AAA633 Visual Computing (Fall 2022)
Instructor: Prof. Won-Ki Jeong
Due date: Dec 6, 2022, 11:59 pm.

Assignment 2: Web-based Data Visualization using D3 (100 pts)

In this assignment, you will implement a web-based data visualization application using HTML and D3. The input data is a multi-dimensional table dataset, and you are required to implement various information visualization methods, such as the bar chart, parallel coordinate plot, scatter plot and t-SNE plot.

1. Input data

The provided input data (`penguins.csv`) is a table dataset contains the collection of penguin data from three different species (Adelie, Chinstrap, and Gentoo). In this data, each row is an item representing an individual penguin which consists of five attributes (species, beak length, beak depth, flipper length, and body mass). Below is an example of the contents of csv file:

```
Species,BeakLength,BeakDepth,FlipperLength,BodyMass
Adelie,39.1,18.7,181,3750
Adelie,39.5,17.4,186,3800
Adelie,40.3,18,195,3250
```

To load the given csv file, you can use D3's built-in parser, e.g., d3.cvs(). You may check out the tutorials, e.g., https://www.tutorialsteacher.com/d3js/loading-data-from-file-in-d3js#d3.csv

2. Data Visualization using D3

The main goal of this assignment is implementing different visualizations using D3 APIs and find the best visualization showing a good separation between three different penguin species. For this, you are required to implement the following visualizations:

2.1 Table Visualization (20 pts)

In this visualization, you should visualize the input data in a table format where the first row shows the attributes (each column is an attribute) and the following rows represent individual items. An example is as follows:

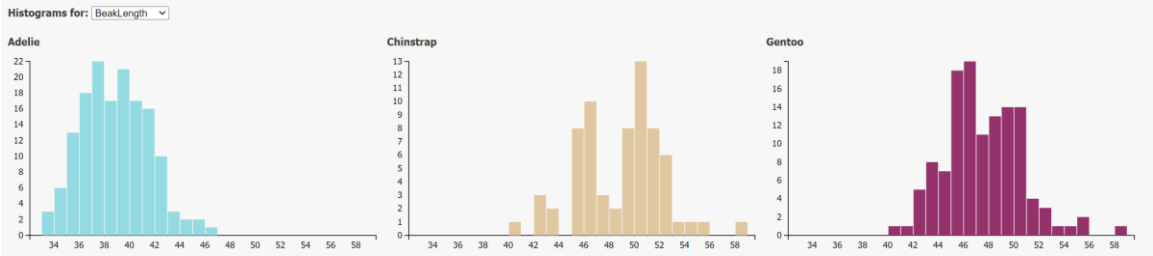
Table

Species	BeakLength	BeakDepth	FlipperLength	BodyMass
Adelie	39.1	18.7	181	3750
Adelie	39.5	17.4	186	3800
Adelie	40.3	18	195	3250
Adelie	36.7	19.3	193	3450
Adelie	39.3	20.6	190	3650
Adelie	38.9	17.8	181	3625
Adelie	39.2	19.6	195	4675
Adelie	34.1	18.1	193	3475
Adelie	42	20.2	190	4250
Adelie	37.8	17.1	186	3300
Adelie	37.8	17.3	180	3700
Adelie	41.1	17.6	182	3200
Adelie	38.6	21.2	191	3800
Adelie	34.6	21.1	198	4400
Adelie	36.6	17.8	185	3700
Adelie	38.7	19	195	3450
Adelie	42.5	20.7	197	4500
Adelie	34.4	18.4	184	3325
Adelie	46	21.5	194	4200

Note that the above table shows a bar chart as well, which is not required in this assignment.

2.2 Histogram Plot (20 pts)

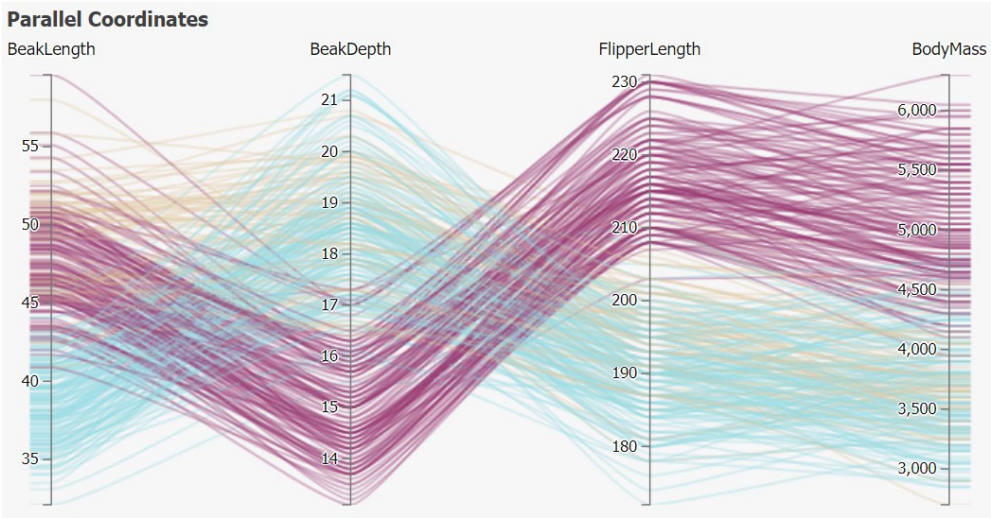
One of the basic visualization methods to reveal the data distribution is the histogram plot. In this assignment, you should implement a top-down menu to select one among the four attributes (beak length, beak depth, flipper length, and body mass) and draw three histogram plots for three penguin species for the selected attribute. You should choose a distinct color for each penguin species, which will be used in the follow-up visualizations. An example of histogram plots is as follows:



2.3 Parallel Coordinate Plot (20 pts)

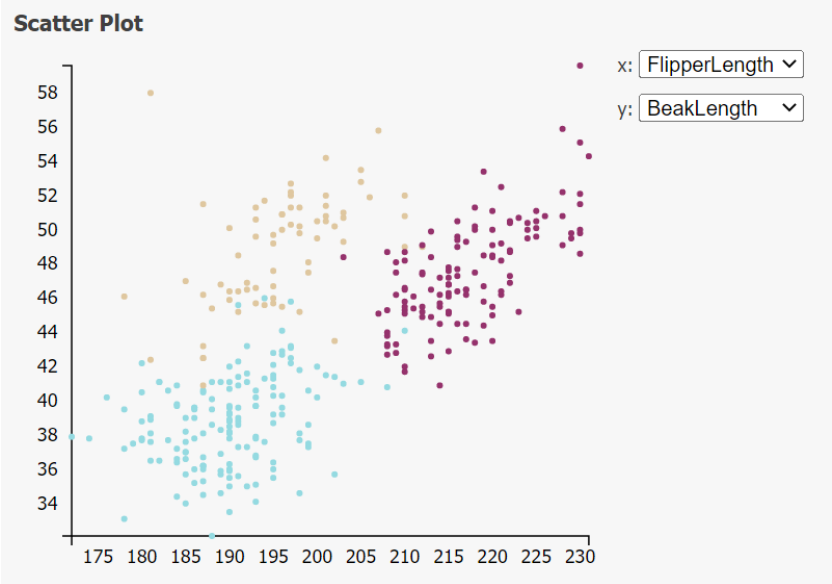
Parallel coordinate plot is commonly used to visualize multi-dimensional data (e.g., data item contains multiple attributes). In this plot, multiple vertical axes are used (each corresponds to a specific attribute). Then, a single data item is represented as a piecewise-linear line graph. Parallel coordinate plot shows the overall trends of the data well (e.g., clustering of the data). Note that you should use the same color

schemes used in the histogram plots as shown in the example below:



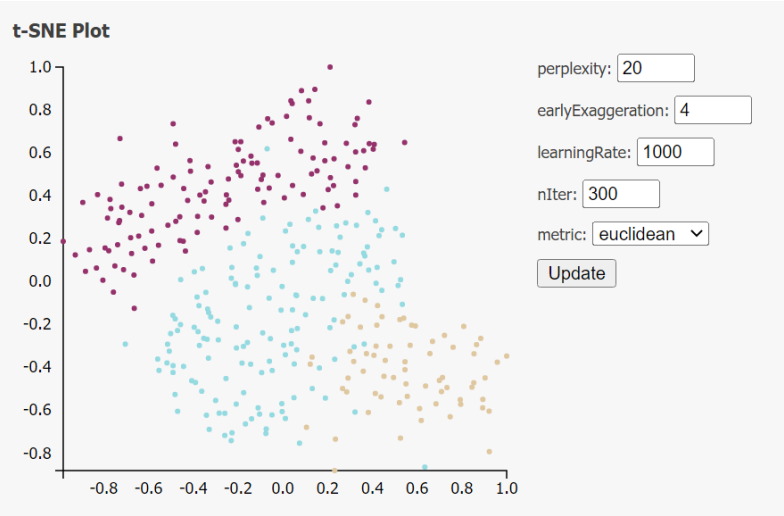
2.4 Scatter Plot (20 pts)

You are required to implement a 2D scatter plot. In this implementation, you should implement two top-down menus to select x- and y-axis attributes. Then, each data item is displayed as a point with a corresponding color on a 2D domain. Note that x- and y-axis should range between min and max of each attribute and make the plot square shape. Find the best combination of attributes that separates three penguin species well. An example is as follows:



2.5 t-SNE Plot (20 pts)

t-SNE is a widely used non-linear dimensionality reduction method. In this assignment, you are required to visualize 2D t-SNE embedding. You can use any t-SNE implementation you want (e.g., scikit-learn), and your visualization should provide a collection of menu to change user-parameters, e.g., perplexity, learning-rate, etc. Find the best t-SNE that separates three penguin species. Use the same color scheme as before. An example of t-SNE plot is as follows:



3. Etc.

You should submit your source code (.html, .css, .js) and the report explaining your work in a single zip file.

You may need some environment setup to make the visualization working, especially on the Chrome browser. Modify the environment setup of your Chrome browser as describes in the following website: https://i5i5.tistory.com/935

Extra points (50 pts): If you are willing to challenge further, you can add a data-filtering interaction to your visualization. For example, you can add "selection" interaction to the parallel coordinate axes (click and drag on one of the axes will select a subset of data for the given range) or the scatter plot (a simple box selection or lasso selection of data points on the 2D space) and visualize only the selected data (update visualizations using the selected data).

You are not allowed to look at the source code of other students in this course, but you may refer to some external D3 examples/tutorials on the web. Make sure you properly cite them in your code and report.

Good luck and have fun!