Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

setwd("~/Desktop/github/RepData\_PeerAssessment1")  
activity<- read.csv("activity.csv")  
head(activity)

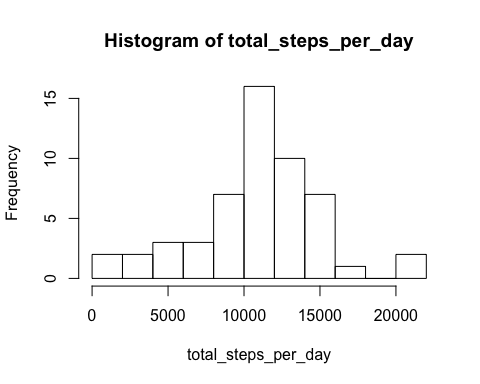
## steps date interval  
## 1 NA 2012-10-01 0  
## 2 NA 2012-10-01 5  
## 3 NA 2012-10-01 10  
## 4 NA 2012-10-01 15  
## 5 NA 2012-10-01 20  
## 6 NA 2012-10-01 25

str(activity)

## 'data.frame': 17568 obs. of 3 variables:  
## $ steps : int NA NA NA NA NA NA NA NA NA NA ...  
## $ date : Factor w/ 61 levels "2012-10-01","2012-10-02",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ interval: int 0 5 10 15 20 25 30 35 40 45 ...

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

total\_steps\_per\_day <- tapply(activity$steps, activity$date, sum)  
hist(total\_steps\_per\_day, breaks = 10)



average\_daily<- mean(total\_steps\_per\_day, na.rm = TRUE)  
median\_daily<- median(total\_steps\_per\_day, na.rm = TRUE)  
average\_daily

## [1] 10766.19

median\_daily

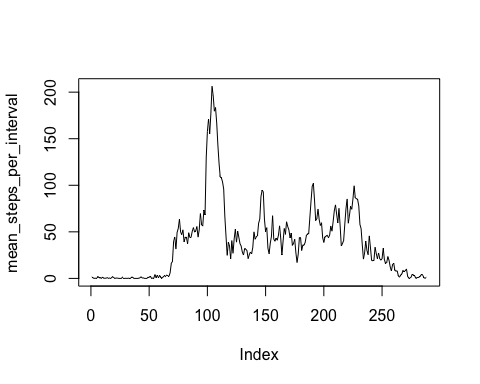
## [1] 10765

## What is the average daily activity pattern?

mean\_steps\_per\_interval<- tapply(activity$steps, activity$interval, mean, na.rm = TRUE)  
str(mean\_steps\_per\_interval)

## num [1:288(1d)] 1.717 0.3396 0.1321 0.1509 0.0755 ...  
## - attr(\*, "dimnames")=List of 1  
## ..$ : chr [1:288] "0" "5" "10" "15" ...

plot(mean\_steps\_per\_interval, type = "l")



max(mean\_steps\_per\_interval)

## [1] 206.1698

my\_obs<- which(mean\_steps\_per\_interval== max(mean\_steps\_per\_interval))  
my\_obs

## 835   
## 104

## Imputing missing values

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