



HUMAN COMPUTATION AND VISUALIZATION LABORATORY

SMITH COLLEGE DEPARTMENT OF COMPUTER SCIENCE
NORTHAMPTON, MASSACHUSETTS 01063

DOC. TITLE: **Smith College Learning Space Assessment Report**

SURF STUDENTS: Sherry Zhenyao Cai '19
Stella Chen Li '19
Chris Xiaoyue Tan '21
Hening Zheng '20
Starry Yujia Zhou '20

FACULTY ADVISER: R. Jordan Crouser, Assistant Professor of Computer Science

PREPARED FOR: Smith College

DATE SUBMITTED: March 24, 2019

Table of Contents

1	Overview	3
2	Methods and Materials	3
2.1	Understanding the Data	3
2.2	Survey Results - Master.xlsx	3
2.3	MyCampus Results.csv	4
3	Analysis of Scores: Sasaki Score vs. HCV Score	5
3.1	Sasaki Score	5
3.2	HCV Score	5
3.3	Combining Two Scores	7
4	Sasaki Score vs. Faculty and Student Preference	7
4.1	Introduction	7
4.2	Data Wrangling	8
4.3	Creating the scatter plot	8
4.4	Discussion	8
5	Other Features that Might Influence Faculty and Student Preference	10
5.1	Number of Seats	10
5.2	Faculty Member's Academic Rank	11
6	General Usage	11
6.1	Frequency of Classrooms Being Used	11
6.2	Average Fullness and Usage Standard Deviation	12
6.3	Chair Usage Efficiency	14
7	Interface	16
7.1	Equipment Checklist	16
8	Future Work	17
A	Code for Wrangling “MyCampus Results.csv”	17

1 Overview

In 2017-2018, Smith College contracted Sasaki Associates, Inc. to conduct a survey of Smith College study space analysis. During the summer of 2018, five undergraduate student researchers in the Human Computation and Visualization Laboratory conducted additional analysis of the collected data under the supervision of Prof. R. Jordan Crouser.

This reanalysis of Smith Space Usage focused on evaluating utilization rates, technology deployment, campus geography and the physical condition of classroom spaces. The goal of this study is to identify potential methods to support faculty with their diverse pedagogy and enhance student learning experiences. The assessment was summarized as the convergence of the following two themes: Teaching, and Allocating. There are several essential goals we hope to achieve at the end of this summer:

- Understand current utilization of classrooms
- Assess adequacy of instructional technology
- Develop a decision-making interface to help resource allocation and room assignment

2 Methods and Materials

2.1 Understanding the Data

The reanalysis was conducted based on multiple data sets from the Learning Space Assessment data provided by Sasaki and the previous classroom assignment from 2009-2018 provided by Smith College registra office. In the Learning Space Assessment data sets, a comprehensive and detailed measurement of Smith College spaces and calculated scores based on Sasaki's weighted algorithm are included in the different data frames. The data sets provided by Smith College Registrar contains information about paired course and classroom, the course type, maximum enrollment, actual enrollment, etc.

2.2 Survey Results - Master.xlsx

The excel document "Survey Results - Master.xlsx", with 810 observations of 173 variables, contains all answers to a survey on learning space usage conducted on Smith campus in spring 2017. We used R to wrangle this dataset. Since answers from students and faculty members were mixed together in this document, we first used the filter function in R to filter out answers from these two groups by looking at the tenth column, which is the participants academic rank if they are a faculty member or blank *na* if they are a student. By doing so, we filtered out 621 responses from students and 187 from faculty members.

Then, to find out classrooms used most frequently by faculty members, we looked into the eighteenth column, which contains answers to the faculty-only question "please identify a room they teach most frequently. We grouped the dataset by room numbers this column, counted the number of times each room was identified, and arranged the result in descending order using `group.by`, `summarise`, and `arrange` functions respectively in R. We also grouped the dataset by faculty members academic department to get each departments preferences.

Next, to find out classrooms most preferred by faculty members, we performed the same procedure as mentioned above on the thirty-ninth column, which contains answers to the faculty-only question "please list up to 3 of your most preferred rooms for teaching. We also grouped the dataset by faculty members academic department to get each departments preferences.

In addition, to find out students and faculty members opinions on how frequently Smith’s teaching spaces are effective for teaching and learning, we grouped both groups’ answers by “OVERALL column, which contains answers to the question “how frequently does Smith’s teaching spaces meet the following criteria: effective for teaching and learning?, counted the number of times each answer (one of often, sometimes, never), and arranged the result in descending order using group_by, summarise, and arrange functions respectively in R. We also grouped the dataset by students/faculty members academic department to get each departments preferences.

Next, to find out students and faculty members opinions on how frequently Smith’s teaching spaces are at a quality level and condition befitting Smith’s academic reputation, we performed the same procedure as mentioned above on the “X_165 column, which contains answers to the question “how frequently does Smith’s teaching spaces meet the following criteria: at a quality level and condition befitting Smith’s academic reputation?. We also grouped the dataset by students/faculty members academic department to get each departments preferences.

2.3 MyCampus Results.csv

The csv document “MyCampus Results.csv”, with 810 observations of 173 variables, contains all answers to a survey on learning space usage conducted on Smith campus in spring 2017. Each row in this document corresponds to an answer to one of the questions in the survey based on its icon_id. Each row has a feature “customdata”, which contains the “buildingId”, “roomId”, and “option” mentioned in the answer. We used Python to wrangle this dataset.

In order to get the room mentioned by each answer, we first used the split function in Python to each extract the buildingId and roomId in each row, then we used pandas.merge to merge it with the sheet “Index in “Room Utilization.xlsx”, finally, by adding the column “Scheduled Buildings and the column “room_id”, we got the room mentioned in each answer in the format buildingName_roomNumber, as we can see in the first part of Appendix A.

Then, we defined a function to wrangle wrangle “MyCampus Results.csv” based on “icon_id”, as we can see in the second part of Appendix A. For any given ‘icon_id”, the function looks into:

- Overall results of rooms on campus
- Overall results of rooms on campus by building
- Overall results of buildings on campus
- Results of rooms on campus by profile (faculty, staff, undergrad, grad)
- Results of rooms on campus by building by profile
- Results of buildings on campus by profile
- Undergrad’s results of rooms on campus by house
- Undergrad’s results of rooms on campus by building by house
- Undergrad’s results of buildings on campus by house
- Faculty’s results of rooms on campus by academic rank
- Faculty’s results of rooms on campus by building by academic rank
- Faculty’s results of buildings on campus by academic rank

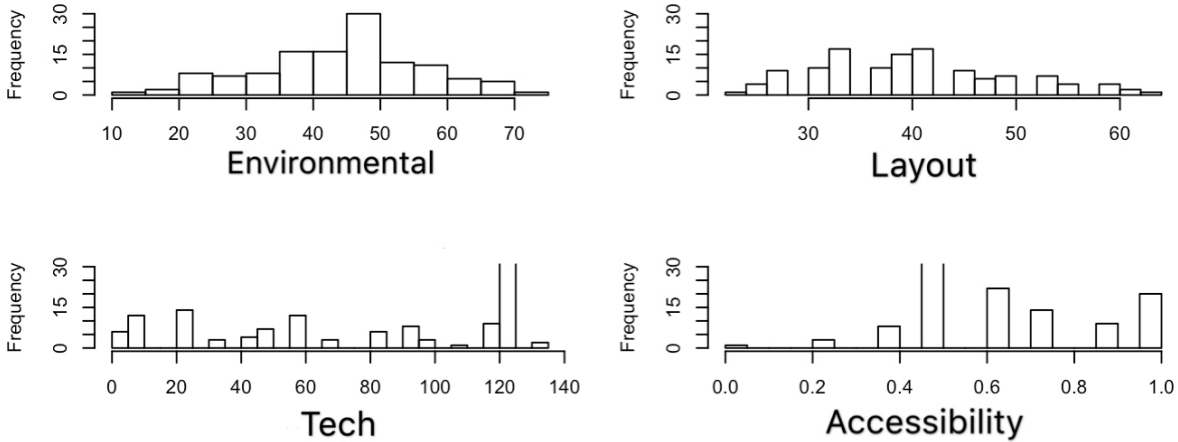


Figure 1: A histogram for the numerical explanatory variables in Sasaki scores.

3 Analysis of Scores: Sasaki Score vs. HCV Score

3.1 Sasaki Score

Sasaki Association provided Smith College a detailed evaluation of 174 learning spaces, giving a evaluation score on specific aspects (room information, general data, location, environment, layout, technology, safety and accessibility) and a weighted total score. Among these 8 predictors, environment [13.33, 73.33], layout [22.22, 63.89], technology [0, 133.33] and accessibility [0, 1] are numerical scores with their ranges in the bracket; and general data (2), location (2), safety (3) and information (4) are categorical variables with their number of levels in the parenthesis. We also conducted an analysis on how internally-related each physical character score is with each other. We listed the correlation values of each pair of numerical scores, and found three pairs of factors that are mildly correlated: technology and information (correlation value: .41), accessibility and layout (.33), accessibility and technology (-.30).

3.2 HCV Score

While these scores are based on careful assessments of many physical characters of the classrooms, we wonder if they correctly represent how likable they are to users and how likely professors are to reuse them. Based on our data of the favorite classroom survey and classroom assignments for faculty members of the last 10 years, we proposed a formula that calculates an HCV score to assess how likable and reusable a classroom is for students and faculty members. Combining each category of the Sasaki scores and HCV scores, we aim to find out whether the physical evaluations by Sasaki indicate how much people actually like and use the classrooms.

There are two components of our HCV score:

P = number of people (both faculty members and students) who list this classroom as their favorite one in 2009 – 2018

Q = average number of total enrolled students in the classroom in 2009 – 2018

Variable 1	Variable 2	Correlation
tech	information	0.41
accessibility	layout	0.33
tech	environmental	0.21
accessibility	environmental	0.17
information	room capacity	0.16
layout	environmental	0.15
accessibility	final score	0.14
tech	room capacity	0.08
tech	final score	0.03
environmental	final score	0.01
environmental	information	0.01
layout	final score	-0.02
accessibility	room capacity	-0.03
layout	room capacity	-0.08
room capacity	final score	-0.09
information	final score	-0.09
accessibility	information	-0.21
environmental	room capacity	-0.22
tech	layout	-0.22
layout	information	-0.26
accessibility	tech	-0.3

Table 1: A table illustrating the correlation of each pair of variables in Sasaki scores.

The rate of a classroom being ones favorite after using it = $\frac{P}{Q}$

S = number of times a classroom was used by a professor X in 2009 – 2018

T = total number of courses taught by a professor X in 2009 – 2018

The index to evaluate how irreplaceable a classroom is for a faculty member = $\frac{S - 1}{T}$

Combining these two factors, we have our HCV score formula as:

$$\text{HCV Score} = \frac{P}{Q} + \frac{S - 1}{T}$$

3.3 Combining Two Scores

To assess whether these two evaluations are correlated, we ran multiple linear regressions on the dataset. Having our normalized HCV scores (ranging from -0.18 to 1.82) as the response variables and the 8 physical character scores as explanatory variables, we saw that only the accessibility factor has a statistically significant effect on the HCV score (on a 90% confidence interval): for each 1 score increase of accessibility factor, the HCV score is expected to increase by -.01 to .35.

We also conducted a principal component regression analysis on the component of HCV score with the 5 numerical predictors (information, environmental, layout, technology, accessibility). The result shows that 5 components predict 99.997% of the training data, 1 component predicts 90.32% of the training data, and the scatter plot for prediction plot is not consistent. While this PCR result shows that the 5 numerical predictors do not make a good prediction model for the HCV score (as we expected it would not), the fact that the component of these 5 numerical predictors only can predict more than 90% of the result is alarming in terms of understanding the shape of Sasaki scores. Because we treated 4 other variables (general data, location, safety, information) as categorical variables, they were not included in the PCR analysis. This means that the combination of these categorical variables together only count less than 10% of the resulting HCV score. Therefore, it is obvious that the each Sasaki score is highly correlated with very limited variance. To gain a better insight of the Sasaki score, we would like to use a classification example to test out if higher and lower HCV scores correspond to these predictors better than all scores do as a group in future analysis.

4 Sasaki Score vs. Faculty and Student Preference

4.1 Introduction

While Sasaki took into account many physical features, such as size, number of chairs, accessibility, etc., when calculating the score for a room, it did not take any faculty or student's preference into consideration. As a result, we started to wonder whether the Sasaki score would match faculty and students' preferences, which intrigued us to look into the correlation between the Sasaki score and people's preferences.

4.2 Data Wrangling

4.2.1 Sasaki's weighted overall room score

First of all, in order to get the Sasaki score, we selected the column 'WEIGHTED OVERALL ROOM SCORE' from the sheet "Smith Survey_Total" in "Data Learning Space Assessment.xlsx".

4.2.2 Faculty members and students' preferences

In addition, in order to find out faculty and students' preferences, we first subsetted all the rows in "MyCampus Results.csv" that were answers to the question on favorite place on campus (i.e. all rows whose 'icon_id' is 'favorite_place'). Then, we further subsetted rows whose 'profile' is 'Faculty member' and 'Undergrad Student' respectively. Next, for faculty members and undergrads, respectively, we grouped their answers by building and room ID and counted the number of people preferring each room. Finally, we put the results into two separate dataframes – one for faculty members, another for undergrads.

4.2.3 Merging Sasaki score and Smith preference

After we wrangled out both the Sasaki score and the Smith preference, we inner merged the Sasaki score with faculty members' preferences and undergrads' preferences, respectively, on building and room id.

4.3 Creating the scatter plot

Now that we have both the Sasaki score and the Smith preference, we generated the following scatter plot using Plotly, with Sasaki's weighted overall room score being the x axis and the number of people preferring the room being the y axis. Points in this scatter plot are divided into two groups: the set of blue points represent rooms that are preferred by faculty members (i.e. in the survey, each of these rooms was listed as at least one faculty member's favorite place), whereas the set of orange points represent rooms that are preferred by undergraduate students.

4.4 Discussion

4.4.1 Faculty and students had similar preferences

As we can see in Figure 2, faculty preference counts are overall lower than student preference counts, since fewer faculty members answered the survey. Nonetheless, it is noteworthy that faculty and students at Smith had similar preferences for classrooms: rooms preferred by more faculty members, such as Ford 240, are also preferred by more undergraduate students; similarly, rooms preferred fewer faculty members, such as Bass 103, are also preferred by fewer undergraduate students.

4.4.2 No direct correlation between Sasaki score and people's preferences

Despite faculty and students' consensus on their favorite classrooms, as we can see from the scatter plot, for both faculty members and undergraduate students, there is no direct correlation between the Sasaki score and their preferences. Rooms that are most preferred by faculty and students, such as Ford 240 and McConnell 103, received mediocre Sasaki scores, varying from 55 to 70. On the other hand, rooms that received very high Sasaki scores, such as Bass 102, are preferred by few faculty members and undergraduate students.

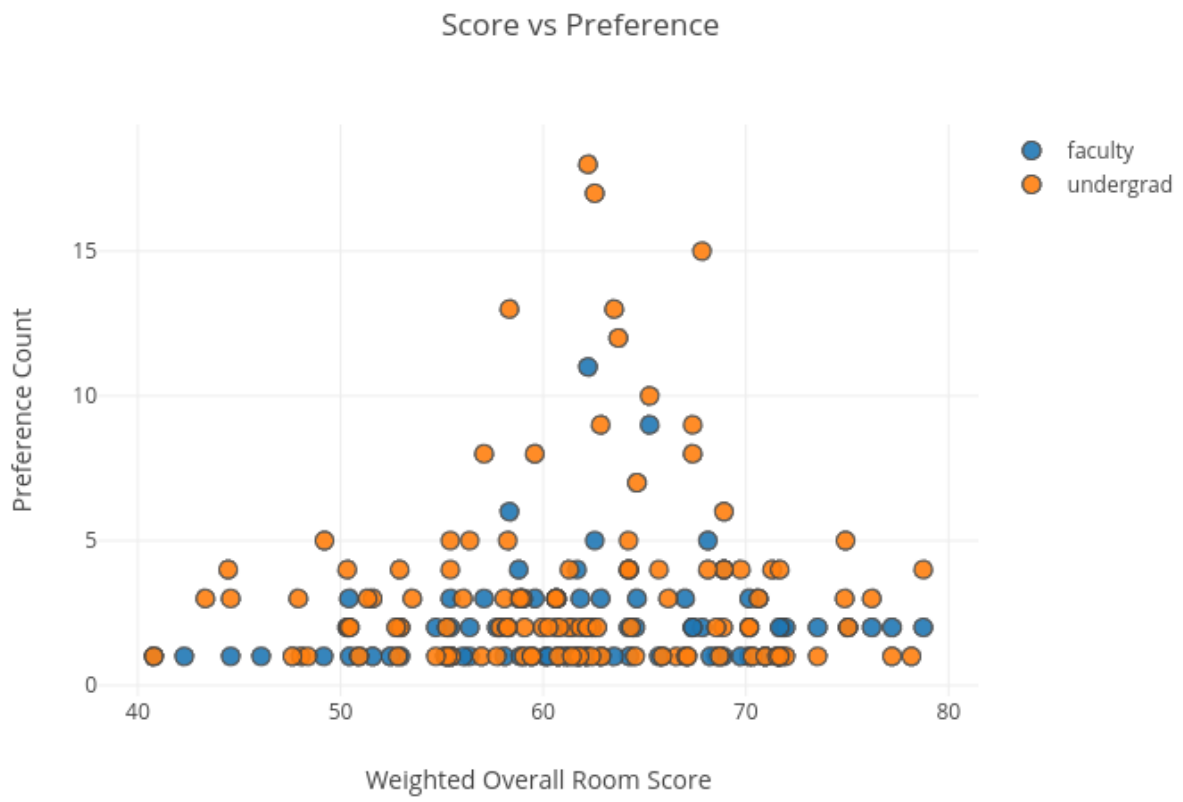


Figure 2: Sasaki Score vs. Faculty and Student Preference

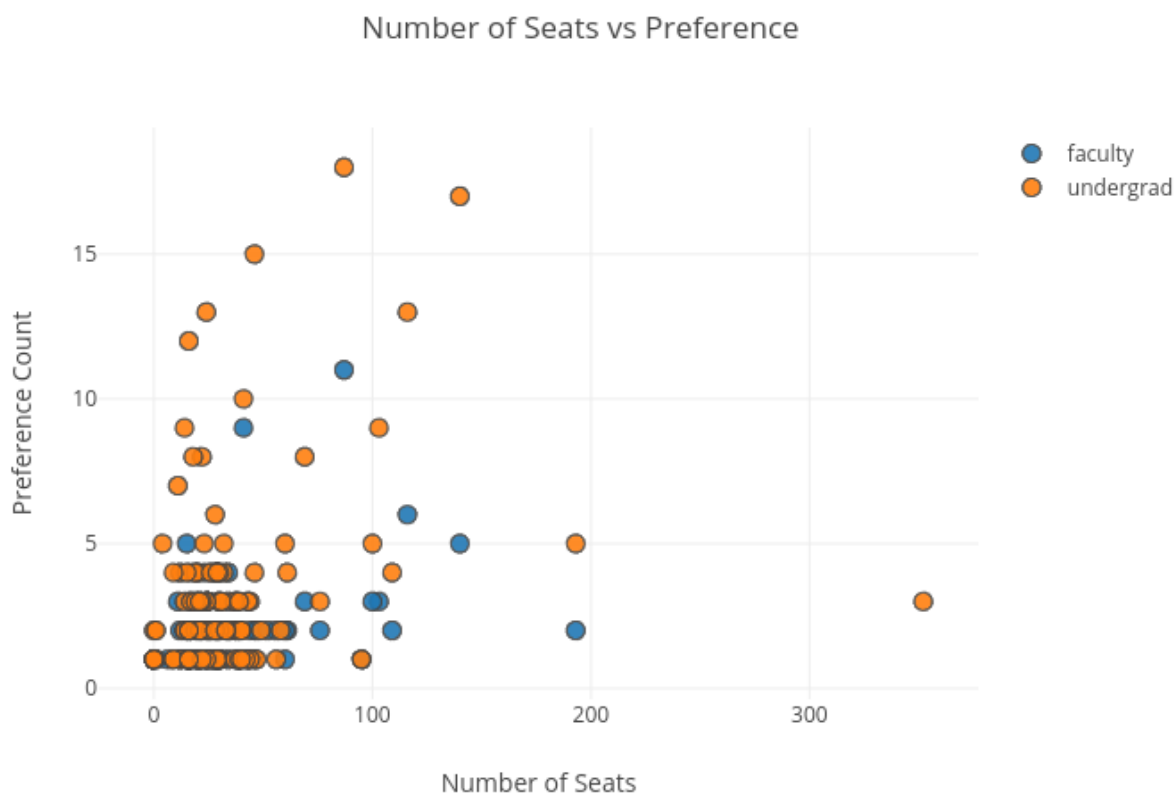


Figure 3: Number of Seats vs. Faculty and Student Preference

5 Other Features that Might Influence Faculty and Student Preference

5.1 Number of Seats

While looking into the correlation between the Sasaki score and faculty and students' preferences, we found out that rooms that are most preferred by faculty and students, such as Ford 240 and McConnell 103, are also the comparatively larger rooms on campus, which intrigued us to look into and visualize the correlation between the size of a room (measured by number of seats in the room) and the number of people who prefer this room.

As we can see in Figure 3, overall, rooms with more seats are preferred by more people; on the other hand, when the number of seats in a room is above 150 (which applies to only two rooms on campus), the higher the number of seats a room has, the fewer the people who prefer this room. This result corresponds to the reality: rooms with larger size have more people who have been there, thus they have a higher possibility of being preferred; however, when a room has more than 150 seats, since there is hardly any class whose size goes above 150, only very few classes will be scheduled in these rooms, thus they have a lower possibility of being preferred. This result is consistent with the overall preference (by faculty, staff, undergrad, and grad students) as well as the preference by faculty/undergrad.

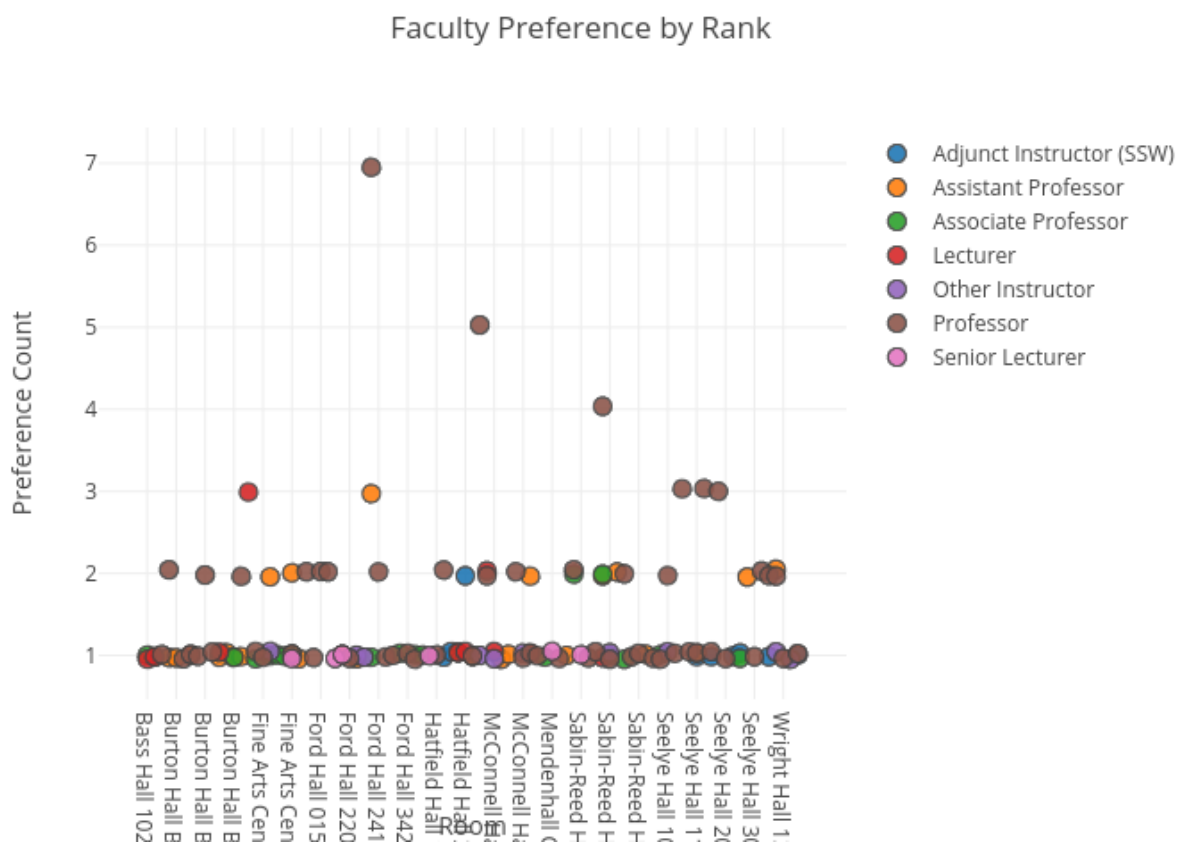


Figure 4: Faculty Preference by Academic Rank

5.2 Faculty Member’s Academic Rank

Aside from the size of a classroom, we were also wondering if faculty members with different academic ranks would have different preferences for classrooms. We tried to look into and visualize the correlation between a faculty member’s academic rank and their preference using Plotly. However, as we can see in Figure 4, the scatter plot generated by Plotly does not clearly reflect the correlation since points in the scatter plot overlap a lot. Due to the small sample sizes, most of the classrooms are preferred by very few (one or two) faculty members of each academic rank, which leads to the overlapping problem. We can see that there are more professors who prefer certain classrooms, such as Ford 240 and McConnell 103. Nonetheless, this might only be a result of the relatively larger sample of professors.

6 General Usage

6.1 Frequency of Classrooms Being Used

While we took into consideration features such as distance and location of departments, equipment, professors’ preference and students’ preference, we also pay attention to how frequently one classroom is used by each professor and each course. Moreover, we further put the frequency of usage

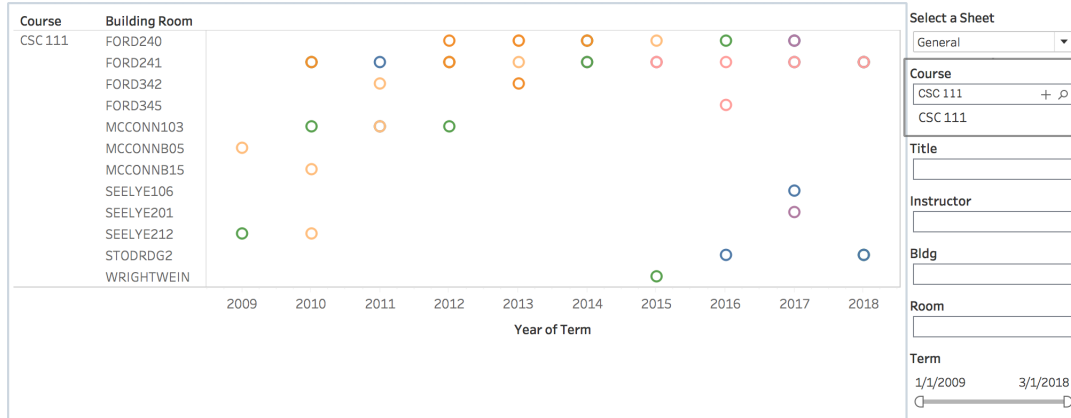


Figure 5: Classroom usage for course CSC 111 from 2009-2018

into time frame, so we can investigate whether professors have changed their preference through years and whether classrooms used frequently before are no longer suitable for some courses. The data we used is from Smith College Course Catalog 2009-2018. It shows the information of instructor and classroom for every course taught each year. We used python, and pandas library to wrangled the data and merged course, classroom, instructor, and time into the same data frame. Furthermore we used Tableau to visualize the data. In Tableau visualization, we set four filters: Course, Instructor, Classroom, and Term, so we can compared the preference of instructors and courses.

To take CSC 111 as an example, we can see that while classrooms have been used frequently in 2009 and 2010, the course now are taken frequently in Ford 241 and Ford 240. Moreover, while professor Marshall(the pink dot) prefers Ford 241, professor O'Rourke prefers Stodrd2.

6.2 Average Fullness and Usage Standard Deviation

HCV scores made us realize the importance of analyzing the usage efficiency of each classroom. Therefore, we further wrangled the usage history of each classroom. We calculated the total enrolled number of each classrooms in the past 10 years as the total number of people who have ever used the classrooms for taking classes. Having the total usage number divided by the number of classrooms used in the past 20 terms (2 terms for 1 year), we have the average number of people using one classroom for classes each term. Then, we calculated the fullness portion using the average usage divided by the capacity of each classroom. As the fullness portion data shows, 7 of the 189 classrooms (Sab-Rd 308, Aniswo-Sqcts, Sab-Rd 328A, Sab-Rd 103, Lyman 110, Sab-Rd 419, Sab-Rd 314) on the list have a fullness above 1 (from 149% to 112%), which means that they usually have more people than the capacity taking classes at the same time.

We also calculated the standard deviation of the usage number of each room over the past 20 terms. The rooms with the most consistent amount of user is Wright 126, Ford 323 and Boat House. The rooms that are used most randomly (the ones whose amount of users vary the most over the terms) are Wright-Wein, Stodrd G2, and Mcconn 103. Meanwhile, many study spaces have significantly low fullness portions (lower than 10%), such as Ainswo-ITT, Scott Gym, Sage-Weeney, due to their very large capacities. Among average sized classrooms, Theatr 114, Theatr 207A, Sage-Recit and Hillyr-Graham have the lowest fullness portion of under 30%. The building with the highest fullness portion is LYMAN with 88% of the rooms being full and Sab-Rd with 77%, and the one with

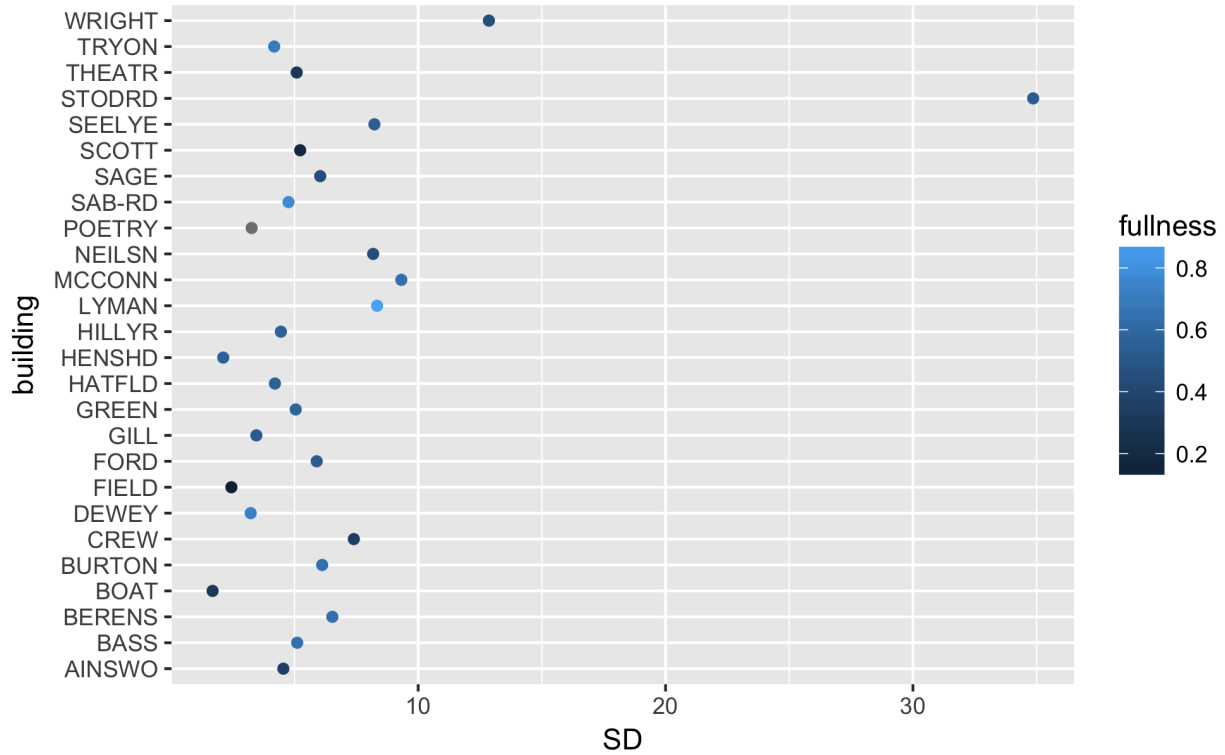


Figure 6: SD of the usage of each building, with the color of each point representing its fullness

the lowest portion is Field and SCOTT with 11% and 17%. However, in the data of 26 buildings, many buildings have very few classrooms data available - for example, Lyman has 4 classrooms, Field has 1, and Scott has only 3. Therefore, the fullness portion ranking may not apply if more or all classrooms in these buildings are included. Meanwhile, with our current available data, the buildings with the lowest average SD for usage amount are Boat (1.69) and Henshd (2.12), and the ones with the highest average SD are Stodrd (34.86) and Wright (12.86), indicating that these buildings have usage amounts that vary vastly among different terms.

We also calculated the standard deviation of the usage number of each room over the past 20 terms. The rooms with the most consistent amount of user is Wright 126, Ford 323 and Boat House. The rooms that are used most randomly (the ones whose amount of users vary the most over the terms) are Wright-Wein, Stodrd-G2, and Mcconn 103.

Building	Standard Deviation	Fullness
AINSWO	4.55	0.35
BASS	5.11	0.64
BERENS	6.53	0.64
BOAT	1.69	0.31
BURTON	6.12	0.64
CREW	7.4	0.34
DEWEY	3.23	0.72
FIELD	2.45	0.11
FORD	5.9	0.53
GILL	3.46	0.53
GREEN	5.05	0.58
HATFLD	4.21	0.56
⋮	⋮	⋮

Building	Standard Deviation	Fullness
HENSHD	2.12	0.57
HILLYR	4.45	0.57
LYMAN	8.34	0.88
MCCONN	9.32	0.64
NEILSON	8.18	0.45
POETRY	3.27	NA
SAB-RD	4.76	0.77
SAGE	6.04	0.44
SCOTT	5.23	0.17
SEELYE	8.23	0.54
STODRD	34.86	0.53
THEATR	5.09	0.3
TRYON	4.18	0.7
WRIGHT	12.86	0.45

6.3 Chair Usage Efficiency

6.3.1 Introduction

In 2017, Smith College Office of Campus Sustainability and Planning sent out an on-line questionnaire to current Smith students and faculties. The questionnaire asks about people’s learning or teaching experience and aims to collect information about people preference and comments for the academic-use classrooms in Smith College. After the questionnaire was sent out, the office received 810 responses and then arranged the results into an excel table named ”Survey Results.xlsx”. The questionnaire has a final question asking survey takers to leave some free thoughts about classrooms; and it is interesting to see that among 214 people who answered the final question, 34 of them complained about chair arrangement. Some of their answers are: ”I think we need more classrooms at smith so we have more options. we need more big classrooms with movable chairs”; ”while a nice option for movement and group work, putting rolling chair desks in smaller rooms is a bad idea.”; ”The furniture is not always in appropriate ratio to class or classroom size.”... The complains and negative comments show that the current arrangement of chairs does not fully satisfy faculties’ pedagogy nor students’ needs for a better learning experience. For a professor whose pedagogy based more on students’ discussion than heavy lectures may prefer to have more spacious classroom with relatively small amount of flexible chairs so that the class can easily move around; for a professor whose teaching style is more lecture-based may feel more comfortable to have smaller classroom with sufficient supply of chairs for each student.

In order to further comprehend the comments on chair arrangement, our lab conducted a research on analyzing chair-usage efficiency. In the research, we try to find out: 1) how much the need for chair is overestimated/underestimated; 2) how can we optimize current chair arrangement system.

6.3.2 Source of Data

The data we need for conducting the study includes: -information about chair condition for each classroom (the number of chairs in a classroom; the flexibility of the chairs), -information about classes held in that classroom (the department of which class belongs to; the number of enrollment of that class; the pedagogy that class uses); -information about the classroom condition (the size of that classroom; de-densified capacity; seat ranges.

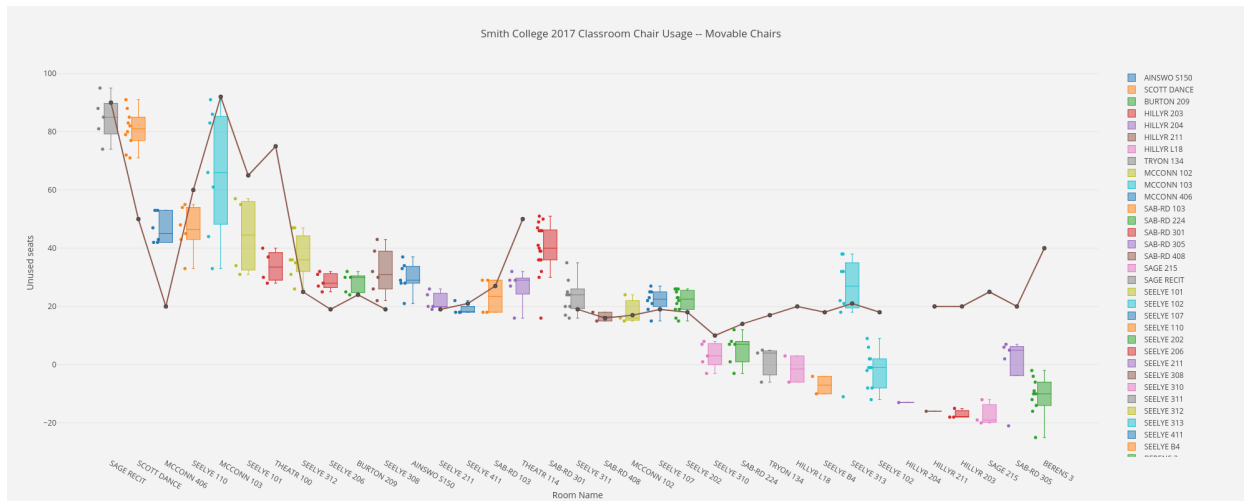


Figure 7: An analysis for classroom chair usage efficiency.

In order to make sure all the data correspond with each other for the same time, our lab mainly used the data from 2017. Here are the sources of data we used:

- Chair information for each classroom
In Sasaki's *Learning Space Assessment/Smith Survey_Total Data*, column "401_furniture_types" indicates the flexibility of the chairs. In Sasaki's *Learning Space Assessment/Utilization*, column "Number of Seats" shows the number of seats placed in a classroom by the time Sasaki did the survey.
- Class schedule for each classroom
Course enrollment information is collected from *Smith College Course Catalog Fall 2017*. It shows, for each time a classroom was used for a class, how many people were in the classroom.
- Classroom condition
Size, seat ranges and de-densified capacity of each classroom are collected from Sasaki's *Learning Space Assessment/Utilization*.

6.3.3 Data wrangling

The tool used for data wrangling is python's data manipulation package *Pandas*. First of all, we merged all the useful information into a single dataframe where each course is the main observation and each the classroom for holding that course is the secondary observation. We then used "number of seats - actual enrollment" to get the number of unused seats for each course. After that, based on the movability of chairs, we first filtered out fixed chairs since classrooms with fixed chairs are not rearrangeable; we then filtered out movables chair whose number of chairs had already been optimized — $0 < \text{number of seats} - \text{actual enrollment} < 15$. Then we focused on the rest of the observations and made two plots — one with "number of unused chairs" as y axis, another with "De-densified Capacity" from Sasaki's assessment as y axis, all in descending order.

6.3.4 The plot

6.3.5 How do we make use of these results?

The next step for having an analysis for square feet per seat would be to provide insight into the seat arrangement for rooms with different type and size, and therefore accommodate faculties with different demands. For example, one goal would be to optimize square feet per seat in order to relieve the congestion and provide more maneuverability within the classrooms.

7 Interface

7.1 Equipment Checklist

Sasaki's Learning Space Assessment provided us with detailed information of whether each classroom has a list of equipments, such as different kinds of writing boards, types of door hardwares, electronic players. To make this information directly helpful to faculty members when they select preferred classrooms or to familiarize themselves with what different classrooms have to offer, we selected some equipments from the list that we consider the most important for faculty members considerations, and created a web interface to for people selecting classrooms with technical requirements as inputs and receive a list of eligible classrooms as output. If there are too few eligible rooms, a professor may decide to de-select some less-needed requirements and get more classrooms as options.

We also have multiple levels of seat ranges and building areas (divided by buildings locations) as options. These two factors include 6 levels of seat ranges: 1-15, 16-25, 26-50, 51-75, 76-150, and 5 divided areas: 19 classrooms in Ainswo, Sage, Theatr, Scott, Crew; 10 classrooms in Berens, Capen, Lyman, Tyron; 62 classrooms in Burton, SAB-RD, Bass, McConn; 20 classrooms in Ford, Stoddard; and 63 classrooms in Seelye, Hatfld, Dewey, Hillyr, Wright.

Our current technology menu includes the following options: data projector, lectern, computer, TV, Apple TV, blue ray, control system, two kinds of cameras (document camera, web camera), various types of writing equipments (electronic whiteboard, chalkboard, smart board, whiteboard), mic support, and movable chairs.

Combining these two aspects, we can help faculty members to get a sense of what classrooms may best fit both the characteristics of the classes (which building/area does the subject usually holds classes in, how many students are expected to enroll in the class) by choosing building areas and capacity, and their own technology pedagogy needs by selecting items on the technology menu.

However, currently, the checklist may not be completely informative due to some missing data issues: among the 174 classrooms in the list, 48 does not have seat range information, and in each technology option, a small number of classrooms does not have a record. So further research and data gathering of the details are needed.

7.1.1 Discussion

In the future, we would like to create a more effective and convenient interface that will meet more detailed needs. For example, we would like to develop our current menu of select and de-select the exact item to tell us how many of these would you like. Because for many faculty members whose requirements cannot be all met and would not mind having one or some less, their requirements can be something like: "I want a classroom in X area with 4 out of these 5 requirements met, and it does not matter which one is not met", then our checklist is to have it generate an order of all classrooms, ranked by how much they met the requirements. The highest ranked classrooms will

be those that meet all the requirements, and then the ones that meet most of the requirements (i.e. in the required area with the required capacity, but with one technology requirement missed), and the lowest ranked classrooms will be those who do not meet any of the requirements. By having that output available for faculty members, they will have a much clearer picture of how well each classroom satisfies their needs, and thereby making choices that enable them to weight the pros and cons in the selection process best.

Furthermore, we would like to advance the interface by also consider time arrangement for each class. Our ultimate goal is to have the interface generate a comprehensive schedule for all classes that optimizes each of their needs, thereby hugely accelerates the classroom selection process for the registrar office.

8 Future Work

In future study on this dataset, we would like to advance the research on the following aspects:

- We would like to further our study on how distance of a classroom to one’s workspace or dorm room influences their preferences of it. We speculate that distance is an important aspect of the accessibility of study spaces for students. However, because the survey on faculties’ and students’ classroom preferences are all anonymous, we cannot conduct further analysis based on this data. Therefore, we would like to suggest Smith College Office of Institutional Research to add questions that address people’s locations in the next survey.
- As mentioned in the discussion chapter of **Interface**, to accelerate the process of classroom assignment, we would like to include schedule into consideration. We would like to utilize To have each class associated with its time block, we could advance the current web interface to work on the scheduling much clearer and more efficient to come up with a more comprehensive schedule.

Appendix A Code for Wrangling “MyCampus Results.csv”

```

1 import numpy as np
2 import matplotlib.pyplot as plt
3 import pandas as pd
4
5 survey = pd.read_csv('/Users/heningzheng/Documents/SURF/ClassroomData/MyCampus
  ↳ Results.csv')
6 room_util = pd.read_excel('/Users/heningzheng/Documents/SURF/ClassroomData/Room
  ↳ Utilization.xlsx', sheet_name='Index')
7 survey['bld_id']=survey['customdata'].str.split(",").str[0].str.split(":").s
  ↳ tr[1]
8 survey['room']=survey['customdata'].str.split(",").str[1].str.split(":").str
  ↳ [1]
9 survey = survey.drop(survey[survey['room'] == 'NOT_SPECIFIED'].index)
10 survey = pd.merge(survey, room_util.drop_duplicates(), how='inner',
  ↳ left_on='bld_id', right_on='CODE')
11 survey['BLD_RM'] = survey['Name']+' '+survey['room']
12 survey['bld_rm_id'] = survey['bld_id']+'_'+survey['room']
13 survey['sbld_rm_id'] = survey['Scheduled Buildings']+'_'+survey['room']

```

```

14 survey['profile'] = survey['profile'].str.replace('{','').str.replace('}','').st
   ↪ r.replace('","').str.replace('?','')
15 survey['house'] = survey['house'].str.replace('{','').str.replace('}','').str.re
   ↪ place('","').str.replace('?','')
16 survey['faculty'] = survey['faculty'].str.replace('{','').str.replace('}','').st
   ↪ r.replace('","').str.replace('?','')
17 survey.rename(columns={'Name': 'Scheduled Buildings', 'Scheduled Buildings':
   ↪ 'Name'}, inplace=True)
18 survey.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/survey.csv')
19
20 def check(icon_id):
21     criterion = survey.loc[(survey['icon_id']==icon_id),:]
22
23     #Overall results of rooms on campus
24     overall = criterion.groupby(['Scheduled Buildings', 'room'])[['icon_id']].cou
   ↪ nt().sort_values(by='icon_id', ascending=False)
25     overall.columns = ['count']
26     overall.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id+'/'
   ↪ overall.csv')
27
28     overall_id = criterion.groupby(['bld_rm_id', 'BLD_RM'])[['icon_id']].count().
   ↪ sort_values(by='icon_id', ascending=False)
29     overall_id.columns = ['count']
30     overall_id.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id
   ↪ +'/overall_id.csv')
31
32     #Overall results of rooms on campus by building
33     by_bld = overall.sort_index(level=[0], sort_remaining=False)
34     by_bld.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id+'/'
   ↪ y_bld.csv')
35
36     #Overall results of buildings on campus
37     bld = overall.groupby(['Scheduled
   ↪ Buildings']).sum().sort_values(by='count', ascending=False)
38     bld.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id+'/'
   ↪ bld.csv')
39
40     #Results of rooms on campus by profile
41     by_profile = criterion.groupby(['profile', 'Scheduled
   ↪ Buildings', 'room'])[['icon_id']].count().sort_values(by='icon_id', ascend
   ↪ ing=False).sort_index(level=[0],
   ↪ sort_remaining=False)
42     by_profile.columns = ['count']
43     by_profile.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id
   ↪ +'/by_profile.csv')
44
45     #Results of rooms on campus by building by profile
46     by_bld_by_profile = by_profile.sort_index(level=[0,1], sort_remaining=False)

```

```

47 by_bld_by_profile.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+
↳ icon_id+'/by_bld_by_profile.csv')
48
49 #Results of buildings on campus by profile
50 bld_by_profile = by_profile.groupby(['profile', 'Scheduled Buildings']).sum()
↳ .sort_values(by='count', ascending=False).sort_index(level=[0],
↳ sort_remaining=False)
51 bld_by_profile.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+ico
↳ n_id+'/bld_by_profile.csv')
52
53 #Undergrad's results of rooms on campus by house
54 undergrad = criterion.loc[(criterion['profile']=='Undergrad Student'),:]
55 undergrad_overall = undergrad.groupby(['Scheduled Buildings', 'room'])[['icon_
↳ _id']].count().sort_values(by='icon_id', ascending=False)
56 undergrad_overall.columns = ['count']
57 undergrad_overall.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+
↳ icon_id+'/undergrad_overall.csv')
58
59 undergrad_id = undergrad.groupby(['bld_rm_id', 'BLD_RM'])[['icon_id']].count(
↳ ).sort_values(by='icon_id', ascending=False)
60 undergrad_id.columns = ['count']
61 undergrad_id.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_
↳ id+'/undergrad_id.csv')
62
63 undergrad_by_house = undergrad.groupby(['house', 'Scheduled
↳ Buildings', 'room'])[['icon_id']].count().sort_values(by='icon_id', ascend
↳ ing=False).sort_index(level=[0],
↳ sort_remaining=False)
64 undergrad_by_house.columns = ['count']
65 undergrad_by_house.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+
↳ +icon_id+'/undergrad_by_house.csv')
66
67 #Undergrad's results of rooms on campus by building by house
68 undergrad_by_bld_by_house = undergrad_by_house.sort_index(level=[0,1],
↳ sort_remaining=False)
69 undergrad_by_bld_by_house.to_csv('/Users/heningzheng/Documents/SURF/Classroo
↳ mData/'+icon_id+'/undergrad_by_bld_by_house.csv')
70
71 #Undergrad's results of buildings on campus by house
72 undergradbld_by_house = undergrad_by_house.groupby(['house', 'Scheduled
↳ Buildings']).sum().sort_values(by='count', ascending=False).sort_index(le
↳ vel=[0],
↳ sort_remaining=False)
73 undergradbld_by_house.to_csv('/Users/heningzheng/Documents/SURF/ClassroomDat
↳ a/'+icon_id+'/undergradbld_by_house.csv')
74
75 #Faculty's results of rooms on campus by academic rank
76 faculty = criterion.loc[(criterion['profile']=='Faculty member'),:]

```

```

77 faculty_overall = faculty.groupby(['sbld_rm_id'])[['icon_id']].count().sort_
   ↪ values(by='icon_id',ascending=False)
78 faculty_overall.columns = ['count']
79 faculty_overall.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+ic
   ↪ on_id+'/faculty_overall.csv')
80
81 faculty_id = faculty.groupby(['bld_rm_id', 'BLD_RM'])[['icon_id']].count().so
   ↪ rt_values(by='icon_id',ascending=False)
82 faculty_id.columns = ['count']
83 faculty_id.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+icon_id
   ↪ +'/faculty_id.csv')
84
85 faculty_by_rank = faculty.groupby(['faculty', 'Scheduled
   ↪ Buildings', 'room'])[['icon_id']].count().sort_values(by='icon_id', ascend
   ↪ ing=False).sort_index(level=[0],
   ↪ sort_remaining=False)
86 faculty_by_rank.columns = ['count']
87 faculty_by_rank.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'+ic
   ↪ on_id+'/faculty_by_rank.csv')
88
89 faculty_by_rank_id = faculty.groupby(['BLD_RM', 'faculty'])[['icon_id']].coun
   ↪ t().sort_values(by='icon_id',ascending=False).sort_index(level=[0],
   ↪ sort_remaining=False)
90 faculty_by_rank_id.columns = ['count']
91 faculty_by_rank_id.reset_index(inplace=True)
92 faculty_by_rank_id = faculty_by_rank_id.pivot(index='BLD_RM',
   ↪ columns='faculty', values='count')
93 faculty_by_rank_id.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'
   ↪ +icon_id+'/faculty_by_rank_id.csv')
94
95 #Faculty's results of rooms on campus by building by academic rank
96 faculty_by_bld_by_rank = faculty_by_rank.sort_index(level=[0,1],
   ↪ sort_remaining=False)
97 faculty_by_bld_by_rank.to_csv('/Users/heningzheng/Documents/SURF/ClassroomDa
   ↪ ta/'+icon_id+'/faculty_by_bld_by_rank.csv')
98
99 #Faculty's results of buildings on campus by academic rank
100 facultybld_by_rank = faculty_by_rank.groupby(['faculty', 'Scheduled Buildings
   ↪ ']).sum().sort_values(by='count',ascending=False).sort_index(level=[0],
   ↪ sort_remaining=False)
101 facultybld_by_rank.to_csv('/Users/heningzheng/Documents/SURF/ClassroomData/'
   ↪ +icon_id+'/facultybld_by_rank.csv')
102
103 check('favorite_place')
104 check('too-hot')
105 check('improvement')
106 check('feels_far_away')
107 check('oversized')

```