Names:

1. What is the unit of observation in the data frame in the slides?

2. Come up with four questions that can be posed using the variables in this dataset:

- One that can be answered via a summary
- One that can be answered via a generalization
- One that can be answered via a prediction
- One that can be answered via a causal claim.

3. Name a variable which has an ambiguous (could be numerical or categorical) type to you. Then, sketch *two* visualizations of this plot: one where the variable is treated numerically, and one where the variable is being treated categorically. In each plot:

- depict a shape which reflects *your expectation* of the phenomenon;
- label your axes and provide a title.

4. What is your guess for the units/format used to record the departure time? Said another way, what would a value of 1517 represent?

5. What filter would you use to extract the flights that left in the springtime destined for Portland, Oregon? Draw a sketch of this smaller data frame showing at least the columns used in the filter.

6. What mutation would you use to create a column that records `TRUE` for flights leaving during or after March 2020 and `FALSE` otherwise?

7. What mutation would you use to create a column that records the average speed of a plane during its flight, in miles per hour? *Hint: think about the units of the variables needed to complete this mutation, and perform any unit conversion that is necessary.*

8. Provide one `dplyr` pipeline that calculates the mean departure delay for Oakland and San Francisco Airports. Provide a sketch of the resulting data structure with correct dimensions.