# Report for Project 6 : Data Visualization

##Summary

This data visualization is based on the data from PISA( Programme for International Student Assessment), which includes information of 500000 students, 636 dimension features for each of them.

I tried to find the relationship between the students’ academic performance and their family situations.

I selected several related features, and use them to construct new features for my investigation.

The result illustrates that, statistically speaking, the students from more wealthy and higher educated family trend to achieve a better academic performance.

##Design

### **design decisions**

At first I intended to use point chart to represent each student with single point.

When I look around the dataset, I find that, most of the academic evaluation data are quantitive value, and family status is in both quantity and quality.

In this way, I decided to use the y axis to represent the academic scores, and x axis for the quantitive family status features. The catagorial family status is shown with different color.

However, the result came to be horrible. Because the student number is too large, all the points are overlapped, and the groups are also messed up together. We can hardly tell the trend or difference.

Then I think, the bubble chart maybe a good idea. Each catagorial feature is grouped into groups. For each bubble, the x and y value is the mean of the whole group. The radius of one bubble shows its group size.

Consequently, it is clearly shown that x and y value is positive correlated for the bubbles.

**###Interaction or animation**

The audience can choose their own choice from the menu bar at the top of the chart. In this way, they can investigate a little about the data, and see relations between the different features.

When the audience hover their mouse over the bubbles, they can find the exact value along the x and y value , and also the group size.

##Feedback

1.From my classmate Tong Wu, “ I think, the y axis along at the middle of the whole chart is really annoying. Is there a better way to display it ? ”

Then I found that, the default y axis position of dimple.js is along the zero value of x axis.

So I scaled the data along the x axis, in order to make the y axis at the very left of the whole chart.

2. From my friend „Tianqi Zhang”, “Actually, what is the meaning of the size of your circles? ”

The radius of each circle represents the total size of each group. But the default variable name of this feature is just „n“. It is confusing for the people who are not familiar with this dataset.

So I changed it into “group\_size”, which is more understandable. It appears whenever the audience hover the circles.

3. From online Feedback “Raj”, “The legends at the top seems to be not organized. They are in a mess.”

I realized that, the order of the legends is important for my audience to understand the quality sequence of the catagorial feature.

So I added a Series’ order rule in my codes. In this way, the legends are well organized in order.

## **Resources**

1. <http://www.oecd.org/pisa/pisaproducts/datavisualizationcontest.htm>
2. <https://github.com/d3/d3/blob/master/CHANGES.md>
3. <https://github.com/PMSI-AlignAlytics/dimple/wiki/dimple.chart#axes>
4. <https://d3js.org/d3.v3.min.js>
5. <https://github.com/brandjamie/udacity_data_visualisation>
6. http://stackoverflow.com/questions/38893691/how-can-i-change-the-y-axis-position-in-dimple#comment65158547\_38893691