# Exploratory Analysis: Project 1

Loading libraries and datasets

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.5.1
```

```
##
## 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)
```

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

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

```
hp <- fread("/users/andrewhu/desktop/Coursera/Data/hp.txt")
```

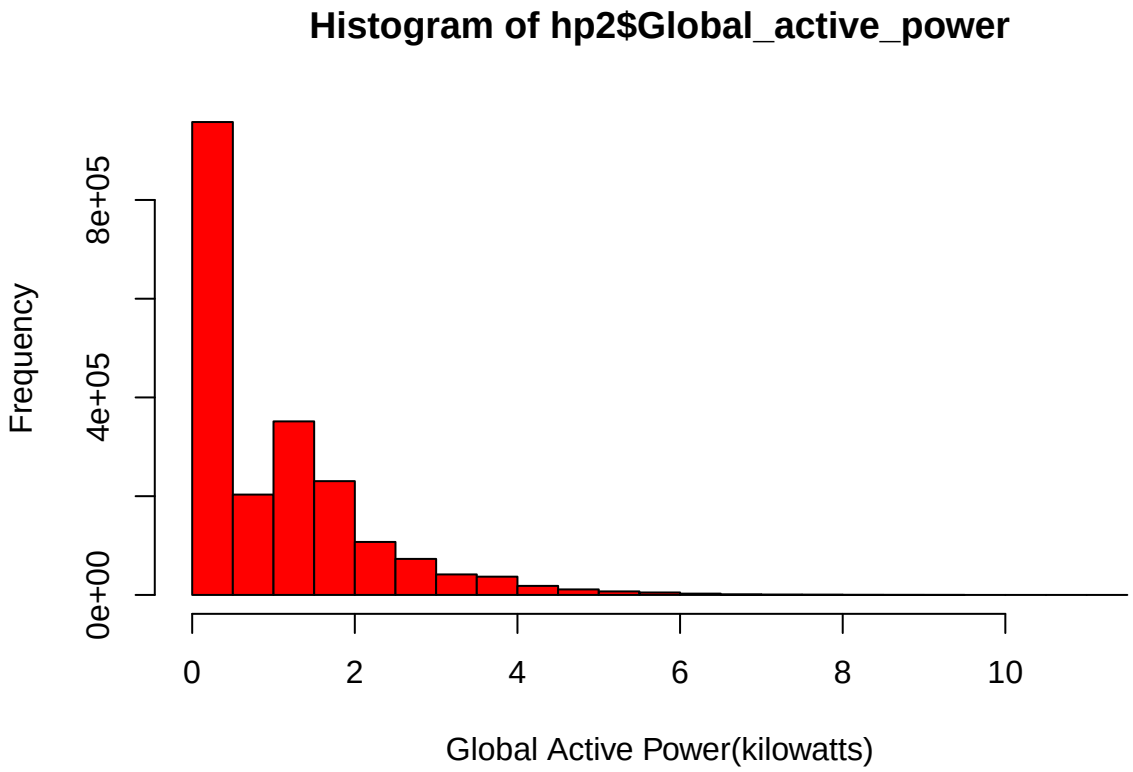
Transforming and subsetting datasets

```
#Trasforming dates and time
hp$dateTime =  as.POSIXct(paste(hp$Date, hp$Time), format = "%d/%m/%Y %H:%M:%S")

hp$Date = as.Date(hp$Date, format= "%d/%m/%Y")


hp2<- subset(hp, dateTime >= "2007-02-01" | dateTime <= "2007-02-02")

#transform to numeric
hp2$Global_active_power<-as.numeric(hp2$Global_active_power)
```

```
## Warning: NAs introduced by coercion
```

Plot 1:

```
#plot1
hist(hp2$Global_active_power, col="red", xlab = "Global Active Power(kilowatts)")
```

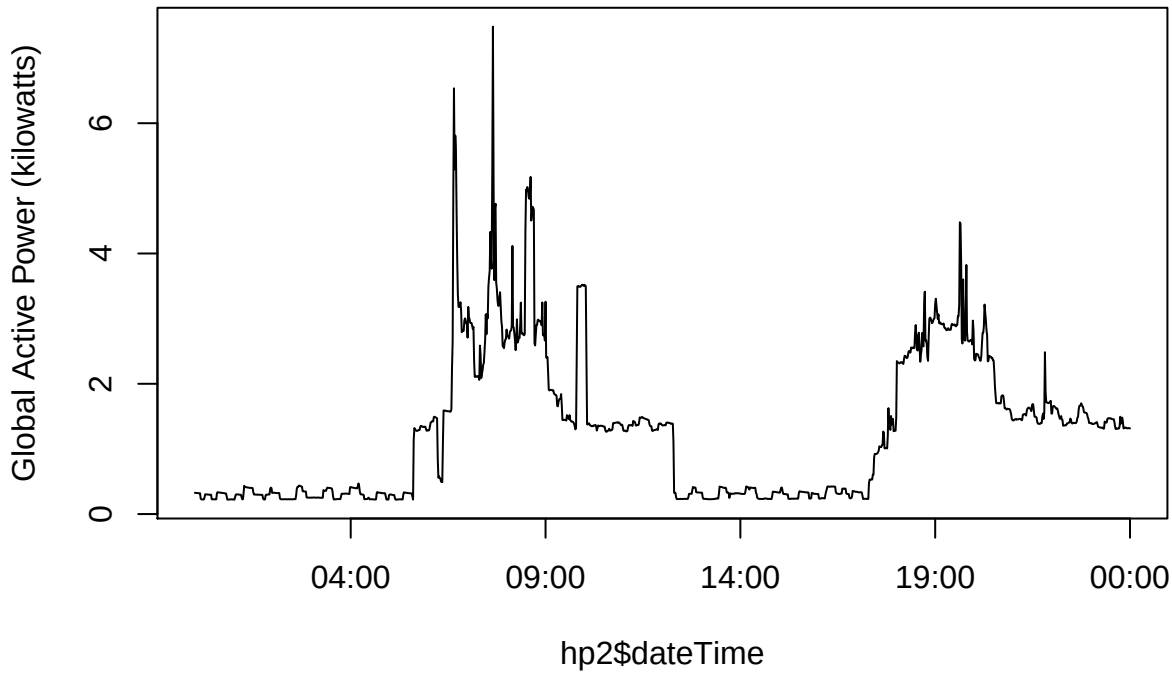



Plot 2:

```
#plot2
hp2$dateTime =  as.POSIXct(paste(hp$Date, hp$Time), format = "%d/%m/%Y %H:%M:%S")
```

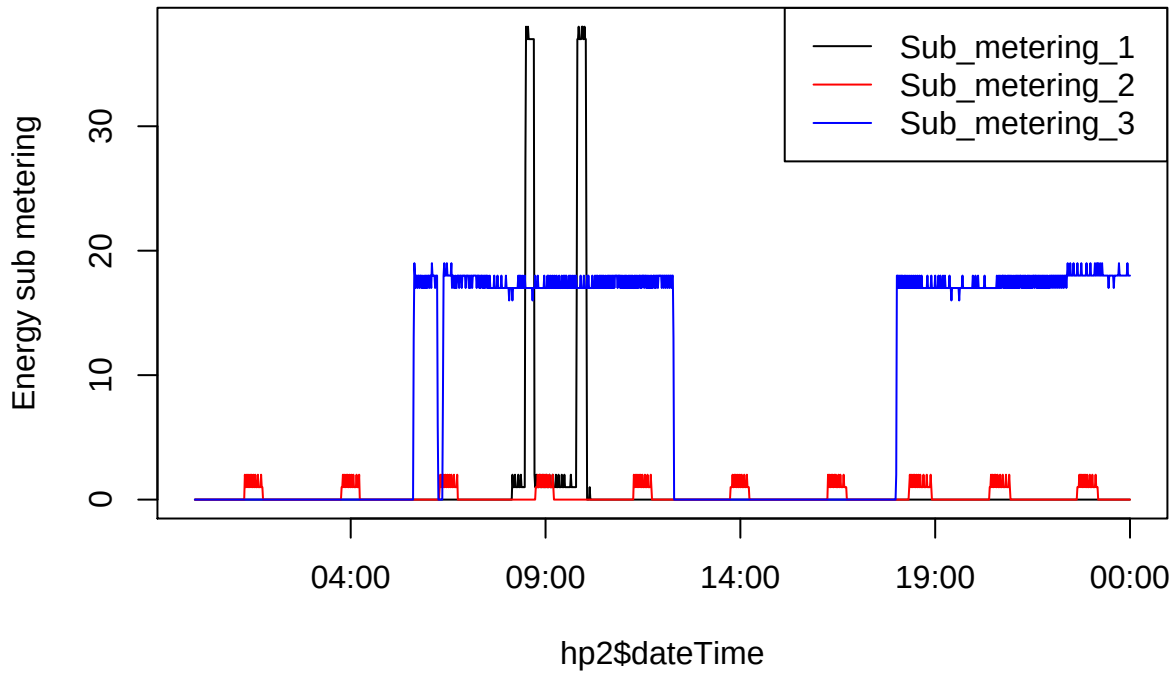
```
## Warning in `[<-.data.table`(x, j = name, value = value): Supplied 2075259
## items to be assigned to 2075019 items of column 'dateTime' (240 unused)
```

```
hp2 <- filter(hp, dateTime >= "2007-02-01" &  dateTime <= "2007-02-02")

plot(hp2$dateTime, hp2$Global_active_power, type="l", ylab = "Global Active Power (kilowatts)")
```



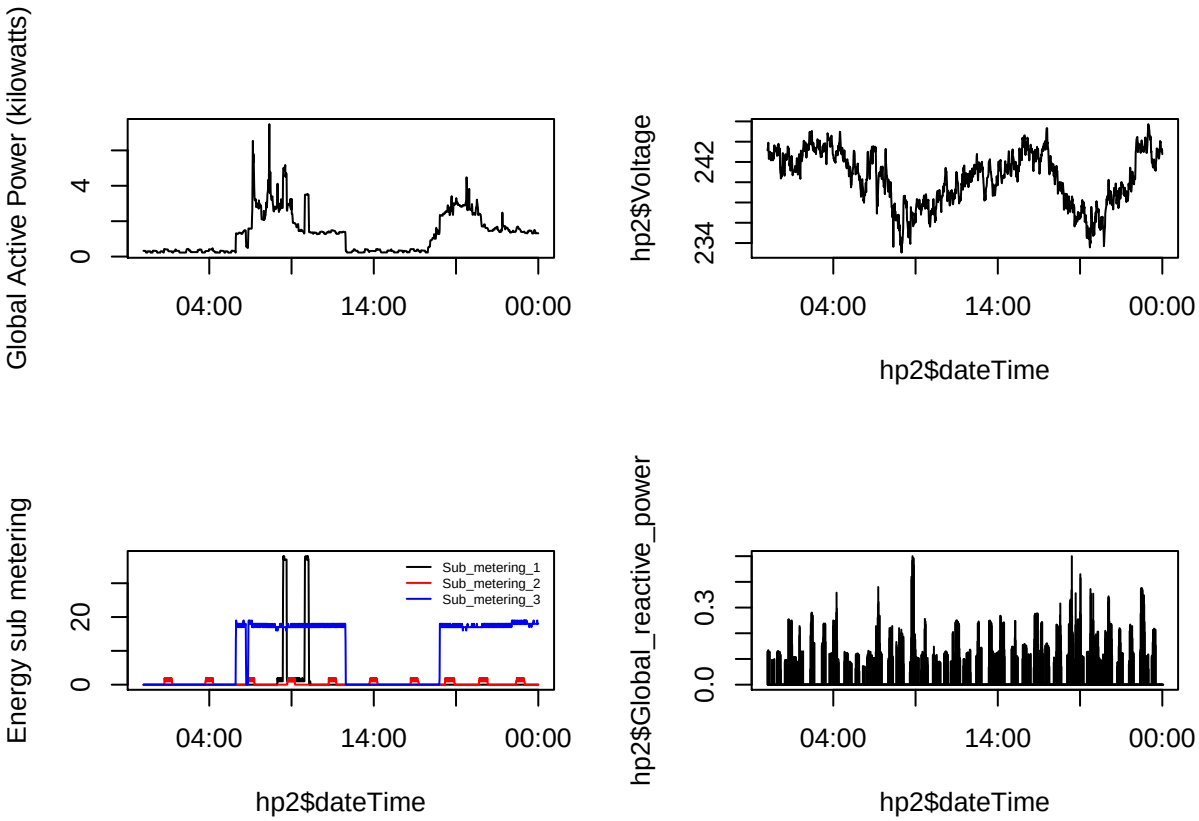

Plot 3:

```
#plot3

plot(hp2$dateTime, hp2$Sub_metering_1, type= "l",ylab = "Energy sub metering")#meaning: don't put the d
lines(hp2$dateTime, hp2$Sub_metering_2, type= "l",col="red")
lines(hp2$dateTime, hp2$Sub_metering_3, type= "l", col="blue")

legend("topright"
       , col=c("black","red","blue")
       , c("Sub_metering_1  ","Sub_metering_2  ", "Sub_metering_3  ")
       ,lty=c(1,1), lwd=c(1,1))
```




Plot 4:

```
#plot4
par(mfrow=c(2,2))
plot(hp2$dateTime, hp2$Global_active_power, type="l",xlab="", ylab = "Global Active Power (kilowatts)")
plot(hp2$dateTime, hp2$Voltage,type="l")


plot(hp2$dateTime, hp2$Sub_metering_1, type= "l",ylab = "Energy sub metering")#meaning: don't put the d
lines(hp2$dateTime, hp2$Sub_metering_2, type= "l",col="red")
lines(hp2$dateTime, hp2$Sub_metering_3, type= "l", col="blue")

legend("topright"
       , col=c("black","red","blue")
       , c("Sub_metering_1  ","Sub_metering_2  ", "Sub_metering_3 "), lty=c(1,1)
       , bty="n"
       , cex=.5 )

plot(hp2$dateTime,hp2$Global_reactive_power,type="h")
```


Global Active Power (kilowatts)
0
4
04:00
14:00
00:00
hp2$Voltage
234
242
04:00
14:00
00:00
hp2$dateTime
Energy sub metering
0
20
Sub_metering_1
Sub_metering_2
Sub_metering_3
04:00
14:00
00:00
hp2$dateTime
hp2$Global_reactive_power
0.0
0.3
04:00
14:00
00:00
hp2$dateTime