Data Visualization Examples

by James Bernhard

Spring 2012
For some ideas and inspiration on your projects, try the following websites:

- Visualizing Economics
- Chart Porn
- Flowing Data
- Infosthetics
- Information is Beautiful
For an example where they detail their process of creating a specific data visualization, try:

Geo bee contests

In all of these, be sure to understand an important distinction

The primary goal may be the communication of information, which is what we are after (often referred to as data visualization or information visualization)

Or the goal may be an artistic or appealing representation of something, which is not what we are after (often referred to as data graphics, information graphics, or infographics)

These terms are not standardized, and different people use them in different ways, but be aware of the different goals of the two types of graphics displays

The above websites contain both types
For some REALLY advanced data visualizations, try Martin Wattenberg’s work at

Martin Wattenberg’s site

As you are designing your project, realize that the grid graphics package will allow you to create a graphic from scratch, not from a template.

However, if you are doing something specialized (such as displaying data on a map or displaying a mathematical graph, with edges and vertices), then there may be other packages that can help you.
For an example of a project data visualization, consider my gradebook visualization

*The problem:* How can I use my gradebook to monitor student progress? In particular, I need to be able to create alerts, etc.

*The data:* I began with letter grades for each assignment, unfolding as the semester goes on

But would I need more? Weights? Comments? The curve?
Initial mock-up: a color-coded “matrix of grades”

Which axis should have students, and which assignments?

What colors should I use? (I opt for a scheme that is easy for me to remember and interpret)

How should I order the students and assignments? (Assignments more through time on the x axis, students are ordered alphabetically on the y axis)

Which assignments should I include? (All of them)

How do I program that? (Grid graphics)
I make a simple mock-up, a “proof of concept”

Will this allow me to solve the original problem? Not exactly: I need to incorporate more data

In particular, I need an overall grade to date

This complicates matters considerably: now I have to have weights for each assignment (even if not explicitly in the visualization)

I also have to have a way of converting from letter grades to numerical grades

_Hmmm...._
I actually designed a method of translating letter grades to numerical grades according to a flexible scale.

The traditional 4-point scale has some serious flaws for computing overall course grades.

In particular, when I give a test that has a numerical grade, I will almost invariably work with a 90-80-70-60 scale, which assigns a much wider range of numbers to the F grade than to any other.

After pondering which scale I think more suitably represents the way that student performance should be assessed as a grade, I decided that I prefer a 90-80-70-60 scale.

For example, if a student completes only half the total amount of work for the class, in my view the student has earned a 50 (F), not a 2 (C).
The 4-point scale has the strange problem that we don’t assign A+ grades here, so there is no way to counter an A- to balance out to an A.

Further thought has reinforced my preference for a 90-80-70-60 scale for other reasons too (e.g., the 4-point scale is too discrete, etc.)
So I set out to find a good way to convert A-F to a 90-80-70-60 scale and realized that while it can be done, I am throwing away useful information by grading an assignment with a letter grade in the first place.

In my mind when I am grading an assignment, I assign a 90-80-70-60 score (or something thereabout) to the assignment, which I had then been converting to A-F and giving to the student.

Correspondingly, both the student and the gradebook are not getting all of the information.

For this reason, I generally work on a 90-80-70-60 scale throughout and only assign an overall letter grade at the end of the course.
For many assignments, this scale is less natural than a point scale (say from 0 to 40), but converting from the point scale to a 90-80-70-60 scale is just a matter of computing percentages.

Although... sometimes percentages don’t really make a good conversion scale, and a *curve* is desirable.

However, that doesn’t need to be built into the visualization so much as the data structure, so we will continue with the visualization for the moment.
I realize that I need to assign a weight to each assignment in order to compute a grade to date on a 90-80-70-60 scale.

This may change some over the course of the semester, but I can hazard a pretty good guess about what the weights will be (especially if I have taught the course before).

But what about individual students who either demonstrate their understanding in a nonstandard way? Or students who must be absent during a quiz? How will I handle that in the gradebook?

I decide that each student should have a separate weighting system, so that I can assign zero weight to an excused missing assignment, etc.

For this purpose, having a comment position available for each assignment for each student might be useful, so I implement that.

(That is still in place today in my gradebook object, but I have found that I rarely use it in practice.)
With the scale and weights decided upon, I can now put an “overall grade” column in my matrix

I opt to put this on the left near the student names, since that is often the first thing that I want to access

I can now order students by overall grade, with the highest on top
I think about how I want to display the 90-80-70-60 scale in the matrix

There are *many* options: as numbers, as colors, as colored numbers, as miniature bar graphs, as line segments, . . .

These all have merits

To help me decide, I think back on the problem that I’m trying to solve: at a glance, I want to see how all students are doing, especially who (if anyone) is falling behind

With this in mind, ironically enough, I decide to display the grade as a color, not displaying the extra information that I added by converting to a 90-80-70-60 scale (except in the overall grade to date column)
I then realize that some students have missing assignments — how should I display those?

I decide to code them as zeroes, so that’s how I recognize them as I am creating the graphic.

Black (or the background color) is one option, which seems like a natural choice to indicate missingness.

However, thinking back to the original problem, missing assignments are one of the most important things to notice in this visualization! They are often the first (and best) indicator of who is falling behind.

I play around with colors such as red but realize that white actually sticks out more to me among the rest of the colorful display.
So I code missing assignments in white, but now I have the issue that all the assignments that haven’t been graded yet are in white. This makes it harder to see missing assignments at a glance, which detracts from the effectiveness of the visualization.

The solution: more data!
I decided to record a table of assignments and whether or not they have been graded.

It also allows me to include only the assignments that have been graded in the overall course visualization.

This allows me not only to compute a grade to date, but to indicate how much of the course is incomplete (as in the online version).

At this point, I have a color-coded matrix of grades.
The next problem: while the overall grade to date makes a nice summary number, the overall pattern is still not really being represented by the colors very well.

I can tell some students who are falling behind by a large amount of red (D or F) or white (missing) assignments, but some students may be doing very well on heavily weighted assignments but not as well on lots of less weighty assignments, or vice versa.

Consequently, the amount of color in a given student’s row doesn’t quite indicate how well the student is doing: some translation is needed.

The solution: maybe I can get information about the weights in there somehow...
Now I completely reconsider what to put in each cell of the matrix.

As before, there are many options: how about a little “time series” line graph, for example? Then the weight could be given by the size of the dot on the line.

Or maybe a miniature bar chart? Then the weight could be given by the width of the bar... but that would leave a lot of blank space between bars, since I don’t think I can make the column widths vary as much as the scale of the assignments.
What I decide on is to make the size of the square indicate the weight of the assignment, and continue to code the grade by color. After some playing with things, I decide that a (round) dot would actually be preferable to a square for each assignment.

Since I want the amount of color to be an immediate indication of the overall grade, I scale the radii of the circle by the square root of the weight instead of the weight.

That way the area of the circle (instead of the radius) is proportional to the weight.
I also decide to write the actual score in the “grade to date” column because that exact number is actually relevant, not just the overall grade.

I continue to color code the grade to date column though.

The visualization that I arrived at is...
This visualization has served its purpose very well: at a glance, I can identify how each student is doing in the class and notice who is doing particularly well or falling behind.

As an aside, I can also see if any assignments were particularly difficult.

I have thought about how to improve this graphic, but I haven’t come up with anything since.