Project 1 - Reproducible Research

### 1. Code for reading in the dataset and/or processing the data

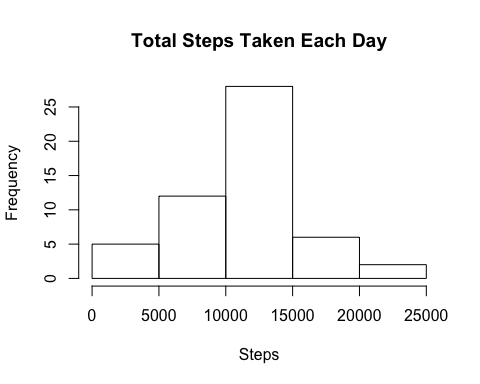
folder\_location <-"/Users/ethanschein/r-programming/RepData\_PeerAssessment1"  
setwd(folder\_location)  
unzip(zipfile = paste(folder\_location, "/", 'activity.zip', sep = ""))  
activity\_DS <-read.csv(file = 'activity.csv', header = T)  
activity\_DS$date <- as.Date(x = activity\_DS$date, "%Y-%m-%d" )

## Warning in strptime(x, format, tz = "GMT"): unknown timezone 'zone/tz/  
## 2018e.1.0/zoneinfo/America/New\_York'

good\_act <- activity\_DS[complete.cases(activity\_DS), ]

### 2. Histogram of the total number of steps taken each day

dta.sum <- aggregate(x = good\_act['steps'], FUN = sum, by = list(Group.date = good\_act$date))  
hist(dta.sum$steps, main = "Total Steps Taken Each Day", xlab = 'Steps')



### 3. Mean and median number of steps taken each day

print(round(mean(dta.sum$steps)))

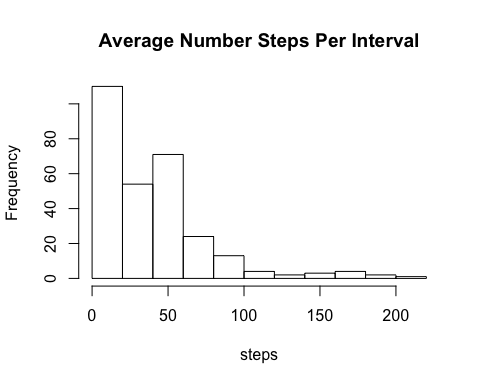
## [1] 10766

print(round(median(dta.sum$steps)))

## [1] 10765

### 4. Time series plot of the average number of steps taken per interval

interval <- aggregate(good\_act['steps'], by = list(good\_act$interval), FUN = mean)  
hist(interval$steps, main = "Average Number Steps Per Interval", xlab = "steps" )



### 5. The 5-minute interval that, on average, contains the maximum number of steps

print(interval[which.max(interval$steps),])

## Group.1 steps  
## 104 835 206.1698

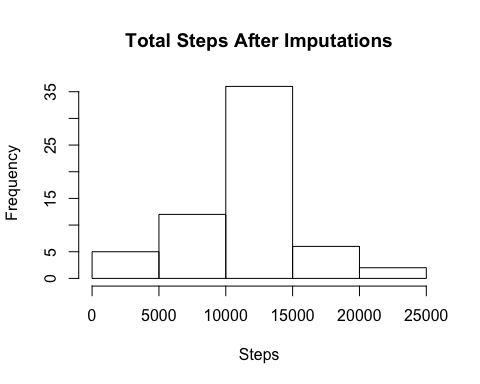
### 6. Code to describe and show a strategy for imputing missing data

for(i in 1:length(activity\_DS$steps)){  
 if(is.na(activity\_DS$steps)[i]){  
 row <- activity\_DS[i,]  
 int <- row[,which(colnames(row)=='interval')]  
 impute <- mean(activity\_DS[which(activity\_DS$interval==int),]$steps, na.rm = T)  
 activity\_DS[i,which(colnames(activity\_DS)=='steps')] <- impute  
 }  
}  
print(which(is.na(activity\_DS$steps)))

## integer(0)

### 7. Histogram of the total number of steps taken each day after missing values are imputed

dta.sum2 <- aggregate(x = activity\_DS['steps'], FUN = sum, by = list(Group.date = activity\_DS$date))  
hist(dta.sum2$steps, main = "Total Steps After Imputations", xlab = 'Steps')



### 8. Panel plot comparing the average number of steps taken per 5-minute interval across weekdays and weekends

index1 <- which(weekdays(activity\_DS$date)=="Saturday" | weekdays(activity\_DS$date)=='Sunday')  
 DS\_weekend <- activity\_DS[index1,]  
 DS\_weekday <-activity\_DS[activity\_DS!=index1,]  
   
 DS\_weekday\_agg <- aggregate(x = DS\_weekday['steps'], FUN = mean, by = list(DS\_weekday$interval))  
 DS\_weekend\_agg <- aggregate(x = DS\_weekend['steps'], FUN = mean, by = list(DS\_weekend$interval))  
  
 par(mfrow=c(1,2))  
 plot(x = DS\_weekday\_agg$Group.1, y = DS\_weekday\_agg$steps, type = "l")  
 plot(x = DS\_weekend\_agg$Group.1, y = DS\_weekend\_agg$steps, type = "l")

