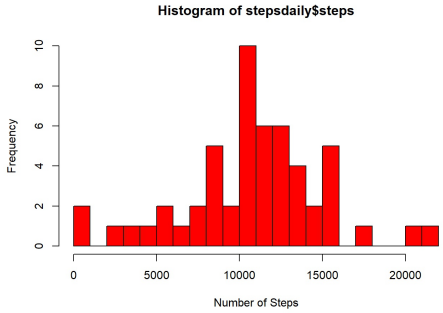
# Reproducible Research: Peer Assessment 1

## Loading and preprocessing the data

```
data <- read.csv("activity.csv")
```

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

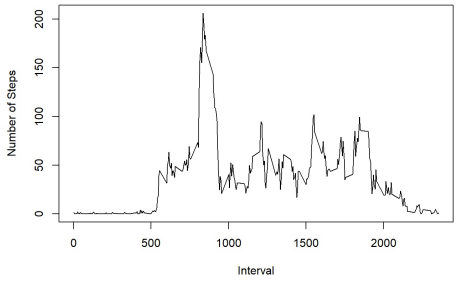
```
stepsdaily<-aggregate(steps~date, data, sum)
hist(stepsdaily$steps,breaks=25, col="red",  xlab="Number of Steps")
```




## What is the average daily activity pattern?

```
steps_by_interval <- aggregate(steps ~ interval, data, mean)
stepsdaily <- aggregate(steps ~ date, data, sum)

plot(steps_by_interval$interval,steps_by_interval$steps, type="l", xlab="Interval", ylab="Number of Steps")
```




```
medsteps<-median(stepsdaily$steps)
meansteps<-mean(stepsdaily$steps)
max_interval <- steps_by_interval[which.max(steps_by_interval$steps),1]
```

The `mean` is 1.076618910^{4} ; the `median` is 10765 with an interval `maximum` interval of 835

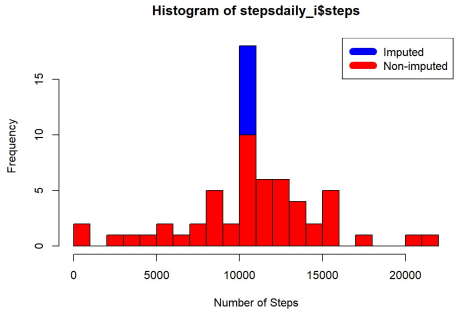
## Imputing missing values

```
incomplete <- sum(is.na(data))
imputed_data <- transform(data, steps = ifelse(is.na(data$steps), steps_by_interval$steps[match(data$interval,
steps_by_interval$interval)], data$steps))
```

Recount total steps by day and create Histogram.

```
stepsdaily_i <- aggregate(steps ~ date, imputed_data, sum)
hist(stepsdaily_i$steps, breaks= 25, col="blue", xlab="Number of Steps")

#Create Histogram to show difference.
hist(stepsdaily$steps, breaks=25, col="red", xlab="Number of Steps", add=T)
legend("topright", c("Imputed", "Non-imputed"), col=c("blue", "red"), lwd=10)
```




Calculate new mean and median for imputed data.

```
meansteps.i <- mean(stepsdaily_i$steps)
medsteps.i <- median(stepsdaily_i$steps)
```

Calculate difference between imputed and non-imputed data.

```
mean_diff <- meansteps.i - meansteps
med_diff <- medsteps.i - medsteps
```

Calculate total difference.

```
total_diff <- sum(stepsdaily_i$steps) - sum(stepsdaily$steps)
```

- The imputed data mean is 1.076618910^{4}
- The imputed data median is 1.076618910^{4}
- The difference between the non-imputed mean and imputed mean is 0
- The difference between the non-imputed median and imputed median is 1.1886792
- By imputing with average amounts– the median and mean remain the similar
- The difference between total number of steps between imputed and non-imputed data is 8.612950910^{4}. Thus, there were 8.612950910^{4} more steps in the imputed data.

*However, a measure of kurtosis shows differences between the two distributions of data– both have very large kurtosis values (data does not follow a normal distribution)

```
library(e1071)
kurtosis(data$steps, na.rm=TRUE)
```

```
## [1] 18.43161
```

```
kurtosis(imputed_data$steps, na.rm=TRUE)
```

```
## [1] 20.82177
```

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

Created a plot to compare and contrast number of steps between the week and weekend. There is a higher peak earlier on weekdays, and more overall activity on weekends.

```
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")
imputed_data$dow = as.factor(ifelse(is.element(weekdays(as.Date(imputed_data$date)),weekdays), "Weekday", "Weekend"))

steps_by_interval_i <- aggregate(steps ~ interval + dow, imputed_data, mean)

library(lattice)

xyplot(steps_by_interval_i$steps ~ steps_by_interval_i$interval|steps_by_interval_i$dow, main="STEPS_BY_DAY",xlab="Interval", ylab="Steps",layout=c(1,2), type="l")
```