JobQuest - Process Book

1. Basic Info

Project Title: JobQuest: Interactive Visualization of Job Opportunities in the USA

Group Members:

- Vidhi Agarwal, Email: vidhi.agarwal@utah.edu, UID: u1448738
- Pavitra Mehra, Email: pavitra.mehra@utah.edu, UID: u1472969

Project Repository:

https://github.com/dataviscourse2024/group-project-jobquest-visualization

2. Overview and Motivation:

The tech industry in the USA is booming, with constant demand for skilled workers across various roles such as software engineers, data scientists, and cybersecurity experts. However, understanding which specific skills are in demand, which regions offer the most job opportunities, and what salary trends look like can be overwhelming due to the sheer volume of job postings and industry data available. This project aims to visualize tech job opportunities in an interactive way to make this data more accessible.

We chose this project because of our personal and academic interest in data visualization and the tech industry. As aspiring software engineers, understanding the dynamics of the tech job market will not only benefit us but also provide insights to other job seekers, recruiters, and policymakers. Additionally, visualizing this data will help us hone our skills in working with real-world data, modern web frameworks, and interactive visualization techniques.

3. Project Objectives:

This project seeks to answer the following primary questions through interactive visualization:

- What tech job roles are most in demand across the USA?
- How do job opportunities vary across different states and cities?
- What are the salary ranges for various tech roles across different regions?

4. Dataset

- **Kaggle Dataset:** We discovered a dataset on Kaggle scraped from LinkedIn for the years 2023-2024, containing over 124,000 job postings. We can utilize this dataset to complement the real-time data collected through APIs or web scraping. By leveraging this dataset, we can fill in any gaps in data from current job postings.
  The dataset can be found here: LinkedIn Job Postings (2023 - 2024)
  We will be mainly using postings.csv for this task.

- **BLS Dataset:**
  To analyze current trends in job opportunities, unemployment rates, and layoffs, we are utilizing the Bureau of Labor Statistics (BLS) dataset. This publicly available and regularly updated dataset provides comprehensive insights into the U.S. job market, allowing us to visualize trends and shifts in employment across the nation.
  For finding the dataset we used the following website
  https://www.bls.gov/data/home.htm
  We are using the dataset from the employment, unemployment and employment projections sections of BLS to visualize the current trends in the job market.

5. Data Processing:

- **Preprocessing of Kaggle Dataset:**
  - Data Loading: The dataset (postings.csv) was loaded, keeping only relevant columns like job_id, company_name, title, max_salary, location, normalized_salary etc.
  - Handling Missing Values:
    - Rows with missing values in the company_name column were dropped.
    - Rows with missing values in the state column (derived from the location preprocessing step) were also dropped, as they are required for state-level visualizations.
  - Location Preprocessing:
    - The location column contained entries with varying formats, including cities, states, and countries.
    - A mapping of U.S. state codes and full names was used to standardize the state values.

        - Locations containing "United States" or "USA" were mapped to the corresponding state code using a preprocessing function, while maintaining information about the city and country_code.
    - Job Count per State: A derived column, total_jobs_in_state, was created by counting the number of job postings (job_id) for each state using a group-by operation on the state column.
    - Top Job per State: A new column, top_job_in_state, was generated by identifying the most common job title (title) for each state based on the number of job postings. This was achieved using a group-by operation on the state and title columns, and finding the most frequent job title for each state.
    - Salary Outlier Removal: Outliers in normalized_salary were filtered using the Interquartile Range (IQR) method.
    - Industry Column: The postings.csv was joined with companies.csv using the company_id column to enrich the job postings data with the industry column. This allowed analysis of job postings by industry, enabling industry-specific visualizations and insights.
    - Final Data: Unnecessary columns were removed from the DataFrame, keeping only the fields required for analysis and visualizations, such as job_id, company_name, title, normalized_salary, city, state, country_code, total_jobs_in_state, and top_job_in_state.
    - After preprocessing we have approx 100k rows.

- **Preprocessing the BLS Dataset**

    - **Data Loading**: We selected and loaded three key files from the BLS dataset: jobrate.xlsx, unemploymentrate.xlsx, and occupationmatrix.csv.
    - **Filtering for Tech Roles**: In the occupation matrix, we filtered for tech-related roles using keywords such as *software*, *data science*, and *computers*. These were then merged into a consolidated occupation matrix covering relevant roles.
    - **Attribute Selection**: For the occupation matrix, only the most relevant attributes were retained:
        - *Industry Title*
        - *2023 Percent of Industry*
        - *Projected 2033 Percent of Industry*
        - *Projected 2033 Employment*
    - Entries with an industry percentage below one were filtered out as they dont hold any meaningful data.

- **Unemployment and Job Rate Indexing**: For the unemployment and job rate data, we indexed records by month, enabling visualizations with the month on the x-axis to show trends over different years.
- **Attribute Renaming**: Attribute names were standardized and renamed for ease of coding and interpretation.

6. Design Evolution:

Ideas:

We aim to create an intuitive, interactive visualization that allows users to:

- Explore job opportunities geographically (e.g., job density on a US map).
- Compare salary ranges and job demand for different roles.
- View trends in job postings over time.

Idea 1: Choropleth Map with Pop-ups

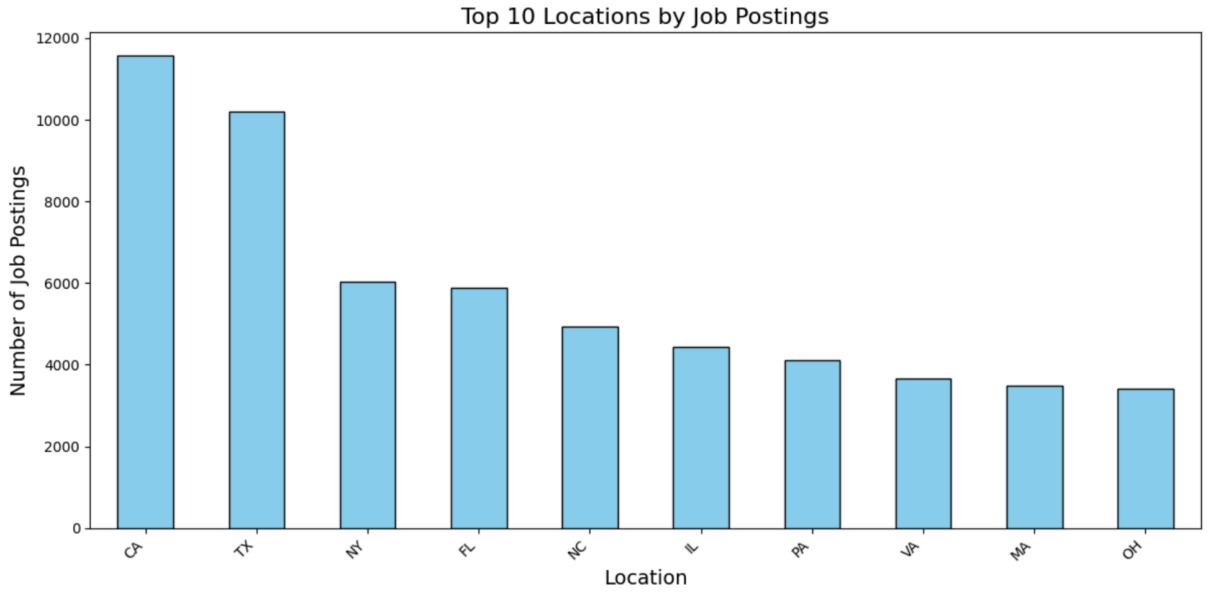
We have decided to add a choropleth map of the U.S., color-coded by the density of tech jobs in each state. The darker the shade, the more job opportunities in that area. Users can hover over states or cities to view a pop-up with more detailed information (e.g., job titles, salary ranges, number of job openings). There will also be filters to narrow down the data by job type, role, or salary.

This design leverages geographic encoding to visually represent data density, making it intuitive for users to identify high-concentration areas for tech jobs. It's particularly effective for users interested in location-based job opportunities.

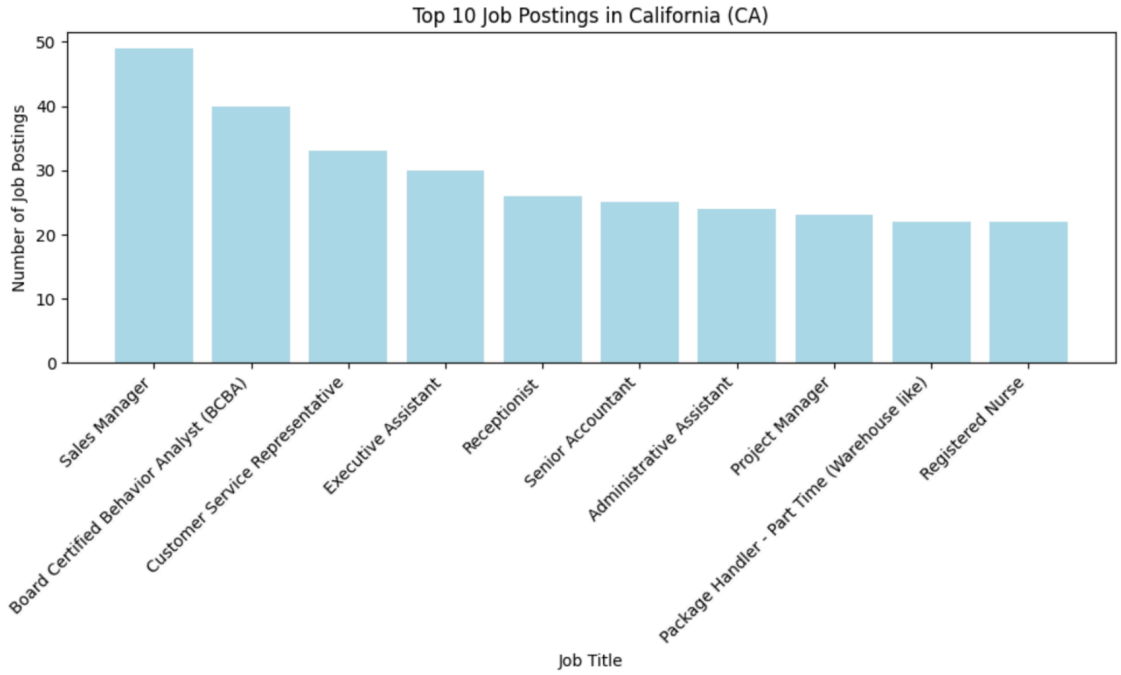
Idea 2: Bar chart

We thought of adding a bar chart as it provides a clear and effective way to visualize job data across various dimensions. Bar charts are particularly useful for comparing categorical data, such as job roles, locations, or salaries. Through our exploration, we explored multiple types of bar charts to address different insights.

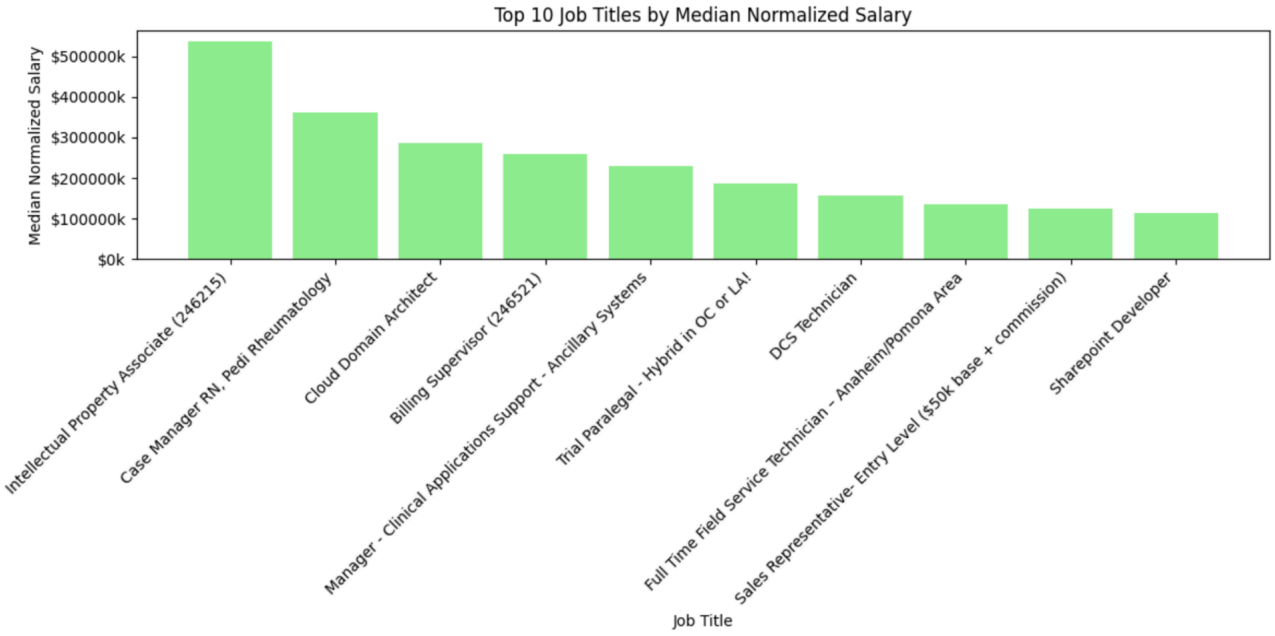
Bar Chart 1 (Job Postings by Location): This bar chart visualizes the top 10 locations with the most job postings. We realized that it would give users a quick overview of the states or cities with the highest demand for jobs, offering an at-a-glance understanding of where the most opportunities are. This design also serves as a complementary view to the choropleth map by quantifying job density.



Bar Chart 2 (Top Jobs in a State): We thought of creating this bar chart to display the top job roles within a specific state, such as California. The idea is to offer a deep dive into what roles are most in demand in a selected state. We plan to add interactivity, such as a dropdown or toggle, allowing users to dynamically change the state and explore the top roles in different regions without refreshing the page.



Bar Chart 3 (Median Normalized Salary by Job Title): We found that visualizing the top 10 job roles by median normalized salary would help users quickly identify high-paying roles. This bar chart offers a direct way to compare the salary distribution across roles, which can help users or job seekers prioritize specific positions. It also highlights which industries offer competitive pay.



Idea 3: Bubble chart

For the occupation matrix, we explored different visualization options to compactly represent projected growth rates for various tech job roles in the coming years. Initially, we considered using pie charts and spider plots to display percentage growth and specific job details. However, we determined that a bubble chart would better serve our goals, as it allows for both comparison and individual insights. In a bubble chart, each job role can be represented by a bubble, with size indicating the growth rate. Additionally, interactive features like hovering would enable users to view specific job details, making this approach well-suited for our data and visualization needs.

Idea 4: Line Plot

To visualize trends in the current job market, specifically the number of job opportunities and the unemployment rate, we chose a line plot as it effectively illustrates changes in values over time. We initially explored the Kaggle dataset to check if it had enough

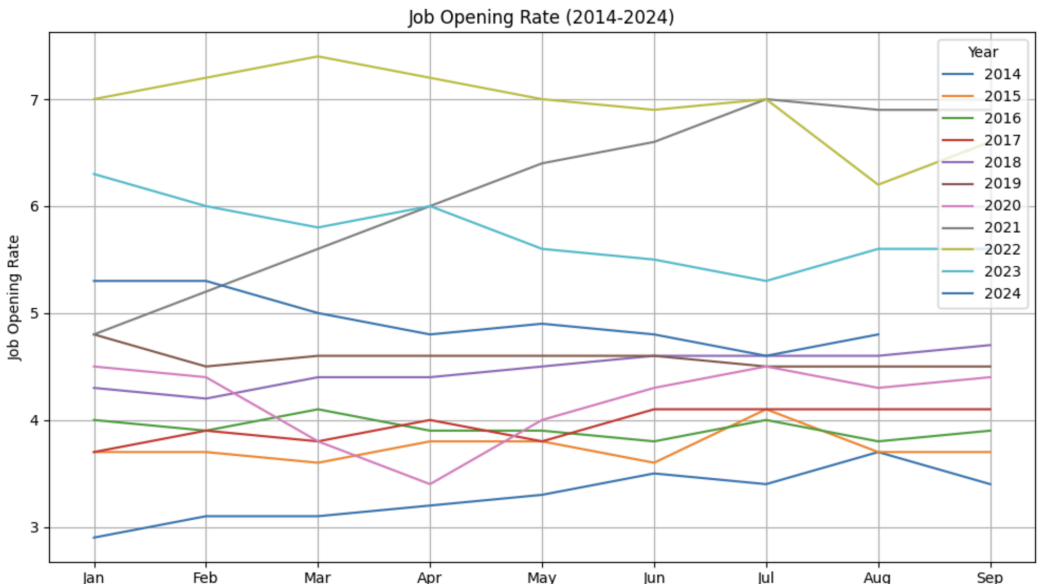
time-based data for creating a meaningful line plot, as it included a timestamp column. However, after further exploration, we discovered that the dataset primarily covered data for a single month, limiting its usefulness for visualizing trends over time. To address this, we expanded our search and found more suitable datasets, including the BLS data, which provided a more comprehensive time series for analyzing trends effectively.

By plotting months on the x-axis, we can track fluctuations in employment rates and job availability across different years, allowing for easy comparisons within the same graph. This approach provides a clear view of how these metrics evolve throughout the year.

We created these initial line plots as a way to explore potential visualizations for interactive implementation in D3. By plotting the job opening rates and unemployment trends over time, we aimed to get a sense of patterns, and how best to present these insights interactively. These exploratory plots helped us determine key trends and design choices that would allow us to create D3 visualizations that are both insightful and user-friendly.

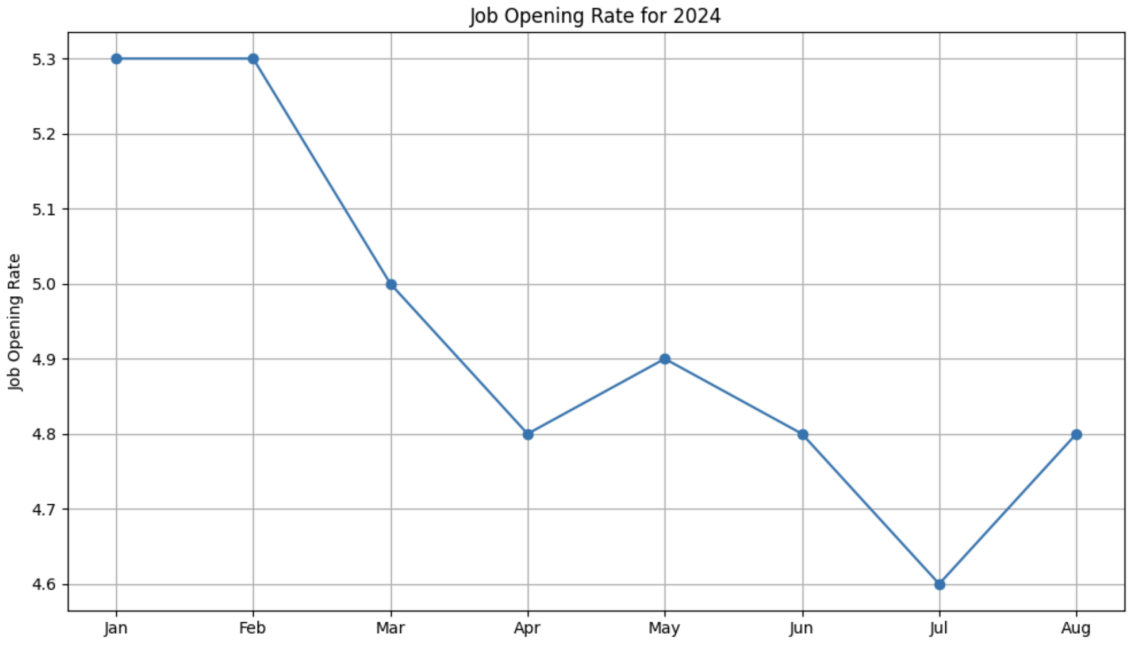
Line Plot 1: Monthly Job Opening Rate (2014-2024)

In this visualization, we aimed to capture trends in the monthly job opening rate from 2014 to 2024. We chose a line plot to show each year's monthly progression on the same graph, allowing us to compare trends in job availability across different years. By setting months on the x-axis and job opening rates on the y-axis, we could highlight seasonal variations and year-over-year differences. This approach lets us see any recurring patterns in job openings over the months and identify potential growth or decline trends over time. The legend categorizes each year for easy reference, while gridlines enhance readability across the timeline.



Line Plot 2 : Individual Year Line Plot: Job Opening Rate for 2024

To explore alternative visualization approaches, we considered displaying data for each year separately. This plot focuses on the job opening rate for 2024, with each month on the x-axis to show the year's trend in isolation. By visualizing one year at a time, we aimed to gain insights into specific monthly patterns and fluctuations without the visual overlap of multiple years. This approach helped us evaluate if isolating individual years might offer clearer insights in an interactive D3 setup.

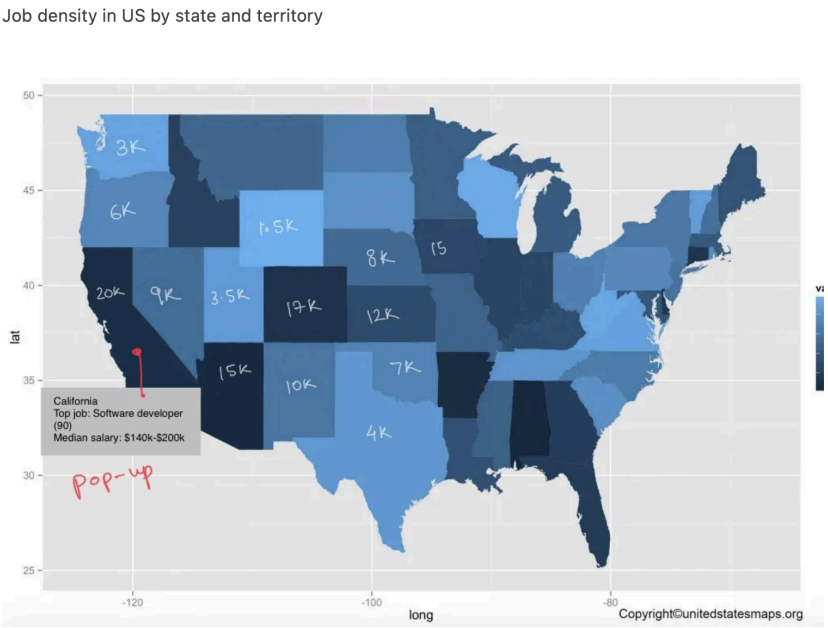


Final Design:

For the final design of the project, we have decided on four key visualizations that provide comprehensive insights into job opportunities, salary trends, and growth rates across various roles and regions. These visualizations include a Choropleth Map, an Interactive Bar Chart, a Bubble Chart, and a Line Plot. Each visualization has been designed with interactivity in mind, ensuring that users can dynamically explore the data without refreshing the page. Additionally, we have designed a user-friendly website layout to present all visualizations in one place.

Visualization 1: Choropleth Map with Pop-ups

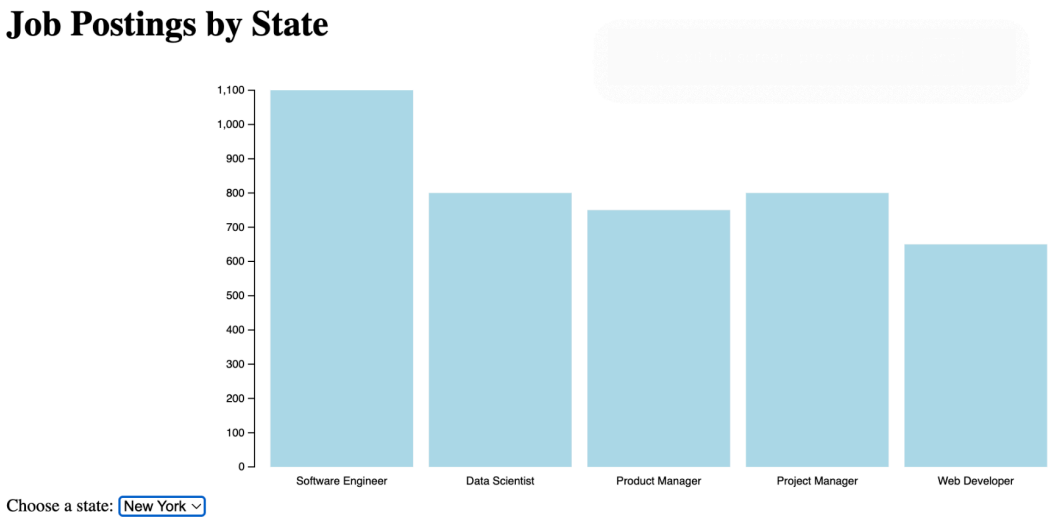
The Choropleth Map provides a geographic overview of tech job opportunities across the United States. Users can hover over individual states to see pop-ups displaying key details such as the number of job openings, top job roles, and salary ranges. This map allows users to quickly identify job hotspots and explore opportunities based on geographic location.



Visualization 2: Bar Chart 2 - Top Jobs in a State

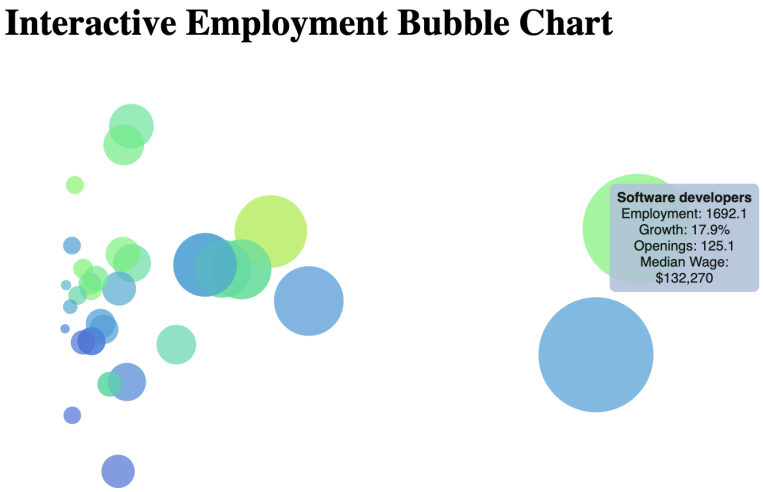
We will add an **interactive bar chart** that displays the top job roles within a selected state. For instance, users can explore which job roles are most in demand in California by using the dropdown or toggle options. The bar chart dynamically updates to reflect the top job roles in the chosen state, allowing users to dive deeper into the specific job demand across different regions. This chart complements the Choropleth Map by

providing more detailed insights into the roles available within each state. The interactivity helps users seamlessly switch between states and compare job roles across regions.



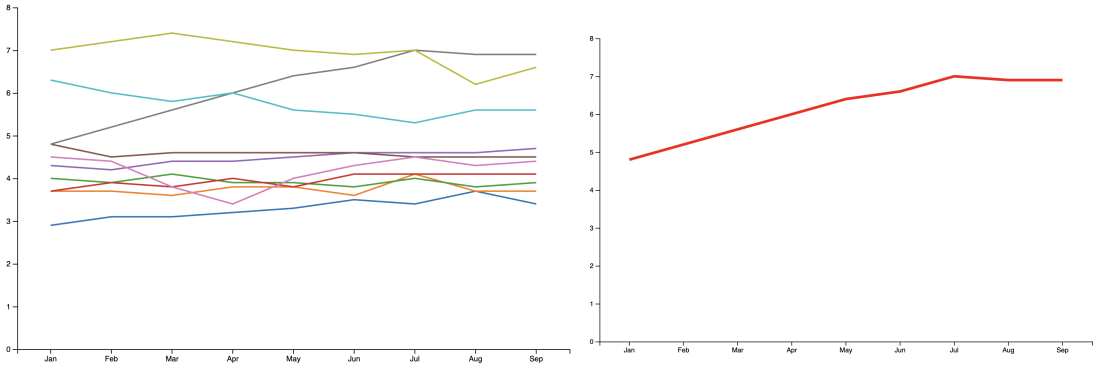
Visualization 3: Bubble Chart - Job Growth Rates by Industry

To represent projected growth rates for tech job roles, we selected a **bubble chart**. Each job role is represented by a bubble, with the size of the bubble corresponding to the projected growth rate. This visualization allows for easy comparison between job roles while highlighting those with the most significant growth potential. Users can hover over each bubble to get detailed information about the job role, including salary and projected growth.



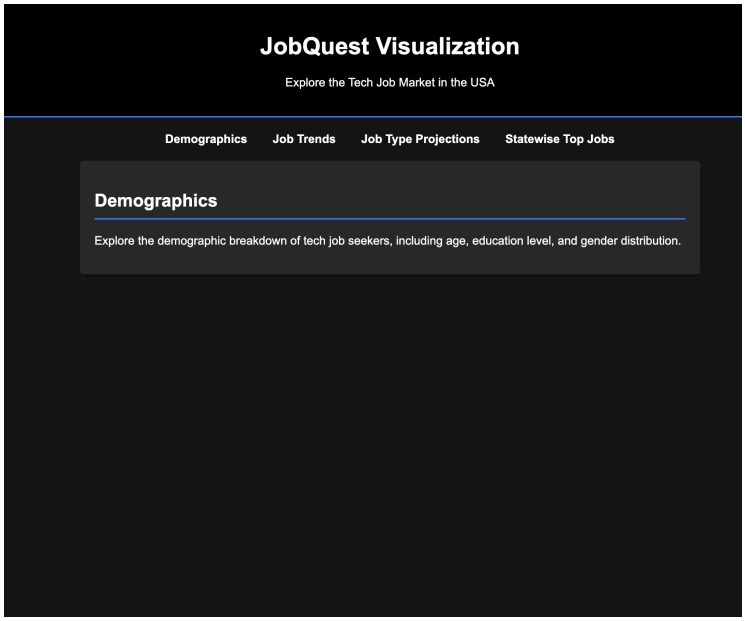
Visualization 4: Line Plot - Job Opening Rate Trends (2014-2024)

A **line plot** was chosen to visualize trends in job openings from 2014 to 2024. This plot highlights seasonal trends and year-over-year differences in job availability, making it easy to identify patterns in hiring cycles or periods of growth and decline. The plot uses monthly data to show trends in job openings for multiple years, providing a detailed timeline that helps users explore how job availability fluctuates over time. We also included a **second line plot** for individual years (e.g., 2024) to give users a more focused view of job openings during specific periods.



User can click on one of the line and select that line to show data for that particular line, we will be adding information about the particular year and the option to show different trends

Website Layout



To house all these visualizations, we designed a **clean and responsive website** with intuitive navigation. The website consists of the following key sections:

- **Navbar:** A sticky navbar at the top of the page allows users to navigate seamlessly between the different visualizations. The navbar links to each section, enabling easy access to the Choropleth Map, Bar Chart, Bubble Chart, and Line Plot.
- **Home Section:** The homepage introduces the project, explaining the purpose of the visualizations and how they can help users explore job opportunities and trends.
- **Demographics:** This section contains the interactive map, allowing users to explore job densities across the U.S. by state.
- **State Wise Top Jobs:** This section contains the interactive bar chart where users can select a state and view the top job roles for that region.
- **Job Type Projections:** This section houses the bubble chart, displaying projected growth rates for tech jobs, with interactive hover features for detailed insights.
- **Job Trends:** The line plots show job opening trends over time, with toggles to switch between multi-year and individual-year views.

Conclusion

We selected these visualizations because they provide a broad yet detailed view of the job market in the tech industry. The **Choropleth Map** offers geographic insights, while the **Bar Chart** allows users to dig deeper into specific job roles in selected states. The **Bubble Chart** efficiently represents growth projections, and the **Line Plot** captures trends over time. Together, these visualizations form a cohesive tool for users interested in exploring job opportunities, salary trends, and hiring patterns across the U.S.

We have started with the initial implementation and plan to enhance our visualizations further. This includes adding detailed information about the axes, refining the visual representations, and incorporating interactivity features that will allow users to engage more deeply with the data. By doing so, we aim to create a more user-friendly experience that facilitates better decision-making in the tech job market.