Review: Numerical and graphical summaries of one variable

- Numerical summaries:
  - Measures of center
  - Measures of spread
  - Counts [why would we use these?]
- Graphical summaries
  - Of one numeric variable
  - Of one categorical variable
  - Of one numeric and one categorical
  - Of two numeric variables
  - Of two categorical variables
- Descriptions of distributions
  - Center
  - Shape
  - Spread

Review: Simple linear regression I have some data about my daily activity that comes from both my Fitbit and my Leaf. They both try to quantify how much I've moved in a day by counting my steps, but they give me different information. Lets look at some numeric and graphical summaries of the model of my Fitbit steps by my Leaf activity summaries.

First, here's what my data looks like:

```
head(steps)
```

```
##         days    fb leaf weekday calories
## 1 2015-09-07 12672  114       1     2152
## 2 2015-09-08 10943   96       1     1995
## 3 2015-09-09  9875  109       1     2075
## 4 2015-09-10 10492   64       1     2274
## 5 2015-09-11  9177   80       1     1996
## 6 2015-09-12  9033   81       0     1958
```

```
dim(steps)
```

```
## [1] 18  5
```

Now, I can run a model,

```
m1 <- lm(fb~leaf, data=steps)
coef(m1)
```

```
## (Intercept)        leaf
##  4841.46241    55.09286
```

```
cor(fb~leaf, data=steps)^2
```

```
## [1] 0.7373445
```

- Write the equation for the linear model

- Interpret the coefficients, β_0, β_1

- Interpret the R^2 value

More on multiple regression Now, lets work on a multiple regression problem.

```
m2 <- lm(fb~leaf+weekday, data=steps)
coef(m2)

## (Intercept)        leaf     weekday
##  3727.46225    54.66871  1487.97131
```

- Write the equation of the regression line

- Interpret the coefficeints

- Calculate the R^2 value. (To find the R^2, you need a little more information.)

```
sum(residuals(m2)^2)

## [1] 21352934

sum((steps$fb - mean(~fb, data=steps))^2)

## [1] 107503633
```

The equation for multiple R^2 is

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\text{variability in residuals}}{\text{variability in the outcome}}$$

And the equation for adjusted R^2 is

$$R^2_{\text{adj}} = 1 - \frac{SSE/(n-k-1)}{SST} = 1 - \frac{SSE}{SST} \times \frac{n-1}{n-k-1}$$

- Calculate the multiple R^2 value

- Calculate the adjusted R^2 value