James Guan
Data Visualization
CMSC 436

Homework 3

Group Members: James Guan, John Wolf, Abishek Abishek

The problem with the current visualization is that it is very cluttered. Lines usually implies

that there is change over time. However, the lines here are misleading because the columns that

represent the different proteins may be slid around so there's no real correlation along the x-axis.

It does however represent the peaks of what proteins show up the most for each treatment but it

takes a while to understand that is what they are trying to accomplish with the lines.

To improve on this visualization, my group proposed a bar graph to represent the proteins

and the x-axis. There would be a bar graph for each protein and the bar graph would contain the

effectiveness of each treatment by show how much of the protein is present compared to other

treatments.
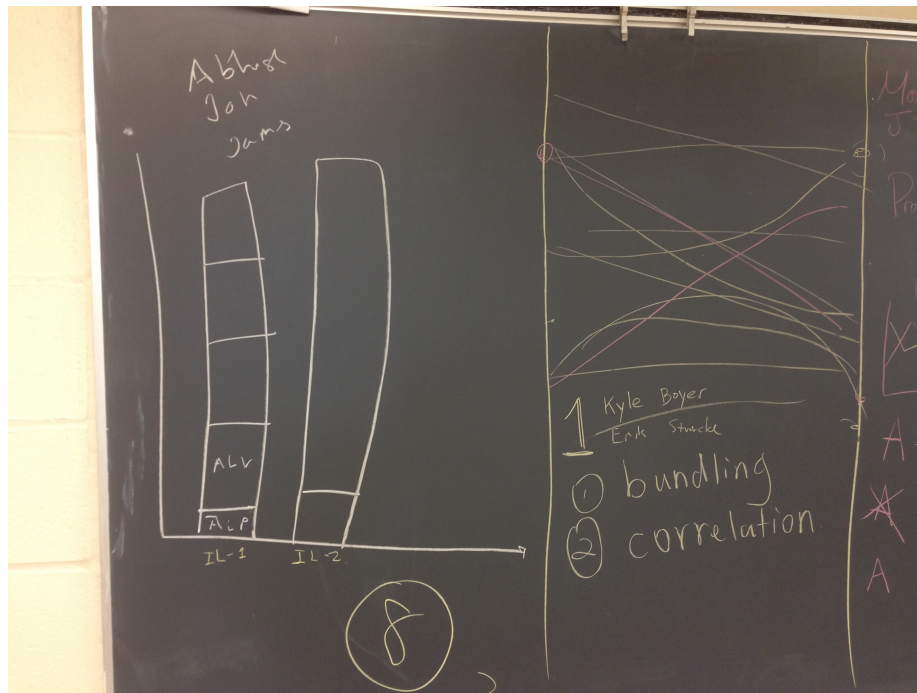
The pros of the bar graph are that it shows what the original visualization was trying to accomplish with the lines which was to show the peaks of the proteins with the certain treatment. The bar graph would reduce a lot of clutter and allow for the data to be shown very visually with the varying sizes. The negatives of this approach is that you can't really do a scale for the x-axis but you would have a legend to correlate the length of the bars with a length and it might be a little hard to tell quantity because with each segment of the bar, 0 in the y-axis would start from the top of the previous bar.
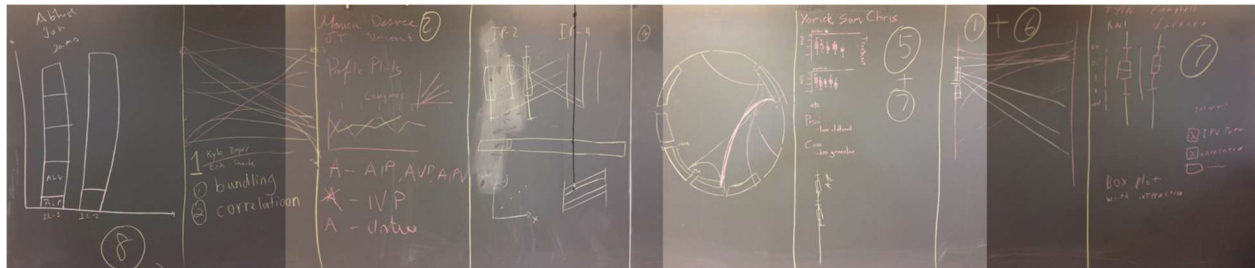
The box and whisker plot seems to be a good idea. It gives you a better picture of the distribution of the data using the percentiles and box and whisker. However, after some tinkering, I still need to do 2 * 6 * (the number of proteins) which means there would still be a clutter of data on the screen, but in this case, box plots.

When I tried to implement it, I had made a javascript file that controlled the functionality of the box plot. I made a separate .css file to add all the good looking elements to it. What went wrong with my implementation of the box and whisker plot is trying to understand how d3.js is truly used. It seems to be a library for css manipulation which could also be done in javascript regularly with more control. Another part is that a box plot only displays one dimension so trying to figure out what the data was trying to show was incredibly confusing. The project description said 12 box plots, and I thought that 12 box plots would be coming from the two types of organs,

the lymph node and the tumor, and one box plot for each treatment. However, figuring out what to correlate the box plot to when you have 9 mice and many columns of proteins. I originally combined all the protein data and mice data for one treatment and tried to display the box plot for each treatment, but I realized I can't just combine all the data points like that or else I'd be throwing out data. I then realized that I needed 12 box plots for every protein. Thus, the data would still be as cluttered as ever but the box plots would be easier to look at.

In the future, I would just try and consolidate the data into just 12 box plots that managed to display all the protein data with it. I would just click on the box plot that would pop up a scatter plot of the proteins and show the more detailed information about the proteins.