Final Dashboard: College Majors Dataset

Analise Burko and Marcos delaTorre

The Dataset

Our chosen dataset is sourced from the FiveThirtyEight database and contains information relevant to college majors for recent university graduates. Taken from the American Community Survey 2010-2012 public use microdata series and contains information on basic earnings and labor force information. Most features are numerical values of employment status by major, additionally containing features such as sex and type of job. All of the original variables are listed in the following table. Identifying mcolumns are 'Major' which indicate the name of the specific major record, and 'Major_category' which groups those majors into higher level categories which often represent departments in a university setting. 'Median_income' is numerical and indicates the median income of those employed from that major-record. All remaining variables are numerical counts indicating different demographics within each major, for instance the total number of people broken down into the number of men and women, unemployed graduates versus employed graduates broken down into full-time employment, graduates employed within/outside their field, and graduates employed in jobs that receive a low-wage. From these original columns we derive percentages that can be seen in the final dashboard.

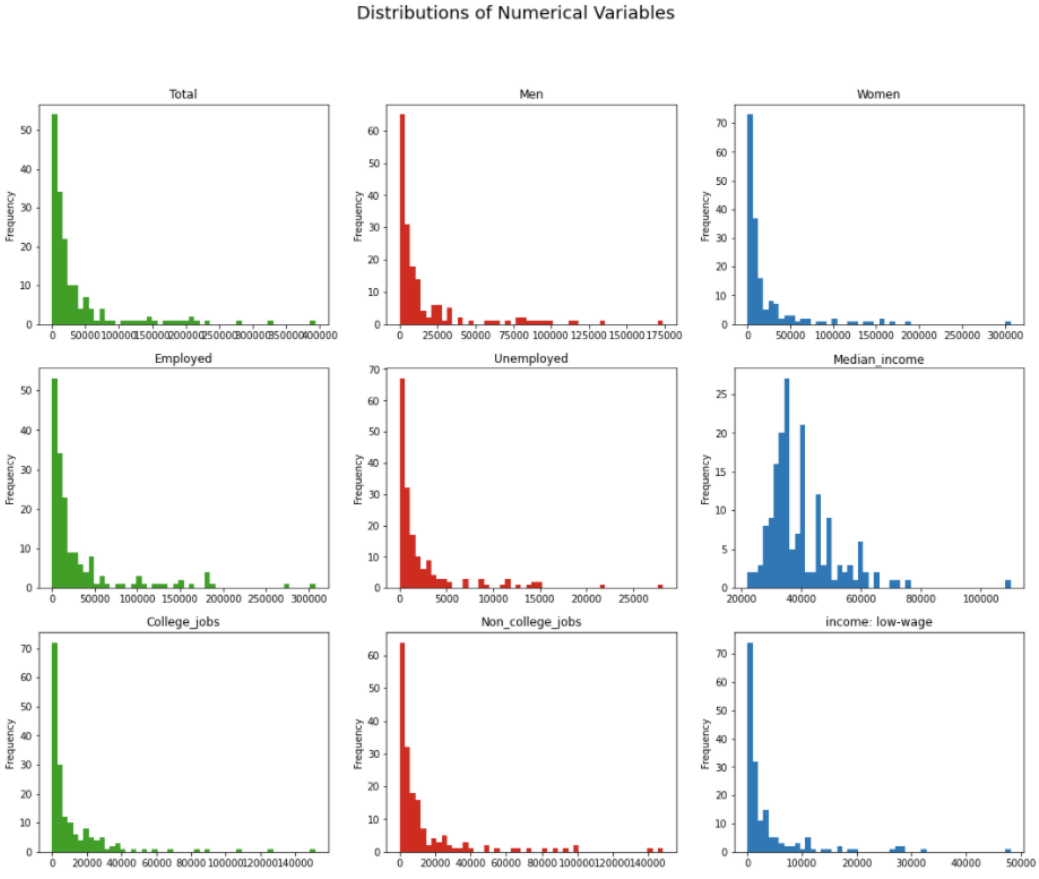
Variable Name	Description	Data type
Major_code	Major code, FO1DP in ACS PUMS	Identifier of the row
Major	Major description	Descriptive
Major_category	Category of major from Carnevale et al	Categorical
Total	Total number of people with major	Numerical (count)
Men	Number of men in major	Numerical (count)
Women	Number of women in major	Numerical (count)
Employed	Number employed (ESR == 1 or 2)	Numerical (count)
Unemployed	Number unemployed (ESR == 3)	Numerical (count)
Full_time_year_round	Employed at least 50 weeks (WKW == 1) and at least 35 hours (WKHP >= 35)	Numerical (count)
Median_income	Median earnings of full-time, year-round workers	Numerical (real)
College_jobs	Number with job requiring a college degree	Numerical (count)
Non_college_jobs	Number with job not requiring a college degree	Numerical (count)
Low_wage_jobs	Number in low-wage service jobs	Numerical (count)

The table below describes key statistical features for each relevant column, including a minimum and maximum. All columns aside from 'Median' represent individual persons and therefore the precision should remain as whole integers. Further, the dataset is taken from census surveys from 2010-2012. It is reasonable to assume a short lifespan considering the constant flux of the labor market, and thereby should ideally be updated each year to maintain accuracy.

	Total	Men	Women	Employed	Full_time_year_round	Unemployed	Median_income	College_jobs	Non_college_jobs	Low_wage_jobs
count	172.000000	172.000000	172.000000	172.00000	172.000000	172.000000	172.000000	172.000000	172.000000	172.000000
mean	39370.081395	16723.406977	22646.674419	31355.80814	19798.843023	2428.412791	40076.744186	12387.401163	13354.325581	3878.633721
std	63483.491009	28122.433474	41057.330740	50777.42865	33229.227514	4121.730452	11461.388773	21344.967522	23841.326605	6960.467621
min	124.000000	119.000000	0.000000	0.00000	111.000000	0.000000	22000.000000	0.000000	0.000000	0.000000
25%	4549.750000	2177.500000	1778.250000	3734.75000	2474.750000	299.500000	33000.000000	1744.750000	1594.000000	336.750000
50%	15104.000000	5434.000000	8386.500000	12031.50000	7436.500000	905.000000	36000.000000	4467.500000	4603.500000	1238.500000
75%	38909.750000	14631.000000	22553.750000	31701.25000	17674.750000	2397.000000	45000.000000	14595.750000	11791.750000	3496.000000
max	393735.000000	173809.000000	307087.000000	307933.00000	199897.000000	28169.000000	110000.000000	151643.000000	148395.000000	48207.000000

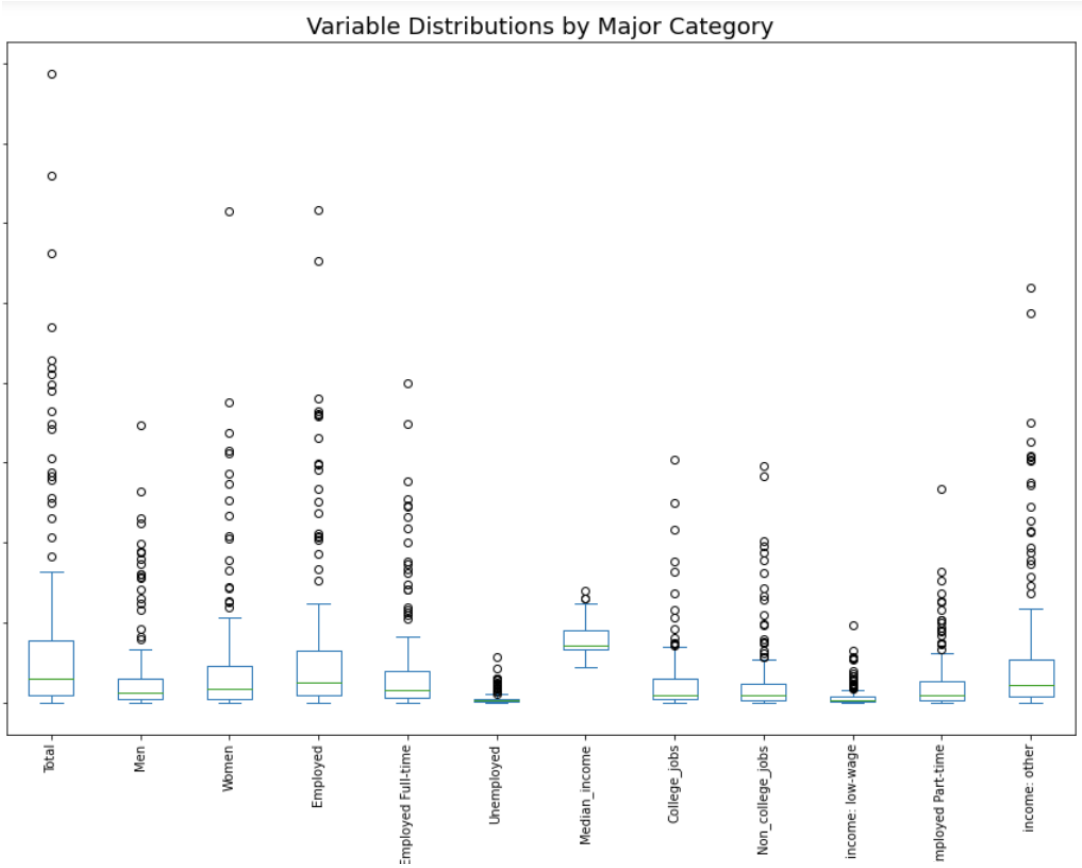
Distribution

The first figure shows representation of major categories, where areas such as engineering and education have much higher occurrence than communications and interdisciplinary majors shown left. The mean Median income for major categories is shown on the right. The figure 'Distributions of Numerical Variables' plots histograms for each variables, revealing left-skews for all variables aside from 'Median' which takes on a more normal distribution.

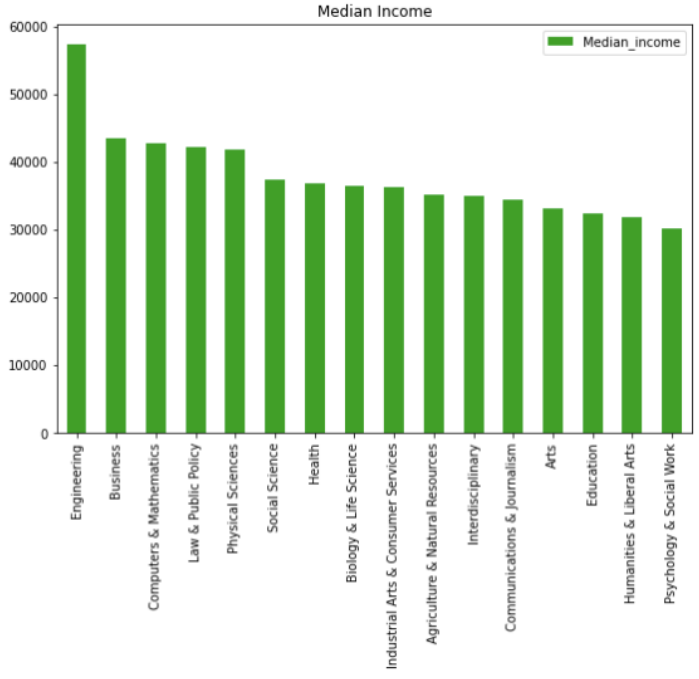


Outliers & Groupings

The prominent grouping of this dataset is by 'Major_category' to compare records of the same fields internally and against other groups. As previously discussed, demographic groupings are Men/Women/Total, Employed/Unemployed/Full_time_year_round/ Low_wage_jobs, and College_jobs/Non_college_jobs. Looking at the box plots of numerical variables by Major category leads to some initial intuition of outliers. Variables 'Total', 'Women', 'Unemployed', and 'Non_college_jobs' have obvious outliers.



The key variable of interest is Median_income, which can be seen averaged and sorted by major category at right. The category with the highest global average in Engineering, which may be considered an outlier. The lowest global average is ‘Psychology & Social Work’, though overall the trend between categories is fairly stable and we do not witness any major-jumps after the top-earning category.



Audience

The audience of these charts are not expected to be experts in the usage of charts. They will access the charts from a PC or a smartphone. They will consult the visualization occasionally. The main message of these charts is that the kind of job that graduates may obtain depends strongly on their major. An additional message is that there may be differences between male and female students in their choice of majors. Our dashboard will target students who are contemplating which major to pursue in their own academic journeys. By painting a full-picture of all the majors available to them in terms of university demographics and also long term impacts, students will be empowered in their decision-making processes.

General goal: to convey major categories and individual majors in terms of demographics, popularity, employability (represented by type, and also full versus part-time employment), and future economic success effectively. The dashboard will concisely and effectively communicate said information to provide insights to drive decisions by our audience.

We have identified the following typical kinds of users personas::

> **Persona 1: Prospective student**
> **Goal**: Help them choose a major, by understanding what impact of their choice may have in their future job opportunities and earnings.
>
> **Persona 2: Faculty administration**
> **Goal**: Understand which majors are more demanded by students and by the job market, and take it into account for planning the offering of places and hiring of new professors.
>
> **Persona 3: Government official**
> **Goal:** Investigate possible actions to improve social mobility, by helping children of poor families to obtain a degree that may lead to well-paid jobs. Investigate majors in which the rate of female students is very low, to possibly promote actions against negative stereotypes. Investigate which majors typically result in jobs requiring a college degree or not.

Based on the goals indicated previously, a number of questions that the charts should help to answer are:

- How do median earnings of recent graduates vary depending on their kind of major?
- Which are the majors with the highest earnings?

- Which are the majors with the most graduate students?
- Is there a difference for chosen majors between male and female students?
- Which majors/major categories are more likely to result in a job requiring a college degree?

Scenario

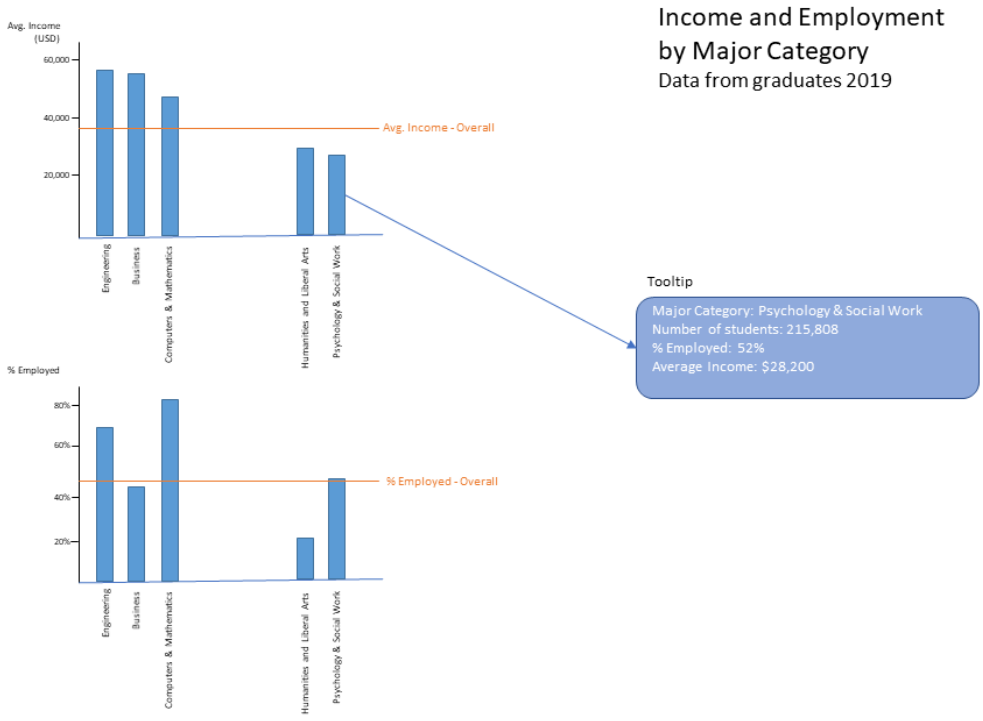
Maria has just finished high school and is wondering what degree she should study. In principle she is interested in Mathematics and also in Psychology. Maria hears about the dashboard that compares job opportunities for graduates of different majors, and she decides to consult it.

In the top-level dashboard she sees that the area of Mathematics offers much better chances to get a job than Psychology, as well as considerably higher salaries. Within Mathematics, she checks the drill-down by specific major and realises that the major of Applied Mathematics, which she is attracted to, offers good prospects to obtain qualified jobs (that require a degree) as well as an outstanding expected income. This helps her decide to major in this subject.

Napkin Designs

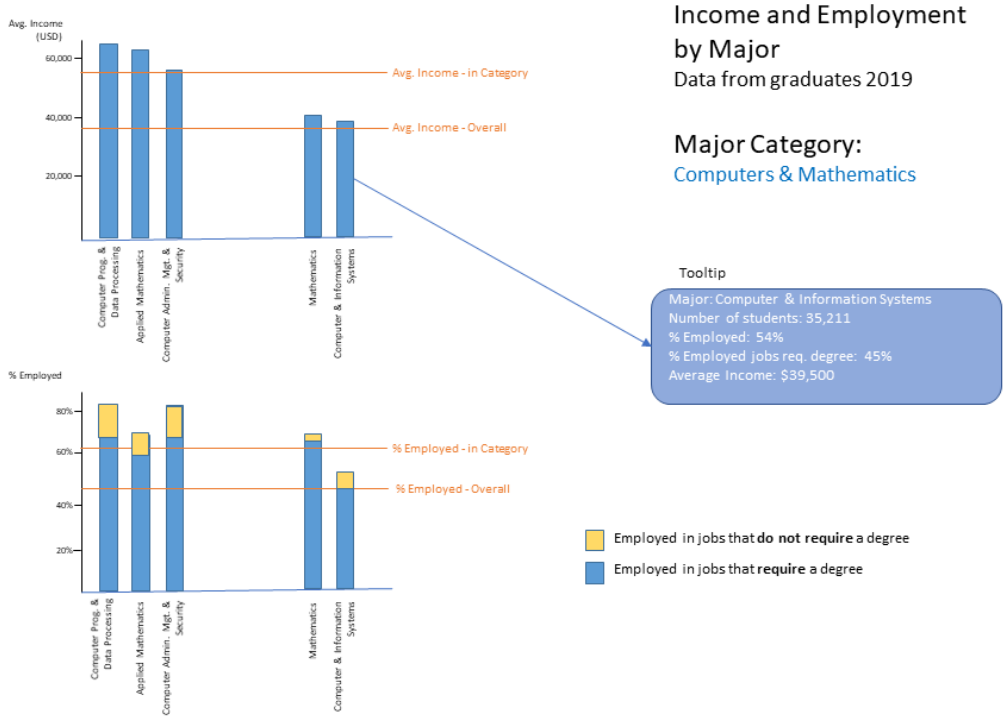
<u>Level 1</u>

Level 1 shows data by major category. For each category, the average income and the % of employed students are displayed in two parallel bar charts. A tooltip is available to view the exact data.



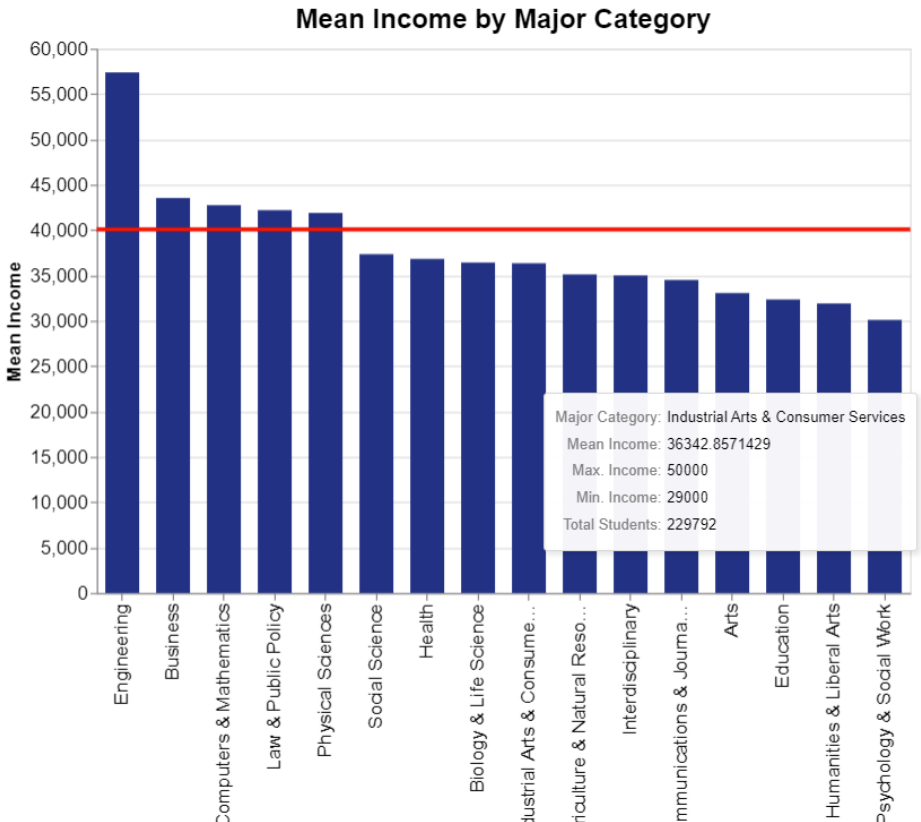
Level 2

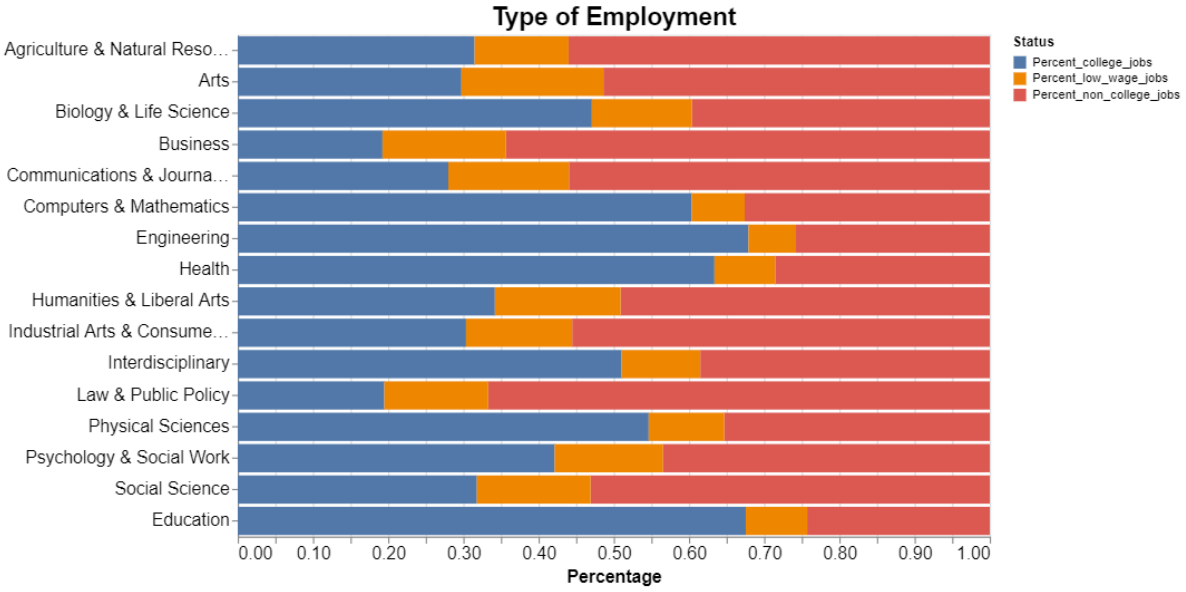
By clicking on the bar of a major category, we drill down into level 2. Level 2 shows data by major filtered by the selected major category. For each major, the average income and the % of employed students are displayed in two parallel bar charts. The % of employment is divided between jobs that **require a degree** and those that don't. We can see as context the average values for the category and overall. A tooltip is available to view the exact data.



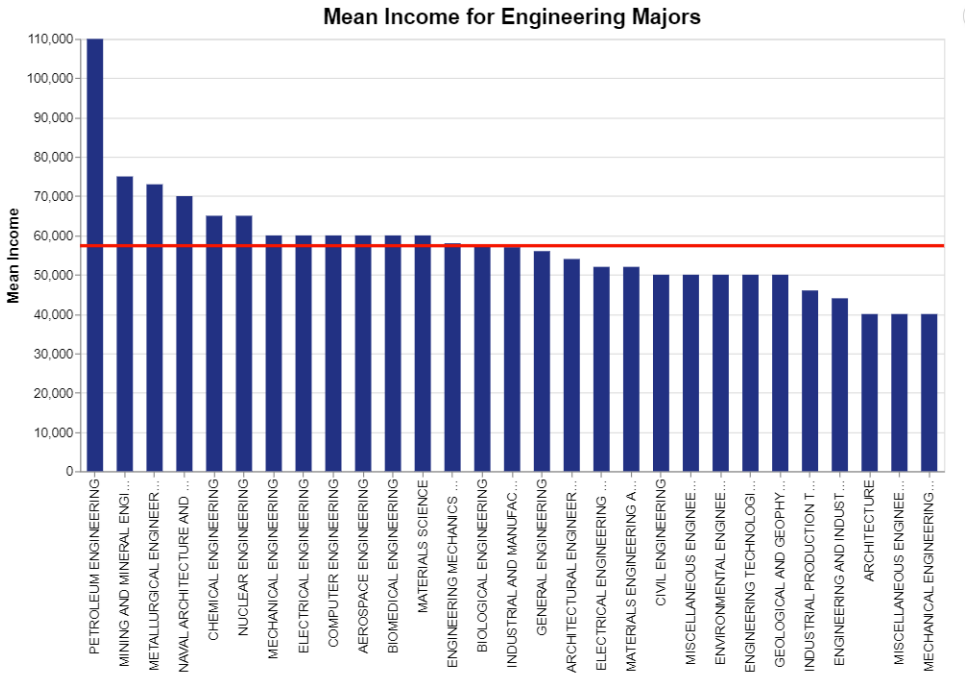
Proof of Concept

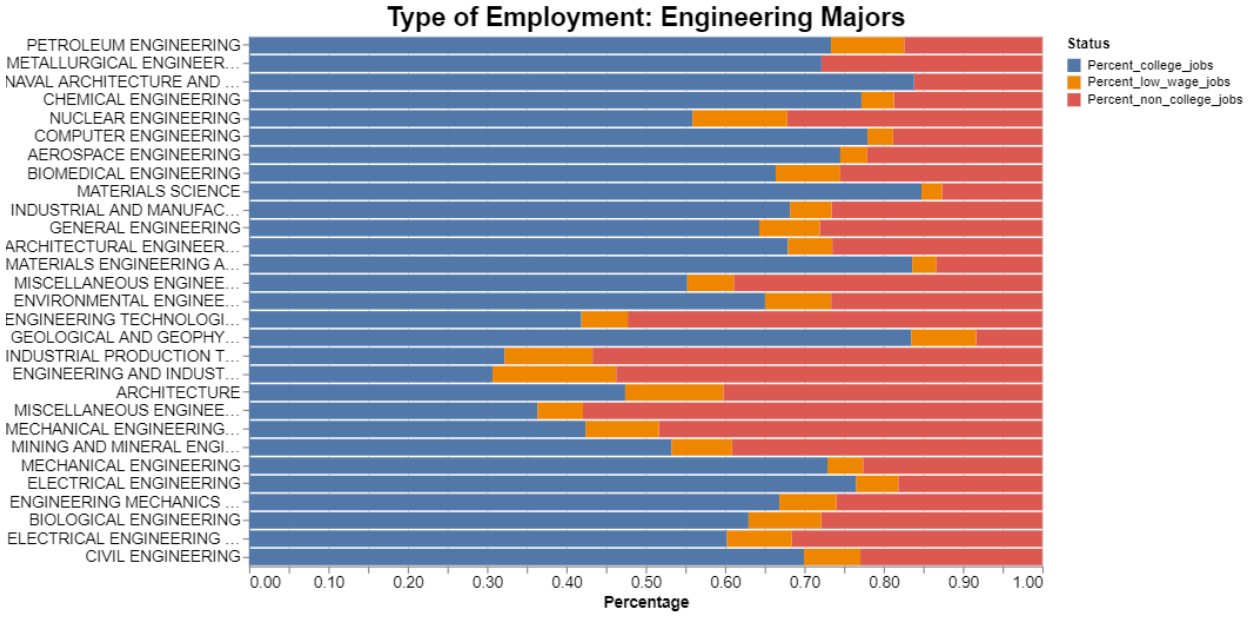
Level 1




Type of Employment
Agriculture & Natural Reso...
Arts
Biology & Life Science
Business
Communications & Journa...
Computers & Mathematics
Engineering
Health
Humanities & Liberal Arts
Industrial Arts & Consume...
Interdisciplinary
Law & Public Policy
Physical Sciences
Psychology & Social Work
Social Science
Education
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
Percentage
Status
Percent_college_jobs
Percent_low_wage_jobs
Percent_non_college_jobs

Level 2


Mean Income for Engineering Majors
Mean Income
110,000
100,000
90,000
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0
PETROLEUM ENGINEERING
MINING AND MINERAL ENGI...
METALLURGICAL ENGINEER...
NAVAL ARCHITECTURE AND ...
CHEMICAL ENGINEERING
NUCLEAR ENGINEERING
MECHANICAL ENGINEERING
ELECTRICAL ENGINEERING
COMPUTER ENGINEERING
AEROSPACE ENGINEERING
BIOMEDICAL ENGINEERING
MATERIALS SCIENCE
ENGINEERING MECHANICS ...
BIOLOGICAL ENGINEERING
INDUSTRIAL AND MANUFAC...
GENERAL ENGINEERING
ARCHITECTURAL ENGINEER...
ELECTRICAL ENGINEERING ...
MATERIALS ENGINEERING A...
CIVIL ENGINEERING
MISCELLANEOUS ENGINEE...
ENVIRONMENTAL ENGINEE...
ENGINEERING TECHNOLOGI...
GEOLOGICAL AND GEOPHY...
INDUSTRIAL PRODUCTION T...
ENGINEERING AND INDUST...
ARCHITECTURE
MISCELLANEOUS ENGINEE...
MECHANICAL ENGINEERING...


Type of Employment: Engineering Majors
PETROLEUM ENGINEERING
METALLURGICAL ENGINEER...
NAVAL ARCHITECTURE AND ...
CHEMICAL ENGINEERING
NUCLEAR ENGINEERING
COMPUTER ENGINEERING
AEROSPACE ENGINEERING
BIOMEDICAL ENGINEERING
MATERIALS SCIENCE
INDUSTRIAL AND MANUFAC...
GENERAL ENGINEERING
ARCHITECTURAL ENGINEER...
MATERIALS ENGINEERING A...
MISCELLANEOUS ENGINEE...
ENVIRONMENTAL ENGINEE...
ENGINEERING TECHNOLOGI...
GEOLOGICAL AND GEOPHY...
INDUSTRIAL PRODUCTION T...
ENGINEERING AND INDUST...
ARCHITECTURE
MISCELLANEOUS ENGINEE...
MECHANICAL ENGINEERING...
MINING AND MINERAL ENGI...
MECHANICAL ENGINEERING
ELECTRICAL ENGINEERING
ENGINEERING MECHANICS ...
BIOLOGICAL ENGINEERING
ELECTRICAL ENGINEERING ...
CIVIL ENGINEERING
0.00 0.10 0.20 0.30 0.40 0.50 0.60 0.70 0.80 0.90 1.00
Percentage
Status
Percent_college_jobs
Percent_low_wage_jobs
Percent_non_college_jobs

Considering Perception

Income: the most important variable is the **income**. This is displayed with vertical bars using bar length.

Advantages:

- Type of chart is well known and understood
- Easy to perceive: **length** is a powerful preattentive property
- Easy to compare between bars (majors and major categories)

Employment/Employment college jobs: other variables to be shown and compared are "% Employment" and "% Employment college jobs". For these, stacked bars have been selected, with each bar adding up to 100%.

Advantages:

- Type of chart is well known and understood
- Easy to perceive: **length** is a powerful preattentive property. Besides, the different sections of the bars are differentiated with **color**, which is also preattentive. The colors (blue, orange, red) are from Altair's default palette and are accessible for color-blind people.
- Easy to compare between bars (majors and major categories)

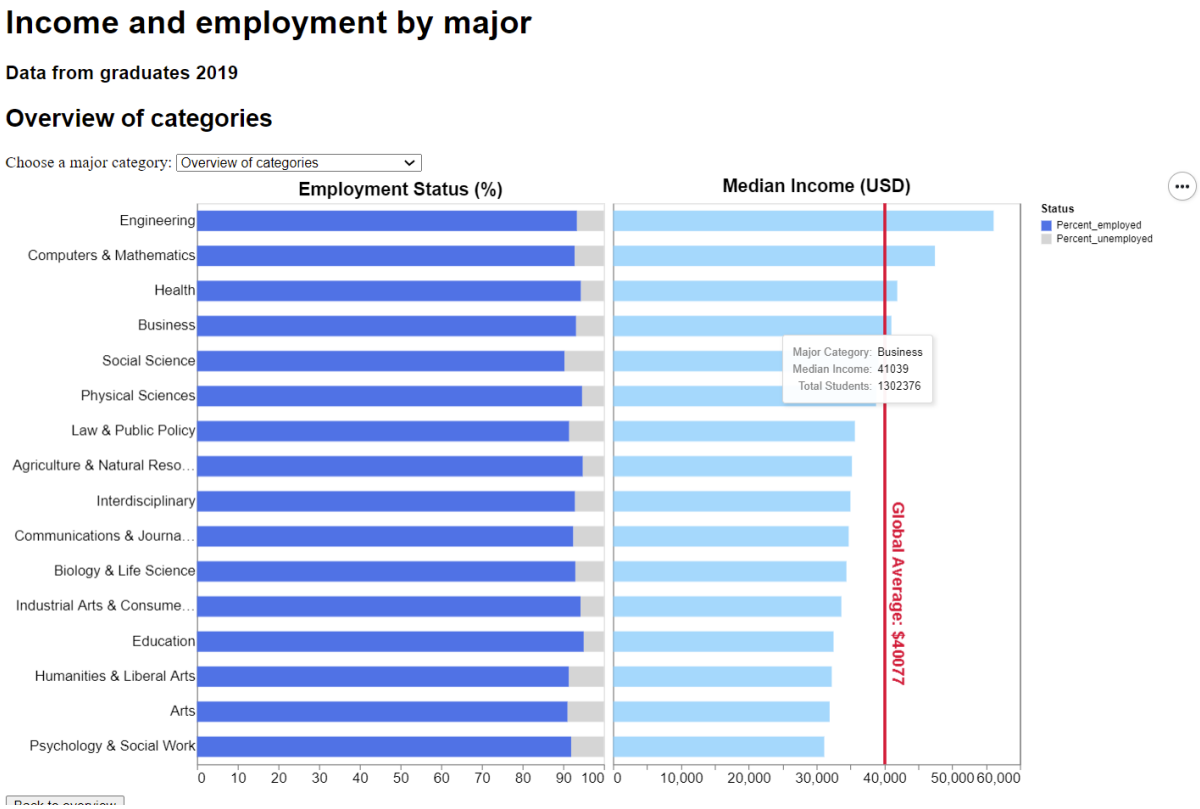
Potential disadvantages:

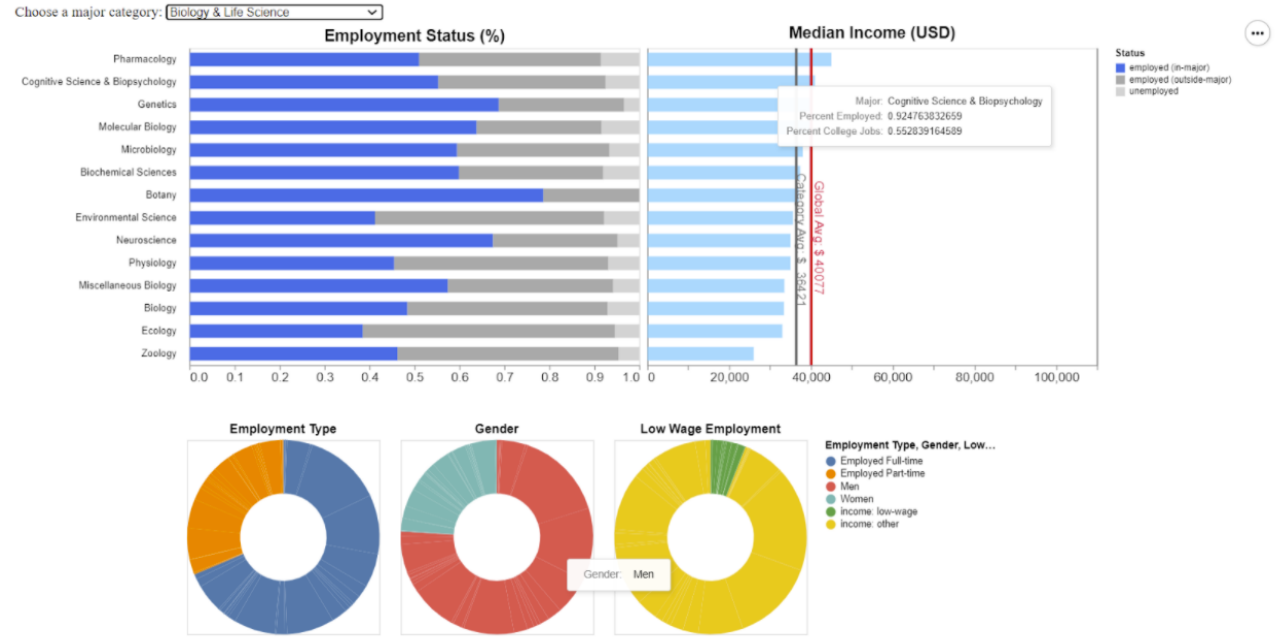
- The bar sections that do not start at the bottom cannot be compared visually. This means that the percent of unemployment and percent of non-college jobs are not easy to compare.
- This is however considered a minor issue, as the users are mostly interested in comparing Percent of employment and Percent of jobs requiring college, and these are visually comparable. In addition, the tooltip shows the numeric values if needed.

Context: Global and category average

In order to provide context, the global average and the average for the category are shown as horizontal lines. This relies on the preattentive property of position

Overview of Dashboard





Dashboard Description

As shown on the previous page, the finalized dashboard has two levels that are similarly structured. The top level provides an overview of all the data, and thereby represents all majors and major categories in the dataset to provide a broad impression of the different major categories that can then drive the audience's decision when seeking out more information on the second-level of the dashboard. Visualizations consider user perception in all components to enhance the viewing experience and transmit the message of the dashboard more effectively. Overall we aimed to display as much useful information as possible in a way that is concise and easy-to-understand.

The main visualizations of both dashboard levels are two horizontal bar charts, the left a stacked bar chart that displays a breakdown for employment status, and the right displays median income. These charts are intentionally oriented and aligned so that the user can read each major and consolidate redundant information that existed in previous iterations, mainly being the y-axis labels. The charts were also refined to be easier to interpret by our audience. We added white-space between the bars and removed the grid-lines in the background to be able to better distinguish between each segment. Additionally the label names were shortened and resized to make them more legible, allowing for the full names to fit within the designated space.

The color scheme was also altered from the default colors to make the entire dashboard more cohesive and interpretable. The bar charts are displayed in blue and grey tones to reflect their connectivity, allowing the viewer to read the label then attributes from left-to-right in a continuous fashion. For the stacked bar chart, percent employed and percent employed (in-major) are the segments of importance for our viewers, with the other segments regarding unemployment and employment (outside-major) offering supplementary information to the graphic, this distribution of importance is emphasized with the bright-blue bringing attention to the employment segments to focus on, and the grey-tones understating the less-important information. On the level 2 dashboards, the donut charts at the bottom are colored in contrast to the main visualization to create separation between the two sections. Each of the sections in each donut chart is colored to be on the opposite side of the color-wheel as the other, creating contrast and emphasizing the duality in each visualization.

For the overall layout of the level 3 dashboard, donut charts are smaller and at the bottom to act as supporting information if the user is interested in gaining further insights after understanding the broader and more critical information that is consumed first. This is a reasonable assumption as it is shown in western nations the user will perceive the visualizations from left to right and top to bottom, in the same fashion that text is read.

Supporting attributes such as the average lines and tooltips offer the user extra assistance in interpreting the graph while minimizing the amount of clutter throughout the dashboard. The global income average line stands out in red, creating a universal reference point and connection between both dashboards and all categories. Having a universal reference to compare the category average line against (on Level 2) gives the user better ability to compare different categories against each-other without needing to look at all of the available graphs at once. The tooltips reinforce the information provided on the axes while offering a small-amount of additional information, this style of reinforcement increases user confidence in his/her own interpretation of the data presented.