

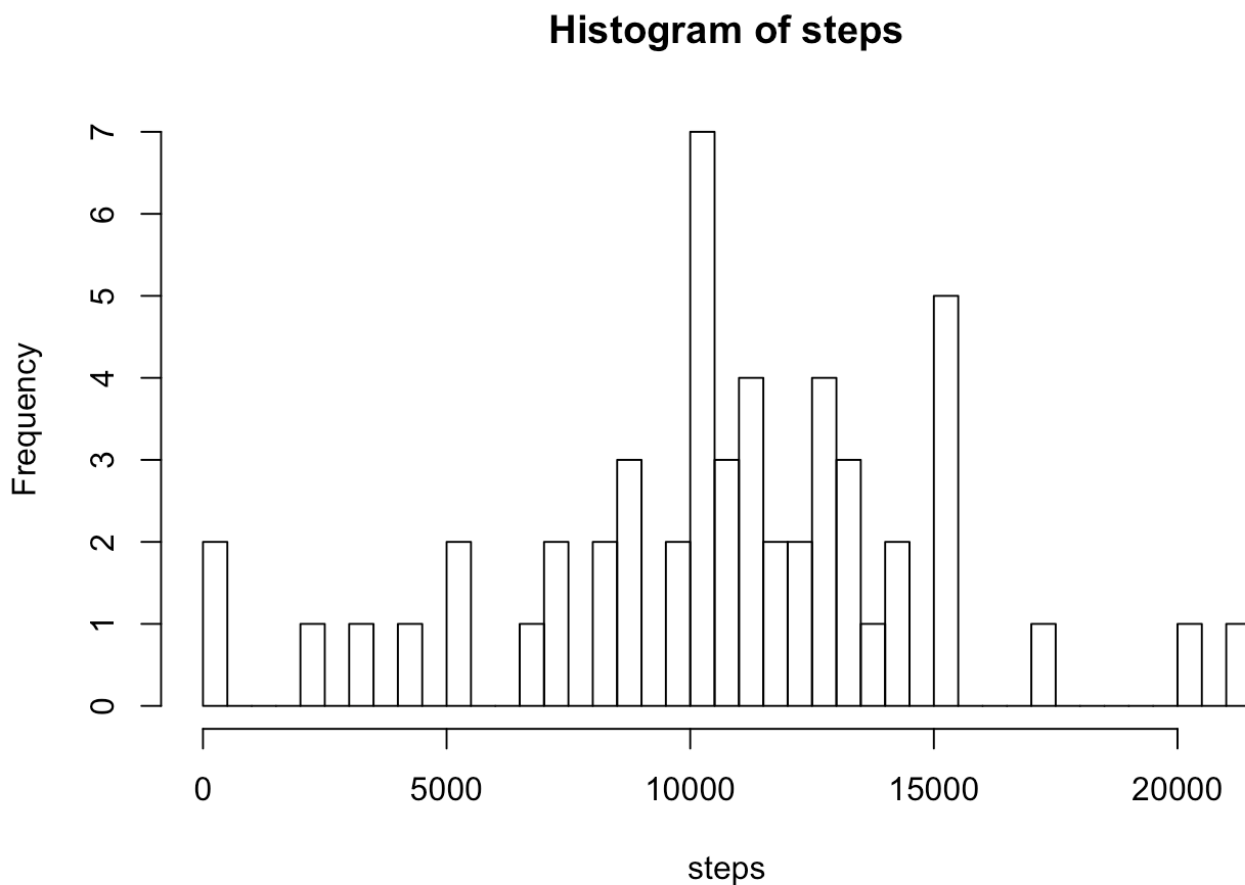
Reproducible Research Assignment 1

1. Code for reading in the dataset and/or processing the data

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

2. Histogram of the total number of steps taken each day

```
steps <- tapply(data$steps, data$date, sum)  
hist(steps, breaks=50)
```



3. Mean and median number of steps taken each day

```
mean(steps, na.rm=TRUE)
```

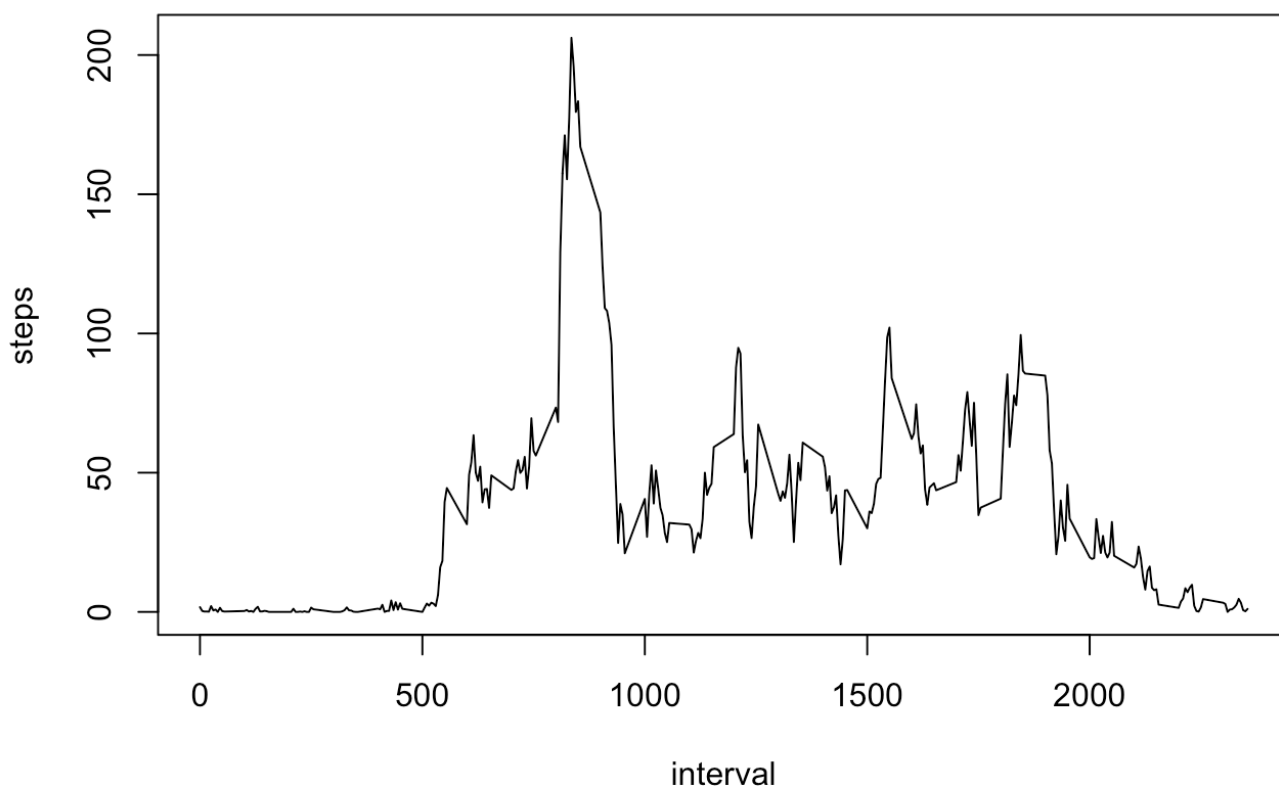
```
## [1] 10766.19
```

```
median(steps, na.rm=TRUE)
```

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

4. Time series plot of the average number of steps taken

```
aveSteps <- aggregate(x=list(steps=data$steps),  
                       by=list(interval=data$interval),  
                       mean,  
                       na.rm=TRUE)  
plot(aveSteps, type='l')
```



5. The 5-minute interval that, on average, contains the maximum number of steps

```
aveSteps[which.max(aveSteps$steps), ]
```

```
##      interval      steps
## 104         835 206.1698
```

6. Code to describe and show a strategy for imputing missing data

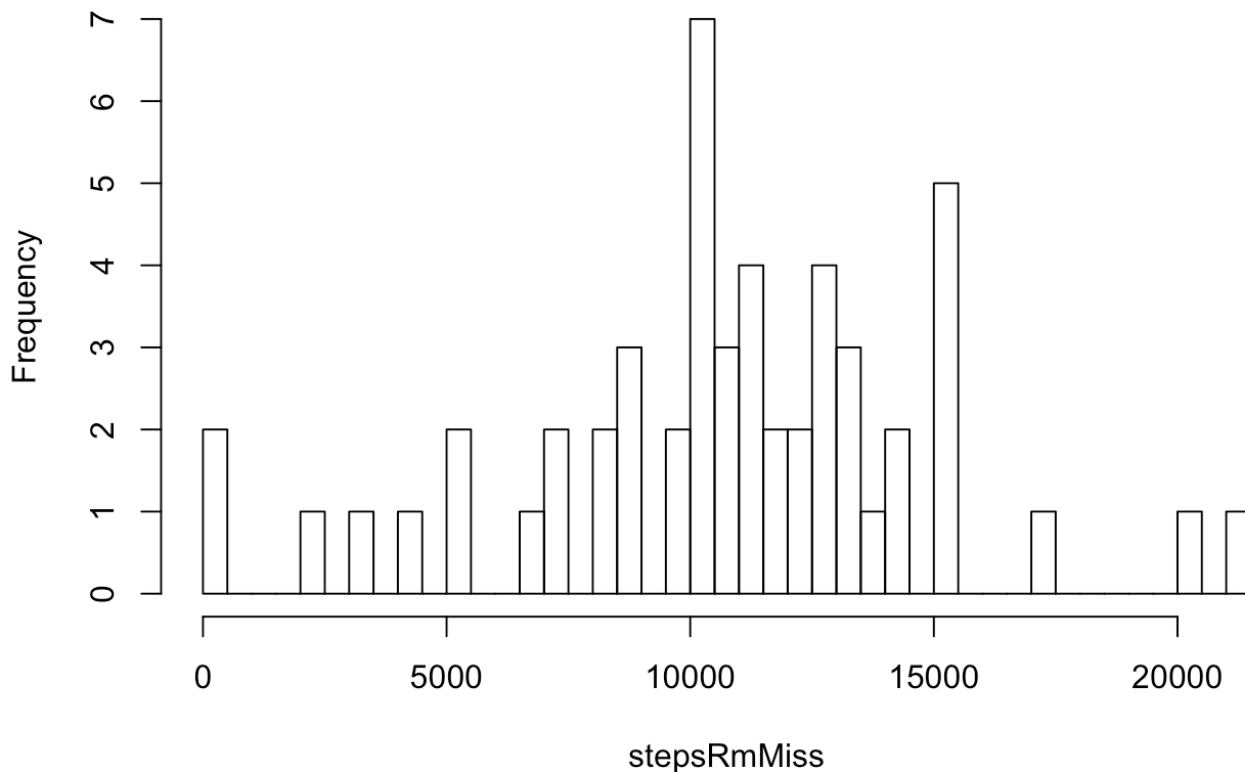
```
missingData <- is.na(data$steps)
table(missingData)
```

```
## missingData
## FALSE  TRUE
## 15264  2304
```

7. Histogram of the total number of steps taken each day after missing values are imputed

```
dataRmMiss <- data[!missingData,]
stepsRmMiss <- tapply(dataRmMiss$steps, dataRmMiss$date, sum)
hist(stepsRmMiss, breaks=50)
```

Histogram of stepsRmMiss



8. Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

```
dataRmMiss$day <- 'weekdays'
dataRmMiss$day[weekdays(as.Date(dataRmMiss$date), abb=TRUE) %in% c('Sat', 'Sun')] <- 'weekends'
aveStepsWeek <- aggregate(x=list(steps=dataRmMiss$steps),
                           by=list(interval=dataRmMiss$interval, day=dataRmMiss$day),
                           mean,
                           na.rm=TRUE)

library(lattice)
xyplot(steps ~ interval | day, data=aveStepsWeek, type='l',
       grid=TRUE, layout=c(1,2),
       ylab='steps', xlab='intervals')
```

