Data visualization

I *Scientific and mathematical basics*

II ***Design principles***

Design principles

In order to achieve the greatest possible significance with the help of a representation, the right choice of design is required. A wrong choice of the design can turn the statement of a representation completely around and lead to wrong conclusions. Visual design must pursue the following characteristics.

- Visual design must be informative.
- Visual design must be accurate.
- Visual design must be descriptive.

Storytelling with data

A proper and structured approach is the basis for telling stories with data. In Chapter I Scientific and Mathematical Basics, point 2, the correct choice of diagram, has already been dealt with. In the Design Principles chapter, points 3 - 6 are covered.

1. build data understanding
2. correct chart selection
3. **data cleaning**
4. **focus**
5. **develop designer mindset**
6. **storytelling**

Exploratory and explanatory analysis

In data analysis, a basic distinction is made between explorative and explanatory analysis. In addition, descriptive and causal data analysis can be mentioned, which are based on explorative and explanatory analysis. Depending on which research status is available for a topic or data set, this has a decisive influence on the approach.

Exploratory analysis

Exploratory analysis is flexible in structure and is used in areas about which little or no knowledge exists. It is the starting point to get an overview. Exploratory analysis is the basis for descriptive and causal research, which is not flexible.

- Searching for relationships and features in data structures.
- Formulation and testing of hypotheses.

Explanatory analysis

The explanatory analysis derives hypotheses from the analysis, tests them or provides findings, and then links them to a "story". These stories provide the answer to the question formulated at the beginning. Often, the explanatory analysis or its presentation is the product of the exploratory analysis.

- Elaboration of insights
- Basis of storytelling or answers the original question

Data analysis

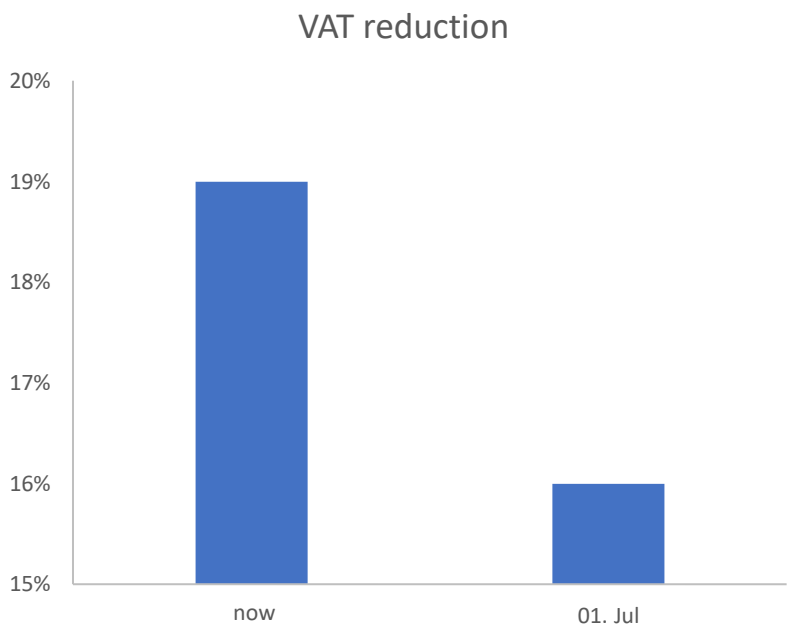
Data analysis consists of five 5 subtasks that must be performed recurrently as the data set expands or changes.

- **Extraction** from the data source
- **Cleaning** of the data
- **Exploration** of the data
- **Analyzing** the data
- **Sharing** the findings

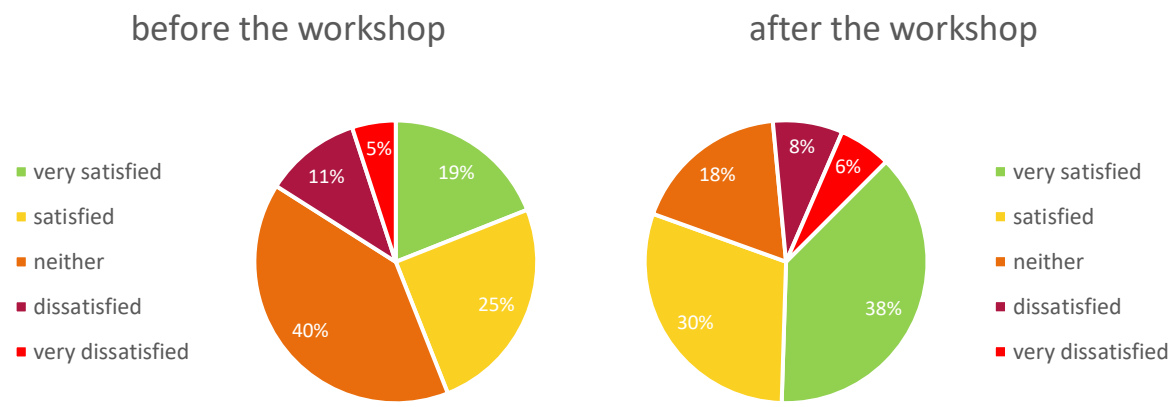
Poor visualization

But what makes a visualization a bad visualization? One speaks of bad visualization when the core message is misleading, covering up or distracting.

Using the example of the VAT reduction on July 1, 2020, one can explain a misleading visualization very well. The reduction looks much clearer in the depicted representation than it actually turns out to be.



Another example deals with employee satisfaction before and after a workshop. The pie chart makes it very difficult for the reader to draw the right conclusions quickly and easily. The number of different colors and the correct assignment of the individual areas cover up or distract from the actual core message.



Visualizations should have only as much additional information as necessary. All superfluous information that does not support the core message must be removed. Elegance and simplicity should be paramount.

Visual Encoding

The term visual encoding refers to the mapping of data within a diagram. This can be done in several ways or the existing data structure can determine the type of mapping. By the choice of the representation type, different conclusions result.

The following visual encoding types exist.

- **Position** on the x and y axis
- **Size** of points and bars
- **Shape** and **structure** of diagram elements
- Use of **angles** and **lengths**

A special characteristic to be mentioned is the use of colors. People are strongly attracted to color contrasts, yet the use of **colors** should be limited. The use of colors should be used as an aid in focusing on a particular chart element or inference derived from it.

Graphic waste

Graphic garbage or superfluous diagram elements confuse the viewer and distract from the actual core message. A visualization should have only as many visual elements as possible or as few as necessary to allow clear conclusions.

The following points must be avoided

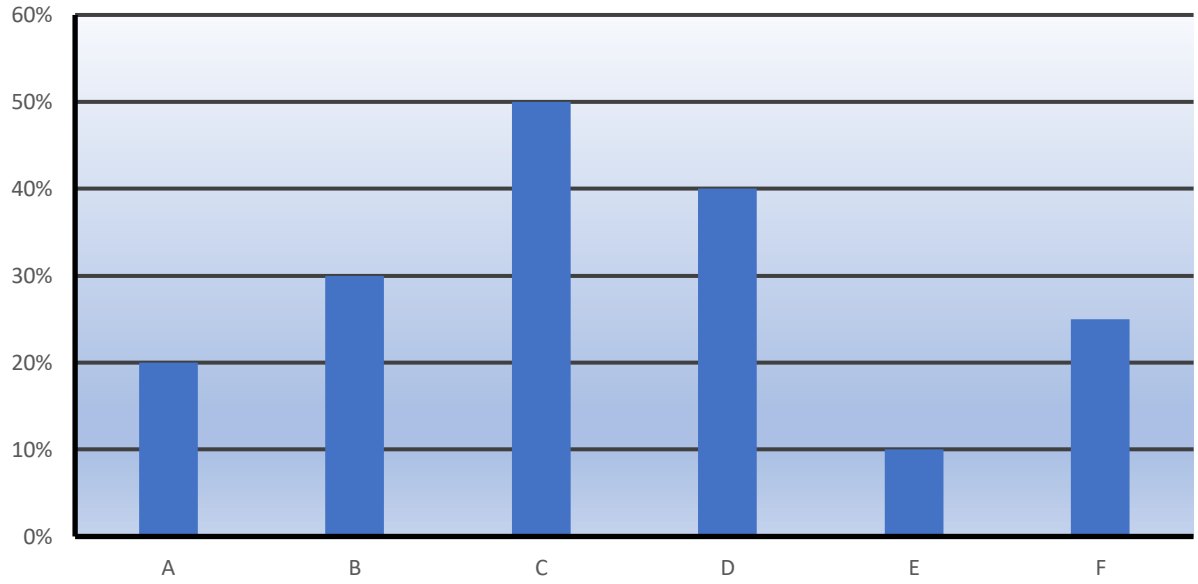
- Excessive dark or thick guides
- Unnecessary text that provides no information
- Images in diagrams
- 3D model or shadow

Data-ink ratio

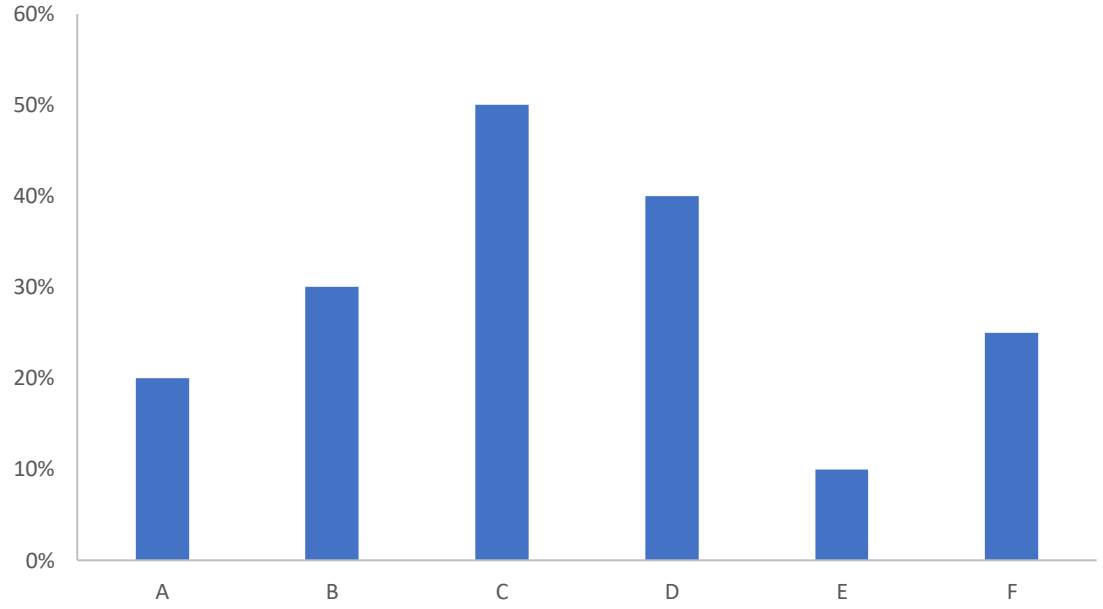
Data-ink ratio describes how much ink can be replaced by pixels. The higher the ratio, the better the data description.

$$Data - Ink = \frac{ink\ used\ for\ data\ declaration}{ink\ used\ to\ explain\ the\ rest}$$

low data-ink value



high data-ink value

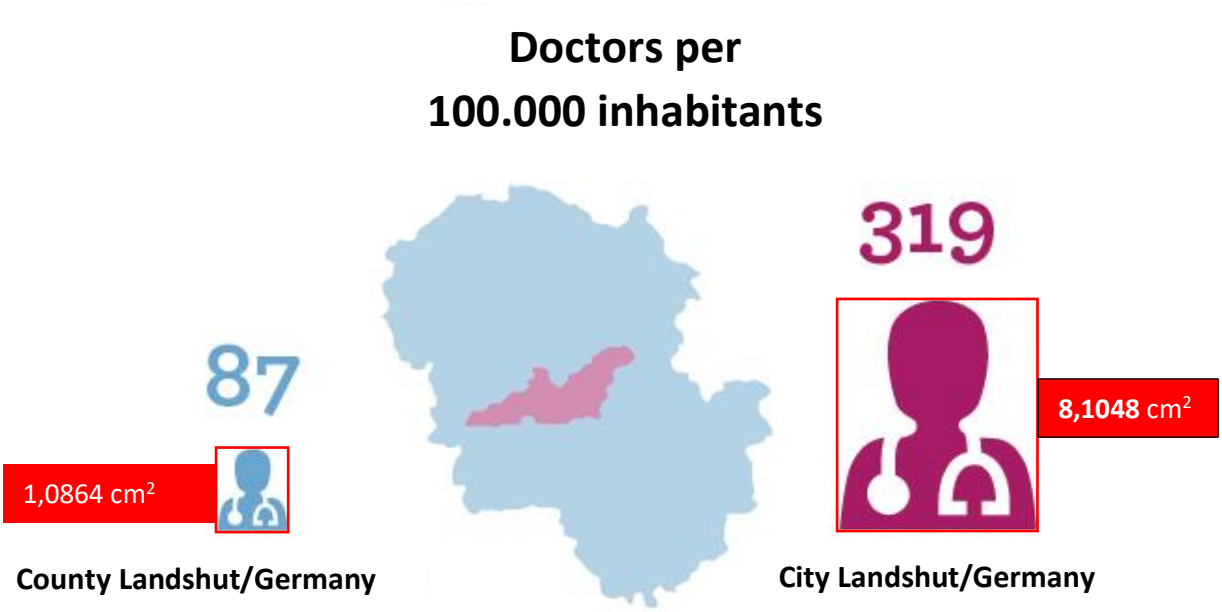


The chart with the low data-ink value, unlike the visualization with the higher data-ink value, uses much more ink (labels) to explain the chart, which only unnecessarily confuses or distracts the viewer.

Bias factor

The bias factor, which is part of data integrity and thus describes the correctness, completeness and consistency of data, indicates how much the visual representation differs from its actual data basis. Often, comparisons between two values are graphically supported by symbols, yet these do not often correspond to the symbols of their actual value variables. The distortion factor indicates how much the representation differs from its value. A distortion factor greater than 1 can mislead the viewer.

$$Bias\ factor = \frac{\Delta\ visual/visual_{start}}{\Delta\ data/data_{start}}$$



Quelle: https://www.lass-dich-nieder.de/angebote/zahlen-fakten/hausaerzte-sind-in-sachsen-am-juengsten.html

$$chart\ number = \frac{319 - 87}{87} = 2{,}67$$

$$chart\ area = \frac{8{,}1048 - 1{,}0864}{1{,}0864} = 6{,}46$$

$$bias\ factor = \frac{6{,}46}{2{,}67} = 2{,}42$$

The bias factor of 2.42 says that the chosen size of the symbols does not reflect the ratio of the data comparison and thus misleads the viewer.

Color

The use of color can be both useful and not useful. Three effective tips for the sensible use of color.

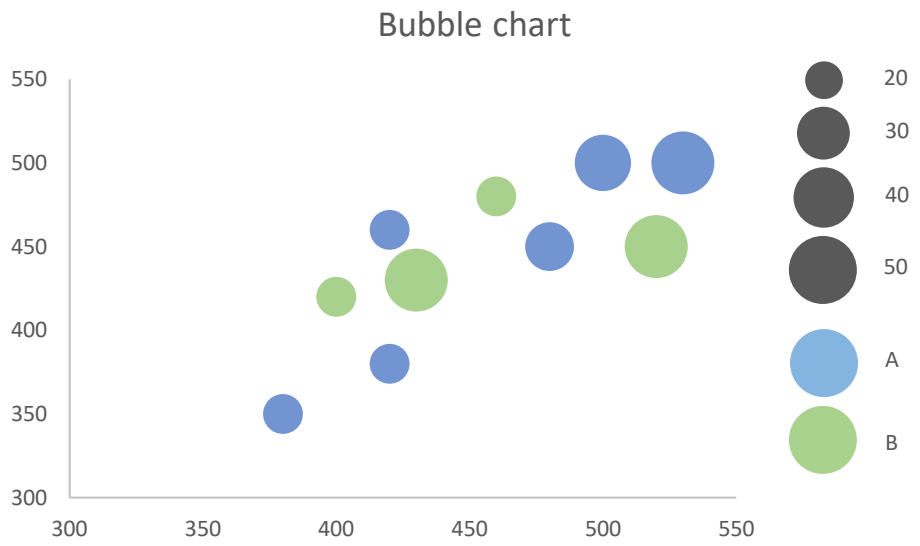
- Before using color, check whether black and white tones are not sufficient?
- The use of natural pastel tones makes it more pleasant for the observer to look at the visualization for a longer period of time. The use of bright and intense colors unnecessarily strains the human eye.
- Color should be used as a means of communication or to emphasize data.

Note: When using color, always keep in mind that about 10% of men and 1% of women are color blind and thus have a green-red deficiency. Therefore, it is always advisable to use black, gray and white tones.

Shapes and sizes

Often you have a lot of quantitative data and to present all this information in a visualization you need additional tools. Therefore, it is advisable to include shapes, as well as their size and color in the visualization.

Colors and shapes can best support categorical variables. In contrast, the diameter of a shape is used to represent quantitative variables. The example of this bubble chart helpfully illustrates the use of shapes, sizes and color.



Note: Before attempting to represent a variety of different information in one visualization using shapes, sizes and colors, it is advisable to divide the information and the core message into several diagrams.

General tips

To avoid bad visualization, you need to follow these tips.

1. adherence to a high data-ink value and avoidance of superfluous information.

2. choosing the right visual encoding to highlight the highlights

3. maintaining data integrity in diagrams.

If the viewer does not immediately draw the right conclusions from the visualization, then the visualization is incorrect or too complicated in structure. Remove the excess of information and focus on the highlights.

Storytelling

To come to the right conclusions, you need good visualization as well as a good story. One cannot lead to the right conclusions without the other. When building the story, the following points should be considered.

1. formulation of the problem or the question arising from the problem.
2. justification of the questions with further examples (deep dive).
3. working out the solution and answering the initial question.
4. deduction of decisions and fields of action resulting from the solution.

Summary

The correct use of design principles has a decisive influence on the conclusion or on the answer to the initial question. The correct use of visual encoding and the avoidance of unnecessary information in visualizations can increase the message enormously. Finally, the presentation should always go hand in hand with a good story to derive the right conclusions or decisions.