Plot4

Harold Pollack

9/15/2018

This is the markdown file that creates plot4. I didn’t print out all the intermediate products that are the same as those for plot one.

I first read the UCI data. Then save the dates 2/1/2007 and 2/2/2007. Note that the date variable is dd/mm/y. So some care must be taken. ALso cache=True given size of the dataset.

mydata <- read.csv("/Users/haroldpollack/Documents/coursera\_EDA/pS1/household\_power\_consumption.txt",header=TRUE,sep=";")  
mydata$date\_formatted <-as.Date(mydata$Date,format="%d/%m/%Y")  
str(mydata)

## 'data.frame': 2075259 obs. of 10 variables:  
## $ Date : Factor w/ 1442 levels "1/1/2007","1/1/2008",..: 342 342 342 342 342 342 342 342 342 342 ...  
## $ Time : Factor w/ 1440 levels "00:00:00","00:01:00",..: 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 ...  
## $ Global\_active\_power : Factor w/ 4187 levels "?","0.076","0.078",..: 2082 2654 2661 2668 1807 1734 1825 1824 1808 1805 ...  
## $ Global\_reactive\_power: Factor w/ 533 levels "?","0.000","0.046",..: 189 198 229 231 244 241 240 240 235 235 ...  
## $ Voltage : Factor w/ 2838 levels "?","223.200",..: 992 871 837 882 1076 1010 1017 1030 907 894 ...  
## $ Global\_intensity : Factor w/ 222 levels "?","0.200","0.400",..: 53 81 81 81 40 36 40 40 40 40 ...  
## $ Sub\_metering\_1 : Factor w/ 89 levels "?","0.000","1.000",..: 2 2 2 2 2 2 2 2 2 2 ...  
## $ Sub\_metering\_2 : Factor w/ 82 levels "?","0.000","1.000",..: 3 3 14 3 3 14 3 3 3 14 ...  
## $ Sub\_metering\_3 : num 17 16 17 17 17 17 17 17 17 16 ...  
## $ date\_formatted : Date, format: "2006-12-16" "2006-12-16" ...

# summary(mydata)  
# table(mydata$Date)  
#  
# now extract the two dates called for in the assignment  
#  
data\_for\_graphing <-subset(mydata,(Date== "1/2/2007" | Date == "2/2/2007"))  
# str(data\_for\_graphing)  
# summary(data\_for\_graphing)  
# table(data\_for\_graphing$Date)  
table(data\_for\_graphing$date\_formatted)

##   
## 2007-02-01 2007-02-02   
## 1440 1440

#  
# now figure out weekday  
#  
data\_for\_graphing$weekday <- weekdays(data\_for\_graphing$date\_formatted)  
table(data\_for\_graphing$weekday)

##   
## Friday Thursday   
## 1440 1440

Now manipulate date/time. I used code found here <http://biostat.mc.vanderbilt.edu/wiki/pub/Main/ColeBeck/datestimes.pdf>

data\_for\_graphing$dts1 <- paste(data\_for\_graphing$date\_formatted, data\_for\_graphing$Time)  
time\_scr <-strptime(data\_for\_graphing$dts1,format = "%Y-%m-%d %H:%M:%S")  
#  
# time\_scr uses strptime. I just included this to learn.  
#  
str(time\_scr)

## POSIXlt[1:2880], format: "2007-02-01 00:00:00" "2007-02-01 00:01:00" ...

data\_for\_graphing$dts2 <- as.POSIXct(data\_for\_graphing$dts1, format = "%Y-%m-%d %H:%M:%S")  
str(data\_for\_graphing$dts1)

## chr [1:2880] "2007-02-01 00:00:00" "2007-02-01 00:01:00" ...

str(data\_for\_graphing$dts2)

## POSIXct[1:2880], format: "2007-02-01 00:00:00" "2007-02-01 00:01:00" ...

make the multi-facet graph and save as plot4.png Note that I shrunk the legend using CEX

#  
# note that data\_for\_graphing$Global\_active\_power is a factor. It has to be changed to numeric for graphing  
#  
# Global\_active\_power is a factor. So one must be careful converting it to numeric  
#  
data\_for\_graphing$Global\_active\_power\_numeric <- as.numeric(as.character(data\_for\_graphing$Global\_active\_power))   
# hist(Global\_active\_power\_numeric,col="red",xlab="Global Active Power (kilowatts)",main="Global Active Power",ylab="Frequency")  
data\_for\_graphing$Sub\_metering\_1\_numeric <- as.numeric(as.character(data\_for\_graphing$Sub\_metering\_1))   
summary(data\_for\_graphing$Sub\_metering\_1\_numeric)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.4062 0.0000 38.0000

data\_for\_graphing$Sub\_metering\_2\_numeric <- as.numeric(as.character(data\_for\_graphing$Sub\_metering\_2))   
summary(data\_for\_graphing$Sub\_metering\_2\_numeric)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.0000 0.0000 0.0000 0.2576 0.0000 2.0000

summary(data\_for\_graphing$Sub\_metering\_3)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.000 0.000 8.501 17.000 19.000

str(data\_for\_graphing$Sub\_metering\_1\_numeric)

## num [1:2880] 0 0 0 0 0 0 0 0 0 0 ...

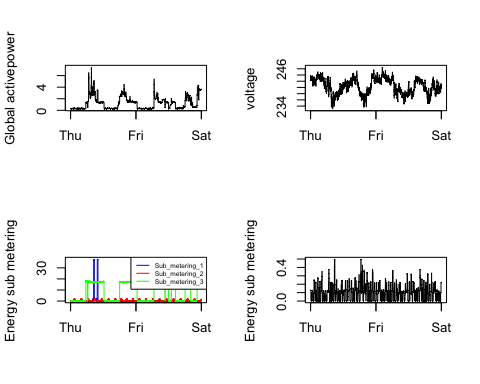
str(data\_for\_graphing$Sub\_metering\_2\_numeric)

## num [1:2880] 0 0 0 0 0 0 0 0 0 0 ...

str(data\_for\_graphing$Sub\_metering\_3)

## num [1:2880] 0 0 0 0 0 0 0 0 0 0 ...

#  
# time series by day  
#  
#  
# template here https://www.sixhat.net/plotting-multiple-data-series-in-r.html  
#  
par(mfrow=c(2,2))  
with (data\_for\_graphing, {  
 plot(dts2,as.numeric(as.character(Global\_active\_power)),type="l", xlab=" ", ylab = "Global activepower")   
 plot(dts2,as.numeric(as.character(Voltage)),type="l", xlab=" ", ylab = "voltage")  
 plot(dts2,Sub\_metering\_1\_numeric,type="l", xlab=" ", ylab = "Energy sub metering", col="blue")  
points(dts2,Sub\_metering\_2\_numeric, type='l',col="red")  
points(dts2,Sub\_metering\_3, type='l',col="green")  
legend ("topright",lty=1, col=c("blue","red","green"),legend = c("Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3"),cex=0.5)  
   
 plot(dts2,as.numeric(as.character(Global\_reactive\_power)),type="l", xlab=" ", ylab = "Energy sub metering")  
 })



#  
# now save the png file  
#  
dev.copy(png,'/Users/haroldpollack/Documents/coursera\_datascience3/Plot4/plot4.png')

## quartz\_off\_screen   
## 3

dev.off()

## quartz\_off\_screen   
## 2