Section C1 Group 1: Aarom Guillaume, Aziza Barkuschwabe, HaoTian Liu, Jaskaran Singh.

How Do Boston University Undergraduate Students Commute on the Charles River Campus?

Abstract

Transportation is central to the daily life of Boston University undergraduate students. This paper seeks to understand the patterns of usage and satisfaction with transportation among this population, as well as if there could be a relationship between transportation satisfaction and academic performance. Data collection was conducted using a survey that was distributed on social media and through direct in person interactions. 135 undergraduates responded. We analyze this data using descriptive and inferential statistics to investigate the state of transportation usage by BU undergraduates. Unexpectedly, our findings indicate that BU undergraduates who use the T and MBTA more are no more or less satisfied with transportation, and on or off campus housing location may not significantly impact use of the T and the BU Shuttle. Additionally, GPA was not shown to correlate with transportation satisfaction. These results could imply that BU undergraduate students view and use transportation similarly, independent of other factors, and that by studying the broad trends the most effective strategy for improvement could be identified.

Introduction

Due to the expansive campus, as well as student desire to live off-campus, Boston University's undergraduate population employs various mediums of transportation to travel to class as well as to their housing. This study hopes to get a holistic view and understanding of the different modes of transportation used as well as the frequency of which they are used, particularly by full-time undergraduate students at Boston University's Charles River campus. This will entail understanding the factors that influence transportation use, the limitations, as well as the overall student satisfaction with each mode of transport.

The aim of this study is to gain a better understanding of the transportation methods so that improvements can be made where necessary. We sought to analyze the variance in the mode of transportation used including those methods provided by the University, the city of Boston, and self-transport methods. Additionally, we wanted to see if there would be any correlation between where a student lives in relation to the University, and what method or methods of transportation they preferably use.

By further looking into the data, we hope to find patterns and connections between housing, GPA, and method of transport used within the undergraduates. The hypotheses we aim to investigate are as follows:

1. Frequency of T and MBTA bus usage is correlated to satisfaction with transportation overall.

2. Students who live off campus ride the T and BUS more times per week than students who live on campus.
3. There is a correlation between overall transportation satisfaction and GPA.

Experimental methods

The population of interest in this study are the ~17,000 full-time undergraduates at BU, and the sample will be the group of students who respond to the survey. In total, we contacted 135 undergraduates over four weeks in October and November of 2022.

Before beginning our final study, we conducted a brief pilot study and surveyed 36 respondents. After the pilot study, some immediate issues with the survey itself became apparent, and data cleaning was necessary for many variables. We resolved many of the problems by refining our questions, including by switching to a numeric scale (as opposed to two day ranges) for our frequency questions, and by more clearly instructing respondents in how to take the survey, and what questions they didn't need to answer. We also added a demographic question about class year so that we could have a better grasp on the representativeness of our sample.

A Google Form was used for data collection. Multiple qualitative and quantitative variables were considered. Firstly, the modes of transportation used by a given student, as well as the frequency of use for each mode of transportation, as in times used in the last week. Other basic information we collected for comparison includes GPA, class year, school(s) of study, major(s), and housing location. Additionally, we asked students to rate their satisfaction with their transportation on/to the BU Charles River campus overall, as well as satisfaction with the specific transportation modes or infrastructure that they used, on a numerical scale.

To avoid volunteering bias as much as possible, we contacted people directly with the request to complete the survey. Selection bias is inherent in this method, especially because we are a group of only a few students who are limited in our time and resources. Our group contacted friends and acquaintances, both via social media and in person, and we also conducted two surveying sessions on different days at the George Sherman Union (GSU) wherein we estimate we contacted 60 undergraduates unknown by our group. We chose the GSU because it is a place frequented by BU undergraduates and is not populated by a specific subset of the student population, which might be the case in an area such as a dining hall or T stop.

Unfortunately, we did not avoid sampling bias and our sample is not representative of BU's undergraduate population, particularly due to an overabundance of Sophomore respondents, which will be described more fully in the following section. As such, despite our efforts to avoid bias, our sample is not random.

Data visualization and results

Figure 1. Comparison of usage of modes of transportation in the last 30 days.

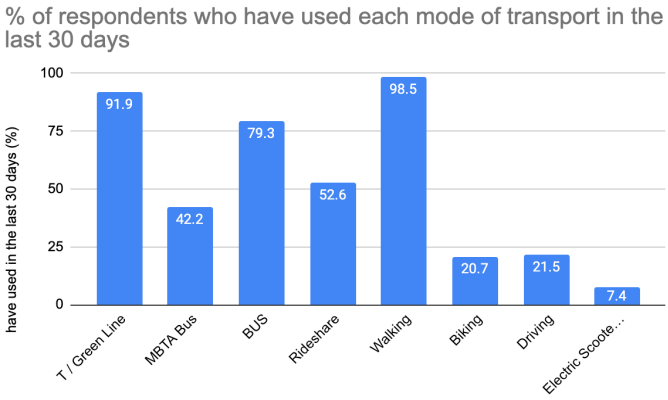
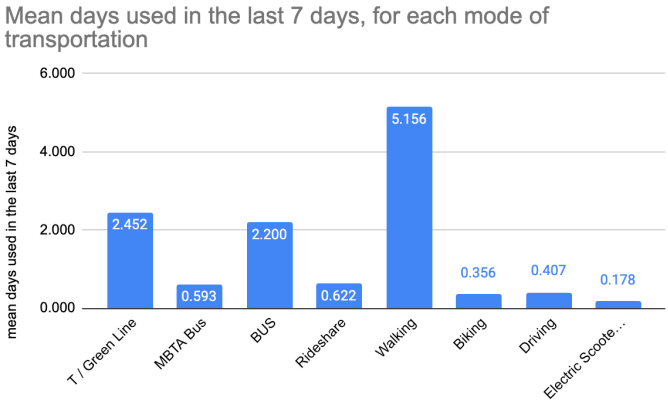


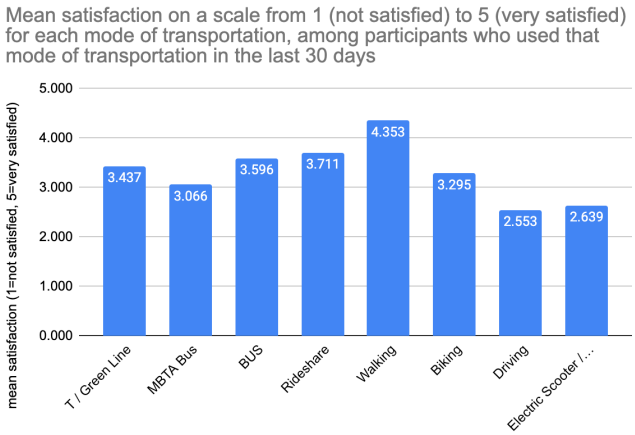
Figure 1 shows the proportion of respondents who said they had used each respective mode of transportation in the last 30 days. There is a mix of different modes used, the most common being walking, riding the T, and the BU Shuttle. Rideshare apps were the next most popular, then the MBTA bus, then biking and driving and finally the least common mode of transportation was electric scooter or skateboard. This distribution, while not altogether unexpected, reveals the areas where BU's transportation infrastructure may be the most important to this sample.

Figure 2. Comparison of mean frequency of use for each mode of transportation.



Similarly to the data shown in Figure 1, Figure 2 shows that the most frequently used modes of transportation are walking, riding the T and the BUS. This data reveals that not only is it common to use those modes of transportation, those modes are often used more often in the day-to-day lives of the respondents, and that walking by far is the most preferred mode of transportation for day-to-day use.

Figure 3. Comparison of mean satisfaction with each mode of transportation, among respondents who used each respective mode of transportation in the last 30 days.



From the data shown in Figure 3, it seems that the population on average (here, showing the mean) has the greatest satisfaction with walking. Recall that walking is the most commonly and frequently used mode of transportation (Figure 1, Figure 2) which is significant because it seems that the needs and wants of students who walk to commute are

likely being met, at least for this sample. Most other mean satisfaction values fell between 3 and 4, so medium to satisfactory (but not high satisfaction), except for driving and commuting by electric scooter / skateboard, which both fell around 2.5. It's important to note that for each of these values, only respondents who actually used each mode of transportation were asked to rate their satisfaction.

Figure 4. Overall satisfaction distribution.

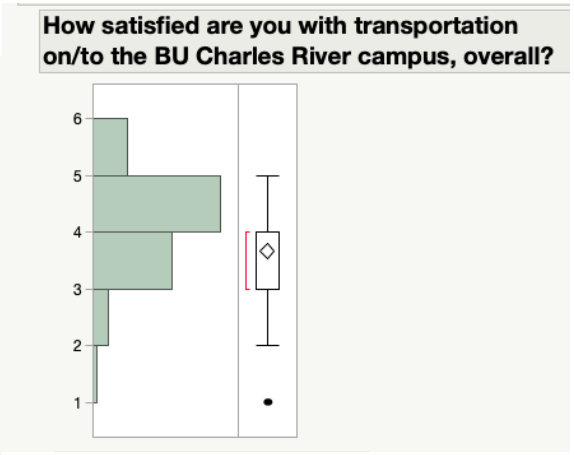


Figure 4 shows that overall, respondents were satisfied with their experience of commuting on the Charles River campus. The median rating was 4 and the mean was 3.667. There is a possible outlier, rating 1, which two respondents listed as their level of satisfaction.

Figure 5. GPA distribution.

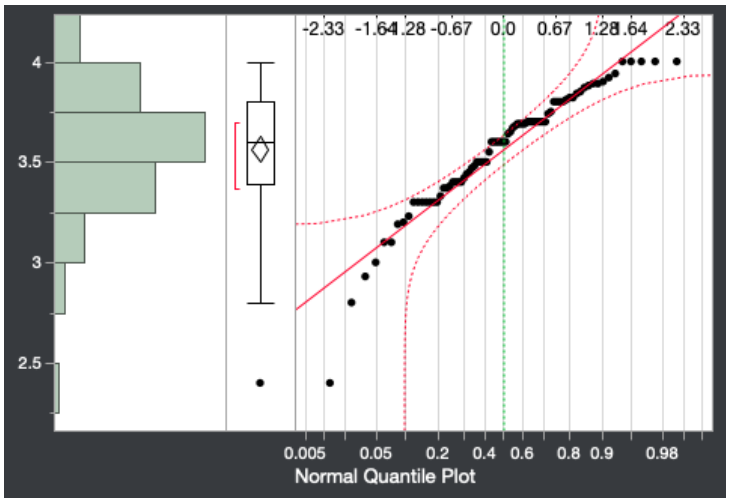
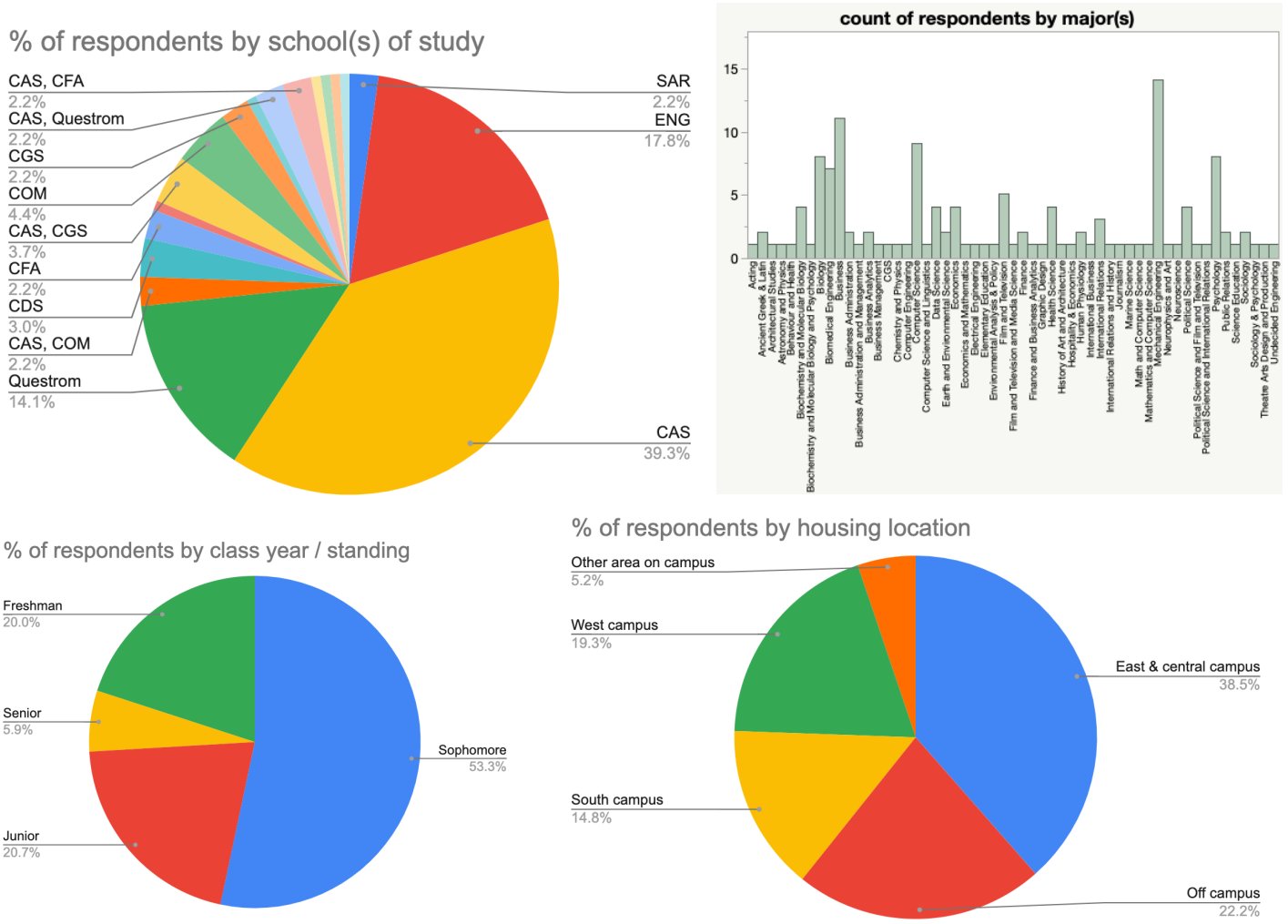


Figure 5 shows that the distribution of Grade Point Averages (GPAs) among our sample is skewed to the left, with a relatively high mean and median of 3.565 and 3.6 respectively. There is a possible outlier of 2.4. Only 81 of the 135 respondents listed a GPA, so this is not likely to be representative of our sample's academic performance. Assessment of normality: The distribution for this variable is roughly mound-shaped, but it is not symmetric. With the exception of the outlier, however, it appears somewhat normal. Examining

the QQ (or Normal Quantile) Plot, we see that there is at least one point that falls outside of the lines, deviating from the line of slope=1. Thus, we cannot conclude that the data are normally distributed.

Figure 6. Demographic data, including school(s) of study, major(s), class year, and housing location proportions.

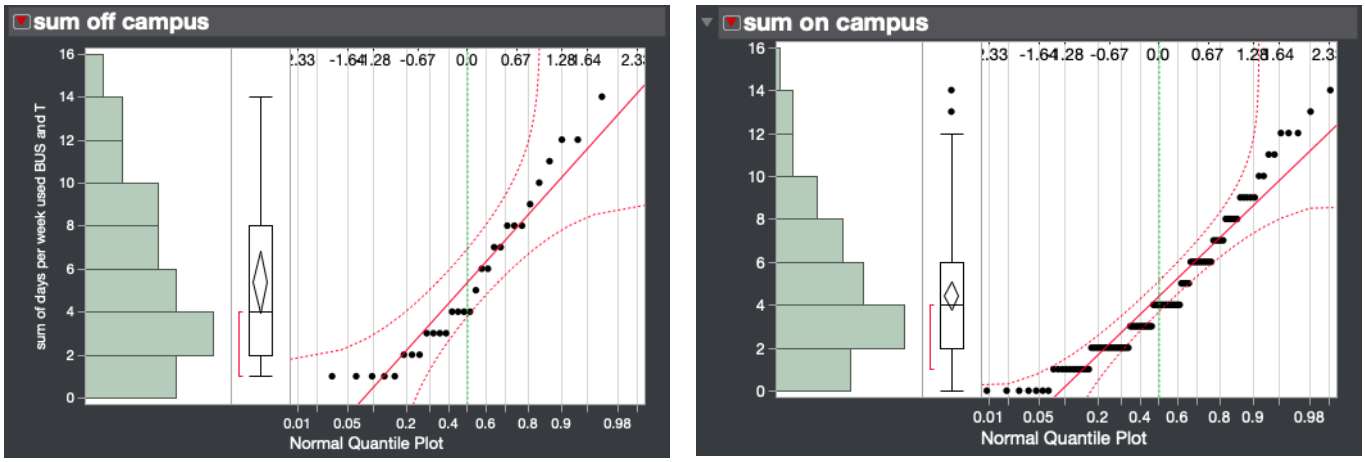


There are students from every one of BU's undergraduate schools and colleges represented in the sample. A plurality of respondents are students in CAS and the other largest groups are ENG and Questrom. Each school/college is represented by at least one student in our sample (note that SHA is represented by only one student, who is also a student at another college). Similarly, the distribution of majors is wide, with only a few significant repeats among popular majors (Psychology, Biochemistry & Molecular Biology, and others).

Unfortunately, the data we have collected shows a clear sampling bias. Specifically, our demographic data shows that 53.3% of our respondents were Sophomores, while only 5.9% were

Seniors. This is far from a representative sample of the BU undergraduate population, which is in reality split approximately evenly between the four class years, with some students who don't fit into any of those categories. From the data provided by BU to the City of Boston[1] in 2021, 6263 (961 living at home plus 5302 not at home) of the 17334 (17526 total minus 192 studying abroad) full-time undergraduates not studying abroad live off campus, which is approximately 36.1% of students. Only 22.2% of our respondents reported living off campus, indicating another possible source of bias. Based on these findings, we cannot claim to have a representative sample.

Figure 7: Distribution of sum of T and BUS usage frequency for off and on campus residents.



This figure shows the sum of frequencies of T and BUS usage (in days per week, in the last week) for two subsets of our sample: off campus residents (30 respondents) and on campus residents (105 respondents). The mean value for off campus students is 5.4 days total and the median is 4 days total. For on campus students, the mean frequency is approximately 4.438 days while the median is also 4 days total. There are two possible outliers, with values 13 and 14 days total respectively, in the on campus distribution. Since those are still reasonable values (it is completely possible to ride both the T and BUS to commute every day of the week), we have chosen not to eliminate them from our data. Both distributions are right-skewed.

Assessments of normality: For the off campus residents, the data is roughly normal for the sample size as illustrated by the QQ (or Normal Quantile) plot, and the histogram which while considerably right-skewed, does have some normal character. However, the off campus distribution is not normal, as shown by the QQ plot wherein multiple data points fall outside of the lower percentile line, deviating significantly from the reference line.

[1] "Student Housing Trends: 2021." City of Boston Mayor's Office of Housing, 2021. Page 8. https://www.boston.gov/sites/default/files/file/2022/08/Student%20Housing%20Report%2C%202021.pdf

Confidence intervals: The following confidence intervals are all based on a significance level of 95%, or $\alpha = 0.05$.

Class year (proportion): With 95% confidence, we expect that the population proportions of students belonging to each category will fall within the following confidence intervals, based on our data.

Freshman (0.14125, 0.275351) | Sophomore (0.449423, 0.6154) | Junior (0.147577, 0.283428) | Senior (0.03033, 0.112577)

Logically, the proportion of each grade should be about a quarter or 0.25. However, our results indicate that there is sampling bias, as discussed previously, because the expected proportion (~0.25 for each year) falls within only the Freshman and Junior confidence intervals, so the proportions of data here are far from accurate of BU's actual undergraduate population.

Housing location (proportion):

Off campus (0.160329, 0.299486) | East & central campus (0.307354, 0.46937) | South campus (0.097996, 0.21777) | West campus (0.134956, 0.26724) | Other area on campus (0.025341, 0.103161)

With 95% confidence we expect that the population proportion of students who live off campus, for example, falls within (0.160329,0.299486). In this case, with 95% confidence, we expect that the population proportion of students living in different house locations will fall within the respective confidence intervals. From the intervals above, we can deduce that most of the students live in East & central campus, and the number of students who live in other areas on campus is the least.

GPA (mean): The 95% confidence interval of the sample mean of GPA values (out of 4.0) is [3.4993849, 3.6297509]. However, as indicated by Figure 5, the distribution of GPAs of our sample is not normal. However, the sample size is 135, which is greater than 30. The sampling distribution of the population mean can be approximated by normal according to the central limit theorem.

Overall satisfaction (mean): With 95% confidence, we can expect the population mean of the overall satisfaction on a scale from 1 (not satisfied) to 5 (very satisfied) to fall within the respective confidence intervals, which is (3.5240952, 3.8092382).

Frequency (days used in the last week) per mode of transportation (mean):

Sum of T and MBTA Bus (2.5832252, 3.5056637) | T / Green Line (2.0942309, 2.8094728) | MBTA Bus (0.3994047, 0.7857804) | BUS (1.8393195, 2.5606805,) | Rideshare (0.4573324, 0.7871121) | Walking (4.741264, 5.5698471) | Biking (0.1785231, 0.532588) | Driving (0.2022761, 0.6125387) | Electric Scooter / Skateboard (0.0384795, 0.317076)

With 95% confidence, we can expect the population mean of students who take different transportation methods to fall within the respective confidence intervals. As we can see in the

data, most students prefer to take the T / Green line, and very few use an electric scooter or skateboard.

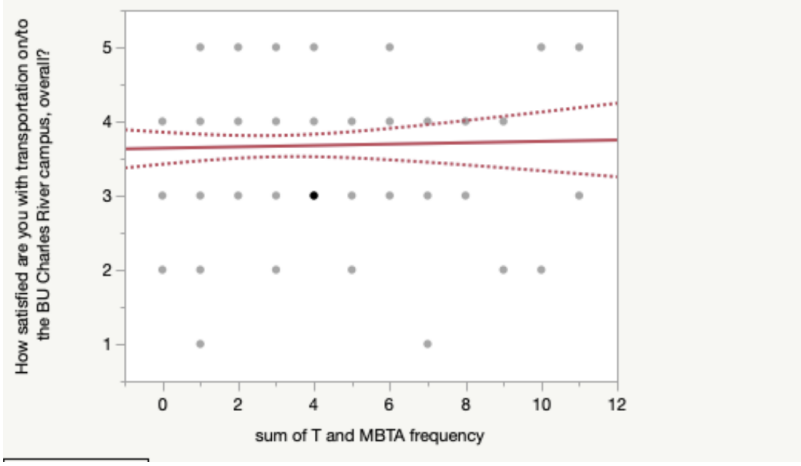
Hypothesis testing:

Hypothesis 1

For our first hypothesis, we set α equal to 0.05 and claimed that the frequency of the Green Line and the MBTA Bus is inversely related to the overall satisfaction with transportation. Our null hypothesis is that $\beta_1 = 0$, implying that there is no correlation between MBTA use and satisfaction and our alternative hypothesis is that $\beta_1 \neq 0$, implying that there is correlation because MBTA use and satisfaction. β_1 is a part of the equation $y = \beta_1(x) + \beta_0 + \varepsilon$, which shows that there is a linear relationship between the two variables. Our hypotheses are as followed:

Null Hypothesis: $\beta_1 = 0$
Alternative Hypothesis: $\beta_1 \neq 0$

After doing the test, we've found that β_1 is equal to 0.0091488 and the p-value is equal to 0.7333. Because the p value is greater than α, we fail to reject the null hypothesis because there is insufficient evidence to prove that there is correlation between the frequency of MBTA use and overall satisfaction. Also our R^2 value is 0.000876. Since this value is very close to 0, it concludes that there is little to no correlation between the two variables. In our graph, we included confidence bands to the line of best fit and none of the values are within the confidence bands, revealing to us that there is no correlation.



R^2(level of correlation)	α	MBTA Use(β_1)	P-Value
.000876	.05	.0091488	.7333

Hypothesis 2

Our second hypothesis was that students who live off campus will use the T and BUS more frequently than on campus residents. The null hypothesis was that there is no difference in frequency of T and BUS use between on and off campus students. To test this, we used the sum of T and BUS usage (times per week) for each student. We compared the mean frequency for off campus students (μ_{off}) to the mean frequency for on campus students (μ_{on}) using a two sample hypothesis test of independent samples.

Null hypothesis: $(\mu_{off} = \mu_{on}) = (\mu_{off} - \mu_{on} = 0)$
Alternate hypothesis: $(\mu_{off} > \mu_{on}) = (\mu_{off} - \mu_{on} > 0)$

The following table shows the values we found to describe our data, after having split our responses into two samples, on campus residents and off campus residents.

Housing location	Sample mean ($\bar{x}$)	Standard deviation (s)	Count (n)
On campus	4.438095238	3.278495813	105
Off campus	5.4000	3.783813074	30

We calculated the test statistic using the following formula for the z-value, which we chose because both our samples had $n \geq 30$ and unknown true standard deviations.

$$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \qquad z = \frac{(5.4000 - 4.4381)}{\sqrt{\frac{(3.7838)^2}{30} + \frac{(3.2785)^2}{105}}} = 1.2635$$

Test statistic (z-value)	P-value (from the Standard Normal distribution)
1.263469457	0.1038

Next, we conducted a one-tailed z-test for a confidence level of 95% (or $\alpha = 0.05$). The p-value, approximately 0.1038, is greater than 0.05. Due to this result, we fail to reject the null hypothesis.

However, this inference is only meaningful if both of our samples are normally distributed. By the Central Limit Theorem, since both samples have $n \geq 30$, we can state that they are approximately normal and thus can apply the z statistic and approximate the population standard deviations using the sample standard deviations. However, examining the data, reveals that the on campus sample is not normally distributed (see Figure 7). Thus, our hypothesis test is fundamentally flawed and the result cannot be considered useful without further investigation.

Hypothesis 3

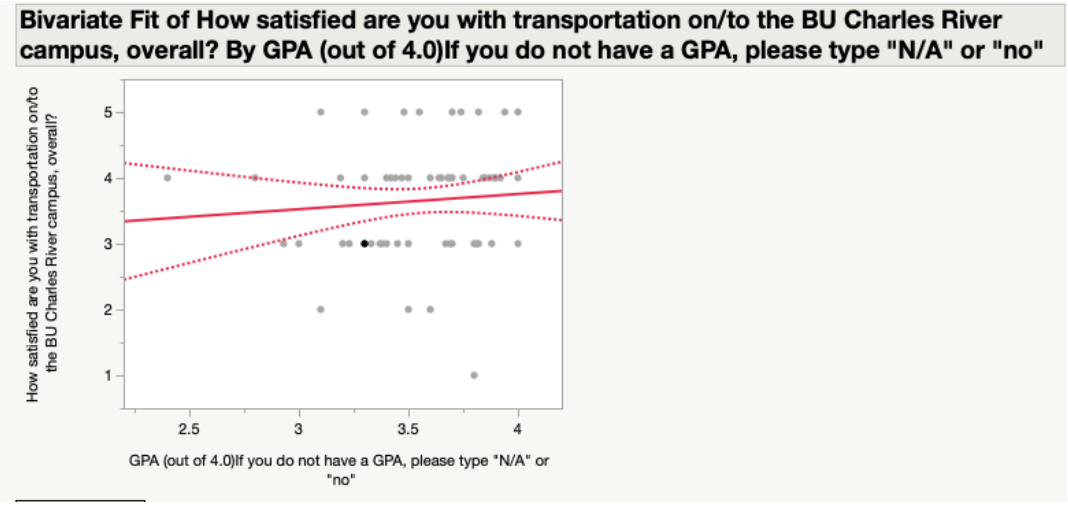
For our third hypothesis we also set α to be 0.05 , we proclaimed that the satisfaction mode of transportation used would have no correlation to the respondents GPA. Our null

hypothesis is that $\beta_1 = 0$, which would mean that GPA and overall satisfaction with the method of transport is not correlated. Our alternative hypothesis is that $\beta_1 \neq 0$, which would mean that as overall satisfaction with transportation increases, the respondent GPA will also increase. β_1 is used in the analytical equation $y = \beta_1(x) + \beta_0 + \varepsilon$, which shows a linear relationship between the two variables. Our hypothesis are as follows:

Null Hypothesis: $\beta_1 = 0$

Alternative Hypothesis: $\beta_1 \neq 0$

After testing, we found β_1 to be equal to 0.2312858 and that the p value of the statistic is 0.4708. Due to the fact that the p value is greater than the α value of 0.05, we fail to reject the null hypothesis due to the lack of evidence to prove the correlation between overall transportation satisfaction and GPA. The R^2 value of 0.006603 being very close to 0, allows us to conclude that there was very little correlation between the two variables being compared. In including confidence bands on our graph of the line of best fit, it was evident that there was no correlation as none of the values were within the confidence bands.



R^2(level of correlation)	**α**	**Transportation Satisfaction (β_1)**	**P-Value**
0.006603	0.05	0.2312858	0.4708

Discussion

In conclusion, our data does not support the hypothesis that the frequency of MBTA usage is correlated with the overall satisfaction of transportation on campus overall. This result was unexpected since we predicted that students who use the MBTA more often tend to be dissatisfied with the service. It's possible that students who take the MBTA frequently may not like the service but they would rather use transportation than stick to walking if they live off

campus and far from classes. Furthermore, our data also does not support our third hypothesis, which states that increased satisfaction with transportation would correlate to an increased GPA. We are not surprised by this outcome since the usage of transportation doesn't have a direct effect on grades. It's true that a day of transportation may lead to someone being late to class, but that wouldn't ultimately impact their performance in the class since most of the notes are online. For our second hypothesis, we claimed that students who live off campus would use transportation more often than students who live on campus. Even though we fail to reject the null hypothesis that the mean usage of transportation is equal to each other, the on campus sample of data is not normally distributed so we cannot on a solid basis claim to reject or fail to reject the null hypothesis.

Our goal was to understand how BU undergraduates commute, and we definitely have collected some useful data that allowed us to begin to analyze what factors may (or as indicated by our data, may not) influence different students' experiences. However, sampling bias was a major problem for this project. To obtain more accurate results in the future, a larger sample size and more representative proportions of each class year and housing location ought to be surveyed, which can be obtained by spending more time surveying in locations such as the GSU. One other possible way to improve investigation into this topic would be to ask respondents to share their thoughts about transportation. Sometimes, their satisfaction or dissatisfaction may be due to their personal reasons. Taking that into consideration may improve the accuracy of our study, compare trends between groups, and allow us to talk a bit more on how to improve the transportation system in order to make students in Boston University more comfortable and satisfied.

Contributions

This report was written by Aarom Guillaume, Aziza Barkuschwabe, HaoTian Liu, and Jaskaran Singh. Hypothesis testing was done by Aarom Guillaume, Aziza Barkuschwabe, and Jaskaran Singh, and data analysis was also done by HaoTian Liu.

Deliverable 4 was written by Aziza Barkuschwabe, HaoTian Liu, and Jaskaran Singh. Data analysis was done by Aarom Guillaume and Aziza Barkuschwabe. Data collection was done by all members of the group. Deliverable 3 was written by Aarom Guillaume, Jaskaran Singh and Aziza Barkuschwabe. Data analysis was done by HaoTian "Kris" Liu, Jaskaran Singh, Aarom Guillaume and Aziza Barkuschwabe. Deliverable 2 was written by Aarom Guillaume and Aziza Barkuschwabe, the survey for our pilot study was designed by Jaskaran Singh, HaoTian Liu and Aziza Barkuschwabe, and all members of the group contributed their ideas and participated in data collection. Deliverable 1 was written by Aziza Barkuschwabe, with ideas and questions also contributed by Aarom Guillaume and Jaskaran Singh.