Final Project Report

(due June 14th 11:59p.m)

Instructions on Implementation

1. Data Preprocessing: If the dataset you chose is not uniform/workable (has missing values,etc.) , you can use any programming languages of your choice before reading the data into d3.js. You can clean it with javascript too (and of course that's the ideal workflow). Once the data is cleaned, it is mandatory to use d3.js to read the data and visualize the plots for all the tasks given below.

   Note) You are not allowed to generate outputs from d3's generators during the preprocessing stage. You still have to use d3's generators (path generators, scale generators, etc.).

2. Written Report: It can be identical to your project progress report 2. However if you have made any change on your responses to the questions (except for the prototype itself, since you are building a real product this time anyways), indicate it on this report in **red** font, so that we can identify what has changed since your progress report 2.

3. The resulting html should deliver the story clearly and follow the narrative genre/structure of your choice. (e.g. proper headline, texts, placement of plots, not-so-crammed around a specific spot on the html).

4. Minimum 3 visualizations - table-based, network-based (i.e. a vis with a link data), and geometry-based (i.e. a vis with a position data) - should be included.

   a. Include interactivity in at least two visualizations. Refer to the Interactivity lecture slides for types of interactivity and examples. Your visual should match the interactive intention (e.g. correct element should be highlighted, tooltip should be positioned correctly).

   b. All the plots will be considered complete if they have the following:
      i. Axes (if required)
      ii. Appropriate dynamic scale
      iii. Legend (if required)
      iv. Title
      v. Axis Labels (if required)

What to Submit on Gradescope

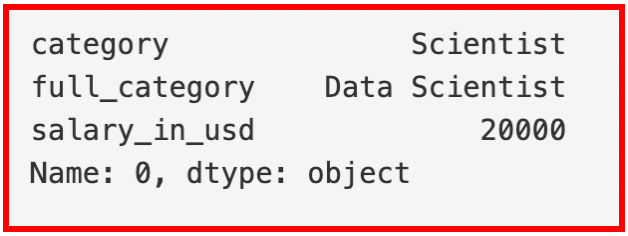
You **must submit all the files** listed below to get credit.

1. Your project progress report II (in pdf). If your final project has deviated from the latest/approved proposal, indicate so in **red** font on this pdf.
2. PDF-printed version of your resulting html page.
3. All your files (dataset, javascript, html, css, etc.) in a single zip file. When unzipped, the graders should be able to change directory into the root of that directory, run live-server, and everything should work without any modification on your zip file. If any file is missing and/or we cannot replicate what we see on your PDF, you will get 0pt.

Part 1 - Story and Narrative

Link to the dataset	https://ai-jobs.net/salaries/download/ https://datahub.io/core/country-list/r/0.html

Example item from the dataset	Salaries dataset:
	`1 data.iloc[0]`
	✓ 0.0s
	work_year 2023
	experience_level SE
	employment_type FT
	job_title Principal Data Scientist
	salary 80000
	salary_currency EUR
	salary_in_usd 85847
	employee_residence ES
	remote_ratio 100
	company_location ES
	company_size L
	Name: 0, dtype: object
	
	Country_code dataset:
	`1 country_code.iloc[0]`
	✓ 0.0s
	Name Afghanistan
	Code AF
	Name: 0, dtype: object
	
	Data_by_categories dataset (generated from the preprocessing step):

	
Story you want to deliver	**Facts:** ● Data science jobs are in high demand, with a growing need for skilled professionals. ● Salary increases for data science jobs tend to be substantial, especially as experience and expertise increase. ● Geographic location plays an important role in determining salary levels, with certain regions and countries offering higher compensation for data science positions. ● Remote work opportunities are becoming more common in the data science field, allowing professionals to work from anywhere. ● Data science offers a diverse range of job titles and career paths, including roles such as data analyst, data engineer, machine learning engineer and data scientist. **Insights**: ● The high demand for skilled data science professionals indicates a bright job market and ample career opportunities. ● Salary growth in data science jobs motivates individuals to invest in continuous learning and professional development. ● Geographic factors can affect earnings and should be considered when evaluating job opportunities or considering relocation. ● The availability of remote work options provides flexibility and potential work-life balance for data science professionals. **Messages**: ● The field of data science holds immense potential for individuals seeking rewarding and well-compensated careers. ● Strategic career planning, taking into account geographic factors and investing in educational qualifications, can pave the way for higher earning potential and career advancement in data science. ● The availability of remote work opportunities in this field provides the added benefits of flexibility and work-life balance, opening up new possibilities for professionals.

	• By understanding and leveraging data science job trends, individuals can make informed decisions and strategically shape their career paths for long-term success and financial growth.
Describe your target audience.	The intended audience is those who aim to work in the data science field and want to understand the salary growth of data science related jobs, including data science major students. **1. Familiarity with your topic? If not, how do we catch them up?** They will have an understanding of the growing demand for data science professionals today. If not, I will provide a brief overview of data science, its applications, and the growing demand for data science professionals. **2. Do they care? Why? Why not?** Audiences care because they are interested in pursuing a career in data science. They want to understand the potential for wage growth and career prospects in the field. They are seeking information that can guide their career decisions and help them effectively plan their educational path. **3. What do you want them to take away? Key points?** They can learn about wage growth trends in data science at different experience levels, identify the potential impact of geographic factors on earning potential and job opportunities and understand the value of educational levels for data science career advancement. **4. What do they know about visualization? Are your techniques standard?** The audience may have some familiarity with data visualization techniques. But I will use standard types of visualizations such as bar charts, line graphs, scatter plots, and pie charts to represent different aspects of the data. Avoid overly complex visualizations and provide annotations at some graphs. **5. How do they encounter your visualization?** They will interact with my visualization on the website by clicking some buttons to have a better view of data. **6. Mathematical background? Are you assuming too much? Too little?** I will assume the audience have a general understanding of mathematical terms such as mean, min, max, etc. But I will provide explanations of some higher level statistical terms such as standard deviation, etc.

	7. Device? Mobile phone, computer, print media... The audience may use mobile phones and computers to open this website. The canvas of the graphs will be adjusted with the screen size.
The goal of your project outcome. And why?	**Explanatory**: The goal of the project outcomes is to provide explanatory insights into data science job trends, salary growth, and related factors. The audience can learn the cause-and-effect relationships and develop a deeper understanding of the factors that influence compensation and job opportunities in data science. It's because it helps the audience comprehend the dynamics of the industry, make informed decisions, and navigate their career paths effectively.
Narrative structure you plan to use	Drill-Down
Elaborate your choice of narrative structure.	The project will begin by presenting an overview of the data science job trends, salary growth, and related factors through a high-level visualization. This initial visualization will provide a broad understanding of the overall patterns and trends in the data. As the audience interacts with the visualizations, they will find out details about those data.
Narrative genre you plan to use	Partitioned poster
Elaborate your choice of narrative genre.	The visualizations will be divided into different sections, each focusing on data science job trends, wage growth, ~~correlates~~, etc. Each partition will contain relevant visualizations, data summaries, and explanatory text to convey key findings and information effectively.

Part 2 - Outline

Story you want to deliver	**Facts:** ● Data science jobs are in high demand, with a growing need for skilled professionals. ● Salary increases for data science jobs tend to be substantial, especially as experience and expertise increase. ● Geographic location plays an important role in determining salary levels, with

	certain regions and countries offering higher compensation for data science positions. ● Remote work opportunities are becoming more common in the data science field, allowing professionals to work from anywhere. ● Data science offers a diverse range of job titles and career paths, including roles such as data analyst, data engineer, machine learning engineer and data scientist. **Insights**: ● The high demand for skilled data science professionals indicates a bright job market and ample career opportunities. ● Salary growth in data science jobs motivates individuals to invest in continuous learning and professional development. ● Geographic factors can affect earnings and should be considered when evaluating job opportunities or considering relocation. ● The availability of remote work options provides flexibility and potential work-life balance for data science professionals. **Messages**: ● The field of data science holds immense potential for individuals seeking rewarding and well-compensated careers. ● Strategic career planning, taking into account geographic factors and investing in educational qualifications, can pave the way for higher earning potential and career advancement in data science. ● The availability of remote work opportunities in this field provides the added benefits of flexibility and work-life balance, opening up new possibilities for professionals. ● By understanding and leveraging data science job trends, individuals can make informed decisions and strategically shape their career paths for long-term success and financial growth.
Specifications on each plot **in the order of how you lay out on your project**	1. Plot 1 1) Task: This plot a) compares the proportion of remote work opportunities in different years, and b) have the user see the annotation of the percentage values. 2) Attributes used: work_year, remote_ratio 3) Marks: Area mark 4) Channels: - Aligned vertical position channel for work_year - Encoded area channel for the percentage of remote_ratio

- Color channel for remote_ratio

5) How this plot adds to the story:
This plot adds to the story by visually demonstrating the trends and changes in the proportion of remote work opportunities over time, allowing the audience to easily interpret the data and understand the growth or decline of remote work in the field.

2. Plot 2

1) Task: This chart a) compares average salary for full time data science jobs of different levels of experience and b) analyzes trends in change of average salary over time.
2) Attributes: work_year, salary_in_usd, experience_level
3) Marks: line mark
4) Channels:
   - Aligned horizontal position channel for work_year
   - Aligned vertical position channel for average salary
   - Color channel for experience_level
5) How this plot adds to the story:
My visualization demonstrates the relationship between experience and salary, adding to the understanding of the salary growth potential in the data science field.

3. Plot 3

1) Task: This map helps the audience to lookup average salary of ~~different level of~~ data science jobs of a given country
2) Attributes: salary_in_usd, company_location, ~~experience_level~~
3) Marks: area
4) Channels:
   - Position channel for locating the countries on the map
   - Color channel for representing the average salary
5) How this plot adds to the story:
This map visualization allows the audience to easily look up and compare the average salary of ~~different levels of~~ data science jobs in various countries. It allows for a quick visual comparison between countries.

4. Plot 4

1) Task: This plot links a job category to many sub-categories that are related to this job category and can compare their average salaries.
2) Attributes: job_title, salary_in_usd
3) Marks: Connection

	4) Channels: - Color channel for distinguishing different job categories - Area channel for demonstrating the amount of average salaries associated with a node 5) How this plot adds to the story: The plot demonstrates which job pays more in each category, allowing people to take into consideration when choosing a job type.
Elaborate the choice of their marks and channels for each vis	Plot 1: Using area markers helps to visualize the proportion of remote jobs, where the size of the area represents the percentage value. The aligned vertical position channels make it easy to compare different work_year, while the color channels help distinguish between different levels of remote_ratio. This combination of markers and channels effectively communicates trends and changes in remote work opportunities over time. Plot 2: Using line markers allows comparison of average wages at different experience levels. The aligned horizontal position channel represents the work years, facilitating the analysis of trends over time. The aligned vertical position channels convey average wage values, while the color channels distinguish between experience levels, allowing for clearer differentiation of data points. Plot 3: Using area markers on the map helps visualize the average salary for different levels of data science work across countries. Location channels are used to locate countries on the map, while color channels represent average salaries, providing a visual comparison between countries. This combination of markers and channels allows the audience to quickly identify countries with higher or lower average salaries. Plot 4: Using links markers helps establish relationships between job categories and their related subcategories. Color channels are used to distinguish different job categories and make them easily identifiable. The area channel indicates the average wage associated with each subcategory, providing a visual comparison of wages within job categories.

Following sample answer about a single plot shows how detailed your answers to part 2 should be.

1. Plot 1
    1) Task: This chart a) analyzes trend between Height and Weight of patients with heart diseases and b) locates outliers within the patients
    2) Attributes: Height, Weight
    3) Marks: point mark
    4) Channels:
        - aligned vertical position channel for Height
        - aligned horizontal channel for Weight
    5) How this plot adds to the story:
       My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.

Part 3 - Prototype

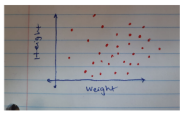
Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

A basic, barebones sample prototype for this project

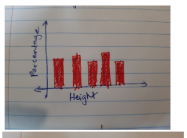
Heart Disease in the United States

Lading cause of death for men, women, and people of most racial and ethnic groups in the United States.
One person dies every 33 seconds in the United States from cardiovascular disease
About 695,000 people in the United States died from heart disease in 2021—that's 1 in every 5 deaths.
Heart disease cost the United States about $239.9 billion each year from 2018 to 2019.

We will look into how height and weight plays a role in heart disease.



Here, we can see that weight overpowers height in terms of influence on heart disease.
Larger weight compared to height can lead to a person having a larger chance of having a heart disease.
(explanation continues)



From this plot, we can observe that height does not have a significant impact on heart disease.
Height is primarily determined by genetic factors and influenced by nutrition and overall health during childhood and adolescence.
It is important to note that height itself does not directly influence the functioning of the heart or the development of heart disease.
(explanation continues)

On the contrary, weight plays a significant role.
Weight puts a person at risk for type-2 diabetes, sleep apnea, metabolic dysfunction, high blood pressure, which in turn gives rise to heart disease.
(explanation continues)

Heart disease continues to be a leading cause of mortality globally, but the good news is that it is often preventable.
By understanding the role of weight in heart disease prevention, you can take proactive steps towards improving your cardiovascular health and enhancing your overall quality of life.

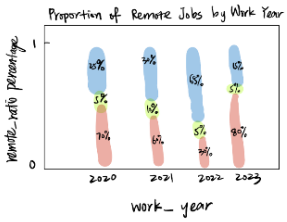
We can use the following strategies for the prevention of heart disease

- Adopt a healthy eating pattern
- Engage in regular physical activity
- Maintain portion control
- Limit sugary drinks and alcohol

Stay healthy, stay happy!

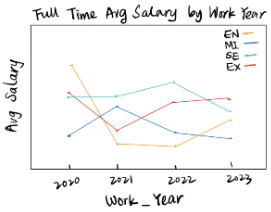
Status and Prospects of Data Science Careers

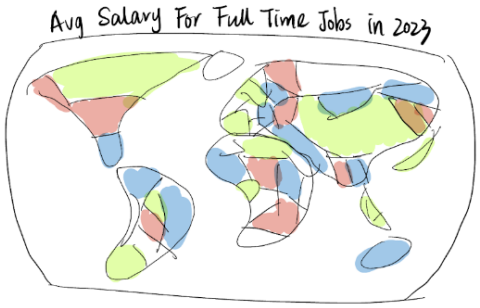
The field of data science has experienced tremendous growth and is in high demand, with organizations across industries increasingly relying on data-driven insights for decision-making. As a result, data science jobs have become sought after by individuals seeking rewarding and well-compensated careers. It is important to understand the industry landscape and make informed career choices now.



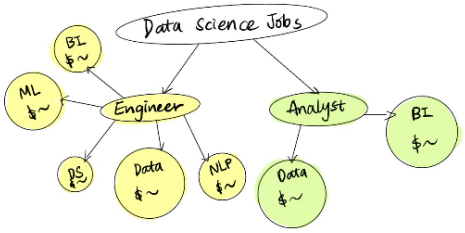
With the spread of COVID-19 virus since 2020, many companies have adopted remote work mode or hybrid work mode. However, as the severity of the pandemic diminishes, many positions have started transitioning back to offline offices starting from 2022.

From this graph, we can see that __ level job salary is increasing, while __ level job salary is decreasing. It means there are lots of people entering tthe job market, causing _____





(The audience can move mouse over each region to see the average salary of each experience level jobs of that country.)



(The audience can click on each node to see the average salary of each kind of jobs.)