##

## Author: ESTEVAN GOMEZ

##

## Course Project Assignment: Exploratory Data Analysis, Project 1

## 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:

## 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 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.

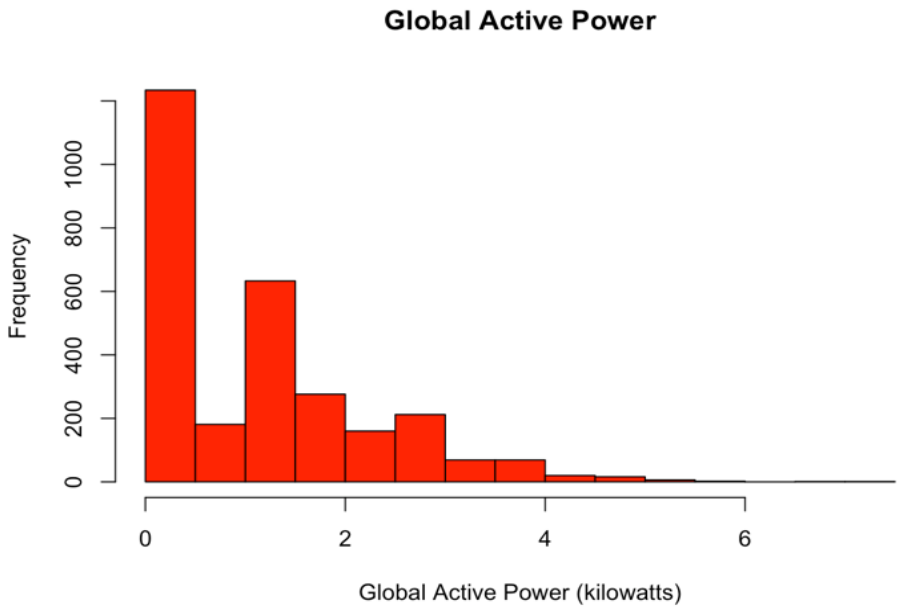
##

## 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.

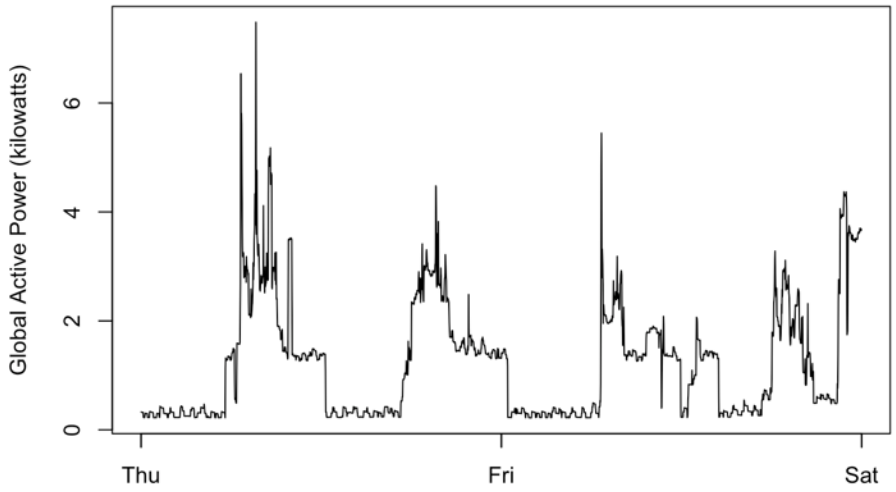
```
##
## Note that in this dataset missing values are coded as ?.
##
## Making Plots
##
## Our overall goal here is simply to examine how household energy usage varies
## over a 2-day period in February, 2007. Your task is to reconstruct the
## following plots below, all of which were constructed using 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
##
## Construct the plot and save it to a PNG file with a width of 480 pixels and a
## height of 480 pixels.
##
## Name each of the plot files as plot1.png, plot2.png, etc.
##
## 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.
## 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("~/Desktop/BigDataAnalytics/ExData_Plotting1")
```

```
downloadURL <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip"

downloadFile <- "./Data/household_power_consumption.zip"

householdFile <- "./Data/household_power_consumption.txt"

##

if (!file.exists(householdFile)) {

  download.file(downloadURL, downloadFile, method = "curl")

  unzip(downloadFile, overwrite = T, exdir = "./Data")

}

##

plotData <- read.table(householdFile, header=T, sep=";", na.strings="?")

## set 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)

##

## Generating Plot 1

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



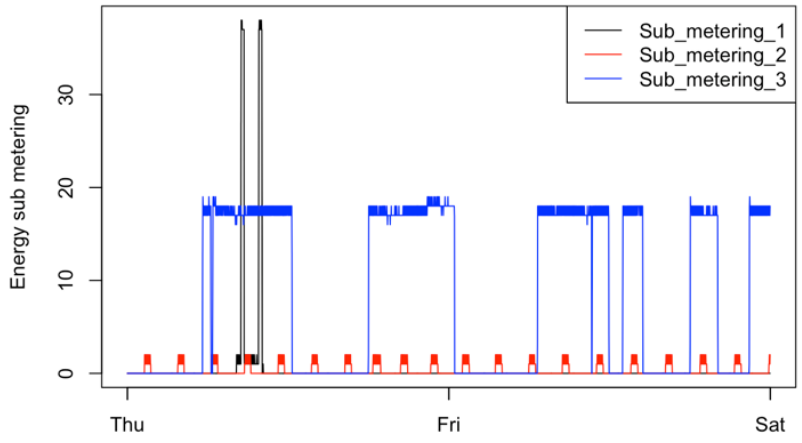
## Generating Plot 2

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




## 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")
```




## Generating Plot 4

```
labels <- c("Sub_metering_1","Sub_metering_2","Sub_metering_3")

columnlines <- c("black","red","blue")

par(mfrow=c(2,2))

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")
```