Rachel Berman Robert Petteruti Heather Sabel
Big Data Final Project Report
ECON1660 Spring 2017

Big Data Final Project Report

The Hunger Games?

Housing at Brown:
Inferential and Predictive Analysis of Dormitory Demand and Lottery 'Success'

Abstract:

Many Brown students would say that the annual Housing Lottery is one of the most stressful yet distinctly Brown experiences they've had in college. While those lucky enough to land Caswell or 315 Thayer think back on their lottery experiences with gratitude and relief, other slightly less fortunate dwellers in the concrete halls of Grad Center or the remote neighborhood of Perkins Hall reminisce on their trying days fighting the lottery with utter disdain. Such varied experiences among sophomores and seniors alike has sparked our interest in studying this enigmatic process for the future benefit of our peers. What conclusions can we make about lottery 'success'? How can quantifying demand and popularity for specific dorms tell us about common lottery outcomes? Welcome to our inferential and predictive investigation of the housing lottery and the factors that make it a wealth of information to study with our newly-developed big data skills.

Introduction:

Overview and History of the Housing Lottery Process

Brown University allocates housing differently than most other universities. Students that want to live on campus form their own groups and enter into the Housing Lottery - a process through which each group is assigned a random number (will priority given to those that have completed more semesters). When the group's number is called, the group gets complete control over their housing, picking exactly which rooms they'd like to live in, given the remaining available options. Given that the quality, or at least, perceived quality, of housing on campus differs so widely, some housing is much more desirable than other housing, and consequently runs out very early on in the Housing Lottery selection process.

The Housing Lottery used to be held in person in Sayles Hall, with a big projector showing all of the remaining available housing, where buildings would disappear when all of the rooms had been selected. Many alumni have compared this process to the Hunger Games, because it was so brutal and cutthroat to watch the group in front of you take the housing that your group wanted. In 2014, the Housing Lottery transitioned to being held online. In order to limit our results to variation based solely on year/dorm, we only use data through 2013, the last year that the lottery was held in person in Sayles Hall, in case moving the lottery online had an effect on the demand for housing.

The way that lottery numbers are assigned is as follows:
Each person gets randomly assigned a priority number within a block of people of the same semester level that are entering the lottery, with higher semester levels getting preference. I.e. if there are 100 semester level 7 students and 100 semester level 6 students, the semester level 7 students would be randomly assigned a number from 1-100, and the semester level 6 students

would be randomly assigned a number from 101-200. Then, the students that have formed groups get all of their priority numbers averaged together. The average priority numbers are then ranked across everyone in the lottery, and each group is assigned their lottery pick number based on that ordering. Of course, there are some problems with this methodology, mathematically speaking. Because of the averaging, we tend to see bigger groups land toward the middle of the pack, and most people who enter the lottery in singles and doubles wind up either at the very beginning or very end of the list. Because of this effect, the ranking system for sophomores was changed this year, and the ranking system for juniors and seniors will be changed next year, but the number assignment system was consistent from 2006-2013, which are the years for our data.

Sophomores have designated housing, and many seniors live off campus - meaning that effectively disjoint groups pick over the same housing options each year (ie. I pick from Sophomore-only housing when I'm a sophomore, Junior/Senior housing when I'm an upperclassmen, and likely become part of the ⅔ of the senior class that lives off campus (or else get to pick from premium housing that likely would have run out when I picked as a junior, leading me to be able to select from effectively a completely different subset of housing). This led us to one of our initial research questions: does demand for the same housing look the same across years? In other words, is there some demand curve that's intrinsic to the particular dorm, or does each class value different attributes, and therefore different dorms, differently?

The way that Brown Residential Life renovates buildings is rotational, meaning that each building gets renovated or at least retouched once every 5-10 years, typically over summer break. Students pick housing in the spring (before renovations happen but after they are announced). We were curious what effect, if any, building renovations has on demand for the building. If students have perfect information, we would expect to see higher demand for a building in anticipation of renovations. If not, we would expect to see a one year delay between renovations and effect on demand. We may also see no, or very small, effect, as the most important attributes

of buildings may be related to location and layout as opposed to how recently it's been updated. However, to do this analysis, we first need to determine if there is a characteristic demand curve for each building across years.

Objectives and Methodology:

There are three primary objectives of this analysis:

1. Determine the characteristic demand for each dorm or group of dorms on campus and analyze whether or not demand is consistent for a particular building across years.
2. Assess demand through derived popularity metrics and construct predictive model for dorm placement success given several parameters
    a. Other ways to assess demand
    b. Our early predictive model (in order of discussion)
3. Our Dorm-Specific Predictive model

The data we will used comes from the Office of Residential Life at Brown. Each year, they post raw data (for students to peruse to get an idea of what rooms might be available at their given lottery number) from the Housing Lottery in previous years. Each year is in its own excel booklet, and includes group number, what building/room combination was chosen, and the time at which it was chosen. For our purposes, since times are not more informative than numbers, (each group is allotted exactly the same amount of time to choose their room, so there's a perfect correlation between number and pick time) we will use pick number for simplicity. We had to clean the data - there are some anomalies in the picking process (groups can drop down to join with other groups with a lower number if they want to pick a four person suite instead of a double, for example) that we had to exclude, and we combined like dorms such as New Pembroke, Young Orchard, and Grad Center, that appear in the data separated by building, when we did not expect the granularity to be informative (Grad Center A should be no different than

Grad Center B, for example). Methodology specific to each subsequent analysis will be discussed in more detail in subsequent sections.

Analysis and Discussion

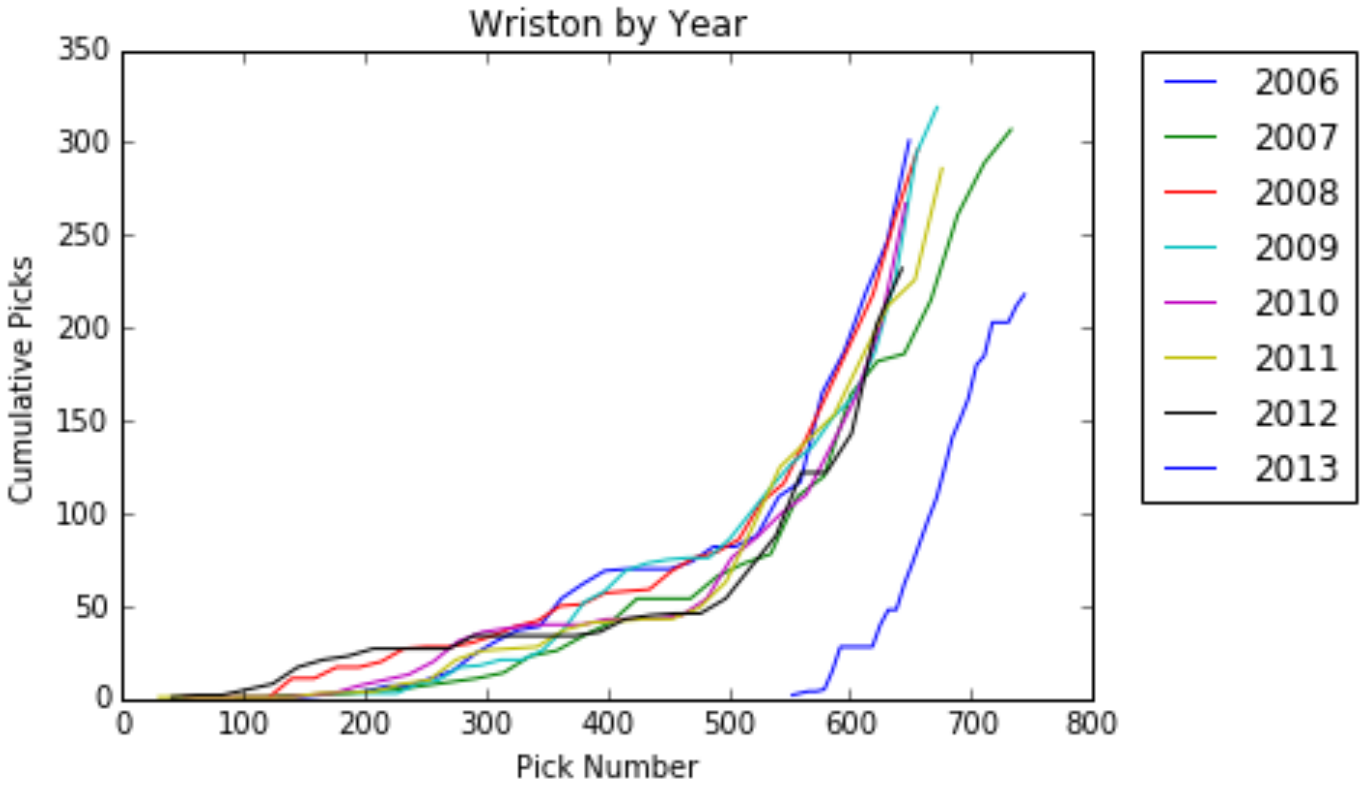
1. Characteristic Demand of Each Dorm/Group of Dorms

Methodology -

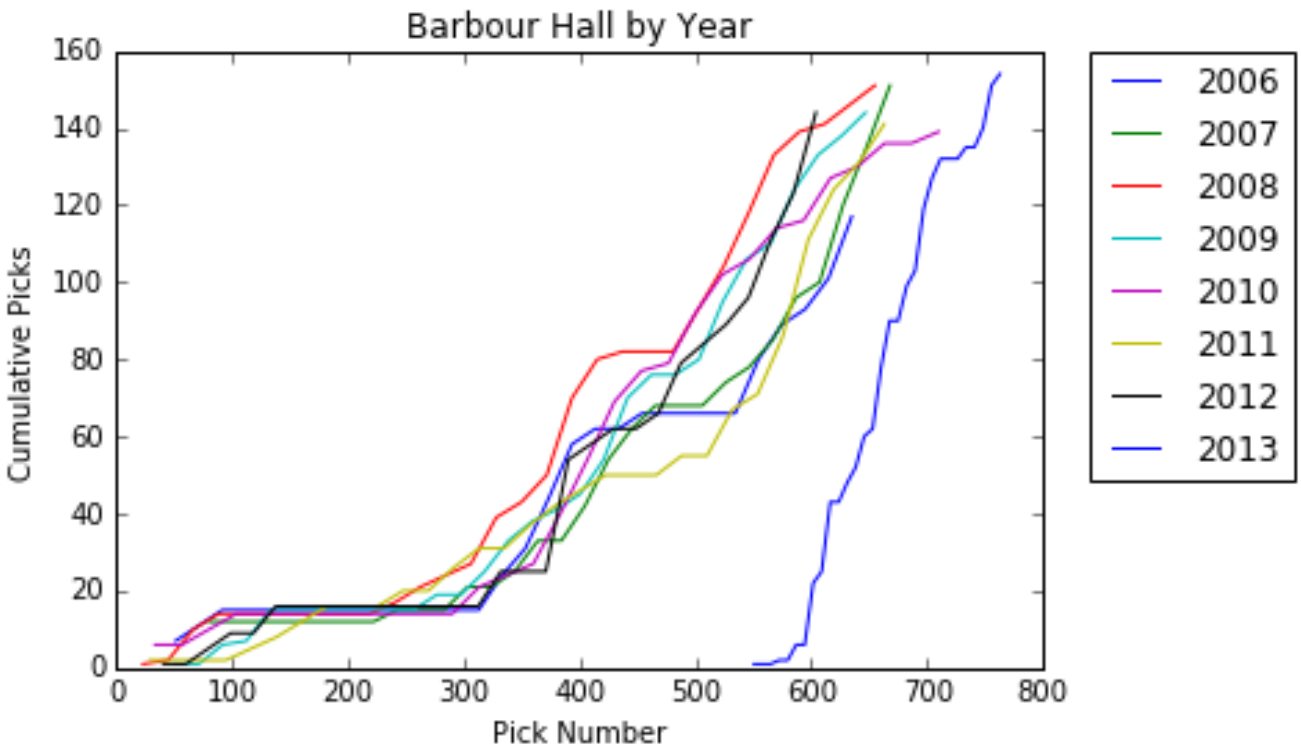
Our goal was to get a sense of the demand curves associated with each building or group of buildings (as well as determining if there was such a demand curve, or if dorm popularity varies across years if different classes of students choose to value different attributes as more or less valuable than in other years). To obtain what we will henceforth refer to as the "demand curve" for each housing option, we plotted what was essentially a CDF of number of rooms chosen in each building, using the pick numbers as a metric of time on the x-axis. In this way, we can see how sharply the graph rises (if lots of rooms are chosen in a particular building very early on in the lottery) or how flat the curve is (if a few rooms are chosen in early rounds, but mostly the rooms get chosen when not many options remain). Once we plot the demand curve across years, it became clear that there was, in fact, a similar pattern of demand across years for a given building.

"Wriston Quad" is a designated common group of dorms that border Wriston Quad and Patriot's Court. They are grouped together because they are all similar in nature - they are typical dorm-style rooms, with a hallway of single or double rooms and a common bathroom on each hallway. These houses are primarily allocated to the Greek and Program houses that take up most of the space in these buildings, but when there aren't enough students in a given house that want to live "in house" with their community, the extra rooms go to "independents", or students who select the open rooms through the housing lottery. Mostly the demand curves for housing on Wriston are consistent from 2006-2012, but we see that 2013 is an outlier from the rest of the

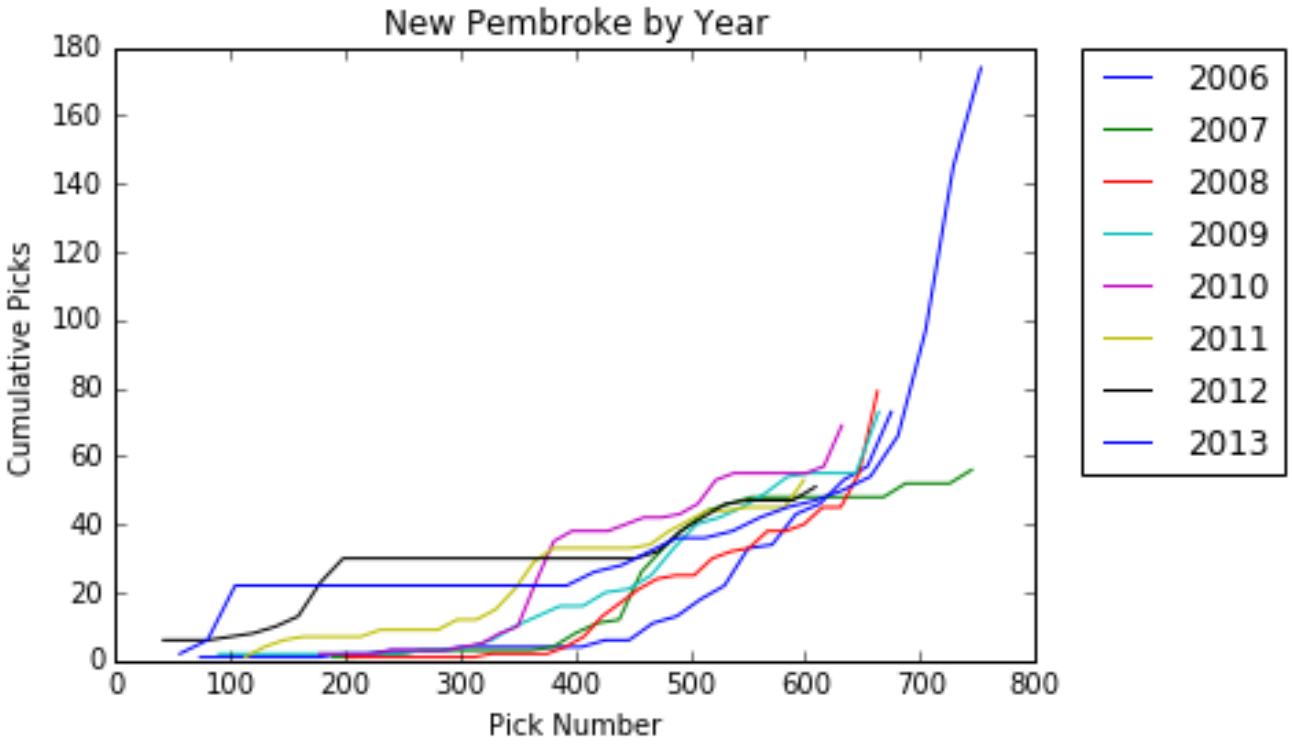
group. As it turns out, there was a major reorganization of dorms effective in 2013, as far as which dorms were designated as only available to sophomores (as well as which dorms were set aside to be allocated to first years - more on that later). In 2013, only sophomores could select independent housing on Wriston Quad, whereas anyone could select housing on Wriston Quad in previous years. Since sophomores have much higher lottery numbers, we don't see the first pick in 2013 until after 500 (the first sophomore pick in 2013 was 536).



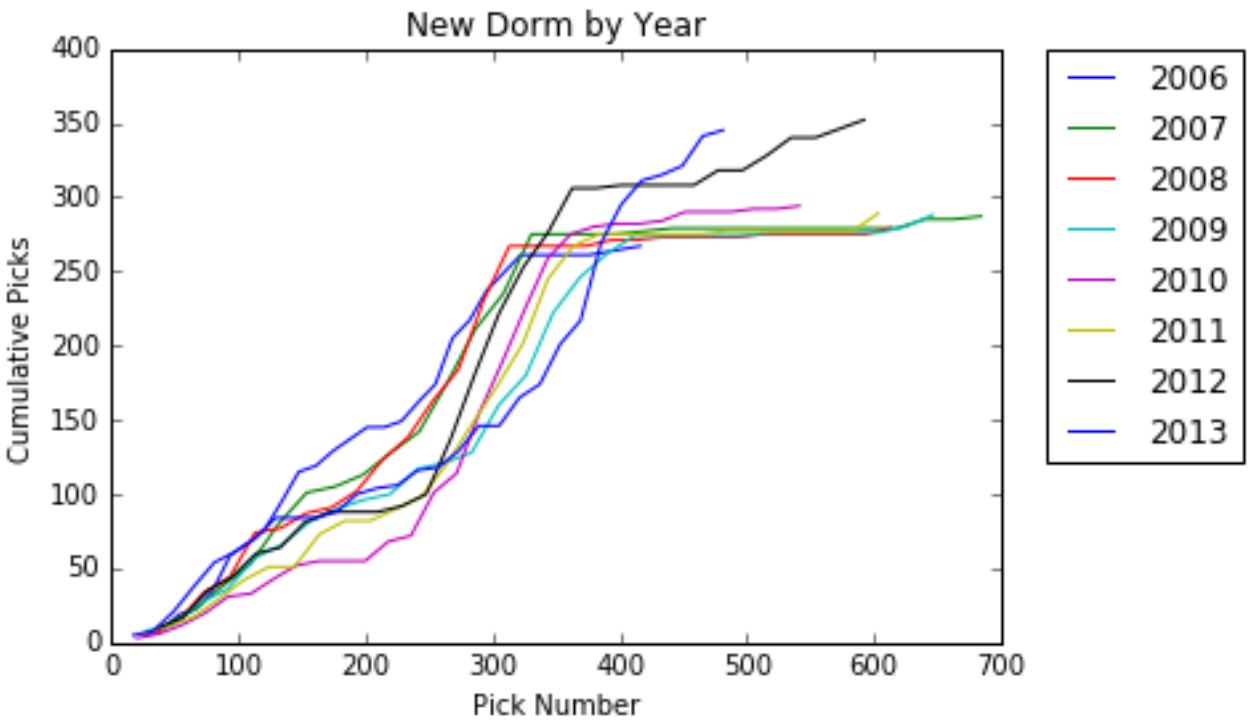
Next, we examined demand for Barbour Hall, which, unlike housing on Wriston Quad, is laid out in mostly suites with private bathrooms that house either 3 or 4 people. We see a similar clump of demand curves from 2006-2012, with a similar outlier for 2013 (the same thing happened to Barbour Hall in 2013 as Wriston Quad - it was designated as Sophomore only).



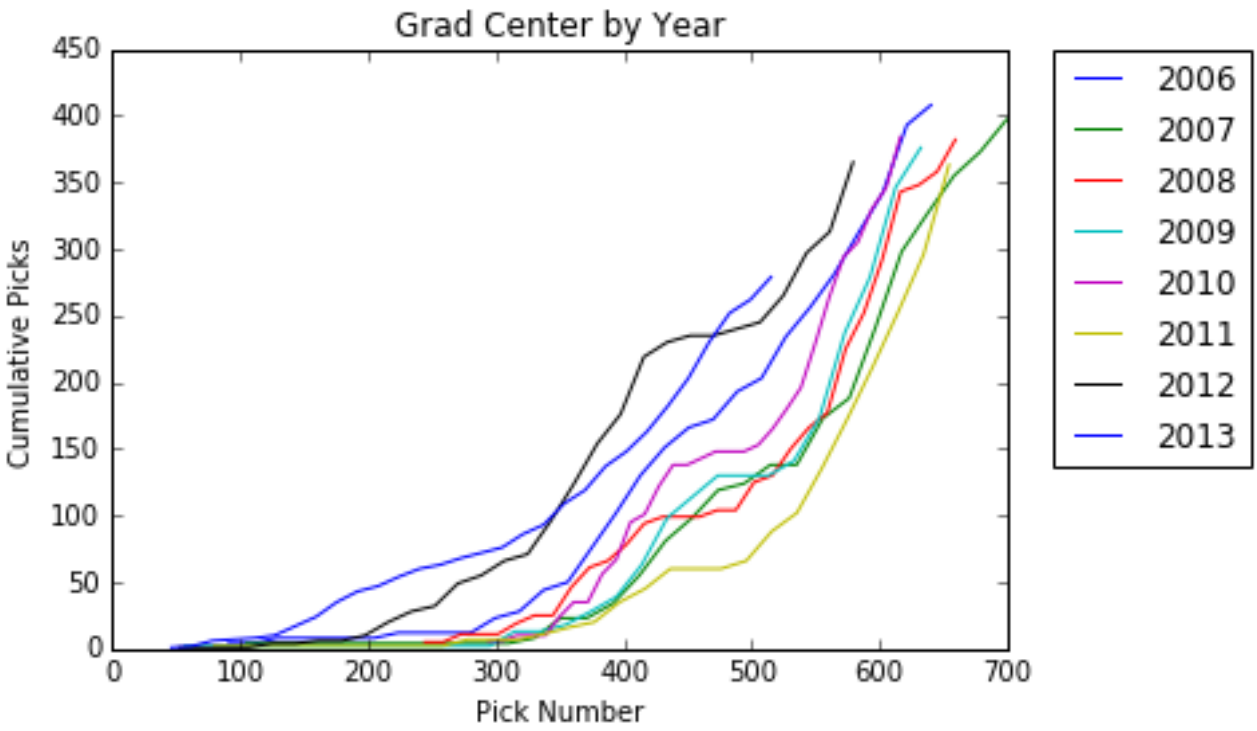
New Pembroke is a group of three buildings up on the north side of campus. It's mostly singles and doubles (with hall bathrooms, in a similar style as Wriston Quad). Some students think of it as being very far away from everything happening in the main area of campus, but some students (particularly athletes) value the location because it's the closest housing to the athletic complex. The one curve that's much longer than the others is a result of another New Pembroke building being added to the Housing Lottery options in 2013. Additionally, we can see the impact of renovations to New Pembroke - 2012 and 2013 curves for the group show more demand earlier on in the lottery than in previous years. It is interesting to note that renovations of New Pembroke lounges and hallways occurred in the summer of 2010. We see a small lift in the 2011 demand curve and a larger lift in 2012-2013, but no change in 2010 than previous years, which indicates that either students didn't know there would be renovations, or they didn't know to what extent the renovations would improve New Pembroke, and as such did not adjust their valuation.



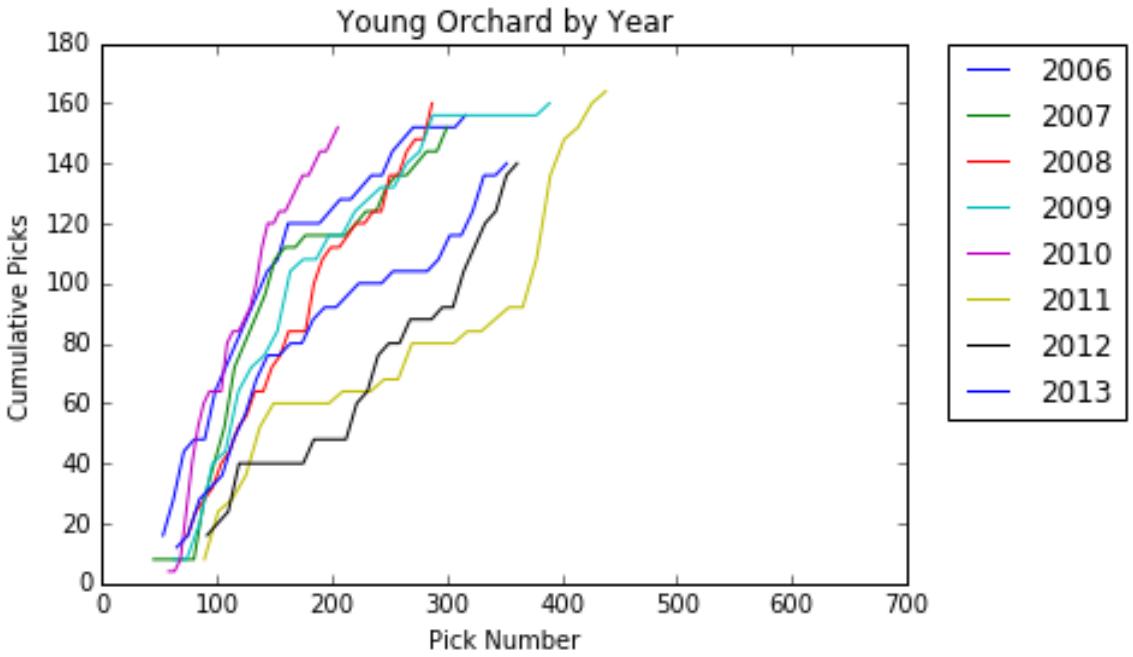
New Dorm is one of the most desirable housing options at Brown. Its location is proximate to campus, and it's laid out in four person suites, each of which has four single rooms, with a large common space. Unlike Barbour, it has shared bathrooms for each hallway, but it's much newer and closer to campus, which students must value because it has much steeper demand than other dorms, including Barbour. It was renovated during this time period (the same time as New Pembroke - the summer of 2010) but there are no obvious improvements in popularity, possibly because it was taken so quickly to start with, it wouldn't really have been possible for the rooms to be picked much earlier.



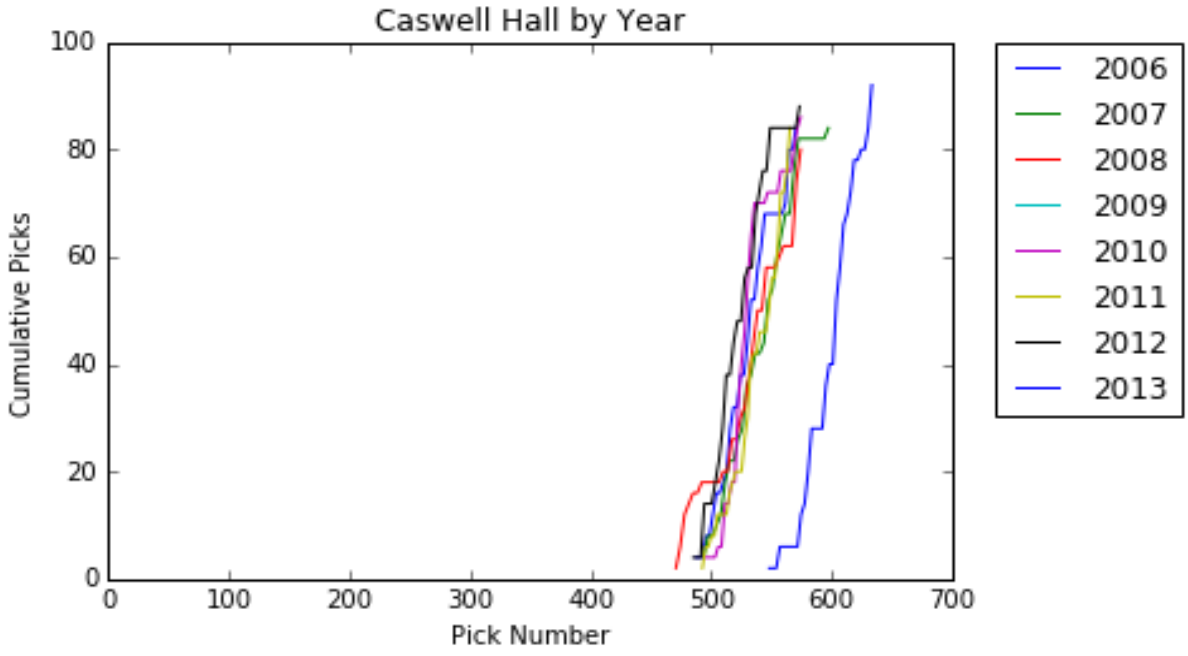
Grad Center is one of Brown's oldest housing structures. It's comprised of four towers, each of which has a couple of "suites" per floor. The "suites" are not the same as in Barbour or New Dorm, though - they're five single rooms that share a private bathroom and a hallway behind a locked door. As a result, they are not assessed the "suite fee" that suites in Barbour, New Dorm, Young Orchard, and some others have, meaning for groups of students that want to live together, they're a less expensive option. We see a similarly shaped but wider band of demand curves for Grad Center across years - possibly because of how differently students value having a nicer, complete suite (as in New Dorm, Barbour, etc) against having to pay extra to live in those spaces. From personal experience, some students we know have gone into the lottery wanting to avoid Grad Center at all costs, and some consider it to be a decent location and a way to live in a group with friends without having to pay the suite fee.



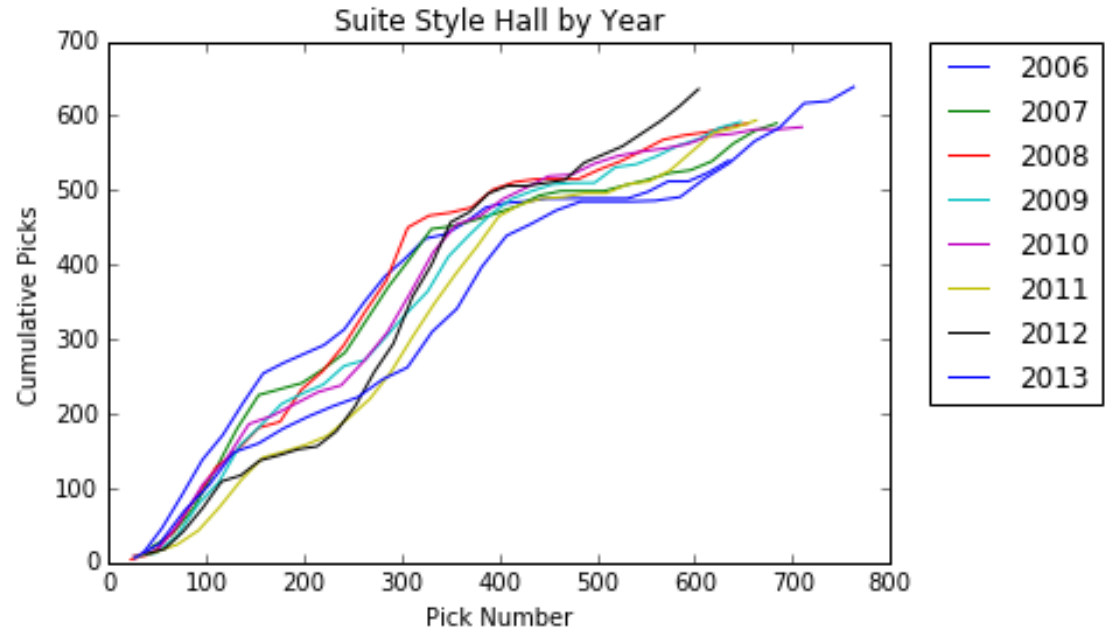
Young Orchard is made up of three buildings, each of which has suites that have a private common area, kitchen, and bathroom. For those that want to live off meal plan and cook their own food, it's an ideal place to live; however, some students don't like how far away it is from everything else on campus. As a result, we see a similar pattern as we saw for Grad Center - similarly shaped demand curves, that are steep early on in the lottery, but form a wider band than we see for dorms like Barbour. Young Orchard also runs out very early on in the lottery - it's clearly very desirable housing.

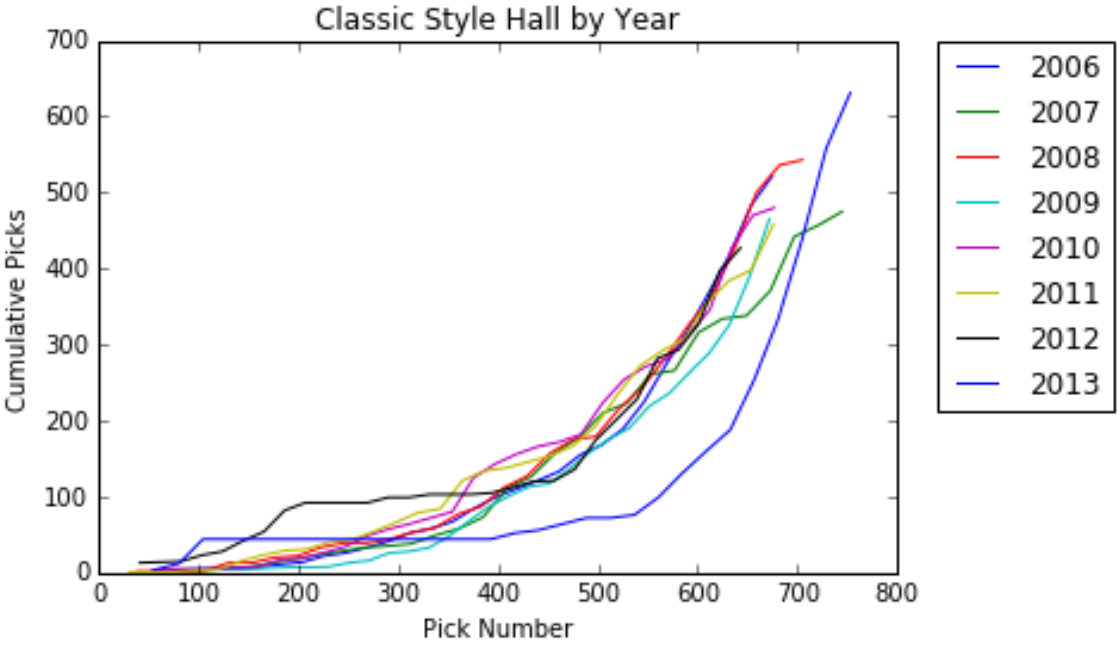


Caswell Hall has always (for the years 2006-2013 for which we have data) been designated sophomore-only, but it has an extremely steep demand curve. It's arranged in classic dorm style with doubles and a shared hallway bathroom. Its biggest appeal is its proximity - it's one of the most centrally located dorms on campus.



Finally, we were curious whether or not there was a different demand curve for some styles of housing versus others. We grouped together all of the suite-style housing into one bucket and all of the classic-style housing (singles or doubles on a hallway with a shared bathroom). It appears that there is, in fact, different demand curve shapes for the different housing styles - the suites seem to be taken much more quickly, whereas the classic dorms seem to have much flatter initial demand. This could have to do, in part, with the way that bigger groups are more concentrated towards the middle of the lottery - maybe suites of certain sizes or configurations run out at some point, and demand for singles and doubles picks up as students have fewer suite options but still want to live near each other.





No matter which dorm you look at, or even if you separate the dorms into suite-style versus classic style, there is definitely a difference in demand for each dorm/style of housing that is consistent across years. We see very similar patterns of demand across years for the same dorm, and different dorms have very different patterns of demand. Because housing at Brown is so different, and each dorm has a very different combination of features, it's difficult to determine exactly which features Brown students are prioritizing, especially since different

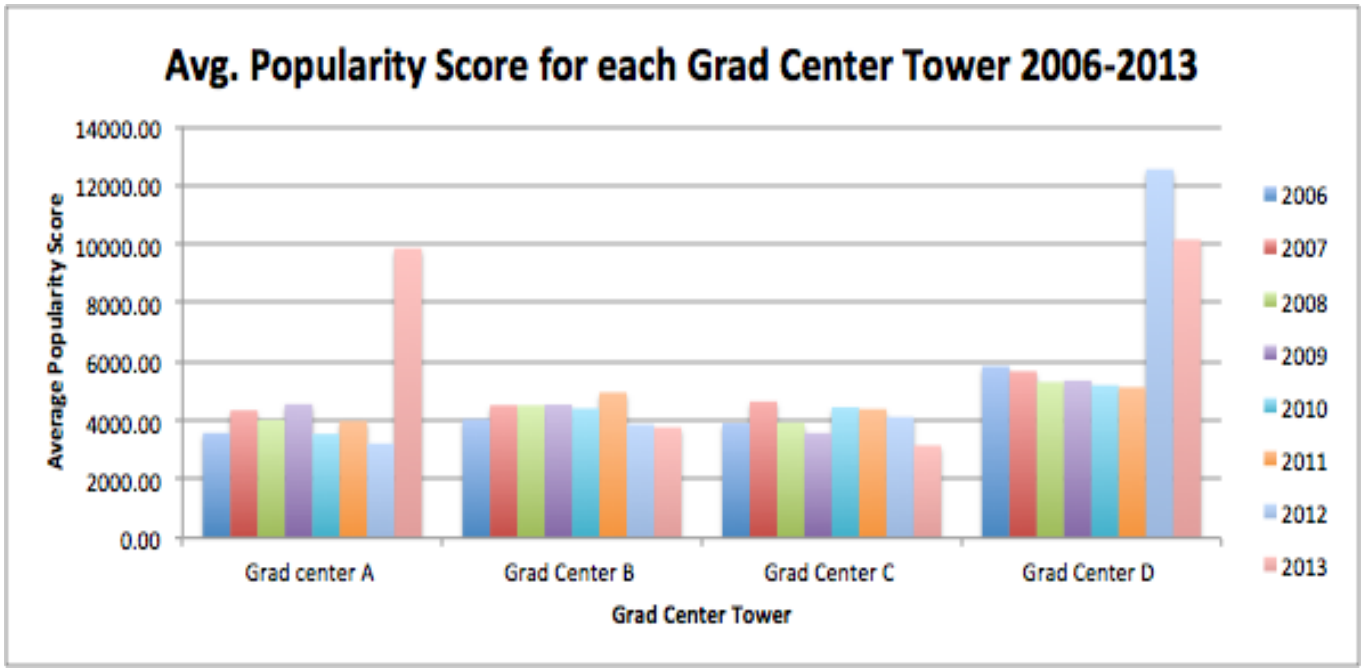
students will prioritize different features (like location, recency of renovation, layout, common space, etc). The fact that there is such consistency across years would also lead us to believe that there is some element of hearsay and reputation that goes into housing selection, too - students hear from their friends which dorms are reputed to be the best, and that reputation lives on across years.

2. Inference II

a) Popularity scoring and trends

Our analysis so far examines the way demand for each dorm fluctuates from 2006-2013 based on observations related to the presence of each dorm in each year's housing lottery list. But the second part of our analysis dives into a customized set of metrics used to describe the popularity and demand of different dorms for each year. We do this by defining what we call a **popularity score**, derived programmatically by finding all instances that a particular dorm appears in the list of dorms and for each year finding the average lottery number for each dorm. This gives us a decent measurement of the place in a particular year's lottery a particular dorm was requested by an average housing group. The way to interpret popularity scores are such that the lower the scores, the more preferred or "popular" the dorm was, since it was picked earlier in the lottery relative to other dorms. So therefore, popularity scores are technically inversely proportional to actual dorm popularity, as the higher the score, the less popular the dorm is. We also use lottery numbers for a particular dorm within a single year to observe the demand of a particular dorm at different points in a given lottery. For example, it just so happens that Grad Center's lottery numbers plateaued at the end of the lotteries in several years because it is the one of the least sought-after dorms and therefore gets chosen by many groups with less fortunate numbers toward the end of the lottery where these groups choose housing. Furthermore, we later compare popularity scores across different years which gives us a good sense of how popularities for different dorms changes during the 8 years we have data. Before we go into some important assumptions we made in this analysis, here is a graphical representation of popularity scores for

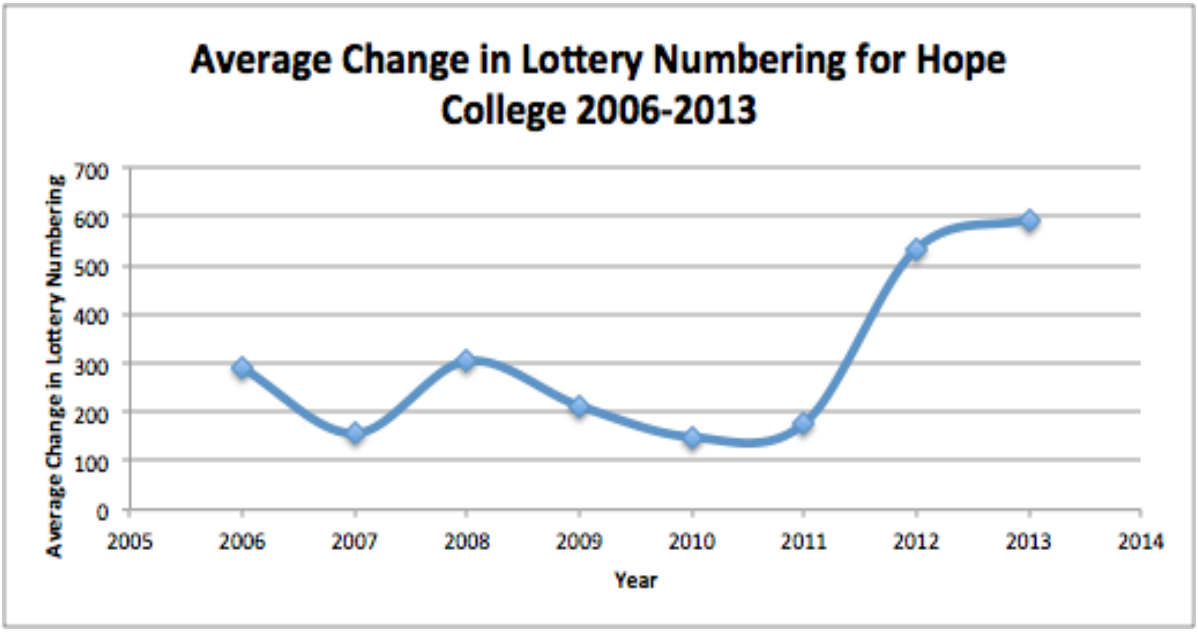
the four towers in grad center during the 8 years we have data. Visualizations like these are useful in determining and observing trends in popularity for different dorms.



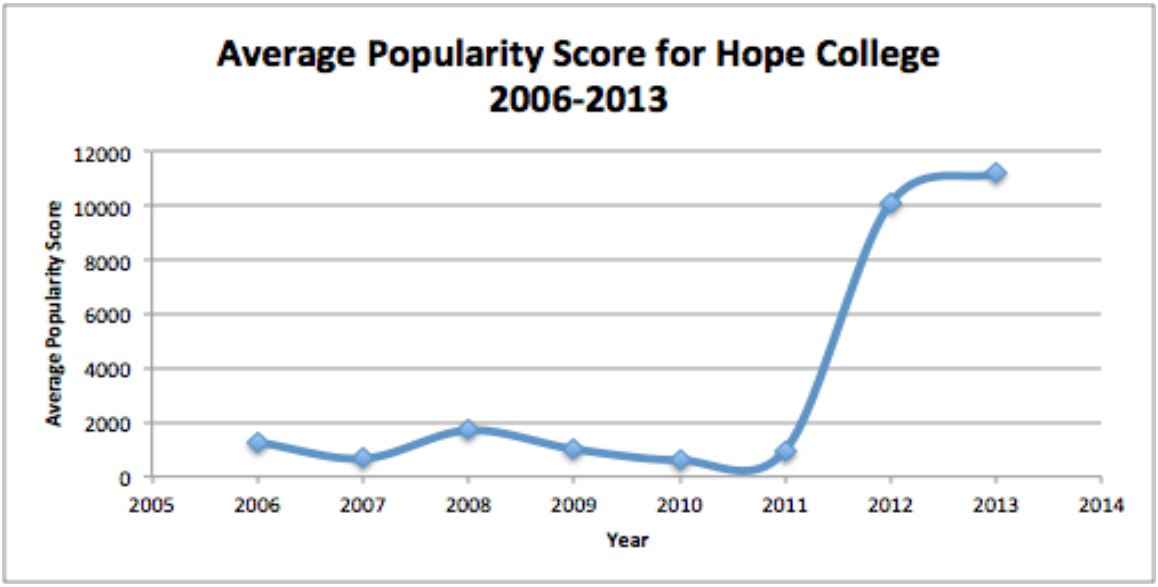
Views like this also give us indications of various developments that occurred on Brown's campus that could have affected the popularity of grad center dorms. For example, we know from our background research that most of the suites in graduate center tower D are a lot smaller than those in other grad center towers, which explains the very high popularity score for grad center tower D relative to other towers.

Popularity score trends are a very important metric for us as they assess the rate of change of popularity scores across the years from which we have housing lottery data. For example, we can take into account the changing popularity of Hope College from 2006 to 2013 and observe its change in popularity among students after its grand renovations that occurred in 2012 when the university administration embarked on an archeological dig outside the building to find traces of the university's past (involvement in the slave trade and other artifacts from Brown's founding). Changes in its scoring can then be observed with this historical knowledge

in mind and compared to changes in lottery numbering for a particular dorm across a series of years like below:



Before we interpret this chart, it is important to note how average change in lottery numbering can be interpreted itself. For each year 2006-2013, the lottery numbers that correspond to student groups selecting Hope College are listed. The differences in each number from the one before is computed and all the differences in numbers for each year are averaged out and called the average change in lottery number for a particular year. The interest part, however, is comparing the changes in lottery numbering in the chart above to changes in the popularity scoring for a particular dorm, in this case Hope College, shown below:



For the dorms in our datasets, we expect this chart to tell a similar story to the one above. For example, the chart showing the average rate of change of lottery numbers can tell us a few things about Hope College's demand at different points in the lottery for certain years. A huge increase in average difference in lottery numbering from 2011 to 2012, for example, could mean that in 2012, Hope College may have been chosen at more varied points in the 2012, perhaps by groups in the beginning of the lottery and then again by other groups in the end of the lottery. A small or negligible increase in average differences in lottery numbering between 2010 and 2011, for example, could mean that Hope College was selected at roughly the same places in the 2010 lottery as it was in the 2011 lottery. The average popularity score for Hope College gives more of an absolute picture of the changes in dorm popularity from 2006 to 2013, but could still add to the story told by the first chart. For example, looking at the decrease in dorm popularity from 2011 and 2012 (signified by a drastic increase in the popularity scoring) tells us that Hope College was chosen later in the lottery in 2012, a parallel conclusion to the one made from the first chart showing average changes in lottery numberings. Likewise, looking at the very small different in popularity scoring from 2012 to 2013 indicates that Hope College neither increased or decreased in popularity during this time at all but merely was sought after by lottery

participants in much the same way as 2012. Disruptions to the quality of the dorm, such as Brown's archeological dig outside the building, may have decreased the dorm's popularity among students in 2012 and lasted until 2013.

b) A predictive model of dorm 'success'

An important part of our exploration in this project was constructing a statistical model based on our historical data that can be used by students in future runs of the lottery to predict specific dorm placement given several parameters. Data-points of interest to us in this part of our analysis were the occupancy of the housing group (the number of people in each group) as well as the name of the particular dormitory that the housing group desires to live in. We then make the critical assumption that lottery number assignments happen on a bell curve with a normal distribution to allow us to utilize the our known equations of standard deviation and z-scoring in our analysis to predict success with 95% confidence intervals.

i) Dealing with occupancy data:

The first step in this part of our analysis was to look at all past instances of a particular dorm from 2006 to 2013 and for all of the years in the dataset compute the average occupancy (size of the group) that requests this dorm in the lottery. This average is used together with the standard deviation of the occupancies for a particular dorm to arrive at what we call an occupancy probability. The standard deviation for the occupancy is calculated using the formula for standard deviation for a normal distribution:

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}}$$

Doing this programmatically on python gives the standard deviation for occupancy for each dorm. We then use this and average occupancy to arrive at a probability of matching with that particular dorm using a 95% confidence interval.

ii) Using 95% confidence intervals as a basis for assigning **occupancy probabilities:**

We know from our knowledge of normal distribution that 1.96 standard deviations to the left and right of the middle of the distribution covers exactly 95% of the distribution. We then looked at the difference between the requested group size and the average occupancy computer earlier. Using the deviation from the mean that the requested group size is, we can then assign a probability based on how many standard deviations the deviation is. However, some changes have to be made to make the final probability as accurate as possible and the best as possible reflection of the data that we have. One of the crucial last steps the model computes is to determine the average size of a lottery given all data sets incorporating some of the important assumptions we must make throughout this process that we will describe in more detail shortly. By dividing the average number computer earlier with the average size of a lottery in any given year, we arrive at a typical place in the lottery that the particular house and group size show up in. Multiplying this by our occupancy probability from earlier, we arrive at the final probability of placement in a particular dorm with a given group size.

For example, if I were to request 315 Thayer for a group size of 3, the probability that I get placed in this frat house is approximately 34%. There are a few very important assumptions that we had to incorporate while we came up with this early predictive model. The most important assumption was that each housing group only occupies one housing option in the lottery as opposed to a few. The reason why this assumption is important to be made is because it is typical for a group of 5 people to occupy a housing selection of occupancy 3 and then another double. The predictive model that we developed did not take these biases into account (partly because we did not have the data to tell us which groups started out as one larger group) so we have to assume away with that real-life occurrence. The second important assumption we made throughout this part of our analysis was that there were a few dorms that did not appear in the housing lottery after a certain year, such as Bronson house, which was included as one of the freshman housing units and therefore taken off the lottery for sophomores through rising seniors. To deal with such discontinuities, the algorithm in this early predictive model employed a series

of checks to make sure it was averaging out 8 sets of data-points for each dorm (1 for every year we have data) to obtain as accurate results as possible that closely indicate the future.

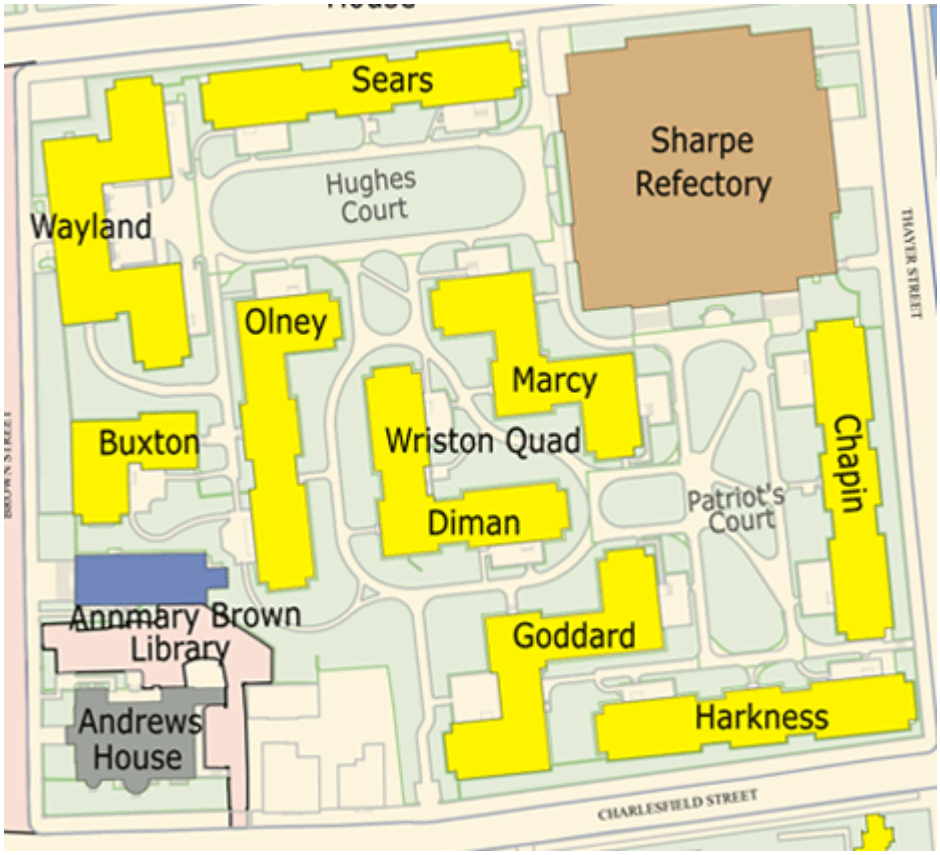
3. Our Dorm-Specific Predictive model

Greek Neighbors

At Brown University, there are fraternities, sororities, program houses and literary societies. With the exception of King House and those that do not have a residential space, they are all within Brown University dorms on Wriston Quad. Typically, all dorms on Wriston Quad contain two housing organizations (fraternities, sororities, program houses and literary societies), one in each side of the building, and a "buffer zone" of independents (students not associated with either housing organization) in between. These "buffer zones" are filled through the housing lottery. In this section, I intend to analyze housing lottery data to try and identify whether people are more willing to live next to some housing organizations than others, and, if so, how desireable as a neighbor each housing organization is.

I begin with the assumption that students don't particularly care about small differences in location and so the difference in dorm preference among dorms on wriston is due to what organizations share the dorm. This is a reasonable assumption since the building plans of most wriston dorms are identical except for the building shape (room size and structure is consistent throughout Wriston) and if brown students cared about small differences in location, we would expect to see that there was also a difference in preferences among Grad Center A,B and C which there isn't (Analysis detailed in another section showed no difference in preferences among grad center buildings A,B, and C which are identical in all ways but location. (Grad Center D does differ in popularity due to it having smaller suites)

Map of Wriston Quad



As a control, I've included Wayland House, which is on Wriston but does not contain any housing organizations. The OLS regression (which regressed the average number a room was chosen in the lottery on the occupancy of the room, it's room number and dummy variables for all of the dorms in Wriston Quad (except Buxton, the international house, which was never part

of the lottery) found coefficients for each dorm. The results are as follows:

Source	SS	df	MS		
				Number of obs	= 909
				F(9, 899)	= 619.98
Model	38120126.4	9	4235569.6	Prob > F	= 0.0000
Residual	6141815.22	899	6831.83005	R-squared	= 0.8612
				Adj R-squared	= 0.8599
Total	44261941.6	908	48746.6317	Root MSE	= 82.655

avgNum	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
occupancy	5.605757	2.009712	2.79	0.005	1.661485	9.55003
ChapinHouse	55.21367	14.7137	3.75	0.000	26.33647	84.09087
DimanHouse	24.15665	13.93498	1.73	0.083	-3.192236	51.50553
GoddardHouse	54.43434	12.49249	4.36	0.000	29.91649	78.95218
HarknessHouse	33.61135	14.56672	2.31	0.021	5.022605	62.20009
MarcyHouse	61.17499	15.01062	4.08	0.000	31.71505	90.63494
OlneyHouse	31.21261	14.91344	2.09	0.037	1.943398	60.48182
SearsHouse	45.94959	13.45995	3.41	0.001	19.53301	72.36616
WaylandHouse	0	(omitted)				
number	1.132885	.0203982	55.54	0.000	1.092851	1.172919
_cons	23.42821	10.67802	2.19	0.028	2.471468	44.38496

The limitations of this approach are that you can't differentiate between two housing organizations in the same building. For example, you don't know if Marcy House's desirability is due to the presence of Zete or AEPi, and unless one of them moves, it will be impossible to distinguish. So I've calculated desirability by pair.

Wayland, the control with no housing organizations in it, was by far the most desirable. Among the others, Diman is probably the most desirable. It's the only house whose coefficient lacks definitive statistical significance. It was home to Interfaith House and the sorority ΚΑΘ (commonly known as Theta). Next, and the top with statistical significance, is Olney, with a coefficient of 31.2. During the sample period, contained Delta Tau and Sigma Chi, two fraternities. It is followed by Harkness House, which contained two program houses, Technology House and the now defunct Art House. Sears House is the fifth most desirable dorm on Wriston according to the regression. It contained Alpha Chi Omega, a sorority, and Phi Kappa Psi, a fraternity. Next is Goddard House, which housed Alpha Delta Phi, a co-ed literary society, and

Delta Phi, a fraternity. Close after Goddard is Chapin, which contains the Theta Delta Chi fraternity and Harambee House, a program house. Dead last is Marcy House, which is home to the fraternity Beta Rho Phi (then known as Alpha Epsilon Pi) and Zeta Delta Xi, a co-ed fraternity.

While I was expecting to find that people were less willing to live near organizations that I knew would be forcibly disbanded or suspended by the university after the span of the data set (Delta Phi and Phi Kappa Psi) this was not the case. The statistics show that Brown students clearly and strongly prefer some Wriston dorms over others. While some of this preference may be due to factors such as location, a significant portion of it seems to be due to the different housing organizations occupying the dorms. This is supported by the fact that Wayland, which contains no housing organizations, is by far the most popular. That the most and least popular dorms, Diman and Marcy, (among those with housing organizations) are right next to one appears to further discredit the idea that location within Wriston is a significant influence. However, as always, there may be unobserved variables and factors that have not been identified or considered.

Prediction of desirability of rooms not seen in the data set

After adding available metadata on the rooms, we tried predicting the desirability of rooms using different types of regressions and a random forest. I first created a dataset I that included as variables the order chosen of all rooms close to it, timestamps of times chosen in each year, and all other things that could conceivably be used as a variable and I could find to put in the database. I then used a tree algorithm but instead of finding the most relevant variables it found the least and deleted them from the data set to narrow down what variables were the most important and what weren't to avoid overfitting with irrelevant information. With this reduced dataset, I ran a random forest and the three types of regressions on it and compared the accuracy of the different models. We found that the most accurate model was the Lasso Regression. But I

was still underwhelmed by the accuracy so, using the results of the models to figure out which variables were the most important predictors, I created my own predictive model I've named the adjusted nearest neighbor. This model predicts desirability to be the desirability score of the rooms closest to it that had the same occupancy, adjusting for how the rooms directly above and below it vary from the rooms closest to them. Specifically, the prediction for desirability of a target room is equal to average desirability of rooms next to it (restricting to rooms it has adequate data on and have the same occupancy) + difference in desirability between averaged rooms and target room on the floors above and below it. This makes sense as a formula since, in most dorms, rooms are generally similar to others on the same floor, but have some variation in size and shape. The size and shape of room is generally consistent between in rooms with the same number on different floors. So this prediction method says the desirability of a room should be similar to others like it on the same floor, adjusting for size and shape by using the rooms above and below the room as proxies. It outperforms all other models used to created to predict it.

Conclusion

The first part of our analysis focused on determining whether demand for certain dorms was consistent across years and specific to each dorm. We found that yes, across years, the same dorm or group of dorms had a very similar pattern of demand. Additionally, we found that there was a very distinct demand pattern for each dorm - no two demand curves were the same. Not only were there differences between particular dorms, there were differences between housing types: suite-style dorms had a very different demand pattern than classic-style dorms.

The second part of our analysis first examined the trends associated with popularity scores for each dorm across the different years we have data (2006-2013). By observing changes in these scoring metrics and using background knowledge we have on chronological developments and renovations on the Brown campus, we can make educated attempts at

explaining student behavior in the housing lottery to different dorms given these outside conditions. Our predictive model then utilizes data on group sizes and computes probabilities of successfully getting placed in a particular dorm with a given occupancy. The result is a useful predictive model that can be used by future participants in the housing lottery (of course assuming housing options stay roughly the same as they were from 2006-2013.

In analyzing with regressions housing pattern choices in Wriston Quad, we looked into our hypothesis that what student's dorm choices are affected by what housing organizations occupy that space, and used the housing lottery results to determine how people felt about different housing organizations on campus.

We also created a new model using information gleaned from the results of various models and strategies taught in class to build a predictive model to determine the desirability of rooms that we had no previous data for that outperformed all methods used in class.