CPTR420
September 26, 2023 Assignment
Due September 28

Nycflights13 data is a package that contains datasets related to NYC flights in 2013
Flights – has flight data for flights from New Your City in 2013
Airlines - has the full carrier name from its abbreviated code
Planes – has data about each plane, identified by its tailnum
Weather - weather gives the weather at each NYC airport for each hour:
Airports - has data about each airport, identified by the airport code.

1. Install and load the nycflights13 package. Load each of the 5 datasets into appropriately named dataframes

   Alternatively, you may read the files from the zipped folder you downloaded from learningHub and load each into an appropriately named dataframe

2. View the flights data
   a. Column names,
   b. data types
   c. descriptive stats

3. Split the time_hour column into date and time. Do not remove the original column

4. For each df, decide what you think is the key variable(s) and check to see whether there are any duplicates

5. Combine the flight day, month, year into a single field called flightDate. Similarly combine the year, month, day for the weather df into a single column called obsDate

6. Create a new df called dailyWeather that will have the average temp, average wind_speed, and average humidity for each origin airport each day

7. Create a new df called full_flights. This should have (flight, origin, destination, … all the data from flights as well as the appropriate carrier name, average temperature, average wind speed and average humidity

8. Is there a relationship between the age of a plane and its delays?

9. What weather conditions make it more likely to see a delay?

10. Filter flights to only show flights with planes that have flown at least 100 flights.

11. Combine `fueleconomy::vehicles` and `fueleconomy::common` to find only the records for the most common models.

12. Find the 48 hours (over the course of the whole year) that have the worst delays.