

# Multicollinearity and model minimalism

Previously, we discussed the concepts of multicollinearity and variable selection. As you continue learning about and applying multiple linear regression and these concepts, it is important to consider different scenarios. In your work as a data professional, you will encounter cases where you'll have plenty of data, but you'll have to consider the context and/or use advanced techniques to find the best model that only includes some of the data available.

For this discussion prompt, consider the following:

Why is it important to not include every variable possible in a multiple linear regression?

What are some scenarios where you might have a lot of variables available? Are there variables you think would be important? Are there variables you think might be less important? Why would you exclude them?