# Exploratory Analysis Week 1 project

### download and unzip data

```
url="https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip"
download.file(url,destfile = "data.zip")
data<-unzip("data.zip")
```

### check the dimention and characteristic of each variables

```
table<-read.table(data,header = TRUE,stringsAsFactors = T,sep=";")
sapply(table,class)
```

```
##                  Date                  Time   Global_active_power
##              "factor"              "factor"              "factor"
## Global_reactive_power               Voltage      Global_intensity
##              "factor"              "factor"              "factor"
##        Sub_metering_1        Sub_metering_2        Sub_metering_3
##              "factor"              "factor"             "numeric"
```

```
head(table,5)
```

```
##         Date     Time Global_active_power Global_reactive_power Voltage
## 1 16/12/2006 17:24:00               4.216                 0.418 234.840
## 2 16/12/2006 17:25:00               5.360                 0.436 233.630
## 3 16/12/2006 17:26:00               5.374                 0.498 233.290
## 4 16/12/2006 17:27:00               5.388                 0.502 233.740
## 5 16/12/2006 17:28:00               3.666                 0.528 235.680
##   Global_intensity Sub_metering_1 Sub_metering_2 Sub_metering_3
## 1           18.400          0.000          1.000             17
## 2           23.000          0.000          1.000             16
## 3           23.000          0.000          2.000             17
## 4           23.000          0.000          1.000             17
## 5           15.800          0.000          1.000             17
```

### strip the Date column to Date format

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

```
## [1] "Date"
```

### select the data from 2007-02-01 and 2007-02-02

```
new_table<-table[table$Date=="2007-02-01"|table$Date=="2007-02-02",]
dim(new_table)
```

```
## [1] 2880    9
```

```
head(new_table)
```

```
##             Date     Time Global_active_power Global_reactive_power
## 66637 2007-02-01 00:00:00               0.326                 0.128
## 66638 2007-02-01 00:01:00               0.326                 0.130
## 66639 2007-02-01 00:02:00               0.324                 0.132
## 66640 2007-02-01 00:03:00               0.324                 0.134
## 66641 2007-02-01 00:04:00               0.322                 0.130
## 66642 2007-02-01 00:05:00               0.320                 0.126
##       Voltage Global_intensity Sub_metering_1 Sub_metering_2
## 66637 243.150            1.400          0.000          0.000
## 66638 243.320            1.400          0.000          0.000
## 66639 243.510            1.400          0.000          0.000
## 66640 243.900            1.400          0.000          0.000
## 66641 243.160            1.400          0.000          0.000
## 66642 242.290            1.400          0.000          0.000
##       Sub_metering_3
## 66637              0
## 66638              0
## 66639              0
## 66640              0
## 66641              0
## 66642              0
```

### change time into time variable and other columns into numeric variables

```
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

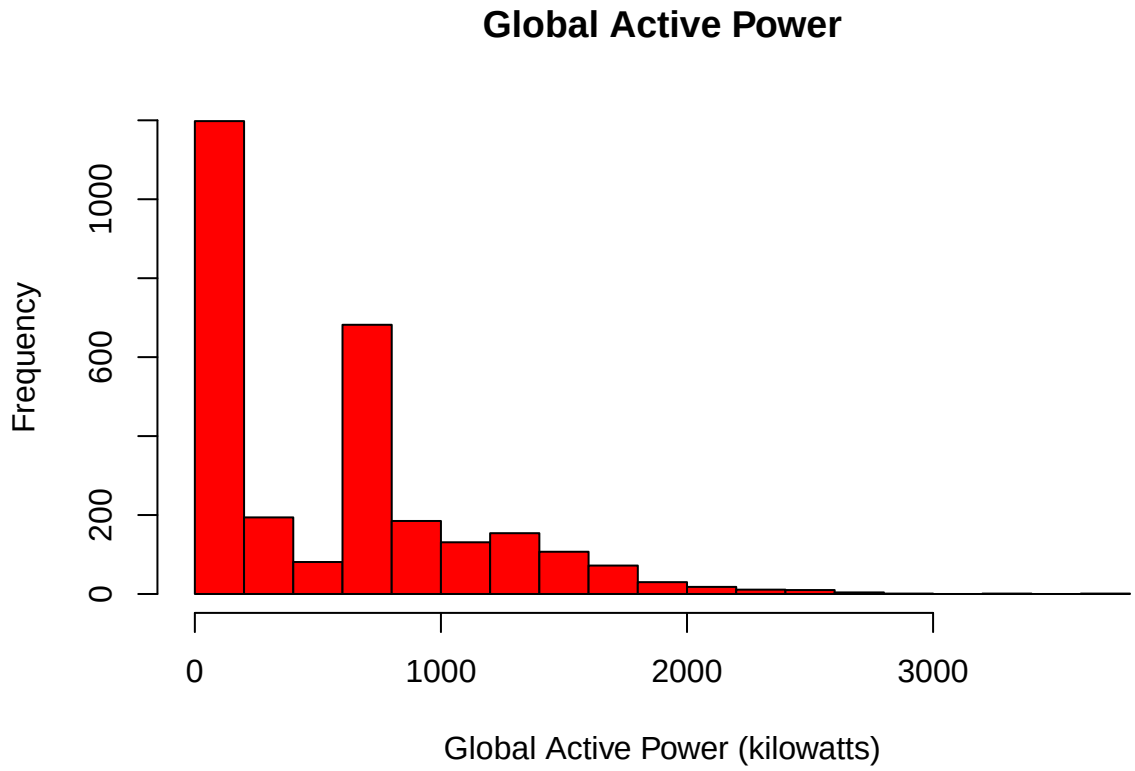
```
new_table<-transform(new_table,timestamp=as.POSIXct(paste(Date, Time)), "%d/%m/%Y %H:%M:%S")
new_table$Global_active_power<-as.numeric(new_table$Global_active_power)
new_table$Global_reactive_power<-as.numeric(new_table$Global_reactive_power)
new_table$Voltage<-as.numeric(new_table$Voltage)
new_table$Global_intensity<-as.numeric(new_table$Global_intensity)
new_table$Sub_metering_1<-as.numeric(new_table$Sub_metering_1)
new_table$Sub_metering_2<-as.numeric(new_table$Sub_metering_2)
head(new_table)
```

```
##             Date     Time Global_active_power Global_reactive_power
## 66637 2007-02-01 00:00:00                 127                    44
## 66638 2007-02-01 00:01:00                 127                    45
## 66639 2007-02-01 00:02:00                 126                    46
## 66640 2007-02-01 00:03:00                 126                    47
## 66641 2007-02-01 00:04:00                 125                    45
## 66642 2007-02-01 00:05:00                 124                    43
##       Voltage Global_intensity Sub_metering_1 Sub_metering_2
## 66637    1823                8             2             2
## 66638    1840                8             2             2
```

```
## 66639    1859                8              2              2
## 66640    1898                8              2              2
## 66641    1824                8              2              2
## 66642    1737                8              2              2
##       Sub_metering_3           timestamp X..d..m..Y..H..M..S.
## 66637              0 2007-02-01 00:00:00    %d/%m/%Y %H:%M:%S
## 66638              0 2007-02-01 00:01:00    %d/%m/%Y %H:%M:%S
## 66639              0 2007-02-01 00:02:00    %d/%m/%Y %H:%M:%S
## 66640              0 2007-02-01 00:03:00    %d/%m/%Y %H:%M:%S
## 66641              0 2007-02-01 00:04:00    %d/%m/%Y %H:%M:%S
## 66642              0 2007-02-01 00:05:00    %d/%m/%Y %H:%M:%S
```

### Plot 1

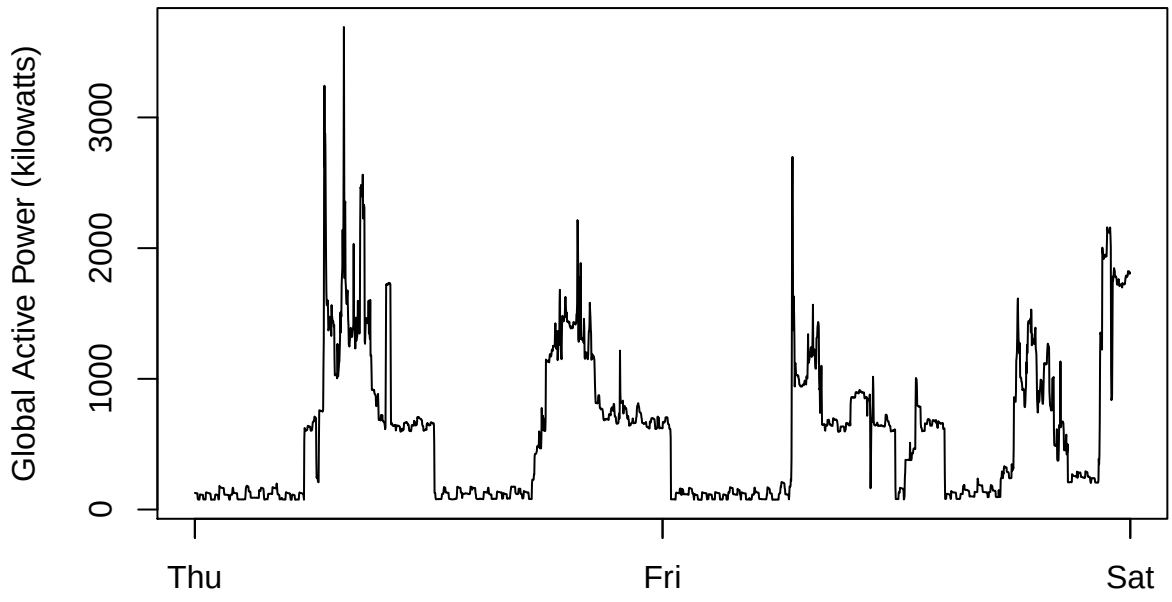
```
plot1 <- function() {
        hist(new_table$Global_active_power, main = paste("Global Active Power"), col="red", xlab="Global
        dev.copy(png, file="plot1.png", width=480, height=480)
        dev.off()
        cat("Plot1.png has been saved in", getwd())
}
plot1()
```




```
## Plot1.png has been saved in C:/Users/ilvb9/Dropbox/Data Science Specialization class/R exercise/Explo
```

### Make plot 2

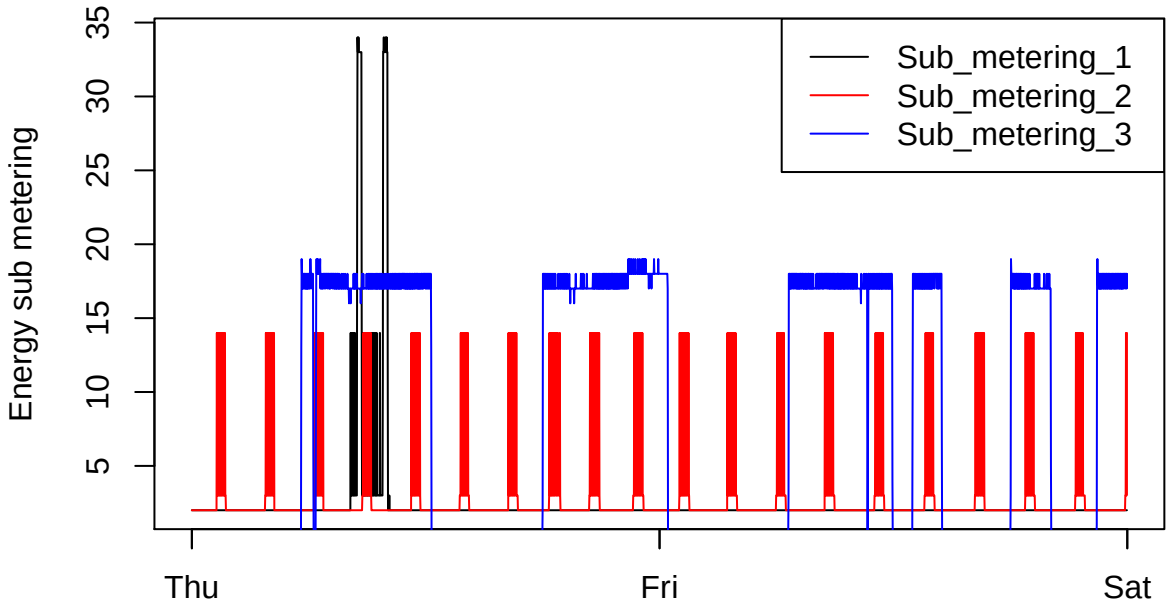
```
plot2 <- function() {
        plot(new_table$timestamp,new_table$Global_active_power, type="l", xlab="", ylab="Global Active P
        dev.copy(png, file="plot2.png", width=480, height=480)
        dev.off()
        cat("plot2.png has been saved in", getwd())
}
plot2()
```




```
## plot2.png has been saved in C:/Users/ilvb9/Dropbox/Data Science Specialization class/R exercise/Explo
```

### Make Plot 3

```
plot3 <- function() {
        plot(new_table$timestamp,new_table$Sub_metering_1, type="l", xlab="", ylab="Energy sub metering
        lines(new_table$timestamp,new_table$Sub_metering_2,col="red")
        lines(new_table$timestamp,new_table$Sub_metering_3,col="blue")
        legend("topright", col=c("black","red","blue"), c("Sub_metering_1  ","Sub_metering_2  ", "Sub_m
        dev.copy(png, file="plot3.png", width=480, height=480)
        dev.off()
        cat("plot3.png has been saved in", getwd())
}
plot3()
```

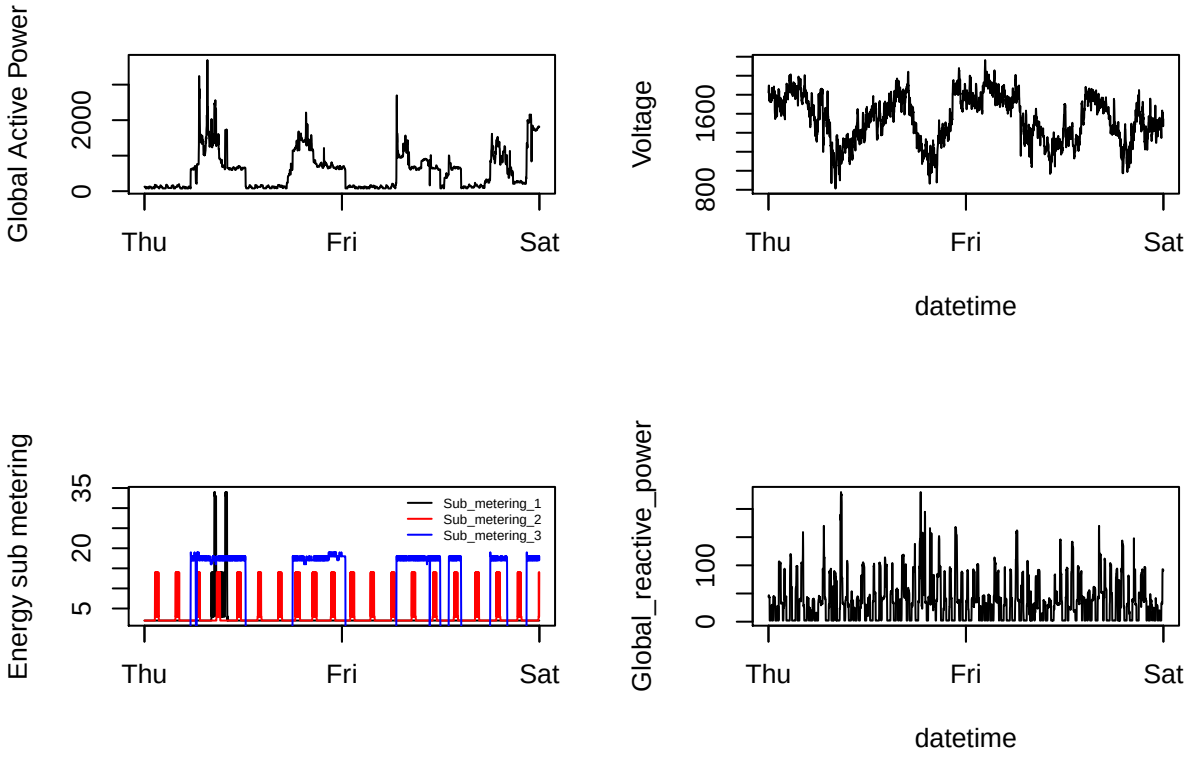



```
## plot3.png has been saved in C:/Users/ilvb9/Dropbox/Data Science Specialization class/R exercise/Explo
```

### Make plot 4

```
plot4 <- function() {
        par(mfrow=c(2,2))

        ##PLOT 1
        plot(new_table$timestamp,new_table$Global_active_power, type="l", xlab="", ylab="Global Active P
        ##PLOT 2
        plot(new_table$timestamp,new_table$Voltage, type="l", xlab="datetime", ylab="Voltage")

        ##PLOT 3
        plot(new_table$timestamp,new_table$Sub_metering_1, type="l", xlab="", ylab="Energy sub metering"
        lines(new_table$timestamp,new_table$Sub_metering_2,col="red")
        lines(new_table$timestamp,new_table$Sub_metering_3,col="blue")
        legend("topright", col=c("black","red","blue"), c("Sub_metering_1  ","Sub_metering_2  ", "Sub_me

        #PLOT 4
        plot(new_table$timestamp,new_table$Global_reactive_power, type="l", xlab="datetime", ylab="Globa

        #OUTPUT
        dev.copy(png, file="plot4.png", width=480, height=480)
        dev.off()
        cat("plot4.png has been saved in", getwd())
```

```
}
plot4()
```




```
## plot4.png has been saved in C:/Users/ilvb9/Dropbox/Data Science Specialization class/R exercise/Explo
```