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	Flight:https://www.kaggle.com/datasets/rulyjanuarfachmi/domesticusairflight2016-2018 Airport:https://data.humdata.org/dataset/ourairports-usa/resource/03caba63-41a9-4638-9596-2dc7ada2a541?view_id=9ab8da5b-6f81-459e-b53c-4f5002c974e2
Example item from the dataset	Since the original dataset include every single commercial flight in us between 2016 to 2018, which is too large and will crash the browser, so I summarize monthly information between every two airport. Notice that Alaska and Hawaii are exclude. For major city selection in second chart, data are restricted to top 30 cities with most population. Year: year of flight, int Month: month of flight, int ~~Day:day of month of flight, int~~ ArrDelay: total Delay of arrive in minutes in that month, int DepDelay: total Delay of departure in minutes in that month, int Origin: departure airport in IATA form, string Dest: arrival airport in IATA form, string Cancelled: total number flight cancelled, int TypeO: type of departure airport, string TypeD: type of arrival airport, string latitudeO: latitude of departure airport, float longitudeO:longitude of departure airport, float latitudeD: latitude of arrival airport, float longitudeD:longitude of arrival airport, float StateD: state of arrival airport, string CityD: city of arrival airport, string StateO: state of departure airport, string CityO: city of departure airport, string Totalflight: total number of flight from origin to destiny in that month, int Example of item: columns:Year,Month,Origin,Dest,TypeO,latitudeO,longitudeO,CityO,StateO,TypeD,latitudeD,longitudeD,CityD,StateD,ArrDelay,DepDelay,Cancelled,totalflight item:2016,1,KABE,KATL,medium_airport,40.651773,-75.442797,Allentown,PA,large_airport,33.6367,-84.428101,Atlanta,GA,2.5,13.742424242424242,0.030303030303030304,66

Story you want to deliver	Fact: -The dataset contains detailed information about every commercial flight in the US from 2016 to 2018, including data on flight delays, cancellations, airport types, and geographical information. -The dataset allows for the analysis of flight numbers, delays, and cancellations over time. -The dataset includes geographical information about the airports, including their latitude and longitude. Insight: -The interconnectivity of airports within the US can be visualized to show the patterns of commercial flights over the given time period. -Trends can be identified, such as peak travel times, busiest airports, and common flight routes. -The geographical distribution of flights can be visualized to show the concentration of commercial flights in certain regions. Message: -By understanding these patterns, travelers can make informed decisions about their travel plans, such as avoiding certain routes or times that are prone to delays or cancellations. -Aviation industry professionals can use this information to improve operations, such as scheduling flights at less busy times or improving services at busy airports. -Researchers can use this information to study the impact of commercial flights on different regions, such as the economic benefits or environmental impacts. - By understanding the common causes of flight cancellations, airlines can take measures to reduce cancellations and improve customer satisfaction.
Describe your target audience.	**Travelers**: Individuals planning to travel within the US could use this data to optimize their travel plans. They could avoid routes or times that are prone to delays or cancellations and choose more efficient routes based on the data. **Aviation Industry**: This includes airline operators, airport management, and aviation policymakers. They could use the insights from this data to improve operations, manage flight schedules, and enhance customer satisfaction.

	General Public: Individuals interested in flight patterns, trends, and efficiency could find this data visualization informative and engaging.
The goal of your project outcome. And why?	The goal of this project is primarily explanatory. The aim is to communicate specific findings and insights from the dataset to the audience in a clear and engaging manner. The data visualization will guide the audience through the story, highlighting key points and providing context where necessary.
Narrative structure you plan to use	Martini Glass Structure
Elaborate your choice of narrative structure.	The Martini Glass Structure is suitable for this project because it allows us to first present the key insights and trends in a structured way, guiding the audience through the main findings. This is important given the complexity of the dataset and the need to clearly communicate the main messages. Once the audience has a good understanding of the main findings, they can then explore the data more freely, looking at different time periods, airport types, or other variables of interest. This exploratory phase allows the audience to engage more deeply with the data and discover additional insights that are particularly relevant to them.
Narrative genre you plan to use	Annotated Chart
Elaborate your choice of narrative genre.	The Annotated Chart genre is suitable for this project because it allows for a detailed and comprehensive visualization of the flight data. Given the complexity and multi-dimensionality of your dataset (involving time, location, flight delays, cancellations, etc.), a single, complex visualization can effectively represent the data in an integrated manner. Annotations can be used to guide the audience through the visualization, explaining the various components, highlighting key insights, and providing additional context where necessary. This can make the data more understandable and accessible to

	the audience, helping them to grasp the main messages and insights.

Part 2 - Outline

Story you want to deliver	Fact: -The dataset contains detailed information about every commercial flight in the US from 2016 to 2018, including data on flight delays, cancellations, airport types, and geographical information. -The dataset allows for the analysis of flight numbers, delays, and cancellations over time. -The dataset includes geographical information about the airports, including their latitude and longitude. Insight: -The interconnectivity of airports within the US can be visualized to show the patterns of commercial flights over the given time period. -Trends can be identified, such as peak travel times, busiest airports, and common flight routes. -The geographical distribution of flights can be visualized to show the concentration of commercial flights in certain regions. Message: -By understanding these patterns, travelers can make informed decisions about their travel plans, such as avoiding certain routes or times that are prone to delays or cancellations. -Aviation industry professionals can use this information to improve operations, such as scheduling flights at less busy times or improving services at busy airports. -Researchers can use this information to study the impact of commercial flights on different regions, such as the economic benefits or environmental impacts. - By understanding the common causes of flight cancellations, airlines can take measures to reduce cancellations and improve customer satisfaction.
Specifications on each plot **in the order of how you lay out on your project**	**For every graph cancellation will be represent by cancellation rate and delay will be represent by average departure delay time for each flight** **Plot 1**: Bar Chart of Flight Data

	Task: This plot allows users to analyze the overall US airline number, delay, and cancellation data for selected years and months. Attributes: Year, month, number of flights, flight delays, and cancellations. Marks: Bars. Channels: Height (to represent the number of flights, delays, or cancellations), ~~color (to differentiate between different years or months).~~ How This Plot Adds to the Story: This plot provides a summary of the flight data, allowing users to easily compare the number of flights, delays, and cancellations across different time periods. **Plot 2**: US Map with Flight Connections Task: This plot visualizes the interconnectivity of airports within the US, showing the patterns of commercial flights over the given time period. Attributes: Origin and destination airports (latitude and longitude), number of flights, flight delays, and cancellations. Marks: Lines (representing flight routes) and points (representing airports). Channels: Position (for the geographical location of airports), color (to represent the number of flights, delays, or cancellations), and line thickness (to represent the volume of flights between two airports). How This Plot Adds to the Story: This plot provides a visual representation of the flight network within the US, highlighting the busiest routes and airports. It also allows users to see the impact of delays and cancellations on different routes. **Plot 3**: US Geography Chart of Flight Data Task: This plot allows users to analyze the overall US airline number, delay, and cancellation data for **each state** for selected years .

	Attributes: Year, number of flights, flight delays, and cancellations, State. Marks: shape, position Channels: color(to represent the value of total flight/average delay/cancellation rate). How This Plot Adds to the Story: This plot provides a summary of the flight data, allowing users to easily compare the number of flights, delays, and cancellations across States in select time periods.
Elaborate the choice of their marks and channels for each vis	Plot 1: Bars are used because they are effective for comparing quantities. The height of the bars can be easily compared, making it clear which time periods have the highest or lowest numbers. Color is used to differentiate between different years or months, making it easier for users to compare data across time periods. Plot 2: Lines are used to represent flight routes because they naturally depict the connection between two points (airports). Points are used to represent airports because they can accurately depict the geographical location of each airport. Color and line thickness are used to encode additional information (flight volume, delays, cancellations) because they can be easily perceived and compared. Plot 3: Shapes and position used to represent each state's shape and relative position on a map. The color will be used to represent the the value of total flight/average delay/cancellation rate, darker the color i, higher the value is.

Part 3 - Prototype

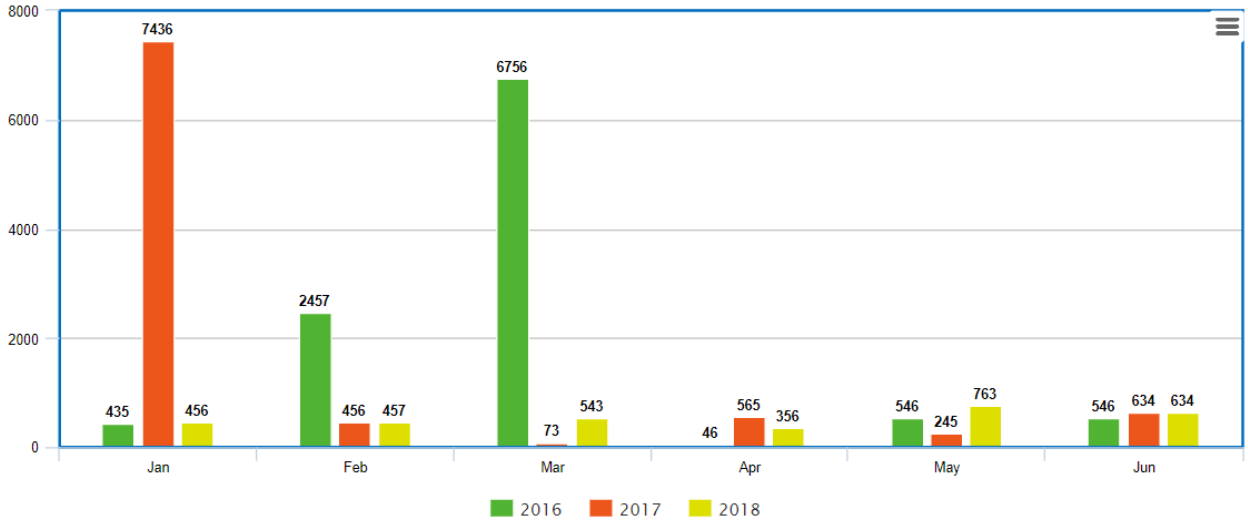
Provide a photo or screenshot of your prototype. A prototype should depict how you place different components of your visualization. You may use pen-paper, or using tools like excalidraw, figma etc.

==========

"Exploring US Commercial Flights: A Data Visualization Study (2016-2018)"

Introduction:
Welcome to my data visualization project, where we delve into the intricate world of US commercial flights from 2016 to 2018. Our study utilizes a comprehensive dataset that includes flight delays(Below it will be represented by average delay time for each flight in minutes), cancellations(Below it will be represented by cancellation rate), airport types, and geographical information. The aim is to provide insights into the patterns and trends of commercial flights during this period. Whether you're a traveler, an aviation professional, a researcher, or simply interested in flight patterns, this visualization offers a unique perspective on US commercial flights. Notice that, we exclude all data related to Alaska and Hawaii.

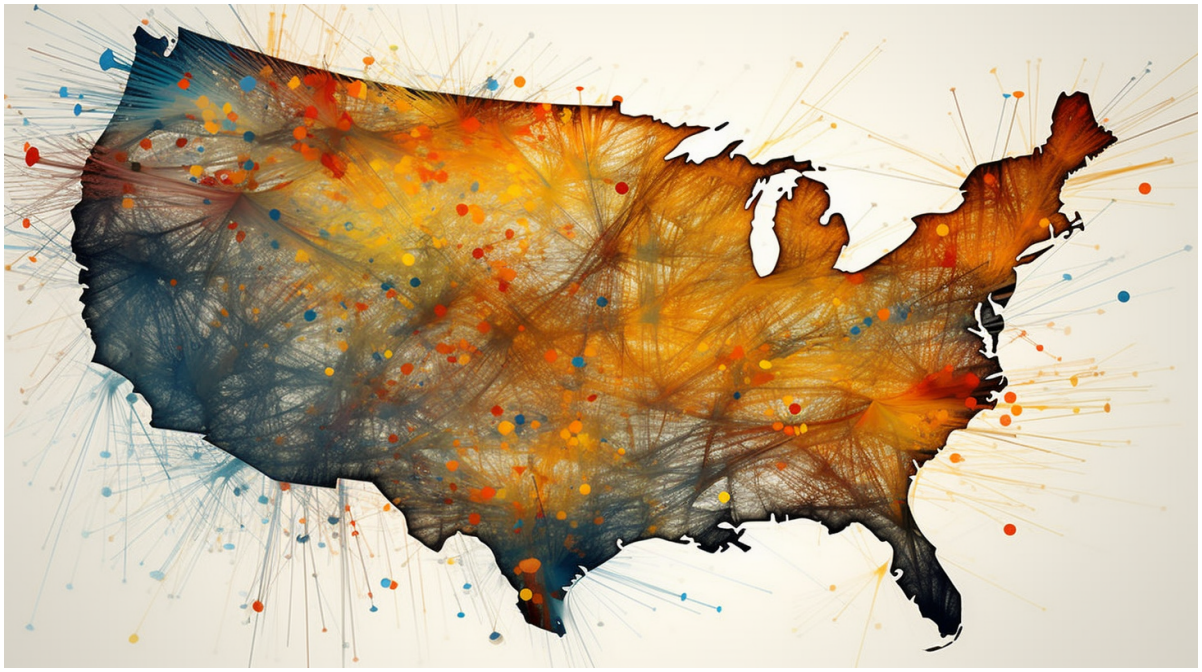
Graph 1 - Bar Chart of Flight Data:
Our first visualization is a neat bar chart representing flight data over time. The bars represent the total number of flights, average delay time for each flight, or cancellation rate, depending on what you choose. This chart provides a summary of the flight data, enabling easy comparison across different time periods. Select a year and type of date to delve deeper into the data and understand the trends of commercial flights during that period.



(This plot is generator by https://www.meta-chart.com/)

Graph 2 - US Map with Flight Connections:

Our second visualization is a stylized map of the United States, highlighting the interconnectivity of airports across the country. Each point on the map represents an airport, with lines connecting them to depict flight routes. The color of the lines vary, representing the volume of flights, average delay times for each flight or cancellation rate. This visualization provides a bird's-eye view of the flight network within the US, allowing you to identify the busiest routes and airports at a glance. Explore the map to discover more about the patterns of commercial flights during 2016-2018. You can also select the Major City option to include only top 30 cities with the most population in the US, notice that some cities may not operate an airport.



(This plot is generated by AI, midjourney)

Graph 3 - US map with each State's flight information

Our third visualization is a color-coded map of the United States, reflecting the aggregate flight number, cancellation rate, or average delay times for each flight data for each state during the selected time period. The color darkness corresponds to the total volume of flights, the cancellation rate, and the average delay time in each state. Darker shades represent higher numbers or longer delays, providing a clear visual summary of state-wise flight activities. This representation allows you to understand which states have the busiest airports, the most cancellations, or the longest delays. Delve into the details of each state by choosing times and type of data on it, unveiling an even deeper level of data analysis for your exploration. This map delivers an expansive view of the trends and patterns that have shaped US commercial flights from 2016 to 2018.