Reroducible Research - Course Project 1

Rassul Kalfane

29/05/2020

Loading and preprocessing the dataLoading and preprocessing the data

First, we load and see some summary statistics of the data:

```
#Code for reading in the dataset and/or processing the data

unzip("repdata_data_activity.zip")
data <- read.csv("activity.csv")
summary(data)
```

```
##      steps                date          interval     
##  Min.   :  0.00   2012-10-01:  288   Min.   :   0.0  
##  1st Qu.:  0.00   2012-10-02:  288   1st Qu.: 588.8  
##  Median :  0.00   2012-10-03:  288   Median :1177.5  
##  Mean   : 37.38   2012-10-04:  288   Mean   :1177.5  
##  3rd Qu.: 12.00   2012-10-05:  288   3rd Qu.:1766.2  
##  Max.   :806.00   2012-10-06:  288   Max.   :2355.0  
##  NA's   :2304     (Other)   :15840
```

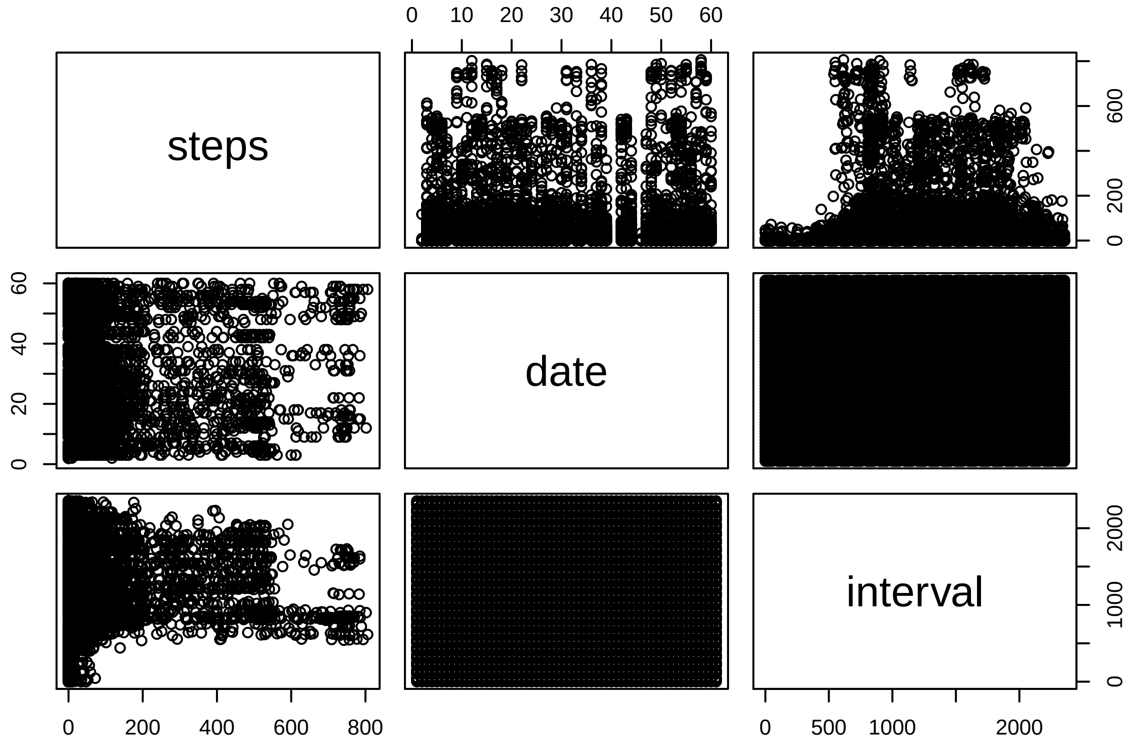
```
names(data)
```

```
## [1] "steps"    "date"     "interval"
```

```
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
```

```
pairs(data)
```



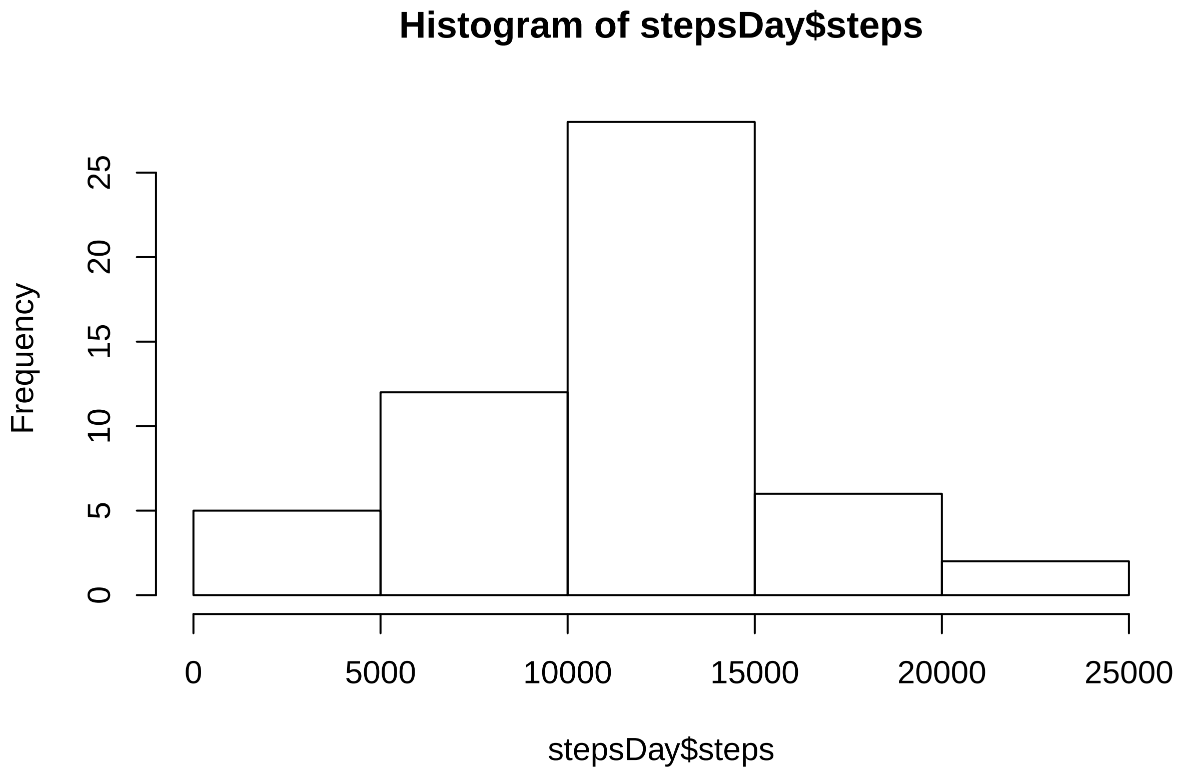
What is mean total number of steps taken per day?

1- Calculate the total number of steps taken per day

```
stepsDay<-aggregate(steps ~ date, data, sum, na.rm=TRUE)
```

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

```
hist(stepsDay$steps)
```



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

```
#Mean and median number of steps taken each day

meanStepsDay<-mean(stepsDay$steps)
meanStepsDay
```

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

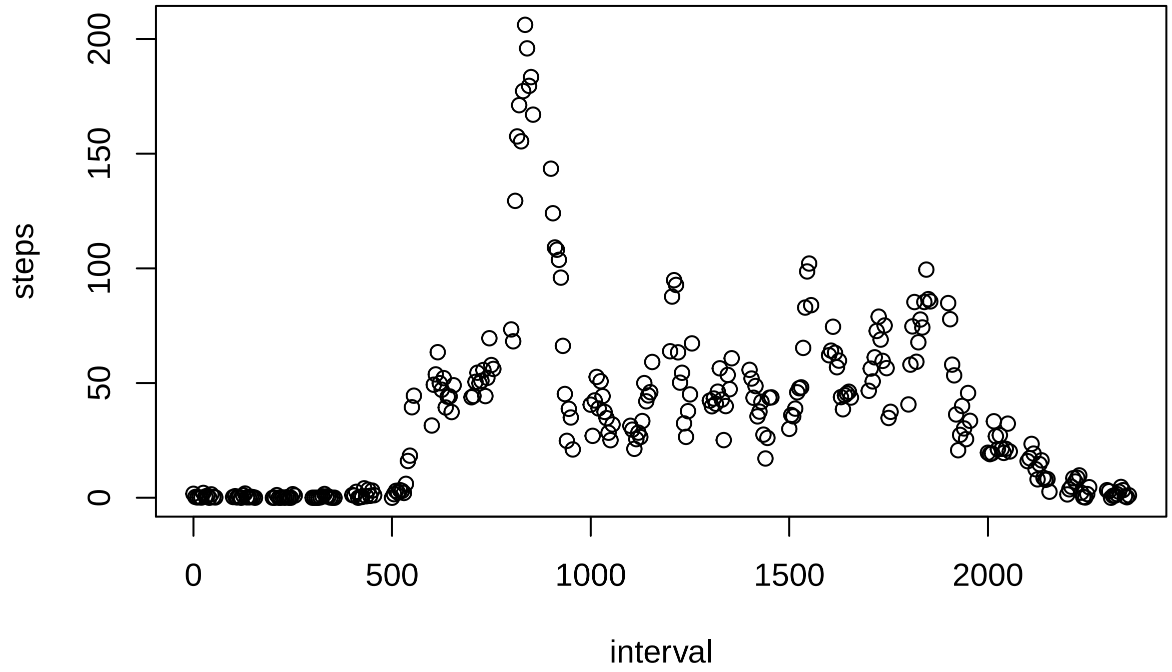
```
medianStepsDay<-median(stepsDay$steps)
medianStepsDay
```

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

What is the average daily activity pattern?

1- Make a time series plot

```
stepsInterval<-aggregate(steps~interval, data=data, mean, na.rm=TRUE)
plot(stepsInterval)
```



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

```
maxInterval<-stepsInterval[which.max(stepsInterval$steps),]$interval
maxInterval
```

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

Imputing missing values

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

1- Calculate and report the total number of missing values in the dataset

numberNA<-sum(is.na(data$steps)) numberNA

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

The strategy does not need to be sophisticated. For example, you could use the mean/median for that day, or the mean for that 5-minute interval, etc.

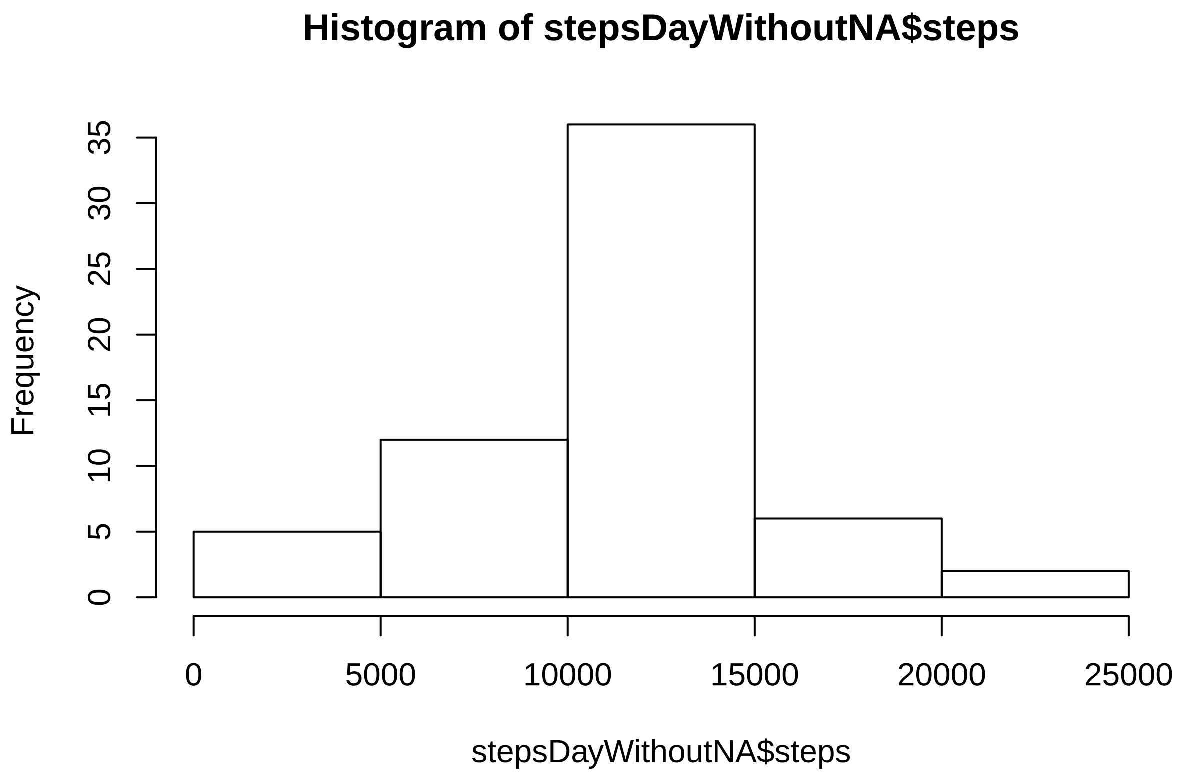
```
intervalMean<-function(interval){
  stepsInterval[stepsInterval$interval==interval,]$steps
}
```

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

```
dataWithoutNA<-data
for(i in 1:nrow(dataWithoutNA)){
  if(is.na(dataWithoutNA[i,]$steps)){
    dataWithoutNA[i,]$steps <- intervalMean(dataWithoutNA[i,]$interval)
  }
}
```

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.

```
stepsDayWithoutNA <- aggregate(steps ~ date, data=dataWithoutNA, sum)
hist(stepsDayWithoutNA$steps)
```



Are there differences in activity patterns between weekdays and weekends?

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.

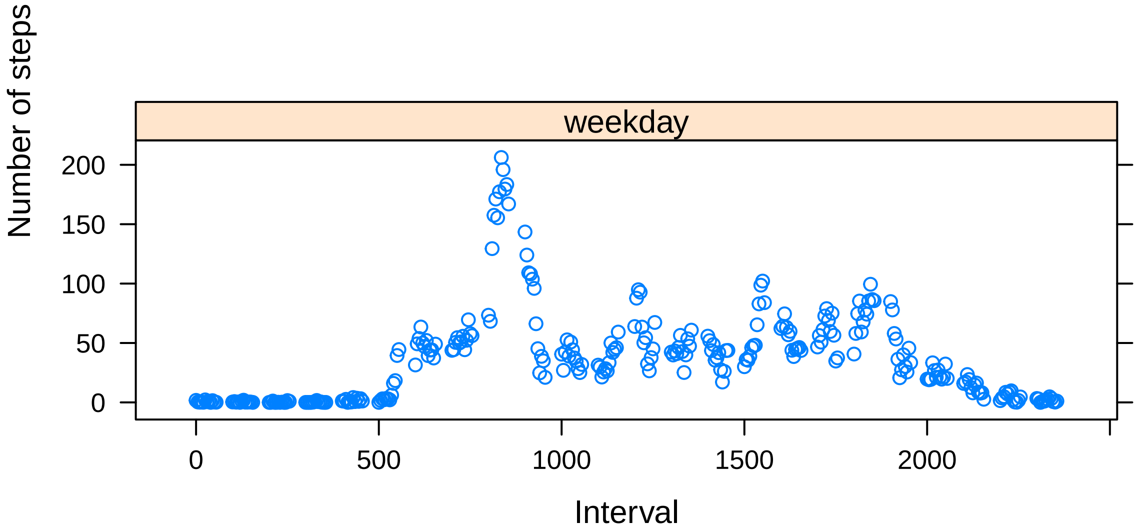
```
#Create a new factor variable in the dataset with two levels - "weekday" and "weekend" indicating wheth

dataWithoutNA$date <- as.Date(strptime(dataWithoutNA$date, format="%Y-%m-%d"))
dataWithoutNA$day <- weekdays(dataWithoutNA$date)
for (i in 1:nrow(dataWithoutNA)) {
  if (dataWithoutNA[i,]$day %in% c("Saturday","Sunday")) {
    dataWithoutNA[i,]$day<-"weekend"
  }
  else{
    dataWithoutNA[i,]$day<-"weekday"
  }
}
stepsByDay <- aggregate(dataWithoutNA$steps ~ dataWithoutNA$interval + dataWithoutNA$day, dataWithoutNA
```

2- Make a panel plot containing a time series plot

```
#Make a panel plot containing a time series plot (i.e. type = "l"\color{red}{\verb|type = "l"|}type = "

names(stepsByDay) <- c("interval", "day", "steps")
library(lattice)
xyplot(steps ~ interval | day, stepsByDay, layout = c(1, 2),
       xlab = "Interval", ylab = "Number of steps")
```


weekday
Number of steps
0
50
100
150
200
0
500
1000
1500
2000
Interval