

CIS 4930/6930-002: Data Visualization (Spring 2018)

Project 7: Using Aggregations as Derived Attributes

1 Objectives

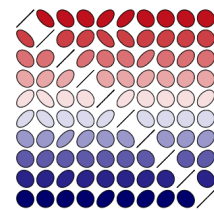
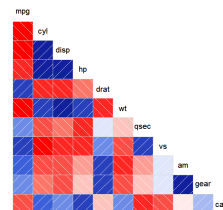
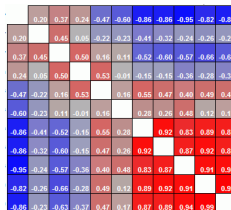
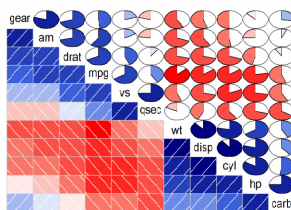
In this assignment you look to address scalability concerns with your previous sketches by using aggregation techniques. Again, take care to use good software engineering practices.

2 Ground Rules

This assignment is intended to be done alone. You may ask others for help with figuring out how details of Processing. However, code must be your own (MOSS will be used!). Furthermore, NO additional libraries (such as giCentre utilities) may be used. Doing so will result in a 0 for those sketches.

3 Assignment Instructions

- Download the data set at <http://www.calvin.edu/~stob/data/srsatact.csv>. This dataset contains standardized scores for all Calvin College 2004 seniors that have taken both the ACT and the SAT, together with their GPAs. There are 271 data points and 4 dimensions.
- Using your existing dashboard implementation, you will add a few new visualizations (you can remove old ones of your choice).
- Create a Histogram¹ for each attribute and display using small multiples.
- Calculate the Pearson Correlation Coefficient² and Spearman's rank correlation coefficient³ for all combination of attributes and show it using a Corrgram. Try to pick a variety beyond just a heatmap (i.e. use extra and/or redundant encodings).
- Calculate the mean and standard deviation and add them using additional visual encodings.



- Make sure your code works with the Calvin College dataset and at least 1 other dataset from Canvas.
- Modify your sketch to use any additional linking, interactions, or visual channels to encoding additional variables. Your selection and their implementation will have an impact on your grade.

¹<https://en.wikipedia.org/wiki/Histogram>

²http://en.wikipedia.org/wiki/Pearson_correlation

³https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient

- Add embellishments of your choice. These can include but are not limited to: axis lines, labels, and tick marks. Your selection and their implementation will have an impact on your grade.
- Make sure your visualizations are robust by designing them to support other data (number of elements or value range) and by designing them to support any size of canvas.

4 Submission

All of your work should be done in your git repository in a directory named **project7**. If you name it anything else, our script will fail (and so will you). Make sure things are labeled well, so that your peers can find them.

As you work on the files, and when you're done, make sure you add the files to the repository (i.e. *git add*), commit the changes (i.e. *git commit*), and push changes to the remote server (i.e. *git push*). If you fail to do this, we won't get your files.

5 Grading and Feedback

Your grade will be combination of objective measures (based on the assignment instructions) and subjective grading by the instructor.

Peer Review will be used to provide feedback. You will review 3 of your peers' submissions, and 3 of your peers will review your work. This should be taken very seriously as it is the primary form of feedback you'll receive.