Reproducible Research: Peer Assessment 1

## Loading and proccesing the data

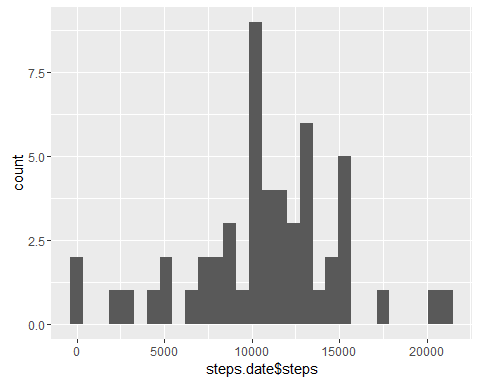
data<-read.csv("activity.csv")

## What is mean total number of steps taken per day?

### 1:Histogram of the total number of steps taken each day

library(ggplot2)  
steps.date<-aggregate(steps~date, data=data, FUN = sum)  
p<-qplot(steps.date$steps, geom="histogram")  
p

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



### 2: Mean and median total number of steps taken per day

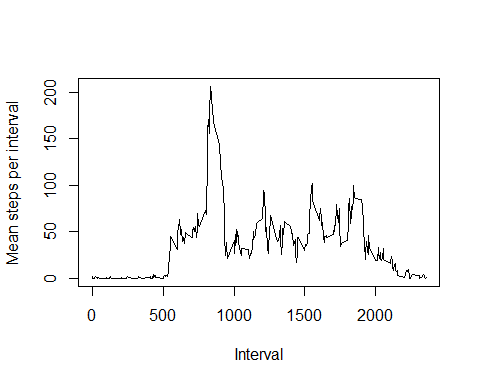
mean.steps<-mean(steps.date$steps)  
median.steps<-median(steps.date$steps)

The mean: 1.076618910^{4}  
The median: 10765

## What is the average daily activity pattern?

1. A plot of the average steps per interval:

steps.interval<-aggregate(steps~interval, data=data, FUN = mean)  
plot(y=steps.interval$steps, x=steps.interval$interval, type = "l",   
 ylab = "Mean steps per interval", xlab = "Interval")



1. Calculate the interval with the highest average steps:

max.steps<-steps.interval[which.max(steps.interval$steps),1]

The interval with the highest average steps is 835

## Imputing missing values

1. The number of NA rows:

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

The number of rows without data is: 2304

1. I will fill the missing data with the average steps per interval:

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

1. The new dataset is named: "data.fill"
2. 4.1:Histogram

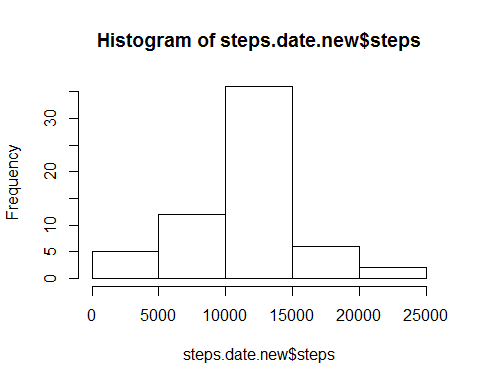
steps.date.new<-aggregate(steps~date, data=data.fill, FUN = sum)  
p<-hist(steps.date.new$steps, geom="histogram")

## Warning in plot.window(xlim, ylim, "", ...): "geom" is not a graphical  
## parameter

## Warning in title(main = main, sub = sub, xlab = xlab, ylab = ylab, ...):  
## "geom" is not a graphical parameter

## Warning in axis(1, ...): "geom" is not a graphical parameter

## Warning in axis(2, ...): "geom" is not a graphical parameter



4.2: mean and median:

mean.steps.new<-mean(steps.date.new$steps)  
median.steps.new<-median(steps.date.new$steps)

The new mean is 1.076618910^{4} The new median is 1.076618910^{4}

Their is no big difference in the mean an in the median, yet there is a diffrence in the histogram, the results make sence becaouse of the way the mean and median is calculate.

## Differences in activity patterns between weekdays and weekends

1: Adding to the date if the activity where in weekdays or in weekend:

weekdays<-weekdays(as.Date((data.fill$date)))  
sabat<-weekdays=="שבת"  
sunday<-weekdays==names(table(weekdays))[2]  
weekend\_day<-sabat+sunday  
weekend<-vector(length = nrow(data.fill))  
weekend[weekend\_day==1]<-"Weekend"  
weekend[weekend\_day==0]<-"Weekday"  
data.fill$weekend<-weekend

2: A plot by Weekend/Weekday

steps.day <- aggregate(steps ~ interval + weekend, data = data.fill, mean)  
library("lattice")  
xyplot(steps ~ interval | weekend, steps.day, type = "l", layout = c(1, 2),   
 xlab = "Interval", ylab = "Number of steps")

