Reproducible Research: Peer Assessment 1

Jianling Peng

April 30, 2019

### set up working directory

setwd("C:/Users/jpeng11/coursera/Reproducilbe Research/RepData\_PeerAssessment1")

## Downlodad data

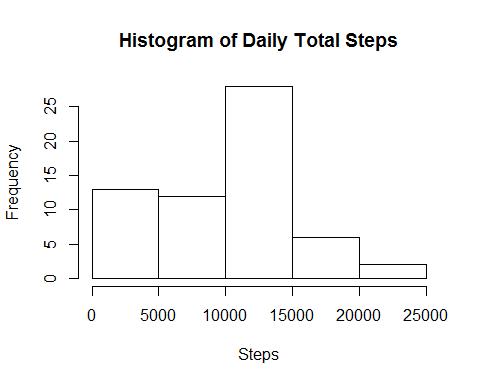
unzip(zipfile = "activity.zip", exdir = ".", unzip = "internal")

## Loading and preprocessing the data

activity <- read.csv("activity.csv", stringsAsFactors = F, header = T)  
head(activity)  
activity$Date <- as.Date(activity$date, format = "%Y-%m-%d")

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

daily\_steps <- tapply(activity$steps, activity$Date, sum, na.rm = T)  
hist(daily\_steps, xlab = "Steps", main = "Histogram of Daily Total Steps")



## What is mean and median number of steps taken per day?

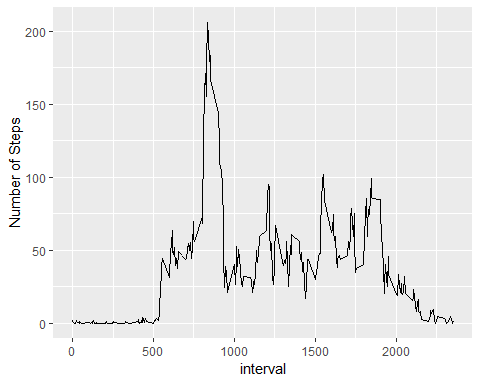
mean\_step <- mean(daily\_steps, na.rm = T)  
median\_step <- median(daily\_steps, na.rm = T)

#### The mean steps of total number of steps taken per day is 9354.2295082

#### The meidan steps of total number of steps taken per day is 10395

## Time series plot of the average number of steps taken

intervals <- activity %>% group\_by(interval) %>%  
 summarise(average = mean(steps, na.rm = T))  
gg <- ggplot(data = intervals, aes(x = interval, y = average)) + geom\_line()   
gg + labs(y = "Number of Steps")



max\_step <- intervals %>% filter(average == max(average))  
max\_interval <- max\_step$interval

#### the interval 835, on average across all the days in the dataset, contain the maximum number of steps, which is 206.1698113.

## Total number of missing values

miss\_num <- sum(!complete.cases(activity))

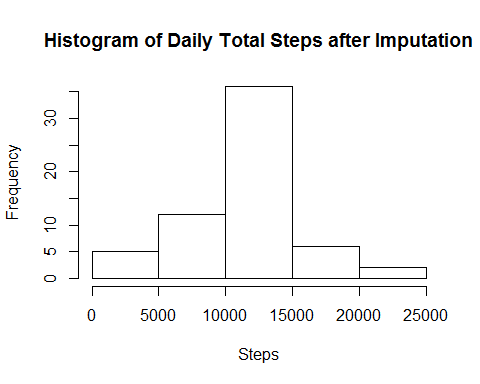
#### Total number of missing values in the dataset is 2304.

## Imputing missing values with mean value for each interval

imputed\_activity <- activity  
for (i in 1:dim(imputed\_activity)[1]) {  
   
 if (is.na(imputed\_activity$steps[i])){  
 int <- imputed\_activity$interval[i]  
 sub <-intervals[intervals$interval == int,]$average  
 imputed\_activity$steps[i] <- sub  
 }  
 #else imputed\_activity$steps <- imputed\_activity$steps  
}

## Total number of steps take per day after imputing missing values

daily\_steps\_imp <- tapply(imputed\_activity$steps, imputed\_activity$Date, sum, na.rm = T)  
hist(daily\_steps\_imp,xlab = "Steps", main = "Histogram of Daily Total Steps after Imputation")



mean\_step <- mean(daily\_steps\_imp)  
median\_step <- median(daily\_steps\_imp)

#### The mean steps of total number of steps taken per day afterimputation is 1.076618910^{4}

#### The meidan steps of total number of steps taken per day after imputaion is 1.076618910^{4}

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

imputed\_activity <- imputed\_activity %>%   
 mutate(wDay = ifelse((weekdays(Date) %in% c("Saturday", "Sunday")),  
 "WEEKEND", "WEEKDAY")) %>%   
 group\_by(wDay, interval) %>%   
 summarise(average= mean(steps))  
  
gg <- ggplot(imputed\_activity, aes(x = interval, y = average)) +   
 geom\_line() + facet\_grid(wDay ~.)  
gg + labs(y = "Average Number of Steps", title = "Average Number of Steps per 5-min Interval" )

