

Visualization Project Proposal

Roy Bastien, Priyanka Parekh

October 23, 2015

1 Basic Info

Title: Visualization of Skill, Salary and Location information

Partner 1: Roy Bastien, u0155098, u0155098@utah.edu

Partner 2: Priyanka Parekh, u0756881, u0756881@utah.edu

Project Website: <https://github.com/roybastienuofu/dataviscourse15-finalproject/>

2 Background and Motivation

The idea behind our project is visualizing the predicted salary for a job based on keywords in the job description.

From our experiences, the majority of the time, job listings are not accompanied by estimated salaries, forcing the job seeker to visit external resources or ask others for information regarding their future salary.

Additionally, individuals often times may consider how salaries for certain skills vary across regions or how developing certain skills will benefit them financially. Our tool will allow for users to enter a skill, for example “Java” and then be able to see what the associated salaries are for that skill. Our decision to pursue this project is based on our personal frustration when searching for jobs and not being able to immediately garner the salary associated with said job.

The initial problem, of predicting salaries based on keywords will be addressed as a part of our project for our machine learning class. We will use the Glassdoor API to collect job descriptions and parse these descriptions for keywords. And then we will create a training set based on salary data that we will also pull from Glassdoor and use that training set to predict future salaries based on provided keywords.

While the work will be doing in ML is very interesting, in our opinion a key extension of this project would be enabling users to query our data set and interact with it. That is why we want to visualize this particular data set as our project for Visualization.

3 Project Objectives

- Visualize salaries associated with particular skills

- See salary variances across regions
- Understand relationships between skill sets and job titles

4 Data

We will be utilizing the Glassdoor API to access their data set. Use of their API simply required the creation of a Glassdoor account and a short approval process which we have already gone through. We will be able to make API calls to see job descriptions as well as salaries. Here is a link to the API overview: <http://www.glassdoor.com/developer/index.htm>

5 Data Processing

Because we want to visualize the "learned" data set we do expect a substantial amount of data processing however that will be completed as a part of our machine learning course. Through the use of the Glassdoor API, and web scraping (if the accessible data is not enough), we should be able to garner enough data to predict a salary based on a feature(skill). The results from our work in for our machine learning course will serve as our input data set for this course.

In the unlikely event that our ML project is unsuccessful we will simply use existing data on Glassdoor to manually associate skills and salaries and visualize this aggregated data set. In this scenario the data acquisition may be time-intensive but we do not believe that data cleaning will be a major concern.

6 Visualization Design

In the following pages you are able to see what we came up using the Five Design Sheet Methodology. You can see below that our choices for visual encoding mainly consist of bar graphs and maps. Due to the nature of our data we believe that these two choices are easy to understand but also aesthetically pleasing. And as was mentioned in class, if a bar chart can properly tell the desired story, then it is an ideal choice.

Color and size will play a large role in our final visualizations for the sake of differentiation and also showing prevalence/magnitude. Users will interact with our page via both the initial search function and via interactive visual encodings. Further details regarding our approach, and more specific considerations can be found below.

Sheet 1

How to search - dropdown w/criteria. Search by:

Job title	Key word
Company	Industry
State	
City	

Display Results: location on map

top company names

jobs available on Glassdoor

job progression -

what jobs did they get next
Salaries of job progression.

jobs within a certain radius

available jobs within a company
Salaries of available jobs

Salaries - national averages
industry averages
title averages

Title: Employment & Salary Visualization
Authors: Roy Bashir & Prayanka Parekh

Date: 10/22/15

Sheet: 1

Sheet 2 Layout

Display's



Focus

Should the visualizations change based on search category? If a state is selected the map should only display that state. If a city is selected, there may not be a map needed.

Title: Employment & Salary Visualization.

Authors: Roy Bastien & Priyanka Parekh

Date: 10/22/15

Sheet: 2

Operations

User selects category from drop box (ie company, state, industry) and enters a search term.

Company

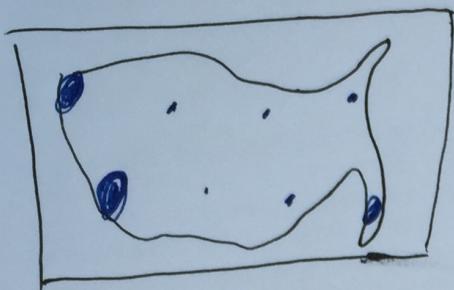
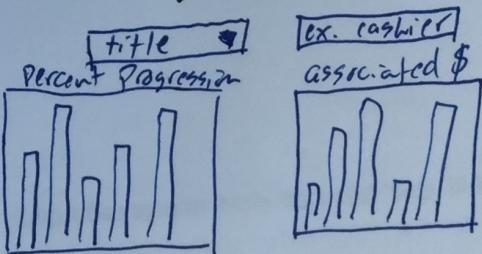
Amazon

Discussion

A dynamic initial visual layout that changes for each search category will be extra work to implement but will add an additional layer of interaction with the user.

Map and bar graphs will be the main display mediums.

Sheet 3 Layout

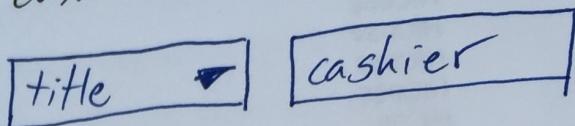


Focus

Within the map, dots change size based on available positions across the country. Progression shows the jobs the title gets next along with associated salaries in a separate(?) bar chart.

Operations

User selects search category from drop menu. The text box displays an example from the category and the visualizations customize for the category



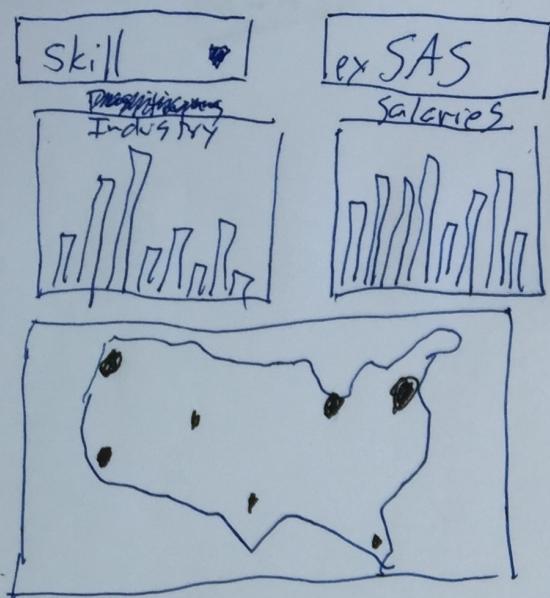
Discussion

Data should be populated in the charts before the user searches for a term. Display data should match the example given in the text field.

Visualizations will change based on search category.

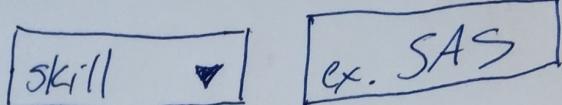
Title: Employment & Salary Visualization
Authors: Ray Bastien & Priyanka Parekh
Date: 10/22/15
Sheet: 3

Sheet 4 Layout



Operations

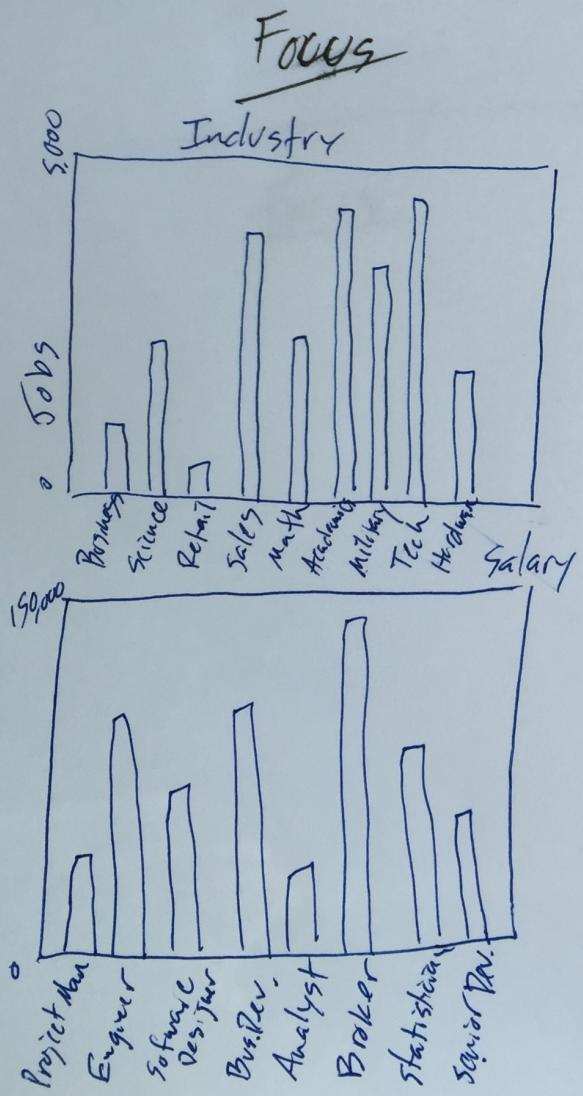
User selects the skill from a dropdown menu and enters text into the search field.



The display examples auto fill with data and predicted job titles.

Discussion

Our algorithm will search job descriptions with the keyword and return the 10 highest paying titles with the entered keywords along with the salaries to display. Response time will be a challenge to effectively engage the user.



Title: Employment & Salary Visualization
Authors: Roy Bastien & Priyanka Parekh
Date: 10/22/15
Sheet: 4

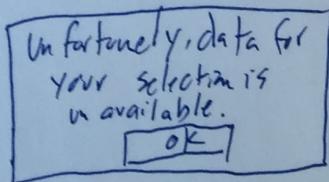
Sheet 5

Layout



Focus

If data is unavailable for a given input, an alert box will be displayed in the center of the screen with an "OK" selection for the user to press to dismiss.



Operations

The final user inputs will consist of just the dropdown menu and a text field. If data is available for the selection, it will be displayed in the plots. Initial data will be displayed in the plots and selections will remain in the plots if user data is unavailable.

Detail

D3.js will be the software used for creating the majority of the necessary choices for visualizing the data. We will need to create a novel algorithm for predicting the industry and job based on skill input. We will use a JSON format to interface between the model, view, and controller.

Title: Employment & Salary Visualization
Authors: Roy Bashir & Priyanka Parekh
Date: 10/22/15
Sheet: 5

7 Must-Have Features

- Allow users to enter a skill. Proceed to show the user information regarding the salaries associated with that skill across regions in a bar chart, or in a map with points that represent different cities, or most likely both. Our final decision on representation mediums will be contingent on the content of our data set. Based on our data set we will choose the medium that best conveys information to our user.
- Our final visualization must have an interactive component. This will provide users with more information on hover or click which will serve to further engage the user.

8 Optional Features

- Provide auto-complete functionality in our drop down menus.
- Understand relationships between skill sets and job titles. Allow users to see what salaries are associated with the job titles that require a specific skill. Potentially have position descriptions to accompany the job titles.
- Visualize the number of jobs associated with that skill. Represent this on a map with the size of a point indicating the prevalence of jobs.

9 Project Schedule

Below is the current plan in terms of dates and self-imposed deadlines. We chose not to have a division of labor, rather we plan to work on all milestones together. Therefore the milestones below represent what we aim to achieve as a team on a weekly basis.

Table 1: Project Schedule

Date	Milestone
10/23	Project proposal due
10/30	Complete data acquisition
11/6	Have data structures in place. Begin work on visualization prototypes.
11/13	Project milestone due. Have working visualization prototype.
11/20	Transition from visualization prototype to predetermined final design. Ensure that all individual elements are in place.
11/27	Have initial interactivity in place.
12/4	Final Project Due. Ensure that all elements are fully functional and that the desired level of interactivity is in place.