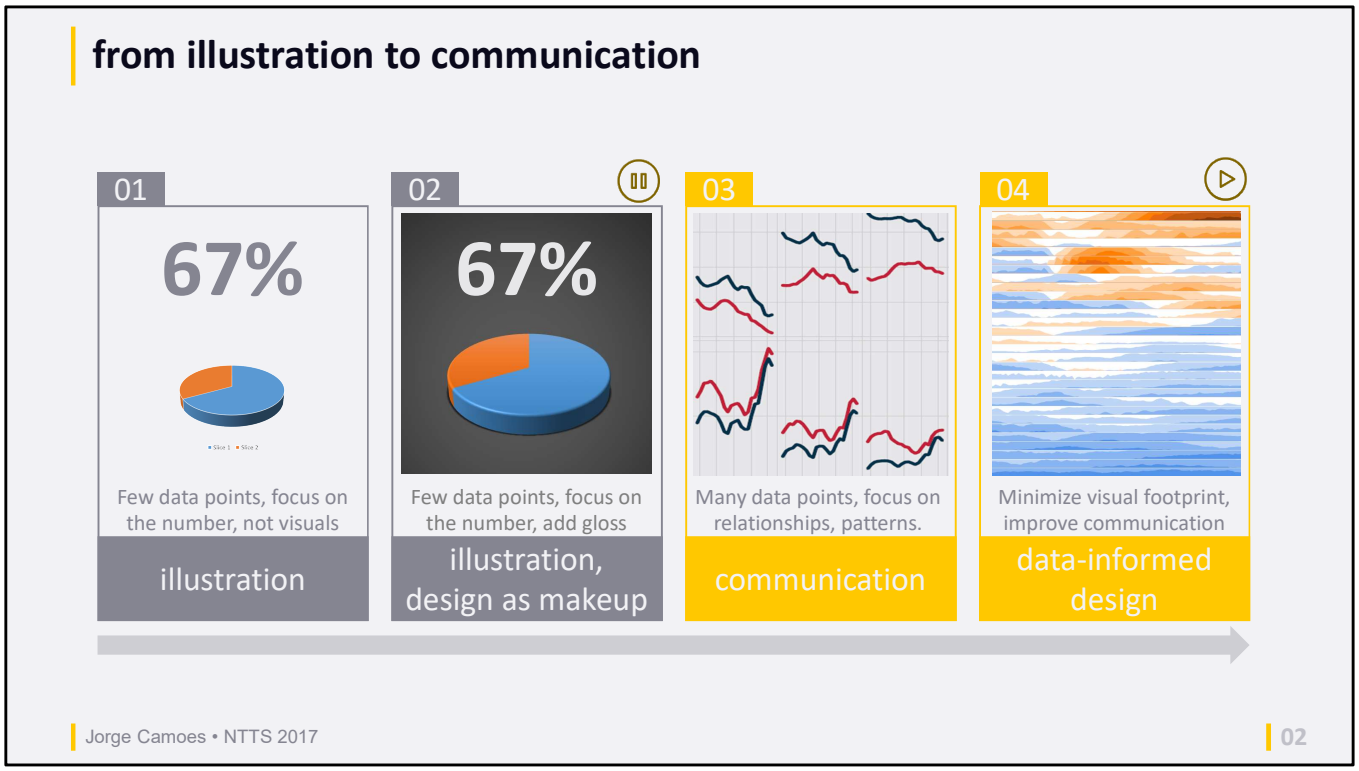
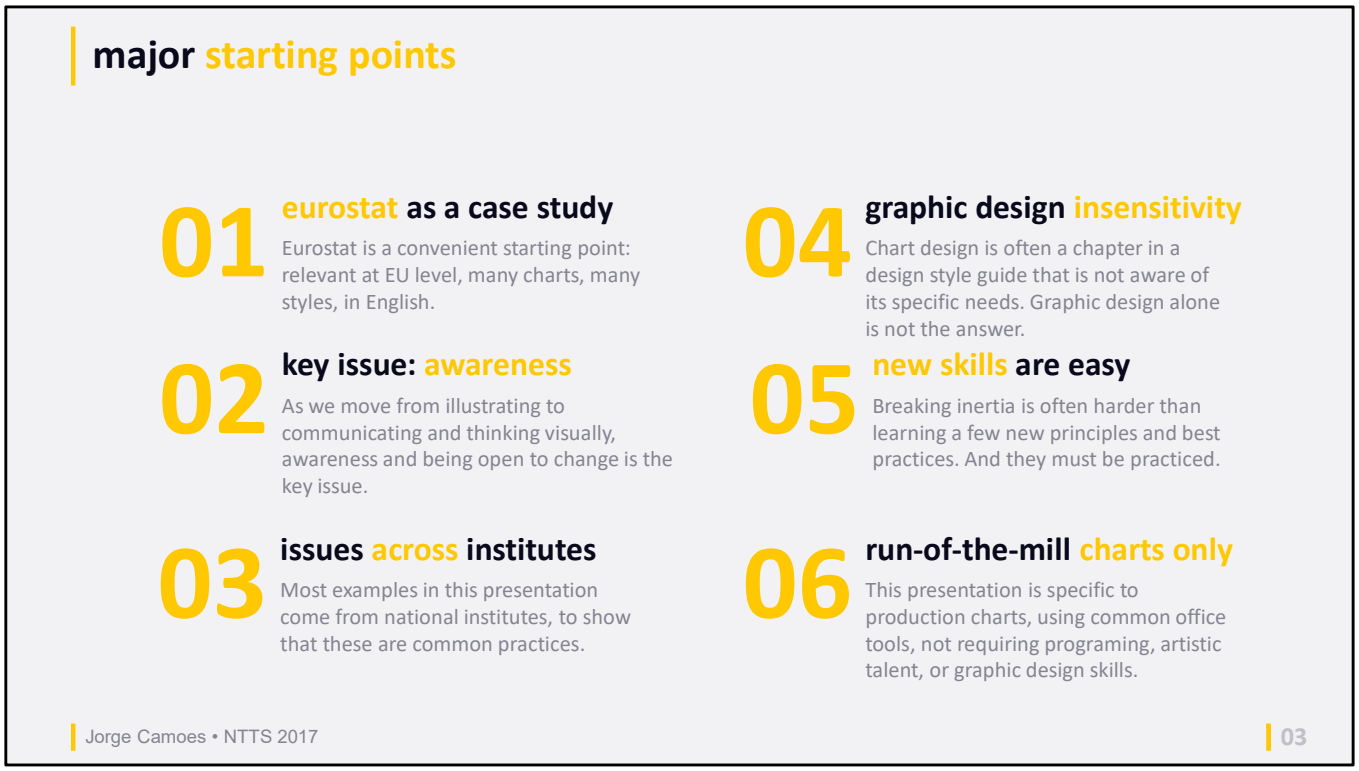
An evaluation of data visualisation practices of statistical institutes

Jorge Camões
NTTS 2017 • Brussels

So, this presentation is titled “An evaluation of data visualization practices of statistical institutes” and I’ll try to show more and tell less.

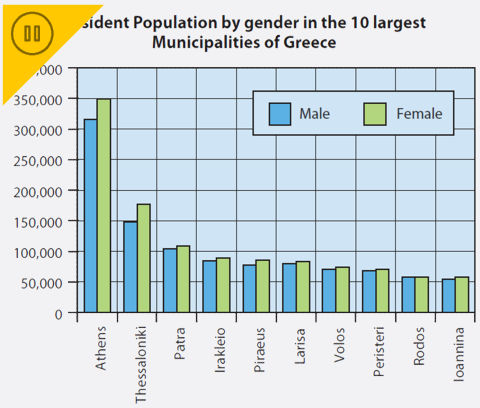


- There was a time when we could focus on individual data points and afford to use colorful but irrelevant visuals (box #1).
- Today, we have many data points and we need to focus on their relationships (box #3).
- Most people are aware of some change, but often they buy into the idea that the answer is adding new makeup to old charts. (box #2).
- Graphic design skills should not be required for run-of-the-mill charts,
- but we can all learn how some functional design choices impact our visual communication (box #4) and apply them effectively.
- So, this will not be about good or bad, right or wrong practices. It will be about switching perspectives, from illustration to communication.
- We just need to pause our current practices, evaluate them, add a few basic rules and continue.
- Finally, the first rule of data visualization is that you do talk about data visualization, you break the rules and you try to come up with the ones that fit the concrete situation better.

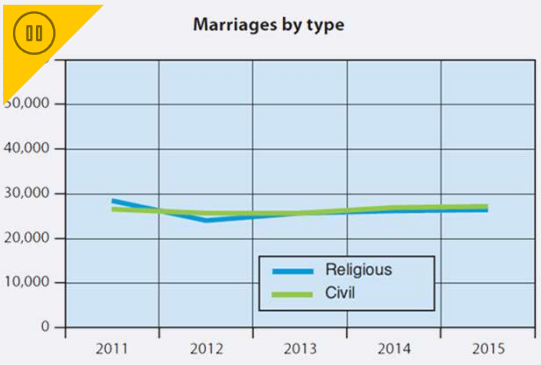


- Since this is a change in a widespread perspective, I assumed it would be visible across institutes.
- I started with Eurostat because of convenience alone.
- But most charts you'll see in this presentation come from other sources, especially national institutes.
- I will not discuss tools, but everything here can be done with Microsoft Excel.
- The discussion applies to run-of-the-mill charts (as oppose to design charts or infographics).
- Programming, artistic talent or graphic design skills are not required.

think about the data



is gender relevant here?

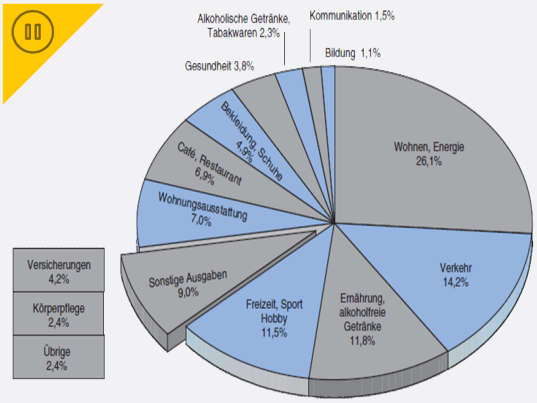


are absolute values relevant here?

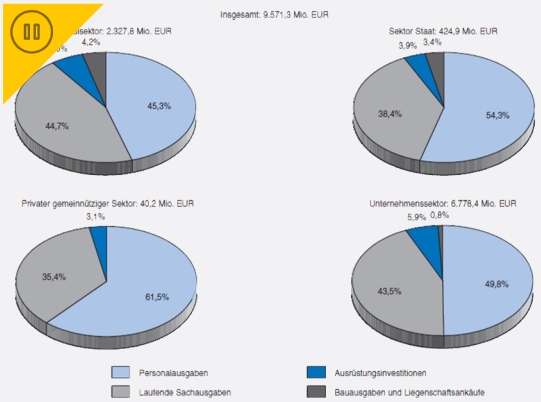
Jorge Camoes • NTTS 2017 04

- The data must always be our starting point
- We need to decide first how relevant or interesting it looks in a chart
- for example, should we present absolute values or focus on ratios or rate of change?

things that belong to the past: everything 3D



3D pie charts

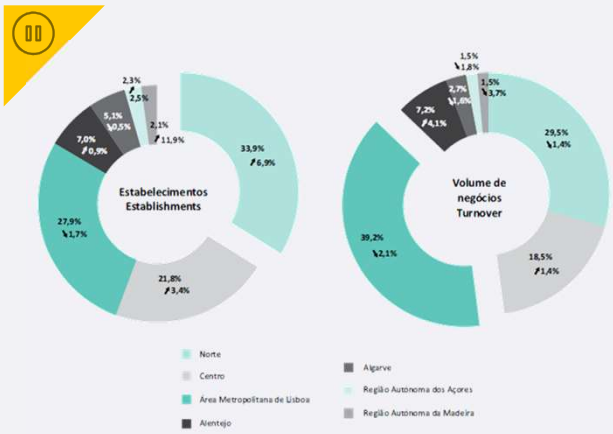


Comparing 3D pie charts

Jorge Camoes • NTTS 2017 05

- Some formatting options are no longer acceptable
- the 3D effects come at the top of the list.

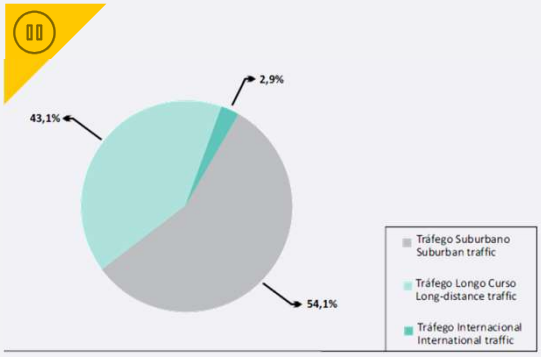
things that belong to the past:
exploded slices, comparing circular charts



Jorge Camoes • NTTS 2017 06

- Exploded slices should also be avoided

things that belong to the past:
legends in circular charts

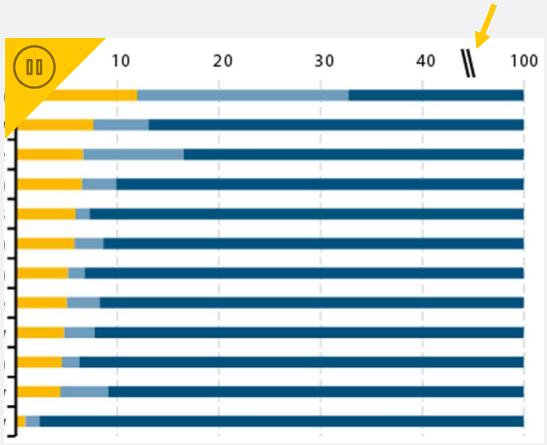


Jorge Camoes • NTTS 2017

07

- Think of a legend as a necessary evil.
- Even more so with circular charts.
- Avoid it if possible
- Because direct labeling is usually better.

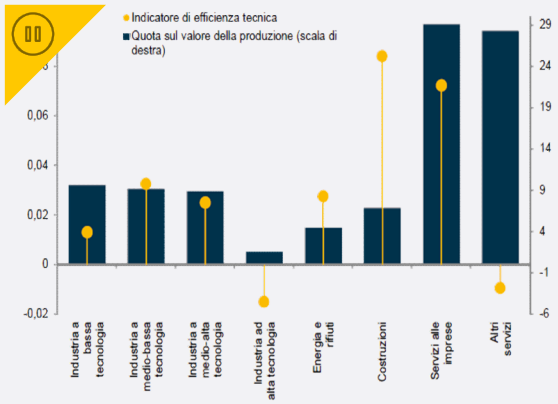
things that can't happen:
breaking scales in part-of-whole charts

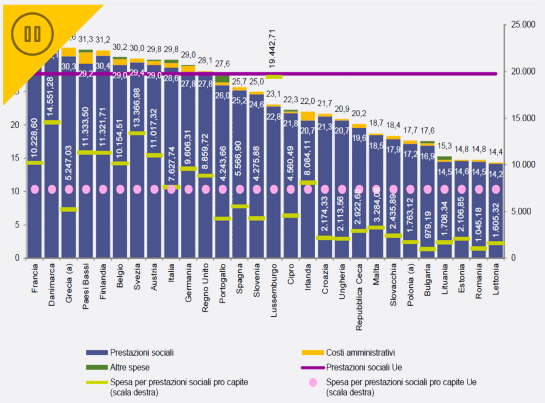


Jorge Camoes • NTTS 2017 08

- Part-of-whole charts mean that the whole must be represented,
- and breaking the scale is not an option.

things that should be avoided:
dual-axis charts



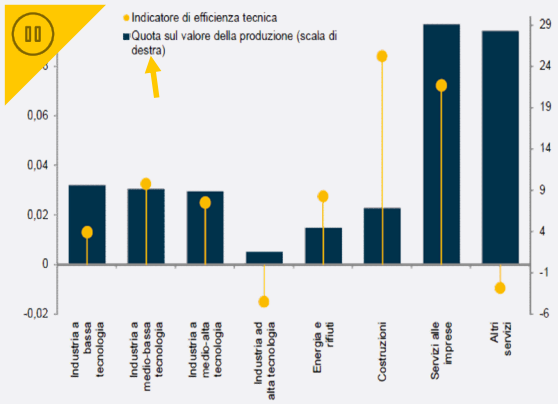


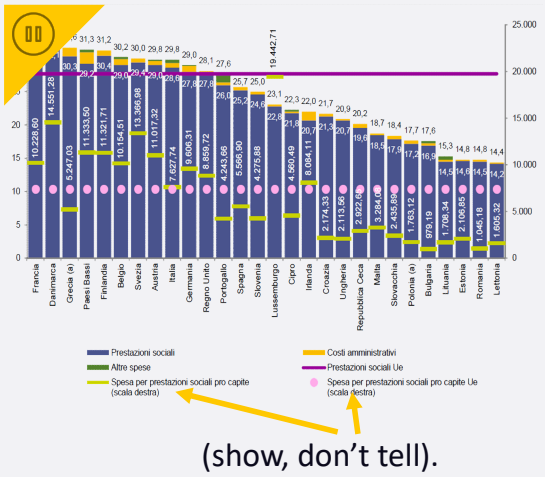
we are tempted to adjust scales to "discover" relationships.

Jorge Camoes • NTTS 2017 09

- Dual-axis are dangerous,
- Because we are tempted to adjust the scales to find relationships that may or may not exist.

things that should be avoided:
in dual-axis charts, not color-coding the axes



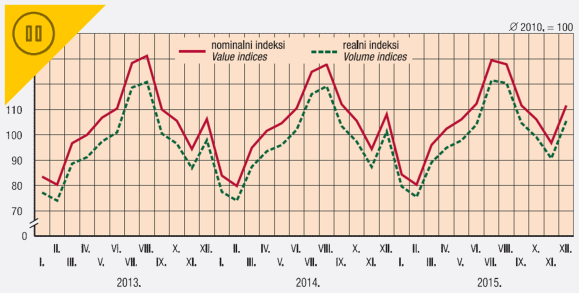


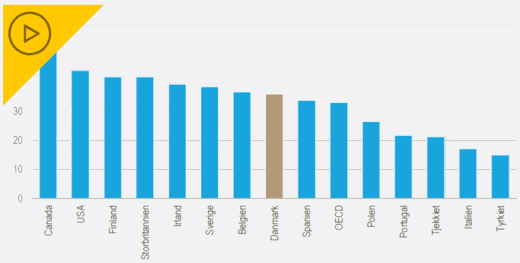
(show, don't tell).

Jorge Camoes • NTTS 2017 010

- If you still need two axes then make sure the readers see which series belong to which axis.
- Don't simply write it in the legend.

things that must be there in a supporting role only:
gridlines too loud, and/or too many





(You probably don't need more than 5 gridlines.)

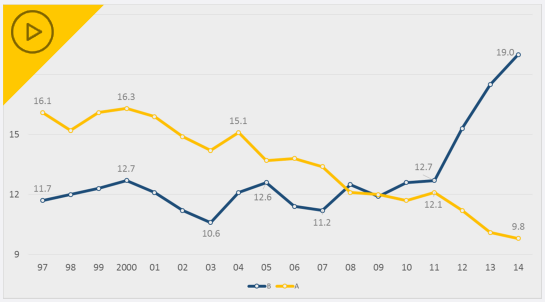
Jorge Camoes • NTTS 2017

011

- The gridlines in the first chart are very loud.
- Gridlines are used in a supporting role,
- so they have to be muted to let the data shine.
- 3 to 5 gridlines are enough in most cases.

things that must be there in a supporting role only:
data labels too loud, and/or too many

Don't turn a chart into a table.
Use a table if you need it.

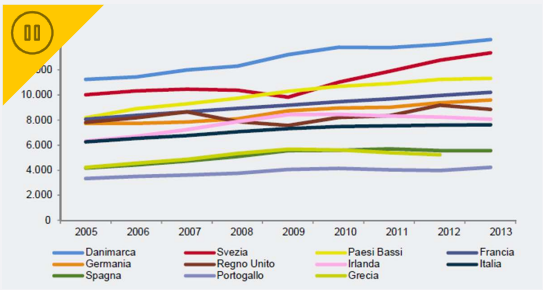


Select the more meaningful data points and use a shade of gray.

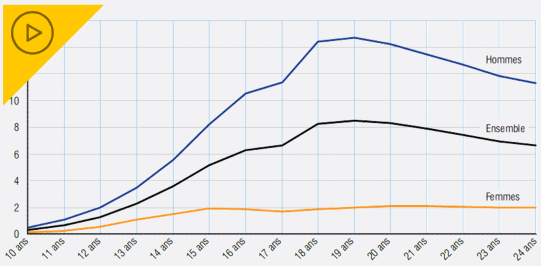
Jorge Camoes • NTTS 2017 012

- A chart is not a table.
- If you need to show the data use a small table close to the chart.
- Otherwise, label only a few relevant data points.

things that must be there in a supporting role only: legends



Too many series.

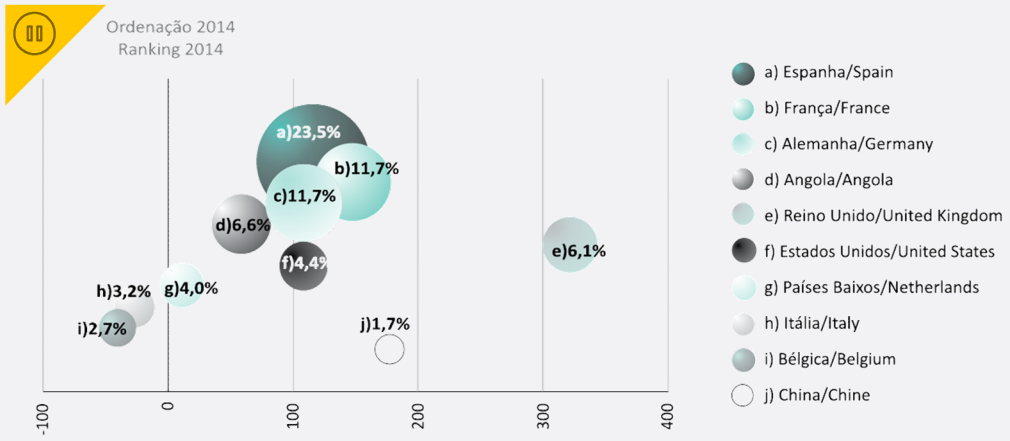


Direct labeling is better than a legend.

Jorge Camoes • NTTS 2017 013

- The first chart is too busy,
- most users will not bother to read because, you know, readers are lazy. Or maybe not.
- If you can't use direct labeling in a line chart probably you are using too many series.

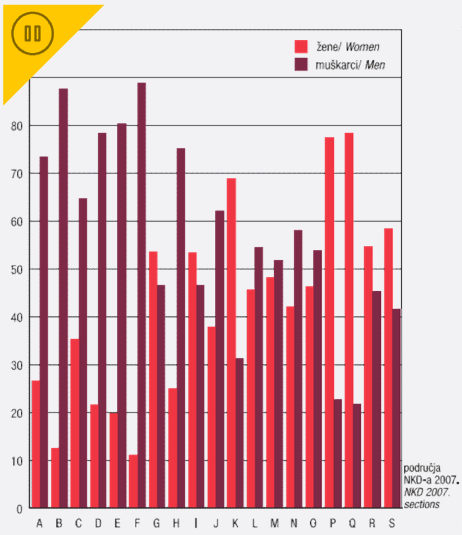
things that must be there in a supporting role only:
legends with a dual authentication system?



Jorge Camoes • NTTS 2017 014

- I called this technique a “dual authentication system”
- because it makes even harder to identify each data point.
- This dual authentication system is something that you clearly need to avoid.

things that must be there in a supporting role only: legends with a dual authentication system?



područja NKD-a 2007.
NKD 2007. sections

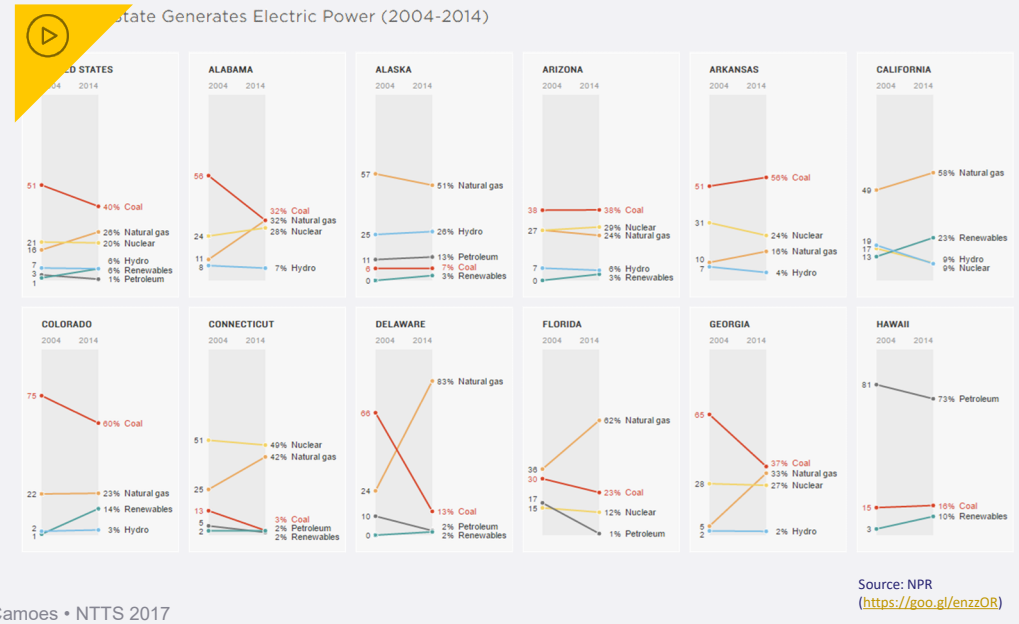
A Poljoprivreda, šumarstvo i ribarstvo
Agriculture, forestry and fishing
B Rudarstvo i vađenje/ *Mining and quarrying*
C Prerađivačka industrija/ *Manufacturing*
D Opskrba električnom energijom, plinom, parom i klimatizacija
Electricity, gas, steam and air conditioning supply
E Opskrba vodom; uklanjanje otpadnih voda, gospodarenje otpadom te djelatnosti sanacije okoliša
Water supply; sewerage, waste management and remediation activities
F Građevinarstvo/ *Construction*
G Trgovina na veliko i na malo; popravak motornih vozila i motocikla
Wholesale and retail trade; repair of motor vehicles and motorcycles
H Prijevoz i skladištenje/ *Transportation and storage*
I Djelatnosti pružanja smještaja te pripreme i usluživanja hrane
Accommodation and food service activities
J Informacije i komunikacije/ *Information and communication*
K Financijske djelatnosti i djelatnosti osiguranja
Financial and insurance activities
L Poslovanje nekretninama/ *Real estate activities*
M Stručne, znanstvene i tehničke djelatnosti
Professional, scientific and technical activities
N Administrativne i pomoćne uslužne djelatnosti
Administrative and support service activities
O Javna uprava i obrana; obvezno socijalno osiguranje
Public administration and defence; compulsory social security
P Obrazovanje/ *Education*
Q Djelatnosti zdravstvene zaštite i socijalne skrbi
Human health and social work activities
R Umjetnost, zabava i rekreacija
Arts, entertainment and recreation
S Ostale uslužne djelatnosti/ *Other service activities*

Also: what is the message?

Jorge Camoes • NTTS 2017 015

- A different version of the “dual authentication system”.
- It’s very hard to learn something from this chart.

busy charts can often be turned into small multiples



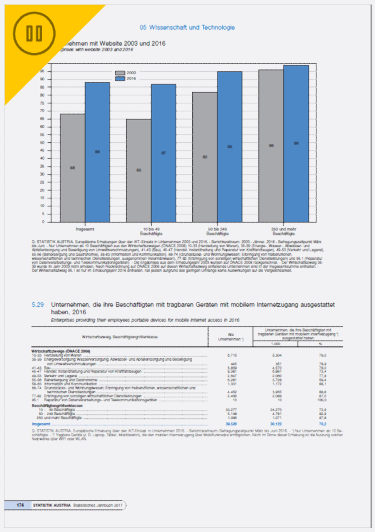
Source: NPR
(https://goo.gl/enzzOR)

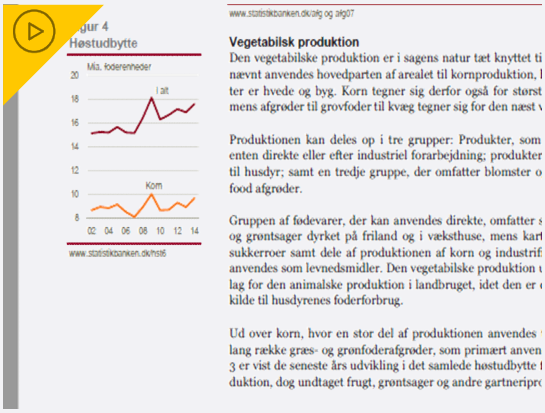
Jorge Camoes • NTTS 2017

016

- The small multiples technique is often a good alternative when a chart is too busy.

size (and aspect ratio) matters:
inefficient page space management





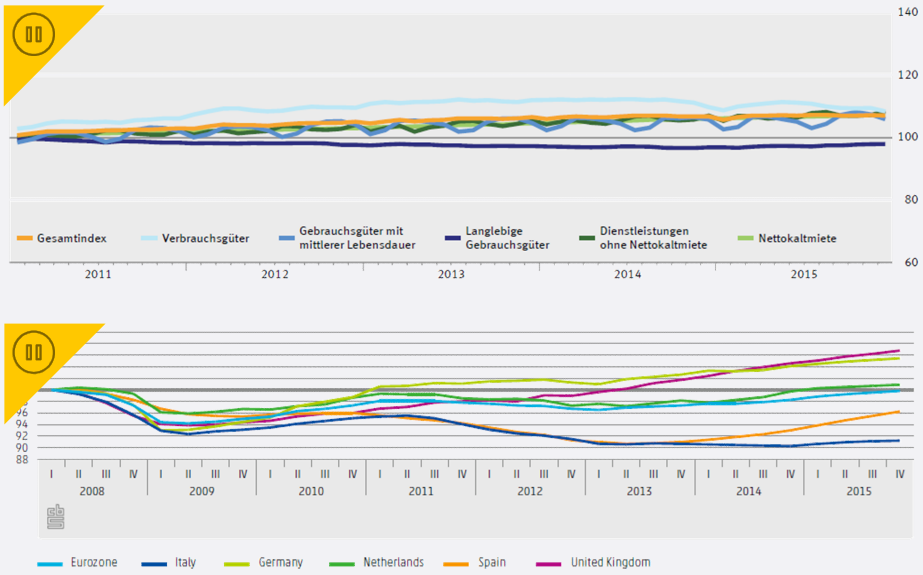
Small charts, next to the text.

Jorge Camoes • NTTS 2017

017

- Most charts are much larger than needed.
- The chart on the left uses almost half a page to display only 8 data points.
- Many charts can be replaced by smaller versions,
- Place them near the text discussing the data.

size (and aspect ratio) matters:
the swimming pool effect

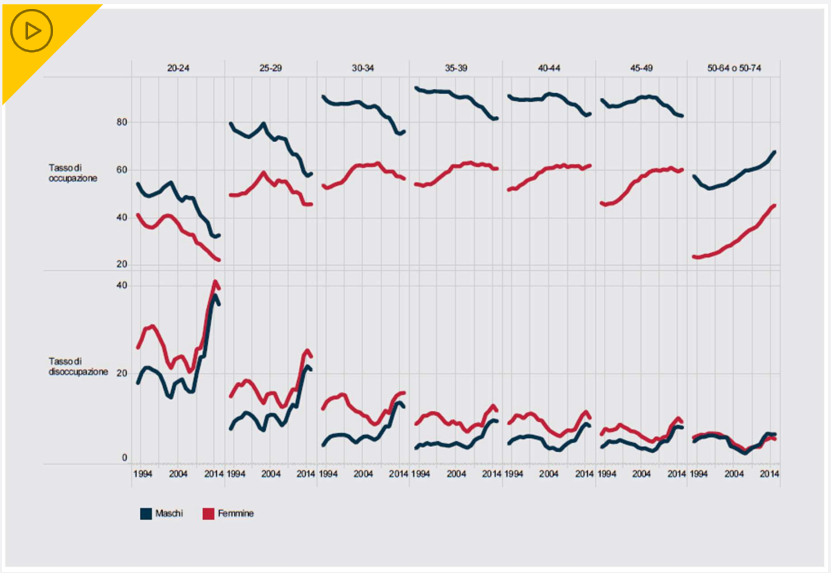


Jorge Camoes • NTTS 2017

018

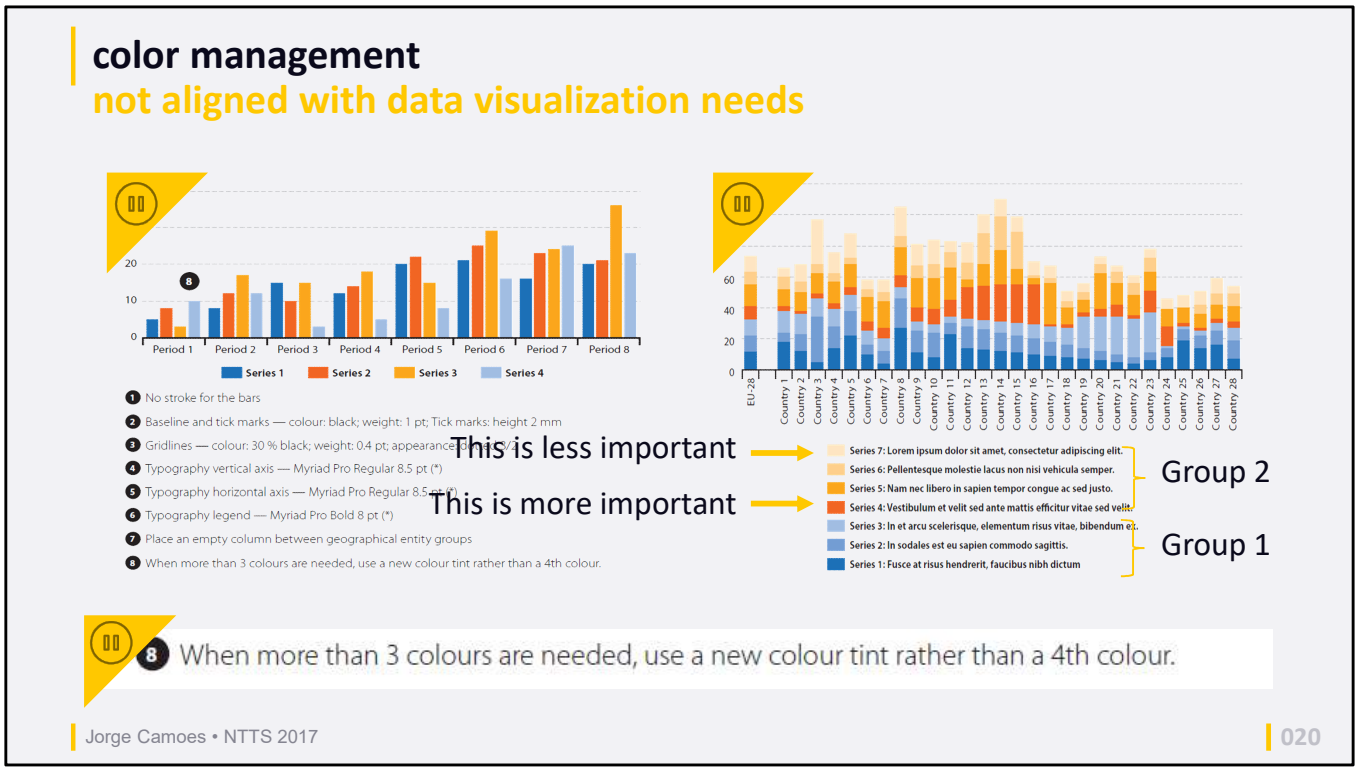
- There are many examples of what I call the swimming pool effect
- By that I mean charts too wide and with an aspect ratio of 4:1 or even more.
- This reduces vertical resolution.
- There are no specific guidelines,
- but taking into account the human field of vision and common aspect ratios in use,
- probably a chart shouldn't be stretched above a 2:1 ratio.

a panel chart is a solution to the swimming pool effect



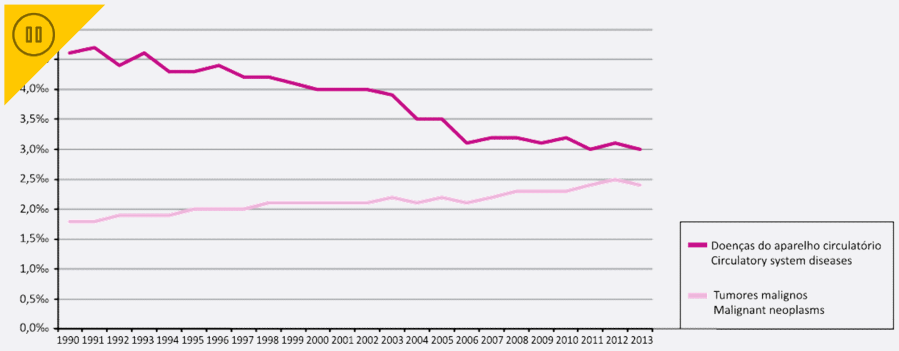
Jorge Camoes • NTTS 2017 019

- Again, small multiples could be an interesting alternative.



- Color is a complex issue.
- Edward Tufte says that our main concern when using it should be avoid catastrophe.
- These rules come from the Eurostat style guide.
- From a design perspective, they may be OK,
- but without more context they are not OK for data visualization tasks.
- If you use two different colors and several tints, readers will probably read each color as a group,
- lighter tints will be read as less important than darker or more saturated ones.
- We should take advantage of this, but it is a data-driven design decision, not simply because we have too many series.

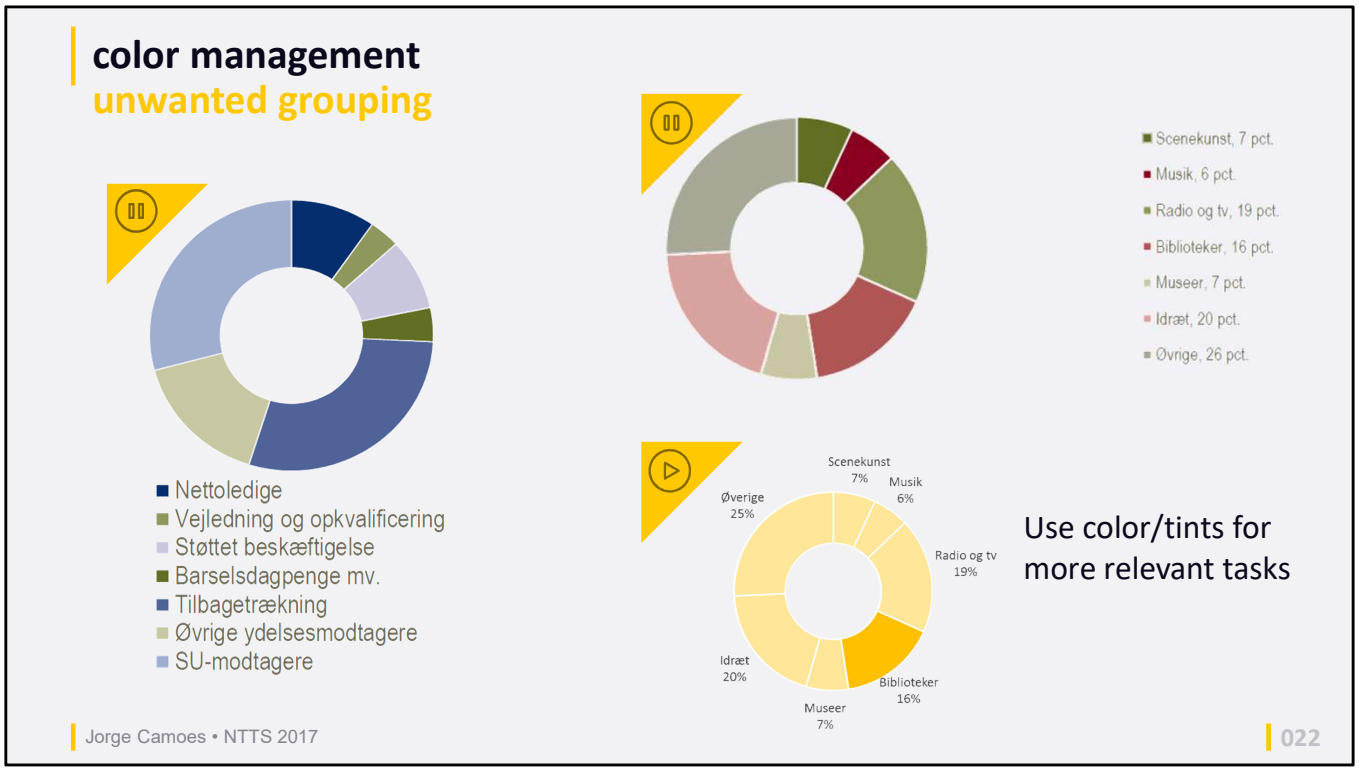
color management
change chromatic intensity to denote relevance (on purpose)



Trend of mortality rate caused by diseases of the circulatory system and malignant neoplasms, 1990 – 2013

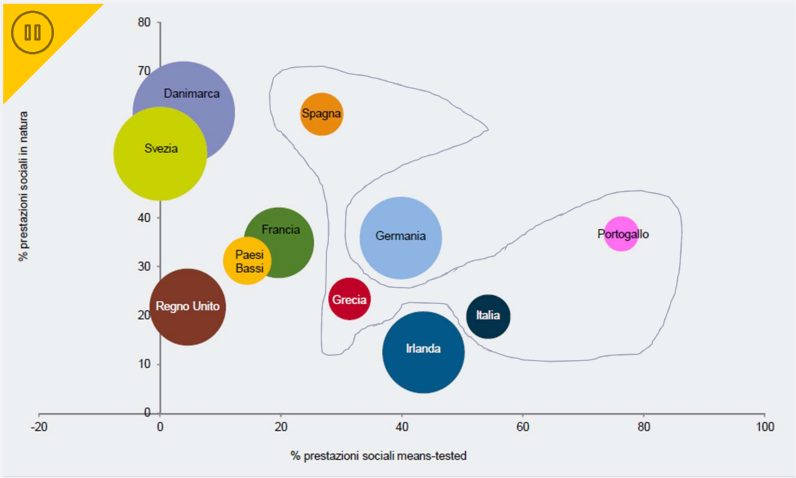
Jorge Camoes • NTTS 2017 021

- Here, the darker series looks more relevant.
- If this is done on purpose, it should be said it in the title.



- The top two charts show examples of unwanted grouping,
- a consequence of the guidelines above.
- A functional use of color would be more effective than defining how many colors and tints people should use.
- Most of the time we use too much color,
- and it is virtually never needed in pies or donuts,
- When we don't need it to identify slices we can use it for emphasis
- as you can see in the chart below.

color management
use it when you need it



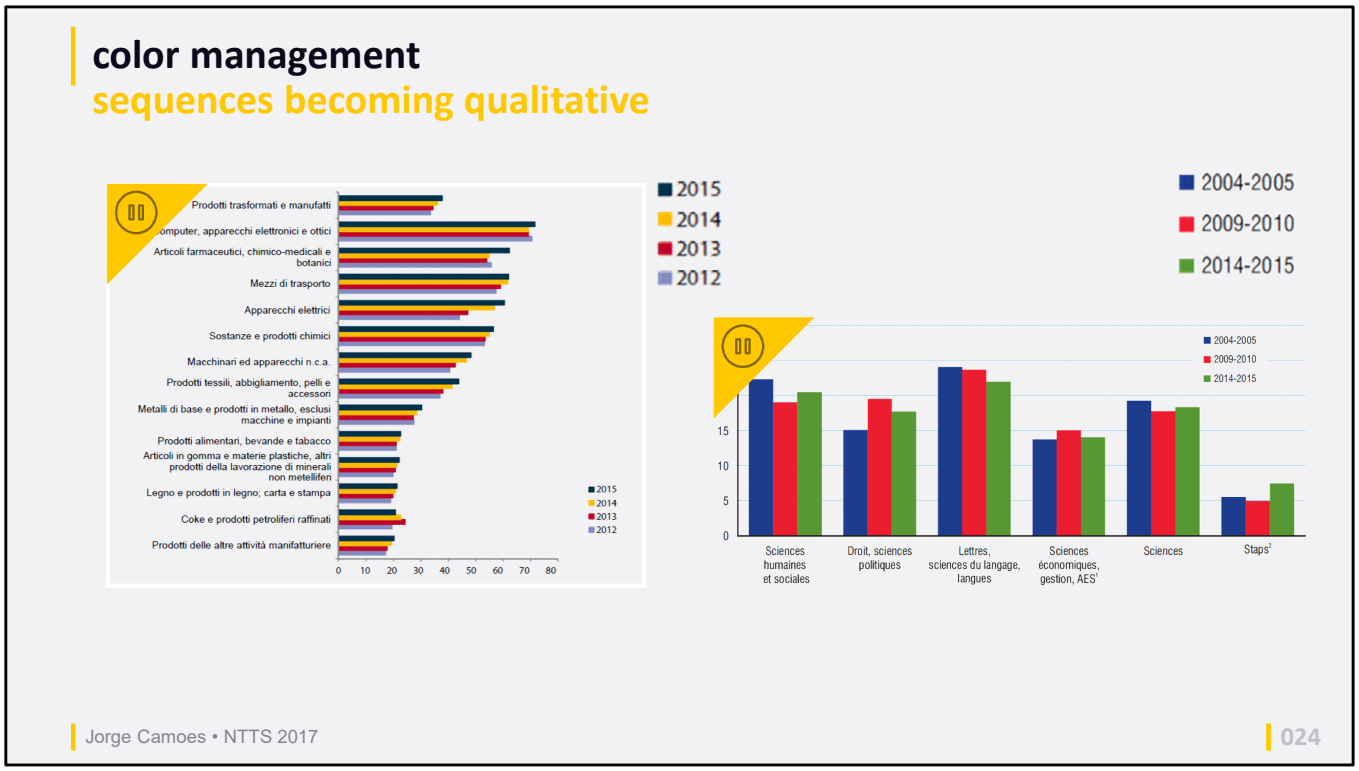
Color-encode these 4 countries using the same color.

Do the same with the other countries.

Explicit the grouping variable.

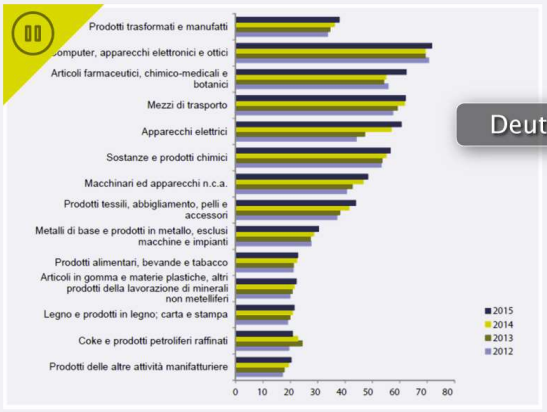
Jorge Camoes • NTTS 2017 023

- This is another example where color is not needed.
- We could use it to group countries.

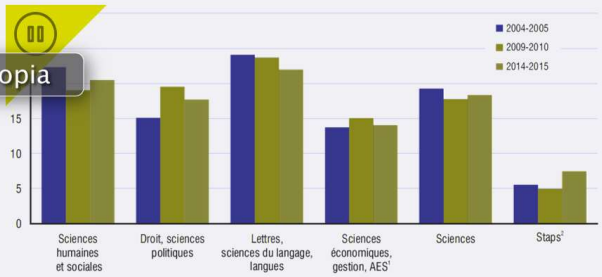


- If there is some kind of sequence, like in a time series, using different colors destroys that sequence.

color management
sequences becoming qualitative



Deuteranopia



Tips:

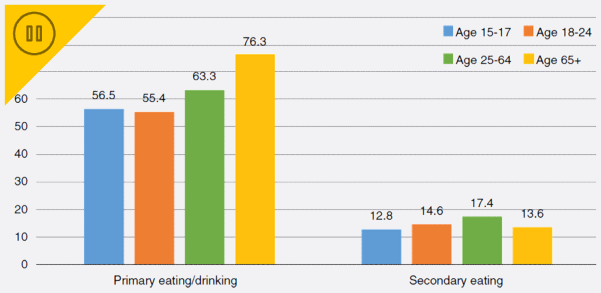
- Test your charts with shades of gray;
- Use a color blindness simulator like Color Oracle to test your colors.

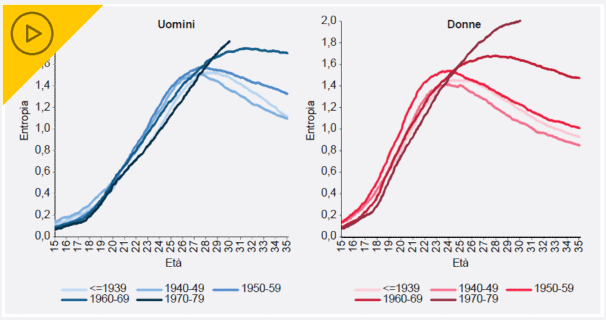
Jorge Camoes • NTTS 2017 025

- This simulates color blindness, and it's something you should be aware of.
- As you can see, you can't tell red and green apart

color management

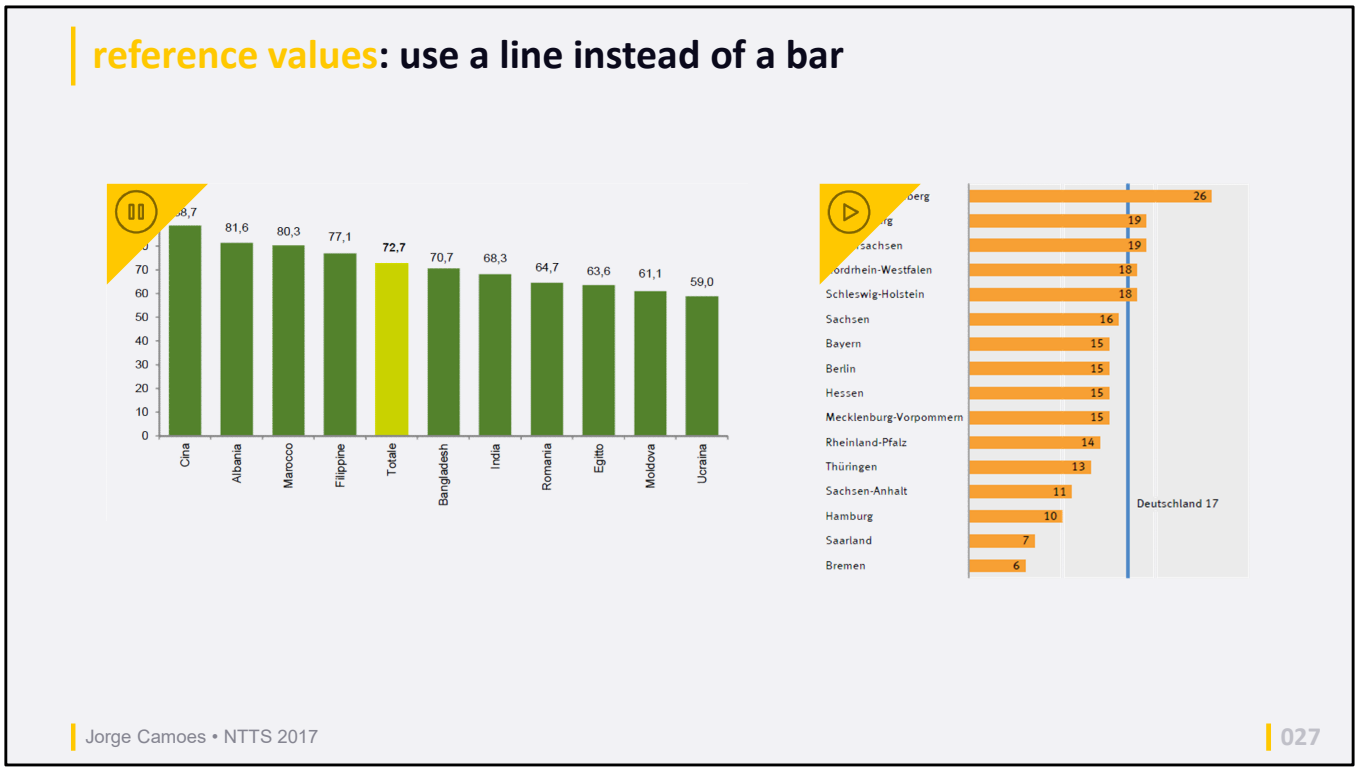
use tints instead of colors to make a sequence more obvious





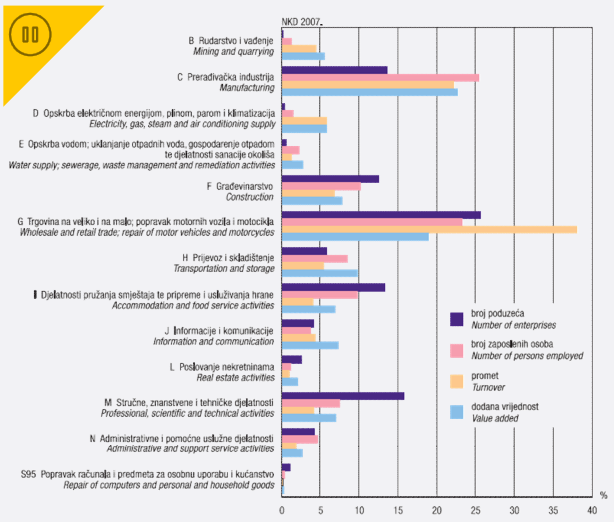
Jorge Camoes • NTTS 2017 026

- The charts on the right shows how it should be done with a simple mnemonic: darker means more recent. Very simple.
- In general, make sure the colors you choose help people reading the chart, minimizing the need for a legend.
- Put color discriminability and stimuli intensity before aesthetics.

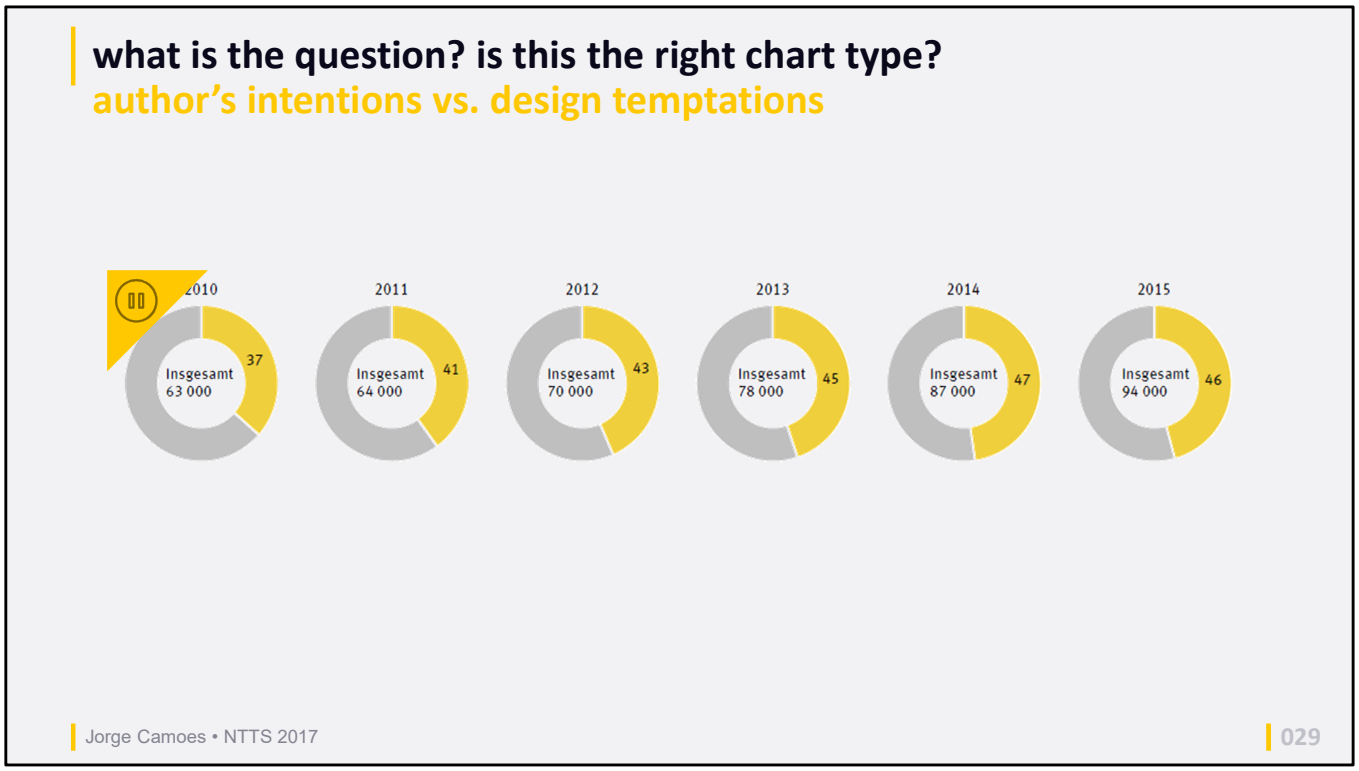


- This is also a common issue.
- When you have a reference value in a bar chart use a line to encode it, like the right chart does.
- Comparing gaps to the reference line is easier than using a bar.

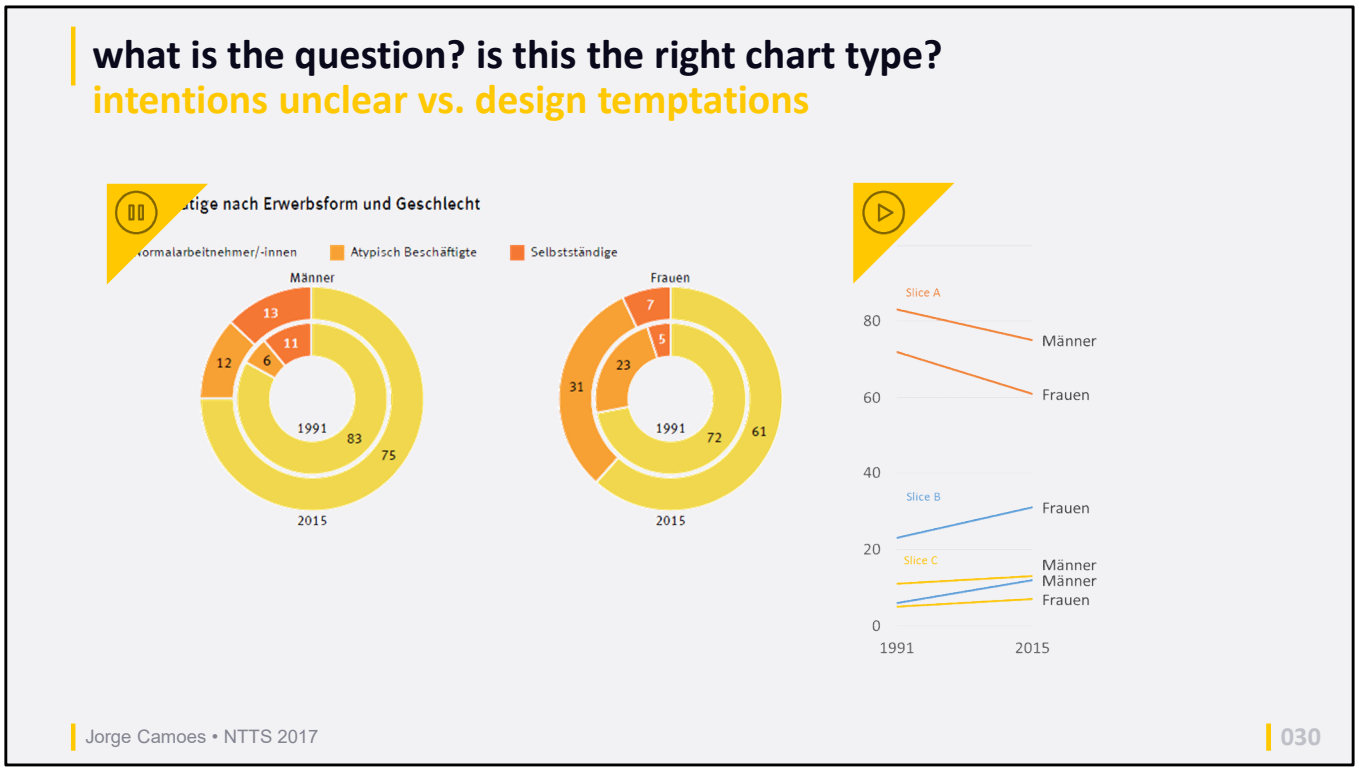
what is the question? is this the right chart type?
author's intentions vs. chart's ability to deliver the message



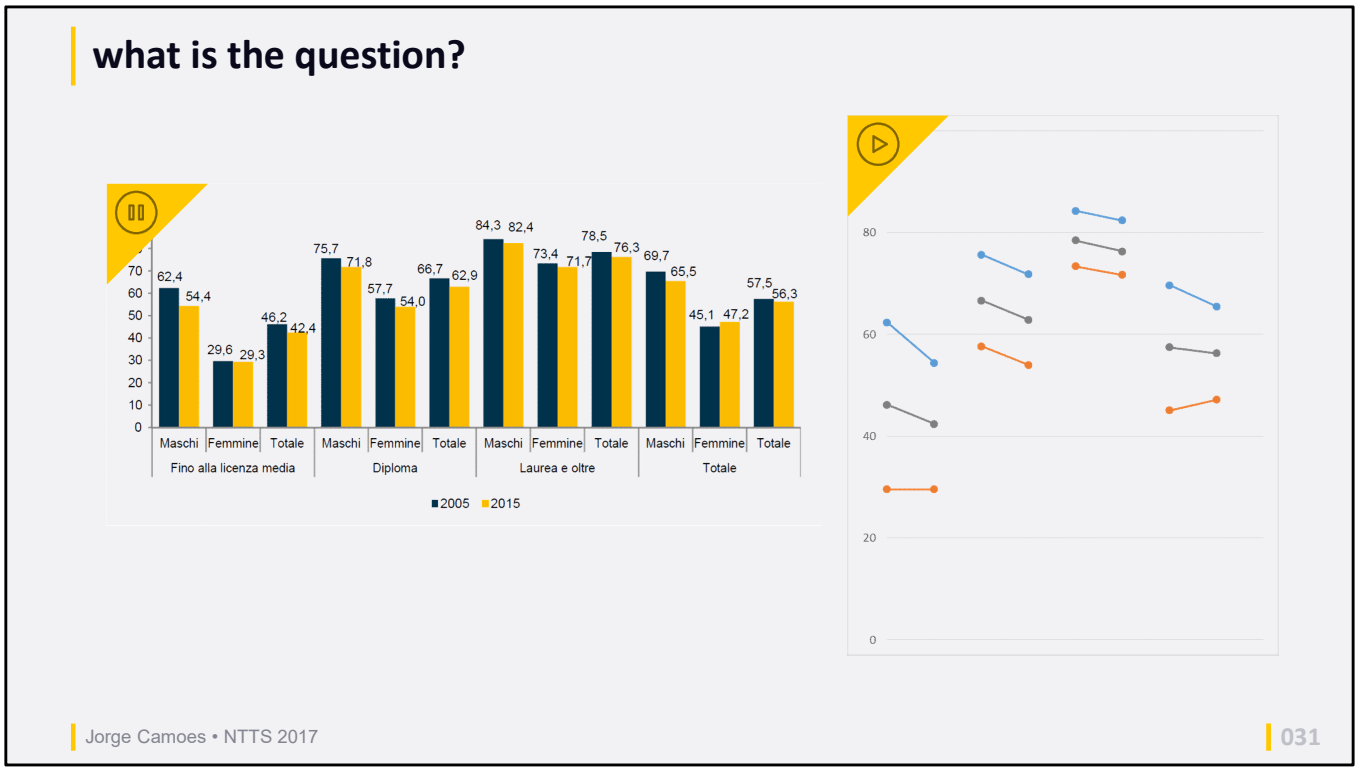
- Think of a chart as a visual answer.
- That means that there should be a clear answer
- but also a clear question.
- Often we can guess the author's intention,
- but in cases like this the answer is not clear.



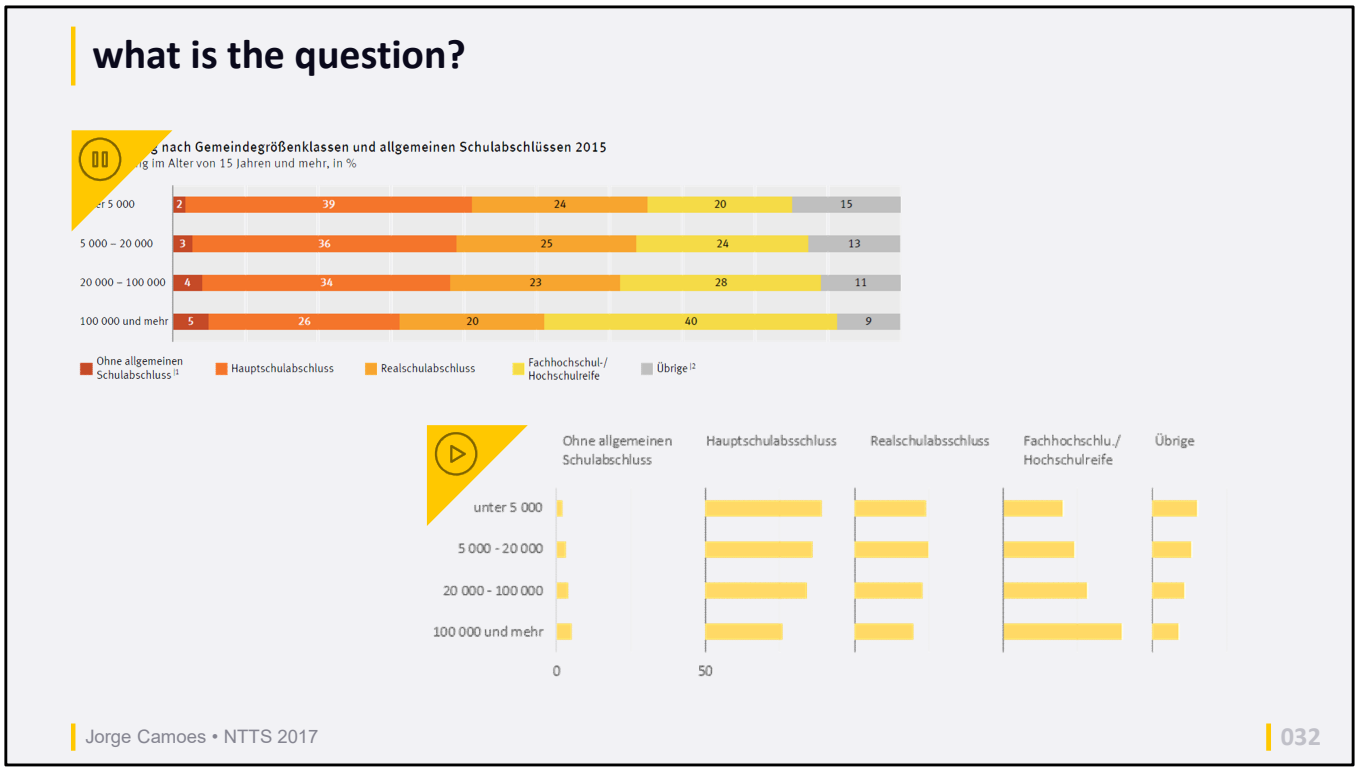
- Here, the intention is clear, but the author couldn't resist having fun with donuts.
- It's a case of, as Stephen Few puts it, "Our Irresistible Fascination with All Things Circular"



- Comparing two or more circular charts should be avoided
- and that includes rings within rings
- The chart on the right simplifies evaluating and comparing the data points.



- Displaying data split by gender is very common.
- Show the difference and the gaps to the total, not the full bars.
- The chart on the right would achieve that and would be a much lighter display



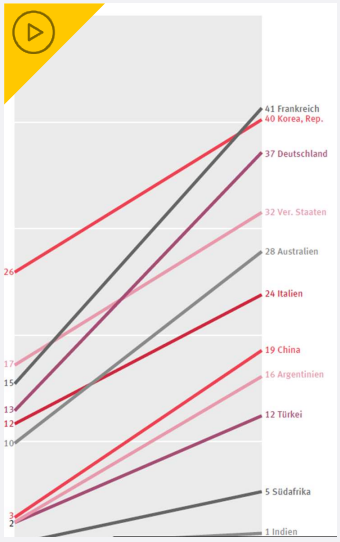
- 100% stacked bars are very ineffective.
- They use too many colors,
- need a legend
- and, above all, makes it very hard to compare the middle series.
- The patterns are much clearer using the chart below.

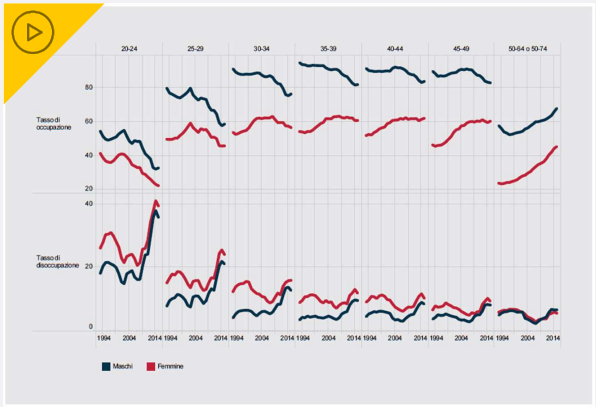
> ***There's a strand of the data viz world that argues that everything could be a bar chart. That's possibly true but also possibly a world without joy.***
>
> ***Amanda Cox, NYT***

https://hbr.org/2013/03/power-of-visualizations-aha-moment

- There are hundreds of chart types,
- and I'm sure you'll find the right one for your data, your question and your audience.
- Bar charts are usually safe,
- but to provide a more precise answer you often have to go beyond them.
- Don't be afraid to experiment, within reason.

should be used more often

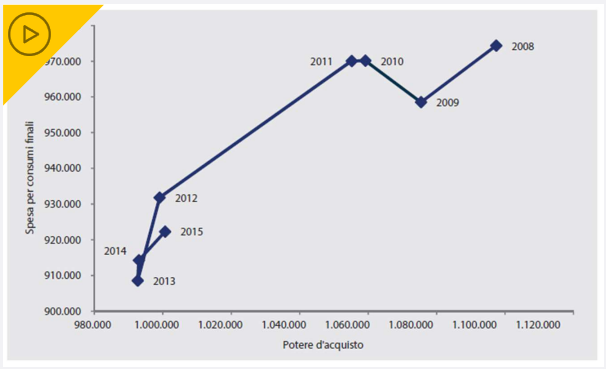




Jorge Camoes • NTTS 2017 034

- Here are a few chart types that could be used more often
- Slope charts and small multiples.

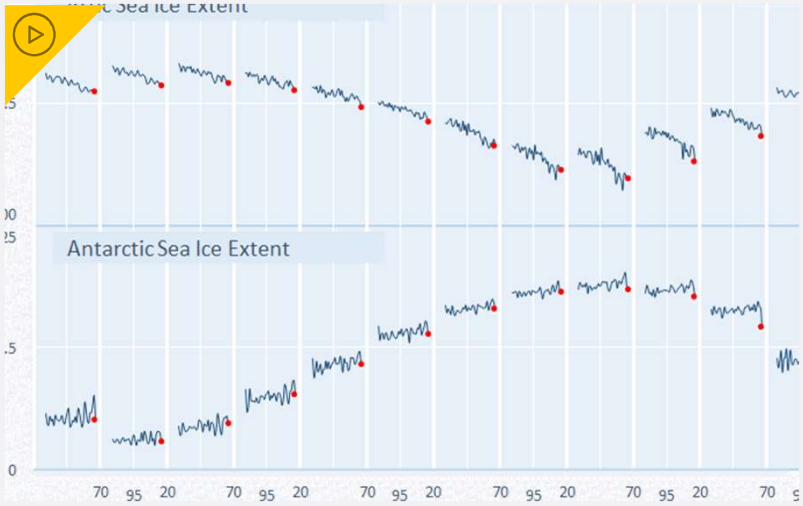
should be used more often



Jorge Camoes • NTTS 2017 035

- Scatterplots are used, but should include more annotations
- connected scatterplots are great to show relationships over time

should be used more often

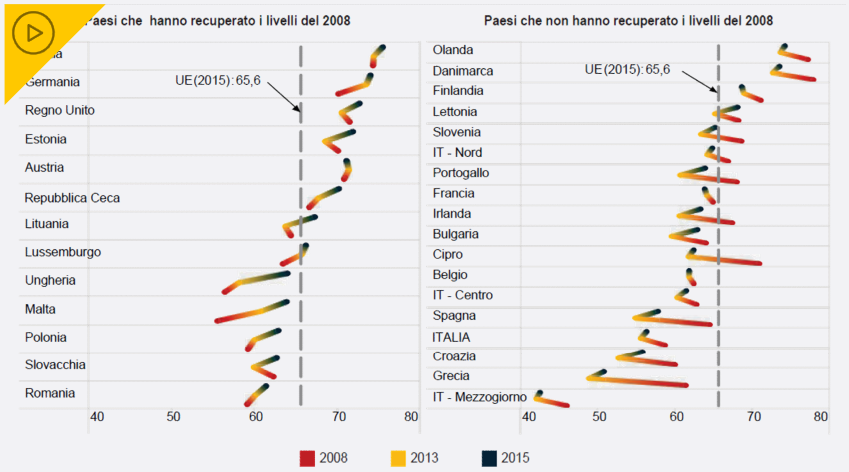


Jorge Camoes • NTTS 2017

036

- Cycle plots for strongly seasonal data

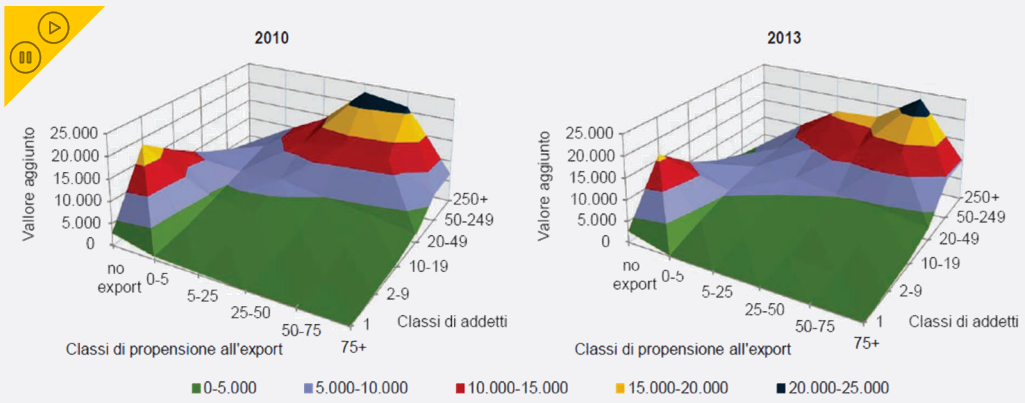
some experiments seem to work



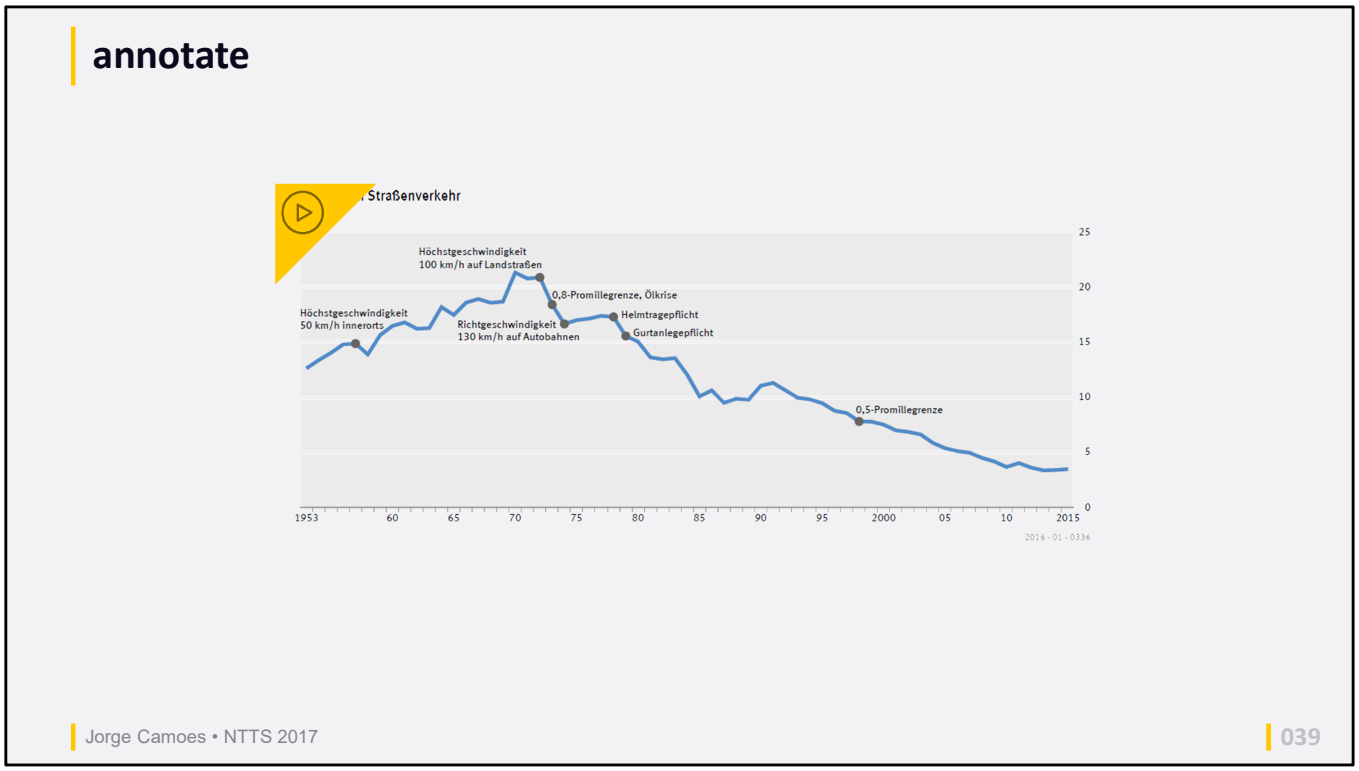
Jorge Camoes • NTTS 2017 037

- Should we call it a boomerang chart?
- A very interesting experiment,
- Could be improved using a different color pallet

others not so much

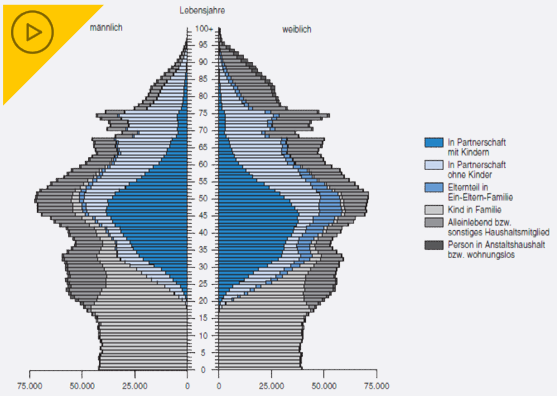


- Others don’t seem to work
- 3D and color pallet issues here



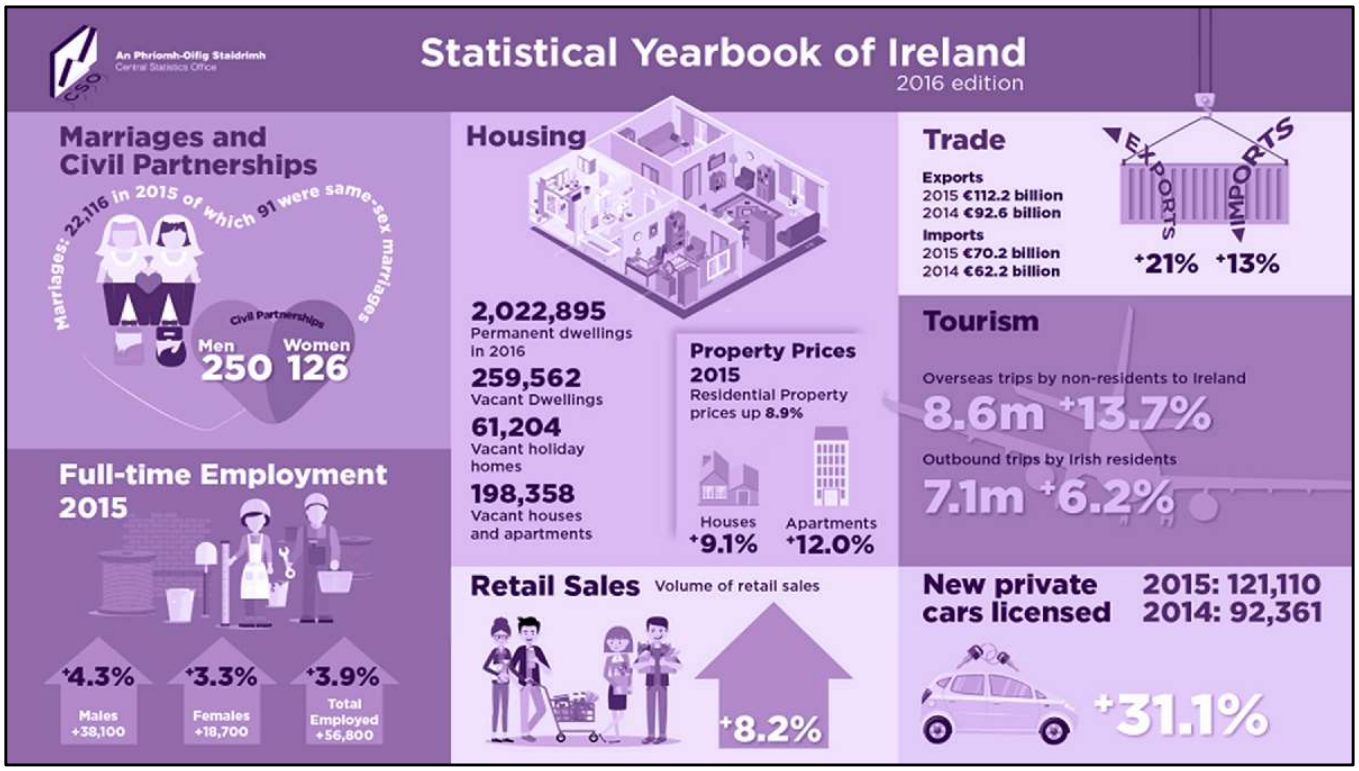
- Annotate relevant data points or outside events that could impact the data.

test sizes, enrich the message

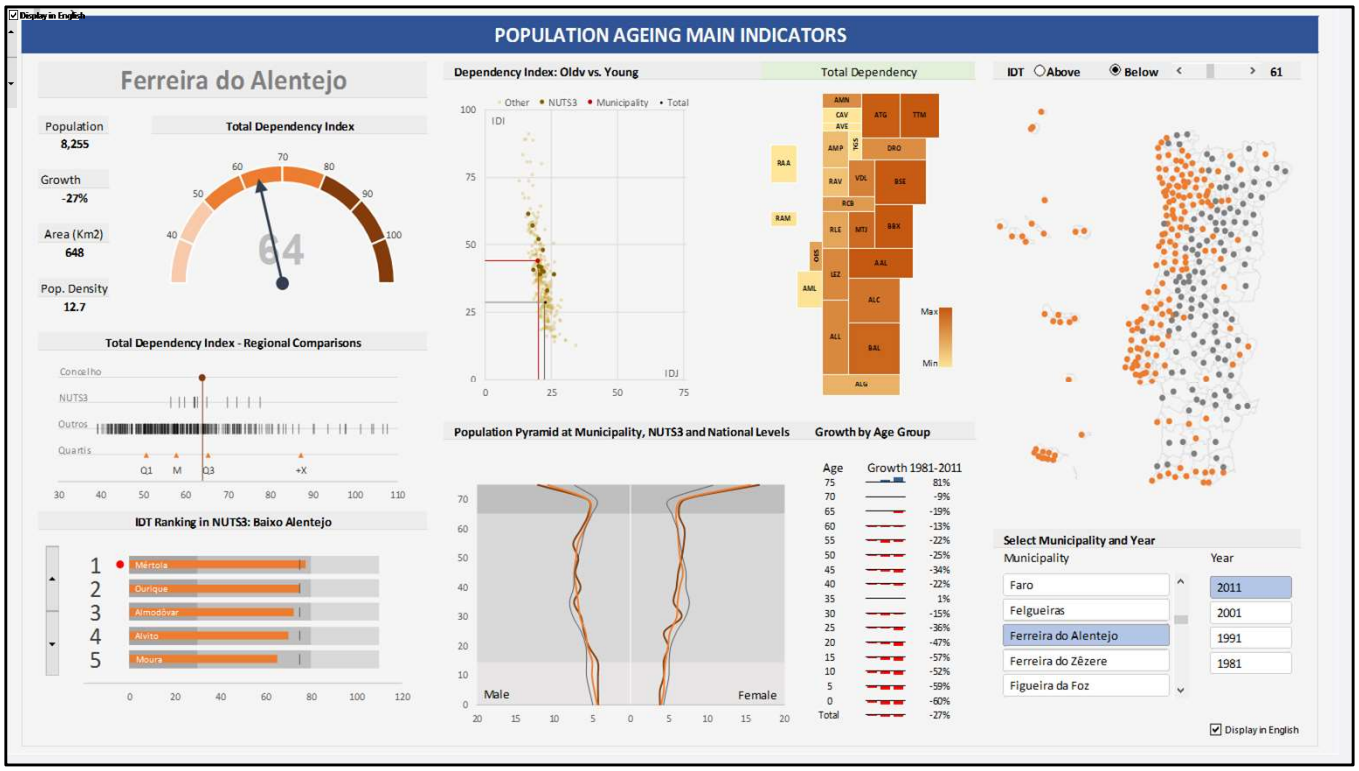


Jorge Camoes • NTTS 2017 040

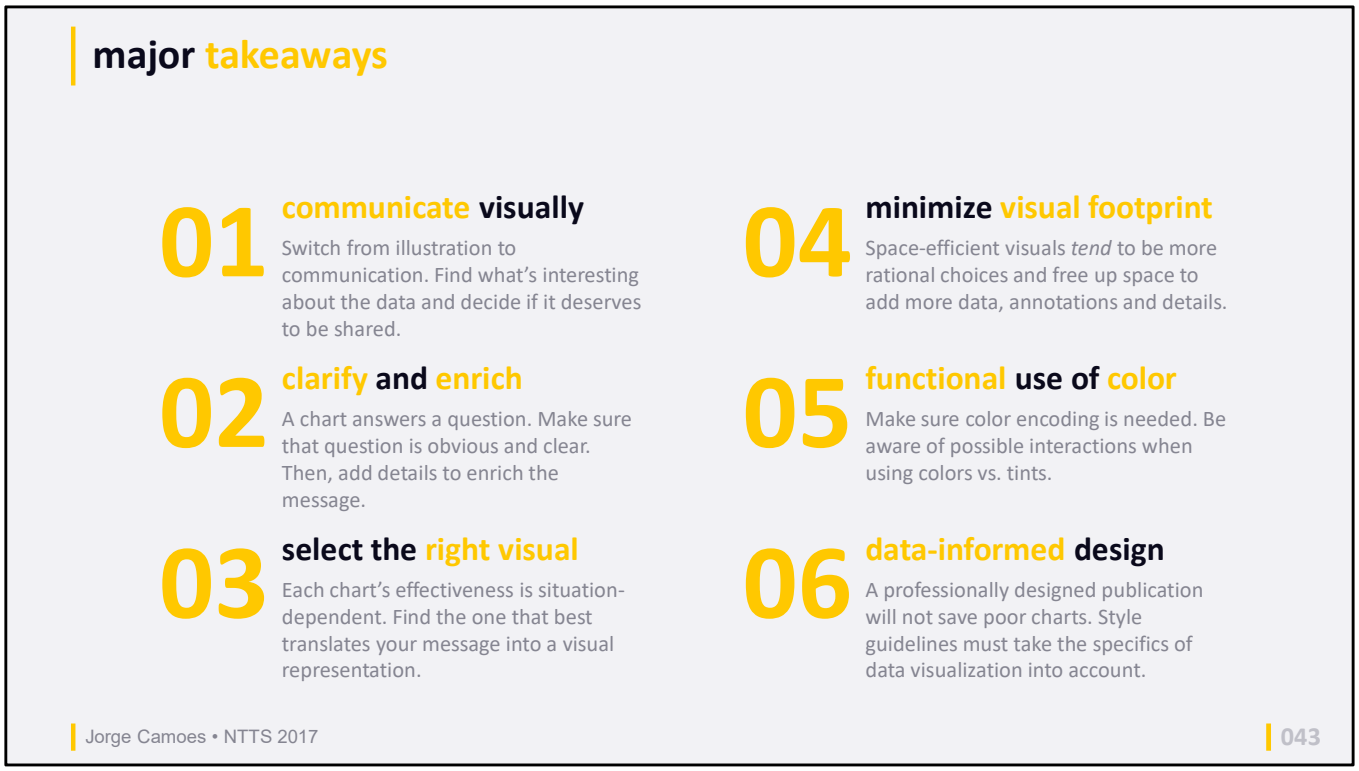
- Test smaller sizes close to the text
- Add detail



- Be aware of what makes your charts more effective,
- and feel comfortable with design choices like color.
- But with more data comes the need for a more complex and consistent message.
- Don’t think of charts, think of communication,
- and for that you have to remove the mental border around each chart.
- Treat a chart as a paragraph in a longer narrative.
- You could try the infographic way, like in this case,
- but infographics require a more design-oriented approach, and we often lack those skills.



- This, on the other hand, is data-oriented,
- and it is nothing more than a few Excel charts
- put together with the purpose of showing related data.
- This is where we need to go after solving single-chart issues.
- I'm sure your results will be much better than this.



- So, here are some major takeaways.
- Use visuals to communicate, not to illustrate
- Clarify and enrich the message
- Select the right visual for your data, your message and your audience
- Select charts that minimize the visual footprint
- Make a functional use of color
- In general, apply a data-informed design
- Thank you.