

RR Peer Assessment 1

Jana

Wednesday, November 11, 2015

```
library("ggplot2")
```

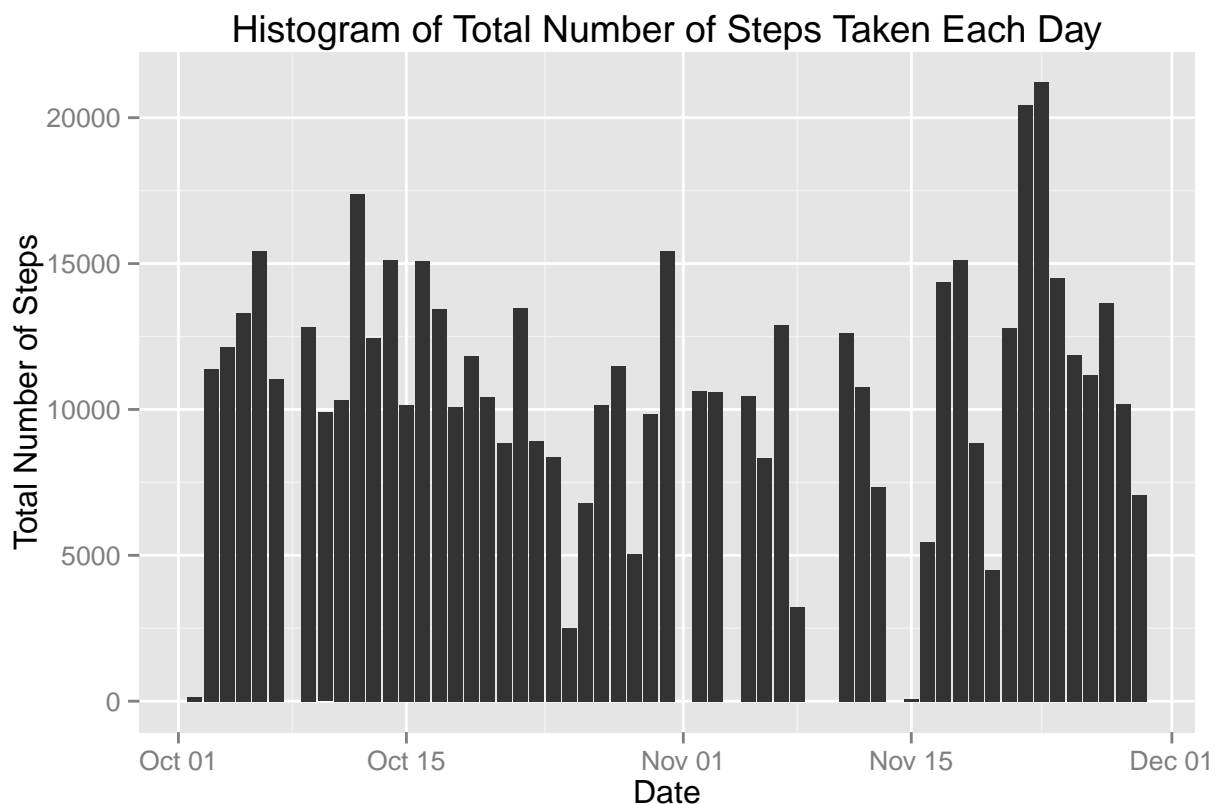
```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

Load Data

```
data <- read.table('activity.csv',sep = ",",header = TRUE, na.strings = "NA",colClasses = c('integer','D
```

Steps taken per day

```
new.data <- na.omit(data)
total.steps <- tapply(new.data$steps, new.data$date, FUN = sum)
plot1 <- ggplot(new.data, aes(date, steps)) + geom_bar(stat = "identity",binwidth = .5) +
  labs(title = "Histogram of Total Number of Steps Taken Each Day",x = "Date", y = "Total Number of Steps")
print(plot1)
```



Steps taken per day (Mean and Median)

```
mean(total.steps)
```

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

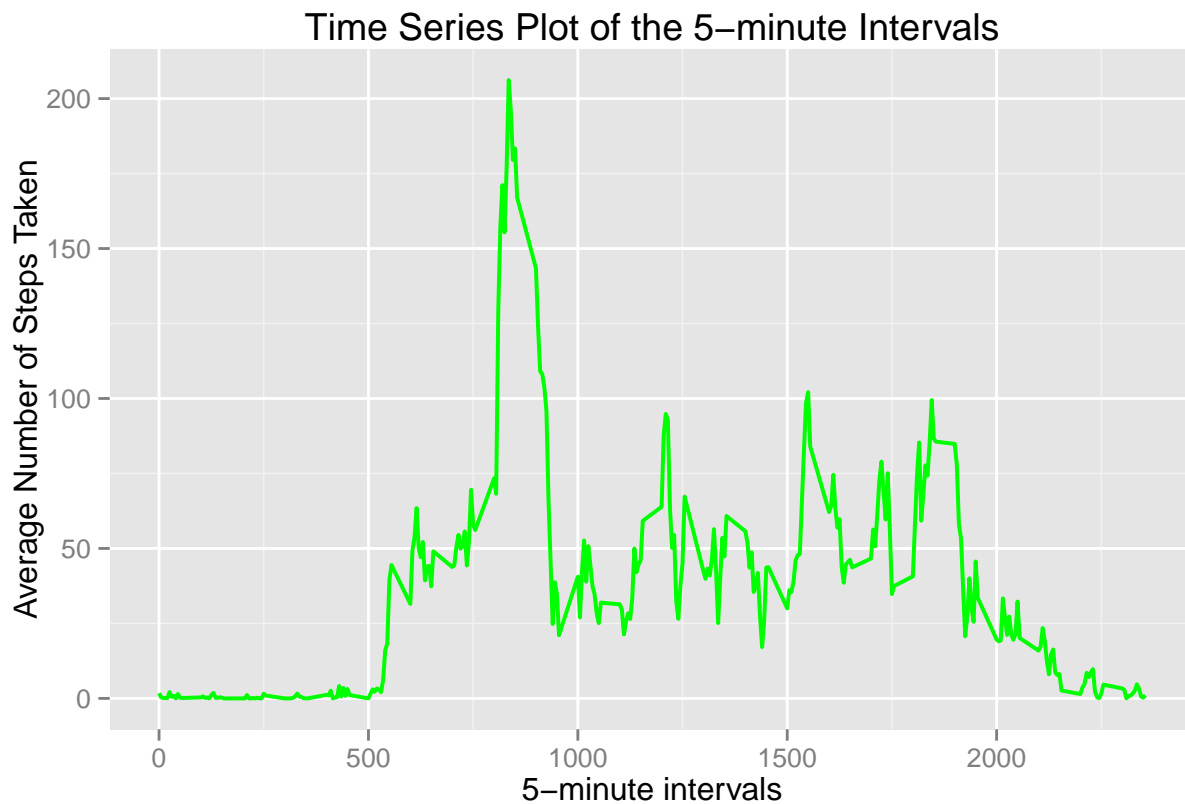
```
median(total.steps)
```

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

Average daily Activity pattern

```
averages <- aggregate(new.data$steps, list(interval = as.numeric(as.character(new.data$interval))), FUN = mean, na.rm = TRUE)
names(averages)[2] <- "Avg.Steps"

plot2 <- ggplot(averages, aes(interval, Avg.Steps)) + geom_line(color = "green", size = 0.7) + labs(title = "Average daily Activity pattern", x = "5-minute intervals", y = "Average Number of Steps Taken")
print(plot2)
```



```
averages[averages$Avg.Steps == max(averages$Avg.Steps),]
```

```
##      interval Avg.Steps
## 104         835  206.1698
```

```
sum(!complete.cases(data))
```

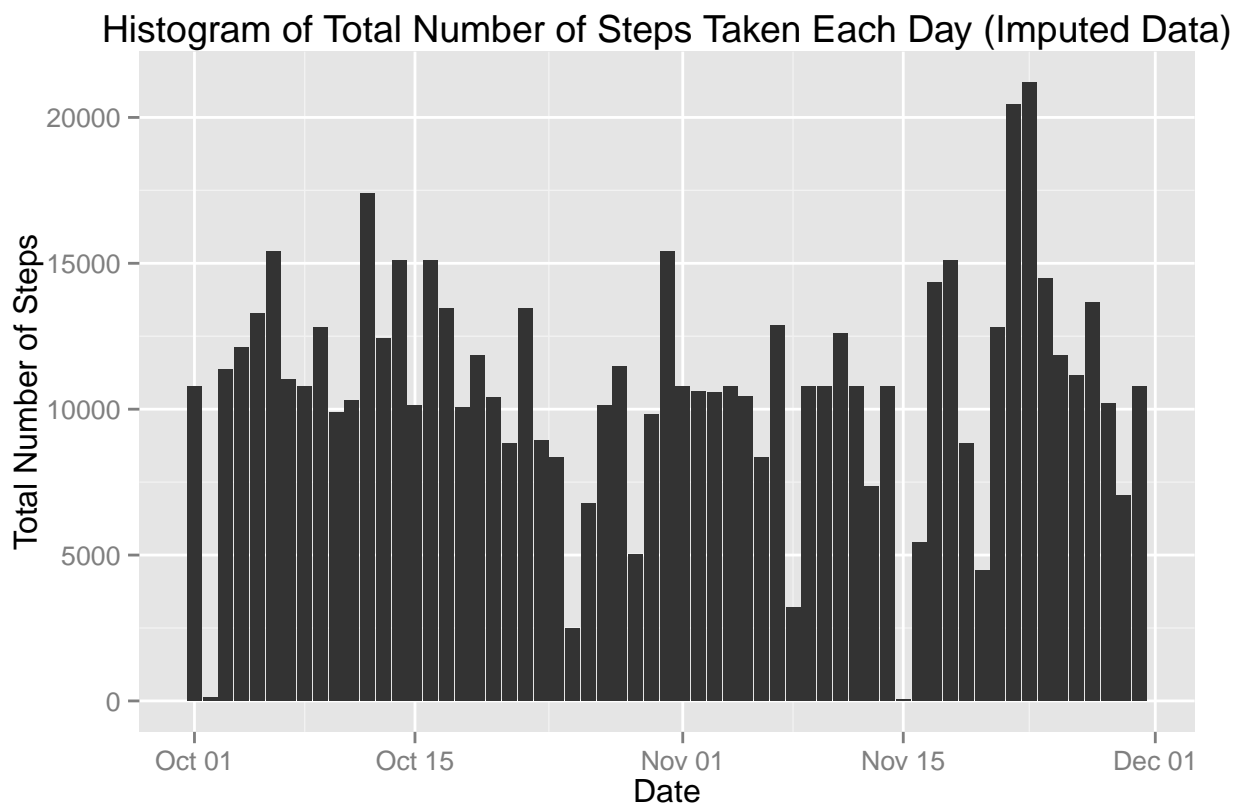
```
## [1] 2304
```

Imputing missing Data

```
impData <- data
for (i in 1:nrow(impData)) {
  if (is.na(impData$steps[i])) {
    impData$steps[i] <- averages[which(impData$interval[i] == averages$interval), ]$Avg.Steps
  }
}
sum(!complete.cases(impData))
```

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

```
plot3 <- ggplot(impData, aes(date, steps)) + geom_bar(stat = "identity", binwidth = .5) +
  labs(title = "Histogram of Total Number of Steps Taken Each Day (Imputed Data)", x = "Date", y =
print(plot3)
```



```
total.steps.impute <- tapply(impData$steps, impData$date, FUN = sum)
mean(total.steps.impute)
```

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

```
median(total.steps.impute)
```

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

```
impData$weekdays <- factor(format(impData$date, "%A"))  
levels(impData$weekdays)
```

```
## [1] "Friday"      "Monday"      "Saturday"    "Sunday"      "Thursday"    "Tuesday"  
## [7] "Wednesday"
```

Finding differences in activity patterns between weekdays and weekends

```
levels(impData$weekdays) <- list(weekday = c("Monday", "Tuesday",  
                                              "Wednesday",  
                                              "Thursday", "Friday"),  
                                weekend = c("Saturday", "Sunday"))  
levels(impData$weekdays)
```

```
## [1] "weekday" "weekend"
```

```
table(impData$weekdays)
```

```
##  
## weekday weekend  
##    12960    4608
```

```
new.averages <- aggregate(impData$steps,  
                          list(interval = as.numeric(as.character(impData$interval)),  
                              weekdays = impData$weekdays),  
                          FUN = "mean")  
names(new.averages)[3] <- "meanOfSteps"  
library(lattice)
```

```
## Warning: package 'lattice' was built under R version 3.1.3
```

```
plot4 <- xyplot(new.averages$meanOfSteps ~ new.averages$interval | new.averages$weekdays,  
               layout = c(1, 2), type = "l",  
               xlab = "Interval", ylab = "Number of steps")  
print(plot4)
```

