Reproducible Research Project 1

September 6, 2017

Loading and Preprocessing Data

```r
library(ggplot2); library(dplyr)

## Warning: package 'ggplot2' was built under R version 3.3.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

zipF <- file.choose()
unzip(zipF)
activData <- read.csv(file.choose(), header=T)

activ2 <- subset(activData, steps >= 0)
activNA <- subset(activData, is.na(steps))
```
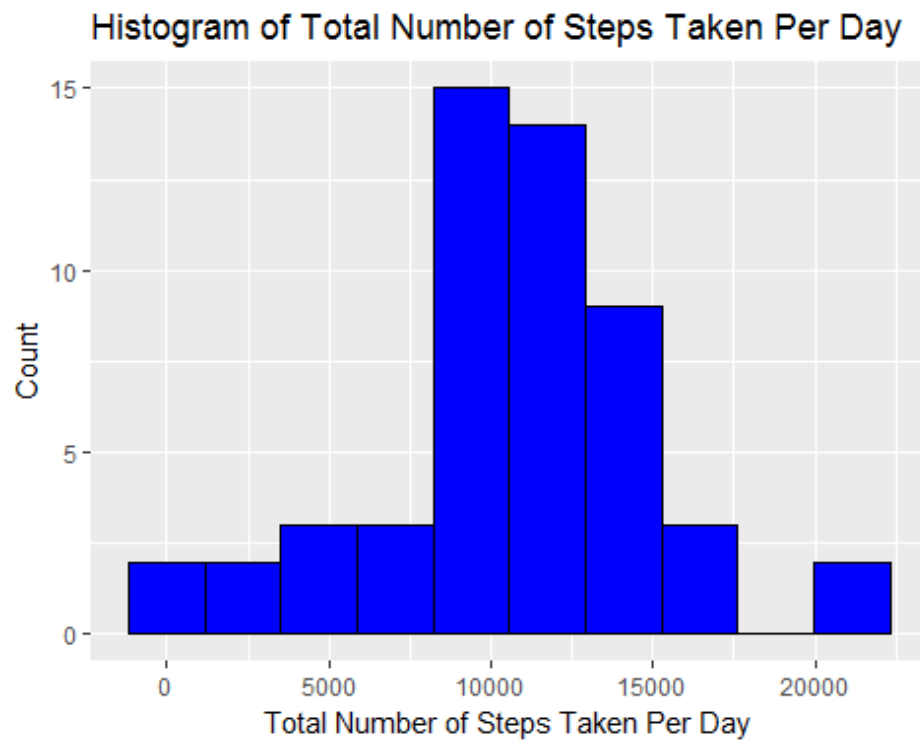
What is mean total number of steps taken per day?

```r
dateSum <- activ2 %>% group_by(date) %>% summarize(totStep = sum(steps))

ggplot(dateSum, aes(x=totStep))+
  geom_histogram(color="black", fill="blue", bins=10)+
  xlab("Total Number of Steps Taken Per Day")+
  ylab("Count")+
  ggtitle("Histogram of Total Number of Steps Taken Per Day")
```

Histogram of Total Number of Steps Taken Per Day


```
mean(dateSum$totStep) # 10766.19

## [1] 10766.19

median(dateSum$totStep) # 10765

## [1] 10765
```
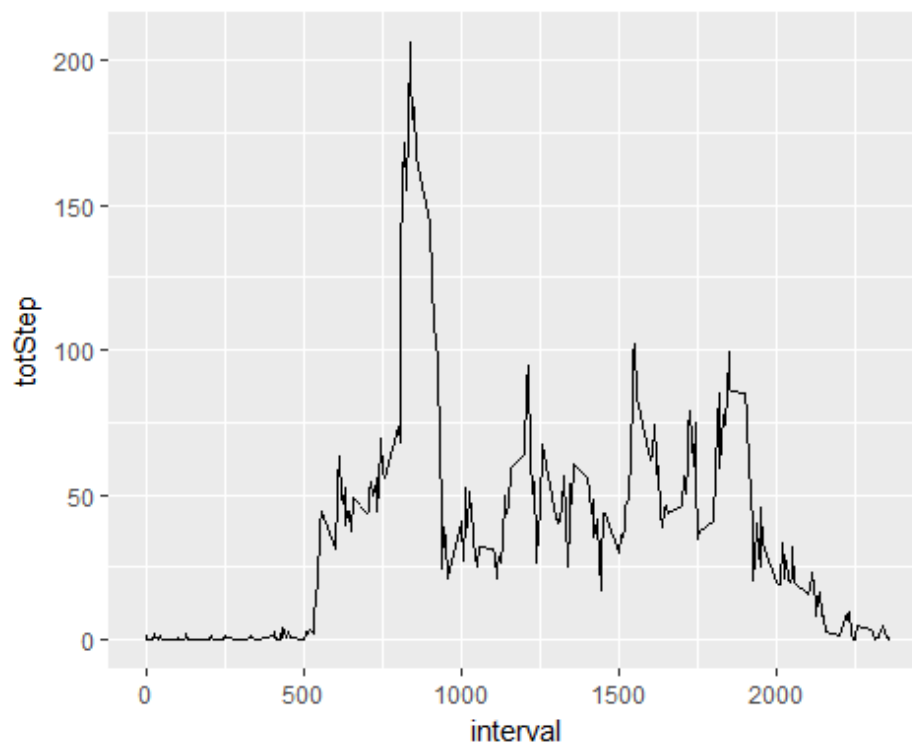
Average Daily Activity Pattern

```
intAvg <- activ2 %>% group_by(interval) %>% summarize(totStep = mean(steps))

ggplot(intAvg, aes(x=interval, y=totStep))+
  geom_line()
```

```
which.max(intAvg$totStep)

## [1] 104

intAvg[104,] # the 835 interval has the highest average total steps

## # A tibble: 1 × 2
##   interval  totStep
##      <int>    <dbl>
## 1      835 206.1698
```

Imputing missing values

```
activNA2 <- merge(activNA, intAvg, by=c("interval"))
count(activNA) # 2304

## # A tibble: 1 × 1
##       n
##   <int>
## 1  2304

activNA2 <- activNA2[,c(4,3,2,1)]
dim(activNA2)

## [1] 2304    4

activNA2$steps <- NULL
colnames(activNA) <- c("steps", "date", "interval")
activNA2$type <- c("imputed")
```
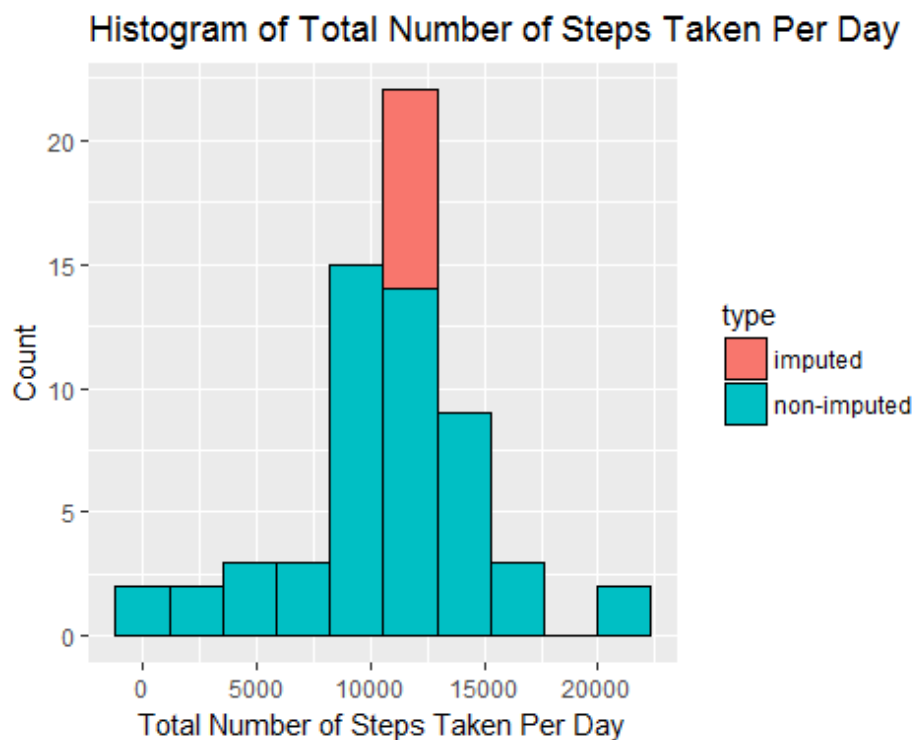
```
activ2$type <- c("non-imputed")

activJoin <- rbind(activNA2, activ2)
joinSum <- activJoin %>% group_by(date, type) %>% summarize(totStep =
sum(steps))

ggplot(joinSum, aes(x=totStep, fill=type))+
  geom_histogram(color="black", bins=10)+
  xlab("Total Number of Steps Taken Per Day")+
  ylab("Count")+
  ggtitle("Histogram of Total Number of Steps Taken Per Day")
```

Histogram of Total Number of Steps Taken Per Day


```
mean(joinSum$totStep)
```

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

```
median(joinSum$totStep)
```

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

Are there differences in activity patterns between weekdays and weekends?

```
activJoin$date <- as.Date(activJoin$date)
weekdays1 <- c('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')
```

```r
activJoin$wkDay <- factor((weekdays(activJoin$date) %in%
weekdays1),levels=c(FALSE,TRUE), labels=c('weekend','weekday'))

intWkDay <- activJoin %>% group_by(interval, wkDay) %>% summarize(totStep =
mean(steps))

ggplot(intWkDay, aes(x=interval, y=totStep))+
  geom_line()+facet_grid(wkDay ~.)
```