

[Portfolio] Component II: Data Visualization

Figure 1

Graphic inquisition

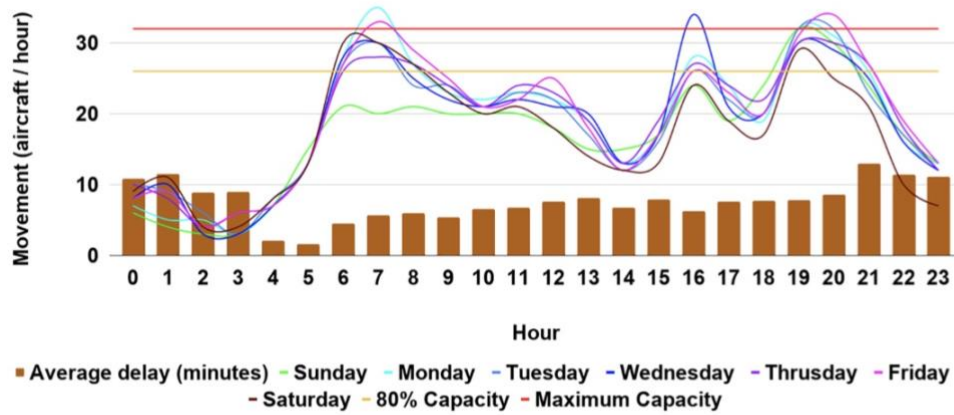


Fig. 9. Maximum actual demand, airport capacity, and average delay at SBGR in 2017.

Note: This image was retrieved from Pamplona and Alves (2020).

Pamplona, D. A., & Alves, C. J. P. (2020). An overview of air delay: A case study of the Brazilian scenario. *Transportation Research Interdisciplinary Perspectives*, 7, 100189. <https://doi.org/10.1016/j.trip.2020.100189>

Graphic inquiry

Gestalt Principles: In the figure (Figure 1) the authors make use of the Gestalt principle, *common fate*. The demand variable for different days is moving in the same direction, so we perceive these as part of the same group. Same principle for the bars measuring delays in minutes.

Decoding and operations: There are grid lines in place to help the viewer decode the values to do further operations. The element lines, used for plotting symbols, are however on top of each making it hard to tell them apart. The color scale also makes decoding hard. The y-axis isn't scaled properly ending at 30 when the value of *Maximum capacity* (red line), and most of the demand variables, indicate higher values than that for some hours of the day. On the y-axis, they're using the same values for two different units. This gets confusing when decoding the values from the bars in the plot because you have one element that's easier to decode (lines) along with an element that's harder to decode (bars) on a common scale measuring in different units.

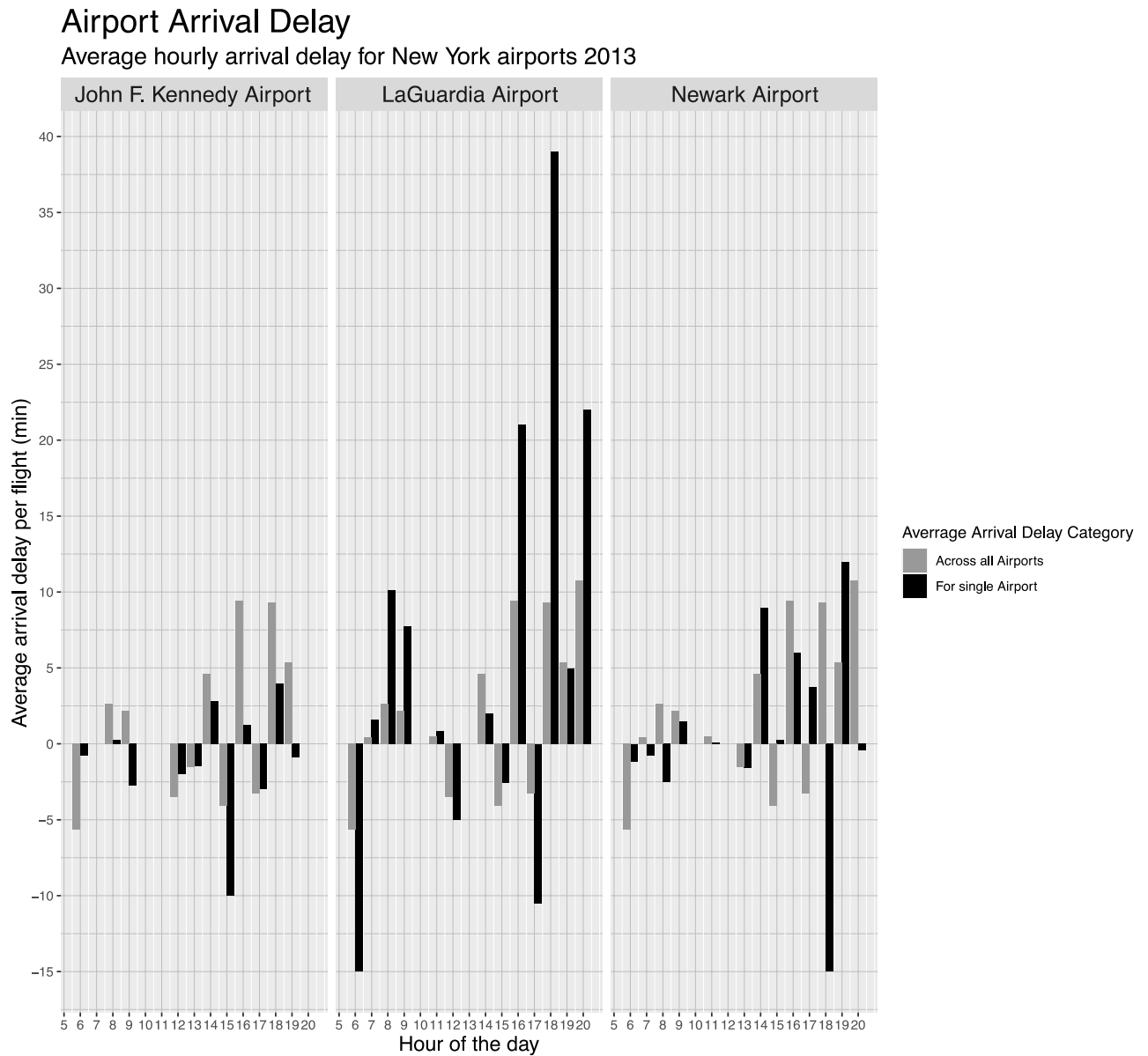
Annotation and stand-alone readability: Most of the variables are marked. The figure title (Figure 1) also provides enough information to be interpreted outside its context (the article). Preferably, the *Maximum Capacity variable* could've been named the same thing in the figure title as it is in the notes.

Graphical data integrity and lie factor: It doesn't seem as if the author has misrepresented their data. The x-axis contains all hours of the day, not creating large increases/decreases vertically. It seems as if they've adhered to the 2H:3L aspect ratio.

Chart junk: The data-to-ink ratio is quite low. Because they include many variables in one panel, they use a lot of colors (=ink). Also, they don't clarify the use of the 80% capacity variable. They might mention a reason in the article, but as a viewer, it's not clear. Also, if they'd scaled the y-axis properly, I would've been able to read the value of the maximum capacity, making computation of the 80 % of the maximum capacity easy.

Figure 2

Graphic design



Note: This graph is generated using a dataset with 150 entries for the year 2013. Outliers in the data have also been removed. An outlier is defined as being $1.5 \times$ Inner quartile away from the 2nd and 1st quartile. After doing that, the remaining number of entries were Newark (58) John F. Kennedy (47), and LaGuardia (34). The average arrival delay per flight is computed over the number of flights remaining after outliers were removed. No data was available for the 10th hour, and for each airport where no flights for giving hour were given, there's also no Average Across all Airports displayed. Data named "flights.txt" was downloaded from the UiO canvas folder.

Graphic design

I chose to display the average arrival delay per flight for each origin airport. I also chose to display the average arrival delay across all airports for comparison.

Decoding and operations: I chose to include gridlines to easily derive values from the figure for further computation. And I put the bars I wish to be compared next to each other so that the viewer doesn't have to move objects mentally, referencing the term, *Superimposition*. I chose a color palette that would be accessible even to those with color blindness.

Annotation and stand-alone readability: I labeled each panel, gave the figure a title, and labeled the legend so that the average category was clear. I noted in the figure text how the data was processed before being visualized.

Graphical data integrity and lie factor: I chose to create a variable mean arrival delay per flight to be able to compare the three airports. One could argue that it's not fair to LaGuardia having the least departing flights and so one single delay, excluding the outliers had a massive effect. But then again, this is not a representation of reality anyway. I note that the number of entries from which the mean was calculated for each airport is in the figure notes (Figure 2).

Gestalt principles: I chose to include the mean for all airports in each of the panels. I think this helps the viewer put the values for each panel into perspective, and I also use the same plotting symbols for all averages conveying that they're symbolizing the same unit (arrival delay per flight (min)). Adhering to the gestalt principle, *similarity*.

Chart junk: I debated including the average across all airports because the viewer can calculate this using the graph. But I kept it, thinking it visualizes a comparison directly to the viewer.