

# dplyr basics

Submit a single neat .html document via Moodle. In general, you should include both code and output in your solutions, but please be judicious in how much of any data frame you display. Use `gt` when possible and `glimpse` otherwise.

These problems refer to the `flights` data set, included in the `nycflights13` package, which you will need to install. Before attempting these problems, you should take a few minutes to look at the data set and its help file to understand what the different variables represent and how they're encoded.

## Problem 1

Obtain a smaller set `flights_sm` consisting only of destinations and the variables involving arrival/departure times and delays. The destinations should be the first column of this data frame, followed by the departure variables, and finally the arrival variables. Glimpse your result.

## Problem 2

Use `flights_sm` for each of the following. Glimpse your results.

- Find all flights that flew to Boston (BOS) or Atlanta (ATL).
- Find all flights that left at least an hour late but arrived early. Create a table of your results.
- Read about the `between()` helper. Use it to find all flights that arrived between 1am and 7am, inclusive.

### Problem 3

- a. Sort `flights_sm` according to departure delay, breaking ties with arrival delay. In each case, the longest delays should be first.
- b. Sort the set `flights_sm` so that all flights with missing departure times are first.

### Problem 4

Add a column to `flights_sm` showing the difference between delays in departure and arrival for each flight. Add a second column indicating if the flight gained time while in the air. The latter should be a logical vector (one consisting of `TRUE` and `FALSE`).