What have we done so far?

➡ Graphical Analysis
- Plot data, summary statistics, etc.
- Does the relationship between y and x look linear?

➡ Simple linear regression
- *Assuming that the relationship between y and x is linear*
- Conducted simple linear regression
- Interpretation of model and coefficients

  Intercept and slope
  CI's for coefficients and p-values
  CI's for mean fitted y and for predictions
  R^2

➡ Assumptions in simple linear regression
- Assumptions: linear relationship between y and x, residuals have constant variance and are normally distributed (also independence)
- So, is the relationship between y and x linear?

  Conduct regression of y on x and look at residual vs. fitted and Normal Q-Q plots
- If not linear relationship – transform?

  Which transformation/s? Do you have a theoretical model in mind?
  Try common transformations (e.g., log, inverse, sqrt, logit for proportions)
  Do the transformations help? Look at residual vs. fitted and Normal Q-Q plots
- Interpretation of coefficients after transforming – e.g., go back to the original scale

  Some that we've done in class:
  log y ~ x
  y ~ log x
  log y ~ log x
  logit(P) ~ x
- R^2 – only can use it to compare models if the response variable is the same
- Formal lack-of-fit F-test to determine if the linear model is appropriate (need repeated x's)

➡ Multiple regression and inference
- Can use x^2 term to "informally" test for lack-of-fit, x^2 allows you to model some curvature
- Dummy/indicator variables
- Extra SS F-test – is a reduced model OK?

  Note: the lack-of-fit F-test that we did is a special type of Extra SS F-test
  Various ways to conduct Extra SS F-test, e.g.:

  i.) run the 2 different models (e.g., reduced and full) and calculate Extra SS
  ii.) use the sequential SS ANOVA table – ORDER of fit is important
  iii.) in S-plus use the "model comparison" option
- F-test vs. t-tests – multiple parameter hypothesis vs. single parameter hypothesis
- R^2 vs. adjusted R^2
- Which variables to include in model? Variable selection – a topic for later

➡ Model checking diagnostics
- Outliers, leverage, influence, etc.