



INTRO TO DATA VISUALIZATION

(FOR DSI - MARCH 2017)

susan.wolfe@generalassemb.ly



- Pie I have taken away
- Pie I have not yet taken away

- What is this data visualization stuff
- Some best practices
- Lots of examples along the way...

QUICK OVERVIEW

USING DATA

TASK



GOAL

Exploration

Searching for significant facts

Discovery

Sensemaking

Examining & making sense of data

Understanding

Communication

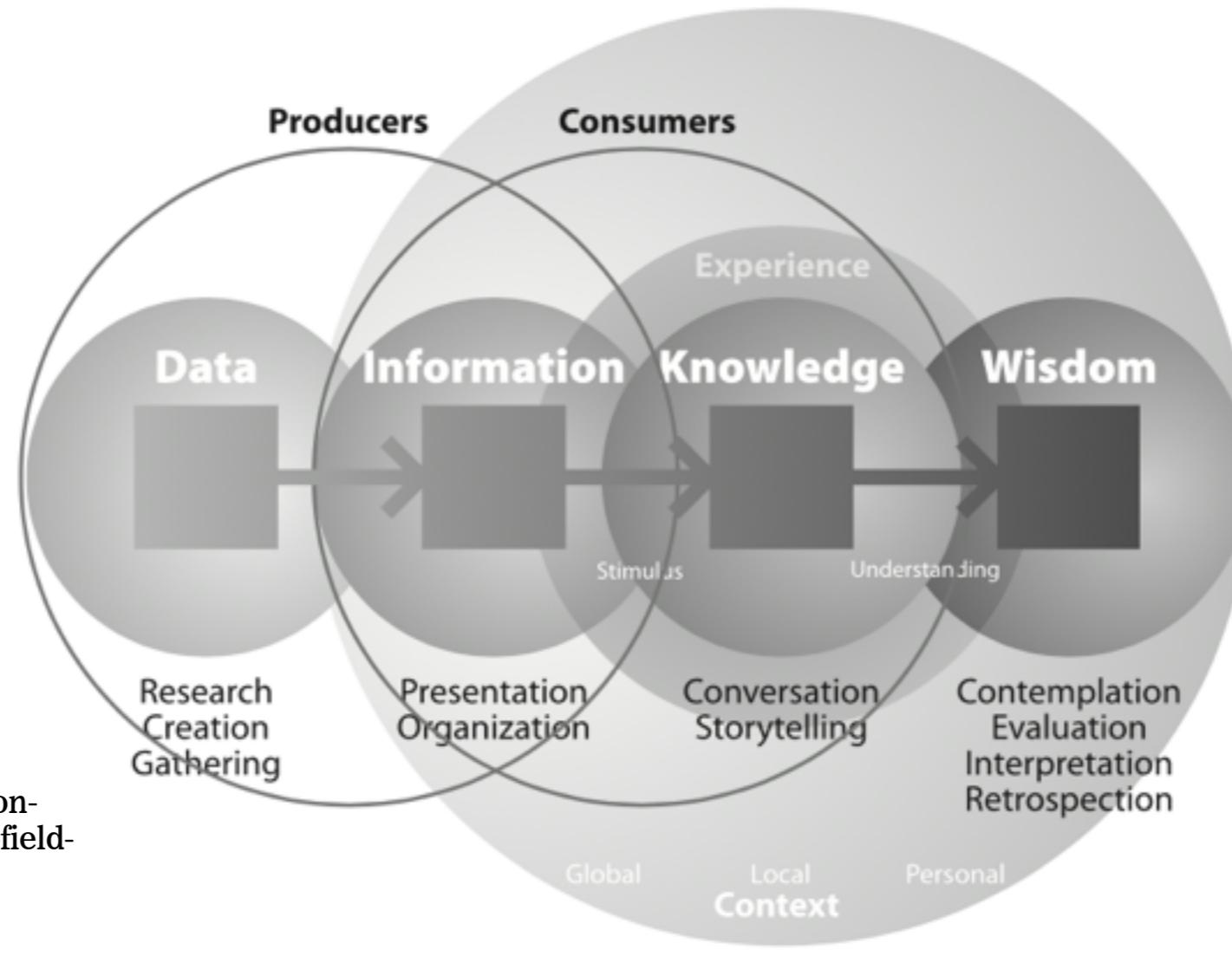
Conveying information to others

Informed decisions

- STEPHEN FEW

<https://www.perceptualedge.com/blog/?p=1897>

REASONS TO VISUALIZE RAW DATA



<http://nathan.com/information-interaction-design-a-unified-field-theory-of-design/>

- NATHAN SHEDROFF

**BUT INSIGHTS EMERGE,
ONLY IF YOU CAN FIND THEM.**

REVEAL INSIGHTS

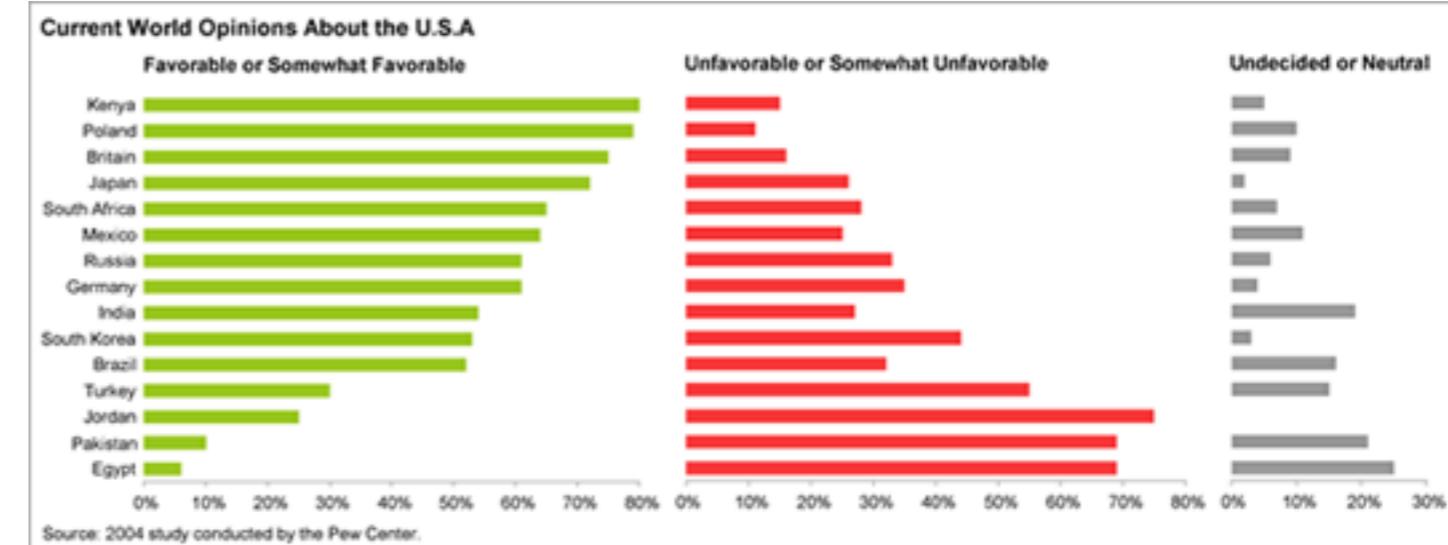
John Snow (15 March 1813 – 16 June 1858) was an English physician and a leader in the adoption of anesthesia and medical hygiene. He is considered one of the fathers of modern epidemiology, in part because of his work in tracing the source of a cholera outbreak in Soho, London, in 1854. His findings inspired fundamental changes in the water and waste systems of London, which led to similar changes in other cities, and a significant improvement in general public health around the world.



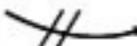
- JOHN SNOW (1854)

VISUALIZATION CAN LEAD TO CLARITY

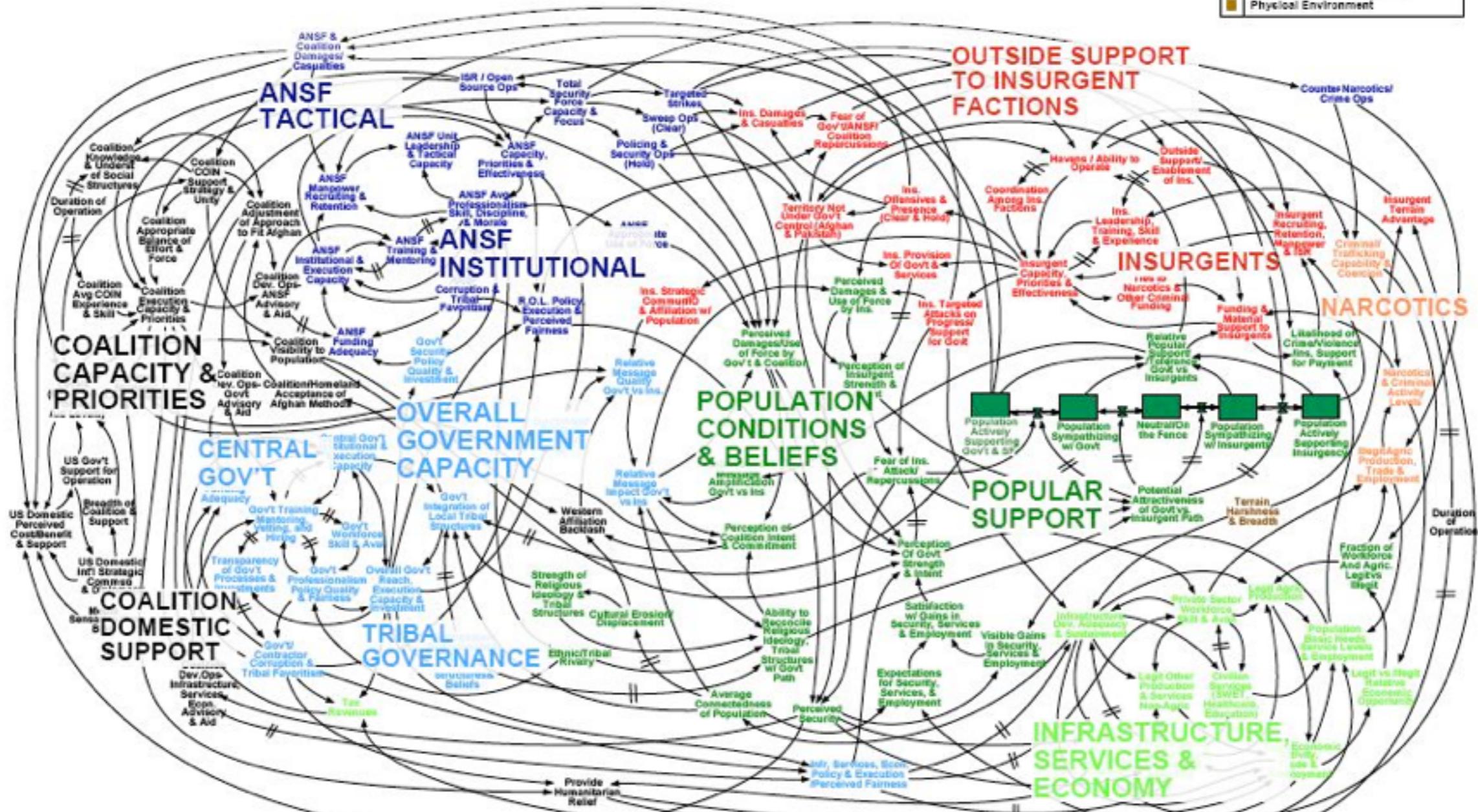
Favorable or Unfavorable View of the U.S.	
Brazil: % with somewhat or very favorable opinion of the U.S.:	52%
Brazil: % with somewhat or very unfavorable opinion of the U.S.:	32%
Mexico: % with somewhat or very favorable opinion of the U.S.:	64%
Mexico: % with somewhat or very unfavorable opinion of the U.S.:	25%
Britain: % with somewhat or very favorable opinion of the U.S.:	75%
Britain: % with somewhat or very unfavorable opinion of the U.S.:	16%
Germany: % with somewhat or very favorable opinion of the U.S.:	61%
Germany: % with somewhat or very unfavorable opinion of the U.S.:	35%
Russia: % with somewhat or very favorable opinion of the U.S.:	61%
Russia: % with somewhat or very unfavorable opinion of the U.S.:	33%
Poland: % with somewhat or very favorable opinion of the U.S.:	79 %
Poland: % with somewhat or very unfavorable opinion of the U.S.:	11%
South Africa: % with somewhat or very favorable opinion of the U.S.:	65%
South Africa: % with somewhat or very unfavorable opinion of the U.S.:	28%
Kenya: % with somewhat or very favorable opinion of the U.S.:	80%
Kenya: % with somewhat or very unfavorable opinion of the U.S.:	15%
India: % with somewhat or very favorable opinion of the U.S.:	54%
India: % with somewhat or very unfavorable opinion of the U.S.:	27%
Japan: % with somewhat or very favorable opinion of the U.S.:	72%
Japan: % with somewhat or very unfavorable opinion of the U.S.:	26%
South Korea: % with somewhat or very favorable opinion of the U.S.:	53%
South Korea: % with somewhat or very unfavorable opinion of the U.S.:	44%
Egypt: % with somewhat or very favorable opinion of the U.S.:	6%
Egypt: % with somewhat or very unfavorable opinion of the U.S.:	69%
Pakistan: % with somewhat or very favorable opinion of the U.S.:	10%
Pakistan: % with somewhat or very unfavorable opinion of the U.S.:	69%
Turkey: % with somewhat or very favorable opinion of the U.S.:	30%
Turkey: % with somewhat or very unfavorable opinion of the U.S.:	55%
Jordan: % with somewhat or very favorable opinion of the U.S.:	25%
Jordan: % with somewhat or very unfavorable opinion of the U.S.:	75%

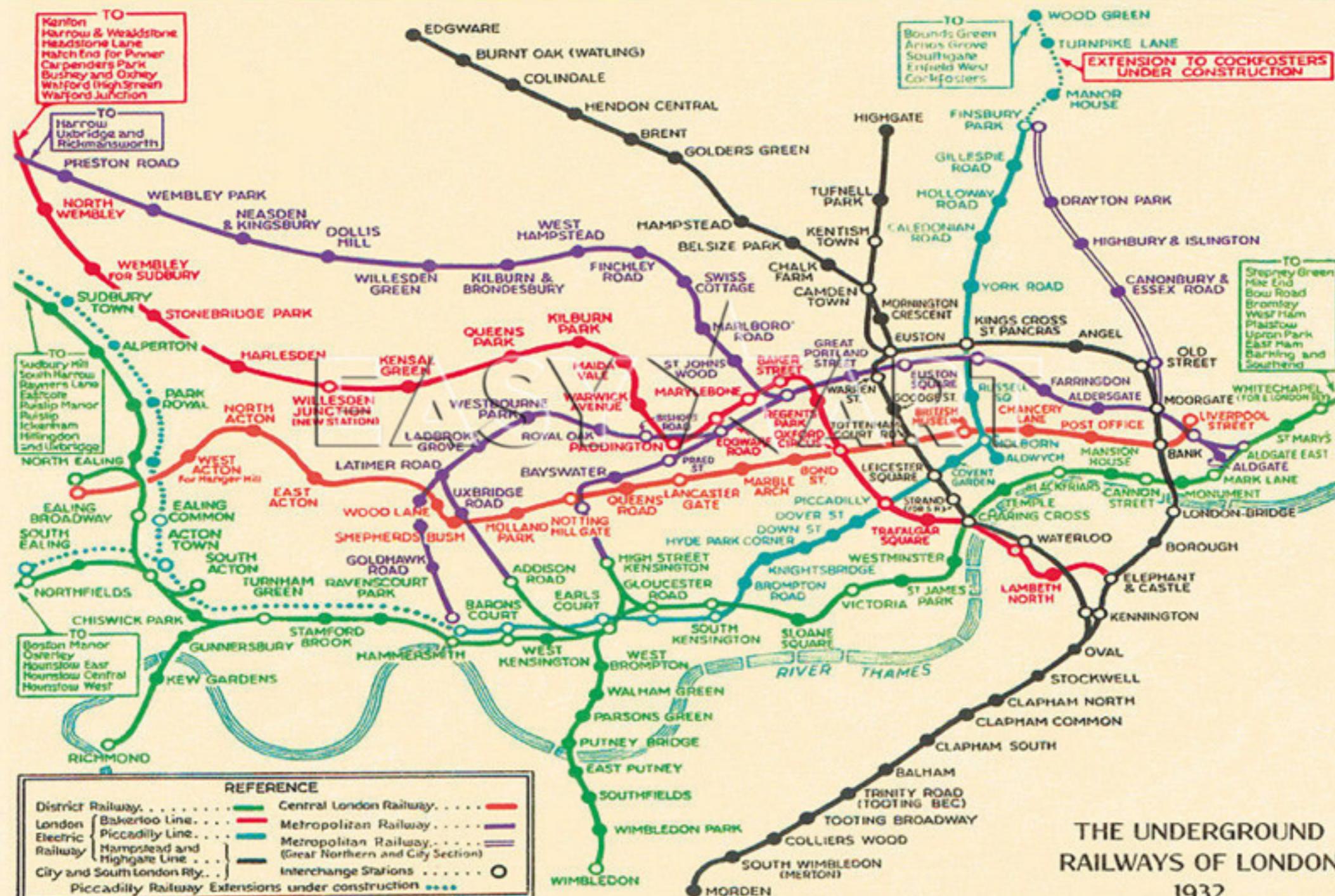


Afghanistan Stability / COIN Dynamics

 = Significant Delay

 Population/Popular Support
 Infrastructure, Economy, & Services
 Government
 Afghanistian Security Forces
 Insurgents
 Crime and Narcotics
 Coalition Forces & Actions
 Physical Environment





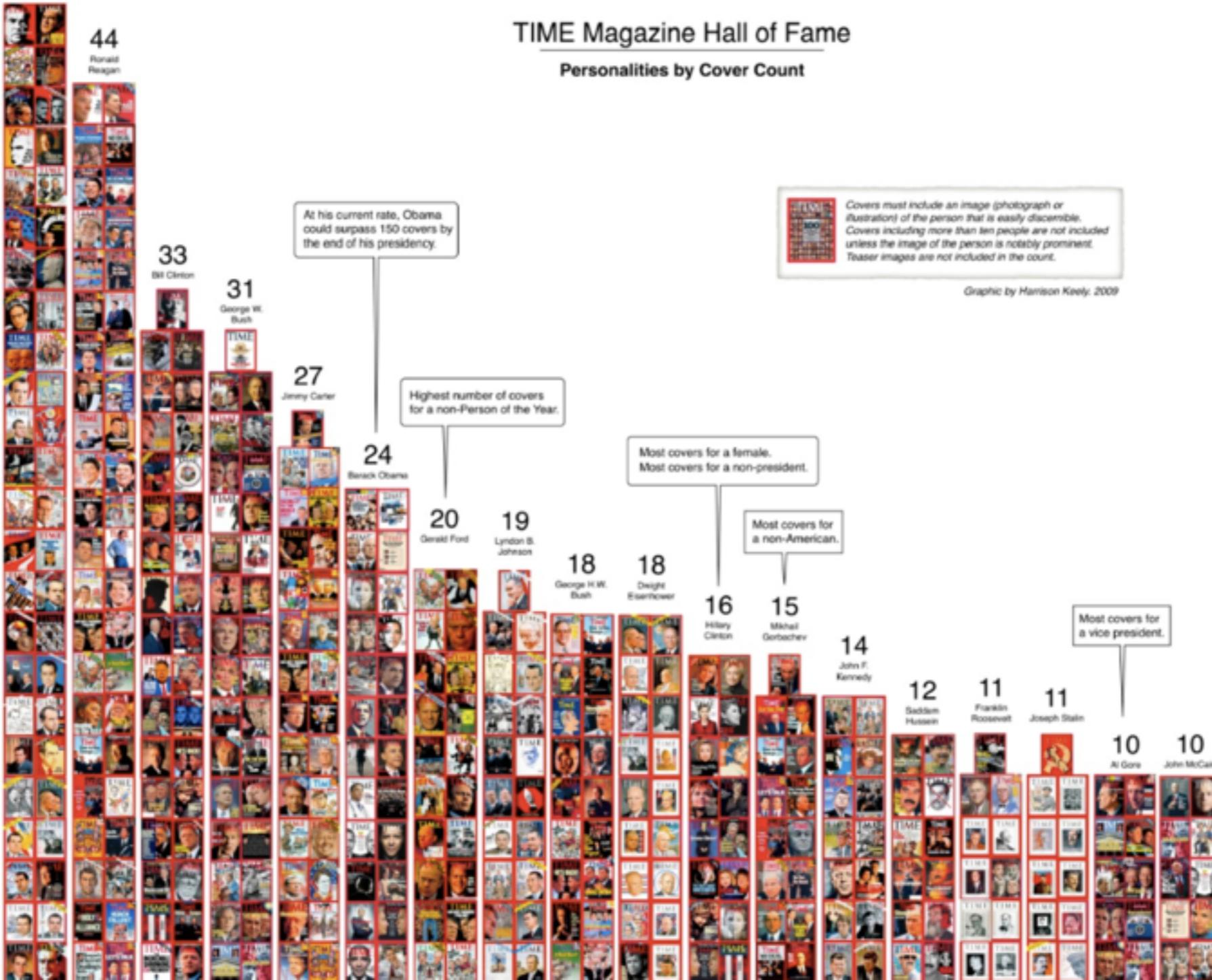
THE UNDERGROUND
RAILWAYS OF LONDON
1932



REASONS TO TURN RAW DATA INTO VISUALIZATION

- See the big picture
- Rapidly compare values
- See patterns
- Compare patterns

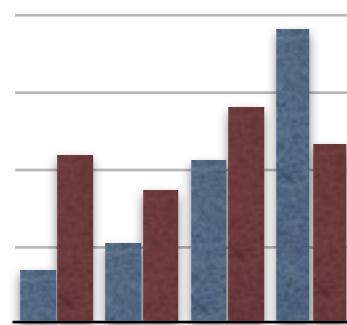
- STEPHEN FEW (2004)



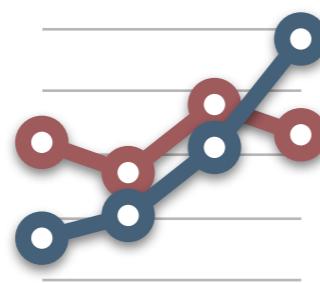


TYPICAL REPRESENTATIONS OF DATA

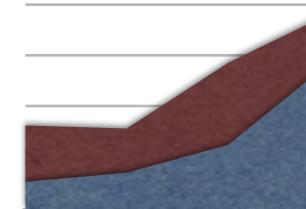
bar



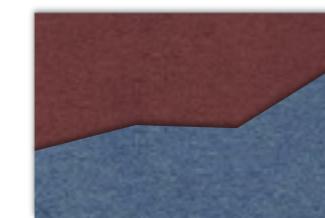
line



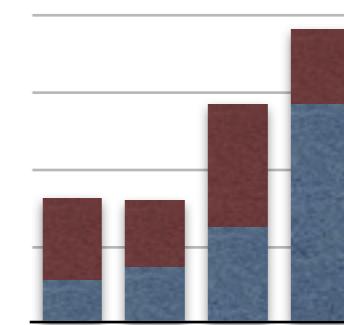
stacked
area



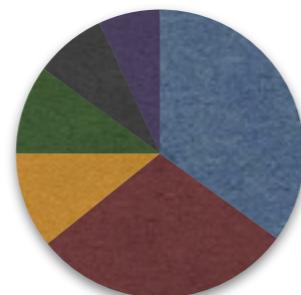
scaled
stacked
area



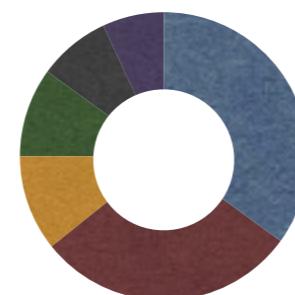
stacked
bar



pie



donut



scatter

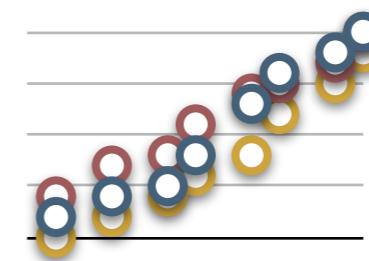


Chart Suggestions—A Thought-Starter

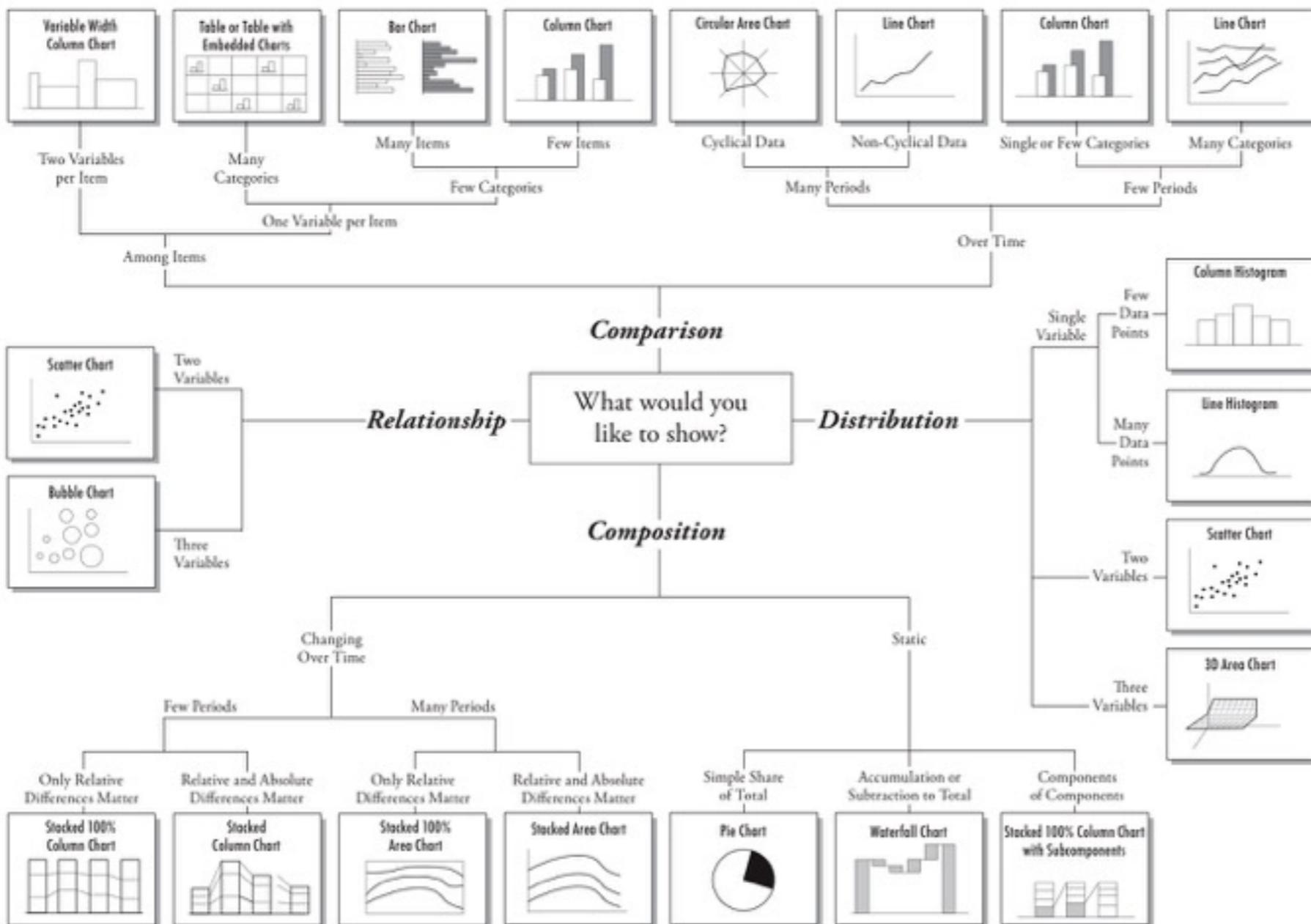
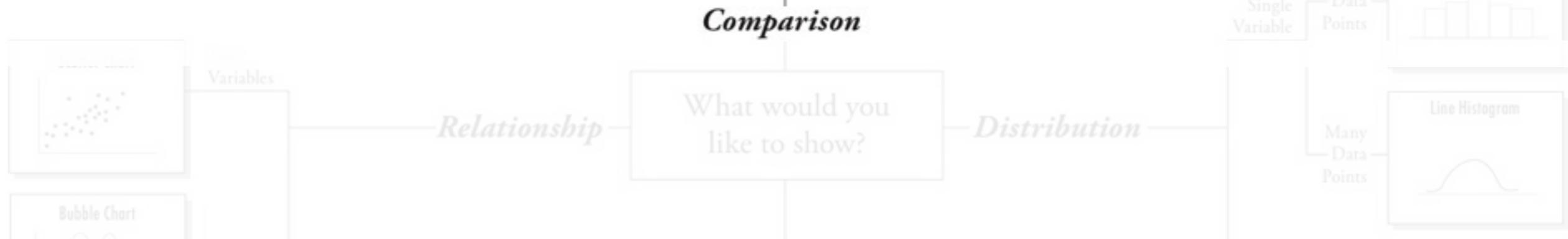
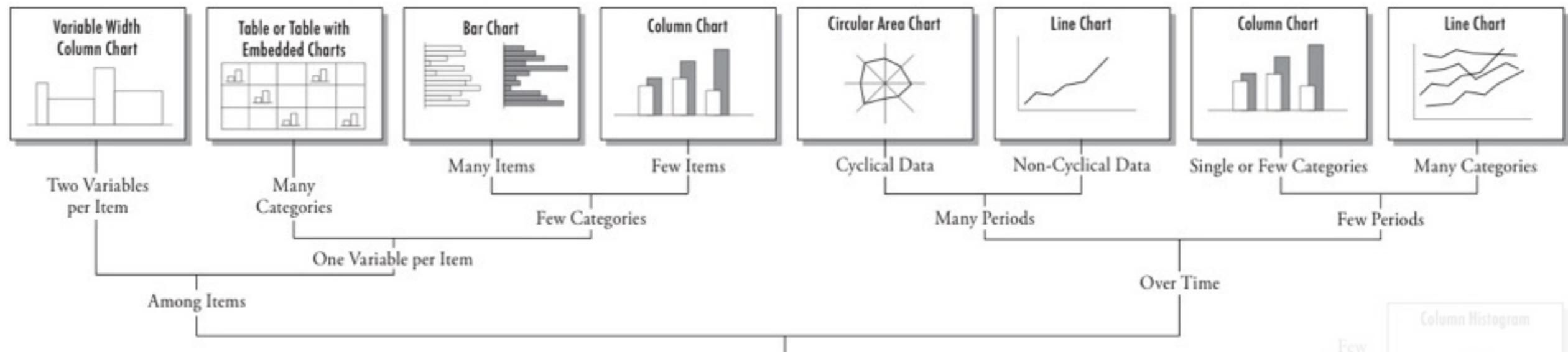
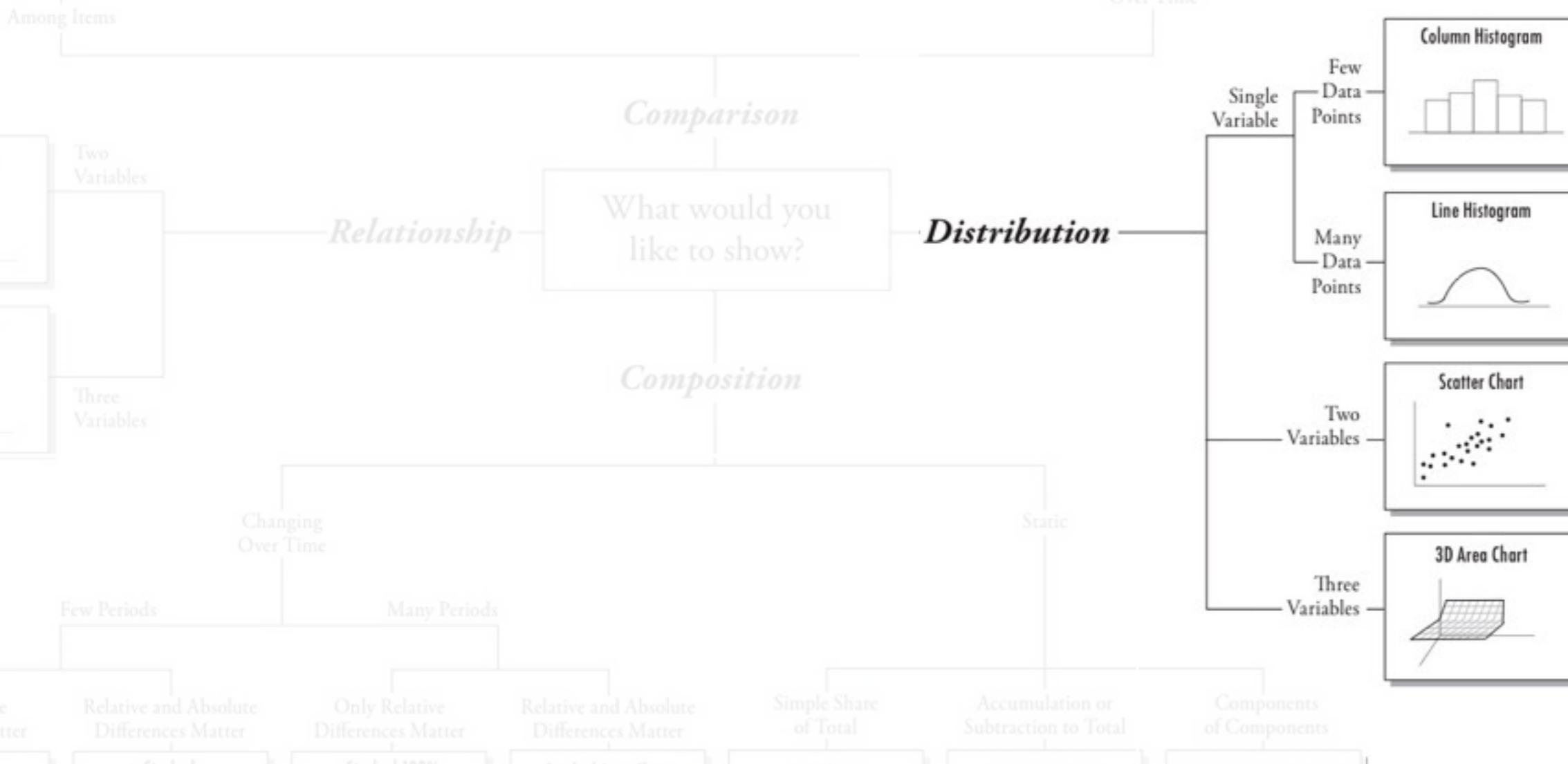
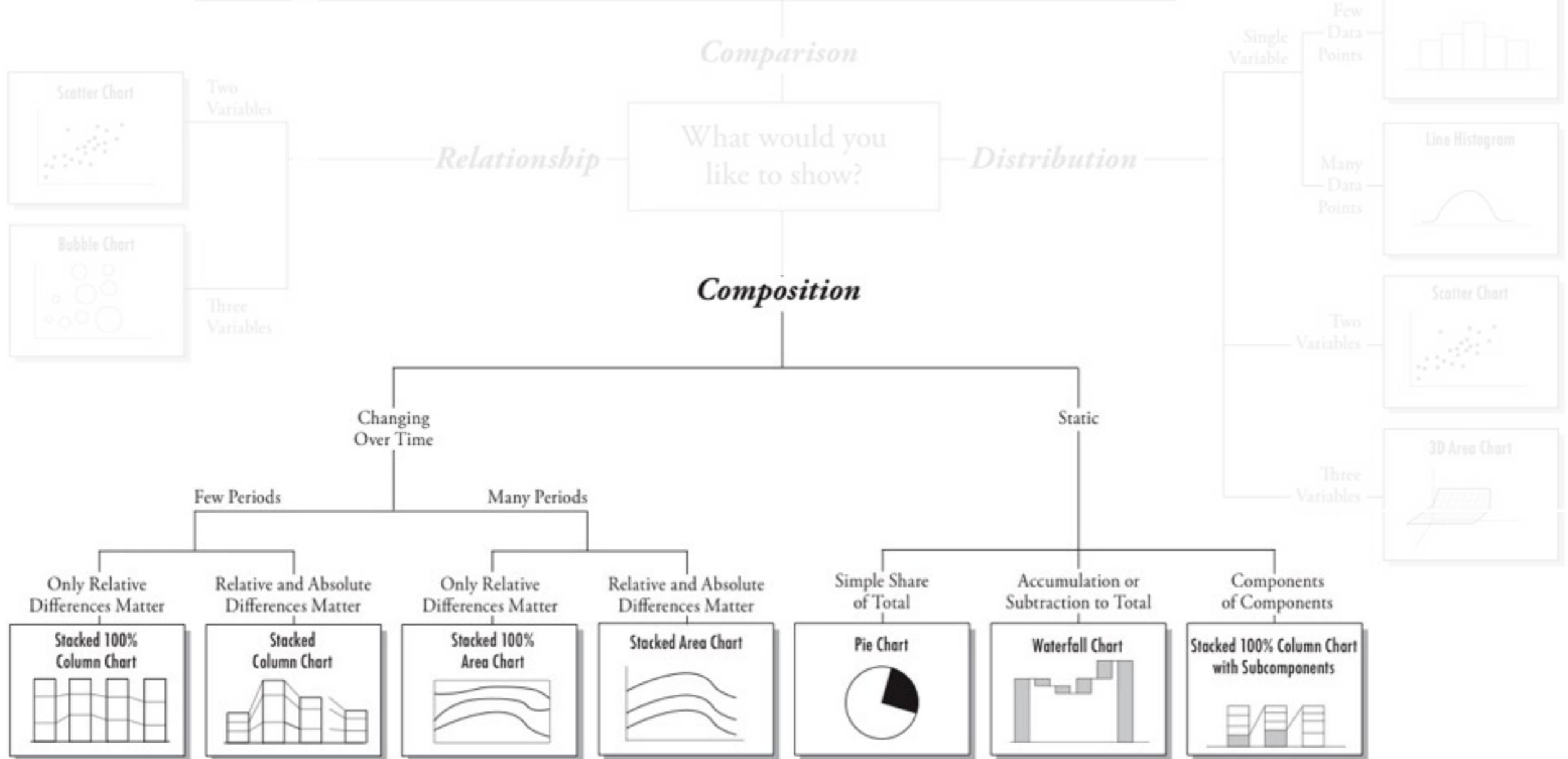


Chart Suggestions—A Thought-Starter







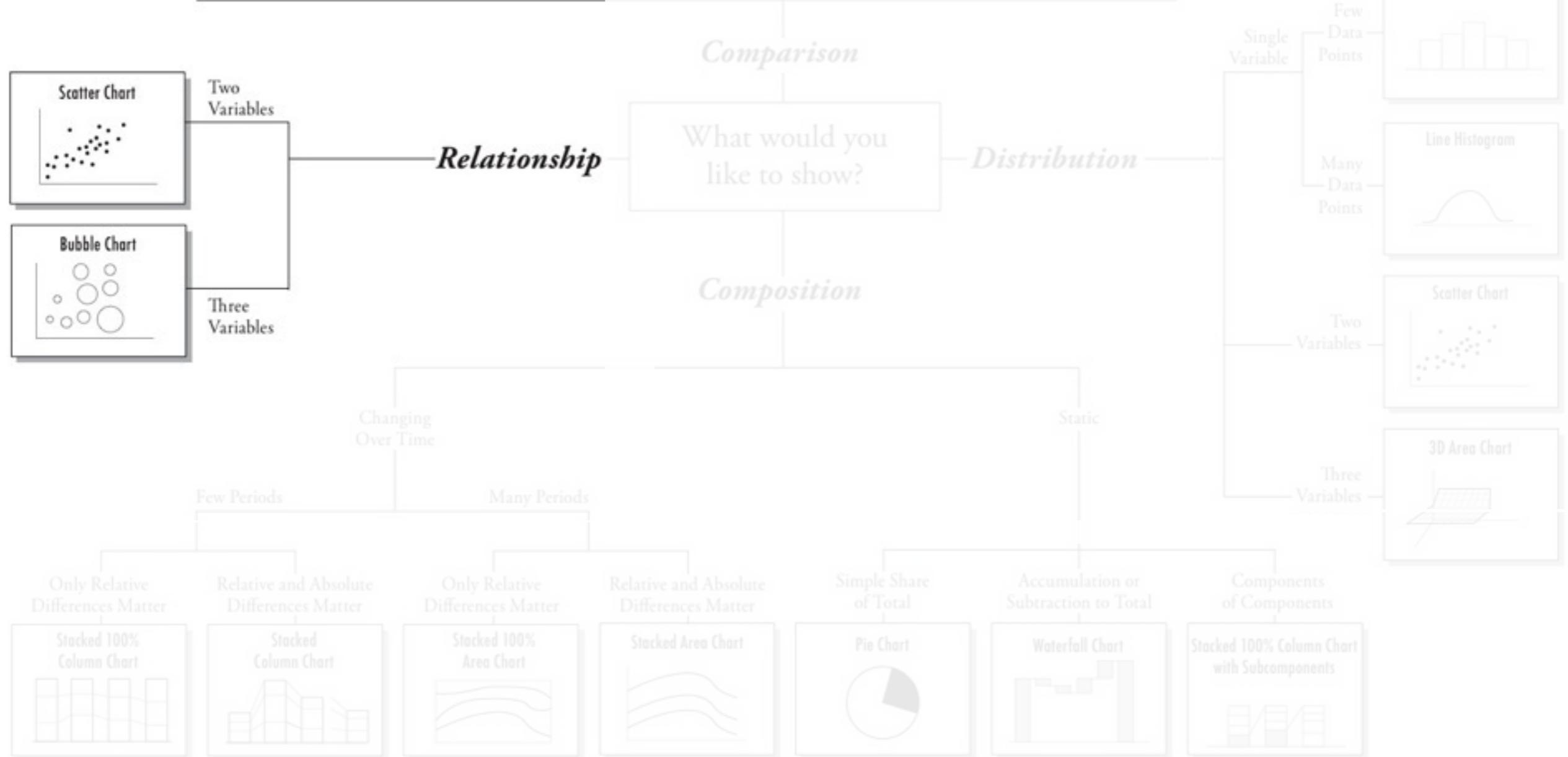
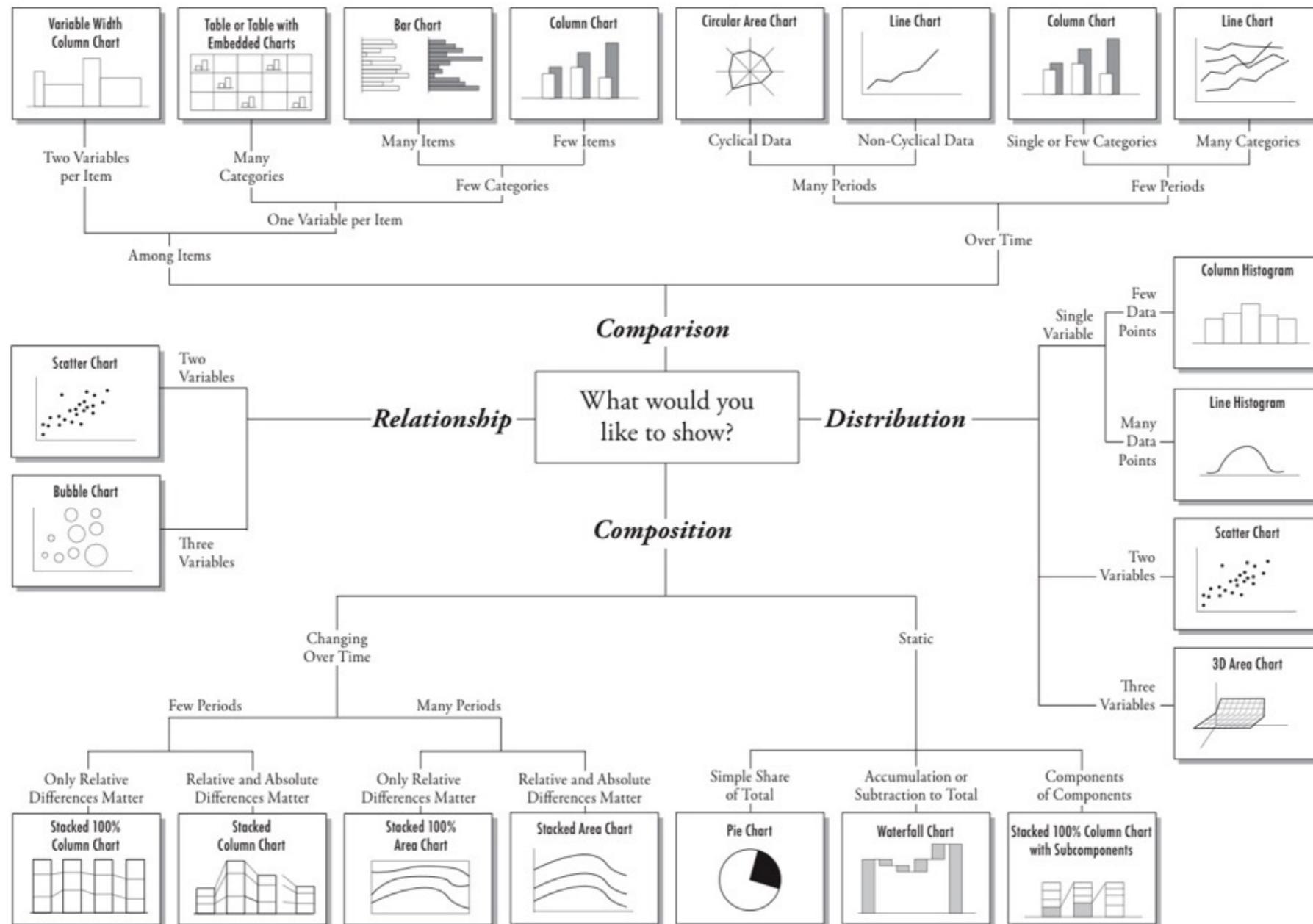
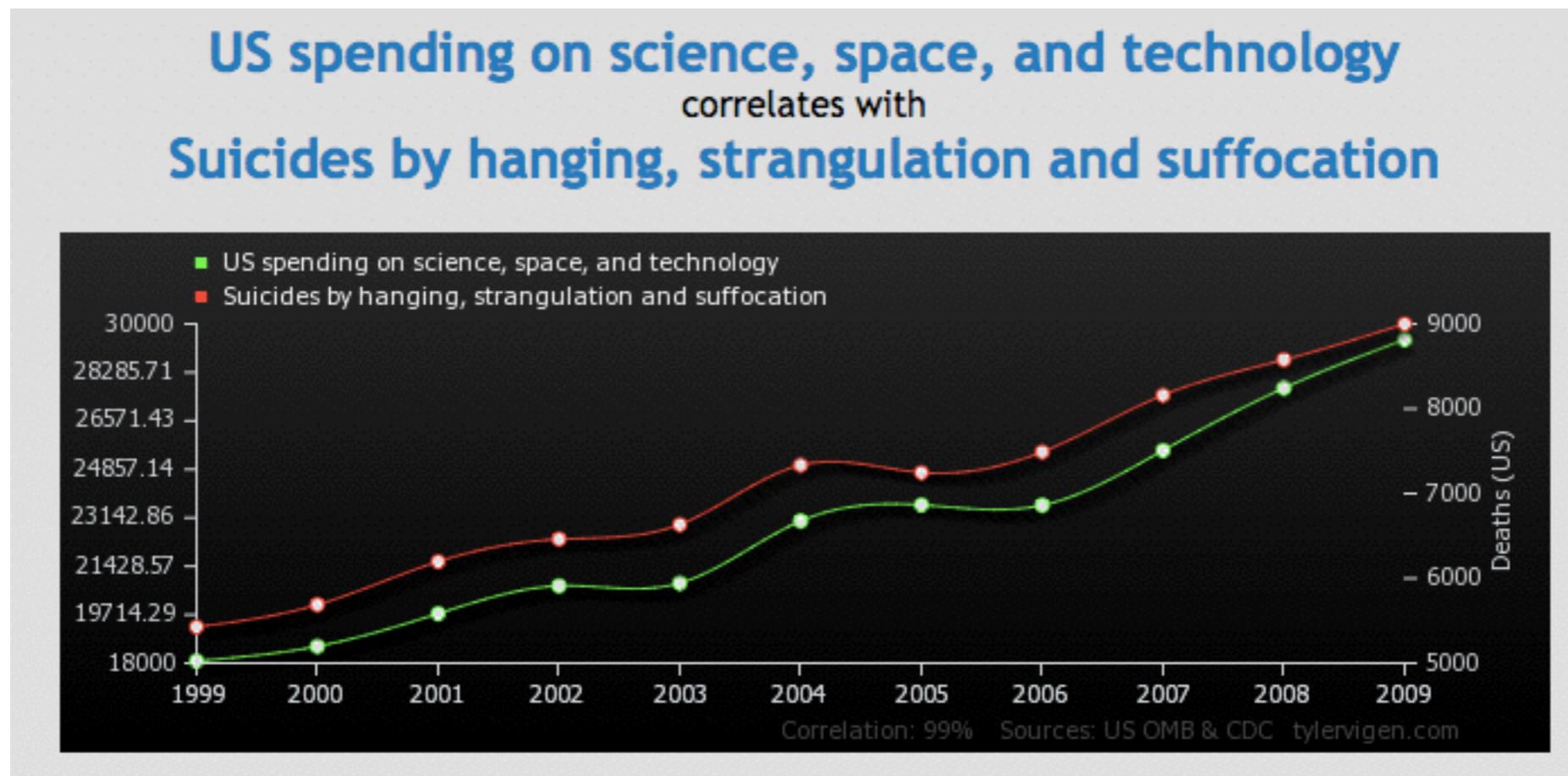


Chart Suggestions—A Thought-Starter



SPURIOUS CORRELATIONS



CHOOSE THE BEST-FIT VISUALIZATION

A PERIODIC TABLE OF VISUALIZATION METHODS

C continuum	G graphic facilitation
Tb table	S Strategy Visualization The systematic use of complementary visual representations in the analysis, development, stimulation, communication, and implementation of strategies in organizations.
Ga garden continuum	I Information Visualization The use of innovative visual representations of data to support cognition. This means that the data is transformed into an image; it is mapped in screen space. The image can be changed by users as they proceed working with it.
Pi pic chart	M Metaphor Visualization Visualizations position information graphically in sequence and structure information. They also connect or link ideas through the representation of information through the key characteristics of the metaphor that is employed.
L line chart	C Compound Visualization The combination of different graphic representation formats in one single scheme or frame.
B bar chart	Hm metro map
Ac area chart	Tm map
R ratio chart	St story template
Pa parallel coordinate	Tr tree
Hy hypercube	Ct cartoon
Ey cycle diagram	
T timeline	
We web diagram	
Mi mapping	
Sq space representation	
Cc category circle	
Ar argumentation	
Sw swim lane diagram	
Gc part chart	
Pm projection diagram	
D diagram	
Pr process flow	
Kn knowledge map	
Hi hierarchy	
Sc scatterplot	
Sa sunburst diagram	
In information lens	
E entity relationship diagram	
Pt petri net	
Fl flow chart	
Cl clustering	
Lc layer chart	
Py meta diagram	
Ge conceptual chart	
Tl thinking map	
Dt data tree	
Cp open central path method	
Cf concept flow	
Co concept map	
Ic icong	
Lm learning map	
Th tally bar plot	
Sp spectrogram	
Da data map	
Tp treemap	
Cn cone tree	
Sy synapse type I	
Df data flow diagram	
Se semantic network	
Sn sensemaking map	
Fo face field diagram	
Io information map	
Pr prior chart	
Ev evolving knowledge map	
V vizi diagram	
Hh hexagon 'I' left chart	
I infogram	
Gy Process Visualization	
Hy Structure Visualization	
○ Overview	Su supply/demand curve
□ Detail	Pe performance charting
○ Detail AND Overview	St strength map
< > Divergent thinking	Oc organization chart
> < Convergent thinking	Ho house of quality
	Fd feedback diagram
	Ft future tree
	Mq magic quadrant
	Ld life-cycle diagram
	Po portfolio for focus
	S s-cycle
	Sm stakeholder map
	Is isolate diagram
	Tc technology road map
	Ed elbow chart
	Pf profile diagram
	Sg strengths/growth
	Mz matrix/organigram
	Z matrix/anthropological tree
	Ad affinity diagram
	Be bounce diagram
	Bm big matrix
	Stc value chain
	Hy hypercycle
	Sr sustainable rating map
	Ta tag
	Sd sway diagram

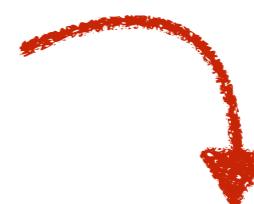
Note: Depending on your location and connection speed it can take some time to load a pop-up picture.
 © Ralph Lengler & Martin J. Eppler, www.visual-literacy.org

version 1.5

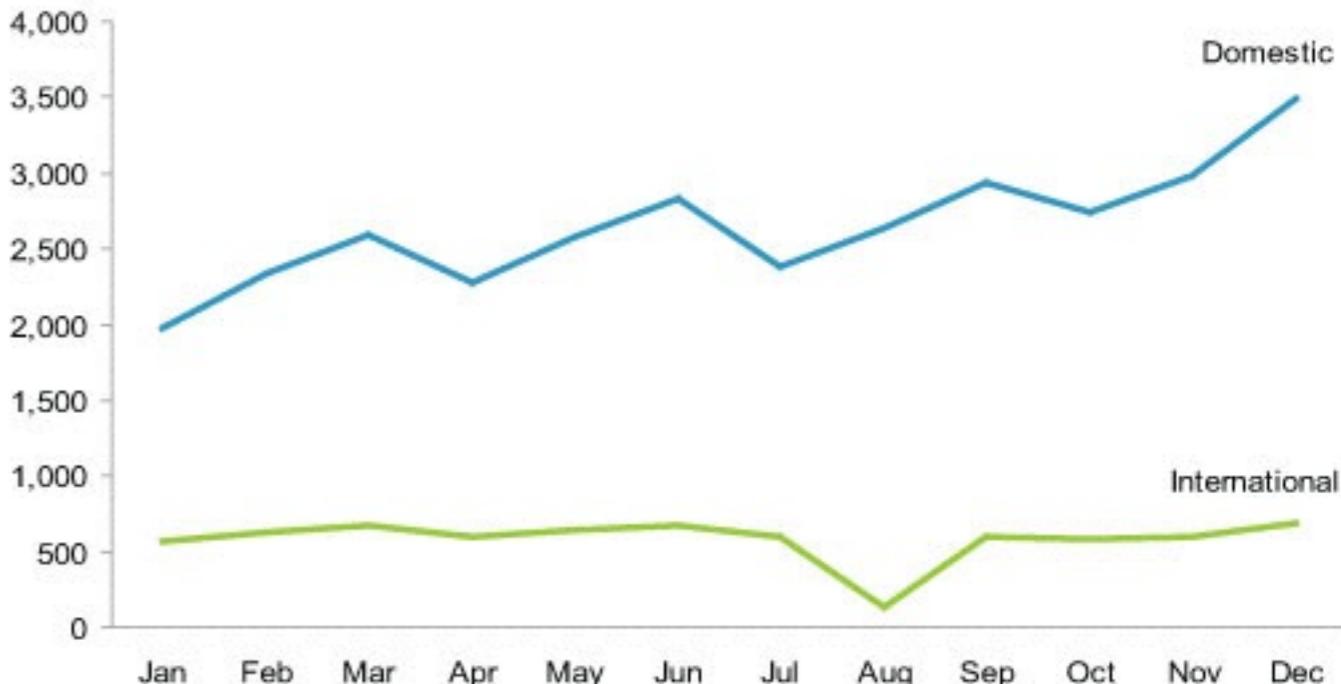
UNDERSTANDING A DATA SET...

Sales

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493
International	574	636	673	593	644	679	593	139	599	583	602	690
	\$2,557	\$2,979	\$3,266	\$2,876	\$3,218	\$3,517	\$2,975	\$2,773	\$3,537	\$3,322	\$3,585	\$4,183



U.S. Dollars
(thousands)



HOW BIG IS AFRICA?

REGION

Africa

United States

Europe

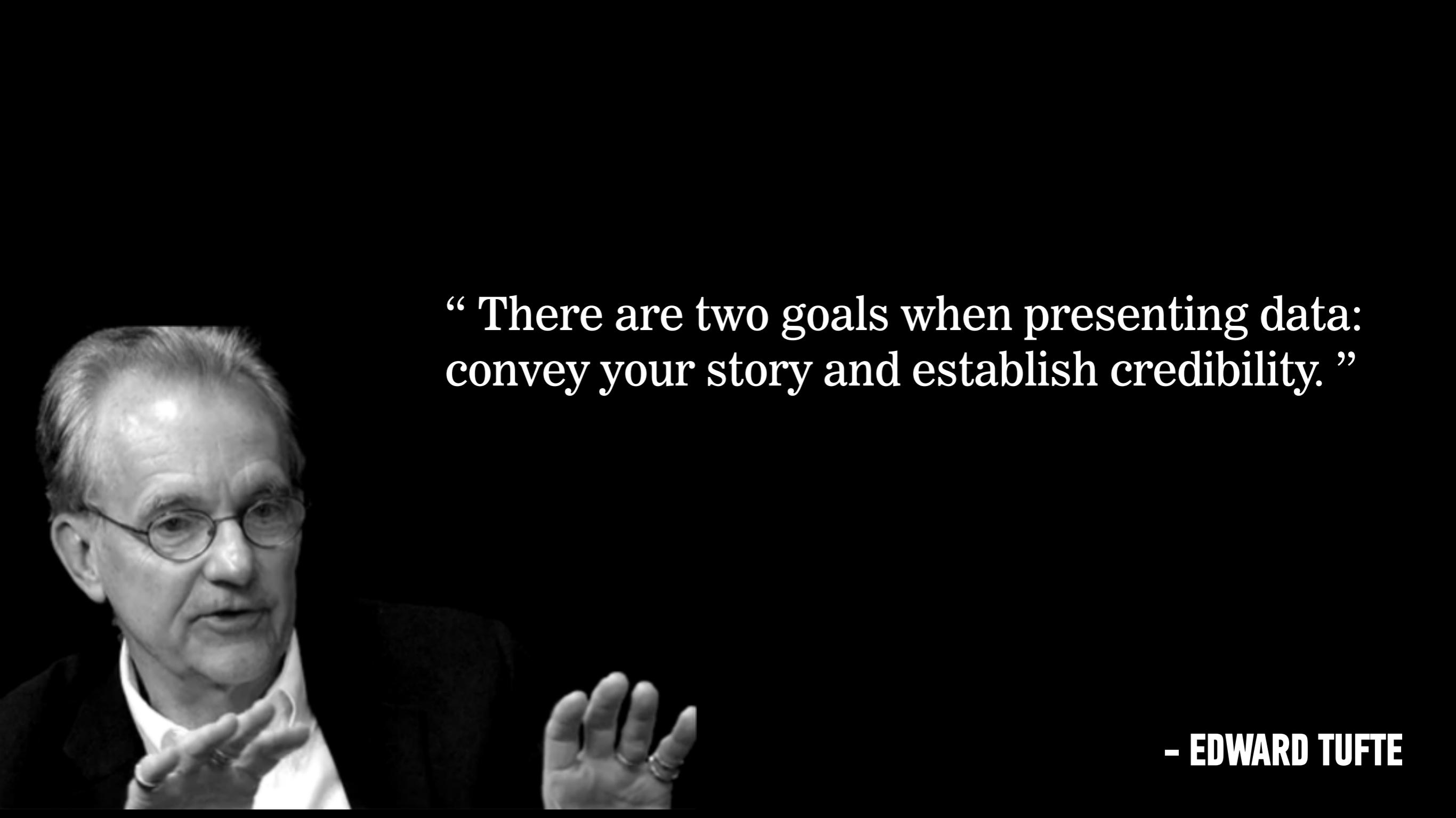
SIZE

11.72 million sq mi

3.7 million sq mi

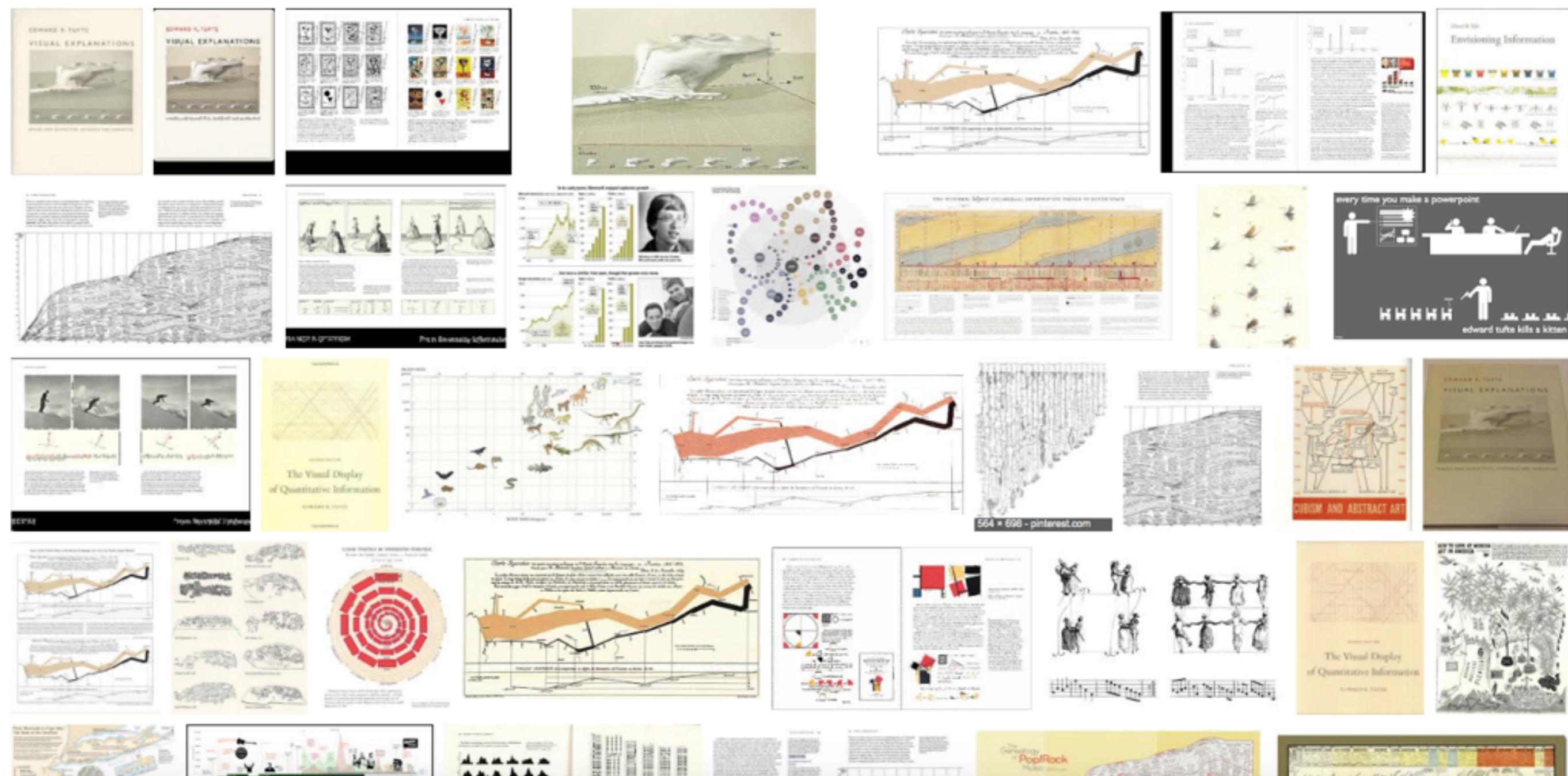
2.3 million sq mi

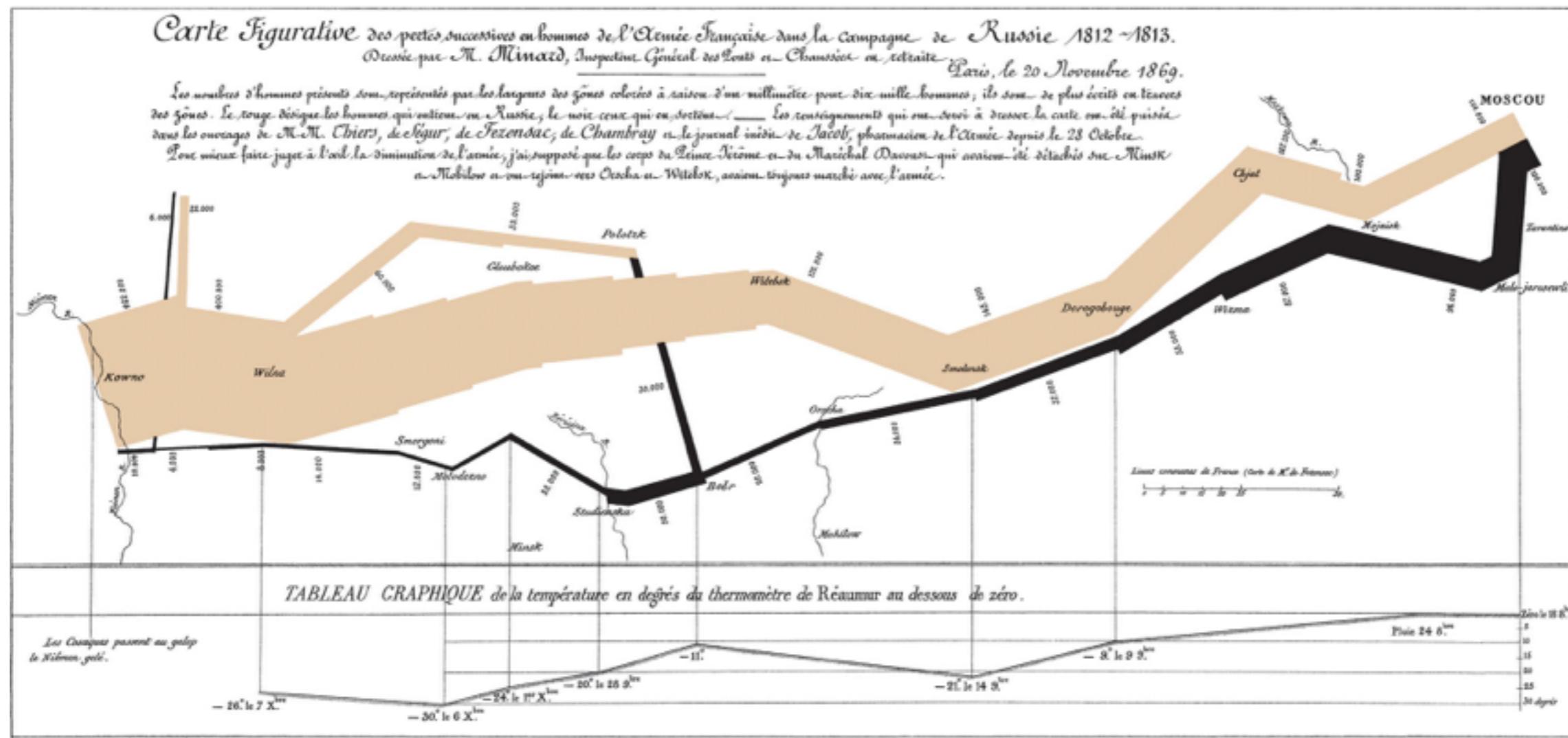


A black and white photograph of Edward Tufte, a middle-aged man with glasses and short hair, wearing a dark suit jacket over a white shirt. He is gesturing with his hands while speaking. The background is dark.

“There are two goals when presenting data:
convey your story and establish credibility.”

- EDWARD TUFTÉ



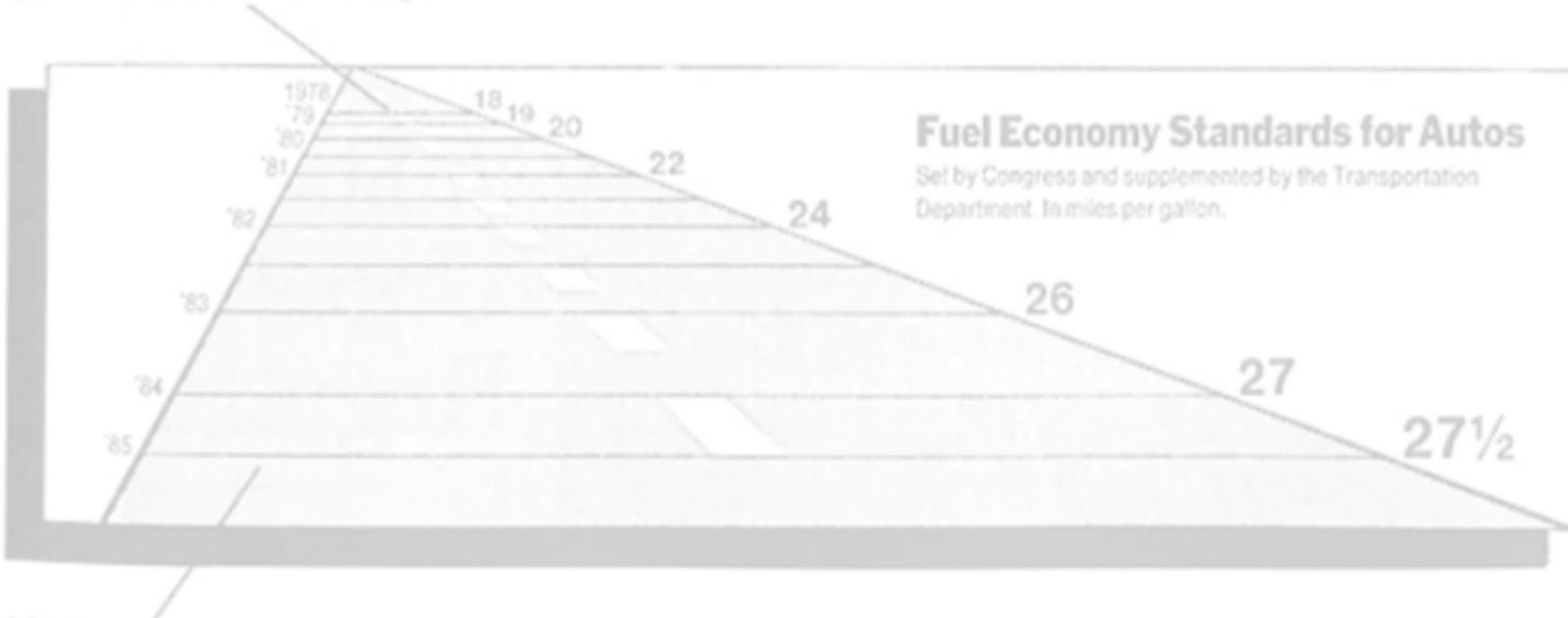


TUFTÉ'S PRINCIPLES OF GRAPHICAL EXCELLENCE

- Show the data
- Induce the viewer to think about the substance of the findings rather than the methodology, the graphical design, or other aspects
- Avoid distorting what the data have to say
- Present many numbers in a small space, i.e., efficiently
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal the data at several levels of detail, from a broad overview to the fine structure
- Serve a clear purpose: description, exploration, tabulation, or decoration
- Be closely integrated with the statistical and verbal descriptions of the data set

From E. R. Tufte. *The Visual Display of Quantitative Information*, 2nd Edition. Graphics Press, Cheshire, Connecticut, 2001.

This line, representing 18 miles per gallon in 1978, is 0.6 inches long.

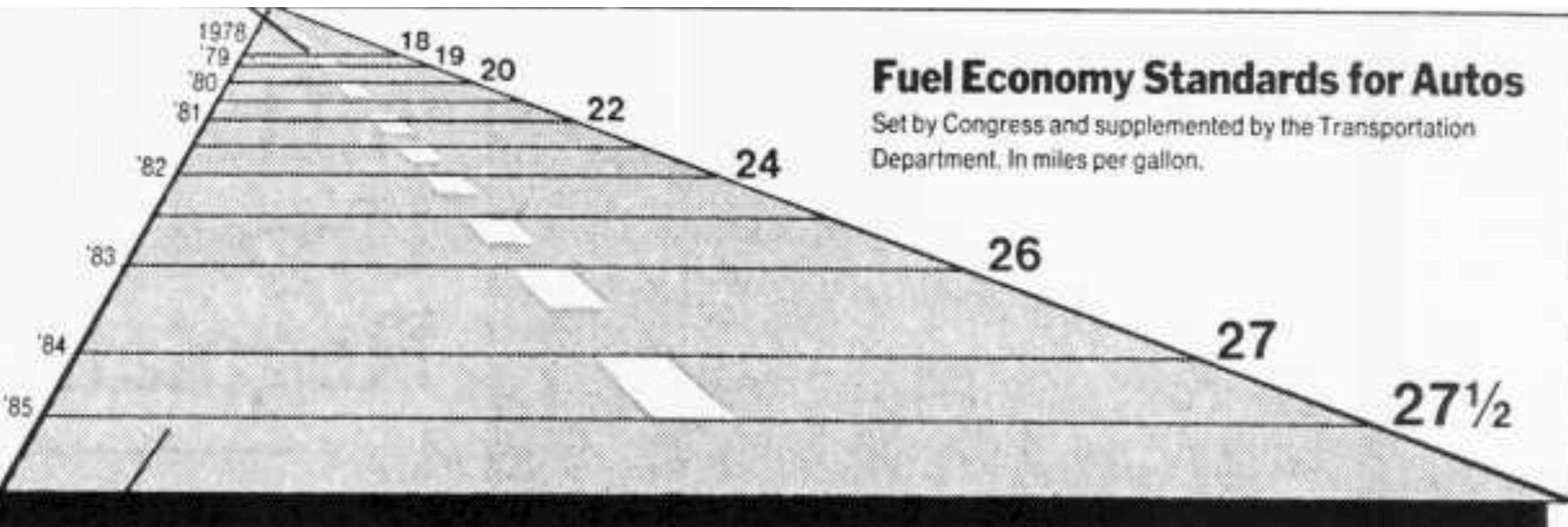


This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.

New York Times, August 9, 1978, p. D-2.

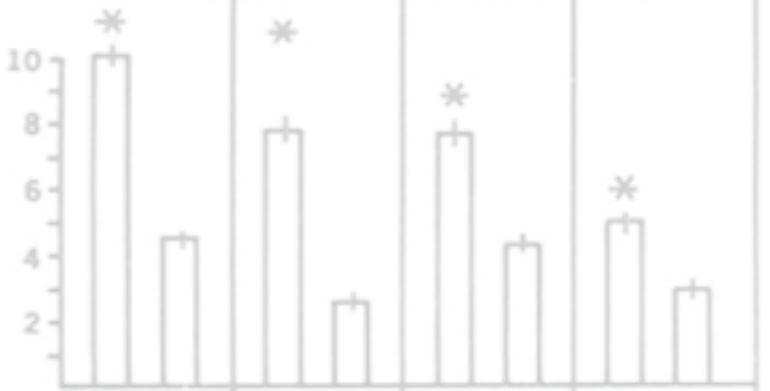
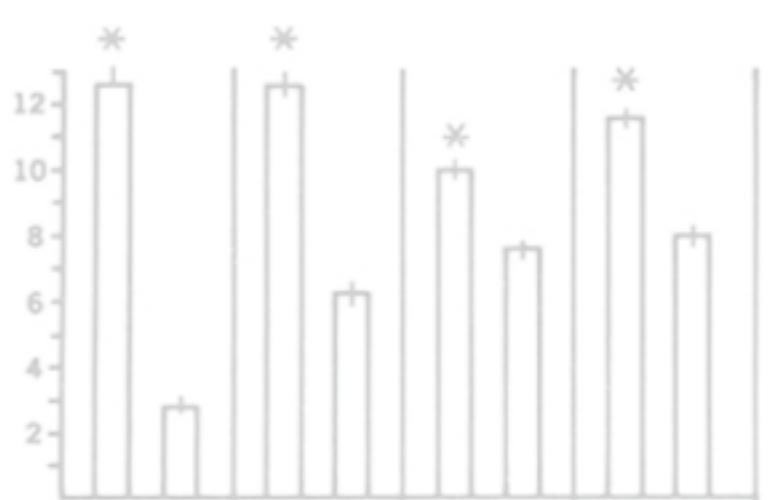
Graphical Integrity

Avoid distortion of numbers by graphic devices, whether accidental or intentional. Show data variation in context, and label them.



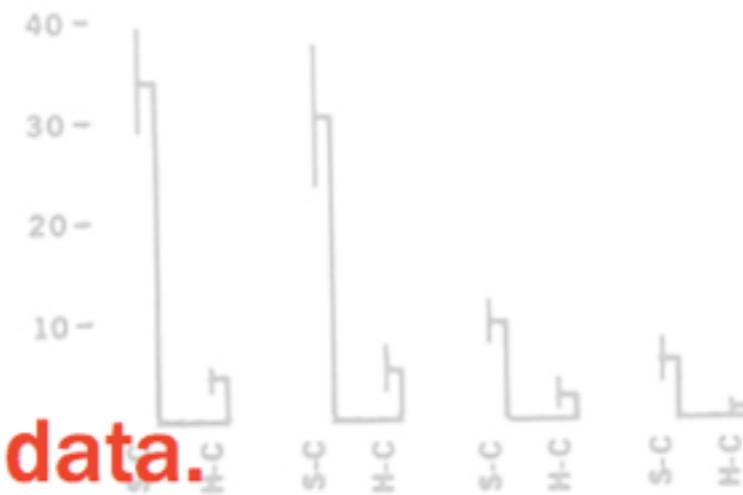
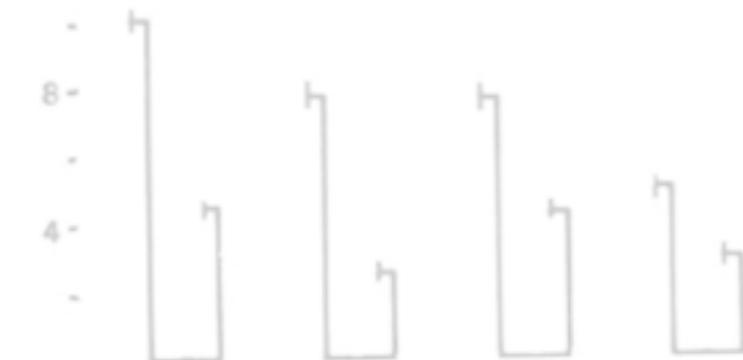
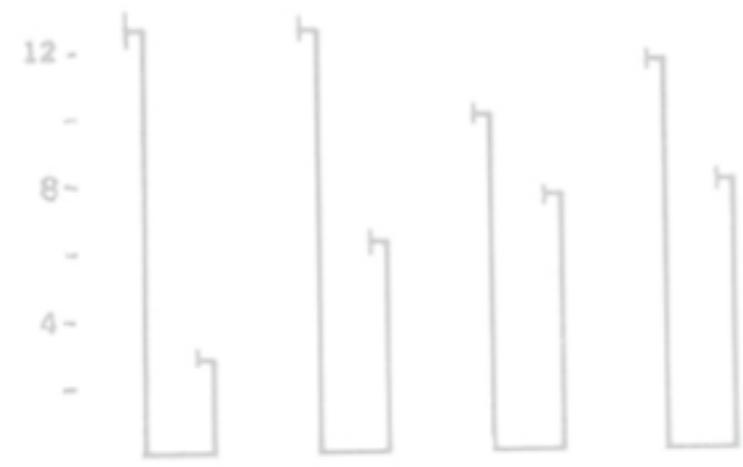
Fuel Economy Standards for Autos

Set by Congress and supplemented by the Transportation Department. In miles per gallon.



Data-ink
Above all else, show the data.
Erase ink that does not report data.

pre GS post GS pre tea post tea

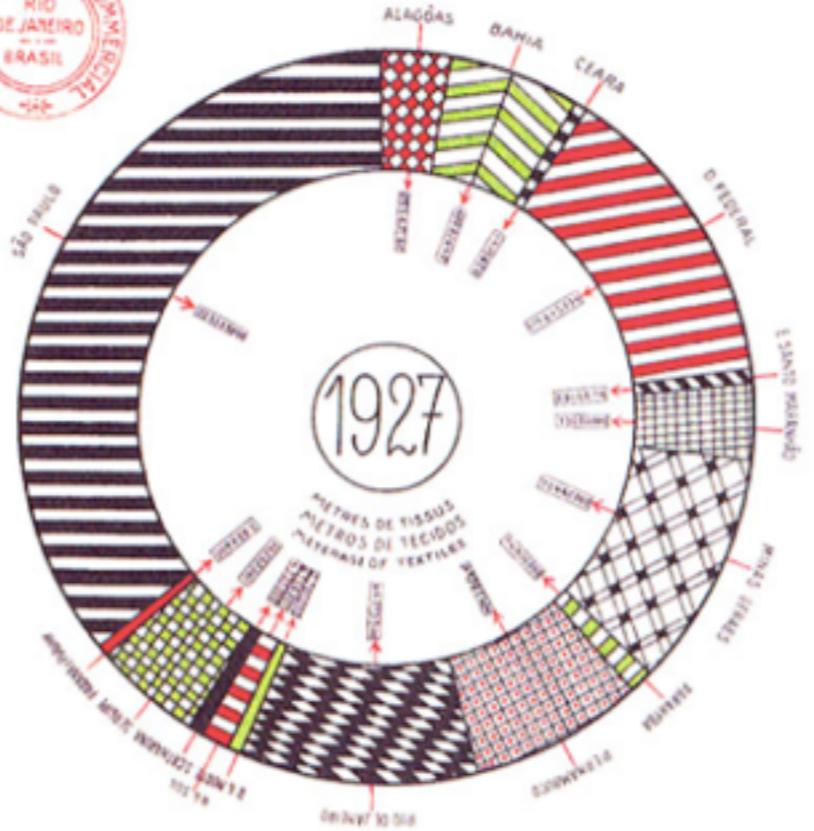
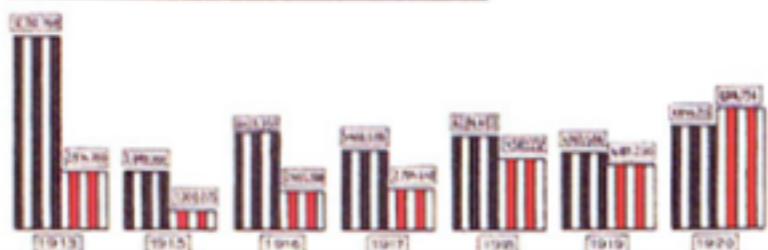
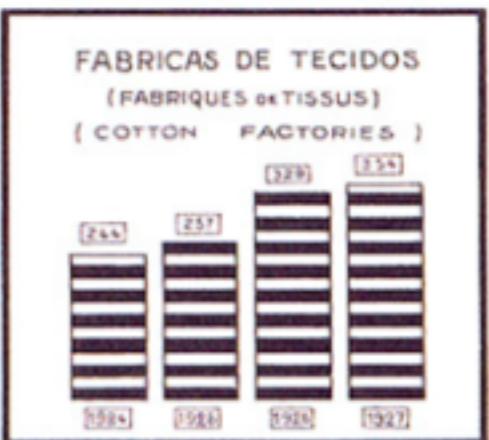


TECIDOS DE ALGODÃO

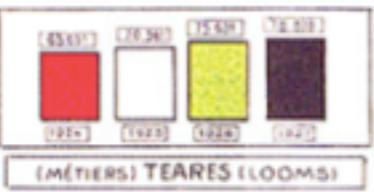
(COTONNADES) (COTTON TEXTILES)



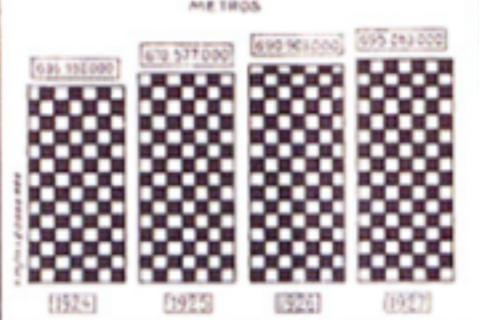
FABRICAS DE TECIDOS
(FABRIQUES de TISSUS)
(COTTON FACTORIES)



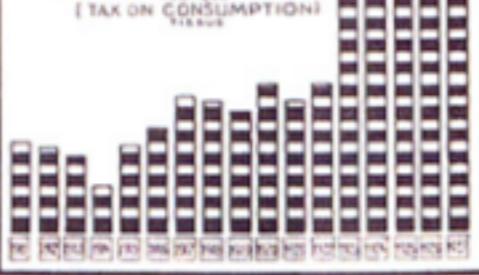
(IMPORTATION)
IMPORTAÇÃO



PRODUÇÃO DE TECIDOS PRODUCTION OF TISSUES (TEXTILE PRODUCTION)

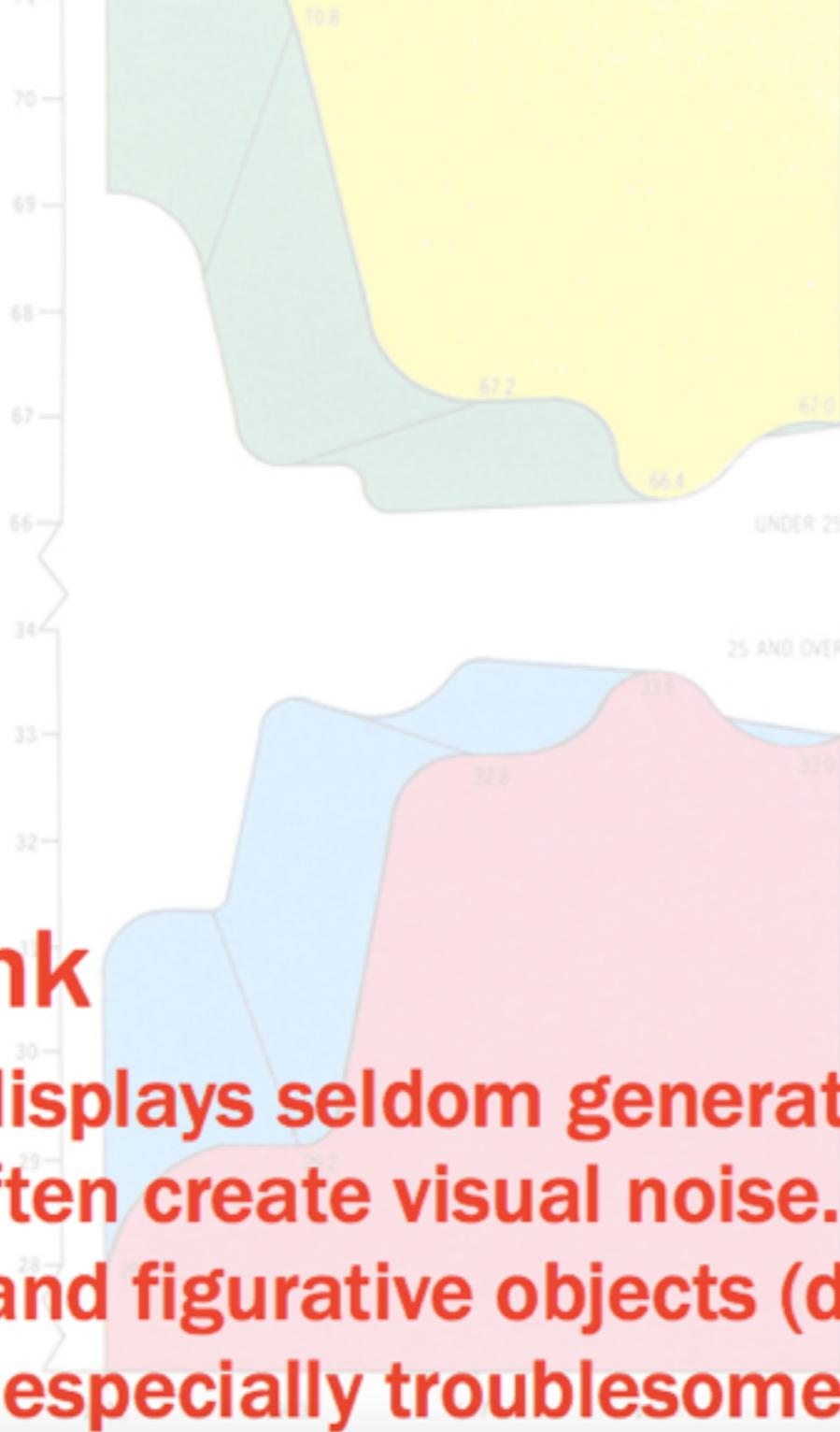


IMPOSTOS DE CONSUMO TECIDOS
(IMPOSTS DE CONSOMMATION TEXTILES)



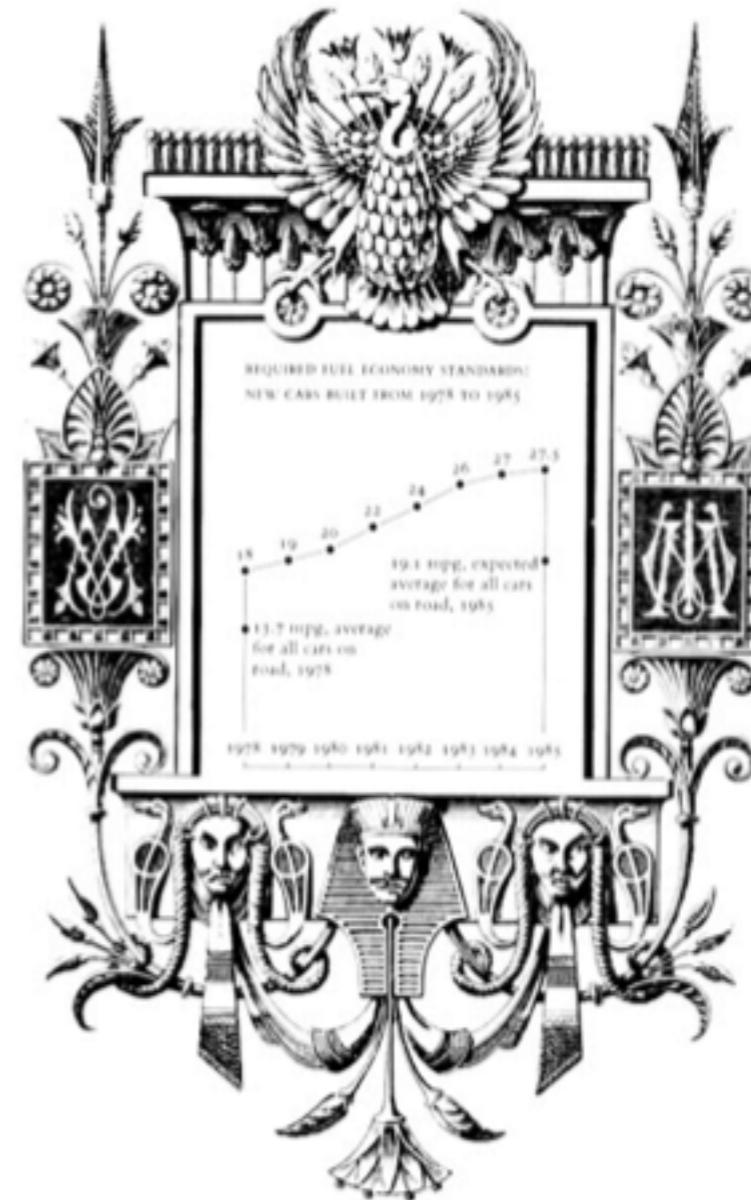
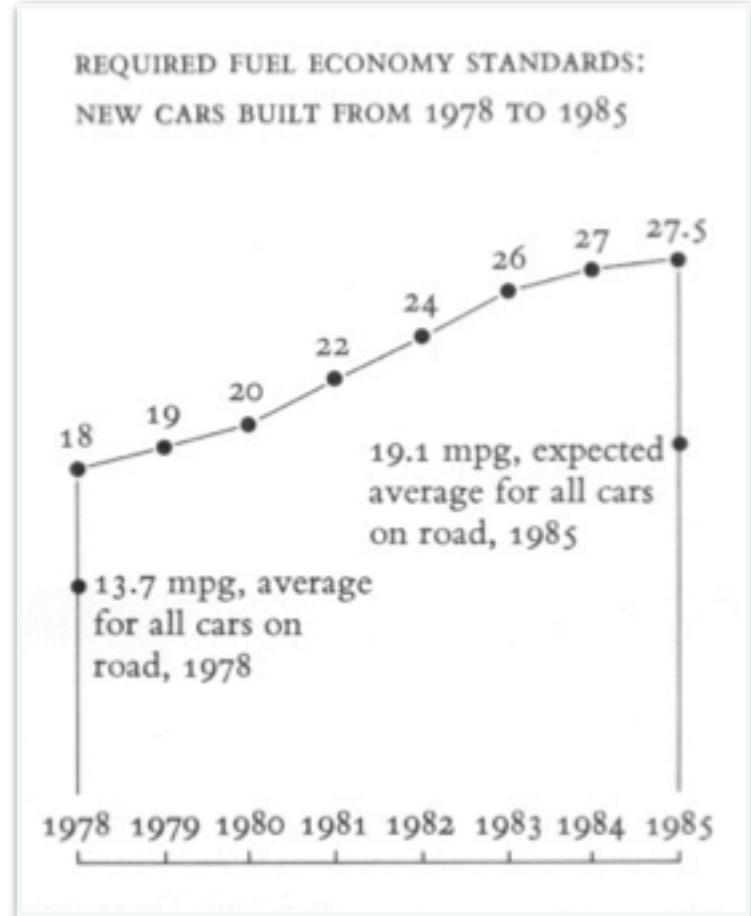
**(IMPORTATION)
IMPORTAÇÃO**

Year	1921-1922 (Black)	1923-1924 (White)	1925-1926 (Red)	1927 (Grey)
1921	271.12	251.12	251.12	251.12
1922	401.38	304.60	304.60	304.60
1923	528.87	371.88	371.88	371.88
1924	716.91	445.90	445.90	445.90
1925	887.75	740.85	740.85	740.85
1926	887.75	640.85	640.85	640.85
1927	887.75	631.85	631.85	631.85



Chartjunk

Inventive displays seldom generate interest, but they often create visual noise. Grids, hatching, and figurative objects (ducks) tend to be especially troublesome.





Micro-macro

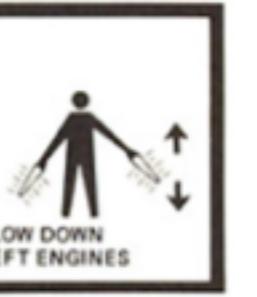
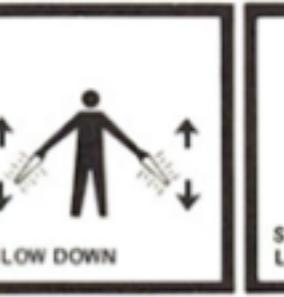
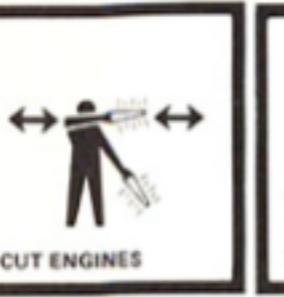
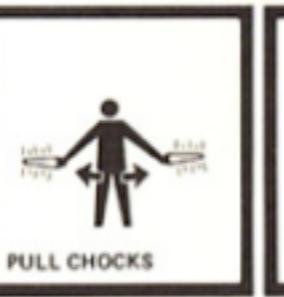
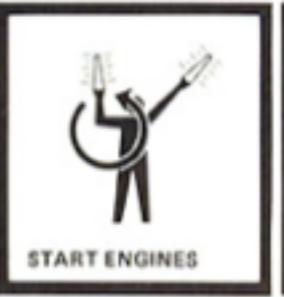
Graphics should be readable in whole images and in parts, and that should help manage levels of detail.



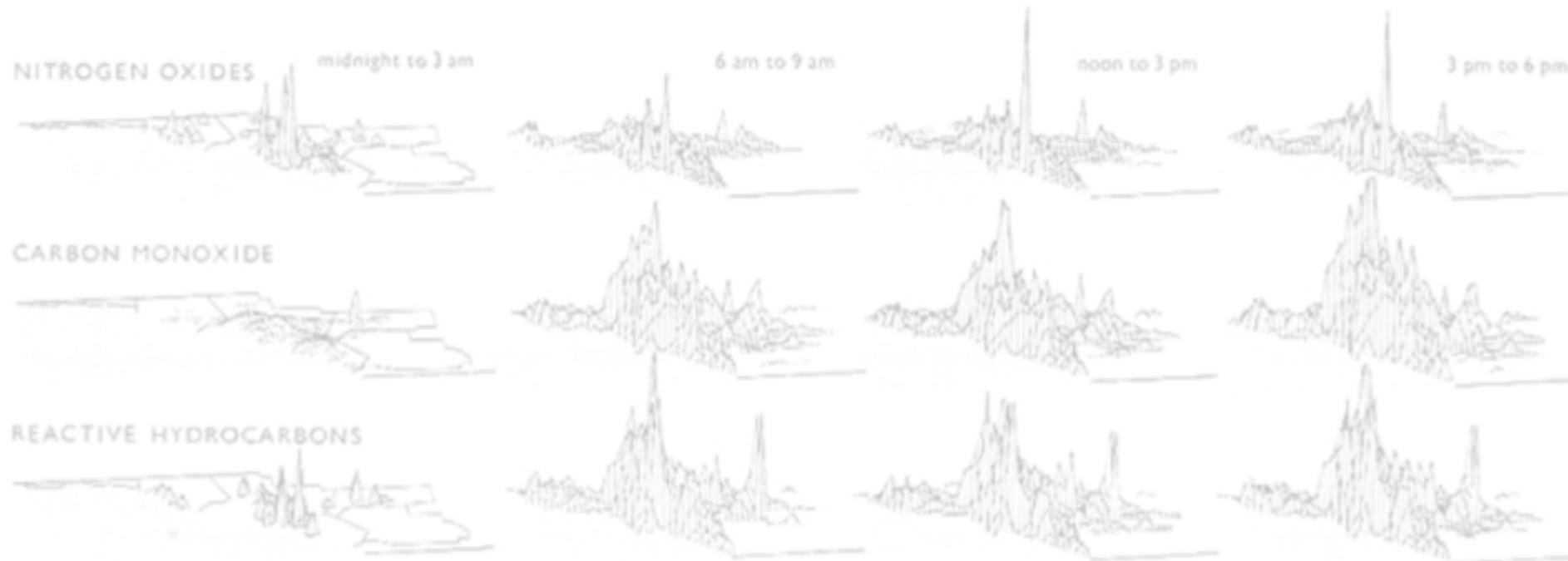


Layering

Reveal the complexity of data, and don't blame it for any confusion. Use graphic devices to separate categories. Give layers hierarchy and levels of detail.



MARSHALLING SIGNALS	PROCEED; WATCH SIGNALS	THIS WAY	PROCEED TO NEXT SIGNALMAN	TURN LEFT	TURN RIGHT	MOVE AHEAD
STOP	START ENGINES	INSERT CHOCKS	PULL CHOCKS	CUT ENGINES	SLOW DOWN	SLOW DOWN LEFT ENGINES

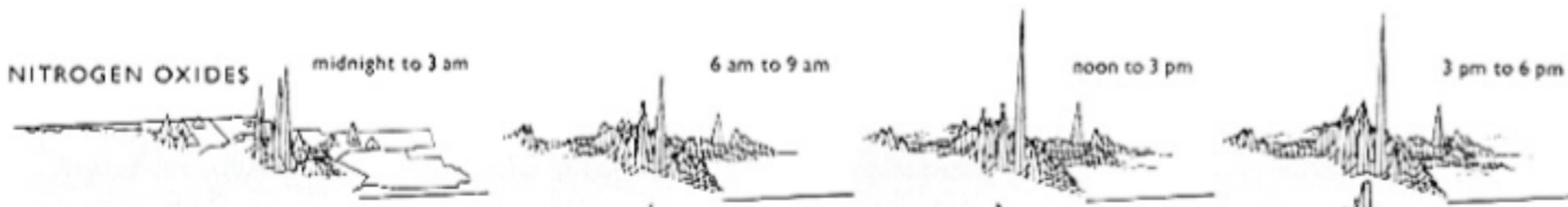


Small multiples

High-quality information graphics portray many numbers per square inch. Small multiple, comparative images work especially well for this.

NITROGEN OXIDES

midnight to 3 am



CARBON MONOXIDE



REACTIVE HYDROCARBONS



BEST PRACTICES

AVOID CLUTTER & DISTRACTIONS

“Everyone spoke of an information overload, but what there was in fact was a non-information overload.”

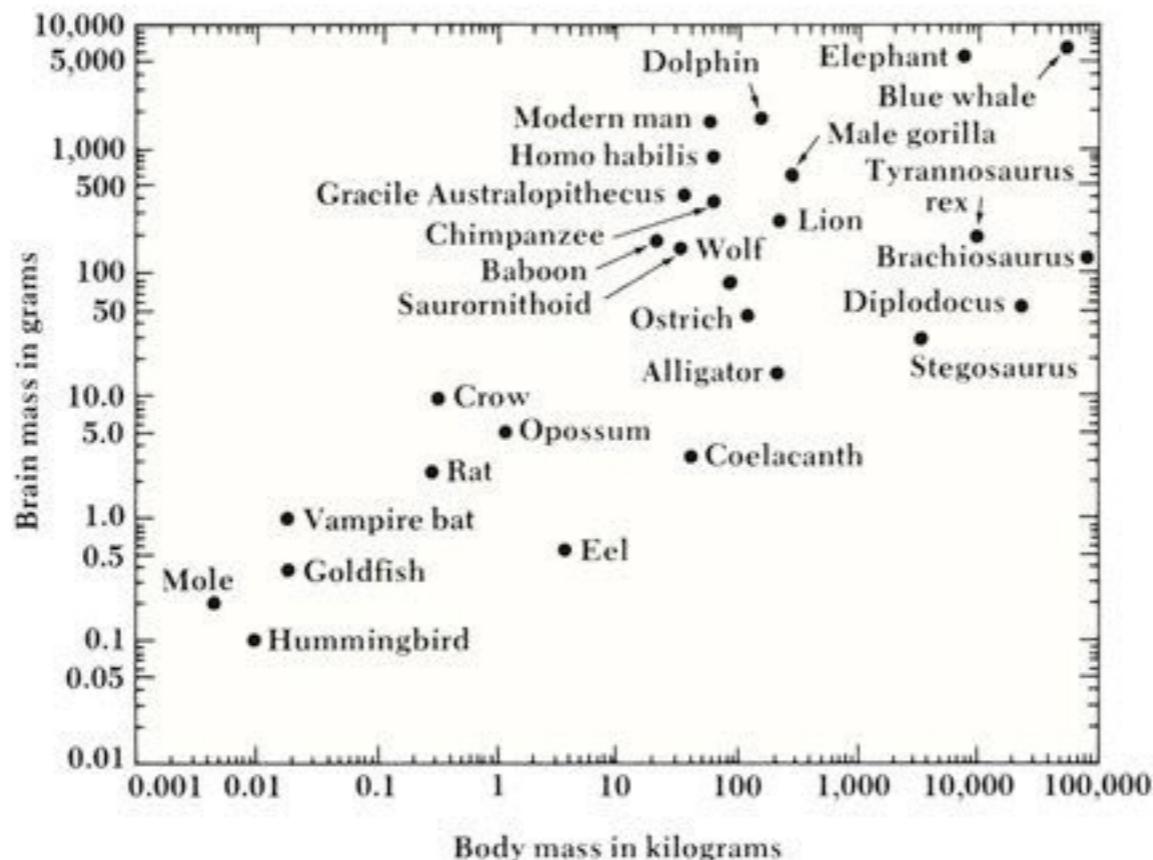
– RICHARD SAUL WURMAN, 1976

American architect and graphic designer
Creator of Access guides
Created the TED conferences

LESS IS MORE

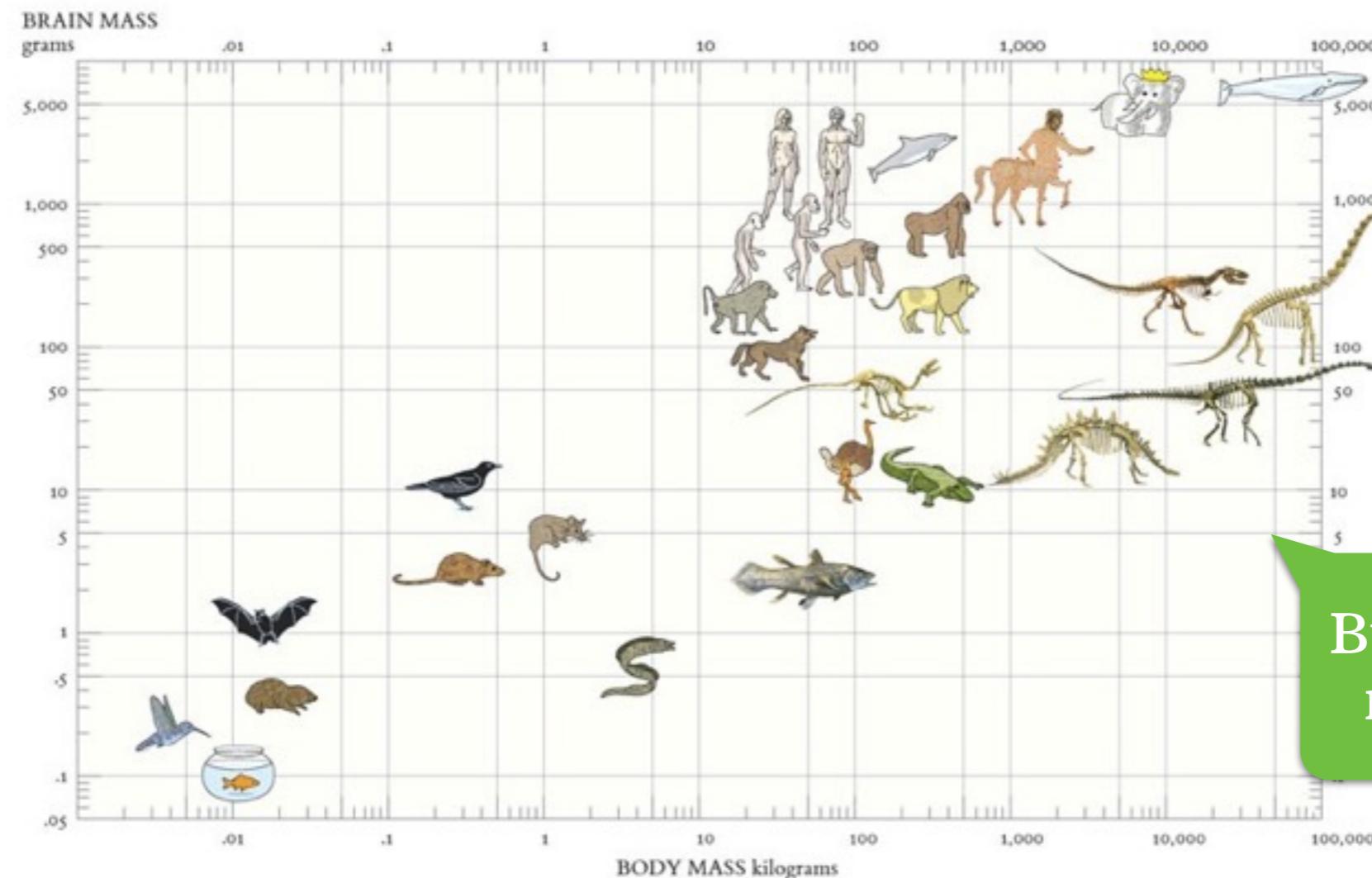
Remove
to improve
(the **data-ink** ratio)

CLARITY THROUGH CREATIVITY



- EDWARD TUFTE

CLARITY THROUGH CREATIVITY



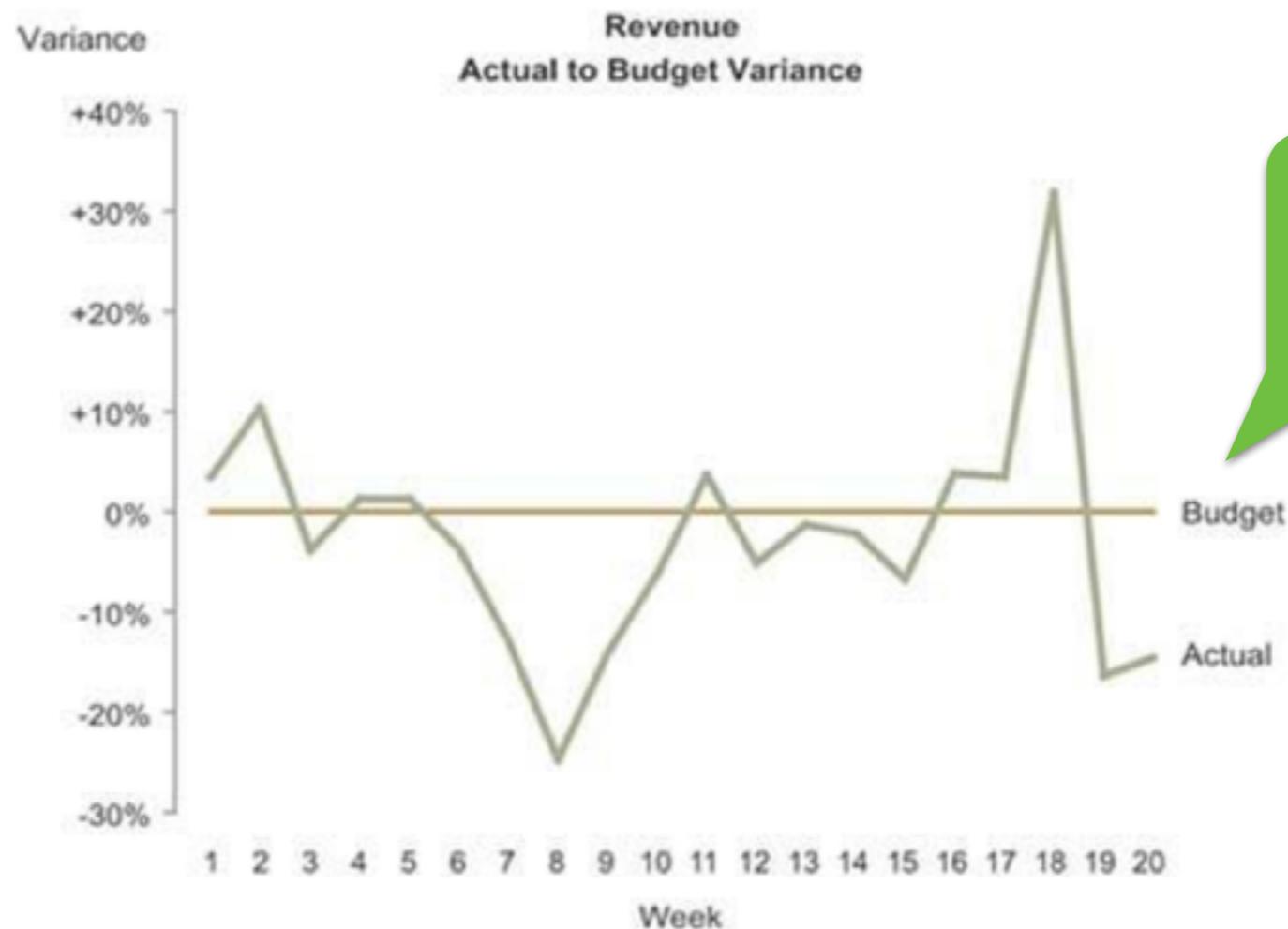
But what data is missing now?

TELL A STORY WITH YOUR DATA

- With (non-distracting) imagery
- Make comparisons (product, goals, competition)
- Show the effect of time

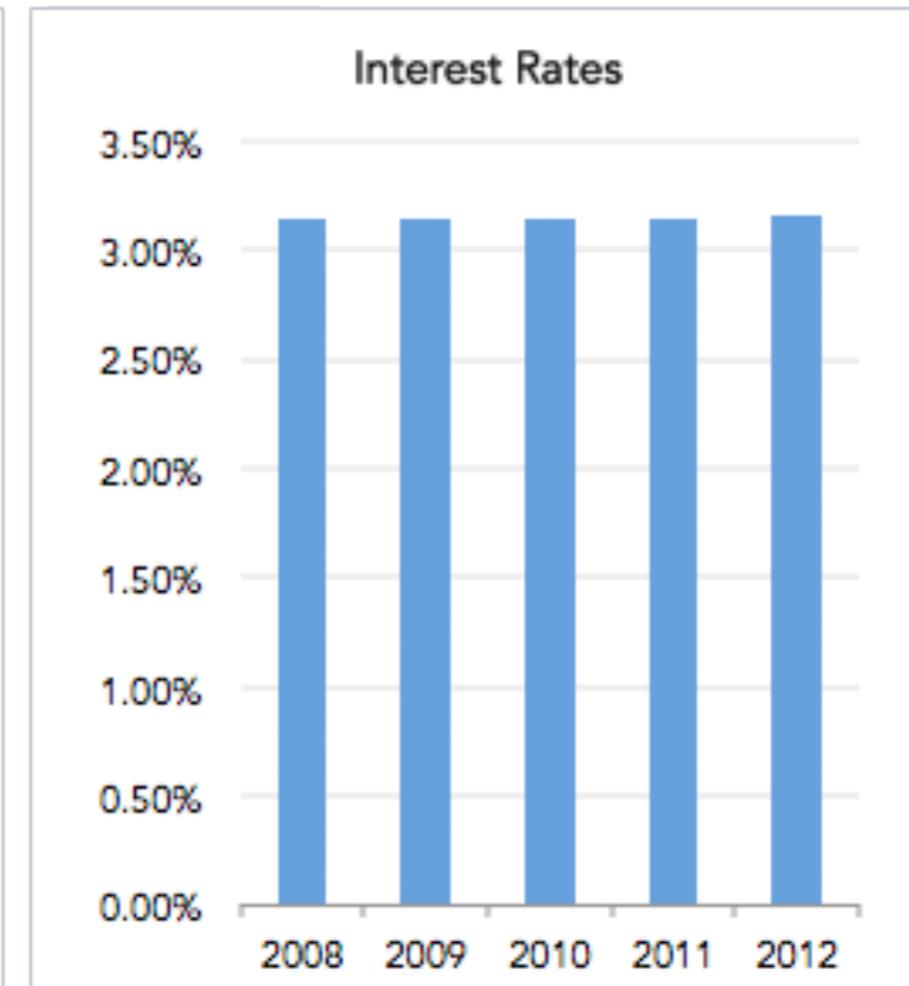
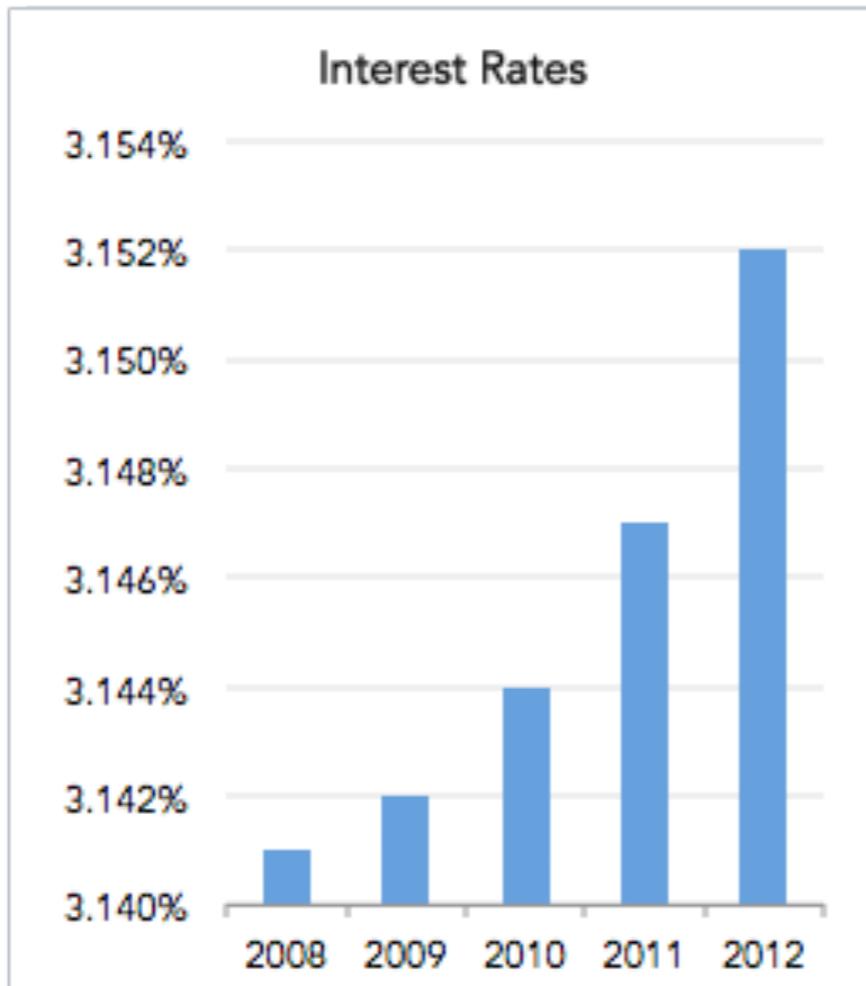


CHOOSE THE RIGHT MEASURE

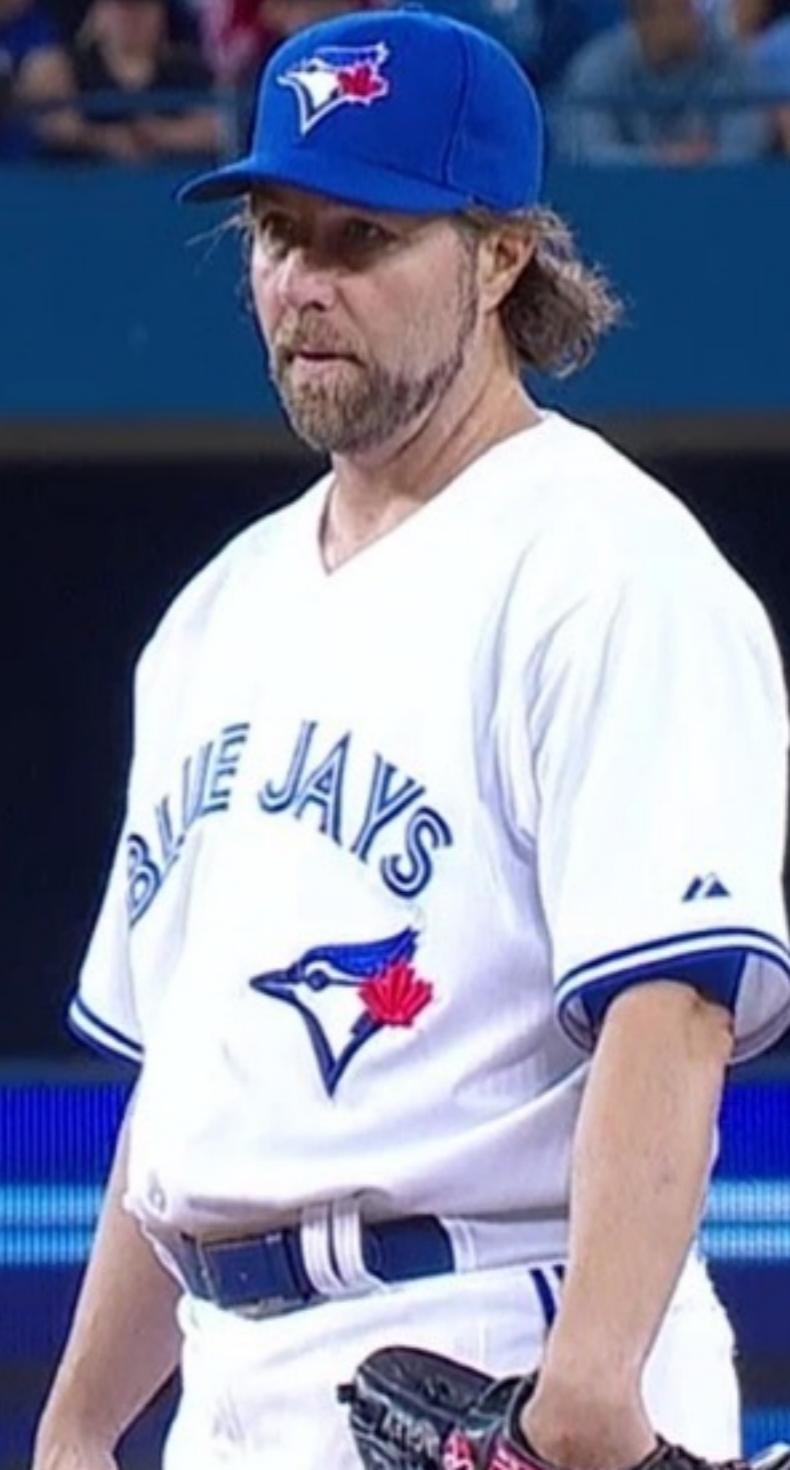
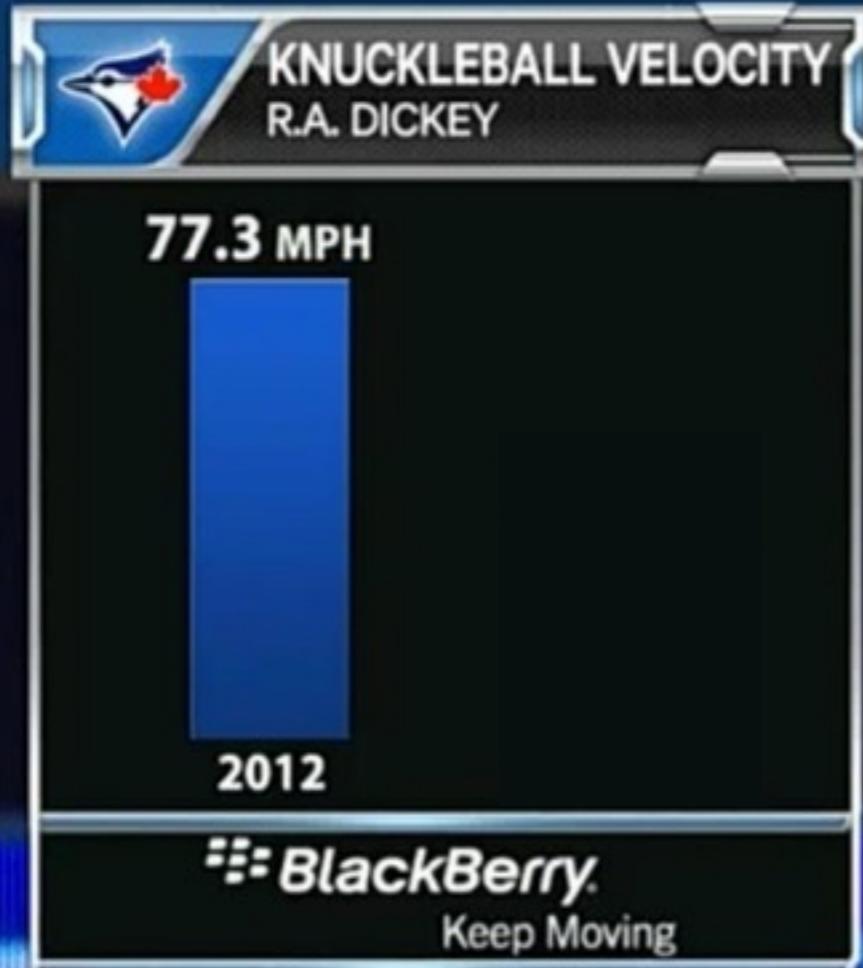


Did this project
come in budget?

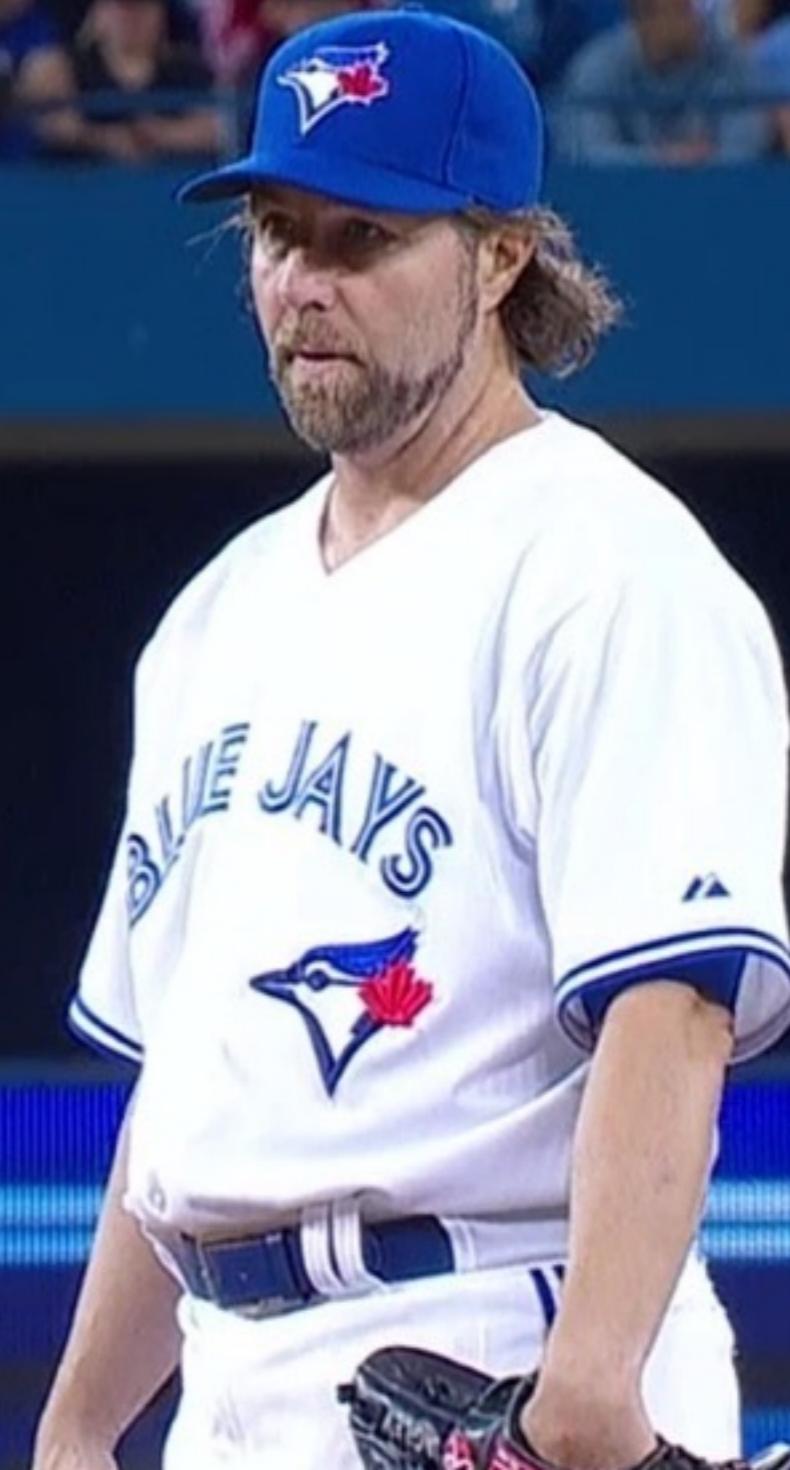
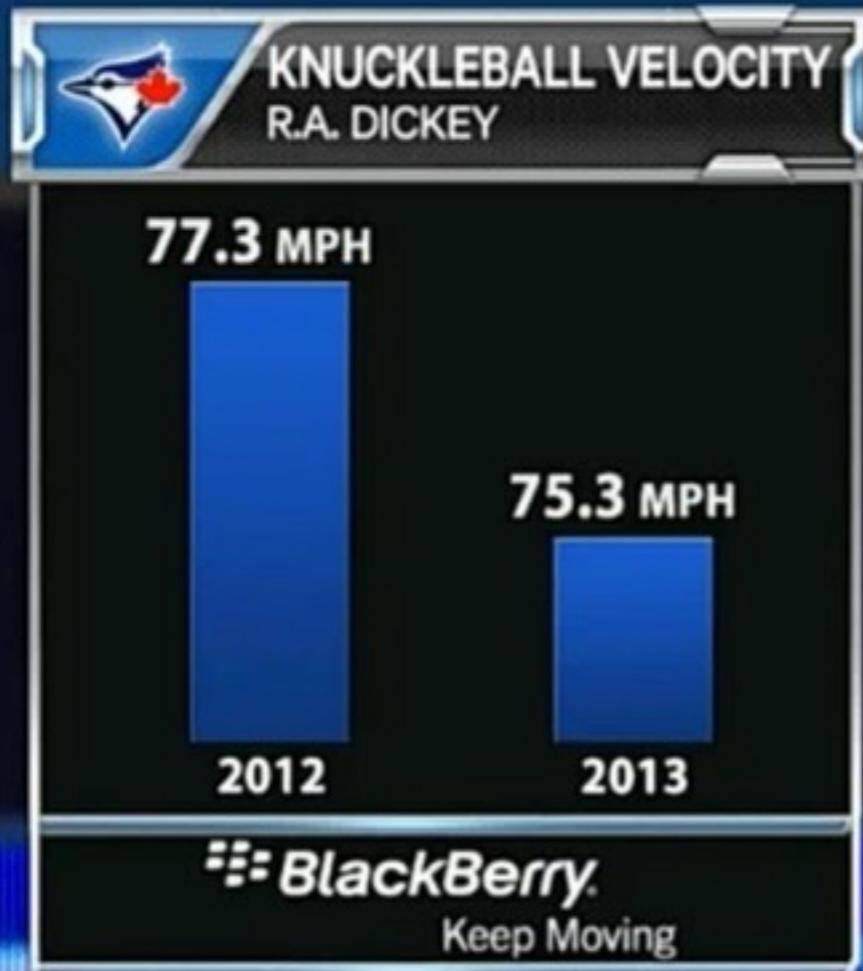
BE HONEST



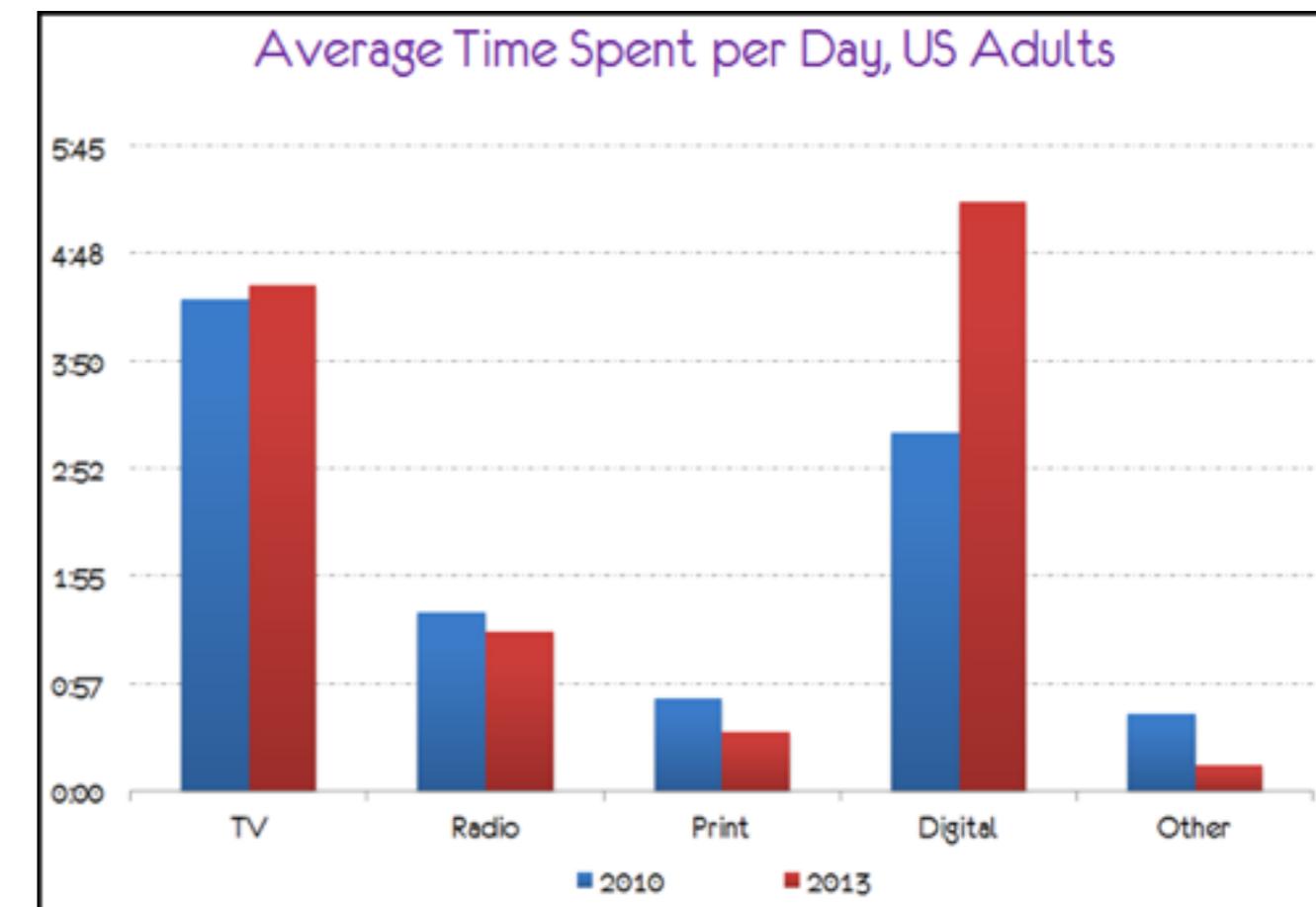
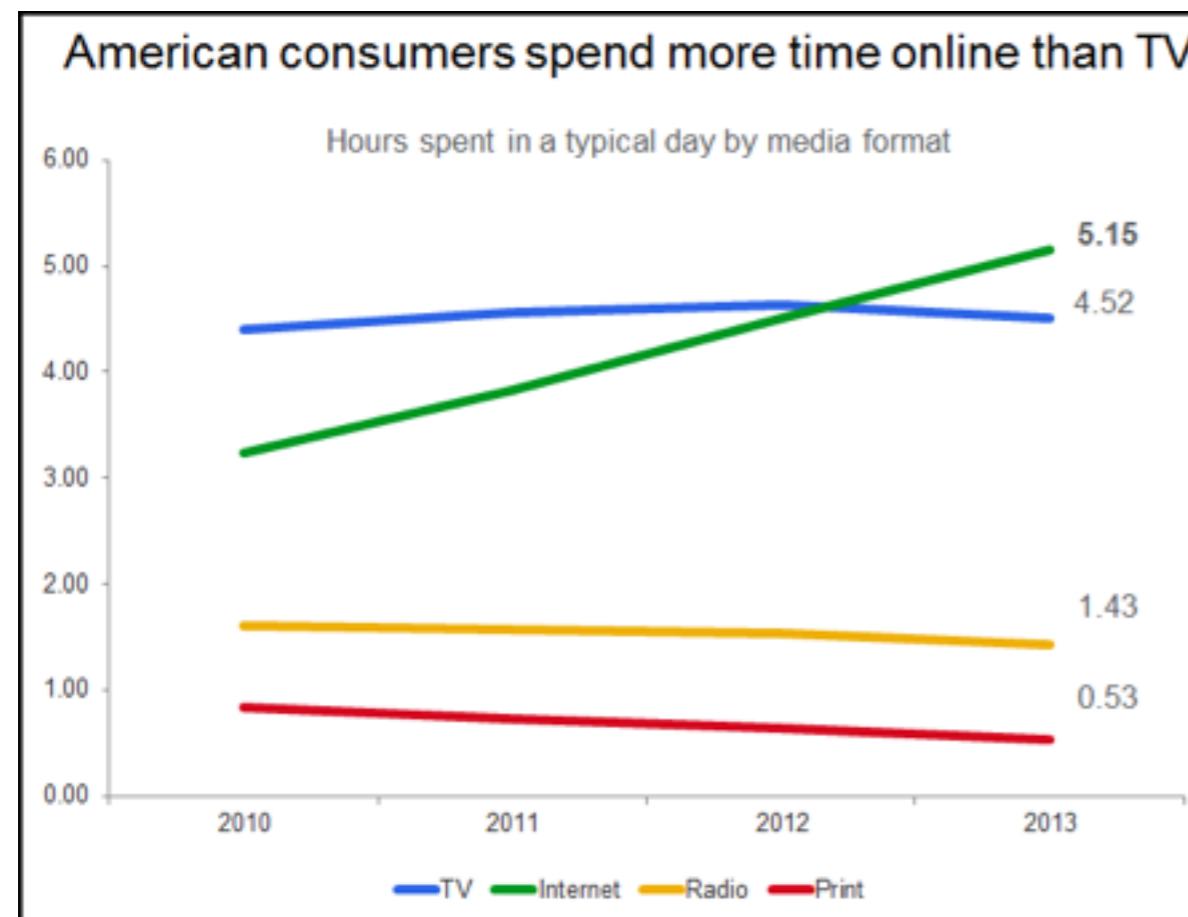
BE HONEST



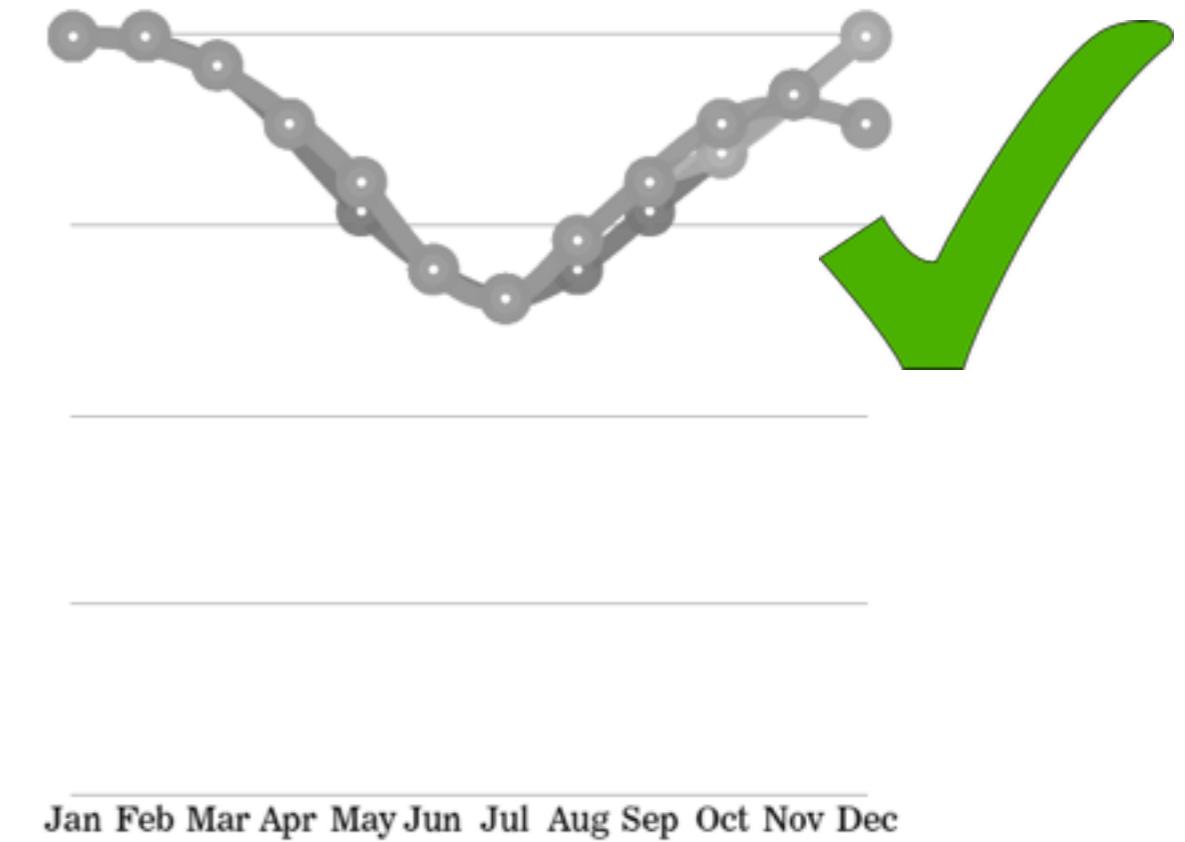
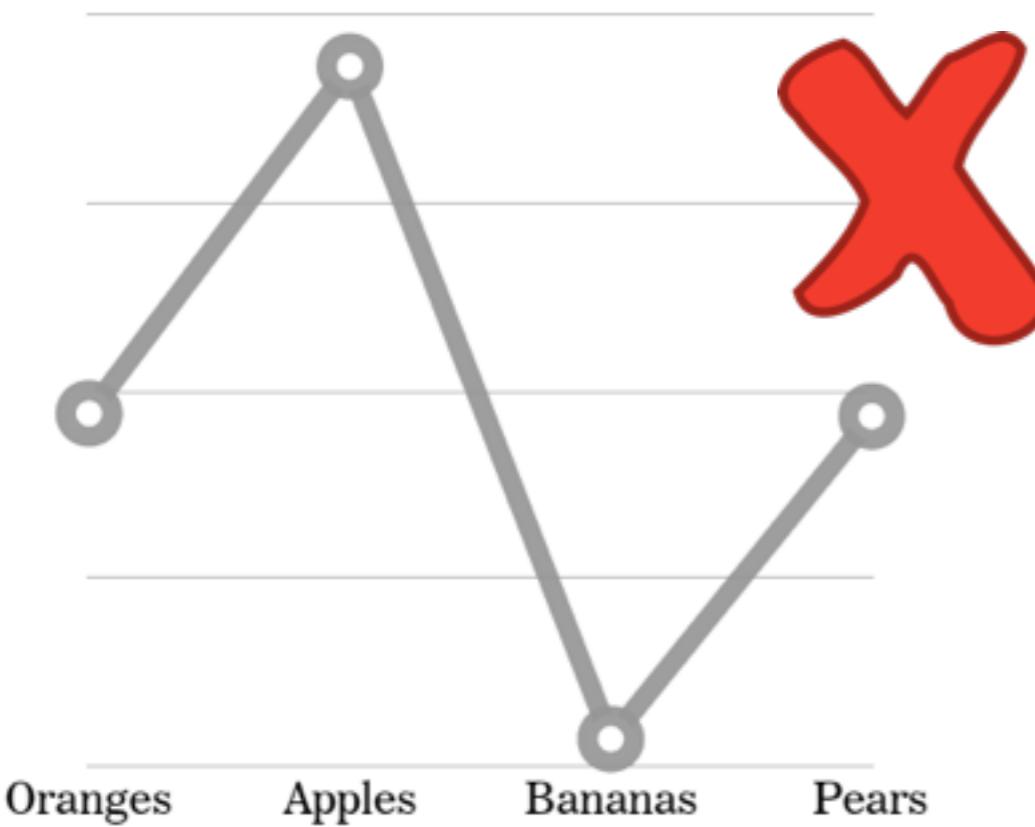
BE HONEST



DIFF VISUALIZATIONS HIGHLIGHT DIFF MESSAGES



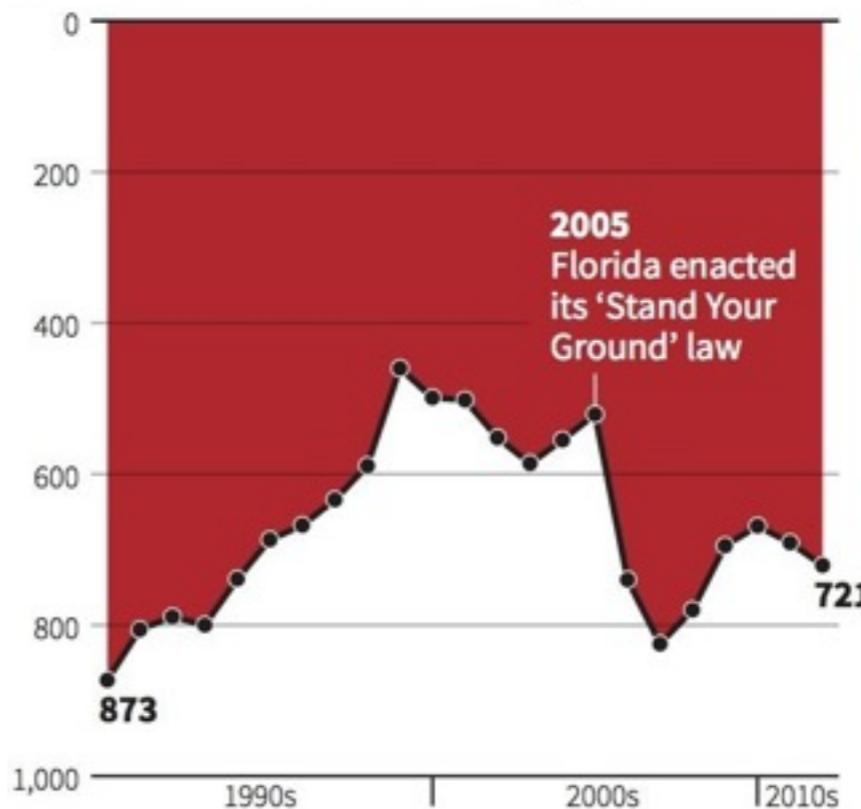
LINES FOR CONTINUOUS SPECTRUMS, LIKE TIME!



USE CONVENTIONS

Gun deaths in Florida

Number of murders committed using firearms

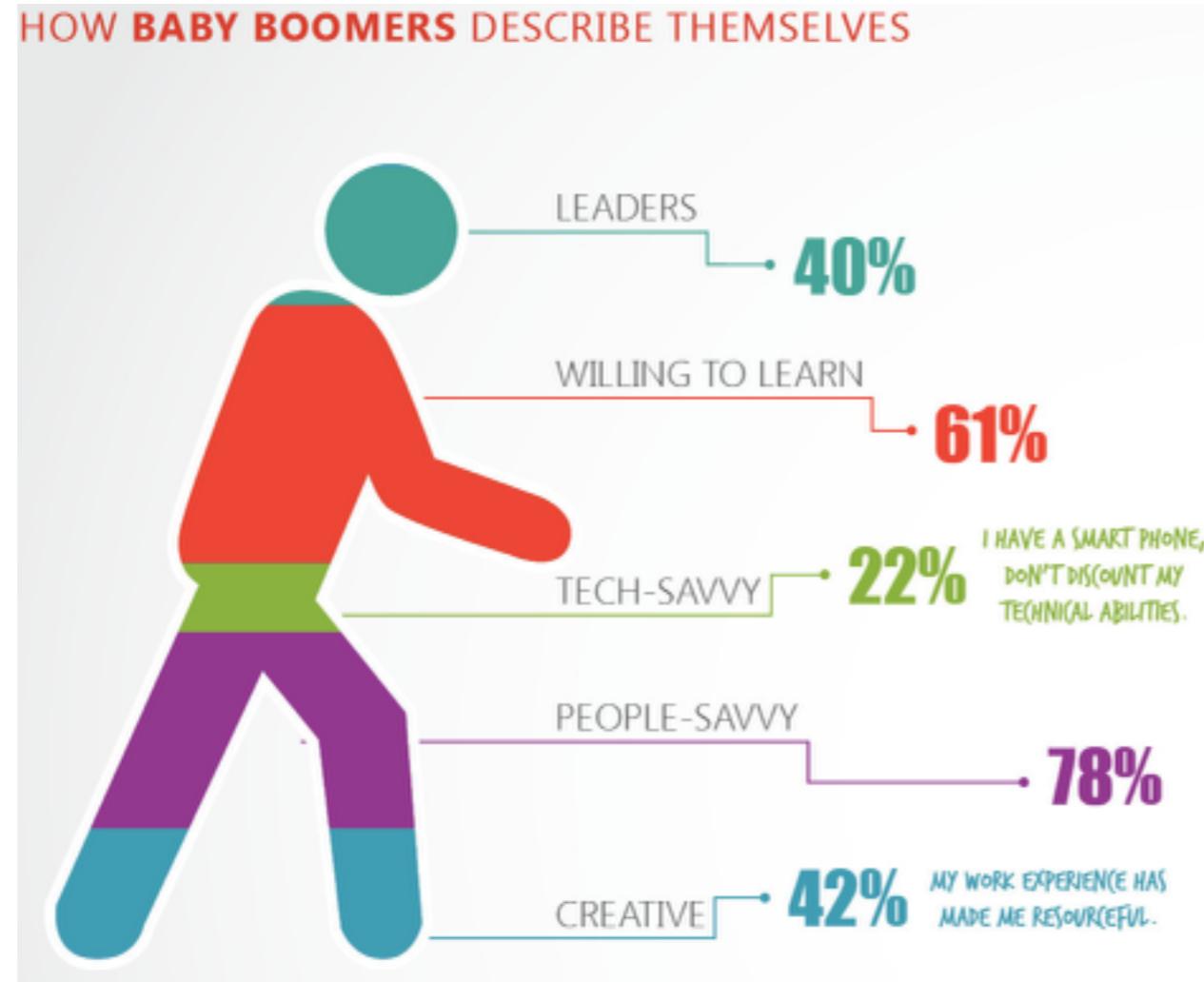


Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

DESCRIBE PARTS OF WHOLE



VIEWERS HAVE TROUBLE QUICKLY COMPARING

- 2-dimensional sizes
- Especially if angled
(eg. pie graphs)

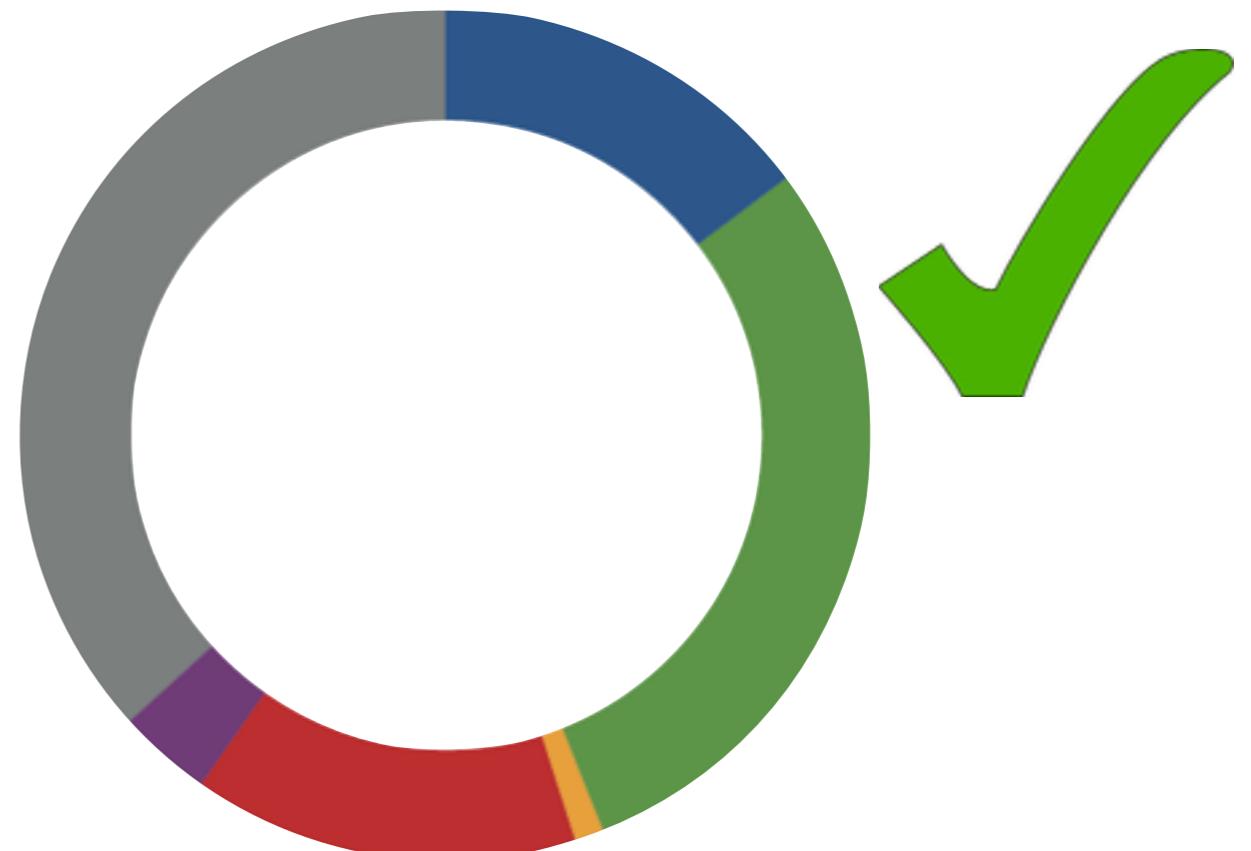
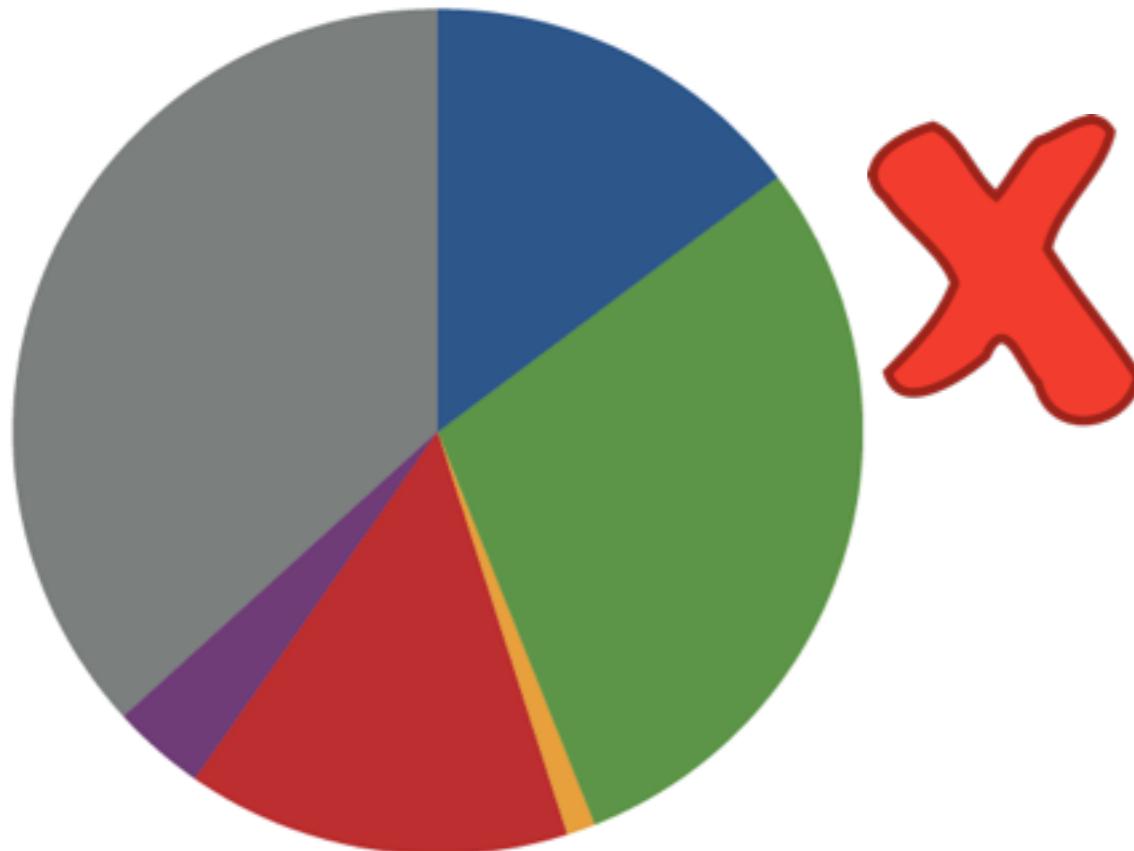


The main takeaway?

Visual-Area math is hard for humans, especially sector slices

arcadenw.org/article/scientific-visualization-unaidstreemaps-and-more

VIEWERS HAVE TROUBLE QUICKLY COMPARING



SO WHAT DOES THIS ALL MEAN?

- Make it neat, clean, and clear
- Tell a story
- Avoid pie charts
- Choose the best-fit visualization
- Follow conventions
- Include & understand context
- Focus — include only what's important

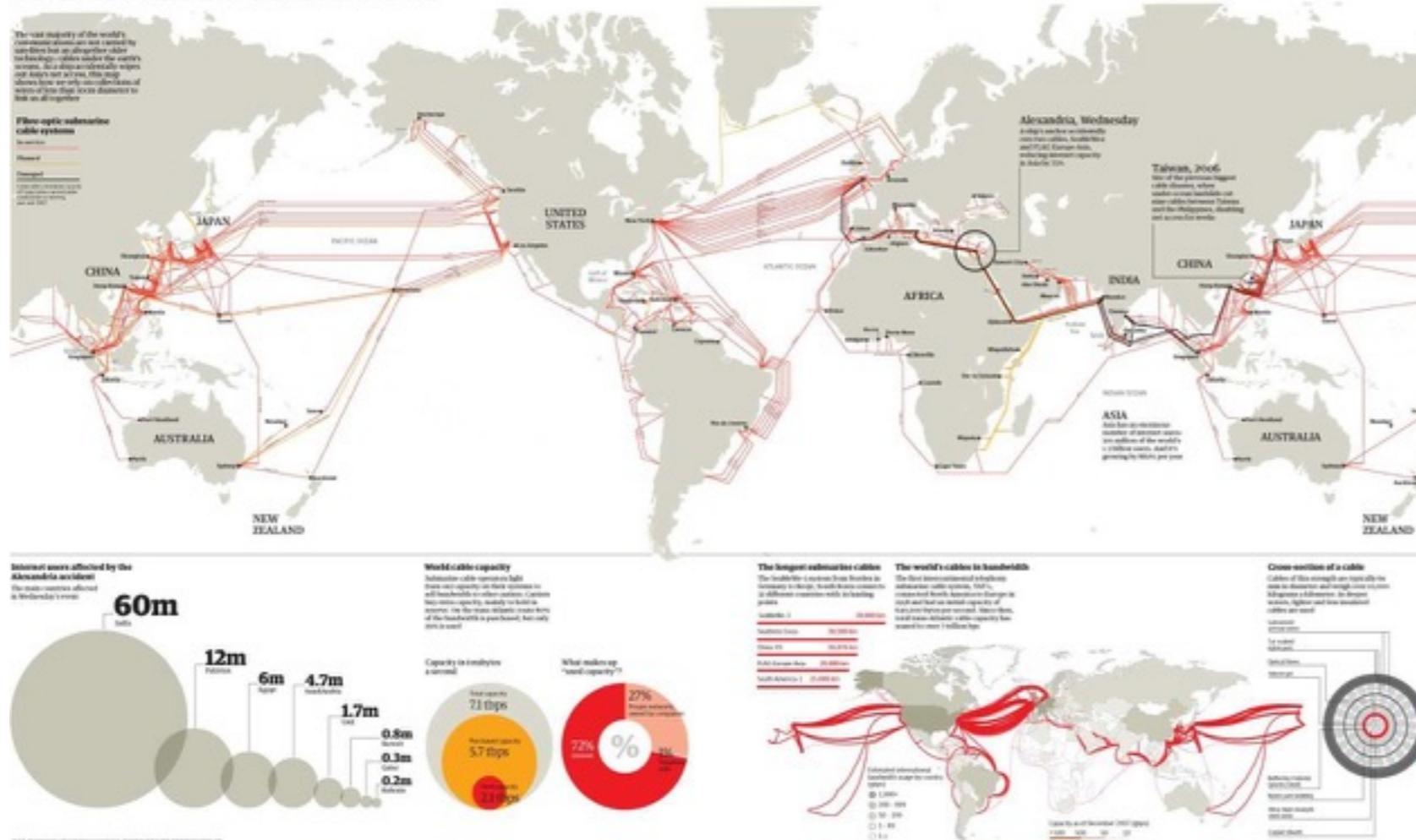
OTHER TYPES OF DATA VISUALIZATION

WORD CLOUDS



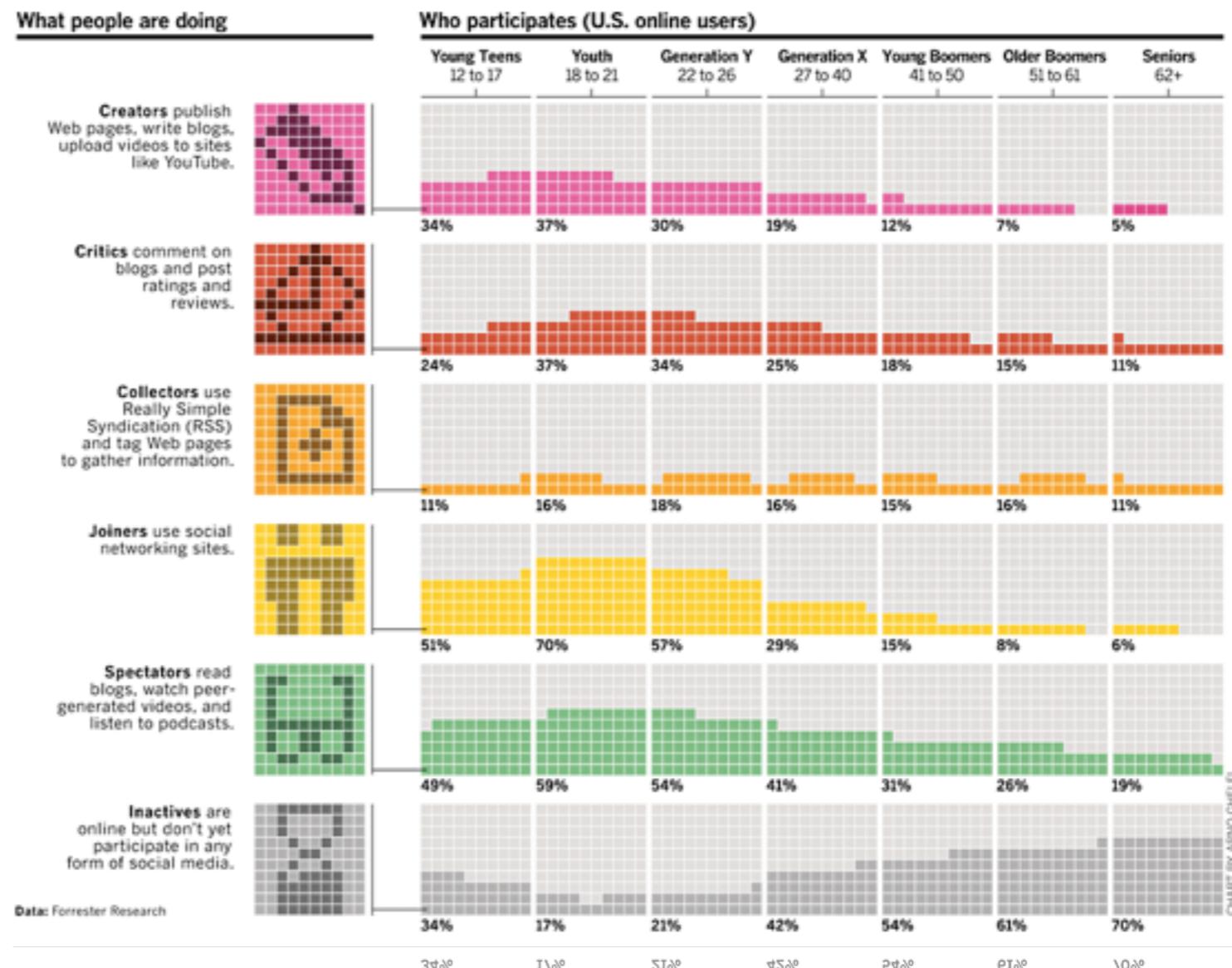
INFOGRAPHICS

The internet's undersea world



COMPLEX DATA

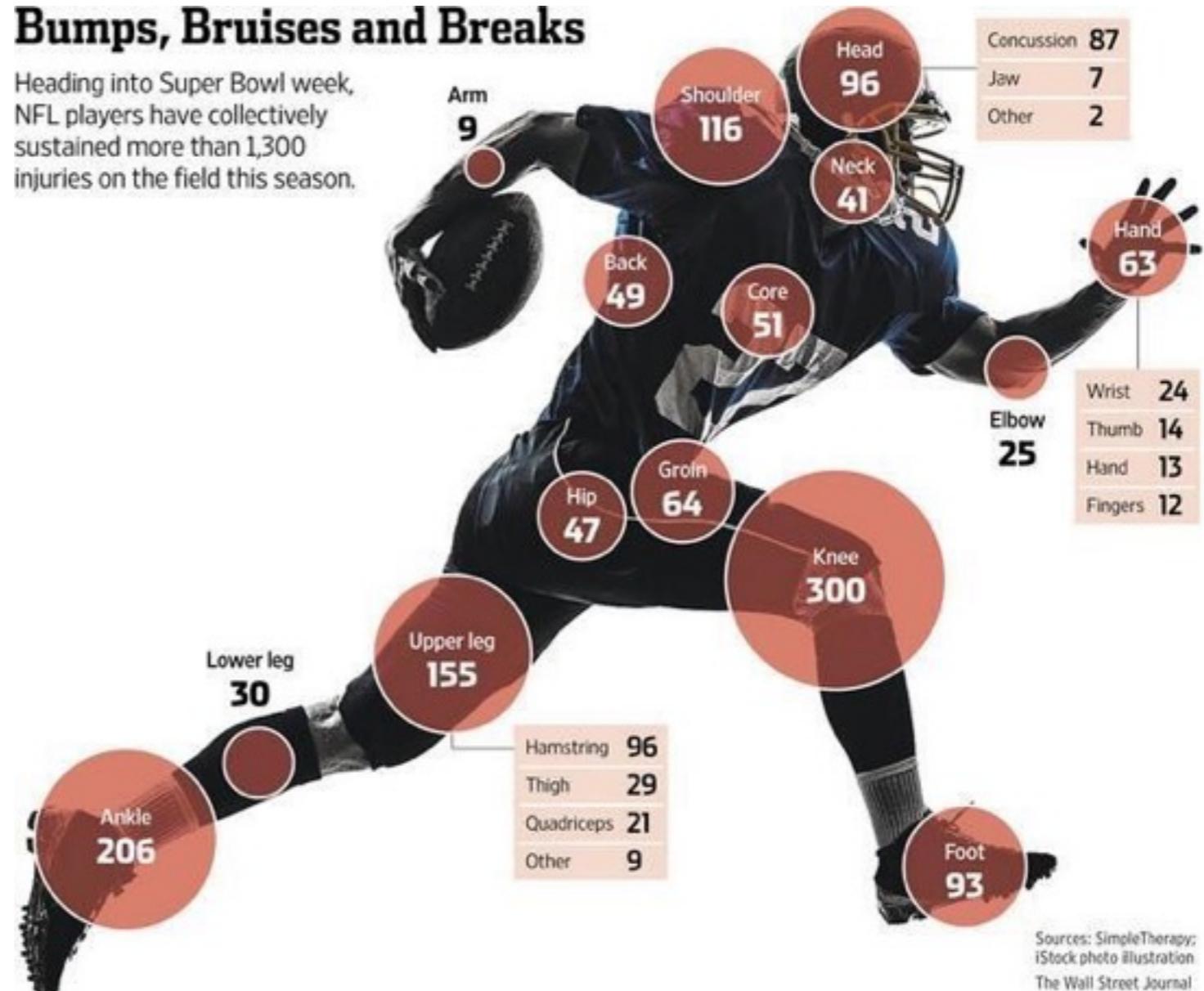
Complex data visualizations
visualizations
need more
explanation
(but what's too much?)



LAYERING DATA

Bumps, Bruises and Breaks

Heading into Super Bowl week, NFL players have collectively sustained more than 1,300 injuries on the field this season.



OTHER TYPES OF DATA VISUALIZATION

65

INTERACTIVITY

Jobs Charted by State and Salary

BY NATHAN YAU / POSTED TO DATA UNDERLOAD / TAGS: JOBS, SALARY



By FlowingData.com | Source: Bureau of Labor Statistics

INTERACTIVITY



Homeless Australians

Some 105,200 Australians are without a home.

This Fact File, based on the latest research in Australia, is one of the first using geographical areas to reveal how the patterns of homelessness have changed over 15 years.

Homelessness is a loss of a sense of security, stability, privacy, safety or the ability to control your living space.

Indigenous Australians and single parents are over-represented in the numbers, with domestic and family violence the main reasons they seek help from homelessness services.

The fastest growing group of homeless people are those living in severely **overcrowded accommodation**.

Rates of homelessness are highest in northern remote and very remote regions, followed by inner cities and some growth corridors.

Twenty-seven per cent of Australia's homeless live in New South Wales, 22 per cent in Victoria, 19 per cent in Queensland, 15 per cent in the Northern Territory, 9 per cent in Western Australia, 6 per cent in South Australia, 2 per cent in Tasmania and 1 per cent in the ACT.

ⓘ Click on a hexagon for more information about a region. Zoom in to see more hexagons and for a closer look.



SOME INSPIRATION

information is beautiful

ideas, issues, knowledge, data – visualized!

Home About Blog Data Books Workshops Contact



Popular topics: interactive belief food health ideas science web-tech



<http://www.informationisbeautiful.net/>

FURTHER RESOURCES

<http://www.vislives.com/2011/10/xkcd-visualizations.html> - fun

<http://viz.wtf/> - bad charts

flowingdata.com

informationisbeautiful.net

Blog Visually ([Learning From Mistakes](#), [Communicating Quantities](#))

ALSO JUST READ EVERYTHING EDWARD TUFTE

IF YOU TAKE NOTHING ELSE AWAY

Use visualizations to **simplify**, not jazz things up

Visualization makes non-visual data, such as numbers, **relatable**

Be mindful of how you convey the data

“ What is to be sought in designs for the display of information is the clean portrayal of complexity. Not the complication of the simple; rather the task of the designer is to give visual access to the subtle and difficult - that is, the revelation of the complex. ”



- EDWARD TUFTE