Load the data

```
rawdata <- read.csv("activity.csv", header = T, stringsAsFactors = F)
dim(rawdata)
```

```
## [1] 17568     3
```

Transform the data

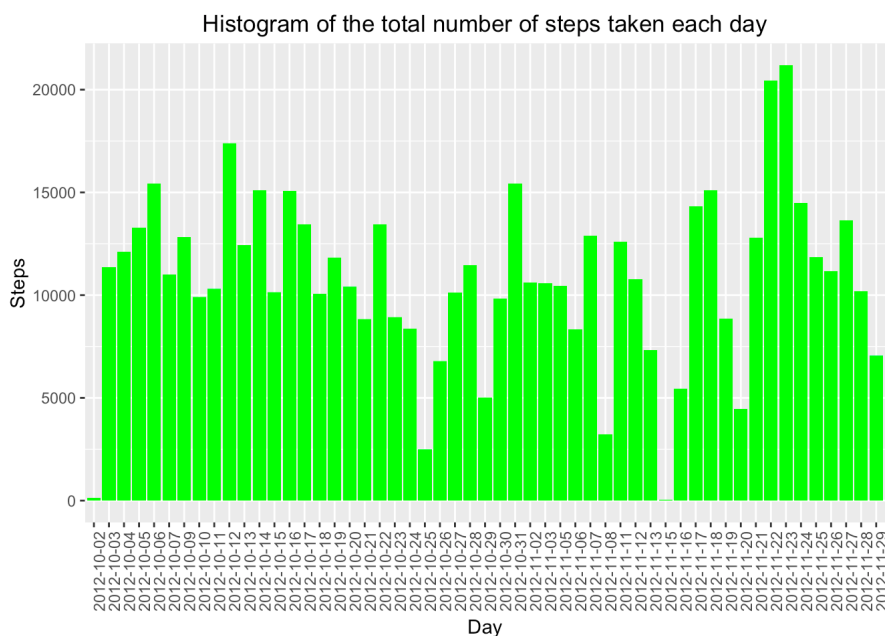```
noNA <- rawdata[complete.cases(rawdata),]
dim(noNA)
```

```
## [1] 15264     3
```

mean total number of steps taken per day

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```r
ggplot(noNA, aes(as.factor(date),steps))+geom_bar(fill="green", stat="identity")+    xlab("Day") + ylab("Steps")+ggtitle("Histogram
```

Histogram of the total number of steps taken each day

Calculate the mean of the total number of steps taken per day

```r
meanFreq <- as.integer(mean(aggregate(steps~date,noNA, FUN=sum)$steps))    meanFreq
```
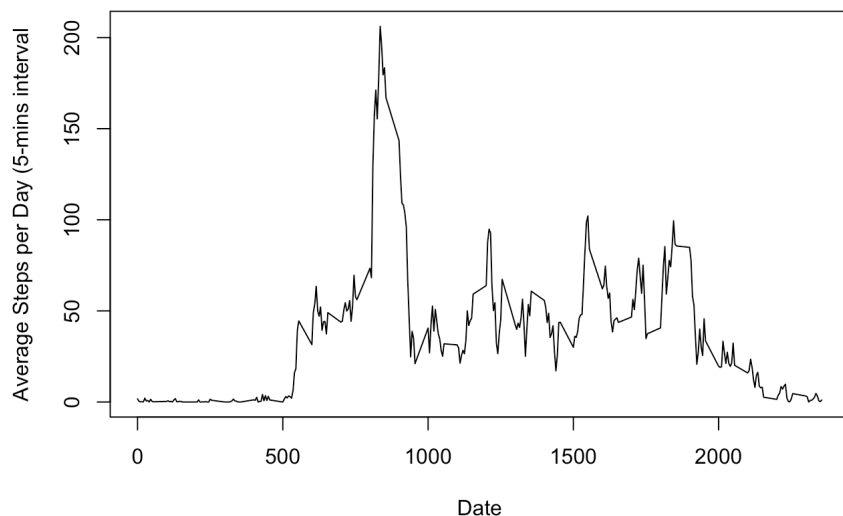
```
## [1] 10766
```

Calculate the median of the total number of steps taken per day

```
medianFreq <- median(aggregate(steps~date,noNA, FUN=sum)$steps)
medianFreq
```

```
## [1] 10765
```

What is the average daily activity pattern

```r
q2 <- aggregate(steps~interval, noNA, FUN=mean)  plot(q2$steps ~ q2$interval, type="l", xlab="Date", ylab="Average Steps per Day (
```

Imputing missing values

```
numNA <- nrow(rawdata) - nrow(noNA)
imputedata <- rawdata
meanNA <- aggregate(steps~interval,imputedata, FUN=mean)
for(i in 1:nrow(imputedata)){
  if(is.na(imputedata[i,c("steps")])){
    imputedata[i,c("steps")] <- meanNA[meanNA$interval==imputedata[i,c("interval")], c("steps")]
  }
}
ggplot(imputedata,aes(as.factor(date),steps))+geom_bar(fill="green", stat="identity")+xlab("Day") + ylab("Steps")
+ggtitle("Histogram of the Total Number of Steps Taken Each Day (Imputed Data)")+ theme(axis.text.x = element_tex
t(angle = 90, hjust = 2))
```
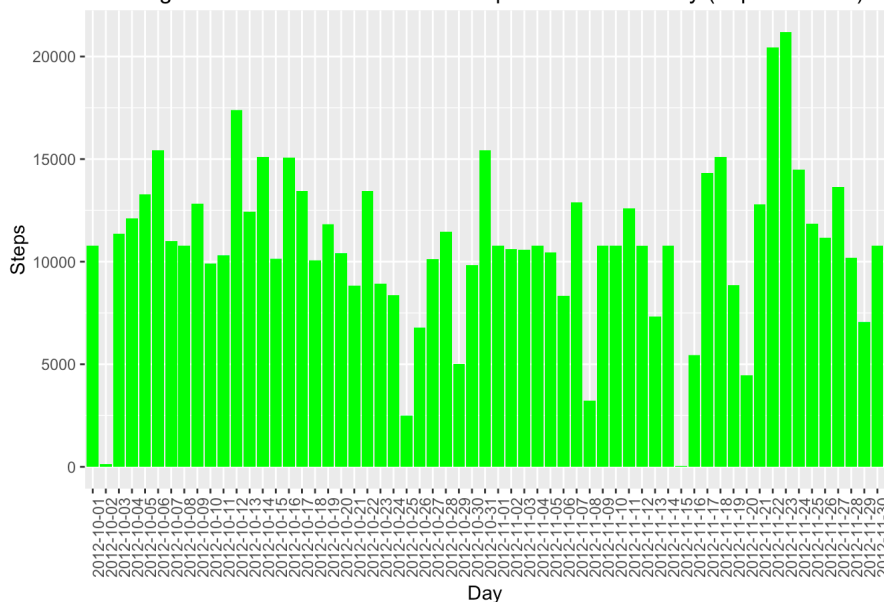

Calculate the mean and median total number of steps taken per day

```
newmeanfreq <- as.integer(mean(aggregate(steps~date, imputedata, FUN = sum)$steps))
newmeanfreq
```

```
## [1] 10766
```

```
newmedianfreq <- median(aggregate(steps~date, imputedata, FUN = sum)$steps)
newmedianfreq
```

```
## [1] 10766.19
```

1 There is no diffrence between old meanFreq value and new meanfreq value. 2 There is .66 diffrence between old median value and new median value.

differences in activity patterns between weekdays and weekends

```
weekDayEnd <- data.frame("dayType"=character(0))
weekDayEnd <- NULL

#Check wether the day is a weekday or weekend (Saturday)
weekInd <- function(day){
  if (weekdays(day)=="Saturday") {
    out <- "Weekend"
  } else {
    out <- "Weekday"
  }
  out
}

for (i in 1:nrow(imputedata)){
  date <- imputedata[i,c("date")]
  newObs <- data.frame("dayType"= weekInd(as.Date(date)))
  weekDayEnd <- rbind(weekDayEnd, newObs)
}
```

Add the day type column in the data set

```
imputedata <- cbind(imputedata, weekDayEnd)
averageWeekDayEnd<-aggregate(steps~interval + dayType, imputedata, FUN=mean)
ggplot(averageWeekDayEnd,aes(interval,steps, color=dayType))+geom_line()+facet_wrap(~dayType, ncol=1)+xlab("5-min Interval")+ylab("Steps")+guides(fill=FALSE)
```