Reproducible Research: Course Project 1

1. Loading of data,preprocessing data, splitting data by date and plotting by a histogram of total number of steps taken each day

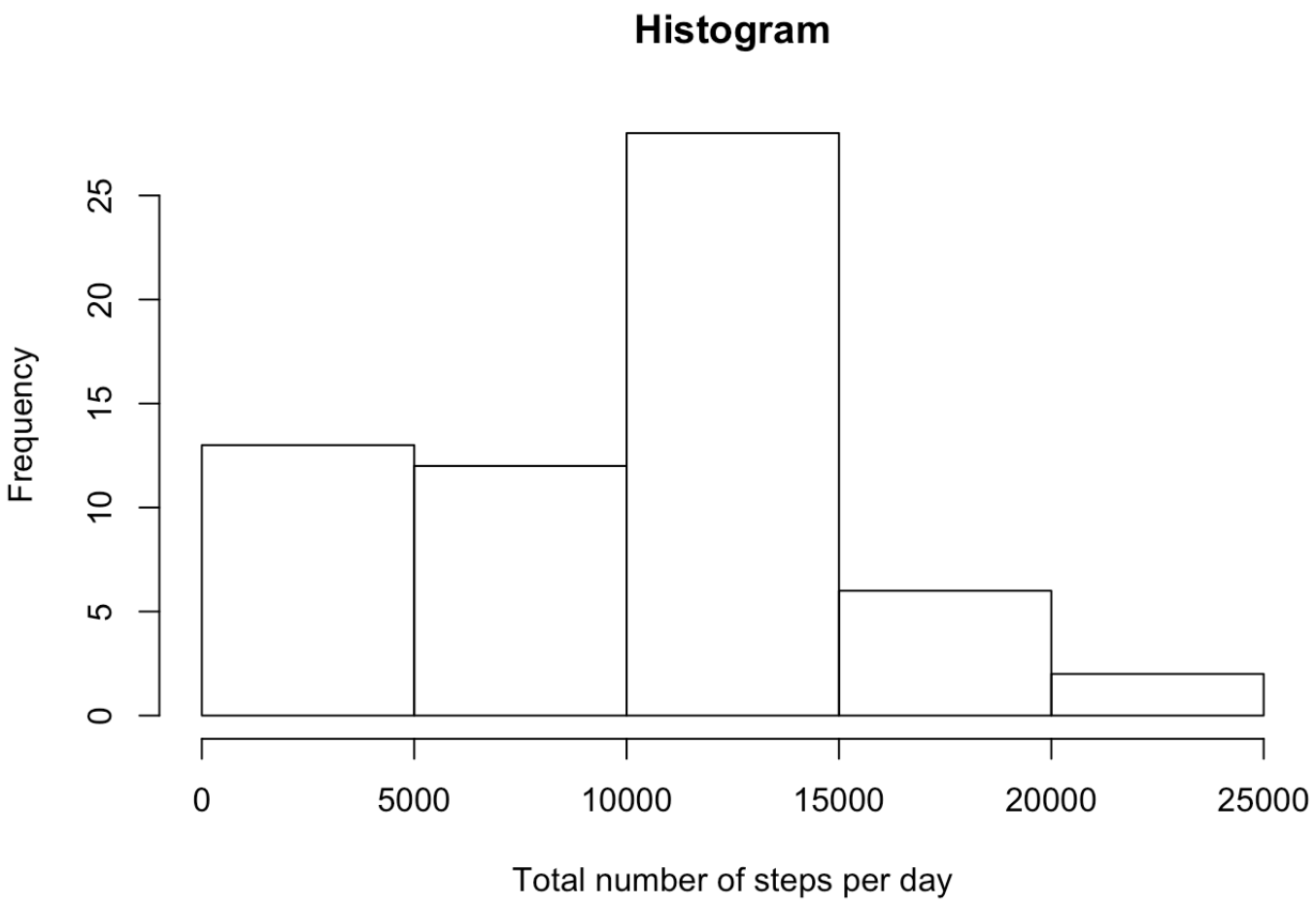
```
data = read.csv('activity.csv')
databydate = split(data,data$date)
Result=sapply(databydate, function(x) sum(x[,1],na.rm=T))
head(Result)
```

```
## 2012-10-01 2012-10-02 2012-10-03 2012-10-04 2012-10-05 2012-10-06
##          0        126      11352      12116      13294      15420
```

```
Todataframe = data.frame(Result)
head(Todataframe)
```

```
##            Result
## 2012-10-01      0
## 2012-10-02    126
## 2012-10-03  11352
## 2012-10-04  12116
## 2012-10-05  13294
## 2012-10-06  15420
```

```
hist(Todataframe$Result, xlab ='Total number of steps per day', main='Histogram')
```



2. Calculating the mean of total number of steps per day

```
mean(Todataframe$Result, na.rm=T)
```

```
## [1] 9354.23
```

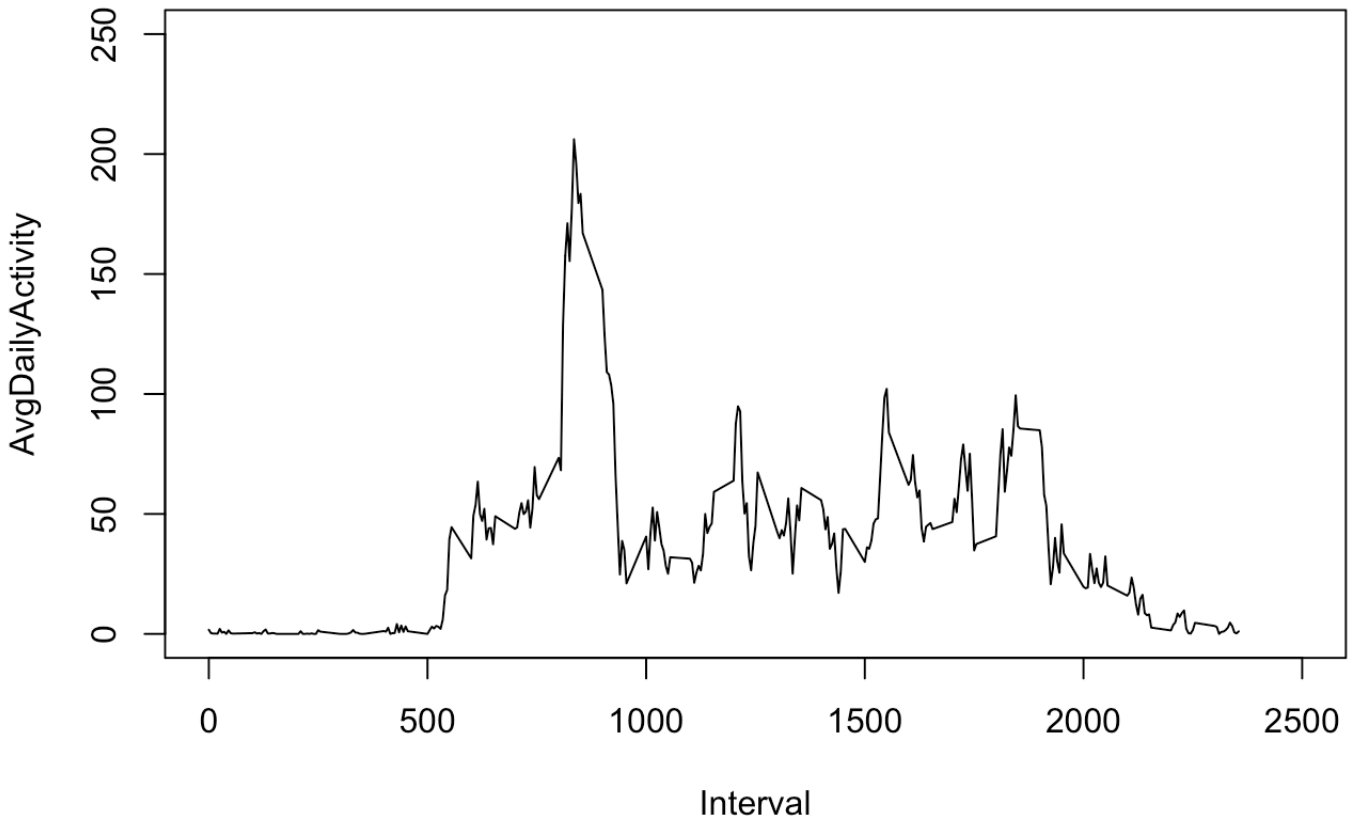
3. Calculating the median

```
median(Todataframe$Result, na.rm = T)
```

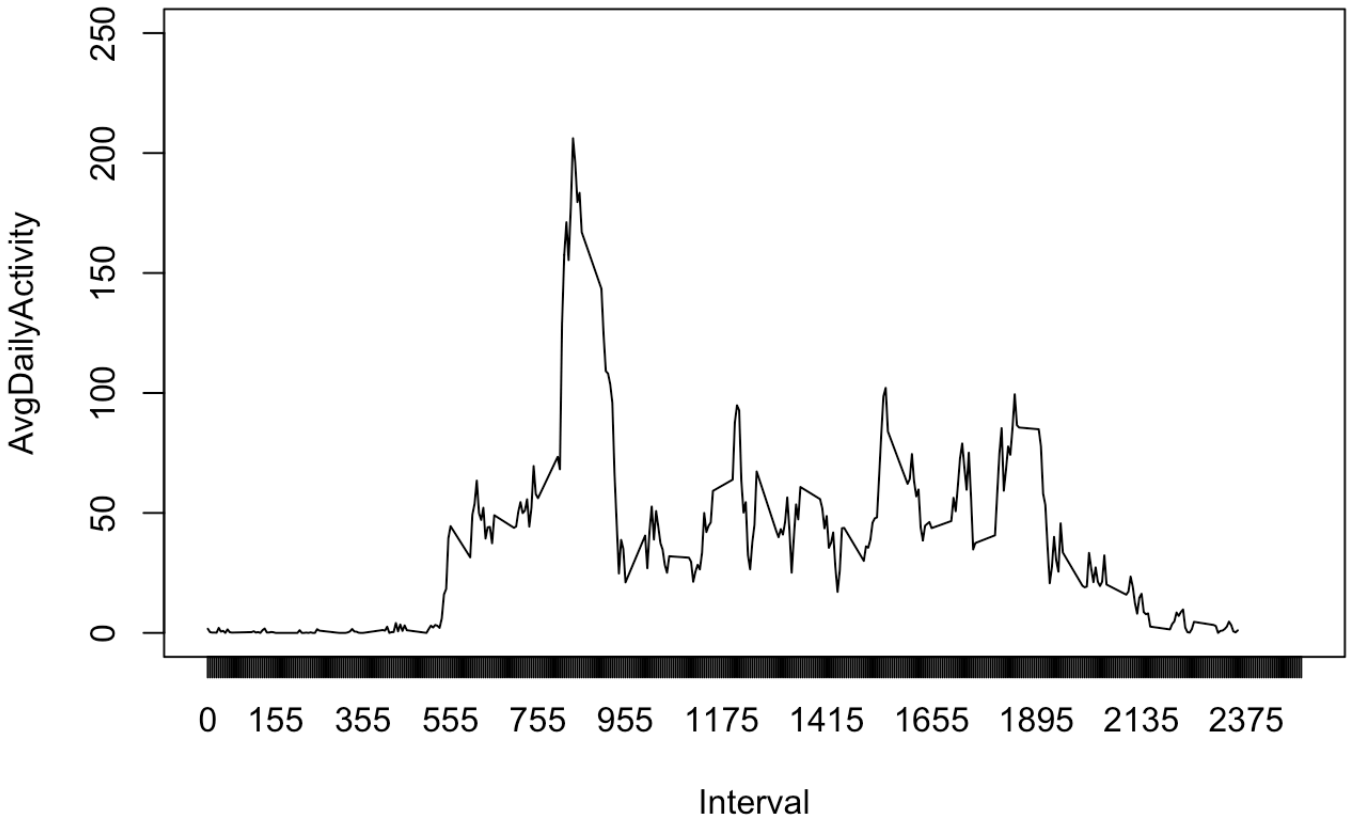
```
## [1] 10395
```

4)Calculating Average daily activity pattern

```
ByDate = split(data, data$date)
ByInterval = split(data, data$interval)
Result=sapply(ByInterval, function(x) mean(x[,1],na.rm=T))
data2 = data.frame(Result)
Name = names(Result)
Name2 = as.numeric(as.character(Name))
Name3 = as.data.frame(Name2)
Newdata = cbind(Name3,data2)
colnames(Newdata)[1]= 'Interval'
colnames(Newdata)[2]= 'AvgDailyActivity'
plot(Newdata, xlim=c(0,2500), ylim=c(0,250), type='l')
```



```
a = seq(from=0, to = 2500, by = 5)
plot(Newdata, xlim=c(0,2500), ylim=c(0,250), type='l', axes=F)
axis(side=1, at=seq(0,2500,by= 5),labels=a )
axis(side=2, at=seq(0,250,by= 50), labels=c('0','50','100','150','200','250'))
box()
```



5) Find the maximum daily activity

```
which.max(Newdata$AvgDailyActivity)
```

```
## [1] 104
```

```
Newdata[104,]
```

```
##     Interval AvgDailyActivity
## 835      835         206.1698
```

6. Find the number of data that has NA values

```
colSums(is.na(data))
```

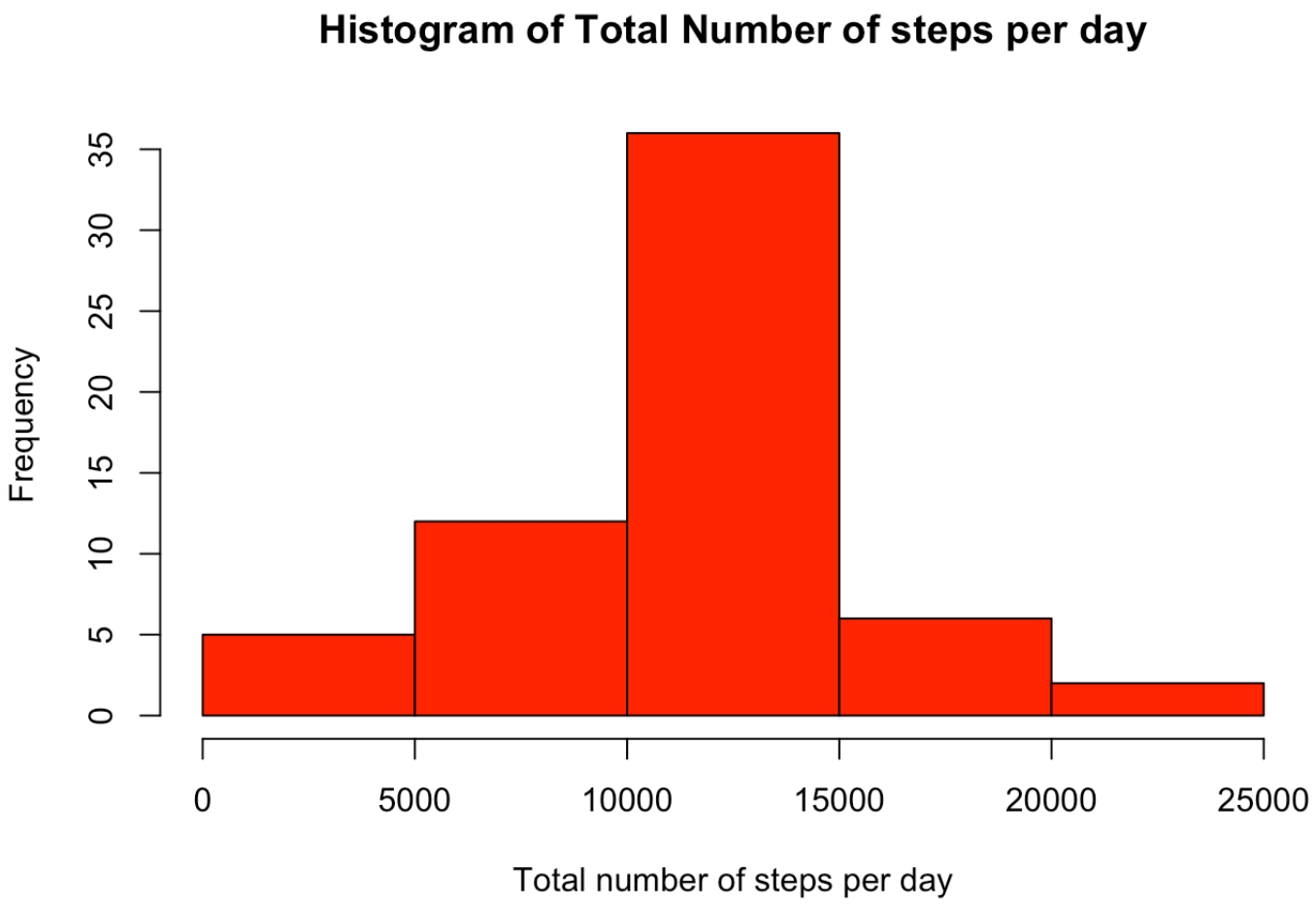
```
##    steps     date interval
##     2304        0        0
```

7. Are the difference in activity patterns between weekdays and weekends?

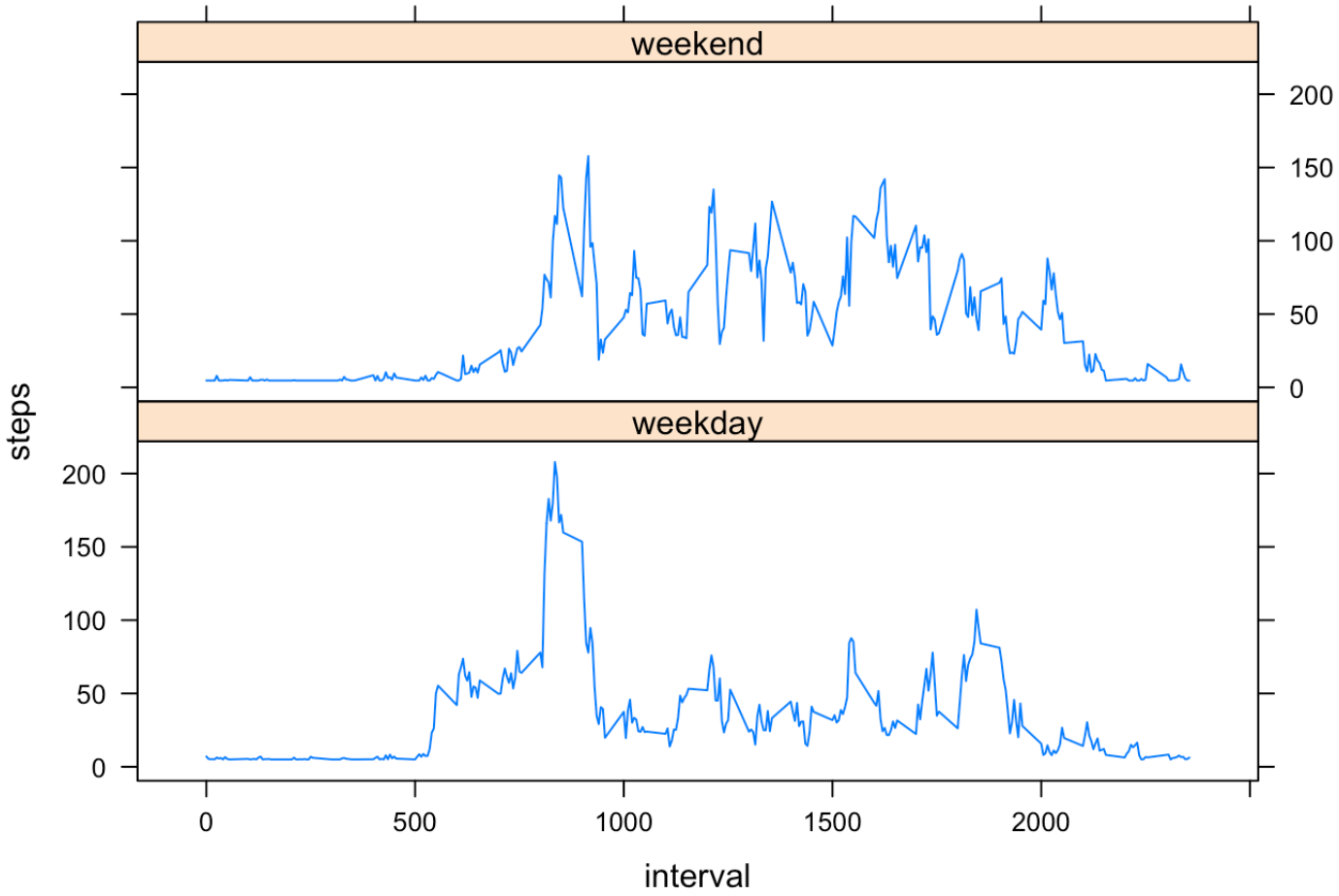
```
Ans =mean(data$steps,na.rm=T)
data2 = data
data2$steps[is.na(data2$steps)] = Ans
Byday = split (data2,data2$date)
Result = sapply(Byday, function(x) sum(x[,1]))
Result2 = data.frame(Result)
hist(Result2$Result, col='red', xlab= 'Total number of steps per day', main='Histogram of T
otal Number of steps per day')
data2$Weekday= weekdays(as.Date(data2$date))
data2$Weekday[(data2$Weekday=='Monday')] = 'weekday'
data2$Weekday[(data2$Weekday=='Tuesday')] = 'weekday'
data2$Weekday[(data2$Weekday=='Wednesday')] = 'weekday'
data2$Weekday[(data2$Weekday=='Thursday')] = 'weekday'
data2$Weekday[(data2$Weekday=='Friday')] = 'weekday'
data2$Weekday[(data2$Weekday=='Saturday')] = 'weekend'
data2$Weekday[(data2$Weekday=='Sunday')] = 'weekend'
data2$Weekday = factor(data2$Weekday)
library(plyr)
Newdata = ddply(data2,c('interval','Weekday'), function(x) apply(x[1],2,mean))
head(Newdata)
```

```
##   interval Weekday    steps
## 1        0 weekday 7.006569
## 2        0 weekend 4.672825
## 3        5 weekday 5.384347
## 4        5 weekend 4.672825
## 5       10 weekday 5.139902
## 6       10 weekend 4.672825
```

```
library(lattice)
```




```
xyplot(steps~interval | Weekday, data= Newdata, layout=c(1,2), type='l')
```


weekend
weekday
steps
interval
0
50
100
150
200
0
500
1000
1500
2000