Exploratory Data Analysis Assignment 1

Myron Keith Gibert Jr

8/4/2020

Introduction

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.

The following descriptions of the 9 variables in the dataset are taken from the UCI web site:

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.

I had to re-generate the 4 example figures using the provided dataset. The example figures can be found (here)[https://www.coursera.org/learn/exploratory-data-analysis/peer/ylVFo/course-project-1].

I created 4 individual R scripts (plot1.R, plot2.R, plot3.R, plot4.R) to regenerate the example plots. Each file downloads and unzips the data if the directory does not already exist. It then reads in the data, subsets it according to what is needed for each plot, generates the appropriate plots, and outputs a corresponding Portable Network Graphics (PNG) file (plot1.png, etc.).

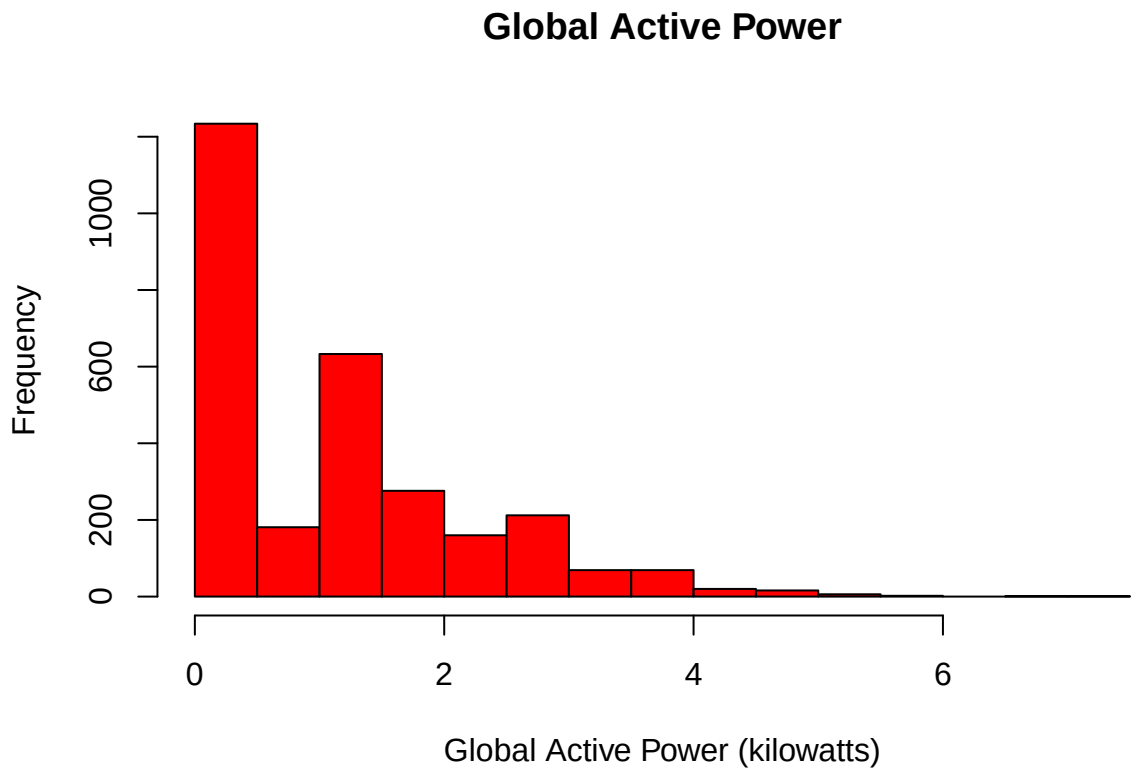
Figure 1

```
if(!file.exists("exdata_data_household_power_consumption.zip")){
  download.file(
"https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip",
```

```
"exdata_data_household_power_consumption.zip", quiet = TRUE)}

if(!dir.exists("household_power_consumption.txt")){
  unzip("exdata_data_household_power_consumption.zip")
}

data <- read.csv("household_power_consumption.txt", header=T, sep=';', na.strings="?",
                 nrows=2075259, check.names=F, stringsAsFactors=F,
                 comment.char="", quote='\"')
data2 <- subset(data, Date %in% c("1/2/2007","2/2/2007"))
data2$Date <- as.Date(data2$Date, format="%d/%m/%Y")
hist(data2$Global_active_power, main="Global Active Power",
     xlab="Global Active Power (kilowatts)", ylab="Frequency", col="Red")
```




```
dev.copy(png, file="plot1.png", height=480, width=480)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

Including Plots

You can also embed plots, for example:

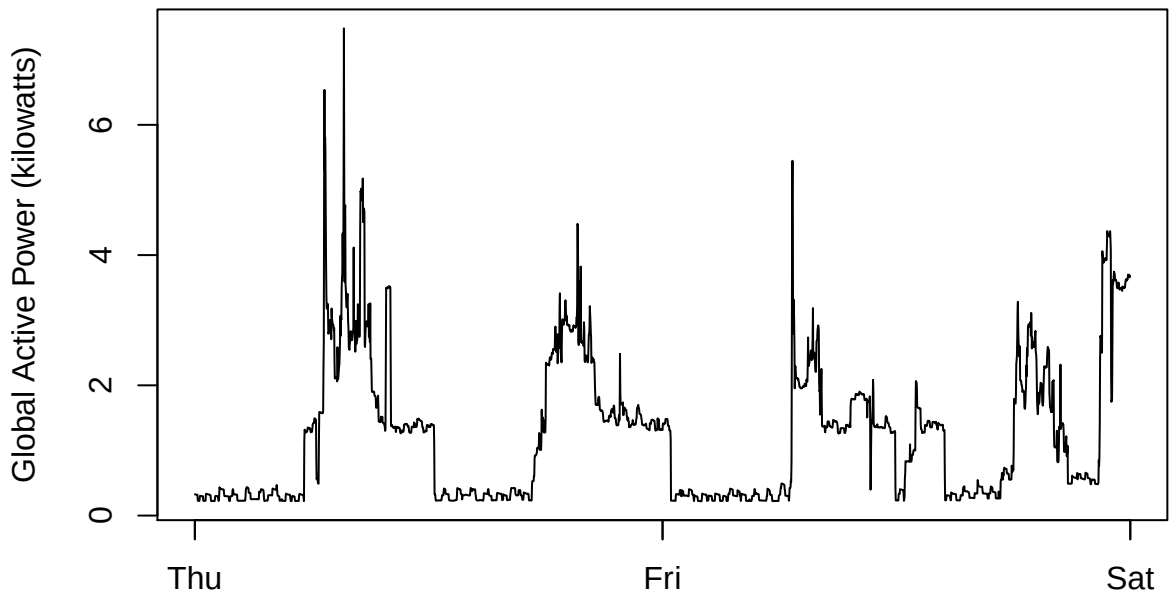
```
if(!file.exists("exdata_data_household_power_consumption.zip")){
  download.file(
"https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip",
"exdata_data_household_power_consumption.zip", quiet = TRUE)}

if(!dir.exists("household_power_consumption.txt")){
  unzip("exdata_data_household_power_consumption.zip")
}

data <- read.csv("household_power_consumption.txt", header=T, sep=';', na.strings="?",
                 nrows=2075259, check.names=F, stringsAsFactors=F,
                 comment.char="", quote='\"')
data2 <- subset(data, Date %in% c("1/2/2007","2/2/2007"))
data2$Date <- as.Date(data2$Date, format="%d/%m/%Y")

datetime <- paste(as.Date(data2$Date), data2$Time)
data2$Datetime <- as.POSIXct(datetime)

## Plot 2
with(data2, {
  plot(Global_active_power~Datetime, type="l",
       ylab="Global Active Power (kilowatts)", xlab="")
})
```




```
dev.copy(png, file="plot2.png", height=480, width=480)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

Figure 3

You can also embed plots, for example:

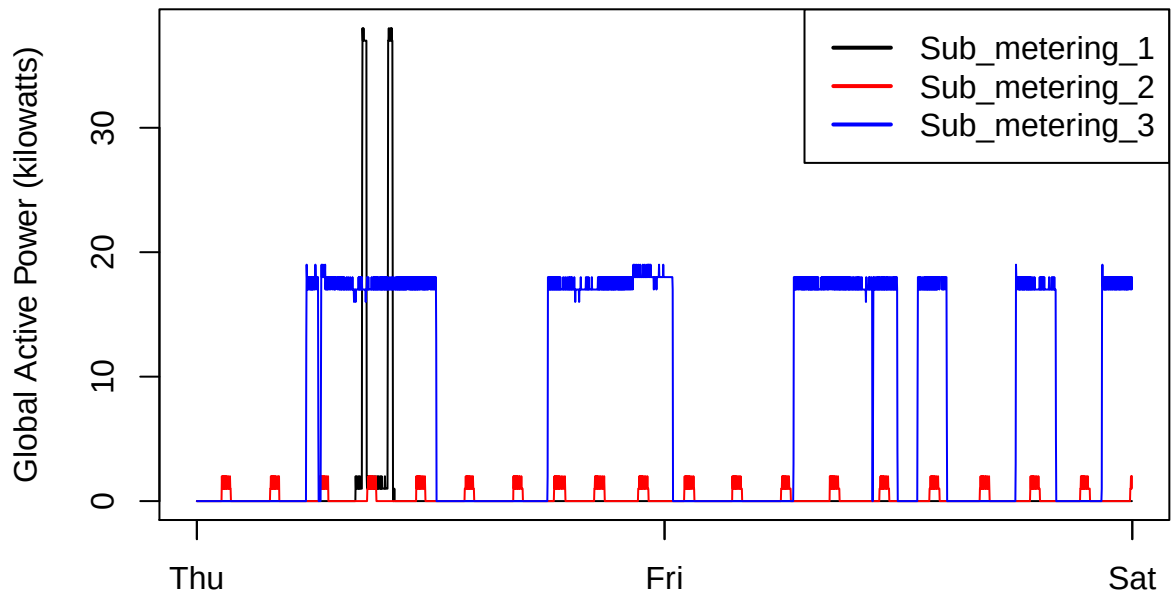
```
if(!file.exists("exdata_data_household_power_consumption.zip")){
  download.file(
"https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip",
"exdata_data_household_power_consumption.zip", quiet = TRUE)}

if(!dir.exists("household_power_consumption.txt")){
  unzip("exdata_data_household_power_consumption.zip")
}
```

```
data <- read.csv("household_power_consumption.txt", header=T, sep=';', na.strings="?",
                 nrows=2075259, check.names=F, stringsAsFactors=F,
                 comment.char="", quote='\"')
data2 <- subset(data, Date %in% c("1/2/2007","2/2/2007"))
data2$Date <- as.Date(data2$Date, format="%d/%m/%Y")

datetime <- paste(as.Date(data2$Date), data2$Time)
data2$Datetime <- as.POSIXct(datetime)

with(data2, {
  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"))
```




```
dev.copy(png, file="plot3.png", height=480, width=480)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

Figure 4

You can also embed plots, for example:

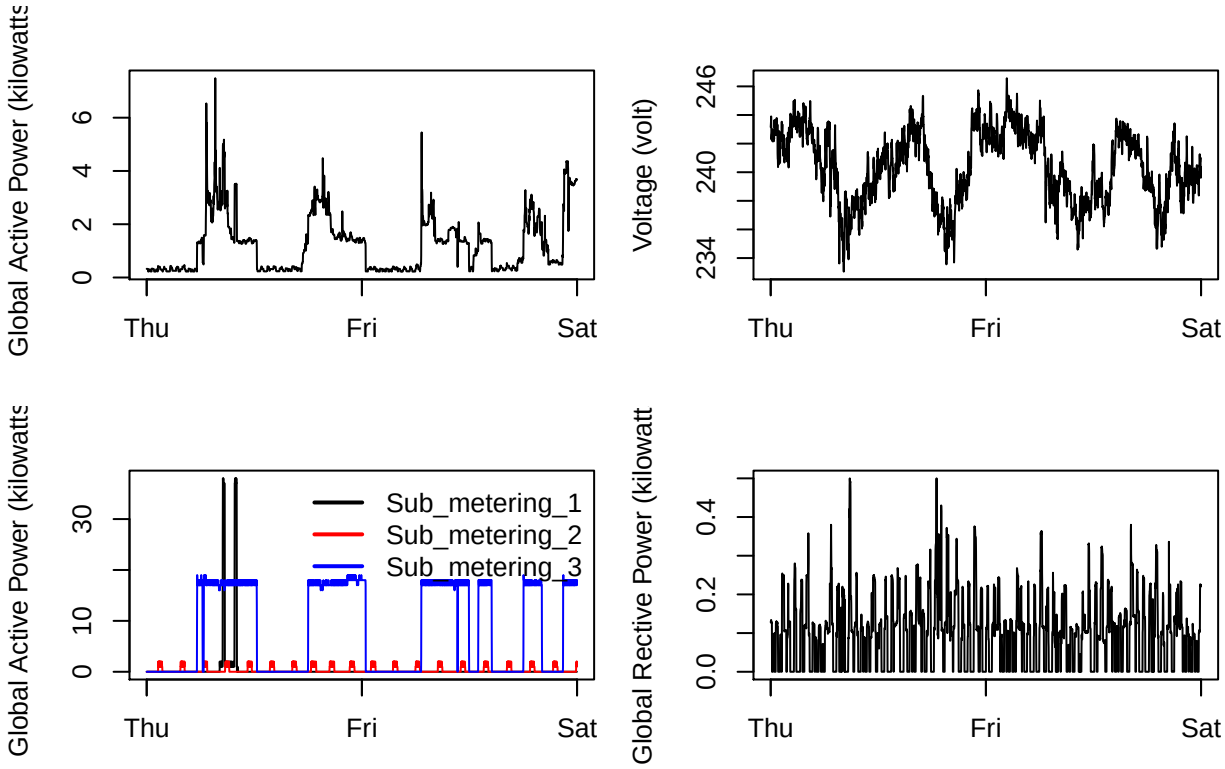
```
if(!file.exists("exdata_data_household_power_consumption.zip")){
  download.file(
"https://d396qusza40orc.cloudfront.net/exdata%2Fdata%2Fhousehold_power_consumption.zip",
"exdata_data_household_power_consumption.zip", quiet = TRUE)}

if(!dir.exists("household_power_consumption.txt")){
  unzip("exdata_data_household_power_consumption.zip")
}

data <- read.csv("household_power_consumption.txt", header=T, sep=';', na.strings="?",
                 nrows=2075259, check.names=F, stringsAsFactors=F,
                 comment.char="", quote='\"')
data2 <- subset(data, Date %in% c("1/2/2007","2/2/2007"))
data2$Date <- as.Date(data2$Date, format="%d/%m/%Y")

datetime <- paste(as.Date(data2$Date), data2$Time)
data2$Datetime <- as.POSIXct(datetime)

par(mfrow=c(2,2), mar=c(4,4,2,1), oma=c(0,0,2,0))
with(data2, {
  plot(Global_active_power~Datetime, type="l",
       ylab="Global Active Power (kilowatts)", xlab="")
  plot(Voltage~Datetime, type="l",
       ylab="Voltage (volt)", xlab="")
  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, bty="n",
         legend=c("Sub_metering_1", "Sub_metering_2", "Sub_metering_3"))
  plot(Global_reactive_power~Datetime, type="l",
       ylab="Global Rective Power (kilowatts)",xlab="")
})
```




```
dev.copy(png, file="plot4.png", height=480, width=480)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```

```
dev.copy(png, file="plot4.png", height=480, width=480)
```

```
## png
##   3
```

```
dev.off()
```

```
## pdf
##   2
```