

Jeong Yong Yang
CDS DS 594
Professor Chamberas
Feb 22, 2024

“Flight Delay Analysis” Critique

“Flight Delay Analysis” uses a range of graphs to showcase various types of delay categories, delays by airport, and delays by airline. Despite the four graphs effectively displaying the delays, the graphs can be enhanced in multiple ways.

The author wisely used a vertical bar chart in “Top Delays by Delay Components”. When comparing the frequency or number of various categories (reasonings behind the delays for this specific graph), bar charts are generally effective since viewers can easily read the biggest, smallest, and incremental differences between categories. Specifically, this chart included the total number of minutes on top of every bar, allowing the audience to easily figure out the highest and lowest delay components with exact numbers. Additionally, the design choice to have a zero baseline to avoid false visual comparison and the appropriate thickness of bars (wider than the white spaces between them) contribute to its effectiveness. However, “Top Delays by Delay Components” can be improved by introducing various colors, transitioning into a horizontal bar graph, and ordering the graphs by size. The addition of different colors for each category can make the graph more visually appealing since it allows the audience to interpret the visualization more easily. Switching to a horizontal bar chart would allow the audience to read the categories initially since most of them are more comfortable with the left-to-right reading flow. In other words, they are introduced to the category names before the actual data, allowing them to already know what each bar represents by the time they get to the data. Sorting the data from highest to lowest (or lowest to highest) would enable the audience to quickly identify the smallest and largest categories, simplifying the analysis in one glance.

“Top Delays by Time of Year” displays the same categories as “Top Delays by Delay Components” but presents the minutes of delay per month by utilizing area charts. Although the visualization uses a horizontal layout, the decision to use an area chart may not be optimal. Area charts are more useful for demonstrating cumulative data over time, but this chart simply compares delay categories per month, making a line graph more effective to avoid the implication of cumulative data from the areas filled beneath the lines. Moreover, the area chart has a different y-axis number for each graph. For instance, Airline Delay has a y-axis of 1000, and Air System Delay has a y-axis of 400, which can confuse the audience when comparing delays across categories since they might misinterpret the severity of delays due to the different scalings.

“Top Delays by Airport” and “Top Delays by Airline” show a comparison of different metrics by airline and airport through multiple series of horizontal charts. When analyzing the relationships between different metrics, scatter plots are generally a great choice because they are

effective when showing relationships between the metrics, especially when there exists numerical paired data (examining if and how one metric affects another). Finally, if we want to compare all metrics to each other, we can use a pair plot, which is another variation of a scatterplot, since it allows a comparison of three or more metrics across one or more categories.

“Flight Delay Analysis” displays four different data visualizations regarding flight delays based on categories per year, categories per month, airlines, and airports. While the charts effectively convey the extent and nature of delays, they contain several opportunities for development to improve clarity and interpretation.