APPLIED STATISTICS AND VISUALIZATION FOR ANALYTICS: **REDESIGN PROJECT**

MOST POPULAR COLLEGE MAJORS

BY GROUP 2:

AUSTEN HENDRICKSON

VENKATA SRI ATHULYA GOPISHETTY

WEI WANG

Introduction

Nowadays, choosing which major to pursue is a huge challenge for many students, because it affects not only their knowledge acquisition, but also their future careers. Existing research shows that high school students generally lack an understanding of what is expected for college majors. Lisa Y et al. (2002) pointed out that more than half of the students lacked insight into what would be expected of them in the majors they wished to apply for in the future. Additionally, Jamin D. Speer (2017) showed that there are large gender gaps in college majors. Most importantly, career paths are generally dictated by which major is chosen. For these reasons, our group decided to explore the most popular college majors for different students in order to help them make more informed decisions about which major they may want to take. Unfortunately, the chart designed by Owlguru (owlguru.com/most-popular-college-majors-for-men-and-women/) is confusing and information cannot be easily interpreted. To make the information clearer, we have redesigned the chart so that it looks more informative. However, there will be some disconnects between the redesigned charts we produced and the original visual. This is because we could not find the data set for the year that the visual is referencing but we were able to find the data for the following year.
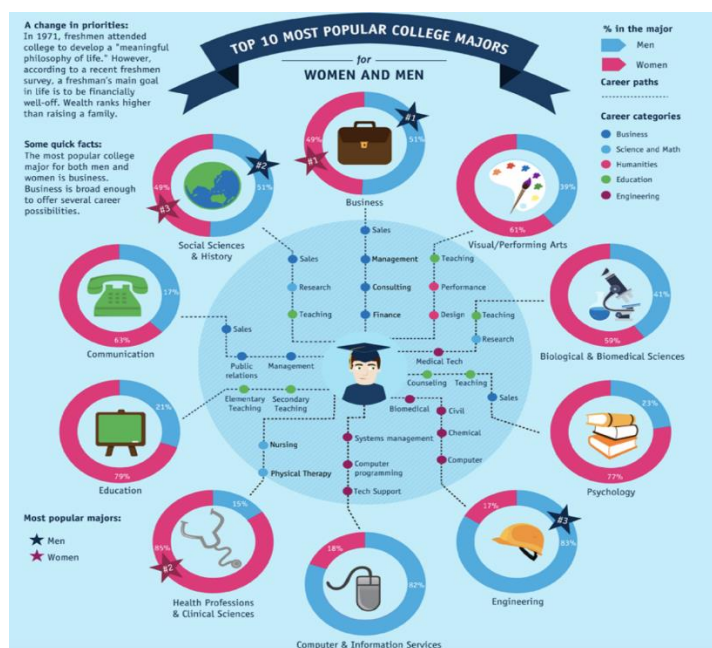
The Bad Graph

Figure 1. Top 10 most popular college majors for male and female (owlguru)

Figure 1 is the original graph to be redesigned. The original graph designed by Owlguru institute demonstrates the top10 most popular college majors for males and females. As shown above, every doughnut chart represents a major category, the blue part of every doughnut chart describes the percentage of male students in the major category and the red part shows the percentage of female students. In order to make the information more accurate, the figure is also marked with specific numbers. Besides, the dotted line represents career path and the color dots represent different career categories. The dots on the dotted line represent different majors in this major category, and the color dots

2

correspond to different career categories. In addition, the graph also marked the top three most popular majors for men and women with blue and purple stars. As shown in the graph, for men, the top three most popular majors are Business, Social Science& History, and Engineering. For Women the top three are Business, Health professions & Clinical Sciences, and Social Science& History. However, these classifications are not real college majors, but instead are major categories. Finally, this graph also gives some quick facts, like business is the most popular college major for both men and women and that most freshmen go to college to be financially well off.

The strengths of the graph:

· **Beautiful color scheme**
The rich and beautiful color scheme is the most intuitive advantage of this graph. The graph mainly adopts three colors of red, blue and green, with blue representing men and red representing women. In general, the design is attractive.
· **Abundant elements**
The design of the graph uses different kinds of elements, such as a circle chart to represent the proportion, colored dots to indicate the career category, dashed lines to represent the career paths, and colors to indicate the gender, which makes it easy to compare the differences between groups of data.
· **Informative**
This graph integrates multiple pieces of information such as gender, college majors, major categories, career categories, career path, etc., and displays it in a visual form. Additionally, information in the graph is explained and summarized through legend.

The weakness of the graph:
Despite the above advantages, this visualization also has some disadvantages.
· **The subject is unclear**
The title of this graph is Top 10 most popular college majors for males and females, but the graph lists the most popular major categories. Designers ignored the distinction between college major and major categories, resulting in confusing information across the entire graph.
· **Ignoring important information.**
The designer did not rank the Top 10 most popular college majors, and with the current information in the visual we only know what the top three majors for each gender are. Another issue is that the visual uses the total number of students to decide what the Top10 majors are but that does not mean the Top10 majors for males are the same as females. Furthermore, the title adds to this confusion and is not consistent with the actual information displayed. The visual also fails to offer any insight as to why the Top10 majors are what they are which is a question many viewers could have.
· **Information redundancy.**
Firstly, the layout of the graph is so crowded that it is difficult for readers to visually get the information they need from the graph. Secondly, the central part of the graph contains 10 large circles and the cartoon icon corresponding to different college majors. Although it makes the picture more vivid, the large area of the icon makes other information more crowded. In addition to being crowded, over-complicated elements such as colored dots and dotted lines weaken the readability of the whole picture, which is another challenge for readers. In fact, the information they represent is not very important. Finally, the legend in the upper left corner is redundant because the quick facts are obvious.

Redesign schema

The original graph could not convey all of the information at the same time clearly. According to the original graph, we extracted the key information it intended to show career(major) category, college major and gender, etc. Based on this, we also considered the reasons for choosing a major, such as employment rate, salary and other related factors and redesigned it using 5 different aspects:

1. Major category & Major
2. Major Category & Gender
3. Gender & Major
4. Salary
5. Employment

Redesigned Graph

Figure2. Number of males & females per field

The first redesign of the visual was developed using R in the statistical computing environment to display the popularity of different major categories among men and women. To make the data more intuitive, we used a bar chart to show the gender composition for the most popular major categories. The X-axis is the gender, and the Y-axis is the number of students. Red represents female and blue represents male. This chart covers 16 major categories, and is so brief and simple that we can easily compare which categories are most popular as well as the distribution of men and women in each category. This provides a much clearer picture than the original visualization did and shows some interesting information. For instance, the most popular major category overall is Biology & Life Sciences however it is not the most popular for either men or women, those being Social Science and Education respectively. Unfortunately, the chart also has some disadvantages, for example there is no way to split the initial total amount of students in each category by gender and so this redesign is only representative of those that graduated. Also, it can only directly compare the differences between males and females based on the major category, but we

still need to know the relationship between popular majors within those categories, so we will move on to the next step of designing, showing popular majors in various major category.
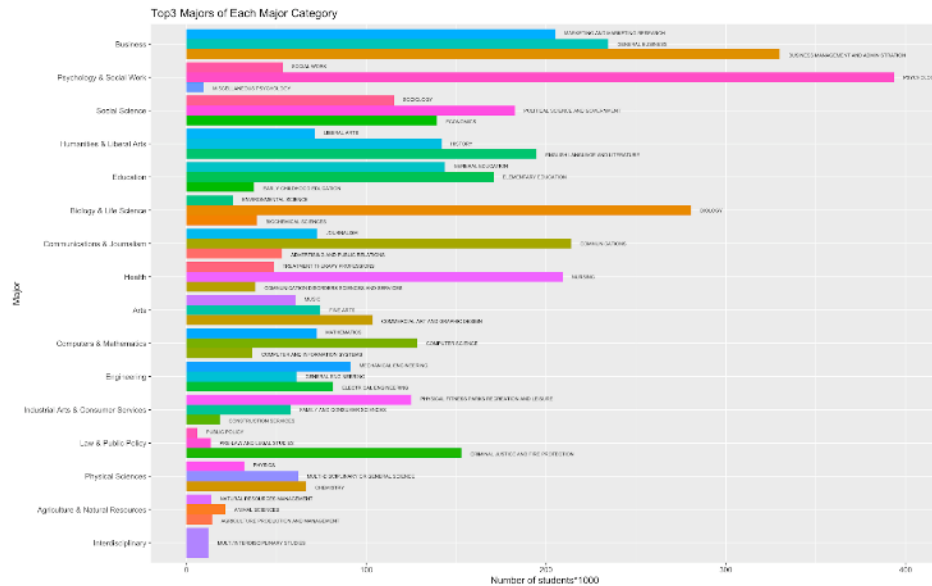

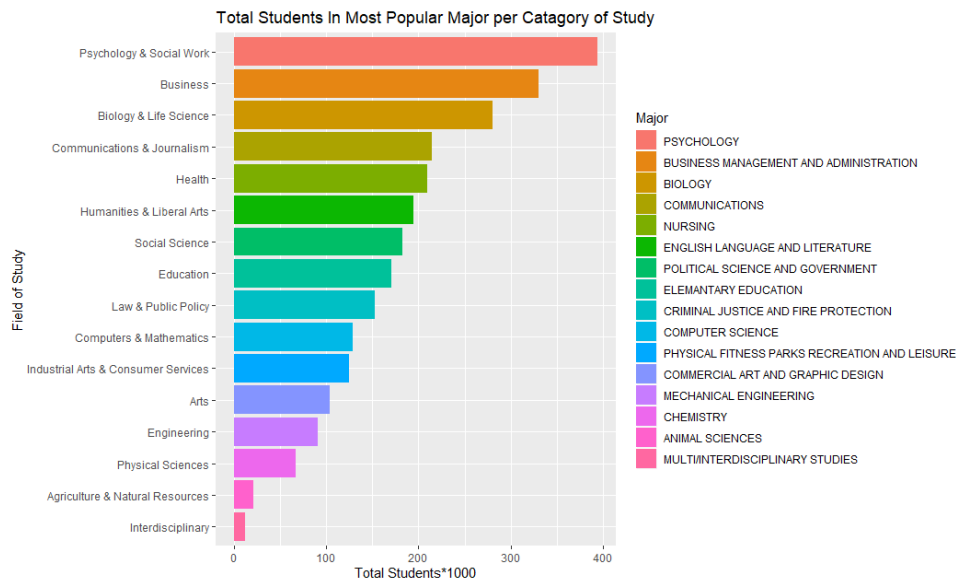**Figure 3. Top 3 Majors of each Major Category**

Figure 4. Top1 major of each Major Category

To show the most popular majors in each major category we designed figure3. Figure3 shows the Top3 most popular majors in each major category with a bar chart. Where, the X-axis represents the number of students, and the Y-axis represents major categories and individual majors. Although this graph contains the information about the most popular majors, it contains too many colors and is not easy to make comparisons between majors. This makes it difficult for readers to get the information. Figure 3 also cannot show the difference between the popular majors among different genders. In order to make the most popular major in each category clearer we designed figure 4. Figure 4 is another horizontal bar chart which shows the most popular major in each category. Where, the X-axis represents the number

of graduates, the Y-axis represents the major categories, and different colors are used to directly relate each major with its associated category. Figure 4 is much more concise and intuitive than figure 3 but as we mentioned before it still does not include the gender differences between majors. In order to show which majors were most popular based on gender we decided to design two more graphs.
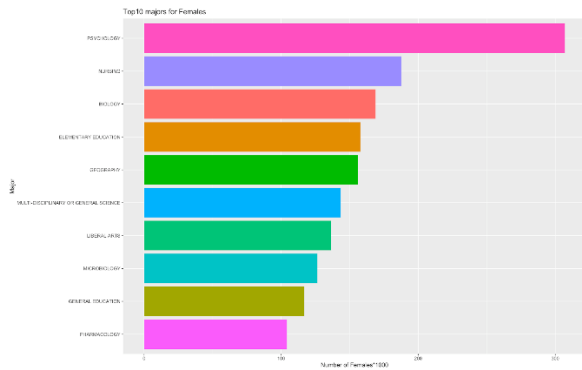

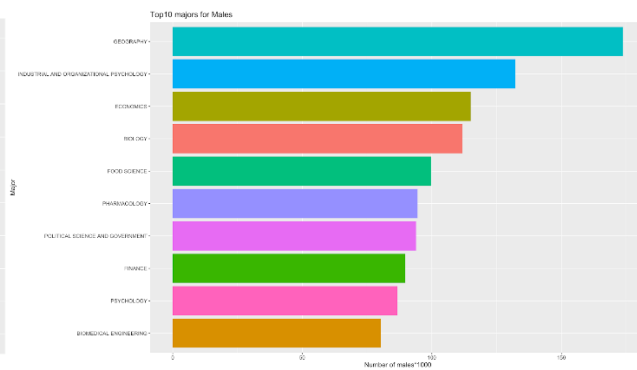Figure 6. Top10 majors for Males.   Figure 6. Top10 majors for Females

In order to illustrate what the Top10 majors based on gender were, we designed Figures 5 and 6. Figures 5 and 6 shows the most popular college majors in two independent bar charts and ranks them lowest to highest. Where, the X-axis represents the number of students and the Y-axis represents college majors. This side by side comparison provides more useful and intuitive information than the previous diagrams, it allows students to which majors tend to be more popular among different genders. In addition, this graph directly shows the original purpose of our design, to find the top ten majors based on gender. Through our design process, readers can visualize what the most popular majors are from multiple perspectives. For instance, one can now see what the most popular major categories are overall, by gender, or by individual major. The reason we designed numerous visualizations is that our group found that the real purpose of designing a data visualization is not only to display the data but also to improve the readability and clarity of the data in order to assist the reader's judgment. In the next part, we will use data visualization to analyze some of the reasons for major's popularity and what factors we can use to judge the major chosen trends
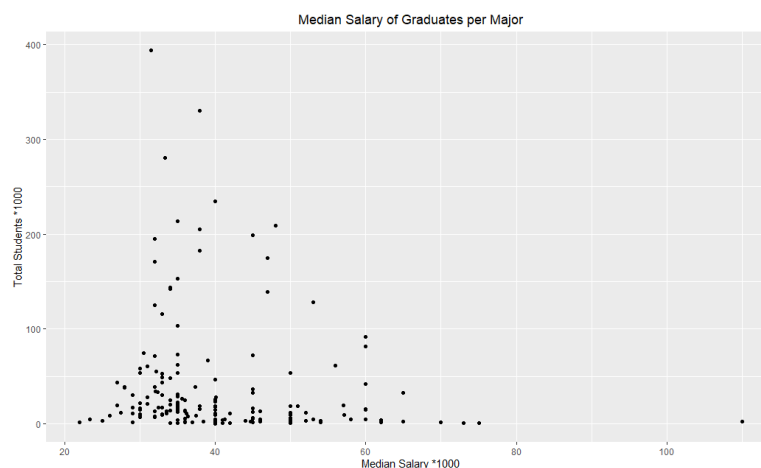

Figure 7. Correlation between median salary & total graduates

6

As mentioned previously one question that probably pops into everyone's mind is why are these the most popular majors, something that the initial design offered no insight to. In order to provide some context as to why certain majors were more popular, we decided to look at how the total number of students related to Salary and Employment. Figure 7 describes the relationship between the Median Salary of graduates per major and the number of graduates in that major with a scatter graph. Where, the X-axis represents the Median Salary and the Y-axis represents the number of graduates. For the purpose of seeing if there is a relationship between number of students in a major and the salary of those who graduate it is not important to show the individual majors. This also would've added a lot of clutter to the graph and so we decided not to label which major corresponded to what dot. It can be seen from the figure that there is no linear correlation between the two variables. This result was surprising, because our initial assumption was that salary could be one of the motivations behind choosing a certain major and that a higher salary would result in higher popularity. Due to our surprise we decided to re-analyze the relationship from another angle.
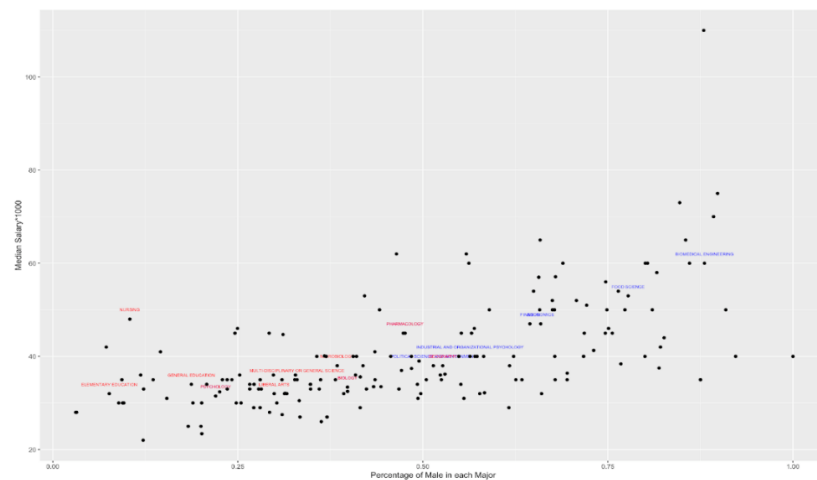

Figure 8. Scatter plot for median salary& percentage of Male

In order to further find out the relationship between salary and popularity of college major, we conducted a correlation analysis between the percentage of men in each major with Median Salary, and designed Figure 8 based on the results. Using a scatter diagram, Figure 8 describes the relationship between graduate Median Salary and the popularity of majors (Percentage of Males). Where, the Y-axis represents the Median Salary, the X-axis represents the popularity of different majors for males as a percent of the total. For some additional reference the red words represent the top10 majors for women, and the blue represents the top10 majors for men. As can be seen from the graph, there is a slightly positive correlation between popularity of majors as a percentage of men and the Median Salary, and that median salary of majors mostly comprised of men is generally higher than that of majors mostly comprised of women. We guess the reason may be same as Carmichael, S. G. (2017) pointed out that many college majors that lead to high-paying roles in geography and industrial and organization psychology are male dominated, while majors that lead to lower-paying roles in psychology and nursing tend to be female dominated, placing men in higher-paying career pathways. From this information one can see that men tend to choose majors with higher paying salaries.
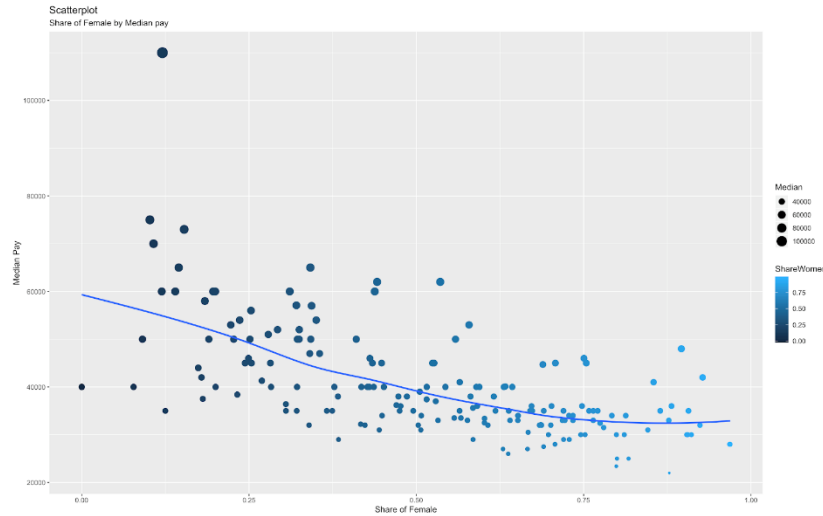
Figure 9. Scatter plot for median salary & percentage of females per major

We also decided to look at the correlation between average salary and the percent of women in each major. Figure 9 uses Ggplot2 with LOESS regression technique, and a trendline is fitted to show the relationship between percentage of women per major and the average salary of that major. As is evident in the graph there is a slightly negative correlation between the two showing that women tend to pick majors with average to low salaries. The median salary in the plot is represented by different size dots to help further the distinction as it decreases and the percentages of women are colored coded to also better help the viewer quickly see the trend from low to high percentage rates.
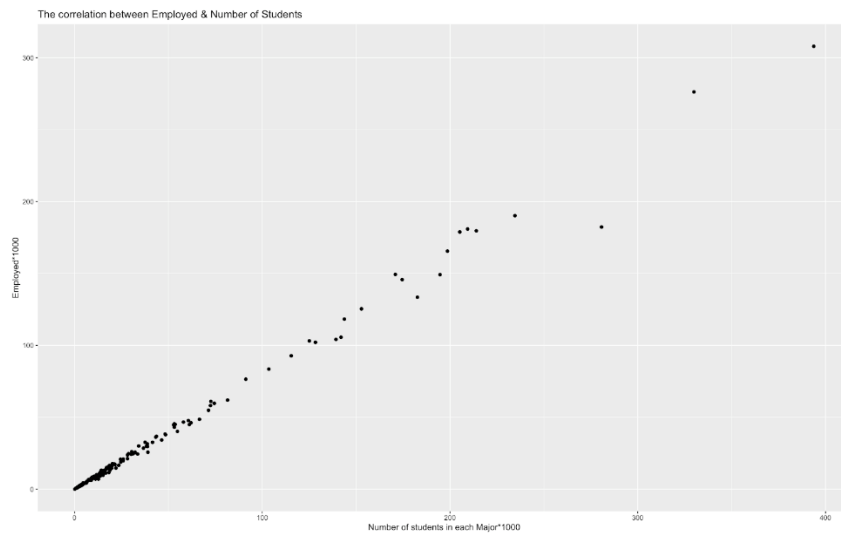

Figure 10. Correlation between employed & total

Along with salary we also wanted to see how employment related to each major. Figure 10 is a scatter plot showing the correlation between the number of graduates for each college major and the number of employed graduates. Where, the X-axis represents the number of graduates for each college major, and the Y-axis represents the number of employed graduates for each college major. As is evident in figure

10, the scatter plot shows a strong positive correlation, which leads us to believe that employment is a good indicator of why certain majors are more popular than others.

Conclusion

Rank	TOP 10 Major For Males	Salary(M)	Employed
1	GEOGRAPHY	38K	76%
2	INDUSTRIAL AND ORGANIZATIONAL PSYCHOLOGY	40K	77.70%
3	ECONOMICS	47K	74.80%
4	BIOLOGY	33.4K	64.90%
5	FOOD SCIENCE	53K	72.20%
6	PHARMACOLOGY	45K	64.90%
7	POLITICAL SCIENCE AND GOVERNMENT	38K	73.07%
8	FINANCE	47K	83.49%
9	PSYCHOLOGY	31.5K	78.20%
10	BIOMEDICAL ENGINEERING	60K	67.18%
Rank	TOP 10 Major For Females	Salary(M)	Employed
1	PSYCHOLOGY	31.5K	78.20%
2	NURSING	48K	86.40%
3	BIOLOGY	33.4K	64.90%
4	ELEMENTARY EDUCATION	32K	87.40%
5	GEOGRAPHY	38K	76%
6	MULTI-DISCIPLINARY OR GENERAL SCIENCE	35K	74.40%
7	LIBERAL ART	32K	76.80%
8	MICROBIOLOGY	38K	63.60%
9	GENERAL EDUCATION	34K	82.30%
10	PHARMACOLOGY	45K	64.90%

Figure11. The top 10 majors for Male & Female with salary and Employed

Finally, we redesigned a table containing the top 10 most popular college majors by gender and included the average salary and employment rate of each major so that it's easier to convey information in an effective and efficient manner. Because of the design limitations of the R language, we integrated the results of the previous data analysis and produced summary tables using Excel and marked the data with different colors. This table visually shows the information that readers are most concerned about, and provides detailed data support, which can help parents and students with choosing a major if they're unsure. In addition to using excel for our final design our group also used it to condense and manipulate some of the original data before uploading it to R. This allowed for a much easier time producing some of the visuals that were provided. Although we believe that the designs, we have provided are much better representations of the original there still may be some shortcomings with some of the visuals. For example, in our designs we focused on how salary and employment may influence the popularity of a major. But that doesn't take into account those who choose the major because of interest or a tuition constraint, or even those who purely pick a major because their parents urged them to do so. The information that parents and students could be paying attention to may not just be salary and employment rate, but could also be about development trends of related industries. If we had more time and access to data relevant to these possible factors, we would compare them to popularity as well. Another thing we would like to have implemented if we had more time would have been an interactive table where one could choose to look at a particular major and compare it to another using the factors shown in figure 11. This would be a nice way to compare each major in a nice concise manner instead of listing out hundreds of majors in one place.

References:

Llc, E. O. (2020, February 25). What's the Best Pre-Med Major?: Earnest. Retrieved from
https://www.earnest.com/blog/whats-the-best-major-for-medical-school/

Carmichael, S. G. (2017, September 20). Women Dominate College Majors That Lead to Lower-Paying
Work. Retrieved from https://hbr.org/2017/04/women-dominate-college-majors-that-lead-to-lower-
paying-work

CollegeFactual, T. (2017, November 22). The Most Popular Majors for Women & Men. Retrieved
from https://inside.collegefactual.com/stories/the-most-popular-majors-for-women-men

Lisa Y.Flores, Karen M.O'Brien.The career development of Mexican American adolescent women A test of
social cognitive career theory.Journal of Counseling Psychology, 2002, (1) :14.

Jamin D.Speer.The Gender Gap in College Major:Revisiting the Role of Pre-College Factors[J]. Labour
Economics, 2017, (44):69.

Most Popular College Majors for Men and Women. (2015, October 27). Retrieved from
https://www.owlguru.com/most-popular-college-majors-for-men-and-women/