

Chuting Liu. A Visual Analysis to Help College Freshmen's Career Choosing – via Aspects of Education Level and Salary. A Master's Paper for the M.S. in I.S degree. April, 2018. 40 pages. Advisor: David Gotz

This study is mainly about constructing a visualization project to help college students have a clear and intuitive view of general circumstance of occupations, so that to help them with future career choosing hopefully. Firstly, the users' potential information needs are discussed and different kinds of personas are built, then comes the visualization. The visualization interface mainly contains three graphs showing information of general occupation types, education level and salary of different positions, followed by some brief verbal analysis of the data and then the evaluation part, which is in the form of online survey. Data used in this project is mainly second-hand data accessed via official websites of government.

Headings:

Data Visualization

Visual Analysis

Web Development

A VISUAL ANALYSIS TO HELP COLLEGE FRESHMEN'S CAREER CHOOSING –
VIA ASPECTS OF EDUCATION LEVEL AND SALARY

by
Chuting Liu

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

April, 2018

Approved by:

David Gotz

Table of Contents

Introduction.....	2
Literature review	4
Methods.....	9
Data collection.....	9
Research Design.....	10
Evaluation part -- survey	11
Discussion of visualization results	12
Constructing the project.....	13
Building the personas	13
Data processing	16
Data visualization and building the interface (webpage)	17
Some conclusion from the data and graphs.....	22
Evaluation	25
Participants recruitment and obtaining data	25
Feedback from the participants	26
Discussion of evaluation results.....	27
Discussion and conclusion.....	28
Future work -- Potential draft of second version.....	28
Lessons learned	30
Limitations	31
References.....	32
Appendix.....	36
Evaluation survey	36

Introduction

Many people always have the intent to pursue higher level of education, either for getting better jobs or attaining self achievement. With a diploma of Bachelor's degree or higher, they can get into fields which require more knowledge and training (such as becoming a lawyer or physician) as well as having likely higher pay. It is surely admirable to attain higher education, but from the aspect of finding jobs, the situation of job market now seems not so optimistic towards students who spend 4 years or longer in college, and they may not have certain advantages when looking for jobs comparing to others who only have a high school diploma.

Future income is another inevitable part when thinking about careers, and it is reasonable for students to consider it as a main factor of job choosing other than personal interest. As a result, for many high school students and university freshmen who are still struggling about their future majors or career choices, they have indeed plenty of things to consider, such as the input-output ratio between education cost and salary income, or whether their desire for high achievement in specific fields is strong enough to be put in the first place instead of compensation.

Based on these career choice concerns, students are motivated to consider a variety of employment-related questions such as listed below:

1. What is the link among career profession, education level and salary? How are they related to each other, and how this affects students' career choices?

2. Whether well-trained college students can become more competitive and qualified job candidates in specific industries than those with lower education levels?
3. What industries need more skillful and trained workers so that they tend to require people with higher education levels? How well do those workers get paid?
4. Is higher education level a necessary condition to get higher pay? Do workers with higher education level always get higher pay?

With those considerations of future career choosing, students need some better ways to get informed of relevant information. Other than digging directly from plain texts and numbers, a visualization project can potentially be one of the better solutions, which can help students have a clear and intuitive view of general job situations in different industries so that keep themselves informed and make a better career choice.

For university freshmen, they can compare the career professions with required education level and salary, to consider which major they are going to choose for the job; even for those students who have a clear career goal already, looking at how well it is paid in the field and what education level is more competitive for getting the job is still helpful. Besides, students can also think about the balance between the tuition cost and potential future income of their career interests to judge the necessity of getting a college degree.

To construct the visualization project, a visualization pattern is firstly designed based on users' requirements, and then an interface is built from that design. After finishing the interface, a user study is conducted to test the efficiency of the visualization in order to see how well it can help users obtain the information.

Literature review

After searching for academic papers, there are some related fields which have some relevant studies, such as career choice and salary, career choice and other factors like self accomplishment, data visualization and data analysis. Following are the theories and opinions which are interested and necessary to look at.

Career choice

Career Choice and Salary

When it comes to the decision of picking a future career, it is inevitable but reasonable to put salary level into consideration. Kinouani et al. (2016) conduct a research about a group of medical students' career choice, and find out that for the career of doctors, the main factors affecting the students' choice are occupational factors, which also include the salary level. Fatima et al. (2015) send out a career choices questionnaire to hundreds of undergraduate students in South Africa, and the analyzing results shows that financial factor weighs heavily when students decide their careers, and the potential for personal development as well as future high earnings are some of the most important reasons.

Other than the students who have currently not made final decision of their career, salary is still a big issue for people who are already at work to determine whether still remain in the current career or not. Peters (2015) looks into the trend showed by 2015 Pharmaceutical Technology Europe annual survey, saying that European bio/pharma employees tended to have a greater desire to change jobs than before, because of the more working hours with few salary increases. Although people may usually hold the

hypothesis that higher education level will receive higher level of earnings and there is a trend to go even upper, Yang et al. (2015) actually argue that the postdoctoral experience has actually not much impacted on the salary after completing the degree a decade later, but will have significantly positive contribution towards people's self accomplishment.

Career Choice and other factors such as self accomplishment

Except for the practical aspect of compensation, still plenty of other factors have great impact on students' career choice, such as self accomplishment and strong career interest. Ojeda et al. (2009) do some research on undergraduate freshmen and prove the theory that many people think entering university means self accomplishment and higher social status in the perspective of career choice. Schulz et al. (2016) indicate that choosing the field of study will greatly affect students' professional choices in the future, and earnings largely depends on students' major; also, confidence is important for selection, students in some certain majors tend to be over-confident while those in some other majors are under-confident.

On the other hand, the situation varies for people from different cultural background. Peri et al. (2011) find out that it can be quite different for highly educated native and foreign-born workers to choose new occupations, as the foreign-borns tend to specialize in quantitative and analytical skills for occupations, while the natives do better in occupations demanding interactive and communication skills. Abrahamsen et al. (2014) study the case of ethnic minority students in Norway, and find that they have higher career ambitions but their expectations are not correspondingly high, as ethnicity is both a resource and an obstacle when pursuing professions.

Data visualization

Data visualization is a very broad concept, and according to the definition by Azzam et al. (2013), it is a process which is based on qualitative and quantitative data and results in an image that represents the raw data, and the image is readable by viewers and supports exploration, examination and communication of the data, while the theory of this definition is originally from Kosara (2007). Mallon (2015) introduces several useful visualizing tools in the paper, such as Piktochart, Mapbox, etc., and gives some examples of their usage. Many researchers also work on building new framework of data visualizing, for example, Kai (2013) has a new way to create dynamic visualization by using maps or even more than two dimensions, while Munzner (2015) provides a systematic and comprehensive framework which considers visualization in terms of principles and design choices.

While the technique of data visualization gets much developed and advanced, there are still concerns and challenges of this field. Kovalerchuk (2017) studies the lossless visualization of n-D data and considers it better than 2-D in perceptual and cognitive abilities for discovering visual pattern, but human has inability to discover patterns in n-D data using a naked eye, which therefore motivates the development of 2-D visual representations of n-D data. Blas (2014) et al. reveal a crucial challenge at the user interface level – the question of how to convey all turn-takings efficiently between user and the system.

As a powerful tool of data demonstration, data visualization has been applied to various of fields. Pomarede et al. (2017) use a visualization software to help the advances of extragalactic cosmography, while Wang et al. (2017) use visual analytics of student

data to improve education and invent a special visualization tool for supporting. Other than that, Logre et al. (2015) indicate in the paper that visualization dashboards can support end-user decision making process, while the current designs of dashboards are still unsatisfactory and facing challenges, and the solution can be aligning them with modular paradigms. Data visualizing is also useful when it comes to data extraction and export, as Klomklao et al. (2016) develop a tool for the users to observe data of tweets' hashtags and export the results needed for further analysis.

Data analysis

Data analysis appears everywhere, which is vital to both everyday life and academic studies. Devitt (2015) writes that the goal of data analysis is to organize and condense large amounts of data, making them comprehensible to the readers, and analyzing quantitative data can highly support certain studies. However, Lucko et al. (2010) say that in some specific fields as constructing and management, archival data sets are not always correct and consistent, which may cause difficulties on the step of data preparation.

Different types exist within the field of data analysis, such as object oriented data analysis mentioned by Marron et al. (2013) in the paper, which is the statistical analysis of populations of complex objects, providing a framework to help solve many modern complex data analyses; Kohli et al. (2014) talk about R language related to data analysis, believing that it gives an opportunity to work on real time data from social networking sites, which can possibly be helpful.

On the other hand, big data has become a popular topic recently, thus the needs of techniques of analyzing large scale of data are urging. Fisher (2016) discusses some of the major challenges of data analysis when dealing with data at scale, saying that using those techniques of big data processing involves a lot of difficult design trade-offs. Norvig (2011) talks about Internet scale data analysis, including some big data tools like MapReduce and Pregel, and it is also related to data mining and cleaning.

As for practical usage, researchers have used data analysis to build or improve many useful systems, such as Basole et al. (2015) do on learning healthcare system, Becker (1984) on designing the S system, Milo et al. (2016) present a system to help users with personalized recommendation when the users are non-experts, and Shyr et al. (2014) do something on automated data analysis, which is important to big data analytics, and they discuss how possible the further research directions will be of it.

As a conclusion, many articles and books have already discussed about the factors affecting people's (especially students') career choices, and about data visualization and data analysis, which is very informative for the basic context of this project, especially the part of data visualization which provides lots of useful concepts and techniques.

When discussing the relationship of career choosing and salary, many studies do indicate that salary is one of the key factors, but the context of those case studies are mainly outside the United States but in regions such as Europe and Africa, so the trend here in the US is still lack of adequate studying. And when it comes to the link between education level and career professions, there are not much study about this part, and most of the related papers talk about other aspects such as ethnicity.

On the other hand, there is few visualization which helps illustrate career choosing by showing education level and salary, relevant works are mainly with tables and raw data only, which may not be intuitive enough.

Methods

This project includes both the quantitative and qualitative methods, quantitative for constructing the visualization and qualitative for evaluation.

Data collection

Data used for the visualization is accessed from the webpage in Bureau of Labor Statistics of the United States Department of Labor, including the distribution of workers' education levels and salary levels in different industries. Those data are officially published by the Bureau of Labor Statistics, which are publicly accessible to people and with no restriction. Some of the important data set of the project are as following:

Data set 1: National Occupational Employment and Wage Estimates

This is the most important dataset for the project, which includes the values of employment numbers, median hourly wage, mean hourly wage and annual mean wage of different occupations, etc., and the occupations can be sorted by industries. And the data from 1988-2016 can be accessed via another linked website, so there will be enough data to draw the trend by years. In this project, however, data of the most recent three years of 2014-2016 is mainly used.

Data set 2: Educational attainment for workers 25 years and older by detailed occupation (2015-16)

This is another important data set of the project, with 7 classes (columns) and 820 records (rows), but still needed to categorize it into a visualizing-friendly version.

Those data sets are sorted by detailed occupations instead of industries, but according to categories provide by North American Industry Classification System (NAICS) (<https://www.bls.gov/oes/current/oesrci.htm>), those occupations can be put into groups of broader scales as industries so that can make a clearer and more concise view. In this project, the category of 22 general occupation types is used for grouping.

Research Design

Data analysis and visualization

The main contribution of this project is a visualization interface to show the relationship among career professions, education level and salary, so that the data analysis is mainly about extracting and filtering data by different attributes (most parts are done manually, having tried using Python and KNIME the data analysis tool as well) and then write codes in JavaScript (mainly applying D3 library) to construct the interface. Besides, 3 personas are needed to better “tell the story” of visualization, and they should be built based on those three different kinds of information needs of users, which also helps find out what questions the visualization should answer specifically.

Other than cleaning the data and finding out the attributes needed for visualization, the design of visualization/interface is also one of the essential parts, which

is followed by actually working on the code. The interface would contain a few graphs showing different aspects of the dataset (general information about the occupations, salary level and education level of different job types), along with some text explaining the usage.

What's more, after building the visualization, some analysis in text follows up as illustrating findings and trends of this visualization, mainly from the aspects of different industries.

Evaluation part -- survey

As for constructing a data visualization project, it is necessary to have a final part as evaluation to test the efficiency of using this visualization and also how well it can help users get more relevant information and a better view of the field. For this part, a survey (questionnaire) is sent out with the visualization webpage link to ask people about the general view of the visualization interface and give some specific but comparatively easy tasks (mostly one-line questions, for example: what's the annual average salary of lawyer in 2014) to see how the users feel and how efficient it is when using the visualization results to solve problems. The goal of those tasks is not to retrieve people's answers and make judgment, but rather for the purpose of helping people become familiar with the visualization and see whether they can get the answers fast and correctly.

General questions of visualization have the answers recorded in a Likert Scale (using the type of "disagree-agree", with the range of 1 to 5). Other questions such as those about the tasks include:

1. asking how long does it take to finish the tasks;

2. is it easy to explore around different graph sections and find the information needed;
3. whether the user answer it correct or not (answers of the task questions will be provided), etc.

The questions about users' feelings to the visualization can be quite subjective but still helpful for evaluation, so the survey also has a few questions of this kind to get some subjective judgment, for instance, how's the users' overall impression of the graphs, is the graph sections easy to use/understand, which section/view of the visualization interface interests the users most.

Participants: the subjects group is the students in UNC-Chapel Hill. Sampling methods are convenience sampling and random sampling, since the flyers of this study are posted in campus and the participating request is sent out via emails by those students who are interested, while some other subjects are recruited by me sending requests to some UNC students I already knew or via some of my friends' help. The expected size of sampling group is 10.

Discussion of visualization results

After conducting the visualization and doing simple usability tests by survey, there is a part of verbal analysis about some basic trends based on the data and visualization, such as:

- a. the connection among education level, salary and positions;
- b. picking some popular positions with high level of salary or specific level of education and trying to see why they are leading the trend of job market;

- c. finding out about some positions whose education level seems not match to salary level or vise versa, and discussing why the situation occurs combining the social environment and other conditions.

Other than the questions above, how the data and visualization can help the 3 personas' examples is also discussed.

Constructing the project

Building the personas

As stated in the introduction part at the beginning, career choosing has always been an important issue to students and there are many factors to be considered. Undeniably one of the top concerns is about wage, which is essential and may directly affect people's decision of whether going on higher education levels such as going to college or post-graduate study. Family factor is surely one of the key points, financial situation and education level of parents would have certain effect on the decision of getting education and planning for future career. On the other hand, for certain occupations it is more competitive for candidates who has higher education degrees. Thus, practical purposes do play an important role in career choosing except for personal interest, and those compose users' information need.

For the purpose of building a visualization project to help career choosing, potential users of this project will mostly be university freshmen who still consider about their future careers, or even high school students if extending the range:

1. university freshmen: to see which major(s) to pick, as what job to do in the future; or to see the salary trend and the competitive education level for getting the job;
2. high school students: to see whether going to college or not.

For the convenience of construction and evaluation, the aimed users are therefore defined to be UNC freshmen who still does not define his/her major and not sure about what job to do in the future. For those students, there are three possible status of their potential information need:

- a. have interests in some industries, but not sure about what education level is competitive for entering specific career professions;
- b. want to know how are the salary situations (starting salary, mid-long term salary prospect, etc.) of some professions, and to see whether the salary “equivalent” to the effort and cost for attaining higher level of education;
- c. don’t have specific interests, but want to have a look at the general trend of salary and education levels of employees in different industries.

Therefore, the visualization interface should include enough information about general industries, detailed occupations, salary and education level, etc. to help users find out the answers of those potential need above.

Following are three examples of persona which is built based on those three status of information need above:

Persona 1

Name	John Doe
------	----------

Year of School	First year
Field of Interest	Laws, accounting, public policies
Education level of parents	Father: PhD; mother: Master's degree
Annual total income of family	~\$200,000
Expectation of education level	N/A

In the case of John, being interested in the fields such as laws and policies, he is not sure what degree is competitive for getting a relevant future career. For potential factors of his decision, his parent got comparatively high education and the finance situation of his family is able to afford further education if he wants.

Persona 2

Name	Jane Doe
Year of School	First year
Field of Interest	Computer science, mathematics, psychology
Education level of parents	Both have a Bachelor's degree from university
Annual total income of family	~\$100,000
Expectation of education level	Bachelor's degree

For Jane, she likes the field of computer science and mathematics and expects to get a Bachelor's degree from university, yet she isn't sure whether she should continue further studying afterwards. For potential factors of her decision, the situation of hiring in those fields needs to be considered in order to be competitive in it, and it may also

depend on the salary level of those occupations as she needs to invest more time and money for the higher education, which can be a tricky case for her family.

Persona 3

Name	Jack Smith
Year of School	First year
Field of Interest	N/A
Education level of parents	Father: Associate degree; mother: high school diploma
Annual total income of family	~\$60,000
Expectation of education level	N/A

For Jack, he has got involved in university but not sure about his field of interest yet. For potential factors of his decision, the financial situation of his family as well as family background would potentially make an impact on his choice of career field or even whether getting the bachelor's degree or not.

Data processing

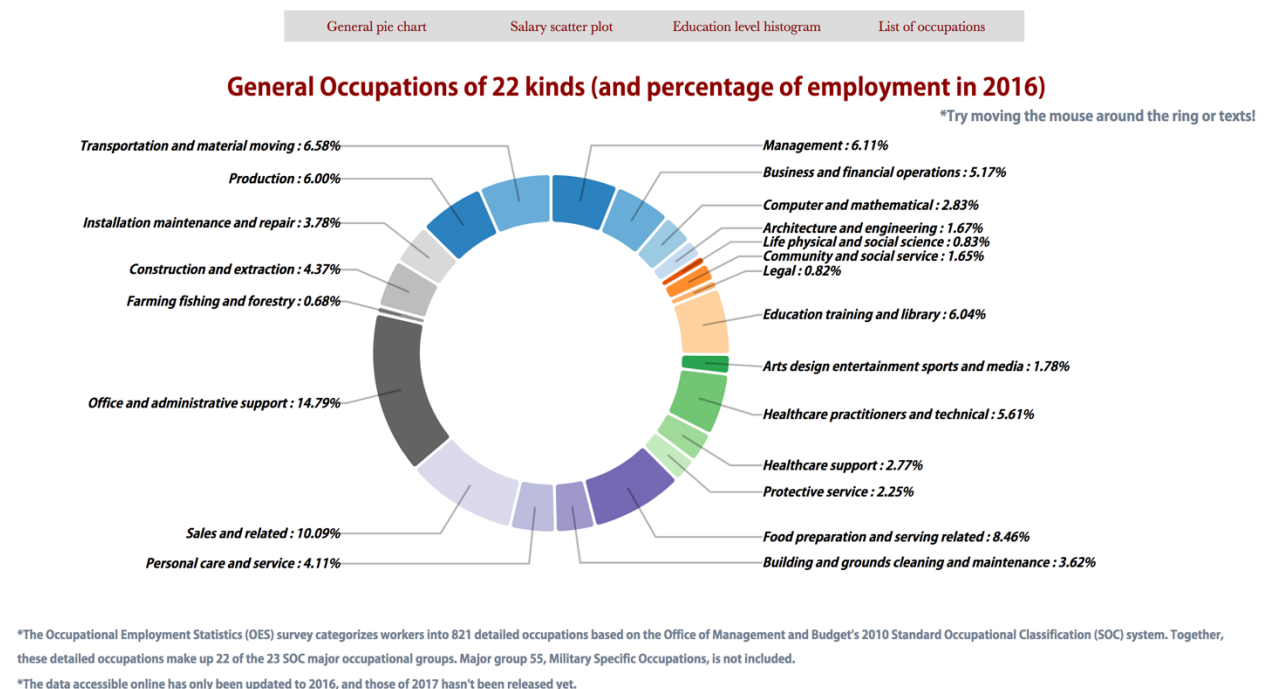
For the visualization part, the datasets were accessed from the official website of Bureau of Labor Statistics, and those were downloadable CSV files. In order to make it easier for constructing the visualization graphs, those data needed to be cleaned. I mainly used Microsoft Excel to look up the data records and editing the columns, and I created four new CSV files to be imported into the JavaScript code of the graphs.

Except for Excel, I also tried using other methods such as Python and the analysis tool of KNIME to clean up the unnecessary data then convert the file format.

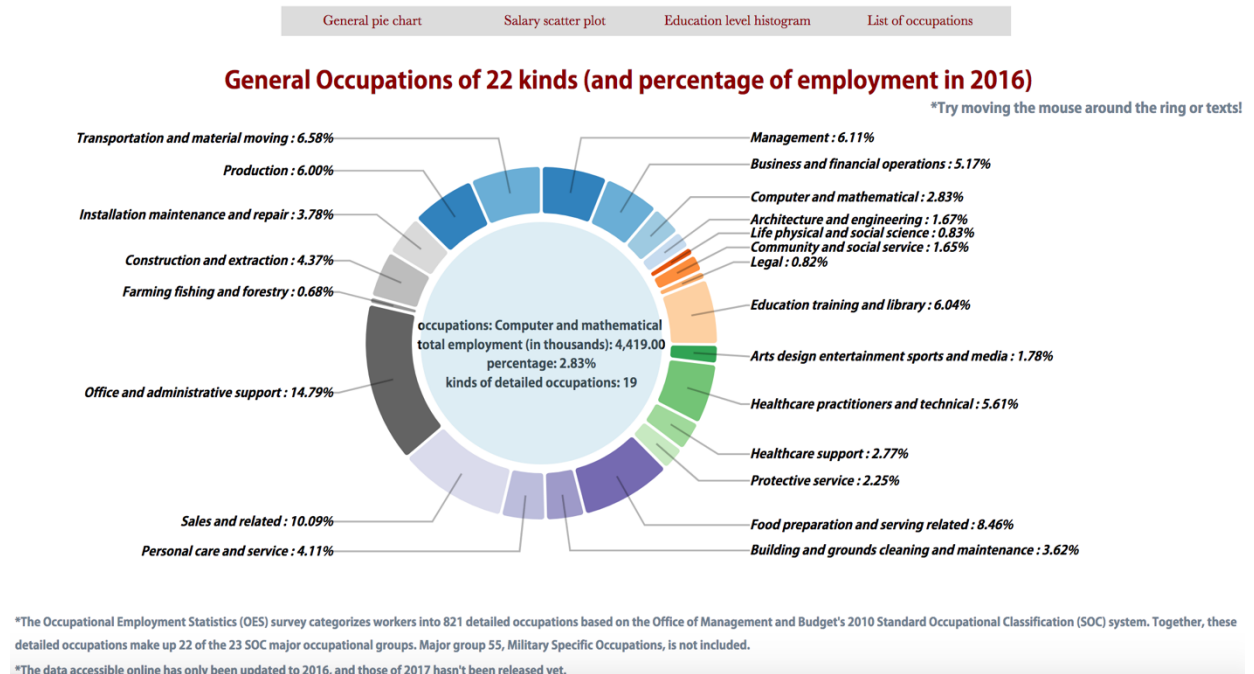
Data visualization and building the interface (webpage)

According to the users' potential information need, three graphs are made to help them out:

1. a pie chart showing all the general occupation types and also details about hiring numbers and percentage of the specific type, which serves as the main graph of the visualization so that users will first see this graph when opening up the interface; when moving the mouse to specific parts of the pie chart or the text, details will show up in the middle of the chart;

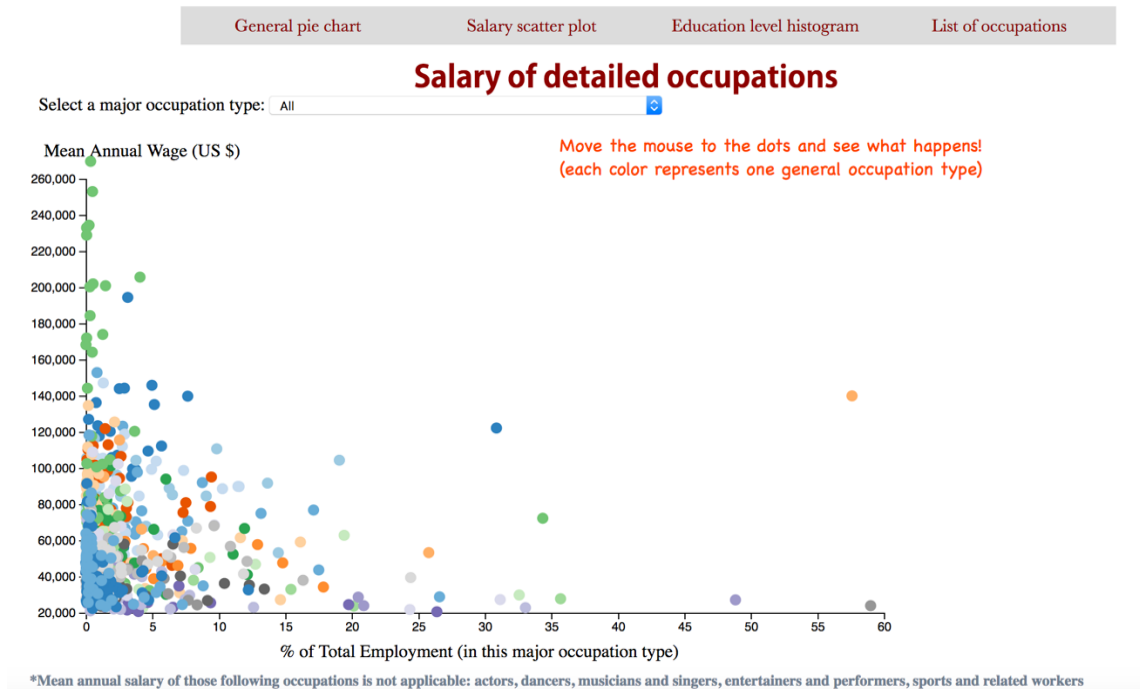


graph 1: General pie chart (default setting)

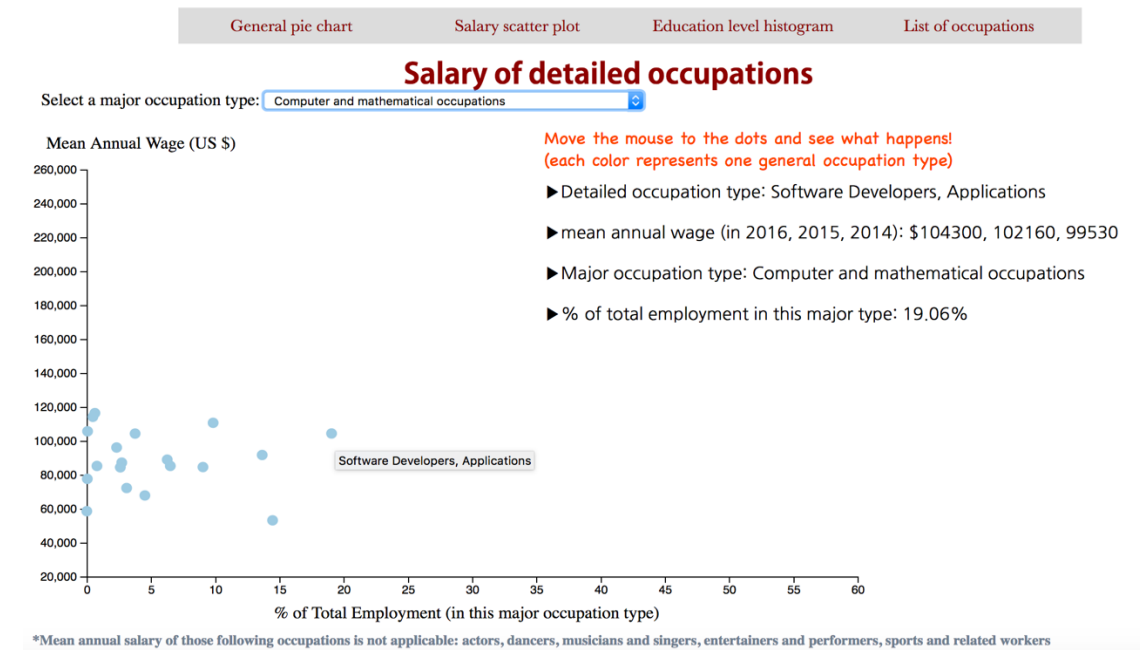


graph 2: General pie chart (moving the mouse to a specific section)

2. a scatter plot showing the salary level of different general occupation types, users can look into salary of detailed occupations and also employment percentage by selecting a general occupation type from the drop down bar; when moving the mouse to the dots, text of details will pop up at the top right corner;



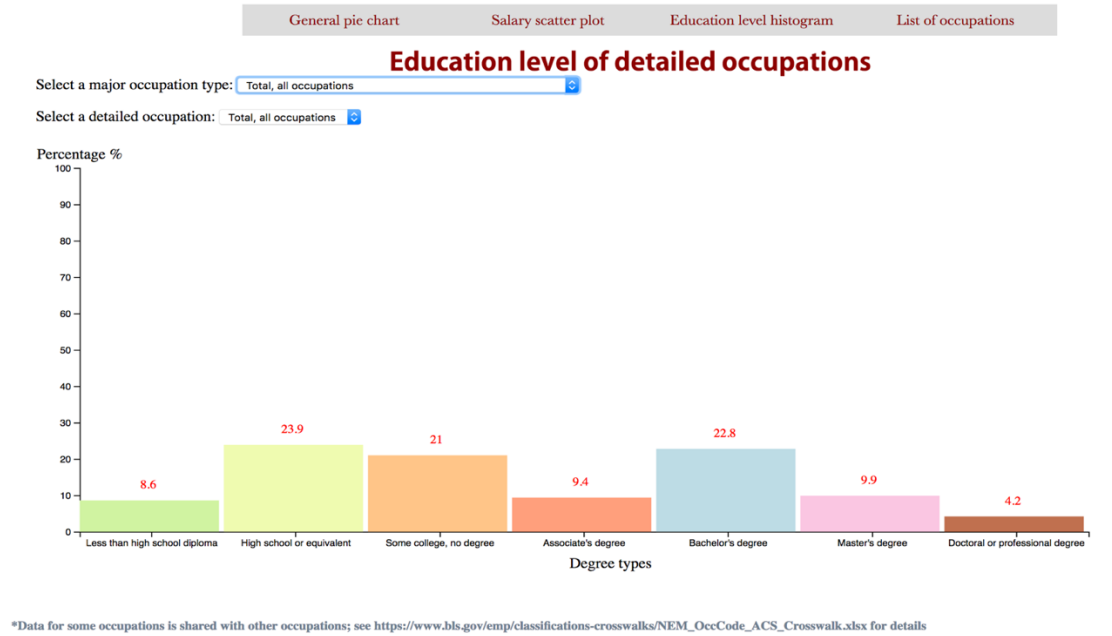
graph 3: Salary scatter plot (default setting)



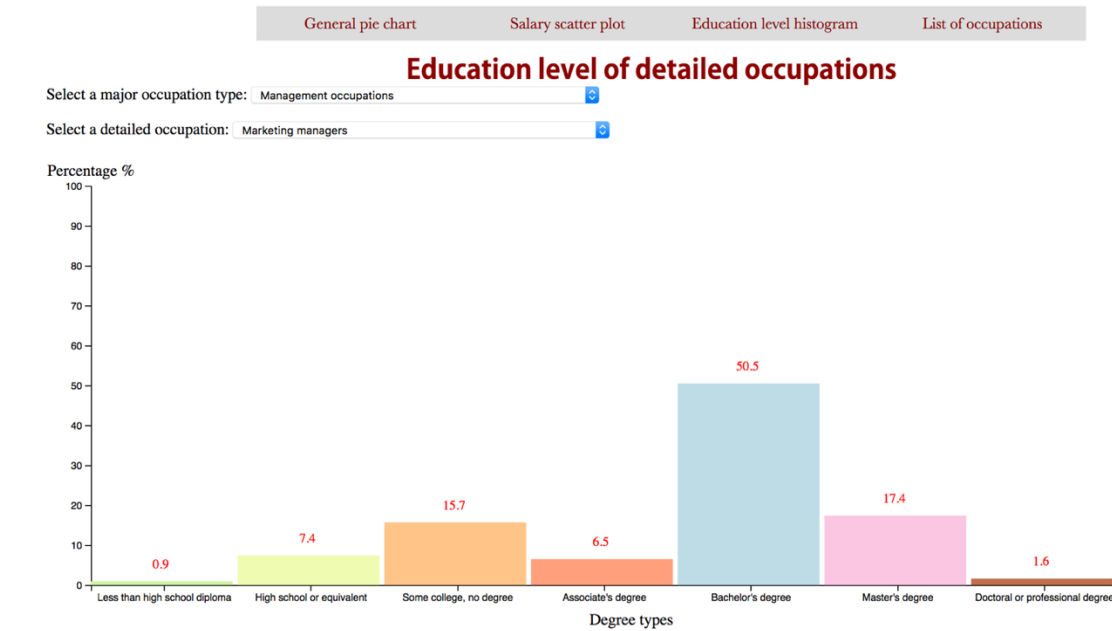
graph 4: Salary scatter plot (selecting a specific type of occupations)

3. a histogram showing the education level of specific detailed occupations. Having two drop down bars correlated together, users can first select a type of general occupation

from the first drop down bar and then the second one will automatically display only those detailed occupations included in this general type. By selecting different options in either of the drop down bars, users can see the histogram below changing simultaneously.



graph 5: Education level histogram (default setting)



graph 6: Education level histogram (selecting a specific detailed occupation)

For the coding part, I mainly used the D3 library (Data-Driven Documents) of JavaScript to build the graphs and HTML5 and CSS3 for constructing the webpage, while also using some functions of general JavaScript to implement the two correlated drop down bars in the histogram.

In order to make it easier for users to find out what occupation they are looking for, the fourth page of website which includes a table of general and detailed occupations is also placed in the interface as an appendix, so that users can search for the general type of some detailed occupations they want in that table or just browse through to get some general ideas.

General pie chart	Salary scatter plot	Education level histogram	List of occupations
-------------------	---------------------	---------------------------	---------------------

List of detailed occupations

(22 general types, 820 detailed occupations in total)

*Try Ctrl+F to search for what you want!

General types	Detailed occupations
Management occupations	Chief Executives
	General and Operations Managers
	Legislators
	Advertising and Promotions Managers
	Marketing Managers
	Sales Managers
	Public Relations and Fundraising Managers
	Administrative Services Managers
	Computer and Information Systems Managers
	Financial Managers
	Industrial Production Managers
	Purchasing Managers
	Transportation, Storage, and Distribution Managers
	Compensation and Benefits Managers
	Human Resources Managers
	Training and Development Managers
	Farmers, Ranchers, and Other Agricultural Managers

graph 7: List of detailed occupations (appendix of the interface)

After finishing the visualization interface, I put it on the web server of SILS's (<https://opal.ils.unc.edu>) to host the link for evaluation.

Reasons for the design

For the main graph which presents the general view of occupations, pie chart was chosen not only due to the fact that it can clearly show the ratio of each occupation by separating and coloring different parts, but also a complete circle can give people the impression that the whole dataset is combined together. Besides, rational visual effect is one of the key points when designing data visualization – especially the main graph, and pie chart can be more impressive to serve as the main one comparing to other types of graphs, which is proved by users in the following part of evaluation.

The dataset of scatter plot includes information of many aspects, so that it needs a type of graph which can help with displaying that; on the other hand, scatter plot can show the cluster of dots, which is one of the main purposes of visualizing this dataset of salary – to compare and see the trend of different positions' compensation.

Using histogram to show the education level of each detailed occupations can help directly compare different parts, and the two drop down bars strongly support the function of filtering to make the graph usable.

As for the color, I applied the standard color sets which could be accessed online, and chose the lighter but still contrasting colors to avoid causing potential visual fatigue of the users.

Some conclusion from the data and graphs

According to the graphs, there are certain patterns of positions requiring different education levels and offering varied ranges of salary, which is mostly similar to people's common sense (or “stereotype” in some degree). For example, many people nowadays

tend to focus more on the field of computer science, and related occupations do offer a comparatively higher level of salary. People can get to work in this field by either going studying for it in college or self learning, but many companies – the general trend of the computer science field, in other words – still prefer candidates with a bachelor's degree or higher; which means that no matter how well someone has learned about coding by himself/herself, he/she may have much fewer chances to get a job of software developer if only got a high school diploma. Why employers choose to believe in degrees is probably highly due to the credibility of universities and practical experience during the companies' daily operations that candidates with higher education level can better handle the job than those without.

Also for the positions like lawyers and some healthcare practitioners (e.g.: many kinds of doctor), they have a high annual mean salary while they require candidates obtain a doctoral or professional degree to be qualified for the job, since large amount of professional knowledge is essential especially to these kinds of occupation, which is hard to be systematically acquired outside colleges. Those fields are examples that education level, salary and positions are highly linked together.

On the other hand, however, higher education level is not always necessarily come with high salary. As the case of community and social service occupations, many workers (in some specific positions even over 70 percent) have a bachelor's degree or higher according to the data, yet the salary is much lower than positions in other fields like business and financial operations or computer and mathematical, which has similar distribution of workers' education level. This situation may due to the property of the work field, since social workers often work in NGOs (Non-Governmental Organization)

which are mostly also non-profit organizations, so it largely depends on people's interest and their career goals. But for most of the other occupations, education level is positive correlated to salary, which means people's input of education cost will likely match the output of salary income.

Fulfilling the information need of personas' example

As for considering those three personas mentioned above (on page 14), the visualization can provide some help:

a. for the first one, John, who is not sure what degree is competitive for getting a relevant future career, he can look at the histogram of education level and select his/her interested fields and positions, which visually shows the proportion of different degrees;

b. for the second one, Jane, who is confused by whether to continue pursuing a higher education level while may be restricted by the financial condition of her family, she may then need to take a look at the salary scatter plot about the related field to know more about the compensation situation, and also check the histogram to see what degree is competitive enough to get the ideal position. Other than that, she can calculate and compare the cost of education and income of this occupation's salary to decide if it is worthy to get further education of Master's degree or higher. As a side note, as her study goes further, education cost and salary level might not be the only reasons she would use to make her decision, it may also depend on the classes she takes and how much more she still wants to learn of this field, but nonetheless the information from the visualization can still be helpful;

c. for the third one, Jack, who does not have his interested field yet and has certain financial restriction of his family, he would need to think about whether continuing the four years' education for Bachelor's degree is the best choice for him or find some other solutions instead, since objectively speaking it can be a hard time for him as well as his family to pursue the degree. It largely depends on the potential income of his future job; and since he is not sure about what field he is interested in, he can widely explore around different graphs of the interface to obtain general ideas and have deeper thoughts.

Evaluation

Participants recruitment and obtaining data

This project is mainly for helping university freshmen get ideas about occupations' information and have some idea of their future career, so for the evaluation part I also focused on recruiting first-year students to do the online survey at first. Due to some objective limitations such as being short of time and not having enough sources to get more freshmen involved, I eventually extended the range from only freshmen included to a broader one, with UNC-Chapel Hill students in other years who wants to know more about occupations and is actively looking into his/her interested career fields if applicable.

For the survey, each participant needed to explore around the visualization interface and tried answering the questions. Two types of questions were given, including single-choice and short text of open-ended. Data of the evaluation survey was collected using Google Form, and then converted into CSV file for further analysis.

Feedback from the participants

In general, 12 people who was eligible for the study participated in the survey. For the survey questions, five different simple task questions were given to the participants, which mainly helped testing the efficiency of the visualization. In the following questions of Likert scale and open-end questions, participants showed different opinions which were very helpful. For the time cost of answering the questions, over half of the participants agreed or strongly agreed that it took them a comparatively short period of time (less than 10 minutes) to finish the questions. For the easiness of exploring around the interface and correctness of answers, half of the participants held positive opinions, while the rest of them were either neutral or negative. Those task questions were set to be comparatively distinct and the answers could be easily find out without making too many operations by users. From the feedback, most of the mistakes occurred in those questions related to the scatter plot. In consequent, from the efficiency aspect the interface still needs further improvement of user experience.

For the overall impression of three different graphs, participants were most impressed by the pie chart which serves as the main graph, followed by the histogram and scatter plot. For the easiness of using/understanding the graphs, both the pie chart and histogram got higher positive evaluation while some participants indicated that it was harder to find out the information needed in the scatter plot. By the question of whether the visualization was helpful for getting some general idea of future career choosing, two thirds of subjects agreed or strongly agreed that this interface did give some help.

On the other hand, however, it was interesting to see in the open-end question that almost half of the participants interested in the scatter plot the most, although they found

it harder to use as well as the questions related to it got the lowest rate of correctness; and the pie chart got the most comments of visually appealing among the three graphs. Many people mentioned that the search function should be added in the visualization so that it could be easier to get the information people needed, and also diverse designs of font and color as well as linking contents in different pages could help users have better experience of using the visualization.

Discussion of evaluation results

Testing the efficiency can help prove the usefulness of the visualization interface, since the questions of whether it is useful for getting information needed or easy to explore the interface are subjective and answers vary from person to person, but the correctness of task answers and the time cost are some of the objective proofs of the performance of visualization.

From the result of tasks, most participants could get the information they needed for the questions quickly and correctly, which can help indicate that the graph shows explicitly about the link between salary and positions, and the one between education and positions. The subjects also showed agreement that this visualization helped them get ideas about future career choosing, which has been the original motivation of constructing the project.

Besides, why people have been interested the most in the scatter plot is probably because it provides information about salary of different positions, while it also compares the salary situation within a general field of occupations. Comparatively more information is contained there than the other two graphs, so that it requires a more

efficient and less confusing way to link the data to the graph for display and interaction. According to the evaluation, the idea of drawing a scatter plot is accepted by the participants and it does help people get relevant information and answer their questions from the information need, yet the way of showing data needs further improvement.

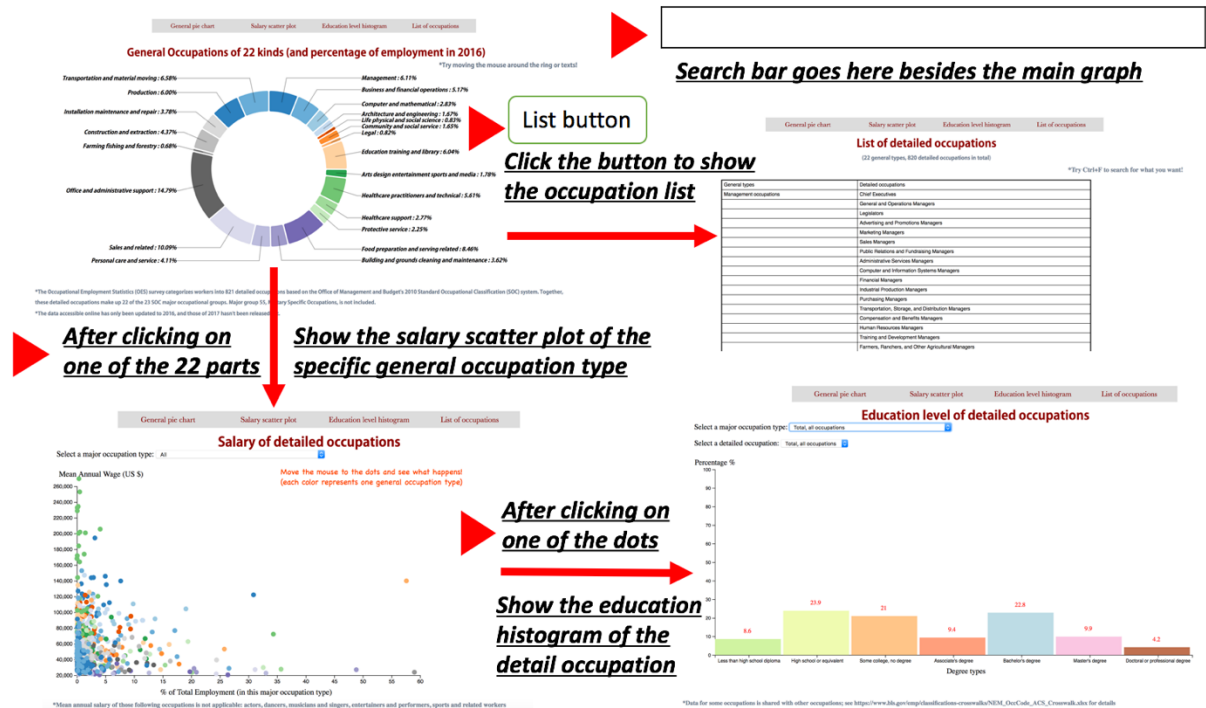
Discussion and conclusion

Future work -- Potential draft of second version

After summarizing the suggestions from fellow participants, the second version of visualization interface (the graph is on the next page) would have those potential updates which can possibly make the interface more helpful:

- a. add a search bar at the top of the page for users to search for either general occupation type or detailed occupation, and the three graphs should be put on the same page, which can give a more general impression and be easier for users to obtain different aspects of information of the same occupation. After clicking on the search button, all the three graphs will be updated to show data of that search term:
 - i) if searching for a general type, then the specific part of pie chart will be highlighted, and scatter plot will show the graph of that type, while histogram will show the first term within the type as default;
 - ii) if searching for a detail occupation, then histogram will show the graph of that type, while the general type which this detail occupation belongs to will be highlighted in pie chart and shown in scatter plot.

- b. add a list button – when users click the button it can show the occupation list to the right of the interface for users to browse through, which can make it easier for the user to search for the position they want instead of jumping from page to page;
- c. link pie chart with scatter plot – after clicking on one of the 22 parts, scatter plot below will automatically show the graph using relevant data, reducing the trouble of repeatedly selecting the same search term from the drop down bar;
- d. link scatter plot with histogram – after clicking on one of the dots, histogram on the right will show the graph using relevant data, also for reducing the trouble of repeatedly selecting the same search term from the drop down bar.



graph 8: draft of second version (updates added)

Lessons learned

Technical aspect

This project is mainly about constructing a data visualization interface using HTML and JavaScript, so during the process my ability of web development and visual analytics improved a lot, getting used to the D3 library and knowing more about JavaScript. I came across quite some trouble when drawing the graphs, such as cleaning and formatting the data file for better usage of the graph, having trouble showing the data in the graph as I wanted, matching the axis with the data, etc. As the project finally finished, I would say I solved most of the questions but it would still take a long way to go on the road of data visualization; and due to the time limitation of the project it is hopefully the best I could do for it.

Study aspect

The study is aimed to provide the visualization as a tool to help college freshmen finding out their potential career interest or get to know more about that if they already have the goal. No matter what kind of visualization is going to be constructed, the first and most important thing is always about the users. Questions can be plenty: which group of people will potentially be the users of this visualization, what information do they need and possibly wish to get from here, what factors of the design of this visualization will effect their user experience, etc. There is a lot to think about, and I have tried my best to cover most of these questions and more concerns during finishing the project so now I understand more about data visualization from both the aspect of users (which is the basic and essential part) and technical.

Limitations

As for the data collection, while using second hand data set there will be certain limitation that some of the required data are missing, or it may not be representative enough to carry out the result. For solving this problem, I tried to combine the fields of available data into different sets of values for visualizing and analyzing, which could hopefully maximize their usage and reduce the restrictions to minimum.

For the technical part of constructing the data visualization graphs, there were some difficulties for coding and finally implementing the design plan of the graphs, and there surely still have many places within the visualization which needs to be improved.

For the evaluation part, the sampling of evaluation has certain limitation, due to the fact that the group of students to be recruited for study may not equally distributed by ethics and genders, which might cause it to be less representative towards a more general conclusion. Besides subjective reasons, this is also one of the limitation of convenient and random sampling.

Also, it is subjective for each individual to regard the usefulness of information resources, so it is hard for the survey questions to cover the whole point, and other factors such as the process of recruiting participants as well as the number of people getting into the evaluation all did impact on the result. Other than the fact that resources were limited while doing the evaluation, yet it is still a way to test whether the visualization is efficient/helpful or not.

References

1. Kinouani, S., Boukhors, G., Luaces, B., Durieux, W., Cadwallader, J.-S., Aubin-Auger, I., & Gay, B. (2016). Private or salaried practice: how do young general practitioners make their career choice? A qualitative study. *BMC Medical Education*, 16(1), 231. <http://doi.org.libproxy.lib.unc.edu/10.1186/s12909-016-0754-6>
2. Fatima, A., Rukhsana, J., Burger, L. (2015). Factors influencing the career choice of undergraduate students at a historically disadvantaged South African university. *Industry and Higher Education*, 29(3), 209-219
3. Yang, L., & Webber, K. L. (2015). A decade beyond the doctorate: The influence of a US postdoctoral appointment on faculty career, productivity, and salary. *Higher Education*, 70(4), 667-687. doi:10.1007/s10734-015-9860-3
4. Peters, R. (2015). Facing Crucial Career Choices. *Pharmaceutical Technology Europe*. 27(12), 18-21
5. Ojeda, Beatriz Sebben, Creutzberg, Marion, Feoli, Ana Maria Pandolfo, Melo, Denizar da Silva, & Corbellini, Valéria Lamb. (2009). Nursing, nutrition and physiotherapy students: career choice. *Revista Latino-Americana de Enfermagem*, 17(3), 396-402. <https://dx.doi.org/10.1590/S0104-11692009000300018>
6. Schulz, J., & Thoni, C. (2016). Overconfidence and career choice. *Plos One*, 11(1), e0145126. doi:10.1371/journal.pone.0145126

7. PERI, G., & SPARBER, C. (2011). Highly educated immigrants and native occupational choice. *Industrial Relations: A Journal of Economy and Society*, 50(3), 385-411. doi:10.1111/j.1468-232X.2011.00643.x
8. Abrahamsen, B., & Drange, I. (2015). Ethnic minority students' career expectations in prospective professions: Navigating between ambitions and discrimination. *Sociology*, 49(2), 252-269. doi:10.1177/0038038514542494
9. Azzam, T., Evergreen, S., Germuth, A. A. and Kistler, S. J. (2013), Data Visualization and Evaluation. *New Directions for Evaluation*, 2013: 7–32. doi:10.1002/ev.20065
10. Kosara, R. (2007, July). Visualization criticism—The missing link between information visualization and art. *Information Visualization, IV'07. 11th International Conference* (pp. 631–636). doi: 10.1109/IV.2007.2
11. Melissa Mallon (2015) Data Visualization, *Public Services Quarterly*, 11:3, 183-192, DOI: 10.1080/15228959.2015.1060147
12. Olsen, Kai A. "Data Visualization." *Computer Sciences*, edited by K. Lee Lerner and Brenda Wilmoth Lerner, 2nd ed., Macmillan Reference USA, 2013. Science in Context, link.galegroup.com/apps/doc/CV2642250148/SCIC?u=unc_main&xid=495a6d14. Accessed 6 Sept. 2017.
13. T. Munzner (2015). *Visualization analysis and design*. CRC Press
14. Kovalerchuk, B. (2017). *Visual Cognitive Algorithms for High-Dimensional Data and Super-intelligence Challenges*. Cognitive Systems Research.
15. Nicoletta Di Blas , Mirjana Mazuran , Paolo Paolini , Elisa Quintarelli , Letizia Tanca, Exploratory computing: a challenge for visual interaction, *Proceedings of the*

2014 International Working Conference on Advanced Visual Interfaces, May 27-29, 2014, Como, Italy [doi>10.1145/2598153.2600037]

16. Pomarede, D; Courtois, HM; Hoffman, Y (2017). Cosmography and Data Visualization, PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, Volume 129, Issue 975

17. Wang Y, White W, Andersen E (2017). PathViewer: Visualizing Pathways through Student Data. Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp: 960-964

18. Logre I, Mosser S, Riveill M (2015). Composition challenges for sensor data visualization. Companion Proceedings of the 14th International Conference on Modularity, pp: 25-26

19. Klomklao T, Ratanarungrong P, Phithakkitnukoon S (2016). Tweets of the nation: tool for visualizing and analyzing global tweets. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct, pp: 1349-1357

20. Devitt, P.(2015). Research essentials: Understanding quantitative data analysis. Nursing Children and Young People, Volume 27, Issue 7

21. Lucko, Gunnar; Mitchell, Zane W (2010). Quantitative Research: Preparation of Incongruous Economic Data Sets for Archival Data Analysis. Journal of Construction Engineering and Management. 01/2010, Volume 136, Issue 1

22. Marron, J. S. and Alonso, A. M. (2014), Overview of object oriented data analysis. Biom. J., 56: 732–753. doi:10.1002/bimj.201300072

23. Shruti Kohli and Himani Singal. 2014. Data Analysis with R. In Proceedings of the 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing (UCC '14). IEEE Computer Society, Washington, DC, USA, 537-538.
DOI=<http://dx.doi.org/10.1109/UCC.2014.81>
24. Danyel Fisher. 2016. Big data exploration requires collaboration between visualization and data infrastructures. In Proceedings of the Workshop on Human-In-the-Loop Data Analytics (HILDA '16). ACM, New York, NY, USA, Article 16, 5 pages.
DOI: <https://doi.org/10.1145/2939502.2939518>
25. Peter Norvig. 2011. Internet scale data analysis. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '11). ACM, New York, NY, USA, 2-2. DOI: <https://doi.org/10.1145/2020408.2020412>
26. Rahul C. Basole, Mark L. Braunstein, and Jimeng Sun. 2015. Data and Analytics Challenges for a Learning Healthcare System. *J. Data and Information Quality* 6, 2-3, Article 10 (July 2015), 4 pages. DOI: <https://doi.org/10.1145/2755489>
27. Richard A. Becker and John M. Chambers. 1984. Design of the S system for data analysis. *Commun. ACM* 27, 5 (May 1984), 486-495.
DOI=<http://dx.doi.org/10.1145/358189.358078>
28. Tova Milo and Amit Somech. 2016. REACT: Context-Sensitive Recommendations for Data Analysis. In Proceedings of the 2016 International Conference on Management of Data (SIGMOD '16). ACM, New York, NY, USA, 2137-2140. DOI: <https://doi.org/10.1145/2882903.2899392>
29. Shyr, J. and Spisic, D. (2014), Automated data analysis. *WIREs Comput Stat*, 6: 359–366. doi:10.1002/wics.1318

Appendix

Evaluation survey

Part 1 -- Tasks time!

Following are some simple questions in order to help you explore around the visualization, each of them only have 1 correct answer and will take mostly just a few minutes to find out the answer; and the answer will be provided below.

1. Q1: How many kinds of major occupations are there in total? *

Check all that apply.

- ☐ 12
- ☐ 22
- ☐ 32

2. Q2: What is the mean annual wage of lawyers in 2016? *

Check all that apply.

- ☐ 95,000
- ☐ 120,210
- ☐ 139,880

3. Q3: Within the major occupation type of Healthcare Practitioners and Technical Occupations, how many detailed occupations have the mean annual wage more than 200,000 dollars? *

Check all that apply.

- ☐ 9
- ☐ 15
- ☐ 20

4. Q4: What is the percentage of total employment of Computer Systems Analysts? *

(within the major occupation type of Computer and Mathematical Occupations)

Check all that apply.

- ☐ 10.5%
- ☐ 13.7%
- ☐ 25%

5. **Q5: Which occupation prefers to hire people with bachelor's degree or higher education levels more, Sales Engineers or Retail salespersons? ***

(within the major occupation type of Sales and Related Occupations)

Check all that apply.

- ☐ Sales Engineers
☐ Retail salespersons

Answers

Q1: 22

Q2: 139,880

Q3: 9

Q4: 13.7%

Q5: Sales Engineers

Part 2: Questions about the process of finishing the tasks above

6. **1. It took me not too much time (less than 10 minutes) to finish all the tasks ***

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

7. **2. I found it easy to explore around different sections of the graphs and find the information I need**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

8. **3. I answered all the questions correctly**

Mark only one oval.

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

Part 3: General questions about the visualization interface

1. How would you rate the overall impression of the following graph sections? (5 as the highest and 1 as the lowest) *

	1	2	3	4	5
Main graph (pie chart)	☐	☐	☐	☐	☐
Scatter plot	☐	☐	☐	☐	☐
Histogram	☐	☐	☐	☐	☐

2. Did you find the sections easy to use/understand? *

	1	2	3	4	5
Main graph (pie chart)	☐	☐	☐	☐	☐
Scatter plot	☐	☐	☐	☐	☐
Histogram	☐	☐	☐	☐	☐

3. Did you find the visualization helpful for getting some general idea of future career choosing? *

	1	2	3	4	5	
Strongly Disagree	☐	☐	☐	☐	☐	Strongly Agree

4. Overall, what part(s) do you like the best / interests you most in this visualization?

Your answer

5. What part(s) do you think needs to be improved?

Your answer

6. Any suggestions of how to improve the visualization?

Your answer