**-days household energy usage pattern**

***Nicholas kelcz***

***Sunday, July 03, 2016***

**Introduction**

This assignment uses data from the [UC Irvine Machine Learning Repository](http://archive.ics.uci.edu/ml/), a popular repository for machine learning datasets. In particular, we will be using the “Individual household electric power consumption Data Set”.

**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.

The following descriptions of the 9 variables in the dataset are taken from the [UCI web site](https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption):

1. Date: Date in format dd/mm/yyyy
2. Time: time in format hh:mm:ss
3. Global\_active\_power: household global minute-averaged active power (in kilowatt)
4. Global\_reactive\_power: household global minute-averaged reactive power (in kilowatt)
5. Voltage: minute-averaged voltage (in volt)
6. Global\_intensity: household global minute-averaged current intensity (in ampere)
7. Sub\_metering\_1: energy sub-metering No. 1 (in watt-hour of active energy). It corresponds to the kitchen, containing mainly a dishwasher, an oven and a microwave (hot plates are not electric but gas powered).
8. Sub\_metering\_2: energy sub-metering No. 2 (in watt-hour of active energy). It corresponds to the laundry room, containing a washing-machine, a tumble-drier, a refrigerator and a light.
9. Sub\_metering\_3: energy sub-metering No. 3 (in watt-hour of active energy). It corresponds to an electric water-heater and an air-conditioner.

**Loading the data**

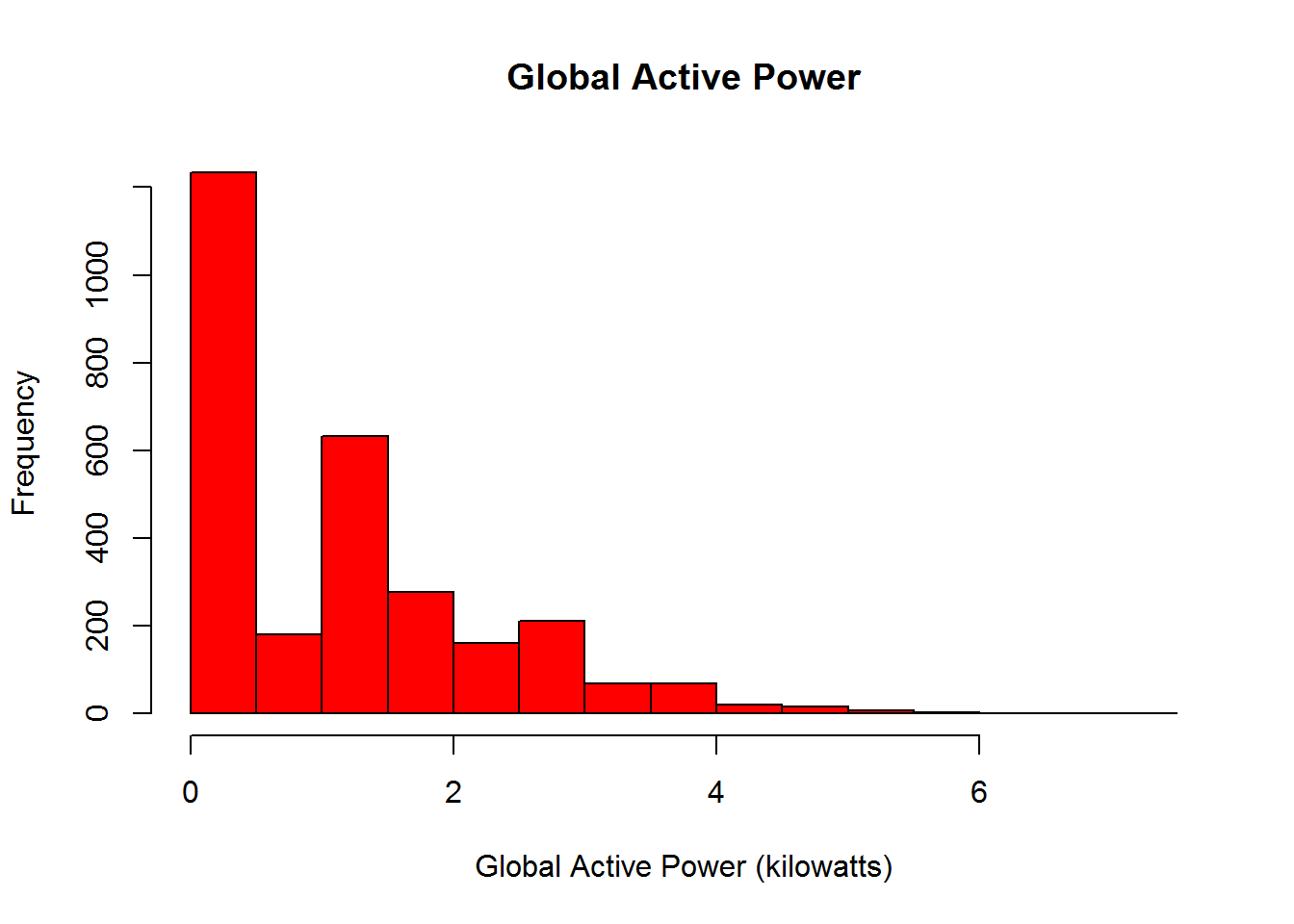
When loading the dataset into R, the following to be considered:

* The dataset has 2,075,259 rows and 9 columns. First a rough estimate of how much memory the dataset will require in memory before reading into R should be calculated. Computer should have enough memory (most modern computers should be fine).
* We will only be using data from the dates 2007-02-01 and 2007-02-02. One alternative is to read the data from just those dates rather than reading in the entire dataset and subsetting to those dates.
* It may be useful to convert the Date and Time variables to Date/Time classes in R using the strptime() and as.Date() functions.
* In this dataset missing values are coded as ?.

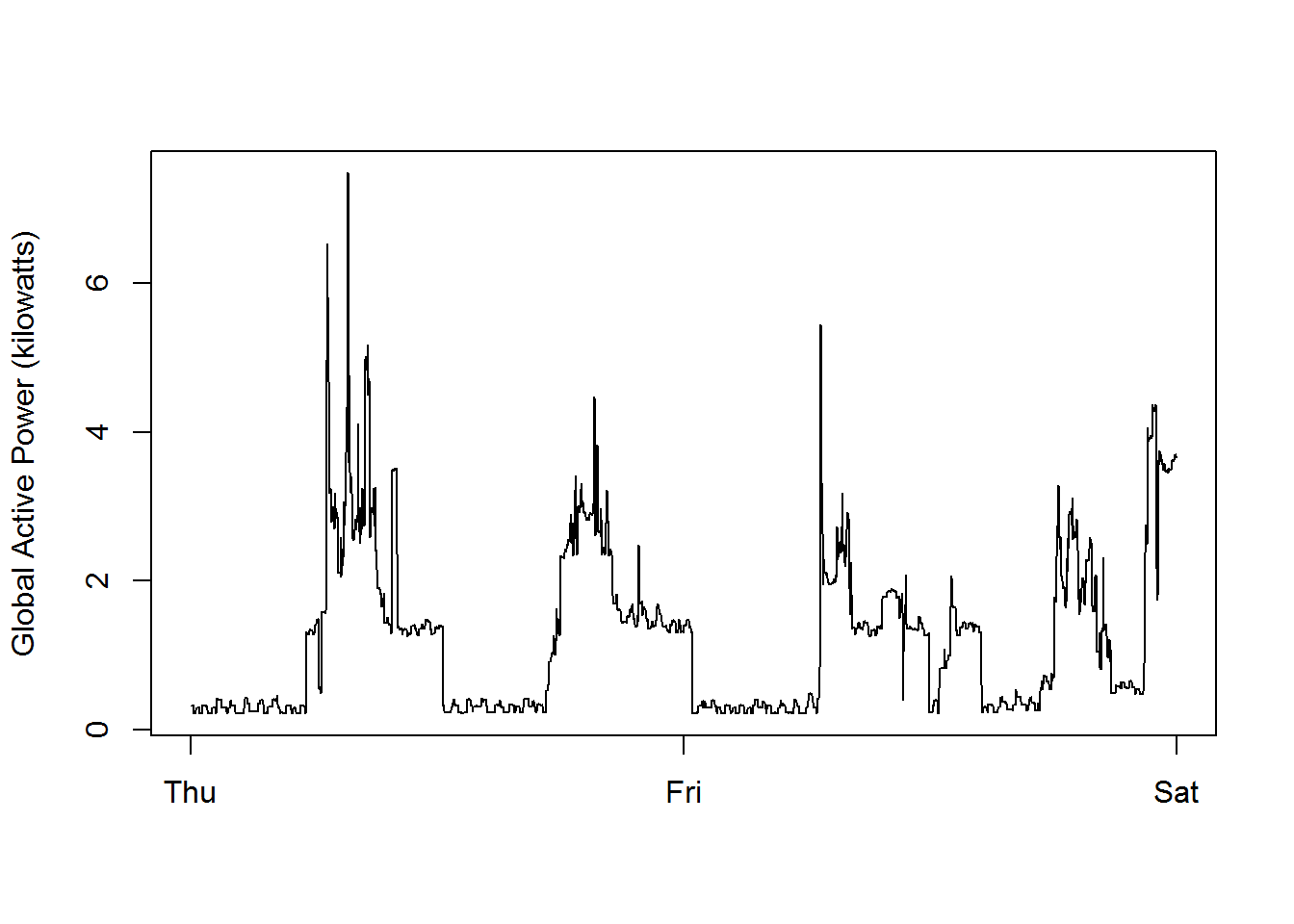
**Making Plots**

Our overall goal here is simply to examine how household energy usage varies over a 2-day period in February, 2007. Our task is to construct the plots using the base plotting system.

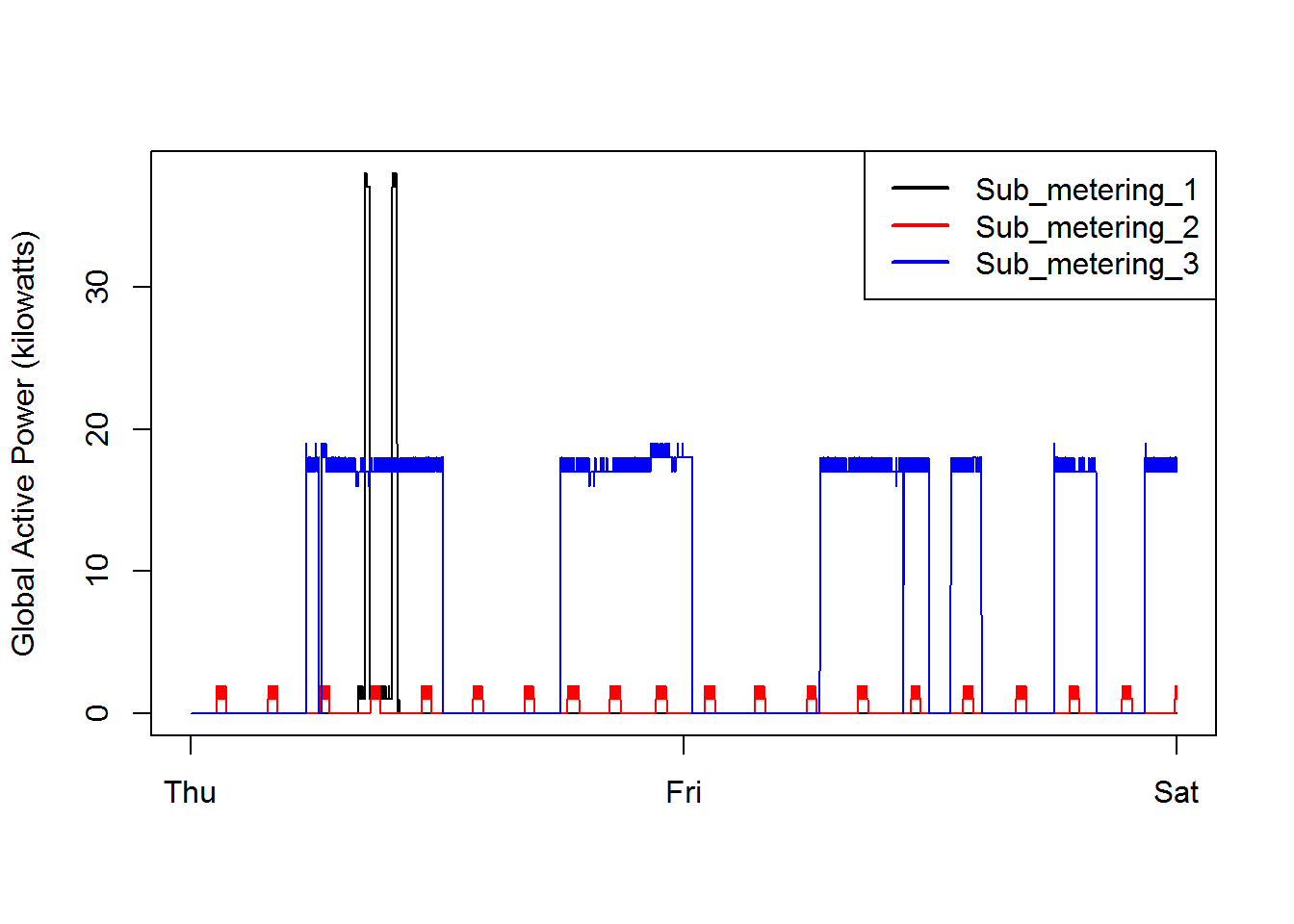
fh <- file("household\_power\_consumption.txt")  
ba <- read.table(text = grep("^[1,2]/2/2007", readLines(fh), value = TRUE), col.names = c("Date", "Time", "Global\_active\_power", "Global\_reactive\_power", "Voltage", "Global\_intensity", "Sub\_metering\_1", "Sub\_metering\_2", "Sub\_metering\_3"), sep = ";", header = TRUE)  
  
*# Generating Plot 1*  
hist(ba$Global\_active\_power, col = "red", main = paste("Global Active Power"), xlab = "Global Active Power (kilowatts)")



*## Getting full dataset*  
data\_full <- read.csv("household\_power\_consumption.txt", header = T, sep = ';',   
                     na.strings = "?", nrows = 2075259, check.names = F,   
                     stringsAsFactors = F, comment.char = "", quote = '\"')  
data\_full$Date <- as.Date(data\_full$Date, format = "%d/%m/%Y")  
  
*## Subsetting the data*  
data <- subset(data\_full, subset = (Date >= "2007-02-01" & Date <= "2007-02-02"))  
rm(data\_full)  
  
*## Converting dates*  
datetime <- paste(as.Date(data$Date), data$Time)  
data$Datetime <- as.POSIXct(datetime)  
  
*## Generating Plot 2*  
plot(data$Global\_active\_power ~ data$Datetime, type = "l",  
    ylab = "Global Active Power (kilowatts)", xlab = "")



*## Generating Plot 3*  
with(data, {  
   plot(Sub\_metering\_1 ~ Datetime, type = "l",   
   ylab = "Global Active Power (kilowatts)", xlab = "")  
   lines(Sub\_metering\_2 ~ Datetime, col = 'Red')  
   lines(Sub\_metering\_3 ~ Datetime, col = 'Blue')  
})  
legend("topright", col = c("black", "red", "blue"), lty = 1, lwd = 2,   
      legend = c("Sub\_metering\_1", "Sub\_metering\_2", "Sub\_metering\_3"))



*## Generating Plot 4*  
par(mfrow = c(2,2), mar = c(4,4,2,1), oma = c(0,0,2,0))  
with(data, {  
    plot(Global\_active\_power ~ Datetime, type = "l",   
    ylab = "Global Active Power", xlab = "")  
    plot(Voltage ~ Datetime, type = "l", ylab = "Voltage", xlab = "datetime")  
    plot(Sub\_metering\_1 ~ Datetime, type = "l", ylab = "Energy sub metering",  
         xlab = "")  
    lines(Sub\_metering\_2 ~ Datetime, col = 'Red')  
    lines(Sub\_metering\_3 ~ Datetime, col = 'Blue')  
    legend("topright", col = c("black", "red", "blue"), lty = 1, lwd = 2,   
            bty = "n",  
            legend = c("Sub\_metering\_1", "Sub\_metering\_2", "Sub\_metering\_3"))  
    plot(Global\_reactive\_power ~ Datetime, type = "l",   
         ylab = "Global\_rective\_power", xlab = "datetime")  
})

