Hi Sam –

Here is what I am thinking regarding the ordination about which I emailed you.

A little background:

CFANS will be undergoing reorganization next year to improve delivery of academic programs – both grad and undergrad. I am on a small (n=5) team that is developing the plan. We will be creating 3 – 5 schools within which the programs will reside. In most cases, a program and department are aligned (FWCB and the FWCB major). Sometimes it is a bit looser (Con Sci and FWCB). We are at the brainstorming level of how to organize the schools and everyone is suggesting grouping based on their hunch. I want to add some fact-based analysis to it and, for me, that means I’m tacking a community ecology point of view.

Overall Goal:

To group undergraduate programs based on the coursework that students are taking.

Data:

I requested a data set of 1) student identifier, 2) major, 3) course taken for AY ’22-’23. Each student has ~5 – 10 records depending on how many courses they took last year. We have >20,000 records!

Privacy:

The only way we could pull this data is by student name. Please keep student names confidential. (I know that you probably aren’t interested in student names whatsoever, but please keep this in mind if you ever share the raw data).

Suggested Analysis Process:

1. Trim data set.
   1. Perform a pivot table for course and count.
   2. Create a Pareto chart to see the length of the ‘tail’ to determine what courses to drop from the analysis. E.g. do we eliminate singletons? Doubletons? Classes taken by 10 or fewer students? I don’t really know since I haven’t seen the data. However, I suspect the sparse courses will only confuse the analysis.
   3. Chop the courses from the tail from the data and create a new data set to analyze.
   4. There may be an uninformative start to the chart, too. If everyone takes a particular course, it won’t help the ordination. We can chop out those as well.
2. Ordinate!

NOTE: THESE ARE MY IDEAS BASED ON A 30-YEAR-OLD KNOWLEDGE OF ORDINATION. I SUSPECT THAT YOU MAY KNOW BETTER, MORE CONTEMPORARY, METHODS.

* 1. Run an NMDS based on student name and course, with the dots colored by major.
  2. Run a constrained ordination, such as CCA by student name and course that is constrained by major.
  3. Run a hierarchical cluster analysis based on major being the case and course being the data to create a dendrogram. I don’t have an outgroup (e.g. kids who are getting engineering degrees) but you shouldn’t need one.

1. Plot!
   1. Plot a bivariate graph for the NMDS with dots colored by major.
   2. Plot a bivariate graph for the CCA with the vectors identified as the majors.
   3. Plot a dendrogram.

Let me know what you think. Since I’m not doing it, my brain says that it is pretty straightforward. I know that you will see the problems with what I suggest. Feel free to pursue other paths that you think are better.

Thanks,

Rob