



Universidad
Internacional
Menéndez Pelayo

MÁSTER UNIVERSITARIO EN INVESTIGACIÓN EN
INTELIGENCIA ARTIFICIAL

TRABAJO SERIES TEMPORALES

*Datos temporales
y complejos*

Laura Rodríguez Navas

Diciembre de 2020

rodrigueznava@posgrado.uimp.es

Índice general

Índice de Figuras	II
Índice de Tablas	III
1. Análisis de la serie temporal	1
1.1. Introducción	1
1.2. Datos de la serie temporal	2
1.3. Características de la serie temporal	3
1.4. Componentes de la serie temporal	4
1.4.1. Análisis de autocorrelación y autocorrelación parcial	5
1.4.2. Descomposición de la serie	6
2. Búsqueda de outliers	7
2.1. Conclusiones	8
3. Método de predicción	9
4. Aplicación del método de predicción	10
A. Apéndice: Título del Apéndice	11
A.1. Primera sección	11
B. Apéndice: Título del Apéndice	12
B.1. Primera sección	12
Bibliografía	13

Índice de figuras

Índice de tablas

Capítulo 1

Análisis de la serie temporal

1.1. Introducción

Una empresa tecnológica cuya área de negocio es la inteligencia artificial es contratada por una empresa eléctrica para que diseñe un sistema de recomendación que haga ofertas personalizadas a sus clientes sobre los paquetes energéticos más adecuados a sus consumos. Para ello, la empresa primero debe llevar a cabo un análisis exhaustivo de los consumos energéticos y diseñar una técnica de predicción que sea capaz de predecir dichos consumos con un horizonte temporal dado. La empresa eléctrica suministra a la empresa tecnológica para dicho cometido los consumos eléctricos desde el 01 de enero de 2015 hasta el 31 de diciembre de 2015 medidos con una frecuencia temporal de 10 minutos.

Se pide:

1. Analizar la serie temporal de consumos eléctricos y describir brevemente las principales características de esta.
2. Realizar un estudio para determinar si la serie temporal presenta *outliers*. Describir brevemente el estudio realizado y las conclusiones alcanzadas.
3. Seleccionar un artículo publicado en una revista científica de prestigio internacional en el que se presente un método de predicción. Estudiar de manera detallada el método publicado y describirlo de forma resumida.
4. Aplicar el método seleccionado en el apartado anterior para obtener una predicción del consumo eléctrico con un horizonte temporal de 4 horas para los días de la semana desde el lunes 8 de junio hasta el domingo 14 de junio. Para ello, se debe implementar un método de predicción en el lenguaje de programación que se estime oportuno o bien usar software libre disponible como WEKA, R, KEEL, etc. Una vez obtenidas las predicciones del periodo indicado, visualizar los resultados (predicciones, errores, etc.).

El trabajo de investigación se realiza utilizando el lenguaje de programación **R** y todos los datos y el código necesarios para reproducir el estudio se proporcionan dentro de este documento para garantizar que éste sea completamente reproducible.

Se han utilizado los siguientes paquetes para el análisis:

```
library(tidyverse)
library(ggplot2)
library(lubridate)
library(xts)
library(forecast)
```

1.2. Datos de la serie temporal

Los datos están disponibles para descargar [aquí](#). Una vez descargados, los cargamos en nuestro espacio de trabajo.

```
data <- read.csv("Demanda_2015.csv", header = FALSE, sep = ",")
```

Para su uso posterior, el marco de datos de la serie temporal se formatea.

TODO

```
#
colnames(data) <- c("date", "time", "demand")
```

```
#
head(data, 10)
```

```
##           date time demand
## 1  01/01/2015 0:00  25459
## 2  01/01/2015 0:10  25591
## 3  01/01/2015 0:20  25531
## 4  01/01/2015 0:30  25453
## 5  01/01/2015 0:40  25329
## 6  01/01/2015 0:50  25247
## 7  01/01/2015 1:00  25093
## 8  01/01/2015 1:10  24853
## 9  01/01/2015 1:20  24678
## 10 01/01/2015 1:30  24391
```

```
#
summary(data)
```

```
##           date           time           demand
## Length:52560      Length:52560      Min.    :17985
## Class :character   Class :character   1st Qu.:24392
## Mode  :character   Mode  :character   Median :28566
##                                     Mean   :28349
##                                     3rd Qu.:31664
##                                     Max.   :40648
```

Tenemos la variable *date* como *character*, la variable *time* como *character* y la variable *demand* como *integer*.

```
sum(is.na(data))
```

```
## [1] 0
```

```
#
data <- cbind(datetime = paste(data$date, data$time), data)
data$datetime <- parse_date_time(data$datetime, "dmY HMS",
                                truncated = 3, tz = "UTC")
```

```
#
data$date <- NULL
data$time <- NULL
```

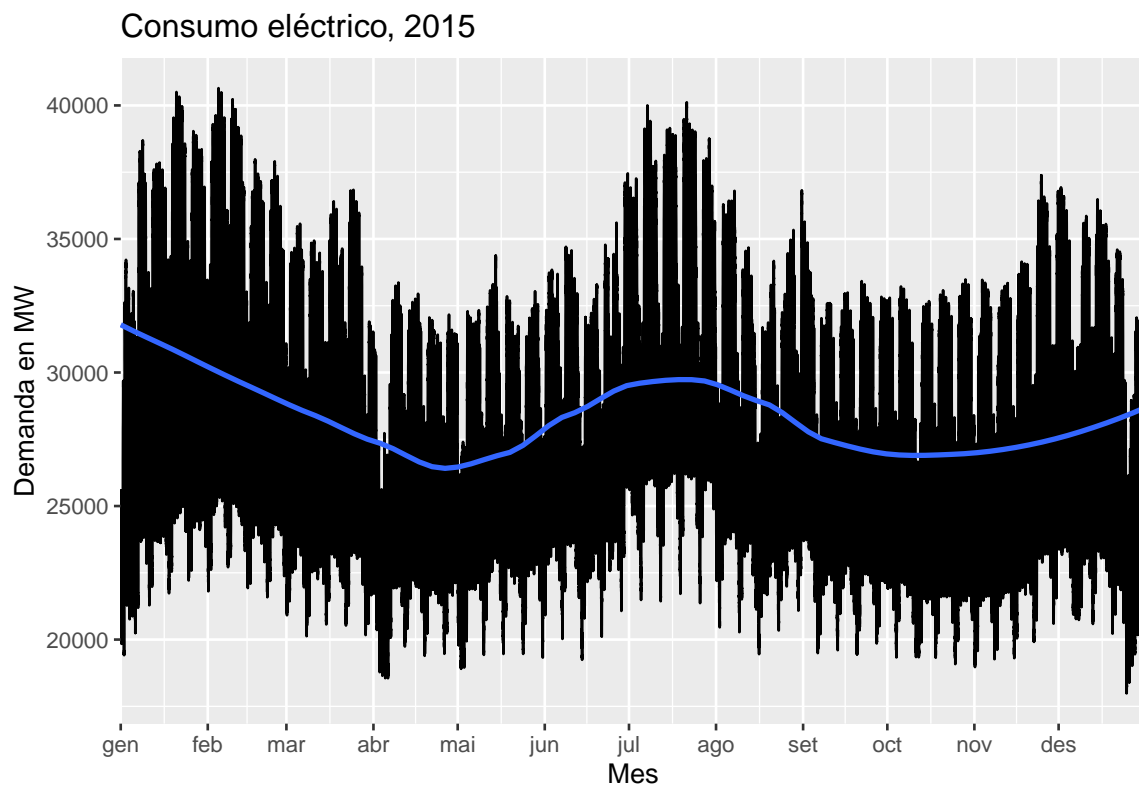
```
#
head(data, 10)
```

```
##          datetime demand
## 1 2015-01-01 00:00:00 25459
## 2 2015-01-01 00:10:00 25591
## 3 2015-01-01 00:20:00 25531
## 4 2015-01-01 00:30:00 25453
## 5 2015-01-01 00:40:00 25329
## 6 2015-01-01 00:50:00 25247
## 7 2015-01-01 01:00:00 25093
## 8 2015-01-01 01:10:00 24853
## 9 2015-01-01 01:20:00 24678
## 10 2015-01-01 01:30:00 24391
```

1.3. Características de la serie temporal

La siguiente figura ayuda a visualizar la evolución de la demanda a lo largo del año 2015.

```
ggplot(data = data, aes(x = datetime, y = demand)) +
  geom_line() +
  geom_smooth(method = "loess", se = FALSE, span = 0.6) +
  scale_x_datetime(date_labels = "%b",
                   breaks = "1 month",
                   expand = c(0, 0)) +
  ggtitle("Consumo eléctrico, 2015") +
  xlab("Mes") +
  ylab("Demanda en MW")
```

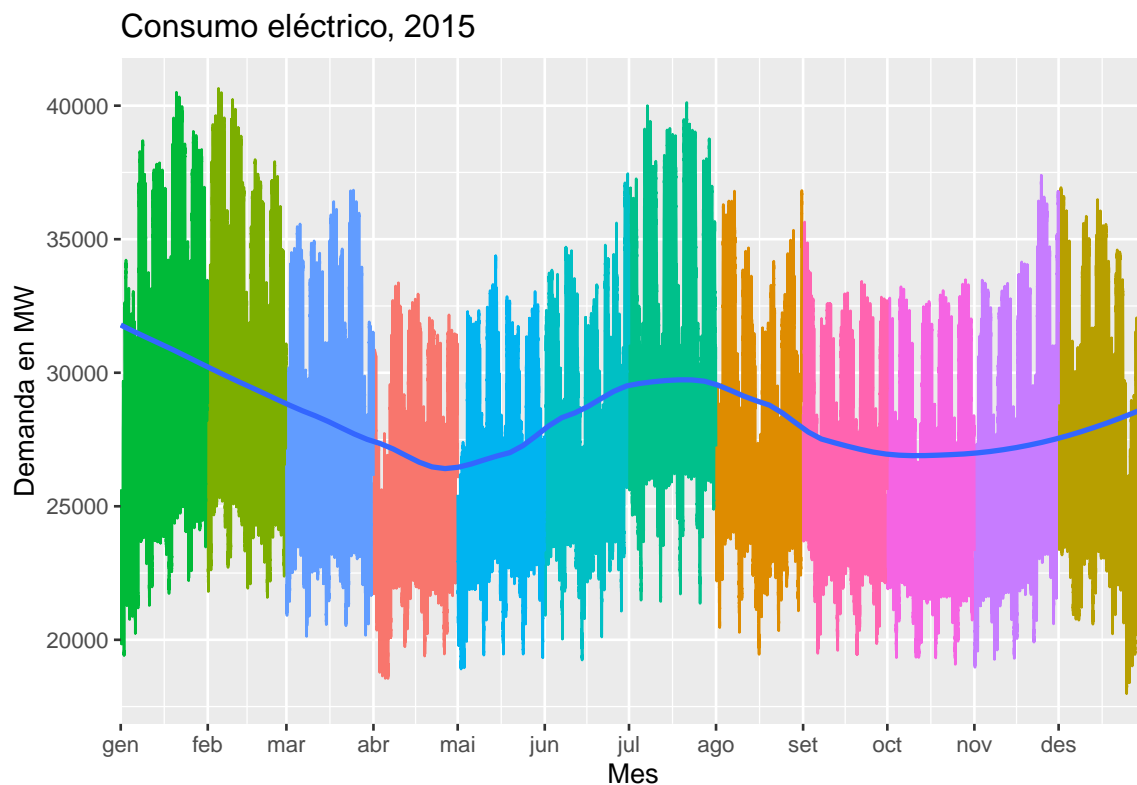


Observando la figura se puede detectar una dependencia estacional de la demanda, aunque pueden existir otros factores que pueden afectarla, como los días festivos, los fines de semana, etc.

Si queremos comparar mejor la distribución de la demanda de energía para cada mes, la

siguiente figura es muy útil.

```
ggplot(data, aes(x = datetime, y = demand)) +
  geom_line(aes(colour = month)) +
  geom_smooth(method = "loess", se = FALSE, span = 0.6) +
  scale_x_datetime(date_labels = "%b",
                   breaks = "1 month",
                   expand = c(0, 0)) +
  ggtitle("Consumo eléctrico, 2015") +
  xlab("Mes") +
  ylab("Demanda en MW") +
  theme(legend.position = "none")
```



Durante el invierno y el verano la demanda es de manera clara mayor salvo, probablemente, en los períodos vacacionales. Como se puede observar en el mes de agosto. Podemos deducir de la figura que en sí mismos los datos siguen un patrón de subidas y caídas en una tendencia que varía sumo en los meses de invierno. Esto se debe a que las personas usan la electricidad para enfriar y calentar sus viviendas respectivamente.

1.4. Componentes de la serie temporal

Es frecuente analizar las series temporales desde el punto de vista de sus componentes estructurales: *Serie observada = Tendencia + Efecto estacional + Residuos*.

Si queremos que **R** trate un objeto como una serie temporal, tenemos que determinar apropiadamente sus características con el comando *ts*. Para definir la serie correctamente escribimos:

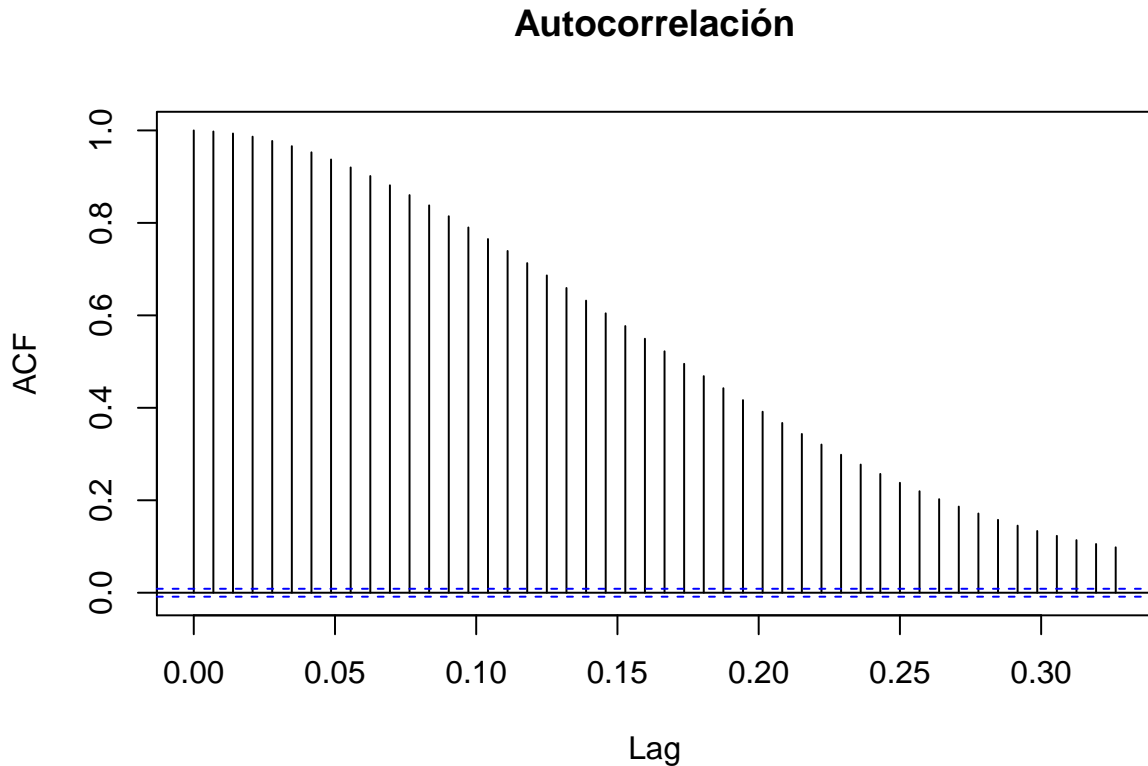
```
ts <- ts(data$demand, frequency = 24*60/10)
```

El argumento *frequency* se utiliza para indicar la periodicidad de la serie (en este caso de 10 minutos).

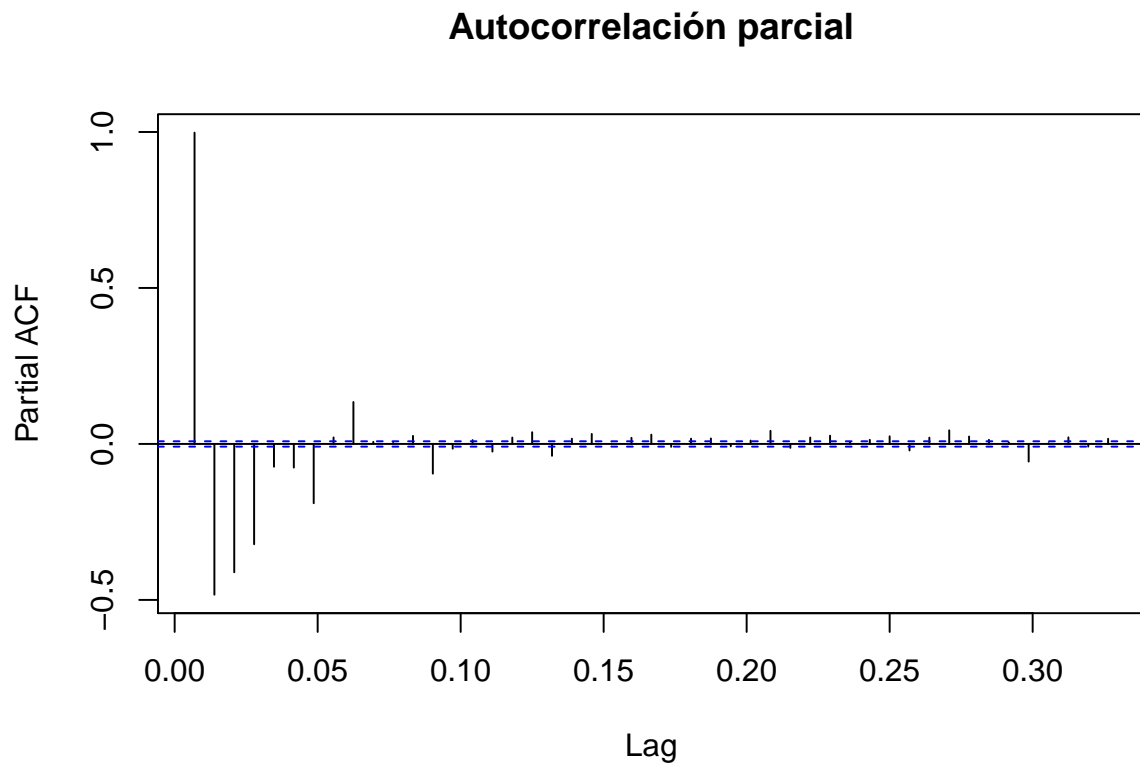
1.4.1. Análisis de autocorrelación y autocorrelación parcial

El análisis de correlación y autocorrelación parcial se usa para examinar la dependencia de la serie.

```
acf(ts, main = "Autocorrelación")
```



```
pacf(ts, main = "Autocorrelación parcial")
```



La función de autocorrelación muestra un proceso no estacionario altamente autocorrelacionado,

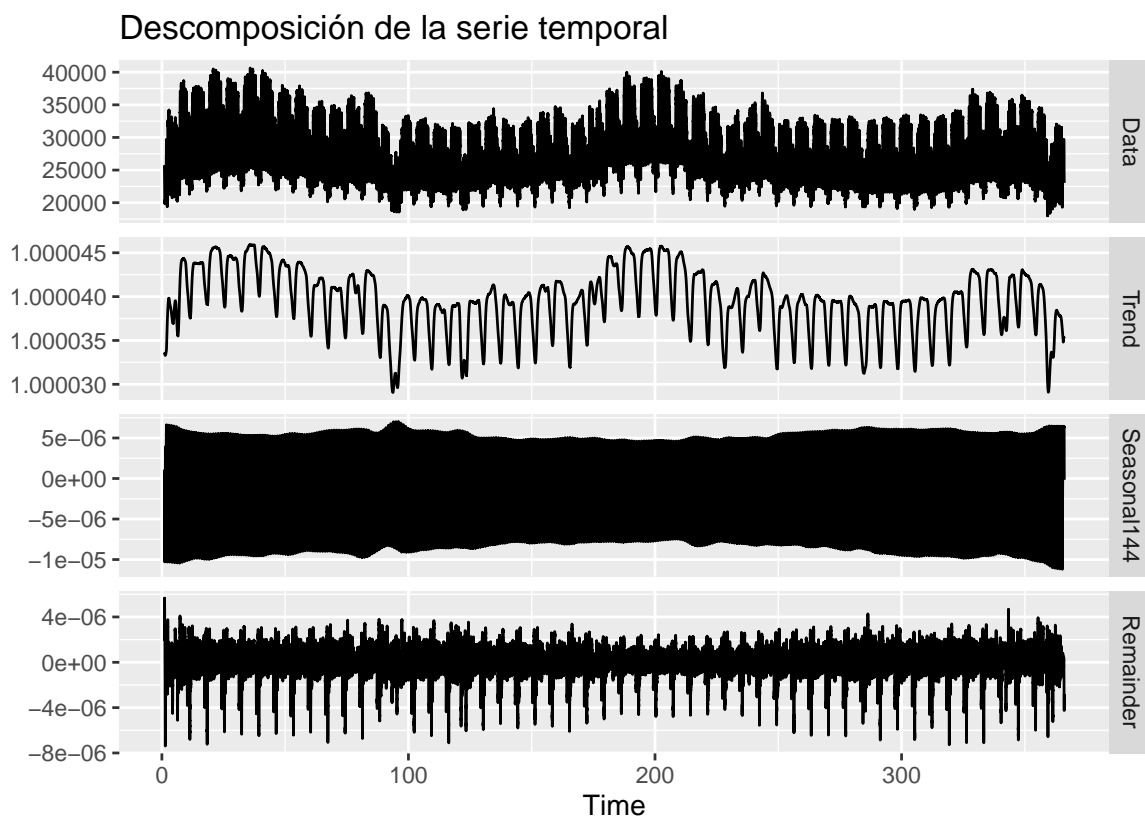
y destaca la estacionalidad. La función de autocorrelación parical es más difícil de interpretar, pero debido a la forma de *tail-off*, donde los valores cada vez se van haciendo más pequeños, parece un comportamiento típico de los modelos **ARIMA**.

1.4.2. Descomposición de la serie

En **R** es muy sencillo y existen diferentes opciones para obtener la descomposición estructural de la serie temporal. Por ejemplo, se puede usar la combinación de los paquetes *mstl* y *autoplot*.

A continuación se representa la descomposición de la serie temporal:

```
# Múltiple estacionalidad
mstl(ts, lambda = "auto") %>%
  autoplot(main = "Descomposición de la serie temporal")
```



Como puede verse, los datos de la serie temporal *Data* son descompuestos en:

- **Trend** o tendencia, se calcula con una media móvil. En la figura, la tendencia general de los datos varía suavemente.
- **Seasonal144**, es el efecto estacional fijado.
- **Remainder**, son los residuos de la serie, se obtienen restando a la serie temporal observada *Data* las dos componentes anteriores. En la figura, parece que existen bastantes residuos, será difícil que un modelo estacionario pueda modelar bien si no se transforma la serie temporal de manera que el modelo estacionario sea apropiado para la serie transformada.

Capítulo 2

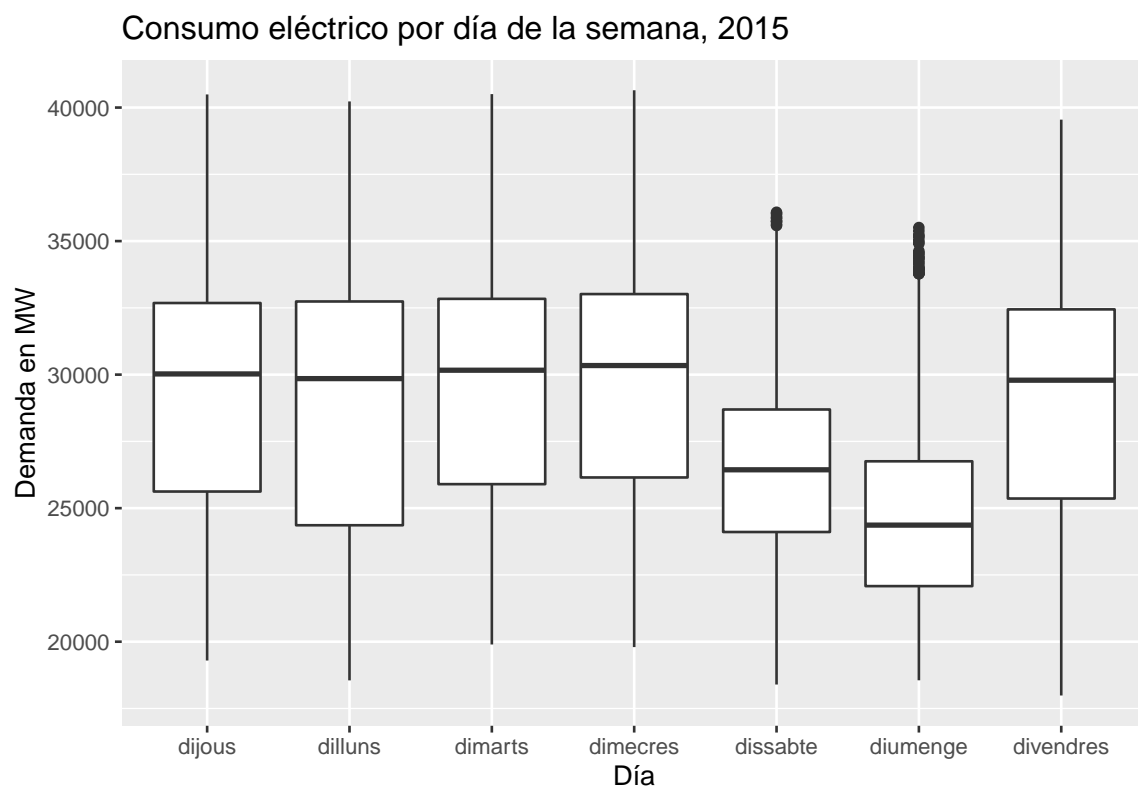
Búsqueda de outliers

Se han utilizado los siguientes paquetes para la búsqueda de outliers:

```
library(tsoutliers)
```

Durante los fines de semana, la demanda disminuye considerablemente en comparación con el resto de días de la semana, además observamos la presencia de *outliers*.

```
ggplot(data, aes(day, demand)) +  
  geom_boxplot() +  
  ggtitle("Consumo eléctrico por día de la semana, 2015") +  
  xlab("Día") +  
  ylab("Demanda en MW")
```



Como hemos comentado en el capítulo anterior,

```
# outliers <- tsoutliers(ts, iterate = 2, lambda = "auto")  
# plot(outliers)
```

2.1. Conclusiones

Capítulo 3

Método de predicción

Capítulo 4

Aplicación del método de predicción

Se aplica el método de predicción descrito en el capítulo anterior para obtener una predicción del consumo eléctrico con un horizonte temporal de 4 horas para los días de la semana desde el lunes 8 de junio hasta el domingo 14 de junio.

Apéndice A

Apéndice: Título del Apéndice

A.1. Primera sección

Apéndice B

Apéndice: Título del Apéndice

B.1. Primera sección

Bibliografía

Pedro L. Luque-Calvo. *Escribir un Trabajo Fin de Estudios con R Markdown*, 2017.