

PA1_template

Tuesday, October 14, 2014

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.

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

#Read data into R

```
data <- read.csv("c:/rWork/activity.csv", header=T,  
                stringsAsFactors = FALSE)
```

#Install ggplot2

```
library("ggplot2", lib.loc="~/R/win-library/3.1")
```

#Order data by date

```
dataord <- data[order(data$date, decreasing=FALSE), ]
```

#Set X and Y axis values

```
a <- c(1, 10750, 21500)
```

```
b <- c("1-Oct-2012", "31-Oct-2102", "30-Nov-2012")
```

```
c <- c(0, 3.5, 7)
```

```
d <- c("0", "450", "900")
```

#Sum steps by day

```
stepsXdate <- aggregate(x=dataord$steps, by=list(dataord$date), FUN="sum")
```

#Order file stepsXdate by date

```
dataord2 <- stepsXdate[order(stepsXdate$x, decreasing=FALSE), ]
```

```

#Calculate the mean and median of the total number of steps taken per day
mn <- mean(stepsXdate$x, na.rm=T)
mn

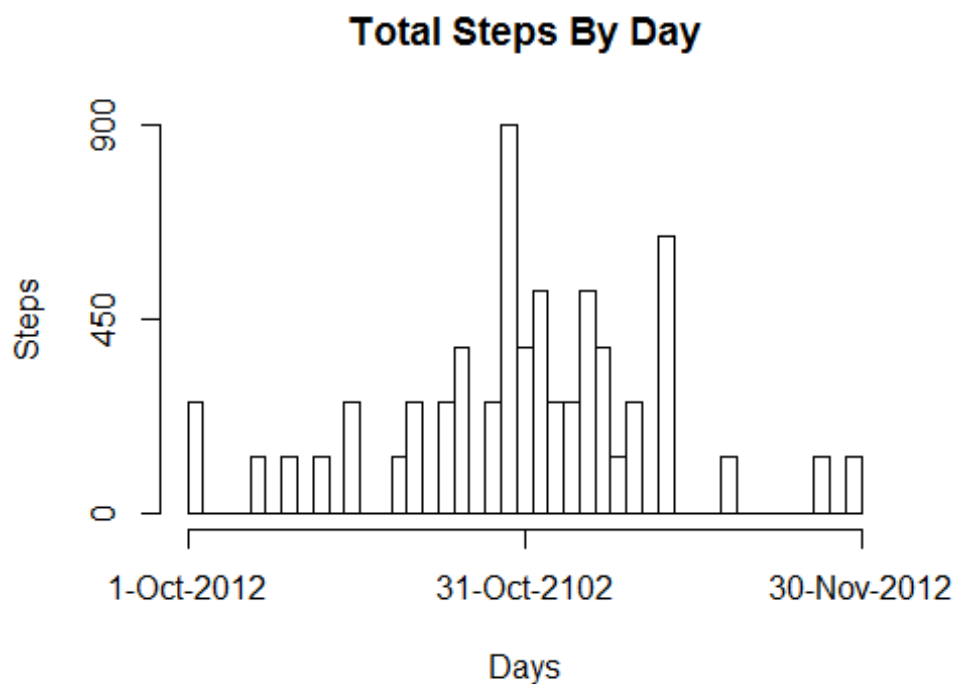
## [1] 10766

md <- median(stepsXdate$x, na.rm=T)
md

## [1] 10765

#Create a histogram of steps taken per day
hist(stepsXdate$x, main = "Total Steps By Day", xlab="Days", ylab="Steps",
      xaxt="n", yaxt="n", breaks=61)
axis(1, at=a, labels=b)
axis(2, at=c, labels=d)

```

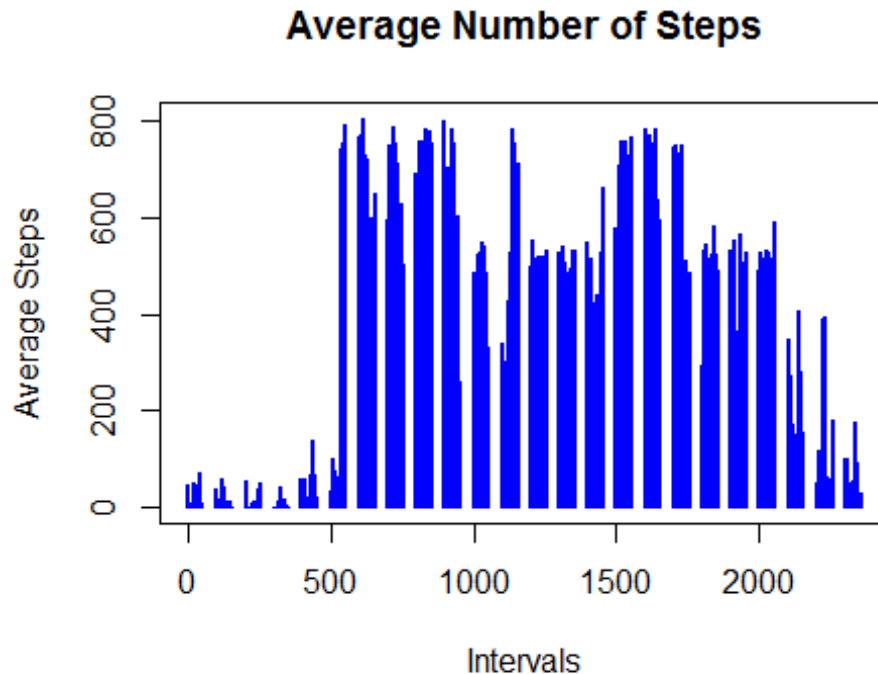


```

#Order data by interval
dataord3 <- data[order(data$interval, decreasing=FALSE), ]

```

```
plot(dataord3$interval, dataord3$steps, type="l", xlab= "Intervals",
      ylab= "Average Steps", col="blue" , lwd=2, main="Average Number of
Steps")
```



```
#Find and print the interval with the maximum number of steps
maxSteps <- max(dataord3$steps, na.rm=T)
maxout <- subset(dataord3, steps==maxSteps, select=c(steps, date, interval))
maxout

##      steps      date interval
## 16492    806 2012-11-27      615

#Calculate and print the total number of NA's
x <- is.na(dataord3$steps); y <- which(x); z <- length(y)
z

## [1] 2304

aggmean <- tapply(dataord3$steps, dataord3$interval, mean, na.rm=T)
taggmean <- mean(aggmean)
dataord3$steps[is.na(dataord3$steps)] <- taggmean
stepsXdateNA <- aggregate(x=dataord3$steps, by=list(dataord3$date), FUN="sum")
```

```

mnNA <- mean(stepsXdateNA$x, na.rm=T)
mnNA

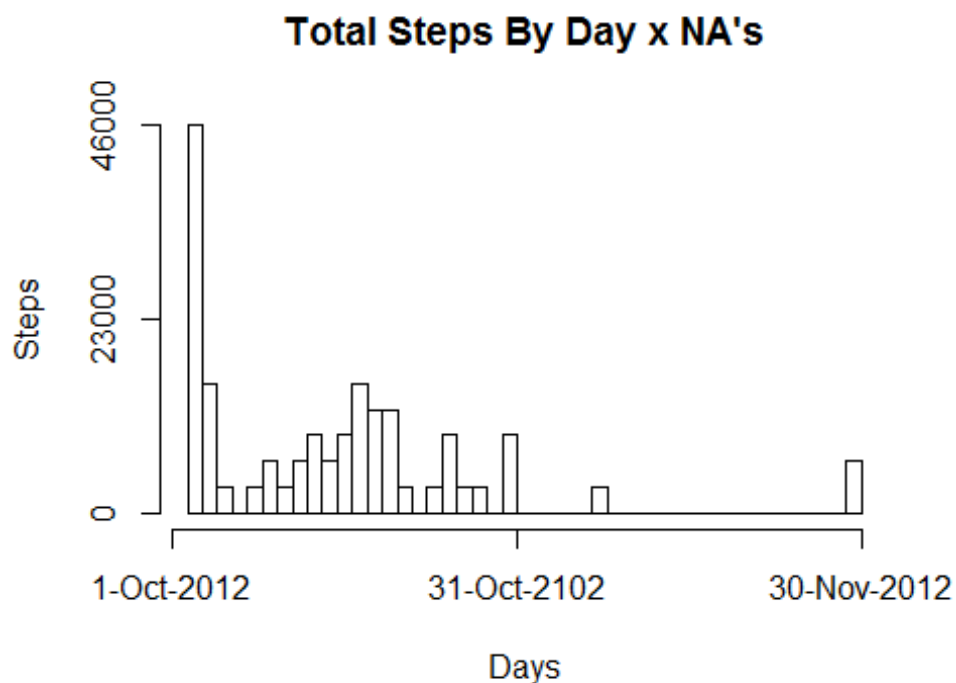
## [1] 10766

mdNA <- median(stepsXdateNA$x, na.rm=T)
mdNA

## [1] 10423

aa <- c(1, 23000, 46000)
bb <- c("1-Oct-2012", "31-Oct-2102", "30-Nov-2012")
cc <- c(0, 7.5, 15)
dd <- c("0", "23000", "46000")
hist(stepsXdateNA$x, main = "Total Steps By Day x NA's", xlab="Days",
      ylab="Steps",
      xaxt="n", yaxt="n", breaks=61)
axis(1, at=aa, labels=bb)
axis(2, at=cc, labels=dd)

```



```

dataord4 <- dataord3
dataord4$day <- "w"
dataord4$day <- ifelse(!weekdays(as.Date(dataord4$date)) %in% c("Saturday",
"Sunday"),
      "weekday", "weekend")

```

```
ggplot(dataord4,aes(interval,steps)) +geom_line() + facet_grid(.~day)
```

