

Decoding the American Dream: Visual Analytics of Job Market in the U.S.

Jingheng Huan Zhiyun (David) Lu Yantao Mei

Duke Kunshan University STATS 401



Introduction

This report delves into the U.S. job market's intricacies using visual analytics, sequentially analyzing industry, company, individual, and geographic levels. Initially, our tree map visualization showcases the industry landscape by mapping companies to sectors based on employee size. Subsequently, a radar map dives into corporate performance, comparing key financial metrics among S&P 500 companies. The Ridge Line plot then transitions to an individual focus, examining the relationship between education levels and salary distributions. The word cloud further analyzes individual prospects by identifying job titles with potentially higher salaries. Lastly, our bubble map geographically contextualizes job opportunities, guiding job seekers towards cities and states with plentiful employment prospects. Through this logical analytical progression, we aim to provide a nuanced understanding of the U.S. job market, assisting international talents in making informed career decisions in their pursuit of the American Dream.

Dataset

In pursuit of a comprehensive analysis, we have sourced two principal datasets from Kaggle and used several other APIs for context and additional information.

- **Glassdoor dataset** provides a wealth of information on salary distributions, education level, and employer reviews.
- **LinkedIn dataset** offers an almost exhaustive snapshot of 15,000+ job postings across the United States within a 48-hour span.
- **Google Map API** provides precise geographical coordinates of U.S. cities, thereby enriching the 'location' attribute in our analysis.
- **Yahoo API** provides financial metrics related to companies listed in the S&P 500.
- **Geojson** provides U.S. county map

Visualizations

Tree Map

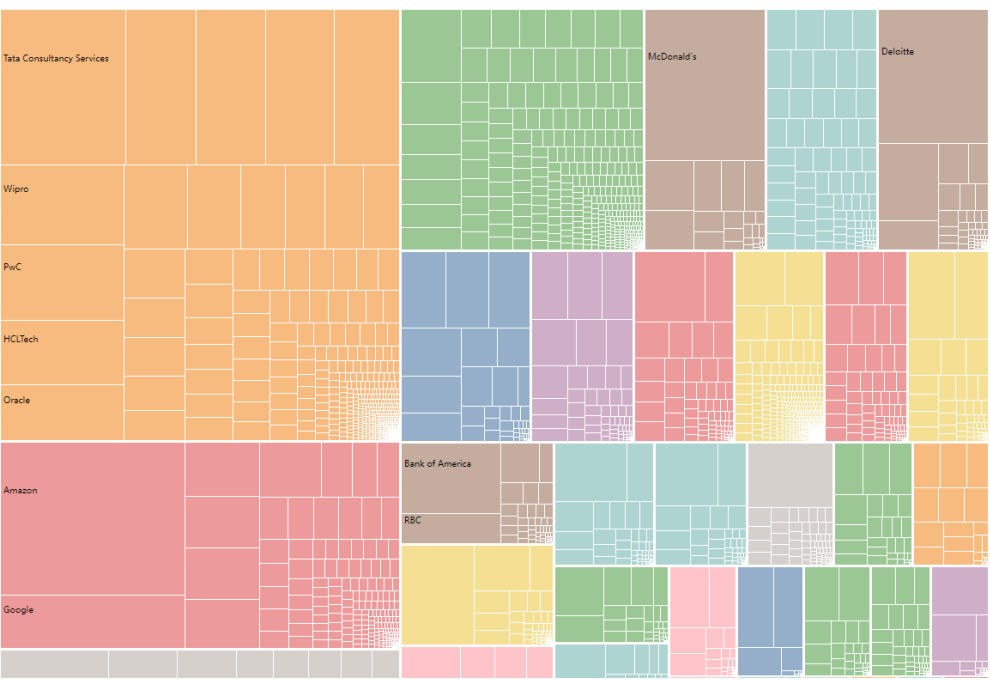


Figure 1. Tree Map

Utilizing the LinkedIn dataset and processing it with Tableau, we crafted a tree map visualization to depict companies and industries by employee size, mapping unique IDs in companies.csv to respective industry and employee count data. We handled multiple entries for a company in employee_counts.csv by considering the highest employee count, acknowledging possible company changes within the year. Executed using D3.js, the interactive tree map enables users to select specific industries, triggering a re-draw to focus on the selected industry's subtree. Company names annotate each cell, with tooltips providing industry and employee count details. Despite successfully meeting its objective, the visualization faces a limitation of overplotting when zoomed into specific industries. This acknowledged shortcoming notwithstanding, the tree map provides a valuable perspective to explore the magnitude and scope of various industries and companies, illuminating the landscape of employee distribution across different sectors.

Radar Map

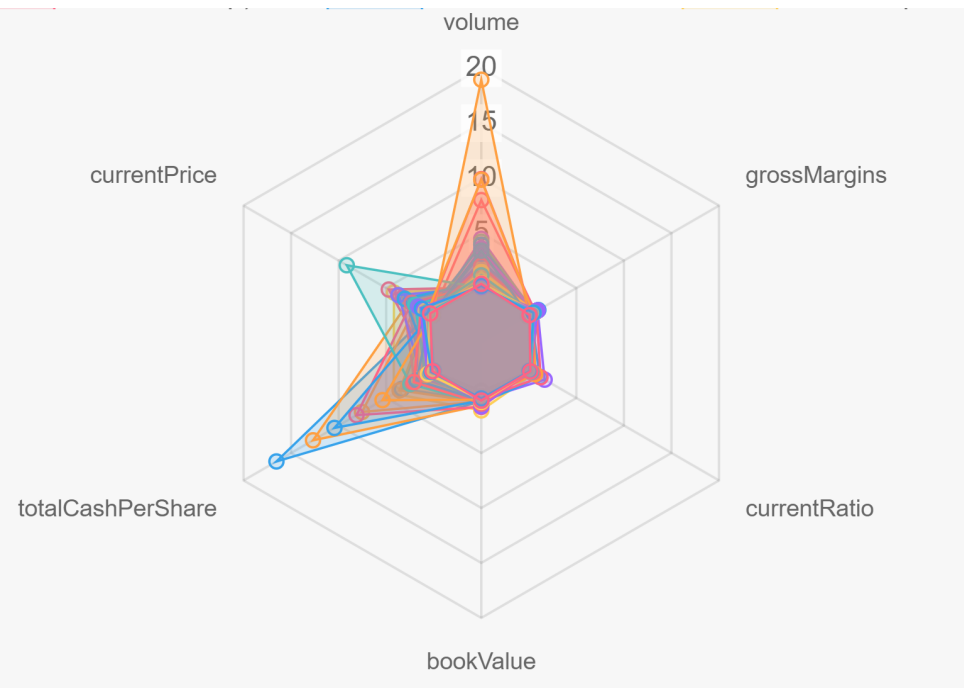


Figure 2. Radar Map

We devised a radar map visualization to delve into corporate performance within the U.S. job market, utilizing data from the Yahoo API and LinkedIn, focusing on S&P 500 listed companies. This visualization showcased key financial metrics like volume, current price, total cash per share, book value, current ratio, and gross margin, offering a holistic glimpse into a company's financial health. A dynamic filter in the visualization allows users to compare multiple companies' financial performance simultaneously. On selection, hues and scales adjust for optimal readability. Tooltips provide instant context for each financial attribute. Our analysis spotlighted notable trends, such as NVR Inc's superior performance across nearly all financial metrics, while Bank of America led in trading volume. These findings serve as a robust resource for potential investors and job seekers, providing a lens into the financial stability and growth potential of prospective employers in the U.S. job market.

Ridge Line Plot

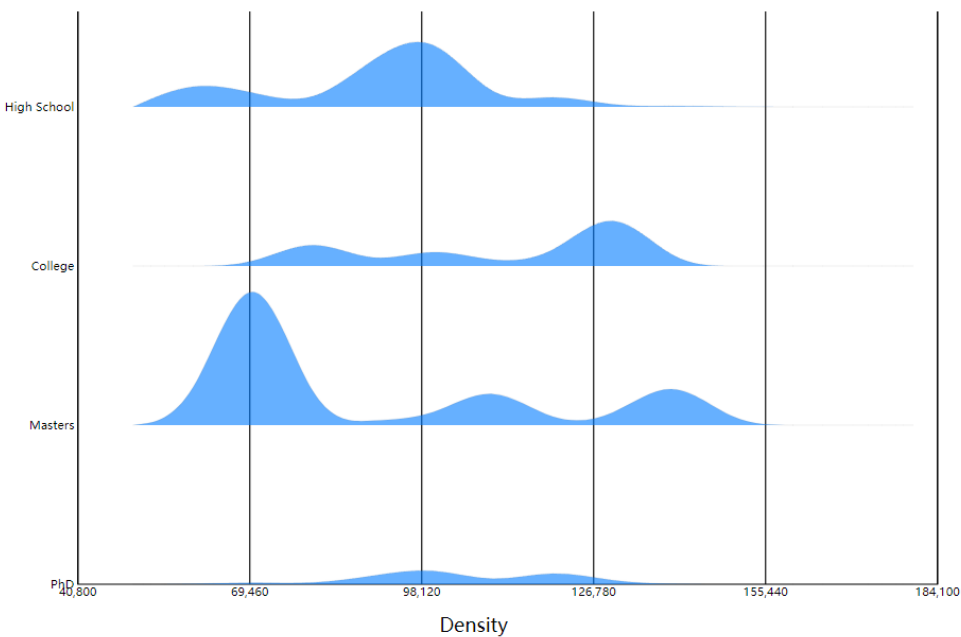


Figure 3. Ridge Line Plot

Utilizing the Glassdoor dataset, we crafted a ridge line plot to probe the correlation between salary and education levels in the U.S. job market, focusing on education levels from high school to PhD. Surprisingly, the salary distributions across education levels weren't as distinct as expected due to factors like the varied academic status of 'PhD' label holders and longer work tenures of those with lesser education. The plot is interactive with a focus on Kernel Density Estimation (KDE), allowing users to tailor the analysis by adjusting KDE parameters—kernel value and smoothing parameter—via sliders. While Glassdoor is a reliable source, the self-reported nature of data like age, salary, and education level can inject variability, possibly affecting the analysis's robustness. This visualization thus not only reveals intriguing salary distribution insights but also highlights the importance of data accuracy in deriving meaningful conclusions.

Word Cloud



Figure 4. Word Cloud

In our analysis of the U.S. job market, we leveraged a Kaggle LinkedIn dataset, focusing on job titles and their frequencies, specifically within full-time positions. Our word cloud visualization aimed to highlight the top 30 job titles on LinkedIn. We incorporated a filter allowing users to sort the word cloud by "frequency," "maximum salary," "median salary," or "minimum salary," enhancing user interaction. Upon selection, a hoverboard feature displays detailed attributes of the chosen job titles, adding value to the visualization. Notable findings include the high frequency of "Sales Director" and "Owner and Operator" titles, with "Quantitative Developer" positions showing the highest salary range. Predominantly, sectors like Engineering, FinTech, and managerial roles were not only common but also had competitive salaries, providing valuable insights for job seekers and industry analysts navigating the U.S. job market.

Bubble Map

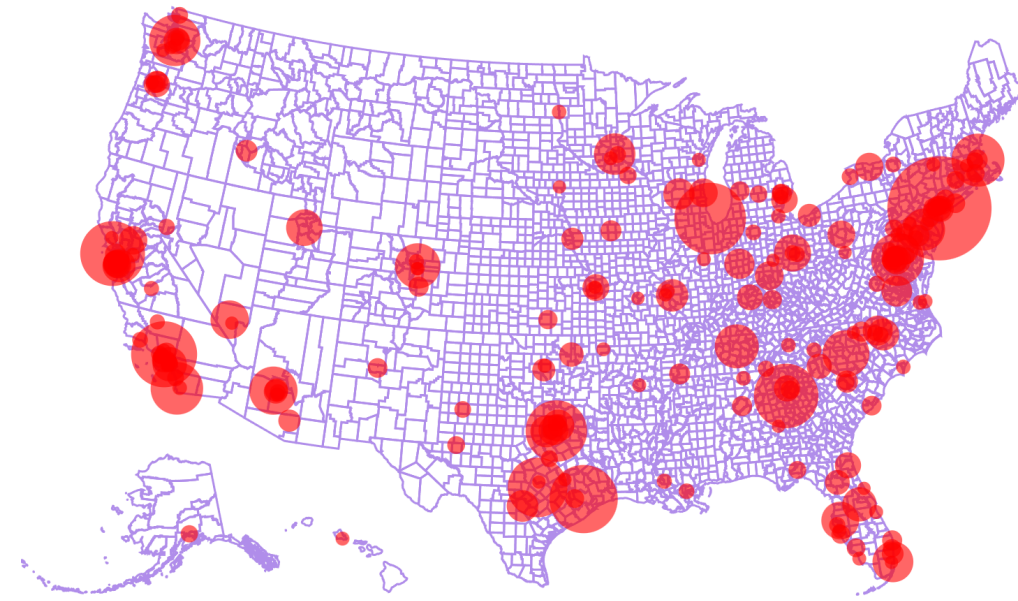


Figure 5. Bubble Map

We created a bubble map visualization using the LinkedIn dataset to geographically contextualize U.S. job opportunities, focusing on the 'location' attribute. Ambiguous locations like "US" were excluded, while similar areas were consolidated, like merging "NYC" with "NYC Metropolitan area." A U.S. county map from geojson and geographical coordinates from the Google Map API enhanced the geographical accuracy. The visualization, interactive in nature, features a filter to categorize cities by job opportunities: "All," "0-50 jobs," "50-150 jobs," and "150+ jobs." A hoverboard feature further provides detailed city information. The map highlights job opportunity concentrations in major cities like New York, Chicago, and Los Angeles, offering crucial insights for job seekers and policymakers on the U.S. job market landscape.