Homework Assignment #1
Scatterplot Visualization

By Shiqi Zhong
Email: szhong4@vols.utk.edu

Objective:

This is a tool project written in JAVA. It could read a CSV file and generate scatterplot graphics based on user selection.

The basic function is to choose variables for X-axis and Y-axis, show the variable names, show max and min values on axes, catch the mouse movement events, etc.

In addition to that, the user could select a third variable and use the circle radius to represent that (If you set the default equal circle radius, choose 'Unset' option for 'Choose Circle Radius').

By using this tool, we could be able to analyze the relationships in the data.

Below is my thinking.

Analysis:

In order to fully explore the relationship between two values at a time, for each generated figure, we could set the circle radius for each point to be equal. But sometimes I notice that add a third variable to the circle radius makes the figure contain more useful information. Thus, figure 1 and figure 2 I choose equal circle radius and choose a third variable for circle radius in figure 3 and figure 4.

Let's start with Figure 1 below. For this figure, we choose X-axis to be MPG and Y-axis to be Acceleration. From the figure, we can tell this is a linear relationship, as MPG goes up, Acceleration goes up. I think this is reasonable because instinctively less MPG means this car to be more powerful (more likely to be a racing car instead of an economy one). More powerful means the car could accelerate faster and result in less acceleration time.
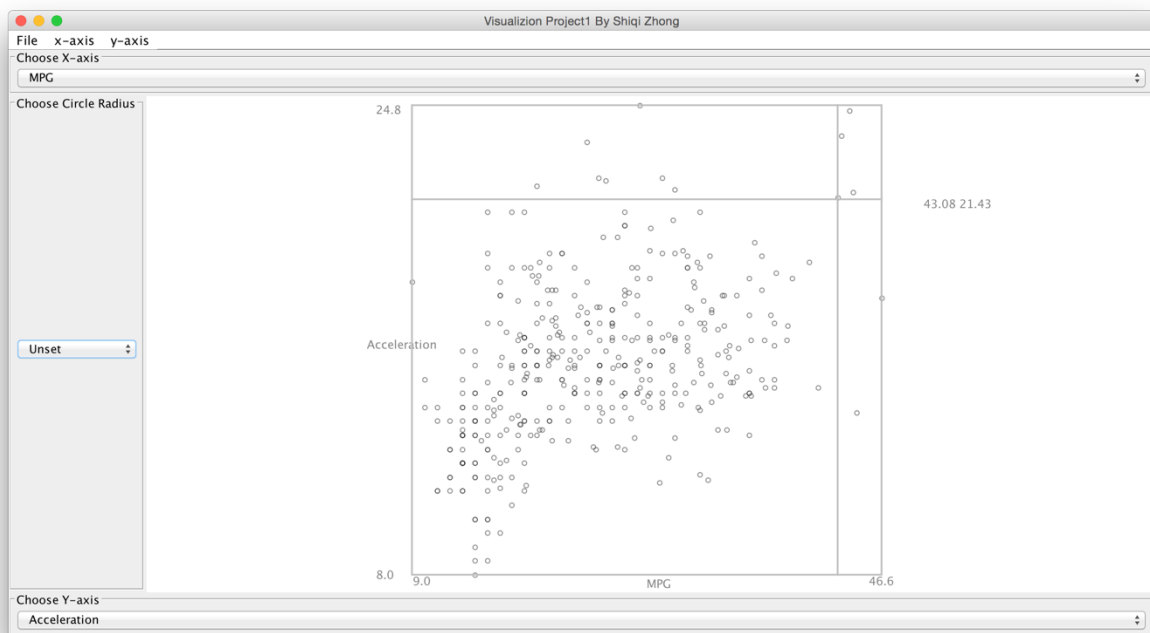

Figure 1

Then let's look at Figure 2 below. For this figure, we choose X-axis to be MPG and Y-axis to be Weight. From the figure, we can tell this is also a linear relationship, as MPG goes up, weight goes down. I think this is very reasonable.

A greater MPG means this car could go further given the same amount of gasoline. How? This figure tells us, reduce the weight of your car. And this also explains to us why Japanese brand cars are usually made in light weight and thus become more economy compare to US made ones.
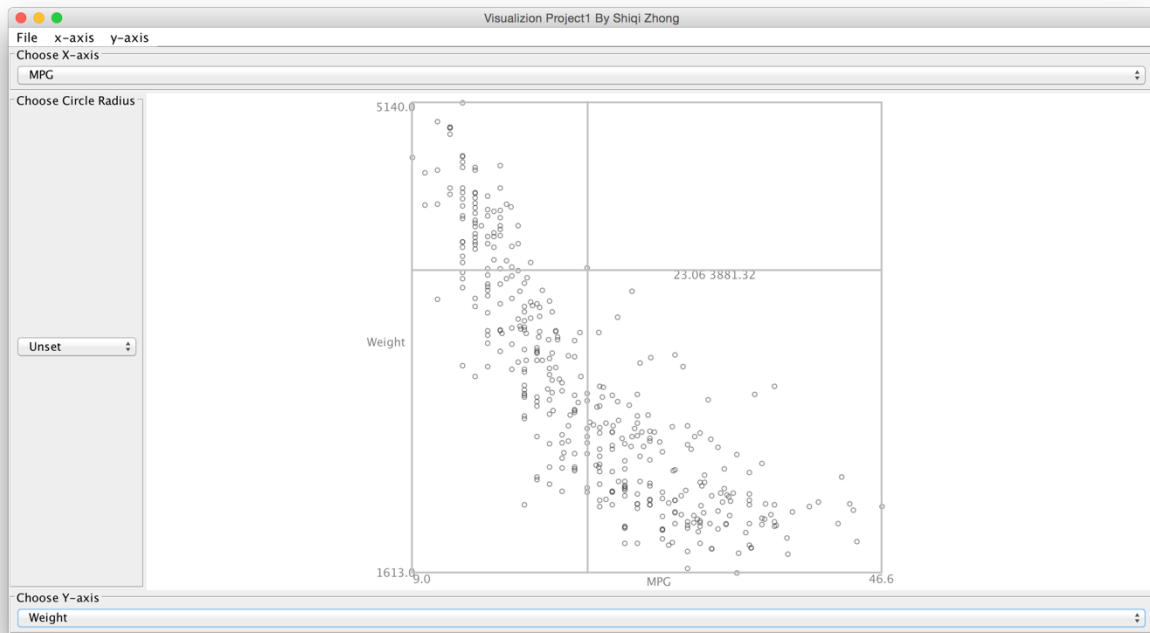

Figure 2

Thirdly, let talk about Figure 3. For this figure, we choose X-axis to be Cylinders and Y-axis to be HP (Horse Power). From the figure, we can tell this is not a linear relationship. But there is still a tendency which tells us more Cylinders, more HP.

Say you want to raise your HP, the first thing might be using a bigger Cylinder but soon you will hit a HP bottleneck. If you want to raise your HP further, this Figure tells us the solution, use more Cylinders.

Also from the circle radius prospective, we find that more Cylinders means more displacements. So, if you decided to have more Cylinders, you may put yourself into a risk of going to the garage more frequently.
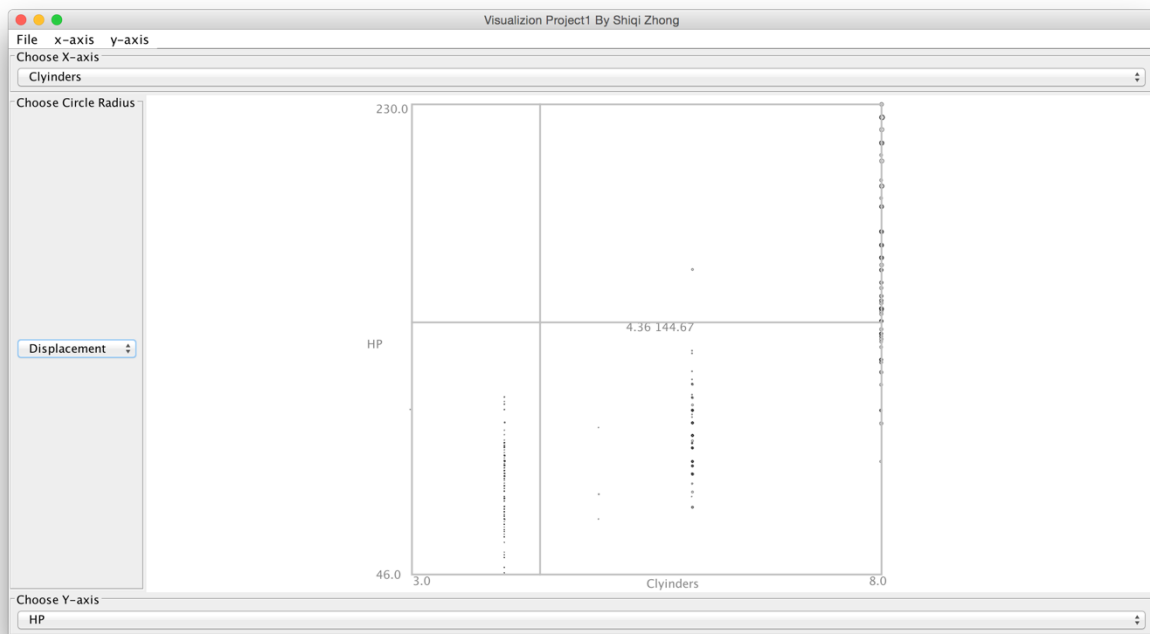

Figure 3

Last one, let's look at Figure 4 below. For this figure, we choose X-axis to be Year and Y-axis to be MPG. From the figure, we can tell this is not a linear relationship. The tendency tells us that people were more likely to buy economy cars as time passed by. I think the reasons for this might include:

1. The Government putted up the federal fuel economy standards which restricted the use of low MPG cars.
2. The gasoline prize raised.
3. The technology development.

Also from the circle radius, we can tell that the car weight is becoming lighter and lighter.
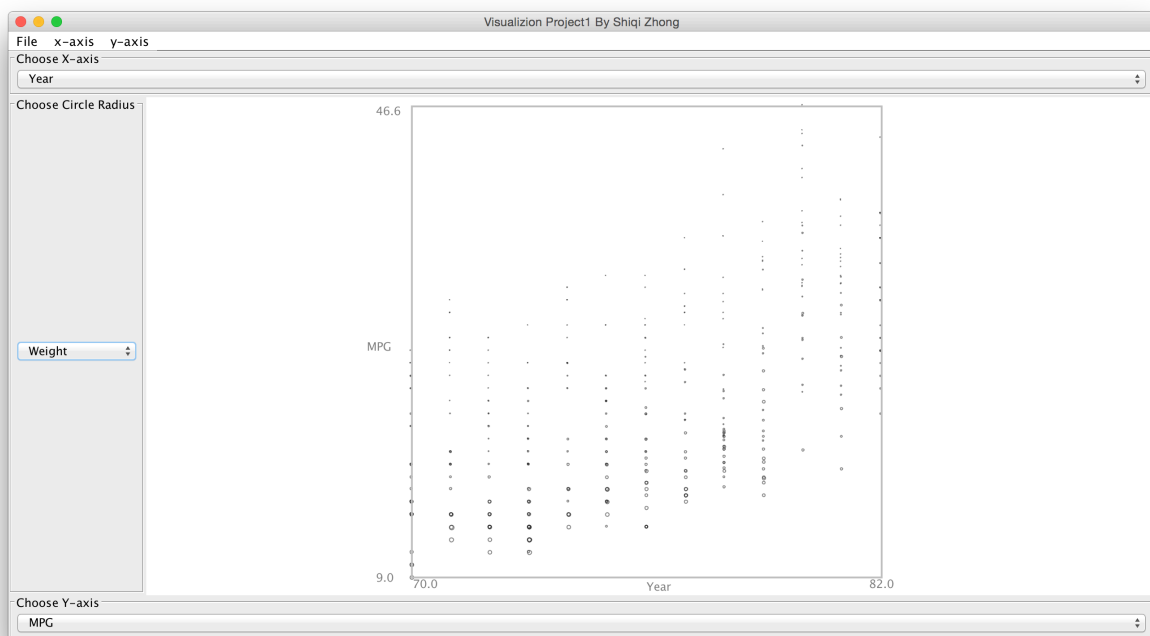

Figure 4

Future Works:

I think in order to make this tool more accurate, it should provide user with some statistic values for the sample data, like the mean, covariance for each variable and the correlation between two variables, etc.