Portfolio for MAE4000
Component 2: Data visualization
Otávio Mattos

Part 1: Graphic inquisition.

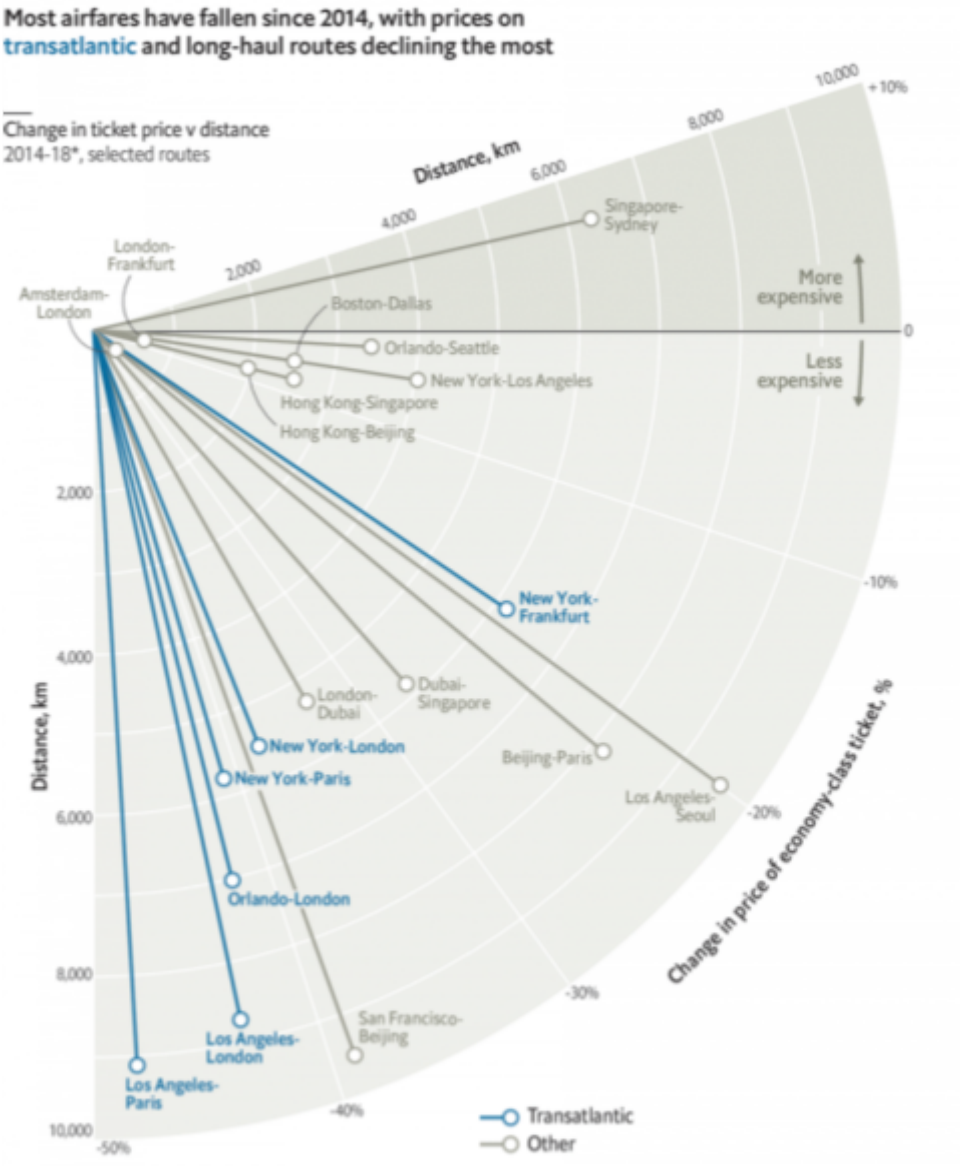

Figure 1. The graph aims to show a relation between distance (in km) and change in price of economy-class tickets (in percentage) ("Why ticket prices", 2018)[1].

[1] Online newspaper article without author, cited according to APA Citation Guide (7th Edition).

The graph in figure 1 was published as an online article from The Economist ("Why ticket prices", 2018). It aims to show a relation between distance (in km) and change in price of economy class flight tickets (in percentage). The graph contains multiple data visualization problems. We summarize some of them below.

First, a key problem is that the graph contains too much visual information to process. For example, we can see 15 lines marking the relation between distance and airfare change (D&FC) for different flights. Different flights, in turn, are indicated through 19 dots and labels. To facilitate the comparison of D&FC across flights, the authors could have used the strategy of "small multiples" (Bock, n.d.) to avoid multiple lines together, or simply used a scatter plot with a regression line where each dot stands for a flight.

The graph also brings additional comparisons, making it overly "busy". The lines were distinguished by color, i.e. blue for "transatlantic" and dark gray for "other", suggesting that we should also compare D&FC between these categories. A line graph or a scatter plot with one regression line for each category could have facilitated cross-category comparisons.

We also think that the "temporal component" of the data could have been implemented differently. Instead of having the variable "airfare change" with percentages, they could have shown different scatter plots with the relation between distance and airfare *per year* (i.e., a scatterplot for each year, from 2014 to 2018). In this way, we would have been able to see how airfare has been falling throughout the years.

Last, but not least, the relation between distance and airfare change in their graph requires us to consider two visual dimensions: the *direction* of the lines (to indicate fare change), and *length* (to indicate flight distance). A standard scatter or line plot would have just required direction, making the "decoding" of the correlation simpler.

Part 2. Graphic design

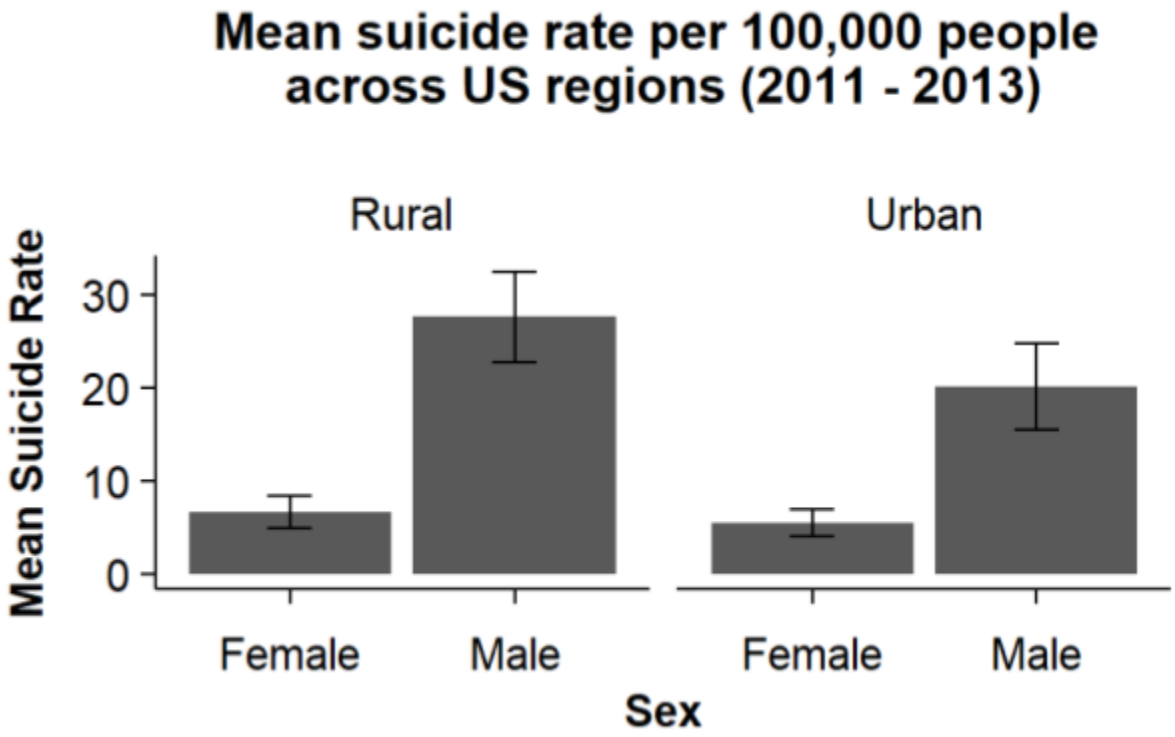

Figure 2. Mean suicide rate per 100,000 people across 10 HHS regions in the United States (HHS: Department of Health and Human Services), between 2011 and 2013. Suicide rate distinguished by sex and rural-urban status. Error bars stand for standard deviation. Data publicly accessible in R under the name of "USRegionalMortality".

Figure 1 intends to show the mean suicide rate per 100,000 people across HHS regions (Department of Health and Human Services) in the US. The data was collected between 2011 and 2013. The bar graph involves three variables: sex, urban-rural status and mean suicide rate. It shows that suicides among men are, on average, higher than among women across HHS regions. The graph also seems to indicate that the suicide rate among men in rural areas is slightly higher than among men in urban areas (note, however, that the standard deviation for rural and urban men partially overlaps; not all rural regions regions have higher male suicide rates than urban regions).

Since our goal was to make cross-category comparisons (sex and urban-rural status), we decided to use bar graphs — ruling out graph options that are especially tailored to explore relations between different numerical variables (e.g., mean suicide rate and time). We could have used box plots instead of bar graphs. Box plots would have allowed us to show the dispersion of regional suicide rates from the median. However, we opted to use bar graphs for the sake of simplicity: we wanted our graph to convey the message that, overall, individuals from some categories are more prone to commit suicide than others, rather than reflect on regional variability within categories.

Finally, we decided to have a black and white graph since color distinctions would be informatively redundant with the graph labels. Finally, we decided to distinguish statuses through panels and sex through bars, instead of the opposite. The main suicide rate difference is between men and women and having sex distinguished through bars highlights this distinction —after all, we show it two times, one for each urban-rural status.

References

Block, T. (n.d.). The Psychology of Small Multiples. *Displayr Blog*. https://www.displayr.com/the-psychology-of-small-multiples/

Why ticket prices on long-haul flights have plummeted. (2018, December 8). *The Economist*. https://www.economist.com/graphic-detail/2018/12/08/why-ticket-prices-on-long-haul-flights-have-plummeted