

# Introducción a R. Ejercicio U02\_E03

Entendiendo la factura de la luz. Segunda parte

## 1. Lectura de ficheros

### Directorio de Trabajo

Cuando queremos leer y escribir ficheros desde R es conveniente especificar el directorio de trabajo.

```
setwd("RutaDirectorioCurso/02IntroduccionR/Ejercicios/")
```

Leer el fichero consumo de electricidad horaria, de un cliente de una compañía eléctrica (Fichero “consumo\_horario\_cliente.csv” en la carpeta Ejercicios de la unidad 2)

```
consumo=read.delim("consumo_horario_cliente.csv",sep=";")
```

## 2. Fechas

La columna datetime es una marca de tiempo, es un dato que debería tener formato fecha-hora. *read.table* o *read.delim* por defecto leen este tipo de datos como texto, más aun por defecto convierten dicho texto a un factor. Para convertir este dato a un tipo fecha-hora (POSIXct) lo podemos hacer durante la lectura o a posteriori

```
consumo$datetime = as.POSIXct(consumo$datetime)
# o con el argumento colClasses de la función read.delim. Hay que especificar el tipo de dato de todas
consumo=read.delim("consumo_horario_cliente.csv",sep=";", colClasses = c("POSIXct","numeric","numeric"))
```

## 3. Calculo de variables derivadas

Vamos a usar el comando *transform* para calcular algunas magnitudes derivadas:

- `coste_eur`: coste en euros de la electricidad consumida cada hora. Se calcula multiplicando el consumo en cada hora por su precio.
- Fecha: usando la conversión del tipo de dato *POSIXct* a tipo *Date*: `as.Date()`
- Hora: usando el comando `format`, que permite cambiar el formato a una fecha en su conversión a cadena de texto. En la ayuda de la función `format.POSIXct` encontrarás como hacerlo.
- Mes del año: usando el comando `format`, que permite cambiar el formato a una fecha en su conversión a cadena de texto
- Día de la semana: usando la función `weekdays`

## 4. Filtrado

Usemos las herramientas de filtrado para conocer:

- Las 5 primeras filas
- Las 3 primeras filas y las dos primeras columnas
- Las últimas 5 filas (el comando `nrow(df)` te dice el número de filas en el data frame)

- Los consumos del 31 de diciembre de 2016.
- Fecha-hora del mayor consumo. ¿Coincide con el mayor coste?
- Menor consumo
- Consumos por debajo de 20 watios-hora (0.02 kwh)
- Calcular el valor medio del consumo y del coste los martes y los viernes

## 5. Reetiquetado de variables

- Convertir la variable *wd* a un factor y especificar los niveles de forma explícita, usando el orden de la semana y no el orden alfabético que elige por defecto (`levels=c("lunes", "martes", "miércoles", "jueves", "viernes", "sábado", "domingo")`)
- Cambiemos los nombres a los niveles, cambiándolos a la letra inicial: L,M,X,J,V,S,D

## 6. Agregación

- Consumo y coste total por mes. ¿Es el mes de mayor consumo el de mayor coste?

## 7. Union de data frames

Cargamos los datos de precio del mercado mayorista, parecidos a los que usamos en la teoría de la unidad 2, del fichero "precio\_md.csv" en el mismo directorio que el enunciado del ejercicio.

Calcula las diferencias entre el precio minorista, el que paga el consumidor, respecto al precio del mercado mayorista y calcula las diferencias medias mensuales