Exploratory analysis project 1

Haichen

1/6/2021

Overview

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”. The dataset has 2,075,259 rows and 9 columns. First calculate a rough estimate of how much memory the dataset will require in memory before reading into R. Make sure your computer has 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.

Get Data

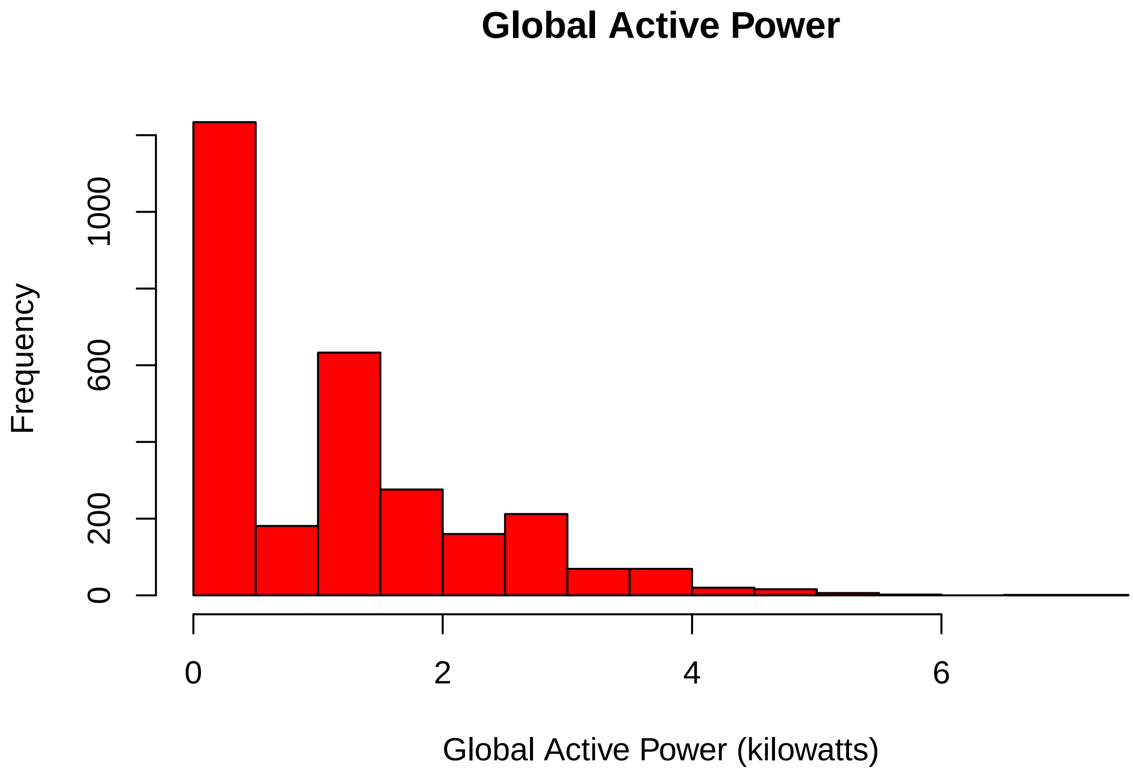
```
setwd("C:/Coursera/Exploratory data analysis")
data <- read.table("household_power_consumption.txt", header= TRUE, sep=";", stringsAsFactors=FALSE, de
subsetdata <- data[data$Date %in% c("1/2/2007","2/2/2007"),]
str(subsetdata)
```

```
## 'data.frame':    2880 obs. of  9 variables:
##  $ Date                 : chr  "1/2/2007" "1/2/2007" "1/2/2007" "1/2/2007" ...
##  $ Time                 : chr  "00:00:00" "00:01:00" "00:02:00" "00:03:00" ...
##  $ Global_active_power  : chr  "0.326" "0.326" "0.324" "0.324" ...
##  $ Global_reactive_power: chr  "0.128" "0.130" "0.132" "0.134" ...
##  $ Voltage              : chr  "243.150" "243.320" "243.510" "243.900" ...
##  $ Global_intensity     : chr  "1.400" "1.400" "1.400" "1.400" ...
##  $ Sub_metering_1       : chr  "0.000" "0.000" "0.000" "0.000" ...
##  $ Sub_metering_2       : chr  "0.000" "0.000" "0.000" "0.000" ...
##  $ Sub_metering_3       : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
globalActivePower <- as.numeric(subsetdata$Global_active_power)
globalReactivePower <- as.numeric(subsetdata$Global_reactive_power)
voltage <- as.numeric(subsetdata$Voltage)
subMetering1 <- as.numeric(subsetdata$Sub_metering_1)
subMetering2 <- as.numeric(subsetdata$Sub_metering_2)
subMetering3 <- as.numeric(subsetdata$Sub_metering_3)
```

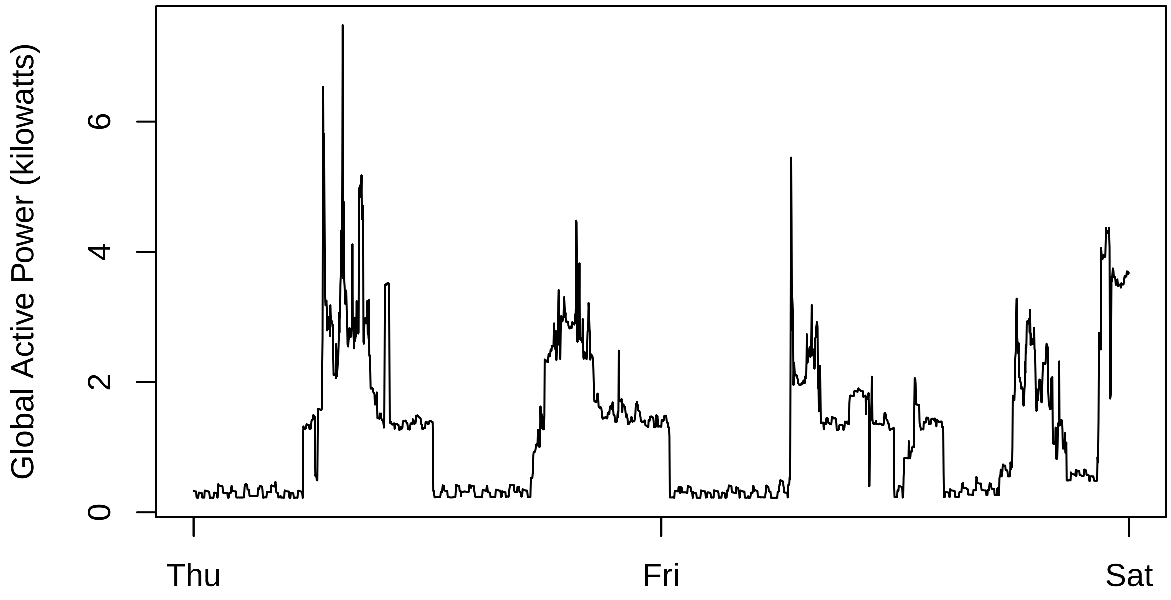
R Histogram

```
hist(globalActivePower, col="red", main="Global Active Power", xlab="Global Active Power (kilowatts)")
```



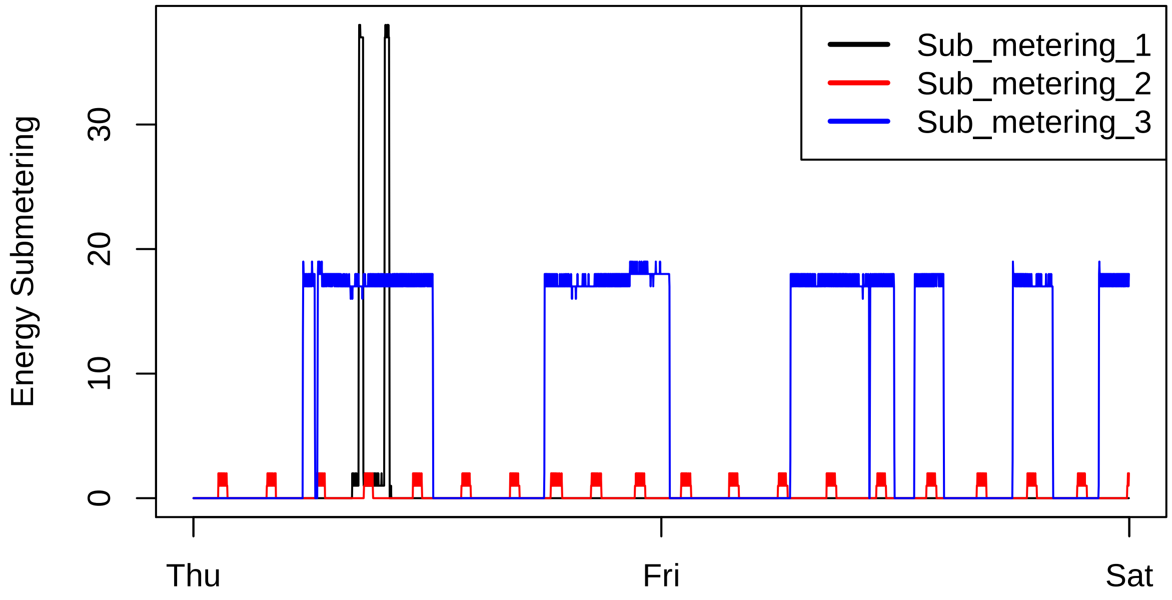
Time Series

```
datetime <- strptime(paste(subsetdata$Date, subsetdata$Time, sep=" "), "%d/%m/%Y %H:%M:%S")
plot(datetime, globalActivePower, type="l", xlab="", ylab="Global Active Power (kilowatts)")
```



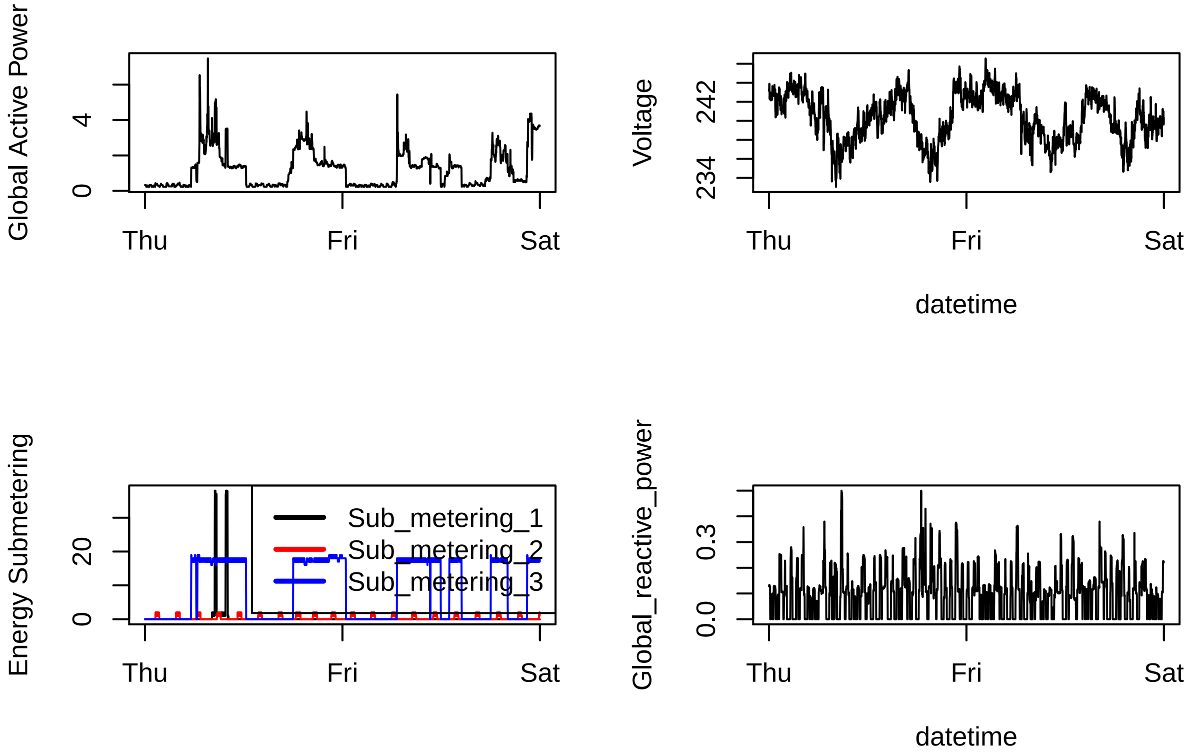
Plot for sub metering

```
plot(datetime, subMetering1, type="l", ylab="Energy Submetering", xlab="")
lines(datetime, subMetering2, type="l", col="red")
lines(datetime, subMetering3, type="l", col="blue")
legend("topright", c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"), lty=1, lwd=2.5, col=c("blac
```



Multiple plot

```
par(mfrow = c(2, 2))
# First plot
plot(datetime, globalActivePower, type="l", xlab="", ylab="Global Active Power", cex=0.2)
# Second plot
plot(datetime, voltage, type="l", xlab="datetime", ylab="Voltage")
# Third plot
plot(datetime, subMetering1, type="l", ylab="Energy Submetering", xlab="")
lines(datetime, subMetering2, type="l", col="red")
lines(datetime, subMetering3, type="l", col="blue")
legend("topright", c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"), lty=, lwd=2.5, col=c("black
# Fourth plot
plot(datetime, globalReactivePower, type="l", xlab="datetime", ylab="Global_reactive_power", cex=0.2)
```


Global Active Power
0
4
Thu
Fri
Sat
Voltage
234
242
Thu
Fri
Sat
datetime
Energy Submetering
0
20
Sub_metering_1
Sub_metering_2
Sub_metering_3
Thu
Fri
Sat
Global_reactive_power
0.0
0.3
Thu
Fri
Sat
datetime