Name:_____ Date:_____

LAB 4F: This model is big enough for all of us!
Response Sheet

Directions: Record your responses to the lab questions in the spaces provided.

Building better models

Divide & Conquer

(1) Start by loading the `movie` data and write and run code splitting it into two sets (see Lab 4C for help).

- A set named `training` that includes 75% of the data.
- A set named `test` that includes the remaining 25%.
- Remember to use `set.seed`.

(2) Write and run code creating a linear model, using the `training` data, that predicts `gross` using `runtime`.

(3) Write and run code creating the MSE of the model by making predictions for the `test` data.

(4) Do you think that a movie's `runtime` is the only factor that goes into how much a movie will make? What else might affect a movie's `gross`?

Including more info

(5) Fill in the blanks below to predict `gross` using `runtime` and `reviews_num`.

```
lm(_____ ~ _____ + _____, data = training)
```

(6) Does this new model make more or less accurate predictions? Describe the process you used to arrive at your conclusion.

Name:_____ Date:_____

LAB 4F: This model is big enough for all of us!
Response Sheet

(7) Write down the code you would use to include a 3rd variable, of your choosing, in your `lm()`.

Own your own

(8) Write down which other variables in the `movie` data you think would help you make better predictions.

(9) Are there any variables that you think would not improve our predictions?

(10) Write and run code creating a model for all of the variables you think are relevant.

(11) Assess whether your model makes more accurate predictions for the `test` data than the model that included only `runtime` and `reviews_num`.

(12) With your neighbors, determine which combination of variables leads to the best predictions for the `test` data.