Reproducible Research Week 2 Course Projec 1

Johnnery Aldana

4/11/2019

1 Loading and preprocessing the data

```
activity<-read.csv("./activity.csv",header = TRUE)
activity$date<-as.Date(activity$date)
```
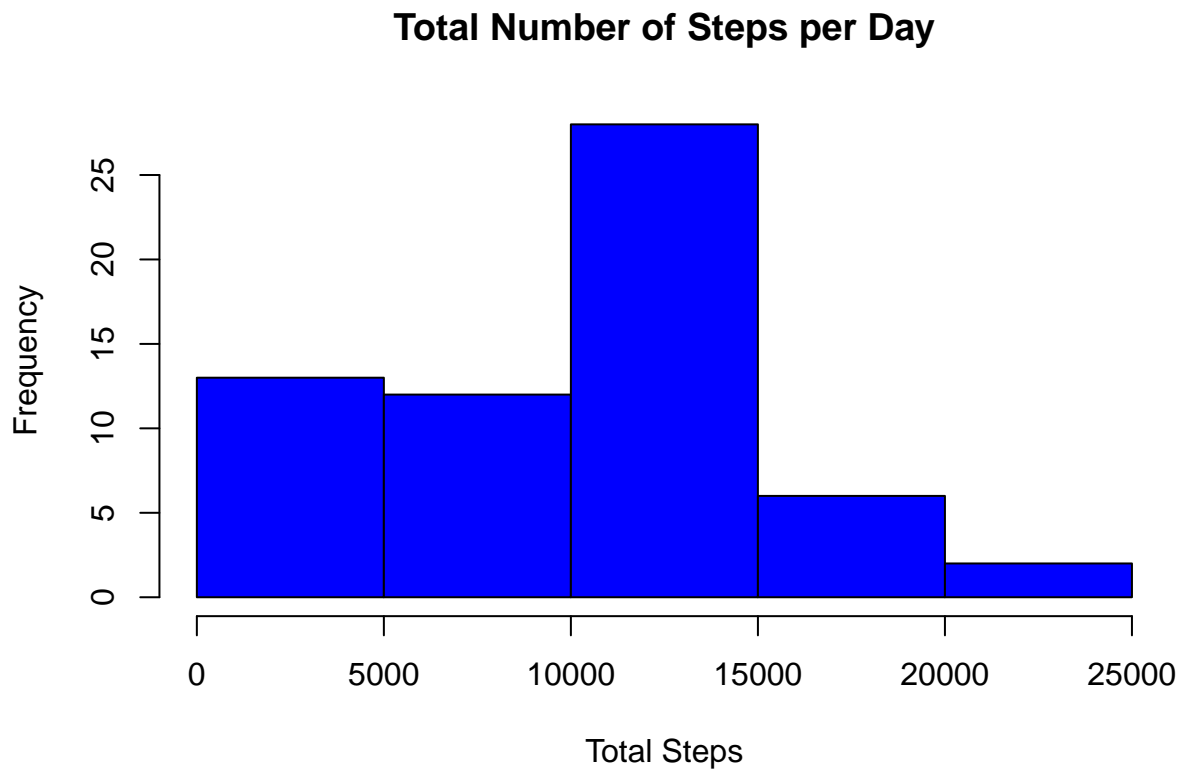
2 What is mean total number of steps taken per day?

2.1 Calculate the total number of steps taken per day

```
stepsday<-with(activity,tapply(steps,date,sum,na.rm=TRUE))
```

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

```
hist(stepsday,col = "blue",xlab = "Total Steps",ylab = "Frequency", main = "Total Number of Steps per Da
```

Total Number of Steps per Day

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

```r
mean(stepsday)
```
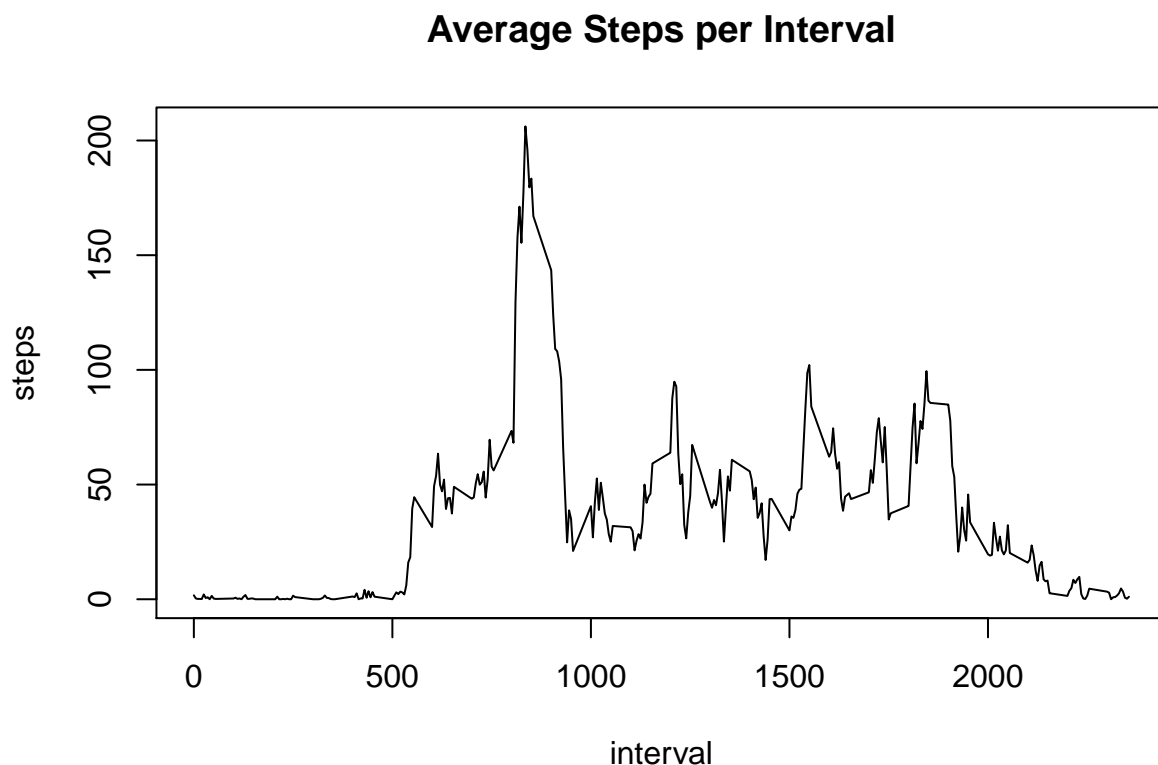
```
## [1] 9354.23
```

```r
median(stepsday)
```

```
## [1] 10395
```

3 What is the average daily activity pattern?

3.1 What is the average daily activity pattern?

```r
averageStepsbyInterval<-aggregate(steps~interval, activity, mean)
with(averageStepsbyInterval, plot(interval, steps, type = "l", main = "Average Steps per Interval"))
```

Average Steps per Interval

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

```
averageStepsbyInterval[which.max(averageStepsbyInterval[,2]),1]
```

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

4 Imputing missing values

4.1 Calculate and report the total number of missing values in the dataset

```
sum(is.na(activity$steps))
```

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

4.2 Devise a strategy for filling in all of the missing values in the dataset.

I decided to fill in all of the missing values in the dataset by the mean number of steps per interval.

```
m<-mean(averageStepsbyInterval$steps)
```

4.3 Create a new dataset that is equal to the original dataset but with the missing data filled in.

```
activity1<-activity
missingIndex<-is.na(activity[,1])
activity1[missingIndex,1]<-m
```

4.4 Make a histogram of the total number of steps taken each day and Calculate and report the mean and median total number of steps taken per day.

```
totalStepsByDay1<-aggregate(steps~date, activity1, sum)
hist(totalStepsByDay1$steps, xlab="Class of Total Number of Steps per day", ylab="Number of Days", main=
```

Number of Steps taken each day after missing values are imputed
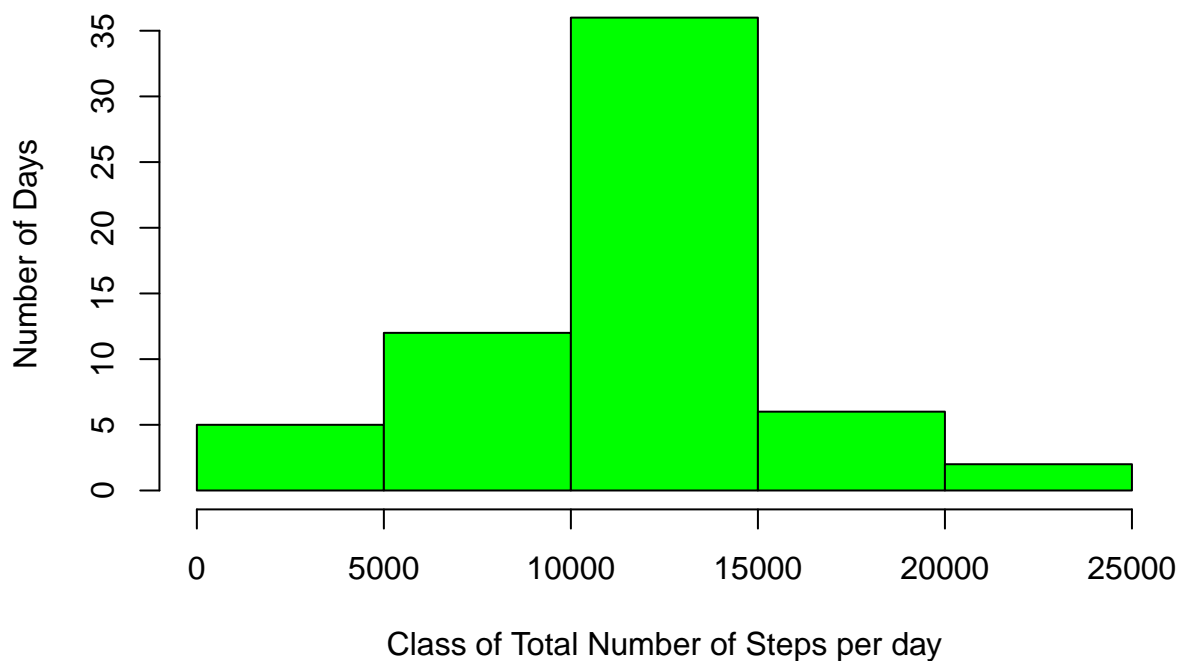

```
totalStepsByDay1<-aggregate(steps~date, activity1, sum)
mean(totalStepsByDay1$steps)
```

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

```
median(totalStepsByDay1$steps)
```

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

5 Are there differences in activity patterns between weekdays and weekends?

5.1 Create a new factor variable in the dataset with two levels – "weekday" and "weekend" indicating whether a given date is a weekday or weekend day.

```
activity2<-activity

activity2$date<-as.Date(activity2$date)

activity2$date <- weekdays(activity2$date)

daytype <- function (x) {

        if (x == "sábado" | x == "domingo"){
          x <- "Weekend"

        } else {
          x <- "Weekday"}

    }

activity2$daytype <- sapply(activity2$date, daytype)

activity_by_date <- aggregate(steps~interval + daytype, activity2, mean, na.rm = TRUE)
```

5.2 Make a panel plot containing a time series plot (i.e. type="l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all weekday days or weekend days (y-axis)

```
library (ggplot2)

plot<- ggplot(activity_by_date, aes(x = interval , y = steps, color = daytype)) + geom_line() + labs(ti

print(plot)
```

Average daily steps by type of date