Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

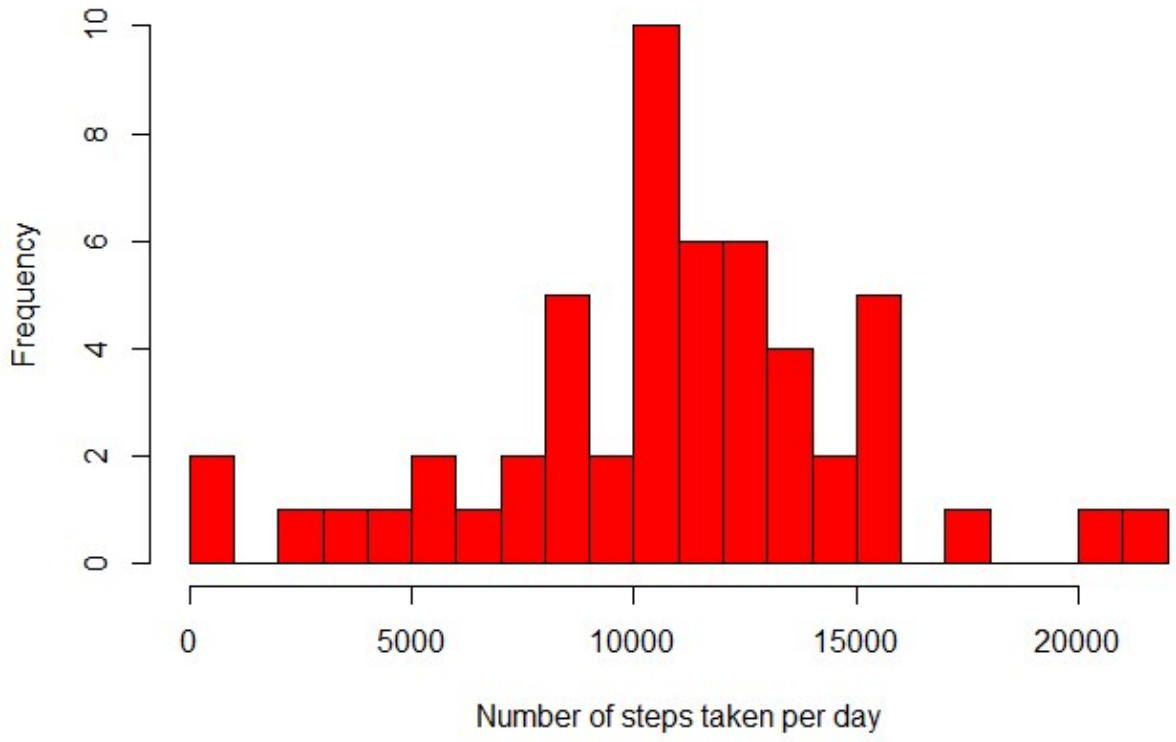
```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
options(scipen=999)
act<-read.csv("D:\\Coursera\\RepData_PeerAssessment1\\activity\\activity.csv")
```

What is mean total number of steps taken per day?

```
sums<-tapply(act$steps,act$date,sum)
sums<-sums[!is.na(sums)]
hist(sums,breaks=20,xlab="Number of steps taken per day",col="red",main=NULL)
```

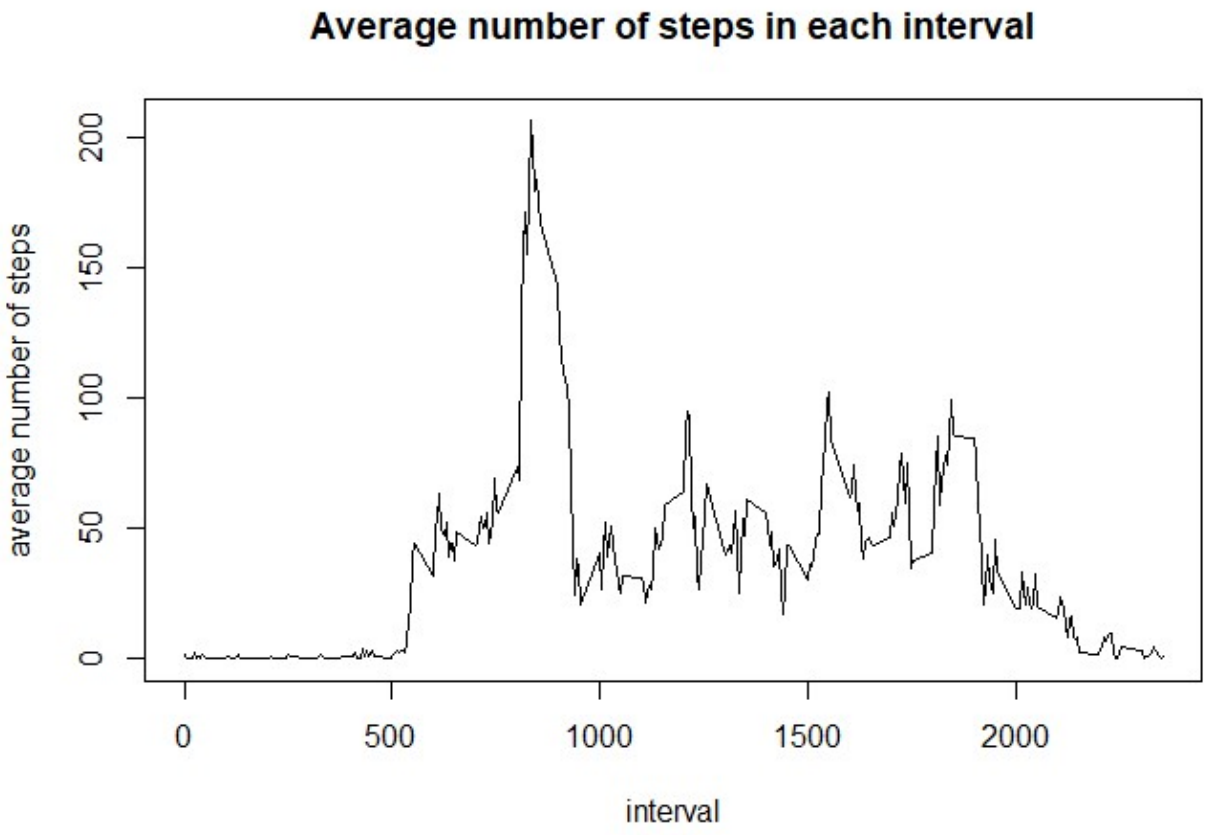



```
me<-mean(sums)
md<-median(sums)
```

Thus we can see the mean number of steps taken per day is 10766.2 and the median number of steps is 10765.

What is the average daily activity pattern?

```
ave<-tapply(act$steps,act$interval,mean,na.rm=T)
plot(unique(act$interval),ave,type="l",xlab="interval",ylab="average number of steps",
main="Average number of steps in each interval")
```




```
maxst<-max(ave)
ind<-match(maxst,ave)
maxint<-names(ave[ind])
```

The highest average number of steps is 206.17 in interval 835

Imputing missing values

```
tbna<-table(is.na(act$steps))
```

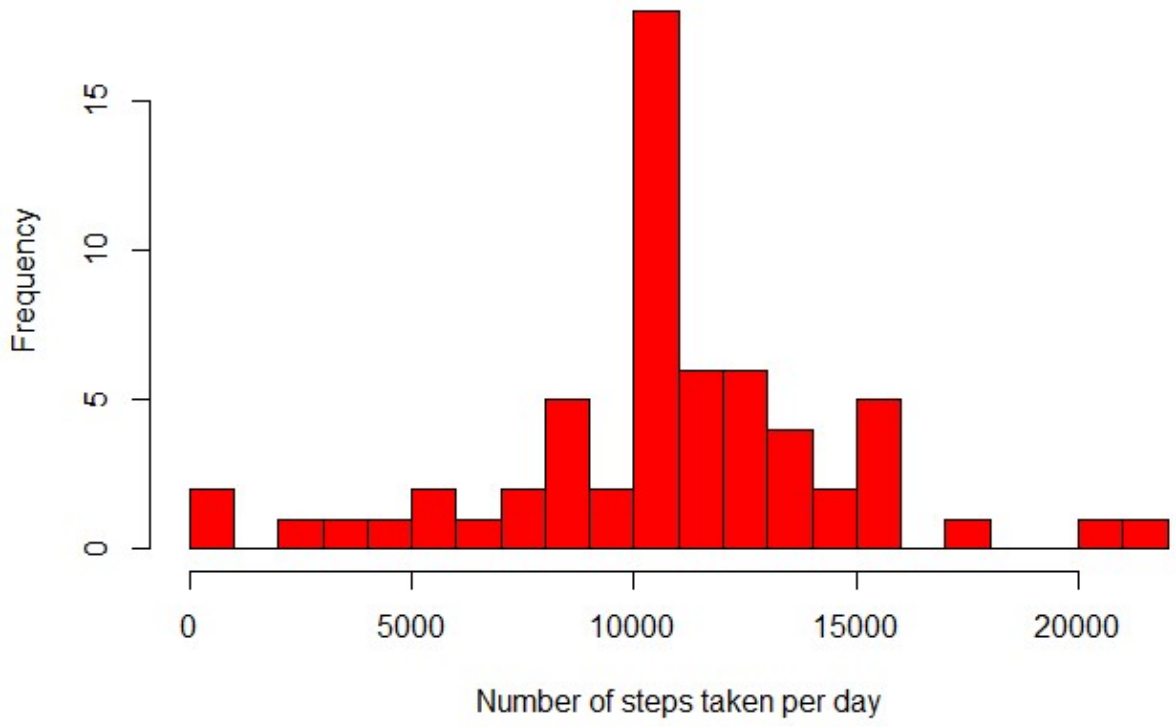
The number of NA values present is 2304.

To replace the missing values we will use the average steps taken for the interval that corresponds to that NA value (since the averages are decomals we use the round function to 0 decimal places).

```
fil<-filter(act,is.na(act$steps))
replNA<-NULL
for (i in 1:length(fil$steps)){
        replNA<-c(replNA,ave[match(fil$interval[i],act$interval)])
}
colRep = act$steps
ind<-which(is.na(act$steps)==TRUE)
for (i in 1:length(ind)){
        colRep[ind[i]]=round(replNA[i])
}

actF<-mutate(act,steps=colRep)

sums<-tapply(actF$steps,act$date,sum)
sums<-sums[!is.na(sums)]
hist(sums,breaks=20,xlab="Number of steps taken per day",col="red",main=NULL)
```



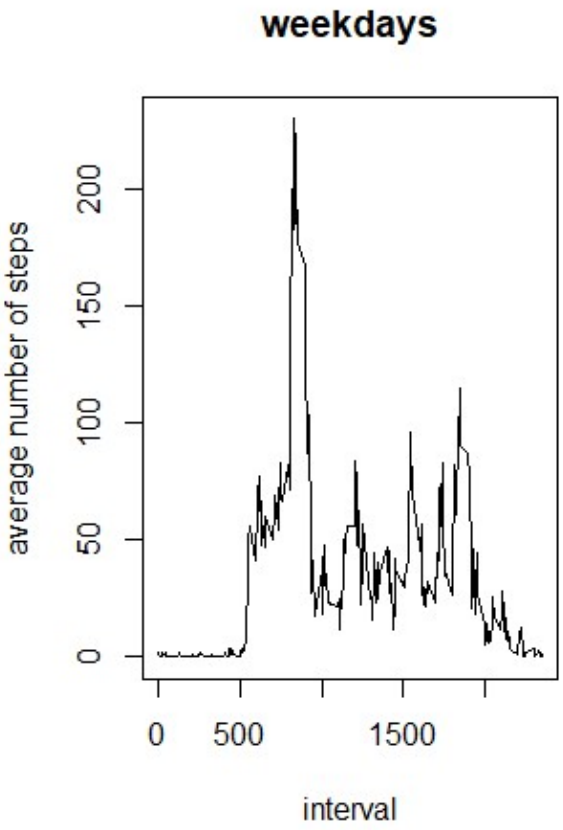

```
me<-mean(sums)
md<-median(sums)
```

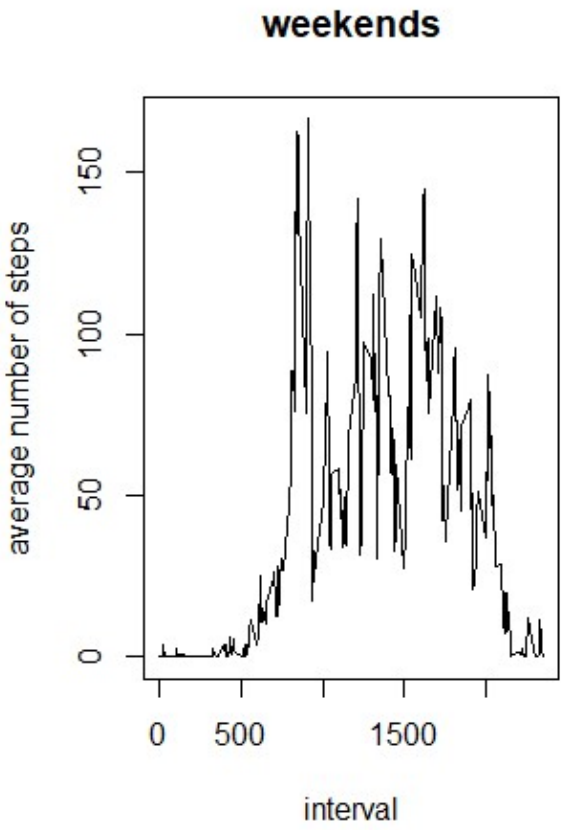
From the new histogram we see that compared to the old histogram the only reall changes we see is that the number of days where between 10 and 11 thousand steps were taken increases signifigantly. This makes sense as we used the avergae value for each interval with a missing value, hence the length of the days that had missing data is also average (and the average is about 10760 steps. The mean is 10765.64 compared to 10766 and the median is 10762 compared to 10765. Thus we see that the mean and median remain relativiely unchanged which makes sense as we used avergae values to replace the missing values (hence we would not expect to see much difference betweent the new and old values)

Are there differences in activity patterns between weekdays and weekends?

```
library(chron)
weeklog<-is.weekend(as.Date(actF$date))
we<-NULL
for (i in 1:length(weeklog)){
        if (weeklog[i]==FALSE){
        we<-c(we,"weekday")
        } else {
          we<-c(we,"weekend")
        }

}
actF<-mutate(actF,day=we)
ave<-tapply(actF$steps,list(actF$interval,actF$day),mean)
par(mfrow=c(1,2))
plot(unique(actF$interval),ave[,1],type="l",xlab="interval",ylab="average number of st
eps",main="weekdays")
plot(unique(actF$interval),ave[,2],type="l",xlab="interval",ylab="average number of st
eps",main="weekends")
```


weekdays
average number of steps
0
50
100
150
200
0
500
1500
interval


weekends
average number of steps
0
50
100
150
0
500
1500
interval