

Assignment #4: All About Data

6/28/2024

100 Points Possible

Attempt 1



In Progress

NEXT UP: Submit Assignment



Add Comment

Unlimited Attempts Allowed

Details

This exercise is a solo exercise. It's a *choose your own adventure*.

Overview

You have two main options for the exercise this week:

- Option 1: Try your hand at a bit of an exploration of some complex data type.
- Option 2: Generate and explore synthetic data.

For either option, I expect well-documented code and lots of explanatory text in the Quarto document.

While this is an open-ended exercise and there are no specific requirements as to what you need to produce, I expect more than just a single figure or 10 lines of code. Since you are getting better each week, it should definitely be at least as comprehensive as last week's exercise 😊.

Setup

This exercise doesn't have an existing folder in your portfolio. Make a new folder, call it `data-exercise`. Copy one of the existing Quarto files into that new folder (e.g., the one called `tidytuesday-exercise.qmd`), rename it to `data-exercise.qmd` and open it (make sure you open your portfolio by clicking on your R project file, then open the Quarto file inside R Studio).

Option 1

Find some data that falls into the categories described in the *Complex data types* unit. It could be some text or images or video or audio or any other complex data. Basically anything that's not rectangular data with rows as observations and columns as variables.

Then look for some suitable R packages that allow you to explore the data. You don't need to do any statistical analysis at this stage. Focus on loading the data, processing as needed, and then do some descriptive/exploratory analysis, e.g. by making plots or tables or... Basically anything that you can think of.

Option 2

Write code that generates a synthetic dataset. This dataset should have multiple variables, and there should be some associations between variables.

Then write code that explores the data by making plots or tables to confirm that your synthetic data is what you expect it to be.

Then fit a few simple models to the data. For instance, use the `lm` or `glm` functions to fit a linear or logistic model. Make sure your model can recover the associations you built into the data. Explore if and how different models might be able to capture the patterns you see.

Take some inspiration from the examples shown in the Synthetic Data module.

Wrap-up

Once you got your exercise done, you need to add it to your website. To that end, open the `_quarto.yml` file. In the section that says `menu` and lists several of your projects, add lines at a `-text:` and `file:` entry that points to your new folder/file. **Be careful to have exactly the same number of empty spaces at the beginning of the line as all the other lines.**

While you are in that file, if you haven't done so, update other parts, e.g. such as the title and the github URL. See the instructions for the first exercise if you need a reminder.

Once all done, render your website to make sure everything looks ok and your new exercise shows up. Once things looks fine on your local computer, push your updates to GitHub and check that they also look ok online.

Since this will be part of your portfolio site, and you already posted a link to that previously, you don't need to post anything, I know where to find it.

Enter Web URL

https://



(<https://utsa.instructure.com/courses/42465/modules/items/2138731>)



(<https://utsa.instructure.com/courses/42465/modules/items/2138731>)

