# RepResearchProj1

*Kyle Davis*
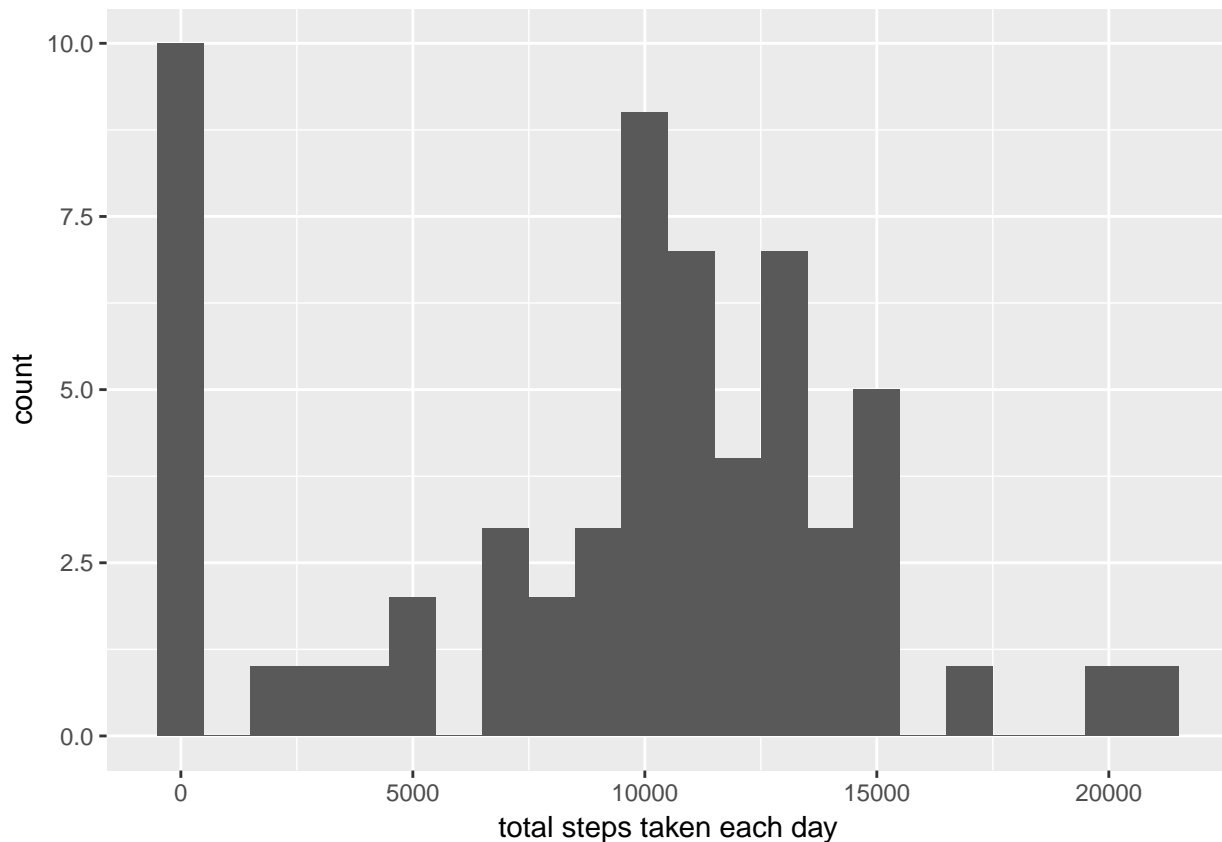
*10/29/2017*

### Loading and preprocessing the data

```
activity <- read.csv("activity.csv")
```

### What is mean total number of steps taken per day?

```
library(ggplot2)
total.steps <- tapply(activity$steps, activity$date, FUN=sum, na.rm=TRUE)
qplot(total.steps, binwidth=1000, xlab="total steps taken each day")
```



### What is the average daily activity pattern?

```
library(ggplot2)
averages <- aggregate(x=list(steps=activity$steps), by=list(interval=activity$interval),
                      FUN=mean, na.rm=TRUE)
ggplot(data=averages, aes(x=interval, y=steps)) +
```

```
    geom_line() +
    xlab("Five minute interval") +
    ylab("average steps taken")
```
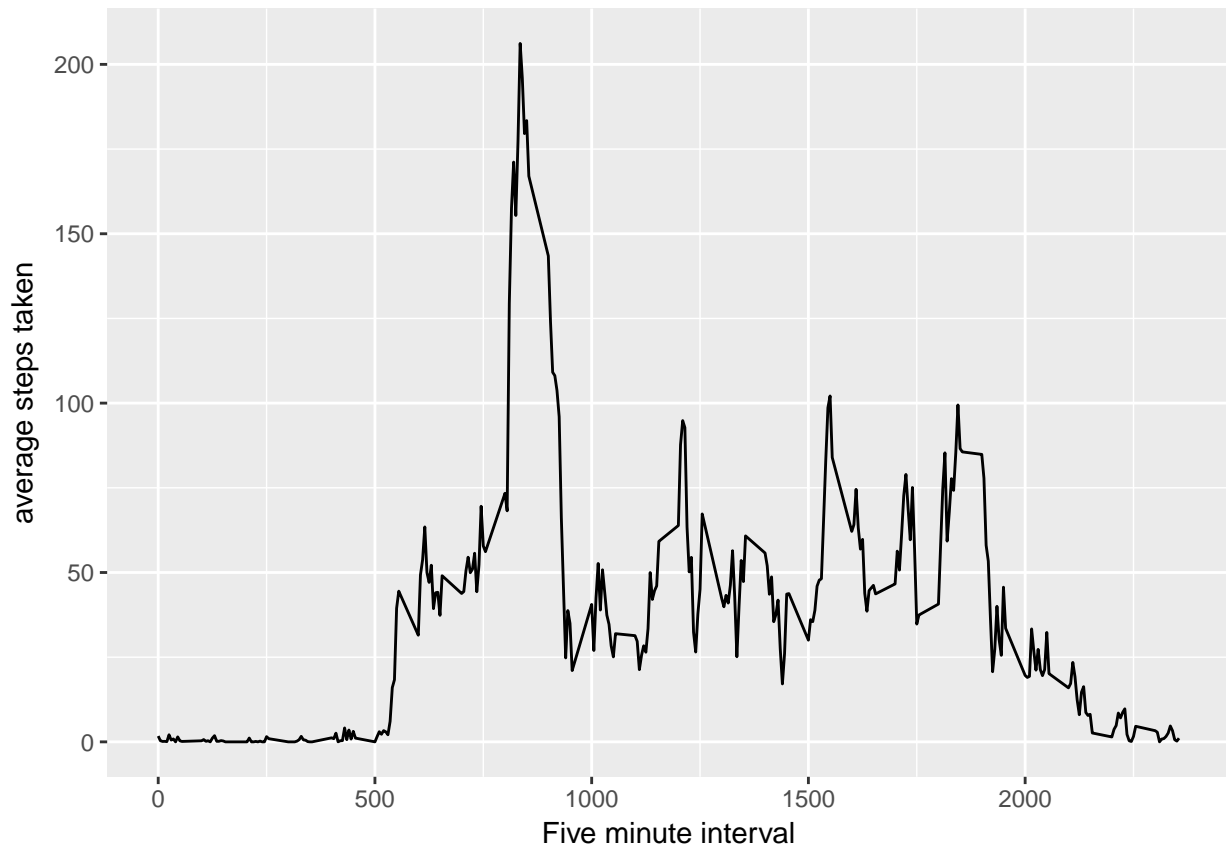


On average across all the days in the dataset, the 5-minute interval contains the maximum number of steps?

```
averages[which.max(averages$steps),]
```

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

## Imputing missing values

There are many days/intervals where there are missing values (coded as `NA`). The presence of missing days may introduce bias into some calculations or summaries of the data.

```
missing <- is.na(activity$steps)
# How many are missing
table(missing)
```

```
## missing
## FALSE  TRUE
## 15264  2304
```

Replace missing values with means from that five minute interval.

```
# Replace each missing value with the means
fill.value <- function(steps, interval) {
```

```
    filled <- NA
    if (!is.na(steps))
        filled <- c(steps)
    else
        filled <- (averages[averages$interval==interval, "steps"])
    return(filled)
}
filled.data <- activity
filled.data$steps <- mapply(fill.value, filled.data$steps, filled.data$interval)
```
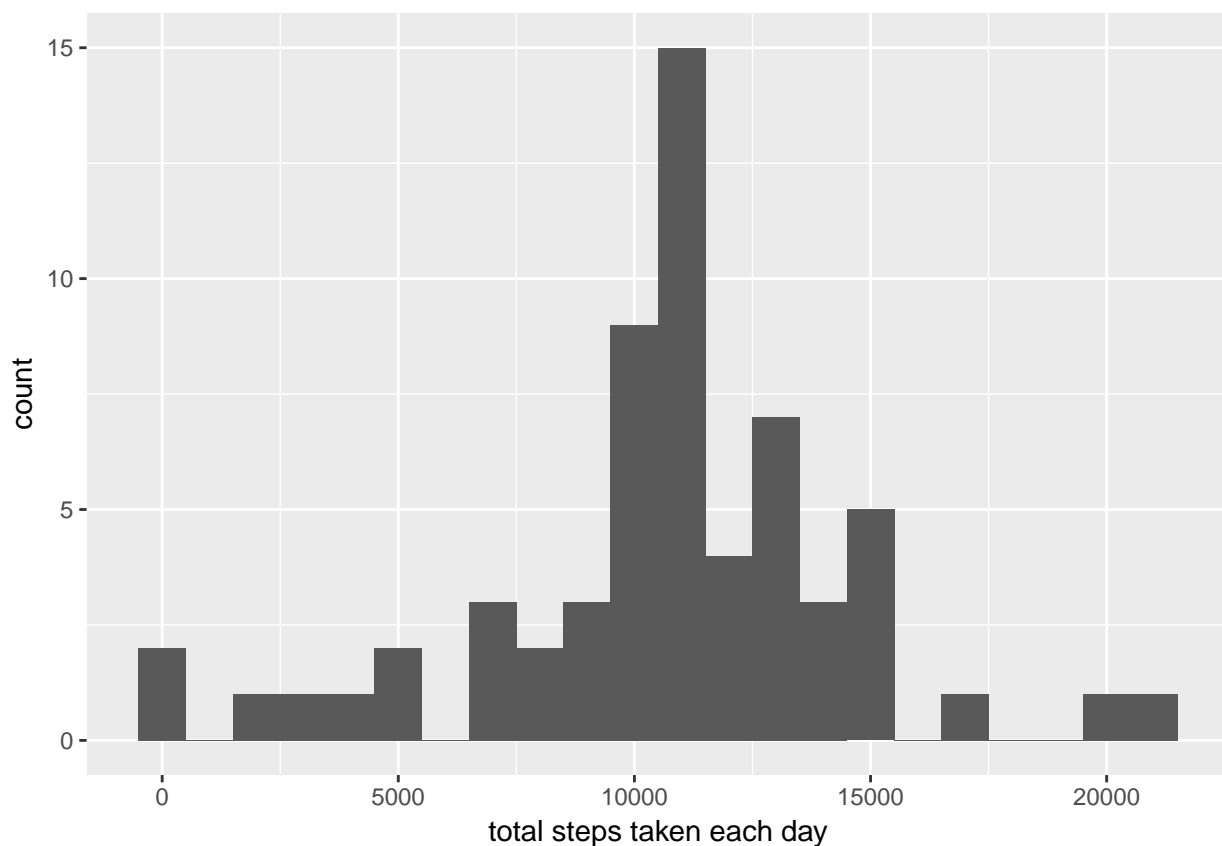
Using the transformed data set, make a histogram of the total number of steps taken per day and calculate the mean and median total number of steps.

```
total.steps <- tapply(filled.data$steps, filled.data$date, FUN=sum)
qplot(total.steps, binwidth=1000, xlab="total steps taken each day")
```



```
mean(total.steps)
```

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

```
median(total.steps)
```

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

## Are there differences in activity patterns between weekdays and weekends?

First, let's find the day of the week for each measurement in the dataset.

```
weekday.or.weekend <- 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")
}
filled.data$date <- as.Date(filled.data$date)
filled.data$day <- sapply(filled.data$date, FUN=weekday.or.weekend)
```

Make a panel plot containing plots of average steps taken on weekdays and weekends.

```
averages <- aggregate(steps ~ interval + day, data=filled.data, mean)
ggplot(averages, aes(interval, steps)) + geom_line() + facet_grid(day ~ .) +
    xlab("Five minute interval") + ylab("Steps Taken")
```