INFO 201 Group Project Proposal

By Alberto Melendez, Daniella Mesler, Erin Chang, Nam Pham (Team Dane)

Project Description

We will be working with the occupational employment and wage data in U.S. obtained through the Bureau of Labor Statistics website, which publishes a large amount of information by occupation, including career information, employment levels and projections, and data on earnings and working conditions. These estimates are calculated with data collected from employers in all industry sectors in metropolitan and nonmetropolitan areas.

Our targeted audience will be university students seeking employment. We anticipate the students to be a lot more interested in future trends in employment since this particular group still has a long journey to go in terms of professional life. We expect these data to provide them insights into each occupations and industries through outlining wages, job stability, future trends, and geographical representation. The data may assist graduating students in making informed decisions about their future employment, and it can also help students without a major to be more mindful of what they want to study.

Our project will be able to answer these specific questions:

- What are the highest paying occupations related to a particular major?
- What are the top 10 states for a particular occupations in terms of number of employment?
- What are the fastest growing occupations or industry in ten years?

Technical Description

The format of our final product will be a Shiny app. The platform entails a map that display the whichever data of interest (occupations, wages, etc.) that user wish to see and graphs that demonstrate current situation and future trends for wages, occupations, industries, or number of employment, etc.

We will be using multiple .csv datasets containing labor statistics information for many different occupations and industries. Therefore the type of data-wrangling we will need to do to our data includes selecting the data we actually want: joining tables, and

filtering. Some libraries that we will be using in this project includes: ggplot2, jsonlite, stringr, shiny, slackr, etc.

We can use statistical analysis to answer questions about which field provides the highest-paying jobs or what are the earning discrepancies between each field. In order to answer this question we have to group occupations into various field (and maybe even machine learning to build a predictive model about the growing trend of the field) and then use statistical analysis in order to compare those fields together.

Some major challenges we anticipate are those related to manipulating and representing the data. There may be issues with the accessing data, resulting in the data being unusable. There are also concerns about possible datasets with insufficient information that prevents the platform from having optimized filtering feature. Also, job data will most likely come in large number, so there will be questions concerning how to effectively extract the data of interest. In addition, there will be concerns about making the platform and visualization clear and intuitive so that the users can observe the data effortlessly and efficiently. Lastly, we expect challenges with working collaboratively through Github since pushing codes have always been a task that can easily go wrong even in individual project.

Project Setup

As group we will decide we want to focus on and complete the data manipulation steps necessary to obtain the data and visualizations we need to answer the specific questions above. We will also begin designing the interface and decide who will be in charge of which feature.

GitHub Issues:

1. Do API request and import JSON data into usable format
   with data of interest - assigned to Erin
2. Create a README and write the description for this project - assigned to Dani
3. Do statistical analyses on the dataset to determine insights, major trends, etc. - assigned to Nam
4. Perform user research on our targeted audience to assist in platform and feature design - assigned to Alberto
5. Design the user interface and create wireframe for menu and other pages - assigned to Erin