Data Visualization Process Book: Flight Delays

Armance Nouvel, Diogo Valdivieso D. C., Tomas Valdivieso D. C.

Group Les Panthères

Introduction

In our increasingly connected world, air travel is essential for linking people and locations. However, the complexity of flight data can be overwhelming for everyday users to analyze and understand. This project seeks to tackle this challenge by delivering insightful analysis and visualization of flight data. Visualizations in this field are particularly valuable as they offer an intuitive means to grasp large amounts of information related to flights, such as routes, frequencies, delays, and passenger volumes.

Our objective is to create a user-friendly interface that allows users to explore detailed flight and airport information, helping in informed decision-making and enhancing overall air travel experiences. We aim to serve individual travelers by providing them with comprehensive and easily accessible flight data.

In this process book you will discover how this project was transformed from an idea into an interactive tool, providing valuable insights into flight patterns, delays, and airport activities.

Path to Final Result

Data Exploration

We are utilizing the "On-Time: Marketing Carrier On-Time Performance" dataset from the Bureau of Transportation Statistics. This dataset provides essential information on commercial US flights, including scheduled and actual departure and arrival times, canceled and diverted flights, causes of delays and cancellations, and air time, among other details. The user-friendly interface allows us to select specific features before downloading, reducing the need for extensive filtering.

Since the data is maintained and sourced by the US government, we anticipate minimal data cleaning. However, during our initial exploratory data analysis, we discovered significant gaps in several variables.

Website Structure

After completing the data exploration, we had a cleaned dataset ready for use. This allowed us to form a clearer vision of the future structure of our website, which we planned to divide into two main parts:

1. Exploration

   This area of the website was planned to integrate an interactive map where users could select departure and arrival airports. It would offer detailed flight and airport information, including departure times, delays, flight frequencies, and cancellations.

2. Analysis:

   This section was intended to present correlations between different flight characteristics.

By merging exploration and analysis, we wanted to provide a robust tool for understanding and improving air travel through effective data visualization."

First sketches and plans

After finalizing the structure of our website, we began sketching ideas for the design of its various components. Our initial concept for the main map includes blue points representing different airports, with their size corresponding to the airport's capacity. Initially, we planned to allow users to select a specific flight and display detailed information about it, as shown in the plot below. However, as the project progressed, we shifted our focus to emphasize the statistics of the selected airport rather than individual flights. We decided to highlight the largest connections between airports when one is selected.

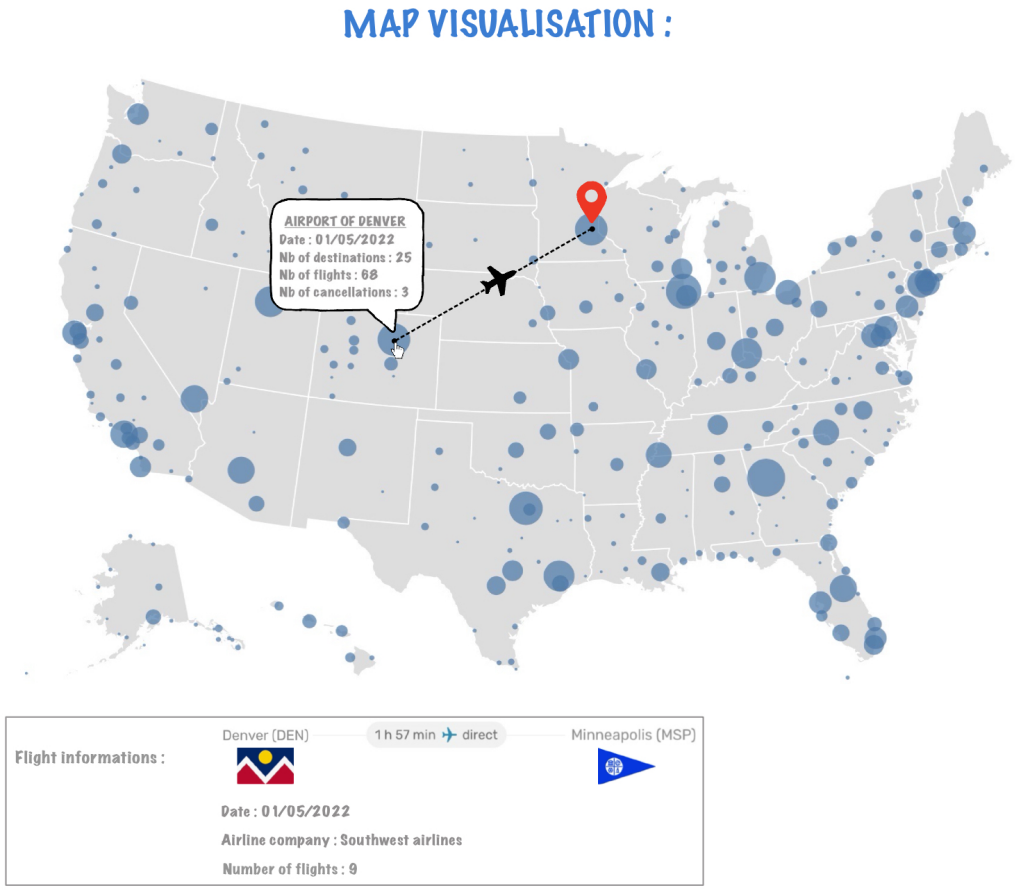

Figure 1: Drawing of our exploration map

For the statistics section of our website, we aimed to provide detailed insights into airport characteristics through comprehensive and clear visualizations. Below is the sketch of the statistics we envisioned for our website:

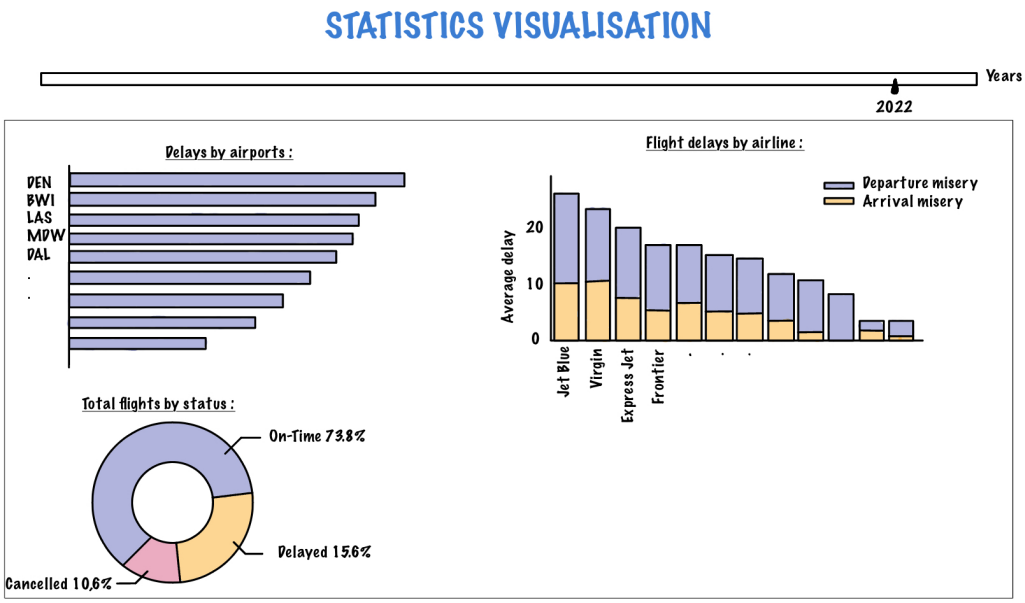

Figure 2: Drawing of our statistics visualisations

First skeleton

Following our structural decision, we transitioned into the coding phase of our website development, utilizing D3.js to construct dynamic maps depicting airport connections, and incorporating chart.js for the statistical analysis.

We have begun implementing the skeleton of our website, laying the foundation for its structure and functionality. This initial phase involves setting up the core components and ensuring a solid framework for future development and enhancements. You can see our first skeleton below:

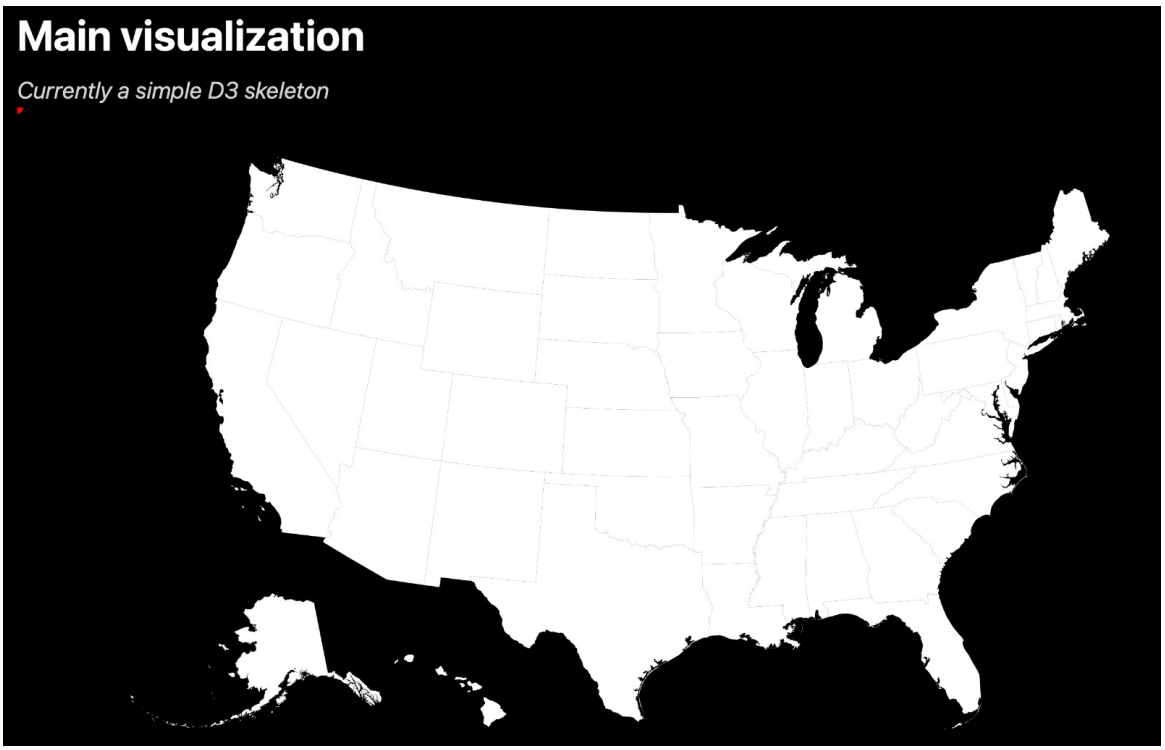

Figure 3: Picture of our initial map skeleton

Final Website

- **Step 1 : the exploration map**

  We opted to spotlight the 100 most prominent airports in the United States. Furthermore, when illustrating connections between airports, we opted to display the top five correlated airports for each, facilitating intuitive comprehension. We designed the map to be interactive, allowing users to click on any desired airport to display its name and characteristics, as shown below:

Figure 4: Picture of our interactive map

- **Step 2 : analysis of the airport**

  Notably, upon selecting an airport, users can navigate to a comprehensive statistical breakdown of delays, encompassing the frequency of delays in minutes across other airports, the distribution of delay causes as a percentage, and an airport's performance score depicted through a spider chart. This intricate visualization method encapsulates multivariate data across diverse dimensions, enabling effortless comparison. Each axis represents a variable scored from 0 to 100, encompassing aspects such as:

  - Cancelled flights
  - Delayed flights
  - Airport size
  - Diverted flights
  - Composite score, which represents the overall performance of the airport, composed of the preceding variables

Here is the outline of our airport analysis:

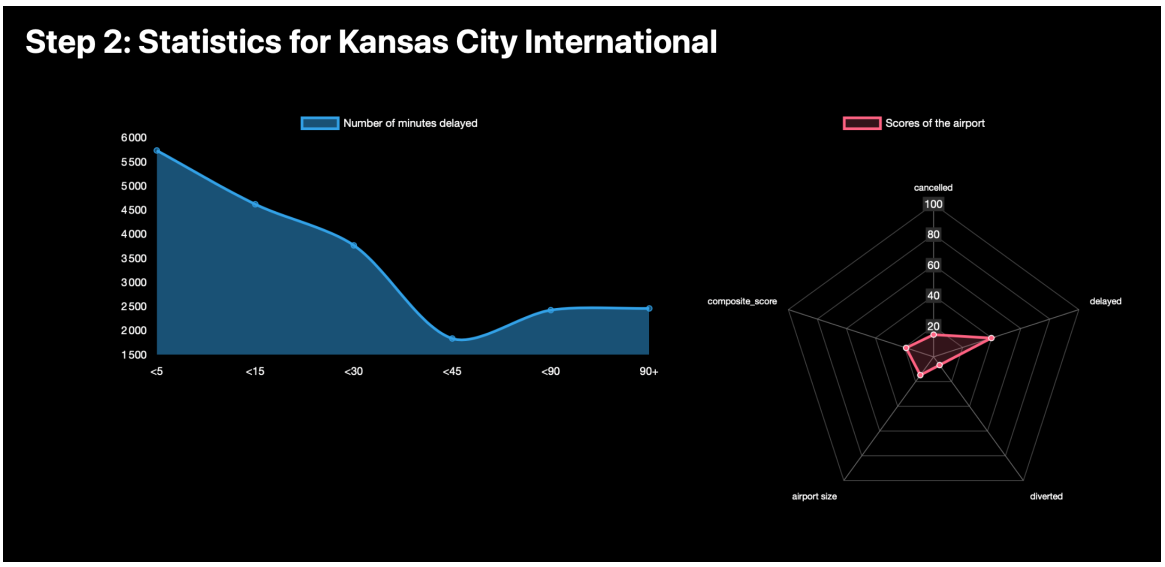

Figure 5: Picture of the statistics part

Challenges We Encountered

Dataset Challenges

Working with the "On-Time: Marketing Carrier On-Time Performance" dataset from the Bureau of Transportation Statistics presented several challenges :

- **Data Completeness**: Despite being a comprehensive dataset, we discovered significant gaps in several variables during our initial exploratory data analysis. These missing values necessitated careful handling to avoid biased or inaccurate insights.
- **Data Volume**: The dataset's sheer size posed computational challenges, especially when performing complex queries or generating visualizations. We decided to use less data to focus our attention on the visualizations.
- **Variable Selection**: With lots of features available, selecting the most relevant variables for our analysis was very important. We focused on variables directly related to flight performance, such as delay causes and delay times. It took some time to properly understand he different acronyms used in the datasets.

Coding Challenges

- **Integrating diverse technologies**: Combining various technologies such as React, Next.js, D3, and Chart.js presented a significant challenge. Ensuring seamless interaction between these frameworks required careful planning and implementation.
- **Optimizing chart load times**: The charts initially took a long time to load, which impacted the user experience. To address this, we reduced the amount of data being processed and introduced a skeleton screen for lazy loading. This approach improved performance and provided visual feedback to users while the charts were loading.
- **Handling opacity issues**: We encountered bugs where some div elements remained active with an opacity of 0, leading to unexpected behavior. Debugging and resolving these issues involved identifying the problematic elements and ensuring they were removed from the DOM when not in use.

Peer Assessment

We worked in a team of three students. Here is a brief introduction of each member and their respective contributions.

- Tomas : Coding of the 2D map and its user-friendly interface - coding of the skeleton - process book report - realisation of the video
- Armance : Coding of the score spider chart - most of the process book report - creation of the sketches
- Diogo : Created the core of the website structure - coding of all the others statistic part - process book report

All parts were reviewed by every team member and we are happy about everyone's contribution. Despite individual task distribution, all choices and decisions were collectively agreed upon by the three team members.

Conclusion

We successfully created an interactive, user-friendly tool for visualizing and analyzing flight data. Using the "On-Time: Marketing Carrier On-Time Performance" dataset, we developed a platform that simplifies access to detailed flight and airport information.

We began with data exploration and cleaning, then designed and built the website structure. Our final product features an interactive map highlighting major U.S. airports and a detailed analysis section of airport performance metrics.

Despite challenges in integrating various technologies, our team collaborated effectively, each member contributing to different project aspects. This project enhanced our technical skills and showcased the value of data visualization in making complex information more accessible.