Javin Unger
Faculty Supervisor: Professor Serita Nelesen
1 August 2015

**Educational Impacts in Computer Science / Program Validation through Simulation**

I worked with Professor Nelesen this summer on two projects: one in which I compiled and analyzed semester and exam grade data of various Computer Science courses and another in which I continued the work of former Calvin students on a program called HTMAD and an associated simulation engine.

Computer Science is - relative to the other STEM fields - a new science and an even newer academia; thus, teaching it can be rather difficult, especially when new developments in the field occur so rapidly. Therefore, understanding how computer science students learn and react to new teaching tools is particularly important in introductory courses. It is to that end that my task for the first part of this summer was to encode many semesters’ worth of grades into spreadsheets and to study them for significant trends not only between semesters, but between sexes, programming language learned, exercises assigned, etc. My main objective was to identify the methods that worked; for example, with the help of Professor Joel Adams, I analyzed course data from the latest CS112 course (spring 2015) in which approximately half the class performed a different parallelism lab from the standard. Did the new lab make a difference? My job was to answer that question via statistical analysis and, in this case, it did indeed make a difference. I also found that, in general, women do better than men in introductory computer science courses, which was rather surprising (and heartening!) given the unfortunate stigma.

The second part of the summer focused on a project completely unrelated to the previous. Rather than data analysis, I worked with existing software created in Python called HTMAD. Its purpose is, admittedly, rather unclear to me because I am not a biologist. However, I do know it works with data produced by a machine called MALDI-TOF, and it exists because data from said MALDI is difficult to explore. HTMAD was developed by a team at Calvin to simplify the processing of MALDI data, and allow data exploration in ways that the machine developers did not envision. After performing code maintenance on HTMAD, we were interested in verifying that HTMAD worked as we desired. Since true biological data is often messy, I used a MALDI simulator (published in 2004) and wrote a script to generate a variety of inputs for HTMAD. I then could check that, even with variation in the data (as generated by the simulator), HTMAD could correctly perform the data analyses we targeted.

I learned a great deal this summer. I’ve worked with spreadsheets before, but never on such a large scale; that is, I’ve never broken down so much data into so many pieces - there are so many factors that could account for the trends I studied, and accounting for all of them was challenging. Learning how to maintain and interpret other people’s code was another new experience and it turned out to be more rewarding than I thought. Beyond that, I’ve also never done so much technical writing in my life; I’m convinced that it is valuable and practical, though, if I’m honest, I’m praying that I won’t have to as much in the future.