

Electric Power Consumption

BROM, P. C.

31 de outubro de 2016

Description

This assignment uses data from the UC Irvine Machine Learning Repository, a popular repository for machine learning datasets. In particular, we will be using the “Individual household electric power consumption Data Set” which I have made available on the course web site:

1. Dataset: Electric power consumption [20Mb]
2. Description: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years.

Different electrical quantities and some sub-metering values are available. The 9 variables in the dataset are taken from the UCI web site.

Loading the data

When loading the dataset into R, please consider the following:

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).

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.

You may find it useful to convert the Date and Time variables to Date/Time classes in R using the `strptime()` and `as.Date()` functions.

Note that in this dataset missing values are coded as `?`.

Making Plots

Our overall goal here is simply to examine how household energy usage varies over a 2-day period in February, 2007. Your task is to reconstruct the following plots below, all of which were constructed using the base plotting system.

First you will need to fork and clone the following GitHub repository: https://github.com/rdpeng/ExData_Plotting1

For each plot you should

1. Construct the plot and save it to a PNG file with a width of 480 pixels and a height of 480 pixels.
2. Name each of the plot files as `plot1.png`, `plot2.png`, etc.
3. Create a separate R code file (`plot1.R`, `plot2.R`, etc.) that constructs the corresponding plot, i.e. code in `plot1.R` constructs the `plot1.png` plot. Your code file should include code for reading the data so that the plot can be fully reproduced. You must also include the code that creates the PNG file.

4. Add the PNG file and R code file to the top-level folder of your git repository (no need for separate sub-folders) When you are finished with the assignment, push your git repository to GitHub so that the GitHub version of your repository is up to date. There should be four PNG files and four R code files, a total of eight files in the top-level folder of the repo.

```
setwd("/home/pcbrom/Dropbox/Trabalho e Estudo/Cursos Livres/Exploratory Data Analysis/p_w1")
plotData = data.table::fread("household_power_consumption.txt",
                             sep = ";", na.strings = "?")
```

```
##
Read 41.4% of 2075259 rows
Read 71.3% of 2075259 rows
Read 99.3% of 2075259 rows
Read 2075259 rows and 9 (of 9) columns from 0.124 GB file in 00:00:05
```

Adjust time variable

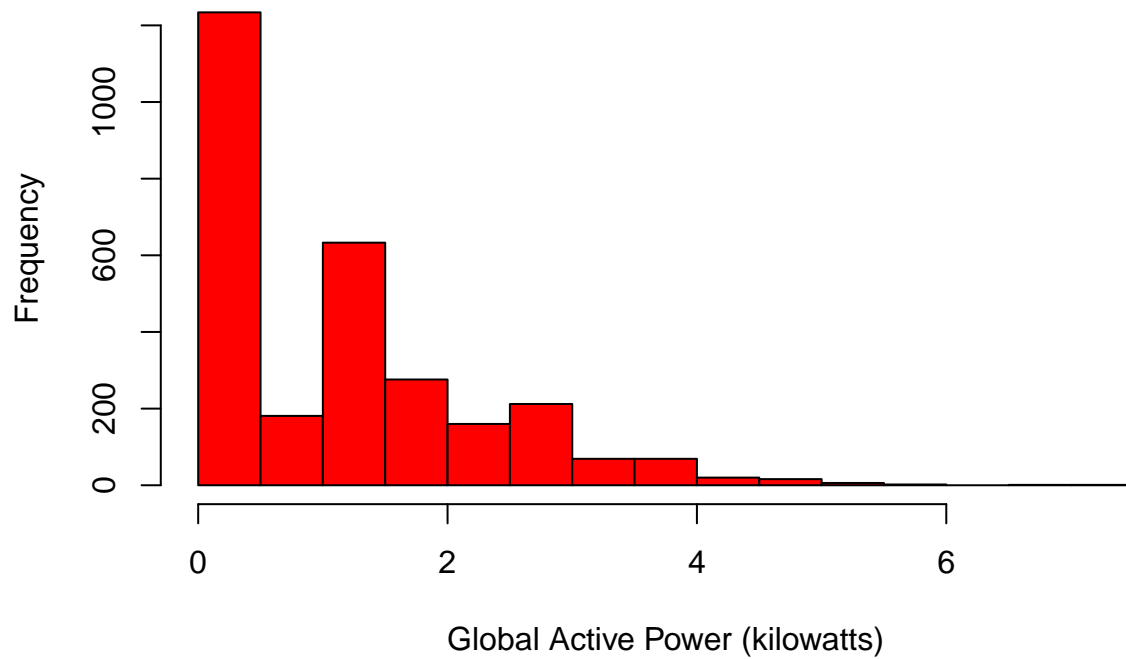
```
finalData = plotData[plotData$Date %in% c("1/2/2007", "2/2/2007"),]
SetTime = strptime(paste(finalData$Date, finalData$Time, sep = " "),
                   "%d/%m/%Y %H:%M:%S")
finalData = cbind(SetTime, finalData)
```

```
## Warning in data.table::data.table(...): POSIXlt column type detected and
## converted to POSIXct. We do not recommend use of POSIXlt at all because it
## uses 40 bytes to store one date.
```

Generating Plot 1

```
hist(finalData$Global_active_power, col="red", main = "Global Active Power",
      xlab = "Global Active Power (kilowatts)")
```

Global Active Power

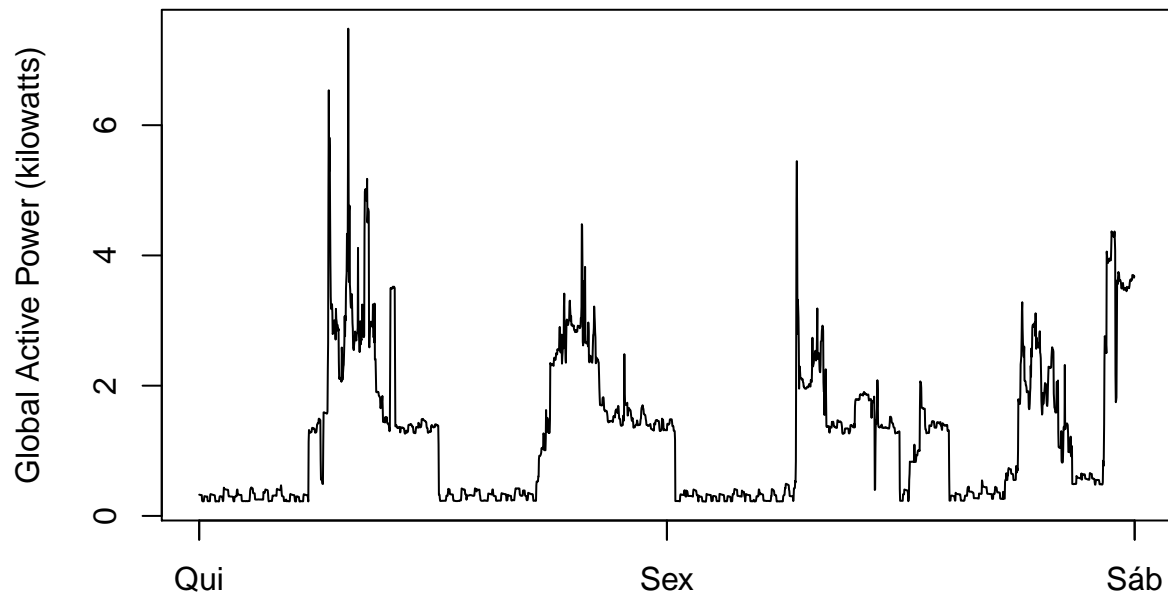


```
# Export plot1
png(filename = "plot1.png", width = 480, height = 480)
hist(finalData$Global_active_power, col="red", main = "Global Active Power",
      xlab = "Global Active Power (kilowatts)")
dev.off()
```

```
## pdf
## 2
```

Generating Plot 2

```
plot(finalData$SetTime, finalData$Global_active_power, type = "l", col = "black",
      xlab = "", ylab = "Global Active Power (kilowatts)")
```

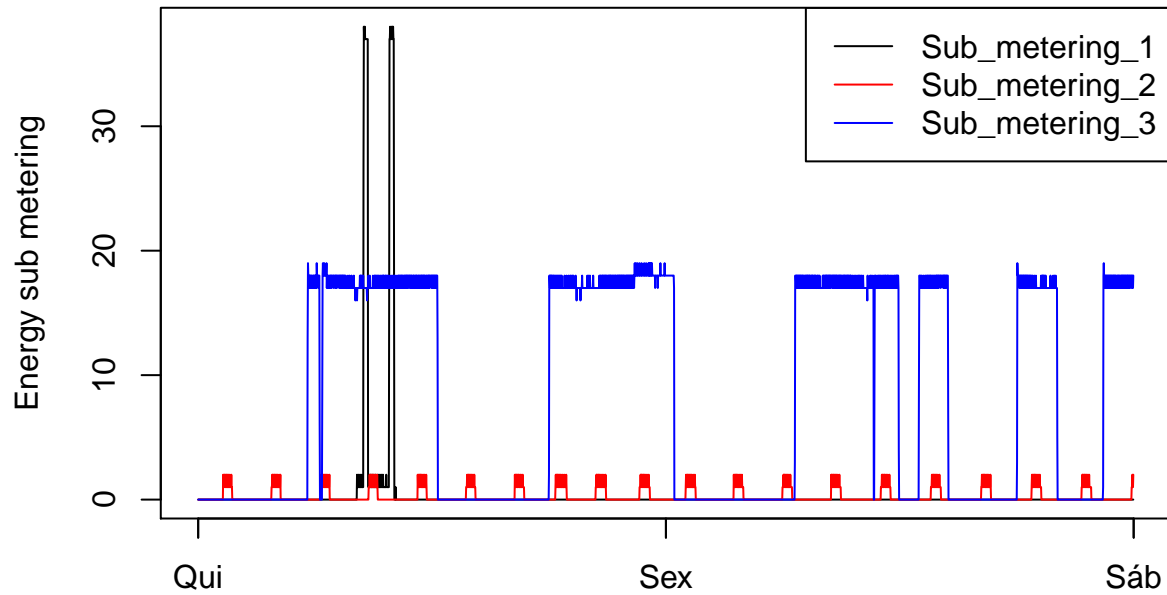


```
# Export plot2
png(filename = "plot2.png", width = 480, height = 480)
plot(finalData$SetTime, finalData$Global_active_power, type = "l", col = "black",
      xlab = "", ylab = "Global Active Power (kilowatts)")
dev.off()
```

```
## pdf
## 2
```

Generating Plot 3

```
columnlines = c("black", "red", "blue")
labels = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3")
plot(finalData$SetTime, finalData$Sub_metering_1, type = "l", col = columnlines[1],
      xlab = "", ylab = "Energy sub metering")
lines(finalData$SetTime, finalData$Sub_metering_2, col = columnlines[2])
lines(finalData$SetTime, finalData$Sub_metering_3, col = columnlines[3])
legend("topright", legend = labels, col = columnlines, lty = "solid")
```

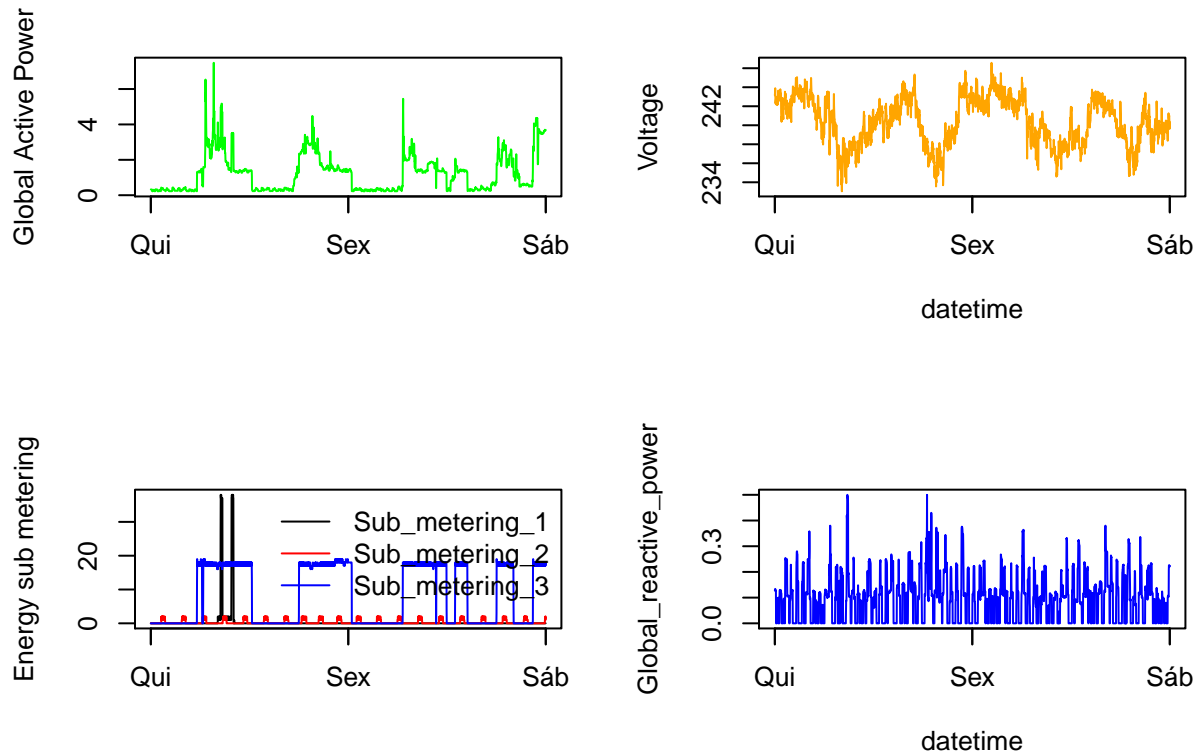


```
# Export plot3
png(filename = "plot3.png", width = 480, height = 480)
columnlines = c("black", "red", "blue")
labels = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3")
plot(finalData$SetTime, finalData$Sub_metering_1, type = "l", col = columnlines[1],
      xlab = "", ylab = "Energy sub metering")
lines(finalData$SetTime, finalData$Sub_metering_2, col = columnlines[2])
lines(finalData$SetTime, finalData$Sub_metering_3, col = columnlines[3])
legend("topright", legend = labels, col = columnlines, lty = "solid")
dev.off()
```

```
## pdf
## 2
```

Generating Plot 4

```
par(mfrow = c(2,2))
labels = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3")
columnlines = c("black", "red", "blue")
plot(finalData$SetTime, finalData$Global_active_power, type = "l", col = "green",
      xlab = "", ylab = "Global Active Power")
plot(finalData$SetTime, finalData$Voltage, type = "l", col = "orange",
      xlab = "datetime", ylab = "Voltage")
plot(finalData$SetTime, finalData$Sub_metering_1, type = "l", xlab = "",
      ylab = "Energy sub metering")
lines(finalData$SetTime, finalData$Sub_metering_2, type = "l", col = "red")
lines(finalData$SetTime, finalData$Sub_metering_3, type = "l", col = "blue")
legend("topright", bty = "n", legend = labels, lty = 1, col = columnlines)
plot(finalData$SetTime, finalData$Global_reactive_power, type = "l", col = "blue",
      xlab = "datetime", ylab = "Global reactive power")
```



```
par(mfrow = c(1,1))

# Export plot4
png(filename = "plot4.png", width = 480, height = 480)
par(mfrow = c(2,2))
labels = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3")
columnlines = c("black", "red", "blue")
plot(finalData$SetTime, finalData$Global_active_power, type = "l", col = "green",
      xlab = "", ylab = "Global Active Power")
plot(finalData$SetTime, finalData$Voltage, type = "l", col = "orange",
      xlab = "datetime", ylab = "Voltage")
plot(finalData$SetTime, finalData$Sub_metering_1, type = "l", xlab = "",
      ylab = "Energy sub metering")
lines(finalData$SetTime, finalData$Sub_metering_2, type = "l", col = "red")
lines(finalData$SetTime, finalData$Sub_metering_3, type = "l", col = "blue")
legend("topright", bty = "n", legend = labels, lty = 1, col = columnlines)
plot(finalData$SetTime, finalData$Global_reactive_power, type = "l", col = "blue",
      xlab = "datetime", ylab = "Global reactive power")
par(mfrow = c(1,1))
dev.off()
```

```
## pdf
## 2
```