Data Visualization - Milestone 2

Flight Delays Data - Seyed Parsa Neshaei, Aryan Ahadinia, Matin Ansaripour

1. Introduction and Goal

Flight delays are a common issue with real-world consequences for millions of travelers. Disruptions caused by flight delays propagate across individuals, airlines, and airport logistics. In this project, we aim to explore how data visualization can help investigate such delays and reveal patterns that might have gone unnoticed otherwise. We aim to create an intuitive and interactive tool that allows prospective passengers and researchers to engage with the trends of delays over time. Our project focuses on the United States due to the lack of data accessible with low costs, but has the potential to be expanded with new datasets to include additional transportation options and other regions in the future. By presenting this information visually, we aim to make navigating complex transportation data more meaningful and accessible for users. Passengers will be able to explore when and where delays are most common, how they change across seasons, and what might be the contributing factors. Finally, we aim for our visualizations to also be useful for airports and airlines; we hope to invoke broader conversations among the stakeholders on the efficiency of transportation, experience of passengers, and the challenges in the aviation industry.

2. Sketches (available in the appendix)

2.1 The core MVP visualization: interactive flight delay visual dashboard (figure 1)

Features: The main page of the website will include a map of the United States with the airports indicated by icons over the map. Users can either click on a certain airport on the map, or use the dashboard on the left side which includes basic instructions and a text-box with autocomplete feature to select a certain airport or airline. After selecting a airport, it is highlighted in the map along with all the connections to other airports, and after selecting an airline, all airports which include flights from that airline will be highlighted. The users can see the aggregated data of delays for the airport or airline, and can filter for a specific year.

Extra features: In case of selecting two airports, a comparative analysis of the flight routes between the two airports will be shown, allowing the passengers to find the airline with the minimum delay that can bring them from the first to the second airport and vice-versa.

Goal: This visualization will enable the users to freely explore and compare airports and airlines regarding delay information in order to gain insights on the optimal paths to take for traveling by plane.

2.2 Delay changes over time (figure 2)

Features: This visualization again includes the map of the United States with the airports represented as circles. The size of the circles indicates the flight volume (i.e., bigger circles for airports with a larger average number of flights per day). The color of the circles indicates the relative delay status (green, blue, yellow, and red for the four quartiles from lower to higher amounts of delay). A timeline slider below the map enables users to see how delays change over time (in units of months).

Extra features: A play button next to the slider enables the playback feature, in which it can be seen how the delays of the part of the map that the users have zoomed in have changed across subsequent years. The sizes and colors of the "circles" will change while the slider on the bottom shows how the time is going forward.

Goal: By interacting with the visualization, users can identify temporal trends, e.g., seasonal patterns in delays. This enables not only the passengers but other flight stakeholders, such as airport and

airline managers, to understand how airport performance has fluctuated over time and to make informed decisions based on historical delay data.

2.3 Delay cause exploration (figure 3)

Features: This visualization includes an interactive stacked bar chart, displaying the distribution of different delay causes (e.g., weather, security, etc.) for airlines and airports that can be added or removed to the plot. By clicking on each delay cause, users can drill down into specific delay types to view more detailed breakdowns in terms of aggregated statistics. Filters below the plot allow users to refine the data by airline, airport, and year, enabling targeted exploration and comparison between multiple airlines or airports.

Extra features: Similar to 2.2, a play button next to the "year" dropdown enables the playback feature, in which it can be seen how the distribution of the delay causes changes over time. The height of the columns and the distribution inside each column will move up and down with animation as time progresses in the playback.

Goal: This visualization helps passengers and stakeholders analyze the root causes of flight delays and find how those are interrelated. This visualization combines high-level aggregated summaries with more in-depth comparisons on request, providing data-driven insights for travelers and policymakers to reduce the more common delay types and improve travel efficiency.

2.4 City network planner (figure 4)

Features: This is an interactive connection planning tool, in which users input names of desired cities, and see a visualization of the paths between cities in terms of a circular layout graph (chord). The graph shows cities as nodes and flight connections as edges, with edge thickness for flight volume and edge color for average delay (using a similar color coding to section 2.2).

Extra features: Users can filter the plot to include or not include certain airlines in the visualization, enabling them to see the connections between cities in case of certain passengers preferring to use or not use certain airlines.

Goal: This visualization helps users explore and plan air travel between cities by offering an overview of connections. By using a circular layout, users can easily identify the most connected cities and the busiest routes. The feature of filtering by airline makes this visualization useful particularly for travelers with specific preferences (or loyalty programs), and also for airline analysts who might aim to understand the coverage of their network for optimizing the flight operations.

3. Tools and Lectures

To develop the website, we use the React framework based on TypeScript (a superset of JavaScript with static typing to allow us write more maintainable and robust codes), and host our website on GitHub pages. We utilize the contents of the Basic web development, Javascript, and D3.js lectures when designing the basics of the website. For the principles of designing the visualizations and the data story, we use the lectures Designing viz, Do and dont in viz, "Marks, channels", Interactions, and Storytelling.

In addition to the lectures mentioned above, we particularly aim to:

- For 2.1 and 2.2, make use of the Maps and Practical Maps lectures.
- For 2.1 and 2.4, make use of the Graphs lecture.
- For 2.2, 2.3, and 2.4, make use of the Perception colors lecture.

Regarding the tools for creating the interactive visualizations, we use D3.js. For visualizations 2.1 and 2.2, which include maps, we use the react-simple-maps library. Finally, for the appearance of the landing page of the website, we use Globe.gl as well.

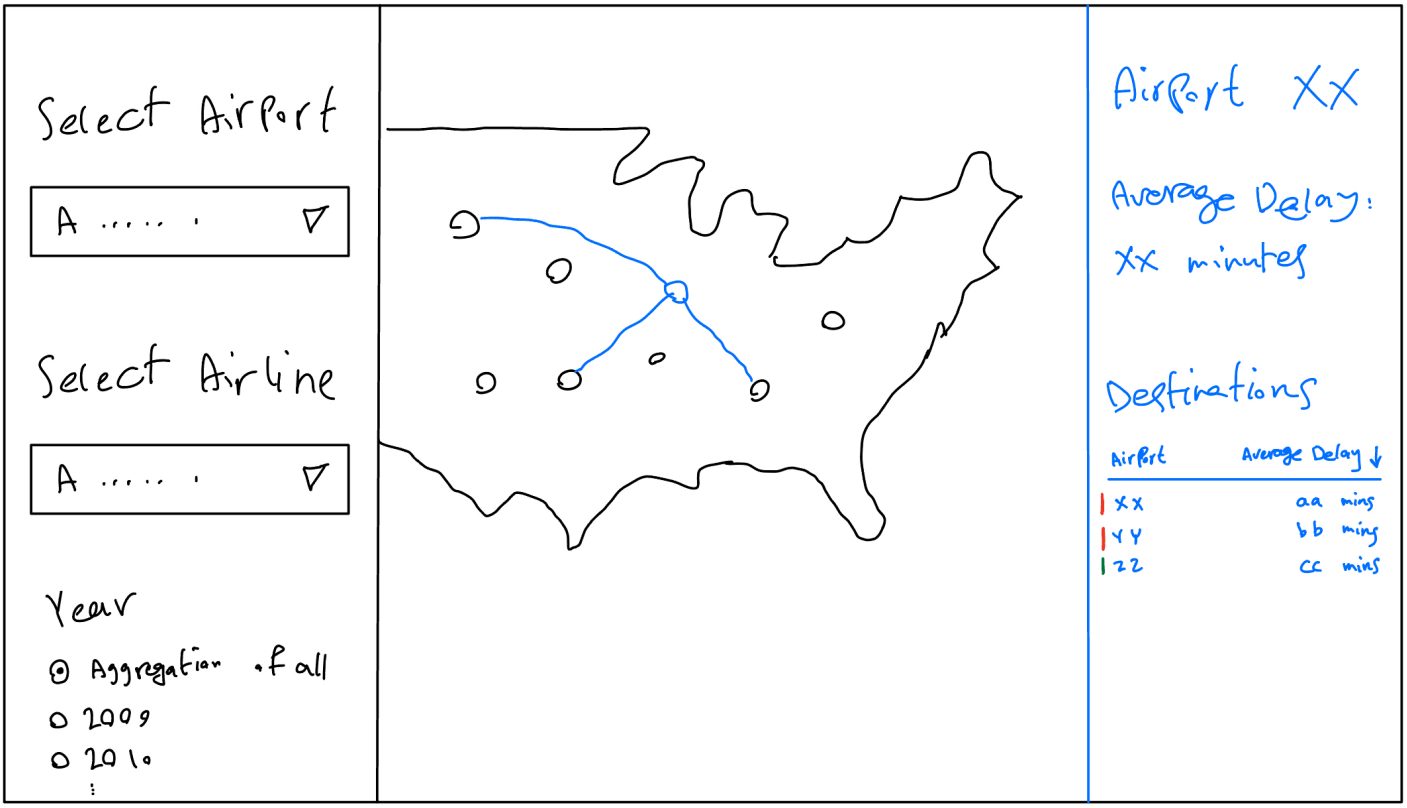

Figure 1: interactive flight delay visual dashboard

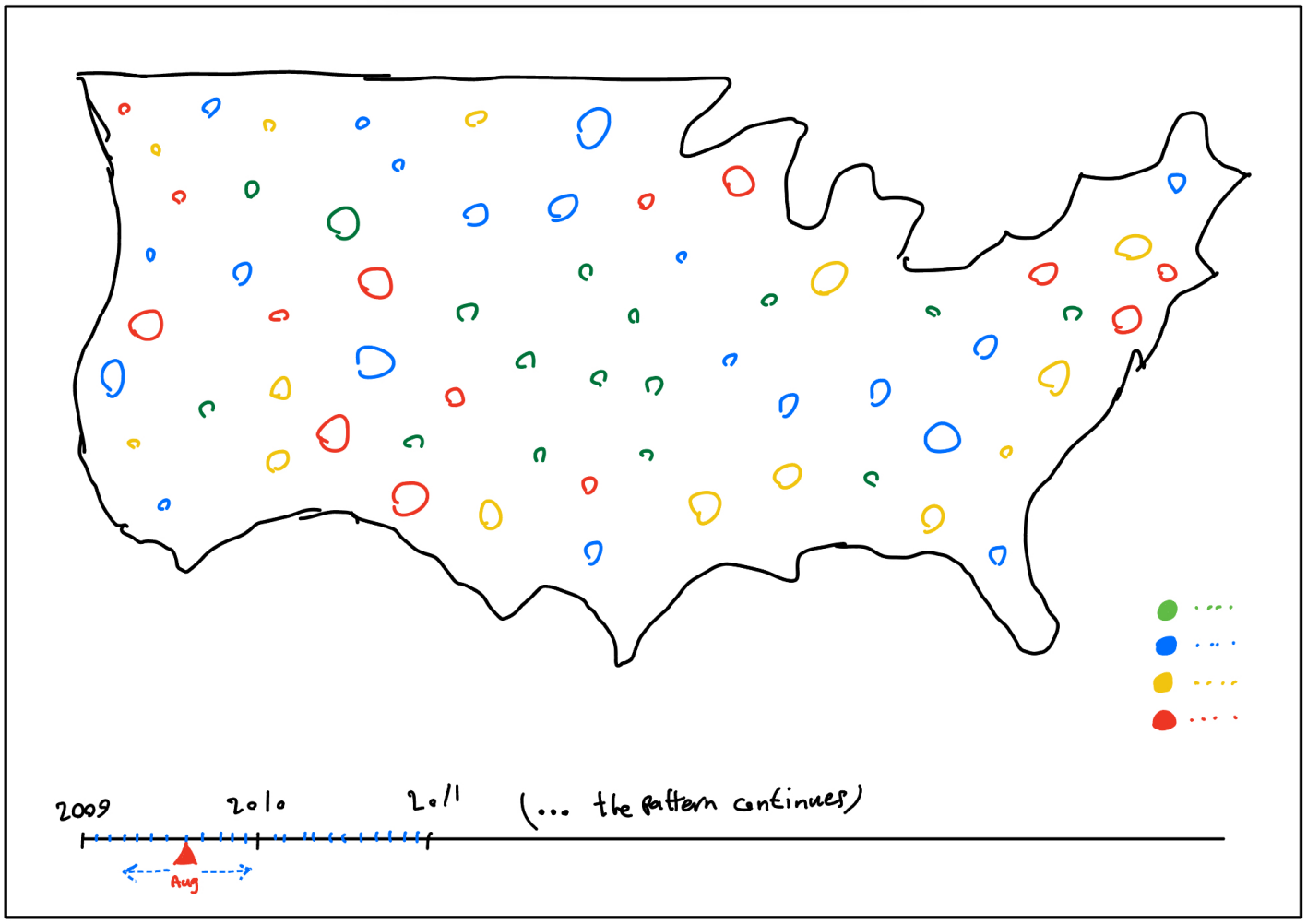

Figure 2: delay changes over time

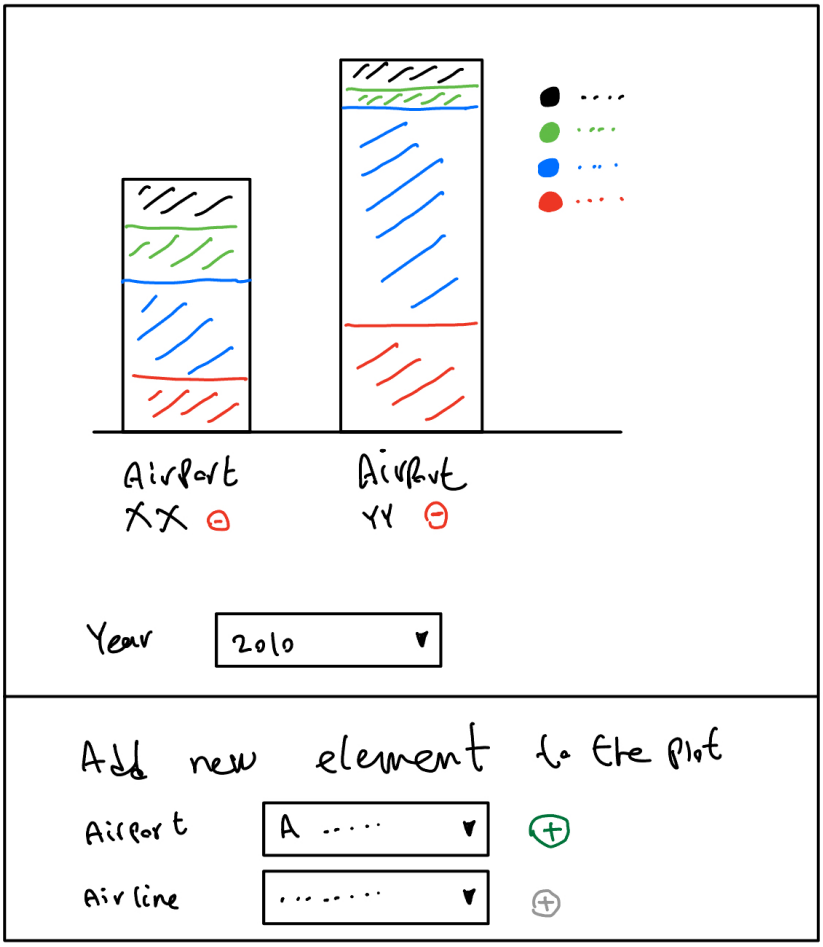

Figure 3: delay cause exploration

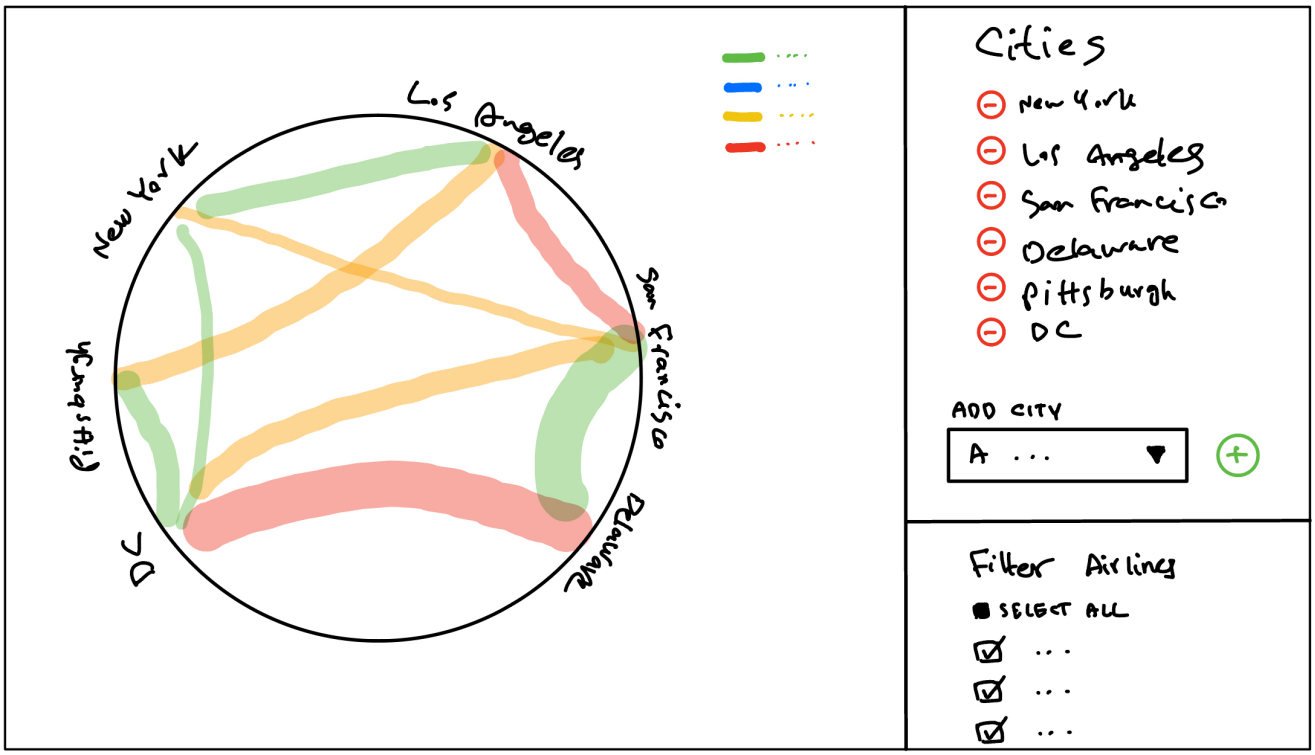

Figure 4: city network planner