

## SI 618 Project Proposal: Student Evaluations

### **1.Evaluation Data**

For this project I will be using data provided by the University of Michigan Registrars Office through a Fellows project I am currently participating in with Kevyn Collins-Thompson (esteemed SI Faculty member). This data set contains all responses to student evaluations from the Fall 2013 semester. The data came in a csv file containing variables including:

- Subject: text
- Class ID: int
- Class Section: int
- Course Type: text, Lecture (LEC), Seminar (SEM), Discussion (DIS), Independent study (IND) or Lab (LAB)
- Evaluation response rate (a column for population and responses):
- Question Key: int
- One column for each of the 6 response types from Strongly agree to Strongly Disagree and NA: int
- The median of the responses (if strongly agree was 5 and strongly disagree was 1): float

There are over 200,000 rows in this file. I have access to this file through the M+Box system.

### **2.Research Questions**

1. Which departments are the most liked and most hated?
2. What class levels (100,200,300,400) do students like the most?
3. Do different departments differ in the class levels that are most liked?
4. How does the size of the class relate to how much it is liked?

### **3.Methods and Visualizations**

1. Breaking down by department and questions that indicate how positive a student felt towards the content of the course. I could show a comparison of departments on the same graph, or many small histograms of the aggregate mean for all of the classes.
2. Bucket all 100, 200... etc. level courses together into factors(levels). I could also choose not to bucket them and see if there is a trend, even within levels (101, 102) towards worse or better classes. I could use smoothing for this with a scatterplot.
3. This combines the approaches I would take for question 1 and 2. I would want to show graphs of each with level of course vs. likert responses being charted.
4. This would use the data on populations and compare it with how liked a class was. This would also be an excellent place to use smoothing to determine whether any trend exists in the data.