

Reproducible Research Course Project 1

Humin Xue

March 27, 2016

Introduction

This assignment makes use of data from a personal activity monitoring device. This device collects data at 5 minute intervals through out the day. The data consists of two months of data from an anonymous individual collected during the months of October and November, 2012 and include the number of steps taken in 5 minute intervals each day.

Reading and Processing the data

```
activity <- read.csv("C:/datascience/reprodice/activity.csv")
processed_activity <- activity[!is.na(activity$steps),]
```

calculate mean, median, total steps and histogram

```
daily_step <- aggregate(processed_activity$steps ~ processed_activity$date, FUN=sum, )
colnames(daily_step)<- c("Date", "Steps")
mean(daily_step$Steps)
```

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

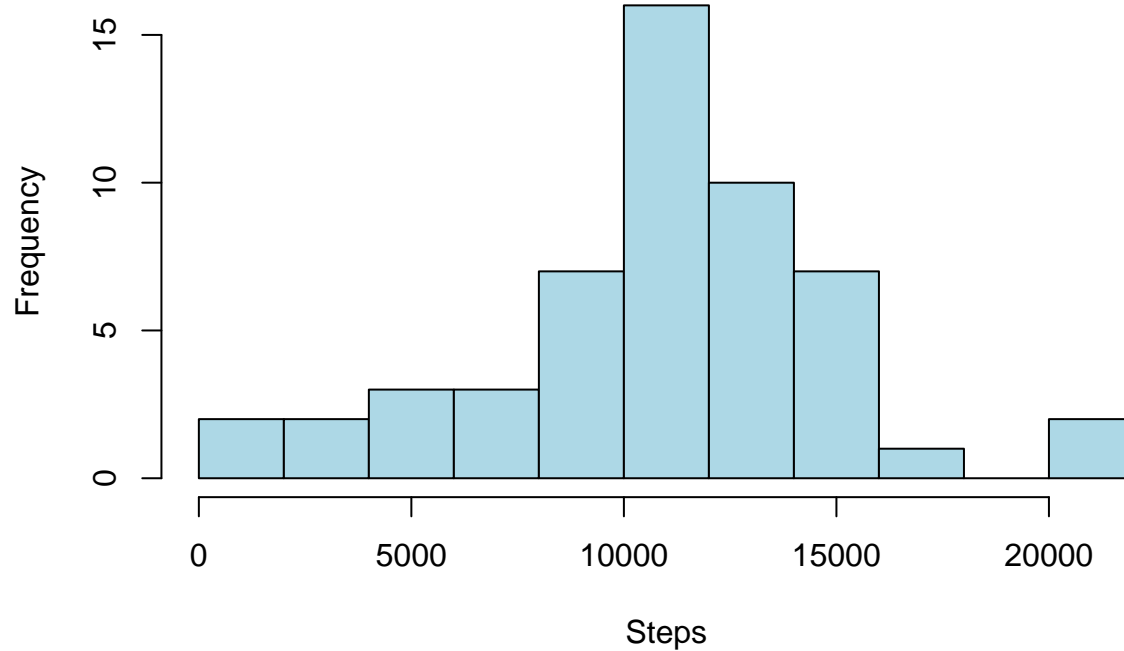
```
median(daily_step$Steps)
```

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

Histogram of the total number of steps taken each day

```
hist(daily_step$Steps, breaks=10, col="LIGHT BLUE", main="Histogram of activity", xlab="Steps", ylab = "Frequency")
```

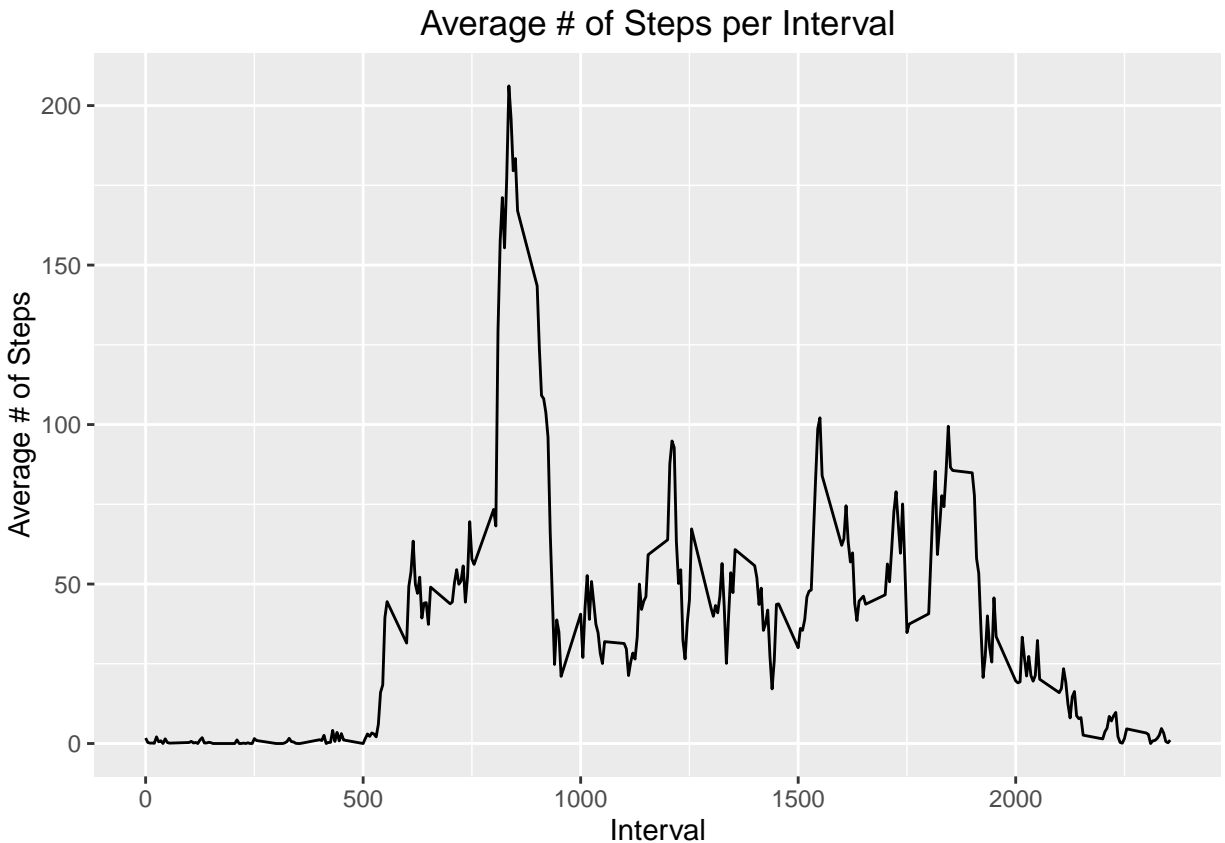
Histogram of activity



Time series plot average number of steps taken and interval

```
library(plyr)
library(ggplot2)
mean_step <- ddply(processed_activity, "interval", summarise, mean = mean(steps))

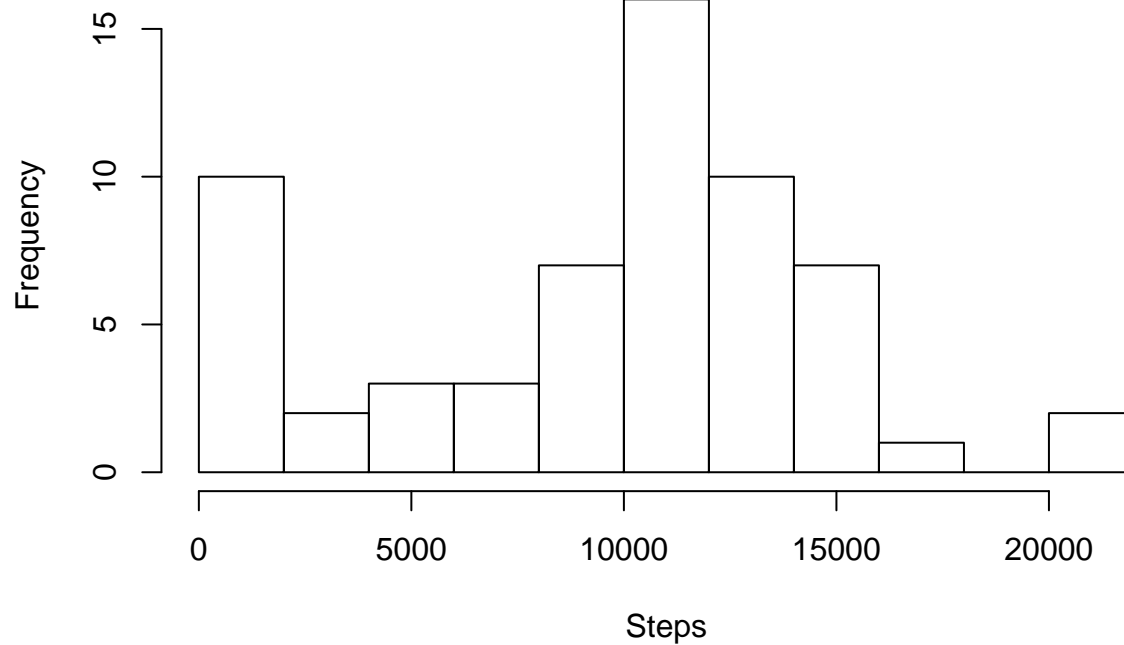
p <- ggplot(mean_step, aes(x=interval, y=mean), xlab = "Interval", ylab="Average # of Steps")
p + geom_line()+xlab("Interval")+ylab("Average # of Steps")+ggtitle("Average # of Steps per Interval")
```



Histogram of the total # of steps taken each day after missing values are imputed
replace NAs with median value

```
activity2 <- activity
activity2$steps[is.na(activity2$steps)] <- median(activity$steps, na.rm=TRUE)
activity2day <- aggregate(steps ~ date, data=activity2, sum, na.rm=TRUE)
hist(activity2day$steps, breaks=10, main="Total Steps per Day with NAs", xlab="Steps", ylab="Frequency")
```

Total Steps per Day with NAs



##comparing the average # of steps taken across weekdays and weekends

```
library(lattice)
processed_activity$date <- as.Date(processed_activity$date)
processed_activity$dayname <- weekdays(processed_activity$date)
processed_activity$daytype <- ifelse(processed_activity$dayname %in% c("Saturday", "Sunday"), "Weekend", "Weekday")

activity2_steps <- ddply(processed_activity, .(interval, daytype), summarize, avg=mean(steps))
xyplot(avg ~ interval | daytype, data=activity2_steps, type="l", layout = c(2,1), main="Comparing Average Steps")
```

Comparing Average Steps per Interval between weekday and weekend

