Reproducible Research: Peer Assessment 1

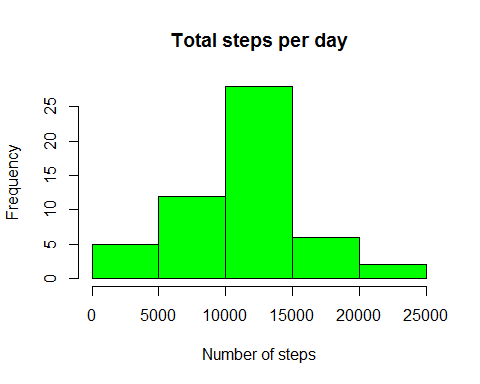
## Loading and preprocessing the data

activity <- read.csv("~/activity.csv", colClasses = c("numeric", "character", "numeric"))

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

First sum steps by day and create histogram.

daily\_steps <- aggregate (steps ~ date, data= activity, sum, na.rm=TRUE)  
hist(daily\_steps$steps, main = paste ("Total steps per day"), col="green", xlab="Number of steps")



Then calculate mean and median number of steps per day...

mean\_steps <- mean(daily\_steps$steps)  
mean\_steps

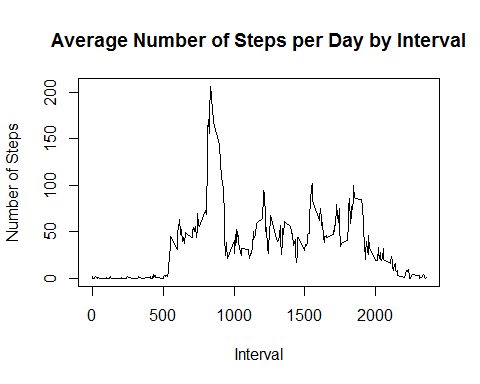
## [1] 10766.19

median\_steps <- median(daily\_steps$steps)  
median\_steps

## [1] 10765

## What is the average daily activity pattern?

steps\_by\_interval <- aggregate(steps ~ interval, data= activity, mean)  
  
plot(steps\_by\_interval$interval,steps\_by\_interval$steps, type="l", xlab="Interval", ylab="Number of Steps",main="Average Number of Steps per Day by Interval")



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

max\_interval <- steps\_by\_interval[which.max(steps\_by\_interval$steps),1]

## Imputing missing values

Determine number of rows with missing values

activity\_NAS <- sum (!complete.cases(activity))  
activity\_NAS

## [1] 2304

Imputing missing values with average of corresponding interval

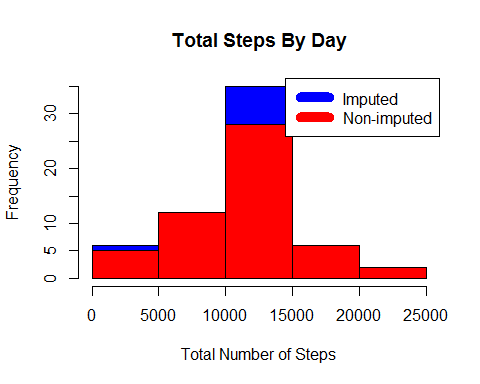
imputed\_Vals <- transform(activity, steps = ifelse(is.na(activity$steps), steps\_by\_interval$steps[match(activity$interval, steps\_by\_interval$interval)], activity$steps))

Particular treatment for 10-01-2012. missing value was replaced by 0 to avoid discrepancies with following days

imputed\_Vals[as.character(imputed\_Vals$date) == "2012-10-01", 1] <- 0

Recount total steps per day and create histogram

daily\_steps2 <- aggregate(steps ~ date, imputed\_Vals, sum)  
hist(daily\_steps2$steps, main = paste("Total Steps By Day"), col="blue", xlab="Total Number of Steps")  
  
#Create Histogram to show difference.   
hist(daily\_steps$steps, main = paste("Total Steps By Day"), col="red", xlab="Total Number of Steps", add=T)  
legend("topright", c("Imputed", "Non-imputed"), col=c("blue", "red"), lwd=10)



compute new mean and median from imputed data

new\_mean <- mean(daily\_steps2$steps)  
new\_mean

## [1] 10589.69

new\_median <- median(daily\_steps2$steps)  
new\_median

## [1] 10766.19

Calculate difference between non-imputed and imputed data

mean\_diff <- new\_mean - mean\_steps  
format(mean\_diff, digits=2, nsmall=2)

## [1] "-176.49"

med\_diff <- new\_median - median\_steps  
format(med\_diff, digits=2, nsmall=2)

## [1] "1.19"

Compute total difference

total\_diff <- sum (daily\_steps2$steps)- sum (daily\_steps$steps)

* The imputed data mean is 1.058969410^{4}
* The imputed data median is 1.076618910^{4}
* The difference between the imputed and non-imputed mean is -176.4948964
* The difference between the imputed and non-imputed median is 1.1886792
* The difference between the total number of steps in imputed and non imputed data is 7.536332110^{4}.

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

Create plot to visualize patterns of activity for weekdays and weekends

weekdays <- c("Monday", "Tuesday", "Friday", "Wednesday", "Thursday", "Friday")  
  
imputed\_Vals$dow = as.factor(ifelse(is.element(weekdays(as.Date(imputed\_Vals$date)),weekdays), "Weekday", "Weekend"))  
  
steps\_by\_intervals <- aggregate(steps ~ interval + dow, imputed\_Vals, mean)  
  
library(lattice)  
  
xyplot(steps\_by\_intervals$steps ~ steps\_by\_intervals$interval|steps\_by\_intervals$dow, main="Average Steps per Day by Interval",xlab="Interval", ylab="Steps",layout=c(1,2), type="l")

