

very nice job

a somewhat prosaic topic but well pursued with interesting results.

nice survey questions, great execution of survey

very nice analysis and writeup

-BJ

Pizza Consumption Behaviors of CMU Undergraduates

Jacob Park • Stafford Brunk • Tony Poor • Atishe Chordia • Henry Wu

36-303: Sampling, Survey, & Society

April 29, 2009

Brian Junker

Table of Contents thx!

I. Introduction	2
• Introduction	
• Research & Prior Work	
• Summary of Results	
II. Methods	4
• Target Population & Sampling Frame	
• Sampling Methodology	
• Sample Questions	
• Post-Survey Adjustments	
III. Research Findings	6
• Most Important Factor in Ordering Pizza	
• Important Times of Business	
• How Important are Price, Freshness, and Business Hours?	
• Comparison of Two Most Popular Pizzerias	
IV. Discussion & Possible Flaws	15
References	16
Appendix A: Full Questionnaire	17
Appendix B: Margin of Error Calculations	25
Appendix C: Post-stratification Weights	26
Appendix D: Mosaic Plots for Analysis of Independence	27
Appendix E: Frequency Tables for Comparison of Pizzerias	31

I. Introduction

The overarching question that this project focuses on is primarily “what characteristics of pizzas and pizza shops appeal to the undergraduate students enrolled in the Pittsburgh campus of Carnegie Mellon University?”

The above question is interesting due to its relevance to a large number of consumers on CMU’s campus. College students in particular would find the results intriguing for several reasons: first, pizza is a traditionally heavy part of their cuisine, and second, many of them are non-local and thus unfamiliar with the wide selection of local pizza restaurants. Our research question could potentially enable us to find some interesting outcomes that we can integrate with CMU’s dining plan by suggesting an on campus dining location or allowing students to use dineX or PlaidCa\$h to make it more convenient. Creating flexibility in the dining system either by adding an extra option with a delivery capability or just by being able to serve at later hours pizza shops will certainly be a popular idea among the student body.

Moreover, there are a tremendous number of pizza options in Pittsburgh, and it is interesting to study why certain pizza places do better than the others and why certain pizza choices are popular among students. As an additional side effect, the survey could benefit small, high-quality pizza places that don't have the name brand to compete with recognized national chains. Pizza shops can clearly benefit from the results from our study by reorganizing to capture a greater market share, increase market value in order to gain long term revenues.

vendors and other food service providers.

The third bracket of stakeholders includes ~~a mix of options~~. Firstly, local food vendors, the trucks, hot dog stand, Tartan’s Pavilion and so on will want to understand why the off campus pizza options might be more popular. CMU’s dining services might want to add another pizza option on to the campus system after this study, and hence the statistical analyses and survey data will be essential in building their case. As expected, we are seeing a large response rate, perhaps indicating that this topic is very popular amongst students.

Research and Prior Work

Several studies and surveys conducted by other organizations are very relevant to this work. For instance, the Zagat Survey displays a comprehensive list of restaurants and allows the respondent to vote on any number of them, giving any restaurant a 4-point score for food, décor, and service (Zagat Products, n.d.). It allows respondents to enter optional information about price and a free-form “comments” field, while offering a large reward for “witty” and quote-worthy comments. These optional fields are likely used as additional information when Zagat releases their retail restaurant guides, rather than in the comparative ranking of restaurants. The criteria that Zagat uses to rank restaurants are interesting, but the survey makes no attempt to determine which of the criteria is actually most important to the respondent. Several online food review sites, such as Yelp and RateItAll, also offer mechanisms for reviewing and rating pizza businesses (Pittsburgh Pizza Search Results, n.d.; Pittsburgh Pizza, n.d.). However, like Zagat, the focus of these sites is to provide a score or rating to an individual restaurant; actual opinions on what makes a restaurant ‘good’ is secondary (coming only in the form of written reviews) and very difficult to quantify.

Meanwhile, other local pizza-ranking surveys have been carried out, including one reported by the Pittsburgh Post Gazette in early 2009 (Jones, 2009). Although the study seems far from statistically rigorous, the article suggests an interesting follow-up question of “how close to your home is the pizza you eat most?” This suggests that proximity may be a factor in people’s preferences.

neat

Summary of Results

Our survey concluded that issues of pizza quality (including factors like freshness of ingredients, heat, etc.) matter the most to students, outweighing factors such as quantity and proximity of location. Most surprisingly to us, we also discovered that students don’t perceive business hours to be extremely important. Furthermore, we compared responses in an attempt to determine what factors cause some students to prefer one popular pizzeria (Papa John’s) over another (Vocelli’s) and found a statistically significant difference that showed that those who preferred Papa John’s were much more concerned about quality issues (like freshness and heat of pizza) than those who preferred Vocelli’s.

II. Methods

Target Population & Sampling Frame

In order to make inferences about a target population consisting of all freshmen, sophomores, juniors, and seniors currently studying in CMU's main campus in Pittsburgh, we drew from a sampling frame that included all Carnegie Mellon University undergraduate students from the Pittsburgh campus who are listed in the online CMU student directory. Note that this includes students (such as freshmen) who are on the dining plan, allowing us to see the potential impact of dineXtra and PlaidCa\$h on consumption behaviors. We are also keen to make possible inferences about differences between subgroups of our population; for instance, how does the most popular pizza choice or the reasons for dissatisfaction vary with major or year?

In order to achieve a margin of error that was about 6.2%, we needed a sample size of 240 people. Since our survey was going out via e-mail, we were also estimating an approximate 30% response rate. This meant that we needed to contact 800 people in order to ensure we reached our desired sample size. (see **Appendix B** for the full calculation)

Sampling Methodology

The sampling was conducted in the following manner:

1. A custom Ruby script was created that enabled us to download the Andrew IDs for every undergraduate that resides on the Pittsburgh campus (Brunk, 2010). These Andrew IDs were randomly sampled. 800 Andrew IDs were sampled and stored in a list. The Andrew IDs were checked to ensure none of them were the Andrew IDs of the researchers.
2. Each Andrew ID in the sample was loaded and hashed with a SHA-1 hash. This hash enables each individual Andrew ID to be mapped to a unique ID but also allows us to go back later to recontact. Since this mapping always happens within software, it is purposefully difficult for researchers themselves to immediately associate a response with a single person.
3. Once a SHA-1 hash had been calculated, each Andrew ID was emailed with a custom link to the survey.

4. Upon the need to recontact, the list of SHA-1 hashes that have answered will be read in. Within software, each Andrew ID in the sample will have its SHA-1 hash recalculated. If that SHA-1 has already taken the survey, it will not be recontacted.

After two waves of emails, we received a response rate of 244 out of the 800 Carnegie Mellon undergraduates that we sampled. Of these, 234 fit our criteria for inclusion into the sampling frame (namely, were undergraduate students of one of the six major Carnegie Mellon schools and were currently studying in the Pittsburgh campus), providing us with a final response rate of 29.25% and a margin of error of 6.27% (see **Appendix B**).

Sample Questions

A full glimpse of our questionnaire (with response summaries) is included in **Appendix A**. For a sample, examine question #16 (“What time of day do you typically order pizza?”) and #6 (“How many times per week do you eat food from a campus dining restaurant?”). The former question involves finding out what time pizzas are approximately ordered. If the major set of responses are in the “odd hours” category (11pm-10am), then we can be sure that there is huge potential for the pizza places to get on campus, since there are not a lot of options available at late hours at night – especially those that deliver. The second question aims at understanding what percentage of the undergraduates in our sampling frame eat frequently on campus and, if so, how frequently.

Post-survey Adjustments

After respondents provided data, we post-stratified based on class and gender to ensure inferences that were statistically representative of the subgroups in the target population. The weights for responses in each individual strata is displayed in **Appendix C**.

III. Research Findings

Most Important Factor in Ordering Pizza

We originally hypothesized that college students would be overly concerned with two variables: the price and quantity of pizza. The true results, however, showed this hypothesis to be only partially correct. As can be seen in Table 1, price and quality of pizza were by far the two factors that were rated as the most important to college students. In fact, 43% of the respondents rated quality to be most important (where quality was factors like freshness, ingredient quality, etc) while 42% of respondents felt price was the most important. This was somewhat surprising and showed that college students do care about the quality of food they are eating. Interestingly, a somewhat popular open-ended response (in the “other” category) expressed desire for dineXtra or PlaidCa\$h options.

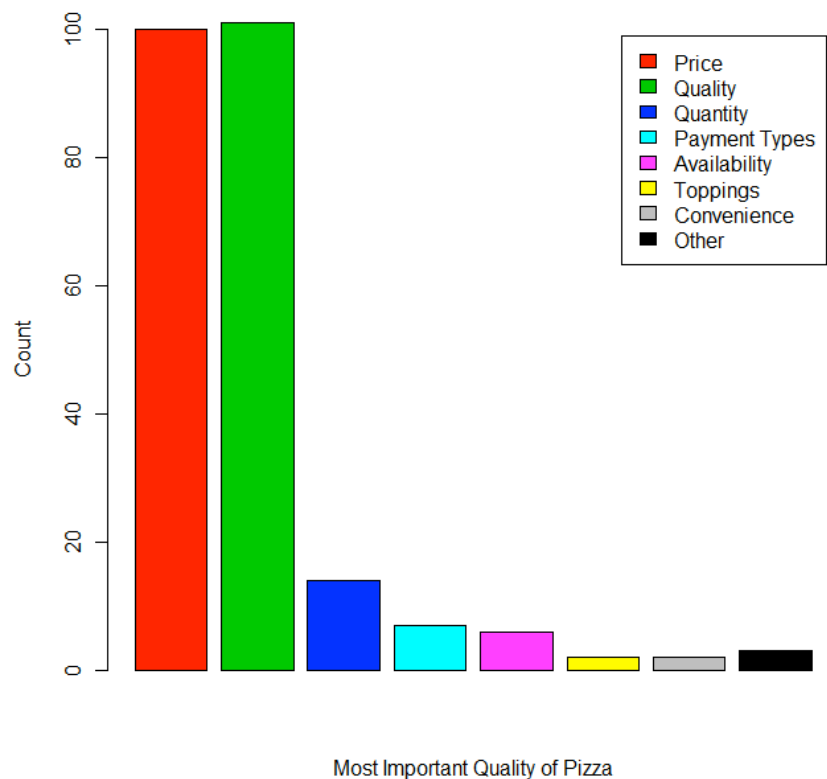


Table 1: The Most Important Factors in Ordering Pizza

Pizza Consumption Patterns

Another of our major goals was to uncover what the most popular times and days of week are for students ordering pizza; for instance, do students order pizza at 4am, when most businesses are closed?

The categorical variable measuring which day(s) of a week subjects typically order their pizza was used to create a new observation for subjects —a variable indicative of how many days of a week subjects typically order their pizza. Transforming this categorical variable into a continuous variable made it possible to create meaningful analyses of pizza consumption trends and patterns. The limitations of this creation of a new continuous variable are stated under **IV. Discussions & Potential Flaws.**

It is clear from Table 2 and Table 3 below that pizzerias receive the most orders on Fridays, and that the overwhelming majority of undergraduate students at CMU order their pizza from 6pm to 4am. Particularly, it is interesting that Monday and Tuesday have the lowest level of pizza orders placed throughout the week, as it they are the start to a new week, and as many groups work throughout the afternoon and evening on those days, easily covering the timeframes from 6pm to 4am. However, one should not forget that most campus events occur throughout the later part of the week, and rarely on Mondays.

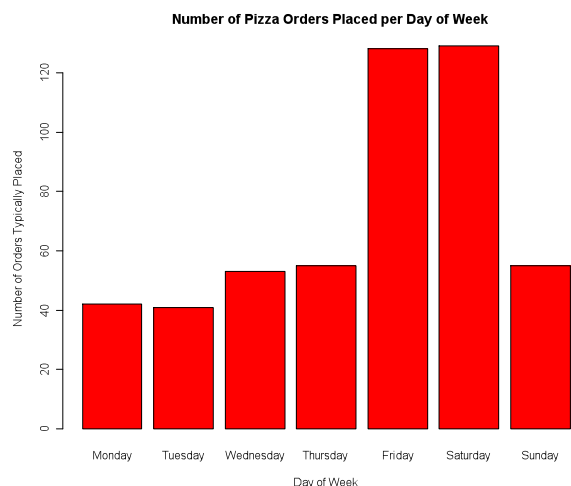


Table 2: Number of Pizza Orders per Day

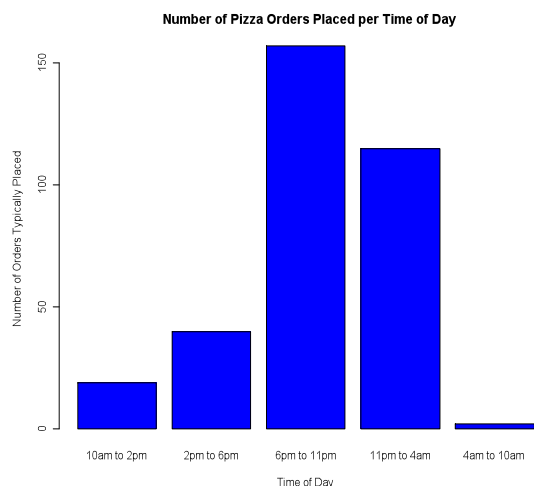


Table 3: Number of Orders per Time of Day

I'd still like to see these combined into a 2-d heat map or 3-d bar chart where the X axis is day of the week, the Y axis is time of day, and the Z axis is count or percent of responses.

Another analysis compared the number of pizza orders placed by the number of on-campus dining experience throughout the week. As shown by the Tables 4 and 5 below, the results vary greatly on whether or not the post-stratification weights are applied to the data. While the mean number of days subjects ordered pizza throughout the week varied very little for the weighted and unweighted data, the third quantiles are very different. Also, there are many more outliers in the weighted data than the unweighted data, and the overall boxplot illustrates that pizza consumptions accounts for a substantial amount of CMU undergraduate students' diet, whether they eat frequently on-campus or off-campus.

While the values greater than 7 in the boxplots are hard to interpret given that there are only seven days in a week, the weighting was necessary to have a more representative data. Lastly, it should be noted that there is an error in overlapping variables ("6 to 10" and "10 to 15" category) that is further explained under **IV. Discussions & Potential Flaws**.

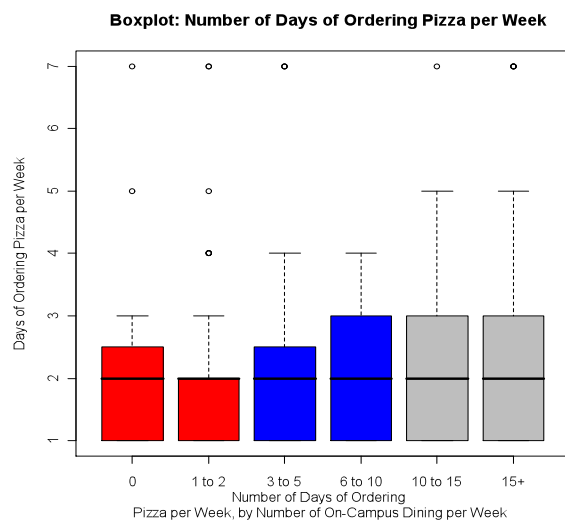


Table 4: Number of Days Subject Ordered Pizza

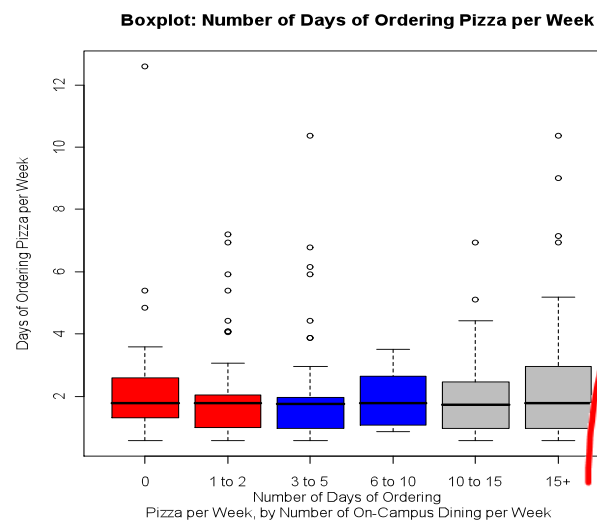


Table 5: Weighted Number of Days Subjects Ordered Pizza

if the max value of the unweighted data was 7 days, then the max value of the weighted data should be 7 days also. perhaps there was just a problem in creating the weighted boxplots.

Check for Relationships Among Categorical Pizza Variables

As it is apparent from our data, the pizza ingredient's freshness and whether the pizza is delivered hot are two variables that influence students' pizza consumption greatly. In further analyzing the variables, it is hypothesized that the importance of pizza's freshness to subjects is dependent on the importance of pizza being delivered hot. Also, it is suspected that the variables measuring the effect

of pizzeria's prices and proximity are independent, that the effect of pizzeria's prices is not dependent upon the importance of pizza being delivered hot, and lastly, that the effect of the pizzeria's prices are not dependent on the importance of the pizza's freshness.

From Table 6 below, the mosaic plot shows that the initial hypothesis is correct — that the importance of pizza's freshness to subjects is dependent on the importance of pizza being delivered hot. While it is apparent that the pizza being hot and the pizza being fresh are important to most of the respondents, it is also shown that the two categorical variables are dependent, shown by the shades on different boxes. (Please see **Appendix D** for an explanation of the mosaic plot alongside a full set of such plots used in our analyses.)

Similarly, as hypothesized, the effect of pizzeria's proximity, importance of pizza being delivered hot, and importance of pizza's freshness are all (separately) independent from the effect of pizzeria's prices on subjects' consumption.

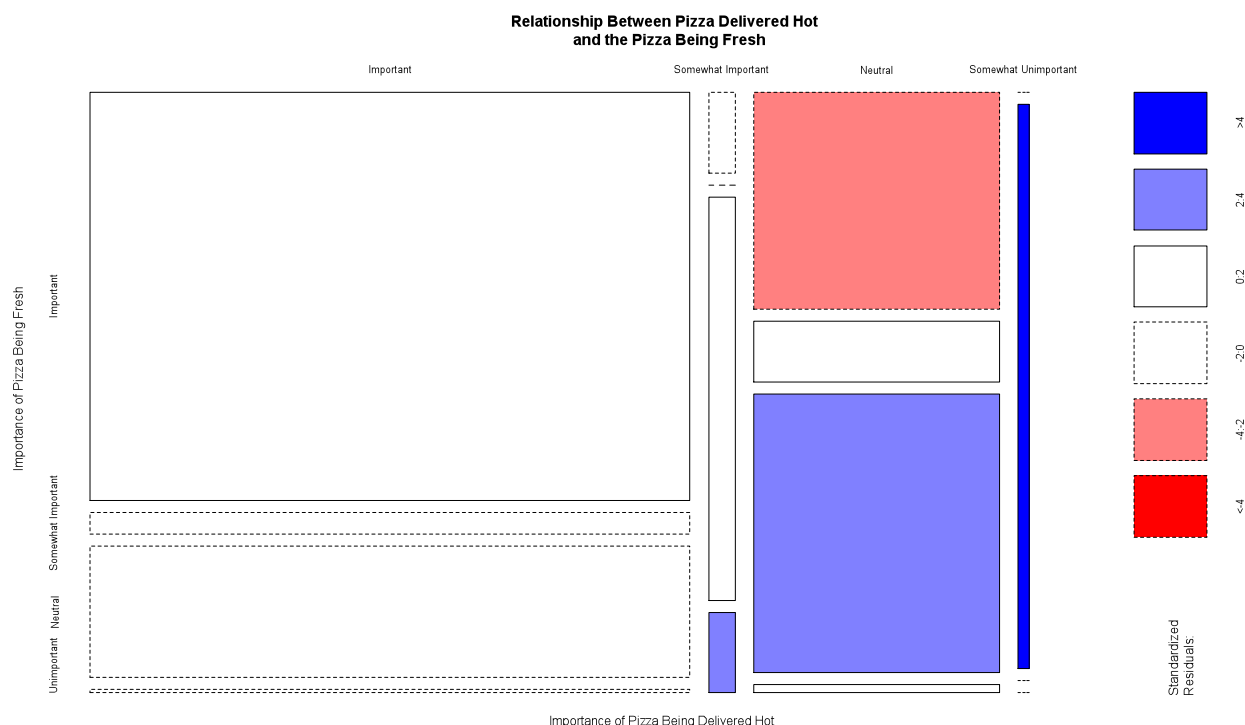


Table 6: Mosaic Plot of Variables for Pizza's Freshness vs. Pizza Delivered Hot

I would call this a figure, but no sweat.

Similarly, the other hypotheses are supported. The mosaic plots in **Appendix D** are mostly white, implying that there are very few cases of deviations from the expected under independence of the two variables. These pairs of variables are independent.

How Important are Price, Freshness, and Business Hours?

In order to ascertain the various levels of importance students held for individual aspects, we asked students to rank price, freshness, and the business hours of the pizzeria on a scale ranging from “None” or “Never” to “Completely” or “Always”. This scale had a total of five levels and was coded after data collection to correspond to the numbers 1-5. The stratified weights were then applied to the coded numbers and a mean for each category was calculated. Tables 7, 8, and 9 show the distribution of these three variables.

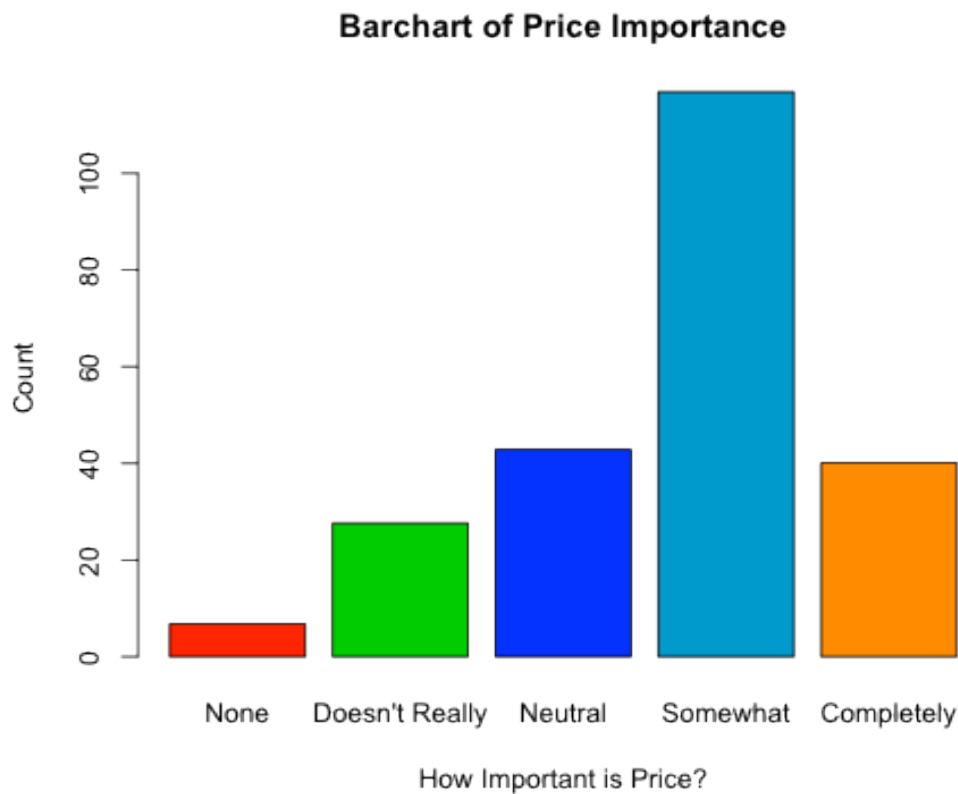


Table 7: Perceived Importance of Price

As can be seen in Table 7, a large majority of the respondents felt that price was only somewhat important. In fact, the weighted mean for this question was 3.67. This would correspond to about the “Somewhat” category in the chart. This result is a bit surprising considering the strong showing for the importance of price in Table 0. Recalling that 42% of respondents felt that price was the most important factor in ordering pizza, the fact that the clear majority of respondents felt that pizza price was only somewhat important is interesting.

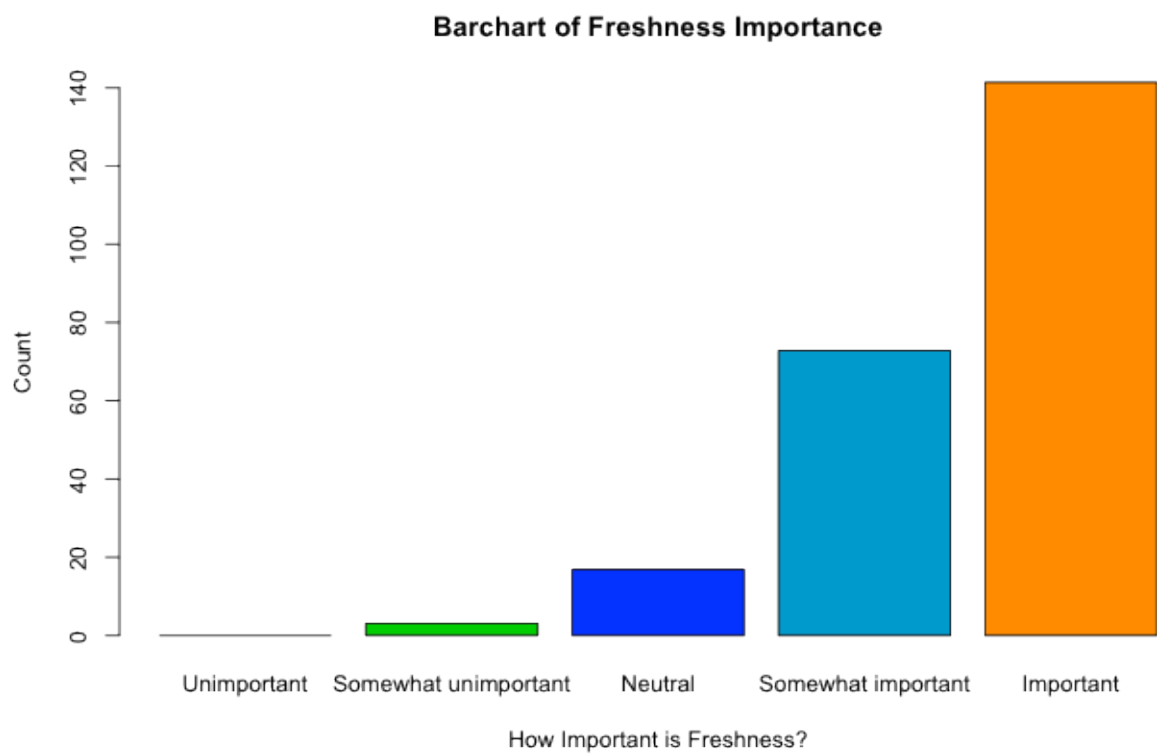


Table 8: Perceived Importance of Freshness

Table 8 shows that pizza freshness is held to be important to the majority of respondents. The weighted mean for this question was 4.51, which corresponds to just over halfway between the “Somewhat Important” and “Important” categories. This makes sense, since a hot, fresh pizza is considerably more enjoyable than reheated or old pizza. This also falls in line with our previous finding that 43% of respondents consider quality of their pizza to be the most important factor of ordering.

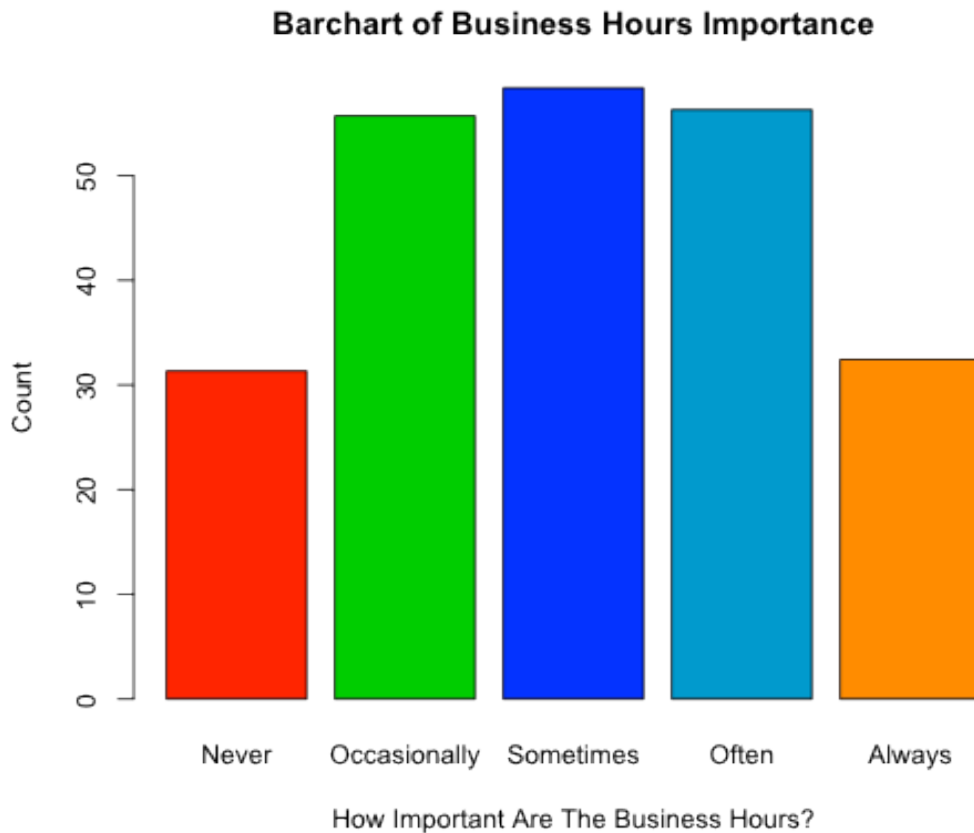


Table 8: Perceived Importance of Business Hours

Finally, in Table 8 we can see that people find the business hours of a pizzeria to be important “Occasionally” or “Sometimes” the most often. The weighted mean for this particular question was 3.01 which signals that the “Sometimes” category is firmly in the center for this particular variable. This is perhaps an expected result, since many college students order from a pizzeria that is open extremely late (for example, Papa John’s) and as such do not have to worry about the hours. In any case, the mean answer for this question is very middle of the road and subsequently shows that the hours of a pizzeria don’t seem to have a large influence on the ordering decision.

Comparing the Two Most Popular Pizzerias

We suspect that there are large variations in the answer choices for various pizzeria qualities, and wish to test whether these differences are statistically significant or not. We assign numerical values for each of the explanatory variables. Importance of price, importance of proximity, and similar

questions have ordered answer choices, and thus are coded with 0 being the bottom most category (never, no importance, etc), and $n - 1$ assigned to the highest category (daily, always, important, etc). The most stable value in unordered variables is assigned to the reference group and given a value of 0, while the other answer choices are given no preferential treatment. For the pizzeria variable, 0 is assigned to Vocelli's Pizza and 1 is given to Papa John's. Over twenty other pizzerias received at least one vote, but none of them enough observations for us to perform meaningful statistical analyses.

We perform weighted chi-square tests using the null hypothesis that the weighted mean of the variable for those who chose Vocelli's is equal to the weighted mean for the Papa John's group, and the alternative hypothesis that the means are unequal. We assign weights based on the breakdown of our sample by gender and class relative to the actual proportions of Carnegie Mellon's undergraduate population. We wrote a loop in R to generate weighted frequency tables for each variable by summing the weights, conditioning on the two pizzeria groups and the variable to be tested. The p-values of the chi-square tests of those frequency tables are displayed below, and the weighted frequency tables of each variable can be found in **Appendix E**.

Variable	p-value
frequency of eating pizza	0.2766
<u>preference type</u>	0.0516
most important factor	0.6217
business hour importance	0.1919
price importance	0.1137
proximity importance	0.6843
acceptable delivery time	0.4105
importance of hot pizza	0.0162
freshness importance	0.0235

Table 9: P-values of Weighted Chi-Square Tests of Vocelli's vs. Papa John's

The weighted chi-square test produced a p-value lower than 0.05 for importance of hot pizza and importance of freshness. We conclude that there is a significant difference in weighted means in the Vocelli's group and Papa John's group for those two variables. The other variables have p-values greater than 0.05, so the weighted means of the two groups for those variables are not significantly different. However, preference type has a p-value just barely above 0.05, so we do not completely dismiss the possibility of preference types having some differences. Most important factor in a

what is 'preference type'? a more descriptive variable name might help (or just remind reader what it is!)

pizzeria and importance of proximity had large enough p-values that we do not consider any possibility them having different weighted means for the two groups.

IV. Discussion & Potential Flaws

Interestingly, our response rate very quickly reached incredibly high numbers; our email-based survey reached over 30% with only one follow-up, easily surpassing the rates observed by our peers conducting similar surveys. We believe that this is due to several factors. First, pizza and the potential for better food choices is a topic of considerable interest to students, and thus our high response rate shows that our survey is indeed very relevant to the student body. Second, our raffle of a gift certificate likely propelled responses.

However, the gift certificate approach is likely not without flaw. It remains to be seen whether the offer of a \$25 Amazon gift certificate significantly biased our respondents towards those who were more concerned about monetary issues. Compounding this issue is the smaller one of question wording: after carrying out the survey, we noticed that one of our questions involved overlapping ranges, with one response for 6–10 and another for 10–15 (see Question #13 in **Appendix A**).

Lastly, there are limitations associated with creating a new continuous variable from the categorical variable measuring which days of the week respondents typically order their pizza. It is possible that the respondents did not consider this from a strictly weekly perspective — it is possible and valid for subjects to report Monday and Tuesday as the days they typically order pizza because they might be recalling from their memories that go back a few weeks. However, the transformation or creation of the new variable allowed for a further analysis of the data that would not have been possible.

Despite these flaws, however, we believe that our findings are still statistically valid as a whole; after all, even if there was a bias towards financially-minded respondents, we still found that other qualities (such as freshness) were perceived as more important. Furthermore, our development of a set of scripts for sampling was an incredibly useful byproduct of our work. A major strength of our study is the ease with which we disseminated our surveys and the use of hashing as a primitive encoding mechanism for privacy. Computerized methods helped us reach out extremely quickly to a larger sample frame, and our scripts have been provided for use by future statistics students.

References

Brunk, Stafford. (April 27, 2010). *Ruby survey sampling tools*. Retrieved April 28, 2010, from github website, <http://github.com/wingrunr21/CMU-Survey-Tools>

Jones, Diana Nelson. (February 25, 2009). *This just in: East End pizza survey's top vote getters revealed*. Retrieved April 28, 2010, from Pittsburgh Post Gazette website, <http://community.post-gazette.com/blogs/citywalkabout/archive/2009/02/25/this-just-in-east-end-pizza-survey-s-top-vote-getters-revealed.aspx>

Pittsburgh Pizza. (n.d.) Retrieved April 28, 2010, from RateItAll website, <http://www.rateitall.com/t-10945-pittsburgh-pizza.aspx>

Pittsburgh Pizza Search Results. (n.d.) Retrieved April 28, 2010 from Yelp website, http://www.yelp.com/search?find_desc=Pizza&ns=1&find_loc=Pittsburgh%2C+PA

Zagat Products. (n.d.) Retrieved April 28, 2010 from Zagat website, <http://www.zagat.com/About/Index.aspx?menu=productsAndServices>

Appendix A: Questionnaire & Answer Summaries

1. What is your age?

There was a roughly even distribution of respondents between the ages of 18 to 21, which is the typical age range for freshmen, sophomores, and juniors. The peak groups were located at ages 19 and 20, with 58 and 59 people selecting them, respectively. Only 42 people in the sample were 18 years of age, which was expected since fewer freshmen are 18 years old during the latter stages of the school year (when we conducted the survey). Only 18 people selected age 22, and 5 were 23 years or older.

2. What is your gender?

- a. Male
- b. Female

There were 127 males and 107 females in our survey. 54.3% were male, which is lower than the percentage of males in Carnegie Mellon's undergraduate population, which sits at 64%.

3. What is your year in school?

- a. Freshman
- b. Sophomore
- c. Junior
- d. Senior

There were 68 freshmen, 55 sophomores, 71 juniors, and 40 seniors in our survey. We expected a large number of freshmen, since they often have excess DineXtra dollars, and one of the area pizzerias, Vocelli's, accepts DineXtra. There was a high spike in the number of junior respondents. Seniors made up the smallest class, may be due to their decreased interest in the topic, since many will graduate and leave Pittsburgh before the results of our analysis makes an impact on local pizzerias.

4. What is your primary school?

- a. CIT
- b. CFA

- c. HSS
- d. TSB
- e. MCS
- f. SCS
- g. Heinz
- h. Other: _____

CIT, with 75 respondents, was the most represented school in our survey, and it was also the largest school in terms of total undergraduate majors. HSS was the next highest total, with 45 people, while CFA came in third with 38. TSB, with the lowest total undergraduate enrollment, also had the lowest representation with only 21 taking the survey. The percentages among each school in our survey fell mostly in line with the breakdown of the undergraduate population, except for SCS, which only had one person more than TSB, but made up a much larger percentage of the undergraduate population.

5. Are you currently studying at Carnegie Mellon's Pittsburgh campus?

Our e-mail script filtered out all andrew IDs of undergraduates who were not enrolled in the Pittsburgh campus, so everyone taking the survey responded "yes".

6. How many times per week do you eat food from a campus dining restaurant (Entropy, trucks, hot dog stand, all other dining locations included)?

- a. 0
- b. 1-2
- c. 3-5
- d. 6-10
- e. 10-15
- f. 15+

The responses to this question had a bimodal distribution, with the most popular response being 1-2 times per week (59 people). The frequency dipped in the 3-5 and 6-10 times per week categories before rising at the second peak at 10-15 times per week (49 people). 18 people do not eat at campus dining locations, and 28 people eat on campus more than 15 times per week.

7. How often do you eat pizza?

- a. Daily
- b. Weekly
- c. Monthly
- d. Annually
- e. Never

these paragraphs are extremely nice and a welcome way to describe results question by question. Might consider augmenting with a table also.

Over half (124) of the 234 respondents said they ate pizza weekly. 98 people ate pizza on a monthly basis. Very few people ate at greater than a weekly rate or less frequently than once per month. Only one person who took the survey said they never ate pizza. We suspected that many of the people e-mailed who did not like pizza chose not to respond to our survey.

8. What kind of pizza do you like the most?

- a. Thin crust
- b. Thick crust
- c. Deep dish
- d. Sicilian
- e. Other: _____

Thin crust came in first with 105 respondents, while 81 preferred thick crust. 31 people liked deep dish pizza the most, but only 6 voted for Sicilian. There were 10 different answers for the 11 people who entered a write in response for “other”. Among those, only stuffed crust received more than 1 vote.

9. What methods do you use to order pizza?

- a. Internet
- b. Telephone
- c. Text
- d. Walk-In Order
- e. Other: _____

The survey takers were allowed to choose as many answers that applied to them. 55 people who only selected one answer choice answered telephone, which was the most popular single response. The most common combination was internet and telephone, where 46 people selected both a and b. Only 12 people selected three answer choices, 10 of which chose internet + telephone + walk-in order.

122 people mentioned the internet as one of their answer choices, while 106 people selected telephone as one of their options.

10. When you order pizza, what factor influences your order the most?

- a. Price
- b. Quantity
- c. Quality
- d. Other: _____

Quality and price were by far the most popular answer choices, with 101 and 100 respectively saying that those were the most important factors. We expected college students to have a stronger attitude towards value and getting the most food for their money, but only 14 said quantity was most important when ordering pizza.

11. When you order pizza, how often do you order from the same pizzeria?

- a. Always
- b. Often
- c. Sometimes
- d. Occasionally
- e. Never

The people in the sample had a general affinity for ordering from the same places. 96 said they always ordered from the same pizzeria, and 109 responded that they often do it. 4 people responded that they never ordered from the same location, while 1 person told us they never order pizza and always ate on campus instead.

12. How much does the proximity of a pizzeria affect your decision to order from it?

- a. Completely
- b. Somewhat
- c. Neutral
- d. Doesn't really
- e. None

Proximity was not too important of a factor. 77 people said that proximity somewhat affected their decision of where to order from, 70 responded that it didn't really matter, and 68 said it had neutral or no effect at all. Only 19 people felt that proximity mattered completely.

13. What delivery time is acceptable to you?

- a. Less than 15 minutes
- b. 15-30 minutes
- c. 31-45 minutes
- d. 46-60 minutes
- e. More than 1 hour (don't really care as long as it comes)

A majority (126 people) told us they could accept delivery times between 31 to 45 minutes. 77 people said they could only tolerate 16-30 minutes. 6 people said they were only okay with deliveries that took fewer than 15 minutes. 3 people were generally apathetic about delivery time, saying they could tolerate longer than a 1 hour wait.

14. How much does the price of the pizzeria affect your decision to order from it?

- a. Completely
- b. Somewhat
- c. Neutral
- d. Doesn't really
- e. None

While price was one of the two most important pizzeria factors in our sample, a majority (118) felt that price had somewhat (but not large) of an influence on their decision of where to order from. 41 said it completely influenced their pizzeria choice, while 41 said it had a neutral impact and 7 said it had no effect at all.

15. How much does the occasion for eating the pizza (ie just for dinner, for a party, etc) affect your decision on what pizzeria to order from?

- a. Completely
- b. Somewhat
- c. Neutral

- d. Doesn't really
- e. None

The answer choices fluctuated in frequency. 63 said the occasion for eating pizza had somewhat of an influence. 62 said it didn't really matter. Fewer people chose answers at the extremes or between the two modes, but each answer choice had at least 32 votes.

16. What time of day do you typically order pizza? Mark all that apply.

- a. 4:00 a.m. – 10:00 a.m.
- b. 10:00 a.m. – 2:00 p.m.
- c. 2:00 p.m. – 6:00 p.m.
- d. 6:00 p.m. – 11:00 p.m.
- e. 11:00 p.m. – 4:00 a.m.

The most popular time of day for eating pizza among our survey takers was between 6 pm to 11 pm, with around 2/3 of our sample selecting that as one of their choices. The late-night option was also very popular, with around 1/2 of our sample mentioning that window in their response. Less than 50 people voted for the afternoon hours between 2 to 6 pm, and very few people ate pizza between 4 am to 2 pm.

17. How much do business hours affect your decision to order from a pizzeria?

- a. Always
- b. Often
- c. Sometimes
- d. Occasionally
- e. None

The importance of business hours had a roughly normal distribution, with 33 saying none, 57 responding occasionally, 62 selecting sometimes (the most frequent response), 54 saying often, and 28 saying that it always influenced their decision of pizzeria.

18. On what day of the week do you typically order pizza? Mark all that apply.

- a. Monday
- b. Tuesday

- c. Wednesday
- d. Thursday
- e. Friday
- f. Saturday
- g. Sunday

Multiple answer responses were allowed for this question. Friday and Saturday were by far the most popular days for ordering pizza, with over 120 people each picking Friday and Saturday. The other days of the week had roughly even distributions, but the next most popular day had fewer than 60 votes. Mondays and Tuesdays were the last popular days for ordering.

19. How do you typically pay for pizza?

- a. Cash
- b. Check
- c. Credit Card
- d. DineXtra
- e. PlaidCa\$h
- f. Other: _____

Credit card was the most common payment method, with 109 respondents. 62 people said they used the DineXtra cash on their meal plan to pay for pizza. 42 said they paid by cash. Only one person told us they paid by cash. Of the write-in responses, 7 said they paid using meal blocks, 1 listed the faculty as their source, and 1 said that others would pay for it.

20. How many of each size pizza do you typically order in one order?

- a. Small: _____
- b. Medium: _____
- c. Large: _____
- d. Extra Large: _____
- e. Extra Extra Large: _____

The most frequent size that the sample ordered was medium and large. A majority of the respondents said they ordered at least one medium pizza, and a majority said they picked at least one large pizza per order. People who ordered multiple pizza ordered large size more than any other type. All of the

people who ordered small pizzas did not typically order any other type or quantity (greater than 1). About $\frac{1}{3}$ rd of those who ordered extra extra large typically ordered multiple pizzas per occasion.

21. Which pizzeria do you feel is the best in Pittsburgh (including on campus dining)?

- a. Vocelli's
- b. Papa John's
- c. Domino's
- d. Pizza Palerma
- e. Tartan's Pavillion
- f. Antoons
- g. Mineos
- h. Lucci's
- i. Other: _____

Vocelli's and Papa John's were the two most popular responses, gaining 60 and 46 votes, respectively. Domino's came in a distant third with 21 votes. Despite Pizza Hut not being among the choices we listed, enough people wrote them in as their favorite pizzeria to get the fourth highest total, with 16 respondents. Over twenty different pizzerias were mentioned in the "other" category, but other than Pizza Hut, no other shop received more than 4 votes.

Appendix B: Margin of Error Calculations

We observed the Carnegie Mellon the total CMU population included in our target population order to estimate the sample size necessary to obtain a 6.2% margin of error, we utilized the following margin of error estimation:

$$ME = .062 \leq 1.96 \times \sqrt{\left(1 - \frac{n}{5219}\right)\left(\frac{.25}{n-1}\right)}$$

$$n \cong 240$$

Once we finalized our set of responses, we had exactly 234 responses, giving us approximately a 6.27% margin of error.

$$ME = 1.96 \times \sqrt{\left(1 - \frac{234}{5219}\right)\left(\frac{.25}{233}\right)} = 6.27\%$$

Appendix C: Post-Stratification Weights

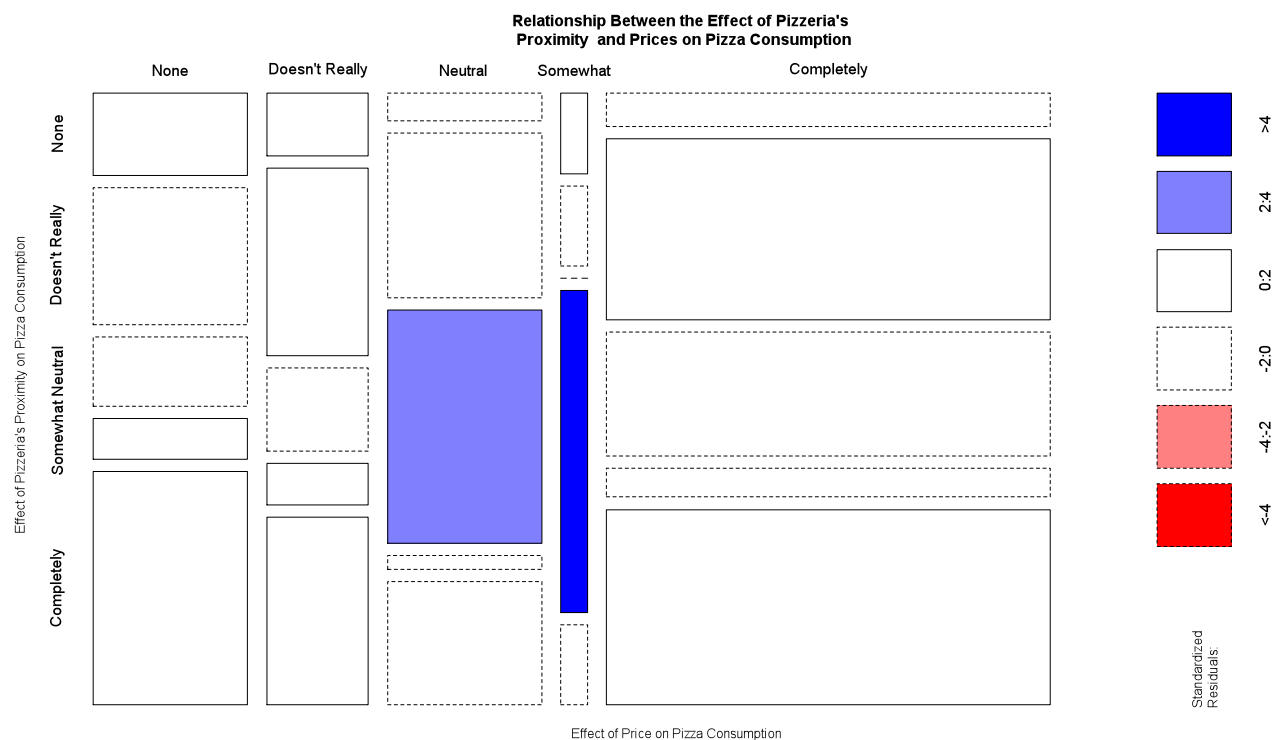
	Freshmen	Sophomores	Juniors	Seniors
Male	0.877792681	1.476144855	0.993868557	1.804656064
Female	0.966901872	0.743662084	0.57518408	1.015539375

The weights above show that, although most groups in our target population were fairly well-represented, females tended to be overrepresented in our responses, and senior males were particular underrepresented.

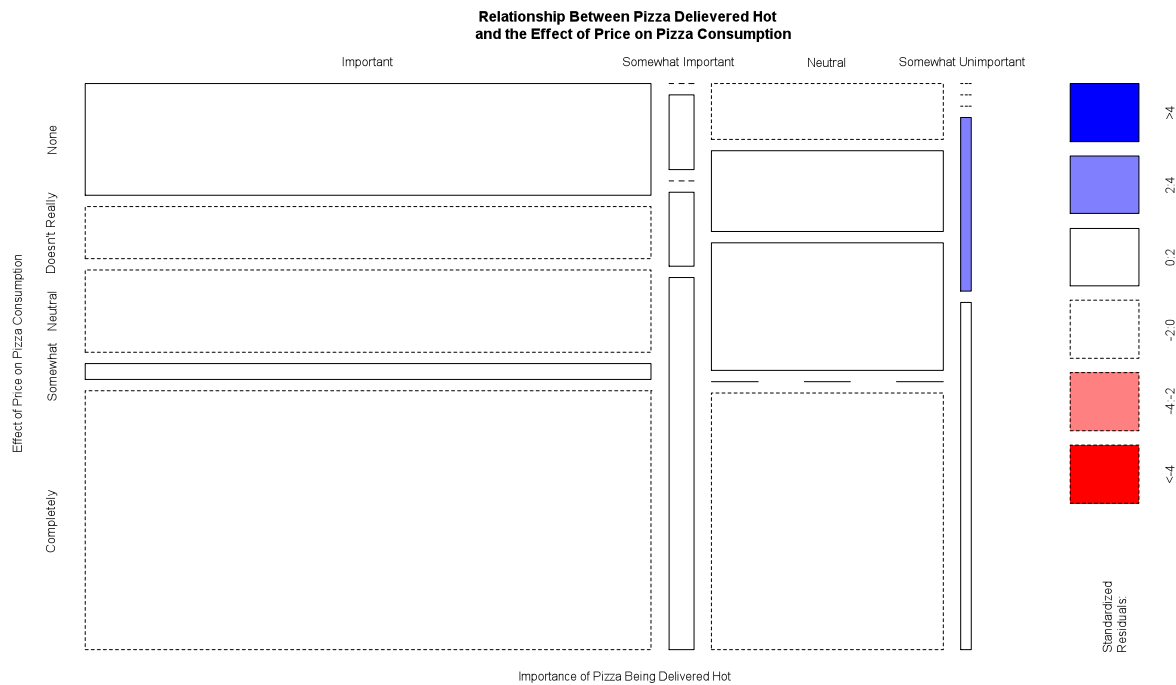
indicating how these are
calculated would be helpful

Appendix D: Mosaic Plots for Analysis of Independence

The mosaic plot is one of the best options in showing whether the categorical variables are dependent or independent. The plot shows the marginal distribution of the first variable (shown by the width of the boxes under each category of that variable), and the conditional distribution of the second variable given the first variable (shown by the height of the variable for categories under each of the first variable’s categories). With the mosaic plot, we can describe the relationship of the two variables. The shades of the mosaic plot are indicative of the deviations from the expected results under independent. It is possible to include three or more categorical variables to the mosaic plot as well, where all variables other than the first will show conditional distributions given the first variable.

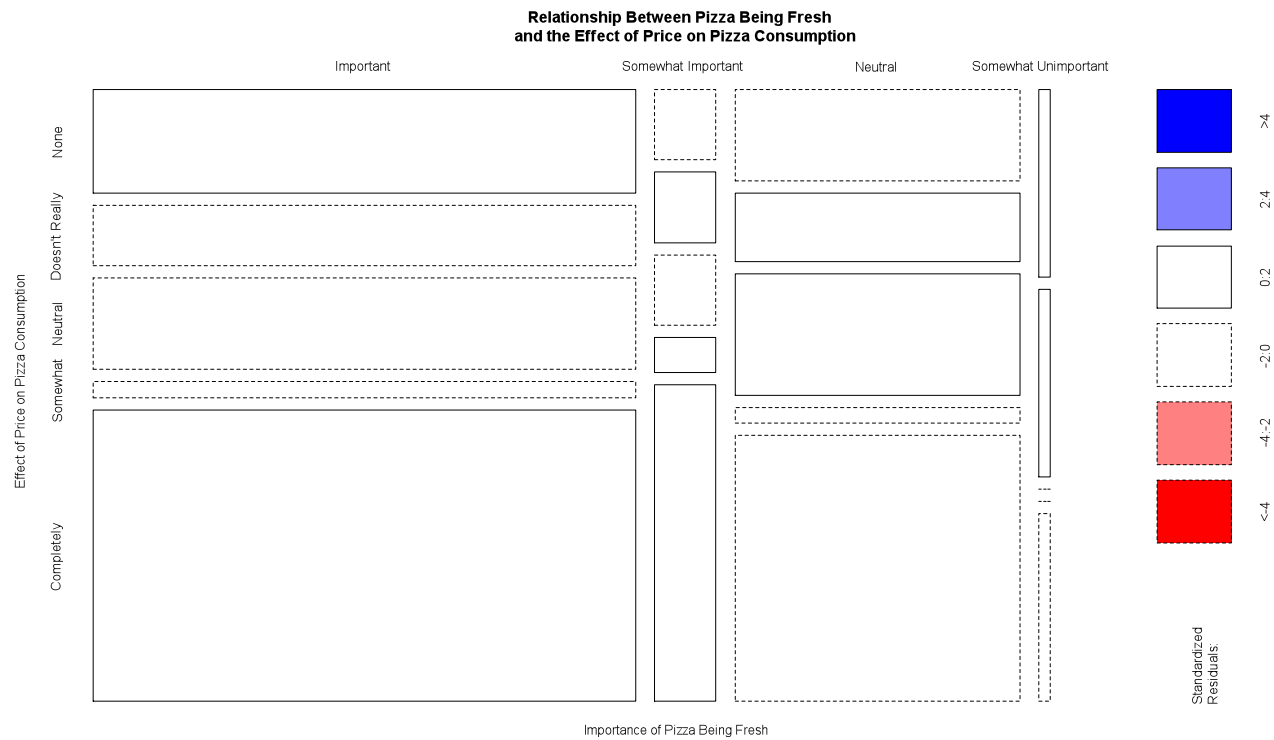


Mosaic Plot of Variables for the Effect of Pizzeria’s Proximity and Prices



M

Mosaic Plot of Variables for the Pizza Delivered Hot vs. Effect of Price on Pizza Consumption



Mosaic Plot of Variables for the Pizza's Freshness vs. Effect of Price on Pizza Consumption

R Code for Mosaic Plots

```
library(MASS)
library(graphics)

finaldata<-read.csv("Pizza Survey rev final2.csv",header=T,sep="," ,quote="\")
names(finaldata)

effectp<-finaldata[,15]
importancehot<-finaldata[,18]
deliverymin<-finaldata[,17]
proximity<-finaldata[,16]
freshness<-finaldata[,19]

importancehot<-factor(importancehot,labels=c("Important","Somewhat Important",
"Neutral","Somewhat Unimportant"))
effectp<-factor(effectp,labels=c("None","Doesn't
Really","Neutral","Somewhat","Completely"))
proximity<-factor(proximity,labels=c("None","Doesn't
Really","Neutral","Somewhat","Completely"))
freshness<-factor(freshness,labels=c("Important","Somewhat
Important","Neutral","Somewhat Unimportant"))

mosaicplot(table(effectp,proximity),shade=T,xlab="Effect of Price on Pizza
Consumption",ylab="Effect of Pizzeria's Proximity on Pizza
Consumption",main="Relationship Between the Effect of Pizzeria's \n Proximity
and Prices on Pizza Consumption",cex.axis=1.2)
mosaicplot(table(importancehot,effectp),shade=T,xlab="Importance of Pizza Being
Delivered Hot",ylab="Effect of Price on Pizza Consumption",main="Relationship
Between Pizza Delievered Hot \n and the Effect of Price on Pizza
Consumption",cex.axis=1.1)
mosaicplot(table(freshness,effectp),shade=T,xlab="Importance of Pizza Being
Fresh",ylab="Effect of Price on Pizza Consumption",main="Relationship Between
Pizza Being Fresh \n and the Effect of Price on Pizza Consumption",cex.axis=1.1)
mosaicplot(table(importancehot,freshness),shade=T,xlab="Importance of Pizza
Being Delivered Hot",ylab="Importance of Pizza Being Fresh",main="Relationship
Between Pizza Delivered Hot \n and the Pizza Being Fresh",cex.axis=.8)

wdays<-finaldata[,11]
oncampus<-finaldata[,23]
ordertime<-finaldata[,25]

oncampus<-factor(oncampus,labels=c("0","1 to 2","3 to 5","6 to 10","10 to
15","15+"))

days0<-wdays[oncampus=="0"]
days1<-wdays[oncampus=="1 to 2"]
days3<-wdays[oncampus=="3 to 5"]
days6<-wdays[oncampus=="6 to 10"]
days10<-wdays[oncampus=="10 to 15"]
days15<-wdays[oncampus=="15+"]

boxplot(days0,days1,days3,days6,days10,days15,names=c("0","1 to 2","3 to 5", "6
to 10","10 to 15", "15+"), xlab="Number of Days of Ordering \n Pizza per Week,
by Number of On-Campus Dining per Week",ylab="Days of Ordering Pizza per Week",
main="Boxplot: Number of Days of Ordering Pizza per Week", col=c(2,2,4,4,8,8))

orderday<-read.table("orderdays.txt",header=TRUE)
```

```
timeofday<-read.table("timeofday.txt",col.names=c("10am to 2pm","2pm to  
6pm","6pm to 11pm","11pm to 4am","4am to 10am"),check.name=FALSE)  
  
barplot(as.matrix(orderday),col=2,xlab="Day of Week",ylab="Number of Orders  
Typically Placed",main="Number of Pizza Orders Placed per Day of Week")  
barplot(as.matrix(timeofday),col=4,xlab="Time of Day",ylab="Number of Orders  
Typically Placed",main="Number of Pizza Orders Placed per Time of Day")
```

Appendix E: Frequency Tables for Comparison of Pizzerias

When comparing responses for those who answered that Vocelli's was their favorite pizza restaurant vs. those who answered Papa John's, we had a total of 106 observations, but a total weighted sum of 103.14. The following frequency tables break down the responses and allowed for the analysis done on page 13.

	annually	monthly	weekly	daily
Vocelli's	0.88	21.18	35.38	0.88
Papa John's	1.32	18.74	21.09	3.67

*Weighted Number of Responses for Pizza Consumption Frequencies
(neither group answered "never")*

	thin crust	thick crust	other
Vocelli's	32.90	17.51	7.91
Papa John's	15.44	16.36	13.02

Weighted Number of Responses for Pizza Type Preferences

	quality	price	other
Vocelli's	25.25	26.87	6.20
Papa John's	22.46	16.40	5.96

Weighted Number of Responses for Most Important Pizza Factors

	never	occasionally	sometimes	often	always
Vocelli's	7.73	15.08	11.05	16.11	8.35
Papa John's	9.99	6.50	14.97	7.74	5.62

Weighted Number of Responses for Importance of Business Hours

	none	somewhat unimportant	somewhat important	completely
Vocelli's	11.24	11.15	25.58	10.35
Papa John's	9.50	1.48	24.46	9.38

Weighted Number of Responses for Importance of Price

	none	somewhat unimportant	somewhat important	completely
Vocelli's	19.38	18.44	17.31	3.19
Papa John's	11.07	13.28	16.52	4.05

Weighted Number of Responses for Importance of Proximity

	0-15 min	16-30 min	31-45 min	46-60 min	>60 min
Vocelli's	0.74	16.03	36.49	3.58	1.48
Papa John's	0.88	16.31	21.87	5.76	0

Weighted Number of Responses for Acceptable Delivery Time

	none	somewhat unimportant	somewhat important	completely
Vocelli's	1.76	1.48	22.85	32.23
Papa John's	0.00	0.74	6.27	37.81

Weighted Number of Responses for Importance of Pizza Arriving Hot

	none	somewhat important	completely
Vocelli's	4.12	23.13	31.07
Papa John's	3.08	6.90	34.84

*Weighted Number of Responses for Importance of Freshness
(neither group responded "somewhat unimportant")*