ETC5512: Assignment 2

Prepared by Dr Emi Tanaka

Due date: 11.55pm, May 14, 2021

👨🏿‍🎓 Learning objectives

This assignment is designed to assess whether you

- have honed your "detective skills" of navigating through new data sources,
- can combine data from different sources to compute information, particularly with spatial data,
- write a reproducible report with your solution for each question, in an informative and readable manner.

📮 Turn-in

Please use `assignment2_template.zip` on Moodle as a template. Produce a **reproducible report** and submit it as a "zip" file containing the *single html file* and a *single Rmd file* that is self-contained and compiles without error when placed in the right location in the project structure given in the template provided. The `html` file you submit should be the result of compiling your `Rmd` file. Your `Rmd` should be named as `FamilyName-GivenName.Rmd` where `FamilyName` and `GivenName` replaced with your family name and given name, respectively. If you have a middle name or preferred name that you would like to include, please add these in between `FamilyName` and `GivenName` separated by a hyphen. **You should NOT include your data as the full downloaded data is too big for easy upload and download from moodle.**

This assignment is worth 40 marks in total. The assignment is marked on the quality of the report and the quality of the analysis. **Show all your code for the analysis in the report**. Computed statistics (including plots) should be computed by code from the data and be reproducible. You may need to search the internet to supplement some of your understanding of the data variables. This is an individual assignment and the report that you submit for **assessment must be your own work**.

🔨 Task

The next Federal election is likely to be held in 2022. Certain elected Victorian members of the House of Representative would like to better understand the make-up of their constituents. These elected members in Victoria have collectively commissioned you as an independent analyst[1] to **answer selected questions about their constituents**. The electoral divisions of interest are **Melbourne**, **Kooyong**, **Macnamara**, **Chisholm**, **Higgins**, **Goldstein**, and **Hotham**, as shown in the map below.
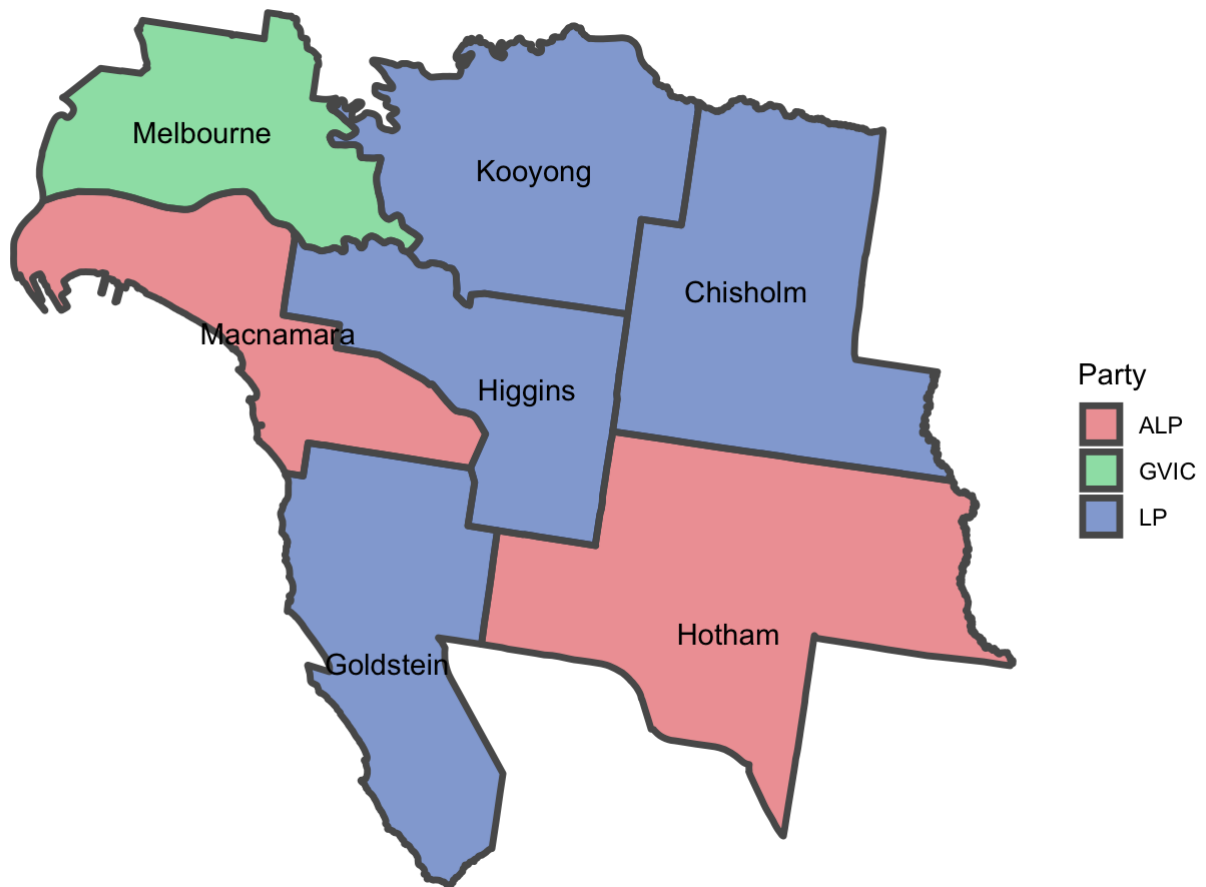
Answer the questions under Task using the *2016* Australian Bureau of Statistics (ABS) Census data and the *2019* Federal electoral boundaries.

Below is a comparison of the SA3 boundaries in 2016 census (in red colour) and the federal electoral boundaries used in 2019 (in black colour). The boundaries do not match up exactly. Some politicians wonder if some gerrymandering occurred.
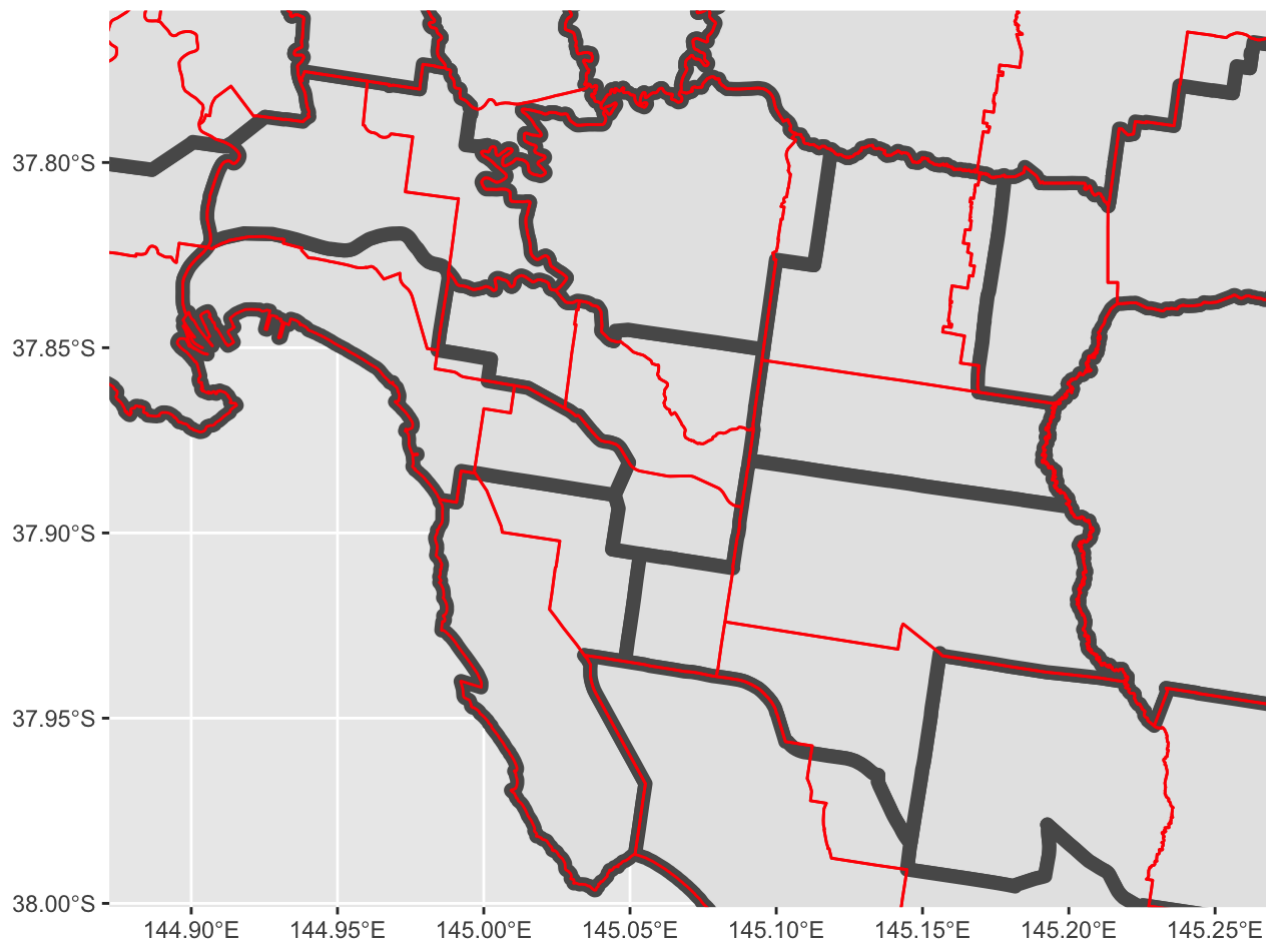
The SA1 boundaries are shown below in red on top of the federal electoral boundaries in black. **For the analysis, use the SA1 census data to compute the information needed. All answers should be with respect to the 2016 census and 2019 electoral boundaries.**
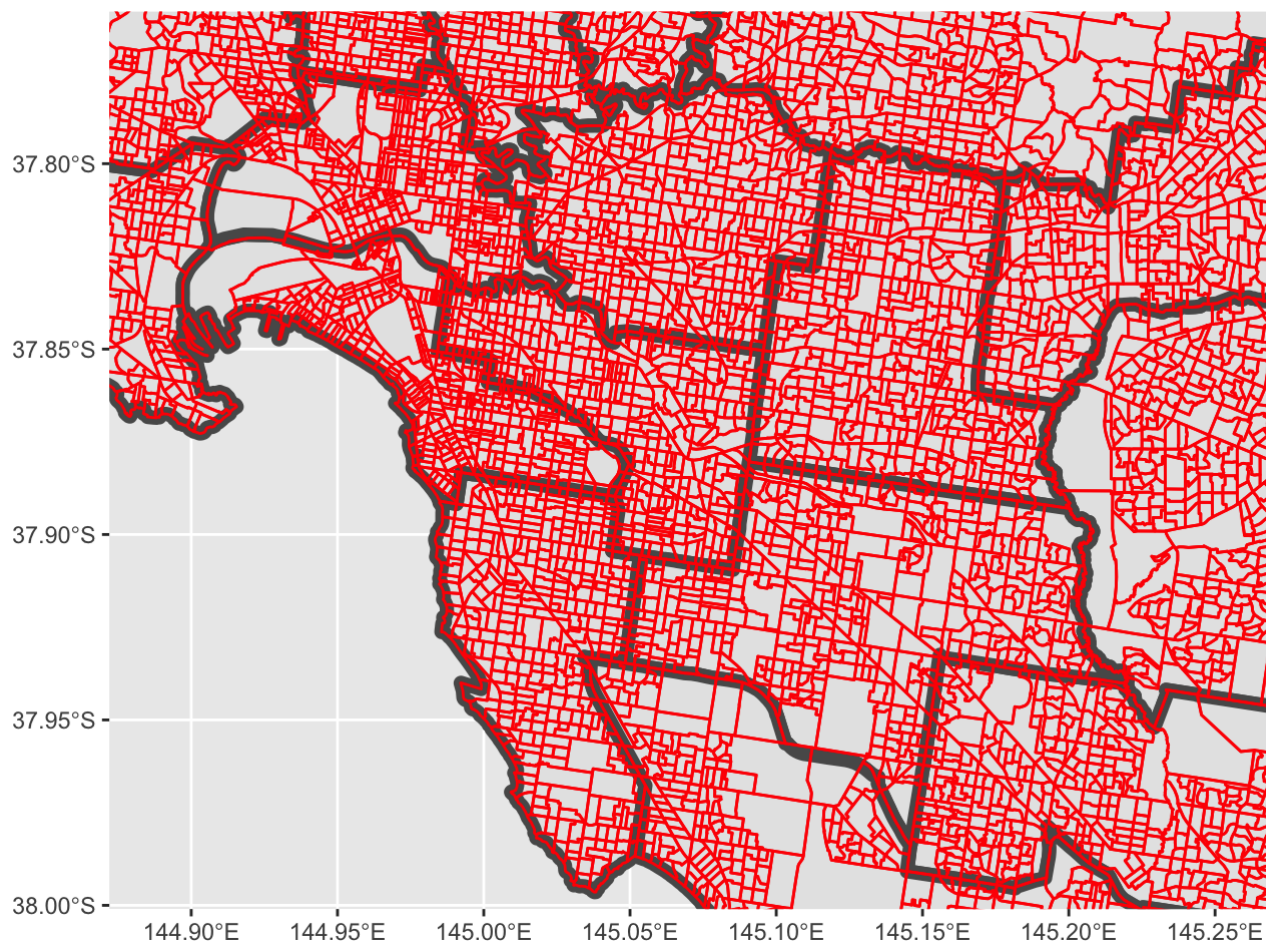
Below code is a hint to match the SA1 census data to an electoral division.

```r
library(tidyverse)
library(sf)
vic_map <- ...
sa1_geomap <- ... %>%
  mutate(centroid = map(geom, st_centroid),
         centroid = st_as_sfc(centroid, crs = st_crs(vic_map)),
         which = as.integer(st_intersects(centroid, vic_map)),
         Elect_div = ifelse(is.na(which),
                            "None",
                            vic_map$Elect_div[which])) %>%
  select(sa1_7digitcode_2016, Elect_div)
```

🔍 Analysis

1. (3 marks) How many people lived in each electoral division of interest?

2. (6 marks) Show the age distribution (omitting those 80 years old or greater) of each electoral division of interest by plotting a barplot like below. State three interesting observations regarding the plot below *with a possibly reason why you see such observations* using knowledge about the electoral division (if you do not know, search the internet to find out).
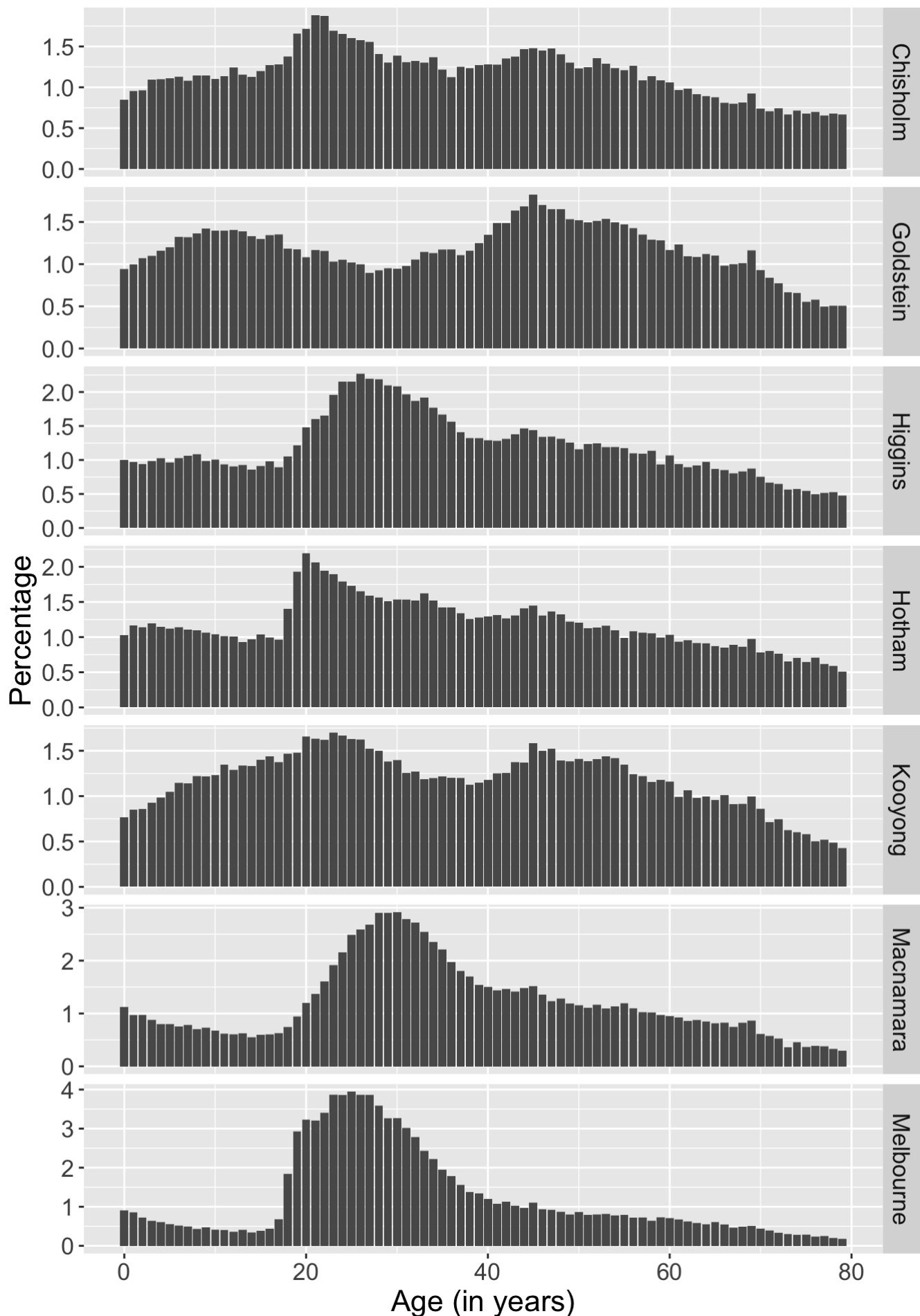
3. (4 marks) What are the percentages of Australian citizens for each electoral division of interest? Why do you think the percentage of Australian citizens is lower for Melbourne, Hotham and Macnamara?

4. (4 marks) What is an estimate of adult (i.e. aged 18 years old or over) Australian citizens in each electoral division of interest? State your assumptions for your estimate.

5. (5 marks) The political members would like to know the composition of ethnic background of their constituents. Show the top 10 reported ancestry for each electoral division of interest with an appropriate graph. State one interesting observation from your graph.

6. (4 marks) What are the distribution of religious background in each electoral district of interest? Show this by using a plot. Report one interesting observation from what you see.

7. (4 marks) What are the level of the highest qualification the person has completed for each electoral division of interest? State one interesting observation with a possible explanation of that observation.

8. (2 marks) List all data **used** for the analysis (including its file name). Do not include any unused dataset.

📄 Report (total of 4 marks)

The report should satisfy the following criteria:

- (1 mark) The report should be written in Rmd and compile to a html report if the data was in it.
- (1 mark) Data should be all located in the `data` folder with no modification of the file content and metadata. While you will not be uploading the data, we will infer this from your output and code.
- (1 mark) Appropriate referencing to all literature, software, and data sources in an academic referencing style.
- (1 mark) Appropriate spelling grammar checks so that the report is high quality.
- (2 marks) Well structured in terms of code placement and style of coding. There should be no redundant code in the analysis.
- (2 marks) A generally well structured report.

Additional resources

If you have questions about writing a report, please consult the Q-manual (https://www.monash.edu/__data/assets/pdf_file/0005/506345/qmanual.pdf).

---

1. This is of course just a hypothetical scenario for your assignment. You are not really commissioned by members of the House of Representative. ↵