Reproducible Research: Peer Assessment 1

Loading and preprocessing the data

```
library("plyr")
library("dplyr")
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:plyr':
##
##     arrange, count, desc, failwith, id, mutate, rename, summarise,
##     summarize
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library("lubridate")
```

```
##
## Attaching package: 'lubridate'
##
## The following object is masked from 'package:plyr':
##
##     here
```

```
library("lattice")
```

Read activity data file

```
where<-getwd()
print(where)
```

```
## [1] "C:/Users/hoffere/RepData_PeerAssessment1"
```

```
activitydata<-read.csv("activity.csv")
adstr<-str(activitydata)
```

```
## 'data.frame':    17568 obs. of  3 variables:
##  $ steps   : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ date    : Factor w/ 61 levels "10/1/2012","10/10/2012",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ interval: int  0 5 10 15 20 25 30 35 40 45 ...
```

```
adhead<-head(activitydata)
print(adstr)
```

```
## NULL
```

```
print(adhead)
```

```
##   steps       date interval
## 1    NA 10/1/2012        0
## 2    NA 10/1/2012        5
## 3    NA 10/1/2012       10
## 4    NA 10/1/2012       15
## 5    NA 10/1/2012       20
## 6    NA 10/1/2012       25
```

What is mean total number of steps taken per day?

sum number steps per day

```
StepsbyDate <-aggregate(activitydata$steps ~ activitydata$date, data=activitydata, FUN=su
m)
colnames(StepsbyDate)<-c("date","steps")
maxSBD<-range(StepsbyDate$steps)
maxval<-maxSBD[2]+5000
breaks= seq(5000,maxval, by=5000)
```

# plot histogram based on steps per day

```
SBD<-StepsbyDate$steps
SBD.cut = cut(SBD,breaks,right=FALSE)
SBD.cut
```

```
##  [1] [5e+03,1e+04)  [1e+04,1.5e+04) [1.5e+04,2e+04) [1e+04,1.5e+04)
##  [5] [1.5e+04,2e+04) [1e+04,1.5e+04) [1.5e+04,2e+04) [1e+04,1.5e+04)
##  [9] [1e+04,1.5e+04) [1e+04,1.5e+04) <NA>            [1e+04,1.5e+04)
## [13] [5e+03,1e+04)  [1e+04,1.5e+04) [5e+03,1e+04)  [5e+03,1e+04)
## [17] <NA>            [5e+03,1e+04)  [1e+04,1.5e+04) [1e+04,1.5e+04)
## [21] [5e+03,1e+04)  [1e+04,1.5e+04) [5e+03,1e+04)  [1.5e+04,2e+04)
## [25] [1e+04,1.5e+04) [1e+04,1.5e+04) [1.5e+04,2e+04) [1e+04,1.5e+04)
## [29] [1e+04,1.5e+04) [1e+04,1.5e+04) [1e+04,1.5e+04) [5e+03,1e+04)
## [33] <NA>            [5e+03,1e+04)  [1e+04,1.5e+04) [1.5e+04,2e+04)
## [37] [5e+03,1e+04)  [1e+04,1.5e+04) <NA>            [1e+04,1.5e+04)
## [41] [2e+04,2.5e+04) [2e+04,2.5e+04) [1e+04,1.5e+04) [1e+04,1.5e+04)
## [45] [1e+04,1.5e+04) [1e+04,1.5e+04) [1e+04,1.5e+04) [5e+03,1e+04)
## [49] [1e+04,1.5e+04) [1e+04,1.5e+04) [5e+03,1e+04)  [1e+04,1.5e+04)
## [53] <NA>
## 4 Levels: [5e+03,1e+04) [1e+04,1.5e+04) ... [2e+04,2.5e+04)
```

```
SBD.freq=table(SBD.cut)
hist(SBD,main="Total Number Steps Taken per Day")
```

Total Number Steps Taken per Day


```
SBDmean<-mean(SBD)
SBDmedian<-median(SBD)
print(SBDmean)
```

```
## [1] 10766.19
```
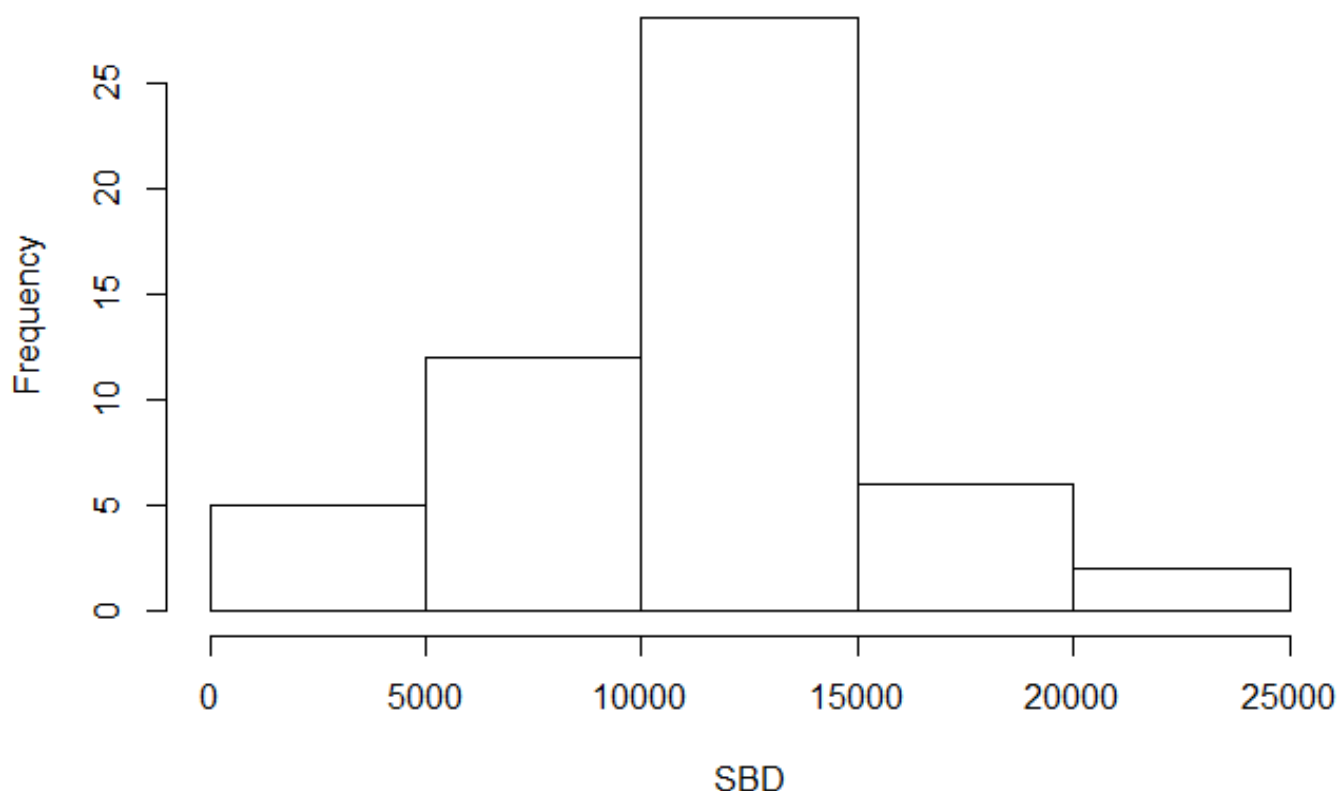
```
print(SBDmedian)
```

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

What is the average daily activity pattern?

```
StepsbyInterval <-aggregate(activitydata$steps ~ activitydata$interval, data=activitydat
a,  FUN=ave)
colnames(StepsbyInterval)<-c("interval","steps")
SBI<-data.frame()
SBI<-cbind(StepsbyInterval$interval,StepsbyInterval$steps[,1])
plot(x=SBI[,1],y=SBI[,2],type="l", main="Average Daily Steps/Activity",xlab="Interval",yl
ab="Average number of steps")
```

Average Daily Steps/Activity
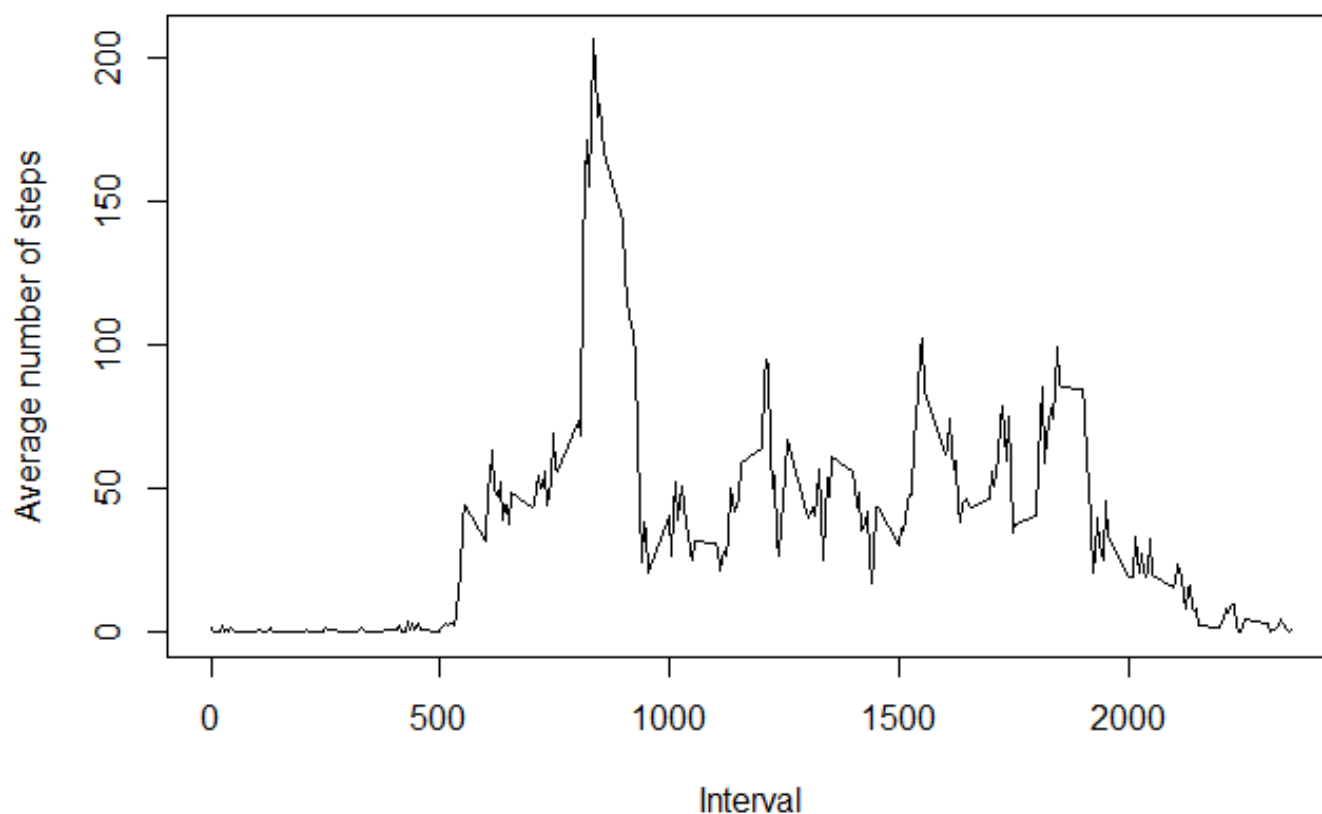
determine interval with max steps

```
MaxSteps<-max(SBI[,2])
MaxInterval<-SBI[which.max(SBI[,2])]
print(MaxSteps)
```

```
## [1] 206.1698
```

```
print(MaxInterval)
```

```
## [1] 835
```

Imputing missing values

determine NA values for Steps, data and interval

```
sum(is.na(activitydata$steps))
```

```
## [1] 2304
```

```
sum(is.na(activitydata$date))
```

```
## [1] 0
```

```
sum(is.na(activitydata$interval))
```

```
## [1] 0
```

```
total<-sum(is.na(activitydata$steps))+sum(is.na(activitydata$date))+sum(is.na(activitydat
a$interval))
```

Total missing values =

```
print(total)
```

```
## [1] 2304
```

create copy of the dataset and fill in NA values with the overall data mean determined above.

```
newval<-0
activitydata2<-activitydata
allobs<-length(activitydata2$steps)
DailyAvg <- activitydata2 %>% filter(steps >= 0) %>%
    group_by(date) %>% summarize(total = sum(steps), avg = mean(steps))
for (i in 1:allobs) {
  if (is.na(activitydata2$steps[i])){
  activitydata2$steps[i]<-DailyAvg[,2]
  } else {
    }
  }
```

recreate the histogram with the new dataset

```
StepsbyDate <-aggregate(activitydata$steps ~ activitydata$date, data=activitydata, FUN=su
m)
colnames(StepsbyDate)<-c("date","steps")
maxSBD<-range(StepsbyDate$steps)
maxval<-maxSBD[2]+5000
breaks= seq(5000,maxval, by=5000)
SBD2<-StepsbyDate$steps
SBD2.cut = cut(SBD,breaks,right=FALSE)
SBD2.freq=table(SBD.cut)
hist(SBD,main="Dataset with Mean Replacement for NA - Total Number Steps Taken per Day")
```


Dataset with Mean Replacement for NA - Total Number Steps Taken per [

```
SBDmean<-mean(SBD)
SBDmedian<-median(SBD)
print(SBDmean)
```

```
## [1] 10766.19
```
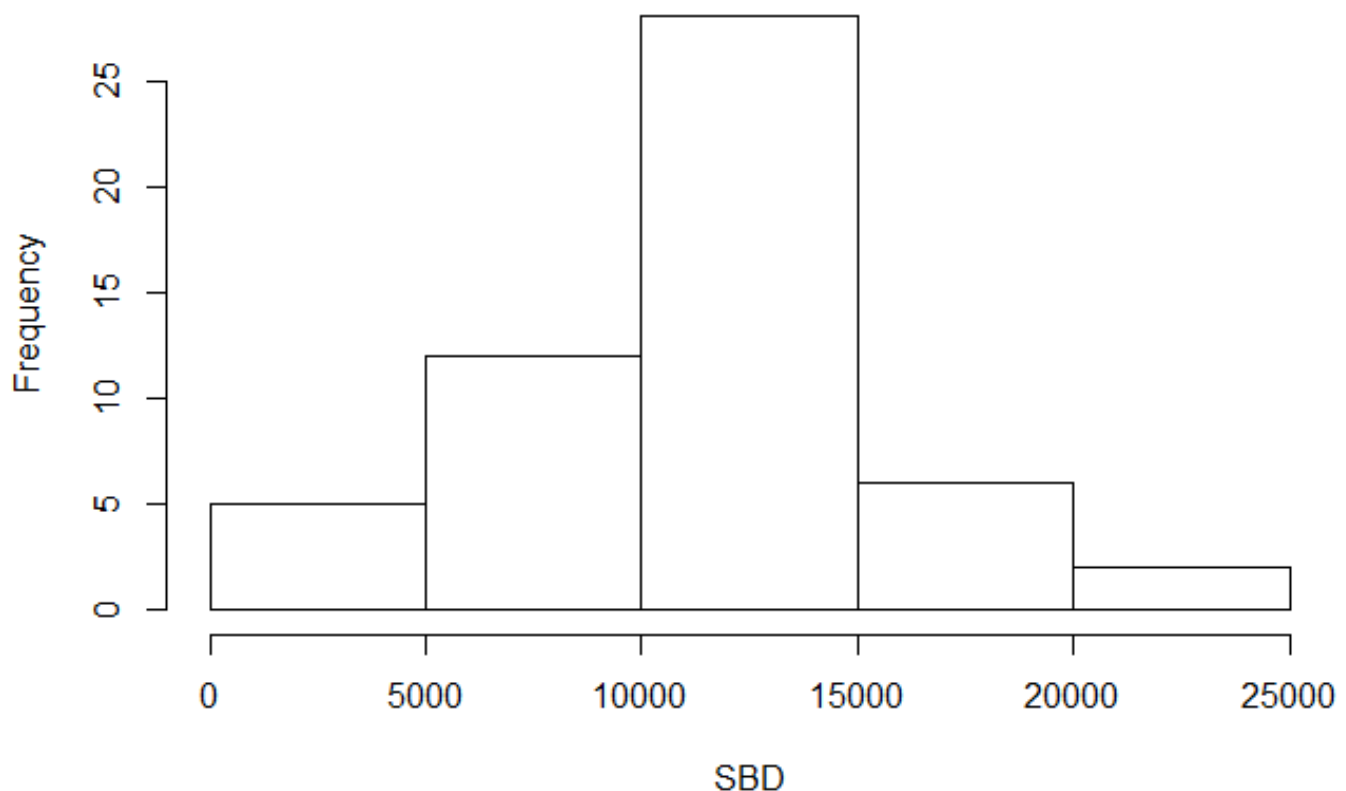
```
print(SBDmedian)
```

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

Are there differences in activity patterns between weekdays and weekends?

separate weekdays and weekend days.

```
activitydata$DayoWeek<-weekdays(as.Date(activitydata$date))

activitydata$DayType<-with(activitydata,ifelse(DayoWeek == "Saturday" | DayoWeek == "Satu
rday", "weekend", "weekday"))
```

plot the steps by time series weekend vs weekday

```
group_by(activitydata,DayType,interval)
```

```
## Source: local data frame [17,568 x 5]
## Groups: DayType, interval
##
##      steps      date interval  DayoWeek DayType
## 1       NA 10/1/2012        0 Wednesday weekday
## 2       NA 10/1/2012        5 Wednesday weekday
## 3       NA 10/1/2012       10 Wednesday weekday
## 4       NA 10/1/2012       15 Wednesday weekday
## 5       NA 10/1/2012       20 Wednesday weekday
## 6       NA 10/1/2012       25 Wednesday weekday
## 7       NA 10/1/2012       30 Wednesday weekday
## 8       NA 10/1/2012       35 Wednesday weekday
## 9       NA 10/1/2012       40 Wednesday weekday
## 10      NA 10/1/2012       45 Wednesday weekday
## ..     ...       ...      ...       ...     ...
```

```
xyplot(activitydata$steps ~ activitydata$interval | activitydata$DayType, xlab = "Interva
l", ylab="Average # of Steps",type="l", layout=c(1,2))
```