CS590V: Data Visualisation & Exploration

Homework 2: Scatterplot
Due: February 18, 2018 11:55PM

Goal: Scatterplots are a very popular and powerful visualization tool for any data scientist. The goal of this assignment is to familiarize yourself with scatterplots. Similar to the first assignment, you will be submitting a bl.ocks.org link on Moodle.

Data set: For this assignment you may choose your own dataset. Do not pick a dataset that is very large as it will need to be uploaded on your github. The dataset should have a minimum of 5 columns and 300 rows.

Submitting homework: You will be submitting a link to your bl.ocks.org. Once you are done creating your file you can upload it to gist.github.com and then view it on bl.ocks.org. It is the link from bl.ocks that you will be submitting. Make sure you are submitting a link to the bl.ock page that contains your code and a preview of your designed webpage, Make sure you are using v4 of D3. You can write your answers in the actual HTML page or you can write them in the README.md section of your block.

Questions:

1. (25 points) Familiarize yourself with scatterplots: write up a summary of what they are [3pts], how they are created [3pts], and how they are used [3pts]. There are hundreds of different variations on scatterplots, so select a few and identify their features and how they could be controlled (provide a list). Example tick marks (one every major point and several in-between as minors, identify units, are they represented differently, …). [1 point for each feature control – could have more than 16 but max for this part is 16 points]
2. (10 points) Identify or make up a new feature or a new scatterplot that is not in the common scatterplots you have seen (no coding necessary but a good clear description or drawing)
3. (5 points) Provide a brief description of your data and a link to its source.
4. (20 points) Using that data set display a scatterplot with a color legend. Your scatterplot should take advantage of at least 4 variables (x, y, shape, size, color etc) to display information about the data. Provide the user the ability to change what each variable represents.
5. (15 points) Allow the user to select the color legend value as a filter (and provide a reset) as well as mouse over a point for more information about the other dimensions of the record being hovered over.
6. (10 points) Allow the user to select a set of points and compute something about the selected points (for example the average of x, y, and some of the other variables). You can use a box or circle selector or even a a lasso selector (more flexible)
7. (10 points) Write a brief description of the surprising trends you can see in the data (if any) and an explanation of what might be causing them. What other visualization may be able to capture that trend? If there are none, justify.

8. (5 points) Provide good code documentation and organization.
9. (10 points) Extra credit – some novel interaction and description