Assignment Template

2024-11-08

The following is a template .rmd RMarkdown file for you to use for your homework submission.

Please Knit your .rmd to a PDF format or HTML and submit that with no identifiers like your name.

To create a PDF, first install tinytex and load the package. Then press the Knit arrow and select "Knit to PDF".

QUESTION 01: Data Visualisation for Science Communication

Create a figure using the Palmer Penguin dataset that is correct but badly communicates the data. **Do not make a boxplot***.*

Use the following references to guide you:

- *https://www.nature.com/articles/533452a*
- *https://elifesciences.org/articles/16800*

Note: Focus on visual elements rather than writing misleading text on it.

a) Provide your figure here:

b) Write about how your design choices mislead the reader about the underlying data (100-300 words).

QUESTION 2: Data Pipeline

Write a data analysis pipeline in your .rmd RMarkdown file. You should be aiming to write a clear explanation of the steps as well as clear code.

Your code should include the steps practiced in the lab session:

- *Load the data*

- *Appropriately clean the data*

- *Create an Exploratory Figure (**not a boxplot**)*

- *Save the figure*

- ***New**: Run a statistical test*

- ***New**: Create a Results Figure*

- *Save the figure*

An exploratory figure shows raw data, such as the distribution of the data. A results figure demonstrates the stats method chosen, and includes the results of the stats test or model.

Between your code, communicate clearly what you are doing and why.

Your text should include:

- *Introduction*
- *Hypothesis*
- *Stats Method*
- *Results*
- *Discussion*
- *Conclusion*

You will be marked on the following:

a) Your code for readability and functionality

b) Your figures for communication

c) Your text communication of your analysis

Below is a template you can use.

Introduction

```
# Make sure your code prints.
```

Hypothesis

Statistical Methods

```
# Make sure your code prints.
```

Results & Discussion

```
# Make sure your code prints.
```

Conclusion

QUESTION 3: Open Science

a) GitHub

*Upload your RProject you created for **Question 2** and any files and subfolders used to GitHub. Do not include any identifiers such as your name. Make sure your GitHub repo is public.*

GitHub link:

You will be marked on your repo organisation and readability.

b) Share your repo with a partner, download, and try to run their data pipeline.

Partner's GitHub link:

*You **must** provide this so I can verify there is no plagiarism between you and your partner.*

c) Reflect on your experience running their code. (300-500 words)

- *What elements of your partner's code helped you to understand and run their data pipeline?*

- *Did it run? Did you need to fix anything?*

- *What suggestions would you make for improving their code to make it more understandable or reproducible, and why?*

- *If you needed to alter your partner's figure using their code, do you think that would be easy or difficult, and why?*

d) Reflect on your own code based on your experience with your partner's code and their review of yours. (300-500 words)

- *What improvements did they suggest, and do you agree?*

- *What did you learn about writing code for other people?*

e) What are the main barriers for scientists to share their data and code, and what could be done to overcome them? (500-700 words)

- Maitner et al. Code sharing increases citations, but remains uncommon. https://doi.org/10.21203/rs.3.rs-3222221/v1
- Trisovic et al. A large-scale study on research code quality and execution. https://rdcu.be/dZB7x
- A Rock-Star Researcher Spun a Web of Lies—and Nearly Got Away with It. https://thewalrus.ca/a-rock-star-researcher-spun-a-web-of-lies-and-nearly-got-away-with-it/
- Gomes et al. Why don't we share data and code? Perceived barriers and benefits to public archiving practices https://doi.org/10.1098/rspb.2022.1113