Pattern SWE Interview - Visualization Take-Home Assignment

Last updated November 23, 2022

Described below is a coding exercise that uses data from JSON files to generate a plot and make various modifications to it. Please use the provided data to answer the questions below using JavaScript.

The purpose of this exercise is simply to gauge how you develop software as a whole, including design and documentation. We'll be paying attention to how you strike that balance.

If you use any other external JavaScript libraries while going through this exercise, please make sure that these libraries are included in your HTML file(s) when you submit this assignment. You should feel free to refer to any resources you wish to complete this coding exercise, including Google and StackOverflow.

How long should you spend on this? The expected time needed to complete this exercise is approximately 3 - 4 hours. However, you may spend as much time on it as you'd like. We know you have other things going on in your life, so if you're not finished after a few hours, that's absolutely fine—submit as much as you have. As previously noted, this exercise will be used to gauge how you develop software as a whole. In particular, we will value a clean, robust, and reusable solution that is well documented over something that is hard to read. Ideally, your solution will properly parse the data and render the plots specified in the problems below. If there are improvements you would make to your code given more time, please document it for our consideration.

If you have any questions along the way please do not hesitate to send them to Ben at bcooley@broadinstitute.org.

Your deliverables should include:
- relevant source code
- documentation on how to run/use your program to demonstrate all of the functions
- any relevant supplementary files such as tests or documentation

When complete, share an accessible git repo or create a tarball or zip archive of your files and email it to bcooley@broadinstitute.org.

Overview

PTEN is a gene that encodes instructions for creating an enzyme that is a known tumor suppressor; its function is commonly lost in human cancer tissues. Here, we have provided expression data of this gene from hundreds of samples in over 50 human tissues. We would like you to write a simple summary plot that visualizes this data.

You have received two files with this assignment:
1. `PTEN_gene_expression.json` - PTEN gene expression levels across different samples. Multiple samples may have been collected for a single tissue site.
2. `tissue_info.json` - Metadata about the tissues in this dataset

Problem 1

Write code that parses the PTEN expression data (`PTEN_gene_expression.json`) and calculate the median expression value of each tissue. Then, visualize the median of each tissue in a chart (e.g. bar chart). You can choose whichever plot type you think would be best. Our team often uses d3.js for visualizations, but you are not required to use d3.js if you are more comfortable with another library.

Problem 2

Looking at the `tissue_info.json` file, you will see that each individual `tissueSiteDetailId` has a color assigned to it. Update the code you wrote and chart you created in problem 1 to use that tissue color as the color for each corresponding bar.

Problem 3

When exploring this sort of data, it's useful to be able to sort it in different ways. Please write code that updates your chart, sorting by:
   a. `tissueSiteDetailId`, ascending alphabetically
   b. Median gene expression, descending by value

For each type of sort, add a button that when clicked, will sort the chart according to that attribute.

Problem 4

Is there anything you can do to ensure this page works well for different screen sizes? How about on mobile devices?

Food for thought

No code is required for these next questions, but we are interested to hear your thoughts about them!

1. A simple chart showing the median for each tissue doesn't let us see sample-level variations in the provided data set. Suggest one or more visualization ideas that will allow us to see how PTEN's expression varies across tissues at the sample level.
2. Researchers may be interested in seeing how the expression levels vary in different samples, depending on donor demographics, e.g. sex or ethnicity. Would you use the same visualization ideas, or something different?