36-303 Team G

Spatial and Analytical Study of Student Housing at Carnegie Mellon University

Ariel Liu, Sam Lavery, Alejandra Munoz, Terra Mack, Shannon Lauricella

Introduction

In the study, “A Quasi-Experimental Approach to Estimating the Impact of Collegiate Housing,” Ryan Yeung observed that students “from the residence halls to off-campus housing…become less integrated into the academic and social systems of the college” (Yeung, 2011). Research sponsored by the U.S. Department of Education upholds this idea that living on campus provides a stronger support system, more engagement in educational practices, and increased social interaction (Schudde, 2011). These key factors provide “the single most consistent within-college determinant of the impact of college”(Pascarella & Terenzini, 2005).

However, the numbers do not seem to match this sentiment. According to figures published in the Digest of Education Statistics, 86.2% of undergraduate students in the United States live off-campus, with 55.2% not living with their parents (Snyder, Dillow, & Hoffman, 2009), suggesting that students and colleges do not necessarily “subscribe to this theory of the benefits of on-campus housing” (Yeung, 2011). Recent articles published by *The Tartan*, Carnegie Mellon University’s student-run newspaper, indicate that more and more students at CMU are actively looking to move off-campus (Fitzgerald, 2006).

This discrepancy between the reported increased well-being of students living on-campus versus the number of students living off-campus served as a motivating factor for our survey. Do students care about integrating the supposed benefits of living on-campus when choosing an off-campus residence? Considering factors like neighborhood, proximity to peer groups, and relative distance to campus and campus amenities, like shuttle stops may increase academic well-being and foster social interaction, perks that are often attributed to living on-campus.

Our survey seeks to answer questions about the dynamics of student housing at CMU. We are particularly interested in investigating the correlation between where students choose to live and what they choose to study. The results of the survey will be valuable for students in finding neighborhoods within the city that are popular with students like themselves as well the university in planning shuttle routes, campus police coverage, and future housing projects

Methods

In order to explore our question, we needed to understand the population to sample, what questions to pose to the population, and how to process the data.

Target Population and Sampling Frame

it would be helpful to give the sizes of the populations and the registrar's data bases here.

The target population in our study is undergraduate and graduate students enrolled at CMU that live off-campus (Main campus Pittsburgh). It is the same population that we are looking to make inferences about from our survey. Our sampling frame includes records of student housing information we were able to obtain from the Office of the Registrar. The target population differs from the sampling frame in that addresses are self-reported to the registrar and students may neglect to update their address. Thus, we will not have access to information for the entire population enrolled, so our sampling frame will include only those students who comply with the registrar's office or volunteered their information to CMU.

Sample Size

We will use our data as census data for further analysis, but we calculated the margin of error for our sample if we were to use a sampling scheme of simple random sample without replacement. From class we know:

$$\frac{Nn_0}{N+n_0}, \ where \ n_0 = \frac{z^2_{\propto/2}(SD)^2}{(ME)^2}$$

$$z_{\propto/2} = 1.96$$

$$SD = \sqrt{p(1-p)}$$

To calculate p, the population proportion, we selected the following question: 'Is this person a member of CIT (Carnegie Institute of Technology)?'. Then, from the CMU Factbook, we used the head count of students in each college enrolled in Fall 2011 in Pittsburgh to calculate the proportion of the student body represented by students enrolled in CIT (p).

$$N_{CIT} = 3217$$

$$N = 10957$$

$$p = \frac{N_{CIT}}{N} = \frac{3217}{10957} = 0.293$$

$$SD = \sqrt{0.293(1 - 0.293)}$$

$$SD = 0.455$$

We choose a narrow margin of error of 0.0098 for our sample size calculation. why?

$$n_0 = \frac{1.96^2\ 0.455^2}{0.0098^2}$$

$$n_0 = 8281$$

According to the CMU Factbook, there are 2,252 undergraduates living off-campus and 5,769 graduates living off-campus.[1] Therefore:

$$N_{off-campus} = 2252 + 5769 = 8021$$

$$n \geq \frac{N_{off-campus} n_0}{N_{off-campus+n_0}}$$

$$= \frac{(8041)(8281)}{8041 + 8281} = 4079.6 \approx 4080$$

So the minimum sample size for a 0.0098 margin of error is 4080 people.

Data Collection

We collected data from the administrative records provided by the Office of the Registrar. We believe surveying data records is a more accurate and reliable method in comparison to asking students directly. This mode of collection and survey can help reduce high non-response and coverage errors.

We were successfully able to obtain off-campus housing records from the University registrar. The records have 891 undergraduate records and 4,036 graduate records. The registrar provided us with all the records that they had. According to the CMU Factbook, there are 2,252 undergraduates living off-campus and 5,769 graduates living off-campus. Clearly, the ratio of undergraduate records to graduate records is not the same as the population ratio, but this could be explained by response errors more relevant to undergraduates.

Possible reasons and sources for apparent bias

Most undergraduates start their CMU careers living on-campus so changing their address to an off-campus location will probably be less likely reported to the registrar (especially if they still use their SMC mailboxes to get mail from the university). Other sources of bias in the collection of data could be the limitation of department information. When looking at clusters of students off-campus according to their major, a student could have more than one major, but the records only indicate one major and one affiliated department per student. Another possible bias is that students may not have reported accurate addresses of zip codes e.g. using abbreviations or interchangeable zip codes. We had to sort through the data to locate these inconsistencies as part of the data cleaning process.

Given that we have obtained all of the records from the registrar for students living off-campus that provided responses, we are going to analyze our sample with two different methodologies: as a census and as a stratified sample. As part of cleaning the data, we noticed that graduate students have a duplicate entry for their offices; therefore we made sure to only report their residences in our results. Other issues we needed to consider when cleaning the data were duplicate records, response missingness, and incorrect forms of address format.

Questionnaire

In general, the questions included in the survey consisted of which department, school, and class year the student belonged to as well as where the student lived and the distance and time it took to travel to campus from their off-campus residence.

A sample of questions included:

- Identification of class
    - Does this record belong to an undergraduate student?
    - Does this record belong to a graduate (Master) student?
- Identification of college/department
    - Does this record belong to a student enrolled in the School of Computer Science (SCS)?
        - Which department?
    - Does this record belong to a student enrolled in the College of Fine Arts (CFA)?
        - Which department?

The full questionnaire can be found in Appendix A.

Post-Survey

Based on all the data we obtained from the office of the registrar, we had to format our data into a uniform coding system so that it could be used for analysis. After reviewing the data, we found that we had to omit 182 records due to 157 students reporting campus addresses, 32 reporting duplicate addresses, and three reporting no addresses to the office of the registrar. We did not include these records in our dataset since our question of interest is related to only assessing off-campus housing for students. We decided to include the first address listed for students who reported two addresses, allowing us to take account for the individual's information in the data.

Our main variables included in the current analysis are address, class year, college, distance to Carnegie Mellon's campus, distance to the closest CMU shuttle stop, distance to the closest Pittsburgh Port Authority bus stop, distance to the closest supermarket, and distance to the closest bar. The class year variable was separated three classes: undergraduate, Master's, and PhD students. The college variable was comprised of eight distinct colleges, which included the College of Fine Arts (CFA), Carnegie Institute of Technology (CIT), Heinz College (HC), Dietrich College of Humanities and Social Sciences (HSS), Mellon College of Science (MCS), School of Computer Science (SCS), Tepper School of Business (TSB) and a representative college for interdisciplinary majors (CMU). The housing variable was comprised of our address list and was coded into street name, apartment, city, and zip code. To estimate the measurements for our distance variables, we used the ArcMap Geographic Information Systems (GIS). Using GIS we were able to obtain a map of all the addresses within the city of Pittsburgh and use it to estimate these distances for each student's off-campus address.

need a citation for this.

Many addresses listed included students who live in cities that are outside the city of Pittsburgh, such as Homestead, Monroeville, etc. We were able to add separate maps of these cities to the Pittsburgh map and estimate their distances to campus using GIS. However we were only able to analyze students' distances to supermarkets, bars, bus stops, and shuttle stops for students living in the city of Pittsburgh since the GIS maps were not able to provide this information for students outside of the city. Therefore the following analyses of distances are based on the addresses that were only in the city of Pittsburgh. Distribution graphs and tables for the full data set and the observations within the city can be found in Appendix B.

The counts by class and college for the full data set of all 4,090 students in comparison to CMU's actual population for 2011-2012 (obtained from CMU Factbook), is shown below.

well done graphs & tables!

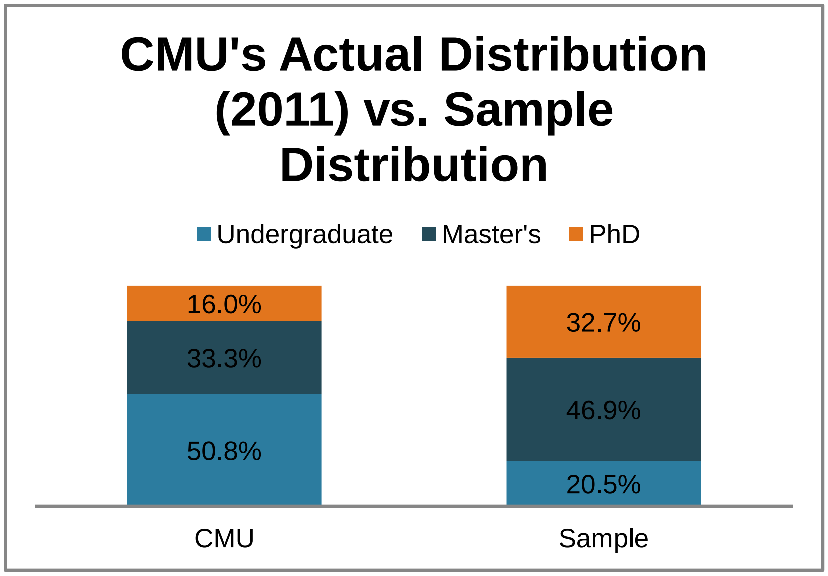


Figure 1: Graph showing distribution by class using all students

Table 1: Showing students distribution by class

	CMU	Sample
Undergraduate	5,843	837
Master's	3,830	1,917
PhD	1,840	1,336
TOTAL	11,513	4,090

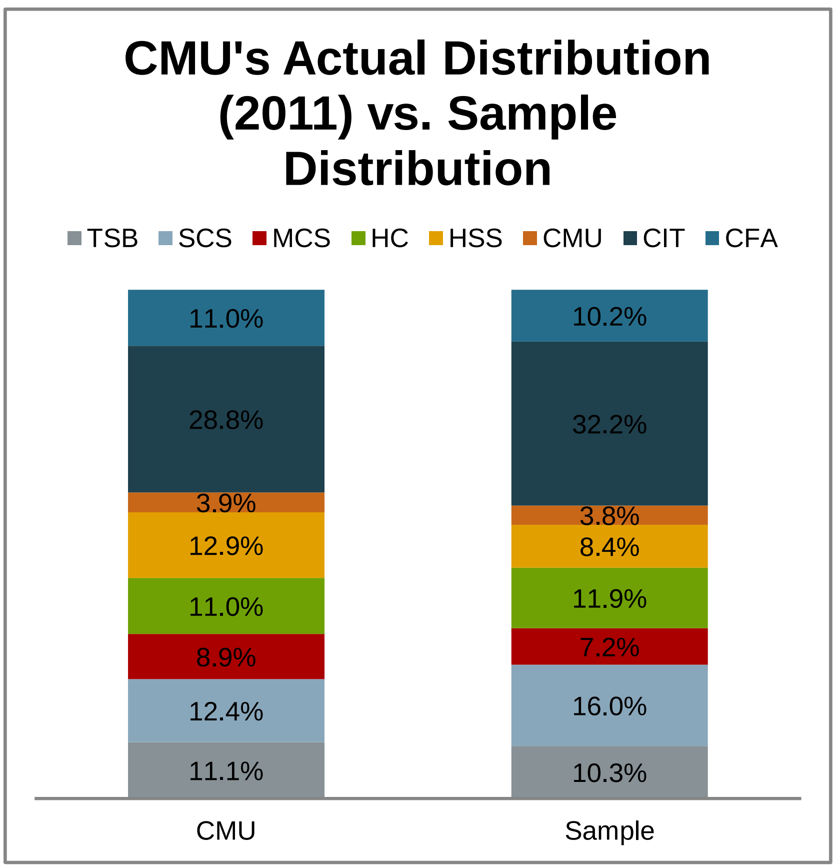


Figure 2: Graph showing distribution by college using all students

Table 2: Showing students distribution by college

	CMU	Sample
CFA	1,267	417
CIT	3,319	1,318
DC	1,485	343
HSS	1,269	488
CMU	452	155
MCS	1,019	294
SCS	1,429	653
TSB	1,273	422
TOTAL	11,513	4,090

Weighting

As can be seen, the data from the sample is not representative of CMU's population. Specifically, undergraduates seem to be particularly underrepresented in the sample. We used post-stratification weighting to account for this misrepresentation in the data. We computed weights by class and weights by college to use in analysis. The results of the weighting are shown in the table below.

Table 3: Showing post-stratification weights by class

	Population	Sample	Population Proportion	Sample Proportion	Weight
Undergraduate	5843	837	0.5075	0.2046	2.47996
Master's	3830	1917	0.3327	0.4687	0.70976
PhD	1840	1336	0.1598	0.3267	0.48927

Table 4: Showing post-stratification weights by college

	Population	Sample	Population Proportion	Sample Proportion	Weight
CFA	1267	417	0.11005	0.10196	1.07938
CIT	3319	1318	0.28828	0.32225	0.89460
CMU	452	155	0.03926	0.03790	1.03596
HC	1269	488	0.11023	0.11932	0.92380
HSS	1485	343	0.12899	0.08386	1.53804
MCS	1019	294	0.08851	0.07188	1.23129
SCS	1429	653	0.12412	0.15966	0.77742
TSB	1273	422	0.11057	0.10318	1.07165

The analysis done from this point on utilizes the post-stratification weights. However, the analysis performed on the unweighted data (taking our sample to be a census) can be found in Appendix C.

good (presenting both analyses)

Observations Outside Allegheny County

There were 10 students in our full data set with residences outside of Allegheny County. We chose to consider these observations to be outliers and thus not include these them in our statistical analysis. The majority of these students were graduate students and all of the students that lived outside of Allegheny County were enrolled in CFA, CIT, HC or TSB. The weighted and unweighted demographics can be seen in the graphs and tables below.

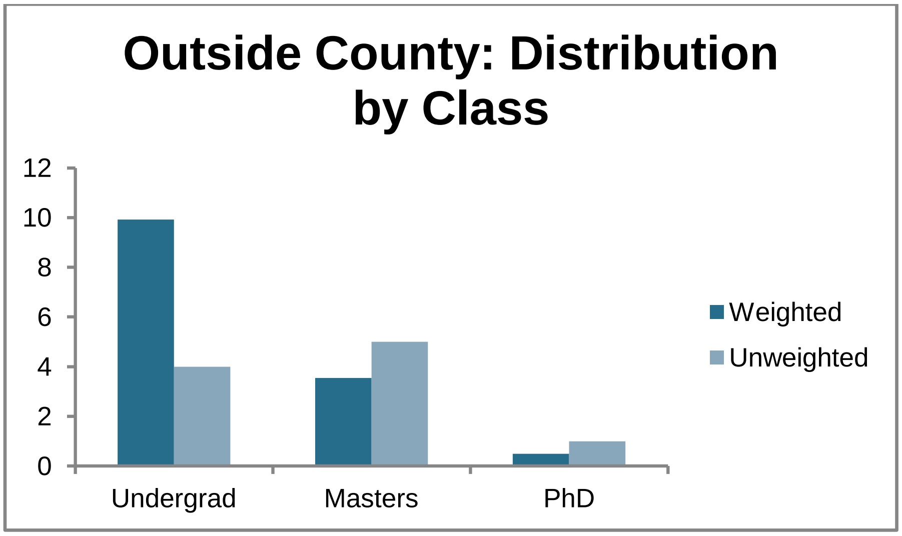


Figure 3: Graph showing distribution by class of students outside of Allegheny County.

Table 5: Distribution of students outside Allegheny County by class.

	Undergraduate	Master’s	PhD
Unweighted	4	5	1
Weighted	9.92	3.55	0.49

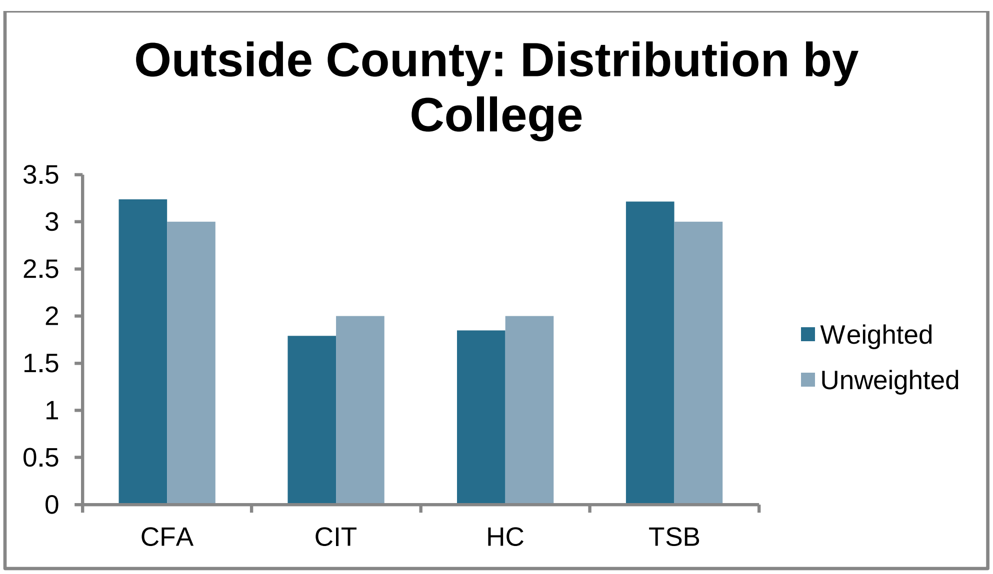


Figure 4: Graph showing distribution by college of students outside of Allegheny County.

Table 6: Distribution of students outside Allegheny County by college.

	CFA	CIT	CMU	HC	HSS	MCS	SCS	TSB
Unweighted	3	2	0	2	0	0	0	3
Weighted	3.24	1.79	0	1.85	0	0	0	3.21

Observations Outside Pittsburgh City Limits

There were 197 observations of students that lived within Allegheny County but outside of the city limits of Pittsburgh. Although we were able to analyze the distance these students lived from CMU's campus, we were limited in further analysis of these observations. The majority of students that lived outside the city of Pittsburgh were Master's students. Their weighted and unweighted demographics can be seen in the graphs and tables below.

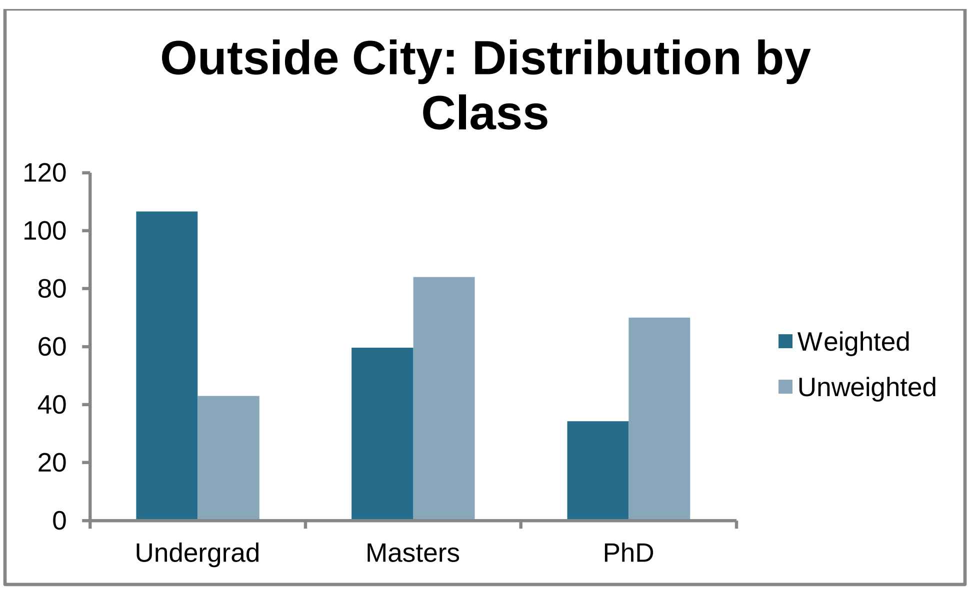


Figure 5: Graph showing distribution by class of students outside of the City of Pittsburgh.

Table 7: Distribution of students outside of the City of Pittsburgh by class.

	Undergraduate	Master's	PhD
Unweighted	43	84	70
Weighted	106.64	59.62	34.25

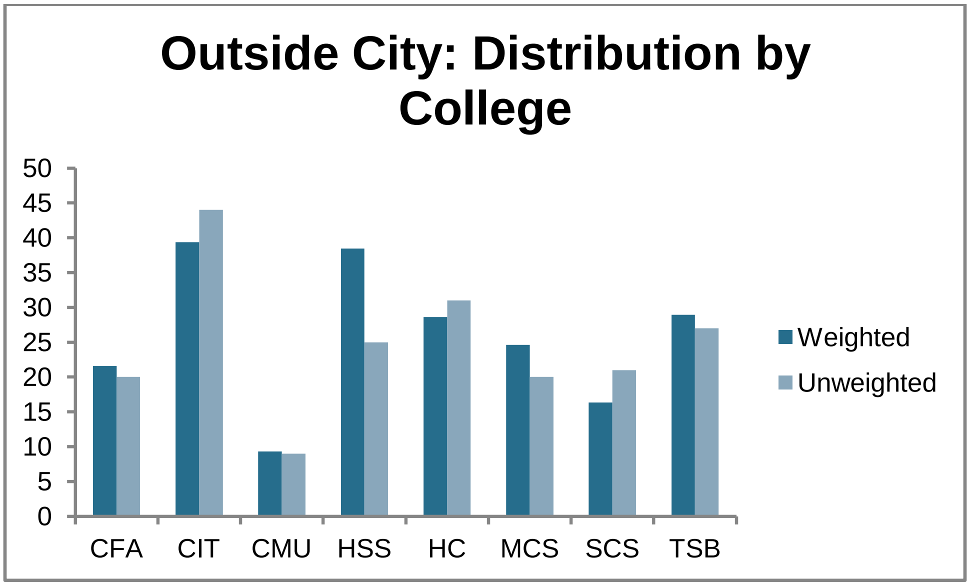


Figure 6: Graph showing distribution by college of students outside of the City of Pittsburgh.

Table 8: Distribution of students outside the City of Pittsburgh by college.

	CFA	CIT	CMU	HC	HSS	MCS	SCS	TSB
Unweighted	20	44	9	25	31	20	21	27
Weighted	21.59	39.36	9.33	28.64	38.45	24.63	16.33	28.93

Results

The primary research question motivating our analysis pertains to finding the relationship, if any, between where students live and what they choose to study.

We found five different distance variables for each observation in our sample of students that lived within the city of Pittsburgh (distance to CMU, distance to closest bus stop, distance to closest shuttle stop, distance to closest supermarket, and distance to closest bar) and compared the mean distance between different groups of students. The details of the mean distances and tests can be seen in Table 9.

Distance to CMU's Campus

Our most important distance variable considered was mean distance to Carnegie Mellon's campus. Visual representation of the spatial distances between classes in relationship to campus can be seen in Figure 7.

these (& following) are lovely graphs.

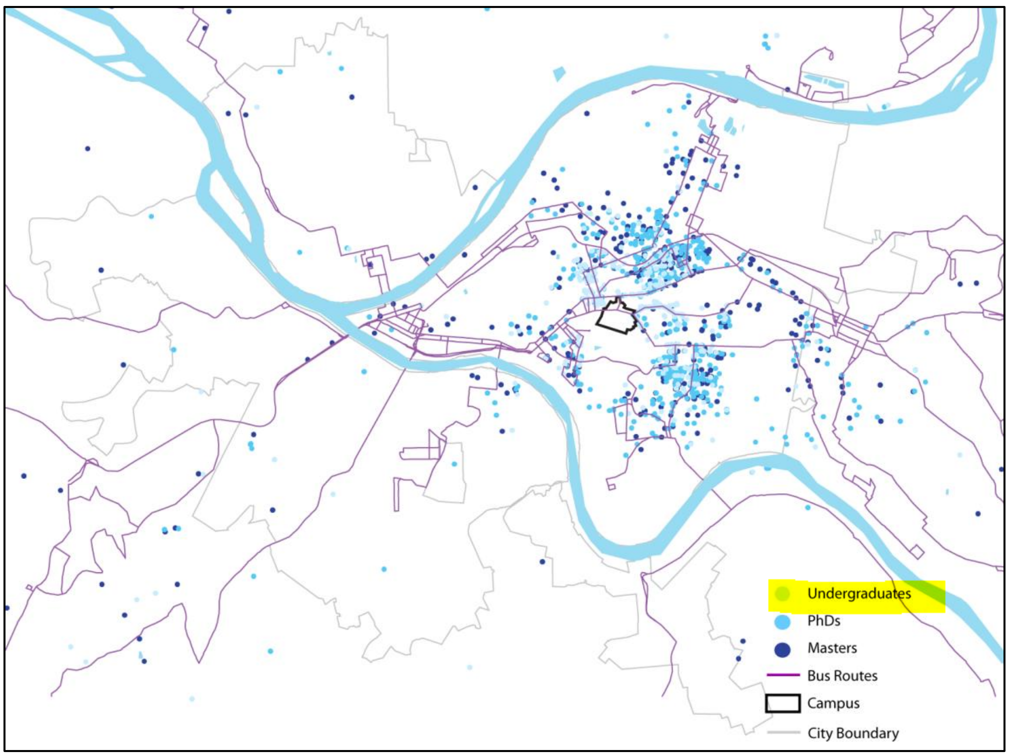
need cite sources for underlying maps, GIS locations, and graphing software (a package in R, and/or google app, I guess?)

Figure 7: Map showing distances to campus for undergraduates, master's, and PhD students

We calculated the mean distances by class (undergraduate vs. Master's vs. PhD) and by the eight distinct colleges. Then, using two sample t-tests, correlation tests, and ANOVA tests, we looked for significance between groups in terms of their respective distances to campus. From the t-tests, we found a significant difference between mean distance to campus for all pairs of groups by class (undergraduate vs. Master's, undergraduate vs. PhD, and Master's vs. PhD). The correlation test provided evidence of a significant difference in distance to campus across all classes. From the ANOVA tests, we found a significant difference in distance to campus across schools of Master's students and PhD students, but no significance across schools of undergraduate students.

Distance to Closest Bus Stop

We used GIS to find the closest Pittsburgh Port Authority bus stop to each observation. A visual map of the location of the bus routes and the addresses can be seen in Figure 8.



kind of unfortunate color for undergrads, hard to see & distin-guish from phds...

Figure 8: Map showing distances to closest bus stop by class.

We used the same tests to look for significance between groups in terms of their respective mean distances to bus stops. We found a significant difference in mean distance to closest bus stop for all pairs of groups by class (undergraduate vs. Master's, undergraduate vs. PhD, and Master's vs. PhD). Again, the correlation test showed a significant difference in mean distance to bus stop across all classes. However, there was no significant difference across schools for any class.

Distance to Closest CMU Shuttle Stop

Carnegie Mellon's shuttle service has four fixed stop routes designed to assist CMU students and faculty in commuting between their residences and campus. Below, in Figure 9 is a map of the shuttle stops with buffer zones indicating distances from the shuttle stops with radii of a quarter mile and a half-mile.

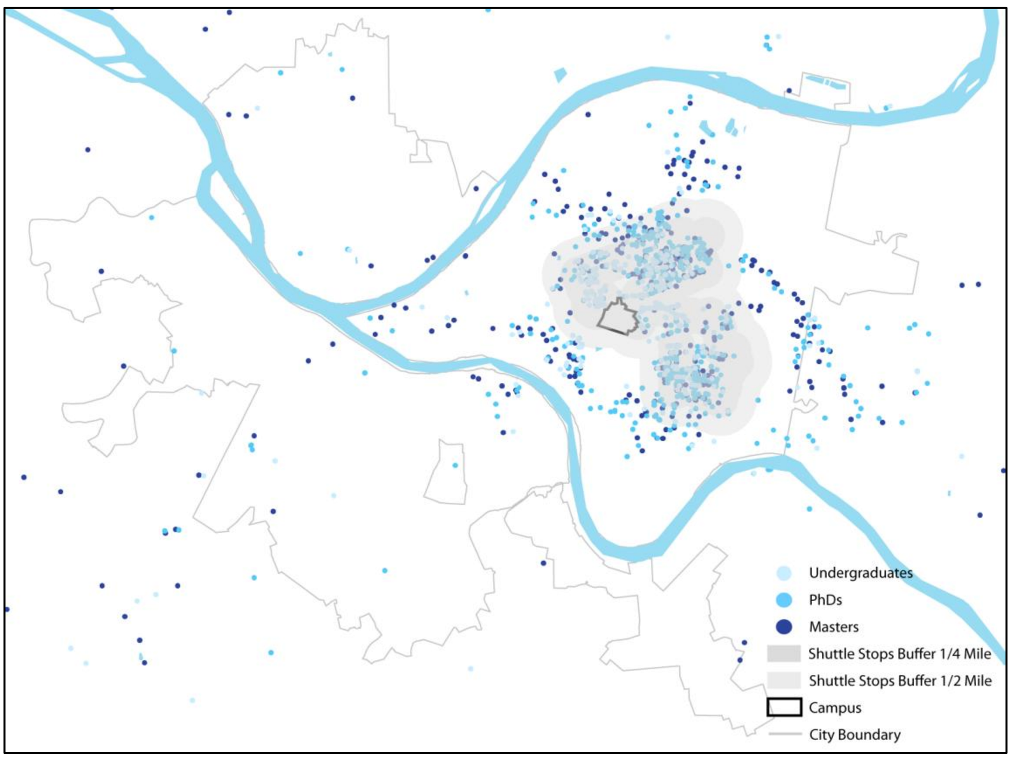


Figure 9: Map showing distances to closest shuttle stop by class.

We were able to use GIS to find the closest shuttle stop to each student's off-campus address. We found a significant difference in mean distance to closest shuttle stop for undergraduates vs. Master's students and undergraduates vs. PhD students. There was no significant difference between Master's students and PhD students. We did find that there is a difference in mean distance to shuttle stop between colleges for Master's students and between colleges for all classes.

Distance to Closest Supermarket

We were interested in the possible relationship between choosing an off-campus residence and the location's proximity to grocery stores, a resource that most college students use. The visual representation of the location of grocery stores can be seen in Figure 10 below.

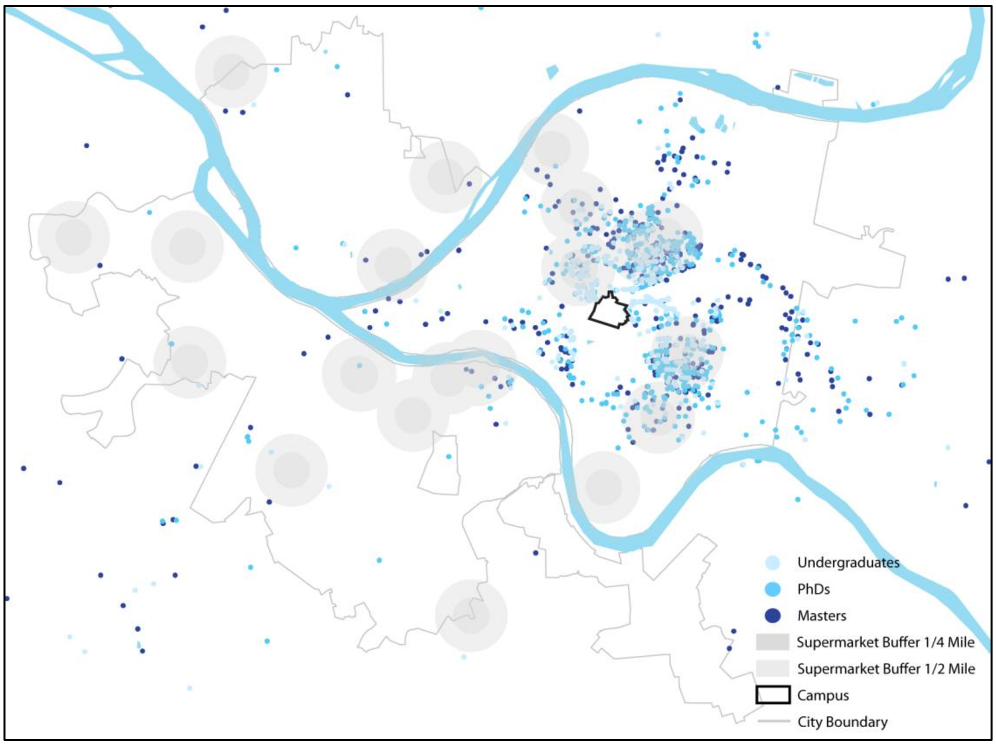


Figure 10: Map showing distances to closest supermarket by class.

From GIS, we were able to find the distance between each residence and the closest supermarket. We found a significant difference in mean distance to closest supermarket for undergraduates vs. Master's students and undergraduates vs. PhD students, but no significant difference for Master's vs. PhD students. The ANOVA tests indicated evidence of a significant difference in mean distance to supermarkets between colleges for Master's students and between colleges for all classes.

Distance to Closest Bar

We were interested in seeing whether proximity to Pittsburgh nightlife had any relationship to where students chose to live off-campus. The visual representation of the location of bars can be seen in Figure 11 below.

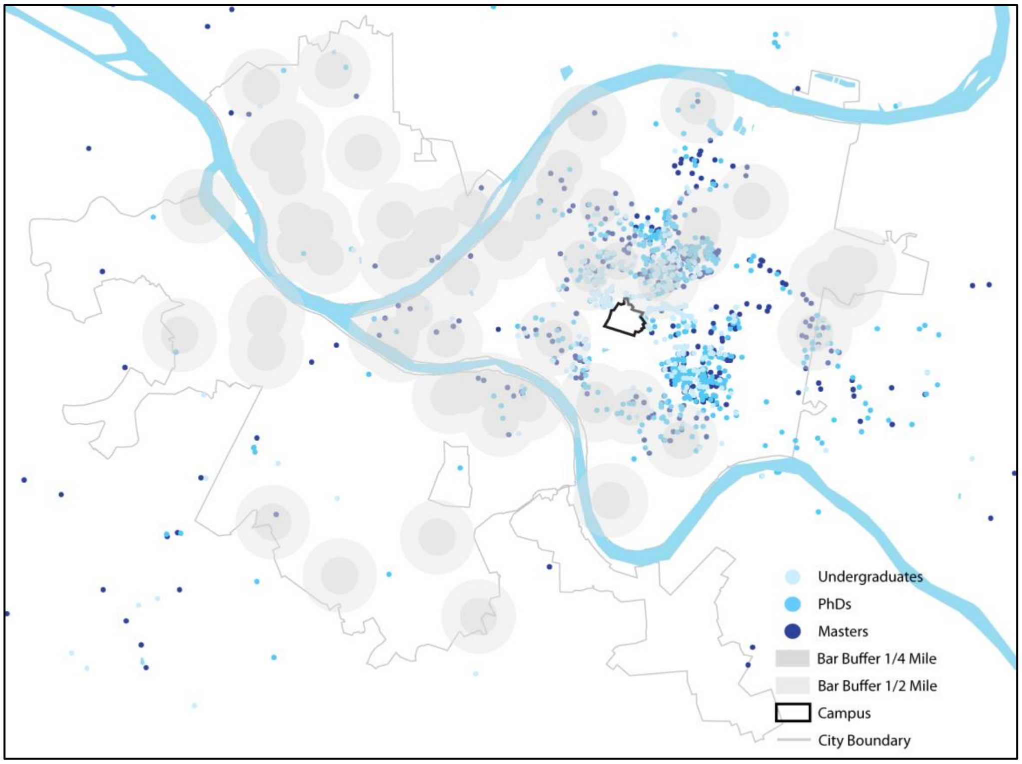


Figure 11: Map showing distances to closest bar by class.

GIS calculated the distance to the closest bar from each observation in our sample. We found a significant difference in mean distance to bars for undergraduates vs. PhD students and Master's vs. PhD students (evidently, PhD students live farther away from bars than other CMU students). From the ANOVA tests, we found evidence of a significant difference in distance to bars between colleges for Master's students and between colleges for all classes.

Table 9: Mean distances, T-Test, Correlation, ANOVA results

Means by Class	Distance to CMU	Distance to Supermarkets	Distance to Bars	Distance to Bus Stops	Distance to Shuttle Stops
Undergrads	0.4776	0.4440	0.4020	0.0591	0.1521
Masters	0.7974	0.3365	0.4054	0.0794	0.2107
PhDs	0.8776	0.3747	0.4471	0.0918	0.2474

Means by College	Distance to CMU	Distance to Supermarkets	Distance to Bars	Distance to Bus Stops	Distance to Shuttle Stops
CFA	0.7186	0.4005	0.4226	0.0803	0.2030
CIT	0.7287	0.3612	0.4228	0.0767	0.1894
CMU	0.8585	0.4273	0.4647	0.0833	0.3387
HC	0.8146	0.3837	0.4119	0.0791	0.2233
HSS	0.7360	0.4690	0.4249	0.0777	0.2410
MCS	0.7770	0.4066	0.4178	0.0830	0.2577
SCS	0.7616	0.3655	0.4436	0.0822	0.1950
TSB	0.7908	0.3814	0.3442	0.0793	0.1933

T-Test	Distance to CMU	Distance to Supermarkets	Distance to Bars	Distance to Bus Stops	Distance to Shuttle Stops
U. vs. M.	<0.01	<0.01	Not significant	<0.01	<0.01
U. vs. PhD	<0.01	<0.01	<0.01	<0.01	<0.01
M vs. PhD	<0.01	Not significant	<0.01	<0.01	Not significant

Correlation Test	Distance to CMU	Distance to Supermarkets	Distance to Bars	Distance to Bus Stops	Distance to Shuttle Stops
Classes	<0.01	<0.01	<0.01	<0.01	<0.01

Anova	Distance to CMU	Distance to Supermarkets	Distance to Bars	Distance to Bus Stops	Distance to Shuttle Stops
U: Colleges	Not significant	Not significant	Not significant	Not significant	Not significant
M: Colleges	<0.01	<0.01	<0.01	Not significant	<0.01
PhD: Colleges	<0.01	Not significant	Not significant	Not significant	Not significant
All: Colleges	Not significant	<0.01	<0.01	Not significant	<0.01

* Not significant at alpha level 0f 0.01

Neighborhood Comparison

From the sample of observations within the city limits of Pittsburgh, there were 47 neighborhoods that Carnegie Mellon students resided in. We collapsed the different neighborhoods to run analysis on the five most populous neighborhoods: Shadyside, Squirrel Hill North, Squirrel Hill South, Oakland, and Bloomfield. A map showing the different neighborhoods in the city of Pittsburgh can be seen in Figure 12 and the counts for the number of students in each of these neighborhoods can be seen in Figure 13.

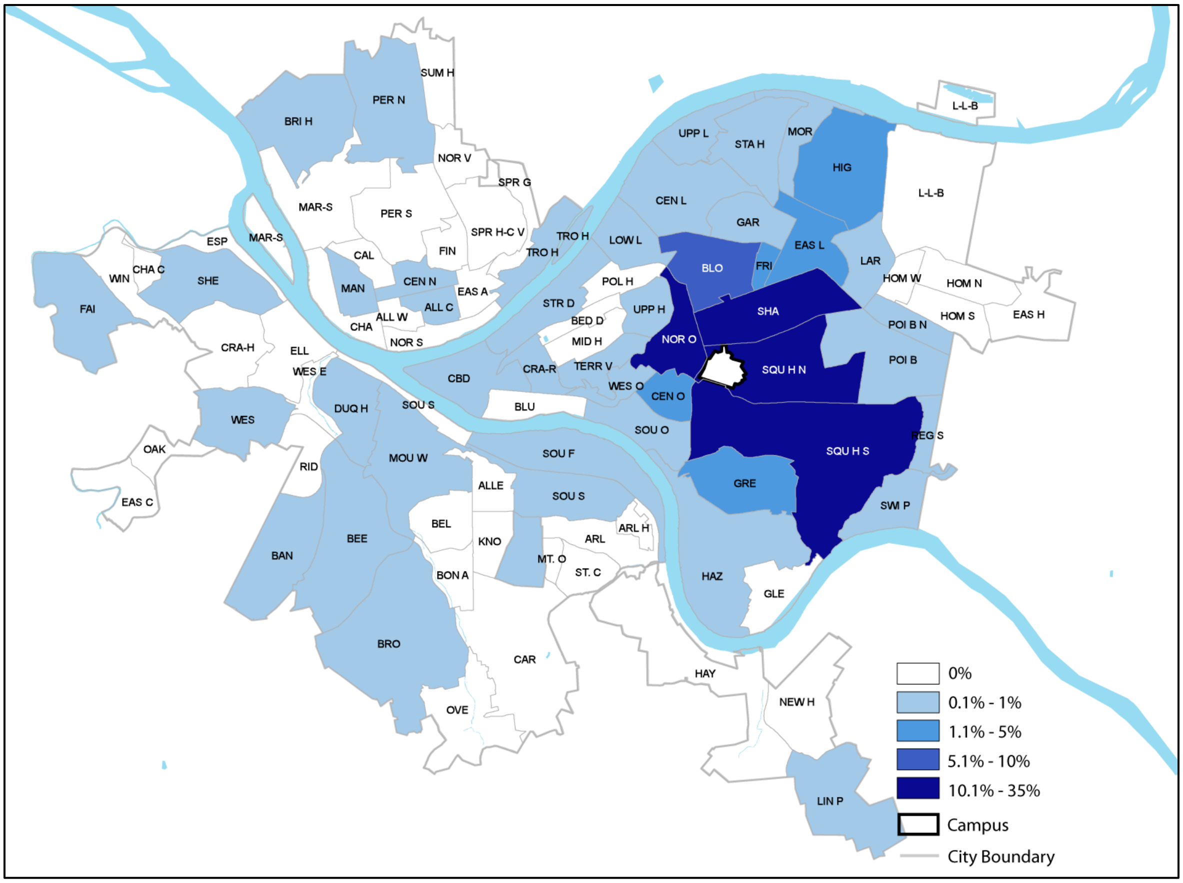


Figure 12: Map showing percentage of students in neighborhoods.

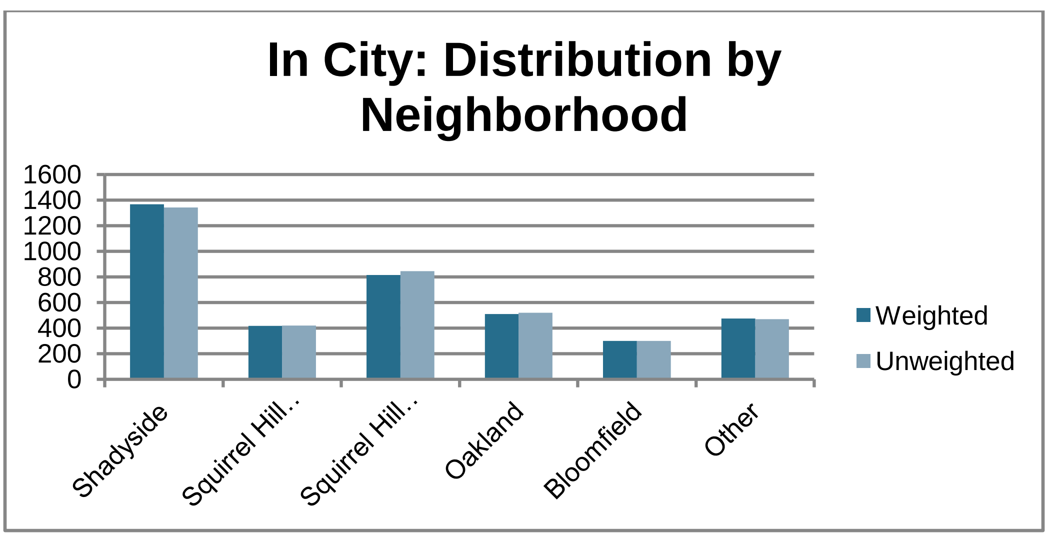


Figure 13: Graph showing distribution of neighborhoods within the City of Pittsburgh.

We used ANOVA tests to check for evidence of significant differences between colleges within each class. We found that the difference in choice of neighborhood between colleges for all classes and between colleges for Master's students was significant. The results are provided in Table 10.

Table 10: ANOVA results of colleges within neighborhoods

ANOVA	
U: Colleges	Not significant
M: Colleges	<0.01
PhD: Colleges	Not significant
All: Colleges	<0.01
Classes	Not significant

* Not significant at alpha level 0f 0.01

Discussion

This survey intended to assess if there was a relationship between what students are studying and where they choose to live. We wanted to look at several different factors that may play into students choosing to live or congregate in a certain area based on class and college. Our main questions included whether distances to Carnegie Mellon's campus, distances to supermarkets, distances to bus stops, and distances to shuttle were indications of where students live. We wanted to see if these differences were also dependent on which year of study or what college students were in.

Our results found that a difference in distances to campus for all three class groups such that Undergraduates students lived the closest to campus, followed by Masters' students, and finally PhD students. Therefore PhD students lived the furthest away from campus. This is to be expected since PhD students are usually older and may have cars to travel to and from campus. Masters' students are also older than undergraduates, but typically younger than PhD students so they may have more variability in their access to cars. There was also a difference in the distances to campus across colleges for Masters and PhD students, but not undergraduates. It is plausible that students who are Masters' and PhD students may have more variability in the in the time they are needed on campus for

good conjectures. along the same lines, I would guess that phd students like quieter nbhds, and these are typically farther from campus.

research or courses across different areas of study. Whereas undergraduates students take many classes outside of their area of study so they do not have this same variability in time needed on campus.

We also found a difference in distances to closest bus stops and shuttle stops for each year of study with undergraduates having a closer average distance. This also supports the result above with undergraduates living closest to campus since undergraduates were the closest to bus stops and shuttle stops, followed by Masters', and PhD students. Since there were no differences across colleges for these variables we can conclude that areas of study are not associated with closeness to bus stops and shuttle stops.

We found that undergraduates tended to live further away from supermarkets than both Master's and PhD students, but there was no difference between Master's and PhD students. This is interesting because it may suggest that supermarkets are not typically close to campus since undergraduates live the closest to campus and yet the furthest from supermarkets. Lastly in looking at distances to the nearest bar, we found that PhD lived further away from bars than Masters' and undergraduates students. However there was no difference between Masters' and undergraduates. This may be in effect because PhD students live further from campus and many bars surround campus. This may also be the case because PhD students tend to be older and therefore may be more focused on research and coursework; so bars may not be an important feature for them when living off-campus.

As for strengths of our survey, obtaining records from the Office of the Registrar allowed us to have a large amount of observations and reduce errors in our survey. Other strengths of our survey included the fact that we were analyzing information based on facts as opposed to opinions. Therefore we possibly obtained more accurate information from students in choosing to look at distances, class, and college information that were recorded. Another strength was that we were able to use the GIS system to calculate distances to campus for our survey. By using GIS we were able to accurately calculate distances rather than using an arbitrary estimate. A weakness in our survey was that data cleaning for all of the records we obtained was very tedious. Whenever a large of data cleaning is necessary, there is a greater possibility for imputation error. Another weakness in our survey was that we were limited to the analysis we could perform based on the information we obtained from the records. We did not have information pertaining to age, gender, or race for example. In future work, this survey could be extended on by assessing other variables that might contribute to the well-being of students who live off-campus. For example research could include looking at housing prices and whether students have a car. Future may want to assess distances by students' majors. We were not able to analyze distances by major since we did not have a large enough sample of students within each major. However future work may be able to group departments together and thus make comparisons and analyze clustering among these departments. Future work could also assess clustering of students based on college, majors, and other demographics. We were not able to analyze clustering of students due to time limitations.

List of References

Fitzgerald, M. R. (2006, September 25). Housing: A Financial Look. *The Tartan*.

Office of Institutional Research and Analysis. (2012). Carnegie Mellon University Factbook 2011-2012. *26*. Pittsburgh: Carnegie Mellon University. Retrieved April 2012, from http://www.cmu.edu/ira/factbook/facts2012.html

Pascarella, E., & Terenzini, P. (2005). *How College Affects Students: A Third Decade of Research (Vol. 2).* San Francisco: Jossey-Bass: A Wiley Imprint.

Schudde, L. T. (2011, Summer). The Causal Effect of Campus Residency on College Student Retention. *The Review of Higher Education, 34*(4), 581-610.

Snyder, T., Dillow, S., & Hoffman, C. (2009). *Digest of Education Statistics, 2008.* Washington, D.C.: National Center for Education Statistics.

Yeung, R. (2011). *A Quasi-Experimental Approach to Estimating the Impact of College Housing*. Syracuse, NY: Syracuse University.

Appendix A: Full Questionnaire

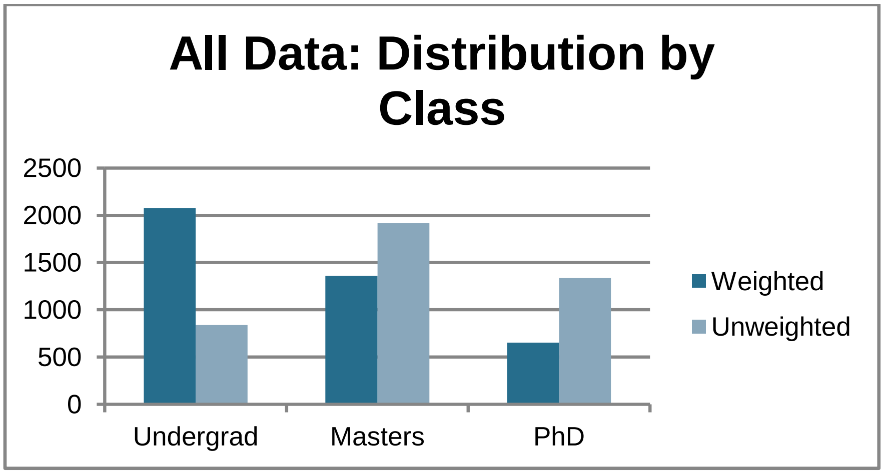
Undergraduate/Graduate
Is the person an undergraduate student?
Is the person a graduate (Master) student?
Is the person a graduate (PhD) student?

College
Is the person a member of Marianna Brown Dietrich College of Humanities and Social Sciences (HSS)?
    Which department?
Is the person a member of Carnegie Institute of Technology (CIT)?
    Which department?
Is the person a member of David A. Tepper School of Business (TSB)?
    Which department?
Is the person a member of School of Computer Science (SCS)?
    Which department?
Is the person a member of College of Fine Arts (CFA)?
    Which department?
Is the person a member of H. John Heinz III College at Carnegie Mellon University (HC)?
    Which department?
Is the person a member of Mellon College of Science (MCS)?
    Which department?
Does the person have an interdisciplinary major (CMU)?

Housing
What is the person's street address?
What city does the person live in?
What neighborhood does the person live in?

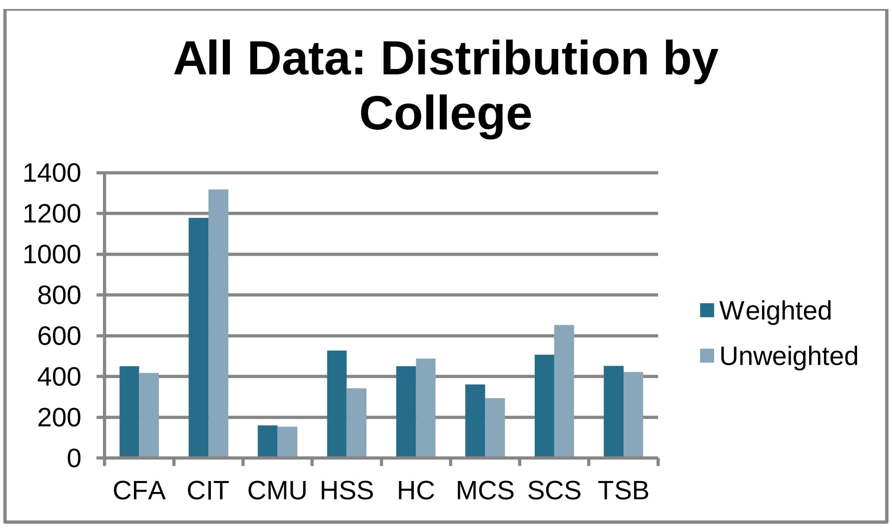
Distances
How far does the person live from campus?
How far is the closest bus stop?
How far is the closest CMU shuttle stop?
How far is the closest supermarket?
How far is the closest bar?

Appendix B: Distribution Graphs and Tables



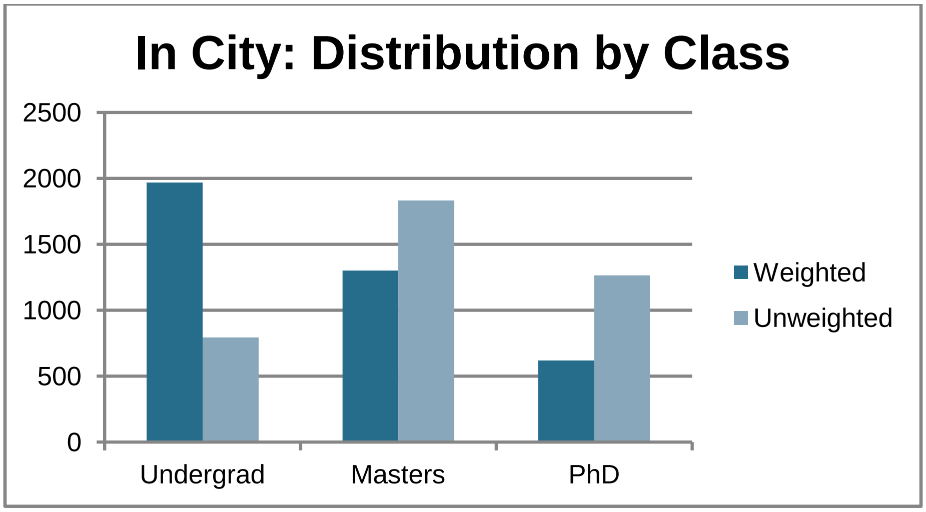
All Data: Distribution by Class

	Undergraduate	Master's	PhD
Unweighted	837	1917	1336
Weighted	2075.73	1360.61	653.66



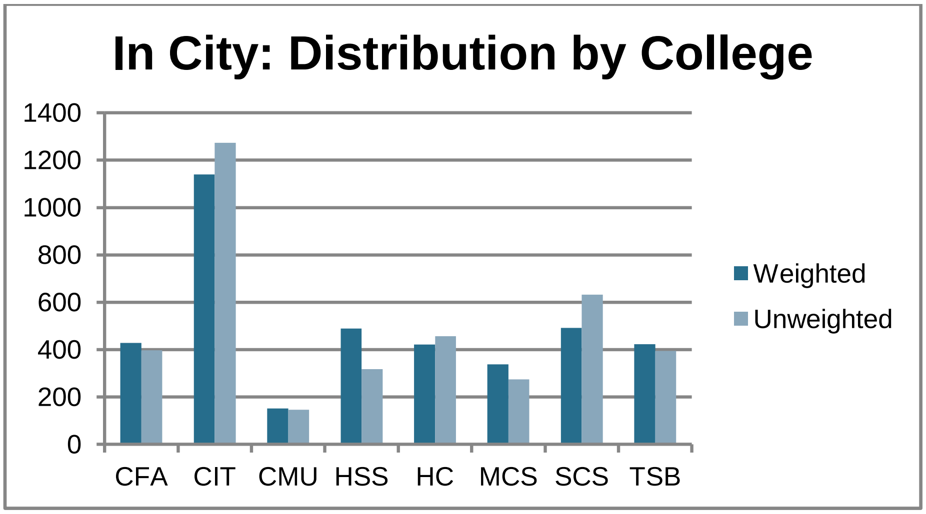
All Data: Distribution by College

	CFA	CIT	CMU	HC	HSS	MCS	SCS	TSB
Unweighted	417	1318	155	488	343	294	653	422
Weighted	450.10	1179.08	160.57	450.81	527.55	362.00	507.65	452.23



In City: Distribution by Class

	Undergraduate	Master's	PhD
Unweighted	794	1833	1266
Weighted	1969.09	1300.99	619.41



In City: Distribution by College

	CFA	CIT	CMU	HC	HSS	MCS	SCS	TSB
Unweighted	397	1274	146	457	318	274	632	395
Weighted	428.51	1139.72	151.25	422.18	489.10	337.37	491.33	423.30

Appendix C: Statistical Tests for Unweighted Data

Means by Class	**Distance to CMU**	**Distance to Supermarkets**	**Distance to Bars**	**Distance to Bus Stops**	**Distance to Shuttle Stops**
Undergrads	0.48	0.44	0.40	0.06	0.15
Masters	0.80	0.37	0.41	0.08	0.21
PhDs	0.88	0.37	0.45	0.09	0.25

Means by College	**Distance to CMU**	**Distance to Supermarkets**	**Distance to Bars**	**Distance to Bus Stops**	**Distance to Shuttle Stops**
CFA	0.72	0.40	0.42	0.08	0.20
CIT	0.73	0.36	0.42	0.08	0.19
CMU	0.86	0.43	0.46	0.08	0.34
HC	0.81	0.38	0.41	0.08	0.22
HSS	0.74	0.47	0.42	0.08	0.24
MCS	0.78	0.41	0.42	0.08	0.26
SCS	0.76	0.37	0.44	0.08	0.19
TSB	0.79	0.38	0.34	0.08	0.19

T-Test	**Distance to CMU**	**Distance to Supermarkets**	**Distance to Bars**	**Distance to Bus Stops**	**Distance to Shuttle Stops**
U. vs. M.	<.01	<.01	Not Significant	<.01	<.01
U. vs. PhD	<.01	<.01	<.01	<.01	Not Significant
M vs. PhD	<.01	Not Significant	<.01	<.01	<.01

Anova	**Distance to CMU**	**Distance to Supermarkets**	**Distance to Bars**	**Distance to Bus Stops**	**Distance to Shuttle Stops**
Classes	<.01	<.01	<.01	<.01	<.01
College	Not Significant	<.01	<.01	Not Significant	<.01

* Not significant at alpha level 0f 0.01

Appendix D: R Code

```
#Import Data
library(RODBC)
channel <- odbcConnectExcel("c:/Users/slaurice/Documents/303/All_Students.csv.xls")
data.all <- sqlFetch(channel, "All_Students")
odbcClose(channel)

channel2 <- odbcConnectExcel("c:/Users/slaurice/Documents/303/WithinCounty.xls")
in.county <- sqlFetch(channel2, "WithinCounty")
out.county <- sqlFetch(channel2, "Outside County")
odbcClose(channel2)

channel3 <- odbcConnectExcel("c:/Users/slaurice/Documents/303/WithinCity.xls")
in.city <- sqlFetch(channel3, "WithinCity")
out.city <- sqlFetch(channel3, "All outside city")
odbcClose(channel3)

in.city<-read.csv("weighted_in_city.csv")

#Subsets each data set by Undergrad, Graduate, Master's, PhD
in.city.UG<-in.city[which(in.city$Class<6),]
in.city.G<-in.city[which(in.city$Class>6),]
in.city.Mas<-in.city[which(in.city$Class==10),]
in.city.PhD<-in.city[which(in.city$Class==20),]

out.city.UG<-out.city[which(out.city$Class<6),]
out.city.G<-out.city[which(out.city$Class>6),]
out.city.Mas<-out.city[which(out.city$Class==10),]
out.city.PhD<-out.city[which(out.city$Class==20),]

in.county.UG<-in.county[which(in.county$Class<6),]
in.county.G<-in.county[which(in.county$Class>6),]
in.county.Mas<-in.county[which(in.county$Class==10),]
in.county.PhD<-in.county[which(in.county$Class==20),]

out.county.UG<-out.county[which(out.county$Class<6),]
out.county.G<-out.county[which(out.county$Class>6),]
out.county.Mas<-out.county[which(out.county$Class==10),]
out.county.PhD<-out.county[which(out.county$Class==20),]

#Collapse Neighborhood Variable
ColNeigh<-vector(length=nrow(in.city))
for (i in 1:length(Class_Cat)){
        if (in.city$Neighborhood[i]=="Shadyside") ColNeigh[i]=0
        else if (in.city$Neighborhood[i]=="Squirrel Hill North") ColNeigh[i]=1
        else if (in.city$Neighborhood[i]=="Squirrel Hill South") ColNeigh[i]=2
        else if (in.city$Neighborhood[i]=="Bloomfield") ColNeigh[i]=3
        else if (in.city$Neighborhood[i]=="Central Oakland") ColNeigh[i]=4
        else if (in.city$Neighborhood[i]=="South Oakland") ColNeigh[i]=4
        else if (in.city$Neighborhood[i]=="North Oakland") ColNeigh[i]=4
        else ColNeigh[i]=5
}

#Renames variables
```

```
names(in.city)=c("ID","Address_1","Address_2","City","State","Zip","College",
"Dept","Class","Dist_CMU","Dist_Bar","Dist_Super","Dist_Bus","Dist_Shuttle",
"Neighborhood")

#Creates categorical variable for Class (Undergrad vs Masters vs PhD)
Class_Cat<-vector(length=nrow(in.city))
for (i in 1:length(Class_Cat)){
	if (in.city$Class[i]<6) Class_Cat[i]=0
	else if (in.city$Class[i]==10) Class_Cat[i]=1
	else Class_Cat[i]=2
}
#Within City -- Distance to Shuttle
t.test(in.city.UG[,14],in.city.G[,14],var.equal=FALSE)
class<-aov(in.city[,14]~Class_Cat)
summary(class)
t.test(in.city.Mas[,14],in.city.PhD[,14],var.equal=FALSE)
UG.school<-aov(in.city.UG[,14]~in.city.UG[,7])
summary(UG.school)
G.school<-aov(in.city.G[,14]~in.city.G[,7])
summary(G.school)
schools<-aov(in.city[,14]~in.city[,7])
summary(schools)
```

good to have found weights package! But as far as I can tell it deals with replication weights, not survey weights....

```
#Weighted Distance to Shuttle
library(weights)
wtd.t.test(in.city.UG$Dist_Shuttle,in.city.Mas$Dist_Shuttle,
weight=in.city.UG$school.weight,weighty=in.city.Mas$school.weight)

wtd.t.test(in.city.UG$Dist_Shuttle,in.city.PhD$Dist_Shuttle,
weight=in.city.UG$school.weight,weighty=in.city.PhD$school.weight)

wtd.t.test(in.city.Mas$Dist_Shuttle,in.city.PhD$Dist_Shuttle,
weight=in.city.Mas$school.weight,weighty=in.city.PhD$school.weight)
wtd.cor(in.city$Dist_Shuttle,Class_Cat,weight=in.city$school.weight)


summary(aov(in.city.UG$Dist_Shuttle~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$Dist_Shuttle~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$Dist_Shuttle~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$Dist_Shuttle~in.city$College,weight=in.city$college.weight))
TukeyHSD(aov(in.city$Dist_Shuttle~in.city$College,weight=in.city$college.weight))


#Within City -- Distance to Supermarket
t.test(in.city.UG[,12],in.city.G[,12],var.equal=FALSE)
class<-aov(in.city[,12]~Class_Cat)
summary(class)
t.test(in.city.Mas[,12],in.city.PhD[,12],var.equal=FALSE)
UG.school<-aov(in.city.UG[,12]~in.city.UG[,7])
summary(UG.school)
G.school<-aov(in.city.G[,12]~in.city.G[,7])
summary(G.school)
schools<-aov(in.city[,12]~in.city[,7])
summary(schools)

#Weighted Distance to Supermarket
```

```
wtd.t.test(in.city.UG$Dist_Super,in.city.Mas$Dist_Super,
weight=in.city.UG$school.weight,weighty=in.city.Mas$school.weight)

wtd.t.test(in.city.UG$Dist_Super,in.city.PhD$Dist_Super,
weight=in.city.UG$school.weight,weighty=in.city.PhD$school.weight)

wtd.t.test(in.city.Mas$Dist_Super,in.city.PhD$Dist_Super,
weight=in.city.Mas$school.weight,weighty=in.city.PhD$school.weight)

wtd.cor(in.city$Dist_Super,Class_Cat,weight=in.city$school.weight)
summary(aov(in.city.UG$Dist_Super~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$Dist_Super~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$Dist_Super~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$Dist_Super~in.city$College,weight=in.city$college.weight))

#Weighted Distance to CMU
wtd.t.test(in.city.UG$Dist_CMU,in.city.Mas$Dist_CMU,
weight=in.city.UG$school.weight,weighty=in.city.Mas$school.weight)

wtd.t.test(in.city.UG$Dist_CMU,in.city.PhD$Dist_CMU,
weight=in.city.UG$school.weight,weighty=in.city.PhD$school.weight)

wtd.t.test(in.city.Mas$Dist_CMU,in.city.PhD$Dist_CMU,
weight=in.city.Mas$school.weight,weighty=in.city.PhD$school.weight)

wtd.cor(in.city$Dist_CMU,Class_Cat,weight=in.city$school.weight)
summary(aov(in.city.UG$Dist_CMU~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$Dist_CMU~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$Dist_CMU~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$Dist_CMU~in.city$College,weight=in.city$college.weight))

#Weighted Distance to Bus
wtd.t.test(in.city.UG$Dist_Bus,in.city.Mas$Dist_Bus,
weight=in.city.UG$school.weight,weighty=in.city.Mas$school.weight)

wtd.t.test(in.city.UG$Dist_Bus,in.city.PhD$Dist_Bus,
weight=in.city.UG$school.weight,weighty=in.city.PhD$school.weight)

wtd.t.test(in.city.Mas$Dist_Bus,in.city.PhD$Dist_Bus,
weight=in.city.Mas$school.weight,weighty=in.city.PhD$school.weight)

wtd.cor(in.city$Dist_Bus,Class_Cat,weight=in.city$school.weight)
summary(aov(in.city.UG$Dist_Bus~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$Dist_Bus~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$Dist_Bus~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$Dist_Bus~in.city$College,weight=in.city$college.weight))

#Within City -- Distance to Bars
t.test(in.city.UG[,11],in.city.G[,11],var.equal=FALSE)
class<-aov(in.city[,11]~Class_Cat)
summary(class)
t.test(in.city.Mas[,11],in.city.PhD[,11],var.equal=FALSE)
UG.school<-aov(in.city.UG[,11]~in.city.UG[,7])
summary(UG.school)
G.school<-aov(in.city.G[,11]~in.city.G[,7])
summary(G.school)
```

```
schools<-aov(in.city[,11]~in.city[,7])
summary(schools)

#Weighted Distance to Bars
wtd.t.test(in.city.UG$Dist_Bar,in.city.Mas$Dist_Bar,
weight=in.city.UG$school.weight,weighty=in.city.Mas$school.weight)

wtd.t.test(in.city.UG$Dist_Bar,in.city.PhD$Dist_Bar,
weight=in.city.UG$school.weight,weighty=in.city.PhD$school.weight)

wtd.t.test(in.city.Mas$Dist_Bar,in.city.PhD$Dist_Bar,
weight=in.city.Mas$school.weight,weighty=in.city.PhD$school.weight)

wtd.cor(in.city$Dist_Bar,Class_Cat,weight=in.city$school.weight)
summary(aov(in.city.UG$Dist_Bars~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$Dist_Bars~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$Dist_Bars~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$Dist_Bars~in.city$College,weight=in.city$college.weight))

in.city<-cbind(in.city,ColNeigh)

#Within City--Neighborhood Comparison
chisq.test(Class_Cat,ColNeigh)
chisq.test(as.numeric(in.city$College),ColNeigh)
chisq.test(as.numeric(in.city.Mas$College),in.city.Mas$ColNeigh)
chisq.test(as.numeric(in.city.PhD$College),in.city.PhD$ColNeigh)
chisq.test(as.numeric(in.city.UG$College),in.city.UG$ColNeigh)

#Weighted Neighborhood Comparison

summary(aov(in.city.UG$ColNeigh~in.city.UG$College,weight=in.city.UG$ps.weight))
summary(aov(in.city.Mas$ColNeigh~in.city.Mas$College,weight=in.city.Mas$ps.weight))
summary(aov(in.city.PhD$ColNeigh~in.city.PhD$College,weight=in.city.PhD$ps.weight))
summary(aov(in.city$ColNeigh~in.city$College,weight=in.city$college.weight))
summary(aov(in.city$ColNeigh~Class_Cat,weight=in.city$school.weight))

#POST-STRAT WEIGHTS (ALL DATA)
#Adds class weight to full data set
school.weight<-vector(length=nrow(data.all))
data.all<-cbind(data.all,school.weight)

for(i in 1:nrow(data.all)){
	if (data.all$Class[i]<10) data.all$school.weight[i]=2.479963173
	else if (data.all$Class[i]==10) data.all$school.weight[i]=0.709759909
	else data.all$school.weight[i]=0.489267275
}

#Adds class/college weight to full data set
ps.weight<-vector(length=nrow(data.all))
data.all<-cbind(data.all,ps.weight)

#Weighting Undergraduates
UG.CFA<-subset(data.all,Class<6 & College=="CFA")
for (i in 1:nrow(UG.CFA)){
	UG.CFA$ps.weight[i]=0.66151
}
```

```
UG.CIT<-subset(data.all,Class<6 & College=="CIT")
for (i in 1:nrow(UG.CIT)){
	UG.CIT$ps.weight[i]=1.197829
}
UG.CMU<-subset(data.all,Class<6 & College=="CMU")
for (i in 1:nrow(UG.CMU)){
	UG.CMU$ps.weight[i]=0.929367
}
UG.HSS<-subset(data.all,Class<6 & College=="HSS")
for (i in 1:nrow(UG.HSS)){
	UG.HSS$ps.weight[i]=1.013957
}
UG.MCS<-subset(data.all,Class<6 & College=="MCS")
for (i in 1:nrow(UG.MCS)){
	UG.MCS$ps.weight[i]=1.209282
}
UG.SCS<-subset(data.all,Class<6 & College=="SCS")
for (i in 1:nrow(UG.SCS)){
	UG.SCS$ps.weight[i]=1.208269
}
UG.TSB<-subset(data.all,Class<6 & College=="TSB")
for (i in 1:nrow(UG.TSB)){
	UG.TSB$ps.weight[i]=0.970305
}
data.all.UG<-rbind(UG.CFA,UG.CIT,UG.CMU,UG.HSS,UG.MCS,UG.SCS,UG.TSB)

#Weighting Masters
M.CFA<-subset(data.all,Class==10 & College=="CFA")
for (i in 1:nrow(M.CFA)){
	M.CFA$ps.weight[i]=0.723581
}
M.CIT<-subset(data.all,Class==10 & College=="CIT")
for (i in 1:nrow(M.CIT)){
	M.CIT$ps.weight[i]=0.704888
}
M.CMU<-subset(data.all,Class==10 & College=="CMU")
for (i in 1:nrow(M.CMU)){
	M.CMU$ps.weight[i]=0.816641
}
M.HC<-subset(data.all,Class==10 & College=="HC")
for (i in 1:nrow(M.HC)){
	M.HC$ps.weight[i]=1.347985
}
M.HSS<-subset(data.all,Class==10 & College=="HSS")
for (i in 1:nrow(M.HSS)){
	M.HSS$ps.weight[i]=1.112272
}
M.MCS<-subset(data.all,Class==10 & College=="MCS")
for (i in 1:nrow(M.MCS)){
	M.MCS$ps.weight[i]=0.719501
}
M.SCS<-subset(data.all,Class==10 & College=="SCS")
for (i in 1:nrow(M.SCS)){
	M.SCS$ps.weight[i]=0.826774
}
M.TSB<-subset(data.all,Class==10 & College=="TSB")
```

```
for (i in 1:nrow(M.TSB)){
	M.TSB$ps.weight[i]=1.453302
}
data.all.M<-rbind(M.CFA,M.CIT,M.CMU,M.HC,M.HSS,M.MCS,M.SCS,M.TSB)

#Weighting PhD
PHD.CFA<-subset(data.all,Class==20 & College=="CFA")
for (i in 1:nrow(PHD.CFA)){
	PHD.CFA$ps.weight[i]=0.760663
}
PHD.CIT<-subset(data.all,Class==20 & College=="CIT")
for (i in 1:nrow(PHD.CIT)){
	PHD.CIT$ps.weight[i]=1.026112
}
PHD.HC<-subset(data.all,Class==20 & College=="HC")
for (i in 1:nrow(PHD.HC)){
	PHD.HC$ps.weight[i]=1.016522
}
PHD.HSS<-subset(data.all,Class==20 & College=="HSS")
for (i in 1:nrow(PHD.HSS)){
	PHD.HSS$ps.weight[i]=1.155138
}
PHD.MCS<-subset(data.all,Class==20 & College=="MCS")
for (i in 1:nrow(PHD.MCS)){
	PHD.MCS$ps.weight[i]=1.02106
}
PHD.SCS<-subset(data.all,Class==20 & College=="SCS")
for (i in 1:nrow(PHD.SCS)){
	PHD.SCS$ps.weight[i]=0.945852
}
PHD.TSB<-subset(data.all,Class==20 & College=="TSB")
for (i in 1:nrow(PHD.TSB)){
	PHD.TSB$ps.weight[i]=0.823986
}
data.all.PHD<-rbind(PHD.CFA,PHD.CIT,PHD.HC,PHD.HSS,PHD.MCS,PHD.SCS,PHD.TSB)
data.all<-rbind(data.all.UG,data.all.M,data.all.PHD)

#Adds college weight to full data set
college.weight<-vector(length=nrow(data.all))
data.all<-cbind(data.all,college.weight)
for(i in 1:nrow(data.all)){
	if (data.all$College[i]=="CIT") data.all$college.weight[i]=0.894595369
	else if (data.all$College[i]=="CFA") data.all$college.weight[i]=1.079382477
	else if (data.all$College[i]=="CMU") data.all$college.weight[i]=1.035956548
	else if (data.all$College[i]=="HC") data.all$college.weight[i]=0.923797119
	else if (data.all$College[i]=="HSS") data.all$college.weight[i]=1.538038253
	else if (data.all$College[i]=="MCS") data.all$college.weight[i]=1.231293699
	else if (data.all$College[i]=="SCS") data.all$college.weight[i]=0.777416674
	else data.all$college.weight[i]=1.071644541
}
#Creates new fully weighted data set for full data
write.table(data.all,file="weighted_data_all.csv",sep=",")

#Adding Weights to Other Data Sets
library(Hmisc)
```

```
#Add class/college post-strat weight to outside county, outside city, in city, in county
ps.weight<-vector(length=nrow(out.county))
for (i in 1:length(ps.weight))
	for (j in 1:nrow(data.all)){
		if (out.county$ID[i]==data.all$ID[j]) out.county$ps.weight[i]=data.all$ps.weight[j]
	}

ps.weight<-vector(length=nrow(out.city))
for (i in 1:length(ps.weight))
	for (j in 1:nrow(data.all)){
		if (out.city$ID[i]==data.all$ID[j]) out.city$ps.weight[i]=data.all$ps.weight[j]
	}

ps.weight<-vector(length=nrow(in.county))
for (i in 1:length(ps.weight))
	for (j in 1:nrow(data.all)){
		if (in.county$ID[i]==data.all$ID[j]) in.county$ps.weight[i]=data.all$ps.weight[j]
	}

ps.weight<-vector(length=nrow(in.city))
for (i in 1:length(ps.weight))
	for (j in 1:nrow(data.all)){
		if (in.city$ID[i]==data.all$ID[j]) in.city$ps.weight[i]=data.all$ps.weight[j]
	}

#Add class weight to in county, in city, out county, out city
school.weight<-vector(length=nrow(in.county))
in.county<-cbind(in.county,school.weight)
for (i in 1:length(school.weight))
	for (j in 1:nrow(data.all)){
		if (in.county$ID[i]==data.all$ID[j]) in.county$school.weight[i]=data.all$school.weight[j]
	}

school.weight<-vector(length=nrow(in.city))
in.city<-cbind(in.city,school.weight)
for (i in 1:length(school.weight))
	for (j in 1:nrow(data.all)){
		if (in.city$ID[i]==data.all$ID[j]) in.city$school.weight[i]=data.all$school.weight[j]
	}

school.weight<-vector(length=nrow(out.county))
out.county<-cbind(out.county,school.weight)
for (i in 1:length(school.weight))
	for (j in 1:nrow(data.all)){
		if (out.county$ID[i]==data.all$ID[j]) out.county$school.weight[i]=data.all$school.weight[j]
	}

school.weight<-vector(length=nrow(out.city))
out.city<-cbind(out.city,school.weight)
for (i in 1:length(school.weight))
	for (j in 1:nrow(data.all)){
		if (out.city$ID[i]==data.all$ID[j]) out.city$school.weight[i]=data.all$school.weight[j]
	}

#Add college weight to in county, in city, out county, out city
college.weight<-vector(length=nrow(in.county))
```

```
in.county<-cbind(in.county,college.weight)
for (i in 1:length(college.weight))
        for (j in 1:nrow(data.all)){
                if (in.county$ID[i]==data.all$ID[j]) in.county$college.weight[i]=data.all$college.weight[j]
        }

college.weight<-vector(length=nrow(in.city))
in.city<-cbind(in.city,college.weight)
for (i in 1:length(college.weight))
        for (j in 1:nrow(data.all)){
                if (in.city$ID[i]==data.all$ID[j]) in.city$college.weight[i]=data.all$college.weight[j]
        }

college.weight<-vector(length=nrow(out.county))
out.county<-cbind(out.county,college.weight)
for (i in 1:length(college.weight))
        for (j in 1:nrow(data.all)){
                if (out.county$ID[i]==data.all$ID[j]) out.county$college.weight[i]=data.all$college.weight[j]
        }

college.weight<-vector(length=nrow(out.city))
out.city<-cbind(out.city,college.weight)
for (i in 1:length(college.weight))
        for (j in 1:nrow(data.all)){
                if (out.city$ID[i]==data.all$ID[j]) out.city$college.weight[i]=data.all$college.weight[j]
        }

#Creates new weighted data sets for in county, in city, out city, out county
write.table(in.county,file="weighted_in_county.csv",sep=",")
write.table(in.city,file="weighted_in_city.csv",sep=",")
write.table(out.city,file="weighted_out_city.csv",sep=",")
write.table(out.county,file="weighted_out_county.csv",sep=",")

#Weighted frequency tables
wtd.table(in.city$College,weights=in.city$college.weight)
wtd.table(in.city$Class,weights=in.city$school.weight)
wtd.table(in.city$Neighborhood,weights=in.city$ps.weight)
wtd.table(in.county$Class,weights=in.county$school.weight)
wtd.table(in.county$College,weights=in.county$college.weight)

wtd.table(out.city$College,weights=out.city$college.weight)
wtd.table(out.city$Class,weights=out.city$school.weight)
wtd.table(out.county$Class,weights=out.county$school.weight)
wtd.table(out.county$College,weights=out.county$college.weight)

wtd.table(data.all$College,weights=data.all$college.weight)
wtd.table(data.all$Class,weights=data.all$school.weight)
##Looking at Distances to Campus

data2<-read.table("WithinCity.csv",header=T,sep=",")
attach(data2)

undergrad2=data2[which(data2$Class<=5),]
masters2= data2[which(data2$Class==10),]
phd2=data2[which(data2$Class==20),]
```

```
mean(Distance.to.CMU[College=="CFA"])
mean(Distance.to.CMU[College=="CIT"])
mean(Distance.to.CMU[College=="CMU"])
mean(Distance.to.CMU[College=="HC"])
mean(Distance.to.CMU[College=="HSS"])
mean(Distance.to.CMU[College=="MCS"])
mean(Distance.to.CMU[College=="SCS"])
mean(Distance.to.CMU[College=="TSB"])

mean(Distance.to.Bar[College=="CFA"])
mean(Distance.to.Bar[College=="CIT"])
mean(Distance.to.Bar[College=="CMU"])
mean(Distance.to.Bar[College=="HC"])
mean(Distance.to.Bar[College=="HSS"])
mean(Distance.to.Bar[College=="MCS"])
mean(Distance.to.Bar[College=="SCS"])
mean(Distance.to.Bar[College=="TSB"])

mean(Distance.to.Supermarket[College=="CFA"])
mean(Distance.to.Supermarket[College=="CIT"])
mean(Distance.to.Supermarket[College=="CMU"])
mean(Distance.to.Supermarket[College=="HC"])
mean(Distance.to.Supermarket[College=="HSS"])
mean(Distance.to.Supermarket[College=="MCS"])
mean(Distance.to.Supermarket[College=="SCS"])
mean(Distance.to.Supermarket[College=="TSB"])

mean(Distance.to.Bus.Stops[College=="CFA"])
mean(Distance.to.Bus.Stops[College=="CIT"])
mean(Distance.to.Bus.Stops[College=="CMU"])
mean(Distance.to.Bus.Stops[College=="HC"])
mean(Distance.to.Bus.Stops[College=="HSS"])
mean(Distance.to.Bus.Stops[College=="MCS"])
mean(Distance.to.Bus.Stops[College=="SCS"])
mean(Distance.to.Bus.Stops[College=="TSB"])

mean(Distance.to.Shuttle.Stops[College=="CFA"])
mean(Distance.to.Shuttle.Stops[College=="CIT"])
mean(Distance.to.Shuttle.Stops[College=="CMU"])
mean(Distance.to.Shuttle.Stops[College=="HC"])
mean(Distance.to.Shuttle.Stops[College=="HSS"])
mean(Distance.to.Shuttle.Stops[College=="MCS"])
mean(Distance.to.Shuttle.Stops[College=="SCS"])
mean(Distance.to.Shuttle.Stops[College=="TSB"])

##Undergraduate vs. Masters (All students)
t.test(undergrad1[,4],masters1[,4],var.equal=FALSE, data=data1)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,4],masters2[,4],var.equal=FALSE, data=data2)

##Undergraduate vs. Masters (Within County)
t.test(undergrad3[,4],masters3[,4],var.equal=FALSE, data=data3)

##Masters vs. PhD  (All students)
t.test(phd1[,4],masters1[,4],var.equal=FALSE, data=data1)
```

```
##Masters vs. PhD (Within City)
t.test(phd2[,4],masters2[,4],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within County)
t.test(phd3[,4],masters3[,4],var.equal=FALSE, data=data3)

##Undergraduate vs. PhD (All students)
t.test(undergrad1[,4],phd1[,4],var.equal=FALSE, data=data1)

##Undergraduate vs. PhD (Within City)
t.test(undergrad2[,4],phd2[,4],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within County)
t.test(undergrad3[,4],phd3[,4],var.equal=FALSE, data=data3)

##a. Schools for all students (All students)
anova1=aov(Distance.to.CMU~College,data=data1)
summary(anova1)

##b. Schools for all students (Within City)
anova2=aov(Distance.to.CMU~College,data=data2)
summary(anova2)

##c. Schools for all students (Within County)
anova3=aov(Distance.to.CMU~College,data=data3)
summary(anova3)

##a. Schools within Undergrad (All Students)
anova4<-aov(undergrad1[,4]~undergrad1[,1],data=data1)
summary(anova4)

##b. Schools within Undergrad (Within City)
anova5<-aov(undergrad2[,4]~undergrad2[,1],data=data2)
summary(anova5)

##c. Schools within Undergrad (Within County)
anova6<-aov(undergrad3[,4]~undergrad3[,1],data=data3)

##a. Schools within Masters (All Students)
master.dist.1<-aov(Distance.to.CMU~College,data=masters1)
summary(master.dist.1)

##b. Schools within Masters (Within City)
master.dist.2<-aov(Distance.to.CMU~College,data=masters2)

##c. Schools within Masters (Within County)
master.dist.3<-aov(Distance.to.CMU~College,data=masters3)

##a. Schools within PhD (All Students)
anova7<-aov(phd1[,4]~phd1[,1],data=data1)
summary(anova7)

##b. Schools within PhD (Within City)
anova8<-aov(phd2[,4]~phd2[,1],data=data2)
summary(anova8)
```

```
##c. Schools within PhD (Within County)
anova9<-aov(phd3[,4]~phd3[,1],data = data3)
summary(anova9)

Class_Cat2<-vector(length=nrow(data2))
for (i in 1:length(Class_Cat2)){
	if (data2$Class[i]<6) Class_Cat2[i]=0
	else if (data2$Class[i]==10) Class_Cat2[i]=1
	else Class_Cat2[i]=2
}

##c. Classes (Within City)
anova10<-aov(data2[,4]~Class_Cat2)
summary(anova10)

##Looking at Distances to Bus Stops
##Undergraduate versus graduate (Within City)
t.test(Distance.to.Bus.Stops~undergraduate2,data=data2)
##Schools for all students (Within City)
newanova1=aov(Distance.to.Bus.Stops~College,data=data2)
summary(newanova2)
## Schools within Undergrad (Within City)
newanova2<-aov(undergrad2[,7]~undergrad2[,1],data=data2)
summary(newanova5)

##Schools within Masters (Within City)
newmaster.dist.2<-aov(data2[,7]~data2[,1],data=masters2)
summary(newmaster.dist.2)

##Schools within PhD (Within City)
newanova8<-aov(phd2[,7]~phd2[,1],data=without.cmu2)
summary(newanova8)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,7],masters2[,7],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within City)
t.test(masters2[,7],phd2[,7],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within City)
t.test(undergrad2[,7],phd2[,7],var.equal=FALSE, data=data2)

anova11<-aov(data2[,7]~Class_Cat2)
summary(anova11)
anova12=aov(data2[,7]~College,data=data2)
summary(anova12)


##Looking at Distances to Bars

##Undergraduate versus graduate (Within City)
t.test(Distance.to.Bar~undergraduate2,data=data2)

##Schools for all students (Within City)
newanova1=aov(Distance.to.Bar~College,data=data2)
```

```
summary(newanova2)

## Schools within Undergrad (Within City)
newanova2<-aov(undergrad2[,5]~undergrad2[,1],data=data2)
summary(newanova5)

##Schools within Masters (Within City)
newmaster.dist.2<-aov(data2[,5]~data2[,1],data=masters2)
summary(newmaster.dist.2)

##Schools within PhD (Within City)
newanova8<-aov(phd2[,5]~phd2[,1],data=without.cmu2)
summary(newanova8)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,5],masters2[,5],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within City)
t.test(masters2[,5],phd2[,5],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within City)
t.test(undergrad2[,5],phd2[,5],var.equal=FALSE, data=data2)

anova10<-aov(data2[,5]~Class_Cat2)
summary(anova10)
anova2=aov(Distance.to.Bar~College,data=data2)
summary(anova2)

##Looking at Distances to Supermarkets

##Undergraduate versus graduate (Within City)
t.test(Distance.to.Supermarket~undergraduate2,data=data2)

##Schools for all students (Within City)
newanova1=aov(Distance.to.Supermarket~College,data=data2)
summary(newanova2)

## Schools within Undergrad (Within City)
newanova2<-aov(undergrad2[,6]~undergrad2[,1],data=data2)
summary(newanova5)

##Schools within Masters (Within City)
newmaster.dist.2<-aov(data2[,6]~data2[,1],data=masters2)
summary(newmaster.dist.2)

##Schools within PhD (Within City)
newanova8<-aov(phd2[,6]~phd2[,1],data=without.cmu2)
summary(newanova8)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,6],masters2[,6],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within City)
t.test(masters2[,6],phd2[,6],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within City)
```

```
t.test(undergrad2[,6],phd2[,6],var.equal=FALSE, data=data2)

anova11<-aov(data2[,6]~Class_Cat2)
summary(anova11)
anova12=aov(data2[,6]~College,data=data2)
summary(anova12)


##Looking at Distances to Shuttle Stops

##Undergraduate versus graduate (Within City)
t.test(Distance.to.Shuttle.Stops~undergraduate2,data=data2)

##Schools for all students (Within City)
newanova1=aov(Distance.to.Shuttle.Stops~College,data=data2)
summary(newanova2)

## Schools within Undergrad (Within City)
newanova2<-aov(undergrad2[,8]~undergrad2[,1],data=data2)
summary(newanova5)

##Schools within Masters (Within City)
newmaster.dist.2<-aov(data2[,8]~data2[,1],data=masters2)
summary(newmaster.dist.2)

##Schools within PhD (Within City)
newanova8<-aov(phd2[,8]~phd2[,1],data=without.cmu2)
summary(newanova8)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,8],masters2[,8],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within City)
t.test(masters2[,8],phd2[,8],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within City)
t.test(undergrad2[,8],phd2[,8],var.equal=FALSE, data=data2)


anova13<-aov(data2[,8]~Class_Cat2)
summary(anova13)
anova14=aov(data2[,8]~College,data=data2)
summary(anova14)


data1<-read.table("All_Students.csv",header=T,sep=",")
attach(data1)
data2<-read.table("WithinCity.csv",header=T,sep=",")
attach(data2)
 data3<-read.table("WithinCounty.csv",header=T,sep=",")
attach(data3)

undergraduate1<-ifelse(data1$Class<=5,1,0)
undergraduate2<-ifelse(data2$Class<=5,1,0)
undergraduate3<-ifelse(data3$Class<=5,1,0)
```

```
undergrad1=data1[which(data1$Class<=5),]
masters1= data1[which(data1$Class==10),]
phd1=data1[which(data1$Class==20),]

undergrad2=data2[which(data2$Class<=5),]
masters2= data2[which(data2$Class==10),]
phd2=data2[which(data2$Class==20),]

undergrad3=data3[which(data3$Class<=5),]
masters3= data3[which(data3$Class==10),]
phd3=data3[which(data3$Class==20),]

#Creates categorical variable for Class (Undergrad vs Masters vs PhD)
Class_Cat1<-vector(length=nrow(data1))
for (i in 1:length(Class_Cat1)){
	if (data1$Class[i]<6) Class_Cat1[i]=0
	else if (data1$Class[i]==10) Class_Cat1[i]=1
	else Class_Cat1[i]=2
}
Class_Cat2<-vector(length=nrow(data2))
for (i in 1:length(Class_Cat2)){
	if (data2$Class[i]<6) Class_Cat2[i]=0
	else if (data2$Class[i]==10) Class_Cat2[i]=1
	else Class_Cat2[i]=2
}
Class_Cat3<-vector(length=nrow(data3))
for (i in 1:length(Class_Cat3)){
	if (data3$Class[i]<6) Class_Cat3[i]=0
	else if (data3$Class[i]==10) Class_Cat3[i]=1
	else Class_Cat3[i]=2
}

Class_Cat1<-vector(length=nrow(data1))
for (i in 1:length(Class_Cat1)){
	if (data1$Class[i]<6) Class_Cat1[i]=0
	else if (data1$Class[i]==10) Class_Cat1[i]=1
	else Class_Cat1[i]=2
}
Class_Cat2<-vector(length=nrow(data2))
for (i in 1:length(Class_Cat2)){
	if (data2$Class[i]<6) Class_Cat2[i]=0
	else if (data2$Class[i]==10) Class_Cat2[i]=1
	else Class_Cat2[i]=2
}
Class_Cat3<-vector(length=nrow(data3))
for (i in 1:length(Class_Cat3)){
	if (data3$Class[i]<6) Class_Cat3[i]=0
	else if (data3$Class[i]==10) Class_Cat3[i]=1
	else Class_Cat3[i]=2
}


table(College,Class)
##Looking at Distances to Campus
##1.)
##a. Undergraduate versus graduate (All students)
```

```
t.test(Distance.to.CMU~undergraduate1,data=data1)
p-value = 0.03347
mean in group 0 = 1.0764488       mean in group 1 = 0.8846147

##b. Undergraduate versus graduate (Within City)
t.test(Distance.to.CMU~undergraduate2,data=data2)
p-value < 2.2e-16

##c. Undergraduate versus graduate (Within County)
t.test(Distance.to.CMU~undergraduate3,data=data3)
p-value = 6.249e-06

##2.)
##a. Schools for all students (All students)
anova1=aov(Distance.to.CMU~College,data=data1)
summary(anova1)
p-value = 0.04514

##b. Schools for all students (Within City)
anova2=aov(Distance.to.CMU~College,data=data2)
summary(anova2)
p-value = 0.0188

##c. Schools for all students (Within County)
anova3=aov(Distance.to.CMU~College,data=data3)
summary(anova3)
p-value = 0.0001565

##3.)
##a. Schools within Undergrad (All Students)
##Remove Heinz College
College.new<-data1$College[data1$College!="HC"]
anova4<-aov(undergrad1[,4]~undergrad1[,1],data=College.new)
summary(anova4)
#p-value = 0.605
anova4<-aov(undergrad1[,4]~undergrad1[,1],data=data1)
#p-value = 0.6058

##b. Schools within Undergrad (Within City)
College.new2<-data2$College[data2$College!="HC"]
anova5<-aov(undergrad2[,4]~undergrad2[,1],data=College.new2)
summary(anova5)
#p-value = 0.9129
anova6<-aov(undergrad2[,4]~undergrad2[,1],data=data2)
summary(anova6)
#p-value = 0.6058

##c. Schools within Undergrad (Within County)

##4.)
##a. Schools within Masters (All Students)
master.dist.1<-aov(Distance.to.CMU~College,data=masters1)
summary(master.dist.1)
p-value = 8.836e-05

##b. Schools within Masters (Within City)
```

```
master.dist.2<-aov(Distance.to.CMU~College,data=masters2)
summary(master.dist.2)
p-value = 0.02216

##c. Schools within Masters (Within County)
master.dist.3<-aov(Distance.to.CMU~College,data=masters3)
summary(master.dist.3)
p-value = 0.5026

##5.)
without.cmu1<-data1$College[data1$College!="CMU"]
##a. Schools within PhD (All Students)
anova7<-aov(phd1[,4]~phd1[,1],data=data1)
summary(anova7)
p-value = 0.3806
##b. Schools within PhD (Within City)
without.cmu2<-data2$College[data2$College!="CMU"]
anova8<-aov(phd2[,4]~phd2[,1],data=data2)
summary(anova8)
p-value = 0.04486
##c. Schools within PhD (Within County)
without.cmu3<-data3$College[data3$College!="CMU"]
anova9<-aov(phd3[,4]~phd3[,1],data = data3)
summary(anova9)
p-value = 0.1607
##6

##Undergraduate vs. Masters (All students)
t.test(undergrad1[,4],masters1[,4],var.equal=FALSE, data=data1)

##Undergraduate vs. Masters (Within City)
t.test(undergrad2[,4],masters2[,4],var.equal=FALSE, data=data2)

##Undergraduate vs. Masters (Within County)
t.test(undergrad3[,4],masters3[,4],var.equal=FALSE, data=data3)

##Masters vs. PhD  (All students)
t.test(undergrad1[,4],masters1[,4],var.equal=FALSE, data=data1)

##Masters vs. PhD (Within City)
t.test(undergrad2[,4],masters2[,4],var.equal=FALSE, data=data2)

##Masters vs. PhD (Within County)
t.test(undergrad3[,4],masters3[,4],var.equal=FALSE, data=data3)

##Undergraduate vs. PhD (All students)
t.test(undergrad1[,4],masters1[,4],var.equal=FALSE, data=data1)

##Undergraduate vs. PhD (Within City)
t.test(undergrad2[,4],masters2[,4],var.equal=FALSE, data=data2)

##Undergraduate vs. PhD (Within County)
t.test(undergrad3[,4],masters3[,4],var.equal=FALSE, data=data3)


##Looking at Distance to Bus Stops
```

```
##1.)
##a. Undergraduate versus graduate (Within City)
t.test(Distance.to.Bus.Stops~undergraduate2,data=data2)
p-value = < 2.2e-16
mean in group 0 = 0.08447041 mean in group 1 = 0.05907557

##2.)
##b. Schools for all students (Within City)
newanova2=aov(Distance.to.Bus.Stops~College,data=data2)
summary(newanova2)
p-value = 0.9292

##3.)
##b. Schools within Undergrad (Within City)
newanova5<-aov(undergrad2[,7]~undergrad2[,1],data=College.new2)
summary(newanova5)
#p-value = 0.9129
newanova6<-aov(undergrad2[,7]~undergrad2[,1],data=data2)
summary(newanova6)

##4.)

##b. Schools within Masters (Within City)
newmaster.dist.2<-aov(data2[,7]~data2[,1],data=masters2)
summary(newmaster.dist.2)
p-value = 0.9292

##5.)
##b. Schools within PhD (Within City)
without.cmu2<-data2$College[data2$College!="CMU"]
newanova8<-aov(phd2[,7]~phd2[,1],data=without.cmu2)
summary(newanova8)
p-value = 0.04486

##6
a. Masters vs. Undergraduate  (All students)
<-aov(Distance.to.CMU~Class_Cat1,data=data1)
p = 0.002314

b. Masters vs. Undergraduate (Within City)
```