

Monash University
FIT5147 Data Exploration and Visualisation
Semester 1, 2023

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**
2. **Instructions & Brief**
3. **Assessment Resources**
4. **Assessment Criteria**
 1. **Grading Rubric**
 2. **Word Count (& Penalties)**
5. **How to Submit**

1. Due Date

Monday Week 6, 3 April 2023, 9:30 AM

2. Instructions & Brief

In this assignment you are required to create an interactive visualisation using R. The visualisation will show both **spatial** and **tabular** data. **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;
6. Implement interactive data visualisations using R and other tools.

Details of task:

The data set for this assignment relates to a survey of the frog population of the state of Victoria. The dataset is a compilation of Frog Census records (citizen science program) and the preceding Frog Watch program for the Port Phillip and Westernport CMA Region

(<https://discover.data.vic.gov.au/dataset/frog-census-records3>). It was supplied by the Melbourne Water Corporation.

For this assignment, you will be examining which frog genera (plural of 'genus') have been observed in certain parts of Melbourne - the local

government areas (LGAs) of Monash, Knox, Whitehorse and Maroondah. The data has been edited to only include records relating to those areas.

The task is to use *R Shiny*, *ggplot2*, and *Leaflet* to create a data visualisation using the provided dataset. You will also have to process the data in various ways. The Shiny application that you create should be based on the following template layout:

| FROGS OBSERVED IN OUTER EASTERN MELBOURNE, 2000-2018 | |
|--|---|
| [Brief description of the context of the data and project] | |
| [VIS 1] | <div>[RANGE SLIDER]</div> <div>[MAP]</div> <div><div><div></div><div></div><div></div><div></div></div><div>[checklist]</div></div> |
| <u>Frequency and location of frogs</u> [Relevant description of VIS1 & MAP] | |
| <u>Observation time</u> [Relevant description of VIS 2] | [VIS 2] |

There is **one (1)** dataset used in this assignment:

- *PE2_frog_data.csv* contains records of the frogs observed in the 4 LGAs between 2000-2018.

You are expected to:

1. Load the data set into RStudio and transform the data into the appropriate format(s) for you to carry out Steps 2-5.
2. Create an interactive proportional symbol map using **Leaflet** that shows the spatial positions of all observed frog genera in the dataset (**MAP**). This map should:
 - a. Use the provided longitude and latitude values for each genus observed in the data to position the symbols. The longitude and latitude values must be *rounded down to only 3 decimal places*.
 - b. Provide a range selection slider to *filter the number of observations in a location* that is required in order for a symbol to be visible. You must allow a *minimum* and *maximum* value to be set by the slider.
 - c. Provide a checklist to *filter which frog genera to display*
 - d. Encode to each symbol's size *the number of frogs observed for a particular frog genus* at that location.
 - e. Use a visual variable to differentiate between the frog genera.
 - f. When a symbol is clicked, show a tooltip that displays the name of the frog genus and the number of observations of that genus at that location.
1. Create a visualisation using **ggplot2** that shows the *number of observations* made of each frog genera and their *different preferred terrain* (**VIS 1**). The visualisation should display the totals for each of the frog genera.
2. Create a visualisation using **ggplot2** that shows the *hours in which observations occurred* for each of the **top four** frog genera, according to VIS 1. (**VIS 2**). This must clearly show the results each hour for each genus. Colour must be used in some way to clearly distinguish between the data .
3. Compose VIS1, VIS2, MAP, and your descriptions of these graphs into a nice looking layout using **Shiny** (*preferably using a fixed layout, instead of fluid layout*). This should resemble the above template, replacing all the green elements with your visual and textual work, but does not need to follow it precisely. The descriptions must refer to the relevant visualisations, their data, their design, their

use and any important interpretations. A very brief description of the context of the project and data must also be provided on the layout.

NOTES:

1. The *PE2_frog_data.csv* is a different version of the survey's dataset to that used in PE1. **Please make sure you are using the correct dataset for this assignment.**
2. No data checking or cleaning is required, but you will need to perform data transformations and some minor calculations in order to prepare the data and create the required visualisations (see Step 1). You can use an R package such as *dplyr* (<https://dplyr.tidyverse.org/>) for this purpose.
3. You should avoid explicitly hardcoding data like lists of locations or genera in your code.
4. Including a legend that describes size of the proportional symbols on the map is not required. However you should include a legend of some sort when you encode any data using colour.
5. There are no requirements on the use of colour palettes, but *color brewer* (https://ggplot2.tidyverse.org/reference/scale_brewer.html) is recommended should you wish to use it.
6. No collusion between students is allowed and any R code that is largely based on any third party code **must cite the original source** in comments within the R scripts(s), including webpages or social media messages. Otherwise your work may be considered to be plagiarising the code of others.

3. Assessment Resources

See the Assessments section on Moodle for the data.

4. Assessment Criteria

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

- Demonstrate the ability to read in and transform data using R [1%]
- 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 [2%]

As part of the grading process, mandatory interviews to discuss your submission will occur during your tutorial in Week 7. If you do not have a satisfactory interview, your mark and feedback for the assignment will be withheld

5. How to Submit

Submit a zip file containing all files required to run your work. Name the zip file in this format:

PE2_[LAST NAME]_[STUDENT ID].zip.

Before submitting your assignment, please 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 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)
 - 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)
 - Use relative paths when reading your dataset (do not use absolute paths that refer to specific drives)

6. Late penalty

See the late penalty guidelines in the Assessments section on Moodle.