Project 1

Stephanie Bruce

Reproducible Research - Course Project 1

This R Markdown document include my code for Project 1.

```
library(readr)
data <- read_csv("/home/rstudio/Reproducible_Research/week2/activity.csv")
```

```
## Parsed with column specification:
## cols(
##   steps = col_double(),
##   date = col_date(format = ""),
##   interval = col_double()
## )
```

Part 1. What is mean total number of steps taken per day

```
total_steps <- aggregate(x=data$steps, FUN=sum, by=list(Group.date = data$date), na.rm=TRUE)
```

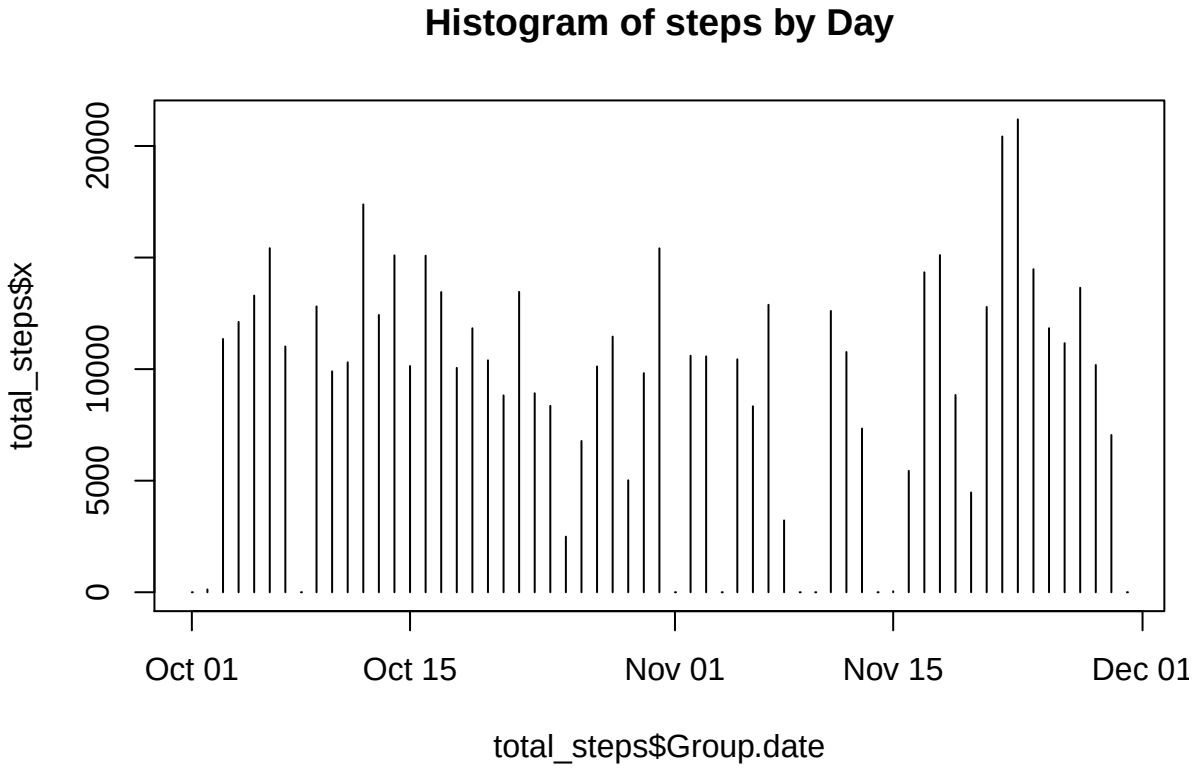
The total number of steps taken per day is:

```
print(total_steps)
```

```
##    Group.date     x
## 1  2012-10-01     0
## 2  2012-10-02   126
## 3  2012-10-03 11352
## 4  2012-10-04 12116
## 5  2012-10-05 13294
## 6  2012-10-06 15420
## 7  2012-10-07 11015
## 8  2012-10-08     0
## 9  2012-10-09 12811
## 10 2012-10-10  9900
## 11 2012-10-11 10304
## 12 2012-10-12 17382
## 13 2012-10-13 12426
## 14 2012-10-14 15098
## 15 2012-10-15 10139
## 16 2012-10-16 15084
## 17 2012-10-17 13452
## 18 2012-10-18 10056
## 19 2012-10-19 11829
## 20 2012-10-20 10395
## 21 2012-10-21  8821
## 22 2012-10-22 13460
```

```
## 23 2012-10-23  8918
## 24 2012-10-24  8355
## 25 2012-10-25  2492
## 26 2012-10-26  6778
## 27 2012-10-27 10119
## 28 2012-10-28 11458
## 29 2012-10-29  5018
## 30 2012-10-30  9819
## 31 2012-10-31 15414
## 32 2012-11-01     0
## 33 2012-11-02 10600
## 34 2012-11-03 10571
## 35 2012-11-04     0
## 36 2012-11-05 10439
## 37 2012-11-06  8334
## 38 2012-11-07 12883
## 39 2012-11-08  3219
## 40 2012-11-09     0
## 41 2012-11-10     0
## 42 2012-11-11 12608
## 43 2012-11-12 10765
## 44 2012-11-13  7336
## 45 2012-11-14     0
## 46 2012-11-15    41
## 47 2012-11-16  5441
## 48 2012-11-17 14339
## 49 2012-11-18 15110
## 50 2012-11-19  8841
## 51 2012-11-20  4472
## 52 2012-11-21 12787
## 53 2012-11-22 20427
## 54 2012-11-23 21194
## 55 2012-11-24 14478
## 56 2012-11-25 11834
## 57 2012-11-26 11162
## 58 2012-11-27 13646
## 59 2012-11-28 10183
## 60 2012-11-29  7047
## 61 2012-11-30     0
```

```r
plot(total_steps$Group.date, total_steps$x, type = "h")
title(main = "Histogram of steps by Day")
```



The mean and median number of steps per day are:

```
mean_steps <- mean(total_steps$x)
median_steps <- median(total_steps$x)
print(mean_steps)
```

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

```
print(median_steps)
```

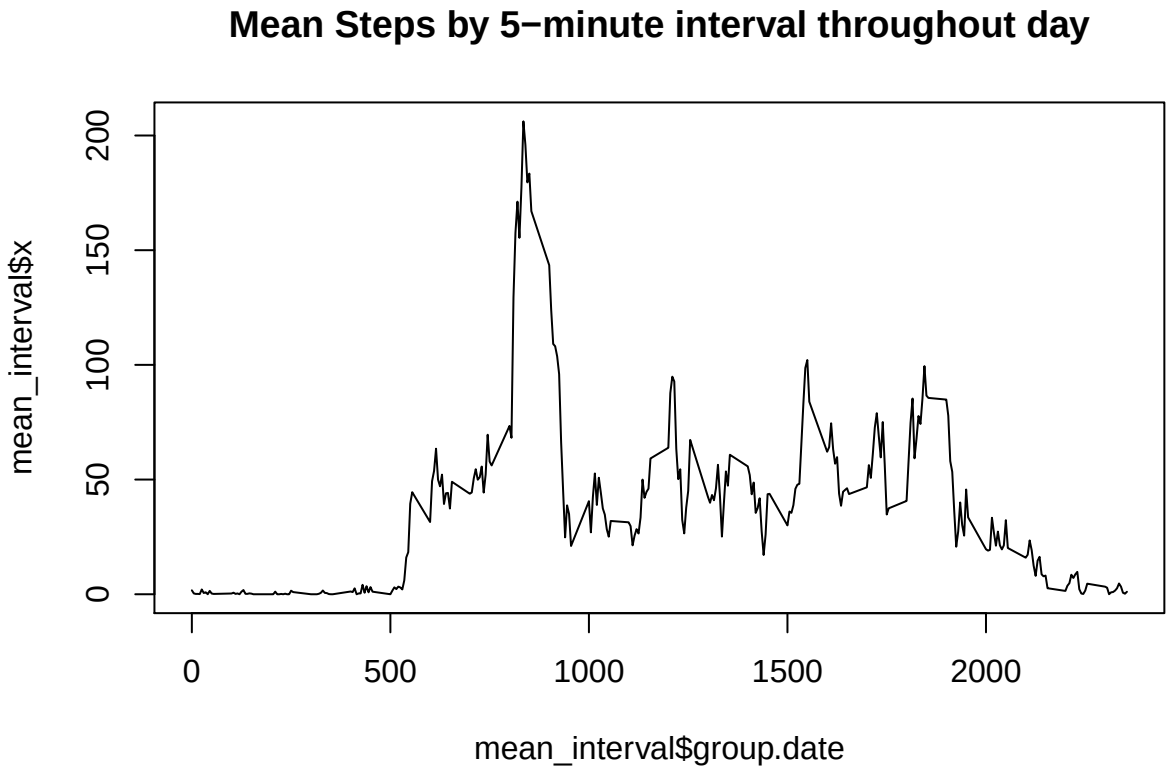
```
## [1] 10395
```

Part 2. What is the average daily activity pattern

```
interval_totals <- aggregate(x=data$steps, FUN=sum, by=list(group.date = data$interval), na.rm=TRUE)
mean_interval <-aggregate(x=data$steps, FUN=mean, by=list(group.date = data$interval), na.rm=TRUE)
```

Time series plot by interval

```
plot(mean_interval$group.date, mean_interval$x, type = "l")
title(main = "Mean Steps by 5-minute interval throughout day")
```



The average maximum number of steps is at interval 825 and the max steps is:

```
print(max(mean_interval$x))
```

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

The daily average number of steps is:

```
daily_mean <- sum(mean_interval$x)
```

Part 3. Imputing missing values

Calculate and report the total number of missing values in the dataset:

```
count_nas <- aggregate(x=data$steps, function(x) {sum(is.na(x))}, by=list(Group.date = data$date))
print(sum(count_nas$x))
```

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

Fill in all the missing values in the dataset (I split the daily average into each interval):

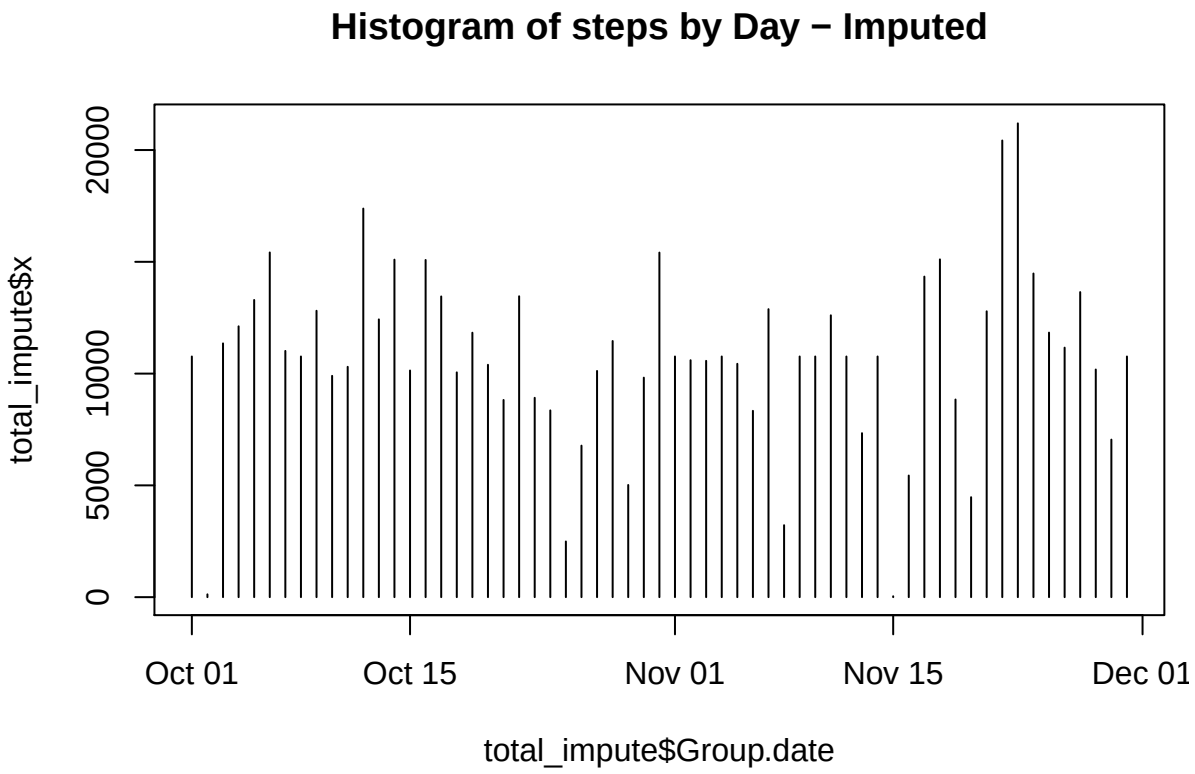
```
data2 <- data
data2$impute <- data$steps
data2$impute[is.na(data2$impute)] <- daily_mean/length(mean_interval$x)
```

Create a new dataset that is equal to the original dataset (new dataset = data2):

```
total_impute <- aggregate(x=data2$impute, FUN=sum, by=list(Group.date = data$date))
```

Histogram:

```
plot(total_impute$Group.date, total_impute$x, type = "h")
title(main = "Histogram of steps by Day - Imputed")
```



Mean and Median steps with imputed data:

```
mean_steps_imp <- mean(total_impute$x)
median_steps_imp <- median(total_impute$x)
print(mean_steps_imp)
```

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

```
print(median_steps_imp)
```

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

Part 4. Are there differences in activity patterns between weekdays and weekends?

Create a new factor

```
dataDoW <- weekdays(data2$date)
weekdays <- c('Monday', 'Tuesday', 'Wednesday', 'Thursday','Friday')
data2$WDay <- factor((weekdays(data$date) %in% weekdays), levels = c(TRUE, FALSE), labels= c('weekday',
```

Finishing defining weekday vs weekend

```
data2$weekday = ifelse(data2$WDay == "weekday", 1,0)
data2$weekend = ifelse(data2$WDay == "weekend", 1,0)
```

Adding information to dataset:

```
data2$wkdy_steps = data2$impute*data2$weekday
data2$wked_steps = data2$impute*data2$weekend
```

Combining information

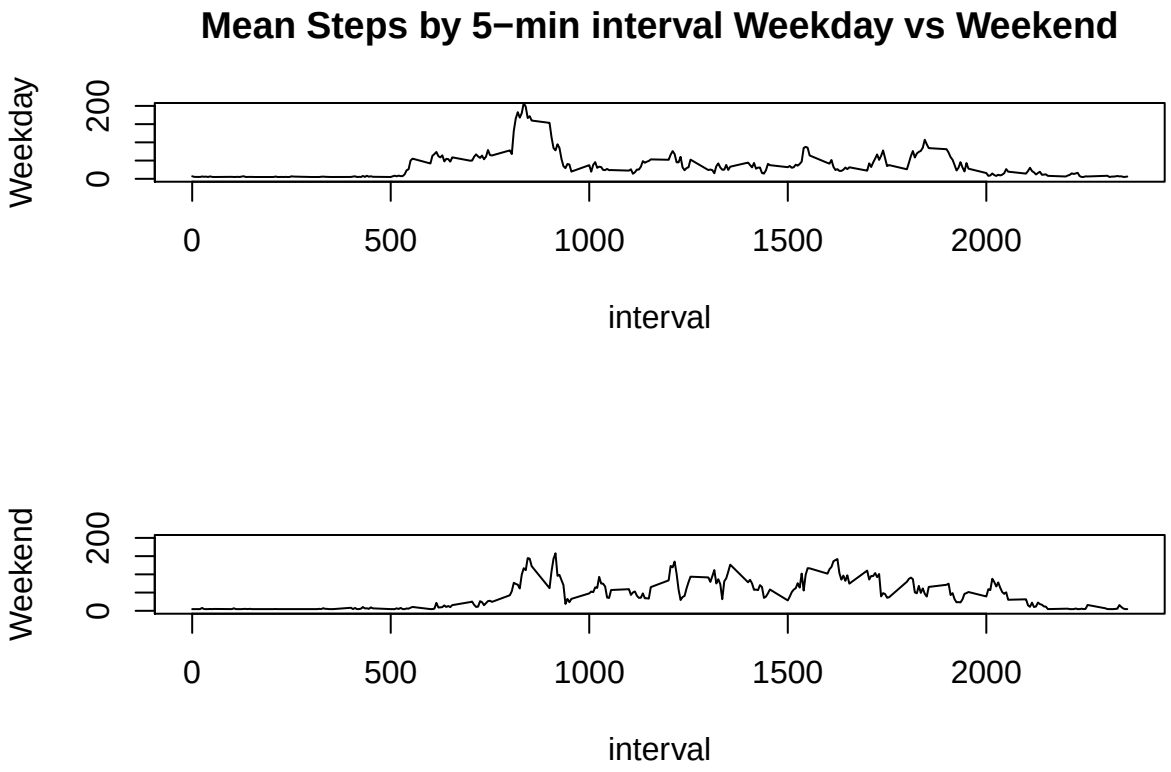
```
interval_wkday <- aggregate(x=data2$wkdy_steps, FUN=sum, by=list(group.date = data2$interval))
interval_wkend <-aggregate(x=data2$wked_steps, FUN=sum, by=list(group.date = data2$interval))
```

Remove 0 out of average from weekend/weekday:

```
interval_wkd <- interval_wkday$x/(sum(data2$weekday)/288)
interval_wke <- interval_wkend$x/(sum(data2$weekend)/288)
```

Panel Plot

```
par(mfrow=c(2,1))
plot(interval_wkday$group.date, interval_wkd, type = "l", ylim = c(0,200), xlab = "interval", ylab = "W
title("Mean Steps by 5-min interval Weekday vs Weekend")
plot(interval_wkday$group.date, interval_wke, type = "l", ylim=c(0,200), xlab="interval", ylab = "Weeke
```



The End