Final Report

Bear Paths

Daphne Jong, Devin Huang, Lily Lin

i247 INFOVIZ SPRING 2018

Table of Contents

Project Goals

Choosing a major at UC Berkeley can be difficult and a typical question is: *where do students end up after graduating from their respective majors?* Currently, the UC Berkeley Career center offers employment statistics and other relevant data for undergraduates to help them make better decisions. However, the current information is buried in PDF files that are inefficient to use.

Our primary goal will be to make the data in the PDF more visually accessible to provide a better and more insightful experience for students when comparing data between majors as well as demonstrate that there are many different pathways available to students. The goal is to answer questions like: *What graduate programs do people in my major go into? What types of industries do they end up in?*

With our data visualization tool, we ultimately hope to help undergraduates at UC Berkeley see what paths are available to different college majors, assisting them in making a decision as to what major is right for them and demystifying the idea that there is only one set path in life.

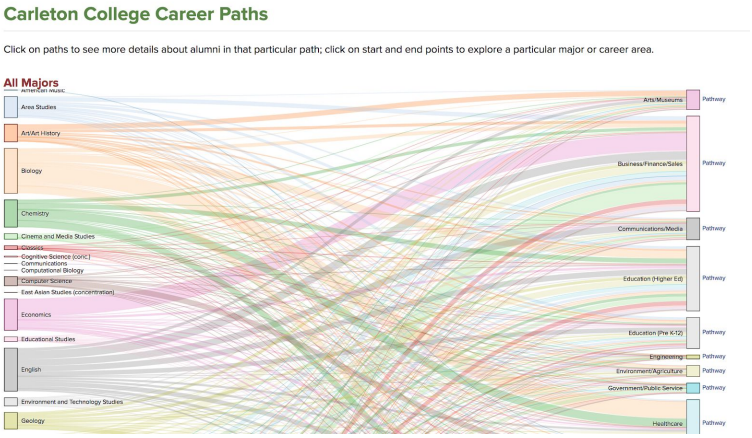
Related Work

Life After College

One interesting work that explores life for recent graduates is an interactive report by Prudential. The report uses graphics effectively to show salary, location, trends for life of graduates after college. While this report is beautiful and informative of trends of the post-college life, it is not particularly useful for choosing a major or career path.

Carleton College Career Paths

Carleton College provides a sankey visualization of different majors and which career sector they end up with, which is a good way to show an overview of the links and the different pathways that are available. We decided to implement a similar visualization but as a Chord Diagram where each major could show links to different sectors.



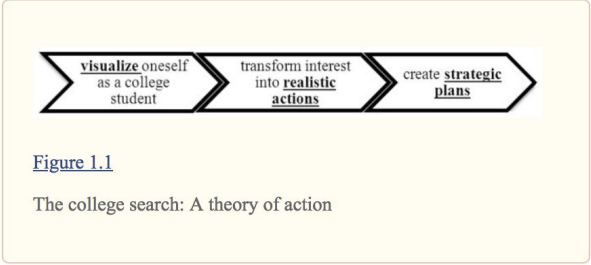
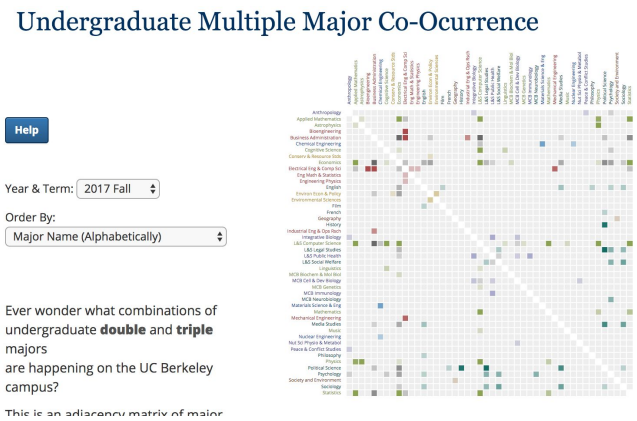


Figure 1.1

The college search: A theory of action

Pathways to college and STEM careers

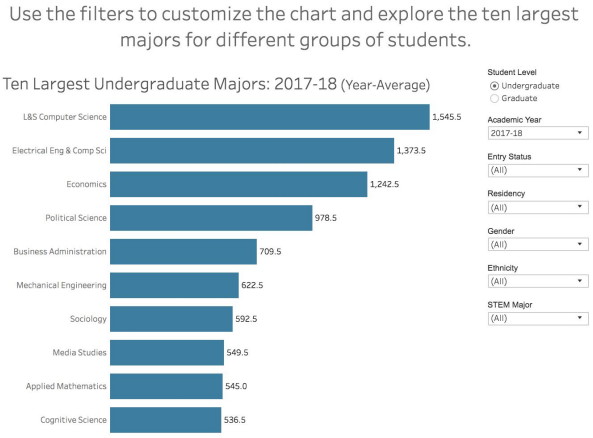
Though the research paper was primarily on how to enhance the high school student pathway to college and STEM careers, it provided grounding and validation that visualization of outcomes plays an instrumental role in what students end up achieving: especially with regards to what they do not already see.



Our Berkeley, Multiple Majors

This innovative visualization uses a adjacency matrix to show major co-occurrence at UC Berkeley. The use of sorting on the same matrix to show trends was the inspiration for the heatmap sorting in our navigation.

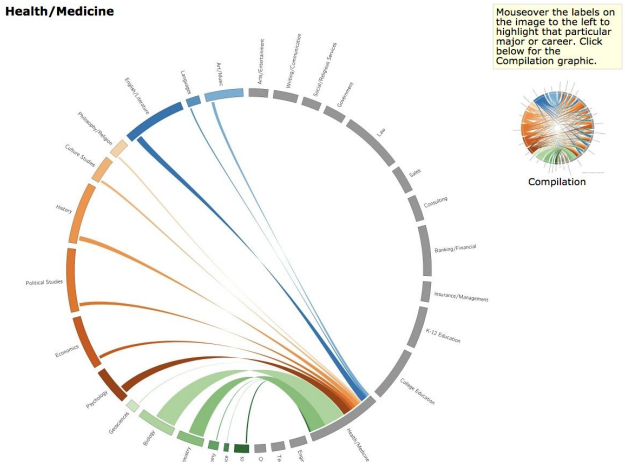
Our Berkeley, Student Headcount by Major and Demographics

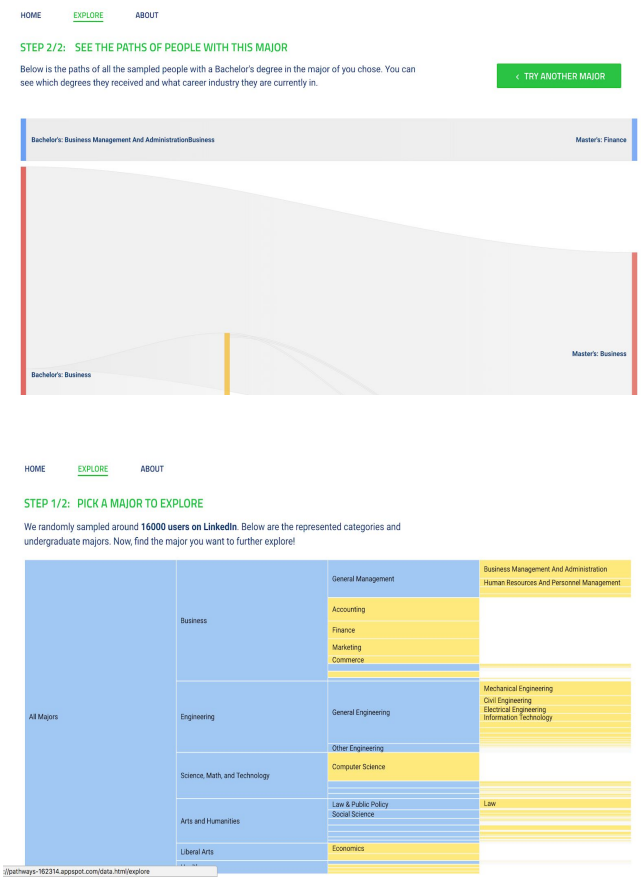


Another project hosted by UC Berkeley's OPA provided us with a visual of one of our early concepts: using bar charts to show facts or trends with regards to the largest majors. We eventually decided against this visualization because 1. it had already been done and 2. having seen an example in action, we felt that while it is effective to show trends of populous majors, it did not serve our needs or goals with regards to helping students explore different potential pathways and majors.

Visualizing the Liberal Arts

An interactive chord diagram connects 15 groups of majors to career groupings. This gives a good overview through the visual interaction's use of highlighting and auto-updating of auxiliary textual information. We considered groupings for the UC Berkeley data but decided that a modified Chord where the groupings are not in a circle would be easier for the eyes to follow and make the trajectories seem more like pathways.





CMU Pathways

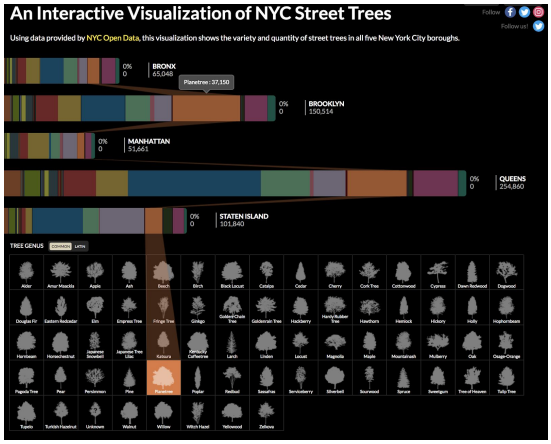
One of the data pipeline projects at CMU used a sankey to visualize where people of each major go but only for whether the students end up doing Masters or PhD degrees. This might not be as informative for students at UC Berkeley as many want to know what jobs or area of industry people end up in as well. In addition, the use of a tree map to let students explore majors makes it hard to navigate as the shape of each sector is sized according to the number of people in the major. We decided to improve on this by interviewing undergraduates about what metrics are most important to them so we can find a way to let students navigate to majors in a more informative way such as using sorting and filtering.

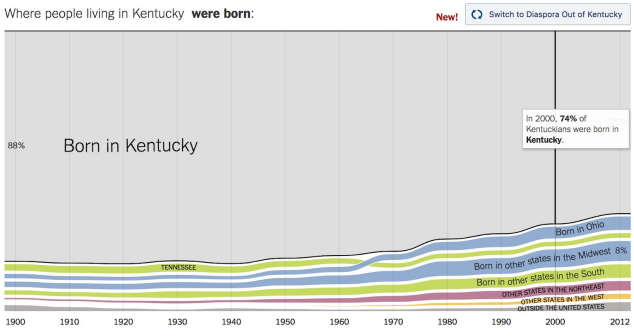
Dungeons and Developers

This website allows users to choose skills in a skill tree that unlock more advanced skills later on. As you level, it displays the relevant job title as an character. We built on top of this idea of skills being more important in a skill-based economy and tried to show that skills gained from majors would lead to different jobs within our narrative.

NYC Street Trees

This interactive visualization showcased a stacked bar chart for different locations. When you hover over a certain type, it highlights that specific genus across all locations. This was interesting for us because it showed a certain type of something across different locations. It could have very easily been applied to certain job titles across different states.





Where We Came From

This visualization by the NY Times was interesting because it showed movement across years. The thickness of the bands also changes to indicate changes in volume. We considered using something similar to show the amount of people within a certain profession over time.

Visualization Description

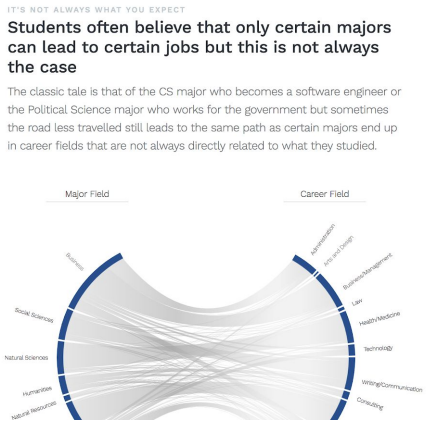
CONTEXTUAL COMPONENT

Homepage

We start with a homepage that introduces the audience to the problem that we are trying to address; that UC Berkeley undergraduates currently find themselves asking themselves what types of careers are available to them after majoring in a certain field.

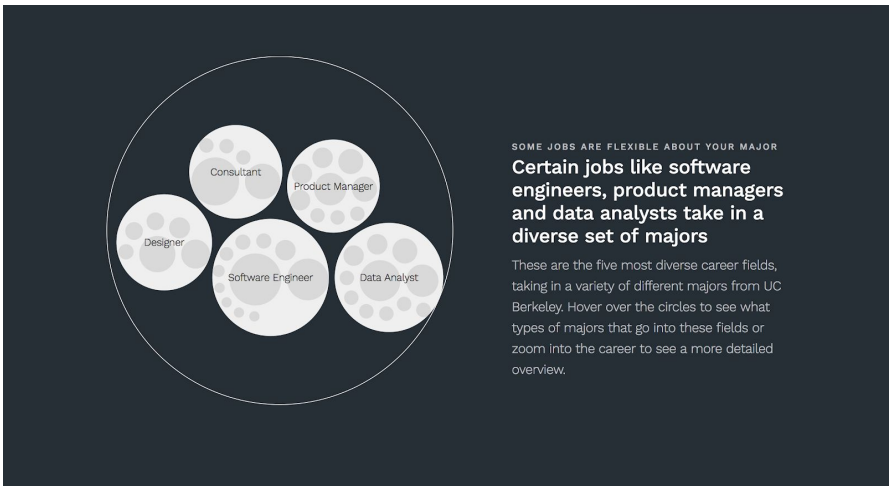
Showing an Overview of Different Pathways

We then try to debunk the myth that only certain majors can lead to certain jobs using a modified chord diagram that shows how one major field can lead to multiple career fields, particular in ways that you might not expect.



Careers that take in diverse majors

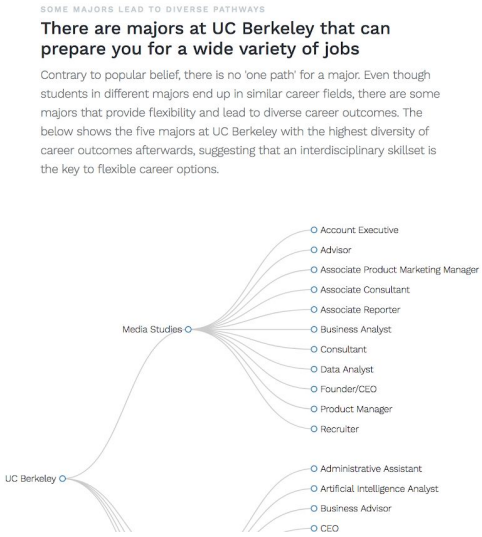
UC Berkeley undergraduates that we spoke to during our exploratory interviews had a fixed mindset in which they only associated certain majors with certain jobs and that most jobs have fixed requirements. In order to show that there jobs don't only take on specific type of major and that there are jobs where a diverse range of majors are welcome, we did some EDA and narrowed down to 5 careers that take in the most diverse majors[1]. We created a cluster diagram that would show one large bubble with bubbles inside it that contain the different majors that are currently in that field, with the size of the bubble representing the number of people from each major. They can zoom into the career to see all the majors that are in that career as well as hover to see the number of people.



[1] See Appendix 2.2 for calculations

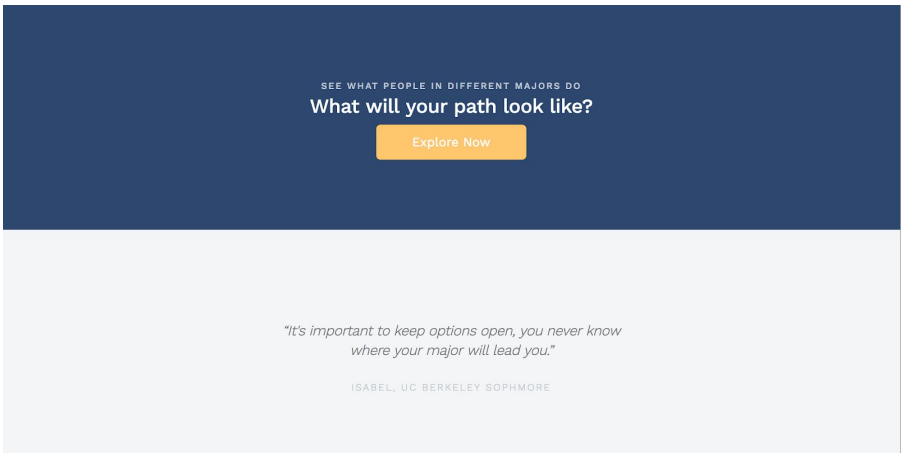
Majors that lead to diverse careers

We then tried to show how diversity can work the other way around in that there are certain majors that can have many options following graduation and that there is not always one ‘set path’ as undergraduates often believe. After conducting EDA using Tableau, we narrowed down to the top 5 majors that had the most diverse career outcomes[2] using a collapsible tree diagram.



EXPLORATION COMPONENT

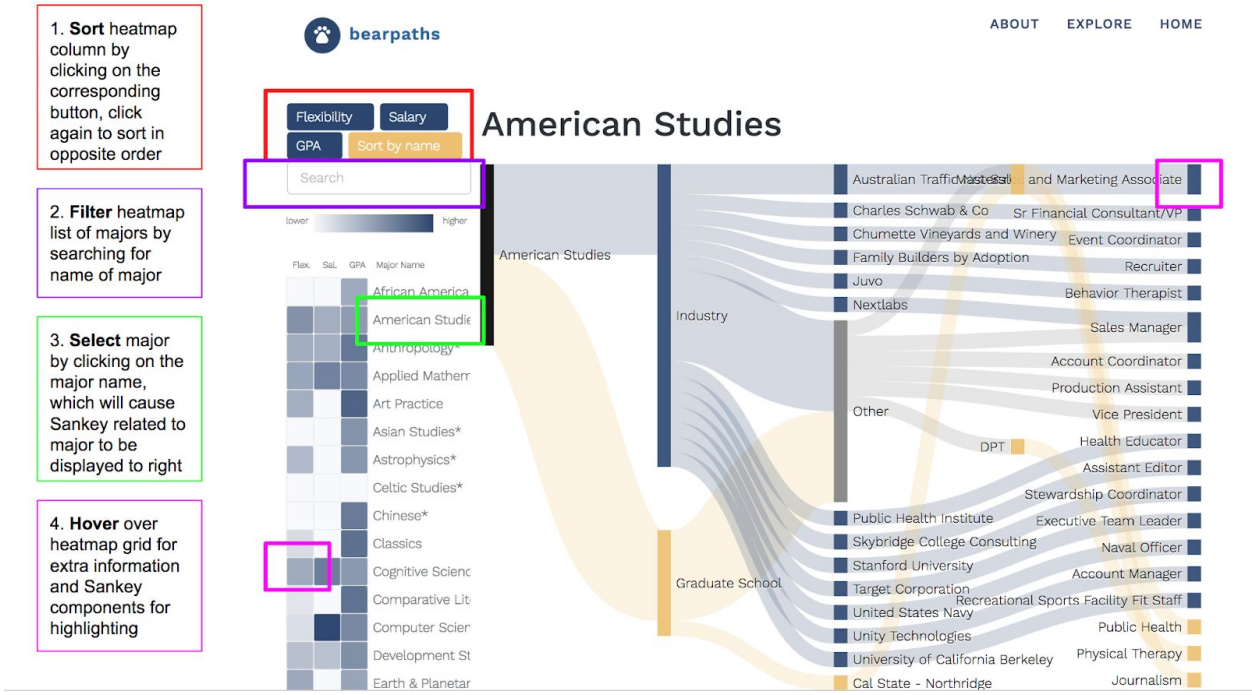
Start exploring button at the end of the home page & quote from user interviews



[2] See Appendix 2.3 for calculations

Exploration Tool

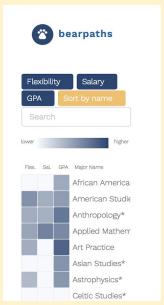
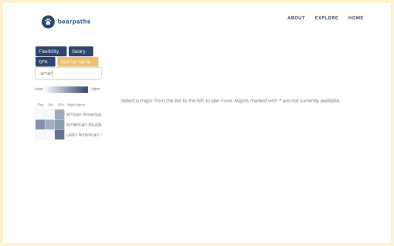
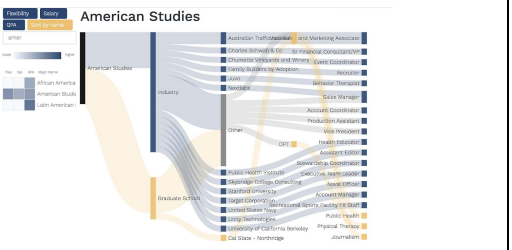
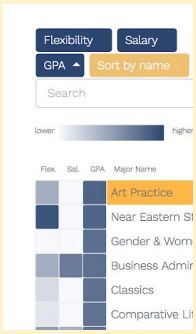
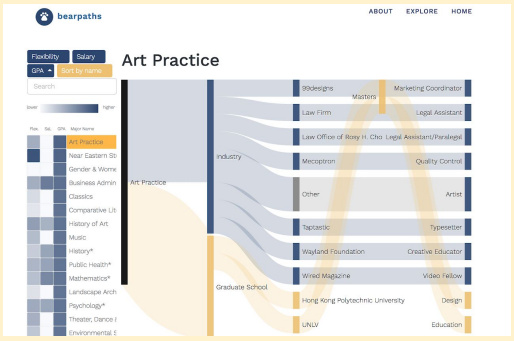
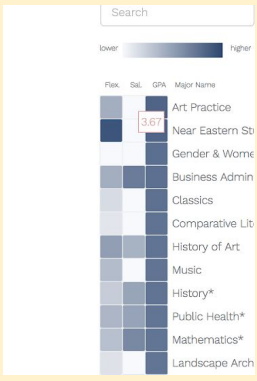
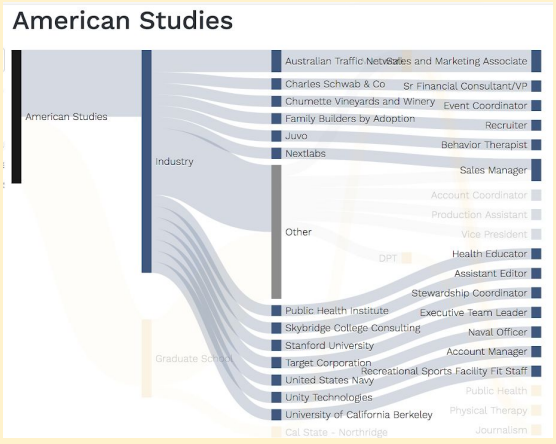
The exploration tool supports four types of interactions, shown below:



Design concepts that we thought was important for the interaction interface to support was 1. to visually provide a way for the user to imagine a pathway between the choice of major and a future outcome, and 2. to suggest alternative ways of navigating to majors other than popularity.

With regards to 1. we choose to pursue the visualization of the Sankey for first year career outcomes as it would allow us to show many paths, create a sense of flow, and show any potential intersections between locations, topics, or job titles.

For 2. we eventually came up with the design of a navigation where the user could sort via more than one metric to get to the major they want: allowing them to use a preferred metric (e.g. gpa) as well as present an alternative (e.g. flexibility) without too much weight given to either one. The navigation is coupled with a heatmap, allowing an user to quickly see increase/decrease without ascribing too much weight on the exact value. The navigation is also searchable by keywords to allow users who already know what they want to get there more quickly. Lastly, hovering is used in the heatmap to display the exact values should user want them and on the Sankey to highlight specific pathways.

Interaction	In practice	
1. Sort		
2. Filter		
3. Select		
4. Hover		

OVERVIEW COMPONENT

About Page

In order to provide more context about the purpose of the website and the approach we used to display our visualizations, we created an About page that could show users how we came up with the metrics that they see as well as provide more information about the user interviews and surveys we used to arrive at the design decisions we made by using a simple bar chart to illustrate our survey results[3].

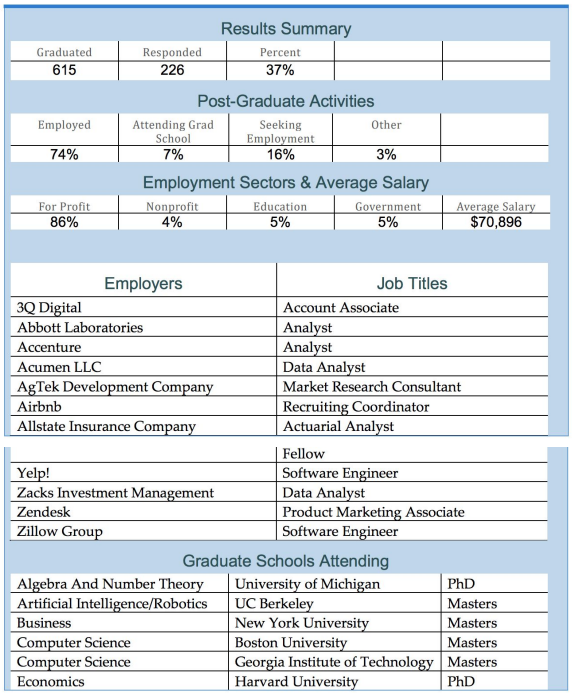
[3] See Appendix Section 1.6 for more details

Implementation

DATA

For Background

Our primary source of information came from UC Berkeley's Career Center's Survey of Career Destination by Major. The survey is conducted for UC Berkeley undergraduates regarding their career outcome within one year of graduating. We used the dataset for students who completed their undergraduate education at UC Berkeley in 2017[4].



Results Summary

Graduated	Responded	Percent		
615	226	37%		

Post-Graduate Activities

Employed	Attending Grad School	Seeking Employment	Other	
74%	7%	16%	3%	

Employment Sectors & Average Salary

For Profit	Nonprofit	Education	Government	Average Salary
86%	4%	5%	5%	$70,896

Employers	Job Titles
3Q Digital	Account Associate
Abbott Laboratories	Analyst
Accenture	Analyst
Acumen LLC	Data Analyst
AgTek Development Company	Market Research Consultant
Airbnb	Recruiting Coordinator
Allstate Insurance Company	Actuarial Analyst
	Fellow
Yelp!	Software Engineer
Zacks Investment Management	Data Analyst
Zendesk	Product Marketing Associate
Zillow Group	Software Engineer

Graduate Schools Attending

Algebra And Number Theory	University of Michigan	PhD
Artificial Intelligence/Robotics	UC Berkeley	Masters
Business	New York University	Masters
Computer Science	Boston University	Masters
Computer Science	Georgia Institute of Technology	Masters
Economics	Harvard University	PhD

The data set is a list of pdfs which contain high level overview of survey results and as well as individual listings of employers and job titles for those who go into industry and field, program type, and school for those who continue in graduate school.

These pdfs were converted into two separate datatables, DataTable 1 General Overview and DataTable 2 First Year Career Trajectories, shown with first four rows in the Appendix.

The General Overview Data was combined with the Average GPA By Major Dataset from the UC Berkeley Office of Planning and Analysis[5] for use in the heatmap navigation dataset. Sankey (explore) and other visualizations (home - context) all used DataTable 2 First Year Career Trajectories. Data used for heatmap navigation and interactive sankey used additional calculations and/or processing which is included in Section 2 of the Appendix.

[4] https://career.berkeley.edu/Survey/2017Majors
[5] https://pages.github.berkeley.edu/OPA/our-berkeley/gpa-by-major.html

Where Data was Used in Final Project

Page	Visualization	Data Set	Source
Home (Contextual Component)	Various	DataTable 2 First Year Career Trajectories Appendix Section 1.5	2017 UCB Major Survey[6]
Explore (Interactive Exploration)	Heatmap	DataTable 3 Navigation	1. 2017 UCB Major Survey 2. UCB OPA: GPA by Major[7]
	Sankey	Appendix Section 2.2	2017 UCB Major Survey

Data available at:
https://docs.google.com/spreadsheets/d/1PsqYJ-hjUR3YqNksEmFY0Wo2IWH6_CLAZBAts2NeGvg/edit?usp=sharing

TOOLS

Tools Used in Design and Development of Project

Activity	Component	Tool
Design	Sketching	Pen and paper
	Graphics	Illustrator, Sketch
	Presentation	Google Slides
Data	Input	Google Sheets
	EDA	Google Sheets, Tableau
	Transformation	Excel
	Generation	Python[8]

[6] https://career.berkeley.edu/Survey/2017Majors
[7] https://pages.github.berkeley.edu/OPA/our-berkeley/gpa-by-major.html
[8] Heatmap navigation and Interactive Sankeys were programmatically generated by custom-written python scripts. More details in Appendix Section 2.

Tools Used in Building of Web Product

Page	Component	Tool (source)
All Pages	Layout	Bootstrap (src) Google Fonts (src)
	Graphics	Illustrator
	Visualizations	D3.js (src)
Home (Contextual Component)	Viz 1: "IT'S NOT ALWAYS WHAT YOU EXPECT"	D3 Stretched Chord Diagram (src)
	Viz 2: "SOME JOBS ARE FLEXIBLE ABOUT YOUR MAJOR"	D3 Cluster Graph (src)
	Viz 3: "SOME MAJORS LEAD TO DIVERSE PATHWAYS"	D3 Collapsible Tree Example (src)
Explore (Interactive Exploration)	Heatmap	HTML, Javascript (built in) JQuery (src) List.js (src)[9]
	Sankey	D3 Sankey Example (src)

IMPLEMENTATION PROCESS

1. **Data Collection and Low Fidelity Sketching**

In order to figure out the overall story we were trying to tell and whether or not it was even feasible, in our early meetings we each created sketches and decided on the best narrative. At the same time, we were searching for any data we could find surrounding student pathways after college in order to determine our constraints. Therefore once we found the PDF's from Berkeley's Career Center, we moved onto the next step.

2. **Data Processing and EDA**

Using the PDF's we manually went through each major and entered in all the values in a Google Sheet. From these values, we were also able to generate additional

[9] Heatmap navigation tool is a custom built tool from HTML Table and JQuery. Combined with the list.js Javascript library for sorting, searching elements in a list.

metrics such as flexibility of a particular major. Once it was in this format, we began using Tableau to explore interesting trends that we could spin into our narrative. From there, we decided on the optimal visualizations that we could use in order to share this insight with others.

3. User Interviews

We conducted user interviews to find out what would be the most effective way to visualize our data and what information mattered most to undergraduates in order to get a better sense of the problem we were trying to solve through our interactive visualization tool. We went to Moffitt Library to interview 5 undergraduates and found out that there was a desire to find out what people end up doing after graduation and that a lot of undergraduates had never seen the PDFs from the career center before; they found that the data was interesting but it could be more visually compelling. Some interesting insights we found were that flexibility mattered a lot to people but that they also found it interesting to see the variety of jobs that people can do. A lot of people thought that a major was 'fixed' but it was nice for them to know that there are many options open to them after post-graduation[10].

4. Initial Prototype

Based off our EDA and user interviews, we began to create user flows and designing the overall structure of the website. For the contextual visualizations on the home and about page, we hard-coded the data that was used. However, for the Sankey, we wanted to make sure that each major would be able to render its own unique sankey. To do this, we modified the 2017 dataset (Appendix 1.4) and created custom python code to auto-generate JSON files for each major with the proper structure to render nodes and their connection widths (Appendix 2.2) . One concern raised by the course instructors was that in our initial prototype, we planned on using size-coded treemap for navigation (Appendix 3.2). This eventually evolved into the use of a heatmap to encode value differences in selected metrics and allows the user to re-order and search through majors to navigate.

After these visualizations were created, we created a website that housed each of them together to tell a compelling narrative. As a final step, we added styling to each

[10] For full notes from the interviews: https://docs.google.com/document/d/1dECM22HLYzhH4xev9iRpGg2plsBfMowgqZGHuD3PRwU/edit?usp=sharing

D3 visualization and the website to ensure that they complemented each other visually.

5. **User Testing and Refinement**

We tested the design we created with user on Demo Day and our first round of changes. Afterwards, we performed another round of usability tests on 5 students. From these tests, we collected feedback to determine how successful our tool was in helping them learn interesting insights about pathways after college. The usability test also allowed us the observe the beneficial effects of iterating on the feedback from Demo Day. Full details are below in the user testing section.

Usability Testing and Results

Overview

In order to understand to understand how successful we were in creating our tool, it was important to perform user testing with different users within our target demographic. We went out and conducted usability tests with 5 prospective users and also collected informal feedback from others during our demo day. All prospective users were students at UC Berkeley (Graduate and Undergraduates). The feedback gathered from these tests were extremely informative and also indicative of degree of success we had in showing our users the different paths that were available after graduation.

With our user testing, we wanted to see how both functional and aesthetic qualities of our designs would hold up in the real world. More specifically, we wanted to determine:

1. If users learned something interesting about pathways after college
2. Whether or not the experience was enjoyable

Test Setup

During the user test, we used a combination of surveys, observation, and a semi-structured interviews. Before users actually saw our tool, we asked them one question about whether or not they were interested in learning where they were going to end up. This was necessary to determine a baseline for their interest levels in their own future. After they answered the initial scenario, they were told to take some time to explore the website and talk out loud. While they explored the website, we wrote down notes on not only their movement through the website but also their thoughts regarding each graph. When they got to the explore page, we asked them to explore different majors and the paths that would result from each. To end, we asked them to fill out their agreement levels to the following 4 statements:

1. I feel like i've learned something interesting about pathways after college
2. This tool is helpful for understanding pathways after college
3. I enjoyed using this tool
4. I would share this tool with someone else

In order to capture the right level of granularity, our scoring guide looked like this:

1 - strongly disagree
2 - disagree
3 - somewhat disagree
4 - undecided
5 - somewhat agree
6 - agree
7 - strongly agree

Result Analysis

In total we were able to run usability tests on 3 undergraduates in Moffitt Library and 2 graduate students in South Hall. From the pretest, we saw that most people agreed that they were interested in where they might end up. From a high level, almost all of the users had a positive experience and were able to learn something interesting about the pathways after college. Not only that, but they stated that they had a really insightful experience and would have shared the tool with another one of their friends. The average response for each of the following statements was in the range of ‘agree’ to ‘strongly agree’.

Pre Test Questions

Question	Average Score
I’m interested in knowing where I might end up after school	6

Post Test Questions

Question	Average Score
I feel like I’ve learned something interesting about pathways after college	6.4
This tool is helpful for understanding pathways after college	6.2
I enjoyed using this tool	6.4
I would share this tool with someone else	6.4

Overall Impression and Usability of the Visualization

After the survey, we asked each participant more exploratory questions in order to capture broader data points. In these discussions we also showed them the original pdfs that were currently being used by the Berkeley Career Center. Afterwards, we asked:

- Which tool do you think would be easier to use to compare majors?
- What was your favorite visualization?
- What were some of the confusing things about this visualization?
- What do you wish this visualization had?
- Do you have any other comments/questions?

Overall, most participants said that they preferred our tool. One indicative comment by an undergraduate regarding the pdfs was “Oh wow, that’s not easy to look at at all”. However, it is interesting to note that we did have one participant say she preferred the original pdfs due to the hard numbers that were present and not apparent in our sankey. The rest of the observations were grouped into categories for each visualization.

Visualization: Stretched Chord Diagram

It was very clear that most people thought it was intuitive to hover over each of the sections. Surprisingly, people spent quite a bit of time exploring each section and where it was connected to. One thing we did notice was that they went straight to the graph and didn’t usually read the text explanation next to it. For a few of the participants, it took them a couple moments to realize that the thickness of the band correlated with the amount of people. In our demo day, people said that the text was too small and we’ve taken steps to make the text easier to read. Several users were interested in results that were unexpected.

Visualization: Cluster Graph

When participants saw that the circle did have zoom capacity, they were intrigued and kept exploring each circle. It was also interesting to see that several of them stayed zoomed in and explored other circle at the same inspection level. Some surprising reactions were hopeful. One participant majoring in economics was very cynical of what she thought would be her future and when she saw that many became data analysts, she was really shocked. We could tell that she had never known this career transition before and that she was curious to learn more. For

future improvements, we are planning on displaying names even for the smaller circles. Also, we want to include visual aids that indicate the hotspots users can click on since some users were unsure of where to click at first.

Visualization: Collapsible Tree

This was one of the highest regarded visualizations in our tool. Users liked that it didn't overwhelm the senses and that they could control how much they wanted to see or keep hidden. One participant noted, "I really like this one because it's really intentional. It doesn't show all the data at once". Another said, "This to me, is really helpful". Not only that but it was surprisingly intuitive even for users not used to advanced visualizations. They were aware of the clickable functionality without being told.

Visualization: Sankey

Even before our first iteration, we were noticing that users loved the aesthetics of this page. Several participants noted "woah, this is so pretty" upon getting to this screen. Before our first iteration of changes, we noticed that some people weren't entirely sure how to interpret the "flexibility" button. Not only that, but they didn't know they were supposed to click on a major in order to have it scroll down to the sankey. As such, we repositioned the sankey to generate right next to the column of majors. Not only that, we've added a legend and indicators that provide status updates on whether or not users are sorting by ascending or descending. The legend also helps to explain that the darker the shade, the higher the quantity. What we found was that there was a dramatic improvement in our second round of user testing. Based on what people said they wished they were able to see, we're considering adding a tooltip that appears for each sankey path and also trying to minimize sankey path crossovers.

Overall Feedback

It was clear that there was a drastic improvement in feedback after our first couple of iterations. From a holistic standpoint, the general feedback from our users was that they really loved the site and that they had learned some interesting lessons along the way (career pivots they weren't expecting and the amount of people who went down each path). Most users were interested enough to take the time to really dive deep into what our visualizations were showing. As such, there were many requests for more information to be shown in the future. From the start, we prioritized showing users insightful paths they might not have been aware of. Now that this has been accomplished, we can also consider how to provide an even greater depth and extend beyond one years worth of data.

Links to Demo and Code

Link to Visualization: http://people.ischool.berkeley.edu/~lilyelin/bearpaths/

Link to Github Repo for Source Code: https://github.com/lilyelin/collegepaths

Link to Data:
https://docs.google.com/spreadsheets/d/1PsqYJ-hjUR3YqNksEmFY0Wo2IWH6_CLAZBAts2NeGvg/edit?usp=sharing

Link to Data Processing Scripts:
https://github.com/lilyelin/collegepaths/blob/master/customscripts/scripts.py

Link to General Data:

https://docs.google.com/spreadsheets/d/1PsqYJ-hjUR3YqNksEmFY0Wo2IWH6_CLAZBAts2NeGvg/edit?usp=sharing

Link to Sankey Data:

https://github.com/lilyelin/collegepaths/tree/master/sankeydata

Distribution of Work

Project Component	Sub Component	Daphne	Devin	Lily
Data Preparation	Data Sourcing	5%	20%	75%
	Data Research & Selection	33%	33%	33%
	Data Structure	0%	0%	100%
Data Analysis	EDA & Dataset Merging	25%	0%	75%
Data Generation	Heatmap (Appendix 1.3, 2.1)	0%	0%	100%
	Sankey (Appendix 2.2)	0%	0%	100%
Design	Website Graphics & Aesthetics	80%	15%	5%
	Front-End Design & Dev.	95%	0%	5%
	User Flow Sketching	33%	33%	33%
	Context Designs	80%	10%	10%
	Explore Interface Designs	15%	15%	70%
Visualization & Programming	About Page Contextual	100%	0%	0%
	Chord Diagram	30%	70%	0%
	Cluster Diagram	100%	0%	0%
	Collapsible Tree Diagram	100%	0%	0%
	Layout (Header, Colors)	80%	10%	10%
	Exploration Sankey	0%	10%	90%
	Heatmap Navigator	0%	0%	100%
	Data Pipeline & Interaction between Sankey and Heatmap (Sort, search, select)	0%	0%	100%
User Testing	Interview and Heuristic Design	50%	50%	0%
	Interviews	33%	33%	33%
	Usability Testing Design	0%	100%	0%
	Usability Testing	0%	100%	0%
Presentation & Report	Presentation	30%	40%	30%
	Report Writing	33%	33%	33%

Appendix

Section 1: Data Set Creation

1.1 DataTable 1: General Overview

Year	College	Major	Graduated	Responded	Percent	Employed	Grad School	Seeking Employment	Other	For Profit	Non profit	Education	Government	Average Salary
2017	L&S	African American Studies	9	4	44	0	100	0	0	0	0	0	0	NA
2017	L&S	American Studies	84	27	32	59	11	26	4	65	24	6	6	47550
2017	L&S	Anthropology	94	31	33	42	13	26	19	59	35	6	0	46678
2017	L&S	Applied Mathematics	232	75	32	64	19	13	4	76	0	20	4	72586

1.2 DataTable 2: First Year Career Trajectories

College	Major	Postpath	Employer	Job Title	Grad School	Type	Program Name
L&S	African American Studies	Grad			Brown University	Masters	African American Studies
L&S	African American Studies	Grad			UC Berkeley	PhD	African American Studies
L&S	American Studies	Industry	Australian Traffic Network	Sales and Marketing Associate			
L&S	American Studies	Industry	Charles Schwab & Co	Sr Financial Consultant/VP			

1.3 DataTable 3: Navigation

college	major	graduate	employed	grad	seeking	salary	gpa	flexibility	advising
L&S	African American Studies	9	0	100	0	NA	2.92	NA	L&S Social Sciences Division
L&S	American Studies	84	59	11	26	47550	3.07	0.5244444444	L&S Undergrad Studies Division
L&S	Anthropology	94	42	13	26	46678	3.45	0.39375	L&S Social Sciences Division
L&S	Applied Mathematics	232	64	19	13	72586	3.26	0.4357446809	L&S Math & Phys Sciences Div

“gpa” was sourced from the UC Berkeley OPA’s “GPA by Major” Dataset. All other data was from the UC Berkeley Career Center’s 2017 UCB Major Survey.

1.4 DataTable 4: Sankey Career Counts

industry.csv

major	employer	title	count
American Studies	Australian Traffic Network	Sales and Marketing Associate	2
American Studies	Charles Schwab & Co	Sr Financial Consultant/VP	1

grad.csv

major	school	type	field	count
African American Studies	Brown University	Brown University	African American Studies	1
African American Studies	UC Berkeley	PhD	African American Studies	1

1.5 DataTable 5: Chord Diagram Overview of Pathways

overview.csv

Major	**Major Field**	**Company Name**	**Job Title**	**Area**
American Studies	Humanities	Chumette Vineyards and Winery	Event Coordinator	Administration
American Studies	Humanities	Stanford University	Stewardship Coordinator	Administration
American Studies	Humanities	Target Corporation	Executive Team Leader	Administration

matrix_overview.csv

Major Field	Administration	Arts and Design	Business Management	Law	Health/Medicine	Technology	Writing/Communication	Consulting	Education	Sales	Government	Engineering	Other
Art	0	13	1	0	0	0	0	2	1	0	2	2	2
Business	2	4	154	0	0	16	1	28	2	7	1	1	10
Social Sciences	44	7	192	16	17	53	2	35	46	16	16	6	40
Science	24	2	80	1	25	47	0	0	40	4	4	8	24
Humanities	51	13	73	13	7	8	11	9	28	21	8	3	29
Natural Resources	11	1	24	0	1	5	0	2	7	1	5	0	8
Engineering	1	0	20	0	1	250	0	12	15	1	0	83	13

We used a sort and filtering system to manually group majors and industries from the original dataset into a more generalized grouping then transformed it into a matrix

1.6 DataTable 6: About Page Metrics

Area	Score
Personal interest	3.5
Reputation of major	2
Skills a major can gain	3.5
Job title post graduation	4
Company post graduation	4
Salary post graduation	4

We conducted interviews with 5 undergraduates and averaged the scores across participants to get a sense of what was most important to them when considering a major to provide more context about why we have chosen to display the navigation in the way that we have, where we allow users to filter by salary whilst also showing the job title and company post graduation in the Sankey.

Section 2: Data Calculations

2.1 For Heatmap Navigation (Explore)

The term “flexibility” was used to describe the likelihood of a major leading to different job outcomes for graduates in industry. In the future, flexibility should be re-calculated to include graduate school outcomes or a separate field regarding graduate studies be included.

The field “flexibility” was calculated according to the following formula:

$$\text{Flexibility} = \text{Job Variability} * \text{Employment Ratio}$$

$$\text{Job Variability} = \textit{distinct counts of} \text{ “Job Titles”} / \textit{ total count of} \text{ “Job Titles”}$$

$$\text{Employment Ratio} = \text{“employed”}/100$$

Thus:

Flexibility = (*distinct counts of* “Job Titles”/ *total count of* “Job Titles”) * (“employed”/100)

Note that *distinct counts of* and *total count of* were calculated via Tableau.

2.2 For Interactive Sankey (Explore)

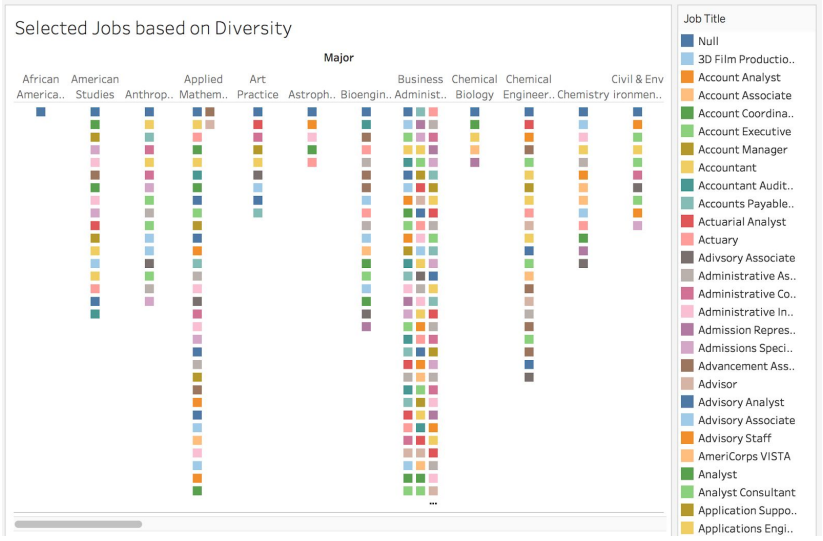
Using DataTable 2, we created a new dataset DataTable 4, which organizes career outcomes by counts of students. This was created via Tableau. The following code was used to generate JSON files for each major.

https://github.com/lilyelin/collegepaths/blob/26dfa77a861d7c4f5aa1a72f1e22843cd9b9c2f6/customscripts/scripts.py#L286

2.3 For Job Diversity Cluster Graph (Home)

Using DataTable2, we calculated the “diversity” of the job by looking at the total distinct number of varying majors that fall under that job, creating a new calculated column called “outcome diversity” using Tableau.

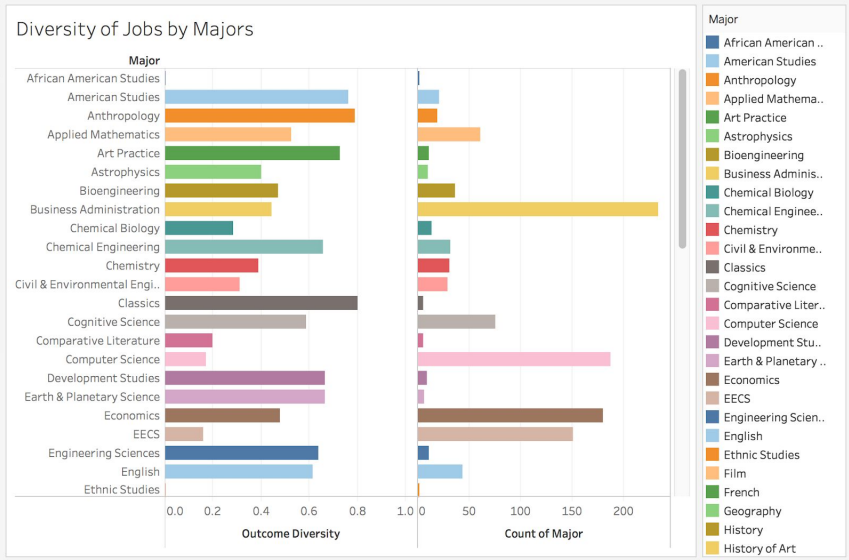
COUNTD([Job Title])/COUNT([Major])



2.4 For Major Diversity Collapsible Tree (Home)

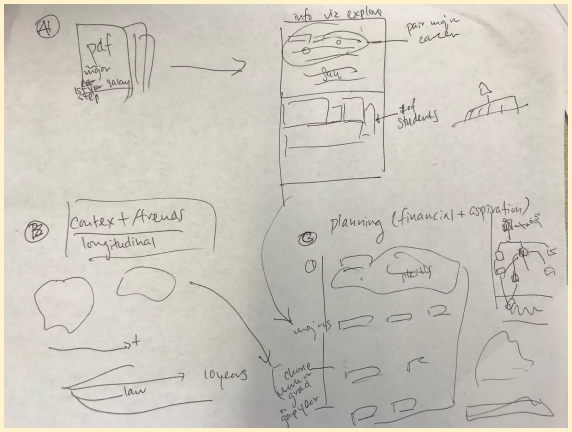
Using DataTable2, we calculated the “diversity” of the major by looking at the total distinct number of varying jobs that the major leads to, creating a new calculated column called “major diversity in jobs” using Tableau.

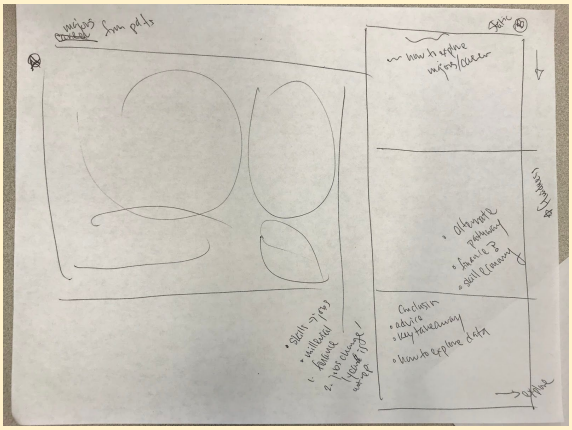
COUNT([Job Title])/COUNTD([Major])



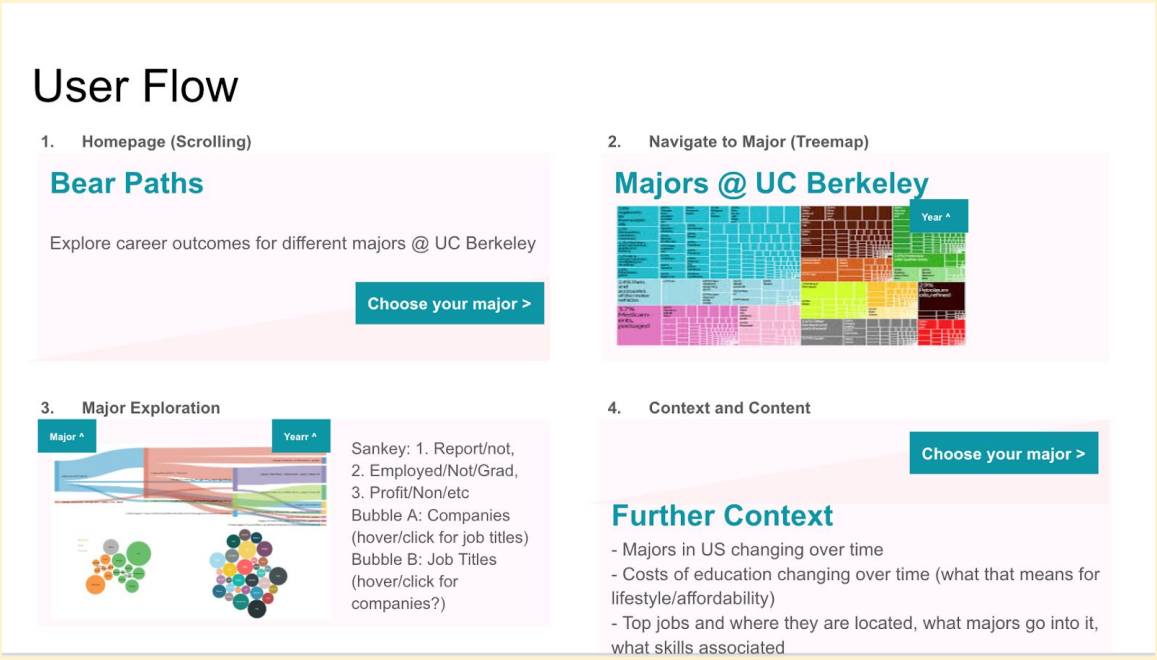
Section 3: Evolution of Design Concepts through Sketches and Mock-ups

3.1

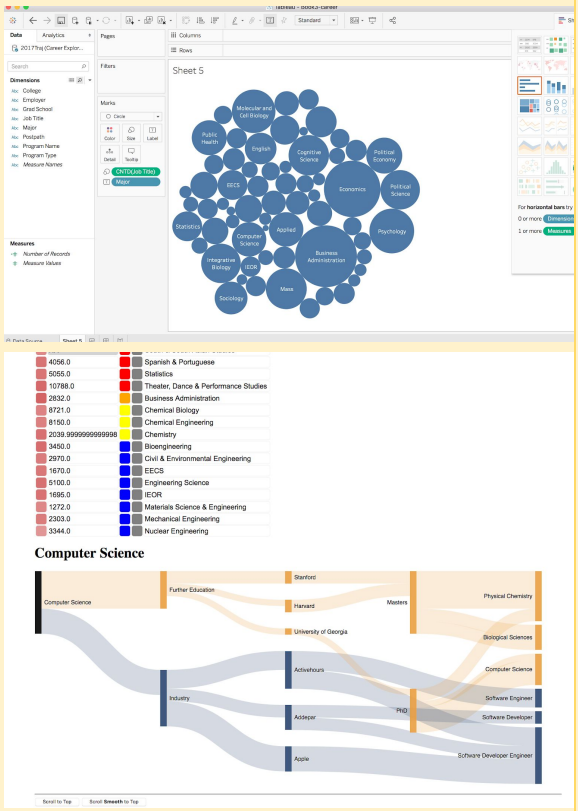


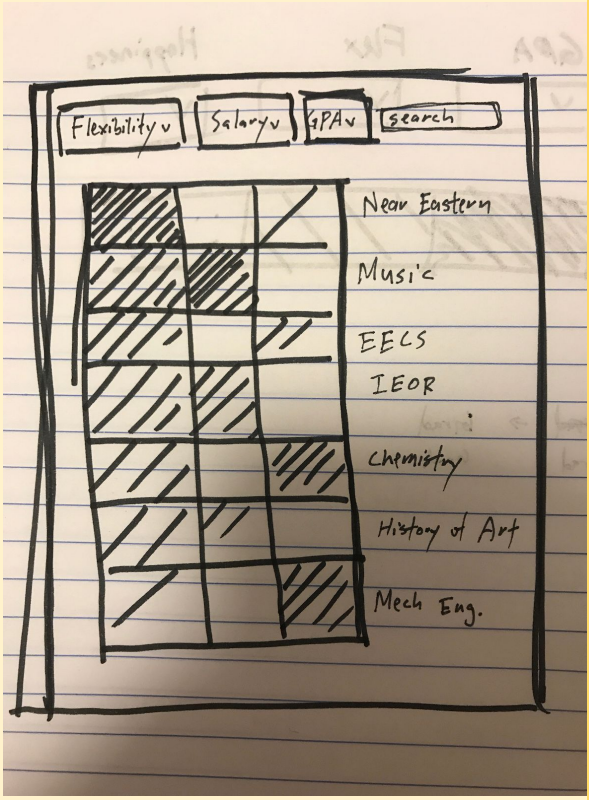


3.2



3.3


Sheet 5
Molecular and Cell Biology
Public Health
English
Cognitive Science
Political Economy
Economics
Political Science
EECS
Statistics
Computer Science
Applied
Psychology
Integrative Biology
IEOR
Business Administration
Mass
Sociology
4056.0 Spanish & Portuguese
5055.0 Statistics
10788.0 Theater, Dance & Performance Studies
2832.0 Business Administration
8721.0 Chemical Biology
8150.0 Chemical Engineering
2039.9999999999998 Chemistry
3450.0 Bioengineering
2970.0 Civil & Environmental Engineering
1670.0 EECS
5100.0 Engineering Science
1695.0 IEOR
1272.0 Materials Science & Engineering
2303.0 Mechanical Engineering
3344.0 Nuclear Engineering
Computer Science
Computer Science
Further Education
Industry
Stanford
Harvard
University of Georgia
Activehours
Addepar
Apple
Masters
PhD
Physical Chemistry
Biological Sciences
Computer Science
Software Engineer
Software Developer
Software Developer Engineer


Flexibility
Salary
GPA
Search
Near Eastern
Music
EECS
IEOR
Chemistry
History of Art
Mech Eng.