If you’re not doing this module straight after the last one, you’ll need to read in the libraries again and create the data frame ‘data’.

Summary is a useful function for giving you an overview of your data. If we run that line we can see the output down in the console. For the columns which are numeric we have the minimum and maximum values, 1st and 3rd quartile, median and mean. For columns with categorical data it will show the count for the first few rows and then a total for how many are in other categories. If there were any missing values in our dataset it would tell us that as well.

For a quick overview of a single column of categorical data, we can use the table function.

If you’re working with data that you’ve imported from a spreadsheet created in Excel, you’ll often find that people have included spaces or used capital letters in the column names. This is something that you can change manually, using the function rename(), which takes the order rename(new\_name = old\_name)

For changing multiple columns, it’s much easier to use a function from the janitor package.

The Janitor package has other helpful functions, including to change the format of dates, remove empty rows, or find duplicate rows. To look at what’s available type ?janitor in the console. If we scroll down to the bottom of the help pane and click on index, we can see the full list of functions in the package.

Q Which expression will change the column name from “number of breaches” to “breaches”?

1. rename(number of breaches = breaches)
2. rename(`number of breaches` to breaches)
3. rename(breaches = `number of breaches`)
4. name(breaches = `number of breaches`)

Answer = c  
The function we need is rename(new\_name = old\_name) with `` around any name with spaces included.