M1 Proposal

Project Title: Visualizing Tech Compensation

Members	Student Numbers
Benjamin Chang	14554159
Joy Xiao	19317171
Joey Shi	41274648

Overview

Our visualization will be tackling the problem of providing transparency on salary ranges for jobs in tech. People entering the field will not have a full understanding of how much they should be paid, and may be exploited by their lack of knowledge and accepting a lower pay than what is fair. By visualizing a dataset of salaries in the tech industry, the people using this visualization can achieve a better understanding of their monetary value and know when they are being offered a low or high salary for their experience and job title.

Data and Data Preprocessing

https://www.kaggle.com/datasets/jackogozaly/data-science-and-stem-salaries
"Salaries for Jobs in Data Science and Tech"

The dataset consists of salary records from top North American tech companies in jobs related to data science and tech. There are about 62,000 items total in the dataset. We will be visualizing the following attributes: company, title, total annual salary, and years of experience. We will also be aggregating and filtering data to show averages of total annual salary by company and/or title and/or years of experience.

Attribute	Type	Cardinality/Range
Company	Categorical	1600
Title	Categorical	15
Total Annual Salary	Quantitative	[10000 - 4980000]
Years of experience	Quantitative	[0 - 69]

Tasks

1. Identify distribution of salaries for different tech roles
2. Identify trends in salaries over time for different experience levels.
3. Locate outliers in salary

Visualizations

Views

- Job Salaries
    - Scatterplot
    - X-axis = time of recording data, y-axis = salary of the company
    - Colour hue encodes the job role
    - Line mark is included to show trend of data on the plot
- Salary by Experience Levels
    - Bar chart
    - 4 bars, each containing data from rows with 0-2, 2-5, 5-10, 10+ years of experience
    - Heights are determined by the average salary in the data of each bar
- Salary distribution
    - Histogram
    - Bin total annual salary data into 10k intervals

View Interactions (not linked)
- Hovering over a point on the scatterplot displays a tooltip with the company, salary, and salary of the data point (nice-to-have).

Linked Interactions
- Clicking on the Job salary legend filters the data in the bar chart and histogram for a specific job role.
- Clicking on a bar in the bar chart filters the data in the scatter plot and histogram for a particular experience level / years of experience range.
- Clicking on a bar in the histogram filters the data in the bar chart and scatter plot for a particular salary range.

After filtering the histogram view through clicking on a particular job role in the scatterplot legend, the histogram shows the distribution of salaries for a particular job role. This addresses our task (1), so this linked interaction (scatterplot -> histogram) is a must-have component.

After filtering the scatterplot view through clicking on a particular experience level in the bar chart, the scatterplot contains only jobs for that particular experience level. The trend line effectively express trends in the salary for a particular job role over time. This addresses our task (2), so this linked interaction (bar chart -> scatterplot) is a must-have component.

The scatterplot is our main visual to showcase the relationship between time and annual salary. When viewing all of the data without filters, users are able to see both a pattern between time or recording and salary for all jobs, as well as any extreme points regarding salary at the top and bottom of the plot to locate outliers. This addresses task (3), so the main view of our scatterplot is a must-have component

The other two views are also must-have components due to their linked interactions with the scatterplot.

Usage Scenarios

Aspiring tech workers can use this visualization to identify which company would provide them the best compensation. The visualization could also help young workers understand an expected range of salaries and prevent them from lowballing their entry salaries.

Example Scenario 1
John is in his final year of his computer science degree, a soon-to-be new grad. He is in the job pipeline with multiple companies and is expecting offers from several of them. He would like to know the expected salary ranges for his role and experience in the industry so that he may gain some negotiating leverage. This can be done by filtering for "New Grad" or 0-2 year compensation entries and particularly for the "Software Engineer" role.

Example Scenario 2
Jane is a freshman and is interested in pursuing a career in tech. The problem is however that she doesn't know what kind of educational stream or discipline she would like to go down. Many of her friends and classmates are intent on pursuing software engineering, but perhaps she would like to use her degree to become a product manager, or go down a more math and statistics heavy degree in order to become a data scientist. Our visualization can help her get an understanding of what she can expect in terms of compensation upon graduation and perhaps down the road in career development.

Example Scenario 3
Johnny is an aspiring tech worker. He has many future plans and endeavors he would like to take on, both personally and professionally and would like to plan for it financially. He knows the rough approximate number he would like to have to be able to live comfortably and pursue his interests but would like to know what kind of role and experience level he will have to be at in order to achieve this. This can be done through our visualization as well, through the histogram of salary distributions, by clicking on the band it will apply a filter across the other views to provide a more detailed view of roles and compensation for that particular salary band.

Team Communication Plan

- Communication between our team will be mostly online through Messenger. We typically expect replies within the same day. In terms of in-person work meetings, we plan on working together in person on the project once a week most likely Tuesday evenings after the 447 lecture. But we are all open and willing to meet in-person additionally if needed, most likely later in the week on a Thursday or Friday (to be determined by everyone's schedule)
- We don't expect any teamwork issues or interpersonal conflict as we are all ex-coworkers now friends but any failure to uphold to the above standards will be handled in the following way:
  - A direct confrontation where expectations and conflicts are laid out explicitly
  - If following this meeting, the issue persists or this meeting did not take place due to "ghosting" then we will go to the teaching staff and request that the team member in question be removed from the project group

Work Breakdown and Schedule

Task	Time Estimate (Hours)	Target Date	Assignment
EDA & Data Wrangling	3	Mar 20	Ben
Add docs + detailed specs for static views	3	Mar 20	Joey
Elaboration on code organization	2	Mar 21	Joey, Ben, Joy
Convert project to typescript and bootstrap with webpack	8	Mar 24	Joey
Implement static histogram	12	Mar 26	Joy
Implement static bar chart	8	Mar 26	Ben
Implement static scatterplot	8	Mar 26	Joey
Refine views, CSS styling according to specs	12	Mar 28	Joey, Joy, Ben

Implement interactive filter for scatterplot (job title)	12	Apr 2	Ben
Implement interactive filter for histogram (salary range)	12	Apr 2	Joy
Implement interactive filter for barchart (company)	12	Apr 2	Joey
Link scatterplot filter -> histogram	8	Apr 4	Joy
Link scatterplot filter -> barchart	8	Apr 4	Ben
Link barchart filter -> scatterplot	8	Apr 6	Ben
Link barchart filter -> histogram	8	Apr 6	Joy
Link histogram filter -> scatterplot	8	Apr 8	Ben
Link histogram filter -> barchart	8	Apr 8	Joey
Implement experience group filter on scatterplot	8	Apr 9	Joey
Link experience grouping -> barchart	8	Apr 11	Ben
Link experience grouping -> histogram	8	Apr 11	Joy
Final touches/refactoring Presentation Prep	4	Apr 12	Ben, Joey, Joy