BMI 706 Homework 2

Learning Goals

- Learn how to link interactive plots using Plotly and the `crosstalk` R library
- Deepen your understanding of the Five Design-Sheet approach

Problem 1 (30 points)

For this problem, you will use R to implement interactive Plotly visualizations that support basic brushing and linking and can be manipulated using external controls such as checkboxes, sliders, and dropdown lists.

Please obtain the following from the [Resources section on Piazza](#):

1. The data set to be used for this problem is the same as for Homework 1:
   Assignment 1 Data - flunet2010_11countries.csv
2. Code to wrangle the data as well as code for a partial solution of Problem 1.1:
   Assignment 2 Sample Code
3. A demo and sample solution that illustrates the interactive plots:
   Assignment 2 Demo - Interactive Visualizations

Please review the demo before starting to work on this problem. The questions below are referring to the demo. And please note that the interactions between `crosstalk` controls and Plotly plots are not bidirectional. Selection of elements or time ranges in a plot is not reflected in the controls, however, any change in the controls is reflected in the plots. This is a limitation of the `crosstalk` library.

The demo and sample code contain references and pointers to methods and settings to be used to implement the solutions. Nonetheless, it is important that you familiarize yourself with the documentation for Plotly and `crosstalk`.

The questions below ask you to create the visualizations shown in the demo. Your solutions can either implement them exactly as shown or contain variations that meet the requirements set forth in the questions.

1.1 (10 points)

Implement two linked plots like the ones shown in *Visualization 1* of the demo to visualize flu cases per week per country (as a line chart) and total flu cases per week (as a bar chart). Use appropriate color maps. The combined plots should share a common x-axis. Note that the bar chart in the demo is a stacked bar chart, even though it does not look like one.

Implement the following interactions:

1. A dropdown list to support selection of the countries to be shown. This is already implemented in the scaffold.
2. A slider to filter the weeks shown in the two plots.
3. Selecting a week and country in either the line chart or bar chart should highlight the corresponding week and country in the other plot.

Notes:

1. Once a selection is made in the country list, it is not longer possible to make selections in the Plotly plots. It is ok for your solution to have this limitation.
2. The stacked bar chart is behaving somewhat unexpectedly when a selection is made, as the selected bar segment is duplicated and placed on top of the current stack. The y-axis is rescaled accordingly. This appears to be a limitation of Plotly subplots.

1.2 (10 points)

Implement a stacked bar chart (detail plot) with a range slider (overview plot) like the one shown in *Visualization 2* of the demo to visualize flu cases per week per country (as a stacked bar chart). Use an appropriate color map.

Implement the following interactions:

1. A list of checkboxes list to support selection of the continents whose countries should be shown.
2. A range slider in the overview plot to filter the weeks shown in the detail plot.

Notes:

1. The stacked bar chart is behaving somewhat unexpectedly when a selection is made, the selected bar segment is duplicated and placed on top of the current stack. The y-axis is rescaled accordingly. This appears to be a limitation of Plotly subplots.

1.3 (10 points)

Implement two linked plots like the ones shown in *Visualization 3* of the demo to visualize flu cases per week per country (as a line chart) and total annual flu cases per country (as a bar chart). Use appropriate color maps.

Implement the following interactions:

1. Selecting a country in either the line chart or bar chart should highlight the corresponding week and country in the other plot.

1.4 (0 points - optional)

Optional: Implement a choropleth map like the one shown in *Visualization 4* of the demo to visualize flu cases of a given week. Can you come up with a better way to do the color scaling? You might also want to try to implement a Plotly range slider to control the week shown on the map instead of the dropdown menu.

What to submit?

1. An R Markdown file with all source code that you used to create the plots.
2. A knitted HTML file including the interactive plots.
3. *Optional:* Short video or GIF that shows the interactive features of your plots.

Problem 2 (20 Points)

In this problem you will be applying the Five Design-Sheet Methodology to explore the design space for a visualization tool that supports cancer researchers (this is purposely broad) in the interpretation and exploration of longitudinal genomics and clinical data (this is purposely broad as well).

You can find short and expanded instructions as well as sample layouts for the three types of design sheets on http://fds.design/index.php/resources-and-publications/.

A sample data set for the problem that you are supposed to address can be found in the Resources section on Piazza and is named "*Assignment 2 Data - Five Design-Sheet Methodology Exercise*". This spreadsheet contains two tabs: one for patient-level data and one for sample-level data. They can be linked on the Patient ID column. Most patients have two samples (taken at two time points), while some have more. The column titles are self-explanatory. IDH1 mutation values are a categorical variable and correspond to changes in the amino acid sequence, where the arginine (R) at position 132 is replaced with either a Cysteine (C), Glycine (G), or Histidine (H).

This data set is derived from Johnson, Mazor *et al.* (Science 2014): https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3998672/. This is only for background if you are interested, we don't expect you to read the paper.

Here are a couple of sample tasks that your visualization tool could address:
- How is treatment related to survival time?
- Is there any correlation between any of the variables, e.g. is there a correlation between patient age and mutation count?
- How many patients have a higher grade tumor in Timepoint 2 than in Timepoint 1?
- What do the distributions of variables look like (e.g. age, mutation count, etc.) and what do they look like if the cohort is stratified by a categorical variable, e.g. treatment?

- Are there any common trajectories for certain groups of patients, e.g. do all patients treated with chemotherapy end up with a high mutation count?
- ...

Note: As long as you follow the instructions for the five design-sheet methodology and apply what you have learned about visual encoding, there is not much that can go wrong in this exercise.

2.1 (4 points)

Complete Sheet 1 (Brainstorming). Here you are generating as many ideas as possible before converging towards a solution on the next sheets.

2.2 (12 points)

Complete Sheets 2, 3, and 4 by developing three of your best ideas from the brainstorming phase in more detail.

2.3 (4 points)

Complete Sheet 5 (Realization Design). This is your best idea developed in depth.

What to submit?
1. Scans or high-resolution photos of your five sheets.

Administrative

When is it due?
- Sunday, April 8th, 11:59 pm

How do I hand in my solutions?
- Email required materials to bmi706.2018@gmail.com.

What if I have questions?
- Please use Piazza (https://piazza.com/harvard/spring2018/bmi706/home) to ask questions about this assignment. Please avoid sending messages for clarification directly to the instructors.