

Peer Graded Assignment: Course Project 1

Eldirdiri Fadol

Tuesday, 7/12/2016

Introduction

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”.

Dataset: [Electric power consumption](#) [20MB]

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 following descriptions of the 9 variables in the dataset are taken from the [UCI web site](#):

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.

Loading the data

When loading the dataset into R, the following to be considered:

- The dataset has 2,075,259 rows and 9 columns. First a rough estimate of how much memory the dataset will require in memory before reading into R should be calculated. Computer should have 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.

- It may be useful to convert the Date and Time variables to Date/Time classes in R using the `strptime()` and `as.Date()` functions.
- In this dataset missing values are coded as `?`.

Making Plots

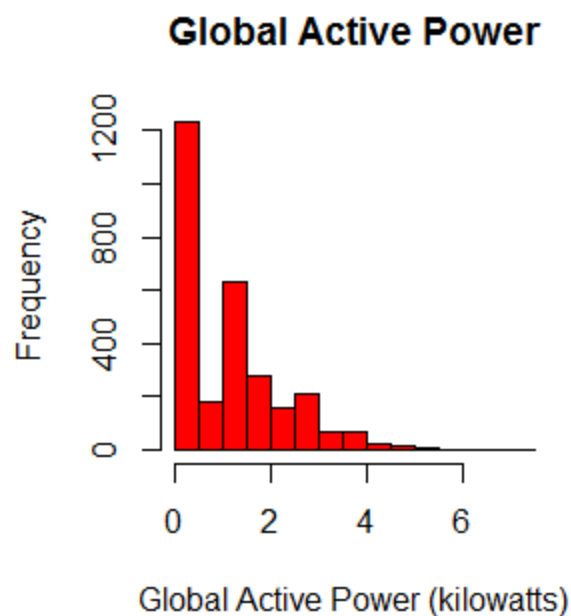
Our overall goal here is simply to examine how household energy usage varies over a 2-day period in February, 2007. Our task is to construct the plots using the base plotting system.

```
fh <- file("household_power_consumption.txt")

ba <- read.table(text = grep("^[1,2]/2/2007", readLines(fh), value = TRUE), col.names = c("Date", "Time", "Global_active_power", "Global_reactive_power", "Voltage", "Global_intensity", "Sub_metering_1", "Sub_metering_2", "Sub_metering_3"), sep = ";", header = TRUE)

# Generating Plot 1

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



```
## Getting full dataset

data_full <- read.csv("household_power_consumption.txt", header = T, sep = ';',
                      na.strings = "?", nrow = 2075259, check.names = F,
                      stringsAsFactors = F, comment.char = "", quote = '\\"')

data_full$Date <- as.Date(data_full$Date, format = "%d/%m/%Y")

## Subsetting the data

data <- subset(data_full, subset = (Date >= "2007-02-01" & Date <= "2007-02-02"))

rm(data_full)

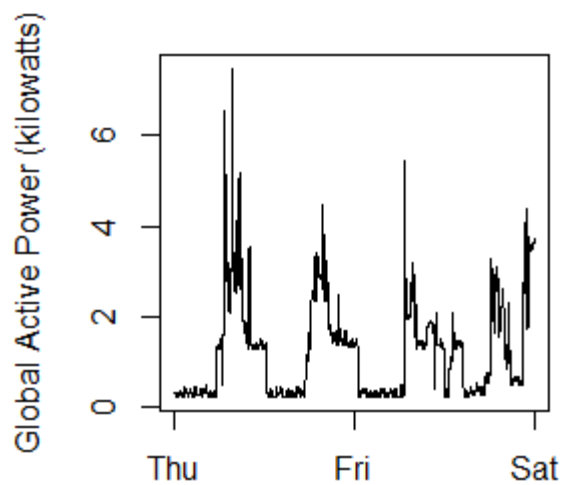
## Converting dates

datetime <- paste(as.Date(data$Date), data$Time)

data$Datetime <- as.POSIXct(datetime)

## Generating Plot 2

plot(data$Global_active_power ~ data$Datetime, type = "l", ylab = "Global Active Power
(kilowatts)", xlab = "")
```



```
## Generating Plot 3

with(data, {

  plot(Sub_metering_1 ~ Datetime, type = "l",

  ylab = "Global Active Power (kilowatts)", xlab = "")

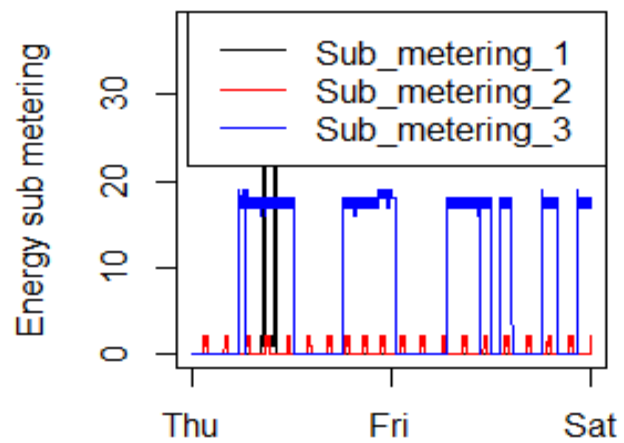
  lines(Sub_metering_2 ~ Datetime, col = 'Red')

  lines(Sub_metering_3 ~ Datetime, col = 'Blue')

})

legend("topright", col = c("black", "red", "blue"), lty = 1, lwd = 2,

  legend = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"))
```



```
## Generating Plot 4

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

with(data, {

  plot(Global_active_power ~ Datetime, type = "l",
        ylab = "Global Active Power", xlab = "")

  plot(Voltage ~ Datetime, type = "l", ylab = "Voltage", xlab = "datetime")

  plot(Sub_metering_1 ~ Datetime, type = "l", ylab = "Energy sub metering",
        xlab = "")

  lines(Sub_metering_2 ~ Datetime, col = 'Red')
  lines(Sub_metering_3 ~ Datetime, col = 'Blue')

  legend("topright", col = c("black", "red", "blue"), lty = 1, lwd = 2,
        bty = "n",
        legend = c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"))

  plot(Global_reactive_power ~ Datetime, type = "l",
        ylab = "Global_reactive_power", xlab = "datetime")

})
```

