

CSCI 1377

Tools for Thought

Visualization III

Expressive and Effective Visualization

“According to *Bartlett’s Quotations*, “a picture is worth 10,000 words” is a Chinese proverb. On inquiry, we find that the Chinese seem not to have heard of it, but the proverb is certainly widely known and widely believed in our culture.”

— Larkin and Simon, “Why a Diagram is (Sometimes) Worth Ten Thousand Words” (1987)

- 🔍

Search
- 📝

New chat
- 📁

Projects
- 🛠️

Artifacts
- 🕒

Scheduled
- ⚙️

Customize
- ⋮

More

- Today
- What makes a text readable?
- Notes on a weekend walk
- Planning a reading routine
- Yesterday
- A short story outline
- Comparing two recipes
- Feedback on a cover letter
- Previous 7 days
- Garden planting calendar
- A gentler morning routine
- Ideas for a neighborhood event
- Explaining orbital motion
- Choosing a film for Friday

Cross-examine

✕

Claude's answer · 5 claims, 0 reviewed

Comprehension Question 2

- 1

Supported

Readability depends on the reader's prior knowledge.

Found across 3 passages

The laundry example, situation model, and readability discussion are relevant to this claim. Read them together before deciding whether they support your wording.

Accept

Revise

Cut
- 2

Your call

Context helps the reader interpret an otherwise unclear description.
- 3

Supported

A reader builds a situation model rather than retaining only the wording.
- 4

Your call

Familiarity with the activity can make the laundry passage easier to understand.
- 5

Your call

The same passage may be more readable for one reader than for another.

Reply to Claude...

+

Sonnet 4.5

▼

↑

Understanding text

252

Found across 3 passages

Laundry passage · p. 252

you understand the passage (given the title Laundry) because the title allows you to retrieve a schema from your real-world knowledge for doing laundry.

Situation model · p. 269

Many theorists argue that constructing a mental representation of the situation or scenario that is described is the principal goal of reading an expository text.

Readability · p. 273

if one wants to measure comprehension, it cannot be done in a vacuum; one needs a theory of what comprehension is

The decision was made more quickly following the passage in 9.27a than following the similar one in 9.27b. Apparently readers found a way to make a causal bridge between the two sentences. When the first concerned dental health, the inference that is made is that Terry called the dentist to make an appointment. However, when the first sentence concerned dental bills, the inference is that Terry called the dentist to complain. Making the bridge in the former case activated pre-existing knowledge that dentists require appointments, which in turn speeded verification.

Elaborative inferences

1 Sometimes inferences are made even before they are needed to make a text coherent. As an example, Garrod and Sanford (1982) had people read one of the sentences in 9.28a followed by 9.28b:

- 2

Explicit: Keith took his car to London.
Implicit: Keith drove to London.

Test: The car kept overheating.

(9.28a)

(9.28b)

The test sentence was read equally fast in both cases. This contrasts with the Clark and Haviland case (*The beer was warm*) discussed earlier. Apparently, the discourse representation constructed in the implicit version of 9.28a included a car (while the picnic supplies did not necessarily include beer). Still, since no car was explicitly mentioned in the text, its existence required an inference—and since its existence was not needed to make the sentence *Keith drove to London* coherent, it could not be considered a bridging inference. It is generally called an elaborative inference.

Such elaborative inferences are not always made. As indicated before, there may just be too many of them. But it is of interest to examine the conditions under which they are made. Consider an example discussed earlier, from Garrod and Sanford (1977), and shown in 9.13 (reproduced again below):

A robin/goose would sometimes wander into the house. The bird was attracted by the larder. (9.13)

- 3

The reading time for the second sentence was longer when the preceding sentence contained the atypical *goose* than when it contained the typical *robin*, suggesting that the typical term was more likely to trigger an inference to the superordinate concept of bird. In addition, reversing the order of the specific and general terms, as in 9.29, increased reading time for the second sentence.

A bird would sometimes wander into the house. The robin/goose was attracted by the larder. (9.29)

It appears that reading *robin* or *goose* in 9.29 requires that the reader add some specific information to the representation of the text, slowing reading compared to 9.13. Reading a word denoting a specific concept may not generally activate a specific example of the category, even a highly typical one (although reading was faster for *robin* than for *goose* in 9.29, suggesting that the inference that it is the bird previously referred to was easier for the more typical instance). However, if the context is constraining enough, readers may make an elaborative inference from a general category to a specific example (Garrod, O'Brien, Morris, & Rayner, 1990). Participants

CONTEXTUALIZE

Some LLM

This is a **constrained instrumental inference**: it goes beyond the text but isn't needed for coherence, since "Keith drove to London" makes perfect sense without picturing any particular vehicle. Readers usually don't make such inferences in advance, so the explanation is that this one is highly constrained. Driving almost always involves a car, so the instrument is nearly part of the verb's meaning, much like stabbed strongly suggesting a knife in the O'Brien et al. (1988) mugger passage.

LEARN

× Clear

Learn: constrained instrumental inference

- Paraphrase**
- This exact term isn't in the text. The author calls this an **elaborative inference**

The text describes readers adding details the sentence doesn't state, like a car for 'drove.' It calls these elaborative inferences, and asks whether readers make them while reading or only when needed.

Steps

- 1

Read the definition of elaborative inferences.
Elaborative inferences (heading)

Jump to passage →
- 2

Study the car example to see the inference in action.
Example 9.28

Jump to passage →
- 3

Read the evidence on when these inferences are made.
Example 9.29 and following paragraph

Jump to passage →

This idea was a paraphrase of 'elaborative inference.'

QUERY *Who spilled the coffee? What colors were the stains on the essay? For each question, use concepts from Chapter 9 to describe the sequence of inferences required to arrive at the answer.*

≡ scenario · Ditto, Eleanor, Will, Jinho · 9 sentences

GEMINI To determine who spilled the coffee, the reader must resolve the pronominal anaphora for the pronoun “He” in the sentence “He spilled black coffee on the laptop”. The preceding sentence states that “Will made lecture slides at the desk, and Eleanor watered the plants next to him.” The reader relies on morphological marking for gender to filter the potential referents; since Eleanor is female (subsequently confirmed by the pronoun “She”) and Ditto is neuter (subsequently referred to as “it”), Will is the only available male antecedent. Furthermore, centering theory posits that “He” acts as a backward-looking center pointing to a high-ranking forward-looking center from the previous sentence. Because Will was the active agent making lecture slides and the immediate referent of the object pronoun “him,” he remains highly salient in the reader’s working memory, allowing the reader to instantly infer that Will spilled the coffee.

Determining the colors of the stains requires the reader to construct a situation model that integrates multiple propositions across the text using causal bridging inferences. When the text states that Jinho noticed the stains, the reader must search backward to identify the cause of the event. The reader makes a bridging inference connecting the newly introduced stains to the preceding action where Ditto used the essay “to dab at the spill,” and then uses nominal anaphora to link “the spill” back to the previously mentioned “black coffee”. This causal chain allows the reader to infer that the spilled coffee created a black stain on the paper.

Simultaneously, the reader must retrieve the earlier proposition that “Ditto graded an essay with a red marker” and incorporate it into the current situation model. Understanding that red marker ink will bleed or mix when a

IN THE CHAPTER rayner_ch9.pdf · p. 246

and his colleagues developed over a long period of time (Kintsch, 1974, 1988, 1994, 1998, Kintsch & van Dijk, 1978, van Dijk & Kintsch, 1983). One of the primary contributions of this theory was to distinguish between different mental representations of a text. These representations are called the surface representation, the propositional representation, and the situation model (Kintsch, 1994). The previous chapter focused on the first two representations, discussing how a reader goes from the printed text to a parsed representation of the form of a sentence (the surface representation) and then to a meaningful interpretation of the sentence (the propositional representation). Text and discourse processing researchers concentrate on how a reader puts the propositional representations of sentences in a text together to arrive at a representation of the situation that the text is describing (the situation model). In some versions of this theory, a distinction is made between "microstructure" and "macrostructure." Microstructure is a level of representation that puts the propositions together into a connected structure; macrostructure is the gist of the text, an edited version of the microstructure tied into schemas from long-term memory.

The typical way of representing a proposition, in approaches like Kintsch's, is as a predicate followed by its arguments. We will illustrate how this is done by sketching part of an analysis of the paragraph shown in Table 9.1. A predicate can be represented by the meaning of a verb (where meaningful concepts rather than words are indicated by UPPERCASE) and the arguments can be

first mention

definition

[open page ↗](#)

The representation of the situation the text is describing.

Will is the most plausible person who spilled the coffee: the reader resolves “him” in “next to him” as Will at the desk, then carries that referent into “He spilled black coffee.” This is pronominal anaphora, supported by the discourse and office context; it is not a rule that a pronoun must refer to the nearest name.

The reader also builds a situation model with Ditto on the couch, Will at the desk, Eleanor nearby, and the laptop involved in the spill. Chapter 9 distinguishes that model of people, objects, locations, and events from the surface wording, which helps explain why context matters when deciding who “He” refers to.

For the stain colors, the reader tracks the same graded essay from Ditto’s grading, through dabbing the spill, to Jinho’s collection, and resolves the final “He” as Jinho. That causal link requires a **bridging inference**: the reader connects the later stains to the earlier cleanup by supplying the unstated transfer of black coffee onto the essay. [note](#)

Ask about these words

Close

bridging inference

What is a causal bridging inference?

Ask

Show the source

erling a
t it refers

clicking another
term brings up a
new source context,
letting you arrange
your ideas spatially
on the canvas like a
mind map.

this sounds right to me
ow, but maybe I
ould add my
efinition to make it
ore detailed?

LLM Answer

Will spilled the coffee. To comprehend this, the reader must resolve pronominal anaphora when encountering the sentence beginning with "He spilled...". The reader uses morphological gender marking to disambiguate the referent, seamlessly mapping the male pronoun "He" to Will rather than the female character Eleanor. Structural factors described by centering theory also facilitate this inference; Will serves as the topic and a high-ranking forward-looking center in the preceding clause ("Will made lecture slides"), making the pronoun "He" an expected and easily processed backward-looking center. Finally, the reader makes a bridging inference to connect the definite noun phrase "the laptop" to Will's previously established location at "the desk," constructing a spatially coherent situation model of the event.

Source Context

Context for "bridging inference"

p. 263 - Definition of bridging inference

open page
in PDF

A bridging inference is made when something has to be added to the discourse model to tie a new proposition into the previous model.

p. 263 - Example of bridging inference

open page
in PDF

Anaphoric inference is an example of a bridging inference; a reader has to figure out who he refers to in John praised Bill because he caught the pass in order to make sense of that sentence. For another example, our discussion of Haviland and Clark (1974) presented the classic example of a bridging inference.

p. 264 - Example of bridging inference

open page
in PDF

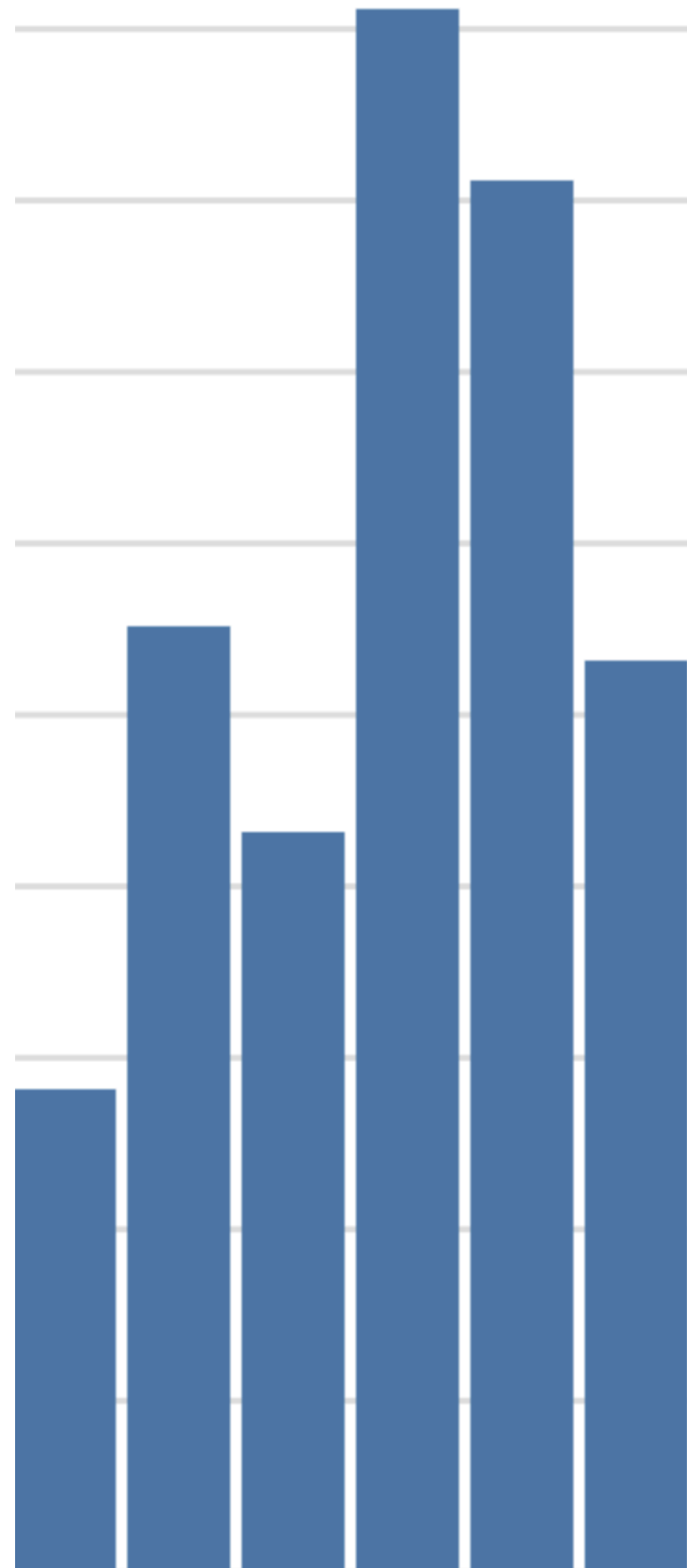
Horace got some beer out of the trunk. The beer was warm. (9.24a)

Horace got some picnic supplies out of the trunk. The beer was warm. (9.24b)

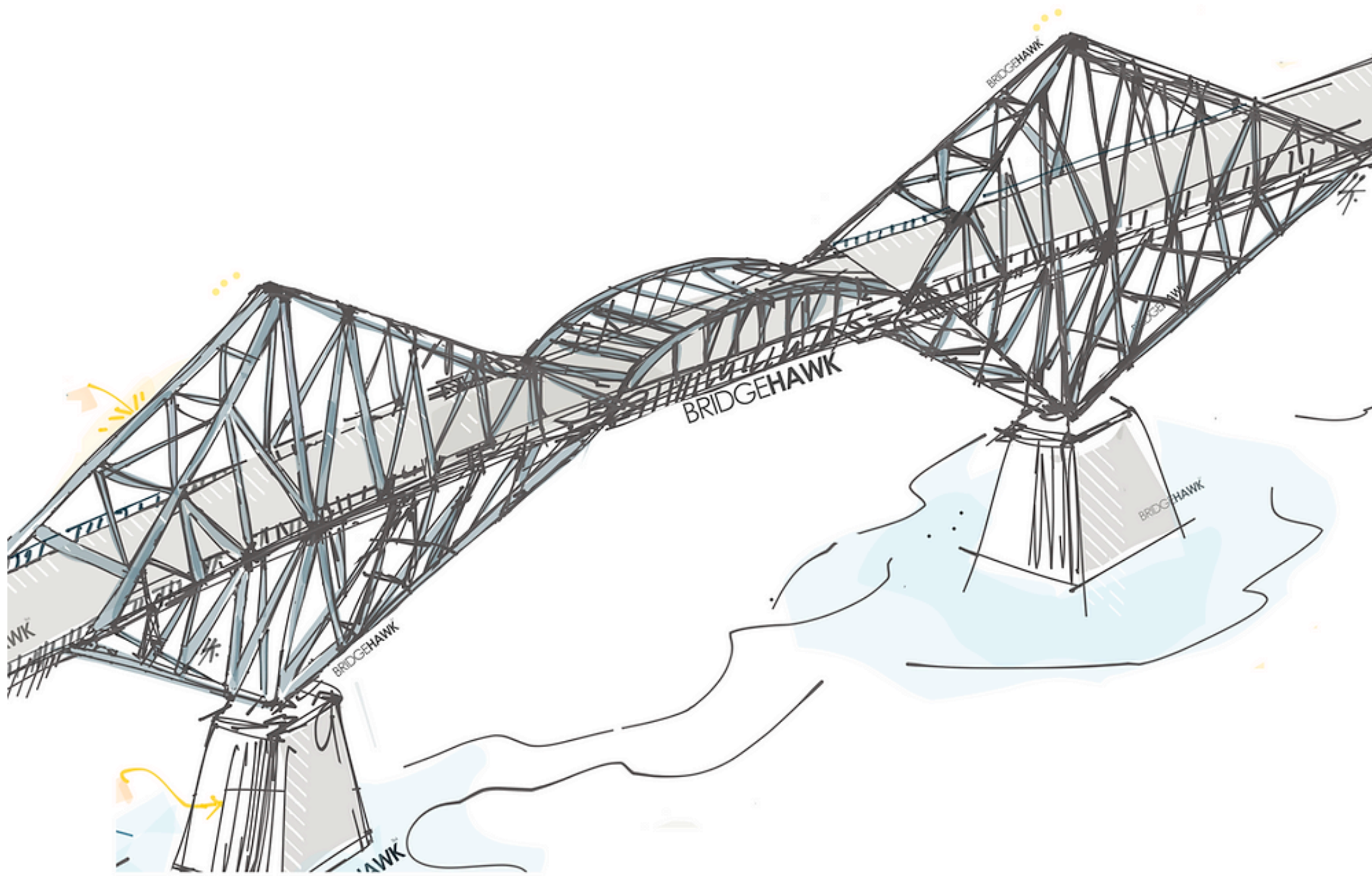
This definition is
confusing, but I th
bridging the gap
between a definite
and a previous ph

This example explain
why "laptop" would
connect to "the desk"

Which of these visualizations is better?

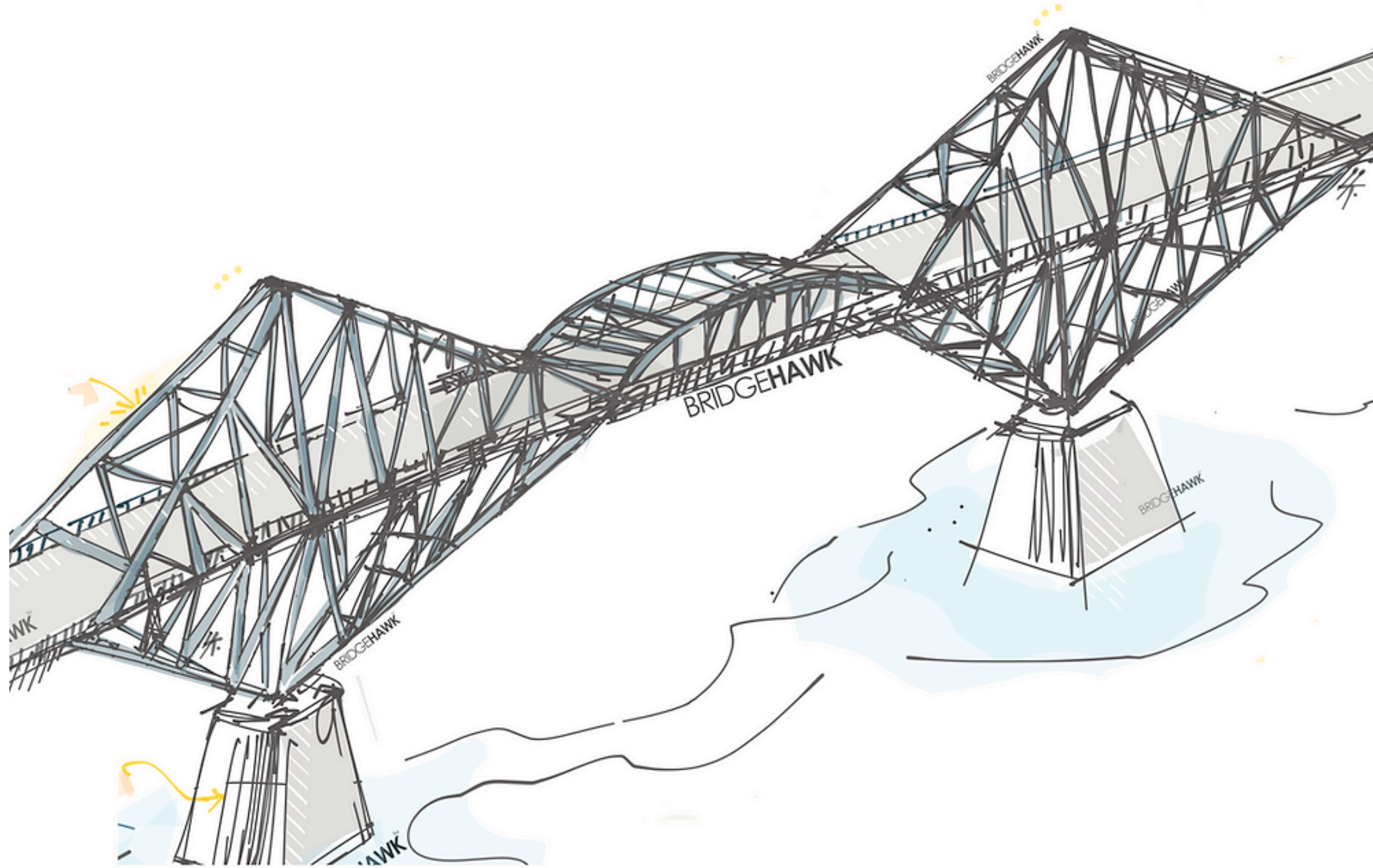


Which of these bridges is better?



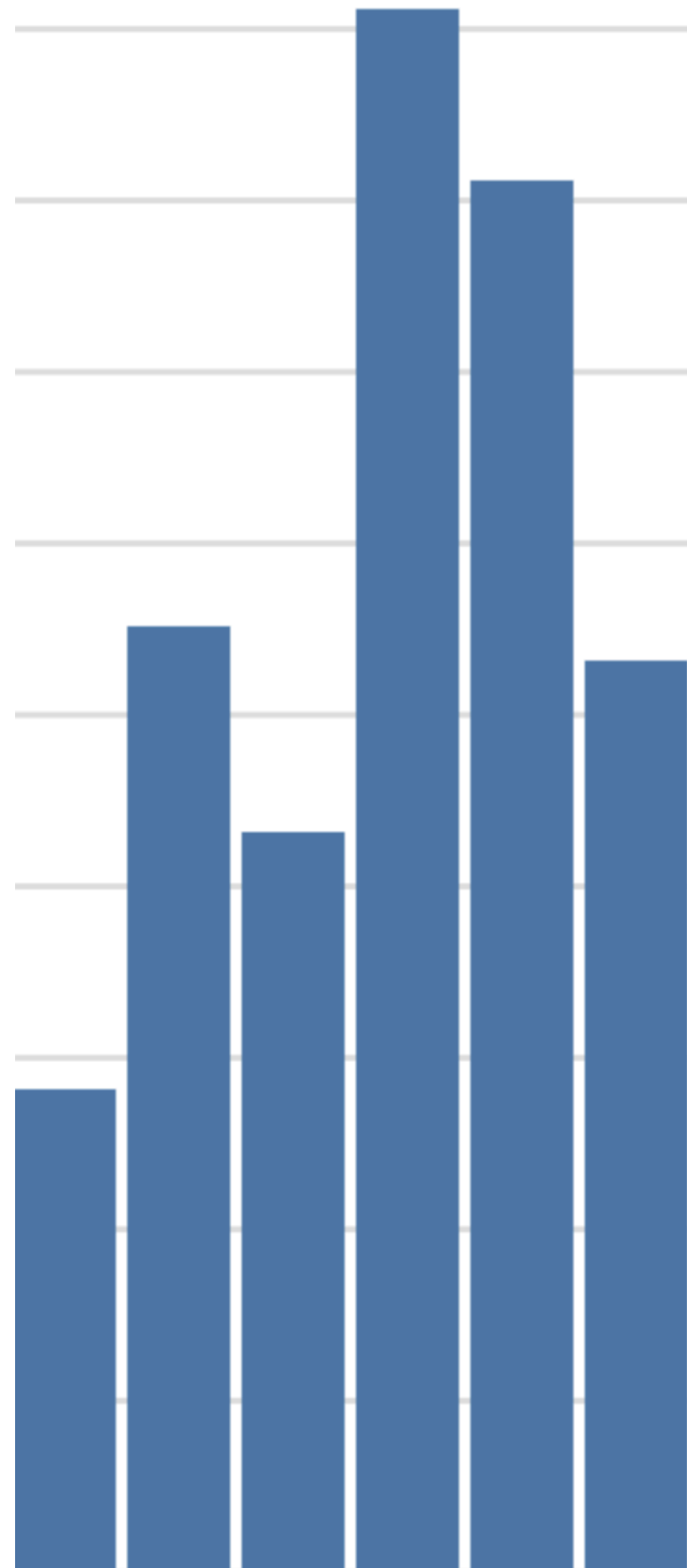
~~Which of these bridges is better?~~

Bad question!



How many cars can it carry? How much will it cost to build and maintain? Where are we building it? How well will it take the local weather?

Which of these visualizations is better?

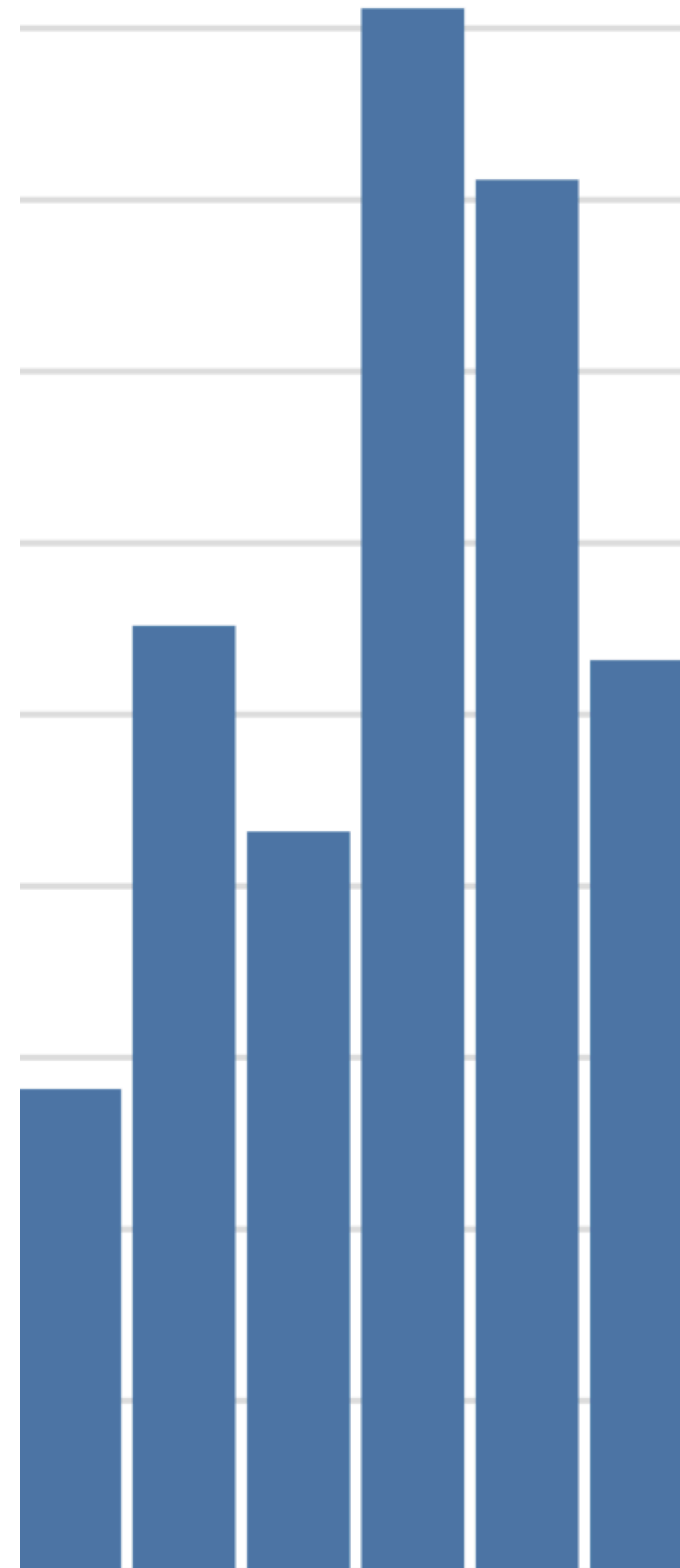


~~Which of these visualizations is better?~~

Bad question!

Which visualization
helps **people**
quickly see
information about
the data?

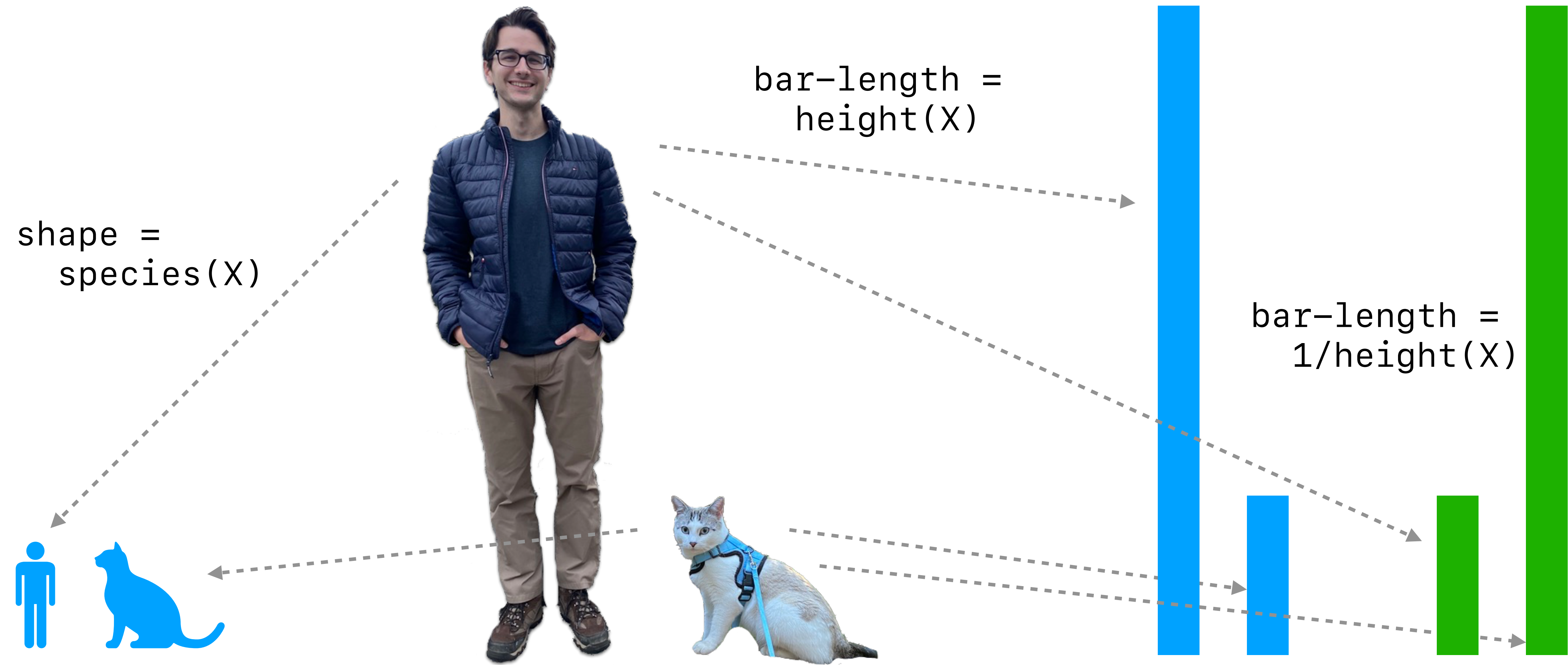
Which **Which** **See**
people? **information?** **how?**



Representations and expressiveness

Which visualization helps people quickly see
information about the data?

Representational systems map information from the world to a representation



Good representational systems preserve relations between domains

$\text{height}(X)$
 $> \text{height}(Y)$

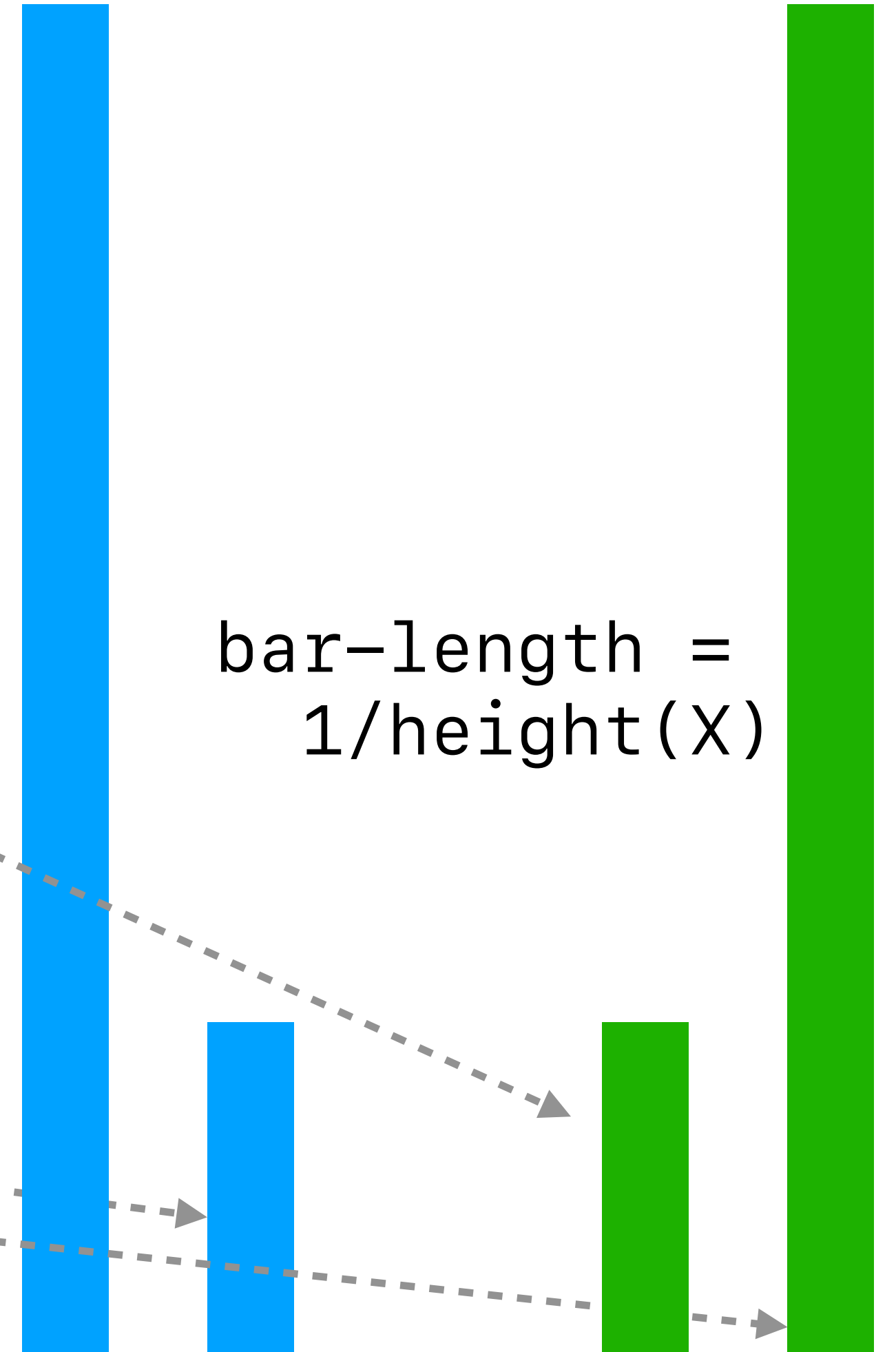


$\text{bar-length}(R(X))$
 $> \text{bar-length}(R(Y))$

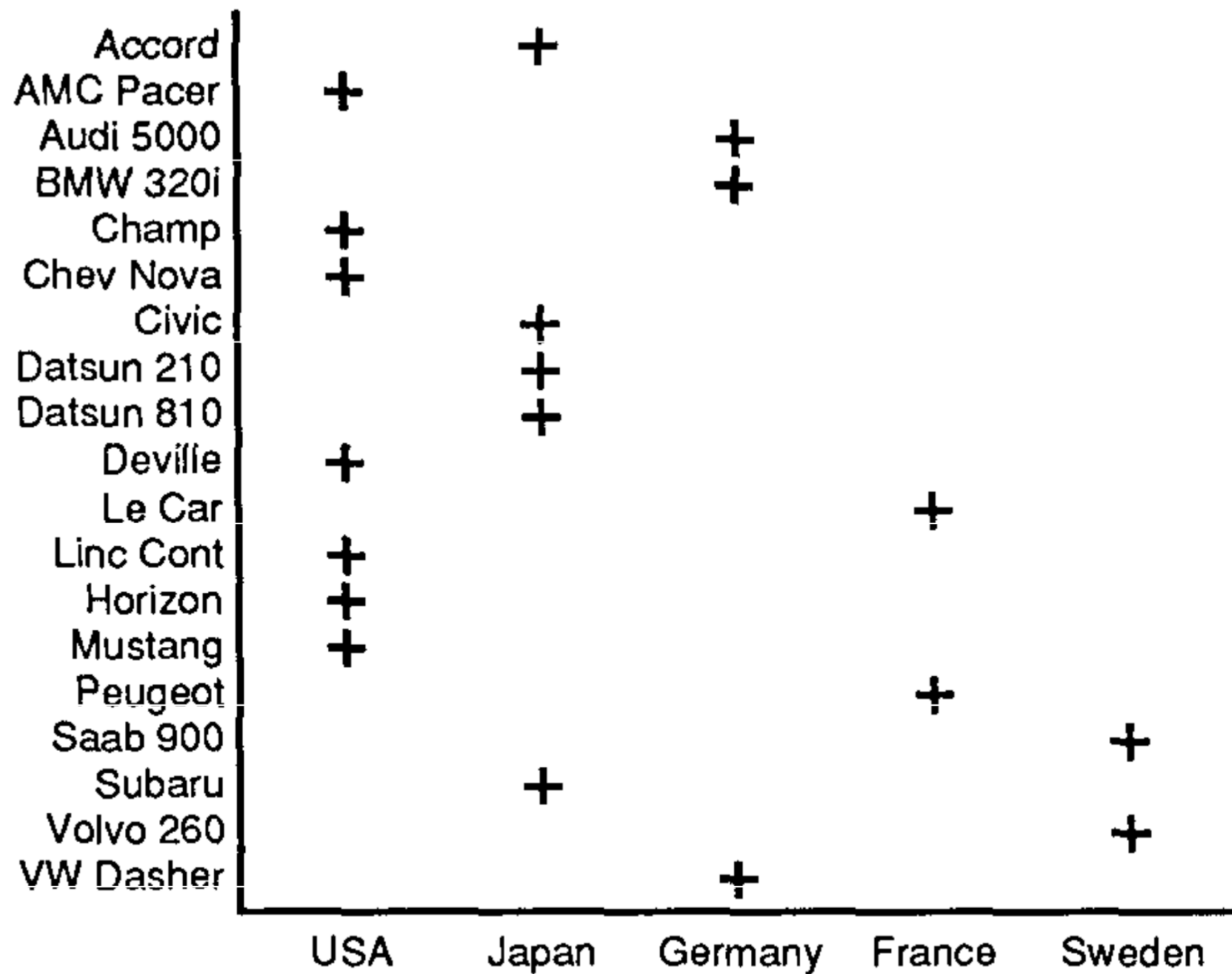
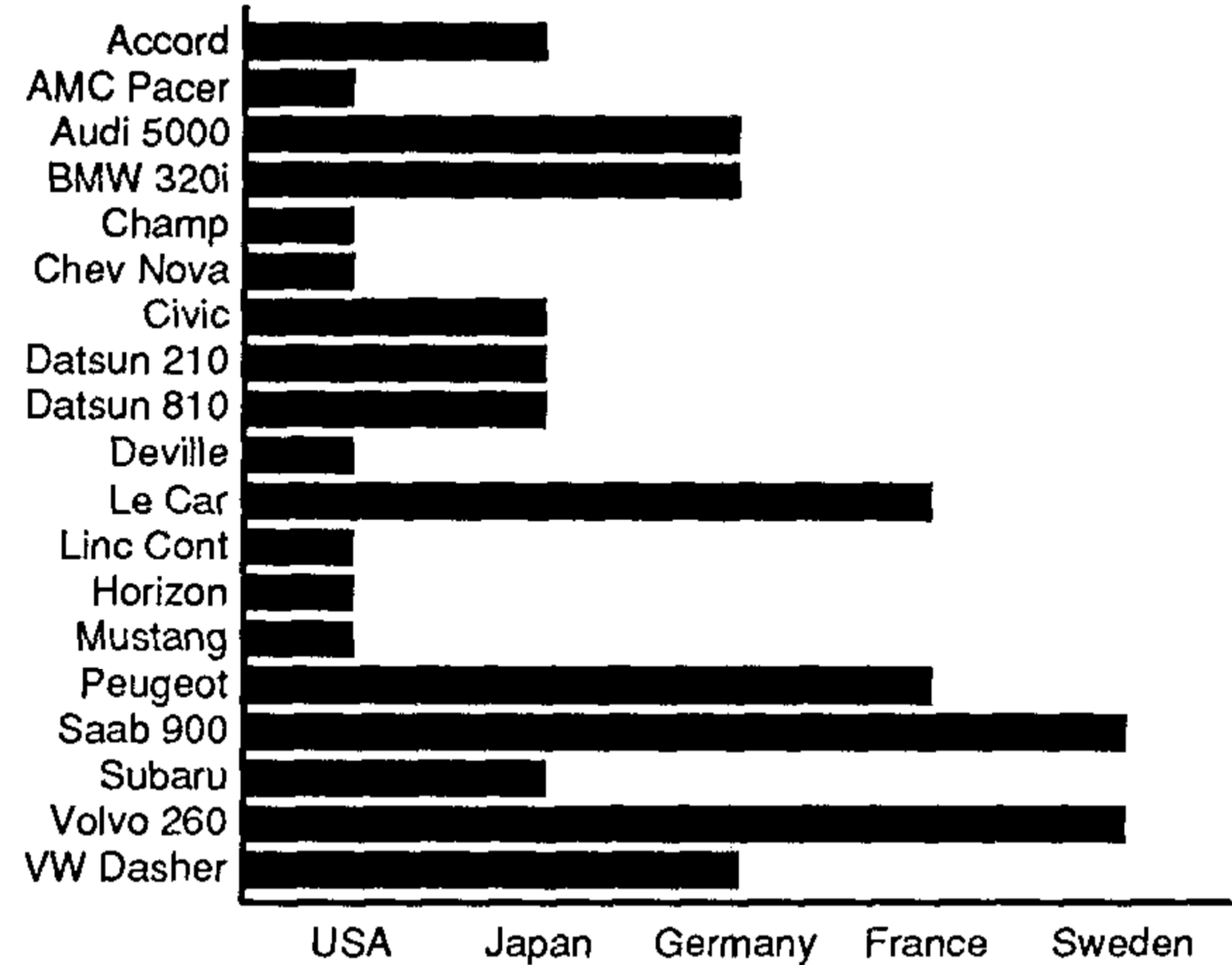


$\text{bar-length} = \text{height}(X)$

$\text{bar-length} = 1/\text{height}(X)$



Which mark best encodes the data?



Expressiveness: “a set of facts is expressible in a visual language if there is a visualization that encodes *all* and *only* the facts in the set.”

What is the mapping
from data to mark?

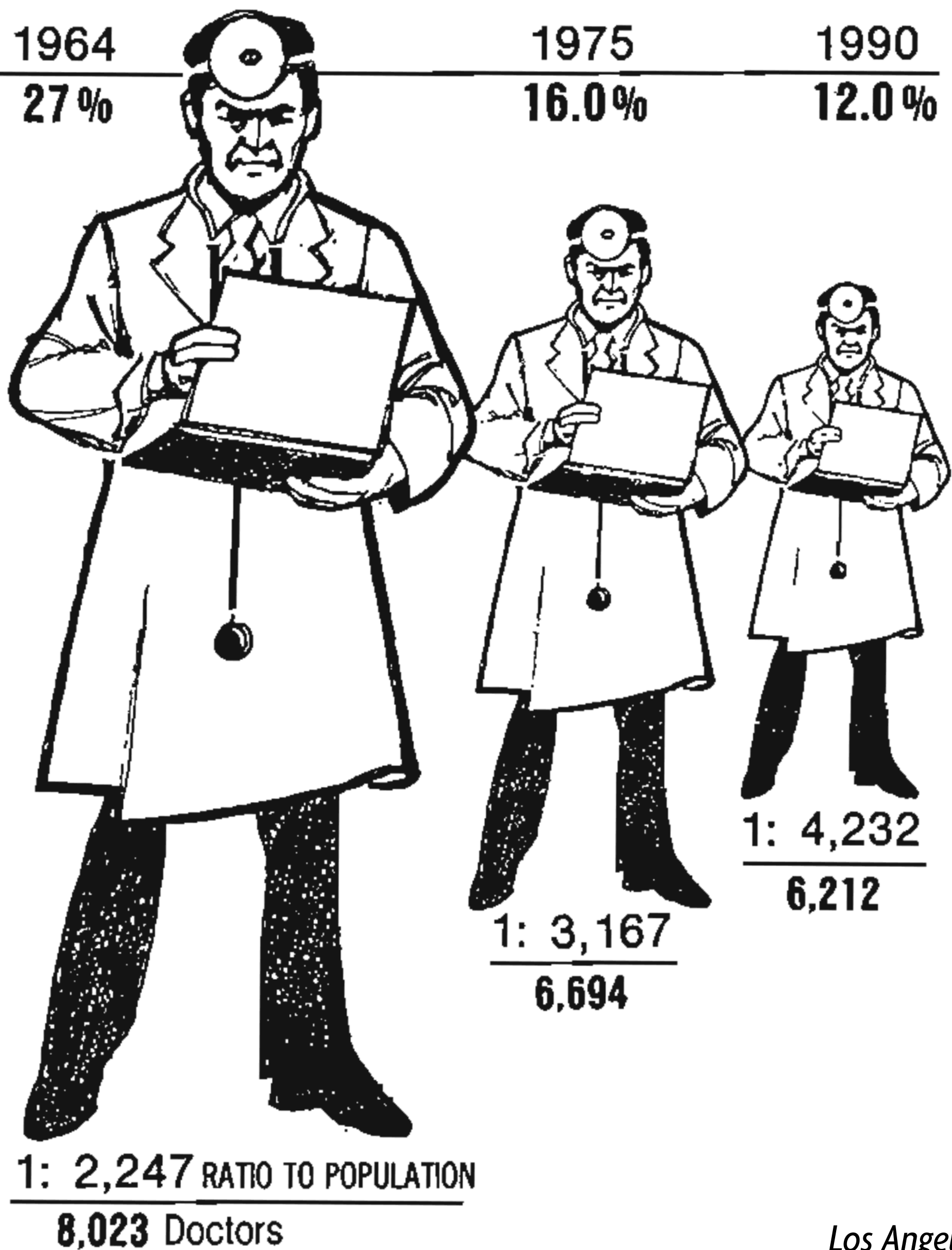
Why does it mislead?

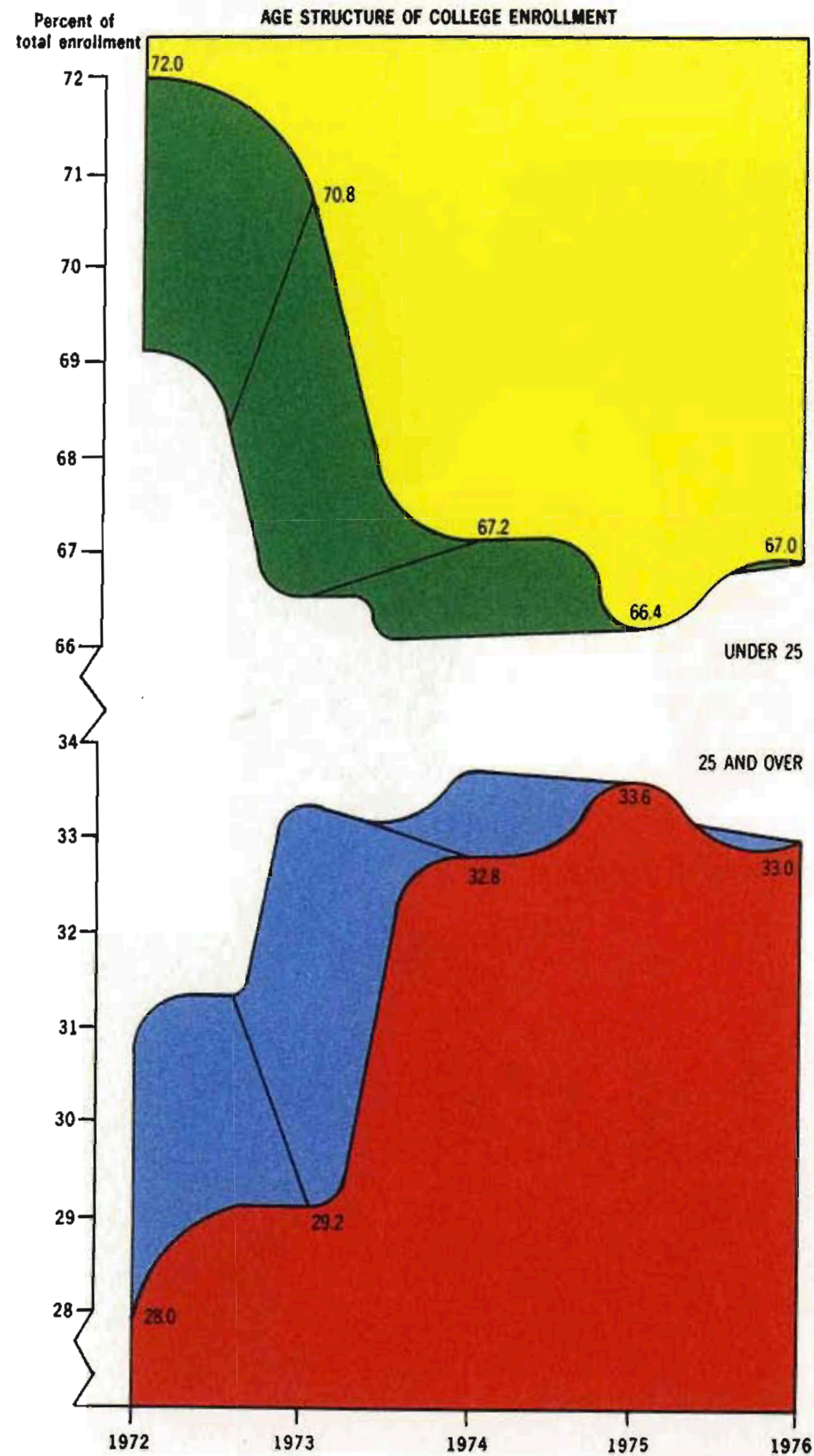
Data	Width	Area
8000	650	1,000,000
7000	400	400,000
6000	275	200,000

THE SHRINKING FAMILY DOCTOR
In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27 %	16.0 %	12.0 %





“This may well be the worst graphic ever to find its way into print” — Tufte

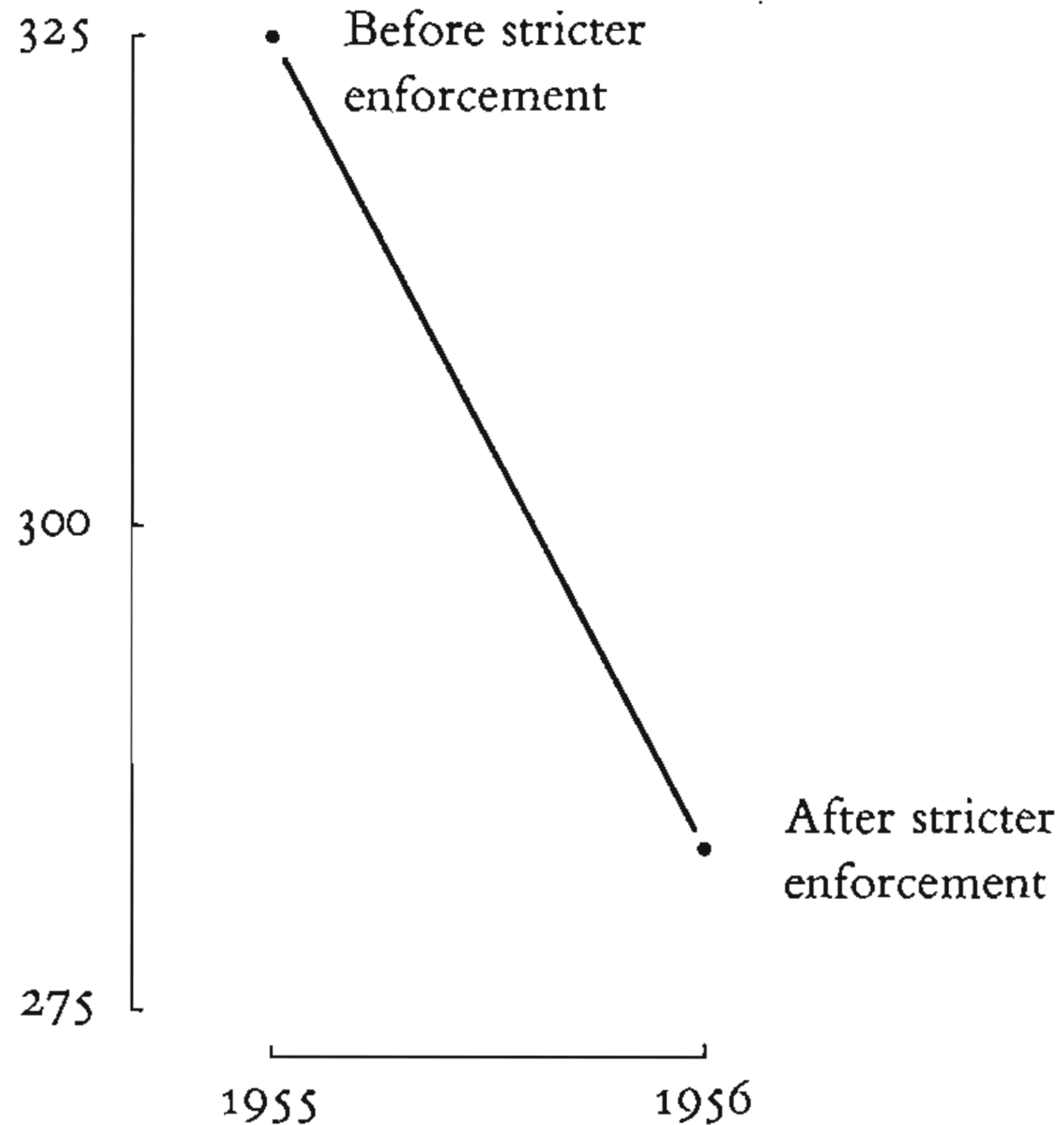
“The emaciated, data-thin design should always provoke suspicion, for graphics often lie by omission, leaving out data sufficient for comparisons. The principle:

Graphics must not quote data out of context.”

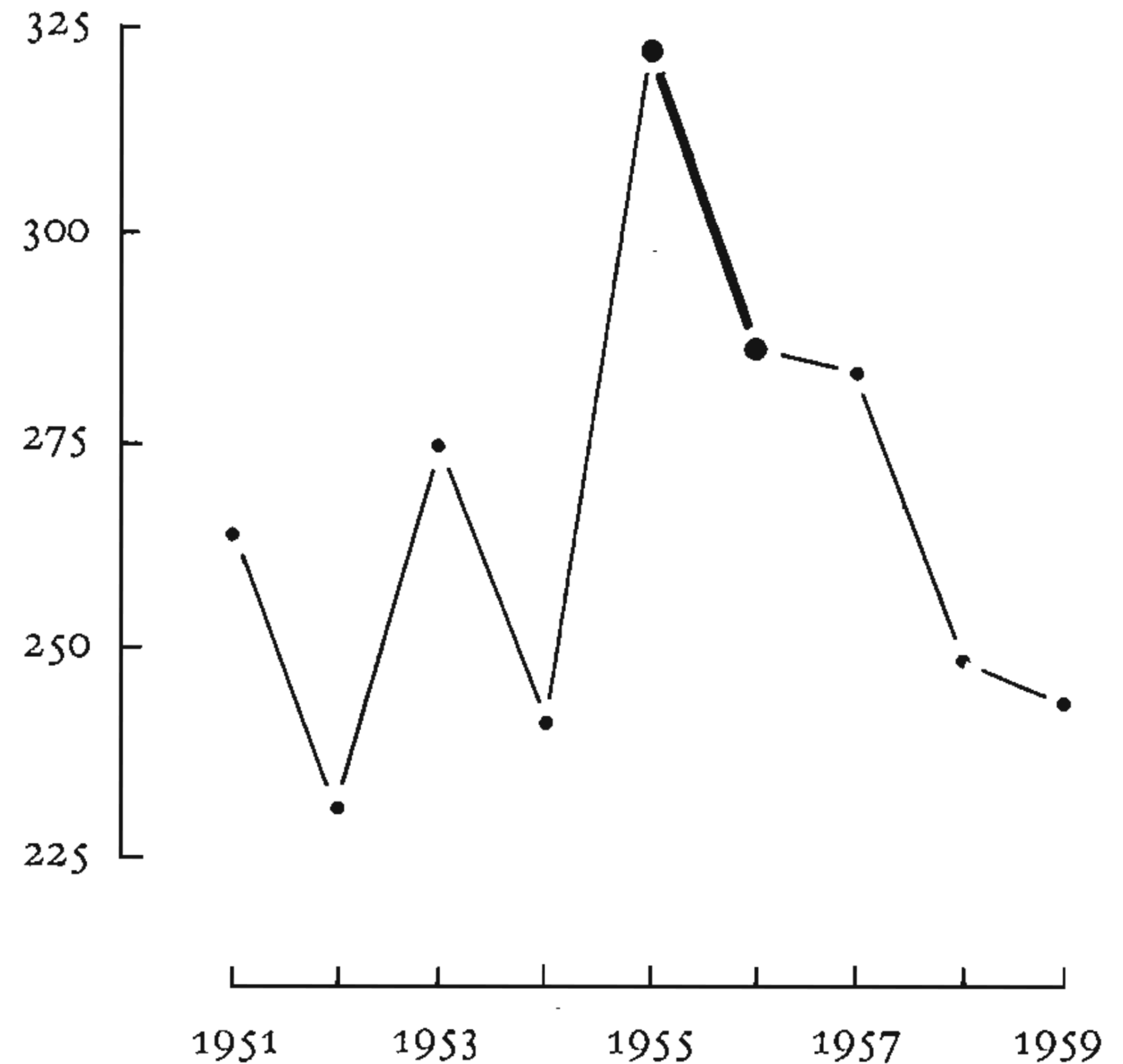


— Edward Tufte, *The Visual Display of Quantitative Information* (1983)

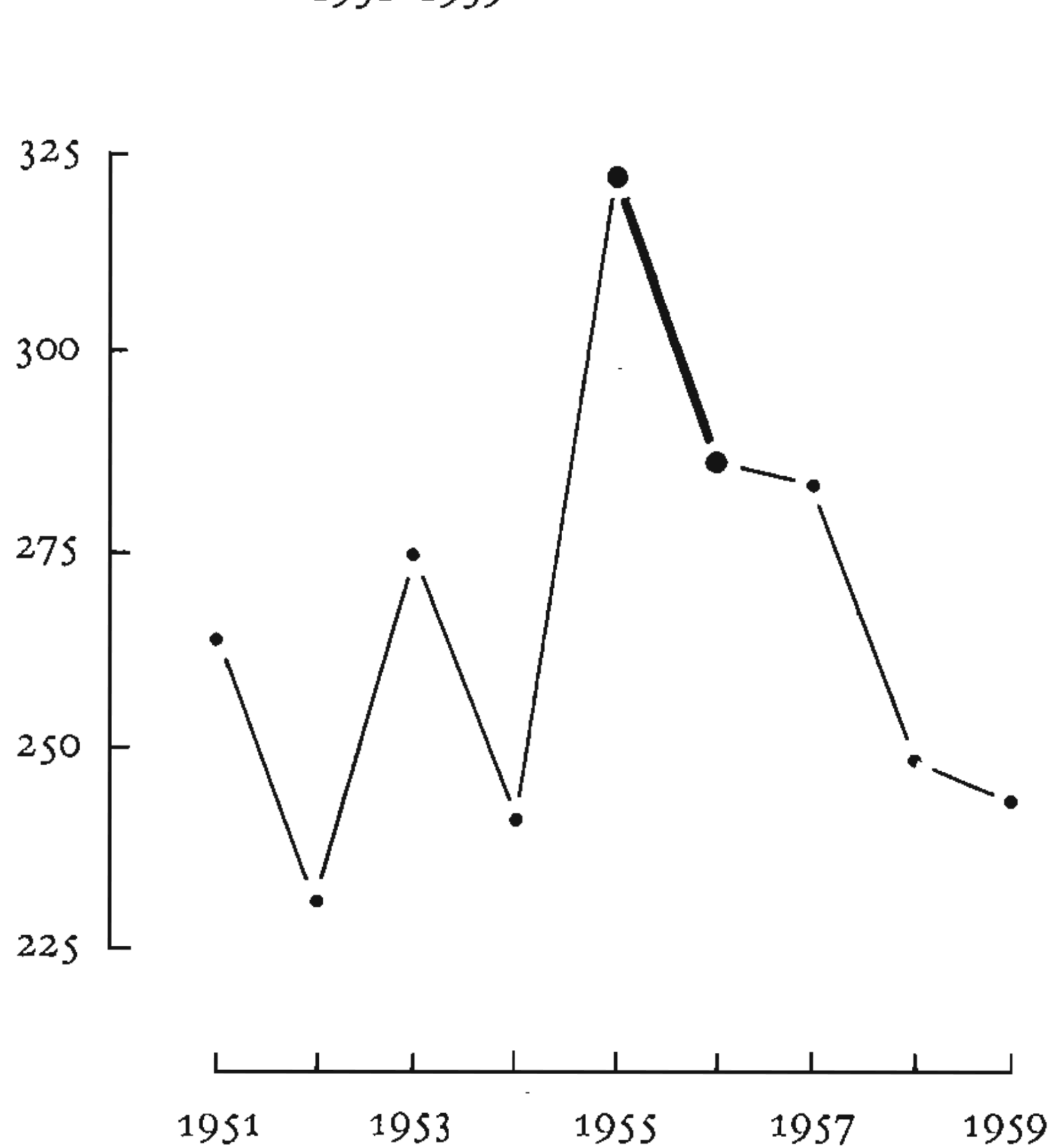
Connecticut Traffic Deaths,
Before (1955) and After (1956)
Stricter Enforcement by the Police
Against Cars Exceeding Speed limit



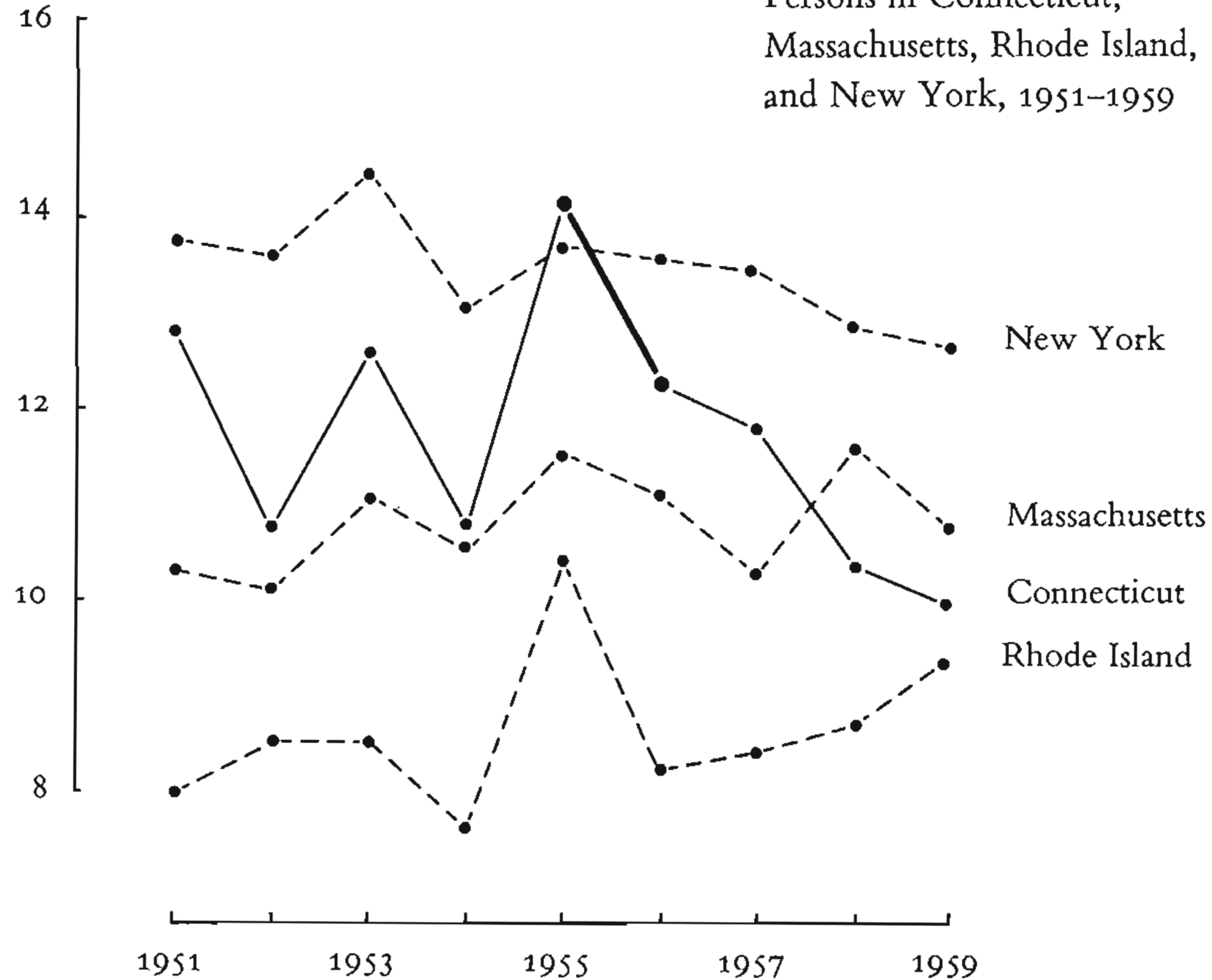
Connecticut Traffic Deaths,
1951–1959



Connecticut Traffic Deaths,
1951–1959



Traffic Deaths per 100,000
Persons in Connecticut,
Massachusetts, Rhode Island,
and New York, 1951–1959



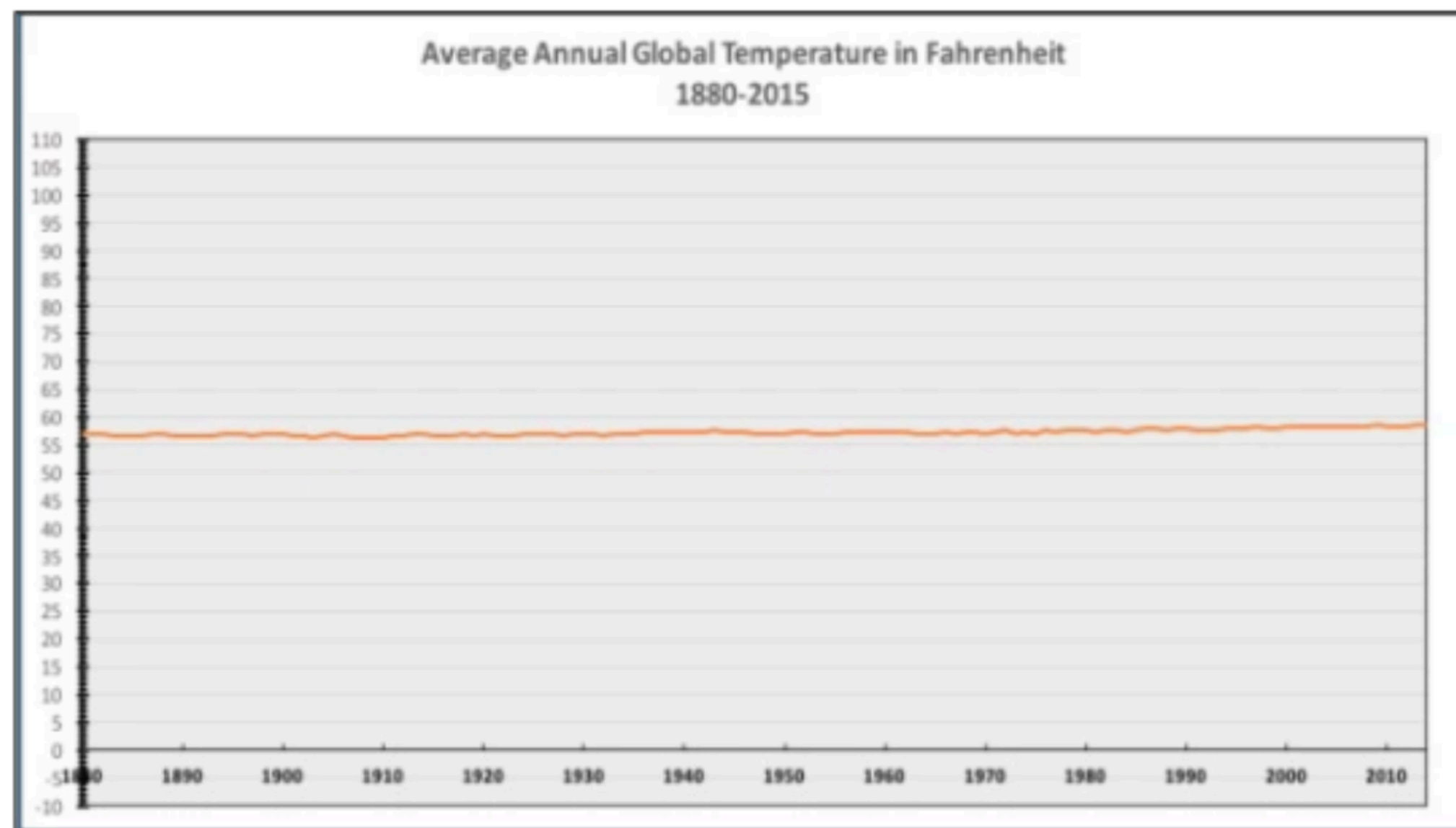
Scale is context, too!



Follow

The only [#climatechange](#) chart you need to see. natl.re/wPKpro

(h/t [@powerlineUS](#))



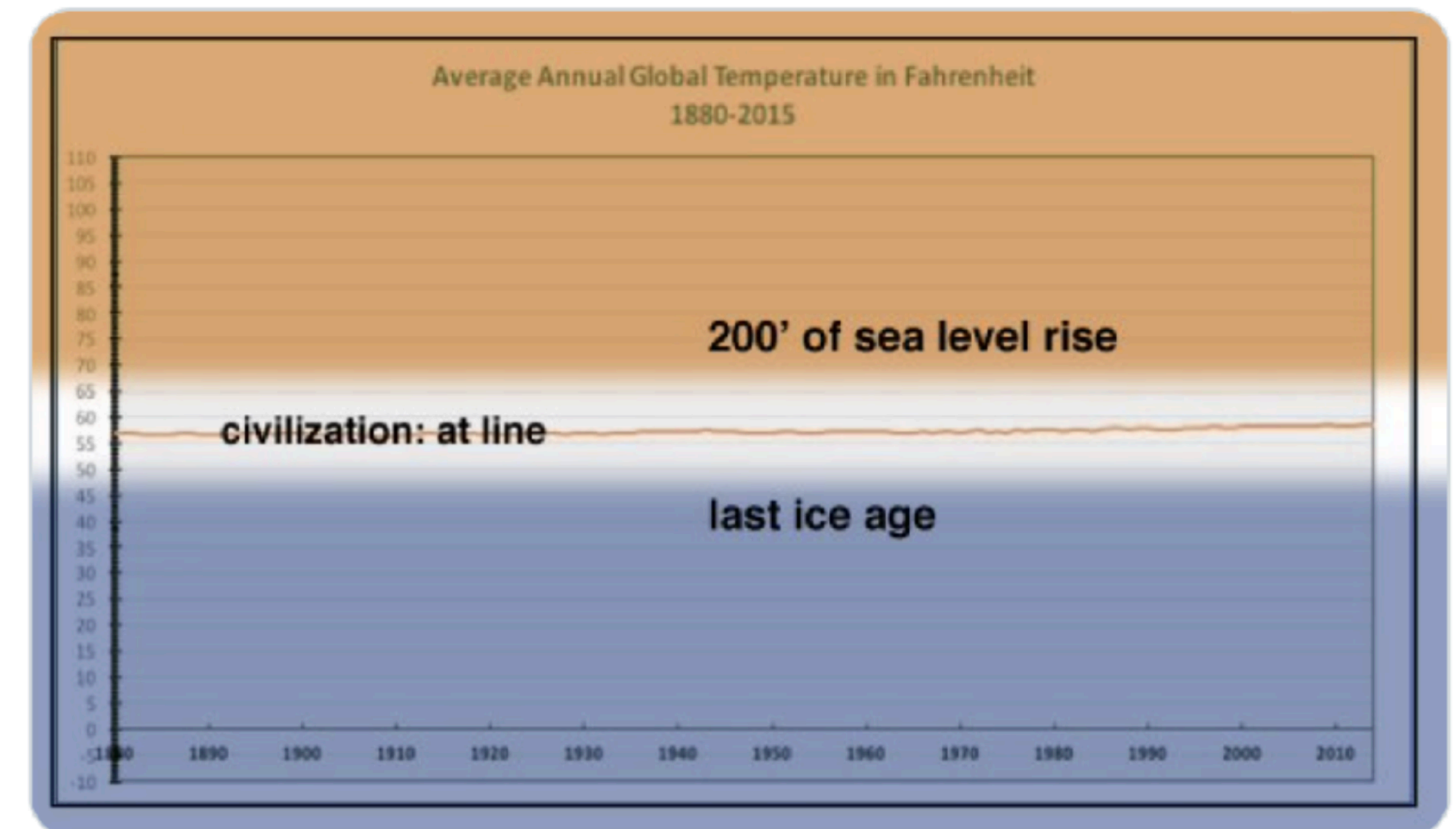
12:36 PM - 14 Dec 2015



City Atlas
@cityatlas

Replying to [@NRO](#)

[.@NRO](#) [@powerlineUS](#) [@bradplumer](#) I'm sure someone else has fixed this for you, but here you go. Great idea, thx --



5:28 PM · Dec 14, 2015

78 Retweets 1 Quote Tweet 208 Likes

Encodings and effectiveness

Which visualization helps **people**
quickly see information about the data?

Signal Detection

Discriminability: how easy is it to tell two things apart?

Magnitude Estimation

Accuracy: how correctly can we read off values?

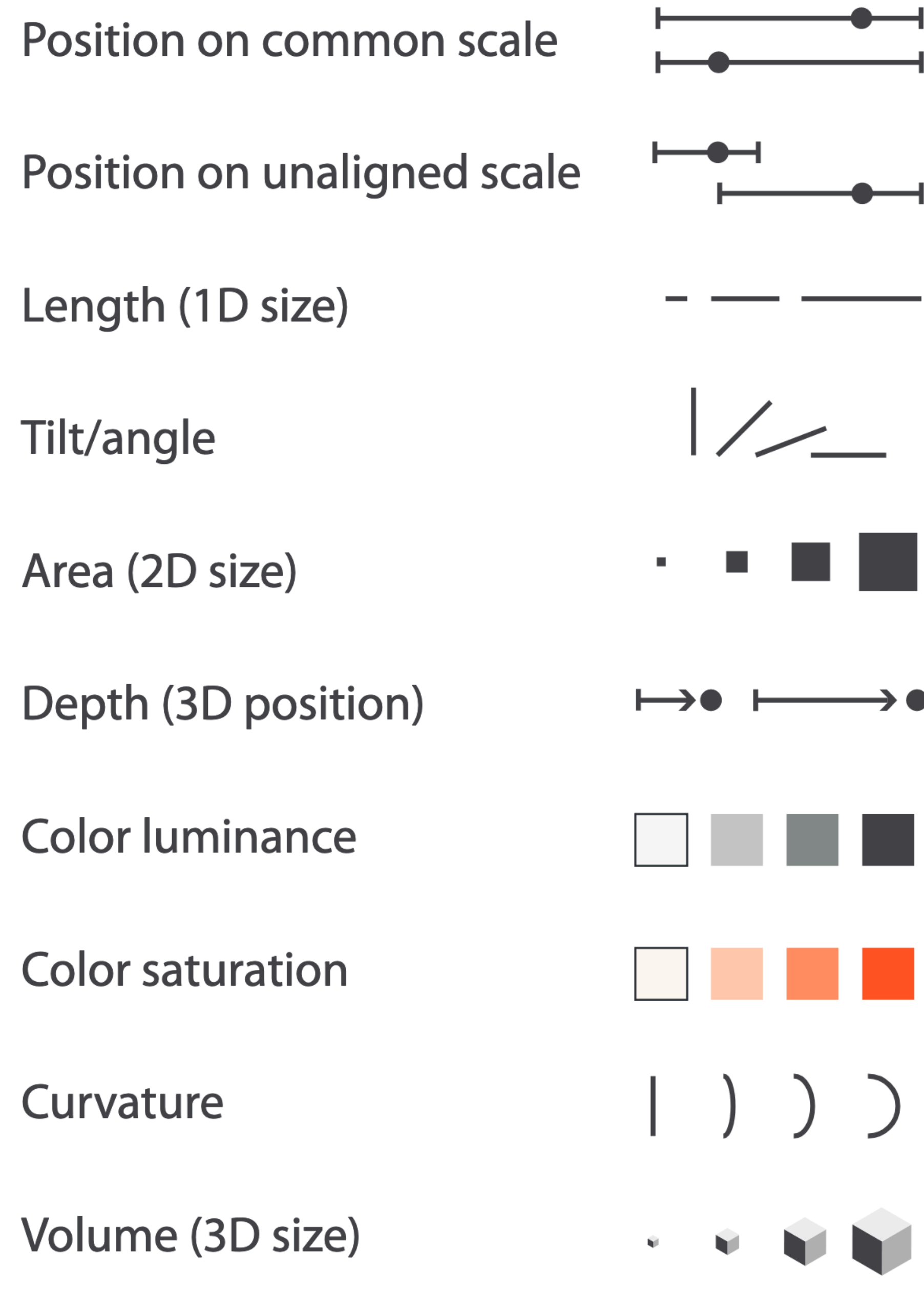
Preattentive Processing

Pop out: how easy is it to spot some values from the rest?

Selective Attention

Separability: how much interaction occurs between attributes?

➔ **Magnitude Channels: Ordered Attributes**



Same

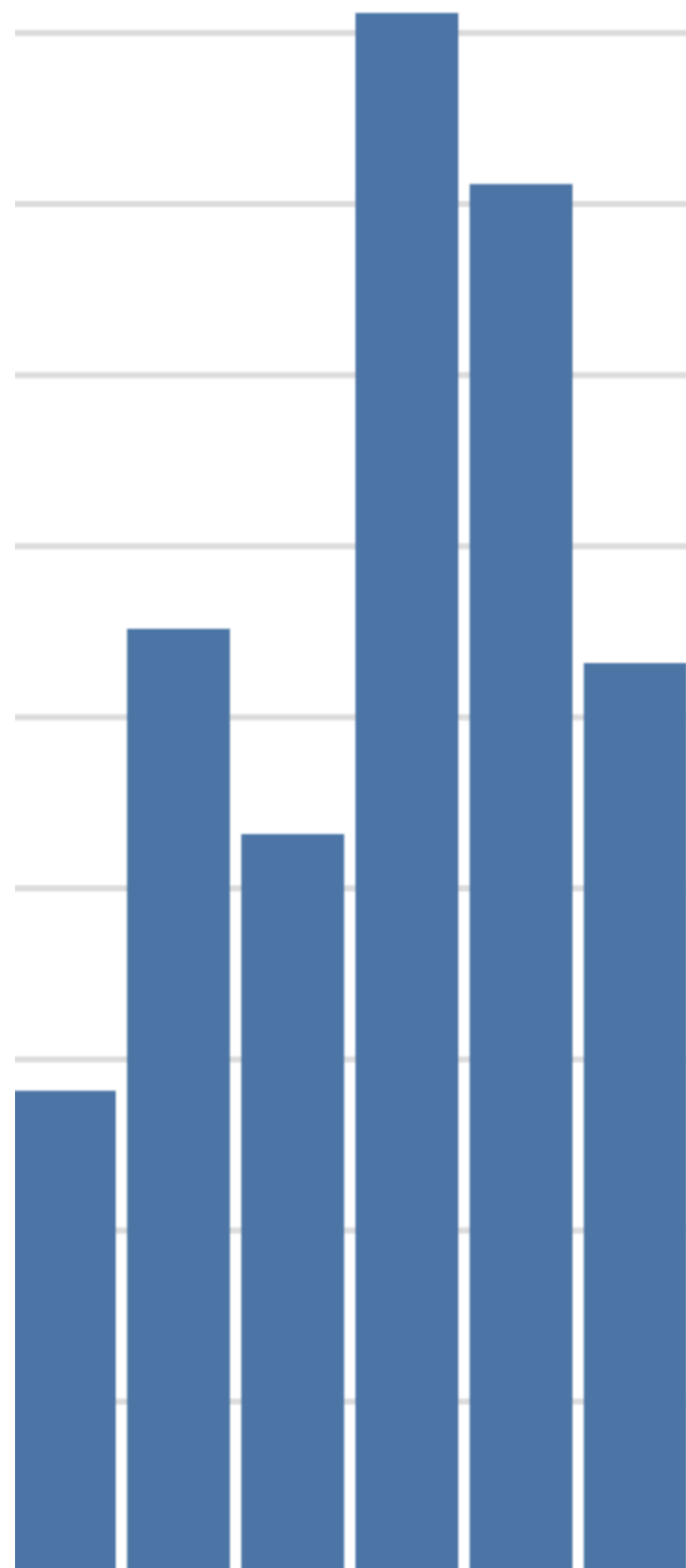
➔ **Identity Channels: Categorical Attributes**



Most
Effectiveness
Least

Which of these visualizations is better?

Which encoding of nominal $\times$ quantitative data facilitates comparisons across categories?



Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance

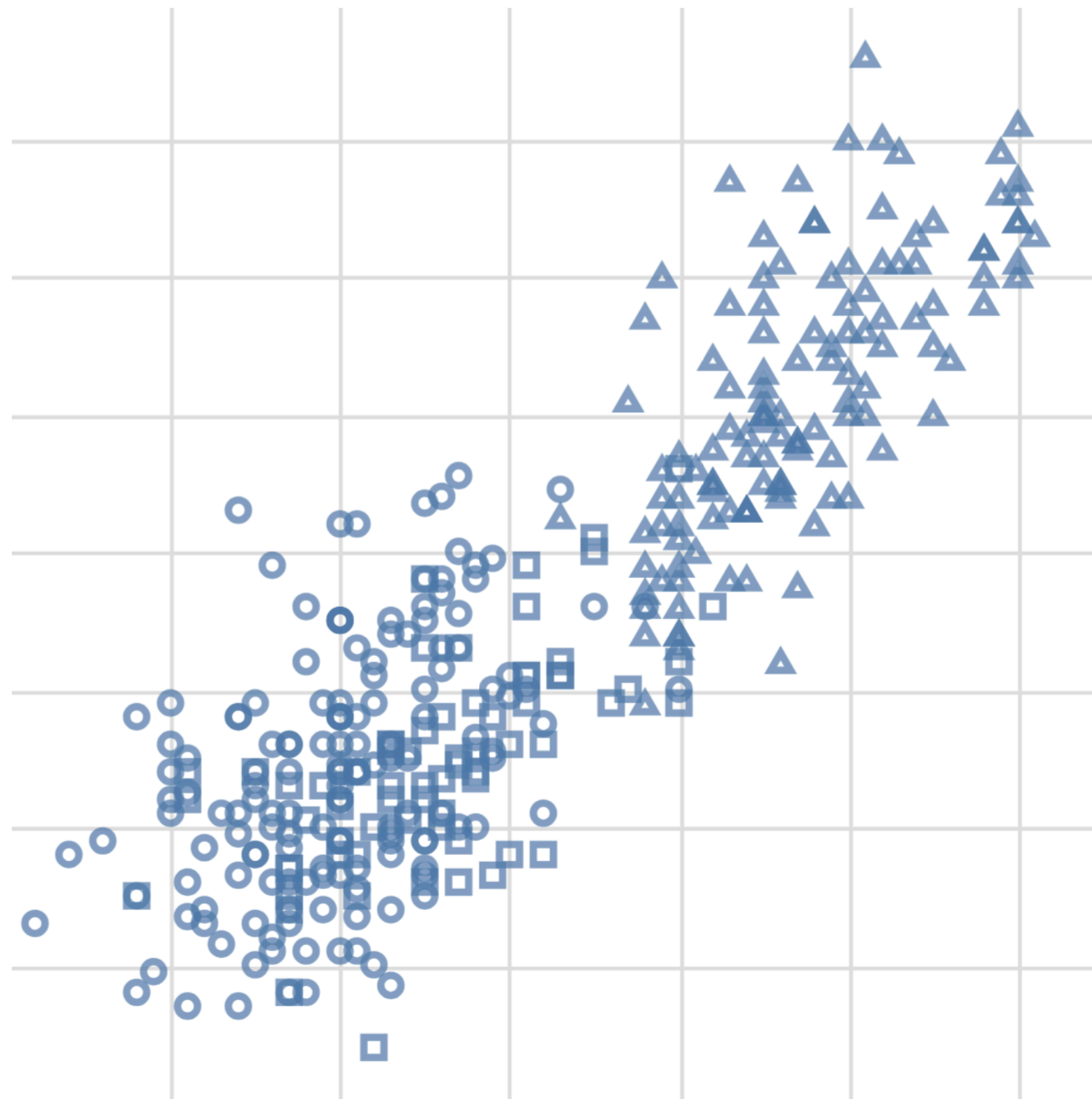
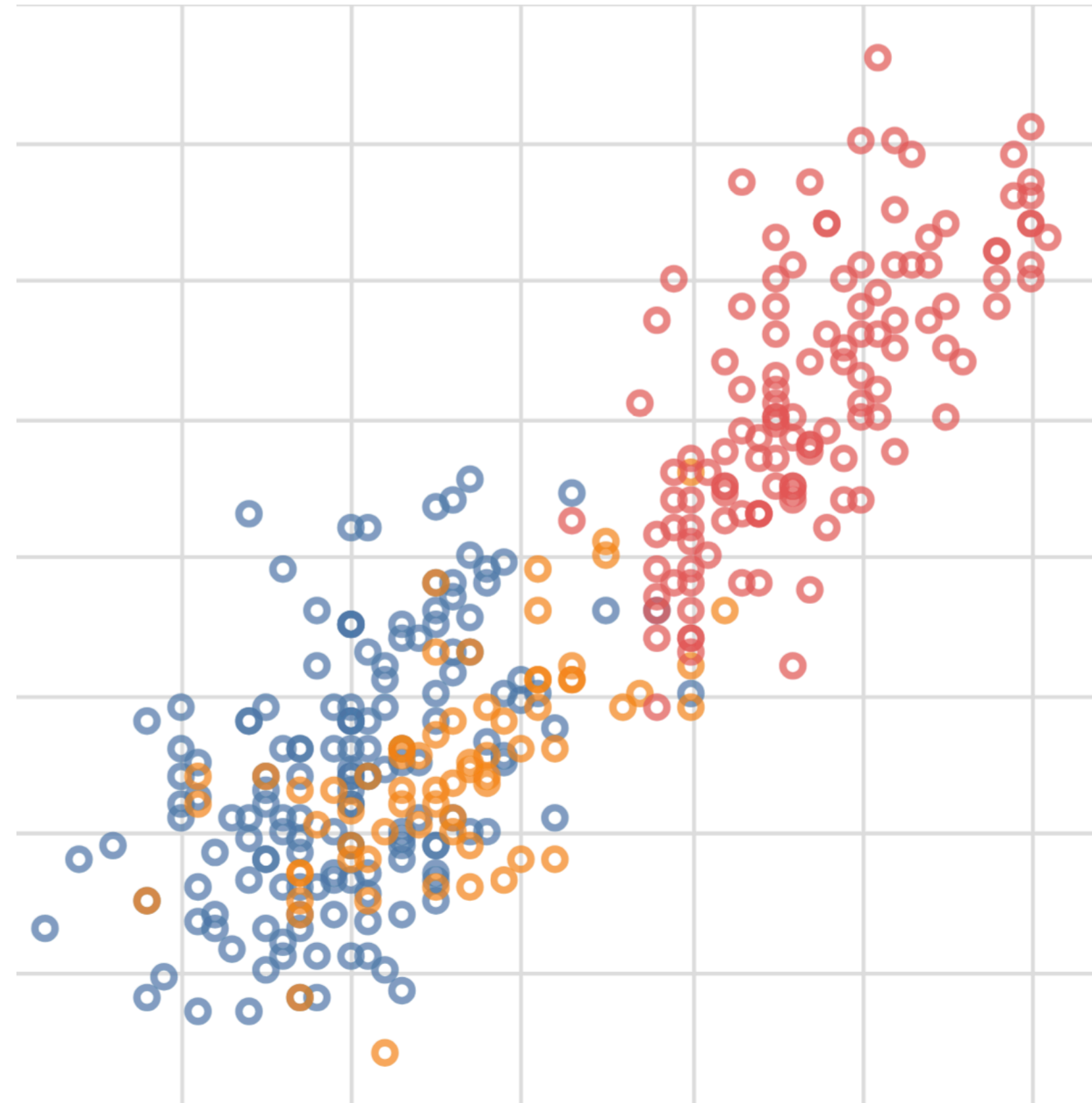


How does Z compare across X and Y ?

Color

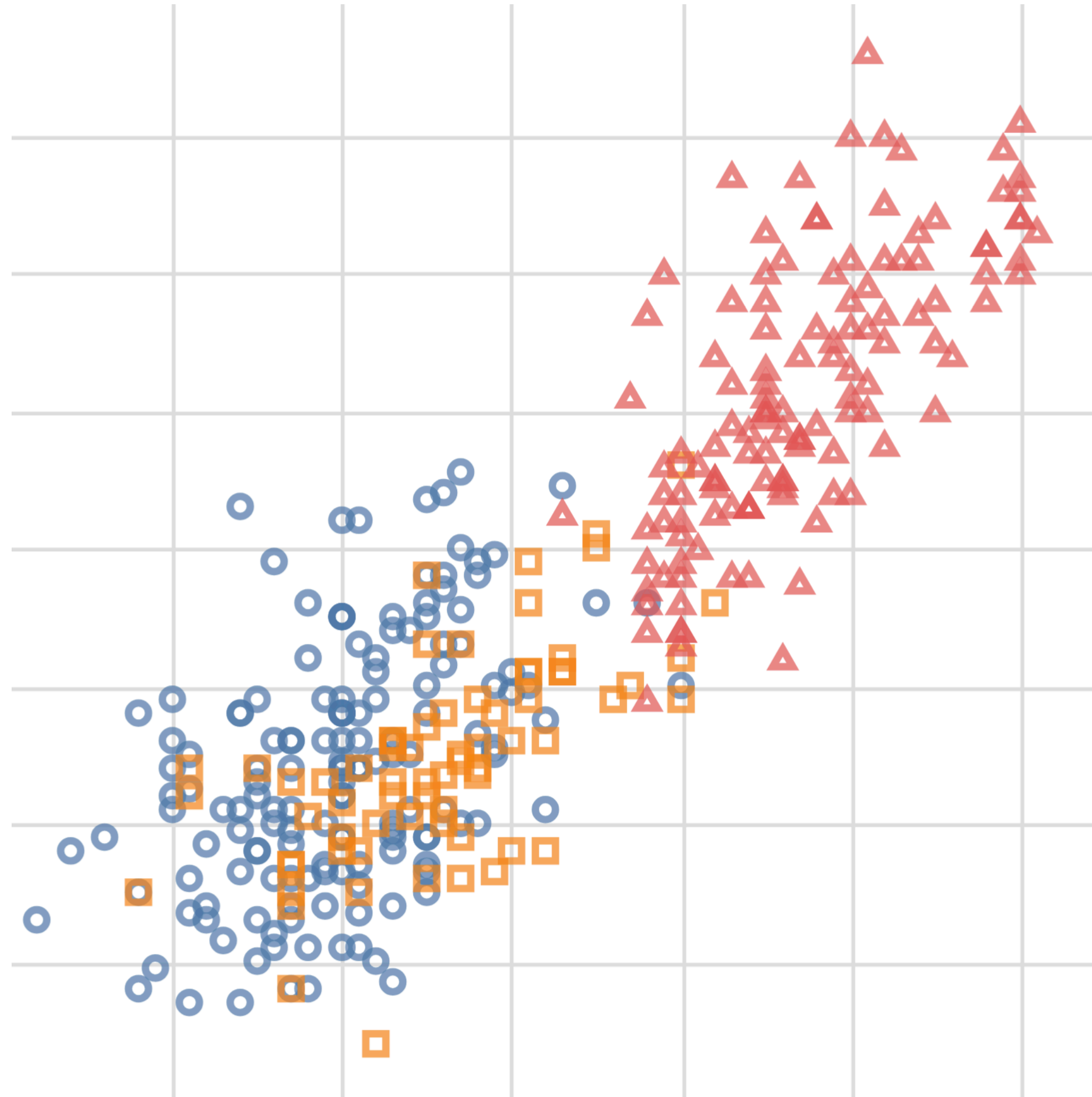


Quantitative Y



Quantitative X

Por que no los dos? Use color + shape for redundant encoding



For a given X , how does Y compare across Z ?

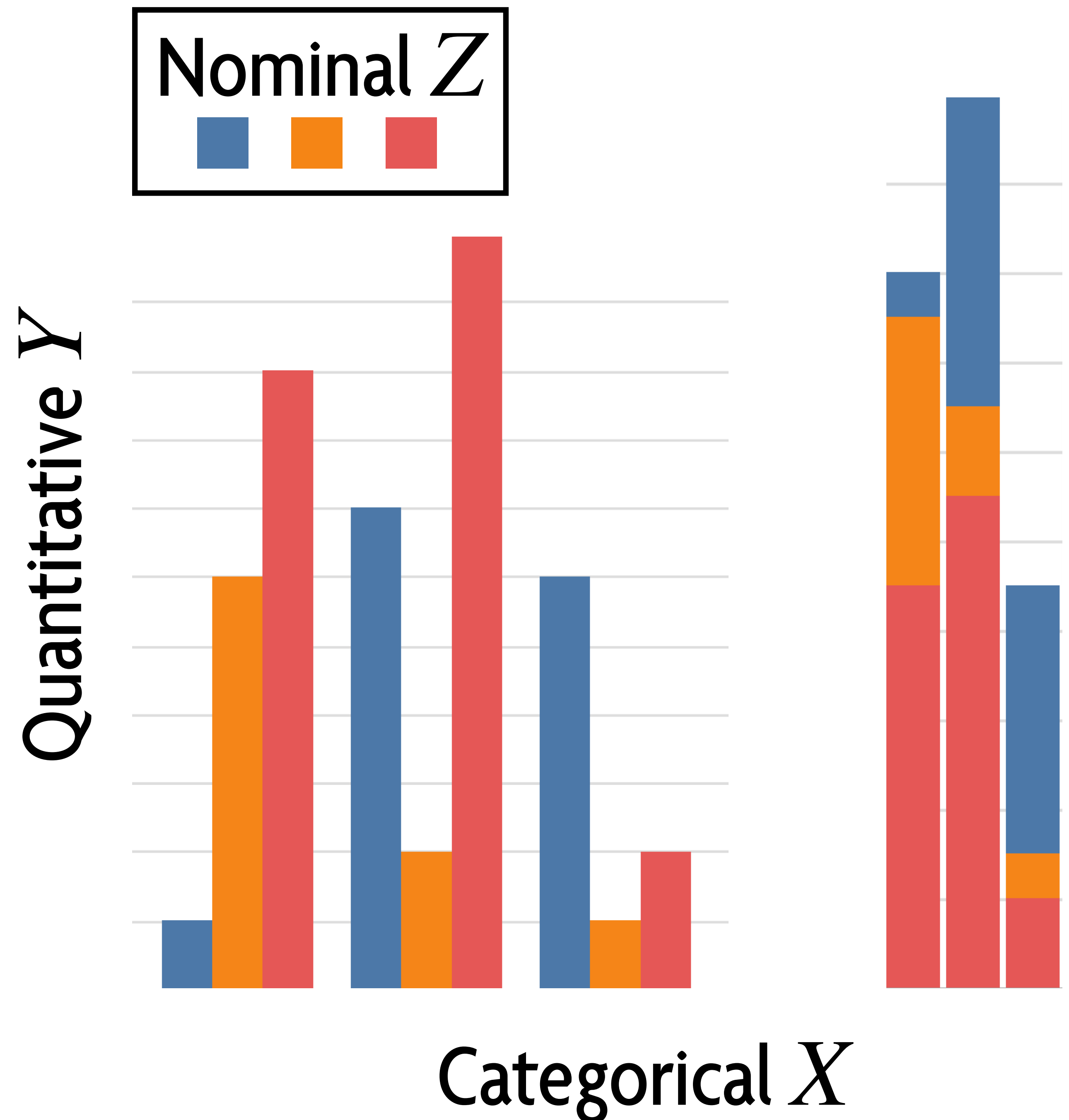
Grouped bar

Combining Z , how does Y compare across X ?

Stacked bar

For a given Z , how does Y compare across X ?

Grouped bar

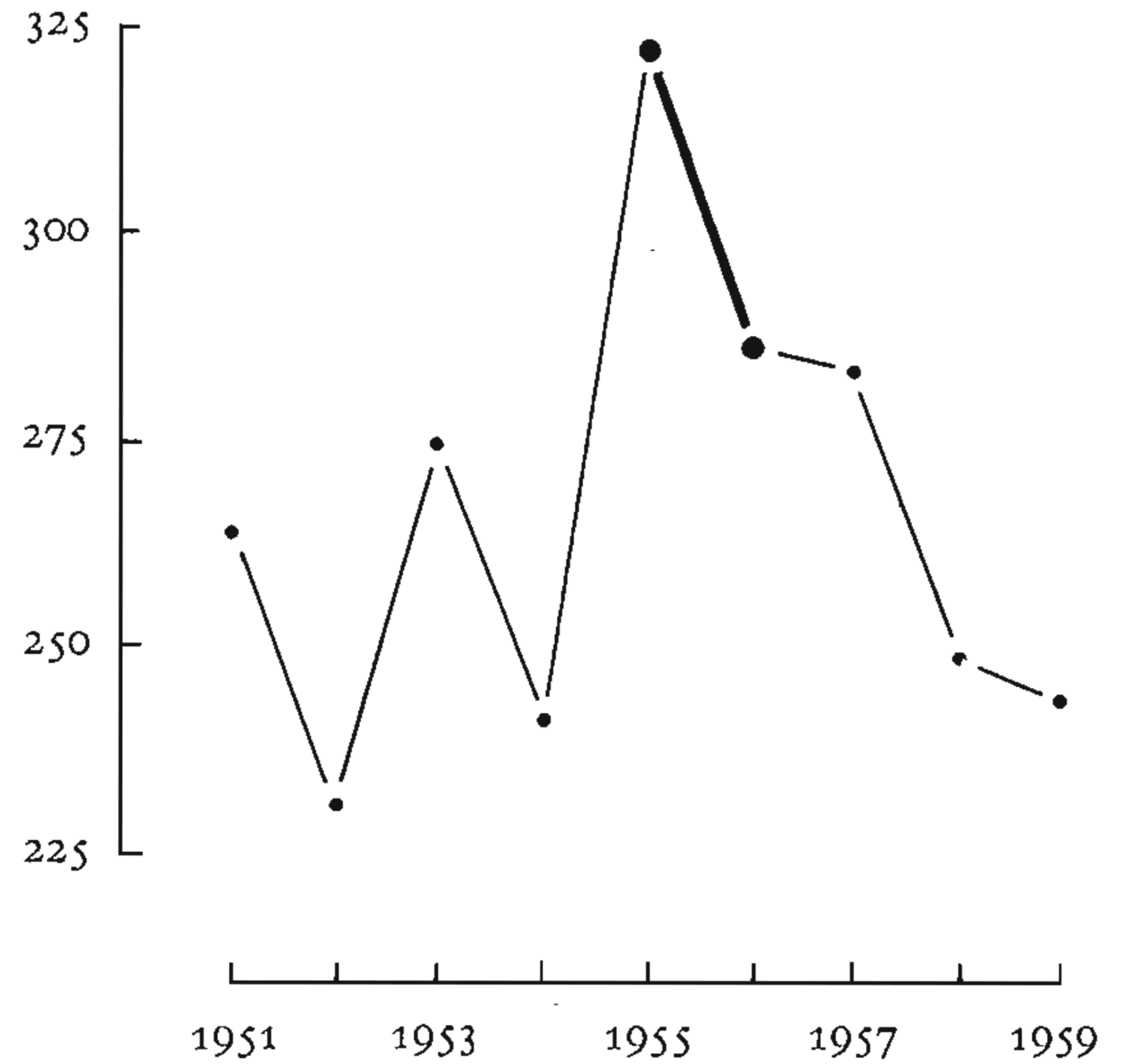


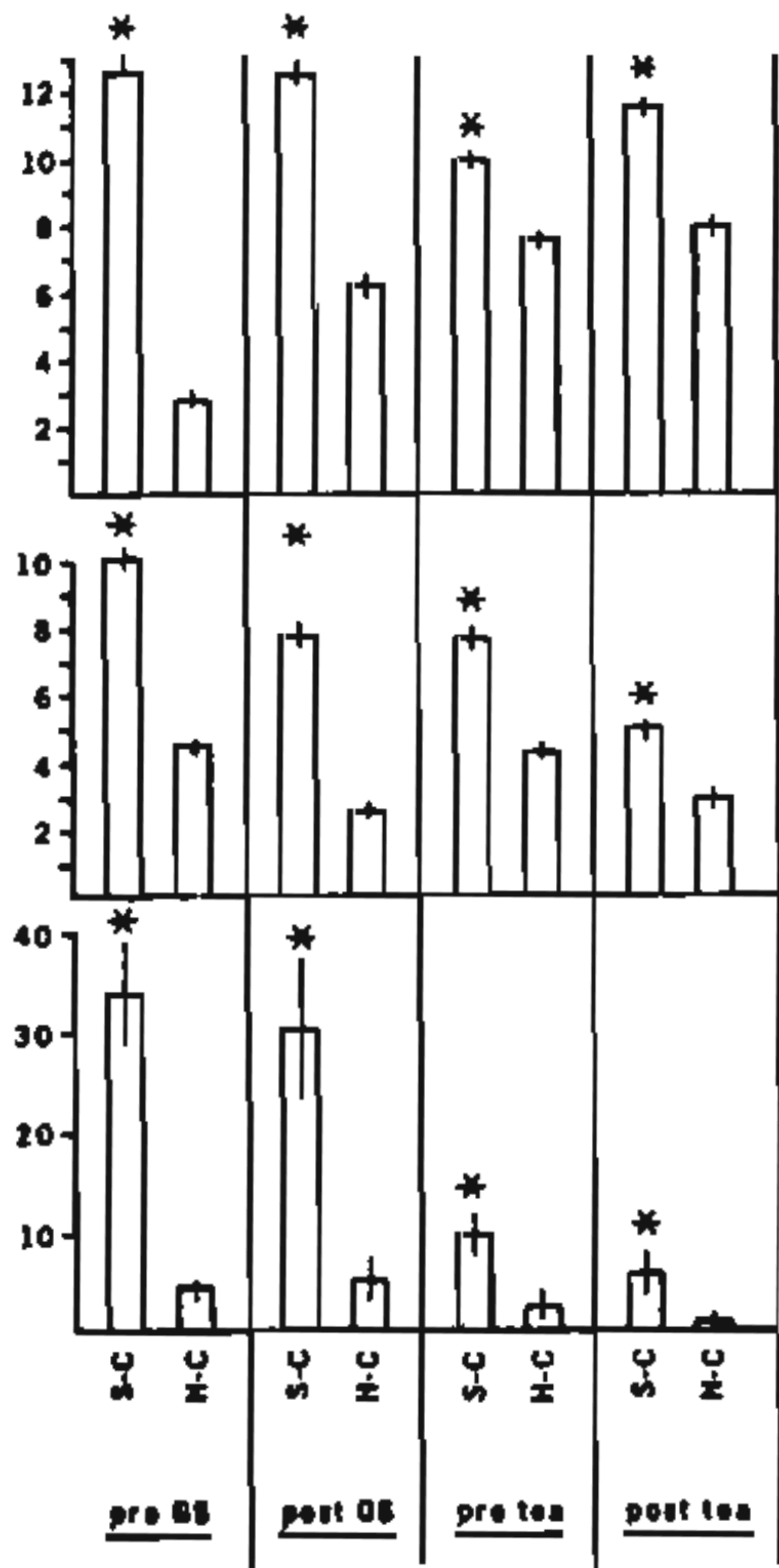
Additional rules of thumb

Data-ink ratio:

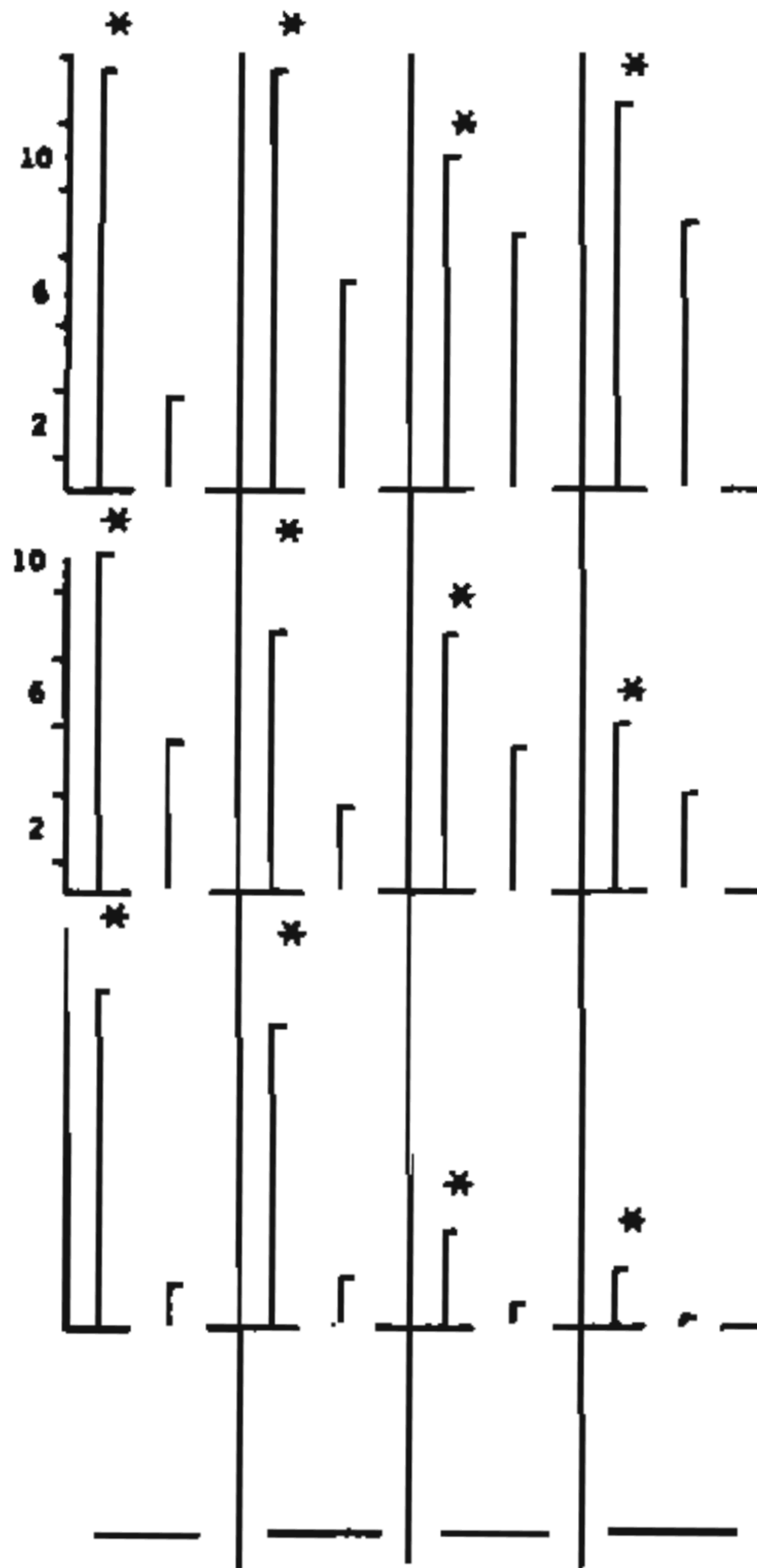
data-ink

total ink used to print the graphics

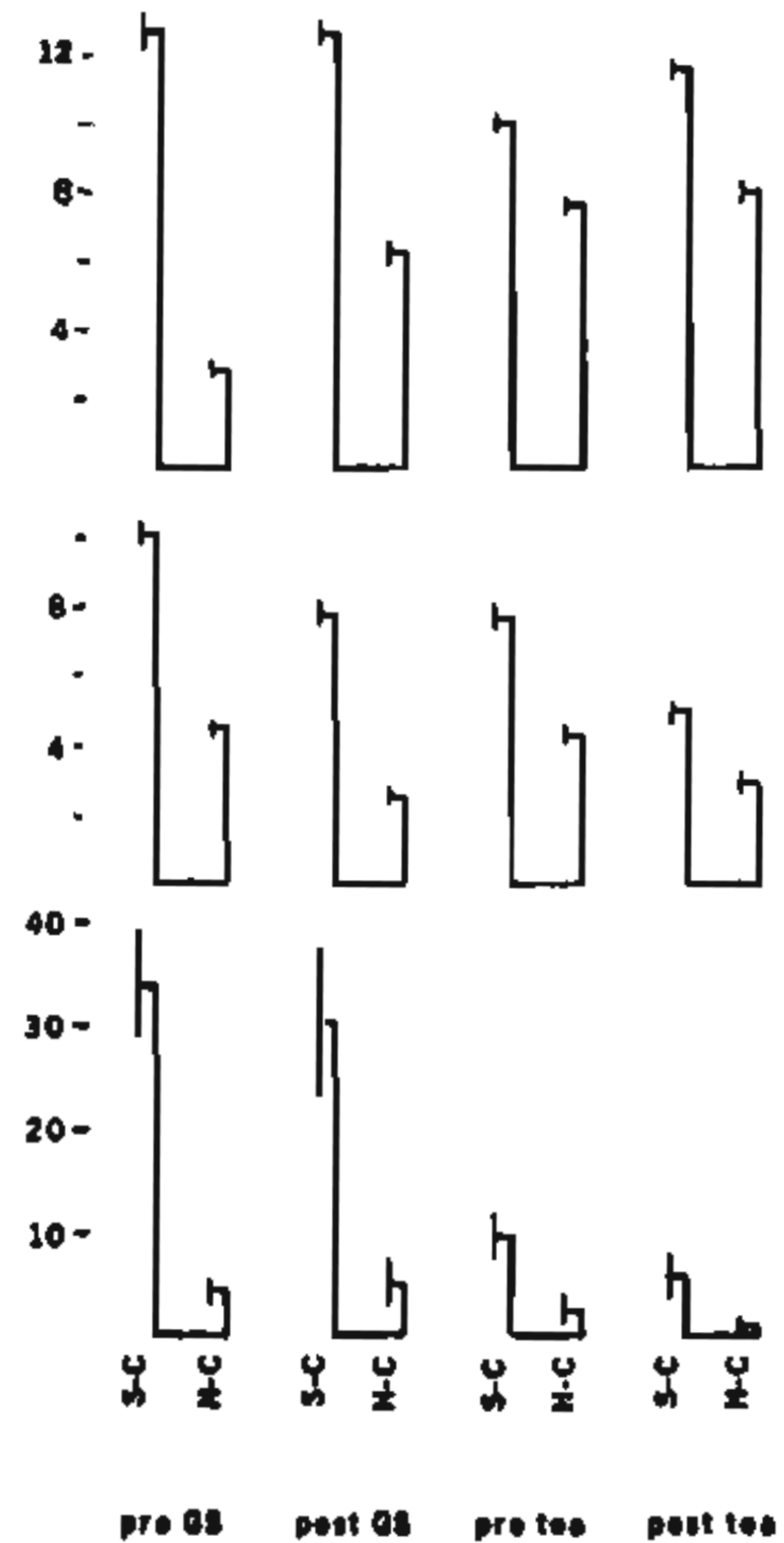




==

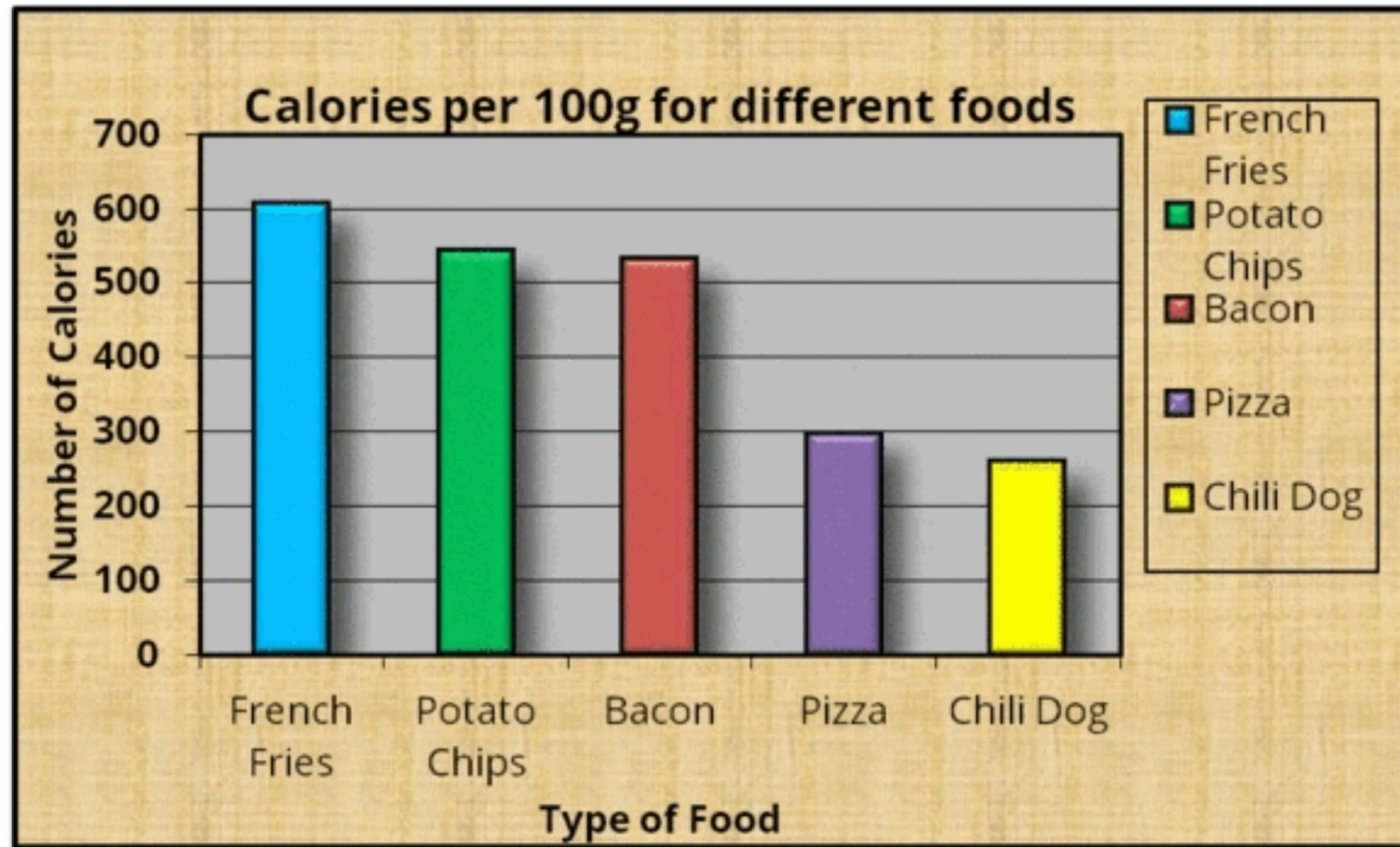
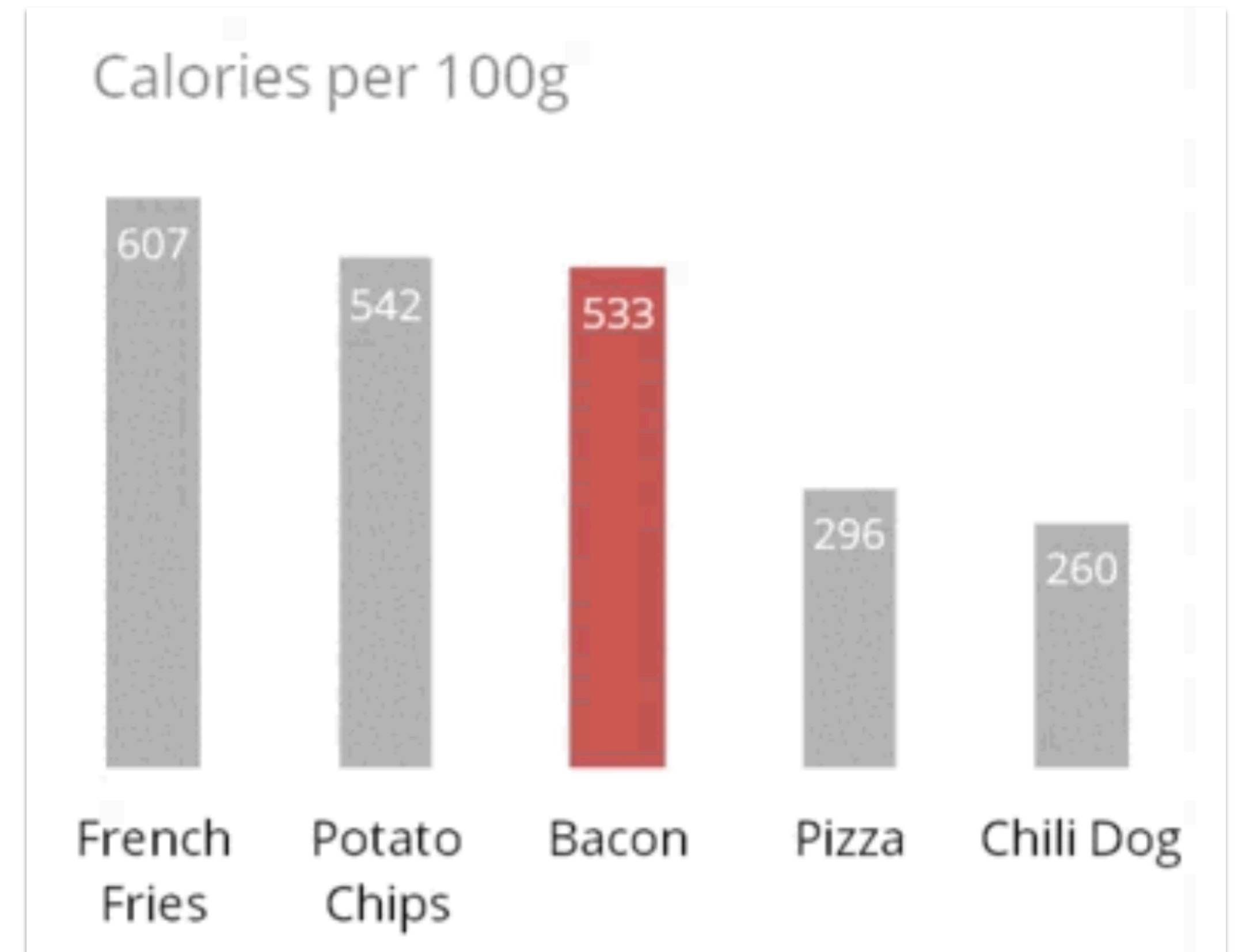


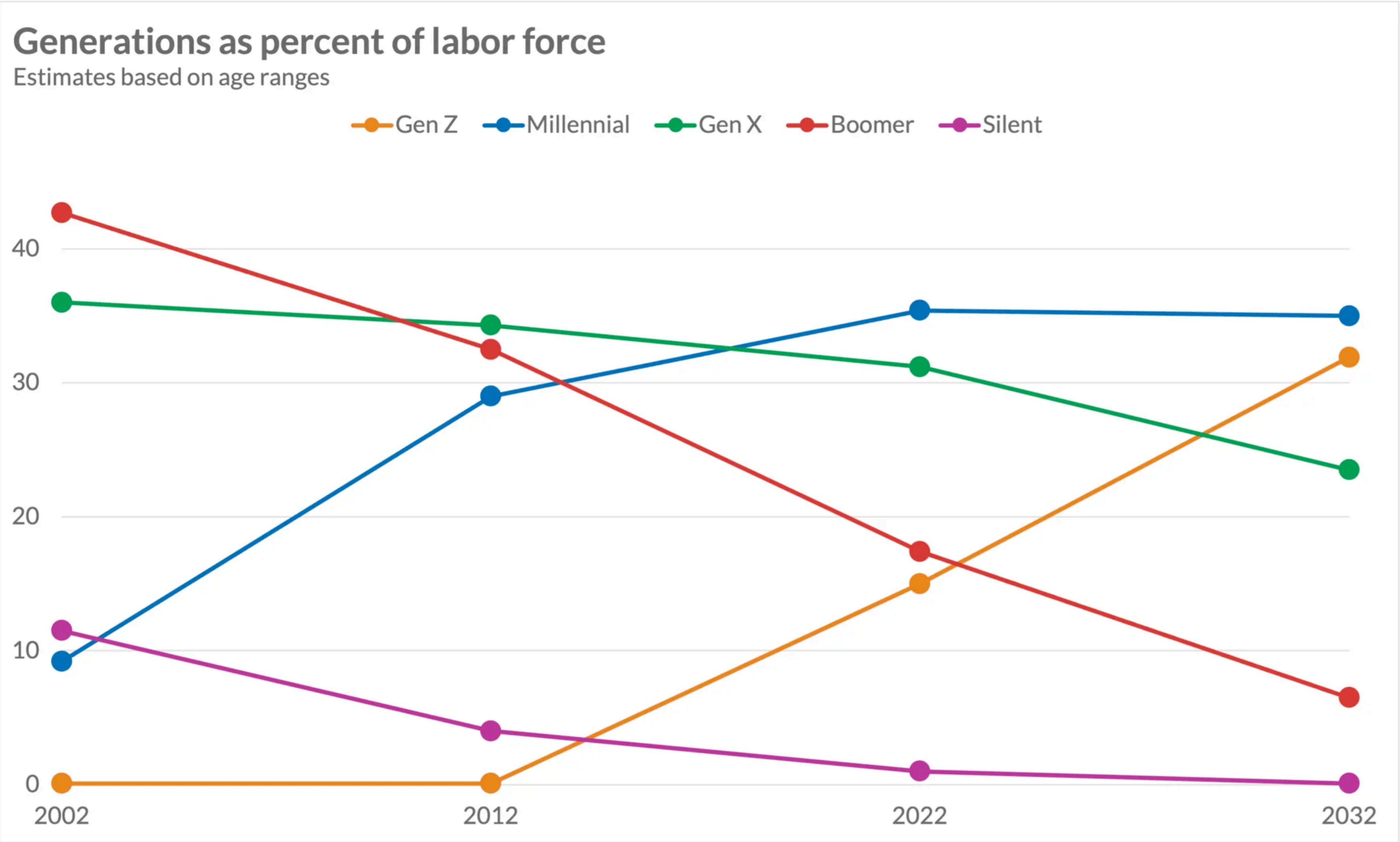
+



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

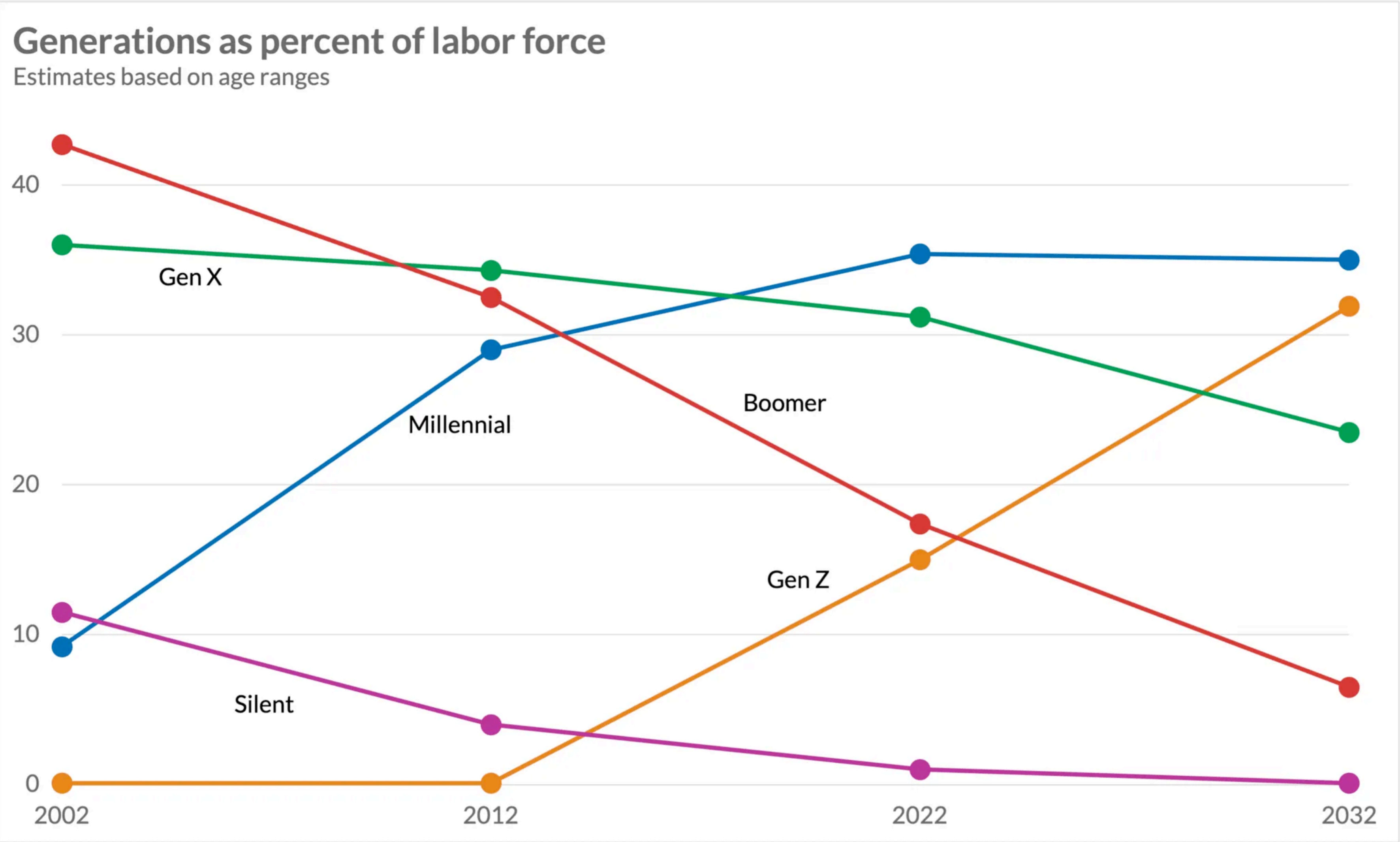
How much non-data-ink do you need?





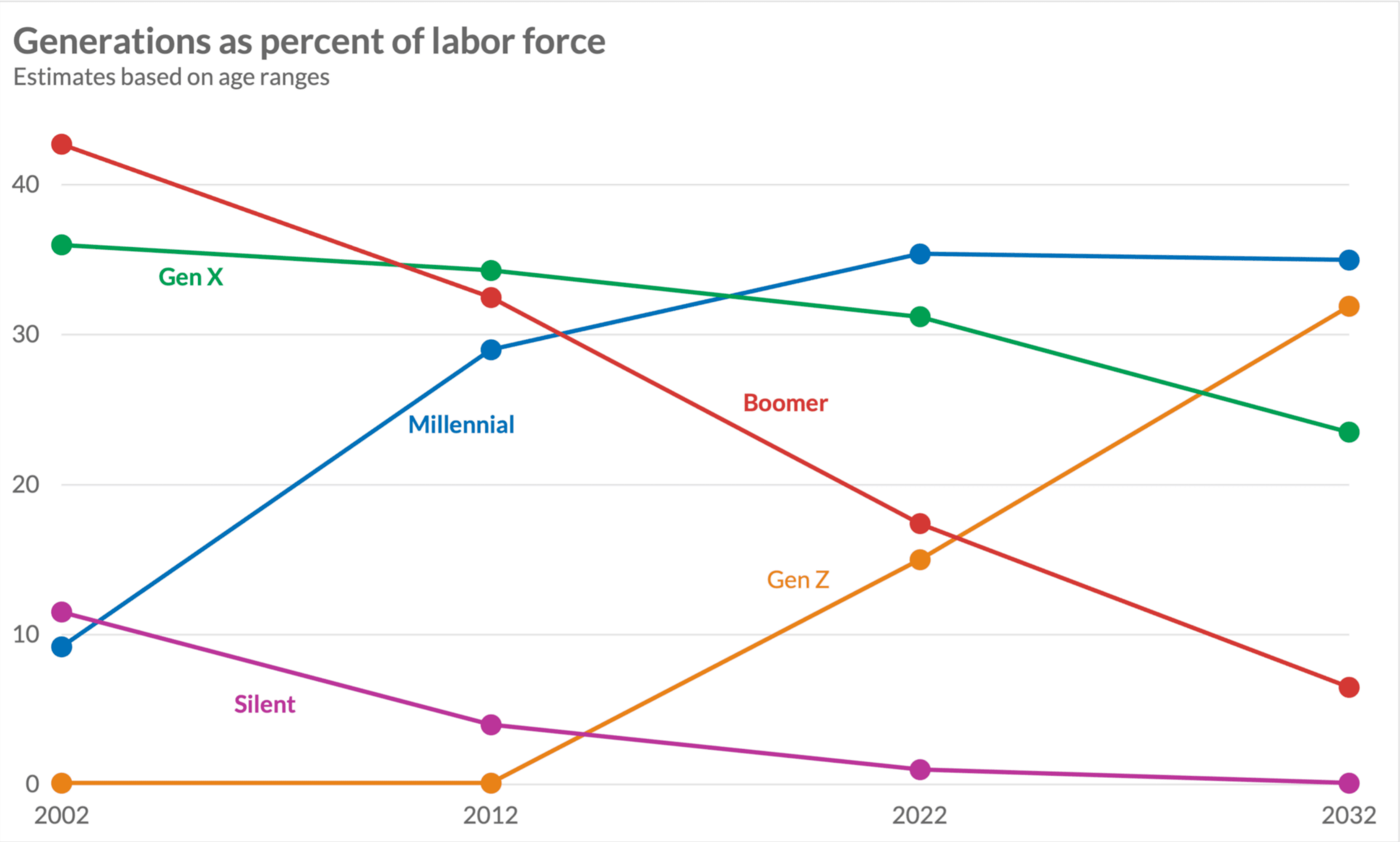
The default legend
for line charts.

Can we do better?



✓ **Direct labeling:** reader doesn't need to split attention between the legend and chart

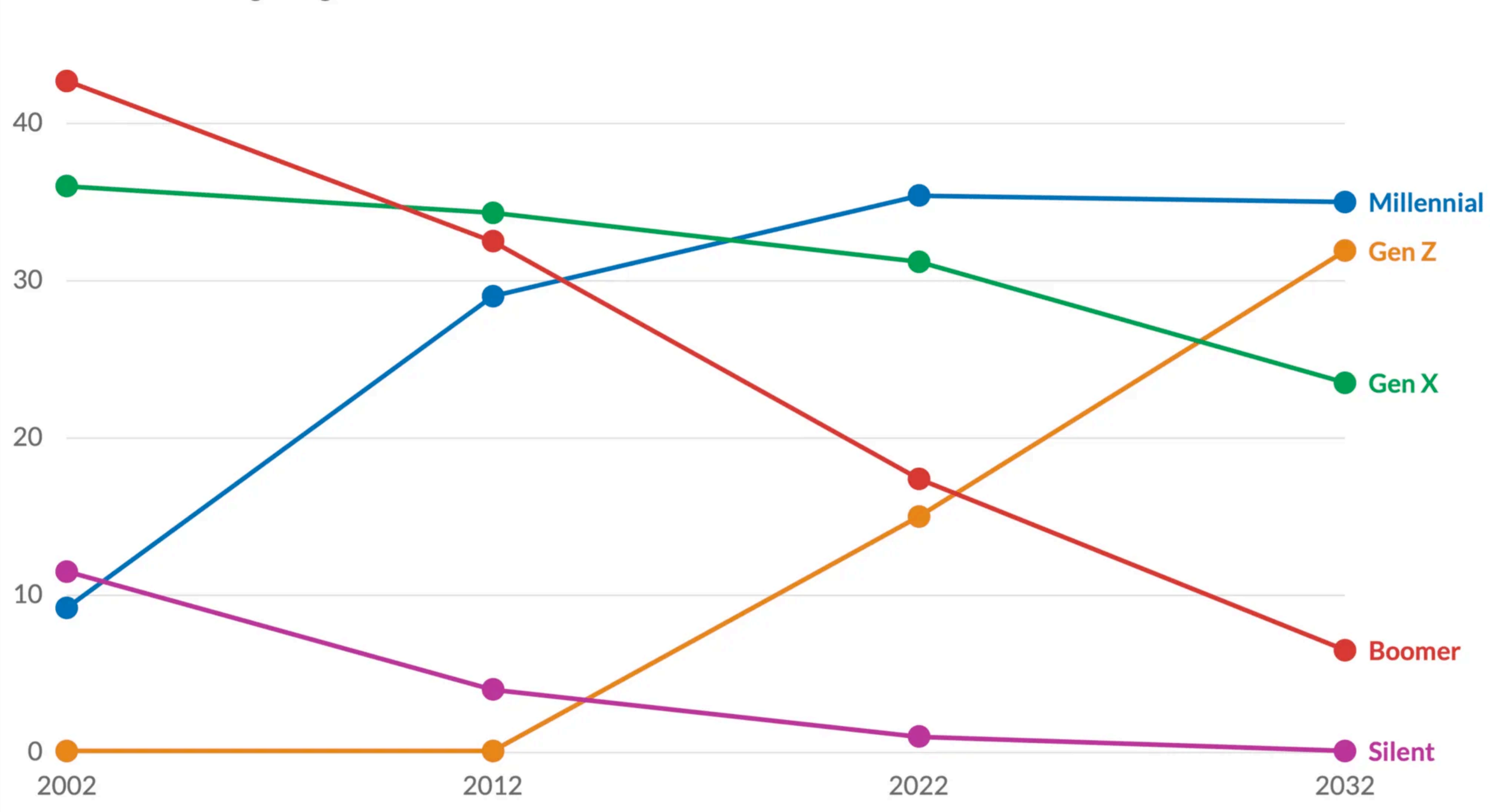
✗ **Label positions are scattered and ambiguous.**



- ✓ **Direct labeling:** reader doesn't need to split attention between the legend and chart
- ✓ **Redundant color encoding** reinforces the relationship between label and line.

Generations as percent of labor force

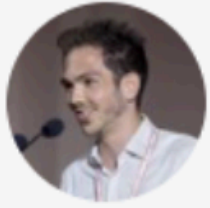
Estimates based on age ranges



✓ **Direct labeling:** reader doesn't need to split attention between the legend and chart

✓ **Redundant color encoding** reinforces the relationship between label and line.

✓ **Spatially align labels** to facilitate reading and skimming.



John Burn-Murdoch @jburnmurdoch · Dec 30, 2023 ...

A good chart doesn't just show, it tells.

“Here's what this means”.

You do this using text. Title, subtitle, labels, annotations. You lose **nothing** by doing this, but add **enormous** value. Your chart is now useful to approximately 1,000x more people than it was before.



4



9



40



4.6K



John Burn-Murdoch @jburnmurdoch · Dec 30, 2023 ...

People who think the purpose of charts is to look nice/colourful/cool/beautiful make a very similar mistake.

Great that looks lovely but just FYI most people's takeaway from looking at it is “ooh lovely picture”, not “ah interesting I see”.

I used to make this mistake a lot!



2



16

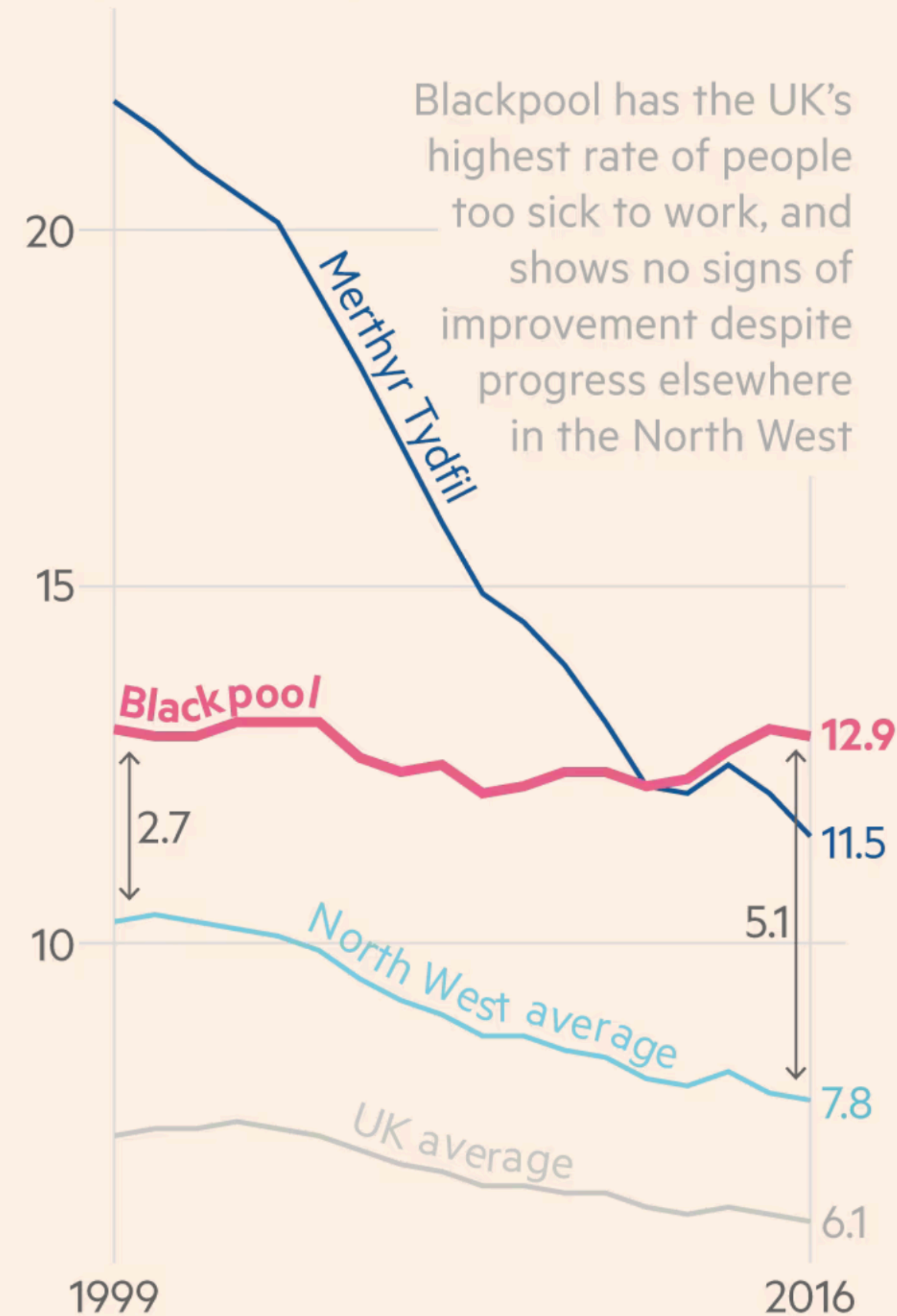


1.5K



Economically and physically, **Blackpool** is unwell

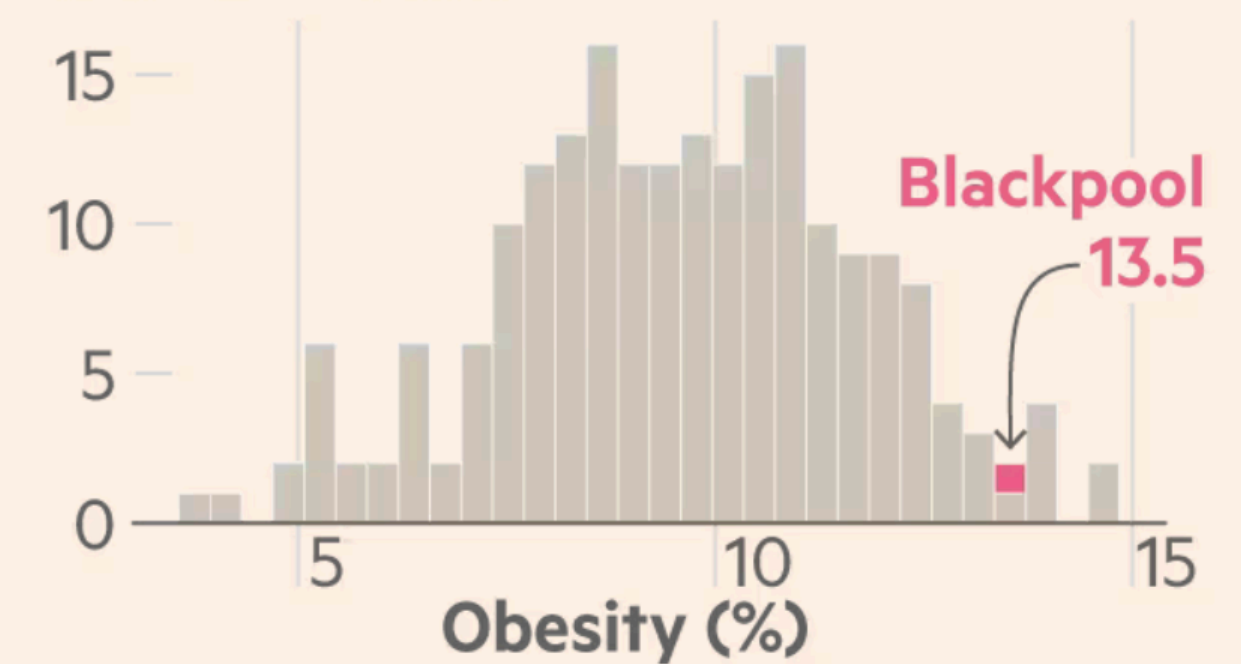
Percentage of working-age population in receipt of incapacity benefits (IB) or employment support allowance (ESA)



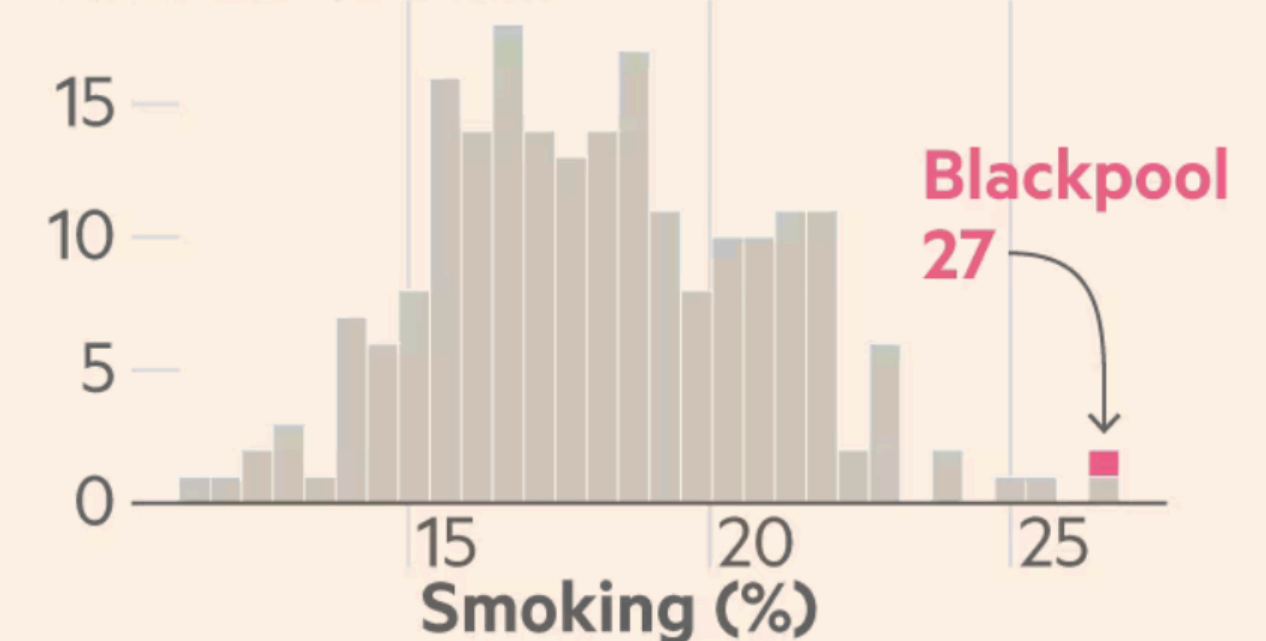
Source: Department for Work and Pensions, Public Health England
Graphic by John Burn-Murdoch / @jburnmurdoch
© FT

Blackpool has some of England's highest rates of obesity and smoking, and the top mortality rate from alcoholic liver disease

Number of areas



Number of areas



Number of areas

