

The point of this lab is to get more practice writing R code and specifically to practice subsetting and writing for loops and working with data structures in R.

Your answer will consist of a file containing R code `lab10.R`. Please *DO NOT* submit anything other than a `.R` file (e.g., DO NOT submit a Word document or a PDF document or an HTML document).

We will work with eight CSV files included in the `immunisation-by-year.zip` (available on Canvas):

```
immunisation-ethnicity-2011-6-months.csv  
immunisation-ethnicity-2012-6-months.csv  
immunisation-ethnicity-2013-6-months.csv  
immunisation-ethnicity-2014-6-months.csv  
immunisation-ethnicity-2015-6-months.csv  
immunisation-ethnicity-2016-6-months.csv  
immunisation-ethnicity-2017-6-months.csv  
immunisation-ethnicity-2018-6-months.csv
```

These files contain data on the number of children “who have received all of the target immunisations on the National Childhood Immunisation Schedule for their age.”¹ The different files provide data on children at the 6 month milestone for different years.

Within these files, every row provides information for a different District Health Board (DHB). For each health board, the counts are divided up by ethnicity, with two columns for each ethnicity: one showing how many children were eligible for immunisation and one showing how many of those children had completed their immunisations.

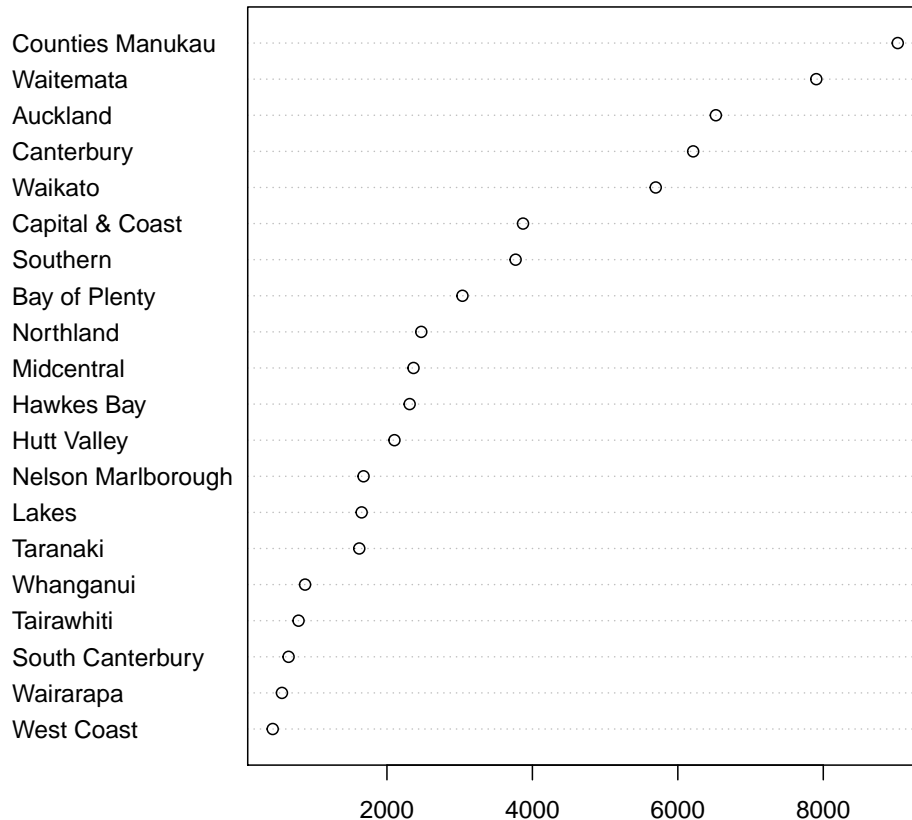
NOTE: You should submit a file containing R code that assigns values to the appropriate symbols. I will run the code in your file and then check the values that have been assigned to the symbols.

NOTE: Your file should **ONLY** contain valid R code, properly **indented**, and with **comments**. You should be able to copy-and-paste your entire file of R code into R and get no errors.

NOTE: You should submit your answers on Canvas.

¹<https://www.health.govt.nz/our-work/preventative-health-wellness/immunisation/immunisation-coverage>

1. We will start by just looking at the data from 2011. The plot below shows the total number of children eligible for immunisation in each DHB. This shows that there are five DHBs that are quite a bit larger than the others.



Write R code to calculate the proportion of immunised children across all “small” DHBs (fewer than 5000 eligible children) and across all “large” DHBs (more than 5000 eligible children).

Your code should produce output as shown below (this is a numeric vector with names).

```

      small      large
0.7082964 0.7106171

```

Hints:

Here are some steps you could take to perform this calculation:

- Subset the “Eligible” columns from the data set.
- Subset the “Immunised” columns from the data set.
- Sum the rows of the two subsets.
- Generate a logical vector that is TRUE when the number of eligible children (per DHB) is less than 5000.
- Subset the per-DHB eligible and immunised counts based on the logical vector.
- sum the subsets and divide immunised by eligible.
- Use `c(small = ??, large = ??)` to give numerics names.

2. Repeat the calculation from the previous question, printing out the result for each year, as shown below.

```

2011: small = 0.71 : large = 0.71
2012: small = 0.74 : large = 0.74
2013: small = 0.78 : large = 0.77
2014: small = 0.78 : large = 0.79
2015: small = 0.79 : large = 0.81
2016: small = 0.8 : large = 0.81
2017: small = 0.77 : large = 0.8
2018: small = 0.76 : large = 0.79

```

3. Write R code to report the number of DHBs where the proportion of immunised children for **ALL** ethnicities is high (greater than 0.7).

```

2011 : 1 DHBs had high immunisation rates for all ethnicities
2012 : 3 DHBs had high immunisation rates for all ethnicities
2013 : 9 DHBs had high immunisation rates for all ethnicities
2014 : 6 DHBs had high immunisation rates for all ethnicities
2015 : 9 DHBs had high immunisation rates for all ethnicities
2016 : 10 DHBs had high immunisation rates for all ethnicities
2017 : 5 DHBs had high immunisation rates for all ethnicities
2018 : 6 DHBs had high immunisation rates for all ethnicities

```

Hints:

The proportion of immunised children is high for all ethnicities if the proportion is high for NZE, **and** the proportion is high for Maori, **and** the proportion is high for Pacific, etc.

[EXTRA for EXPERTS - NO MARKS]

Get the print out from Question 2 to line up properly (extra space or extra zero after the 0.8 values), like this ...

```

2011: small = 0.71 : large = 0.71
2012: small = 0.74 : large = 0.74
2013: small = 0.76 : large = 0.77
2014: small = 0.76 : large = 0.79
2015: small = 0.79 : large = 0.81
2016: small = 0.80 : large = 0.81
2017: small = 0.77 : large = 0.80
2018: small = 0.76 : large = 0.79

```

... or like this ...

```

2011: small = 0.71 : large = 0.71
2012: small = 0.74 : large = 0.74
2013: small = 0.76 : large = 0.77
2014: small = 0.76 : large = 0.79
2015: small = 0.79 : large = 0.81
2016: small = 0.8 : large = 0.81
2017: small = 0.77 : large = 0.8
2018: small = 0.76 : large = 0.79

```