Please pay careful attention to the following instructions for this ECA.

Please submit your responses to all questions as a single MS Word document (in .docx extension) and include your RMarkdown script at the end of the document. You may embed your RMarkdown script as an object in your MS Word document. This is the preferred way for you to include your source code. Alternatively, you may copy-and-paste your RMarkdown codes using the "Courier New" font at the end of your document as an appendix.

Do check that your RMarkdown script can be opened as an embedded object and that your report is reproducible from your codes.

Please ensure that your responses to the questions are contained in the Word document you submit. Marks will **NOT** be awarded if your responses are omitted from your Word document, even if they are contained in your RMarkdown script. Therefore, please ensure that your Word document is readable and can be opened. Corrupted Word document will not be graded and will result in a zero score.

Marks will be deducted if your Word document is not generated from RMarkdown, for instance, if your submission is typed entirely using MS Word. However, formatting works done outside of RMarkdown are permission.

In your response, **do** describe what you have done (say, explain the general data cleaning and data wrangling strategy), and if appropriate, include code snippets as well as intermediate and final outputs as support.

When developing the data visualisations, please pay attention to 1) their design and visual appeal, i.e. the visualisations should look clean and not cluttered, 2) the insights that can be uncovered by these visualisations, and 3) the quality of the data story being told, i.e. the message conveyed by the visualisations should be clear, catchy, and easily understood. In other words, rather than merely creating a data visualisation, where appropriate, the visualisation should try to convey a data insight.

For your work, data preprocessing and data cleaning could be required. Please explore the data, clean them or generate new variables if necessary, before plotting.

The R code chunks in your RMarkdown script should be clearly documented to ensure that the logic and purpose of these codes are clear. Clear documentation is especially important for explaining the steps taken for data preprocessing and data cleaning.

Finally, please remember to cite your sources and include a reference or bibliography at the end of your report using the APA citation format. You may use Generative AI tools for minor edits or coding hints. However, please do not use Generative AI tools to generate your response, for instance, by writing in point form and letting the Generative AI tool to produce the entire output. Your report will undergo both plagiarism and AI detection checks. Reports with high AI detection scores will be flagged for potential plagiarism, which may result in an academic penalty.

(Full marks: 100)

Answer All Questions (100 marks)
Answer all questions in this section.

Question 1

At the People's Action Party (PAP) conference on Nov. 24, 2024, Prime Minister Lawrence Wong highlighted the political challenges Singapore is currently facing. Unlike in the past, he emphasised that the PAP can no longer assume there are "safe seats" in the General Election (GE) or that the party will automatically secure a stable government (see "No chance of 'opposition wipe-out' in next GE: PM Lawrence Wong," 24 November, 2024, in https://mothership.sg/2024/11/opposition-wipe-out-ge2025-lawrence-wong/).

In this assignment, you will compose a report to analyse and document the history and recent developments of parliamentary elections in Singapore from a data analyst's perspective. The objective is to present an objective, data factual, and comprehensive overview of Singapore's parliamentary elections since their inception in 1955. Your primary data source will be data.gov.sg.

This assignment is open-ended, and you are free to explore various aspects of Singapore's electoral history with the objective of informing the general public, who will be the main audience of your article. For instance, you may analyse:

- The number of political parties over time,
- The distribution of votes across parties,
- Trends in political parties' performance,
- Changes in the number of seats won by parties.

For discussion on recent developments in Singapore's parliamentary elections, you may use the Mothership article as a guide and to provide some context. Using data on parliamentary election results (see below) or other relevant data, you could identify key trends or factors that might explain the concerns raised by PM Wong in the article. You are also encouraged to conduct your own careful research to enrich your analysis.

For your study, you will use the elections data from the Elections Department (ELD) provided by data.gov.sg (see https://data.gov.sg/datasets?agencies=Elections+Department+(ELD)&page=1&resultId=d_4e7981c19ae3c1e3b7ce3d9842415c8d). Except for electoral boundaries, where geojson data are provided in the zip folder ElectoralBoundary.zip, all data must be extracted via the data.gov.sg API. You are also free to incorporate other datasets from the data.gov.sg repository to supplement your analysis.

Here is an example of an API to extract data on "Parliamentary General Election Results by Candidate", from the url https://data.gov.sg/datasets/d_581a30bee57fa7d8383d6bc94739ad00/view

```
### Constructing the URL
dataset_id <- "d_581a30bee57fa7d8383d6bc94739ad00"
elections.url                                              <-
paste0("https://data.gov.sg/api/action/datastore_search?r
esource_id=", dataset_id, "&limit=10000")

### Extracting the Data
out.elections <- fromJSON(elections.url, simplifyDataFrame
= T) # Fetching the data
df.elections <- out.elections$result$records # Saving the
records
```

Based on your data analysis, you will write a report titled "From Past to Present: An Analytical Overview of Singapore's Parliamentary Elections."

In this report, you are expected to construct data visualisations using the datasets mentioned above, as well as any other relevant datasets from the ELD contained in data.gov.sg. The report should provide a comprehensive overview of the history of Singapore's parliamentary elections and an analysis of recent developments. You should design a range of visualisations, including those that incorporate election boundary data, to highlight significant events or trends. The visualisations must be clean, structured, and visually compelling. Where appropriate, your visualisations should incorporate strong storytelling elements to emphasise key findings or facts effectively.

You may use R packages not covered in the course. When importing external packages, include comments to explain their purpose and enhance clarity and reproducibility.

The report should have the following structure.

1. Executive Summary. The executive summary is a summary of the report, such as the objectives or motivations of the topics covered, the data, and some key conclusions.
2. Introduction. This section provides the background of the study, a brief discussion about the data and your findings.
3. Data. This section contains a description of the data. You may summarise the features of the datasets, discuss the variables, point out general issues and how you have dealt with them, and other details that are relevant for your research on this topic. This section should be brief. Details on your data wrangling work should be relegated to the appendix.
4. This section will contain your main discussions. You may decide on the title of this section as its subsections. You may first conduct an exploratory data analysis to provide an overview of the two data sets. After doing so, you may drill down into specific insights you have gathered from the exploratory

analysis. Here, you may pull in other data from data.gov.sg to support your analysis.

5. Conclusion. This section summarises the topic, your findings, and conclusion.
6. Data Appendix. This section contains detailed discussions about the data. You may discuss the issues you have found about the dataset, e.g. the names of the locations could be coded differently, locations could be missing, missing values could be coded than left blank, etc. You may discuss the steps you have taken to clean up the datasets. You may also discuss feature engineering (i.e. variable generation) and the reasons for doing so.

The suggested total word limit of the report is 4000 words (excluding references). This includes the main report (less appendix) of about 3000 words (excluding references) and the data appendix of about 1000 words. These word limits are guidelines on the length of the main report and appendix and will not be strictly enforced. In other words, you are free to write slightly less or more than the prescribed word limit.

You are to create both your report and data appendix using RMarkdown. Do not write your report directly in a Word document and separately use the R script for your data work. The report writing and data work must be done within RMarkdown. You may craft out an RMarkdown script for your main report and another for your appendix. You may then knit them into separate Word documents and insert the appendix at the end of the main report. If you prefer, you may carry out the entire work (main report and appendix) in a single RMarkdown file. Formatting works outside RMarkdown are permissible.

The data visualisations you construct and the insights you obtain from them are the key components for assessment. The distribution of the scores is as follows:

- Quality of the report – 40 marks
- Quality of the data work and visualisations – 50 marks
- Quality of codes – 10 marks

Quality of the Report (40 marks)

Executive Summary: Clarity and conciseness in summarising the report.

Logical Flow: Clear and logical progression of ideas and analysis.

Insights and Conclusions: Depth and relevance of insights obtained from the data, and the soundness of the conclusions drawn.

Quality of Data Work and Visualisations (50 marks)

Choice of Visualisations: Appropriateness of the visualisations chosen for various data types.

Clarity: Effectiveness in data storytelling and clarity in presenting the data.

Appearance: Overall aesthetic appeal, including good use of design principles and pre-attentive attributes.

Quality of Codes (10 marks)

Intelligibility: Clarity and readability of code chunks in RMarkdown.