Unit 13: Predictive Analytics: Training and Testing Models

Case Studies:

- To introduce the concept of **using training data to *build* a model** and **using test data to *test* a model for it's predictive capabilities** we will, again, examine the relationship between a:
  - **Categorical response variable:** support for a certain opinion (favor/not in favor) and an
  - **Explanatory variables:**
    - Sex
    - Party, and
    - Age

Summary of Concepts:

- **Definitions/Properties**
  - Training data
  - Test data
- **Modeling**
  - **Why would we not want to "overfit" our dataset?**
    - What does it mean to "overfit" a dataset?
    - What are the downsides of "overfitting" a dataset?
  - **How can we assess how well our "trained" model will do when making predictions with "new" data (that it was not trained with)?**