ITI0209: User Interfaces

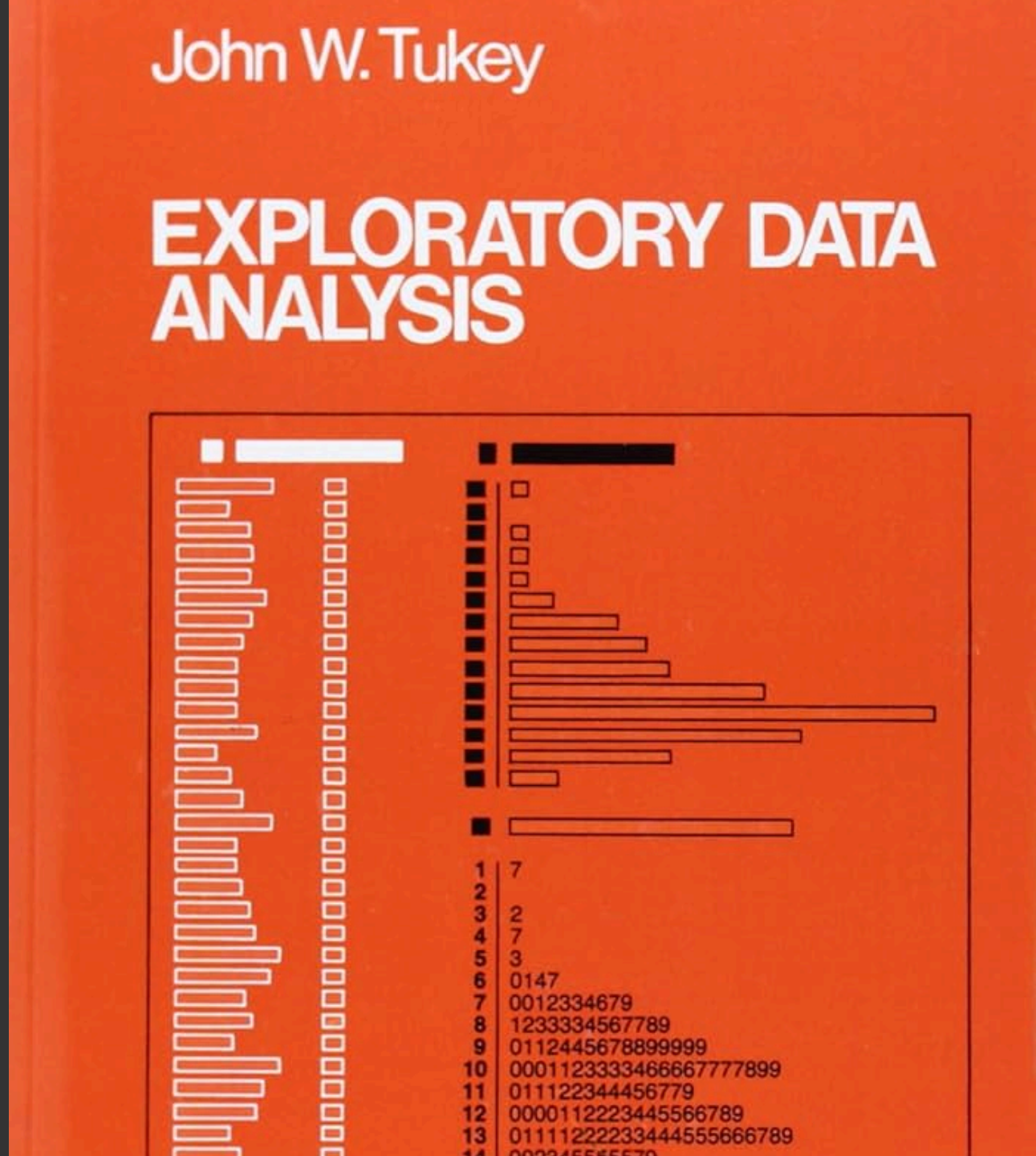
13. Introduction to Data Visualization

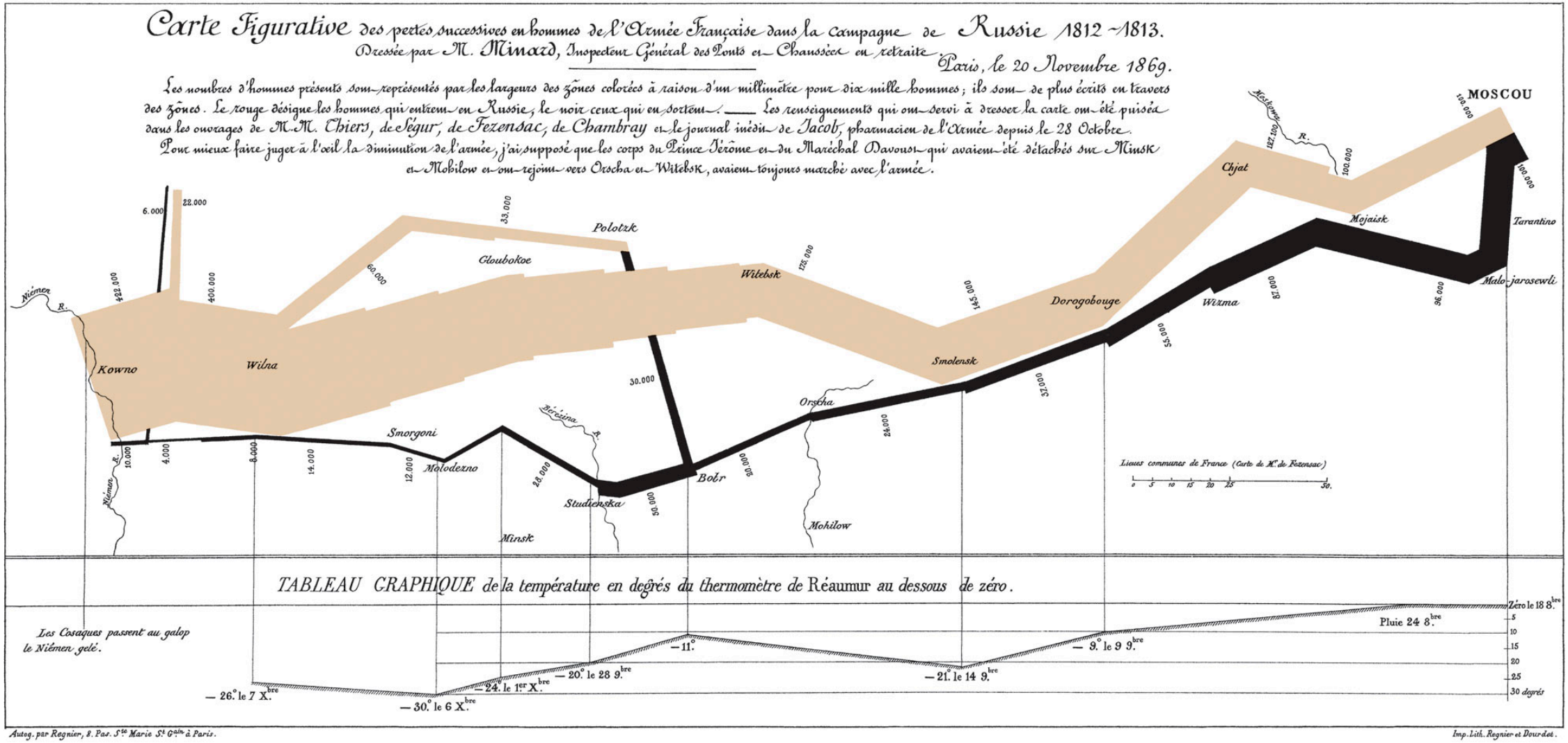
Martin Verrev

Spring 2024

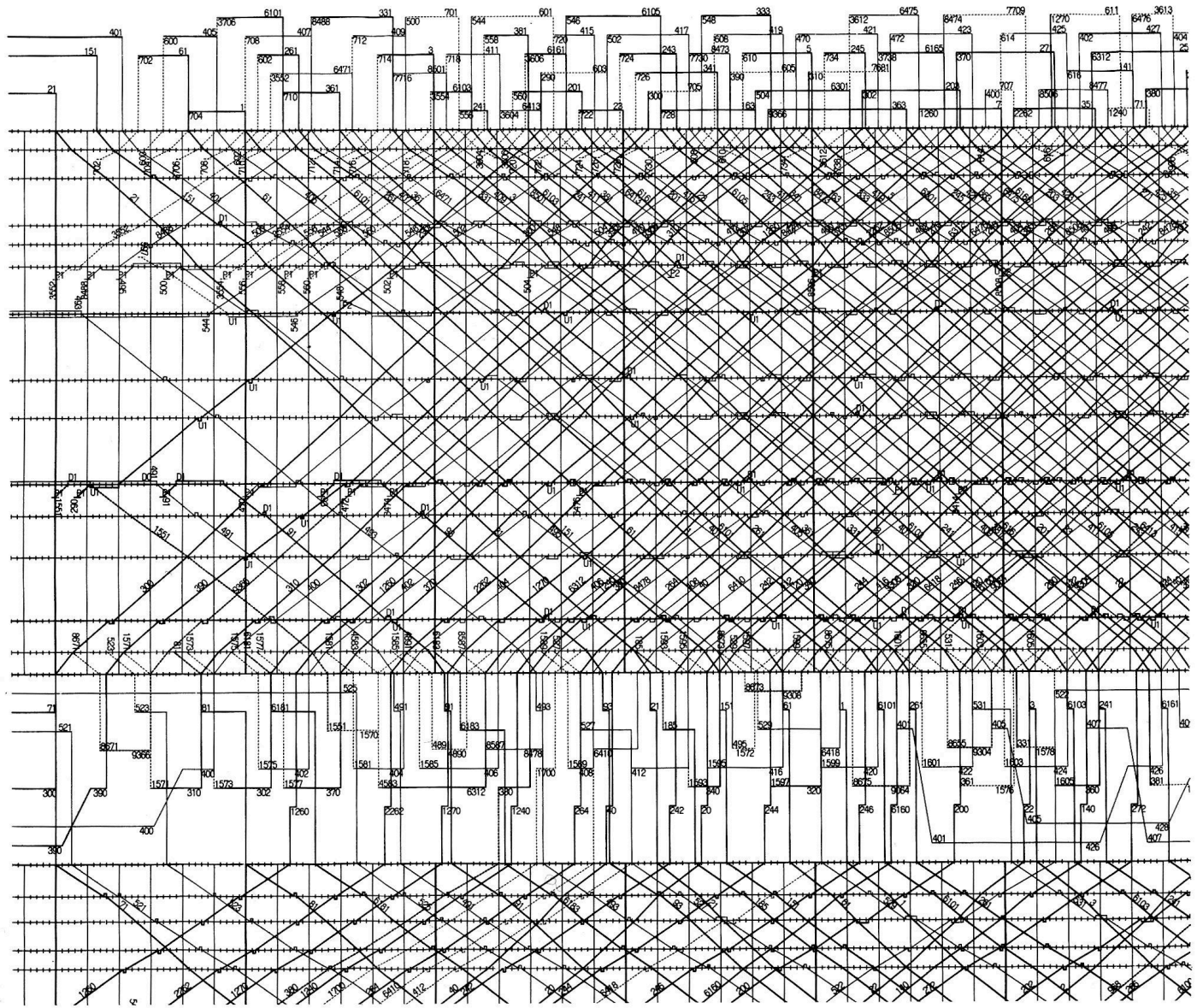
"The greatest value of a picture is when it forces us to notice what we never expected to see."

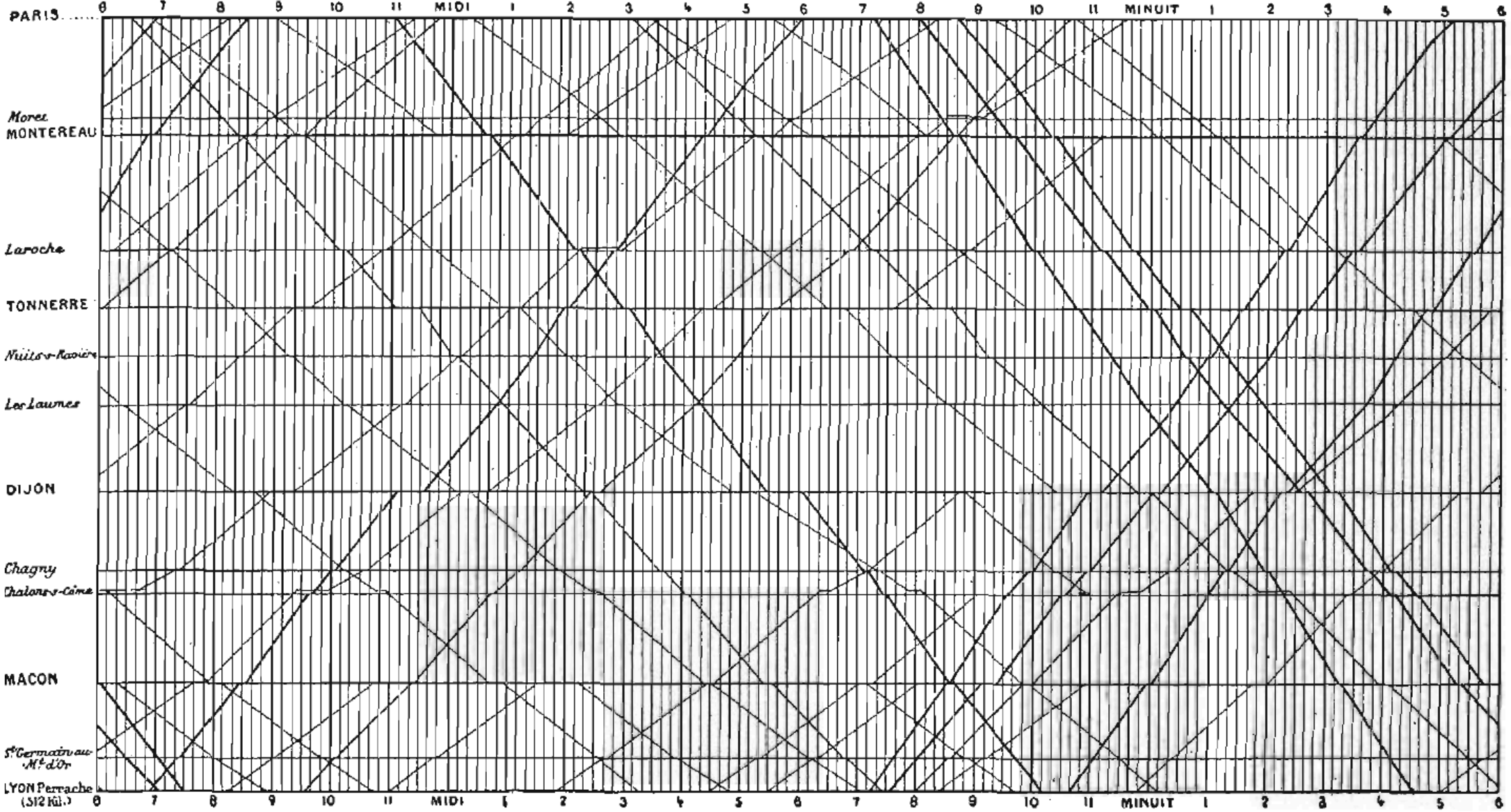
John Tukey, Exploratory Data Analysis





https://www.esri.com/content/dam/esrisites/en-us/esri-press/pdfs/mapping-time-illustrated-minards-map-napoleons-russian-campaign-1812-sample-chapter.pdf





E. J. Marey, *La méthode graphique* (Paris, 1885), p. 20. The method is attributed to the French engineer, Ibry.

Visualizing information is important

The brain doesn’t just process information that comes though the eyes. It also creates mental visual images that allow us to reason and plan actions that facilitate survival. Simply put - a graphical representation is meant to simplify information and make it effortlessly scannable.

The Purpose and Goal

The first and main goal of any graphic and visualization is to be a tool for your eyes and brain to perceive what lies beyond their natural reach.

Check:

https://informationisbeautiful.net/

https://informationisbeautiful.net/beautifulnews/

https://www.reddit.com/r/dataisbeautiful

Why Does it matter?

- **Digestibility:** It makes data easy to digest, turning the daunting into the delightful.
- **Pattern Discovery:** Uncover hidden patterns and correlations within your data, making the complex seem surprisingly simple.
- **Information Compilation:** Imagine having all your information neatly compiled in one place — a visual feast for your analytical appetite.
- **Enhanced Understanding and Retention:** Data visualization not only aids understanding but also boosts retention. Think of it like a well-crafted YouTube video that leaves a lasting impression.

What is Graphical Excellence

1. Well designed presentation of interesting data - a matter of substance,statistics and design
2. Consists of complex ideas communicated with clarity, precision and efficiency
3. That what gives the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space
4. Nearly always multivariate
5. Requires telling the truth about data

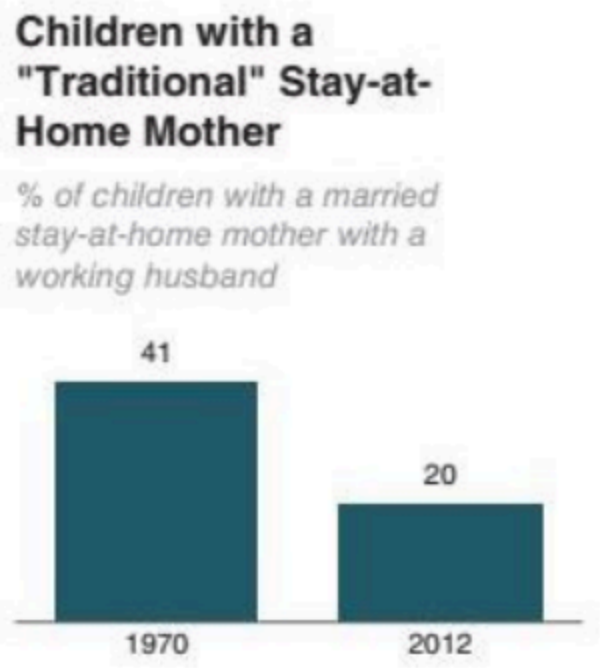
Choosing a visualization

There is an infinite number of different visualization types (check https://d3js.org/). Handful of types will work for the majority of your needs:

Simple Text. Table. Point. Line. Slopegraph. Bars. Area.

Simple Text

If just a number or two to share



Tables

Tables interact with our verbal system, which means that we read them. If you need to communicate multiple different units of measure it is usually easier with a table than a graph.

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Light borders

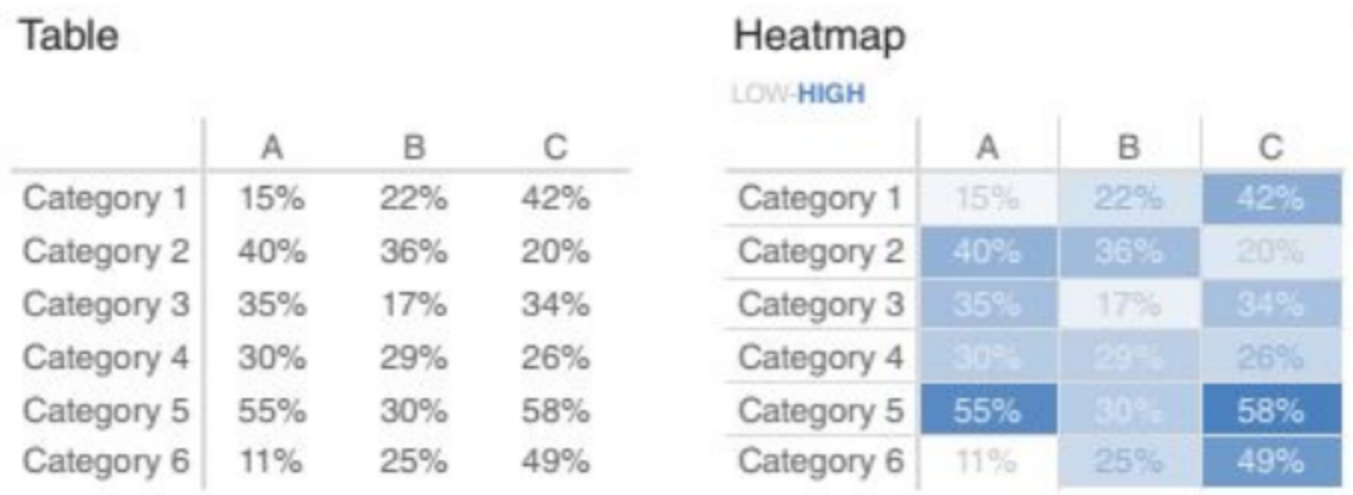
Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Tables. Heatmap

A heatmap is a way to visualize data in tabular format, where in place of (or in addition to) the numbers, you leverage colored cells that convey the relative magnitude of the numbers.

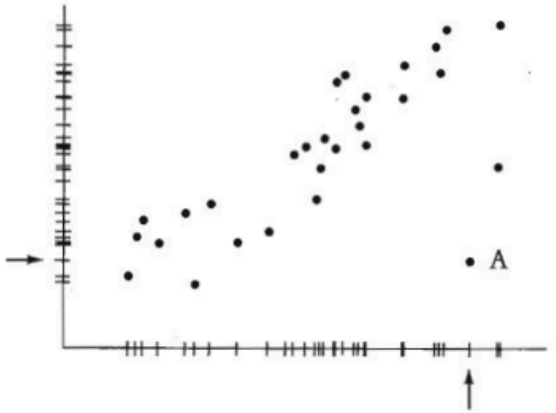


Graphs.

While tables interact with our verbal system, graphs interact with our visual system, which is faster at processing information.

I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89

$N = 11$
mean of X's = 9.0
mean of Y's = 7.5
equation of regression line: $Y = 3 + 0.5X$
standard error of estimate of slope = 0.118
$t = 4.24$
sum of squares $X - \overline{X} = 110.0$
regression sum of squares = 27.50
residual sum of squares of Y = 13.75
correlation coefficient = .82
$r^2 = .67$



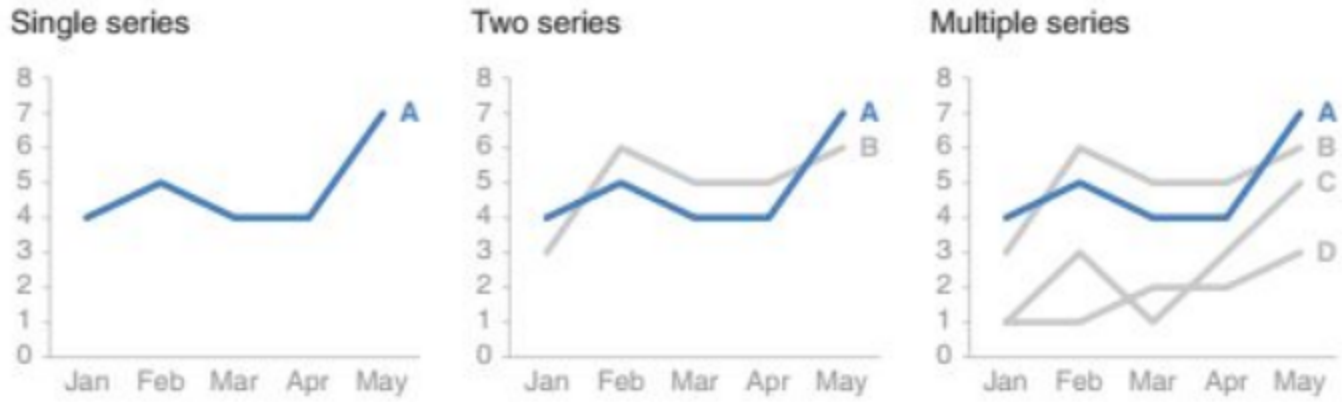
Stephen S. Brier and Stephen E. Fienberg, "Recent Econometric Modelling of Crime and Punishment: Support for the Deterrence Hypothesis?" in Stephen E. Fienberg and Albert J. Reiss, Jr., eds., *Indicators of Crime and Criminal Justice: Quantitative Studies* (Washington, D.C., 1980), p. 89.

Scatterplots

Scatterplots can be useful for showing the relationship between two things, because they allow you to encode data simultaneously on a horizontal x-axis and vertical y-axis to see whether and what relationship exists.

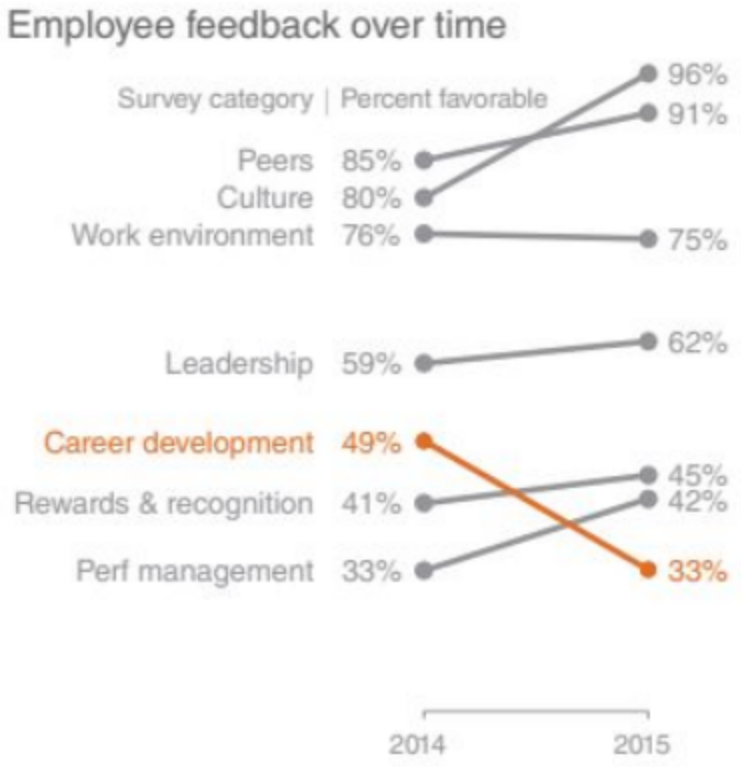
Line

Line graphs are most commonly used to plot continuous data. Because the points are physically connected via the line, it implies a connection between the points that may not make sense for categorical data (a set of data that is sorted or divided into different categories). Often, our continuous data is in some unit of time: days, months, quarters, or years.



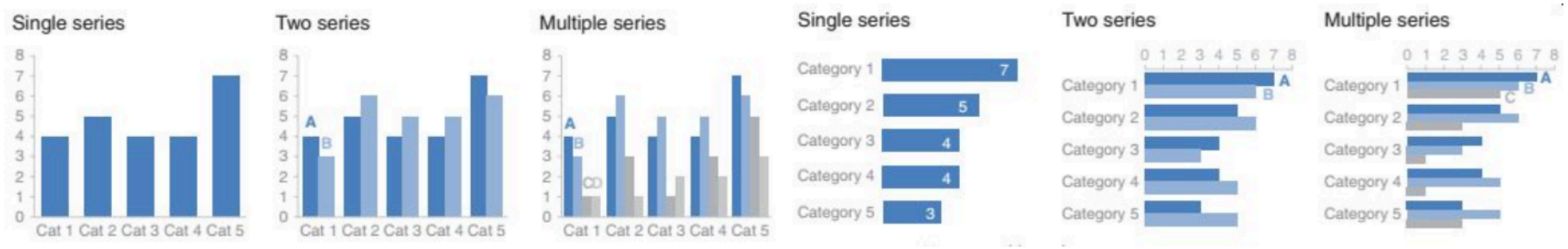
Slopegraph

Slopegraphs can be useful when you have two time periods or points of comparison and want to quickly show relative increases and decreases or differences across various categories between the two data points.



Bar

Bar charts are easy for our eyes to read. Our eyes compare the end points of the bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories. Note that, because of how our eyes compare the relative end points of the bars, it is important that bar charts always have a zero baseline



Area

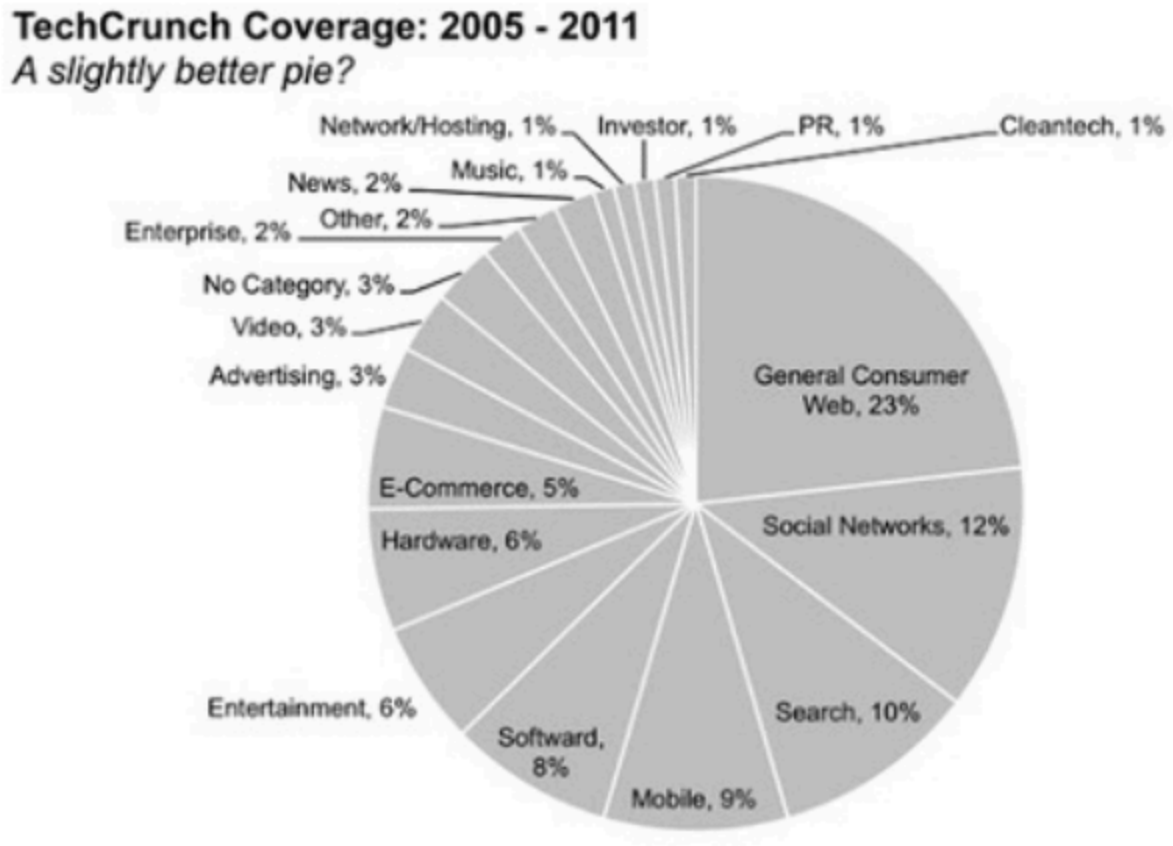
Area graphs are to avoided except when to visualize numbers of vastly different magnitudes. The second dimension you get using a square for this allows this to be done in a more compact way than possible with a single dimension.

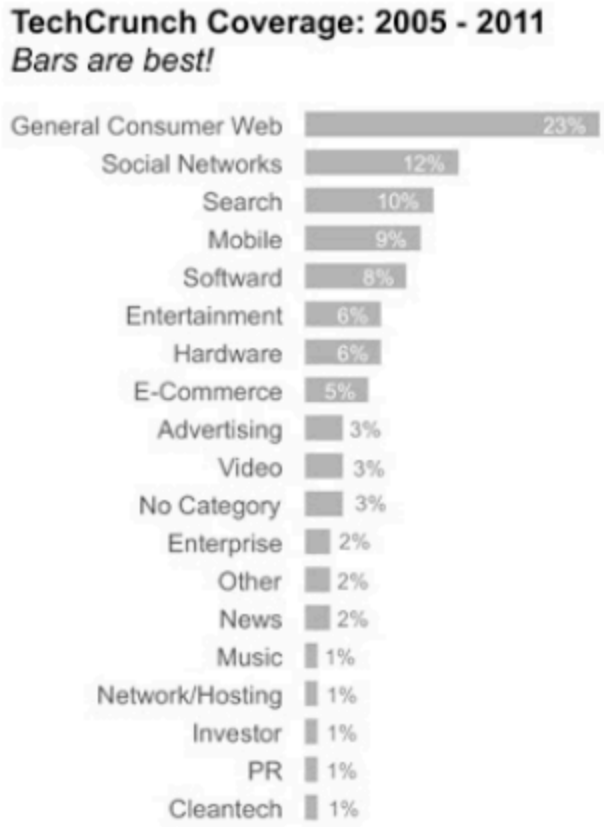
To be avoided

- Pie and donut charts
- 3D unless plotting in 3D
- Secondary Y-axis
- Chartjunk

Pie Charts

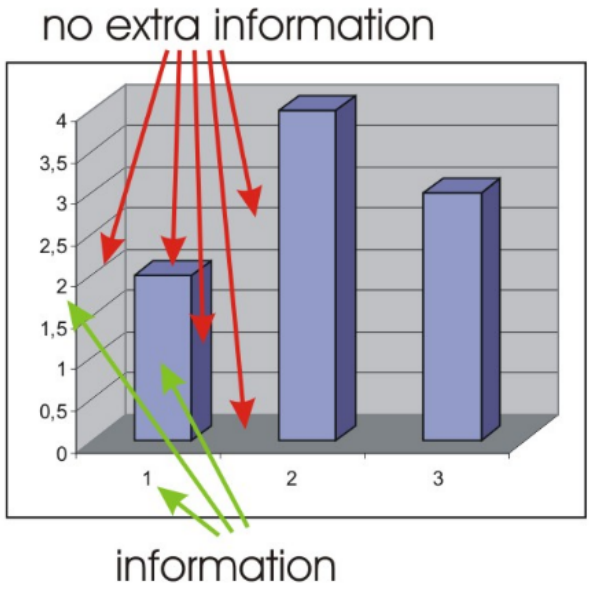
Pie charts are really hard for people to read!





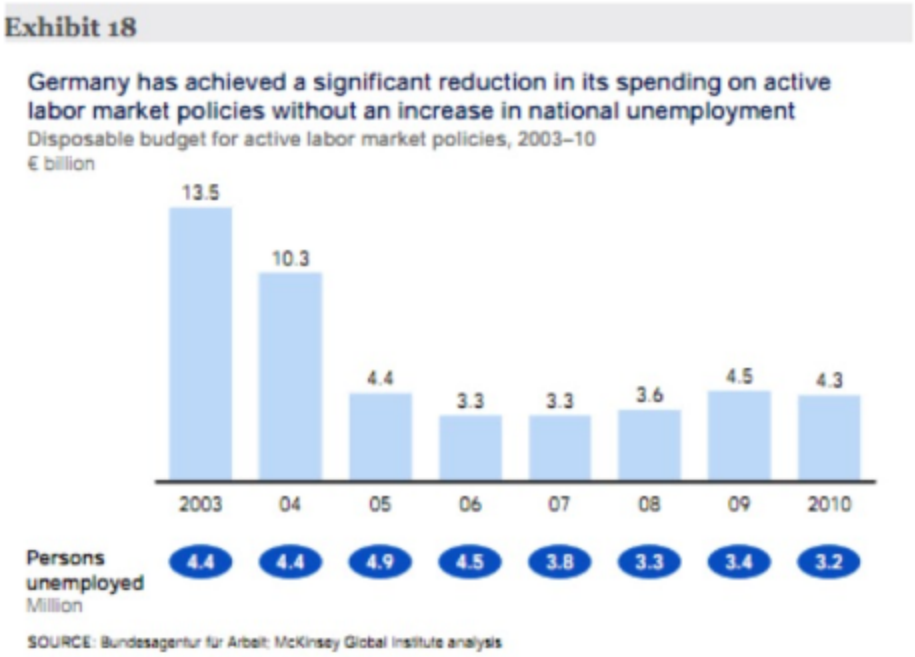
3D

Never use 3D to plot a single dimension. 3D skews our numbers, making them difficult or impossible to interpret or compare. Adding 3D to graphs introduces unnecessary chart elements



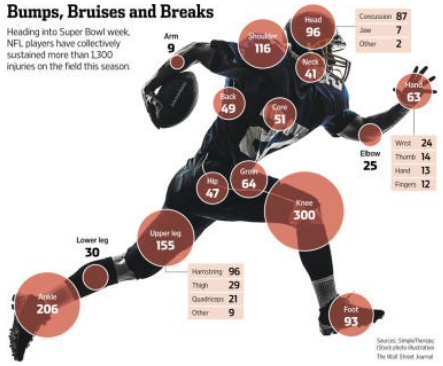
Secondary X or Y-axis

When interpreting the figure, it takes some time and reading to understand which data should be read against which axis.



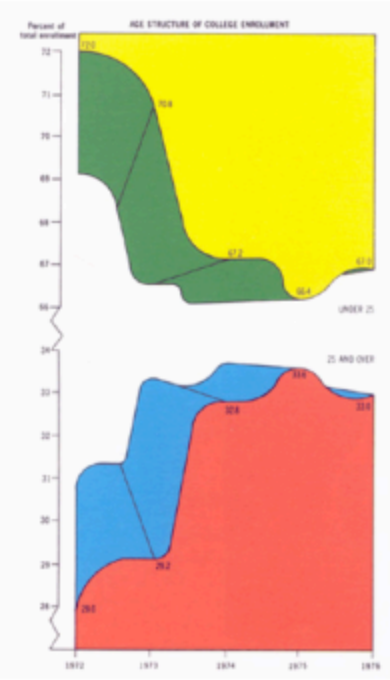
Chartjunk

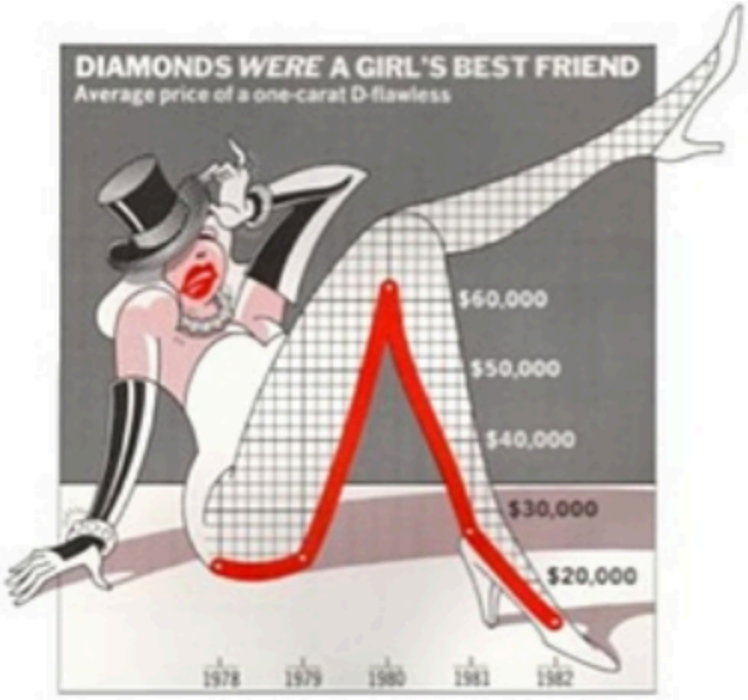
Chartjunk refers to all visual elements in charts and graphs that are not necessary to comprehend the information represented on the graph, or that distract the viewer from this information.



Chartjunk

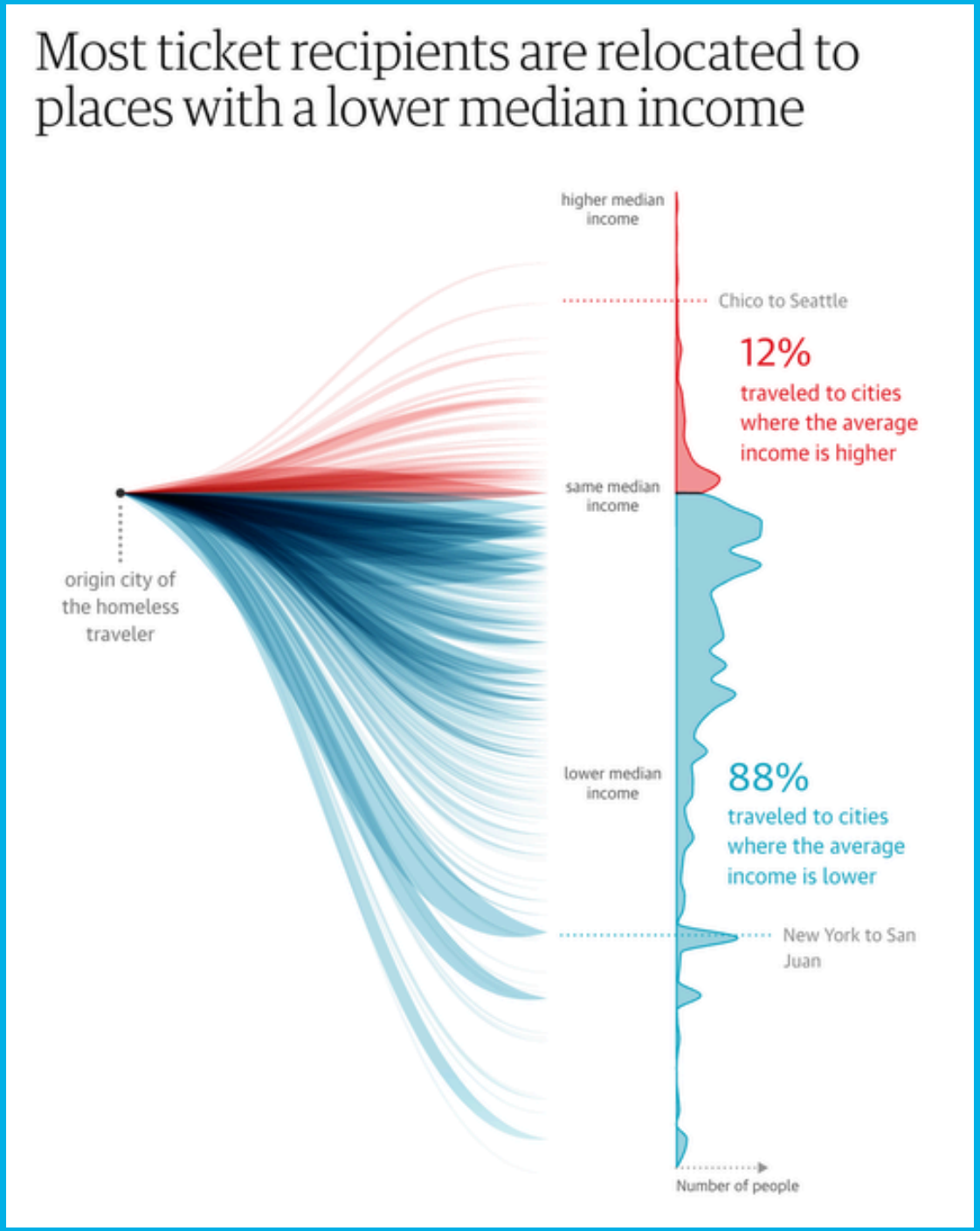
Examples of unnecessary elements that might be called chartjunk include heavy or dark grid lines, unnecessary text, inappropriately complex or gimmicky font faces, ornamented chart axes, and display frames, pictures, backgrounds or icons within data graphs, ornamental shading and unnecessary dimensions.





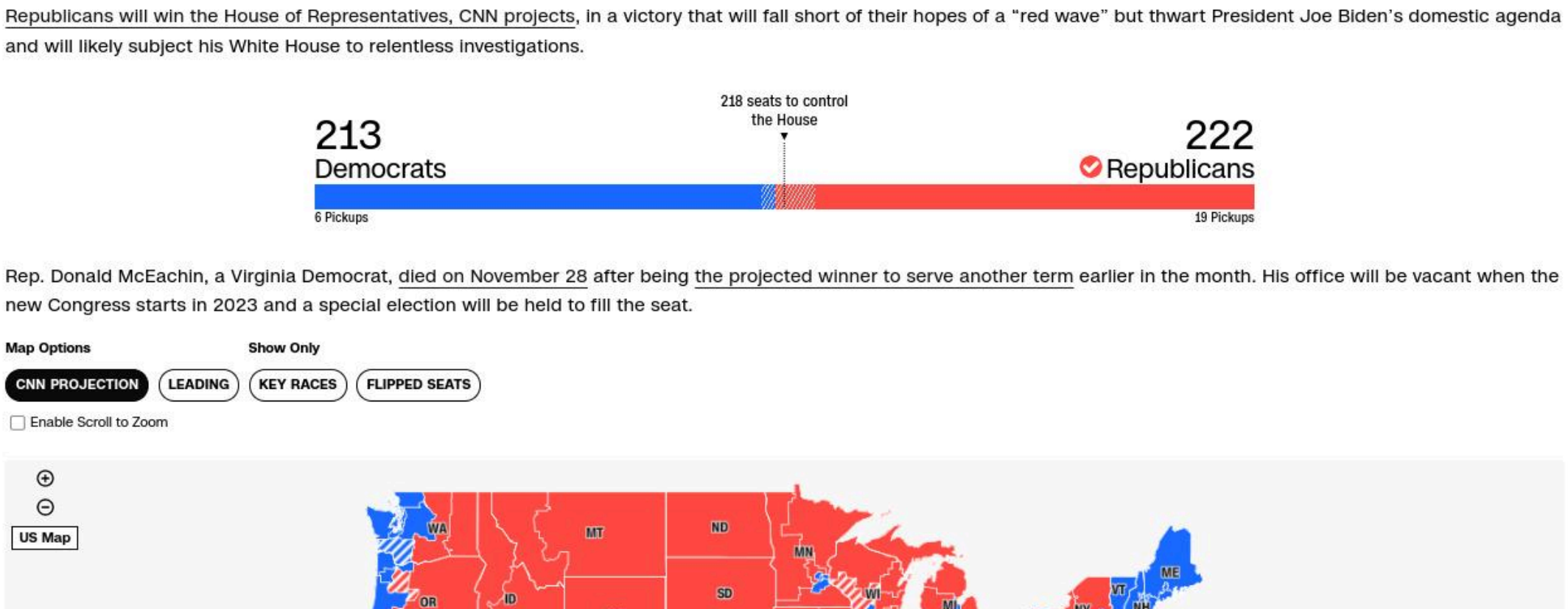
Good Graphs

Good graphs clearly show the important features of the data. They should always have: **title**, **labelled axes** - and a **key**



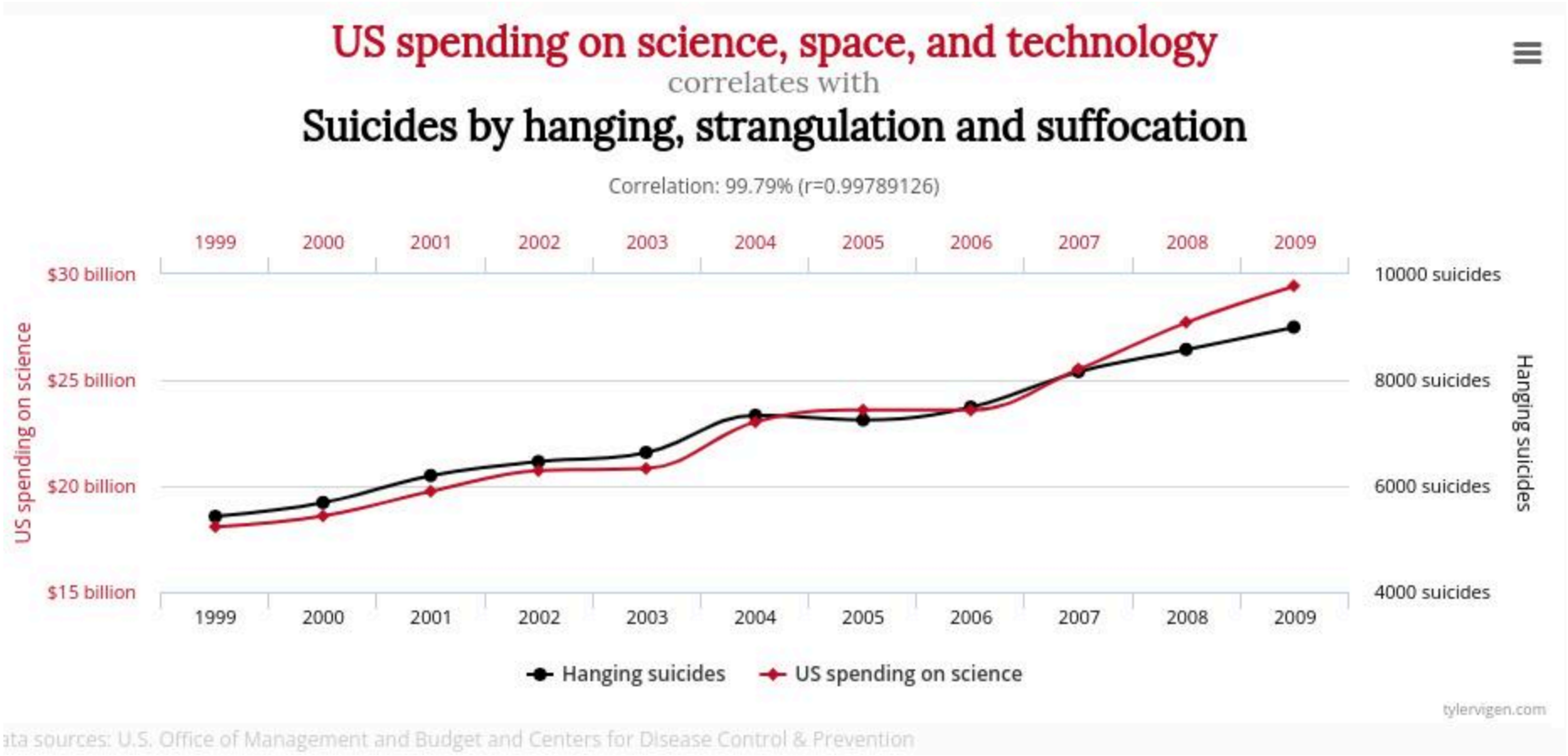
Good graphs

They should *tell a story* - and be **memorable** - but also have a low information to ink ratio - and **not mislead** - the viewer. Choice of colour when designing charts and graphs is also important to allow for colour blindness and black and white printing.



Common Mistakes

- Leaving gaps/changing the scale in vertical axes
- Uneven shading/colours
- Unfair emphasis on some sections
- Distorting areas in histograms (bar widths should always be equal)
- Use of 3-dimensions instead of two
- Misleading use of pictograms



Source: http://www.tylervigen.com/spurious-correlations

Books

- Edward R. Tufte. The Visual Display of Quantitative Information. Graphics 2001. https://www.goodreads.com/book/show/17744.The_Visual_Display_of_Quantitative_Information?ac=1from_search=true&qid=cMoP2LwLen&rank=1
- Cole Nussbaumer Knaflic. Storytelling with Data: A Data Visualization Guide for Business Professionals. Wiley 2015. https://www.goodreads.com/book/show/26535513-storytelling-with-data

Links

- The Few, The Proud: 11 Key Principles of Effective Data Visualization https://www.business2community.com/big-data/the-few-the-proud-11-key-principles-of-effective-data-visualization-02076890
- Data Visualization Hacks https://uxdesign.cc/data-visualization-hacks-75d56d5bfa66
- Data Visualization UX Best Practices (Updated 2024). https://www.designstudiouiux.com/blog/data-visualization-ux-best-practices/
- Line Chart Design Made Simple. https://uxdesign.cc/line-chart-design-made-simple-a1b823510674

Links

- Designing Charts: principles every designer should know (part 2) https://uxdesign.cc/designing-charts-principles-every-designer-should-know-part-2-ce1e06af56fc
- Building Blocks of Good Graphs. https://www.chaione.com/blog/building-blocks-graphs
- Data Visualization Design, Part 4: Removing Chart Junk. https://medium.com/@LauraHKahn/data-visualization-design-part-4-removing-chart-junk-28b3bdd0faa1
- COVID-19 In Charts: Examples of Good & Bad Data Visualization. https://analythical.com/blog/covid19-in-charts

Thank you!