## Introduction

This assignment uses data from the UC Irvine Machine Learning Repository, a popular repository for machine learning datasets. The overall goal is to examine how household energy usage varies over a 2-day period in February, 2007. The "Individual household electric power consumption Data Set" is used for this report. The url is given below.

This report is about exploratory graphs, understanding the data, and developing strategies.The aim is to help establish patterns in data and understand its properties. They suggest modeling strategies and help to debug analyses.

if(!file.exists('data.zip'))  
{  
 url <- "https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold\_power\_consumption.zip"  
   
 download.file(url, destfile = "data.zip")  
}  
unzip("data.zip")  
housePowerConsumption <-read.table("household\_power\_consumption.txt",header = TRUE, sep = ";", na.strings = "?")  
str(housePowerConsumption)

## 'data.frame': 2075259 obs. of 9 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 : num 4.22 5.36 5.37 5.39 3.67 ...  
## $ Global\_reactive\_power: num 0.418 0.436 0.498 0.502 0.528 0.522 0.52 0.52 0.51 0.51 ...  
## $ Voltage : num 235 234 233 234 236 ...  
## $ Global\_intensity : num 18.4 23 23 23 15.8 15 15.8 15.8 15.8 15.8 ...  
## $ Sub\_metering\_1 : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ Sub\_metering\_2 : num 1 1 2 1 1 2 1 1 1 2 ...  
## $ Sub\_metering\_3 : num 17 16 17 17 17 17 17 17 17 16 ...

Removing the NA values and selecting the required data from the dates 2007-02-01 and 2007-02-02

Data1 <- na.omit(housePowerConsumption)  
Data2 <- subset(Data1, Data1$Date == "1/2/2007" | Data1$Date == "2/2/2007" )  
str(Data2)

## 'data.frame': 2880 obs. of 9 variables:  
## $ Date : Factor w/ 1442 levels "1/1/2007","1/1/2008",..: 16 16 16 16 16 16 16 16 16 16 ...  
## $ Time : Factor w/ 1440 levels "00:00:00","00:01:00",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ Global\_active\_power : num 0.326 0.326 0.324 0.324 0.322 0.32 0.32 0.32 0.32 0.236 ...  
## $ Global\_reactive\_power: num 0.128 0.13 0.132 0.134 0.13 0.126 0.126 0.126 0.128 0 ...  
## $ Voltage : num 243 243 244 244 243 ...  
## $ Global\_intensity : num 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1 ...  
## $ Sub\_metering\_1 : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ Sub\_metering\_2 : num 0 0 0 0 0 0 0 0 0 0 ...  
## $ Sub\_metering\_3 : num 0 0 0 0 0 0 0 0 0 0 ...

## Attach time to the date

Data2$Date <- paste(Data2$Date, Data2$Time)  
class(Data2$Date)

## [1] "character"

str(Data2$Date)

## chr [1:2880] "1/2/2007 00:00:00" "1/2/2007 00:01:00" ...

Convert Date to yyyy-mm-dd hh:mm:ss

Data2$Date <- strptime(Data2$Date, "%d/%m/%Y %H:%M:%S")  
class(Data2$Date)

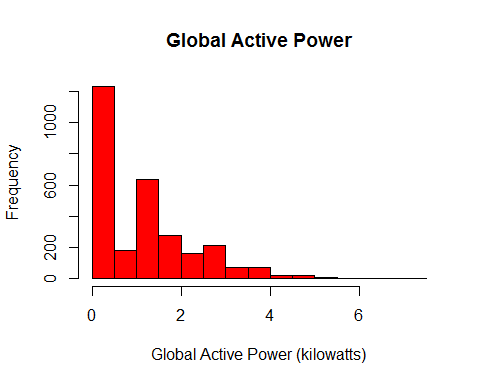
## [1] "POSIXlt" "POSIXt"

str(Data2$Date)

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

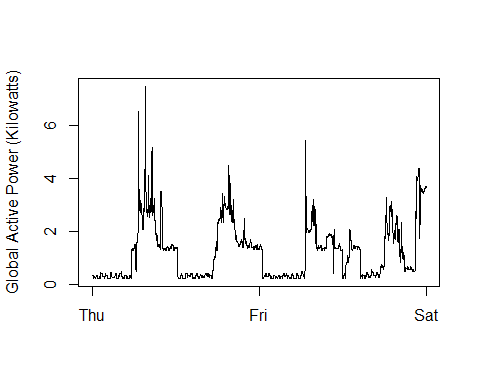
## Plot 1

hist(Data2$Global\_active\_power, main="Global Active Power", xlab="Global Active Power (kilowatts)", col="red")



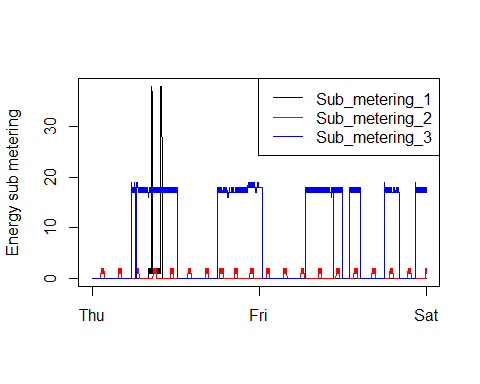
## Plot 2

plot(Data2$Date, Data2$Global\_active\_power,  
 type='l',ylab="Global Active Power (Kilowatts)", xlab="")



## Plot 3

plot(Data2$Date, Data2$Sub\_metering\_1,type='l', ylab ="Energy sub metering", xlab="")  
lines(Data2$Date, Data2$Sub\_metering\_2, type='l', col='red')  
lines(Data2$Date, Data2$Sub\_metering\_3, type='l', col="blue")  
legend('topright', c("Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3"),  
lty = c(1,1,1), col = c("black","red","blue"))



## Plot 4

par(mfcol=c(2,2))  
  
 plot(Data2$Date, as.numeric(as.character(Data2$Global\_active\_power)),type='l',ylab="Global Active Power", xlab="")  
  
plot(Data2$Date, as.numeric(as.character(Data2$Sub\_metering\_1)),type='l', xlab="",ylab ="Energy sub metering")  
 lines(Data2$Date, as.numeric(as.character(Data2$Sub\_metering\_2)),type='l', col='red')  
 lines(Data2$Date, Data2$Sub\_metering\_3,type='l', col="blue")  
 legend('topright', c("Sub\_metering\_1","Sub\_metering\_2","Sub\_metering\_3"),  
 lty=c(1,1,1),col=c("black","red","blue"))  
  
plot(Data2$Date, as.numeric(as.character(Data2$Voltage)),type='l',   
 ylab="Voltage",xlab="datetime" )  
  
plot(Data2$Date, as.numeric(as.character(Data2$Global\_reactive\_power)),type='l',   
 ylab="Global\_reactive\_power",xlab="datetime" )

