

Project Progress Report 2

(due May 24th 11:59p.m)

You can start working on the project once your report is accepted and graded by your TA. The entire final project is worth **35%** of your final grade and this report accounts for **10%**. This project is done individually.

Submission Guideline

Download this google doc, fill the table. **Type** your answers, no handwritten answers will be accepted (except for the very last question). Submit it in **PDF** format on Gradescope.

If you need some inspirations please feel free to take a look at:

[Showcase of Information is Beautiful Awards](#)

[Bloomberg Year In Graphics Review](#)

[The Pudding](#)

[The New York Times](#)

Project Guidelines

Note: The guideline has been further clarified from Progress Report 1, so double-check whether your dataset choice still satisfies the updated guideline below.

1. You may use more than one dataset, however, regardless if you use one or multiple datasets, your visualizations must make use of at least three following data types - **link, position, and attribute**.
2. You cannot use any dataset from the class (Labs, Assignments, Lecture Exercises)
3. You can make your own dataset (Web scrape etc.) provided point 1. is satisfied.

Part 1 - Story and Narrative

Link to the dataset	https://www.kaggle.com/datasets/sogun3/uspollution
Example item from the dataset	Year: 2000 State: Alabama SO2 Mean: 3.3 NO2 Mean: 11.5 CO Mean: 0.3 O3 Mean: 0.02 State Code: AL County Code: 3 Site Num: 5 Address: East Thomas County: Jefferson City: Birmingham Date Local: 2000-01-01 Units of Measure: Parts per million Arithmetic Mean: 4.2
Story you want to deliver	<p>(a story should be in a form of a list of facts, insights, and messages - refer to the lecture slide)</p> <p>Air pollution levels in the United States have been a significant concern over the years. The dataset provides information on various air pollutants in different states and counties. We aim to explore the trends and patterns in air pollution levels across the United States. By analyzing the data, we can identify areas with high pollution levels and examine the changes over time. Our goal is to raise awareness about air pollution and its impact on the environment and public health. Through visualizations and insights, we hope to encourage measures to mitigate pollution and promote a healthier environment.</p>
Describe your target audience.	<p>Familiarity with the topic:</p> <p>The target audience may have varying levels of familiarity with the topic of air pollution in the United States. Some individuals may have a basic understanding of air pollution and its effects, while others may have a deeper knowledge of the subject. It is</p>

	<p>essential to cater to both audiences by providing sufficient context and explanations to catch up those who are less familiar with the topic.</p> <p>Level of interest and care:</p> <p>The target audience may have different levels of care and interest in the topic of air pollution. Some individuals, such as environmentalists, public health professionals, policymakers, or concerned citizens, may be highly invested in understanding and addressing air pollution due to its potential impacts on health, the environment, and social well-being. Others may have a general interest in learning about environmental issues. It is crucial to communicate the relevance and importance of the topic to capture their attention and generate interest.</p> <p>Key takeaways:</p> <p>The goal is for the audience to take away key insights and knowledge about air pollution in the United States. This includes understanding the spatial distribution of pollution levels, identifying trends and patterns over time, recognizing the impact of specific pollutants, and potential contributing factors. The audience should also gain awareness of the significance of addressing air pollution for environmental sustainability and public health.</p> <p>Knowledge about visualization:</p> <p>The target audience may have varied knowledge about visualization techniques. While some individuals may be familiar with common visualization techniques and their interpretation, others may have limited exposure to data visualization. Therefore, it is essential to use clear and intuitive visualizations, accompanied by explanations and context, to ensure understanding for all audience members.</p> <p>Encountering the visualization:</p>
--	---

	<p>The audience may encounter the visualization through various mediums, including websites, online articles, presentations, or reports. It is important to design the visualization to be accessible and responsive across different devices such as computers, tablets, and mobile phones, ensuring a seamless viewing experience.</p> <p>Mathematical background:</p> <p>It is advisable to assume a general understanding of basic mathematical concepts from the target audience. However, it is crucial to avoid assuming advanced mathematical knowledge, as the focus should be on presenting the data and insights in a clear and understandable manner for a broad audience.</p> <p>Device:</p> <p>The visualization should be designed to be responsive and accessible across different devices, including mobile phones, computers, and tablets. Consideration should be given to the screen size and usability on each device to ensure a user-friendly experience.</p>
<p>The goal of your project outcome. And why?</p>	<p>(exploratory vs. explanatory)</p> <p>By taking an exploratory approach, the project aims to uncover insights, patterns, and trends within the data. It seeks to provide a comprehensive understanding of air pollution dynamics across different states and counties in the United States. Exploratory analysis allows for flexible exploration and discovery of relationships and correlations in the data, which can help identify factors contributing to pollution levels and potential areas for intervention and improvement.</p>
<p>Narrative structure you plan to use</p>	<p>Interactive Slideshow</p>

Elaborate your choice of narrative structure.	The chosen chronological and thematic narrative structure provides a clear framework to present the analysis in a logical and easily understandable manner. It allows for a step-by-step exploration of air pollution data, highlighting key insights and patterns at each stage. By organizing the analysis thematically, it becomes easier for the audience to grasp the complex subject matter and follow the narrative flow.
Narrative genre you plan to use	Magazine Style
Elaborate your choice of narrative genre.	It aims to provide a comprehensive overview of air pollution in the United States, raising awareness, informing decision-making, and advocating for environmental action.

Part 2 - Outline

Story you want to deliver	<p>(you can copy/paste from Part 1)</p> <p>Highlight the significance of air pollution and its impact on the environment and public health.</p>
Specifications on each plot in the order of how you lay out on your project	<p>(for each plot, include 1) clear task abstraction, 2) attributes used, 3) marks, 4) channels, and 5) how this plot adds to the story)</p> <p>Plot 1 - Yearly Trends of Pollutants in the US: This plot likely displays the temporal trends of air pollutants (e.g., NO₂, O₃, SO₂, and CO) in the United States from 2000 to 2016. It could use a line chart or area chart to represent the pollution levels over time. The plot helps to analyze the overall trend in air pollution and identify any significant changes or patterns.</p> <p>Plot 2 - Distribution of Pollutants by State: This plot focuses on comparing the distribution of air pollutants across different states in the United States. It may use a bar chart or a map visualization to display the average pollution levels for each state. By examining the differences in pollutant levels, the plot highlights variations in air pollution across states and potentially explores the relationship between pollution and factors such as</p>

	<p>population density and industrialization.</p> <p>Plot 3 - Linked Data State: This plot visualizes the interconnectedness between states based on their O3 Mean pollutant levels. Each state is represented as a node in a connected graph, where states with higher O3 Mean values form one group and states with lower O3 Mean values form another group. The graph demonstrates the connections between these two groups, indicating the relationships between states with varying pollution levels. Interactive features such as zooming and tooltips enhance the exploration of the graph, providing insights into the patterns and connections in pollution levels across different states.</p> <p>Plot 4 - Distribution of Pollutants among States: This plot presents a U.S. map depicting the distribution of air pollutants among states. By excluding Hawaii and Alaska to eliminate volcanic activity influence, it aims to showcase patterns and concentrations of pollutants across the country. The plot could use color coding or shading to represent pollutant levels in different states, allowing for comparisons and identification of regions with higher concentrations.</p>
Elaborate the choice of their marks and channels for each vis	<p>Plot 1 - Yearly Trends of Pollutants in the US: A line chart or area chart can be used to visualize the temporal trends of air pollutants (e.g., NO2, O3, SO2, and CO) in the United States from 2000 to 2016. This plot demonstrates the changes in pollution levels over time. The horizontal position channel represents the date or time period, while the vertical position channel represents the pollution levels. By analyzing the chart, viewers can identify any significant trends, patterns, or fluctuations in air pollution levels throughout the years.</p> <p>Plot 2 - Distribution of Pollutants by State: To compare the distribution of air pollutants across different states in the United States, a choropleth map can be used as the mark for this visualization. The map uses color gradients to represent the pollution levels in each state. Darker colors indicate higher pollution levels, while lighter colors indicate lower pollution levels. This plot enables viewers to observe the variations in air pollution levels across states, providing insights into regional disparities and potential contributing factors.</p> <p>Plot 3 - Linked Data State: This visualization presents an interconnected graph that represents the relationships between states based on their O3 Mean pollutant levels. Each state is depicted as a node, with states having higher O3 Mean values forming one group and states with lower O3 Mean values forming another group. The connections between the nodes illustrate the associations between these two groups of states. The choice of using circles as marks effectively represents the states in</p>

	<p>the graph. Position, color, and potentially size channels are utilized to convey information. The interactive features of zooming and tooltips enhance the exploration of the graph, enabling users to delve deeper into the connections and gain insights into the patterns and relationships in pollution levels across states.</p> <p>Plot 4 - Distribution of Pollutants among States: For visualizing the distribution of air pollutants among states, a bar chart can be used as the mark. Each bar represents a state, and the height of the bar corresponds to the pollution level. The horizontal position channel represents the states, facilitating easy comparison. The vertical position channel represents the pollution levels, allowing viewers to identify states with higher or lower pollutant concentrations. This plot provides a clear understanding of the relative pollution levels among different states, emphasizing regions that require targeted pollution control measures.</p>
--	--

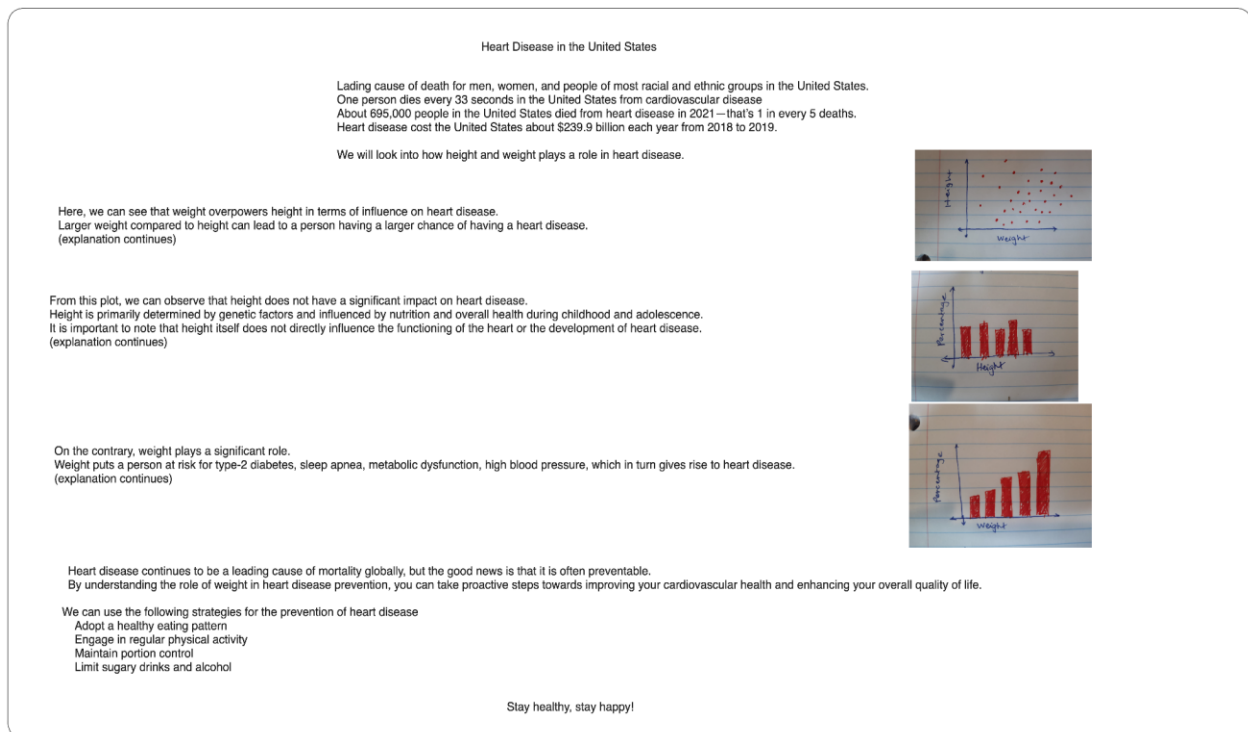
Following sample answer about a single plot shows how detailed your answers to part 2 should be.

<p>1. Plot 1</p> <ol style="list-style-type: none"> 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: <ul style="list-style-type: none"> - aligned vertical position channel for Height - aligned horizontal channel for Weight 5) How this plot adds to the story: <p>My visualizations aim to deliver health characteristics of patients with heart disease. This plot will provide more specific insights on Height and Weight.</p>

Part 3 - Prototype

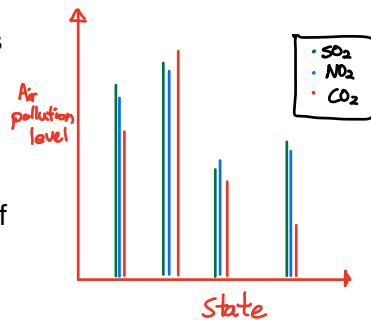
Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

A basic, barebones sample prototype for this project

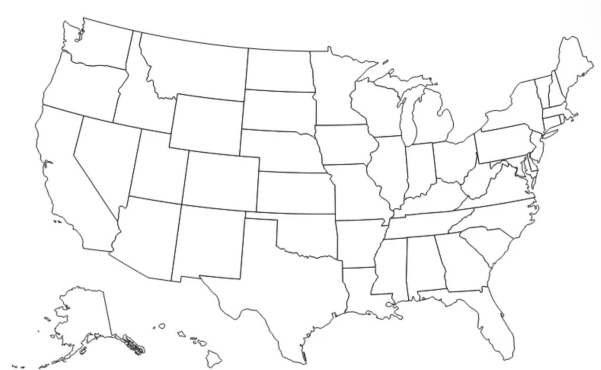


"Welcome to our webpage on air pollution in the United States! Air pollution is a significant environmental issue that affects both the quality of the air we breathe and the health of our communities. Understanding the distribution, trends, and factors influencing air pollution is crucial for developing effective strategies to mitigate its impact. In this webpage, we explore a comprehensive dataset on air pollution in the United States, aiming to provide insights into pollution levels, variations across regions, temporal trends, and potential contributing factors. Join us on this journey as we delve into the data and uncover key findings about air pollution in the United States."

The bar chart provides an overview of air pollution levels in the United States. From this plot, we can observe the variations in pollution levels across different states and counties. The horizontal axis represents the states and counties, while the vertical axis represents the pollution levels. The height of each bar indicates the magnitude of pollution, allowing us to compare the pollution levels between different areas. This plot sets the stage by presenting a broad picture of air pollution across the United States.



In this visualization, we compare the distribution of different air pollutants across states and counties. The grouped bar chart (or stacked area chart) showcases the variations in pollutant levels. Each bar (or area) represents a specific pollutant, and the height (or the area) of each segment indicates the pollution level. By using different colors for each pollutant, we can easily distinguish between them and identify the areas where specific pollutants are more prevalent. This plot allows us to gain insights into the relative distribution of various pollutants across different regions.



The line chart (or area chart) illustrates the temporal trends in air pollution levels over time. By plotting the pollution levels against the dates, we can observe how pollution has changed over a specific time period. The vertical axis represents the pollution levels, while the horizontal axis shows the timeline. The different lines (or areas) correspond to different pollutants, allowing us to analyze the trends for each pollutant separately. This plot enables us to identify any patterns or fluctuations in pollution levels over time and understand the overall temporal dynamics of air pollution.

