Auckland University of Technology
Department of Mathematical Sciences
COMP824: Week 9 Lab

Some of the exercises in this lab are based on those found in Wickham and Grolemund, R for Data Science.

Getting started

(1) In this lab we will need a number of packages.

- Install them, if required, by uncommenting the following code.

```
# List of packages
packages <- c("tidyverse", "nycflights13")

# Find packages that are not installed
packages_to_install <- packages[!(packages %in% installed.packages()[,"Package"])]

# Install required packages
# if(length(packages_to_install) > 0){ install.packages(packages_to_install,
```

- Load the packages

```
# Load new packages
sapply(packages, library, character.only=TRUE)
```

Tip

You can get help in R by typing "?" followed by the function name. For example ?ggplot

Joins

(2) Consider the following data sets about student grades.

- Run the following code and inspect the outputs.

```
COMP824exams <- tribble(
  ~studID, ~exam_mark,
     1, "A",
     2, "B",
     3, "C+"
)
COMP824classlist <- tribble(
  ~studentID, ~name, ~programme,
     1, "Charlotte", "CIS",
     2, "Zoe", "CIS",
     3, "Caitlin", "MA",
     4, "Abel", "MSc"
)

STAT800classlist <- tribble(
  ~studentID, ~name,
     1, "Charlotte",
     2, "Zoe",
     6, "Conor",
     7, "Archie"
)
```

Use the appropriate `join` function to create a dataset that can answers the following questions.

- Create a dataset containing the name, ID and exam mark of all COMP824 students. Include students who didn't sit the exam.
- Create a dataset containing the name, ID and exam mark of the COMP824 students who sat the exam.
- Create a dataset containing the name, ID and programme of the COMP824 students who sat the exam.
- Which students didn't sit the COMP824 exam?
- Create a dataset containing the name and ID of all students studying either COMP824 **or** STAT800 or both. Ensure your new dataset has only 2 columns: one for ID and one for name.
- Create a dataset containing the name and ID of all students studying both COMP824 **and** STAT800. Ensure your new dataset has only 2 columns: one for ID and one for name.

Flights

(3) Load the package `nycflights13`. Use the appropriate `join` function to create a dataset that can answer the following questions. You will need to you some other functions well (e.g. `filter`).

- Create a table showing the names of the destinations that JetBlue Airways flies to. Include a column showing the number of flights to each destination.

Start from this code:

```
selected_airlines <- airlines %>% filter(name %in% c("JetBlue Airways"))
```

- What destinations do both JetBlue Airways and United Air Lines Inc (UA) fly to?

(4) Use ggplot to construct a bar plot showing the number of flights per month. Add a title and subtitle to your plot.

(5) Join flights and airports plot flight paths of 10 flights using longitude and latitude.

(6) Challenge: Create a plot showing a line between the 20 most common origin/destination pairs. Use line thickness to represent the frequency with which flights fly between the two locations, and colour to represent the origin.

Further practice

(7) Review the examples in the lecture notes
(8) Read R4DS chapters 13 and do the exercises.