Read the data from the file

activity <- read.csv("C:/Users/Vyjayanthi/Desktop/activity.csv")  
summary(activity)

## steps date interval   
## Min. : 0.00 2012-10-01: 288 Min. : 0.0   
## 1st Qu.: 0.00 2012-10-02: 288 1st Qu.: 588.8   
## Median : 0.00 2012-10-03: 288 Median :1177.5   
## Mean : 37.38 2012-10-04: 288 Mean :1177.5   
## 3rd Qu.: 12.00 2012-10-05: 288 3rd Qu.:1766.2   
## Max. :806.00 2012-10-06: 288 Max. :2355.0   
## NA's :2304 (Other) :15840

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

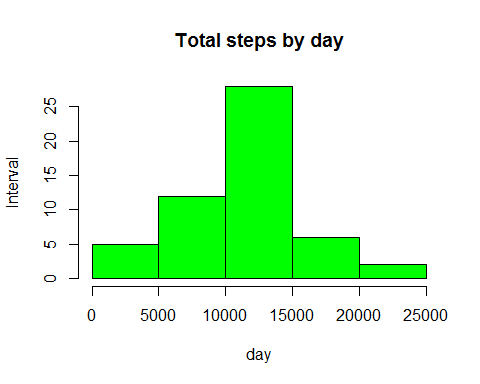
library(lattice)  
activity$date <- as.Date(activity$date, "%Y-%m-%d")

To draw the histogram of total no. of steps taken

totalSteps <- aggregate(steps ~ date, data = activity, sum, na.rm = TRUE)  
summary(totalSteps)

## date steps   
## Min. :2012-10-02 Min. : 41   
## 1st Qu.:2012-10-16 1st Qu.: 8841   
## Median :2012-10-29 Median :10765   
## Mean :2012-10-30 Mean :10766   
## 3rd Qu.:2012-11-16 3rd Qu.:13294   
## Max. :2012-11-29 Max. :21194

hist(totalSteps$steps,   
 main = "Total steps by day",   
 xlab = "day",   
 ylab = "Interval",  
 col = "green")



The mean of the total steps taken

mean(totalSteps$steps)

## [1] 10766.19

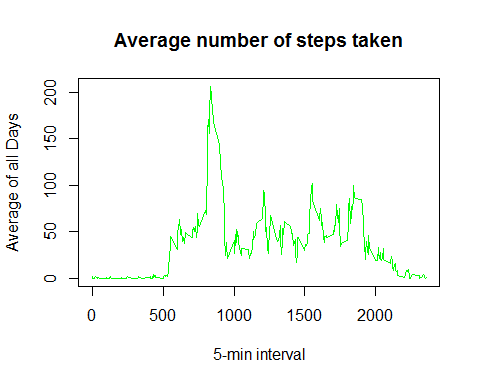
The median of the total steps taken

median(totalSteps$steps)

## [1] 10765

time series plot of the steps taken with 5 min. interval with average steps taken

time\_series <- tapply(activity$steps, activity$interval, mean, na.rm = TRUE)  
  
plot(row.names(time\_series), time\_series, type = "l",   
 main = "Average number of steps taken",   
 xlab = "5-min interval",   
 ylab = "Average of all Days",   
 col = "green")



Maximum no. of steps taken during

max\_interval <- which.max(time\_series)  
names(max\_interval)

## [1] "835"

To impute missing values, first find out how many missing values are there

activityNA <- sum(is.na(activity))  
activityNA

## [1] 2304

Replace the missing value by mean in 5 min interval

StepsAverage <- aggregate(steps ~ interval, data = activity, FUN = mean)  
fill\_na <- numeric()  
for (i in 1:nrow(activity)) {  
 obs <- activity[i, ]  
 if (is.na(obs$steps)) {  
 steps <- subset(StepsAverage, interval == obs$interval)$steps  
 } else {  
 steps <- obs$steps  
 }  
 fill\_na <- c(fill\_na, steps)  
}

New data set with all the filled in missing data

activity2 <- activity  
activity2$steps <- fill\_na  
head(activity2)

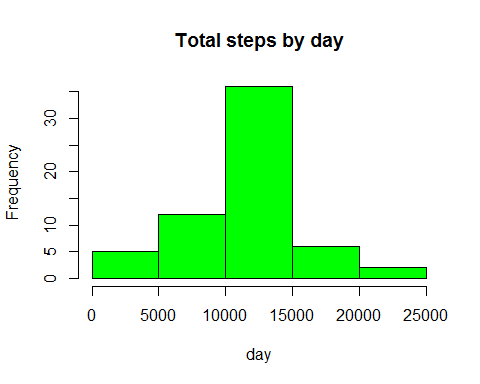
## steps date interval  
## 1 1.7169811 2012-10-01 0  
## 2 0.3396226 2012-10-01 5  
## 3 0.1320755 2012-10-01 10  
## 4 0.1509434 2012-10-01 15  
## 5 0.0754717 2012-10-01 20  
## 6 2.0943396 2012-10-01 25

totalSteps2 <- aggregate(steps ~ date, data = activity2, sum, na.rm = TRUE)  
head(totalSteps2)

## date steps  
## 1 2012-10-01 10766.19  
## 2 2012-10-02 126.00  
## 3 2012-10-03 11352.00  
## 4 2012-10-04 12116.00  
## 5 2012-10-05 13294.00  
## 6 2012-10-06 15420.00

Histogram of total no. of steps taken each day

hist(totalSteps2$steps, main = "Total steps by day",   
 xlab = "day",   
 col = "green")



Calculate mean of total steps taken per day

mean(totalSteps2$steps)

## [1] 10766.19

Calculte median of total steps taken per day

median(totalSteps2$steps)

## [1] 10766.19

Differences in activity patterns between weekdays and weekends