

Lab 4: Mini-Assignment

Census Lab Assignment

Aim

The aim of this practical assignment is to investigate a question of your choosing about a population from the 2016 Census data provided by the Australian Bureau of Statistics (ABS), and present your findings.

The population may be Australia as a whole or a smaller region of your choice.

Within that population, you may choose the data of your interest, and investigate an aspect of that data, that helps answer your question.

Learning Outcomes

This assignment demonstrates competencies in:

- sourcing information from a public data repository
- extracting and cleaning information needed to answer a question or hypothesis about the data
- analysing and interpreting the data
- presenting a visualisation of the data to aid understanding and communicate results

Submission

Your assignment (including all explanations and code) must be provided in the CoCalc notebook `CensusAssignment.ipynb` in the current directory. It should contain headings and explanations in markdown cells, and executable code in python cells.

The code will be executed with a fresh kernel for marking, so (as usual) you should ensure that it runs with a clean kernel.

Any supporting files that you use (data files, images if relevant) must be in a text file in the same directory. (Data files should not be more than 5MB. Other supporting files, if relevant, should be no more than 1MB.)

Rubric

The assignment is worth 14 prac marks, contributing to the 35% practical component.

The assignment will be marked for clarity and professionalism of both the exposition and the coding.

Structure & Content [7 marks]

The assignment should be structured as follows:

- **Context** - What is the question that you are seeking to answer? You may provide a short amount of additional information on why this is an important/relevant question.
- **Data acquisition** - A reader should be able to replicate your work. It should be clear what data is used and where it can be sourced. You may include web link(s). If you are extracting only part of the data your code should be accompanied by a brief description of what you are extracting and why.
- **Data cleaning/conversion** - Your data cleaning steps should be accompanied by a brief description of any steps you took to get the data from its raw form into a form you can use.
- **Data analysis/interpretation** - Briefly describe, along with your code, any additional analysis performed on the data (eg sorting, calculations, etc).
- **Data visualisation** - Provide the results of your investigation using a suitable chart.
- **Conclusion** - State your findings in relation to your original question.

Coding [7 marks]

For this assignment you may make use of the `matplotlib` and `numpy` libraries. (There is no requirement to use `numpy`.)

There is no single "right" way to write the code. However, the following should be considered:

- **Clarity** - Your code should be easy to read and comprehend. Considerations should include:
 - use of meaningful variable names
 - use of comments and/or docstrings for key steps/blocks (you do not need to comment every line, this tends to obscure the key steps)
 - use of functions (see below)
- **Conciseness** - You should not pare the code down to a bare minimum at the expense of clarity and readability. However you should try to avoid extraneous code that is unnecessary.
- **Efficiency** - You do not need to achieve ultimate efficiency at the expense of writing clear logical code. However you should avoid obvious inefficiencies such as using an algorithm that takes time n^2 when an algorithm that takes time n will do.
- **Functional decomposition** - Code is preferred where key tasks are implemented as functions, which are combined for larger tasks.

Deadline

The deadline is **10:00 am, Tuesday 8th September**. (This provides an extra weekend on the usual lab deadline in case its needed.)