**Exploratory Data Analysis Project 1**

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](https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip) [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")]

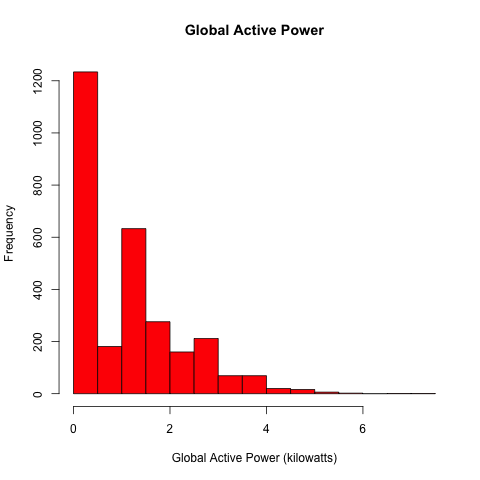
png("plot1.png", width=480, height=480)

## Plot 1

hist(powerDT[, Global\_active\_power], main="Global Active Power",

xlab="Global Active Power (kilowatts)", ylab="Frequency", col="Red")

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project1/plot1.png)

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)

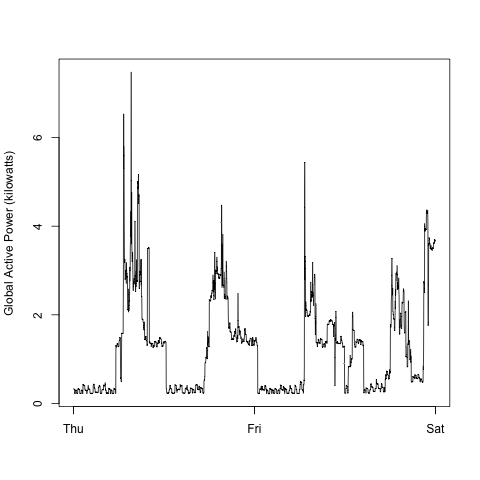
## Plot 2

plot(x = powerDT[, dateTime]

, y = powerDT[, Global\_active\_power]

, type="l", xlab="", ylab="Global Active Power (kilowatts)")

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project1/plot2.png)

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")

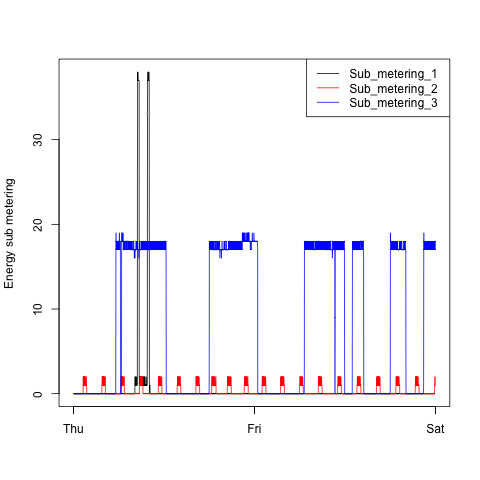
legend("topright"

, 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()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project1/plot3.png)

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)

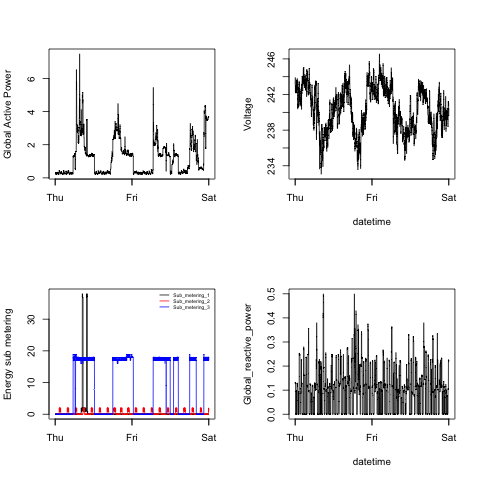
, bty="n"

, cex=.5)

# Plot 4

plot(powerDT[, dateTime], powerDT[,Global\_reactive\_power], type="l", xlab="datetime", ylab="Global\_reactive\_power")

dev.off()

[](https://github.com/mGalarnyk/datasciencecoursera/blob/master/4_Exploratory_Data_Analysis/project1/plot4.png)