Jobs and Skills in the Bay Area

Project Goals

For my final project, I decided to work with a dataset of jobs and skills data that was collected as part of my Master's Final Project. In our Master's Final Project application, called Pivot, users navigate by searching for job titles, skills, and industries to find and compare new career options. My overarching goal for this information visualization project was to design and build a tool that provides a more visual exploration of career options and skill needs. I decided on three ways that I wanted to visualize the data: (1) at a high-level, showing overall in-demand jobs and skills in the Bay Area; (2) at a job level; and (3) at a skill level. Specifically, the tasks that I wanted to support were:

1. Allow people to explore which jobs and skills are most in-demand in the Bay Area right now, so that they can find a job or skill that looks interesting to them
2. Allow people to filter jobs and skills by industry
3. Allow people to see, for a specific job, what skills it uses, and what are related jobs that they can explore
4. Allow people to see, for a specific skill, which jobs use it and what skills are commonly required along with it

Together, I thought that supporting these tasks would help a job-seeker, particularly someone who is looking for a new type of role, quickly understand which jobs and skills are most in-demand, and drill into their specific skillsets to see where else they could apply them and what other types of skills they should learn.

Related Work

In this section, I highlight three visualizations that also allow for exploration of job characteristics and skill needs.

LinkedIn Skills by Region Visualization

I wanted my visualization to allow people to explore in-demand skills for Bay Area jobs. LinkedIn made a skills-based visualization (*Figure 1*) that can be found at: http://blog.linkedin.com/2014/10/22/defining-a-city-by-its-professional-skill-set-with-data-from-linkedin/. This visualization shows a map of major cities in the US, colored with dots that represent one of twenty-one skill categories that most uniquely defines that city. Highlighting or clicking on a skill category in the legend highlights the cities whose jobs require that skill category more than most cities. You can then click on a city on the map to see a pop-up of the specific skills needed in that city within that skill category.

This visualization is related to mine in that both allow exploration of skills by industry. LinkedIn's is more focused on the regional aspect, and seems more about highlighting interesting trends that providing a tool for job seekers to figure out where they might relocate. I like the idea behind this visualization, and LinkedIn clearly has a lot of very interesting data on skill requirements, but the interactions are limited and it doesn't allow for a deeper-dive than just seeing the top 10 skills in one industry in a particular city.

Figure 1: LinkedIn Skills by Region Visualization

Payscale Job Meaning Visualization

Payscale.com has a scatter-plot based visualization (*Figure 2*) made with Tableau to show job median pay vs job meaning (the percentage of people that report that their job makes the world a better place). It can be found at http://www.payscale.com/data-packages/most-and-least-meaningful-jobs/interactive-chart.

There are 454 jobs featured in the visualization, and each is color-coded by one of twenty-two job categories/industries. You can filter the scatterplot in a number of ways, many of which I found to be confusing. For example, there is a filter for experience level (starting, mid-career, and senior), but it's unclear how this should affect the data, because the job titles are very general and would seem to span all of these seniority categories. The most useful interaction is being able to click on a job category to highlight jobs in that category.

There are also two bar plots below the scatterplot to show jobs with the highest and lowest meanings (

Figure 3), but unfortunately these do not update with any of the scatterplot features.

This visualization relates to mine through the subject manner, and also because both use industry as a color choice and filter for job titles. I thought that this was an interesting visualization idea but was poorly executed.

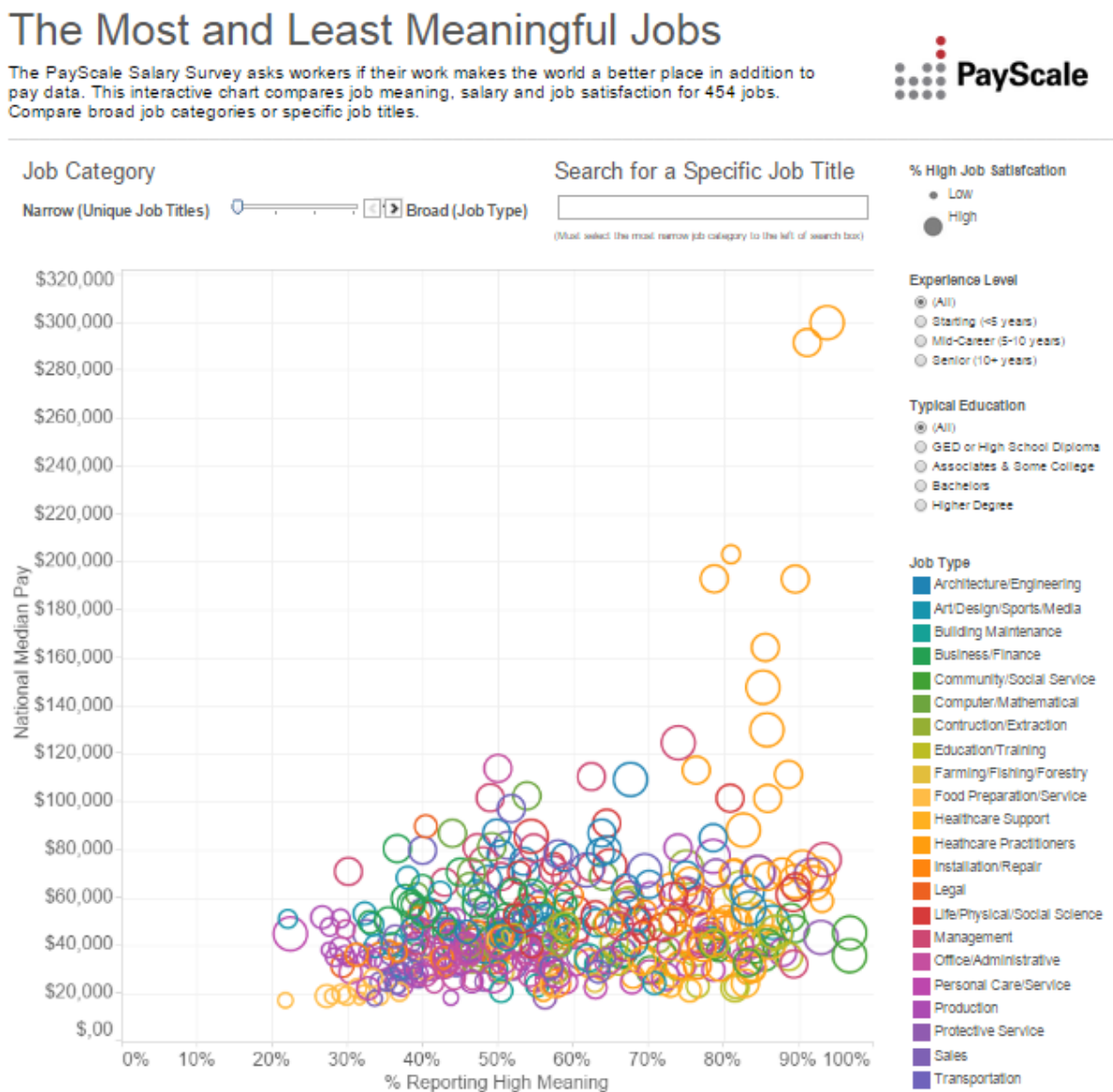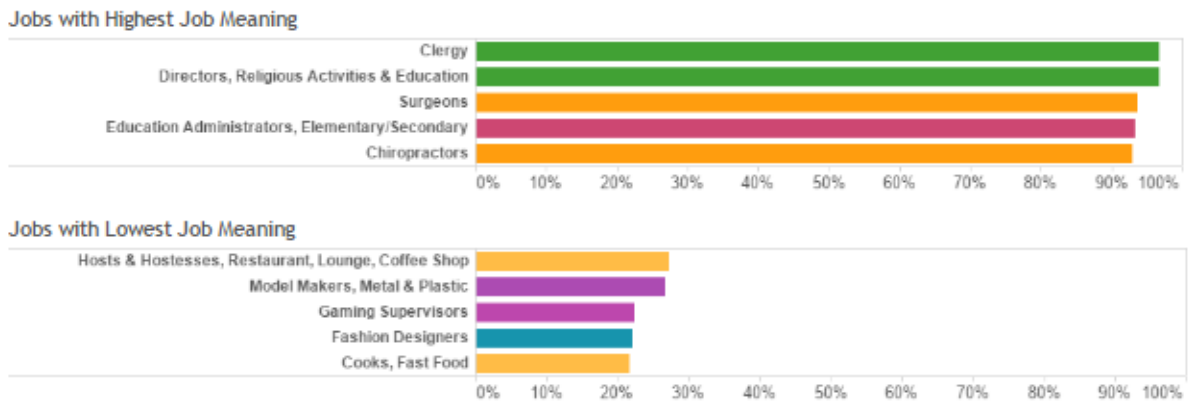Figure 2: Payscale Meaningful Jobs Visualization- Scatterplot

Figure 3: Payscale Meaningful Jobs Visualization- Bar Charts

Monster.com Career Path Tool

Monster.com has recently built a visualization to help people map out different career options. The entry-point to the visualization is a plot of nodes, organized and color-coded by industry and sized by the number of job opportunities (*Figure 4*). You can click on an industry to zoom into that job cluster. Then, you can click on a node on that cluster to see links to the most common jobs that people transition to. There is a side-bar that updates with information about the growth rate of the job that you are looking at.

Figure 4: Monster.com Career Path Tool
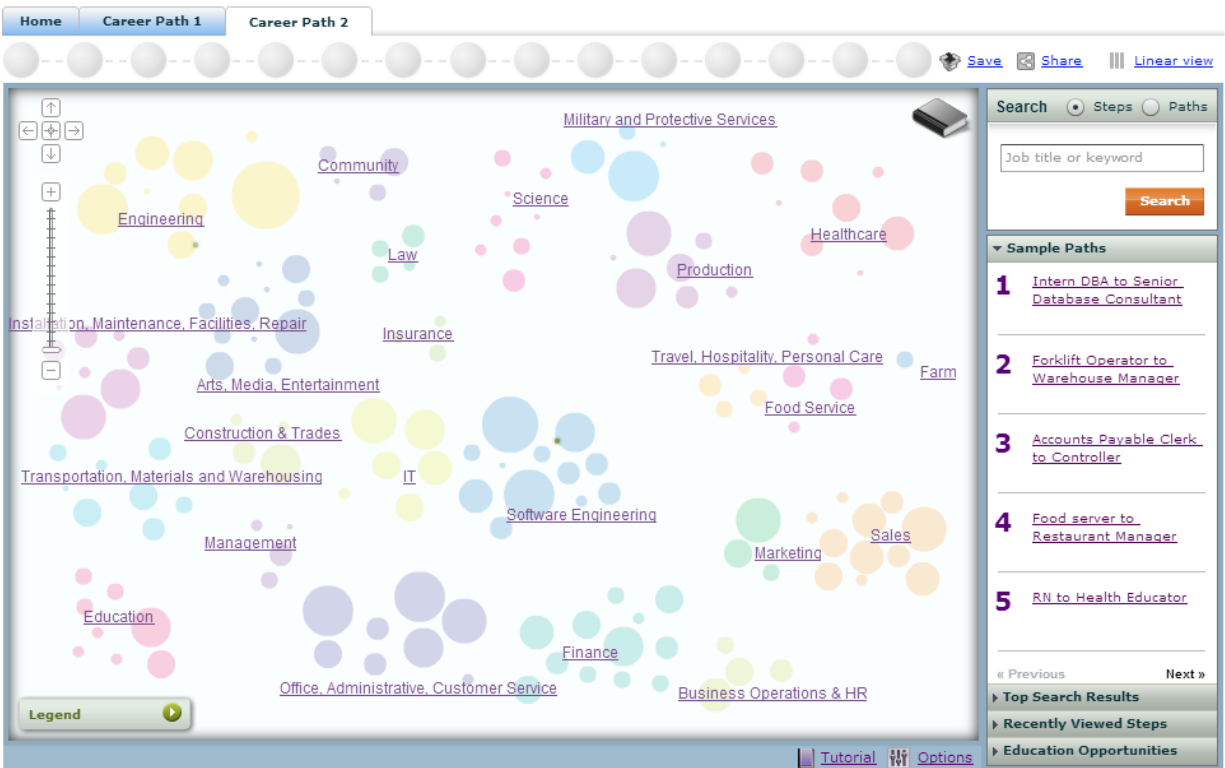
I tried using this tool for my own career path, by navigating through engineering, civil/structural/environmental engineers, then clicking on the civil engineer node (*Figure 5*). It suggested structural engineer and draftsman as common career paths. I was able to click to add Civil Engineer to my path. I didn't quite know what to do next so I searched for data scientist, but this wasn't an option. I experimented with adding web applications developer to my path, but I didn't see any updates with information on how I could transition from Civil Engineer to Web Developer, and how many people have made such transitions.

This visualization seems like a really good idea if it could provide some insight into which skills are transferrable in careers along your path. It related to my visualization most directly through my feature that shows related jobs for a particular job. Mine, however, is not based on path information, but rather on overlapping skills and industries.

Figure 5: Monster.com Career Path Tool, Civil Engineer Node
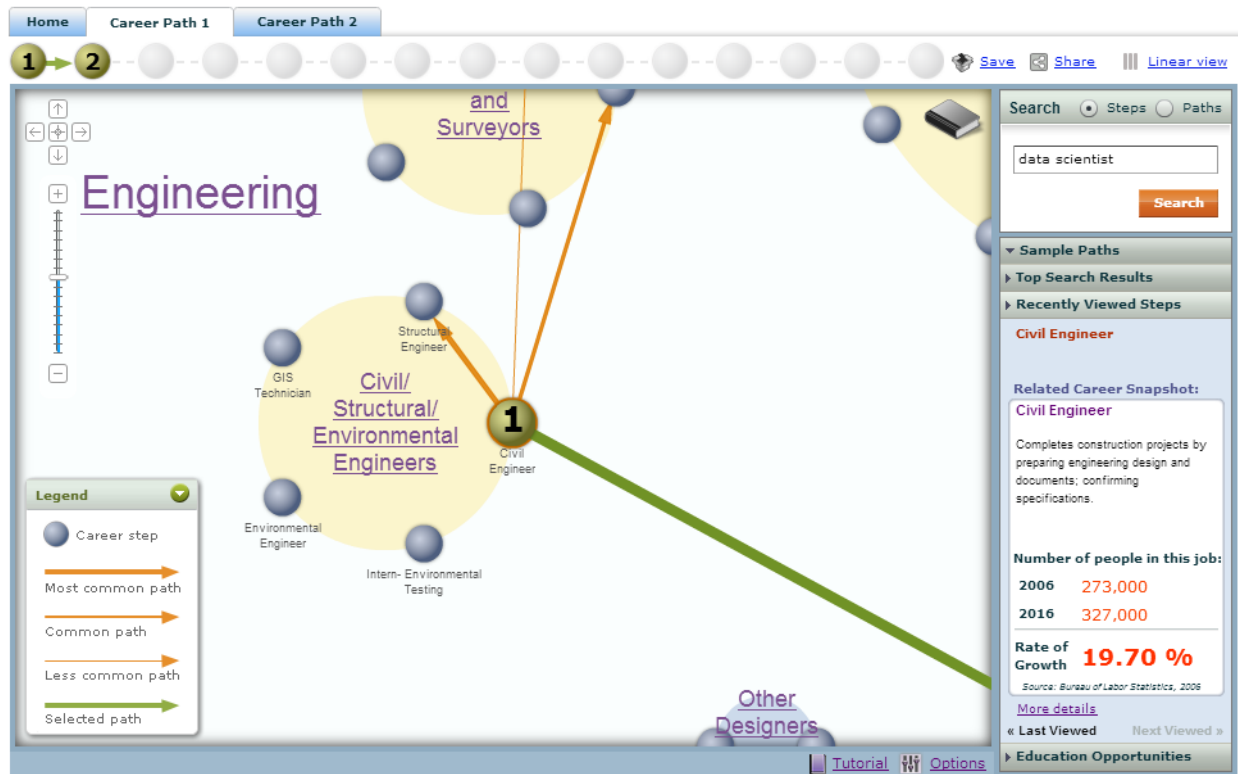
Visualization Description

My visualization consists of three web pages: (1) an overview/homepage; (2) a job page; and (3) a skill page.

Overview Page

The home page of the visualization features two stacked bar charts. The one on the bottom-left of the page shows the top 50 in-demand jobs in the Bay Area, ranked by the number of job postings that I collected. The bar chart on the bottom-right of the page shows the top 50 in-demand skills in the Bay Area, ranked by the number of job postings that mention the skills. While my dataset contains over 200 jobs types and 2,000 skills, I only show the top 50 to make the page load quickly so that the user doesn't get frustrated.

When first arriving on the page, the charts are stacked bar-charts, colored by job industry, to show the proportion of job openings and skill requirements by industry (Figure 6). While stacked bar charts make it difficult to do exact quantitative comparisons, I designed the visualization in this way to quickly show people which skills and jobs are only applicable to a certain industry, versus those that offer opportunities in a lot of different industry. The colors in the plots map to the colors in the industry icons that I placed at the top of the page. The industry icons are clickable so that you can filter the top jobs and skills to only those in that industry (for example,

Figure 7 shows results filtered for Sales and Marketing). When you click on an industry, the bar charts update to only that industry color, and the labels of the charts change to reflect that the user is now looking at jobs and skills in that industry. The user can click on another industry to remake the plot, or reset to see all industries again.

On the bar plots, the axis labels are clickable to take the user to a visualization that provides further detail on that job or skill. In designing how the user filters the overall results, my hope was that people would be able to find job title and/or skills that they were interested in learning more about so that they could click over to the job or skill page.

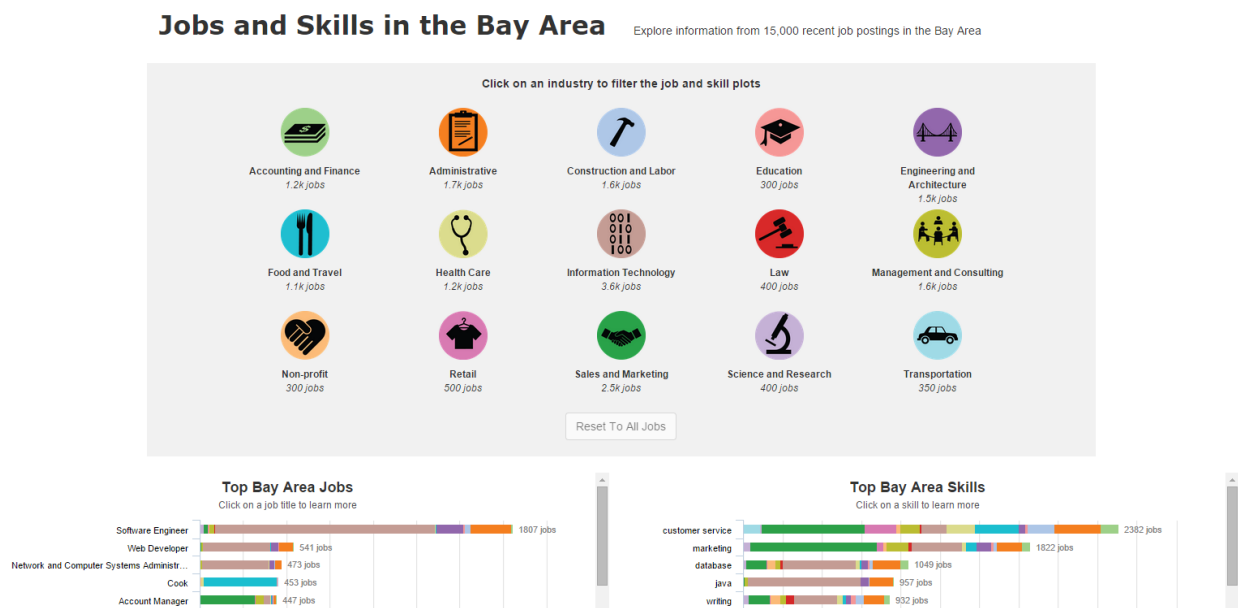Figure 6: Overview Page, Top Bay Area Jobs and Skills

Figure 7: Top Bay Area Jobs and Skills in Sales and Marketing

Job Page

Clicking on a job title takes the user to a visualization for that job. I made three charts for this visualization: a top industries bar chart, a top skills bar chart, and a related jobs radial plot. All are shown for the Product Manager job in *Figure 8*.
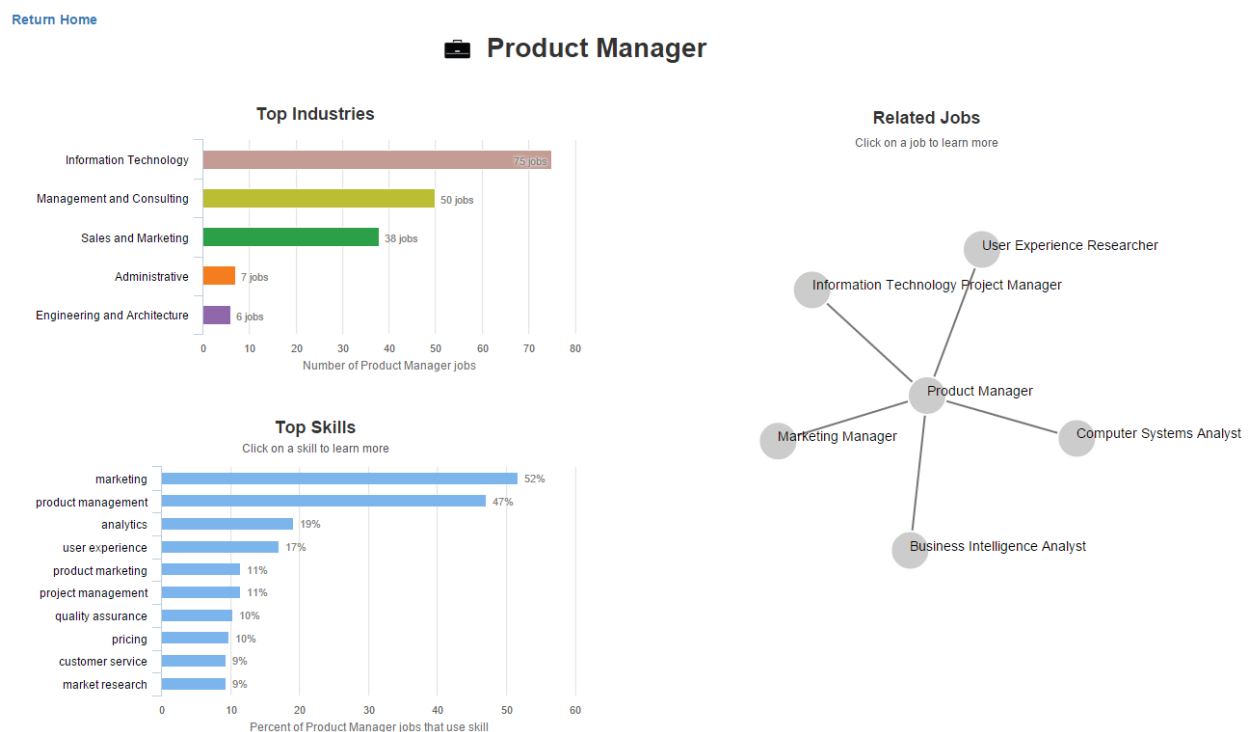
The Top Industries chart shows the top five industries that have openings for this job. The colors of this bar chart correspond with the industry colors on the homepage. My intention for this diagram was to show people whether most of the opportunities for this job are concentrated in a single industry, or whether there is flexibility to work across different industries.

The Top Skills chart shows the top 10 skills that are required for this jobs. Rather than quantifying this value in job openings, I quantified it as the percentage of job postings that mention this skill, to show users the likelihood that they would get to/have to use that skill for that job. I struggled with deciding how to color this plot. I did not want to do a stacked bar chart, but a particular skill can cut-across industries. I decided on a neutral color that was not used for any of the industries. In future work, I would consider coloring them by the dominant industry. The skills are clickable and take the user to a visualization for that skill.

The Related Jobs plot shows the current job in the middle, with links to the five closest 'related jobs'. Related jobs were calculated from skills and industry overlap. Clicking on a node in the graph reloads the page with the visualization for that job. My intention with this design was to provide users with an easy way to navigate from the current visualization to the same visualization for other jobs that they would likely be interested in.

Figure 8: Job Page, Product Manager

Skill Page

Visually, the skill page is the same as the job page. There is a Top Industries chart and a Top Jobs chart. All are shown for the Analytics skill in *Figure 9*. The Top Jobs chart shows the top 10 jobs that use this skill, ranked by the percentage of postings that mentioned the skill. Choosing to rank by the percent of job postings for that job that mention the skill rather than the number of postings for that job that mention the skill makes a big different in the results that the user sees. In the analytics example that I show below,

Search Marketing Strategist is the top job, even though there were less than 10 openings for this job. I chose to do percent because I wanted to show people who are really interested in a particular skill the jobs in which they have the best chance of being able to apply it. In future work, I'd like to add a toggle button to choose between ranking by percent and ranking by total number of postings.

Figure 9: Skills Page, Analytics
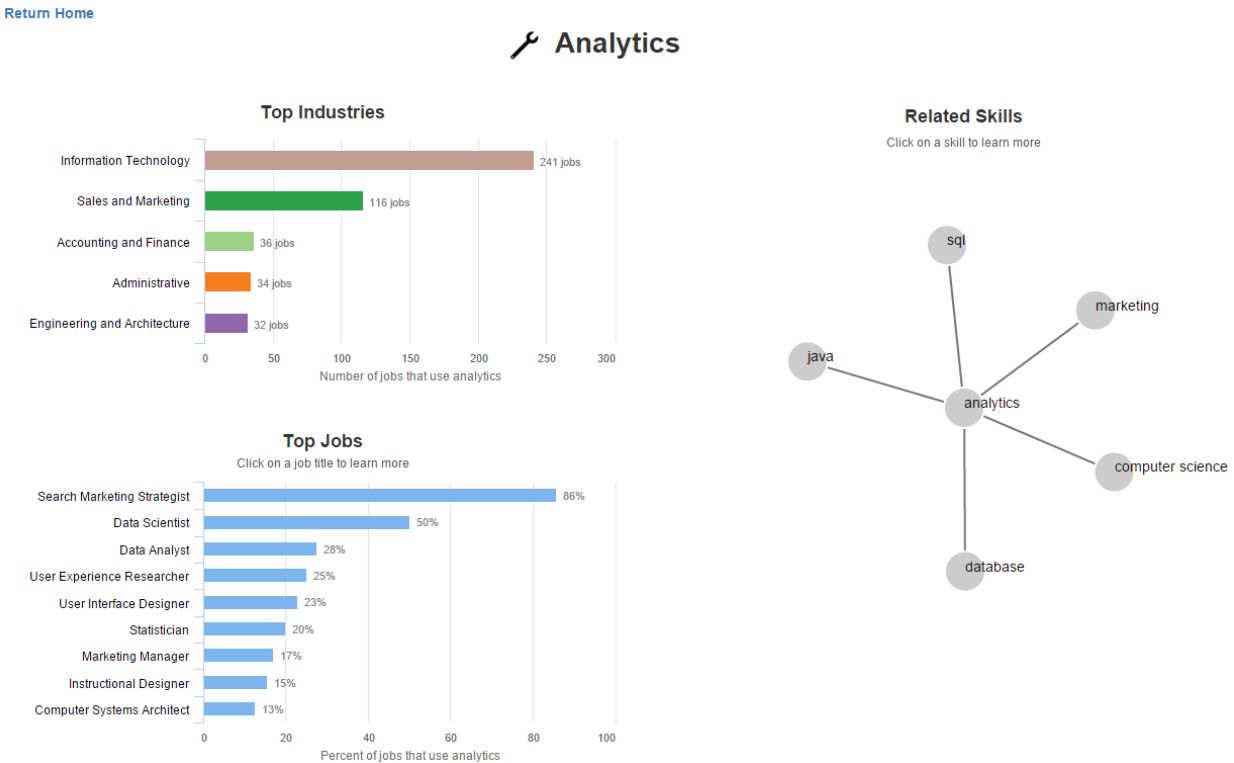
Data

The data used in my project was gathered from online job postings obtained from the CareerBuilder.com API, the SmartRecruiters API, and Craigslist via web scraping. The visualization was based on data collected from 15,000 job postings. My visualization features industries, job titles, and skills. Each collected job posting has a single job title, a single industry, and one or more skills mentioned in the job posting. A brief description of how this data was pulled together is listed below. Since most of this work was done as part of my Master's Final Project, more details can be found in that Final Report.

1. **Industries** were obtained from the APIs and from the Craigslist ad tags. The different data sources combined had over 100 different industry category names. For this project, I mapped them to 15 standardized industry categories to achieve what I considered an appropriate granularity for filtering results. The Careerbuilder data allowed a single job to be assigned multiple categories. To simplify this for the purposes of my visualization, I assigned each of these jobs a single industry category by choosing the category that was the most common for that job title.

2. **Job titles** were obtained from APIs or extracted from the raw text. A title standardization process mapped specific company titles to general job titles (e.g. Java Engineer III to Software Engineer). This visualization features 203 different job titles.
3. **Skills** were obtained from APIs or extracted from the raw text. There were more than 1,000 different skills featured in this visualization. The skills range from general knowledge areas (such as marketing) to specific tools/software packages (like Microsoft Powerpoint). Some of the skills are subsets of others (for example, Python and Java are subsets of Programming). It would be nice to impose and visualize a hierarchy of the skills, but that was not done in this project

Tools

I used the following tools to design and develop my visualization.

- Python, mostly using the Pandas and Numpy libraries, for data processing, ultimately outputting data into JSON formats to be used in the visualization
- Powerpoint and Tableau for prototyping design and interactions
- Illustrator to make the industry, skill, and job icons (which were obtained from the Noun Project)
- Lodash library for javascript data manipulation
- Jquery for website interactions
- Bootstrap for website layout
- Highcharts for the bar charts
- d3 for the force-directed radial graph

Steps

This section describes my project steps, which were: brainstorming, prototyping and iteration, data preparation, implementation, and testing.

Brainstorming

I spent a lot of time trying to conceptualize the best way to use my jobs data set in a visualization. The first decision to make was whether I wanted to tell a narrative story with my visualization, or whether I wanted to make something exploratory. While I thought the narrative story would be easier to accomplish as a team of 1, given the current lack of interesting job and skill visualizations, I thought that an exploratory visualization would be more useful.

I struggled to come up with creative ideas for visualizing this data set, largely because it lacks a lot of the features of datasets that I am more accustomed to visualizing. For example, there is no temporal dimension to it, so it's not possible to explore trends over time. Additionally, it contains no numeric data other than counts.

I thought of a number of design alternatives, sketching them out and showing them to my class project partner and my Master's Final Project partners to get feedback. My initial pitch was to do a chord diagram that shows skills co-occurrences for a particular job title. I got good reactions to this idea, but there were some concerns about its usability, and I could not figure out an easy way for people to navigate to a job title that they wanted to investigate. Another idea that I considered was designing and implementing a 'skill builder' visualization, where people could drag and drop skills into a box to see charts on which jobs and industries use that particular combination of skills. I think that this could have turned out well, but

with over 1,000 different skills, I was struggling with how to provide users with access to all the possible skills to create their customized skill set.

In the end, I decided that I could make a compelling visualization using relatively simple charts by providing people easy ways to drill down into their industries, job titles, and skills of interest to explore opportunities in a visual way.

Prototyping and Iteration

Once I had settled on a concept, I began prototyping it using Tableau for the bar charts, and powerpoint to mock-up anything that I wanted to do using d3. There were a number of things that changed from my beginning design to my final design, mostly involving how to interact with the industry filters. I wanted the industries to serve as filters for the plots below, but I also wanted to somehow convey the quantity of job opportunities in an industry. My initial idea was to use bubble charts for the industries, sized by the number of open positions (shown in *Figure 10* and *Figure 11*). I got feedback that this was difficult to use, because the arrangement of the bubbles was random, and the industry names were not visible in the smaller bubbles. I also got feedback that there were too many industry categories, so I grouped similar categories together (for example, the Sales category and the Marketing category got combined to Sales and Marketing).

I sketched out a number of alternatives to the bubble charts. I tried using bubbles, but arranged in a line alphabetically, but I was worried about visually misrepresenting job availability because we don't perceive area accurately. I also tried using icons with isotypes below to indicate the number of open positions; people found this to be too cluttered. In the end, on the homepage, I just use icons with the industry name and total number of positions listed below the icon. People liked the icons and selected colors for representing industry. On the job and skill pages, I represent industry opportunities using a bar plot, though I received a good suggestion to also use the icons on these page.

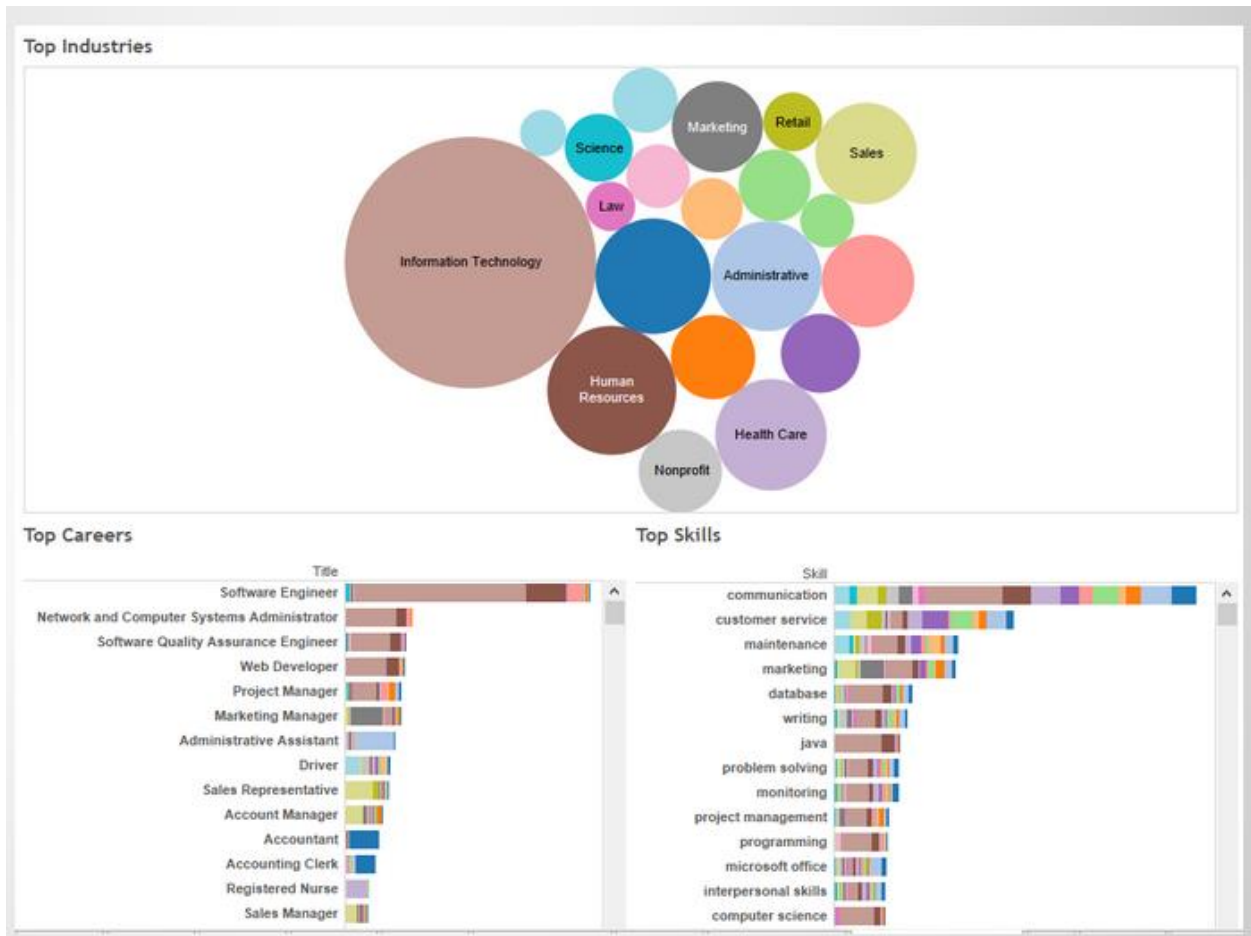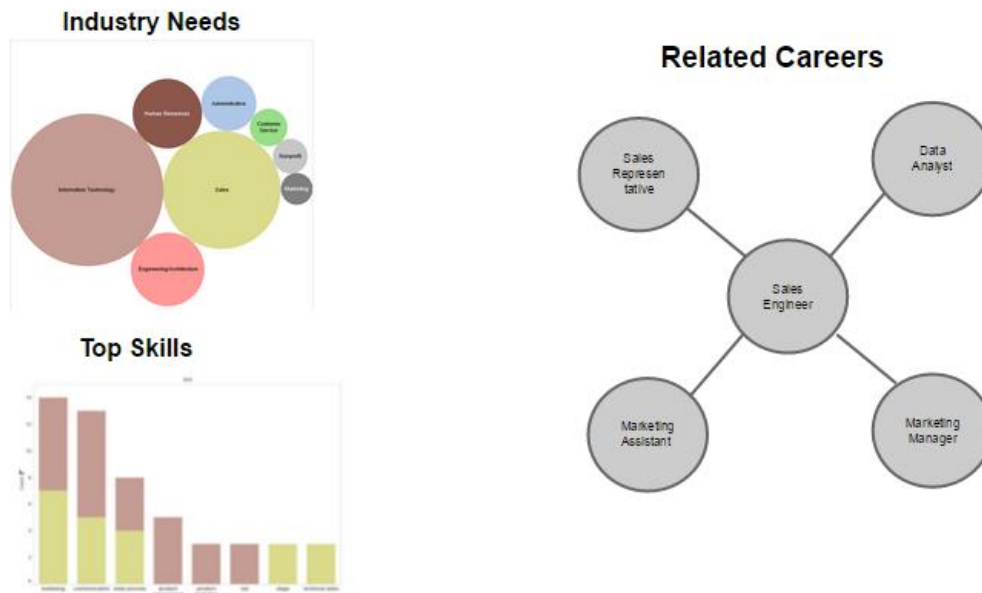Figure 10: Initial Prototype of Overview Page with Bubble Charts

Figure 11: Initial Prototype of Overview Page with Bubble Charts and Stacked Bar Charts

Data Preparation

I read up on the lodash javascript library functions, and used Python to create json files that would be easy to manipulate in javascript. In total, my application uses 6 json files. I used 6 separate files rather than one large file to decrease page loading times.

Implementation

In the implementation, I created three HTML pages, using Bootstrap for its grid layout and class styles. I had initially planned to code everything in d3, but switched to HighCharts for all of the bar charts after being unable to implement a simple barchart in d3. I did use d3 to make the related skills/related jobs plot, imposing a radial style to make sure that people know that the current job is in the center. I used jquery for the webpage interactions.

The implementation step was challenging because it was a lot of work for one person, and I don't have much experience with making and styling web pages. I definitely learned a lot by doing all of these steps, and am glad I had the opportunity to do them before I graduate.

Testing and Refinement

It took me awhile to build out all three of the web pages such that they interacted with one another, so a lot of my testing was done on individual pages, particularly the overview page. Most people understood how the filters worked, but suggested that I make the interaction more obvious by including text within the icon box explaining how to use the icons. In the final design, users can filter down to a single industry to view top jobs and top skills, or can view this data across all industries. I asked two users whether they wanted the functionality to be able to view data for a custom set of industries (for example, Information Technology and Health Care), but they were not that interested in that functionality, and it would have been a lot of work to make this kind of interaction obvious.

When I tested the job and skill pages, I initially had the Top Jobs and Top Skills as stacked bar charts colored by industry, just like on the homepage. In this context, users found it confusing, so I made these regular bar charts with a single, neutral color.

I also got a lot of great feedback from people at the project showcase, which was the first place I was able to test the whole application working together. Unfortunately, I have not had time to incorporate most of the suggestions, but would like to in the future. The following list summarizes my results.

1. *On the homepage, people reacted well to the industry icons, colors, and filtering abilities*. For the most part, people were able to figure out how everything worked without me explaining it, and enjoyed clicking around to see results for different industries. I got good feedback from two people to make it more obvious which industry you are clicking on, and I would like to implement that in the future.
2. *People didn't immediately realize that you could click on a job title or skill to navigate to a different page*. I have subtitles on these plots instructing people to click on the axis labels to learn more, but I need to think of how to make this interaction more obvious.
3. *People tried clicking on bars in the bar charts, which doesn't do anything*. I should think more about what types of interactions this could/should support.
4. *On the jobs and skill pages, people really liked the d3 radial plot*. I was wondering if people would find the animation interesting or distracting, but people really liked it, possibly because it's a nice contrast to all of the static bar plots. Everyone tried clicking on the nodes in these graphs, and liked how they updated the graph and the rest of the page to be centered on that node. The success of this feature suggests that people really like seeing relationships between different jobs and skills. The colors of this graph are currently gray, but Marti had a great suggestion to consider coloring these by the dominant industry, possibly using the icons on the overview page.
5. *I didn't see confusion in people regarding whether they were on the job or the skill page*. The job and skill pages have the same design, so I added a briefcase in the header of the jobs page and a wrench in the header of the skills page to help orient users. I did not see anyone experience confusion over which page they were on, but I could use color to further distinguish between the two pages.

I would like to do further testing among people who are currently looking for jobs to see whether and how they find value in this visualization.

Links

My visualization is at http://people.ischool.berkeley.edu/~tbarkley/infoviz/. My Master's Final Project, which uses the same data, is at pivot.meteor.com.

Work Split

As a team of one, I did all of the work described here. I performed the data collection and extraction of skills, titles, and industries as part of my Master's Final Project. The remainder of the work was done by me for this class, although I'd like to eventually incorporate this visualization into our Master's Final Project application.