Socio-Informatics 348
Practical 7

Submission Instructions

- Submit your completed practical as `studentnumber.qmd` on SocSciLearn.
- Submissions are checked for completeness, not correctness.
- At least 80% of exercises must be attempted to receive 1% towards AF assessment.
- Attendance of at least one practical session per week is required to earn the 1% for that week's practical.

Deadline

Friday 3 October, 17:00 (submit on SocSciLearn)

Chapters Covered:

- R4DS: Chapter 19

Exercises

1. Using the dataframes available in the `nycflights13` package join the `weather` and `airports` dataframes. Which columns are the primary and foreign keys in this relationship?
2. Is there a relationship between the age of a plane and its delays?
    - Calculate the average `dep_delay` and `arr_delay` for each plane (`tailnum`).
    - Join this with the `planes` dataframe to determine its age.
    - Plot year against delay (separate facets for arrival delay and departure delay).
3. What does it mean for a flight to have a missing `tailnum`? What do the tail numbers that don't have a matching record in `planes` have in common? (Hint: one variable explains ~90% of the problems; use `anti_join`).

4. Can you explain what's happening with the keys in the following equi–joins? Why are they different?

```
x |> full_join(y, join_by(key == key))
#> # A tibble: 4 × 3
#>     key val_x val_y
#>   <dbl> <chr> <chr>
#> 1     1 x1    y1
#> 2     2 x2    y2
#> 3     3 x3    <NA>
#> 4     4 <NA>  y3

x |> full_join(y, join_by(key == key), keep = TRUE)
#> # A tibble: 4 × 4
#>   key.x val_x key.y val_y
#>   <dbl> <chr> <dbl> <chr>
#> 1     1 x1        1 y1
#> 2     2 x2        2 y2
#> 3     3 x3       NA <NA>
#> 4    NA <NA>      4 y3
```