Socio-Informatics 348
Practical 2

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 22 August, 17:00 (submit on SocSciLearn)

Exercises

Section 1: NYC Flights (from `nycflights13`)

1. What happens if you specify the name of the same variable multiple times in a `select()` call?

2. Why does the following code not work?

```
flights |>
  select(tailnum) |>
  arrange(arr_delay)
```

3. Which flights traveled the farthest distance? Which traveled the least distance?

4. 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"))
```

5. How do delays vary over the course of the day. Illustrate your answer with a plot. (Hint: Use the `hour` variable.)

6. Sort flights to find the fastest flights. (Hint: You need to calculate a `speed` variable, measured in miles per hour.)

7. Which carrier has the worst average delays? Challenge: can you disentangle the effects of bad airports vs. bad carriers? Why/why not? (Hint: think about `flights |> group_by(carrier, dest) |> summarize(n())`).

Section 2: `table 2` sand `table 3` (from `tidyverse`)

8. For each table, write the code needed to calculate the correct `rate` of TB cases.
   For each table, you will need to:

   - Get the number of TB cases per country per year in its own column.
   - Get the population per country per year in its own column.
   - Use `mutate()` to calculate the rate of TB cases per 10 000 people.

   Hint for `table3` – use `separate_wider_delim()`:
   Type `?separate_wider_delim()` into the console to see the documentation for this function.

Section 3: Non-syntactic names

In R, non-syntactic names are names that do not conform to the rules for valid R variable names. Examples of these are names that begin with a number, or names that contain spaces. To use non-syntactic names, you must enclose them in backticks (‘).

9. Practice referring to non-syntactic names in the following data frame:

```
annoying <- tibble(
  `1` = 1:10,
  `2` = `1` * 2 + rnorm(length(`1`))
)
```

   Complete the following:

   - Extract (`select`) the variable called `1`.
   - Plot a scatterplot of `1` vs. `2`.
   - Create a new column called `3`, which is `2` divided by `1`.
   - Rename the columns to `one`, `two`, and `three`.