

Activity monitoring Project

Summary

This project 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.

Load the data

```
# Load the data
activityData <- read.csv("activity.csv")
head(activityData)
```

```
##      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
```

What is mean total number of steps taken per day?

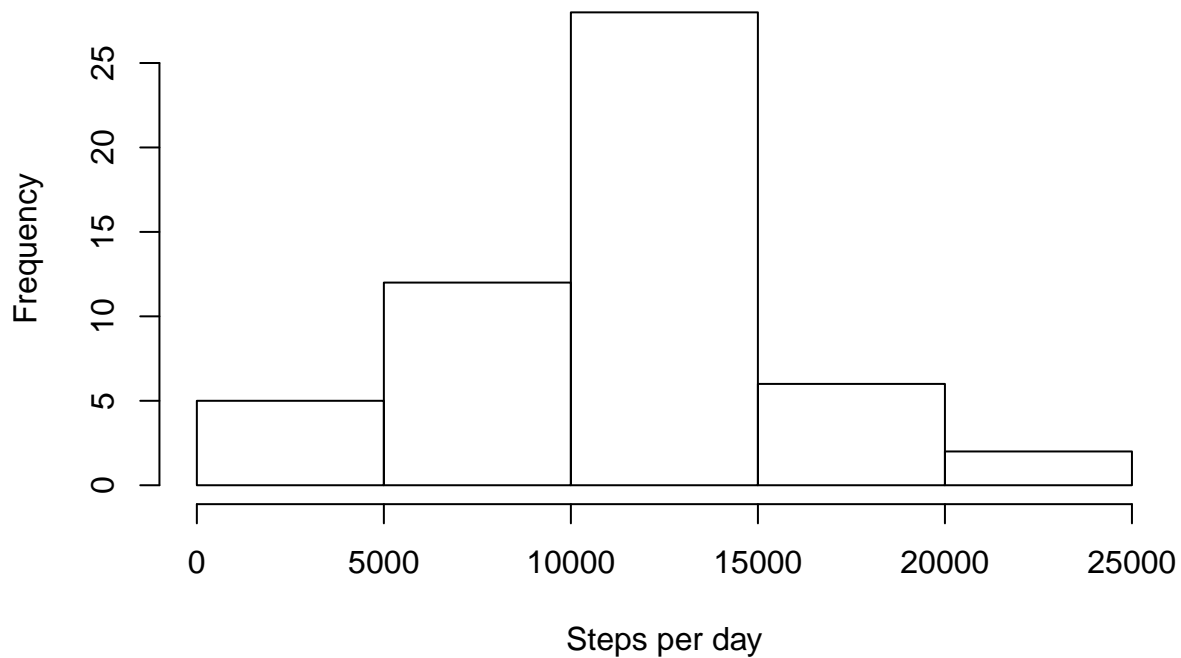
```
# calculate total steps per day
StepsPerDay<-tapply(activityData$steps,activityData$date,sum)

#convert stepsPerDay to a dataframe
StepsPerDay<-data.frame(StepsPerDay)
names(StepsPerDay)<-c("steps")
```

Make a histogram of the total number of steps taken each day

```
stepsFreq<-hist(StepsPerDay$steps,main = "Histogram of total number of steps per day", xlab = "Steps per day")
```

Histogram of total number of steps per day



Calculate mean and median of the total number of steps taken per day

```
summary(StepsPerDay)
```

```
##      steps
## Min.   : 41
## 1st Qu.: 8841
## Median :10765
## Mean   :10766
## 3rd Qu.:13294
## Max.   :21194
## NA's   :8
```

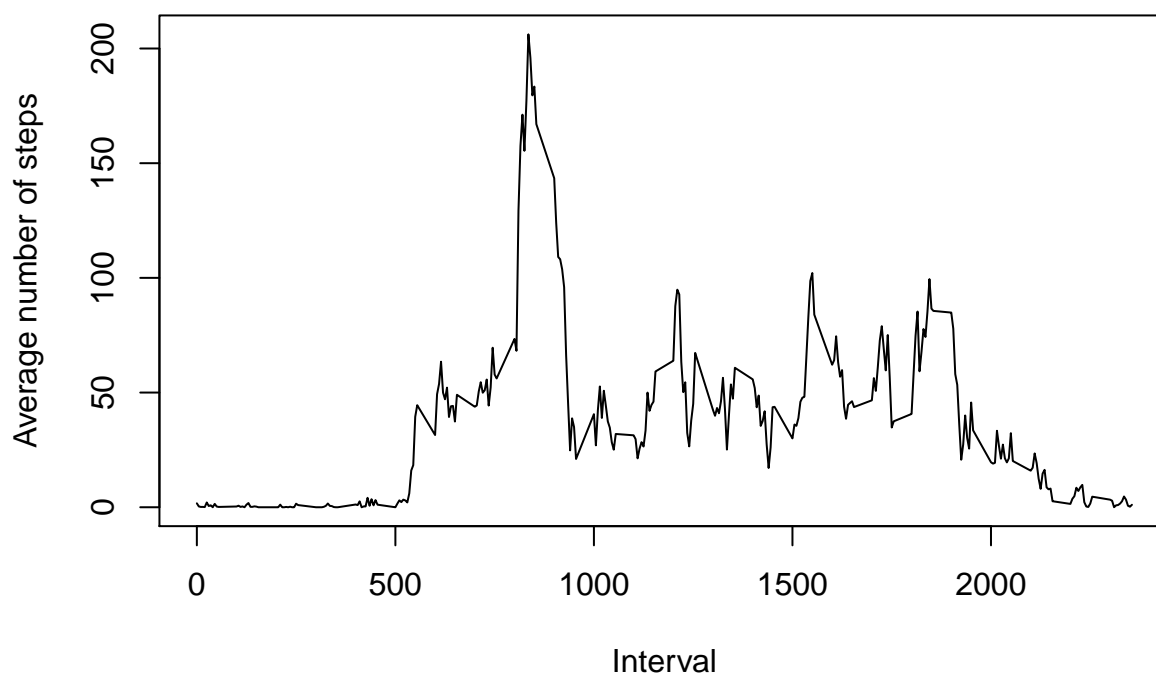
Median :10765, Mean :10766

What is the average daily activity pattern?

```
interval_steps <- aggregate(steps ~ interval, activityData, mean)

plot(interval_steps$interval, interval_steps$steps, type='l', col=1,
     main="Average number of steps averaged across all days", xlab="Interval",
     ylab="Average number of steps")
```

Average number of steps averaged across all days



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps

```
MaxNumOfSteps <- which.max(interval_steps$steps)

# get the interval with maximum average number of steps in an interval
interval_steps [MaxNumOfSteps, ]
```

```
##      interval      steps
## 104         835 206.1698
```

Imputing missing values

```
#Calculate the total number of missing values in the dataset
sum(!complete.cases(activityData))
```

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

replace na values with the mean for that 5-minute interval

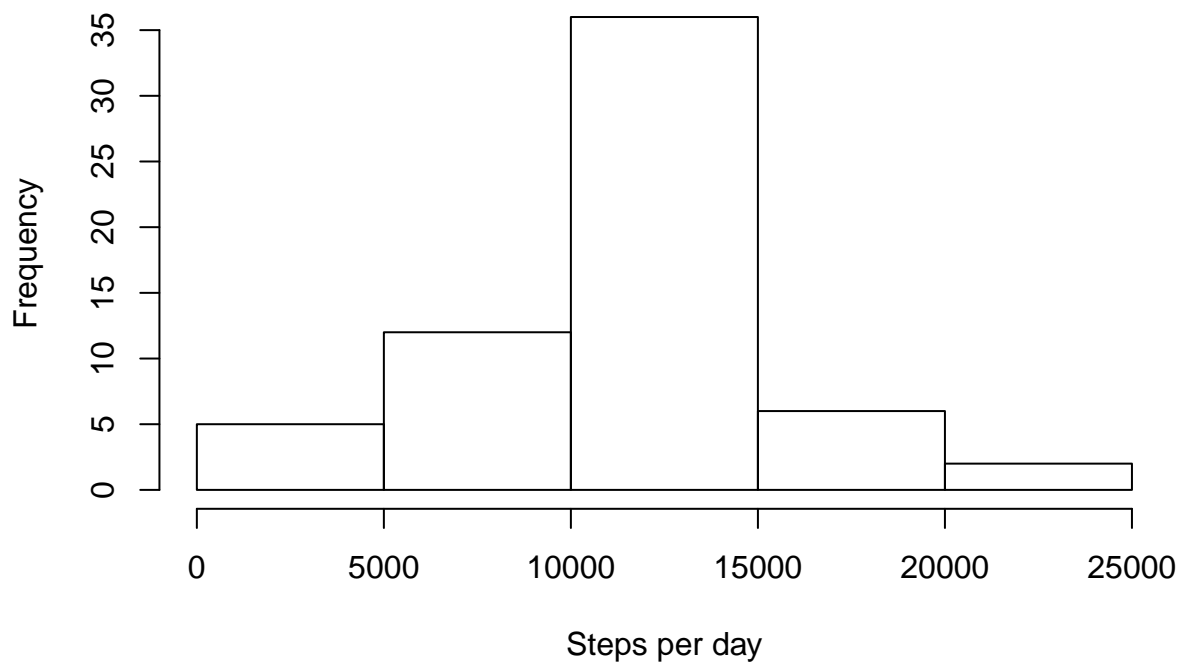
```

for (i in 1:nrow(activityData)) {
  if (is.na(activityData$steps[i])) {
    Corresponding_interval<-activityData$interval[i]
    activityData$steps[i]<-interval_steps$steps[interval_steps$interval==Corresponding_inte
  }
}
#Create a new dataset that is equal to the original dataset but with the missing data filled in.
new_activityData<-activityData

#Make a histogram of the total number of steps taken each day
new_StepsPerDay<-tapply(new_activityData$steps,new_activityData$date,sum)
new_StepsPerDay<-data.frame(new_StepsPerDay)
names(new_StepsPerDay)<-c("steps")
new_stepsFreq<-hist(new_StepsPerDay$steps,main = "Histogram of total number of steps per day", xlab = "Steps per day", ylab = "Frequency", col = "black", border = "black")

```

Histogram of total number of steps per day



```

# get the mean and meadian
summary(new_StepsPerDay)

```

```

##      steps
##  Min.   : 41
## 1st Qu.: 9819
## Median :10766
## Mean   :10766
## 3rd Qu.:12811
## Max.   :21194

```

there is a slight difference in median value. # Are there differences in activity patterns between weekdays and weekends?

```
# Create a new factor variable in the dataset with two levels - "weekday" and "weekend" indicating whether
day <- weekdays(as.Date(new_activityData$date))
day_type <- vector()
for (i in 1:nrow(new_activityData)) {
  if (day[i] == "Saturday") {
    day_type[i] <- "Weekend"
  } else if (day[i] == "Sunday") {
    day_type[i] <- "Weekend"
  } else {
    day_type[i] <- "normalday"
  }
}
new_activityData$day_type <- day_type

# convert day_type from character to factor
new_activityData$day_type <- as.factor(new_activityData$day_type)

# Make a panel plot containing a time series plot of the 5-minute interval (x-axis) and the average number of steps
new_interval_steps <- aggregate(steps ~ interval + day_type, new_activityData, mean)
library(lattice)

# create the panel plot
xyplot(steps ~ interval | day_type, new_interval_steps, type = "l", layout = c(1, 2),
       xlab = "interval", ylab = "Number of steps")
```

