Reproducible Research\_Course Project 1

## Loading and preprocessing the data

activity <- read.csv("C:/Users/JJQ/Documents/data/activity.csv", header = TRUE, stringsAsFactors = FALSE)  
summary(activity)

## steps date interval   
## Min. : 0.00 Length:17568 Min. : 0.0   
## 1st Qu.: 0.00 Class :character 1st Qu.: 588.8   
## Median : 0.00 Mode :character Median :1177.5   
## Mean : 37.38 Mean :1177.5   
## 3rd Qu.: 12.00 3rd Qu.:1766.2   
## Max. :806.00 Max. :2355.0   
## NA's :2304

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

library(dplyr)

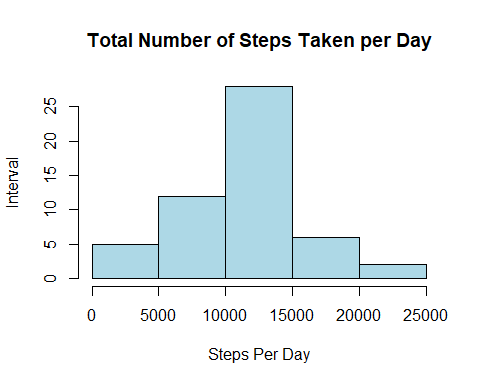
## Warning: package 'dplyr' was built under R version 3.5.1

##   
## 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

steps <- activity %>% select(steps, date) %>% group\_by(date) %>% summarise(steps\_per\_day = sum(steps))  
  
hist(steps$steps\_per\_day, main = "Total Number of Steps Taken per Day", xlab = "Steps Per Day", ylab = "Interval", col = "light blue")



mean\_dailysteps <- mean(steps$steps\_per\_day, na.rm = TRUE)  
mean\_dailysteps

## [1] 10766.19

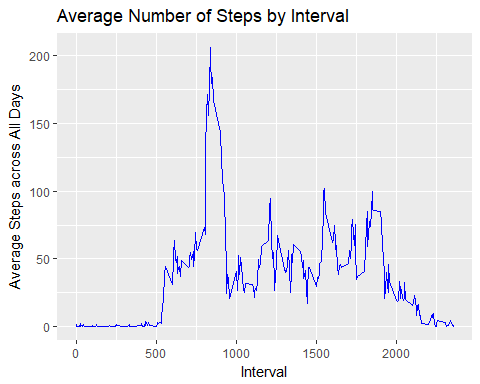
median\_dailysteps <- median(steps$steps\_per\_day, na.rm = TRUE)  
median\_dailysteps

## [1] 10765

Answer: The mean of the total number of steps taken per day is equal to 10766.19 The median of the total number of steps taken per day is equal to 10765

## What is the average daily activity pattern?

library(ggplot2)  
  
average <- group\_by(activity, interval) %>% summarise(average\_steps = mean(steps, na.rm = TRUE))  
  
ggplot(average, aes(x = interval, y = average\_steps, type = "l")) + geom\_line(color = "blue") +   
 ggtitle("Average Number of Steps by Interval") + xlab("Interval") + ylab("Average Steps across All Days")



interval\_max\_steps <- filter(average, average\_steps == max(average$average\_steps))  
interval\_max\_steps

## # A tibble: 1 x 2  
## interval average\_steps  
## <int> <dbl>  
## 1 835 206.

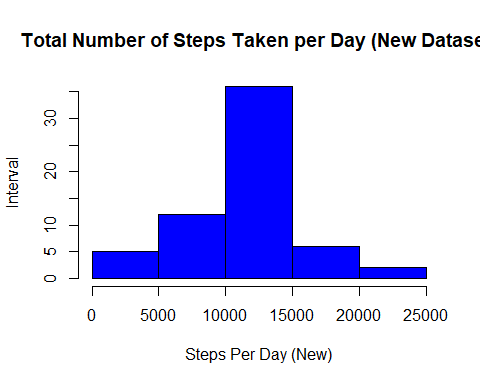
Answer: interval 835 contains the maximum number of steps. The person is very active at around 2:00pm (to be accurate at 1:55pm)

## Imputig missing values

summary(is.na(activity))

## steps date interval   
## Mode :logical Mode :logical Mode :logical   
## FALSE:15264 FALSE:17568 FALSE:17568   
## TRUE :2304

activity[is.na(activity)] <- c(average$average\_steps)  
activity\_new <- activity  
  
steps\_new <- activity\_new %>% select(steps, date) %>% group\_by(date) %>% summarise(steps\_per\_day\_new = sum(steps))  
  
hist(steps\_new$steps\_per\_day\_new, main = "Total Number of Steps Taken per Day (New Dataset)", xlab = "Steps Per Day (New)", ylab = "Interval", col = "blue")



mean\_dailysteps\_new <- mean(steps\_new$steps\_per\_day\_new, na.rm = TRUE)  
mean\_dailysteps\_new

## [1] 10766.19

median\_dailysteps\_new <- median(steps\_new$steps\_per\_day\_new, na.rm = TRUE)  
median\_dailysteps\_new

## [1] 10766.19

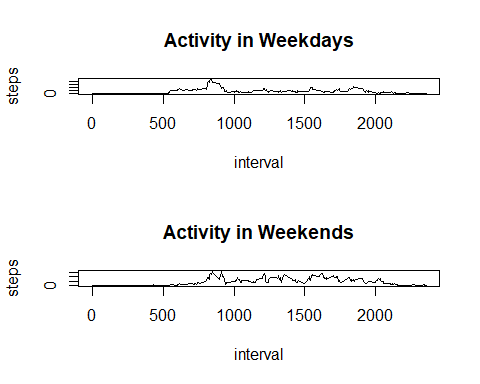
Answer: There are 2304 missing value in total in the dataset The mean of the total number of steps taken per day didn’t change after imputing, still equal to 10766.19 The median of the total number of steps taken per day increased slightly to 10766.19 from 10765, after imputing

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

wkd <- weekdays(as.Date(strptime(activity\_new$date, format = "%Y-%m-%d")))  
wkd <- as.factor(wkd)  
  
activity\_by\_date <- cbind(activity\_new, wkd)  
  
levels(activity\_by\_date$wkd) <- list(weekday = c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday"),   
 weekend = c("Saturday", "Sunday"))  
levels(activity\_by\_date$wkd)

## [1] "weekday" "weekend"

par(mfrow = c(2,1))  
with(activity\_by\_date[activity\_by\_date$wkd == "weekday", ], plot(aggregate(steps ~ interval, FUN = mean), type = "l", main = "Activity in Weekdays"))  
with(activity\_by\_date[activity\_by\_date$wkd == "weekend", ], plot(aggregate(steps ~ interval, FUN = mean), type = "l", main = "Activity in Weekends"))



Answer: The person is more activate in weekends than in weekdays