Broad study at Marlboro: a first play with the data

Matt Ollis

November 7, 2012

1 Introduction

Last February I looked at some data relating to how many students take math classes and natural science classes and, for those that did, at what stage of their academic career they took their first one. After conversation in the First Two Years Committee, I promised to expand this to look more generally at the question of broad study. A mere nine months later, here's a first play with some data that Ian Kozak provided.

For each of the four graphs the data set contains all non-transfer students with graduation dates from May 2006 to December 2011. There are 318 such students.

In the next section we look at the data from the perpective of the four main areas: Arts (ART), Humanities (HUM), Social Sciences (SSC) and Natural Science and Mathematics (NSC). In Section 3 we look at it from a by-student perspective, using a dubious measure of how broadly a particular student has studied.

This is definitely only a first pass at this. Ideas for different ways to extract interesting and useful information from the data very welcome. There are plenty of warnings and disclaimers too that I'll drop in at the appropriate points.

2 By area

Do most students take science classes? By what semester? The first graph, Figure 1, charts the number of students who have taken at least one course in an area by particular semesters.

First batch of disclaimers:

- I've stripped courses with Cross-Disciplinary Studies (CDS) and World Studies Program (WSP) codes out of the data. These courses presumably contribute to a student engaging with one or more of the four main areas, but I have no way of determing which one(s).

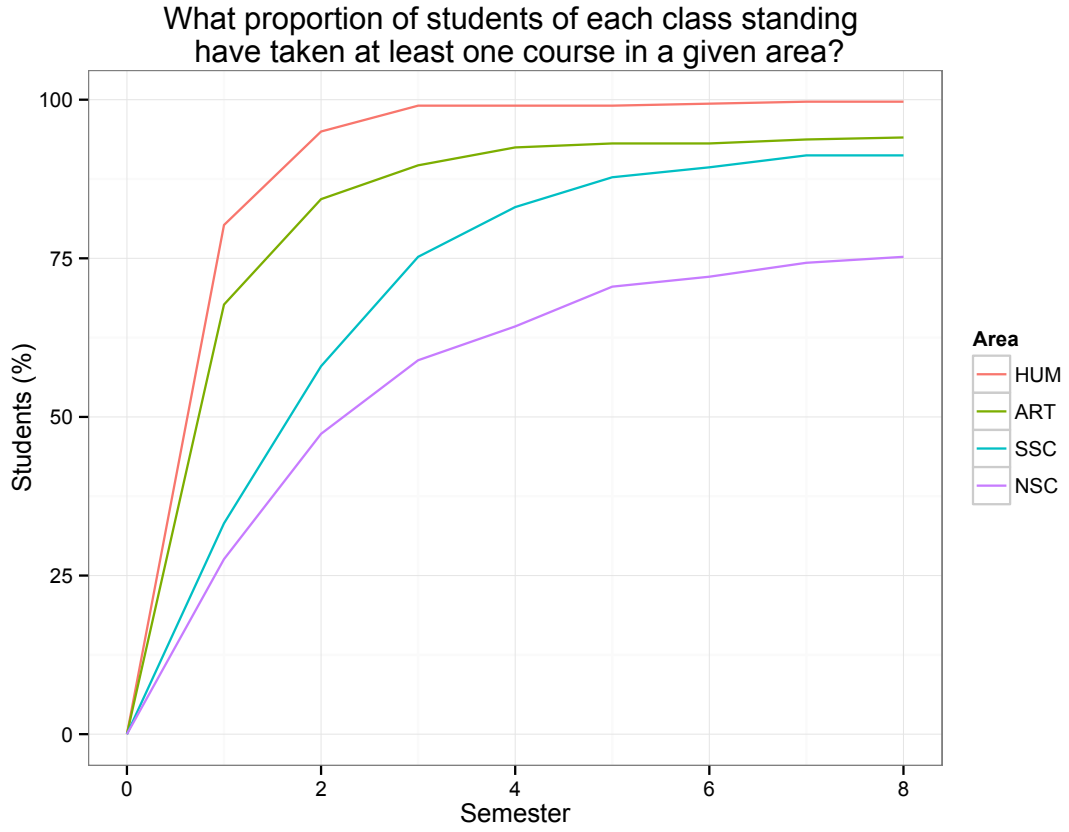

Figure 1: *The percentage of students within each nominal class standing (1 being Fr1 and 8 being Sr2) who have taken at least some credit in an area. If you're reading in black-and-white, the lines on the graph go HUM, ART, SSC, NSC from top to bottom.*

- Main code designations: I don't know how much thought goes into them. Evidently I don't put enough in: when requesting a new course I don't mention the code that it should have. Looking through the data for the previous document on math study, I was surprised to learn that I'd offered a course with an SSC code. So it's not at all clear to me that a student taking a course with a particular code has engaged with that area for that amount of credit or, especially, that it doesn't constitute engaging with another area too.

- Do areas mean much? Anything? If so, do we have the best division of courses into areas? Might there not be multiple overlapping areas that are better considered when asking questions about broad study?

What does the graph tell us? Looking at the right of the picture we see that all students in the data set take at least one course in the humanities (although one student managed to avoid doing so until his/her Jr2 semester). About 90% take arts and social sciences and about 75% take natural sciences. That last figure was discussed in the previous document: if 25% of our students don't take and natural sciences (and 50% don't take math), how do we respond to NEASC's standard stating that "[g]raduates successfully complet-

ing an undergraduate program demonstrate... the ability for scientific and quantitative reasoning"?

Are the corresponding figures at around 10% for arts and social sciences sufficiently large to be worrisome?

The other thing this graph shows us is when students first engage with an area. I was pleasantly surprised to see the NSC and SSC lines growing after the So2 semester. My concern based on my impressions of how students approached their years on plan was that many students were not adding much, if any, breadth to their study at this point. These two lines point to the opposite (at least a little). The ART and HUM lines are roughly horizontal in this region. The HUM one is certainly not a cause for concern: it's not rising because everyone has already taken a HUM course. Is the ART one an issue worth looking at more? Are the NSC and SSC lines rising as fast as we'd like in this region?

3 Breadth of Study

The next graph looks at which of the four main areas a student has taken fewest credits in. We saw in the last section that there are students who take no credits in an area, but how prevalent is it to take just a couple of credits in one area? Figure 2 gives the data.

Look at those outliers. Two students managed 20+ credits in each of the four areas. Wow.

But what is the bulk of the graph telling us? If we take that clump at 1–4 credits as mostly representing taking a single class in an area, then we have a split into three roughly equal groups: students who take no class in one or more areas; students who take a class in each area, but exactly one class in at least one of them; and students who take two or more classes in each area. If we're interested in investigating this further from this point of view I can draw a similar graph but for classes rather than credits.

Is that a distribution we'd hope for? Accept as reasonable?

How do we define broad study? Here's a really bad way, but one that I think might offer some insight into the situation. More disucssion later on its merits and what we should (or, more importantly, shouldn't) do with it.

Give each student a *breadth of study score*, defined to be the sum of the number of credits that they have taken in each of the four main areas but with a cap of 10 credits coming from any one area. So a student who has taken at least 10 credits in each of the four areas scores the maximum of 40. A student with 5 ART credits, 9 HUM credits, 10 NSC credits and 28 SSC credits scores $5 + 9 + 10 + 10 = 34$.

Figure 3 is a box-and-whisker plot of the distributions of breadth of study score at each class standing.

Quick reminder/primer for reading box-and-whisker plots. Look at the Sr2 one. There's

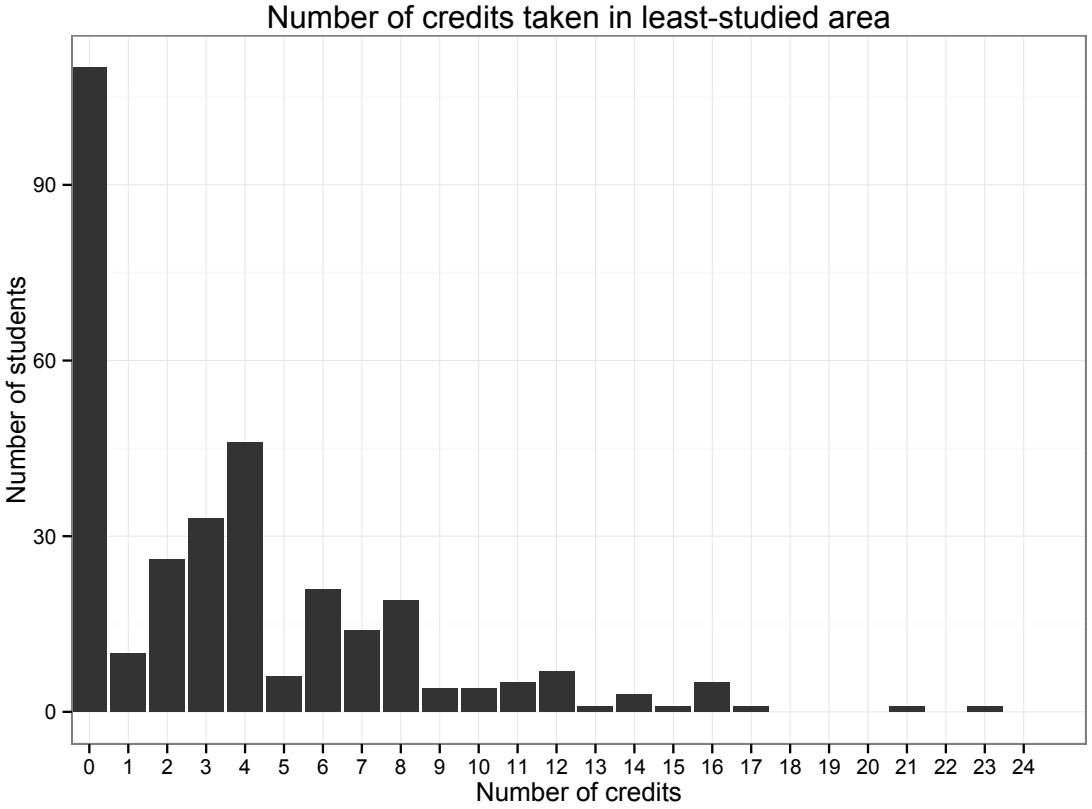

Figure 2: *Bar chart showing how many students in the data set have taken how many credits in the area in which they've studied least.*

a heavy line through the centre of the box; this marks the median. It's at 31, this means that half of our graduates have a score of at least 31 and half have a score of at most 31. The top and bottom levels of the box mark the interquartile range. The upper quartile (top of the box) indicates that a quarter of our graduates score at least 35 and the lower quartile (bottom of the box) indicates that a quarter of our graduates score at most 28. Looked at differently, half of our graduates are in that box, scoring between 28 and 35. The whiskers indicate the extent of the data, excluding outliers; for graduates this runs from 19 to 40. There are a few outliers (data points that seem oddly far, for some statistical definition of oddly far, from the centre of the distribution) in the teens.

Box-and-whisker plots are horribly last century. All the cool data visualisation kids use violin plots now. The same data that generated Figure 3 is used to make such a thing in Figure 4. The idea is that more nuance is immediately apparent than in the box-and-whisker version. The dot indicates the median, just as the thick line did before. The rest of the shape of the distribution is now represented by the splodge (a.k.a. violin) with thicker regions corresponding to higher density of data points.

Never mind the data visualisation nerdery, what are the pictures telling us?

Let's deal with the outliers in Figure 3 first. The 0s for Fr1 through So1 are, presumably, students coming in with AP credit or similar and so skipping Marlboro class standings.

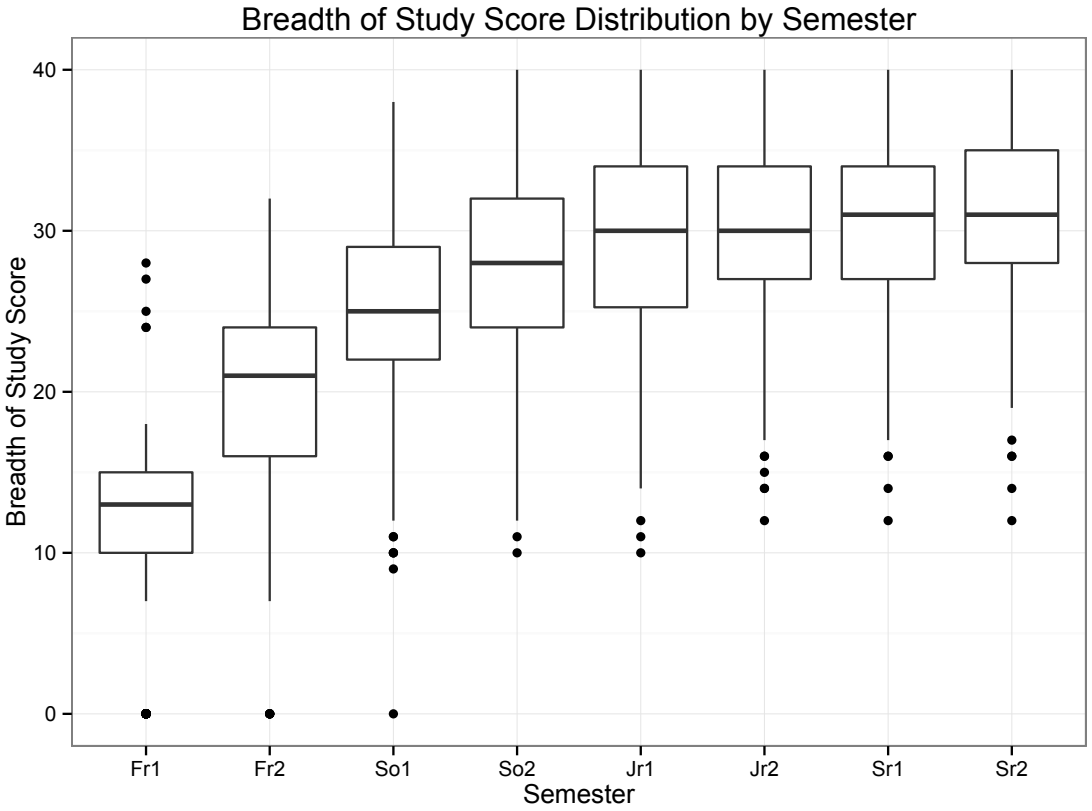

Figure 3: *Box-and-whisker plot giving the distribution of breadth of study score at each class standing.*

My guess for the high end outliers in Fr1 (20+ credits as a Fr1?!) is that they're students who earn less than 12 credits in their first Fr1 semester but are not dismissed from the college and have another semester with Fr1 status. (Probably a good place for a warning that should have come earlier: all of these data use the nominal class standings; students may skip or repeat them and earn 0 or more than 18 for any of them.)

The outliers just below the whisker for So1 through Sr2 I'm less sure of. Is this a handful of students who are genuinely studying very narrowly, or maybe they're studying more broadly and the approach I've taken to the data has hidden it. Possibilities: a lot of WSP/CDS credits; the ART/HUM/NSC/SSC distinction being misapplied or inadequate to capture breadth; not-for-credit breadth-enhancing activity; I messed up the numbers somewhere.

Related to the outliers, but sneaking into the whiskers, are scores at 20 or lower. Such a score means either that the student has not taken any credits in two of the four areas or that there are three of the four areas with a total of 10 or fewer credits. That such students exist is, I think, worrying even bearing in mind the above alternative possibilities for a low score. More generally, my hunch for a reasonable baseline that we'd expect everyone to reach (unless the above, or other, possiblilities kick in for them) is somewhere in the mid-to-high twenties. Maybe even 30 is not unreasonable. What do others think? (But

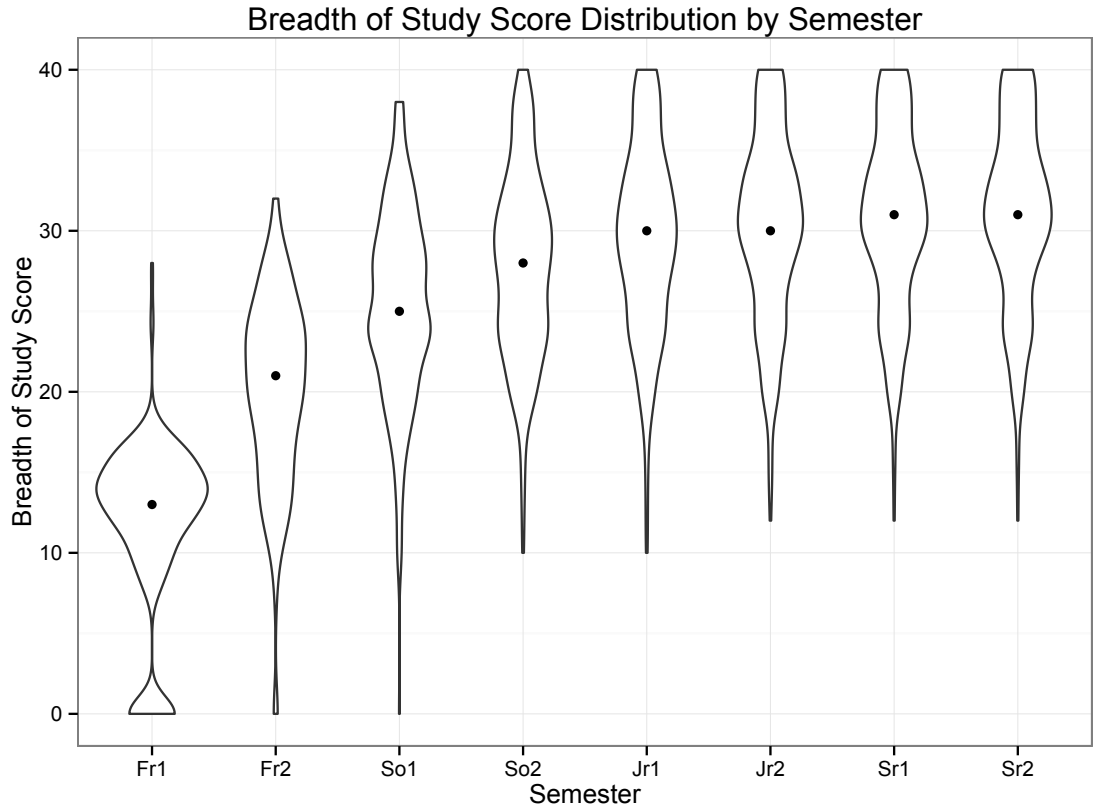

Figure 4: *Violin plot giving the distribution of breadth of study score at each class standing.*

see the next section for some cautionary notes when thinking about the consequences of picking any such number.)

Looking at the Sr2 violin, it has characteristics that we want to see. A flat top indicates that a noticeable fraction are maxing out at 40. It's also bulges nicely near the top—most students are scoring highly—and the tail is narrow—few students are getting low scores. The questions are whether the flat top is broad enough, the bulge is high enough, and the tail is as narrow and short as we'd like.

I think the change with class standing is much as we'd expect. Quick growth with a lot of spread early on and then slowly edging upwards and bunching.

4 Some final comments

What to do next with all of this?

Worst idea. Let's institute the breadth of study score as an advising tool and graduation requirement. It circumvents the need for those time-consuming conversations between faculty and students about the merits of and routes to broad study. By captur-

ing breadth of study in a single number we simultaneously make it measurable, and so easily assessable, and remove any disagreement about what broad study is.

Better idea. Let's use these graphs and data as one more way to think more about broad study at Marlboro. However much numerical data we gather (the spreadsheet I'm working from has $1,522 \times 94 = 143,068$ entries, many of them unused in this first analysis) it might turn out that broad study is not hidden somewhere in it.

What else do the graphs here suggest?

Separate from how to interpret what we have so far is the question of where, if anywhere, to go next. Here are some possibilities:

- Change over time. For these graphs I took the last five years as a clump, hoping that that was a big enough and recent enough dataset to give us a reasonable snapshot of where we are now. The data go back further; we could look to see if things (what things?) are changing or staying about constant.
- Finer, or different, approaches to dividing the curriculum. We've looked at the math graphs already and I have data to do similar work on languages and environmental studies (though in the latter case I don't trust the course designations sufficiently to think anything meaningful will emerge).
- Retention. How broadly do students who leave the college study while they are here? Perhaps too broadly and so not laying sufficient foundation for plan; perhaps too narrowly and they leave in part because they can't focus on what they are most passionate about (perhaps both, which would manifest in the dara as a larger spread).
- Same but better. Other ideas for extracting broad study information from students' academic data? A better definition, or multiple definitions that addess different aspects, of broad study perhaps?
- Does breadth of study correlate with final grade/honours? Does it differ depending on the students' plan areas?