THE HOUSING GAMES

A CS1951A FINAL PROJECT BY

MATT COOPER, MICHAEL GILLETT, ALI PAUL AND SAM DUNIETZ

Introduction

Brown's housing lottery is a notoriously messy, some might even say traumatic, process. In 2012, Blog Daily Herald went as far as to call the housing lottery The Housing Games, making light of the stressful process by equating it with the fictional Hunger Games. Our goal was to transform this process using data. Using data collected during the 2006-2013 housing lotteries, we created a variety of tools that enable students to make more informed decisions about their future housing.

Data

We used the housing lottery data collected from 2006-2013 lottery results, which can be found here: http://www.brown.edu/Student_Services/Residential_Council/lottery/results/. These data include all lottery results from 2006-2015, excluding those from 2014. The dataset from 2015 is formatted differently than those collected from 2006-2013 and was extremely difficult to integrate with the other years. Therefore, we analyzed the data collected over the longest consecutive period of time (2006-2013).

Results

To help students navigate the lottery in future years, we created multiple tools to visualize and analyze past housing data. These include:

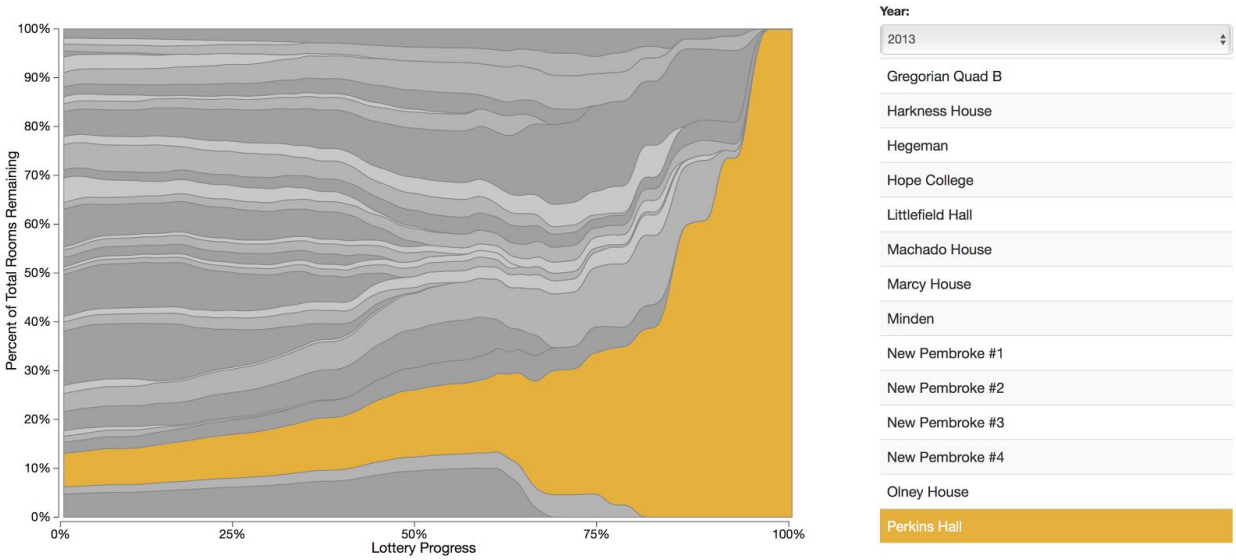
Percent Total Rooms by Dormitory

We created a stacked area chart to visualize what percentage of the total rooms belonged to each dorm at a given time slice of the lottery. This allows students to compare availability of every dorm for a given year at the same time, and enables students to understand which residence halls will likely remain at a given point in the lottery.

In the chart, each dorm's height is the percentage of remaining rooms belonging to that dorm. The popular dorms quickly lose percentage until they disappear and generally before the 50% progress mark of the lottery. On the other hand, unpopular dorms tend to grow in

size as the lottery progresses because a large proportion of the remaining rooms belong to these dorms. On the right side of the visualization is a table of all available dorm rooms for the given year. Clicking on one of the rows will highlight that dorm's section of the graph.

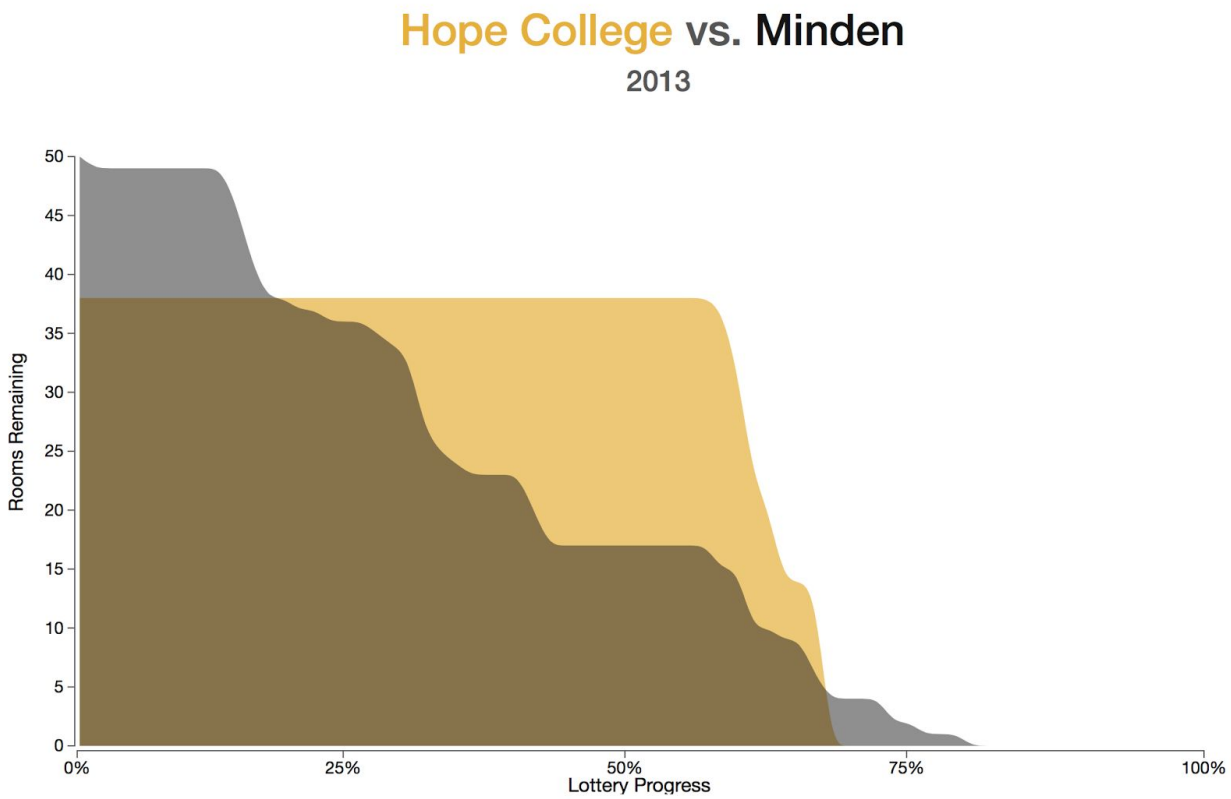
In the visualization below, Perkins Hall is highlighted in the graph. As we all know, Perkins is the least popular dorm on campus and the visualization supports this claim up. The percentage of remaining rooms that belong to Perkins explodes at the end of the lottery as students frantically claim rooms in other dorms.



Number Rooms Remaining Comparison

Often, students want to see how quickly rooms are taken in one dorm vs. another. To aid this, we made a visualization that allows students to select two dorms and compare the number of rooms remaining over time for each. This allows students to quickly see the relative popularity of two dorms over time. It also helps to visualize how many rooms each dorm has, relative to the others.

Results were generally in line with our expectations. For example, each year, Perkins Hall has many rooms remaining very late in the lottery, and is one of the few dorms with rooms remaining even at the very end. We believe this is because groups who have no choice but Perkins decide to enter summer assignment, as they think they have nothing to lose. On the other hand, more preferable dorms, like 315 Thayer, go relatively quickly.



Room Ranking Lookup

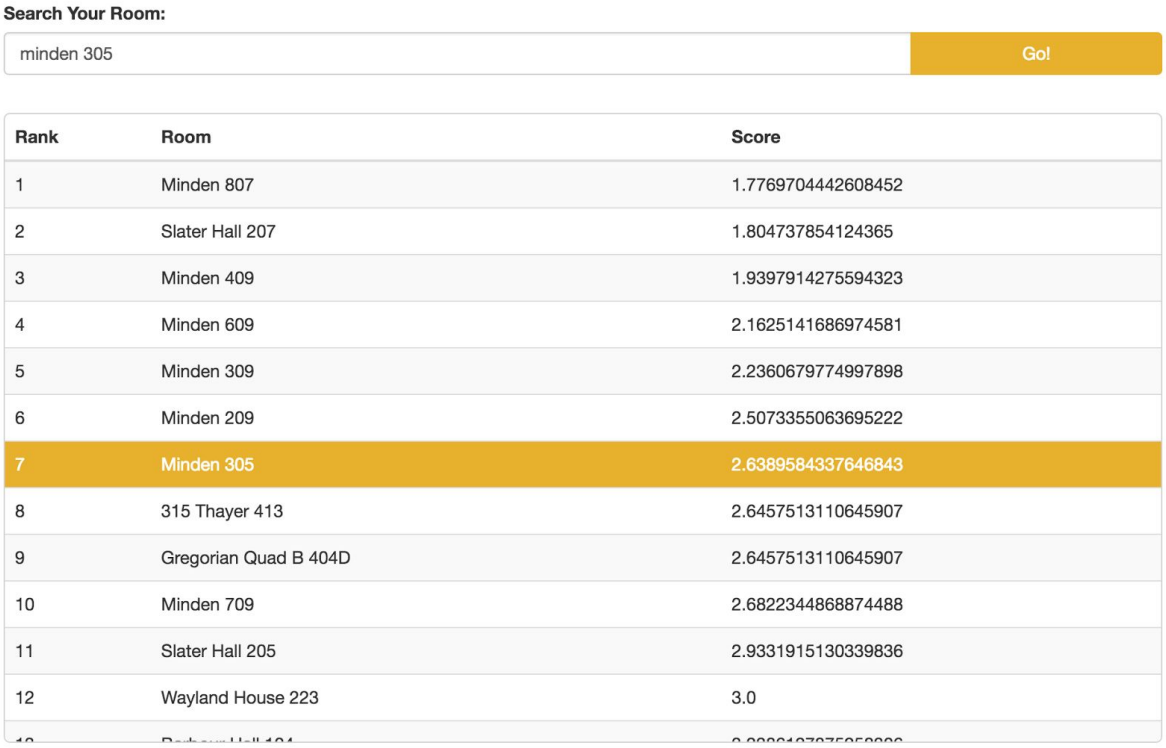
In an attempt to give students an easy way to see which rooms are generally regarded as the best, (and also to enable them to brag to each other about whose room is objectively better,) we ranked each dorm room based on a weighted average of when it was picked in past years' lotteries.

Every room was assigned a score based on the average "pick number" (i.e. a room had a pick number of 3 if the group that picked third in a certain year selected that room.) A smaller score implies greater desirability. We weighted each pick number on a square root scale to reflect that a small discrepancy in earlier picks (e.g. 1st to 10th pick) likely matters more than the same difference in later picks (e.g. 1712th to 1722nd).

The results were as we anticipated. Most of the top rooms are singles in central dorms such as Minden. All of the lowest picks were doubles in Perkins, the residence hall that is farthest from central campus. You can see an example of the results in the table below.

Rank	Score	Room
1	1.776970444	Minden 807
2	1.804737854	Slater Hall 207
3	1.939791428	Minden 409
4	2.162514169	Minden 609
5	2.236067977	Minden 309
...		
1718	27.94637722	Perkins Hall 338
1719	27.96426291	Perkins Hall 343
1720	28.03569154	Perkins Hall 325
1721	28.03569154	Perkins Hall 329
1722	28.0713377	Perkins Hall 220

Using these rankings, we created an interface that allows students to look up specific dorm rooms and discover their rankings. This interface is featured in the image below.

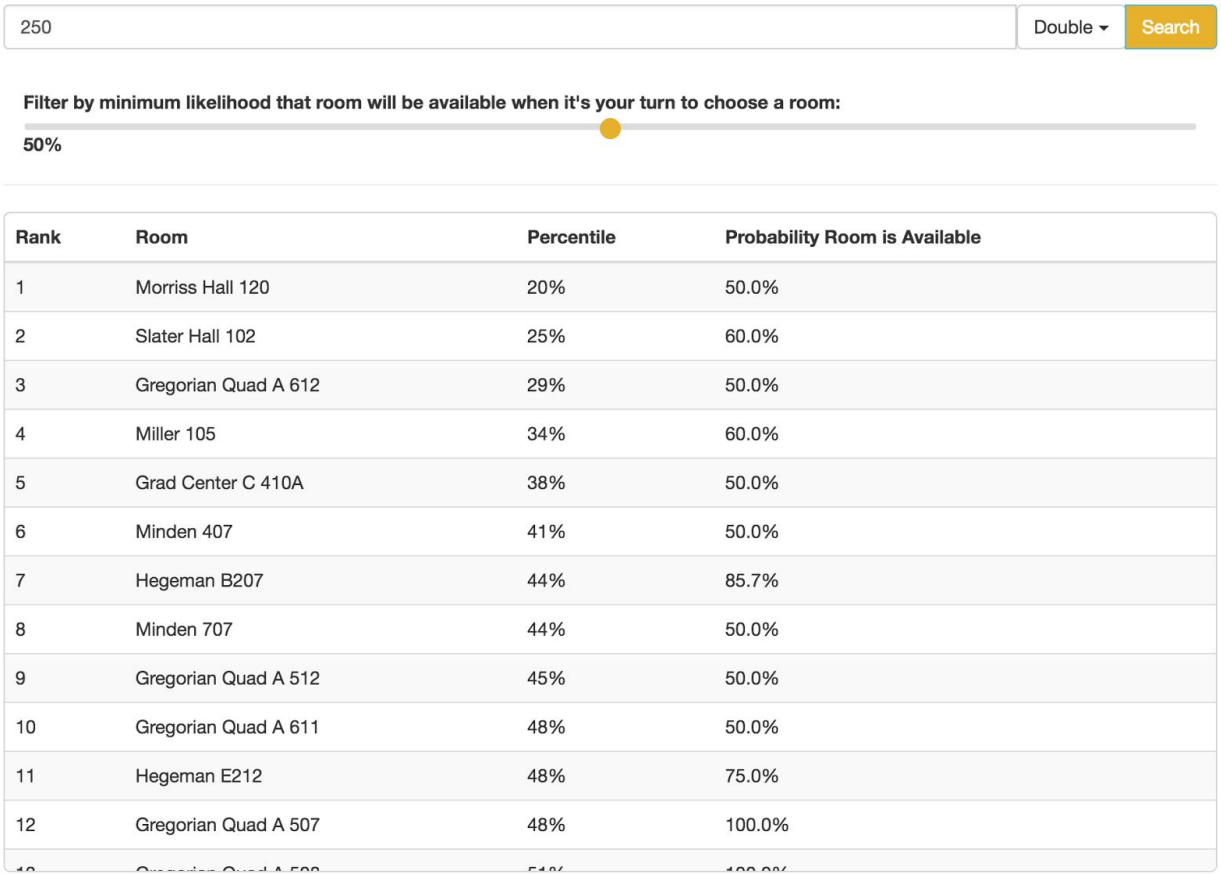


Room Suggestion Generator

We created a room suggestion generator that outputs room suggestions based on lottery number, preferred room capacity, and the percent chance that the rooms remain based on your lottery number. You can adjust the slider to update suggestions to only include rooms that have a certain percent chance of remaining based on your lottery number.

Rooms are ordered in terms of desirability. The **percentile** found next to each listing indicates a room's desirability relative to other rooms of the same capacity. For example, if a room is in the 15th percentile, it is more desirable than 85% of rooms of the same size. Desirability is based on the room rankings discussed in the previous section.

In the image below, you can see the results a student would see if they looked up suggestions for a lottery number of 250 with a preference for a double, and filtered for a minimum of 50% chance that the room would be remaining when it was their turn to select a room.



Rank	Room	Percentile	Probability Room is Available
1	Morriss Hall 120	20%	50.0%
2	Slater Hall 102	25%	60.0%
3	Gregorian Quad A 612	29%	50.0%
4	Miller 105	34%	60.0%
5	Grad Center C 410A	38%	50.0%
6	Minden 407	41%	50.0%
7	Hegeman B207	44%	85.7%
8	Minden 707	44%	50.0%
9	Gregorian Quad A 512	45%	50.0%
10	Gregorian Quad A 611	48%	50.0%
11	Hegeman E212	48%	75.0%
12	Gregorian Quad A 507	48%	100.0%

Important Feature Learning

For the machine learning aspect of this project, we ran linear regression to find out which features most strongly predict students' dorm preference. (i.e. What makes a dorm most desirable?)

We extracted features from the housing lottery data, and used linear regression to predict average lottery score based on those features. We used occupancy and distance from the Nelson, the Ratty, and the Main Green.

Feature weights are as follows (all features were first normalized):

Feature	**Weight**
Distance to Nelson Fitness Center	0.379
Distance to the Main Green	-0.154
Distance to the Ratty	-0.370
Occupancy of Room	1.336

Occupancy turned out to be the most important feature. The positive weight of occupancy means that students prefer rooms with lower occupancy. This matched our hypothesis, as singles are usually the first rooms to be selected, whereas later rooms are all doubles and triples. Negative weights on Distance to the Main Green and Ratty mean that students prefer dorms closer to these locations. One surprising result was that Distance to the Nelson has such a strong positive weight. An explanation for this could be that the "Distance" features are not independent, and the Nelson is on the opposite side of campus from the Main Green and Ratty, which seem to be attractors.

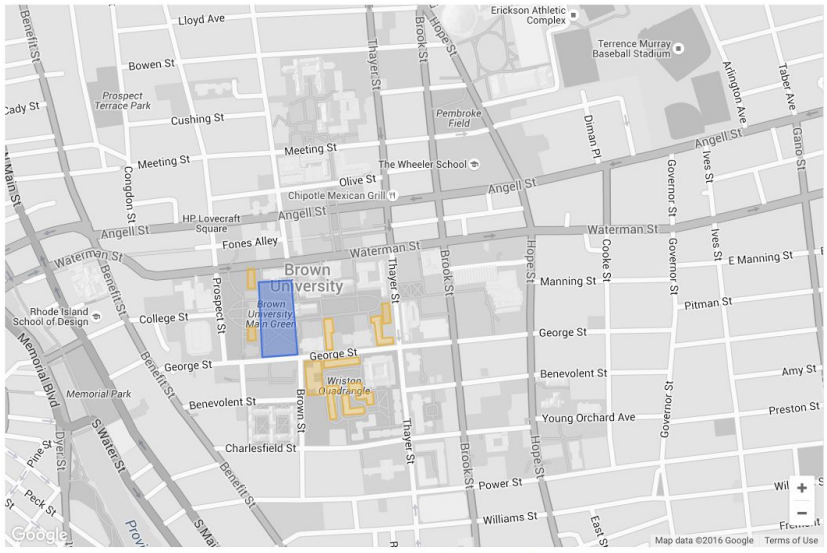
We wanted to include square footage data for each room, which we believe would be one of the most important features; however, we found that the only way to gather those data was to hand-copy the notations in the floor plan images of each room. Extracting these data from the Reslife floor plans for only a couple hundred rooms took several hours, and we did not have time to collect the rest.

Optimal Dorm Location Lookup

Oftentimes, students will pick their dorm largely based on its location to places around campus that are important to them, such as the CIT and the Ratty. Currently the best way to accomplish this is by opening Google Maps and looking where the dorm is in relation to where the important place or places are.

In response to this need, we decided to create a tool that allows student to select important locations around campus and then give them the 10 dorms that minimize the distance to all selected locations. To gather the polygon data we went into Google Maps and got the latitude and longitude coordinates for each corner of every dormitory and important place. To calculate the distance, we simply find the centroid of all of polygon points for each building and then use the Haversine formula to calculate the distance between each important place for each dorm. We then sort the dorms based on the smallest summed distance to each important place and display the top 10.

In the image below, you see what a student would see if they selected the Main Green as the single location most important to them.



#	Dorm
1	Slater Hall
2	Hope College
3	Littlefield Hall
4	Wayland House
5	Sears House
6	Olney House
7	Hegeman
8	Caswell Hall
9	Marcy House
10	Diman House

Reflection

Our proposal stated that creating an animated map that displays the number of rooms remaining in each dorm over time would be a large component of our final project. Since submitting the proposal, we pivoted and instead focussed on creating a variety of different visualizations. We decided that a stacked area chart was more effective for showing the number of rooms remaining in each residence hall over time than showing this change on a map. We instead created a map visualization of Brown's campus that allows students to select locations on campus (i.e. Main Green, Ratty, Nelson, etc.) to discover the optimal residence halls for them based on distance to these locations.

With this exception, we achieved all of our final project proposal and more. In addition to creating a room suggestion generator, we also created an interface that allows students to look up individual rooms and discover their rankings relative to other available rooms. The new map visualization was also beyond the scope of our proposal. We ran linear regression to find out which features most strongly predict students' dorm preference. All in all, we are proud of our work on this project and feel that we achieved our goal of demystifying the housing lottery. We compiled these tools into an interactive website that can be found at **bit.ly/housing-games**. We feel that this site will be very useful for students entering the housing lottery in upcoming years, and that it is far more informative than the simple floor plans that can be found on Residential Life's website.

Further, the data used for this project presented some interesting challenges.

- **Capitalization**
  - There were inconsistencies in the way residential halls were capitalized. For example, Grad Center D was entered as Grad Center D, Grad Center d, and Grade center D within the data.
- **Non-Existent Residence Halls**
  - In some years, a residence hall called Fake Building was listed for several entries in the data, which is slightly puzzling. We recently discovered this is how Brown's system handles students entering "Summer Assignment," which is the process of being randomly assigned a room over the summer.
- **No Shows**
  - "No Shows" (students who did not select a dorm during the allotted lottery time) were recorded in the data.
- **Changes in Room Allocation**
  - Room allocation at Brown has changed since 2006, and continues to change every year. For instance, rooms in Andrews, Miller, and Metcalf are no longer available for selection during the lottery, as they are now freshmen only dorms.

The same problem exists for rooms that are now allocated to program housing, such as “Social Action House” in Diman.

Future Work

If we had more time to work on this project we would improve it in multiple ways:

- We would spend more time gathering additional data for features to improve our ML algorithm. For example, having square footage data for every room would likely show room size to be one of the most important features that students select on.
- We would work to incorporate the new data format for 2015 and request the missing 2014 data from Brown. Additionally, we would set up an automated scraping system to update our tools whenever new lottery data is released each year.
- We would like to try to work with Residential Life to access real time data in future housing lotteries, so that we can make tools that reflect real time changes in available housing.

Dedicated to those who lost their dignity in the 2006-2013 Housing Games.