DATA VISUALIZATION TRAINING SESSION

Summary of the data visualization training session held at the HELCOM Secretariat on 28th November 2018

11th December 2018

Organized by:
Manuel Frias

Table of contents

page 3
Why this training and what's next

page 4
Part 1 What is an effective data visualization

page 6
Part 2 What is the most appropriate graph for a dataset

page 11
Part 3 How can we improve the style of a graph

page 12
Three book recommendations

page 13
Sweet end

Why this training and what's next

What is an effective graph and how can we improve our data visualization skills? HELCOM publishes lots of graphs in many reports every year but we admit many of them have room for improvement.

Unfortunately, none of us at the HELCOM Secretariat has been taught data visualization skills. Despite this fact, we publish countless pages with innumerable graphs and visualizations.

The aim of this training session was to reflect about data visualization, particularly graphs. Are there any basic rules? How can we improve our skills to make HELCOM reports clearer to understand and more attractive?

This training was two hours long divided in three parts:

1.What is an effective data visualization

2. What is the most appropriate graph for a dataset

3.How can we improve the style of a graph

We decided to hold informal meetings in the future to get feedback from our visualizations.

The presentation used in the training is on Slideshare

1 - What is an effective data visualization

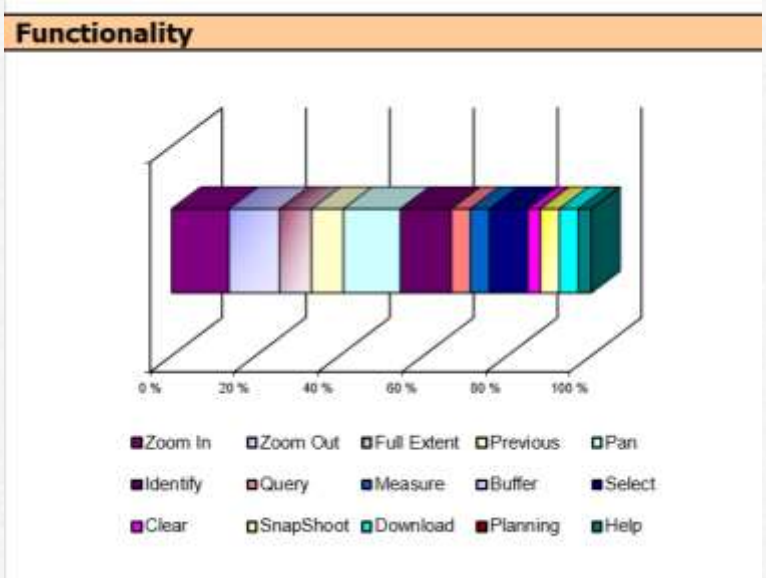


Book recommendations at the end

The session started by defining an effective data visualization. According to designer and educator Alberto Cairo, it should be:

- Truthful
- Functional
- Beautiful
- Insightful
- Enlightening

I showed a graph I made some years ago and we together discussed whether it met those criteria. We criticized the use of 3D, the stacked bar, the colors… All in all, it is virtually impossible to get something out of it.



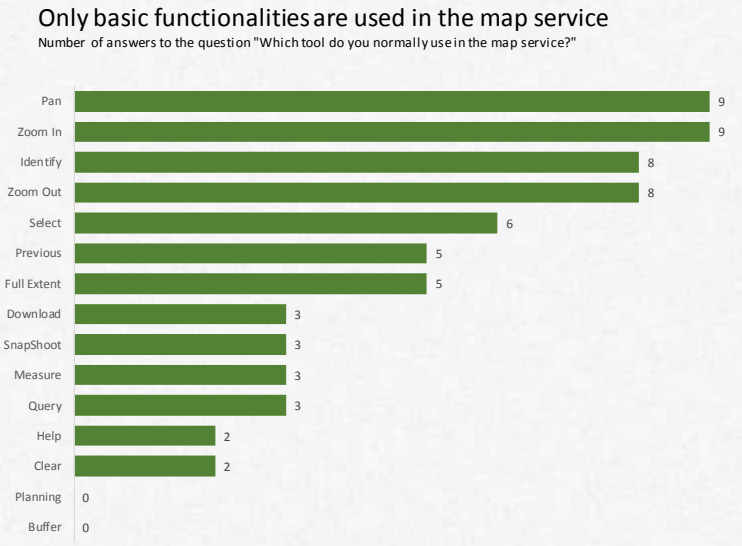
Icon by Arif Arif TheNounProject

We agreed that the graph had many problems and that it definitely didn't meet Cairo's criteria, except for being (or trying to be) truthful.

I showed a remake with a simple bar chart and we all agreed that it communicated the message much better than the previous one.

What was I thinking when I made the first one? Probably I didn't think much. I just wanted to make an attractive graph without thinking in the viewer.

What happened between both graphs? I just applied basic principles of data visualization.



2 - What is the most appropriate graph for a dataset

How do you go from the first horrendous version to the second simpler but more effective visualization?

First, we should focus on finding the message. What do we want to say with the graph? For whom is it ? Do we want to emphasize anything? Do we want to show comparison, deviation…? In my first graph I skipped this step entirely.

Second, we should find the shape of the data. To do this, we have two alternatives.

The first one is to use a chart chooser. I recommended three. However, some have been criticized by authors like Stephen Few for offering some charts types that are not effective. I think, however, that they could be a good starter for newbies.

https://datavizcatalogue.com/



https://www.data-to-viz.com/



http://ft.com/vocabulary



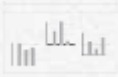
The second option, recommended by authors like Cairo, is to use the Cleveland McGill scale. This is arguably the most famous scale in the data visualization world.

It shows what kind of graphs are better to perform accurate and general comparison. Not surprisingly, the most effective graphs to show accurate comparisons are bar and line graphs.

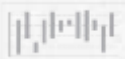
Notice that this research was done only for statistical charts. Shading and hue can be good for showing change in maps.



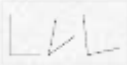
Position along common scale



Position along common but unaligned scales



Length



Angle



Are
a



Volume



Shading



Hue

Enable accurate comparisons

Enable general comparisons

Cleveland, McGill, 1984 adapted from Alberto Cairo, 2016: The truthful art

Bar, line, point (scatterplots) and boxes charts are, in fact, the most appropriate ways to perform quantitative analyses, according to Stephen Few.

Oh but wait… Where are pie charts that are used so much in HELCOM? Bad news—according to cognitive science, our brains are not very good at comparing areas, as you see in the Cleveland McGill scale.

Some authors think they are excellent… for eating them! Nevertheless, I think they can be effective in some counted situations depending on your message.

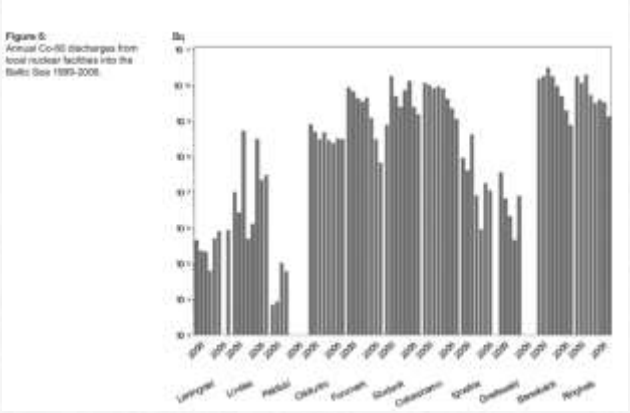


Source: unknown

Next, we made three groups to analyze three charts that have been published in HELCOM. Our aim was to answer two questions:

- Is it effective? Why?
- Is there a better alternative? Sketch it

After thinking together 20 minutes we came up with three new more effective visualizations. Among the solutions proposed were: avoiding clutter, choosing another type of graph and dividing the information in pieces.



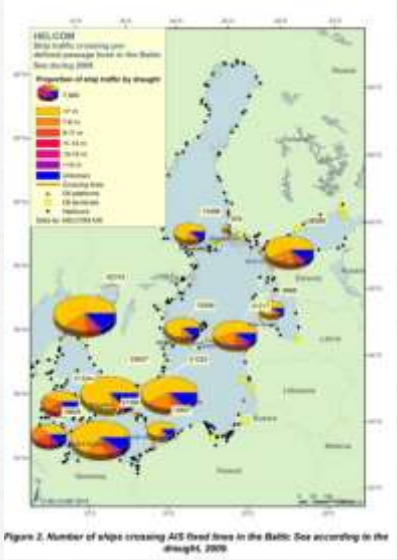


Figure 2. Number of ships crossing AIS fixed lines in the Baltic Sea according to the drought, 2009

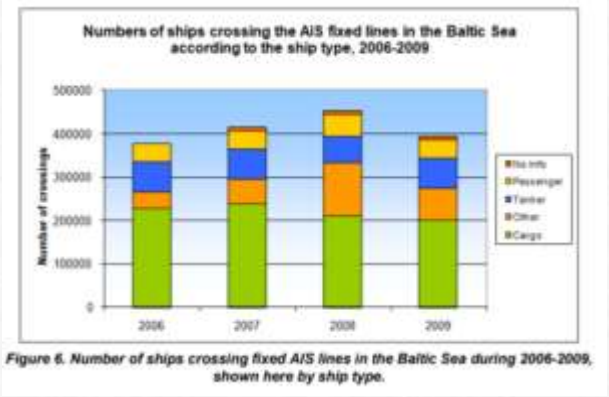


Figure 6. Number of ships crossing fixed AIS lines in the Baltic Sea during 2006-2009, shown here by ship type.

After the exercise, we discussed alternative ways to visualize data like, for example, slope graphs, dot plots and chord diagrams. The challenge with theses graphs is that they are not so straightforward to do in Excel 2016. It is, however, not impossible as designer Jorge Camões nicely demonstrates in his book (reference at the end).

Figure 1.15. INTERACTION BETWEEN PORTS (ARRIVALS AND DEPARTURES) IN THE BALTIC SEA
Passenger ship traffic in 2015: 50 biggest ports, minimum of 500 trips



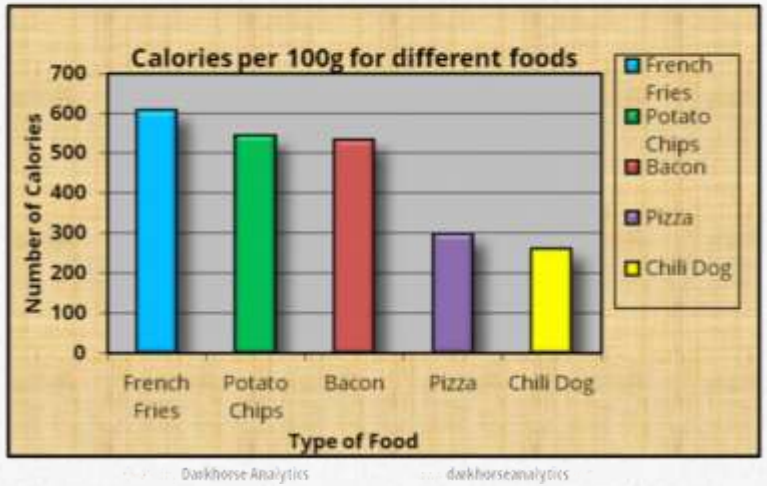
3 - How can we improve the style of a graph

In this section, we discussed how to improve the styles of the graphs we make at the Secretariat. A graph can be more visually attractive if we follow some basic design principles like, avoiding clutter, apply proper alignment, use contrast efficiently, etc.

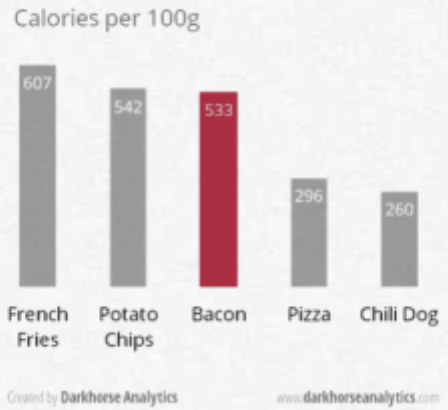
I showed a gif that become popular a while ago in social media and that summarizes nicely how to make a more effective graph in some simple steps.

Finally we talked about alternative tools for exploring (Inzight) and visualizing data (Flourish and D3)

Before



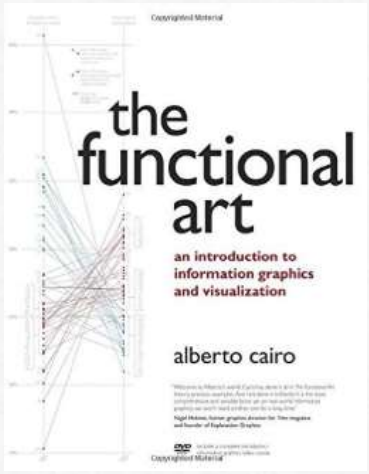
After



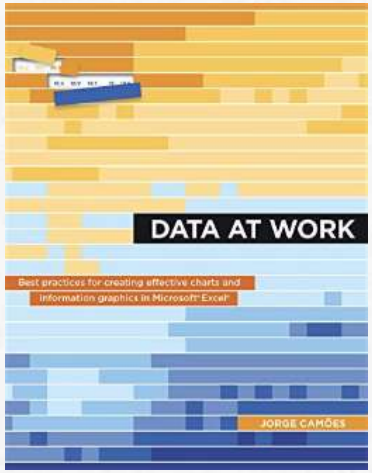
Source: Darkhorse Analytics

Three book recommendations

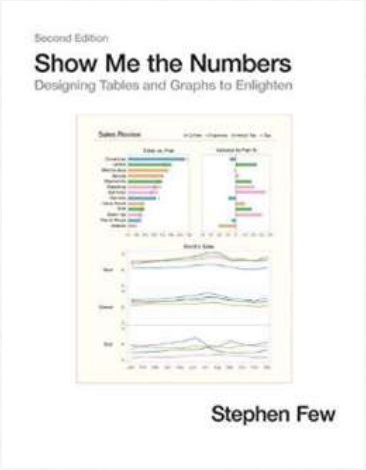
There are endless books about data visualization out there. Here there are three that might interest you.



The functional art
by Alberto Cairo
Clear introduction



Data at work
by Jorge Camões
Anything is possible in Excel



Show me the numbers
by Stephen Few
(Practical) bible for dataviz

Sweet end

The session felt better with some buns. Unfortunately, I was late to take a a pic and these are the leftovers. It was eaten by:

Alexey Bakhtov, Jana Wolf, Joni Kaitaranta, Juuso Haapaniemi, Katarzyna Drożdziel, Owen Rowe, Sanna Saari, Manuel Sala Perez and Florent Nicolas

