Monash University

FIT5147 Data Exploration and Visualisation

Semester 1, 2022

Programming Exercise 1: Tableau Public (5%)

Programming Exercise 1: Tableau Public

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, 18 March 2022, 11:55 PM

---

2. Instructions & Brief

In this assignment you are required to read in some data and explore and visualise it using Tableau Public/Desktop, then submit a brief report showing your findings and the visualisations you used. **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.

For this assignment however, we will focus on reviewing the data collecting process in the original study. As such, it is a study of the study! The data used for this assignment is not identical to the original data as it has been edited in various ways (so we don't recommend you look at the original data or findings). Firstly, the data has been simplified by omitting various fields and records. Secondly, a few known errors in the original data have been removed or corrected**.** Thirdly, some of the values have been merged and simplified. For this PE1 assignment, the resulting data describes where and when zero or more butterflies (of any type) were seen, and what stage of their walk the researcher was at when they recorded the data. For you, the most crucial change to the data is that **three new types of data errors have also been included in the data**. Part of this assignment is for you to find, describe and handle them. This modified dataset can be found on Moodle in the Assessments section under the Programming Exercises heading.

The task has two components: **data exploration using Tableau**, and a short **written report**.

Data Exploration using Tableau
You are expected to:

1. Load the dataset in Tableau Public/Desktop
2. Use data visualisation to **check for and find** the three aforementioned errors in the dataset
3. **Correct** these errors using any tool of your choice (e.g., Excel, Python, R, Tableau)
4. Using visual analysis, answer the following questions:
    - **Q1: Compare and contrast the number of butterflies observed in each area (Site). In what ways does this show that the sites are similar and/or different to each other? Consider this on both an hourly and day-of-the-week timescale.**
    - **Q2: Compare and contrast the number of records for each area (Site) per hour. What does this tell you about the use of the sites for the study? In what ways does this support, challenge or change your conclusions for the first question?**
5. Write a report that describes this data exploration process (see below for details)

Written Report
Once you have finished your data exploration, write a report that contains the following information:

1. Data checking and cleaning (i.e., Steps 1 to 3)
    - An image to show what your data looks like after loading it in Tableau
    - A brief explanation (max one paragraph per error) and accompanying image of each of the errors that you have found, showing how you found them using Tableau, and explaining & justifying how you resolved them

2. Data exploration (i.e., Step 4)
    - Your answers to Q1 and Q2 with accompanying images of the data visualisations that you used to support your analysis, and a brief explanation of why you have used those type of visualisations

The report should also:

- Be submitted as a PDF file
- Be no more than 5 pages in length, including figures, with a minimum font size of 10 (title page is excluded from page limit)
- Be properly structured with headings, subheadings, figure captions, page numbers, and references (if appropriate)
- Have high quality images of your visualisations with clearly readable and legible text/labels
- Not include any code snippets

3. Assessment Resources

Butterfly_biodiversity_survey_2017_PE1.csv - https://lms.monash.edu/mod/resource/view.php?id=10059467

---

4. Assessment Criteria

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

- Demonstrated ability to check and clean data and read into Tableau Public [1%]
- Demonstrated ability to visually explore data using Tableau Public [2%]
- Demonstrated ability to see trends/patterns in data [1%]
- Quality of report [1%]

As part of the grading process, mandatory interviews to discuss your submission will occur during the Tutorial in Week 5.

---

5. How to Submit

Once you have completed your work, take the following steps to submit your work.

1. Save your report as .pdf.
2. Name you file using the following structure **PE1_Surname_StudentID**
3. Click the **Add Submission** button below to submit and upload your report

*Please note that your assignment MUST show a status of "Submitted for grading" before it can be marked. Please read the **Submission Status** section under Assessments for more information. Drafts will not be marked.*

6. Word Count & Late penalty

The report must not be more than 5 pages of graded material including figures (min. font size 10). 1 additional page may be used as a title page. 1 mark will be deducted if the report does not meet these requirements.

The penalty of 0.5 marks (out of a total of 5 marks) per day will be applied for all late submissions, including weekends.