Interactive-Visual Data Analysis

Zejie Guo

November 12, 2023

1 The terrible visualization

The poorly designed visualization I found is from RFI News. It presents data on grain exports from Ukrainian ports in the Black Sea since August 2022. See RFI News.

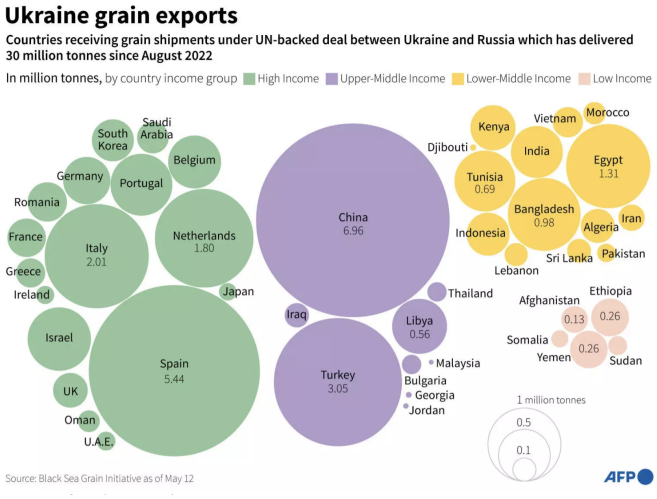

Figure 1: Poor VIS

1.1 Data and task (what and why)

The data in this visualization comprises three attributes: the amount of grain shipments, the name of the country, and the World Bank income level. These attributes, in turn, belong to quantitative, categorical, and ordinal attribute types.

The main task is to identify the primary beneficiaries of grain shipments across different income groups and within each income group.

Users can accomplish several tasks using this visualization. For instance, they can look up the largest grain-importing country in the high-income group. They can also locate the largest grain-importing country by observing the area and label of the circle mark.

1.2 Decoding (how is it shown)

The area of the circle marks is used to encode the quantity of grain shipments, and three different circle areas are specified in the legend (see the legend at the bottom right). However, there are far more than three different values to show in this visualization, which makes it hard to perform the comparison task both within and across the income groups. In Munzner's book, she mentioned that "[the linewidth] would be a poor choice for dozens or hundreds of values. The key factor is matching the ranges: the number of different values that need to be shown for the attribute being encoded must not be greater than the number of bins available for the visual channel used

to encode it"[1]. This also applies to the circle area. In this visualization, the number of circles for the quantitative data being encoded is far greater than the number of circles available for the visual channel used to encode it, in this case, there are only three.
In addition, the circles in each group were randomly placed, which made it more difficult to compare shipments within groups.
Though some value labels were put on top of the circle marks, as shown in the mark for Spain, China, etc., for the marks that are smaller than the one that represents 0.1 million tons in the legend, there are no labels at all, for example, the mark for Malaysia.
Also, just putting the name of the country on top of and around the marks makes the task of locating the smaller import countries very difficult, and some small circles and their labels are placed too close to discern which label is for which circle.
Color hue is used to encode the ordinal attribute of income levels in this visualization. However, color hue is the identity channel, which is used to encode the categorical attributes as mentioned in Munzner's book[1].
In the representation of country labels, the design violates the interface design rule of 'striving for consistency', as Shneiderman points out, "consistent color, layout, capitalization, fonts, and so on, should be employed throughout"[2]. For instance, some countries, such as the United Arab Emirates, are abbreviated (e.g., U.A.E.), while others are presented in their full names. Even though abbreviations like 'UK' for the United Kingdom are familiar to most people, the same cannot be said for 'U.A.E.' I decided to maintain consistent labeling by using the full country name for all countries.

2 The improved visualization

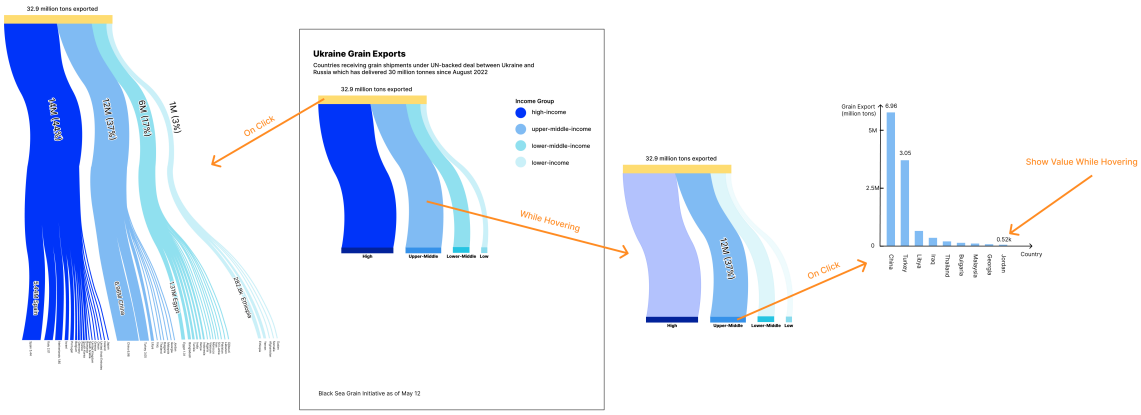

Figure 2: Improved VIS

See Figma Demo.

2.1 Decoding

I used the vertical layout hierarchical flow to represent the overview of grain export combine with four bar charts to show the grain shipments for each income group.
In the Overview, the hierarchical flow shows the movement directions of the trading data, and the amount of grain shipments is encoded by the width of the lines, with colour saturation representing four income levels. This is according to the expressiveness principle that ordered data should be shown in a way that our perceptual system intrinsically sense as ordered [1].
By clicking on each of the income groups, the grain shipment information will be displayed as a bar chart, which

encodes two attributes using a line mark with the vertical spatial position channel for the quantitative data, the grain shipments, and the horizontal spatial position channel for the categorical attribute, the countries. When hovering over the bars, the value is displayed at the top of the bar.
In this way, the user can easily distinguish between shipments to different countries. In addition, I have sorted the countries by shipments, so it is easier to find the largest or smallest grain importers than in the initial visualization.

2.2 Motivation behind design choices

In the overview, the highlighting and the labeling of lines could help users effortlessly obtain the most important information, — specifically, the primary beneficiaries of grain shipments both within and across income groups.
When examining the export details for each income group, the length of bars proves to be a more effective means for users to accurately pinpoint a particular country compared to the original circle marks. Additionally, based on the error rates across visual channels[1], it's evident that the accuracy of the length channel surpasses that of the area. That's why I chose length as the encoding mechanism for shipment quantities. Conversely, in the overview, I utilized area of the flow (line) to maintain design consistency, given that the primary purpose of the overview is to illustrate the main direction of shipments.
Color saturation is employed to signify the ordered income levels, minimizing cognitive load for users. They can readily perceive that lighter colors correspond to lower income levels.

References

[1] T. Munzner, *Visualization Analysis and Design*, ser. AK Peters Visualization Series. CRC Press, 2014. [Online]. Available: https://books.google.ch/books?id=dznSBQAAQBAJ

[2] B. Shneiderman, C. Plaisant, M. Cohen, S. Jacobs, N. Elmqvist, and N. Diakopoulos, *Designing the user interface: strategies for effective human-computer interaction.* Pearson, 2016.