Cathedral Computer Labs
Team mates: Robert, Zach, Manasi, Jay

- Problem Domain:
    - Computing and Printing Services inside the Cathedral of learning

- Specific Questions:
    - 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 into 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?

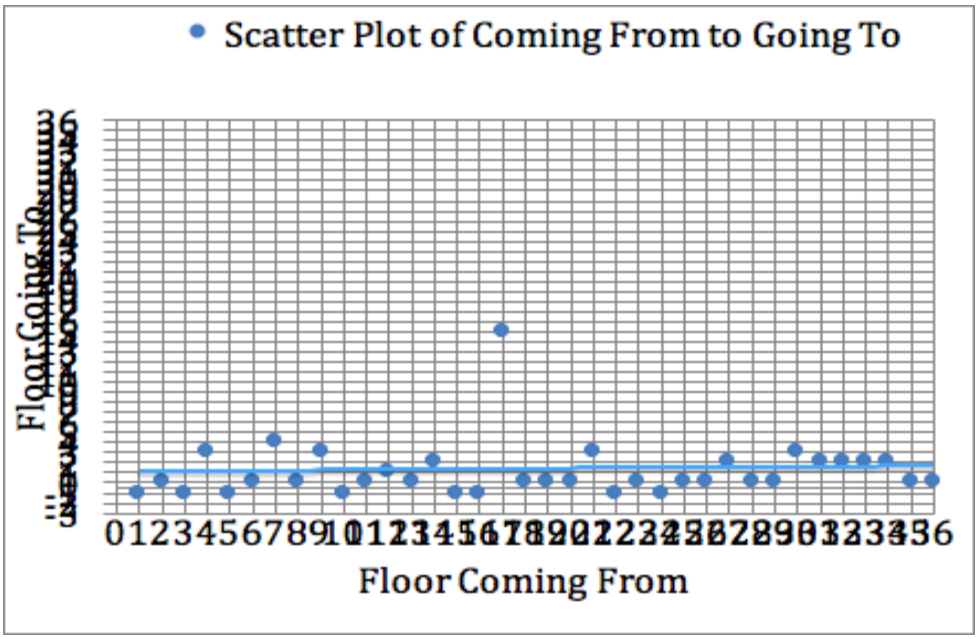
- Why these problems?
    - Scarcity of Resources
        - For the entire Cathedral, there are only two general-purpose computing labs
    - Demand on Resources
        - With over two thousand rooms, the Cathedral faces one of the largest throughputs of any University building
    - Difficulty of Movement
        - Due to the size of the building and its vertical nature, movement through the Cathedral is difficult, making access to the computing labs difficult

- Data Collection:
    - Hand-collected data during one week from both Cathedral labs
    - Each collection session took place over one hour, with sessions placed at different hours of the day
    - Data collection included:
        - Interviewing printer users to find (if applicable) what floor they came from and what floor they were going to
        - Counting the number of computer lab users who interacted with the help desk attendants

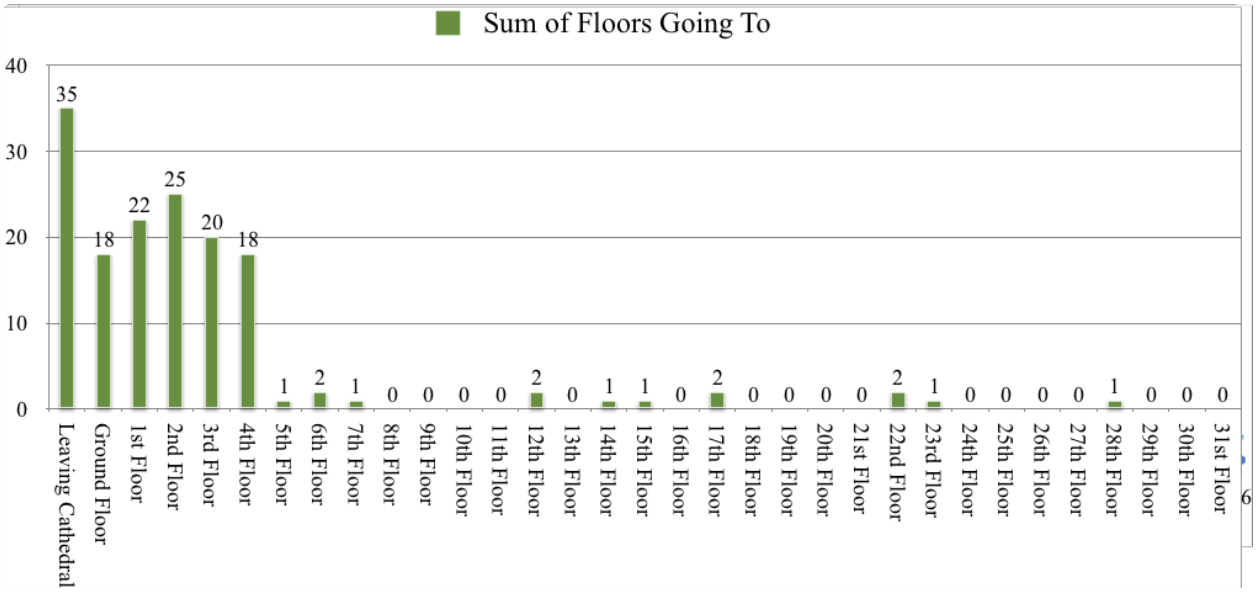
        - Counting the number of computer lab users who used a computer alongside counting the number of computers used
            - Person A and B both use Computer 1 = 2 users, 1 computer
        - Counting the number of users per each available operating system

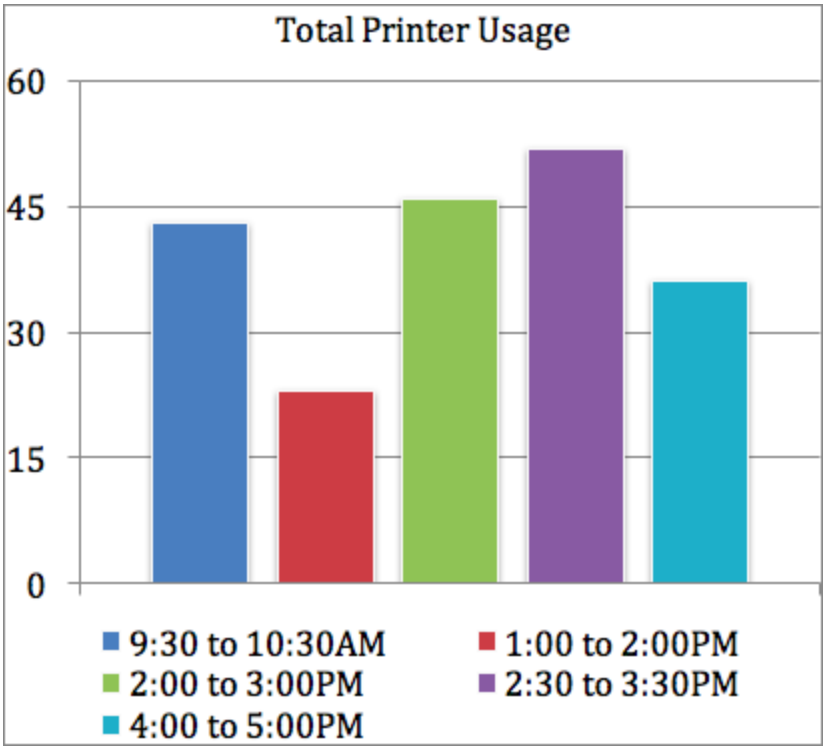
- Model Conceptualization:
    - Each question was modeled separately, as the server of interest changed with each question and there wasn't dependence between servers
    - Users: University Student/Faculty
    - Servers:
        - Printers
        - Computers (Physical Computer/Operating System)
        - Desk Attendants

- Data Visualization:
    - Data presented some difficulty in visualization.
    - When taking data, we collected users, who came down to use printer but because of some reason, we weren't able to get information about what floor they are coming from and what floor they are going to. So there weren't many data points to being the simulation, plus we had to ignore these datapoints where we didn't have the information that we were interested in. Because of this we weren't able to see any clear distribution on our plots in the beginning.
    - Eventually discovered floors coming from is distributed uniformly.



- Floors going to is distributed using right skewed normal distribution if we just ignore the first bin.





- Experimental Design:
  - To avoid over complicating or confounding results, each question was separated into a different script.

- Trial Composition
  - Printers: Single User, traveling to and from lab
  - Computers: Multiple Users, seated, working, and leaving from computers through a single simulated day
  - Desk Attendants: Multiple Users, arriving and leaving from desk through a single simulated day

- Printer Simulation:

Case	Average Floors Traveled
Base Case All Printers on Ground Floor	24.2236
Best Case 1st, 5th, 8th, 9th, 12th, 13th, 16th,33rd	14.7636
Worst Case All Printers on 38th Floor	81.0161

- Disk simulation:
  - Number of Simulated Attendants: **4**
  - Average number of desk visits in a simulated day: **8.6953**
  - Number of minutes the lab is open: **870**
  - Average number of those minutes utilized: **17.395022556286616**
  - Average Maximum Utilization: **0.004998**

- Computer Usage Simulation:
  - Average Concurrent Usage: 32.461
  - Total Number of Lab Computers: 67
  - Average Computer Utilization: .48449

- Operating System Bias Simulation:

Number of Windows PCs: 44	Number of MacOS PCs: 21
Average Number of Windows Users: 341.579	Average Number of Mac Users: 110.156

Average Number of Concurrent Windows Users: 25.786	Average Number of Concurrent Mac Users: 11.014
Average Windows Utilization: 0.586	Average Mac Utilization: 0.524

- Conclusions:
    - Printer
    - Since the number of combinations for 38C8 is 48903492, it was not feasible due to time and processor constraints to brute force the best answer. 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.

    - Help desk
    - From the collected data it was established that the help desk was used at a rate of 0.6 people per hour, or 0.01 people per minute, and an assumed average duration of that desk use was 2 minutes. This assumption was made due to the inability to time the duration accurately while also trying to collect all of the other forms of data needed. The labs are open from 7:30 am to 10 pm (870 minutes) on Monday through Thursday and there are four attendants on duty. The help desk utilization is less than one percent, and thus we can conclude four people for the two labs are unneeded.

    - Computer utilization
    - From our data collection, it was established that for our sample, students come into the labs to use the computers at an average rate of 0.52 people per 60 minutes. This simulation was difficult to quantify 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 of 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 are 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.

- If it is assumed that the people who just drop by to print are the most plentiful, then this scenario would match up well with an exponential distribution. The resulting average computer utilization from that perspective was 0.473, which is not a statistically significant difference from the normal distribution computer utilization.

- Operating system distribution
- The PC and Mac Utilization at pitt are fairly even with respect to how many of each the computer labs have, 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.

- Lessons learned:
  - Collecting different forms of data at once can be hard. Do it with a friend!
    - Collecting more data would have helped us see which distribution the dataset belongs to. Also, we should think about how we are going to collect data well in advance. In our example, there were some people coming and using printers, but we just were not able to get the information needed from them.
    - That actually made finding out which distribution these datasets come from little hard and inaccurate.

        - Also, for future reference we should definitely try and automate this data collecting step using some app or something better than actually having to go to people and ask for such information.
    - You don’t always know what all data you need until after you’re done with the data gathering step, and then might have to make some unfounded assumptions to proceed with what you have
        - To fix this issue, we should make sure to start data collecting and model conceptualization in parallel. This gives us better idea of what is exactly needed when collecting data.
    - Sometimes you don’t need to simulate multiple users of a system at once, depending on your interest in the system.