Earn What You Learn

Ru Yu, Yi Wei, Longfei Li

Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, 90007

Abstract. This project is an interactive website that visualizes the data of salaries by majors and schools and was designed with Cairo's wheel of design principles. The visualization includes a barchart, parallel coordinate plot and a choropleth map.

1 Introduction

This paper summarizes our project for the visualization of salary data categorized by majors and schools. The motivation for this project is to inform high school and college students about salary information of different majors and schools, so they may plan their career path better. Moreover, we want students to carefully consider their academic focus in order to avoid dissatisfaction in their future job salaries. This paper lists and discusses the different visualizations that we used to show these salary data and also the location of schools and their graduates' income, so prospective and current students may use our projects as a reference.

2 Dataset

2.1 Data

The datasets those was used to make the project was from kaggle.com[5]. It has salary data categorized by college, region, and academic major.

degrees-that-pay-back.csv:

1. Starting Median Salary: starting median salary for fresh graduates.
2. Mid-Career Median Salary: mid-career median salary for people with professional experiences.
3. Percent change from Starting to Mid-Career Salary: percent of change of salary from graduation to mid-career. (not used)
4. Mid-Career 10th Percentile : 10 percent of the salary data
5. Mid-Career 25th Percentile : 25 percent of the salary data
6. Mid-Career 75th Percentile : 75 percent of the salary data
7. Mid-Career 90th Percentile : 90 percent of the salary data

salaries-by-region.csv:

1. School Name: the name of the schools of the salary data.
2. Region: the region of the school.
3. Starting Median Salary: starting median salary for fresh graduates.
4. Mid-Career Median Salary: mid-career median salary for people with professional experiences.
5. Percent change from Starting to Mid-Career Salary: percent of change of salary from graduation to mid-career. (not used)
6. Mid-Career 10th Percentile : 10 percent of the salary data
7. Mid-Career 25th Percentile : 25 percent of the salary data
8. Mid-Career 75th Percentile : 75 percent of the salary data
9. Mid-Career 90th Percentile : 90 percent of the salary data

usa2.json:

This is a United State GeoJSON data we obtained from a website[4] that hosts GeoJSON and KML data for the United States

2.2 Data Processing

We used Python to post-process the original data we obtained. Since the original data only has universities categorized by region but not by state, we have added state information to salaries-by-region.csv with Python. Also, since the original data's salary amount are strings, we used Python to convert all the salary data in both csv files into integers for easier use.

3 Design

In this section, we will illustrate and explain the design of our data visualizations, as well as the interactivity each visualization offers on our website.

Our project is also responsive to viewing devices, so the website will resize automatically to be viewed on smart phones and tablets.

3.1 Bar Chart

We designed a bar chart to visualize the starting and mid-career median salary data, categorized by majors. The bars are vertical and are color-coded to better differentiate between starting and mid-career median salaries.

The bars are, by default, grouped side by side, but there is also an option to display bars in stacked style, to better visualize the total income. There are also three buttons which allows user to sort the bars by Starting Salary, Mid-career Salary or Alphabetical order of major names. User may mouse over any bar at any time to display a tooltip that indicates the major and salary information.

We used D3's stack shape layout to create the stacked bar chart. When viewing in mobile devices, the vertical bars will become horizontal bars for better readability.

3.2 Parallel Coordinates

We designed a parallel coordinates plot of mid-career salaries for different majors and categorized by 10th, 25th, 75th and 90th percentile. Each axis is added a brush component, by clicking and dragging along any axis, you can specify a filter for the range of salary in that percentile you wish to compare. Ranges those are not select will hide the data lines.

There is also a table below the parallel coordinate plot which shows the whole dataset by default and it will update its data as user selects and filters the axis in the plot.

3.3 Choropleth Map

We designed a choropleth map to visualize the data by state and by region. We used GeoJSON data to draw the Unite States map. The default viewing option is by state, and each state is color-coded with a shade of green to indicate the median starting salary. Darker shade means less income and lighter shade means more income. When users mouse over any state, it will show a tooltip, which contains the information of that state's starting and mid-career median salary. After user clicks on a state, the bar chart on the bottom will show the ranks of universities in that state by starting median salary. User can also mouse over any bars to see the tooltip that contains the University name and starting and mid-career median salary.

User may also choose to view the data by regions, in which the states and their schools are categorized into five regions: Western, Mid-Western, North-Eastern, Southern and California. When click on a region, each state in that region is colored with a distinguishable color and the bar chart also shows all the universities in that region, with bars the same color as its university's state. Tooltip feature is also available in this viewing mode.

4 Related Work

4.1 Works by others

There are similar visualizations created by user Chris[3] on kaggle.com to explore the college salaries by major. He mainly built scatterplot to visualize the data and intended to perform exploratory analysis by focusing on part of dataset.

US News[1] website also offers ranking of universities by national, regional and major ranks, and it also contains income informations. However, they do not have an interactive map to compare universities income information by region or by state.

4.2 Originality

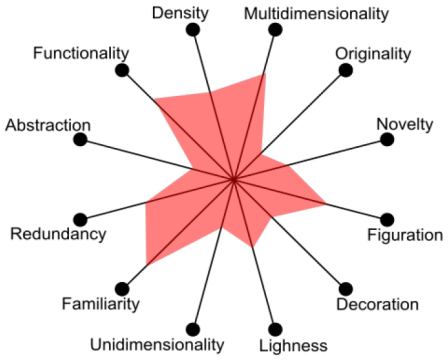
We have created this project as a centralized place for students to explore income data, so they do not have to search everywhere online for such informations.

We found the dataset we obtained contains many useful information, but it is not easy for people to formulate comparisons with the raw data, so we visualized the whole dataset by building stacked and grouped bar chart. This chart offers rankings of income by majors so users can compare the majors more efficiently by sorting the bars. We then showed the percentile of salary incomes for mid-career by building parallel coordinates graph, which the user can drag on axis and select the range of income to compare with. Lastly, our dataset contains a map that visualized the salaries by state and by region, so user can mouse over any state or region to see the median salary. There is also a bar chart below the map, sorted by starting median salary for each universities.

5 Development Plan

5.1 Design Principles

This is our Visualization Wheel:



We based our designs on Alberto Cairo's Visualization Wheel in his book *The Functional Art*[2].

We focused more on **functionality** than on **decoration** for we do not want extra images to further misguide the users, since we already have many charts and different colors on the website.

Our **redundancy** is higher than **novelty** for we want the user to explore the data in many different ways with the same visualizations, so we offered multiple viewing options on the same bar chart and map.

We did not want the users to be baffled by fancy visualizations, so we focus on improving **familiarity** and reducing **originality** by using well-known visualization techniques such as bar chart and map.

Since there are multiple visualizations on the same dataset, our **multidimensionality** is higher than **unidimensionality**, so user can see the same data in different ways and better compare the data of interest.

5.2 Tools

We used D3 to visualize the data with different visualization techniques. The website was built using HTML and CSS. Javascript was used to load the data

and apply D3's data binding for creating visualization. To make the website responsive, we used Bootstrap. Lastly, we used Python to process and to clean up the data.

6 Conclusion

Our objective to create this project is for students to use as a reference guide when applying to universities or switching majors. We saw many article online stating that people in their mid-careers are not satisfied with their job and the leading cause is salary. Therefore, we wish to help students to avoid such mid-career crisis by getting informed about salary and majors before they start college or graduate from colleges. By visualizing data, we found that attending universities in the Mid-West leads to the lowest salary both at graduation and at mid-career, and salaries are higher if graduated from North-Eastern or California universities. Surprisingly, Physician Assistant and Chemical Engineer have, opposite to public impression, higher starting salary than Computer Science.

References

1. Best Colleges — College Rankings and Data — US News Education, US News, www.usnews.com/best-colleges.
2. Cairo, Alberto. The Functional Art: An introduction to information graphics and visualization. New Riders, 2012.
3. Exploration of College Salaries by Major, Chris, www.kaggle.com/cdelany7/exploration-of-college-salaries-by-major.
4. GeoJSON and KML Data For The United States, Eric Celeste, www.eric.clst.org/Stuff/USGeoJSON.
5. Where It Pays to Attend College, www.kaggle.com/wsj/college-salaries.