

Chapter 4: Data Management

Exercises

Brian Fogarty

Exercise 1

Read in the following datasets:

- `simd2020.csv` and name it as `simd`.
- `scottish health boards by datazone.xlsx` and name it as `healthboard`.
- `covid total by health board.xlsx` and name it as `covid`.

Check whether the number of observations and variables match what is displayed in the chapter for each dataset.

Exercise 2

Use the `count()` function to view and sort the values (by magnitude) for the following variables:

- `Council_area` in the `simd` dataset.
- `health_board` in the `healthboard` dataset.
- `health_board` in the `covid` dataset.

Why are the number of observations for **Greater Glasgow and Clyde** different in the variable `health_board` between the `healthboard` and `covid` datasets?

Exercise 3

Using the `simd` dataset, determine how many missing and non-missing values exist for the following variables:

- Broadband
- Employment_rate
- drive_GP

Exercise 4

Using the `simd` dataset,

- Subset the data to only include keep the variables: `Intermediate_Zone`, `Council_area`, `Broadband`, `Employment_rate`, and `drive_GP`. Save the subsetted data as a new tibble object named `simd_sub`.
- Filter the tibble `simd_sub` by the **City of Edinburgh**. How many datazones are there?
- Filter the tibble `simd_sub` by the **City of Edinburgh** and for datazones with 25% or more of its residents classified as employment deprived. How many datazones are there?
- Filter the tibble `simd_sub` by the **City of Edinburgh** and for datazones with 25% or more of its premises without access to “superfast” broadband. How many datazones are there?
- How many datazones in the **City of Edinburgh** have 25% or more of its residents classified as employment deprived AND 25% or more of its premises without access to “superfast” broadband?

Exercise 5

Using the `inner_join()` function, merge the `simd`, `healthboard`, and `covid` datasets that were read-in for Exercise 1. Remove any duplicate variables.

For the answers see **Chapter 4 - Answers to Exercises**.