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. Which variables are categorical?

4. Which variables are numerical?

5. Do any of the variables have ambiguous data types to you? Write them here.

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

7. Write the code 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.

8. What line of code would you use to create a new column that records `TRUE` for flights leaving during or after March 2020 and `FALSE` otherwise?

9. What line of code 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.*

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