Course Project 1

Aaron Calderon

2023-07-30

Procesamiento

```r
knitr::opts_chunk$set(echo = TRUE)
setwd("D:/ONLINE/COURSERA/EXPLORATORY DA/COURSERA_EDA")
unzip("exdata_data_household_power_consumption.zip")
data <- read.csv("household_power_consumption.txt",  sep = ";")

data$Date2 <- as.Date(as.character(data$Date), format = "%d/%m/%Y")
data$Datetime <- as.POSIXct(paste(data$Date, data$Time), format = "%d/%m/%Y %H:%M:%S")

data[data$Global_active_power=='?', "Global_active_power"] = NA
data[data$Global_reactive_power=='?', "Global_reactive_power"] = NA
data[data$Voltage=='?', "Voltage"] = NA
data[data$Global_intensity=='?', "Global_intensity"] = NA
data[data$Sub_metering_1=='?', "Sub_metering_1"] = NA
data[data$Sub_metering_2=='?', "Sub_metering_2"] = NA
# data[data$Sub_metering_3=='?', "Sub_metering_3"] = NA

data$Global_active_power <- as.numeric(data$Global_active_power)
data$Global_reactive_power <- as.numeric(data$Global_reactive_power)
data$Voltage <- as.numeric(data$Voltage)
data$Global_intensity <- as.numeric(data$Global_intensity)
data$Sub_metering_1 <- as.numeric(data$Sub_metering_1)
data$Sub_metering_2 <- as.numeric(data$Sub_metering_2)
data$Sub_metering_3 <- as.numeric(data$Sub_metering_3)

dataf <- subset(data, Date2 >= as.Date("2007-02-01") & Date2 <= as.Date("2007-02-02"))
```

Variables

URL : Dataset Electric power consumption

- Date: Date in format dd/mm/yyyy

- Time: time in format hh:mm:ss

- Global_active_power: household global minute-averaged active power (in kilowatt)

- Global_reactive_power: household global minute-averaged reactive power (in kilowatt)

- Voltage: minute-averaged voltage (in volt)

- Global_intensity: household global minute-averaged current intensity (in ampere)

- Sub_metering_1: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).

- Sub_metering_2: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.

- Sub_metering_3: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

```
head(dataf, 5)
```

```
##           Date     Time Global_active_power Global_reactive_power Voltage
## 66637 1/2/2007 00:00:00               0.326                 0.128  243.15
## 66638 1/2/2007 00:01:00               0.326                 0.130  243.32
## 66639 1/2/2007 00:02:00               0.324                 0.132  243.51
## 66640 1/2/2007 00:03:00               0.324                 0.134  243.90
## 66641 1/2/2007 00:04:00               0.322                 0.130  243.16
##       Global_intensity Sub_metering_1 Sub_metering_2 Sub_metering_3      Date2
## 66637              1.4              0              0              0 2007-02-01
## 66638              1.4              0              0              0 2007-02-01
## 66639              1.4              0              0              0 2007-02-01
## 66640              1.4              0              0              0 2007-02-01
## 66641              1.4              0              0              0 2007-02-01
##                  Datetime
## 66637 2007-02-01 00:00:00
## 66638 2007-02-01 00:01:00
## 66639 2007-02-01 00:02:00
## 66640 2007-02-01 00:03:00
## 66641 2007-02-01 00:04:00
```

```
as.data.frame(colSums(is.na(dataf)))
```

```
##                       colSums(is.na(dataf))
## Date                                      0
## Time                                      0
## Global_active_power                       0
## Global_reactive_power                     0
## Voltage                                   0
## Global_intensity                          0
## Sub_metering_1                            0
## Sub_metering_2                            0
## Sub_metering_3                            0
## Date2                                     0
## Datetime                                  0
```
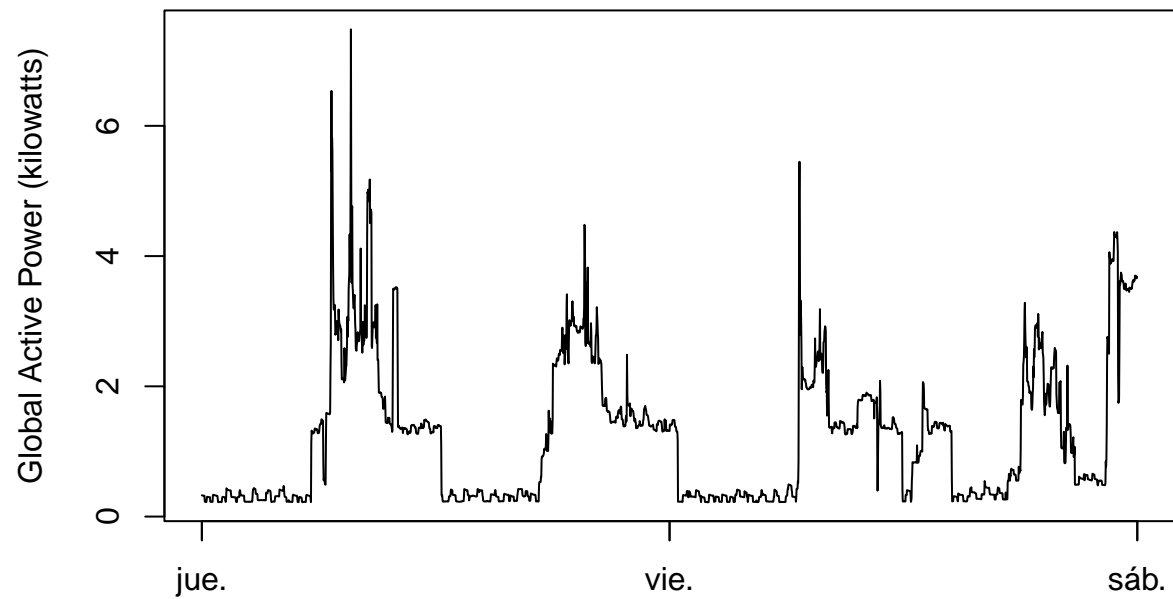
Plot 1

```
hist(x = dataf$Global_active_power,
     breaks = 12,
     col = "red",
     main = "Global Active Power",
     xlab = "Global Active Power (kilowatts)"
     )
```

Global Active Power
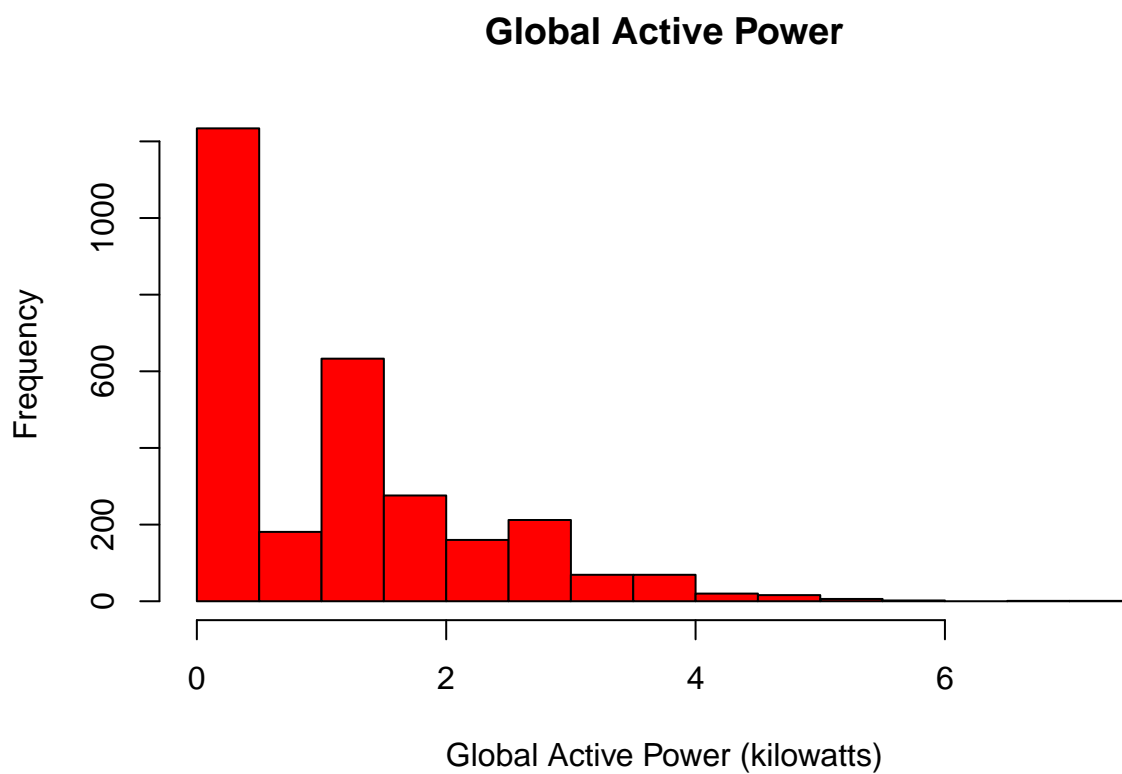
Plot 2

```
# dev.new(width = 480, height = 480, unit = "px")
plot(x = dataf$Datetime,
     y = dataf$Global_active_power,
     type = "l",
     main = "",
     ylab = "Global Active Power (kilowatts)",
     xlab = ""
     )
```
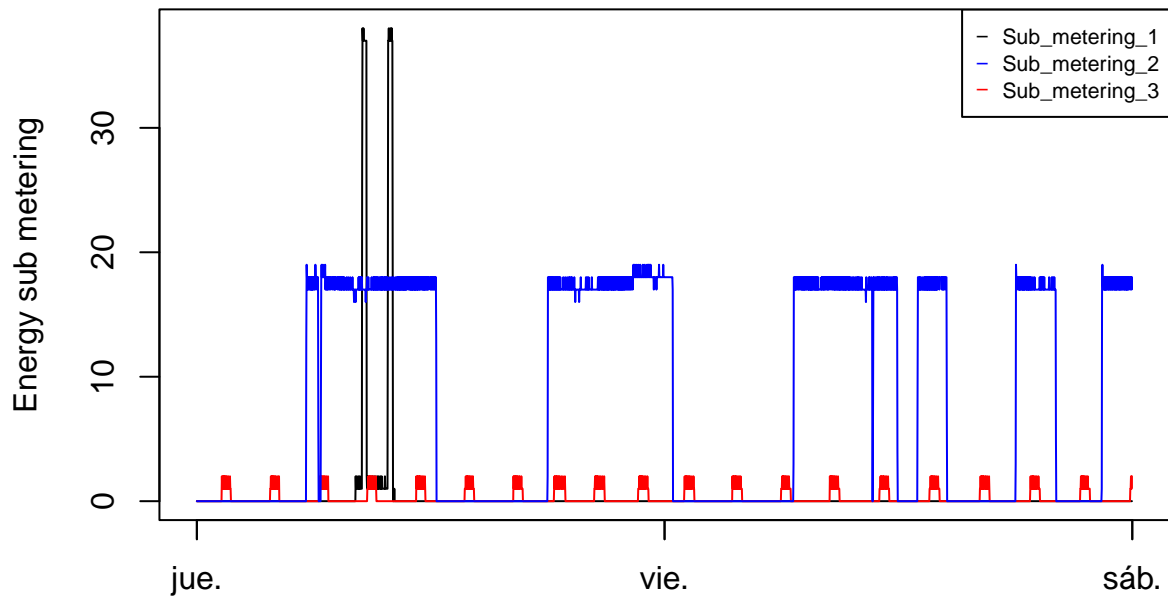
Plot 3

```r
# dev.new(width = 480, height = 480, unit = "px")
plot(x = dataf$Datetime,
     y = dataf$Sub_metering_1,
     type = "l",
     main = "",
     ylab = "Energy sub metering",
     xlab = "",
     col = "black"
     )
points(x = dataf$Datetime,
     y = dataf$Sub_metering_2,
     type = "l",
     col = "red"
     )
points(x = dataf$Datetime,
     y = dataf$Sub_metering_3,
     type = "l",
     col = "blue"
     )
legend(x = "topright",
       cex = 0.7,
```

```
      pch = "-",
      col = c("black", "blue", "red"),
      legend = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3")
      )
```


Plot 4

```
# dev.new(width = 480, height = 480, unit = "px")
par(mfrow = c(2, 2), mar = c(4, 4, 2, 1), oma = c(0, 0, 2, 0))
with(dataf, {
  plot(Datetime, Global_active_power, ylab = "Global Active Power", type = "l", xlab = "")
  plot(Datetime, Voltage, ylab = "Voltage", type = "l")
  plot(x = dataf$Datetime,
      y = dataf$Sub_metering_1,
      type = "l",
      main = "",
      ylab = "Energy sub metering",
      xlab = "",
      col = "black"
      )
  points(x = dataf$Datetime,
        y = dataf$Sub_metering_2,
        type = "l",
```

```
        col = "red"
        )
  points(x = dataf$Datetime,
         y = dataf$Sub_metering_3,
         type = "l",
         col = "blue"
         )
  legend(x = "topright",
         cex = 0.6,
         pch = "-",
         col = c("black", "blue", "red"),
         legend = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"), bty = "n"
         )
  plot(Datetime, Global_reactive_power, ylab = "Global Reactive Power", type = "l")
})
```