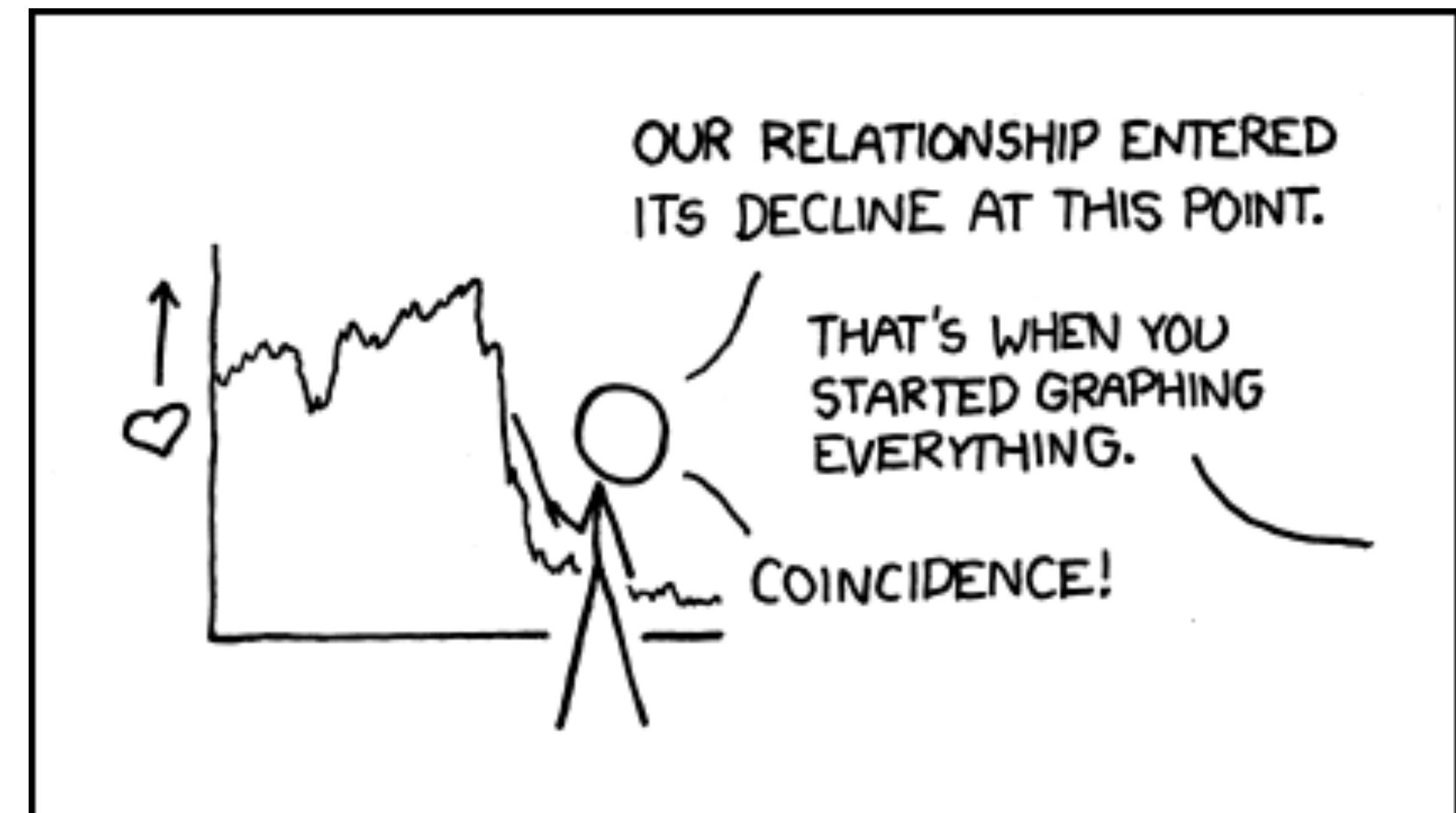


CS-5630 / CS-6630 Visualization for Data Science Design Guidelines

Alexander Lex
alex@sci.utah.edu



Design Guidelines

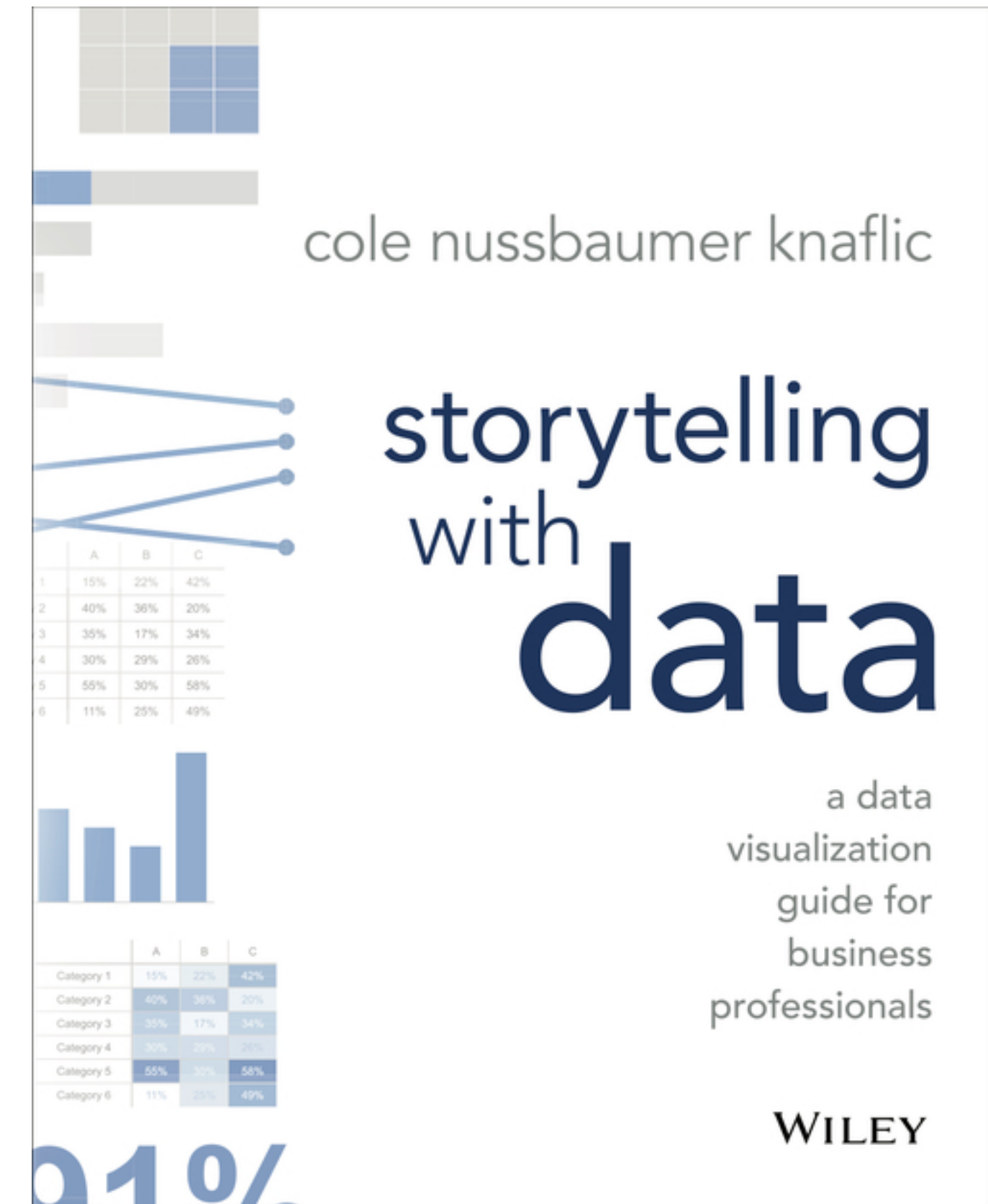
**Rule #1: Use the Best Visual
Channel Available
for the Most Important
Aspect of your Data**

**Rule #2: The visualization
should show all of the data,
and only the data**

Book Recommendation

Great book with simple design guidelines

Not a “Visualization” book, but a “charting” book



Tufte's Integrity Principles

Show **data variation**, not design variation

Clear, detailed, and thorough **labeling** and **appropriate scales**

Size of the **graphic effect** should be **directly proportional to the numerical quantities** (“lie factor”)

Scales

The Lie Factor

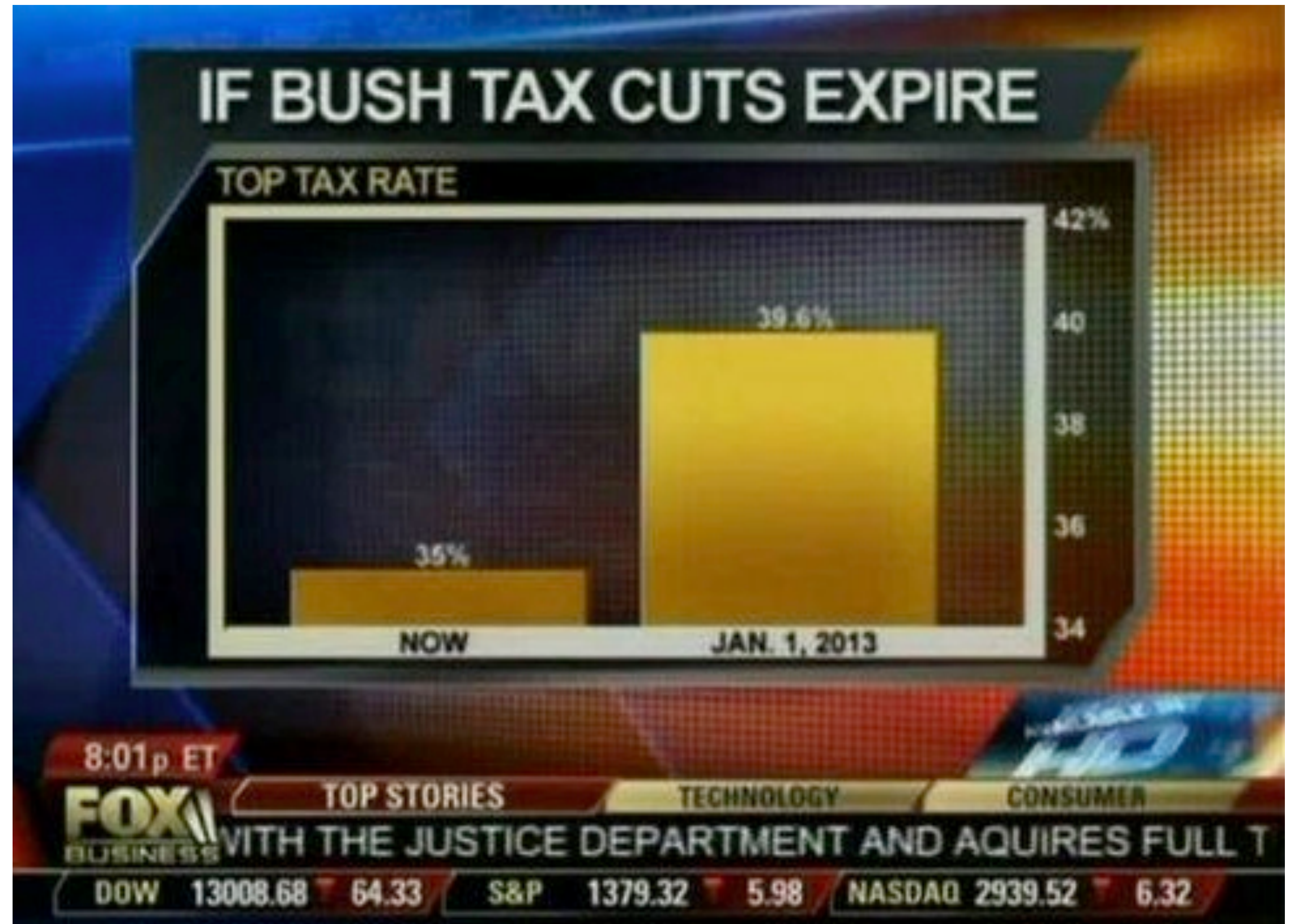
Size of effect shown in graphic

Size of effect in data

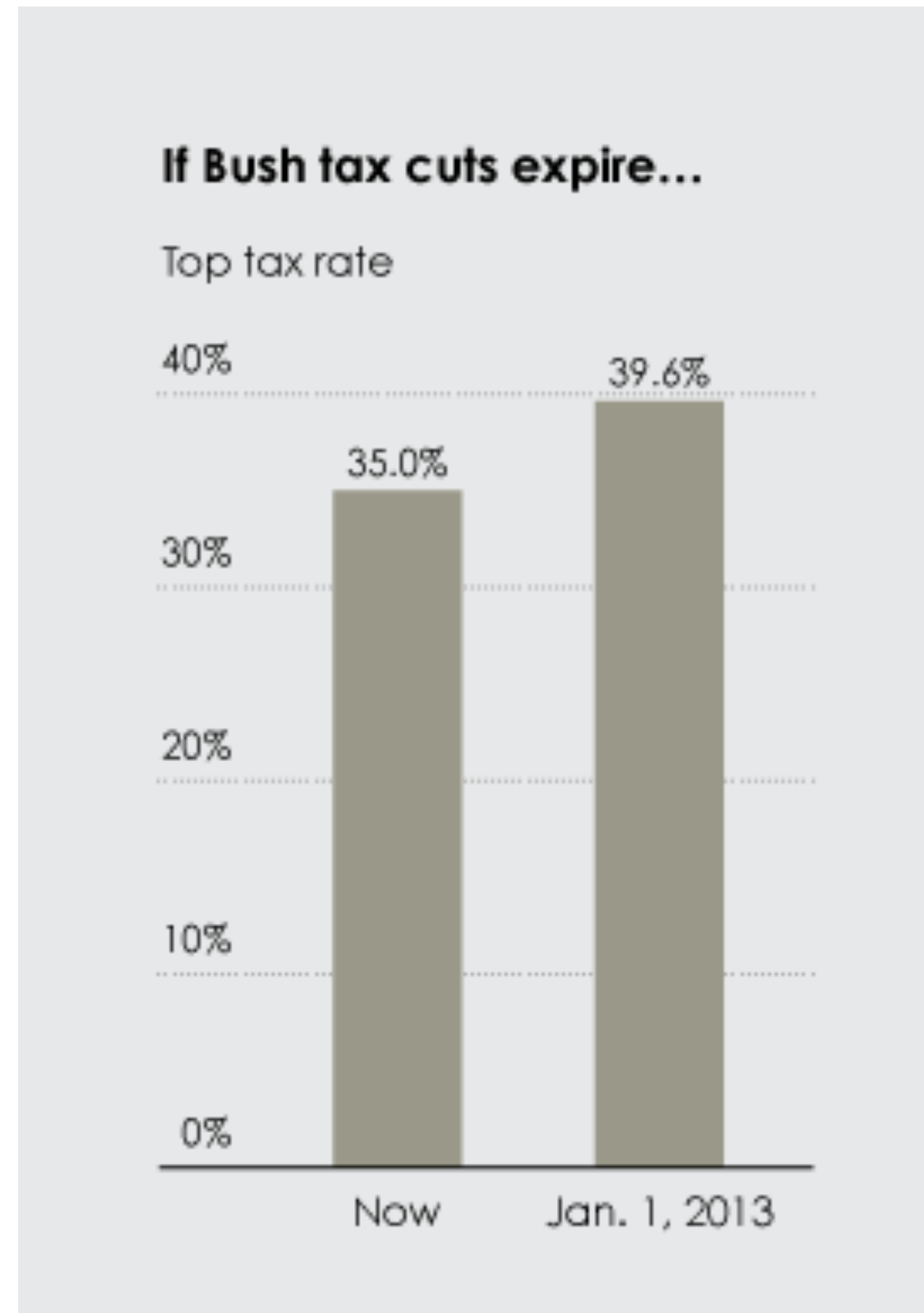
Lie Factor - Graphical Integrity

Magnitude in data
must correspond to
magnitude of mark

Effect in Data: factor 1.14
Effect in Graphic: factor 5
Lie Factor: $5/1.14 = 4.38$



Scale Distortions



What's wrong?



Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

What's wrong?



Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um



Viele Bezieher mit "ungeklärter Staatsbürgerschaft"

Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

1
2
3
Asyl
Opf
frei
"Krc
Räd
geh
Zwe
Mit
Wol
Am
Stre
Mes
Abe

What's wrong?

Grafik der Kronenzeitung



Zusätzlich geht die Mindestsicherung in Wien auch an 1314 Deutsche, 369 Italiener, 66 Schweden, 59 Schweizer, zehn Kanadier, dazu an einen Liechtensteiner, einen Isländer sowie an einen Bürger von Andorra.

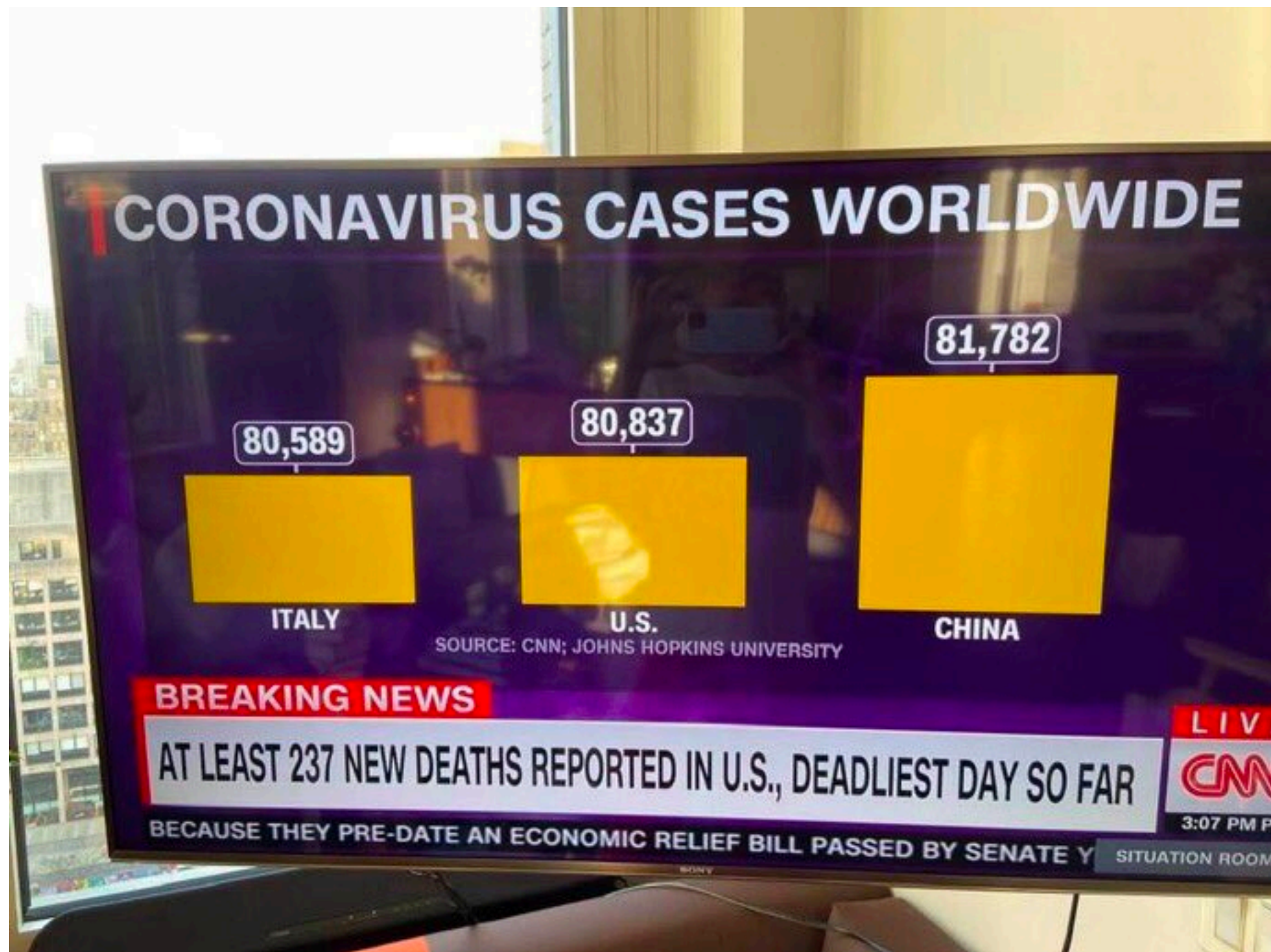


Viele Bezieher mit "ungeklärter Staatsbürgerschaft"
Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um

Grafik in echt



Viele Bezieher mit "ungeklärter Staatsbürgerschaft"
Die größte Gruppe in der Liste der Mindestsicherungsbezieher ist aber jene der "ungeklärten Staatsbürgerschaft". Dass es sich bei den 16.712 Personen um



OBAMACARE ENROLLMENT

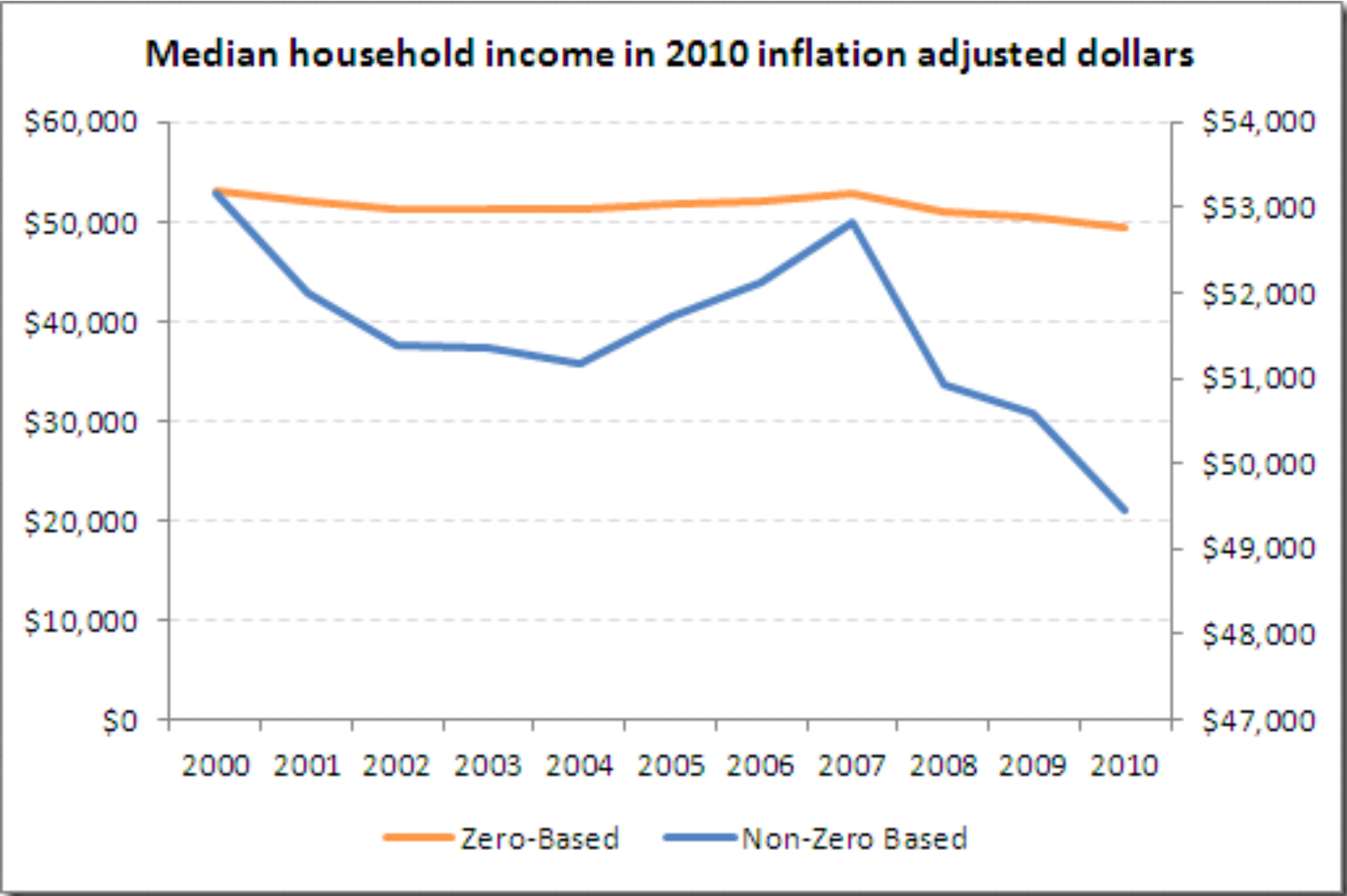


ACTUAL
ENROLLMENT

GOAL



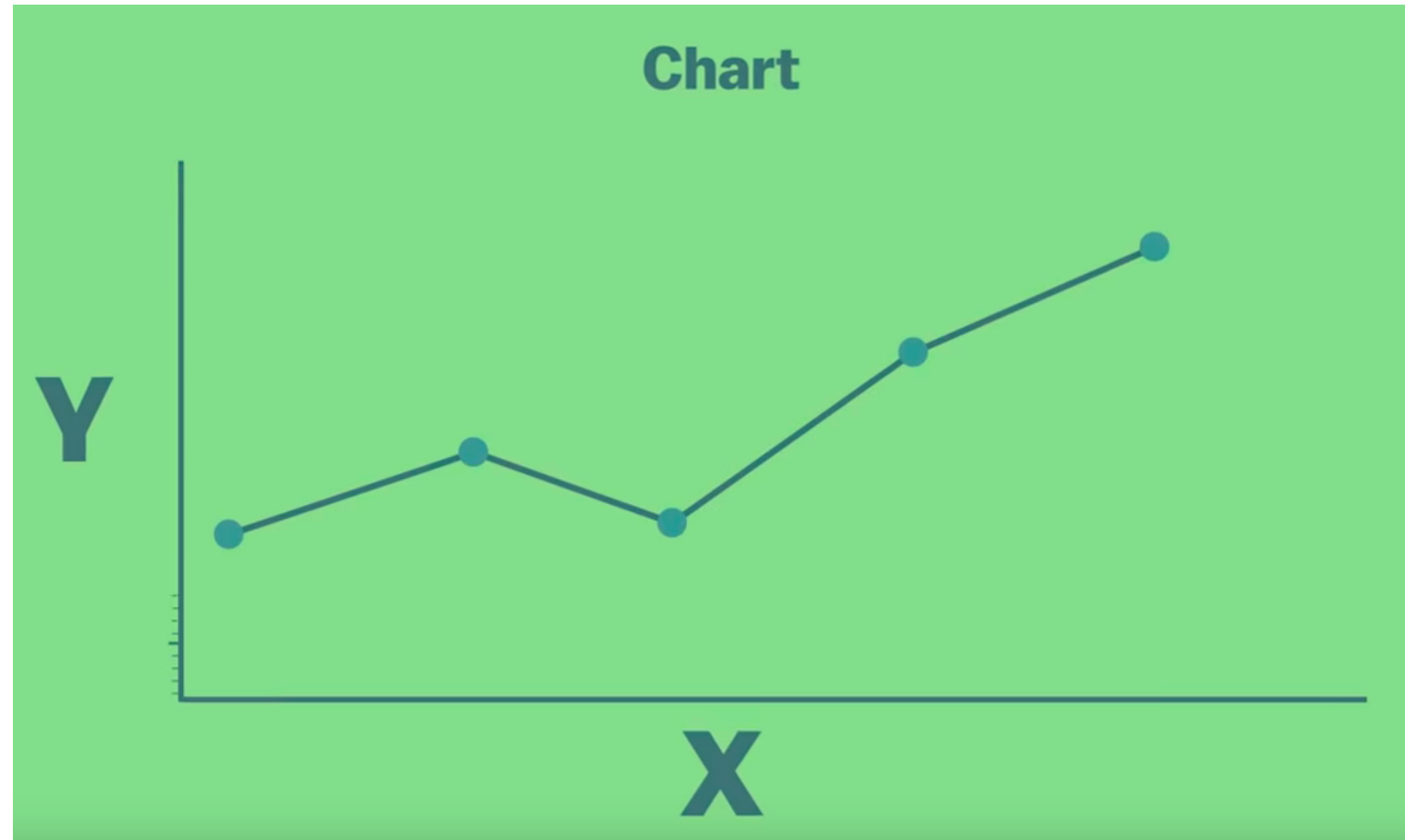
Start Scales at 0?



Use a baseline that shows
the data, not the zero-point.

Think about: what is a meaningful baseline?

Scales at 0

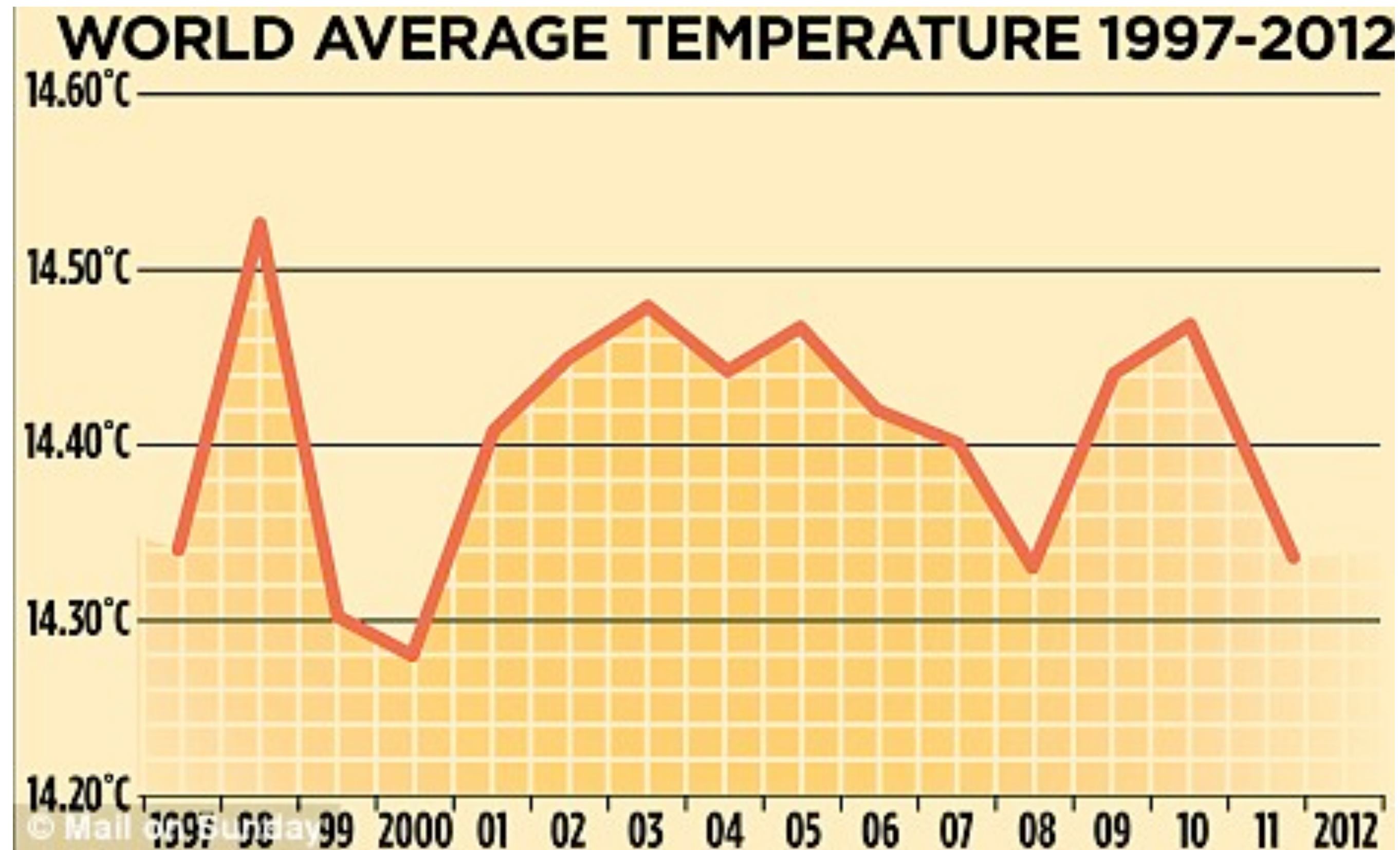


Framing

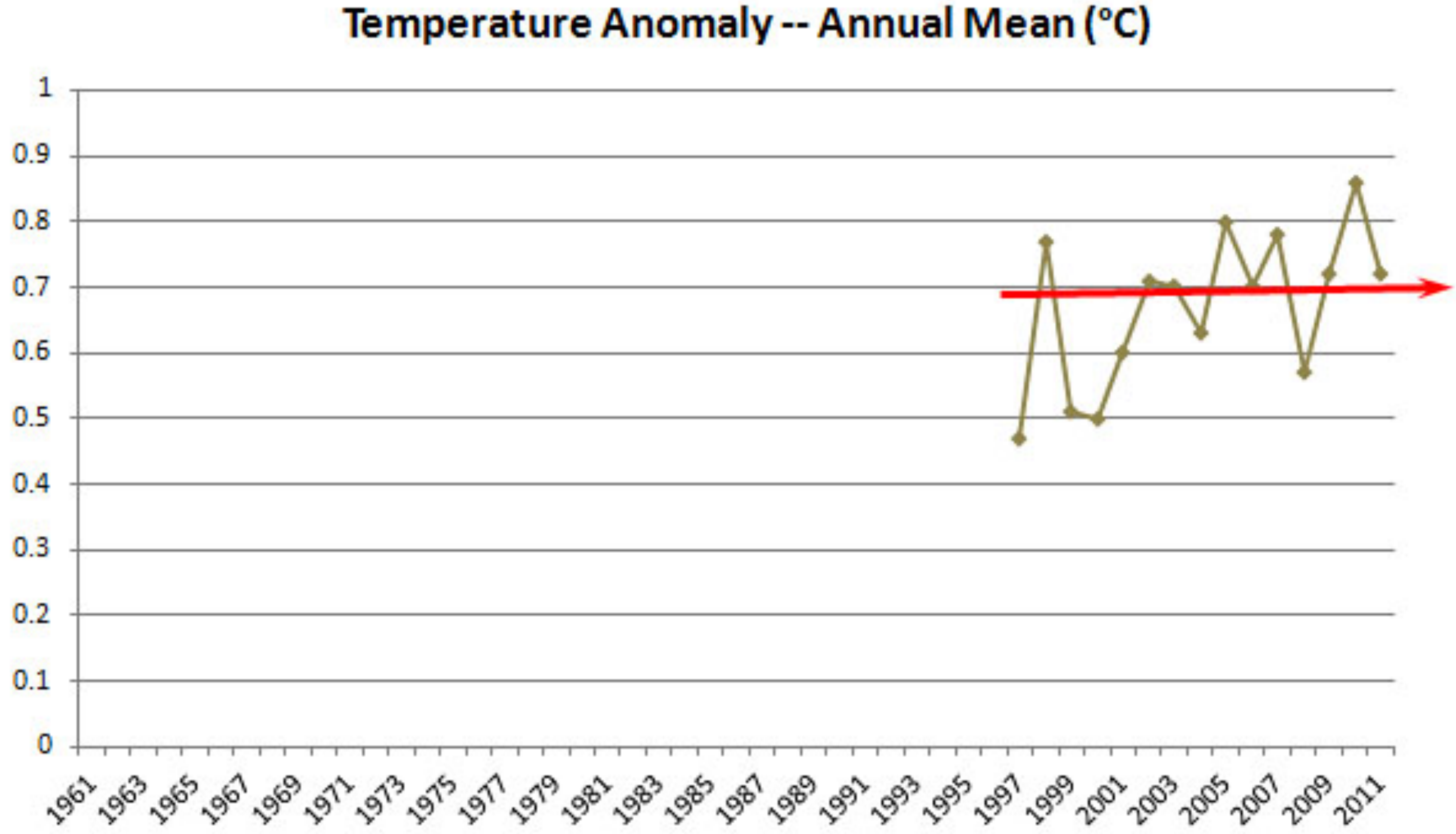
Vis can be used to lie
just as language or statistics

When showing something, make sure that you're faithful to
the data

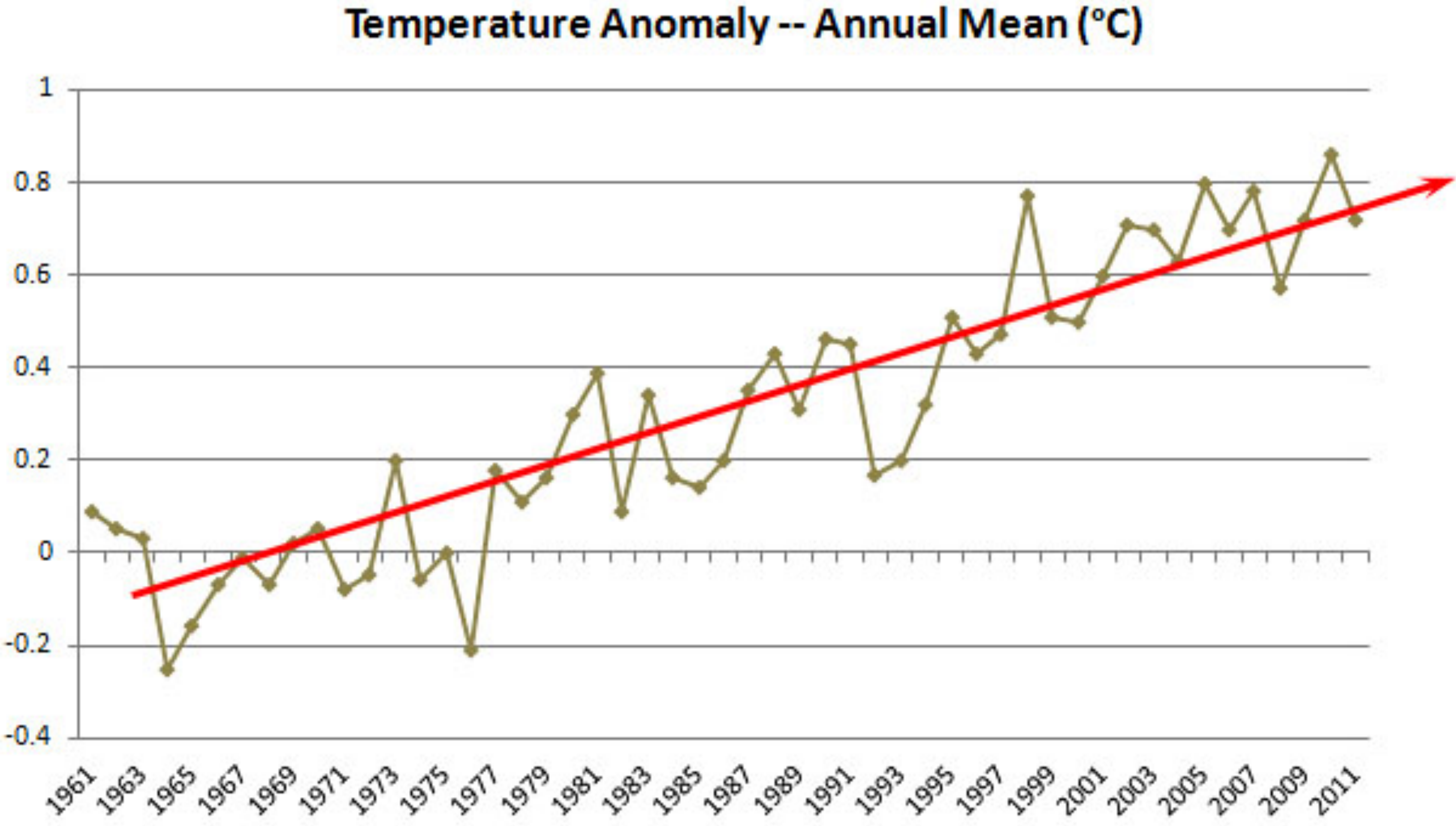
Global Warming?



Global Warming?



Global Warming - Frame the Data



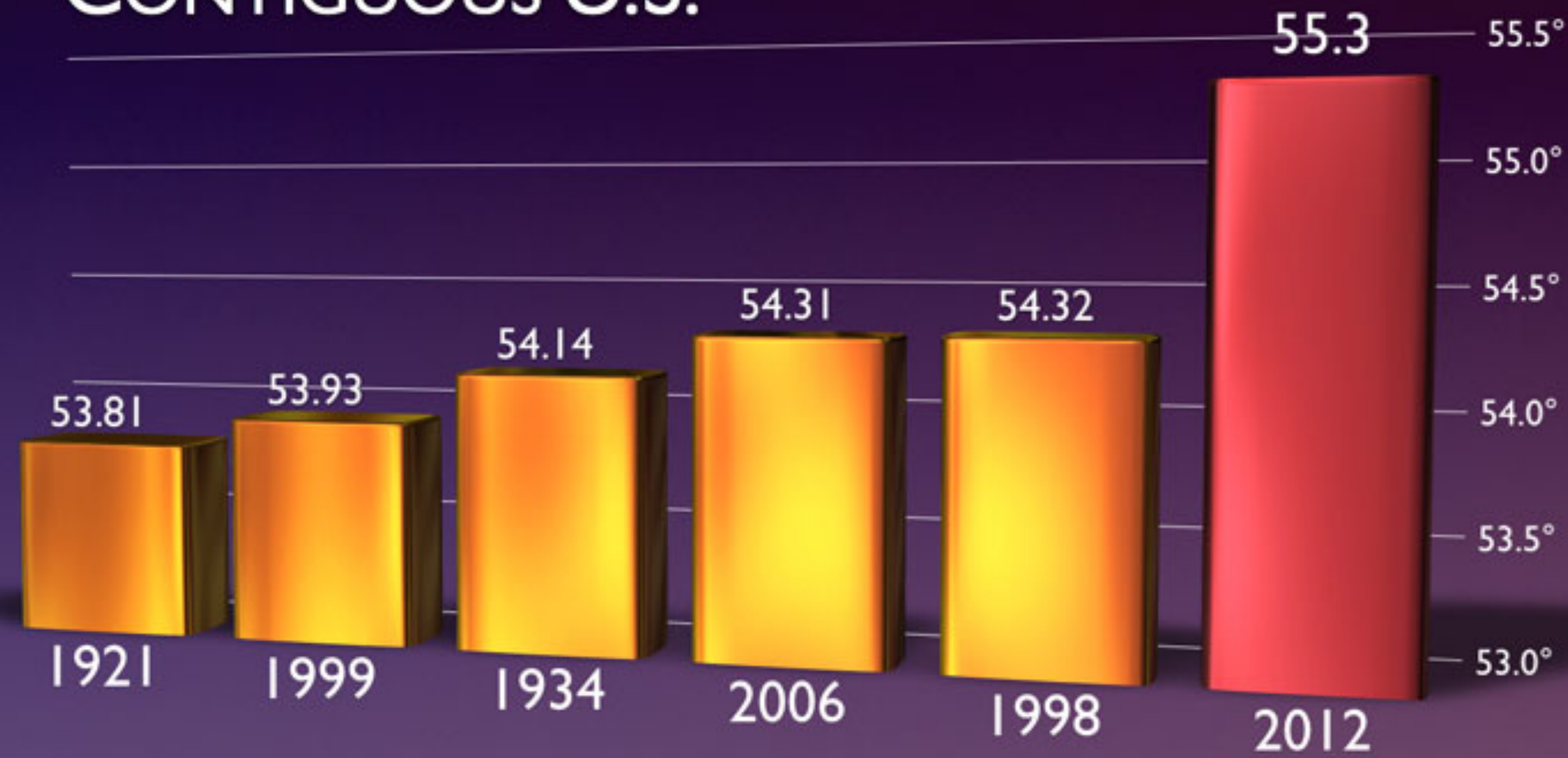
Also see: [USA Temperature: can I sucker you?](#)

What's wrong?

HOW 2012 STACKS UP

THE WARMEST YEARS ON RECORD

CONTIGUOUS U.S.



Source: NOAA's National Climatic Data Center - State of the Climate National Overview

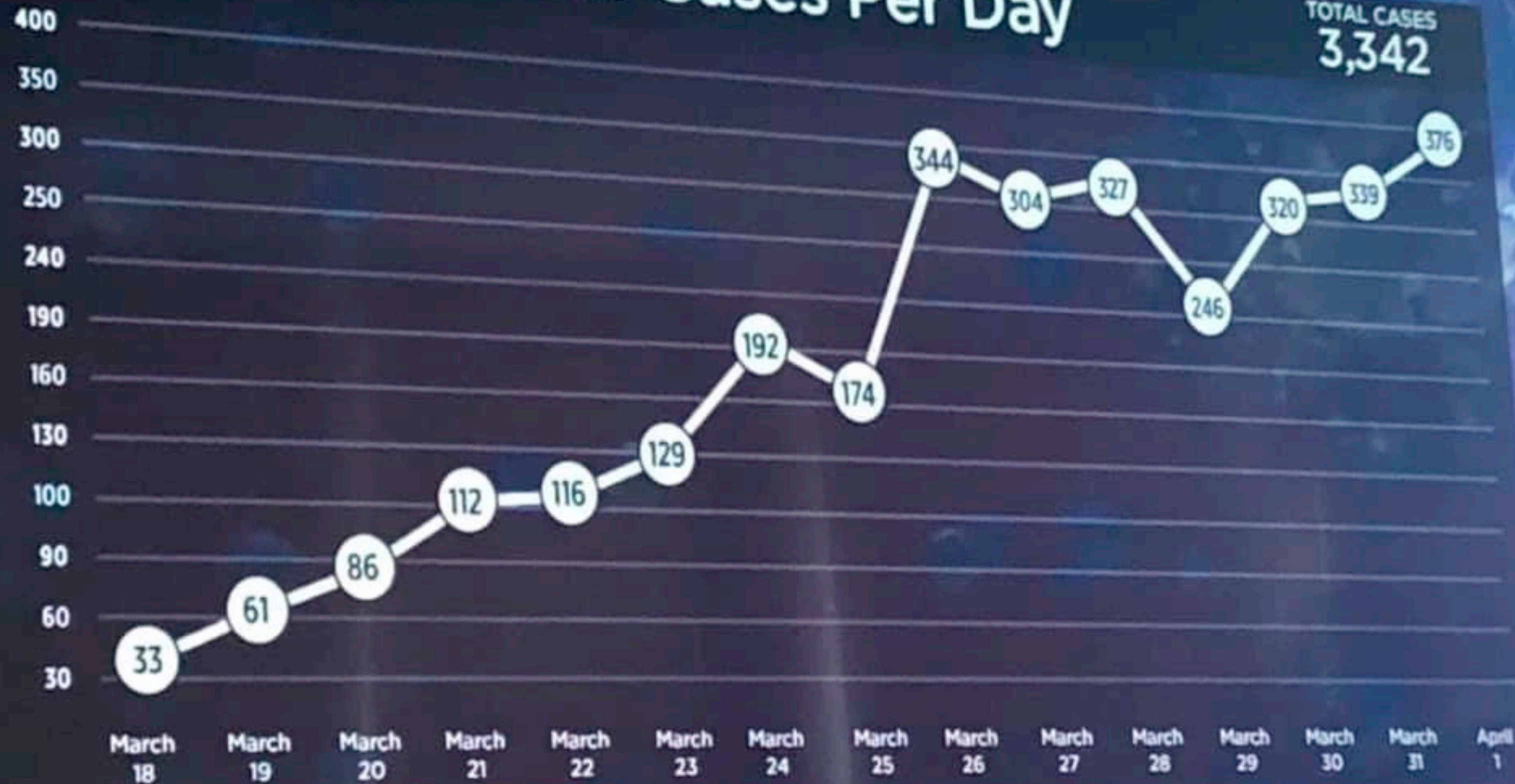
CLIMATE  CENTRAL

Scale Distortions in Temporal Data



New Cases Per Day

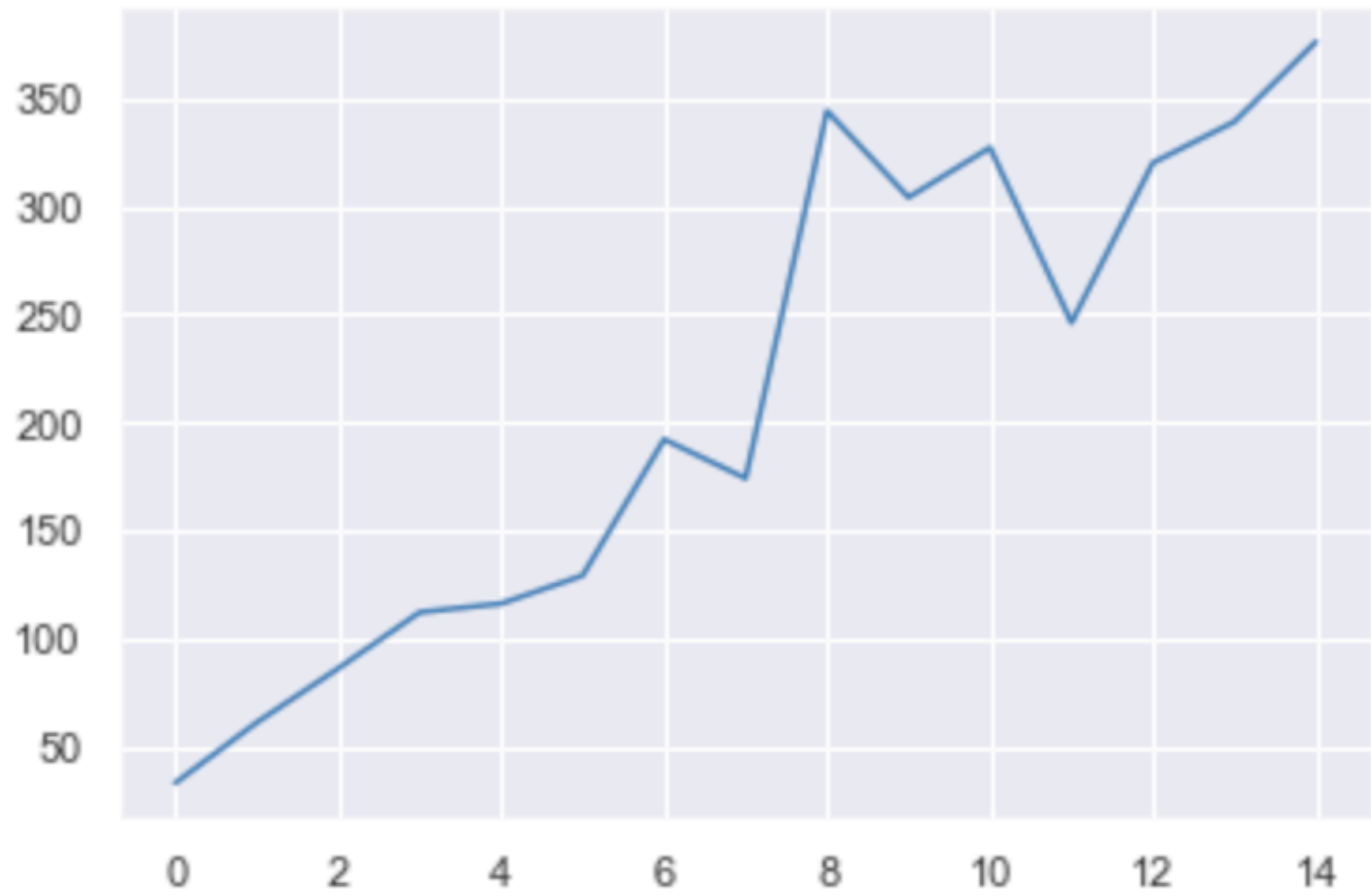
TOTAL CASES
3,342



FOX

31

:01
570



Log Scales

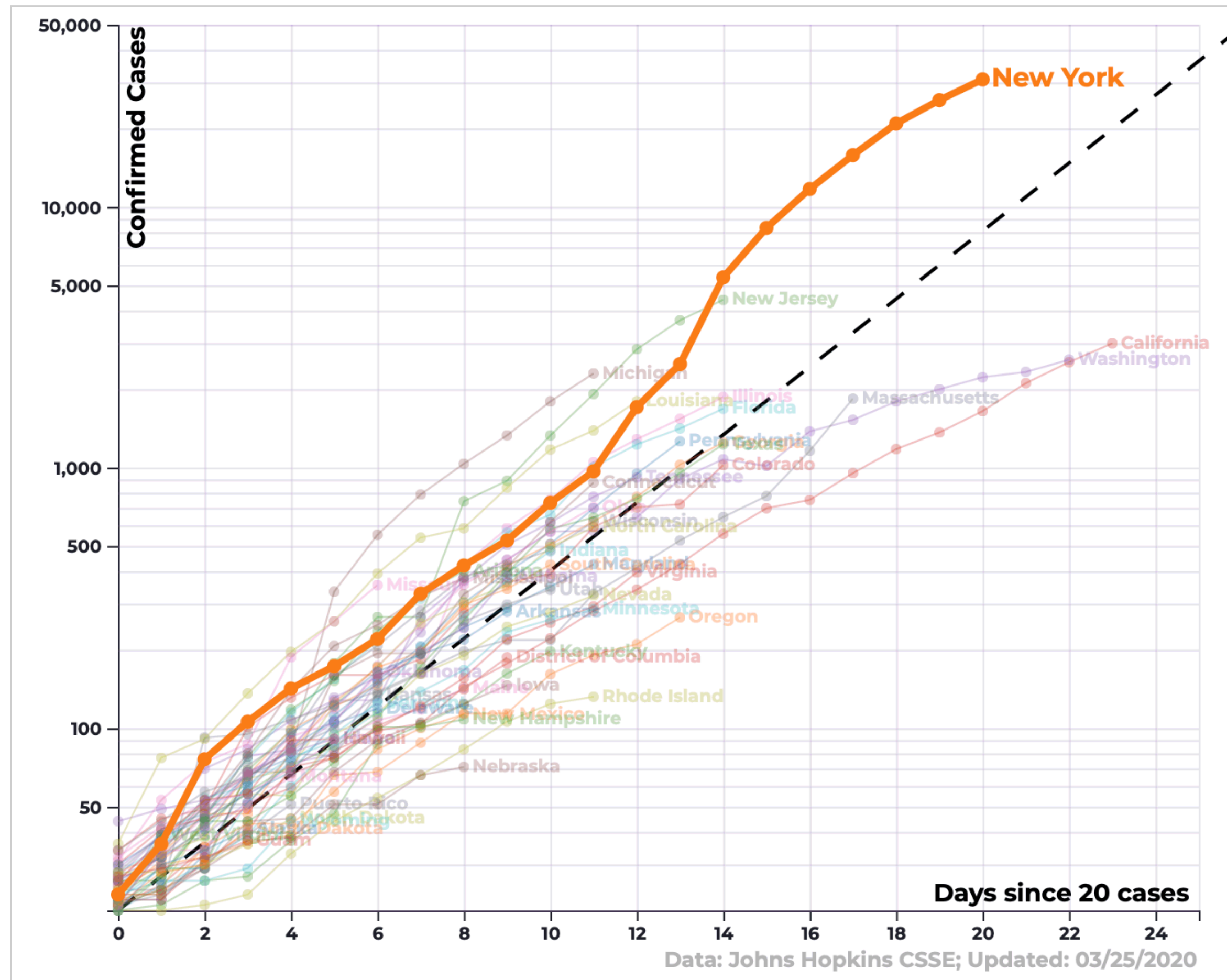
Use log scales if the underlying data warrants it

Typical use case: exponential growth curves

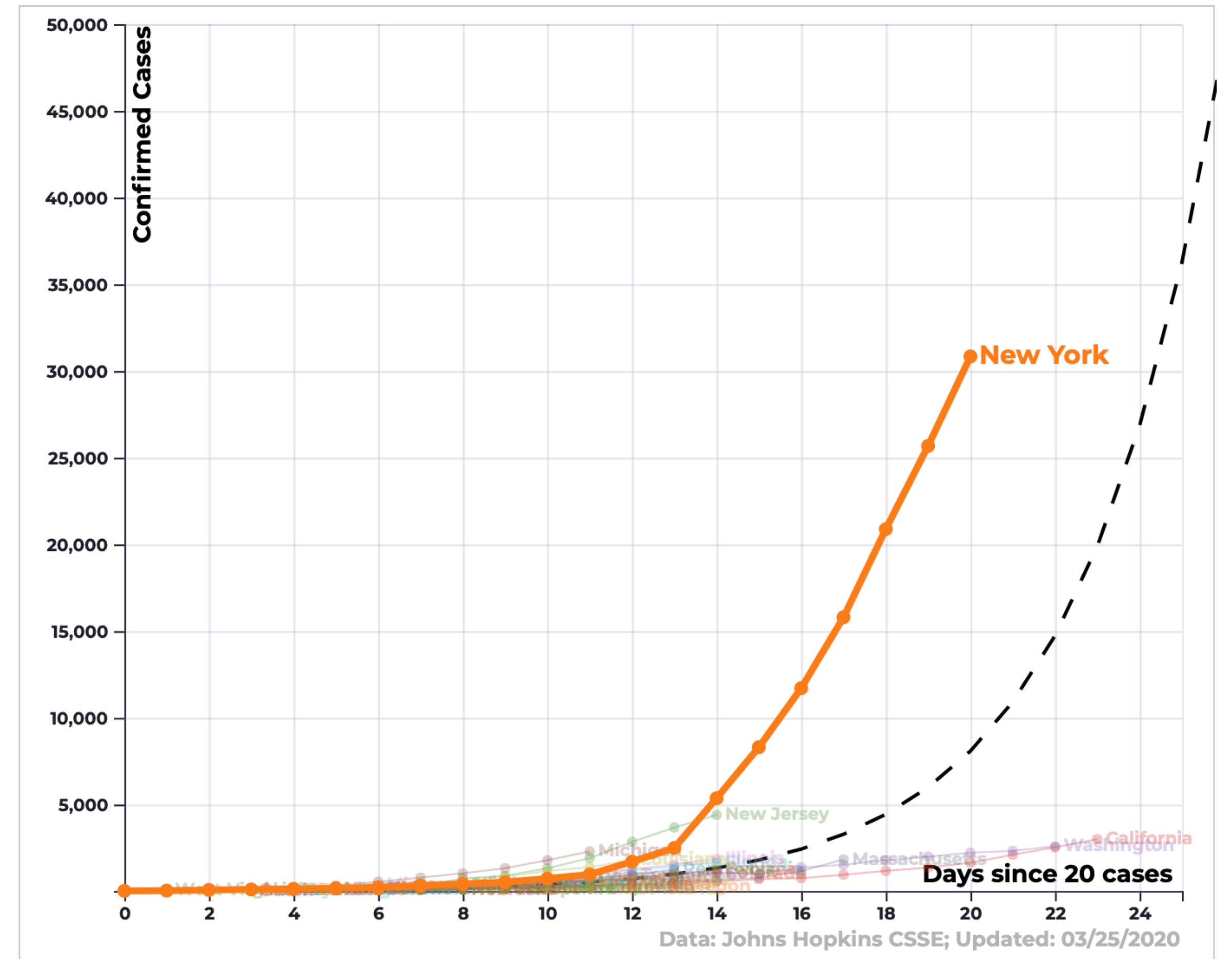
In practice: an expert tool

What are some interpretations?

COVID-19 Cases by US States/Territories



COVID-19 Cases by US States/Territories

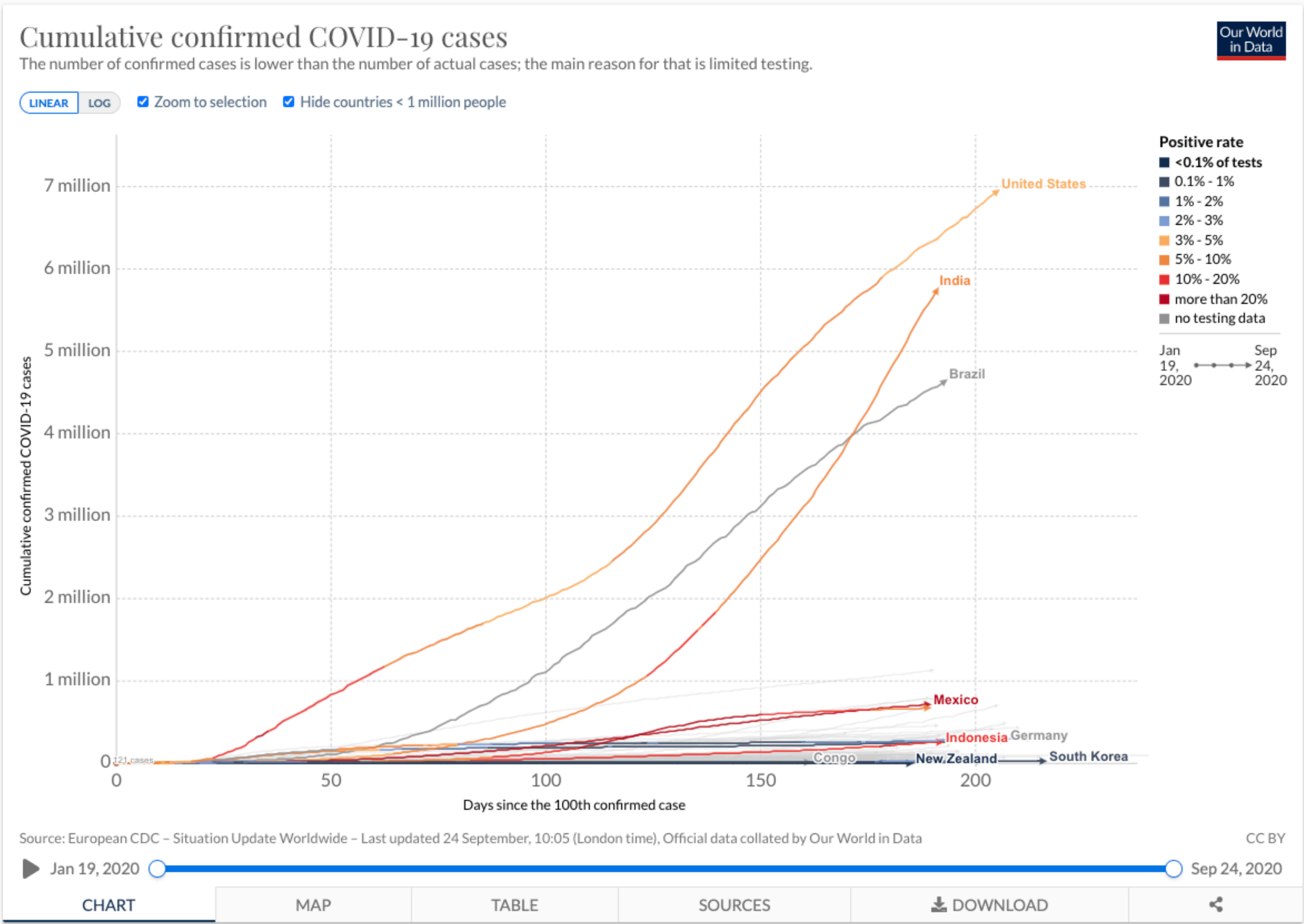


Normalization

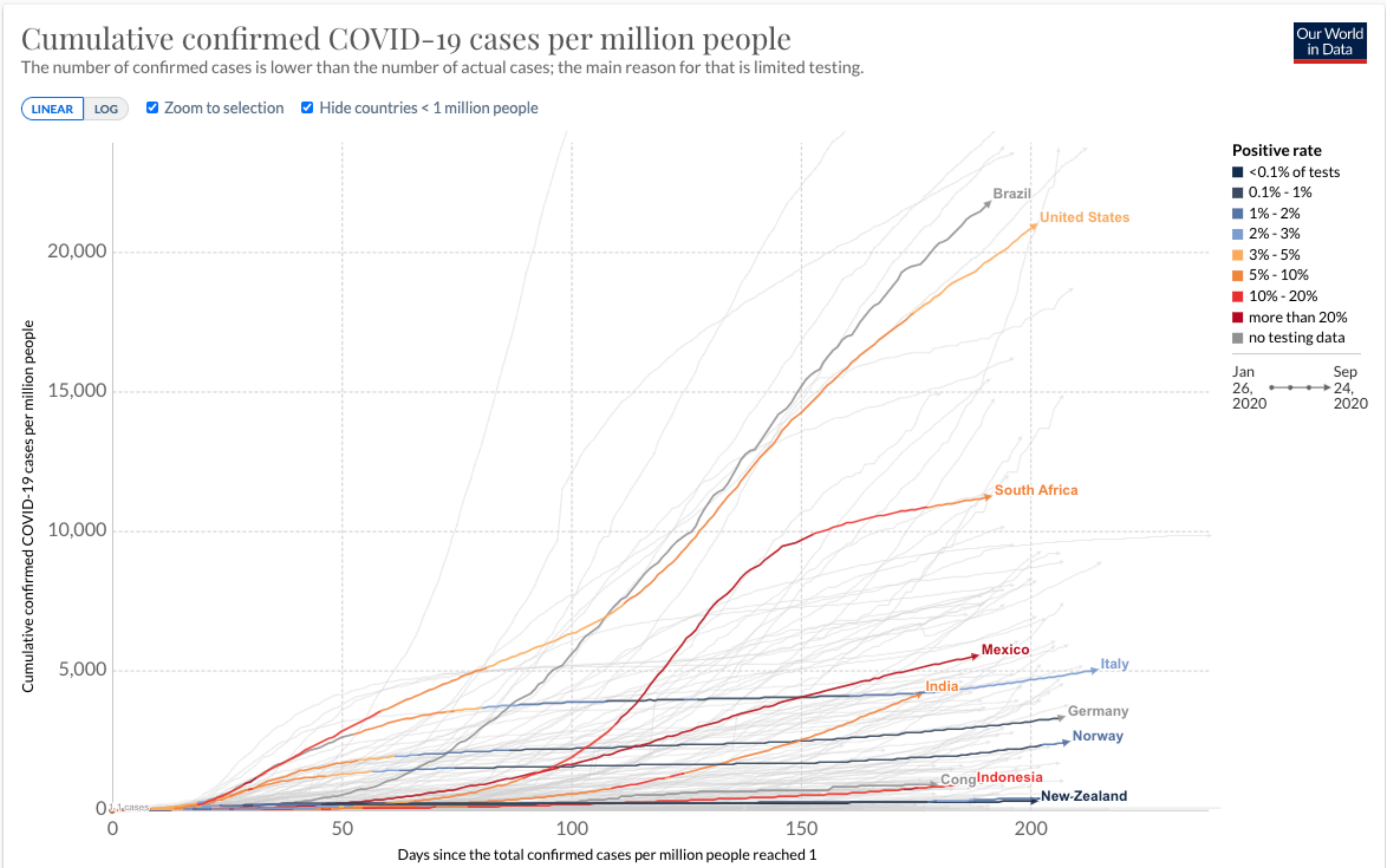
Comparing Apples to Apples

When we compare things that are different, we need to account for that difference. Normalize your data!

Cumulative Cases



Cumulative Cases Per Million



Different Perspectives

To get the full picture, you might look at more than one chart: <https://ourworldindata.org/coronavirus>

[HOME](#)

[TEAM](#)

[PUBLICATIONS](#)

[BLOG](#)

[CULTURE](#)

[POSITIONS](#)



visualization
design lab

The Case Against Dashboards (when Visualizing a Pandemic)



Alexander Lex
July 6, 2020

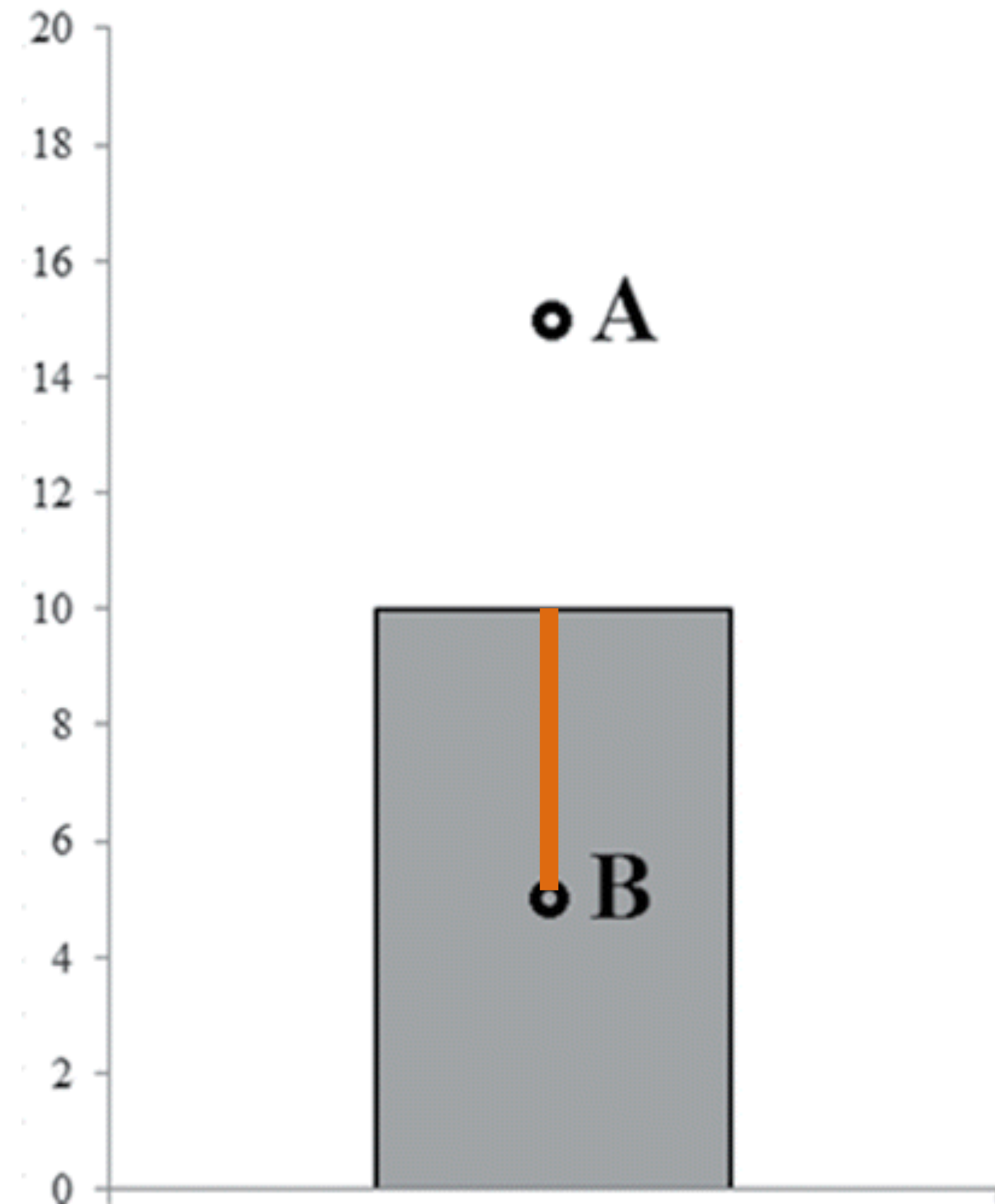
tldr: Using dashboards comes with risks: they leave out critical context by over-simplifying and hence give false certainty. A more nuanced approach including interpretation by experts, and showing multiple perspectives is needed when visualizing data for something as complex as the COVID-19 pandemic.

The COVID-19 pandemic of 2020 has negatively impacted our lives in many ways. The anxiety felt by many is amplified by the obsessive consultation of the latest numbers and statistics about cases, testing rates, deaths, and so on. Both the public and experts have

Distributions

Height of the Bar encodes
mean of a distribution

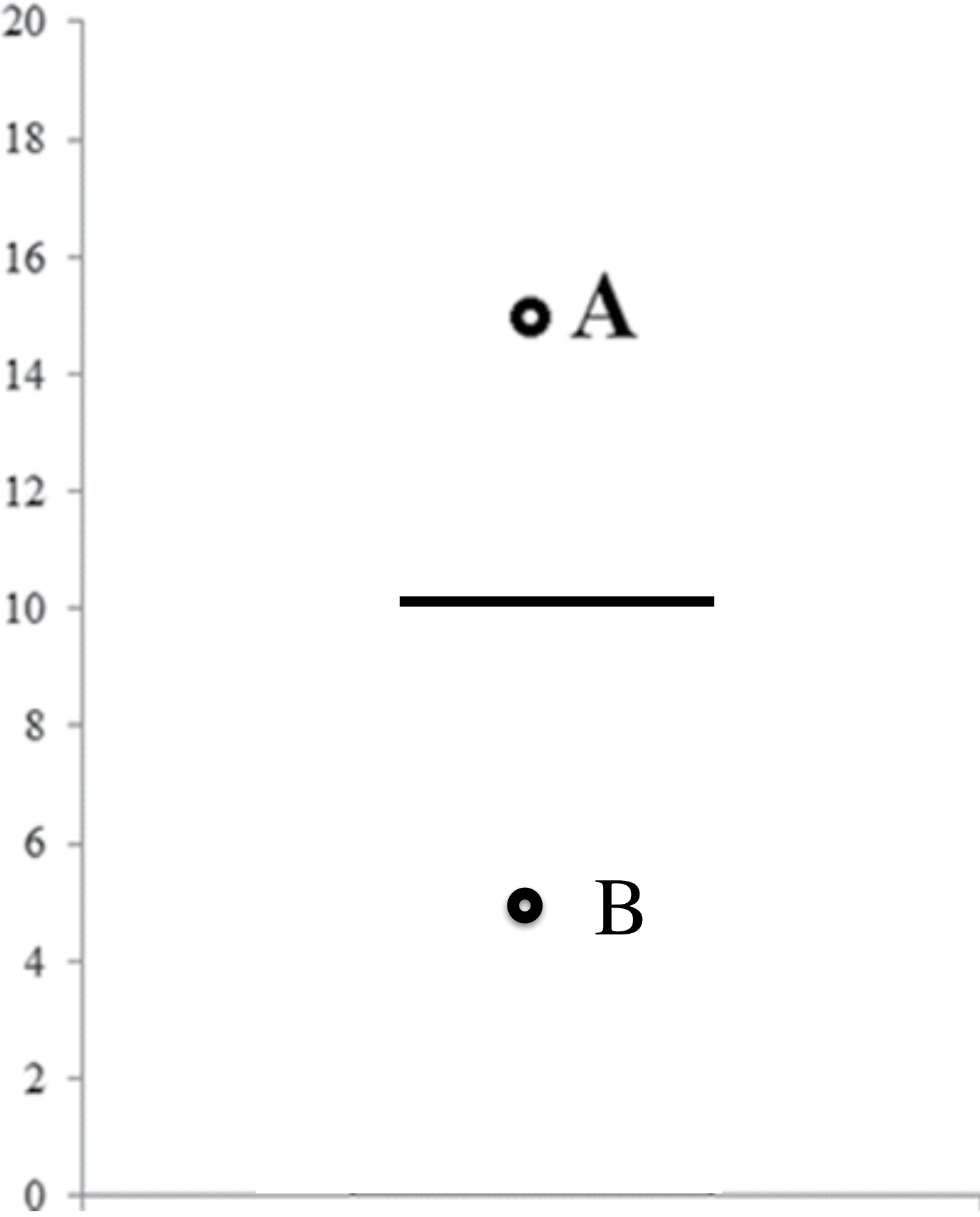
Which value is more likely to
belong to the distribution?
A or B?



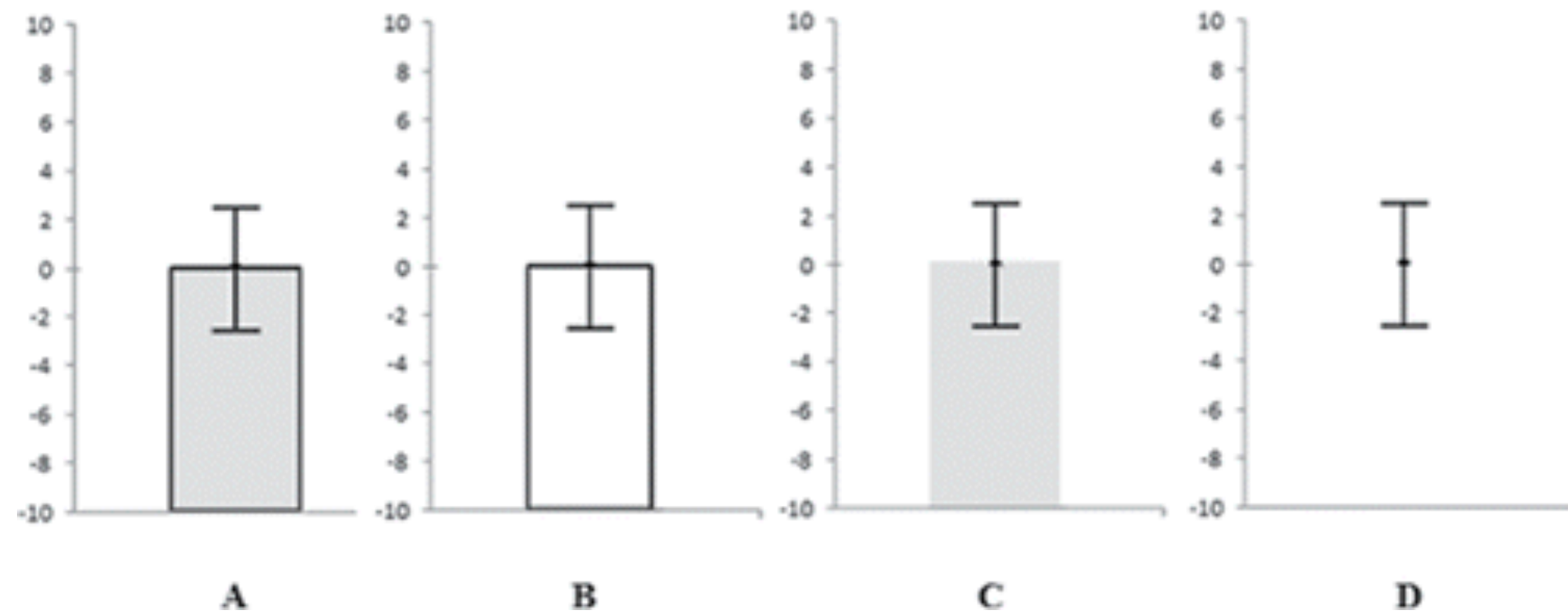
Biases

We can plot the data faithfully, but still perceive it wrongly!

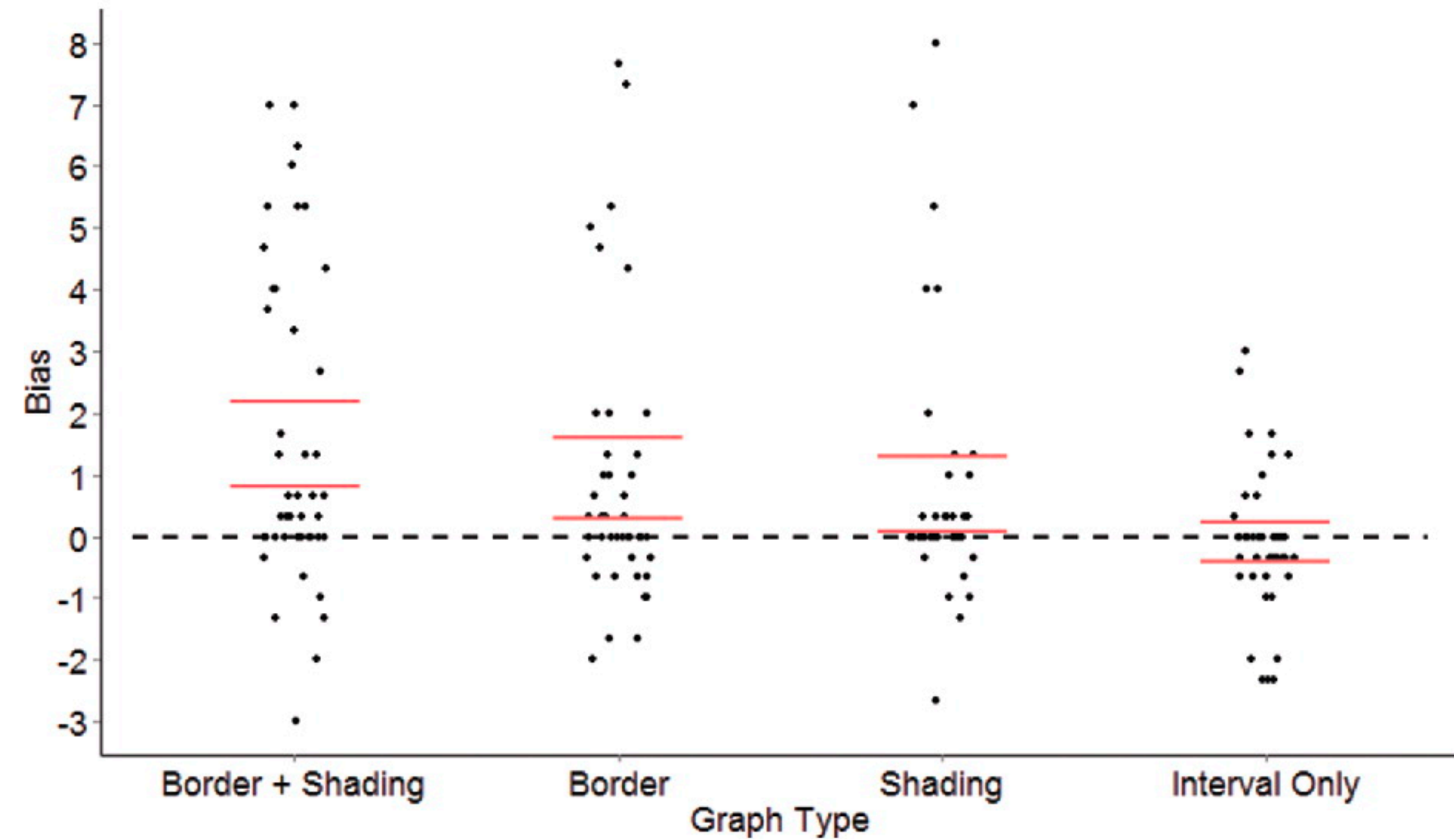
What about now?



Within the Bar Bias

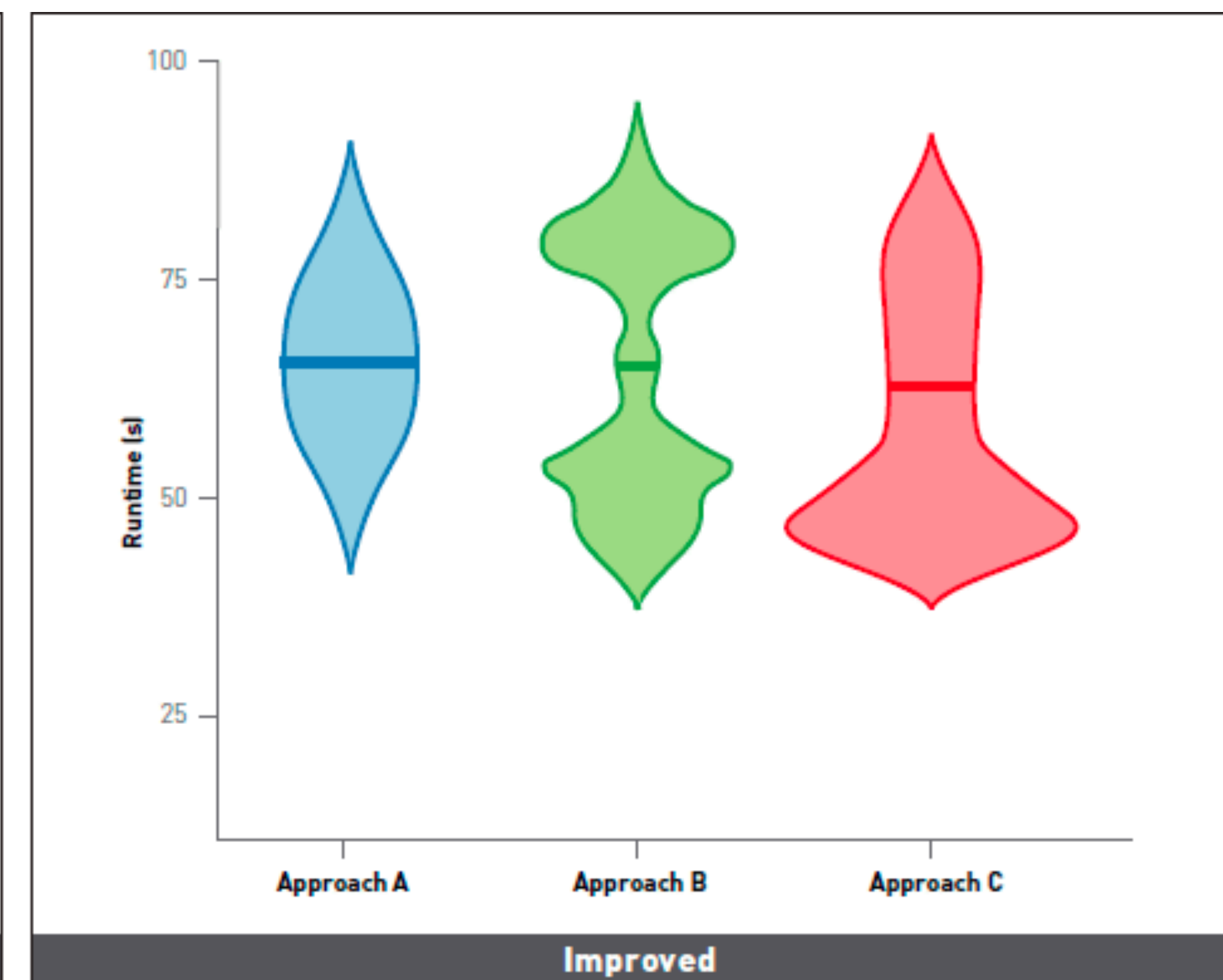
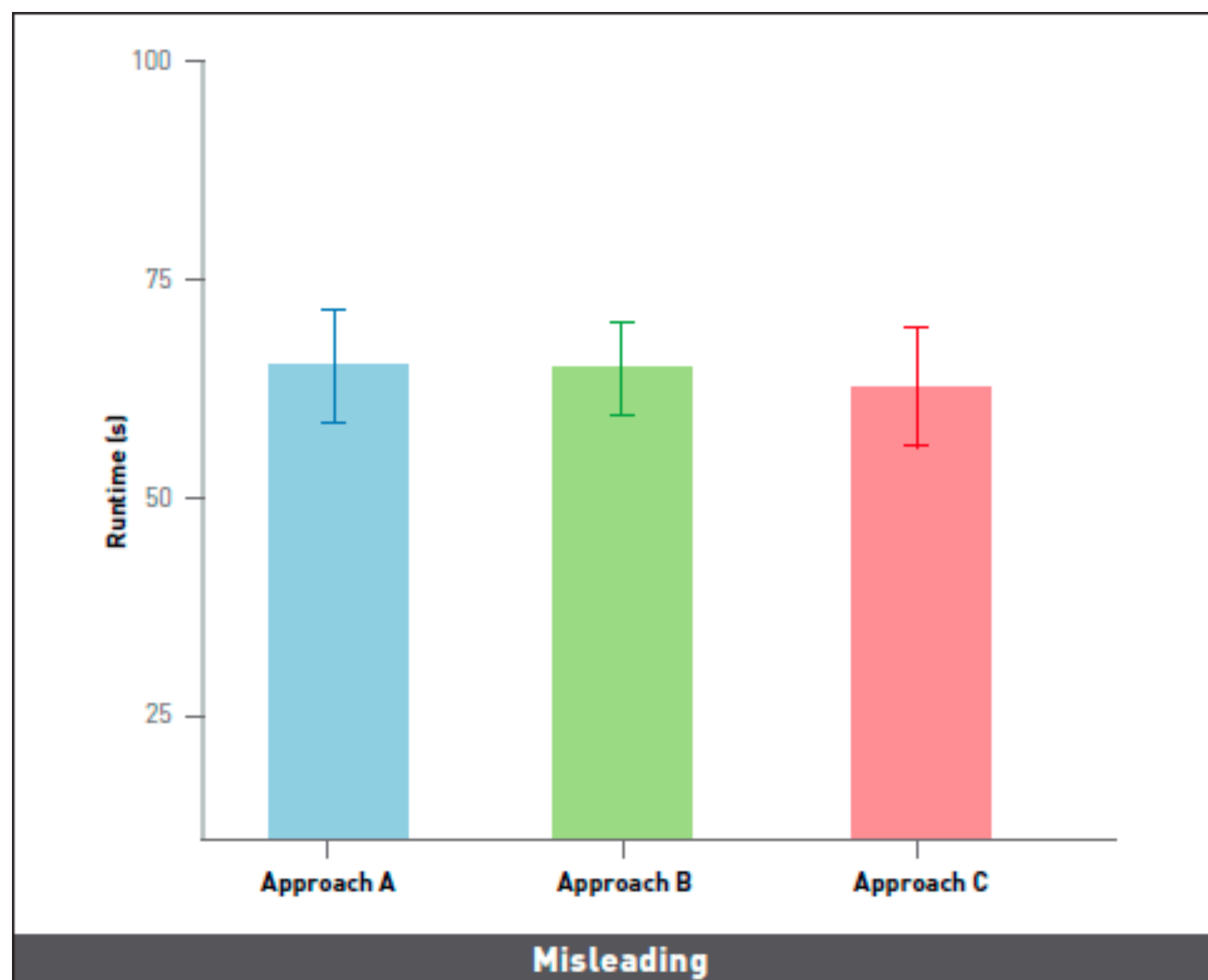


Experimental Conditions

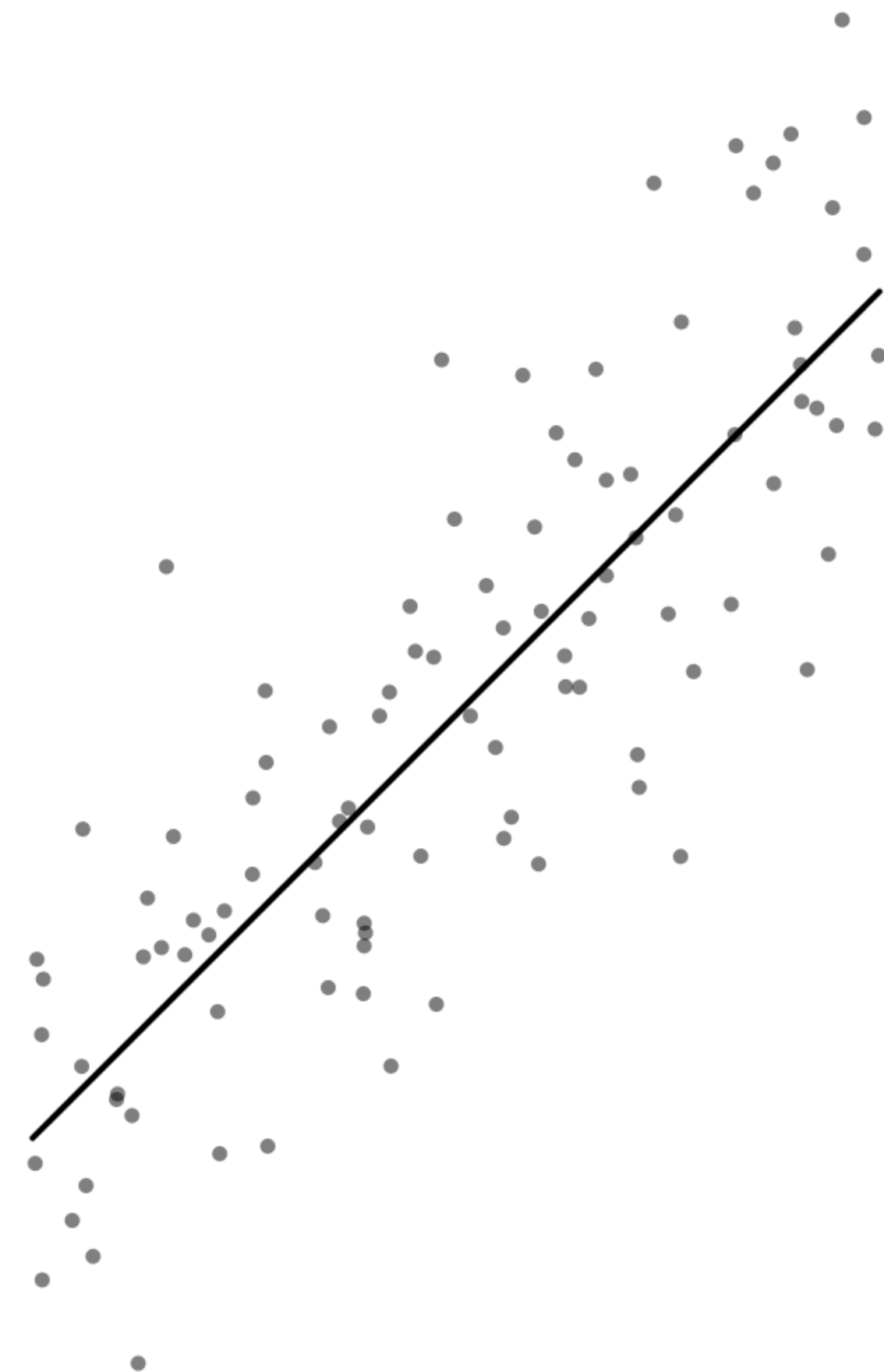
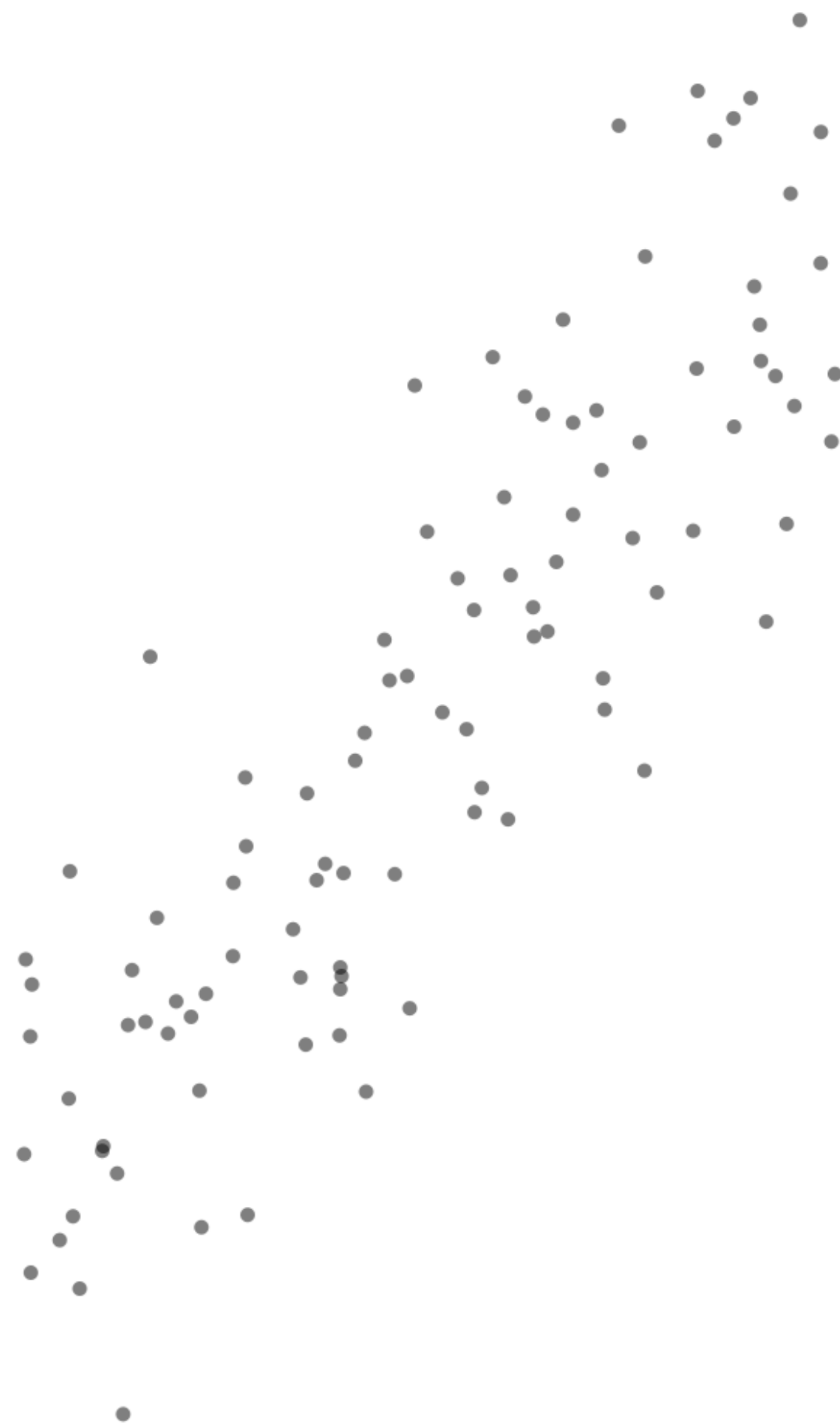


Results

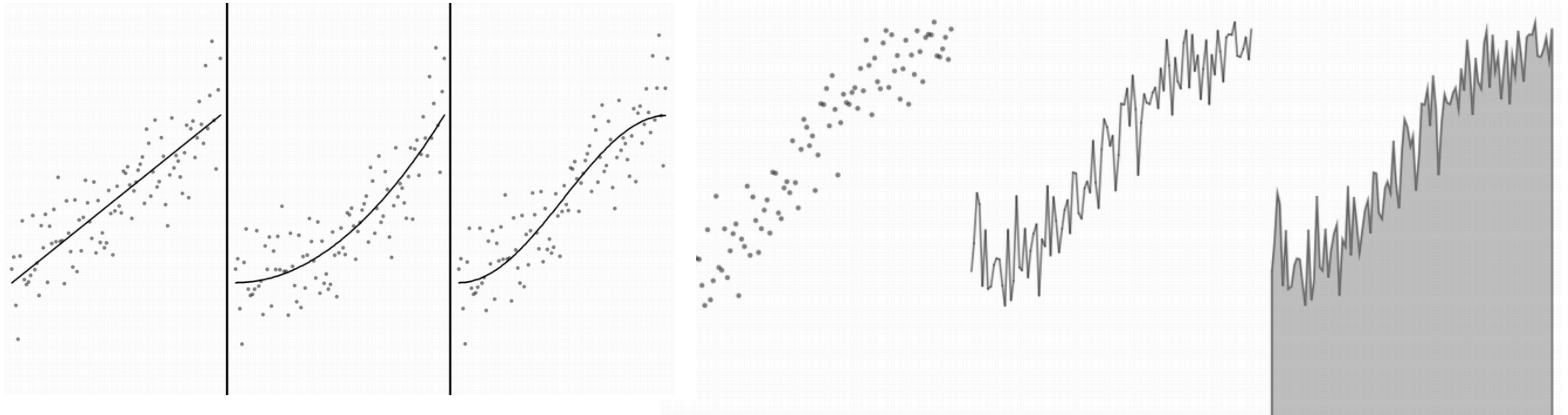
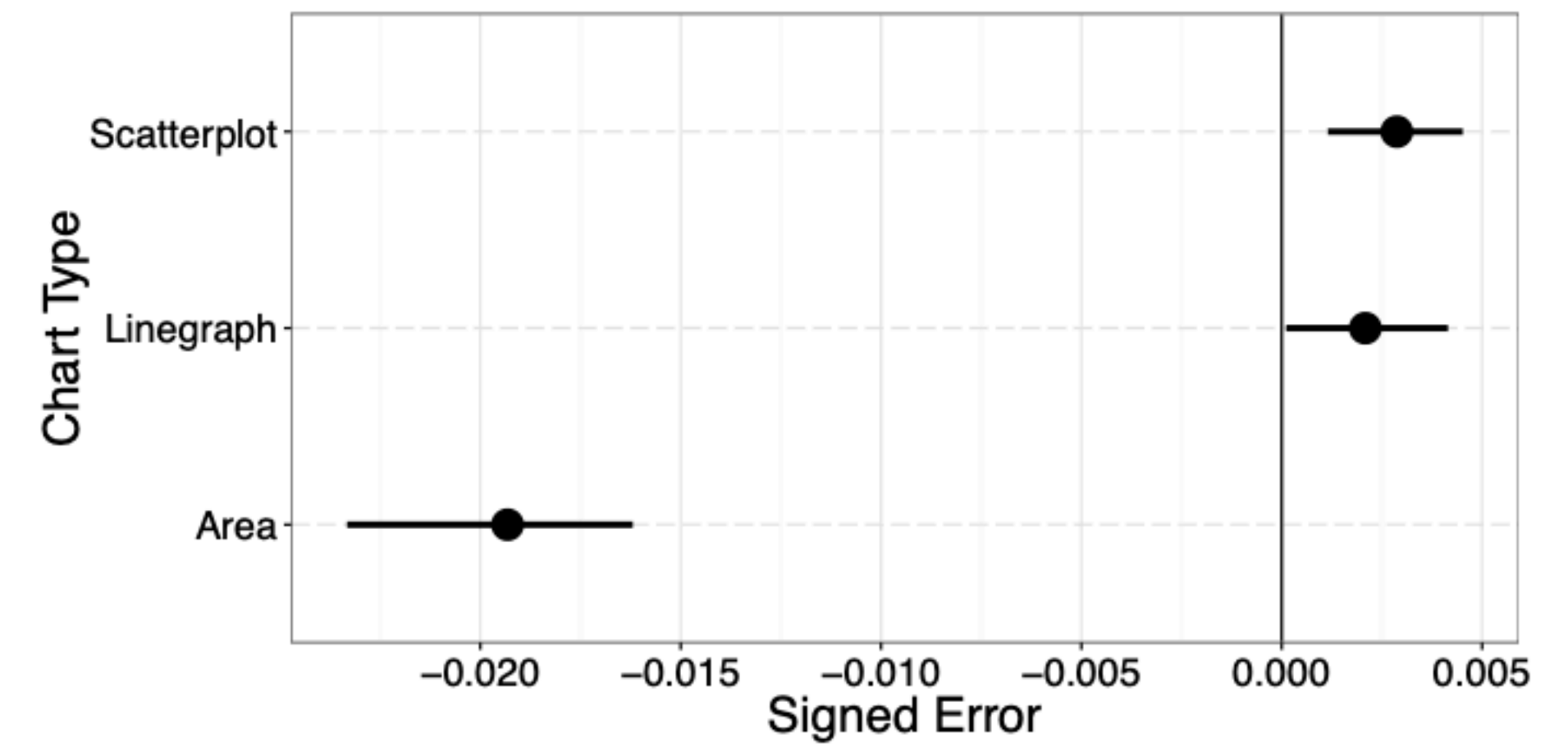
Careful when designing aggregated charts



What's the Trendline?



Regression by eye



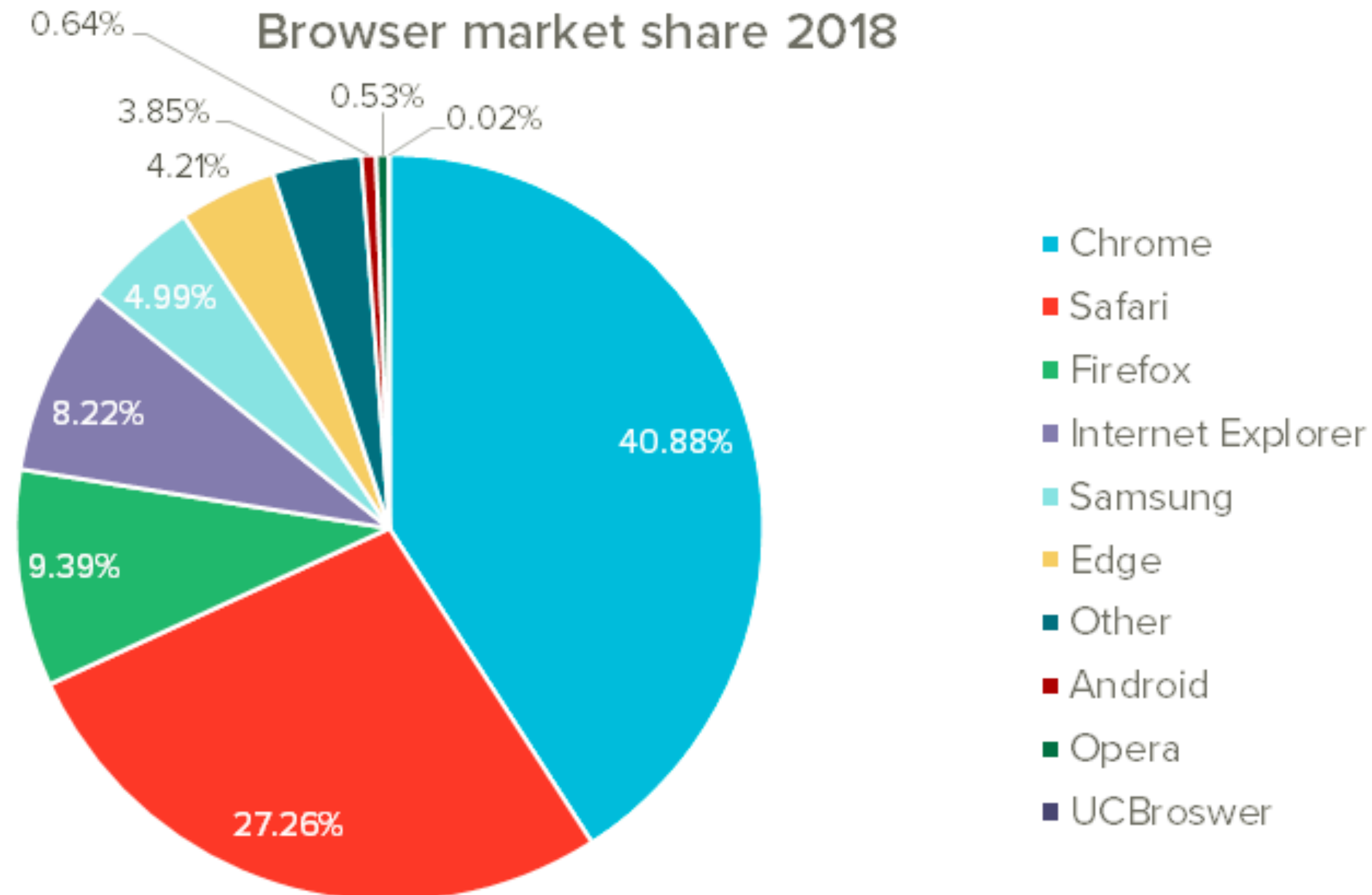
We're good at spotting trends

But the wrong vis technique can deceive us

Pie Charts

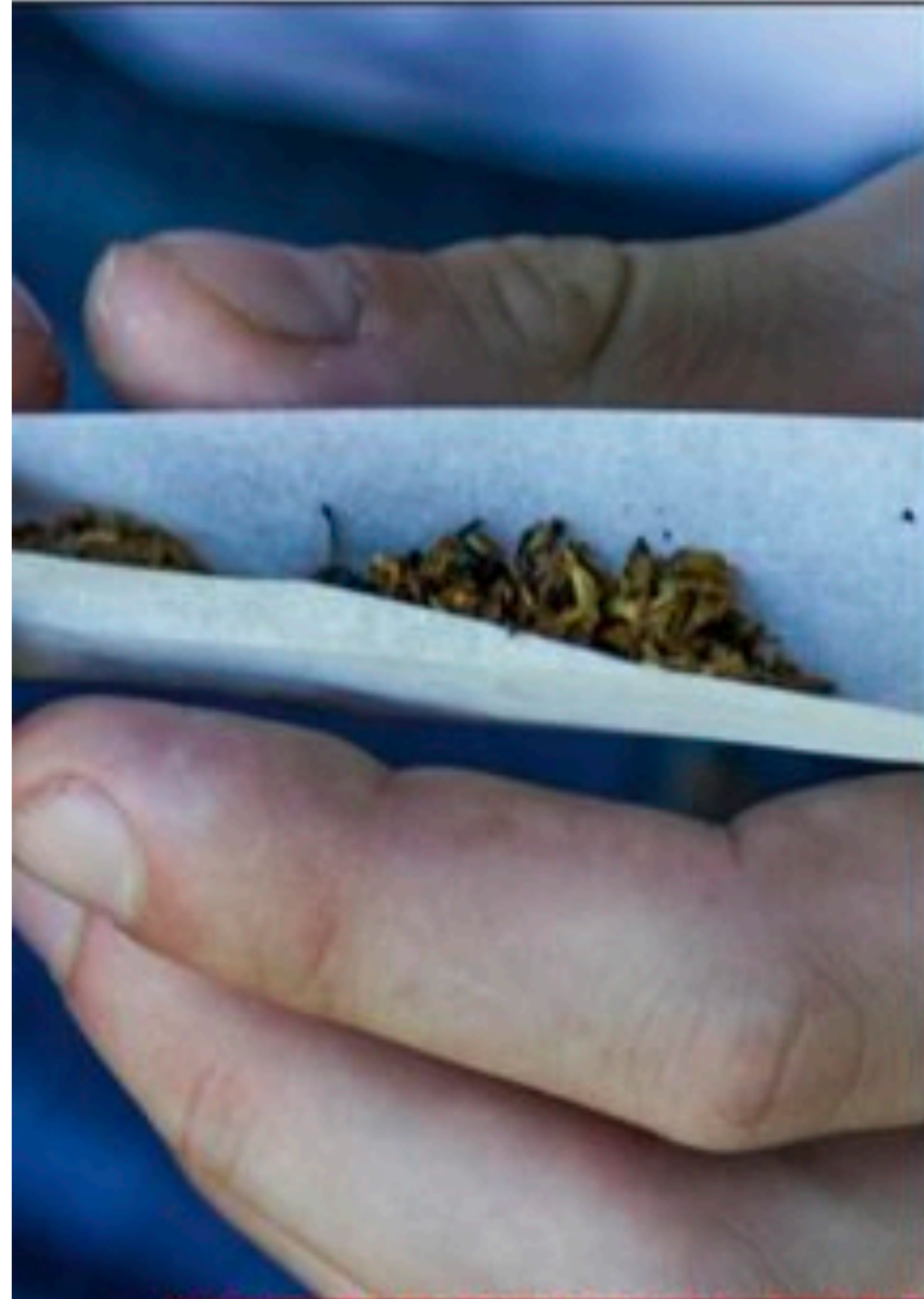
Why Pie Charts?

Show Part-of-Whole Relationships



How can we make this better?

- Label the wedges directly, get rid of color scale
- Fewer segments: put more into “other”
- Make sure labels have contrast



AMERICANS WHO HAVE TRIED MARIJUANA

CBS NEWS POLL

51% TODAY

43% LAST YEAR

34% 1997



Source: MOE +/- 4%

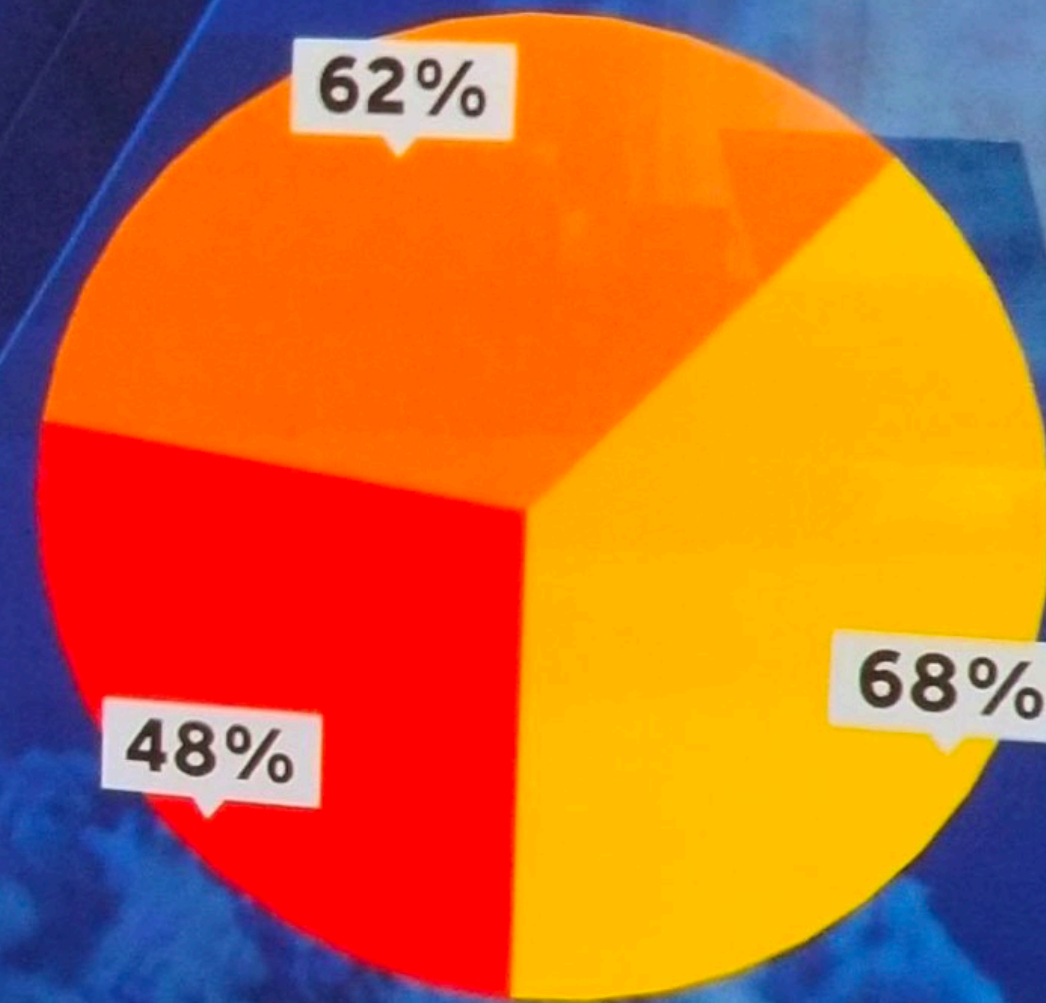
HIGH SUPPORT FOR LEGALIZING MARIJUANA
MORE THAN HALF OF AMERICANS SAY THEY'VE TRIED POT



LIVE
CBSN

BIGGEST COVID-19 WORRIES

- GETTING IT
- FAMILY GETTING IT
- THE ECONOMY



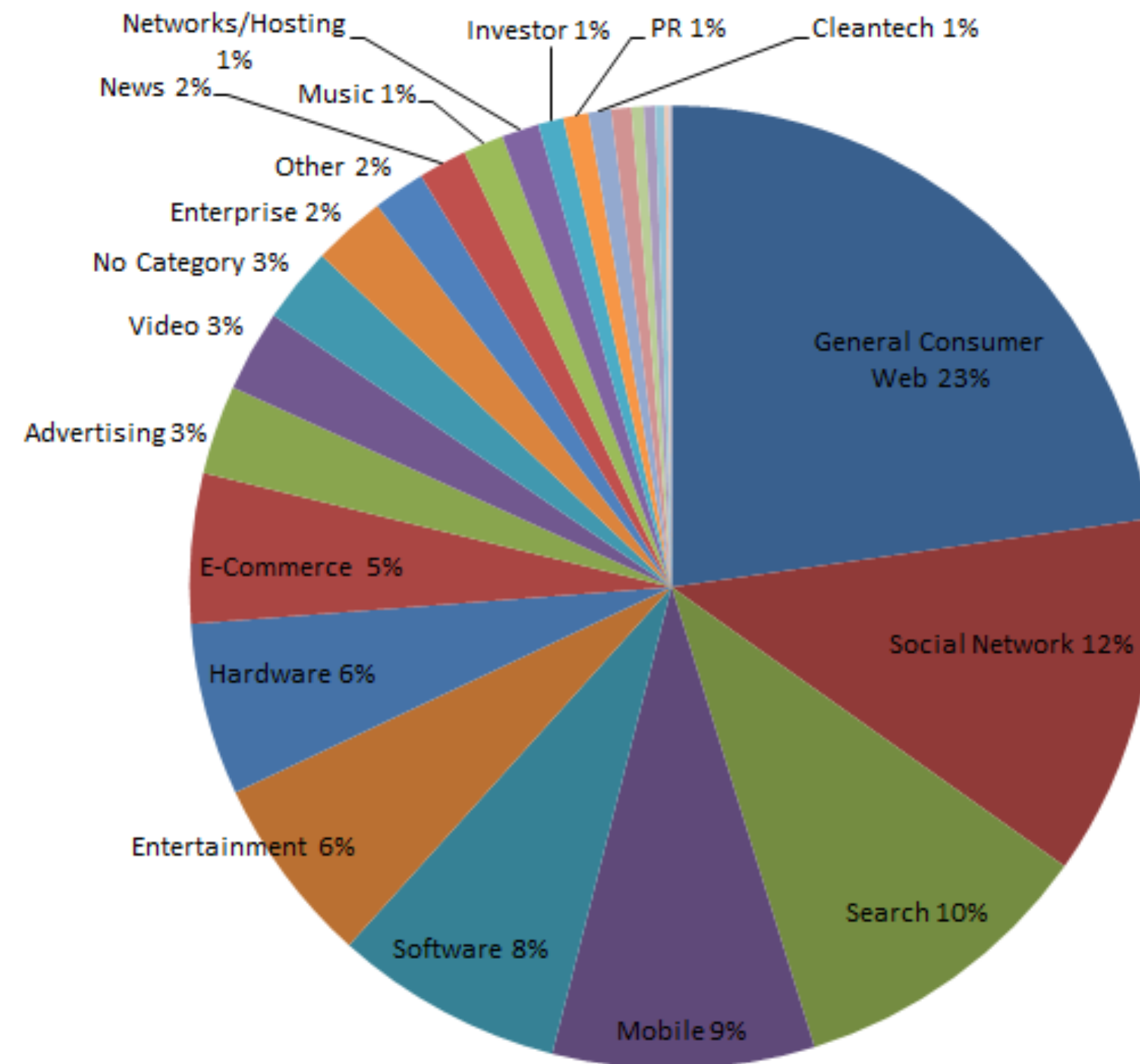
CORONAVIRUS
IMPACT

3 CASES
IN DUKES & NANTUCKET COUNTY



43°
5:49

Death to Pie Charts

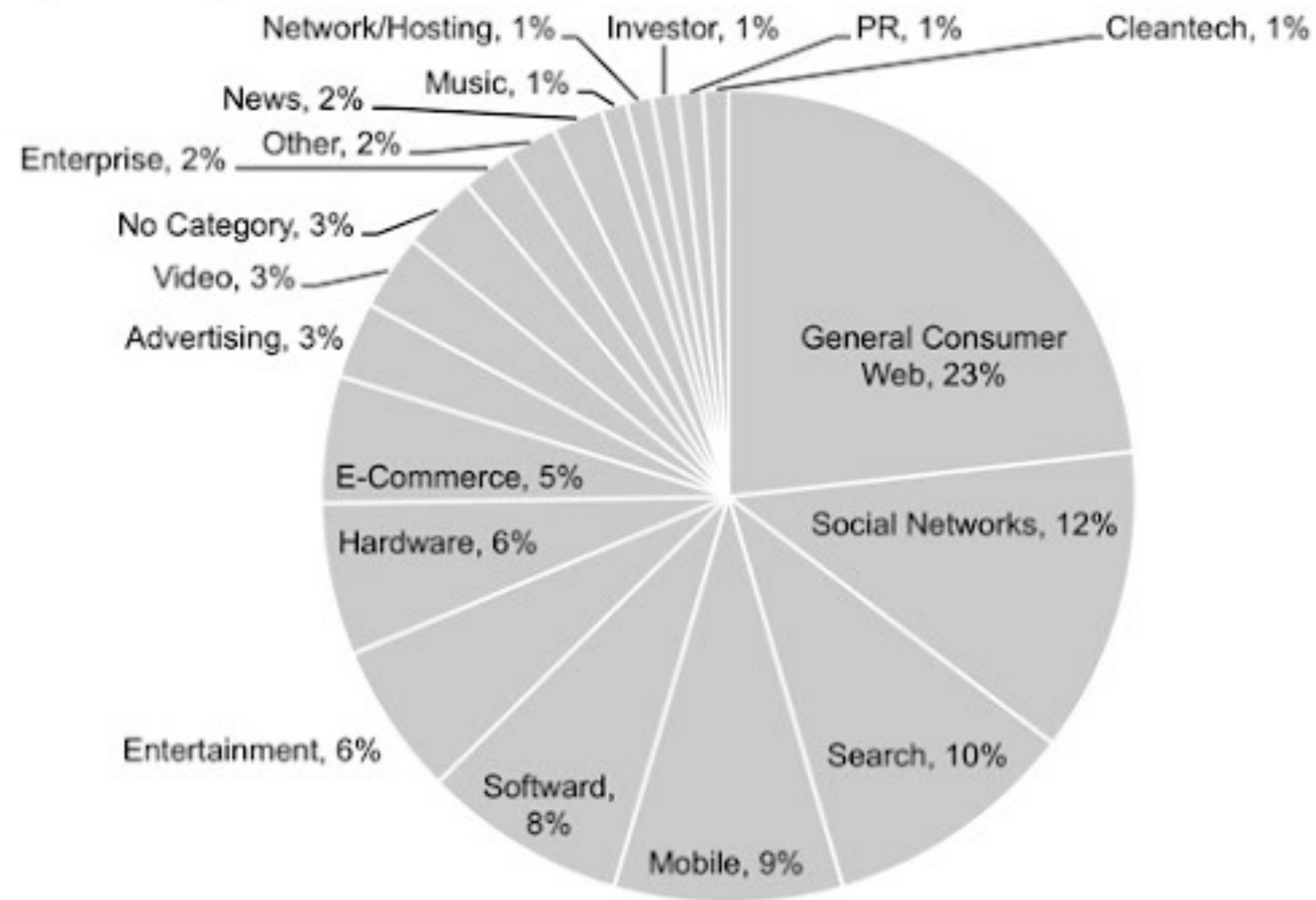


Share of coverage
on TechCrunch

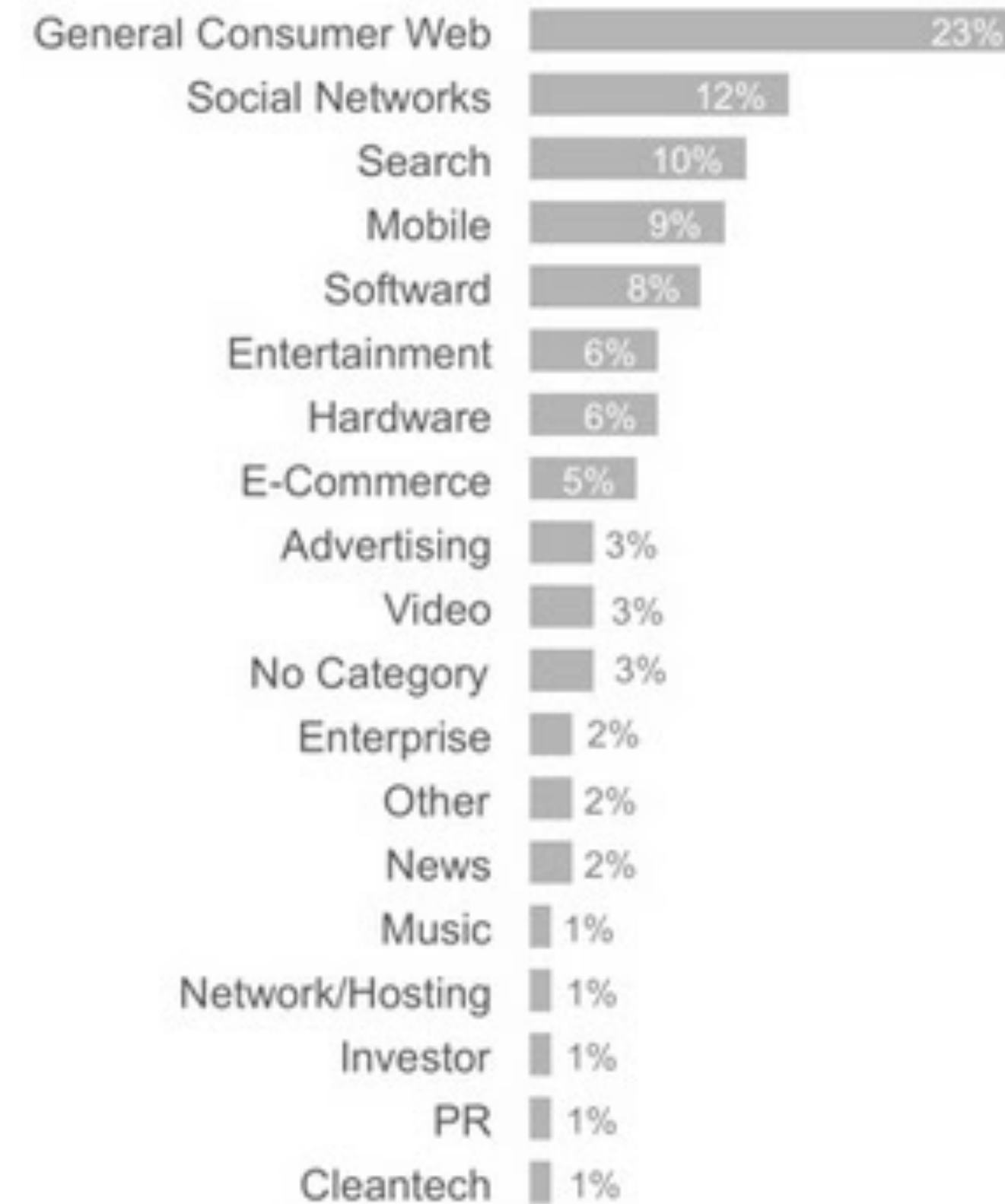
“I hate pie charts.
I mean, really hate them.”

Redesign

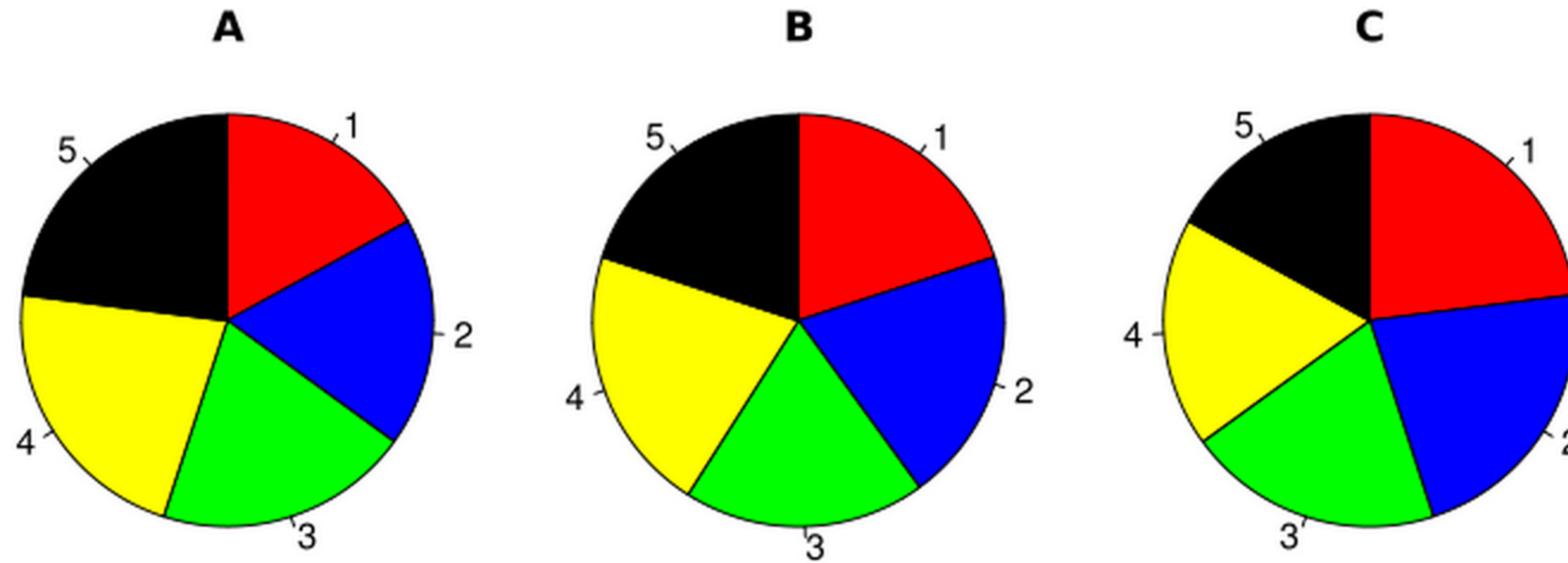
TechCrunch Coverage: 2005 - 2011
A slightly better pie?



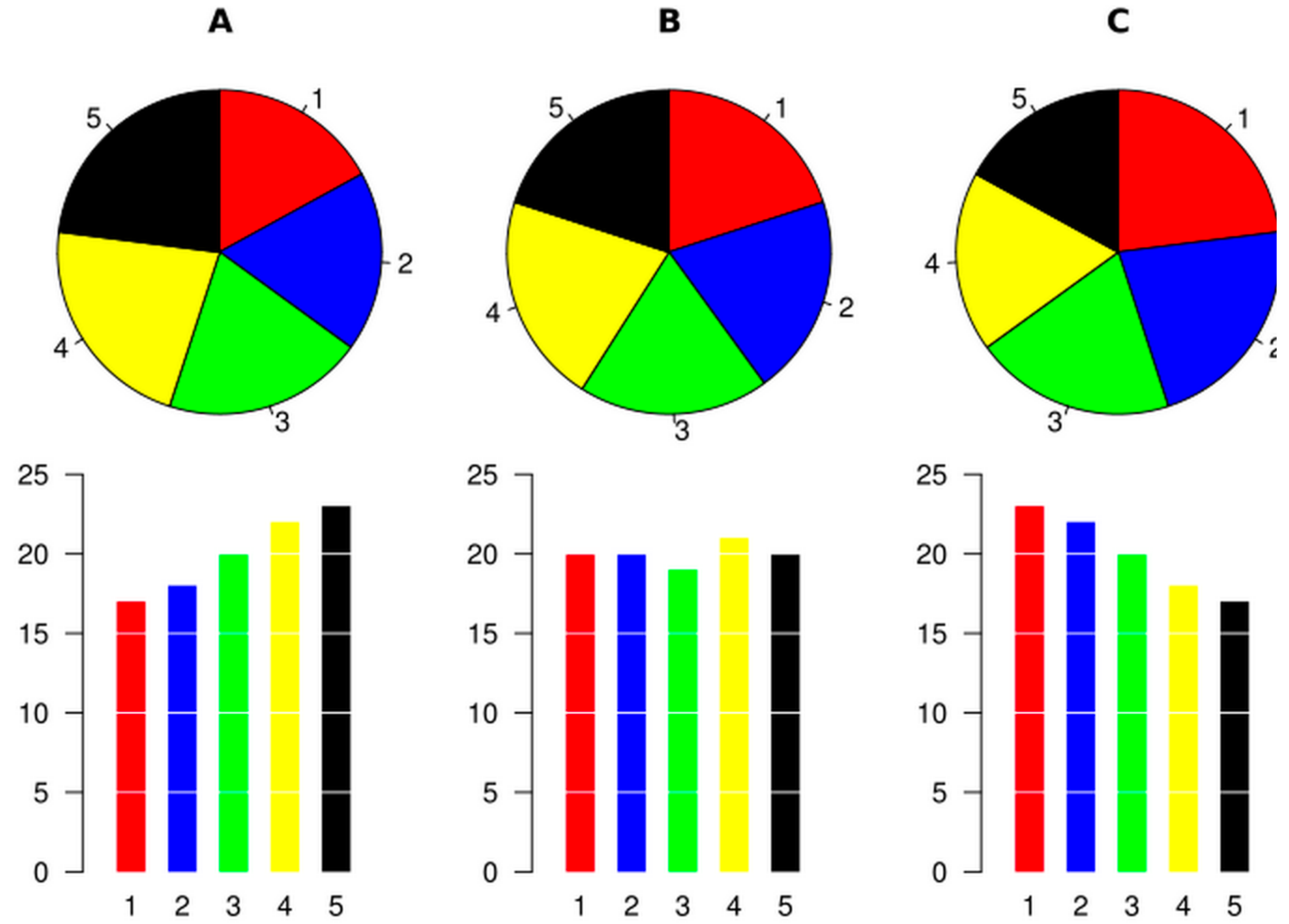
TechCrunch Coverage: 2005 - 2011
Bars are best!



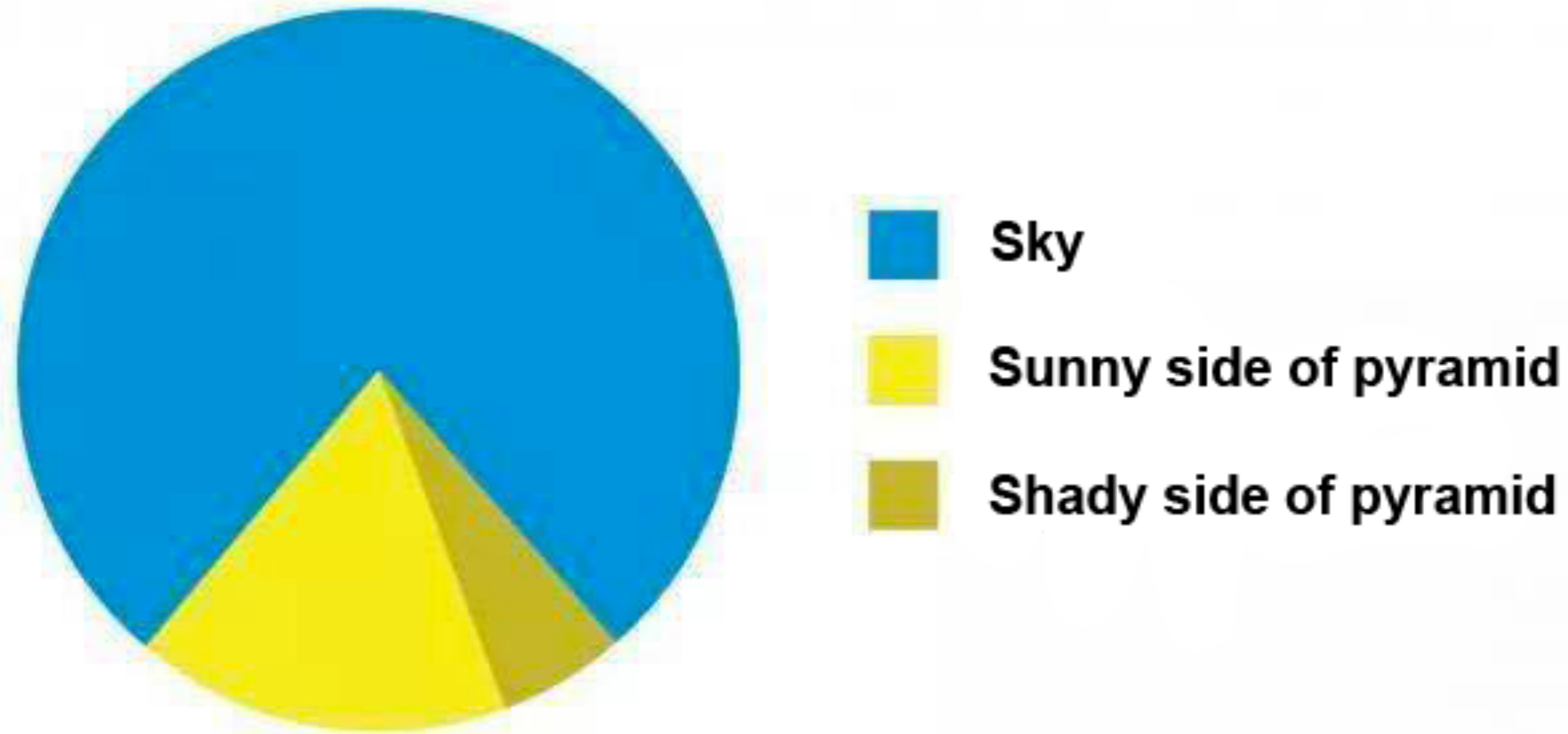
Can you spot the differences?



Can you spot the differences?



My favorite pie chart



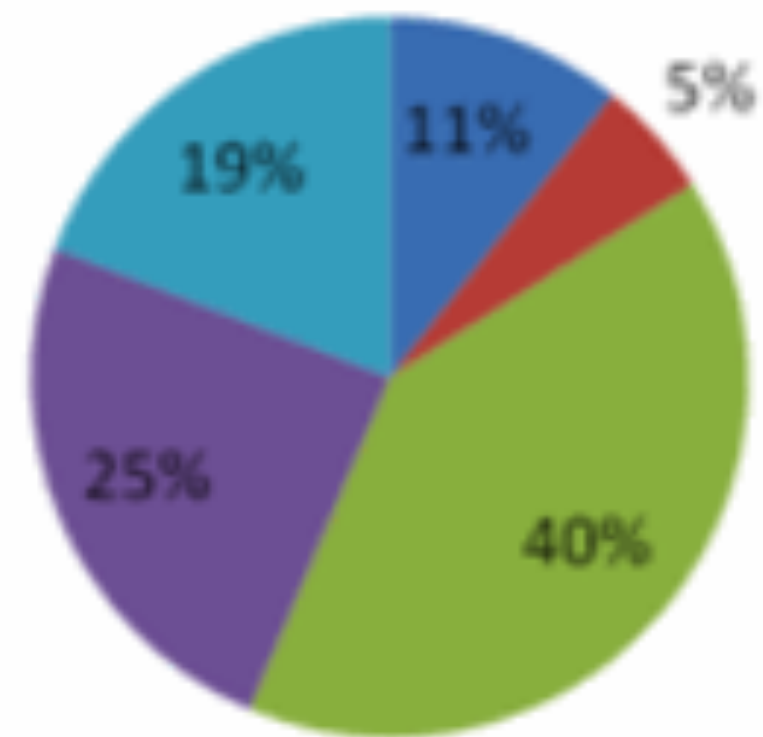
My second favorite pie chart



So, what to use instead?

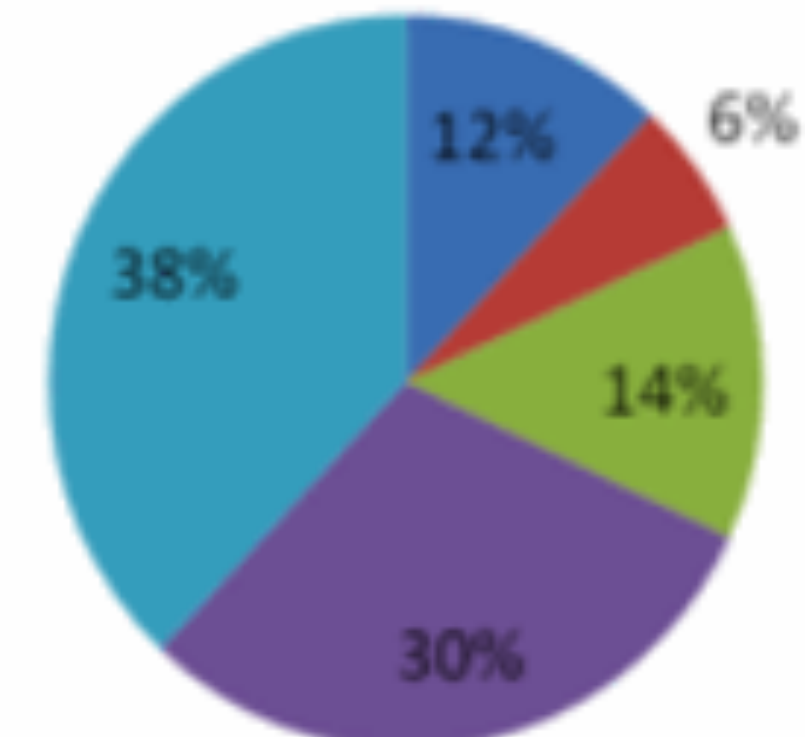
PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



imagine you just completed a pilot summer learning program on science aimed at improving perceptions of the field among 2nd and 3rd grade elementary children

Alternative #1: Show the Number(s) Directly

After the pilot program,

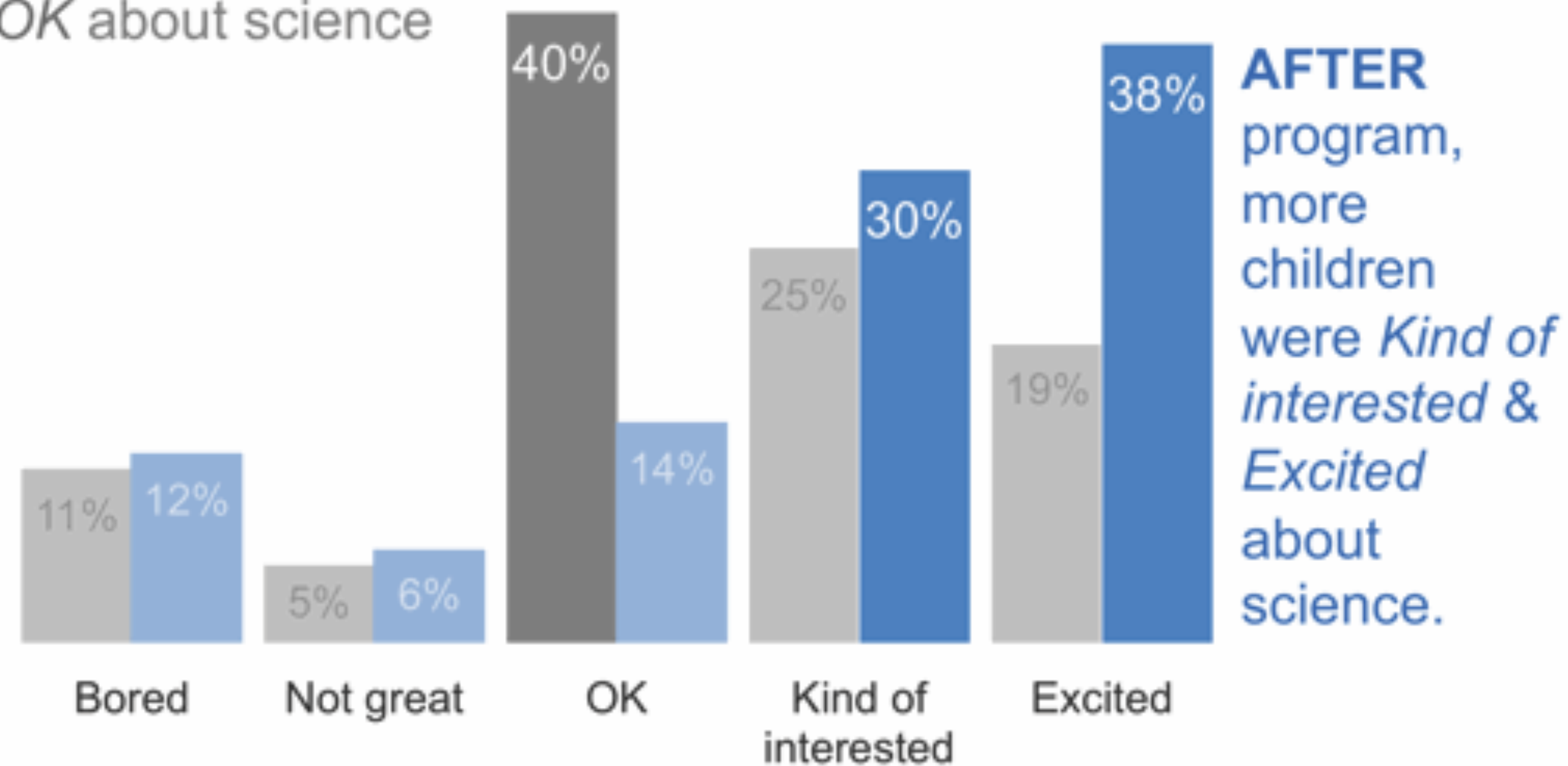
68%

of kids expressed interest towards science,
compared to 44% going into the program.

Alternative #2: Simple Bar Graph

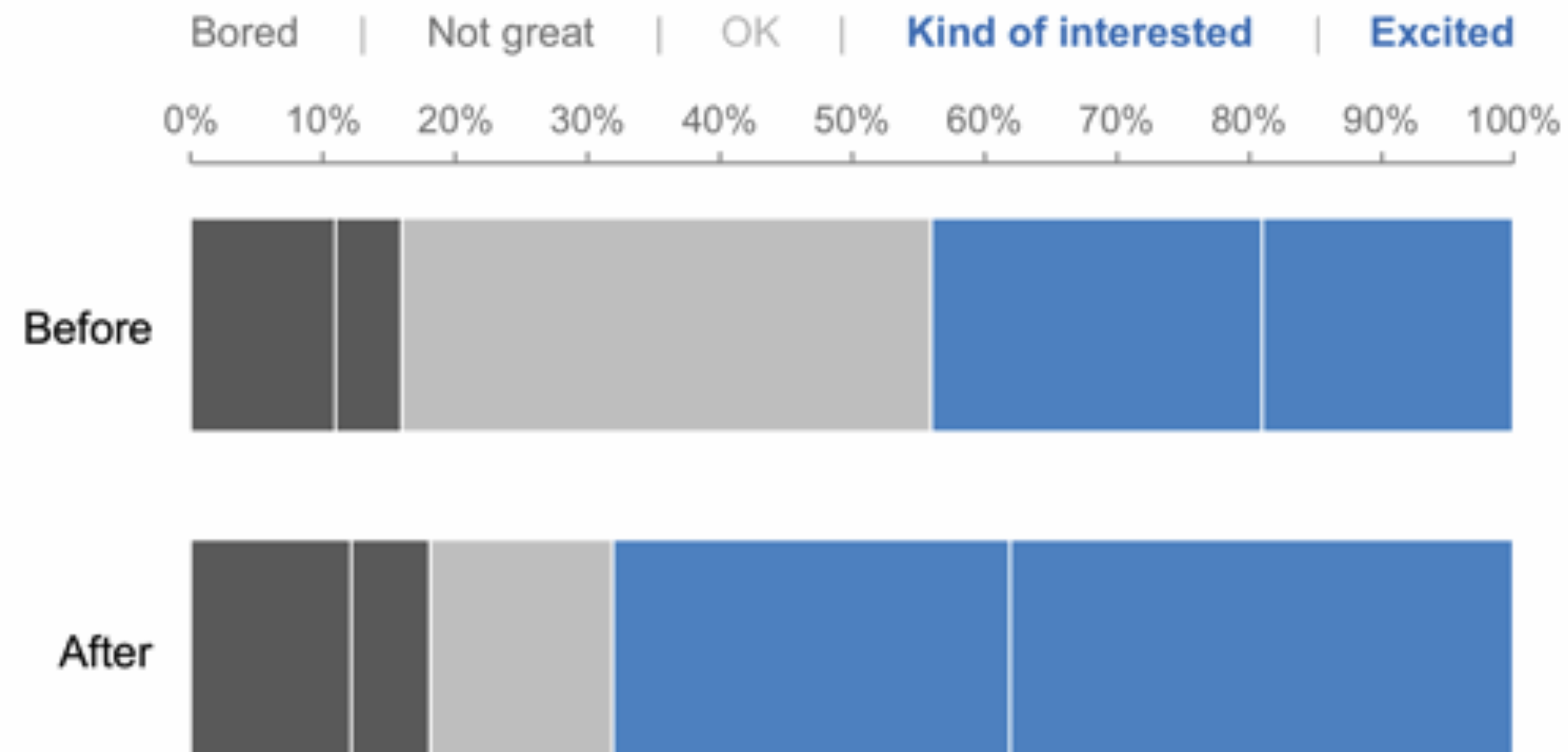
How do you feel about science?

BEFORE program, the majority of children felt just *OK* about science



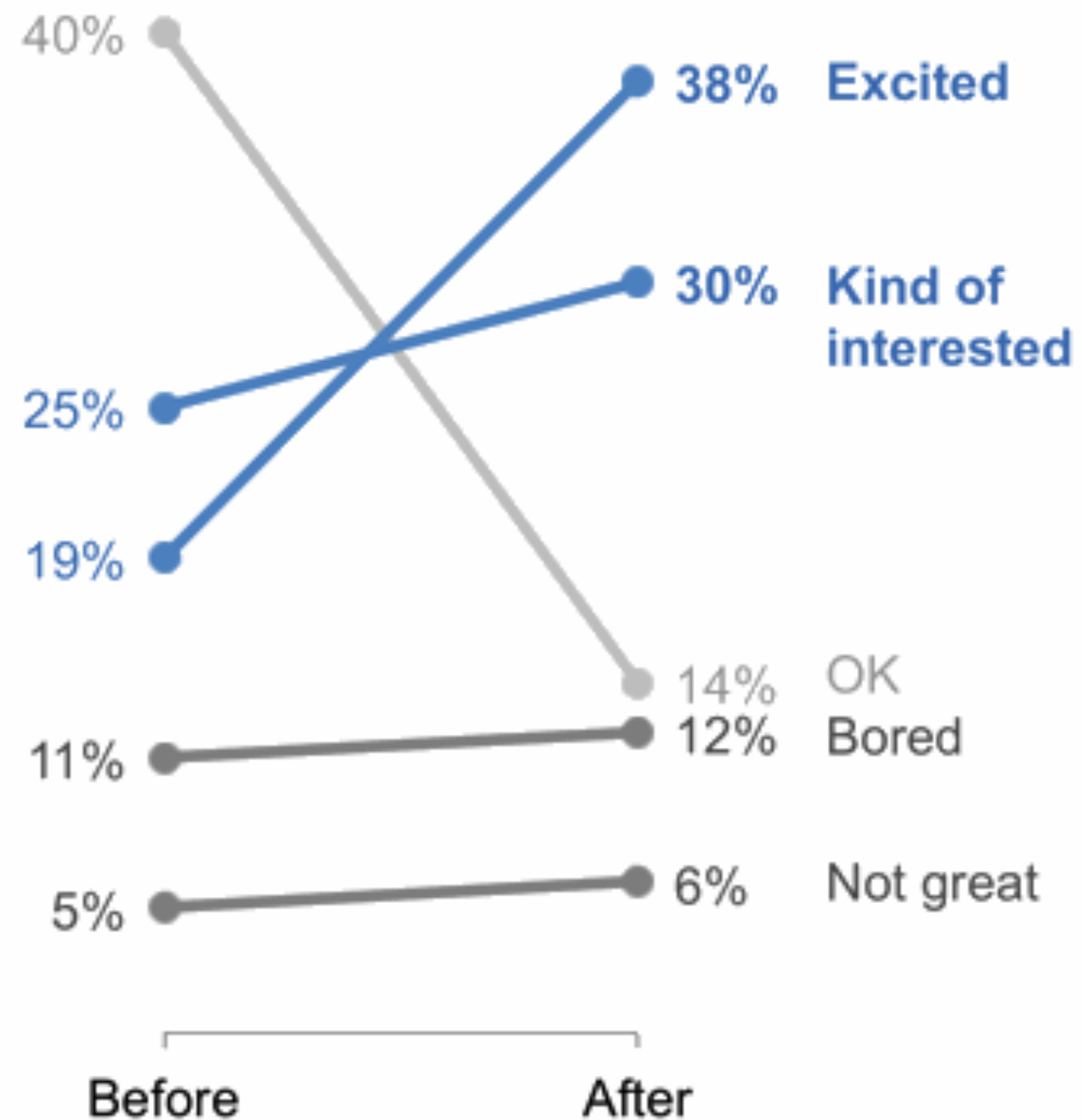
Alternative #3: 100% Stacked Horizontal Bar Graph

How do you feel about science?



Alternative #4: Slopegraph

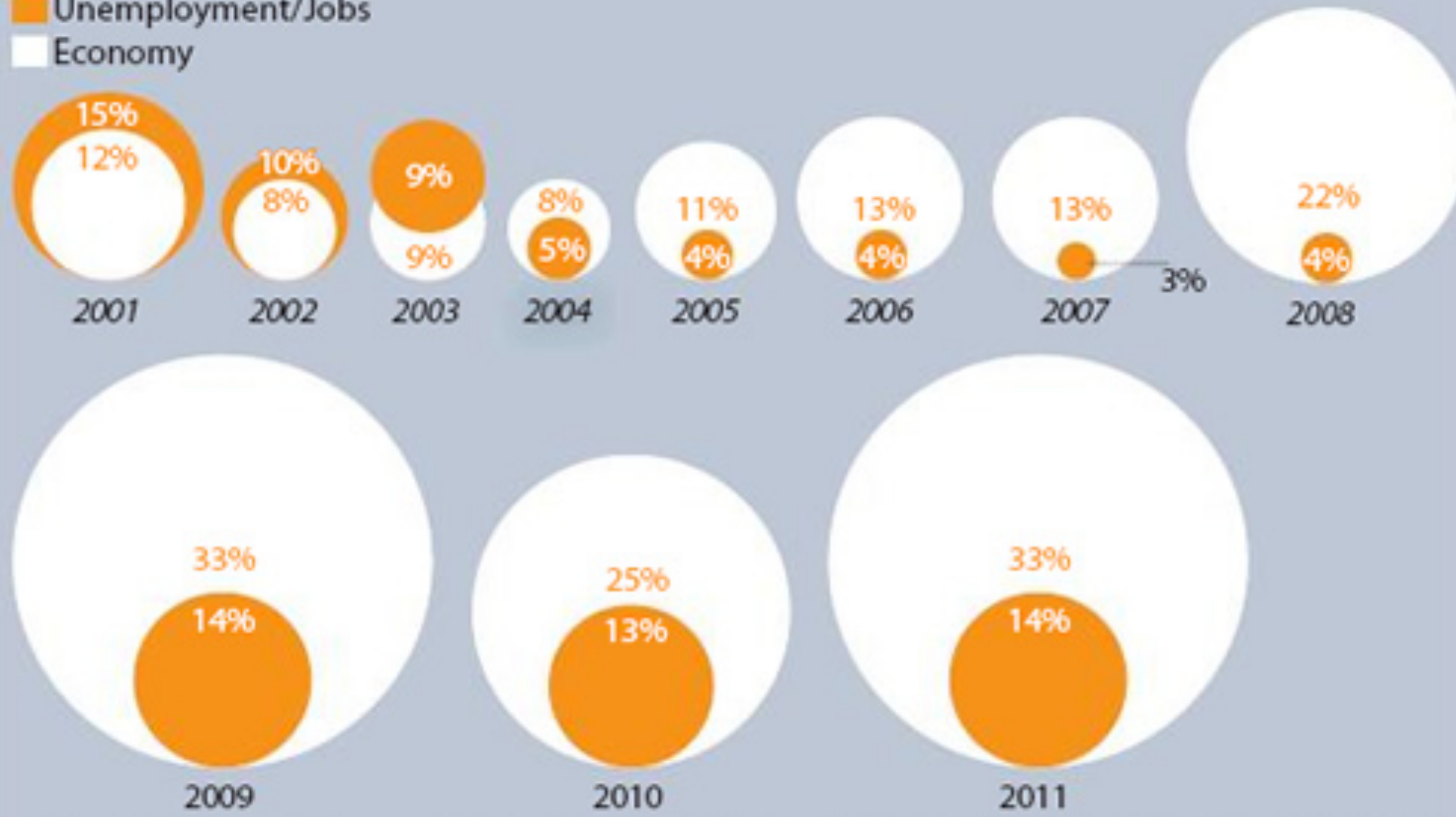
How do you feel about science?



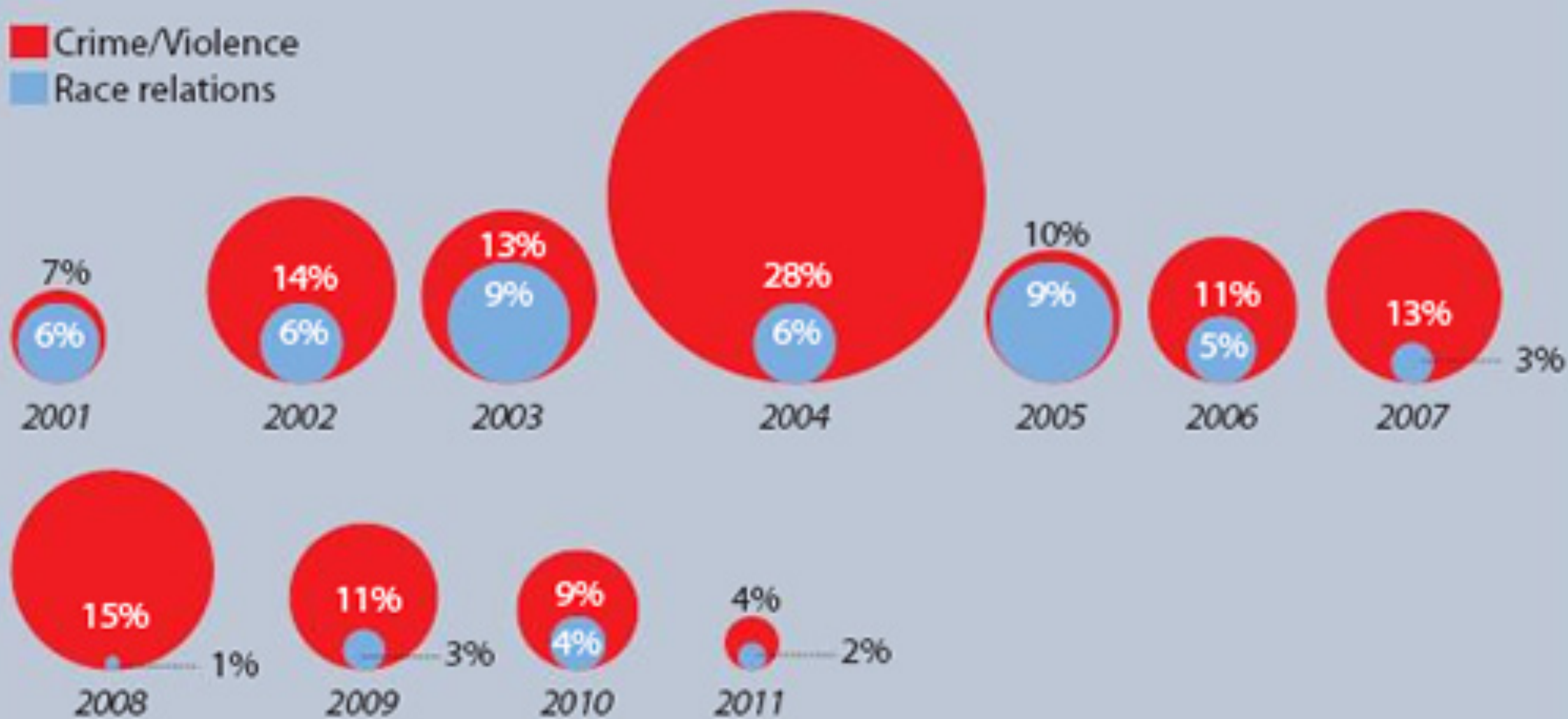
Most important issues

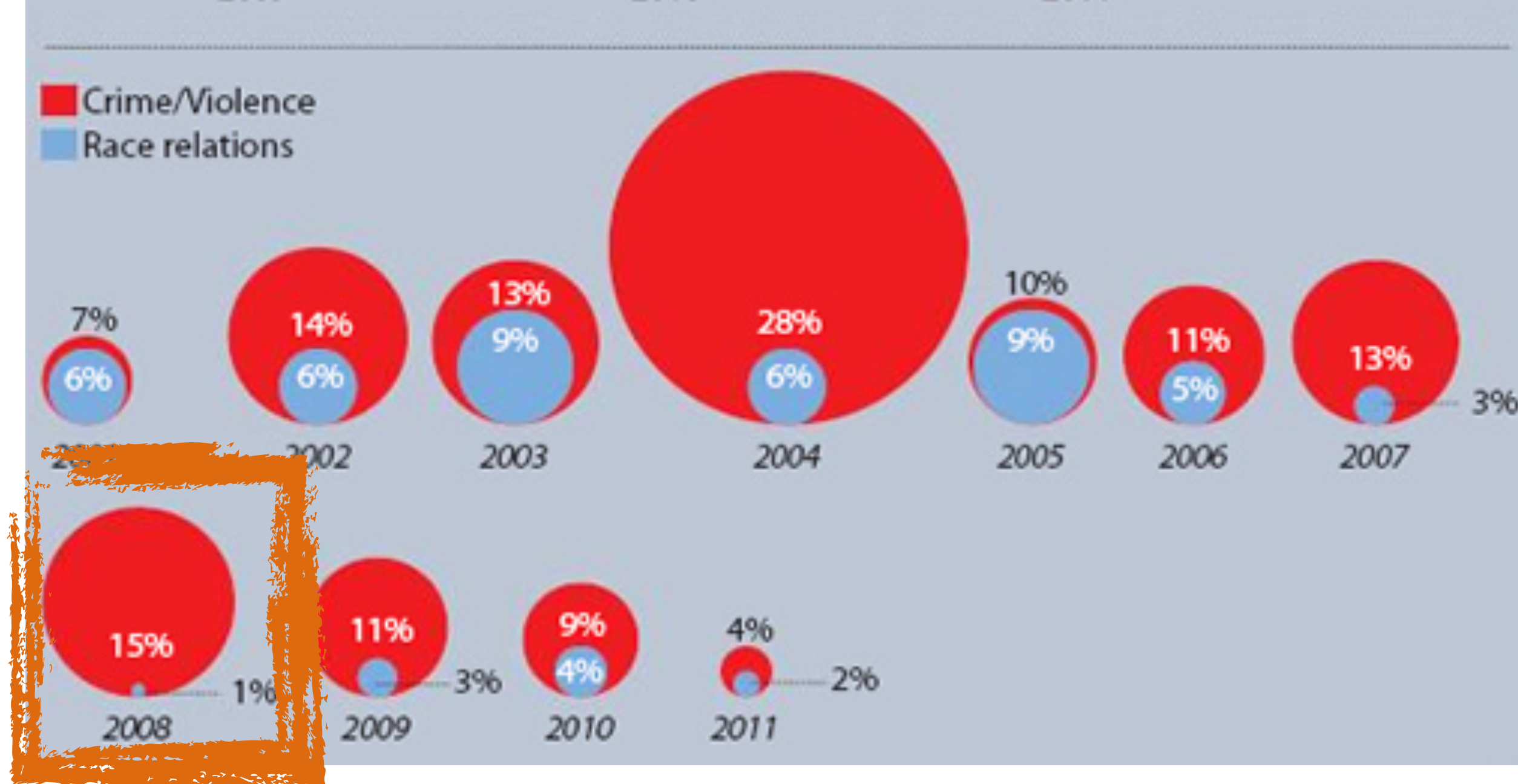
What do you think is the most important problem facing New Zealand today?

■ Unemployment/Jobs
■ Economy



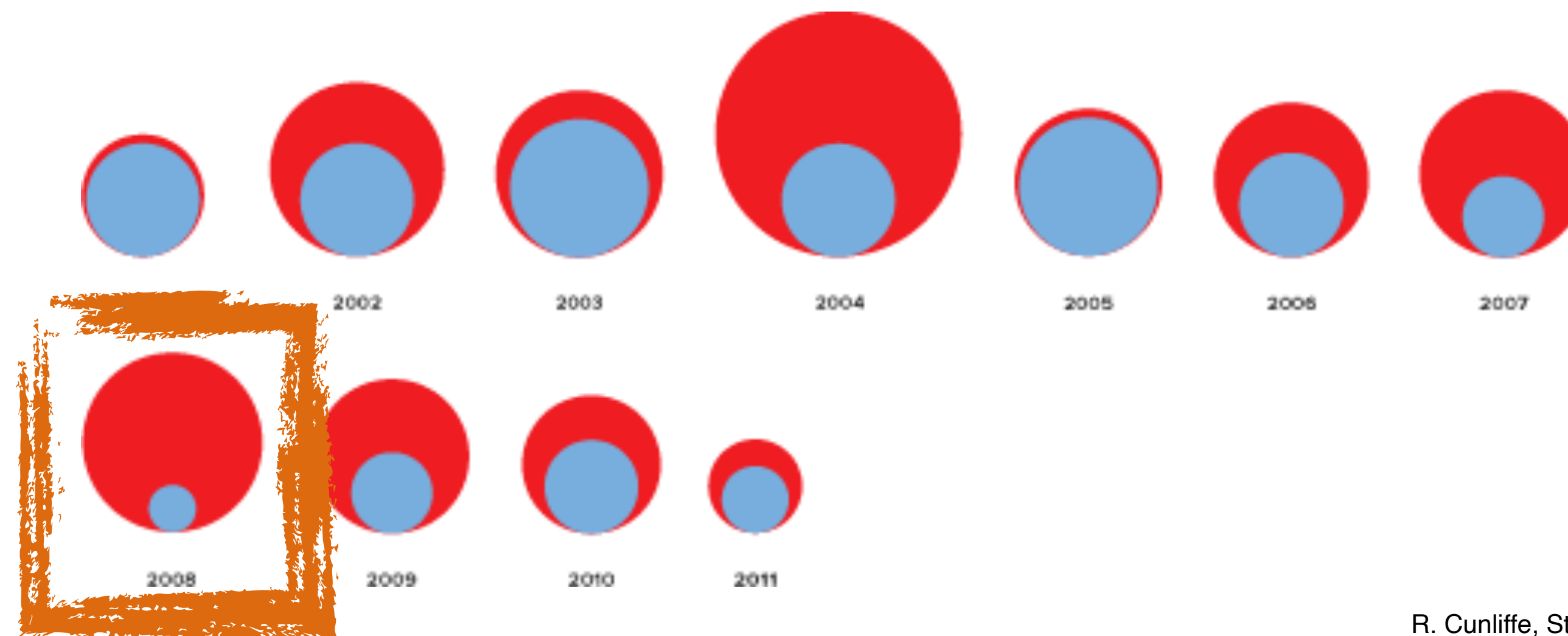
■ Crime/Violence
■ Race relations

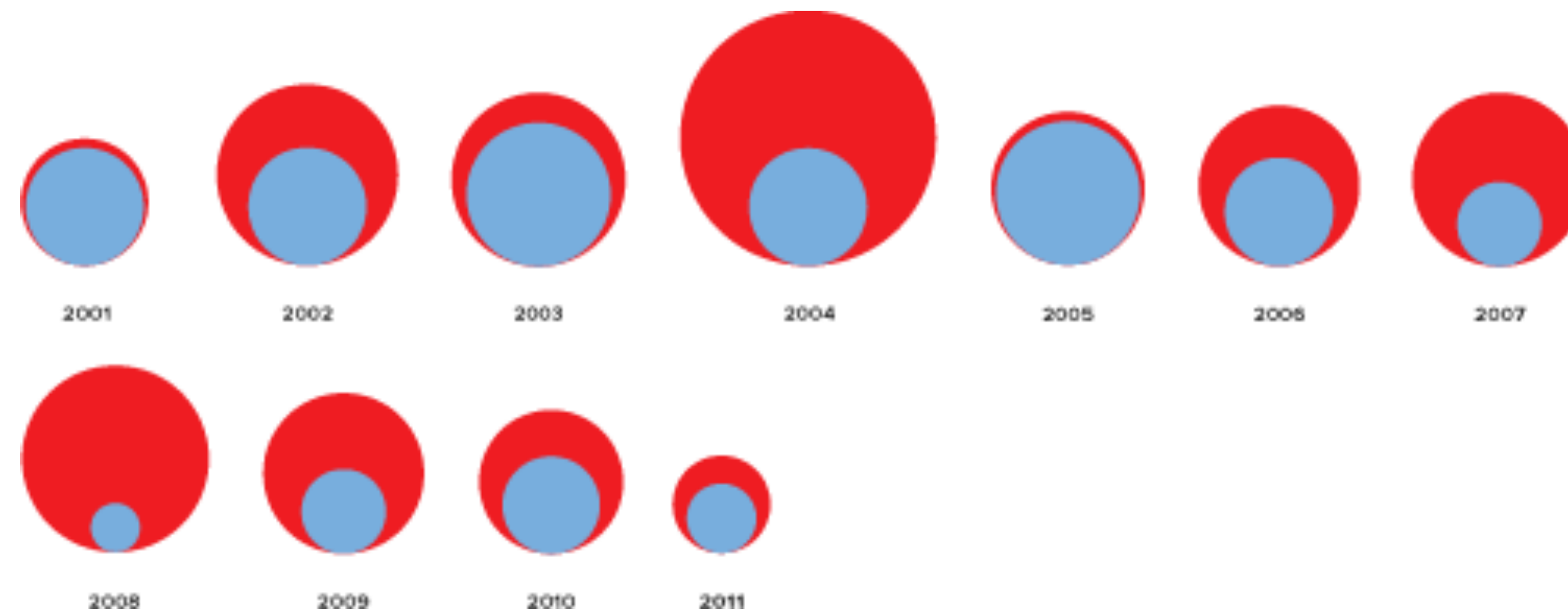




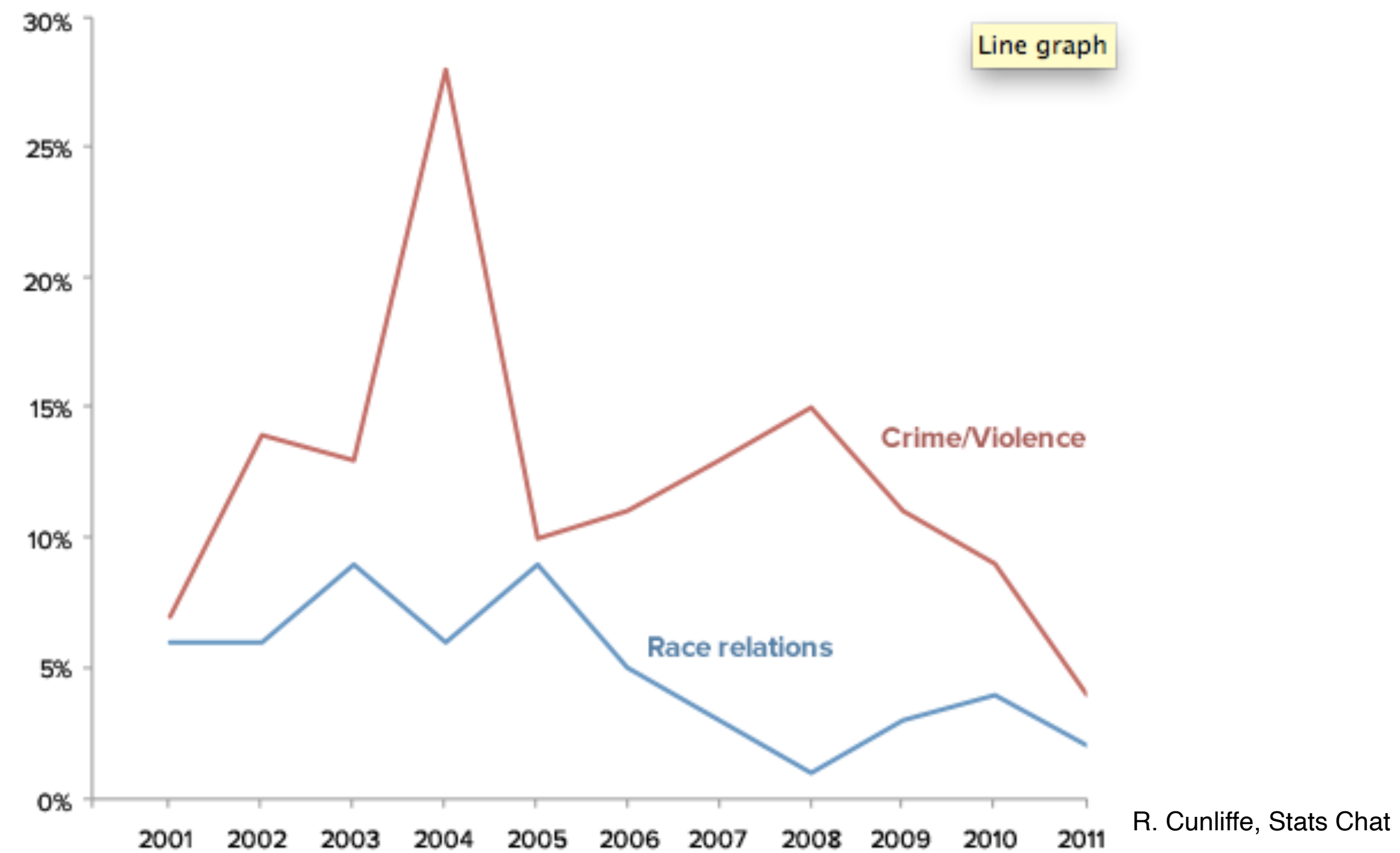
Quantity encoded by diameter, not area!

Fixing that:



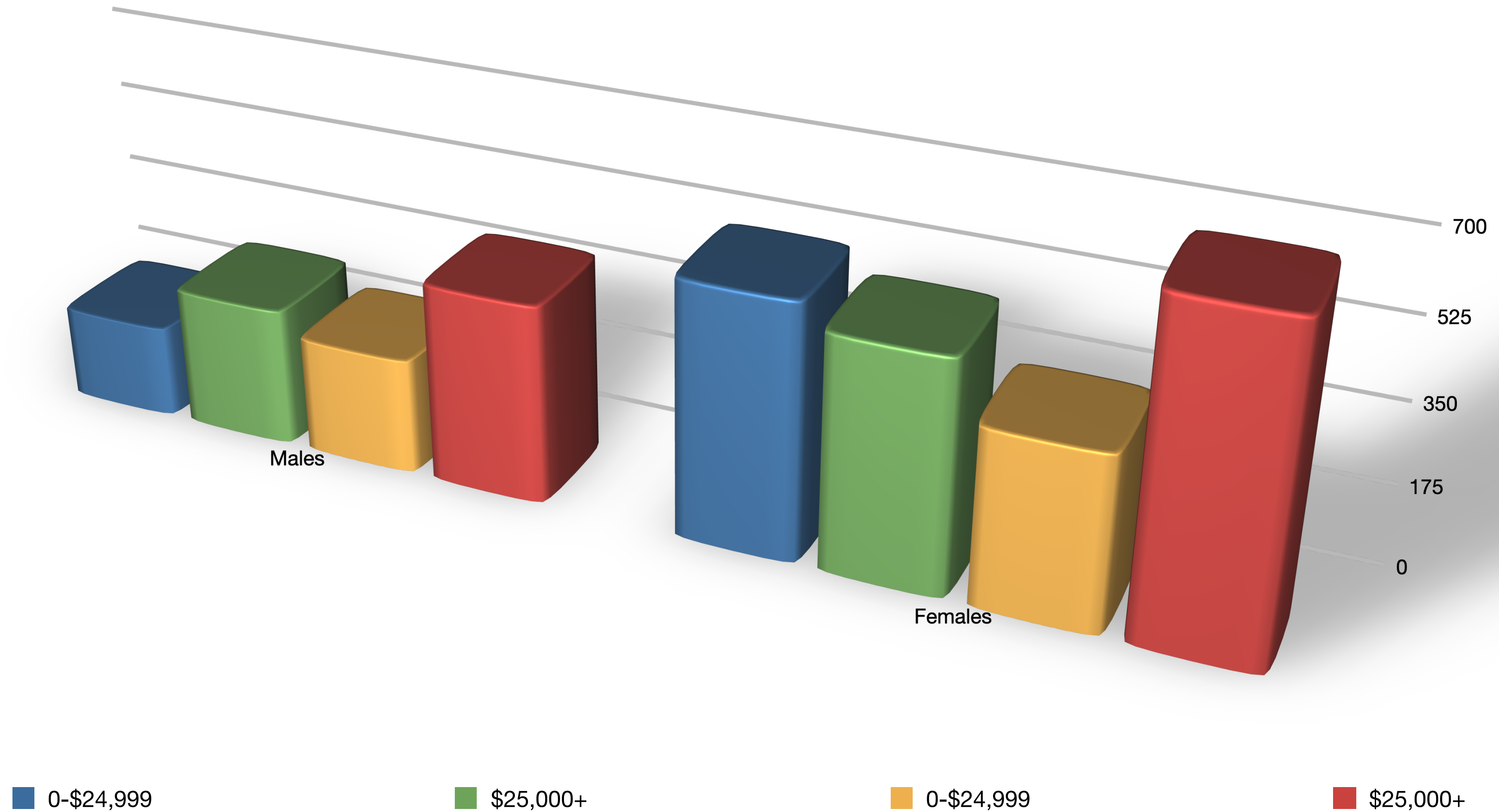


But is this visual encoding appropriate in the first place?

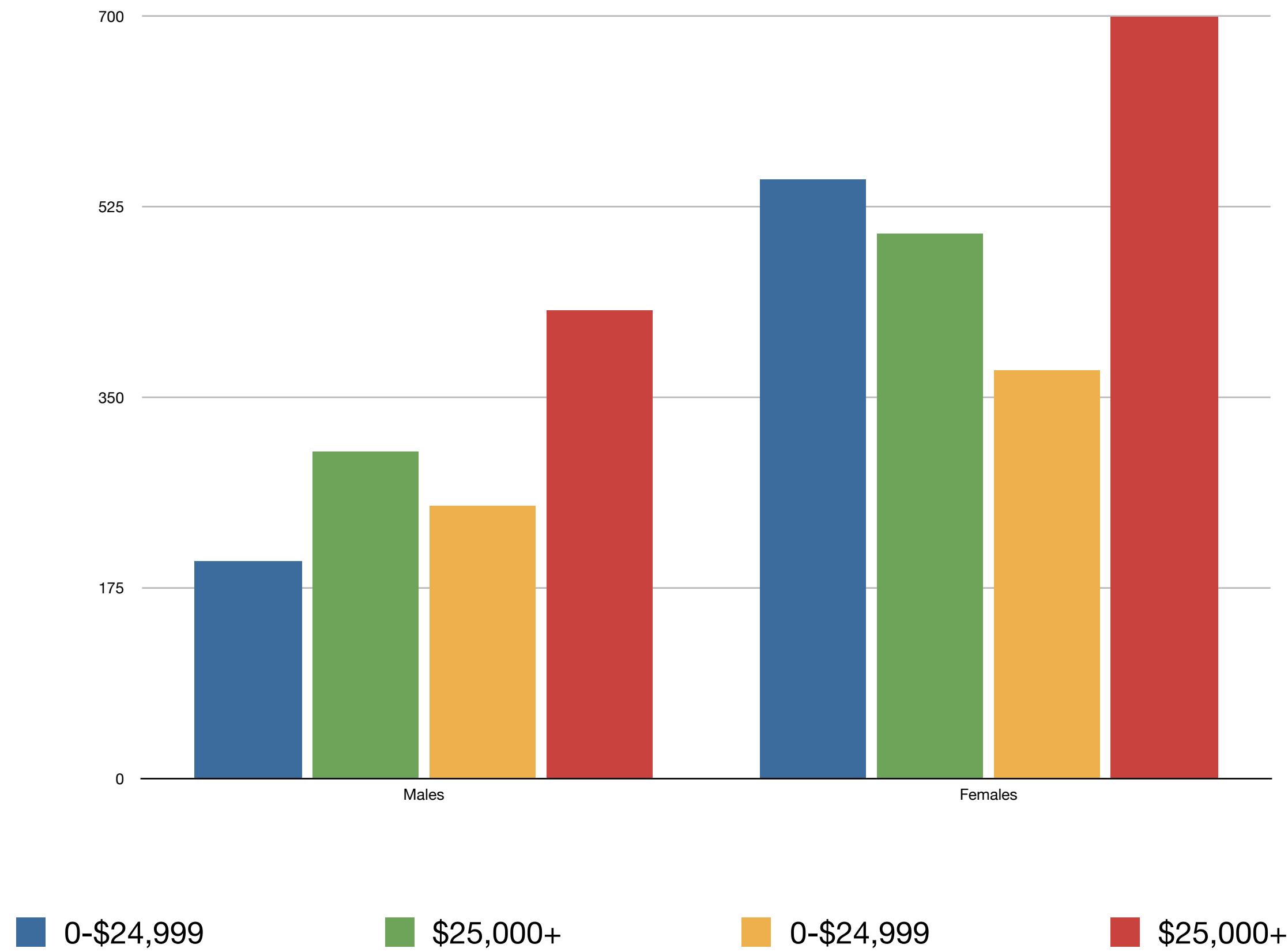


Clean vs Embellished

Maximize Data-Ink Ratio

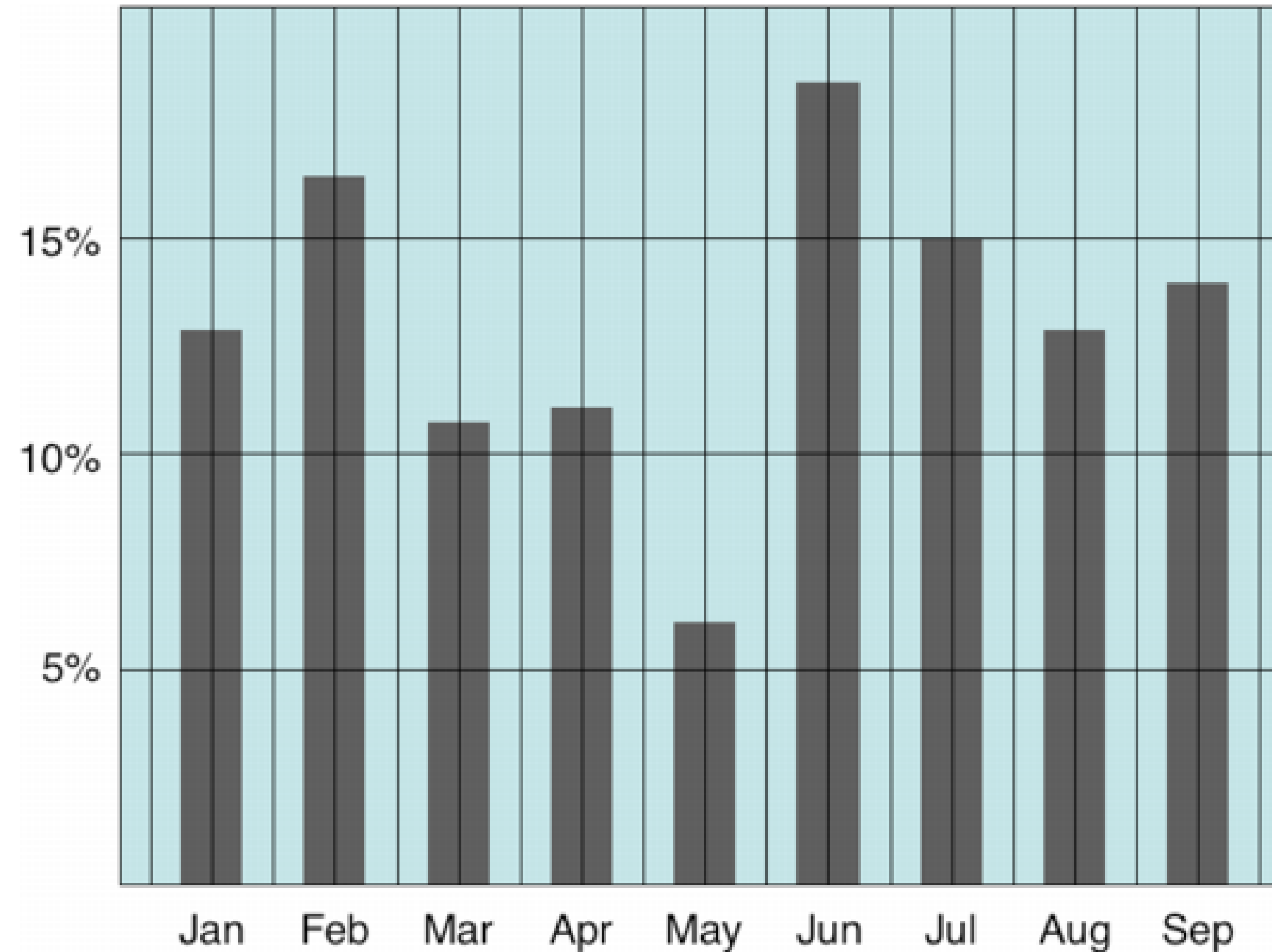


Maximize Data-Ink Ratio

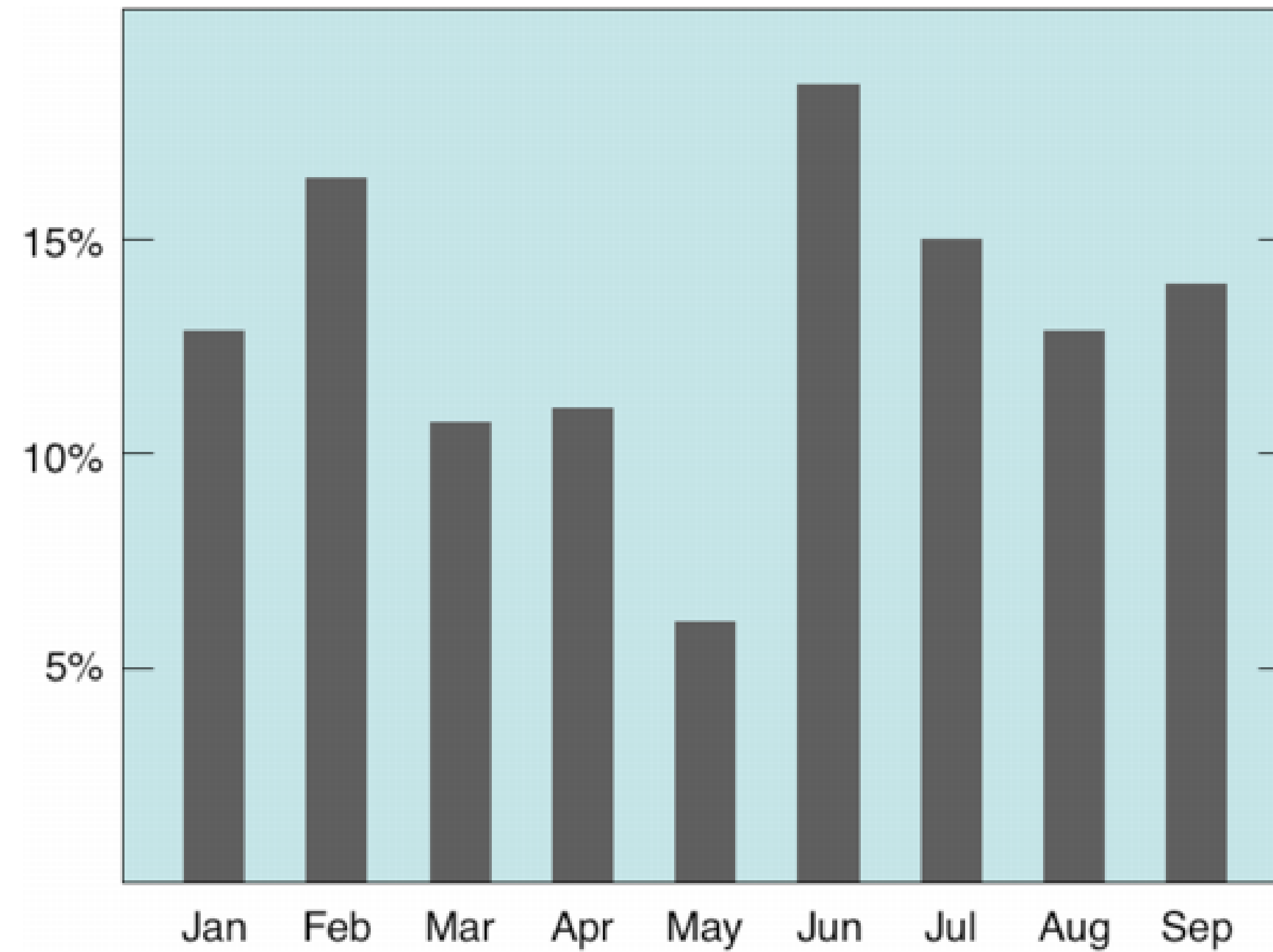


Avoid Chart Junk

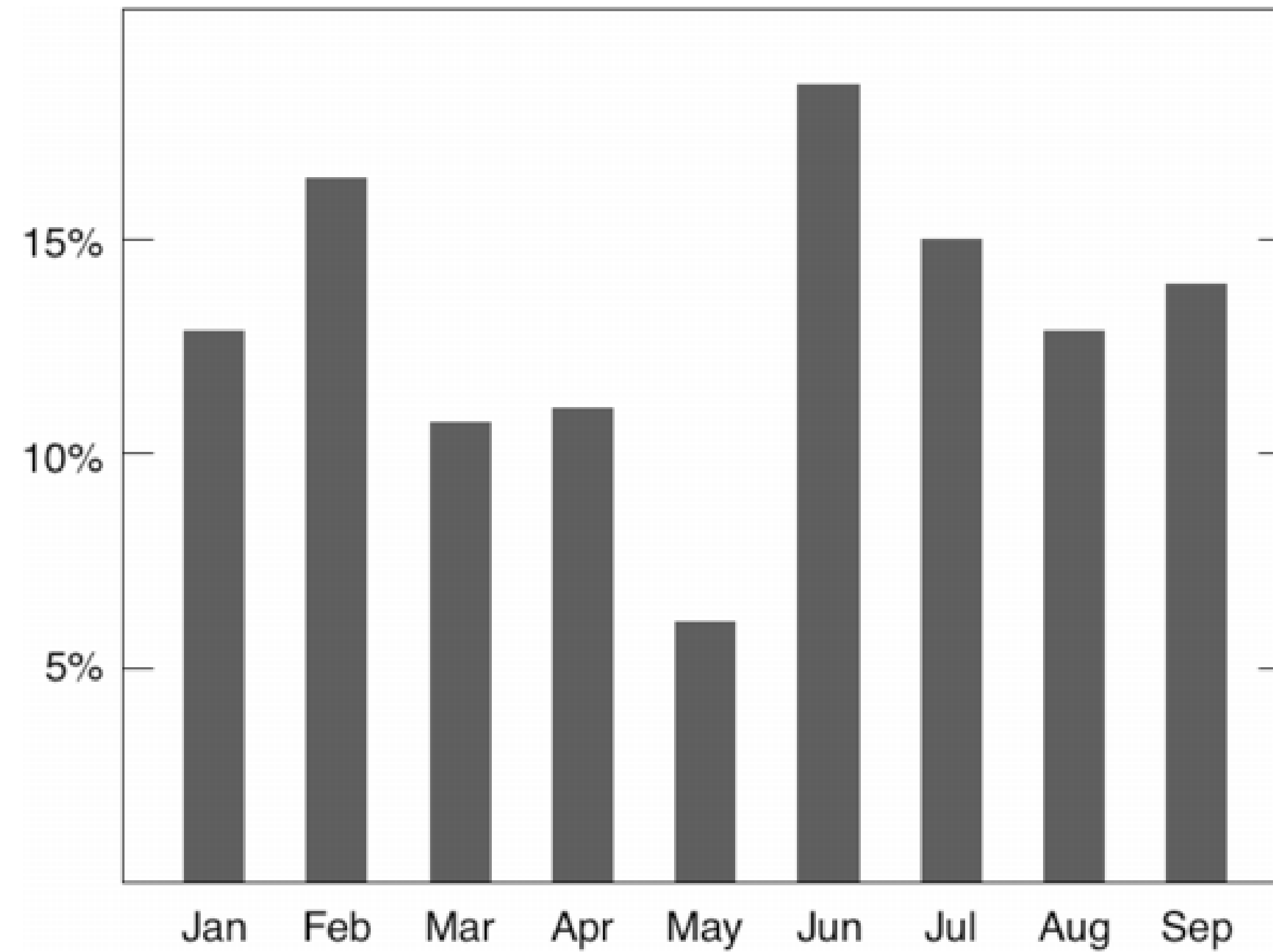
Extraneous visual elements that distract from the message



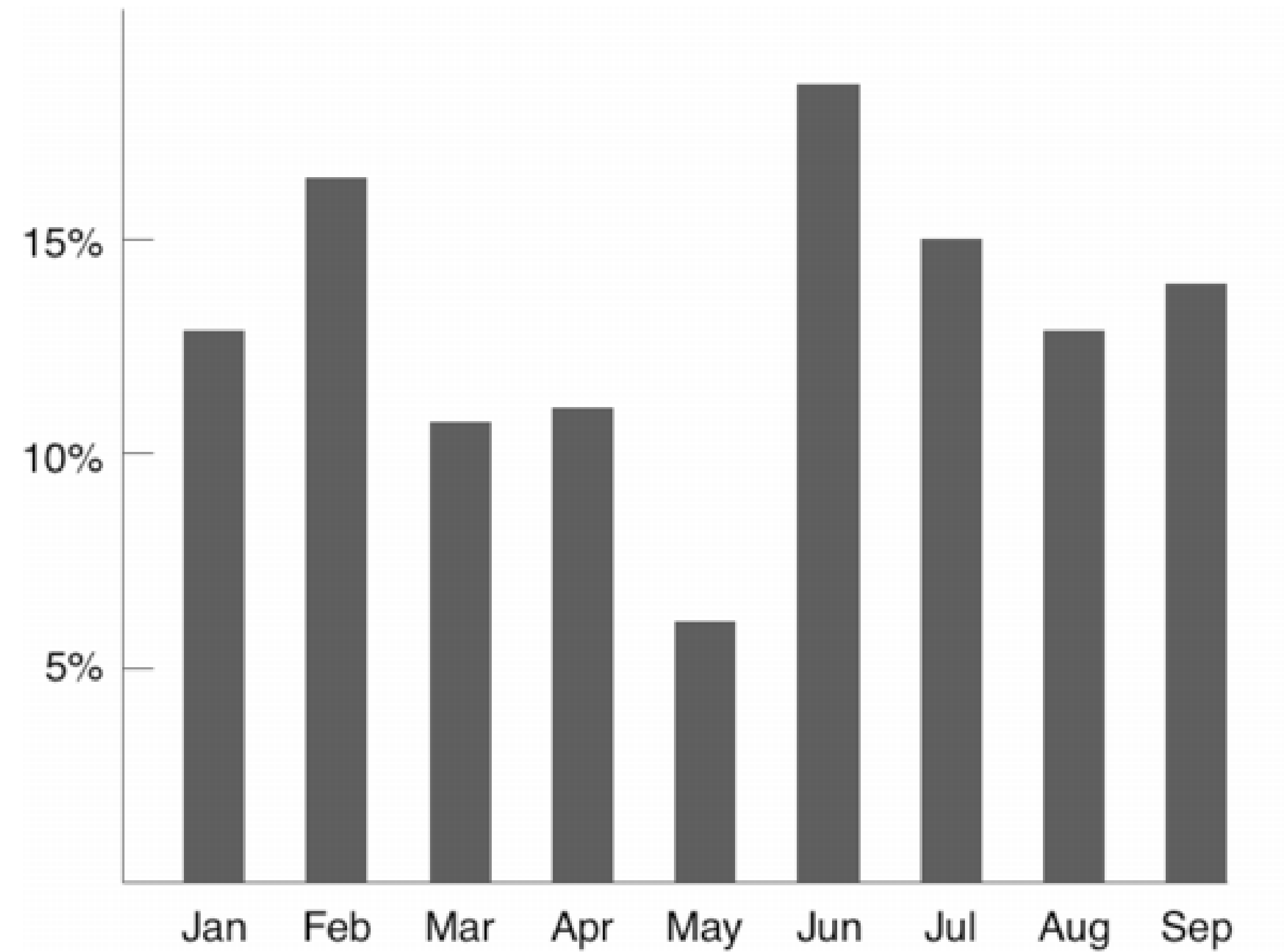
Avoid Chart Junk



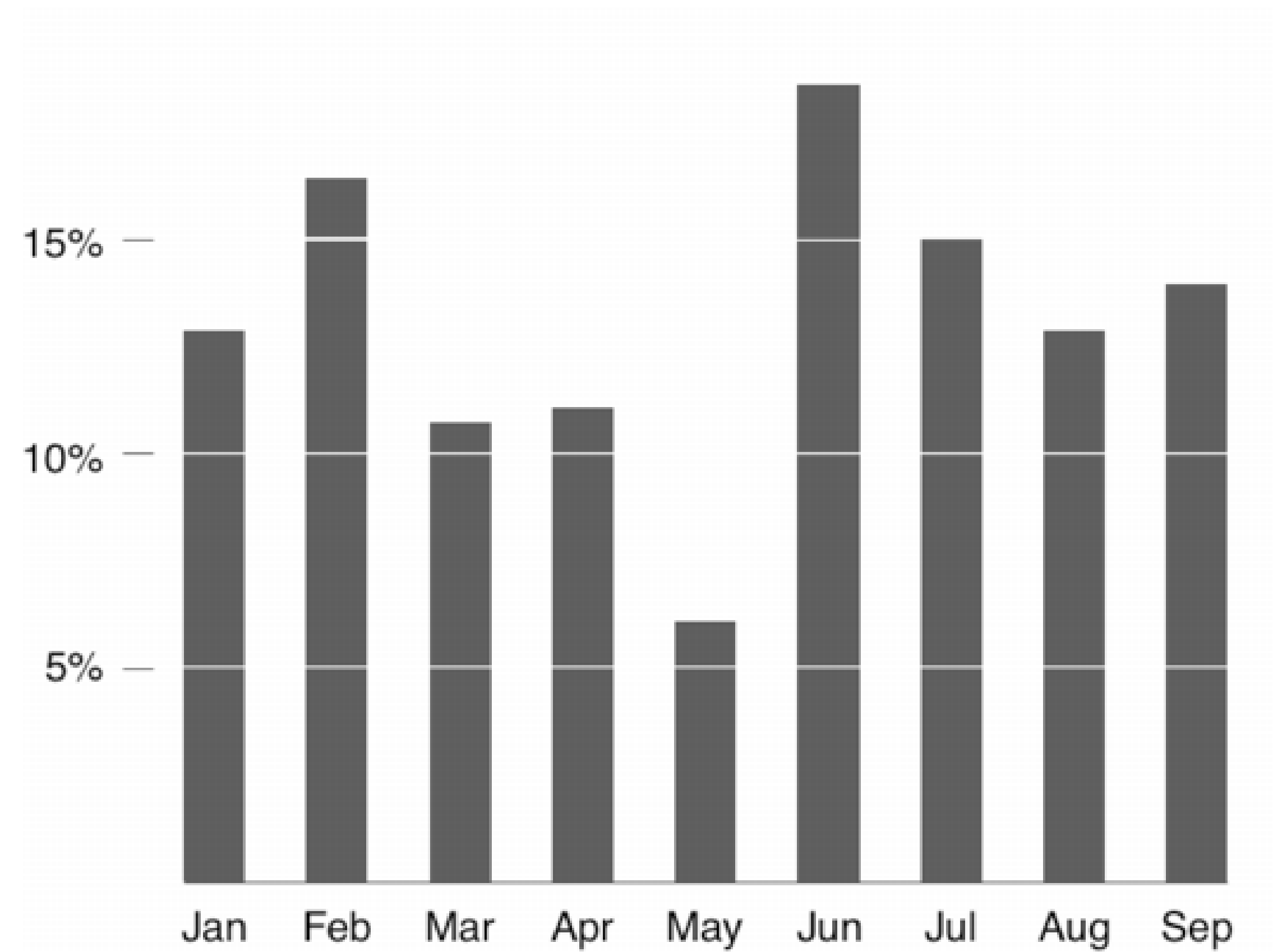
Avoid Chart Junk



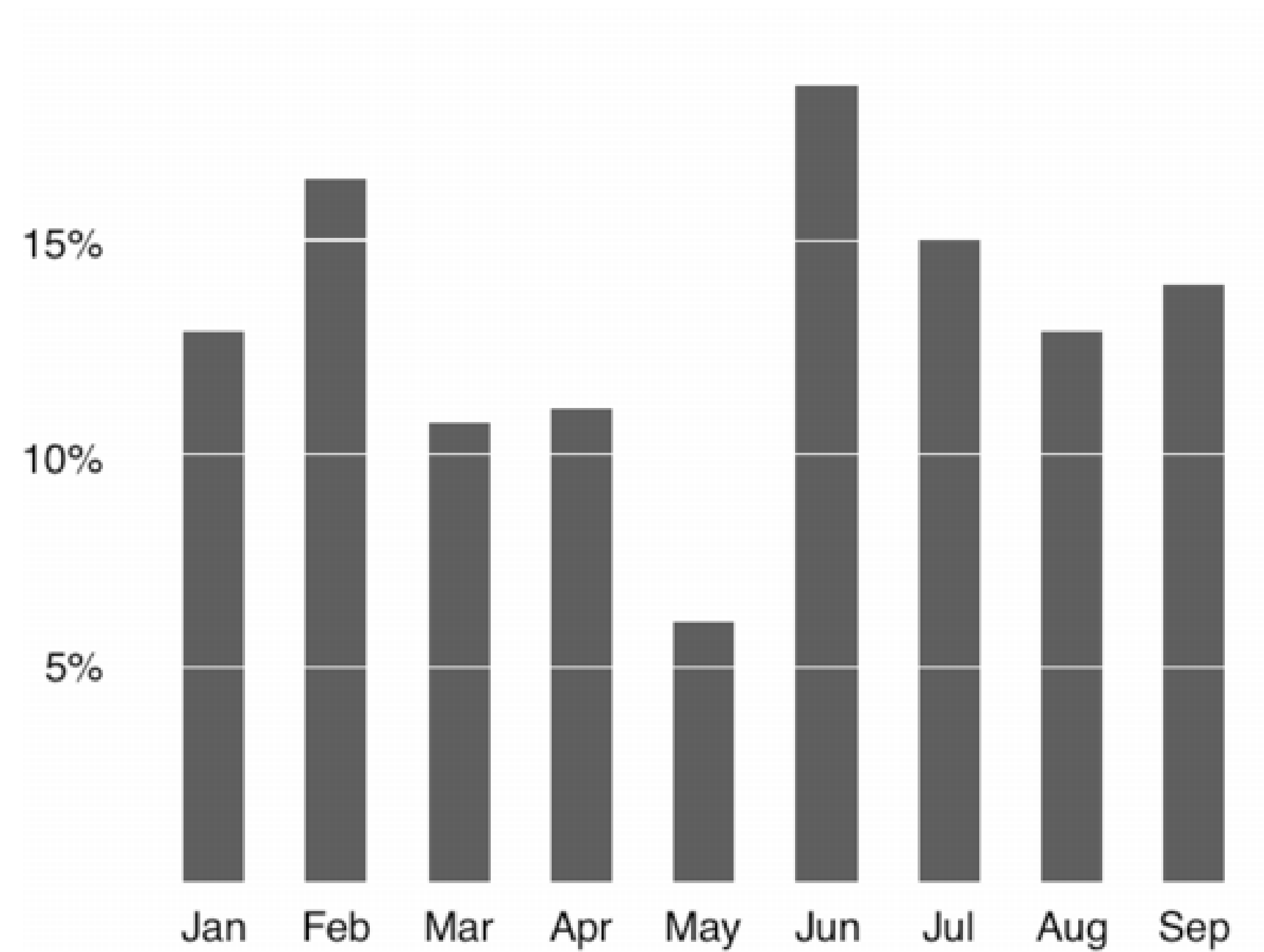
Avoid Chart Junk



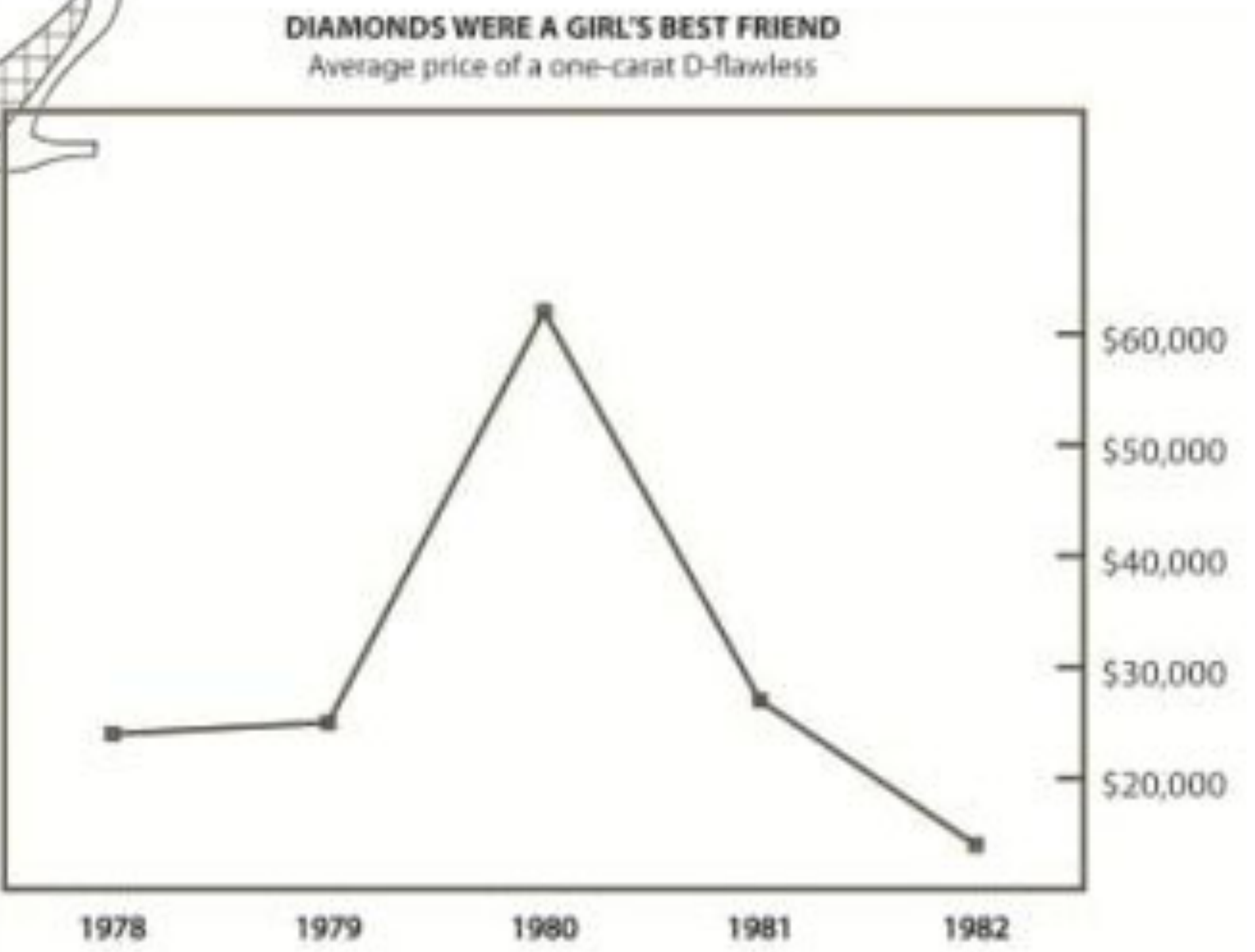
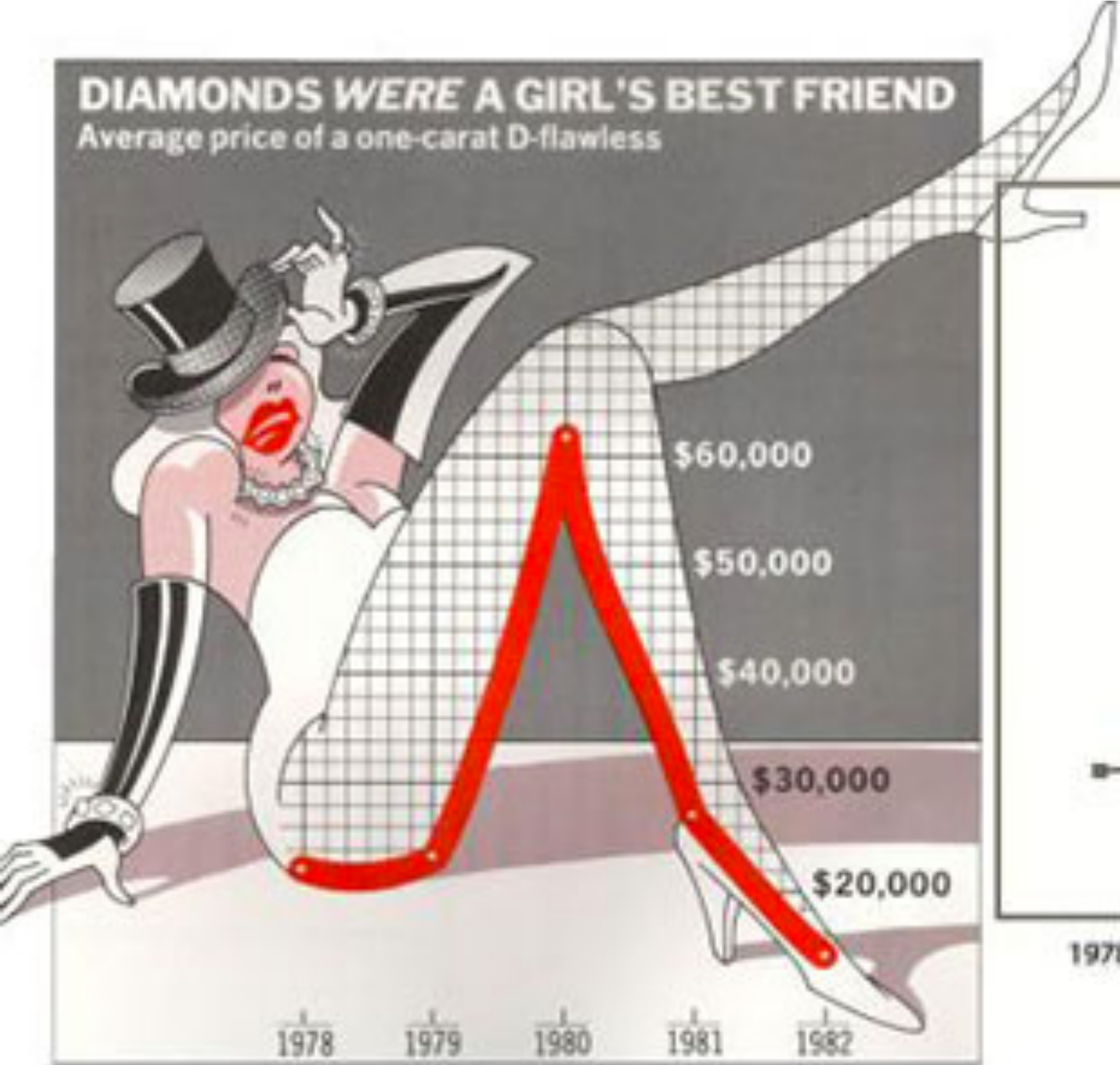
Avoid Chart Junk



Avoid Chart Junk

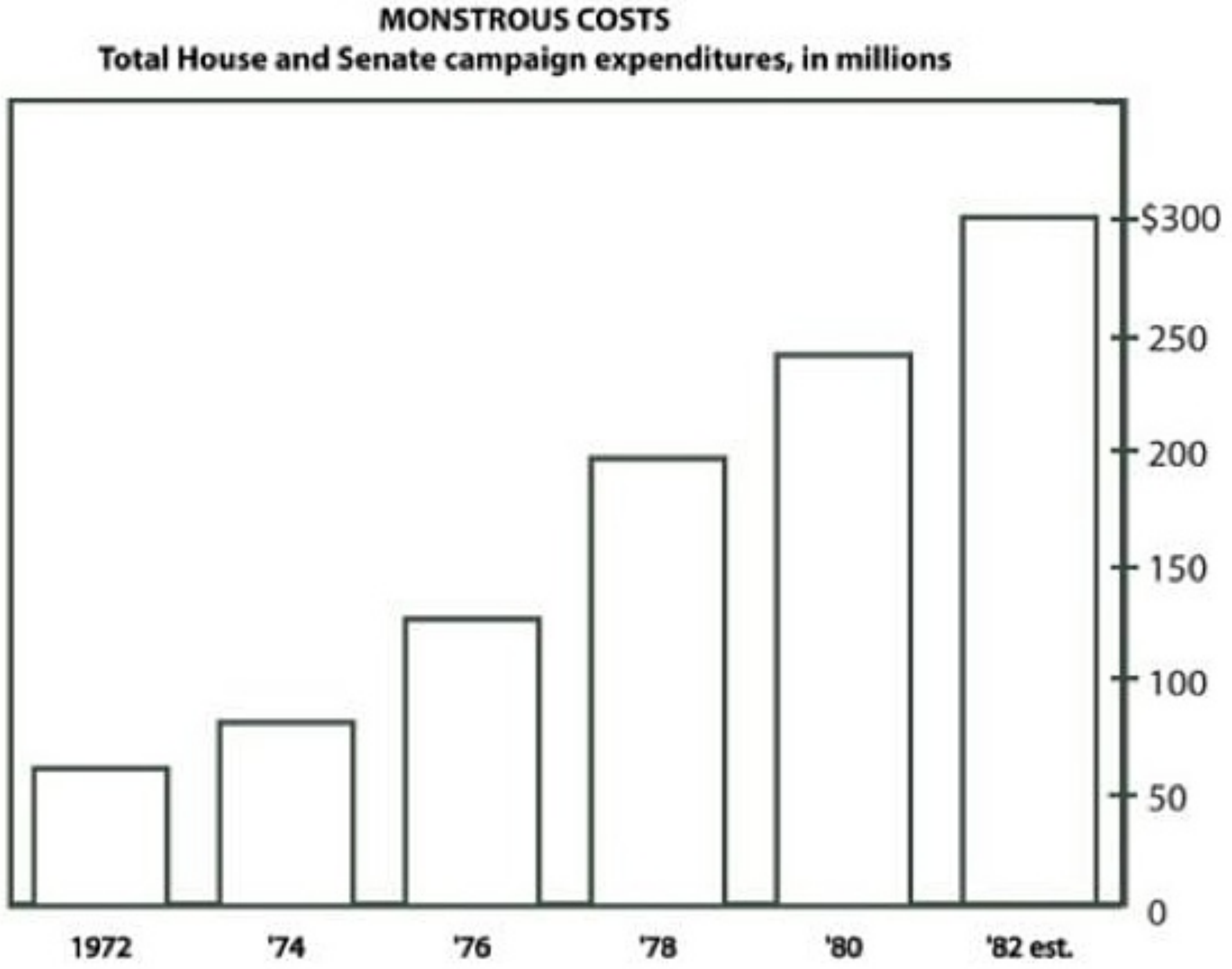


Which is better?



[Bateman et al. 2010]

Which is better?



[Bateman et al. 2010]

<https://eagereyes.org/criticism/chart-junk-considered-useful-after-all>

Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts

Scott Bateman, Regan L. Mandryk, Carl Gutwin,
Aaron Genest, David McDine, Christopher Brooks

Department of Computer Science, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
scott.bateman@usask.ca, regan@cs.usask.ca, gutwin@cs.usask.ca,
aaron.genest@usask.ca, dam085@mail.usask.ca, cab938@mail.usask.ca

ABSTRACT

Guidelines for designing information charts often state that the presentation should reduce ‘chart junk’ – visual embellishments that are not essential to understanding the data. In contrast, some popular chart designers wrap the presented data in detailed and elaborate imagery, raising the questions of whether this imagery is really as detrimental to understanding as has been proposed, and whether the visual embellishment may have other benefits. To investigate these issues, we conducted an experiment that compared embellished charts with plain ones, and measured both interpretation accuracy and long-term recall. We found that people’s accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better. Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

Author Keywords

Charts, information visualization, imagery, memorability.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

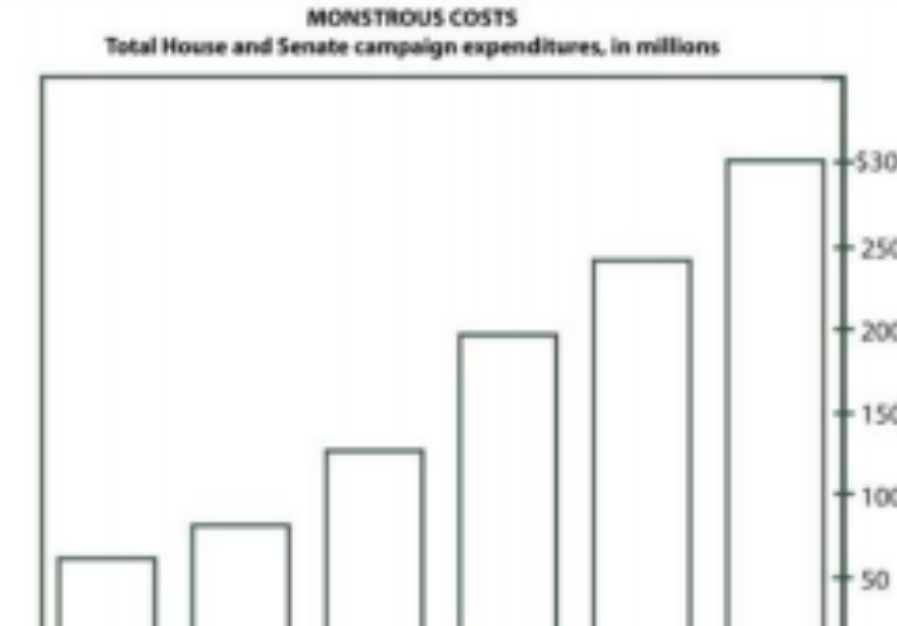
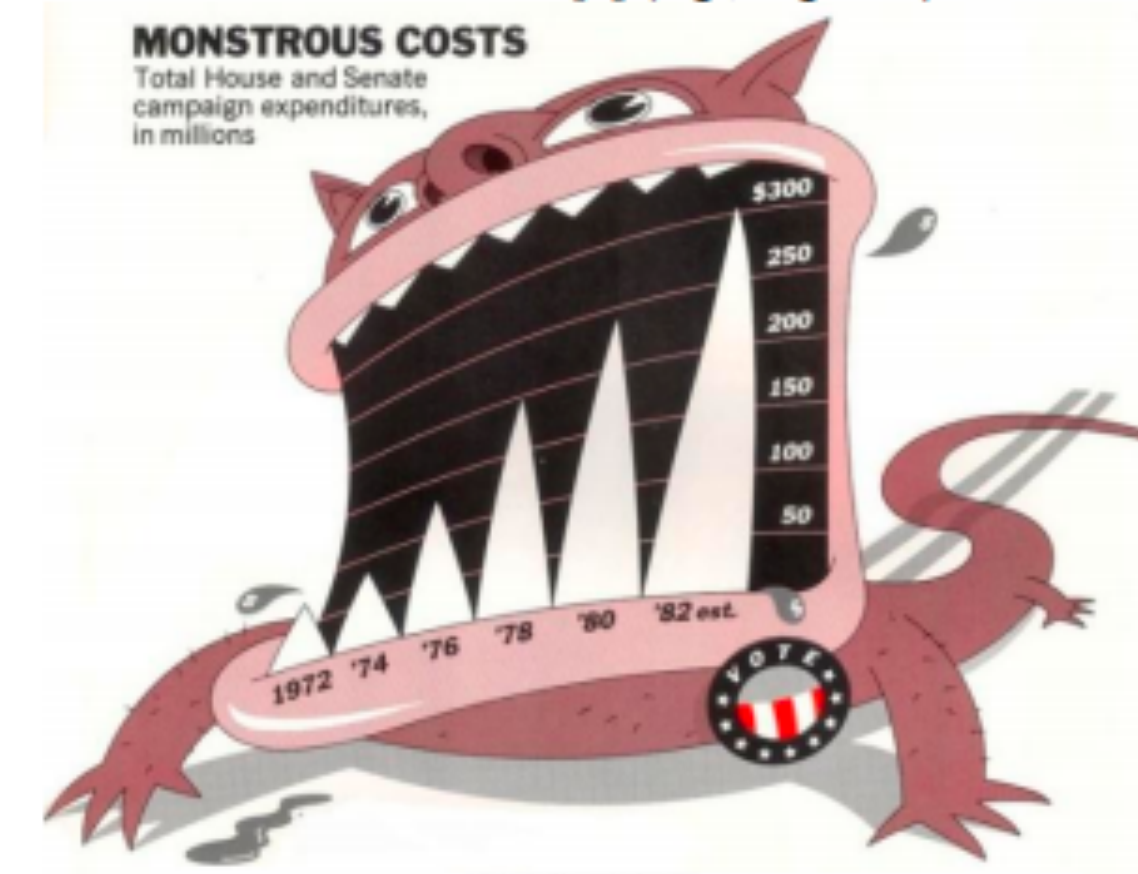
Design, Human Factors

INTRODUCTION

Many experts in the area of chart design, such as Edward Tufte, criticize the inclusion of visual embellishment in charts and graphs; their guidelines for good chart design often suggest that the addition of *chart junk*, decorations and other kinds of non-essential imagery, to a chart can make interpretation more difficult and can distract readers from the data [22]. This *minimalist* perspective advocates

data-ink – or the ink in the chart used to represent data.

Despite these minimalist guidelines, many designers include a wide variety of visual embellishments in their charts, from small decorations to large images and visual backgrounds. One well-known proponent of visual embellishment in charts is the graphic artist Nigel Holmes, whose work regularly incorporates strong visual imagery into the fabric of the chart [7] (e.g., Figure 1).



EXPERIMENTAL RESULTS

1. No difference for **interpretation accuracy**
2. No difference in **recall accuracy after a five-minute gap**
3. Significantly **better recall for Holmes charts** of both the chart topic and the details (categories and trend) **after long-term gap (2-3 weeks)**.
4. Participants **saw value messages** in the Holmes charts significantly more often than in the plain charts.
5. Participants found the Holmes charts **more attractive, most enjoyed** them, and found that they were **easiest and fastest to remember**.

Use Chart Junk? It depends!

PROS

persuasion

memorability

engagement

CONS

biased analysis

trustworthiness

interpretability

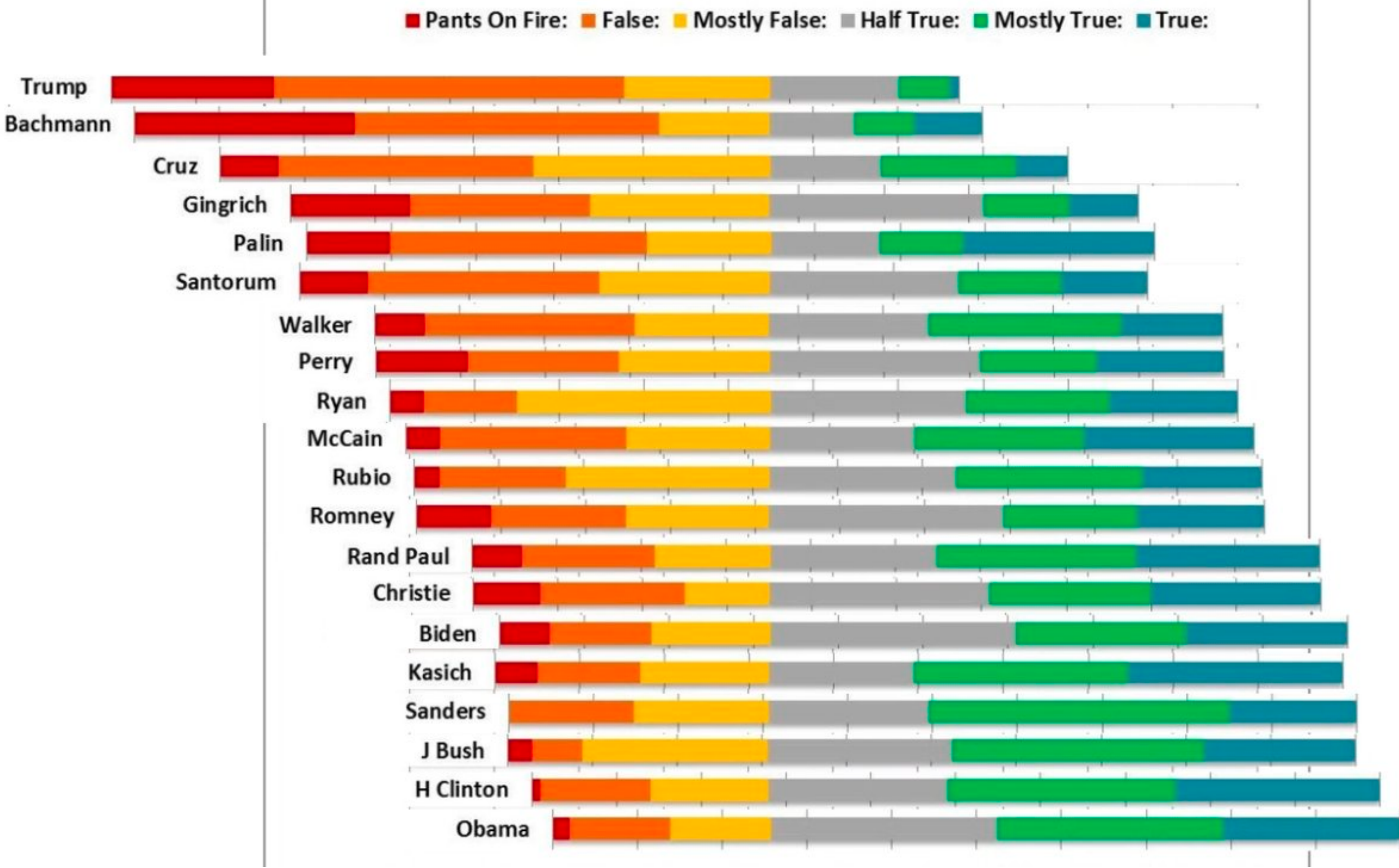
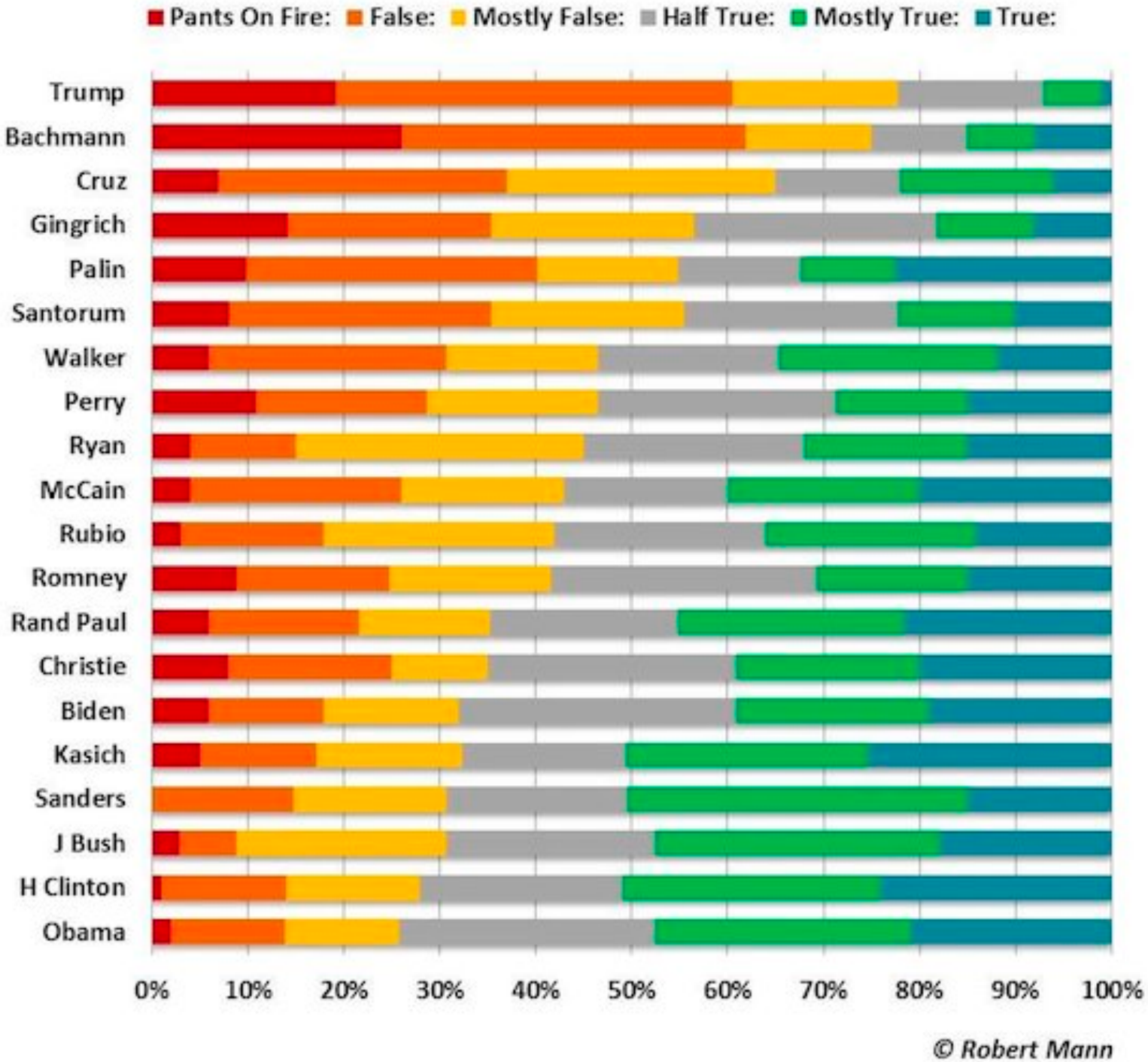
space efficiency

effort

Alignment Matters

Who Lies More: A Comparison

PolitiFact, an independent fact-checking website, has graded more than 50 statements since 2007 from each of these candidates. Here is how they rank.



3D

No Unjustified 3D

Depth judgment is bad

$$N = 0.67 \text{ Sensation} = \text{Intensity}^N$$

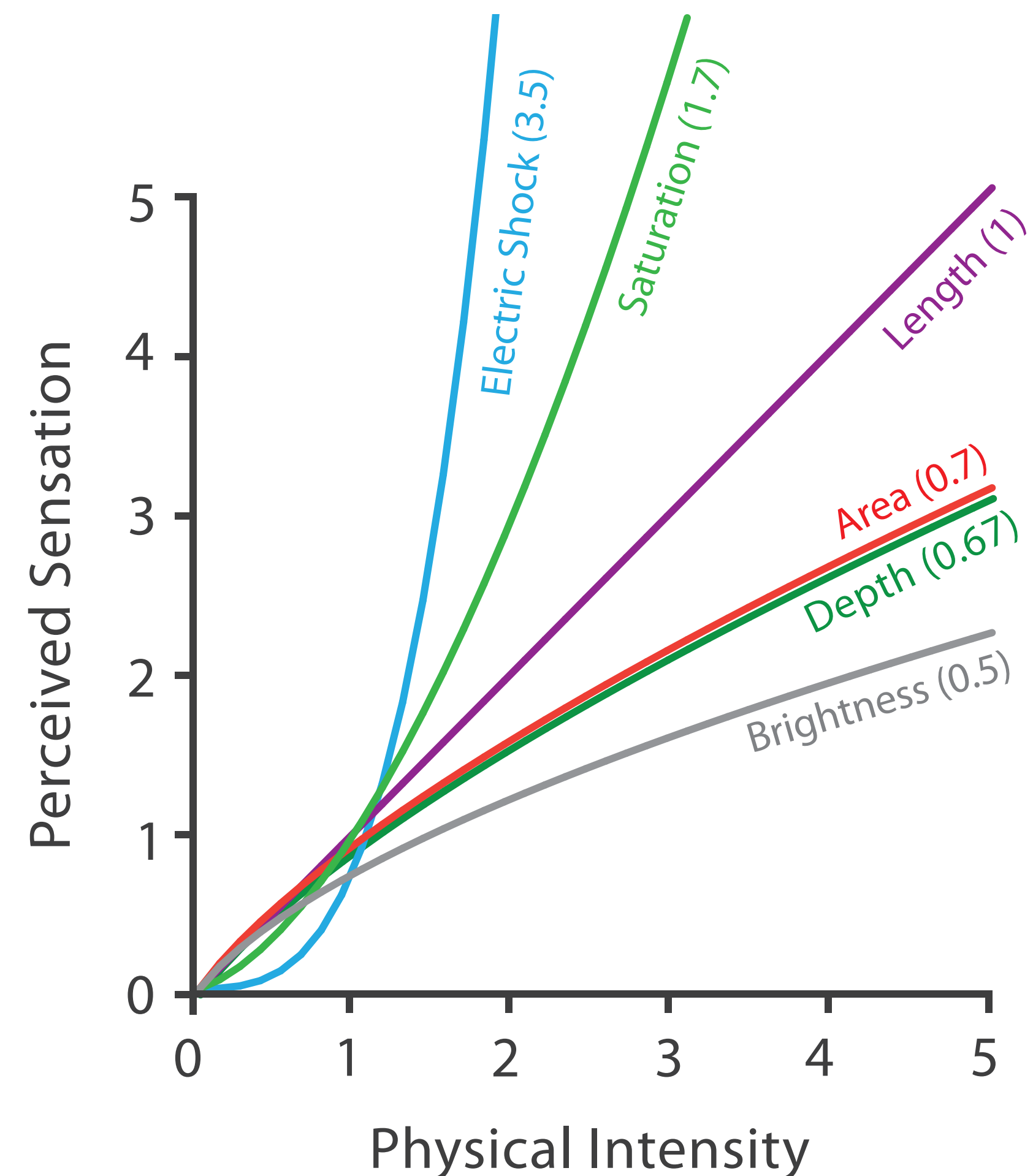
Occlusion

Perspective Distortion

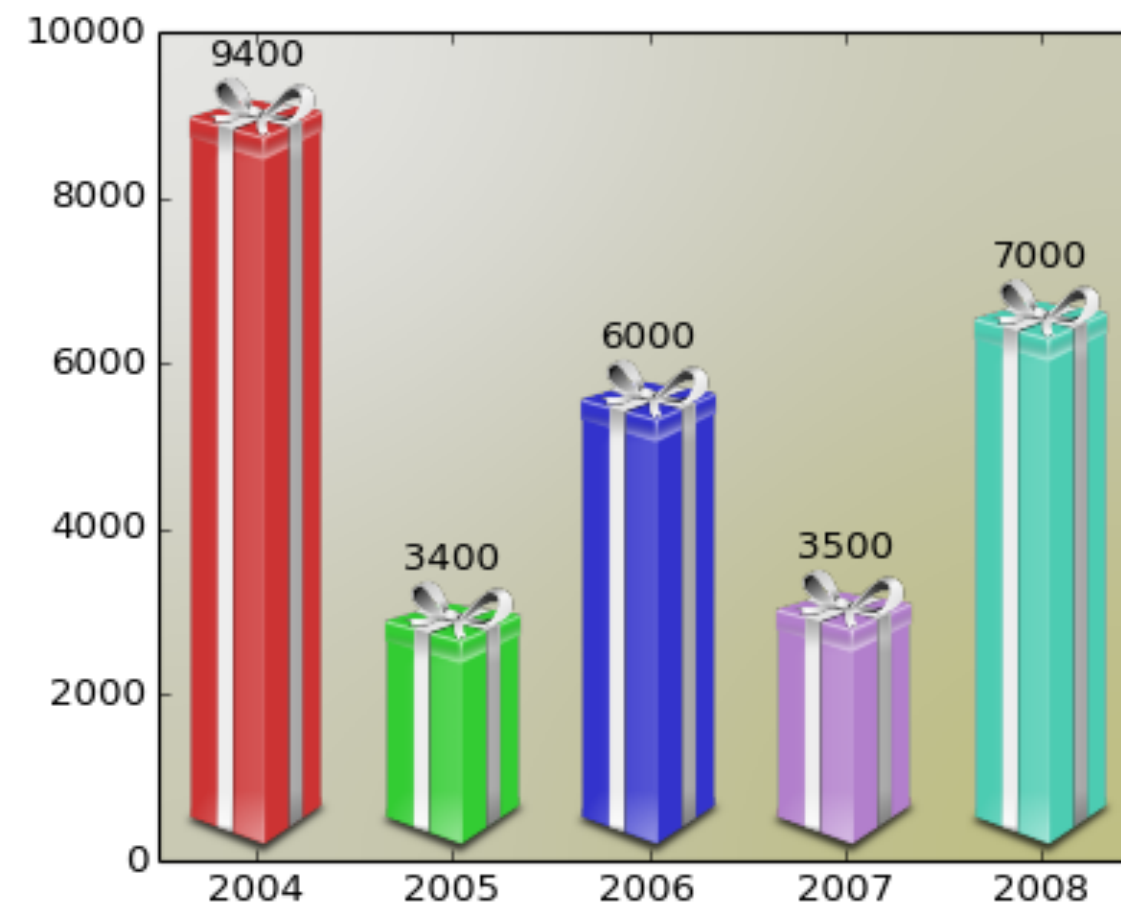
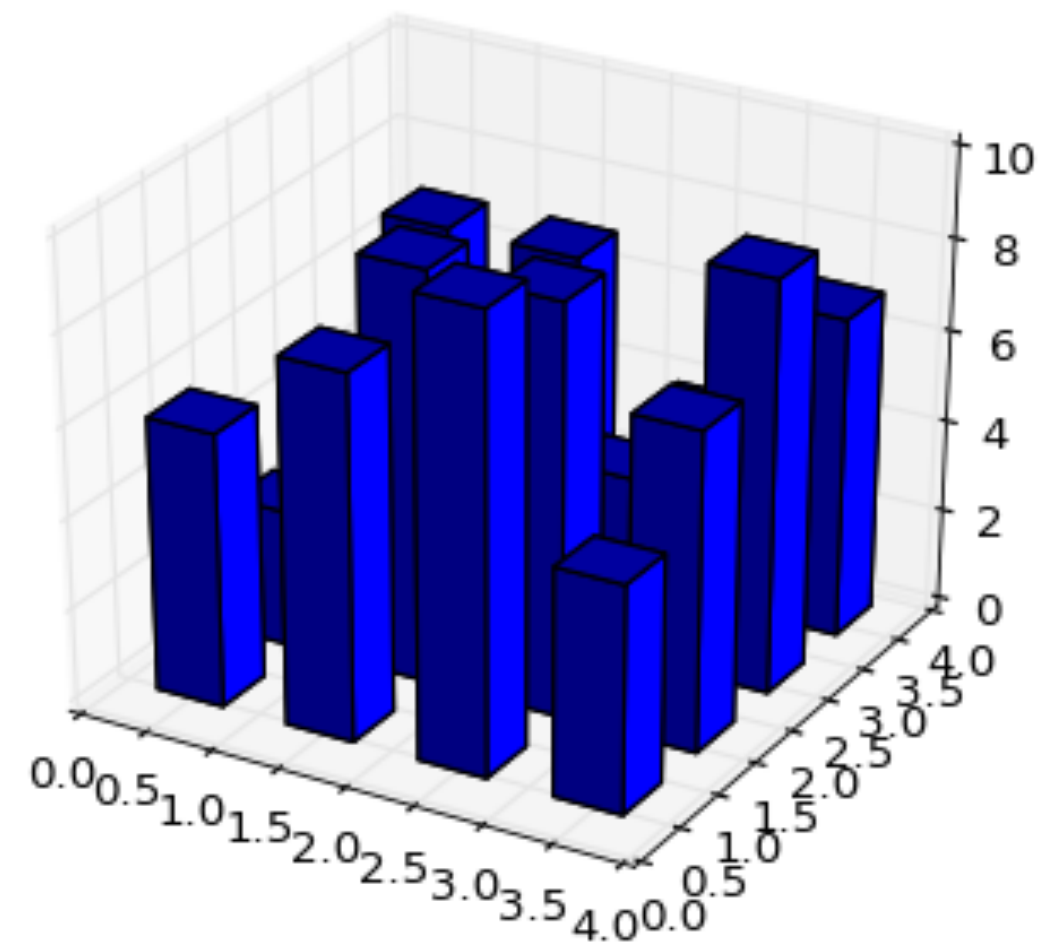
Color: Lighting / Shadows /
Shading

Tilted Text illegible

Steven's Psychophysical Power Law: $S = I^N$

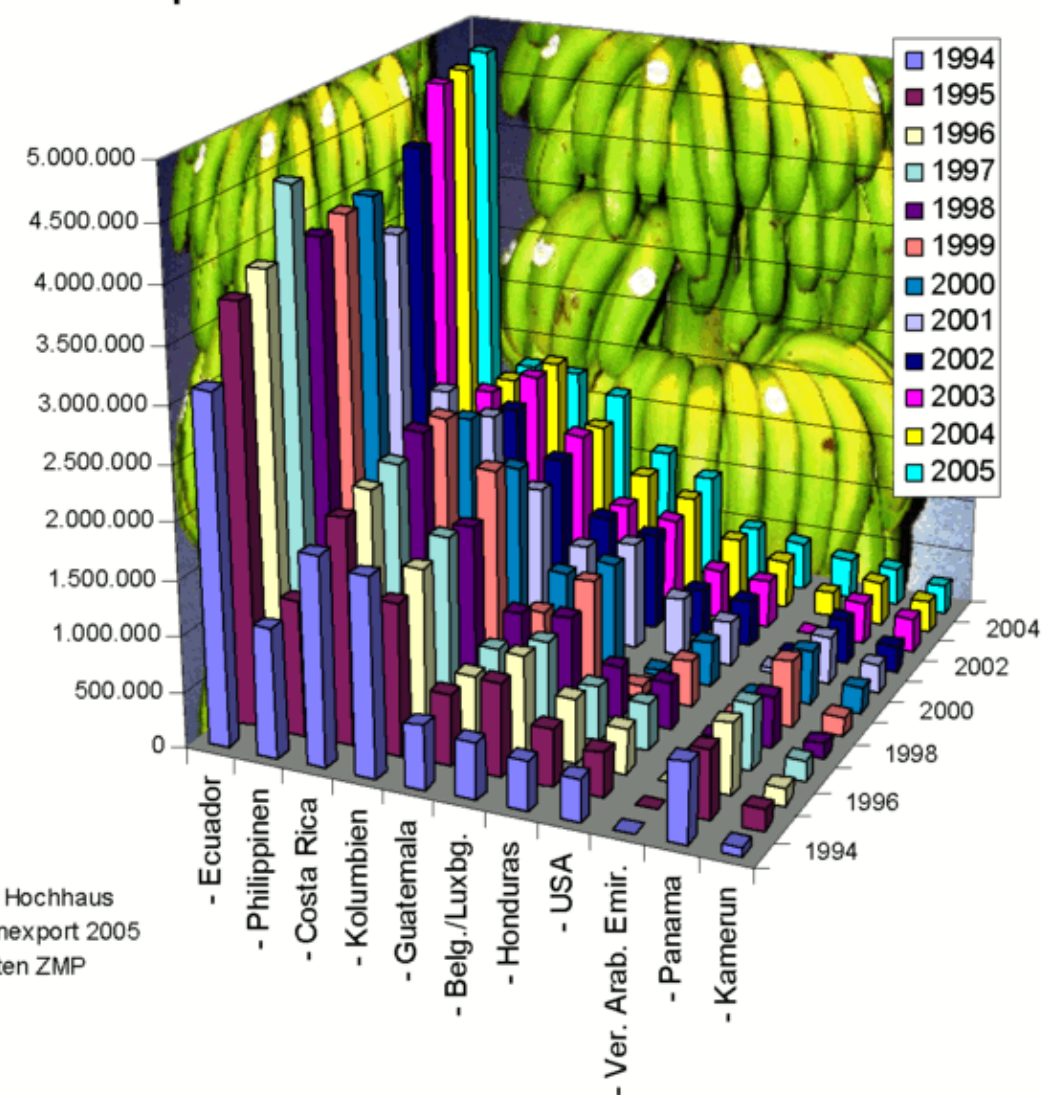


Don't

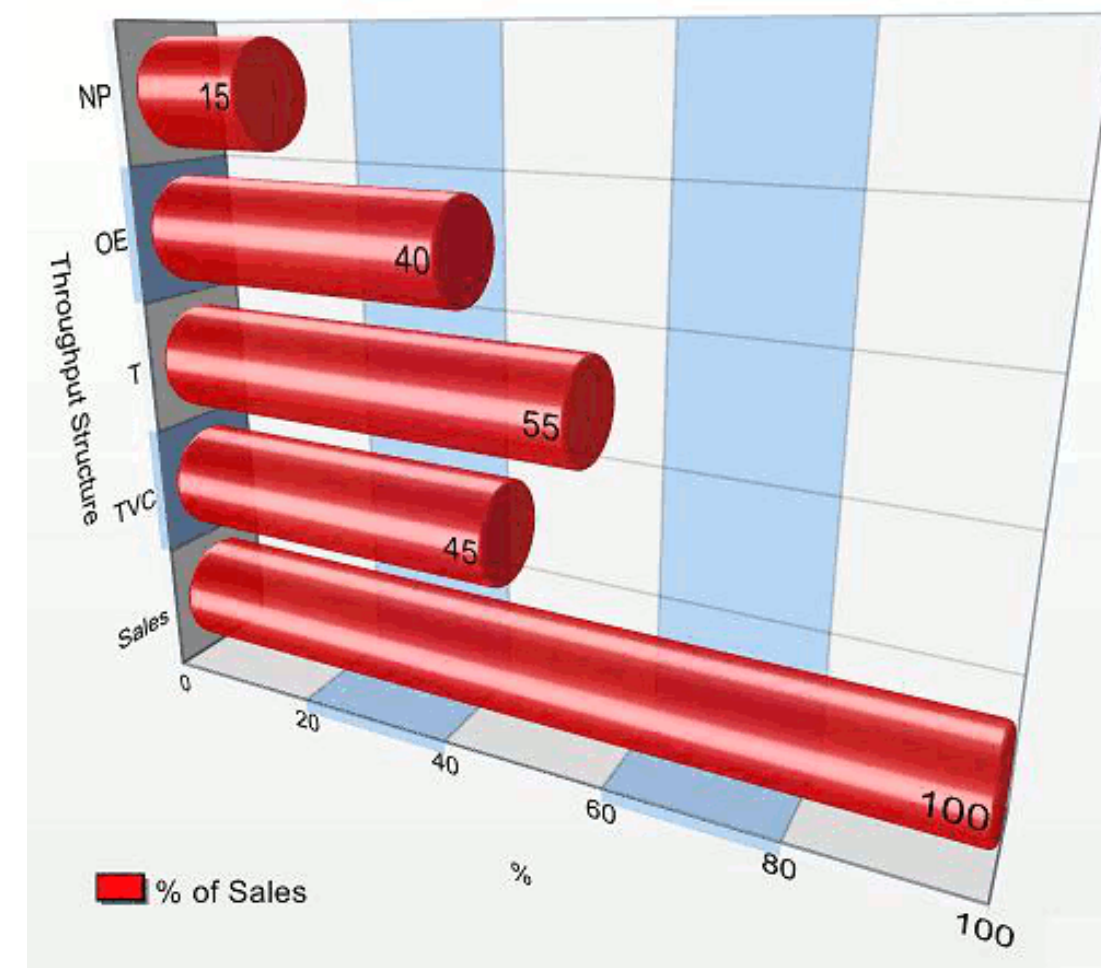


matplotlib gallery

Export von Bananen in Tonnen von 1994-2005



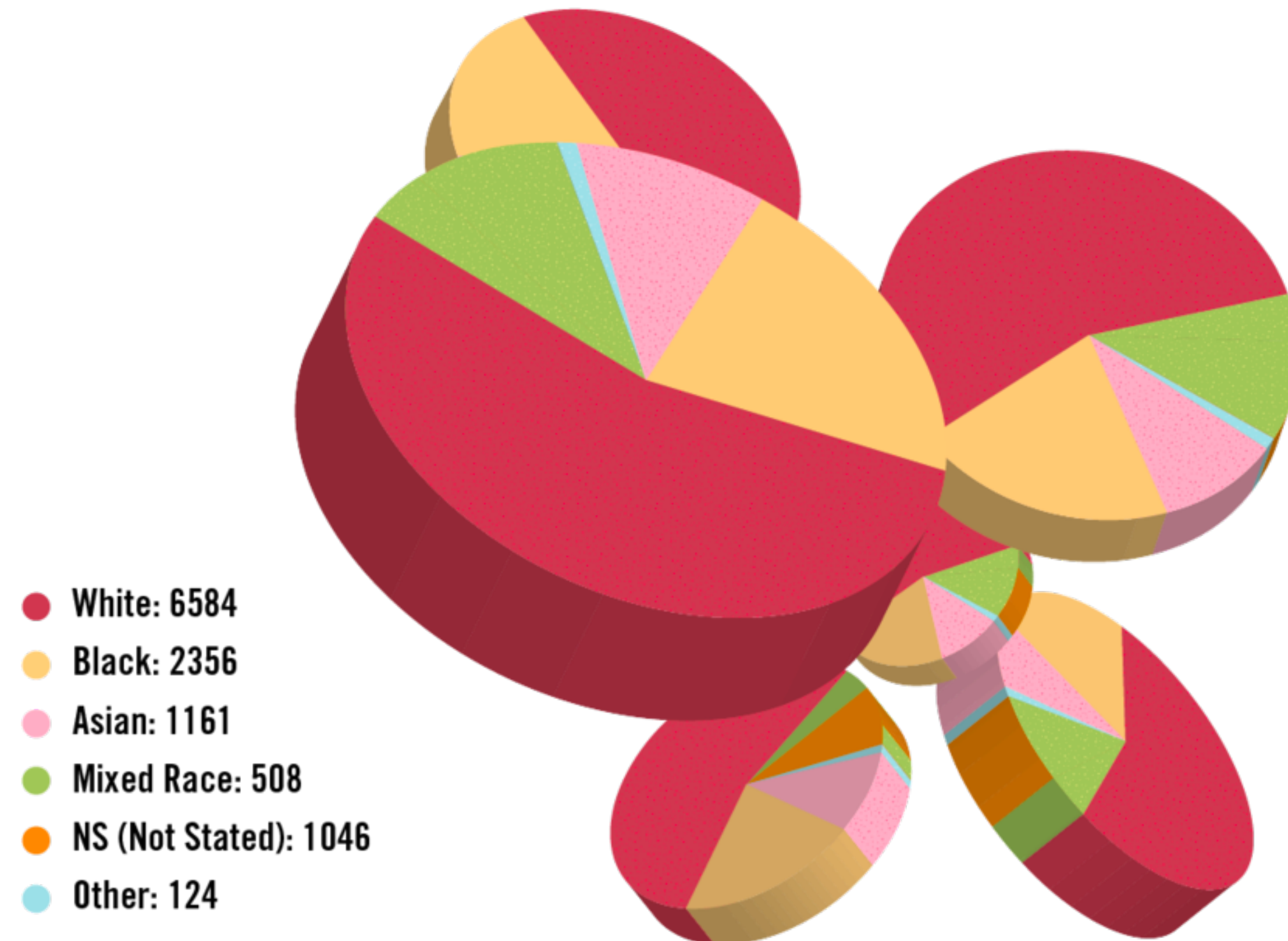
Dr. Hochhaus
Banexport 2005
Daten ZMP



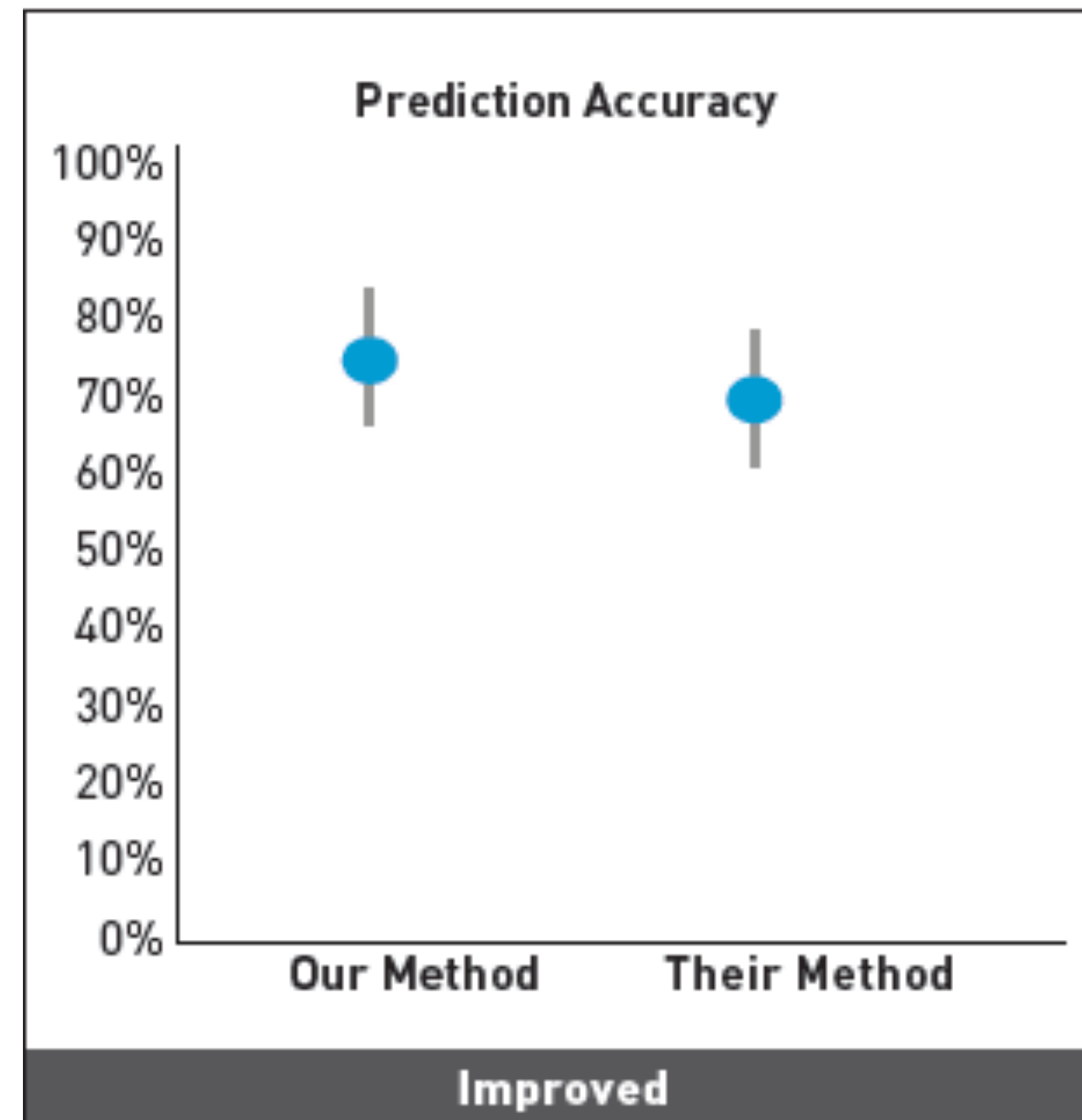
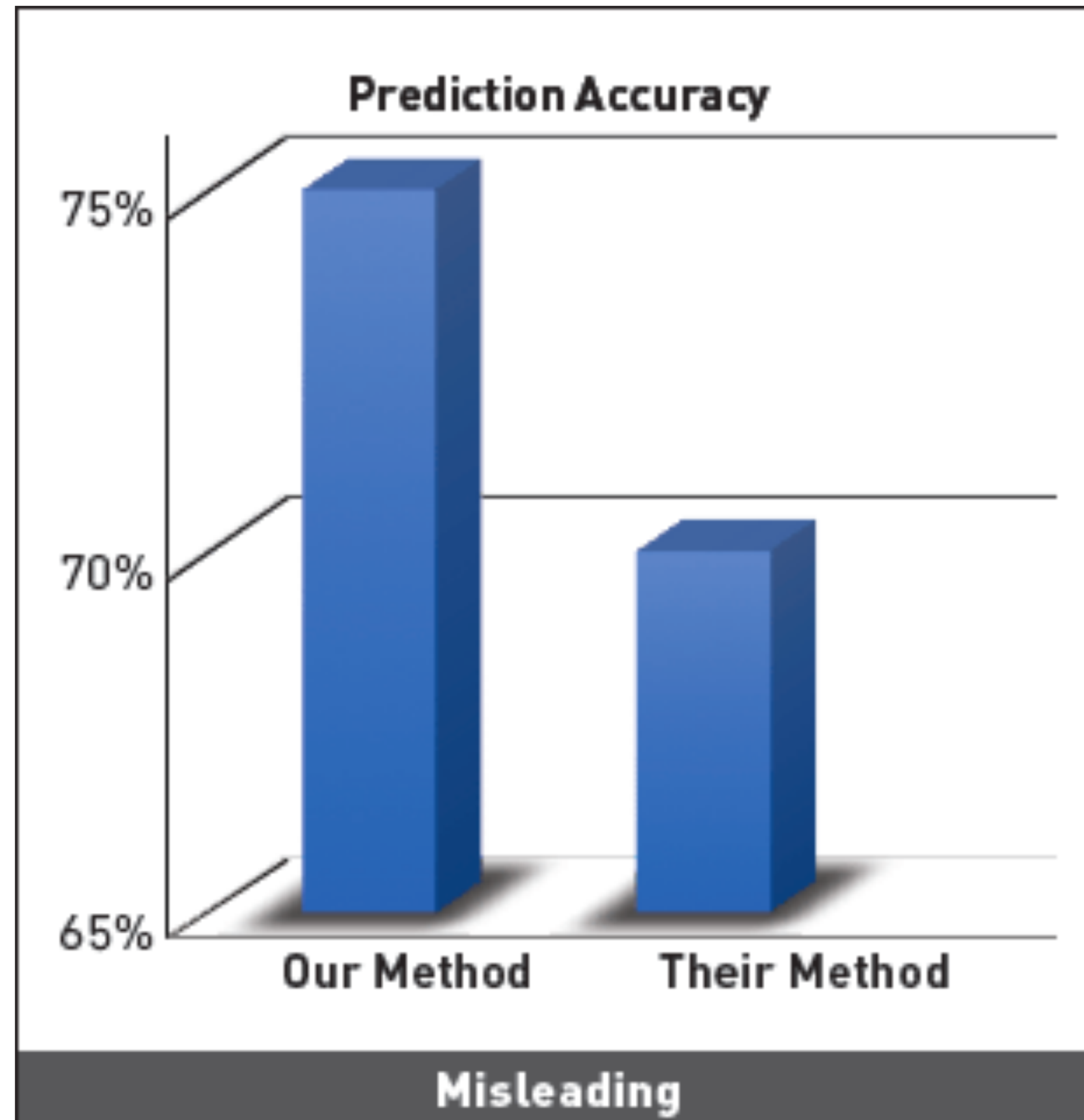
Excel Charts Blog

Don't

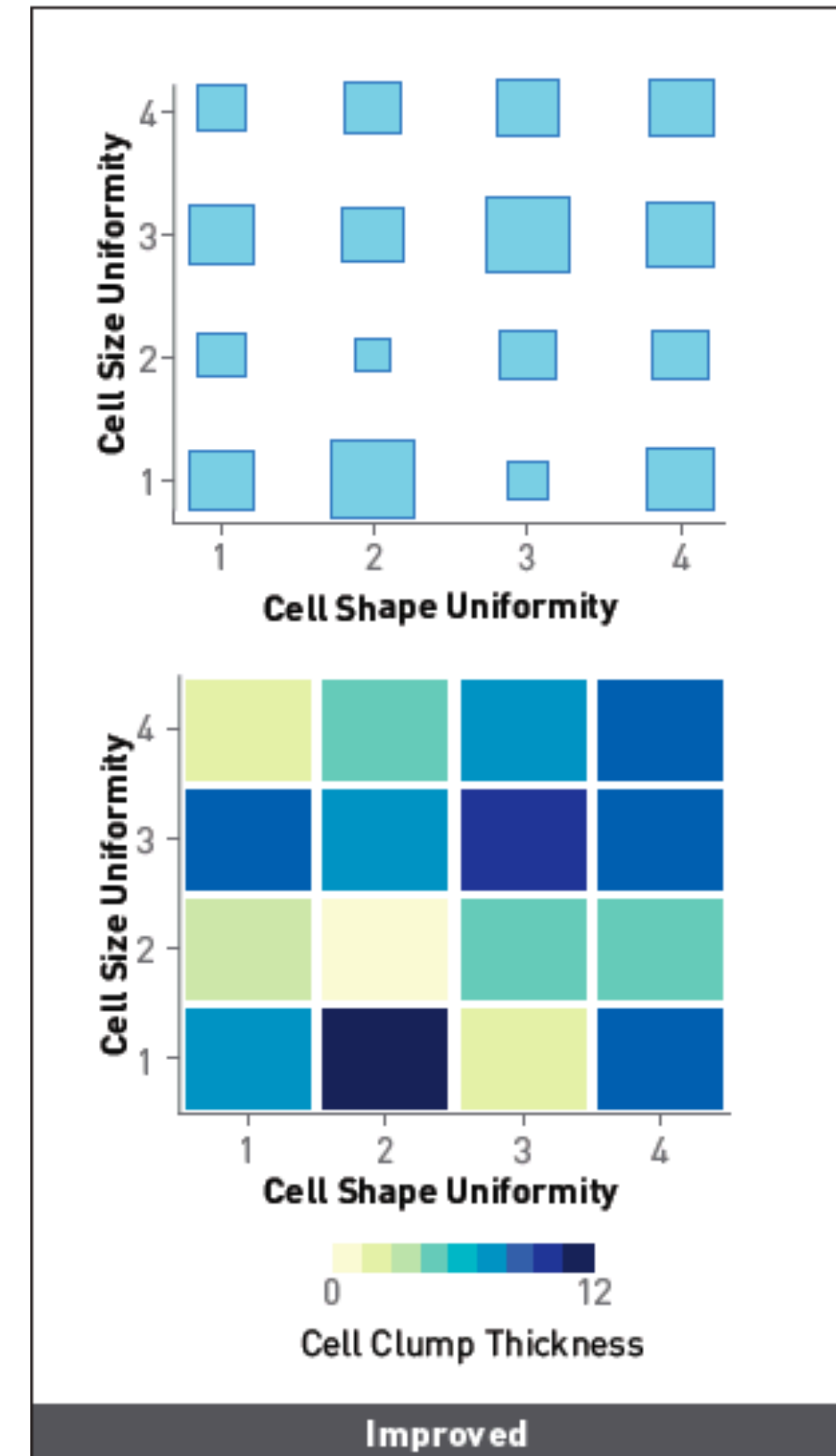
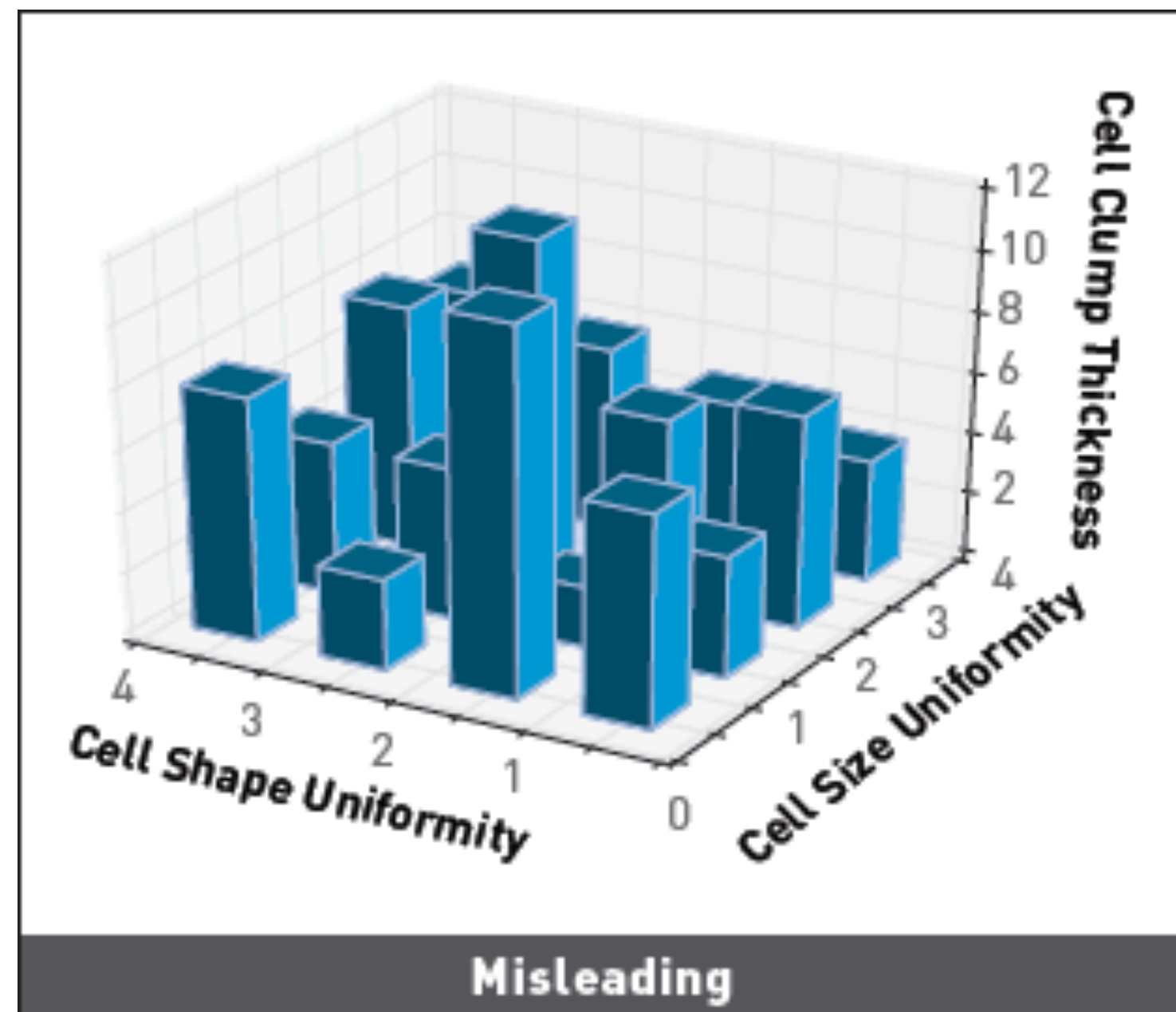
Convictions in England and Wales for class A drug supply.



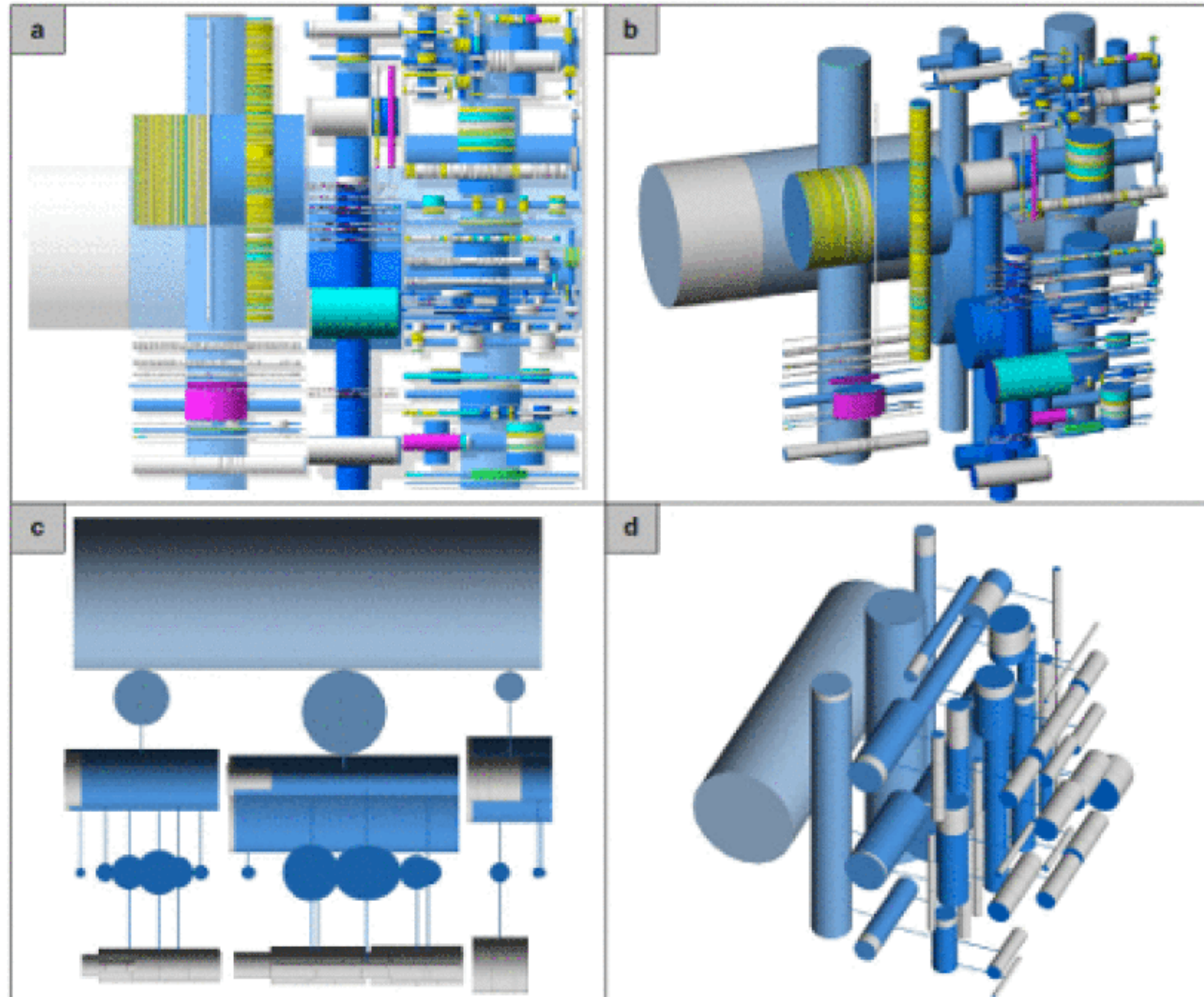
3D Design Alternatives



3D Design Alternatives



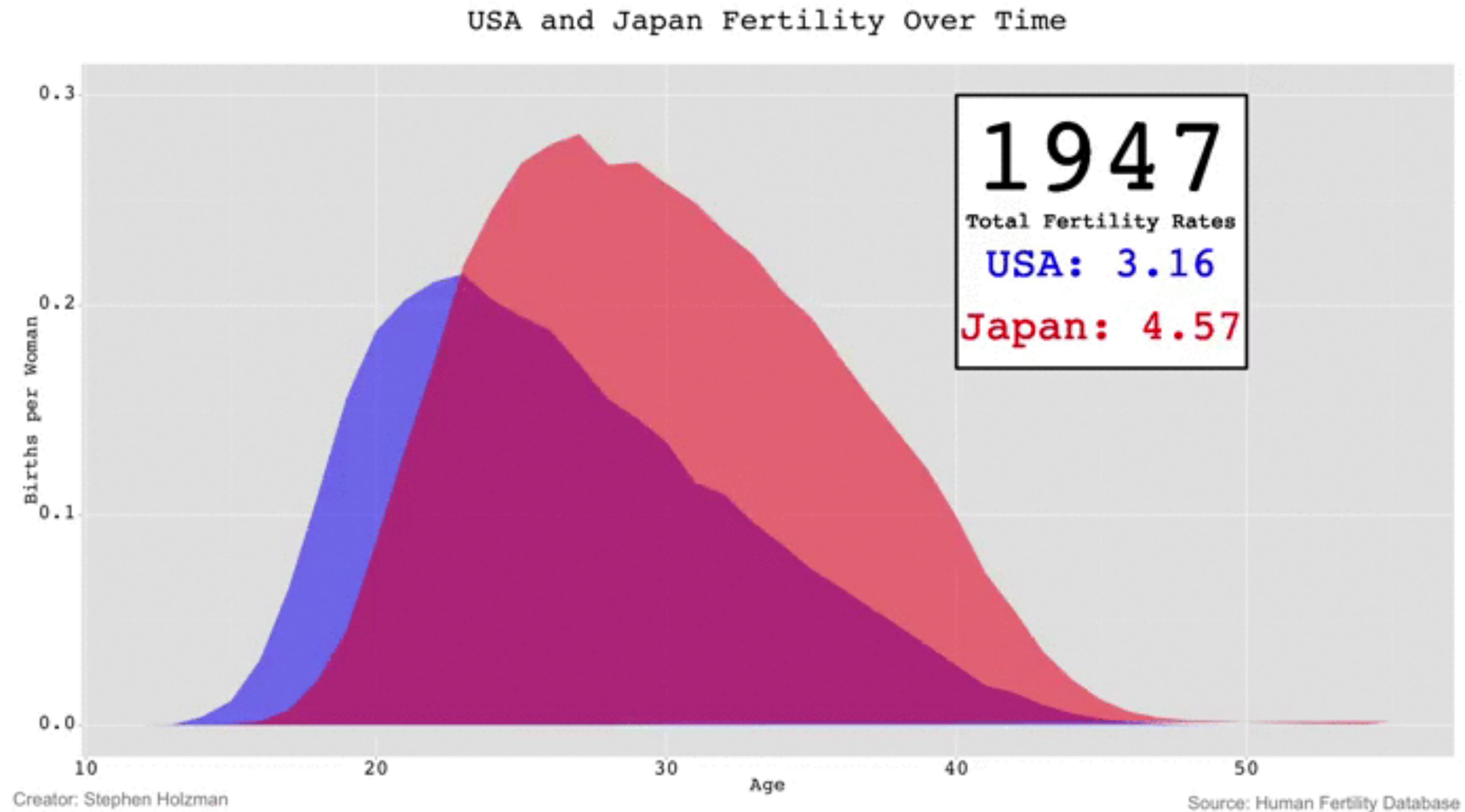
Example: Hierarchy Visualization



More data than fits one
chart:
Animation, Multiple Views

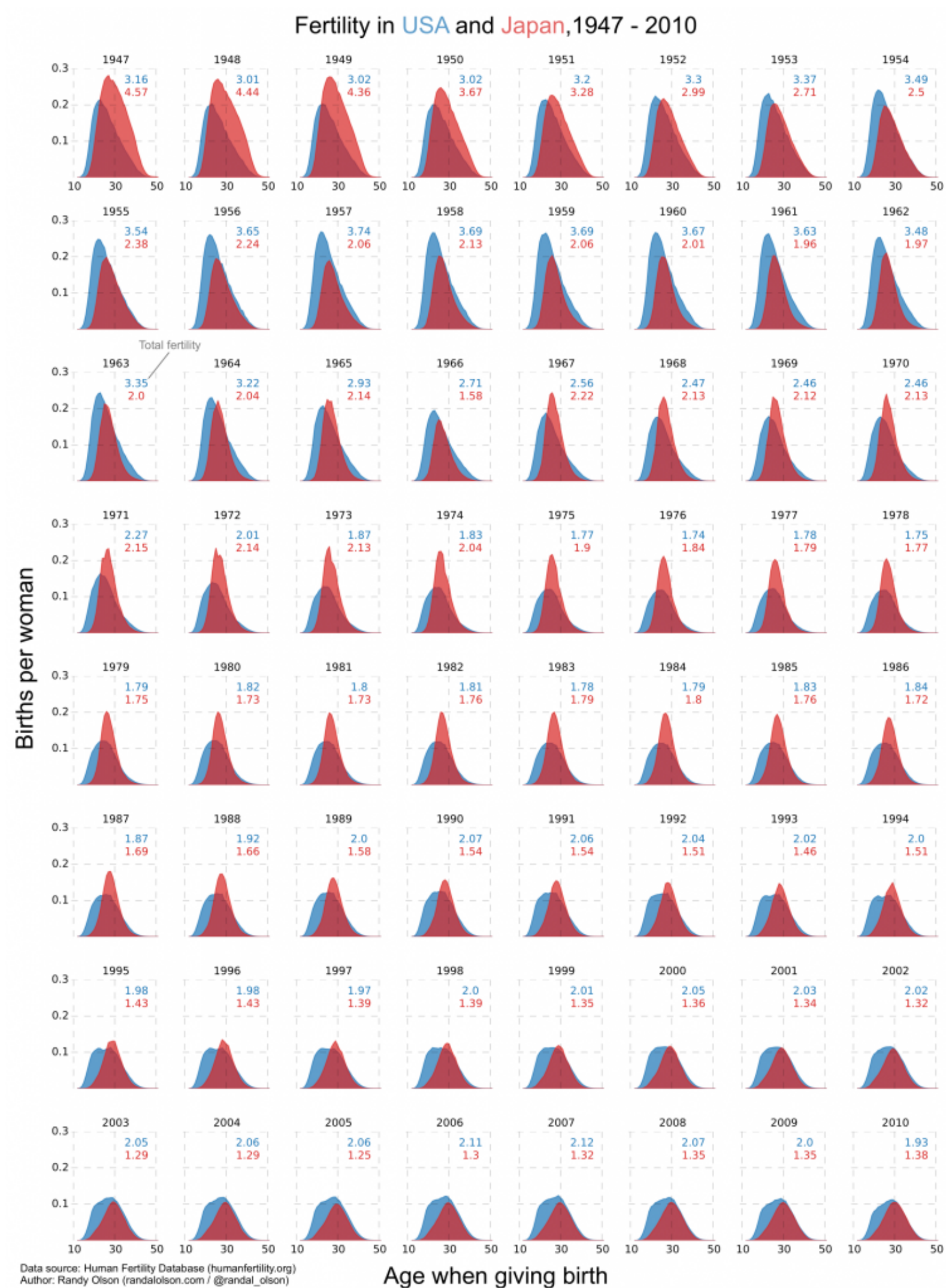
Eyes Beat Memory

Don't make people memorize: Show them



What can we do
differently?

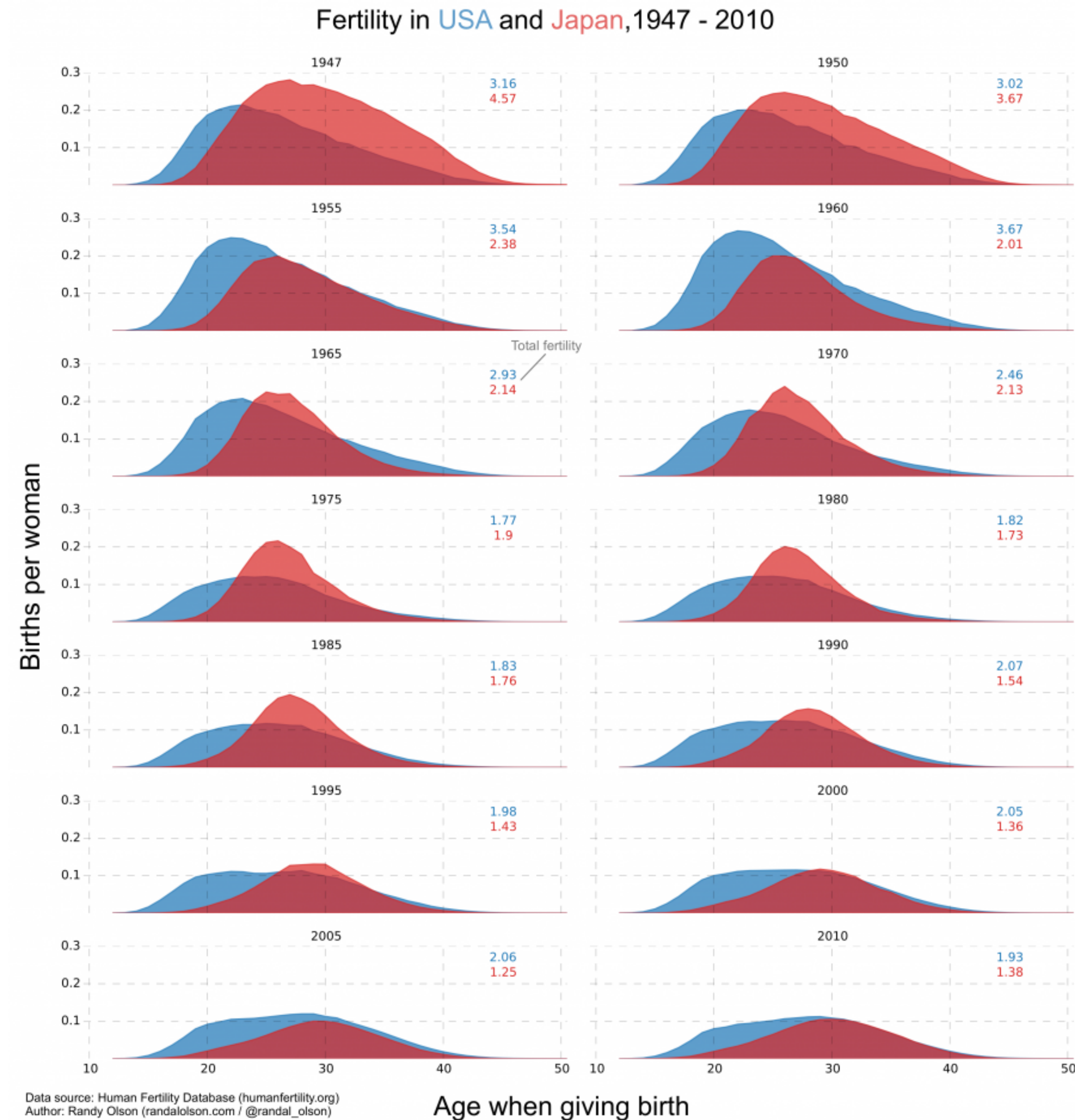
Eyes Beat Memory: Small Multiples



A lot of charts

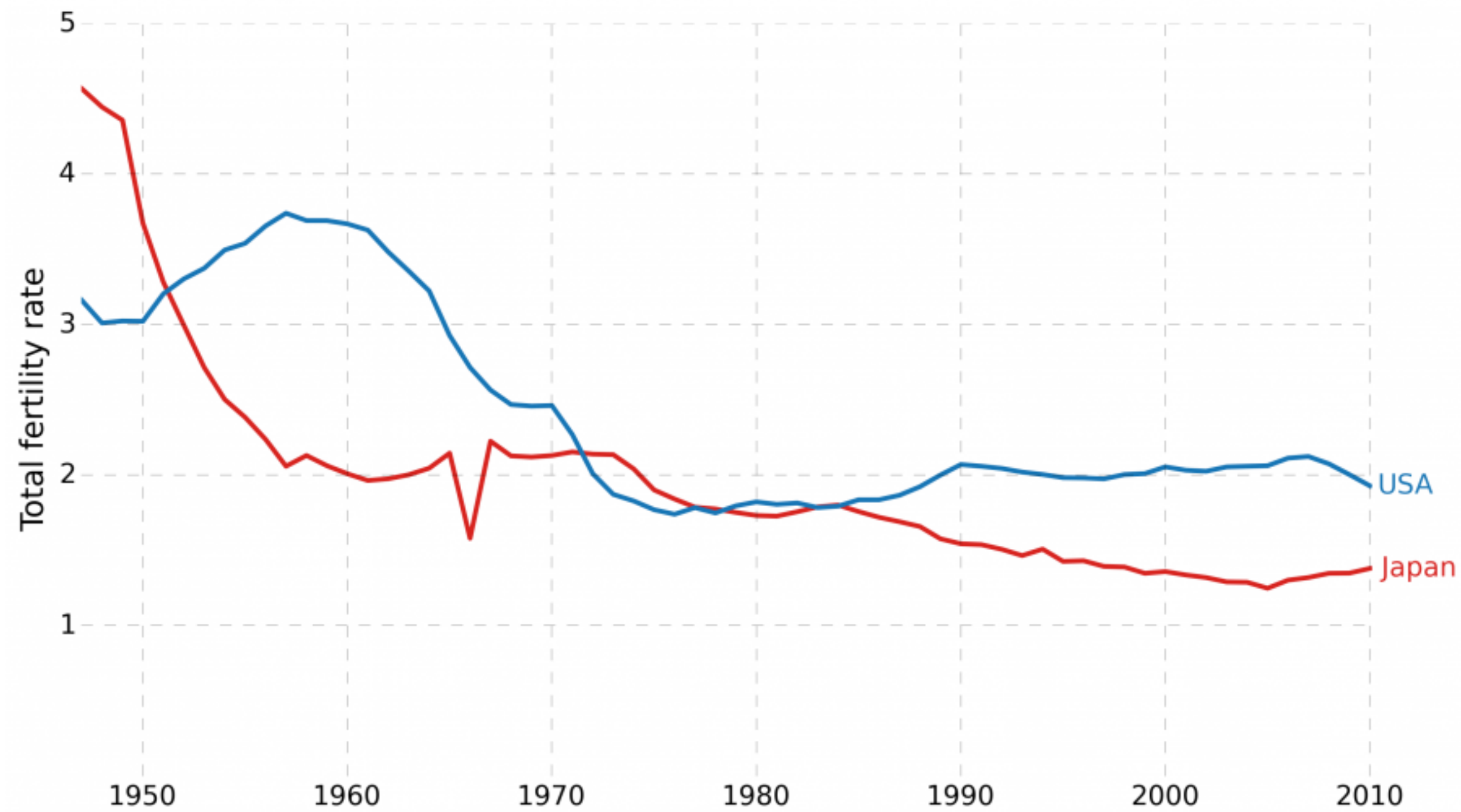
Do we need all of them?

Eyes Beat Memory: Small Multiples



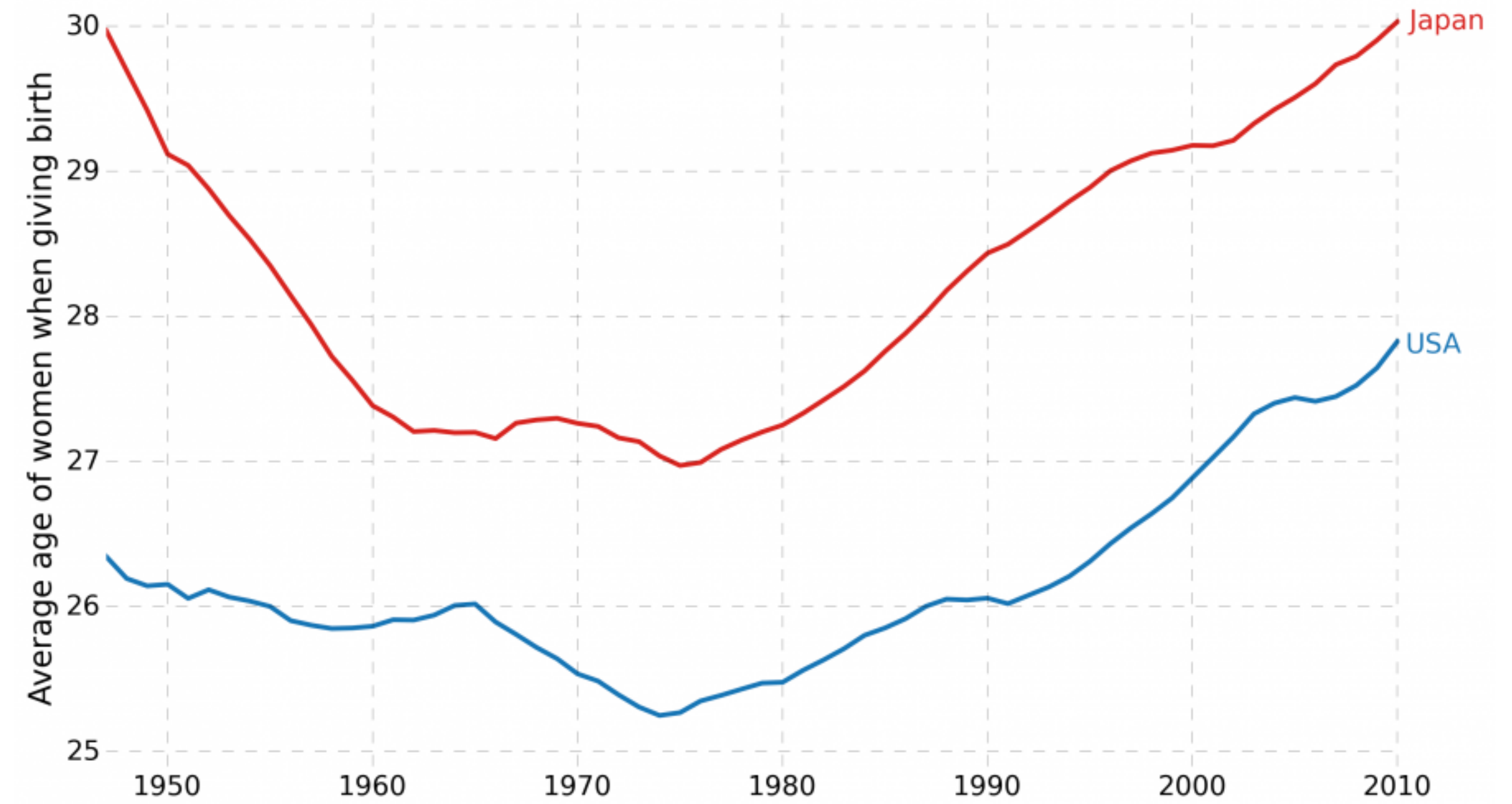
Simplify!

Total fertility rate in USA and Japan, 1947 - 2010



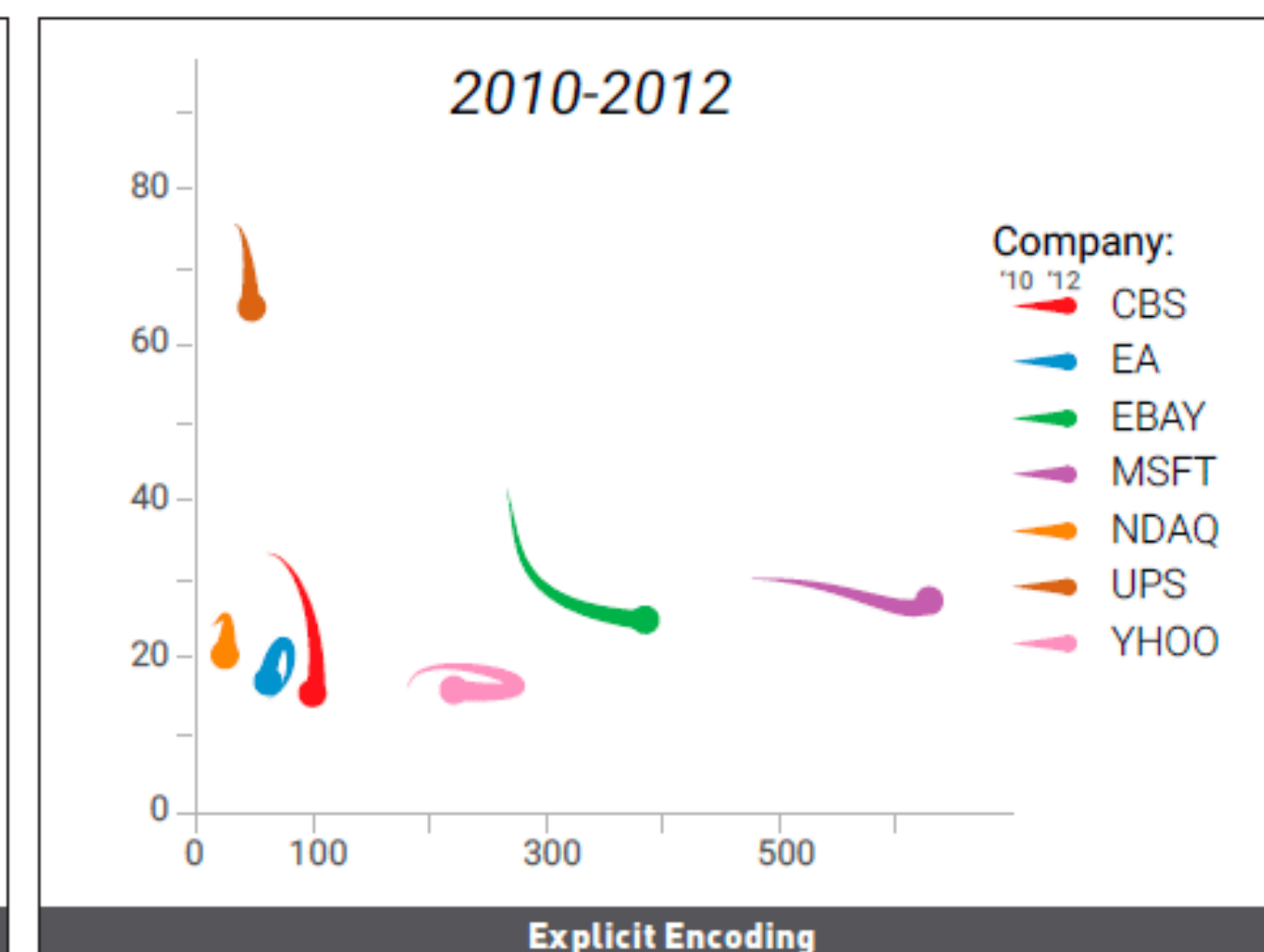
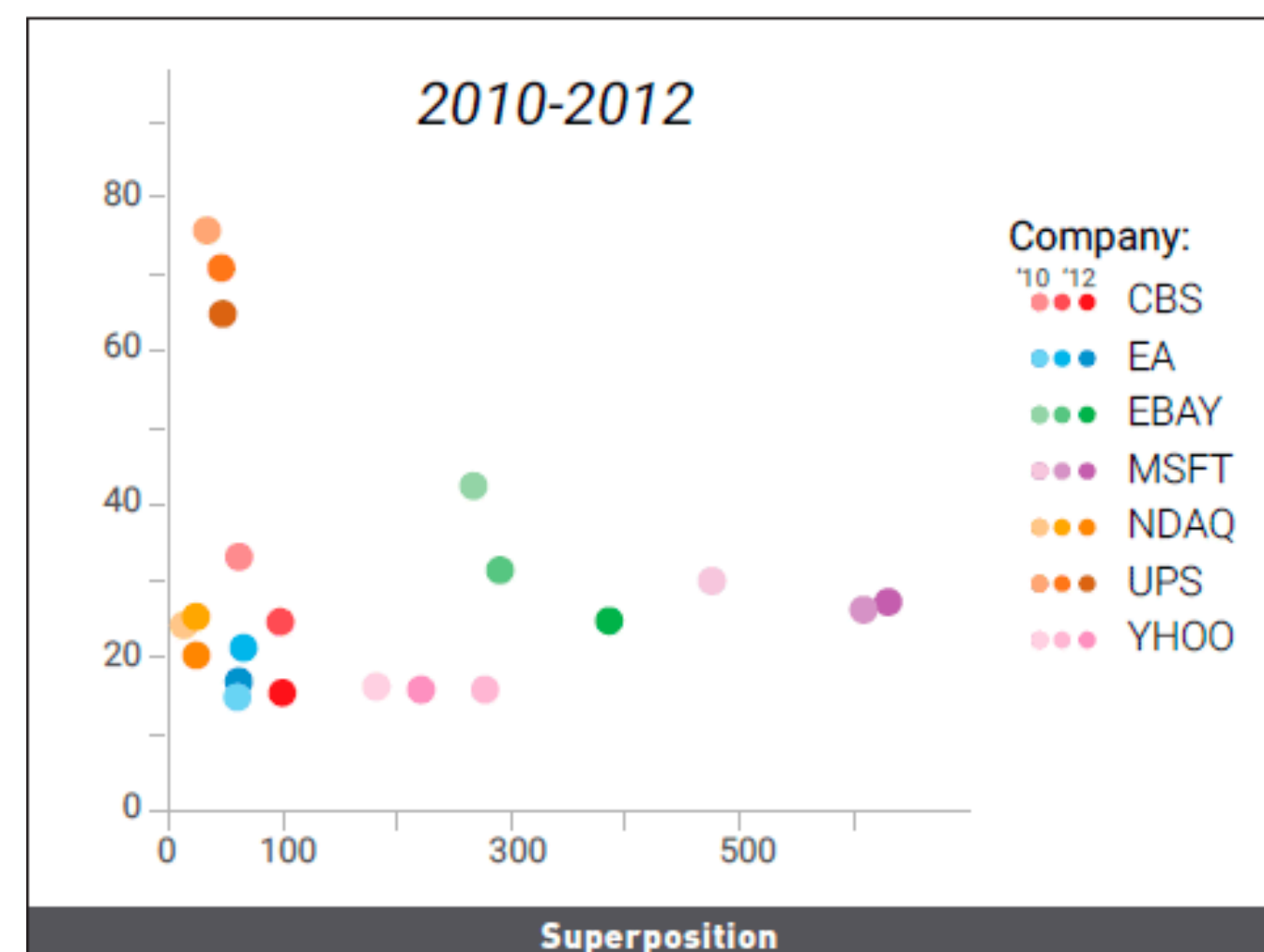
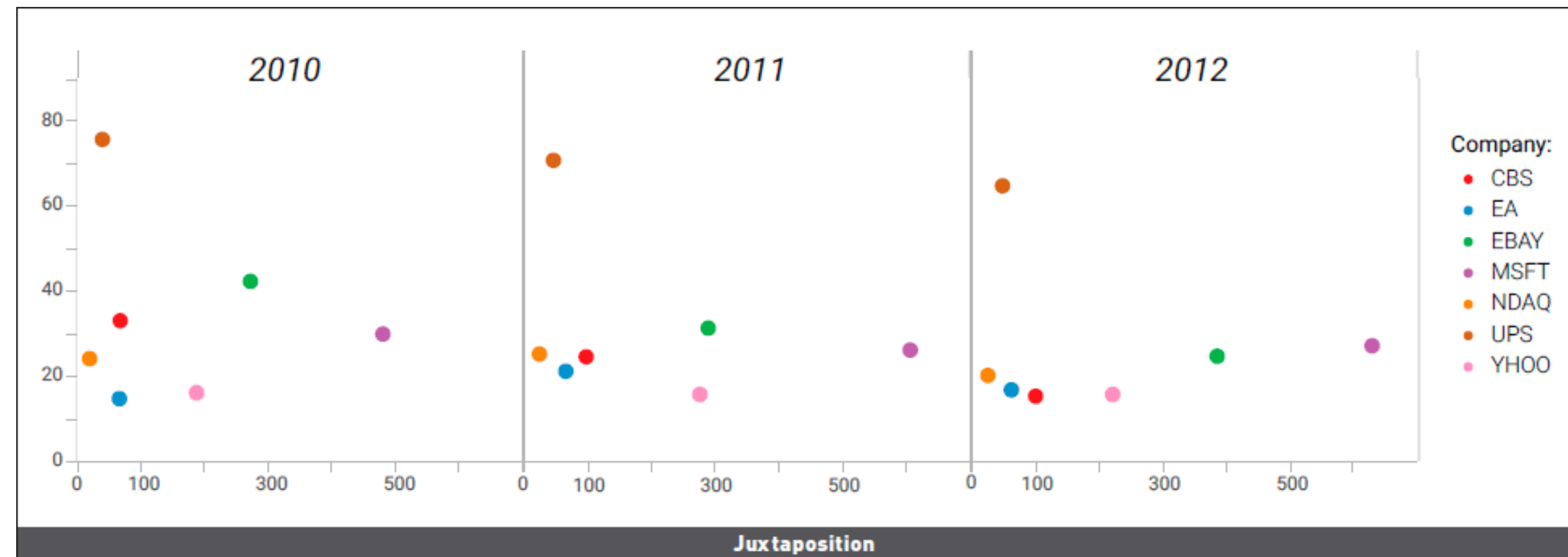
Data source: Human Fertility Database (humanfertility.org)
Author: Randy Olson (randalolson.com / @randal_olson)

Average age when giving birth in USA and Japan, 1947 - 2010



Data source: Human Fertility Database (humanfertility.org)
Author: Randy Olson (randalolson.com / @randal_olson)

Small Multiple Design Alternatives



Design and insights

Design choices shape insights

Chart being faithful to data is necessary but not sufficient

I show you a chart:

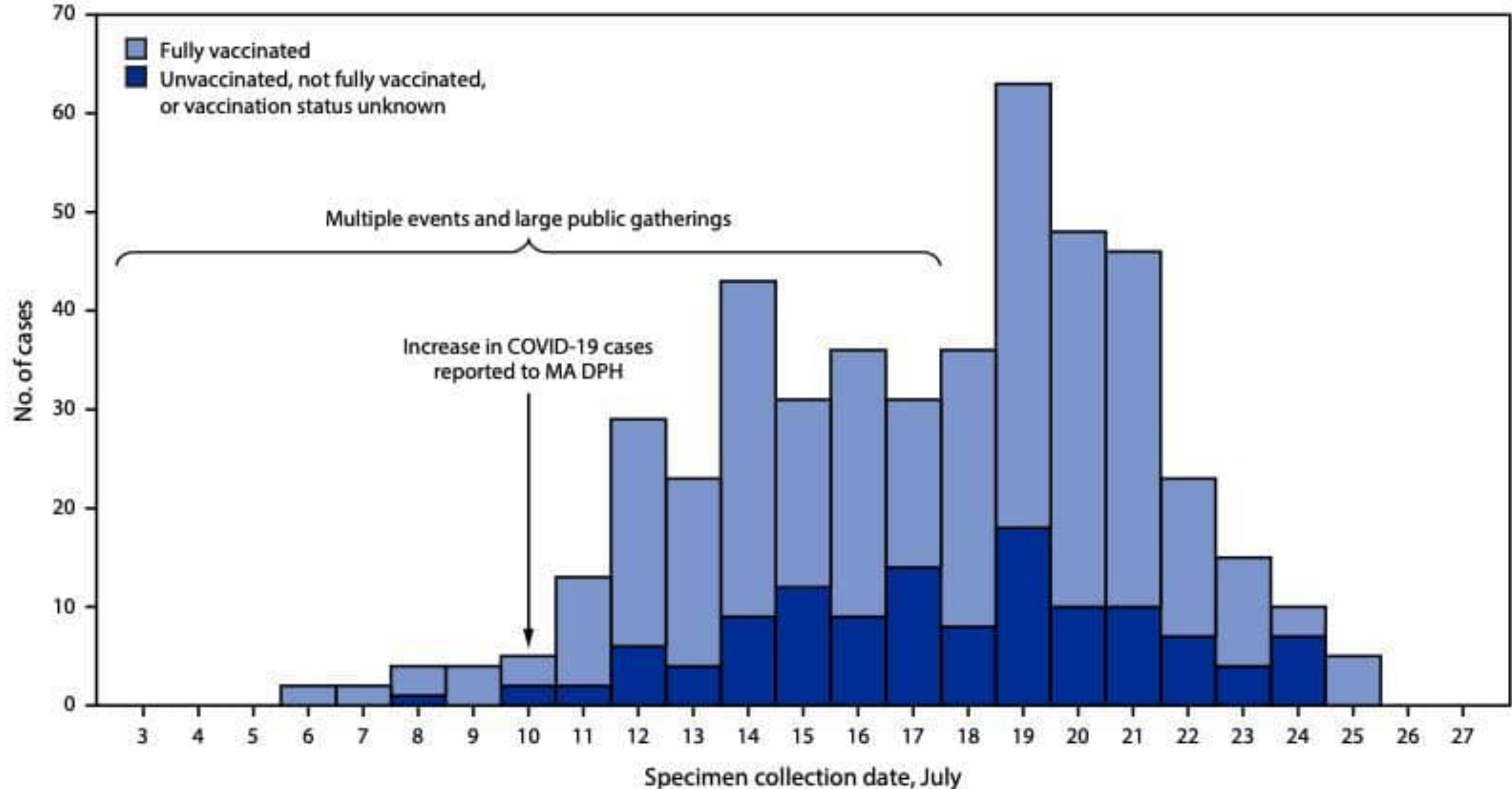
You tell me what you see

You are allowed to speculate and guess

You are allowed to “look outside the visualization” (i.e. use world knowledge)

CDC report

FIGURE 1. SARS-CoV-2 infections (N = 469) associated with large public gatherings, by date of specimen collection and vaccination status* — Barnstable County, Massachusetts, July 2021



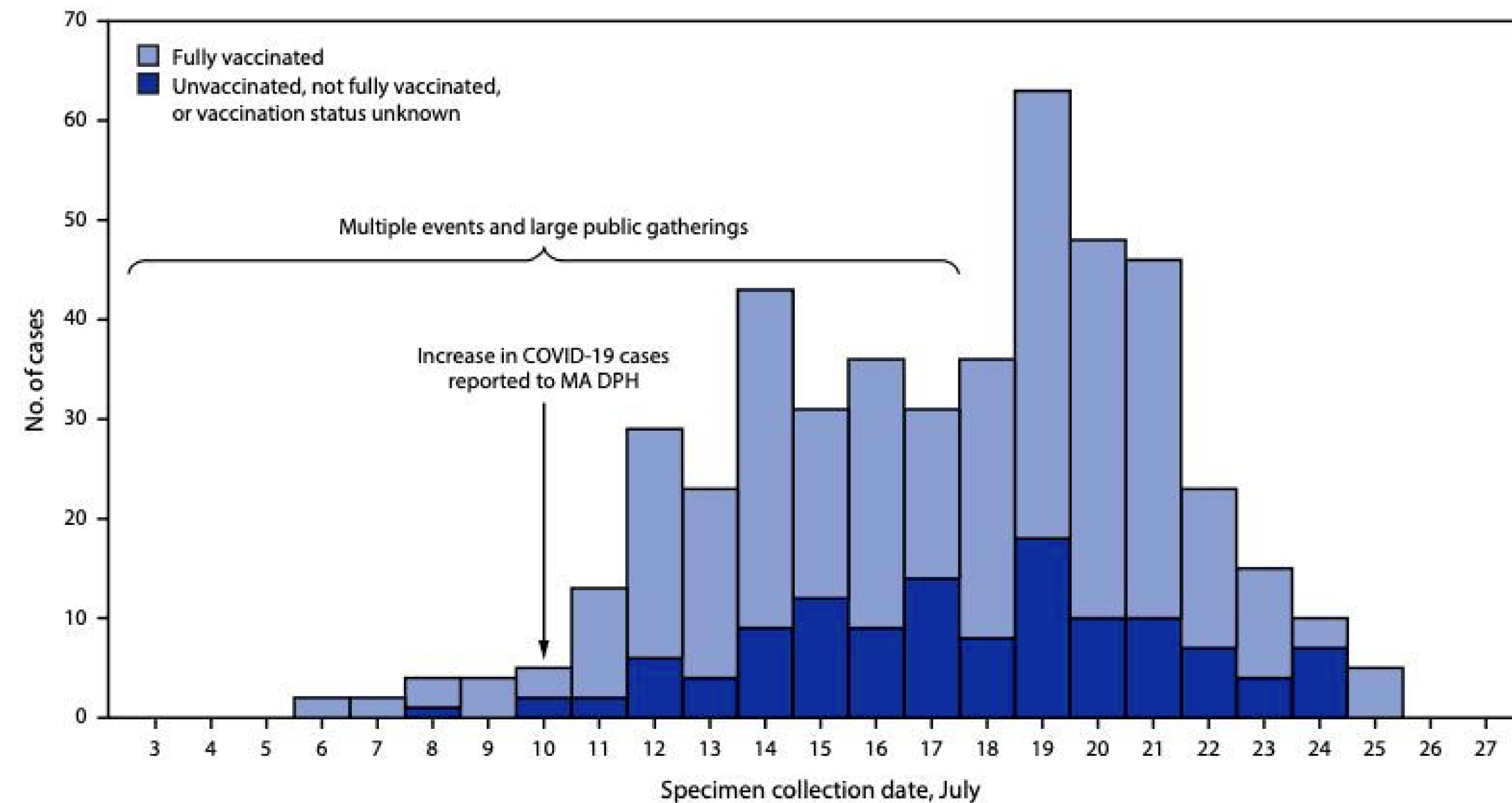
CDC report

According to official data, the vaccinated are the super-spreaders.

<https://cdc.gov/mmwr/volumes/7...>

<https://t.me/EARTH20GENESIS...>

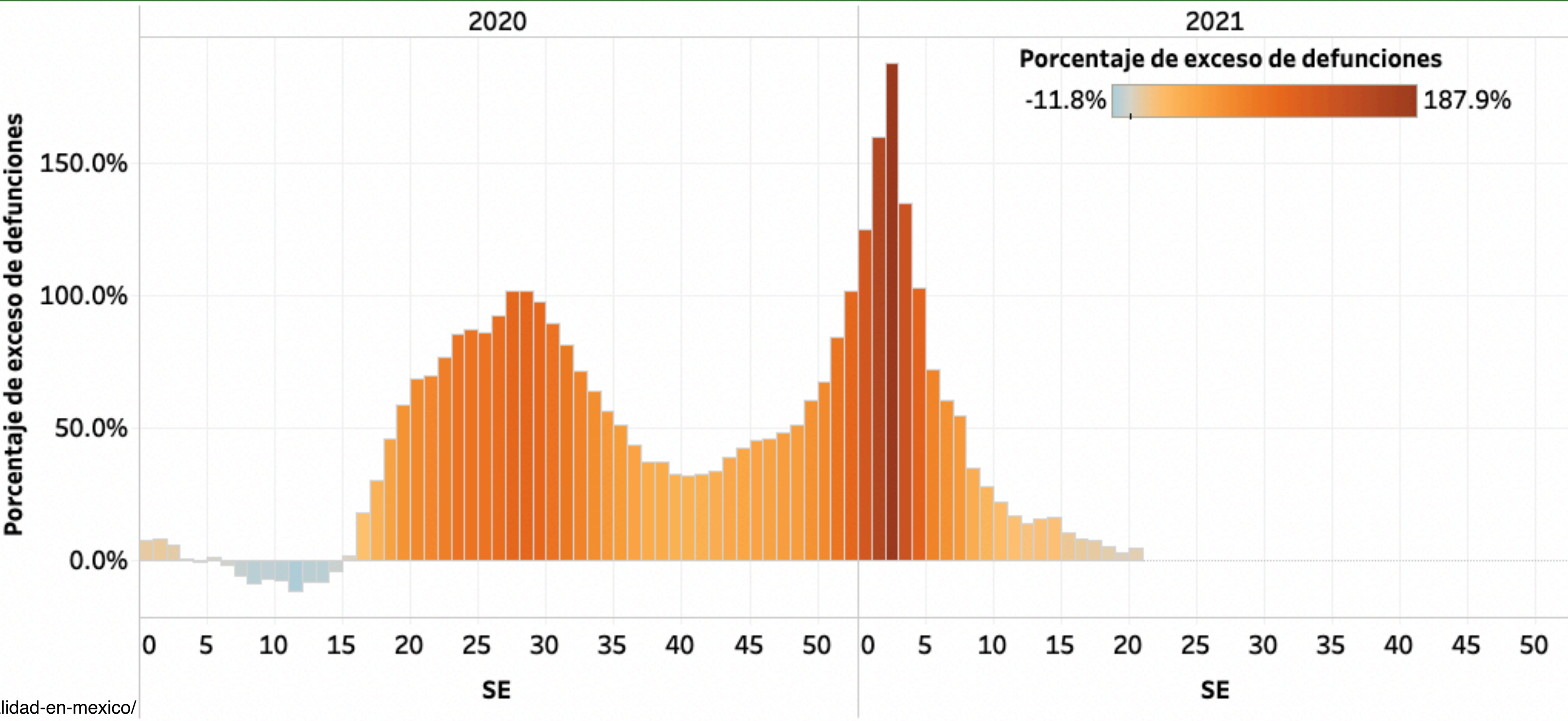
FIGURE 1. SARS-CoV-2 infections (N = 469) associated with large public gatherings, by date of specimen collection and vaccination status* — Barnstable County, Massachusetts, July 2021



Government dashboard (Mexico)

- Inicio
- Datos
- COVID-19
- Público general
- Personas con discapacidad
- Personal de salud
- Vacuna COVID-19
- Contacto
- Histórico

Porcentaje semanal de exceso de mortalidad por todas las causas 2020 - 2022 (Modelo)



Government dashboard (Mexico)

The effect of [#ivermectin](#) in [#Mexico](#) needs no explanation.

[#ivermectinworks](#)

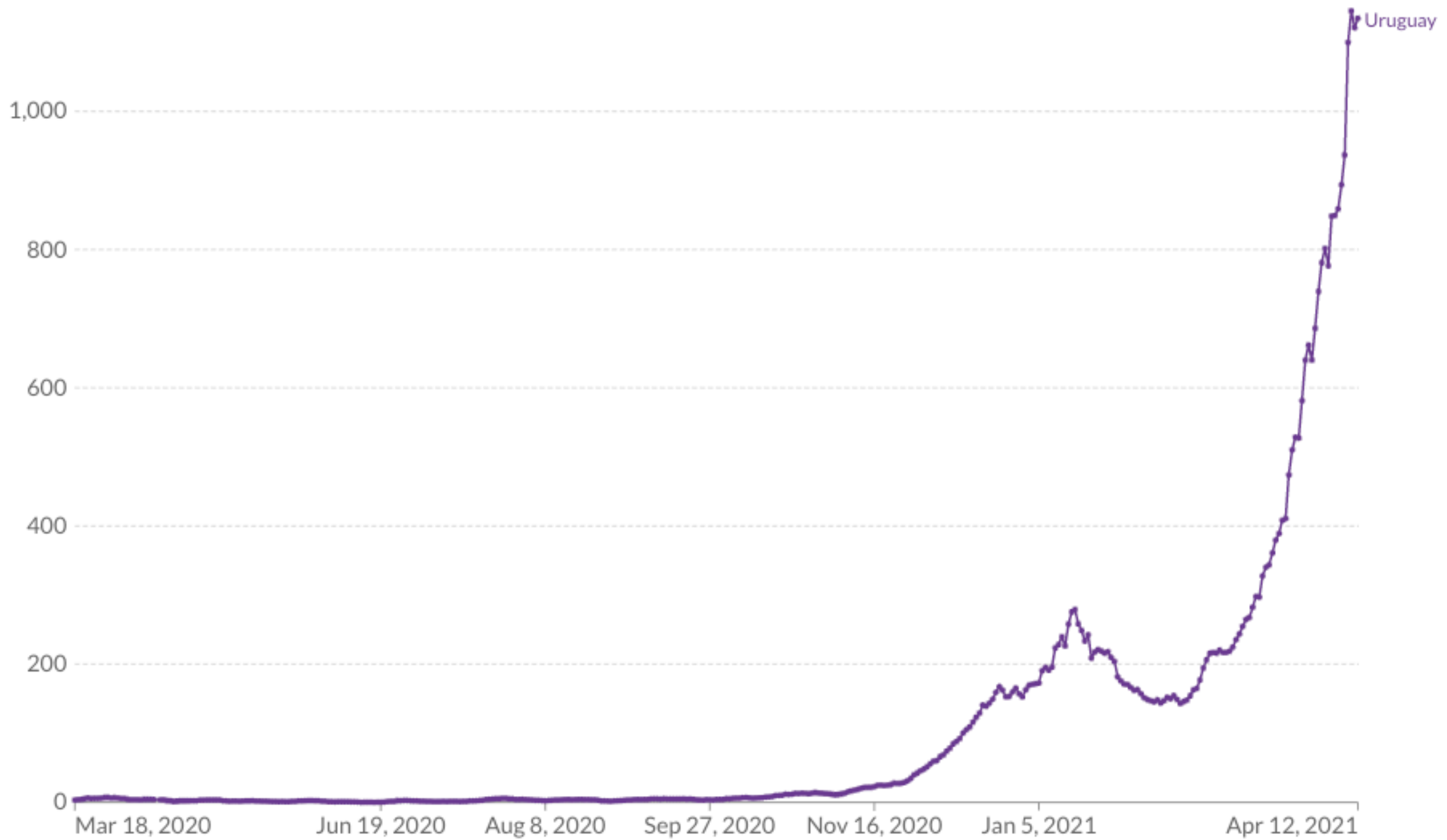


OurWorldInData

Daily new confirmed COVID-19 cases per million people
7-day rolling average. Due to limited testing, the number of confirmed cases is lower than the true number of infections.



LINEAR LOG



Source: Johns Hopkins University CSSE COVID-19 Data

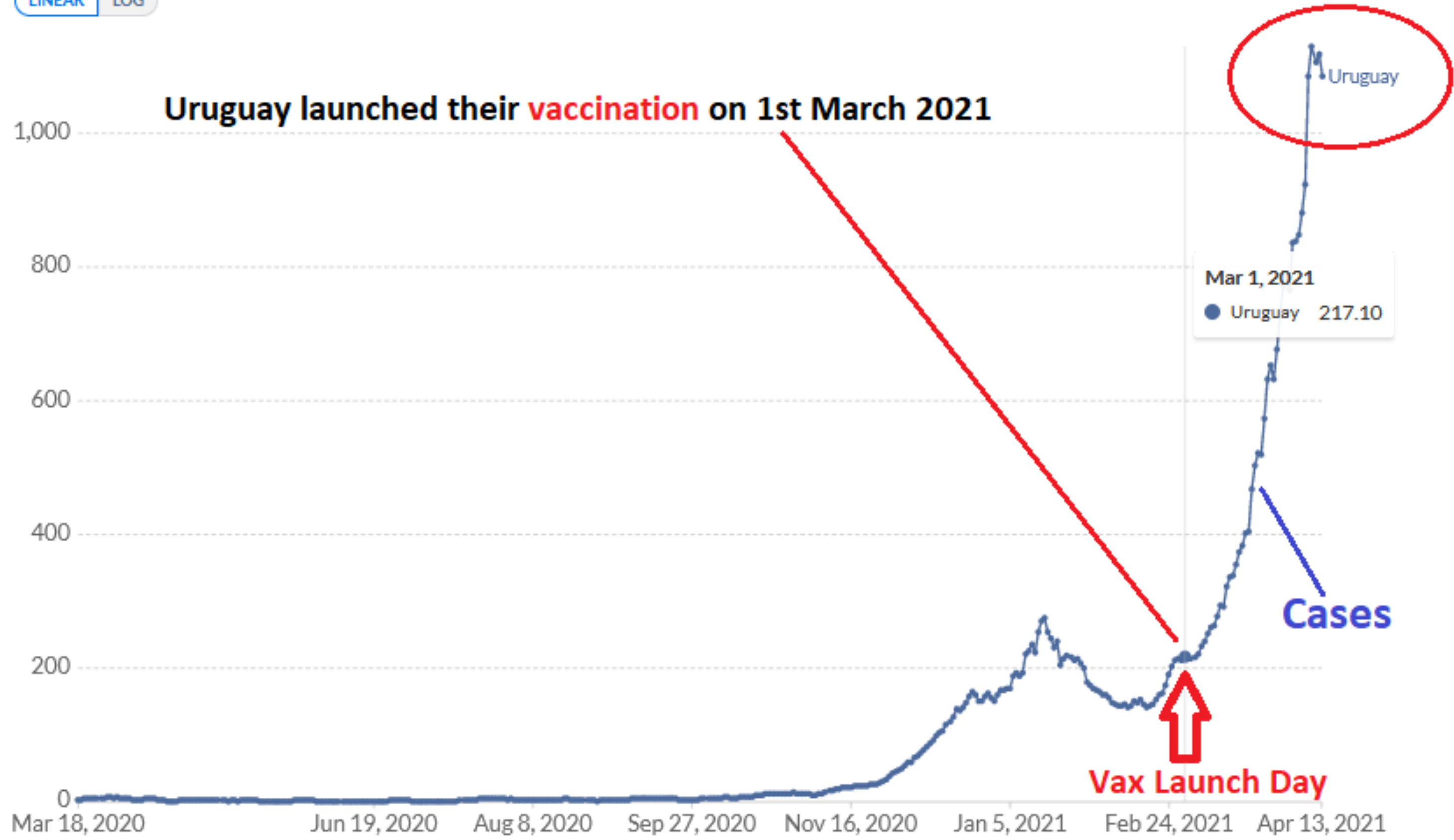
OurWorldInData

Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.



LINEAR LOG



Source: Johns Hopkins University CSSE COVID-19 Data

CC BY



CHART

MAP

TABLE

SOURCES

DOWNLOAD



What shapes our insights?

Major salient features: data, encoding

Framing: text, annotations

Context: world knowledge, expert knowledge

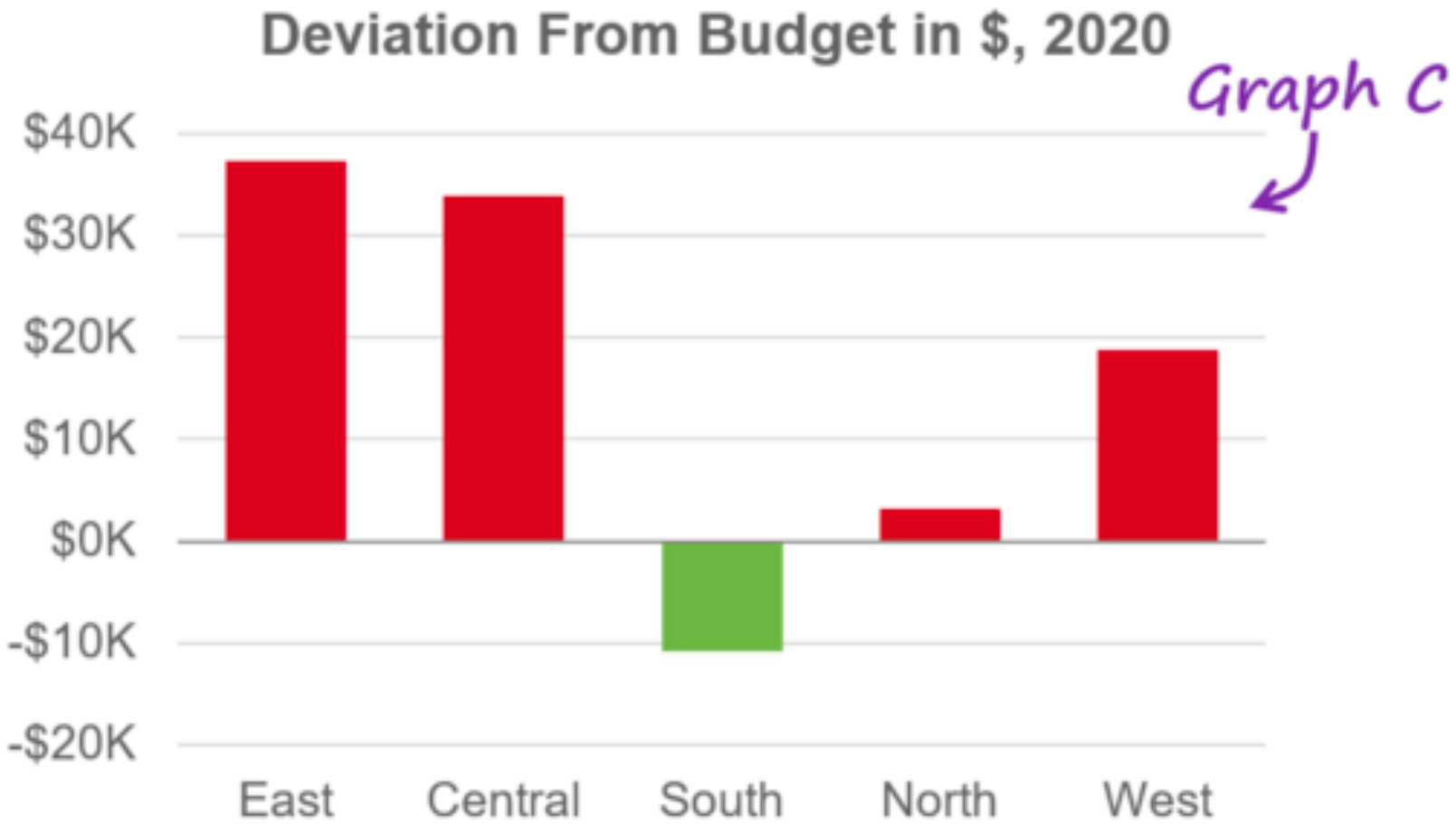
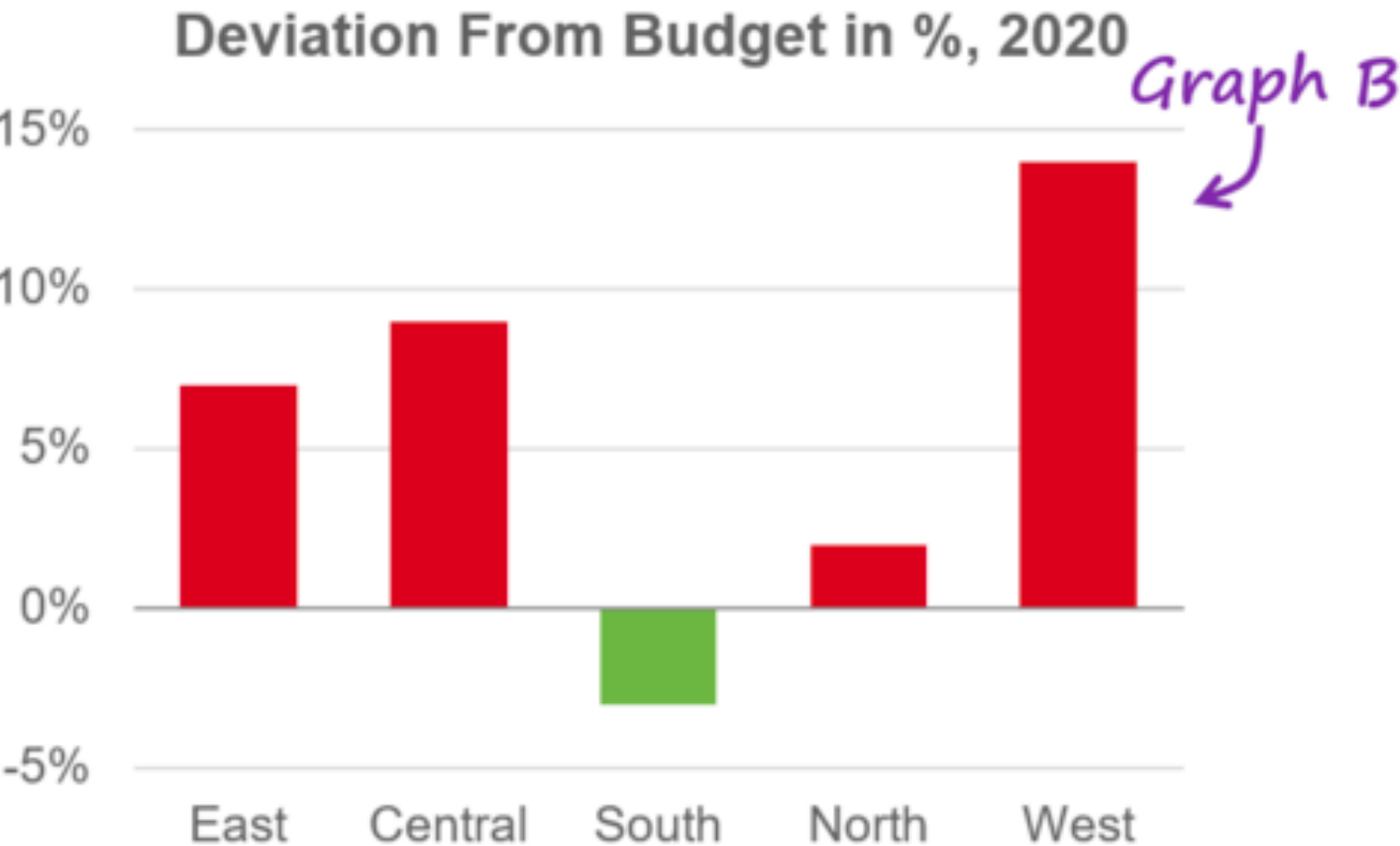
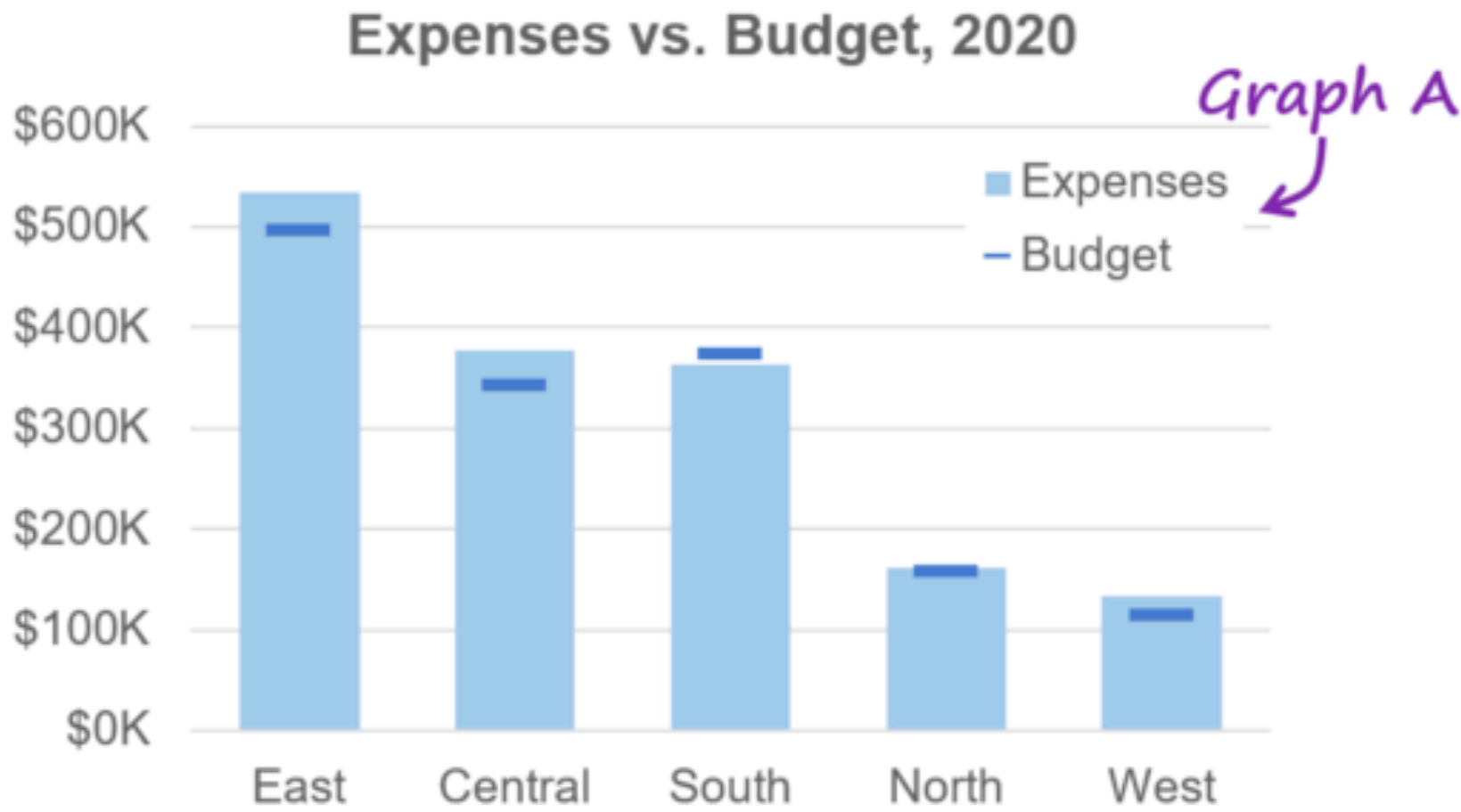
Personal biases

Neutral vs directive, narrative vs exploration

There's no one "best" way to visualize data...

Expenses vs. Budget, 2020

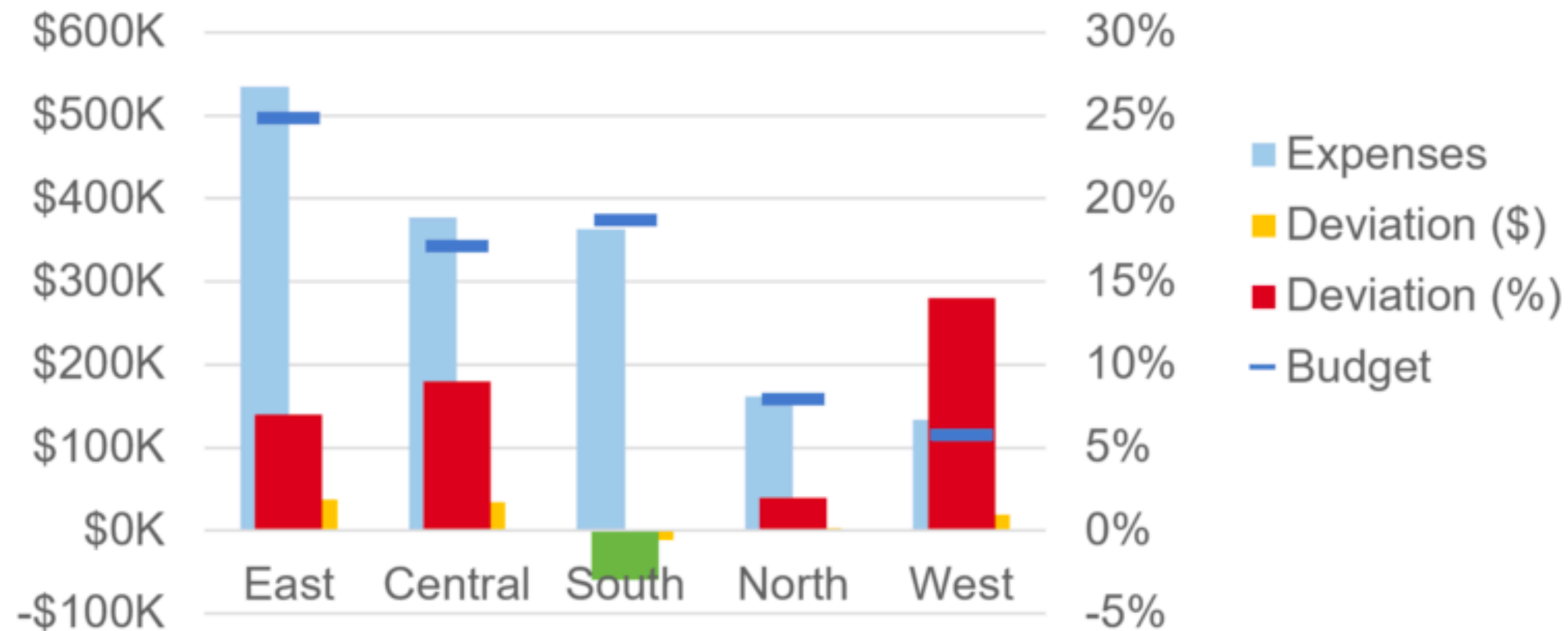
Region	Expenses	Budget
East	\$534,942	\$497,496
Central	\$377,123	\$343,182
South	\$363,122	\$374,016
North	\$161,994	\$158,754
West	\$133,800	\$115,068



“Charts never ‘show the data’, they always just say *a few specific things about the data.*”

...we really wish there was

Expenses vs. Budget, Deviation from Budget (\$), Deviation from Budget (%), 2020



Thread on design considerations

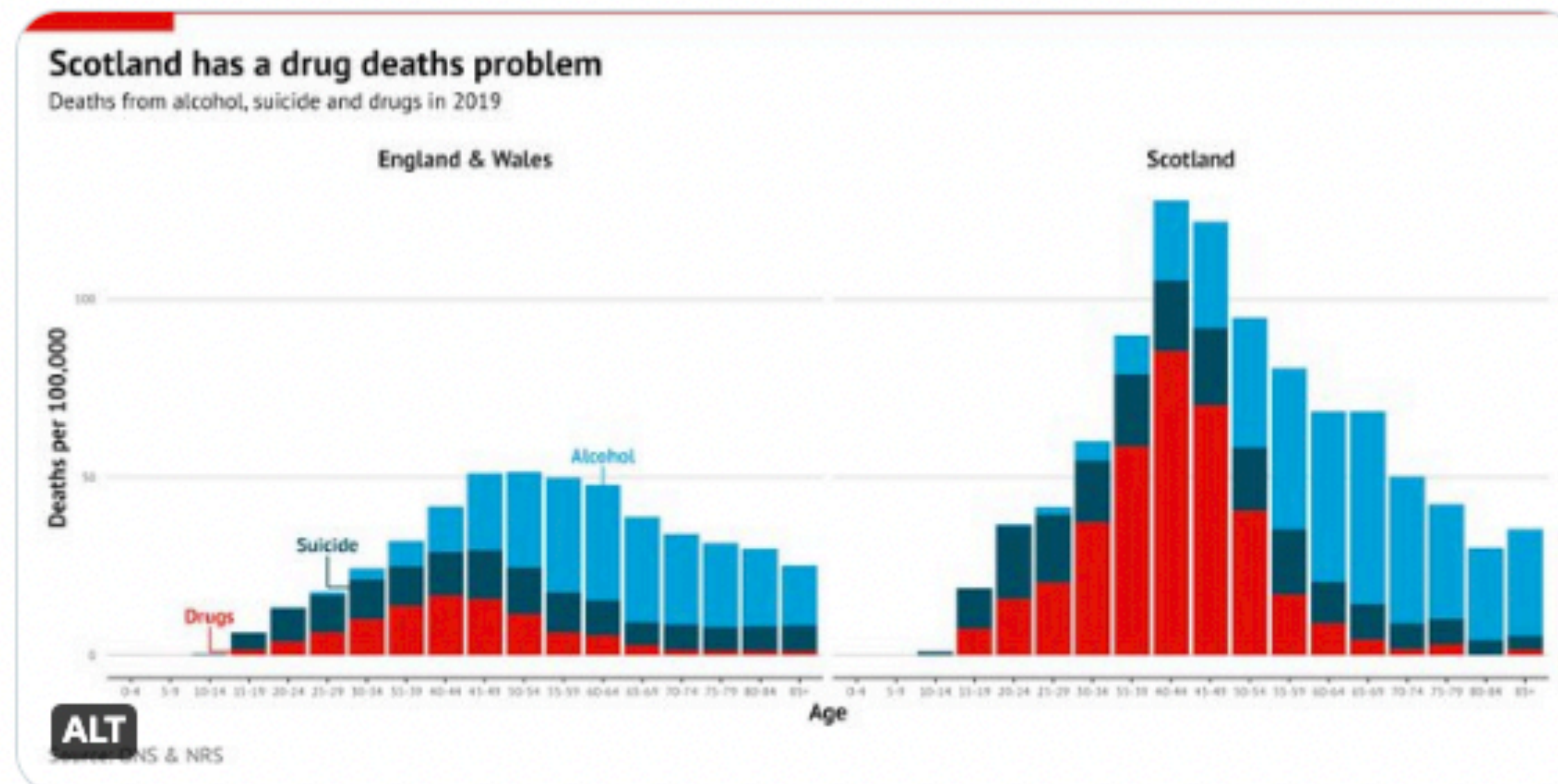


Colin Angus
@VictimOfMaths



I posted this graph yesterday as part of the [#30DayChartChallenge](#) and a few people have asked some reasonable questions about why I chose to present this data in this way.

So I thought I'd write a thread to explain my thought process...



“...people will always find their own messages in graphs, so why not help them find the 'right' one?”

Help find the 'right' message, steer away from the 'wrong' one

