Our problem domain consisted of the computing and printing services available to students and faculty within the Cathedral of Learning. Specifically, the usage of and availability of those resources. To that point, our questions were as follows:

Cathedral Computer Labs

Robert Colleran, Zach Luettgen, Manasi Thakkar, Jay Thompson

* + - * + Are computing/printing services most effective when concentrated on the ground floor of the Cathedral, or should they be dispersed between different floors?
        + Are the stationary computers in the lab being utilized, or should the computing services offer more laptops/tablets and renovate the space in to a general study room?
        + Is the current spread of operating systems within the lab facilitating user needs, or should the spread be changed to focus on a different operating system?
        + Are the current staffing levels at the assistance desk in the computing lab necessary, or should they be adjusted?

Our model was divided into four separate models (and thus separate programs). Across all the models, however, the users remained the same: the University students and faculty. The simulated data of interest changed for each model, as did the users’ interaction with the system. For the printers question, our servers were the eight printers, placed throughout the 38 student-available floors of the Cathedral. For the stationary computers question, our servers were the sixty-seven computers available within a lab. For the operating systems question, our servers were again the sixty-seven computers available within the lab, though with a focus on their operating system rather than just on their physical usage. And for the assistance desk question, our servers were the CSSD staffers.

To develop each model, we needed four different sets of data. For the printers question, we needed a set of tuples from printer users, describing what location they were coming from and what location they were going to. For the computer usage question, we needed to collect two counts. The first was the number of users who came in and used a computer in the lab. The second was the number of computers used by those users. This differentiation is important as, depending on their arrival, every user who came to use a computer in the lab may have used the same computer. This would have severe implications for our model and expectations. For the operating systems question, we focused on two counts: the number of Windows computers used and the number of Mac computers used. Finally, for the desk attendant question, we focused on one count: the number of users who approached the service desk for assistance.

We planned to collect our data through a single week at the computing lab. We did not ensure we gathered data from both computing labs on the ground floor of the Cathedral, though incidentally we ended up doing so. Each group member was scheduled for a handful of hour-long collection sessions at different times of day, on different days of the week. This was to ensure that our data was not biased by any predilection for high or low usage during any certain time of day. During a data collection session, a group member would remain in either computing lab and attempt to monitor and gather all four needed sets of data. This included interviewing printer users as they arrived, tracking the computers that had already been used in the lab and all new users who arrived, including whether they were using an untracked computer or a tracked computer. In addition to tracking computer usage, that group member also had to tally how many of each operating system was used during their session, and how many users approached the assistance desk for help. All of this was collected on paper and entered into an Excel spreadsheet later, or entered directly into the Excel spreadsheet. Each printer user’s data was recorded as a separate row, while the other counters were aggregated and entered within a single row.

The input modeling for our simulations was more difficult than expected, possibly due to the size of our dataset. However, we eventually discovered the distributions we needed, specifically for the printers problem. The tuple that was needed for each simulated user was described by two distributions. The first was a uniform distribution, which describes the ‘floor coming from’ variable. The second was a right-skewed normal distribution, which describes the ‘floor going to’ variable, after we ignored the first bin of the data. The mean floor going to came to be 2.44, with a standard deviation of 4.75 and a degree of skewness of .7255.

Since the number of combinations for 38 combine 8 is 48,903,492, it was not feasible due to time and processor constraints to brute force the best answer for the printer question. As such, the method done to get this result was to make two lists, append every list of printer setups to one, append the average floors traveled to the other, then find the index of the minimum average traveled floors and the associated setup in the other list. The default setup has an average of 24 floors traveled. It does not appear possible to have a setup where the average floors traveled is smaller than 14. Most range from 15 to 18.  Ultimately we can conclude that the printer setup in the Cathedral is not optimal under any circumstances, and could stand to have some printers moved to other floors.

From our data collection it was established that students in our sample come into the labs to use the computers at an average rate of 0.52 people per 60 minutes. This simulation for computer usage was difficult to quantify, and so some assumptions were made based on what was known. In general, people are using a lab computer for one of three reasons. Printing something out, to waste time between classes, or to work on a bigger project. Printing something takes under 2 minutes usually, wasting time lasts on average 40 minutes or so, and project people tend to stay for over an hour, which makes sense intuitively, but is only based on anecdotal evidence. Assuming that the midrange of people is most common, this was thought to follow a normal distribution. The simulation resulted in an average server utilization of  0.484, and thus it was concluded that under these conditions, Pitt could stand to remove several computers from the lab.

The PC and Mac utilization at Pitt are fairly even with respect to how many of each the computer labs are equipped with, such that the operating system distribution of the labs is fine and does not need to be changed. However, this simulation also supports the general computer usage data, in that utilization for individual operating systems being around 50% lines up well with the overall computer utilization also being around 50%, and as such this simulation supports cutting down on the overall number of computers. It further shows that the thinning of computer systems in the Cathedral labs would be better done in a way that is proportional to the operating system distribution.

The first lesson we learned was in the value of data. The input modeling for our various problems would have been much easier had we had more data with which to visualize the potential distributions for our problem. Thus, the procedure and schedule for data collection should be established and deliberated well in advance to avoid such problems in the future. A specific idea for the problem of interviewing a large number of people could be to design some sort of app or supply a small handout at the real-world service location. Another lesson is in the uncertainty of wanted data. We realized after completing our data collection step that some of our models would have been improved had we also gathered other pieces of data. Because we lacked this newly desired data, however, we were forced to create assumptions for our models, sacrificing accuracy. Thus we realized the reason why model conceptualization and data collection occur in tandem—so that one may be adjusted by realizations made in the other.