Monash University

FIT5147 Data Exploration and Visualisation

Semester 1, 2025

Programming Exercise 2: R (5%)

Please carefully review all the requirements below to ensure you have a good understanding of what is required for your assessment.

1. Due Date

Wednesday 9 April, 4:30 PM

2. Brief

In this assignment you will demonstrate your capability in creating an interactive visualisation page with simple narrative elements using R Shiny. It is an individual assignment and worth 5% of your total mark for FIT5147.

Relevant learning outcomes for FIT5147

1. Perform exploratory data analysis using a range of visualisation tools;
2. Critically evaluate and interpret a data visualisation;
3. Choose an appropriate data visualisation;
4. Implement interactive data visualisations using R and other tools.

3. Details of Task

3.1 Dataset

The data set used in this assignment is based on the AusStage online resource. We introduced this dataset in the Week 1 Workshop and Programming Exercise 1 but have prepared a specific data file for this assignment. This data was initially gathered on 28 February 2025.

To enhance your understanding of the context and metadata, you can check the data source link: https://www.ausstage.edu.au/pages/learn/search-ausstage. Using the various interactive tools for the data source of the full dataset may help enrich your

visual analysis. If you discuss or replicate the visualisations or metadata provided by AusStage, be sure to reference these correctly in your report[1].

We have wrangled the data for this assignment, so you must use the data that we provide on Moodle. This is different from the PE1 data. The name of the data set is **AusStage_S12025PE2.csv**. It contains fewer rows, different columns and it is restructured for the ease of the assignment. You may further wrangle the data if you wish.

For this PE2 assignment, the data set provides information about events performed in the State of Victoria. In this activity we will explore the use of a few attributes:

Column	Description
Event Identifier	A unique number identifying an event in AusStage.
Event Name	The title or name of an event.
First Year	The year of the event's first public presentation, including previews
Last Year	The year of the event's final public presentation.
Primary Genre	The kind of event, as defined by its main mode of performance.
Venue Identifier	A unique number identifying the venue where an event happens.
Venue Name	The name of the venue where an event happens.
Suburb	The suburb or local district where the event happens.

---

[1] AusStage does include a citation guide, but you will have to adapt the formats they supply to those specified for this unit.

Longitude	Geographical Location (longitude) of the venue
Latitude	Geographical Location (latitude) of the venue

Table 1: Fields of the "AusStage_S12025PE2" data set.

References: AusStage. (n.d.). *AusStage: About*. AusStage. Retrieved February 27, 2025, from http://www.ausstage.edu.au/pages/learn/about

3.2 Design Brief

The task is to use R Shiny, ggplot2, and Leaflet to create a **data visualisation page** using the provided dataset and following a specified layout mockup and design requirements.

For this assignment, you will need to create two visualisations: **VIS1** and **MAP**, combined into one layout.

The mock-up shown in Figure 1 shows the expected layout of the visualisation page; and the expected content in each section. In order to create this layout you are expected to use a **fixedPage** (not a fluidPage) layout and are expected to position the visualisation elements in an appropriate number of rows and columns.
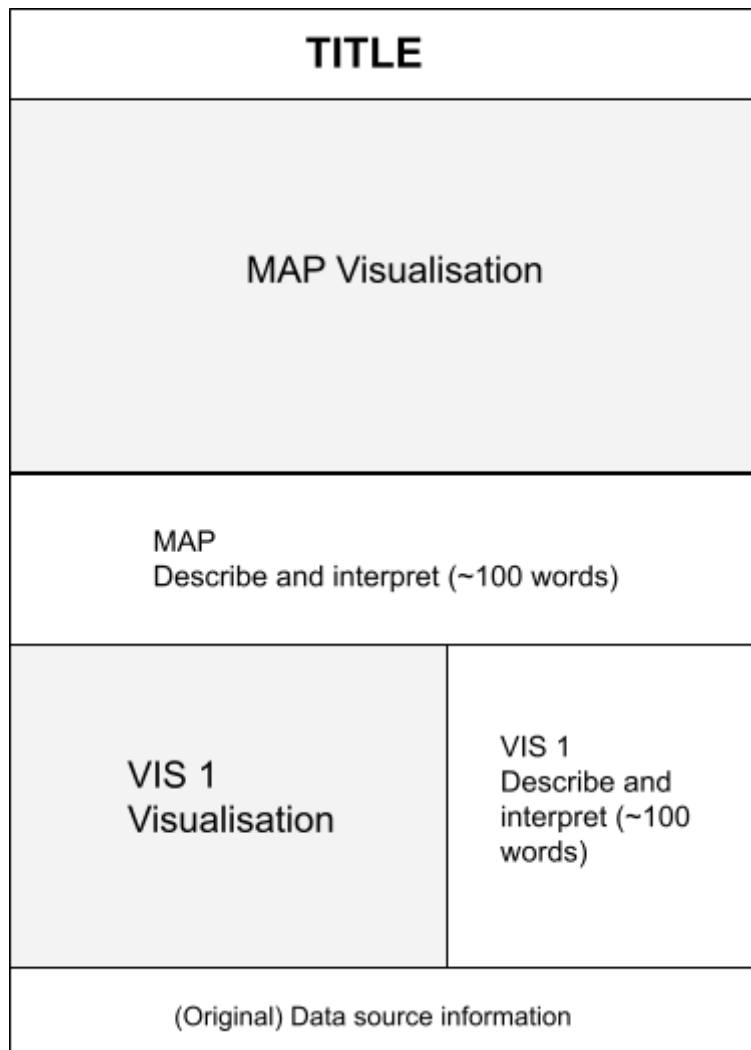
Figure 1. Mock-up showing the approximate position for your two visualisations (Vis 1 and Map) and two descriptions. Word counts are approximate.

3.3 VIS 1

For this assignment, this visualisation should be **static** (not interactive) with the following requirements:

1. **VIS 1** should show the **top 10 most commonly used venues** of events according to the number of *Event_Identifier* values for each *Venue_Name* value. The visualisation must show the **magnitude of the usage** for each venue. Order your VIS 1 from most events to least events.

2. The visualisation must show the **breakdown** of the *Primary_Genre* values for each venue, using a suitable visual variable of your choice.

3. VIS 1 must be created using **ggplot2**.

3.4 MAP

For the **map component**, you create an **interactive** proportional symbol map using *Leaflet*. You must follow the following requirements:

1. Plot the venues on your map using circle markers or equivalent, with the following data mappings and design aspects:

    a. **Each symbol** on the map is for a separate *Primary_Genre*.

    b. **Colour** should be mapped to *Primary_Genre*. You can choose an appropriate colour palette for the type of data.

    c. **Radius** should be mapped to the number of events for that genre at that venue (you may need to scale the size so as to reduce the data overlap on the map, or use opacity, but some overlap is expected)

    d. Provide a **colour legend** for your map.

2. Implement the following interactive features:

    a. Provide a **tooltip** for each circle marker on mouse hover-over that shows the name of the venue, the suburb, the genre, and the number of events for that genre*.*

    b. Add a **numerical slider** (a slider to set a maximum and minimum value) for filtering how many years back the events occurred*.* By default, all venues should be shown on the map, i.e., the sliders' settings should be equal to the maximum and minimum available values. The slider should state which years will be plotted on the map.

3.5 Data Visualisation Narrative

1) Provide a descriptive **title** for your visualisation.

2) The descriptive text in the description boxes should both **describe and interpret** the related visualisations. They should help the viewer see some data insights you have identified, especially when using the interactive features.

3) Information on the original (not the file we are using, but the one it is adapted from) **data source** should be provided for all data visualisations[2]. In the relevant layout location (see Figure 1) briefly provide information about the original data source, including the:

    a) name of the data;
    b) URL to the data;
    c) name of the licensor;
    d) date of the version of the original data used for the visualisation.

3.6 Reflection

Using the provided template, write a brief justification and reflection of your completed visualisation based on the visualisation and visual analytics theory you have learnt in the unit so far. This should include justification of your choice of visualisation, choice of visual variable/s, and a reflection on your visualisation layout and typography. As shown in the template, this needs to be completed for each of the two components: VIS 1 and MAP.

Notes

1. You should avoid explicitly hard coding any of the data such as lists of genre categories or years.
2. Minimise the amount of data processing done by the server code when it reacts to new user interactions.
3. The word counts are a guide. For this textual content we are looking for concise descriptions.
4. The text descriptions are to help your visualisation reader see what you see in the data visualisations, i.e., tell the data story. The reflection is to help your markers understand your design choices using the data visualisation theory and practice taught so far (Weeks 1 to 5).
5. The design choices for VIS 1 and MAP will be included in the evaluation of your work. Make sure you meet the requirements of the assignment, choose any visual elements wisely and try to ensure each component of the layout, including typography, can be justified.
6. Whilst a map legend is usually expected for all map elements, a legend for the size of the proportional symbols on the map is not a requirement for this assignment. (You may provide one if you wish for completeness.)

---

[2] Providing the data source allows users of your visualisation to find and access the data themselves and allows them to better understand the original purpose/ownership of the data.

7. No data checking or cleaning is required, but you may need to perform data transformations. You could use an R package such as *dplyr* (https://dplyr.tidyverse.org/) for this purpose, but it is not a requirement.

8. You may find these settings useful for creating your map (feel free to adapt them for your map needs and design choices): Centre point of latitude -37.8162, longitude 144.962, zoom level 12 and map provider CartoDB.Positron.

9. Whilst there are no requirements on the use of specific colour palettes for this assignment, appropriate use of colour is essential and choice of visual variable should be justified. We recommend the use of *ColorBrewer* palettes: (https://ggplot2.tidyverse.org/reference/scale_brewer.html).

10. The text for your description text boxes, titles and any visualisation element content should be easy to read. Choose a clear font and appropriate font size for your submission and be sure to check your grammar and spelling.

11. Your reflection and code is run through academic integrity checks. No collusion between students is permitted and any text or R code that is largely based on any third party code **must cite the original source** in comments within the R scripts(s) (or as a reference, or footnote in your reflection), including webpages or social media messages. Otherwise your work may be considered to be plagiarising the code of others. No code provided by generative AI can be used in this work. No part or description of this assessment task may be input into a generative AI system.

4. Assessment Resources

See the Assessments section on Moodle for the:
- Data: *AusStage_S12025PE2.csv*
- Reflection Template: *PE2_reflection_template_S12025.docx*

5. Assessment Criteria

The following outlines the criteria which you will be assessed against.

- Demonstrate the ability to read in and manage data effectively using R [0.5%]
- Demonstrate the ability to create static visualisations in R using ggplot2 [1%]
- Demonstrate the ability to create a data map in R with Leaflet [1%]
- Demonstrate the ability to create an interactive visualisation in R with Shiny [1.5%]
- Demonstrate the ability to justify and critically reflect on interactive and static data visualisations [1%]

6. How to Submit

Submit two files:

- **A PDF file** containing the justification and reflection. This reflection must use the template provided. Name the PDF file in this format:
  PE2_[LAST NAME]_[STUDENT ID].pdf.

- **A ZIP file** containing all files required to run your Shiny application. Name the ZIP file in this format:
  PE2_[LAST NAME]_[STUDENT ID].zip.

Before submitting your assignment, double check that your Shiny application runs correctly. To do so, clear objects from the workspace by clicking on the "Broomstick" icon on the top-right section of RStudio. Afterwards, make sure your application is still working by clicking the "Run App" button on RStudio.

The files that you need to include in your ZIP submission are:
- The **one dataset** supplied for this assignment
- **R script(s)** for the final Shiny application (you can use a single R script, or two scripts for UI and Server)
  - o Have all required "library(xxx)" or "require(xxx)" statements at the beginning of your R files (you do not need the code to install the packages)
  - o Use relative paths when reading your dataset (do not use absolute paths that refer to specific drives)

Do not include your PDF file in your ZIP file.

7. Report Word Limitations and Late Penalty

The reflective report is expected to have approximately 100 to 200 words per section, totaling approximately 500 words. The supplied template must be used.

As per Monash policy: All late submissions will receive a penalty of 5% per day (0.25 marks per day out of a total of 5 marks) late inclusive, including weekends. Work submitted more than seven days after the due date will not be marked.