

Effective Data Visualization for Scientific Researchers

Spanbroek, April 11th, 2022

Irene van den Broek, PhD

 @IrenevdBroek

 @JeBentWatJeMeet

 @IreneVDB

Hi there, I'm Irene!



Irene van den Broek

irenevandenbroek@gmail.com

I have a background in biomarker research

Liquid chromatography coupled to tandem mass spectrometry for the quantitative bioanalysis of bioactive and potential biomarker peptides



Irene van den Broek 2010



2005-2010
Utrecht University

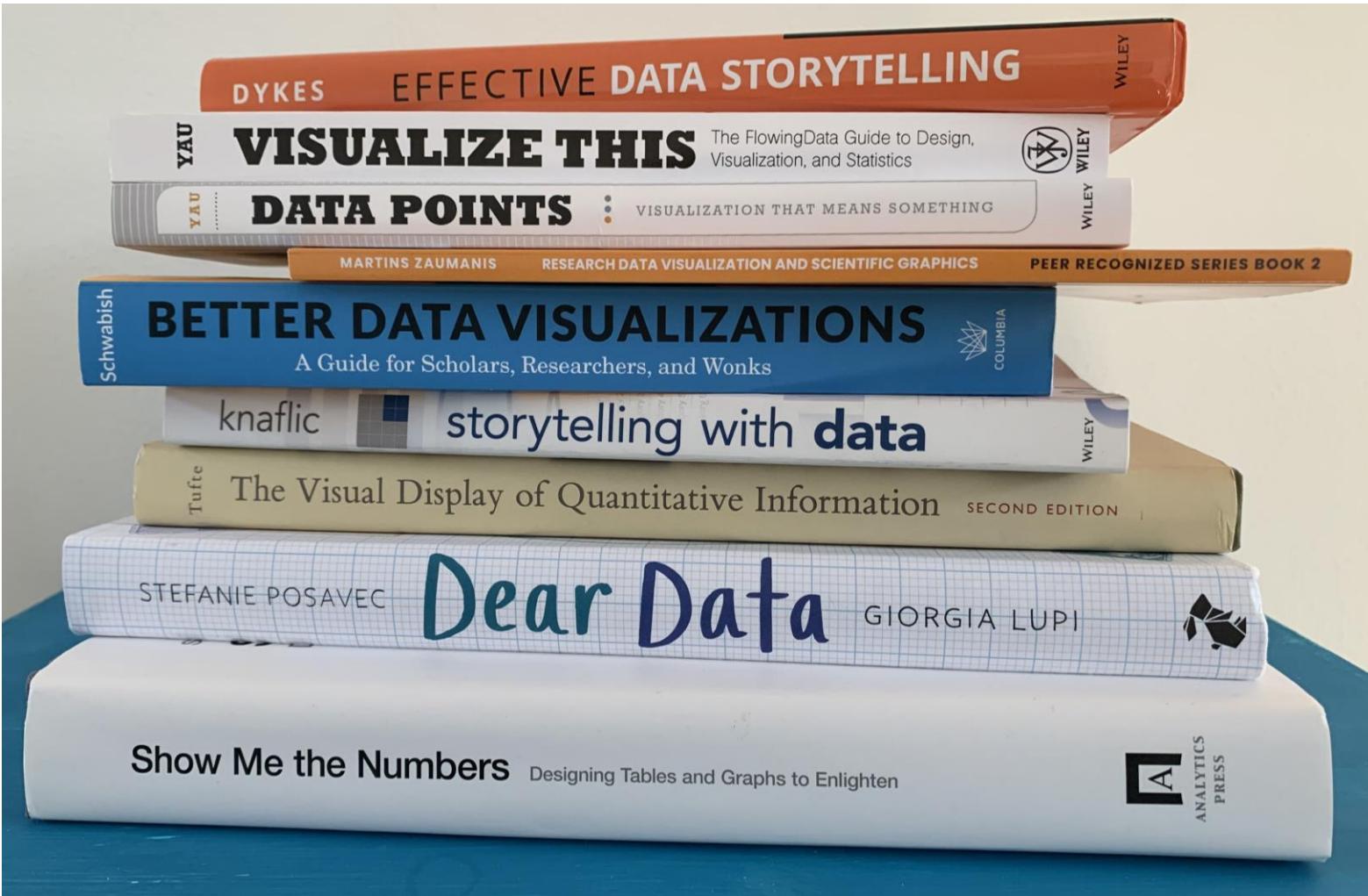


2013-2015
Leiden University
Medical Center



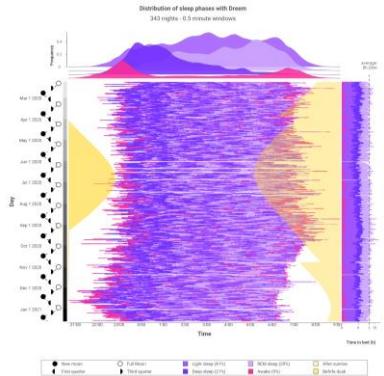
2015-2019
Cedars-Sinai
Medical Center

But when I learned Data Viz is a field on its own, I was hooked!

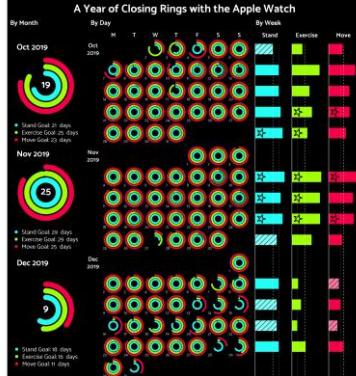


I started visualizing my Health and Habits.

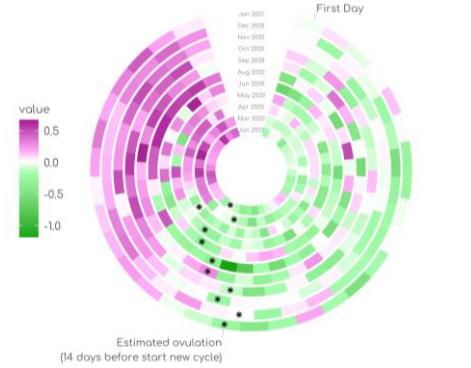
Sleep



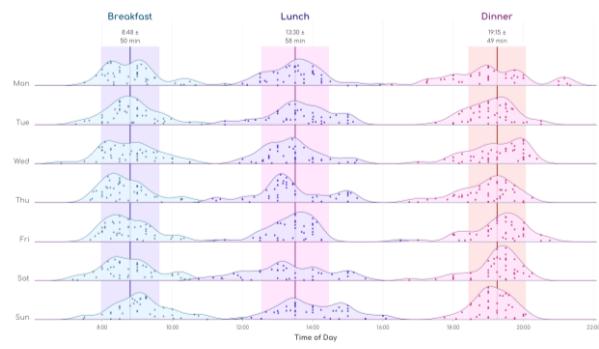
Activity



Body



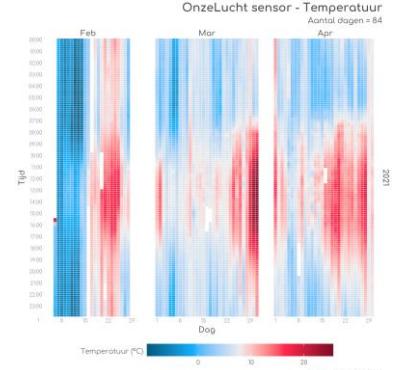
Food



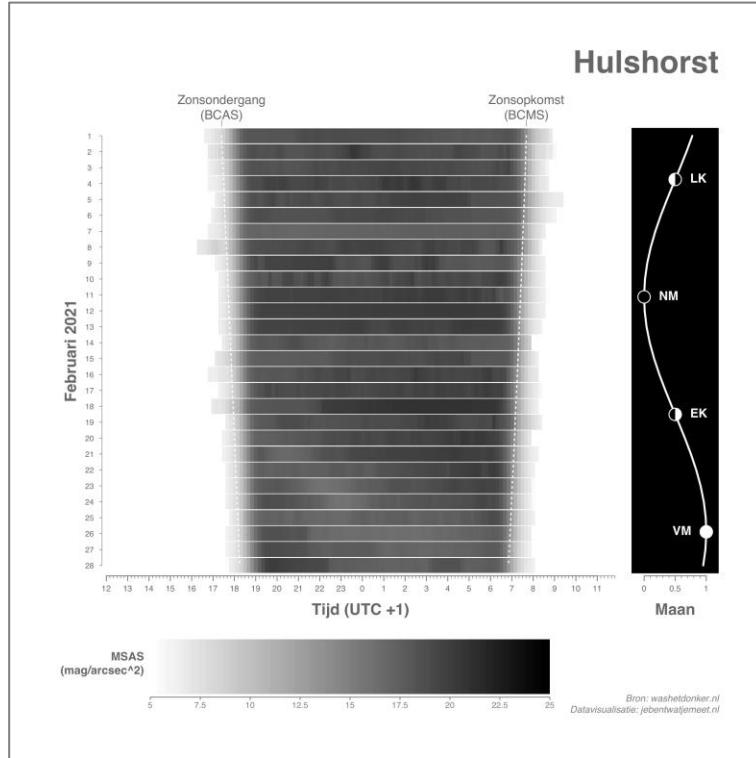
Habits



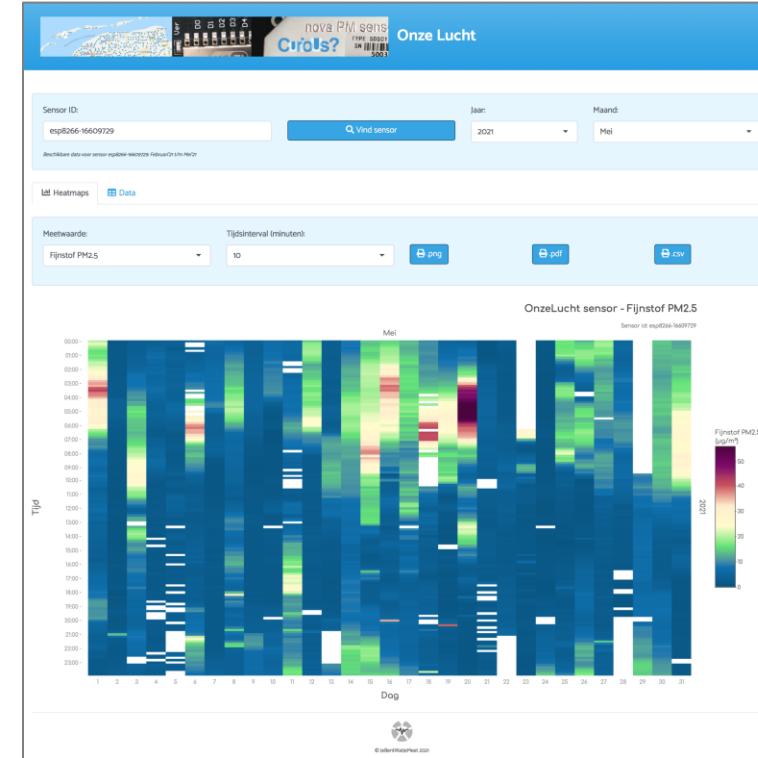
Environment



And now combine my passions for teaching and data visualization

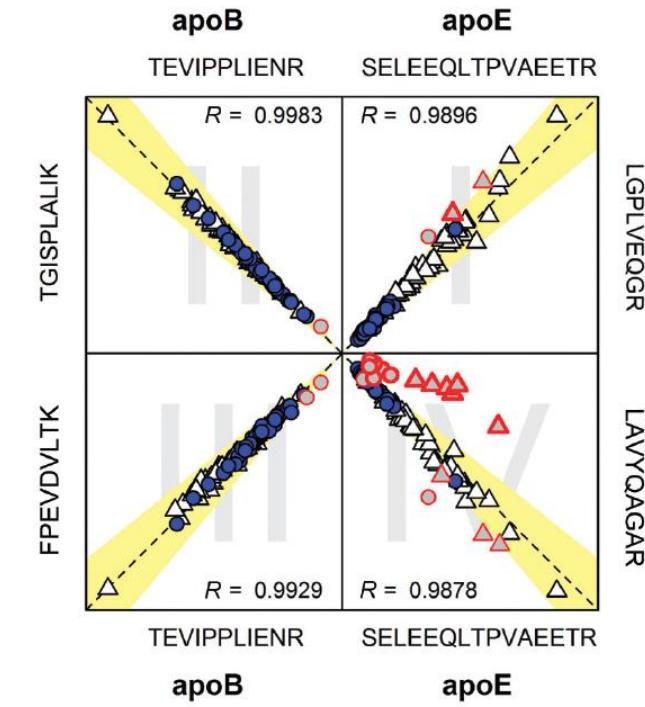
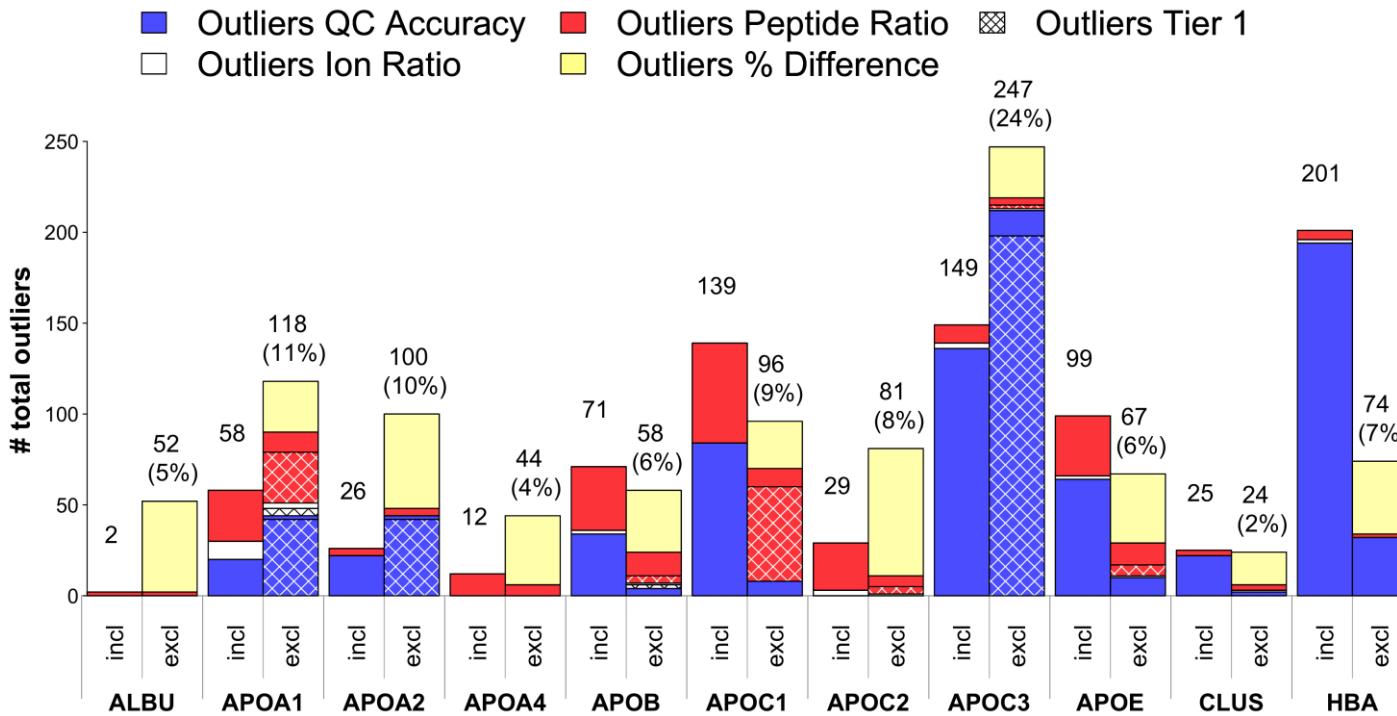


www.washetdonker.nl

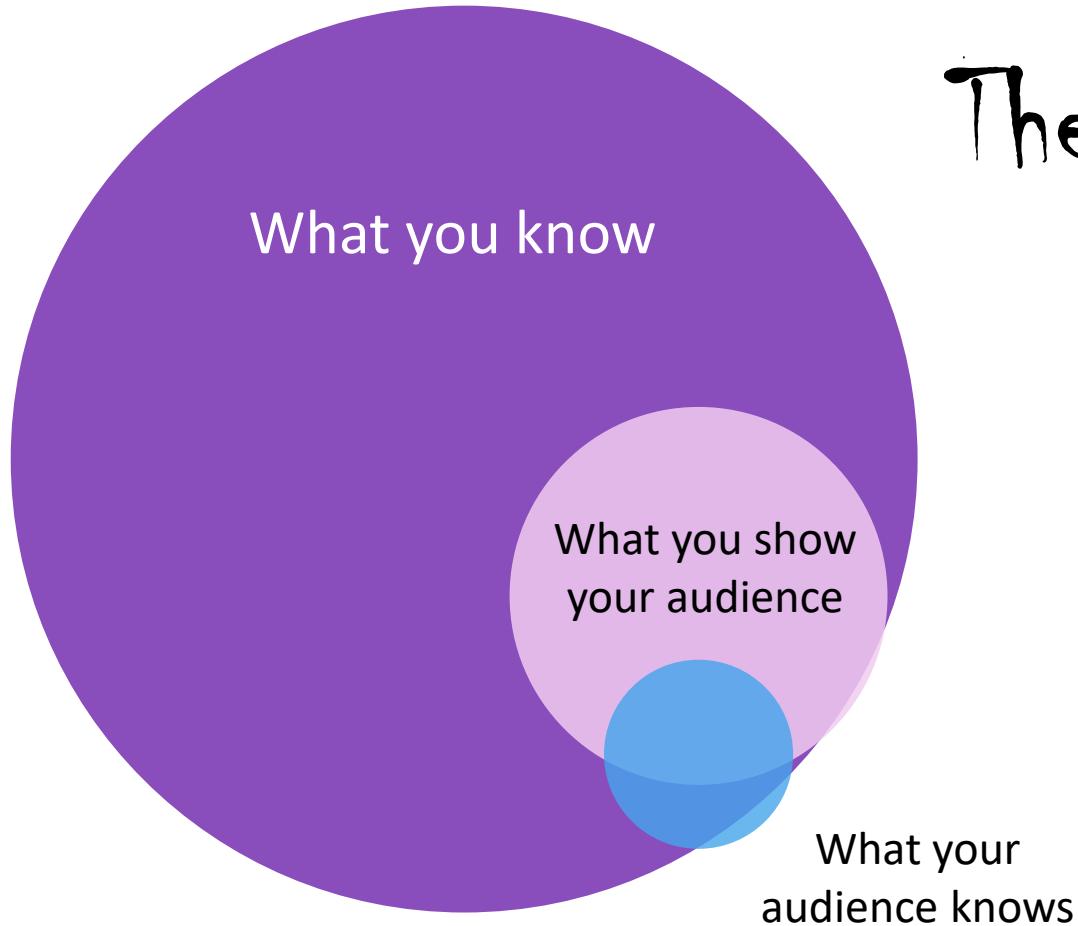


www.onzelucht.nl

I'm excited to share with you what I wish I'd learned during my PhD!



Don't underestimate how smart you are!



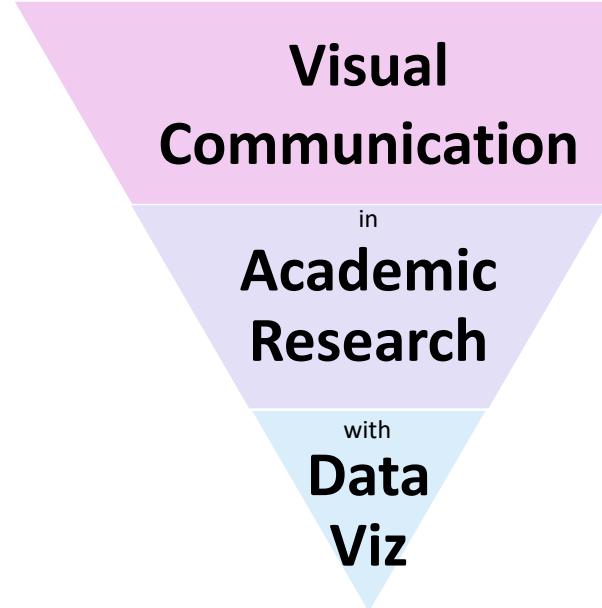
The curse of knowledge

Resist the urge to show *everything*!



An overview for today:

1

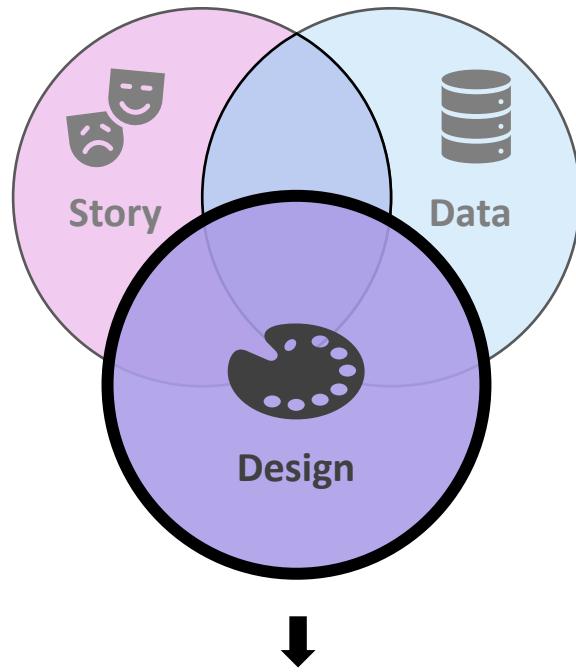


WHY?

WHAT?

HOW?

2



🧠 Understand the brain

👍 Do's and 👎 Don'ts

✍ Practice ⏴

Warm-up!

And stand up

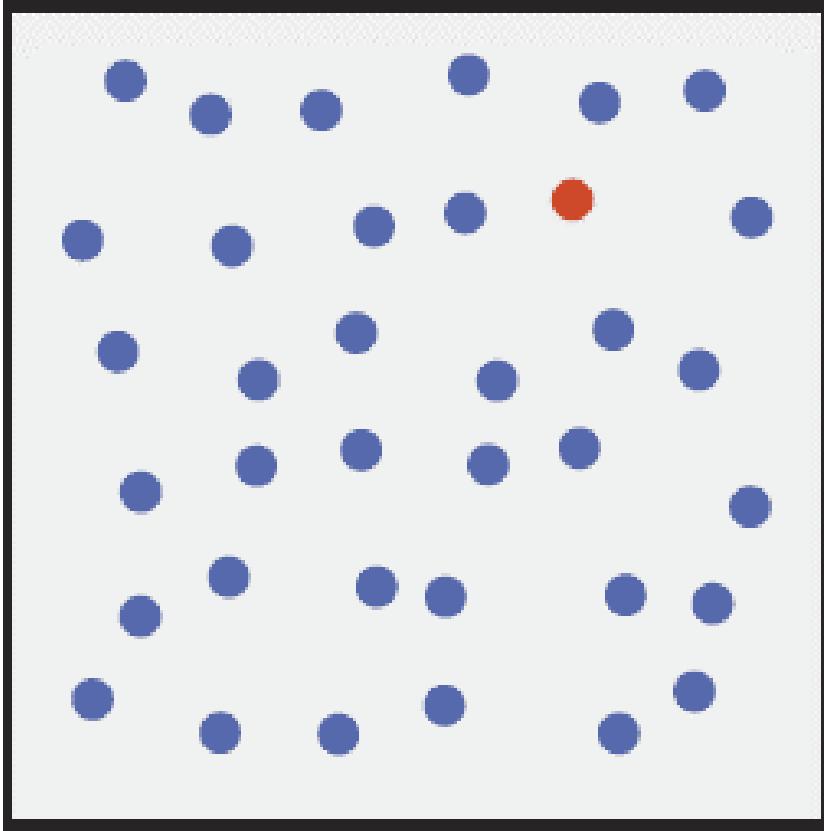
LEFT?

YES?

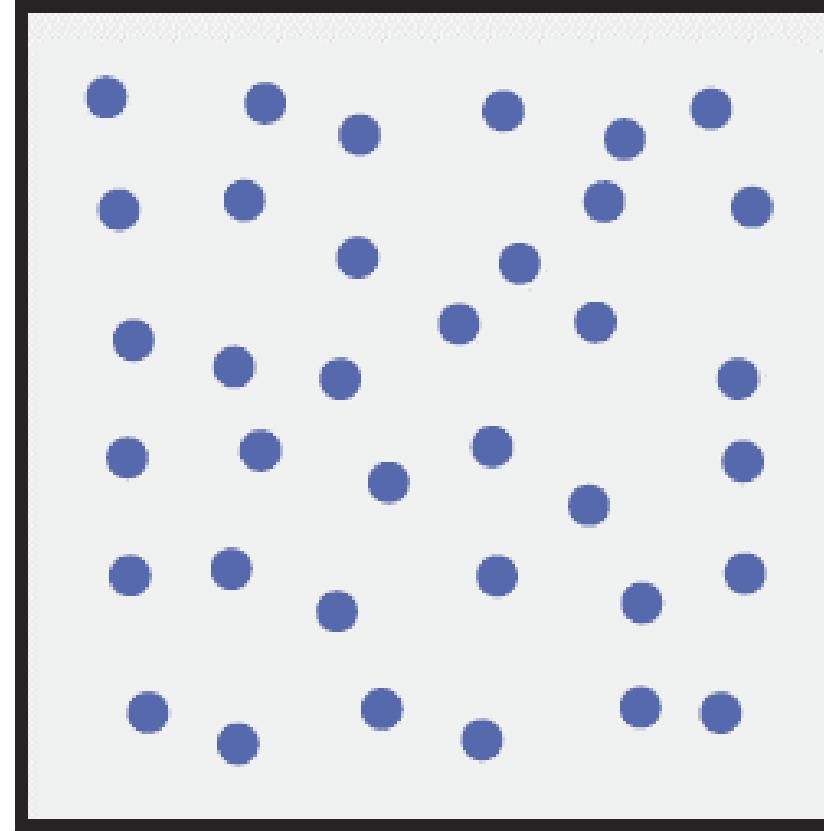


Which panel has a red dot?

A



B



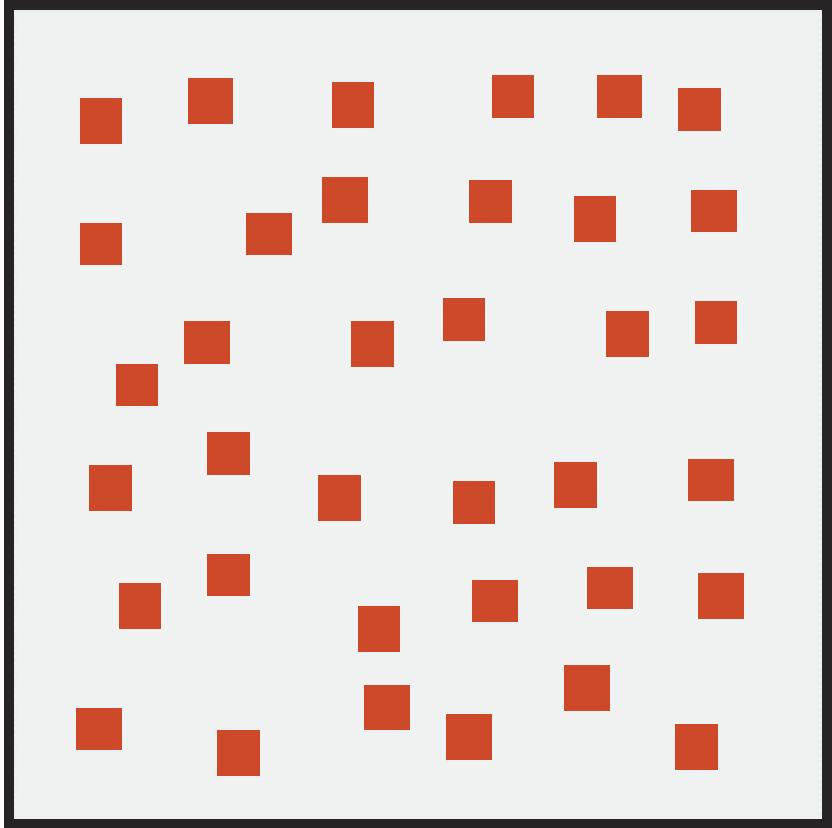
LEFT?

=

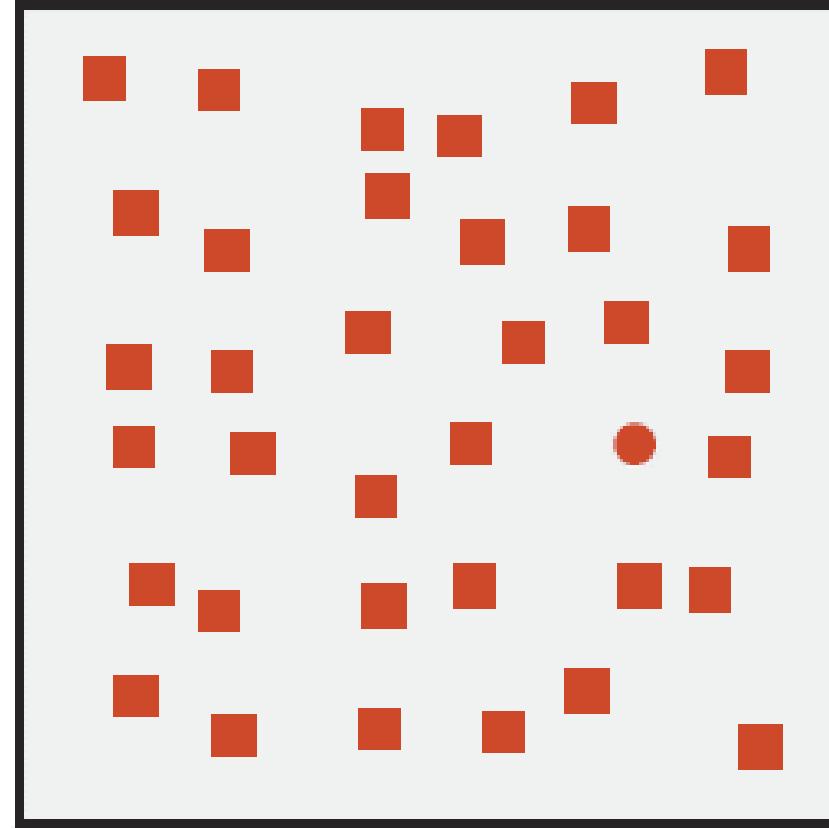
UP!

Which panel has a red dot?

A



B



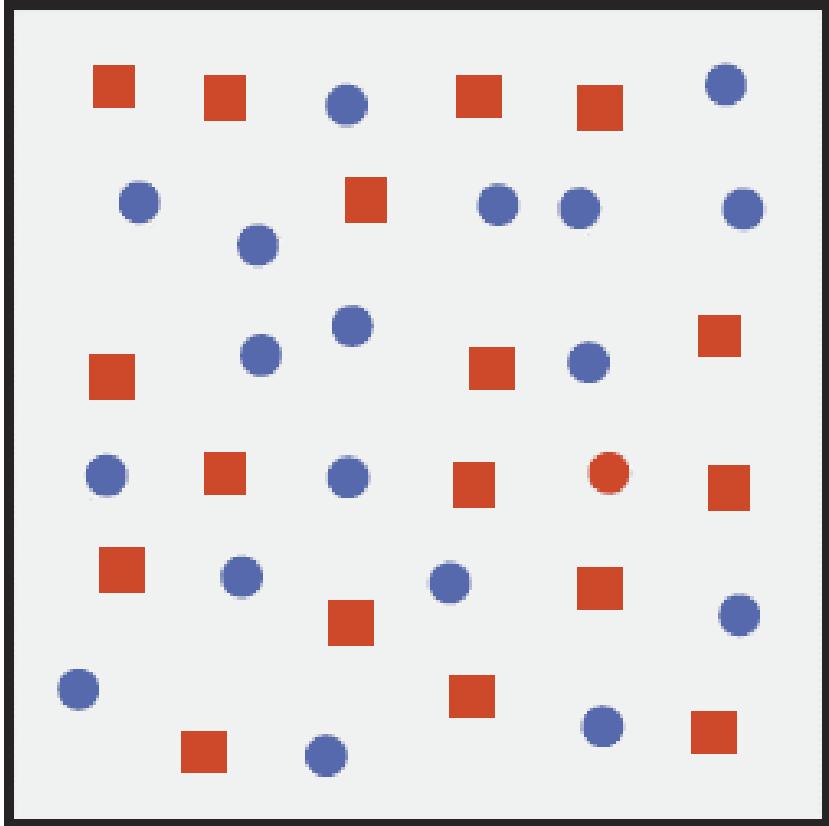
LEFT?

=

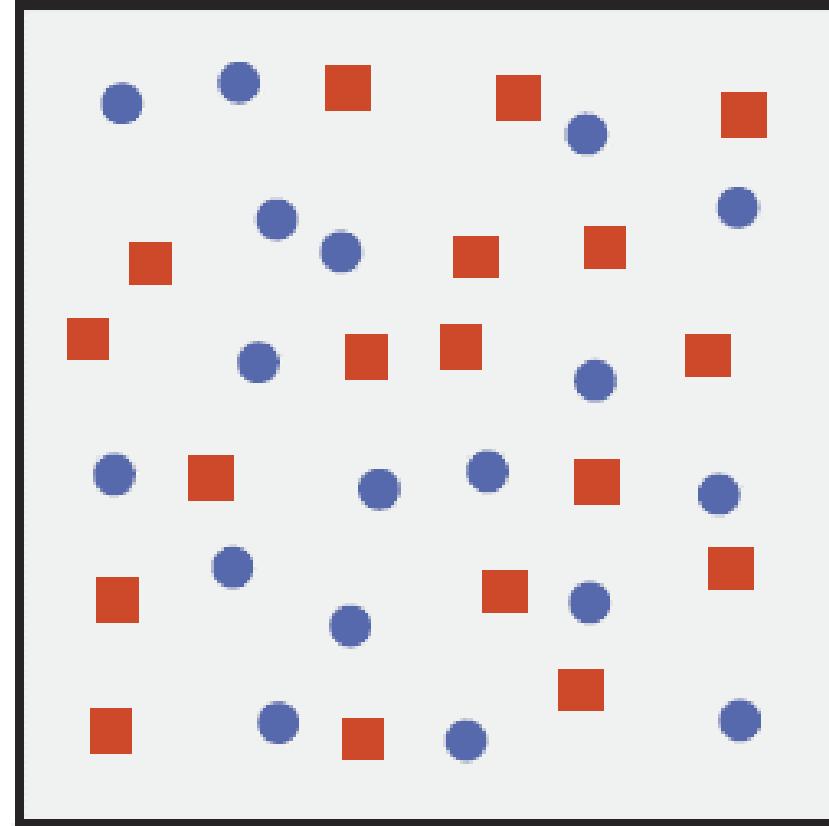
UP!

Which panel has a red dot?

A



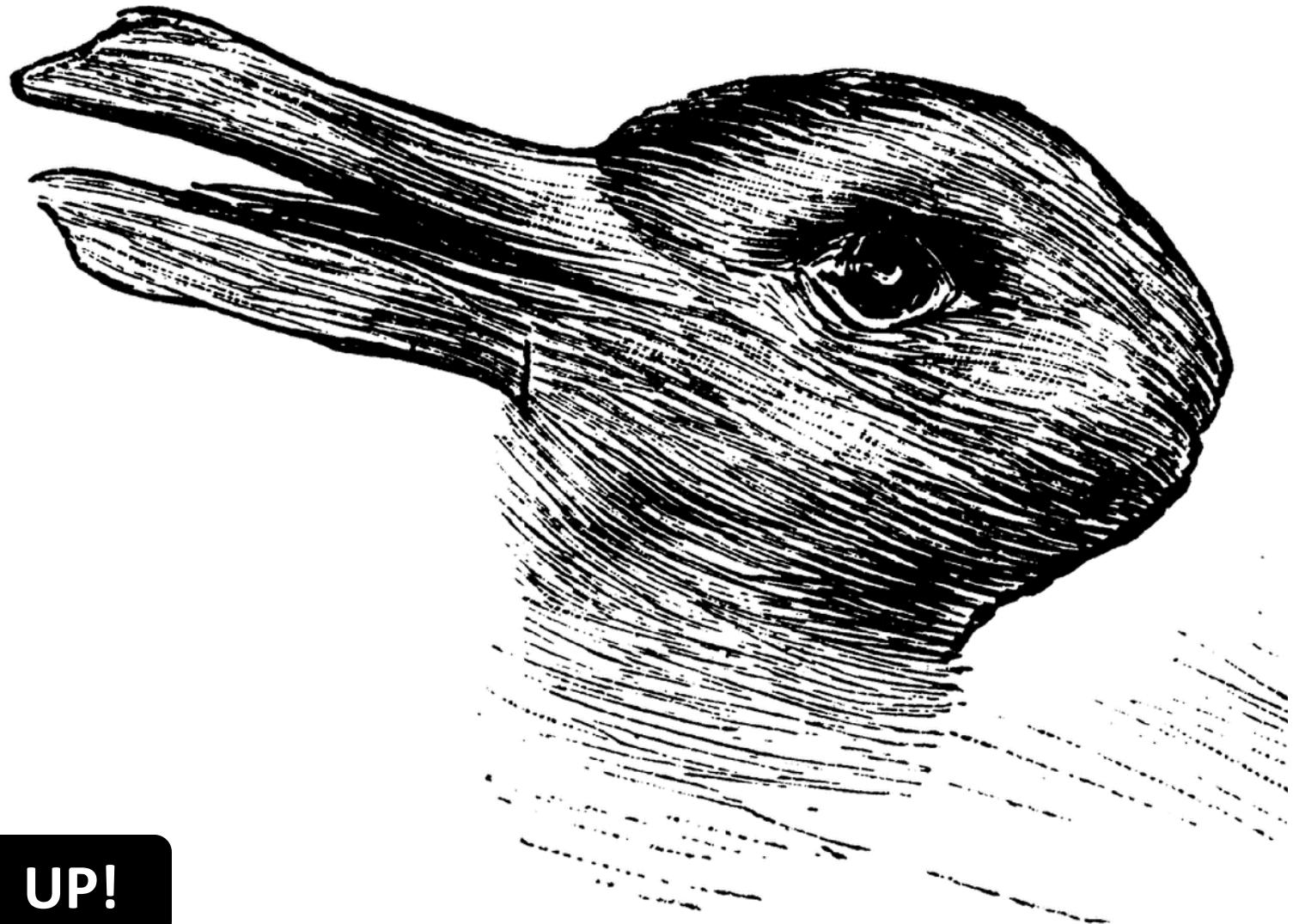
B



LEFT?

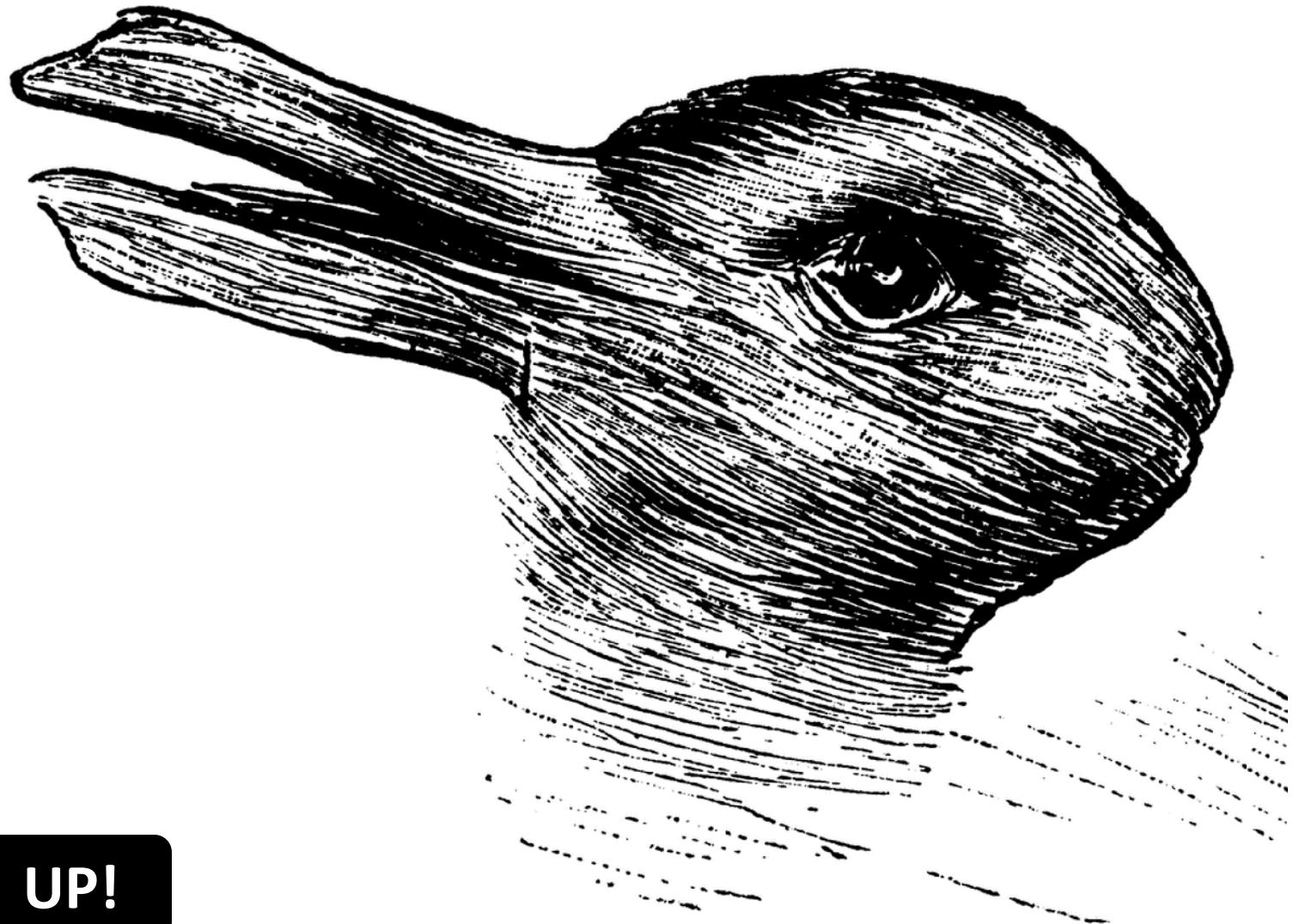
= UP!

Do you see a Rabbit?



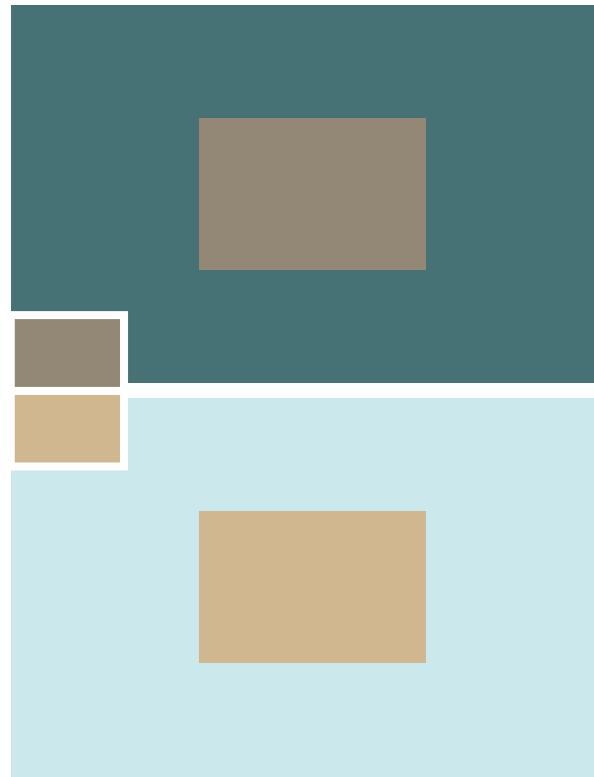
YES? = UP!

Do you see a Duck?



YES? = UP!

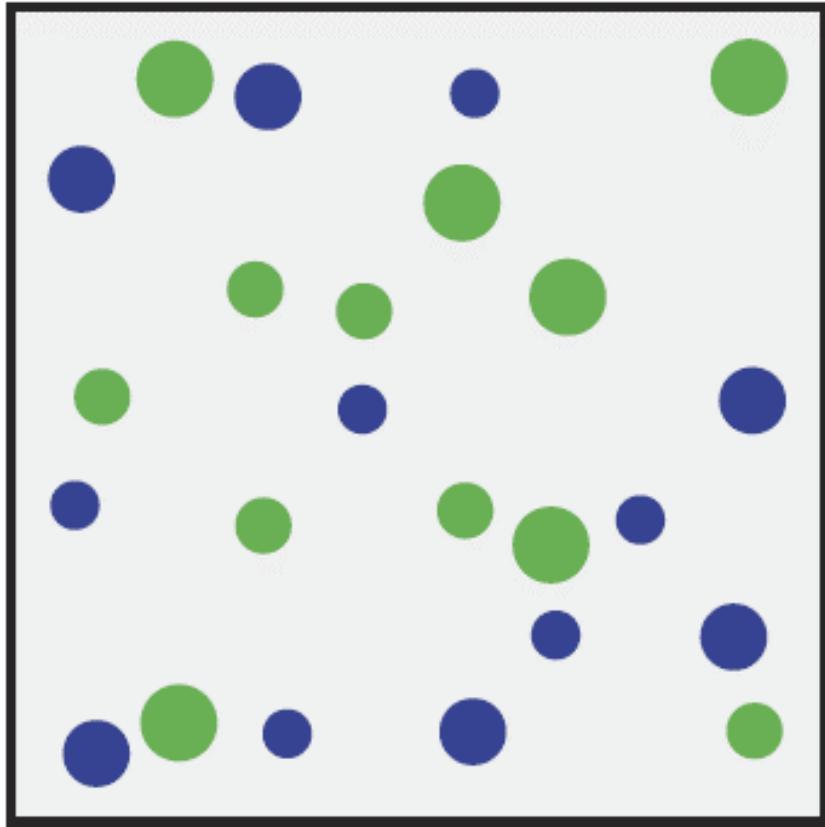
The inner squares have the same color



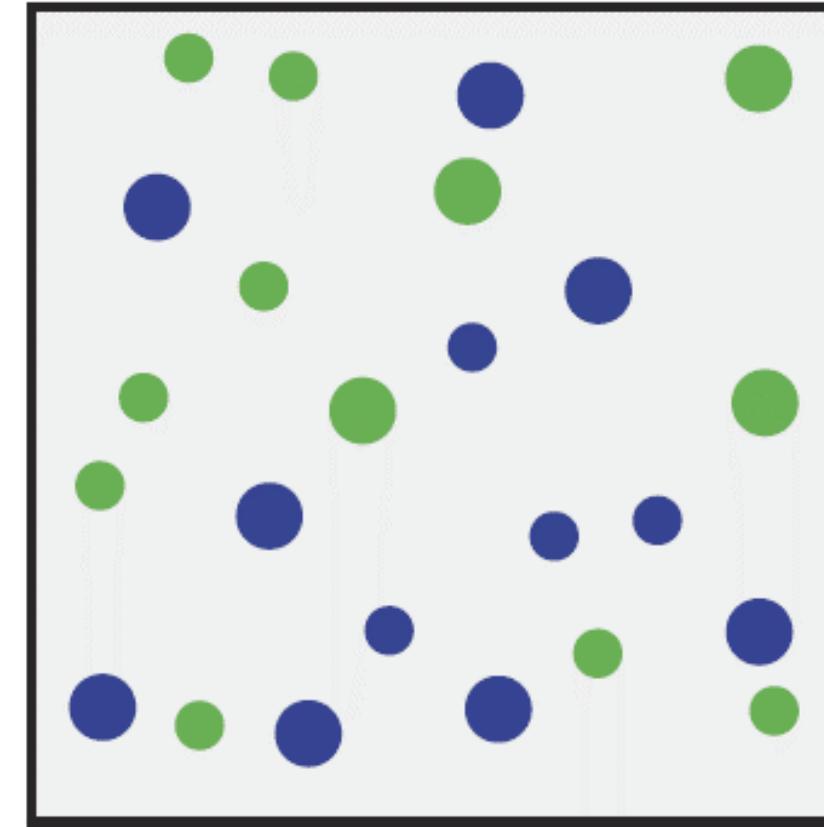
YES? = **UP!**

Which panel has the **biggest green** dots?

A



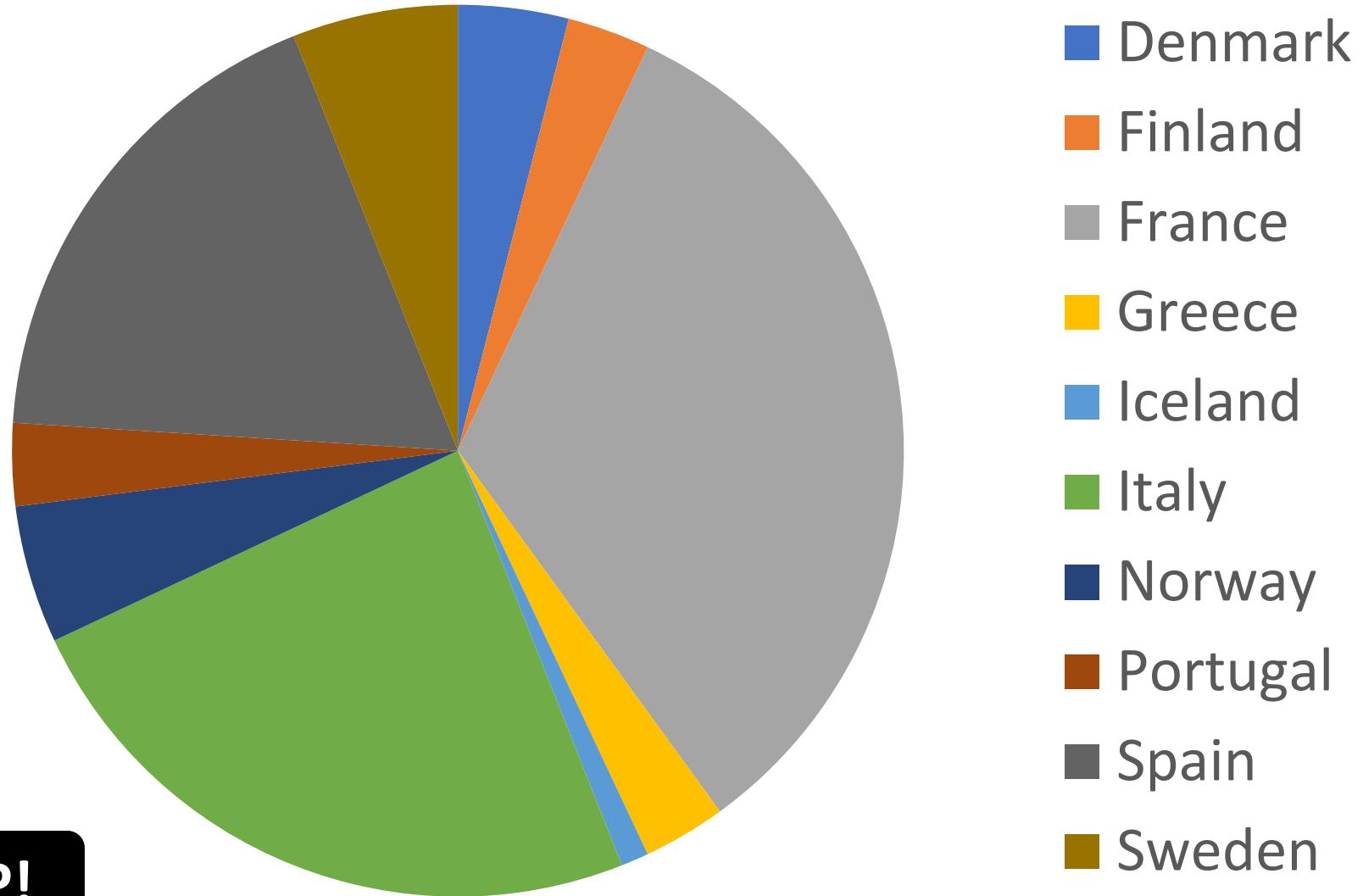
B



LEFT? =

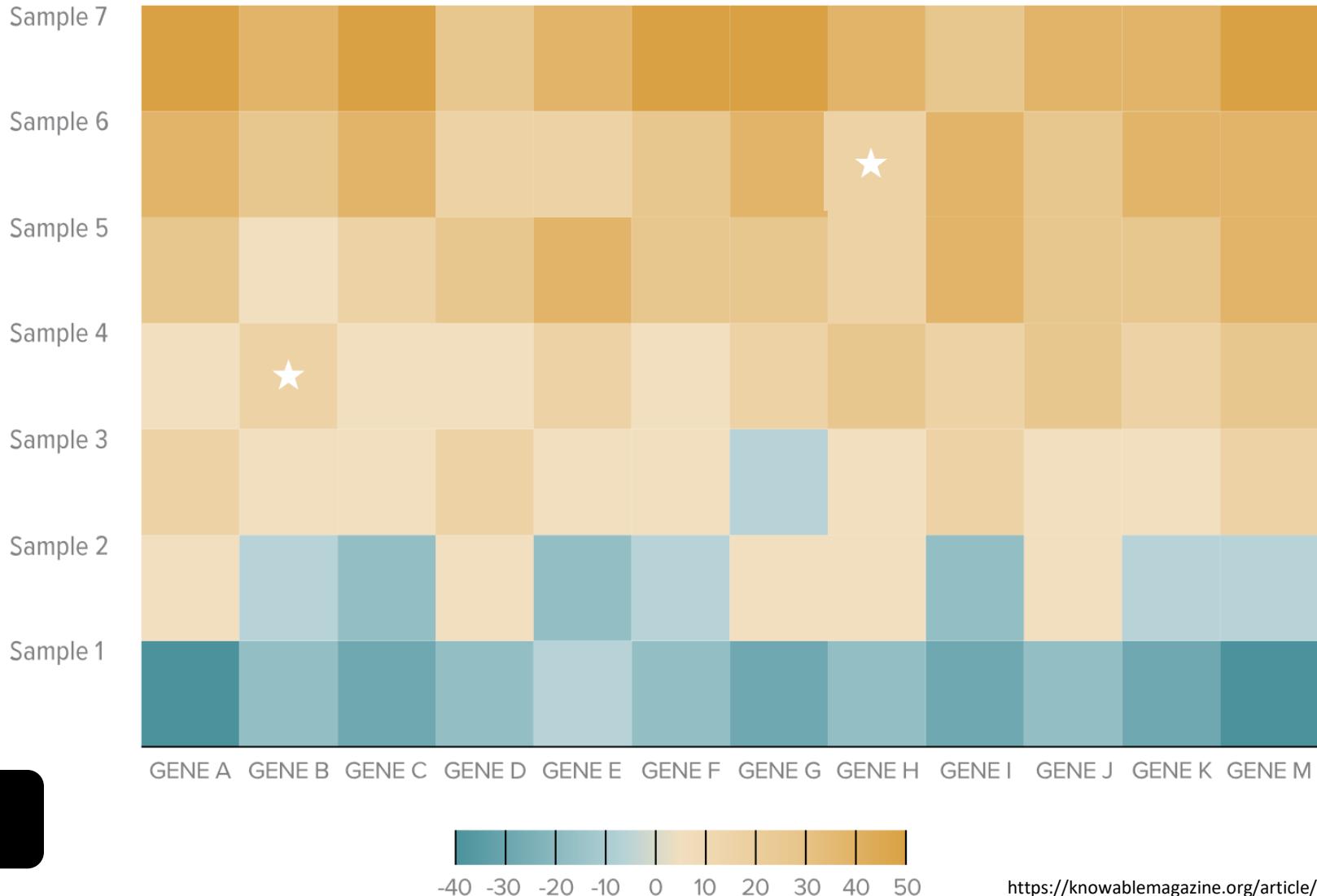
UP!

Norway has more widgets than Finland

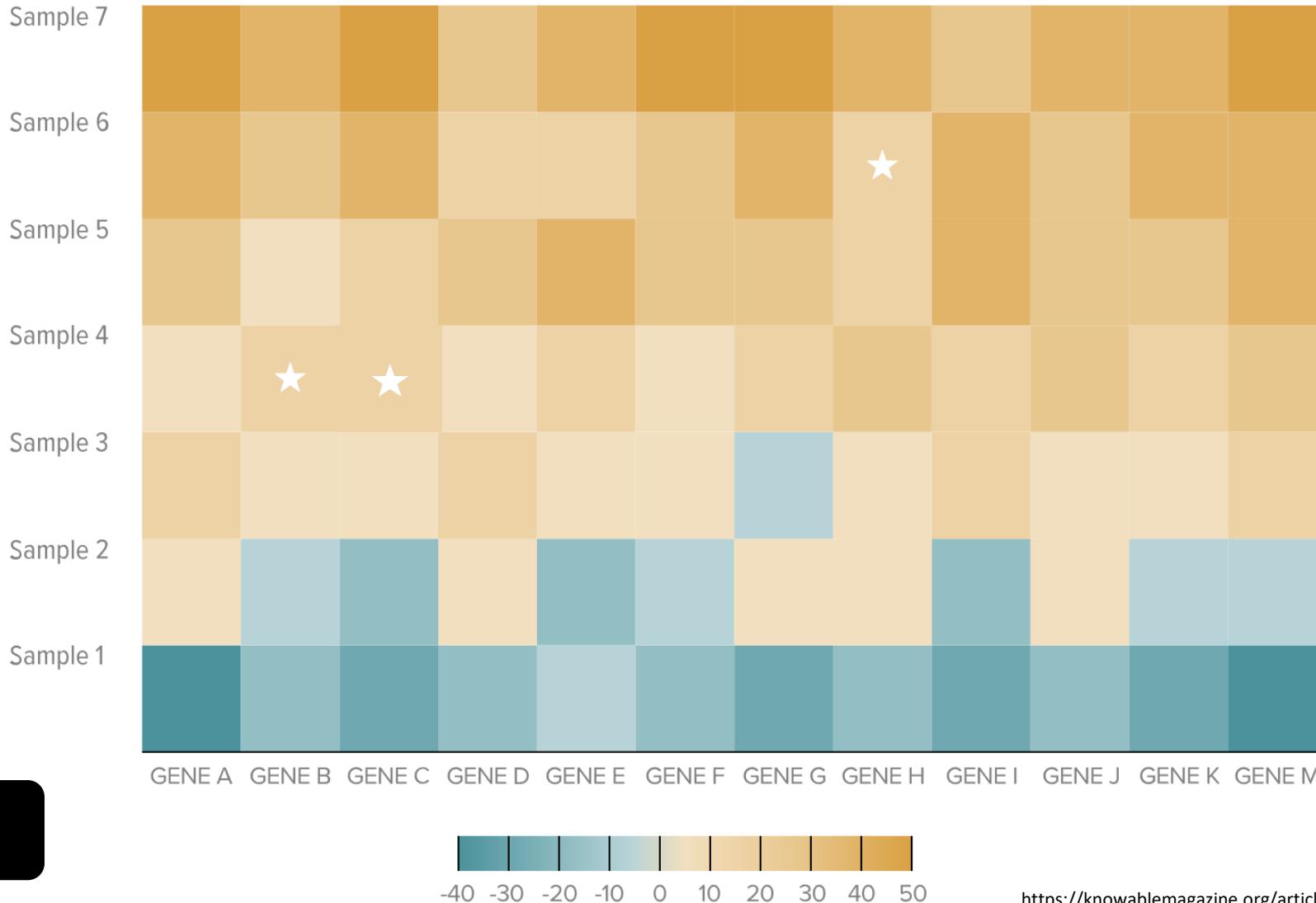


YES? = UP!

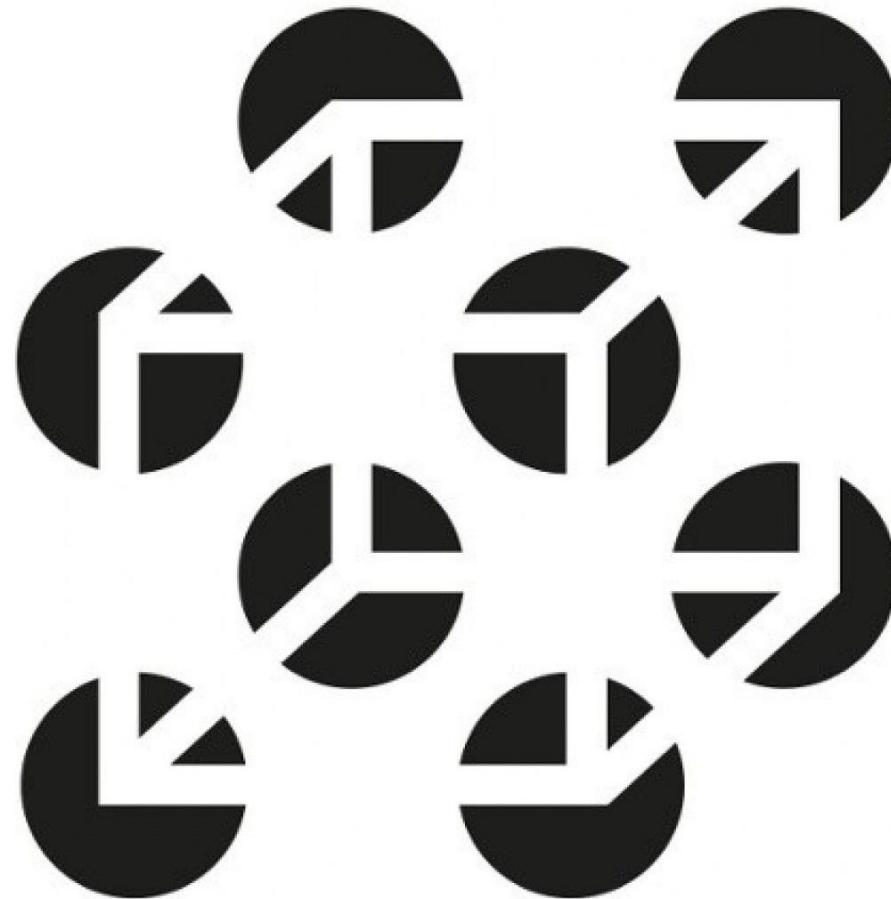
The areas with a star have the same color



The areas with a star have the same color

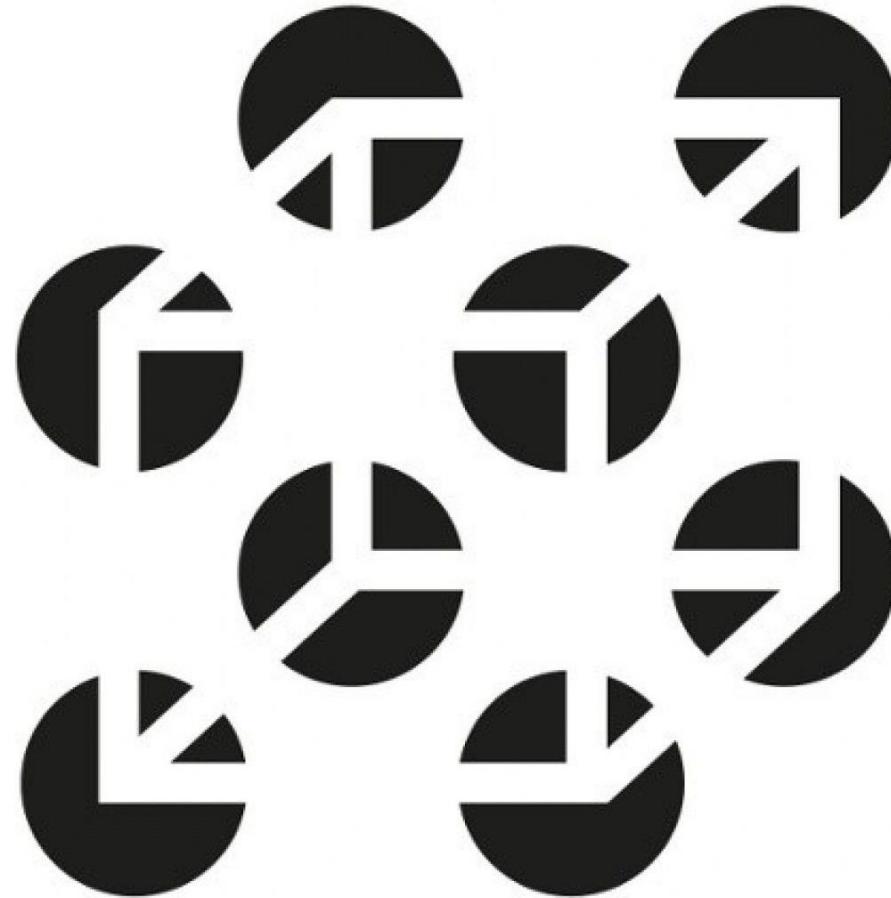


Do you *see* a cube?



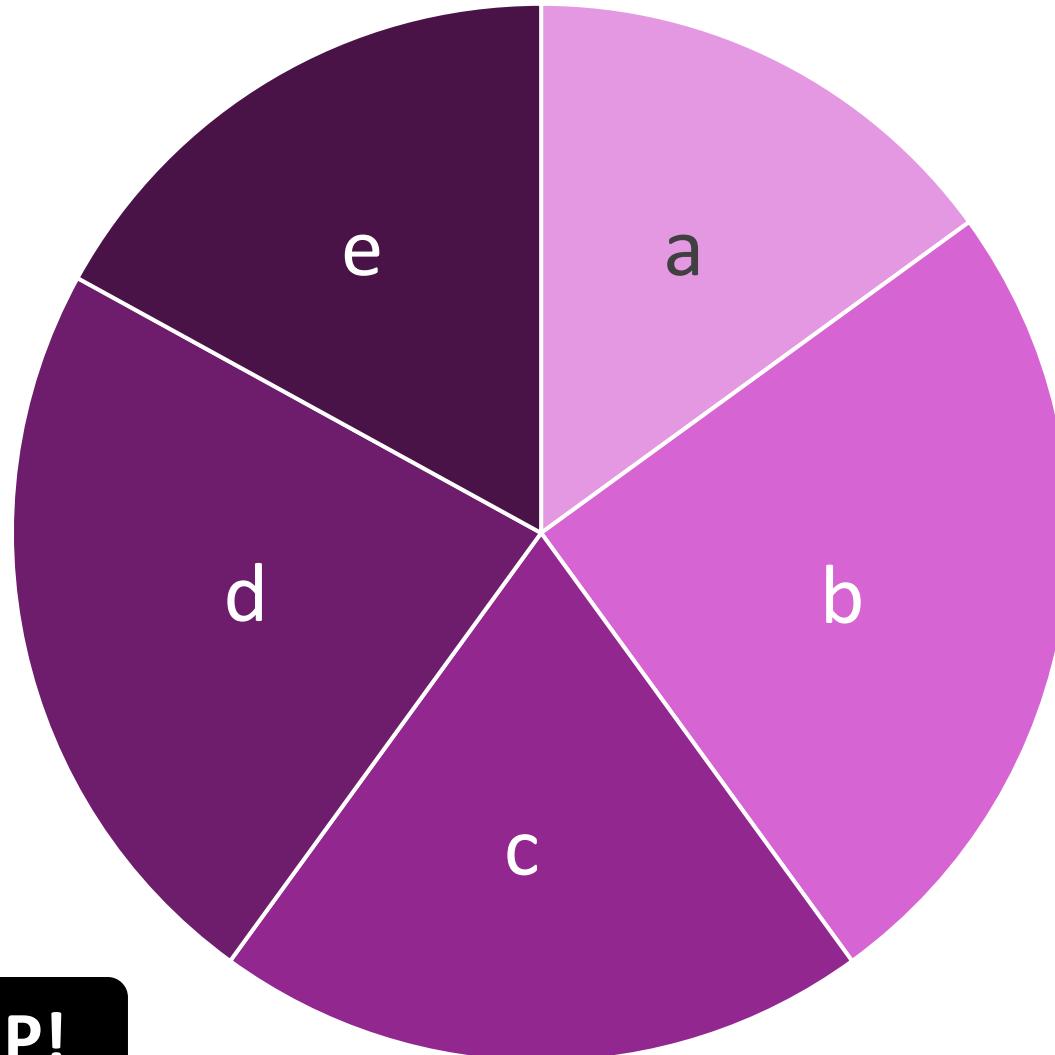
YES? = UP!

Is there a cube?



YES? = UP!

The order from high to low is b > d > c > e > a



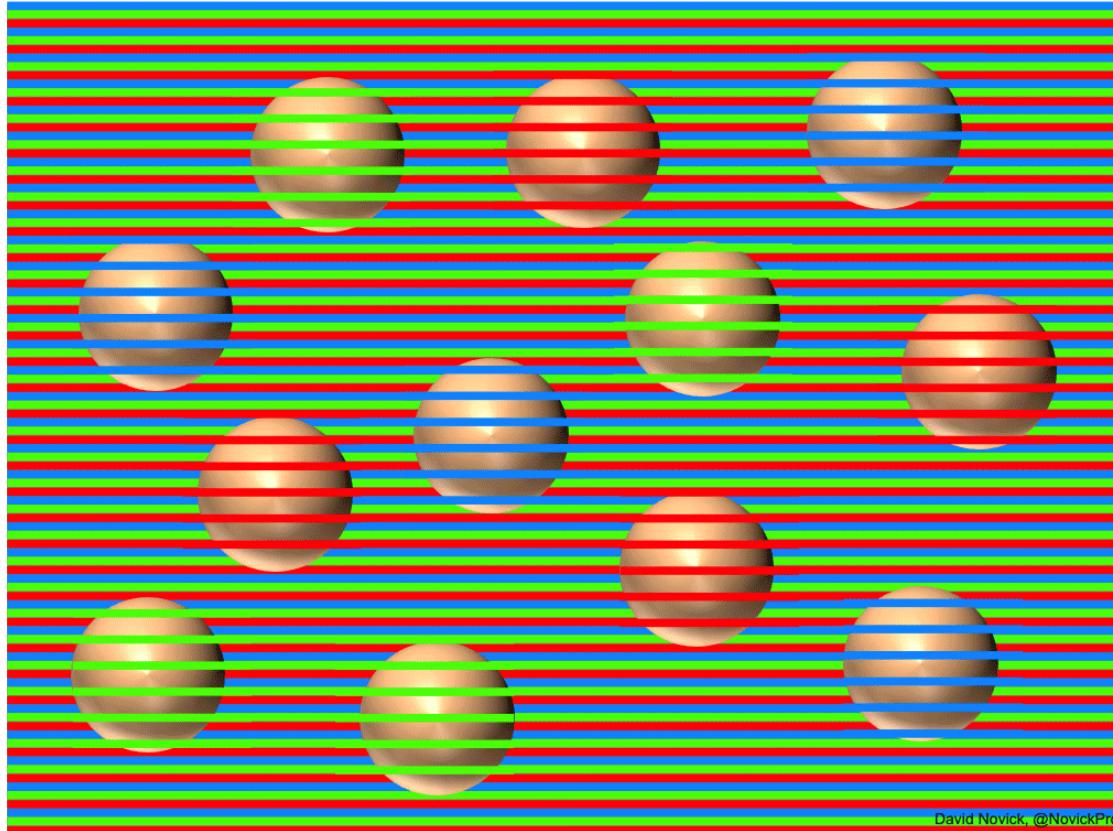
YES? = UP!

These spheres have the same color



YES? = **UP!**

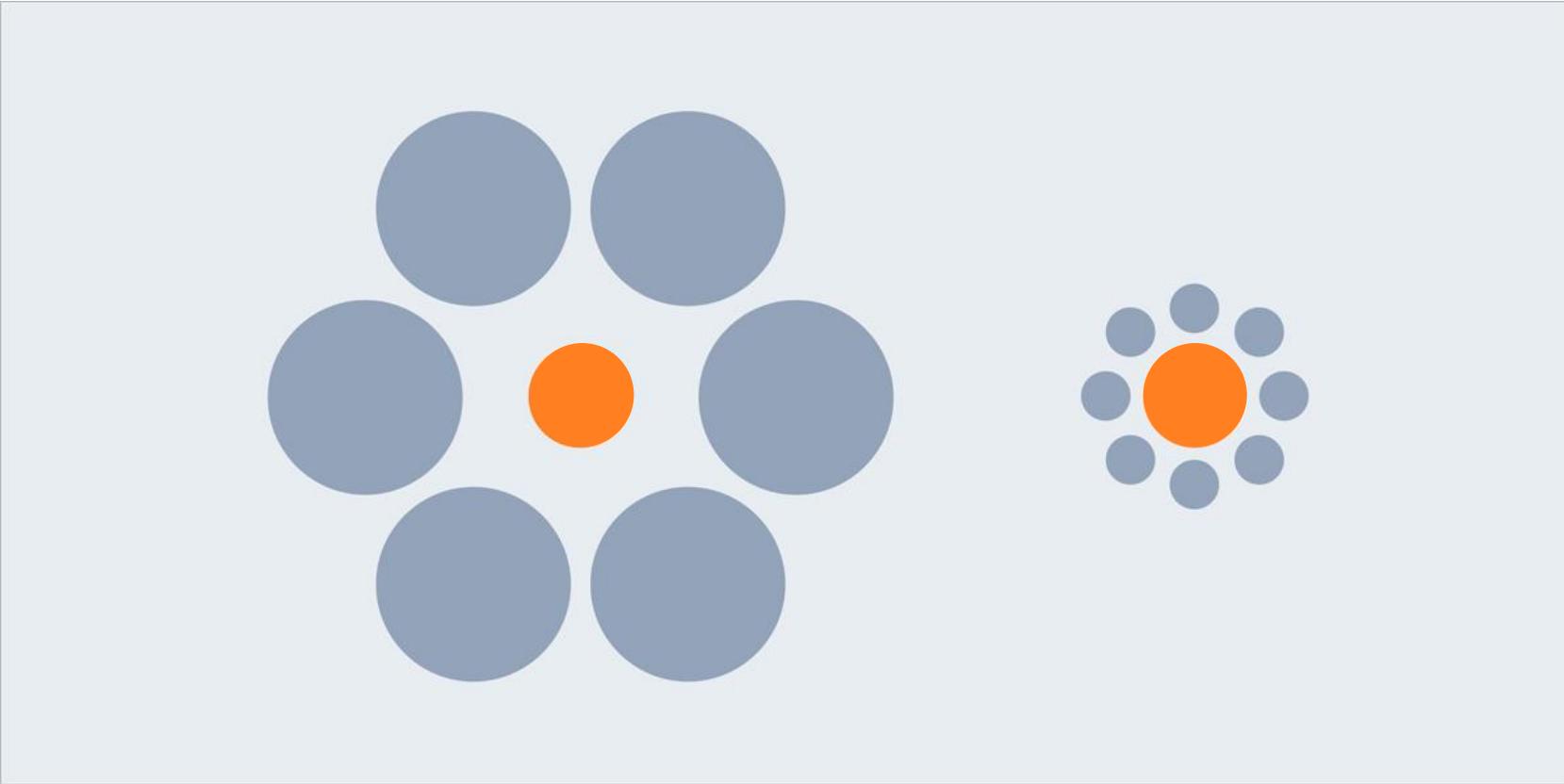
These spheres have the same color



David Novick, @NovickProf

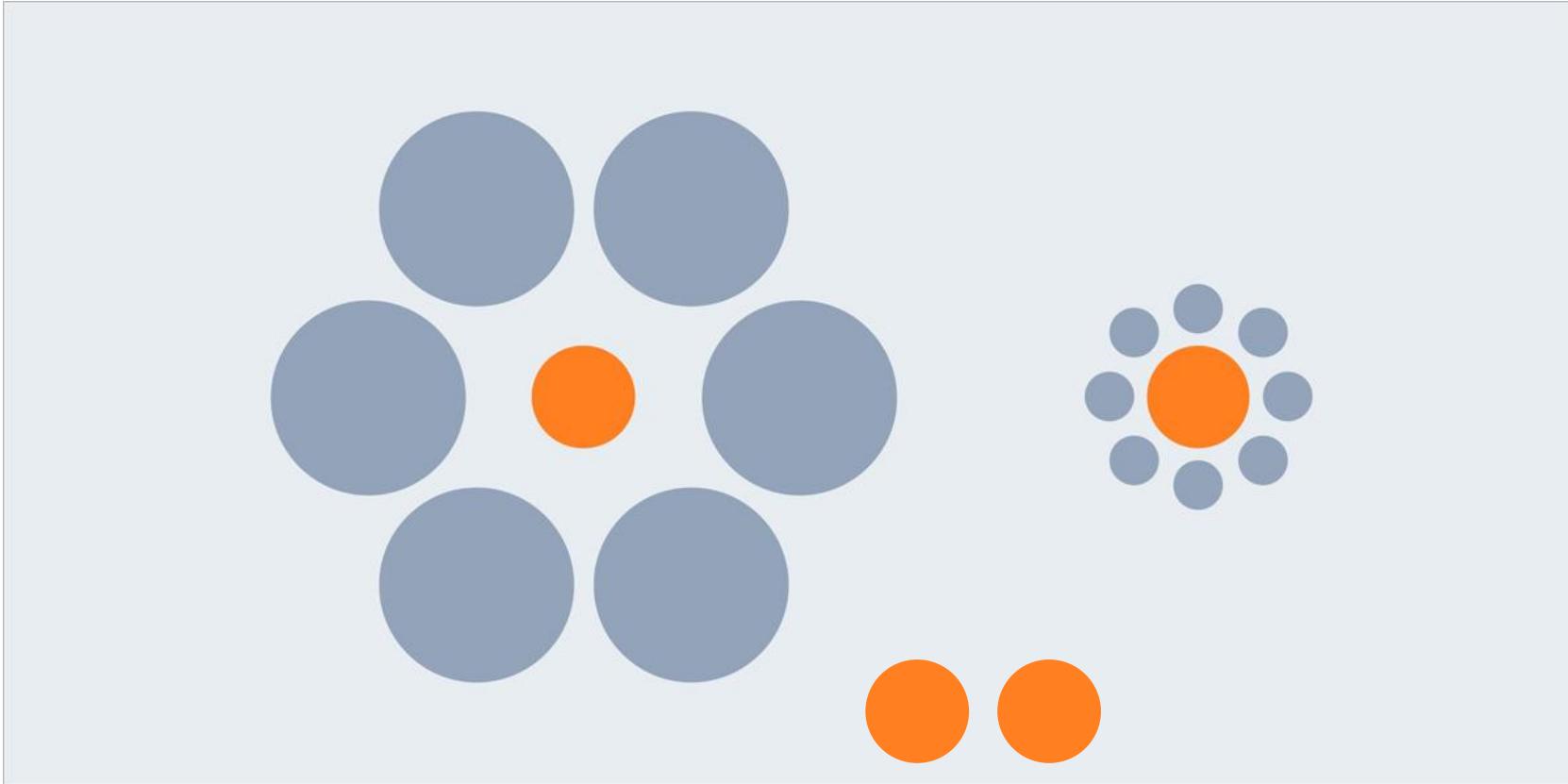
YES? = **UP!**

Which orange circle is smaller?



LEFT? = UP!

Which orange circle is smaller?



LEFT? = UP!



*“It’s not what you look at that
matters, it’s what you see”*

Henry David Thoreau

Why

**Visual
Communication**

in

**Academic
Research**

?

with

**Data
Viz**

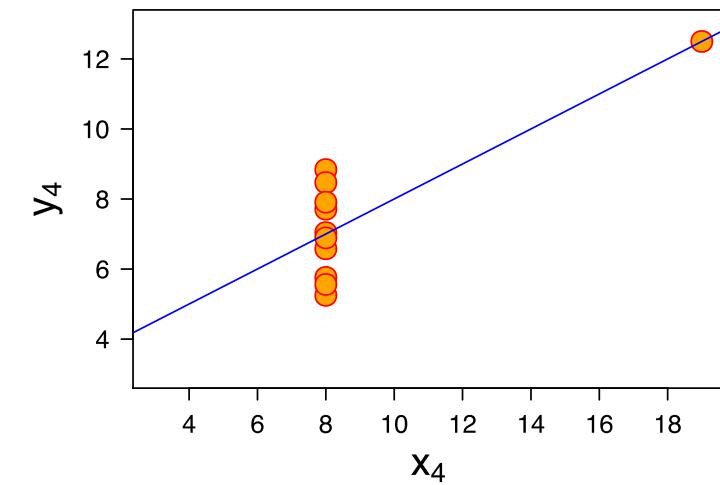
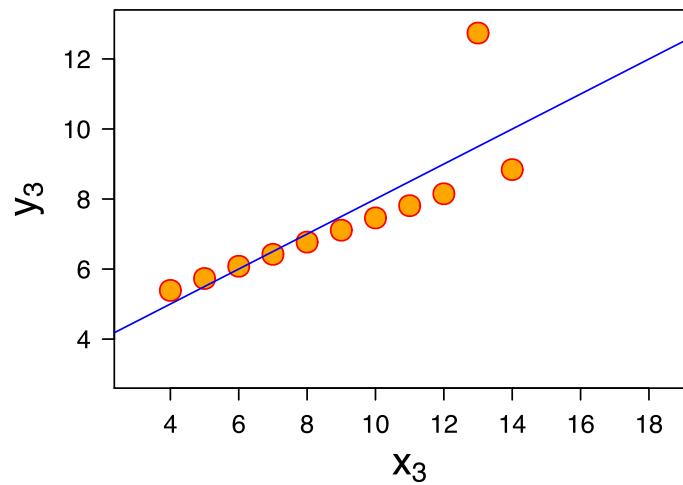
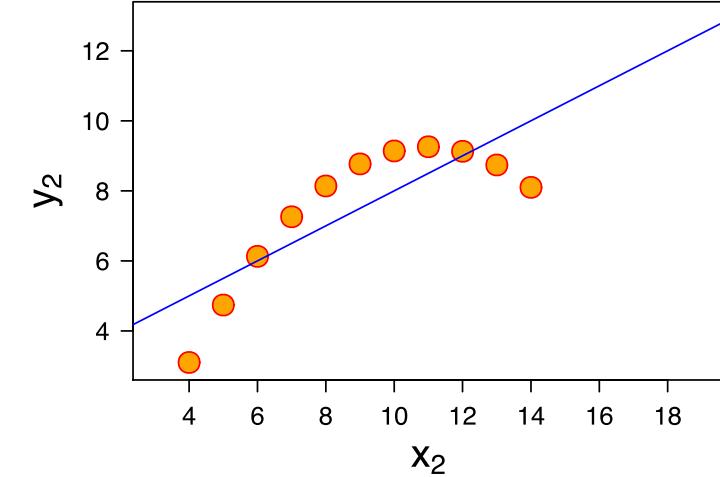
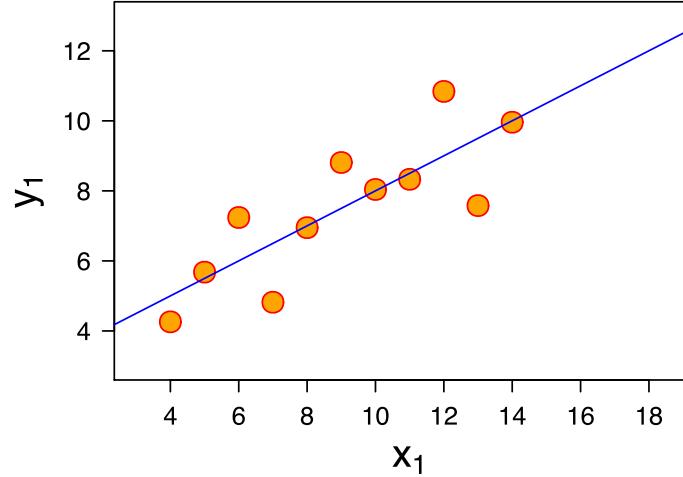
What can you see in these numbers?

Anscombe's quartet:
4 datasets
11 x-values
11 y-values
Same total
Same mean
Same standard deviation

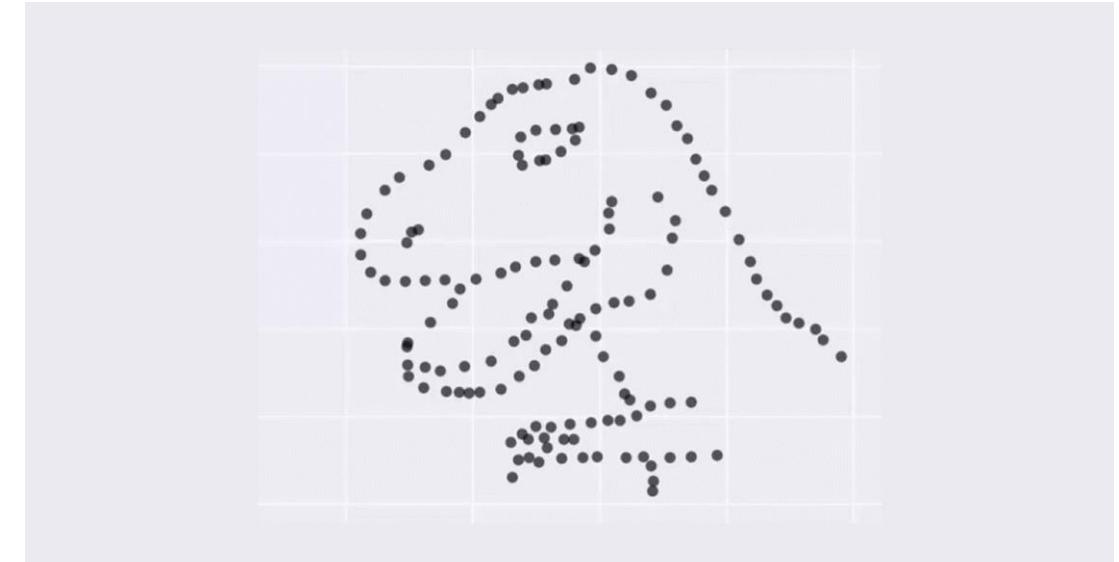
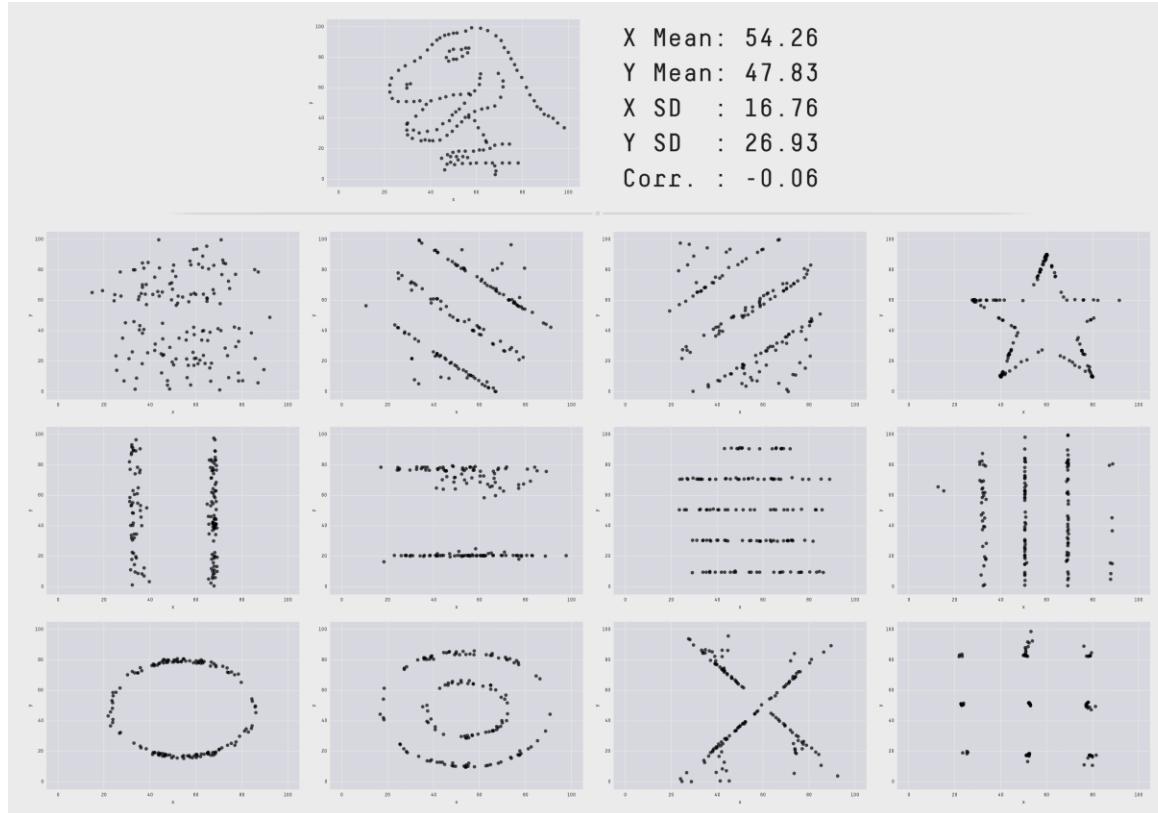
	I	II	III	IV				
	x	y	x	y	x	y	x	y
	10	8,04	10	9,14	10	7,46	8	6,58
	8	6,95	8	8,14	8	6,77	8	5,76
	13	7,58	13	8,74	13	12,74	8	7,71
	9	8,81	9	8,77	9	7,11	8	8,84
	11	8,33	11	9,26	11	7,81	8	8,47
	14	9,96	14	8,1	14	8,84	8	7,04
	6	7,24	6	6,13	6	6,08	8	5,25
	4	4,26	4	3,1	4	5,39	19	12,5
	12	10,84	12	9,13	12	8,15	8	5,56
	7	4,82	7	7,26	7	6,42	8	7,91
	5	5,68	5	4,74	5	5,73	8	6,89
SUM	99,00	82,51	99,00	82,51	99,00	82,50	99,00	82,51
AVG	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
STDEV	3,32	2,03	3,32	2,03	3,32	2,03	3,32	2,03

A picture is worth a thousand words!

Anscombe's quartet:
4 datasets
11 x-values
11 y-values
Same total
Same mean
Same standard deviation



The Datasaurus Dozen 😊: Never trust summary statistics alone!



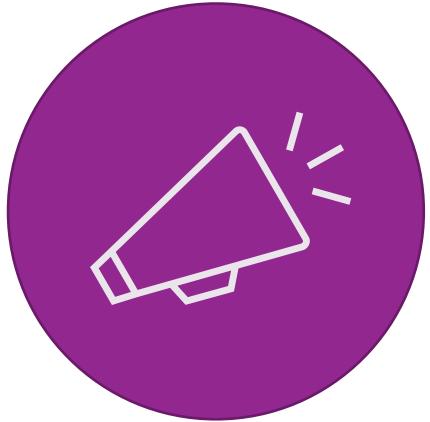
Matejka J & Fitzmaurice G (2017). *Same Stats, Different Graphs: Generating Datasets with Varied Appearance and Identical Statistics through Simulated Annealing*.
<https://www.autodeskresearch.com/publications/samestats>.

Why use visuals in academic research?



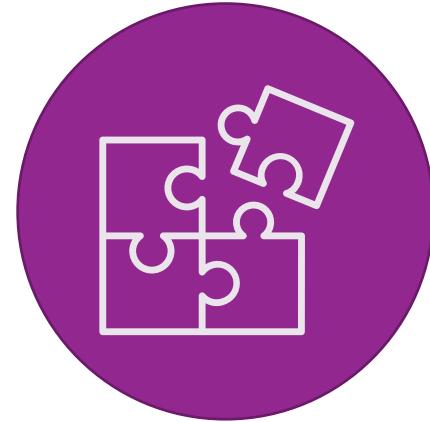
Provide **insight**

-
- Publications
 - Oral presentations
 - Poster presentations



Grab **attention**

-
- Graphical abstracts
 - Social media
 - General media



Show **competence**

-
- Grant applications
 - Portfolio / CV
 - Pitch

WHAT is

Visual
Communication

in

Academic
Research

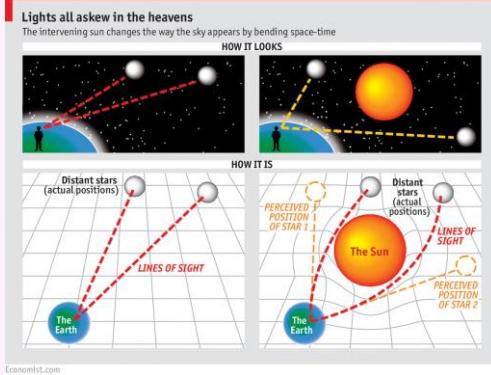
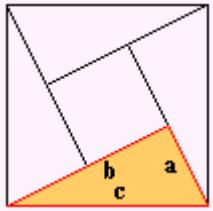
with

Data
Viz

?

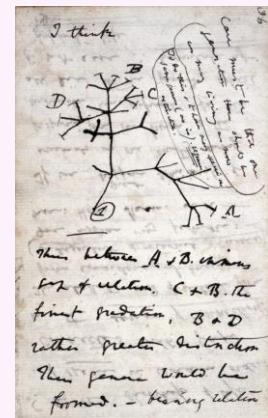
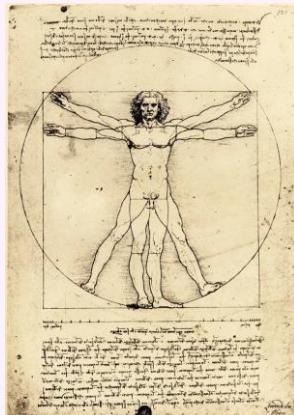
Four types of visualizations

Idea illustration



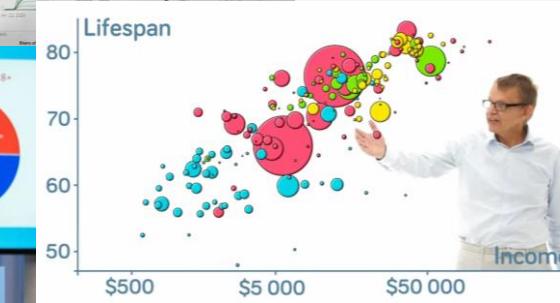
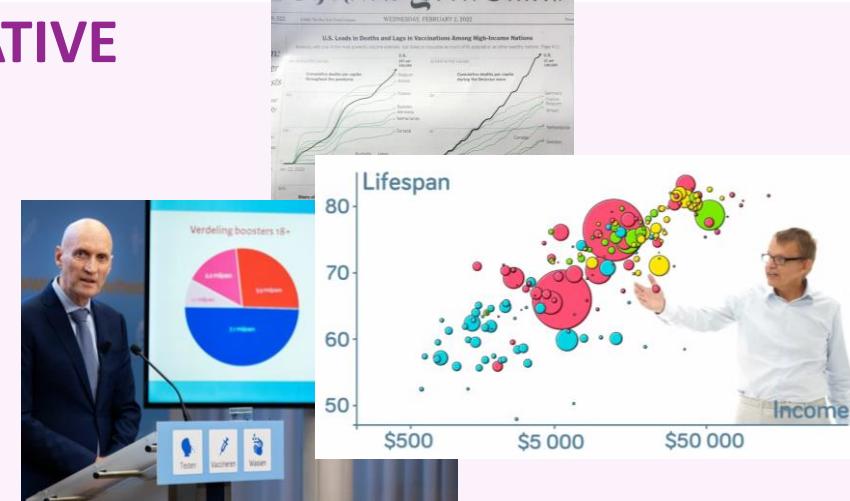
DECLARATIVE

Idea generation



CONCEPTUAL

EXPLORATORY



DATA-DRIVEN

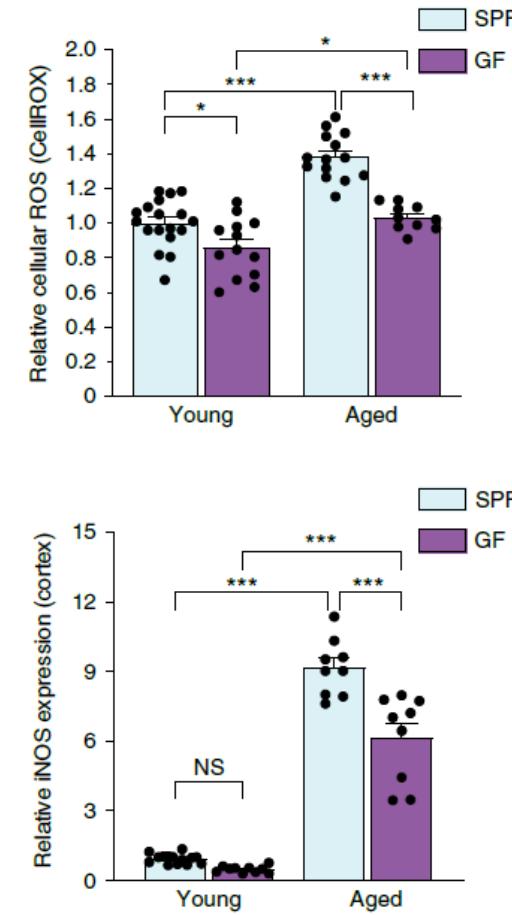
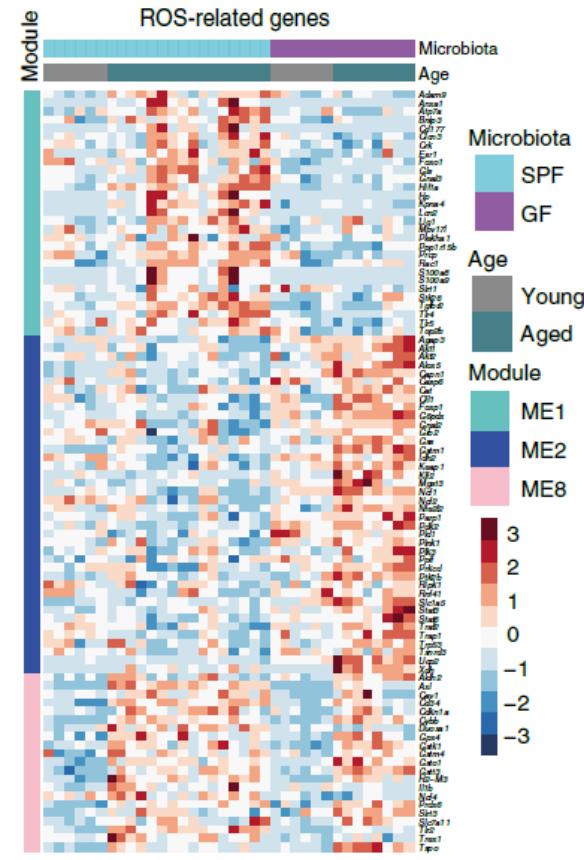
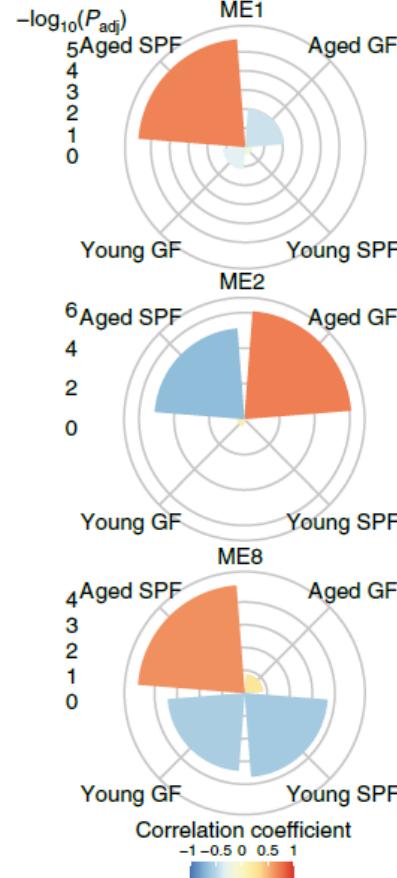
'Everyday'
DataViz

Visual
discovery



sebastiz.github.io/

Visual communication in academic research
is not *just* **data** charts...



You can also think of graphical abstracts or posters...

How does sleep disturbance affect hemodialysis patients?

Kidney Medicine

Methods and Cohort

Adults on maintenance HD with OSA ($n = 36$)

Polysomnogram & Epworth Sleepiness Scale (ESS)

Interview to explore patient experience ($n = 26$)

Findings

Severity of sleep apnea did not affect patients' sleep duration, sleep efficiency or ESS.

However,
70% reported broken sleep

62% felt unrefreshed upon wakening

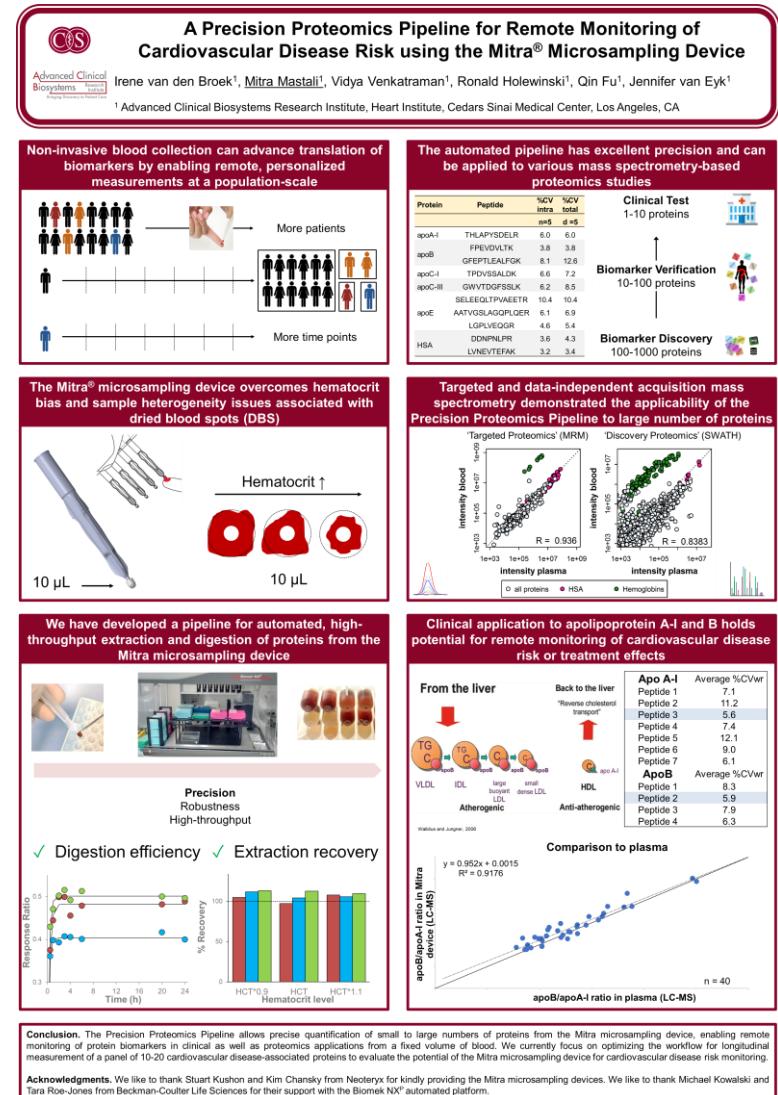
Themes from Interview

- Broken sleep
- Feeling unrefreshed
- Impact of poor sleep
- Having to "soldier on"

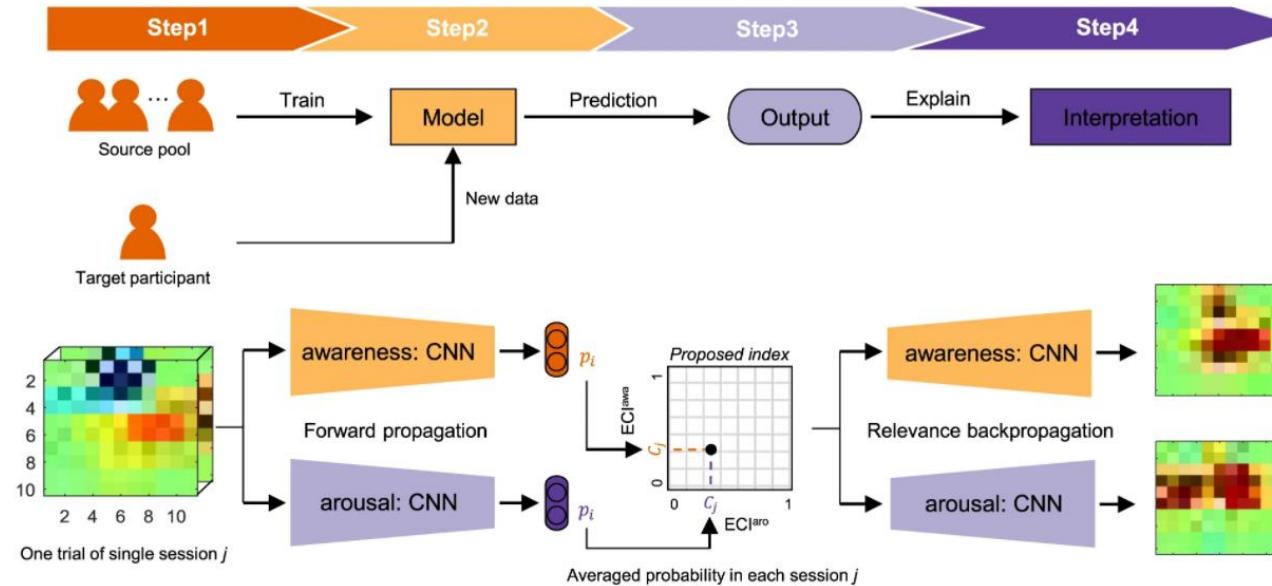
Conclusion: Sleep disturbance is common and has a profound impact on health and QoL of hemodialysis patients. The conflicting message between patient interview and self-reported questionnaires indicate a need for multidisciplinary approaches and improved patient communication to truly capture the health needs of individuals.

Reference: Chu G, Price E, Paech G, Choi P and McDonald V. Sleep apnea in maintenance hemodialysis: a mixed methods study. *Kidney Medicine*, 2020

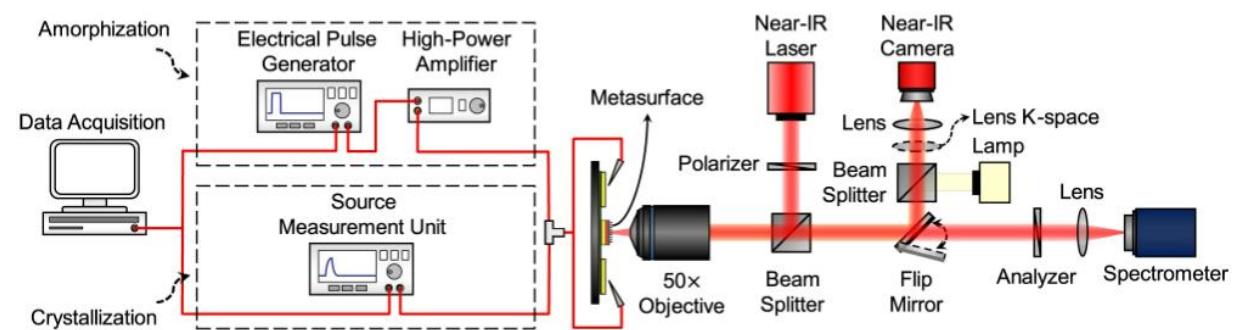
Visual Abstract by Anna R Gaddy, MD @AnnaGaddy



Flow charts or schematics...

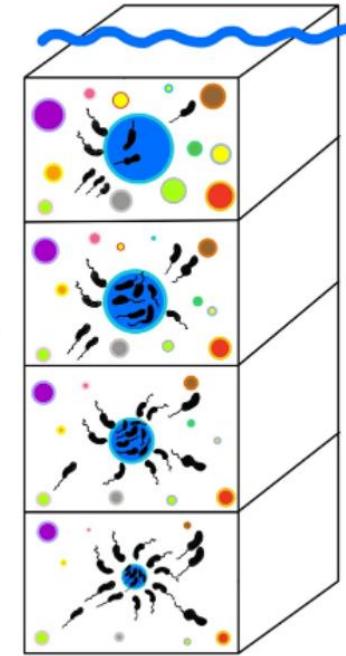
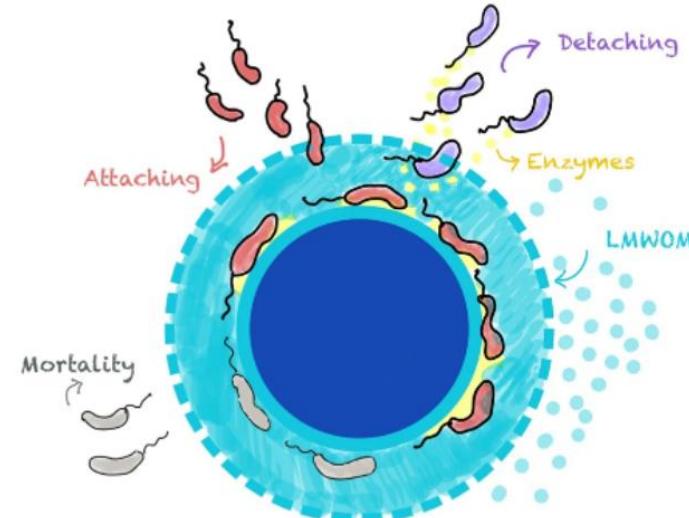
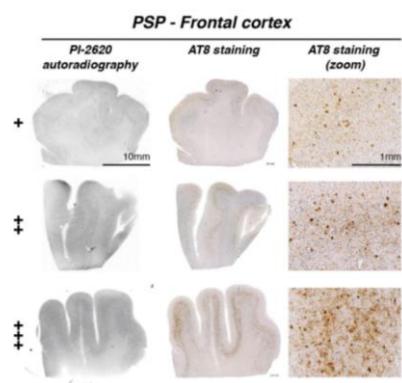
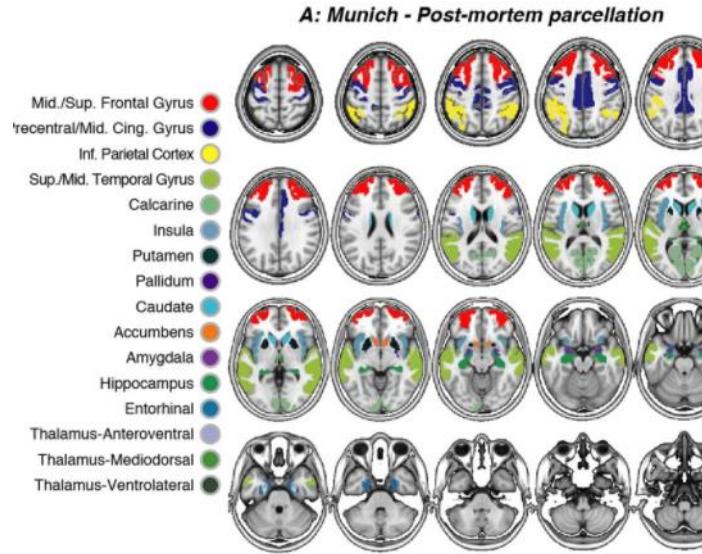


Lee M et al. (2022) Nat Commun **13**, 1064.



Abdollahramezani, S et al. Nat Commun **13**, 1696 (2022).

Illustrations, drawings, images...



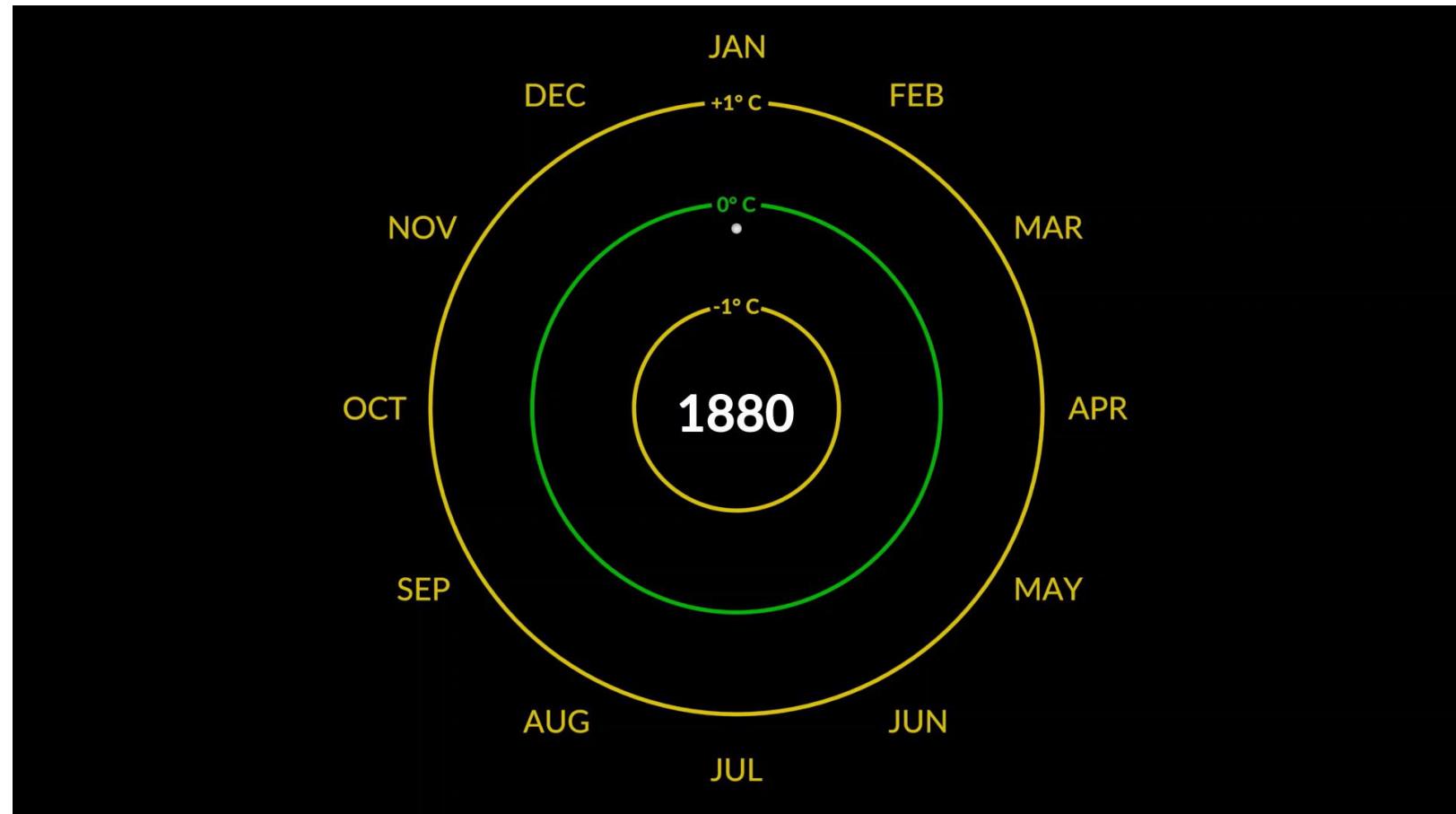
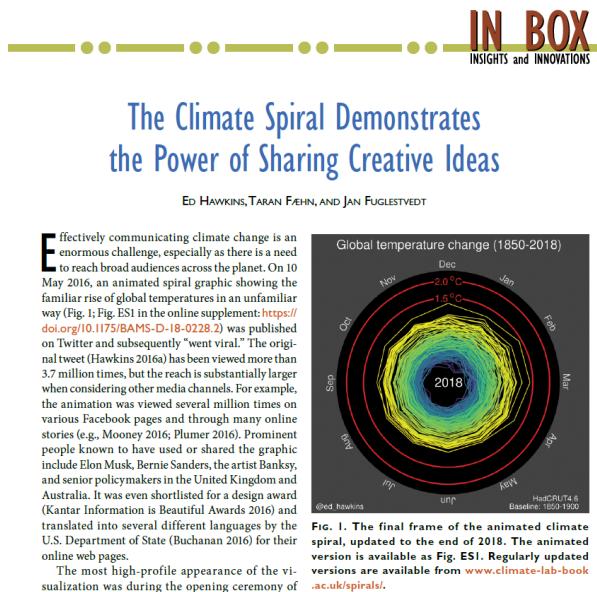
Talking about reaching a broad public!

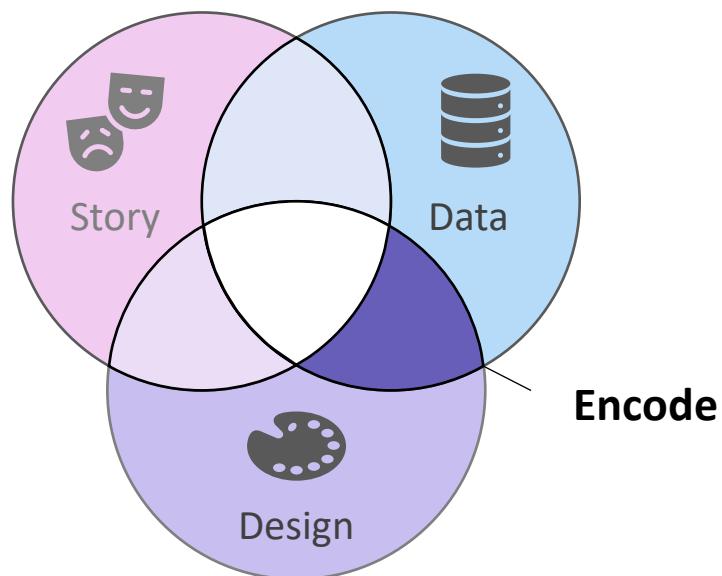


Prof. Ed Hawkins
#showyourstripes



Or animations...





How

**Visual
Communication**

in

**Academic
Research**

with

**Data
Viz**

?

Exercise

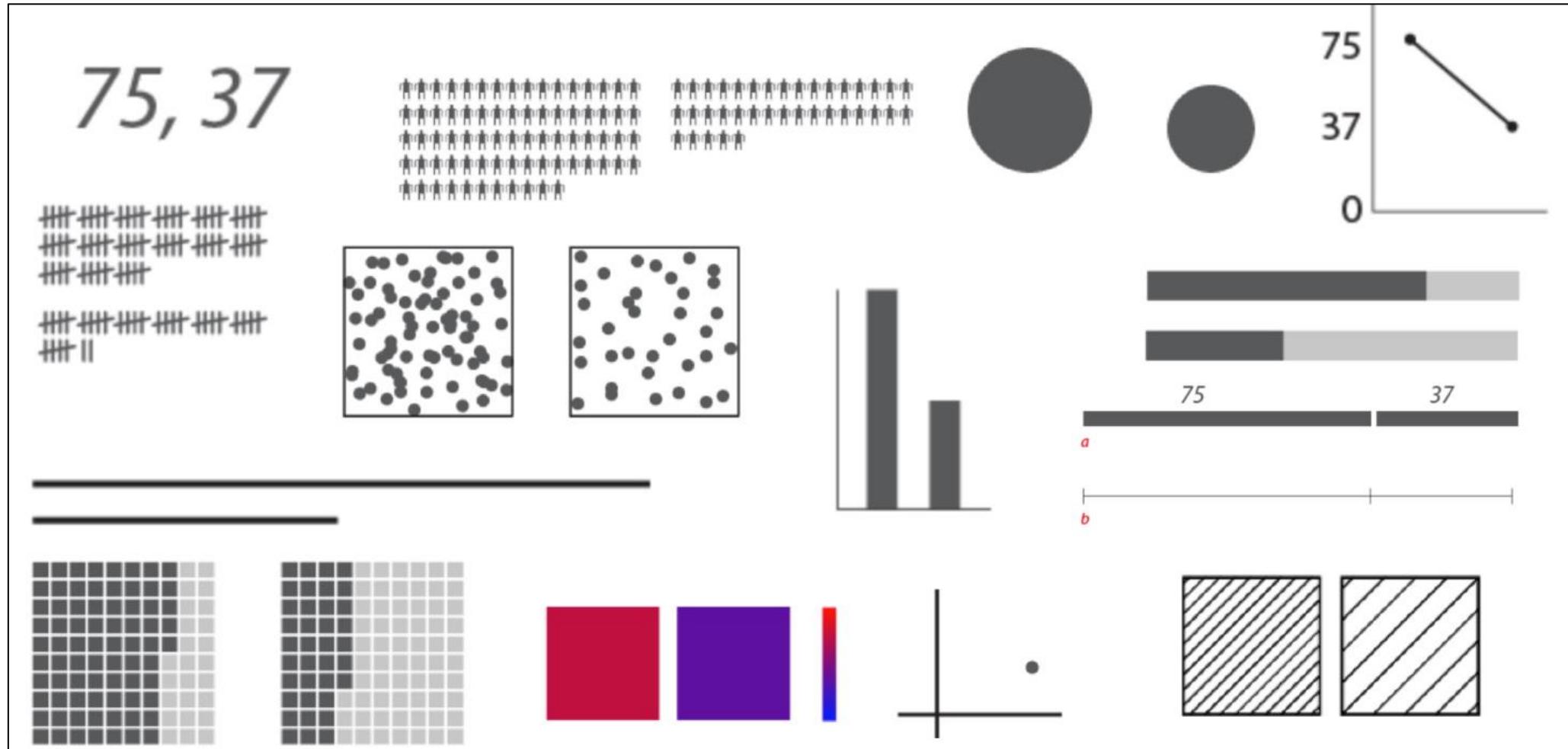
Exercise: In how many ways can you visualize two numbers?

75 and 37



5 min

Some of 45 ways to communicate two quantities



So many ways: how to choose?





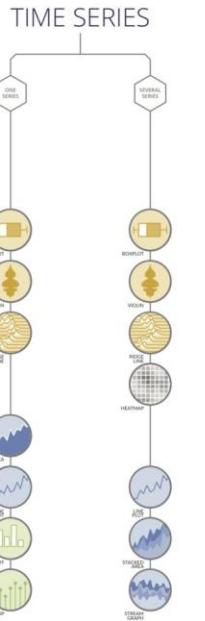
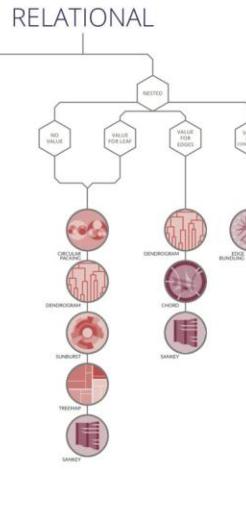
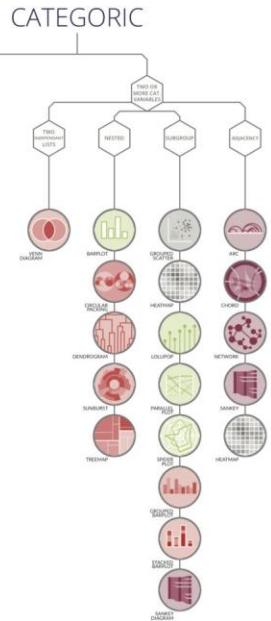
from Data to Viz

'From Data to Viz' is a classification of chart types based on input data format. It will help you find the perfect chart in three simple steps :

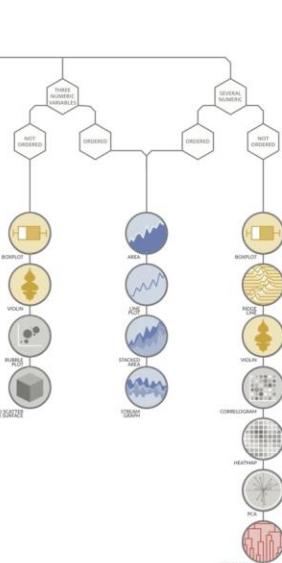
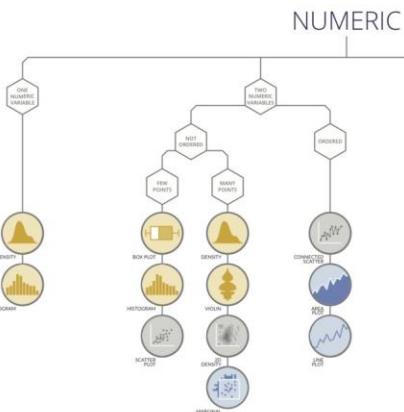
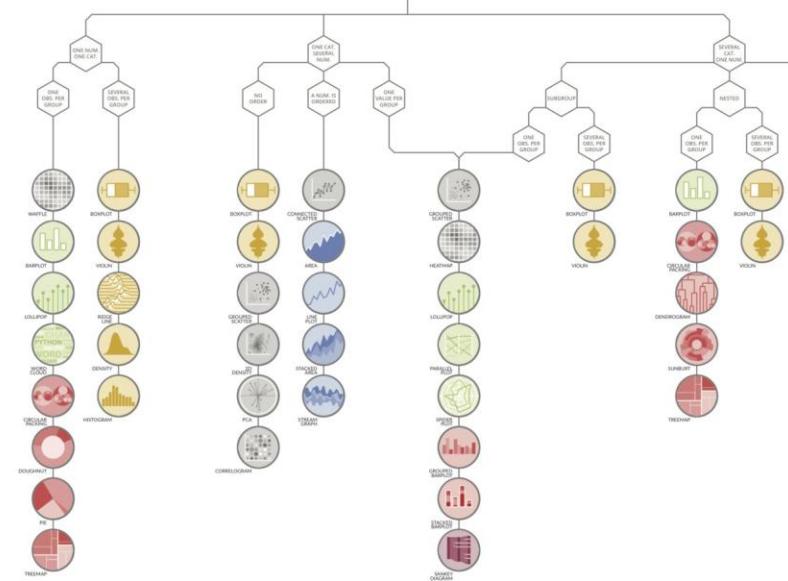
- 1 Identify what type of data you have.
 - 2 Go to the corresponding decision tree and follow it down to a set of possible charts.
 - 3 Choose the chart from the set that will suit your data and your needs best.

Dataviz is a world with endless possibilities and this project does not claim to be exhaustive. However it should provide you with a good starting point. For an interactive version and much more, visit:

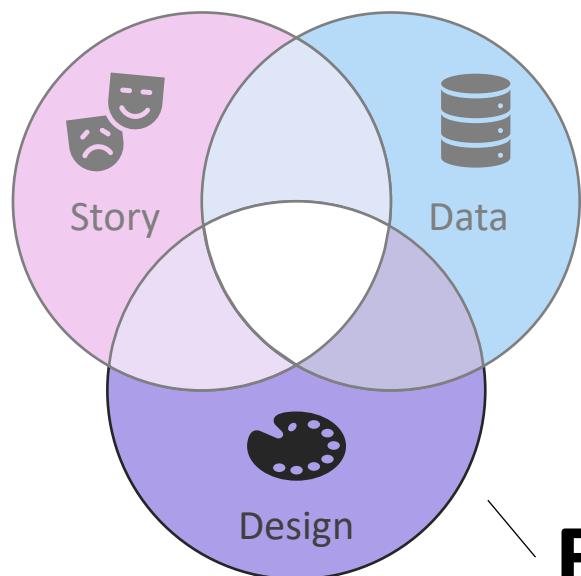
data-to-viz.com



CATEGORIC AND NUMERIC



How



Visual
Communication

in

Academic
Research

with

Data
Viz

?

Enhance: The *strengths* of our brain

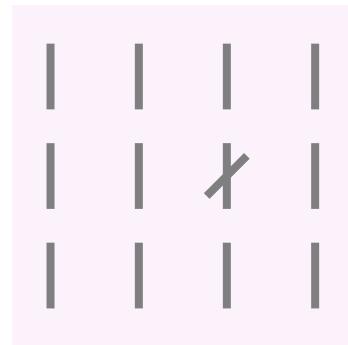
How many fives do you see?

321654643216465132168461321
646432136748651349687463123
668432616266984892536496874
651687964649843616984616957
465167412323161689312439873
421432875943869234987596387

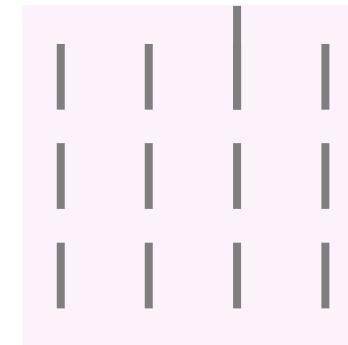
How many fives do you see?

321654643216465132168461321
646432136748651349687463123
668432616266984892536496874
651687964649843616984616957
465167412323161689312439873
421432875943869234987596387

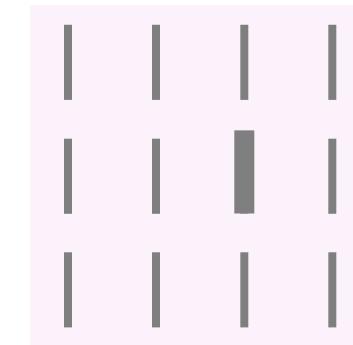
Pre-attentive attributes make your audience see what you want them to see before even seeing it!



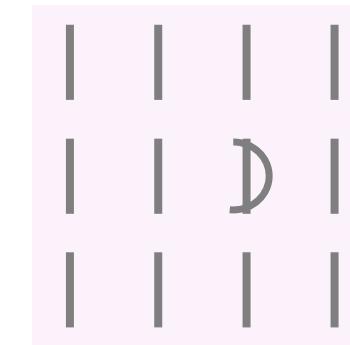
Orientation



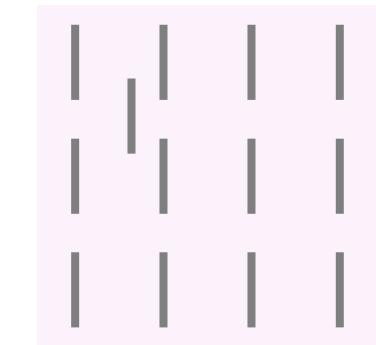
Length



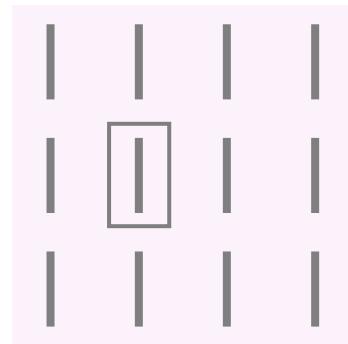
Width



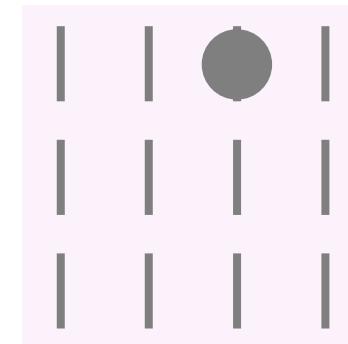
Curvature



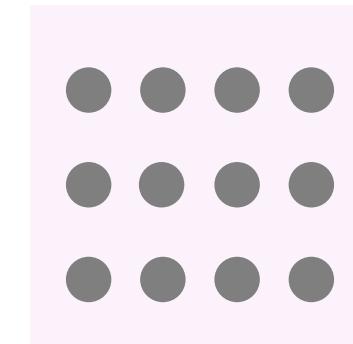
Position



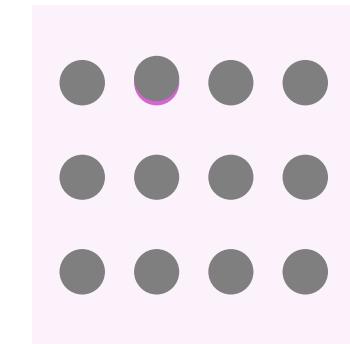
Enclosure



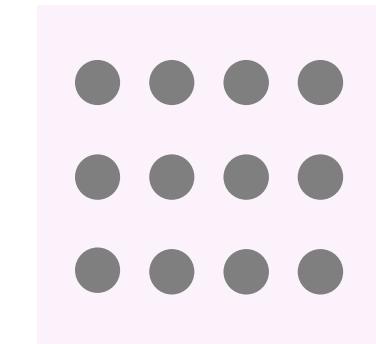
Shape



Size



Color



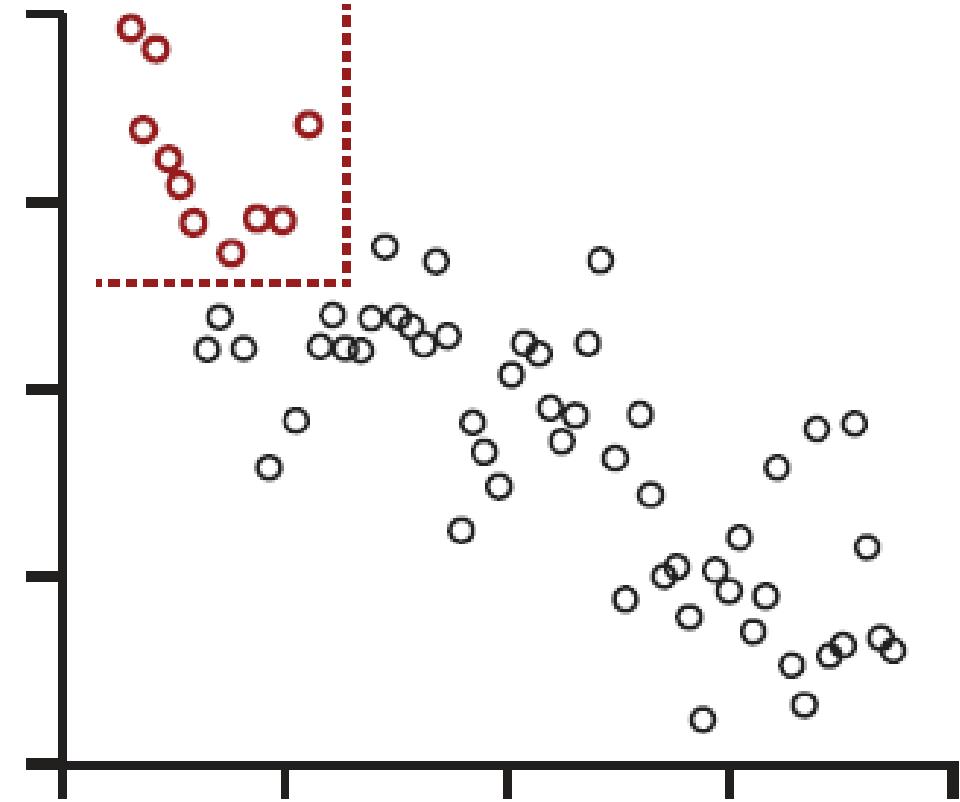
Motion

You can use pre-attentive attributes in tables

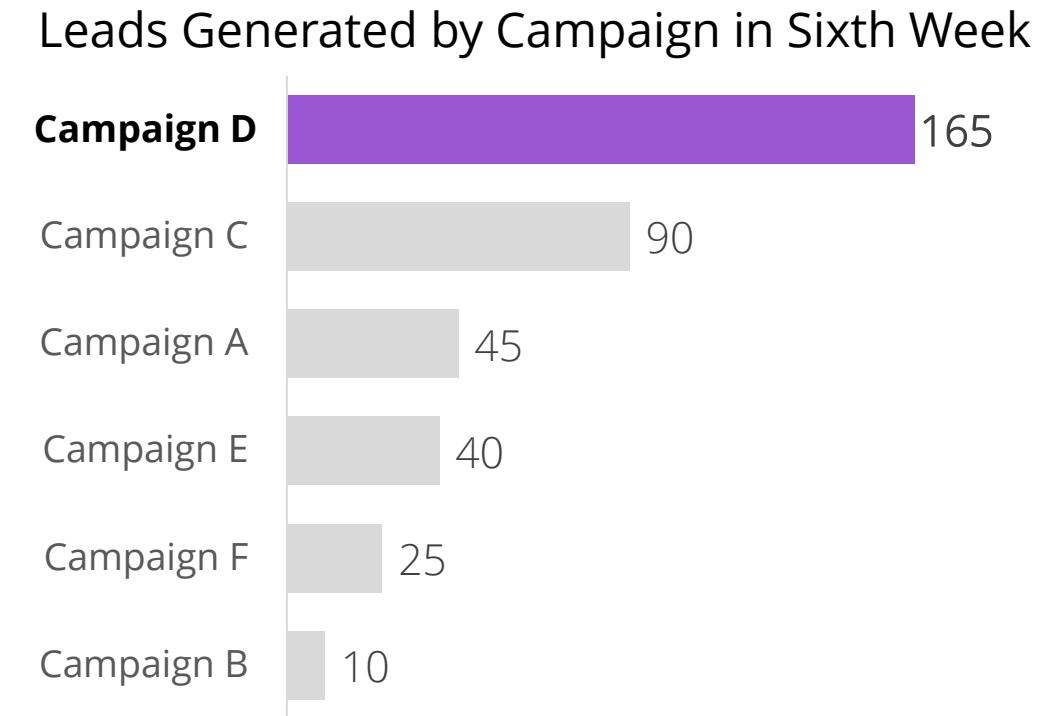
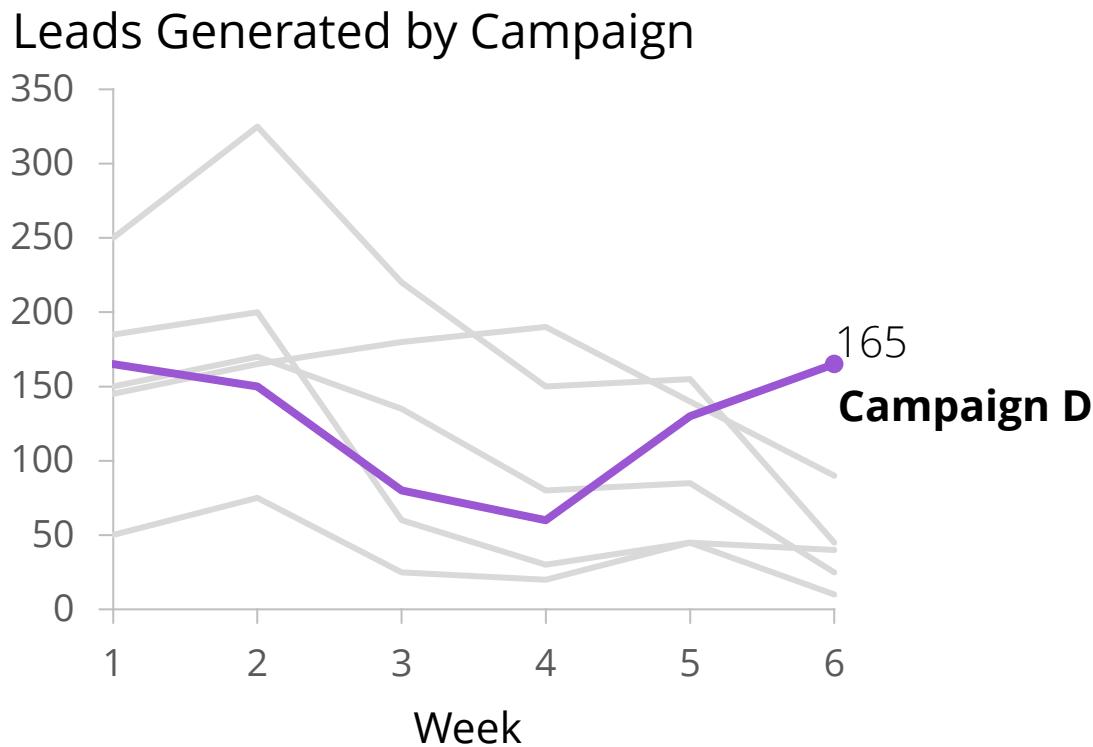
Color name	RGB (1-255)
Black	0, 0, 0
Orange	230, 159, 0
Sky blue	86, 180, 233
Bluish green	0, 158, 115
Blue	0, 114, 178
Vermillion	213, 94, 0

You can use pre-attentive attributes in tables and charts

Color name	RGB (1-255)
Black	0, 0, 0
Orange	230, 159, 0
Sky blue	86, 180, 233
Bluish green	0, 158, 115
Blue	0, 114, 178
Vermillion	213, 94, 0



To draw attention to what's important



You can also use pre-attentive attributes in text.

How can you leverage the use of pre-attentive attributes?

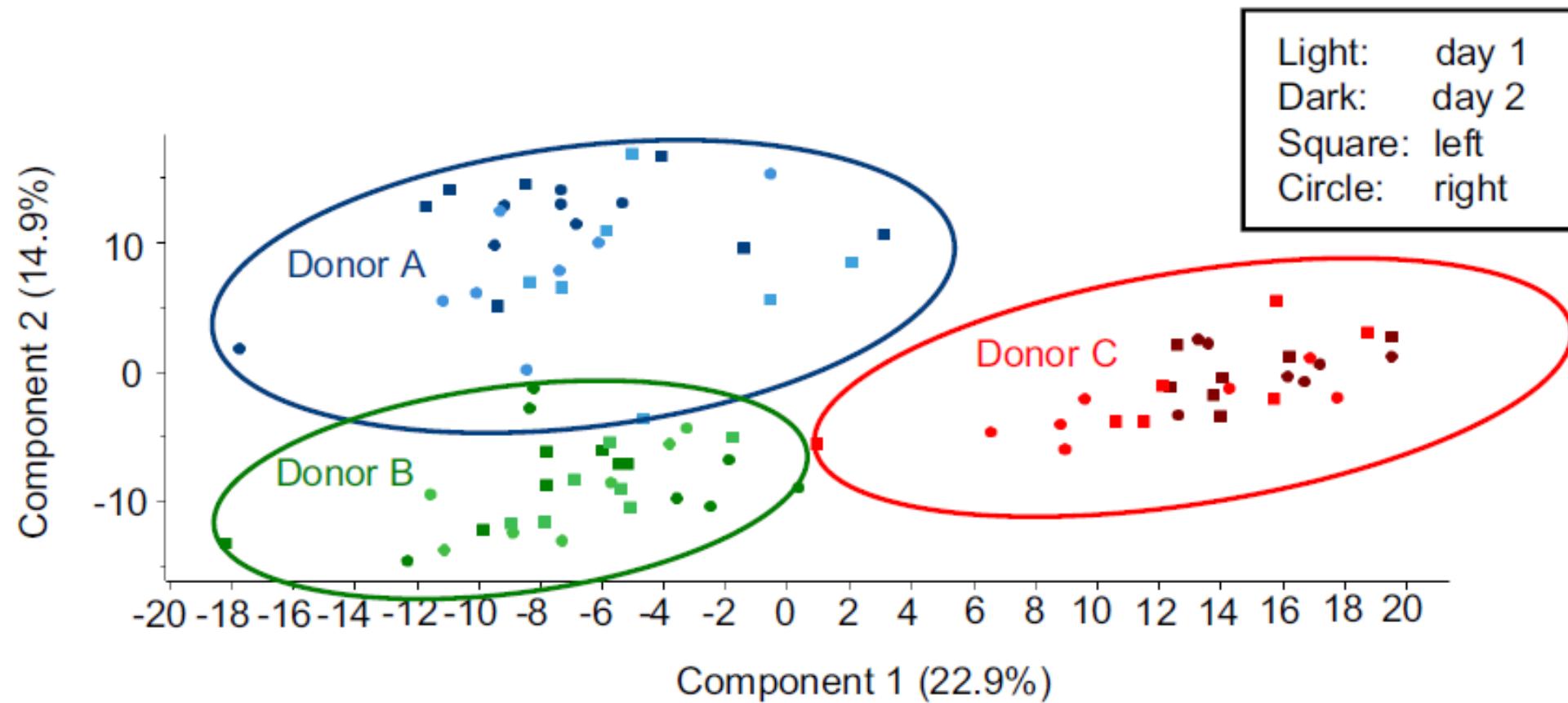
Without pre-attentive attributes you just have plain text. Or a plain figure. There is no guidance for the reader or viewer.

With pre-attentive attributes, you can:

- (1) *Draw your audience attention to where you want them to look*, and
- (2) *Create a visual hierarchy of information*

This counts for **graphs** as well as **text**!

Pre-attentive attributes can create a visual hierarchy

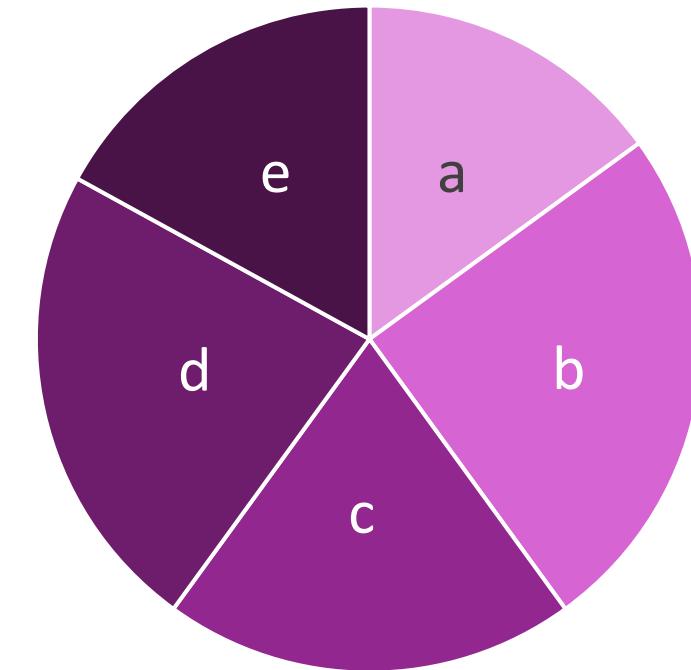


Pre-attentive attributes can be used to map quantities, e.g. by length or by angle

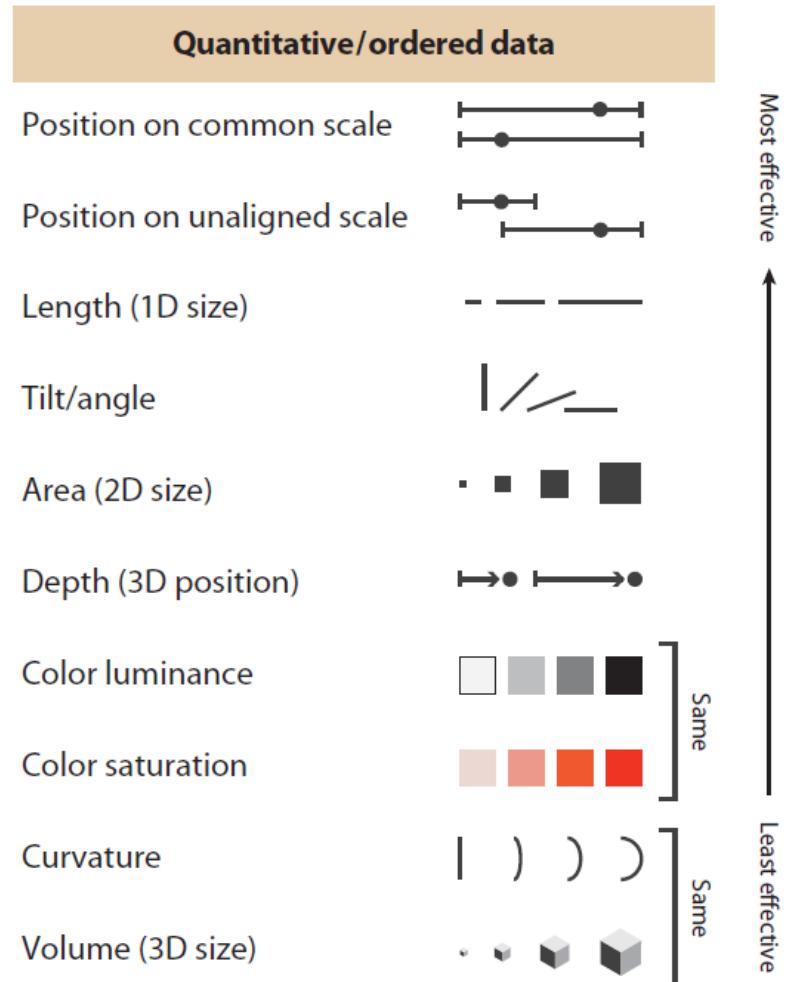
Which bar is highest?



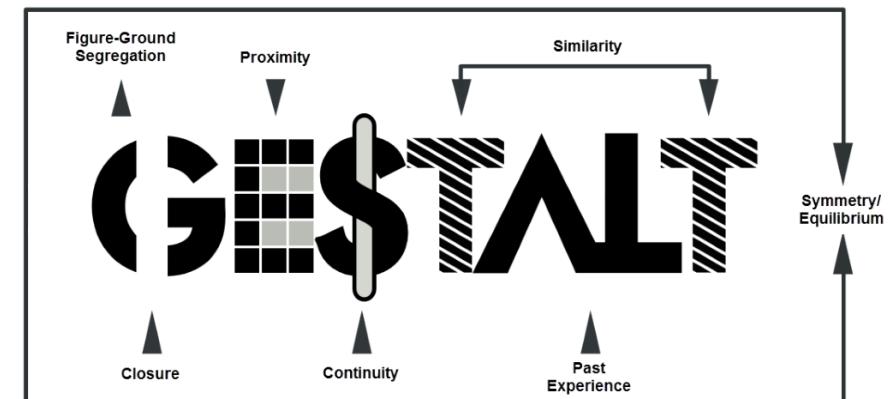
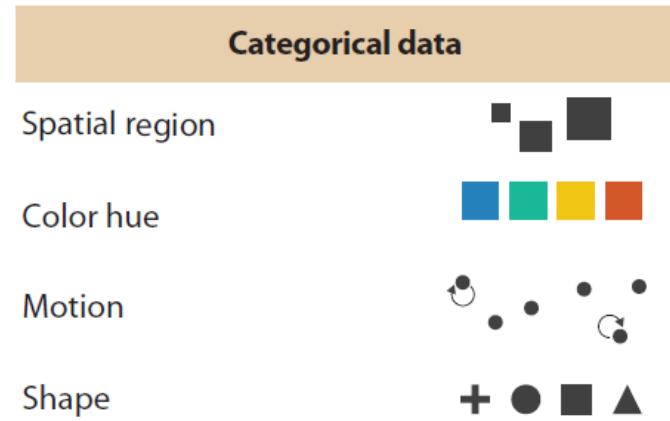
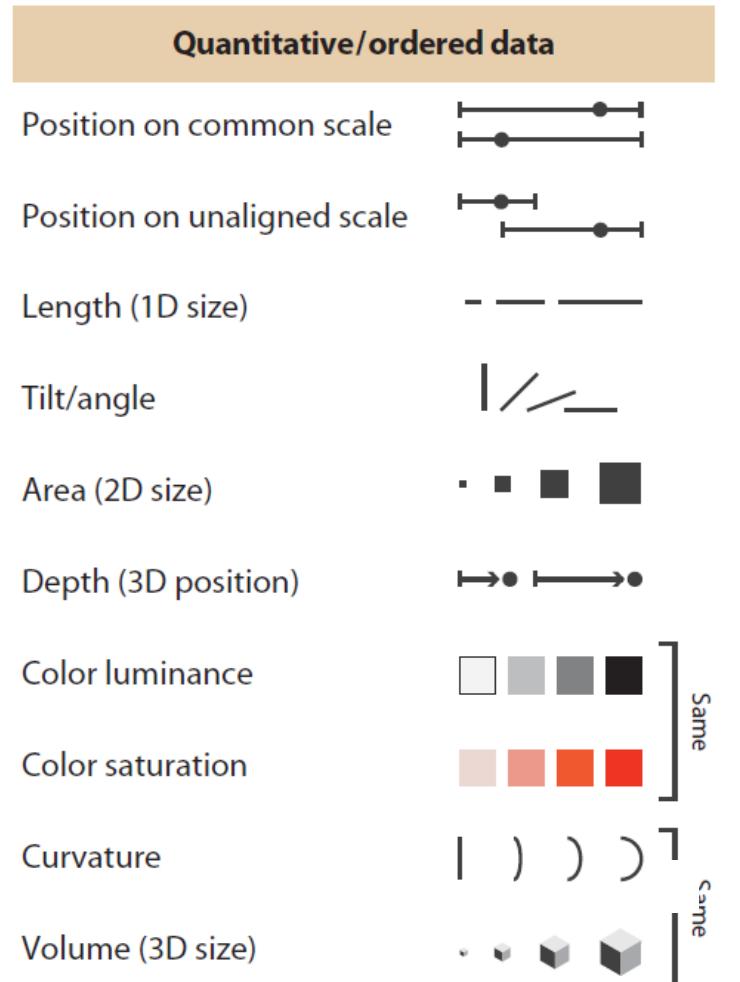
Which slice is biggest?



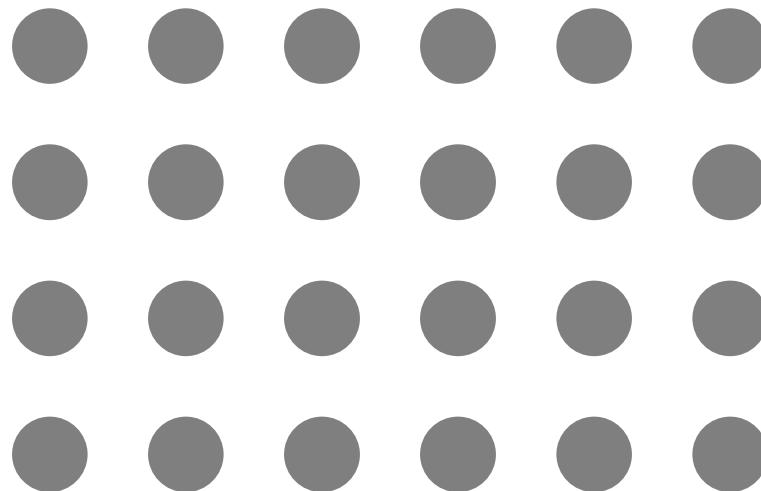
Some attributes are more effective to differentiate **quantities**



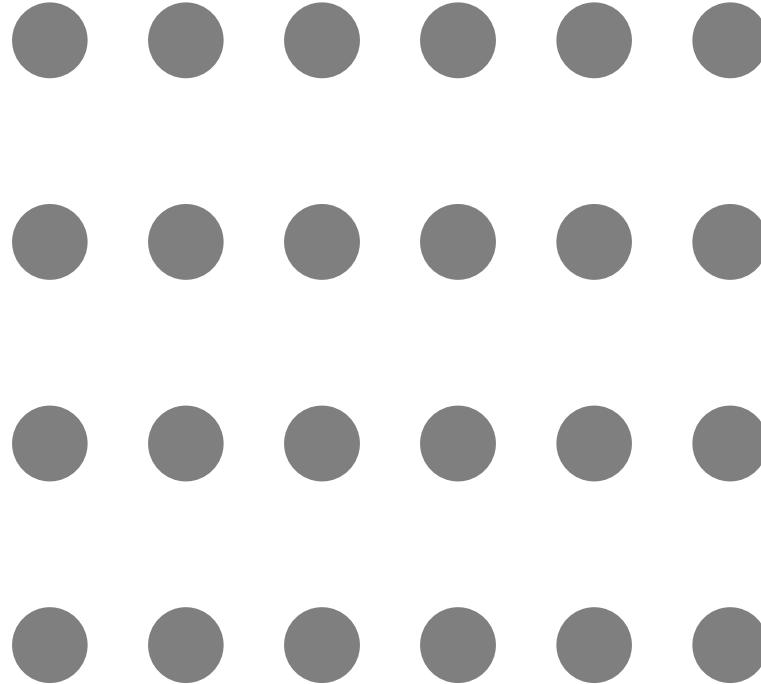
Some attributes are more effective to differentiate **quantities** or **categories**



The **Gestalt** principles describe how our brain groups information

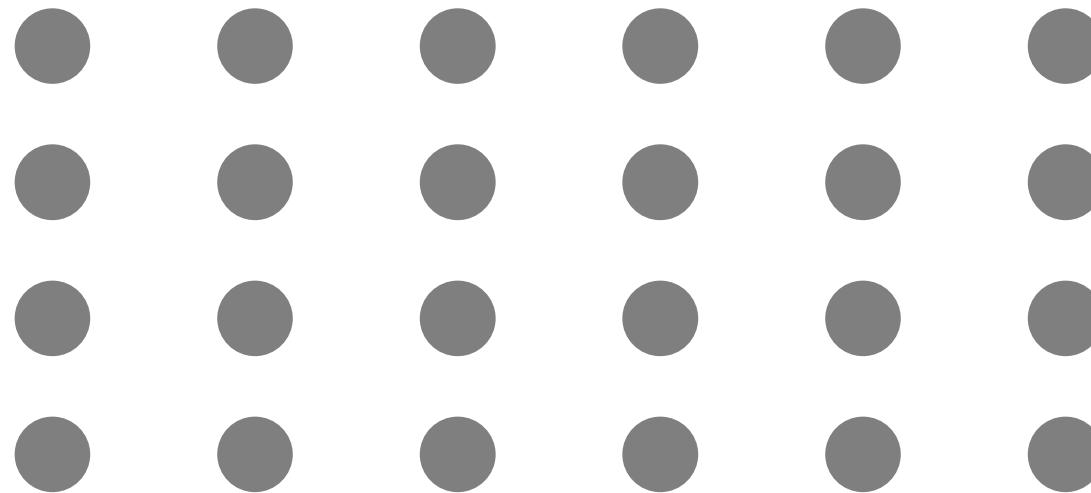


The **Gestalt** principles describe how our brain groups information



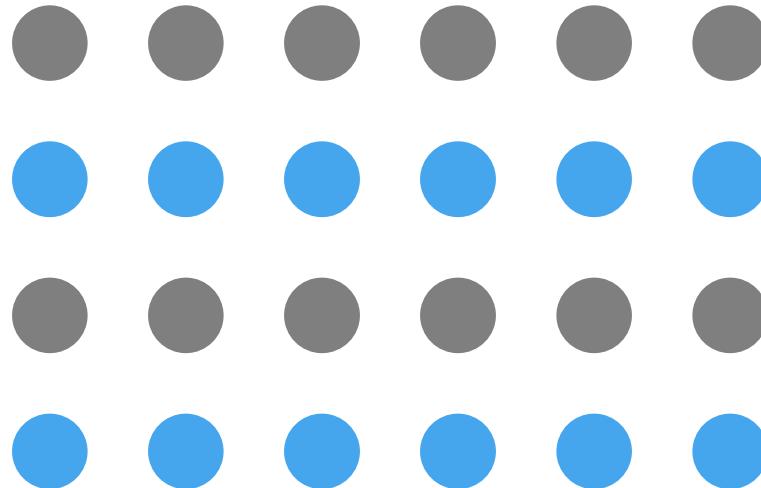
Proximity

The **Gestalt** principles describe how our brain groups information



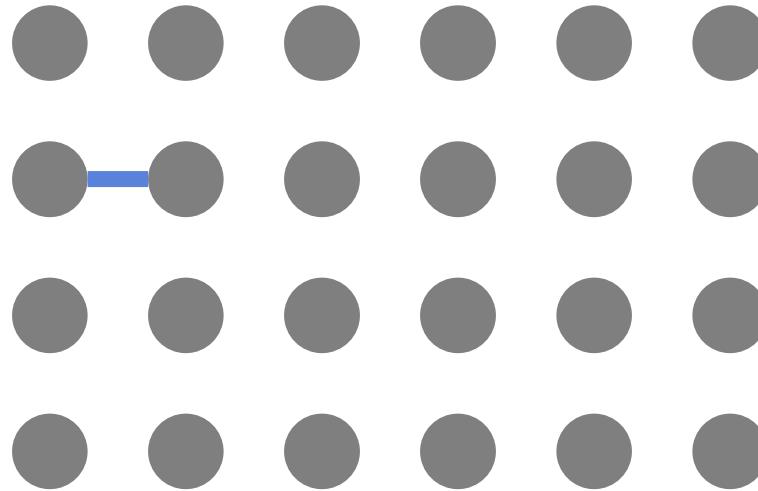
Proximity

The **Gestalt** principles describe how our brain groups information



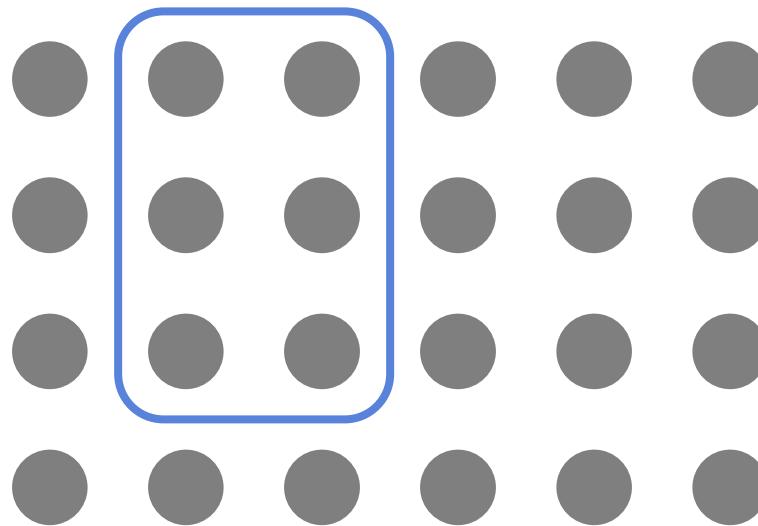
Similarity

The **Gestalt** principles describe how our brain groups information



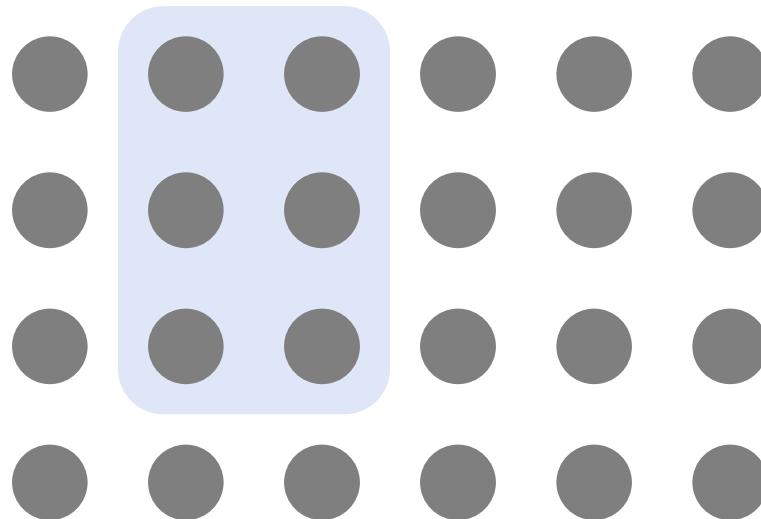
Connection

The **Gestalt** principles describe how our brain groups information



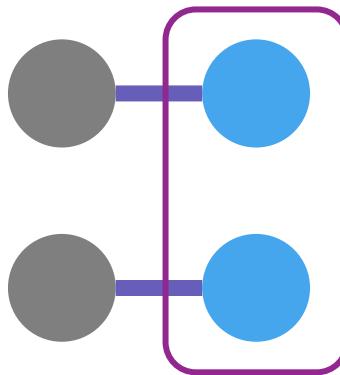
Enclosure

The **Gestalt** principles describe how our brain groups information

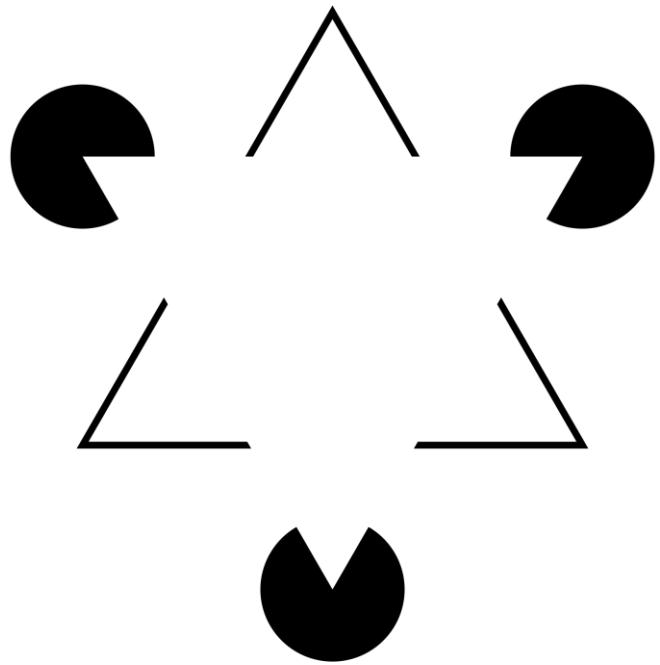


Enclosure

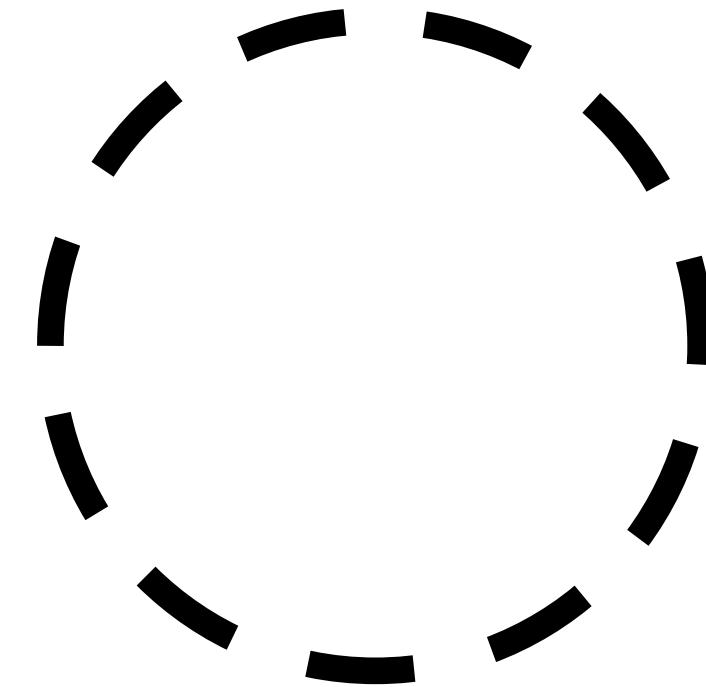
Some visual properties have stronger associative values than others...



We're also good at seeing things that may not explicitly exist!

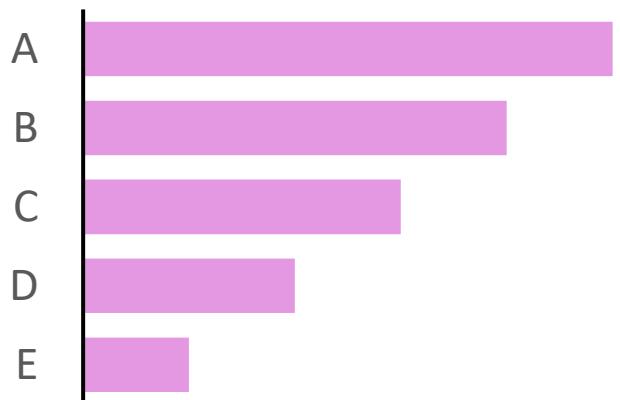
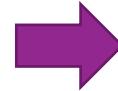
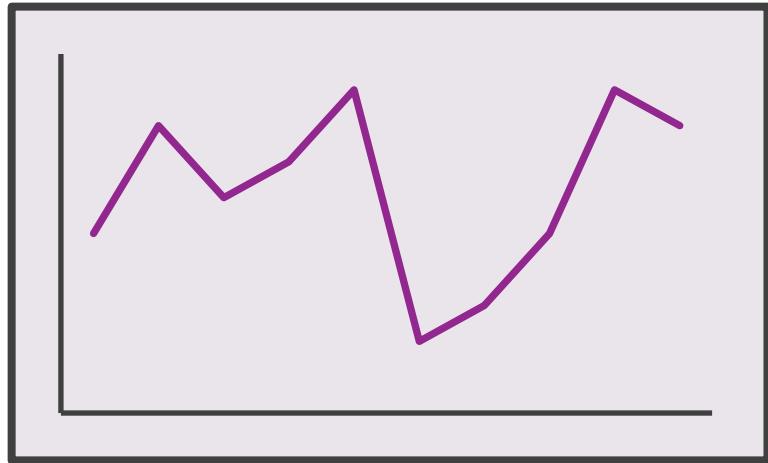


Continuity

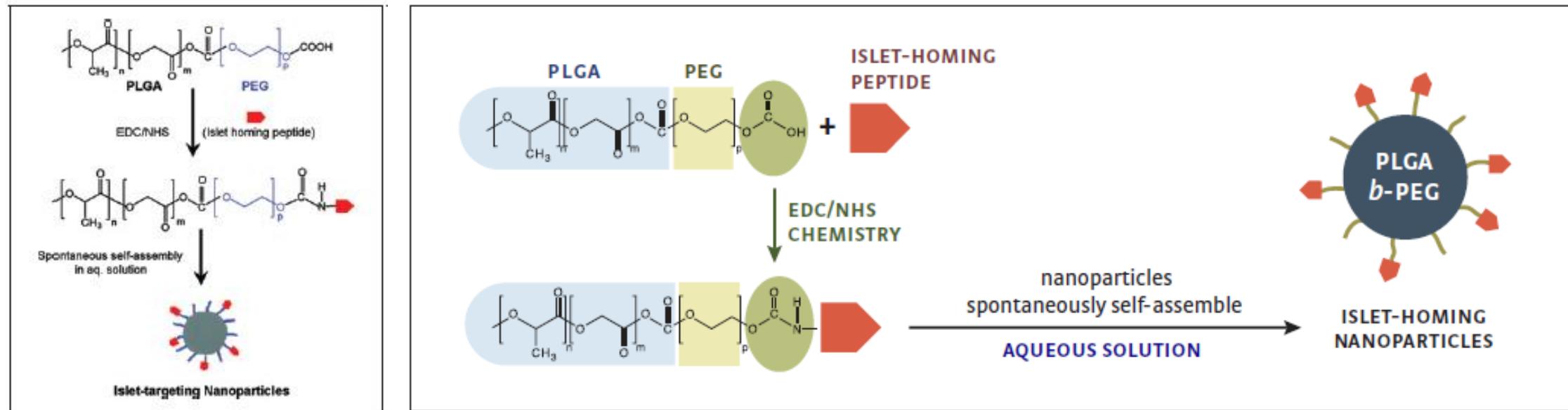


Closure

You can use the **Continuity** and **Closure** principle
to remove unnecessary elements in your chart



The Gestalt principles also apply to graphical abstracts



Shapes and colors used to identify similar structures.

But...



Exercise

I Notice. I Wonder.
A look at your work.



I Notice. I Wonder.



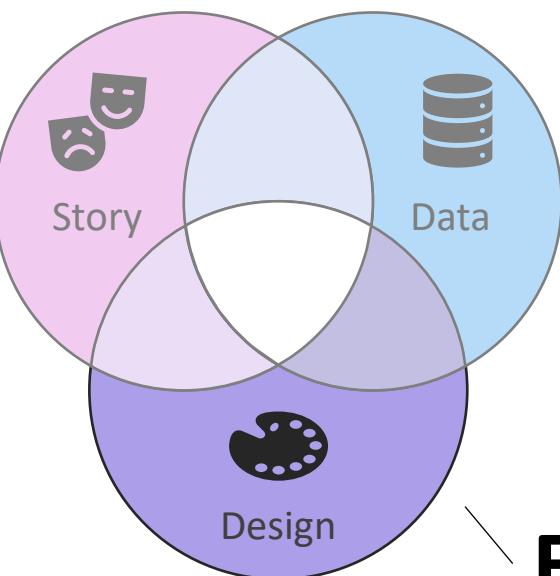
Notice: (<10 sec)

- What is the first thing that catches your eye?
- What is your first impression?



Wonder: (< 1 min)

- Do you get an idea about the message? What is it?
- What questions come to mind?



How

Visual
Communication

in

Academic
Research

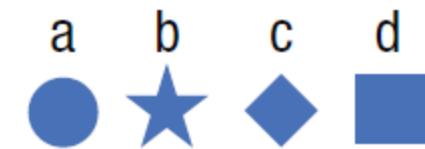
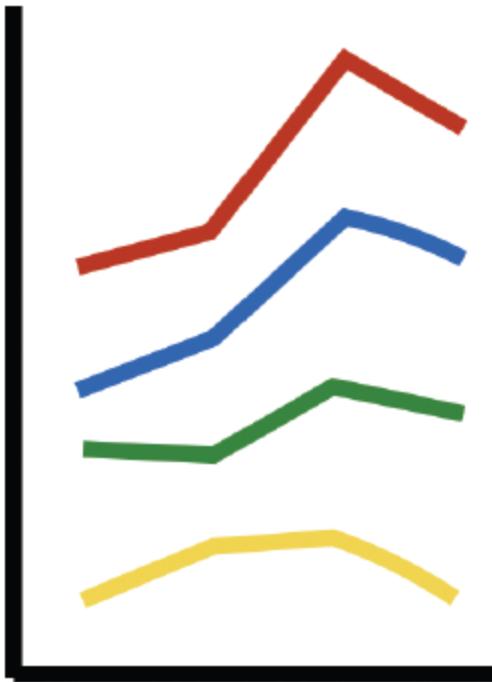
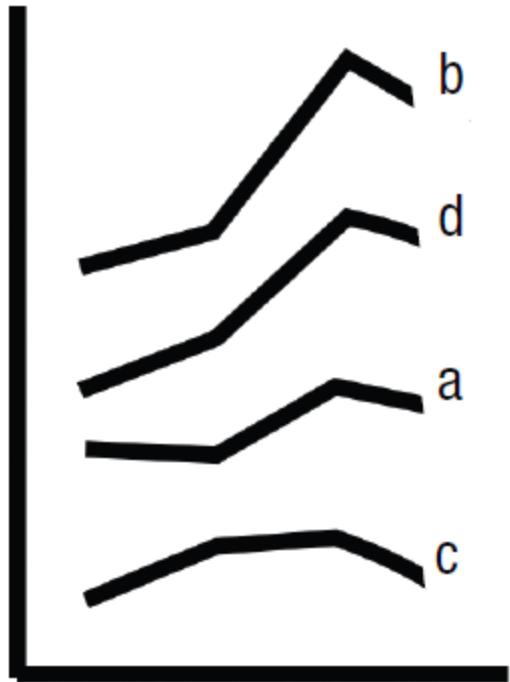
with

Data
Viz

?

Enhance: The *weaknesses* of our brain

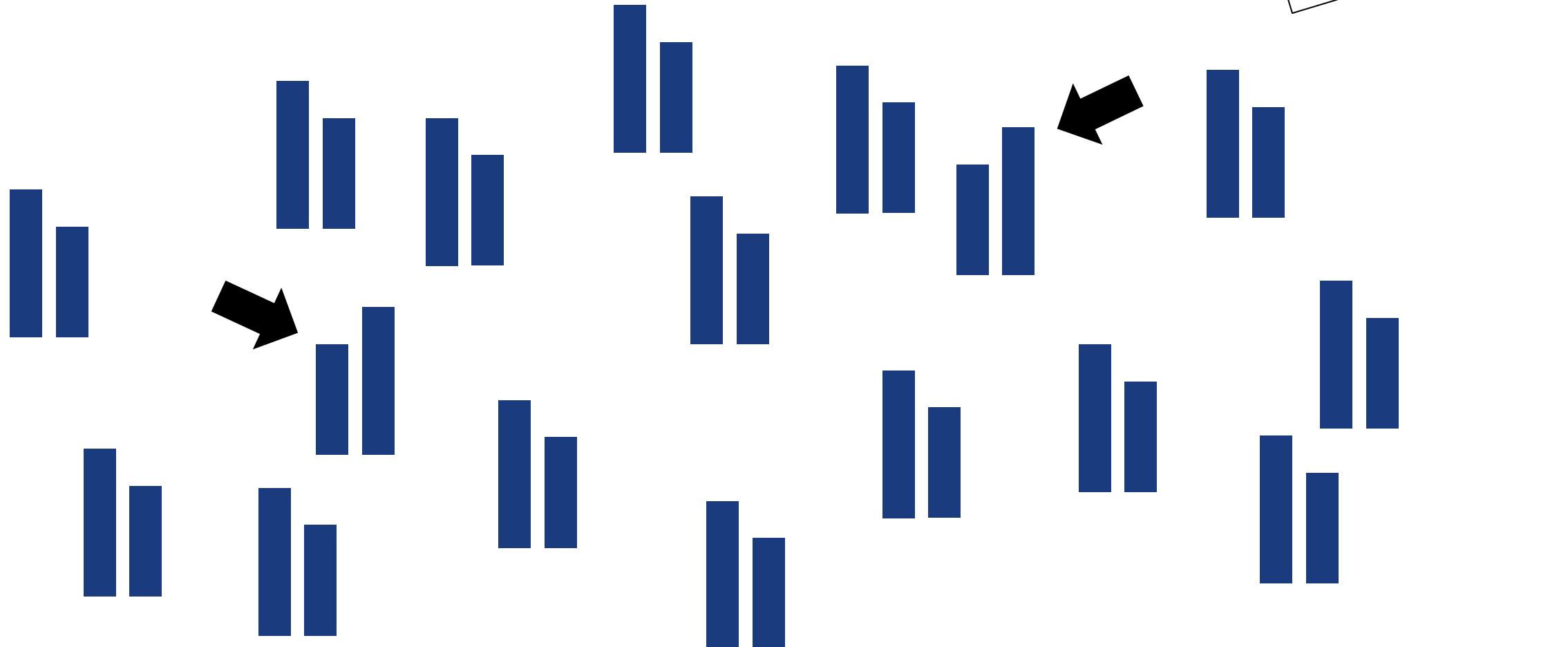
What is the order from top-to-bottom?



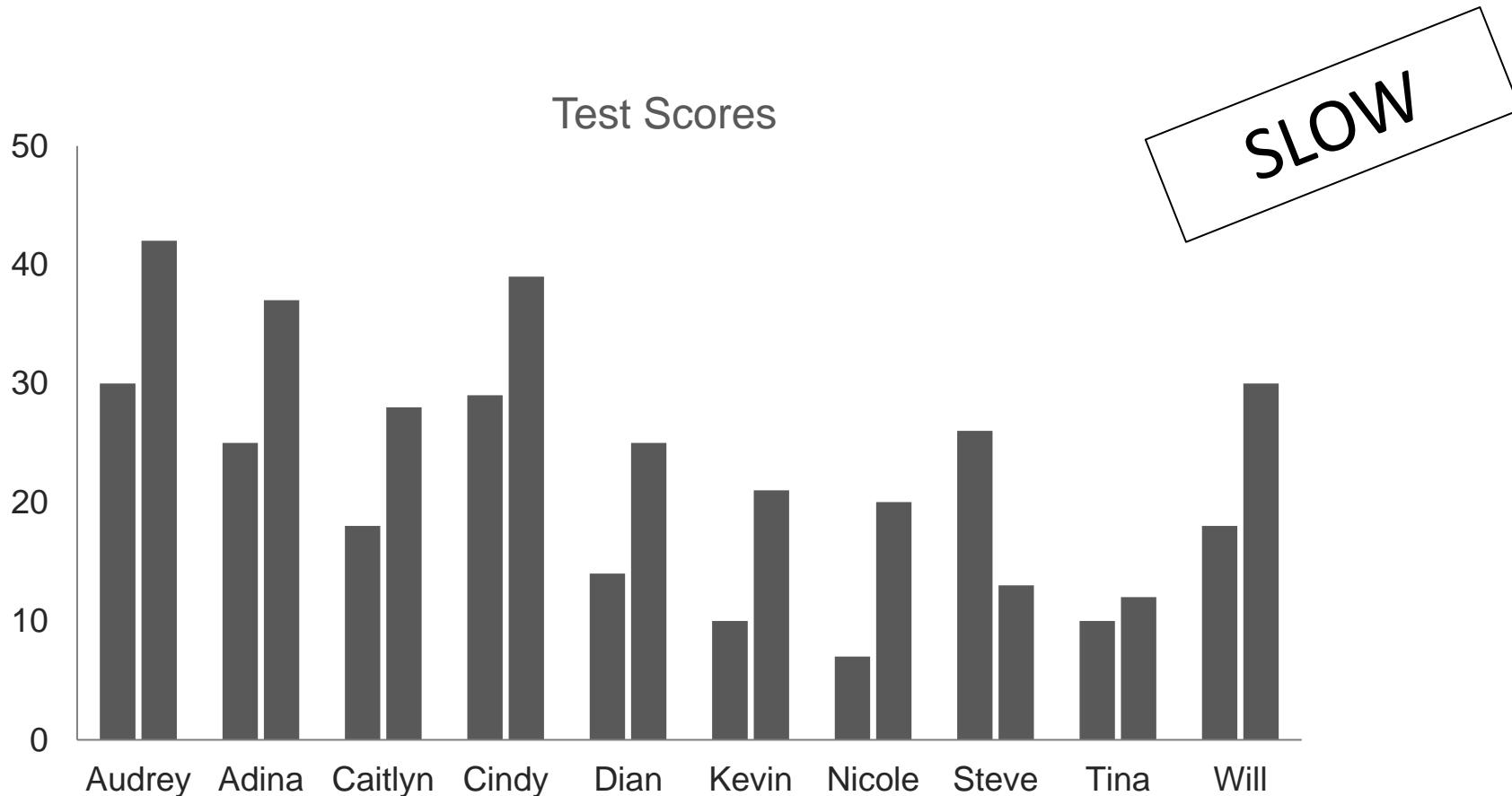
Do you see small-big or big-small?



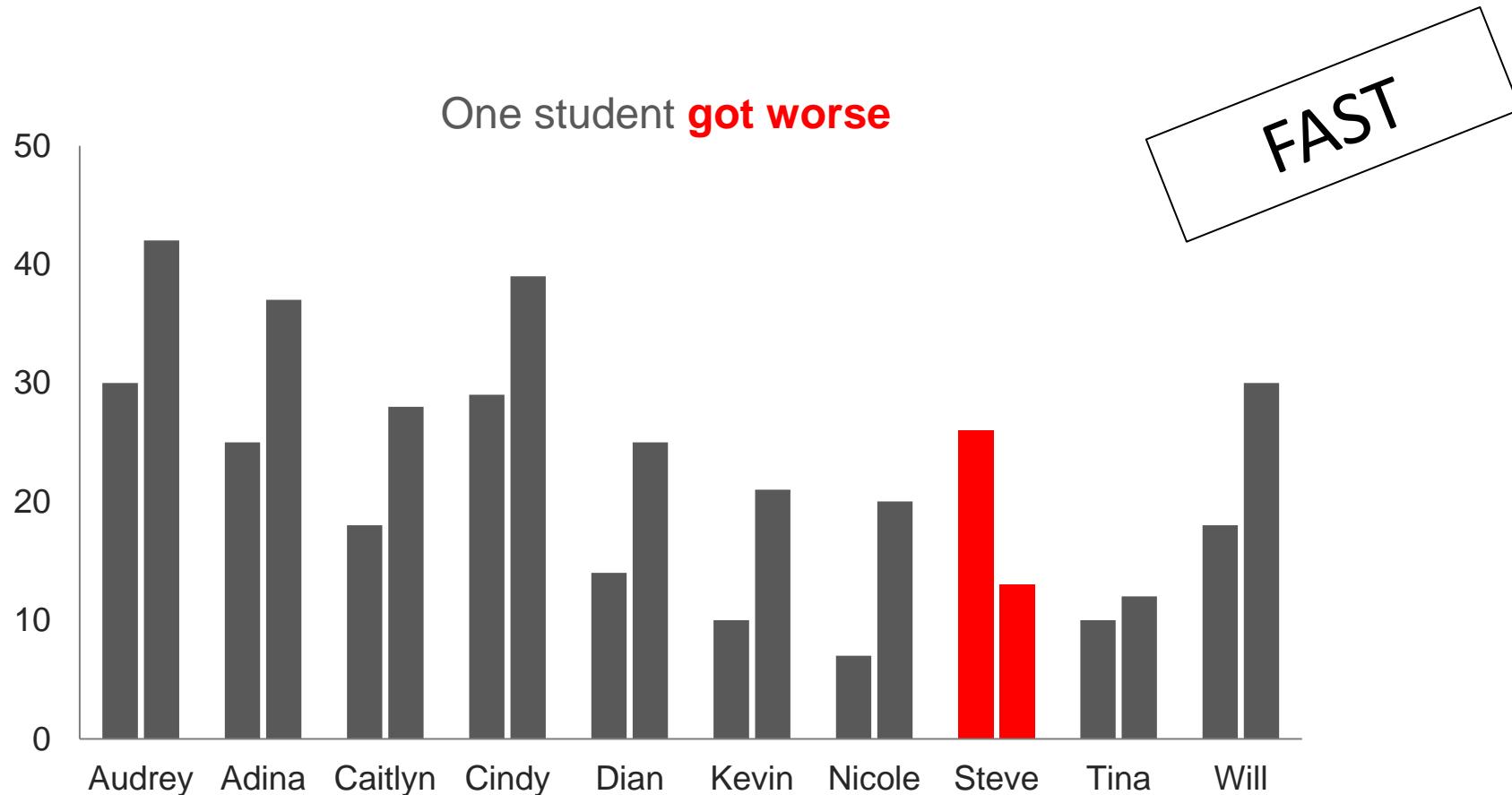
Which one is not like the others?



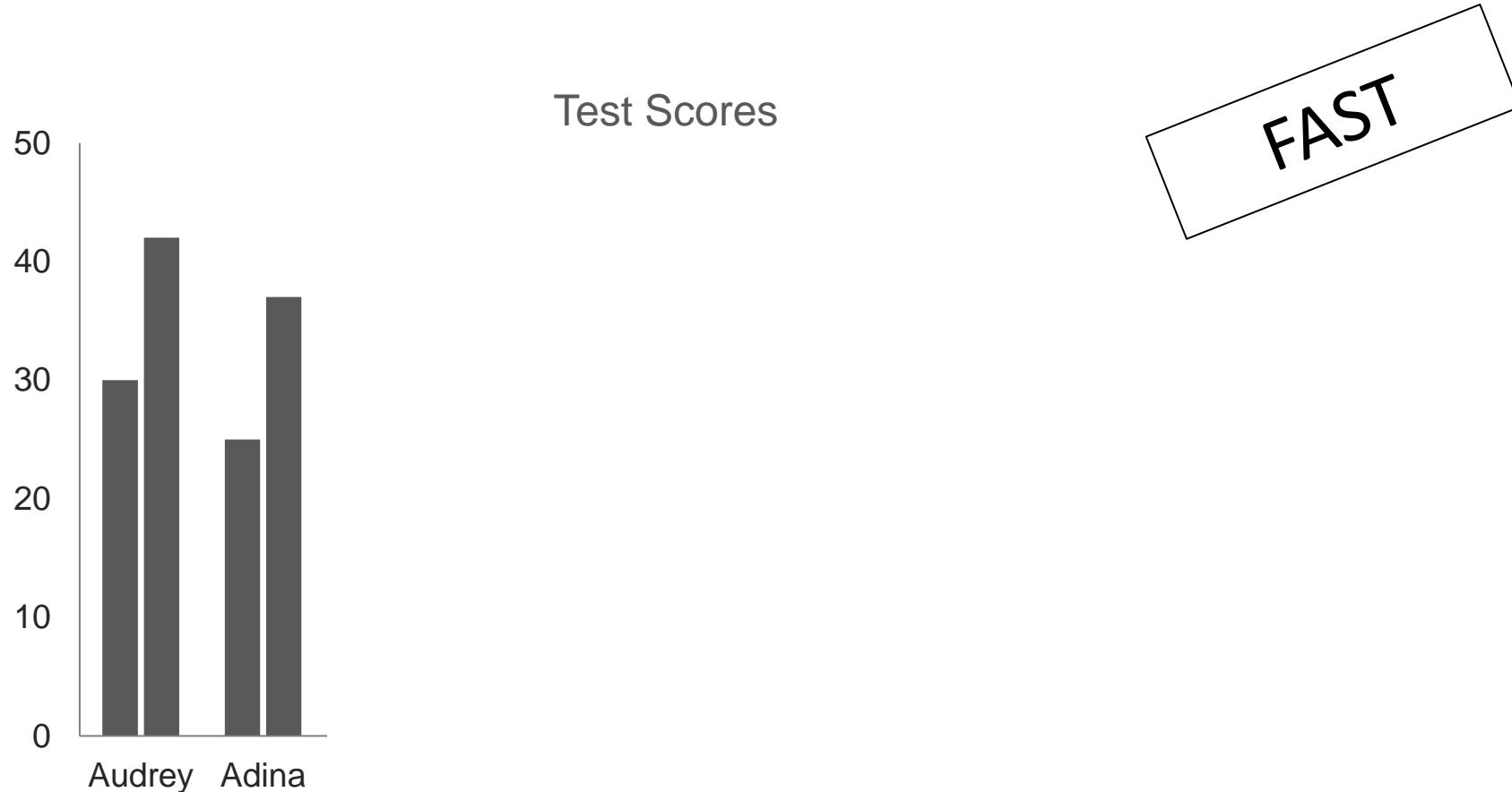
Which student got worse?



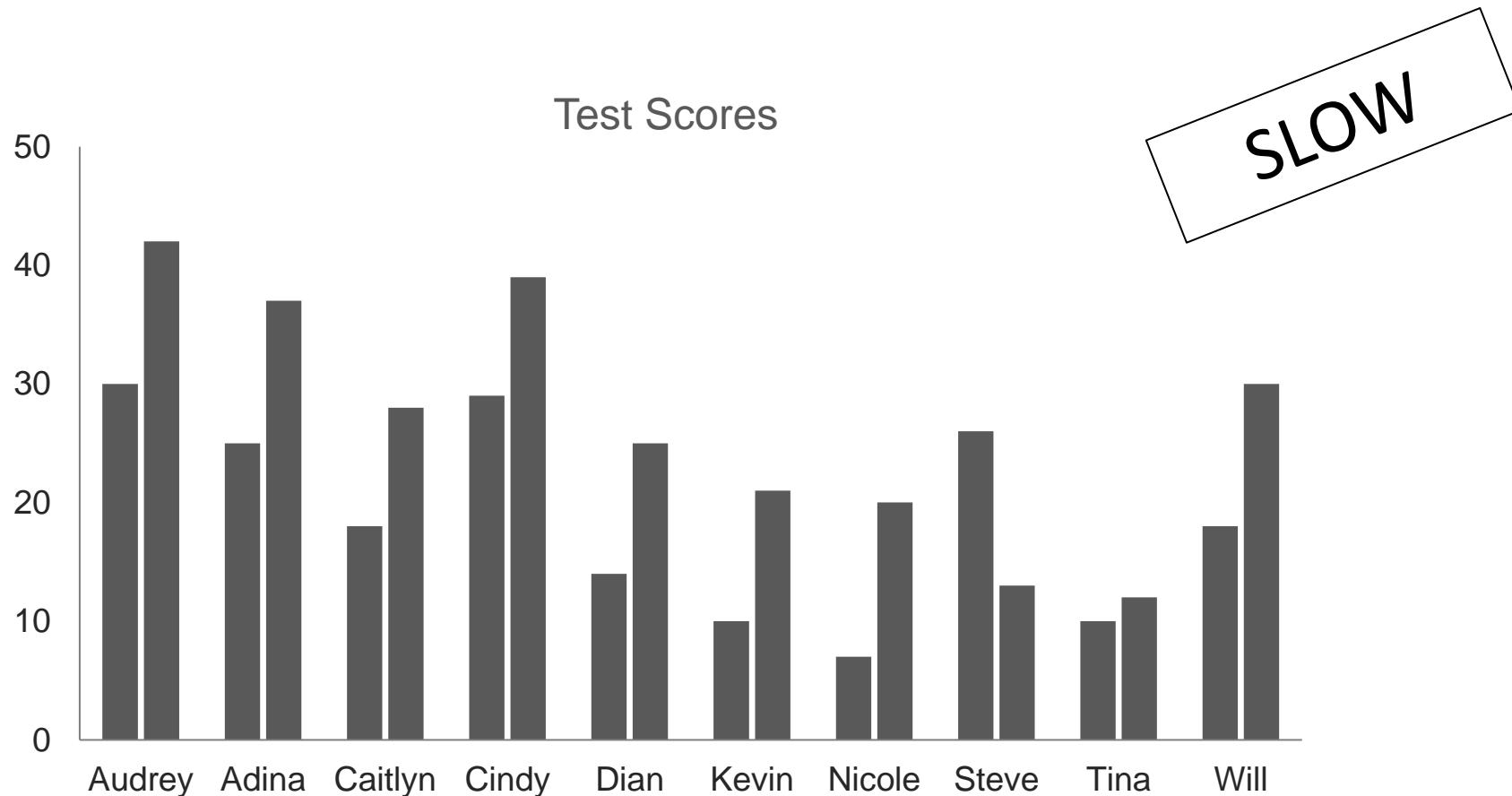
Which student got worse?



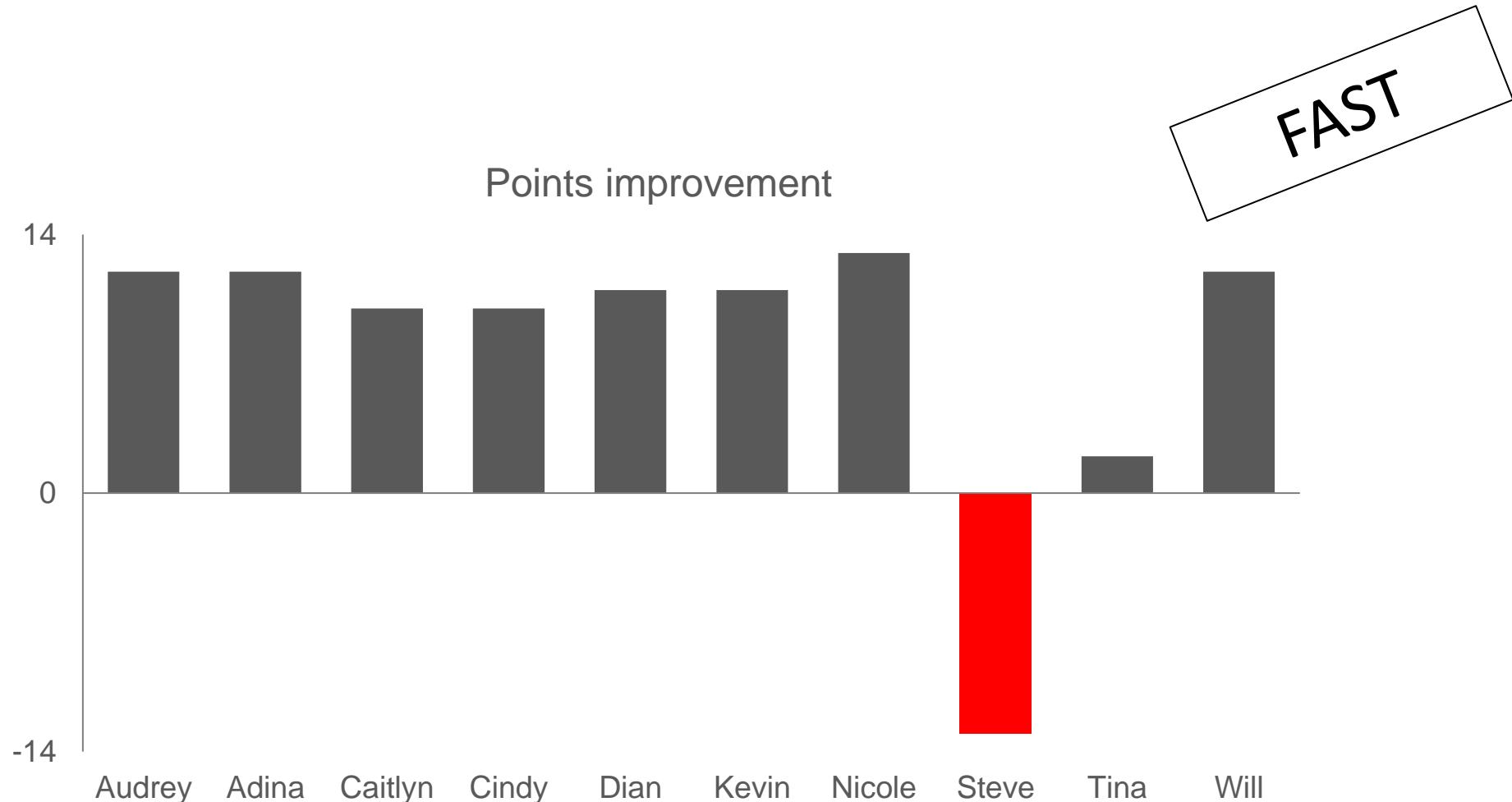
What is the average improvement?



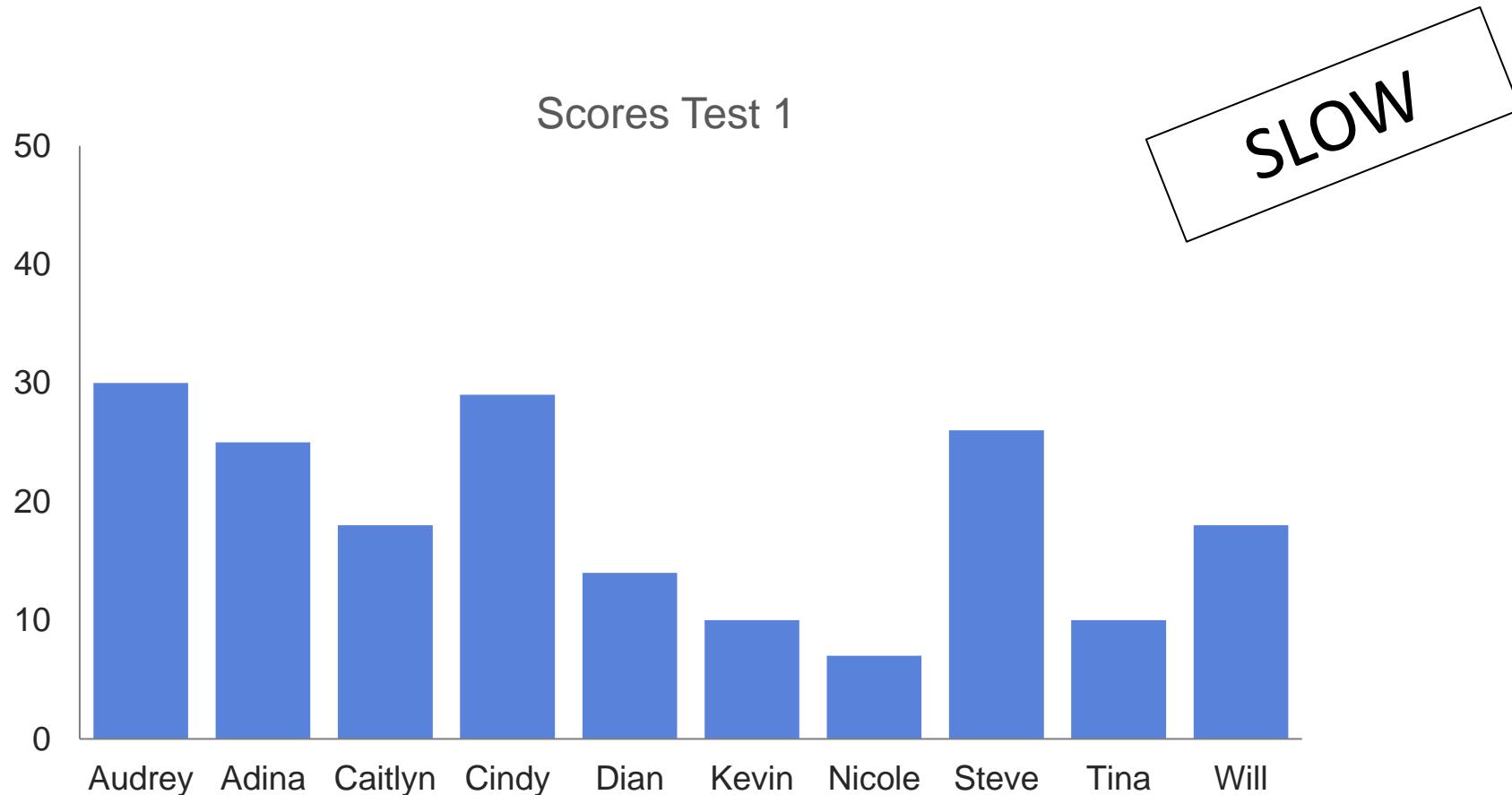
What is the average improvement?



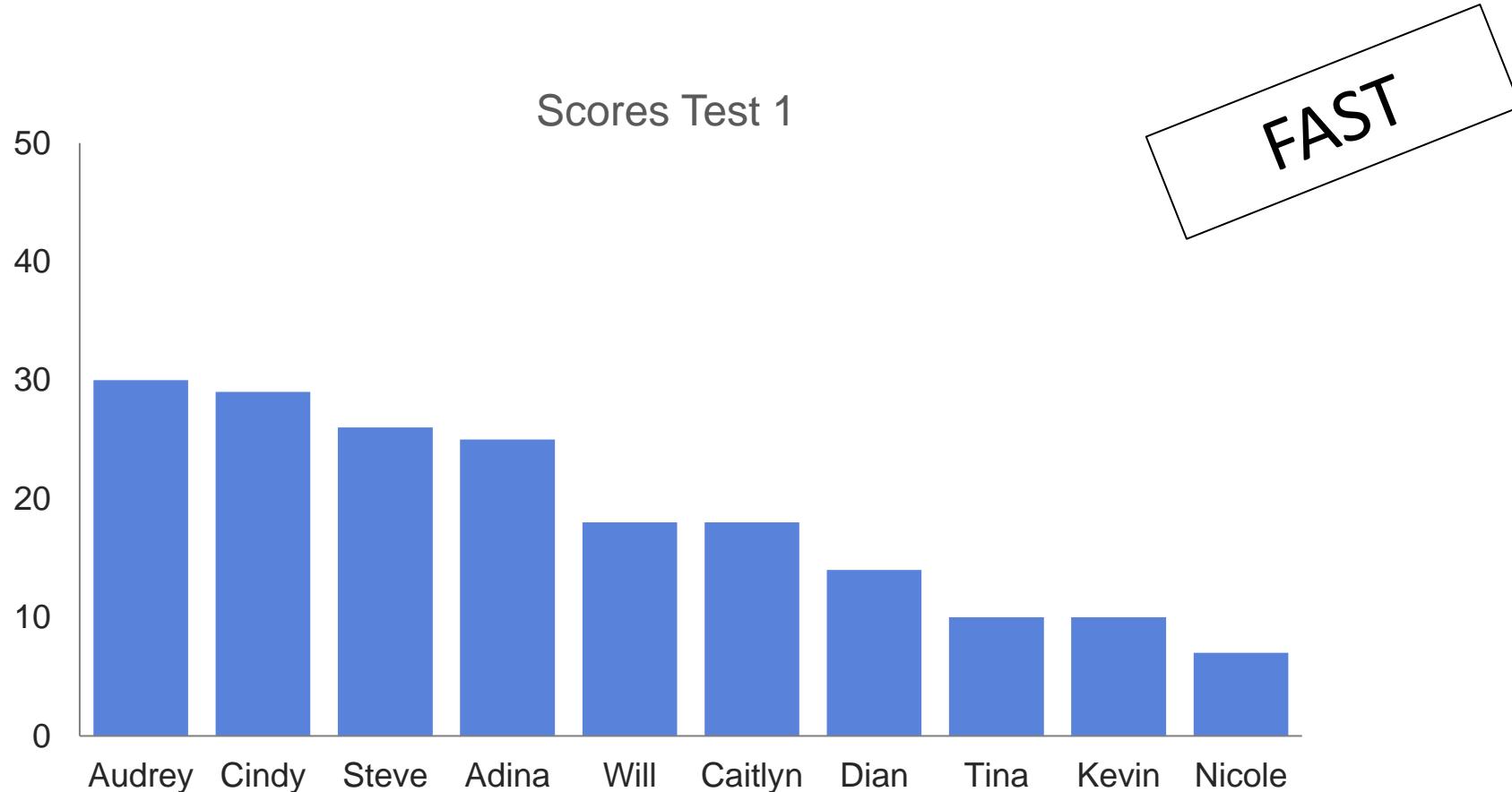
What is the average improvement?



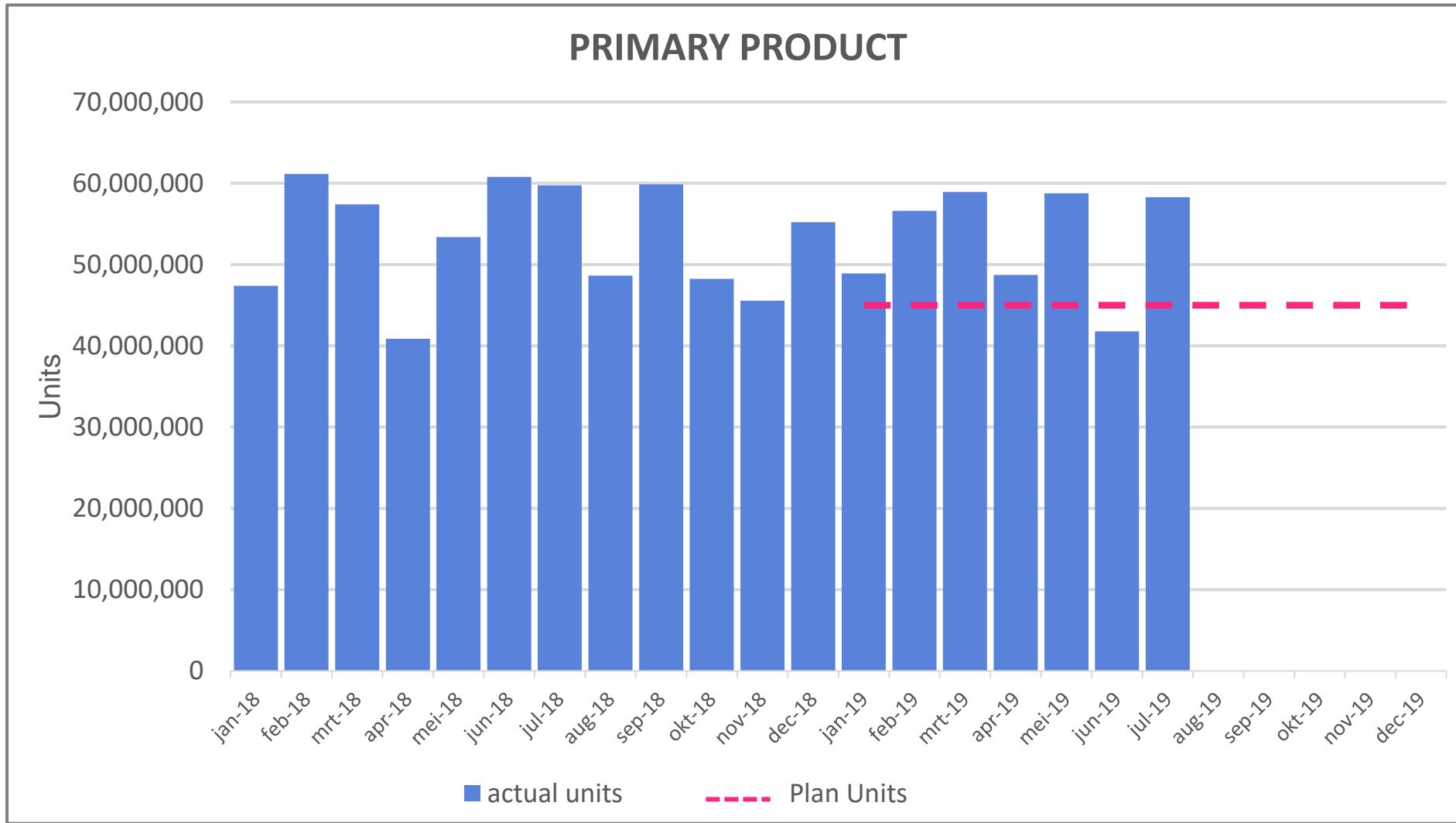
What are the top 5 students?



What are the top 5 students?



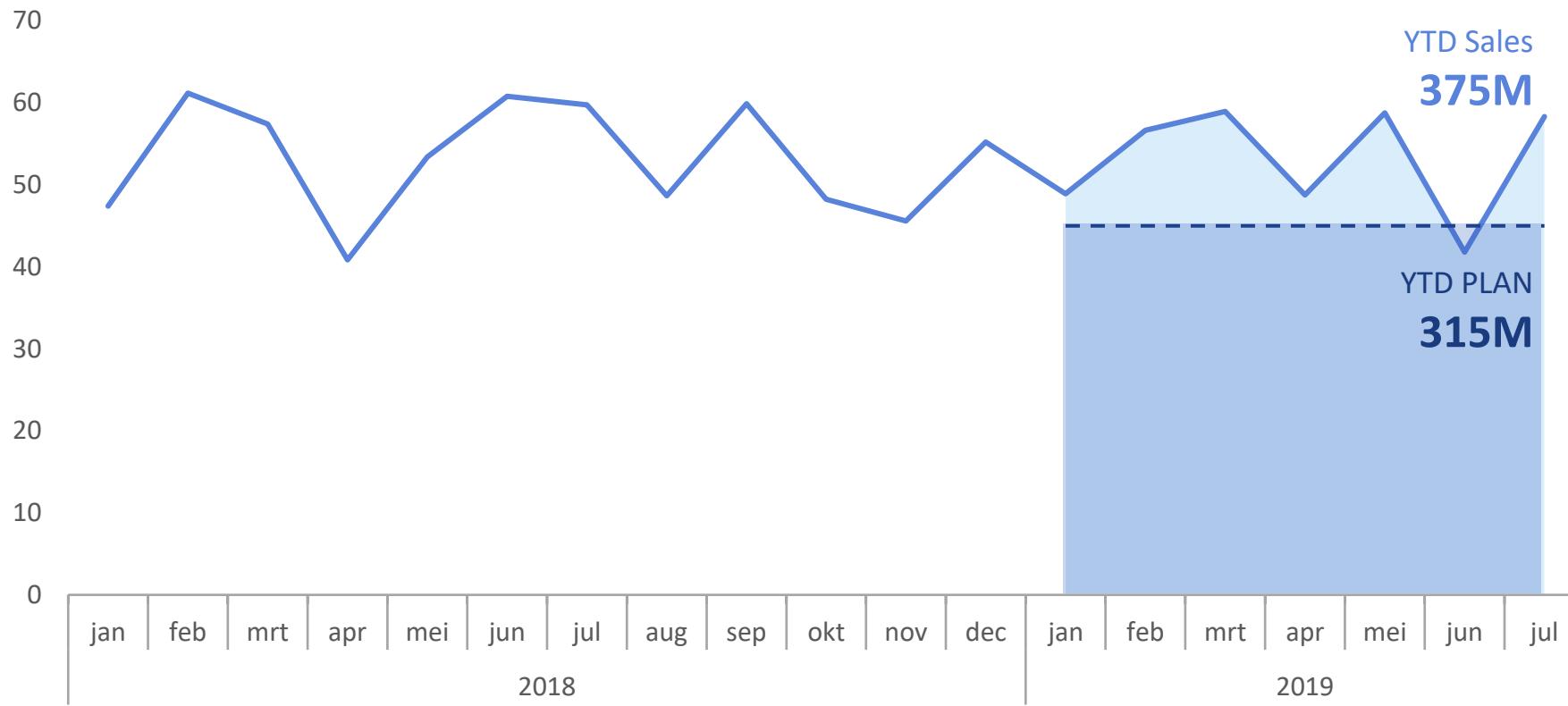
What is the message?



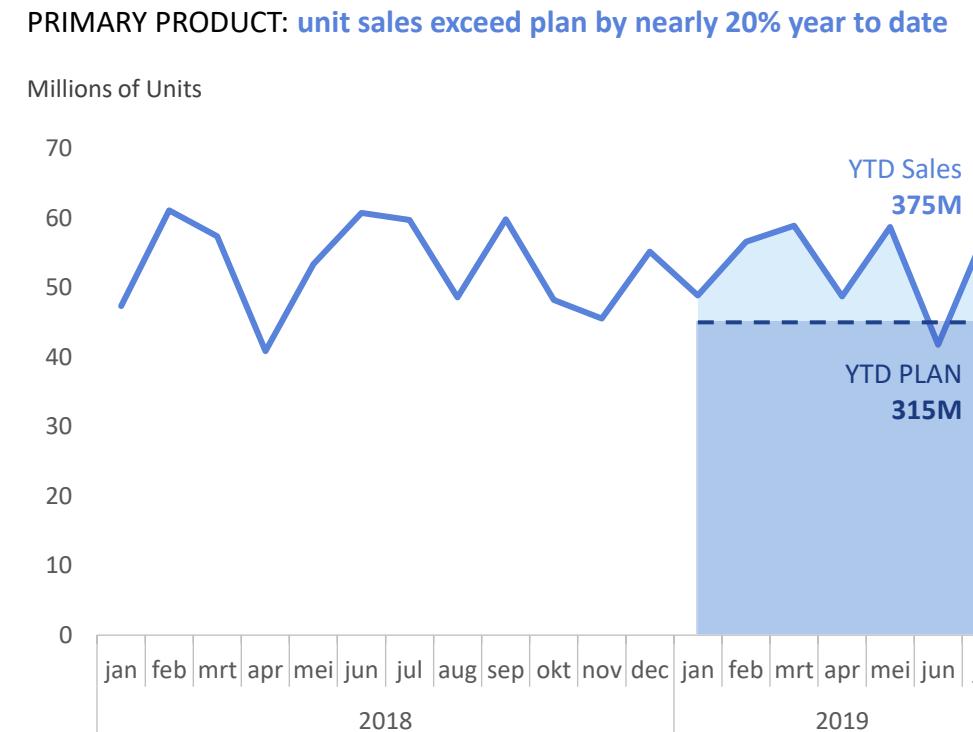
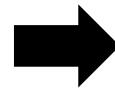
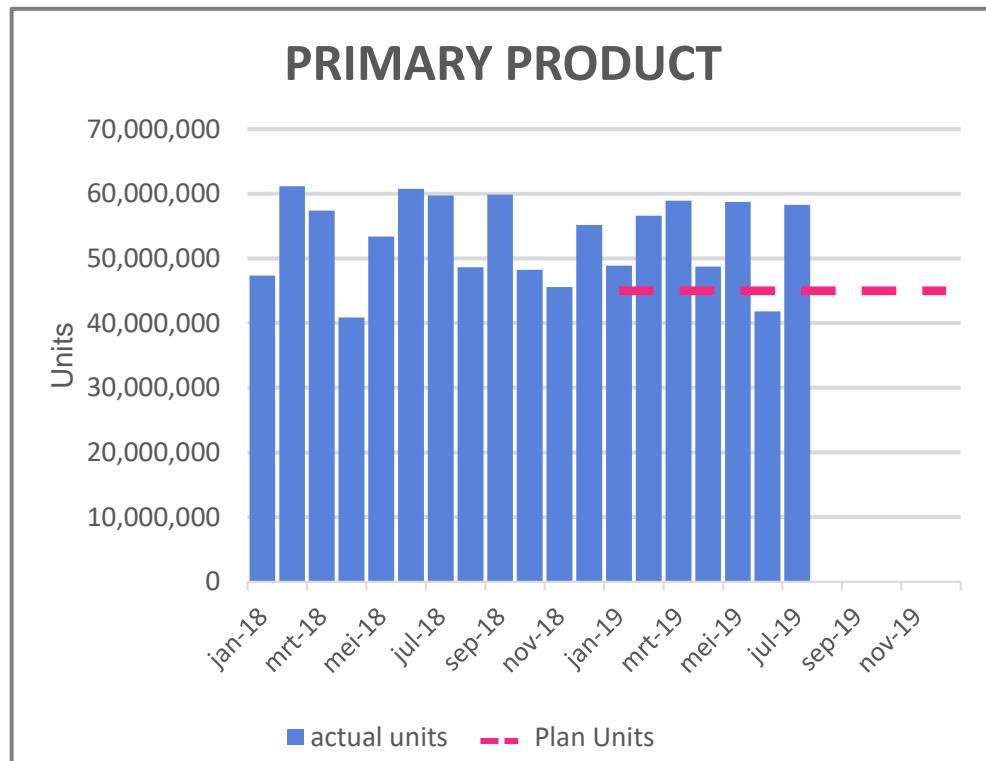
What is the message?

PRIMARY PRODUCT: unit sales exceed plan by nearly 20% year to date

Millions of Units



Make-over: Simplify, declutter and add focus



Minimize distraction: reduce ‘non-data ink’



Effective. Not optimal.

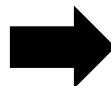
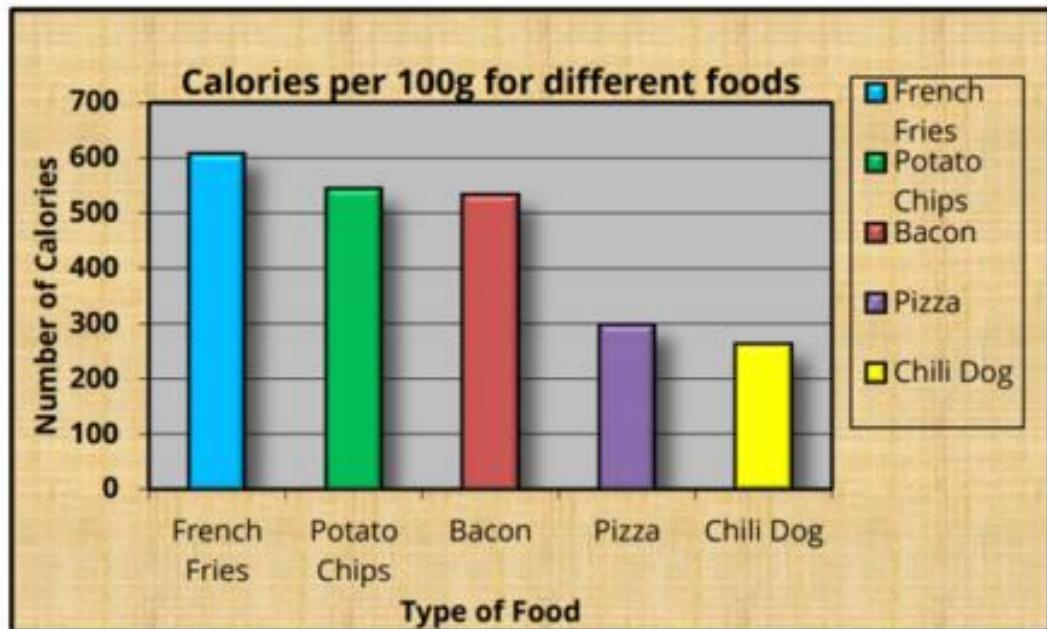


Simple, effective, optimal.

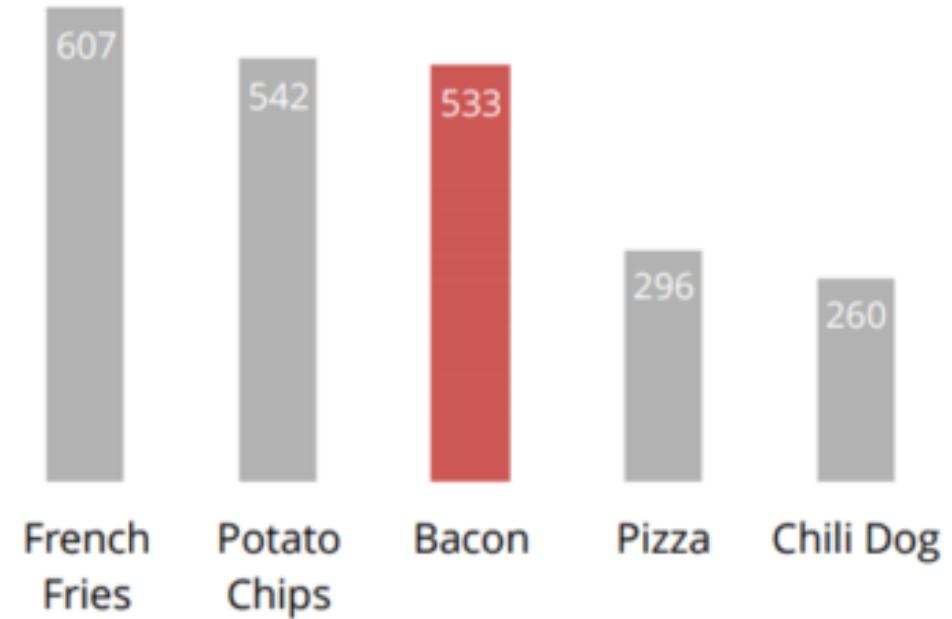


Bad.

Minimize distraction: maximize data-ink ratio



Calories per 100g

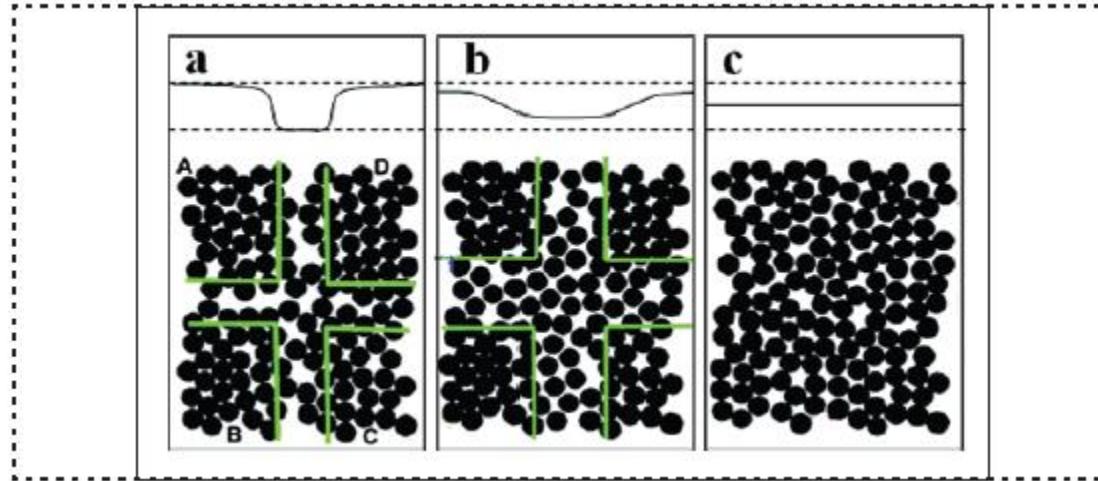


Data-ink ratio also applies to tables

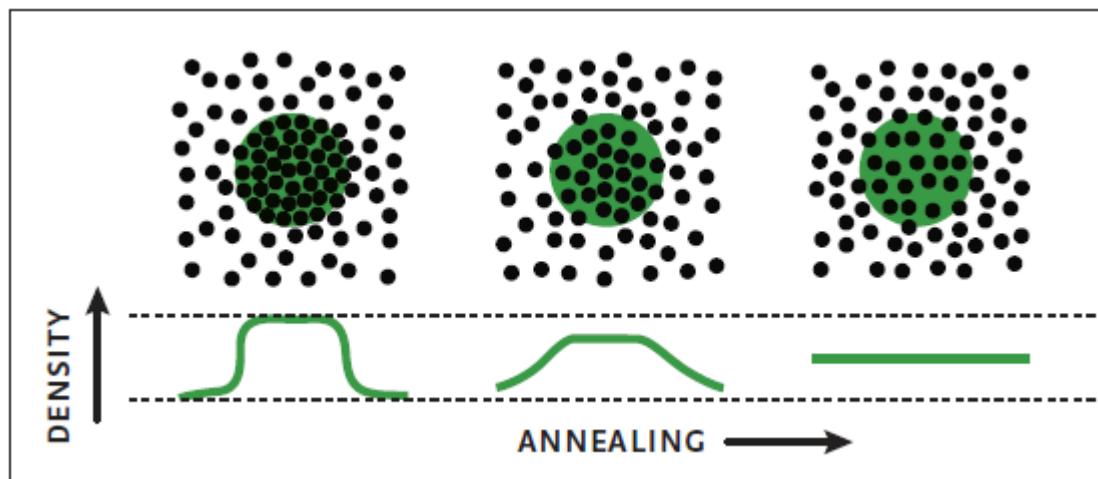
Student	Test1	Test2	Difference
Audrey	30.00	42.00	12.00 pts
Adina	25.00	37.00	12.00 pts
Caitlyn	18.00	28.00	10.00 pts
Cindy	29.00	39.00	10.00 pts
Dian	14.00	25.00	11.00 pts
Kevin	10.00	21.00	11.00 pts
Nicole	7.00	20.00	13.00 pts
Steve	26.00	13.00	-13.00 pts
Tina	10.00	12.00	2.00 pts
Will	18.00	30.00	12.00 pts

Student	Test1	Test2	Difference (pts)
Audrey	30	42	12
Adina	25	37	12
Caitlyn	18	28	10
Cindy	29	39	10
Dian	14	25	11
Kevin	10	21	11
Nicole	7	20	13
Steve	26	13	-13
Tina	10	12	2
Will	18	30	12

And to graphical abstracts

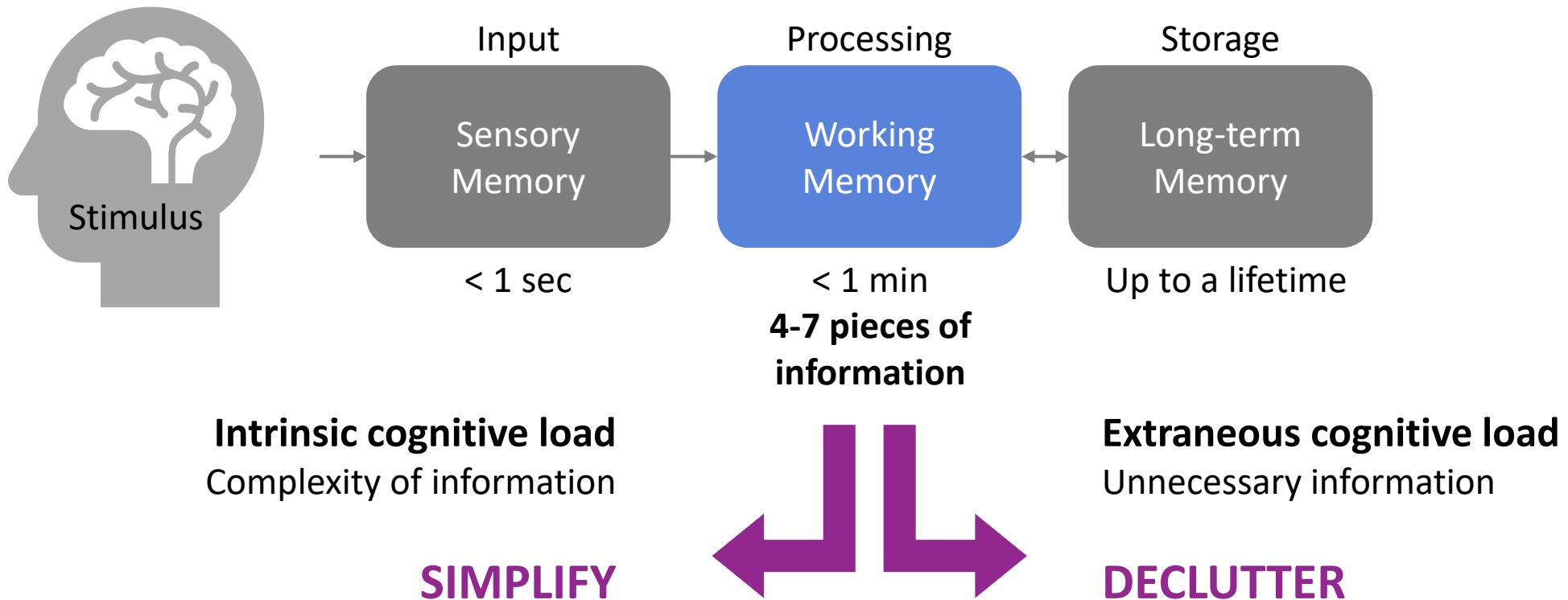


Too many interior boxes clutter the layout.



Boxes removed and arrows simplified.

Optimize visual perception: minimize *intrinsic* and *extraneous cognitive load*



Exercise

Notice. Wonder. Improve.

A look at your work.



Notice. Wonder. Improve.



Notice: (<10 sec)

- What is the first thing that catches your eye?
- What is your first impression?



Wonder: (< 1 min)

- Do you get an idea about the message? What is it?
- What questions come to mind?



Improve: Simplify and Declutter

- Is the most appropriate chart type used to present the data/message?
- What takes up most time to process? Could this process be simplified?
- Are there any unnecessary, redundant or distracting elements that can be removed/reduced?
- How could the message be more emphasized?
- Are all text and labels legible?

Color!

Notice. Wonder. Improve.

A look at your work.



Color issues are common in science communications!

Hydrol. Earth Syst. Sci.

797 papers from 2005-2020

Papers with color issues
(total % per journal)



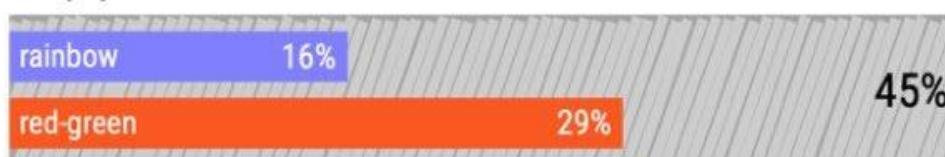
Nature Scientific Reports

Top 100 papers from 2019

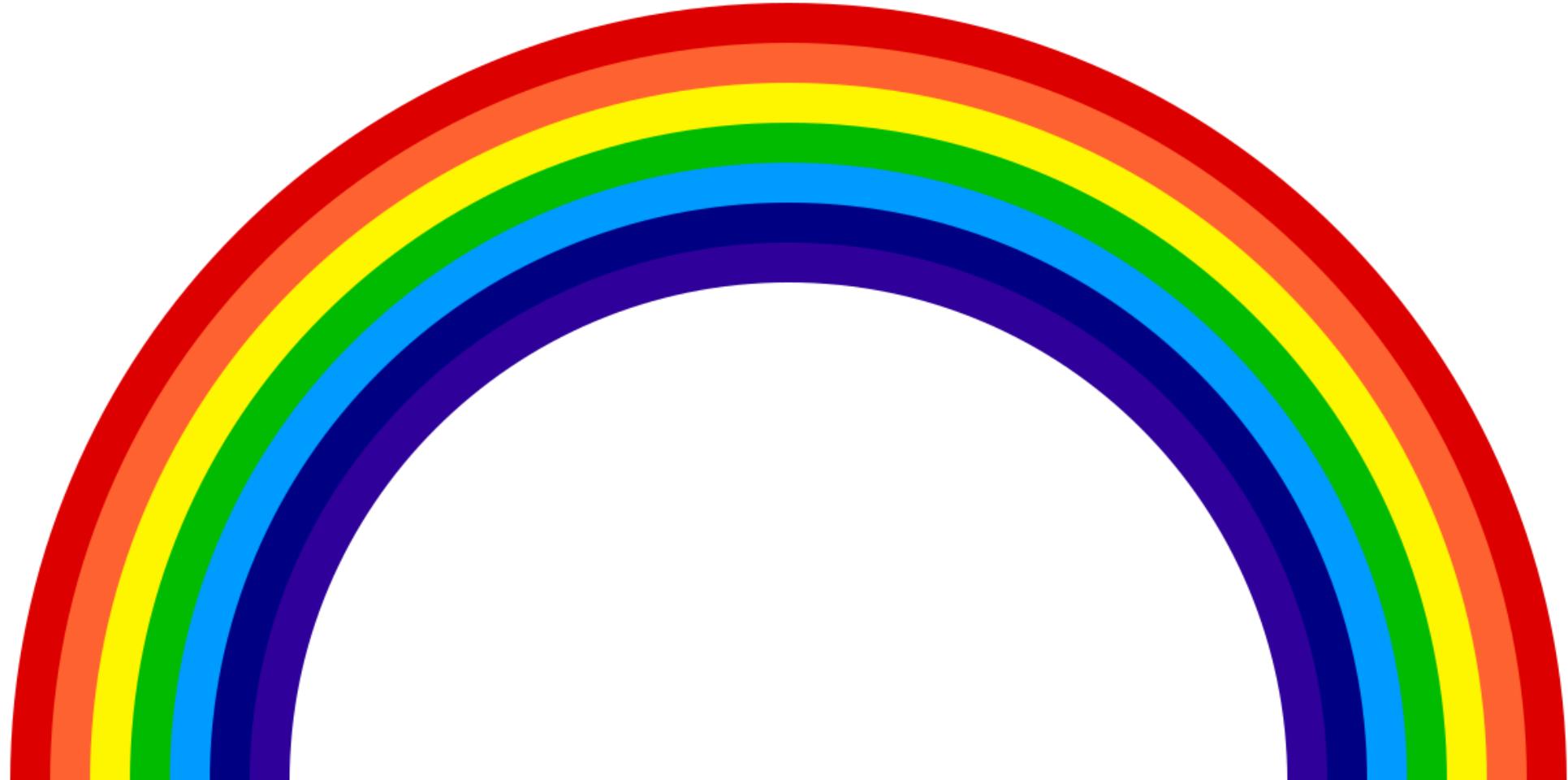


Nature Communications

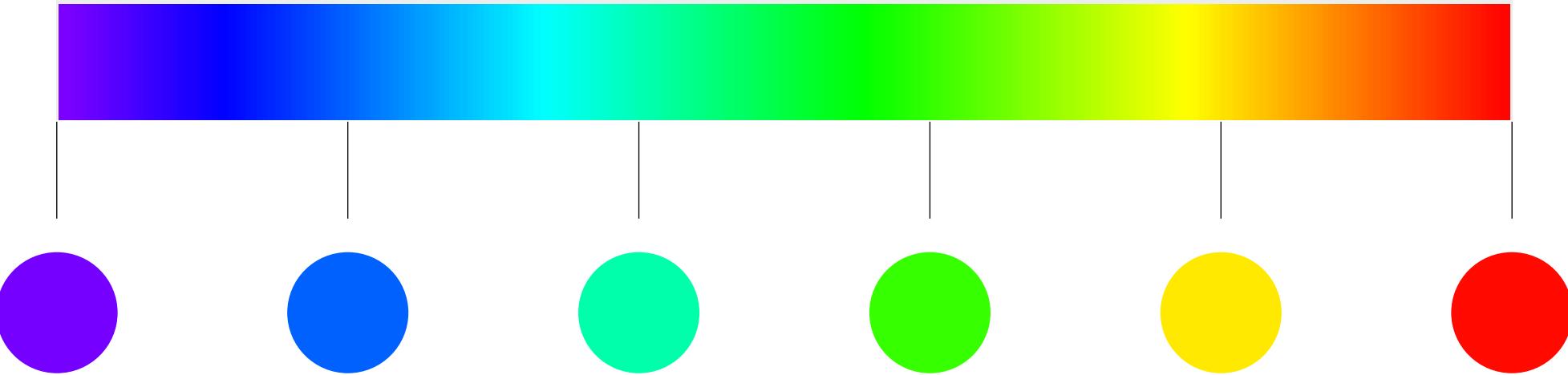
100 papers from 4-6 Nov 2020



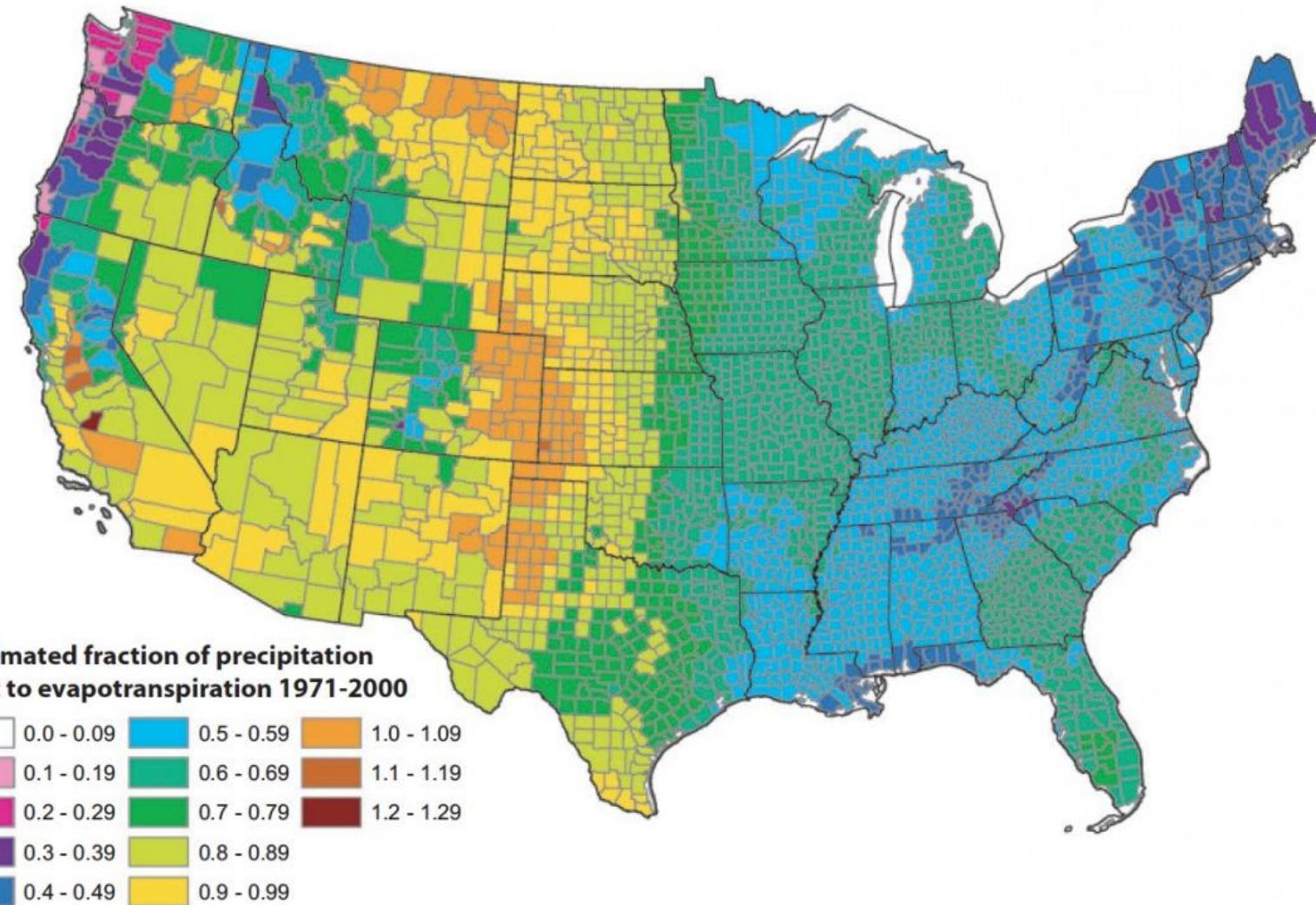
What's wrong with the rainbow color scale?



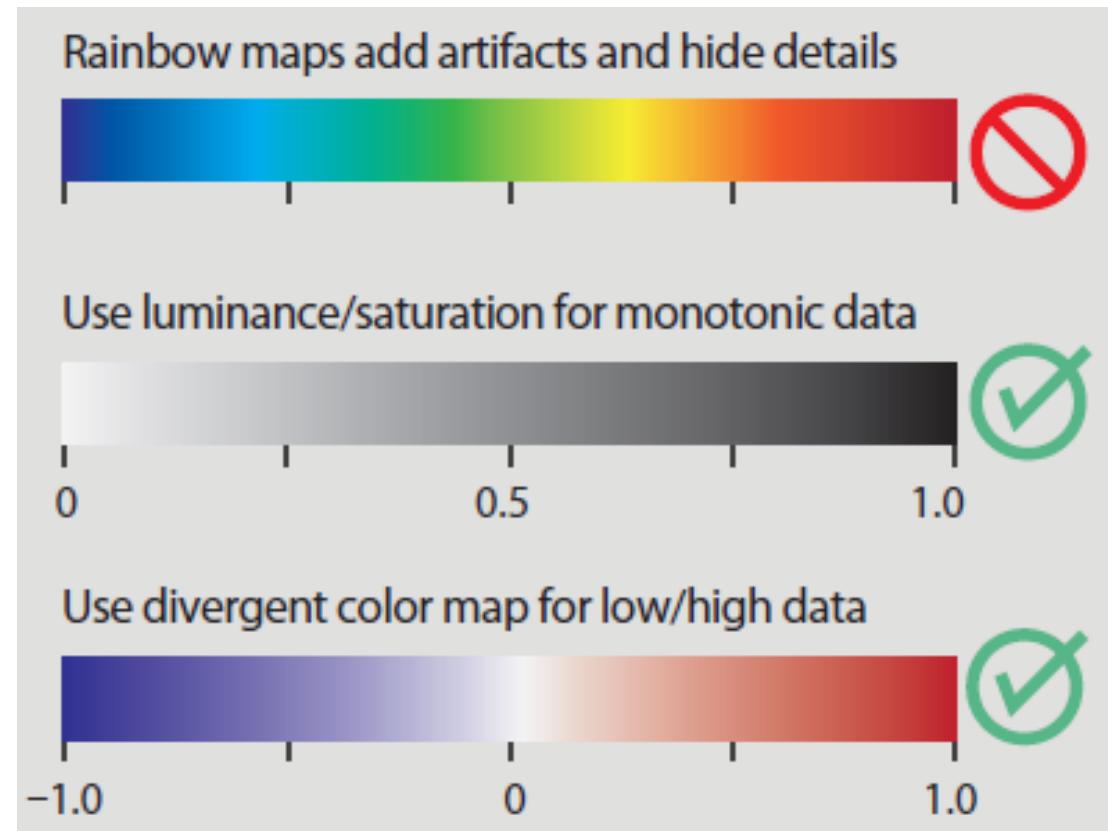
What's wrong with the rainbow color scale?



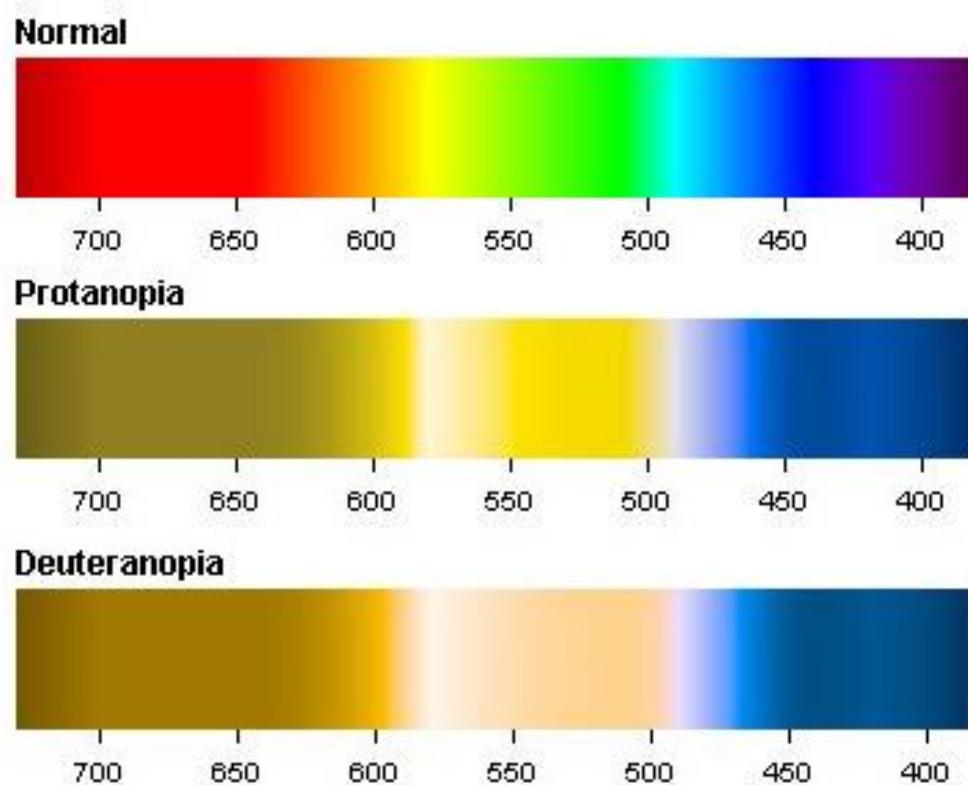
Avoid rainbow color maps for quantities!



Use a monochromatic or a diverging color scale instead



Another issue with color: color vision deficiency

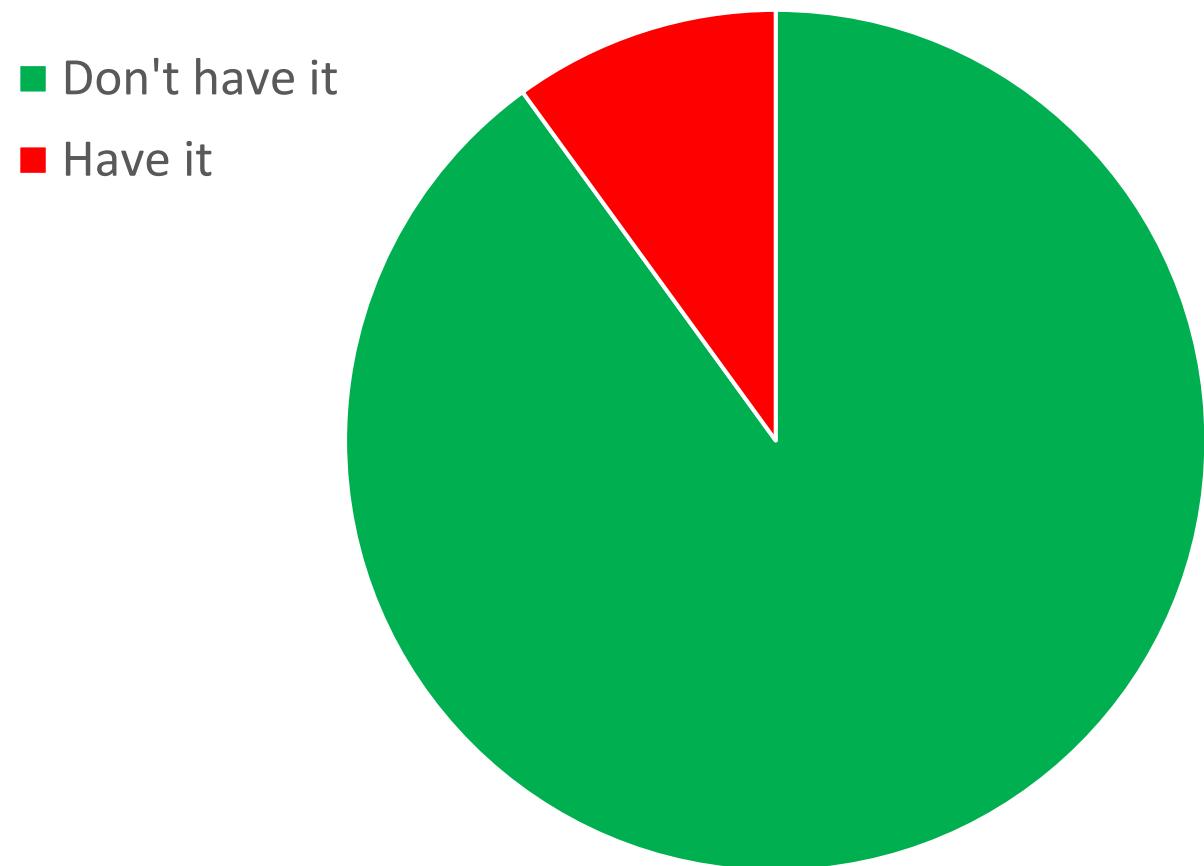


~2% male population

~6% male population

Avoid red-green color palettes

Red-green color blindness in men



Use a color blindness simulator or colorblindness-friendly color scales instead



Color Oracle

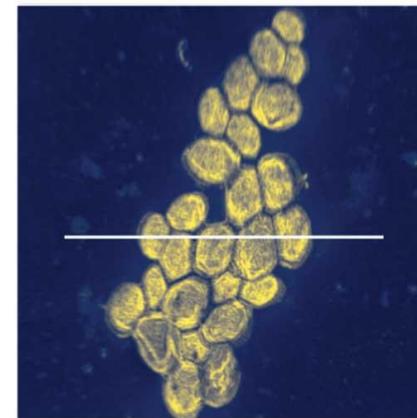
Design for the Color Impaired



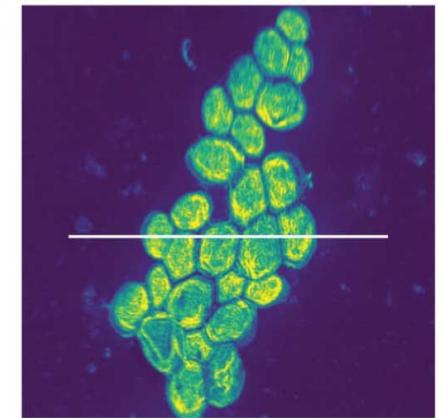
Color Oracle is a free color blindness simulator for Windows, Mac and Linux. It takes the guesswork out of designing for color blindness by showing you in real time what people with common color vision impairments will see.

Color Oracle applies a full screen color filter to art you are designing, independently of the software in use. Eight percent of all males are affected by color vision impairment – make sure that your graphical work is readable by the widest possible audience.

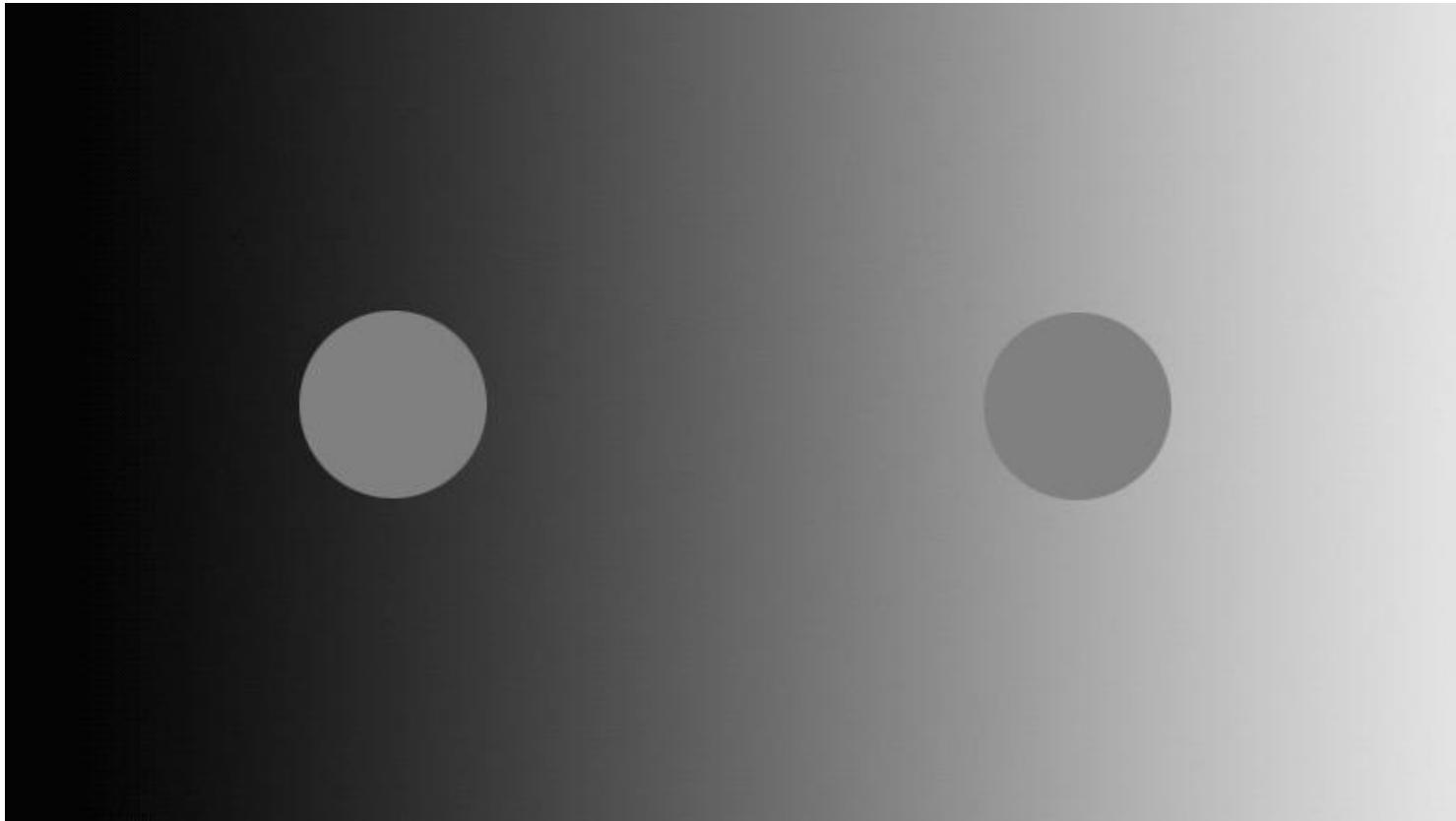
Cividis



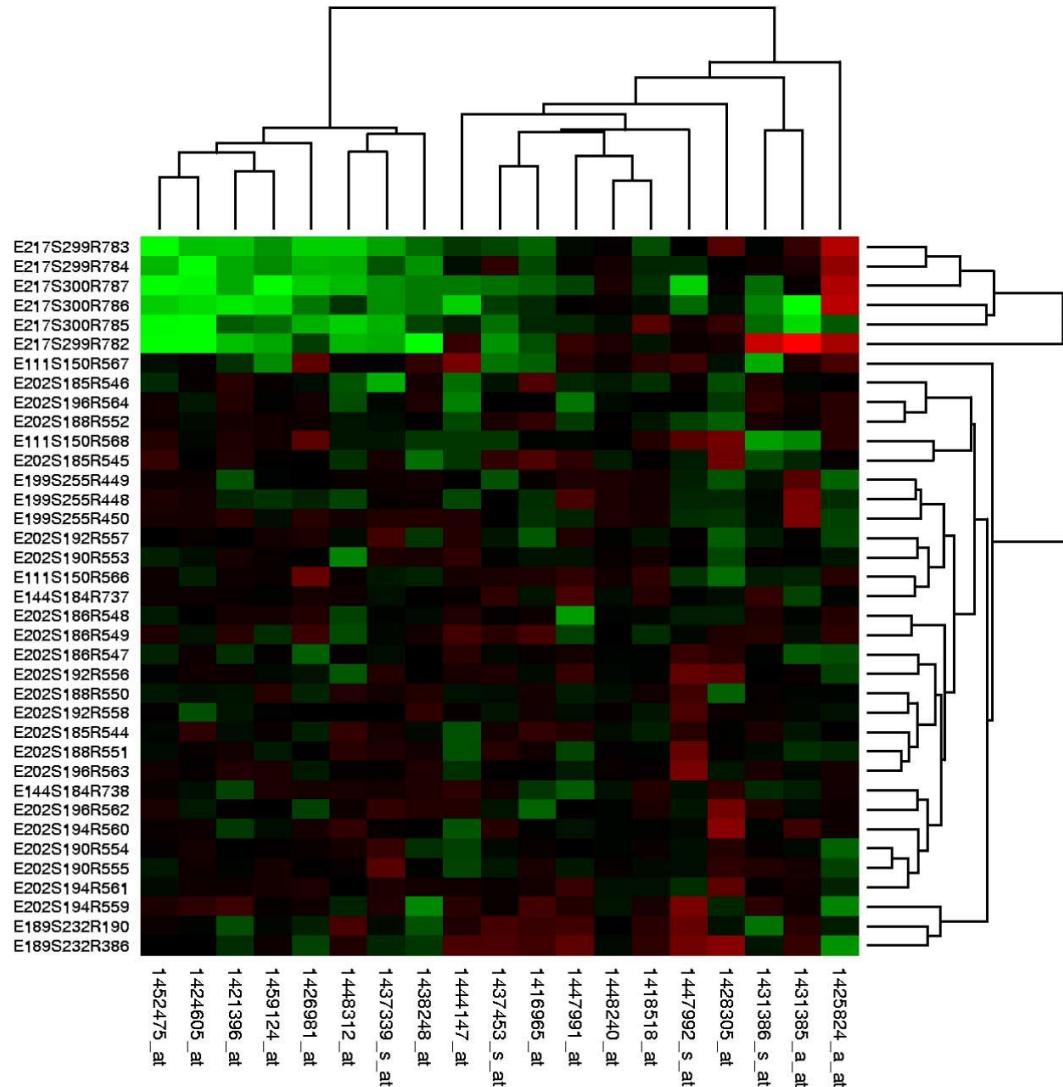
Viridis



Which dot is darker? Left or right?



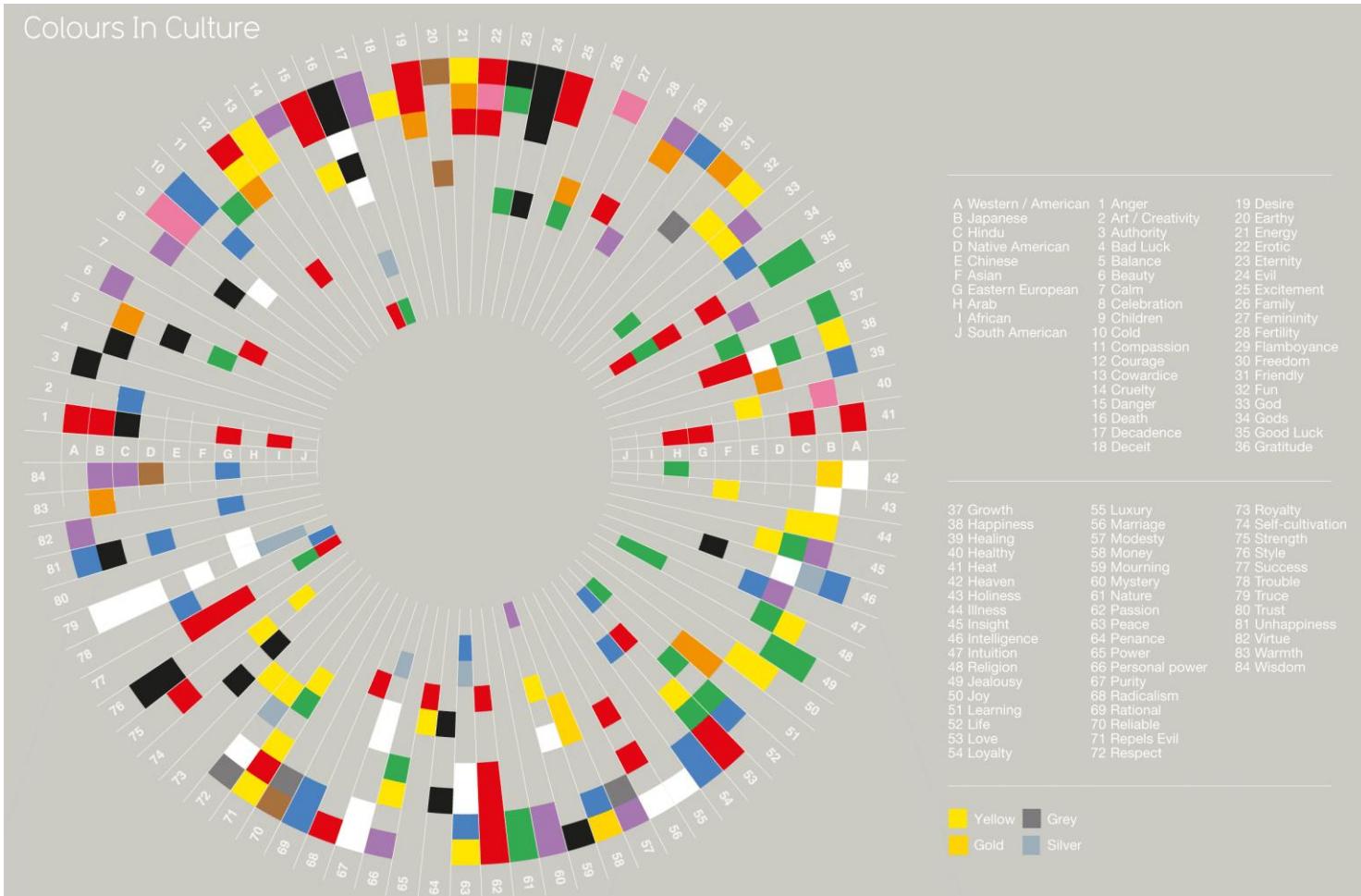
Be aware: contrast can create illusions!



Lastly, be aware how color is associated to feelings...



...and differs between cultures

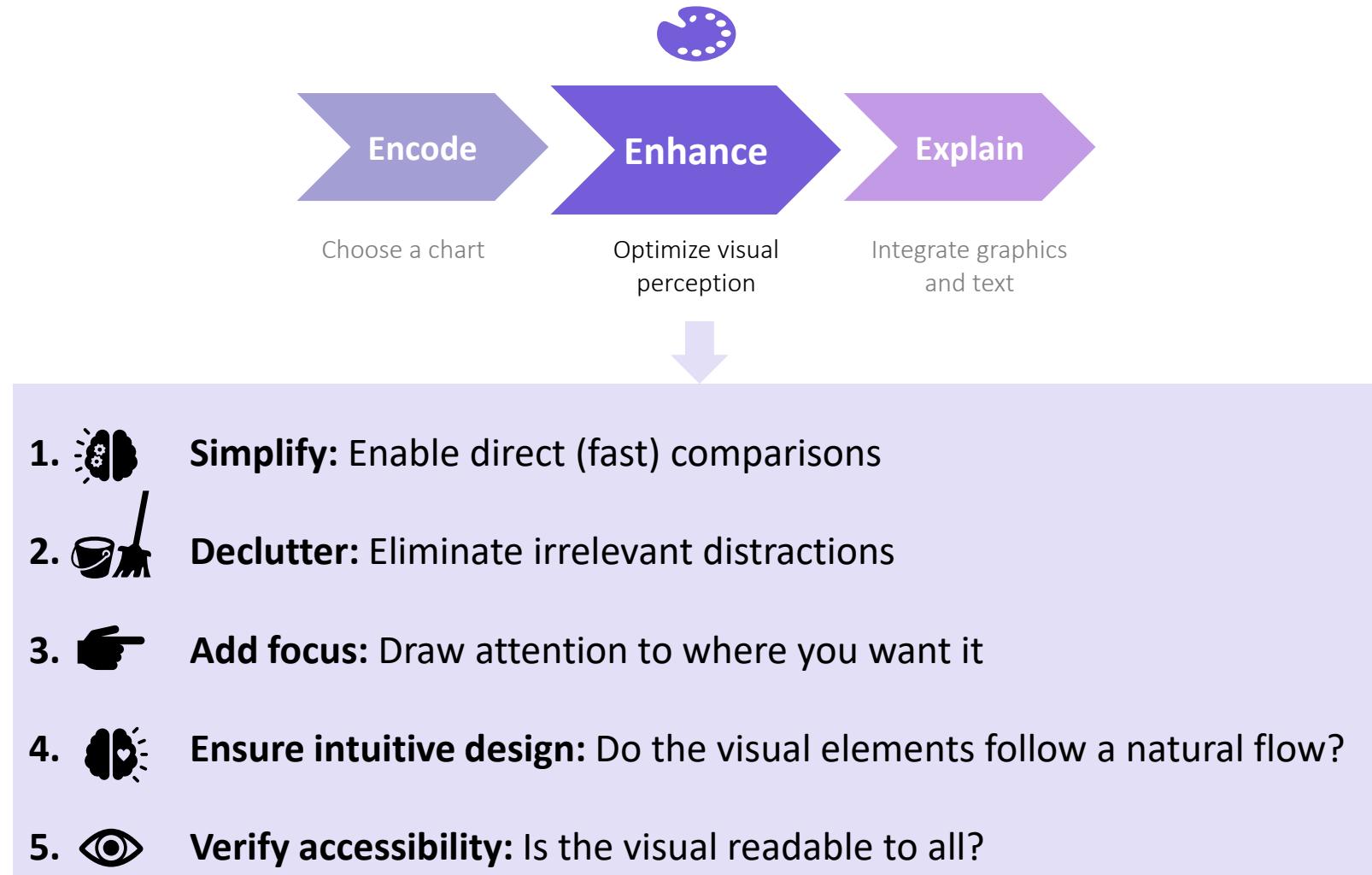


Wrapping up

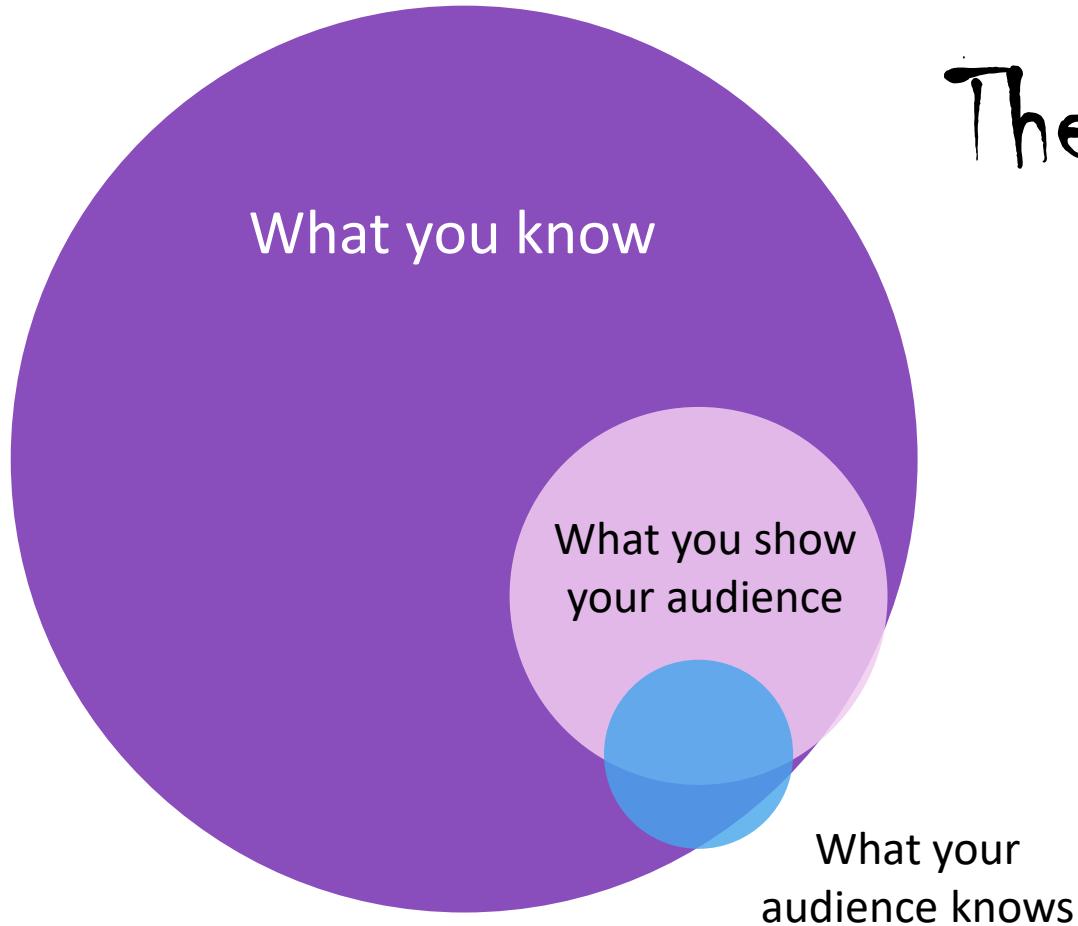
The 5 key steps in visual design
Do's and Don'ts



Pulling it all together: **5 steps** to optimize visual perception



Don't: Forget how smart you are!

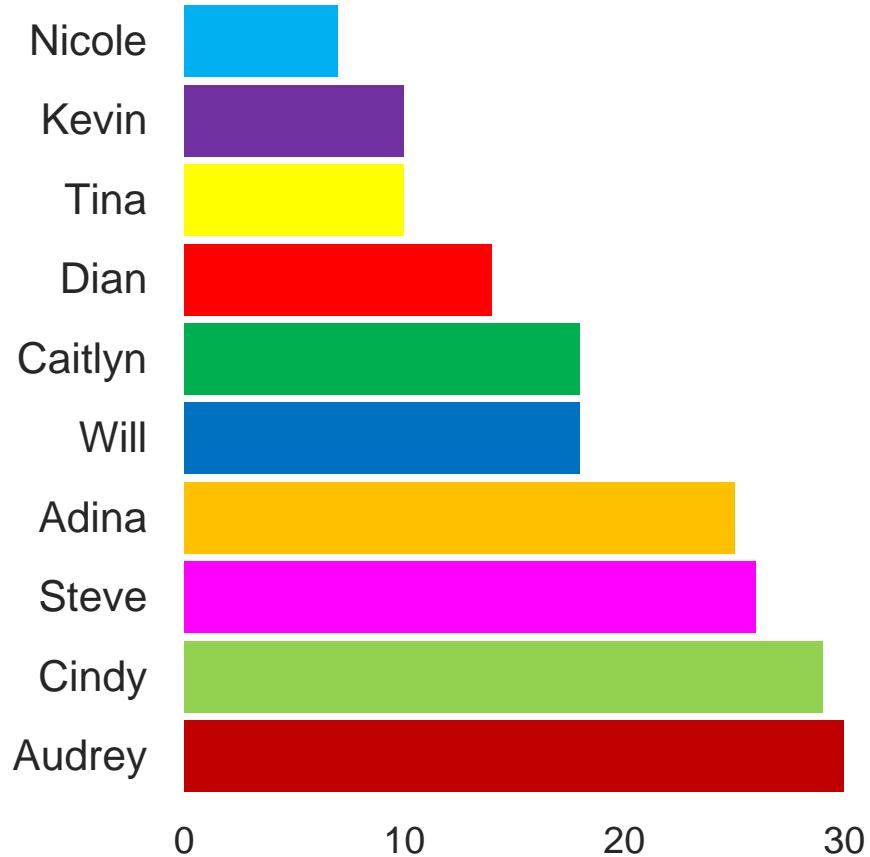


The curse of knowledge

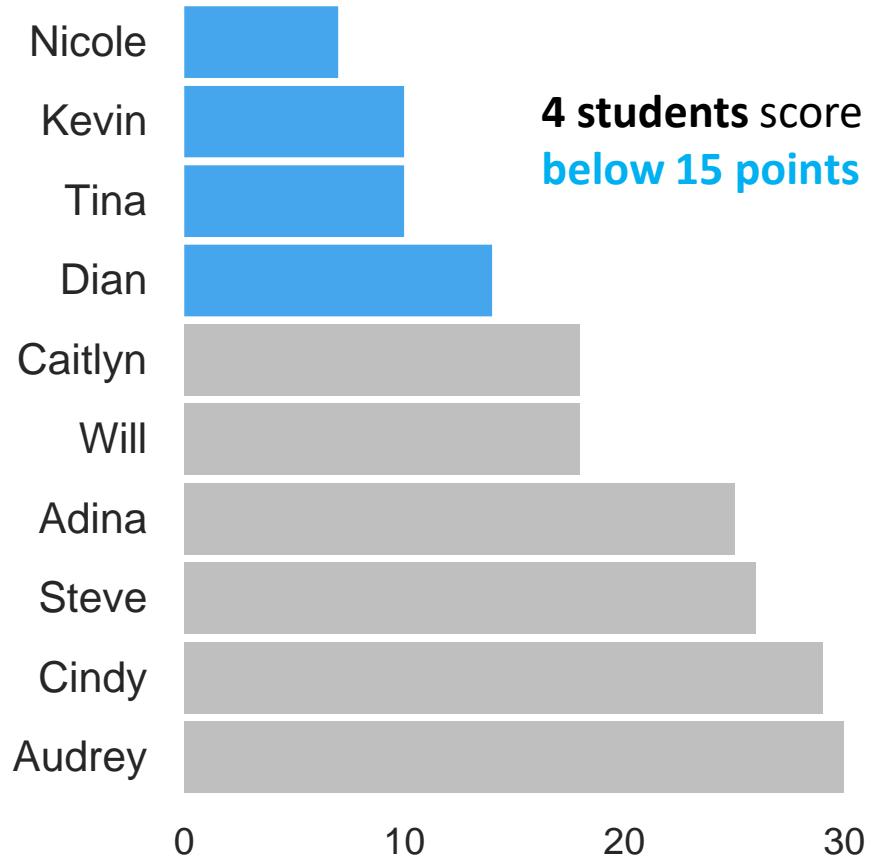
Don't: Let your audience play Where's Wally



Don't: Use too many (useless) colors

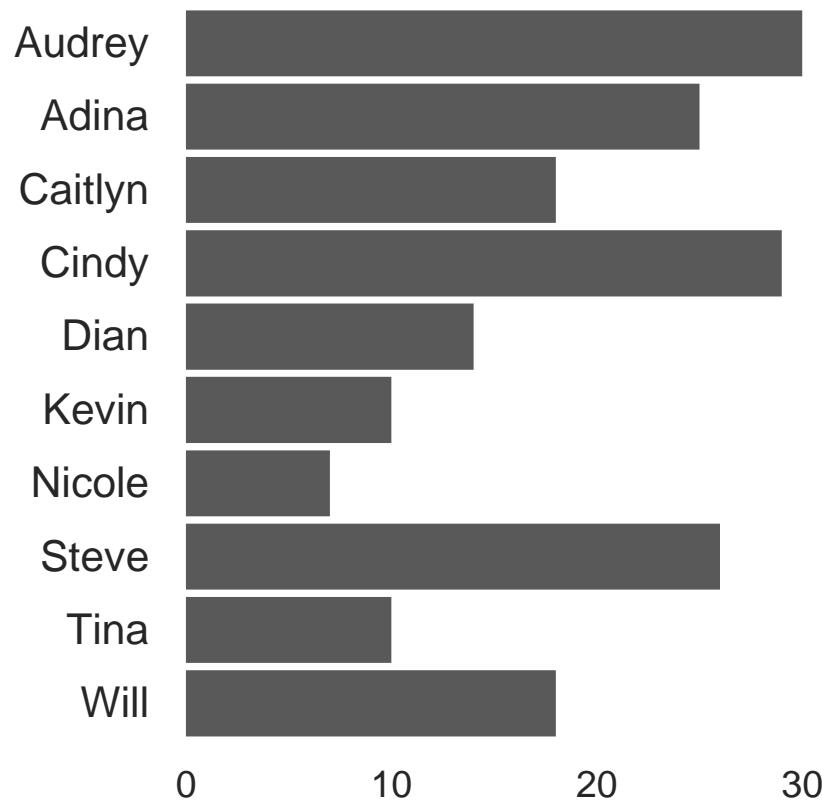


Do: Highlight and annotate your key findings

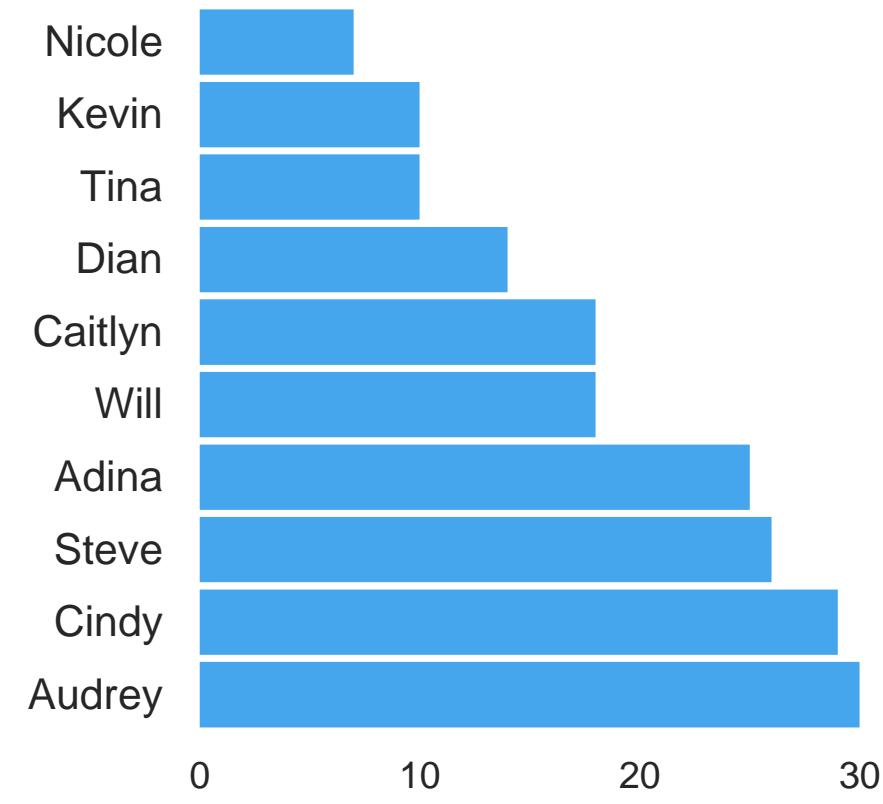


Do: Order your data logically, e.g., alphabetically or by number

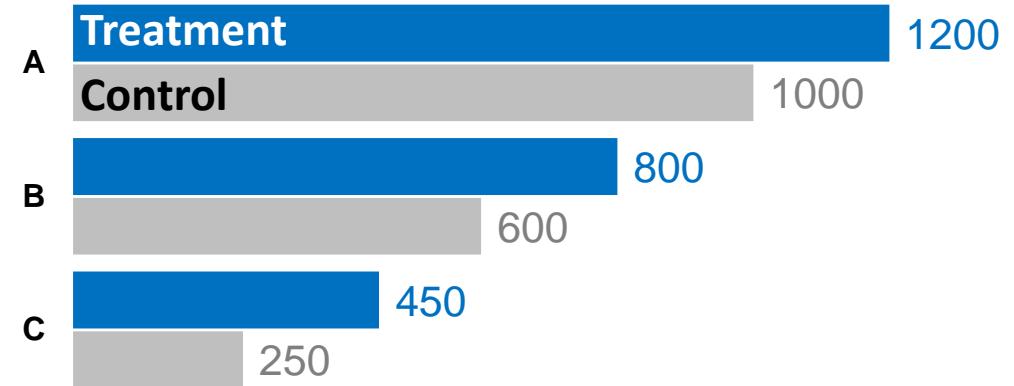
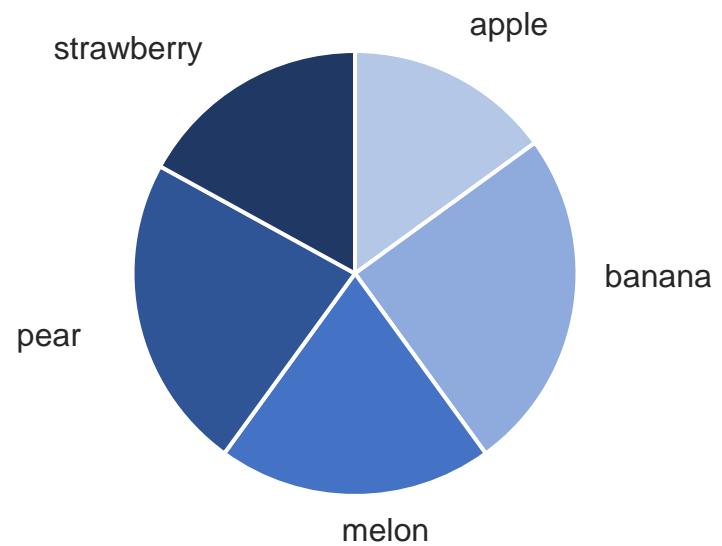
Student's scores for test 1



Student's scores for test 1



Do: Label your data directly

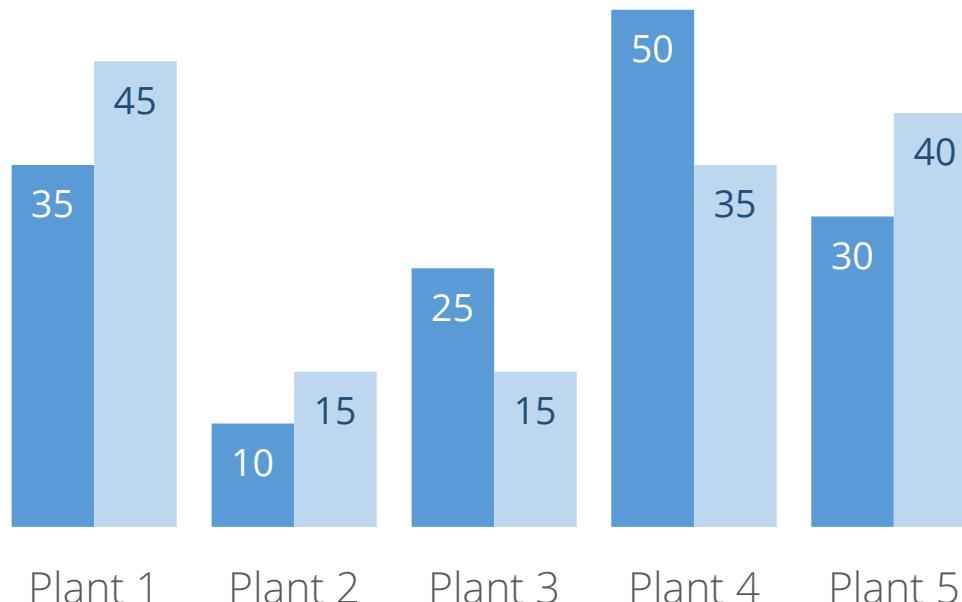


Don't: Let your audience do mental math

Safety Incidents by Plant

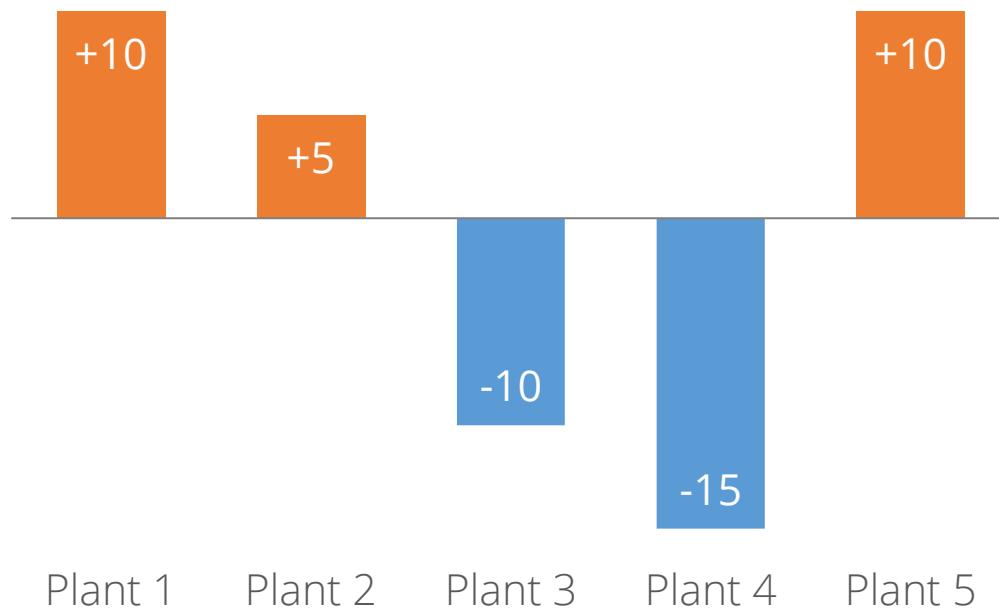
■ 2017 ■ 2018

*What's the mean difference
between 2017-2018?*



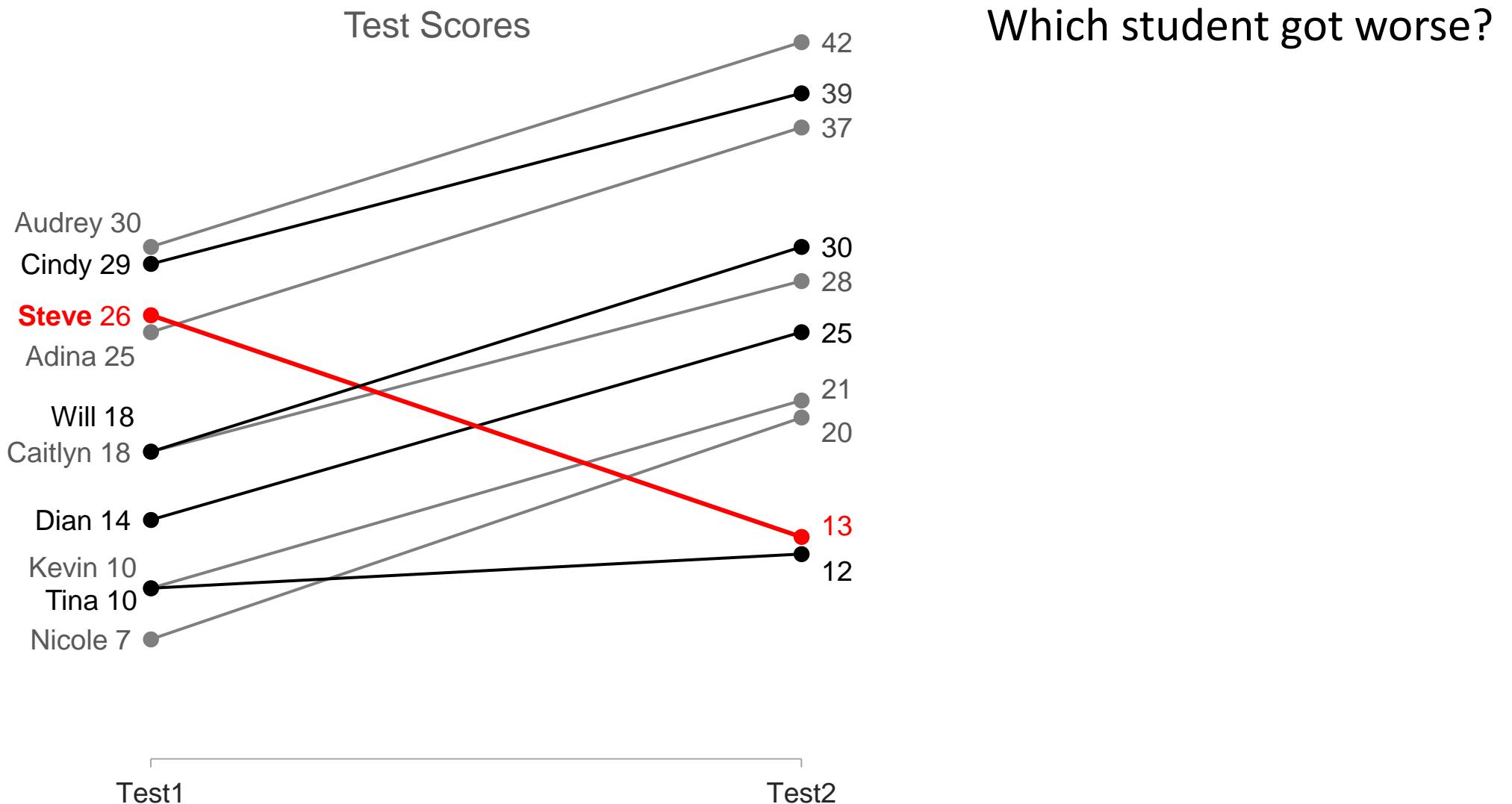
Do: Show differences directly

Year-to-Year Variance in Safety Incidents by Plant (2017-2018)



What's the mean difference between 2017-2018?

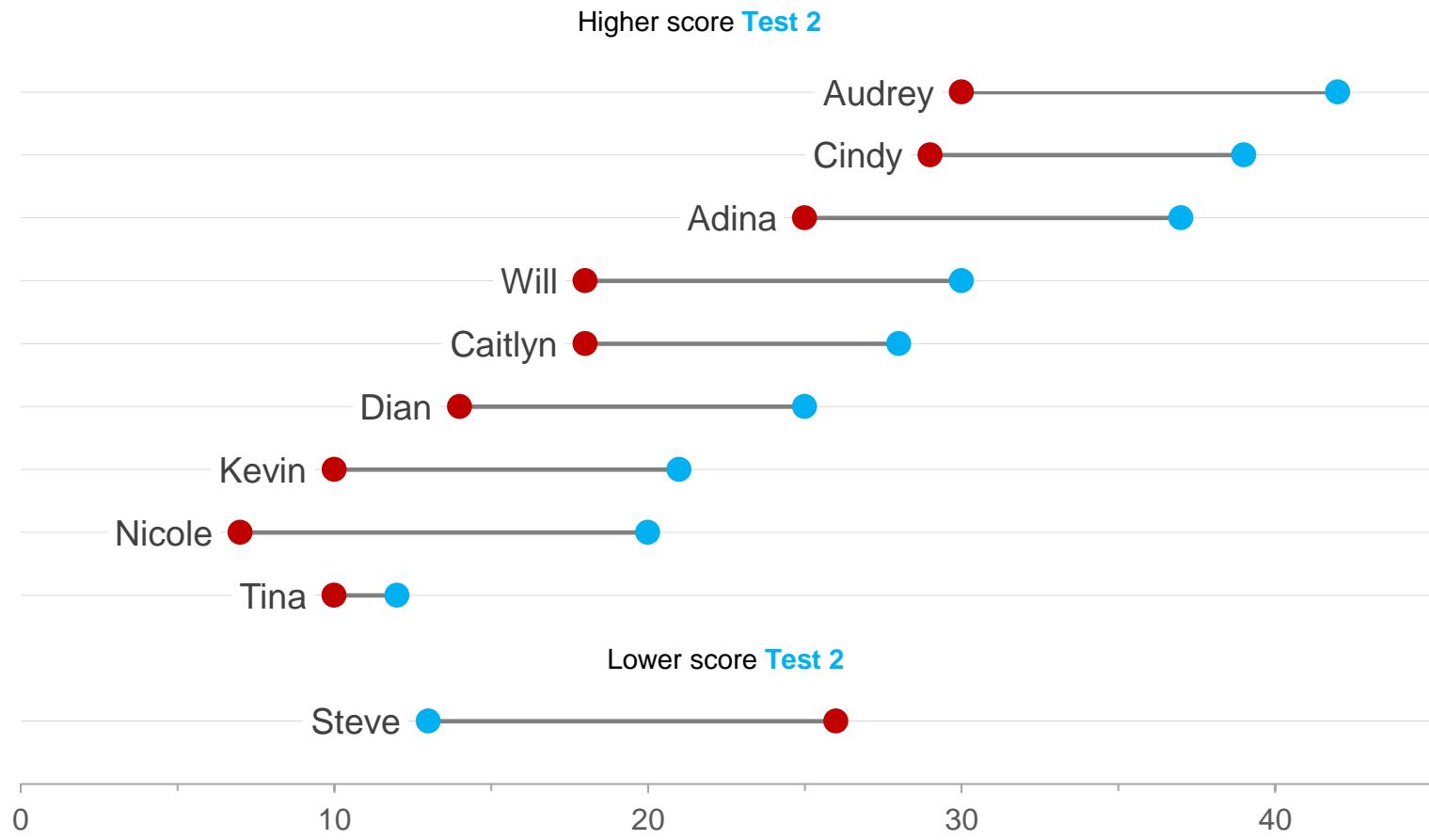
Do: Enable direct comparisons



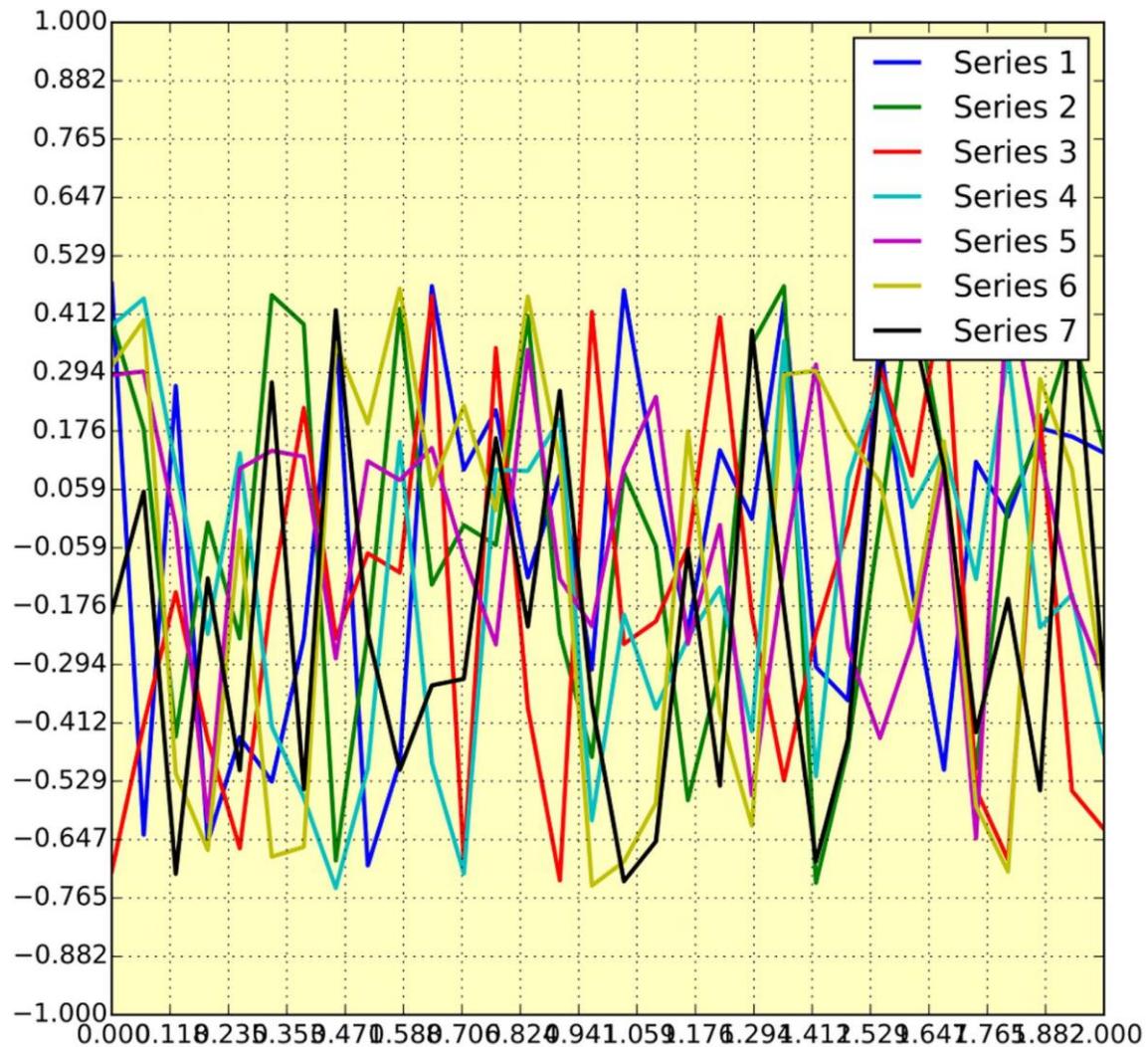
Do: Enable direct comparisons

Scores for **Test1** and **Test2**

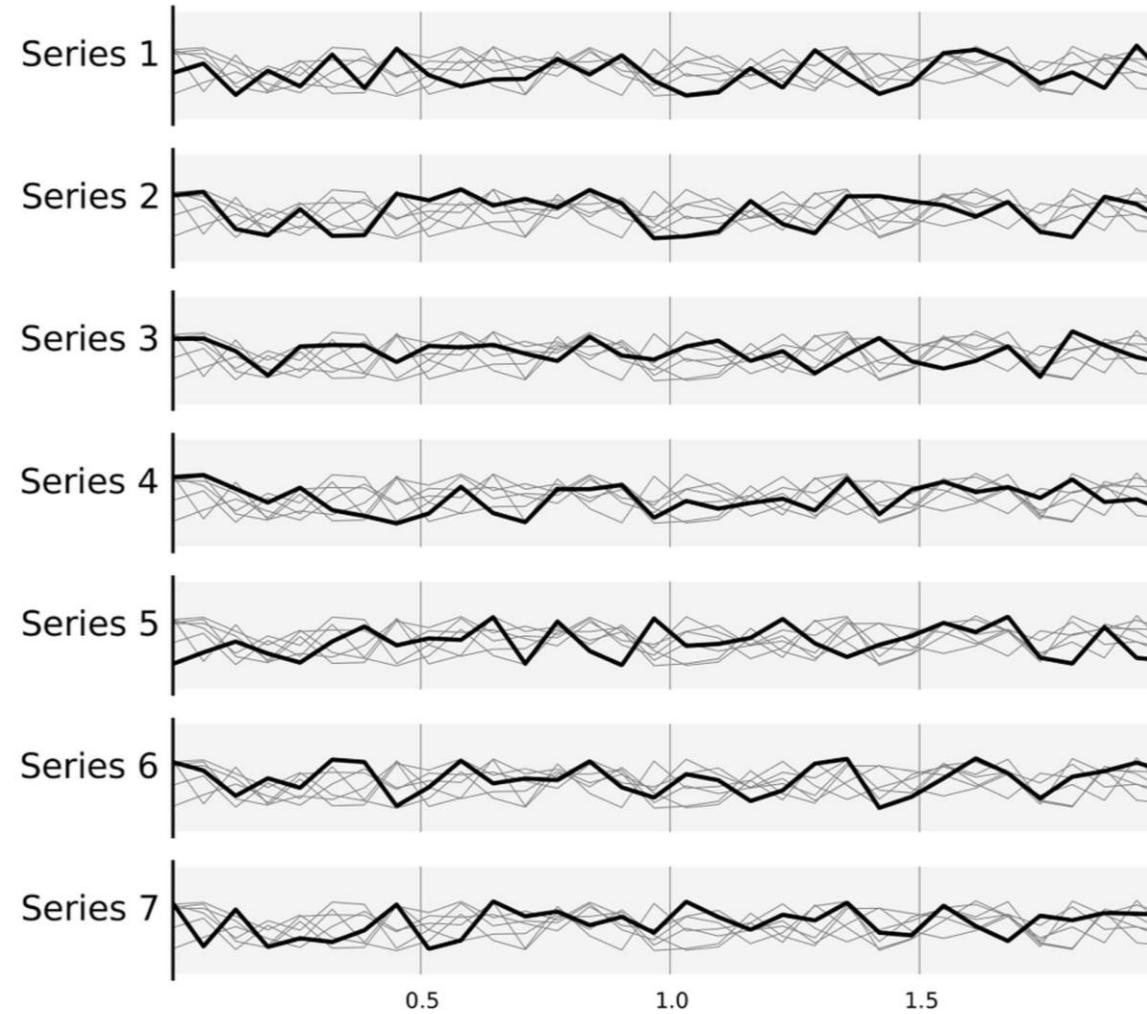
Which student got worse?



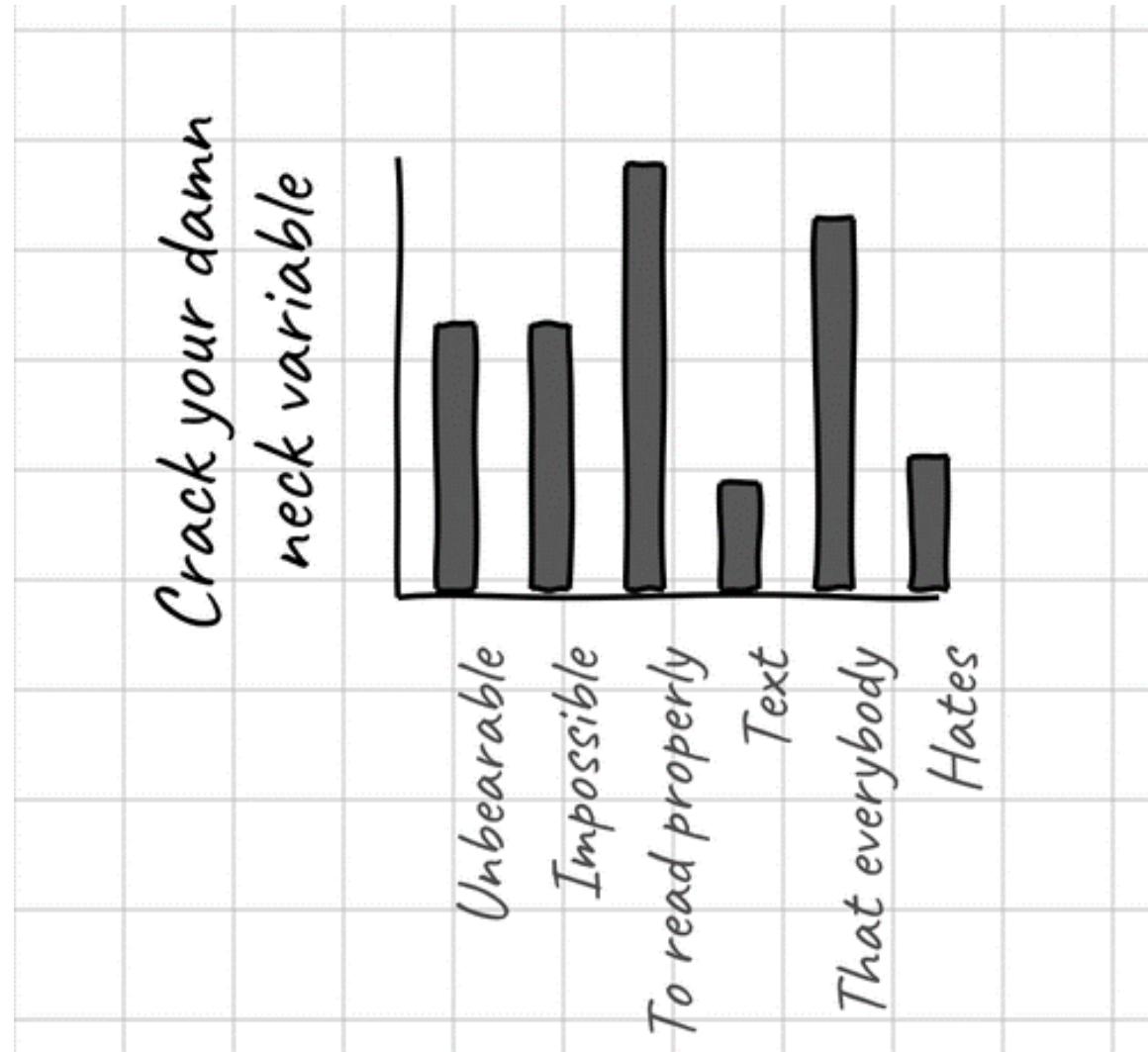
Don't: Use a spaghetti chart



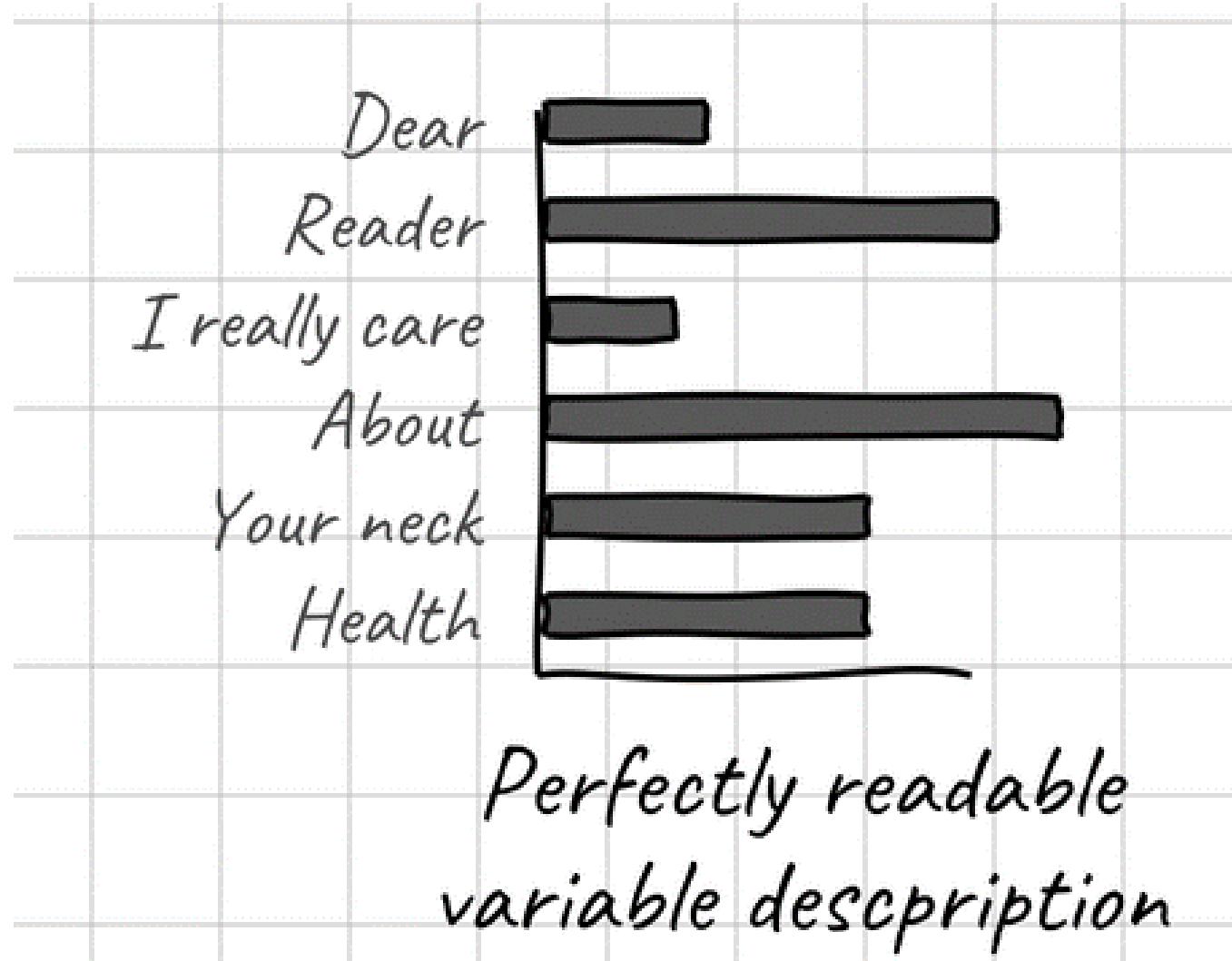
Do: Use small multiples instead



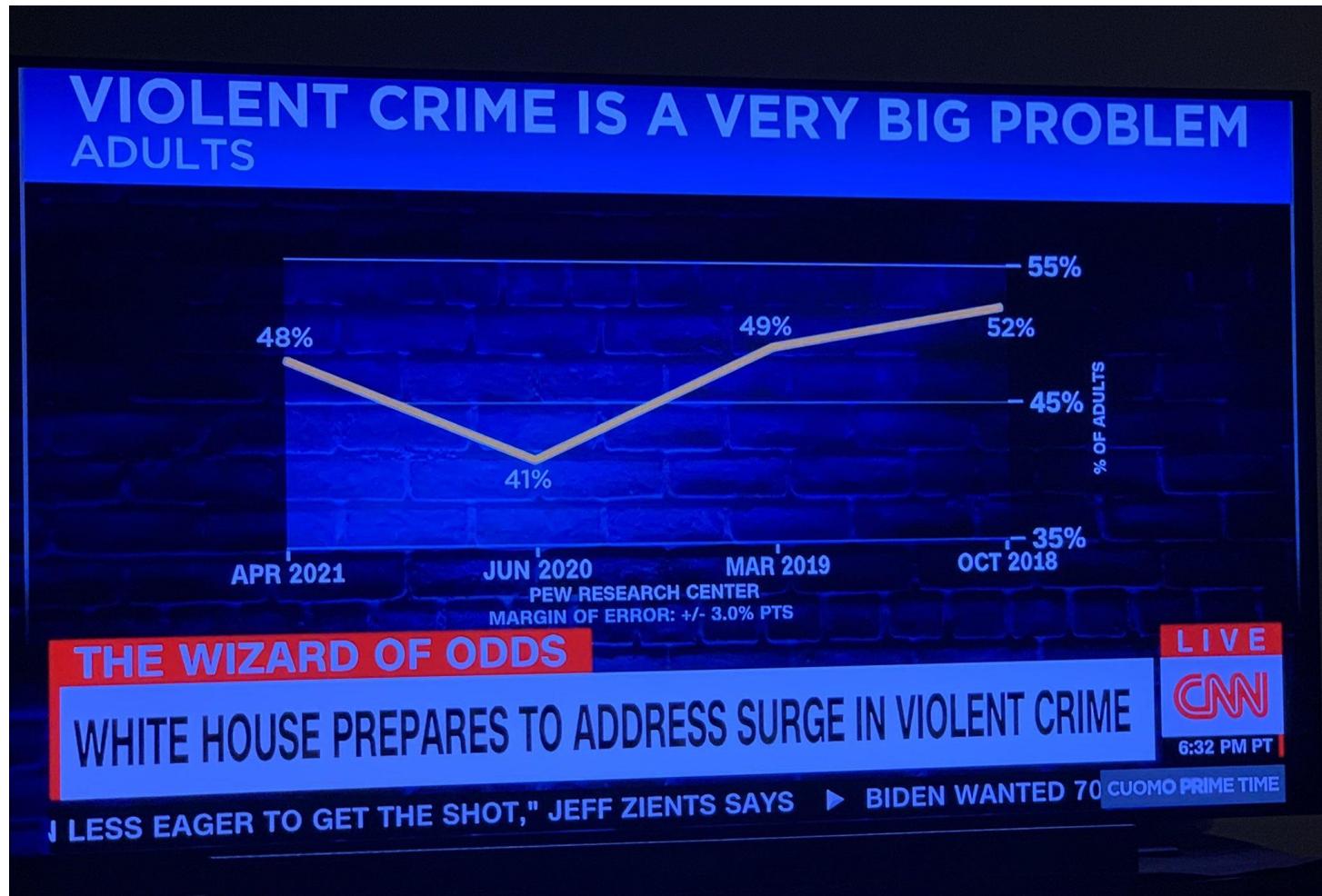
Don't: Make your audience crack their neck!



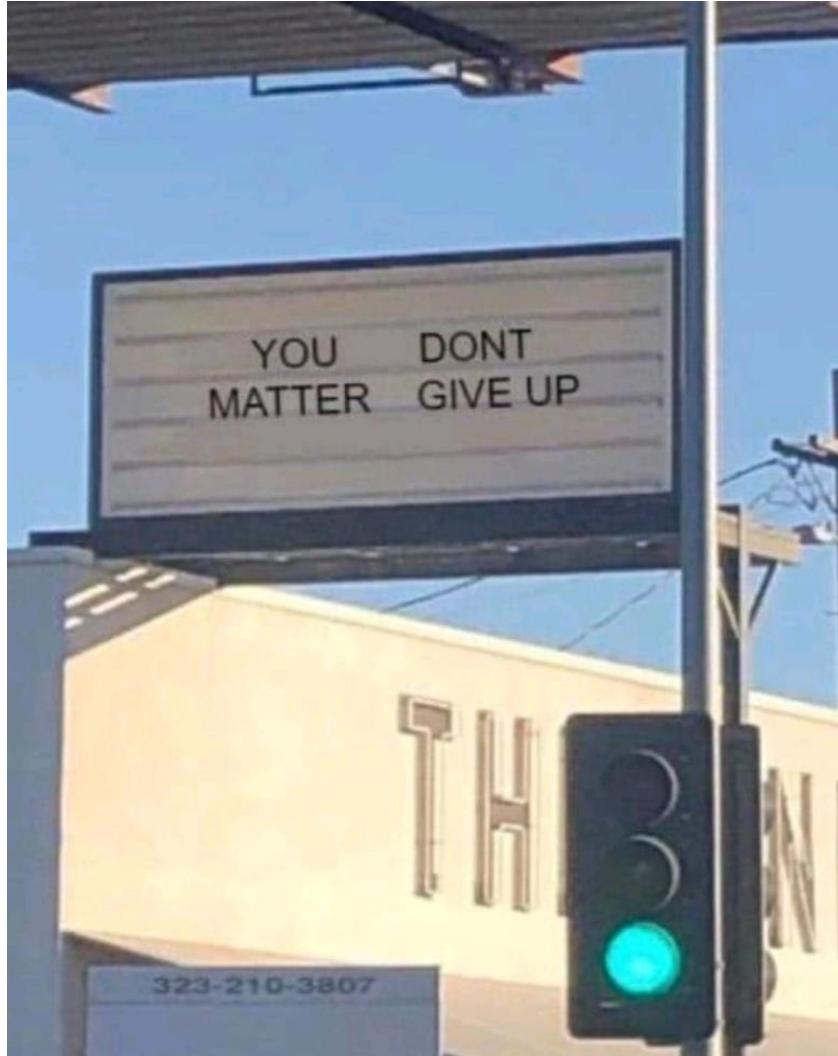
Do: Align text horizontal



Don't: Read from right-to-left



Do: Read from left-to-right...

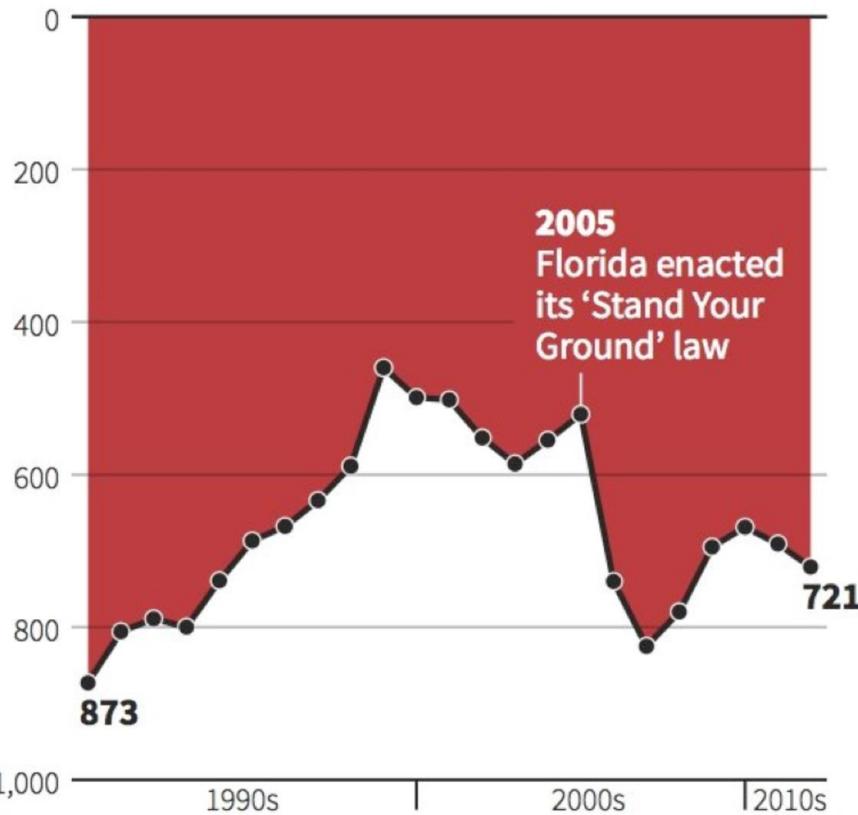


..right?

Don't: Flip the y-axis either

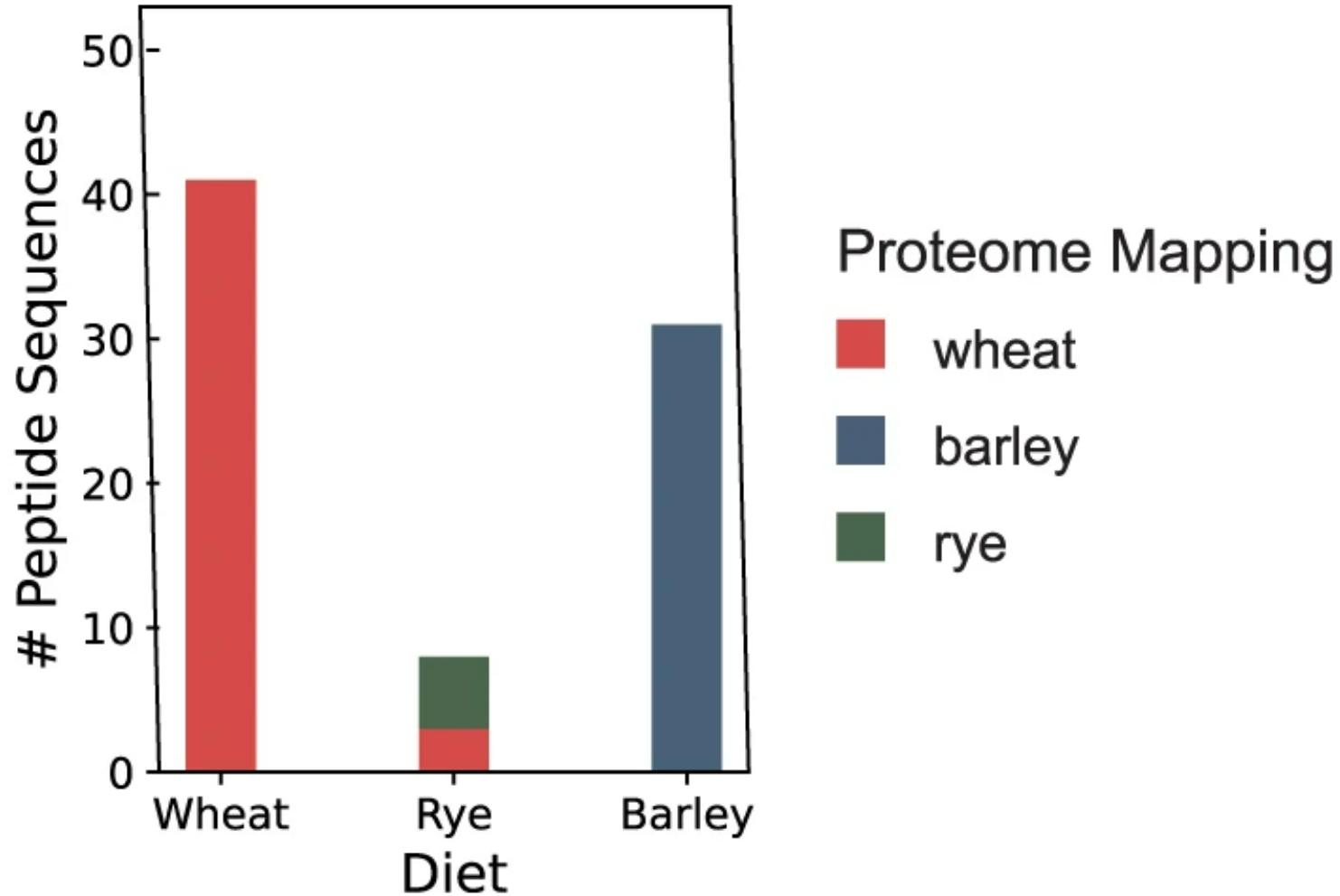
Gun deaths in Florida

Number of murders committed using firearms

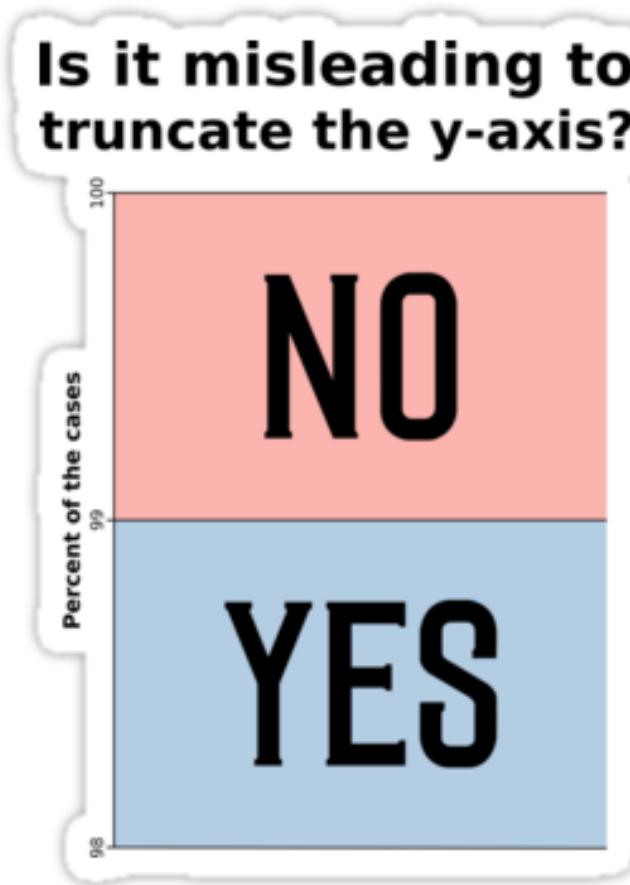


Source: Florida Department of Law Enforcement

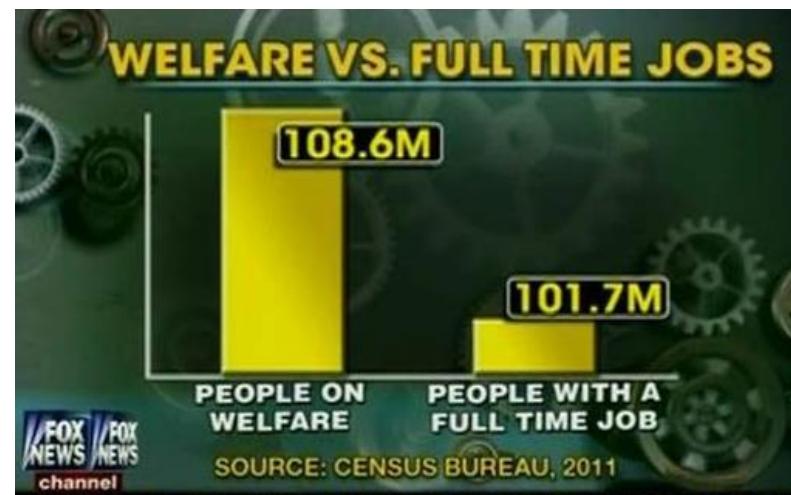
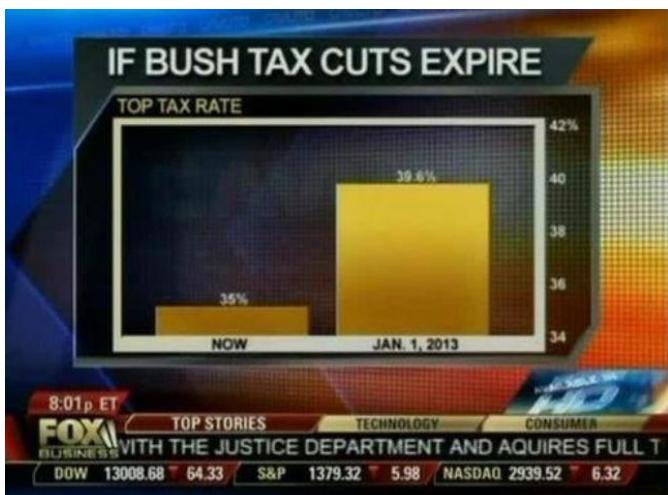
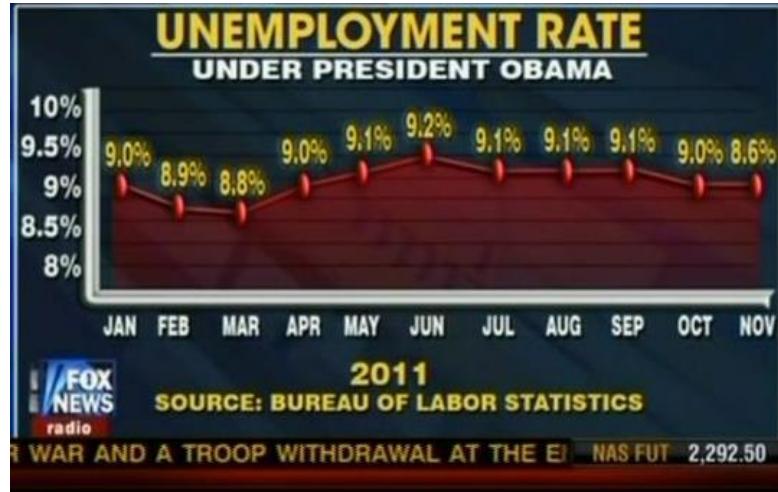
Don't: *Tilt* (?) the y-axis either



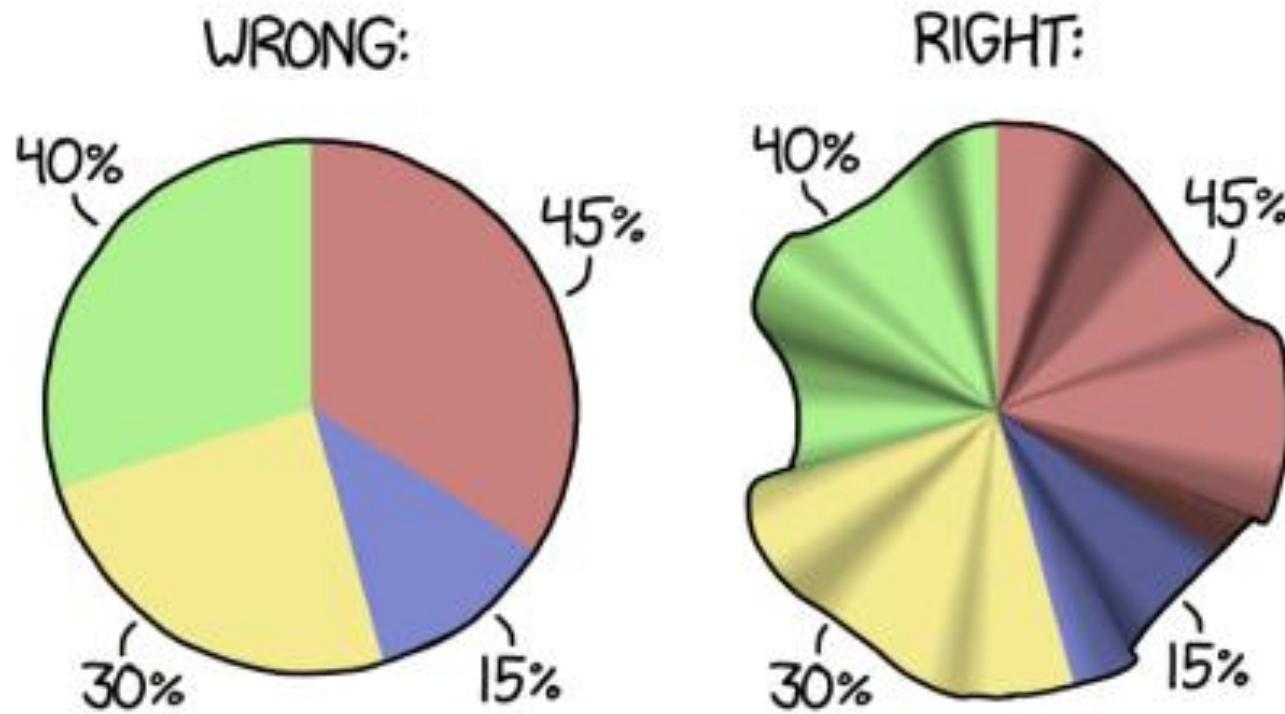
Don't: Truncate the y-axis either



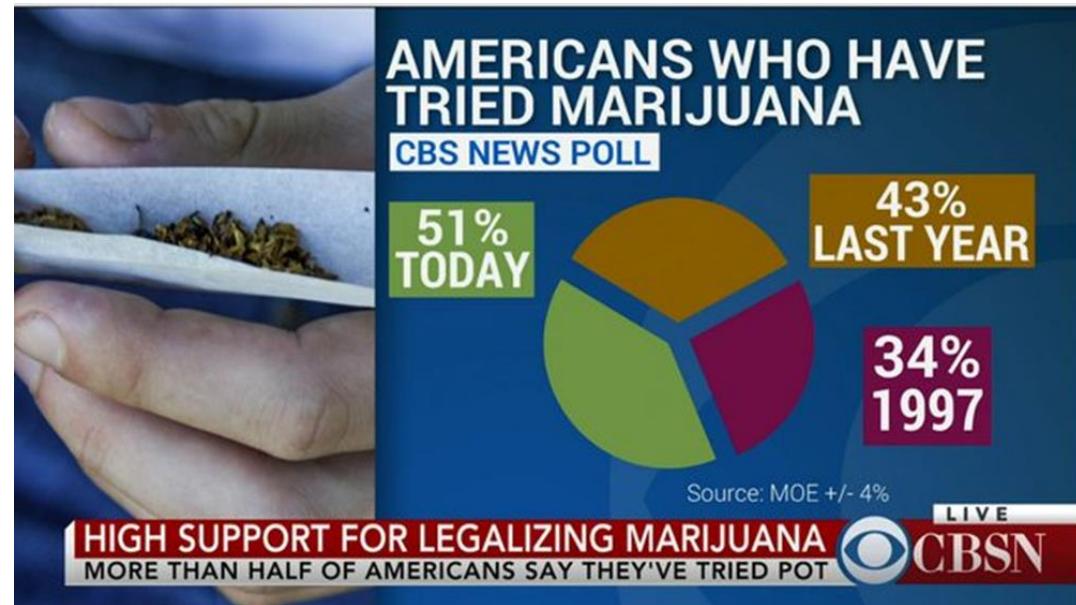
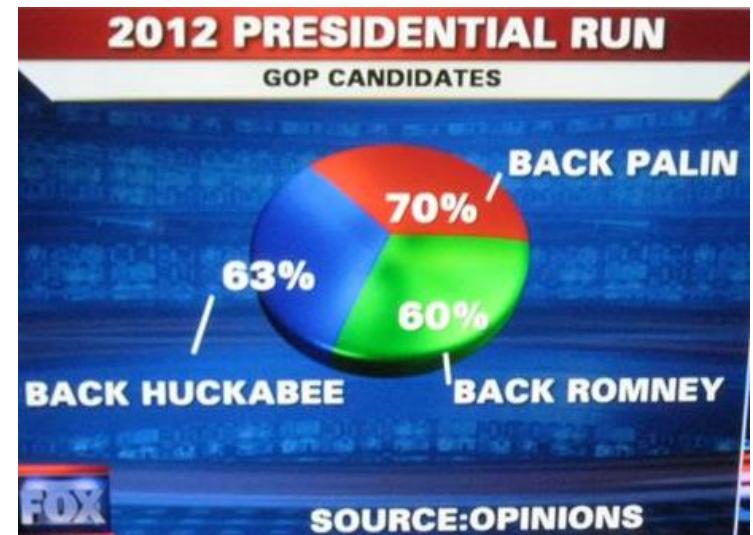
Who does that, you say?



Don't: Use pie charts where the numbers don't add up



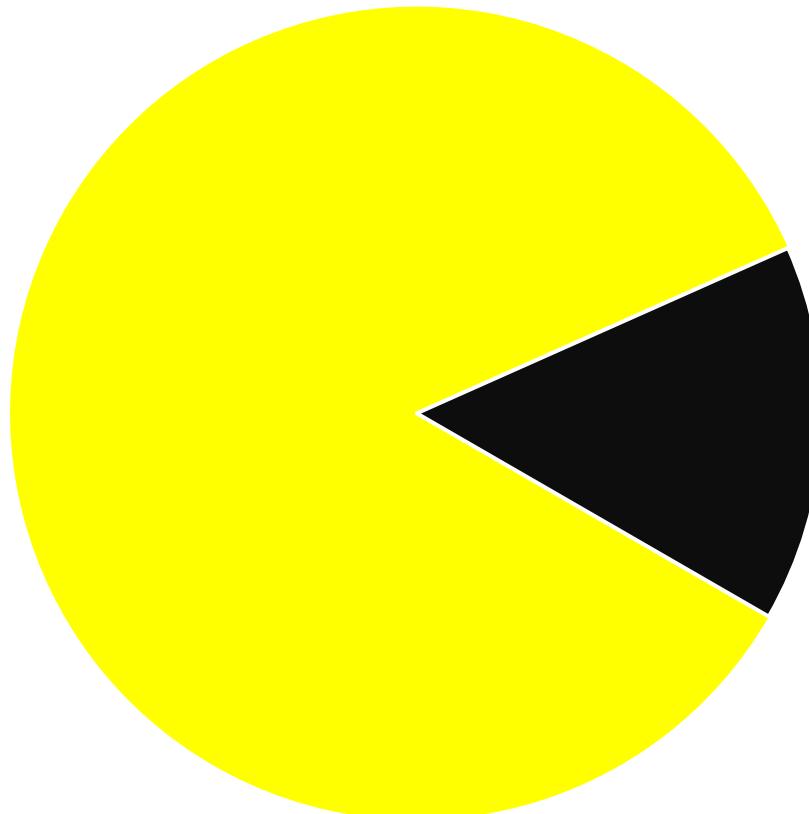
Who does that, you say?



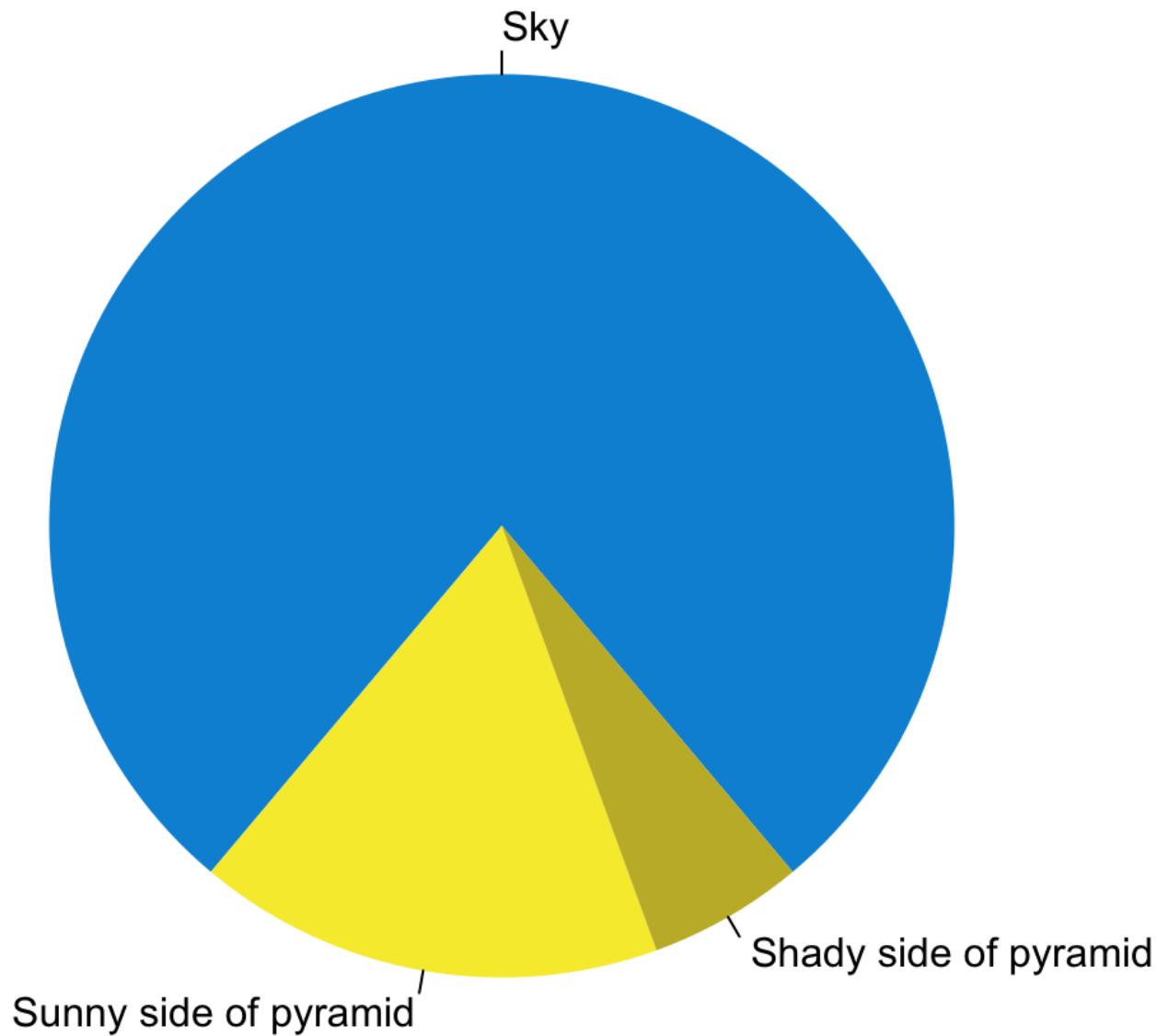
Do: Use pie charts to show the part-to-whole relation of 2 categories

Percentage of chart that looks like Pacman

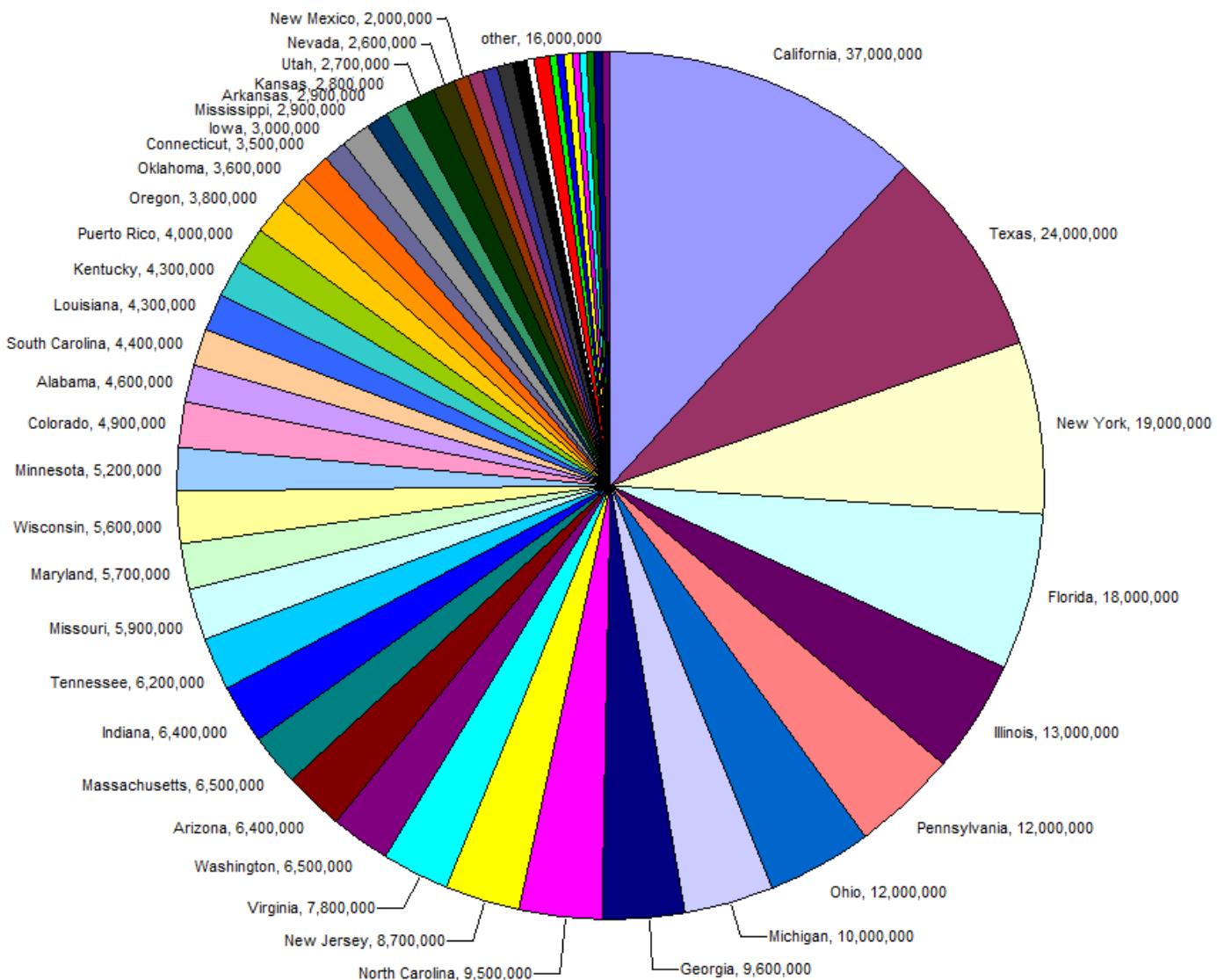
- Pacman
- Not Pacman



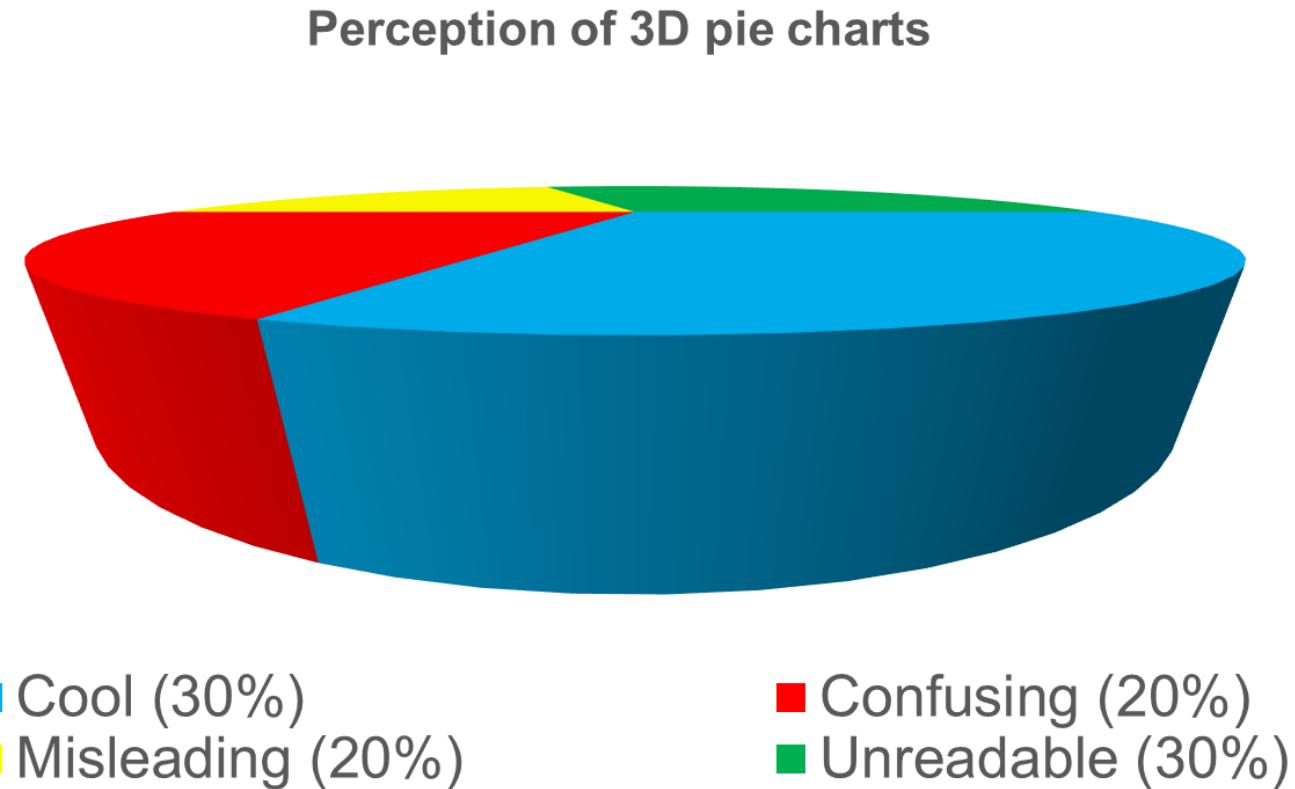
Do: Pie charts for 3 categories is okay too



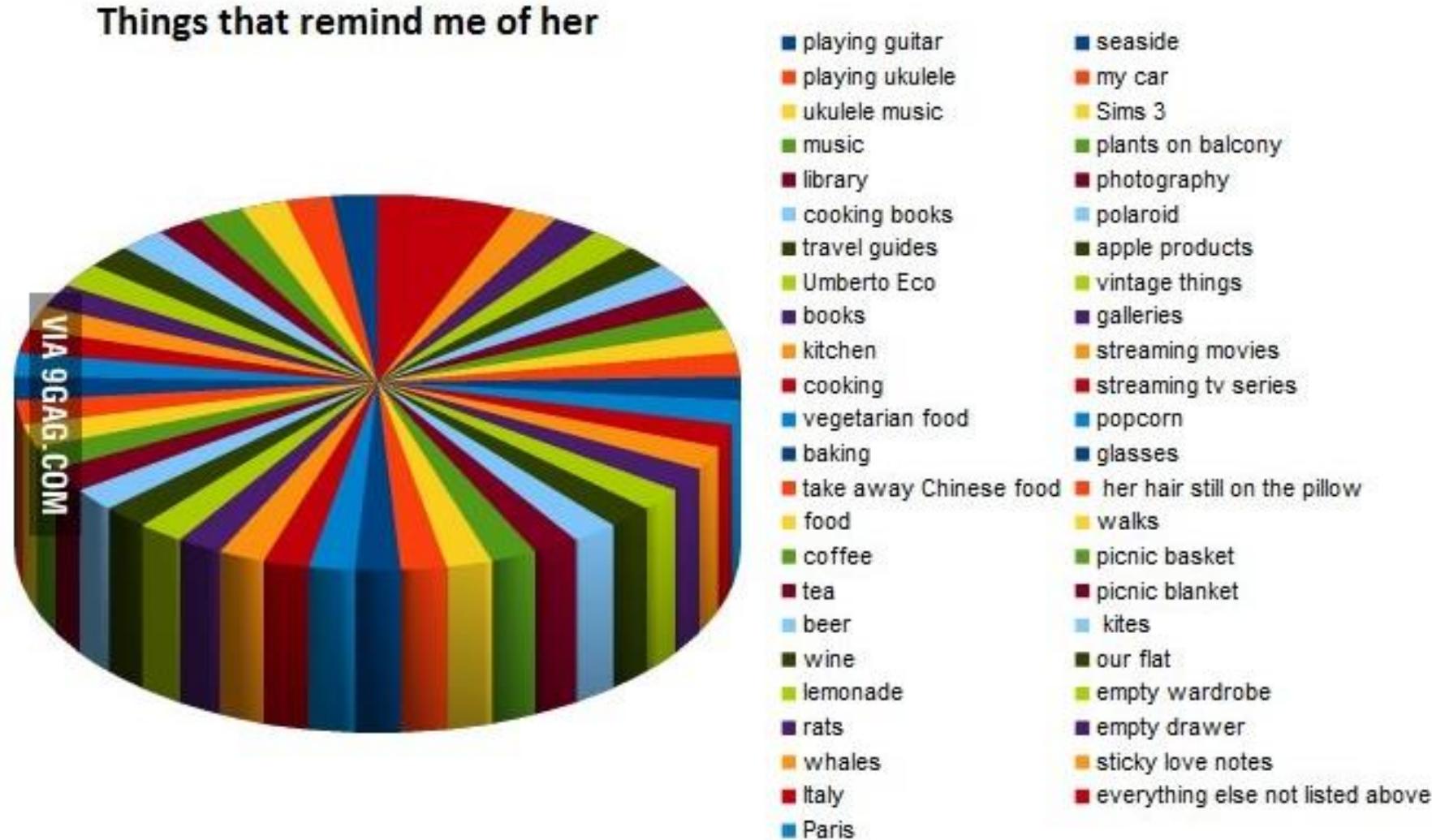
Don't: Use pie charts with >5 categories



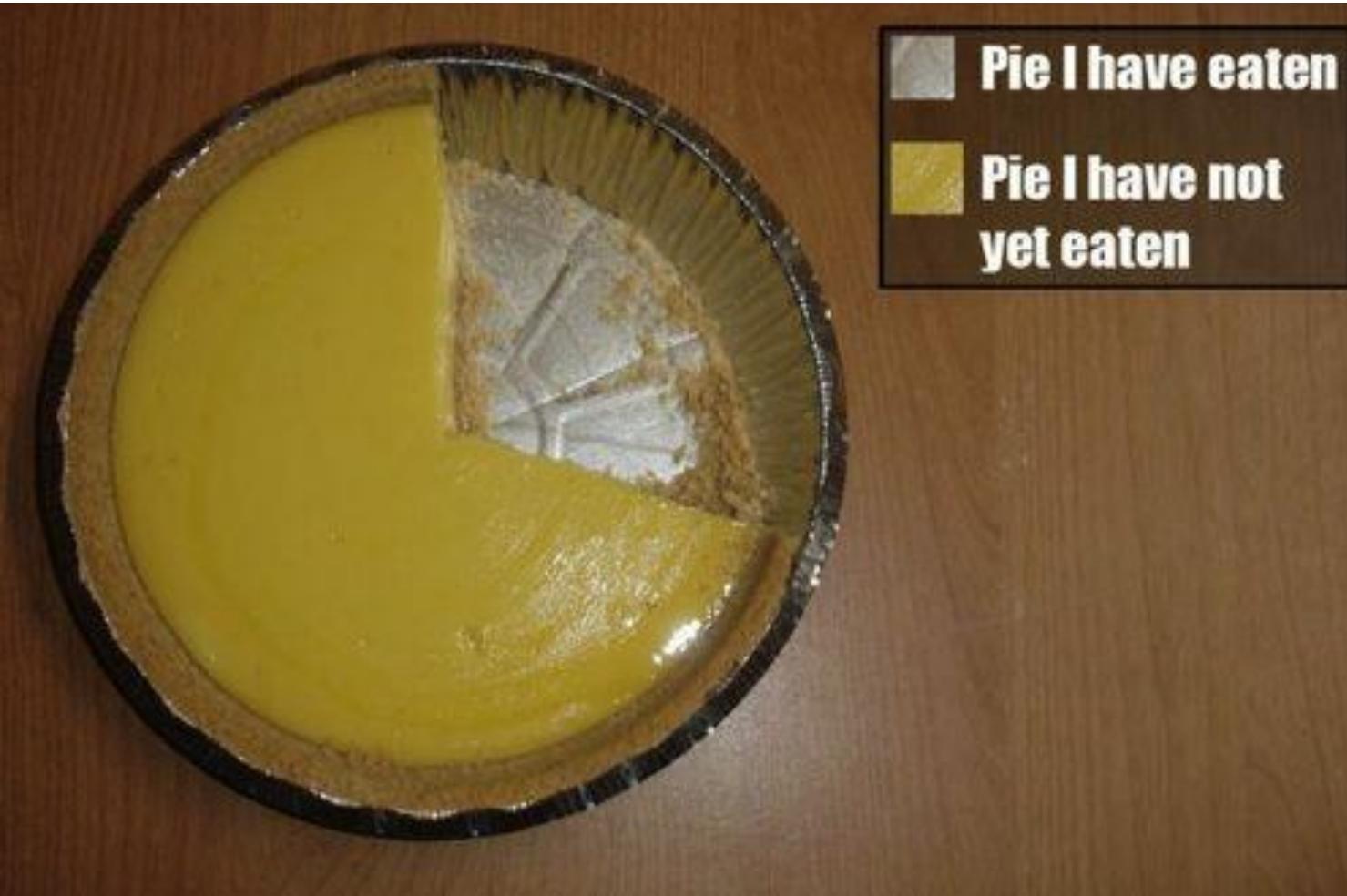
Don't: Use 3D pie charts



Don't: Use 3D pie charts with >5 categories



Do: Know when pie charts are *the best!*



Thank you!



Irene van den Broek, PhD

✉ Irenevandenbroek@gmail.com

LinkedIn: @IrenevdBroek

Twitter: @JeBentWatJeMeet

Happy Viz-ing!



<https://www.jebentwatjemeet.nl>

<https://irenevdb.rbind.io>