

Activity

This assignment to analyze data from a personal activity monitoring device.

```
activity <- read.csv("Y:/personal/personal/Work/teaching/Data Science/Coursera/Report Writing/activity.csv")
```

What is mean total number of steps taken per day?

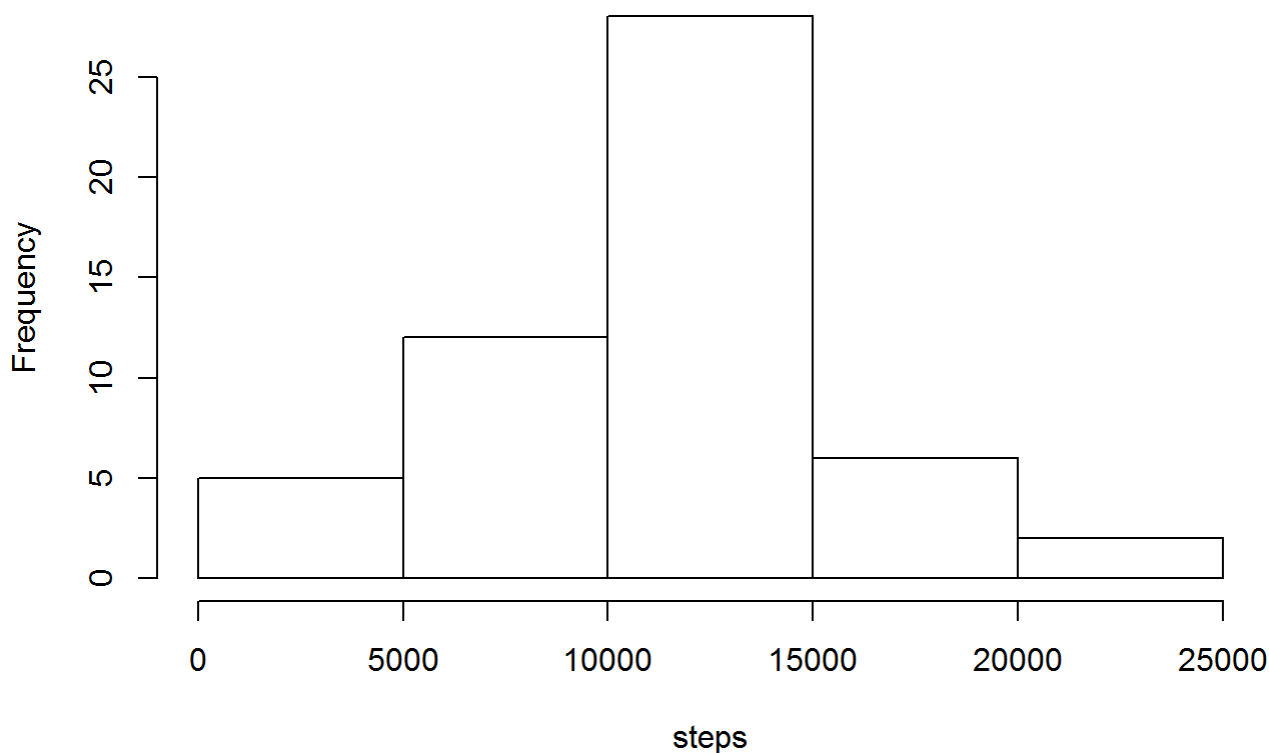
1. First calculate the total number of steps taken per day

```
tsteps<-aggregate(activity$steps, list(date=activity$date), sum)
```

2. Make a histogram of the total number of steps

```
hist(tsteps$x, main= "Total steps taken each day", xlab="steps")
```

Total steps taken each day



3. Calculate and report the mean and median of the total number of steps taken per day

```
mean(tsteps$x, na.rm=T)
```

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

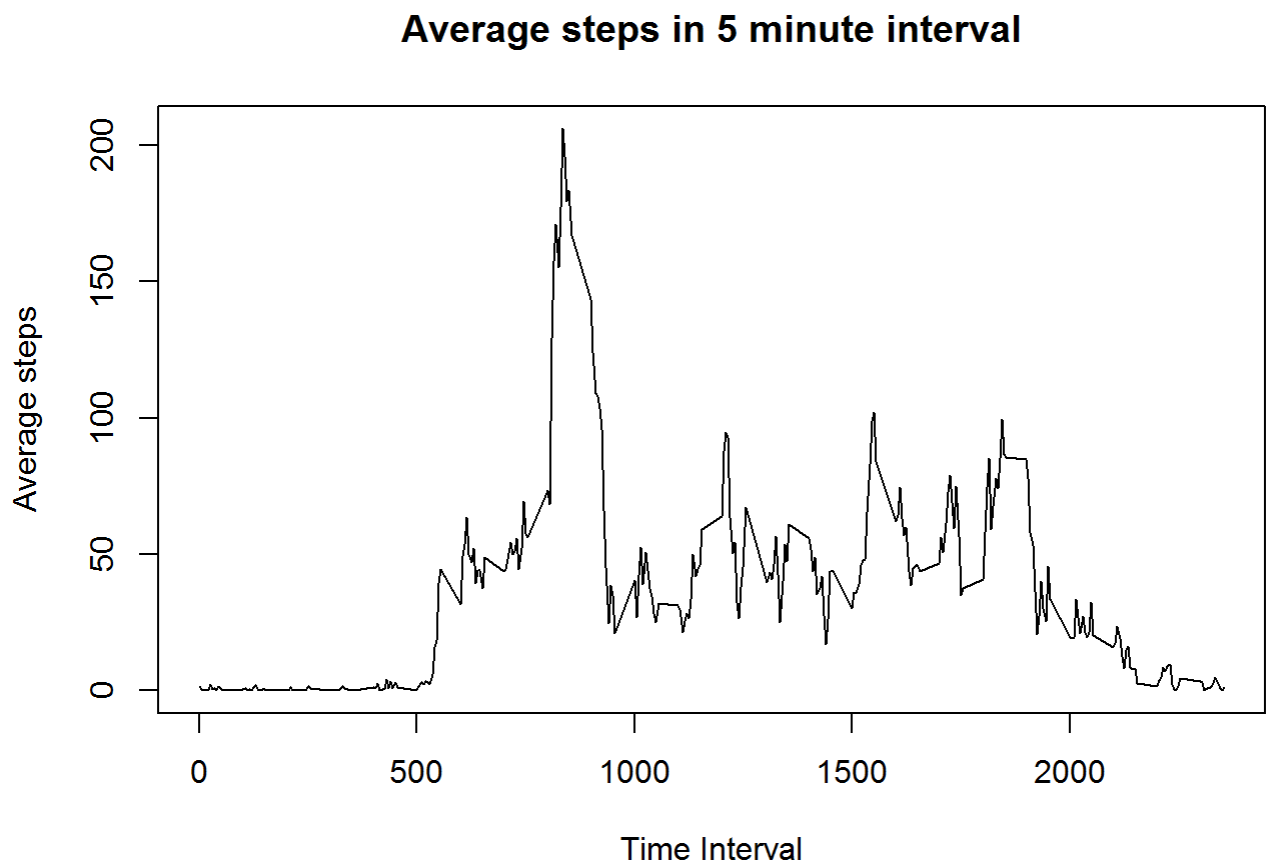
```
median(tsteps$x,na.rm=T)
```

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

What is the average daily activity pattern?

- 1. Make time series plot of the 5 -minute interval

```
isteps<-aggregate(activity$steps, list(time=activity$interval), mean,na.rm=T)
plot(isteps$time,isteps$x, type="l", main="Average steps in 5 minute interval", xlab="Time Int
erval", ylab= "Average steps")
```



- 2. Which time period contains the maximum number of steps?

```
isteps$time[isteps$x==max(isteps$x, na.rm=T)]
```

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

Imputing missing values

- 1. Calculate and report the total number of missing values

```
sum(is.na(activity$steps))
```

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

2. Devise a strategy for filling in all of the missing values in the data using mean for that 5-minute interval

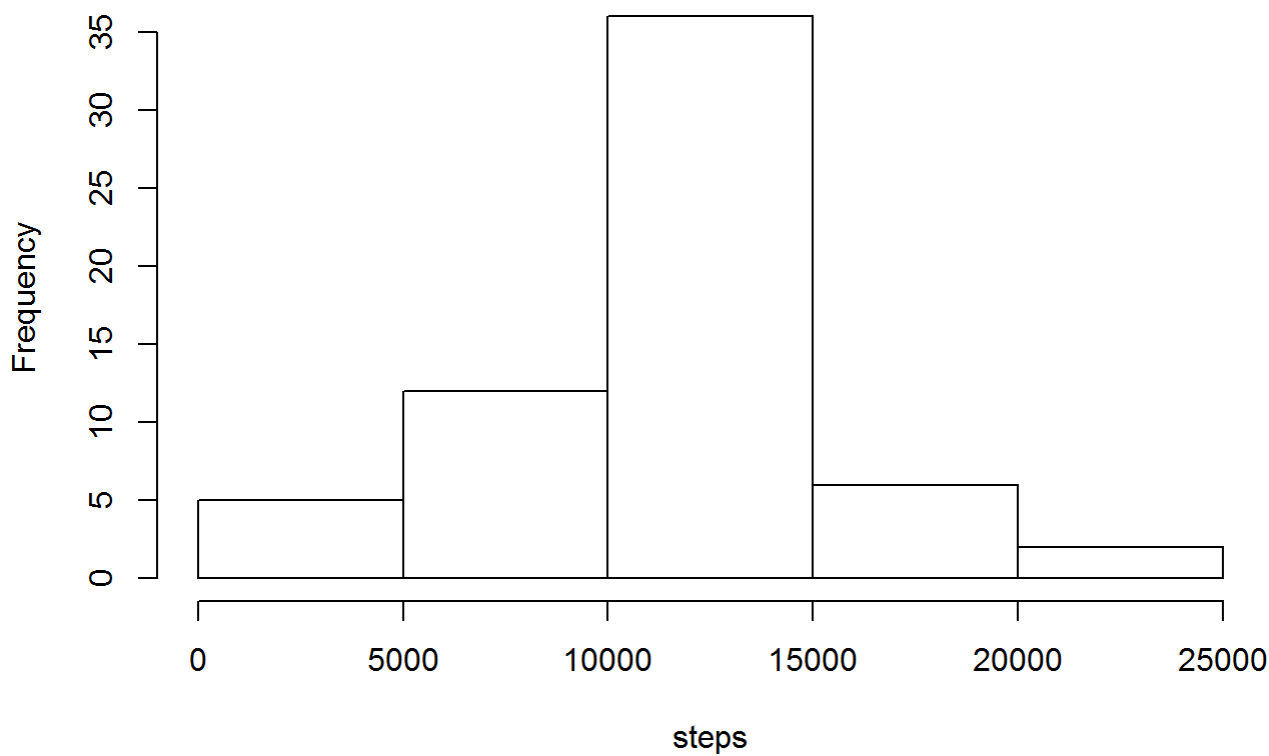
```
activitynona<-activity
activitynona$steps[is.na(activitynona$steps)]<-isteps$x
```

3. Create a new data that the missing data filled in It is already done in 2 and the new file name is acitivitynona.

4. Make a histogram of the total number of steps taken each day and report the mean and median total number of steps taken per day.

```
tstepsona<-aggregate(activitynona$steps, list(date=activitynona$date), sum)
hist(tstepsona$x, main= "Total steps taken each day with missing value imputation", xlab="steps")
```

Total steps taken each day with missing value imputation



```
mean(tstepsona$x,na.rm=T)
```

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

```
median(tstepsona$x,na.rm=T)
```

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

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

The mean and median estimates from the first part of the assignment are 10766.19 and 10765 and the estimates with missing value implimentation based on 5-minute interval mean are 10766.19 and 10766.19. These estimates are very similar.

What is the impact of imputing missing data on the estimates of the total daily number of steps?

It did not affect too much on the estimates. There were only slight difference in median values.

Are there differences in activity patterns between weekdays and weekends?

1. Created a new factor variable weekday1 with two levels. TRUE means weekday and FALSE means weekend

```
library(timeDate)
```

```
## Warning: package 'timeDate' was built under R version 3.1.3
```

```
weekday1<-isWeekday(activitynona$date)
weekdayf<-factor(weekday1, labels=c("weekend", "weekday"))
```

2. Time series plot

```
library(lattice)
istepsnona<-aggregate(activitynona$steps, list(interval=activitynona$interval, weekdayf=weekda
yf), mean, na.rm=TRUE)
xyplot(istepsnona$x~istepsnona$interval|istepsnona$weekdayf, data=istepsnona, type="l", layout
=c(1,2), ylab="Number of steps", xlab="Interval")
```

