power\_analysis.R

jennywang

Tue Oct 11 20:20:15 2016

##This assignment uses data from the UC Irvine Machine Learning Repository,   
## a popular repository for machine learning datasets.  
  
## load packages  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(data.table)

## -------------------------------------------------------------------------

## data.table + dplyr code now lives in dtplyr.  
## Please library(dtplyr)!

## -------------------------------------------------------------------------

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, last

library(datasets)  
  
##1. mannually download the zip file and unzip  
setwd("~/Desktop/Coursera/Course 4--Assignment 1") ##set a folder for this assignment  
list.files() ##to obtain the zipfile's name

## [1] "household\_power\_consumption.txt" "plot1.png"   
## [3] "plot2.png" "plot3.png"   
## [5] "plot4.png" "power\_analysis.html"   
## [7] "power\_analysis.R" "power\_analysis.spin.R"   
## [9] "power\_analysis.spin.Rmd"

unzip(zipfile="exdata-data-household\_power\_consumption 11.11.53 AM.zip")

## Warning in unzip(zipfile = "exdata-data-household\_power\_consumption  
## 11.11.53 AM.zip"): error 1 in extracting from zip file

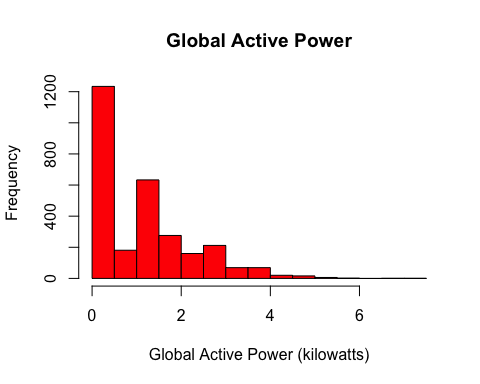
list.files() ## to obtain the unzip file's name

## [1] "household\_power\_consumption.txt" "plot1.png"   
## [3] "plot2.png" "plot3.png"   
## [5] "plot4.png" "power\_analysis.html"   
## [7] "power\_analysis.R" "power\_analysis.spin.R"   
## [9] "power\_analysis.spin.Rmd"

##2. read the file and change the date format  
HPC<- read.table("household\_power\_consumption.txt", header=T, sep=";")   
## use all the argument otherwise u will only see in varialbe "V1"  
dim(HPC) ## returns [1] 2075259 9

## [1] 2075259 9

HPC$Date <- as.Date(HPC$Date, format="%d/%m/%Y") ## change the date format  
  
##3.set working data from the dates 2007-02-01 and 2007-02-02  
power<-subset(HPC, Date=="2007-02-01" | Date=="2007-02-02") ##2880 obs of 9 variables  
  
##4 to make plot1, data has to be numeric  
power$Global\_active\_power<-as.numeric(as.character(power$Global\_active\_power))  
hist(power$Global\_active\_power, main="Global Active Power",   
 xlab="Global Active Power (kilowatts)", col="red")



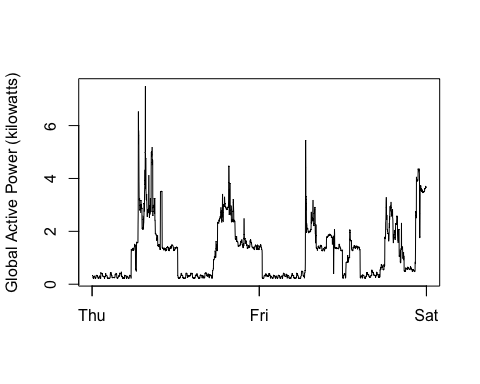
dev.copy(png, file="plot1.png", width=480, height=480)## Plot1.png has been saved in folder

## quartz\_off\_screen   
## 3

dev.off() ## returns to screen

## quartz\_off\_screen   
## 2

## to make plot2  
power<-transform(power, newtime=as.POSIXct(paste(Date, Time))) ## set a new varialbe  
plot(power$newtime, power$Global\_active\_power, type="l",   
 xlab="", ylab="Global Active Power (kilowatts)")



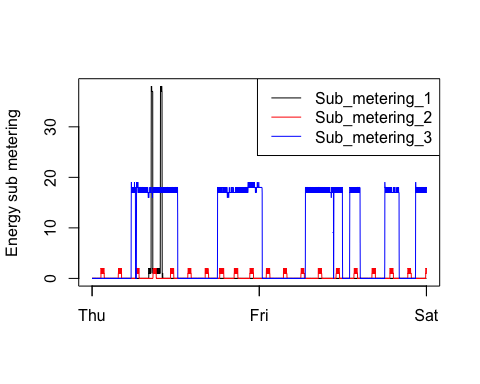
dev.copy(png, file="plot2.png", width=480, height=480)## Plot2.png has been saved in folder

## quartz\_off\_screen   
## 3

dev.off() ## returns to screen

## quartz\_off\_screen   
## 2

## to make plot3  
power$Sub\_metering\_1<-as.numeric(as.character(power$Sub\_metering\_1))  
power$Sub\_metering\_2<-as.numeric(as.character(power$Sub\_metering\_2))  
power$Sub\_metering\_3<-as.numeric(as.character(power$Sub\_metering\_3))  
## ready to plot:  
plot(power$newtime, power$Sub\_metering\_1, type="n", xlab="", ylab="Energy sub metering")  
lines(power$newtime, power$Sub\_metering\_1, type="l",col="black")  
lines(power$newtime, power$Sub\_metering\_2, type="l",col="red")  
lines(power$newtime, power$Sub\_metering\_3, type="l",col="blue")  
legend("topright", col=c("black","red","blue"), legend=c("Sub\_metering\_1","Sub\_metering\_2", "Sub\_metering\_3"),  
 lty=c(1,1), lwd=c(1,1))



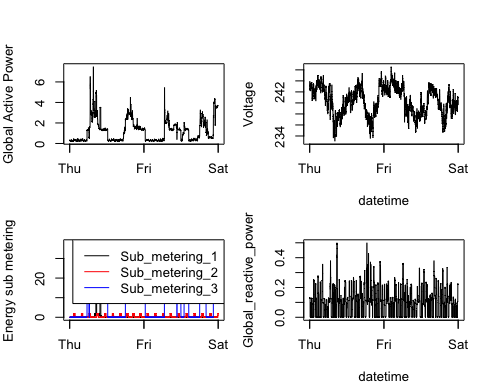
dev.copy(png, file="plot3.png", width=480, height=480)## Plot2.png has been saved in folder

## quartz\_off\_screen   
## 3

dev.off() ## returns to screen

## quartz\_off\_screen   
## 2

##to make plot4  
par(mfrow = c(2,2), mar = c(4,4,2,1), oma = c(0,0,2,0)) ##set up canvas  
plot(power$newtime, power$Global\_active\_power, type="l",   
 xlab="", ylab="Global Active Power") ## 1st plot  
## 2nd plot  
power$Voltage<-as.numeric(as.character(power$Voltage))  
plot(power$newtime, power$Voltage, type="l", xlab="datetime", ylab="Voltage")  
## 3rd plot  
plot(power$newtime, power$Sub\_metering\_1, type="n", xlab="", ylab="Energy sub metering")  
lines(power$newtime, power$Sub\_metering\_1, type="l",col="black")  
lines(power$newtime, power$Sub\_metering\_2, type="l",col="red")  
lines(power$newtime, power$Sub\_metering\_3, type="l",col="blue")  
legend("topright", col=c("black","red","blue"), legend=c("Sub\_metering\_1","Sub\_metering\_2", "Sub\_metering\_3"),  
 lty=c(1,1), lwd=c(1,1))  
## 4th plot  
power$Global\_reactive\_power<-as.numeric(as.character(power$Global\_reactive\_power))  
plot(power$newtime, power$Global\_reactive\_power, type="l",   
 xlab="datetime", ylab="Global\_reactive\_power")



## copy to png file  
dev.copy(png, file="plot4.png", width=480, height=480)

## quartz\_off\_screen   
## 3

dev.off() ## returns to screen

## quartz\_off\_screen   
## 2

## Knit the R, and the end