

Name: _____

Consider the multiple regression model:

$$E[Y] = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \cdot X_2$$

Y = film rating (scale of 1 to 4), *Rating*
 X_1 = year film was made, *Year*
 X_2 = long / short film, *long*

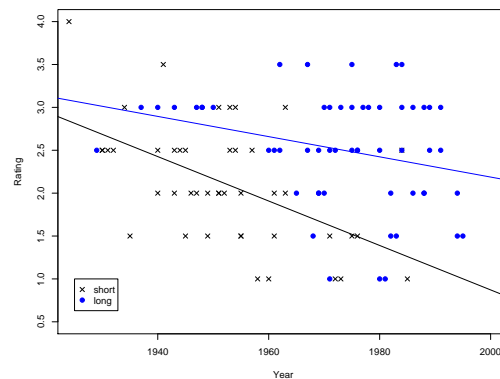
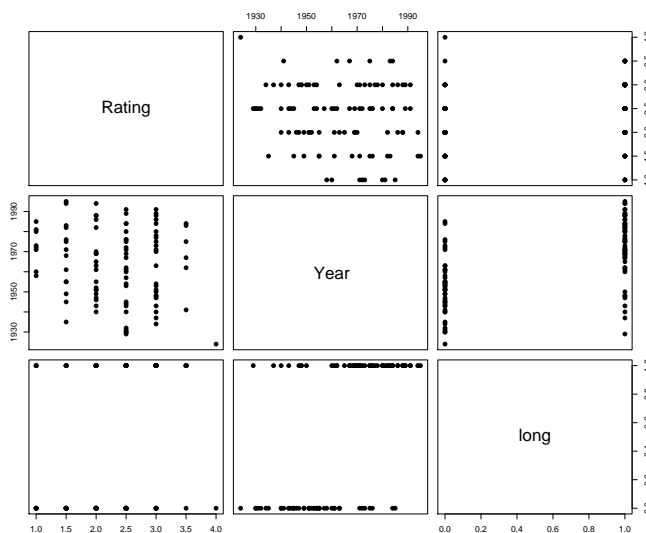
Note: *long* takes the value 1 if the film is longer than 90 minutes, otherwise the variable takes the value 0.

```
lm(formula = Rating ~ Year * long)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	52.686270	12.739579	4.136	7.58e-05	***
Year	-0.025907	0.006526	-3.970	0.000139	***
long	-26.927468	16.643930	-1.618	0.108975	
Year:long	0.014122	0.008489	1.663	0.099478	.

Give an interpretation of each of the three coefficients (not the intercept).



Note the following linear models:

```
lm(formula = Rating ~ Year + long)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	36.395472	8.221878	4.427	2.51e-05	***
Year	-0.017562	0.004212	-4.170	6.63e-05	***
long	0.758189	0.153053	4.954	3.08e-06	***

```
lm(formula = Rating ~ Year)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	13.525097	7.576427	1.785	0.0773	.
Year	-0.005699	0.003858	-1.477	0.1429	

Solution:

- The most important thing to notice here is that the coefficients (**and t-tests**) are based on the fact that the other variables are in the model. We can't possibly interpret the coefficients without considering the entire model.
- Additionally, keep in mind that the movies only get made once. That is, they can't be made in one year and then also in the next year. You should keep this idea in mind when interpreting variables that *can* increase in value for an individual observation (for example, height or age). We are *not* saying that for an individual observation at one level, the response will change when the explanatory goes up by one. We are saying that for the group at one level the average predicted value will be different than the group at one level + 1.

estimate of β_1 Given a movie is short, we estimate that the mean predicted rating will be 0.026 points lower for movies made one year later.

estimate of β_2 Given a movie is long, we estimate that the mean predicted rating for a movie made in the year 0 will be 25.76. Note that neither the rating nor the year are meaningful. Additionally, the prediction extrapolates outside of our range of data (which we do not want to do). A better interpretation would be that the linear model predicts the y-intercept for the long movies to be lower than the y-intercept for the short movies by 26.93 points. Be careful here, this does *not* mean that the rating for long movies are, on average, lower!!!

estimate of β_3 Given a movie is long, we estimate that the mean predicted rating will be 0.012 points lower for movies made one year later.