Matt Fertakos worked w/ Bonnie and John

Question 1: In the context of a dataset (real or made up), describe the inherent conflict between using a complicated model that minimizes the unexplained variation and using a simple model that is easy to communicate.

There are many tradeoffs between increasing the complexity of your model to minimize unexplained variation and using a more simple model. Let's say, for example, in my research I have a dataset that lists where an invasive plant species was sold in the United States, and where it is located today. A simple model explaining this relationship could just be how many places a certain species was sold in relationship to their range today. The model could be made more complicated by adding more parameters to explain this possible relationship, like climate variables, how many years the plant was sold, and more. While the more complicated model might account for more variability, the simple model relating to the number of places the species was sold to the span of its range today is a much more intuitive relationship.

Question 2: (A)water and (B)nitrogen

Question 3: -1.7+(0*0.043)+(0*0.192)+(0*(-0.027))=-1.7g (lacking data near 0 can make a nonsensical intercept value)

Question 4: -1.7+10*0.043+(30*0.192)+(20*(-0.027))=3.95g

Question 5: A simple linear regression uses one continuous predictor and a continuous response, while a 1-way ANOVA has one categorical predictor (with 3+ levels) and a continuous response. ANOVAs must use strictly categorical predictors as opposed to simple linear regressions. Also, a simple linear regression assumes a linear relationship which ANOVA does not. The null hypothesis of a 1-way ANOVA is that all of the observations from the group came from the same population (not significantly different from one another), while the null hypothesis of a simple linear regression is that there is no statistically significant linear relationship between predictor and response variables (slope = 0).

Question 6: α (y-intercept), $\beta 1$ (slope), and xi (independent variable value)

Question 7: ε, the error term of the linear model is the stochastic component.