

csci297b Exercise 6

dplyr

For this project, use the `dplyr` library and pipes.

1. Open `project_03_yourname`.
2. Open a new R markdown file in the project called `exercise07.Rmd`
3. Load the `nycflights13` dataset
4. Compare `dep_time`, `sched_dep_time`, and `dep_delay`. How would you expect those three numbers to be related?
5. Brainstorm as many ways as possible to select `dep_time`, `dep_delay`, `arr_time`, and `arr_delay` from `flights`.
6. What happens if you specify the name of the same variable multiple times in a `select()` call?
7. What does the `any_of()` function do? Why might it be helpful in conjunction with this vector?

```
variables <- c("year", "month", "day", "dep_delay", "arr_delay")
```

8. Does the result of running the following code surprise you? How do the `select` helpers deal with upper and lower case by default? How can you change that default?

```
flights |> select(contains("TIME"))
```

9. Rename `air_time` to `air_time_min` to indicate units of measurement and move it to the beginning of the data frame.

10. Why doesn't the following work, and what does the error mean?

```
flights |>
  select(tailnum) |>
  arrange(arr_delay)
#> Error in 'arrange()':
#> i In argument: '...1 = arr_delay'.
#> Caused by error:
#> ! object 'arr_delay' not found
```