840 Jobs Process Book

Basic Info

- Title: 840 Jobs
- Email: Matthew.Schroeder@utah.edu
- UID: u0202874
- Project Repository: https://github.com/MatthewSchroeder/840Jobs

Background and Motivation

As a labor market economist for the state I track occupational information and forecast occupational performance to help jobseekers and employers. In particular, I have a desire to help students make informed decisions about their future careers. In my experience I have found that most students have a very limited understanding of the numerous career opportunities that even exist, much less the relative wages they might earn and whether they will even be able to find a job.

My goal is to create a visualization that will encourage students and other jobseekers to explore and learn about the occupational possibilities that exist and then help them compare those options based on data and ultimately make a more informed decision.

Project Objectives

The primary questions I am trying to answer with this visualization are these :
- What occupations are out there that I may be interested in?
- How well does each occupation pay in comparison to others?
- What is the demand for the occupations I'm interested in?
- What are my prospects for career growth in a given occupation?
- Where would I be most likely to be able to find work?
- What industries would be most likely to hire me?
- What are some similar occupations that I might be interested in?
- What knowledge, skills, and abilities do I need to be successful?

Aside from the questions I hope to help jobseekers answer, I would like to learn something from this project as well. I work with occupational data for Utah regularly, but I will be doing this project using national data as well as data from the other fifty states. I hope to gain insights into how occupations are demanded differently across geographies and, as a result, better characterize Utah's local labor force in comparison to other regions.

If I can help even one student make a better choice for their future based on what they learn by using this tool then I will be thoroughly satisfied. Among the benefits of having a data-based understanding of occupations before choosing a career path are:

- Greater job satisfaction
- Higher wages
- Easier time finding a job
- More job security
- Greater potential for career growth and advancement
- Cost savings by avoiding re-training after an initial poor choice
- Cost savings by avoiding low-return educational investments

Related Work and Inspiration

In my work at the Utah Department of Workforce Services, I created aTableau visualization for comparing occupations. Broadly speaking, the visualization is useful, but Tableau is very rigid and I was unable to implement more effective methods of encoding of the data. My desire to improve this tool (and expand it to include the other 49 states), is why I chose this subject. Below is an image of the Tableau visualization (http://jobs.utah.gov/wi/uocd/dashboard.html):
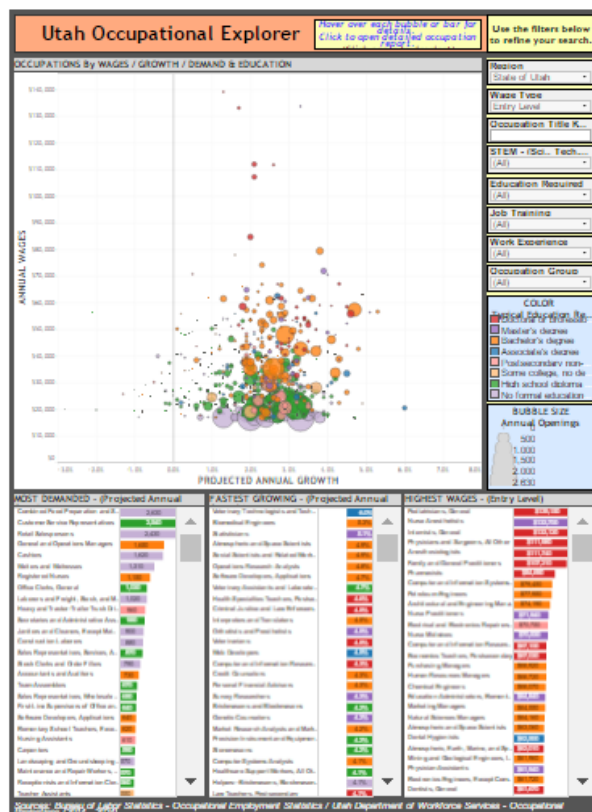
I wanted to improve upon the messiness of the scatterplot and make it more conducive to exploration. I also wanted to make is possible users to cluster jobs by different categories and then show transitions to the new cluster so that the relationships between say, education and higher paying jobs becomes more apparent.

I also wanted the initial view to be catchy and encourage users to explore and I found inspiration in one of Mike Bostock's examples – the Force Cluster Layout lll (https://bl.ocks.org/mbostock/7881887).
I used some of his code along with some from Shan Carter who built upon this idea in D3 4.0 (https://bl.ocks.org/shancarter/f621ac5d93498aa1223d8d20e5d3a0f4). I've included images of these two below:

Clustered Force Layout III

Clustered Force Layout 4.0
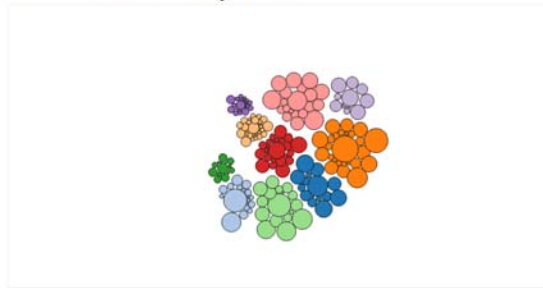
Data Sources

1) Occupational wage data and descriptions are available for the nation, by state, and by industry from the Bureau of Labor Statistics (BLS) – Occupational Employment Statistics (OES)program: http://www.bls.gov/oes/

2) Occupational projections for the nation and by industry are available from the BLS – Employment Projections program: http://www.bls.gov/emp/

3) Occupational projections by state are available from Projections Central (BLS in partnership with each state): http://www.projectionscentral.com/Home/Index

4) Occupational knowledge, skills, and abilities (KSAs) data are available from O*Net (Department of Labor and Employment & Training Administration): http://www.onetcenter.org/database.html?p=2

Data Processing

I did not expect there to be too much data prep work, but I should have known better. Combining all of these data sources resulted in a considerable amount of linking and re-structuring. The final dataset is a combination of data sources 1, 3, and 4 and includes a unique record for each state/occupation pair. The work was done in Excel and required a lot of index(match()) functions and pivoting, etc. The final file has nearly 40,000 rows and about 150 columns. I save the final dataset as a .csv.

Visualization Design

I had previously planned to do two separate top-level version of this visualization – one from a geographic (state) perspective, and one from an industry perspective. Given the amount to data prep required, and based on feed-back from an in-class session, I decided to let go of the industry version. Given my target audience of jobseekers and students, the most relevant information is related to geography. The industry version would be very useful for other users, such as businesses, so I may come back to it in the future, but not for this project.

The general idea for the design has four views:

View 1 – Occupational Cluster Chart
For the first view is a force-layout cluster chart of occupations that can be clustered by different categories. The intent of this chart is to hook the viewer and encourage interaction and exploration, so the interpretation of the visual encoding does not need to be precise. I mostly want it to give viewers an idea of how many jobs there are, and how they are related (i.e. through their clusters). The force-layout cluster charts are fun to play with, and I think it will encourage viewers to engage and start exploring.

The user can choose how to cluster the occupations by using a drop-down menu on the right. Cluster options will include major occupation group (default), education required, STEM/Non-STEM, and possible a couple others. Each cluster will be labeled and each bubble inside will represent an occupation. Each bubble will have a tool-tip on-mouseover with the occupational title and some general information (wages, demand, and growth). The size of the bubbles will encode demand (i.e. the number of projected openings). The color of the bubbles will encode median wage (single color – more saturation = higher wage). I would like to encode the relative wages within each cluster along the vertical axis, so bubbles with higher median wages will float toward the top.

I considered using blocks of small squares to encode the clusters, but decided to go with the bubbles because I can use bubble size to visualize relative demand. I also realized that I might be able to encode the aggregate projected growth of each cluster along the horizontal axis, so the clusters would be in a horizontal row – although I'm not sure I want to encode any positional behavior at the cluster level for fear it may bias a user's choices and they may miss a high-growth occupation hidden within a low-growth cluster.

When the user selects different cluster groupings from the drop-down menu I would like there to be nice transitions so they can follow each occupation to its new cluster. Enlightening insights will be gained by this (i.e. watching the STEM occupations move mostly to the higher education clusters).

I might also like to have the visualization zoom in on an individual cluster on-click, with an additional encoding for growth on the horizontal axis. This will reduce the number of number of bubbles and put them in an X, Y plot so that the viewer can now really see the growth dimension… but I worry that this may be too much info to effectively absorb. Perhaps I will leave the growth encoding to the following views.

View 2 – Ranking Chart

I'm still not sure exactly the best way to encode this view. The goal of this view is to rank and compare all of the occupations across the three primary measures of interest – wages, demand, and growth. I considered several options.

First I considered a simple set to three vertical bar charts, each sorted in descending order of rank, and the value of the dimension encoded in the bar length. The problem with this is that there may be up to 840 occupations in each which is far too many to fit in a single view effectively.

The second idea I considered was to use a parallel coordinates chart. This chart would do a much better job of fitting the high number of occupations into a single view, but it can look cluttered and may be hard to interpret.

Finally, I considered using a distribution chart for each dimension that shows where the highlighted occupation shows up in relation to the median and in relation to the rest of the density function. My concern with this one is that it's overly statistical and may put some viewers off, but I'm thinking if I can label it well and create good tool-tips, it may be the most effective option.

Ultimately I just want viewers to be able to tell how each occupation stacks up against the others across all three measures, and I think the density function chart may be the simplest way to summarize so much information. Additionally, it might be cool if I could implement a brush on each density function that would highlight the occupations that fall within it on the Cluster Chart.

This view should filter to show only the occupations within a cluster on-click of the Cluster Chart. It should also highlight across all three dimensions on-mouseover (connected to on-mouseover of the Cluster Chart also).

View 3 – State / Industry Chart

The intention of this view is to answer one of two questions. If the viewer chose to compare across states at the beginning, then this view will answer which states have the highest demand/growth for a selected occupation. If the viewer chose to compare across industries, then this view will answer which industries have the highest demand/growth for a selected occupation.

I chose to use a simple horizontal bar chart for this view. It will be in descending order of projected openings which will be encoded in the bar height and will have the projected growth of occupational share encoded with color. I considered using colored blocks for this view (in the shape of a map for the states), but in that case I would have to use labels to encode the projected openings, and I would lose the ability to sort. I think the bar chart will do a much better job of quickly conveying the answers to these questions.

View 4 – Individual Occupation Info Panel
This view is intended to provide additional information and details about a selected occupation. It will be primarily text, and will potentially have some simple graphics associated with the primary measures.

Primary information will be the occupational title, a description, and the values of the primary measures. Additional information will include similar occupations (which the user can click on to update the viz), and the KSAs needed for the selected occupation.

Must-Have Features
- Cluster Chart
  - Cluster labels
  - Dynamic cluster transitions
  - Higher wages float to top / redundantly encoded with color
  - Tool-tips on-mouseover
  - Highlight Rank Chart on-mouseover
  - Filter Rank Chart on click
  - Update State/Industry Chart on-click
- Ranking Chart
  - Summarizes relative performance of selected occupation
  - Connected to cluster chart (filter and highlight)
  - Highlights across all three measures on-click
  - Tool-tip on mouse-over
  - Labels present for selected occupation on each measure
  - Smooth transitions
- State / Industry Chart connected to cluster chart
  - Bars sorted by demand
  - Connected to Cluster Chart and Ranking Chart (on-click)

- o   Tool-tip on-mouseover
- o   Smooth transitions
- Individual Occupation Info Panel
  - o   Contains all the relevant info

Optional Features.

List the features which you consider to be nice to have, but not critical.

- Cluster Chart
  - o   Zoom on clusters
- Ranking Chart
  - o   Density function brush filter
- State / Industry Chart connected to cluster chart
  - o   Bars sorted by demand

Project Schedule

Week ending:

- Oct 29th – Complete data collection, structuring, linking, and derivations. Finalize source data files.
- Nov 5th – Prototype of views 1 and 2 up and running.
- Nov 12th – Prototype of views 3 and 4 up and running.
- Nov 19th – Adjust and add additional features.
- Nov 26th – Finalize visualization. Create website and screencast.

840 Jobs

Name: Matt Schroeder; E-Mail: matthew.schroeder@utah.edu; UID: 0202874

Name: Matt Schroeder; E-Mail: matthew.schroeder@utah.edu; UID: 0202874

Name: Matt Schroeder; E-Mail: matthew.schroeder@utah.edu; UID: 0202874