

Time for visualization!

① Visit <https://bit.ly/ds-dojo-2024>, Click **Data visualization**

② Visit **Feltron Report and **Dear Data****

Wednesday

- [Data analysis \(pandas\) and APIs](#)
- [Scraping and more pandas](#)

Thursday

- [Data Visualization, PDFs and AI](#)

Friday

Troubleshooting and project work time

Monday

Project presentations

About the instructor

Slides

- Visu

Data visualization

- **Sites to review:** [Feltron Report](#) and [Dear Data](#)
- [Designing Viz](#), data viz tips
- [Datawrapper](#), a platform for data viz
- [Flourish](#), a platform for data viz

PDFs

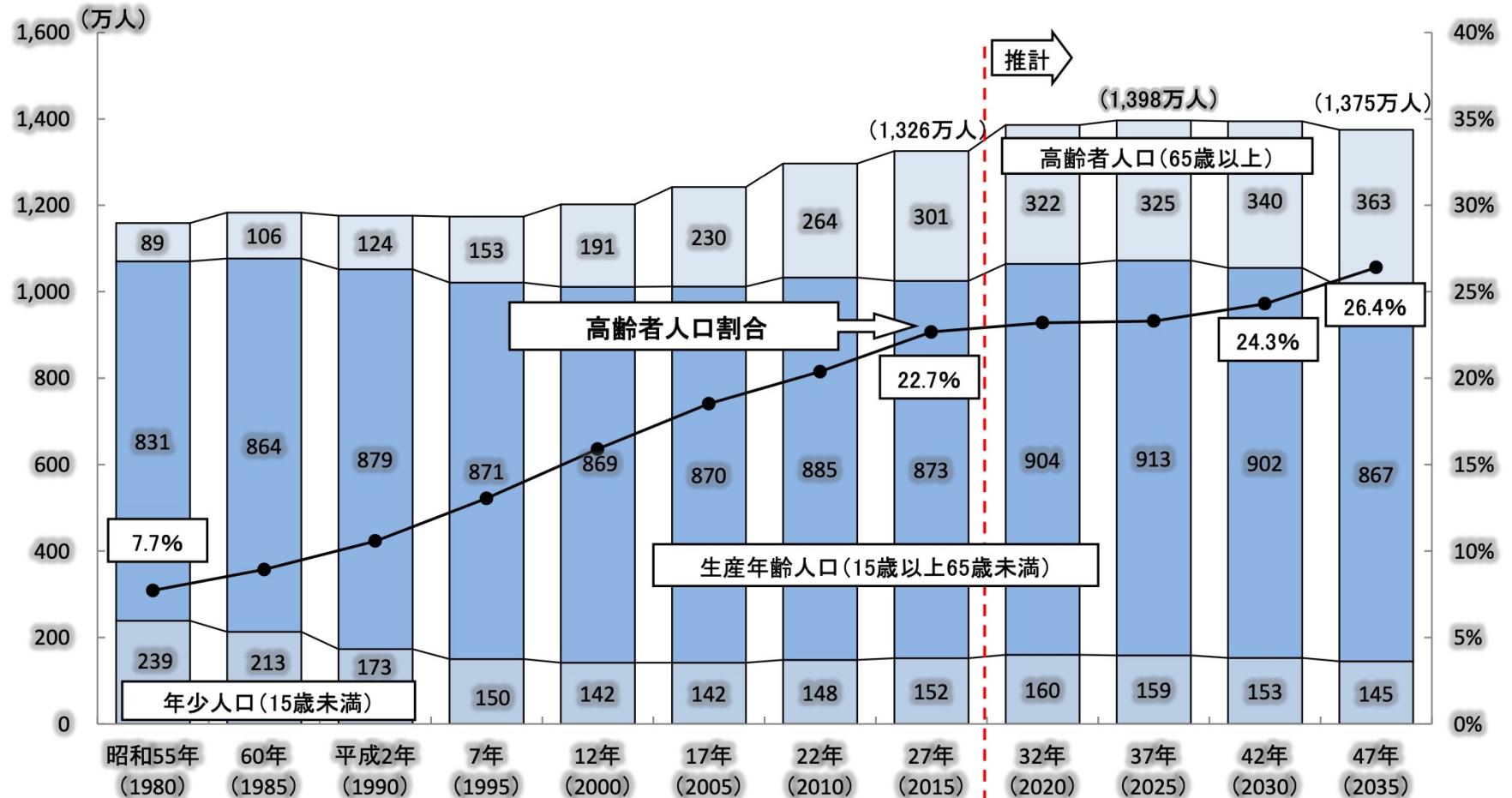
- [Tabula](#), a great tool for extracting data from PDFs
- [Comparison of PDF data extraction tools](#)
- [Comparison of OCR tools](#)

Artificial intelligence

Let's talk about
communication
through design

人口の推移(東京都)

東京都における高齢化率は、総人口がピークを迎える平成37年には23.3%であり、平成42年には24.3%とおよそ4人に1人が高齢者になると推計されます。



(注) ()内は総人口。1万人未満を四捨五入しているため、内訳の合計値と一致しない場合がある。

出典: 総務省「国勢調査」[昭和55年～平成27年]、東京都政策企画局による推計[平成32年～47年]

The Feltron Report

Nicholas Felton

feltron.com



AN ACCOUNTING OF THE YEAR IN

Photos

ANALOG & DIGITAL

Flickr Views:

14,702

WWW.FLICKR.COM/PHOTOS/FELTRON

Percent of Photos Posted to Flickr:

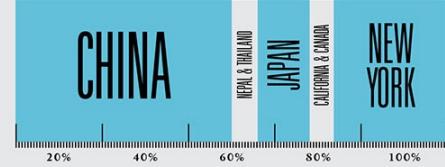
3%

201 PHOTOS

DIGITAL PHOTOS: ANALOG PHOTOS:

6,115 648

Photos by Location:

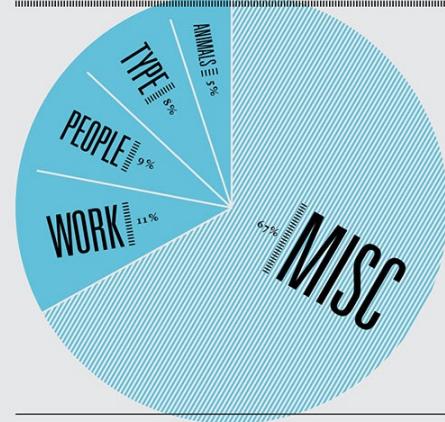


Last Photo:

RYAN, BONNIE & SARA

PATRIOT SALOON, 11:58 PM, DECEMBER 31ST

Photos by Subject:



Elemental, 2014,
PhotoViz, Reporter,
Skillshare, BikeCycle,
TypeCon, 2013, B5B,
Eyeo, 2012, Facebook,
2011, 2010, 2009,
Daytum, 2008,
Editorial, 2007, 2006,
2005.

DAYTUM

https://daytum.com

USER NAME: PASSWORD:

Sign in (REMEMBER ME ON THIS COMPUTER)

HELLO...

DAYTUM HELPS YOU COLLECT, CATEGORIZE AND COMMUNICATE YOUR EVERYDAY DATA.

START USING DAYTUM

BEGIN COLLECTING AND EXPLORING YOUR DATA
TO REVEAL THE BIGGER PICTURE. [LEARN MORE...](#)

CREATE YOUR ACCOUNT

WHAT'S NEW?

IPHONE APP 

DEC 31, 2010: THE DAYTUM IPHONE APP HAS
BEEN RELEASED AND IS AVAILABLE FOR

WHO USES DAYTUM?

 [RAEO](#) USES DAYTUM TO TRACK HIS
274 MUSIC PURCHASES

 [SPOUTDOORS](#): (2 MINUTES AGO)
ADDED 0.5 TR OUTSIDE CLIMBING TO "WORKOUT"

 [SPOUTDOORS](#): (3 MINUTES AGO)
ADDED 0.33 TRAD CLIMBING TO "WORKOUT"

 [SPOUTDOORS](#): (4 MINUTES AGO)
ADDED 0.05 LIGHT LIFTING TO "WORKOUT"

 [SPOUTDOORS](#): (4 MINUTES AGO)

PHOTOGRAPHS PER COUNTRY

1158



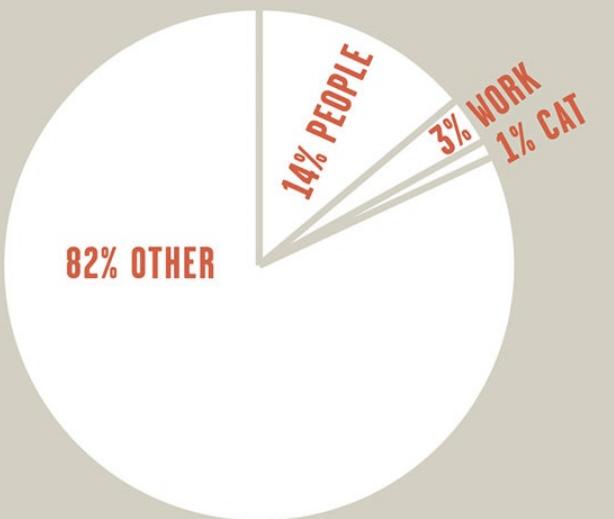
DIGITAL PHOTOS

3,754

ANALOG PHOTOS

0

PHOTOGRAPHIC SUBJECTS



1980-1989
MILL VALLEY & LARKSPUR

THE EIGHTIES

706 ENTRIES

CALENDAR



JUL 14, 1980
51 YEARS
AND 10 DAYS



MILES SAILED
ON THE
VIRAGO FROM
VICTORIA, BC
TO SAUSALITO

870
AUGUST 9-20, 1988

1985
MOST VISITS
TO A WORK
LOCATION

80
THE STATE
OFFICE
BUILDING, SF

NOV 3, 1987

AFTER 14 YEARS, 2 MONTHS
AND 3 DAYS OF MARRIAGE

1984-1990
DAYS OF WORK

268
AND 35 DAYS JOB HUNTING

PHYSICAL
ACTIVITIES

TWELVE

CYCLING, DANCING, FISHING,
HIKING, ICE SKATING, KAYAKING,
RAFTING, SAILING, SKIING,
TAI CHI AND WORK

MOVIES
VIEWED

THIRTEEN

6 WITH CHILDREN

1986-1987
DAYS TAI CHI
PRACTICED

11

PARTIES
ATTENDED

TWENTY

2 CHRISTMAS AND 1 CHANUKAH



1970-2010

CAL

4,412 RECORDS

ENTRI

NUMBER OF
CALENDARS

3

YEAR WITH
THE MOST
ENTRIES

1985
269 DAYS

FIRST
CALENDAR

1970
150 DAYS

NUMBER OF
WORDS

17,5
AVERAGE

615 MEALS

EATING

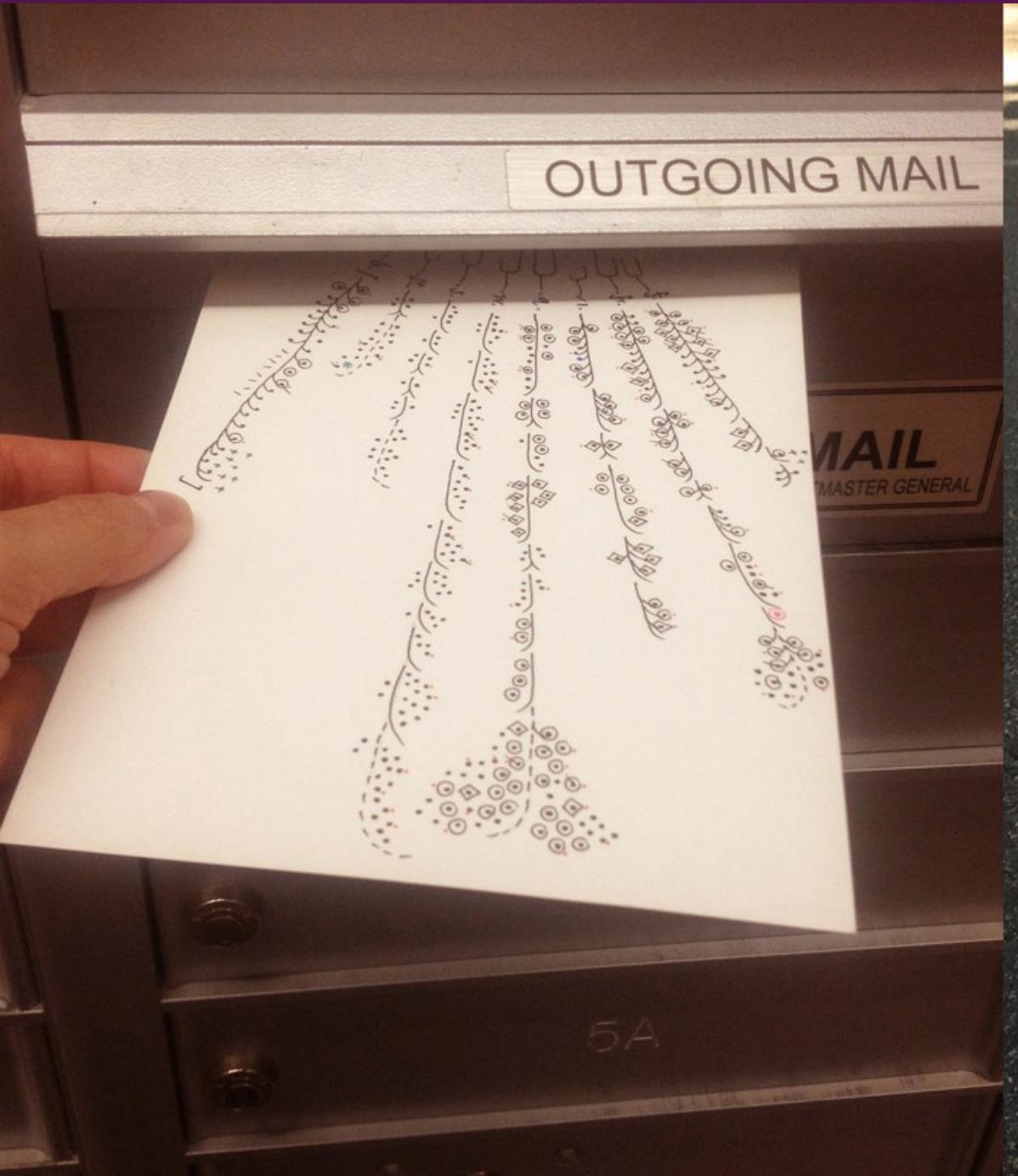
1966

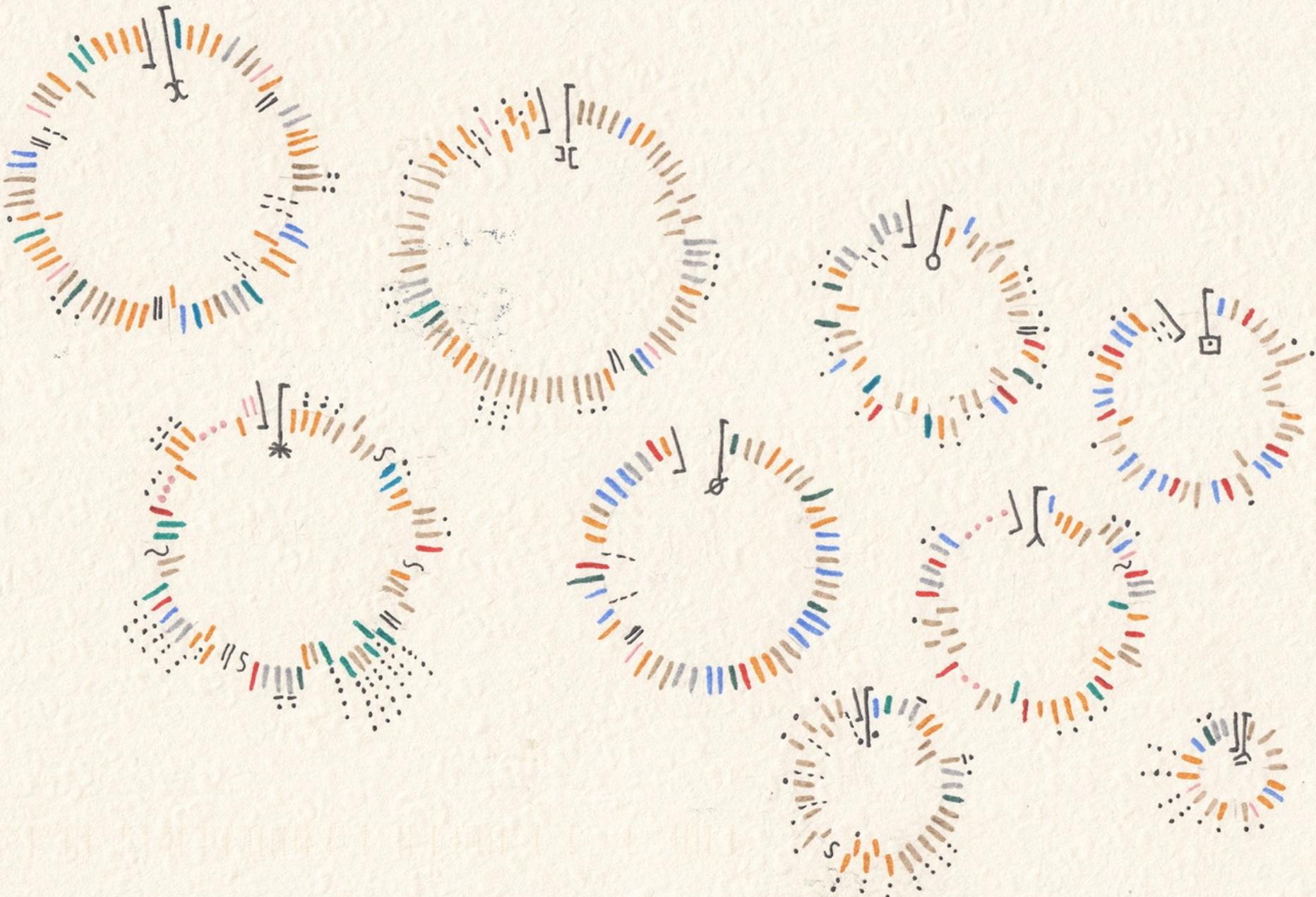
1974

Dear Data

Giorgia Lupi and Stefanie Posavec

dear-data.com





66 DEAR DATA

WEEK 08: PHONE ADDICTION!

HOW TO READ IT:



PLACES / sit.

✗ while walking

* while working

⌚ while waiting
for s.thng or
s.body

∅ in the Bathroom

○ on the couch

□ on the bed

^ other places at
home

%. cafe / restaurants
shops

≈ public
transportation

Every circle represents a PLACE or SITUATION
where I checked my phone, somehow
ordered from left to right according
to how many times I did it in that

Every single LINE is a SINGLE TIME place.
I interacted with my phone, ordered
chronologically per each place.

COLORS: the reason why I picked it:

- text / email
- social media
- other APPS
- check the time
- check the weather
- phone call
- text with somebody
who was in the room
- to charge it
- text / email with you
- take pictures of,
our postcards!
- ... dots = while with
others (at home)
- 1 = used others' phone

ATTRIBUTES:

→ outside =
I picked it
PURPOSELY

→ inside =
Because of an
alert

~ = turned the
phone facing the
table not to see it

--- didn't pick it
because I didn't
want to report

= thought it was
ringing but
wasn't!

NEW YORK CITY
FROM:
GIORGIA LUPI
US NOW



11249 BROOKLYN
- NY - USA

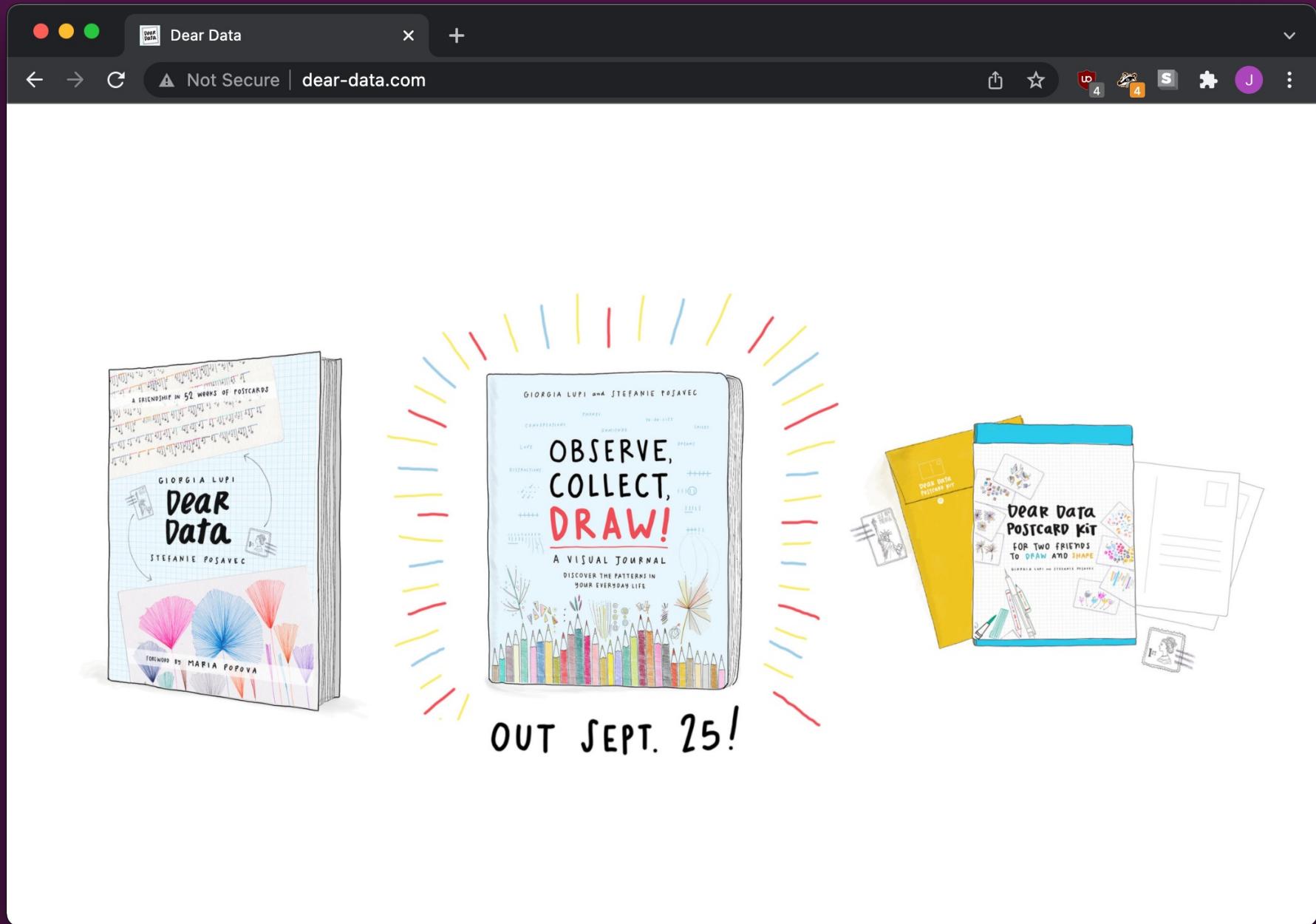
SEND TO:

STEFANIE POSAVEC

LONDON

- UK -

ENGLAND



What's the point?

In this case, maybe **art vs information?**

Why or why not?

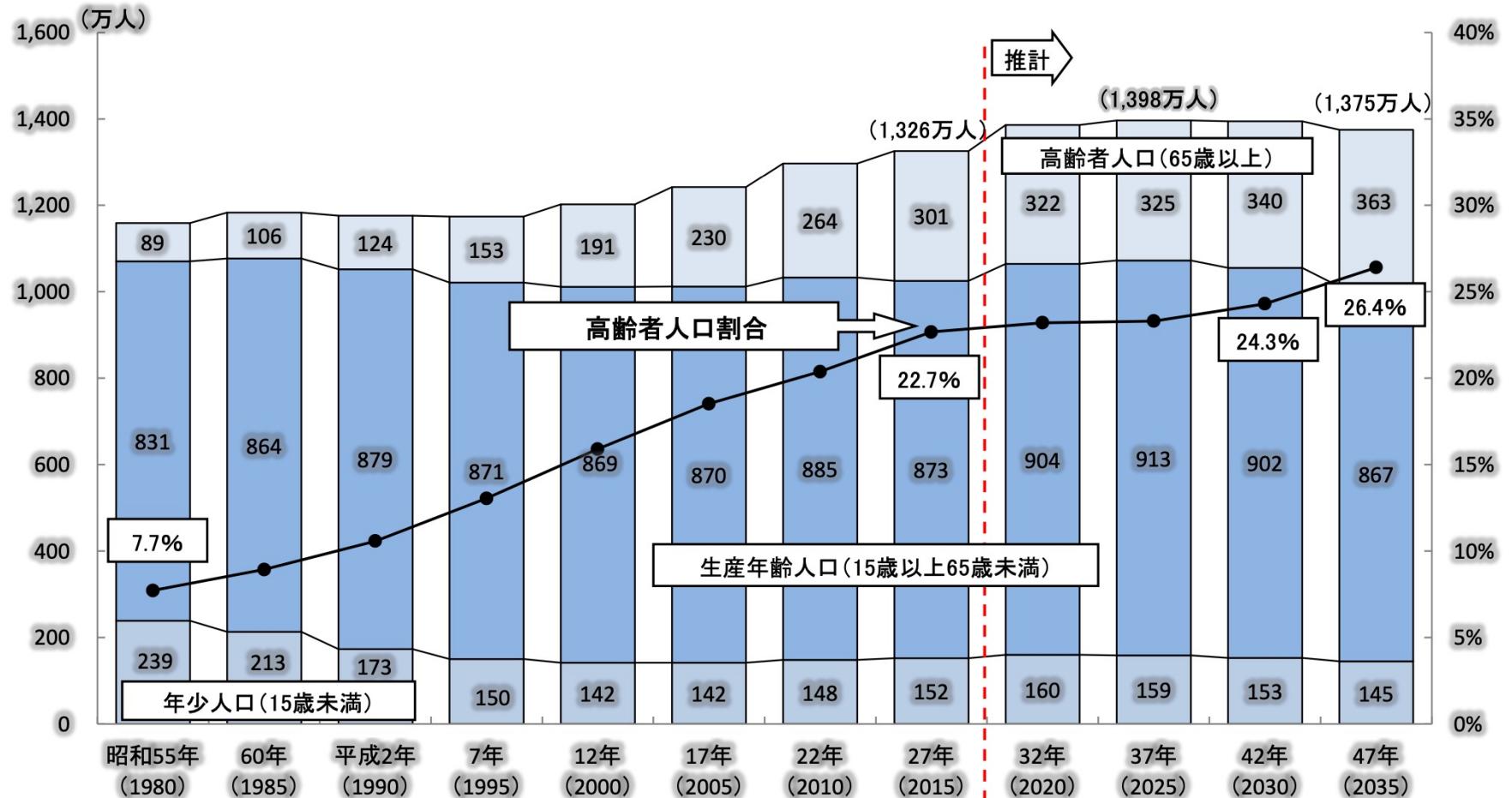
Everything should be on purpose.

Every drop of ink should have a reason.

Every color should have a story.

人口の推移(東京都)

東京都における高齢化率は、総人口がピークを迎える平成37年には23.3%であり、平成42年には24.3%とおよそ4人に1人が高齢者になると推計されます。



(注) ()内は総人口。1万人未満を四捨五入しているため、内訳の合計値と一致しない場合がある。

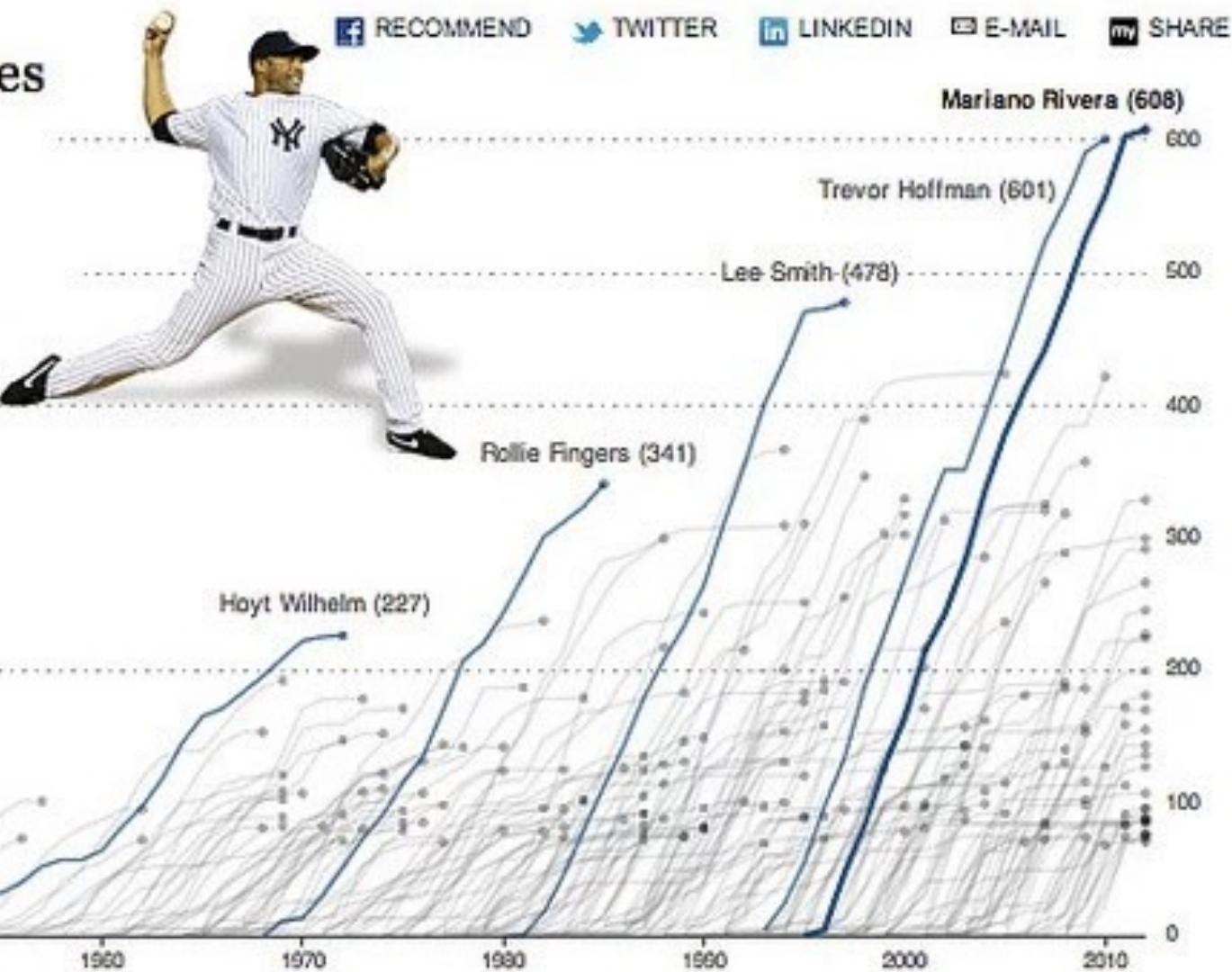
出典：総務省「国勢調査」[昭和55年～平成27年]、東京都政策企画局による推計[平成32年～47年]

DATA IS THE ENEMY

**...THE MORE DATA YOU SHOW,
THE LESS INFORMATION THE USER CAN SEE**

How Mariano Rivera Compares to Baseball's Best Closers

Mariano Rivera tore his right A.C.L. while shagging fly balls on Thursday, possibly ending his career. Considered the best closer in baseball history, Rivera has more saves than any other pitcher. Below, the cumulative saves of the pitchers with 100 or more.



The closers who broke new hundred-save milestones:

Firpo Marberry (101)
The first reliever to get to 100 cumulative saves, done at a time before relief pitchers were commonplace. (Marberry also started 180 games.)

Hoyt Wilhelm (227)
In addition to being the first pitcher to break the 200 save mark, Wilhelm pitched a no-hitter against the Yankees in 1958.

Rollie Fingers (341)
Known for his handlebar moustache, Fingers was the second relief pitcher inducted into Baseball's Hall of Fame.

Lee Smith (478)
From 1983 to 1995, Smith averaged 35 saves a season, saving no fewer than 25 in any season.

Trevor Hoffman (601)
Hoffman was the first to break the 500 and 600 save marks, despite a 1994 shoulder injury that forced him to change his pitching style.

FOCUS

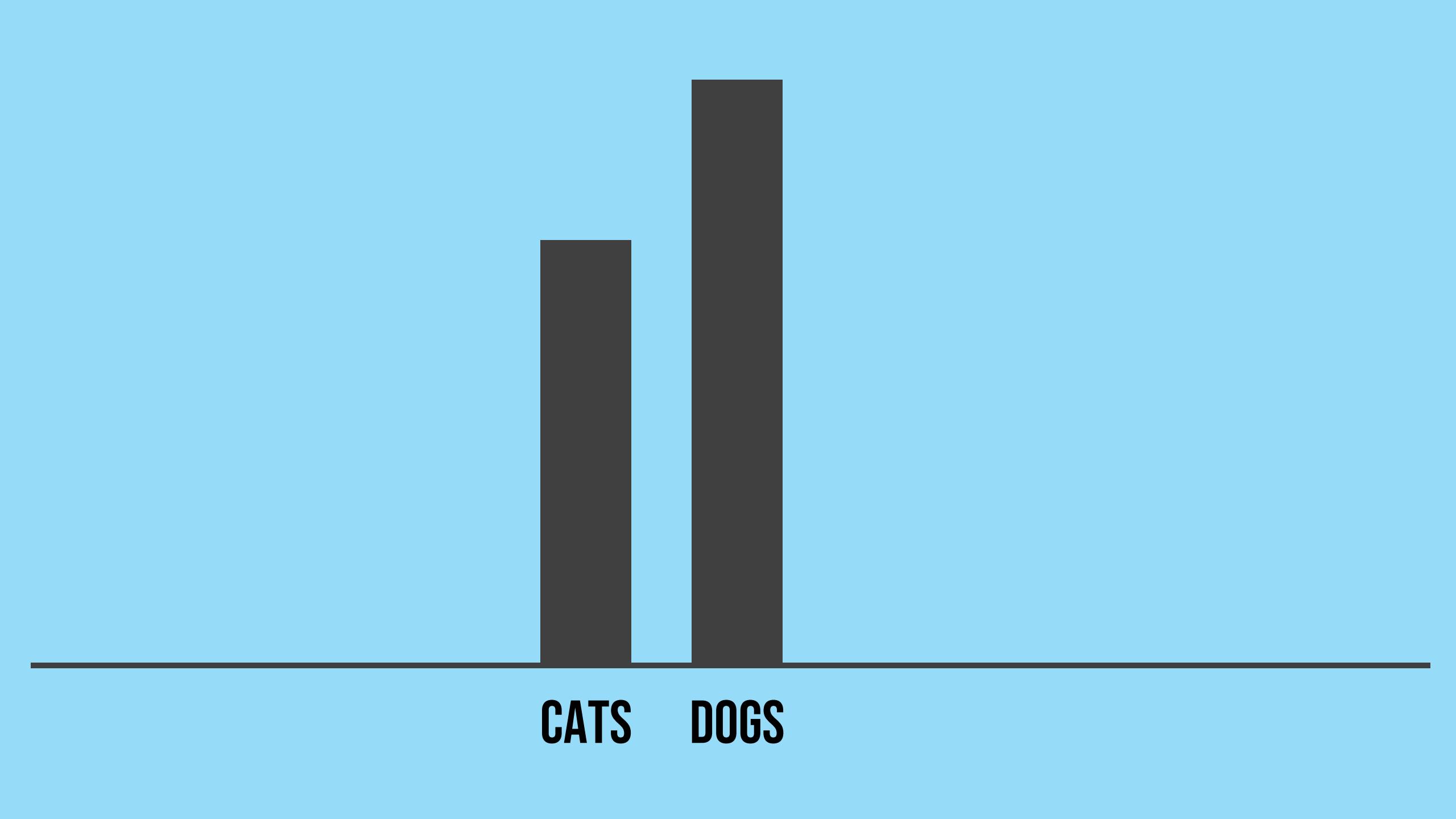
what should the reader pay attention to?

SMALL DATA

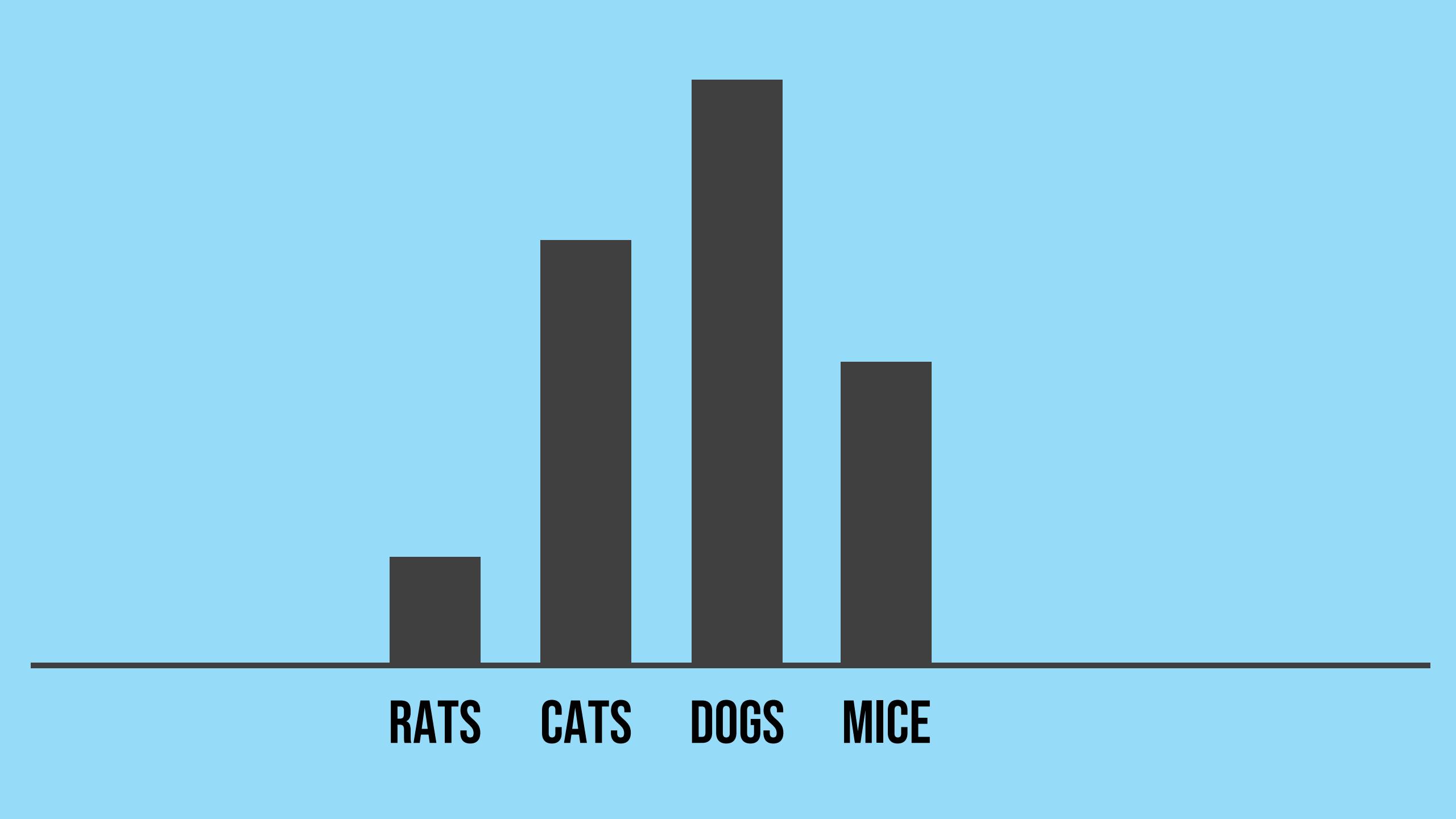
VS

LARGE DATA

...DON'T LET YOUR USER GET DISTRACTED



CATS DOGS



A bar chart comparing the counts of four animal species. The x-axis categories are RATS, CATS, DOGS, and MICE. The bars are dark gray. DOGS have the highest count, followed by CATS, then MICE, and finally RATS with the lowest count.

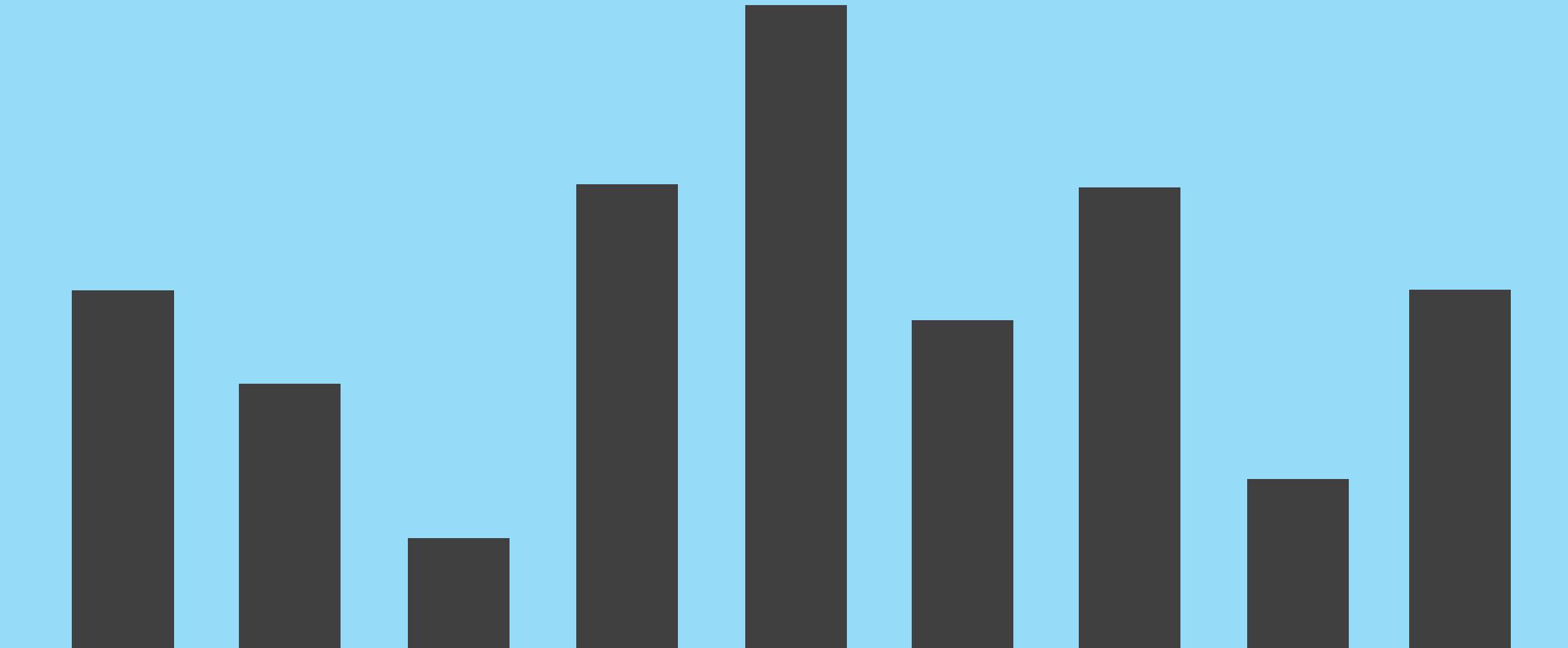
Animal	Count
RATS	1
CATS	3
DOGS	6
MICE	2

RATS

CATS

DOGS

MICE



BIRDS

OWLS

RATS

CATS

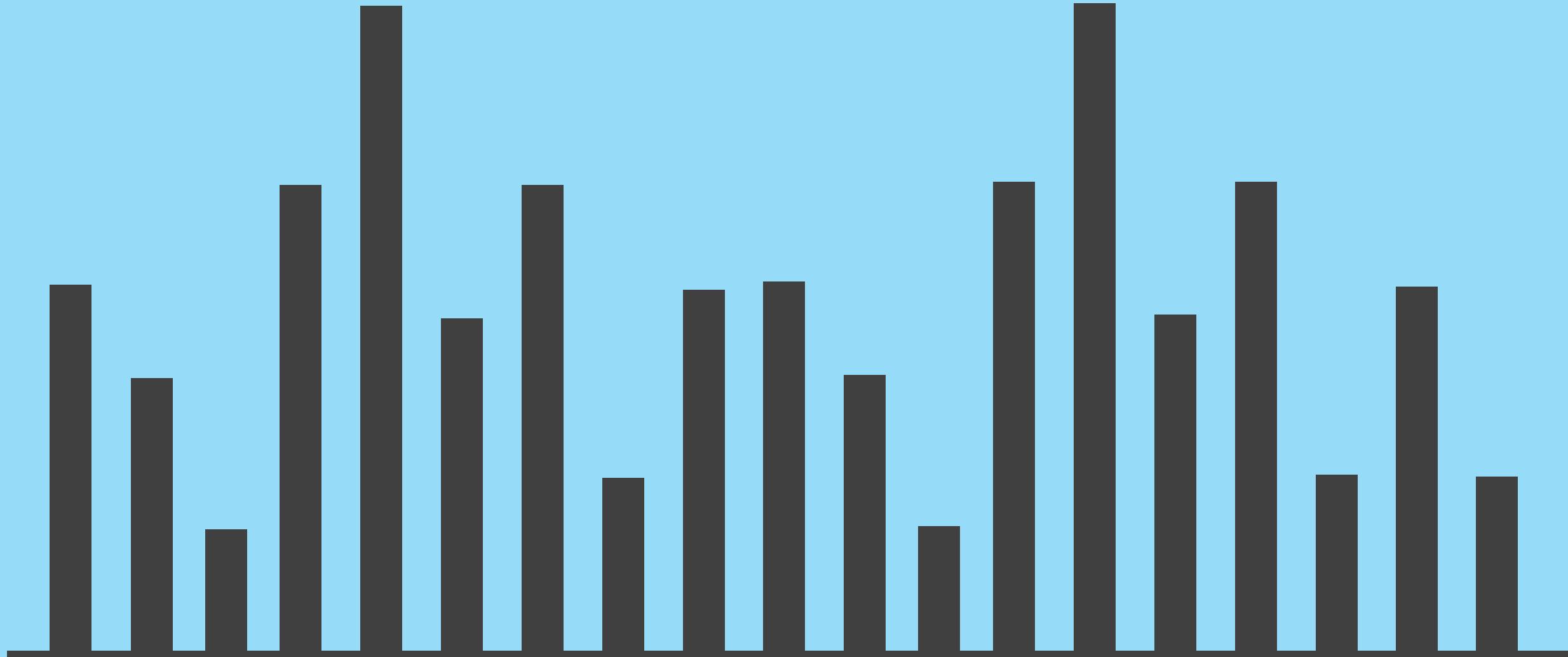
DOGS

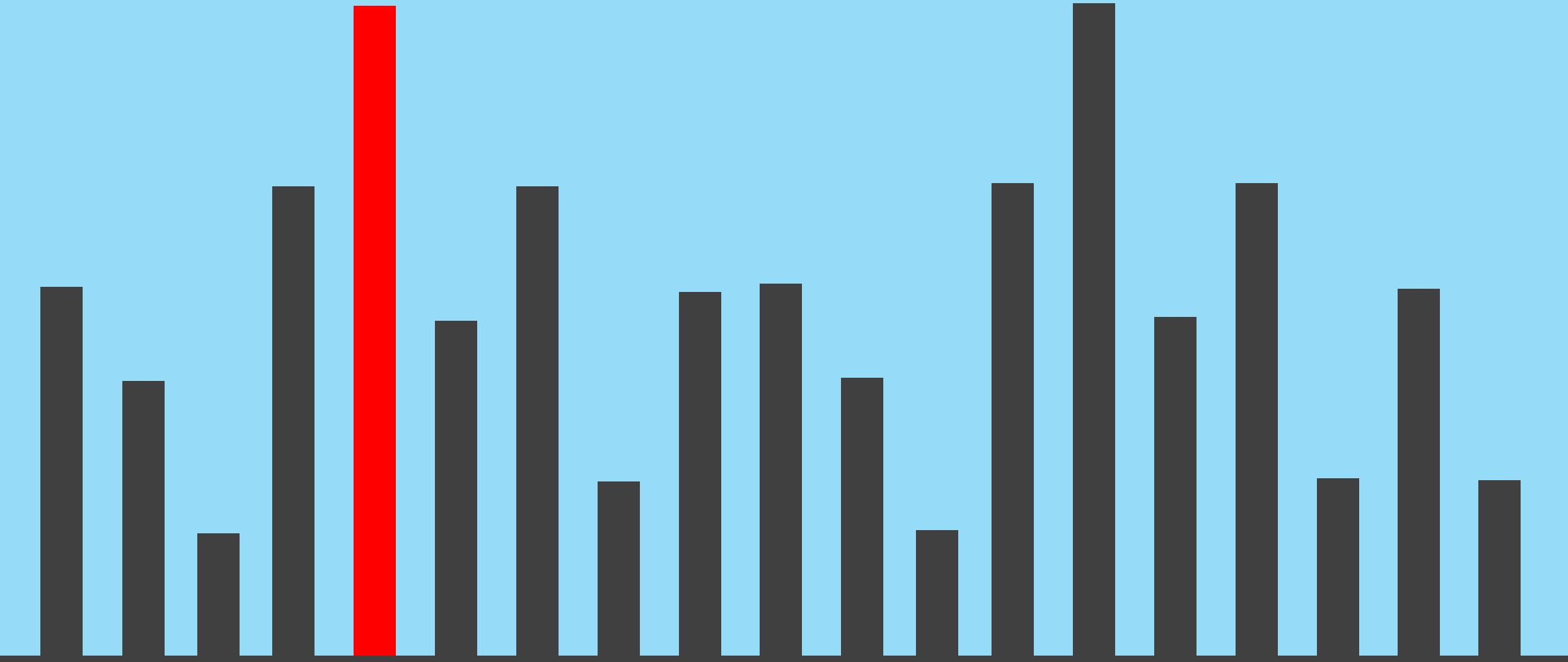
MICE

TIGERS

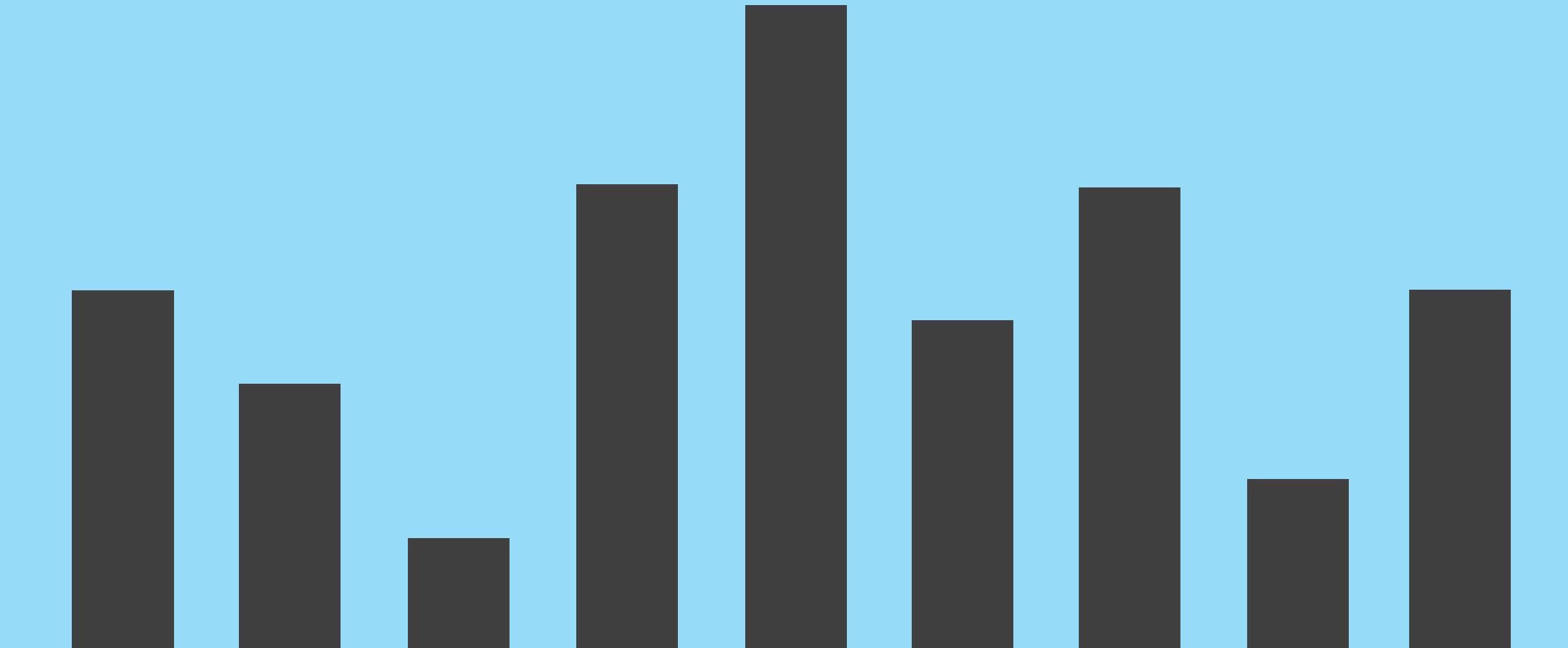
LIONS

BUGS





DIRECT FOCUS TO WHAT'S IMPORTANT



BIRDS

OWLS

RATS

CATS

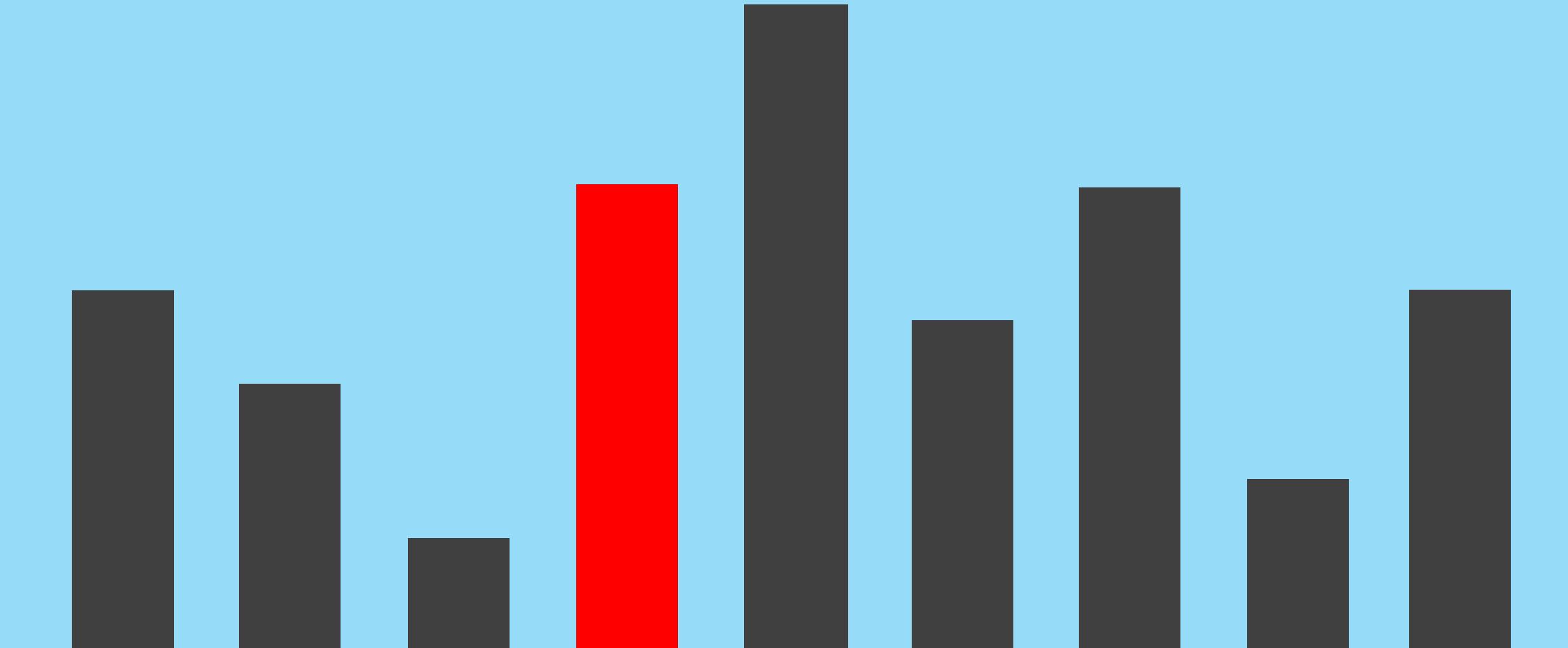
DOGS

MICE

TIGERS

LIONS

BUGS



BIRDS

OWLS

RATS

CATS

DOGS

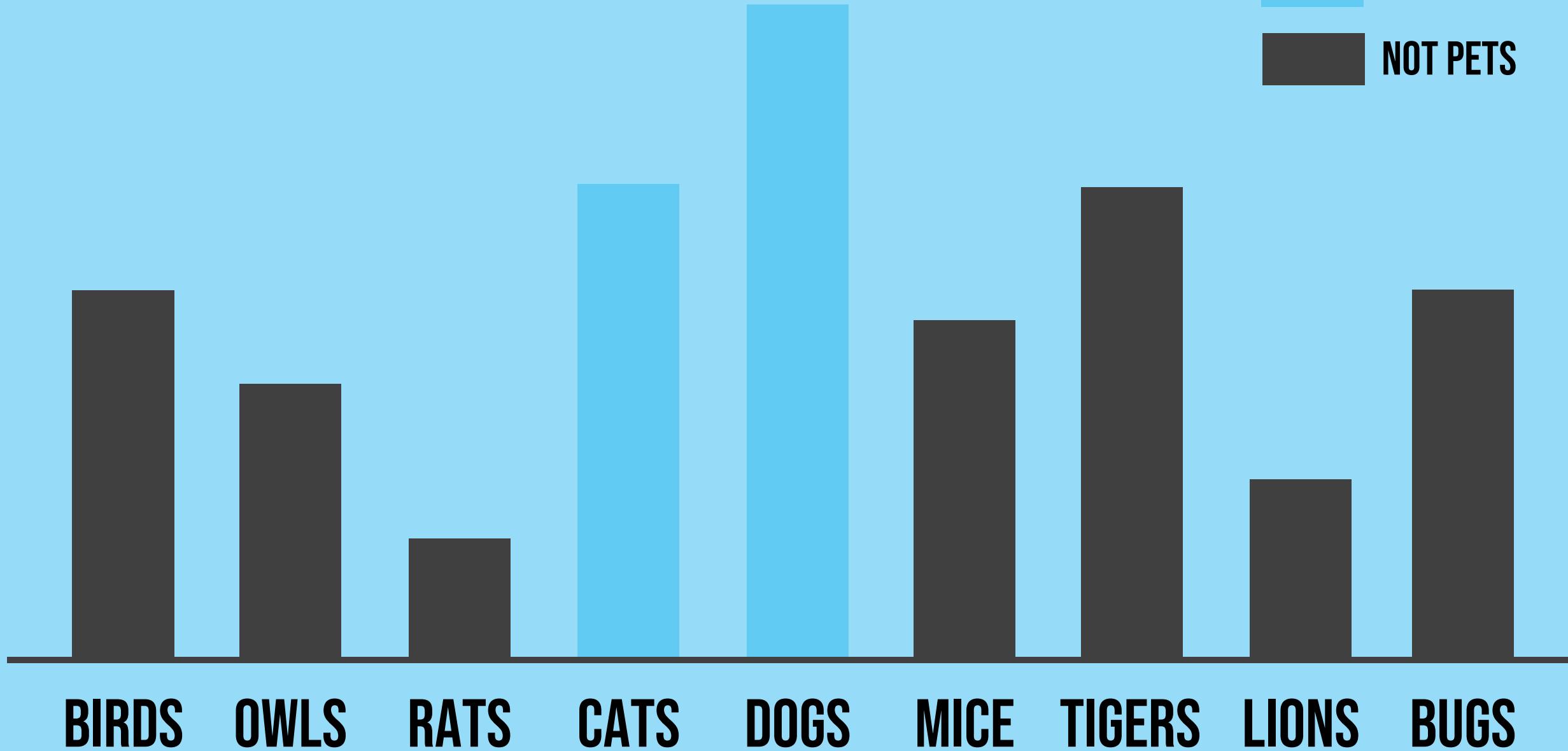
MICE

TIGERS

LIONS

BUGS

PETS
NOT PETS



MORE NUMBERS HERE, TOO

TEN

TWENTY

THIRTY

FORTY

FIFTY

SIXTY

SEVENTY



MORE NUMBERS HERE, TOO

TEN

TWENTY

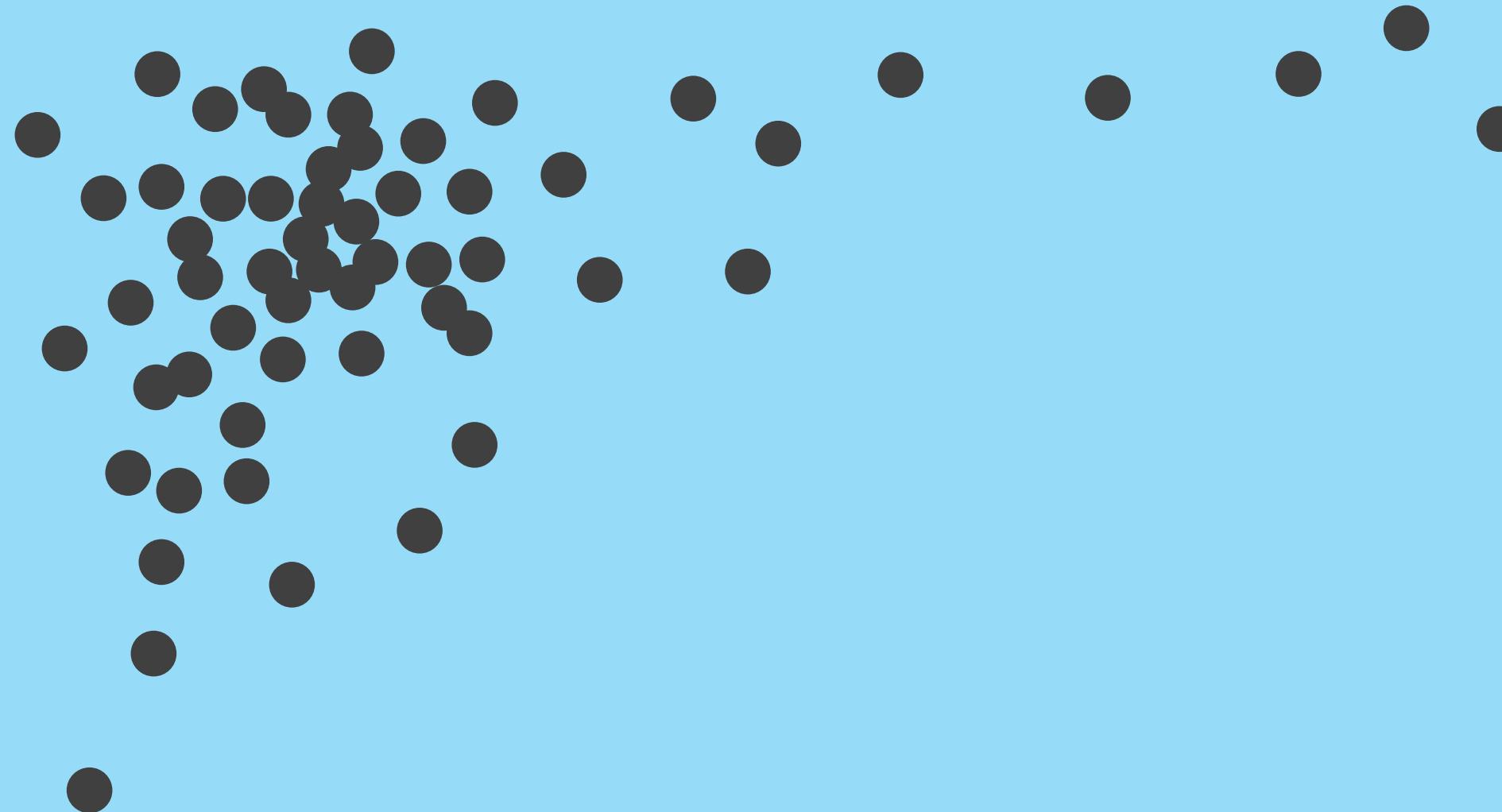
THIRTY

FORTY

FIFTY

SIXTY

SEVENTY



MORE NUMBERS HERE, TOO

TEN

TWENTY

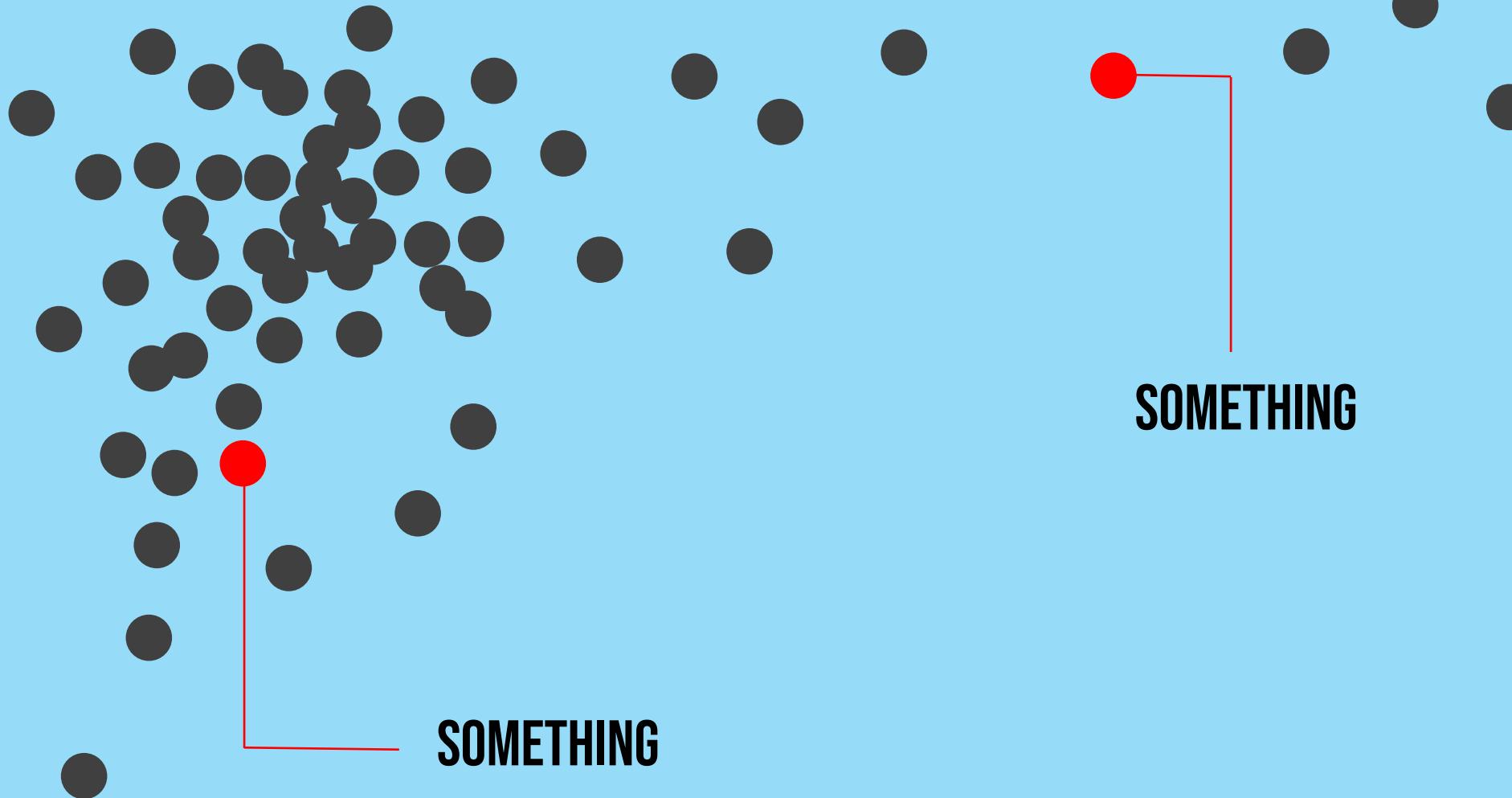
THIRTY

FORTY

FIFTY

SIXTY

SEVENTY



MORE NUMBERS HERE, TOO

TEN

TWENTY

THIRTY

FORTY

FIFTY

SIXTY

SEVENTY

**"BACKGROUND"
DATA WITH GREY**



MORE NUMBERS HERE, TOO

TEN

TWENTY

THIRTY

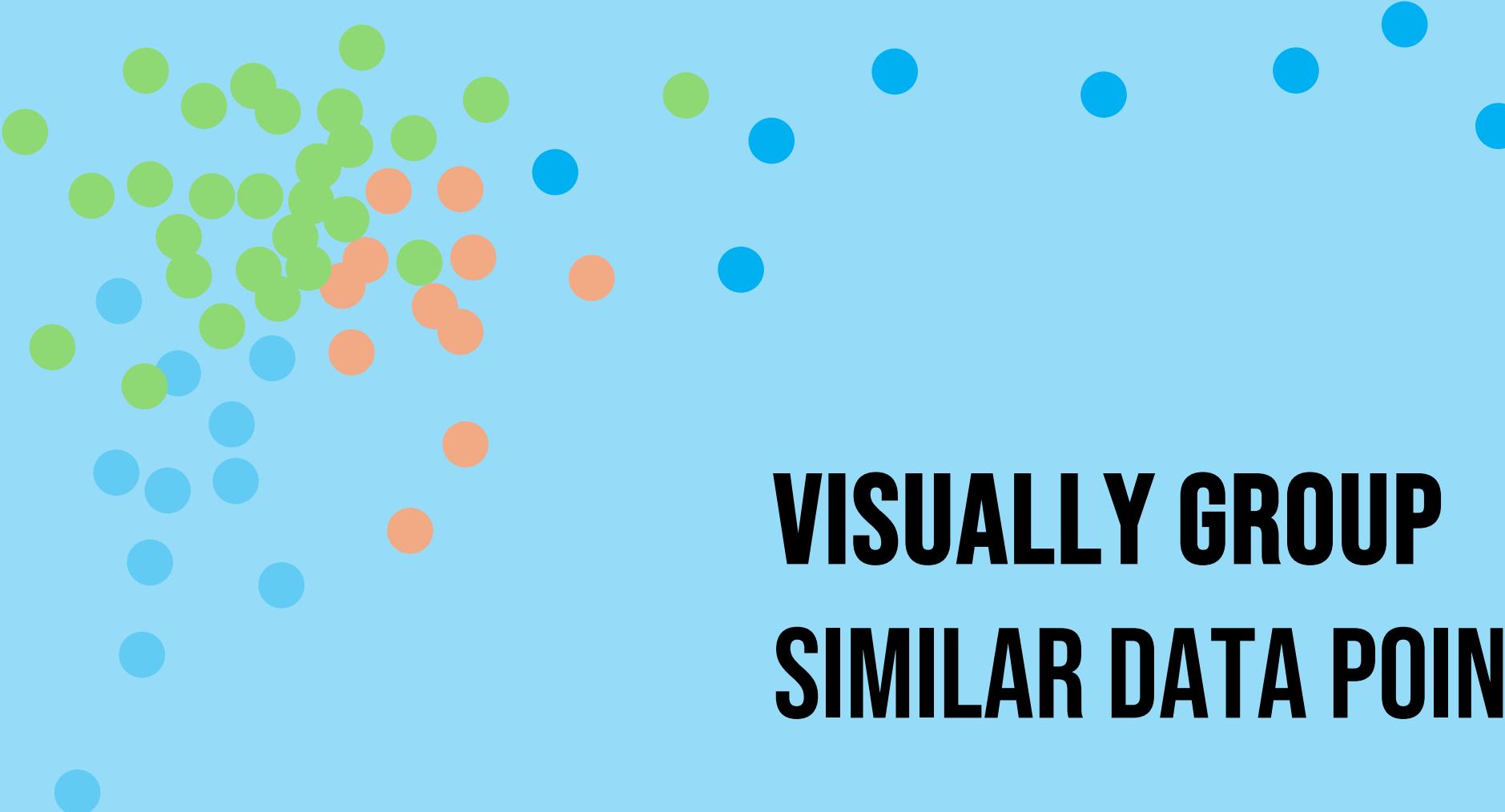
FORTY

FIFTY

SIXTY

SEVENTY

VISUALLY GROUP
SIMILAR DATA POINTS



MORE NUMBERS HERE, TOO

TEN

TWENTY

THIRTY

FORTY

FIFTY

SIXTY

SEVENTY

DON'T DO THIS



MORE NUMBERS HERE, TOO

TEN

TWENTY

THIRTY

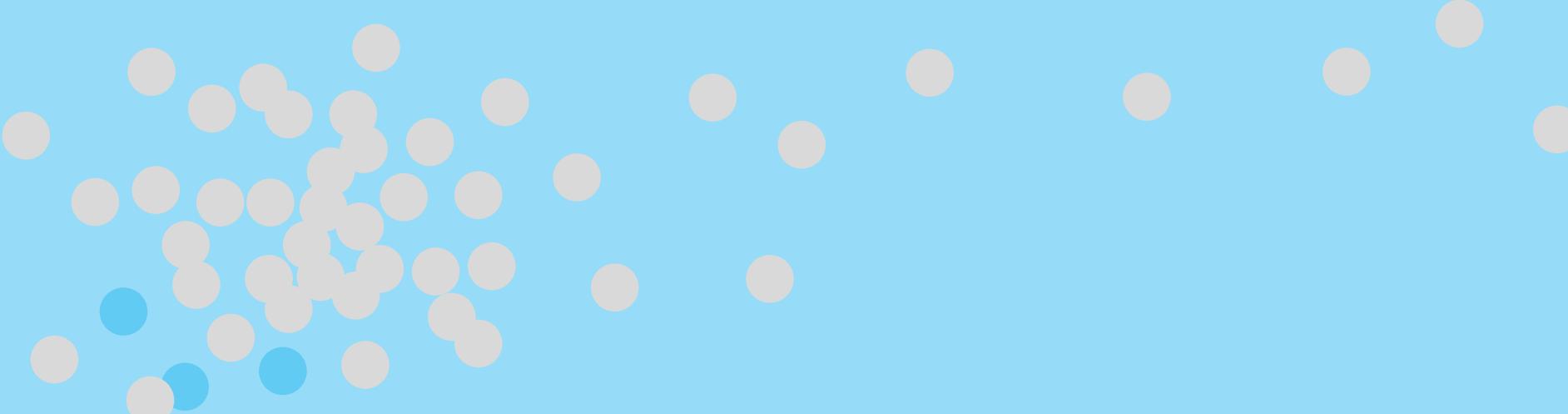
FORTY

FIFTY

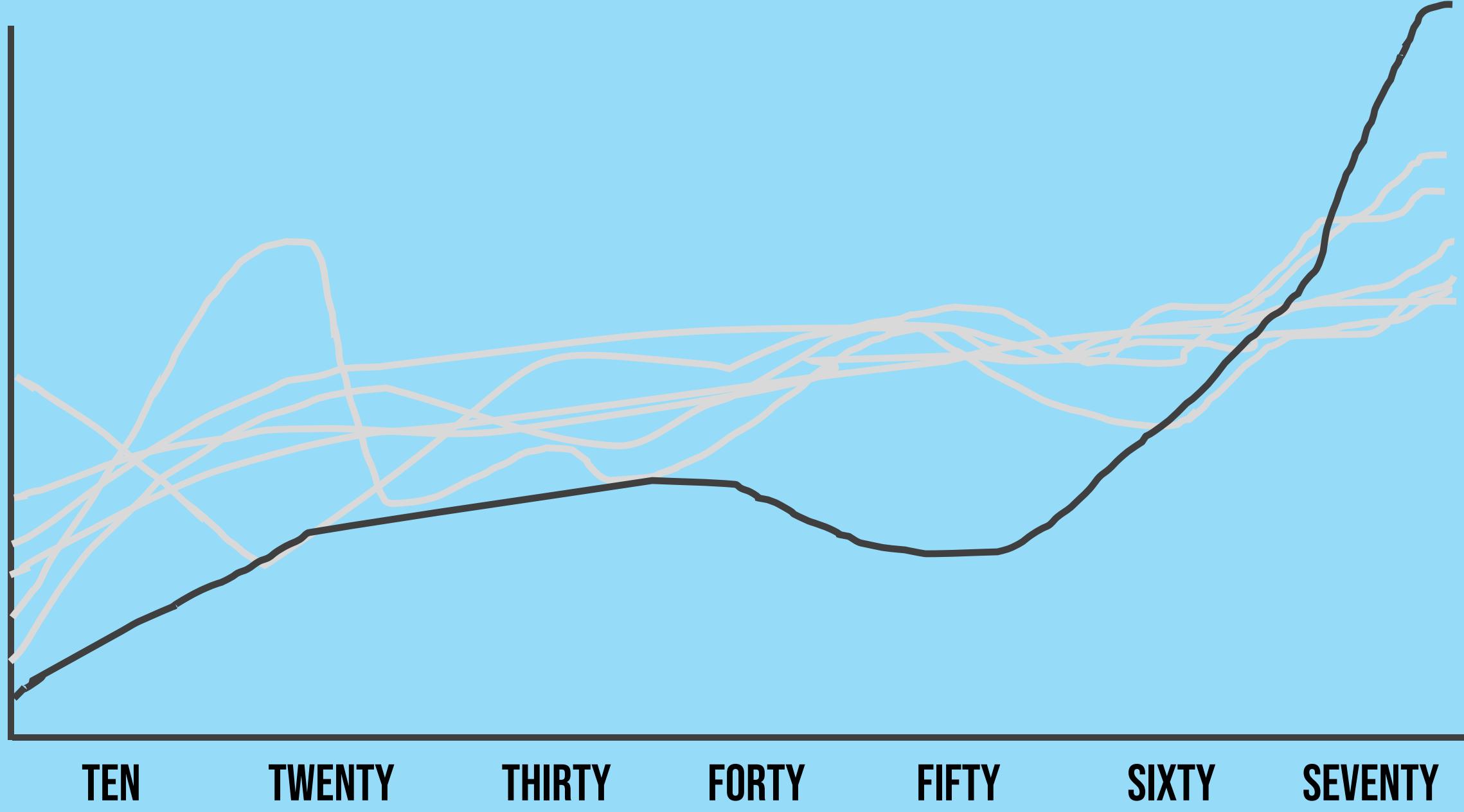
SIXTY

SEVENTY

INSTEAD DO THIS

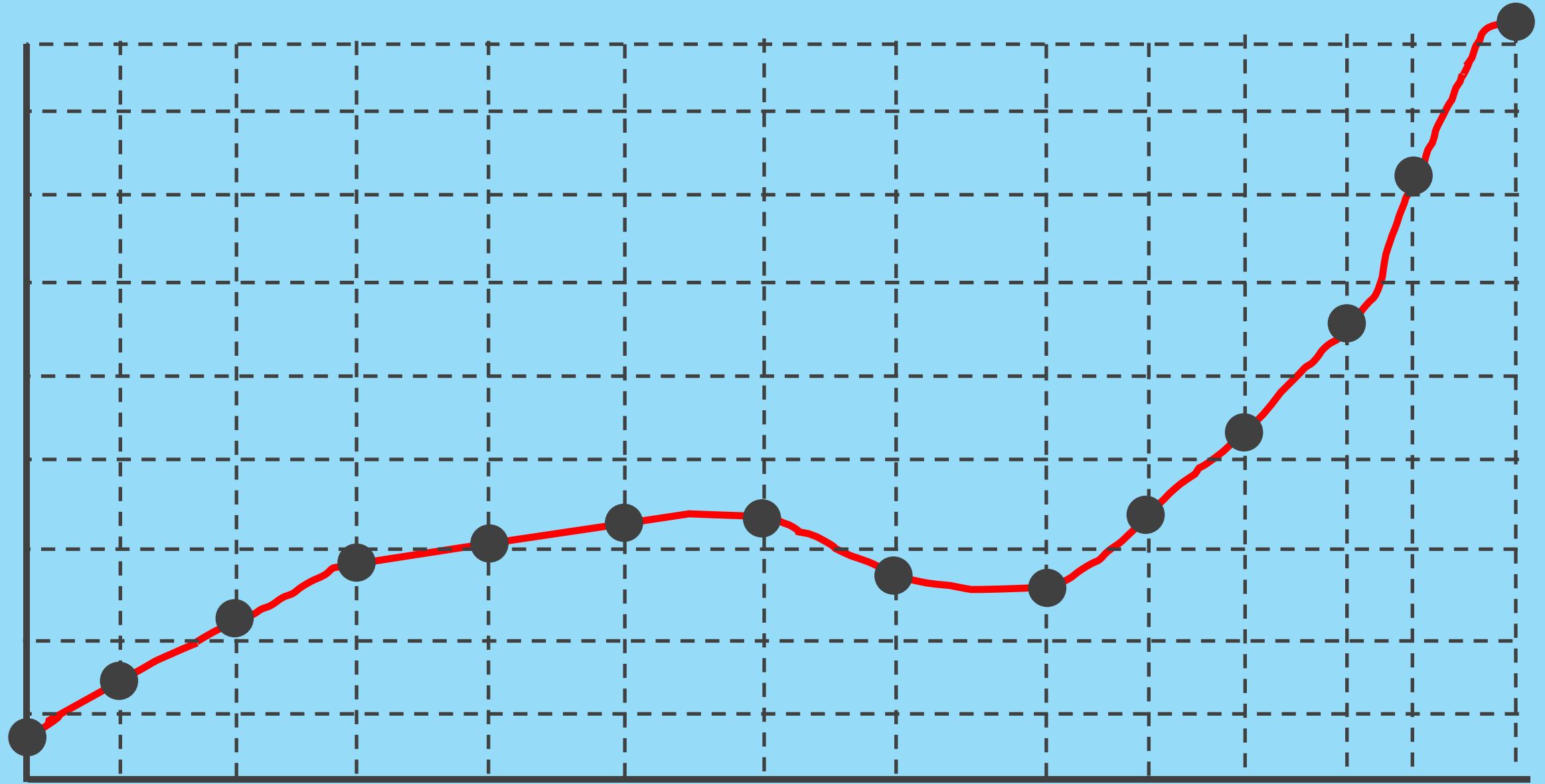


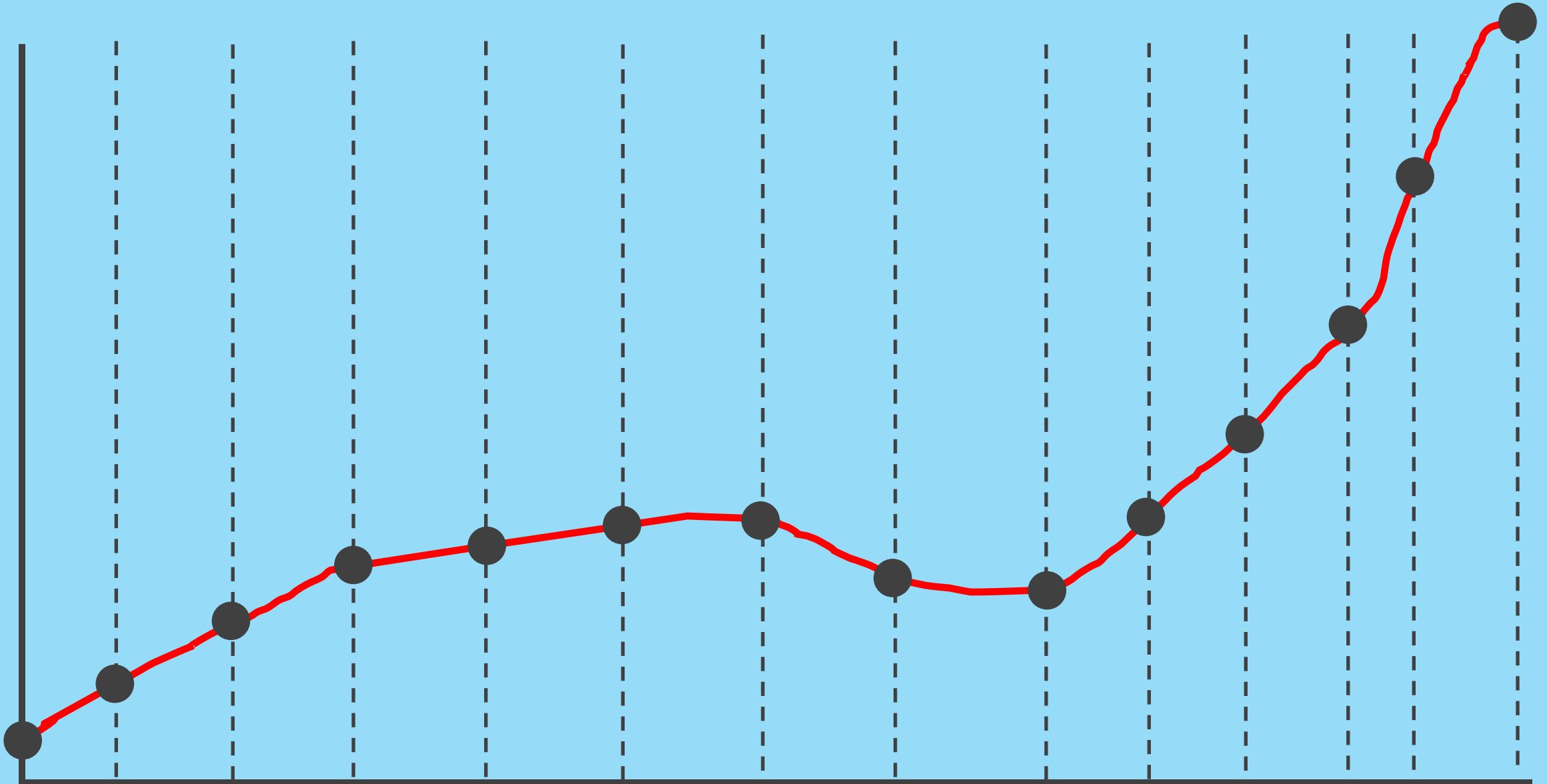
MORE NUMBERS HERE, TOO



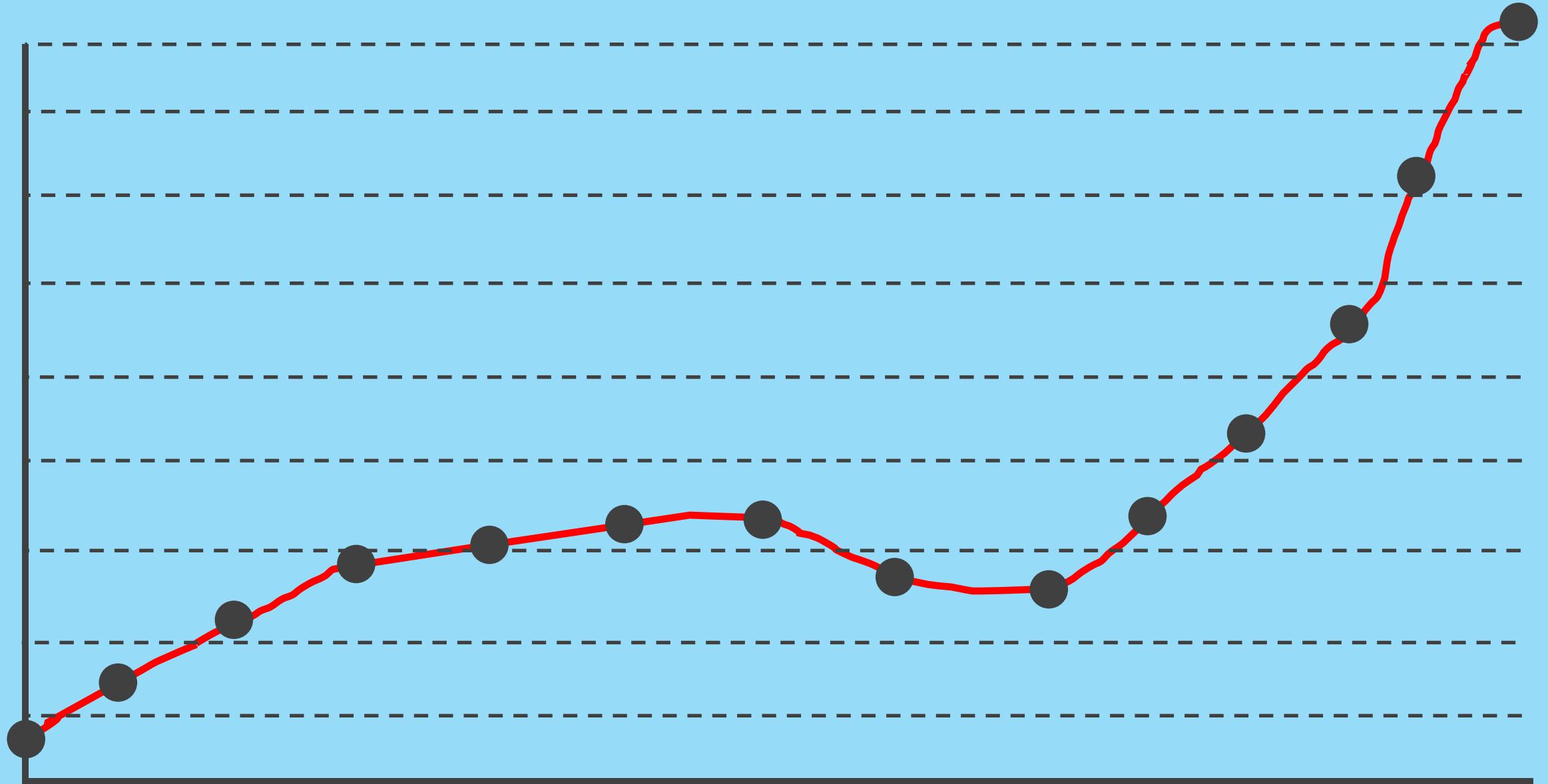
KEEPING YOUR LINES IN LINE

...SHOW YOUR USER ONLY WHAT THEY NEED TO KNOW.

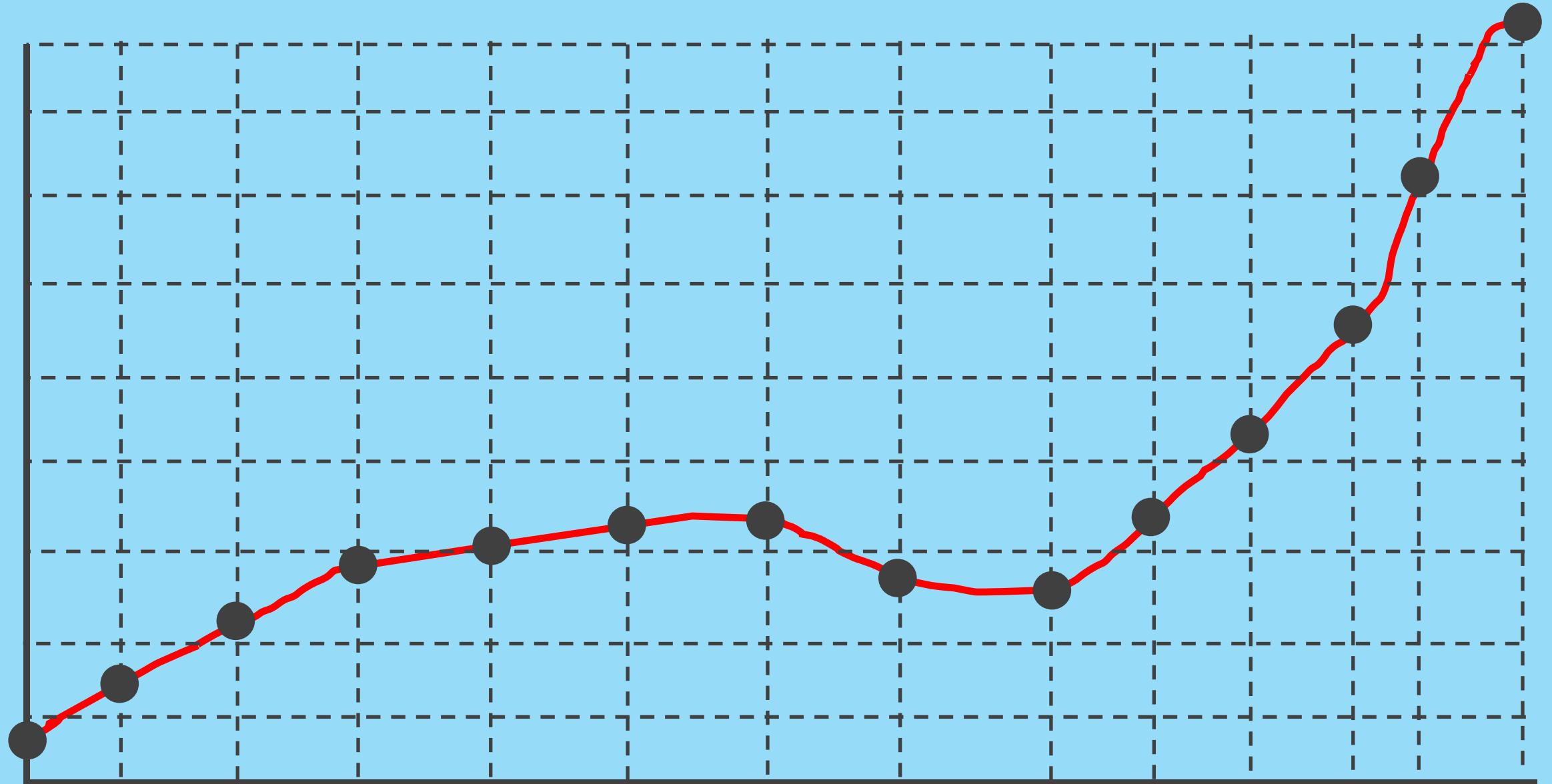




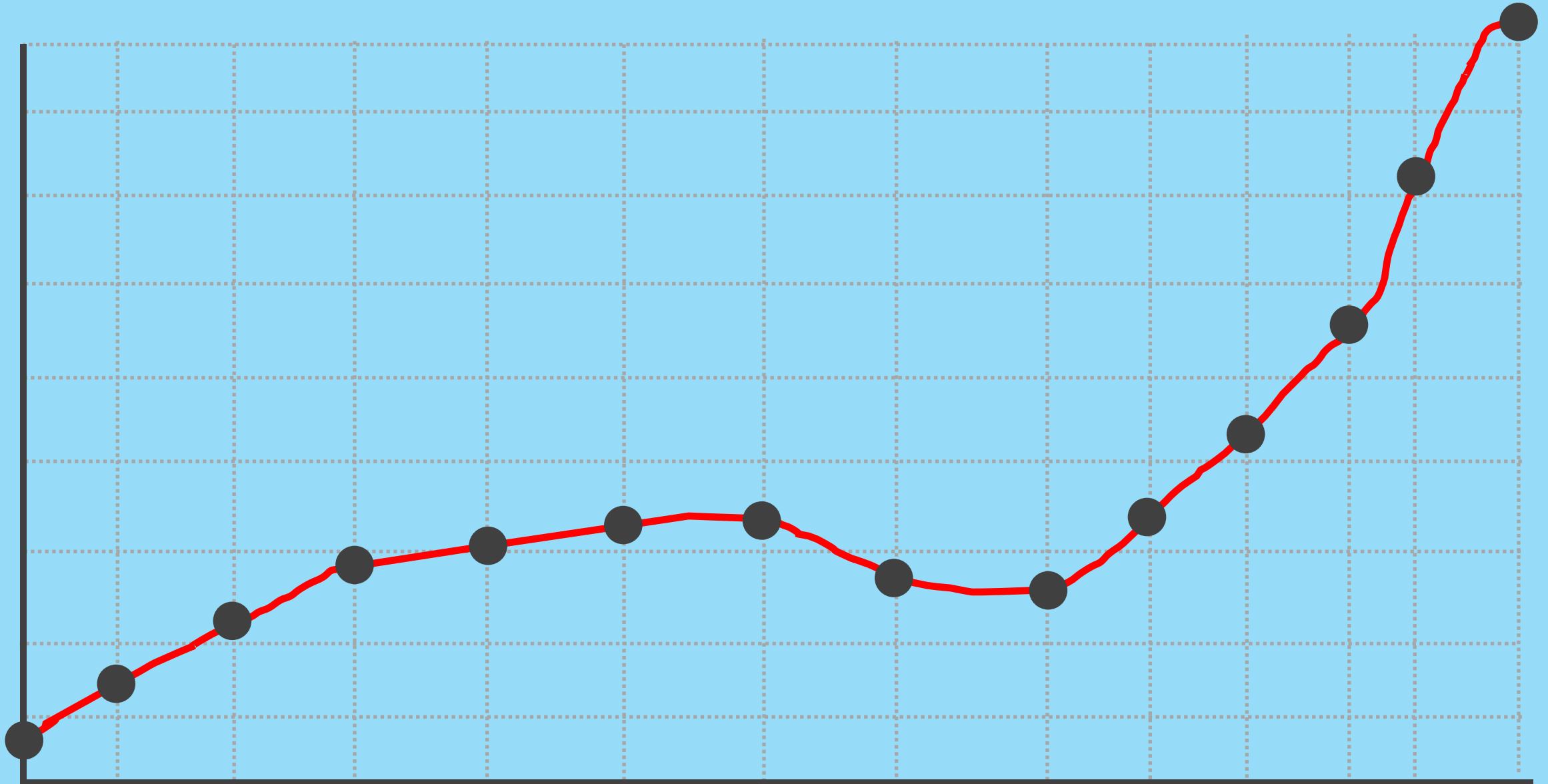
DO YOU NEED TO KNOW EXACTLY WHERE EACH POINT IS?



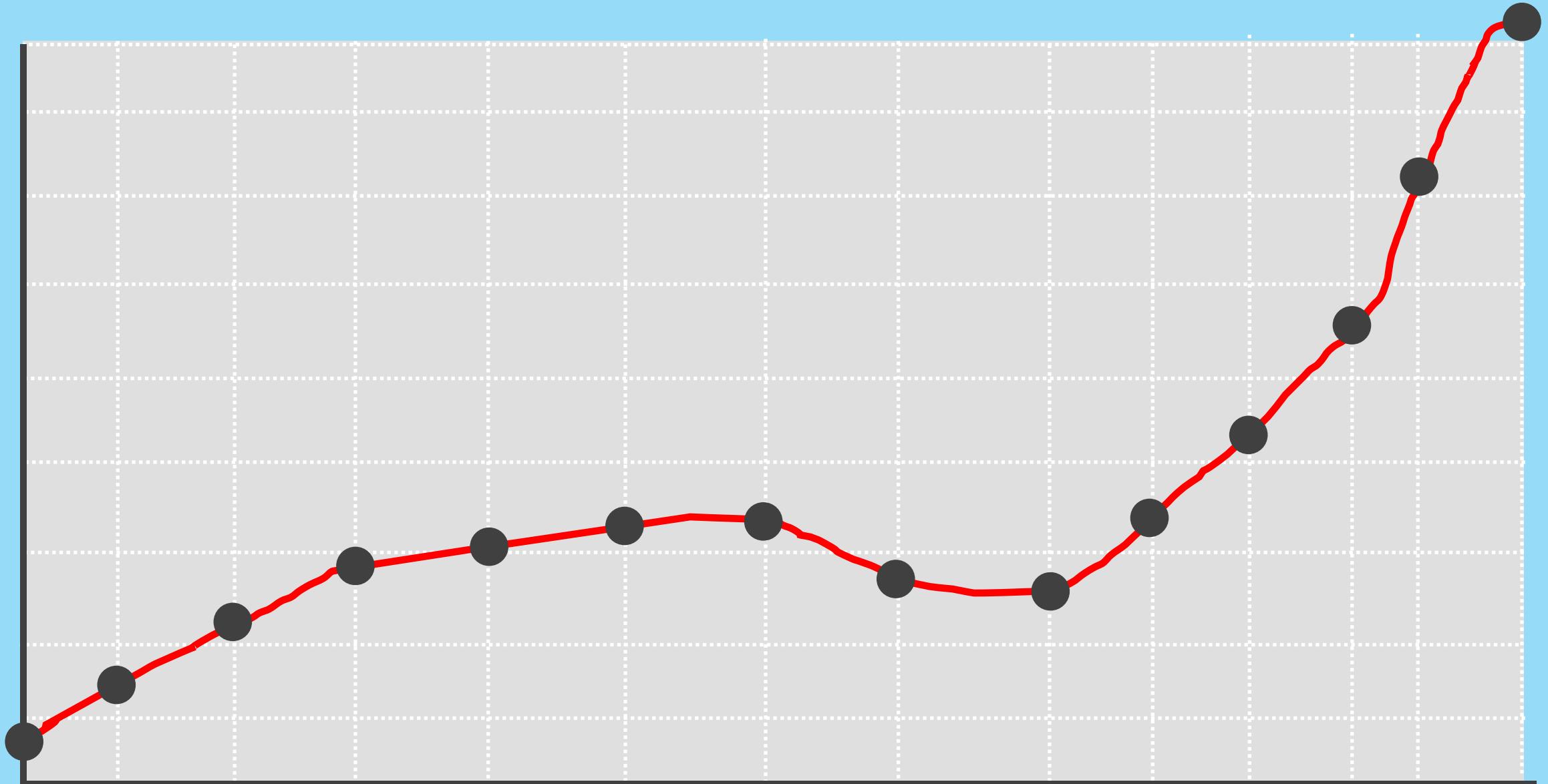
DO YOU NEED TO KNOW EXACTLY WHAT EACH VALUE IS?



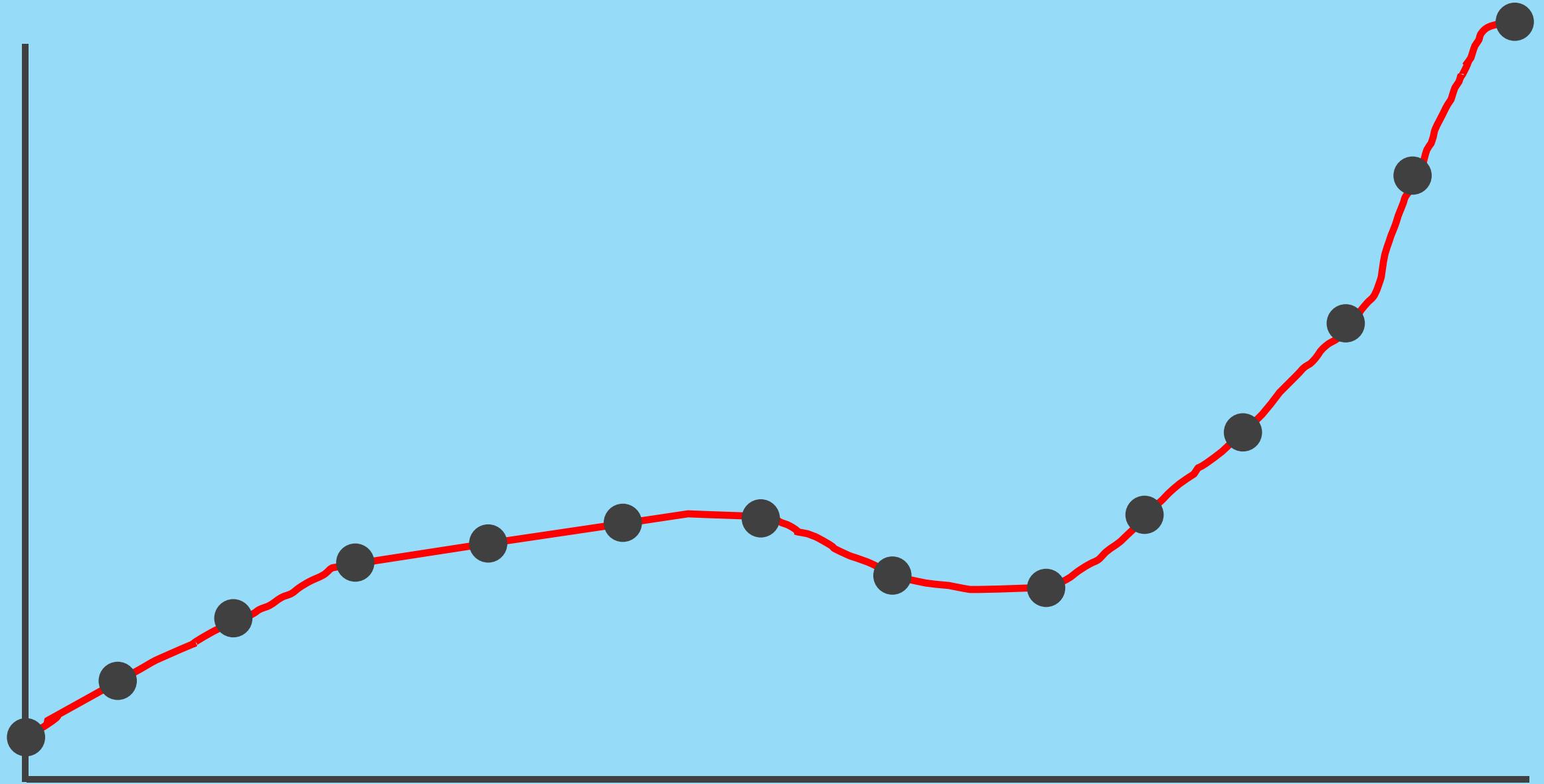
DO YOU NEED TO KNOW EVERYTHING?

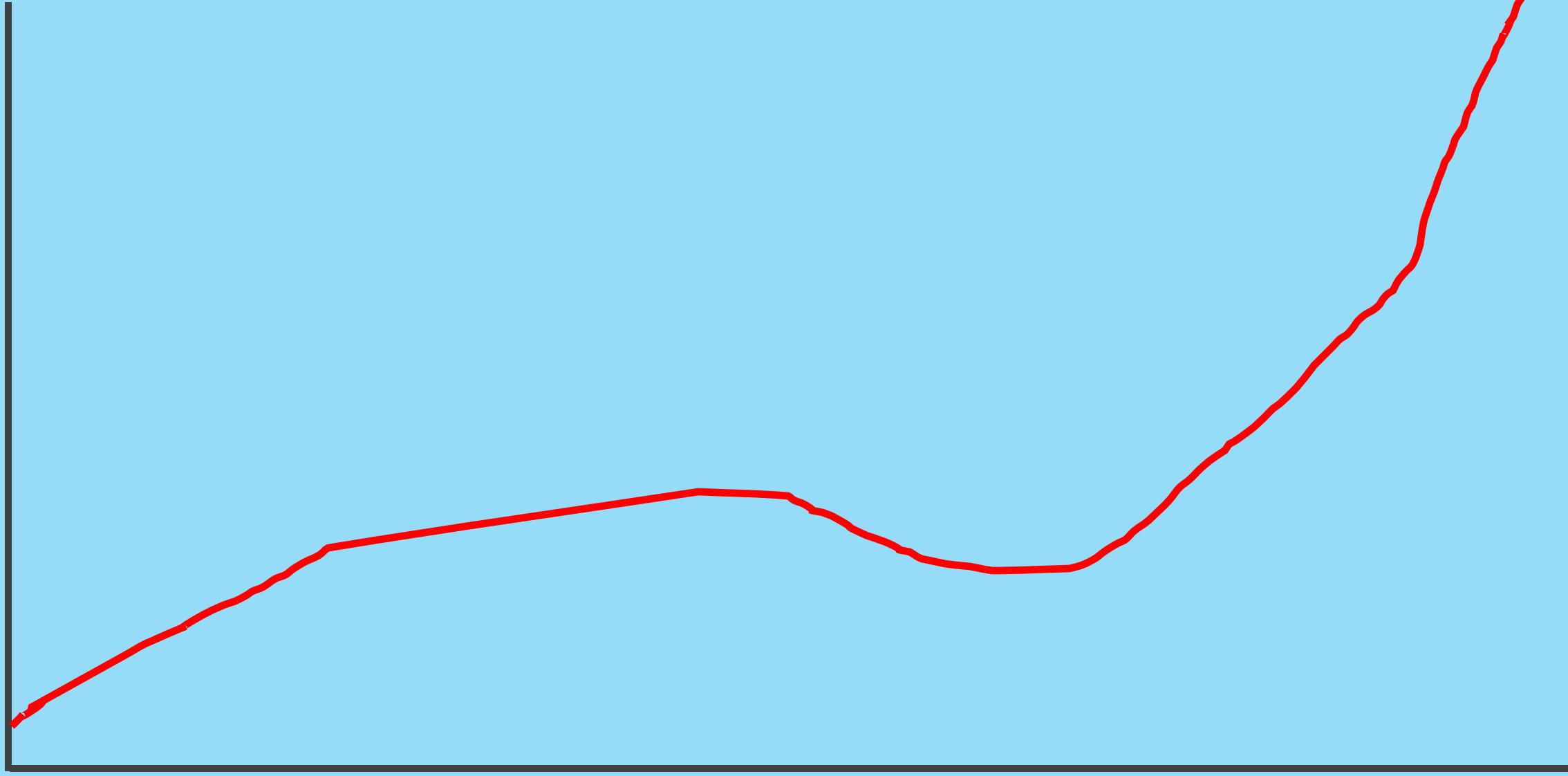


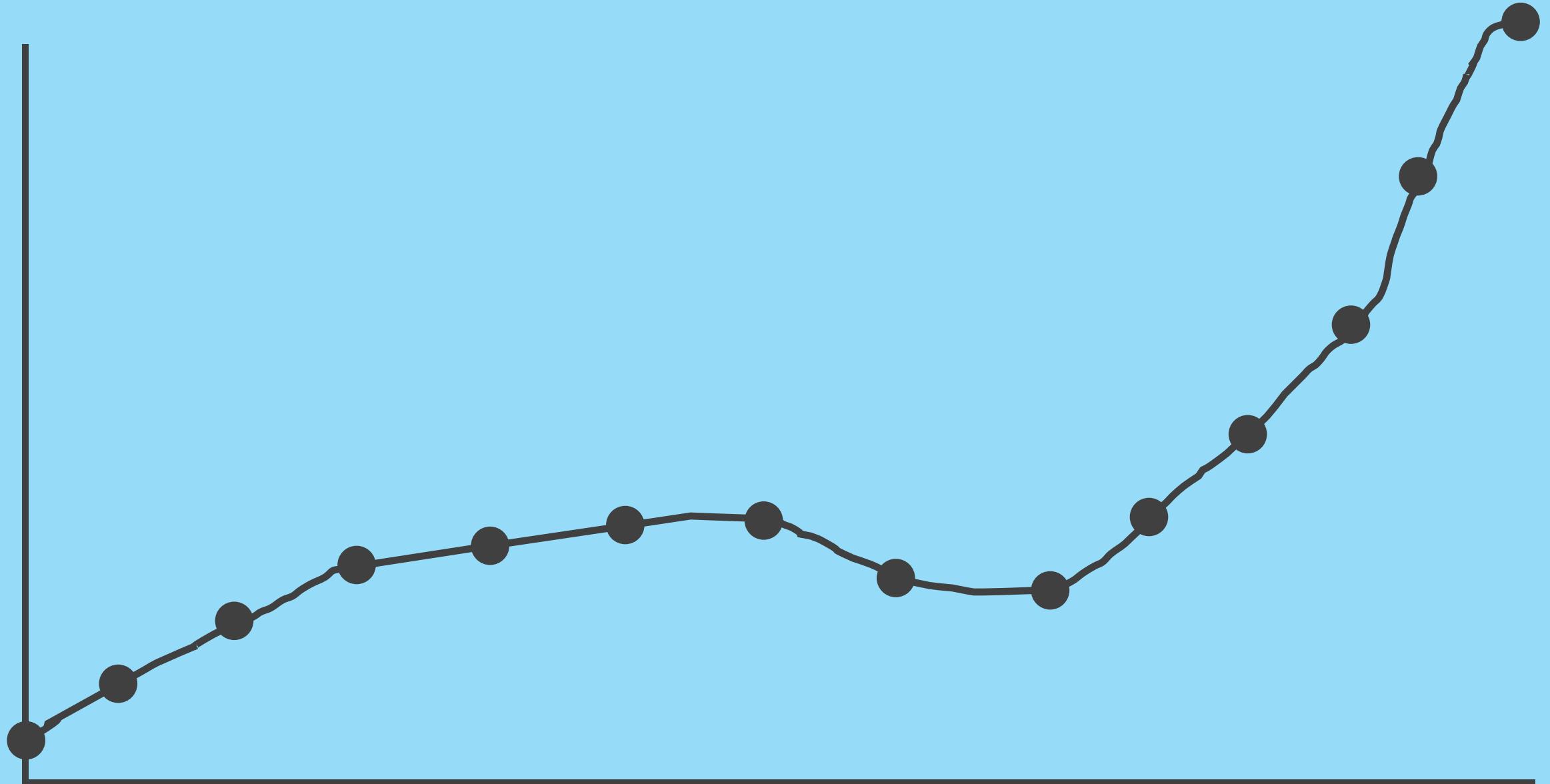
HIDE YOUR GUIDE LINES TO SIMPLIFY



ALTERNATIVE: WHITE OVER GREY







MATCH YOUR LINES AND YOUR DOTS



MATCH YOUR LINES AND YOUR DOTS

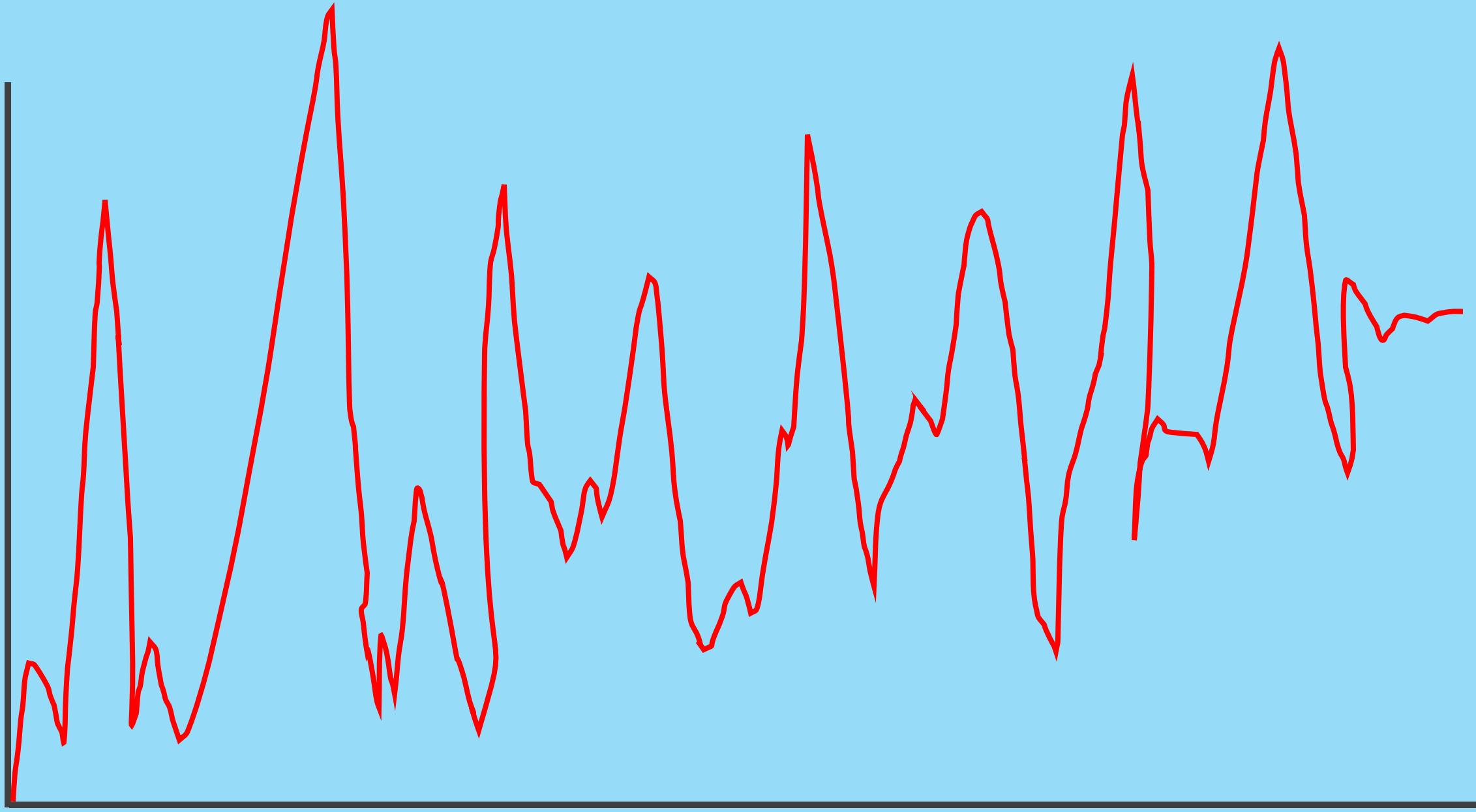


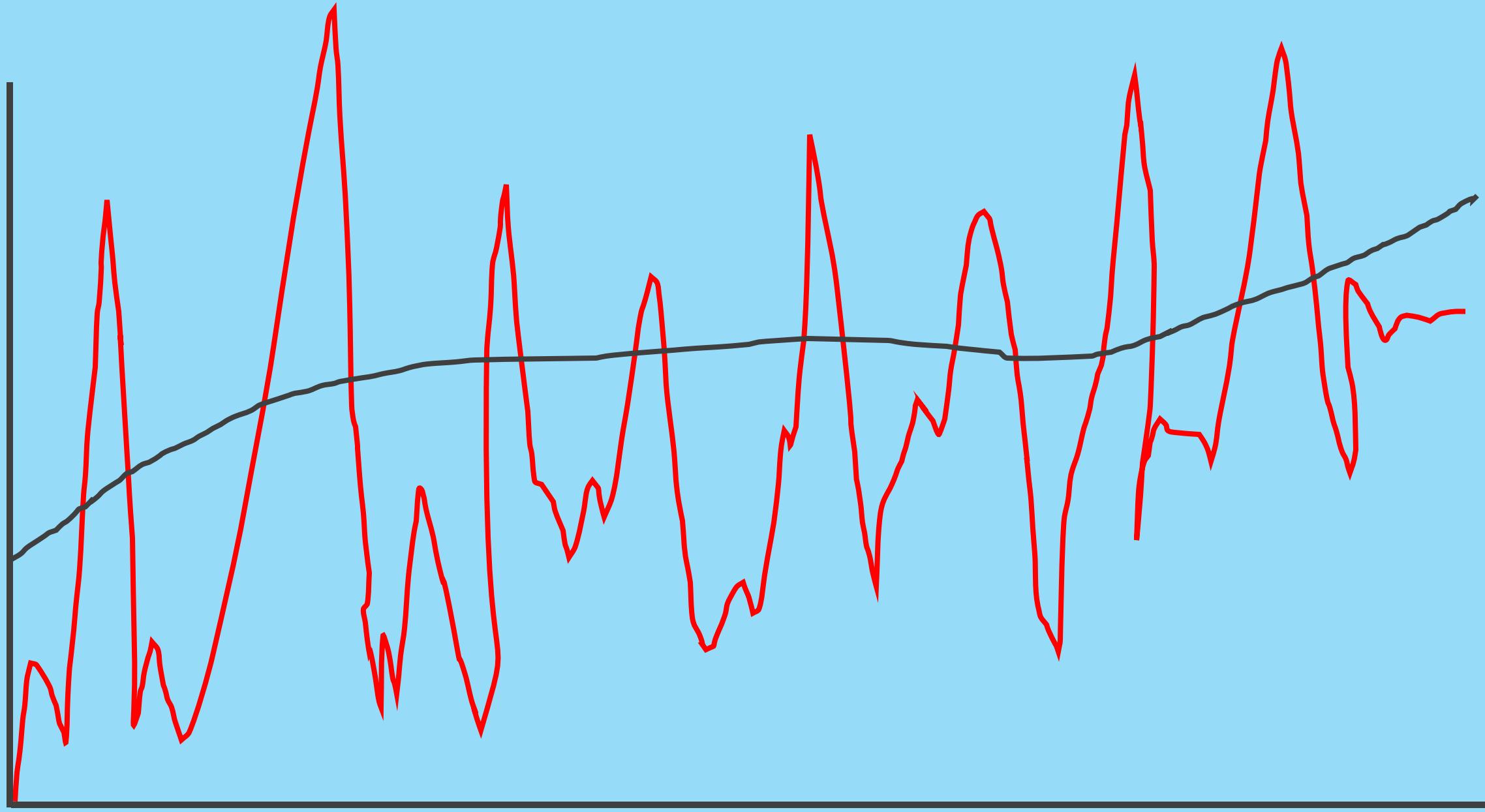
HERE IS ANOTHER NOTE

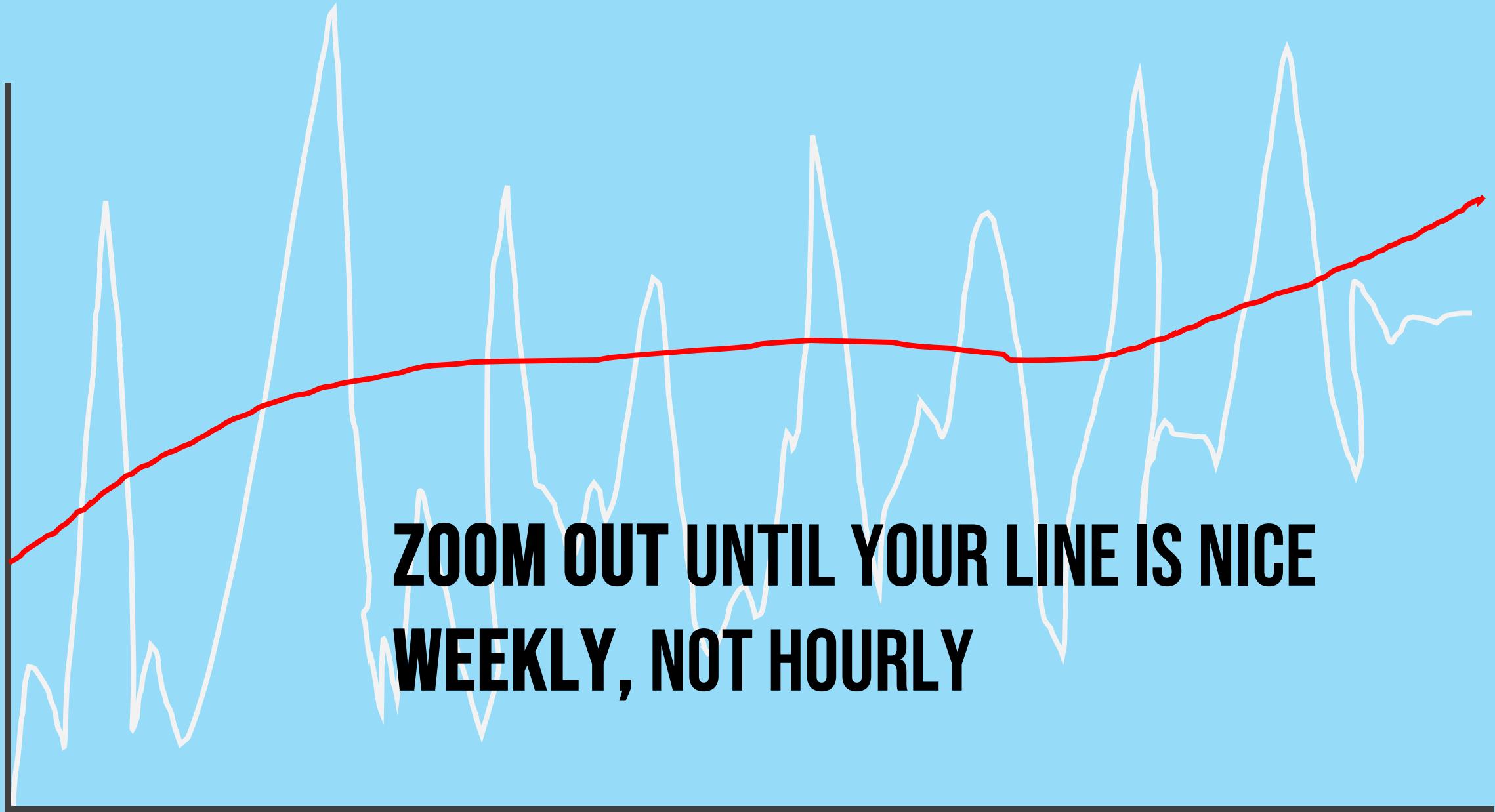
HERE IS A NOTE

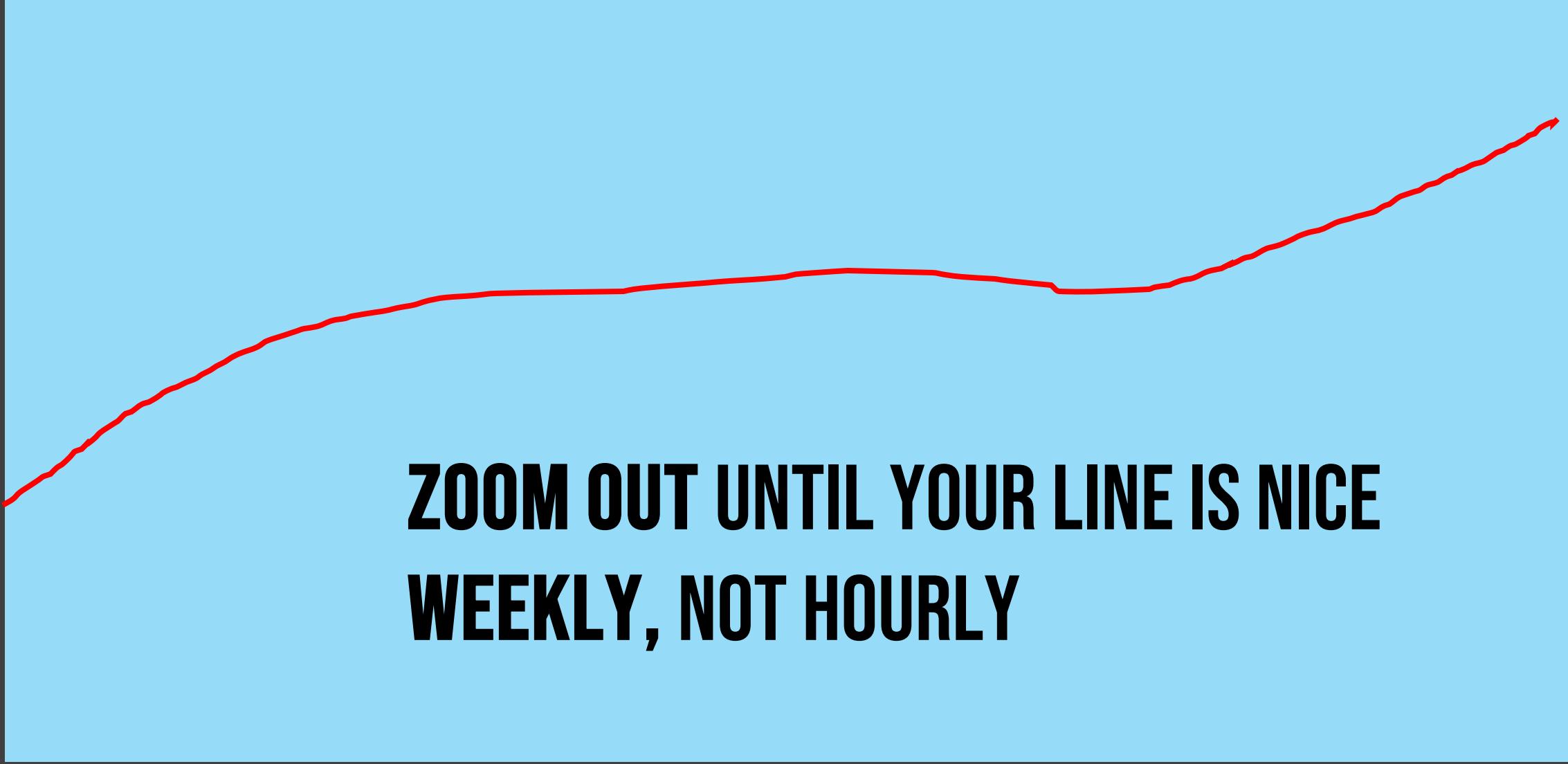
NOTE

MATCH YOUR LINES AND YOUR DOTS





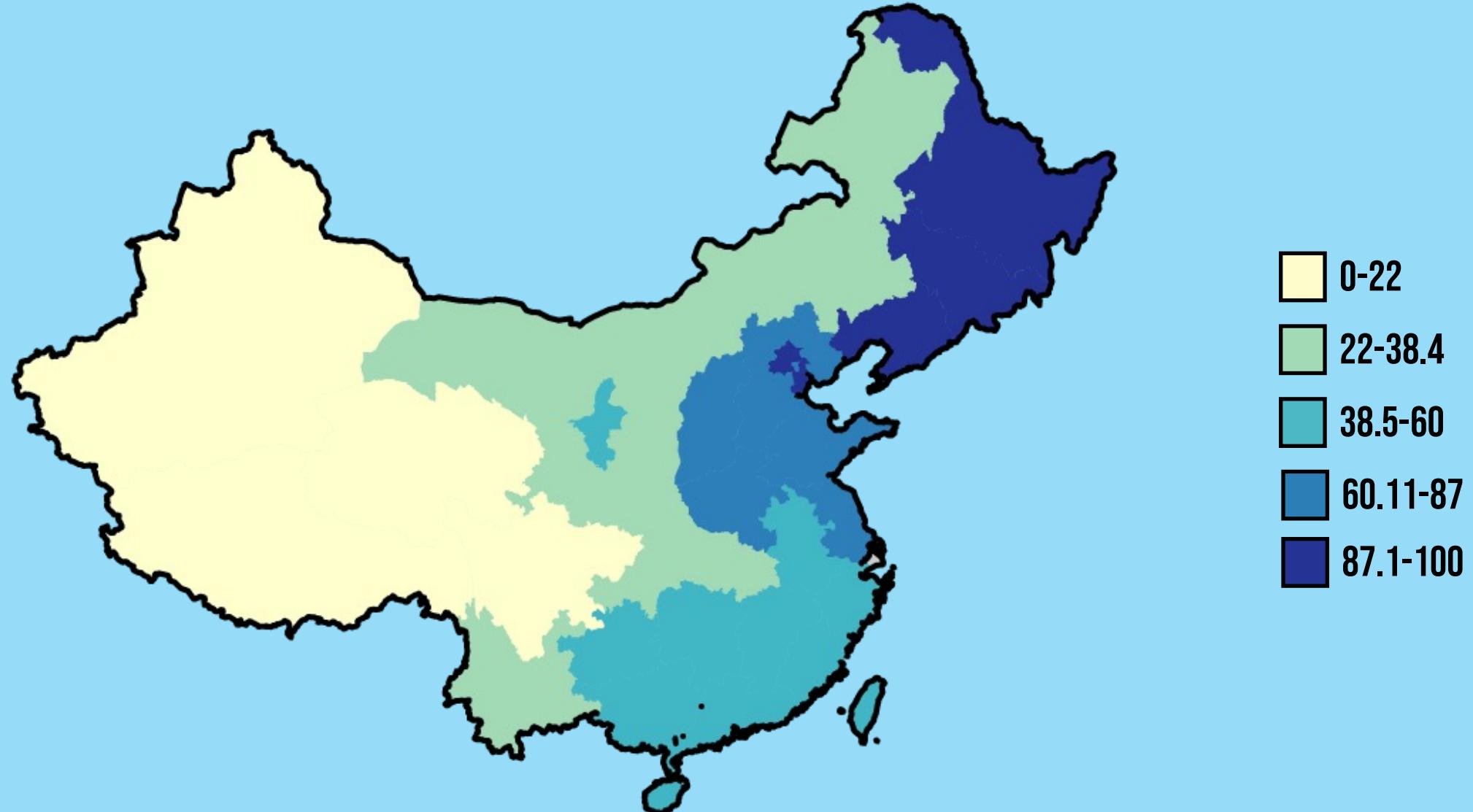




**ZOOM OUT UNTIL YOUR LINE IS NICE
WEEKLY, NOT HOURLY**

BEAUTIFUL LEGENDS

...SIMPLIFY, SIMPLIFY, SIMPLIFY

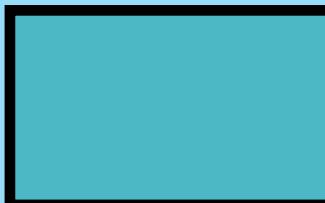




0-22



22-38.4



38.5-60



60.11-87



87.1-100



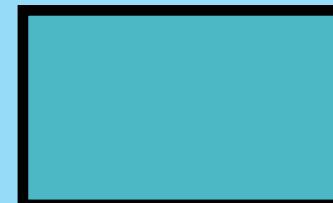
ROUND OFF TO "REAL" NUMBERS



0-20



21-40



41-60

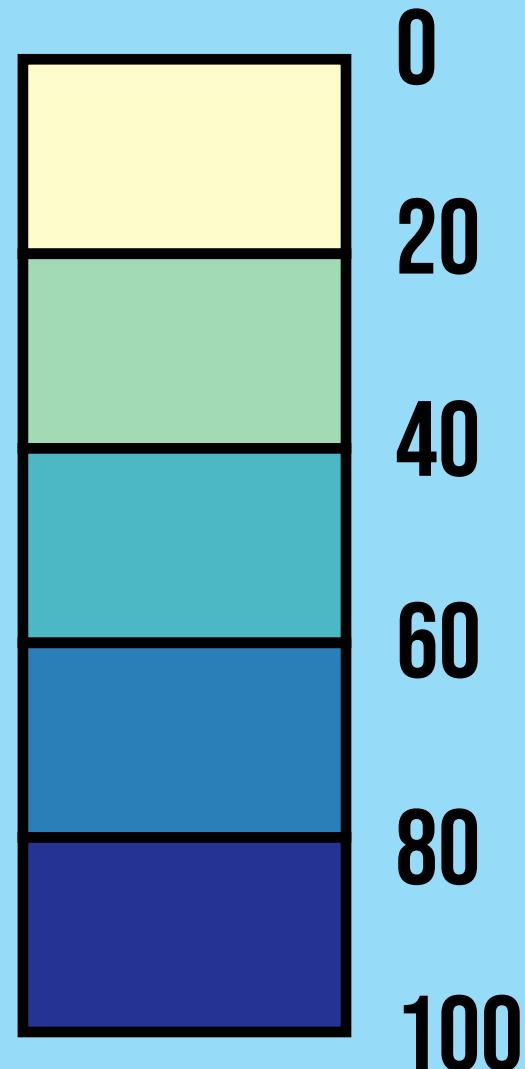


61-80

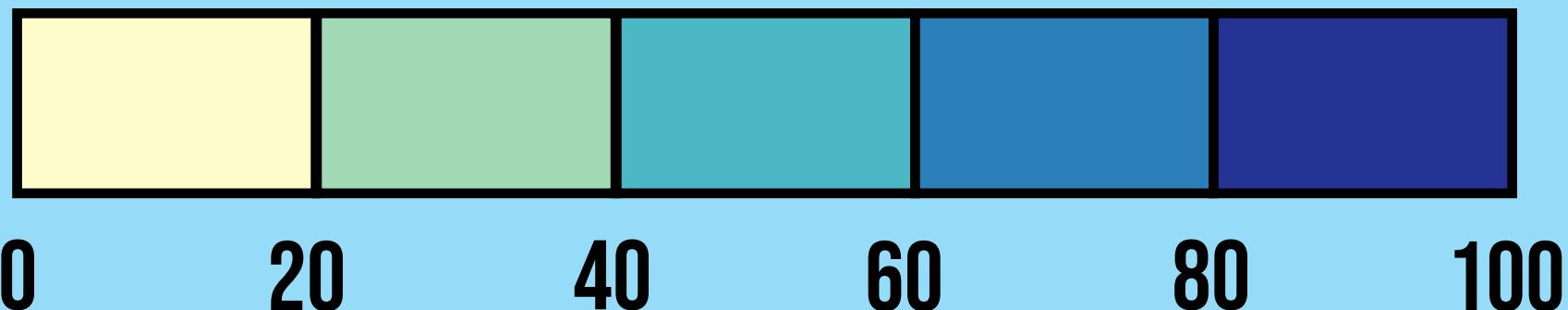


81-100

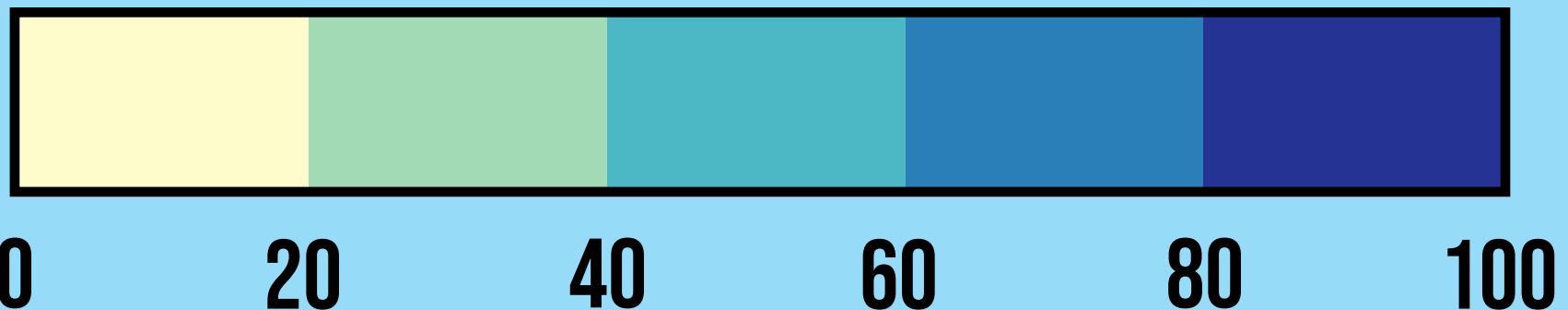
MARK BOUNDS, NOT RANGES



NUMERIC SCALES GO HORIZONTAL



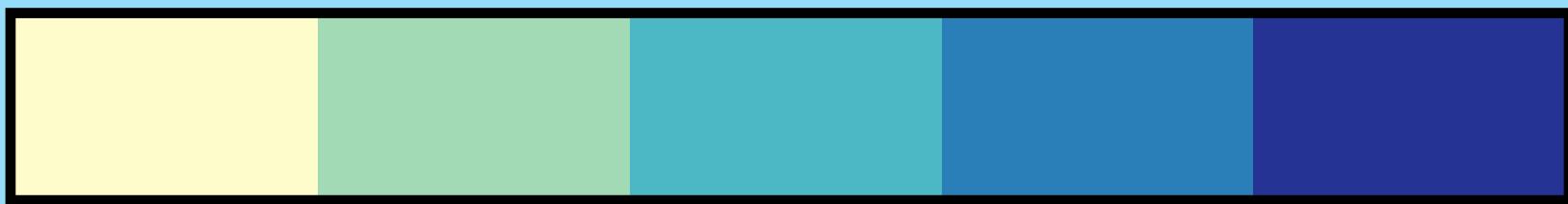
TOO MANY OUTLINES ARE UGLY



0 20 40 60 80 100



DO YOU NEED EVERY NUMBER?



20

40

60

80

0

100



DO YOU NEED EVERY COLOR?



20

40

60

80

0

100

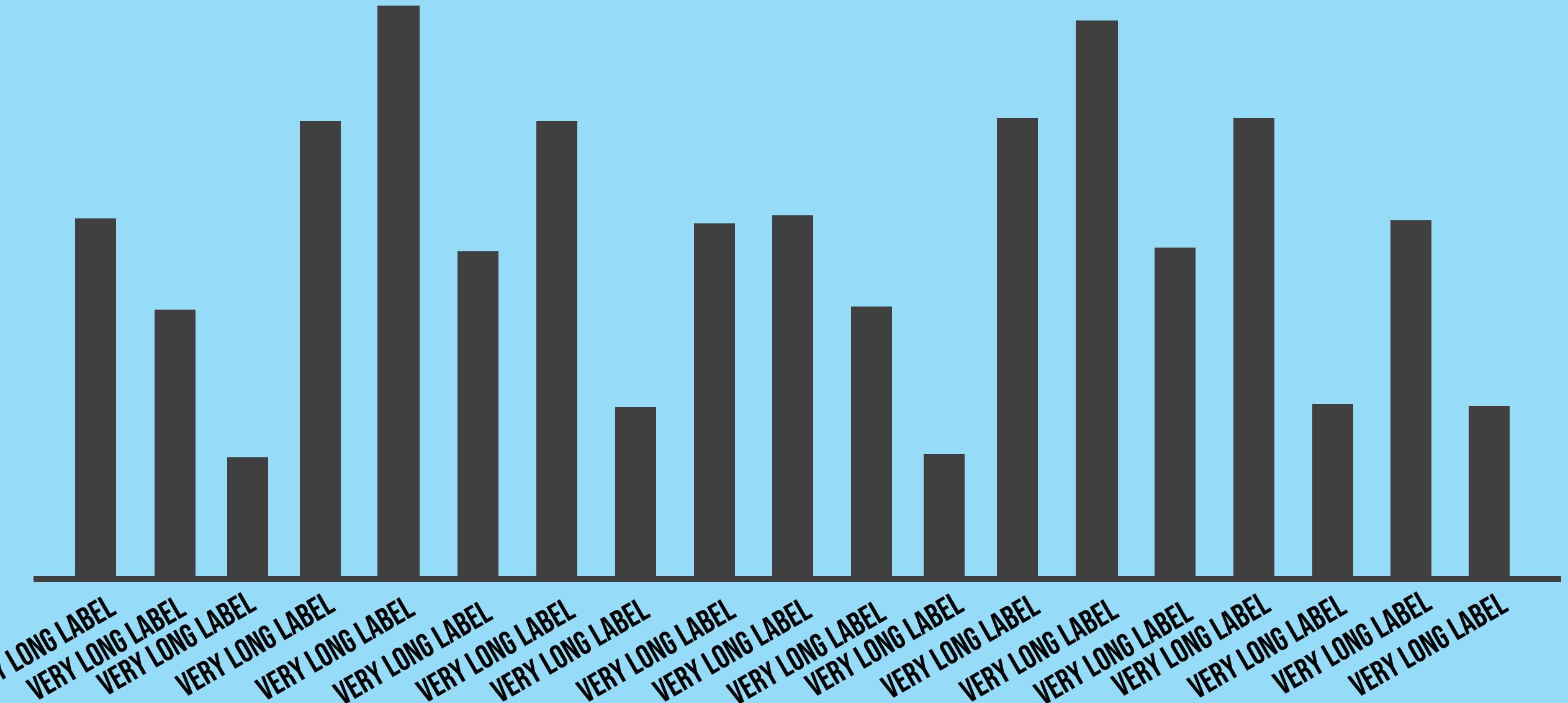
(THESE ARE UGLY, I HATE THEM)



BAR AND COLUMN GRAPHS

...ARE SIMPLE TO MAKE AND SIMPLE TO IMPROVE

LONG LABELS + VERTICAL COLUMNS = CLUTTERED



LONG LABELS GET HORIZONTAL BARS

A VERY VERY VERY LONG LABEL

MANY MANY WORDS CAN BE HERE

AN EASY-TO-READ LABEL

ANOTHER EASY-TO-READ LABEL

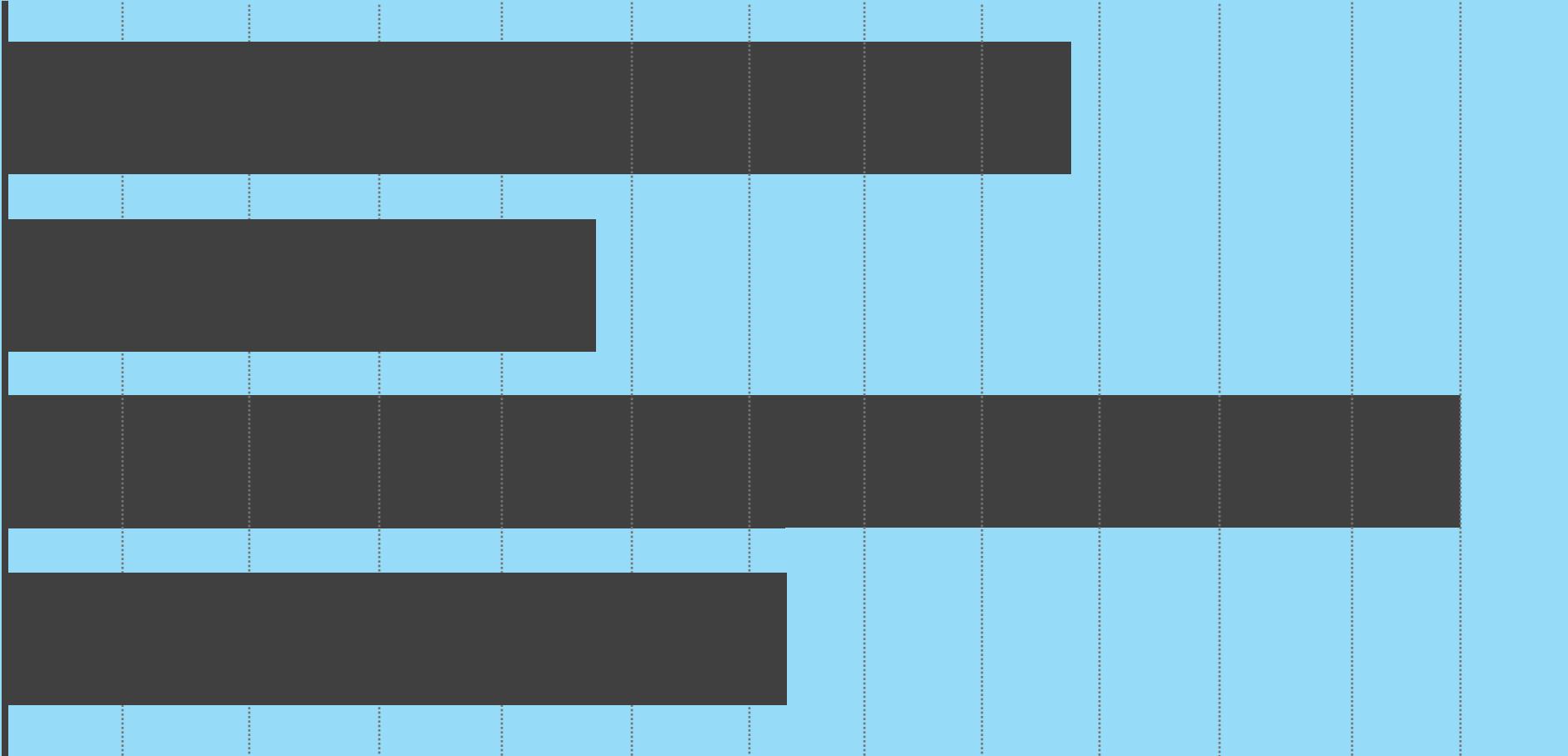
LABEL LABEL LABEL LABEL LABEL

DOGS

BIRDS

CATS

MICE



10

20

30

40

50

60

70

80

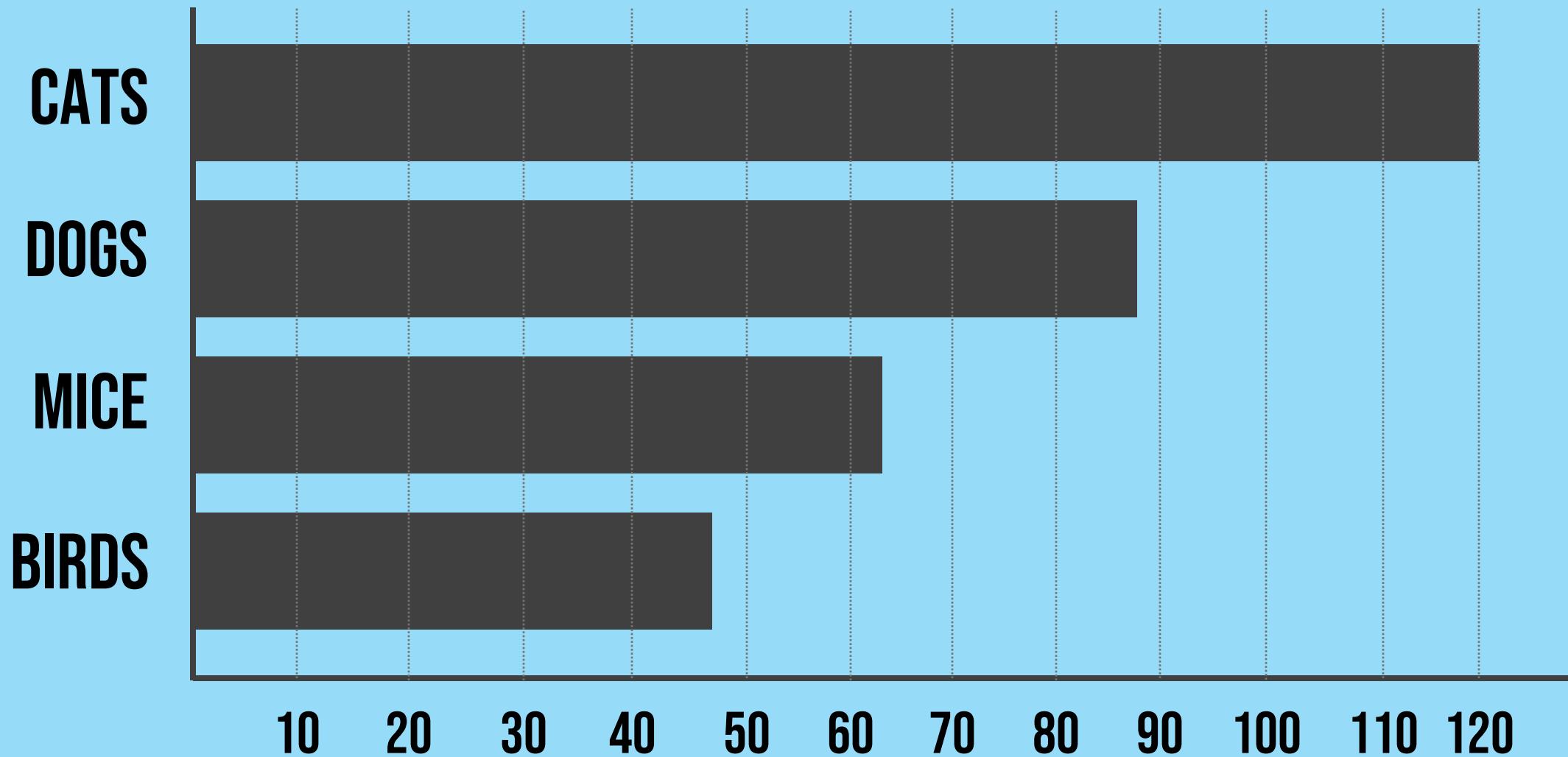
90

100

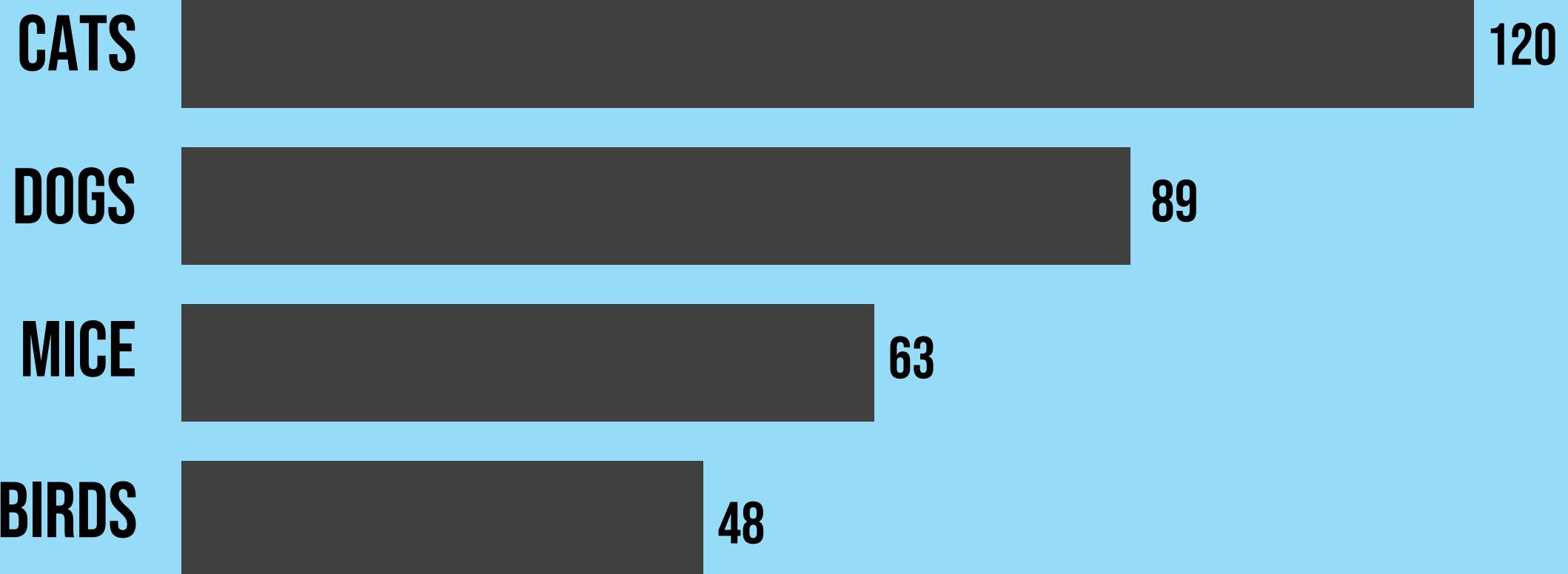
110

120

KEEP YOUR BARS IN ORDER



DIRECTLY LABEL IF YOU HAVE FEW DATA POINTS



YOU CAN GO INSIDE, TOO!

CATS



120

A horizontal bar chart with four bars of increasing length from bottom to top. The bars are dark grey with white numerical labels at their right ends. The first bar (bottom) is labeled 'BIRDS' and has a value of 48. The second bar is labeled 'MICE' and has a value of 63. The third bar is labeled 'DOGS' and has a value of 89. The fourth bar (top) is labeled 'CATS' and has a value of 120.

DOGS

89

MICE

63

BIRDS

48

MORE NUMBERS HERE, TOO

TEN

TWENTY

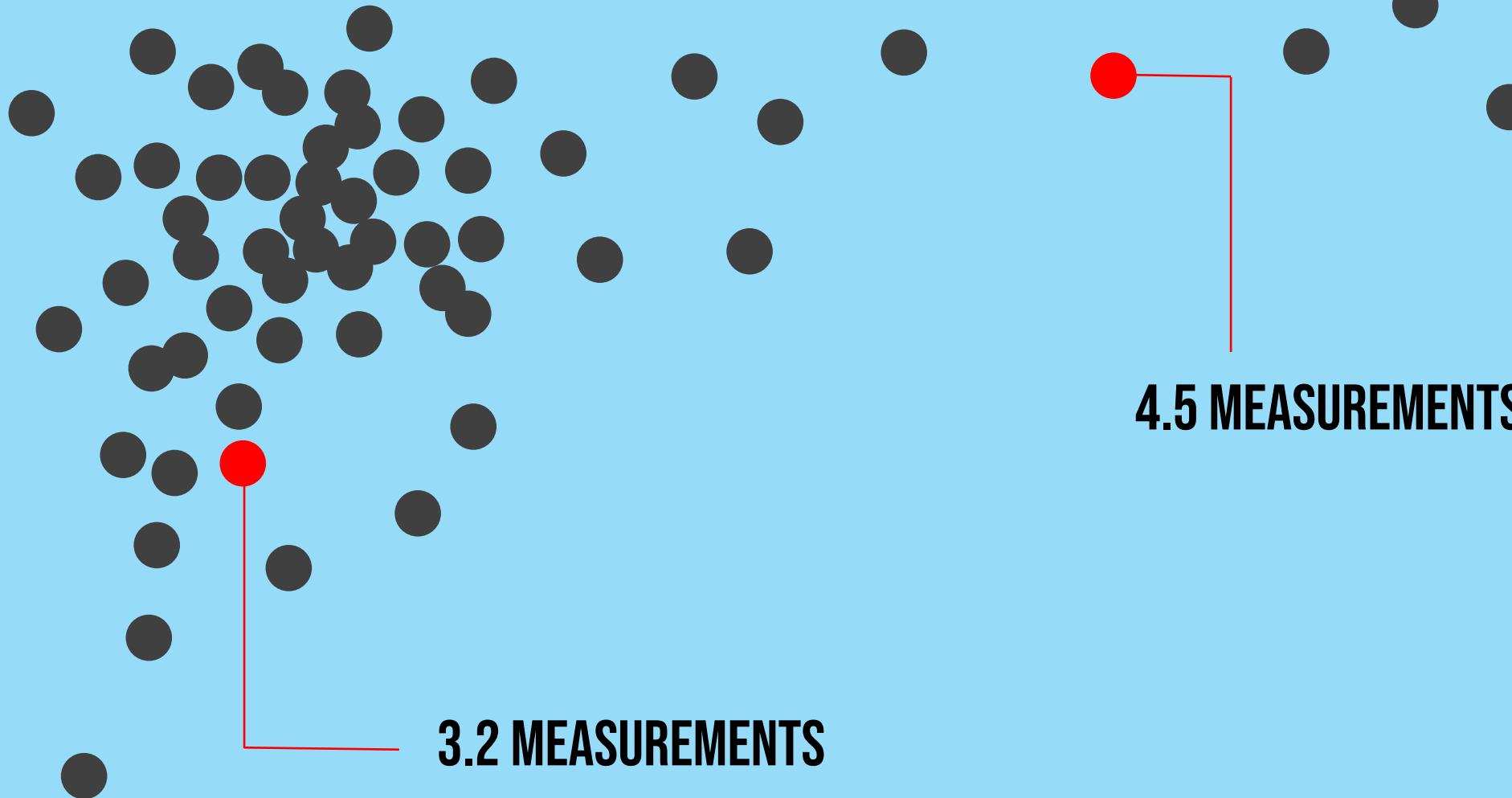
THIRTY

FORTY

FIFTY

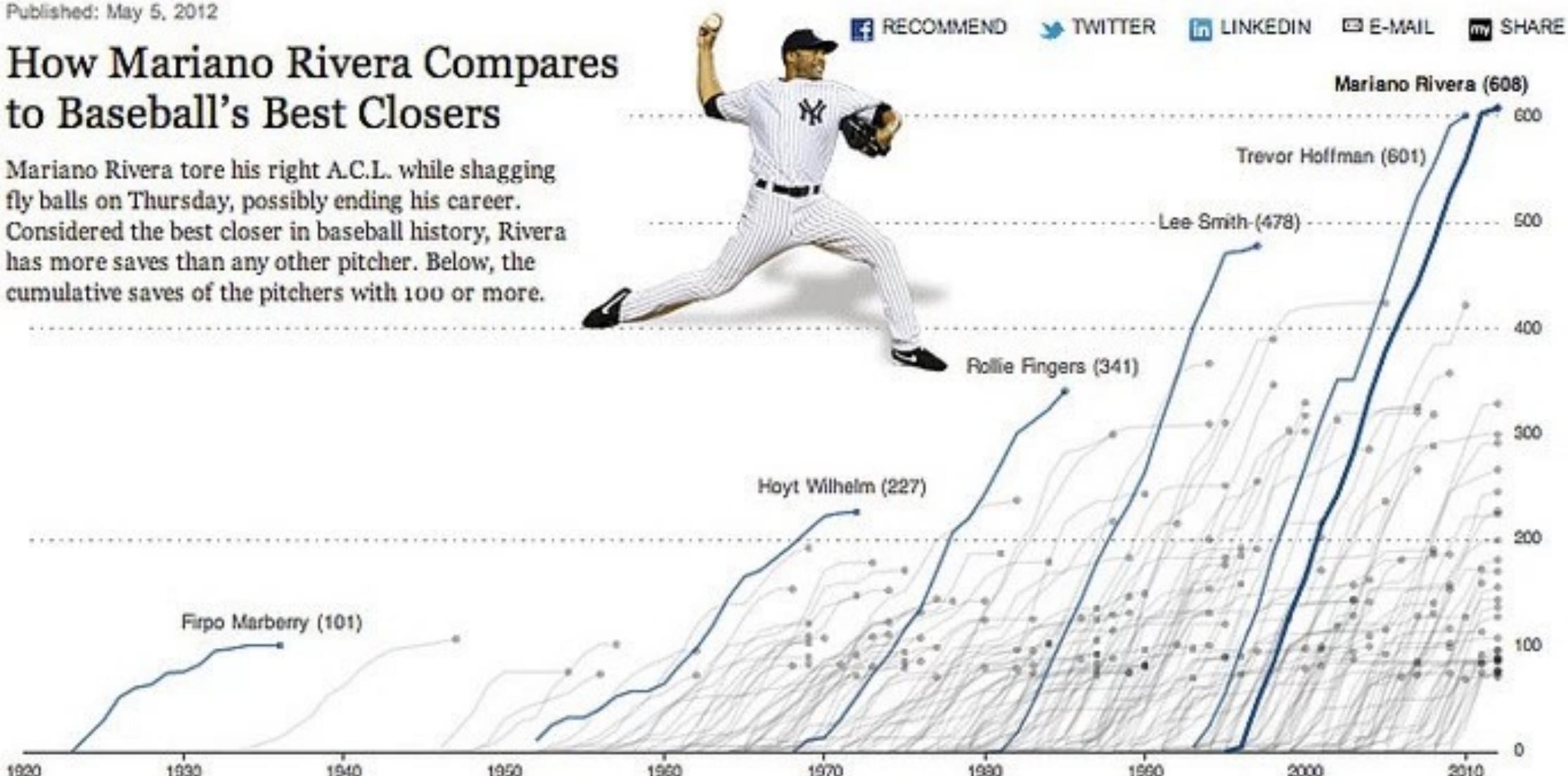
SIXTY

SEVENTY



How Mariano Rivera Compares to Baseball's Best Closers

Mariano Rivera tore his right A.C.L. while shagging fly balls on Thursday, possibly ending his career. Considered the best closer in baseball history, Rivera has more saves than any other pitcher. Below, the cumulative saves of the pitchers with 100 or more.



The closers who broke new hundred-save milestones:

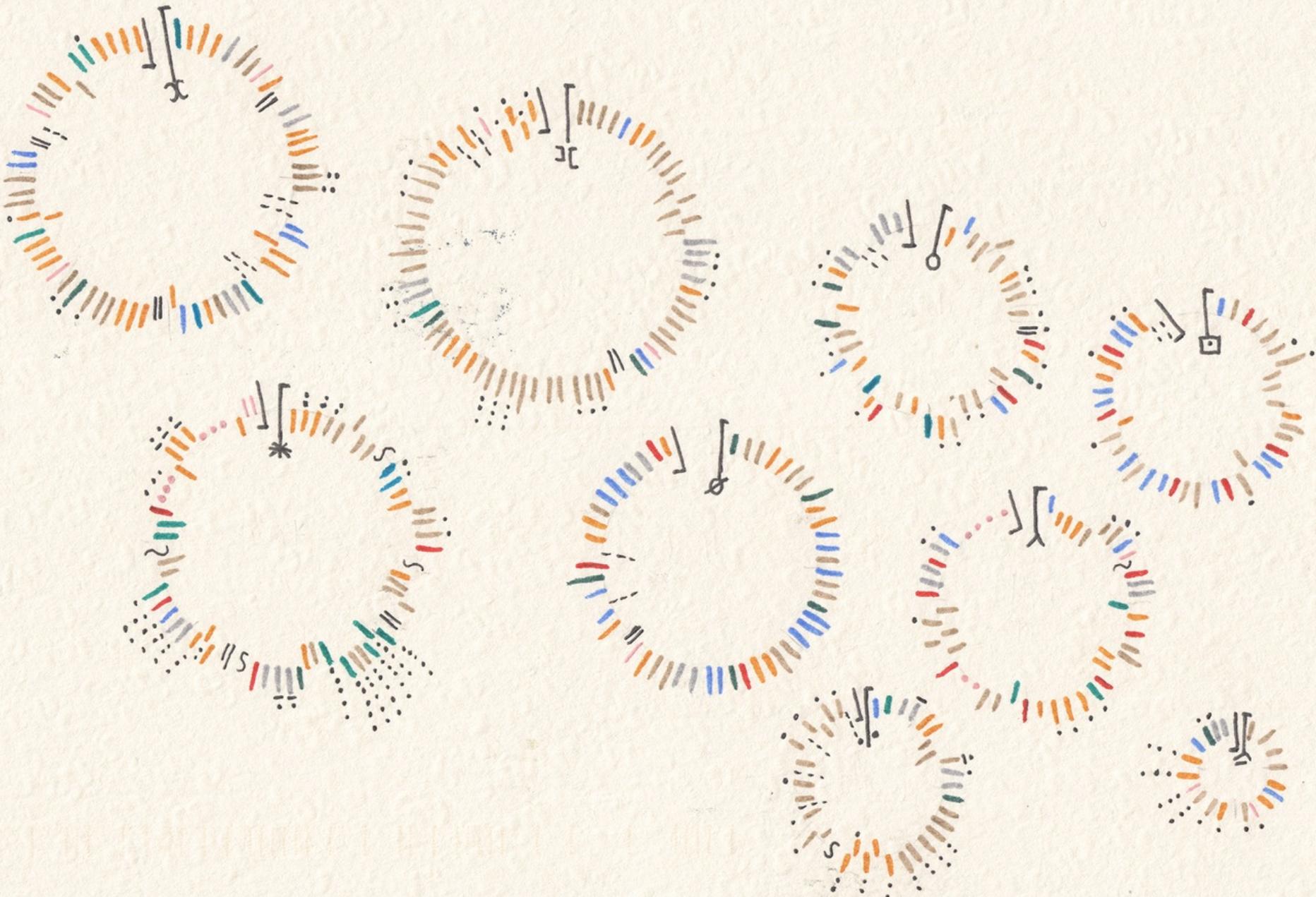
Firpo Marberry (101)
The first reliever to get to 100 cumulative saves, done at a time before relief pitchers were commonplace. (Marberry also started 186 games.)

Hoyt Wilhelm (227)
In addition to being the first pitcher to break the 200 save mark, Wilhelm pitched a no-hitter against the Yankees in 1958.

Rollie Fingers (341)
Known for his handlebar moustache, Fingers was the second relief pitcher inducted into Baseball's Hall of Fame.

Lee Smith (478)
From 1983 to 1995, Smith averaged 35 saves a season, saving no fewer than 25 in any season.

Trevor Hoffman (601)
Hoffman was the first to break the 500 and 600 save marks, despite a 1994 shoulder injury that forced him to change his pitching style.



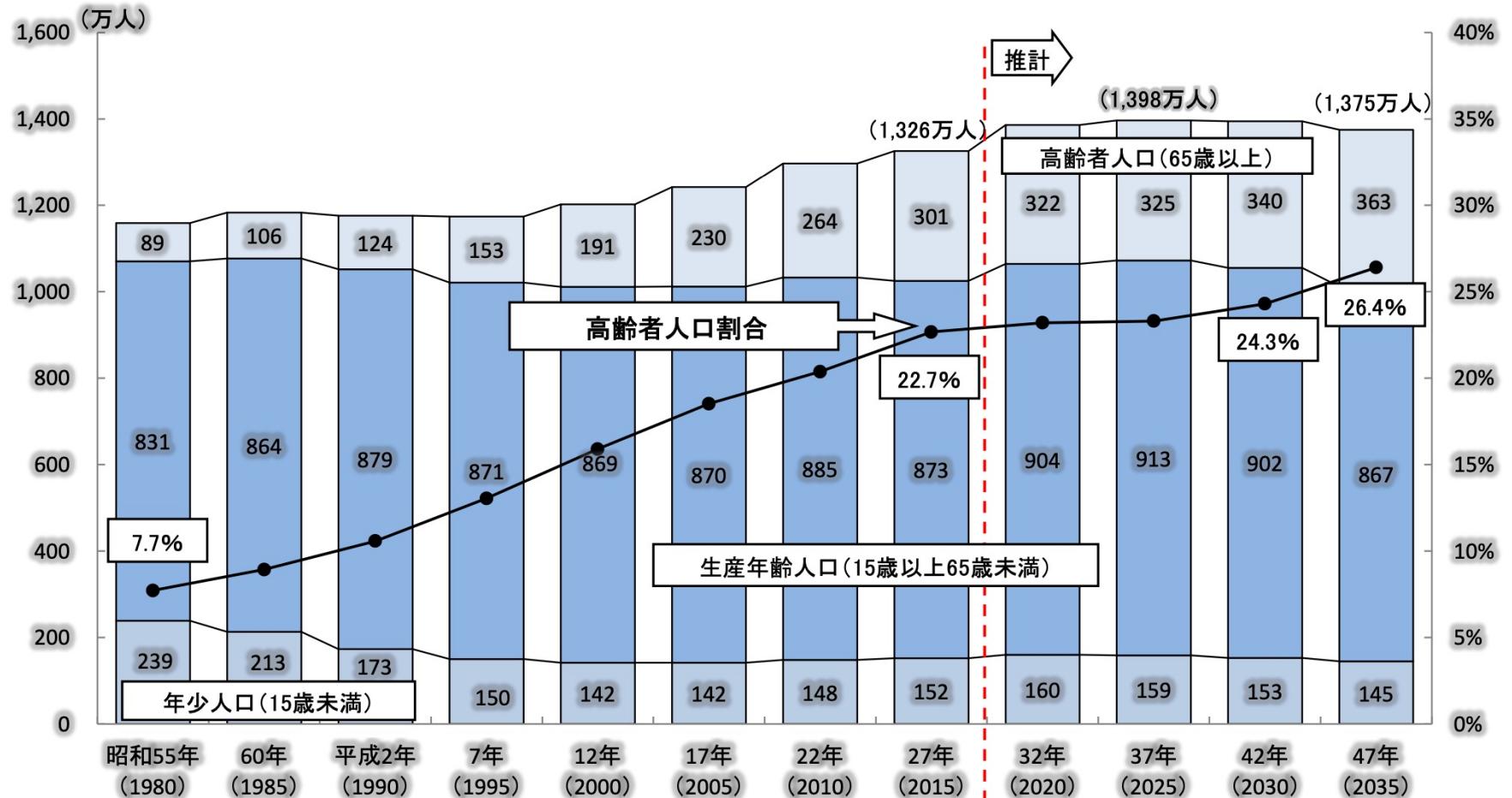
important topics



- Why annotate
- What to annotate
- When to annotate
- How to annotate

人口の推移(東京都)

東京都における高齢化率は、総人口がピークを迎える平成37年には23.3%であり、平成42年には24.3%とおよそ4人に1人が高齢者になると推計されます。



(注) ()内は総人口。1万人未満を四捨五入しているため、内訳の合計値と一致しない場合がある。

出典: 総務省「国勢調査」[昭和55年～平成27年]、東京都政策企画局による推計[平成32年～47年]

This is a blog written by people who work at Datawrapper. We love charts, and we created this place to talk about them. Here you'll find:

- Latest Datawrapper features
- Data Vis Do's & Don'ts
- Color in Data Vis
- Opinions
- Maps
- Weekly Charts
- Data Vis Dispatch
- Data Vis Book Club
- How others use us

Find a [chronological list of our articles](#), and learn about our newsletters or [about us](#).

Value labels in line charts

August 26-30, 2024

436.8K

428.6K

413.1K

392.4K

391.1K

393.1K

August 19-23, 2024:
Summer break for students of public schools in Washington D.C.

\$5.06

\$2.91

\$2.09

Unless gaso per g \$3.52

Datawrapper Blog

New: Automatically label data points in line charts

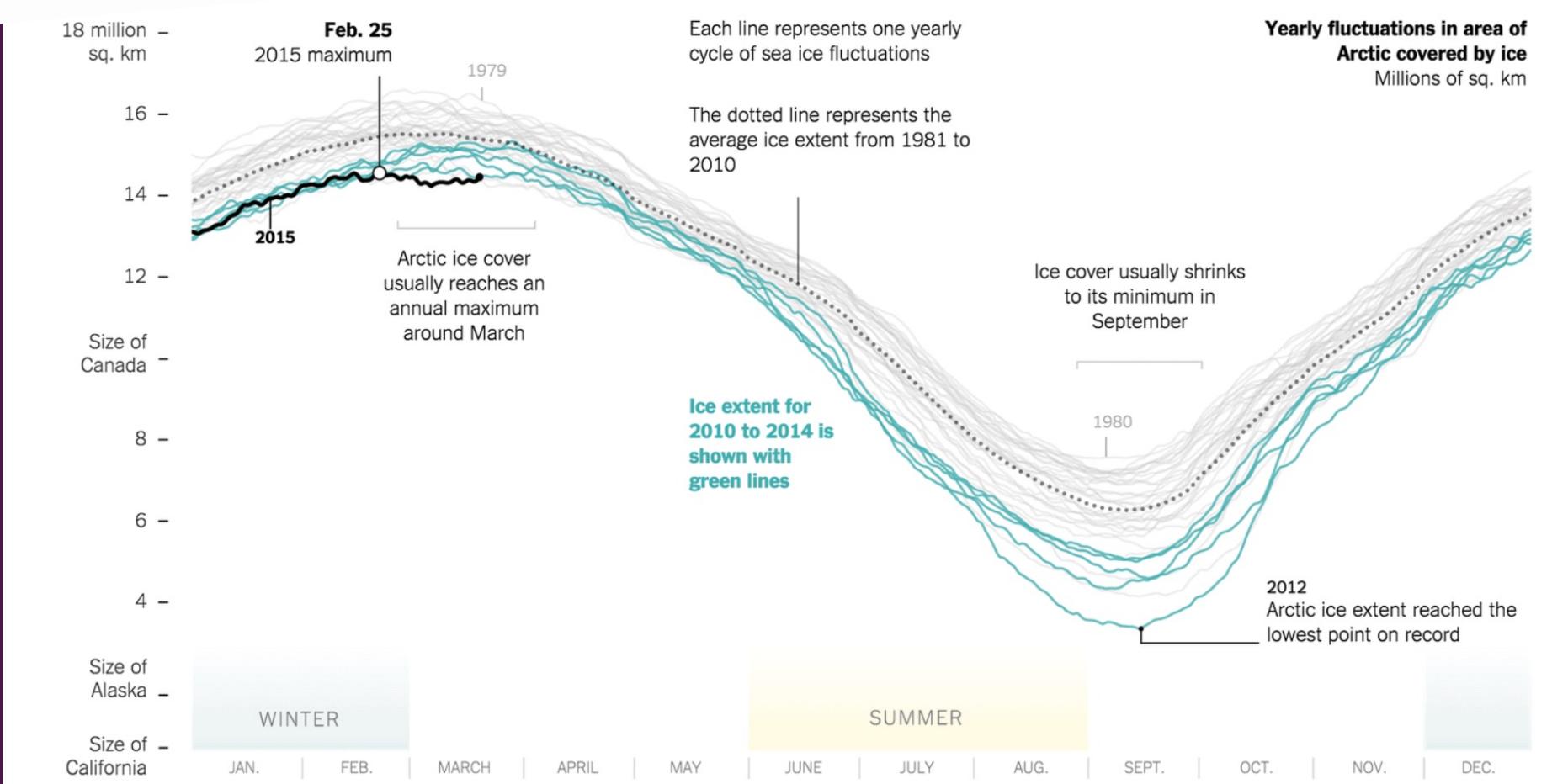
At Datawrapper, we're constantly striving to help you create better data visualizations. A big part...

September 18th, 2024 by Luc Guillemot



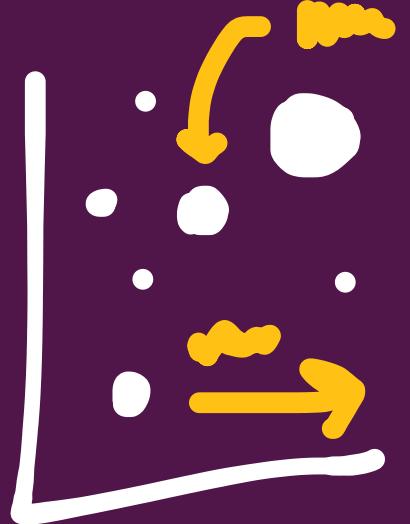
Lisa Charlotte Muth

Sometimes, I give workshops. They are mostly introductions to data visualization. And if you've ever visited one of these workshops, you've seen this chart before:



<https://blog.datawrapper.de/readers-time/>

Why annotate?



For readers: they help them understand the graphic

- We talked about color and position to direct attention
- Annotations do a better job with slightly less elegance

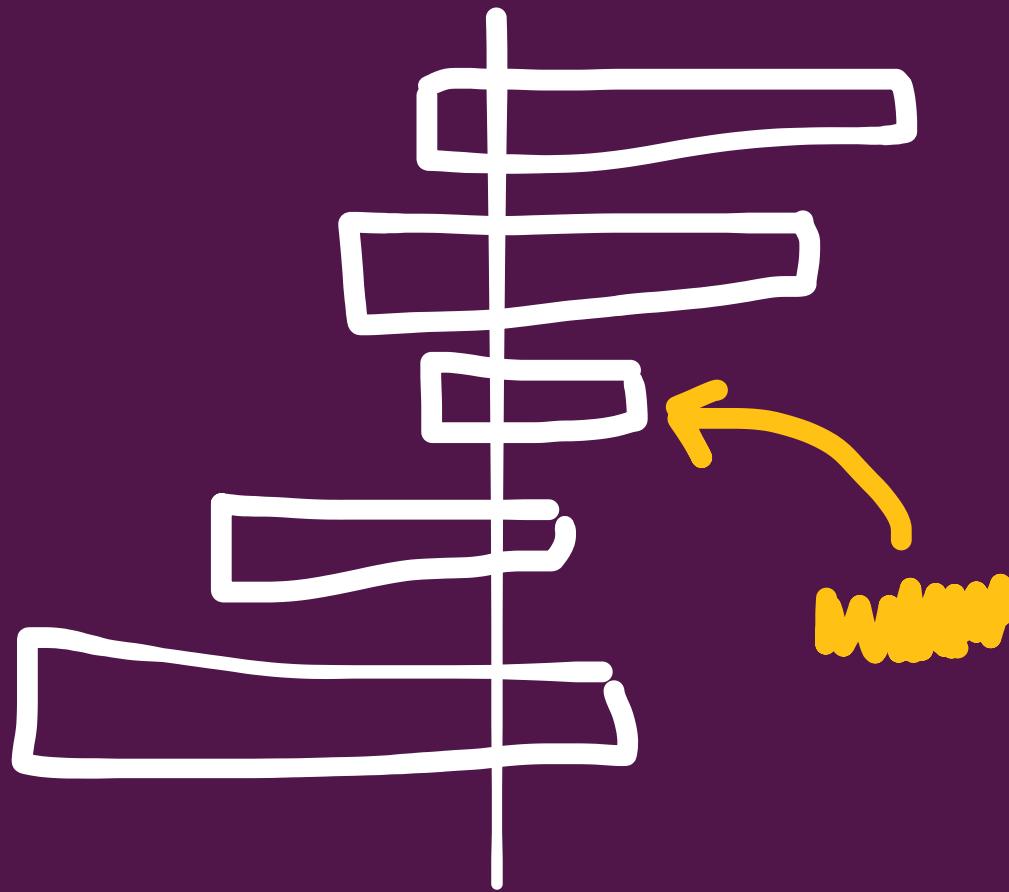
For you: they force you understand the visual a little better

- Just like with titles, you need to determine **what's important**
- There has to be something worth directing attention to

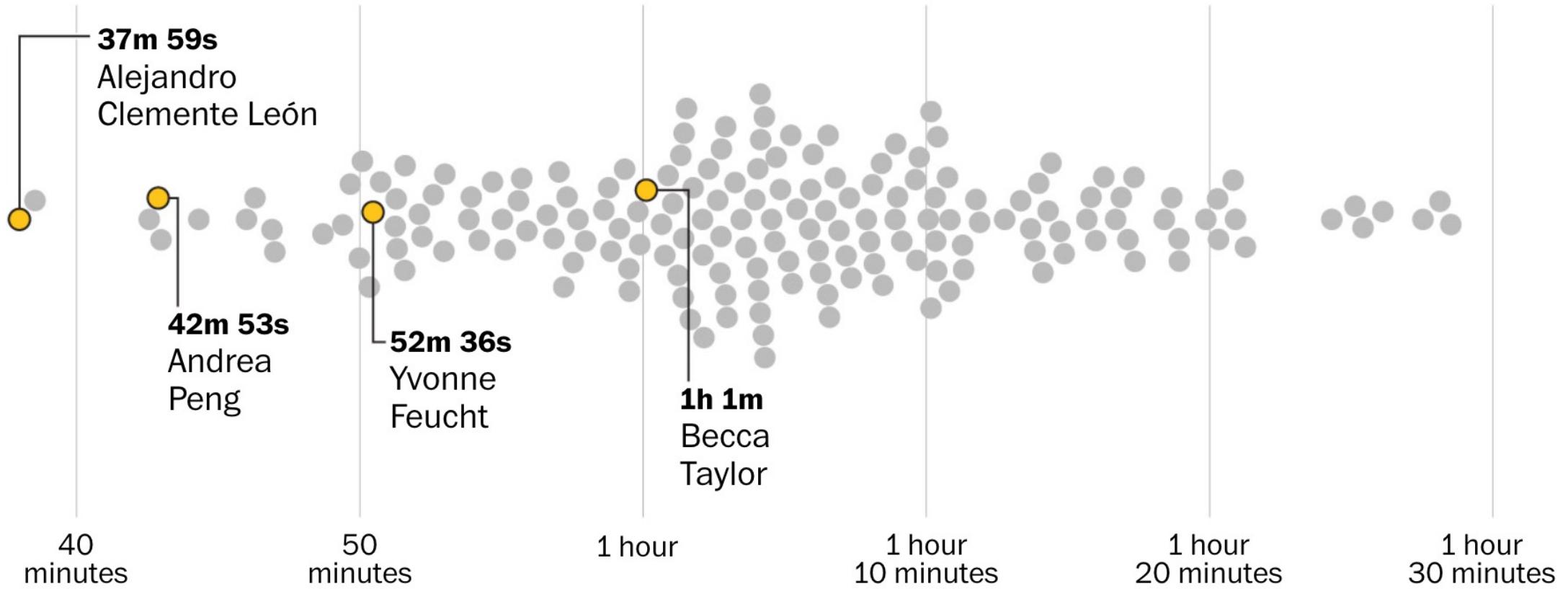
What to annotate?

- **Point out data points** that you want noticed
- Explain why data points **look like they do**
- Explain or support **design decisions**
- **Provide specifics** or follow-up from title

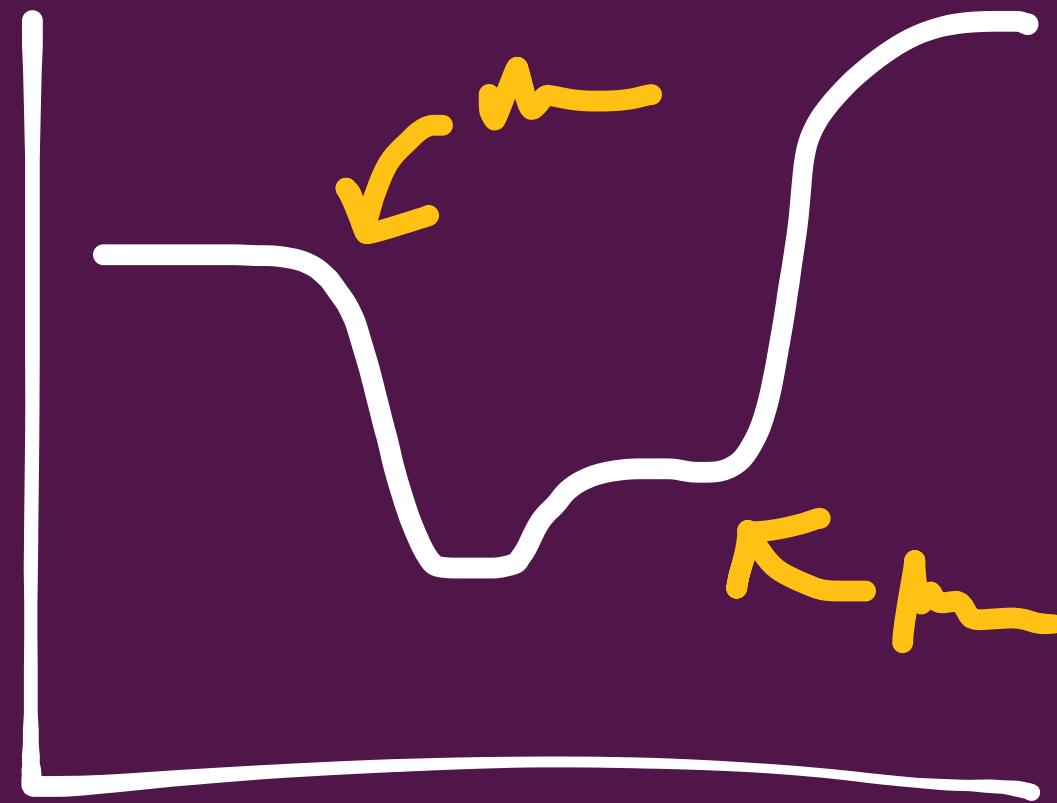
Point out data points you want noticed



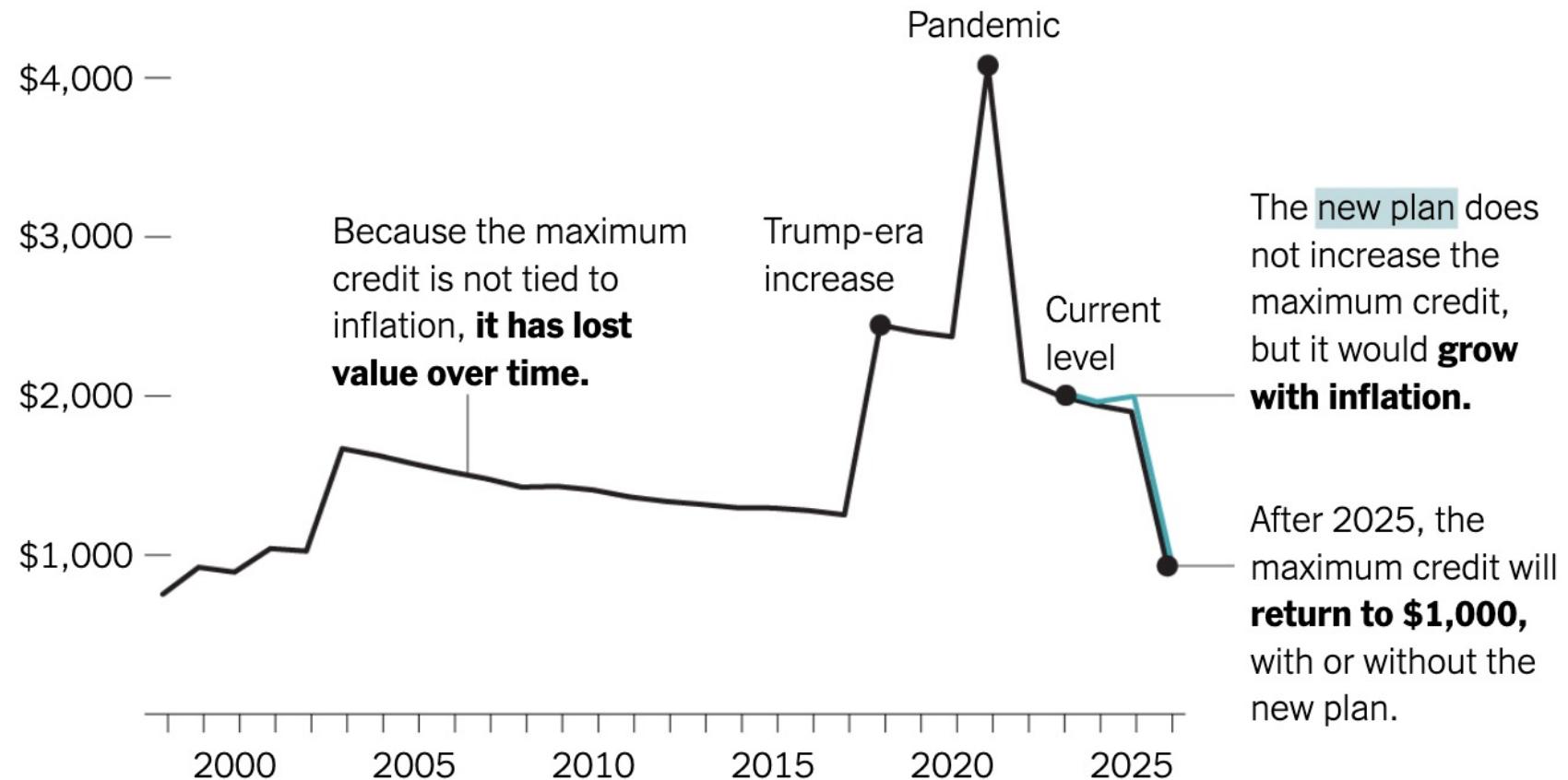
Finishing times for the individual, 500-piece competition
during the 2023 World Jigsaw Puzzle Championship



Explain why the data looks that way



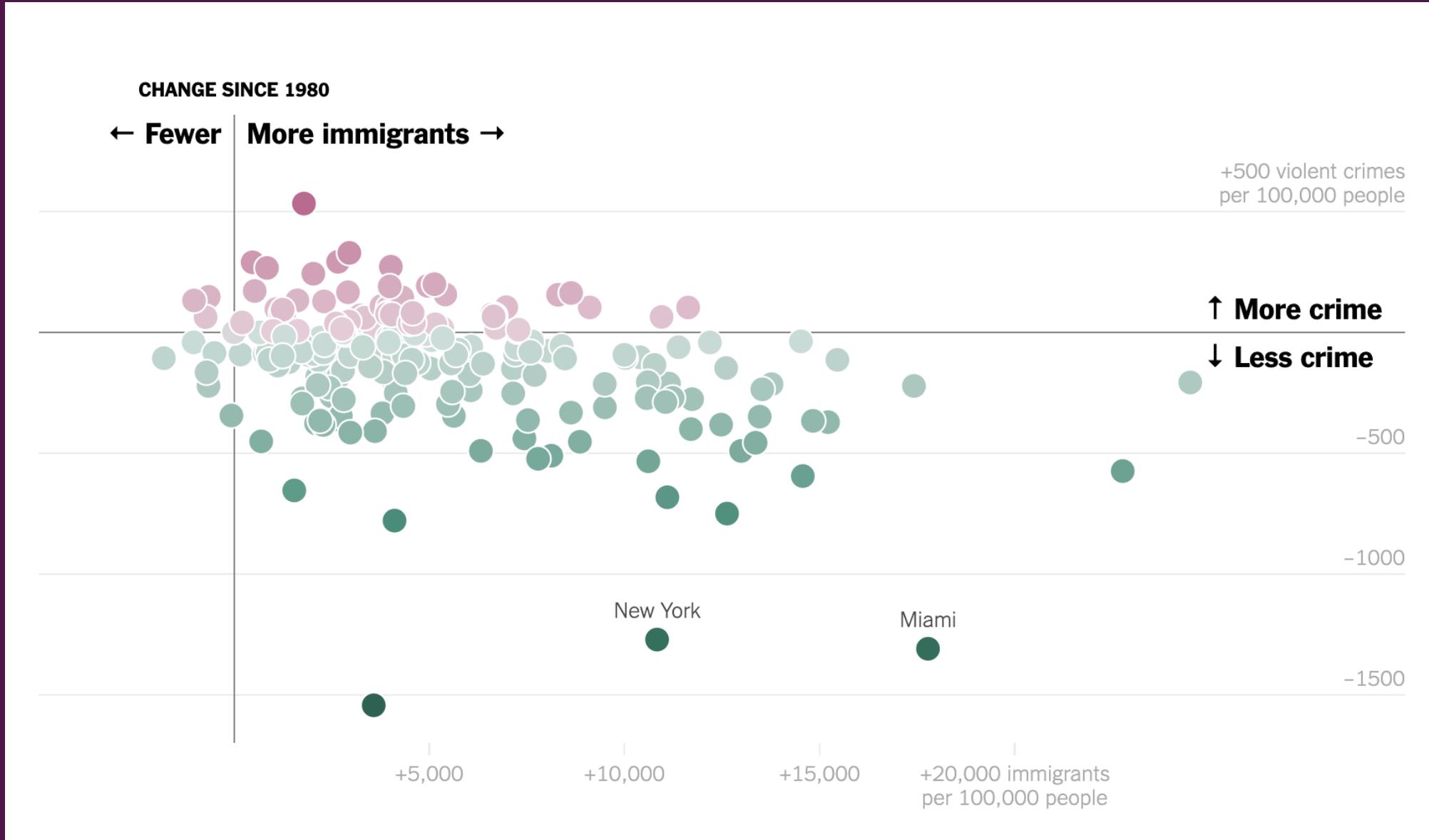
Maximum Child Tax Credit, in 2023 Dollars



Note: The increase shown around the pandemic was for children under six; older children received a slightly lower amount. • Sources: Congressional Research Service; Federal Reserve Bank of St. Louis

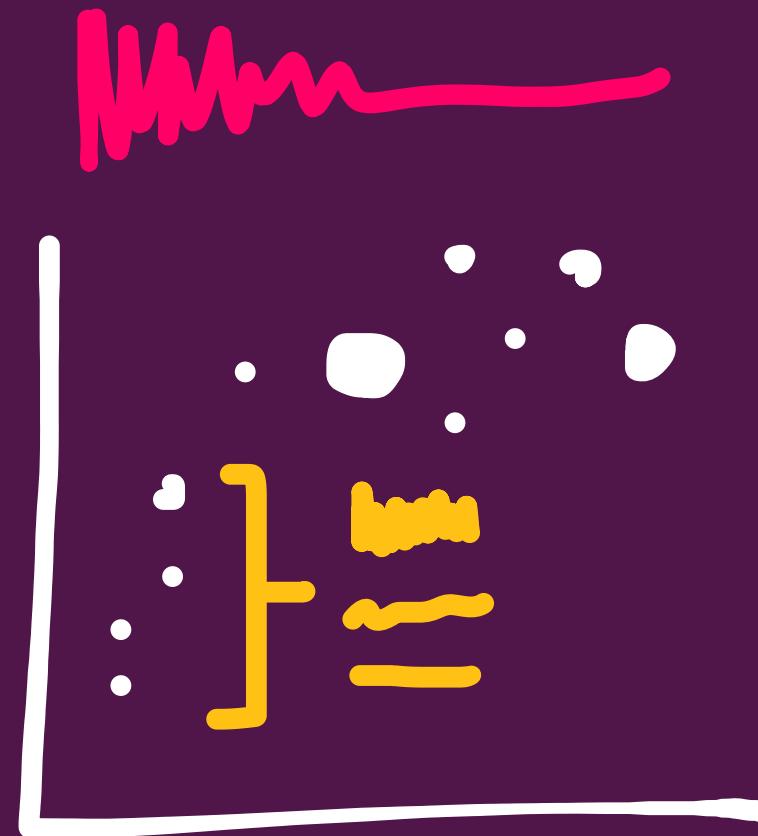
Explain or support design decisions



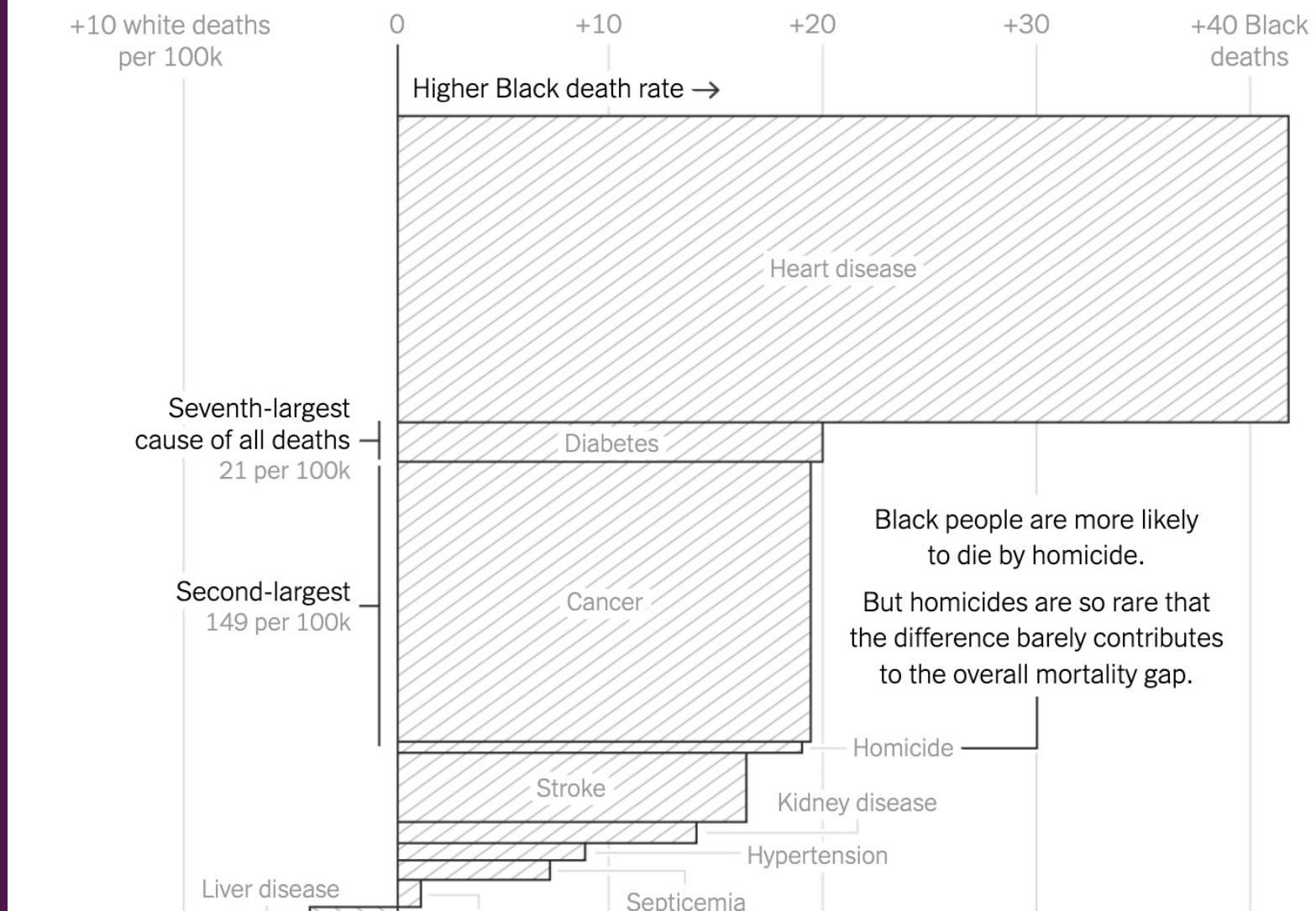


<https://www.nytimes.com/interactive/2018/03/30/upshot/crime-immigration-myth.html>

Provide specifics or follow-up from title



Gaps between Black and white mortality rates for the top 15 causes of death



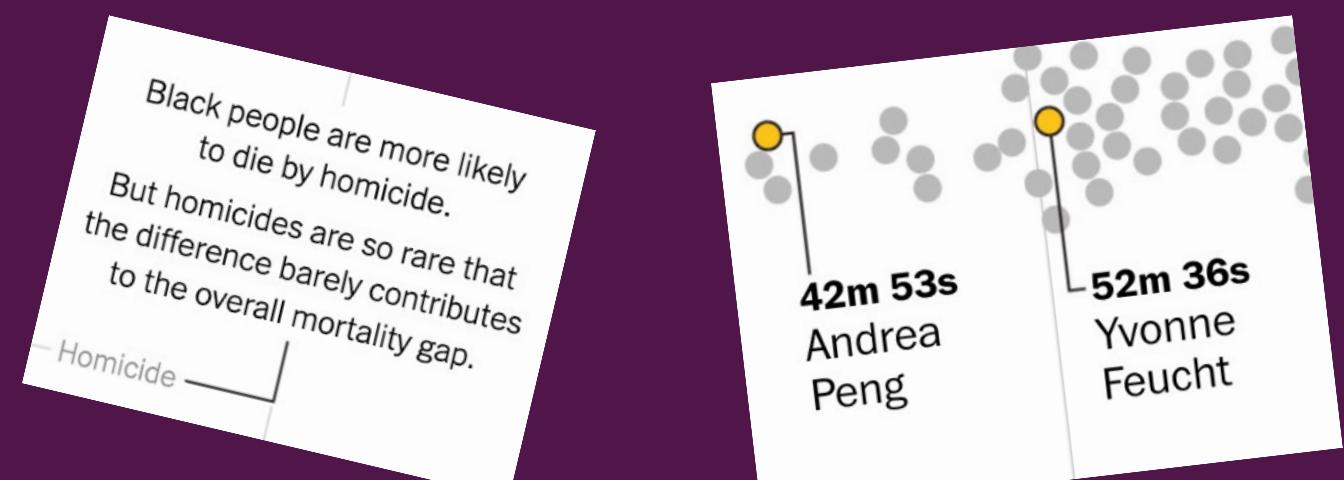
“what that is” vs “why that is”

Here's *what* this data point is: **Texas**

Here's *why* this data point is here: **Texas has a lot of cattle farms**

A little bit of both: **Texas, 4lb**

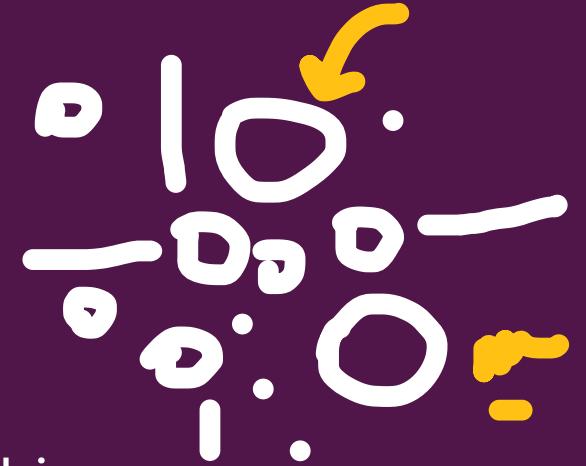
You're deciding between **narrative** or **data-driven** (either is fine!)



When to annotate

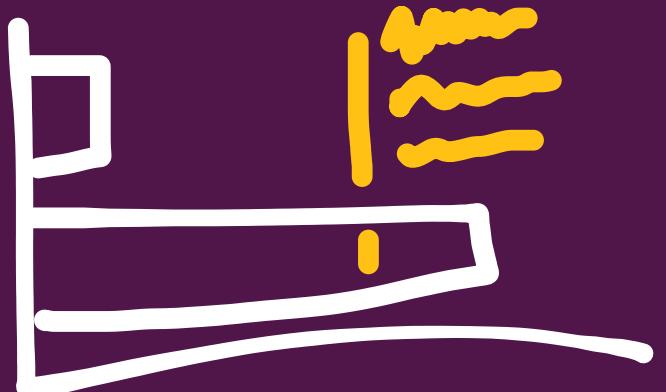
Complicated graphics?

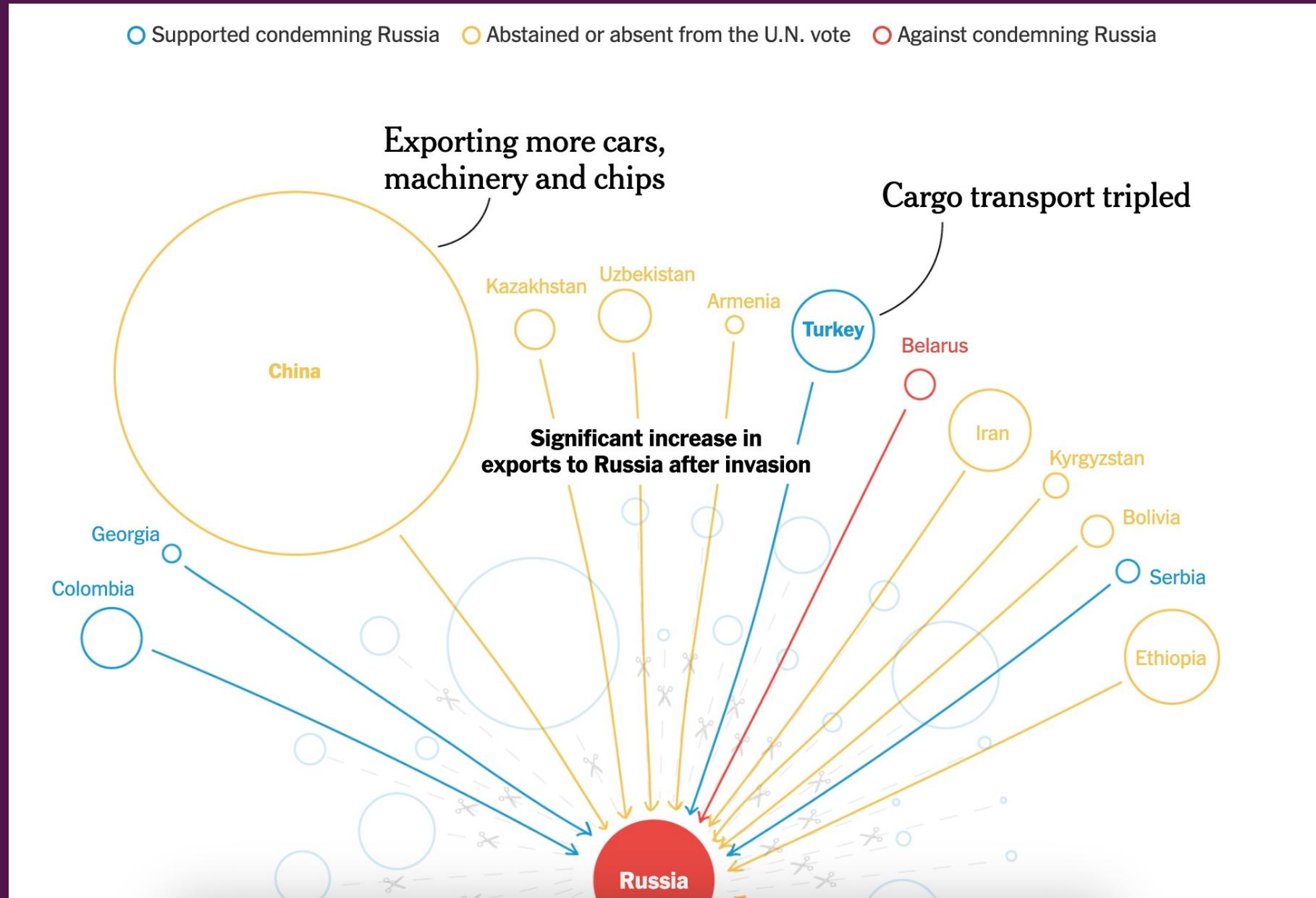
- No, you should probably just split them up
- Annotations are *not* an excuse for lack of curation
- I bet every one of your visuals could be 2-3 smaller graphics
- But sure, okay, annotate it, it'll help



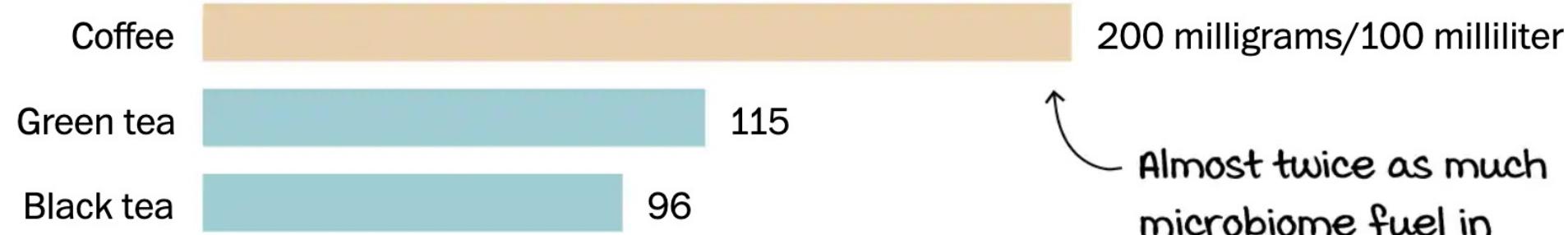
Simple graphics?

- Honestly, you can annotate anything
- Provides a “thought about” or “lived in” a feel





Polyphenol concentration



How to annotate

Flourish: Use the pencil icon - [details here](#)

Datawrapper: Annotations tab – [details here](#)

- **Bonus:** you can also use them to add measurement types for axes

Altair: `mark_text` (potentially with `transform_filter`)

A few links at <https://jonathansoma.com/everything/visuals/annotations/>

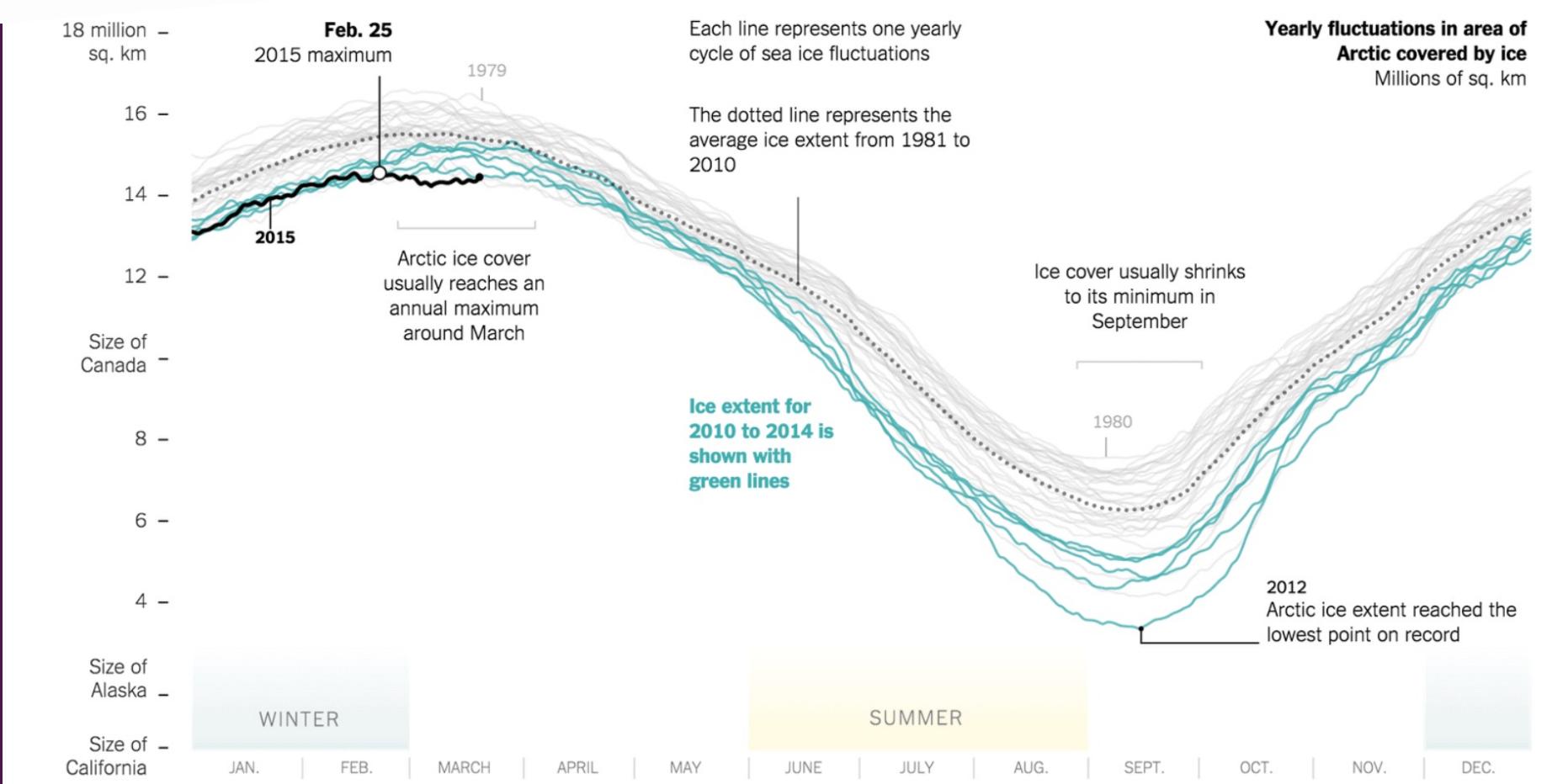
Bonus wisdom

Annotations are the primary reason why
you move away from tooling

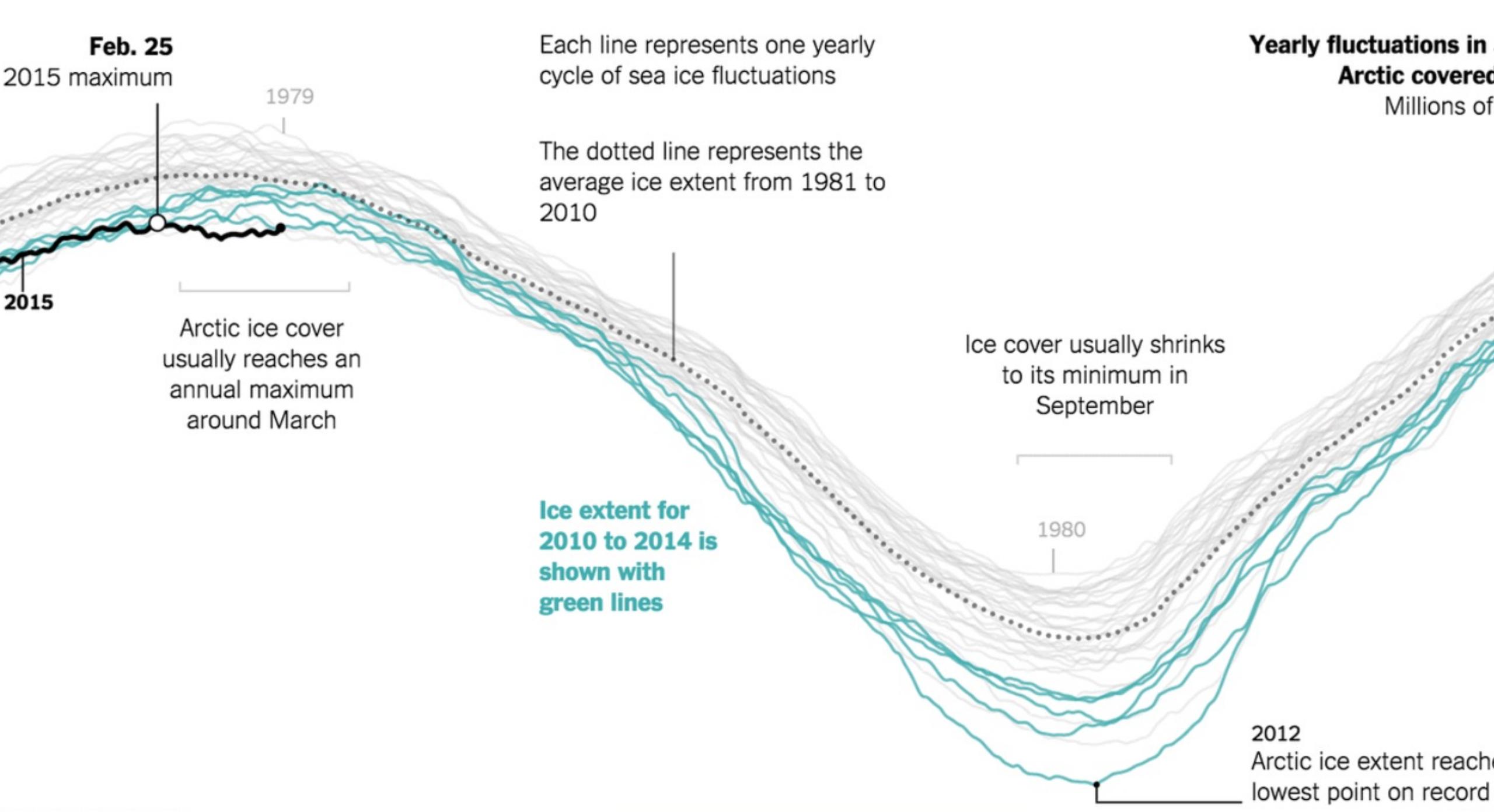


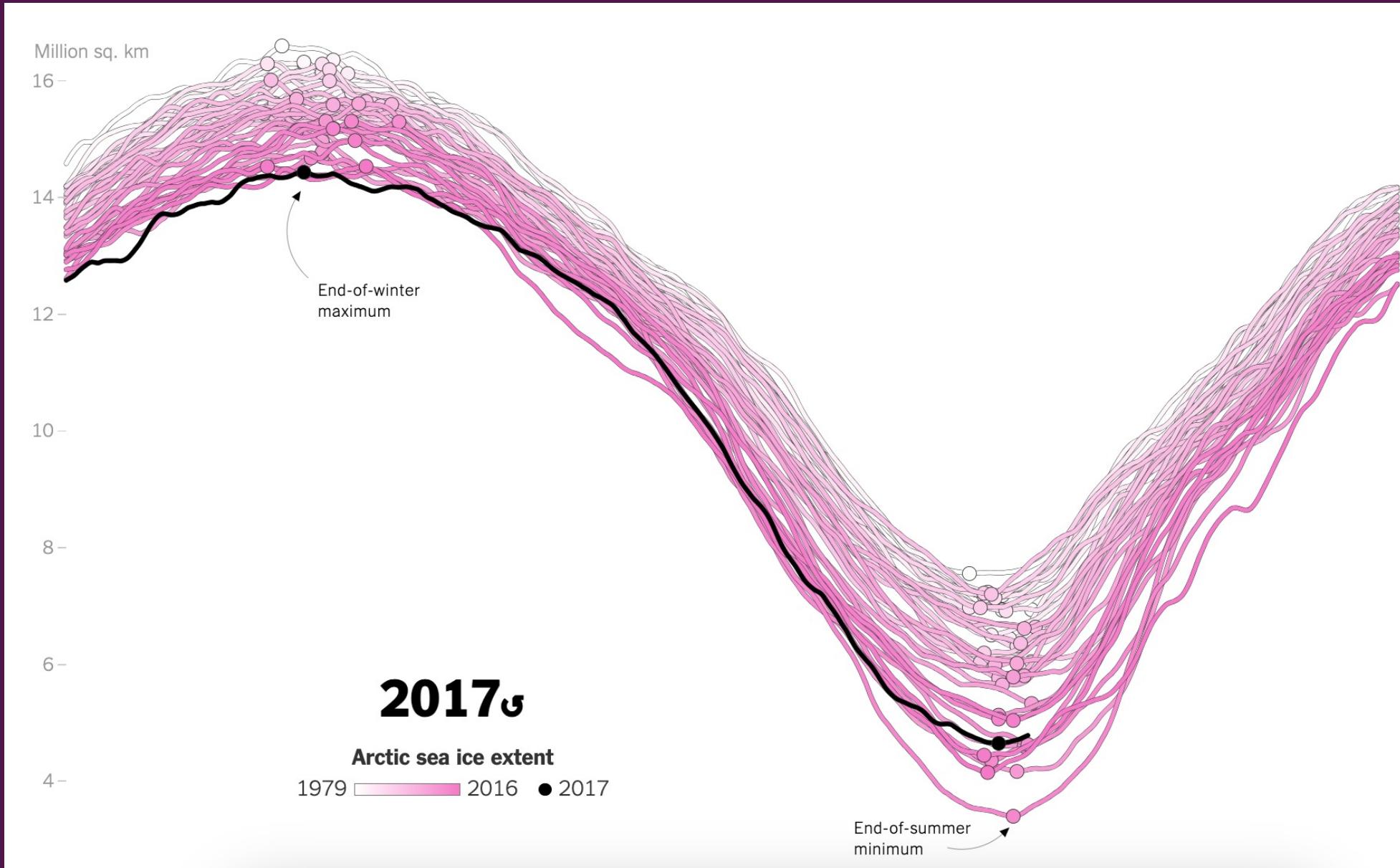
Lisa Charlotte Muth

Sometimes, I give workshops. They are mostly introductions to data visualization. And if you've ever visited one of these workshops, you've seen this chart before:

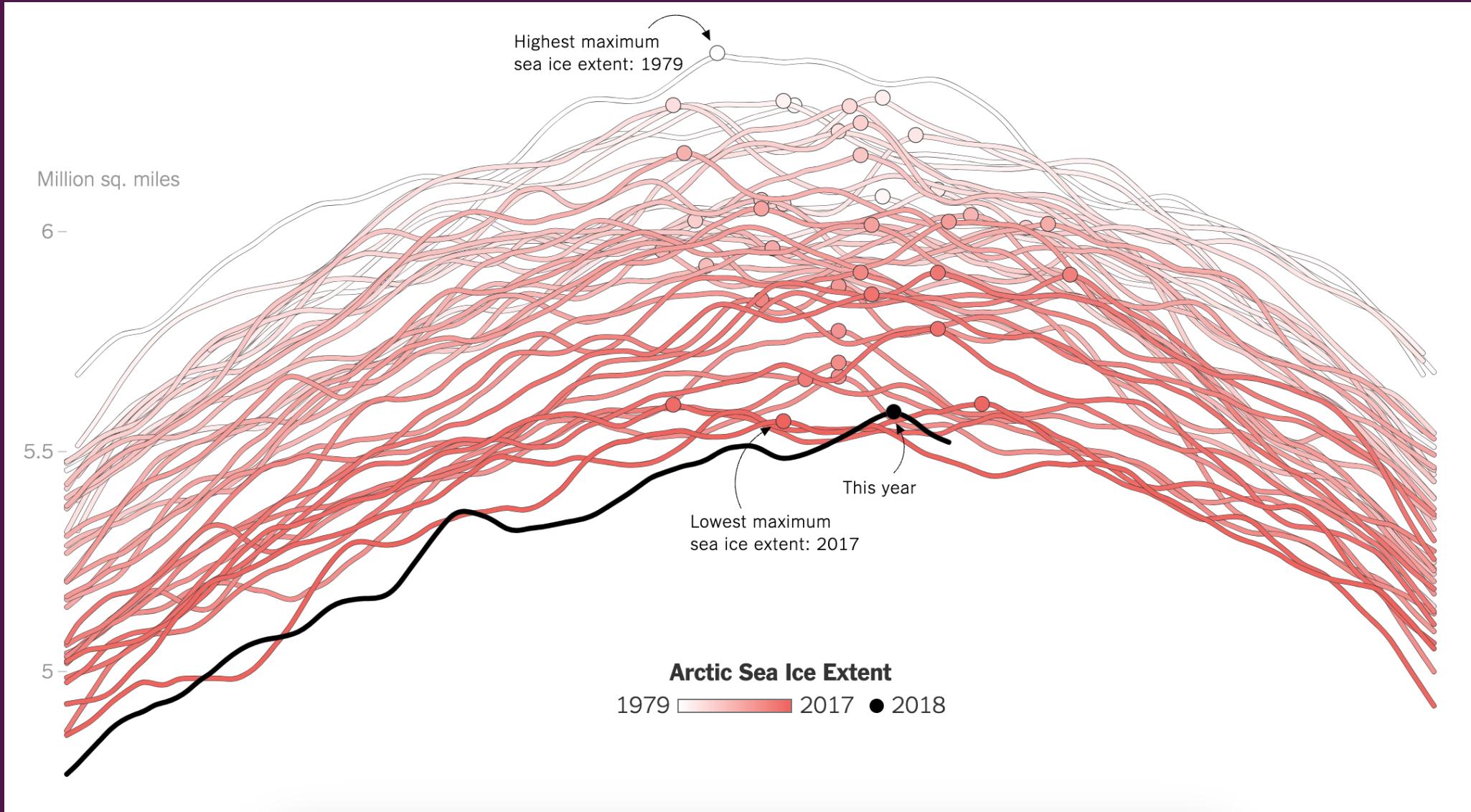


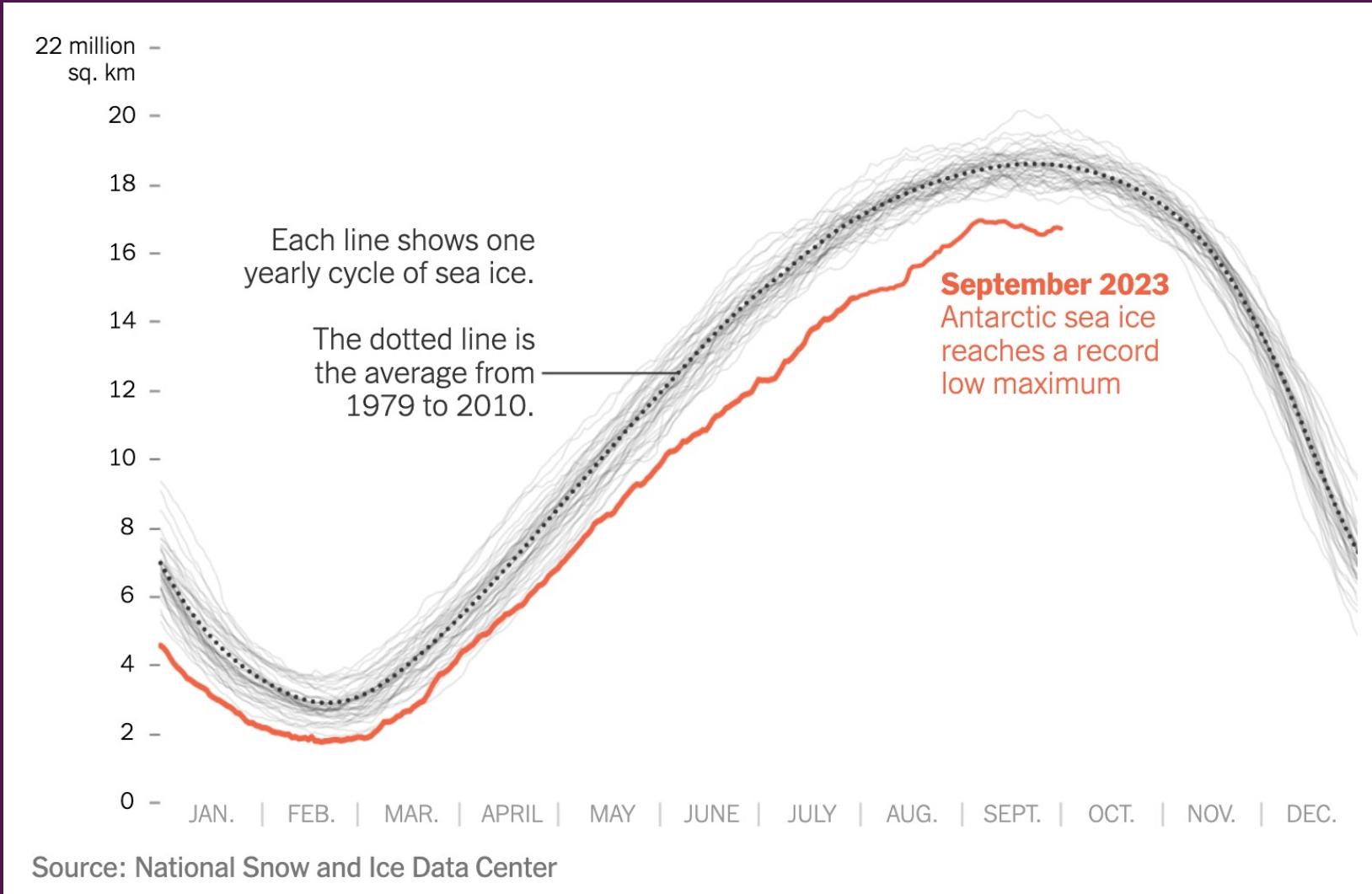
<https://blog.datawrapper.de/readers-time/>





<https://www.nytimes.com/interactive/2017/09/22/climate/arctic-sea-ice-shrinking-trend-watch.html>





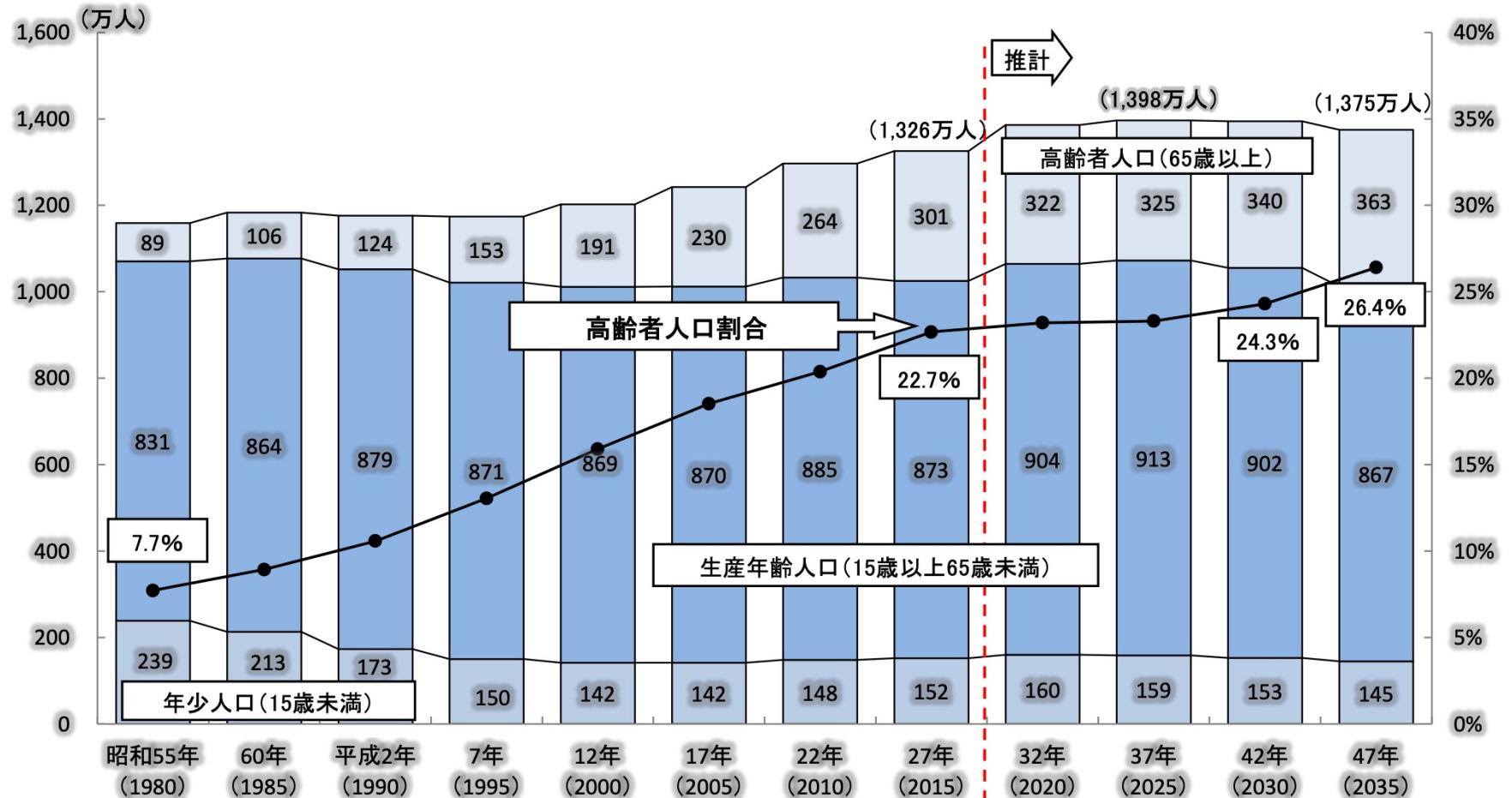
today's topics



- Why annotate
- What to annotate
- When to annotate
- How to annotate

人口の推移(東京都)

東京都における高齢化率は、総人口がピークを迎える平成37年には23.3%であり、平成42年には24.3%とおよそ4人に1人が高齢者になると推計されます。



(注) ()内は総人口。1万人未満を四捨五入しているため、内訳の合計値と一致しない場合がある。

出典: 総務省「国勢調査」[昭和55年～平成27年]、東京都政策企画局による推計[平成32年～47年]

What would you do?

Visual Vocabulary

Deviation

Emphasises outliers (x) from a fixed reference point. Typically the reference point is zero but it can also be a target or a long-term average. Can also be used to show variance (positive/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar
A single standard bar chart that can handle both negative and positive magnitude values.

Diverging stacked bar
Perfect for presenting survey results which involve a balance (eg disagree/neither/agree).

Spine
Shows a single value into two components (eg male/female).

Surplus/deficit flow
The shaded area of these charts allows a balance to be shown – between a baseline or between two series.

Correlation

Show the relationship between two variables. Be mindful that unless you tell them otherwise, many readers will assume the relationships you show them are causal (i.e. one causes the other).

Example FT uses
Inflation and unemployment, income and life expectancy

Scatterplot
The standard way to show the relationship between two continuous variables, each of which has its own axis.

Columns + line timeline
A good way of showing the relationship between an amount (column) and a rate (line).

Connected scatterplot
Usually used to show how the relationship between 2 variables has changed over time.

Bubble
Like a scatterplot, but adds additional detail by sizing the circles according to a third variable.

XY heatmap
A good way of showing the patterns between 2 variables. Very effective at showing fine differences in amounts.

Ranking

Use where an ordering makes more sense than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar
Standard bar charts display the ranks of values much more easily when sorted into order.

Ordered column
See above.

Ordered proportional symbol
Used for showing big variations between values and/or seeing fine differences when data is not so important.

Dot strip plot
Dots placed in order on a strip are a space-efficient way of showing out ranks across multiple categories.

Slope
Perfect for showing how ranks have changed over time or very between categories.

Lollipop
Lollipops draw more attention to the data value than standard bar/column and can also be used to show values effectively.

Bump
Effective for showing changing rankings across multiple dates. For large numbers, consider grouping lines using colour.

Distribution

Show whether the data is ordered and how often they occur. The shape (or 'tail') of a distribution can be a memorable way of highlighting the lack of uniformity or equality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram
The standard way to show a statistical distribution - keep gaps between bins small to highlight the 'shape' of the data.

Dot plot
A simple way of showing the change or growth of data across multiple categories.

Dot strip plot
Good for showing individual value in a distribution, can be a problem when too many dots have the same value.

Barcode plot
Like a dot strip plot, good for displaying all the data in a table. This works best when highlighting individual values.

Boxplot
Summarise multiple distributions by their median (median) and range of the data.

Violin plot
Similar to a boxplot but more effective with complex distributions (data that cannot be summarised with simple averages).

Population pyramid
A standard way for showing the age and sex breakdown of a population distribution; effectively back to back histograms.

Cumulative curve
A good way of showing how unequal a distribution is. If x axis is always cumulative frequency, x axis is always a measure.

Frequency polygons
For displaying multiple distributions of data. Like a regular line chart, best limited to a maximum of 3 or 4 datasets.

Beeswarm
Use to emphasise individual points in a distribution. Points can be sized to an appropriate scale. Best with medium-sized datasets.

Change over Time

Show how things have changed over time. These can be short time steps (movements or extended series traversing decades or centuries). Usually these show a 'cumulative' time period (which is important to provide context for the reader).

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line
The standard way to show a changing time series. If data are continuous, consider markers to represent date points.

Column
Columns work well for showing change over time – but are best with only one series of data at a time.

Column + line timeline
A good way of showing the relationship over time between an amount (column) and a rate (line).

Slope
Good for showing changing data as long as the data can be visualised with 3 or 4 points without missing a key part of the story.

Area chart
Use with care – these are good at showing trends but showing change in components can be very difficult.

Candlestick
Usually focused on day-to-day activity, these charts show opening/closing and high/low points of each day.

Fan chart (projection)
Used to show the uncertainty in future projections – usually this grows the further forward to projection.

Connected scatterplot
A good way of showing changing data for two variables whenever there is a very clear pattern of progression.

Calendar heatmap
A great way of showing temporal patterns (daily, monthly) – at the expense of showing precision in quantity.

Priestley timeline
Great when date and duration are key elements of the story in the data.

Circle timeline
Good for showing date and duration of events.

Radar
A space-efficient way of showing multiple variables – but make sure they are organised in a way that makes sense to the reader.

Parallel coordinate
An alternative to radar charts – again, the arrangement of variables is important. Useful for highlighting values.

Circle timeline
Good for showing date and duration of events.

Bullet
Good for showing % information, they work best when used on a single axis and work well in small multiple layer form.

Magnitude

Show how magnitudes can be relative (not being able to see the difference) or absolute (able to see the difference). Usually these show a 'cumulative' time period (which is important to provide context for the reader).

Example FT uses
Commodity production, market capitalisation, volumes in general

Stacked column/bar
A simple way of showing part-to-whole relationships but can be difficult to read with more than a few components.

Marimekko
A good way of showing the total and proportions of data at the same time – as long as the data are not too complicated.

Paired column
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series.

Bar
See above. Good when the data are not time related and have long category names.

Pie
A common way of showing part-to-whole data – but be aware that it's difficult to accurately compare the size of the segments.

Donut
Similar to a pie chart – but the centre can be a good way of making specific data more apparent about the data (eg total).

Treemap
Use for hierarchical part-to-whole data – it can be difficult to read when there are many small segments.

Proportional symbol
Use when there are big variations between values and/or seeing fine differences that data is not so important.

Isotype (pictogram)
Excellent solution in some instances – use only with whole numbers (do not slice off a decimal).

Lollipop
Lollipop charts draw attention to the data value than standard bar/column – does not have to start at zero (but preferable).

Veronoi
A way of turning data into grid – any point within each area is closer to the central point than any other centroid.

Arc
A tessellate, often used for visualising parliamentary composition by number of seats.

Gridplot
Good for showing % information, they work best when used on a single axis and work well in small multiple layer form.

Heat map
Grid-based data values mapped with an appropriate colour scale. As shown here – but not suited to an admin/political unit.

Venn
Generally only used for schematic representation.

Waterfall
Can be useful for showing part-to-whole relationships where some of the components are negative.

Spatial

Show how reader values differently of movement between two or more places or conditions. These might be logical sequences or geographical locations.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact; catchment areas, variation in election results

Basic choropleth (rate/ratio)
A simple approach for putting data on a map – should always be aware of what the base geography.

Proportional symbol count (magnitude)
Use for totals rather than rates – be wary of the base process, typically budgets. Can include +/- components.

Waterfall
Designed to show the sequencing of data through a flow process, typically budget. Can include +/- components.

Chord
A complex network diagram which can illustrate 2-way flows (and net winners) in a matrix.

Contour map
For showing areas of equal value on a map. Can use deviation colour schemes for showing +/- values.

Equalised cartogram
Converting each unit on a map to a regular and equal size – good for representing voting regions with equal value.

Scaled cartogram
Stretching and shrinking areas so that one area is scaled according to a particular value.

Dot density
Used to show the location of individual events/locations – remember to annotate any patterns the reader should see.

Heat map

designingviz.com

Data Visualization Design Gui x + ⌂

designingviz.com Data Viz Design Guide Contribute on GitHub

A Step-by-Step Design Guide for Data Visualizations



Use these simple checklists to turn terrible visualizations into **stunningly adequate** ones.

Axes

- Using full-width grid or tick lines
- Backgrounding grid or tick lines
- Thick baselines
- Selecting a lower bound
- Selecting an upper bound
- Placing your annotations

Color

- Make unimportant things grey
- Choosing colors

Legends

- Rounding off numbers

Bars

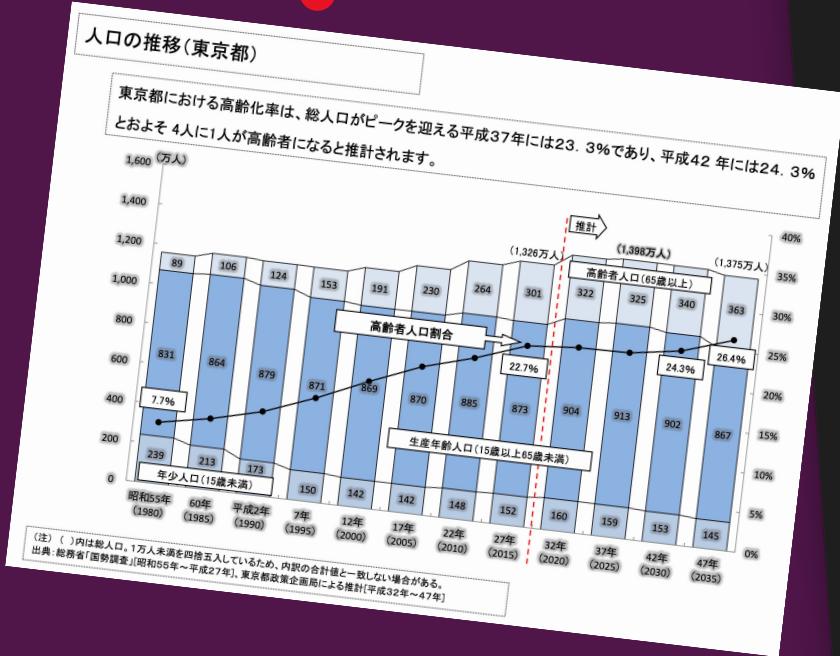
- Removing borders on bars
- Ordering bars
- Don't use diagonal labels
- Thinning out columns
- Extra-wide, 100% columns

Interactivity

- Planning for lazy users
- Tell a story
- Tooltip directions
- Layering lines
- Dense circles
- Close enough should be good enough

Lines

- Using a line graph



ChatGPT 4o

JS

Find the decade that a photo is from

Experience Seoul like a local

Make a recommendation based on my data

Thank my interviewer

Message ChatGPT

ChatGPT can make mistakes. Check important info.

✖️ ⓘ Tokyo_Population_Projection.csv

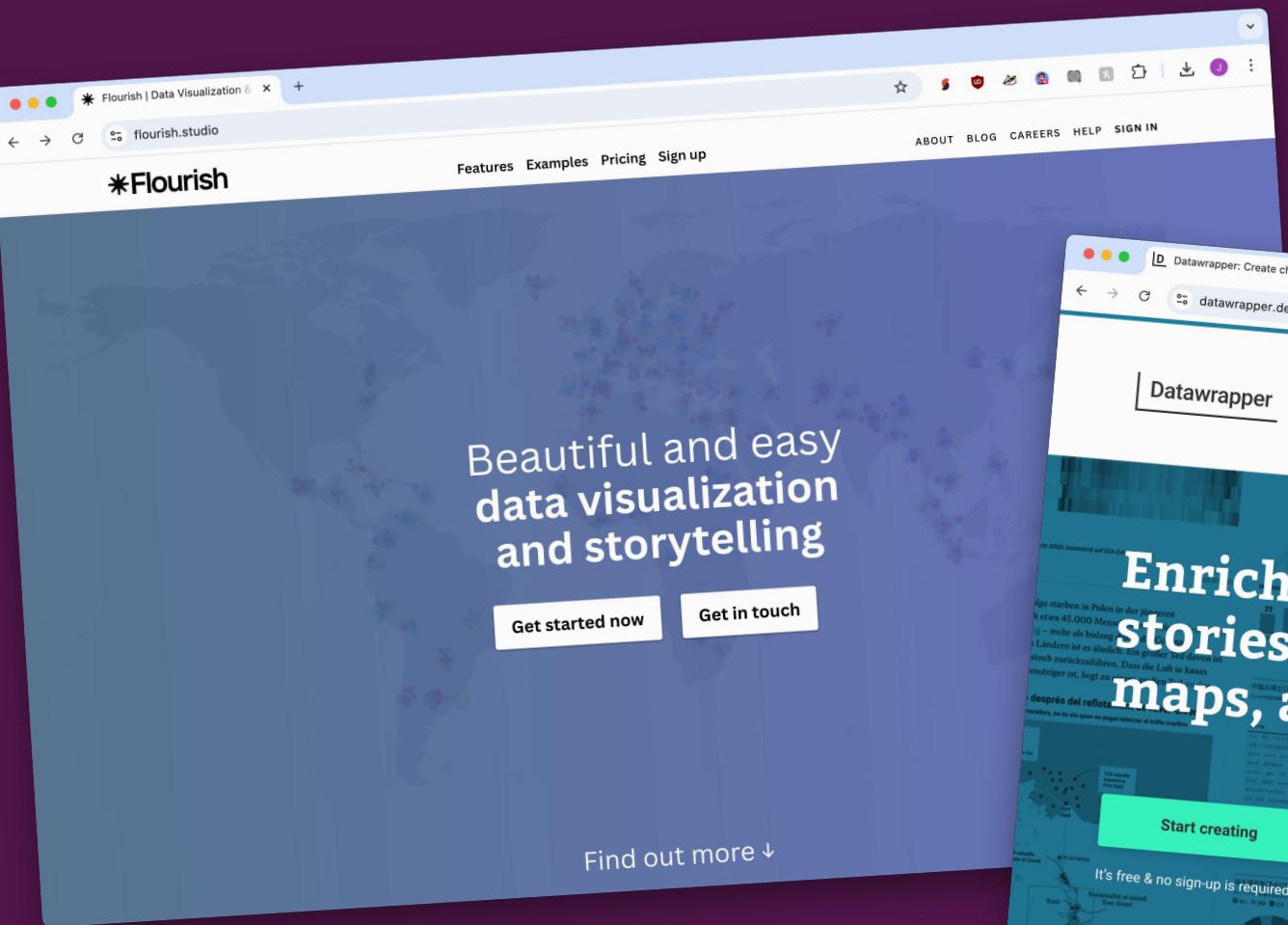
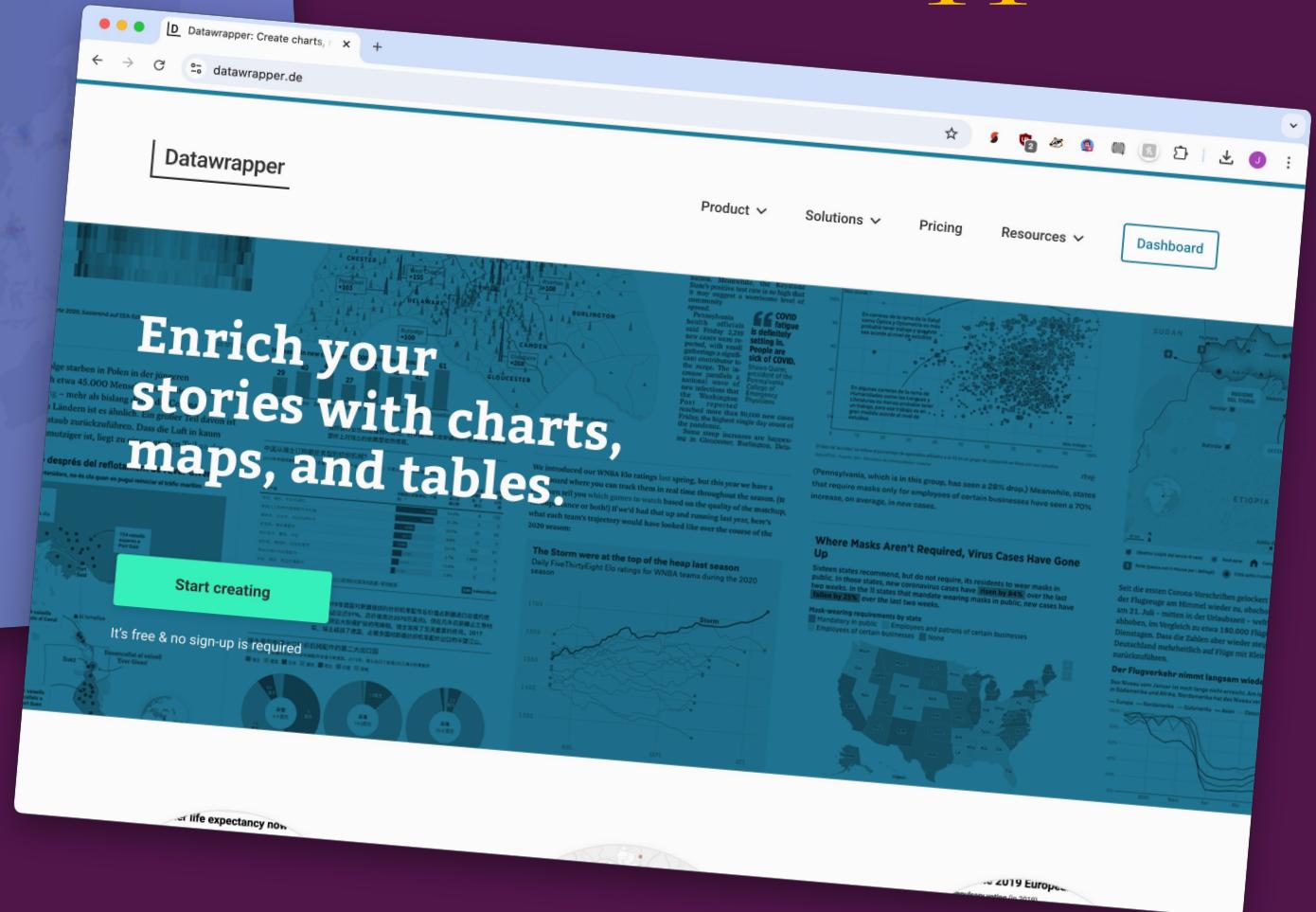


Open with Numbers

Year	Young Population (万人)	Working-age Population (万人)	Elderly Population (万人)	Total Population (万人)	Elderly Population %
1980	239	831	104	1174	7.7
1990	213	816	126	1155	10.9
2000	173	785	154	1112	13.9
2010	150	710	230	1090	21.1
2015	140	664	264	1068	24.7
2020	124	640	322	1086	29.7
2025	145	602	325	1072	30.3
2030	140	570	340	1050	32.4
2035	137	536	353	1026	34.4
2040	125	513	360	998	36.1
2045	114	495	363	972	37.4

Copy the image into ChatGPT,
and ask for a table of data

Datawrapper



Flourish

① Visit <https://bit.ly/ds-dojo-2024>
Click **Data Visualization**

② Download the **PDF data** file

Wednesday

- [Data analysis \(pandas\) and APIs](#)
- [Scraping and more pandas](#)

Thursday

- [Data Visualization, PDFs and AI](#)

Friday

Troubleshooting and project work time

Monday

Project presentations

About the instructor

Session content

Code and data

Please download these files:

- [Visualization data](#)
- [PDF data](#)
- [More PDFs for research](#)

Slides

- [Visualization](#)
- [AI](#)