

RepData Assignment 1

Mark Munsell

May 15, 2015

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 these data remain under-utilized both because the raw data are hard to obtain and there is a lack of statistical methods and software for processing and 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.

```
stepData <- read.csv("../activity.csv")
stepData$date <- as.Date(stepData$date, "%Y-%m-%d")

stepData <- data.frame(date=stepData$date,
                       weekday=toupper(weekdays(stepData$date)),
                       steps=stepData$steps,
                       interval=stepData$interval)

stepData <- cbind(stepData,
                  daytype=ifelse(stepData$weekday == "SATURDAY" |
                                stepData$weekday == "SUNDAY", "WEEKEND", "WEEKDAY"))

stepDataFinal <- data.frame(Date=stepData$date,
                            Weekday=stepData$weekday,
                            DayType=stepData$daytype,
                            Interval=stepData$interval,
                            Steps=stepData$steps)

head(stepDataFinal)
```

```
##           Date Weekday DayType Interval Steps
## 1 2012-10-01  MONDAY  WEEKDAY         0     NA
## 2 2012-10-01  MONDAY  WEEKDAY         5     NA
## 3 2012-10-01  MONDAY  WEEKDAY        10     NA
## 4 2012-10-01  MONDAY  WEEKDAY        15     NA
## 5 2012-10-01  MONDAY  WEEKDAY        20     NA
## 6 2012-10-01  MONDAY  WEEKDAY        25     NA
```

What is mean total number of steps taken per day?

```
stepDayTotal <- aggregate(stepDataFinal$Steps, by=list(stepDataFinal$Date), FUN=sum, na.rm=TRUE)

names(stepDayTotal) <- c("Date", "Total")

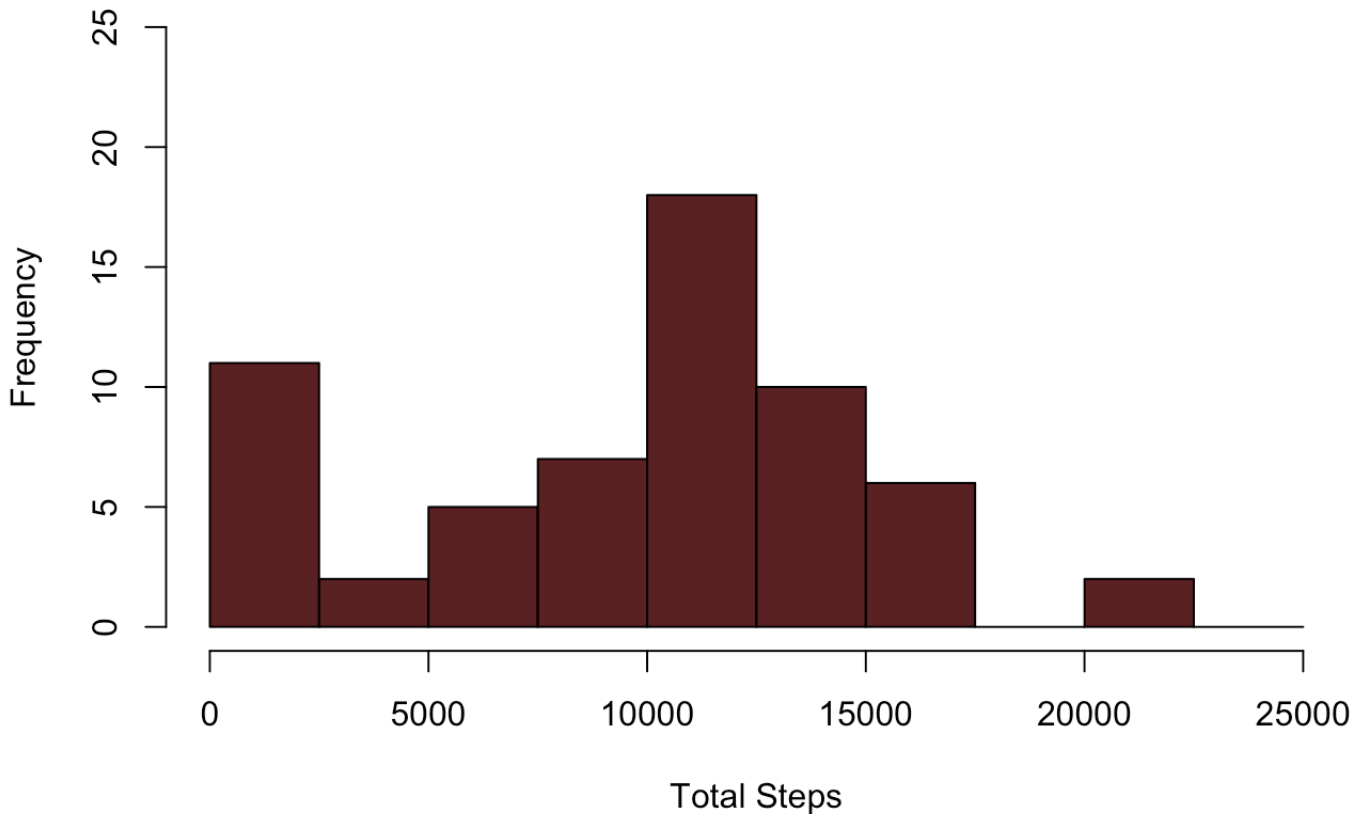
head(stepDayTotal)
```

```
##           Date Total
## 1 2012-10-01      0
## 2 2012-10-02    126
## 3 2012-10-03 11352
## 4 2012-10-04 12116
## 5 2012-10-05 13294
## 6 2012-10-06 15420
```

Histogram of the total number of steps per day:

```
hist(stepDayTotal$Total,
      breaks=seq(from=0, to=25000, by= 2500),
      col="#5a2121",
      xlab="Total Steps",
      ylim=c(0, 25),
      main="Total Steps Taken Each Day")
```

Total Steps Taken Each Day



Here is the calculation for mean and median of total number of steps taken per day.

```
stepMean <- mean(stepDayTotal$Total)

stepMed <- median(stepDayTotal$Total)
```

The mean of the total number of steps is 9354.2295082 and the median is 10395.

What is the average daily activity pattern?

Here is the code to create the data and the summary and first entries.

```

stepIntMean <- aggregate(stepDataFinal$Steps,
                          by=list(stepDataFinal$Interval),
                          FUN=mean,
                          na.rm=TRUE)

names(stepIntMean) <- c("Interval", "Mean")

head(stepIntMean)

```

```

##      Interval      Mean
## 1          0 1.7169811
## 2          5 0.3396226
## 3         10 0.1320755
## 4         15 0.1509434
## 5         20 0.0754717
## 6         25 2.0943396

```

```

summary(stepIntMean)

```

```

##      Interval      Mean
## Min.   : 0.0   Min.   : 0.000
## 1st Qu.: 588.8 1st Qu.: 2.486
## Median :1177.5 Median : 34.113
## Mean   :1177.5 Mean   : 37.383
## 3rd Qu.:1766.2 3rd Qu.: 52.835
## Max.   :2355.0 Max.   :206.170

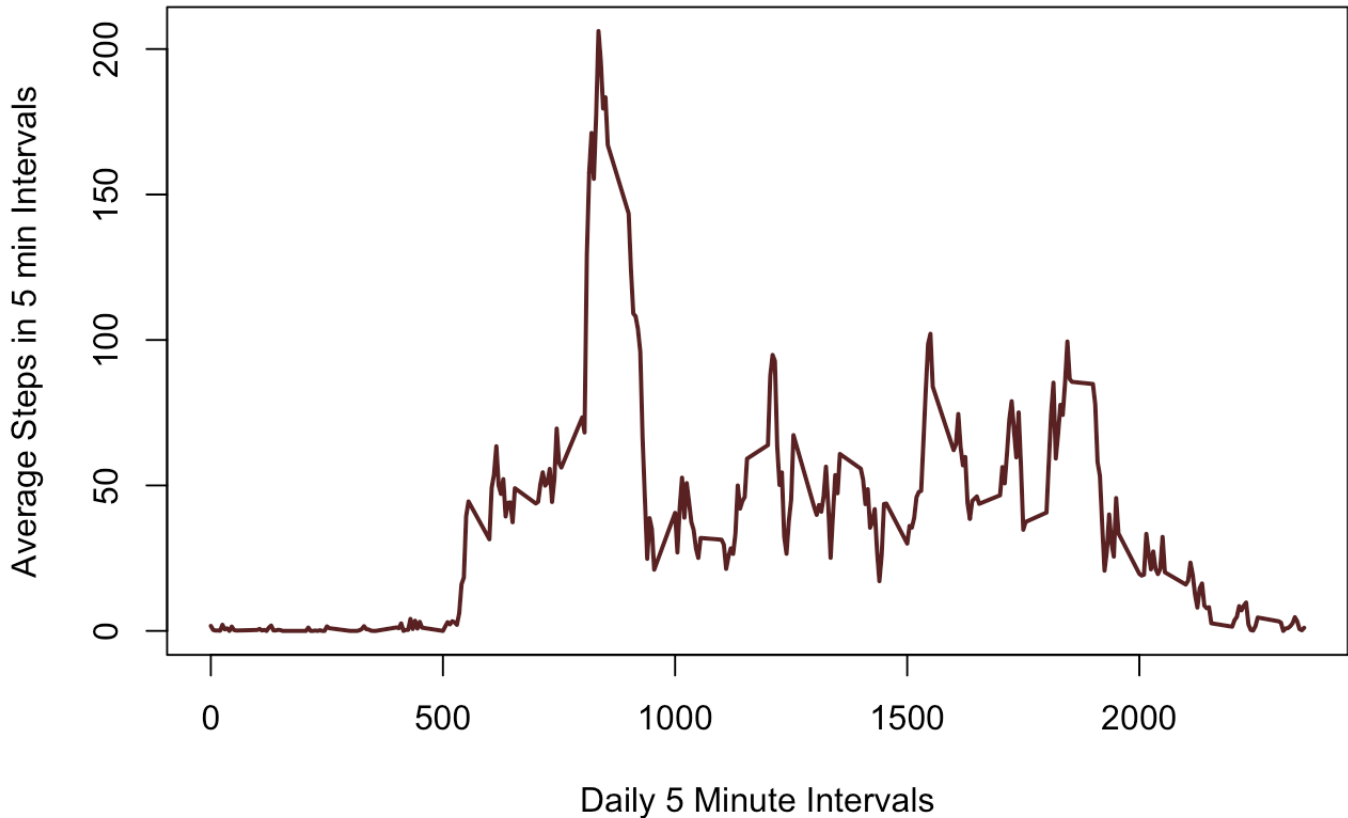
```

```

plot(stepIntMean$Interval,
     stepIntMean$Mean,
     type="l",
     col="#5a2121",
     lwd=2,
     xlab="Daily 5 Minute Intervals",
     ylab="Average Steps in 5 min Intervals",
     main="Average Number of Steps in 5 Minute Intervals per Day")

```

Average Number of Steps in 5 Minute Intervals per Day



Code to determine the interval with the maximum number of steps:

```
maxIntPos <- which(stepIntMean$Mean == max(stepIntMean$Mean))  
  
maxInt <- stepIntMean[maxIntPos,1]
```

The 5-minute interval, on average across all the days in the dataset, that contains the maximum number of steps is 835

Imputing the missing values

Note that there are a number of days/intervals where there are missing values (coded as NA). The presence of missing days may introduce bias into some calculations or summaries of the data.

Code to calculate the missing step data:

```
naCount <- sum(is.na(stepDataFinal$Steps))
```

The number of values coded with NA are: 2304

Here is the code to replace NA with the mean of steps:

```
naPos <- which(is.na(stepDataFinal$Steps))

meanRep <- rep(mean(stepDataFinal$Steps, na.rm=TRUE), times=length(naPos))
```

```
stepDataFinal[naPos, "Steps"] <- meanRep
```

Here are the first few rows of the new stepDataFinal data frame:

```
head(stepDataFinal)
```

```
##           Date Weekday DayType Interval  Steps
## 1 2012-10-01  MONDAY WEEKDAY         0 37.3826
## 2 2012-10-01  MONDAY WEEKDAY         5 37.3826
## 3 2012-10-01  MONDAY WEEKDAY        10 37.3826
## 4 2012-10-01  MONDAY WEEKDAY        15 37.3826
## 5 2012-10-01  MONDAY WEEKDAY        20 37.3826
## 6 2012-10-01  MONDAY WEEKDAY        25 37.3826
```

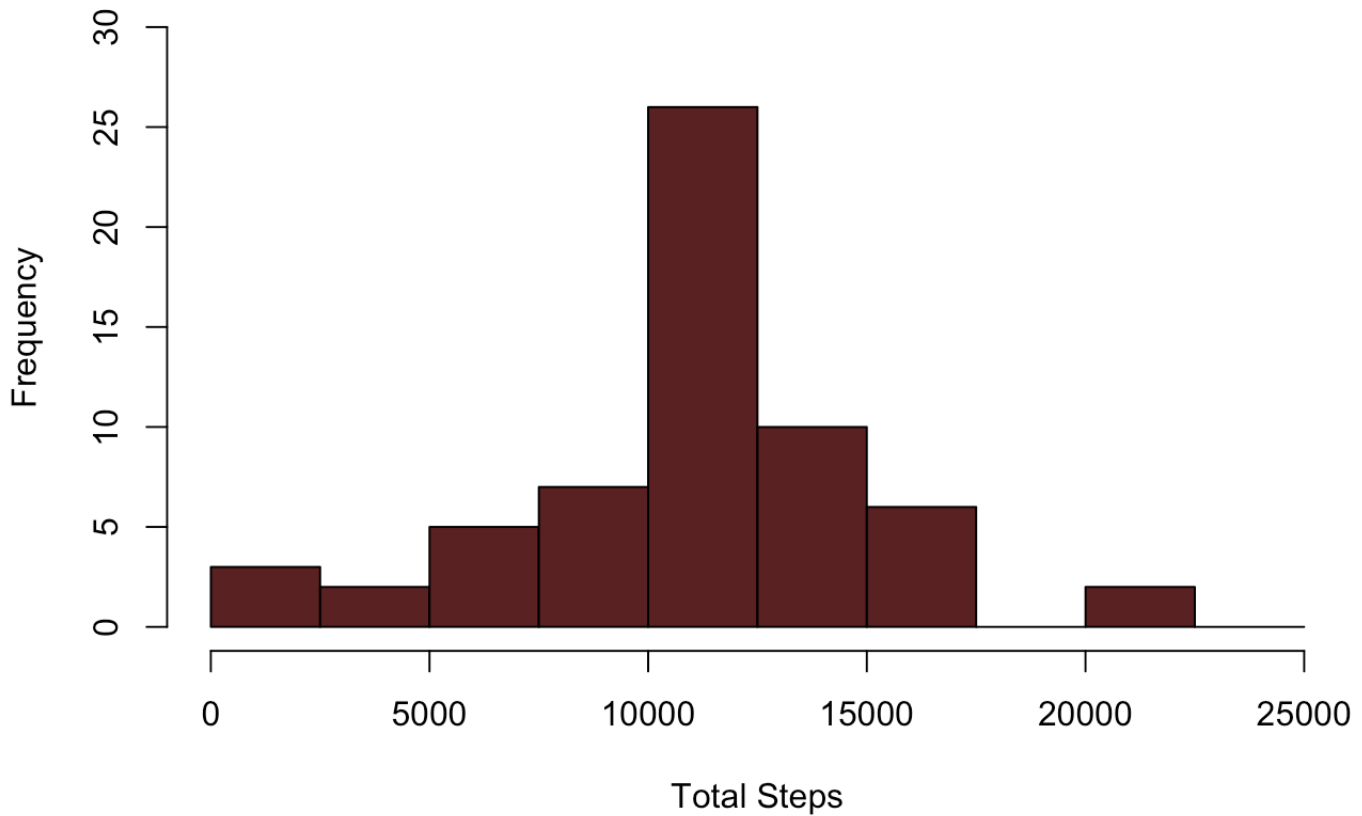
Histogram of the total number of steps per day with NA replaced with step mean value:

```
stepDayTotal <- aggregate(stepDataFinal$Steps, by=list(stepDataFinal$Date), FUN=sum)

names(stepDayTotal) <- c("Date", "Total")

hist(stepDayTotal$Total,
      breaks=seq(from=0, to=25000, by= 2500),
      col="#5a2121",
      xlab="Total Steps",
      ylim=c(0, 30),
      main="Total Steps Taken Each Day(NA replaced by step mean value)")
```

Total Steps Taken Each Day(NA replaced by step mean value)



Here is the new dataframe with new mean and median with NA replaced with steps mean.

```
head(stepDayTotal)
```

```
##           Date      Total
## 1 2012-10-01 10766.19
## 2 2012-10-02   126.00
## 3 2012-10-03 11352.00
## 4 2012-10-04 12116.00
## 5 2012-10-05 13294.00
## 6 2012-10-06 15420.00
```

```
mean(stepDayTotal$Total)
```

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

```
median(stepDayTotal$Total)
```

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

Are there differences in activity patterns between weekdays and weekends?

Here is the dataset with NA replaced by step mean and weekday and day type:

```
head(stepDataFinal)
```

```
##           Date Weekday DayType Interval  Steps
## 1 2012-10-01  MONDAY WEEKDAY         0 37.3826
## 2 2012-10-01  MONDAY WEEKDAY         5 37.3826
## 3 2012-10-01  MONDAY WEEKDAY        10 37.3826
## 4 2012-10-01  MONDAY WEEKDAY        15 37.3826
## 5 2012-10-01  MONDAY WEEKDAY        20 37.3826
## 6 2012-10-01  MONDAY WEEKDAY        25 37.3826
```

```
library(lattice)
```

```
mnData <- aggregate(stepDataFinal$Steps,
                     by=list(stepDataFinal$DayType,
                             stepDataFinal$Weekday,
                             stepDataFinal$Interval),mean)

names(mnData) <- c("DayType", "Weekday", "Interval", "Mean")
```

Here is the dataframe:

```
head(mnData)
```

```
##   DayType Weekday Interval    Mean
## 1 WEEKDAY  FRIDAY         0 8.307244
## 2 WEEKDAY  MONDAY         0 9.418355
## 3 WEEKEND SATURDAY        0 4.672825
## 4 WEEKEND  SUNDAY         0 4.672825
## 5 WEEKDAY THURSDAY        0 9.375844
## 6 WEEKDAY  TUESDAY        0 0.000000
```

Time Series Plot:


```
xyplot(Mean ~ Interval | DayType, mnData,
       type="l",
       lwd=1,
       xlab="Interval",
       ylab="Number of Steps",
       layout= c(1,2))
```

