1. The line of “best fit” in a regression is calculated using the least squares criteria. How would you describe this procedure to someone?
2. A regression model has four notable assumptions (conditions), what are they? For each assumption describe how you assess whether it is violated.
3. When we interpret the coefficients from a linear regression they are in terms of the *mean* of the response variable. Why is this distinction important? How is this different than interpreting the relationship for *one* *specific* observation?
4. In section 3.2.6 the book describes how you can include a categorical variable with two levels as an explanatory variable in a regression. This is essentially a comparison of a difference in means, but with a regression line. How do the conditions for a linear regression relate to the conditions for a difference in means?
5. In section 3.3 you read about outliers and the amount of influence they can have on a regression line. Yet, the textbook states "it is tempting to remove outliers. Don’t do this without a very good reason."