CIS 4930/6930-002: Data Visualization (Spring 2019)

Project 3: Drawing Scatterplots

(Adapted from an assignment created by Carlos Scheidegger)

1 Objectives

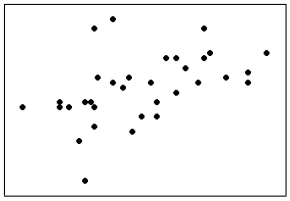
In this assignment you will build one of the most common information visualization interfaces, a scatterplot. Again, take care to use good software engineering practices.

2 Ground Rules

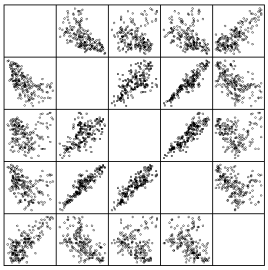
This assignment is intended to be done alone. You may ask others for help with figuring out how details of Processing. However, code must be your own (MOSS will be used!). Furthermore, NO additional libraries (such as giCentre utilities) may be used. Doing so will result in a 0 for those sketches.

3 Assignment Instructions

- Use the provided *srsatact.csv* dataset, which 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.
- Create a sketch with 800x600 resolution that opens a file dialog box (`http://processing.org/reference/selectInput_.html`) and loads a selected data file.
- Create a SCATTERPLOT of SAT's mathematics scores (SATM column) versus SAT's verbal scores (SATV column). In other words, the x-coordinate of the plot should encode the SATM variable, and the y-coordinate should encode SATV. *Hint*: You should consider using the $map()$ function for easier display of the scatterplot (and the scatterplot matrix).



- Add the functionality to switch views (using the keyboard) to a scatterplot with ACT scores (ACT column) on the x-axis and GPA scores (GPA column) on the y-axes.
- Create a 2nd sketch with a SCATTERPLOT MATRIX (SPLOM) that shows all combinations of variables. *Hint*: a SPLOM is just a grid of scatterplots. Reuse your scatterplot code from the first sketch.



- **Grad Req Only:** In addition, test that (at least) ONE (1) of the other datasets provided works with your sketches (and place it in the data directory of your sketches).

- Modify your sketches such that they use additional visual channels to encode additional variables. Consider using color, size, shape, depth, etc. Your selection and their implementation will have an impact on your grade.

- Add embellishments of your choice. These can include but are not limited to: axis lines, labels, and tick marks. Consider the margins for your embellishments (try to pick good values for the tick marks and a good number of them—not too many and not too few). Your selection and their implementation will have an impact on your grade.

- Make your visualizations robust by designing them to support any data (number of elements or value range) and by designing them to support any size or aspect ratio of canvas.

4 Submission

All of your work should be done in your git repository in the directory named **project3**. If put it anywhere 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 make sure you frequently 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.

- Breakdown
  - 2 Visualization - 7 points
    - scatterplot - 3 points
    - scatterplot matrix - 4 points
  - Embellishment & additional Visual Channels - 1.5 points
    - 0.5 points for none used
    - 1.0 point for a few
    - 1.5 points for many
  - Code Professionalism - 1.5 points
    - 0.75 no comments, no classes, "hard coded" values
    - 1.0 minimally commented, few "hard coded" values
    - 1.5 commented, properly used classes, few "hard coded" values

- 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.