Assignment 09

This homework will help you to practice and self assess your knowledge and ability around the course concepts. Use the answer key to assess your work.

1. Load the **tidyverse** & **nycflights13** packages.

2. View the following data sets from the **nycflights13** package:

   - airlines
   - airports
   - flights
   - planes
   - weather

   Look at the help files for these data sets in order to understand their variables.

3. Use left joins to combine all of these data sets into one large data set. Assign the output to an object named `full_flights` (we will use this later).

   - **Hint**: flights have attributes of weather, planes, airports, & airlines. Start with the most granular data set (`flights`) and move to the most broad (`airlines`).
   - **WARNING:** some tables contain the same variable name that is not the same variable. You will likely want to `rename` these variables before joining.

4. Are the number of rows on the new data set sensible compared to the original? Were the matches a one to one or one to many match?

5. Use the combined data set to determine how many flights are flown by each air line.

6. If I am fearful of flying and assume that newer planes are safer I will want to fly an airline that is most likely to have newer planes. Use EDA to make a table/plot that shows the average year of planes within each airline. Order the table/plot descending by average year.

7. Using the mean may result in a loss of information. Plot the distributions of plane build years by airline.

8. Is the build year correlated with arrival delays? Make a plot to investigate this relationship. Be careful of overplotting.