CPSC 446/546
Assignment 2
Due 9/27/2021, 11:59 pm

Upload to Canvas as a zip file named *yourfirstname_yourlastname_*2.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 Assignment 1 do not worry about getting pixel accurate results to match examples. 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. Bar charts (20 pts) :
(CPSC 446 and CPSC 546) Create a file *LetterFrequency.html* that uses D3 to create a bar chart for percent frequency by reading the data in letter_frequency.csv (original source https://people.sc.fsu.edu/~jburkardt/data/csv/csv.html). You should have one bar for each letter of the alphabet, with a label on the horizontal axis for each letter. You should have a vertical axis for the percent frequency. Here is an example (made in a different software system.) Yours does not have to be exactly the same:
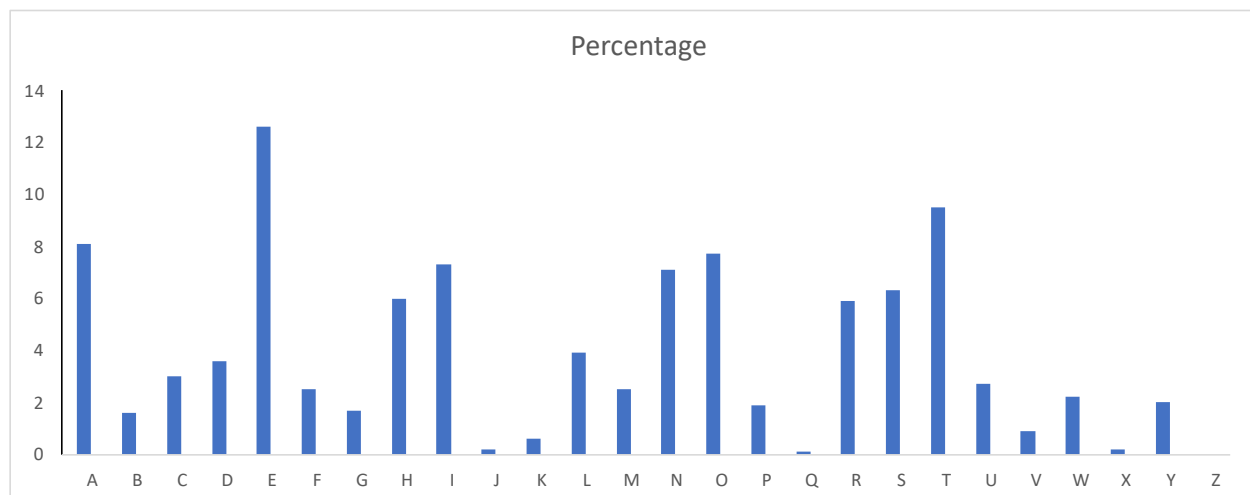
2. Scatter plots (20 pts):
(CPSC 446 and CPSC 546) Create a file *MLB.html* that uses D3 to create a scatter plot from the data about Major League Baseball teams in 2012 that is given in mlb.csv (original source https://people.sc.fsu.edu/%7Ejburkardt/data/csv/mlb_teams_2012.csv) You should plot payroll in millions on the horizontal axis, and number of games won on the vertical. Your plot should have scales and labels. Color code the data according to whether the team is in the National League (N) or American League (A). Here is a poorly scaled, uncolored, unlabeled version I made with different software.


3. Course Table (30 pts):
(CPSC 446 and CPSC 546) In coursetable.com, choose a Yale department, and create a csv file of the department's current courses with their ratings and workload. Note that this will take cutting and pasting -- there is no convenient download option on the site. Create a file *CourseRatings.html* that uses D3 to create a labeled heatmap visualization of rating versus workload that shows the fraction of courses in each of these 25 bins.

To get a 5x5 grid, the bins should range from 1 to 1.5, 1.5 to 2.5, 2.5 to 3.5, 3.5 to 4.5 and 4.5 to 5. Make any labeling changes you feel would improve the legibility of the visualization. You can also make workload horizontal and rating vertical. Be sure to add a title to the visualization.

(CPSC 546 only) **F**rom coursetable.com data, create a file *CourseComp.html* to use D3 to plot a stack of data for the courses, starting with the lowest numbered course on the bottom. Plot a bar to the left for workload, and to the right for the rating. This will show any trends for workload and ratings with course level. The example below was made in Powerpoint. Code your visualization in D3, and choose your own labels and colors.
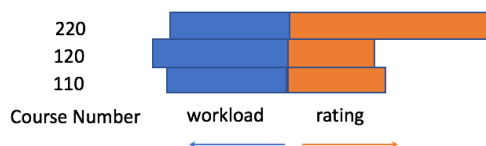
4. Visualizing gradients: (30 pts)
https://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0003nk
(CPSC 446 and CPSC 546) Create a file *CancerSurvival.html* that uses D3 to make an effective visualization similar to that shown in on Tufte's web site illustrating cancer survival rates. The dataset is given in cancer_survival.csv with this assignment. A starter file is given as cancer_survival_starter.html. At a minimum you need to add labels for the values. Make additional changes including (but not limited to) changing the colors and spacing to make the visualization easier to understand. It is ok if there are still some overlapping labels in your solution. Include a text file CancerSurvival.txt explaining the decisions you made modifying the visualization.

(CPSC 546 only) Add some marks to the visualization that indicate the standard error for each survival rate (the spreadsheet columns ending in "se").