An Analysis of Undergraduate College Majors and Future Wages

Part 1: Introduction

Section A: Criterion and Predictor Variables

One of the age-old questions every college student faces is deciding which major to go into and why. Is it for money? Is it for passion? For the average college attendee, it's safe to say many primarily choose the major that will allow them to accumulate the most wealth in their future careers. With this in mind, I will be analyzing data provided by the U.S. Bureau of Labor Statistics. This analysis will be focused on the top 5 most popular majors and the top 5 least popular majors in the United States and the likely potential future wages of these occupations. To determine the highest and lowest wages of the most prevalent occupations, the criterion will be the *level of popularity of the major* from 2022, and the predictor will be the *annual median wage*, assuming popularity in a major is positively correlated with higher wages.

I anticipate a positive correlation between my criterion, the level of popularity of the major and their respective occupations, and my predictor variable, annual mean wage. With this anticipation, under the assumption that there is always the possibility that some students are selecting their major based on their passion rather than monetary value, this will be demonstrated between the level of popularity of the major and the annual median wage as well.

Section B: Sub-Group Identification

Since I am determining how much the most popular majors and least popular majors make in their respective occupations, my sub-group will identify the top 5 most popular public universities in Texas, the University of Texas at Austin, Texas A&M University, Texas Tech, University of Houston, and the University of Texas at Dallas[1]. and their top 3 most popular majors and salaries, as well as their 3 least popular majors and salaries. The reason for picking the top 5 public universities in Texas is to show how the graduates of these universities' salaries compare in their respective majors and fields of occupation.

Since it is speculated that the Aggie Network lands more jobs for those in their desired field of work, it is possible that their annual mean wage may be higher in comparison to other public universities in Texas, in addition to the great population of Aggies in Texas. Because I will be obtaining my data from the internet, it is also possible that the given information showing the economic success of graduates may be biased when showing the wage rates in each major's respective field, potentially skewing the data. I hypothesize that due to the prevalence of A&M's engineering population, their mean will be higher than those at other public Texas universities, with the University of Texas potentially having a greater annual mean wage for business majors due to the university's ranking as a top business school. In regards to Texas Tech, the University of Houston, and the University of Texas at Dallas, since it is likely that both of the aforementioned majors are quite popular, I believe there will be a fairly equal annual mean wage for both business and engineering majors.

Section C: Formal Hypothesis Statement

Popularity is an indicator of future financial success if there is a positive correlation between the level of popularity of the major and the annual median wage. I expect sub-group 2, Texas A&M, to have the strongest correlation between major popularity and annual median wage due to the prevalence of engineering majors and the Aggie Network.

Part 2: Literature Review

Section A: Previous Analysis of Major Popularity

As many analyses suggest, selecting a college major is a fairly important decision during a student's undergraduate studies and the selection has several influential factors. According to an analysis by Adel S. Aldosary and Sadi A. Assaf, these factors include but are not limited to interest in the major, family pressure, academic ability, the major's reputation, job salary, and the major's prestige. In another analysis, Georgetown University has a full report dedicated to the Economic Value Of College Majors. The analysis includes various summaries and data

explorations, drawing attention to STEM and business majors, which are both the most popular majors as well as two of the top highest-paying majors. Georgetown University's Center on Education and the Workforce also includes an interactive web tool to show the varying earnings among majors by state. For example, the median for a general bachelor's business degree earns $65,000 in Texas, $71,000 in California, and $65,000 in New York. However, the median for a general bachelor's engineering degree earns $81,000 in Texas, $84,000 in California, and $76,000 in New York. Georgetown's analysis findings further support my hypothesis that the more popular the major, the higher the future wages for students in those majors.

Works Cited:

Aldosary, Adel S., and Sadi A. Assaf. "Analysis of Factors Influencing the Selection of College Majors by Newly Admitted Students." *Higher Education Policy*, Elsevier, 24 Sept. 2002, www.sciencedirect.com/science/article/abs/pii/S095287339600013X.

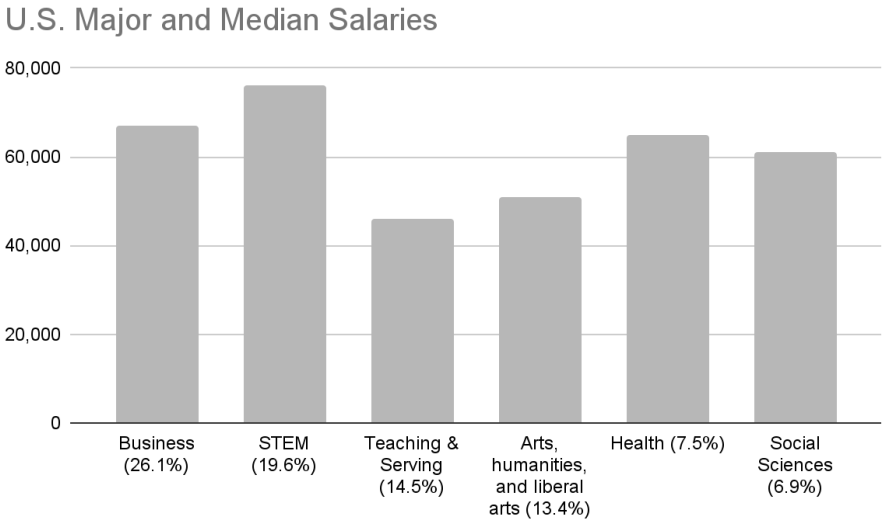
Carnevale, Anthony P., et al. "The Economic Value of College Majors." *CEW Georgetown*, 13 Sept. 2023, cew.georgetown.edu/cew-reports/valueofcollegemajors/.

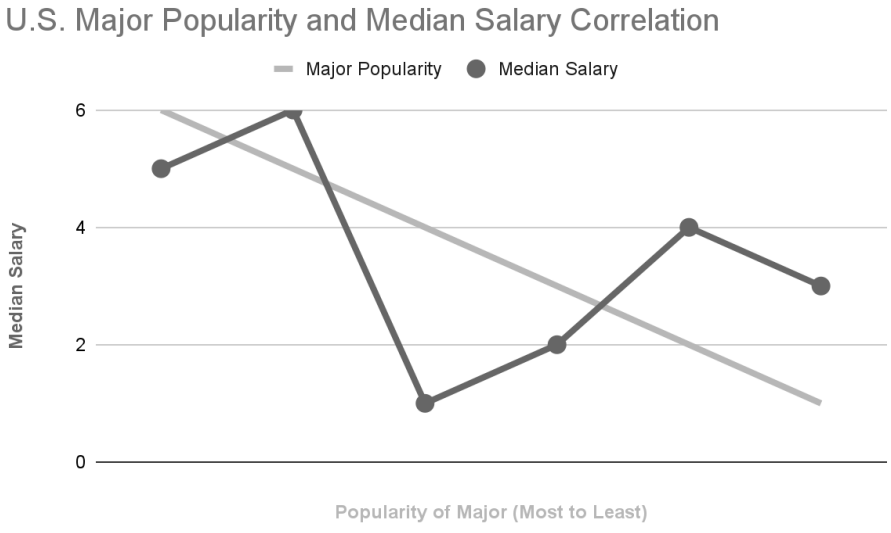
Part 3: Descriptive/Graphical Analysis

Section A: Descriptive Statistics

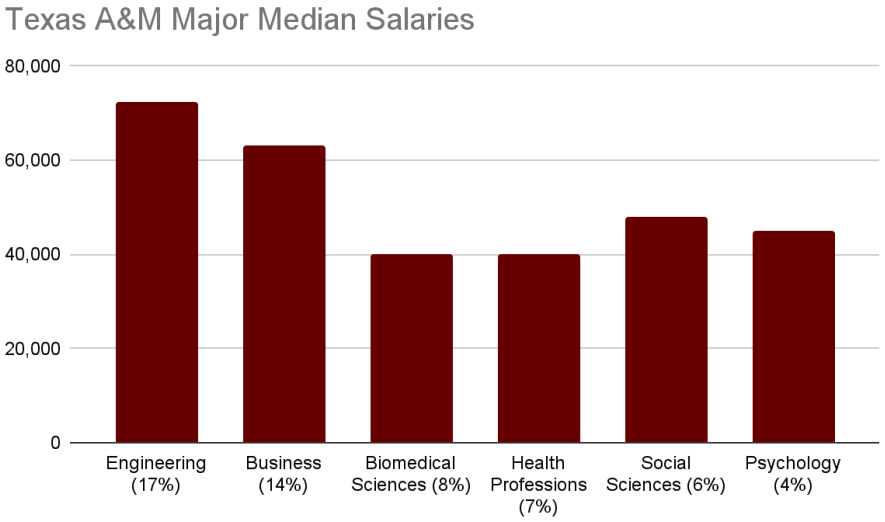
Contrary to my initial hypothesis, it seems that the popularity of a major does not seem to have an extremely strong correlation with the median salary in the overall distribution of majors in the United States. Nonetheless, there still appears to be a correlation of sorts, seeing the most popular major salaries in comparison to the less popular major salaries, however, not as strong as I initially hypothesized. There is significance in seeing that the more popular majors appear to have larger incomes than the less popular majors, for example, comparing STEM majors with those that are in Arts, Humanities, and Liberal Arts.

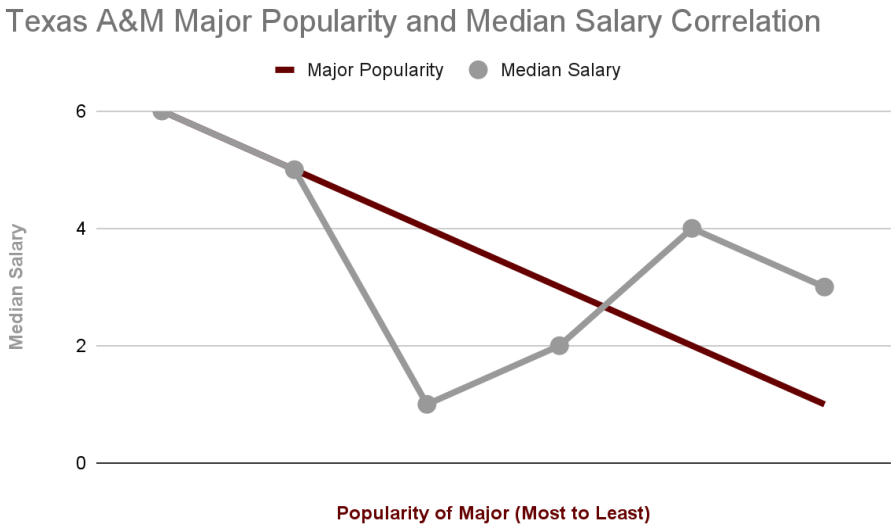
Major (By Supergroup)	% of Graduated Students (2022)	Median Salary
Business	26.1%	$67,000.00
STEM	19.6%	$76,000.00
Teaching & Serving	14.5%	$46,000.00
Arts, Humanities, and Liberal Arts	13.4%	$51,000.00
Health	7.5%	$65,000.00
Social Sciences	6.9%	$61,000.00



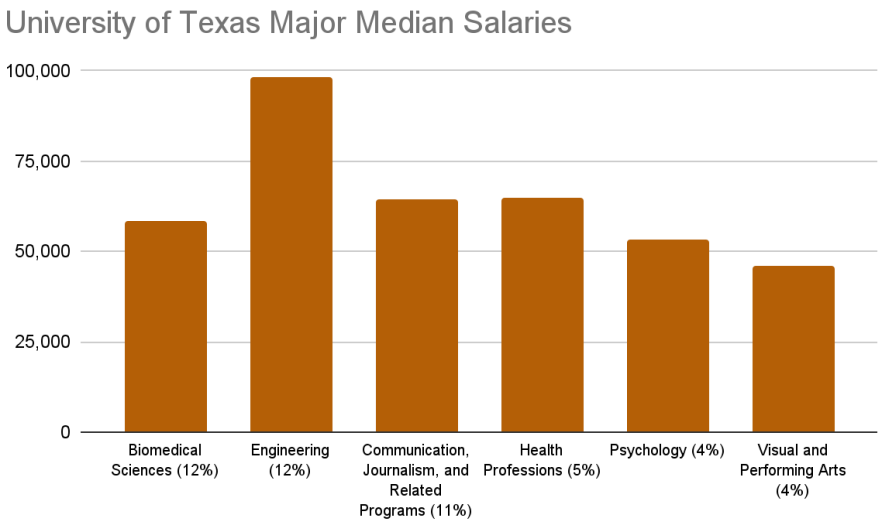


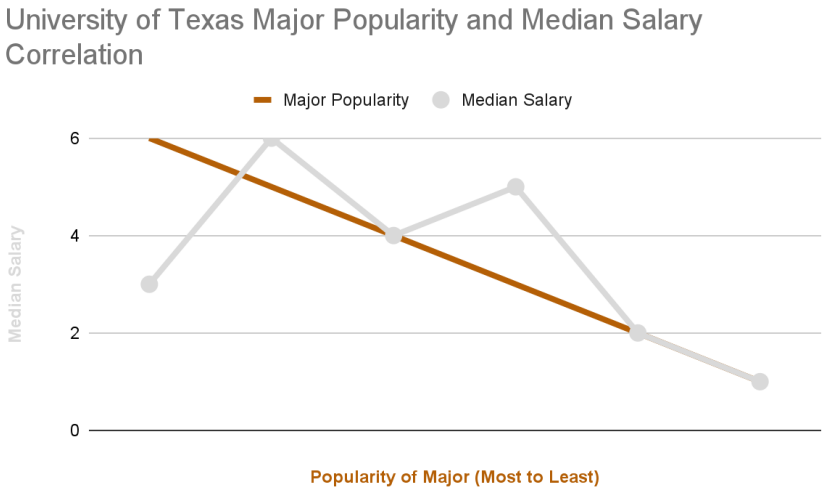
Texas A&M Major	% of Graduated Students (2022)	Median Salary
Engineering	17%	$72,500.00
Business	14%	$63,000.00
Biomedical Sciences	8%	$40,000.00
Health Professions	7%	$40,050.00
Social Sciences	6%	$48,000.00
Psychology	4%	$45,000.00



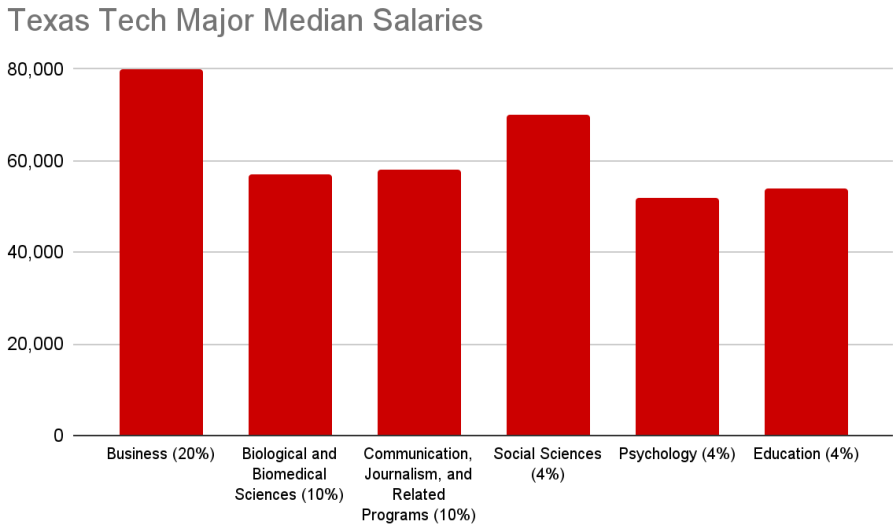


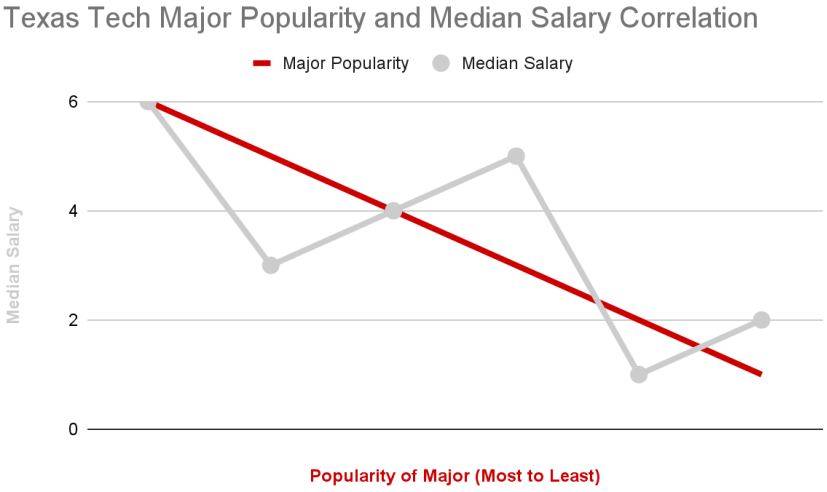
University of Texas Major	% of Graduated Students (2022)	Median Salary
Biomedical Sciences	12%	$58,291.00
Engineering	12%	$98,051.00
Communication, Journalism, and Related Programs	11%	$64,438.00
Health Professions	5%	$64,801.00
Psychology	4%	$53,350.00
Visual and Performing Arts	4%	$46,058.00



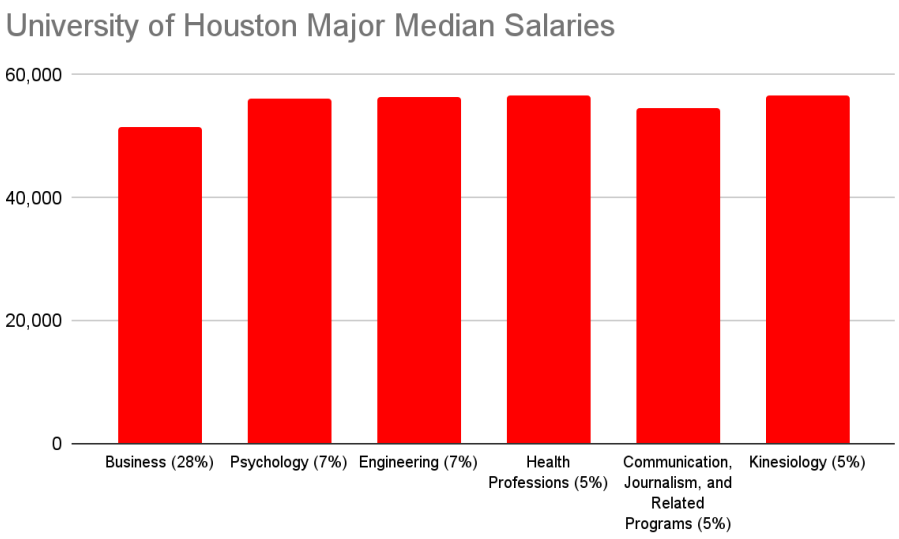


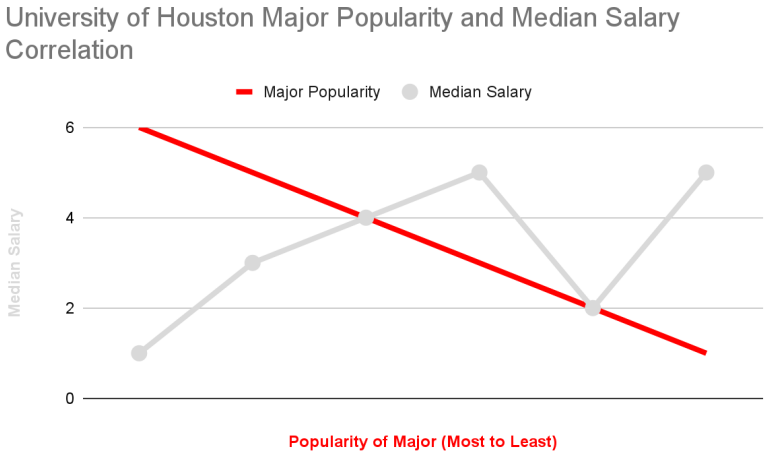
Texas Tech Major	% of Graduated Students (2022)	Median Salary
Business	20%	$80,000.00
Biological and Biomedical Sciences	10%	$57,000.00
Communication, Journalism, and Related Programs	10%	$58,000.00
Social Sciences	5%	$70,000.00
Psychology	4%	$52,000.00
Education	4%	$54,000.00



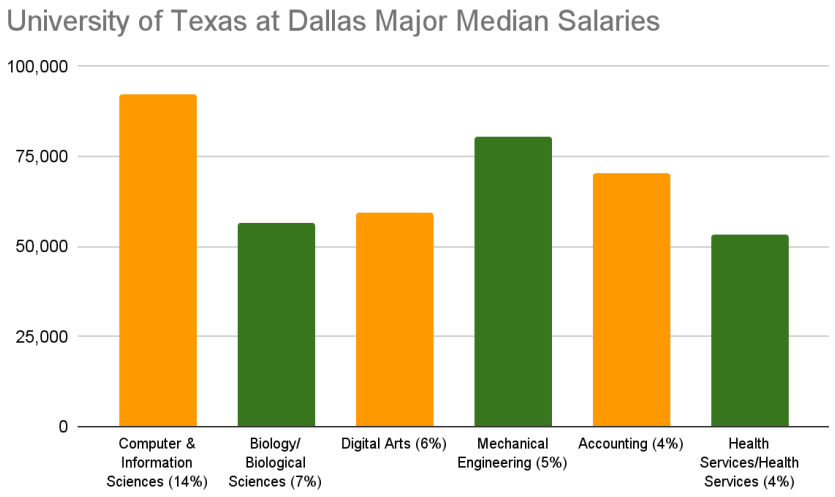


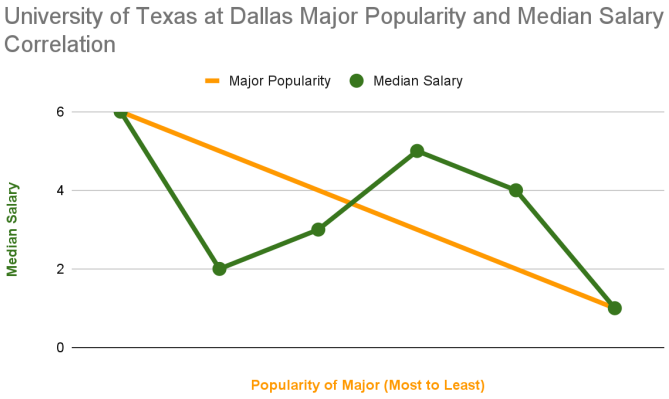
University of Houston Major	% of Graduated Students (2022)	Median Salary
Business	28%	$51,500.00
Psychology	7%	$56,149.00
Engineering	7%	$56,288.00
Health Professions	5%	$56,534.00
Communication, Journalism, and Related Programs	5%	$54,500.00
Kinesiology	5%	$56,534.00





University of Texas at Dallas Major	% of Graduated Students (2022)	Median Salary
Computer and Information Sciences, General	14%	$92,426.00
Biology/Biological Sciences, General	7%	$56,592.00
Digital Arts	6%	$59,248.00
Mechanical Engineering	5%	$80,450.00
Accounting	4%	$70,218.00
Health Services/Allied Health/Health Sciences, General	4%	$53,454.00



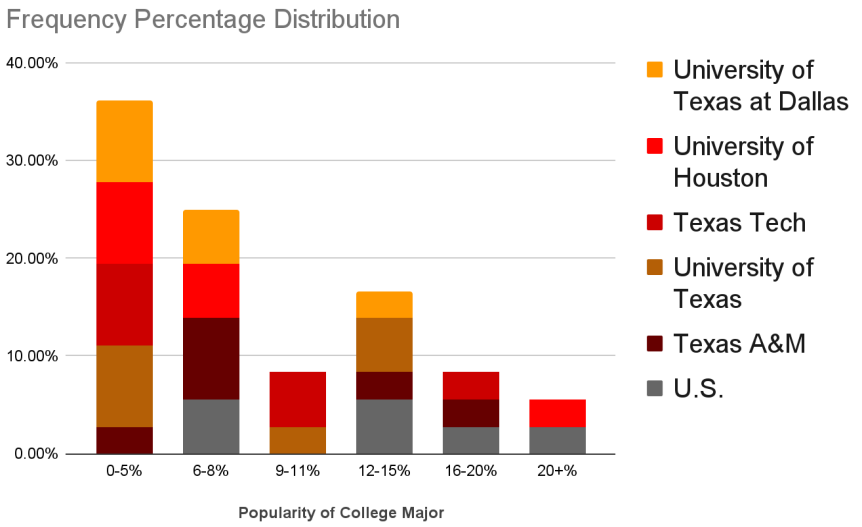


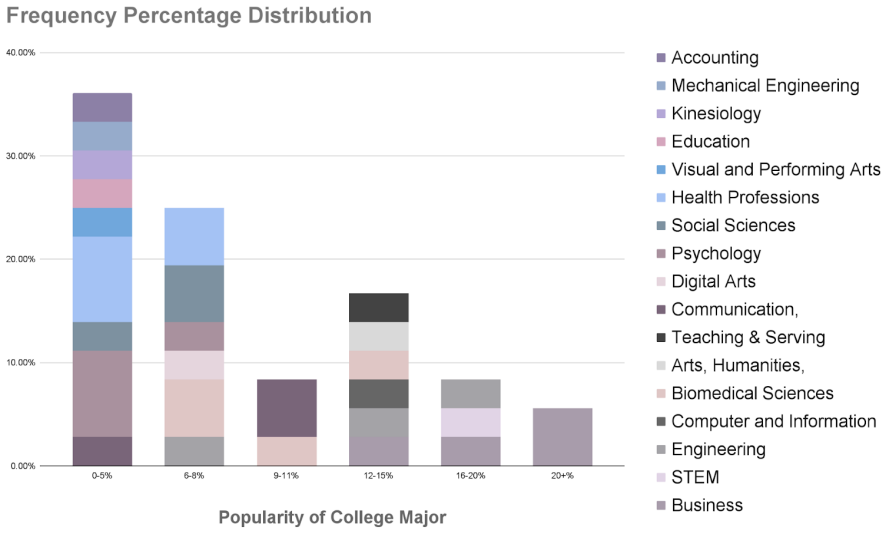
On average, the majority of the subgroups show some form of correlation between popularity and median salary. Texas A&M directly displays that the two most popular majors, Engineering and Business, have the highest median salaries. The University of Texas at Austin displays the two least popular majors, Psychology and Visual and Performing Arts, to have the lowest median salaries. Texas Tech's most popular major, Business, has the highest median salary. The University of Texas at Dallas shows its most popular major, Computer and Information Sciences, to have the highest median salary, and its least popular major, Health Services/Allied Health/Health Sciences, to have the lowest median salary.

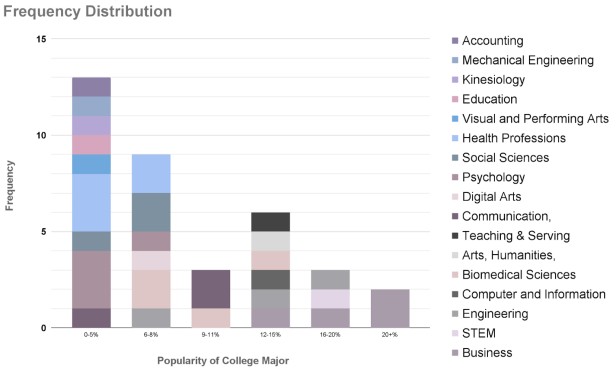
Some of the subgroups appear to show more promise as either the first or second most popular major displays the highest median salary of the six majors. This is true for Texas A&M, the University of Texas at Austin, Texas Tech, and the University of Texas at Dallas. For some of the majors that would be expected to do well financially, such as Biomedical Sciences, I'm led to believe that these majors are expected to pursue higher education such as graduate school in order to get the certifications necessary to allow for proper health practices and ultimately, a higher pay.

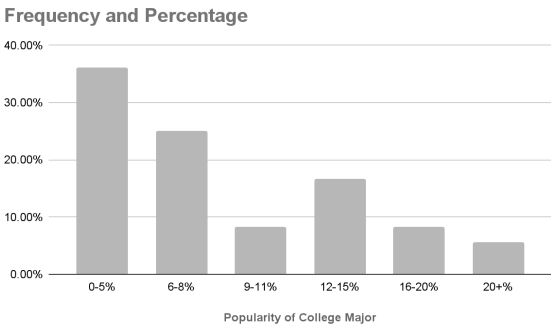
Section B: Frequency Distributions

The discrepancies amongst sub-groups can be seen in the frequency distribution. After cross-referencing the frequency percentages of the universities and their contributions to the popularity of college majors, observations for the overall United States colleges and the University of Houston are more likely to be Business majors, displaying the highest popularity percentage of the major. As for the other colleges, there appears to be a variance in the distribution of major popularity, opposite of my original assumption that the Texas universities would have a similar distribution in terms of the popularity of college majors in comparison to the United States Universities. If my initial assumption was correct, the second chart would have fewer majors, as they would align with the overall United States colleges' most popular and least popular majors, and the most popular majors would have the highest percentage frequency.


Frequency Percentage Distribution
40.00%
30.00%
20.00%
10.00%
0.00%
0-5%
6-8%
9-11%
12-15%
16-20%
20+%
Popularity of College Major
University of Texas at Dallas
University of Houston
Texas Tech
University of Texas
Texas A&M
U.S.

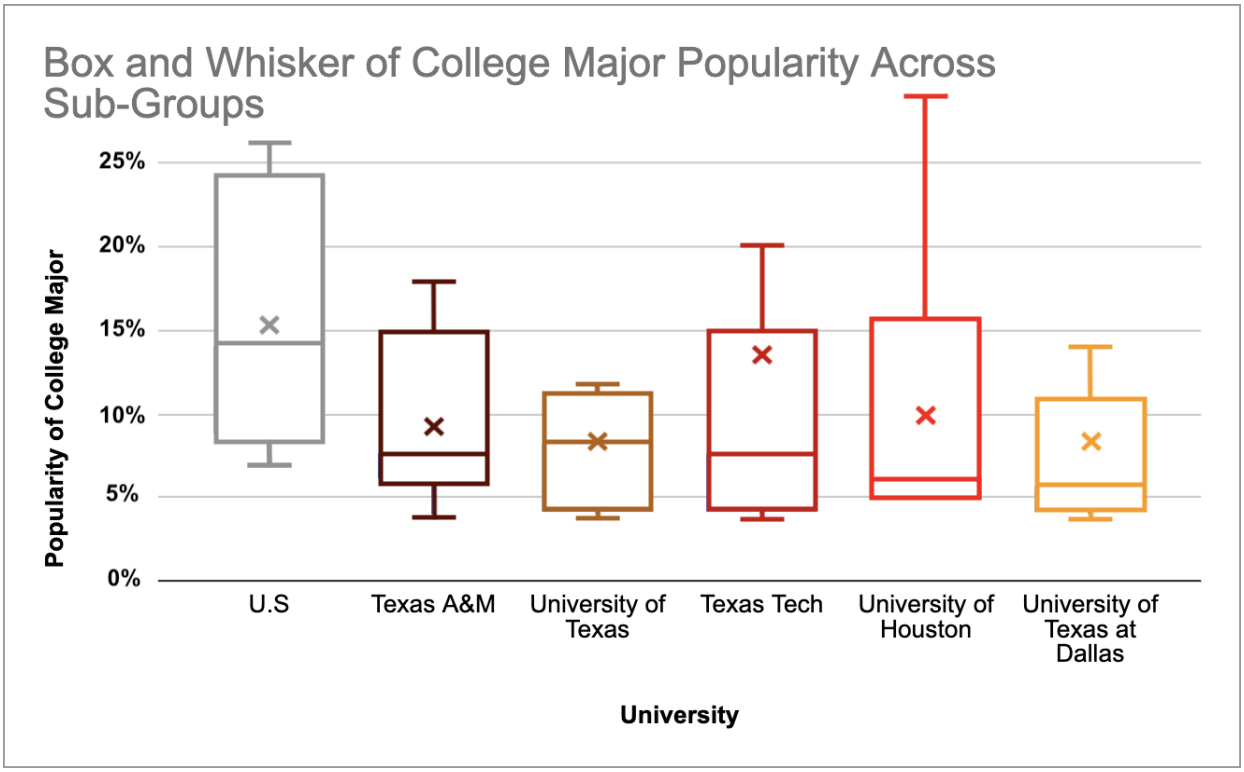

Frequency Percentage Distribution
40.00%
30.00%
20.00%
10.00%
0.00%
0-5%
6-8%
9-11%
12-15%
16-20%
20+%
Popularity of College Major
Accounting
Mechanical Engineering
Kinesiology
Education
Visual and Performing Arts
Health Professions
Social Sciences
Psychology
Digital Arts
Communication,
Teaching & Serving
Arts, Humanities,
Biomedical Sciences
Computer and Information
Engineering
STEM
Business


Frequency Distribution
15
10
5
0
Frequency
0-5%
6-8%
9-11%
12-15%
16-20%
20+%
Popularity of College Major
Accounting
Mechanical Engineering
Kinesiology
Education
Visual and Performing Arts
Health Professions
Social Sciences
Psychology
Digital Arts
Communication,
Teaching & Serving
Arts, Humanities,
Biomedical Sciences
Computer and Information
Engineering
STEM
Business


Frequency and Percentage
40.00%
30.00%
20.00%
10.00%
0.00%
0-5%
6-8%
9-11%
12-15%
16-20%
20+%
Popularity of College Major

Section C: Box and Whisker Plots

Box and whisker plots are utilized to highlight key differences in the mean, median, and variance of popularity of college major percentages. The U.S. Colleges, Texas A&M, Texas Tech, the University of Houston, and the University of Texas at Dallas are all skewed right, whereas the University of Texas is closer to normal distributions. It also shows that the University of Houston has the highest variance in the percentage distribution of college majors out of all of the universities, and the University of Texas has the lowest variance. This plot also suggests that my assumption that students lean towards more popular majors is incorrect, seeing as the median for the majority of the universities is skewed towards the less popular percentages. Since this is true, it is possible that students are choosing their majors based on passion rather than potential future wages.



Part 4: Single Sample Confidence Intervals and Hypothesis Tests

Section A: Confidence Intervals of Sample Means and Sample Variances

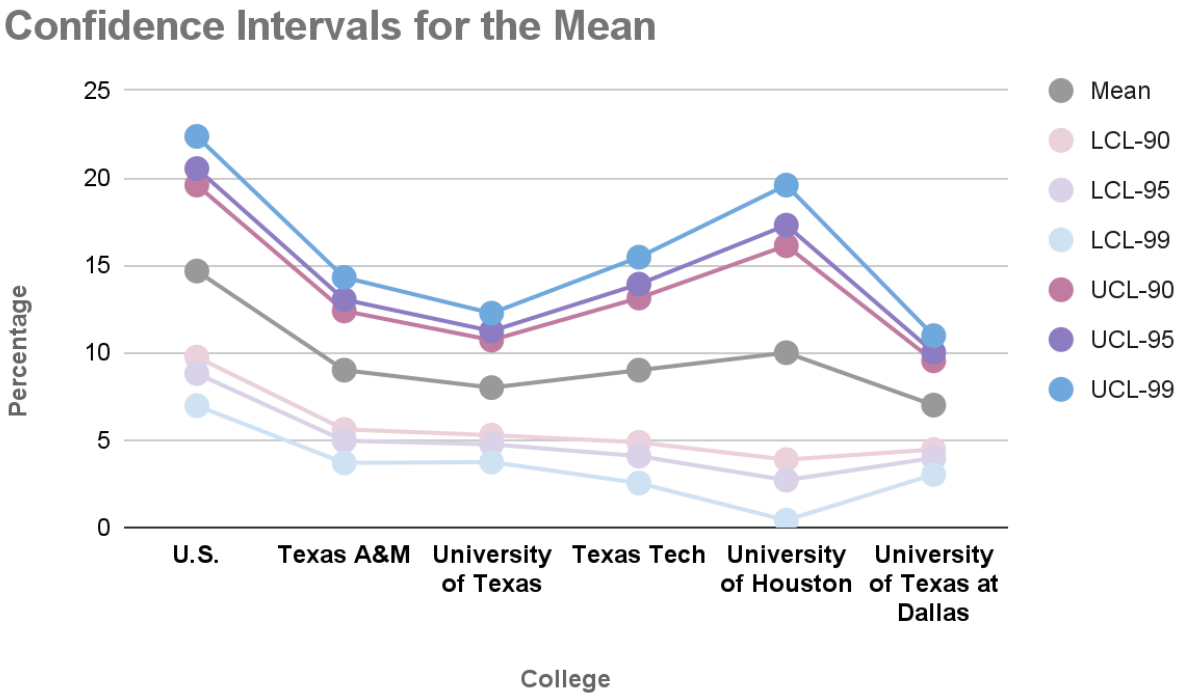
Means

For the overall and sub-group samples, confidence intervals of 90%, 95%, and 99% were calculated for the sample means of my criterion variable. For the overall sample means, 90% of observations fell between 9.7503 – 19.5896; 95% fell between 8.8078 – 20.5321, and 99% between 6.9658 – 22.3741.

Among my sub-groups, I expect 90% of the Texas A&M sub-group's observations to range from 5.6112 – 12.3887, 95% between 4.9621 – 13.0379, and 99% ranging from 3.6932 – 14.3067. The University of Texas observations are expected to range from 5.2805 – 10.7194 at 90% confidence, 4.7596 – 11.2403 at 95% confidence, and 3.7414 – 12.2585 at 99% confidence. For the Texas Tech sub-group, I expect 90% of the observations to range from 4.8732 – 13.1267, 95% to range from 4.0826 – 13.9173, and 99% to range from 2.5375 – 15.4624. The University of Houston observations are expected to range from 3.8785 – 16.1214 at 90% confidence, 2.7058 – 17.2941 at 95% confidence, and 0.4138 – 19.5861 at 99% confidence. For the University of Texas at Dallas sub-group, I expect 90% of the observations to range from 4.4636 – 9.5363, 95% to range from 3.97777 – 10.0222, and 99% to range from 3.02801 – 10.9719.

Confidence Intervals for the Mean

College	Mean (%)	LCL-90	LCL-95	LCL-99	UCL-90	UCL-95	UCL-99
U.S.	14.67	9.7503	8.8078	6.9658	19.5896	20.5321	22.3741
Texas A&M	9	5.6112	4.9621	3.6932	12.3887	13.0379	14.3067
University of Texas	8	5.2805	4.7596	3.7414	10.7194	11.2403	12.2585
Texas Tech	9	4.8732	4.0826	2.5375	13.1267	13.9173	15.4624
University of Houston	10	3.8785	2.7058	0.4138	16.1214	17.2941	19.5861
University of Texas at Dallas	7	4.4636	3.97777	3.02801	9.5363	10.0222	10.9719



This shows a larger difference in 90% to 99% confidence for the University of Houston and Texas Tech sub-group, whereas the University of Texas and the University of Texas at Dallas sub-group confidence intervals don't change as much. This means the University of Texas and the University of Texas at Dallas observations cluster closer to their mean than the other groups or the overall sample.

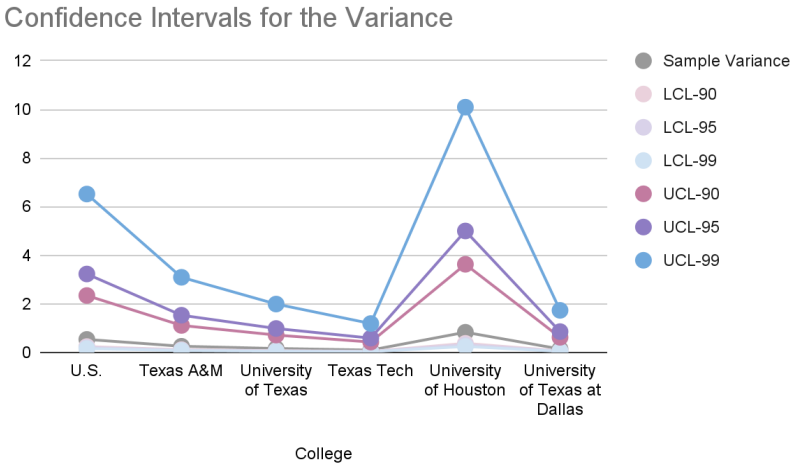
Variance

For the sample variances, the overall data can be expected to have 90% of the observations' variances ranging from 0.2424 - 2.3428, 95% of the data 0.2091 - 3.2287, and 99% of the data 0.1602 - 6.5179. Among the sub-groups, the Texas A&M sub-group shows 90% of the variances ranging from 0.1150 - 1.1116, 95% ranging from 0.0992 - 1.5319, and 99% ranging from 0.0760 - 3.0926. The University of Texas sub-group shows 90% of the variances ranging from 0.0741 - 0.7158, 95% ranging from 0.0639 - 0.9865, and 99% ranging from 0.0489

- 1.9915. The Texas Tech sub-group shows 90% of the variances ranging from 0.0442 - 0.4277, 95% ranging from 0.0382 - 0.5895, and 99% ranging from 0.0293 - 1.1901. The University of Houston sub-group shows 90% of the variances ranging from 0.3753 - 3.6273, 95% ranging from 0.3238 - 4.9987, and 99% ranging from 0.2481 - 10.0913. The University of Texas at Dallas sub-group shows 90% of the variances ranging from 0.0644 - 0.6227, 95% ranging from 0.0556 - 0.8582, and 99% ranging from 0.0426 - 1.7325.

Confidence Intervals for the Variance

College	Sample Variance	LCL-90	LCL-95	LCL-99	UCL-90	UCL-95	UCL-99
U.S.	0.537	0.2424	0.2091	0.1602	2.3428	3.2287	6.5179
Texas A&M	0.255	0.1150	0.0992	0.0760	1.1116	1.5319	3.0926
University of Texas	0.164	0.0741	0.0639	0.0489	0.7158	0.9865	1.9915
Texas Tech	0.098	0.0442	0.0382	0.0293	0.4277	0.5895	1.1901
University of Houston	0.831	0.3753	0.3238	0.2481	3.6273	4.9987	10.0913
University of Texas at Dallas	0.143	0.0644	0.0556	0.0426	0.6227	0.8582	1.7325



The confidence intervals for the sample variances show the largest variance disparity between confidence intervals is for the University of Houston sub-group and the lowest for the Texas Tech sub-group. These confidence intervals further illustrate that the Texas Tech observations are the most clustered, while the University of Houston observations are the most dispersed.

Section C: Single Sample Hypothesis Test Means

Means

Single sample hypothesis tests were conducted with the null hypothesis that subgroup means are equal to the overall mean. For all of the sub-groups, the null hypothesis failed to reject the null hypothesis, which means it was equal to the overall. This means that in all of the sub-groups, 99%, 95%, and 90% of the data displays means equal to the overall. Essentially, the sub-groups model the overall sample in terms of equal means at all confidence levels, indicating that the overall sample means are highly influenced by the sub-groups.

Single Sample Hypothesis Tests for Means: Two-Tailed Tests

College	Mean (%)	T-Stat	T-Crit 90	Concl 90	T-Crit 95	Concl 95	T-Crit 99	Concl 99
U.S.	14.67	**-1.0743**	2.0151	FTR	2.5706	FTR	4.0321	FTR
Texas A&M	9	**-1.1698**	1.6449	FTR	1.9599	FTR	2.5758	FTR
University of Texas	8	**-1.0601**	1.6449	FTR	1.9599	FTR	2.5758	FTR
Texas Tech	9	**-0.6986**	1.6449	FTR	1.9599	FTR	2.5758	FTR
University of Houston	10	**-0.4121**	1.6449	FTR	1.9599	FTR	2.5758	FTR
University of Texas at Dallas	7	**-0.7104**	1.6449	FTR	1.9599	FTR	2.5758	FTR

Part 5: Two-Sample Confidence Intervals and Hypothesis Tests

Section A: Pair-wise Hypothesis Tests of Equal Variances

Ten pair-wise hypothesis tests were conducted for equal variance between the sub-groups: Texas A&M-University of Texas, Texas A&M-Texas Tech, Texas A&M-University of Houston, Texas A&M-University of Texas at Dallas, University of Texas-Texas Tech, University of Texas-University of Houston, University of Texas-University of Texas at Dallas, Texas Tech-University of Houston, Texas Tech-University of Texas at Dallas, and University of

Houston-University of Texas at Dallas. All ten tests result in a P-Value greater than 0.95 but two, Texas A&M-UT Dallas and Texas Tech-UT Dallas, rejecting the null hypothesis that the variances are equal between any two subgroups besides the two pairings previously mentioned.

Variance Test F-Stat

	Texas A&M	**University of Texas**	**Texas Tech**	**University of Houston**
University of Texas	0.003604712			
Texas Tech	0.00093537	0.000934157		
University of Houston	0.003604712	0.0036	0.000934157	
University of Texas at Dallas	0.003613134	0.0036084112	0.00093633971	0.0036084112

P-Values

	Texas A&M	**University of Texas**	**Texas Tech**	**University of Houston**
University of Texas	2.738937828			
Texas Tech	6.650694104	5.953055494		
University of Houston	9.479306558	6.473598354	7.023004348	
University of Texas at Dallas	0.382762689	3.177442387	0.9311260457	2.620848554

Conclusion: Reject if P-value is greater than 0.95 or less than 0.05

	Texas A&M	**University of Texas**	**Texas Tech**	**University of Houston**
University of Texas	Reject			
Texas Tech	Reject	Reject		
University of Houston	Reject	Reject	Reject	
University of Texas at Dallas	Fail to Reject	Reject	Fail to Reject	Reject

The pair-wise tests are used to determine which, if any, of the groups have similar patterns of sample variance. Each of the P-values are well above the calculated F-statistics, indicating that there is close to no likelihood that any sub-group would produce equal variation in their popularity of college major. Each of the sub-groups has a unique pattern of variance which will influence the pair-wise hypothesis tests for differences in means.

Part 6: ANOVA Tests

Section A: Single-factor ANOVA

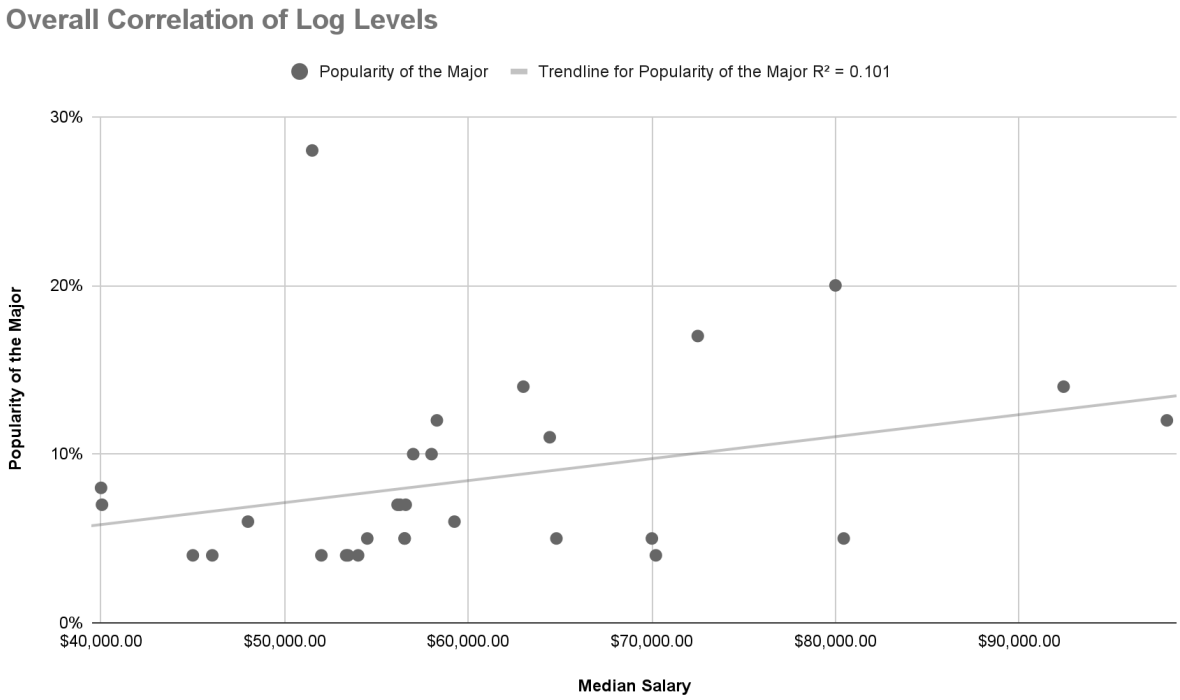
After running a single-factor ANOVA test across the 5 sub-groups, the ANOVA test produced a P-value of 0.919348604, which is slightly more than the cutoff point of 0.05. There is some confidence in the null hypothesis that the means across groups are equal, only due to the fact that this was shown in previous tests, indicating similar means across the sub-groups. This indicates that there is not a significant amount of variation across sub-groups.

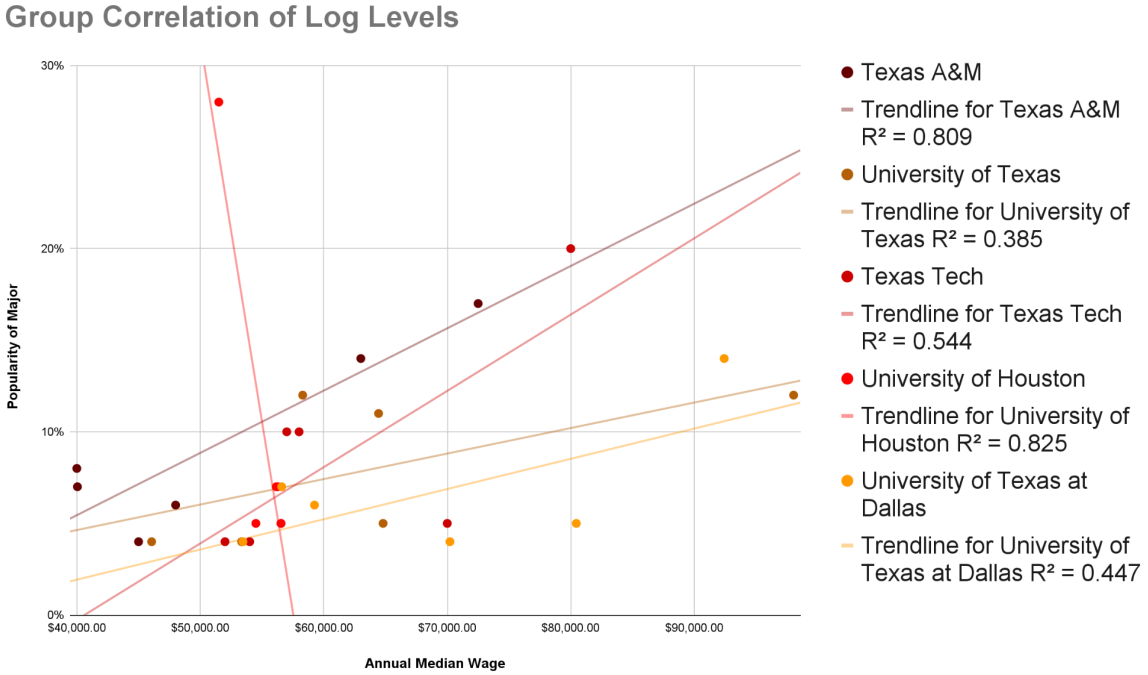
This appears contrary to the earlier test result that suggested that the majority of the sub-groups had a great number of variances besides two; this similarity was determined by testing for a difference amongst variances. The utilization of ANOVA's null hypothesis instead tests for collective equality across subgroup means. The varying results may possibly indicate that the similarity in means of the subgroups outweighs the dissimilarity of the variances between the sub-groups, meaning there is the possibility of an overall greater influence on the entire sample.

Part 7: Correlation Analysis

The criterion and predictor variables are converted to log levels for correlation analysis.

Section A: Scatter Plots and Trend Lines





Section B: Tests of Significant Correlation

For the overall and sub-group samples, the null hypothesis that there is zero correlation between the predictor and criterion variable, H0 : $\rho\rho$XY = 0, was tested with an alpha of 0.05 against the student's t-distribution.

Group	Correlation	R-Squared	Sample Size	zr
U.S.	0.3211872644	0.103	6	0.3329703795
Texas A&M	0.8996790917	0.809	6	1.47053306
University of Texas	0.6203064576	0.385	6	0.7255030609
Texas Tech	0.737731234	0.544	6	0.9454829254
University of Houston	-0.9082679617	0.825	6	-1.517539792
University of Texas at Dallas	0.6685590619	0.447	6	0.8081330337

Tests of Individual Correlation Significance H0: p=0 Two-Tailed Tests, alpha = 0.05

Group	t-calc	t-Critical	Conclusion
U.S.	0.67829744	1.100665317	Fail to Reject
Texas A&M	4.12153267	0.1342630663	Reject
University of Texas	1.58170374	0.5276552326	Reject
Texas Tech	2.18555622	0.3540660198	Reject
University of Houston	-4.34261489	0.1211797513	Fail to Reject
University of Texas at Dallas	1.79803032	0.4544416755	Reject

For the U.S. and University sub-group, I fail to reject the null hypothesis, indicating that there is not a significant level of correlation between the two variables. On the other hand, the Texas A&M, University of Texas, Texas Tech, and University of Texas at Dallas sub-groups

reject the null, showing there is a level of correlation for them. This is true seeing the charts showing correlation, it is apparent that these “negative” sub-groups clearly seem to indicate a trend between major popularity and annual median salary. This shows that the correlation between the popularity of college major and the annual median wage in the Texas A&M, University of Texas, Texas Tech, and University of Texas at Dallas sub-groups could be strong enough that it offsets the low correlations presented in the U.S. and University of Houston subgroup. This is true to my original theory, that the higher the future annual median wage, the more likely a student is to pick that major.

One of the possible reasons for the seemingly lack of correlation between the overall U.S. and the University of Houston, could be due to the generalization of the data rather than specificity when it came to the popularity of the given majors and their respective annual median wages. I noticed when collecting the data that all of the University of Houston’s annual median salaries were extremely similar to each other, potentially altering the correlation of the data results. Since the other universities had a greater difference between each major’s annual median salary, it is possible that this had a positive impact on the correlation between popularity of a major and its respective future salary.

Part 8: Conclusions, Discussions, and Limitations

Before drawing final conclusions, it is important to note that I got all of my data from various sources, which were limited as well. Many of my data sets were derived from each University’s statistic data, therefore, there is likely a possibility that this data could be skewed in favor to illustrate the university’s success rate for students’ future endeavors. However, to compact this potential bias, I also utilized data from third-party sources as well, though this information was limited as well. Although human error is always a potential factor, I did my best to gather as much information as I could from reputable sources. My biggest issue was the annual median wage from the University of Houston. All of its majors seemed to have nearly the

same annual median salary, which seemed potentially inaccurate but I derived the data straight from the University of Houston's website.

Another issue I encountered was the generalization of majors, for example, some schools reported major specific fields, such as Mechanical Engineering, while others categorized them under one umbrella such as General Engineering, which could also potentially affect the overall data interpretation. Ultimately, I did my best to navigate these interpretations, and there was still plenty of data that appeared to prove my initial hypothesis, popularity is an indicator of future financial success if there is a positive correlation between the level of popularity of the major and the annual median wage.

According to the correlation calculations, 4 of 5 sub-groups appeared to have a positive correlation between college major popularity and future annual earnings. As I hypothesized, sub-group 2, Texas A&M, had the strongest correlation between major popularity and annual median wage, with a correlation of 0.8996790917. Texas Tech was the runner-up with a correlation of 0.737731234, which was surprising. The second runner-up was the University of Texas at Dallas with a correlation of 0.6685590619.

Despite the potential inaccuracies of some of the major's potential future incomes, the basis of this project could be expanded and integrated into a data analytic algorithm, much like the Georgetown model. As previously mentioned, the data set was compiled by manually entering observations from each university's data and statistics. It would be great to have a database that could show you all of your state's universities with all majors' potential future salaries. I am confident it would influence many prospective students' college decisions as well as college major decisions. With this being said, I thoroughly enjoyed this project and the knowledge I have accumulated throughout this project have proved especially intriguing, and I'm enthusiastic to share what I've learned with future prospective college students.

References

1: 5 Popular Public Colleges in Texas

https://www.bestcolleges.com/united-states/texas/public-colleges/

Works Cited:

Aldosary, Adel S., and Sadi A. Assaf. “Analysis of Factors Influencing the Selection of College Majors by Newly Admitted Students.” Higher Education Policy, Elsevier, 24 Sept. 2002, www.sciencedirect.com/science/article/abs/pii/S095287339600013X.

Carnevale, Anthony P., et al. “The Economic Value of College Majors.” CEW Georgetown, 13 Sept. 2023, cew.georgetown.edu/cew-reports/valueofcollegemajors/.