Course Project 1

Jerez

11/6/2020

Course: Exploratory Data Analysis

Project 1

#Part 1

Load and clean the table.

```
setwd("~/Google Drive/Coursera") # Set working directory

data <- read.table(file = "household_power_consumption.txt", header = TRUE, sep = ';', colClasses = c('
```

The variables of data are:

1. Date: Date in format dd/mm/yyyy
2. Time: time in format hh:mm:ss
3. Global_active_power: household global minute-averaged active power (in kilowatt)
4. Global_reactive_power: household global minute-averaged reactive power (in kilowatt)
5. Voltage: minute-averaged voltage (in volt)
6. Global_intensity: household global minute-averaged current intensity (in ampere)
7. 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).
8. 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.
9. 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.

Work to do:

a) The dataset has 2,075,259 rows and 9 columns. First calculate a rough estimate of how much memory the dataset will require in memory before reading into R. Make sure your computer has enough memory (most modern computers should be fine).

b) We will only be using data from the dates 2007-02-01 and 2007-02-02. One alternative is to read the data from just those dates rather than reading in the entire dataset and subsetting to those dates.

The script should be like:

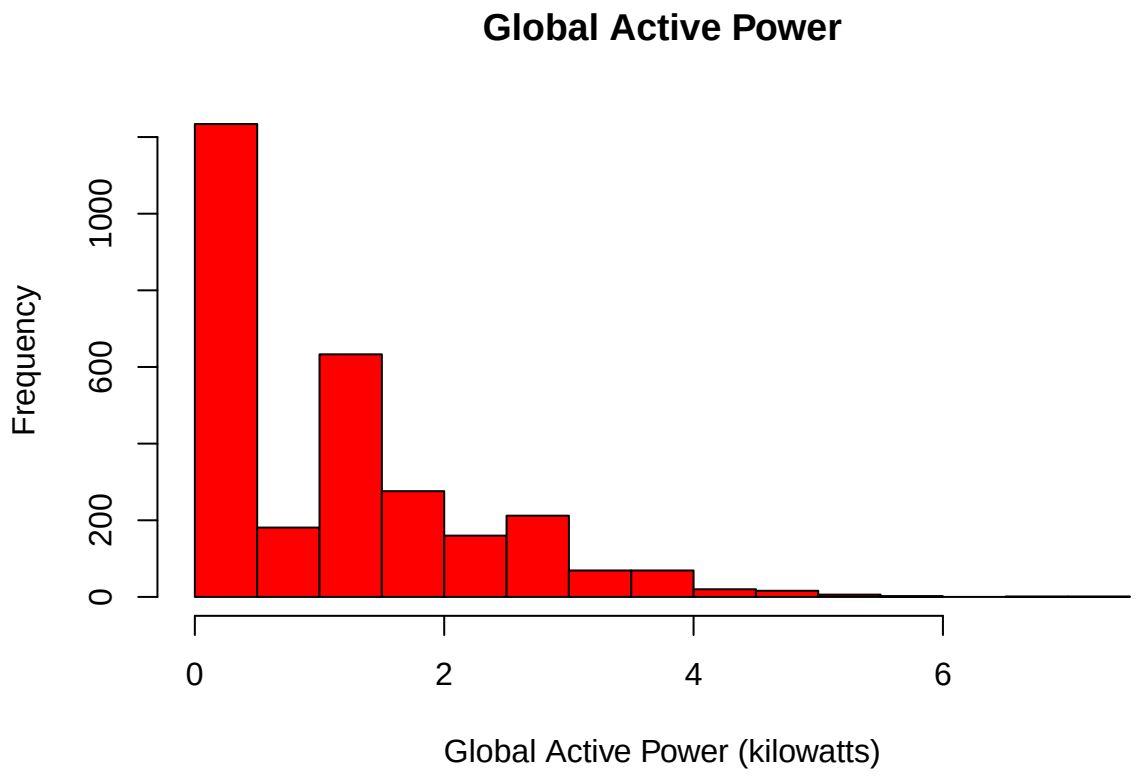
```
data$Date <- as.Date(data$Date, '%d/%m/%Y') # Format date

data.sub  <- subset(data, Date >= as.Date('2007-02-01') & Date <= as.Date('2007-02-02')) # We select th

data.sub  <- data.sub[complete.cases(data.sub), ] # remove the incomplete cases

Date.Time <- paste(data.sub$Date, data.sub$Time)
Date.Time <- setNames(Date.Time, 'Date.Time')

data.sub  <- data.sub[ ,!(names(data.sub) %in% c('Date', 'Time'))] # Remove the old Date and Time

data.sub  <- cbind(data.sub, Date.Time) # Put the new date time format
data.sub$Date.Time <- as.POSIXct(Date.Time) # Change the format
```

#Part 2

Making Plots

Plot 1: Global Active Power

```
hist(data.sub$Global_active_power, col = 'red', xlab = 'Global Active Power (kilowatts)', ylab = 'Frequ
```



```
dev.copy(png, 'plot1.png', width = 480, height = 480)
```

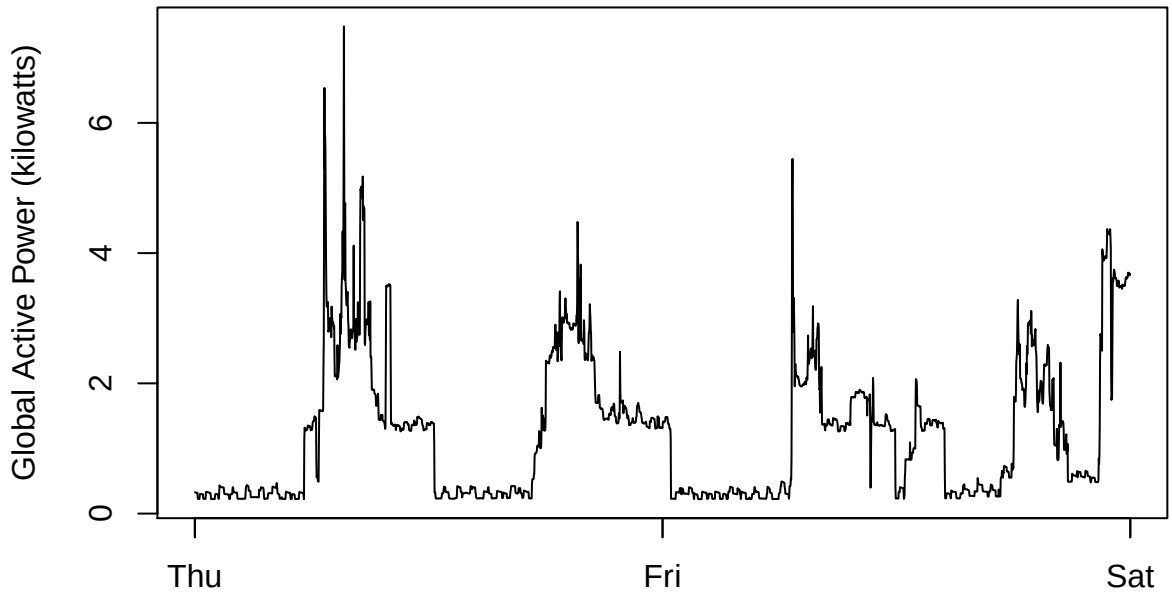
quartz_off_screen 3

```
dev.off()
```

pdf 2

Plot 2:

```
with(data.sub, plot(Global_active_power~Date.Time, type = 'l', col = 'black', xlab = ' ', ylab = 'Global
```




```
dev.copy(png, 'plot2.png', width = 480, height = 480)
```

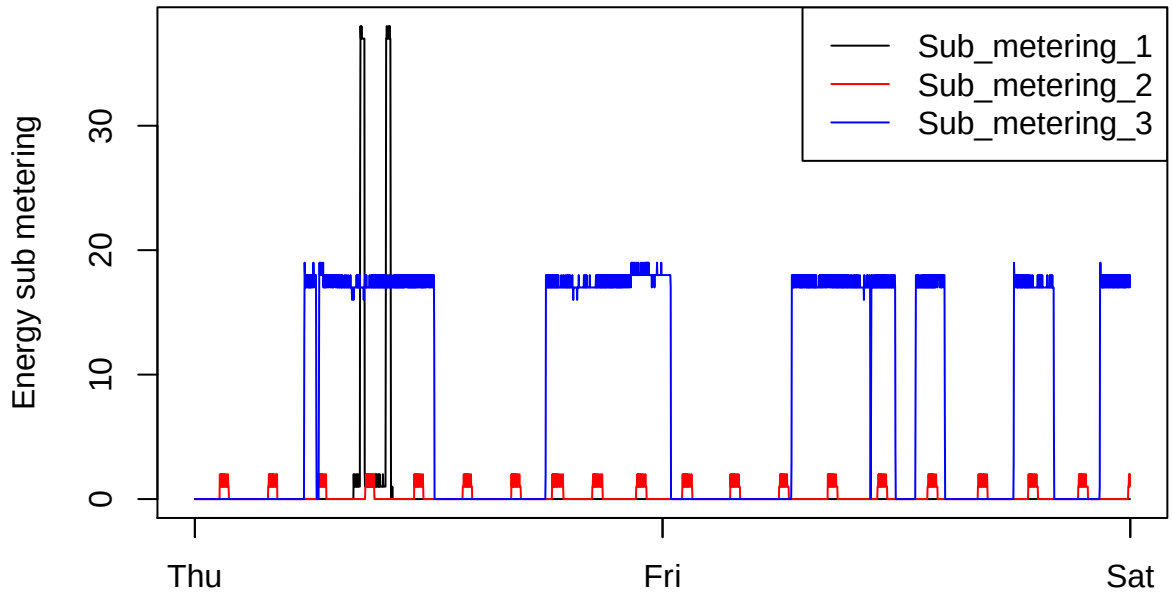
quartz_off_screen 3

```
dev.off()
```

pdf 2

Plot 3:

```
with(data.sub, {
     plot(Sub_metering_1~Date.Time, type = 'l', col = 'black', xlab = ' ',
          ylab = 'Energy sub metering')
     lines(Sub_metering_2~Date.Time, type = 'l', col = 'red')
     lines(Sub_metering_3~Date.Time, type = 'l', col = 'blue')
 })

legend('topright', lwd = c(1,1,1), col = c('black', 'red', 'blue'),
            c('Sub_metering_1',  'Sub_metering_2', 'Sub_metering_3'))
```




```
dev.copy(png, 'plot3.png', width = 480, height = 480)
```

quartz_off_screen 3

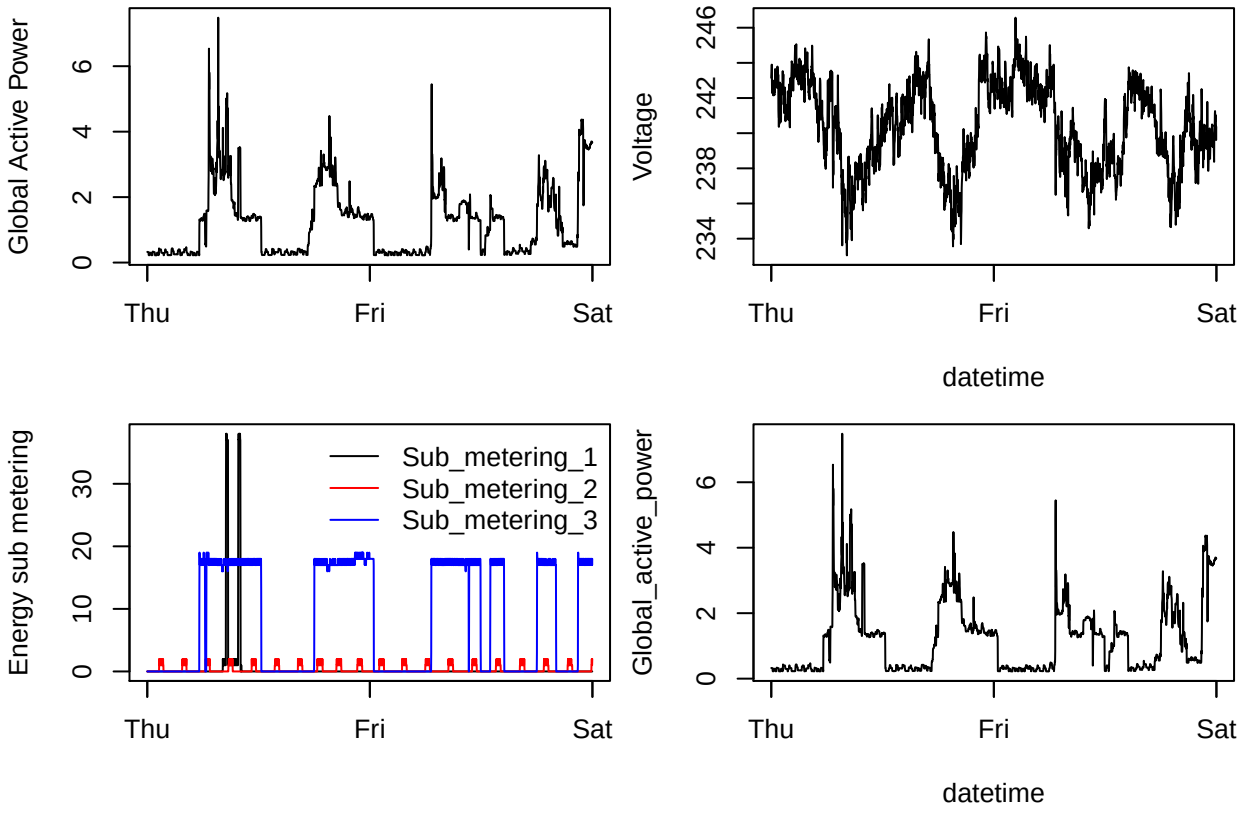
```
dev.off()
```

pdf 2

Plot 4: We need to create a layaout for the four graphs

```
par(mfrow = c(2,2), mar = c(4,4,1,0.5), oma = c(0,0,1,0))

with(data.sub, {
 plot(Global_active_power~Date.Time, type = 'l', xlab = ' ', ylab = 'Global Active Power', col = 'black
```

```
 plot(Voltage~Date.Time, type = 'l', xlab = 'datetime', ylab = 'Voltage', col = 'black')

 plot(Sub_metering_1~Date.Time, type = 'l', col = 'black', xlab = ' ',
          ylab = 'Energy sub metering')
 lines(Sub_metering_2~Date.Time, type = 'l', col = 'red')
 lines(Sub_metering_3~Date.Time, type = 'l', col = 'blue')
 legend('topright', lwd = c(1,1,1), col = c('black', 'red', 'blue'),lty =c(1,1,1),
        bty ='n', c('Sub_metering_1',  'Sub_metering_2', 'Sub_metering_3'))
 plot(Global_active_power~Date.Time, type = 'l', xlab = 'datetime', ylab = 'Global_active_power', col =
})
```




```
dev.copy(png, 'plot4.png', width = 480, height = 480)
```

quartz_off_screen 3

```
dev.off()
```

pdf 2