840 Jobs Process Book

By Matt Schroeder
12/2/ 2016
URL: https://matthewschroeder.github.io/840Jobs/840Jobs.html
Email: Matthew.Schroeder@utah.edu
Project Repository: https://github.com/MatthewSchroeder/840Jobs

Overview

840 Jobs is a data visualization designed to help students and job-seekers make more informed decisions about their career paths. A wealth of labor market information specifically related to occupations is produced by the government on a regular basis, but it is often difficult to access and understand. Making this information available and easily accessible to the people it most benefits is the ultimate goal of this project. This visualization was produced as the final project for a data visualization course at the University of Utah.

Background and Motivation

As a labor market economist for the state of Utah 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 was 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 more informed decisions.

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 are some similar occupations that I might be interested in?

Aside from the questions for jobseekers, 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 a Tableau 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 it 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 required 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. The final source file is saved the final dataset as a .csv.

Design Evolution and Implementation

I had originally planned to do two versions of this visualization depending on the user's perspective – one from a job-seeker centric perspective, and one from an industry perspective. After discussing these ideas in an in-class -feedback session, I realized that the industry based version was going to distract from my primary objective of helping job seekers, so I decided to let go of the industry version. 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 initial concept looked like the picture below. A cluster chart for exploring and initial engagement, a set of distributional charts or histograms to highlight where each occupation ranked within the distribution, a bar chart to encode demand and growth by state or industry and a detailed information panel at the bottom.
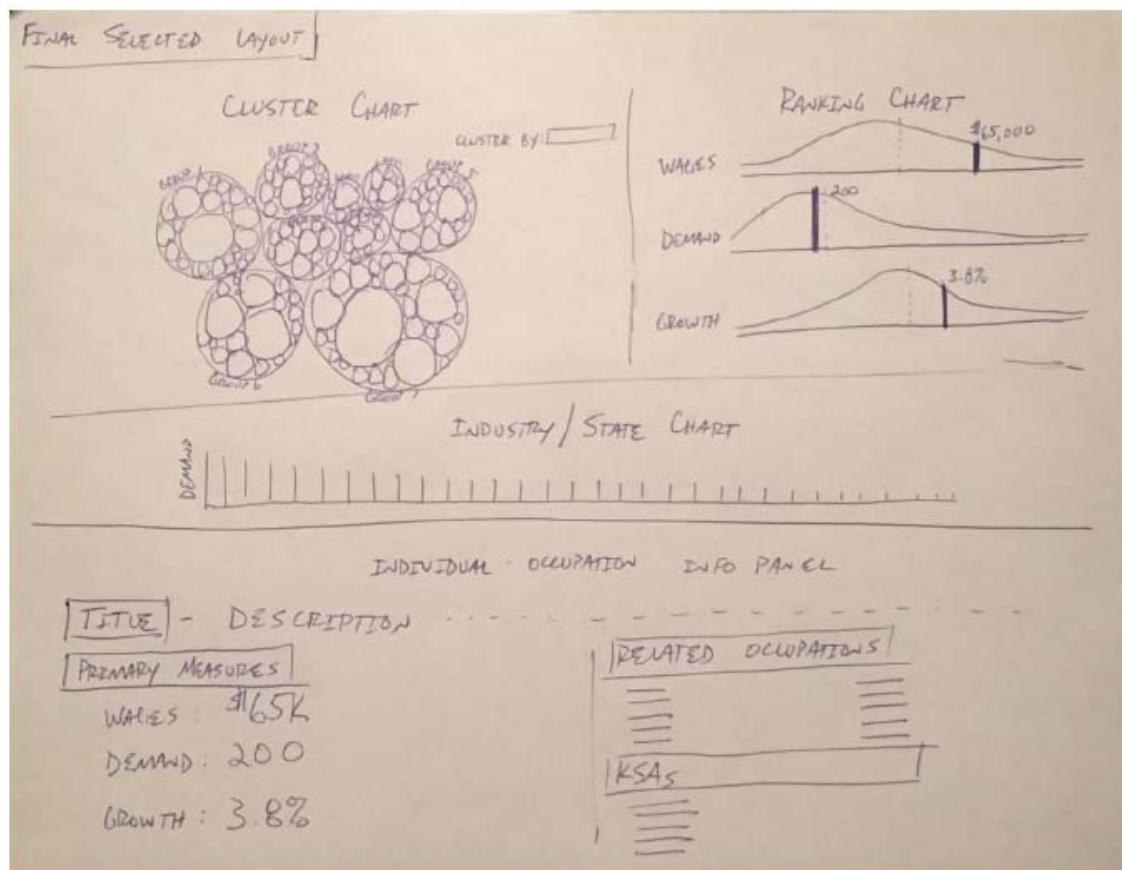
The final design wound up with six views organized into three primary sections. Initially I had conceived of four primary sections, but the fourth was an info panel and I realized as I went that given the amount of screen space the first section required, the user would have to scroll way down the page to see the additional information in the info panel, so it made more sense to create large tooltips that could be opened and closed by clicking on each bubble in the cluster chart.
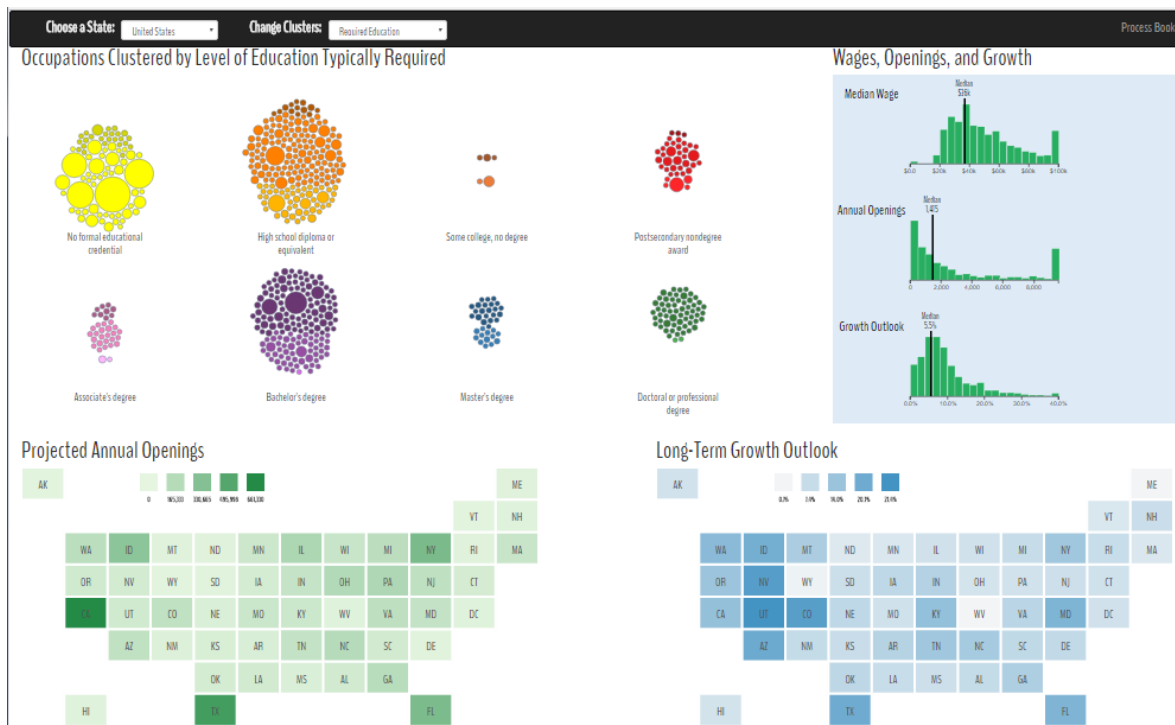
Section 1 – Occupational Cluster Chart

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 did 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 (especially younger viewers) to engage and start exploring.

The user can choose how to cluster the occupations by using a drop-down menu. Cluster options include major occupation groups, education required, STEM/Non-STEM, job training, and work experience. Each cluster is labeled and each bubble represents an occupation. Each bubble has a tool-tip on-mouseover with the occupational title and related occupations. The area of the bubbles encodes demand (i.e. the number of projected openings). The hue of the bubbles encodes the required education level and the brightness/darkness encodes three relative wage bands. Additionally, the wage bands are redundantly encoded by each bubble's vertical position within the cluster - higher wage jobs on top.

I considered using blocks of small squares to encode the clusters, but decided to go with the bubbles so I could use area 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. After trying this, however, I realized that encoding positional behavior at the cluster level was very complicated to convey to users and was not very useful. Further, it had the potential to bias a user's choices and they may miss a high-growth occupation hidden within a low-growth cluster (see below).


When the user selects different cluster groupings from the drop-down there are nice transitions to the new cluster so they can follow each occupation to its new cluster. Enlightening insights can be gained by this (i.e. watching the STEM occupations move mostly to the higher education clusters). Once the transitions were working, however, I realized that unless a bubble was very large it was hard to track among the other 839 bubbles moving to their new locations. I realized that it was necessary to enable highlighting of the bubbles when clicked to make them easier to track.

I also played with the idea of semantically zooming in on individual clusters on-click, and moving the bubbles into a more precise encoding of their relative wages on the Y axis and including an X axis to encode growth. This would reduce the number of bubbles and put them in an X, Y plot so that the viewer could visualize the growth dimension, but it seemed to be too much info to effectively absorb, and it would have been redundant given the encoding of growth in section 2, so I decided not to do that (see below).
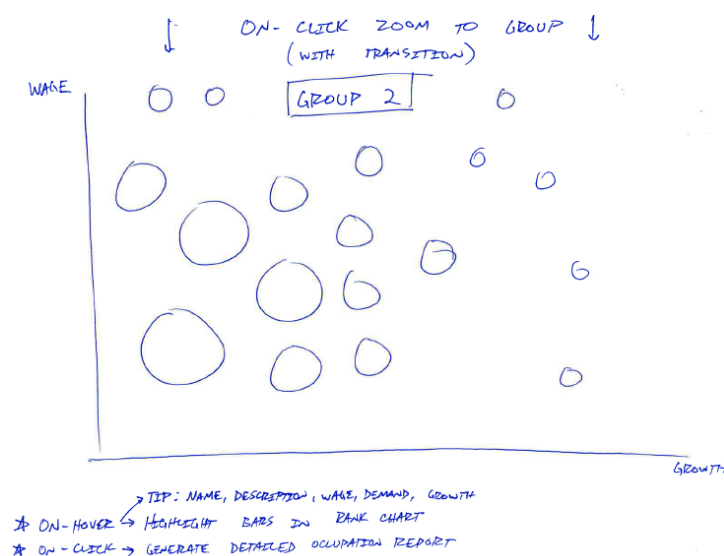
The initial cluster chart had the clusters in a circle. It quickly became clear that this was not a viable option given the large number of categories for some of the options, and the considerable screen space that something like this takes up. The clusters needed to be far enough apart to be easily distinguishable, but as close together as possible to limit screen usage, so I changed to rows of clusters (see below).
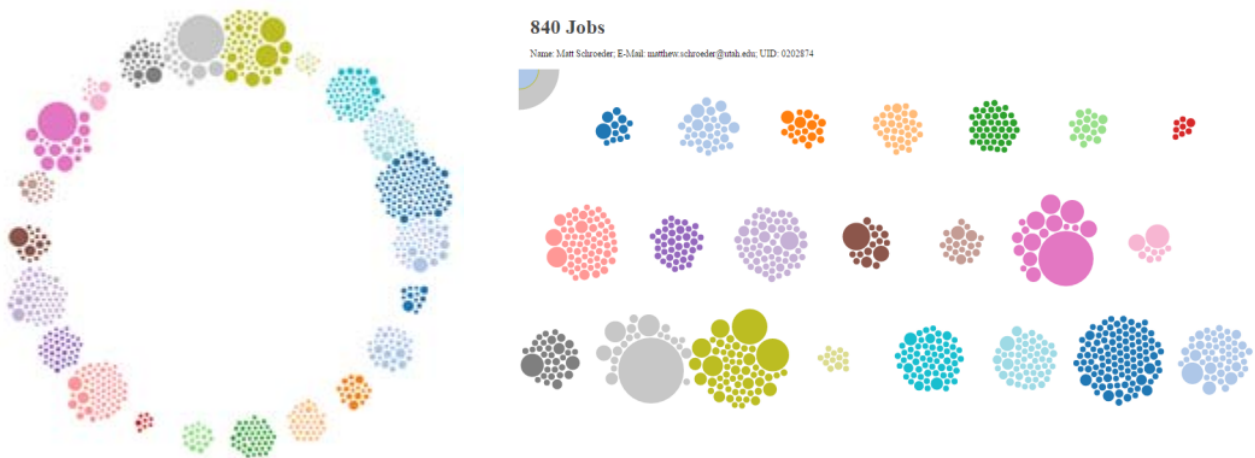
Initially, I thought it would be best to redundantly encode the categories with color, so when the user changed clusters the color of each bubble would change to match its new cluster. Granted, this makes for a prettier picture, but I realized that I was losing two things in doing that. First, the ability to easily track the same bubble from cluster to cluster, and also the original encoding of education level which is among the most important characteristics for each occupation.

As a result, I decided to have the default view open to clusters categorized by education so each cluster would be a uniform hue, and then keep that color upon cluster changes, so that the user could track them. The decision to redundantly encode the wage bands via vertical position arose out the initial results of encoding the wages bands as lighter or darker circles of the same hue. Without the vertical position, the clusters are just a random mixture and the initial encoding becomes useless.


Vs
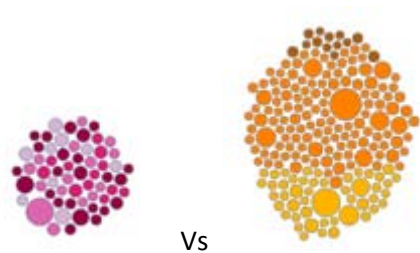
View 2 – Ranking Chart
This was most difficult view to determine how to encode. The goal of this view is to rank and compare all of the occupations across the three primary measures of interest – wages, demand, and growth. I considered several options.

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 (see below).

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.
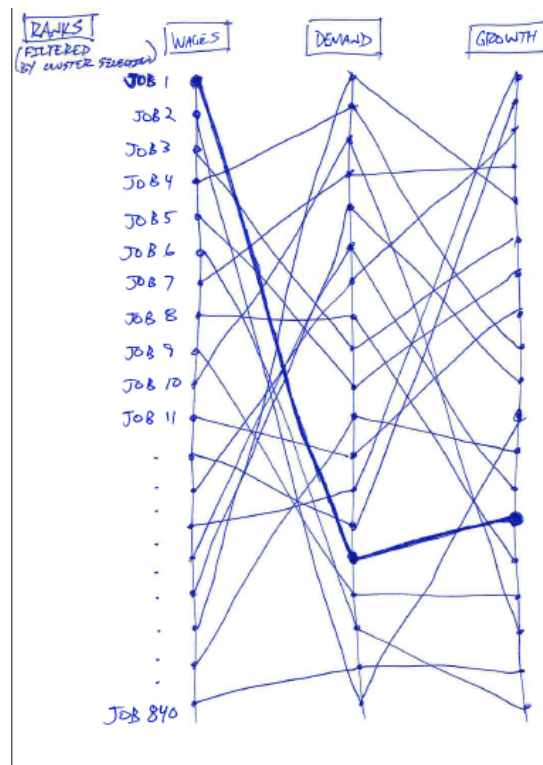
Finally, I considered using a distribution chart for each dimension that shows where the highlighted occupation shows up in relation to the median and in relation to the rest of the distribution. My concern with this one was that it may be overly statistical and may put some viewers off, but I think by explaining it well and creating good labels, it may prove to be the most effective option.
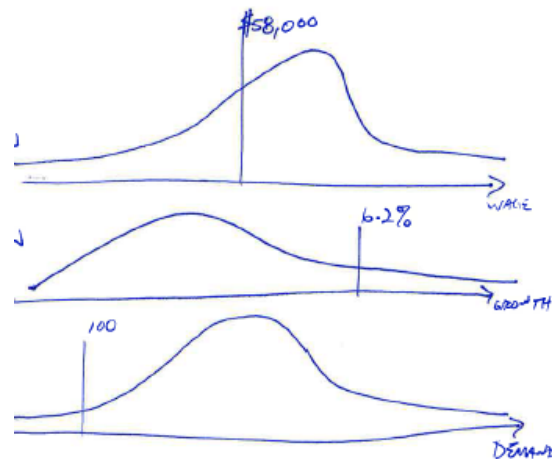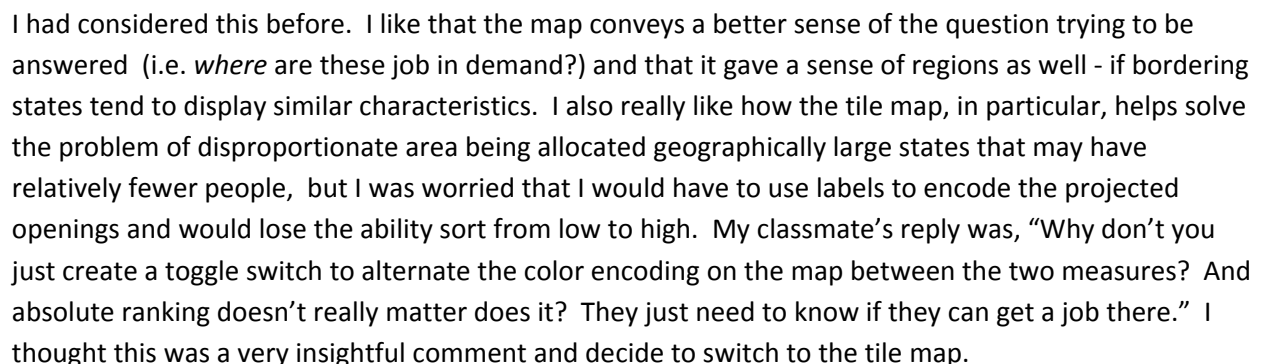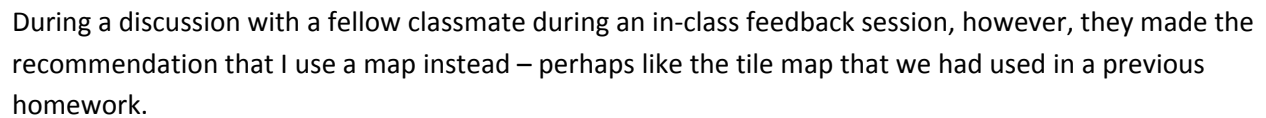

Ultimately I just want viewers to be able to tell how each occupation stacks up against the others across all three measures, and I think the distribution chart is be the simplest way to summarize so much information. Additionally, this option allows for the implementation of a brush on each chart that filters for the occupations that fall within that range.

The final implementation required the use of histograms rather that smooth density functions because all of these measures have much longer tails than would be useful to visualize. The histograms allow for grouping of the all jobs over some threshold into the final bin, whereas the density function would have to be shown in its entirety or cut off at some arbitrary point and the fact that there are still many more occupation still out in the tails would be lost (see below).

View 3 – State / Industry Chart
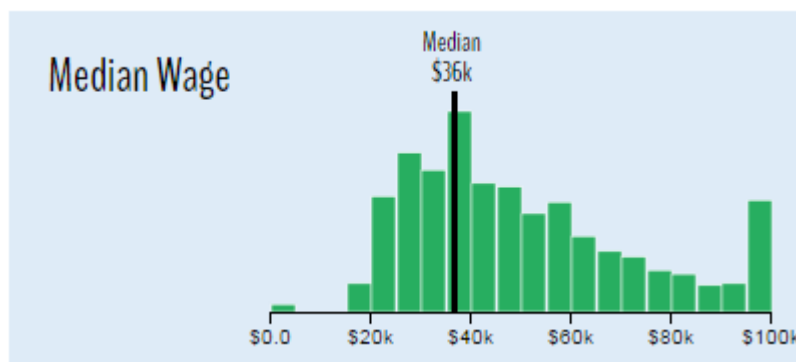
The intention of this view is to answer which states have the highest demand/growth for a selected occupation. Initially I chose to use a simple horizontal bar chart. It would have been in descending order of projected openings which would have been encoded in the bar height and projected growth would have been encoded by color.

During a discussion with a fellow classmate during an in-class feedback session, however, they made the recommendation that I use a map instead – perhaps like the tile map that we had used in a previous homework.

I had considered this before. I like that the map conveys a better sense of the question trying to be answered (i.e. *where* are these job in demand?) and that it gave a sense of regions as well - if bordering states tend to display similar characteristics. I also really like how the tile map, in particular, helps solve the problem of disproportionate area being allocated geographically large states that may have relatively fewer people, but I was worried that I would have to use labels to encode the projected openings and would lose the ability sort from low to high. My classmate's reply was, "Why don't you just create a toggle switch to alternate the color encoding on the map between the two measures? And absolute ranking doesn't really matter does it? They just need to know if they can get a job there." I thought this was a very insightful comment and decide to switch to the tile map.

After first creating a single map that I would toggle between measures I realized that the map was much larger than necessary and by shrinking its size to 50% of the screen width, I could include 2 maps simultaneously, one for openings, and one for growth. This solved two problems – unnecessary and awkward feeling screen use and the need for users to remember what the map looked like for the previous measure when toggling.

Projected Annual Openings
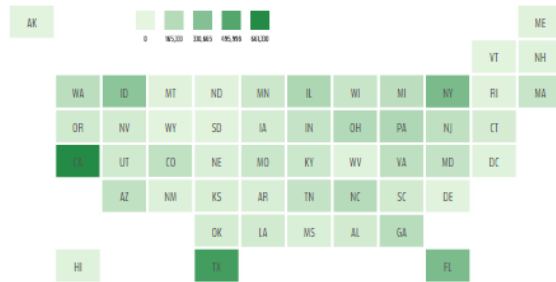
Long-Term Growth Outlook
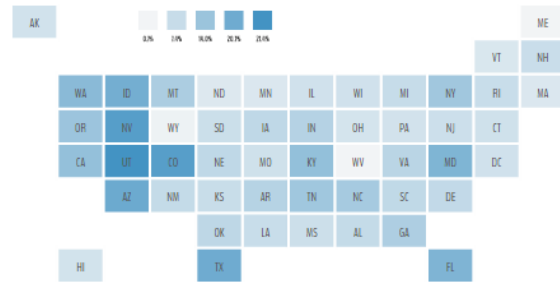
Conclusion

My overarching goal in this project was to improve access to valuable government data on occupations and help job seekers make more informed decisions. I think that goal, broadly speaking, has been met. The questions I had hoped to answer for users can all be found with relative ease and I think it has turned out to be a relatively engaging tool. I hope that you find it useful, informative, and fun as well.