**CodeBook.md**

**Terry L. Jones**

**Monday, October 8, 2018**

**Exploratory Data Analysis – Week 1 Project**

This assignment uses data from the [UC Irvine Machine Learning Repository](http://archive.ics.uci.edu/ml/), 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](https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip) [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](https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption):

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.

**Project Goal:**

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 \color{red}{\verb|plot1.png|}plot1.png, \color{red}{\verb|plot2.png|}plot2.png, etc.
* Create a separate R code file (\color{red}{\verb|plot1.R|}plot1.R, \color{red}{\verb|plot2.R|}plot2.R, etc.) that constructs the corresponding plot, i.e. code in \color{red}{\verb|plot1.R|}plot1.R constructs the \color{red}{\verb|plot1.png|}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.

The code for plotting the coding and the related plot for each are on the following pages:

**Plot 1**

## Plot 1

## Read the consumption data

power <- read.table("household\_power\_consumption.txt", skip=1, sep=";")

names(power) <- c("Date", "Time", "Global\_active\_power", "Global\_reactive\_power",

"Voltage", "Global\_intensity", "Sub\_metering\_1", "Sub\_metering\_2",

"Sub\_metering\_3")

powerUse <- subset(power, power$Date=="1/2/2007" | power$Date == "2/2/2007")

## call in the basic plot funciton

hist(as.numeric(as.character(powerUse$Global\_active\_power)), col="red",

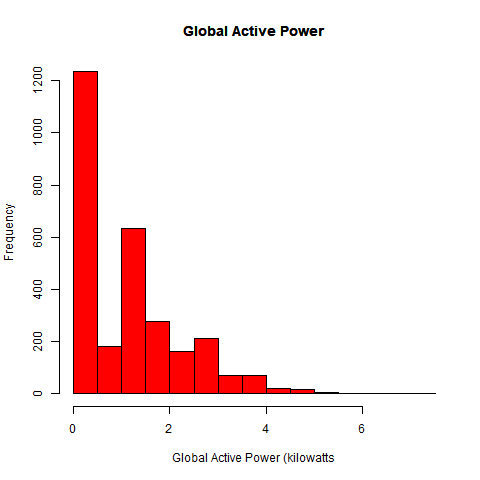
main="Global Active Power", xlab = "Global Active Power (kilowatts")

## add title to graph

title(main = "Global Active Power")

## Create a png file (480 \* 480)

dev.print(png, file="plot1.png", width = 480, height = 480)



**Plot 2**

## Plot 2

## Reading, naming, and subsetting the power consumption data

power <- read.table("household\_power\_consumption.txt",skip=1,sep=";")

names(power) <- c("Date","Time","Global\_active\_power","Global\_reactive\_power","Voltage",

"Global\_intensity","Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3")

powerUse <- subset(power,power$Date=="1/2/2007" | power$Date =="2/2/2007")

## Change the Date/TIme vars from characters into Date and POSIXLt

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

powerUse$Time <- strptime(powerUse$Time, format="%H:%M:%S")

powerUse[1:1440,"Time"] <- format(powerUse[1:1440,"Time"],"2007-02-01 %H:%M:%S")

powerUse[1441:2880,"Time"] <- format(powerUse[1441:2880,"Time"],"2007-02-02 %H:%M:%S")

## Call in the basic plot function

plot(powerUse$Time,as.numeric(as.character(powerUse$Global\_active\_power)),type="l",

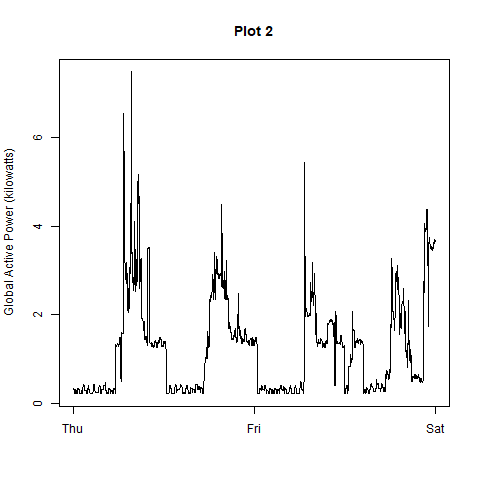
xlab="",ylab="Global Active Power (kilowatts)")

## Add title to graph

title(main = "Plot 2")

## Create a png file (480 \* 480)

dev.print(png, file="plot2.png", width = 480, height = 480)



**Plot 3**

#Plot 3

##Reading, naming, and subsetting the power consumption data

power <- read.table("household\_power\_consumption.txt",skip=1,sep=";")

names(power) <- c("Date","Time","Global\_active\_power","Global\_reactive\_power","Voltage",

"Global\_intensity","Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3")

powerUse <- subset(power,power$Date=="1/2/2007" | power$Date =="2/2/2007")

## Change the Date/TIme vars from characters into Date and POSIXLt

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

powerUse$Time <- strptime(powerUse$Time, format="%H:%M:%S")

powerUse[1:1440,"Time"] <- format(powerUse[1:1440,"Time"],"2007-02-01 %H:%M:%S")

powerUse[1441:2880,"Time"] <- format(powerUse[1441:2880,"Time"],"2007-02-02 %H:%M:%S")

## Call in the basic plot function

plot(powerUse$Time,powerUse$Sub\_metering\_1,type="n",xlab="",ylab="Energy sub metering")

with(powerUse,lines(Time,as.numeric(as.character(Sub\_metering\_1))))

with(powerUse,lines(Time,as.numeric(as.character(Sub\_metering\_2)),col="red"))

with(powerUse,lines(Time,as.numeric(as.character(Sub\_metering\_3)),col="blue"))

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

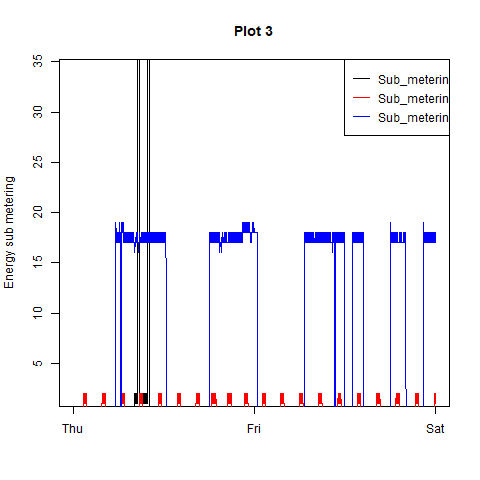
c("Sub\_metering\_1", "Sub\_metering\_2", "Sub\_metering\_3"))

title(main = "Plot 3")

## Create a png file (480 \* 480)

dev.print(png, file="plot3.png", width = 480, height = 480)

## Tried multiple approaches to display the Legend box. However, could not find a cex setting that would work



**Plot 4**

#Plot 4

# Reading, naming, and subsetting the power consumption data

power <- read.table("household\_power\_consumption.txt",skip=1,sep=";")

names(power) <- c("Date","Time","Global\_active\_power","Global\_reactive\_power",

"Voltage","Global\_intensity","Sub\_metering\_1","Sub\_metering\_2",

"Sub\_metering\_3")

powerUse <- subset(power,power$Date=="1/2/2007" | power$Date =="2/2/2007")

# Converting the Date and Time vars from characters into Date and POSIXlt

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

powerUse$Time <- strptime(powerUse$Time, format="%H:%M:%S")

powerUse[1:1440,"Time"] <- format(powerUse[1:1440,"Time"],"2007-02-01 %H:%M:%S")

powerUse[1441:2880,"Time"] <- format(powerUse[1441:2880,"Time"],"2007-02-02 %H:%M:%S")

# Initiating a composite plot for a quad set of graphs

par(mfrow=c(2,2))

#Calling the basic plot function for the different plot functions that build the 4 plots

with(powerUse,{

plot(powerUse$Time, as.numeric(as.character(powerUse$Global\_active\_power)), type="l",

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

plot(powerUse$Time, as.numeric(as.character(powerUse$Voltage)), type="l",

xlab="datetime", ylab="Voltage")

plot(powerUse$Time, powerUse$Sub\_metering\_1, type="n",

xlab="", ylab="Energy sub metering")

with(powerUse, lines(Time, as.numeric(as.character(Sub\_metering\_1))))

with(powerUse, lines(Time, as.numeric(as.character(Sub\_metering\_2)), col="red"))

with(powerUse, lines(Time, as.numeric(as.character(Sub\_metering\_3)), col="blue"))

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

legend=c("Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3"), cex = 0.6)

plot(powerUse$Time, as.numeric(as.character(powerUse$Global\_reactive\_power)), type="l",

xlab="datetime", ylab="Global\_reactive\_power")

})

## Create a png file (480 \* 480)

dev.print(png, file="plot4.png", width = 480, height = 480)

