Reproducible Research: Peer Assessment 1

library (dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.3

library(gridExtra)

## Warning: package 'gridExtra' was built under R version 3.4.3

##   
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':  
##   
## combine

library(lubridate)

##   
## Attaching package: 'lubridate'

## The following object is masked from 'package:base':  
##   
## date

## Loading and preprocessing the data

#list what is in the zip file  
unzip("activity.zip", list=TRUE)

## Name Length Date  
## 1 activity.csv 350829 2014-02-11 10:08:00

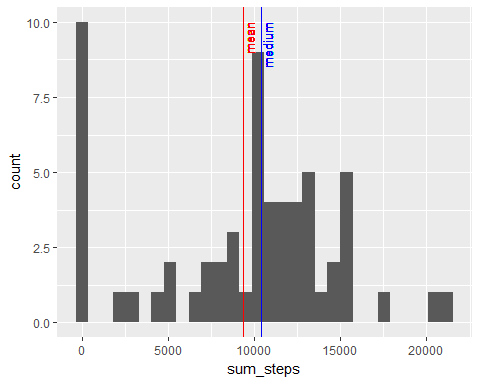
#read the activity.csv  
input\_data <- read.csv(unzip(zipfile="activity.zip"))

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

1. Calculate the total number of steps taken per day
2. If you do not understand the difference between a histogram and a barplot, research the difference between them. Make a histogram of the total number of steps taken each day
3. Calculate and report the mean and median of the total number of steps taken per day

Total\_Daily\_Steps <- input\_data %>%  
 group\_by(date) %>%  
 summarize(sum\_steps = sum(steps, na.rm=TRUE))  
  
mean\_daily = round(mean(Total\_Daily\_Steps$sum\_steps),0)  
median\_daily = round(median(Total\_Daily\_Steps$sum\_steps),0)  
  
#qplot(data=Total\_Daily\_Steps, x=sum\_steps)  
  
ggplot(Total\_Daily\_Steps,aes(sum\_steps)) +   
 geom\_histogram() +   
 geom\_vline(xintercept = mean\_daily, color="red") +   
 geom\_text(aes(x=mean\_daily, y=10), label='mean', color="red", size=3.5, angle=90, hjust=1, vjust=1) +   
 geom\_vline(xintercept = median\_daily, color="blue") +   
 geom\_text(aes(x=median\_daily, y=10),label='medium', color="blue",size=3.5,angle=90, hjust=1, vjust=1)

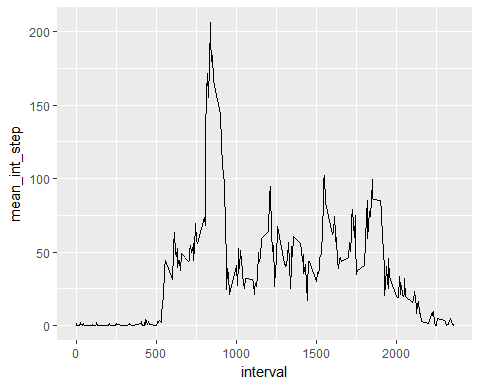
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

 The mean number of steps taken per day is 9354 The median number of steps taken per day is 1.039510^{4}

## What is the average daily activity pattern?

1. Make 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 days (y-axis)
2. Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

Total\_Daily\_Steps$date <- ymd(Total\_Daily\_Steps$date)  
  
  
interval\_mean <- input\_data %>%  
 group\_by(interval)%>%  
 summarize(mean\_int\_step = mean(steps, na.rm=TRUE))  
  
ggplot(interval\_mean, aes(interval, mean\_int\_step)) + geom\_line()



max\_step = interval\_mean[which.max(interval\_mean$mean\_int\_step),'interval']

At interval 835, on average across all the days in the dataset, contains the maximum number of steps.

## Imputing missing values

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