

Data as a Commons Data Visualization

a course for the phd students of GSSI - Gran Sasso Science Institute

Maurizio Napolitano
<napolitano@fbk.eu>

What will we learn today



Beautiful doesn't mean effective

More a graph needs an explanation,
the less the graph is needed.



How will we do it



Horror Gallery

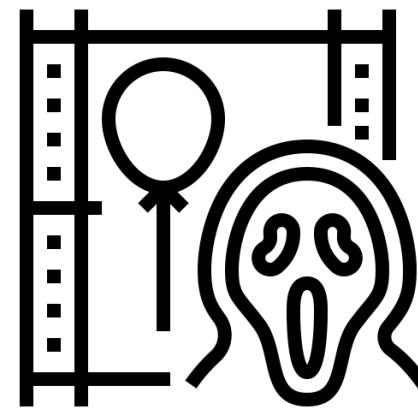
Main errors

Principles of graphic excellence

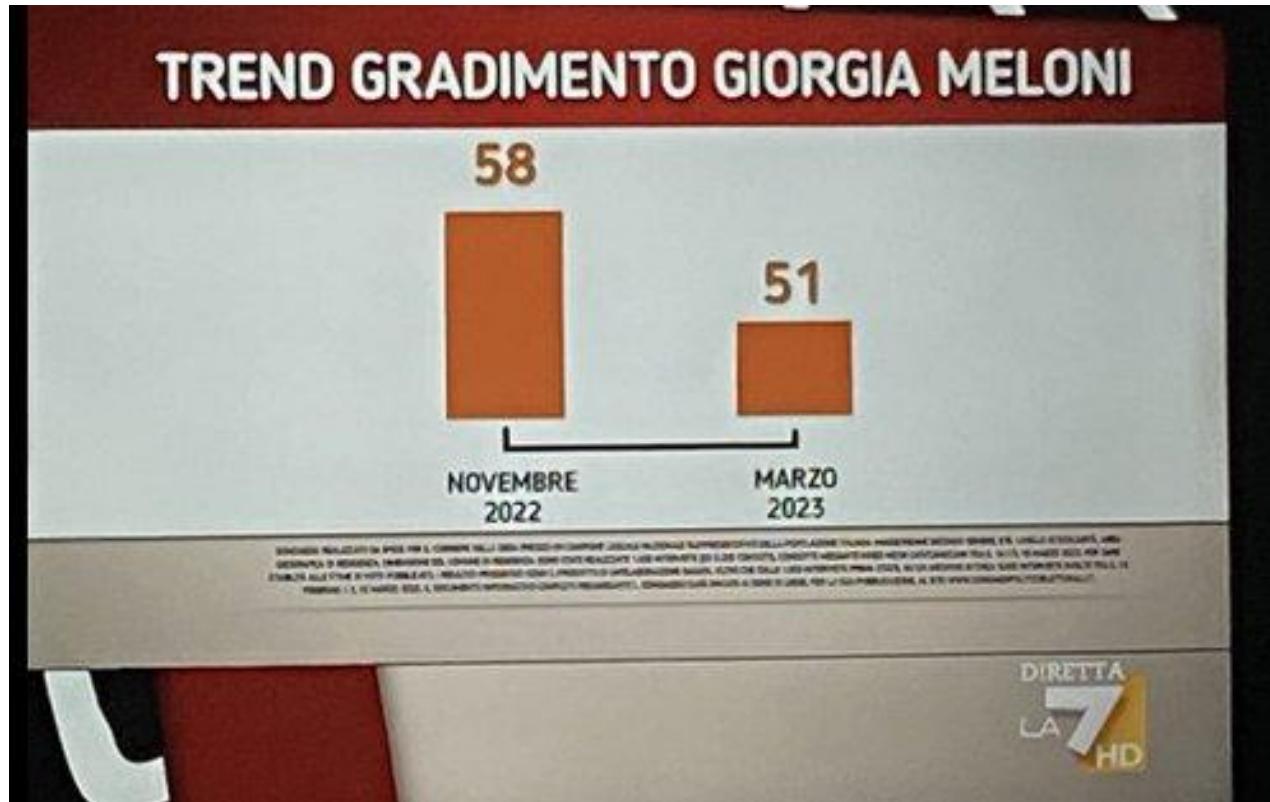
How to choose a chart

Tools

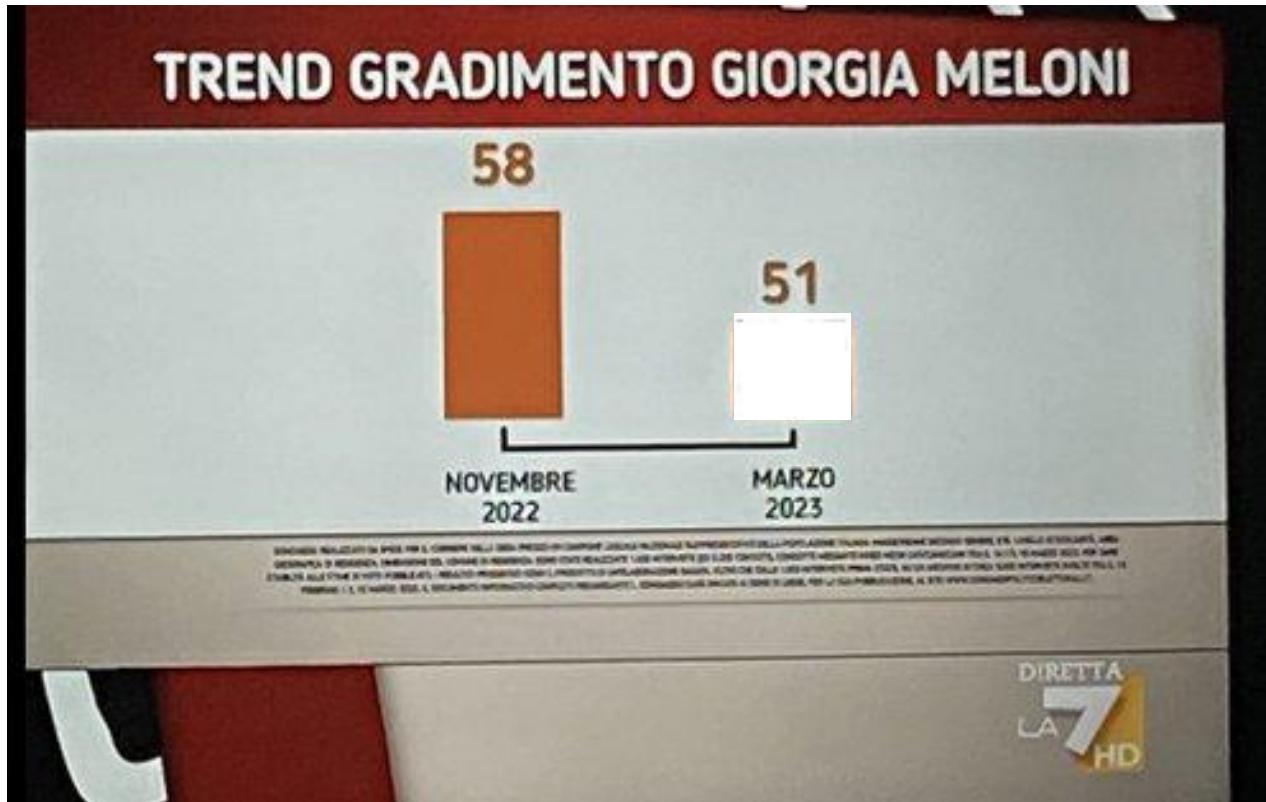
Horror Gallery



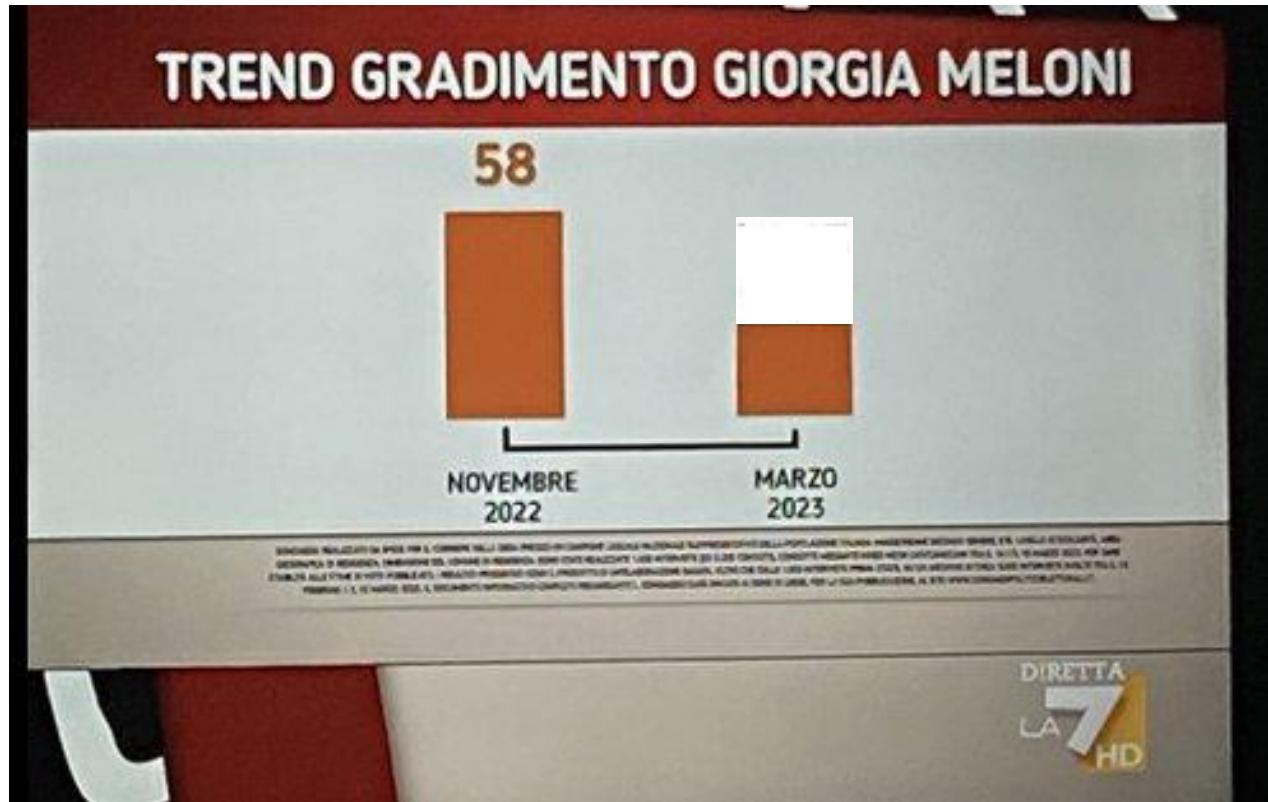
Popularity trend on Giorgia Meloni



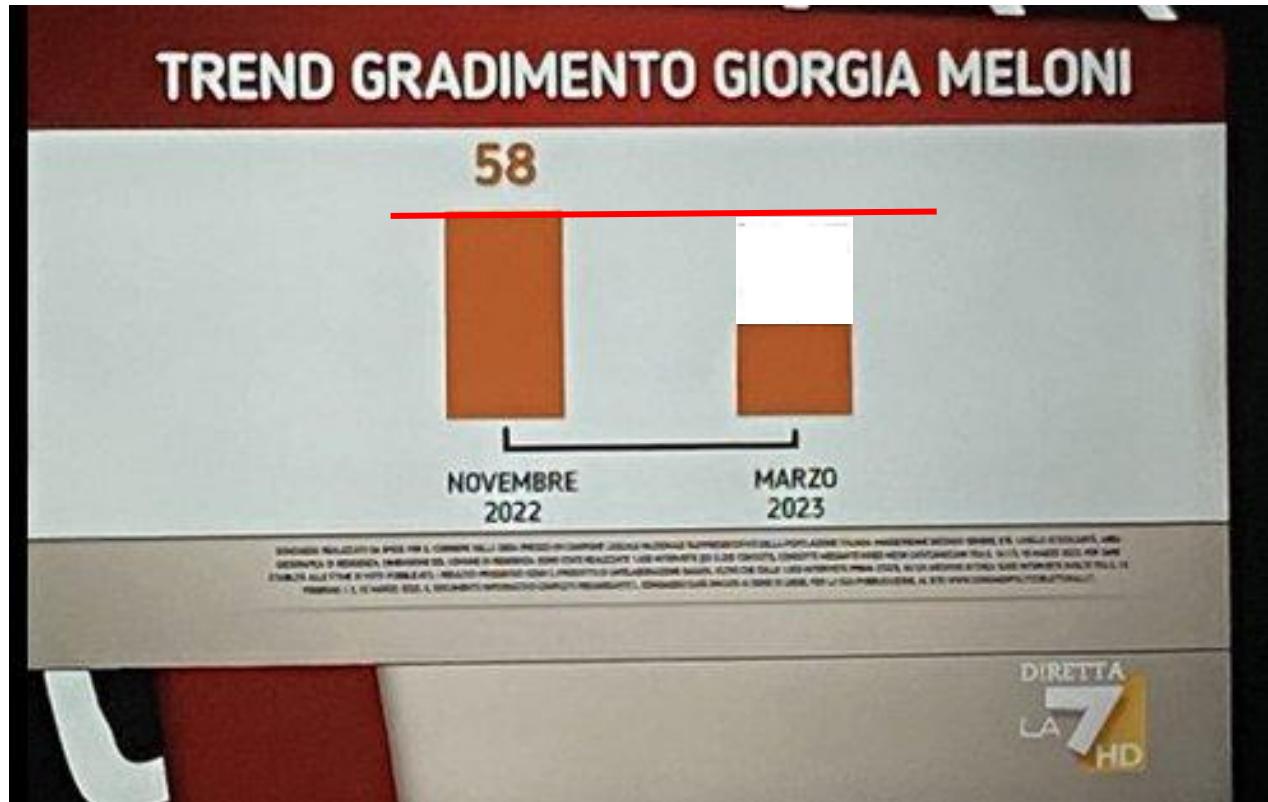
This block is 51%



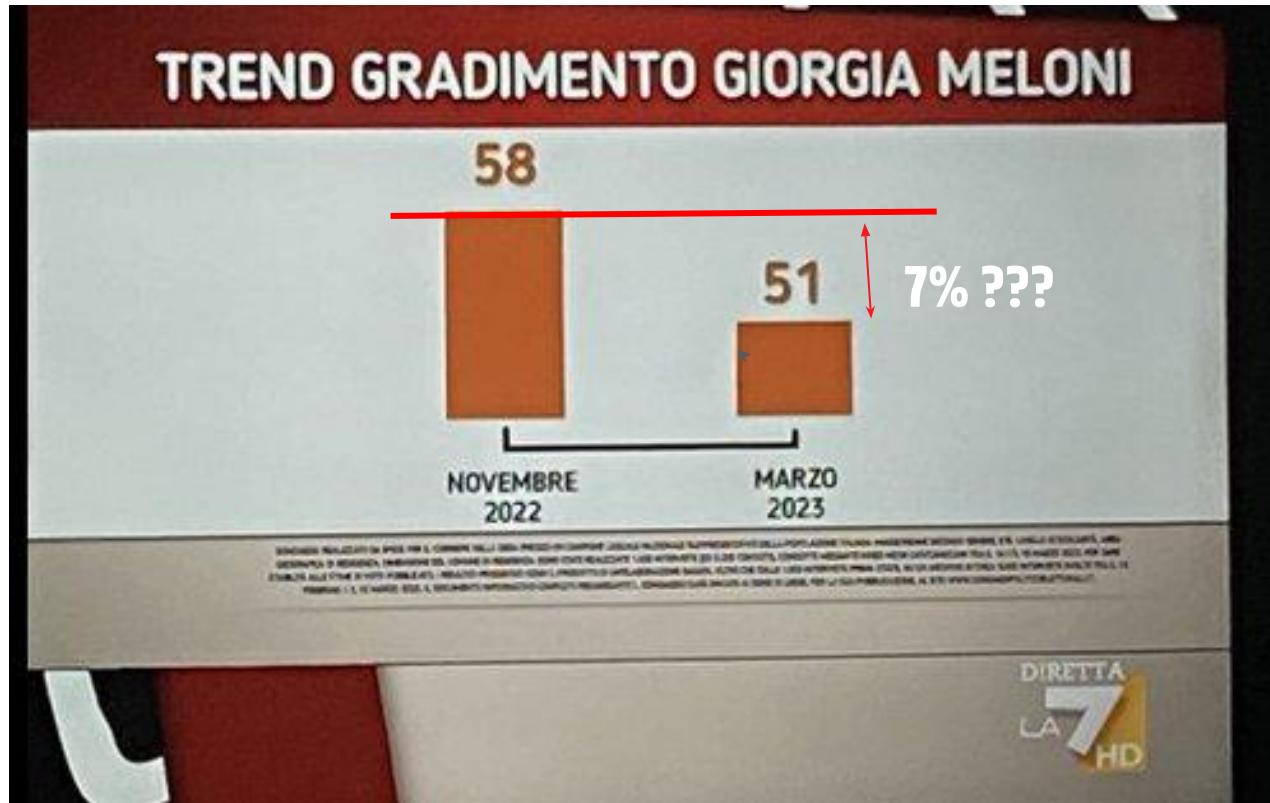
and bring it exactly over (then double)



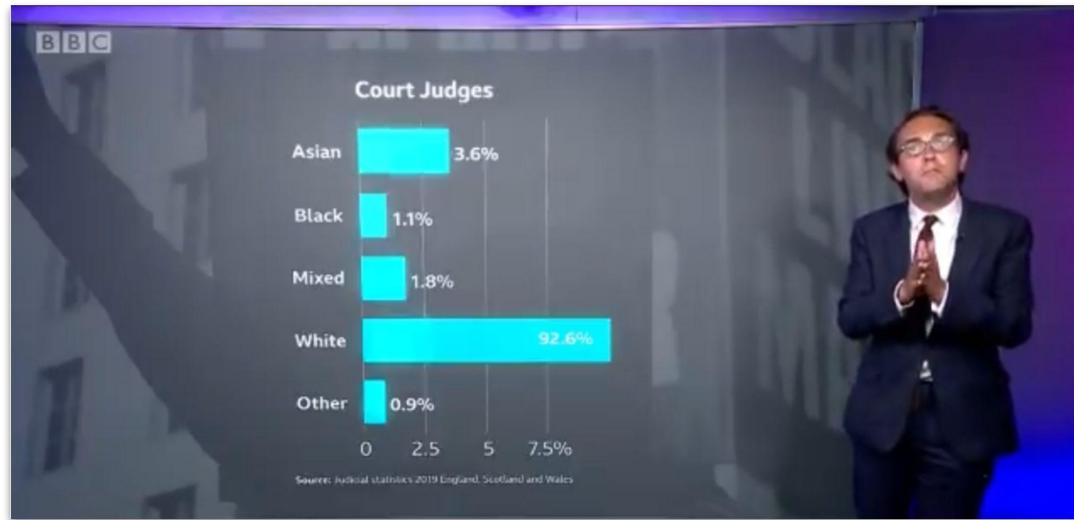
The red line would be 102%...



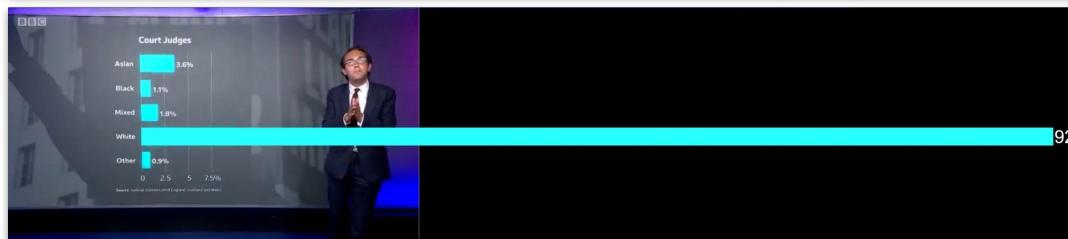
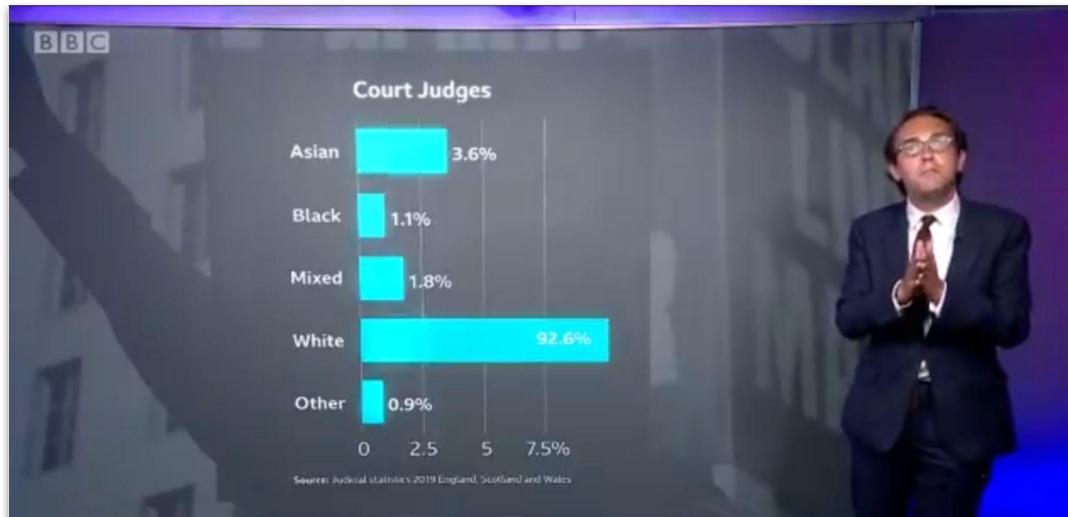
Yet $58-51 = 7$



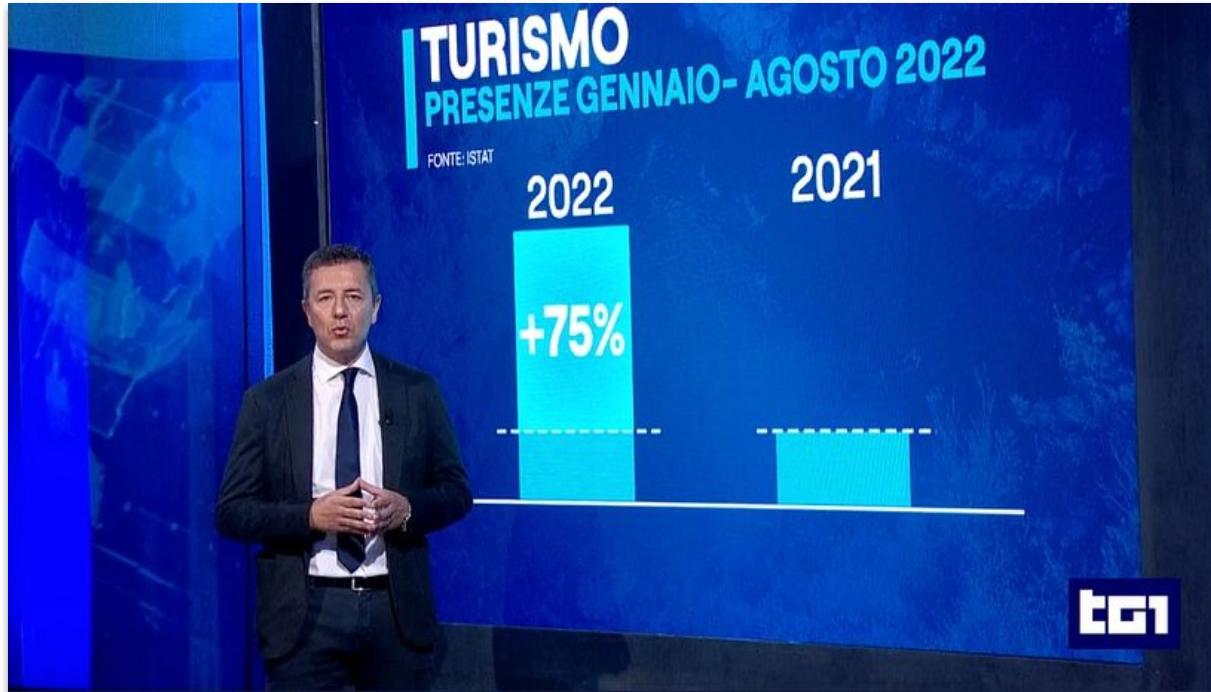
When it doesn't fit on the screen....



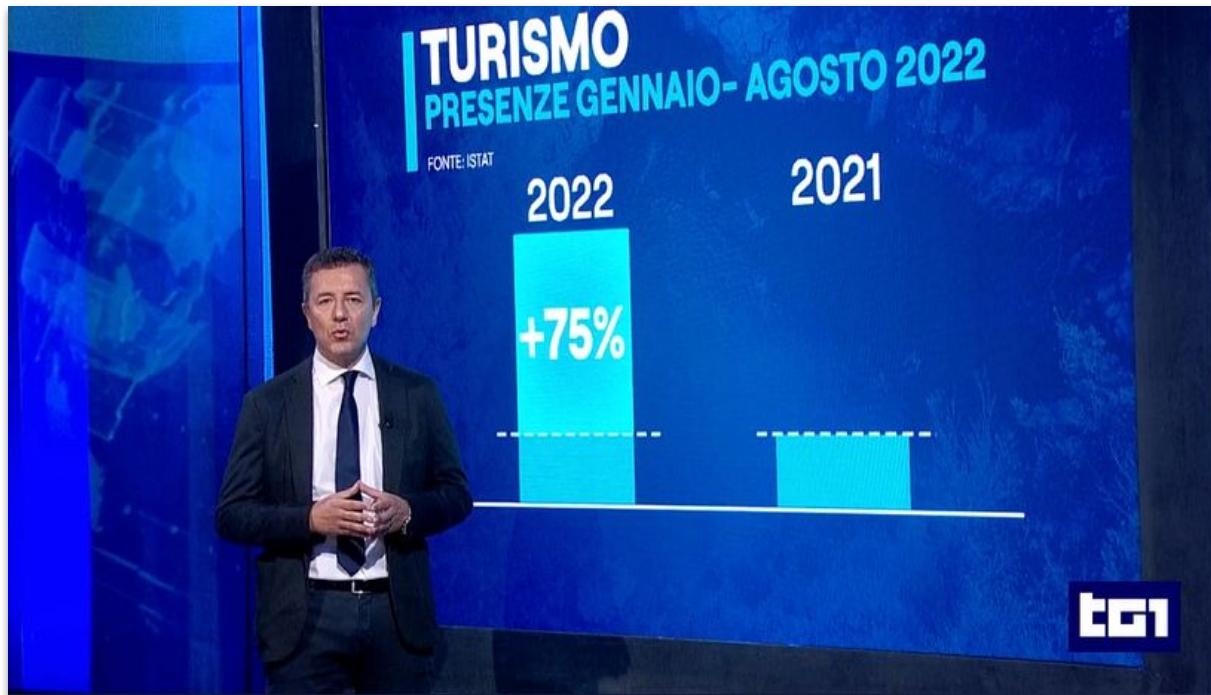
When it doesn't fit on the screen....



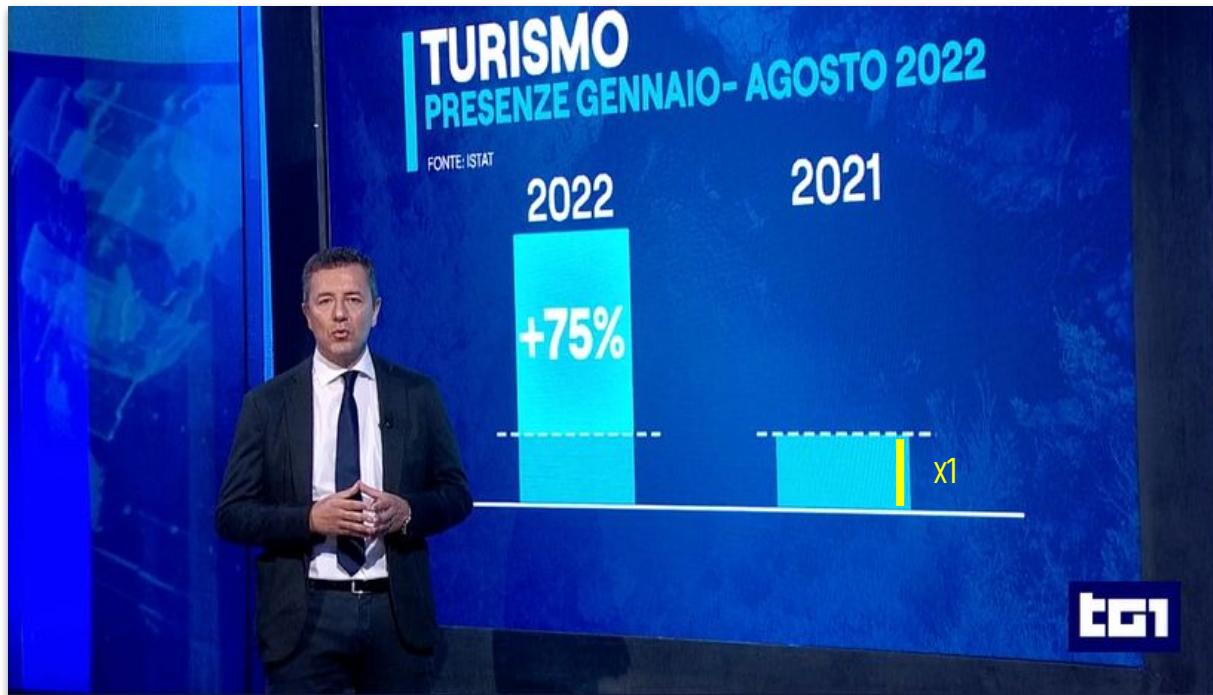
Tourism in Italy increased by 75% from 2021 to 2022



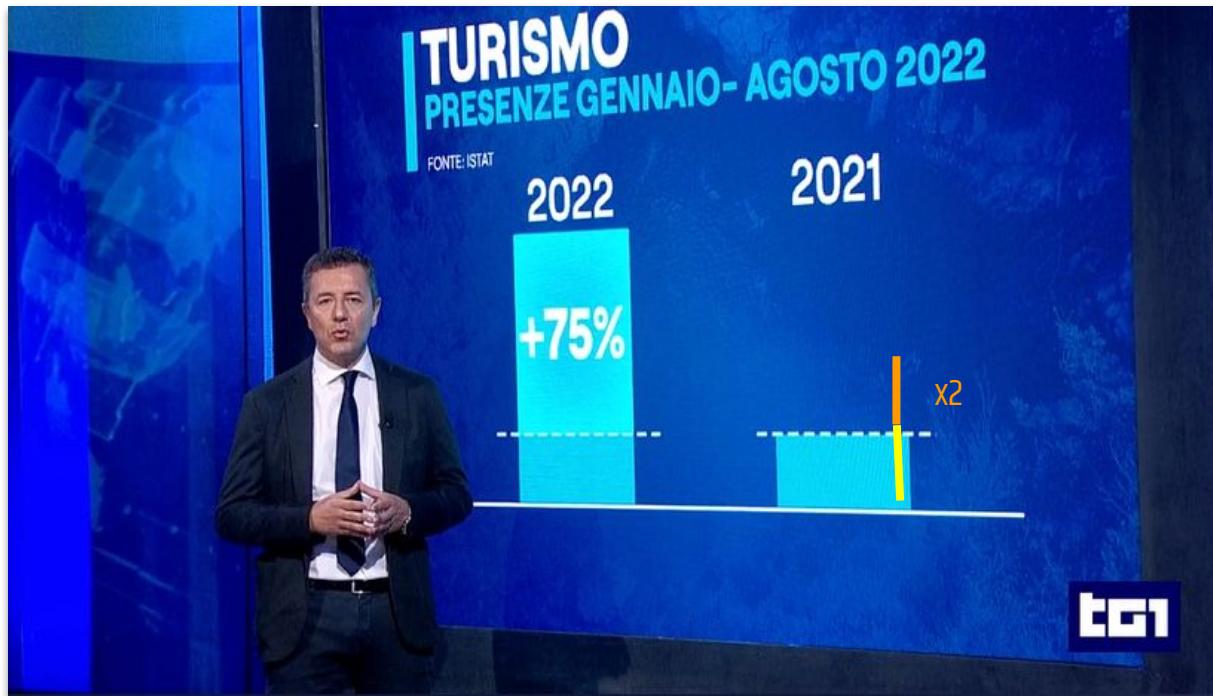
Tourism increased by 75% from 2021 to 2022...
so three quarters more



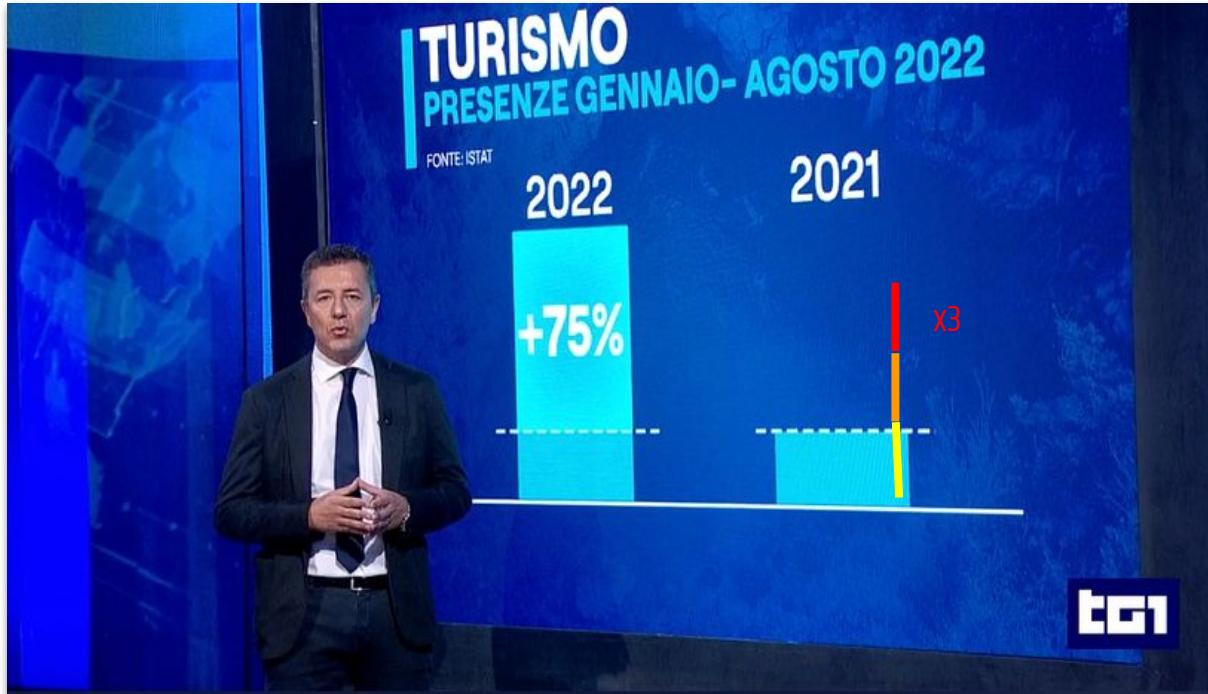
starting unit 2021



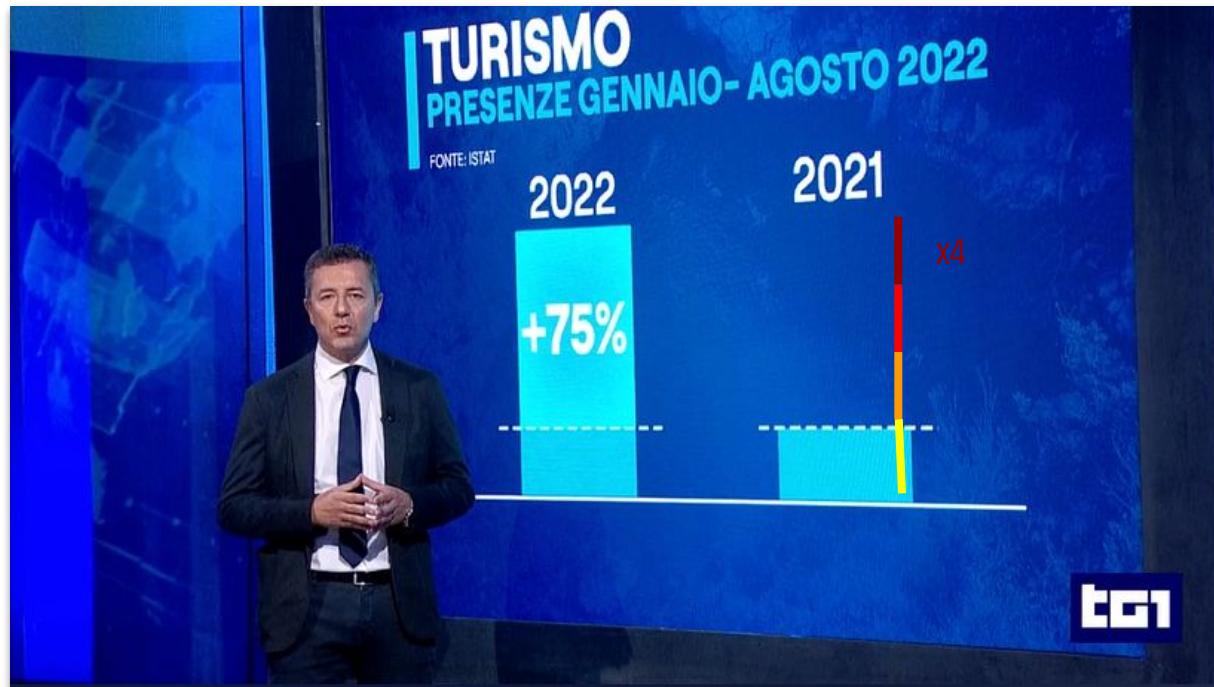
so the double (200%)



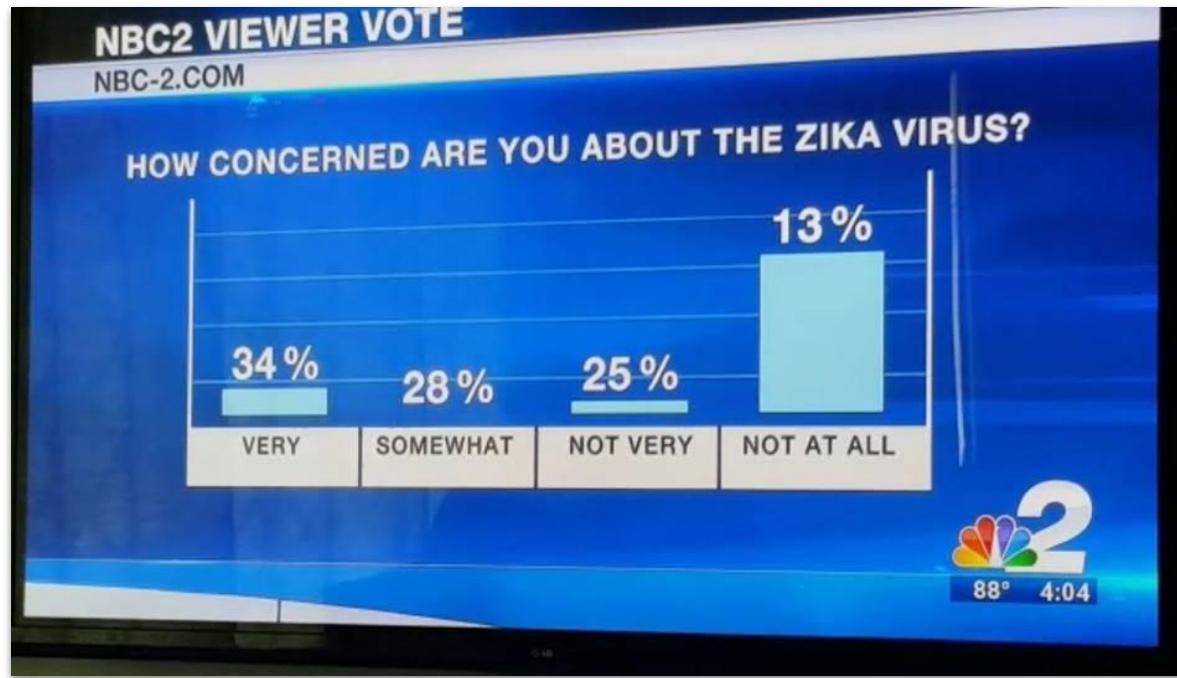
and now the triple (300%)



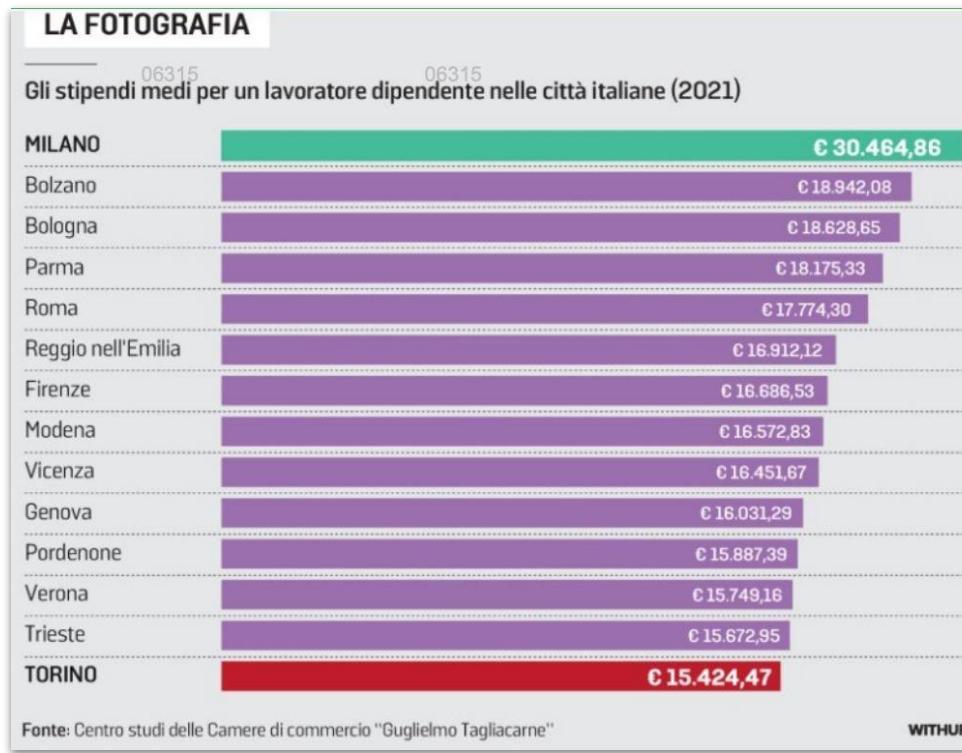
and here quadruple (400%)



A few well confused ideas...



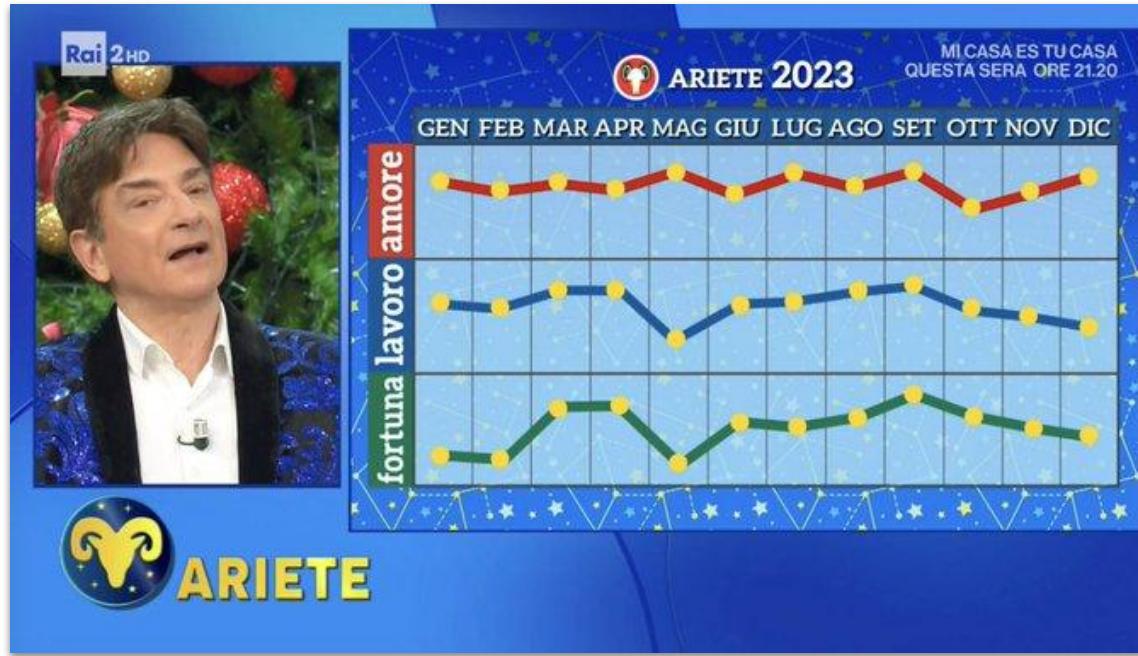
In Milan, the salary is double that in Turin



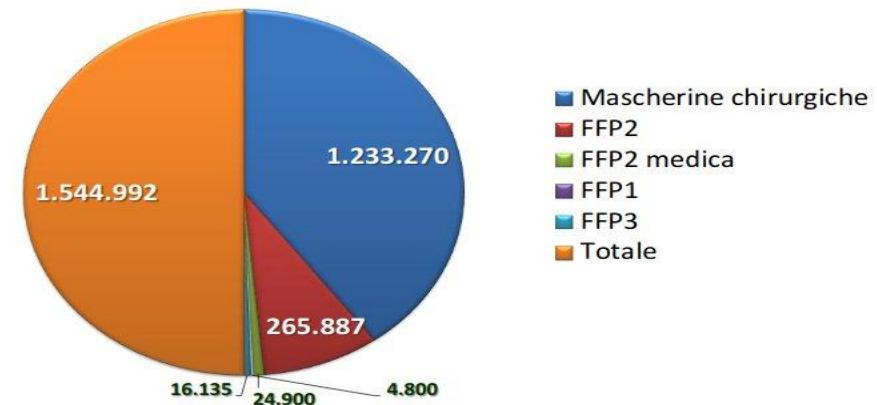
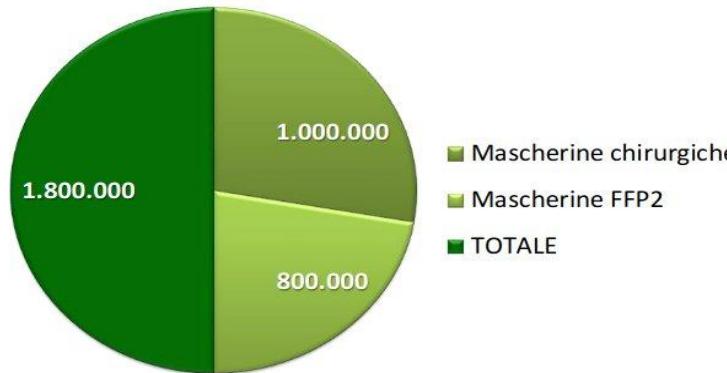
should i look at
the numbers or
the bars?



the progress of the horoscope for the sign of Aries

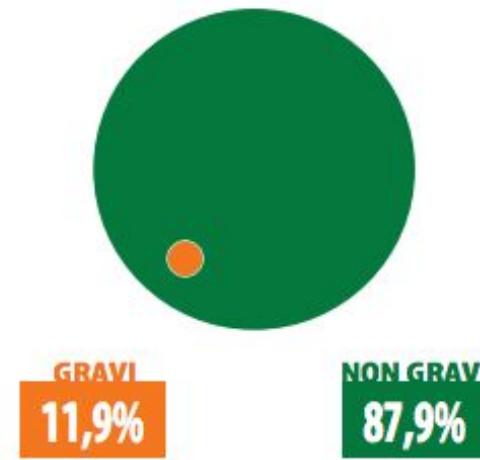


When "total" is a type of mask



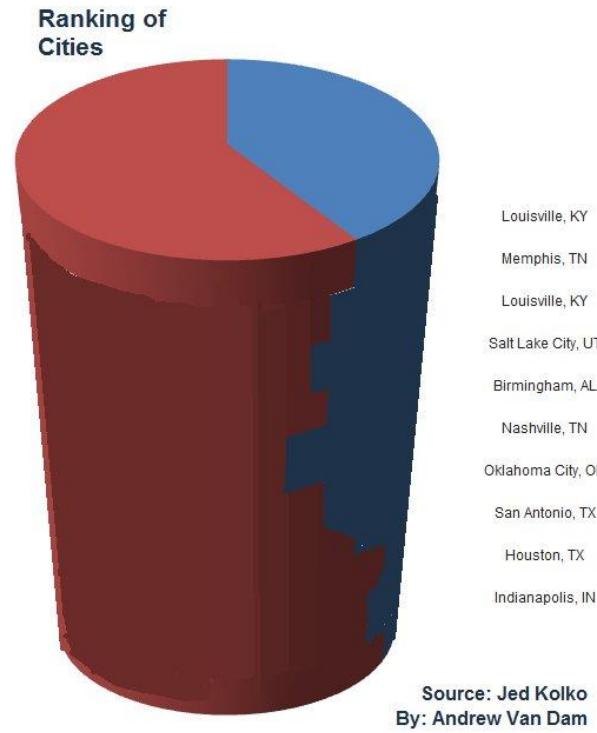
an orange dot that occupies about 12% of the surface???
(about an eighth)

SOSPETTE REAZIONI AVVERSE GRAVI/NON GRAVI

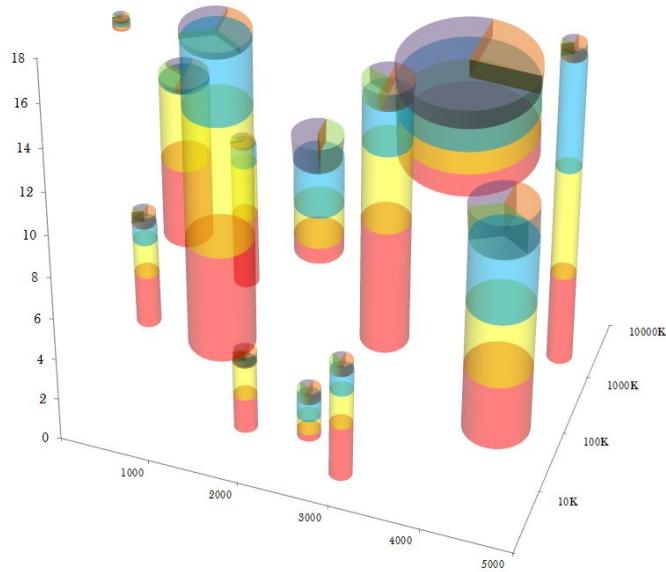


LO 0,2% DELLE SOSPETTE REAZIONI AVVERSE NON È DEFINITO

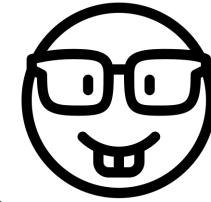
Please help me understand the logic of the ranking...



When you think you are a simplification genius.

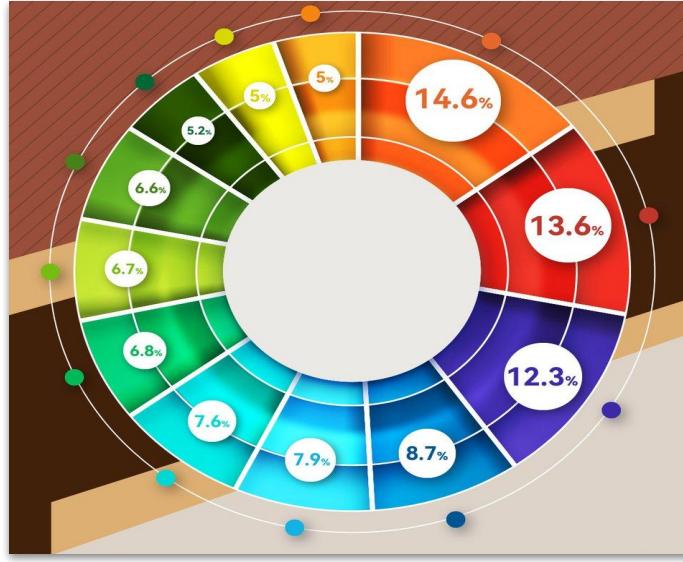


At least seven information-rich variables can be plotted in this new graph module



this guy needs
to take a break

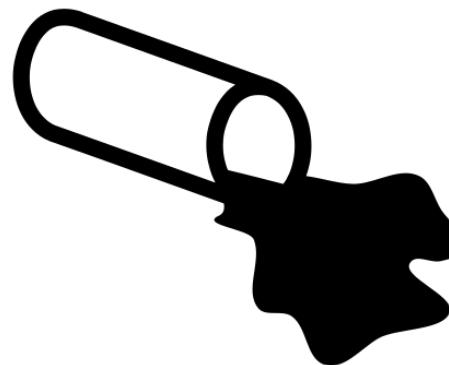
When the legend is used as an explanation (?)



Main Mistakes

no attention
to the ration

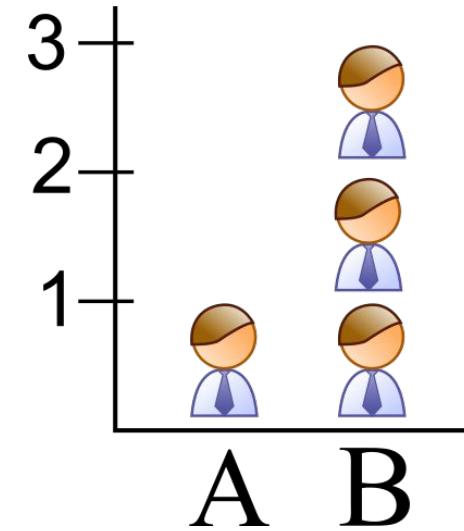
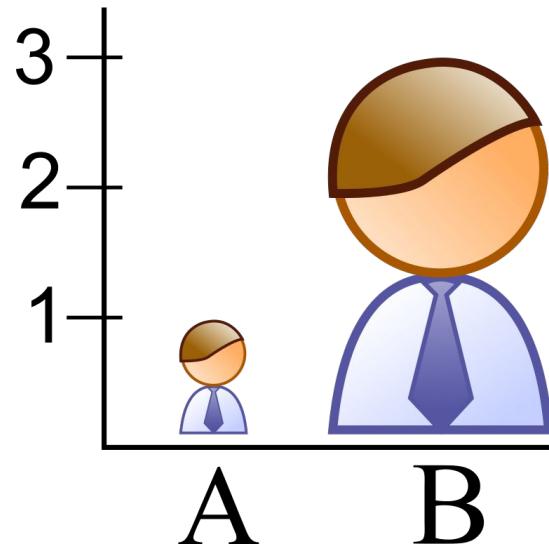
Color abuse



too much
information

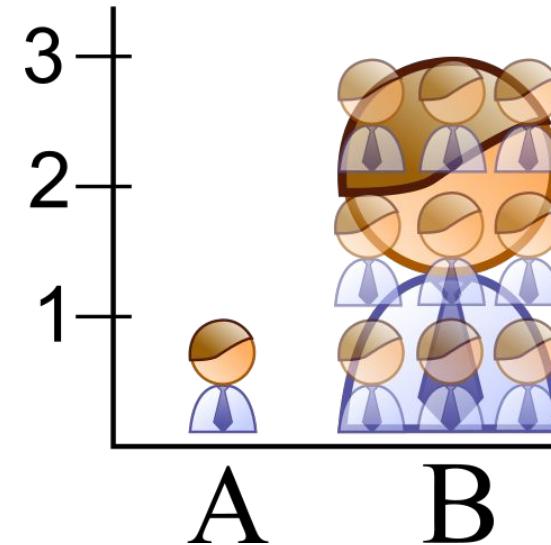
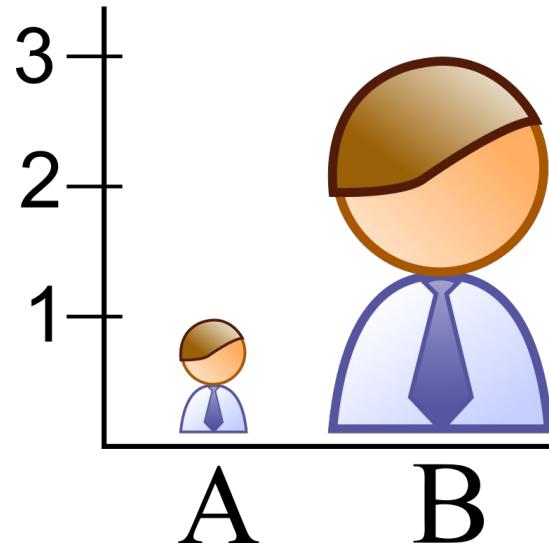
Beautiful
doesn't mean
effective

Respect the ratio



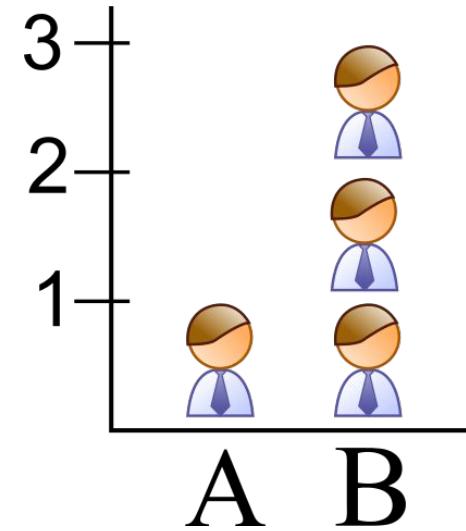
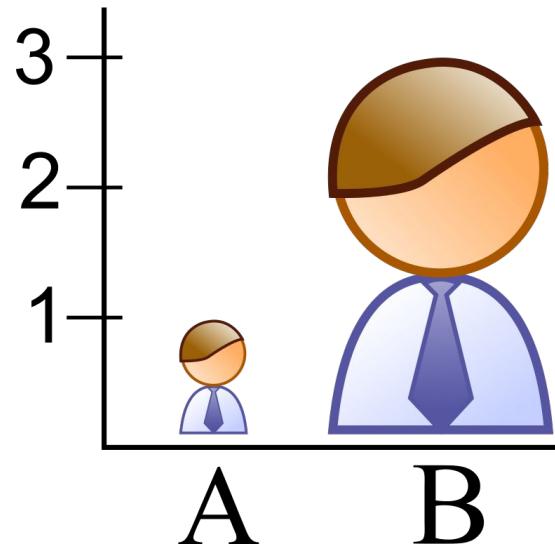
The image on the left should summarize the one on the right

Respect the ratio



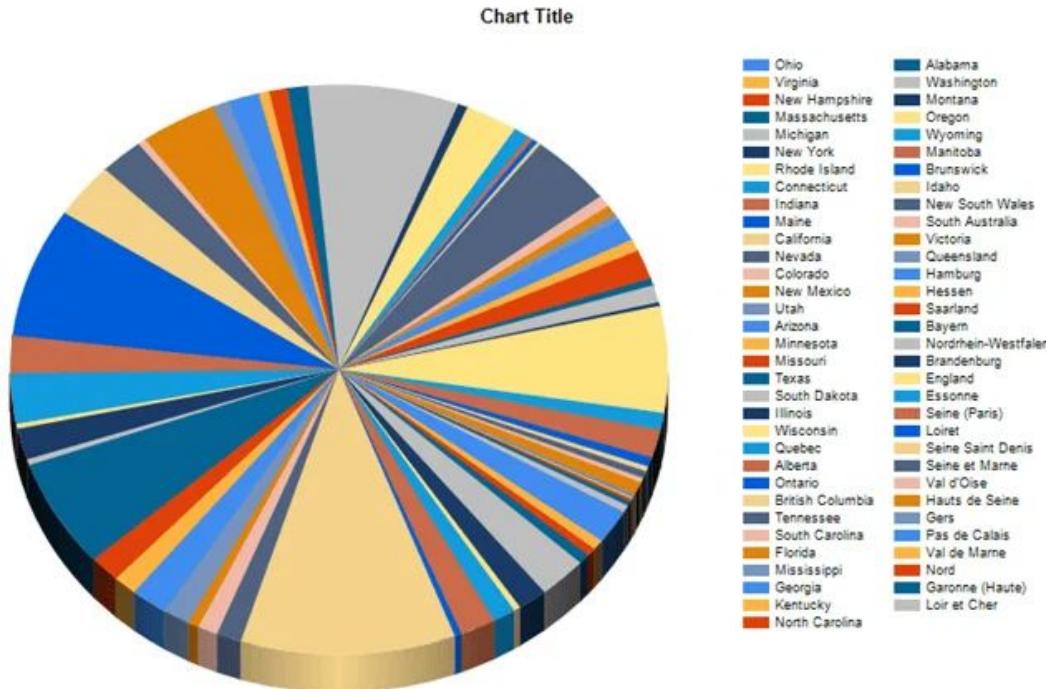
But since it also scales the width ... it is actually representing a wrong value

Use the pictograms well



https://en.wikipedia.org/wiki/Misleading_graph

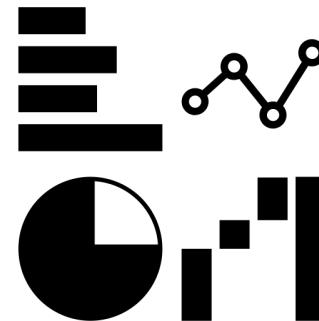
Too much information



Biased Labels

The use of biased or loaded words in the graph's title, axis labels, or caption may inappropriately prime the reader

For example what can you read here? Does it seem simple to you without getting a stiff neck?



EXCESSIVE USE

The **use of graphs** where they are **not needed** can **lead** to unnecessary **confusion**/interpretation.

Generally, the **more explanation a graph needs**, then **less the graph itself is needed**.

Graphs do not always convey information better than tables.

https://en.wikipedia.org/wiki/Misleading_graph

Remove
and move
to improve
the **data tables** edition

Created by Darkhorse Analytics

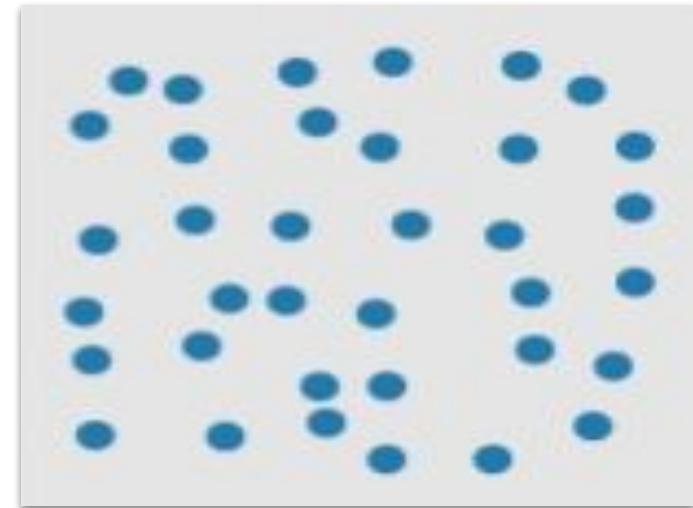
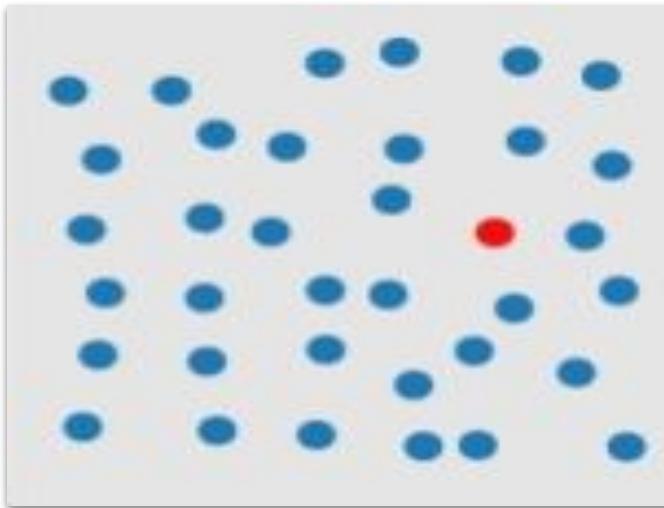
www.darkhorseanalytics.com

<https://www.darkhorseanalytics.com/blog/clear-off-the-table>

Data as a Commons - Maurizio Napolitano



now how to use well the **COLORS**



How many times do you see number three?

432429504454354857609089794218540905454091020
480590439090547601220975090764862459766542119
986775439098550404031029047694421208540878542
570

How many times do you see number three?

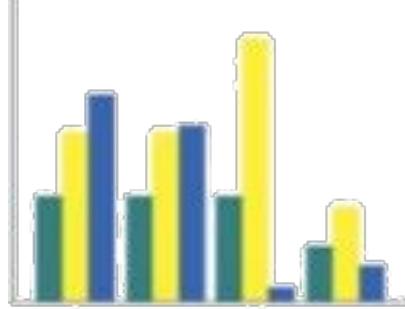
432429504454354857609089794218540905454091020

480590439090547601220975090764862459766542119

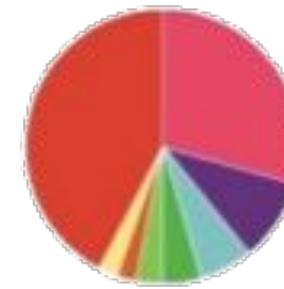
986775439098550404031029047694421208540878542

570

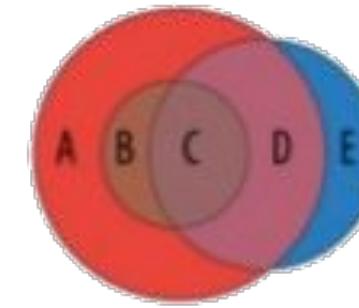
A dominant color



difficulty in distinguishing



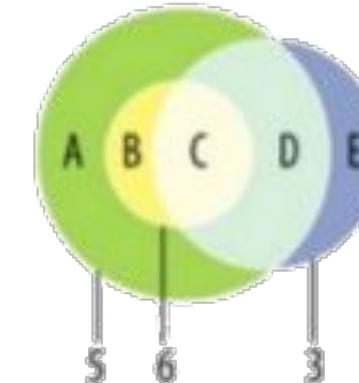
dark colors



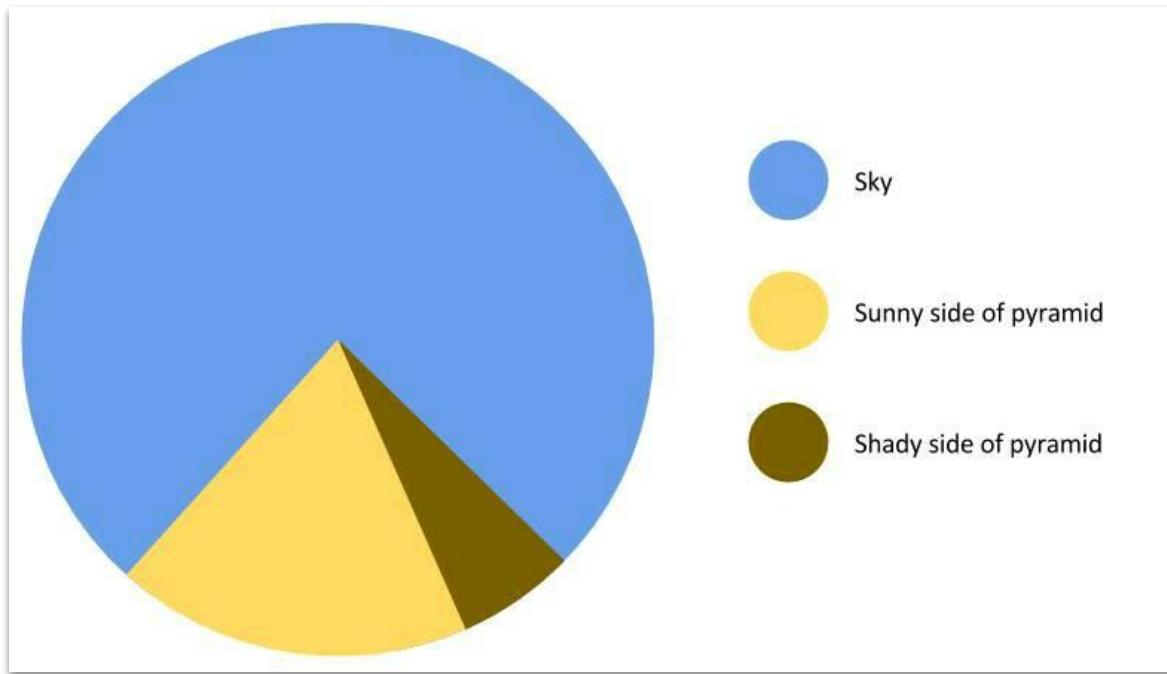
you can correct it after a check on <http://colorbrewer2.org>



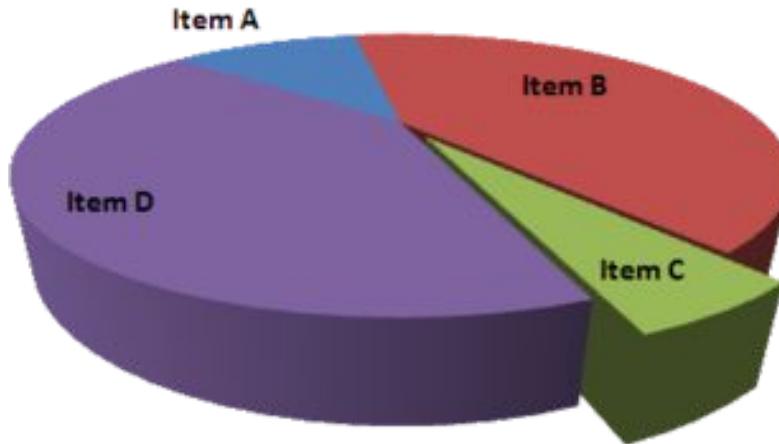
Data as a Commons - Maurizio Napolitano



An egyptian pie chart

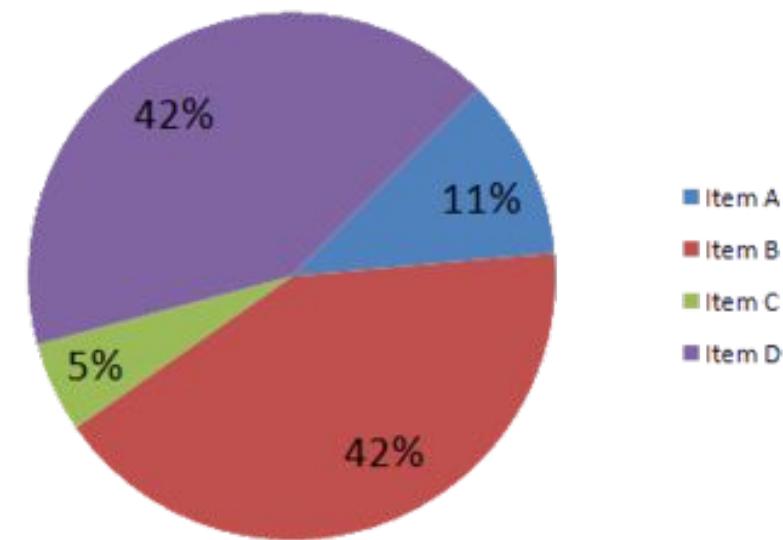


... or a binocular pointed at a pyramid?



The element **C appears** to be at least as **large** as element **A**, while in **reality**, it is **less than half**

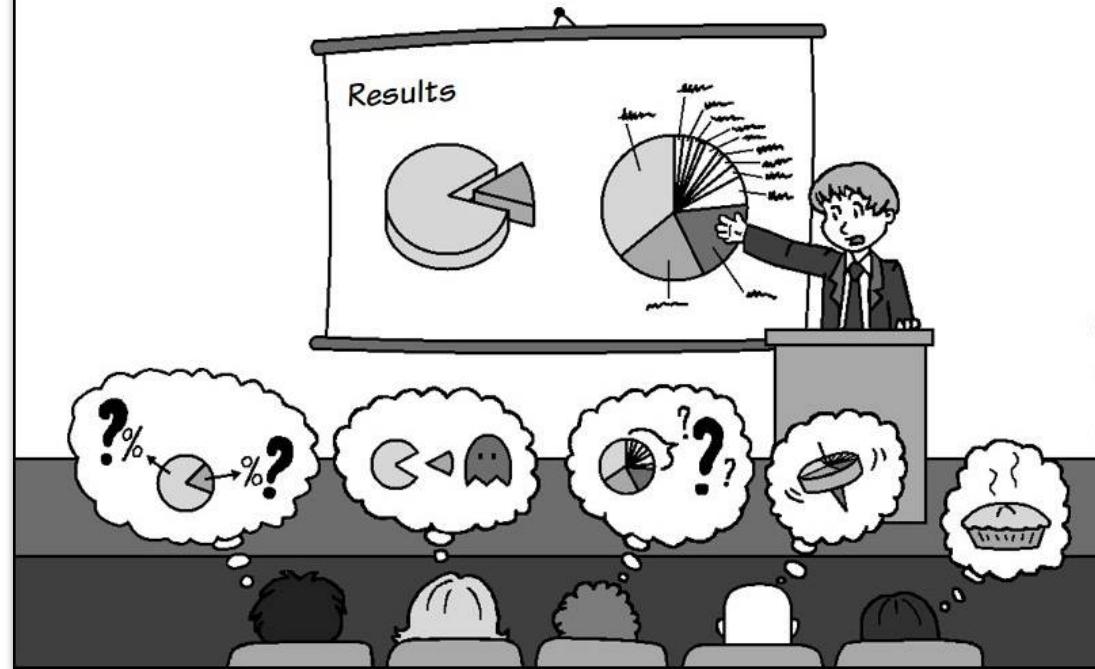
The lies of the 3D pie chart



https://en.wikipedia.org/wiki/Misleading_graph



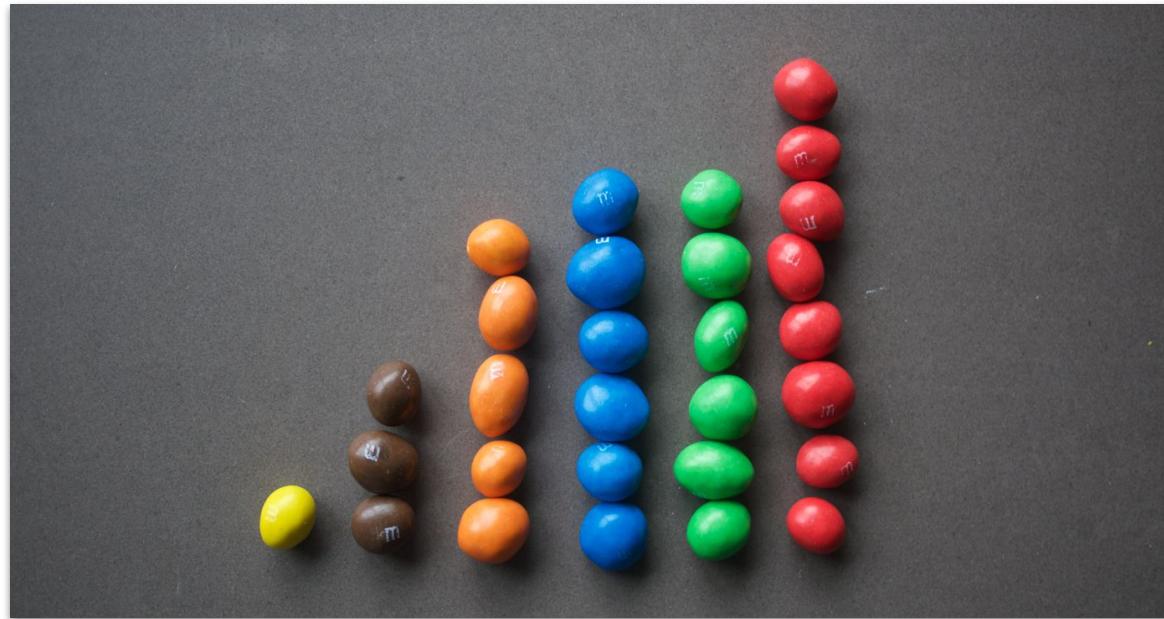
PIE CHART TROUBLE



Which color has the most candies?



Which color has the most candies?

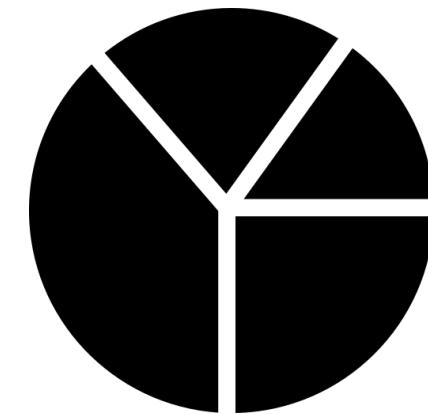


Pie Charts Troubles

Comparing pie charts of **different sizes** could be **misleading** as **people** cannot accurately **read** the comparative **area** of circles.

The usage of **thin slices**, which are hard to discern, may be **difficult to interpret**

The usage of **percentages as labels** on a pie chart can be **misleading** when the sample size is small.

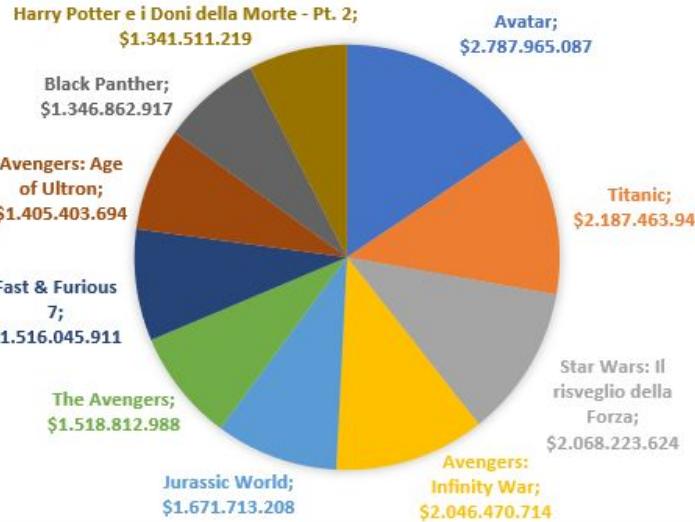


Making a **pie chart 3D** or adding a slant will make **interpretation difficult** due to **distorted effect of perspective**.

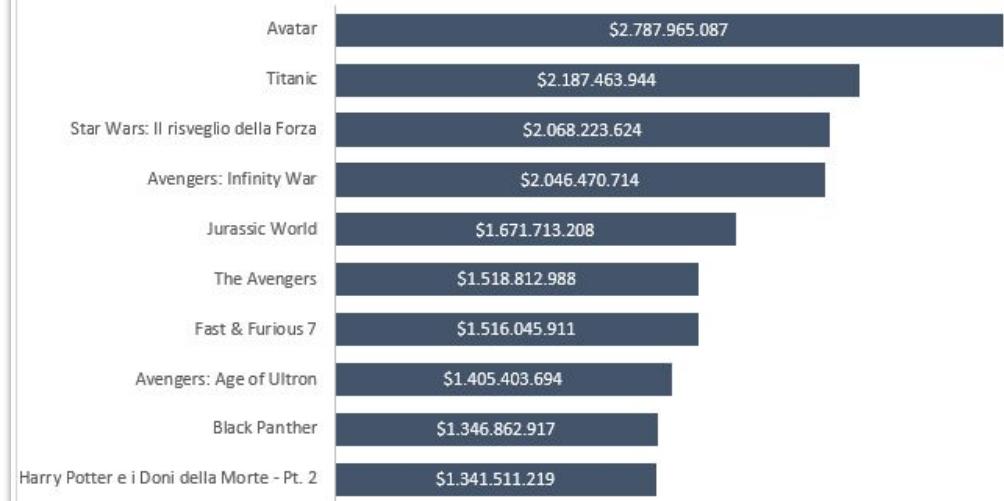
https://en.wikipedia.org/wiki/Misleading_graph

Alternatives: bar charts

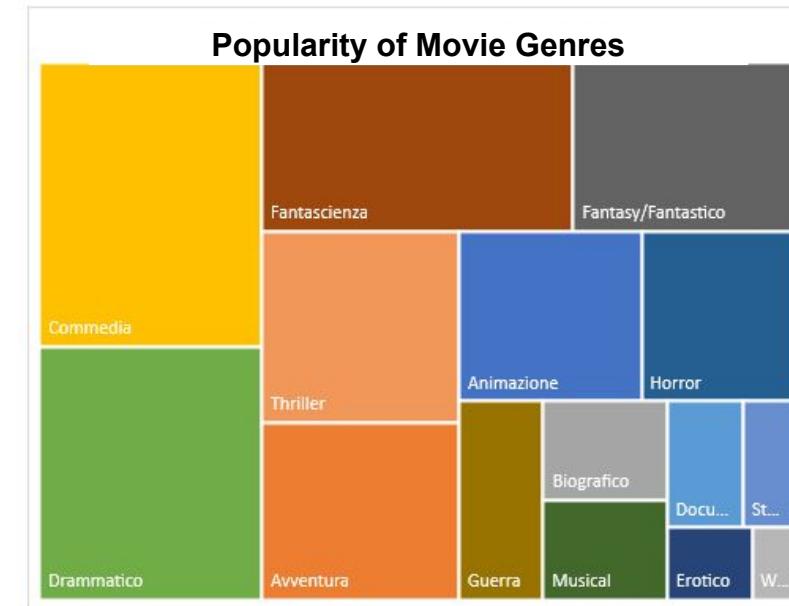
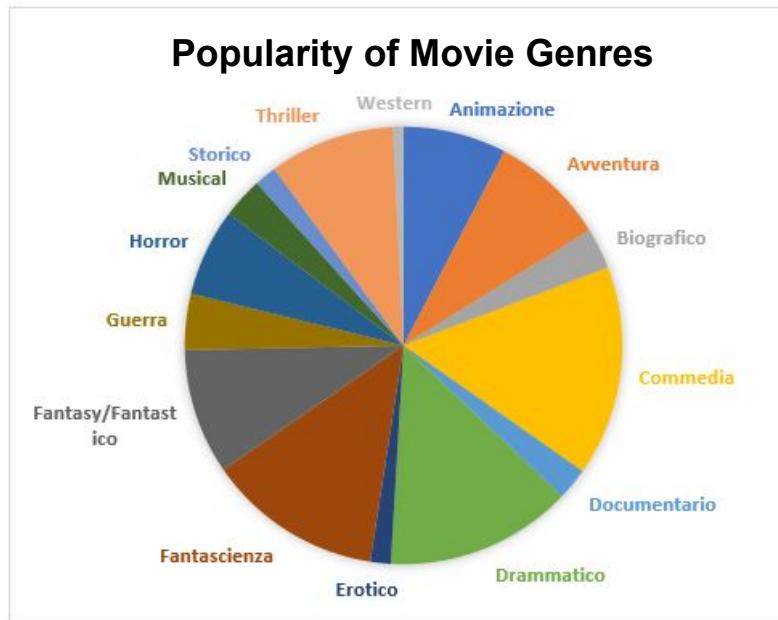
HIGHEST-GROSSING FILMS



HIGHEST-GROSSING FILMS

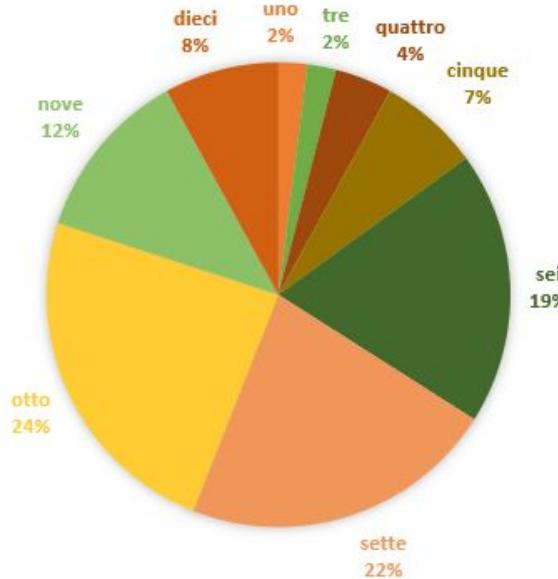


Alternatives: treemap

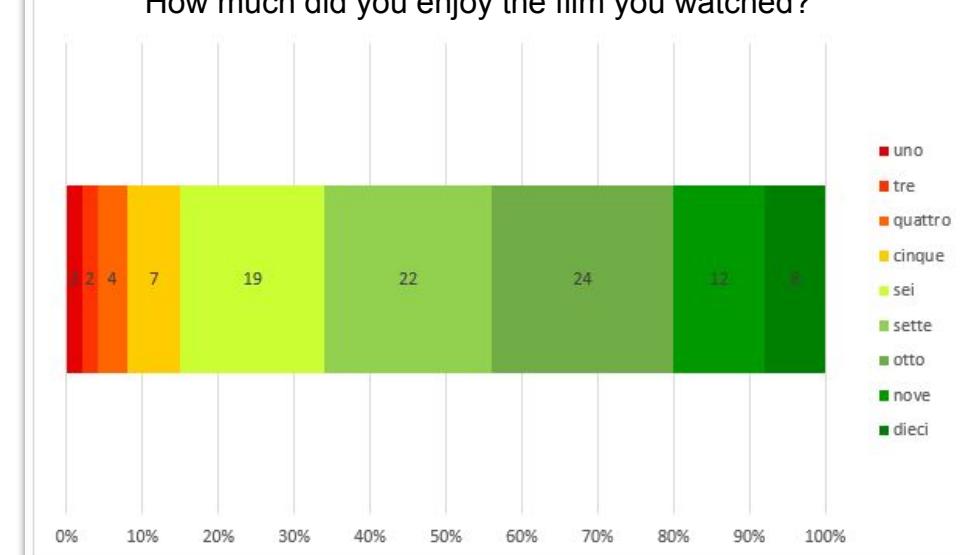


Alternatives: 100% horizontal bar chart

How much did you enjoy the film you watched?



How much did you enjoy the film you watched?



Remove to improve the **pie chart** edition

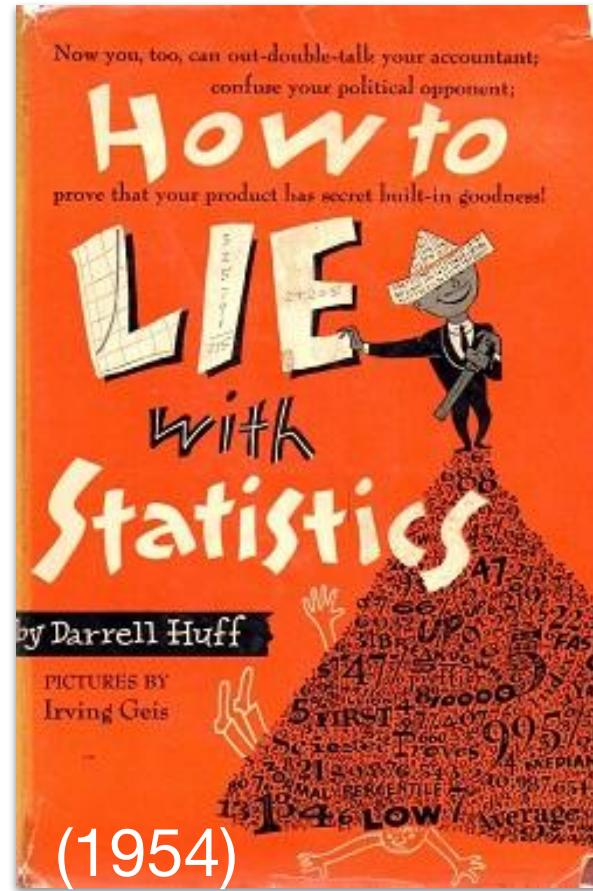
Created by Darkhorse Analytics

www.darkhorseanalytics.com

<https://www.darkhorseanalytics.com/portfolio/2016/1/7/data-looks-better-naked-pie-charts>

Data as a Commons - Maurizio Napolitano





Truncated graph

A truncated graph (also known as a torn graph) has a y **axis** that **does not start at 0**. These graphs can create the impression of important change where there is relatively little change.

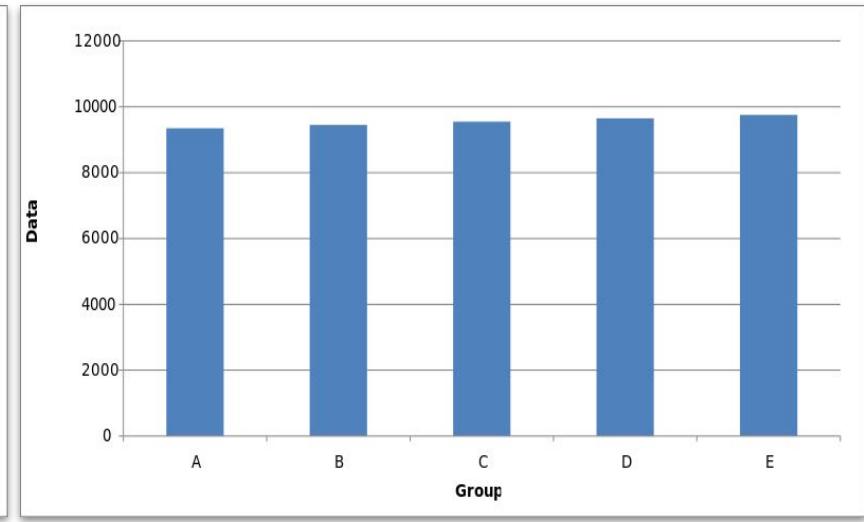
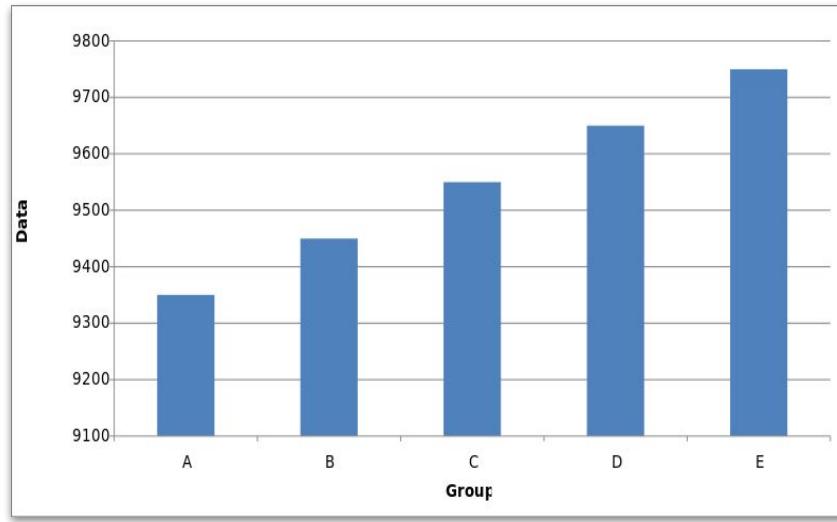
Truncated graphs are **useful** in illustrating **small differences**.

Graphs may also be **truncated to save space**.

Commercial software such as MS Excel will tend to truncate graphs by default if the values are all within a narrow range.

https://en.wikipedia.org/wiki/Misleading_graph

Truncated graph



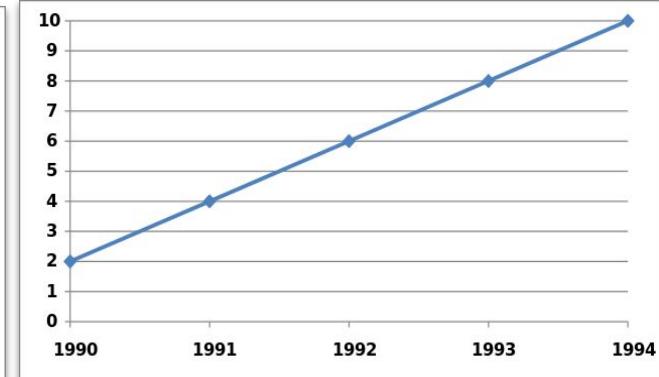
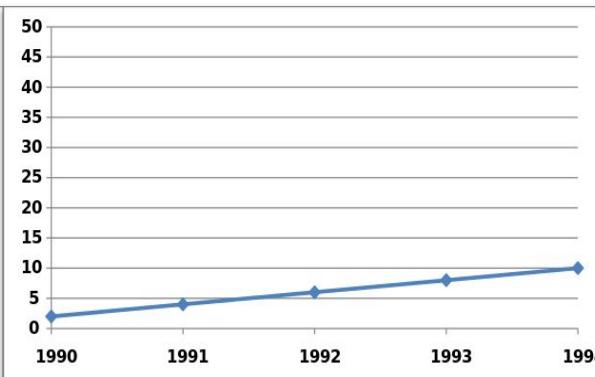
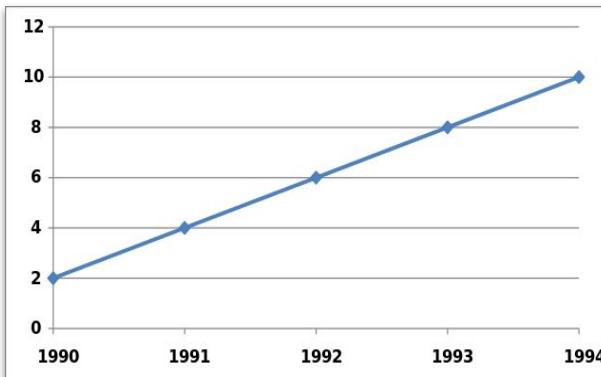
The graphs show identical data

Left: significant differences

Right: the differences are not so significant

https://en.wikipedia.org/wiki/Misleading_graph

Combining Changing the Y Axis



https://en.wikipedia.org/wiki/Misleading_graph

Remove to improve (the **data-ink** ratio)

Created by Darkhorse Analytics www.darkhorseanalytics.com

<https://www.darkhorseanalytics.com/blog/data-looks-better-naked>

The duties of a good data visualization

Answer the questions (or discover them)

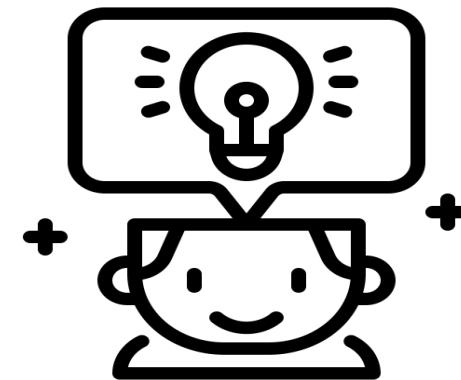
Take decisions

View data in the context

Finding patterns (models)

Present topics or tell a story

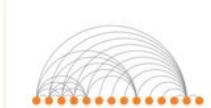
Inspire



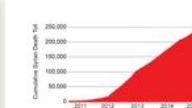
DATA VISUALIZATION & HUMAN RIGHTS KIT PROJECTS RESOURCES LESSONS ABOUT

- Gantt Chart
- Heat Map
- Histogram
- Illustration
- Line Graph
- Marimekko Chart
- Mind Map
- Multi-Set Bar Chart
- Network Diagram
- Pictogram Chart
- Pie Chart
- Population Pyramid
- Proportional Area
- Radar Chart
- Sankey Diagram
- Satellite Image
- Scatter Plot
- Span Chart
- Stacked Area
- Stacked Bar Graph
- Stream Graph

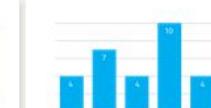
For additional visualization types, visit [The Data Visualisation Catalogue](#).



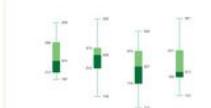
Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Bubble Chart



Bubble Map



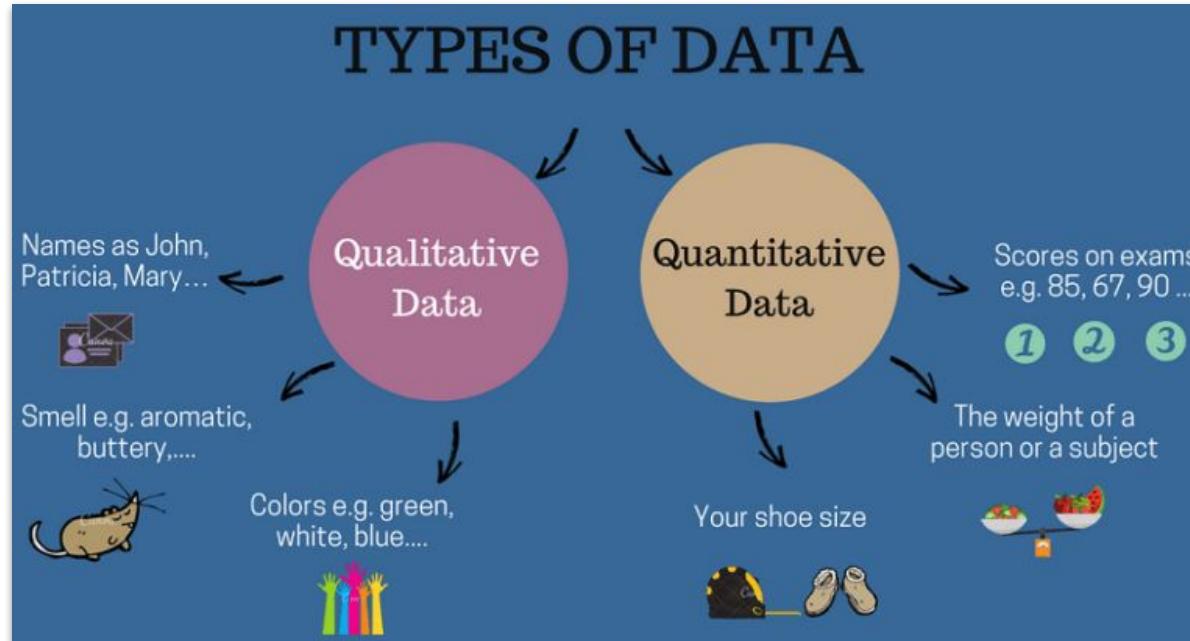
Calendar



Chord Diagram



Choropleth Map



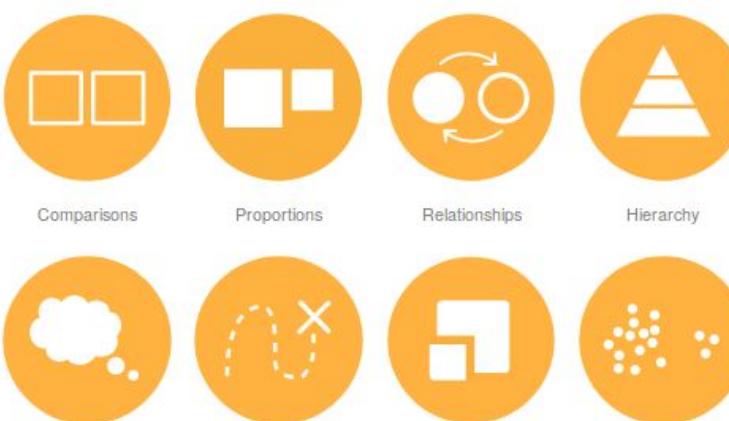
source: <http://intellspot.com/qualitative-vs-quantitative-data/>

The Data Visualisation Catalogue

About • Blog • Shop • Resources

What do you want to show?

Here you can find a list of charts categorised by their data visualization functions or by what you want a chart to communicate to an audience. While the allocation of each chart into specific functions isn't a perfect system, it still works as a useful guide for selecting chart based on your analysis or communication needs.



Comparisons	Proportions	Relationships	Hierarchy
Concepts	Location	Part-to-a-whole	Distribution

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to find the right chart type. If you're not certain in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This is not a guide to be definitive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

© 2014, The Financial Times Ltd. All rights reserved. The Financial Times logo is a registered trademark of The Financial Times Ltd.

ft.com/vocabulary

Deviation
Show the relationship between two variables, highlighting the difference or change from the mean or median.

Correlation
Show the relationship between two variables, highlighting the strength and direction of the correlation.

Ranking
Use when there is a clear order or ranking between items, such as sales, income, or popularity.

Distribution
Show values in a dataset in their order of occurrence, highlighting the tail of outliers, or the distribution's shape.

Change over Time
Show what's changing over time, highlighting trends, seasonality, and other patterns.

Magnitude
Show how large something is, highlighting the scale of differences between values.

Part-to-whole
Show how a specific part relates to the whole, highlighting the proportion of each part to the total.

Spatial
Add the location maps can use to highlight the spatial context of data, such as where it originated, where it's going, or where it's located.

Flow
Show the route or trajectory of data, highlighting the movement of data between locations or geographic regions.

Category	Example	Description
Deviation		A box plot showing the distribution of income across different countries.
Correlation		A scatter plot showing the relationship between age and income.
Ranking		A bar chart ranking companies by revenue.
Distribution		A histogram showing the distribution of ages in a population.
Change over Time		A line chart showing the price of oil over time.
Magnitude		A bar chart showing the magnitude of different earthquakes.
Part-to-whole		A pie chart showing the share of different sectors in the economy.
Spatial		A map showing the location of different cities.
Flow		A network diagram showing the flow of data between different nodes.

Visual Vocabulary	Deviation	Correlation	Ranking	Distribution	Change-over-Time	Magnitude	Part-to-whole	Spatial	Flow	FullList
-------------------	-----------	-------------	---------	--------------	------------------	-----------	---------------	---------	------	----------

Visual Vocabulary - Vega Edition

There are so many ways to visualise data - how do we know which one to pick? Click on a category below to decide which data relationship is most important in your story, then look at the different types of charts within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

Inspired by Financial Times's Visual Vocabulary & Andy Kriebel's ft.
Click any section below to view the charts

Twitter Facebook LinkedIn

Deviation

Emphasise variations (+/-) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show sentiment (positive/neutral /negative)

Correlation

Show the relationship between two or more variables. Be mindful that, unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other)

Ranking

Use where an item's position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Distribution

Show values in a dataset and how often they occur. The shape (or skew) of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data

Change-over-Time

Give emphasis to changing trends. These can be short (intra-day) movements or extended series traversing decades or centuries: Choosing the correct time period is important to provide suitable context for the reader

Magnitude

Show size comparisons. These can be relative (just being able to see larger/bigger) or absolute (need to see fine differences). Usually these show a 'counted' number (for example, barrels, dollars or people) rather than a calculated rate or per cent

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead

Spatial

Used only when precise locations or geographical patterns in data are more important to the reader than anything else.

Flow

Show the reader volumes or intensity of movement between two or more states or conditions. These might be logical sequences or geographical locations

CREATED BY
Pratap Vardhan | @PratapVardhan
CREDITS
Vega Authors & Community, Gramener
<https://gramener.github.io/visual-vocabulary-vega/>
<https://github.com/gramener/visual-vocabulary-vega/>

INSPIRED BY
Andy Kriebel | @VizWizBI (design / data)
FT Graphics: Alan Smith, Chris Campbell, Ian Bott, Liz Faunce, Graham Parish, Billy Ehrenberg-Shannon, Paul McCallum, Martin Stabe
Visual Vocabulary: ft.com/vocabulary

Twitter Facebook LinkedIn

RAWGraphs

The missing link between spreadsheets and data visualization.

USE IT NOW!

FORK IT ON GITHUB



<http://rawgraphs.io>

Data as a Commons - Maurizio Napolitano

[Datawrapper](#)

[River](#)

[Blog](#)

[Academy](#)

[FAQ](#)

[Dashboard](#)

Enrich your stories with charts, in seconds.

[CREATE A CHART](#)

[CREATE A MAP](#)

was a snowstorm in the Northeast and Midwest during the week that the BLS does its survey, which kept some workers at home. Additionally, the "retail apocalypse" of announced store closings meant that more jobs than normal left the economy during the month. This month, the disappointing March number was revised down from 98,000 to 79,000. But the April jobs report provides a bounce back in part because of warmer weather and fewer layoffs. The Labor Department reported gains in hospitality, mining, healthcare, and finance. Including the revisions for the February and March reports, an average of 174,000 jobs were added per month over the last three months.

Monthly Changes in U.S. Employment (Non-Farm), 2006-2017



Source: Bureau of Labor Statistics | [Get The Data](#) | [Embed](#)



3. An interest-rate hike in June now looks more likely.

<http://datawrapper.de>

Data as a Commons - Maurizio Napolitano





FONDAZIONE
BRUNO KESSLER

Books

- "The Functional Art: An Introduction to Information Graphics and Visualization" di Alberto Cairo (2012)
- "The Visual Display of Quantitative Information" di Edward R. Tufte (2001)
- "How to Lie with Statistics" di Darrel Huff (1954)

Web

- <https://www.darkhorseanalytics.com/>
- https://en.wikipedia.org/wiki/Misleading_graph
- <http://callingbullshit.org/>
- <https://www.slideshare.net/janwillemtulp/data-visualization-5724069>
- <https://medium.com/@namwookkim/data-driven-guide-designing-expressive-information-graphics-e068a0c552e9>
- <https://github.com/ft-interactive/chart-doctor/blob/master/visual-vocabulary/Visual-vocabulary.pdf>
- <http://ft-interactive.github.io/visual-vocabulary/>

Tools

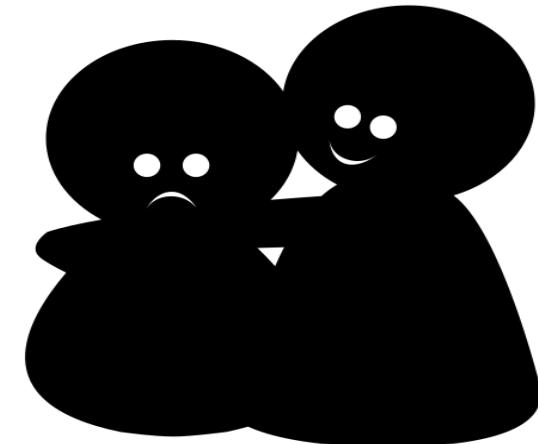
- google spreadsheet - <http://docs.google.com>
- colorbrewer2 - <http://colorbrewer2.org>
- datawrapper - <http://datawrapper.de>
- rawgraphs - <http://rawgraphs.io>
- infogram - <http://infogram.com>
- umap - <https://umap.openstreetmap.de>
- foursquare studio - <https://studio.foursquare.com/>

Follow these hashtags

- #30daychartchallenge
- #horrorchart
- #30daymapchallenge

Conclusion

"A chart should be like a good friend:
it doesn't need many words to make
itself understood"





for contacts:
Maurizio Napolitano



napolitano@fbk.eu



napo

Special thanks
Martina De Sanctis
Matteo Moretti
Andrea Borruso
Lihua Ma
The Noun Project



The presentation is
here

<https://bit.ly/datacommons04>