

772-892 Agricultural Economics Assignment

Now that you have improved your R programming skills, you must apply them.

Step 1: Get your data

- Acquire at least two datasets that contain at least 50 observations each. This can be your own data or data acquired from an online source. Note that you need to join two or more datasets – hence, they need to have a set of suitable ID variables. And the joined dataset must have at least six variables, excluding the join columns.
- Refer to the list of databases below for your options. **Note that you need to provide the dataset's source, and you are not allowed to manipulate the dataset in Excel in any way – all manipulation must be done using R.**
- If you struggle to find a suitable dataset, come see me.

Step 2: Now that you have the data, you need to do the following:

1. [5 marks] Clearly and concisely state your research question that you are trying to answer or enlighten using the steps below. This step does not involve coding
2. [5 marks] You must clean the datasets in preparation for joining.
 - a. Import the datasets,
 - b. reshape the data if needed
 - c. ensure all vectors are in the correct format.
 - d. ensure that all variable names are in lowercase and that they do not contain any special characters, except for spaces that need to be replaced with underscores “_”.
 - e. Join the two or more datasets to create the dataframe called **df**. Larger datasets will get more marks than smaller ones.

Step 3: Once you have joined the data to create **df**, do the following with it:

3. [5 marks] You must create at least two new variables to be used in the subsequent steps.
4. [30 marks] Do a suitable exploratory data analysis on the data. This should include at least three visualisations, but should not be limited to them. Also, this section requires accompanying text to interpret the results.
5. [15 marks] Calculate the relevant summary and other statistics from the data.
6. [15 marks] Create at least two useful plots other than those of the exploratory data analysis.
7. [15 marks] Do a statistical or regression analysis of the data, which could be used for your thesis. Marks will be given for suitability and originality.
8. [10 Marks] These marks are allocated based on the overall impression of your code. The following will be considered:
 - a. Does your code run without a problem?
 - b. How clear and concise your code is
 - c. Are code blocks broken up and well-described

Total: 100

The original data and the accompanying code must be shared with me on OneDrive. If your code fails to run, I will require you to fix it.

Two marks will be subtracted per day for late submissions. Due date: 27 October 23:59

Databases:

- World Bank Databank: <https://databank.worldbank.org/>
- FAO Stat: <https://www.fao.org/faostat/en/>
- Trademap: <https://www.trademap.org/Index.aspx>
- OECD databank: <https://www.oecd.org/en/data.html>
- IMF data portal: <https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42>
- UN data portal: <https://population.un.org/dataportal/>
- South African Reserve Bank Data: <https://www.resbank.co.za/en/home/what-we-do/statistics/releases/online-statistical-query>
- EU Data Portal: <https://data.europa.eu/en>
- Google Dataset Search: <https://datasetsearch.research.google.com/>