

MPDM Guest Lecture

Seb Bate

Intro



Who am I?

- Medical Statistician at Manchester hospitals
- Data visualisation hobbyist

What are we doing today?

1. Menti
2. Guess the data
3. Infographics
4. Simpson's paradox

Join at menti.com | use code 8502 6033

 Mentimeter

Instructions

Go to

www.menti.com

Enter the code

8502 6033



Or use QR code

Warm up activity

- Pick a digital touchpoint that you have hit in the last 24h
- Now think about the hundreds, thousands, millions of hits that your touchpoint has received
- What data has been collected?
- What data can be analysed and how would you analyse it?
- What actions could you take as a marketeer?
- Some bits of data are better quality than others
- Feel free to work quietly in small groups

E.g. Buying a coffee on the Caffe Nero app

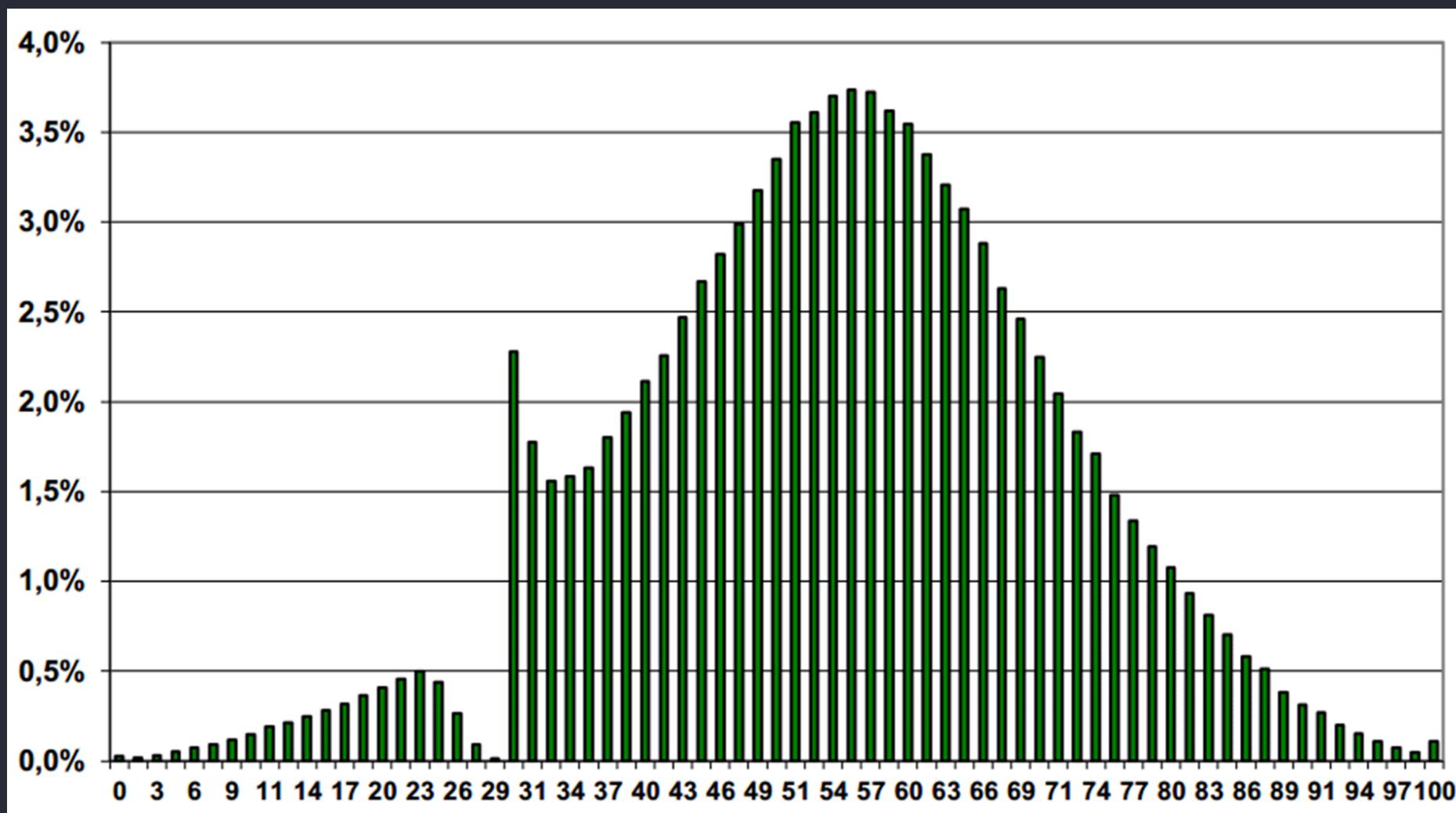
- Time/date
- Location
- Order
- Addons
- Payment method
- Vouchers used
- Click-and-collect or pay at counter
- Sit in or takeaway
- Cart abandonment

Actions that can be taken?

- Giving vouchers for lapsed customers
- More prominent up sell if I never get add-ons
- Push notification if there is a quieter shop nearby

- Pick a digital touchpoint that you have hit in the last 24h
- Now think about the hundreds, thousands, millions of hits that your touchpoint has received
- What data has been collected?
- What data can be analysed and how would you analyse it?
- What actions could you take as a marketeer?
- Some bits of data are better quality than others
- Feel free to work quietly in small groups

Guess the data



YOU WILL READ THIS FIRST.

At some point you may come back to read this line or maybe not.

And then you will read this line next.

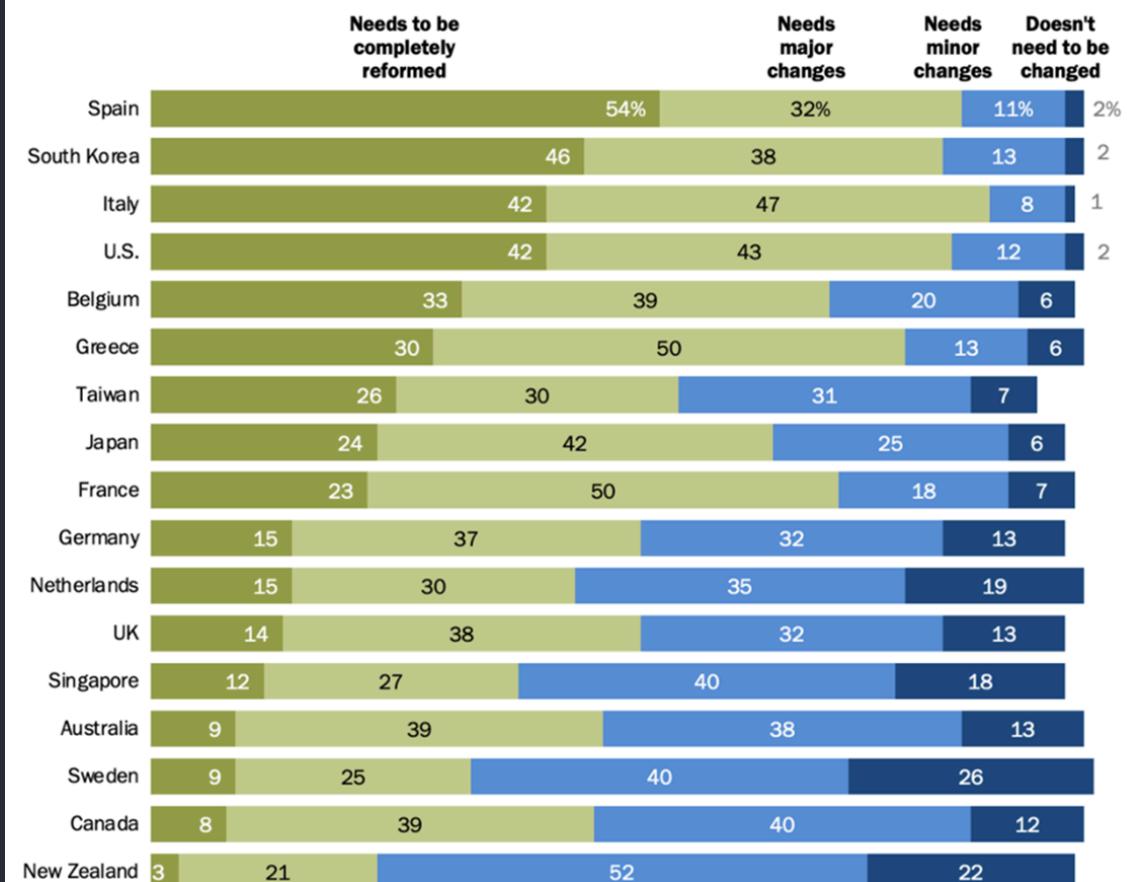
You will go back to read this body copy if you want to know more. It takes the most effort to read because it has a lot of text in a small font in a light weight with tight line spacing. Many people will skip paragraphs like this unless if they aren't engaged right away. This is why it's important to draw attention to your message using visual hierarchy.

**You'll probably
read this before
the paragraph.**

Some best practice

Large shares in many publics say their political system needs reform

% who say the political system in (survey public) ...



Note: Those who did not answer not shown.

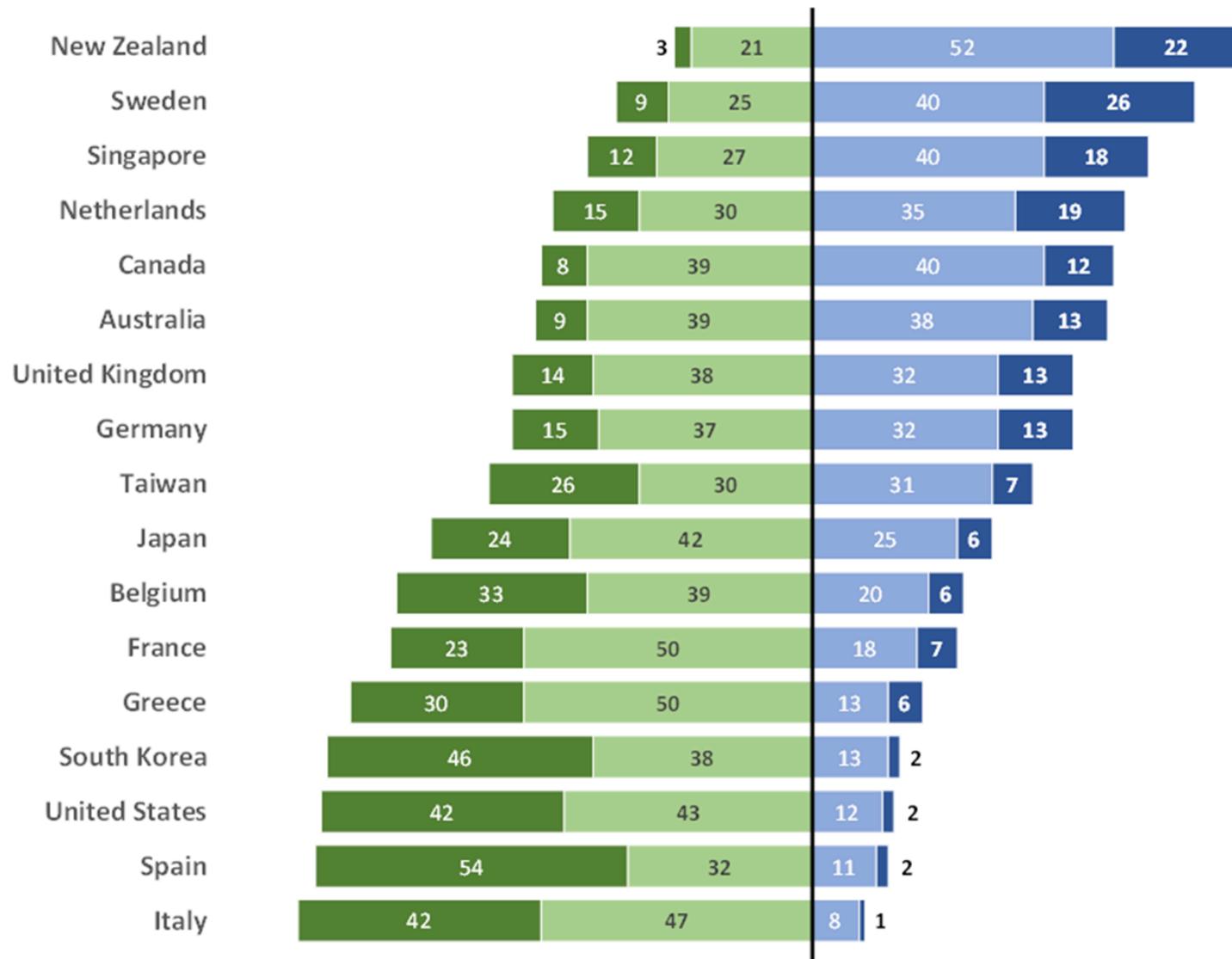
Source: Spring 2021 Global Attitudes Survey, Q13c.

"Citizens in Advanced Economies Want Significant Changes to Their Political Systems"

PEW RESEARCH CENTER

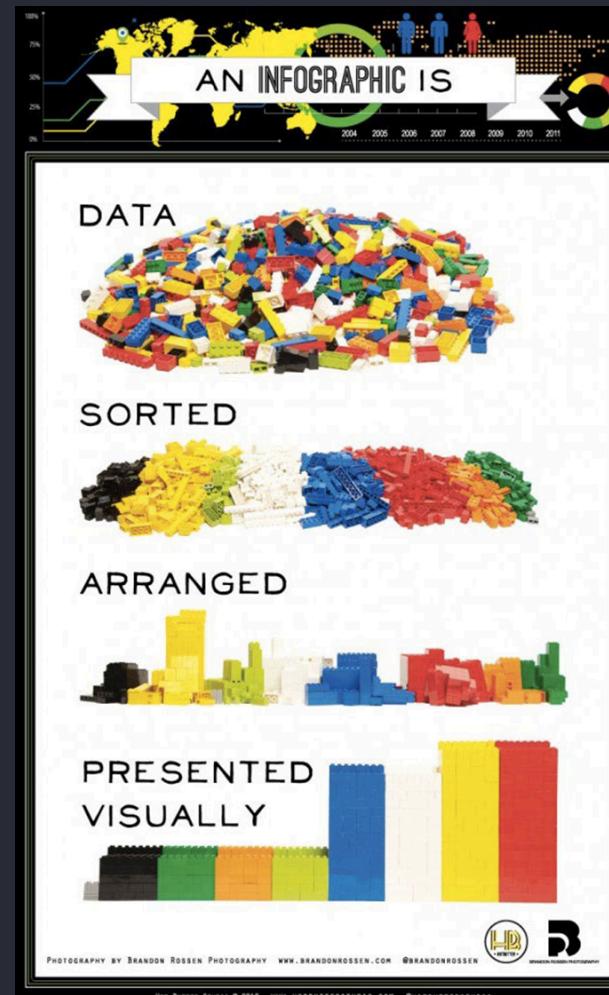
% by country who say the political system...

■ Needs to be completely reformed ■ Needs major changes ■ Needs minor changes ■ Doesn't need to be changed

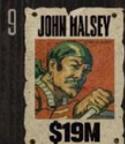
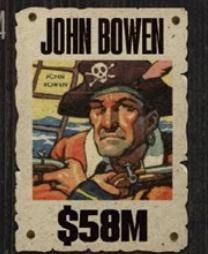
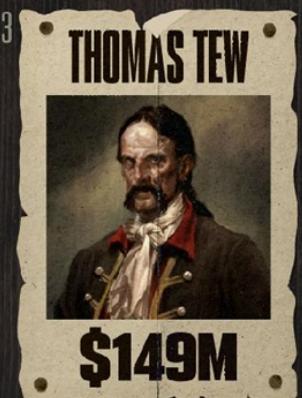
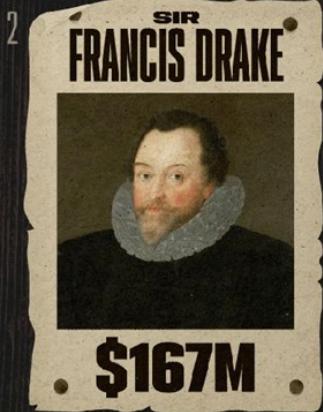


Infographics

- Infographics are more editorialized plots (or combination of plots), often featuring more text and selective labels.
 - Some are more data driven than others
- Most people don't like being told what to think and like to draw their own conclusion
- But this is where you are able to tell the story and at the end make a call to action
- Always be critical



THE WORLD'S RICHEST PIRATES



Avg American's net worth in 2022
\$1.1M

Pirate wealth was adjusted for inflation (March 2024)

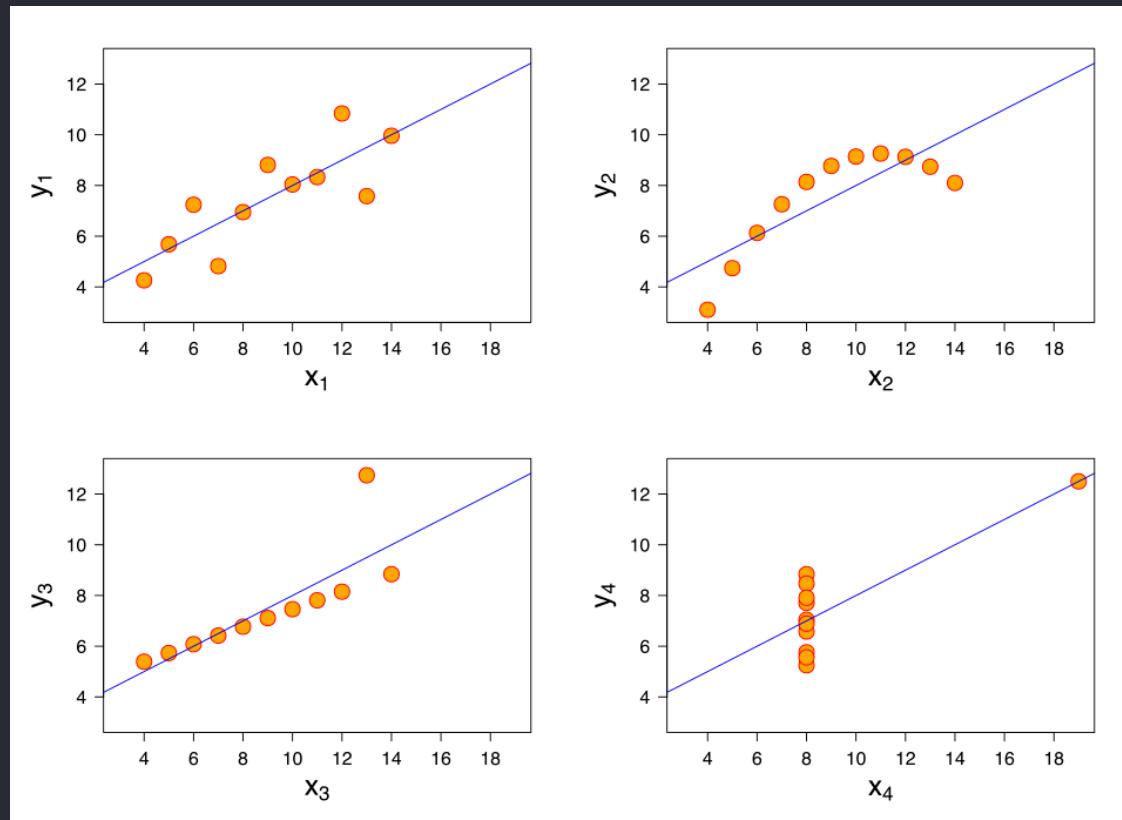
Source: Forbes, U.S. News



Some interesting data

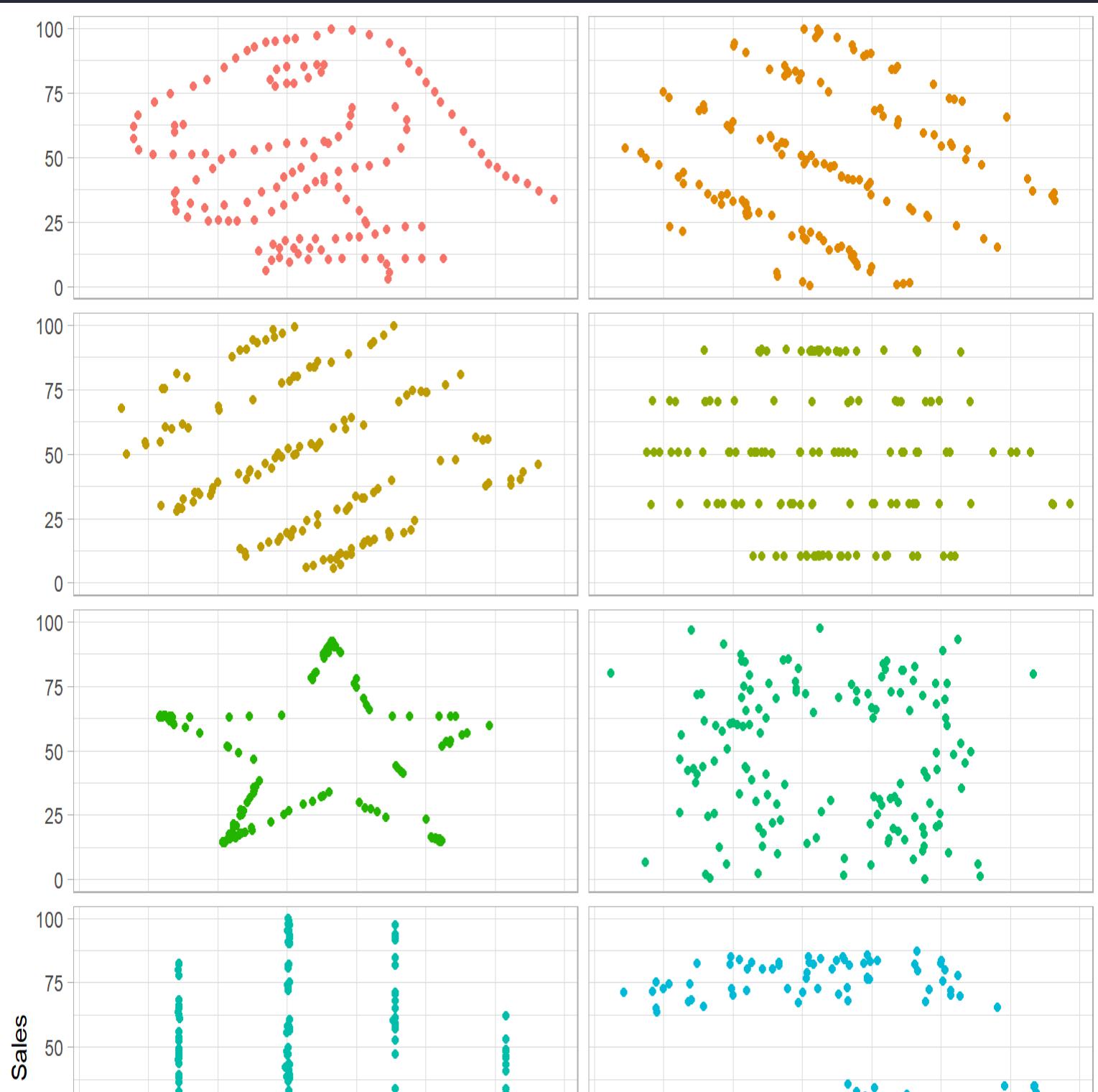
I think I have seen something like this before?

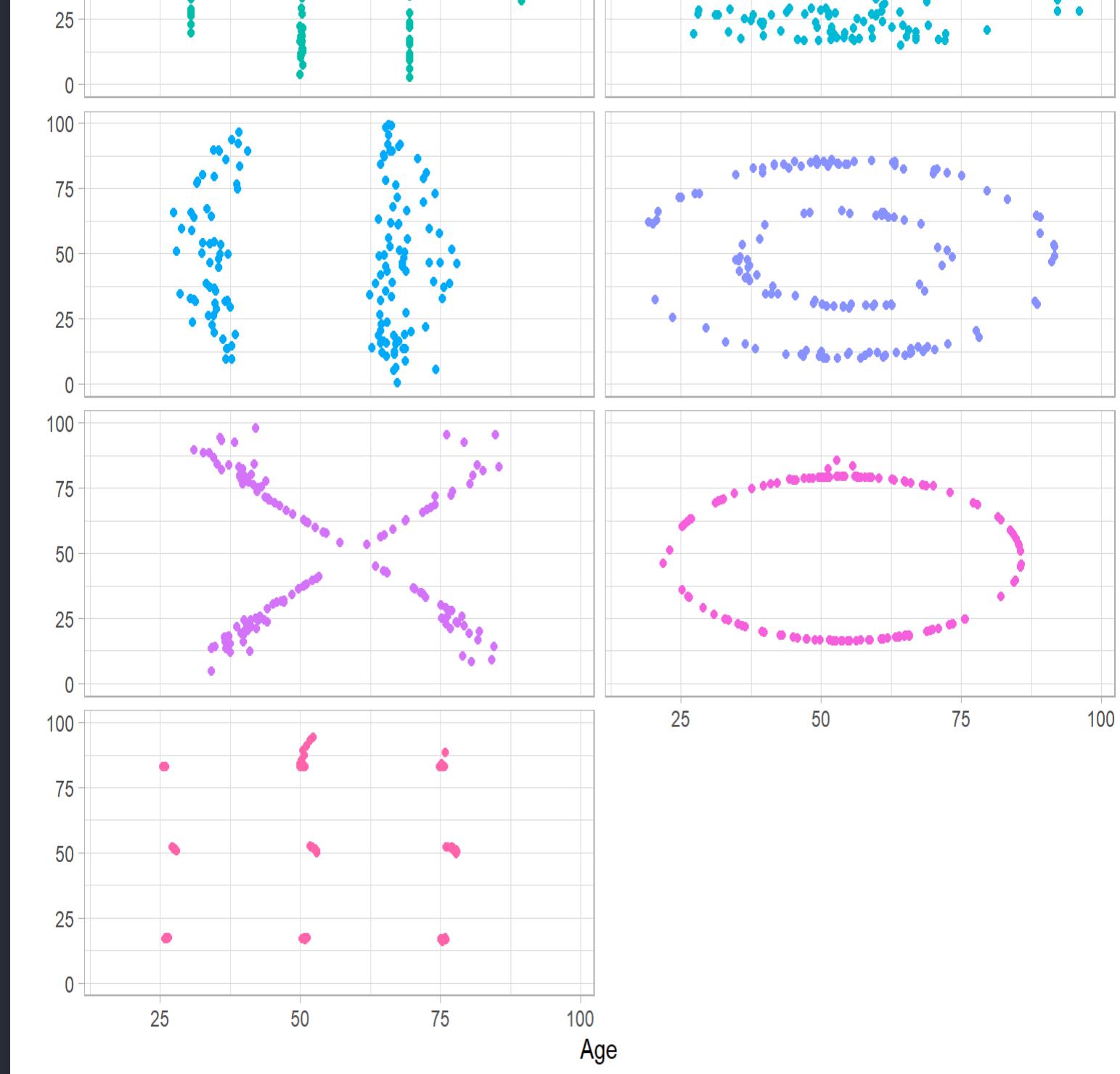
You may have seen Anscombe's Quartet yesterday



Anscombe's Quartet (Anscombe, 1973) image from Wiki (Wikipedia contributors, 2025)

But we can actually extend this idea
to a much larger dataset





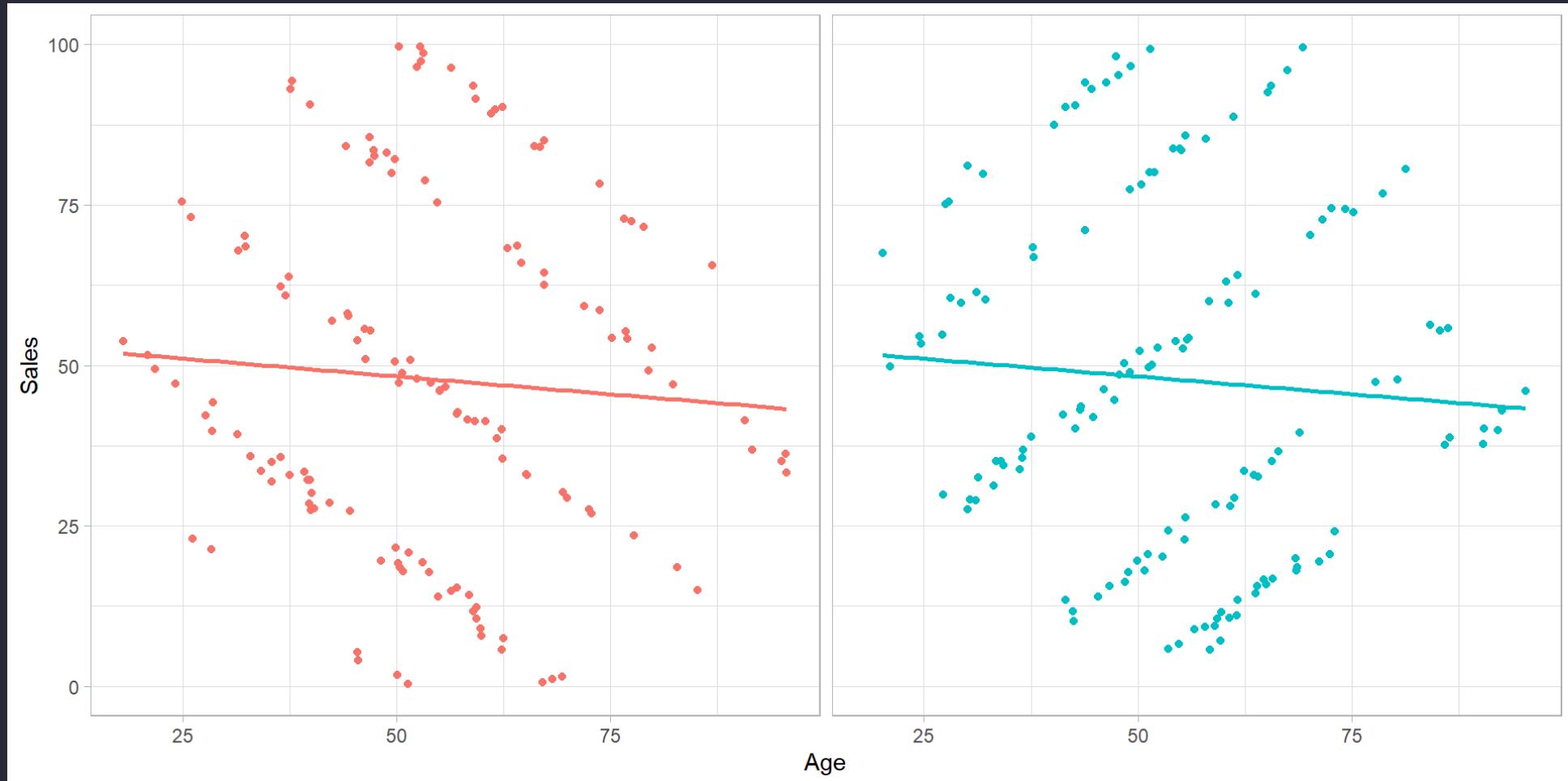
In all of these they have the same properties

- N: 142
- Mean Age: 54.26
- Mean Sales: 47.83
- SD Age: 16.76
- SD Sales: 26.93
- Correlation: -0.06
- Adapted from Gillespie *et al.* (2025)

Simpson's Paradox



What do we notice here?



These are dataset 2 and 3 from the previous slide

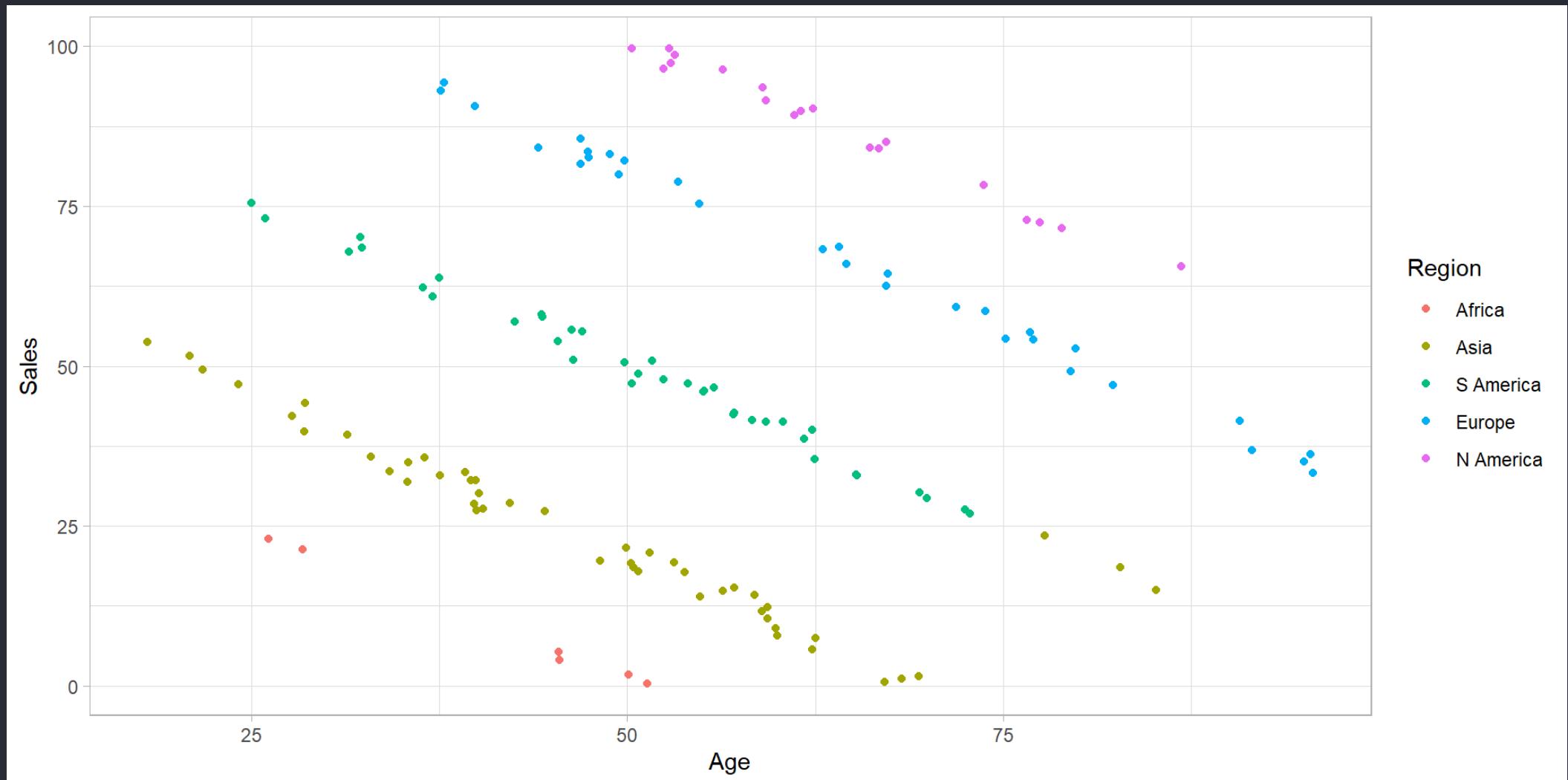
But what happens if we now add labels to the data?

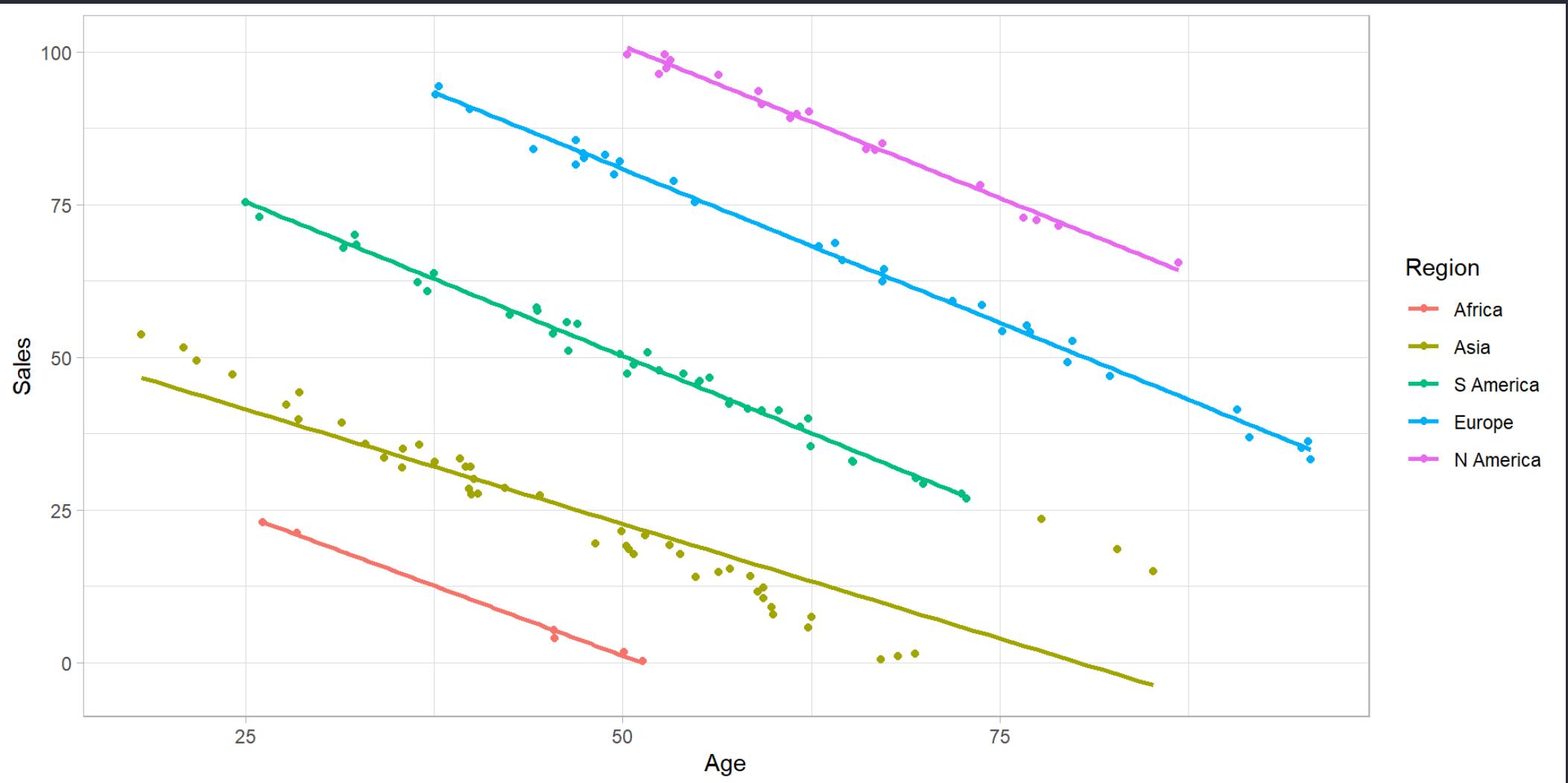
Dataset 2

Our dataset

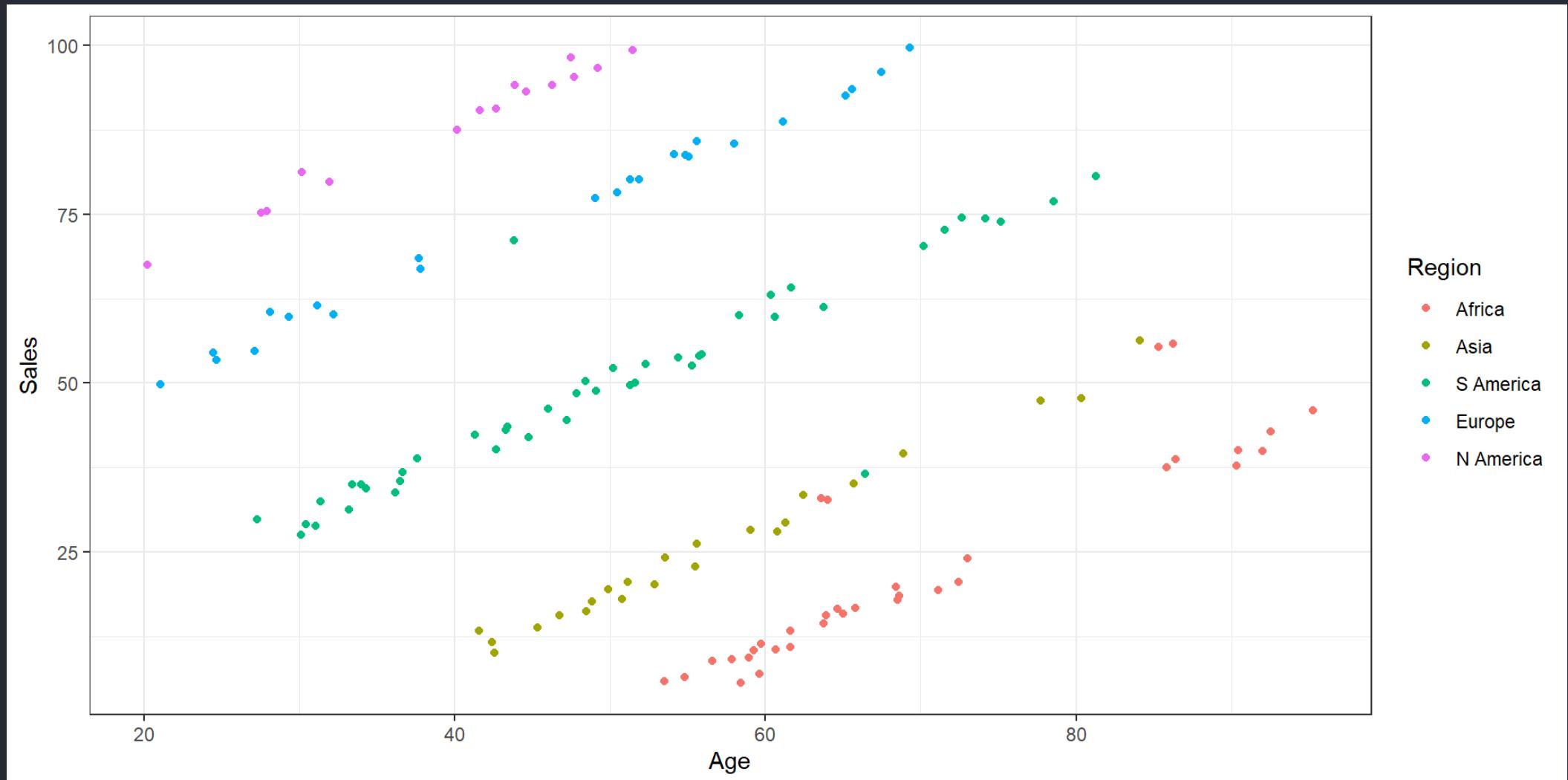
	Age	Sales	Region
1	54.86350	14.009612	Asia
2	59.31099	10.572136	Asia
3	40.00266	27.487510	Asia
4	36.43472	62.294045	S America
5	59.97042	7.913310	Asia
6	55.74124	46.653577	S America
7	50.74172	17.838406	Asia
8	67.30258	64.413463	Europe
9	62.30775	5.749975	Asia
10	46.90954	85.602261	Europe

[1] "Africa" "Asia" "S America" "Europe" "N America"

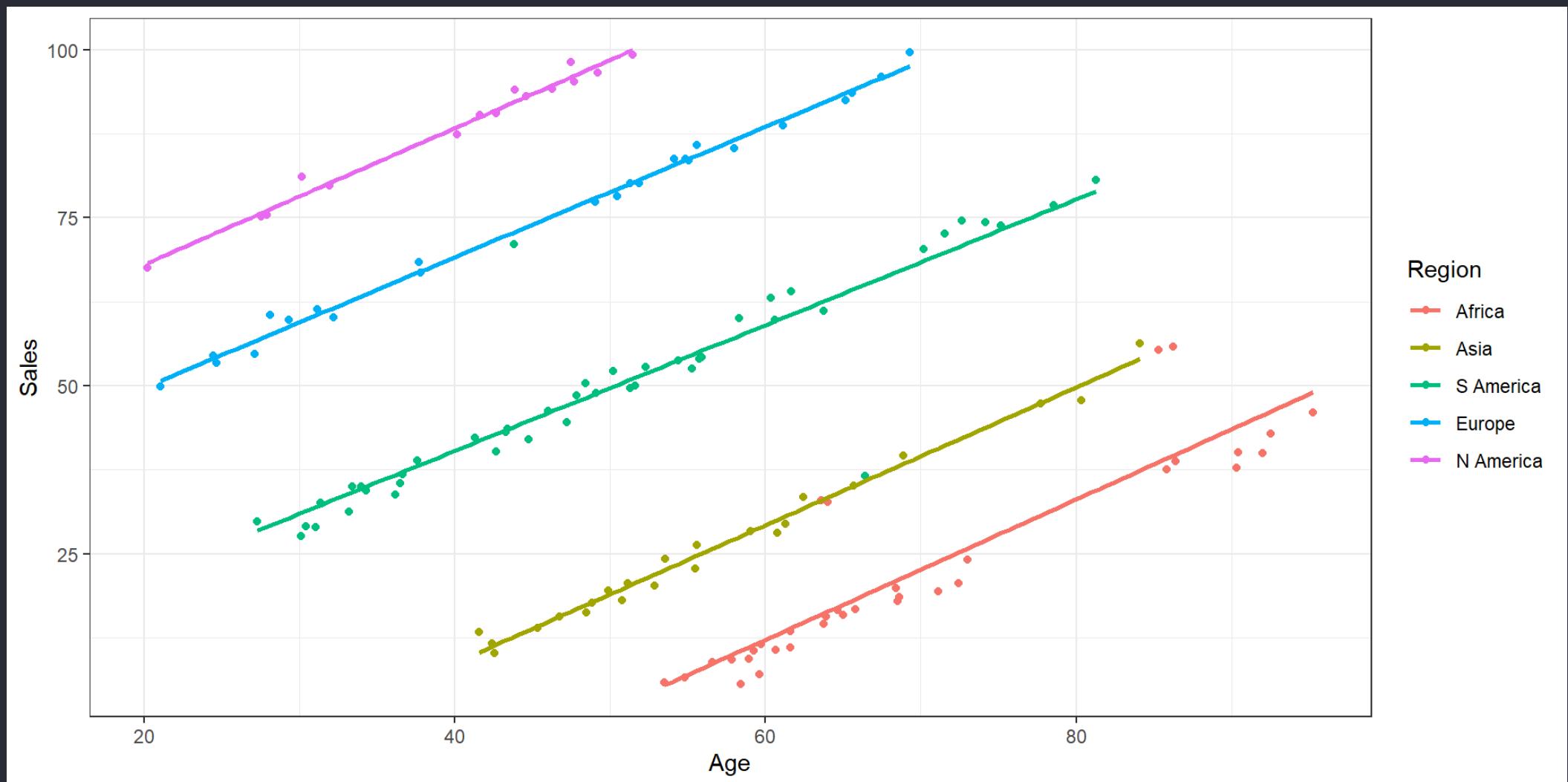




and similarly for dataset 3



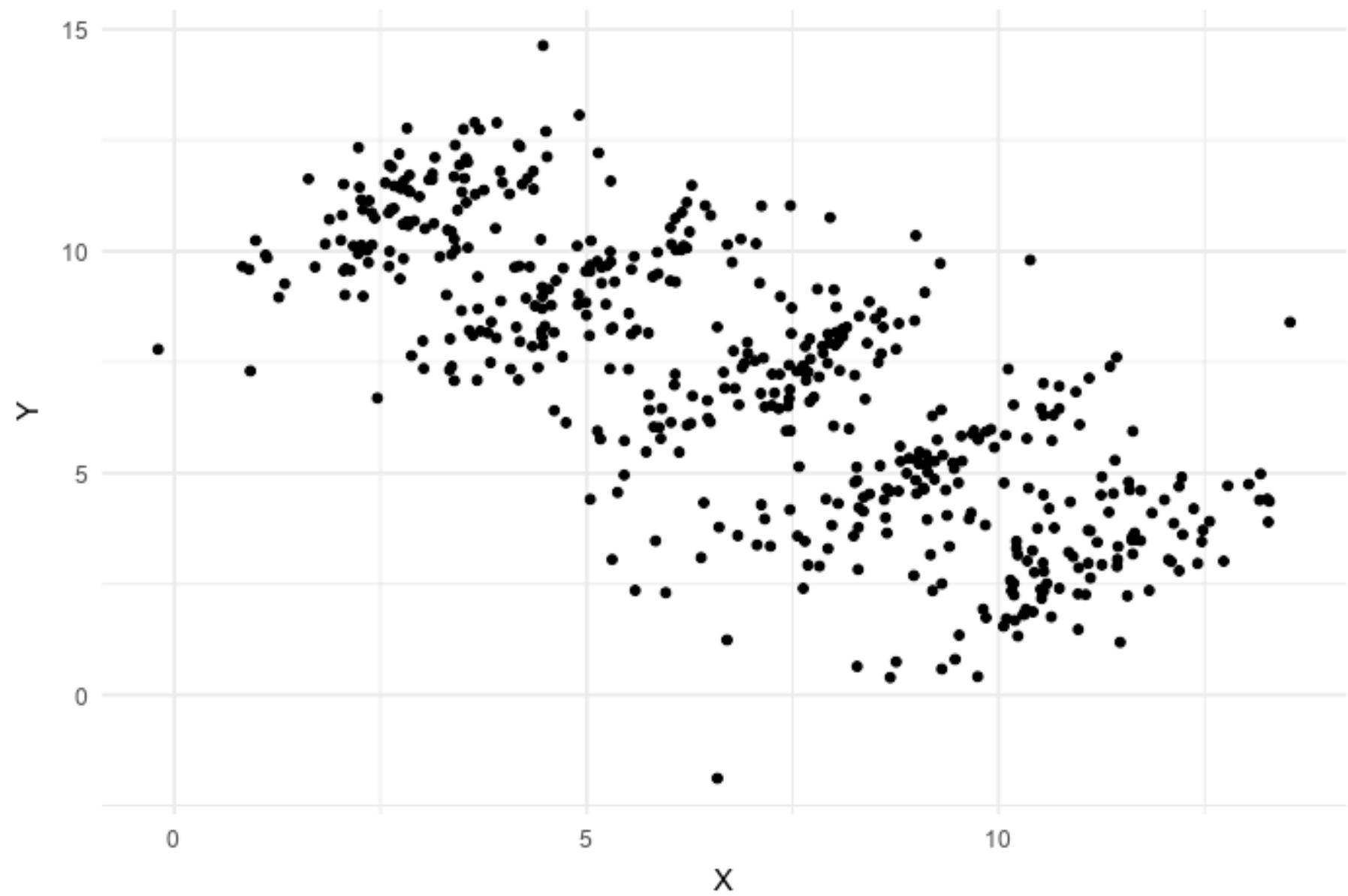
and similarly for dataset 3





- This is what we call Simpson's Paradox (Simpson, 1951; Kievit *et al.*, 2013).
- It is when the direction of the correlation changes when we look at each subgroup

Korrelation:



Crimes against data: a quiz

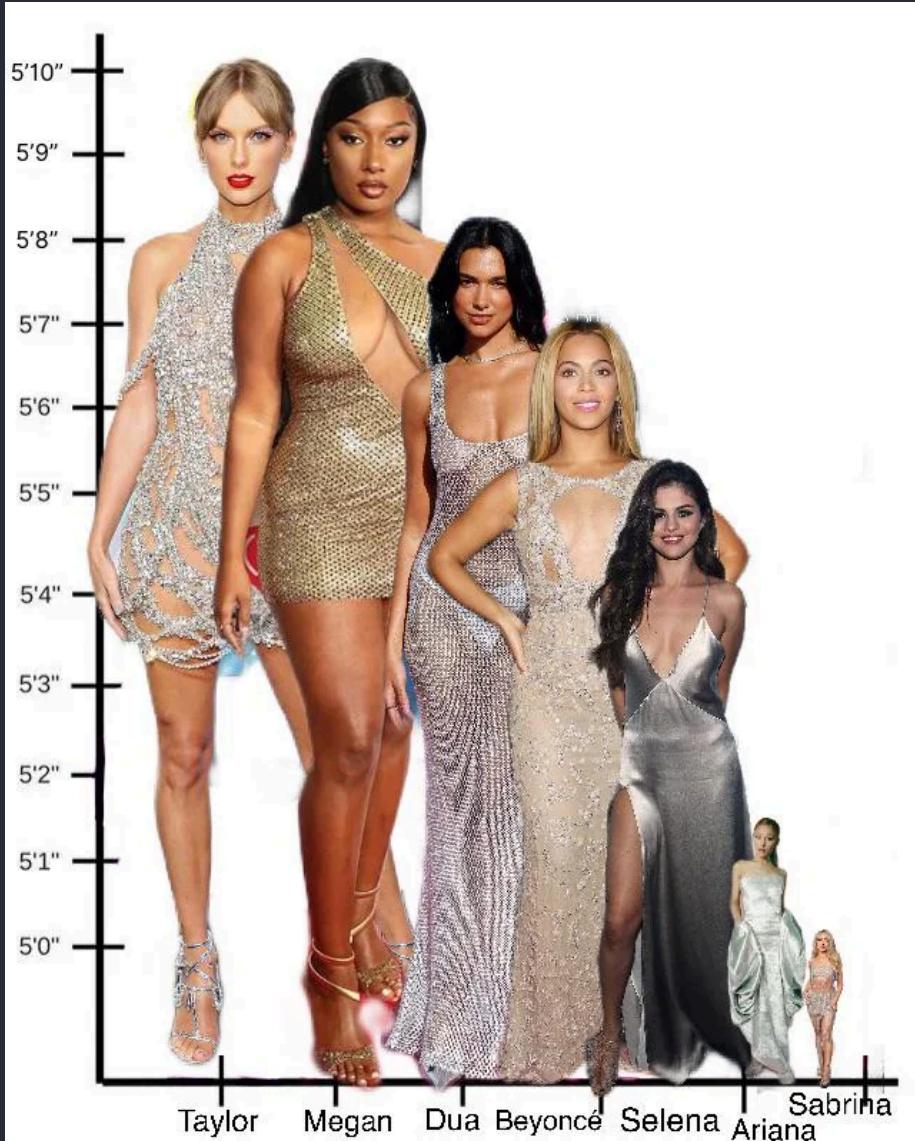


Q12

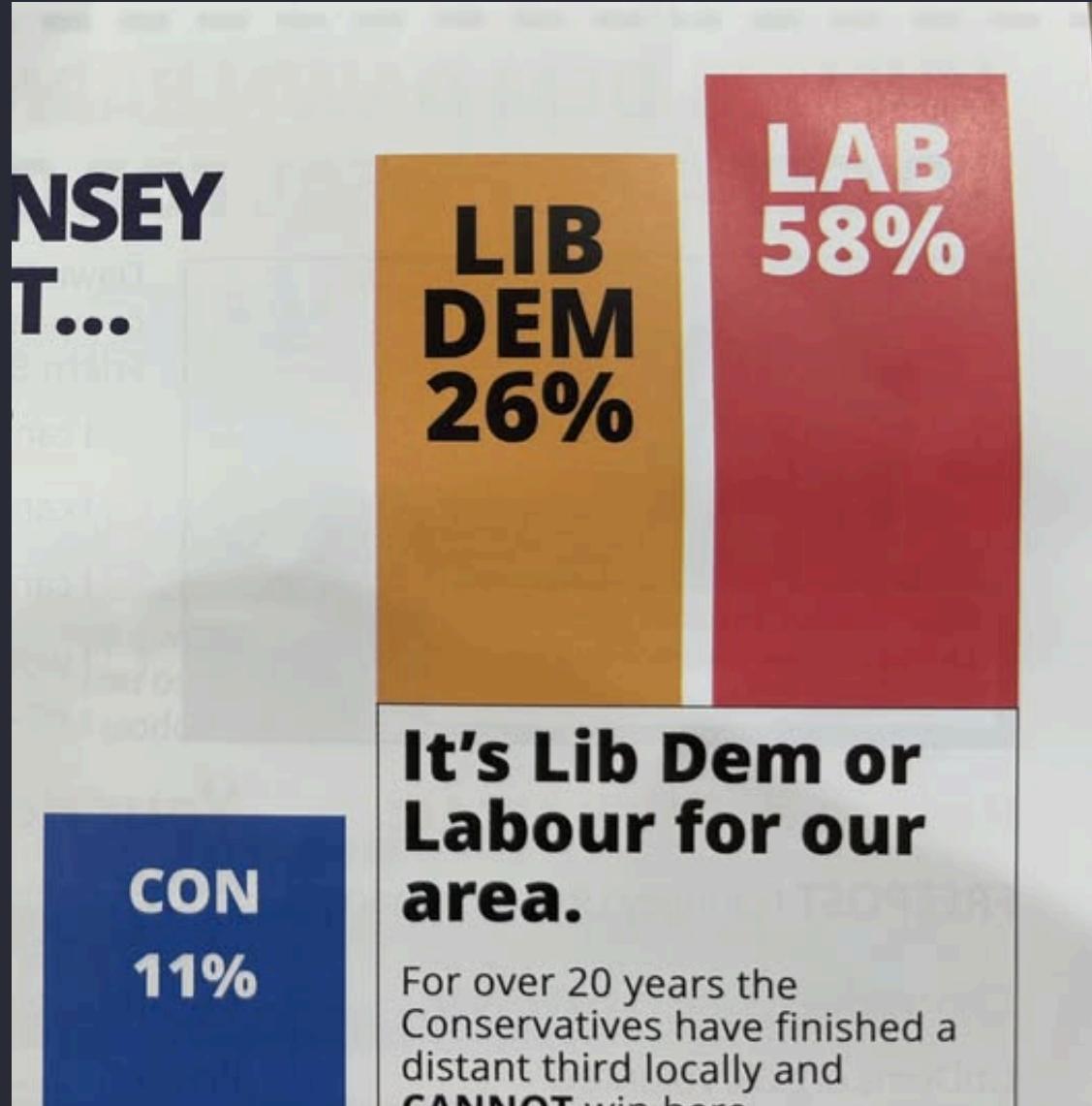
What gender do you identify with?

- Strongly disagree
- Somewhat disagree
- Neither agree nor disagree
- Somewhat agree
- Strongly agree

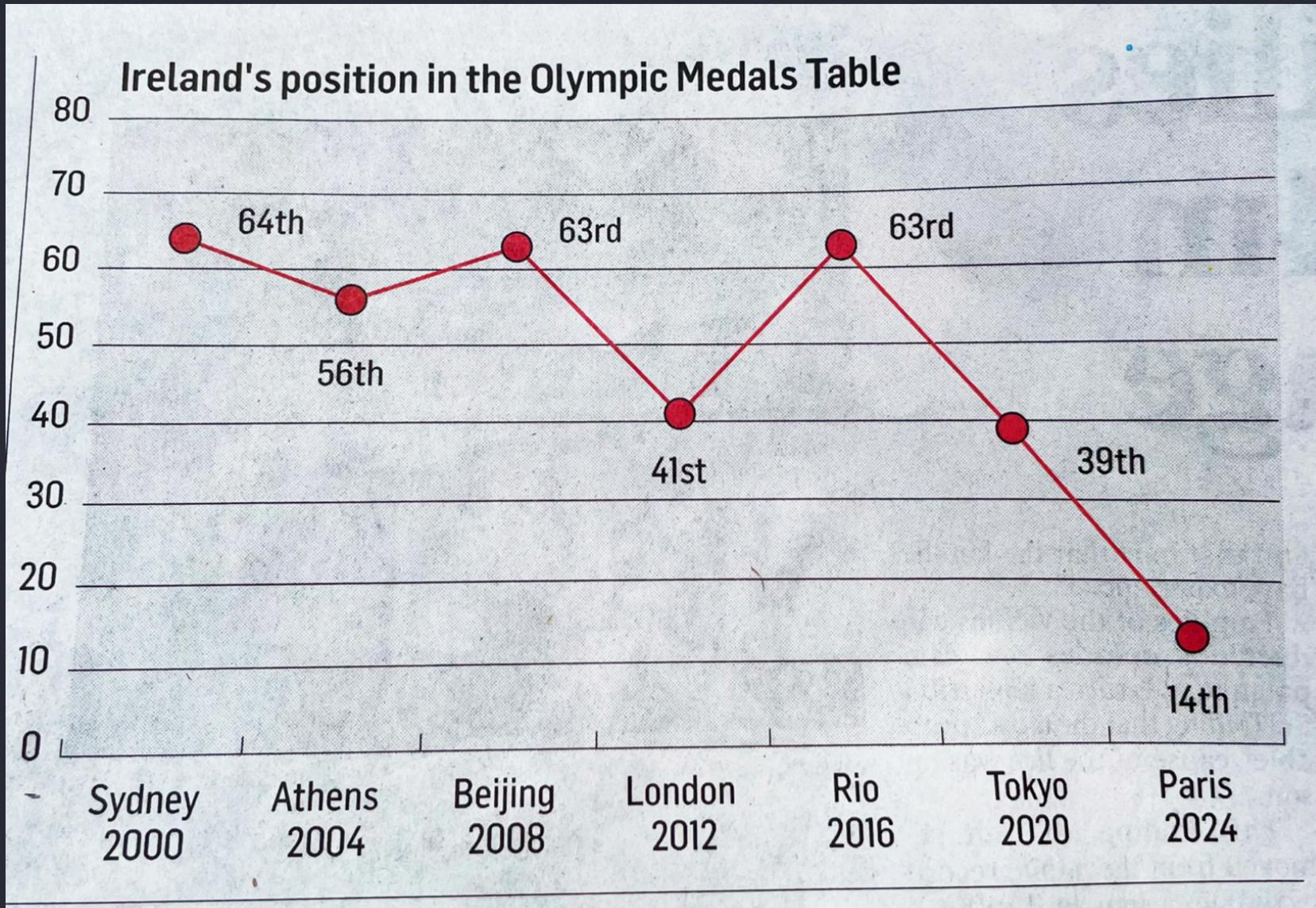
Make sure that your inputs make sense



Don't truncate your y-axis on a bar chart

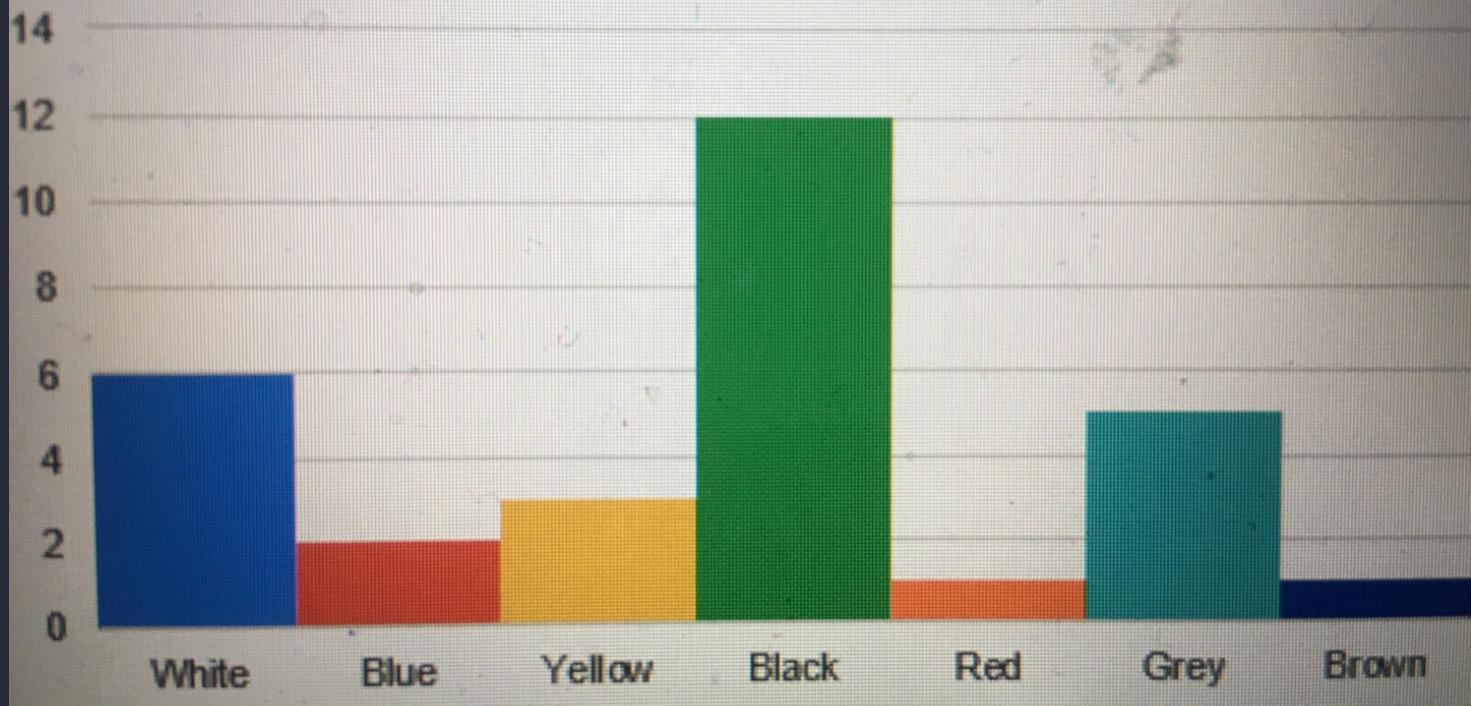


Make your bars proportional



Good should be at the top!!

Shoe color
Frequency



If there is a sensible colour available, use it and use consistently

NEWSMAX SNAP POLL

WHO WON THE DEBATE?

DONALD TRUMP



93%

KAMALA HARRIS



6%

NEWSMAX

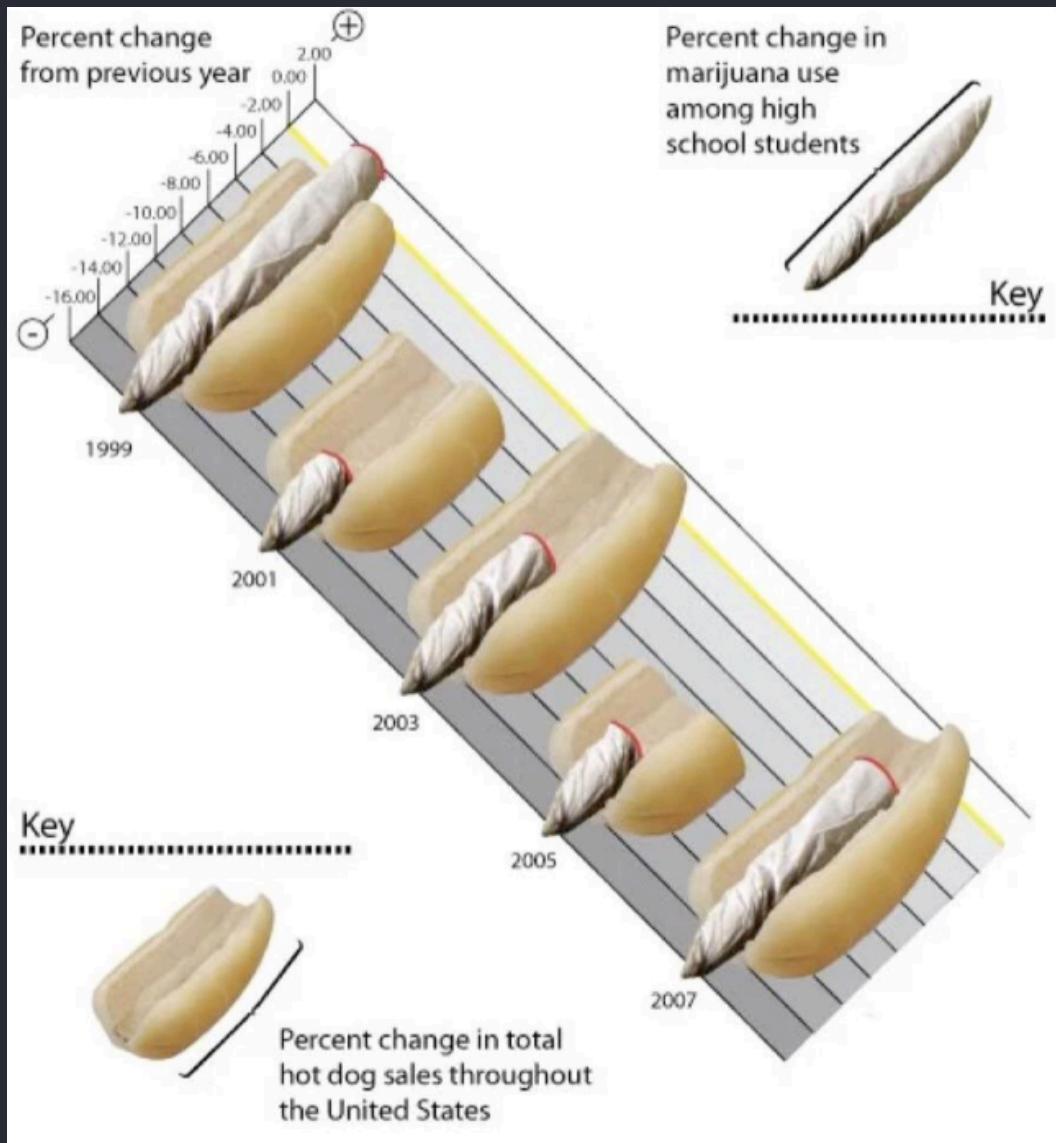
10:13 CT LIVE

VOTE-NEWSMAX.COM/DEBATE

TARKLY DIFFERENT VISIONS IN TENSE, HIGH-STAKES DEBATE . . .

SOME KEY QUOTES

Are your samples representative?



This is just cursed

Thank you

Any questions?

References

- Anscombe, F.J. (1973) "Graphs in statistical analysis," *The American Statistician*, 27(1), pp. 17–21. Available at: <https://doi.org/10.2307/2682899>.
- Gillespie, C. *et al.* (2025) *datasauRus: Datasets from the datasaurus dozen*. Available at: <https://github.com/jumpingrivers/datasauRus>.
- Kievit, R. *et al.* (2013) "Simpson's paradox in psychological science: A practical guide," *Frontiers in Psychology*, 4. Available at: <https://doi.org/10.3389/fpsyg.2013.00513>.
- Simpson, E.H. (1951) "The interpretation of interaction in contingency tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, 13(2), pp. 238–241. Available at: <https://doi.org/10.1111/j.2517-6161.1951.tb00088.x>.
- Wikipedia contributors (2025) *Anscombe's quartet*.
https://en.wikipedia.org/wiki/Anscombe%27s_quartet.