

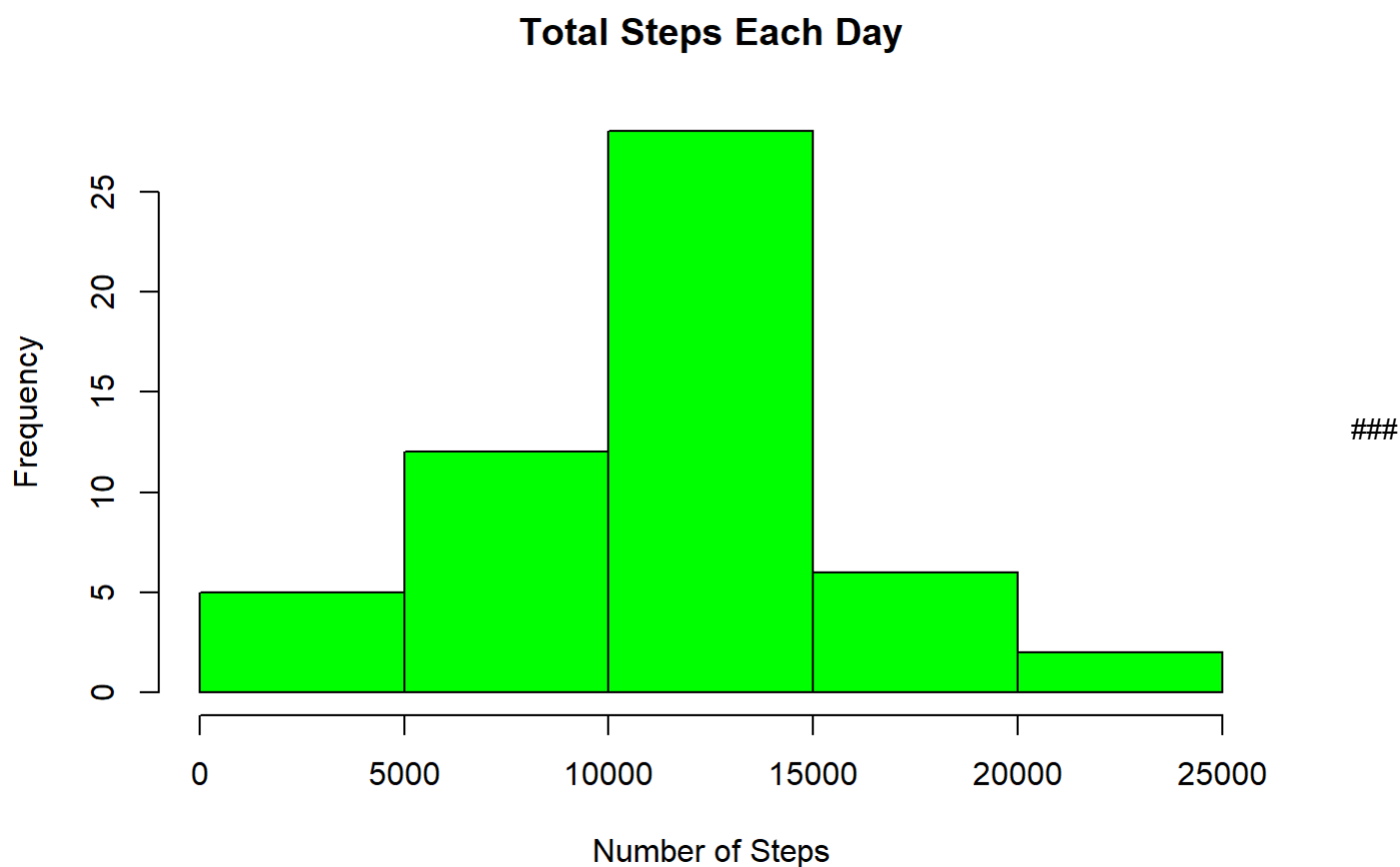
Loading and preprocessing data

```
if (!file.exists("activity.csv")) {  
  dlurl <- 'http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2Factivity.zip'  
  download.file(dlurl,destfile='repdata%2Fdata%2Factivity.zip',mode='wb')  
  unzip('repdata%2Fdata%2Factivity.zip')  
}  
data <- read.csv("activity.csv")
```

What is mean total number of steps taken per day?

Calculating steps per day and histogram

```
steps_per_day <- aggregate(steps ~ date, data, sum)  
hist(steps_per_day$steps, main = paste("Total Steps Each Day"), col="green",xlab="Number of Steps")
```



Mean and Median

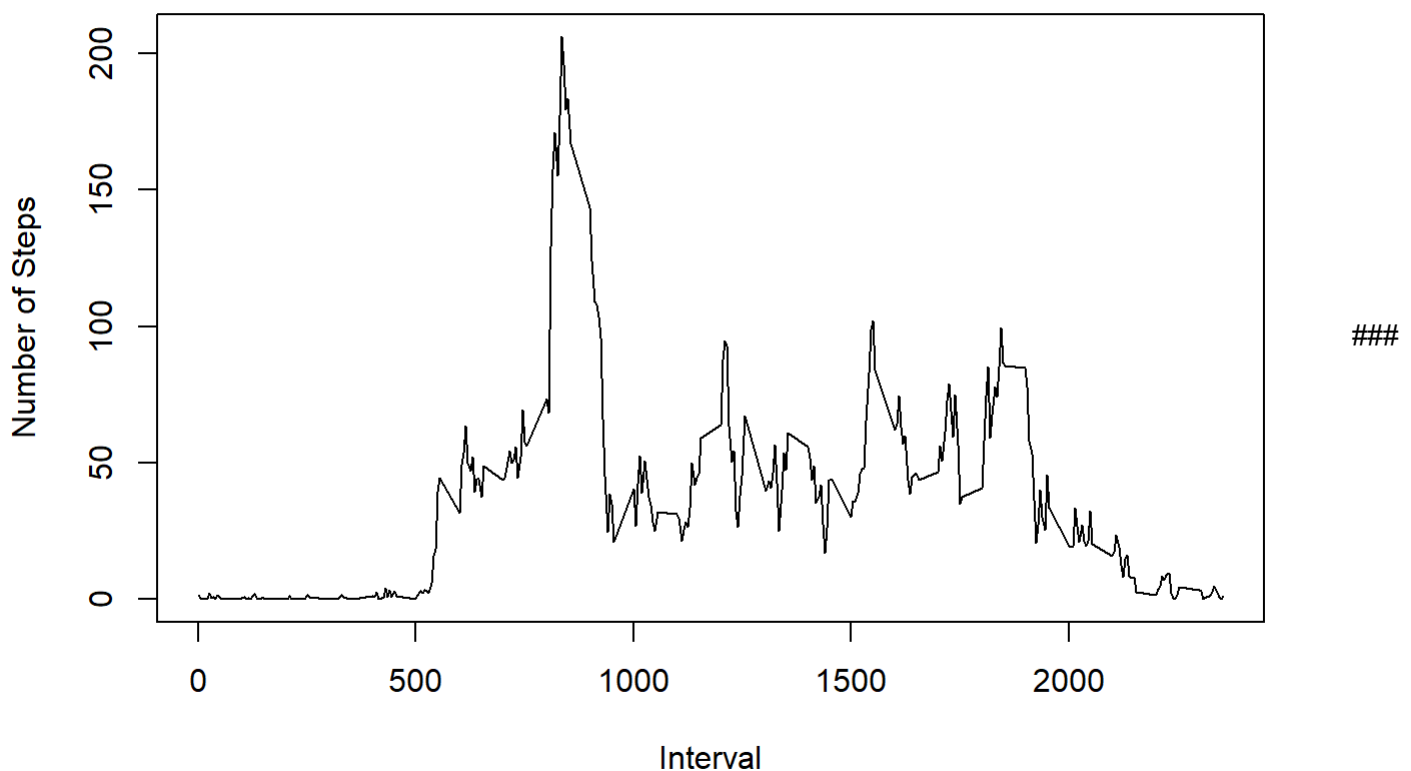
```
rmean <- mean(steps_per_day$steps)  
rmedian <- median(steps_per_day$steps)
```

What is the average daily activity pattern?

Time series plot (i.e. type = "l") of the 5-minute interval (x-axis) and the average number of steps taken, averaged across all days (y-axis)

```
steps_by_interval <- aggregate(steps ~ interval, data, mean)
plot(steps_by_interval$interval, steps_by_interval$steps, type="l", xlab="Interval", ylab="Number of Steps", main="Average Number of Steps per Day by Interval")
```

Average Number of Steps per Day by Interval



Which 5-minute interval, on average across all the days in the dataset, contains the maximum number of steps?

```
steps_by_interval[which.max(steps_by_interval$steps),1]
```

```
## [1] 835
```

Imputing missing values

Calculate and report the total number of missing values in the dataset

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

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

Putting in the mean of the day for missing values

```
StepsAverage <- aggregate(steps ~ interval, data = data, FUN = mean)
fillNA <- numeric()
for (i in 1:nrow(data)) {
  obs <- data[i, ]
  if (is.na(obs$steps)) {
    steps <- subset(StepsAverage, interval == obs$interval)$steps
  } else {
    steps <- obs$steps
  }
  fillNA <- c(fillNA, steps)
}
```

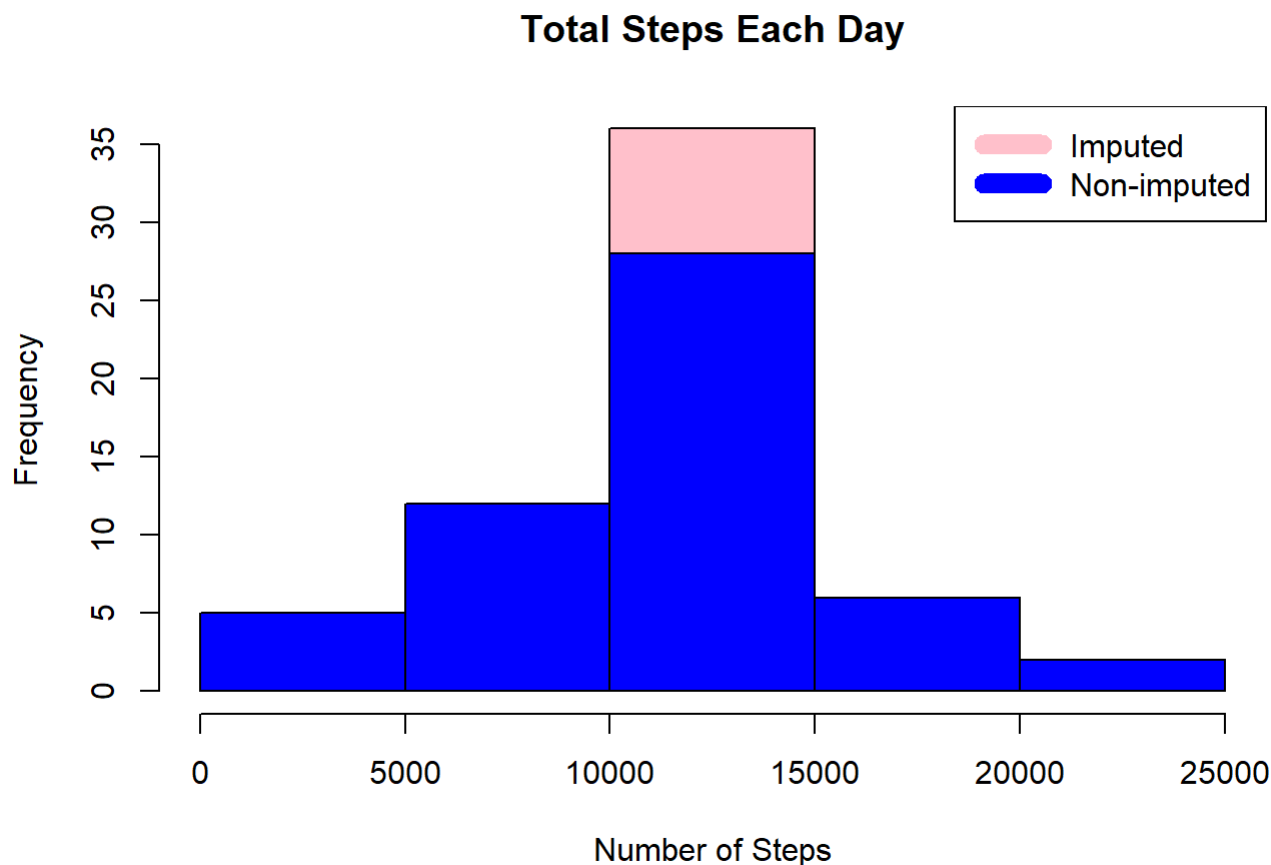
Create new data set

```
new_activity <- data
new_activity$steps <- fillNA
```

Make a histogram of the total number of steps taken each day
What is the impact of imputing missing data on the estimates of the total daily number of steps?

```
StepsTotalUnion <- aggregate(steps ~ date, data = new_activity, sum, na.rm = TRUE)
hist(StepsTotalUnion$steps, main = "Total Steps Each Day", col="pink", xlab="Number of Steps")

hist(steps_per_day$steps, main = "Total Steps Each Day", col="blue", xlab="Number of Steps", add=T)
legend("topright", c("Imputed", "Non-imputed"), col=c("pink", "blue"), lwd=10)
```



Mean and Median total number of steps taken each day

```
rmeantotal <- mean(StepsTotalUnion$steps)
rmediantotal <- median(StepsTotalUnion$steps)
```

Do these values differ from the estimates from the first part of the assignment?

```
rmeantotal - rmean
```

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

```
rmediantotal - rmedian
```

```
## [1] 1.188679
```

Are there differences in activity patterns between weekdays and weekends?

```
weekdays <- c("Monday", "Tuesday", "Wednesday", "Thursday", "Friday")  
weekends <- c("Saturday", "Sunday")
```

Make a panel plot

```
new_activity$dow <- as.factor(ifelse(is.element(weekdays(as.Date(new_activity$date))), weekdays),  
  "Weekday", "Weekend"))  
StepsTotalUnion <- aggregate(steps ~ interval + dow, new_activity, mean)  
library(lattice)  
xyplot(StepsTotalUnion$steps ~ StepsTotalUnion$interval | StepsTotalUnion$dow, main="Average Steps  
per Day by Interval", xlab="Interval", ylab="Steps", layout=c(1,2), type="l")
```

Average Steps per Day by Interval

