**CMSC 436 / 636: Data Visualization**

**Assignment 3: MultiD with Parallel Coordinates- Overview Technique**

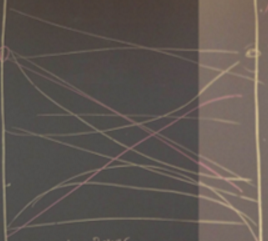
**Instructor: Jian Chen**

**Student: Hao Chen**

**Group Members:** Hao Chen, Tiantian Xie, Benjamin Latt

**Group Idea:** Using data bundles

**Solution Sketch:** Please refer to second column on the page <https://sites.google.com/a/umbc.edu/datavisualization/assignments/assignment-3>



**Pros and Cons:**

Pros- 1. Giving a clean view of the lines

2. See the relationship of each attribute more clear

3. Maintaining the main trend of the data

4. Showing the correlation of data attributes

Cons- 1. Unable to see the trend details

2. Unable to stress outliers

**For the graduate design**

The pros are: It shows much more clear for each therapy, instead of overlap much; it also shows the correlation of each attribute better than the basic one.

The con is: it spreads all over the screen make the view messy.

In the class, we voted for 8 different design choices. Although our group’s idea is not highly voted as the number one, our group and another group though about the solution in a similar way. Our design choice can improve the visual cleanness, and show the correlation of the data attributes. Well, the number one voted solution is the box and whisker plot, which is also what the whole class implemented as the second part of the assignment 3.

For box and whisker plot the good part is obvious, showing the attribute much more clear than the parallel coordinates. It means it will not take much space for each attribute. And because of the color difference for each box and whisker plot, it can distinguish the distribution of each therapy.

However, the box and whisker plot still has some drawbacks. When two therapy data are similar to each other in on attribute domain, the box plot will overlap on each other hardly to show the data distribution. Without line to link each box and whisker plot, it will hardly to see the trend between each attribute.

The steps of produce my box and whisker visualization is like this:

1. Find d3 sources to help me draw box and whisker plot
2. Make this box plot code into a function
3. Import the tumor data and process all the information into groups
4. For each group call box and whisker function to generate the plot
5. Put each group on different attributes to show the result

In the implementation stage, I got a lot of difficulties in representing the box and whisker plots. First, all the box and whisker plots codes online are not applicable to the tumor codes. I have write my own function to draw the box plot. The second one is the multi-dimensional data is hard to process because of date type. The third obstacle in my rendering the data is the logarithmic scale. My function only support the linear scale which is not compatible to the logarithm. The last one is switch on/off function is not working properly.

What I will change in the future inspired me at this moment. The first thing is time arrangement. I thought the online sources could be helpful and it would not be this hard to make all the functions working properly. I will start even more earlier. The second one is the box and whisker plot is not as clean as I expected. Because each attribute axis got so many data, the box and whisker plot is still very messy on the axis. I will change the design and look into the data before I started to do it. The third point is the canvas size is to small. For the complex data sets, the canvas size should be increased dramatically in order to show the details of the data. I will change the canvas not only wider but also longer, if there are huge data to show.