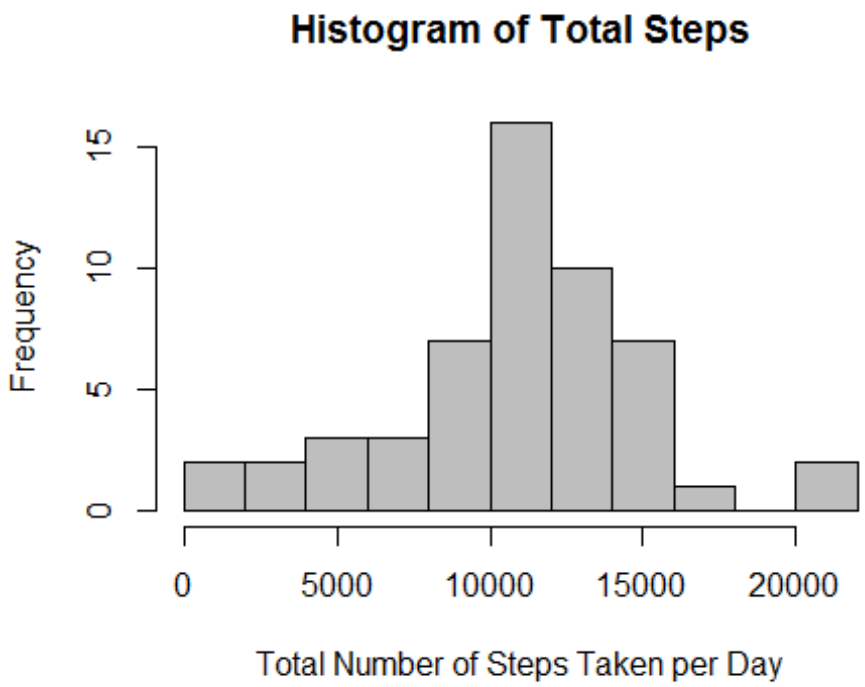
## Loading and processing data

```
##### Load packages
# require(knitr)
# require(markdown)
##### Create .md, .html, and .pdf files
# knit("PA1_template.Rmd")
# markdownToHTML('PA1_template.md', 'PA1_template.html',
options=c("use_xhml"))


setwd("C:/Users/GWANG1/Documents/Wanggu2/DataScience/DS5_ReproducibleResearch
/repdata_data_activity")
dat <- read.csv("activity.csv")

#### convert date to date data type
dat$date <- as.Date(dat$date)

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.4

#### Mean of total number of steps taken per day
datc <- na.omit(dat)
steps.total <-tapply(datc$steps, datc$date,FUN=sum, na.rm=TRUE)
dim(steps.total)

## [1] 53

hist(steps.total,breaks=10,xlab="Total Number of Steps Taken per Day",
col="grey",main="Histogram of Total Steps")
```




```
print("Mean of Total number of steps taken per day")
```

```
## [1] "Mean of Total number of steps taken per day"
```

```
mean(steps.total,na.rm=T)
```

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

```
print("Median of Total number of steps taken per day")
```

```
## [1] "Median of Total number of steps taken per day"
```

```
median(steps.total,na.rm=T)
```

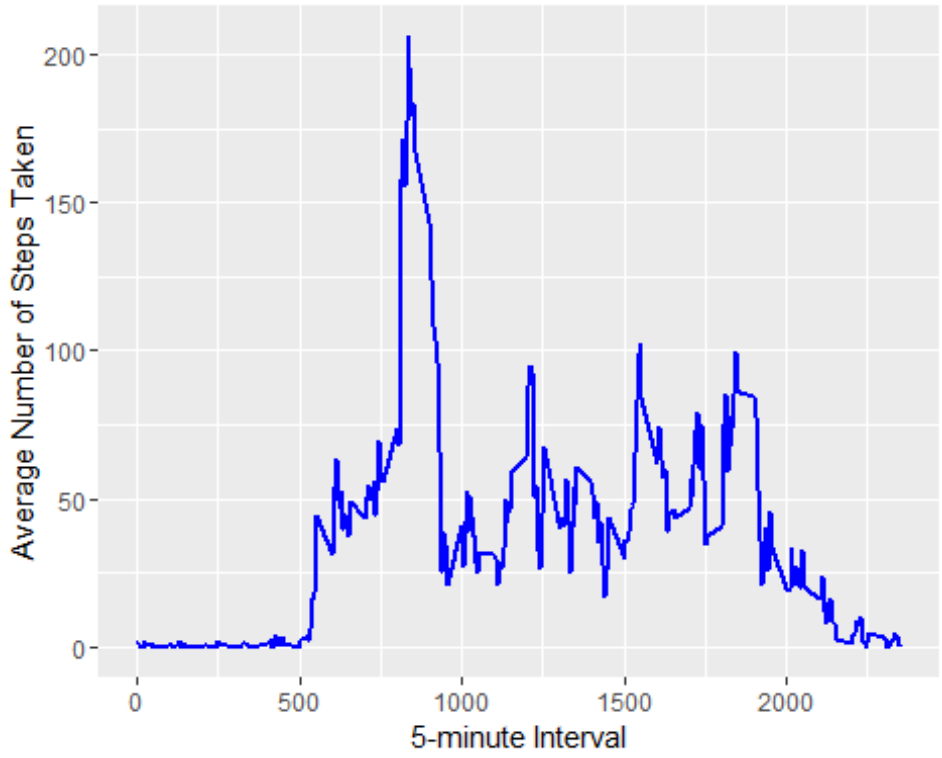
```
## [1] 10765
```

## Average daily activity

```
head(dat[!is.na(dat$steps),])
```

```
##     steps       date interval
## 289     0 2012-10-02        0
## 290     0 2012-10-02        5
## 291     0 2012-10-02       10
## 292     0 2012-10-02       15
## 293     0 2012-10-02       20
## 294     0 2012-10-02       25
```

```
avg <- aggregate(x=list(steps=dat$steps),by=list(interval=dat$interval),
FUN=mean, na.rm=T)
ggplot(data=avg, aes(x=interval,
y=steps))+geom_line(color="blue",size=1)+xlab("5-minute
Interval")+ylab("Average Number of Steps Taken")
```

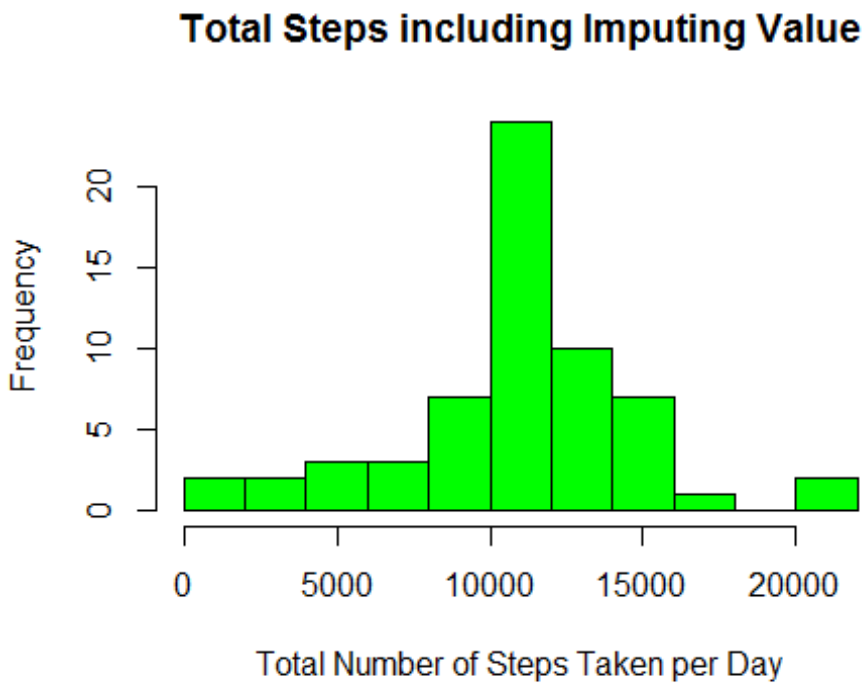



```
print("Maximum number of Steps in 5-minute Interval")

## [1] "Maximum number of Steps in 5-minute Interval"

avg[which.max(avg$steps),]

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

## Imputing missing values

```
#### 1. total number of missing values in dataset
sum(is.na(dat$steps))

## [1] 2304

##### 2. Replace missing value with mean for that 5-minute interval
dat.impmiss <- dat
nas<- is.na(dat.impmiss$steps)
avg_int<- tapply(dat.impmiss$steps, dat.impmiss$interval, mean, na.rm=TRUE,
simplify = TRUE)
```

```
dat.impmiss$steps[nas] <- avg_int[as.character(dat.impmiss$interval[nas])]
names(dat.impmiss)

## [1] "steps"    "date"     "interval"

print("#Missing")

## [1] "#Missing"

sum(is.na(dat.impmiss))

## [1] 0

#### Histogram
steps.imp <-tapply(dat.impmiss$steps, dat.impmiss$date,FUN=sum, na.rm=TRUE)
dim(steps.imp)

## [1] 61

hist(steps.imp,breaks=10,xlab="Total Number of Steps Taken per Day",
col="green",main="Total Steps including Imputing Value")
```




```
mean(steps.imp)

## [1] 10766.19

median(steps.imp)

## [1] 10766.19
```

## Difference in activity patterns between weekdays and weekends

```
#### Function: wwdat
wwdat <- function(date){
        day <-weekdays(date)
        if(day %in% c("Monday","Tuesday","Wednesday","Thursday","Friday"))
                return("Weekday")
        else if( day %in% c("Saturday","Sunday"))
                return("Weekend")
        else
                stop("Invalid Date")
}

dat.impmiss$weektype <- as.Date(dat.impmiss$date)
dat.impmiss$weektype <- sapply(dat.impmiss$weektype, FUN=wwdat)

#### Plot
avg.imp <- aggregate(steps~interval+weektype, data=dat.impmiss, mean)

library("lattice")
xyplot(steps ~  interval|factor(weektype), data=avg.imp,
            type = 'l',
            main="Averaged Across All Weekday Days or Weekend Days",
            xlab="Interval",
            ylab="Number of Steps",layout=c(1,2))
```