Jeff Tao

2/4/2026

CIS 7000: Foundations & Frontiers of HCI

Data Visualization

speculative.tech

Jeff Tao

CIS PhD Student (3rd year)

Current Research: Data-oriented tools for debugging code, ML-based query optimization

Previously: MS Thesis @ Columbia
Thesis topic: *An Execution Engine for Physical Visualization Design*

Pre-previously: 5 years senior software engineer @ MongoDB, Microsoft

Why Study Vis?

Microcosm of HCI

- Perception
- Cognition
- Systems
  - Languages/Formalisms
  - Data transformation
  - Novel interactions/encodings
- Interaction
- Emotion/Affect
- Design conventions
- Art
- Conceptual Frameworks (taxonomies, design dimensions)
- Experimental Design/Evaluation
- Tasks/Workflows
- Context of Use (science, journalism, business analytics, public health, data science)
- Users (scientists, journalists, programmers, data scientists, novices)

Why Study Vis?

Microcosm of HCI

- Perception
- Cognition
- Systems
    - Languages/Formalisms
    - Data transformation
    - Novel interactions/encodings
- Interaction
- Emotion/Affect
- Design conventions
- Art
- Conceptual Frameworks (taxonomies, design dimensions)
- Experimental Design/Evaluation
- Tasks/Workflows
- Context of Use (science, journalism, business analytics, public health, data science)
- Users (scientists, journalists, programmers, data scientists, novices)

Why Visualize?

Your Brain is Good at Pattern Recognition

[10]:

	x	y
0	55.3846	97.1795
1	51.5385	96.0256
2	46.1538	94.4872
3	42.8205	91.4103
4	40.7692	88.3333
...	...	...
137	39.4872	25.3846
138	91.2821	41.5385
139	50.0000	95.7692
140	47.9487	95.0000
141	44.1026	92.6923

142 rows × 2 columns

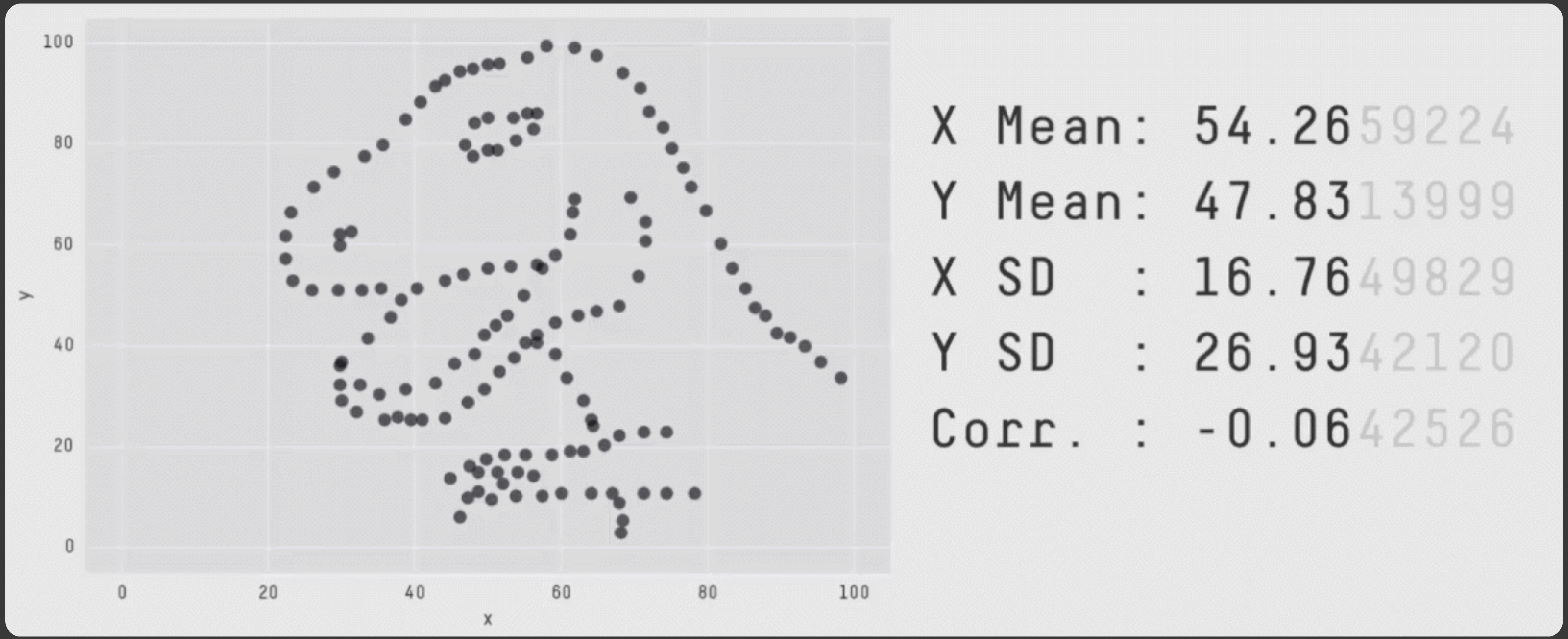
Matejka & Fitzmaurice, CHI 2017

Your Brain is Good at Pattern Recognition

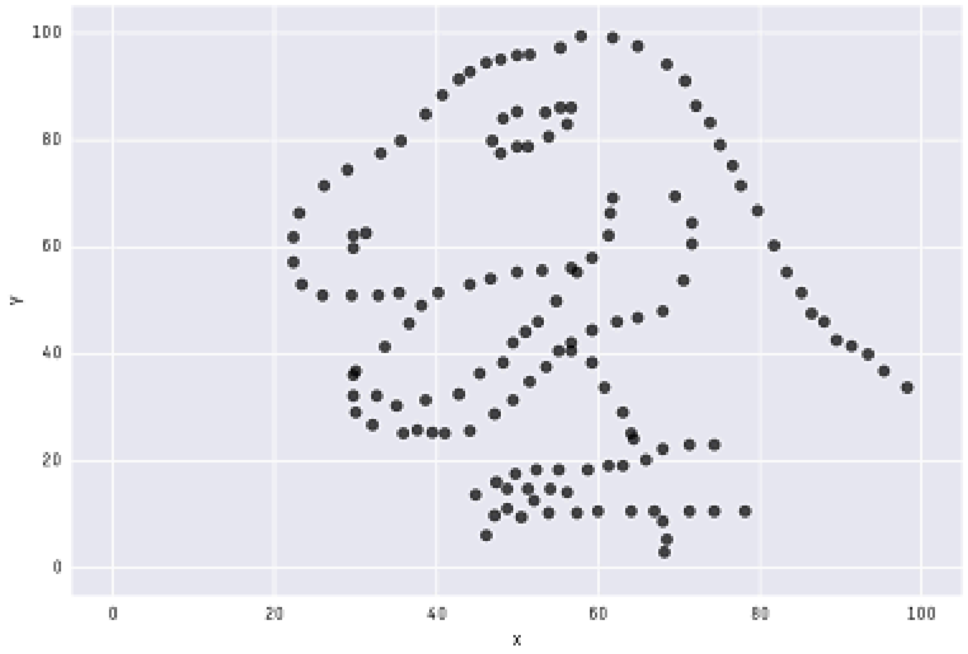
```
X Mean: 54.2659224
Y Mean: 47.8313999
X SD  : 16.7649829
Y SD  : 26.9342120
Corr. : -0.0642526
```

Matejka & Fitzmaurice, CHI 2017

Your Brain is Good at Pattern Recognition



Matejka & Fitzmaurice, CHI 2017



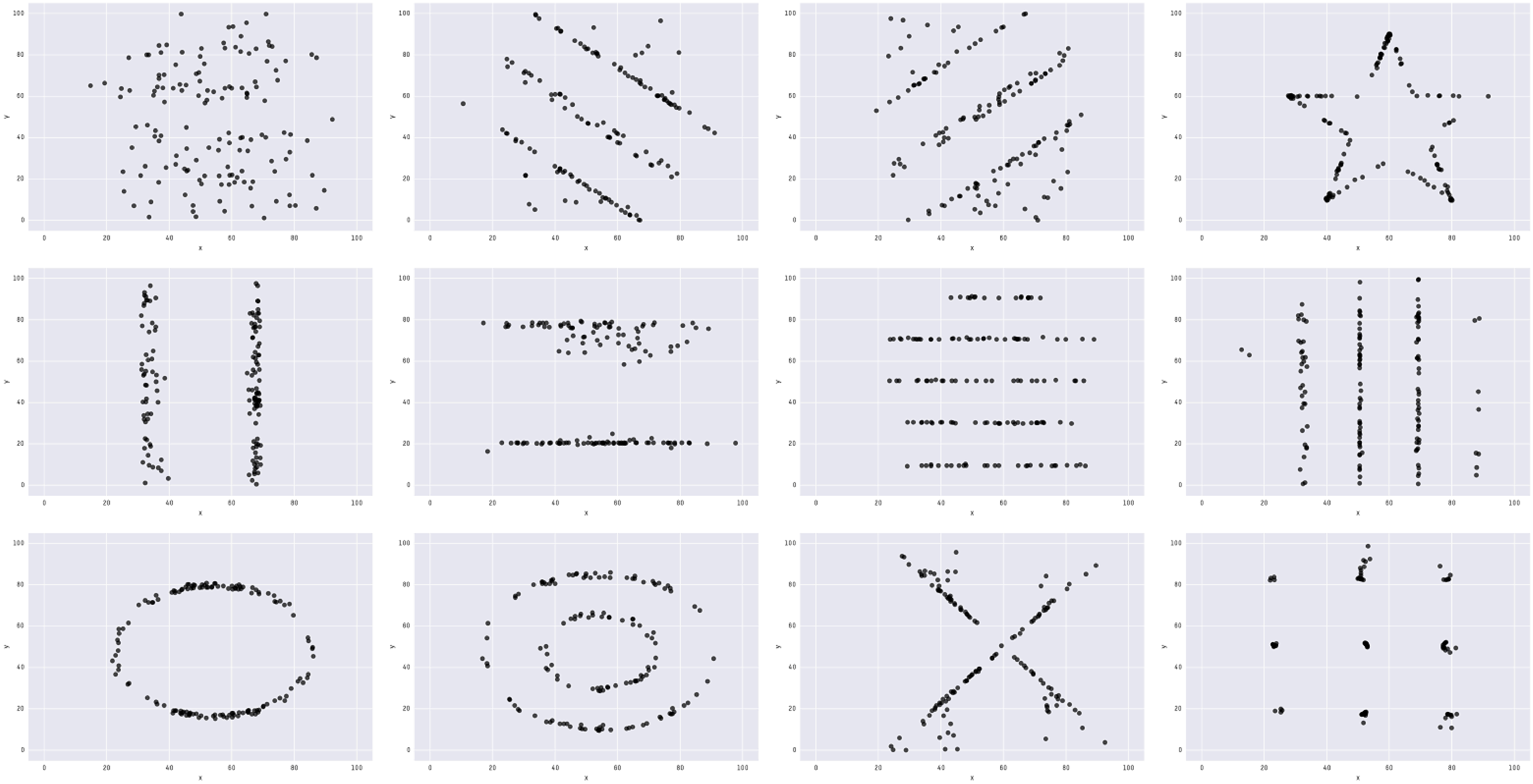
X Mean: 54.26

Y Mean: 47.83

X SD : 16.76

Y SD : 26.93

Corr. : -0.06



Use Cases

To Explore

Your brain is really good at noticing visual patterns!

Use your eyes helps you discover interesting things.

To Debug

“What changed about the data after I ran this code?”

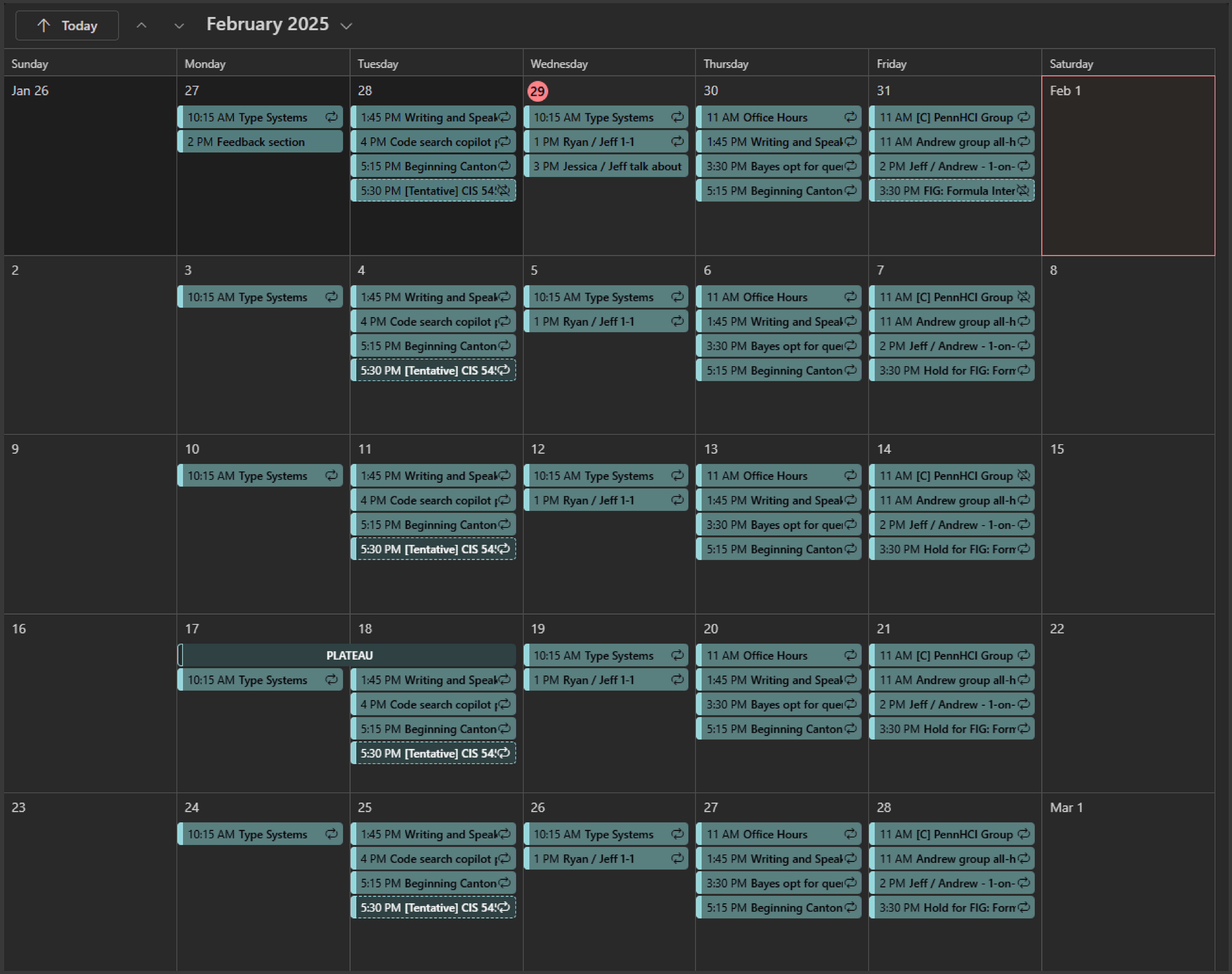
Visual summaries help you decide if the data align with your expectations.

To Present

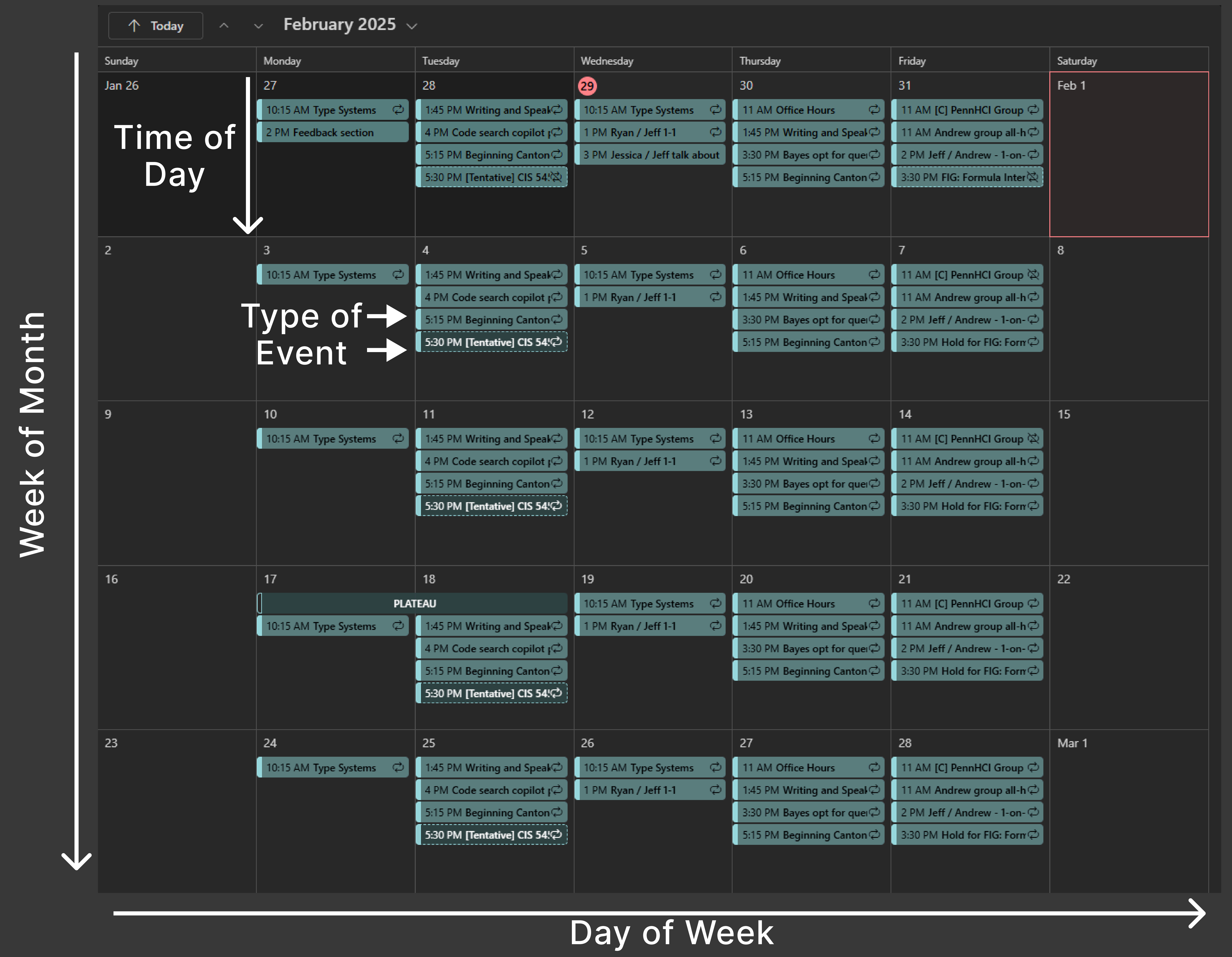
Help *other people* understand what you found in the data.

Tell a *story* about data while letting the viewer verify it.

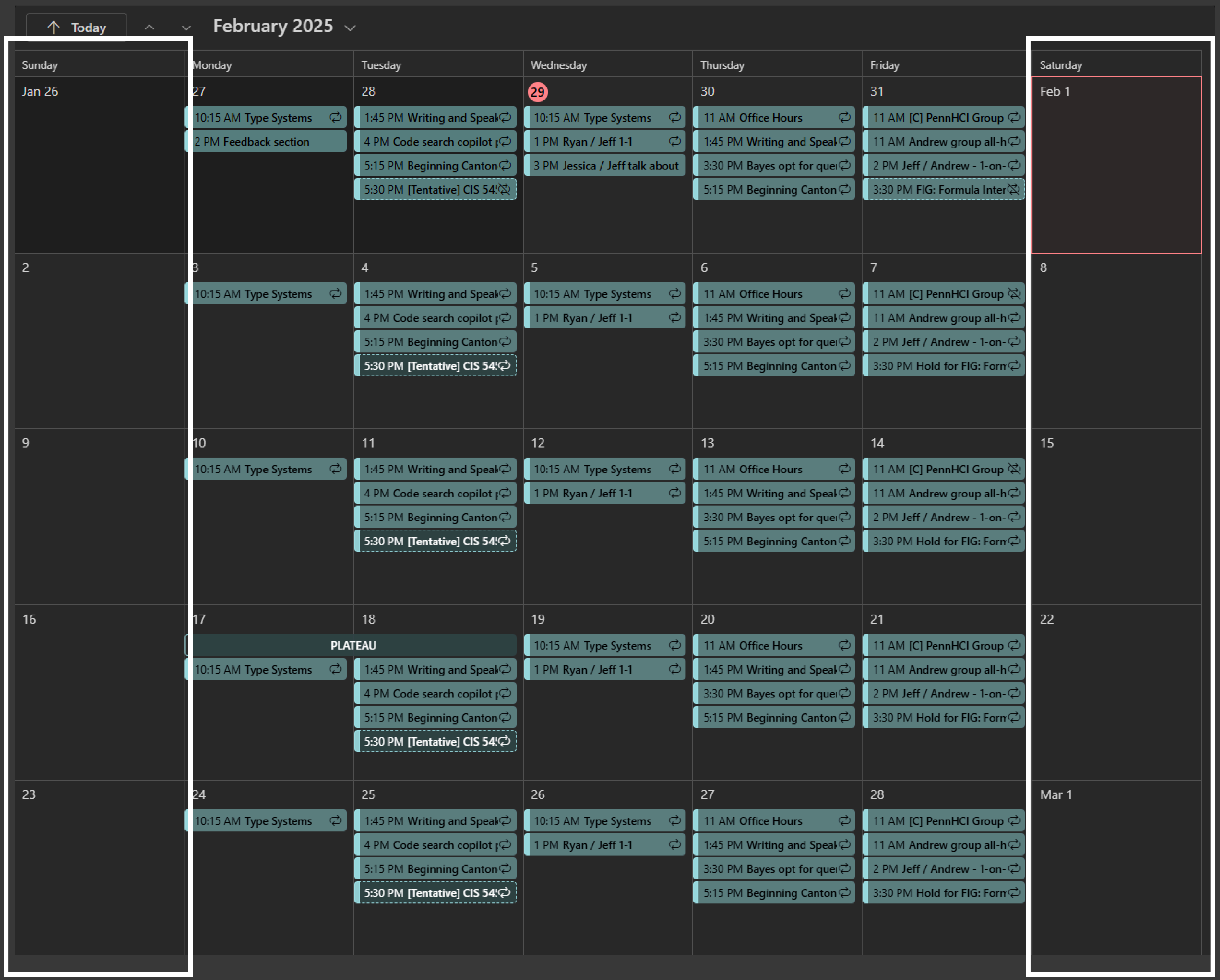
Visualizations Are Everywhere!



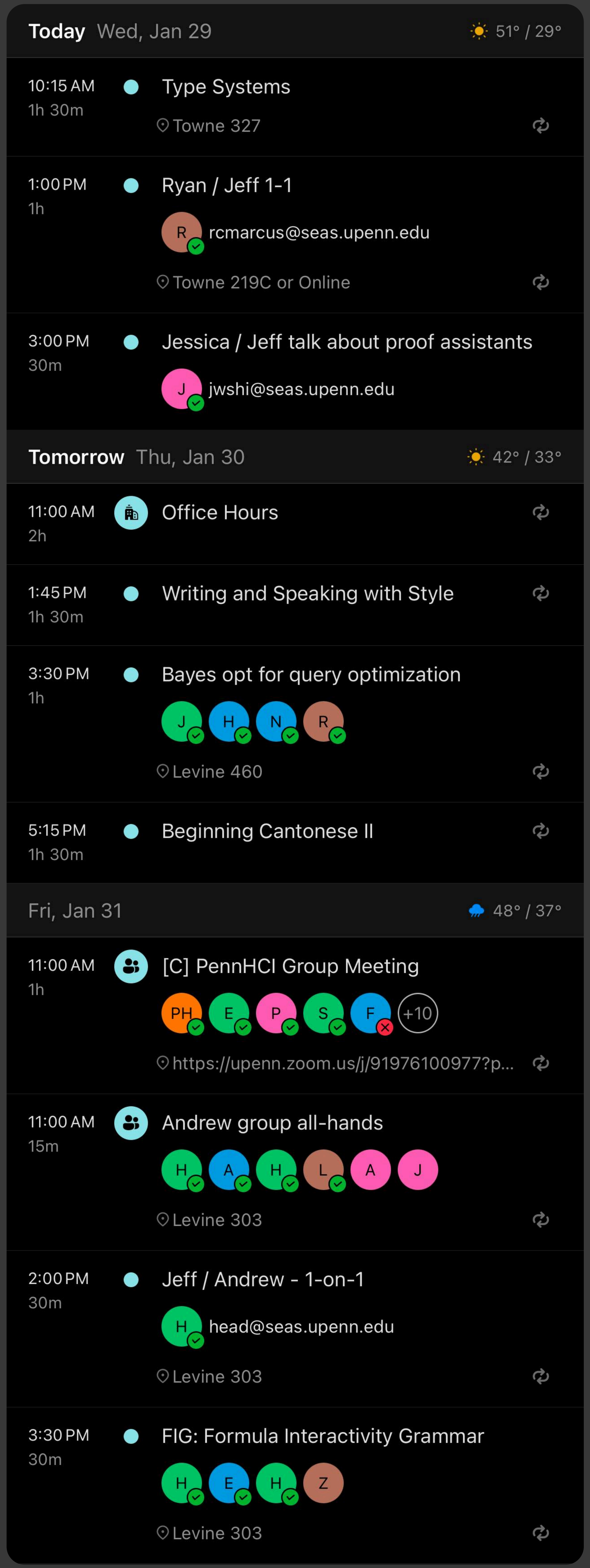
Visualizations Are Everywhere!



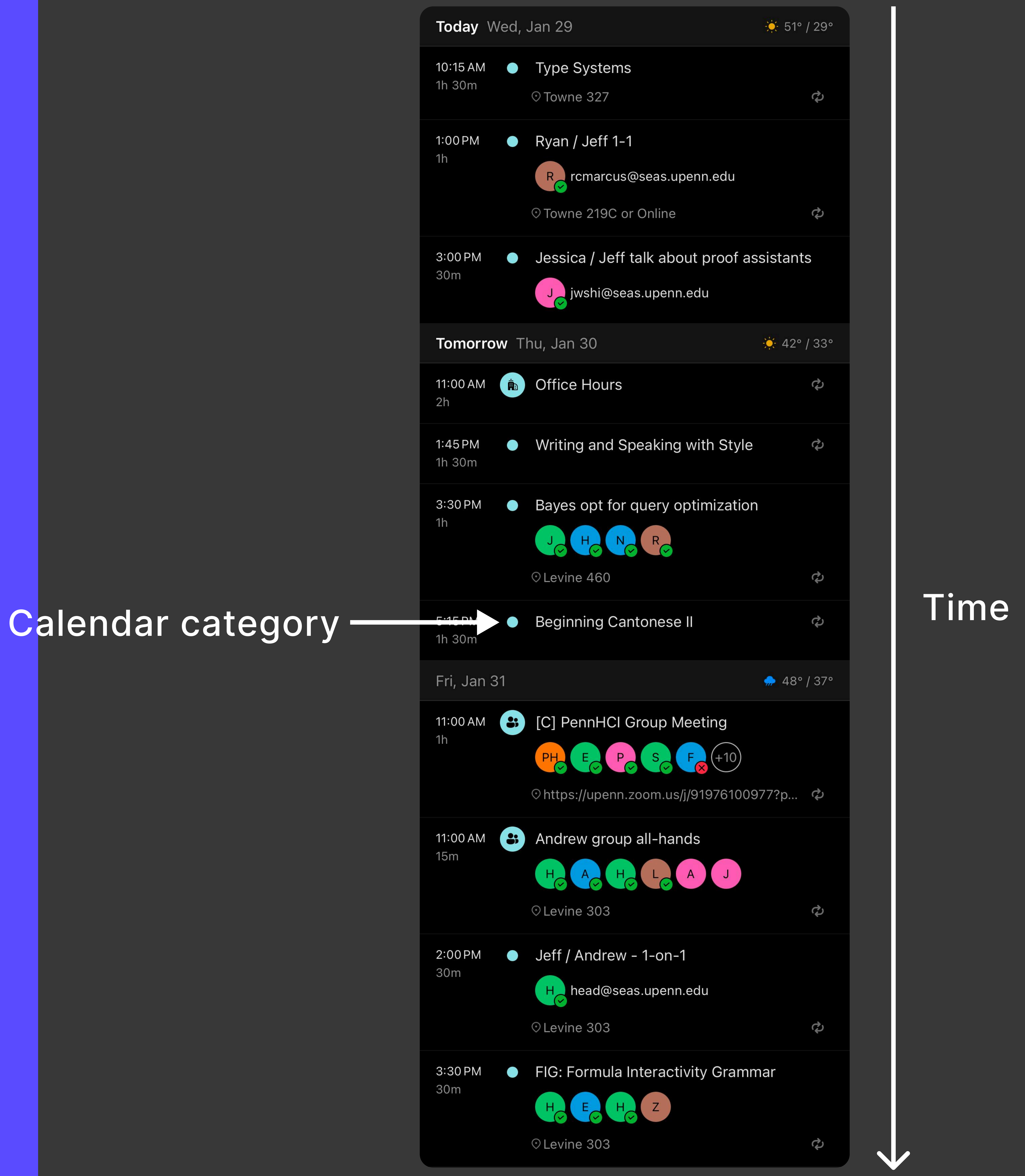
Visualizations Are Everywhere!



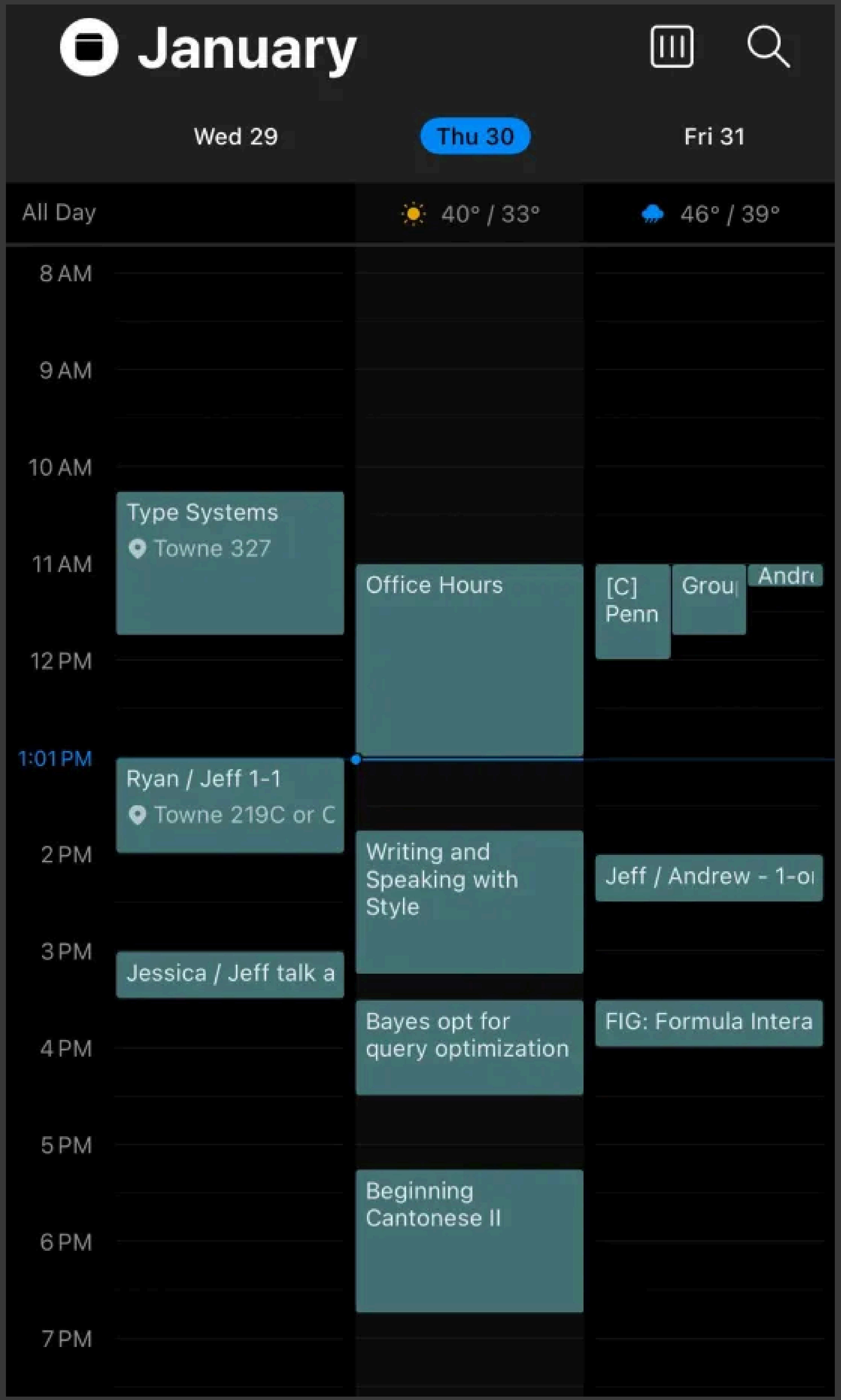
Visualizations Are Everywhere!



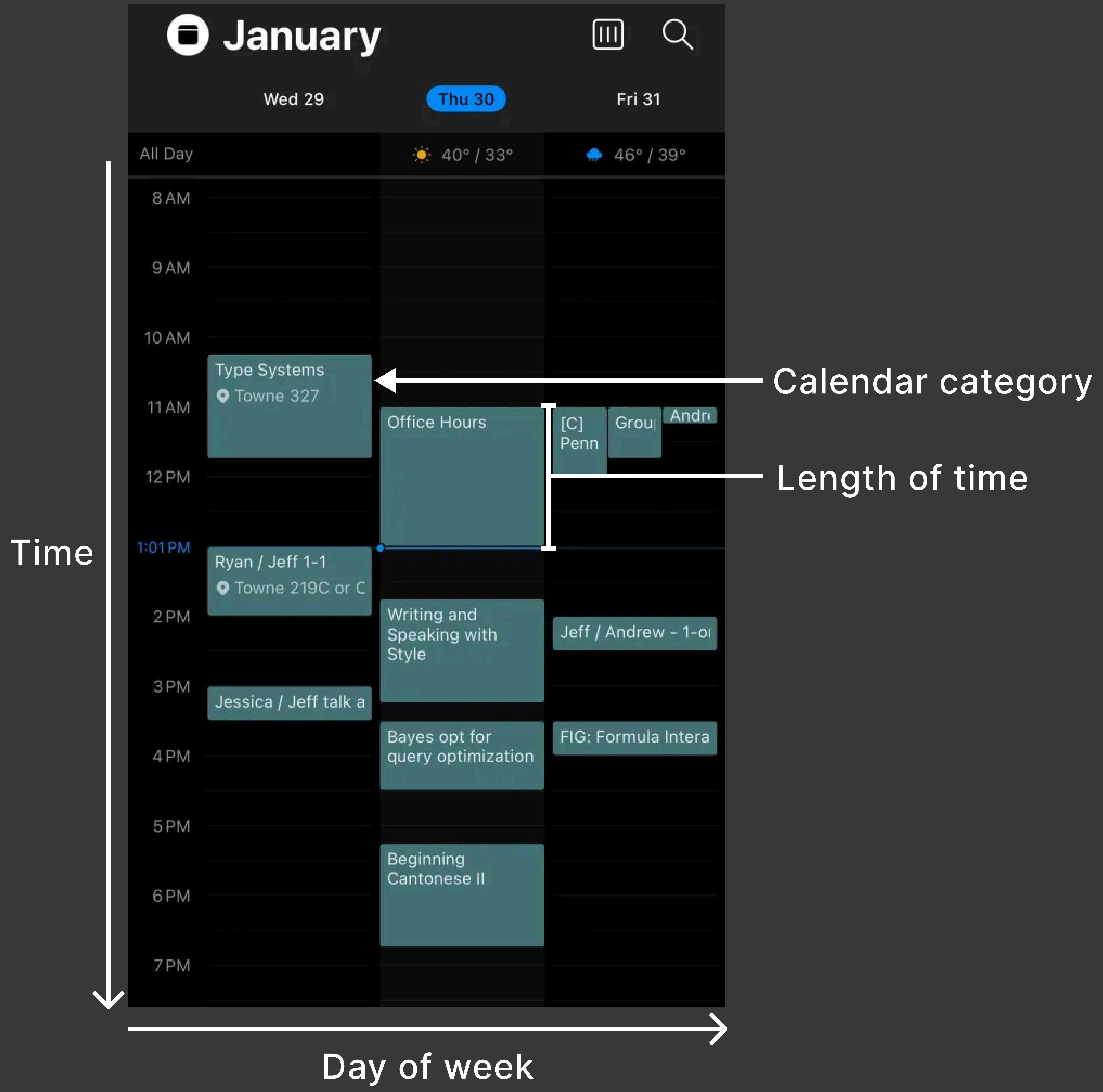
Visualizations Are Everywhere!



Visualizations Are Everywhere!



Visualizations Are Everywhere!



Visualization leverages your *perception* and *pattern recognition* to help explain data.

How to Visualize?

What Type of Data?

Your data has a *type*.

Each visualization pattern only works with certain types of data.

Nominal/Categorical

Can only compare equal/non-equal

- States, Countries
- Type of Student (undergrad, masters, phd)
- Basically anything that isn’t written with numbers
- Could be hierarchical (country → state → city)

Ordinal

Having an order

- Olympic medals
- Likert scales (strongly disagree ↔ strongly agree)
- Level of education

Quantitative

Inherently numeric, you can do math to it

- $ of Revenue
- # of customers
- Intervals
- Ratios

Temporal

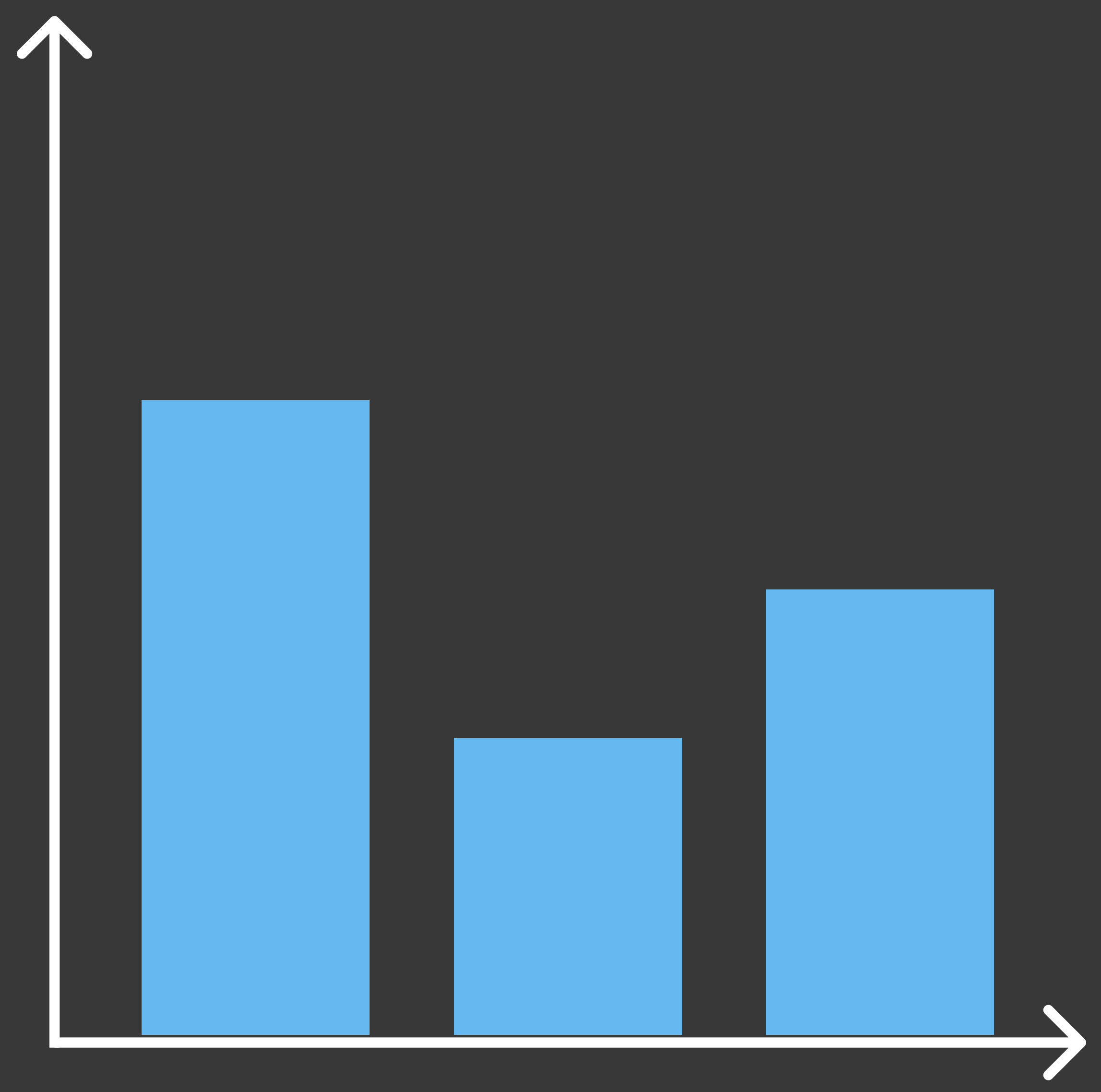
Having to do with dates/time

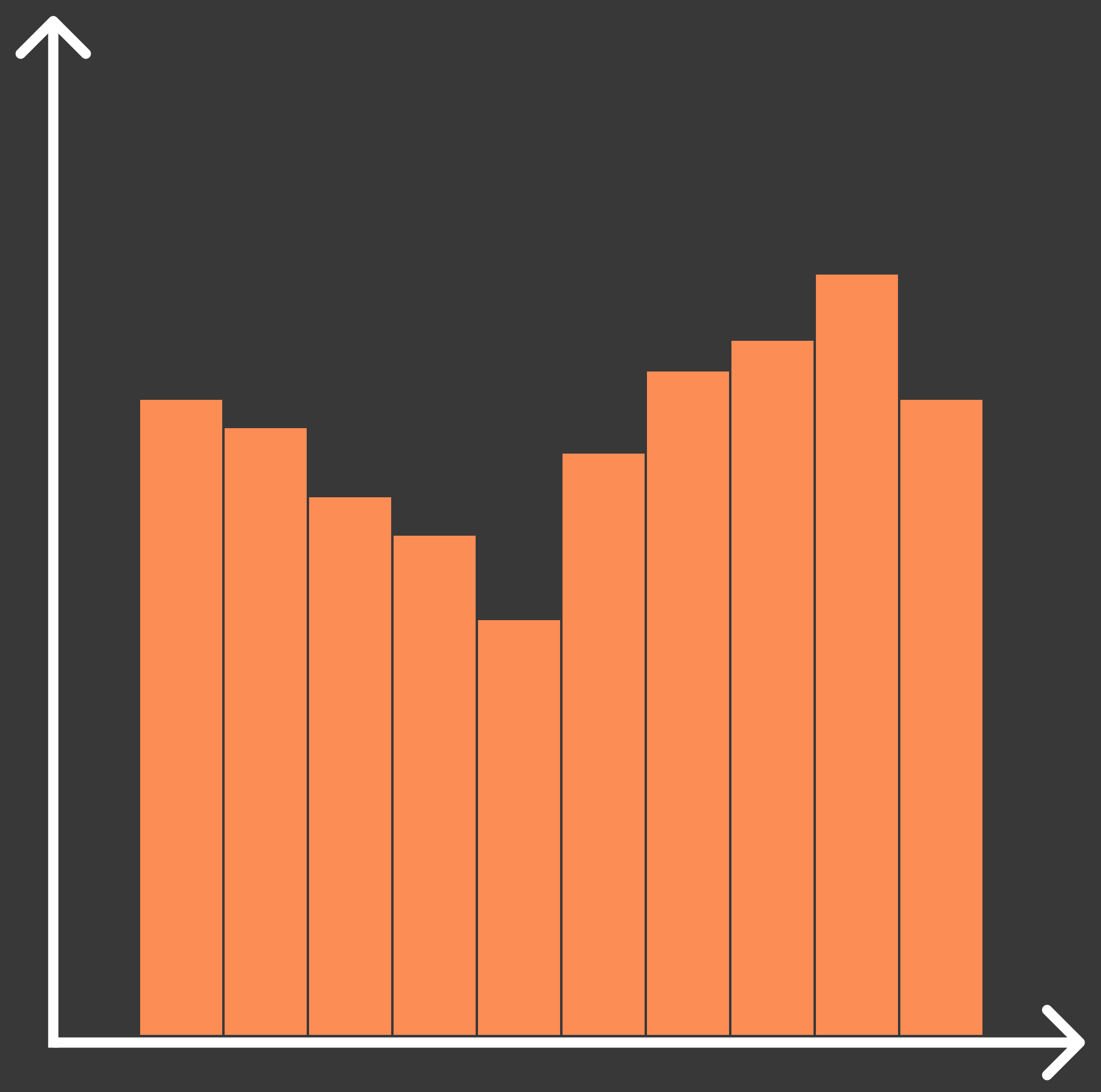
- Year/Month/Day
- Fiscal quarter
- School year
- Ranges/Duration

Marks

Graphical shapes that represents some part of the data

Bars: Compare Position, Length

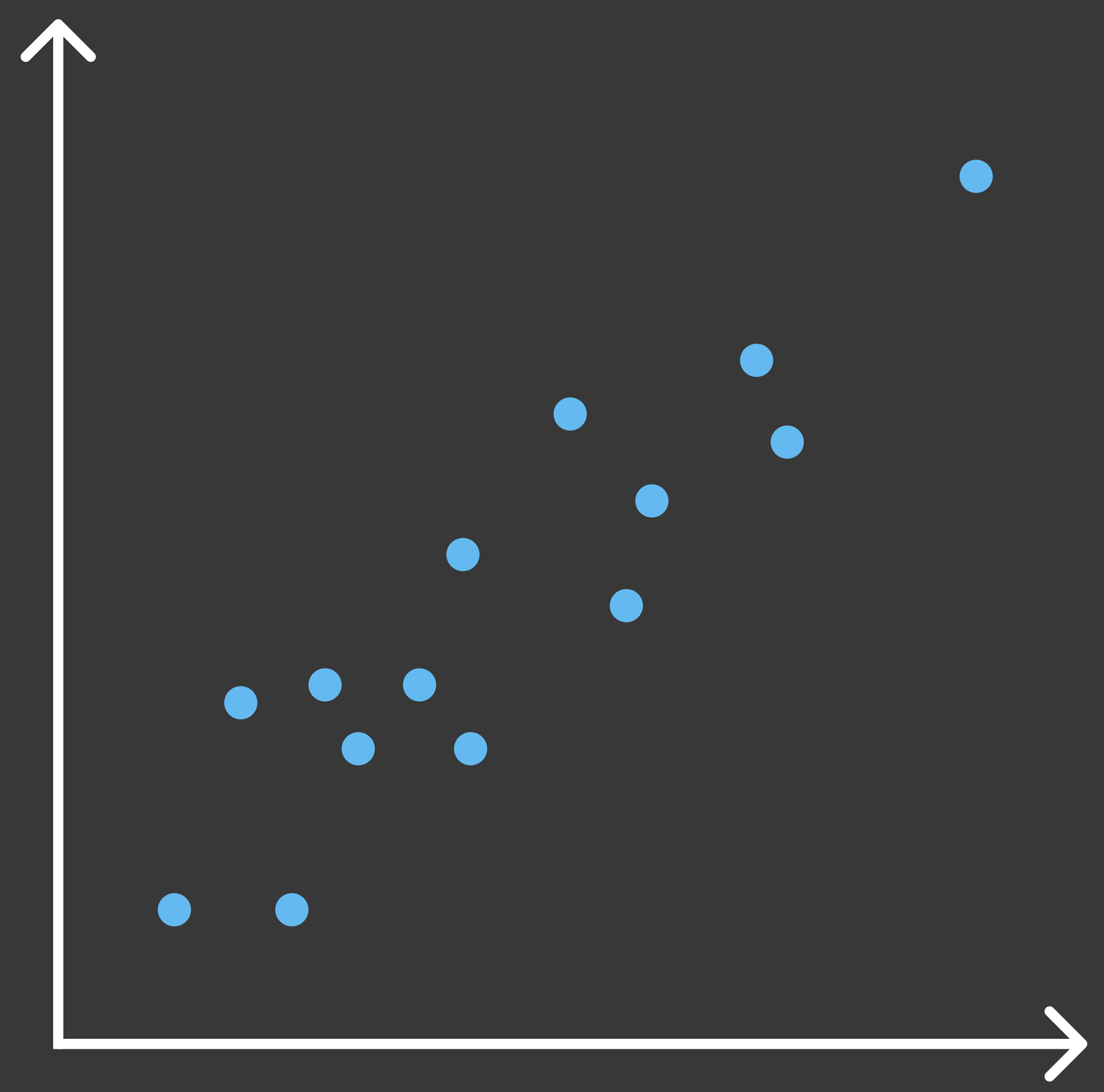


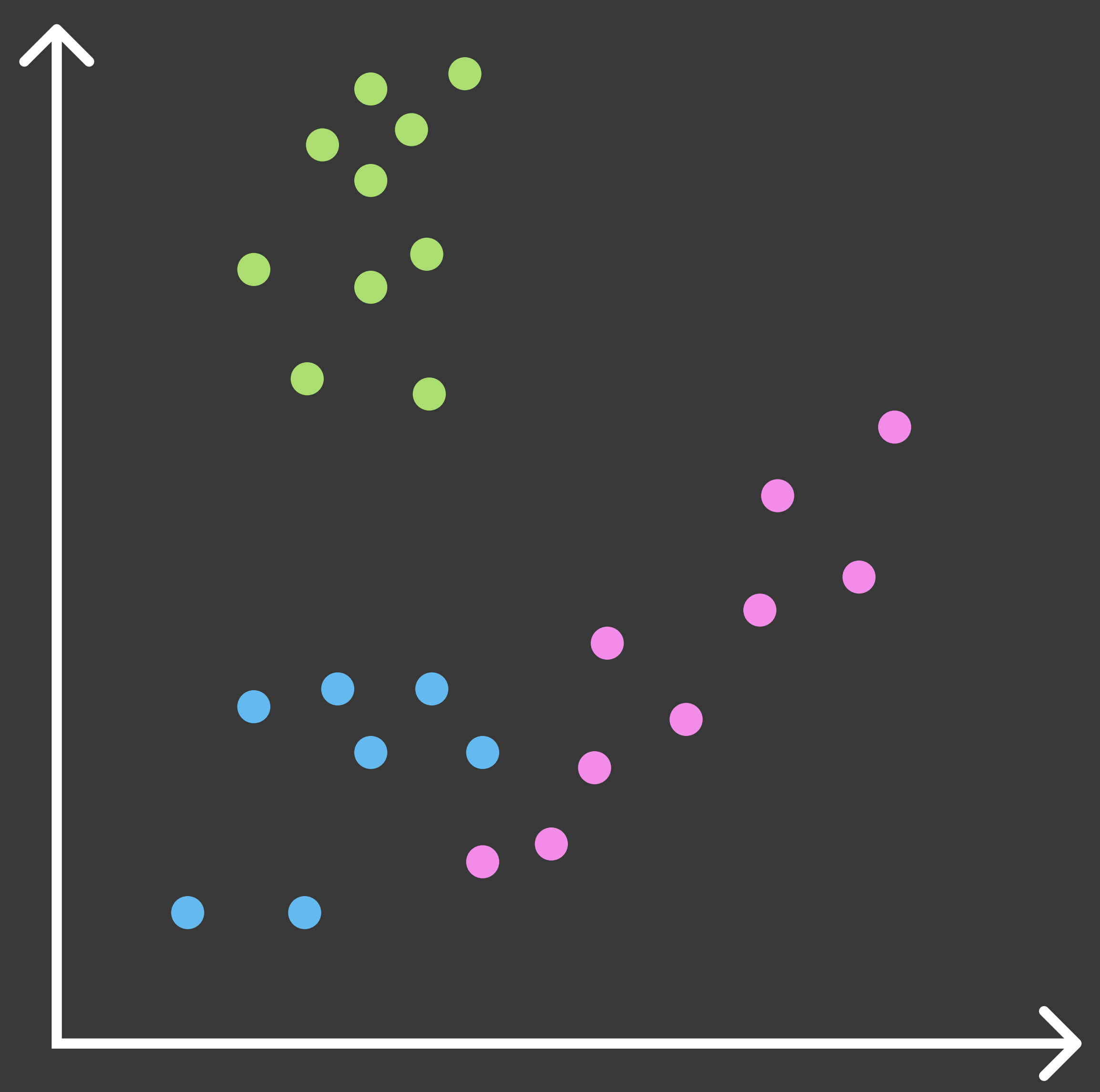


Marks

Graphical shapes that represents some part of the data

Points: Compare Position

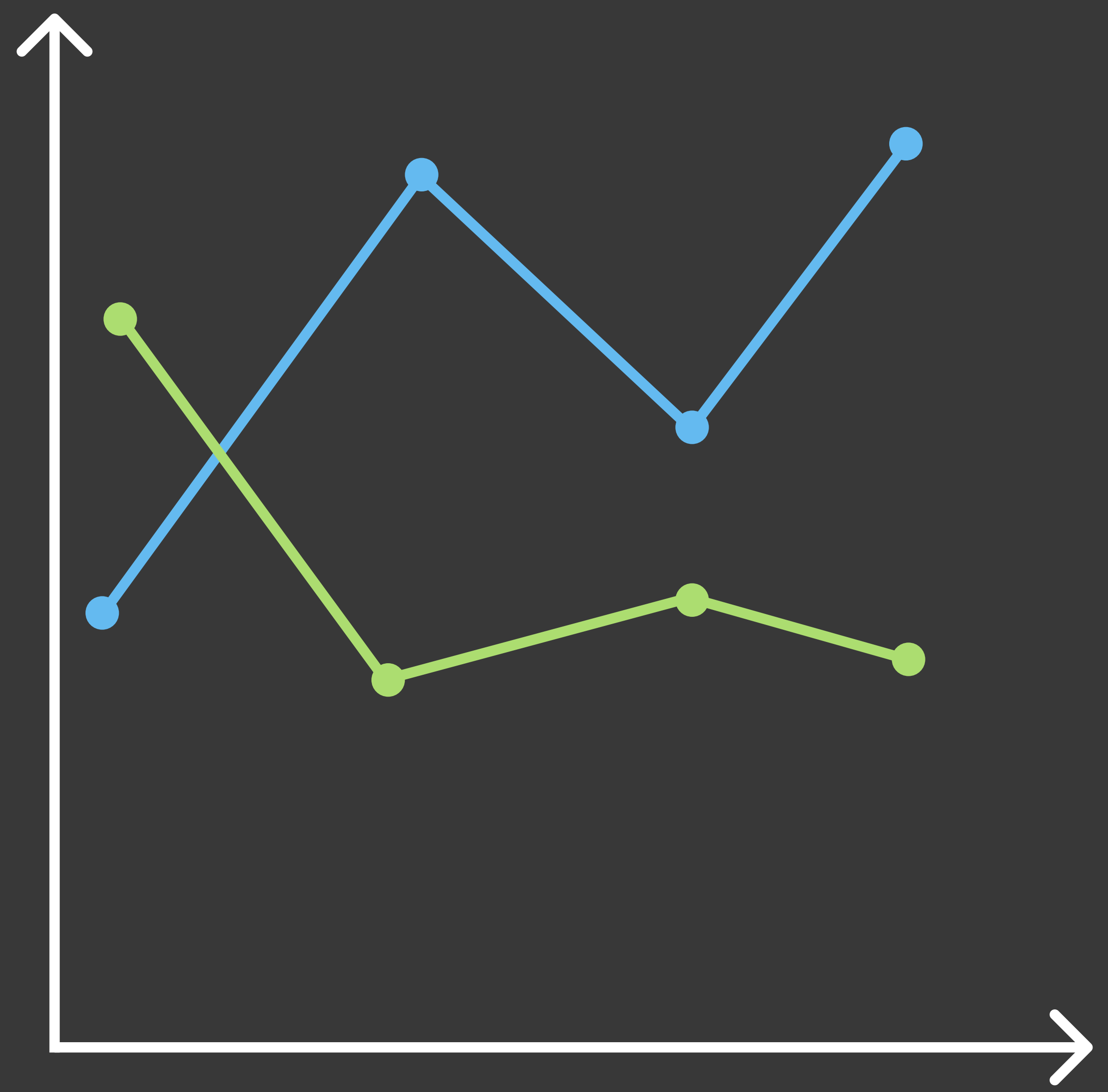




Marks

Graphical shapes that represents some part of the data

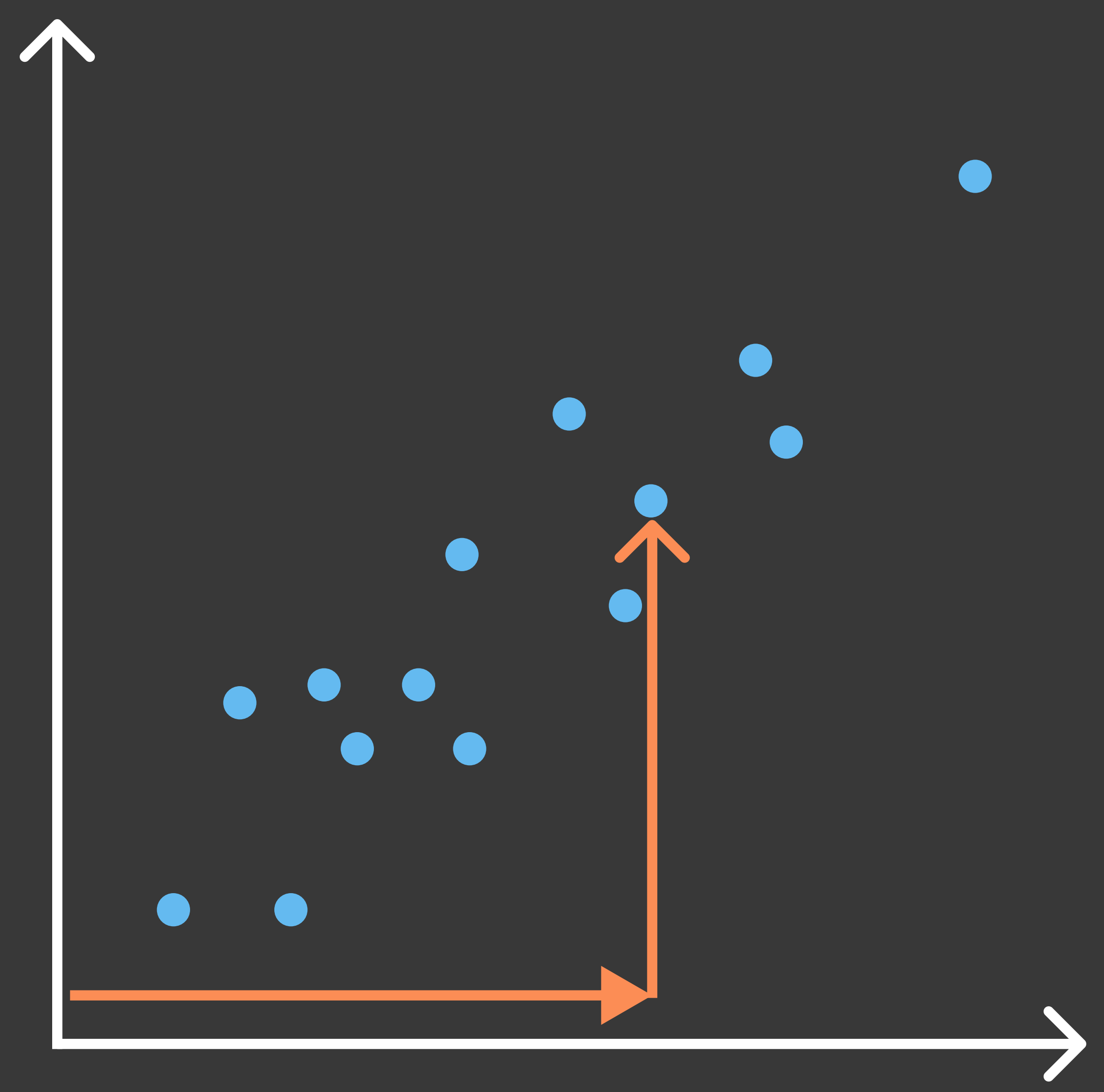
Lines: Compare Position, Length



Encodings

How data maps to perceptual “channels”

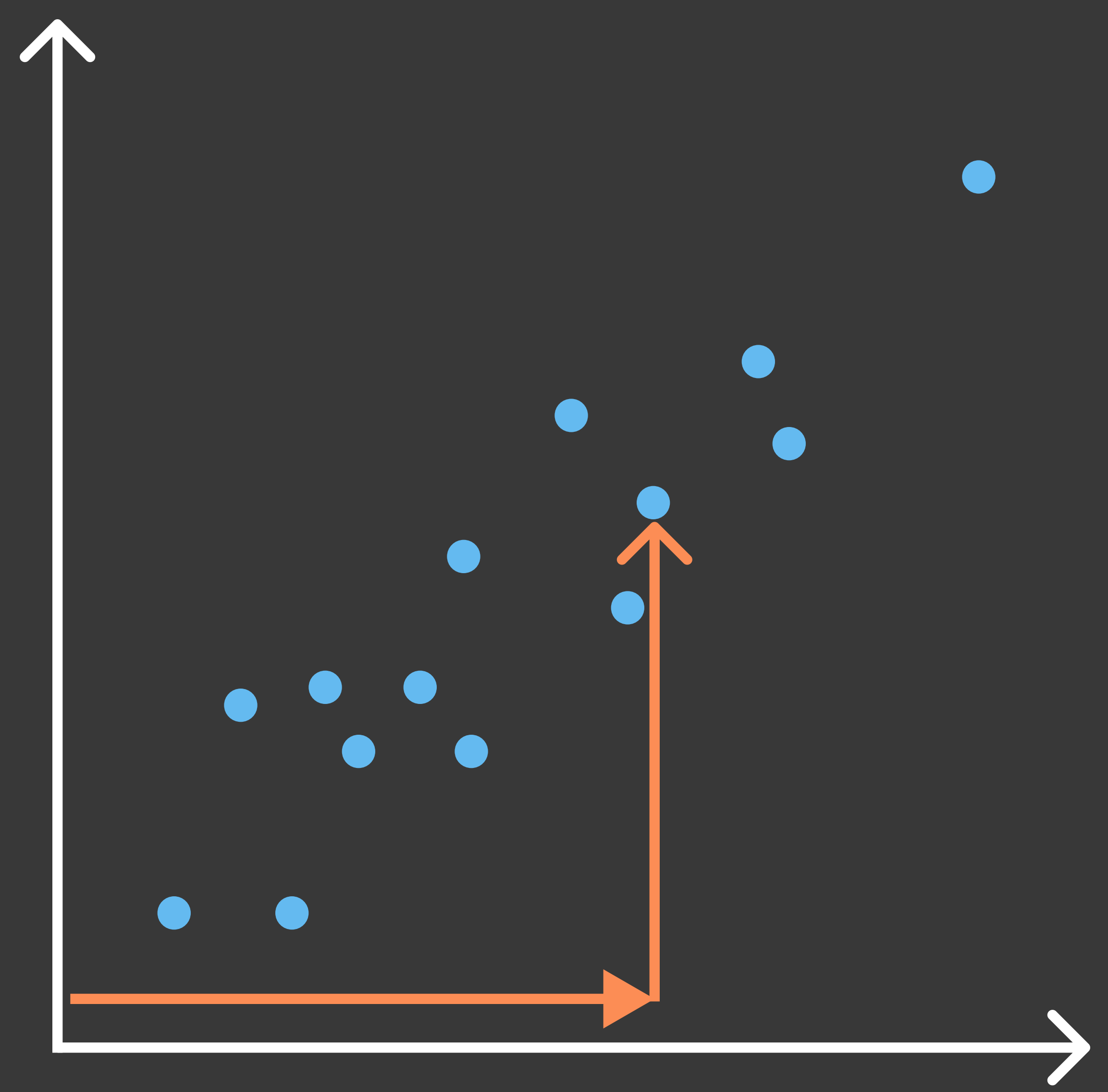
Position: Quantitative, Categorical

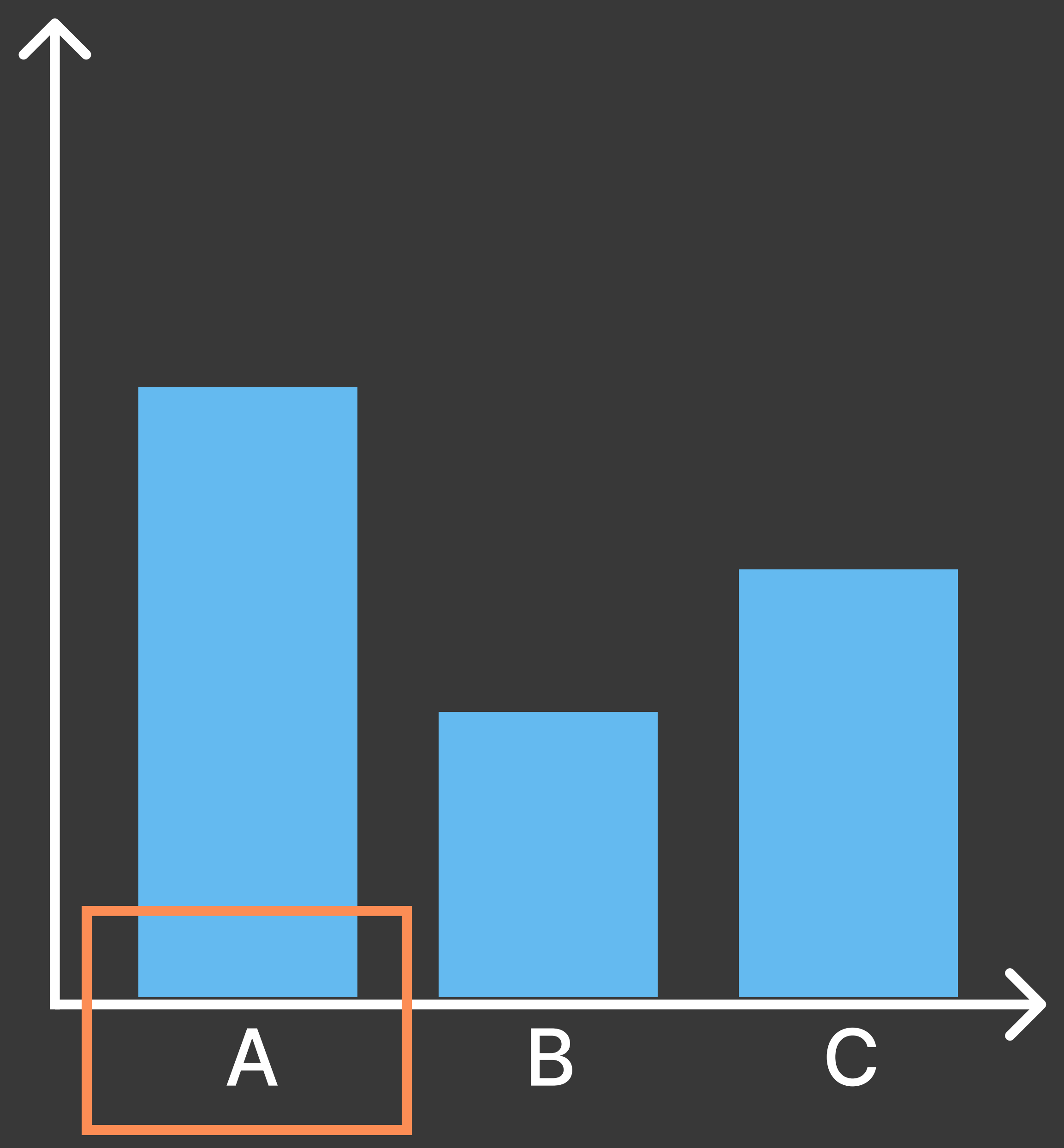


Encodings

How data maps to perceptual “channels”

Position: Quantitative, Categorical

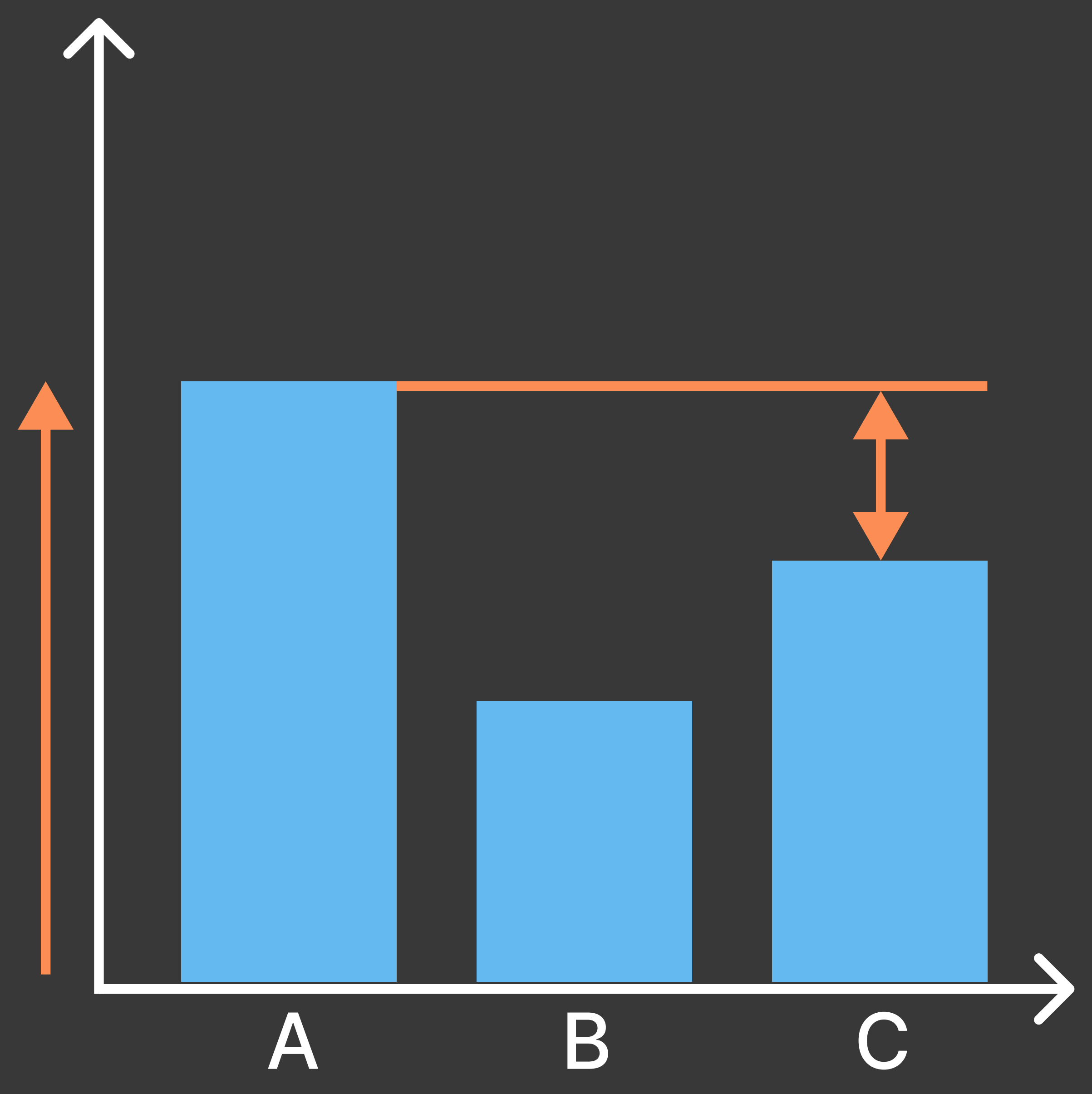




Encodings

How data maps to perceptual “channels”

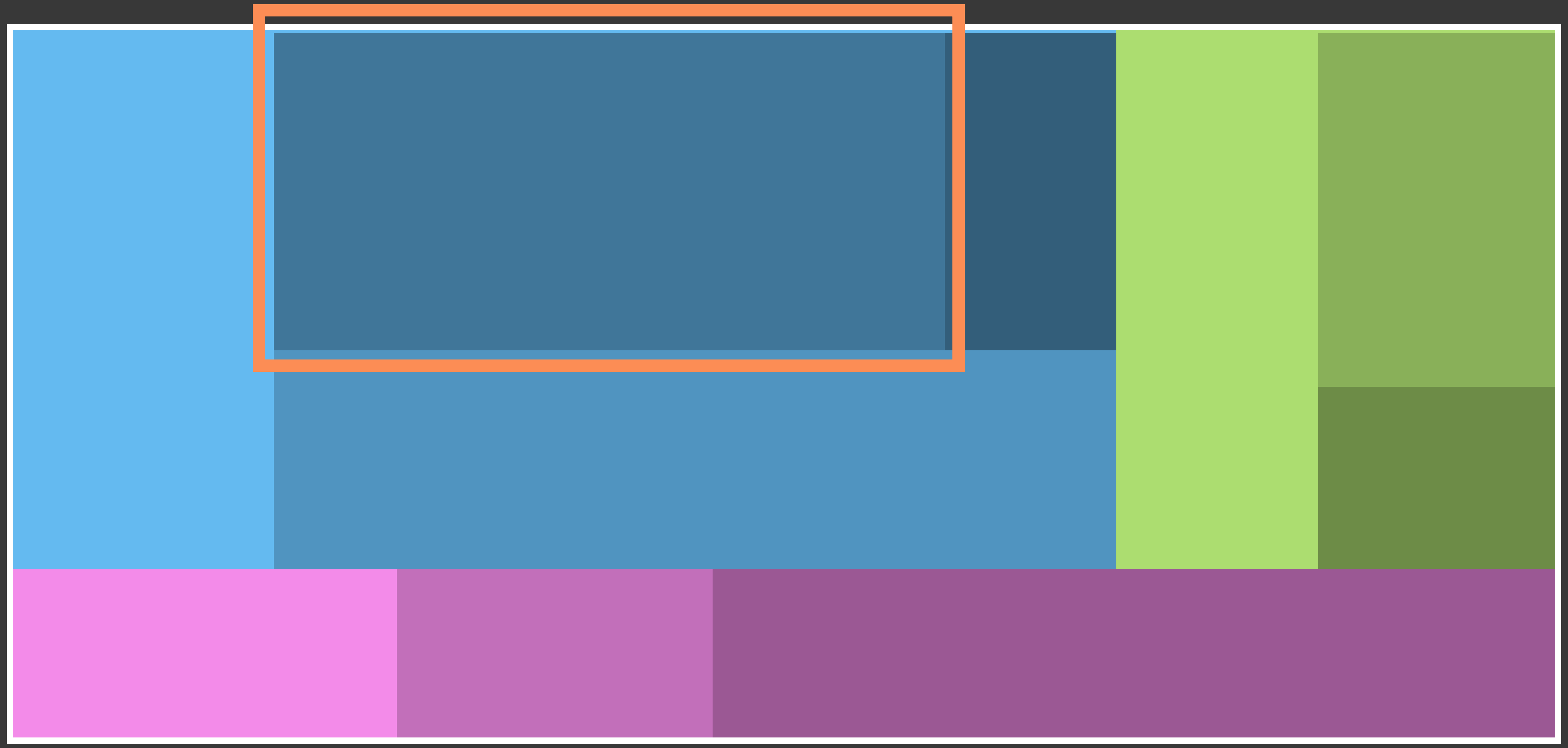
Length: Quantitative

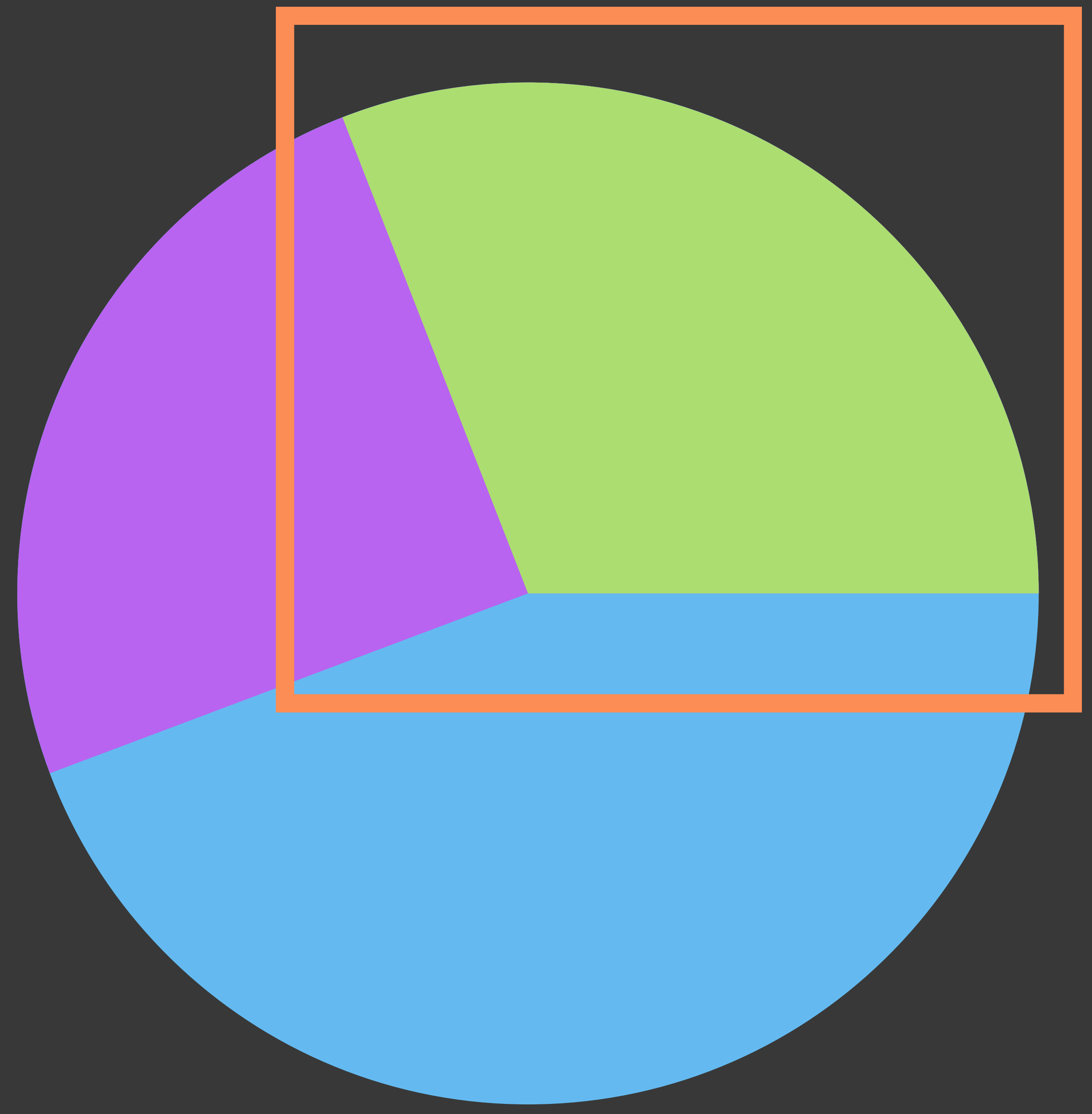


Encodings

How data maps to perceptual “channels”

Area: Quantitative

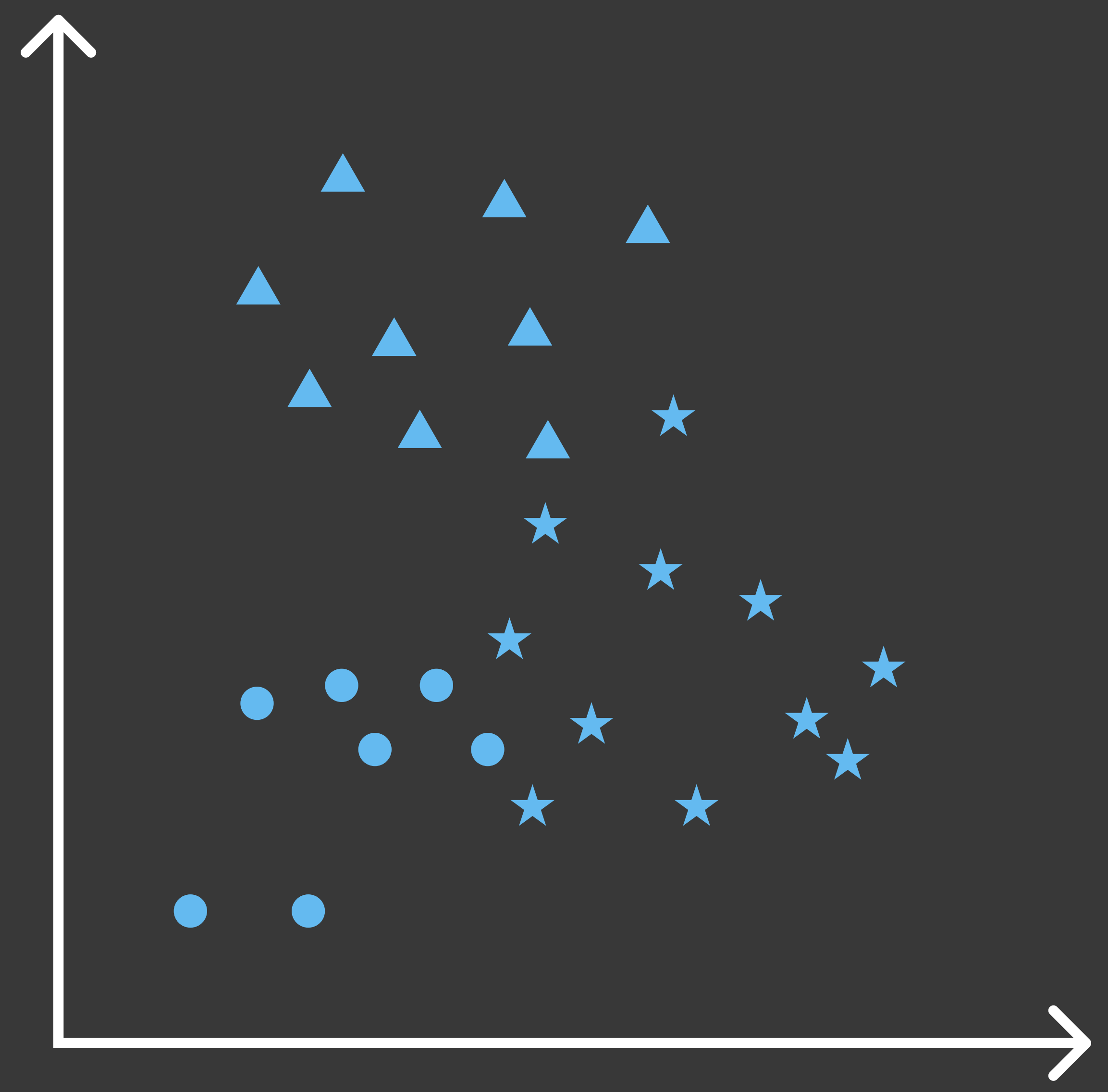


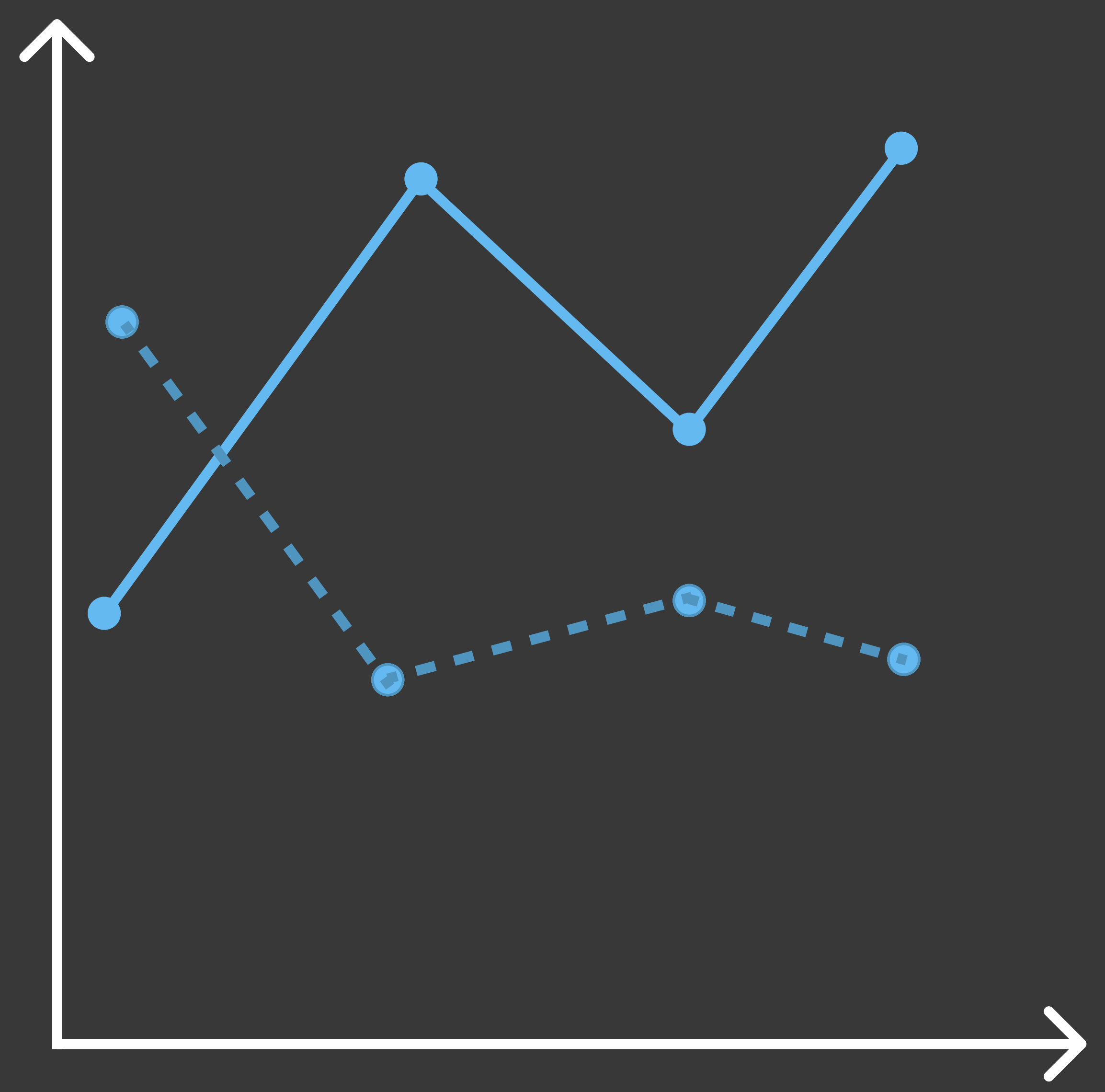


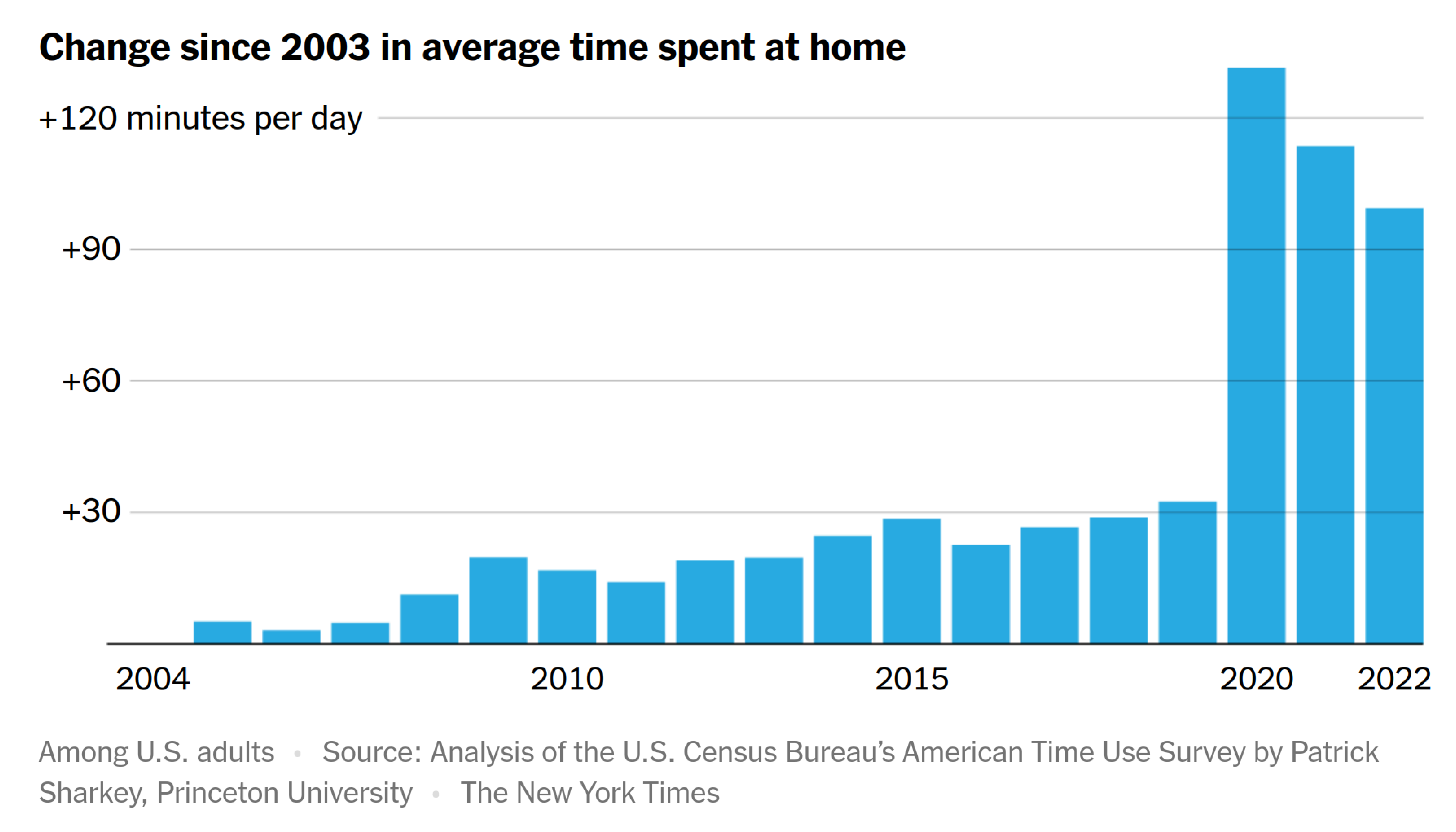
Encodings

How data maps to perceptual “channels”

Shape, Style: Categorical



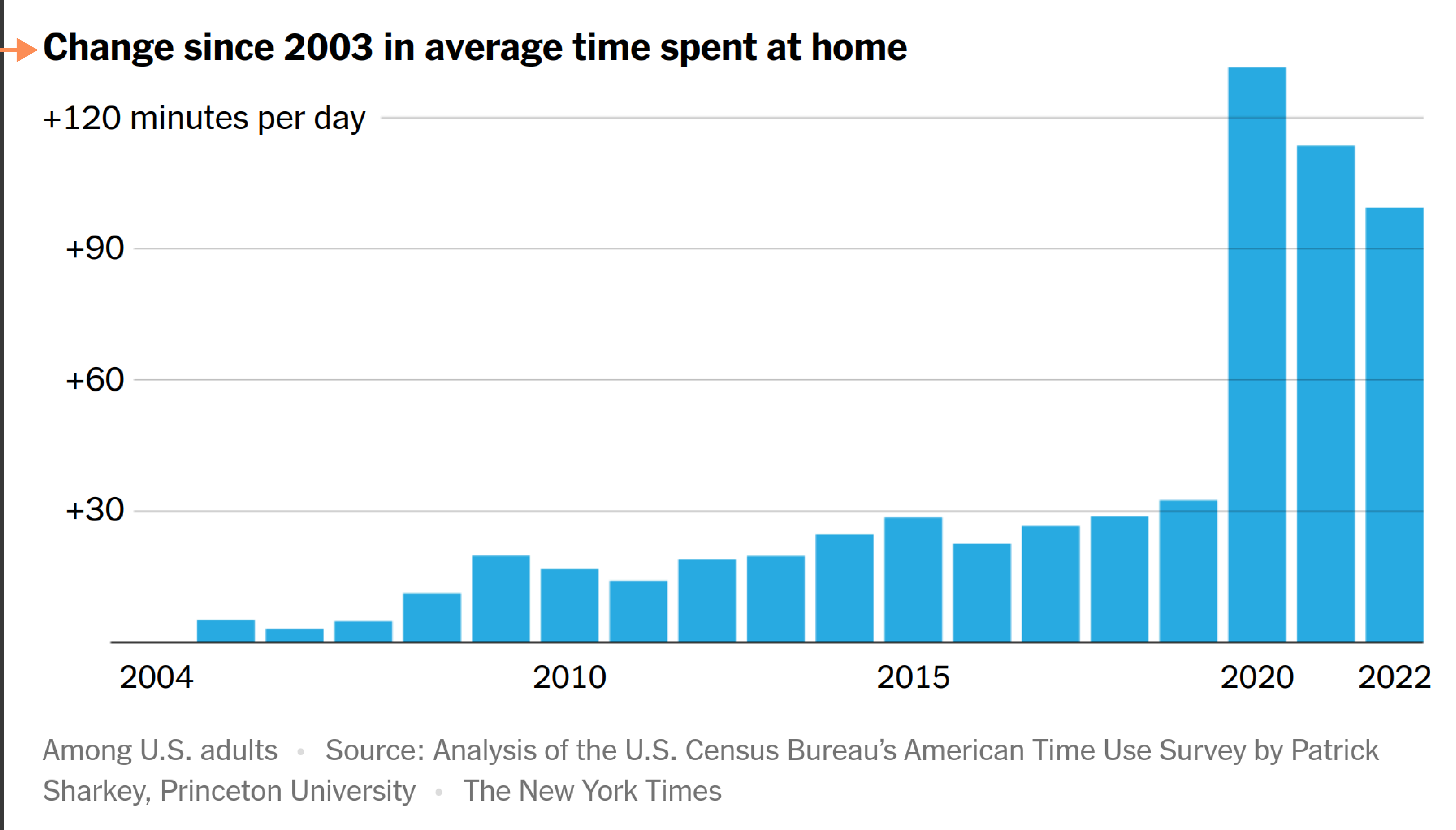




Anatomy of a Chart

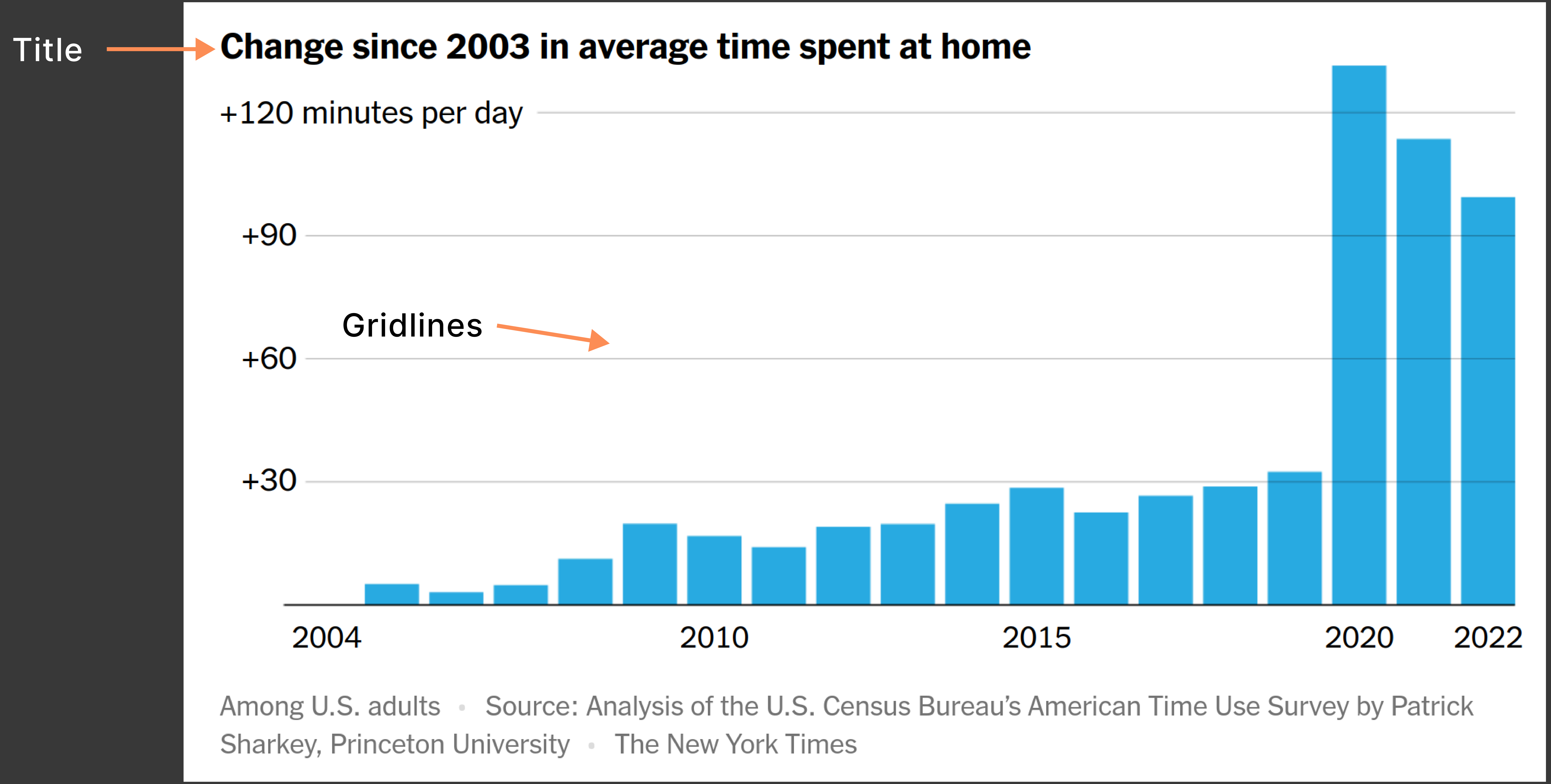
Source: NYT

Title



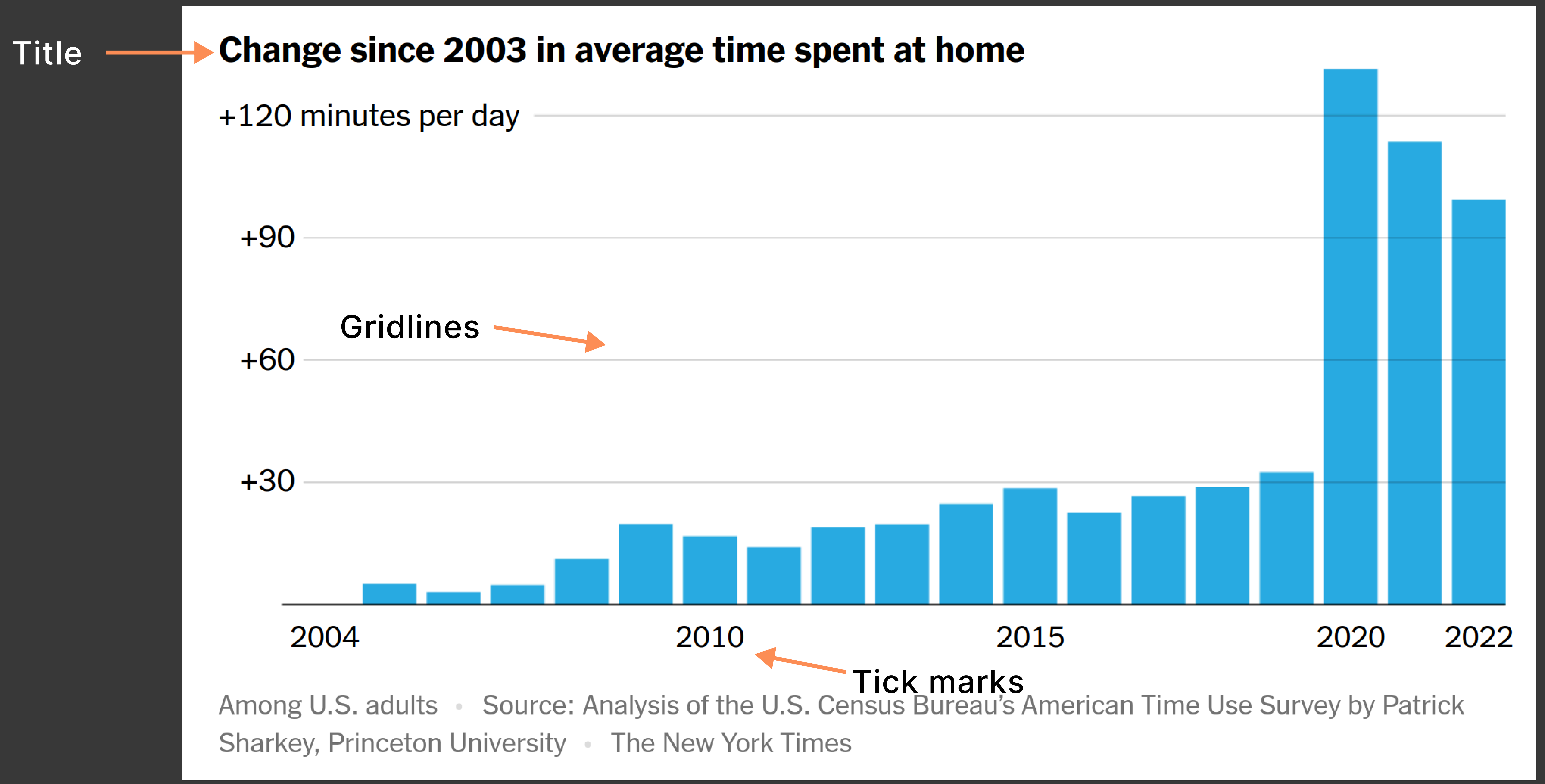
Anatomy of a Chart

Source: NYT



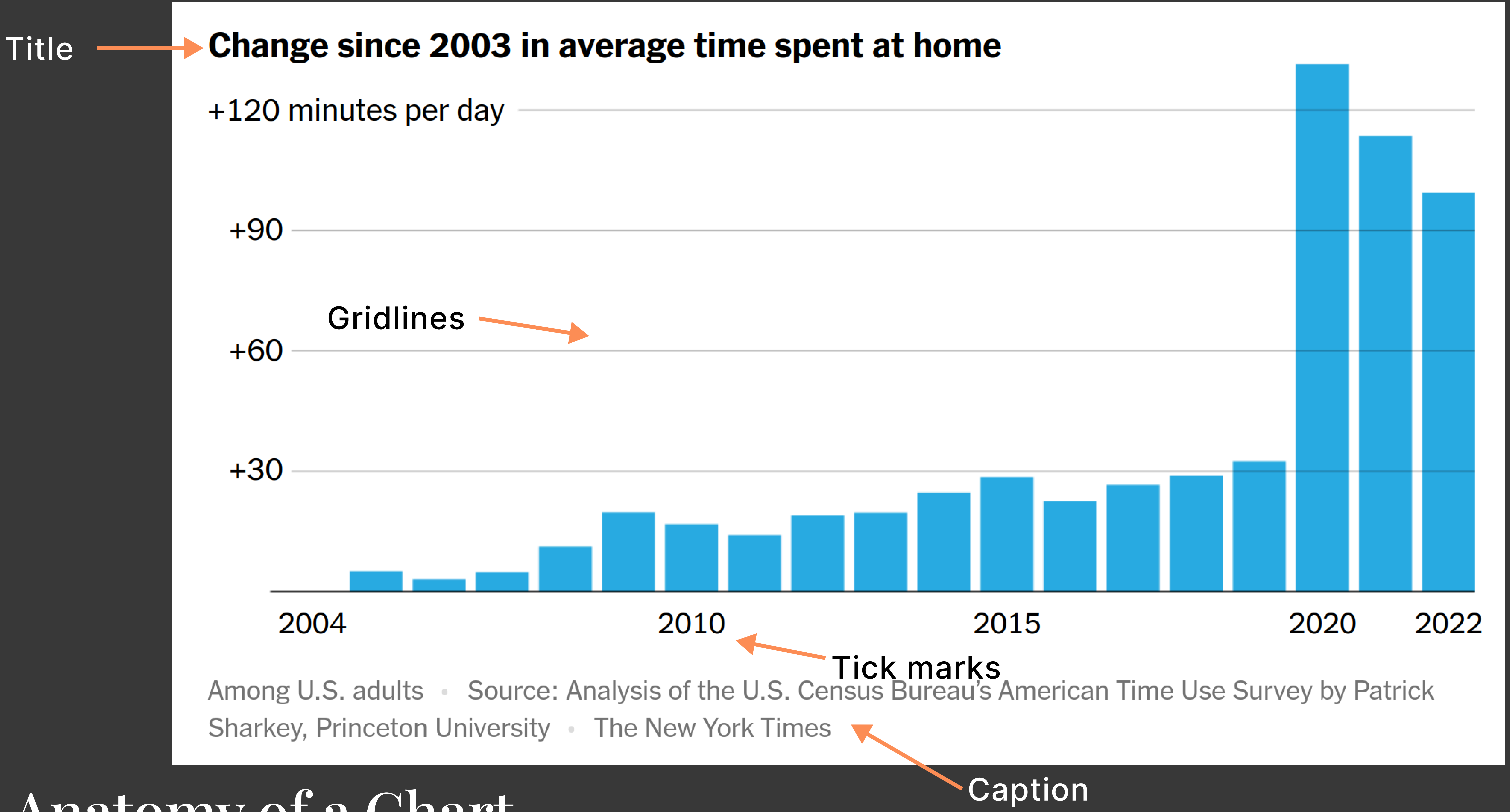
Anatomy of a Chart

Source: NYT



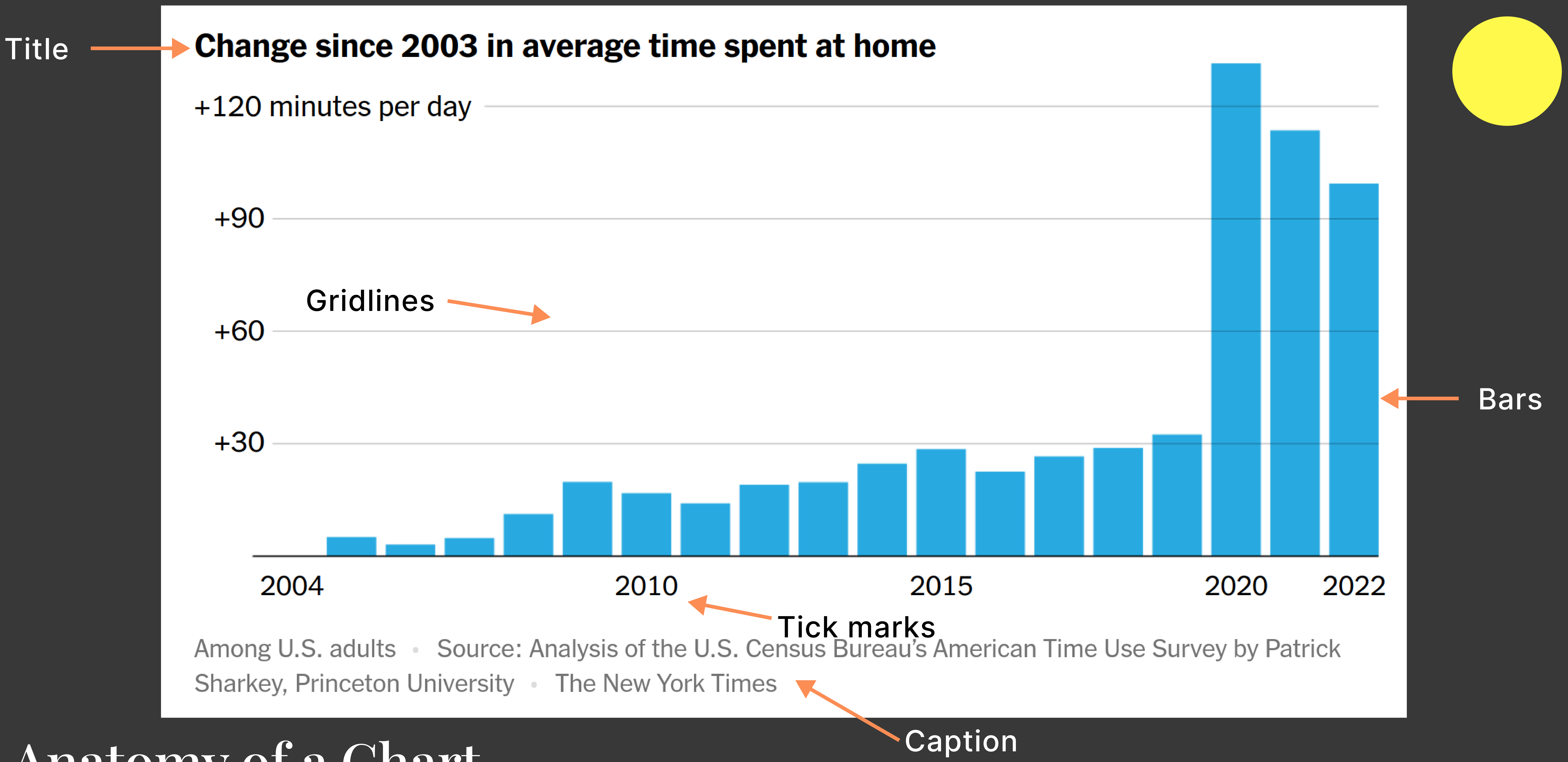
Anatomy of a Chart

Source: NYT



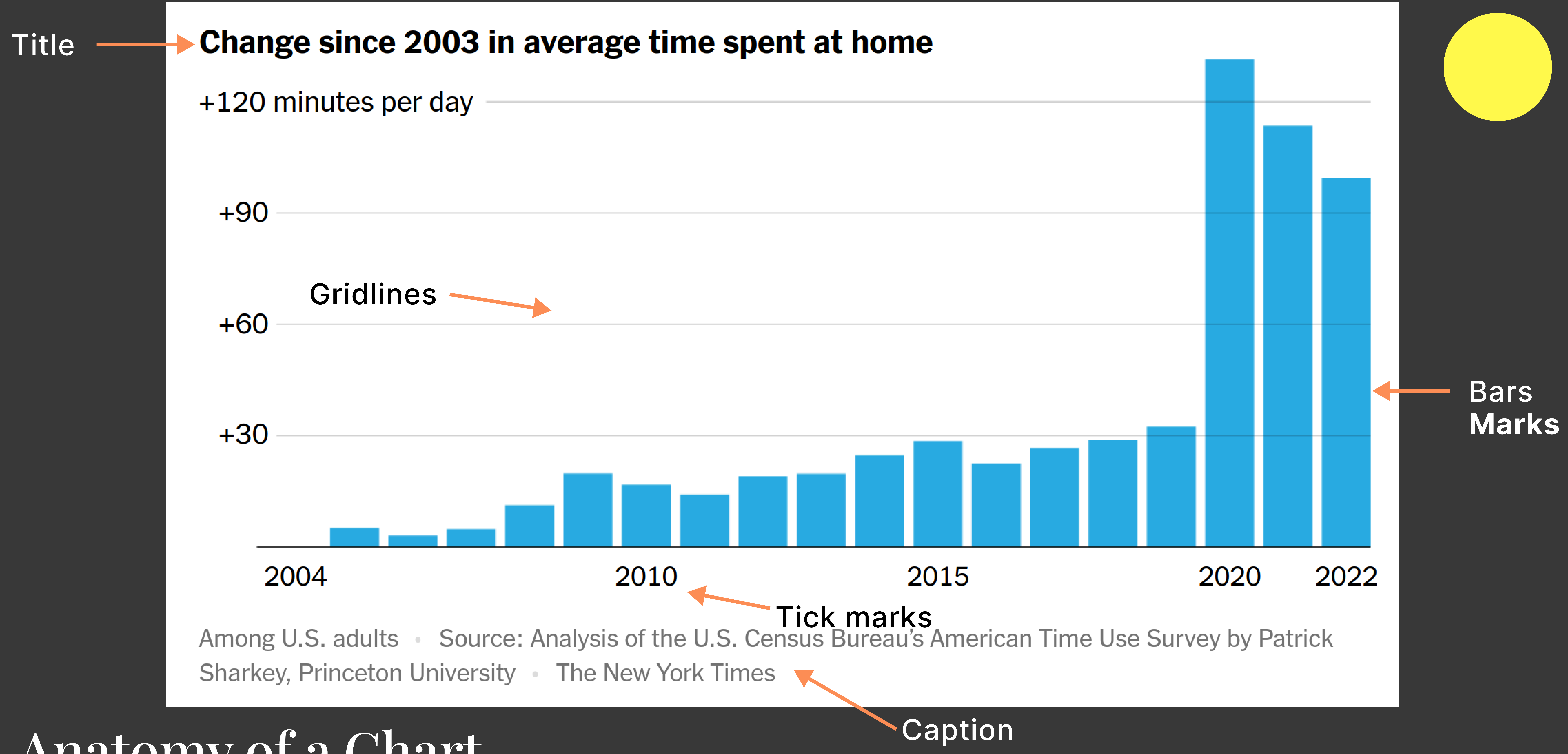
Anatomy of a Chart

Source: NYT



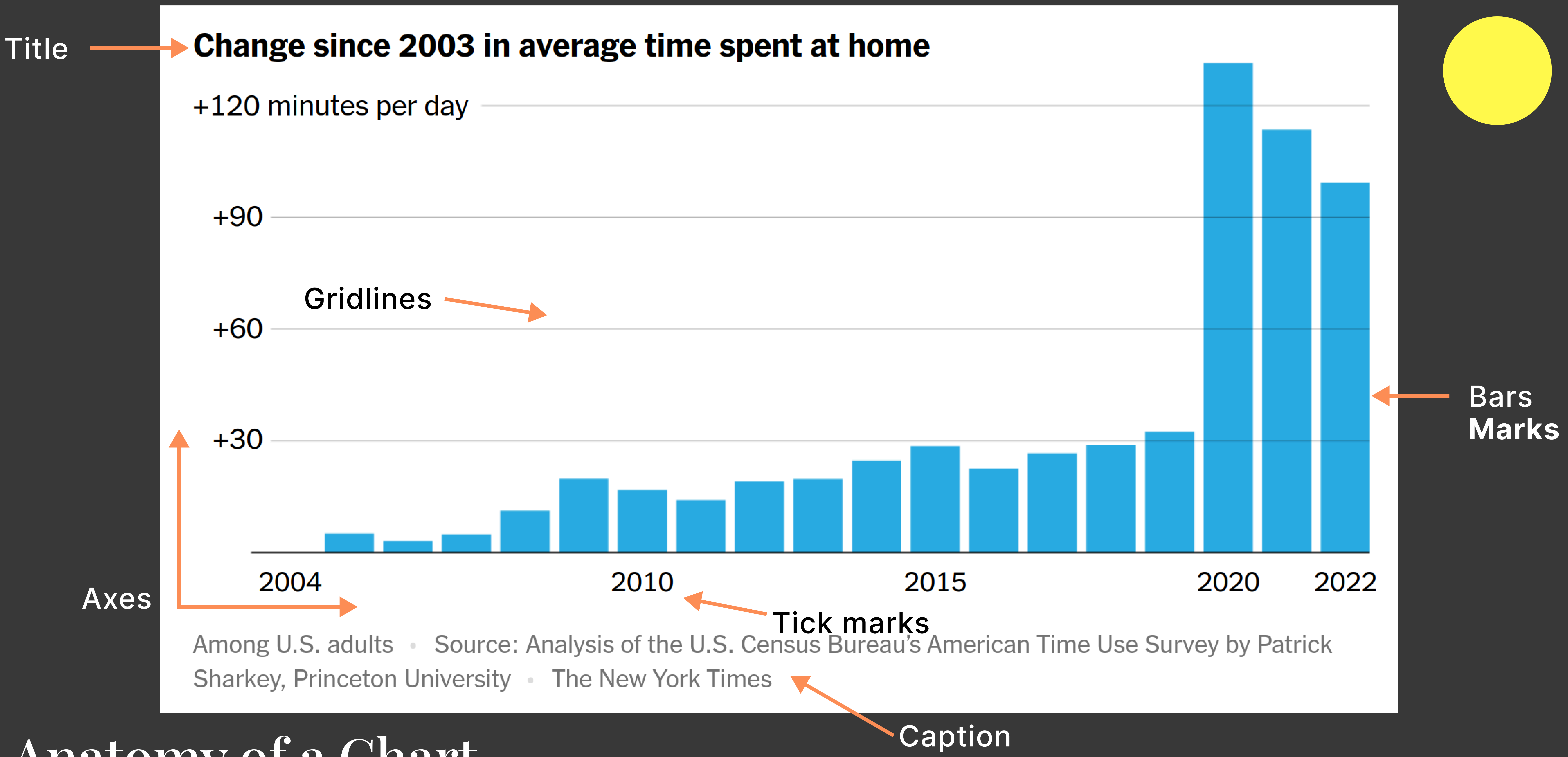
Anatomy of a Chart

Source: NYT



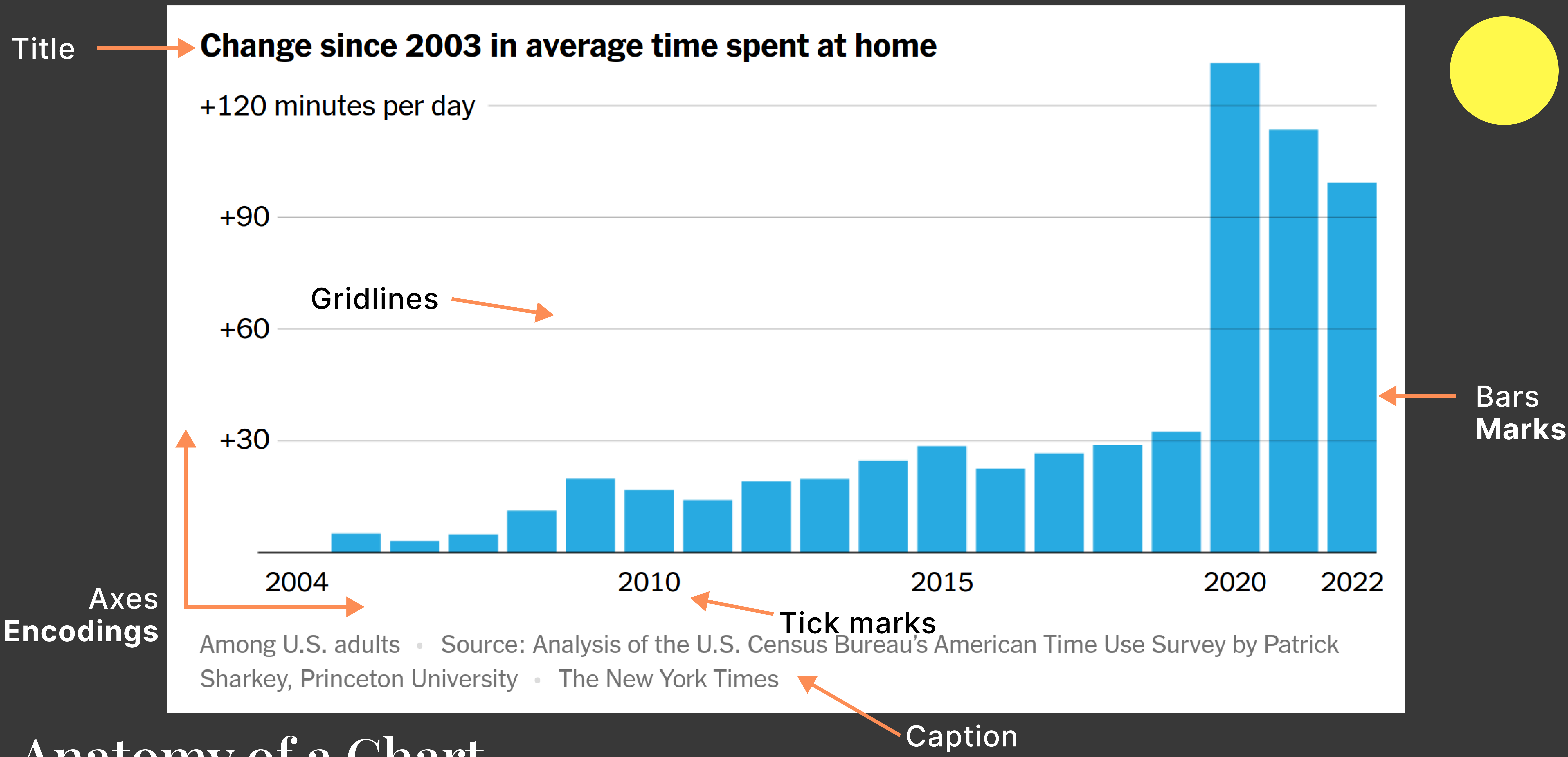
Anatomy of a Chart

Source: NYT



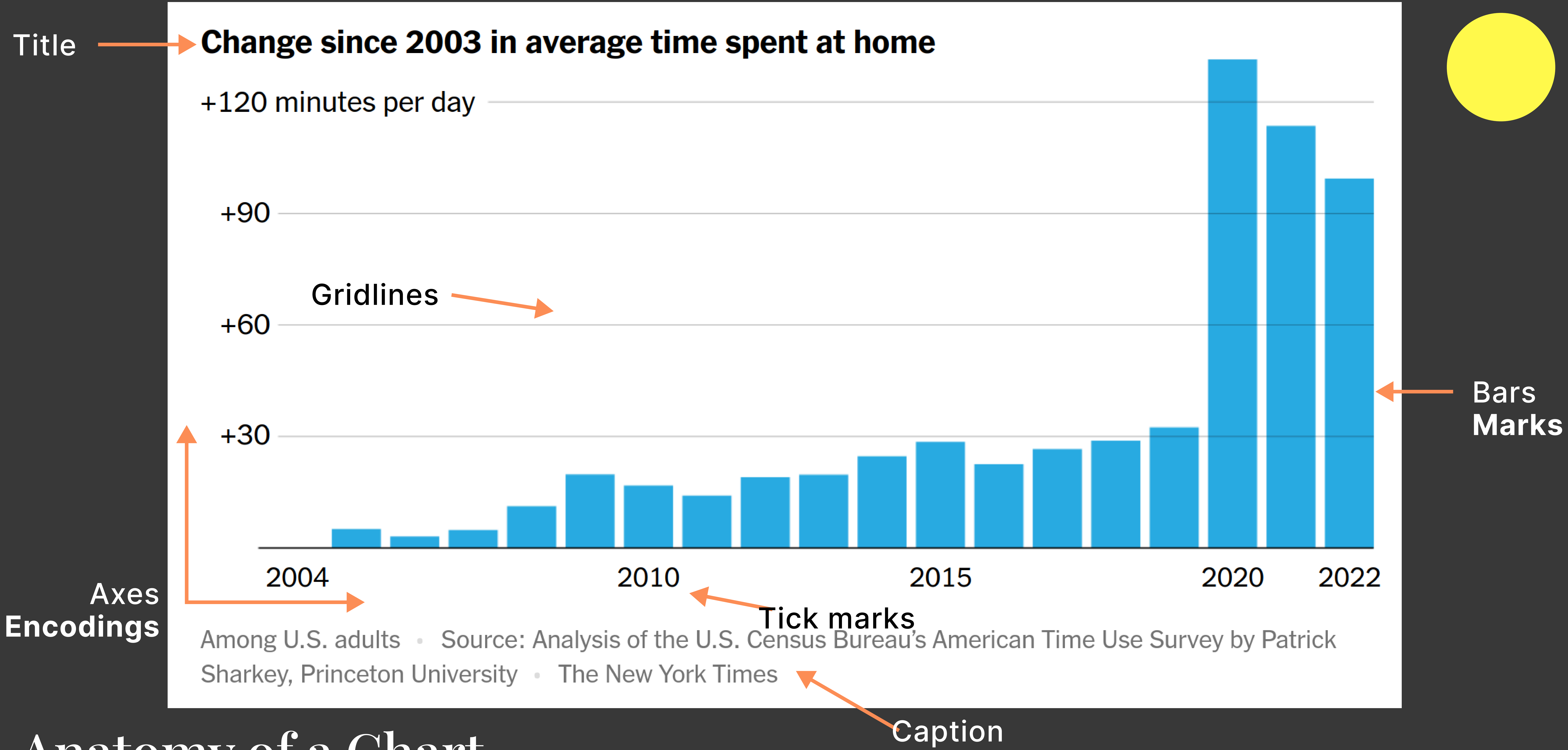
Anatomy of a Chart

Source: NYT



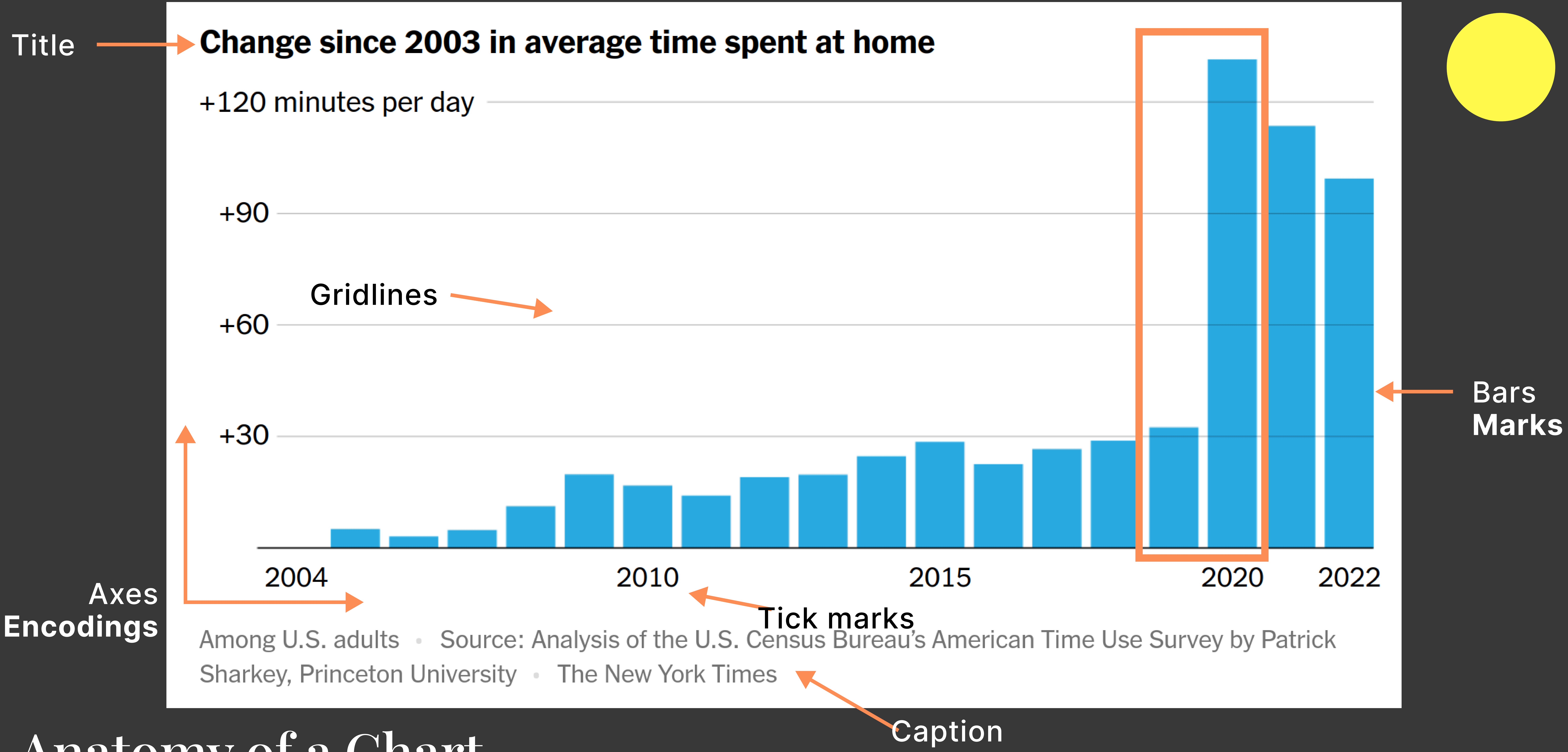
Anatomy of a Chart

Source: NYT



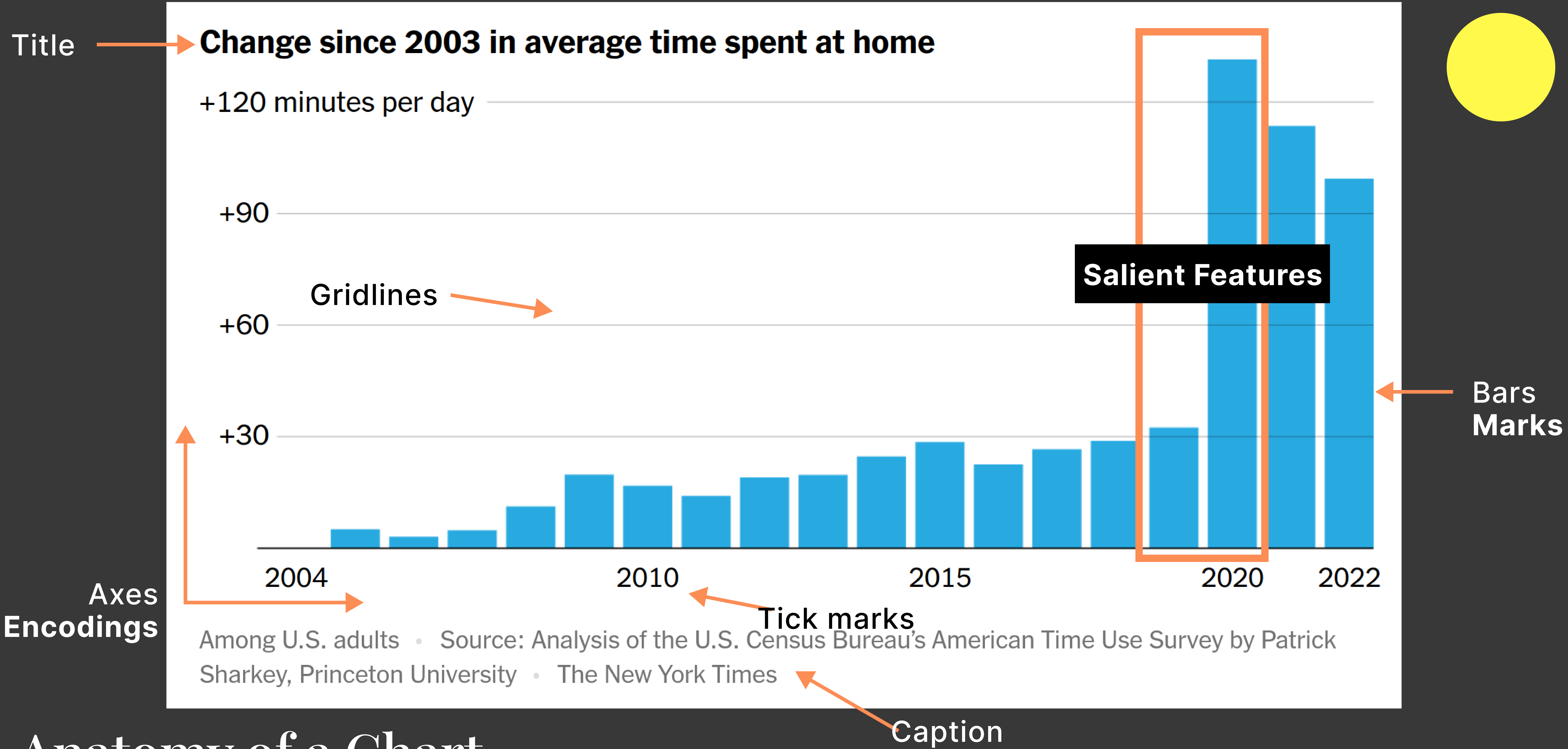
Anatomy of a Chart

Source: NYT



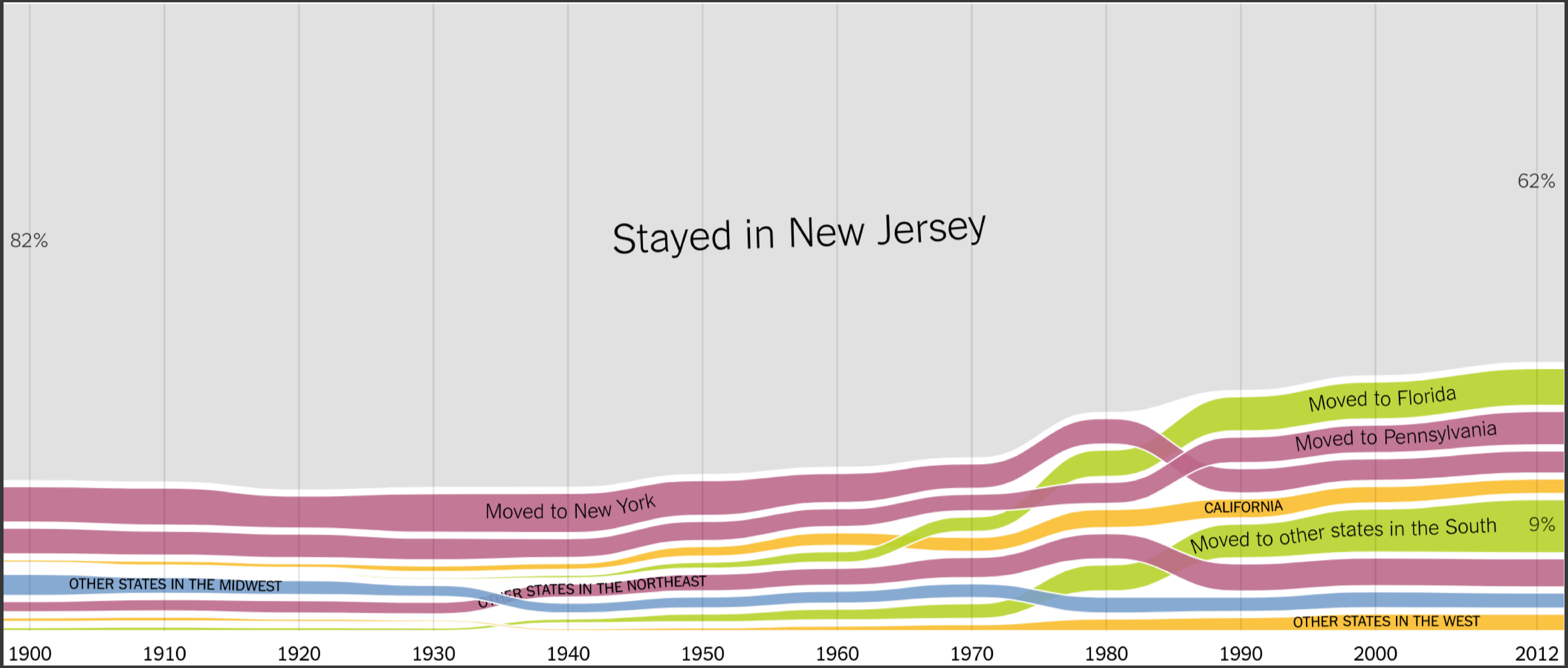
Anatomy of a Chart

Source: NYT



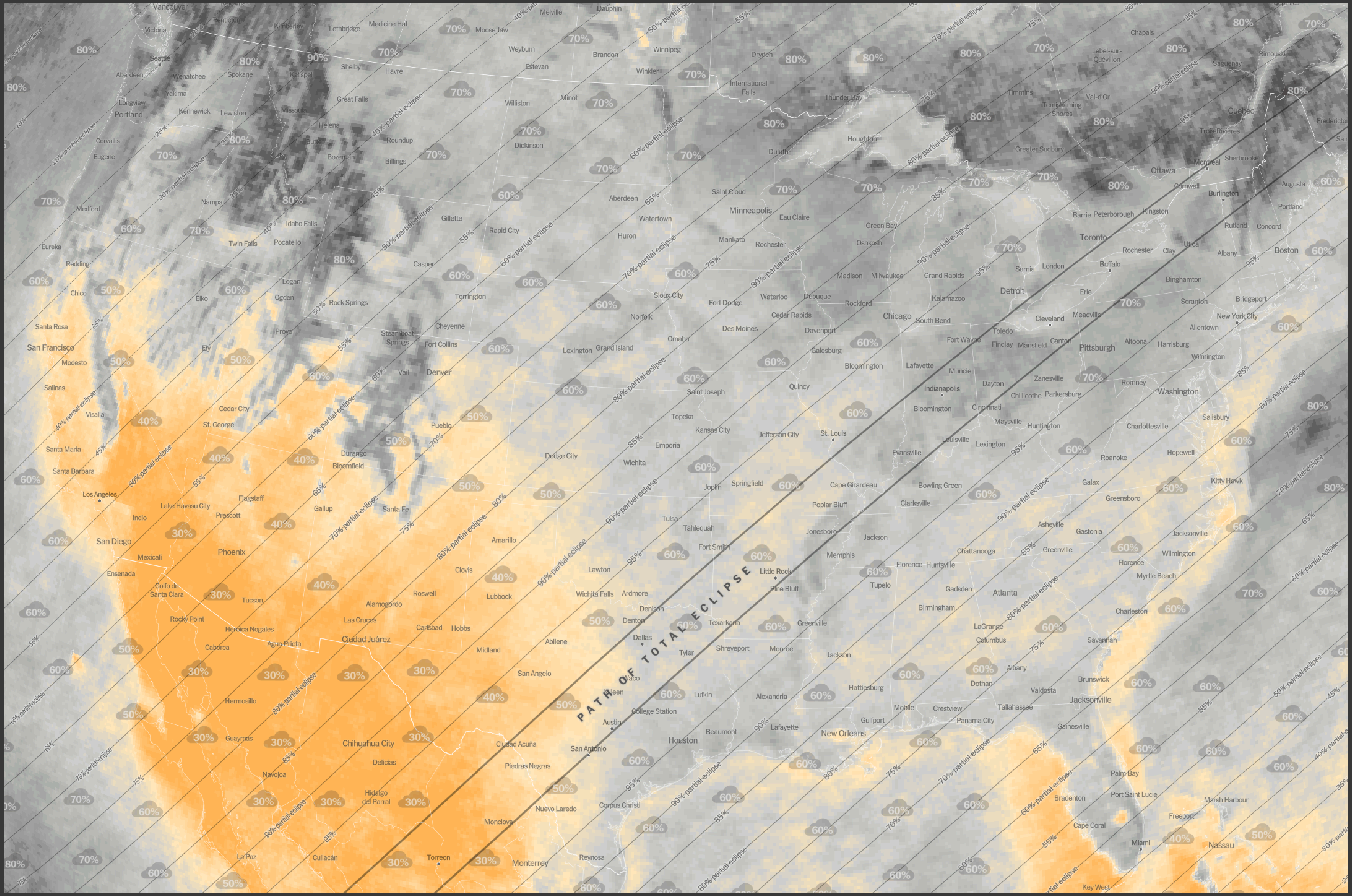
Anatomy of a Chart

Source: NYT



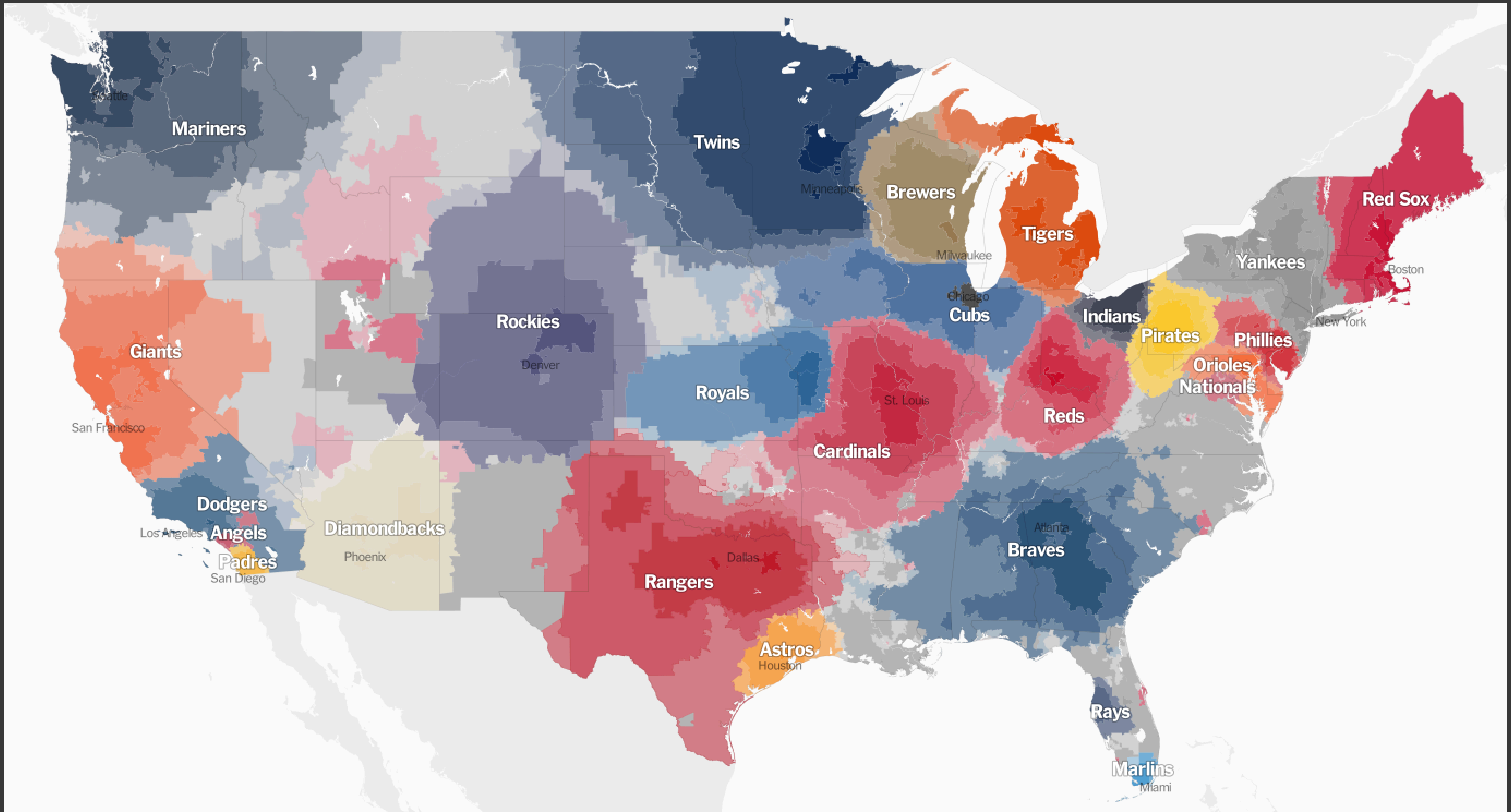
Color: Categorical

Source: NYT



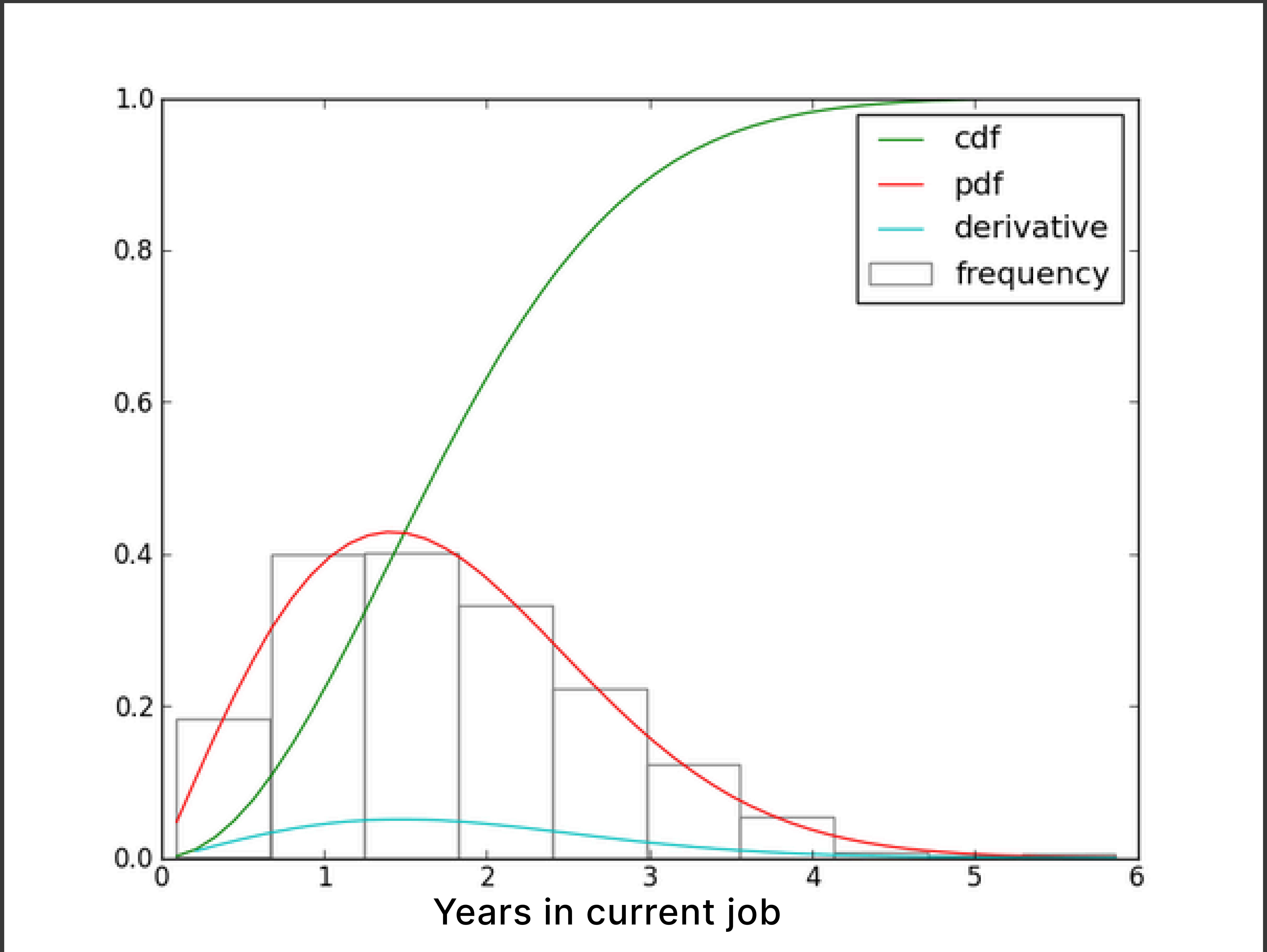
Color: Quantitative

Source: NYT



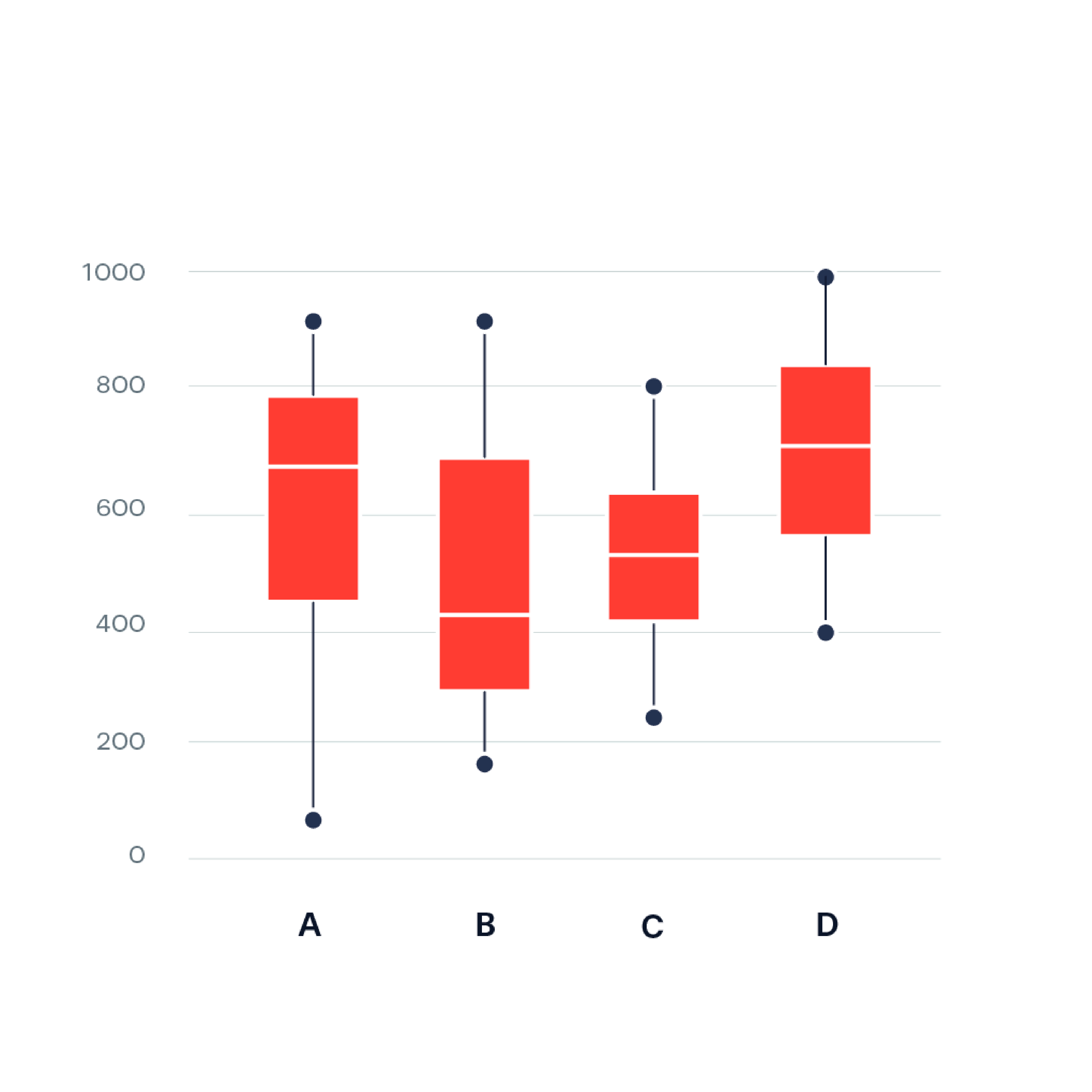
Color: Both?

Source: NYT



Visualizing Distributions

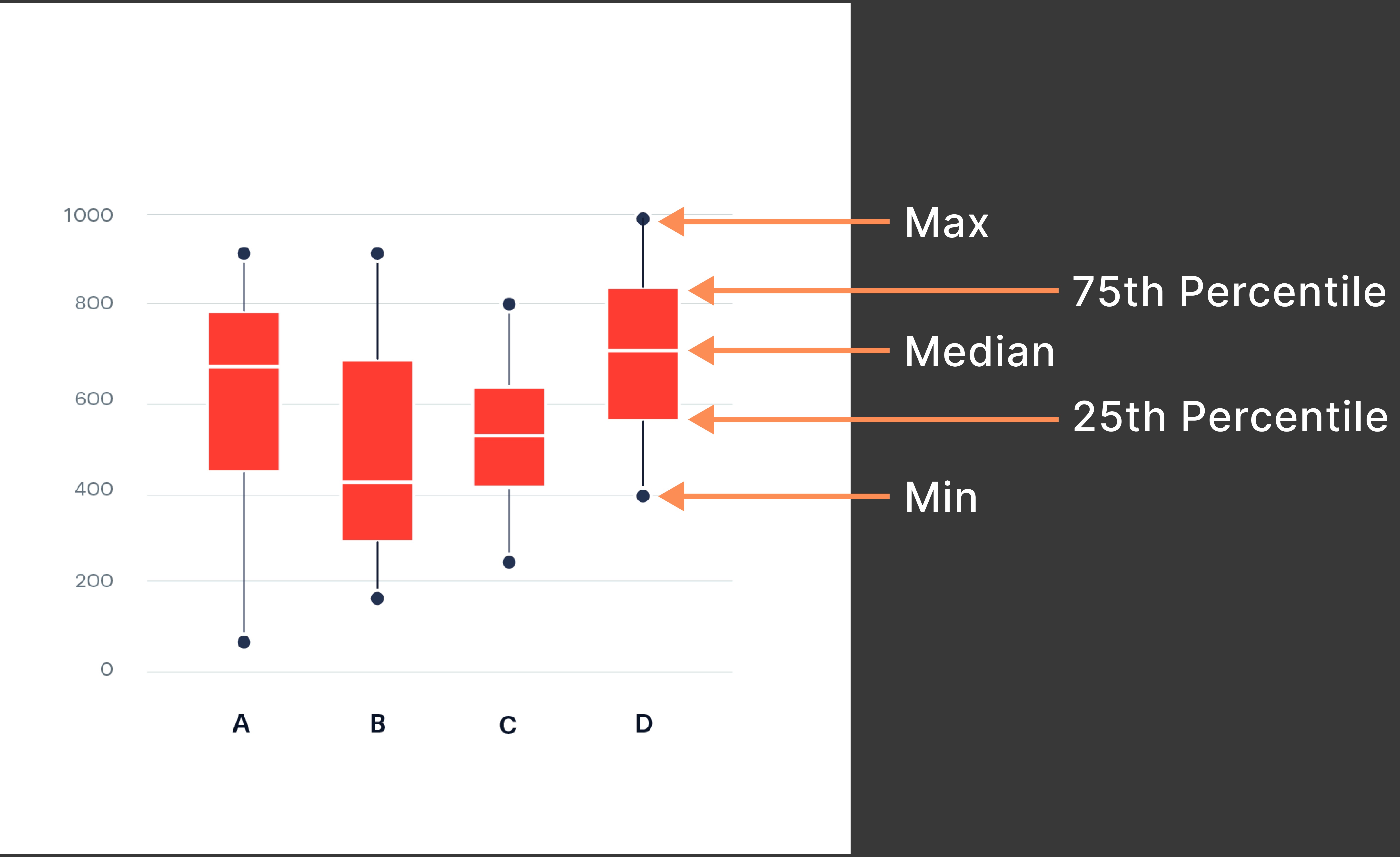
Source



Box Plot

Visualizing Distributions

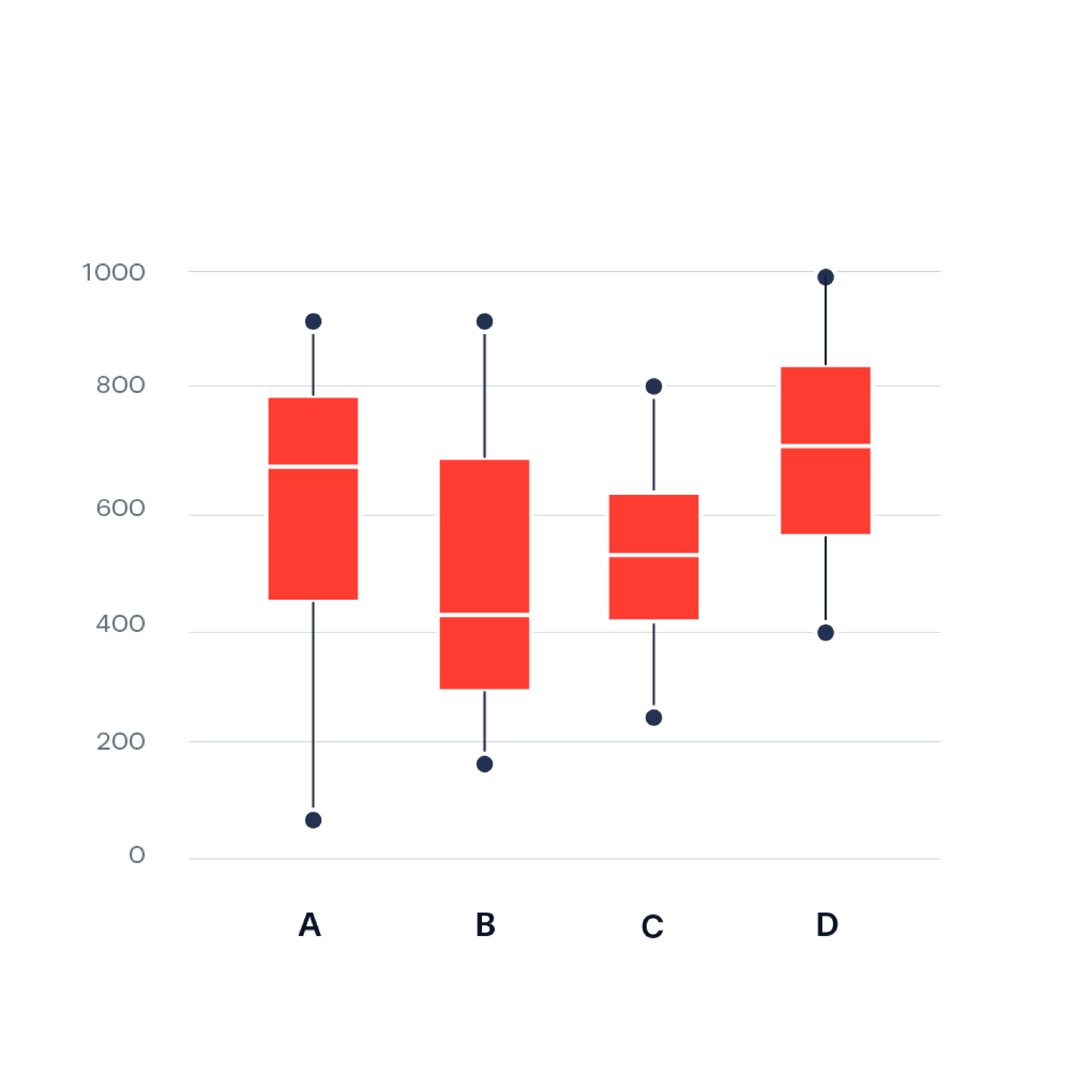
Source: DVP



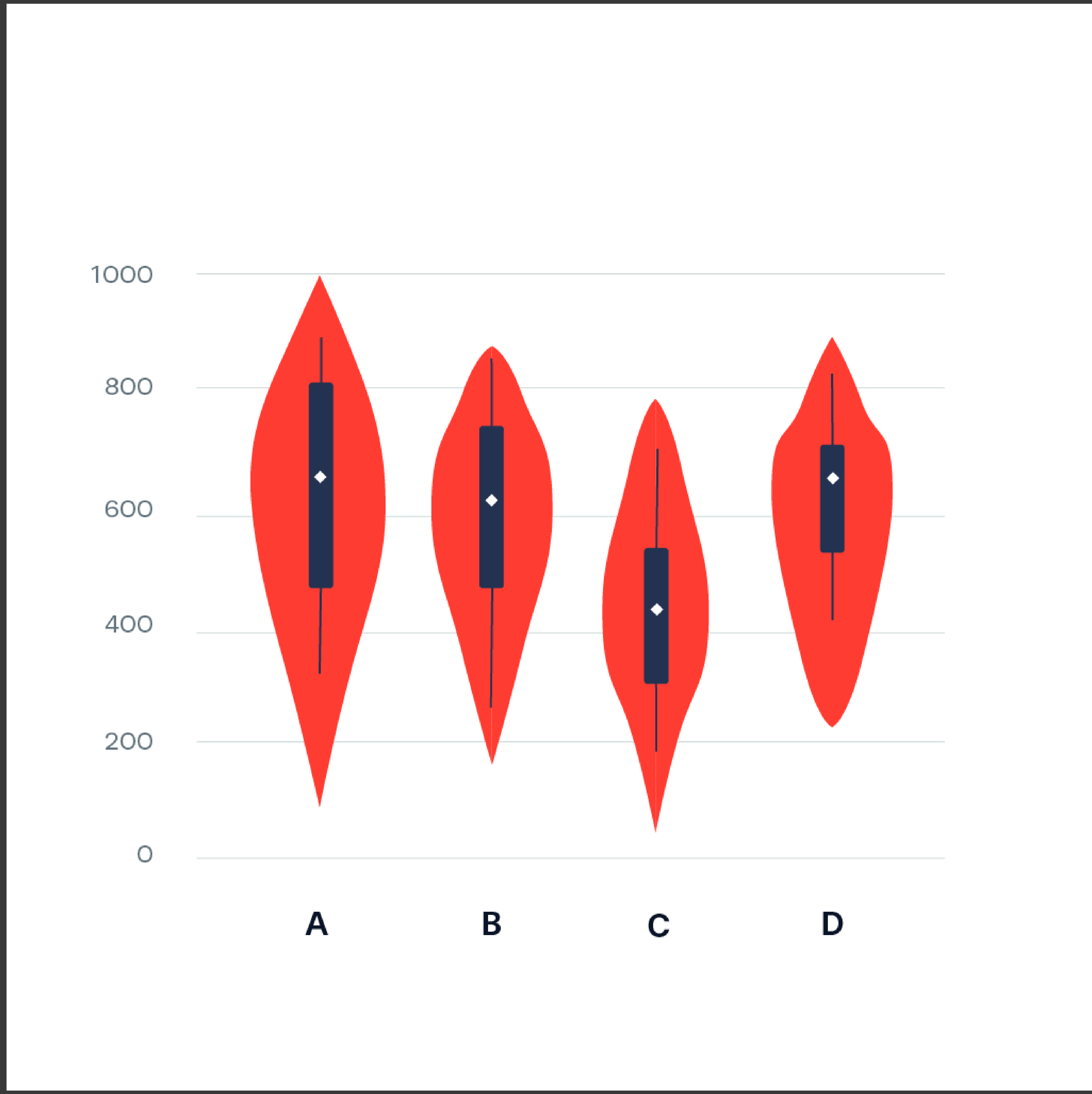
Box Plot

Visualizing Distributions

Source: DVP



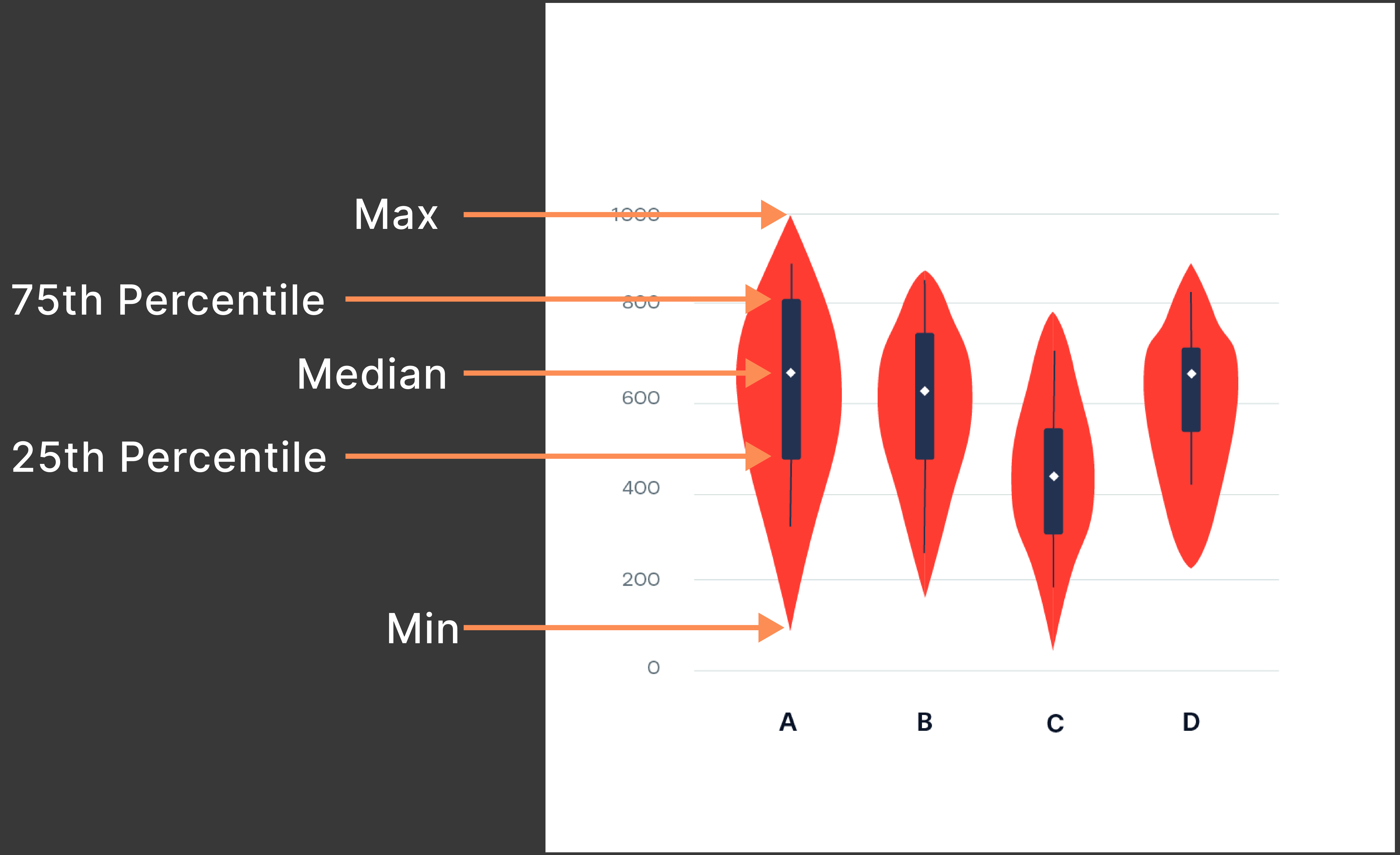
Box Plot



Violin Plot

Visualizing Distributions

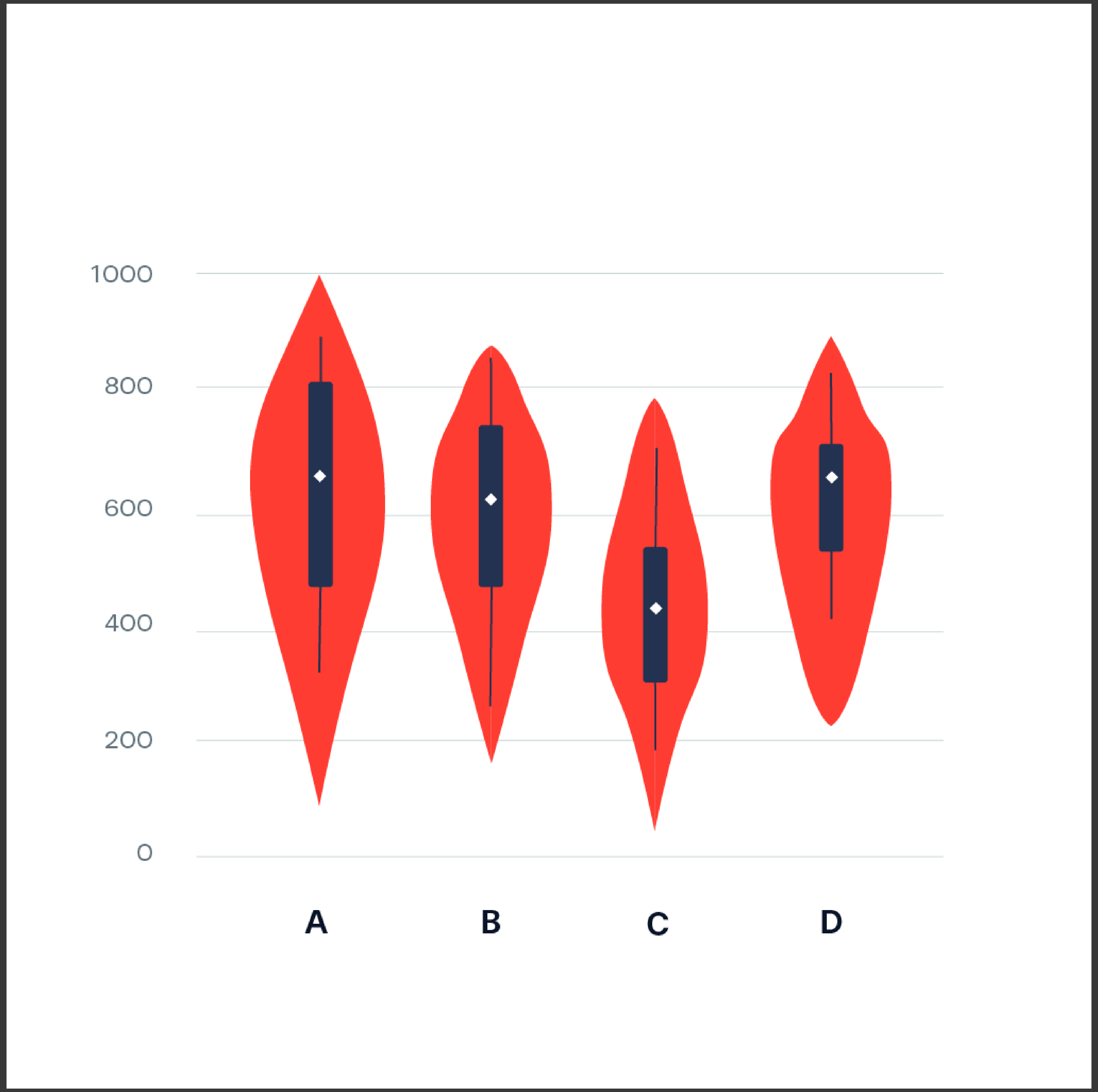
Source: DVP



Violin Plot

Visualizing Distributions

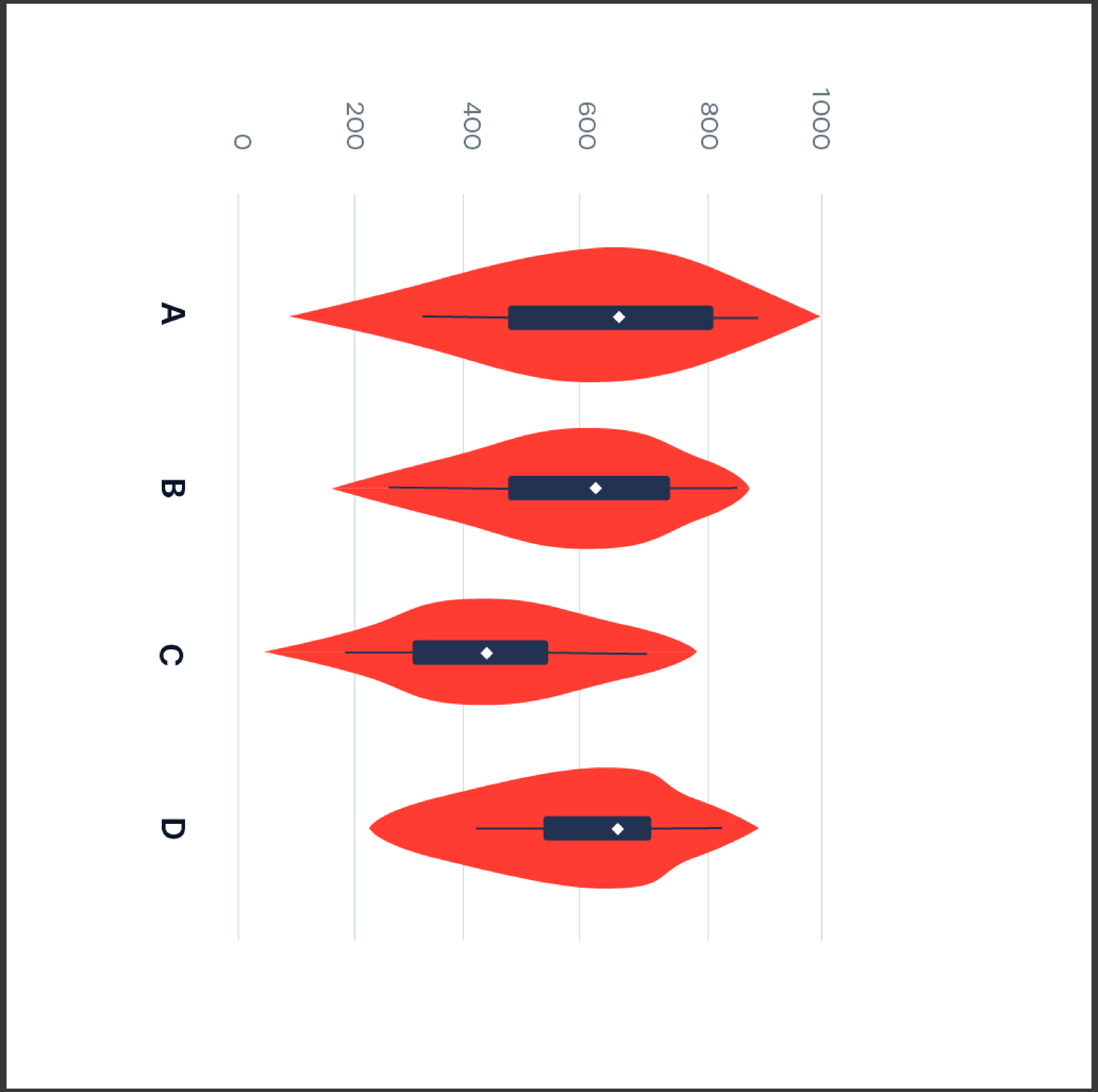
Source: DVP



Violin Plot

Visualizing Distributions

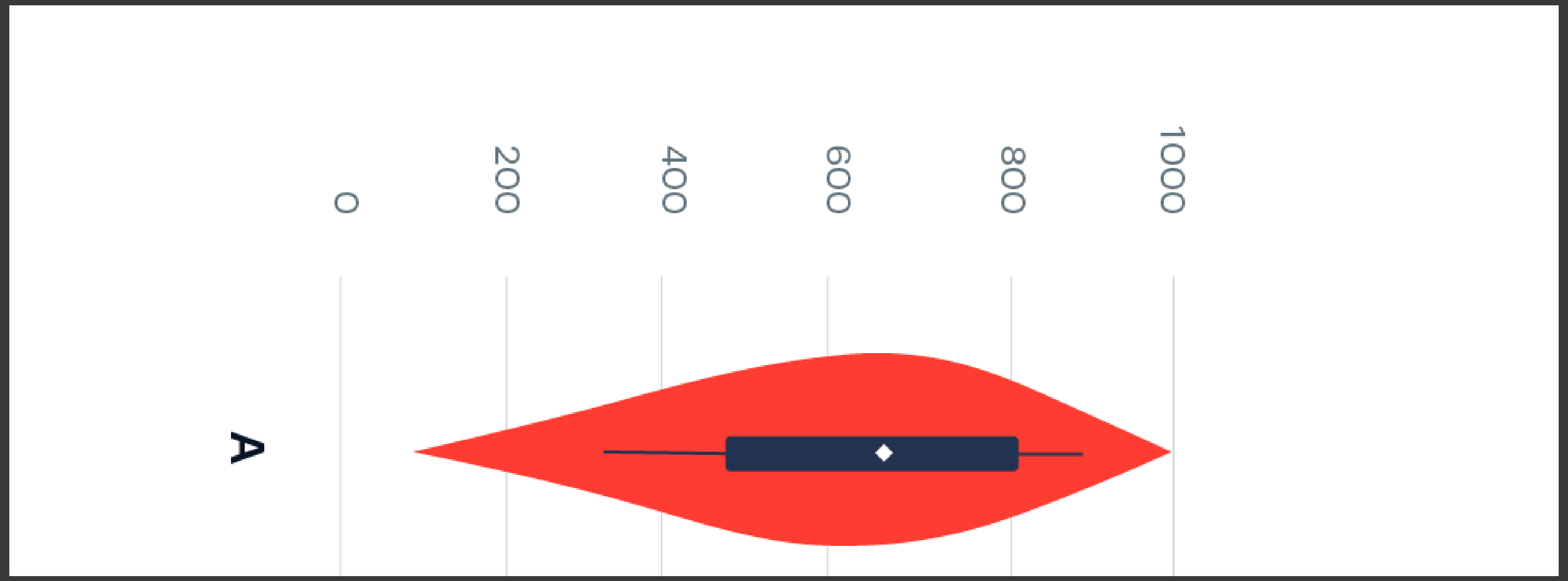
Source: DVP



Violin Plot

Visualizing Distributions

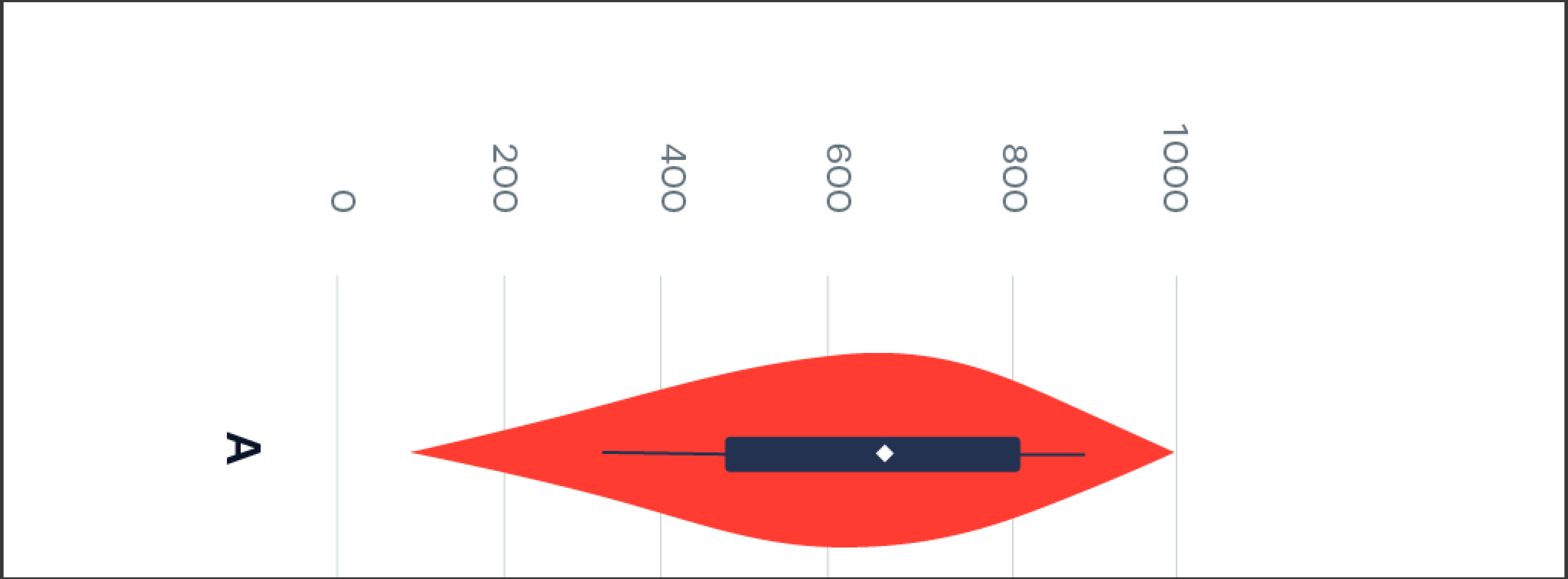
Source: DVP



Violin Plot

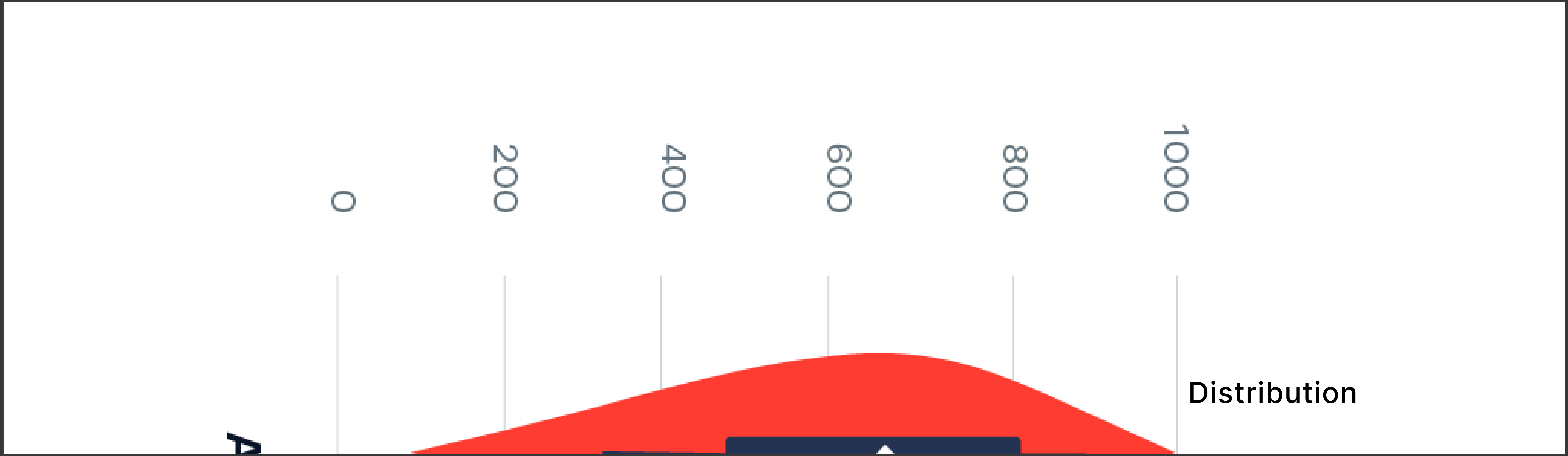
Visualizing Distributions

Source: DVP



Visualizing Distributions

Source: DVP



Visualizing Distributions

Source: DVP

Visualizing Distributions

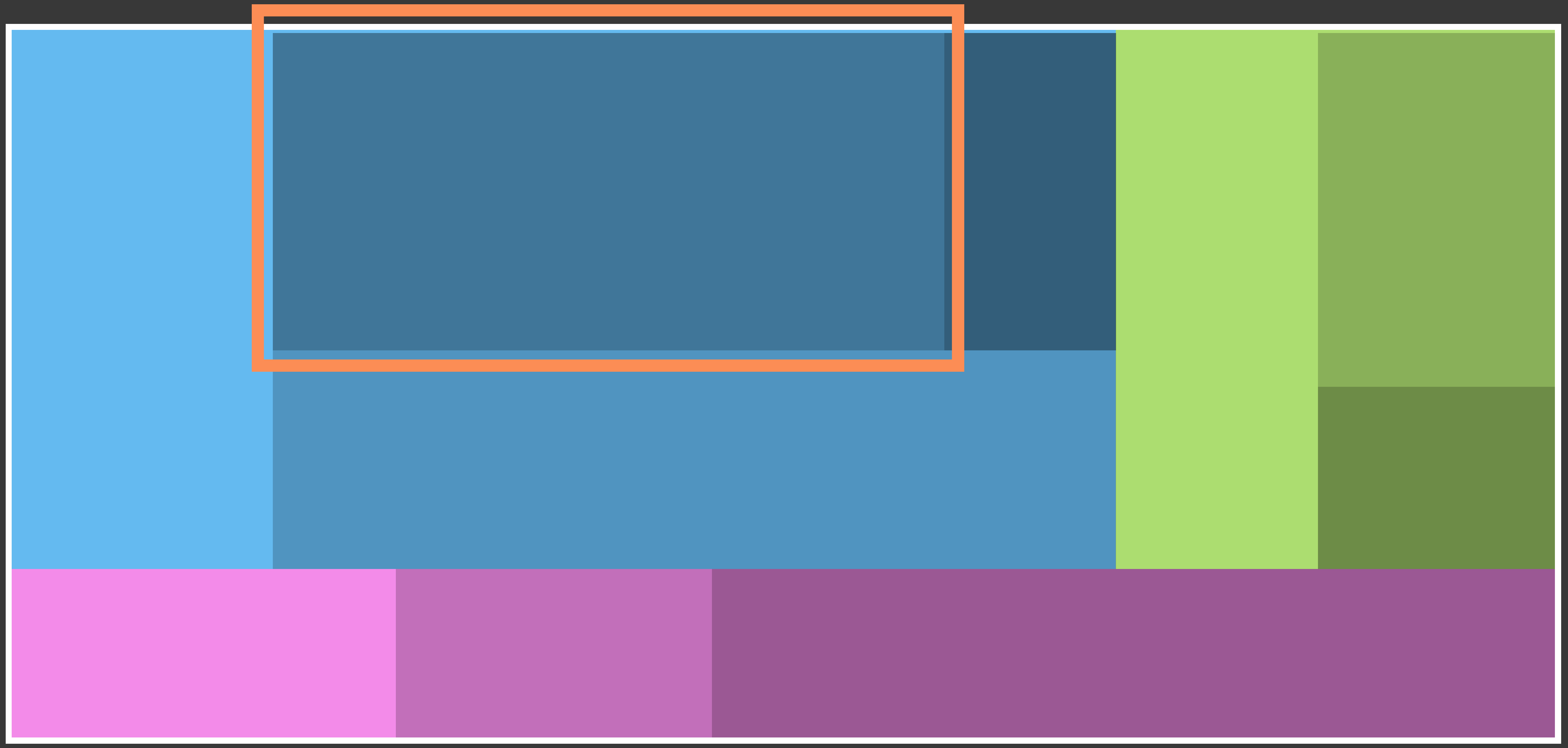
Matejka & Fitzmaurice, CHI 2017

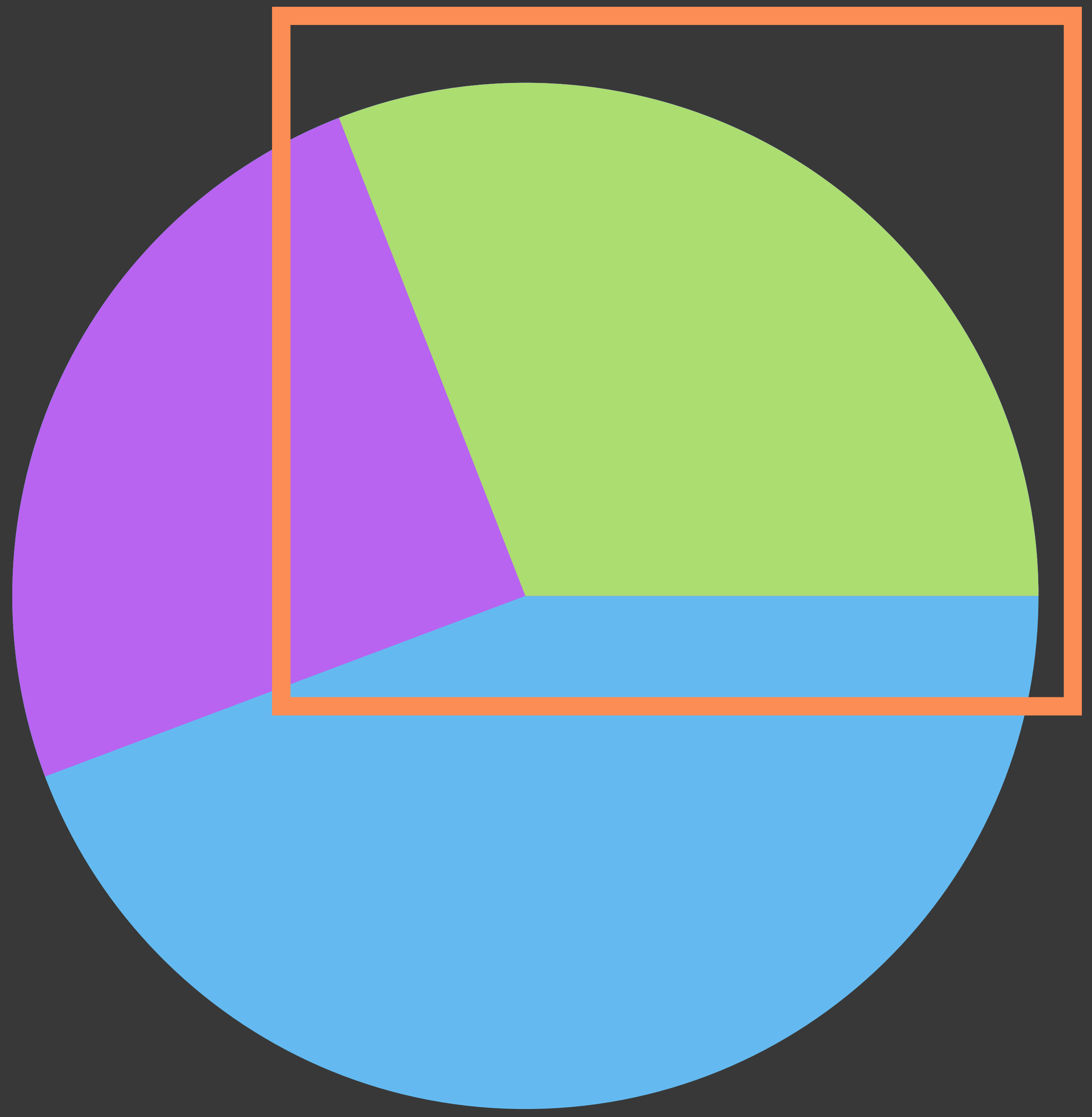
Perception & Comparison

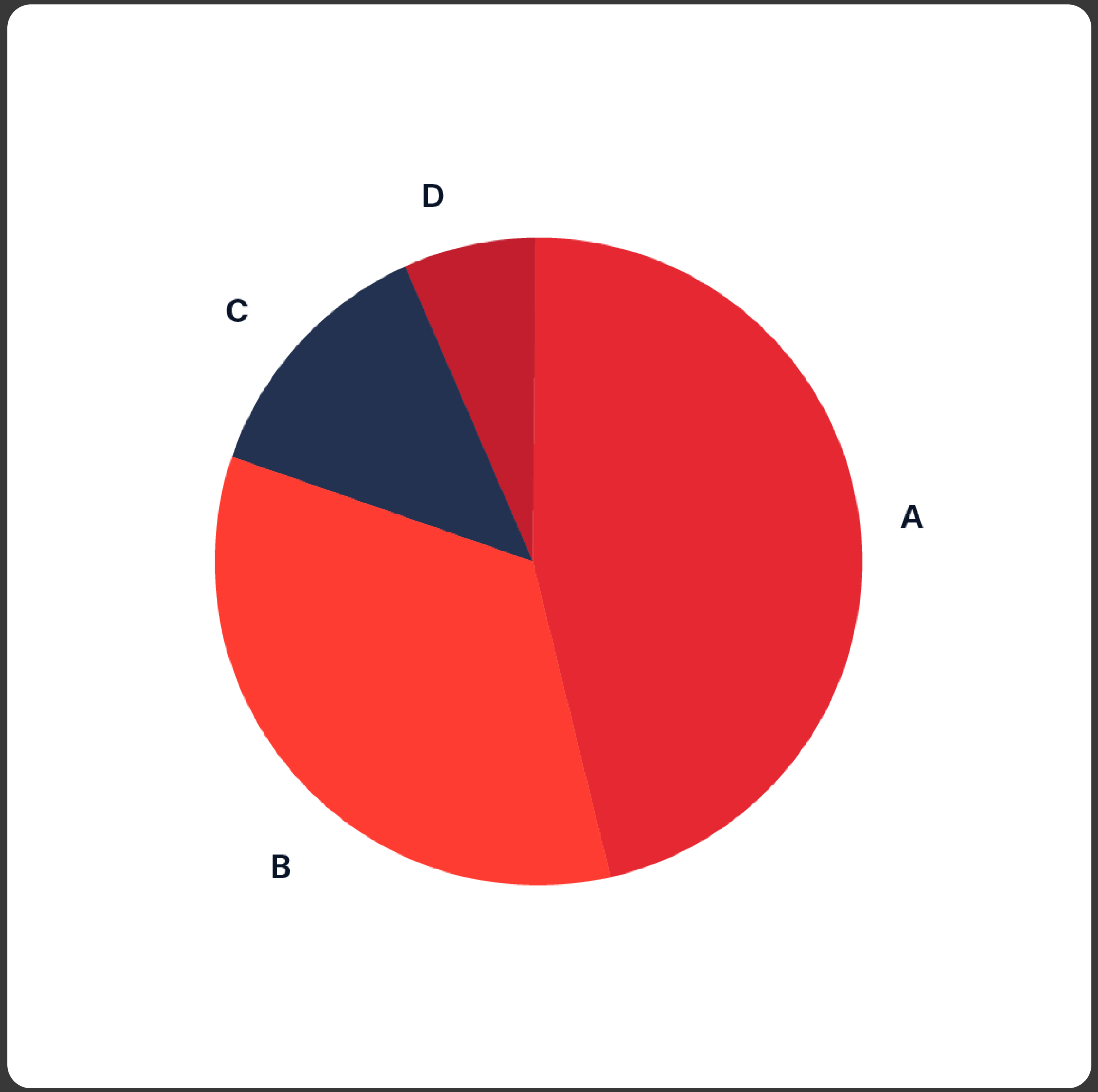
Encodings

How data maps to perceptual “channels”

Area: Quantitative



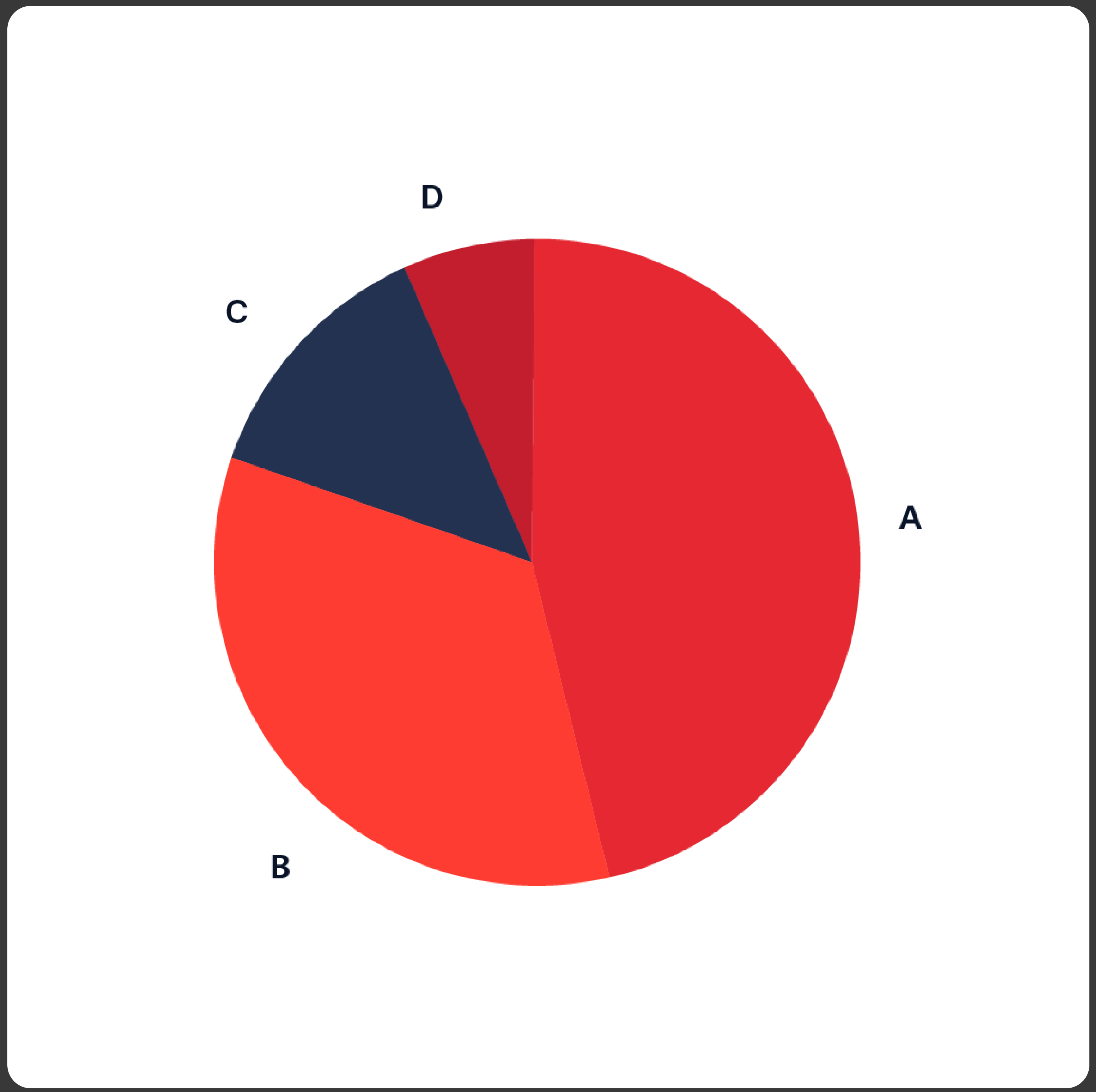


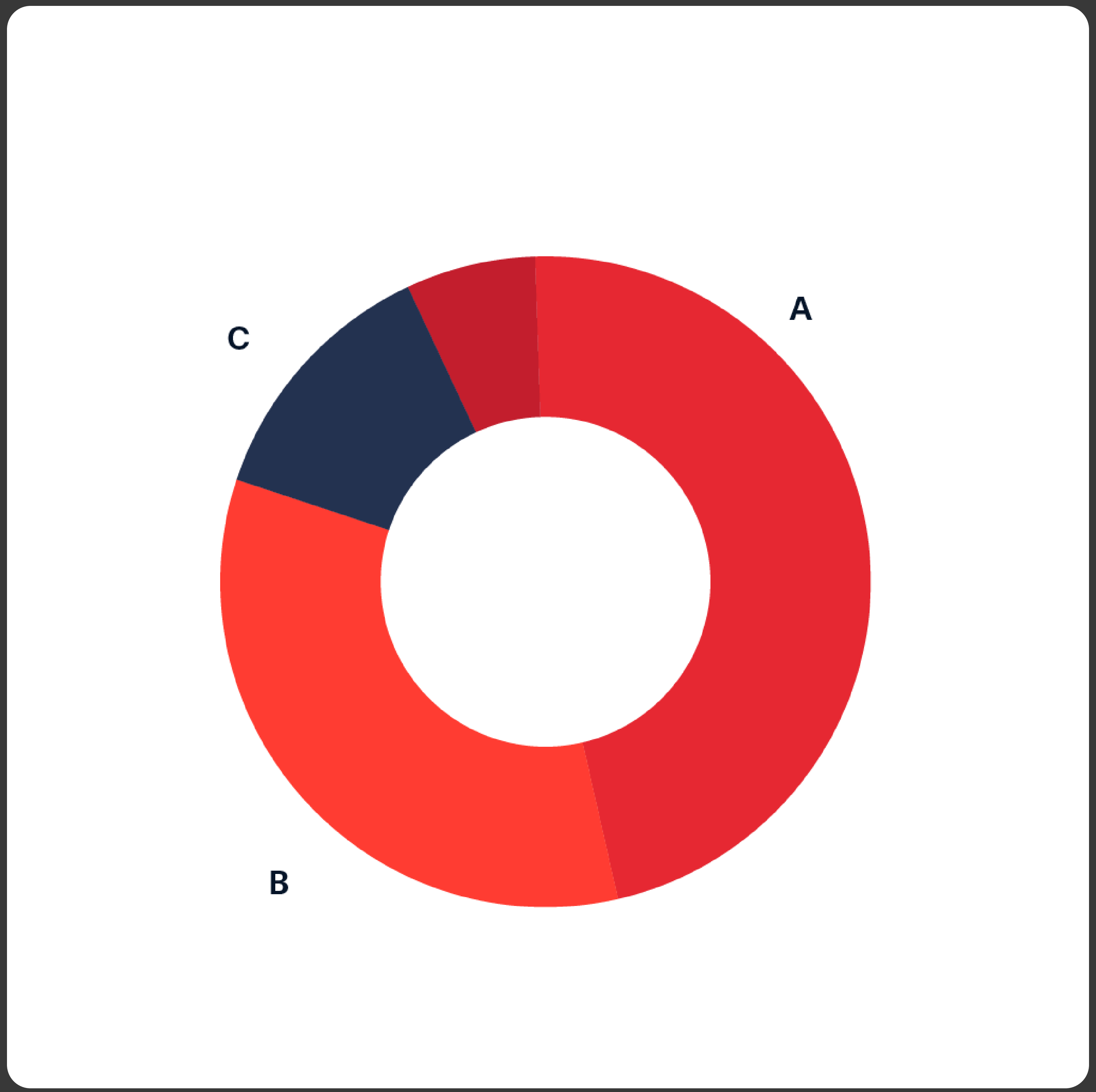


No Pie Charts!

The eye tends to perceive *area* instead of *angle*, making large segments look much larger than small segments.

Source: DVP





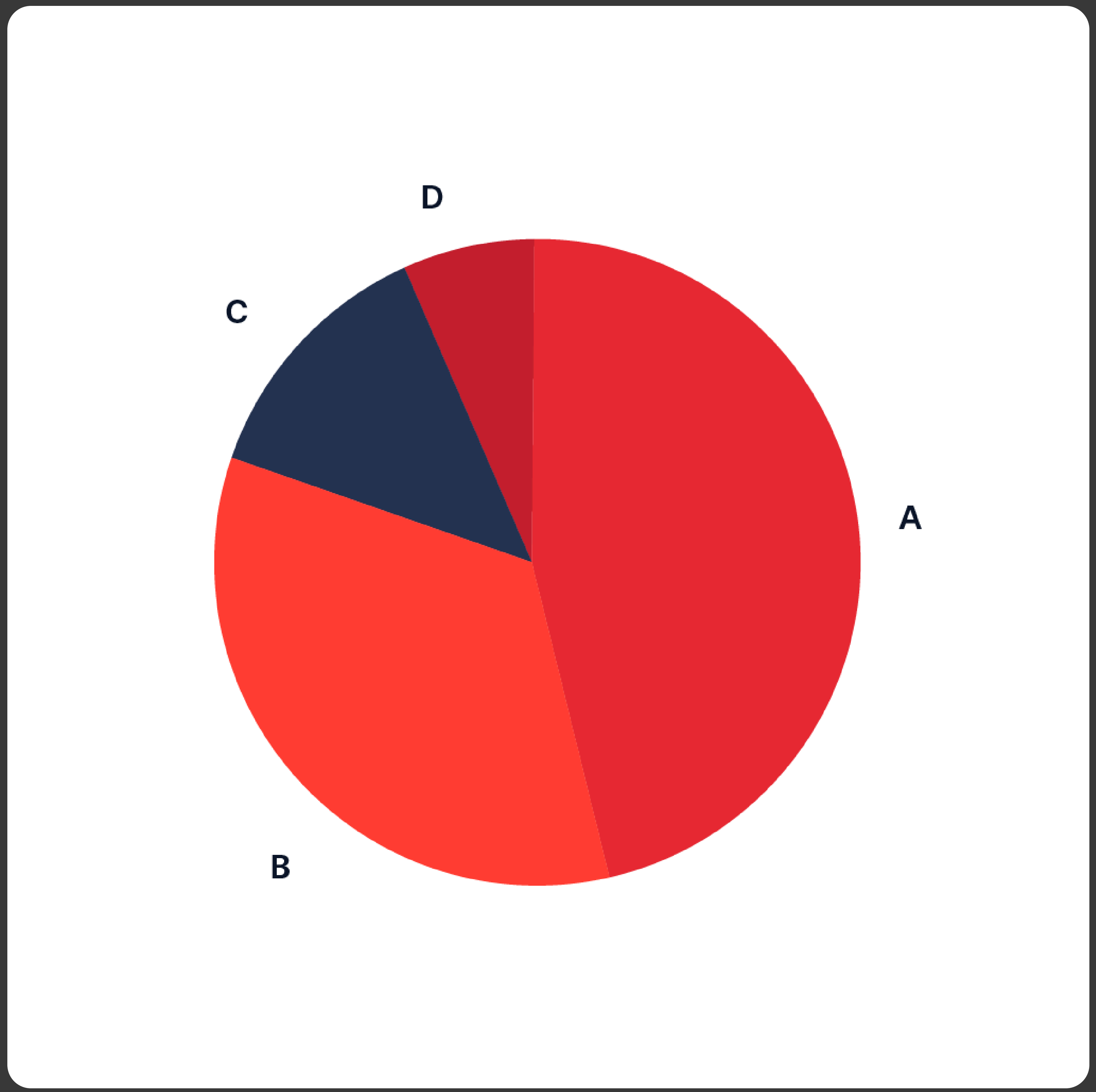
No Pie Charts!

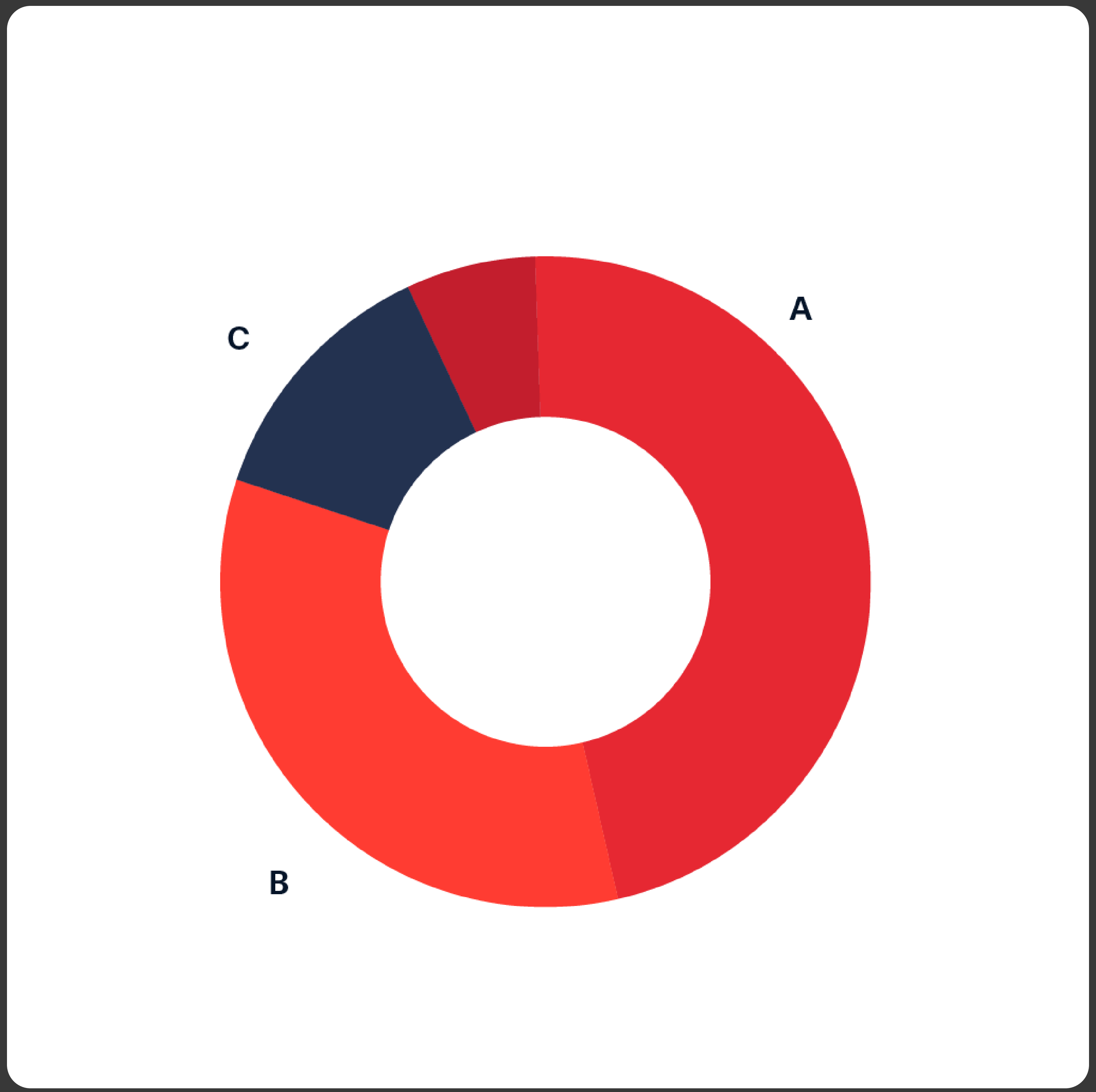
The eye tends to perceive *area* instead of *angle*, making large segments look much larger than small segments.

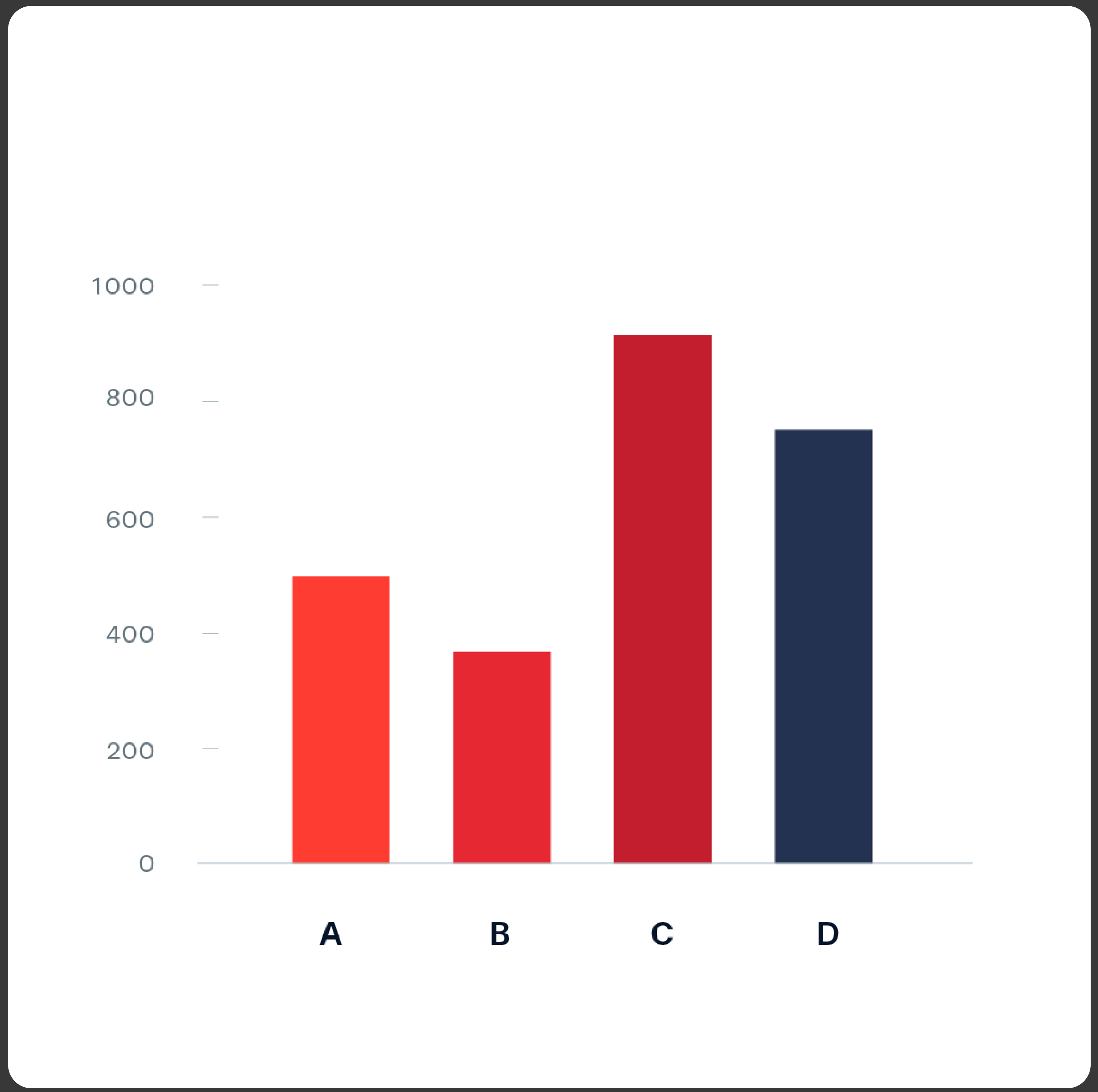
Donut Chart

Encourages viewer to perceive *angle* instead of *area*.

Source: DVP







No Pie Charts!

The eye tends to perceive *area* instead of *angle*, making large segments look much larger than small segments.

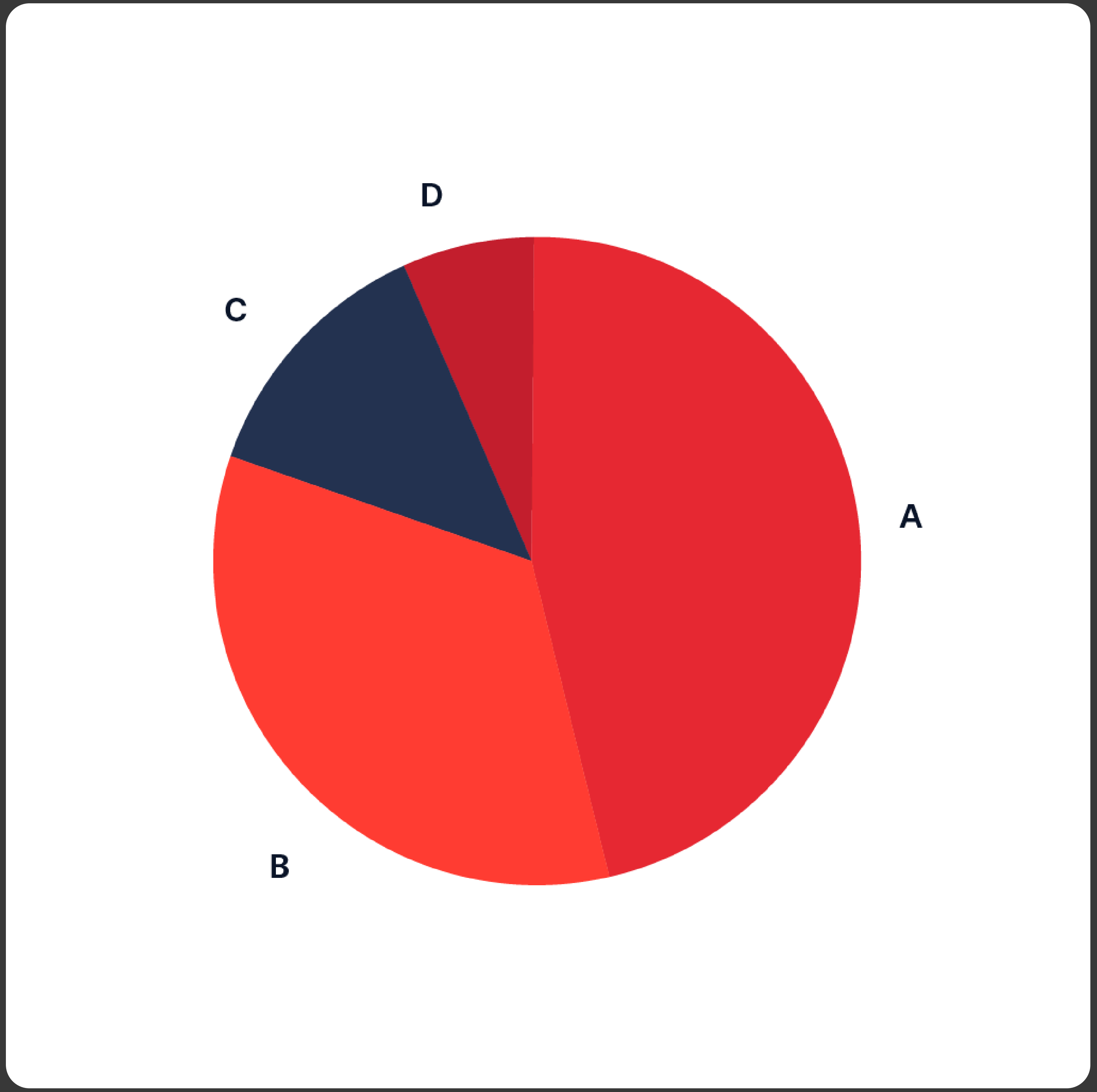
Donut Chart

Encourages viewer to perceive *angle* instead of *area*.

Bar Chart

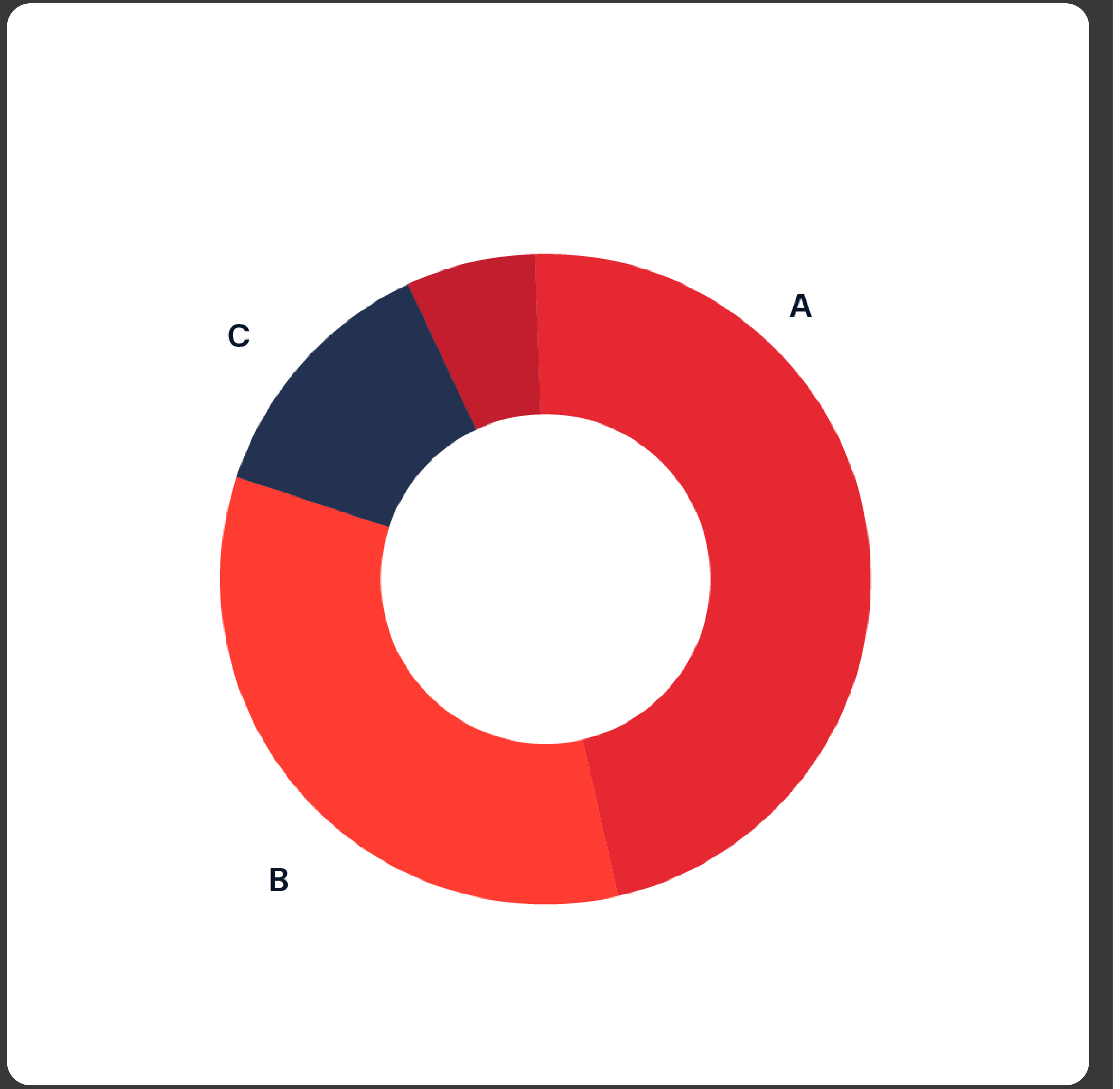
Like a donut chart, but not bendy! Your eye is good at comparing straight lengths!

Source: DVP



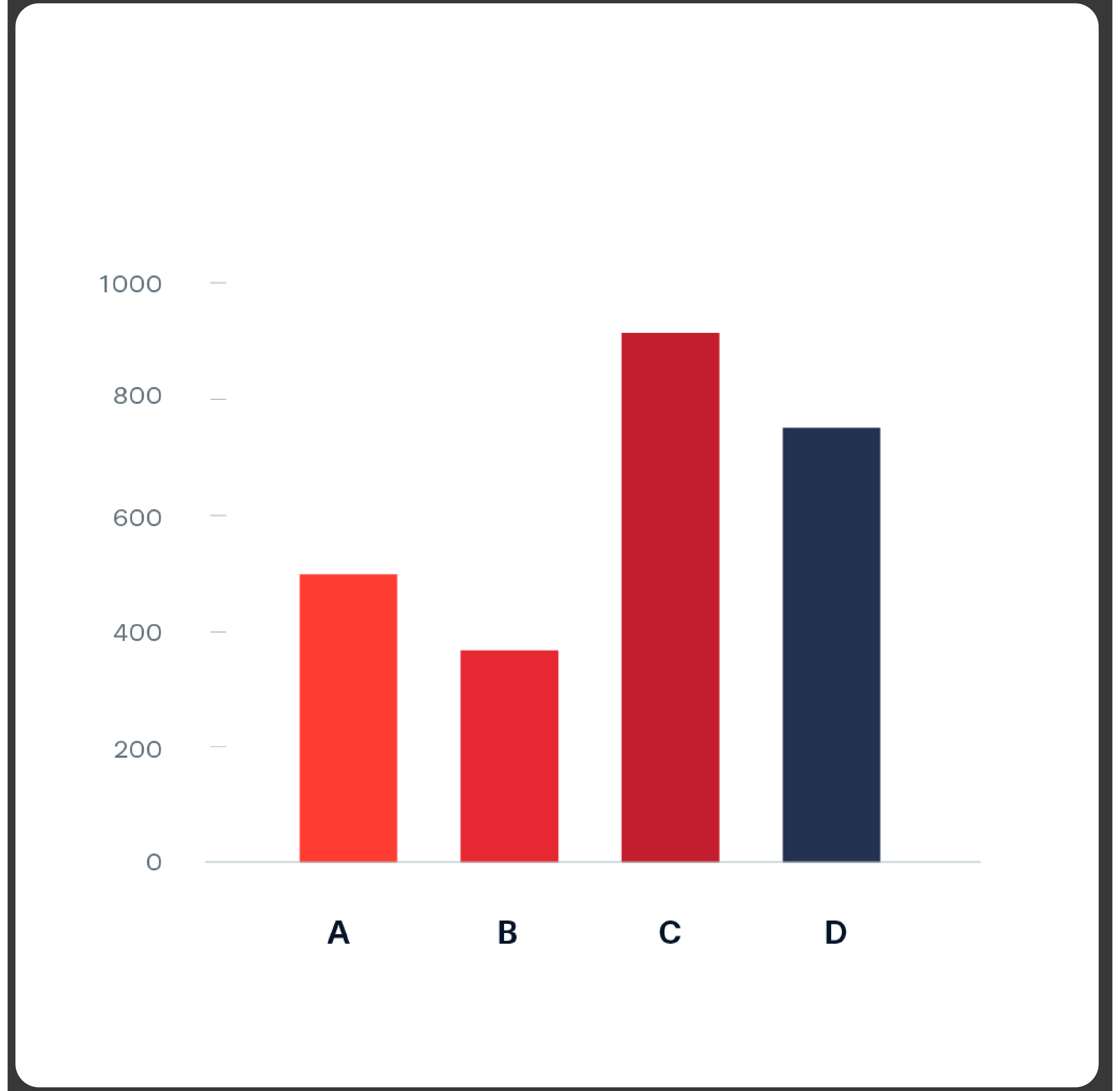
No Pie Charts!

The eye tends to perceive *area* instead of *angle*, making large segments look much larger than small segments.



Donut Chart

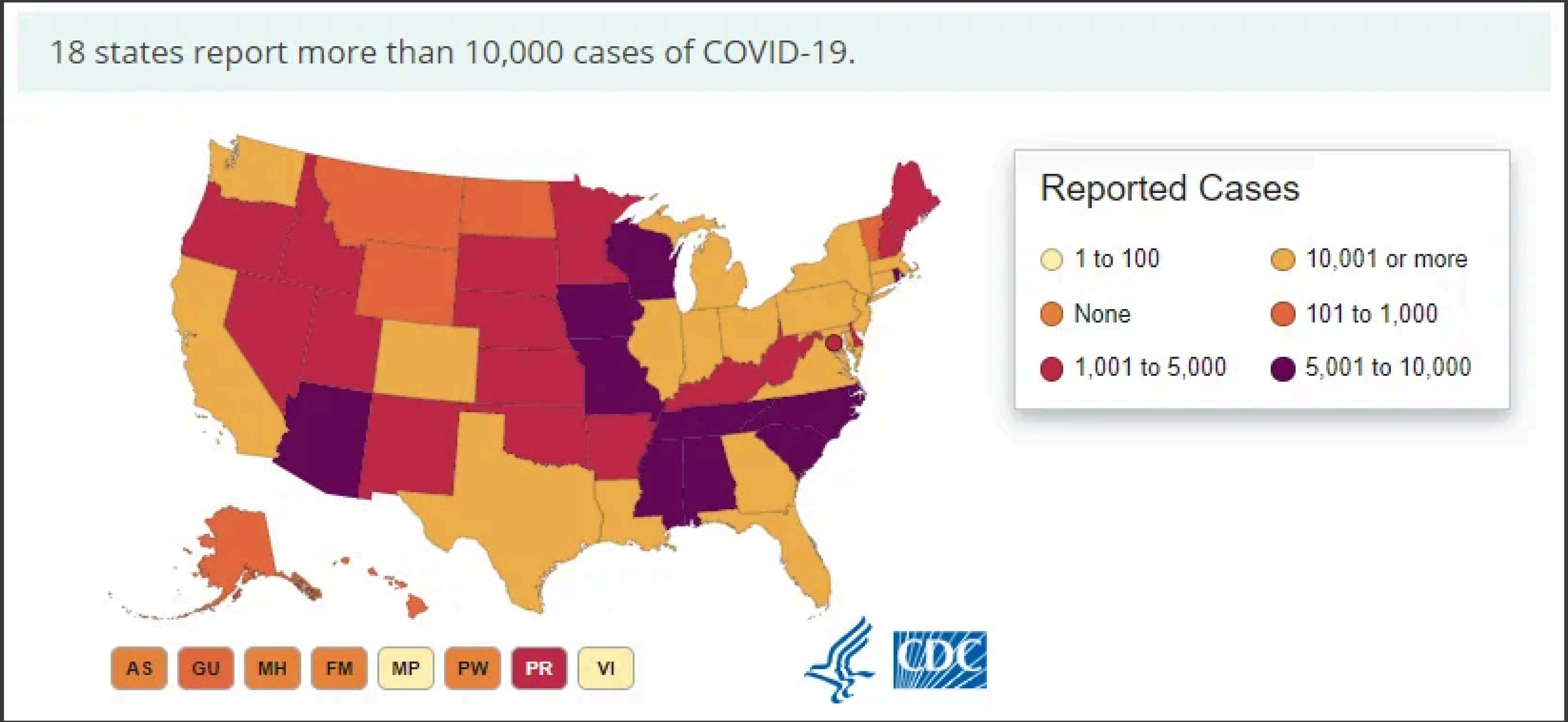
Encourages viewer to perceive *angle* instead of *area*.



Bar Chart

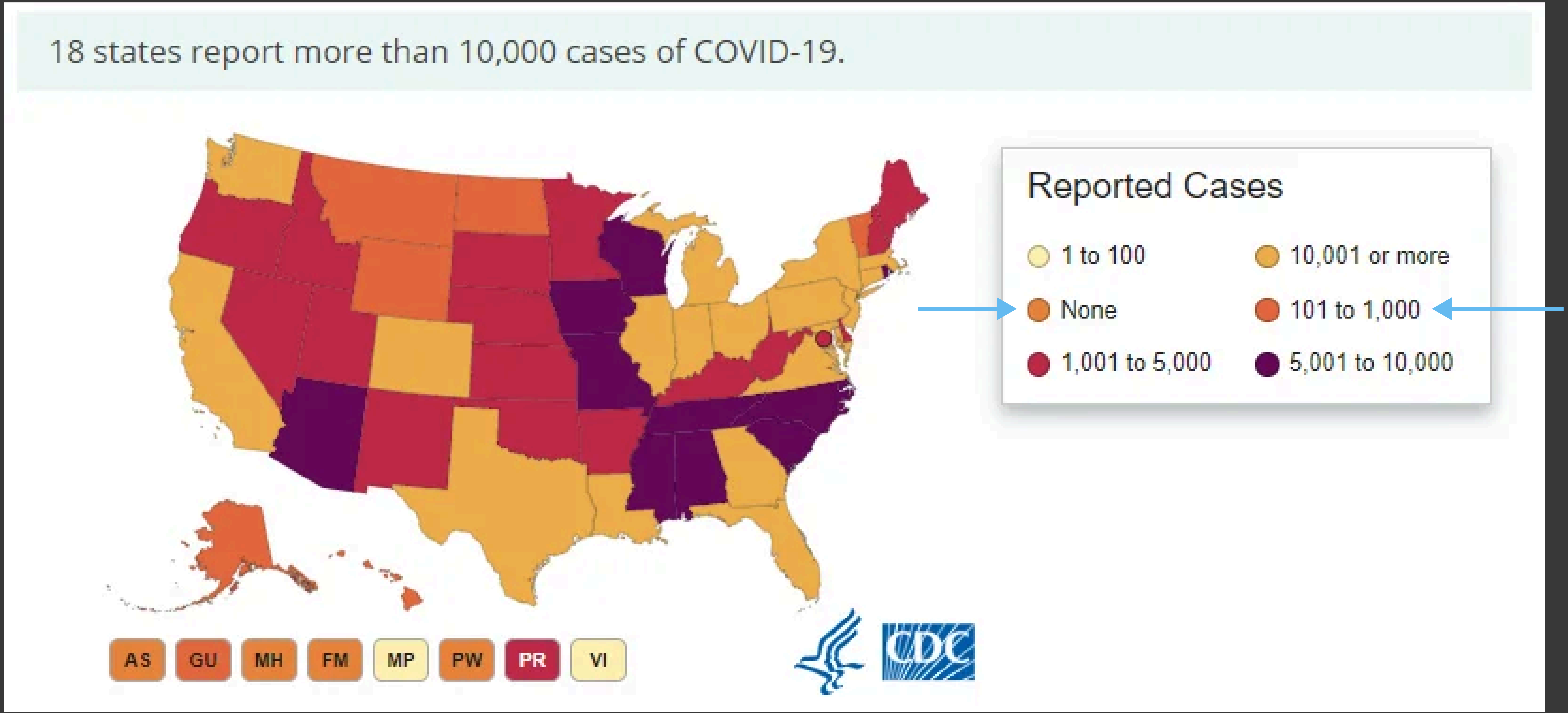
Like a donut chart, but not bendy! Your eye is good at comparing straight lengths!

Source: DVP



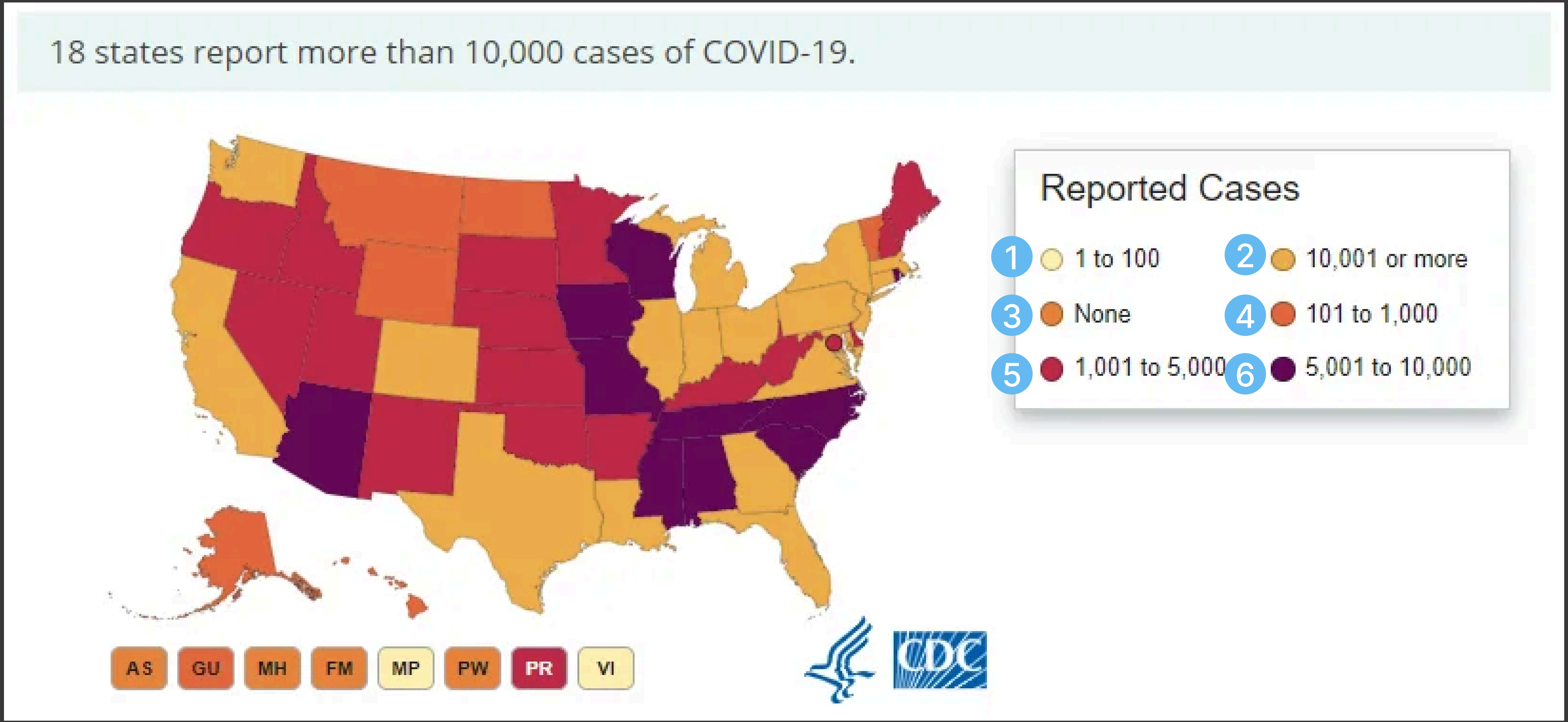
Bad Color Encoding,
Quantitative ↣ Categorical (CDC, 2020)

[Source]()



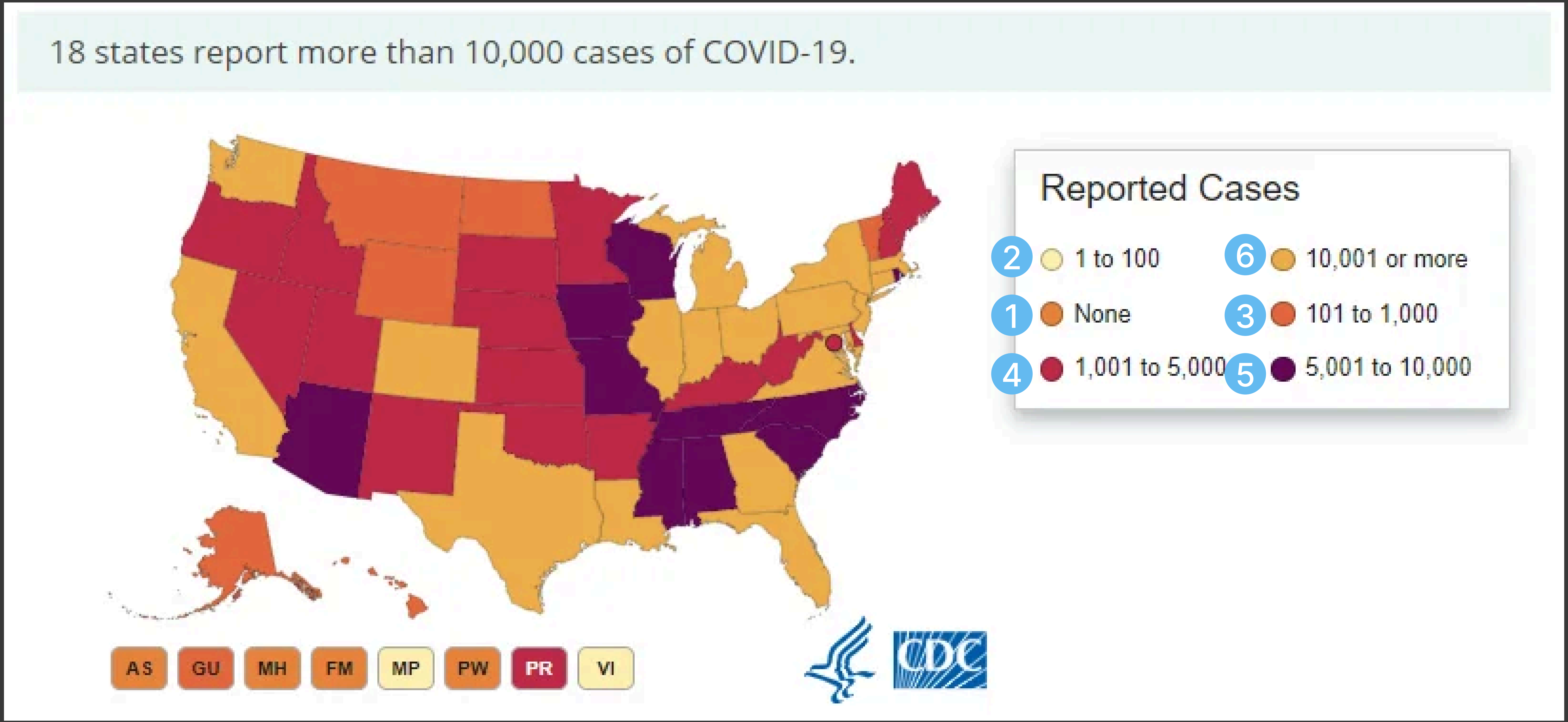
Bad Color Encoding, Quantitative ↣ Categorical (CDC, 2020)

Source



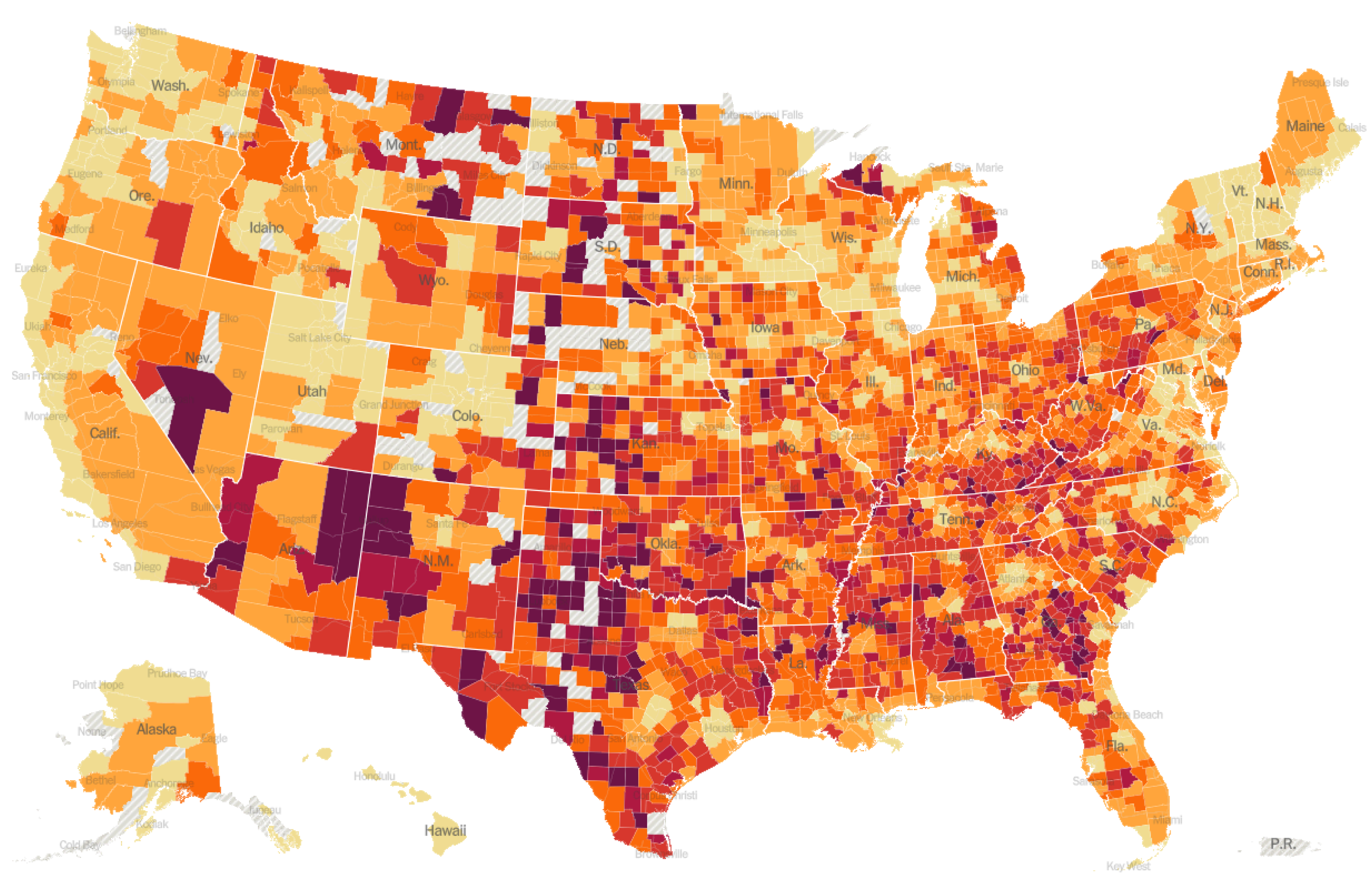
Bad Color Encoding, Quantitative ↣ Categorical (CDC, 2020)

Source



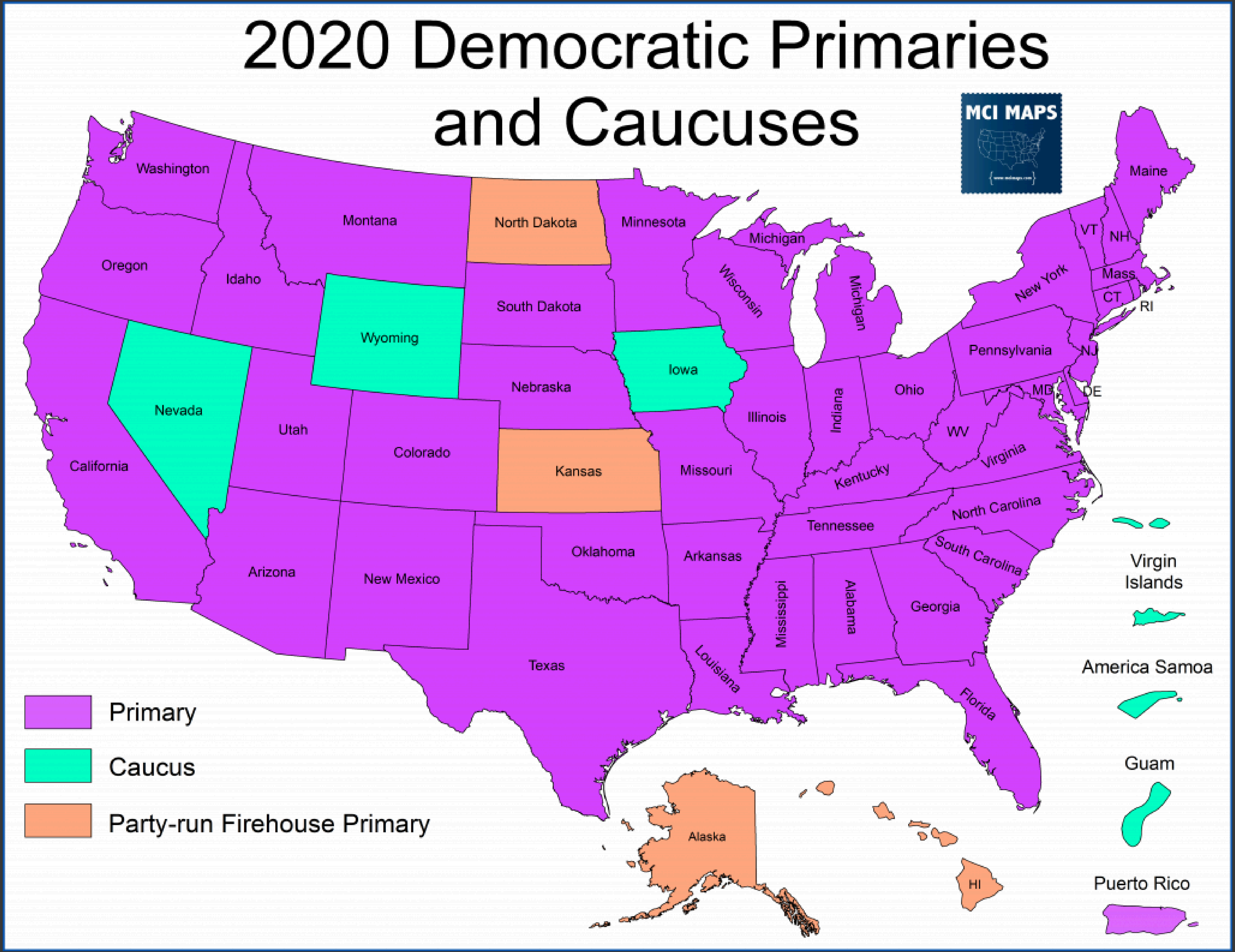
Bad Color Encoding,
Quantitative ↣ Categorical (CDC, 2020)

Source



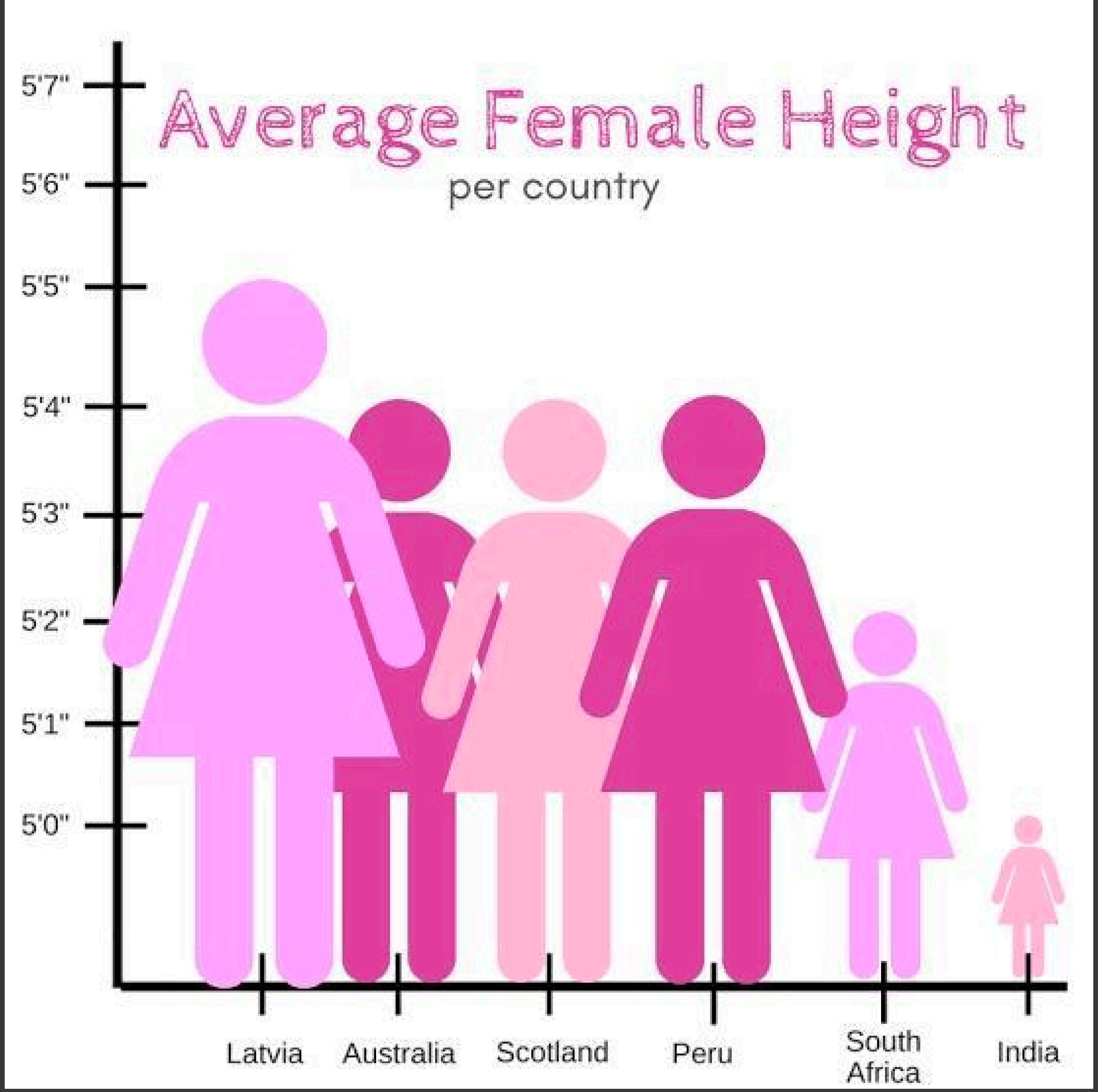
All-Time COVID Deaths (NYTimes, 2023)

Source



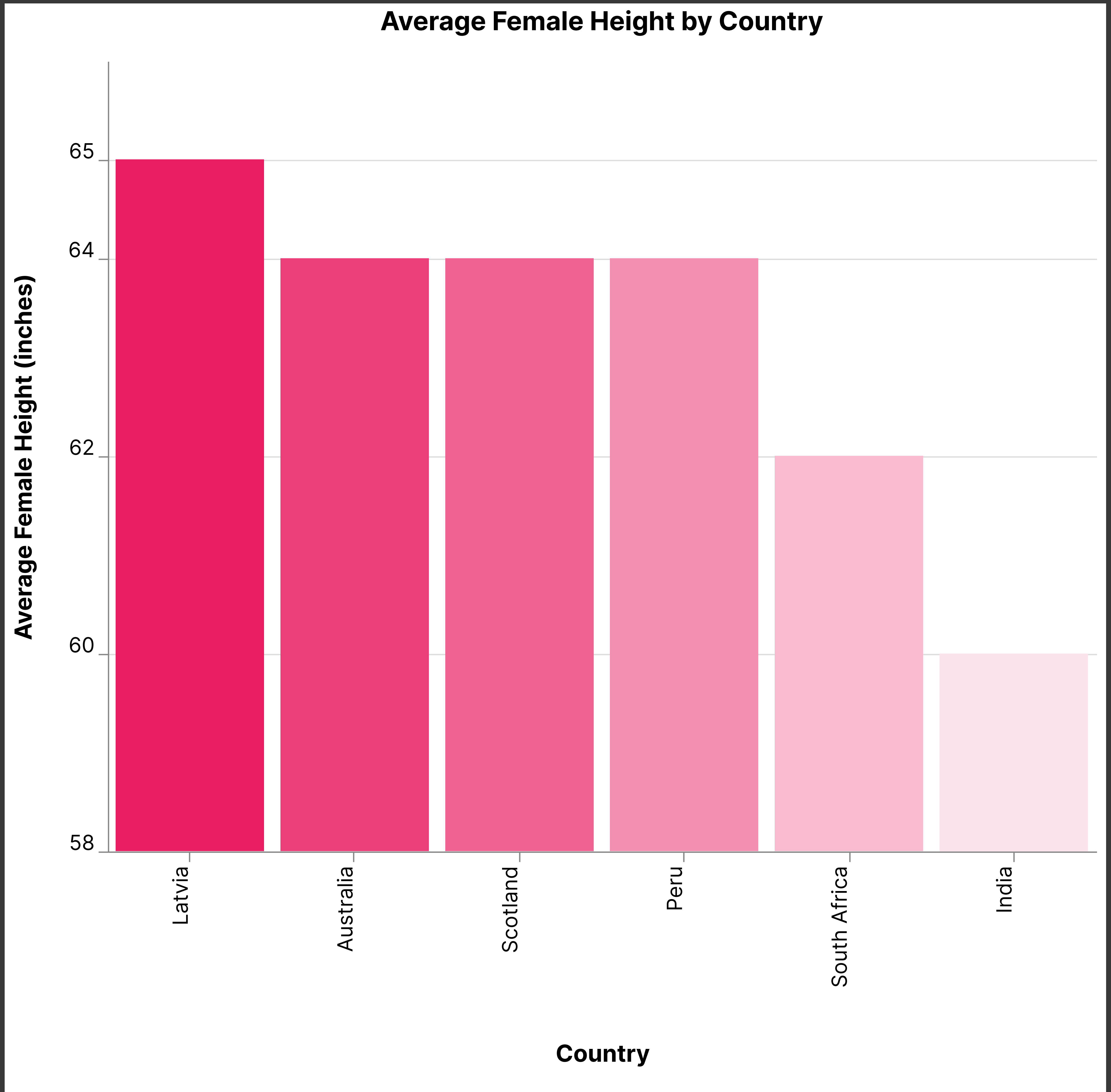
Primaries vs. Caucuses (MCI Maps, 2020)

Source

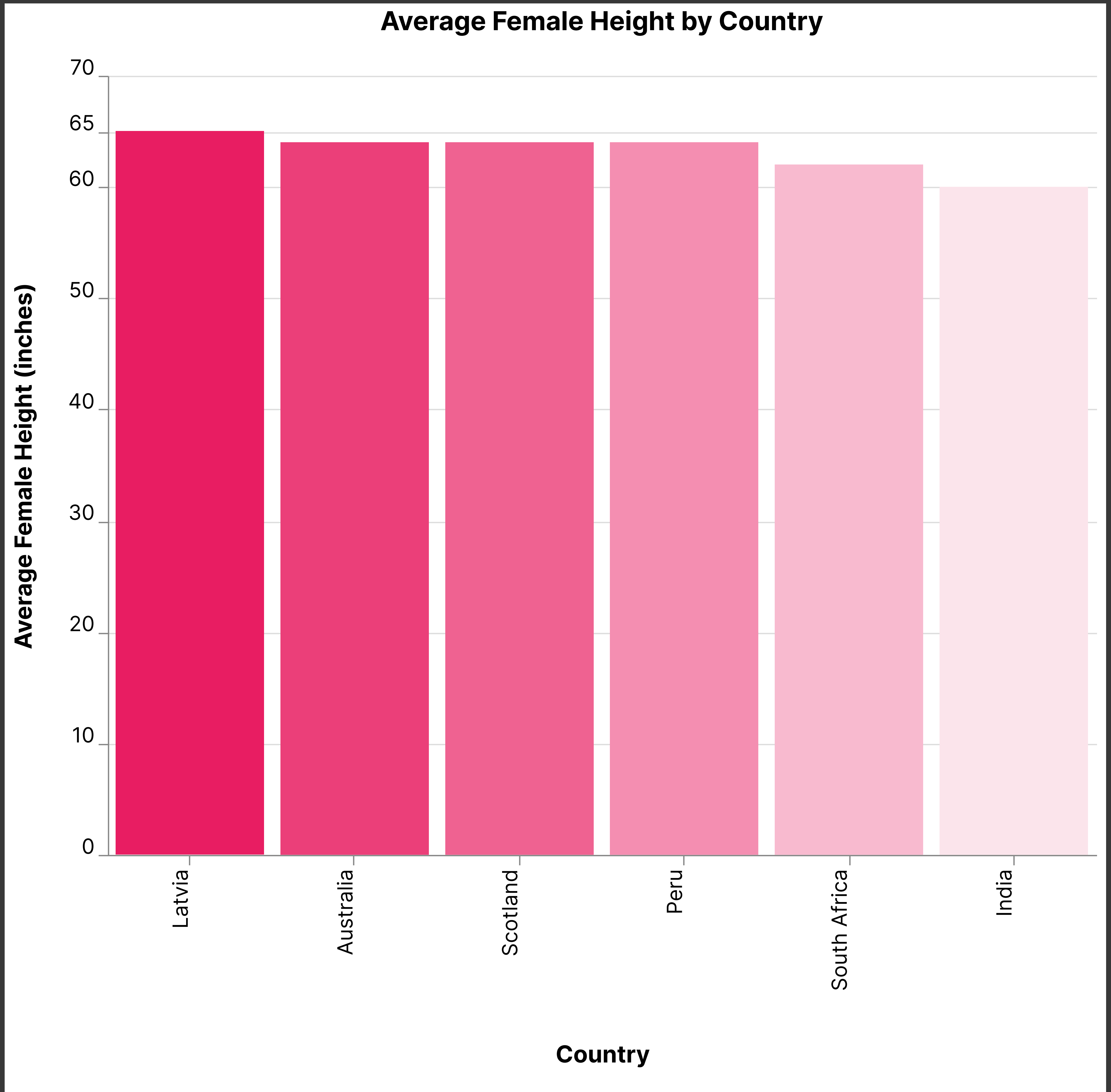


Average Female Height

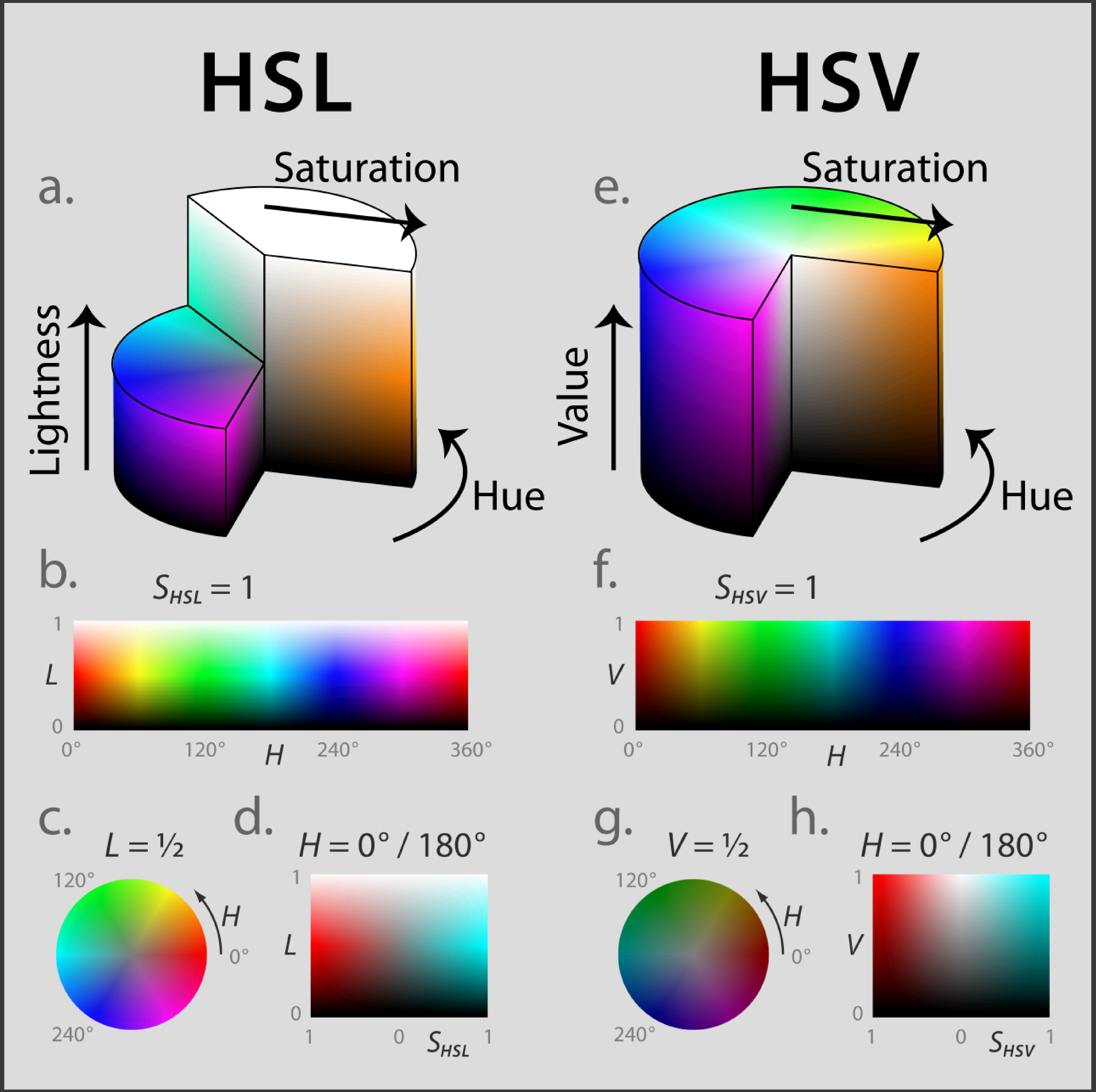
Source



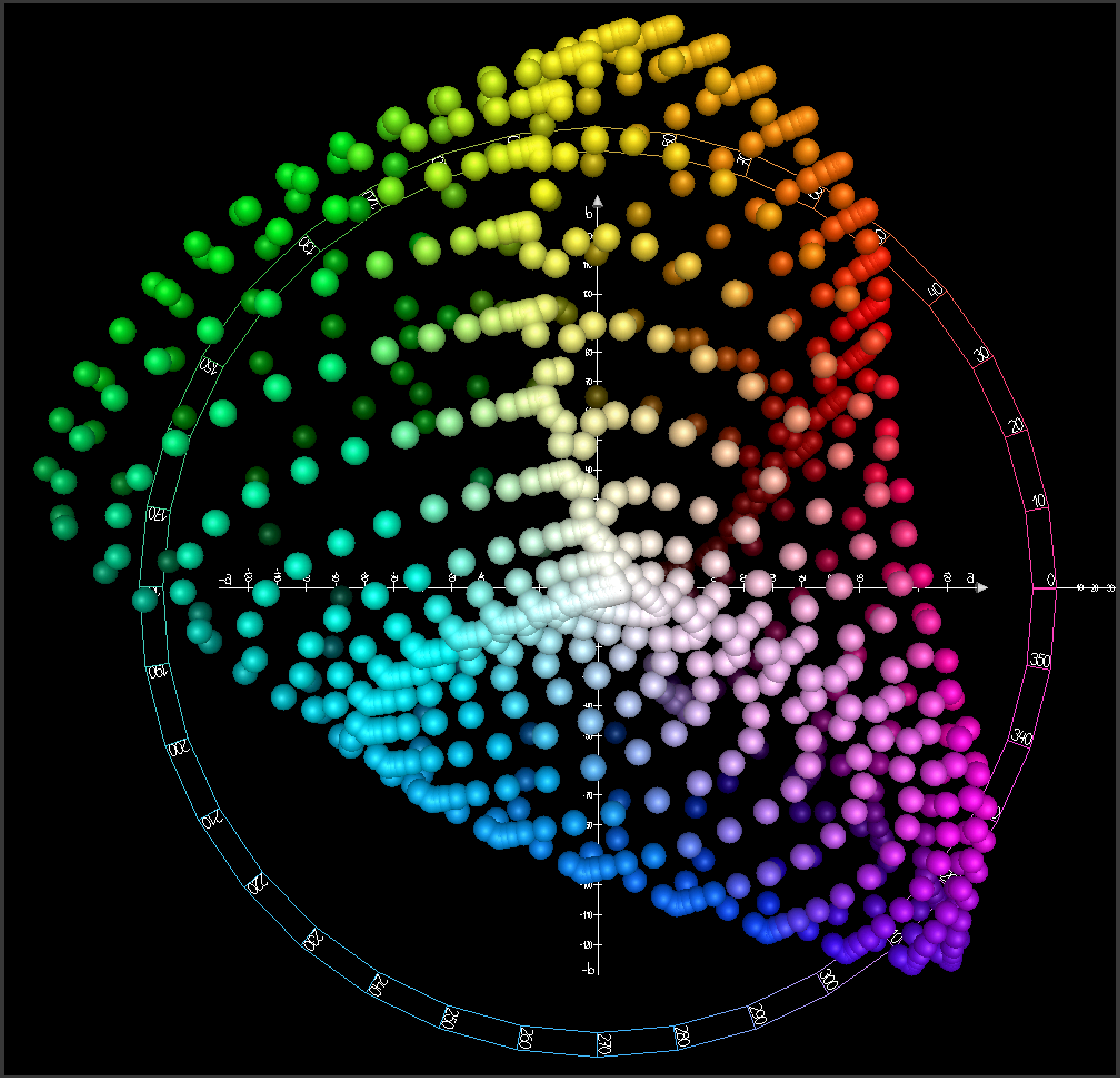
Average Female Height



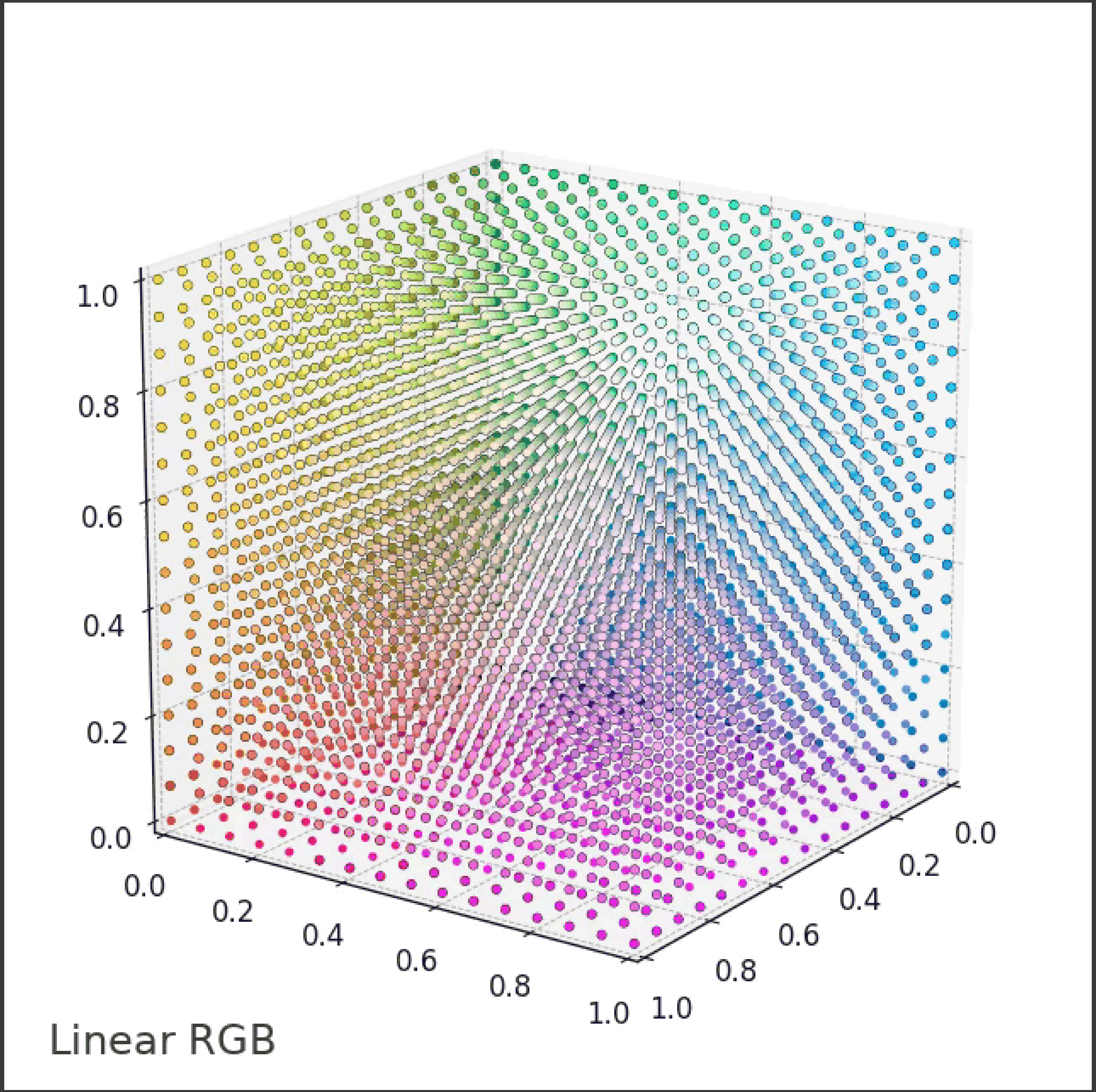
Average Female Height


HSL
HSV
a.
Saturation
Lightness
Hue
e.
Saturation
Value
Hue
b.
$S_{HSL} = 1$
1
L
0
0°
120°
H
240°
360°
f.
$S_{HSV} = 1$
1
V
0
0°
120°
H
240°
360°
c.
$L = ½$
120°
H
0°
240°
d.
$H = 0° / 180°$
1
L
0
1
0
S_{HSL}
1
g.
$V = ½$
120°
H
0°
240°
h.
$H = 0° / 180°$
1
V
0
1
0
S_{HSV}
1

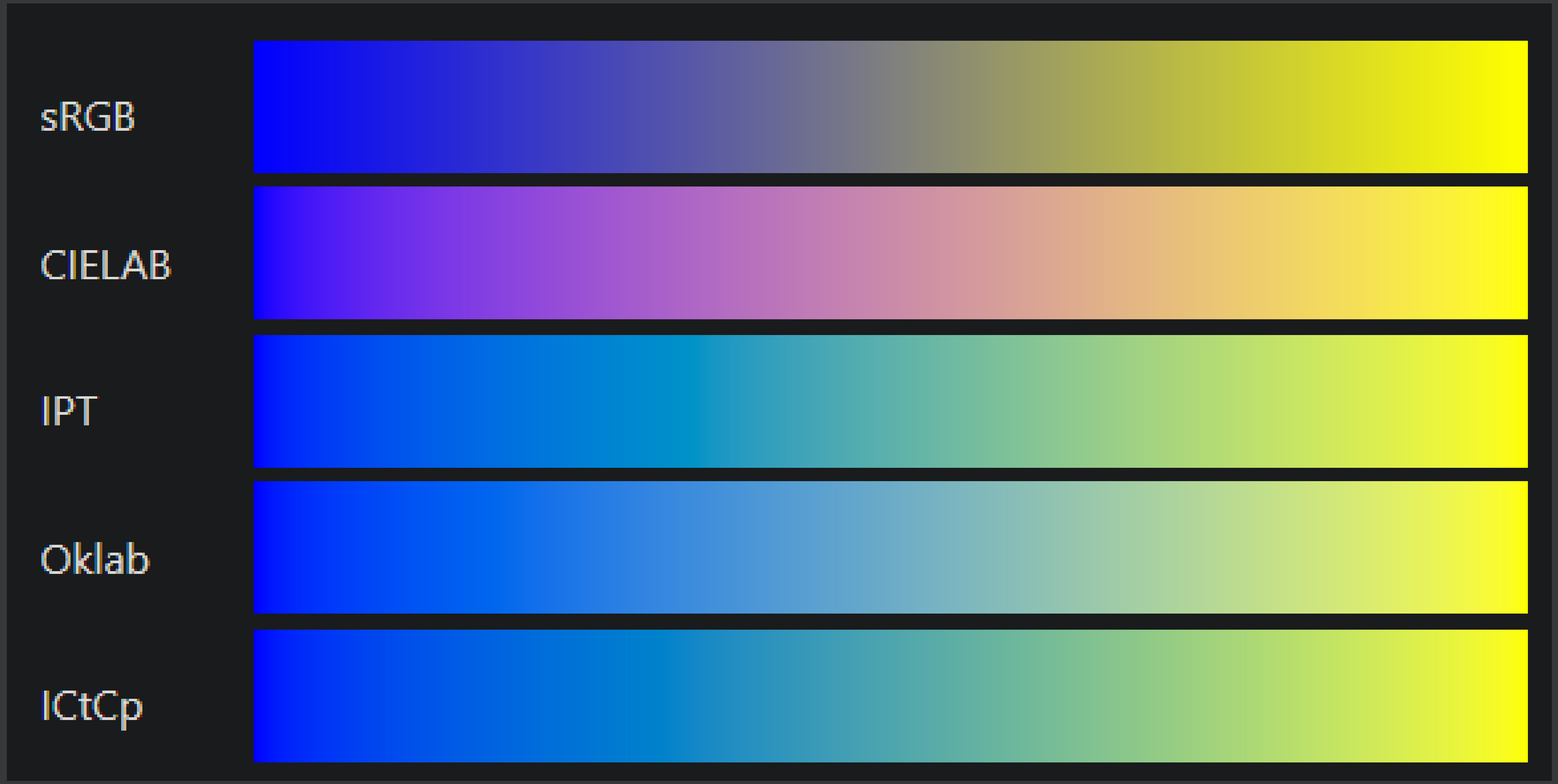
Wikipedia

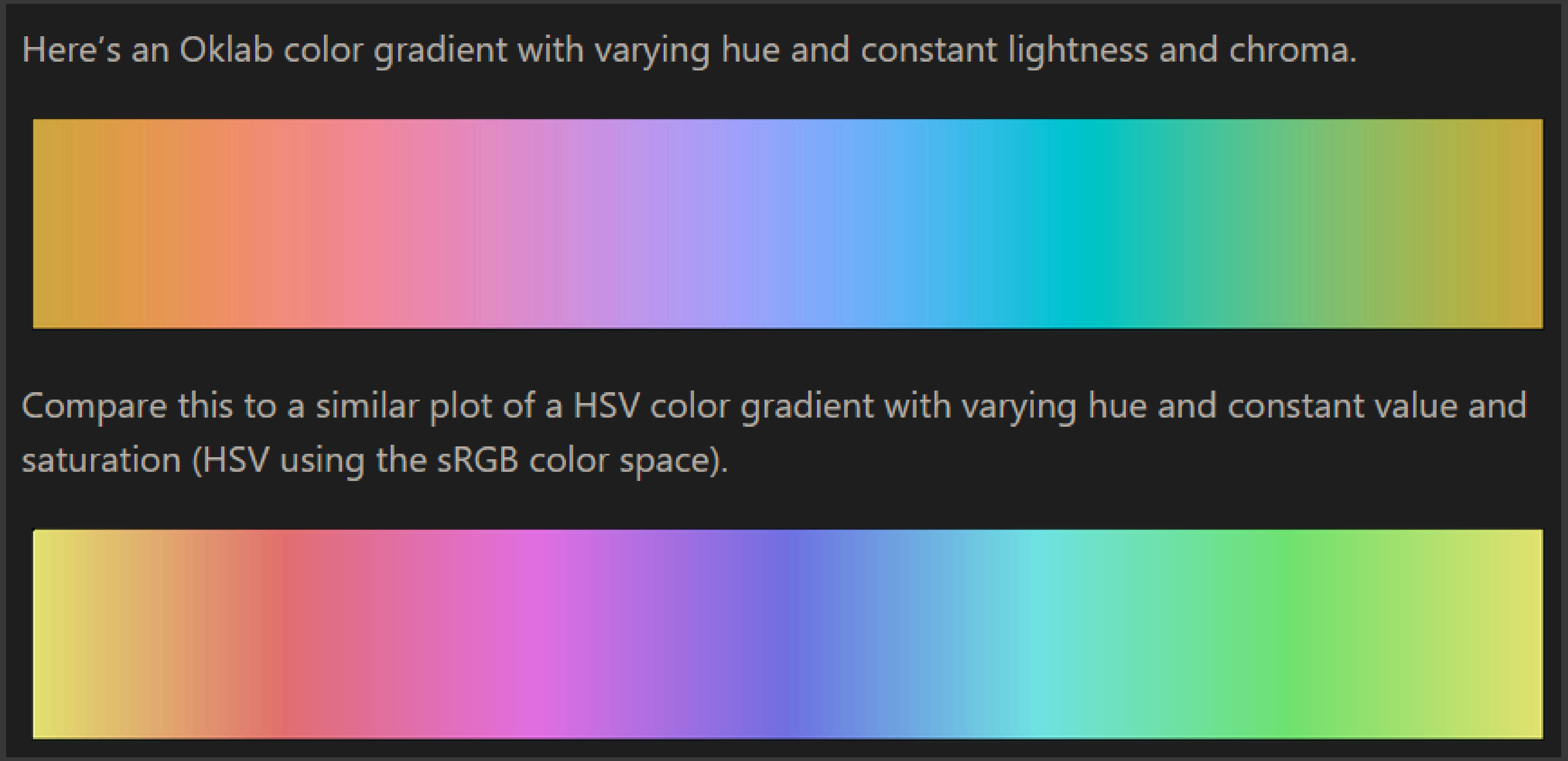


Wikipedia

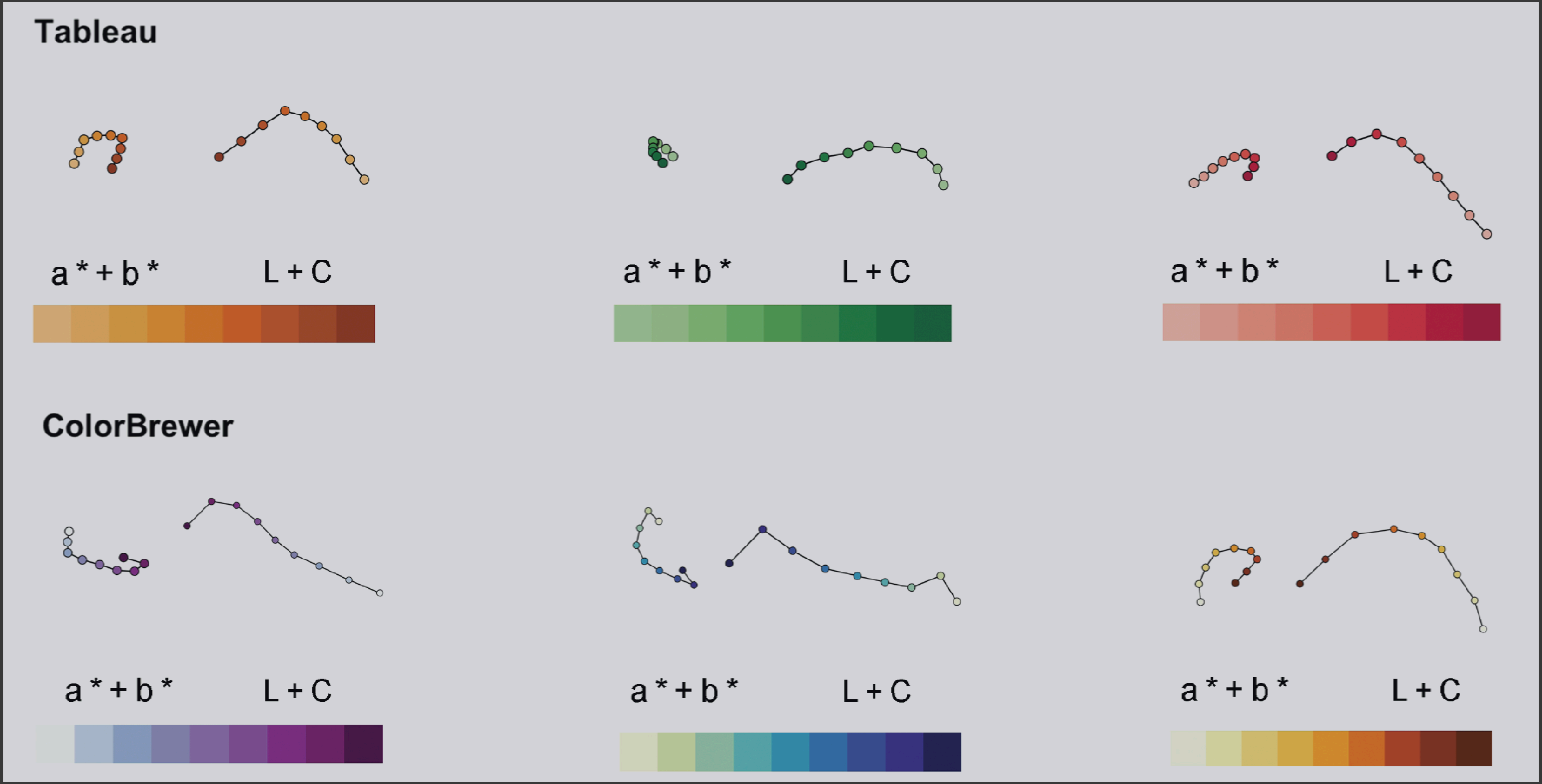


Wikipedia



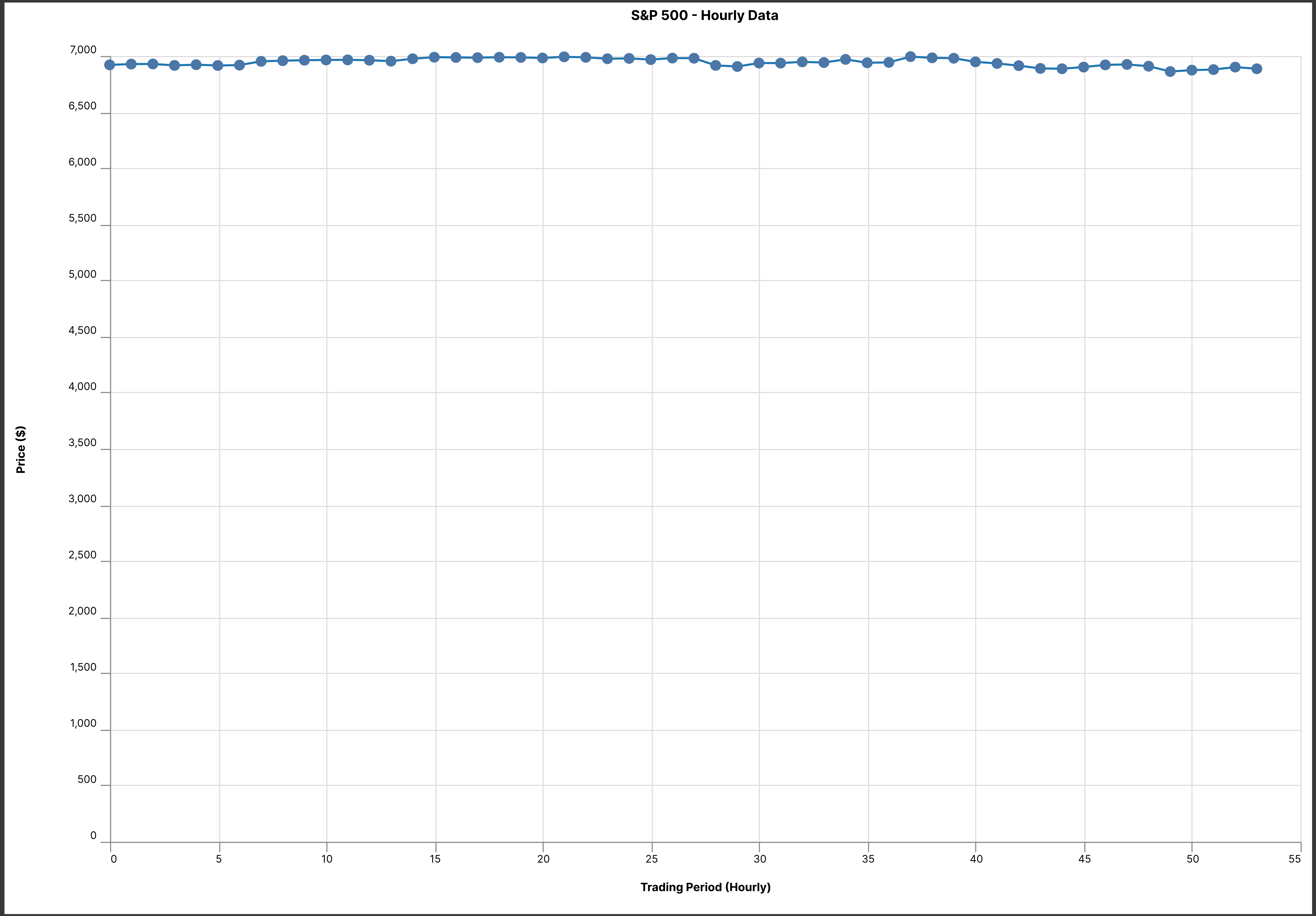


Source: Björn Ottosson,
Raph Linus



Smart et al., VIS 2019

Tasks


S&P 500 - Hourly Data
Price ($)
7,000
6,500
6,000
5,500
5,000
4,500
4,000
3,500
3,000
2,500
2,000
1,500
1,000
500
0
0
5
10
15
20
25
30
35
40
45
50
55
Trading Period (Hourly)

S&P 500


S&P 500 - Hourly Data
Price ($)
7,000
6,980
6,960
6,940
6,920
6,900
6,880
6,860
6,840
−5
0
5
10
15
20
25
30
35
40
45
50
55
Trading Period (Hourly)

S&P 500

> “Overview first, zoom and filter, then details on demand.”

— Bruce Shneiderman, “The Eyes Have It”

There are many visual design guidelines but the basic principle might be summarized as the Visual Information Seeking Mantra:

Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand
Overview first, zoom and filter, then details-on-demand

Each line represents one project in which I found myself rediscovering this principle and therefore wrote it down it as a reminder. It proved to be only a starting point in trying to characterize the multiple information-visualization innovations occurring at university, government, and industry research labs.

Shneiderman, 1996

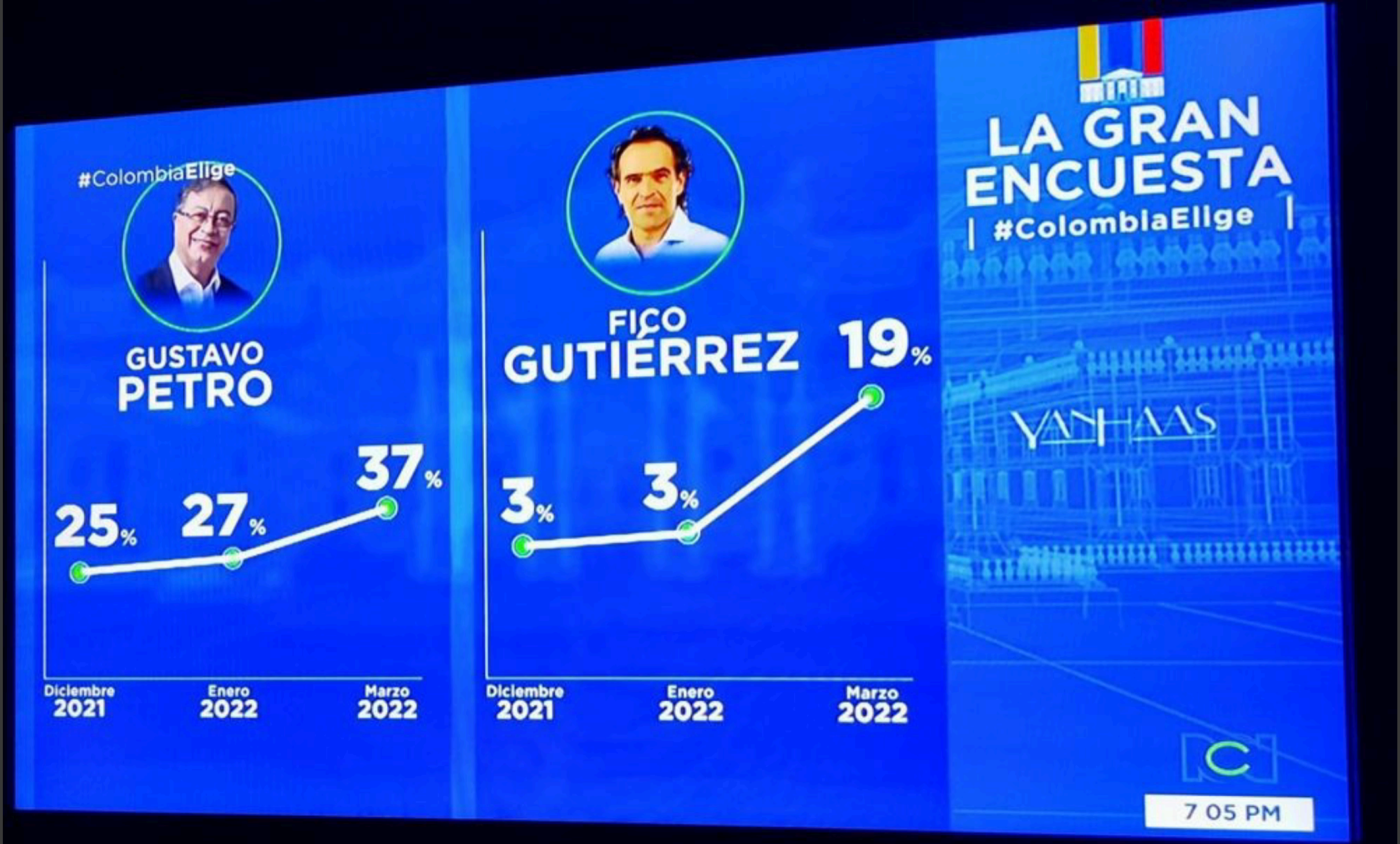

Search Google Maps
Restaurants
Hotels
Things to do
Museums
Transit
Pharmacies
ATMs
CENTENNIAL DISTRICT
EAST PARKSIDE
Checkers
CARROLL PARK
Cathedral Cemetery
KidZooU
Philadelphia Zoo
Skate The Foundry - West Philly
Boathouse Row
Schuylkill River
ANGELINO'S RESTAURANT
Hilltown Tavern Fairmount
The Black Taxi
The Met Philadelphia
Eastern State Penitentiary
YORKTOWN
LUDLOW
The Turkish Brew
Oasis Haven
The National Shrine of St. John Neumann
Jim's West
Save A Lot
HADDINGTON
Monica Restaurant
MILL CREEK
Dollar General
BELMONT
Baz's Hot Chicken & Gyro
MANTUA
HAVERFORD NORTH
Philadelphia Museum of Art
FAIRMOUNT
SPRING GARDEN
Cicala at the Divine Lorraine
Santucci's Original Square Pizza and Bar...
Temporarily closed
Osteria
POPLAR
Piazza Alta
Kaplan's New Model Bakery
Pera Turkish
Cantina Dos Segundos
The Random Tea Room & Curiosity Shop
Edgar Allan Poe National Historic Site
The Kettle Black
Bart's Bagels - West Philadelphia
Fresh Donuts
Top Taste
DUNLAP
Vidas Athletic Complex
POWELTON VILLAGE
Rodin Museum
Dim Sum & Noodle
Cerveau
Love City Brewing
Spring Garden
WEST PHILADELPHIA
CVS
ALDI
WEST POWELTON
Woosa Bubble Tea
Indian Sizzler
Chengdu Famous
King of Tandoor & Biryani House
Sabrina's Cafe
Localhost Philly
William L Sayre High School
West Philadelphia YMCA
46th St
Schmear Bagel Parlor - 36th St
Ramen
Greek From Greece (GFG)
The Franklin Institute
Moustaki Authentic Gyros
Urban Farmer Philadelphia
Race-Vine
Nan Zhou Hand Drawn Noodle House
Sang Kee Peking Duck House
Paul Robeson House & Museum
Kilimanjaro restaurant
Asad's Hot Chicken
Han Dynasty University City
Nanu's Hot Chicken
William H. Gray III 30th Street
30th Street
Thanal Indian Tavern
CHICHA San Chen
Ray's Cafe & Tea House
Tuna Bar
Dim Sum House by Jane G's - University City
Latao Hotpot and BBQ
Sang Kee Noodle House
Cucina Zapata
Vernick Coffee Bar
Suburban
Kyuramen x TBaar Philadelphia
Dim Sum Garden
WALNUT HILL
Saad's Halal Restaurant
Temporarily closed
U Town
Molly Tea (UPenn)
Louie Louie
Masala Kitchen Kati Rolls & Platters
Maison Sweet
DuJour Cafe & Catering
Healthy Picks
Jefferson
Bubblefish
Elfreth's Alley Museum
COBBS CREEK
GARDEN COURT
SPRUCE HILL
Pelicana Chicken and Boba Tea...
Lemon Grass Thai
DaMò Pasta Lab
Harper's Garden
Nom Nom Ramen
Schmear Bagel Parlor
8th & Market Street
Philly Style Bagels Old City
Tomo Sushi & Ramen
Hospital of the University of Pennsylvania...
Tacos Don Memo
University of Pennsylvania
Rosy's Taco Bar
The Dandelion
Giuseppe & Sons
Spice 28
ICI Macarons & Cafe
CEDAR PARK
Doro Bet
Renata's Kitchen
Penn Museum
Ambrosia Ristorante
1-900-ICE-CREAM
Luna BYOB
Double Knot
El Vez
Morimoto
La Famiglia Ristorante
The Franklin Fountain
Clark Park
Vietnam Café
Fu-Wah Mini Market
Penn Medicine
Friday Saturday Sunday
Schlesinger's
Grandma's Philly
Amada Philadelphia
Malooga - Old City
FITLER SQUARE
Trattoria Carina
Miss Saigon Vietnamese Restaurant
Lounge
Knead Bagels
Café Walnut
Buk Chon Korean Cuisine
Café Lutécia
UNI Japanese Cuisine | Japanese, Hibachi &...
Bud & Marilyn's
Middle Child
Bleu Sushi
Zahav
ANGORA
The Woodlands
SouthGate
ILLATA
WASHINGTON SQUARE WEST
Spruce Street Harbor Park
SQUIRREL HILL
Chester Av & 49th St
Four Worlds Bakery
Kingsessing Recreation Center
DEVIL'S POCKET
Bistro La Baia BYOB
Rex at the Royal
The Breakfast Den
Philadelphia's Magic Gardens
Southwest Philadelphia Seventh-day Adventist...
Hardy Williams Elementary & High...
SCHUYLKILL
SOUTHWEST CENTER CITY
Ishkabibble's II
Woodrow's Sandwich Shop
Jim's South St
Pizzeria Stella
Queen & Rook
Cobbs Creek Park
FORGOTTEN BOTTOM
McDonald's
Dock Street Brewery South
Dante & Luigi's
Angelo's Pizzeria
Ralph's Italian Restaurant
Mawn
Isgro Pastries
Sakana Omakase Sushi
KINGSESSING
Philadelphia Trolley Works
Italian Market 9th Street
John's Water Ice
Royal Sushi & Izakaya
The Bagel Place
Sky Cafe
Phở 75
Pho Ga Thanh Thanh
Taco Heart
GRAYS FERRY
Ricci's Hoagies
South Philly Barbacoa
Sophie's Kitchen
MT MORIAH
Google Maps
Layers

1. Overview
2. Zoom
3. Filter
4. Details on demand
5. Relate
6. History
7. Extract

Shneiderman, 1996

Brehmer & Munzner, TVCG 2013

Affect & Salience


#ColombiaElige
GUSTAVO PETRO
25%
27%
37%
Diciembre 2021
Enero 2022
Marzo 2022
FICO GUTIÉRREZ
3%
3%
19%
Diciembre 2021
Enero 2022
Marzo 2022
LA GRAN ENCUESTA
#ColombiaElige
YANHAAS
7 05 PM

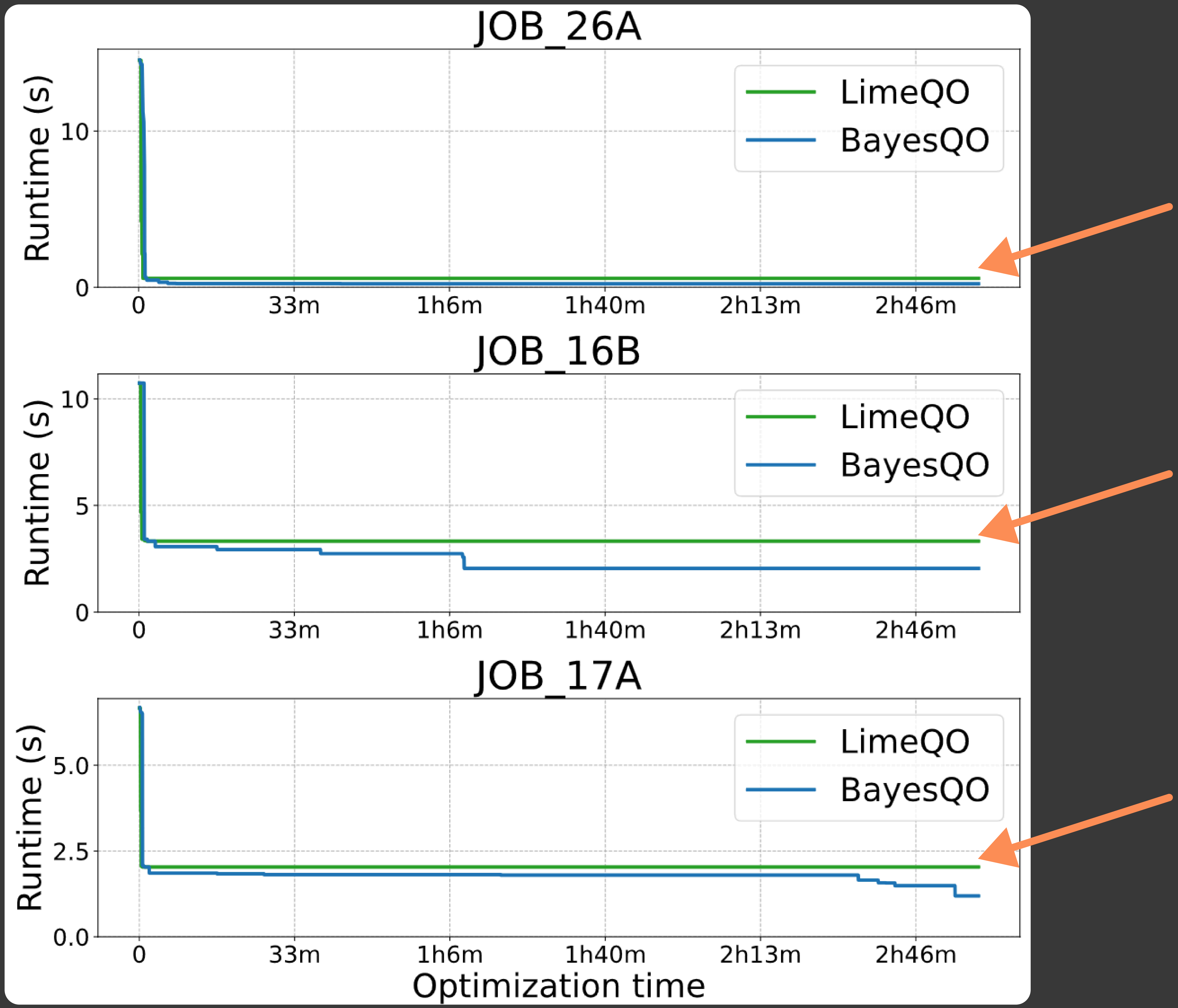
A Personal Tale

Visualization is *narrative*

Visualization is a form of storytelling.

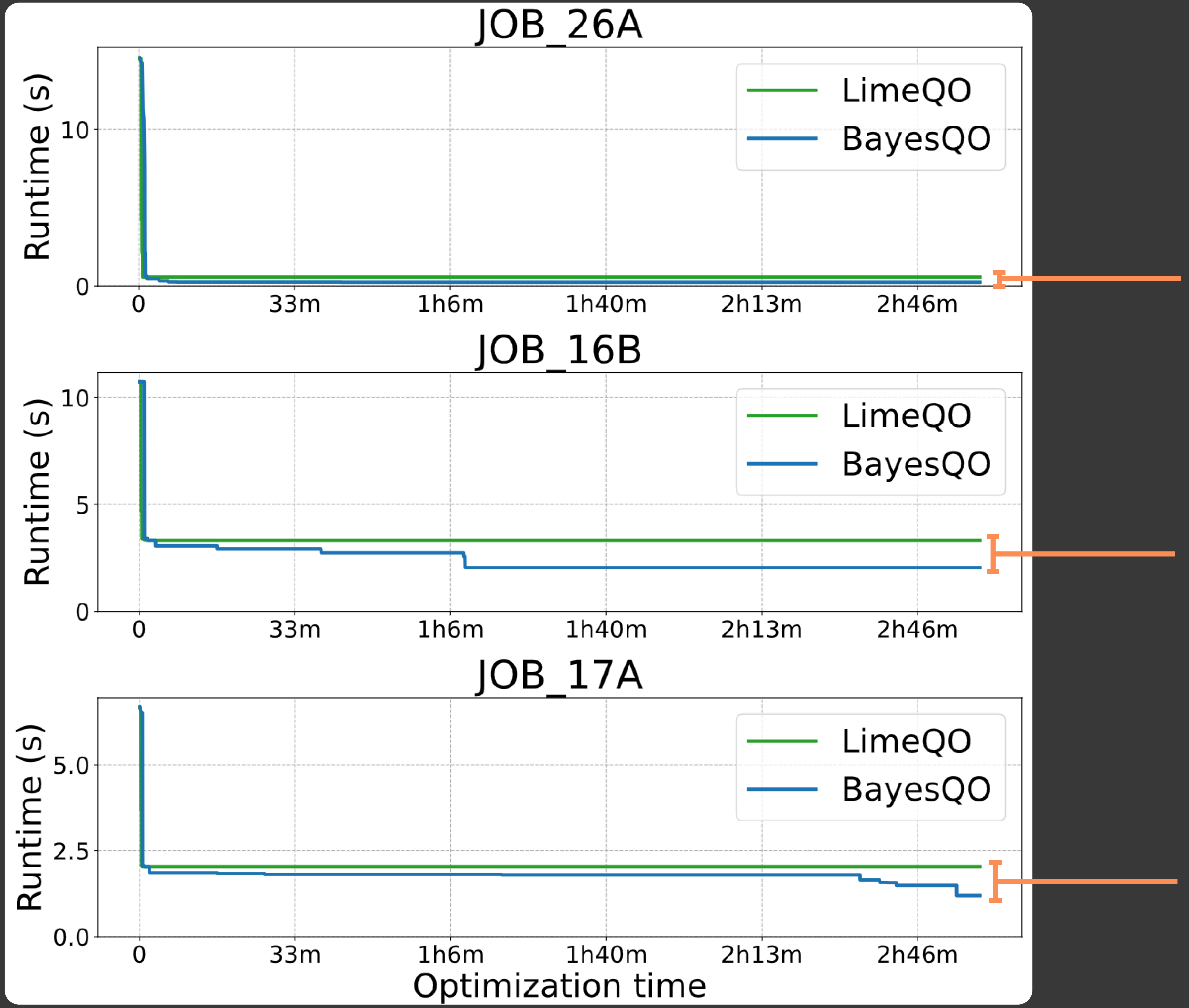
But what you intend and think is salient is not the same as what the viewer will think is salient!

Case Study: My Own Research Paper



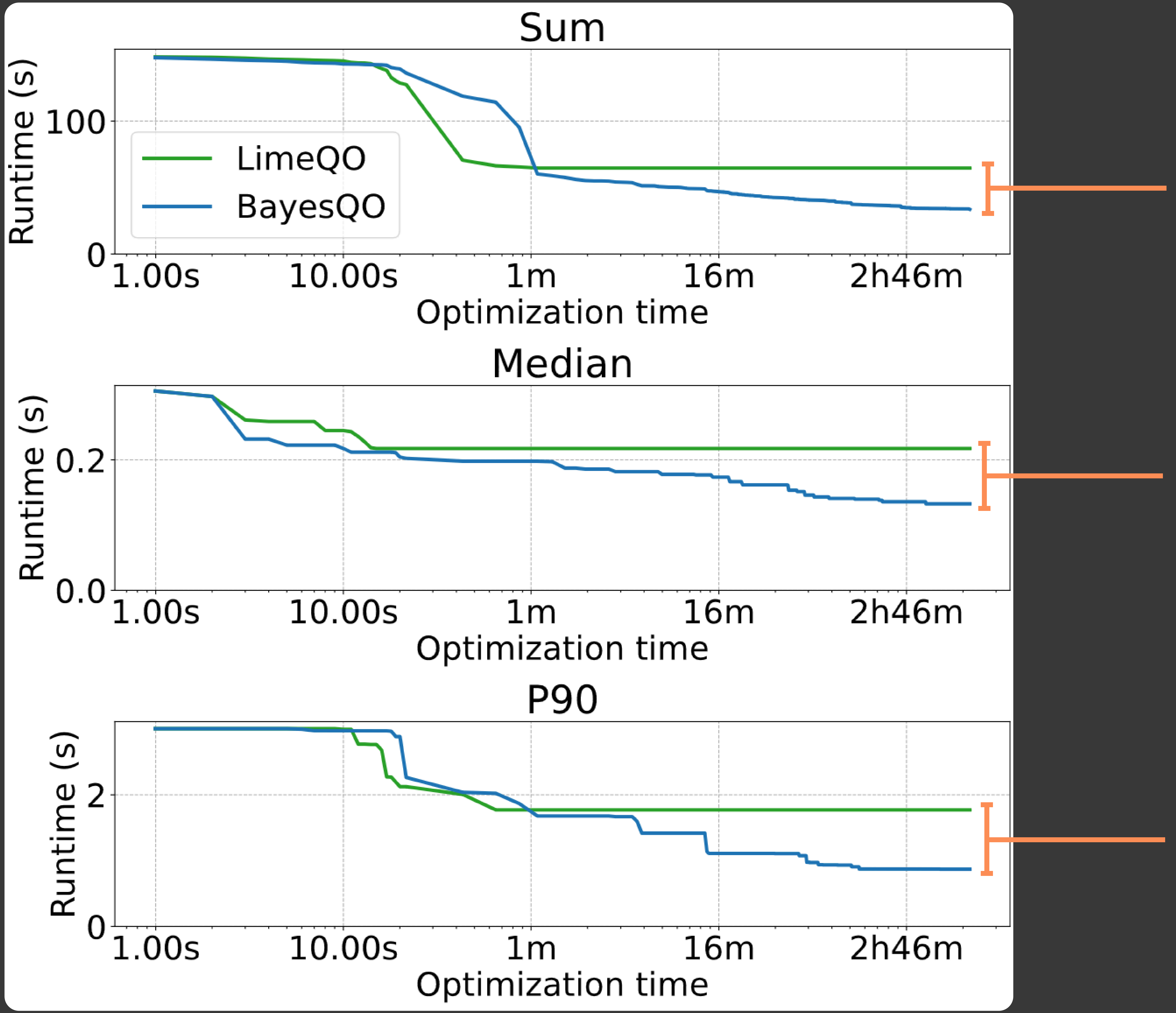
Us: “Look, the green line bottoms out and then doesn’t get any better!”

Case Study: My Own Research Paper

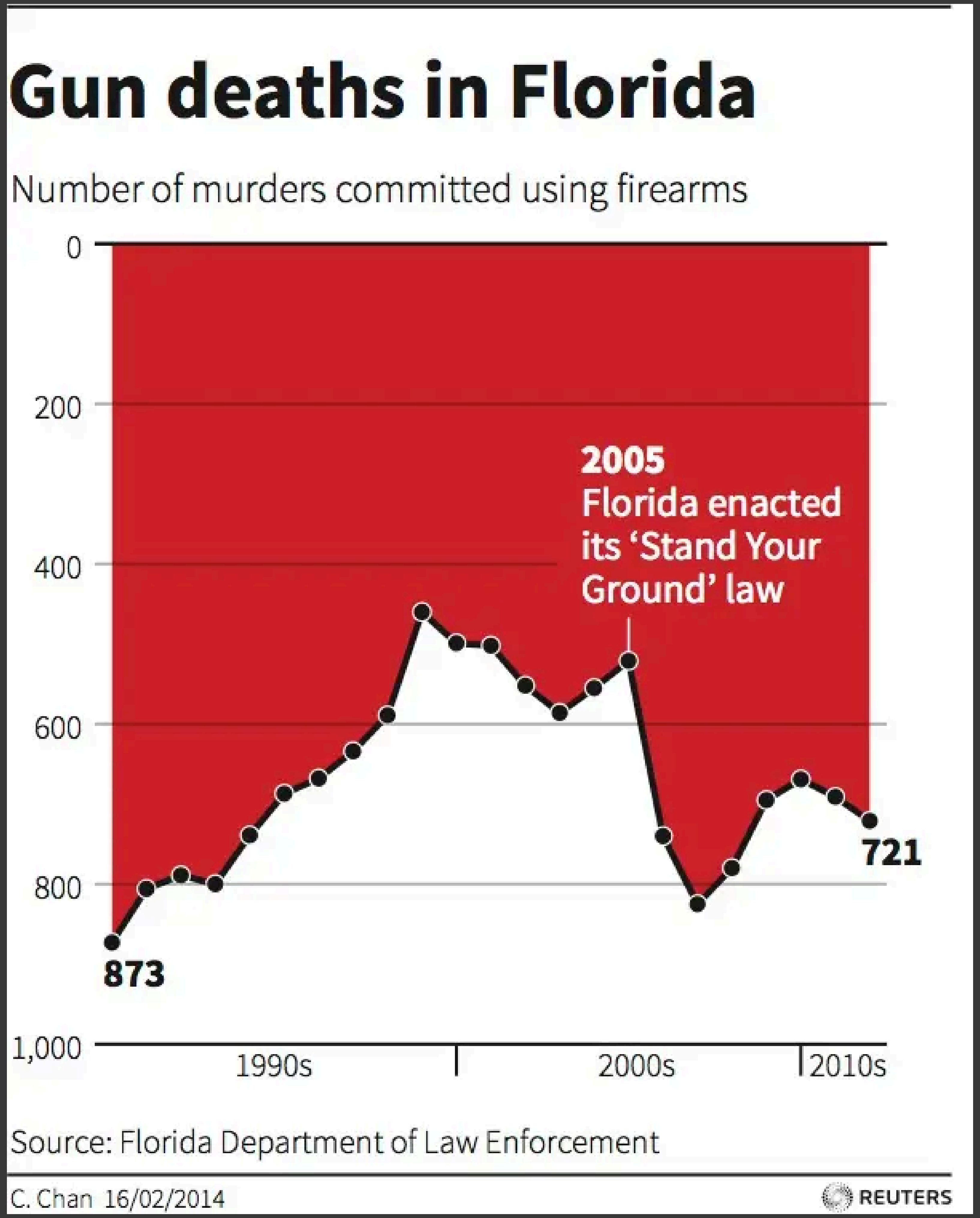


Them: “Wow, your method isn’t that much better!”

Case Study: My Own Research Paper

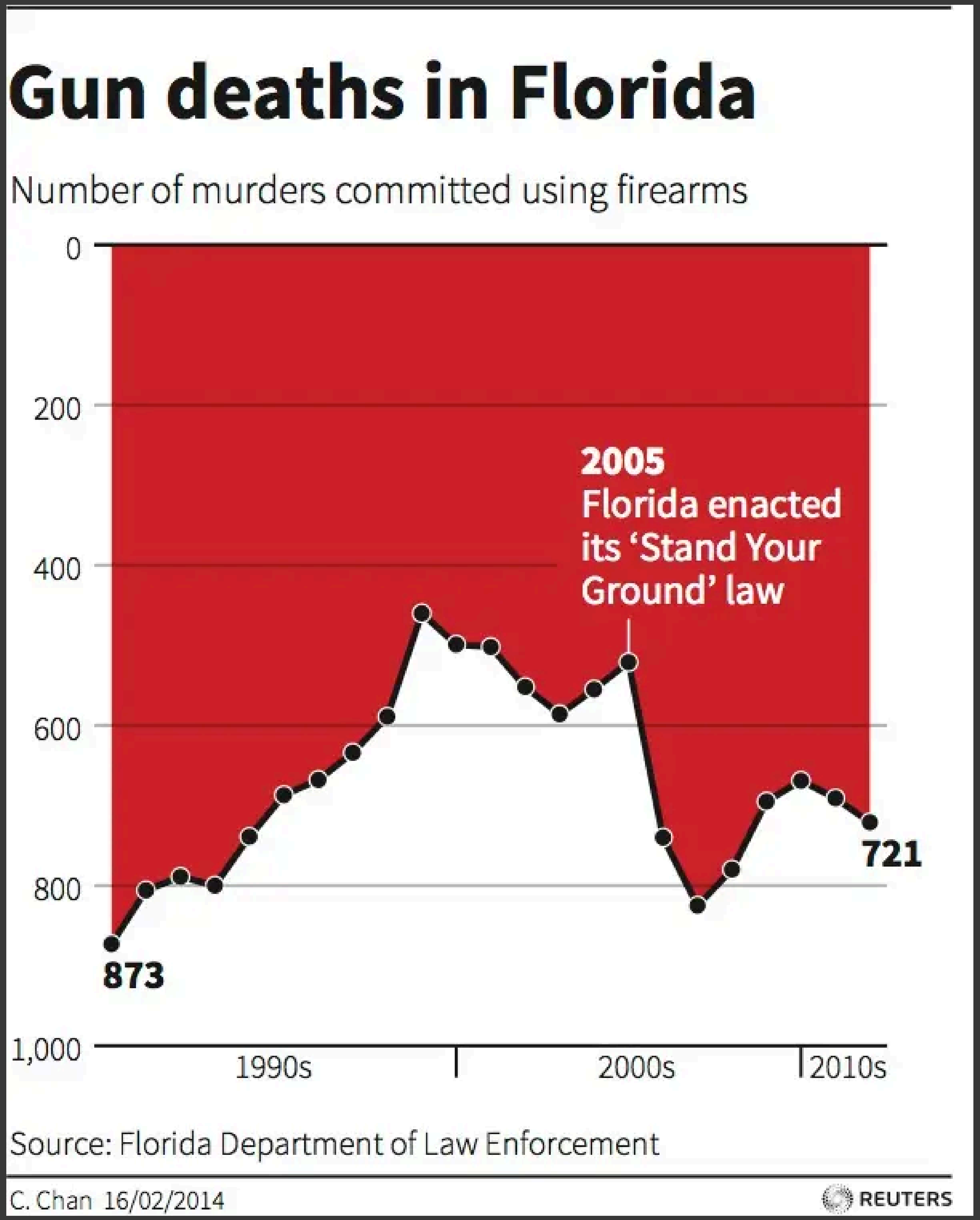


Us: “Look, the gaps are bigger now!”

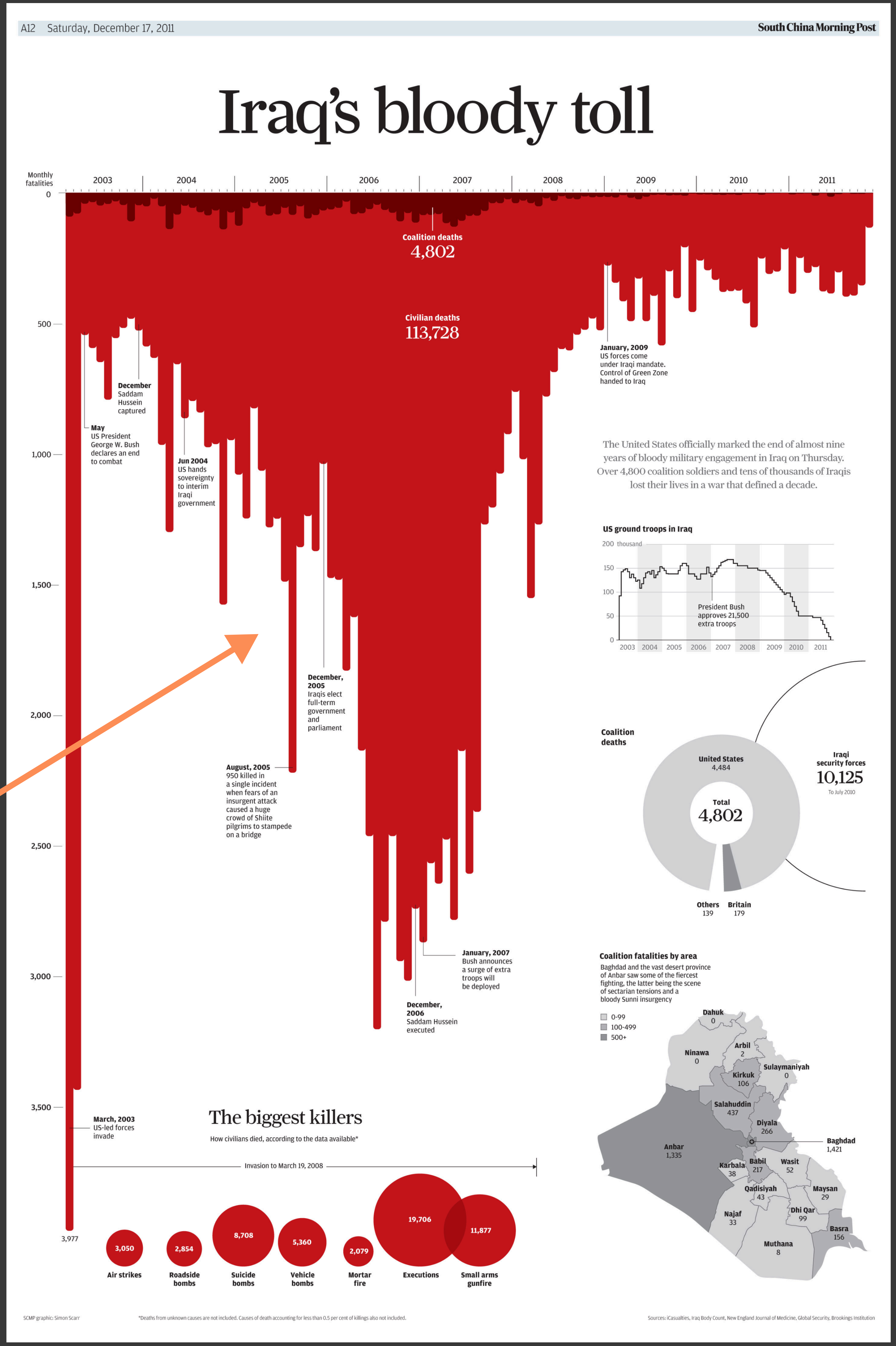


Conventions? Affect? Art?

Source: Reuters



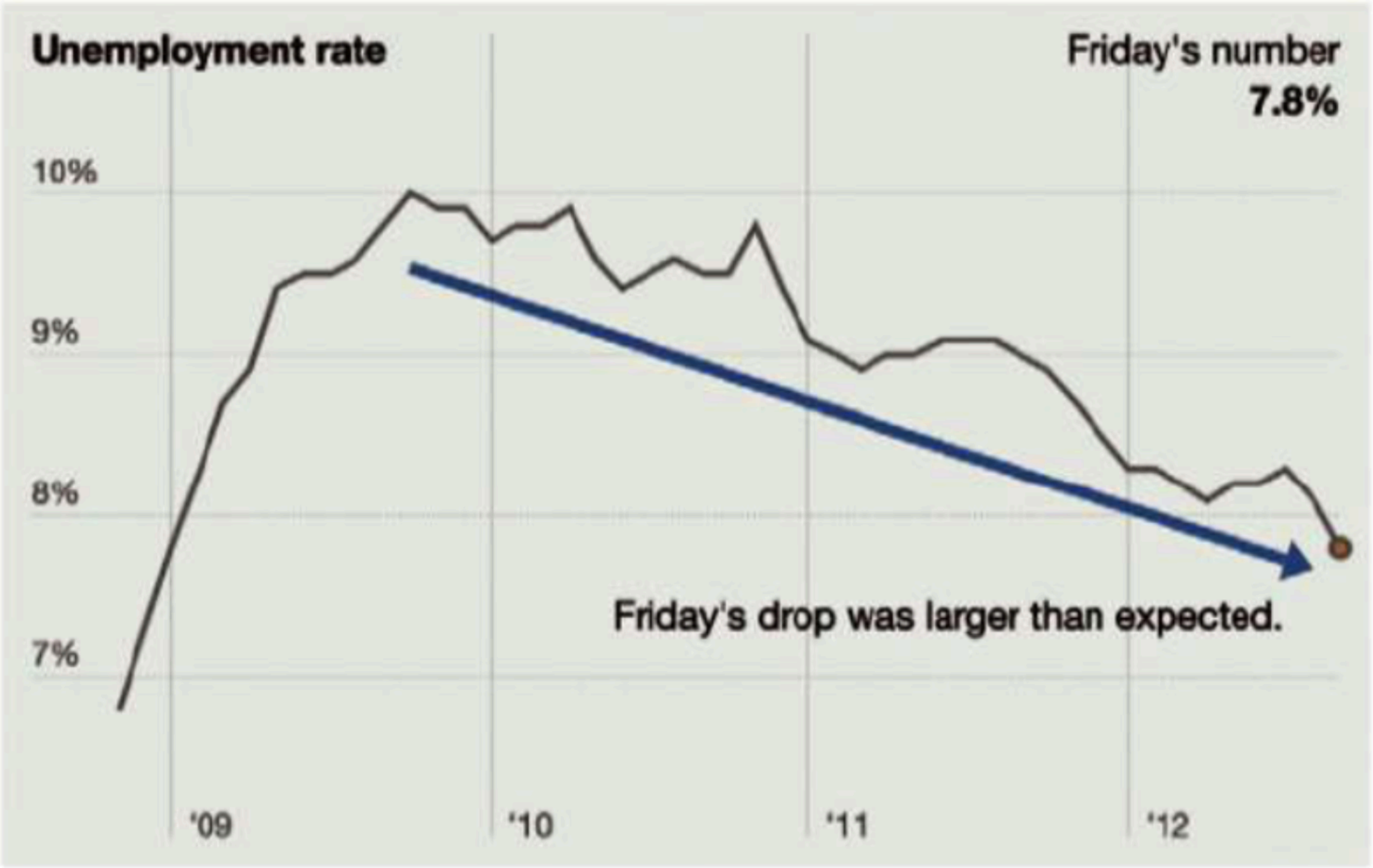
Won multiple awards!



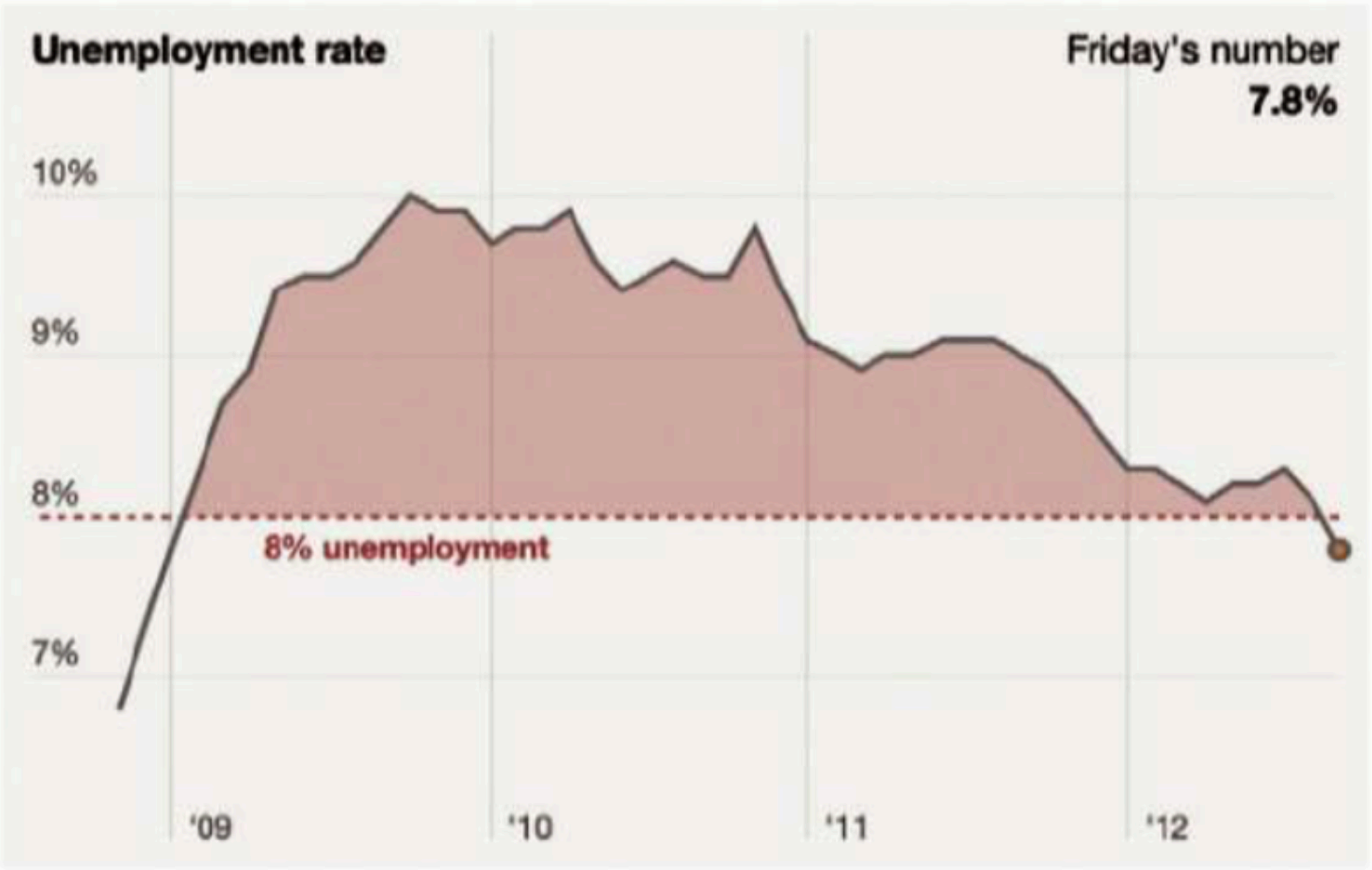
Conventions? Affect? Art?

Source: Reuters, SCMP

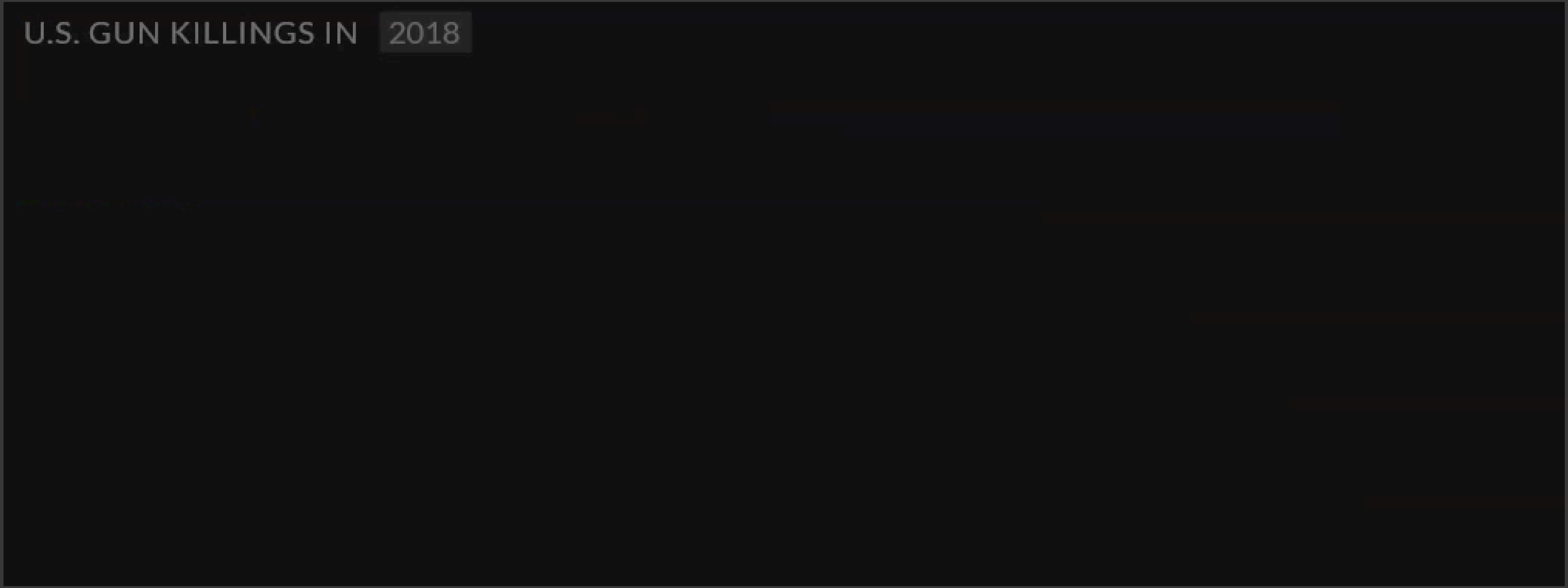
The rate has fallen more than 2 points since its recent peak.



The rate was above 8 percent for 43 months.



Source: Lan et al., VIS 2023



Source: Periscopic, 2018

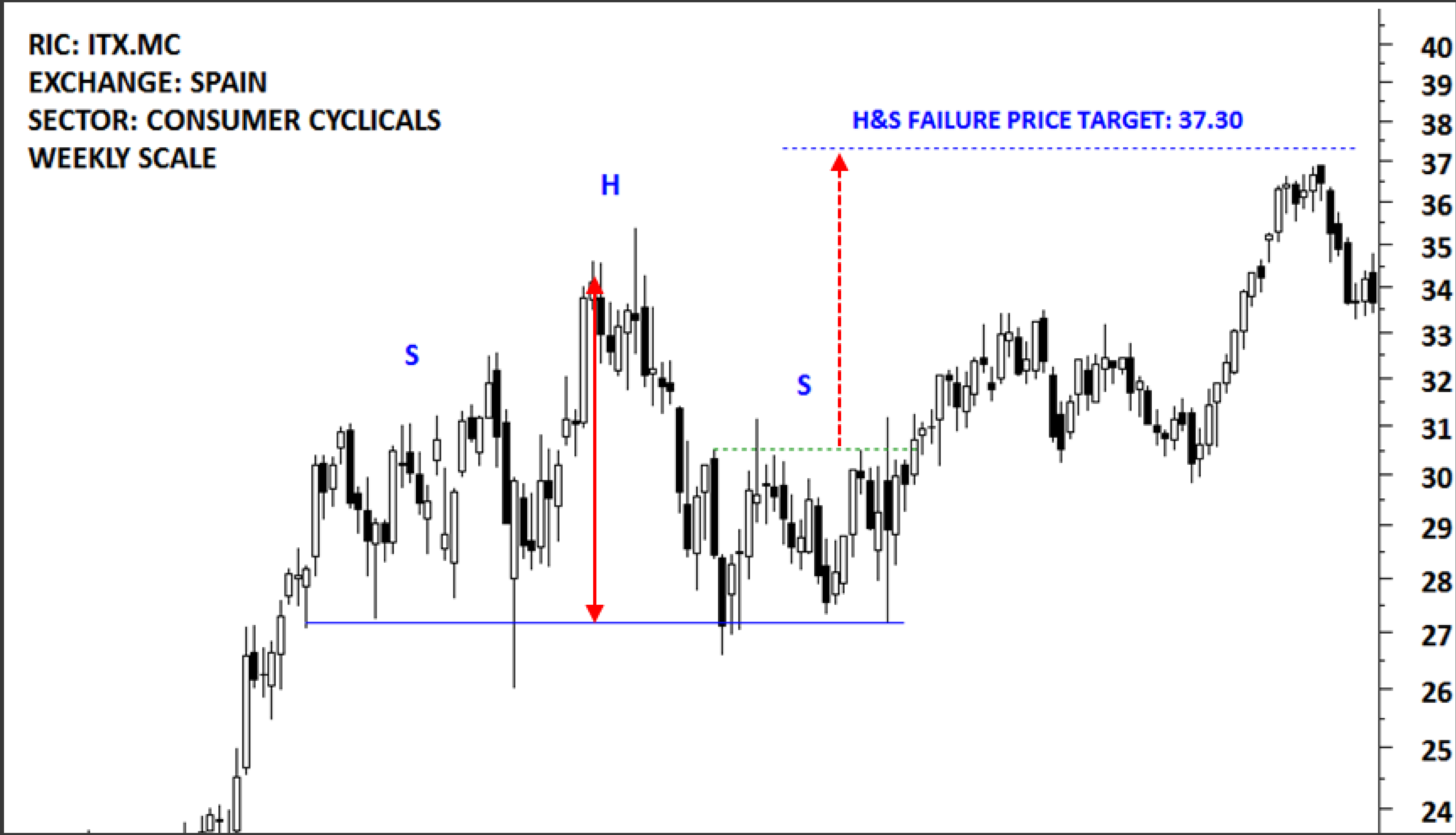
Telling Stories Too Well

Pareidolia

Source

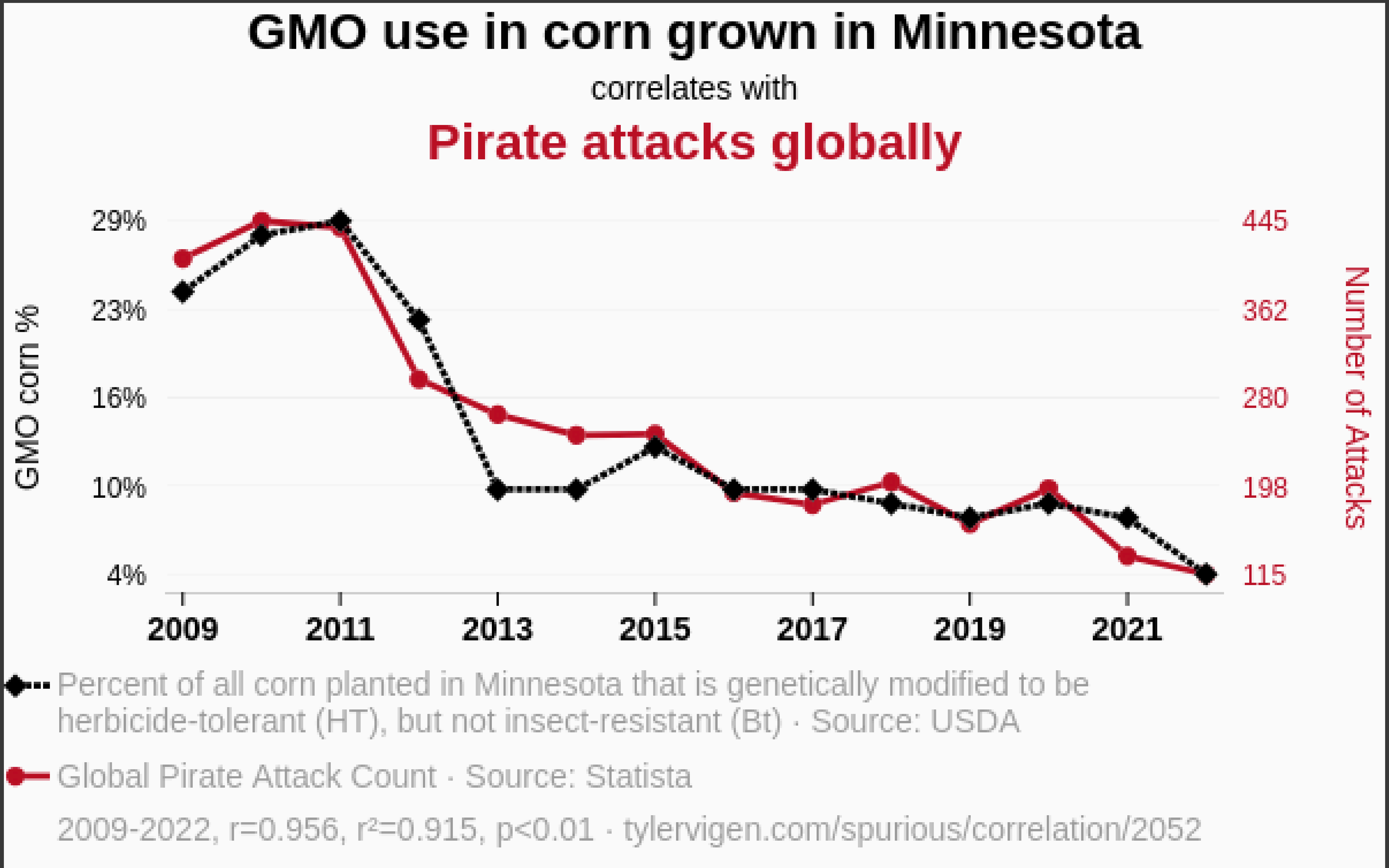
“Technical Analysis”

Source: TrendSpider



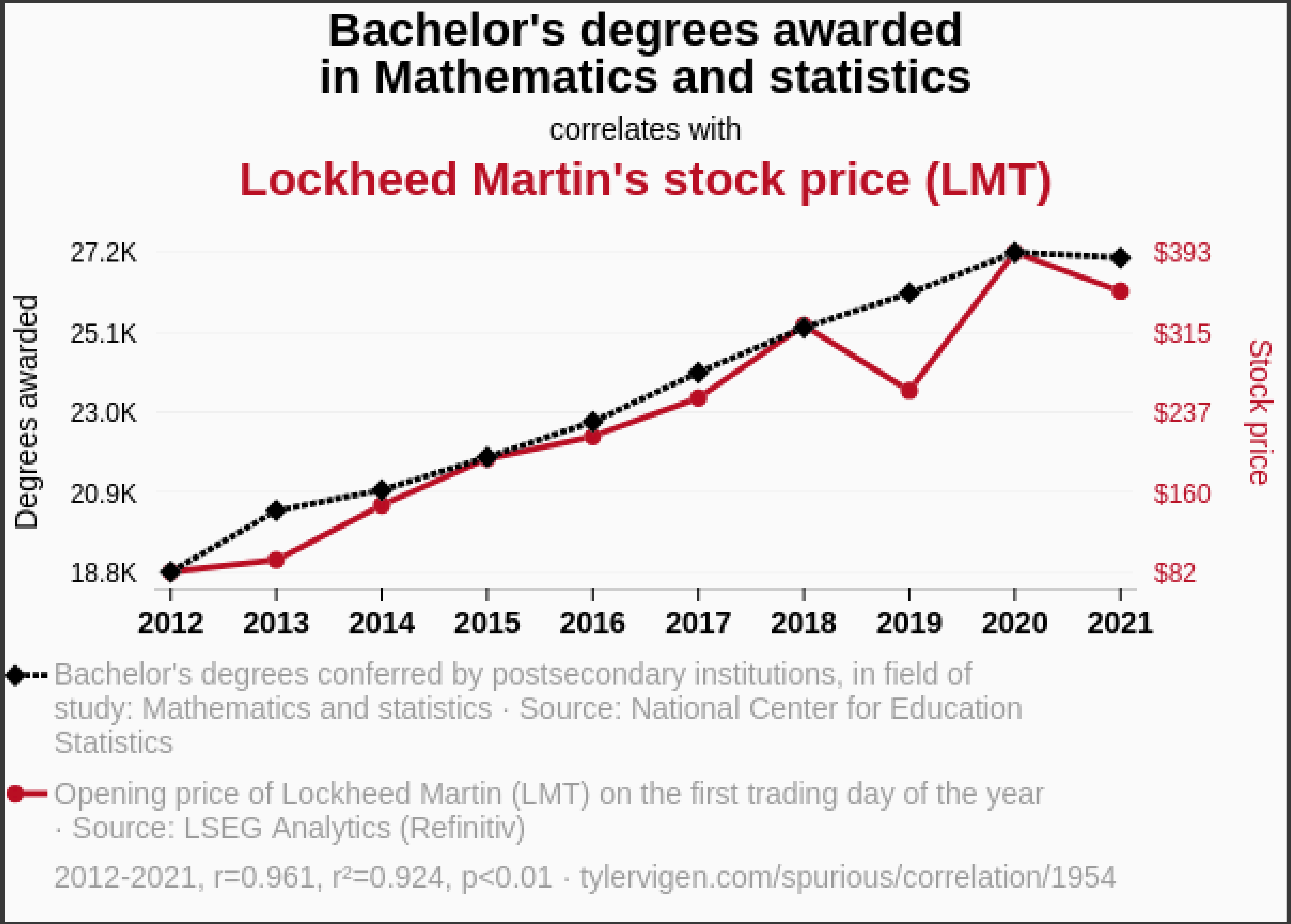
"Technical Analysis"

Source: TechCharts



Spurious Correlations

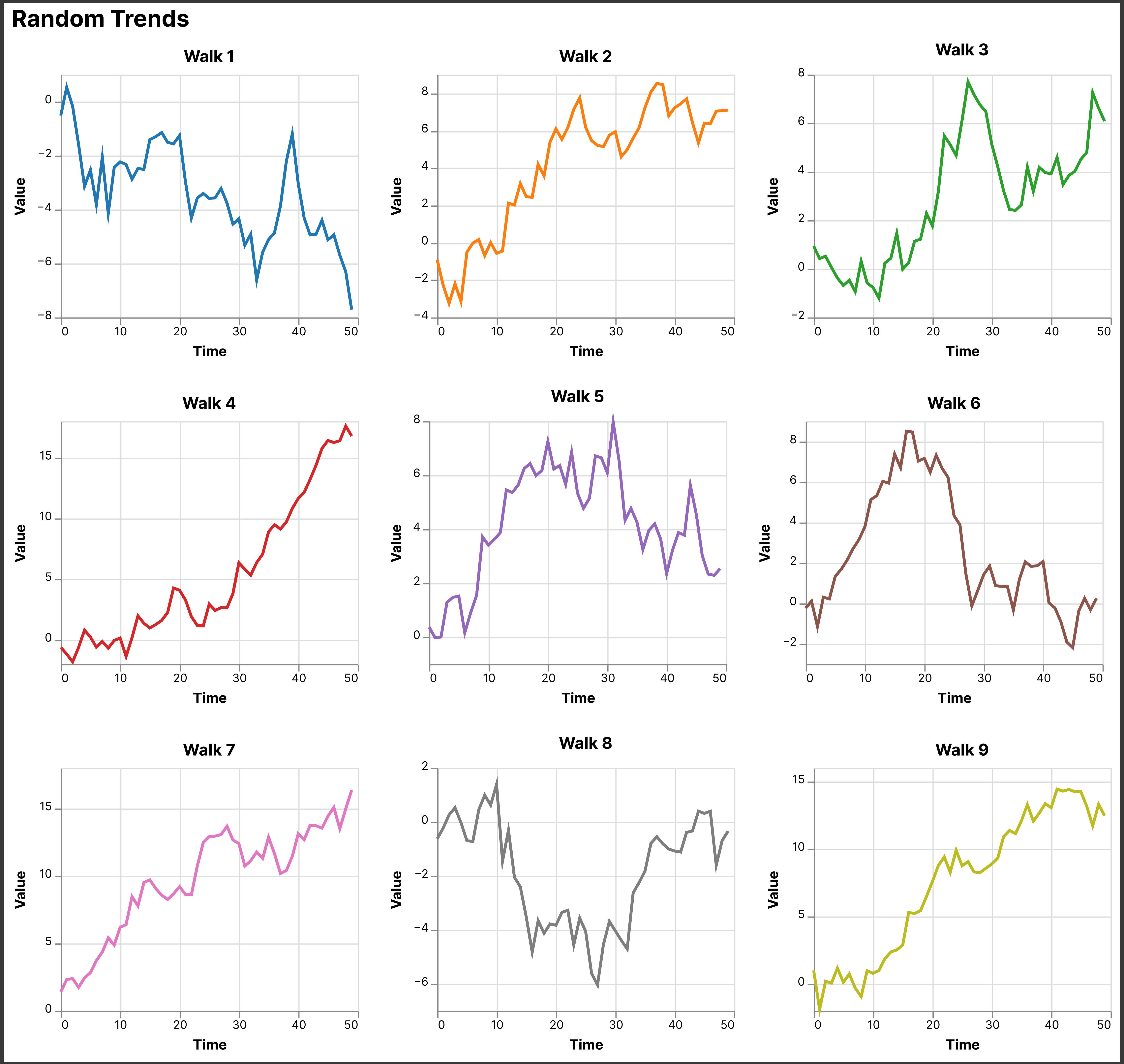
Source: Spurious Correlations



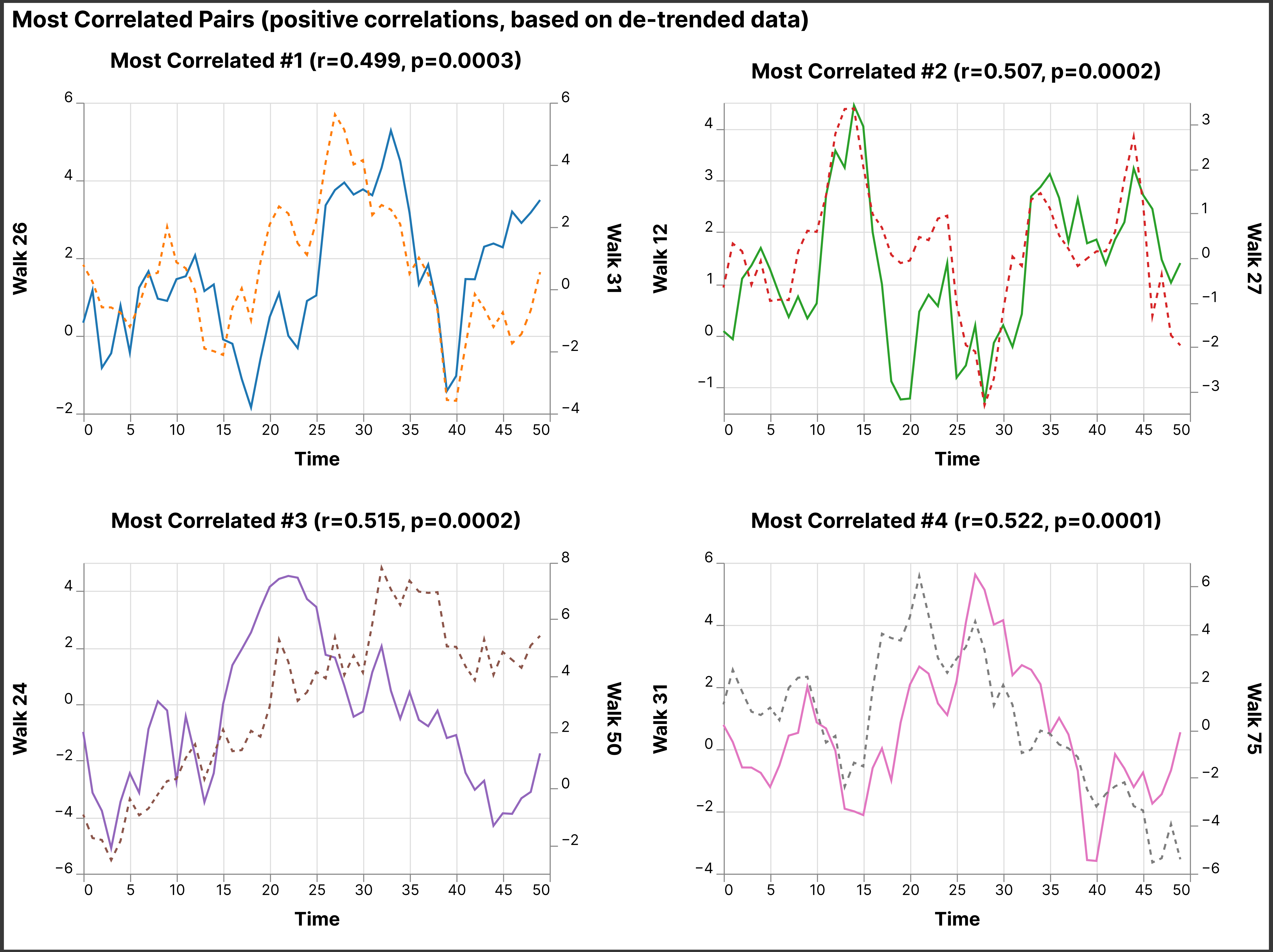
Spurious Correlations

Source: Spurious Correlations

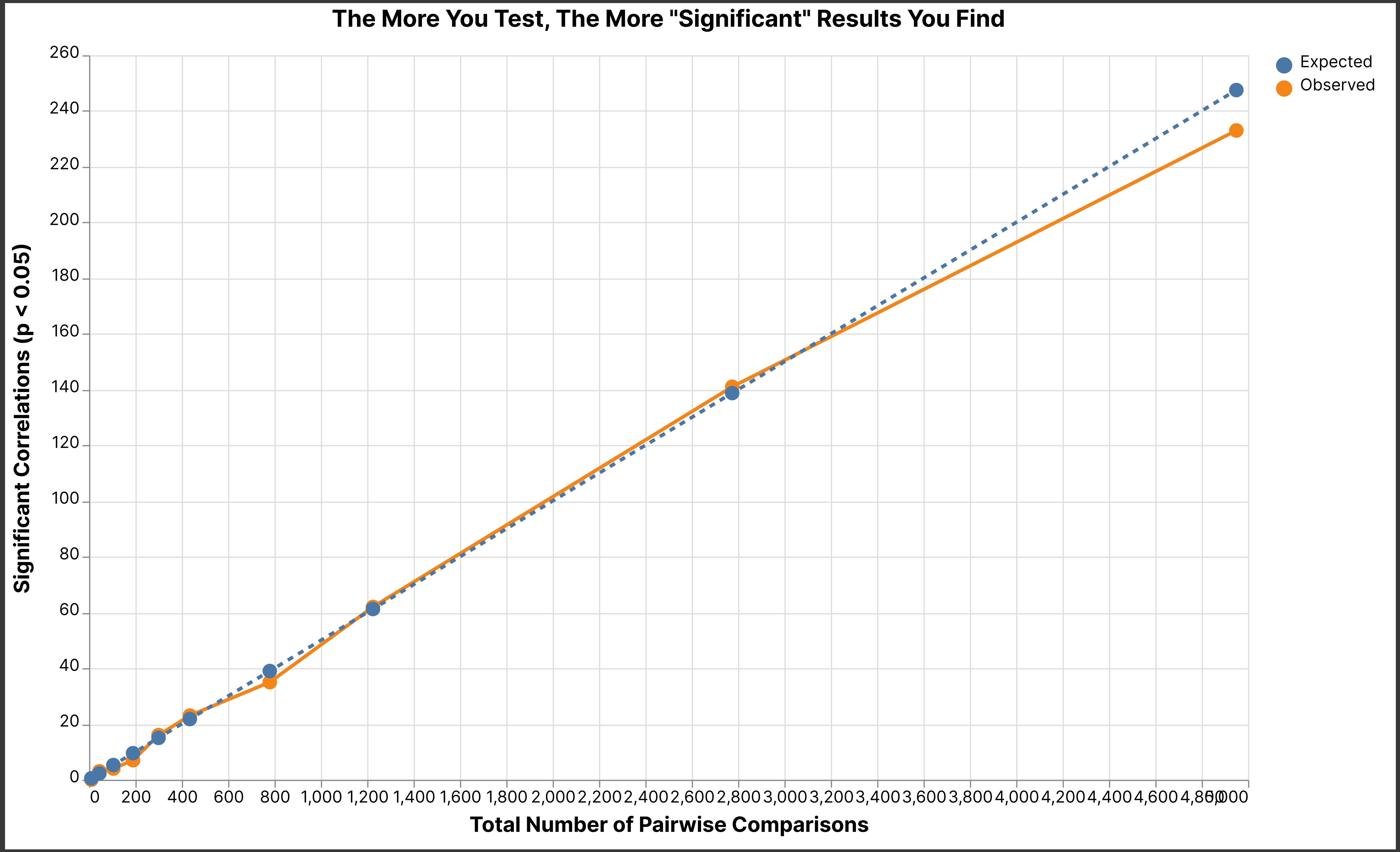
Multiple Comparisons



Multiple Comparisons



Multiple Comparisons



If $p = 0.05$, then 5% of all comparisons I do will be significant, whether or not they make sense!

Summary

- The basics: Data Types, Marks, Encodings
- We visualize data to take advantage of our *perception* and to make *cognition* easier
- Choosing the right encodings can be tricky and depends on *task*, context, and even the shape of the data itself
- Vis can be *affective*/emotional, and that doesn't make it fake
- But vis is so convincing that you can make yourself believe things that aren't true!