there is a model. ** P-value **, is well below 0.005, confirming that there is a model (H1 is met) Now that we know there is a model, let's study how good our model is.

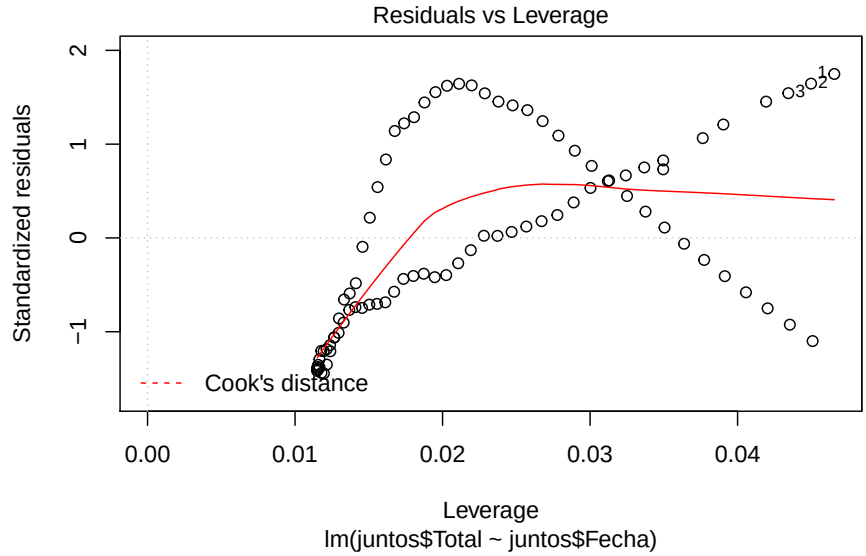
We look at R^2, which has a value of 0.9498, a pretty good value, since **94.98%** of the cases are collected with this model. We also see that there is almost no difference between adjusted R^2 and R^2, so *there is no overfitting* in our model and that the values are relevant (indicated by ***) Let's continue studying the model, for this we will see the graphs of the adjusted values and residuals.

```
plot(modFC)
```


Residuals vs Fitted
Residuals
1000
0
−2000
0
5000
10000
15000
Fitted values
lm(juntos$Total ~ juntos$Fecha)


Normal Q−Q
Standardized residuals
1.5
0.5
−0.5
−1.5
−2
−1
0
1
2
Theoretical Quantiles
lm(juntos$Total ~ juntos$Fecha)





In these graphs we can see how it begins adjusting to the values. However, as cases increase, there is more waste. This may be due to the results of the contingency measures that were taken. With all this we could say that we have a pretty good model. ## Model of the number of people who were hospitalized with respect to the number of infected.

```
mHospitalizados <- lm(formula = Hospitalizados ~ Total_confirmados, data = juntos)
mHospitalizados
```

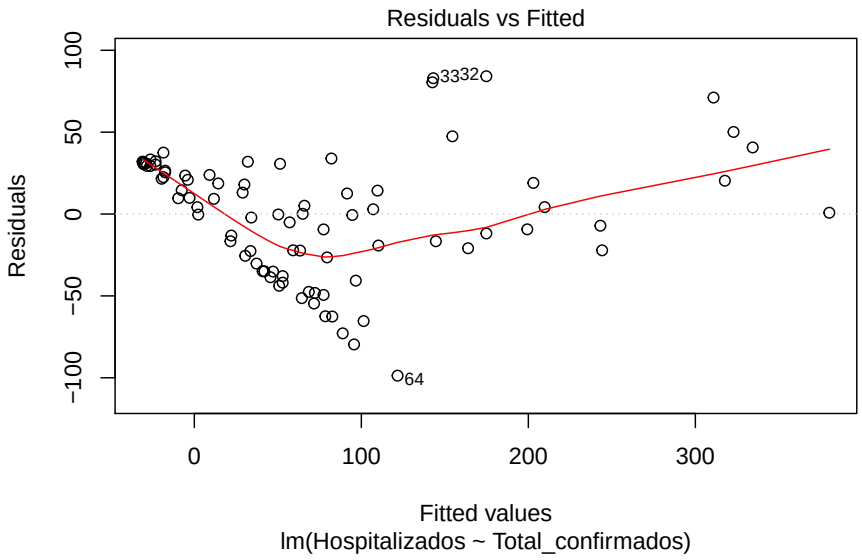
```
##
## Call:
## lm(formula = Hospitalizados ~ Total_confirmados, data = juntos)
##
## Coefficients:
##       (Intercept)  Total_confirmados
##          -31.5285             0.5212
```

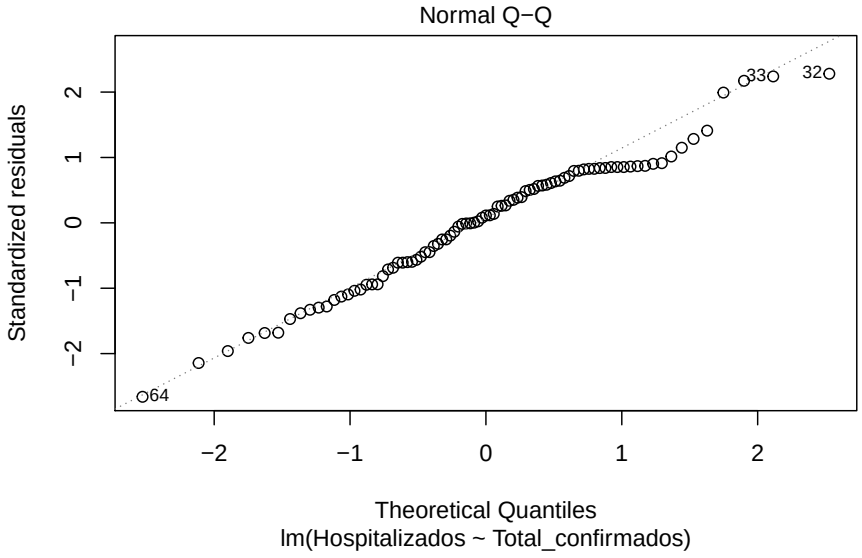
```
summary(mHospitalizados)
```

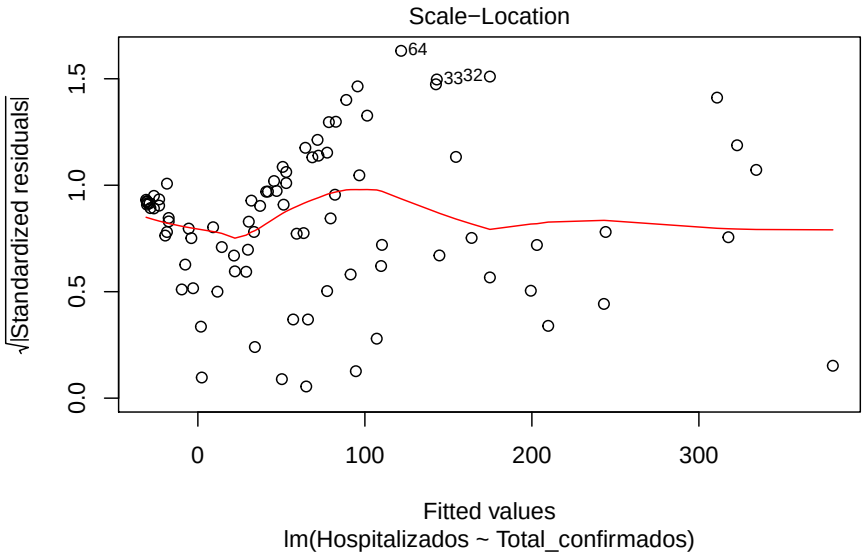
```
##
## Call:
## lm(formula = Hospitalizados ~ Total_confirmados, data = juntos)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -98.696 -24.055   4.173  29.359  84.144
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -31.52846    5.98438  -5.268 1.02e-06 ***
## Total_confirmados   0.52117    0.02246  23.200  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 37.37 on 85 degrees of freedom
## Multiple R-squared:  0.8636, Adjusted R-squared:  0.862
## F-statistic: 538.3 on 1 and 85 DF,  p-value: < 2.2e-16
```

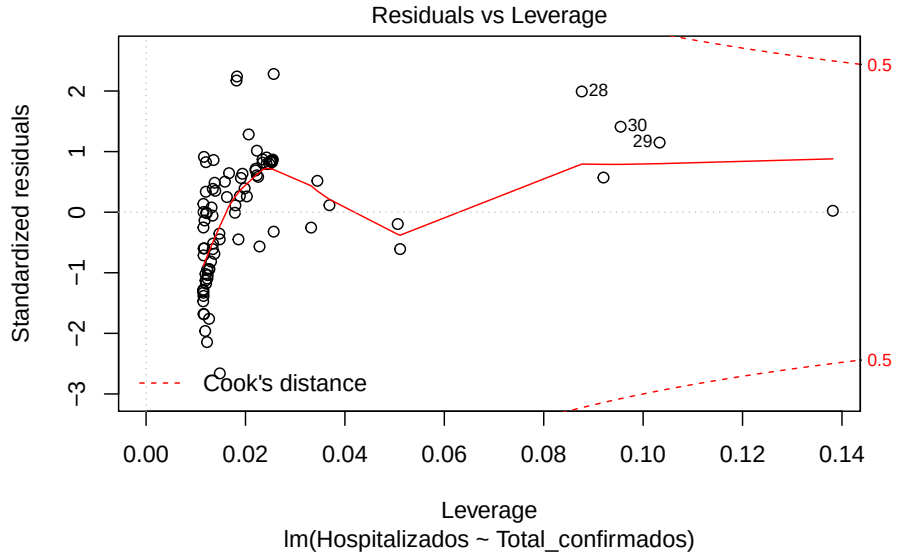
We start by checking if there is a model, for this we look at the values of F-statistic and p-value. F-statistic is quite far from 1 (538), so it indicates that there is a model. P-value, is well below 0.005, which confirms that there is a model (H1 is met) Now that we know there is a model, let's study how good our model is. We look at R^2, which has a value of 0.8636, a pretty good value, since 86.36% of hospitalization cases are included in this model. We also see that there is almost no difference between adjusted R^2 and R^2, so there is no overfitting in our model and that the values are relevant (indicated by ***) Let's continue studying the model, for this we will see the graphs of the adjusted values and residuals.

```
plot(mHospitalizados)
```









In the graphs we can see how there is dispersion of the values, and how it fits better at the beginning than at the end. We also observed that at the beginning of the period the number of hospitalized was very close to the number of cases. But as time progresses, the number of cases increases considerably, but not so much the number of hospitalized. This may be due to the fact that

the population was informed of the first symptoms and the measures they had to take, so that an infected person could be detected in the early stages and monitored so that the severity could be reduced.

```
library(readxl)
library(tidyverse)
datos <- read_excel("cs_export.xls")
datos <- na.omit(datos)
names(datos) <- c(names(datos[1:2]),"Confirmados_PCR",names(datos[4:7]),"Total_confirma
varPred <- names(datos[c(3:6,8)])
datos$`Fecha declaración` <- as.Date(datos$`Fecha declaración`, "%d/%m/%Y")
datos <- arrange(datos, `Fecha declaración`)
filasandalucia <- filter(datos, Territorio=="Andalucía" )
```

We are going to produce a multivariate regression model that explains the variable *Defunction* using an iterative technique in which we will add each time the variable whose R-adjusted is greater.

We define a function to calculate the linear model of a sum of variables

```
linearAdjust <- function(df, y, x){
  mod1 <- lm(str_c(y,"~",str_c(x,collapse="+")),df)
}
calcModR2 <- function(df,y,x){
  mod <- linearAdjust(df,y,x)
  summary(mod)$adj.r.squared
}
```

We are adding variables while increasing the value of adjusted R^2.

```
encontrarMejorAjuste <- function(df,varPos){
  bestVars <- character(0)
  aR2 <- 0
  repeat{
    aR2v <- map_dbl(varPos,~calcModR2(df, "Defunciones",c(bestVars,.)))
    i    <- which.max(aR2v)
    aR2M <- aR2v[i]

    if(aR2M<=aR2 ||length(varPos)<1) break
    #Valor del r-ajustado añadido y nombre de la variable elegida
    cat(sprintf("%1.4f %s\n",aR2M,varPos[i]))
    aR2 <- aR2M
    bestVars <- c(bestVars,varPos[i])
    varPos <- varPos[-i]
  }
```

```
  mod <- linearAdjust(df,"Defunciones",bestVars)
  list(vars=bestVars,mod=mod)
}
bestMod <- encontrarMejorAjuste(filasandalucia, varPred)
```

```
## 0.9499 Hospitalizados
## 0.9521 UCI
## 0.9530 Curados
## 0.9566 Total_confirmados
## 0.9642 Confirmados_PCR
```

```
bestMod
```

```
## $vars
## [1] "Hospitalizados"    "UCI"               "Curados"
## [4] "Total_confirmados" "Confirmados_PCR"
##
## $mod
##
## Call:
## lm(formula = str_c(y, "~", str_c(x, collapse = "+")), data = df)
##
## Coefficients:
##       (Intercept)     Hospitalizados                UCI            Curados
##           0.26562            0.11219            0.35878           -0.21053
## Total_confirmados    Confirmados_PCR
##           0.10853            0.08473
```

```
summary(bestMod$mod)
```

```
##
## Call:
## lm(formula = str_c(y, "~", str_c(x, collapse = "+")), data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -11.2602  -2.2208  -0.3433   1.8867  13.9259
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.26562    0.82376   0.322 0.747941
## Hospitalizados     0.11219    0.03151   3.561 0.000623 ***
## UCI                0.35878    0.10317   3.478 0.000816 ***
```

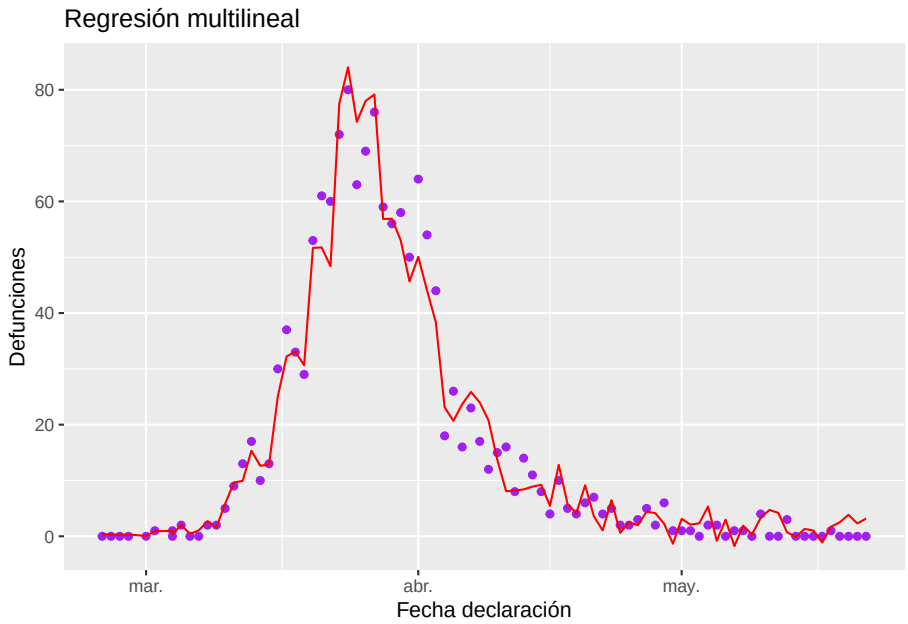
```
## Curados           -0.21053    0.03937  -5.347 8.07e-07 ***
## Total_confirmados  0.10853    0.02412   4.499 2.26e-05 ***
## Confirmados_PCR    0.08473    0.01978   4.284 5.02e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.285 on 81 degrees of freedom
## Multiple R-squared:  0.9663, Adjusted R-squared:  0.9642
## F-statistic: 463.9 on 5 and 81 DF,  p-value: < 2.2e-16
```

The model obtained would be of the form:

$$Defunciones = 0.2656235 + 0.1121855 \cdot Hospitalizados + 0.3587845 \cdot UCI + -0.2105297 \cdot ConfirmadosPC1$$

Finally we can represent the graph of the data (purple) with the regression obtained superimposed (red)

```
g <- ggplot(filasandalucia, aes(x=`Fecha declaración`,y=Defunciones))+
  ggtitle("Regresión multilineal")+
  geom_point(colour="purple")+
  geom_line(aes(`Fecha declaración`, predict.lm(bestMod$mod)), color="red")

g
```



Chapter 5

Final Words

As shown thorugh this book, we can find several approaches to analyze the same dataset. Some of them might work and show interesting results, but others might show nothing interesting - and even sometimes nothing at all!