**CPSC 446/546**
**Assignment 3**
**Due 10/13/2021, 11:59 pm**

Upload to Canvas as a zip file named *yourfirstname_yourlastname*_3.zip.

This assignment requires you to develop visualizations using D3. Do your own coding using code provided with the assignment and examples given in the Scott Murray textbook. **Do not use any code from the internet that you may find that creates visualizations similar to those required in problems.** If we find that code you use for a solution is taken from an internet source, you will receive a zero for the entire assignment.

Be sure to reference d3.js in the same directory, i.e. use this in your html:

```
<script type="text/javascript" src="d3.js"></script>
```

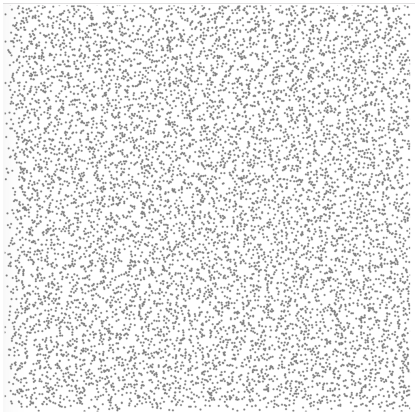
As in previous assignments you may use
https://developer.mozilla.org/enUS/docs/Web/JavaScript/Reference or https://javascript.info/
to do tutorials and look up syntax.

Note that some parts of the questions are for all students and some are additional work for **CPSC 546** only.

1. Color Interpolation (10 pts)
**(CPSC 446 and CPSC 546)** The colors that are in-between two end point colors when linear interpolation is used depend on what system is used to represent the colors. Running the file *color_interpolation.html* allows you to input two colors as RGB and view the colors interpolated between them in RGB and in HSL. You can also control the number of discrete color bars used N. To answer these questions report your answers along with screenshots of the application supporting your answers in a *docx* or *pdf* file.
a.) Find a pair of beginning and ending colors that show a significant difference in the interpolated values in the two color systems. The beginning and ending colors should be significantly different -- the beginning and ending values of R, B, and/or G should differ by at least 100. Which interpolation would be best for visualizing a continuous scalar variable? Why?
b.) Find a pair of beginning and ending colors that show NO significant difference in the interpolated values in the two color systems. The beginning and ending colors should be significantly different -- the beginning and ending values of R, B, and/or G should differ by at least 100.
c.) What is the minimum number of color bars N that you need to use to make the interpolation look smooth? Does it depend on the beginning and ending color values? Is N the same for interpolation in both color spaces?

2.Pre-attentive Effects (25 pts) :
**(CPSC 446 and CPSC 546)** Using D3 create two different scatter plots for the data in testp2.csv (you can just do this by appending circles to the svg.) In each scatterplot use the given x and y values to position circles for each data point, and then use another attribute to encode the third value p. In one of your scatter plots use color, in the other use size. In both cases make your

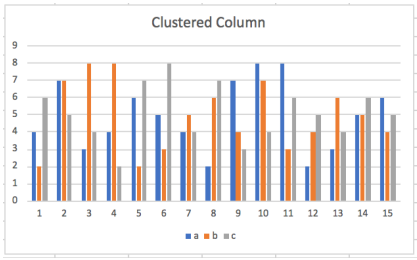
visualization interactive so you can click on circles and see the data that they are encoding. You should examine your two results to find:
a) Two distinctly different regions. What is the shape of the boundary?
b) Three outlier points. Where are they, and what are their values?
c) Which encoding (color or size) was most useful for answering a) and b)? What decisions did you need to make for each plot to get good results?
Here is a plot with the data plotted with uniform circle size and color. The two regions and outliers are not visible.

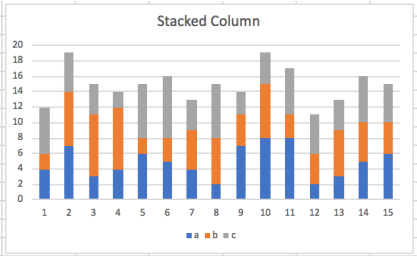


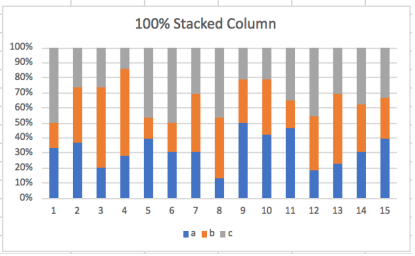
**(CPSC 546)** Use a third type of encoding (other than circle size or color) to encode the value of p. Show your results and comment on whether the two regions and the outliers are more or less visible with this third encoding.

3. Transitions (25 pts):

**(CPSC 446 and CPSC 546)** Using D3 create a visualization of the data in the file barinput.csv. Your visualization should transition from a clustered column, to stacked column, then to 100% stacked column when a user clicks on the page. You should include scales. The goal of these transitions is to make it clear to the viewer how the representation of each data element in one visualization maps to its representation in the next. Set your timings to make these mappings clear. Here are examples of what the column visualization styles look like (these were made in Excel).




4. Nobel Prize Winners (40 pts):

**(CPSC 446 and 546)** The file Jones_Weinberg_2011_PNAS.csv contains data that accompanies the paper “Age dynamics in scientific creativity” by Jones and Weinberg (paper available under Files/Readings). Without recreating the plots used by the authors, develop at least three *different* visualizations in D3 of the data that examine some of the issues discussed by the authors – the

age of the researcher when they did prize winning work as a function of when the prize was given, the field they work in and whether the work was experimental (0 in the spreadsheet) or theoretical (1 in the spreadsheet). Your visualizations can be different idioms (bars, scatter, heat map etc.) and/or can be different combinations of the given attributes. Your visualizations should be interactive so that additional detailed data is available in each visualization by either hovering over or clicking on elements in the visualization.

Women who have won Nobel prizes in these fields (chemistry, physics and medicine) are listed at https://www.nobelprize.org/prizes/lists/nobel-prize-awarded-women-3-2/
Modify one of your visualizations to examine whether these women were outliers with respect to the age effects found for winners overall.

For each visualization, explain:

1. What question(s) are you trying to answer? (Domain situation) For example, are you comparing trends in the different fields, or looking for outliers across all fields?
2. What data do you need to answer the question and did you need to perform any data transformations? (Data/task abstraction)
3. How did you choose to display your attributes? (Visual Encoding and Interaction Idiom)

**(CPSC 546 only)** Provide additional visualizations using the data in Jones_Weinberg_2011_PNAS-citations.csv to consider the issue discussed in the paper.