

# Análisis Exploratorio de los Datos DETECCIÓN CÁNCER DE MAMA , CERVIX UTERINO Y VACUNACIÓN

Alicia Perdices Guerra

20 de mayo, 2021

## Contents

- 1. ANÁLISIS EXPLORATORIO POR PAISES.
  - 1.1 EN RELACIÓN CON PREVENCIÓN
    - \* 1.1.1 Análisis Descriptivo
    - \* 1.1.2 Visualización y Distribución de la variable “Value”
      - DETECCIÓN DE CÁNCER DE MAMA Y CERVIX UTERINO.
      - VACUNACIÓN A LA POBLACIÓN MAYORES DE 65 AÑOS.
    - \* 1.1.3 Normalidad de la variable “Value”
      - DETECCIÓN DE CÁNCER DE MAMA Y CERVIX UTERINO.
      - VACUNACIÓN A LA POBLACIÓN MAYORES DE 65 AÑOS.

**1. ANÁLISIS EXPLORATORIO POR PAISES** Se procede en primer lugar a cargar todos los archivos para poder realizar el análisis.

```
deteccion_cu<-read.csv("C:/temp/Deteccion_Cancer_Mama_Cervix_clean.csv",sep= ",")
vacunacion<-read.csv("C:/temp/Vacunacion_+65_clean.csv",sep= ",")
```

### 1.1.- EN RELACIÓN CON LA PREVENCIÓN SANITARIA

- 1.1.1 Análisis Descriptivo

Se procede a realizar el análisis descriptivo:

```
summary(deteccion_cu)
```

```
##      TIME      GEO      UNIT      SOURCE
## Min.   :2010   Length:1440   Length:1440   Length:1440
## 1st Qu.:2012   Class :character   Class :character   Class :character
## Median :2014   Mode  :character   Mode  :character   Mode  :character
## Mean   :2014
## 3rd Qu.:2017
## Max.   :2019
##      ICD10      Value      Value_imp
## Length:1440    Min.    : 0.04   Mode :logical
```

```
## Class :character 1st Qu.:53.65 FALSE:525
## Mode :character Median :67.70 TRUE :915
## Mean :60.79
## 3rd Qu.:73.59
## Max. :90.40
```

```
summary(vacunacion)
```

```
##      TIME      GEO      UNIT      Value
## Min.   :2010   Length:380   Length:380   Min.    : 0.90
## 1st Qu.:2012   Class :character Class :character 1st Qu.:15.55
## Median :2014   Mode  :character Mode  :character Median :40.04
## Mean   :2014                                     Mean  :35.50
## 3rd Qu.:2017                                     3rd Qu.:50.00
## Max.   :2019                                     Max.   :74.00
## Value_imp
## Mode :logical
## FALSE:261
## TRUE :119
##
##
##
```

Se filtra el dataframe para que la variable GEO aparezcan solo los países objeto de estudio. (Para cada archivo relacionado con La Prevención y unificamos la información). Además se selecciona la información relevante de las variables ICD10 (tipo de cáncer), SOURCE (Survey Data)

```
#En relación con la Prevención
#=====
```

```
#Detección del Cáncer de Mama
#=====
```

```
# Se selecciona la fuente: SOURCE="Survey data"
deteccion_cu<-filter(deteccion_cu,SOURCE=="Survey data")
```

```
#Se filtra por tipo de cancer: Cáncer de mama: "Malignant neoplasm of breast"
deteccion_cmama<-filter(deteccion_cu,ICD10=="Malignant neoplasm of breast")
nrow(deteccion_cmama)
```

```
## [1] 360
```

```
head(deteccion_cmama$Value)
```

```
## [1] 67.3 70.0 75.4 75.4 71.3 62.0
```

```
#Detección del Cáncer de Cérvix Uterino.
#=====
```

```
#Se filtra por tipo de cancer: Cáncer de Cérvix Uterino: "Malignant neoplasm of breast"
deteccion_ccu<-filter(deteccion_cu,ICD10=="Malignant neoplasm of cervix uteri")
nrow(deteccion_ccu)
```

```
## [1] 360
```

```
#Vacunación a mayores de 65 años.  
#=====  
  
vacunacion<- filter(vacunacion,  
  +(GEO!="European Union - 27 countries (from 2020)")&  
  +(GEO!="European Union - 28 countries (2013-2020)"))  
nrow(vacunacion)
```

```
## [1] 360
```

Se crea un Dataframe con toda la información:

```
year<-(vacunacion$TIME)#Columna Year  
country<-(vacunacion$GEO)#Columna Países  
  
#Dataframe con toda la información relacionada  
#con la prevención:cáncer de mama y cérvix uterino, y vacunación  
prevencion<-data.frame("TIME"=year,"Pais"=country,  
  "Deteccion_cancer_mama"=  
    deteccion_cmama$Value,  
  "Deteccion_cancer_cervix_uterino"=  
    deteccion_ccu$Value,  
  "Vacunacion"=  
    vacunacion$Value)  
  
max(prevencion$Deteccion_cancer_mama)
```

```
## [1] 90.4
```

```
max(prevencion$Deteccion_cancer_cervix_uterino)
```

```
## [1] 87.2
```

```
max(prevencion$Vacunacion)
```

```
## [1] 74
```

```
#Generamos el fichero filtrado para utilizarlo en el siguiente análisis.  
write.csv(prevencion, file="Prevencion_Analisis.csv", row.names = FALSE)
```

Se reescalan los datos:

```
prevencion["Deteccion_cancer_mama_norm"]<-  
  rescale(prevencion$Deteccion_cancer_mama, to=c(0,1))  
prevencion["Deteccion_cancer_cervix_uterino_norm"]<-  
  rescale(prevencion$Deteccion_cancer_cervix_uterino, to=c(0,1))  
prevencion["Vacunacion_norm"]<-  
  rescale(prevencion$Vacunacion, to=c(0,1))
```

- 1.1.2 Visualización y Distribución de la información"

Se visualiza las variable que nos dan información sobre la prevención realizada en función de TIME, y País.

```
#Prevención_GRÁFICAS DE BARRAS
#=====
#Gráfica de barras de la información sobre la Detección del Cáncer de Mama"
plot1=ggplot(data=prevencion)+
  geom_col(aes(x=TIME,y=Deteccion_cancer_mama))+
  theme(axis.text.x = element_text(angle = 45))+
  scale_y_continuous(limit=c(0,100))+
  ggtitle("% Deteccion \n Cáncer Mama")+
  theme (plot.title = element_text(size=rel(0.5), hjust = 0.5))

#Gráfica de barras de la información sobre la Detección del Cáncer de Cérvix Uterino"
plot2=ggplot(data=prevencion)+
  geom_col(aes(x=TIME,y=Deteccion_cancer_cervix_uterino))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,100))+
  ggtitle("% Deteccion \n Cáncer Cérvix Uterino")+
  theme (plot.title = element_text(size=rel(0.5),hjust = 0.5))

#Gráfica de barras del porcentaje de Vacunados (>65)"
plot3=ggplot(data=prevencion)+
  geom_col(aes(x=TIME,y=Vacunacion))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,100))+
  ggtitle("% Vacunación+65")+
  theme (plot.title = element_text(size=rel(0.5),hjust = 0.5))

#Prevención_GRÁFICAS DE PUNTOS
#=====
#Gráfica de puntos de la información sobre la Detección del Cáncer de Mama por Países"
plot4=ggplot(data=prevencion)+
  geom_point(aes(x=Pais,y=Deteccion_cancer_mama))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,100))+
  ggtitle("% Deteccion \n Cáncer Mama")+
  theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

#Gráfica de puntos de la información sobre la Detección del Cáncer de Cérvix Uterino por Países"
plot5=ggplot(data=prevencion)+
  geom_point(aes(x=Pais,y=Deteccion_cancer_cervix_uterino))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,100))+
  ggtitle("% Deteccion \n Cáncer Cérvix Uterino")+
  theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

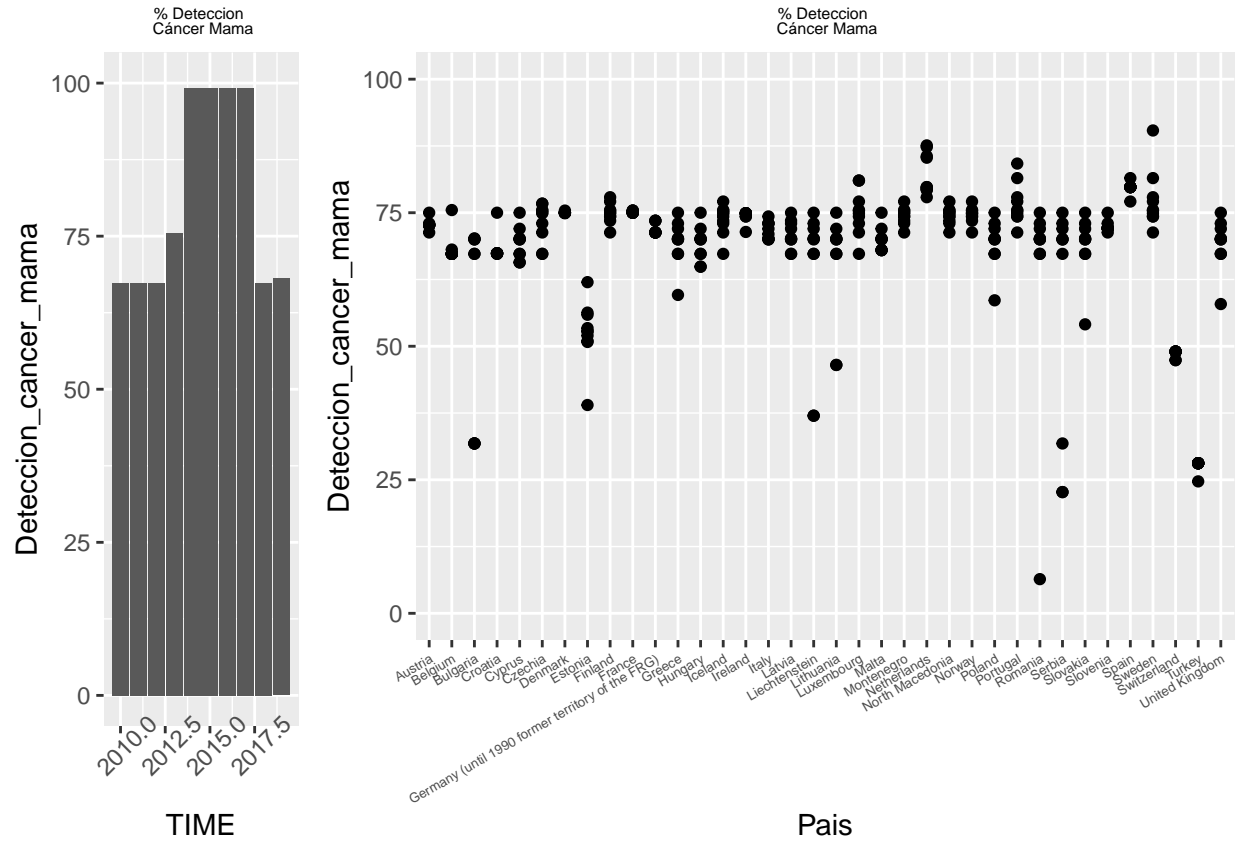
#Gráfica de puntos del porcentaje de Vacunados (>65) por Países
plot6=ggplot(data=prevencion)+
  geom_point(aes(x=Pais,y=Vacunacion))+
  theme(axis.text.x = element_text(size= 5,angle = 30,vjust=1,hjust = 1))+
  scale_y_continuous(limit=c(0,100))+
```

```

ggtitle("% Vacunación+65")+
theme (plot.title = element_text(size=rel(0.5),hjust=0.5))

grid.arrange(plot1,plot4,widths=c(1,3), ncol=2)

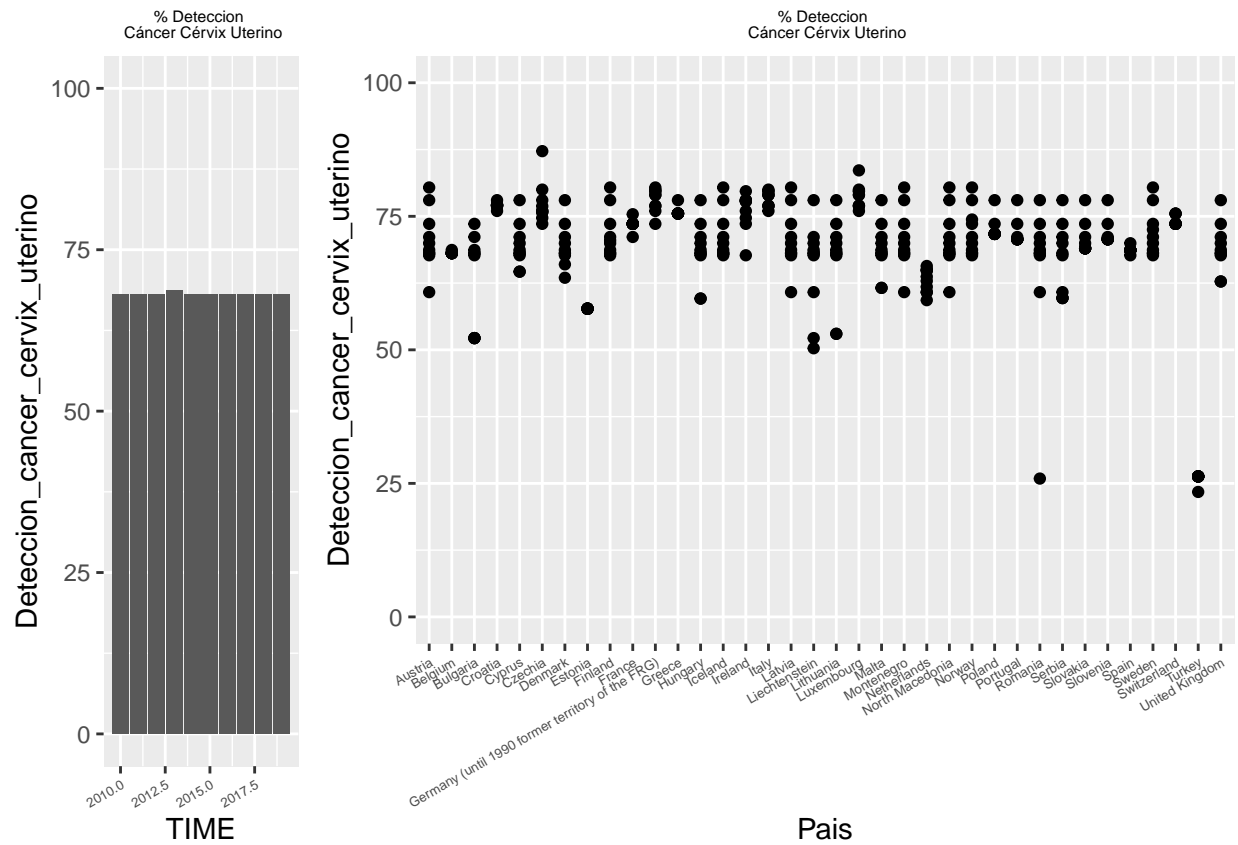
```



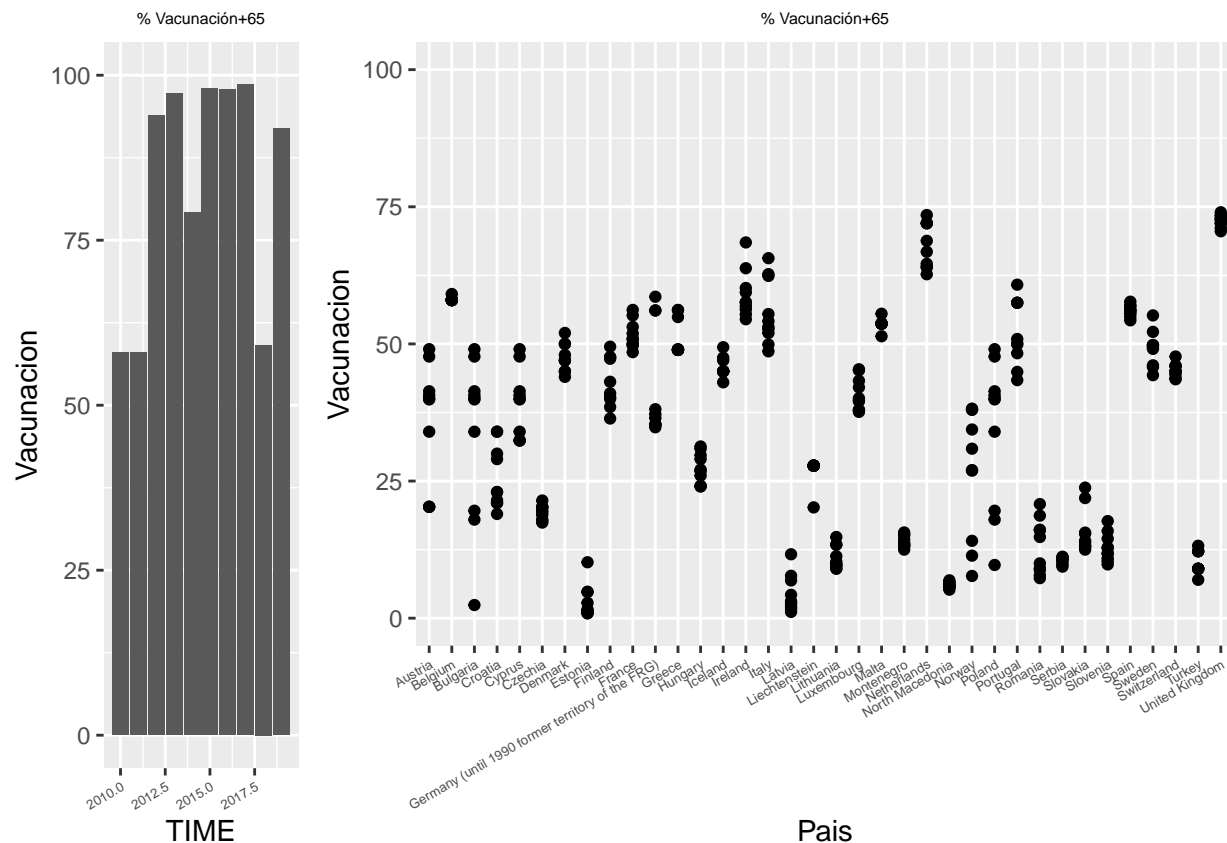
```

grid.arrange(plot2,plot5,widths=c(1,3), ncol=2)

```



```
grid.arrange(plot3,plot6,widths=c(1,3), ncol=2)
```



Se obtienen los 5 países con mayor porcentaje en Detección de Cáncer de Mama, Cérvix Uterino y población vacunada mayor de 65 años:

- **Deteccion\_cancer\_mama**

```
#####
#Para "Deteccion_cancer_mama"#
#####

#Se filtra por Año
y_2010<-filter(prevenccion, TIME==2010)
y_2011<-filter(prevenccion, TIME==2011)
y_2012<-filter(prevenccion, TIME==2012)
y_2013<-filter(prevenccion, TIME==2013)
y_2014<-filter(prevenccion, TIME==2014)
y_2015<-filter(prevenccion, TIME==2015)
y_2016<-filter(prevenccion, TIME==2016)
y_2017<-filter(prevenccion, TIME==2017)
y_2018<-filter(prevenccion, TIME==2018)
y_2019<-filter(prevenccion, TIME==2019)

#Se ordena por "Deteccion_cancer_mama"

prevencion_5países_2010<-y_2010[with(y_2010, order(-y_2010$Deteccion_cancer_mama)),]
prevencion_5países_2011<-y_2011[with(y_2011, order(-y_2011$Deteccion_cancer_mama)),]
prevencion_5países_2012<-y_2012[with(y_2012, order(-y_2012$Deteccion_cancer_mama)),]
```

```

prevencion_5países_2013<-y_2013[with(y_2013, order(-y_2013$Deteccion_cancer_mama)),]
prevencion_5países_2014<-y_2014[with(y_2014, order(-y_2014$Deteccion_cancer_mama)),]
prevencion_5países_2015<-y_2015[with(y_2015, order(-y_2015$Deteccion_cancer_mama)),]
prevencion_5países_2016<-y_2016[with(y_2016, order(-y_2016$Deteccion_cancer_mama)),]
prevencion_5países_2017<-y_2017[with(y_2017, order(-y_2017$Deteccion_cancer_mama)),]
prevencion_5países_2018<-y_2018[with(y_2018, order(-y_2018$Deteccion_cancer_mama)),]
prevencion_5países_2019<-y_2019[with(y_2019, order(-y_2019$Deteccion_cancer_mama)),]

```

*#Se crea una tabla para cada año sobre la Detección del Cáncer de Mama  
#de los 5 Países con un valor más alto.*

```

kable(prevencion_5países_2010[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2010")

```

Table 1: Países con mayor % en Detección de Cáncer de Mama en 2010

	País	% CMama
19	Netherlands	87.6
9	Spain	79.8
3	Czechia	75.4
4	Denmark	75.4
10	France	75.4

```

kable(prevencion_5países_2011[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2011")

```

Table 2: Países con mayor % en Detección de Cáncer de Mama en 2011

	País	% CMama
19	Netherlands	85.6
9	Spain	77.1
16	Luxembourg	77.1
22	Portugal	77.1
26	Finland	77.1

```

kable(prevencion_5países_2012[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2012")

```

Table 3: Países con mayor % en Detección de Cáncer de Mama en 2012

	País	% CMama
19	Netherlands	87.3
9	Spain	79.8
4	Denmark	75.0
10	France	75.0



	País	% CMama
7	Ireland	71.4

```
kable(prevision_5países_2013[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2013")
```

Table 4: Países con mayor % en Detección de Cáncer de Mama en 2013

	País	% CMama
19	Netherlands	85.3
9	Spain	79.8
1	Belgium	75.5
3	Czechia	75.5
16	Luxembourg	75.5

```
kable(prevision_5países_2014[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2014")
```

Table 5: Países con mayor % en Detección de Cáncer de Mama en 2014

	País	% CMama
27	Sweden	90.40
22	Portugal	84.20
16	Luxembourg	81.02
9	Spain	79.80
19	Netherlands	79.80

```
kable(prevision_5países_2015[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2015")
```

Table 6: Países con mayor % en Detección de Cáncer de Mama en 2015

	País	% CMama
9	Spain	79.8
19	Netherlands	79.6
10	France	75.0
3	Czechia	74.9
4	Denmark	74.9

```
kable(prevision_5países_2016[0:5,c(2,3)], col.names = c("País", "% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2016")
```

Table 7: Países con mayor % en Detección de Cáncer de Mama en 2016

	País	% CMama
9	Spain	79.8
19	Netherlands	77.9
22	Portugal	77.9
26	Finland	77.9
27	Sweden	77.9

```
kable(prevision_5países_2017[0:5,c(2,3)], col.names = c("País","% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2017")
```

Table 8: Países con mayor % en Detección de Cáncer de Mama en 2017

	País	% CMama
9	Spain	81.48
22	Portugal	81.48
27	Sweden	81.48
16	Luxembourg	81.02
19	Netherlands	79.30

```
kable(prevision_5países_2018[0:5,c(2,3)], col.names = c("País","% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2018")
```

Table 9: Países con mayor % en Detección de Cáncer de Mama en 2018

	País	% CMama
9	Spain	79.8
19	Netherlands	79.8
4	Denmark	75.0
10	France	75.0
7	Ireland	74.3

```
kable(prevision_5países_2019[0:5,c(2,3)], col.names = c("País","% CMama"),
      caption = "Países con mayor % en Detección de Cáncer de Mama en 2019")
```

Table 10: Países con mayor % en Detección de Cáncer de Mama en 2019

	País	% CMama
9	Spain	79.8
19	Netherlands	79.3
4	Denmark	75.0
10	France	75.0
7	Ireland	74.3

A continuación, se agrupa toda la información por países en una tabla:

```
a1<-group_by(prevenccion,Pais)#Se agrupa por paises
#Se selecciona las variables Pais y Deteccion_cancer_mama
a2<-select(a1,Pais:Deteccion_cancer_mama)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Deteccion_cancer_mama)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la informacion
kable(a5[0:5,c(1,2)],
      col.names = c("País", "% C.Mama"),
      caption = "Países con la mayor media en Detección de Cáncer de Mama en 2010-2019")
```

Table 11: Países con la mayor media en Detección de Cáncer de Mama en 2010-2019

	País	% C.Mama
23	Netherlands	82.150
32	Spain	79.698
33	Sweden	77.258
27	Portugal	76.638
10	France	75.120

- Deteccion\_cancer\_cervix\_uterino

```
#####
#Para "Deteccion_cancer_cervix_uterino"#
#####

#Se ordena por "Deteccion_cancer_cervix_uterino"

prevencion_5países_2010<-y_2010[with(y_2010, order(-y_2010$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2011<-y_2011[with(y_2011, order(-y_2011$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2012<-y_2012[with(y_2012, order(-y_2012$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2013<-y_2013[with(y_2013, order(-y_2013$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2014<-y_2014[with(y_2014, order(-y_2014$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2015<-y_2015[with(y_2015, order(-y_2015$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2016<-y_2016[with(y_2016, order(-y_2016$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2017<-y_2017[with(y_2017, order(-y_2017$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2018<-y_2018[with(y_2018, order(-y_2018$Deteccion_cancer_cervix_uterino)),]
prevencion_5países_2019<-y_2019[with(y_2019, order(-y_2019$Deteccion_cancer_cervix_uterino)),]

#Se crea una tabla para cada año sobre la Detección del Cáncer de Cérvix Uterino
#de los 5 Países con un valor más alto.

kable(prevencion_5países_2010[0:5,c(2,3)], col.names = c("País", "% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2010")
```

Table 12: Países con mayor % en Detección de Cáncer de Cérvix en 2010

	País	% C Cérvix
3	Czechia	75.4
5	Germany (until 1990 former territory of the FRG)	71.3
7	Ireland	74.9
11	Croatia	67.4
12	Italy	70.0

```
kable(prevenccion_5países_2011[0:5,c(2,3)], col.names = c("País","% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2011")
```

Table 13: Países con mayor % en Detección de Cáncer de Cérvix en 2011

	País	% C Cérvix
3	Czechia	76.7
5	Germany (until 1990 former territory of the FRG)	71.3
11	Croatia	67.4
12	Italy	70.1
16	Luxembourg	77.1

```
kable(prevenccion_5países_2012[0:5,c(2,3)], col.names = c("País","% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2012")
```

Table 14: Países con mayor % en Detección de Cáncer de Cérvix en 2012

	País	% C Cérvix
11	Croatia	67.4
12	Italy	70.0
16	Luxembourg	71.3
8	Greece	70.0
31	Switzerland	47.4

```
kable(prevenccion_5países_2013[0:5,c(2,3)], col.names = c("País","% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2013")
```

Table 15: Países con mayor % en Detección de Cáncer de Cérvix en 2013

	País	% C Cérvix
5	Germany (until 1990 former territory of the FRG)	73.5
11	Croatia	67.4
12	Italy	70.2
16	Luxembourg	75.5
8	Greece	70.2

```
kable(preencion_5países_2014[0:5,c(2,3)], col.names = c("País", "% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2014")
```

Table 16: Países con mayor % en Detección de Cáncer de Cérvix en 2014

	País	% C Cérvix
3	Czechia	76.70
16	Luxembourg	81.02
5	Germany (until 1990 former territory of the FRG)	73.50
14	Latvia	73.50
20	Austria	72.70

```
kable(preencion_5países_2015[0:5,c(2,3)], col.names = c("País", "% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2015")
```

Table 17: Países con mayor % en Detección de Cáncer de Cérvix en 2015

	País	% C Cérvix
5	Germany (until 1990 former territory of the FRG)	71.3
12	Italy	72.0
16	Luxembourg	74.9
3	Czechia	74.9
4	Denmark	74.9

```
kable(preencion_5países_2016[0:5,c(2,3)], col.names = c("País", "% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2016")
```

Table 18: Países con mayor % en Detección de Cáncer de Cérvix en 2016

	País	% C Cérvix
3	Czechia	73.0
5	Germany (until 1990 former territory of the FRG)	71.3
12	Italy	73.0
16	Luxembourg	73.0
7	Ireland	74.9

```
kable(preencion_5países_2017[0:5,c(2,3)], col.names = c("País", "% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2017")
```

Table 19: Países con mayor % en Detección de Cáncer de Cérvix en 2017

	País	% C Cérvix
5	Germany (until 1990 former territory of the FRG)	73.50
12	Italy	73.00
16	Luxembourg	81.02
7	Ireland	74.90
11	Croatia	75.00

```
kable(prevenccion_5países_2018[0:5,c(2,3)], col.names = c("País","% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2018")
```

Table 20: Países con mayor % en Detección de Cáncer de Cérvix en 2018

	País	% C Cérvix
5	Germany (until 1990 former territory of the FRG)	71.3
12	Italy	74.3
16	Luxembourg	74.3
7	Ireland	74.3
11	Croatia	67.3

```
kable(prevenccion_5países_2019[0:5,c(2,3)], col.names = c("País","% C Cérvix"),
      caption = "Países con mayor % en Detección de Cáncer de Cérvix en 2019")
```

Table 21: Países con mayor % en Detección de Cáncer de Cérvix en 2019

	País	% C Cérvix
5	Germany (until 1990 former territory of the FRG)	71.3
12	Italy	73.0
16	Luxembourg	67.3
7	Ireland	74.3
11	Croatia	67.3

A continuación, se agrupa toda la información por países en una tabla:

```
a1<-group_by(prevenccion,País)#Se agrupa por países
#Se selecciona las variables País y Deteccion_cancer_cervix_uterino
a2<-select(a1,País:Deteccion_cancer_cervix_uterino)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Deteccion_cancer_cervix_uterino)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la información
```

```
kable(a5[0:5,c(1,2)],
      col.names = c("País", "% C Cérvix"),
      caption = "Países con la mayor media en Detección de Cáncer de Cérvix en 2010-2019")
```

Table 22: Países con la mayor media en Detección de Cáncer de Cérvix en 2010-2019

	País	% C Cérvix
20	Luxembourg	78.790
16	Italy	78.330
11	Germany (until 1990 former territory of the FRG)	78.129
6	Czechia	77.401
4	Croatia	77.042

- Vacunacion

```
#####
#Para "Vacunacion"#
#####

#Se ordena por "Vacunacion"

prevencion_5países_2010<-y_2010[with(y_2010, order(-y_2010$Vacunacion)),]
prevencion_5países_2011<-y_2011[with(y_2011, order(-y_2011$Vacunacion)),]
prevencion_5países_2012<-y_2012[with(y_2012, order(-y_2012$Vacunacion)),]
prevencion_5países_2013<-y_2013[with(y_2013, order(-y_2013$Vacunacion)),]
prevencion_5países_2014<-y_2014[with(y_2014, order(-y_2014$Vacunacion)),]
prevencion_5países_2015<-y_2015[with(y_2015, order(-y_2015$Vacunacion)),]
prevencion_5países_2016<-y_2016[with(y_2016, order(-y_2016$Vacunacion)),]
prevencion_5países_2017<-y_2017[with(y_2017, order(-y_2017$Vacunacion)),]
prevencion_5países_2018<-y_2018[with(y_2018, order(-y_2018$Vacunacion)),]
prevencion_5países_2019<-y_2019[with(y_2019, order(-y_2019$Vacunacion)),]

#Se crea una tabla para cada año sobre la Vacunación a mayores de 65 años
#de los 5 Países con un valor más alto.

kable(prevencion_5países_2010[0:5,c(2,3)], col.names = c("País", "% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2010")
```

Table 23: Países con mayor % Vacunación a mayores de 65 años en 2010

	País	% Vacunacion
32	United Kingdom	70.0
19	Netherlands	87.6
12	Italy	70.0
7	Ireland	74.9
1	Belgium	67.3

```
kable(prevision_5países_2011[0:5,c(2,3)], col.names = c("País", "% Vacunación"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2011")
```

Table 24: Países con mayor % Vacunación a mayores de 65 años en 2011

	País	% Vacunación
32	United Kingdom	70.1
19	Netherlands	85.6
12	Italy	70.1
1	Belgium	67.3
9	Spain	77.1

```
kable(prevision_5países_2012[0:5,c(2,3)], col.names = c("País", "% Vacunación"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2012")
```

Table 25: Países con mayor % Vacunación a mayores de 65 años en 2012

	País	% Vacunación
19	Netherlands	87.3
32	United Kingdom	70.0
12	Italy	70.0
5	Germany (until 1990 former territory of the FRG)	71.3
1	Belgium	67.3

```
kable(prevision_5países_2013[0:5,c(2,3)], col.names = c("País", "% Vacunación"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2013")
```

Table 26: Países con mayor % Vacunación a mayores de 65 años en 2013

	País	% Vacunación
32	United Kingdom	70.2
19	Netherlands	85.3
7	Ireland	74.9
1	Belgium	75.5
9	Spain	79.8

```
kable(prevision_5países_2014[0:5,c(2,3)], col.names = c("País", "% Vacunación"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2014")
```



Table 27: Países con mayor % Vacunación a mayores de 65 años en 2014

	País	% Vacunacion
32	United Kingdom	57.9
19	Netherlands	79.8
7	Ireland	74.9
1	Belgium	67.3
9	Spain	79.8

```
kable(preencion_5países_2015[0:5,c(2,3)], col.names = c("País","% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2015")
```

Table 28: Países con mayor % Vacunación a mayores de 65 años en 2015

	País	% Vacunacion
32	United Kingdom	72.0
19	Netherlands	79.6
1	Belgium	67.3
9	Spain	79.8
7	Ireland	74.9

```
kable(preencion_5países_2016[0:5,c(2,3)], col.names = c("País","% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2016")
```

Table 29: Países con mayor % Vacunación a mayores de 65 años en 2016

	País	% Vacunacion
32	United Kingdom	73.0
19	Netherlands	77.9
1	Belgium	67.3
22	Portugal	77.9
9	Spain	79.8

```
kable(preencion_5países_2017[0:5,c(2,3)], col.names = c("País","% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2017")
```

Table 30: Países con mayor % Vacunación a mayores de 65 años en 2017

	País	% Vacunacion
32	United Kingdom	75.00
19	Netherlands	79.30
22	Portugal	81.48
1	Belgium	67.30
7	Ireland	74.90

```
kable(prevenccion_5países_2018[0:5,c(2,3)], col.names = c("País", "% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2018")
```

Table 31: Países con mayor % Vacunación a mayores de 65 años en 2018

	País	% Vacunacion
32	United Kingdom	67.3
7	Ireland	74.3
19	Netherlands	79.8
1	Belgium	67.3
22	Portugal	74.3

```
kable(prevenccion_5países_2019[0:5,c(2,3)], col.names = c("País", "% Vacunacion"),
      caption = "Países con mayor % Vacunación a mayores de 65 años en 2019")
```

Table 32: Países con mayor % Vacunación a mayores de 65 años en 2019

	País	% Vacunacion
32	United Kingdom	67.3
19	Netherlands	79.3
1	Belgium	68.1
7	Ireland	74.3
22	Portugal	74.3

A continuación, se agrupa toda la información por países en una tabla:

```
a1<-group_by(prevenccion,País)#Se agrupa por países
#Se selecciona las variables País y Vacunación
a2<-select(a1,País:Vacunacion)
#Se muestra la información por cada país, con Value=suma de
#los valores de cada país en los 10 años.
a3<-(summarize(a2,suma=sum(Vacunacion)/10))
a4<-data.frame(a3)#Se convierte la información en un dataframe.
#Se ordena el DataFrame por la variable Suma de forma descendente.
a5<-a4[with(a4,order(-a4$suma)),]

#Se crea una tabla con toda la información
kable(a5[0:5,c(1,2)],
      col.names = c("País", "% Vacunacion"),
      caption = "Países con la mayor media en Vacunación a mayores de 65 años en 2010-2019")
```

Table 33: Países con la mayor media en Vacunación a mayores de 65 años en 2010-2019

	País	% Vacunacion
36	United Kingdom	72.455
23	Netherlands	68.040

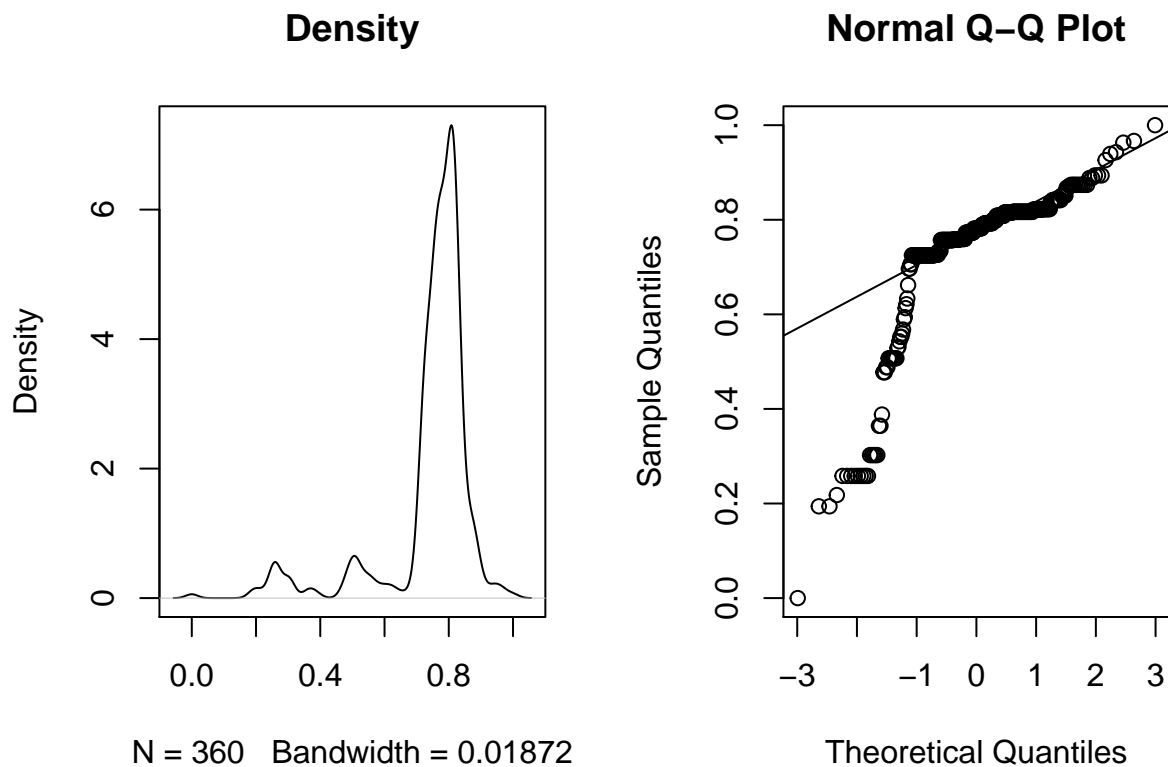
	País	% Vacunacion
15	Ireland	59.020
2	Belgium	58.110
32	Spain	56.050

### • 1.1.3 Normalidad de la variable “Value”

Se comprueba con métodos visuales si la variable tiene una distribución normal.

#### DETECCIÓN DE CÁNCER DE MAMA

```
par(mfrow=c(1,2))
plot(density(prevenccion$Deteccion_cancer_mama_norm) ,main="Density")
qqnorm(prevenccion$Deteccion_cancer_mama_norm)
qqline(prevenccion$Deteccion_cancer_mama_norm)
```



Para estudiar si una muestra proviene de una población con distribución normal, se disponen de tres herramientas:

- Histograma o Densidad
- Gráficos cuantil cuantil (QQplot)
- Pruebas de hipótesis.

Si en la prueba de Densidad se observa sesgo hacia uno de los lados de la gráfica, sería indicio de que la muestra no proviene de una población normal. Si por otra parte, sí se observa simetría, **NO** se garantiza que

la muestra provenga de una población normal. En estos casos sería necesario utilizar otras herramientas como **QQplot y pruebas de hipótesis**.

En la gráfica Densidad de la variable “Deteccion\_cancer\_mama\_norm”, se observa claramente sesgo hacia la derecha, por lo que no se considera normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico. Si se tuviese una muestra distribuída normalmente, se esperaría que los puntos del gráfico quantil quantil estuviesen perfectamente alineados con la línea de referencia, y observamos que para este caso, “Deteccion\_cancer\_mama\_norm” no se alinea.

Por otro lado, se realizan las pruebas de hipótesis:

- \$h\_0\$: La muestra proviene de una población normal.
- \$h\_1\$: La muestra NO proviene de una población normal.

Se aplica la prueba Shapiro-Wilk:

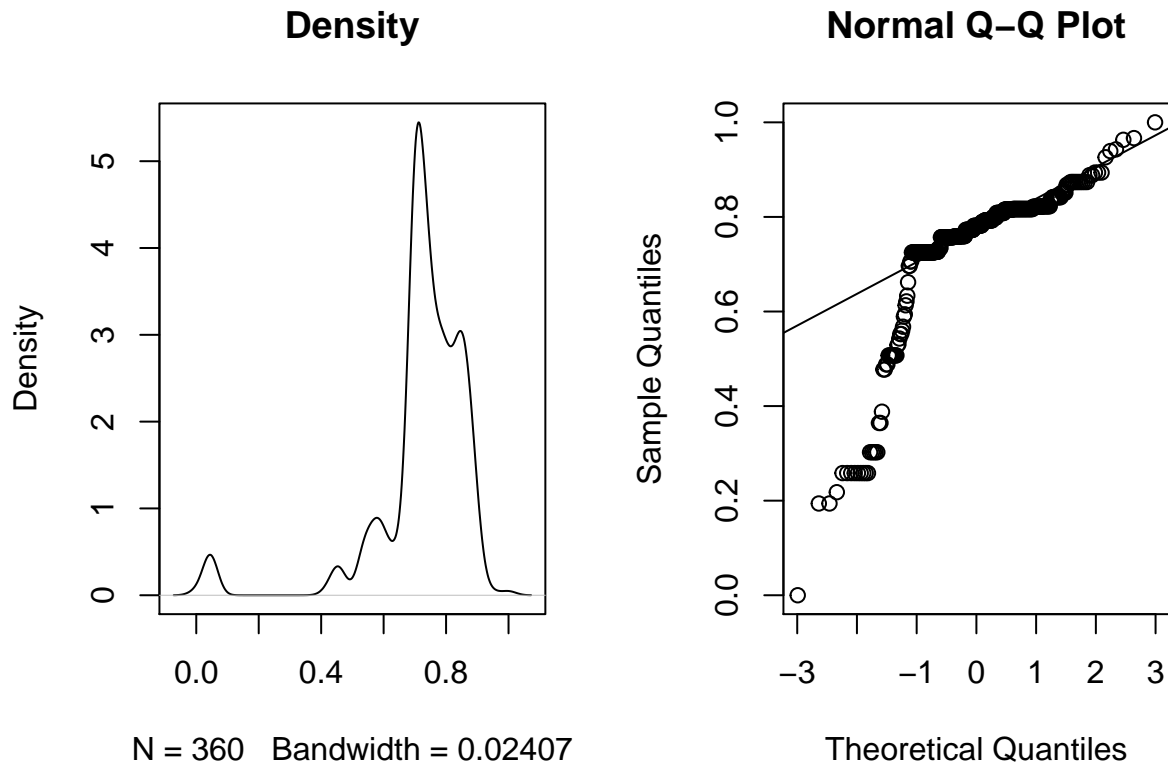
```
shapiro.test(prevenccion$Deteccion_cancer_mama_norm)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  prevenccion$Deteccion_cancer_mama_norm  
## W = 0.69359, p-value < 2.2e-16
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.

## DETECCIÓN DE CÁNCER DE CÉRVIX UTERINO

```
par(mfrow=c(1,2))  
plot(density(prevenccion$Deteccion_cancer_cervix_uterino_norm) ,main="Density")  
qqnorm(prevenccion$Deteccion_cancer_mama_norm)  
qqline(prevenccion$Deteccion_cancer_mama_norm)
```



En la gráfica Densidad de la variable “Deteccion\_cancer\_cervix\_uterino\_norm”, se observa claramente sesgo hacia la derecha, por lo que no se considera normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico, no se alinea con los puntos de los valores de la variable “Deteccion\_cancer\_cervix\_uterino\_norm”.

Tras aplicar la prueba Shapiro-Wilk:

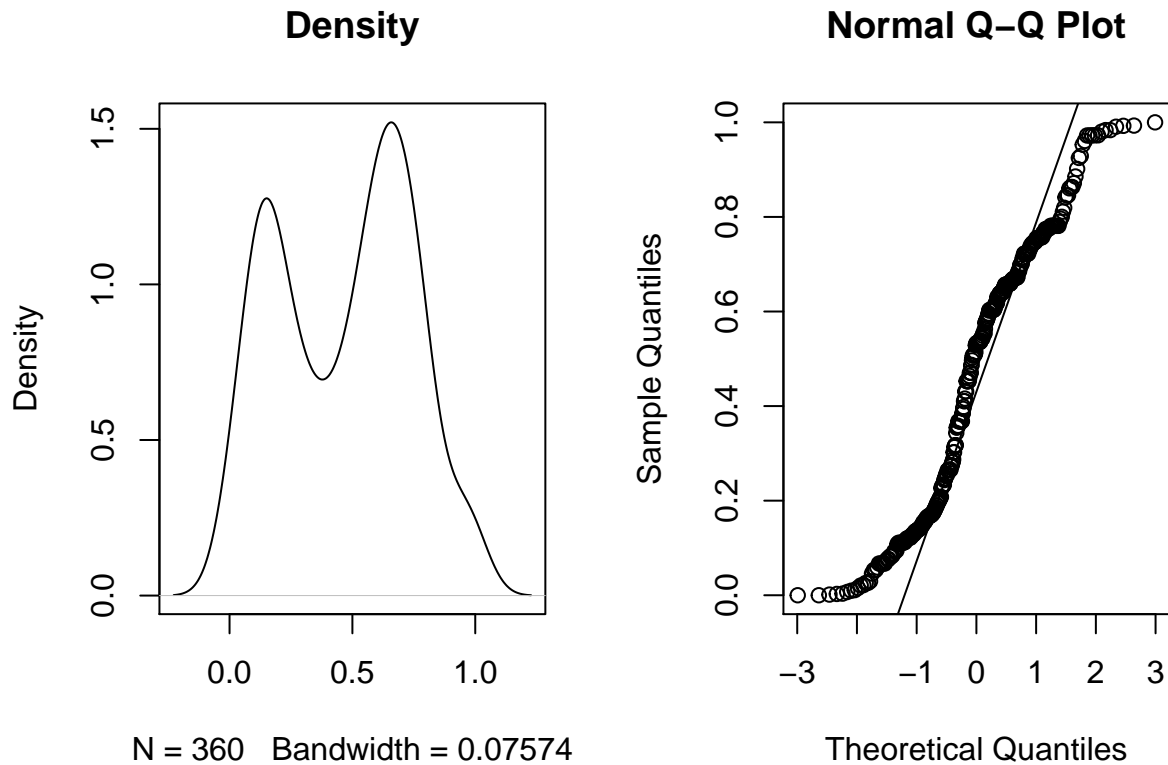
```
shapiro.test(prevenccion$Deteccion_cancer_mama_norm)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  prevenccion$Deteccion_cancer_mama_norm
## W = 0.69359, p-value < 2.2e-16
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.

## VACUNACION A LA POBLACIÓN MAYOR DE 65 AÑOS

```
par(mfrow=c(1,2))
plot(density(prevenccion$Vacunacion_norm) ,main="Density")
qqnorm(prevenccion$Vacunacion_norm)
qqline(prevenccion$Vacunacion_norm)
```



En la gráfica Densidad de la variable “Vacunacion\_norm” , se observa claramente una figura que no se aproxima a la curva de normalidad. Se puede confirmar observando la gráfica QQplot en la que la línea que grafica qqline sirve de referencia para interpretar el gráfico, no se alinea con los puntos de los valores de la variable “Vacunacion\_norm”.Tan solo, existe una alineación central.

Tras aplicar la prueba Shapiro-Wilk:

```
shapiro.test(prevision$Vacunacion_norm)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  prevision$Vacunacion_norm
## W = 0.94426, p-value = 2.148e-10
```

Se observa un p-value muy pequeño, mucho más pequeño que cualquier nivel de significación ( $\alpha=0.5$ ) por lo que se rechaza la hipótesis nula y asumimos **No normalidad** en la muestra.