

How (not) to fall for “bad science”

CS 7123, Spring 2025

Maryam Tabar, Ph.D.

Used/adopted materials from:

[1] Slides of the “Research Design: Interdisciplinary Research Methods for Information Sciences and Technology” course, taught by Prof. Dongwon Lee, at Pennsylvania State University

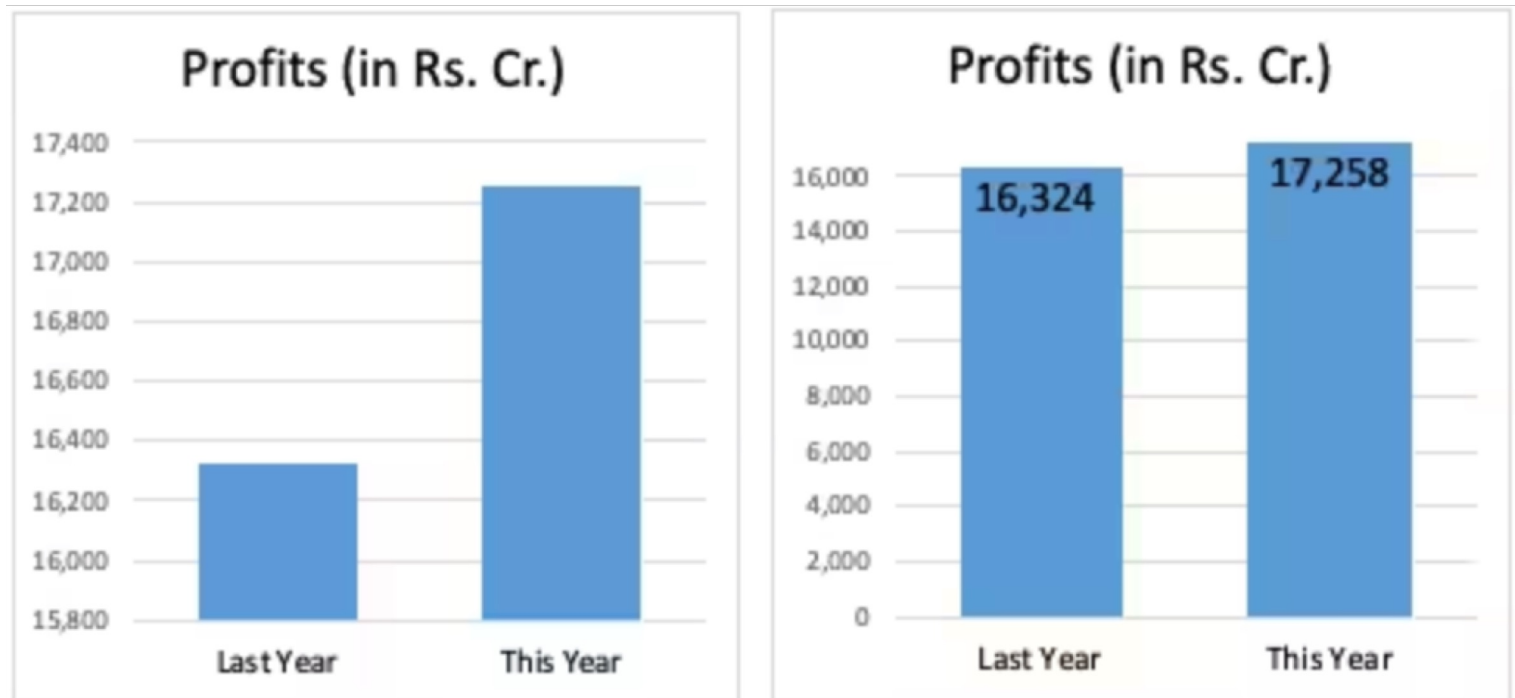
[2] Slides of the Data Science course, taught by Prof. Steven S. Skiena at Stony Brook University (The Data Science Design Manual, by Steven S. Skiena. ISBN: 9783319554433)

Agenda

- Visualization Tricks
- Correlation vs. Causation

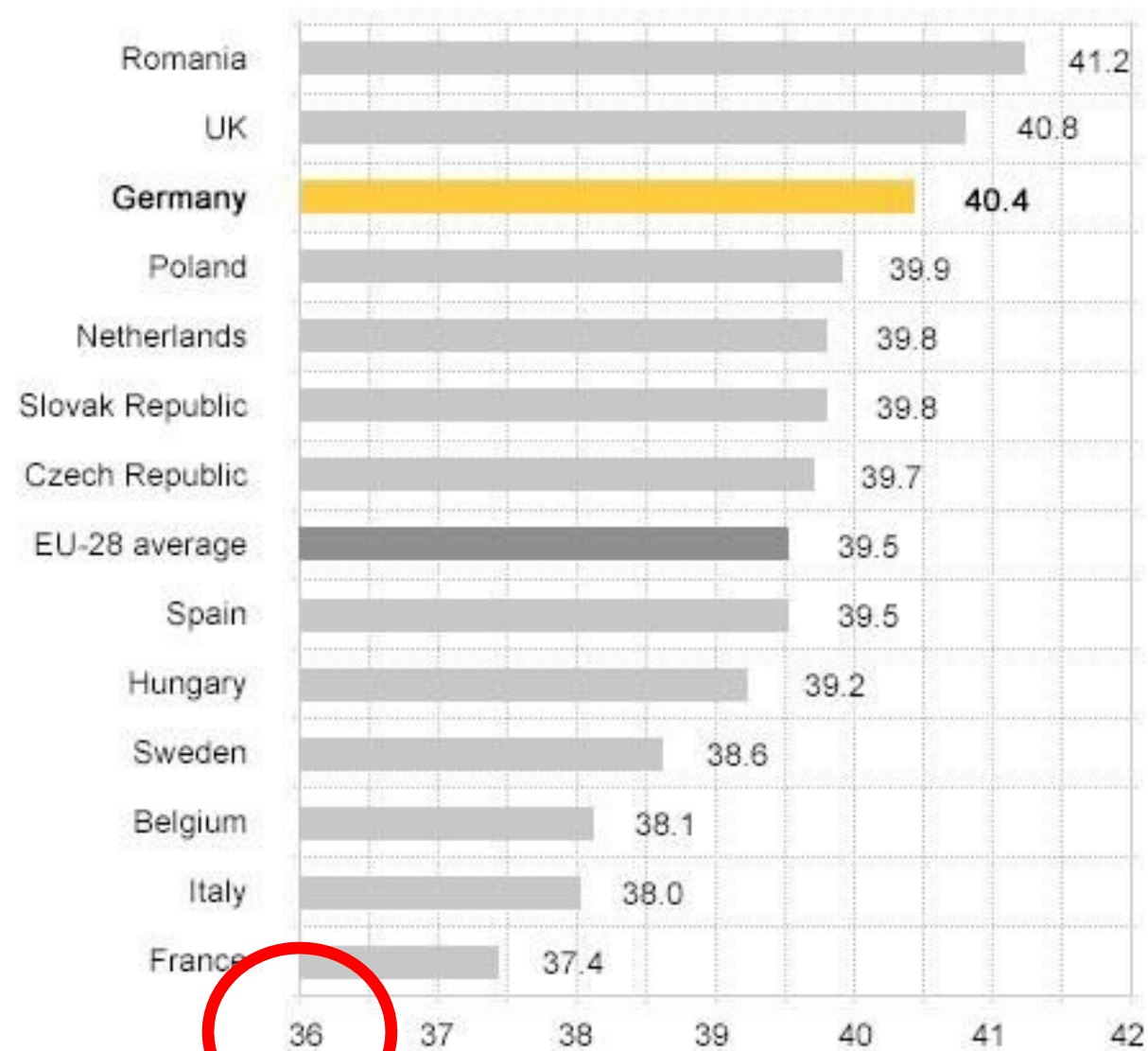
Scale Distortion

- Begin bar graphs from zero



Source: <https://www.moneycontrol.com/news/trends/features/5-costliest-mistakes-stock-market-investors-make-because-of-misleading-graphs-10852961.html>

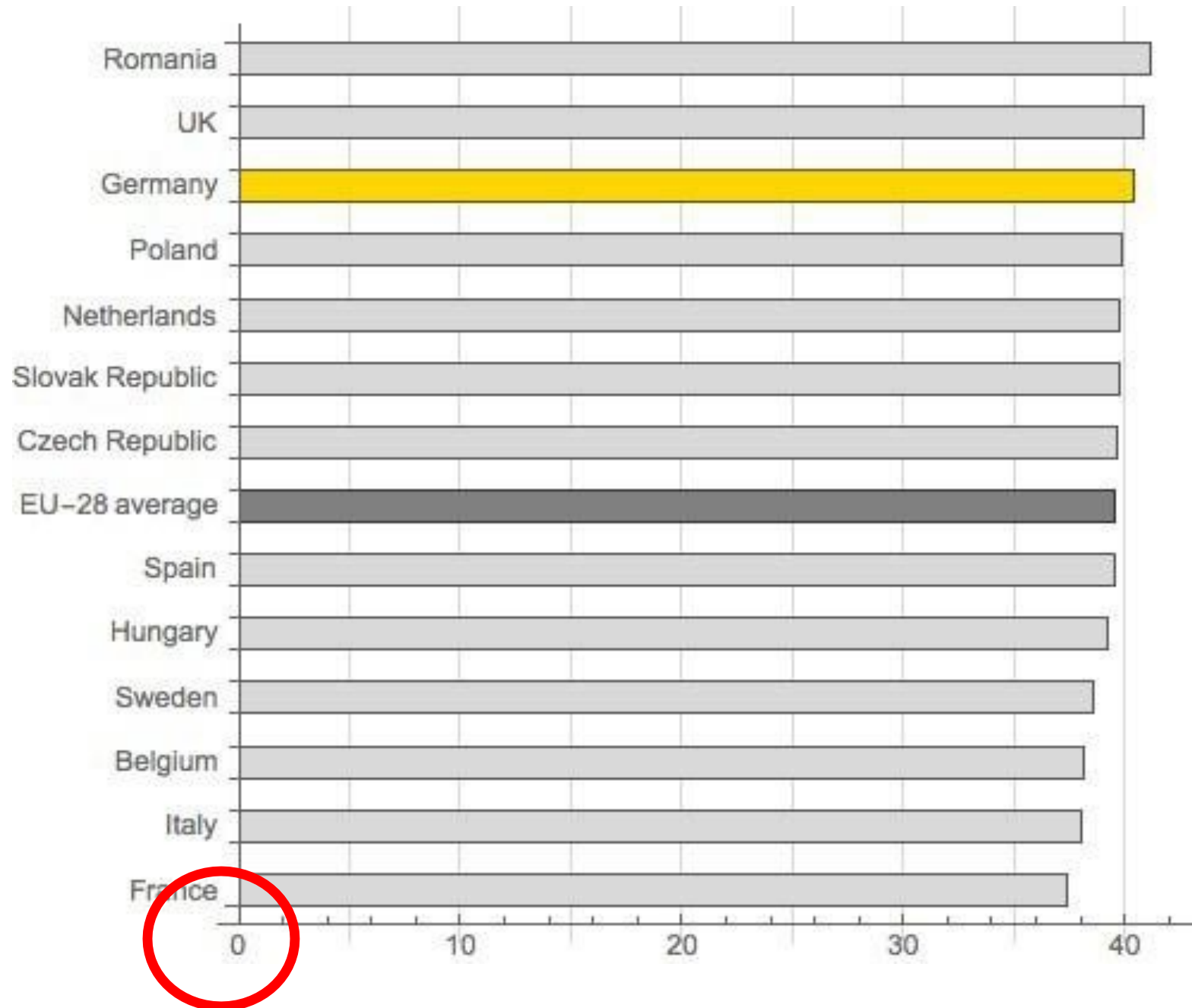
Average number of actual weekly hours of work in main job, full-time employees, 2013



Source: Eurofound 2014

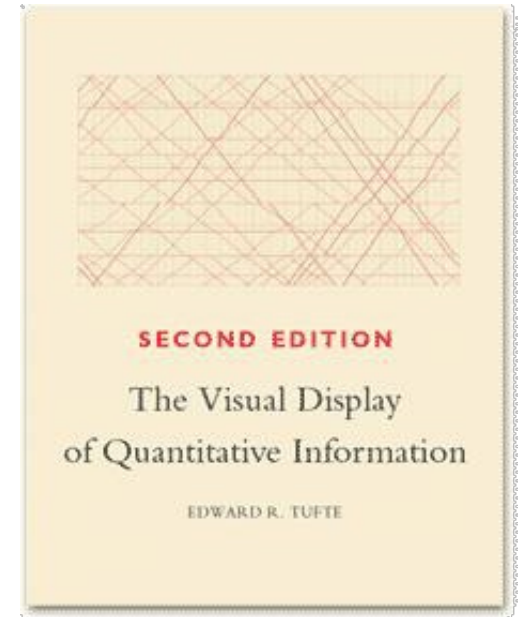
<https://tinyurl.com/y86tuvro>

Correct Version

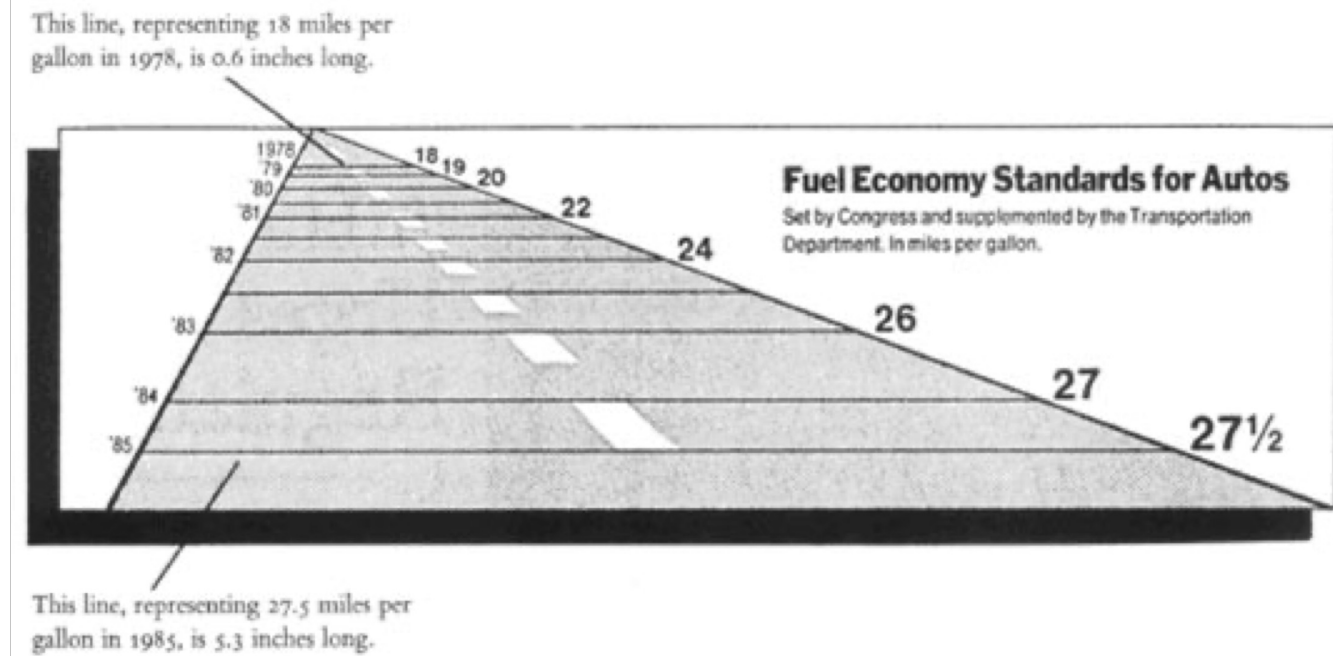
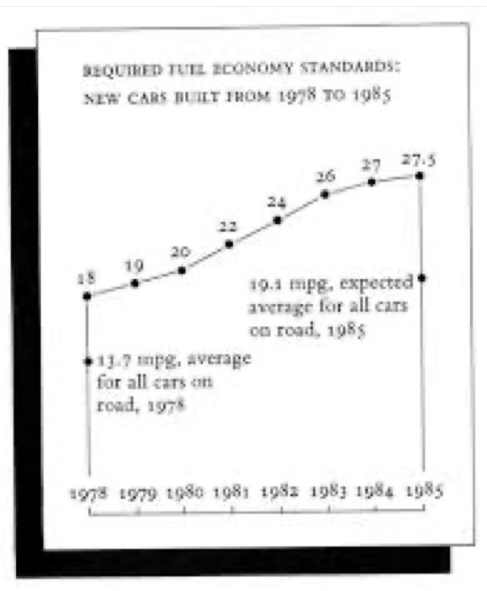


Edward Tufte, *The Visual Display of Quantitative Information*, 2001

*“The representation of numbers, as physically measured on the surface of the graphic itself, should be directly **proportional to** the numerical quantities represented.”*

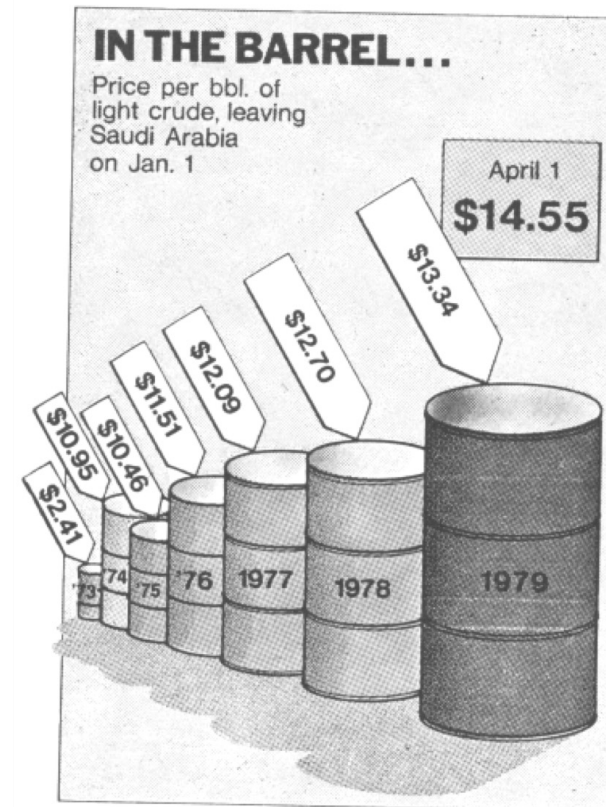
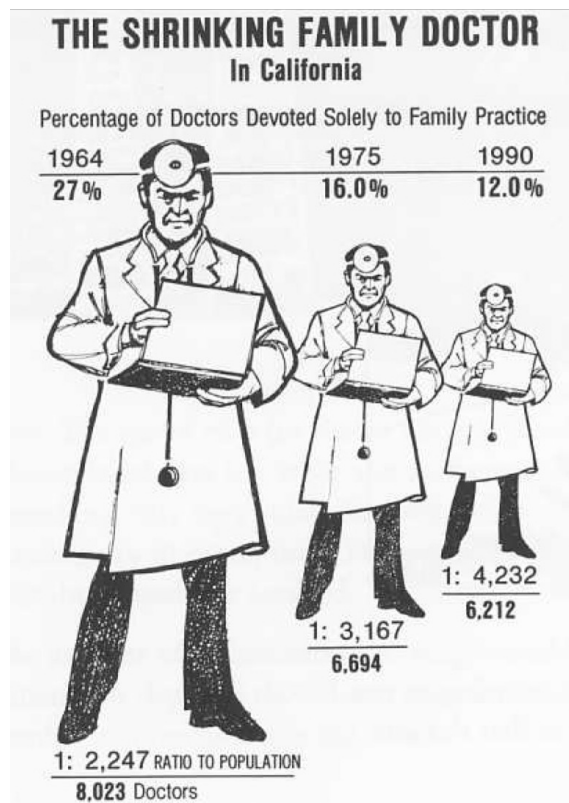


Are they the representation of the same data?



The Lie Factor: Dimensionality

Area/volume increase non-proportionally to length.



It's Who You Know

By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these **personal social networks** begin to decline over time.

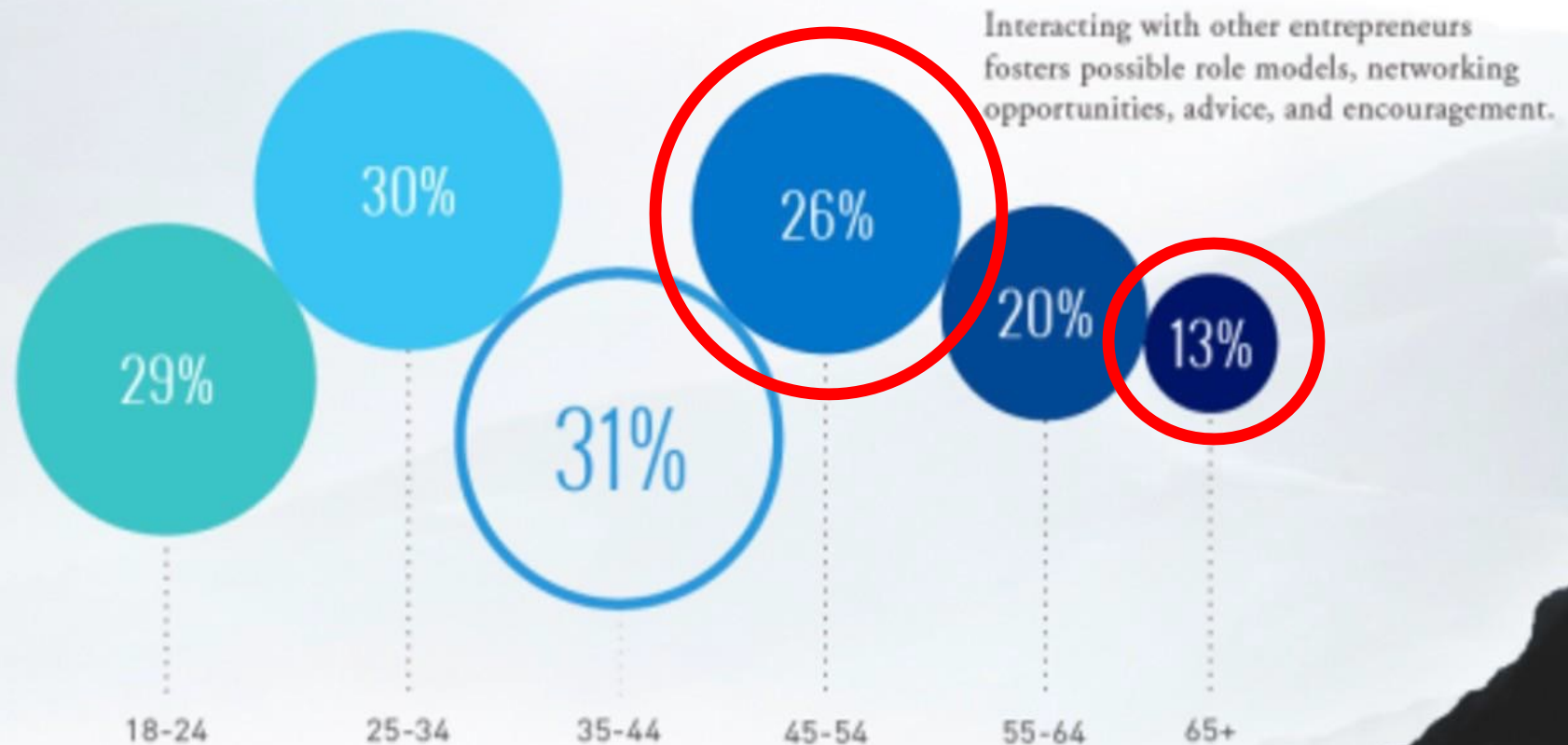
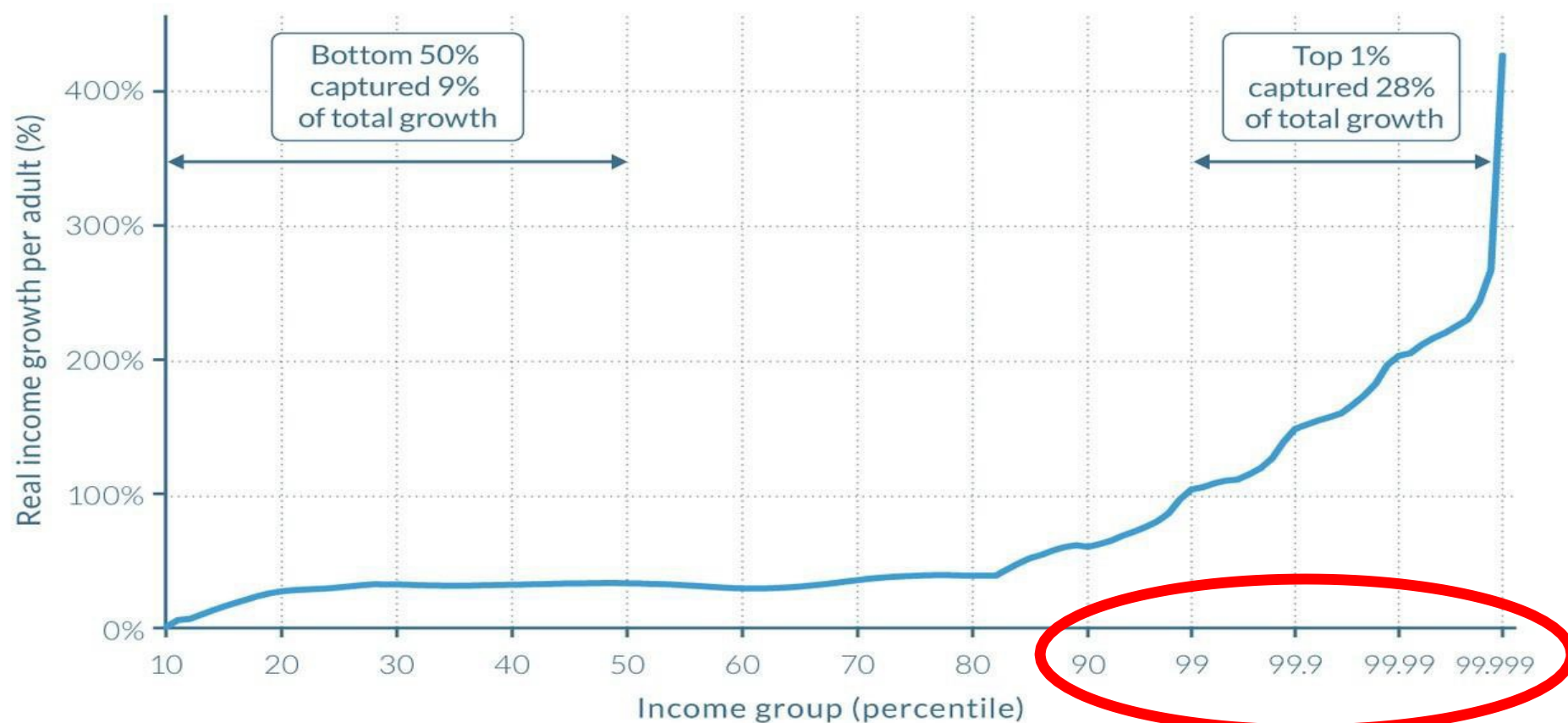


Figure 2.1.2

Total income growth by percentile in US-Canada and Western Europe, 1980–2016

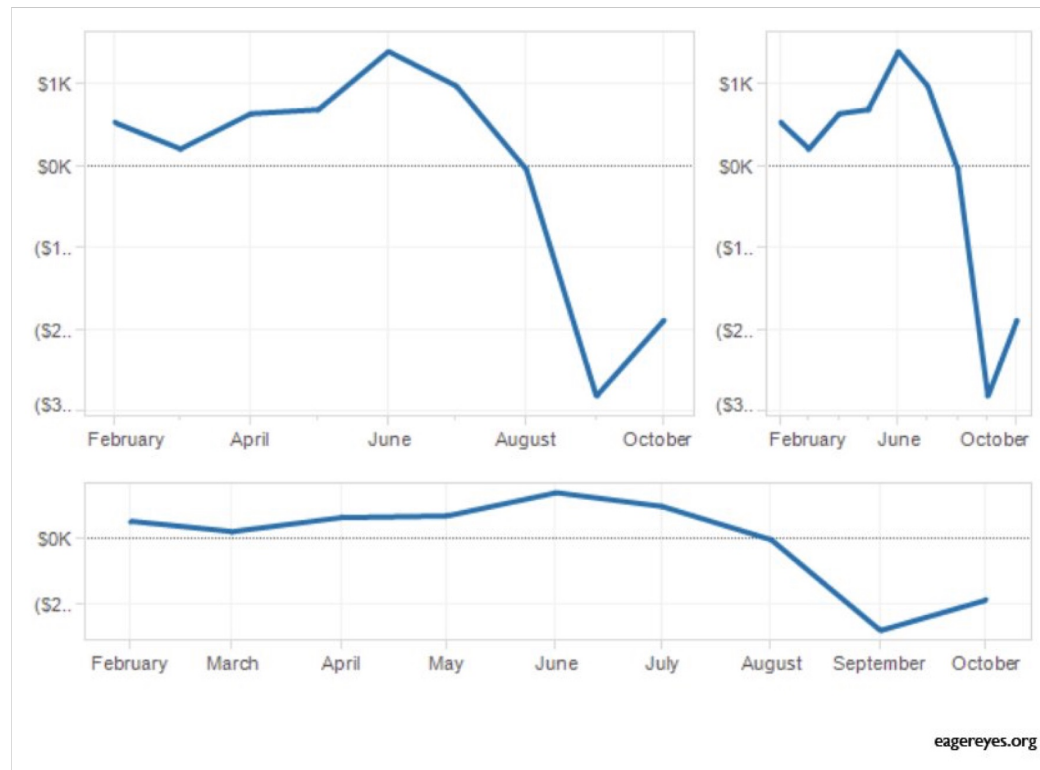


Source: WID.world (2017). See wir2018.wid.world/methodology.html for data series and notes.

On the horizontal axis, the world population is divided into a hundred groups of equal population size and sorted in ascending order from left to right, according to each group's income level. The Top 1% group is divided into ten groups, the richest of these groups is also divided into ten groups, and the very top group is again divided into ten groups of equal population size. The vertical axis shows the total income growth of an average individual in each group between 1980 and 2016. For percentile group p99p99.1 (the poorest 10% among the world's richest 1%) growth was 104% between 1980 and 2016. The Top 1% captured 28% of total growth over this period. Income estimates account for differences in the cost of living between countries. Values are net of inflation.

Aspect Ratios and Lie Factors

- The steepness of cliffs corresponds to the aspect ratio.

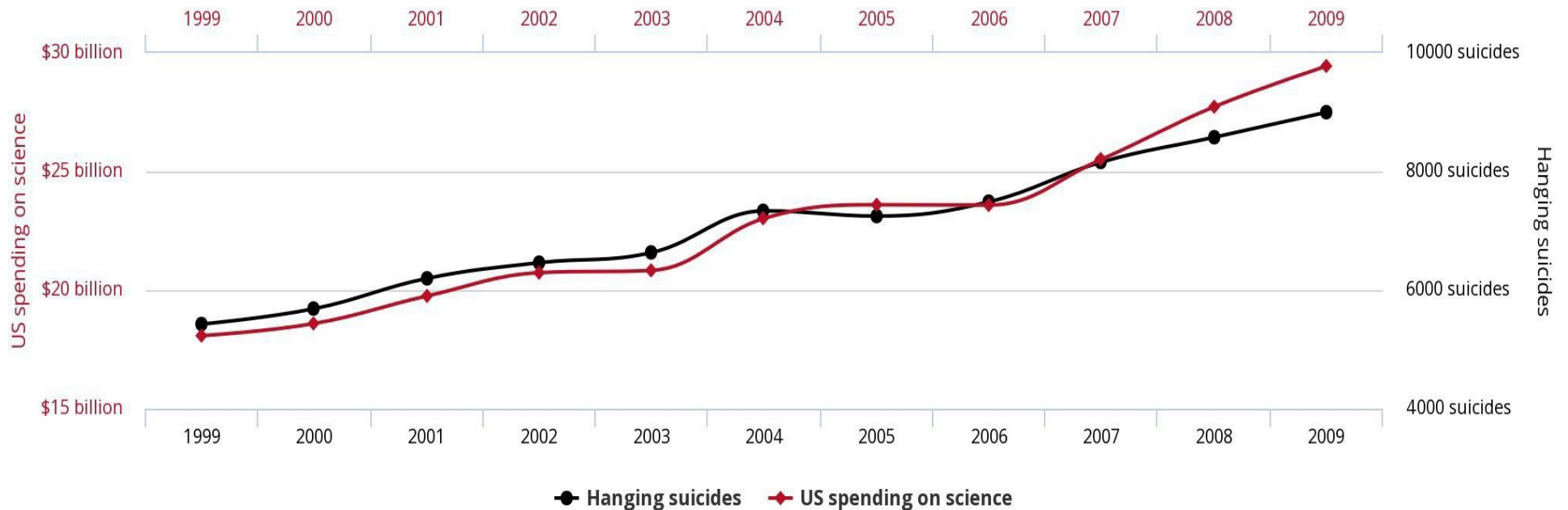


Agenda

- Visualization Tricks
- Correlation vs. Causation

Suicide vs. STEM \$

US spending on science, space, and technology
correlates with
Suicides by hanging, strangulation and suffocation



tylervigen.com

Correlation: 99.79% (r=0.99789126)

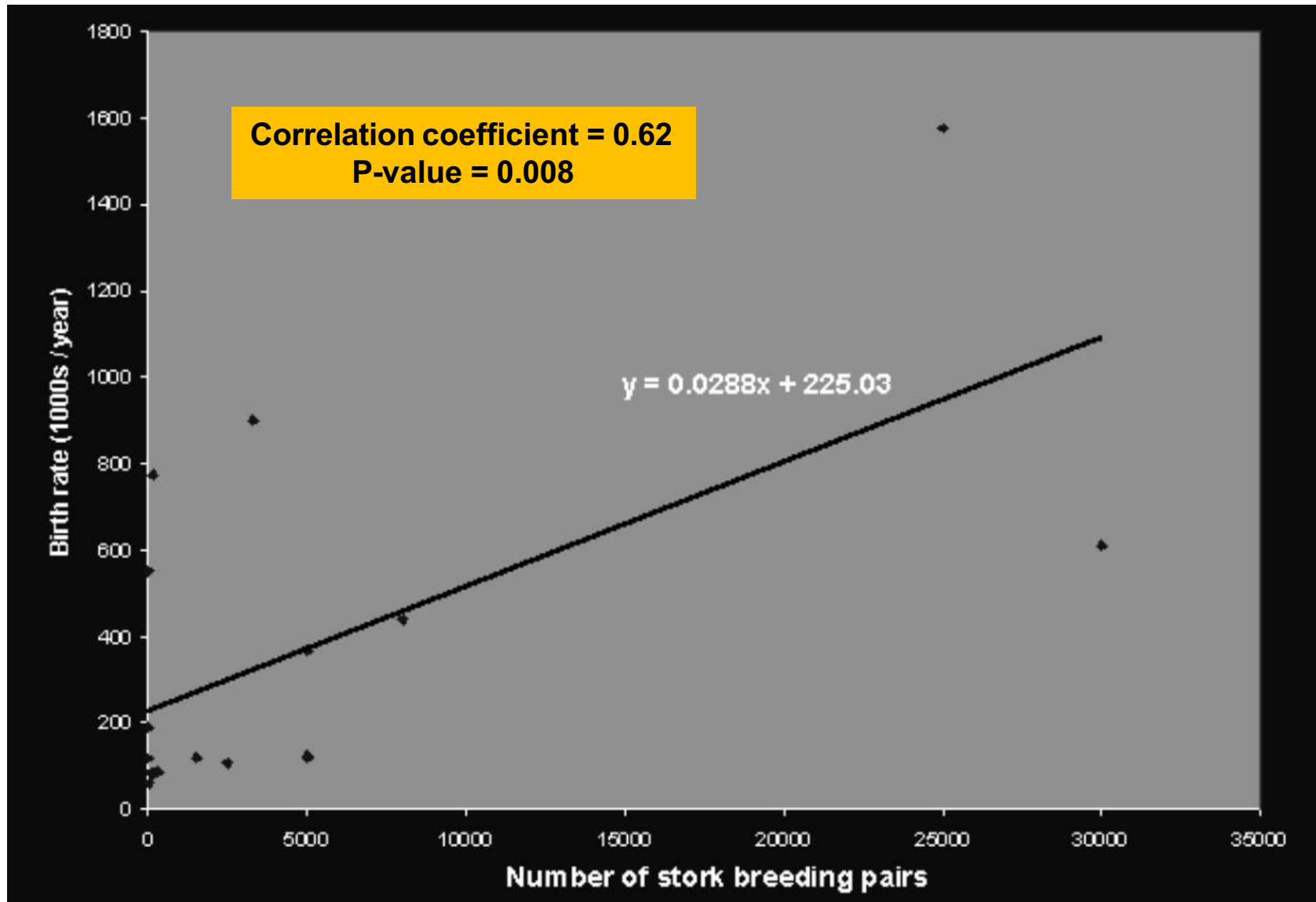
|Storks| vs. |Babies|



Robert Matthews,
Teaching Statistics,
2000

Country	Area (km ²)	Storks (pairs)	Humans (10 ⁶)	Birth rate (10 ³ /yr)
Albania	28,750	100	3.2	83
Austria	83,860	300	7.6	87
Belgium	30,520	1	9.9	118
Bulgaria	111,000	5000	9.0	117
Denmark	43,100	9	5.1	59
France	544,000	140	56	774
Germany	357,000	3300	78	901
Greece	132,000	2500	10	106
Holland	41,900	4	15	188
Hungary	93,000	5000	11	124
Italy	301,280	5	57	551
Poland	312,680	30,000	38	610
Portugal	92,390	1500	10	120
Romania	237,500	5000	23	367
Spain	504,750	8000	39	439
Switzerland	41,290	150	6.7	82
Turkey	779,450	25,000	56	1576

|Storks| vs. |Babies|



Definition

- **Correlation**

- Two processes or events are observed together
- X increases → The increase/decrease of Y is observed
- Eg,
 - high school students with high grades also had high SAT scores
 - Students with more absences tend to have lower grades

- **Causation**

- When one process “causes” another process as an “effect”
- Cause → Effect
- Eg,
 - If I pinch you, you get pain
 - Brain damage causes mental illness
- Often very difficult to establish, if not impossible

A Case Study

- Eating breakfast has long been believed to be correlated with success in school for elementary school children
- Q: Does eating breakfast **cause** elementary students to be better learners?