

# DigitalAgri - Tarea Fundamentos de R

Alberto Gómez Juan

10/12/2021

## Índice

Objetivos . . . . .	1
Introducción . . . . .	1
Datos utilizados . . . . .	1
Metodología . . . . .	2
Interpretación de resultados . . . . .	3
Gráficas . . . . .	5

## Procesos y Casos de Innovación y Digitalización y Vigilancia en el Sector Agroalimentario, y Forestal y del Desarrollo Rural

### Tema: Fundamentos del programa R

## Objetivos

- Analizar un conjunto de datos de la FAO teniendo en cuenta los conceptos estadísticos vistos en clase.
- Aplicar los fundamentos del programa R para el análisis de datos.
- Verificar si existe relación entre área cosechada, producción, rendimiento de aceitunas en España y las exportaciones de aceite de oliva virgen entre los años 1980 y 2020.

## Introducción

España es el primer país productor de aceitunas de mesa del mundo, seguido por Egipto, Turquía, Argelia, Grecia, Siria, Marruecos, Perú y Argentina. La producción media mundial de las últimas cinco campañas asciende a 2.847.700 toneladas, de las cuales 561.100 se produjeron en España, es decir, el 19,7 % del total<sup>1</sup>.

Según datos del COI (Consejo Oleícola Internacional), el consumo medio de las cinco últimas campañas asciende a 2.814.100 toneladas y se reparten por países según el siguiente gráfico.

## Datos utilizados

Los datos utilizados para el estudio fueron obtenidos de los datos abiertos de la FAO (<http://www.fao.org/faostat/es/#home>) y corresponden al reporte de area cosechada, produccion, rendimiento y exportación de aceite de oliva virgen entre los años 1980 y 2019.

---

<sup>1</sup><https://asemesa.es/informacion-general-del-sector/>

## Exportaciones Mundiales Aceituna de Mesa Media campañas 2015-20

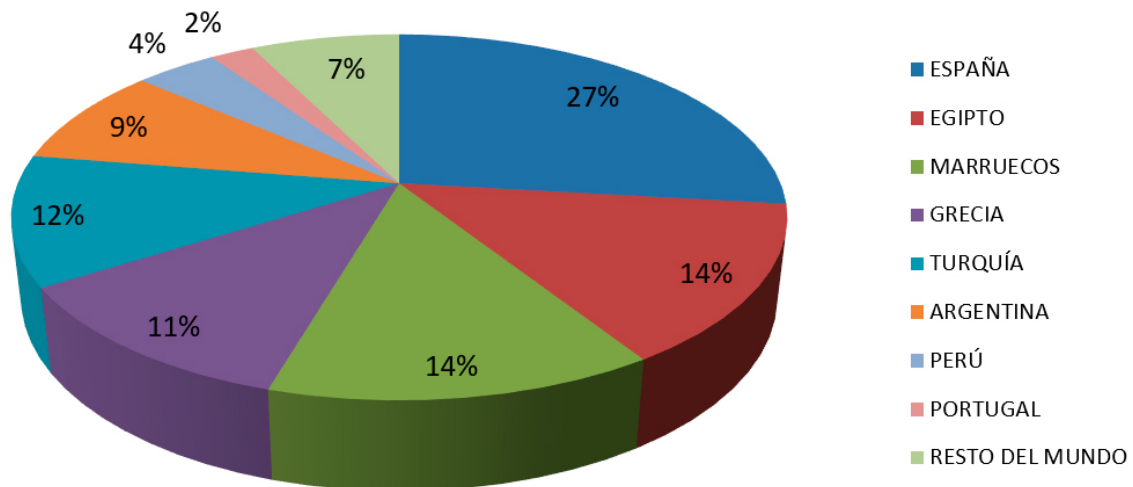


Figura 1: Fuente: Consejo Oleícola Internacional

## Metodología

Para empezar se incluyen las librerías utilizadas para el análisis de los datos:

```
# Importar las librerías utilizadas
library(tidyverse)
library(cowplot)
library(ggplot2)
library(dplyr)

#Gráficos de correlación
library(PerformanceAnalytics)
```

Seguido a esto, se leen los datos de desde archivos CSV para conformar un dataset con los datos organizados para los procedimientos estadísticos.

```
# Lectura de datos a partir de un archivo CSV.
areacosechada_origen <- read.csv("datos/FAOSTAT_data_12-18-2021_area.csv")
rendimiento_origen <- read.csv("datos/FAOSTAT_data_12-18-2021_rendimiento.csv")
produccion_origen <- read.csv("datos/FAOSTAT_data_12-18-2021_produccion.csv")
exportaciones_origen <- read.csv("datos/FAOSTAT_data_12-18-2021_exportaciones.csv")
```

Se realiza la limpieza de los datos para extraer unicamente los necesarios y se descarta el resto de información descargada de la FAO.

```
# Extracción de las columnas correspondientes a año y producción.
areacosechada <- areacosechada_origen[, c(12)]
rendimiento <- rendimiento_origen[, c(12)]
produccion <- produccion_origen[, c(12)]
exportaciones <- exportaciones_origen[, c(12)]

# Lista de años
anios <- produccion_origen[, c(10)]
```

El nuevo dataframe con la información importante para el análisis se puede ver en la siguiente tabla:

```
# Convertir en dataframe
olivar <- data.frame(areacosechada, produccion, rendimiento, exportaciones, row.names = anios)

data(olivar)
summary(olivar)
```

```
## areacosechada      produccion      rendimiento      exportaciones
## Min.      :1156500  Min.      :1328400  Min.      : 6479  Min.      : 32057
## 1st Qu.:2061500    1st Qu.:2973675    1st Qu.:15124    1st Qu.: 207249
## Median :2253778    Median :4347046    Median :19178    Median : 409442
## Mean    :2257193    Mean    :4646455    Mean    :20083    Mean    : 461844
## 3rd Qu.:2471488    3rd Qu.:6008873    3rd Qu.:25380    3rd Qu.: 672981
## Max.    :2601900    Max.    :9819570    Max.    :38075    Max.    :1120699
```

```
head(olivar, 3)
```

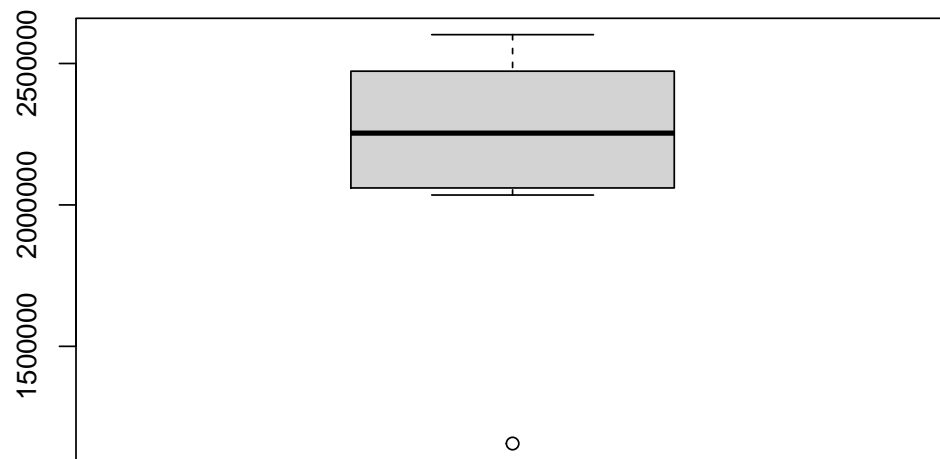
```
##      areacosechada produccion rendimiento exportaciones
## 1980      1156500      2255000      19498      121406
## 1981      2045000      1520800       7437      55495
## 1982      2045800      3337900      16316      32057
```

```
tail(olivar, 3)
```

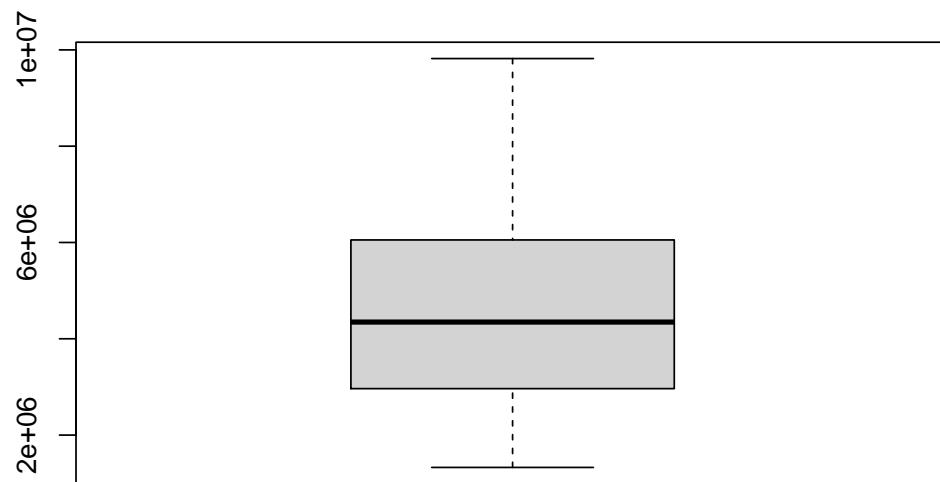
```
##      areacosechada produccion rendimiento exportaciones
## 2017      2554829      6549499      25636      944462
## 2018      2579000      9819570      38075      902210
## 2019      2601900      5965080      22926      1075727
```

## Interpretación de resultados

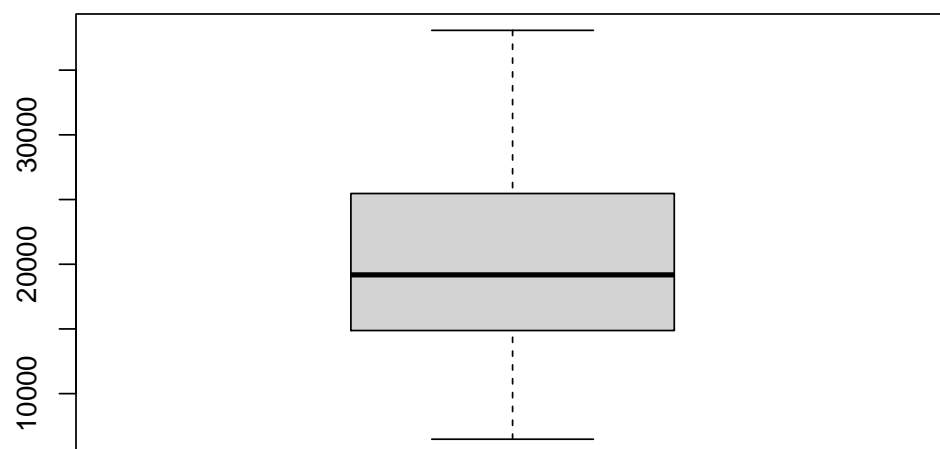
```
#boxplot
b1 <- boxplot(olivar$areacosechada)
```



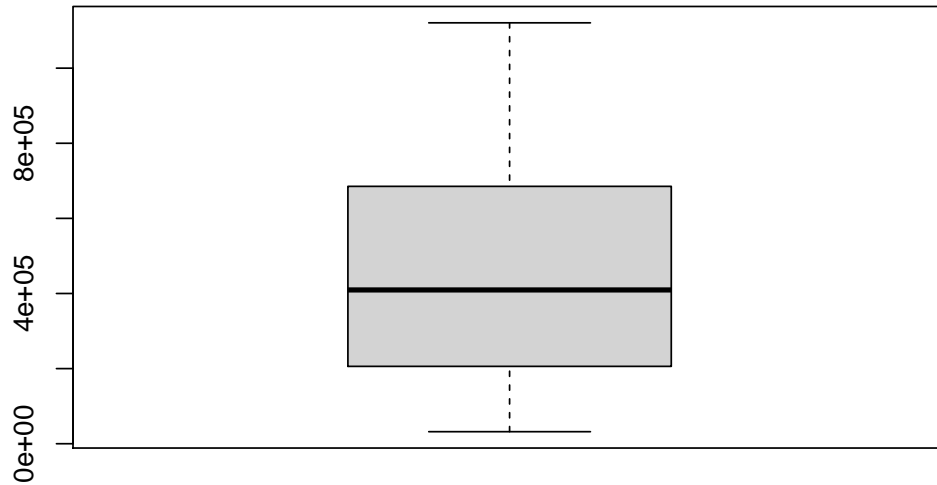
```
b2 <- boxplot(olivar$produccion)
```



```
b3 <- boxplot(olivar$rendimiento)
```

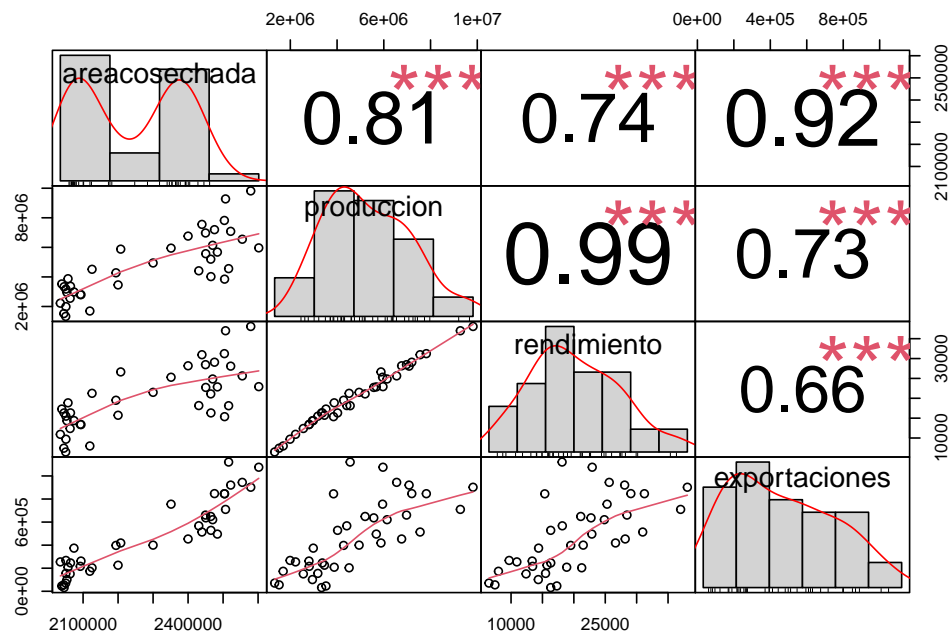


```
b4 <- boxplot(olivar$exportaciones)
```



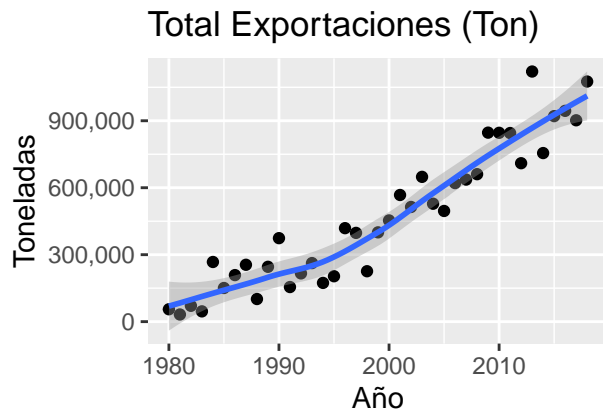
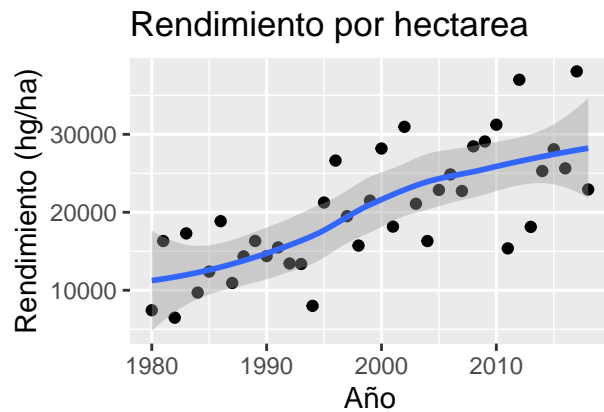
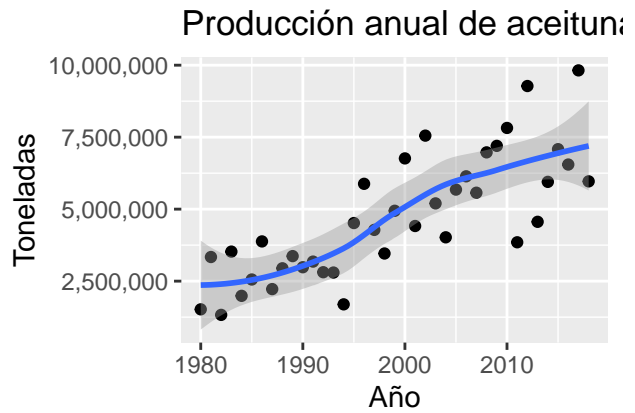
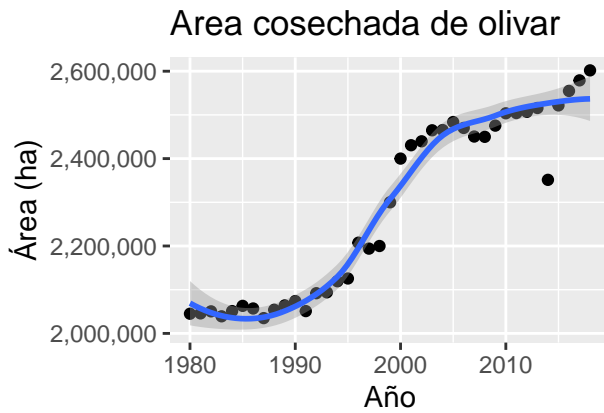
```
#eliminación error grosero
olivar<-subset(olivar,areacosechada!=1156500)

#Gráficos de correlación
chart.Correlation(olivar)
```



## Gráficas

You can also embed plots, for example:



```
cor(olivar$produccion, olivar$rendimiento)
```

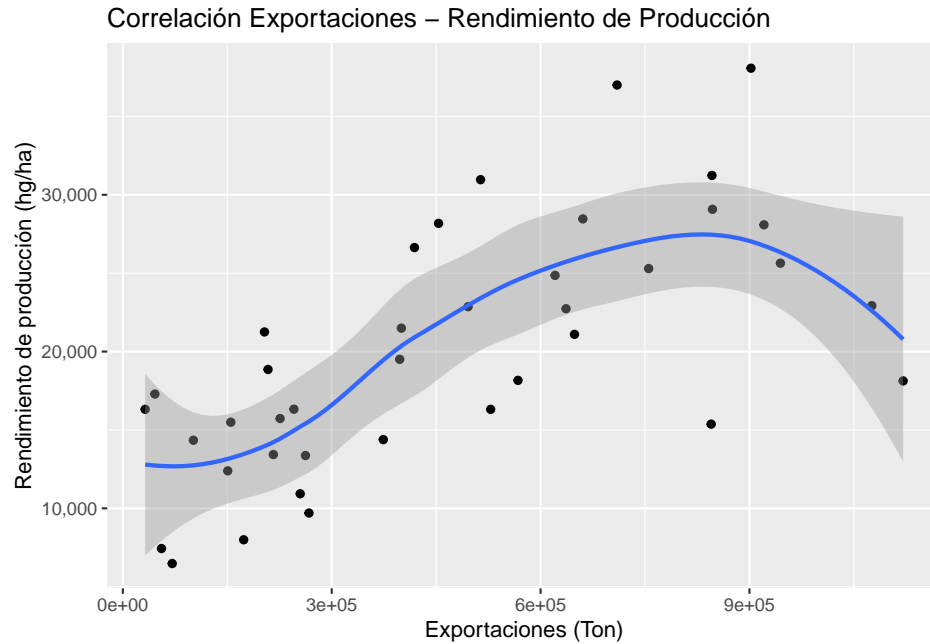
```
## [1] 0.9914391
```

```
modelo <- lm(exportaciones~produccion, data = olivar)
summary(modelo)
```

```
##
## Call:
## lm(formula = exportaciones ~ produccion, data = olivar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -297355 -134592  -36147   98809  665979
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.584e+04  8.478e+04  -0.423   0.675
## produccion   1.076e-01  1.645e-02   6.538 1.18e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 215300 on 37 degrees of freedom
## Multiple R-squared:  0.536, Adjusted R-squared:  0.5235
## F-statistic: 42.74 on 1 and 37 DF, p-value: 1.179e-07
```

```
ggplot(olivar, aes(x = exportaciones, y = rendimiento)) +
  geom_point() +
  stat_smooth() +
  scale_y_continuous(labels = scales::comma) +
  xlab("Exportaciones (Ton)") +
  ylab("Rendimiento de producción (hg/ha)") +
  ggtitle("Correlación Exportaciones - Rendimiento de Producción")
```

## 'geom\_smooth()' using method = 'loess' and formula 'y ~ x'



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.