The following regression-related questions will use the *Salaries* data set, which is in the carData package in R. This data set contains data on academic salaries for Assistant Professors, Associate Professors, and Professors by discipline (A = “theoretical”, B = “applied”), years since PhD, years of service, and sex:

1. First, install and load the carData package
2. Take a look at the first few rows of the *Salaries* data to get a sense of what’s in it
3. We’re going to be taking a look at whether and how discipline, years since PhD, and sex are related to salary. So let’s visualize the data! Using ggplot, create a scatterplot of salary by years since PhD
4. Now create a violin plot of salary by sex (if you want a challange, add a little horizontal jitter to the points so they don’t overlap)
5. And now a violin plot of salary by discipline (ditto with the challenge)
6. Create three separate simple linear models, one for each predictor (years since PhD, sex, and discipline), and store them in an object in R with an intuitive name
7. Look at the summary output for each of the models and report the intercept and slope for all three.
    a. Why are the intercepts different in the three models? Interpret the intercept in each model.
    b. Interpret the slope in each model.
8. Which model explains the most variance? How do you know?
9. Now build a model with all three predictors and look at the summary output.
    a. How much variance is explained by this model?
    b. It went up! Why? Can it ever go down?
    c. How do the residuals compare to the residuals of the other models?
10. It looks like the model with all three predictors provides a better fit than the reduced models. Let’s test that explicitly by conducting model comparisons. Compare each of the three reduced models to the full model and describe your results.
11. You can see from the summary output of the model with all three predictors that it doesn’t look like sex is significant anymore. We often want the most parsimonious model, so now let’s see if we can get away with a slightly simplified model by creating a model without sex as a predictor and comparing it to the full model. This reduced model should have the years since PhD and discipline predictors in it, but not sex, so it differs from the full model only in whether sex is in the model.
12. Is it acceptable to stick with the model with only the two predictors in it [your answer should be yes]? How do you know? What would the output have looked like if the full model provided a better fit?
13. Finally, let’s look at how salaries differ by academic rank (Assistant Professor, Associate Professor, or Professor). Build a model with rank as the only predictor, then build a reduce model with only the intercept, conduct a model comparison, and report the results of the model comparison. What does this comparison tell us? What does this comparison *not* tell us?
14. What is the reference level for the rank variable? How do you know?
15. Now look at the summary output for the better fitting model. How do you interpret the regression coefficients? How much variance is explained by this model?

16. Ah, interesting. But wait, what I'm really interested in is whether a change in rank from Associate Professor to Professor is associated with a significant salary increase. How would you do that? What do the results of that analysis tell you? Why is one of the slope coefficient estimates now negative?