Monash University

FIT5147 Data Exploration and Visualisation

Semester 1, 2022

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

Friday Week 5, 1 April 2022, 11:55 PM

---

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 is based on an Our City's Little Gems study which observed butterfly biodiversity and flower-butterfly interactions in the City of Melbourne between January - March 2017 (https://data.melbourne.vic.gov.au/Environment/Butterfly-biodiversity-survey-2017/kmtd-nvqr). In the original study, the researcher recorded what butterflies they saw, if any, when they walked through various areas of Melbourne. A single site may have different plants and locations in which the study occurred and the researcher visited these sites multiple times during the study. This enabled them to see which types of butterflies they could find, where, when and under which weather conditions. This assignment will only include a modified subset of the original data.

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

SURVEY OF BUTTERFLIES IN MELBOURNE 2017		
[Brief description of the context of the data and project]		
Location of the survey [Relevant description of MAP]	[RANGE SLIDER] [MAP]	
[VIS 1]	**Top Sites for Butterflies** [Relevant description of VIS 1 & VIS 2]	[VIS 2]

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

- *Butterfly_biodiversity_survey_2017_PE2.csv* contains records of the observations within 15 sites around Melbourne during 2017.

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 a visualisation using **ggplot2** that shows only the top 5 sites in the data based on the total number of butterflies observed in 2017 **(VIS 1)**. The visualisation should display these totals for each of the 5 sites.
3. Create a visualisation using **ggplot2** that shows the total number of butterflies observed each day at the same 5 sites in Step 2 over the course of 2017 **(VIS 2)**.
4. Create an interactive proportional symbol map using **Leaflet** that shows the spatial positions of **all** 15 sites in the dataset **(MAP)**. This map should:
    a. Use the average of the provided longitude and latitude values within each site in the data to position the symbols.
    b. Encode the total number of butterflies observed at the site to its symbol's radius.
    c. Provide a range selection slider to filter the visible symbols according to the total number of butterflies that had been observed at the site in 2017. You must allow a minimum and maximum value to be set by the slider.
    d. When a symbol is clicked, show a tooltip that displays the name of the site and the total number of butterflies observed over 2017 at that site.
5. 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, 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 must also be provided on the layout.

NOTES:

1. The *Butterfly_biodiversity_survey_2017_PE2.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 create

the required visualisations (see Step 1). You can use an R package such as *dplyr* (https://dplyr.tidyverse.org/) for this purpose.
3. Including a legend that describes the proportional symbol map's size is not required. However you should include one if you decide to encode data using colour (which is optional).
4. 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.
5. No collusion between students is allowed and any R code that is largely based on third party code must cite the original source in comments within the R scripts(s), including webpages or social media messages.

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.

---

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)

6. Late penalty

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