Instructions

Instructions of Course Project 1

This is an R Markdown document. It simplify the code understanding and help the reader to get the requried information easly.

Introduction

It is now possible to collect a large amount of data about personal movement using activity monitoring devices such as a Fitbit, Nike Fuelband, or Jawbone Up. These type of devices are part of the "quantified self" movement – a group of enthusiasts who take measurements about themselves regularly to improve their health, to find patterns in their behavior, or because they are tech geeks. But this data remains under-utilized because the raw data is hard to obtain and there are limited tools and statistical methods available for interpreting the data. This assignment makes use of data from a personal activity monitoring device. This device collects data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

Data

The data for this assignment can be downloaded from the course web site: Dataset: Activity monitoring data [52K] The variables included in this dataset are: steps: Number of steps taking in a 5-minute interval (missing values are coded as NA) date: The date on which the measurement was taken in YYYY-MM-DD format interval: Identifier for the 5-minute interval in which measurement was taken The dataset is stored in a comma-separated-value (CSV) file and there are a total of 17,568 observations in this dataset.

This Code for reading in the dataset and/or processing the data:

```
#the data is already saved in the project folder
getwd()
```

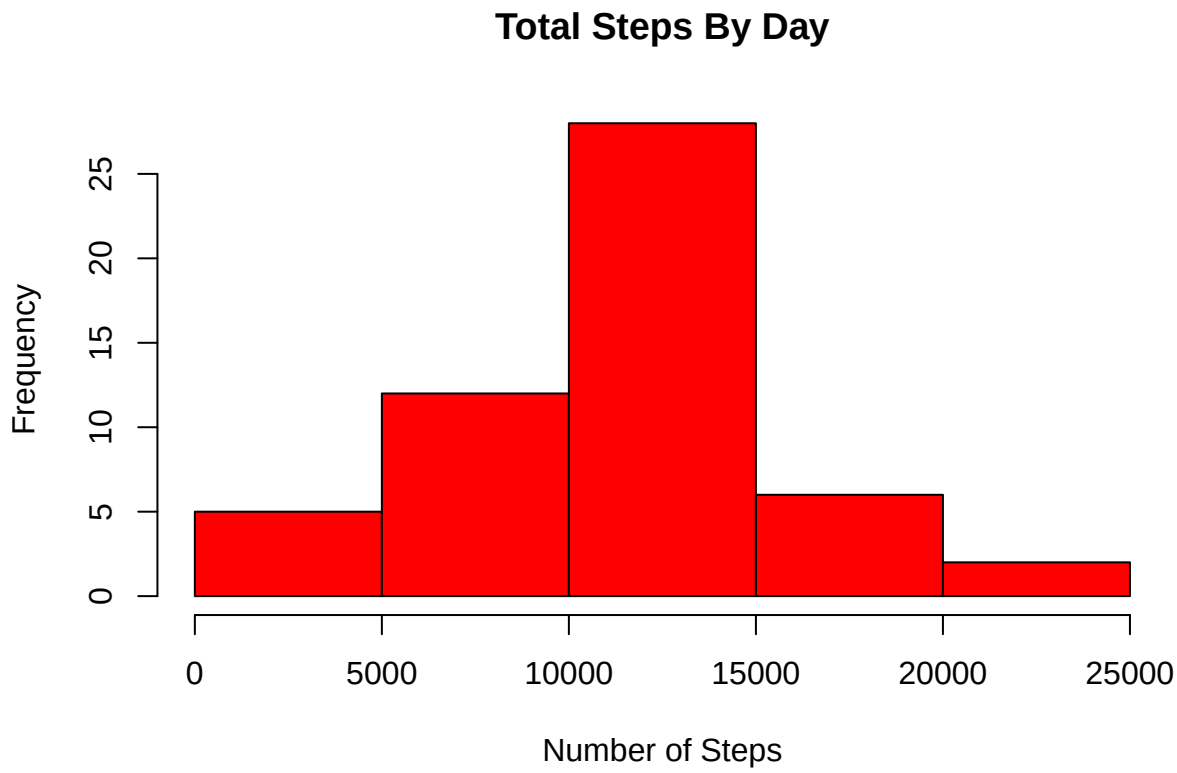
```
## [1] "C:/Users/saeed/Desktop/hti/programnig/R/ass.7"
```

```
data <- read.csv("C:/Users/saeed/Desktop/hti/programnig/R/ass.7/activity.csv")
head(data)
```

```
##   steps       date interval
## 1    NA 2012-10-01        0
## 2    NA 2012-10-01        5
## 3    NA 2012-10-01       10
## 4    NA 2012-10-01       15
## 5    NA 2012-10-01       20
## 6    NA 2012-10-01       25
```

Code for histogram of the total number of steps taken each day:

```
Steps.By.Day<-aggregate(steps ~ date, data, sum)
hist(Steps.By.Day$steps, main = "Total Steps By Day", col="red",xlab="Number of Steps")
```



Code for calculating Mean and median number of steps taken each day:

```
Mean<- mean(Steps.By.Day$steps, rm.na=T)
Mean
```

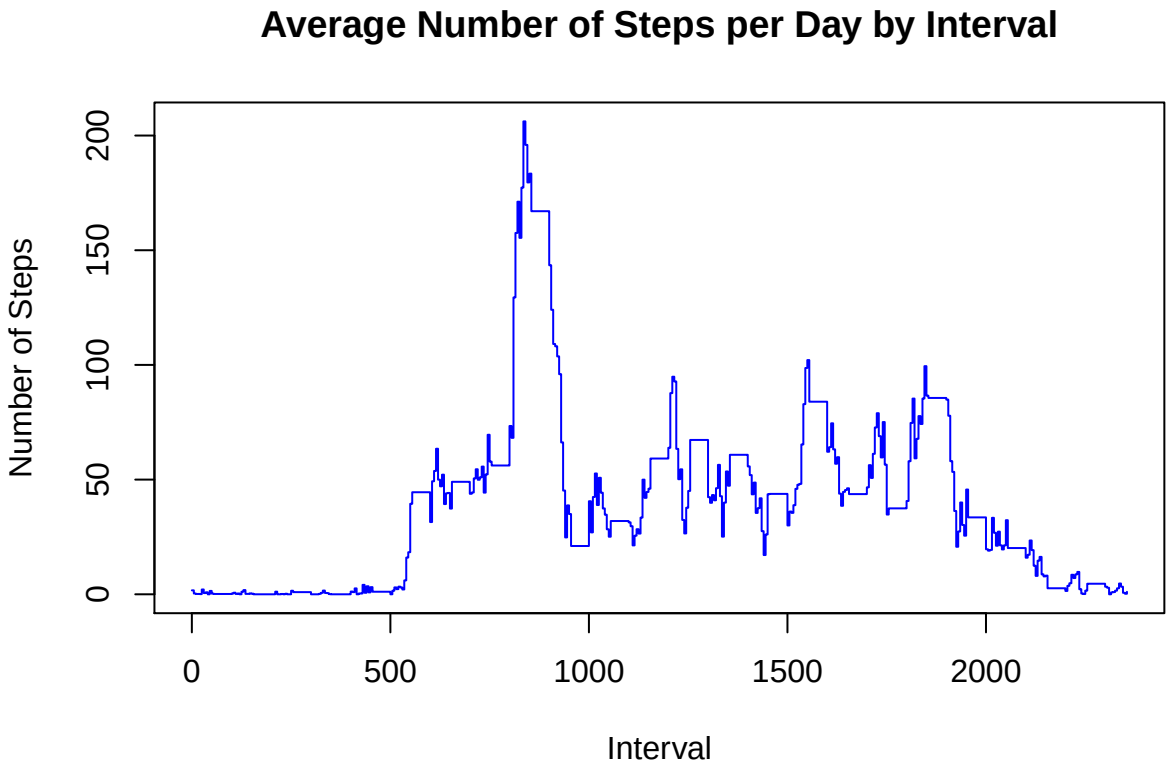
```
## [1] 10766.19
```

```
Meadian<- median(Steps.By.Day$steps, rm.na=T)
Meadian
```

```
## [1] 10765
```

Code for Time series plot of the average number of steps taken:

```r
Steps.By.Interval <- aggregate(steps~ interval, data, mean)
plot(Steps.By.Interval$interval,Steps.By.Interval$steps, type="s",col="blue",xlab="Interval",
ylab="Number of Steps",
main="Average Number of Steps per Day by Interval")
```



Code cacluates Maxmimum number of steps of 5-minute interval that on average:

```r
Max.Interval <- Steps.By.Interval[which.max(Steps.By.Interval$steps),1]
Max.Interval
```

```
## [1] 835
```

Code to describe and show a strategy for imputing missing data:

```r
#Calculate the total number of missing values
NA.Total <- sum(!complete.cases(data))
NA.Total
```

```
## [1] 2304
```

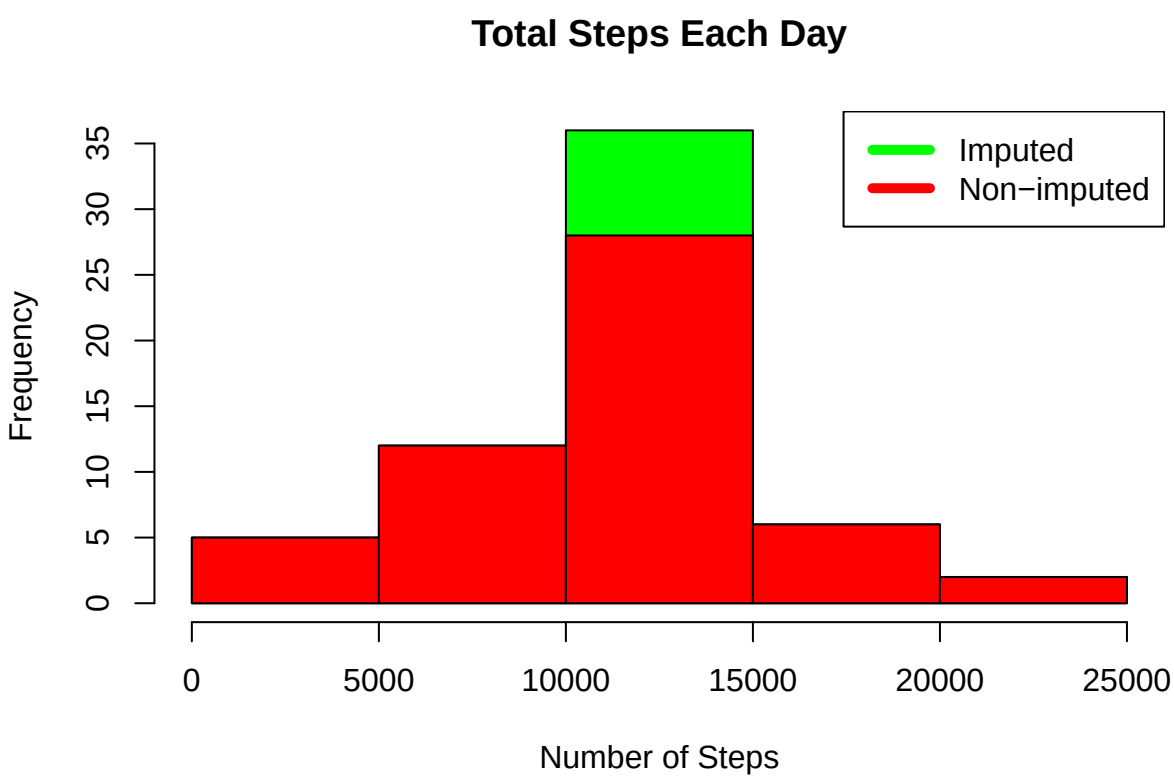
```
#compute missing values using mean
Steps.Average <- aggregate(steps ~ interval,
data = data, FUN = mean)
fill.NA <- numeric()
for (i in 1:nrow(data)) {obs <- data[i, ]
    if (is.na(obs$steps)) {
         steps <- subset(Steps.Average,
          interval == obs$interval)$steps
     } else {
         steps <- obs$steps
     }
     fill.NA <- c(fill.NA, steps)
}
```

```
 #Filling the missing values in new dataset
N.activity <- data
N.activity$steps <- fill.NA
```

Code make a Histogram of the total number of steps taken each day after missing values are imputed:

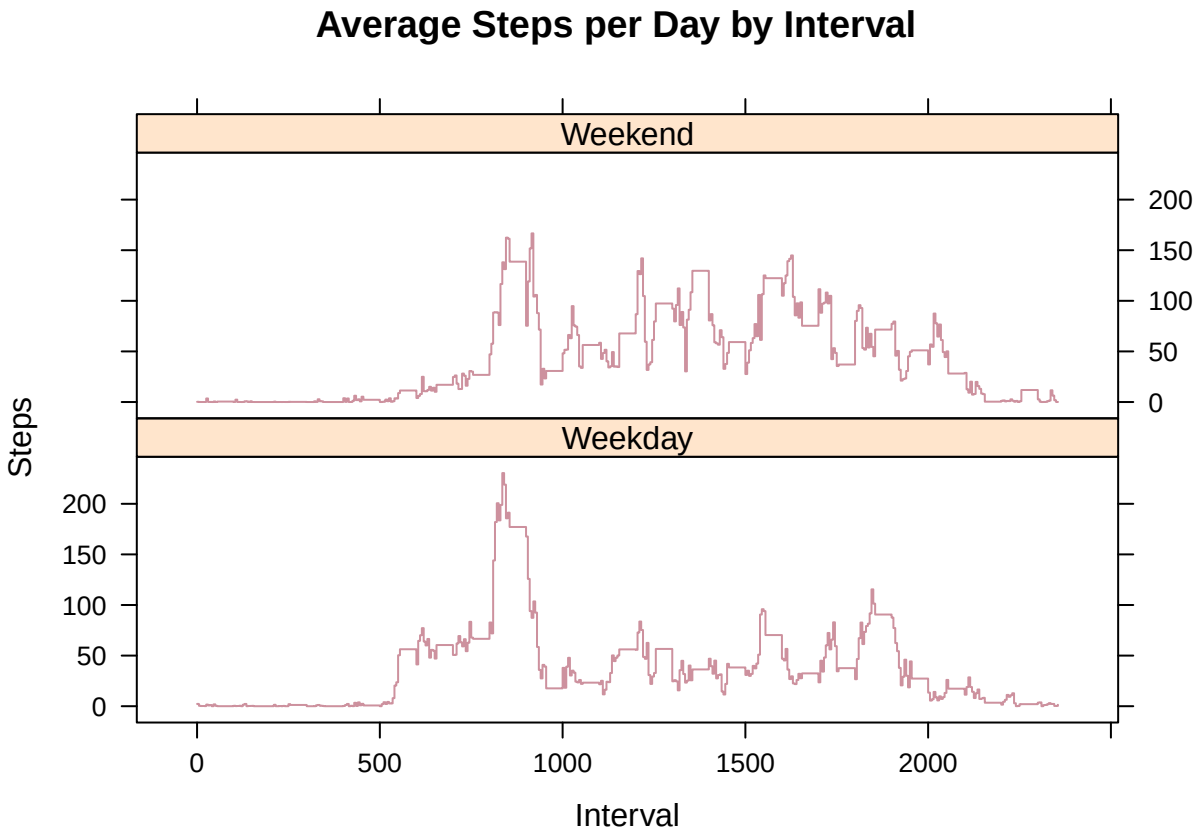
```
 #Filling the missing values in new dataset
N.activity <- data
N.activity$steps <- fill.NA
```

```
Steps.Total.Union <- aggregate(steps ~ date, data = N.activity, sum, na.rm = TRUE)
hist(Steps.Total.Union$steps, main = "Total Steps Each Day",
col="green",xlab="Number of Steps")
#Create Histogram to show difference.
hist(Steps.By.Day$steps, main = paste("Total Steps Each Day"),
col="red",xlab="Number of Steps", add=T)
legend("topright", c("Imputed", "Non-imputed"), col=c("green", "red"), lwd=5)
```



Code for comparing the average number of steps interval across weekdays and weekends illustrated by ploting:

```
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday","Friday")
N.activity$dow = as.factor(ifelse(is.element(weekdays(as.Date(N.activity$date)),
weekdays), "Weekday", "Weekend"))
Steps.Total.Union <- aggregate(steps ~ interval + dow, N.activity, mean)
library(lattice)
xyplot(Steps.Total.Union$steps ~ Steps.Total.Union$interval|Steps.Total.Union$dow, main="Average Steps
xlab="Interval", ylab="Steps",layout=c(1,2), type="s",col="pink3")
```



all the used code:

```
getwd()
data <- read.csv("C:/Users/saeed/Desktop/hti/programnig/R/ass.7/activity.csv")
head(data)
Steps.By.Day<-aggregate(steps ~ date, data, sum)
hist(Steps.By.Day$steps, main = "Total Steps By Day", col="red",xlab="Number of Steps")
```

```
Mean<- mean(Steps.By.Day$steps, rm.na=T)
Mean
Meadian<- median(Steps.By.Day$steps, rm.na=T)
Meadian
Steps.By.Interval <- aggregate(steps~ interval, data, mean)
plot(Steps.By.Interval$interval,Steps.By.Interval$steps, type="s",
col="blue", xlab="Interval", ylab="Number of Steps",main="Average Number of Steps per Day by Interval")
```

```
Max.Interval <- Steps.By.Interval[which.max(Steps.By.Interval$steps),1]
Max.Interval
NA.Total <- sum(!complete.cases(data))
NA.Total
Steps.Average <- aggregate(steps ~ interval, data = data, FUN = mean)
fill.NA <- numeric()
for (i in 1:nrow(data)) {obs <- data[i, ]
```

```
    if (is.na(obs$steps)) {
        steps <- subset(Steps.Average, interval == obs$interval)$steps
     } else {
        steps <- obs$steps
     }
     fill.NA <- c(fill.NA, steps)
}
N.activity <- data
N.activity$steps <- fill.NA
N.activity <- data
N.activity$steps <- fill.NA
Steps.Total.Union <- aggregate(steps ~ date, data = N.activity, sum, na.rm = TRUE)
hist(Steps.Total.Union$steps, main = "Total Steps Each Day", col="green", xlab="Number of Steps")
#Create Histogram to show difference.
hist(Steps.By.Day$steps, main = paste("Total Steps Each Day"), col="red", xlab="Number of Steps", add=T)
legend("topright", c("Imputed", "Non-imputed"), col=c("green", "red"), lwd=5)
```

```
Total.Mean <- mean(Steps.Total.Union$steps)
Total.Mean
Total.Median <- median(Steps.Total.Union$steps)
Total.Median
Mean.Diffrent <- Total.Mean - Mean
Mean.Diffrent
Median.Diffrent <- Total.Median - Meadian
Median.Diffrent
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday","Friday")
N.activity$dow = as.factor(ifelse(is.element(weekdays(as.Date(N.activity$date)),
weekdays), "Weekday", "Weekend"))
Steps.Total.Union <- aggregate(steps ~ interval + dow, N.activity, mean)
library(lattice)
xyplot(Steps.Total.Union$steps ~ Steps.Total.Union$interval|Steps.Total.Union$dow, main="Average Steps
```