

Harvard Concentrations

CS171 Project 3
Spring 2013
By Jerry Chang and Lucas Lin

Project 3 Proposal

Project Title

What Harvard Students Study

Team

Jerry Chang <jerrychang@college.harvard.edu>

Lucas Lin <lucaslin@college.harvard.edu>

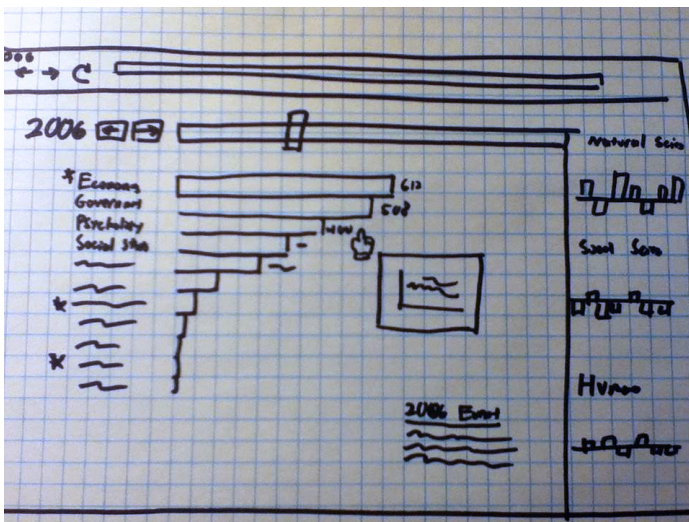
Visualization

We will build upon our bar charts visualization from Project 2 with additional views and explanations. Our visualization for Project 3 will have a similar bar chart to Project 2 except the bars for each concentration will be oriented horizontally and ordered by the number of students in a particular year. This will make it easier to pinpoint the most popular concentrations for that year. This list will also be filterable by category (e.g. Social Science, Natural Science, and Humanities). The drop down to select the year will be modified to only feature a scrollable timeline. When the user hovers over the bars, a tooltip will appear featuring a line graph of how the percent of students in that concentration has changed over time from 2000-2012. For one concentration in each category, we will also display a more detailed story/possible explanations for the trends shown. Tentatively the concentrations we will highlight are Economics, Computer Science, and English. On a separate sidebar, we will display three bar charts of how the categories have changed over time and mention how the social sciences and natural sciences appear to be inversely related. Overall there will be two additional views a line graph within the tooltip and a separate bar graph of the changes in the categories as well as a story module for three of the concentrations.

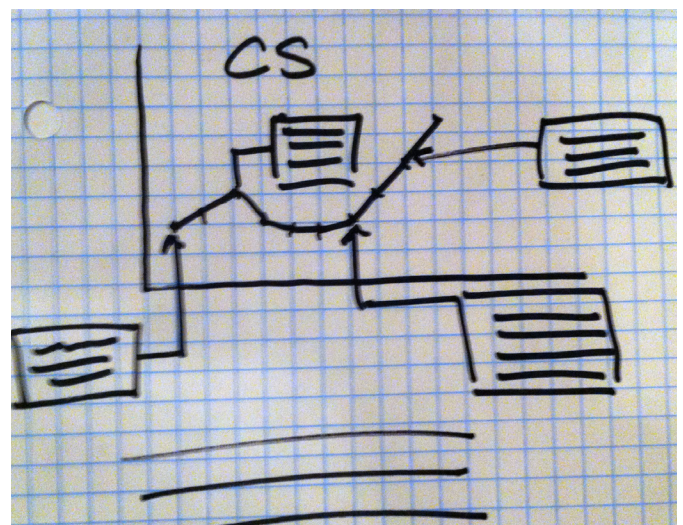
TF Meeting

Based on our meeting with our TF, we decided on the number of additional elements we would include as well as the storytelling elements we would incorporate for Project 3.

Sketches



Full Visualization



Story Visualization

Restructuring Data

In order to easily draw line graphs, our JSON data from Project 2 needed to be modified so that the data for every year is within the dictionary for its concentration. Rather than have a separate json dictionary for every year of a concentration (i.e. 13 dictionaries per concentration), we want to consolidate the data for all years into a single dictionary leaving us with one dictionary per concentration. Thus when we draw the line graphs we will only need to access a single dictionary rather than 13.

To restructure our data we converted the rows from our csv file into columns.

Concentration	Category	Year	Number of Stuc	Change in Num	% of Students	Change in %
African and Afri	Social Sciences	2000	19	0	0.0040783785	0
African and Afri	Social Sciences	2001	15	-4	0.003203075	-0.000873304
African and Afri	Social Sciences	2002	14	-1	0.0029648454	-0.00023823
African and Afri	Social Sciences	2003	8	-6	0.0017338535	-0.001230992
African and Afri	Social Sciences	2004	11	3	0.0024070022	0.0006731487
African and Afri	Social Sciences	2005	9	-2	0.0019612116	-0.000445791
African and Afri	Social Sciences	2006	11	2	0.0023474178	0.0003862062
African and Afri	Social Sciences	2007	15	4	0.0039936652	0.0010462473
African and Afri	Social Sciences	2008	21	6	0.0046604527	0.0012667876
African and Afri	Social Sciences	2009	19	-2	0.004039703	-0.000626482
African and Afri	Social Sciences	2010	11	-8	0.0023295214	-0.001704449
African and Afri	Social Sciences	2011	13	2	0.0027339642	0.0004044429
African and Afri	Social Sciences	2012	30	17	0.0059370671	0.0032031026
Anthropology	Social Sciences	2000	155	0	0.032546664	0
Anthropology	Social Sciences	2001	135	-20	0.0288276746	-0.004426992
Anthropology	Social Sciences	2002	149	14	0.0315544261	0.0027267515
Anthropology	Social Sciences	2003	126	-23	0.0273081925	-0.004246234
Anthropology	Social Sciences	2004	151	25	0.0330415755	0.005733383
Anthropology	Social Sciences	2005	195	44	0.0424929178	0.0094513424
Anthropology	Social Sciences	2006	173	-22	0.0369184806	-0.005574437
Anthropology	Social Sciences	2007	163	-10	0.0368778281	-0.000040653
Anthropology	Social Sciences	2008	126	-37	0.0279627164	-0.008915112
Anthropology	Social Sciences	2009	113	-13	0.0239915074	-0.003971209
Anthropology	Social Sciences	2010	93	-20	0.0196950445	-0.004296463
Anthropology	Social Sciences	2011	69	-24	0.014511041	-0.005184003
Anthropology	Social Sciences	2012	73	4	0.0144486832	-0.000064178
Applied Mather	Natural Science	2000	130	0	0.0278910105	0
Applied Mather	Natural Science	2001	127	-3	0.0271193679	-0.000771643
Applied Mather	Natural Science	2002	125	-2	0.026471834	-0.000647534
Applied Mather	Natural Science	2003	104	-21	0.0225400954	-0.003931739
Applied Mather	Natural Science	2004	101	-3	0.0221006565	-0.000439439
Applied Mather	Natural Science	2005	103	2	0.0224449771	0.0003443207
Applied Mather	Natural Science	2006	114	11	0.0243277849	0.0018828078
Applied Mather	Natural Science	2007	121	7	0.0273755656	0.0030477807
Applied Mather	Natural Science	2008	101	-20	0.0224145584	-0.004961007
Applied Mather	Natural Science	2009	159	58	0.0337579618	0.0113434034
Applied Mather	Natural Science	2010	177	18	0.0374841169	0.0037261551
Applied Mather	Natural Science	2011	196	19	0.0412197867	0.0037956518
Applied Mather	Natural Science	2012	221	25	0.0437939342	0.0025166256
Astrophysics	Natural Science	2000	14	0	0.0030036473	0
Astrophysics	Natural Science	2001	10	-4	0.0021353833	-0.000868264
Astrophysics	Natural Science	2002	6	-4	0.001270648	-0.000864735
Astrophysics	Natural Science	2003	8	2	0.0017393535	0.0004632055
Astrophysics	Natural Science	2004	10	2	0.0021881838	0.0004543303
Astrophysics	Natural Science	2005	9	-1	0.0019612116	-0.000226972
Astrophysics	Natural Science	2006	8	-1	0.001707213	-0.000253999
Astrophysics	Natural Science	2007	3	-5	0.000678733	-0.00102848

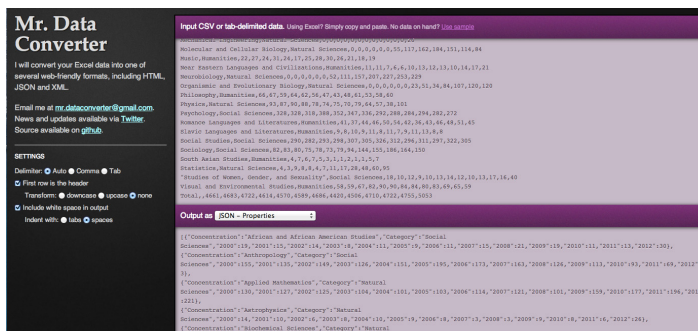
CSV from Project 2

Concentration	Category	2000	2001	2002	2003	2004
African and Afri	Social Sciences	19	15	14	8	11
Anthropology	Social Sciences	155	135	149	126	151
Applied Mather	Natural Science	130	127	125	104	101
Astrophysics	Natural Science	14	10	6	8	10
Biochemical So	Natural Science	262	254	241	234	225
Biology	Natural Science	280	267	297	290	308
Biomedical Eng	Natural Science	0	0	0	0	0
Chemical and F	Natural Science	0	0	0	0	0
Chemistry	Natural Science	78	77	73	89	91
Chemistry and	Natural Science	37	44	38	39	45
Classics	Humanities	41	38	37	34	48
Comparative St	Humanities	43	47	54	41	34
Computer Sciel	Natural Science	203	174	199	98	79
Earth and Plan	Natural Science	25	27	21	24	23
East Asian Stud	Humanities	43	39	48	32	26
Economics	Social Sciences	583	624	666	643	686
Electrical Engin	Natural Science	0	0	0	0	0
Engineering Sci	Natural Science	79	67	80	38	99
English	Humanities	222	231	242	270	246
Environmental &	Natural Science	87	82	77	56	46
Folklore and My	Humanities	14	17	20	17	11
Germanic Lang	Humanities	9	5	3	3	6
Government	Social Sciences	463	547	553	552	549
History	Social Sciences	231	280	280	293	234
History and Lite	Humanities	248	210	197	192	167
History and Sci	Social Sciences	150	139	121	100	92
History of Art	Humanities	47	45	58	57	50
Human Develop	Natural Science	0	0	0	0	0
Human Evolutic	Natural Science	0	0	0	0	0
Linguistics	Social Sciences	31	32	35	27	33
Literature	Humanities	63	52	60	55	53
Mathematics	Natural Science	78	89	69	73	65
Mechanical Eng	Natural Science	0	0	0	0	0
Molecular and	Natural Science	0	0	0	0	0
Music	Humanities	22	27	24	31	24
Near Eastern Lt	Humanities	11	11	7	6	6
Neurobiology	Natural Science	0	0	0	0	0
Organismic anc	Natural Science	0	0	0	0	0
Philosophy	Humanities	66	67	59	64	62
Physics	Natural Science	93	87	90	88	78
Psychology	Social Sciences	328	328	318	388	352
Romance Lang	Humanities	41	37	44	46	50
Slavic Language	Humanities	9	8	10	9	11
Social Studies	Social Sciences	290	282	293	298	307
Sociology	Social Sciences	82	83	80	75	78
South Asian St	Humanities	4	7	6	7	5
Statistics	Natural Science	4	3	9	8	8

Revised CSV for Project 3

Converting CSV to JSON

We used an online JSON converter ([http:// http://shancarter.com/data_converter/](http://http://shancarter.com/data_converter/)) to convert our csv file into JSON.



Adding Data

In order to provide more context, for each concentration we added a course and its difficulty to our csv file. For example for Computer Science we added CS50 along with its average difficulty from the last few years from the Harvard Q Guide. We chose a course to represent each concentration based on the requirements from the Harvard Handbook. This is only an approximate measure of the concentration's difficulty as some concentrations such as Engineering Sciences will have more requirements than the typical Humanities program.

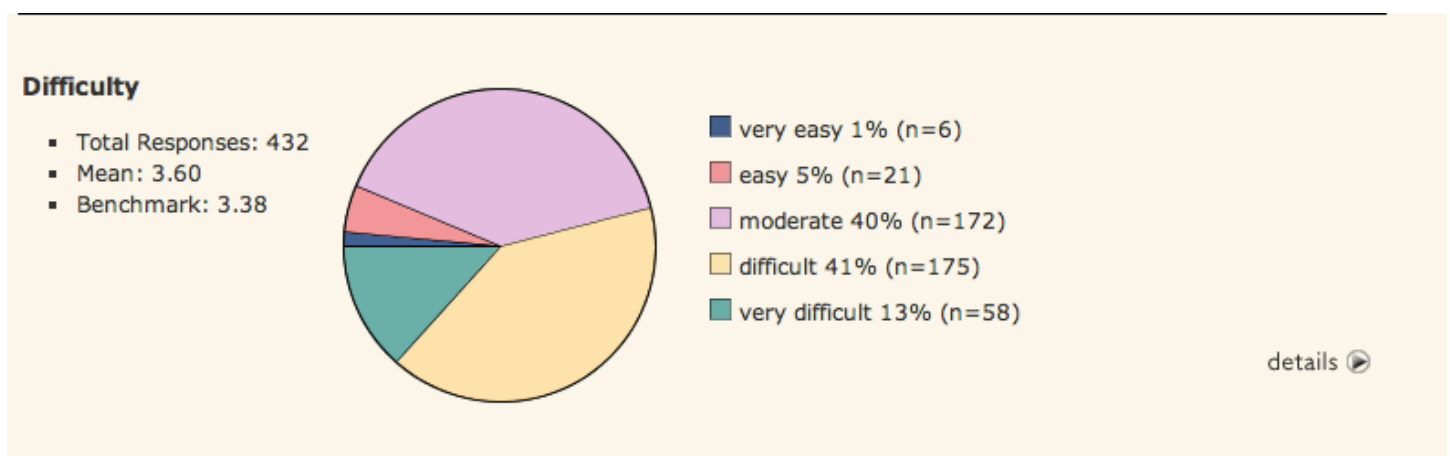
We plan to integrate the difficulty of each concentration into our visualization to make our visual more useful for those comparing concentrations.

REQUIREMENTS

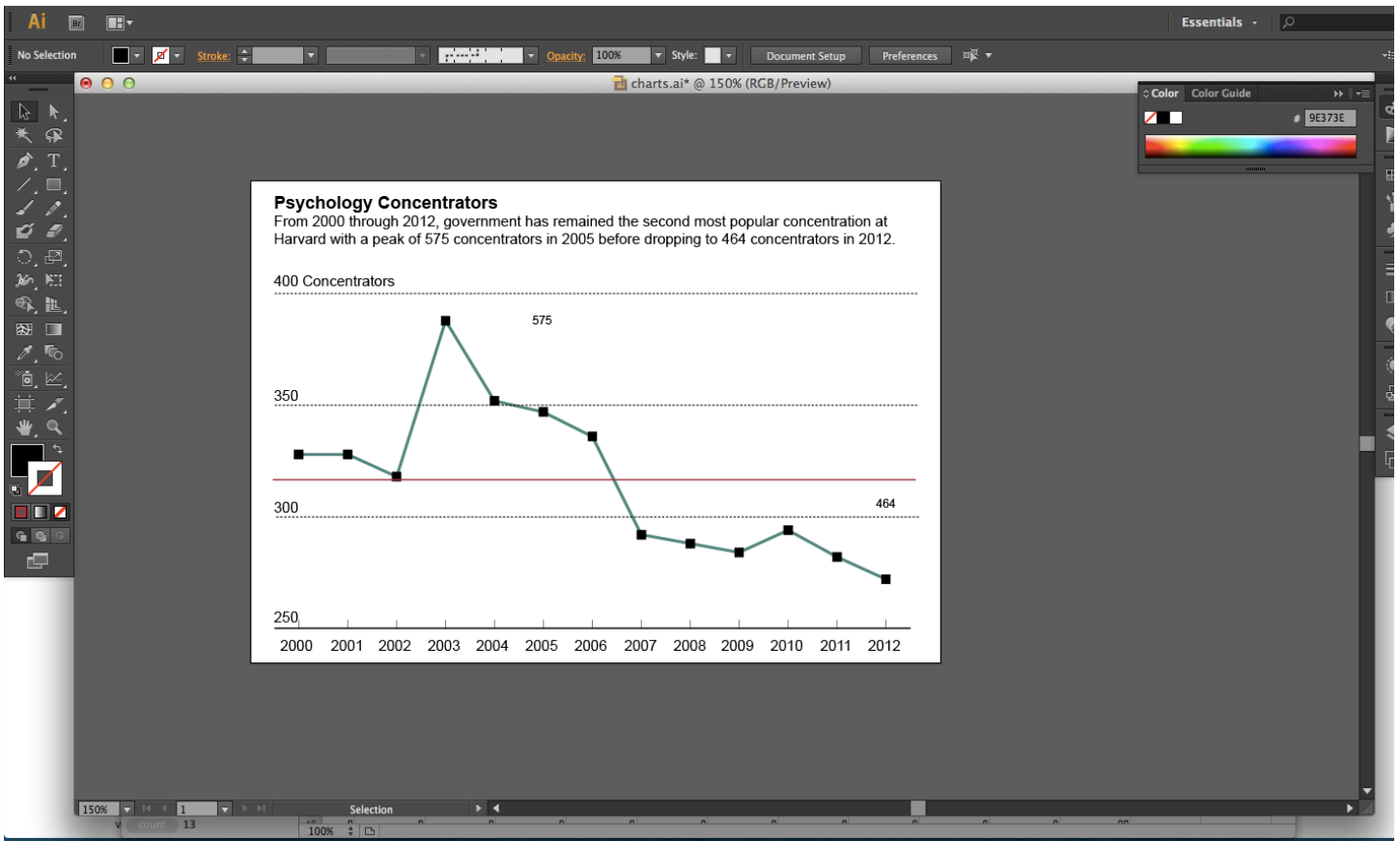
12 half-courses

1. *Required courses:*
 - a. Astronomy 16 and 17. (2 half courses; see 7a. below)
 - b. Physics 15a, 15b, and 15c. (3 half courses; see 7b. below)
 - c. Mathematics 21a and 21b, or Mathematics 23a and 23b, or Mathematics 25a and 25b, or Applied Mathematics 21a and 21b. (2 half courses; see 7c. below)
 - d. Astronomy 98: Research Tutorial, generally taken in the spring semester of the junior year. (1 half course)
 - e. Two additional courses in astronomy. (2 half courses; see 7d. below)
 - f. Two additional courses in astronomy or related fields to complete the requirement of 12 half-courses. (2 half courses; see 7e. below)
2. *Tutorial:* Required, see 1d. above.
3. *Honors Eligibility:* Students who wish to be considered for honors must satisfy requirements 1e. and 1f. by completing Astronomy 99 and/or courses at the 100 level or above. None of the courses satisfying 1e. or 1f. may be taken Pass/Fail. Courses that meet this requirement include:
 - a. Astronomy 99, a full-year course leading to the senior thesis. The Department of Astronomy is located within the Harvard-Smithsonian Center for Astrophysics, one of the world's largest astrophysical research institutes. The Center for Astrophysics offers significant undergraduate research opportunities, which students are encouraged to pursue through the senior thesis.
 - b. Any 100-level or 200-level course in astronomy.
 - c. Physics 143a, 143b, 151, 153, or 181.
 - d. Earth and Planetary Sciences 100, 121, 132, or 150.
 - e. Applied Mathematics 104, 105, 111, or 115.

Requirements for each Concentration



Average Difficulty from Harvard Q Guide

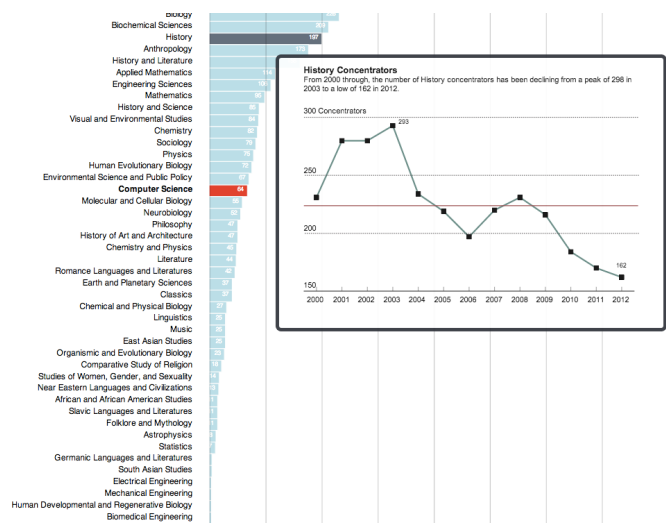


Line Chart in Illustrator

Drawing Line Charts

For each of the concentrations we wanted to show a line chart of the change in the number of concentrators over time. Because these line charts were static and would show up only in the tooltip, we decided to create the SVG elements using Adobe Illustrator instead of d3. Illustrator has a chart tool that allows us to copy the data from our spreadsheet and paste it into a table which creates a chart. We then added a title and text explaining the maximum and minimum of the dataset and drew a line representing the average for the dataset. The chart could then be exported as an svg file and linked to in our web visualization.

Line Chart in Tooltip



Story

At first, we considered a more minimalist start page, so that we could more easily direct the user toward our storytelling components. We wanted a landing that would allow the user to choose either one of our three stories or the main visualization to look at. The three stories we decided to tell were those of Computer Science/Applied Math/Statistics, Economics, and the biological disciplines. We planned on using different visualizations for each of these stories.

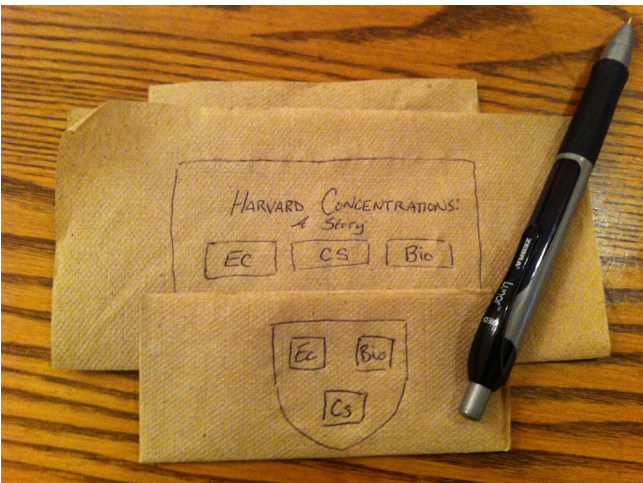
For Economics, we thought of using a bar graph almost identical in style to the one from our main visualization. Since Economics is always the largest concentration, we would put the first few concentration into our graph for context. We would start at 2000, and with each user click, the bar graph would update the year, and a short description of the events that occurred which caused the changes seen.

In our Computer Science/AM/Statistics story, we decided to use an area graph in order to show the general trend of these quantitative concentrations over the years. Like the other stories, the user's clicks guide him or her through explanations of events that affected concentration numbers in different years.

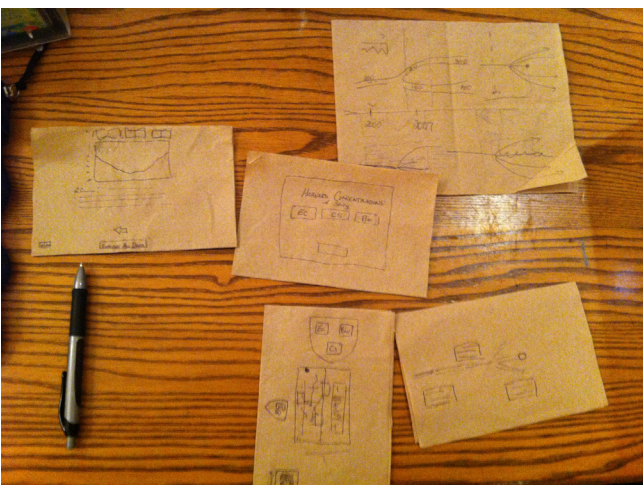
For the Biology story, we chose to use a sort of stream-graph in order to show the flow of concentrators over time as the concentrations changed. The graph allows us to illustrate when concentrations were created and eliminated, and how the total number of concentrators across the area changed.



Napkin Sketch 1



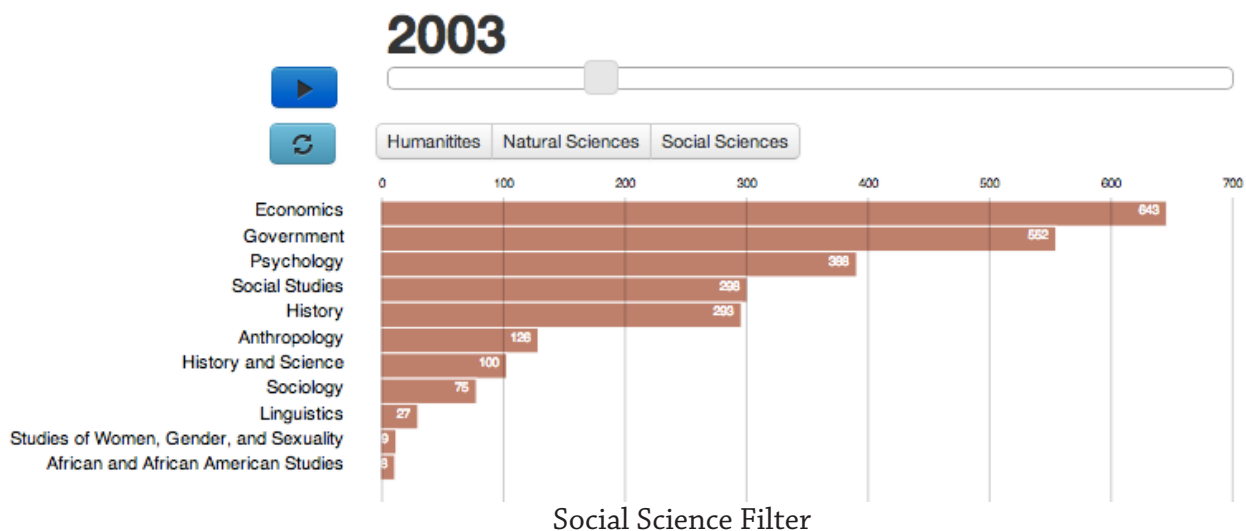
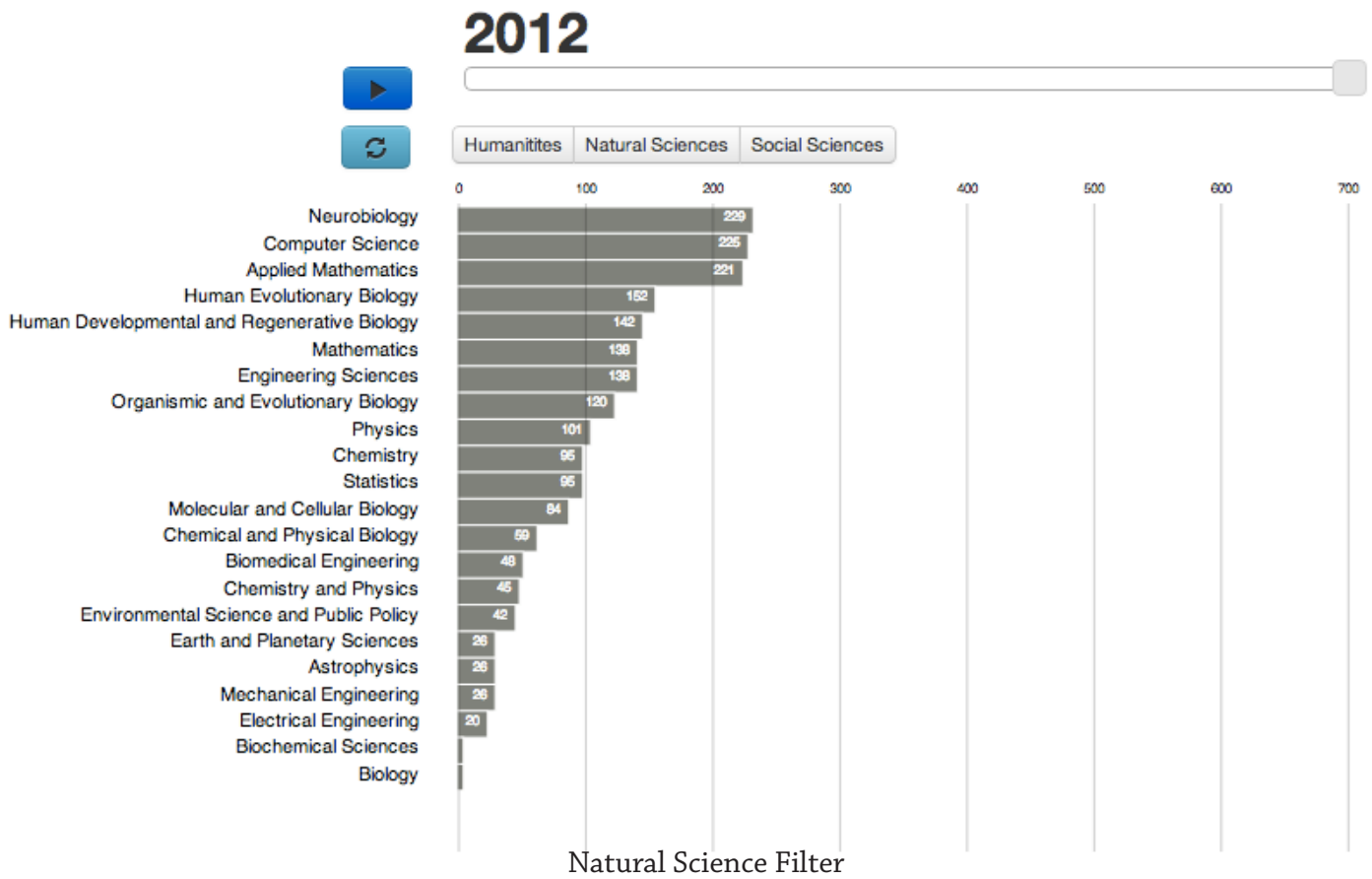
Napkin Sketch 2



Napkin Sketch 3

Filtering

Similar to what we did in project 2, we implemented a filtering function that filtered the bar charts based on the category of the concentrations. The three filters we implemented were Humanities, Natural Sciences, and Social Sciences. When the bar charts are filtered, they disappear and fade in. The quickest way to implement this feature was to remove the svg and then draw a new one based on a different category.

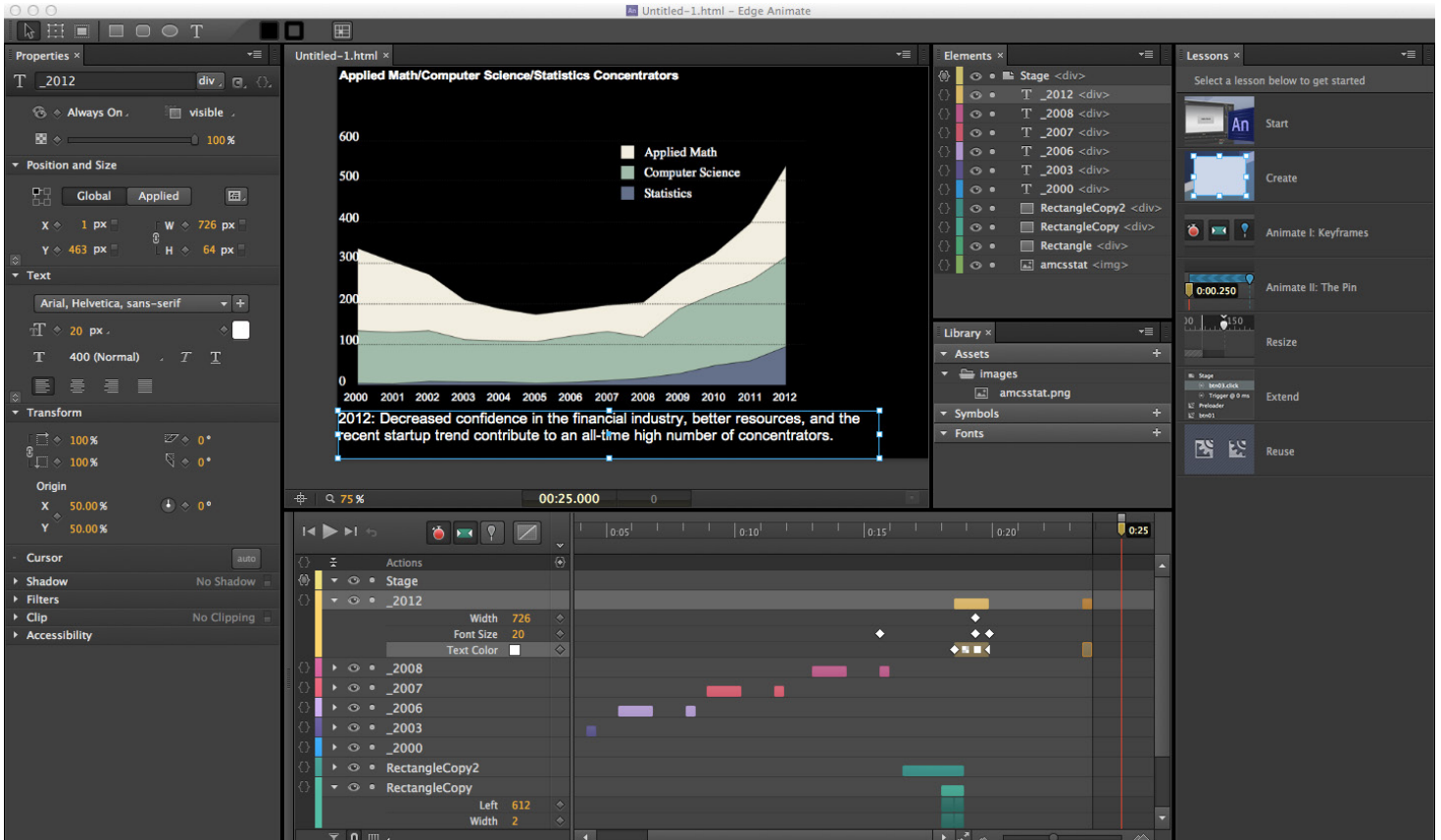


Animating a Story

To convey the stories within our data, we used Adobe Edge Animate to create an animated area chart with text annotations. The area chart was drawn using Adobe Illustrator. Similar to Adobe Flash, Edge allowed us to keyframe motions such as the unfolding of the area graph over time as well as the text fading in and out that could be used in an HTML5 website. Integrating the bar chart into our main visualization was a little harder and we ultimately decided to place the animation into a modal from Twitter Bootstrap. The video plays as soon as the user clicks on it and continues automatically until the end. We believed that animations can be more effective than interactive visualization in communicating a specific story.



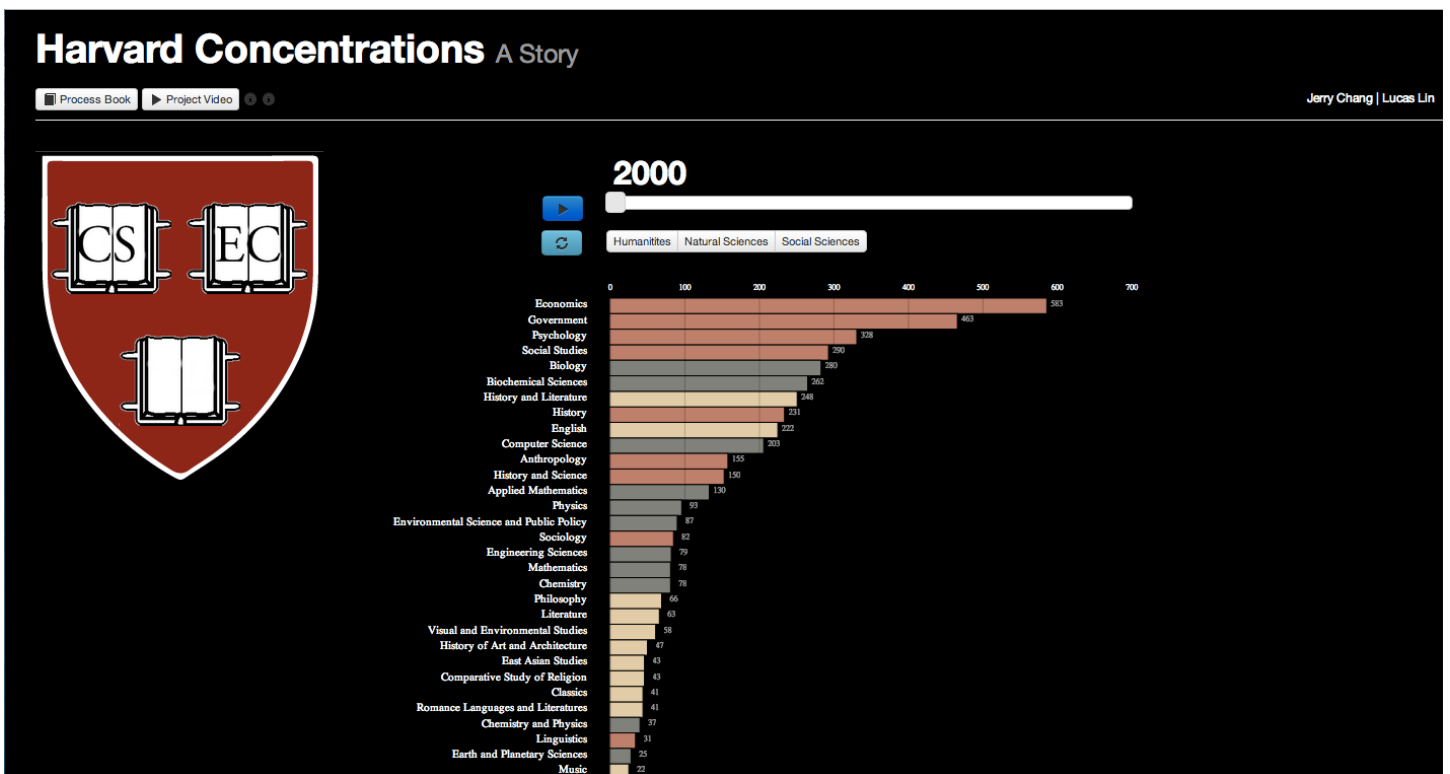
Animation in Modal View



Adobe Edge Animate

Final Visualization

Our final visualization is significantly more complex than our visualization from project 2. We've changed the design to feature a black background and reoriented the bars on the bar charts to be horizontal instead of vertical. We've also added line charts to the tooltip for each concentration and have improved the slider so that it will play automatically when a button is clicked. The color coding based on concentration was maintained and filtering based on category is also available. The final visualization also features two story telling designs based on the changes to the Economics and Computer Science concentrations over time. We think these additions significantly enhance what can be learned from our visualization.



Harvard Concentration Visualization

Sources

We used a variety of sources while making our project. Most notably, we used Twitter's open-source Bootstrap CSS framework. We also used some of the jQuery javascript library. Others are mentioned below.

Slider from <http://api.jqueryui.com/slider/> (Accessed April 22, 2013 1:57PM).

Some bar graph aspects from Hidenari Nozaki at <http://hdnrnz.me/2012/07/04/creating-a-bar-graph-using-d3js/> (Accessed April 20, 2013 11:56PM).

Tooltip from D3noob at <http://www.d3noob.org/2013/01/adding-tooltips-to-d3js-graph.html> (Accessed April 21, 2013 1:48PM).

More bar graph elements from Mike Bostock at <http://mbostock.github.io/d3/tutorial/bar-2.html> (Accessed April 22, 2013, 2:38PM).