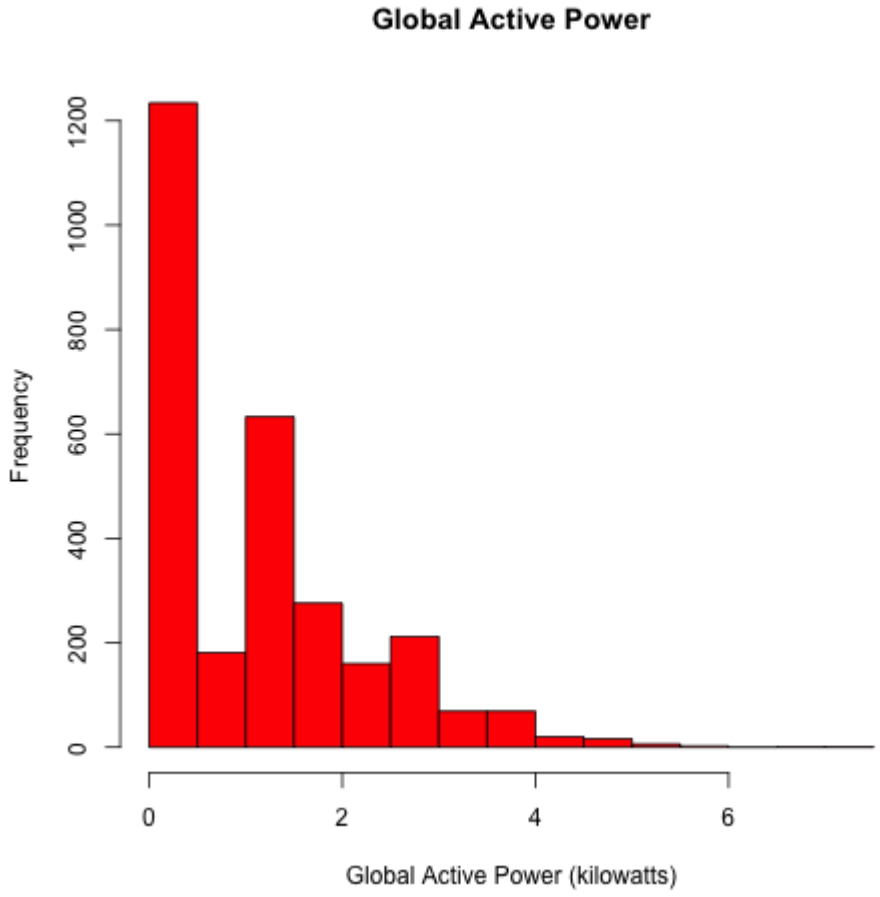
Project Documentation:

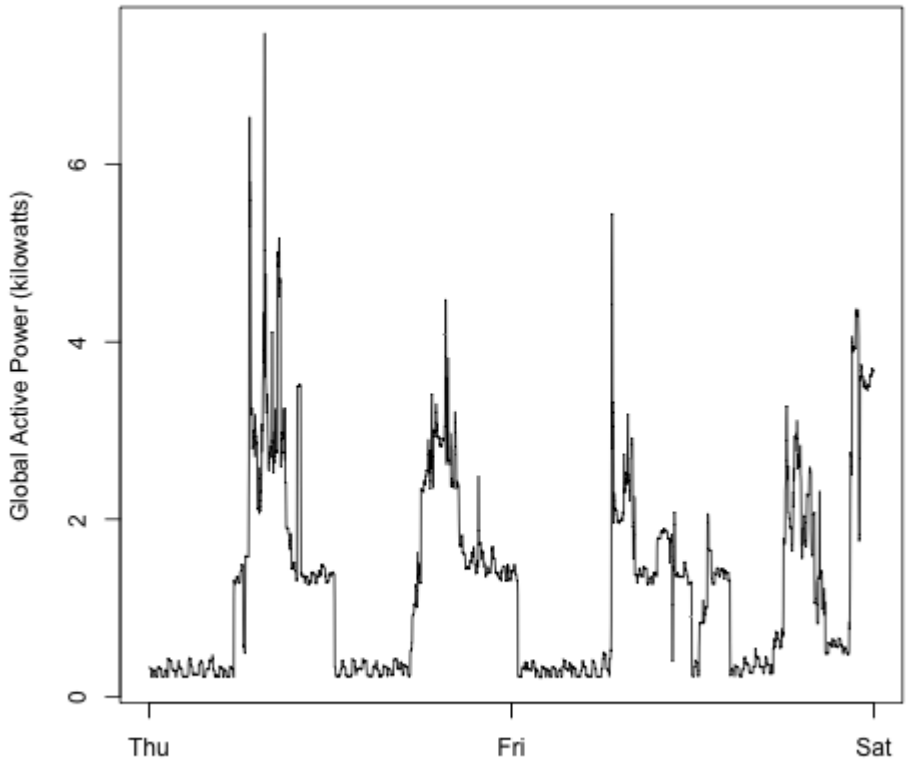
This assignment uses data from the UC Irvine Machine Learning Repository, a popular repository for machine learning datasets. In particular, we will be using the "Individual household electric power consumption Data Set" which I have made available on the course web site:

Dataset: Electric power consumption [20Mb]
Description: Measurements of electric power consumption in one household with a one-minute sampling rate over a period of almost 4 years. Different electrical quantities and some sub-metering values are available.

```
library("data.table")

setwd("~/Desktop/datasciencecoursera/4_Exploratory_Data_Analysis/project/data")

#Reads in data from file then subsets data for specified dates
powerDT <- data.table::fread(input = "household_power_consumption.txt"
                             , na.strings="?"
                             )

# Prevents histogram from printing in scientific notation
powerDT[, Global_active_power := lapply(.SD, as.numeric), .SDcols =
c("Global_active_power")]

# Change Date Column to Date Type
powerDT[, Date := lapply(.SD, as.Date, "%d/%m/%Y"), .SDcols = c("Date")]

# Filter Dates for 2007-02-01 and 2007-02-02
powerDT <- powerDT[(Date >= "2007-02-01") & (Date <= "2007-02-02")]

png("plot1.png", width=480, height=480)

## Plot 1
hist(powerDT[, Global_active_power], main="Global Active Power",
     xlab="Global Active Power (kilowatts)", ylab="Frequency", col="Red")

dev.off()
```

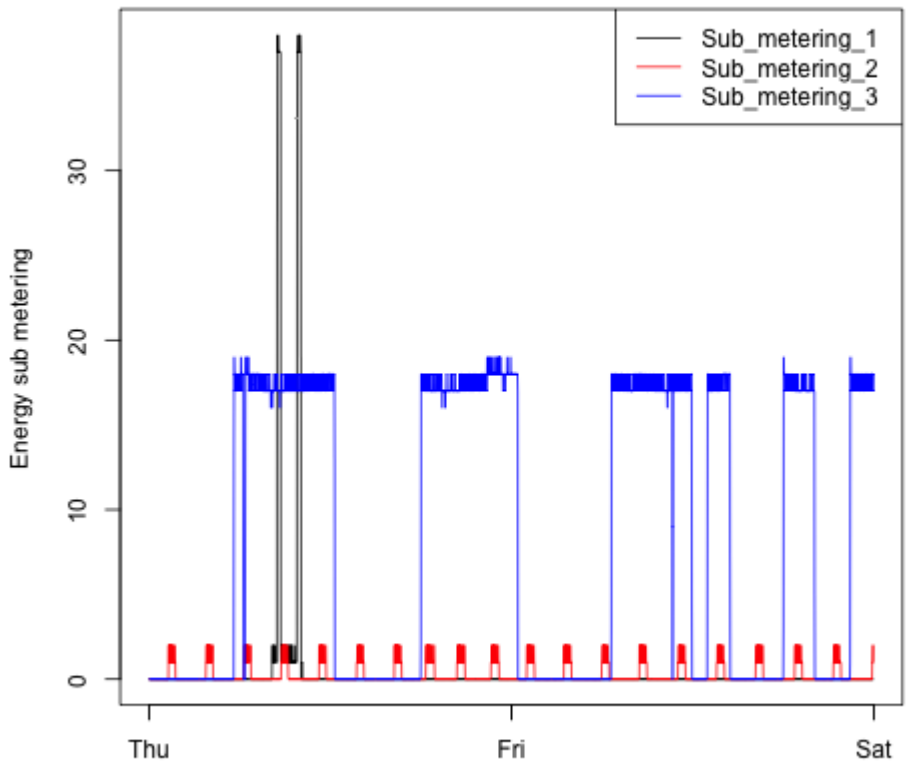



```
library("data.table")

setwd("~/Desktop/datasciencecoursera/4_Exploratory_Data_Analysis/project/data")

#Reads in data from file then subsets data for specified dates
powerDT <- data.table::fread(input = "household_power_consumption.txt"
                             , na.strings="?"
)

# Prevents Scientific Notation
powerDT[, Global_active_power := lapply(.SD, as.numeric), .SDcols =
c("Global_active_power")]

# Making a POSIXct date capable of being filtered and graphed by time of day
powerDT[, dateTime := as.POSIXct(paste(Date, Time), format = "%d/%m/%Y %H:%M:%S")]

# Filter Dates for 2007-02-01 and 2007-02-02
powerDT <- powerDT[(dateTime >= "2007-02-01") & (dateTime < "2007-02-03")]

png("plot2.png", width=480, height=480)

## Plot 2
plot(x = powerDT[, dateTime]
     , y = powerDT[, Global_active_power]
     , type="l", xlab="", ylab="Global Active Power (kilowatts)")

dev.off()
```




```
library("data.table")

setwd("~/Desktop/datasciencecoursera/4_Exploratory_Data_Analysis/project/data")

#Reads in data from file then subsets data for specified dates
powerDT <- data.table::fread(input = "household_power_consumption.txt"
                             , na.strings="?"
)

# Prevents Scientific Notation
powerDT[, Global_active_power := lapply(.SD, as.numeric), .SDcols =
c("Global_active_power")]

# Making a POSIXct date capable of being filtered and graphed by time of day
powerDT[, dateTime := as.POSIXct(paste(Date, Time), format = "%d/%m/%Y %H:%M:%S")]

# Filter Dates for 2007-02-01 and 2007-02-02
powerDT <- powerDT[(dateTime >= "2007-02-01") & (dateTime < "2007-02-03")]

png("plot3.png", width=480, height=480)

# Plot 3
plot(powerDT[, dateTime], powerDT[, Sub_metering_1], type="l", xlab="",
ylab="Energy sub metering")
lines(powerDT[, dateTime], powerDT[, Sub_metering_2],col="red")
lines(powerDT[, dateTime], powerDT[, Sub_metering_3],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))

dev.off()
```




```
library("data.table")

setwd("~/Desktop/datasciencecoursera/4_Exploratory_Data_Analysis/project/data")

#Reads in data from file then subsets data for specified dates
powerDT <- data.table::fread(input = "household_power_consumption.txt"
                             , na.strings="?"
)

# Prevents Scientific Notation
powerDT[, Global_active_power := lapply(.SD, as.numeric), .SDcols =
c("Global_active_power")]

# Making a POSIXct date capable of being filtered and graphed by time of day
powerDT[, dateTime := as.POSIXct(paste(Date, Time), format = "%d/%m/%Y %H:%M:%S")]

# Filter Dates for 2007-02-01 and 2007-02-02
powerDT <- powerDT[(dateTime >= "2007-02-01") & (dateTime < "2007-02-03")]

png("plot4.png", width=480, height=480)

par(mfrow=c(2,2))
```

```
# Plot 1
plot(powerDT[, dateTime], powerDT[, Global_active_power], type="l", xlab="",
ylab="Global Active Power")

# Plot 2
plot(powerDT[, dateTime],powerDT[, Voltage], type="l", xlab="datetime",
ylab="Voltage")

# Plot 3
plot(powerDT[, dateTime], powerDT[, Sub_metering_1], type="l", xlab="",
ylab="Energy sub metering")
lines(powerDT[, dateTime], powerDT[, Sub_metering_2], col="red")
lines(powerDT[, dateTime], powerDT[, Sub_metering_3],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 4
plot(powerDT[, dateTime], powerDT[,Global_reactive_power], type="l",
xlab="datetime", ylab="Global_reactive_power")

dev.off()
```