Write-up: Team BrunchLadies

Our team focuses on visualizing the job searching and job posting statistics in the United States, based on the data from indeed.com. Our product is a Shiny app, which provides several kinds of results: the annual statistics, the dynamic trends, and a Gaussian model.

At the first glance of dataset by using ggpair, we found that the location and industry are the two most important factors in influencing clicks for a specific job. The date of post, on the other hand, influences the average clicks on that day. We’ll visualize these in a second. (Time series model of website clicks. Clearly there is a yearly trend inside, which we could fit a ARIMA/ seasonal ARIMA model if enough data is provided)

Due to the extreme size for the original data, we preprocessed the data by sampling a subset of data based on their job id (500k), which is supposed to be unique. After filtering out outside-state data, we then calculate statistics for “median salary”, “total clicks”, “local/total click ratio” for each states/industry.

We imported geo-coordinate data for states and cities in the U.S. in order to visualize the map. In the Shiny app, you can choose which variable you want to see, and upon clicking on the circles you can get the specific statistics for that city. The results and corresponding insights for some of the plots that we believe are interesting are included below for overall statistics:

1. (Overall local click ratio) a high ratio indicates that this city may be a relatively closed-market. People tends to looks for job locally. More developed cities in general show low local/total click ratio, indicating more frequent migration towards those cities.
2. (Tech-software 2 plots) Computer Science is still the best major to make money and find a job. Tech-Software industry has a very high salary across the states, and the click (potential supply on the current market) is not high, indicating potential opportunities for students.
3. (Job posting by states – monthly trend) Job posting: CA and TX have a lot of opportunities.
4. (Job posting by industry – monthly trend) November to January: increased job posting; July to October: more clicking - more candidates looking for job. There is a time lag between supply and demand which could be fixed.
5. (Gaussian model of average clicks across job age) We guess that indeed has an algorithm which prioritizes new job and old jobs that are about to be removed. Looking at the average clicks over job age, we can see this fits a bathtub curve. This is interesting, and we’d like to explore the mechanism behind it if we have extra time.