

Practica 2

Jaime Pardo y Gabriel Paladines

5/30/2020

- 1 Descripción del dataset. ¿Por qué es importante y que pregunta/problema pretende responder?
- 2 Integración y selección de los datos de interés a analizar
- 3 Limpieza de los datos
 - 3.1 ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?
 - 3.2 Identificación y tratamiento de valores extremos.
- 4 Análisis de los datos.
 - 4.1 . Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).
 - 4.2 Comprobación de la normalidad y homogeneidad de la varianza.
 - 4.3 Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc.
Aplicar al menos tres métodos de análisis diferentes
- 5 Representación de los resultados a partir de tablas y gráficas.
- 6 Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?
- 7 Código: Hay que adjuntar el código, preferiblemente en R, con el que se ha realizado la limpieza, análisis y representación de los datos
- 8 OTROS
- 9 Contribuciones

1 Descripción del dataset. ¿Por qué es importante y que pregunta/problema pretende responder?

El dataset elegido (<https://www.kaggle.com/rushirdx/suicide-rates-from-1986-to-2016/data>) es importante porque contiene diversos parámetros que posiblemente tengan una correlación con la tasa de suicidio. Pretende responder a preguntas como las siguientes: ¿cómo es la evolución temporal de la tasa de suicidios? ¿se producen más suicidios entre los hombres que entre las mujeres? ¿hay una generación especialmente afectada por el suicidio? También a preguntas que combinen varios factores. Por ejemplo: ¿hay una diferencia importante entre la tasa de suicidios en hombres y en mujeres en una determinada generación?

Veamos a continuación los campos que contiene:

```
# Tipología y ciclo de vida de los datos - Práctica 2
# Autores: Gabriel Paladines y Jaime Pardo
# Dataset: https://www.kaggle.com/rushirdx/suicide-rates-from-1986-to-2016/data
library(dplyr)
library(knitr)
library(ggplot2)
library(stats)

# Lectura de datos
data <- read.csv("../data/suicide.csv", header = TRUE, sep = ",", quote="\"", dec=
".", fill = TRUE)
dim(data)
```

```
## [1] 27820    12
```

```
# Tenemos 27820 observaciones y un total de 12 variables. Entre otras:
# La variable sex tiene dos clases: female y male, ambas con la misma cantidad de
registros.
# La variable age tiene 6 clases, todas ellas con la misma cantidad de registros.
# La variable generation tiene 6 clases con diferente cantidad de registros.
summary(data)
```

```

##          country          year          sex          age
## Austria      : 382    Min.      :1985    female:13910    15-24 years:4642
## Iceland      : 382    1st Qu.:1995    male  :13910    25-34 years:4642
## Mauritius     : 382    Median  :2002                    35-54 years:4642
## Netherlands: 382    Mean     :2001                    5-14 years :4610
## Argentina     : 372    3rd Qu.:2008                    55-74 years:4642
## Belgium       : 372    Max.      :2016                    75+ years  :4642
## (Other)       :25548
## suicides_no    population    suicides.100k.pop
## Min.      : 0.0    Min.      : 278    Min.      : 0.00
## 1st Qu.: 3.0    1st Qu.: 97498    1st Qu.: 0.92
## Median : 25.0    Median : 430150    Median : 5.99
## Mean : 242.6    Mean : 1844794    Mean : 12.82
## 3rd Qu.: 131.0    3rd Qu.: 1486143    3rd Qu.: 16.62
## Max. :22338.0    Max. :43805214    Max. :224.97
##
##          country.year    HDI.for.year    gdp_for_year....
## Albania1987: 12    Min.      :0.483    1,002,219,052,968: 12
## Albania1988: 12    1st Qu.:0.713    1,011,797,457,139: 12
## Albania1989: 12    Median :0.779    1,016,418,229 : 12
## Albania1992: 12    Mean :0.777    1,018,847,043,277: 12
## Albania1993: 12    3rd Qu.:0.855    1,022,191,296 : 12
## Albania1994: 12    Max. :0.944    1,023,196,003,075: 12
## (Other) :27748    NA's :19456    (Other) :27748
## gdp_per_capita....    generation
## Min. : 251    Boomers :4990
## 1st Qu.: 3447    G.I. Generation:2744
## Median : 9372    Generation X :6408
## Mean : 16866    Generation Z :1470
## 3rd Qu.: 24874    Millenials :5844
## Max. :126352    Silent :6364
##

```

```

# Tipos de datos
str(data)

```

```
## 'data.frame':    27820 obs. of  12 variables:
##  $ country      : Factor w/ 101 levels "Albania","Antigua and Barbuda",...:
1 1 1 1 1 1 1 1 1 1 ...
##  $ year         : int  1987 1987 1987 1987 1987 1987 1987 1987 1987 1987
...
##  $ sex         : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 1 2 1
...
##  $ age         : Factor w/ 6 levels "15-24 years",...: 1 3 1 6 2 6 3 2 5 4
...
##  $ suicides_no  : int   21 16 14 1 9 1 6 4 1 0 ...
##  $ population   : int  312900 308000 289700 21800 274300 35600 278800 2572
00 137500 311000 ...
##  $ suicides.100k.pop : num   6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56 0.73 0 ...
##  $ country.year   : Factor w/ 2321 levels "Albania1987",...: 1 1 1 1 1 1 1 1
1 1 ...
##  $ HDI.for.year   : num   NA NA NA NA NA NA NA NA NA NA ...
##  $ gdp_for_year.... : Factor w/ 2321 levels "1,002,219,052,968",...: 727 727 72
7 727 727 727 727 727 727 ...
##  $ gdp_per_capita....: int   796 796 796 796 796 796 796 796 796 796 ...
##  $ generation     : Factor w/ 6 levels "Boomers","G.I. Generation",...: 3 6 3
2 1 2 6 1 2 3 ...
```

```
res <- sapply(data,class)
kable(data.frame(variables=names(res),clase=as.vector(res)))
```

| variables | clase |
|--------------------|---------|
| country | factor |
| year | integer |
| sex | factor |
| age | factor |
| suicides_no | integer |
| population | integer |
| suicides.100k.pop | numeric |
| country.year | factor |
| HDI.for.year | numeric |
| gdp_for_year.... | factor |
| gdp_per_capita.... | integer |
| generation | factor |

2 Integración y selección de los datos de interés a analizar

- HDI.for.year tiene demasiados vacíos y country.year es un campo derivado

```
# Eliminar columnas
d_suicides <- select(data, -HDI.for.year, -country.year)
str(d_suicides)
```

```
## 'data.frame':    27820 obs. of  10 variables:
## $ country      : Factor w/ 101 levels "Albania","Antigua and Barbuda",...:
1 1 1 1 1 1 1 1 1 1 ...
## $ year         : int  1987 1987 1987 1987 1987 1987 1987 1987 1987 1987
...
## $ sex          : Factor w/ 2 levels "female","male": 2 2 1 2 2 1 1 1 2 1
...
## $ age          : Factor w/ 6 levels "15-24 years",...: 1 3 1 6 2 6 3 2 5 4
...
## $ suicides_no  : int  21 16 14 1 9 1 6 4 1 0 ...
## $ population   : int  312900 308000 289700 21800 274300 35600 278800 2572
00 137500 311000 ...
## $ suicides.100k.pop : num  6.71 5.19 4.83 4.59 3.28 2.81 2.15 1.56 0.73 0 ...
## $ gdp_for_year.... : Factor w/ 2321 levels "1,002,219,052,968",...: 727 727 72
7 727 727 727 727 727 ...
## $ gdp_per_capita....: int  796 796 796 796 796 796 796 796 796 796 ...
## $ generation    : Factor w/ 6 levels "Boomers","G.I. Generation",...: 3 6 3
2 1 2 6 1 2 3 ...
```

3 Limpieza de los datos

3.1 ¿Los datos contienen ceros o elementos vacíos? ¿Cómo gestionarías cada uno de estos casos?

- Suicides y Suicides/100kpop tienen ceros, pero son valores correctos y no requieren una gestión especial
- HDI.for.year contiene un 70% de vacíos. Es difícil completar con valores fiables, por ello se decide descartarlo.

```
colSums(is.na(d_suicides))
```

```
##          country          year          sex
##          0              0              0
##          age      suicides_no      population
##          0              0              0
## suicides.100k.pop  gdp_for_year.... gdp_per_capita....
##          0              0              0
##          generation
##          0
```

```
colSums(d_suicides=="")
```

```
##          country          year          sex
##          0              0              0
##          age      suicides_no      population
##          0              0              0
## suicides.100k.pop  gdp_for_year.... gdp_per_capita....
##          0              0              0
##          generation
##          0
```

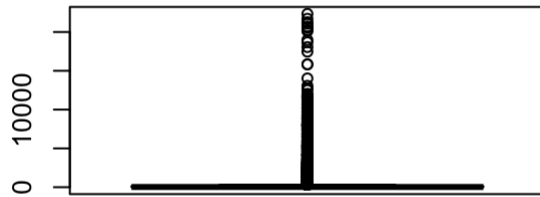
```
colSums(d_suicides=="0")
```

```
##          country          year          sex
##          0              0              0
##          age      suicides_no      population
##          0          4281              0
## suicides.100k.pop  gdp_for_year.... gdp_per_capita....
##          4281              0              0
##          generation
##          0
```

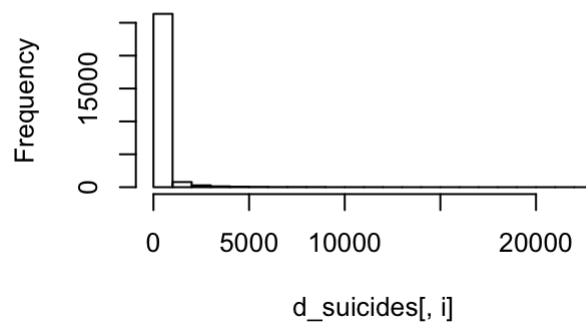
3.2 Identificación y tratamiento de valores extremos.

```
# Análisis mediante diagramas de cajas e histogramas
par(mfrow=c(2,2))
cols <- c("suicides_no","population")
for(i in cols) {
  if (is.integer(d_suicides[,i])){
    boxplot(d_suicides[,i], main = i, width = 100)
    hist(d_suicides[,i], main = i)
  }
}
```

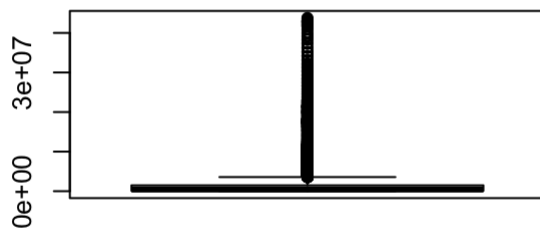
suicides_no



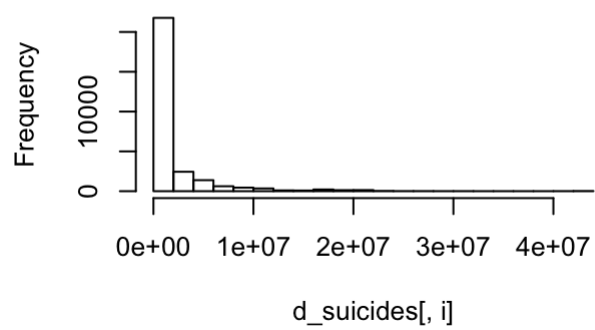
suicides_no



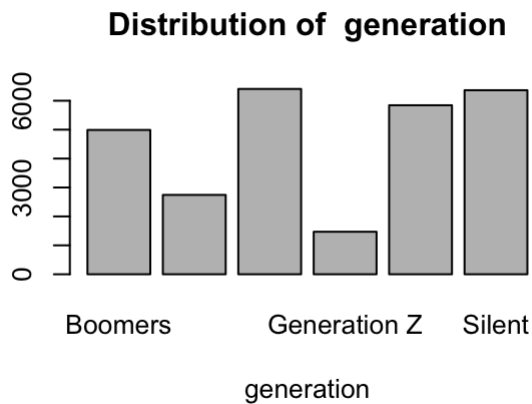
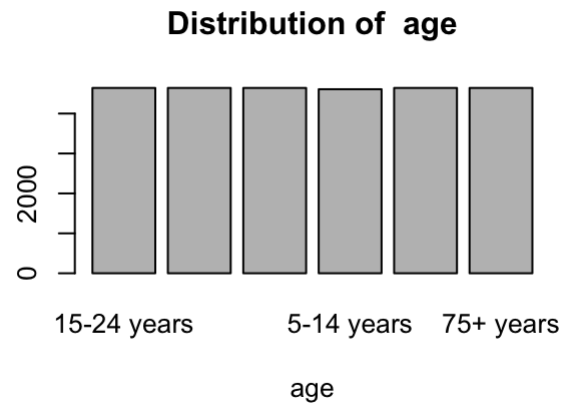
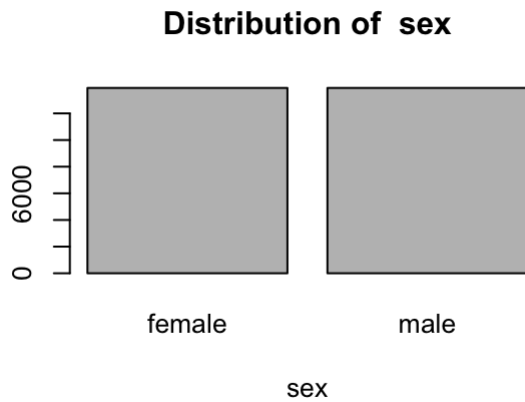
population



population



```
par(mfrow=c(2,2))
cols <- c("sex","age","generation")
for (i in cols){
  if (is.factor(d_suicides[,i])){
    counts <- table(d_suicides[,i])
    barplot(counts,main=paste("Distribution of ", i),
            xlab=i)
  }
}
```



NOTA: Por la naturaleza de las variables population y suicides, no siguen una distribución normal.

4 Análisis de los datos.

4.1 . Selección de los grupos de datos que se quieren analizar/comparar (planificación de los análisis a aplicar).

CASO 1: ¿Es mayor la tasa de suicidios en hombres que en mujeres (en función del porcentaje de población)?

- Seleccionamos atributos: sex, suicides_no, population.
- Calculamos la tasa de suicidios por sexo.
- Representamos gráficamente una comparativa entre la tasa de suicidios de hombres y mujeres.

```
df <- d_suicides %>% select(sex, suicides_no, population)
df[2:3] <- lapply(df[2:3], as.numeric)
report_case1 <- df %>%
  group_by(sex) %>%
  summarise_all(funs(sum)) %>%
  mutate(suicides_100k_pop = suicides_no / population * 100000)
head(report_case1)
```



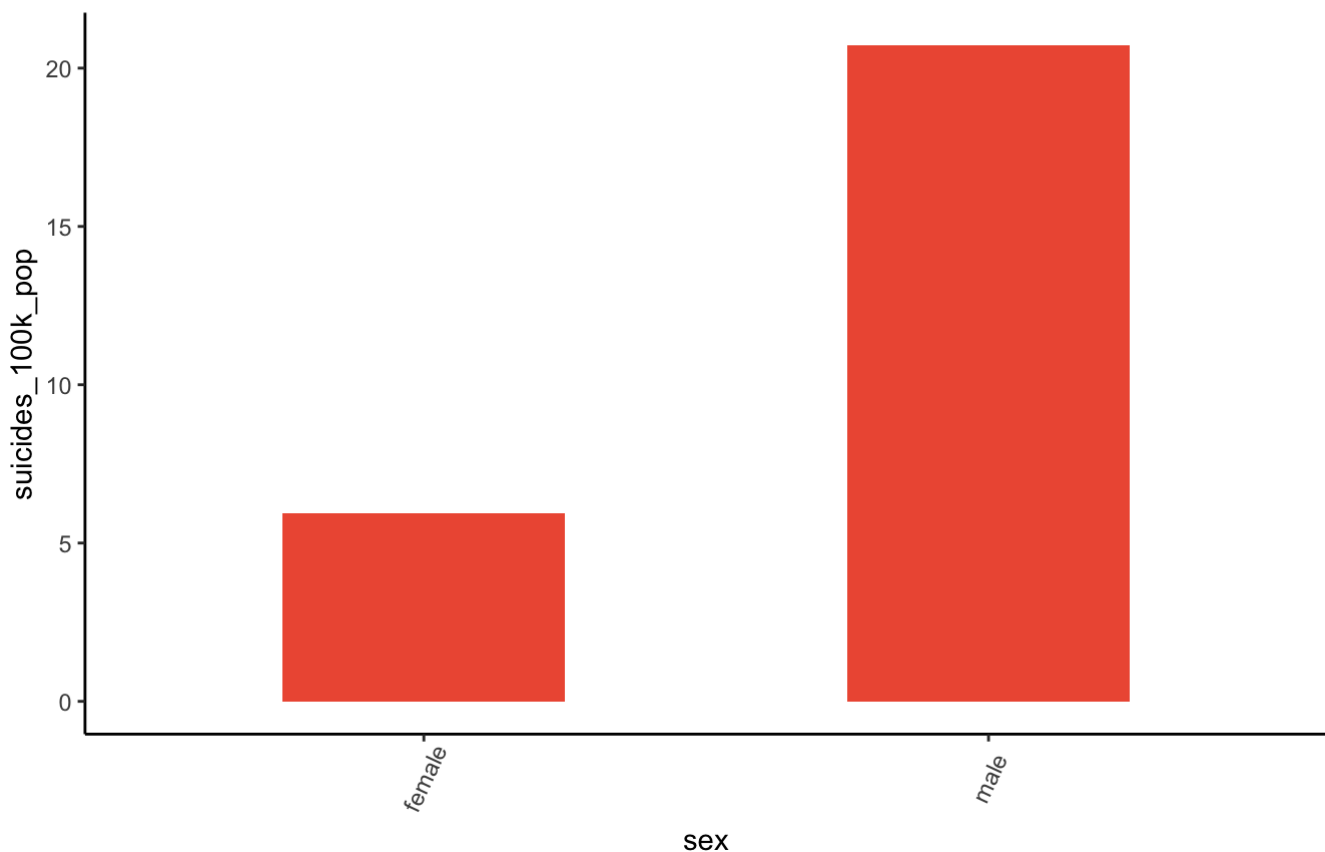
```
## # A tibble: 2 x 4
##   sex      suicides_no  population suicides_100k_pop
##   <fct>      <dbl>      <dbl>      <dbl>
## 1 female    1559510 26272781857         5.94
## 2 male      5188910 25049376579        20.7
```

```
theme_set(theme_classic())
par(mfrow=c(1,1))
g_cas1 <- ggplot(report_cas1, aes(sex,suicides_100k_pop)) + geom_bar(stat="ident
ity", width = 0.5, fill="tomato2") +
  labs(title="CASO 1: Comparativo tasa de suicidios entre hombres y mujeres",
        subtitle="Tasa de suicidios por genero") +
  theme(axis.text.x = element_text(angle=65, vjust=0.6))

g_cas1
```

CASO 1: Comparativo tasa de suicidios entre hombres y mujeres

Tasa de suicidios por genero



Conclusión: la tasa de suicidios en hombres es mayor que en mujeres, más del triple (20.7 vs. 5.94). En total representan el 77,7% de los suicidios.

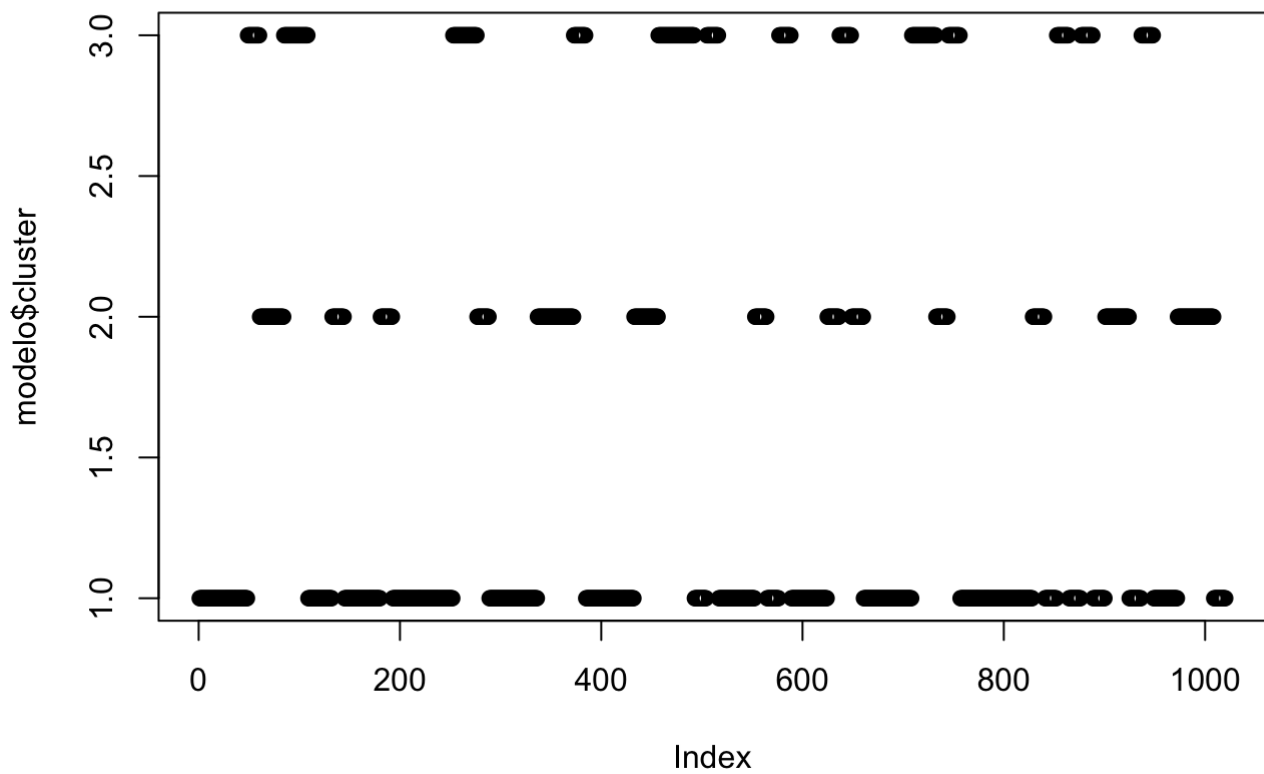
CASO 2: ¿Qué generación tiene mayor tasa de suicidios según el nivel de riqueza de su país en 2008 (crisis económica mundial)?

- Clasificamos los países según el producto interior bruto (gdp_per_capita) para el 2008.

- Calculamos la tasa de suicidios por generación / nivel de riqueza.
- Comparamos por generación y tasa de suicidios para el año 2008.
- Empezamos clasificando los países en 3 niveles de riqueza, mediante el algoritmo k-means.
- Se genera la columna nivel_riqueza
- Atributos: suicides_no, population, generation, gdp_per_capita..., country, year

```
suicides_2008 <- filter(d_suicides, year=='2008')  
  
modelo <- kmeans(suicides_2008$gdp_per_capita..., centers = 3)  
modelo
```

[illegible]



```
suicides_2008$nivel_riqueza <- modelo$cluster
head(suicides_2008)
```

```
## country year sex age suicides_no population suicides.100k.pop
## 1 Albania 2008 male 25-34 years 21 172855 12.15
## 2 Albania 2008 male 35-54 years 37 377119 9.81
## 3 Albania 2008 male 55-74 years 21 228672 9.18
## 4 Albania 2008 female 75+ years 5 59369 8.42
## 5 Albania 2008 female 15-24 years 20 276073 7.24
## 6 Albania 2008 female 25-34 years 9 182663 4.93
## gdp_for_year.... gdp_per_capita.... generation nivel_riqueza
## 1 12,881,352,688 4672 Generation X 1
## 2 12,881,352,688 4672 Boomers 1
## 3 12,881,352,688 4672 Silent 1
## 4 12,881,352,688 4672 Silent 1
## 5 12,881,352,688 4672 Millenials 1
## 6 12,881,352,688 4672 Generation X 1
```

```

report_case2 <- tibble()

for (idx in (1:3)) {
  suicides_2008_riqueza <- filter(suicides_2008, nivel_riqueza==idx)
  df <- suicides_2008_riqueza %>% select(generation, suicides_no, population)

  df[2:3] <- lapply(df[2:3], as.numeric)
  report <- df %>%
    group_by(generation) %>%
    summarise_all(funs(sum)) %>%
    mutate(suicides_100k_pop = suicides_no / population * 100000) %>%
    mutate(nivel_riqueza= paste("Nivel ", idx))

  print(paste("Nivel de Riqueza ", idx))
  print(report)
  report_case2 <- rbind(report_case2, report)
}

```

```

## [1] "Nivel de Riqueza  1"
## # A tibble: 5 x 5
##   generation    suicides_no population suicides_100k_pop nivel_riqueza
##   <fct>          <dbl>      <dbl>          <dbl> <chr>
## 1 Boomers        37105    257826158         14.4 Nivel  1
## 2 Generation X   20265    162811966         12.4 Nivel  1
## 3 Generation Z    1157    164409470          0.704 Nivel  1
## 4 Millenials    16961    176561452          9.61 Nivel  1
## 5 Silent        28427    168182913         16.9 Nivel  1
## [1] "Nivel de Riqueza  2"
## # A tibble: 5 x 5
##   generation    suicides_no population suicides_100k_pop nivel_riqueza
##   <fct>          <dbl>      <dbl>          <dbl> <chr>
## 1 Boomers        31020    187473409         16.5 Nivel  2
## 2 Generation X   10123     85564331         11.8 Nivel  2
## 3 Generation Z     348     79847095          0.436 Nivel  2
## 4 Millenials     7475     87389270          8.55 Nivel  2
## 5 Silent        27550    168891823         16.3 Nivel  2
## [1] "Nivel de Riqueza  3"
## # A tibble: 5 x 5
##   generation    suicides_no population suicides_100k_pop nivel_riqueza
##   <fct>          <dbl>      <dbl>          <dbl> <chr>
## 1 Boomers        18539     97489717         19.0 Nivel  3
## 2 Generation X    7368     49155706         15.0 Nivel  3
## 3 Generation Z     135     34647331          0.390 Nivel  3
## 4 Millenials     3709     38250159          9.70 Nivel  3
## 5 Silent        25265    102120051         24.7 Nivel  3

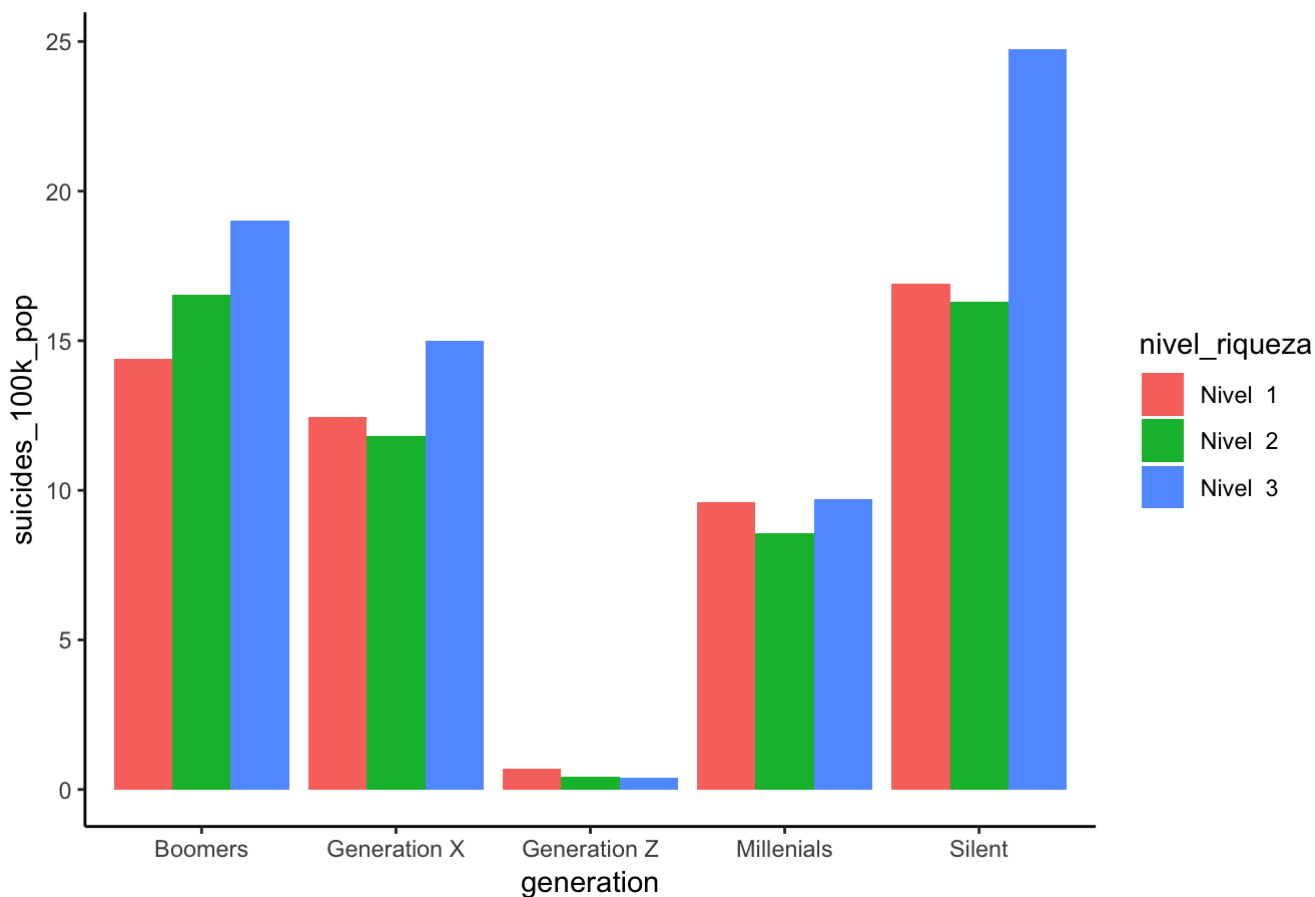
```

```
head(report_case2)
```

```
## # A tibble: 6 x 5
##   generation    suicides_no population suicides_100k_pop nivel_riqueza
##   <fct>          <dbl>      <dbl>          <dbl> <chr>
## 1 Boomers        37105   257826158         14.4 Nivel 1
## 2 Generation X   20265   162811966         12.4 Nivel 1
## 3 Generation Z    1157   164409470          0.704 Nivel 1
## 4 Millenials    16961   176561452          9.61 Nivel 1
## 5 Silent        28427   168182913         16.9 Nivel 1
## 6 Boomers        31020   187473409         16.5 Nivel 2
```

```
g_case2 <- ggplot(report_case2, aes(fill=nivel_riqueza,y=suicides_100k_pop,x=gener
ation)) + geom_bar(position="dodge", stat="identity") +
  labs(title="CASO 2: Comparativo tasa de suicidios de acuerdo a su generacion y n
ivel de riqueza")
g_case2
```

CASO 2: Comparativo tasa de suicidios de acuerdo a su generacion y nivel de ri



CASO 3: ¿Cómo ha evolucionado la tasa de suicidios desde 1985 en España?

- Atributos: suicides_no, population, year, country=Spain

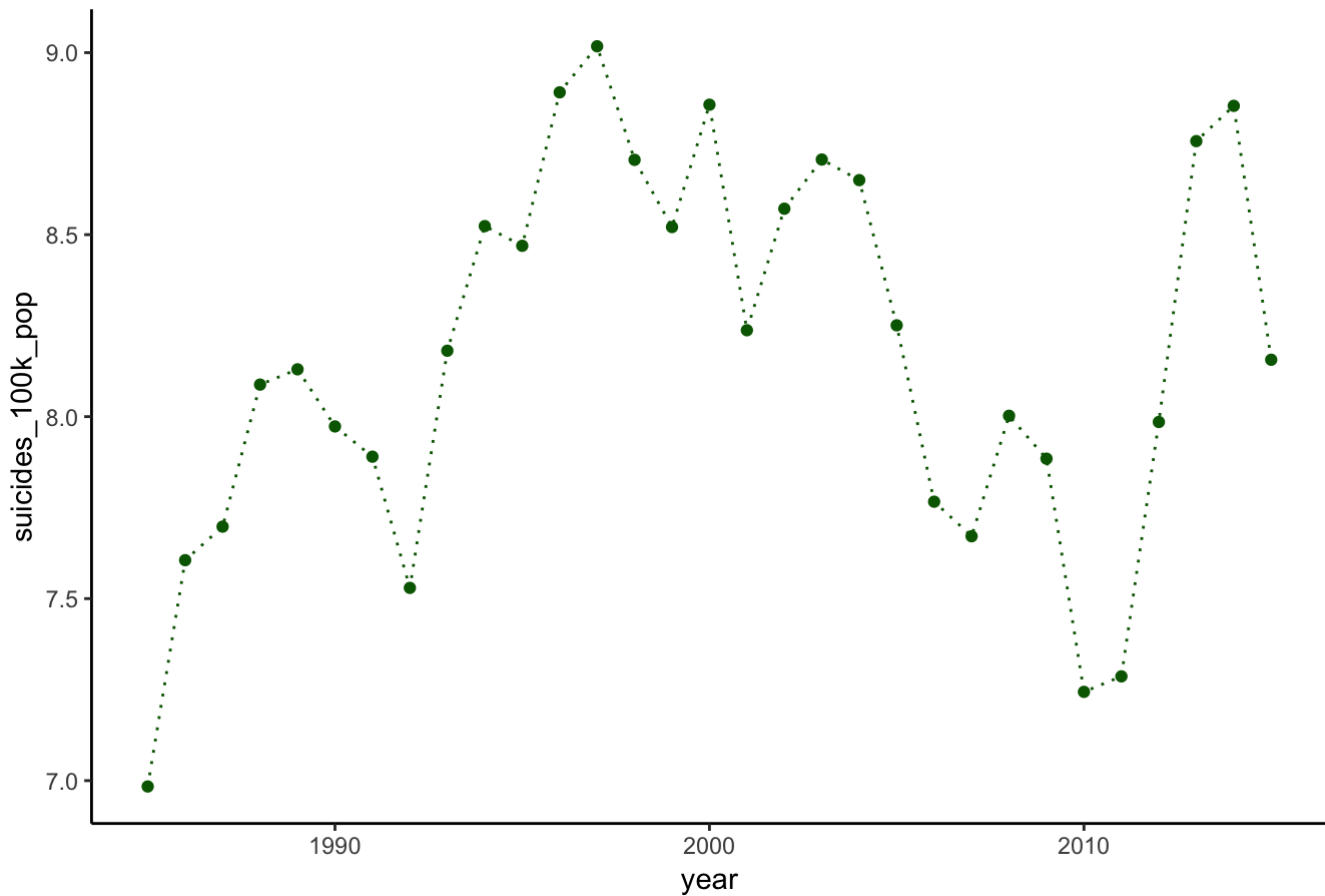
- Creamos una columna que sume el número de suicidios por año (para cada año existen 12 registros, 2 sex x 6 age).
- Creamos una columna que sume la población por año.
- Calculamos la tasa de suicidios por año.
- Graficamos para identificar picos o valores altos.

```
suicides_spain <- select(data, country, year, suicides_no, population)
suicides_spain <- filter(d_suicides, country=='Spain')
df <- suicides_spain %>% select(year, suicides_no, population)
df[2:3] <- lapply(df[2:3], as.numeric)
report_case3 <- df %>%
  group_by(year) %>%
  summarise_all(funs(sum)) %>%
  mutate(suicides_100k_pop = suicides_no / population * 100000)
head(report_case3)
```

```
## # A tibble: 6 x 4
##   year suicides_no population suicides_100k_pop
##   <int>      <dbl>      <dbl>          <dbl>
## 1  1985         2514   35997000          6.98
## 2  1986         2755   36223000          7.61
## 3  1987         2804   36425700          7.70
## 4  1988         2960   36597300          8.09
## 5  1989         2987   36740400          8.13
## 6  1990         2939   36861400          7.97
```

```
par(mfrow=c(1,1))
g_case3 <- ggplot(report_case3, aes(x=year,y=suicides_100k_pop)) + geom_line(stat=
"identity", color="darkgreen", linetype="dotted") +
  labs(title="CASO 3: Evolución de la tasa de suicidios en España desde 1985 hasta
2015") +
  geom_point(color="darkgreen")
g_case3
```


CASO 3: Evolución de la tasa de suicidios en España desde 1985 hasta 2015

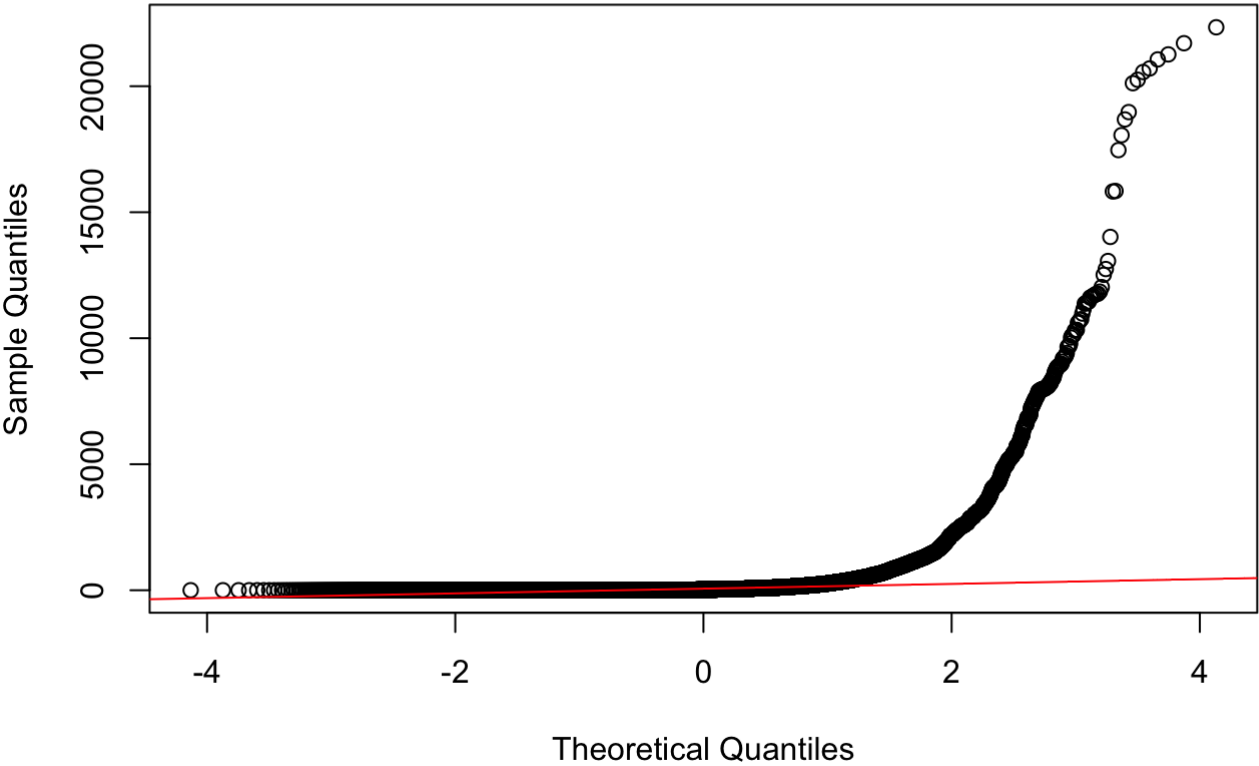


4.2 Comprobación de la normalidad y homogeneidad de la varianza.

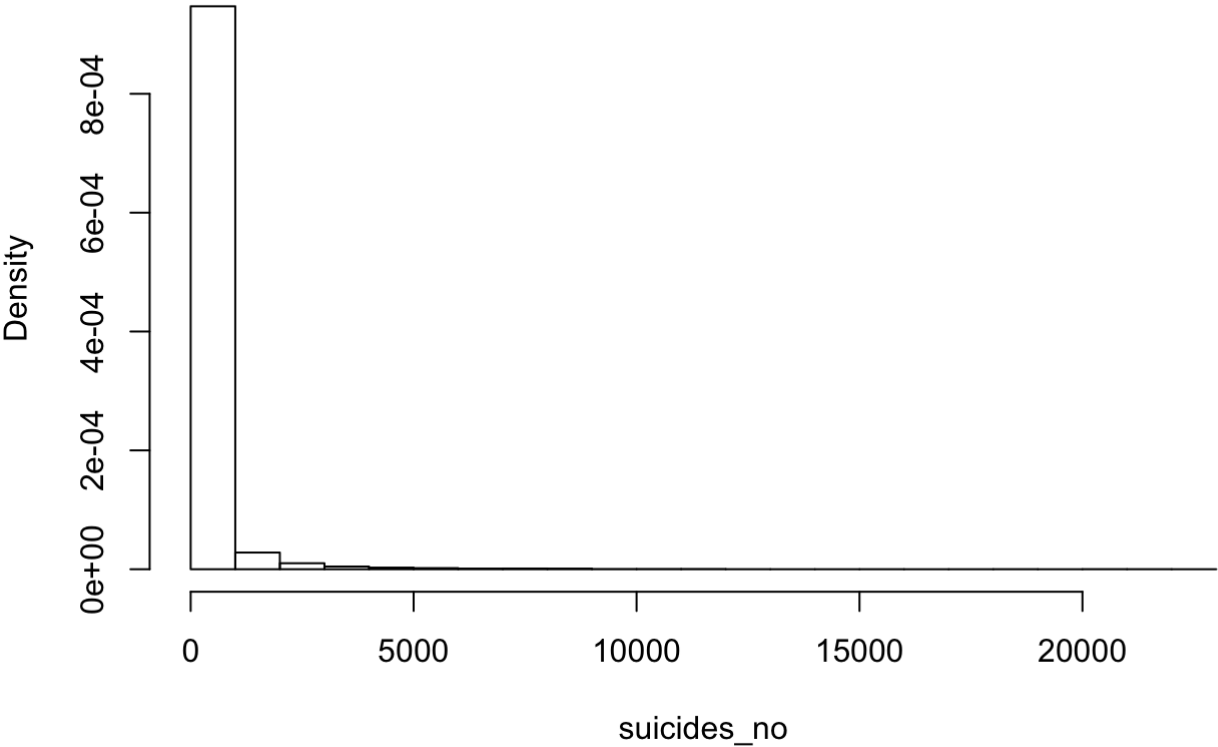
```
# Análisis de normalidad

par(mfrow=c(1,1))
cols <- c("suicides_no", "population")
for (i in cols){
  qqnorm(d_suicides[,i],main = paste("Normal Q-Q Plot for ", i))
  qqline(d_suicides[,i],col="red")
  hist(d_suicides[,i],
       main=paste("Histogram for ", i),
       xlab=i, freq = FALSE)
}
```

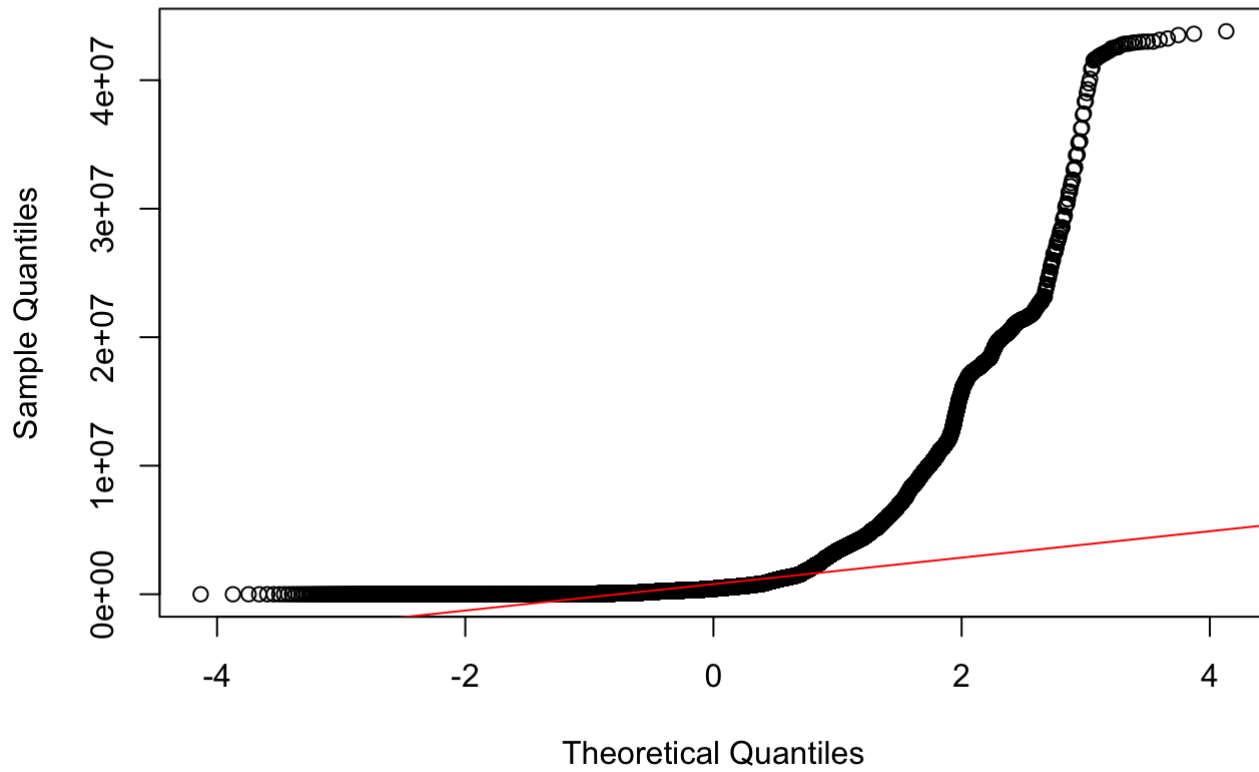

Normal Q-Q Plot for suicides_no



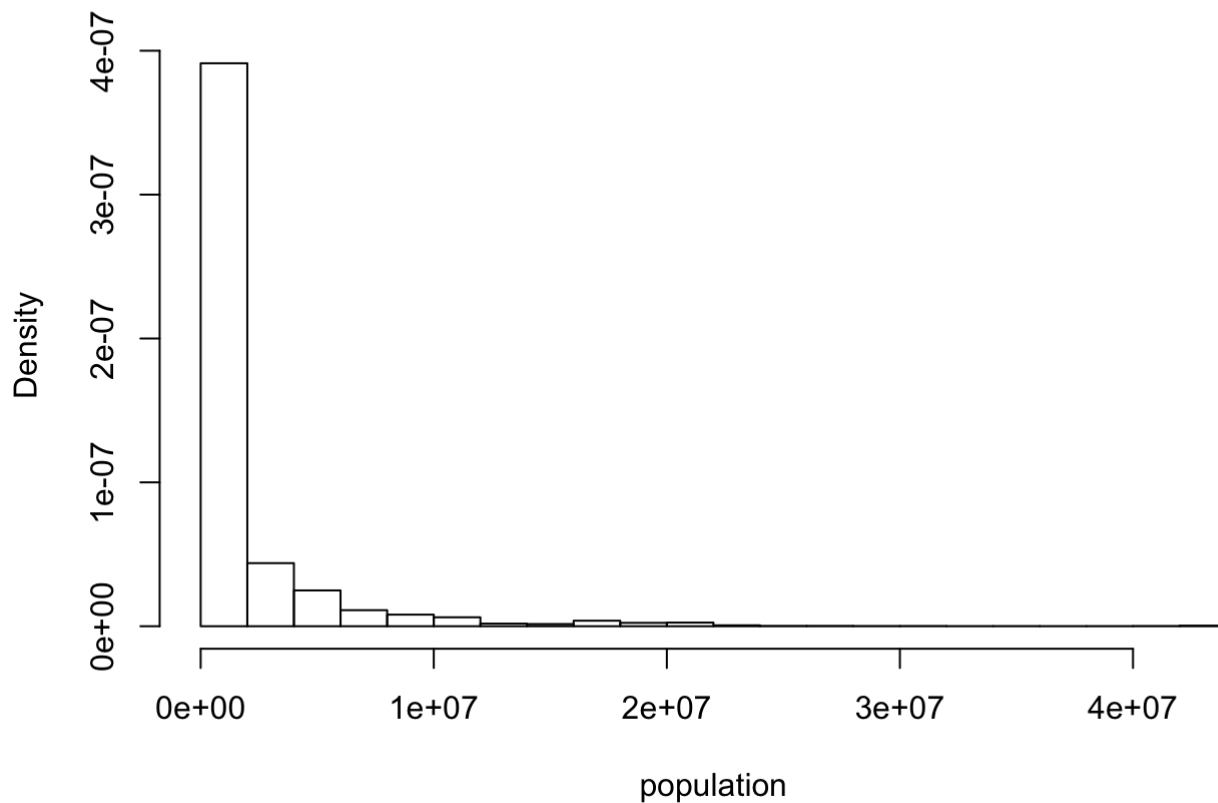
Histogram for suicides_no



Normal Q-Q Plot for population



Histogram for population



CONCLUSIONES:

- Se concluye que no son variables normales.
- Análisis de homogeneidad de la varianza: no aplica.

4.3 Aplicación de pruebas estadísticas para comparar los grupos de datos. En función de los datos y el objetivo del estudio, aplicar pruebas de contraste de hipótesis, correlaciones, regresiones, etc. Aplicar al menos tres métodos de análisis diferentes

5 Representación de los resultados a partir de tablas y gráficas.

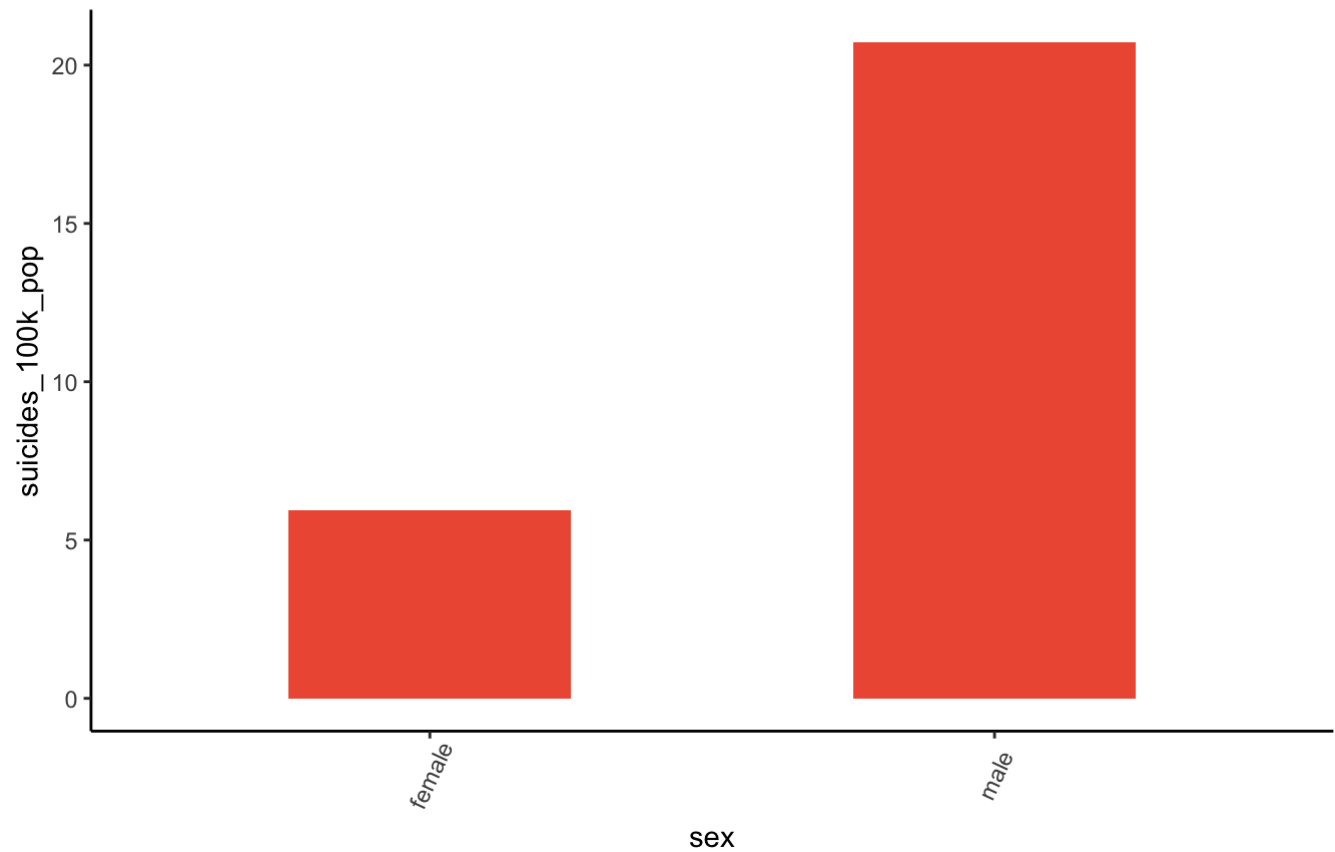
```
par(mfrow=c(2,2))
report_case1
```

```
## # A tibble: 2 x 4
##   sex      suicides_no population suicides_100k_pop
##   <fct>         <dbl>         <dbl>         <dbl>
## 1 female      1559510 26272781857          5.94
## 2 male        5188910 25049376579         20.7
```

```
g_case1
```

CASO 1: Comparativo tasa de suicidios entre hombres y mujeres

Tasa de suicidios por genero



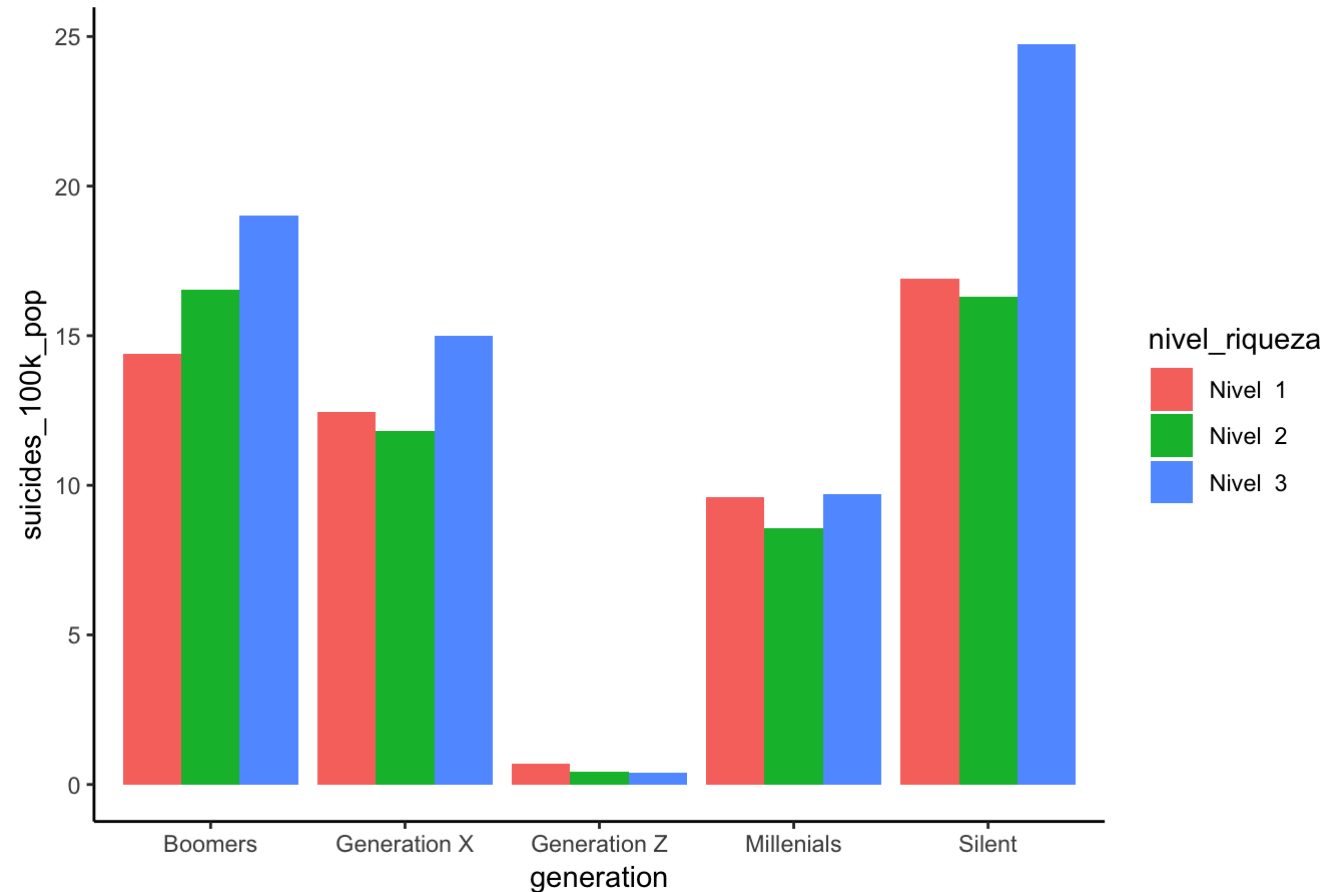
report_case2

```
## # A tibble: 15 x 5
```

| ## | generation | suicides_no | population | suicides_100k_pop | nivel_riqueza |
|----|-----------------|-------------|------------|-------------------|---------------|
| ## | <fct> | <dbl> | <dbl> | <dbl> | <chr> |
| ## | 1 Boomers | 37105 | 257826158 | 14.4 | Nivel 1 |
| ## | 2 Generation X | 20265 | 162811966 | 12.4 | Nivel 1 |
| ## | 3 Generation Z | 1157 | 164409470 | 0.704 | Nivel 1 |
| ## | 4 Millenials | 16961 | 176561452 | 9.61 | Nivel 1 |
| ## | 5 Silent | 28427 | 168182913 | 16.9 | Nivel 1 |
| ## | 6 Boomers | 31020 | 187473409 | 16.5 | Nivel 2 |
| ## | 7 Generation X | 10123 | 85564331 | 11.8 | Nivel 2 |
| ## | 8 Generation Z | 348 | 79847095 | 0.436 | Nivel 2 |
| ## | 9 Millenials | 7475 | 87389270 | 8.55 | Nivel 2 |
| ## | 10 Silent | 27550 | 168891823 | 16.3 | Nivel 2 |
| ## | 11 Boomers | 18539 | 97489717 | 19.0 | Nivel 3 |
| ## | 12 Generation X | 7368 | 49155706 | 15.0 | Nivel 3 |
| ## | 13 Generation Z | 135 | 34647331 | 0.390 | Nivel 3 |
| ## | 14 Millenials | 3709 | 38250159 | 9.70 | Nivel 3 |
| ## | 15 Silent | 25265 | 102120051 | 24.7 | Nivel 3 |

g_case2

CASO 2: Comparativo tasa de suicidios de acuerdo a su generacion y nivel de ri

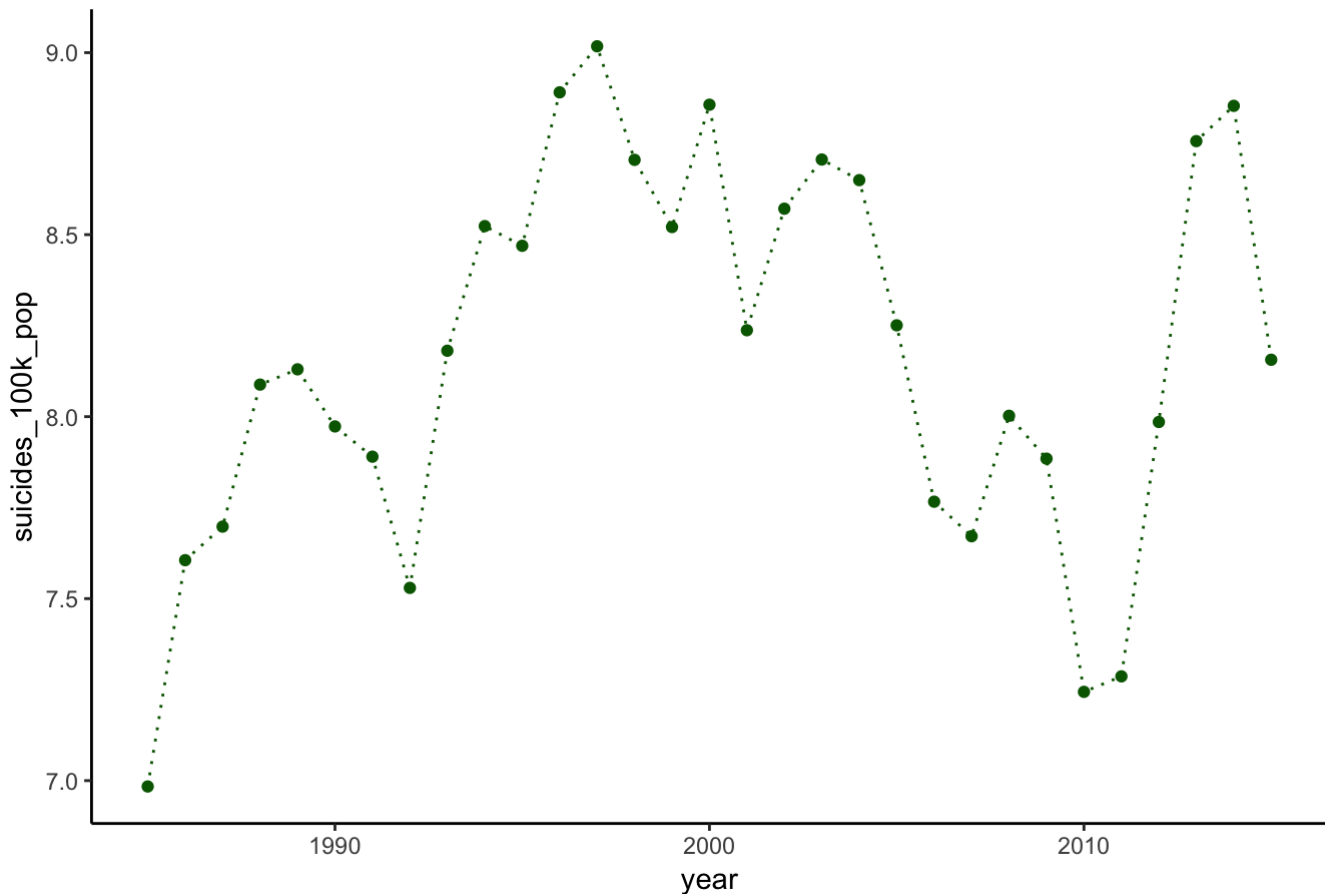


report_case3

```
## # A tibble: 31 x 4
##   year suicides_no population suicides_100k_pop
##   <int>      <dbl>      <dbl>          <dbl>
## 1  1985         2514  35997000         6.98
## 2  1986         2755  36223000         7.61
## 3  1987         2804  36425700         7.70
## 4  1988         2960  36597300         8.09
## 5  1989         2987  36740400         8.13
## 6  1990         2939  36861400         7.97
## 7  1991         2916  36958100         7.89
## 8  1992         2786  37000900         7.53
## 9  1993         3037  37122100         8.18
## 10 1994         3171  37203100         8.52
## # ... with 21 more rows
```

g_case3

CASO 3: Evolución de la tasa de suicidios en España desde 1985 hasta 2015



6 Resolución del problema. A partir de los resultados obtenidos, ¿cuáles son las conclusiones? ¿Los resultados permiten responder al problema?

CASO 1: Tasa de suicidios en hombres y en mujeres

La conclusión aquí es simple: el suicidio afecta sensiblemente más a los hombres, ya que la tasa es más del triple que en las mujeres.

CASO 2: Tasa de suicidios por generaciones y nivel de riqueza

La primera conclusión obvia es que para todas las generaciones salvo una la mayor tasa de suicidios se da entre los países del segundo mundo. El análisis que se puede hacer de esto es que en los países más ricos la crisis no ha golpeado tan fuerte o ha habido más medios para resistir, mientras que en los países más pobres, cuyo nivel de vida previo ya era bajo, la crisis económica posiblemente no supone un gran cambio.

Por otra parte, donde existe una mayor desigualdad es en la generación Silent (nacidos entre finales de los años 20 y mediados de los 40), en la cual destaca claramente el segundo mundo, además de ser la generación más impactada respecto a las demás. En cuanto a lo segundo, se puede concluir que al ser una generación mayor, tiene menor población y por lo tanto la tasa aumenta con facilidad.

Más conclusiones: En la generación Z (nacidos de mediados de los 90 a mediados de los 2000) es en la que claramente se observan menores tasas. Hay que considerar que durante más de la mitad del periodo estudiado no han podido nacer personas de esa generación o bien han sido muy pequeñas como para cometer suicidio, por lo cual es lógico el resultado. Además, es destacable que es el único grupo de población en el que el tercer mundo tiene una tasa más alta. Esto da que pensar que posiblemente los niños del tercer mundo afrontan dificultades que les llevan a cometer suicidios antes que en países más avanzados.

Finalmente se puede observar que las tasas de suicidio han ido disminuyendo por generaciones (en orden cronológico: silent, boomers, x, millenials, z). Siendo un poco optimistas podemos pensar que el mundo ha ido mejorando. Además, las nuevas generaciones han sufrido menor impacto de las guerras y sus consecuencias.

CASO 3: Evolución de la tasa de suicidios en España y su relación con las crisis económicas

Como se puede observar, se confirma la hipótesis de que los suicidios aumentan con las crisis económicas. Los mayores aumentos se producen a raíz de las dos crisis económicas conocidas dentro del periodo estudiado: la de 1993 y la de 2008. En el primer caso, en el que además se parte de una bajada durante 3 años consecutivos, se alcanza en los 3 siguientes la tasa máxima de todo el periodo estudiado. En el segundo caso también tras una progresiva bajada (con algún repunte) se observa la mayor subida puntual de un año para otro.

7 Código: Hay que adjuntar el código, preferiblemente en R, con el que se ha realizado la limpieza, análisis y representación de los datos

acceder código fuente (<https://github.com/gabrielpaladines/uoc-practica2>)

8 OTROS

```
write.csv(d_suicides, file="suicides_datos_finales.csv")
```

9 Contribuciones

| Contribuciones | Firma |
|-------------------------|--------------------------------|
| Investigación previa | Jaime Pardo, Gabriel Paladines |
| Redacción de respuestas | Jaime Pardo, Gabriel Paladines |
| Desarrollo código | Jaime Pardo, Gabriel Paladines |