Today's data concerns house pricing in Los Angeles! We have access to data on all of the houses that sold in four cities around west LA in the course of a single month. Our goal is to predict the price of a home based on a few attributes of that home. Check Ed for a link to the data.

1. Use `ggplot2` to create a scatterplot examining the relationship between square footage and price. Write the code you used below and sketch the relationship that you see.

2. Add two columns onto the `LA` dataset which measure *log* price and *log* square footage. Save the new dataset back into the name `LA`. Write the code you used below.

3. Create another scatterplot using `ggplot2`, this time examining the relationship between *log* square footage and *log* price. Write the code you used below and sketch the relationship that you see. Then, describe the pattern you see and how it might differ from the pattern you saw in the previous sketch between square footage and price in their original units. Is it different/similar in shape? Is it easier/harder to see?

4. Fit two linear models and report the R^2 value for each. Write the code you used for *one* of the models.

- a model which predicts price with square footage
- a model which predicts log price with log square footage

When measuring by R^2, which model has the higher predictive power?

Recently the University of California purchased a new house to serve as the residence of the university President and to host university functions. The address of the house is 2821 Claremont Blvd in Berkeley.

5. Use your linear model to predict the sale price in log USD of this house. (hint: the internet is helpful!). Then, find this price in regular USD. Write the code you used below.

6. Was your model an under- or over-estimate? Why do you think this is? Explain in two to three sentences.