

Data Challenge for Academic Innovation

Yuan Li

1. Distribution of unemployment rates among major categories

To get to know more about the distributions of unemployment rates among major categories, we can first get a summary of the distribution of unemployment rate in general. It shows that average unemployment rate of all majors is about 0.057, based on American Community Survey data (all-ages.csv) from 2010 to 2012. Maximum unemployment rate by major is about 0.15 and the minimum rate is 0.

Metrics	values
mean	0.057355
std	0.019177
min	0.000000
25%	0.046261
50%	0.054719
75%	0.069043
max	0.156147

(Table 1: Summary of statistics of Unemployment_rate¹)

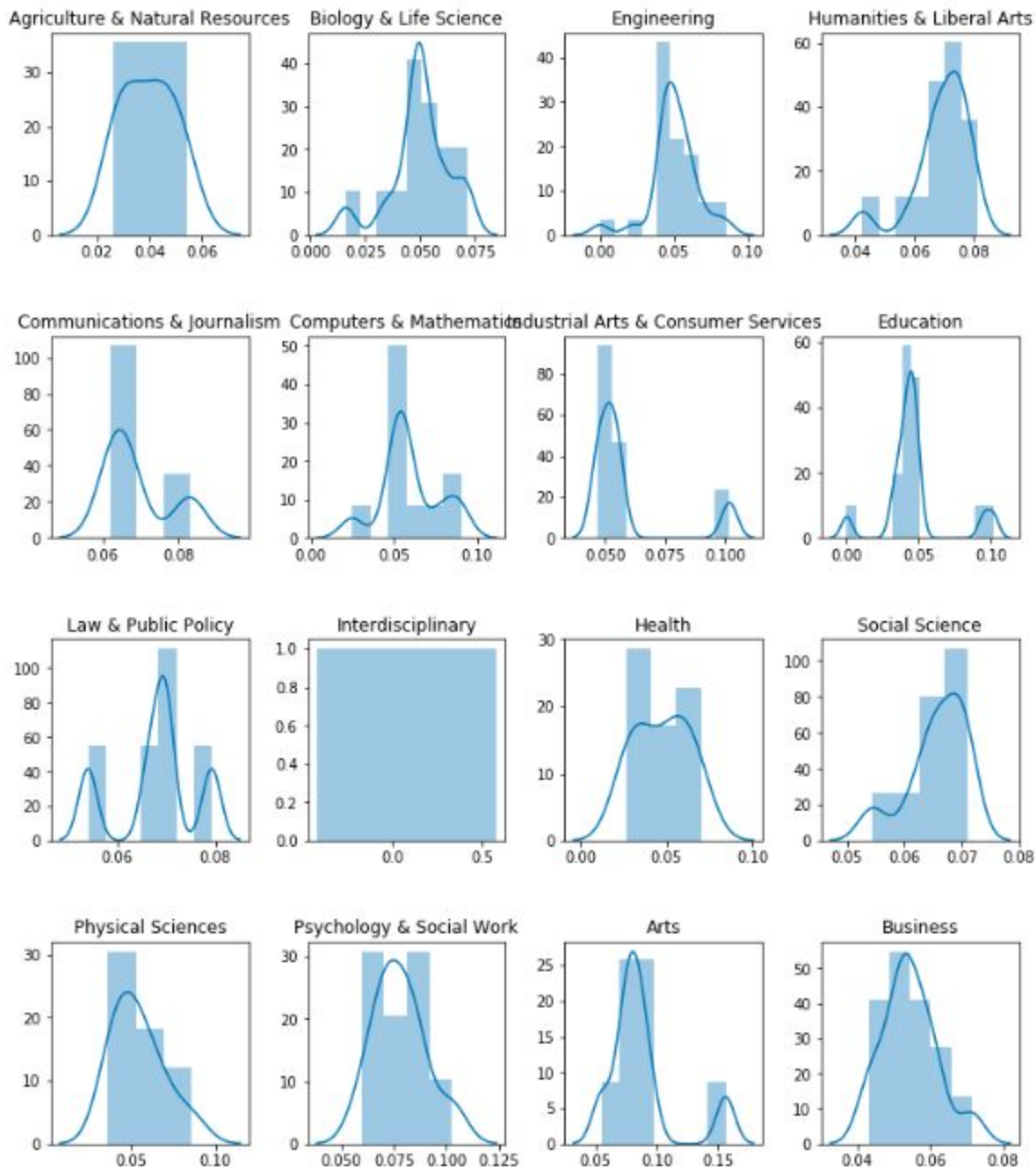
Below shows the distribution of unemployment rate of each major categories, specifically including:

Agriculture & Natural Resources, Biology & Life Science, Engineering, Humanities & Liberal Arts, Communications & Journalism, Computers & Mathematics, Industrial Arts & Consumer Services, Education, Law & Public Policy, Interdisciplinary, Health, Social Science, Physical Sciences, Psychology & Social Work, Arts and Business.

The unemployment rates of certain major categories roughly follow normal distribution, e.g., Psychology & Social Work, yet most categories show some degree of skewedness or bimodal or multimodal distribution, e.g., Law & Public Policy. Moreover, unemployment rate varies

¹ Based on all-ages.csv file

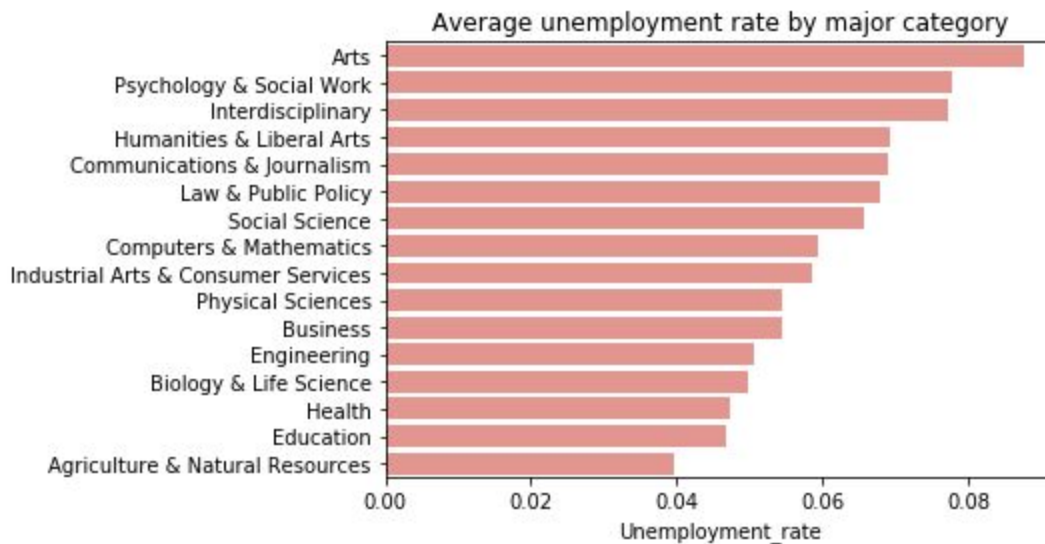
among different major categories. For example, Arts shows a relatively high unemployment rate compared to Business category.



(diagram 1: distribution of unemployment rate of different major categories)

In order to more directly compare the unemployment rate among different major categories, I came up with the bar chart below, which sorts the major categories and the unemployment rates accordingly in descending order. You can tell that categories listing on the top are those with comparatively higher unemployment rate, which include Arts, Psychology & Social Work,

Interdisciplinary, etc. In contrast, categories listed on the bottom are “promising” major categories with lower unemployment rate; to name a few, Agriculture & Natural Resources, Education, Health, etc. Interestingly, Computers & Mathematics seem not to be that “promising” as compared to my previous knowledge, which is ranked in the middle with about 6% of unemployment rate.



(diagram2: average unemployment rate of different major categories)

We can also take a more detailed look if interested in knowing more about the dissection of each major category. For example, below lists the data of each major in Agriculture & Natural Resources category, with its unemployment rate sorted in descending order. Although Agriculture & Natural Resources is relatively low in unemployment rate, its sub-majors such as Natural Resources Management (0.054), Soil Science (0.050), Food Science (0.049) are among those with slightly higher risks of being unemployed. From such data we can have a better understanding of the distribution of unemployment rates within certain major category.

unemployment rate for each major in Agriculture & Natural Resources:

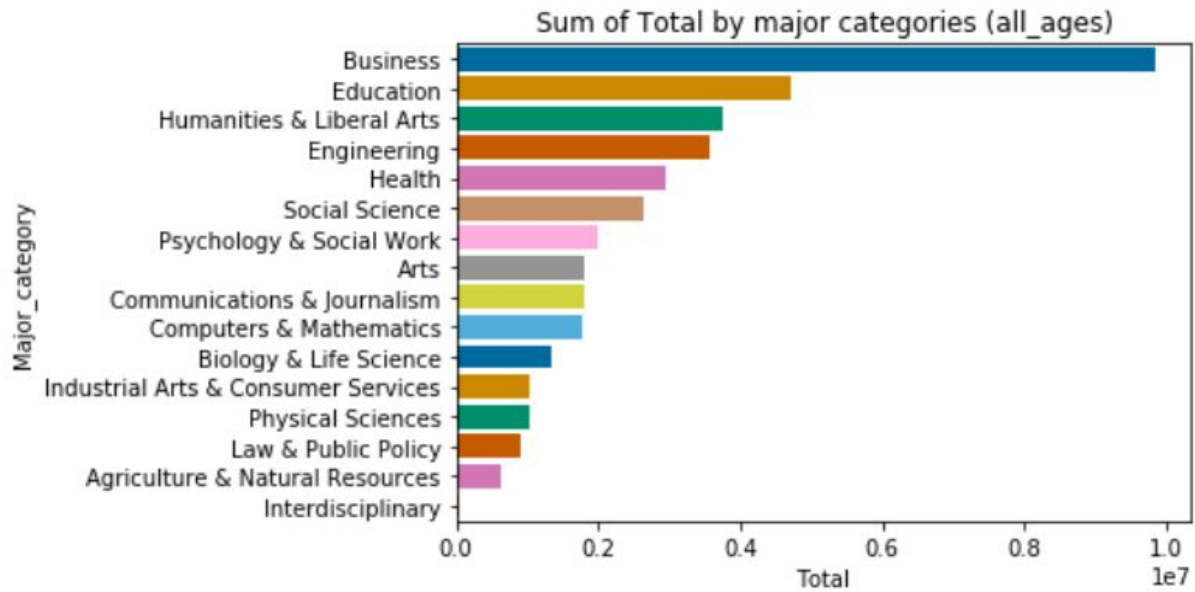
	Major_category	Major	Unemployment_rate
7	Agriculture & Natural Resources	NATURAL RESOURCES MANAGEMENT	0.054341
9	Agriculture & Natural Resources	SOIL SCIENCE	0.050867
3	Agriculture & Natural Resources	FOOD SCIENCE	0.049188
2	Agriculture & Natural Resources	ANIMAL SCIENCES	0.042679
4	Agriculture & Natural Resources	FORESTRY	0.042563
6	Agriculture & Natural Resources	MISCELLANEOUS AGRICULTURE	0.039230
8	Agriculture & Natural Resources	PLANT SCIENCE AND AGRONOMY	0.031791
0	Agriculture & Natural Resources	AGRICULTURAL ECONOMICS	0.030248
1	Agriculture & Natural Resources	AGRICULTURE PRODUCTION AND MANAGEMENT	0.028636
5	Agriculture & Natural Resources	GENERAL AGRICULTURE	0.026147

(Table 2: unemployment rate of each major in category "Agriculture & Natural Resources")

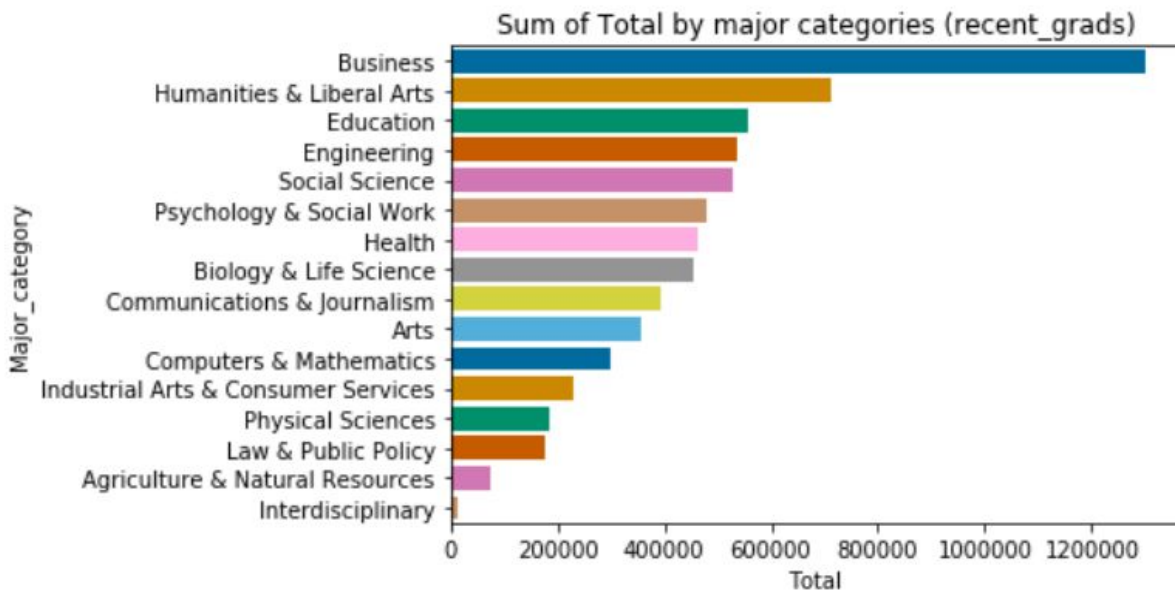
2. Explore recent trends of college majors

2.1 Comparing sum of Toal by different major categories

Visualizations below measure the sum of total number of people within certain major category. It is easy to see that Business category is the most popular within all ages, followed by Education, Humanities & Liberal Arts and Engineering. In comparison, recent grads show a similar pattern in terms of the most popular categories of majors -- the top four of which are the same as all-ages group; however, certain categories of majors, such as Social Science appear to be on an increasing trend as shown in recent grads, while other categories e.g., Health seem to be decreasing.

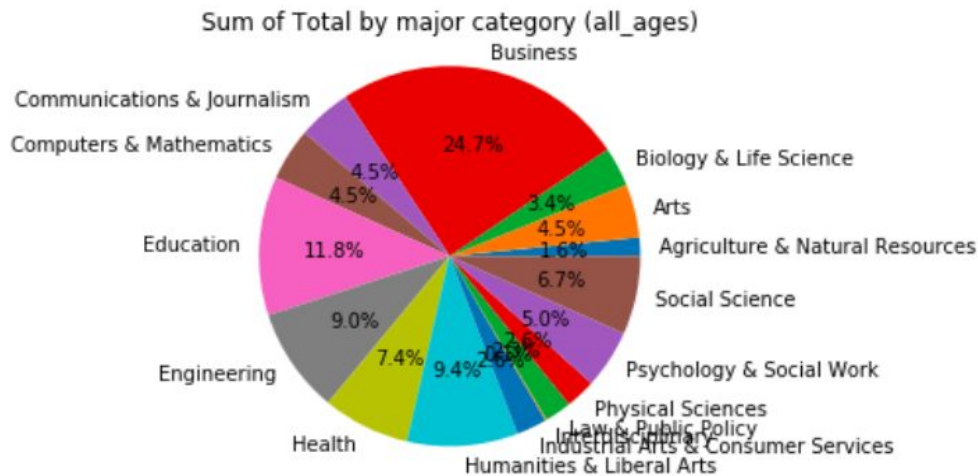


(diagram 3: sum of total num.of people by different major categories)

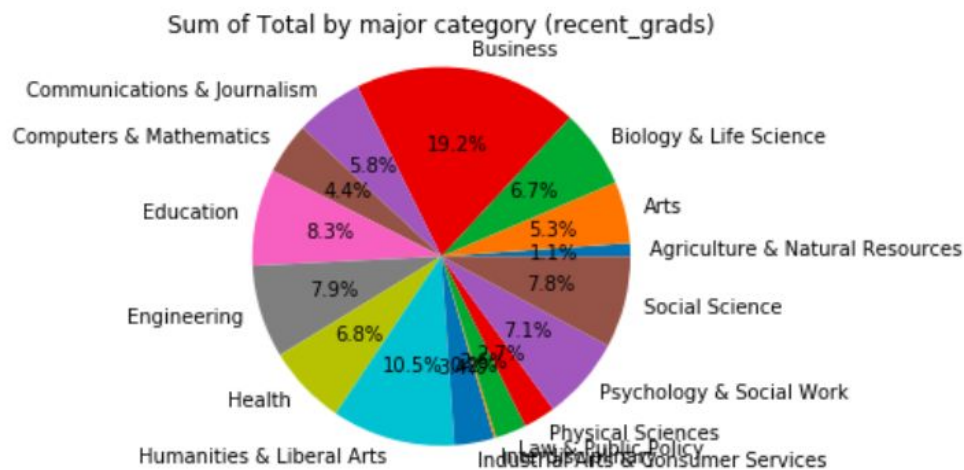


(diagram 4: sum of total num.of recent grads by different major categories)

Piecharts below show the proportion of total number of people within certain category of majors. From the charts, we can tell that the proportions of several categories of majors are shrinking, e.g., Business, Education and Engineering, while quite a few are expanding, e.g., Biology & Life Science, Social Science, Psychology & Social Work. Such changes indicate the change of preferences when people choosing majors in college.



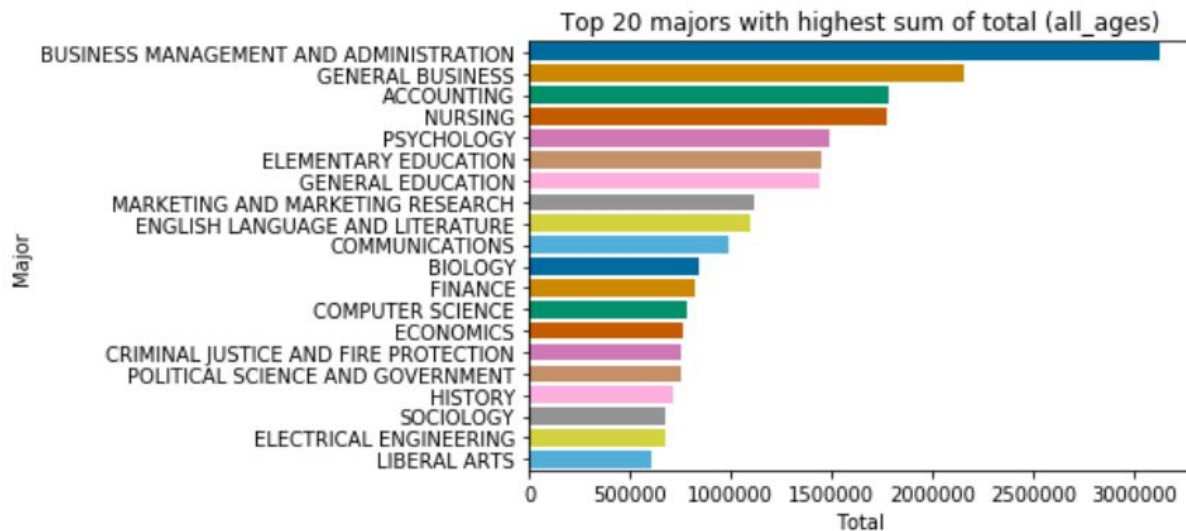
(diagram 5: proportion of total num.of people by different major categories)



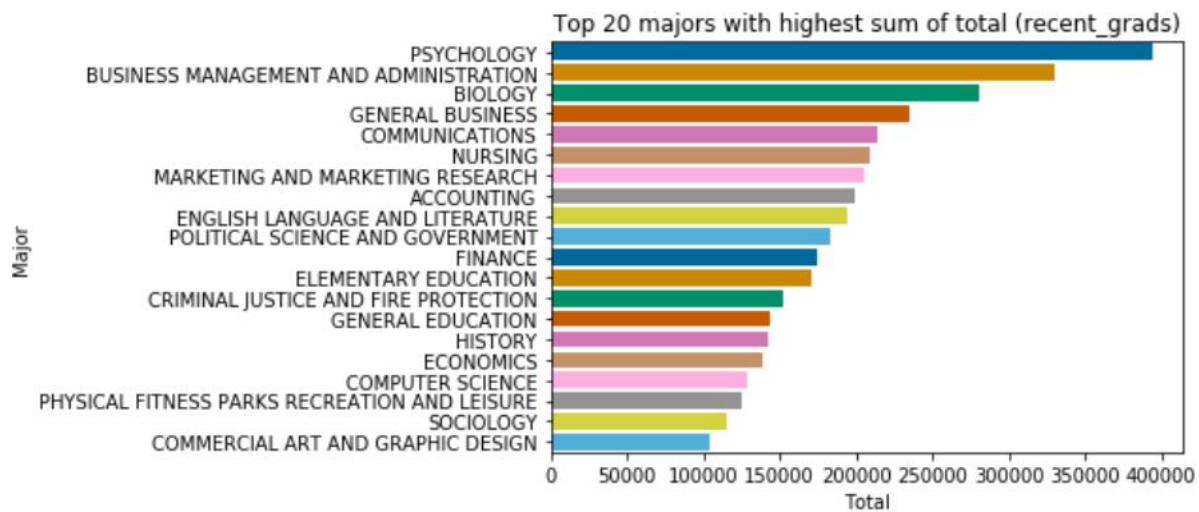
(diagram 6: proportion of total num.of recent grads by different major categories)

2.2 Comparing Top 20 most popular majors

Visualizations below list some of the most popular majors in terms of all-ages group and in recent grads. Surprisingly, Psychology has been on a booming trend of recent grads, which is listed as the most popular major. Other majors that have shown an increase trend include Communications, Biology, etc. There are also several majors that have shown an decrease in term of popularity, e.g., Accounting and Elementary Education, which may not be as popular among young graduates as among their parents' generation. Interestingly, there are also several majors that newly appear in the top ranking list, including Physical Fitness Parks Recreation and Leisure, Commercial Art and Graphic Design,etc. which may not even exist in the older days and now they are becoming more and more popular.



(diagram 7: top 20 majors with most popularity of people)



(diagram 8: top 20 majors with most popularity of recent grads)

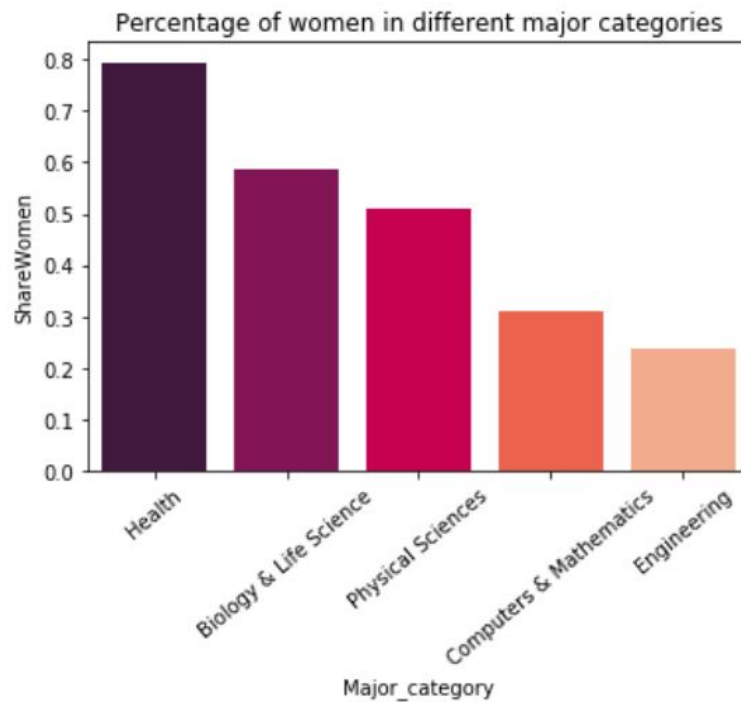
3. More exploration

3.1 Exploration on major references by genders

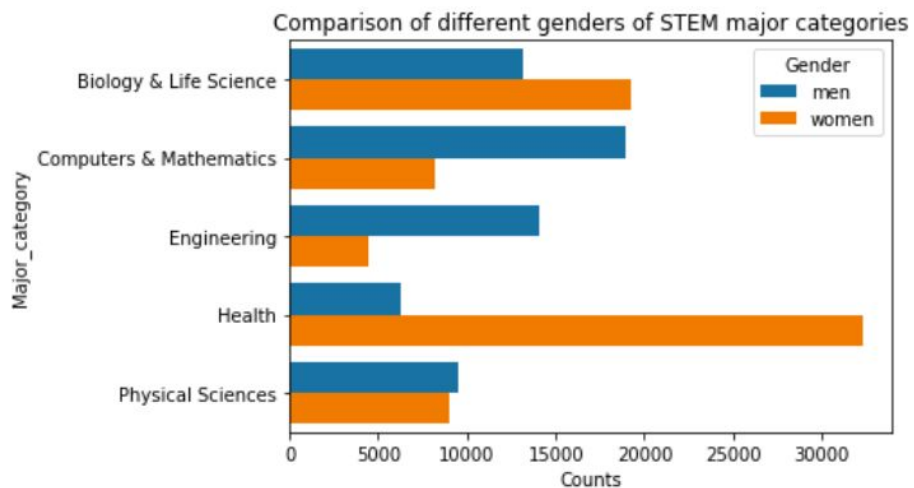
I am also interested in exploring more about women with STEM majors, specifically their preferences in choosing majors, comparison between major preferences in terms of gender, etc. So first I plot the distribution of percentage of women in different STEM domains², including Health, Biology & Life Science, Physical Sciences, Computers & Mathematics and Engineering. It seems that Health and Biology & Life Science are among the most popular category of

²Based on women-stem.csv file downloaded from the github page

women with STEM majors, and Engineering has shown a relatively low popularity among STEM-major women. This also correspond to the comparisons between male and female students majoring in STEM. Computers & Mathematics and Engineering are mostly dominated by males, while in Health and Biology & Life Science, females account for the majority.



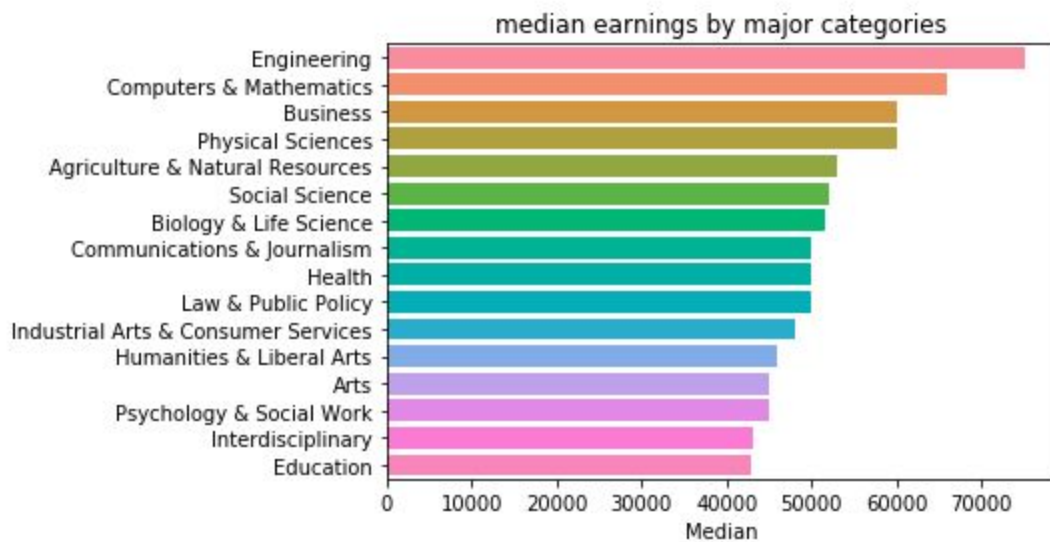
(diagram 9: percentage of women with STEM-major in different major categories)



(diagram 10: comparison of STEM-major categories by gender)

3.2 Exploration on earnings by different categories of majors

Another interesting results that I found is the median earnings of each category of majors³. Not surprisingly, people with Engineering majors seem to have the highest median earnings, followed by Computers & Mathematics, Business and Physical Sciences. Education, Interdisciplinary are among those majors with the lowest median earnings. Interestingly, although recent grads seem to have an increasing interest toward Psychology & Social Work, people with such major do not have a very high earnings in general.



(diagram 11: median earnings of different major categories)

*Additional dataset to consider

One possible dataset I could think of that is somewhat similar to this dataset, is the IPEDS survey data (downloaded link: <https://nces.ed.gov/ipeds/use-the-data>). It gives a detailed listing of each institutions and its data covering almost every aspects, including enrollment data from different years specifically. Several questions I think would be interesting to answer include: are there any preferences in terms of choosing majors in different regions or areas? What trend of major selection can be seen from a timeseries point of view? With the additional data from IPEDS survey, we are possible to answer these questions. For example, I could aggregate the data by regions or areas, and compare the enrollments of different majors across these regions, potentially we can identify the cold spot and hot spot of a specific major (or major category). From a timeseries perspective, we could aggregate data by different years, and get more insights of people's preferences of different majors as year passes by, which is somewhat similar

³Based on all-ages.csv file

to what I have done in comparing the top 20 popular majors between all-ages group and recent grads. These are the questions that I would be very interested in exploring.