Data Analytics Foundations

Module 3: Data visualization

DeepLearning.AI

Data visualization

Module 3 introduction

DeepLearning.AI

- **Data storytelling**
  - Transform raw data into stories
  - Visualization techniques
- **Creating charts**
  - Create common chart types
- **Best practices in data visualization**
  - Choose the right chart
  - Colors, markers, titles, axes, annotations
  - Keep visualizations efficient
- **Data visualization with LLMs**
  - Interpret charts
  - Create charts

Data visualization

What is data storytelling?

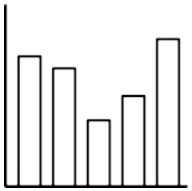
DeepLearning.AI

What is data storytelling?

Results of your analysis

Descriptive statistics

+



Data visualization

→

Meaningful insights

Compelling narrative

Well-visualized data stories can be: 💬 Thought-provoking 🏆 Powerful 😤 Emotional

Key components of data storytelling

Business problem

You're telling this story for a **reason**

Data

Dictates what **kinds of stories** you can tell

Analysis

Extracts **insights** from the data

Visualization

Way you **present data visually** to audience

Data storytelling using descriptive statistics

Example 1️⃣

In the last 150 years, life expectancy has more than doubled from 32 to 71 years in 2021.

Example 2️⃣

Approximately 13% of the US population aged 5 and over speaks Spanish.

Both tell an **interesting** and **complete** data story.

Data storytelling using descriptive statistics

Example 1️⃣

In the last 150 years, life expectancy has more than doubled from 32 to 71 years in 2021.

32 to **71**

in 2021

Example 2️⃣

Approximately

13% 💬

of the US population aged 5 and over speaks Spanish.

Both tell an **interesting** and **complete** data story.

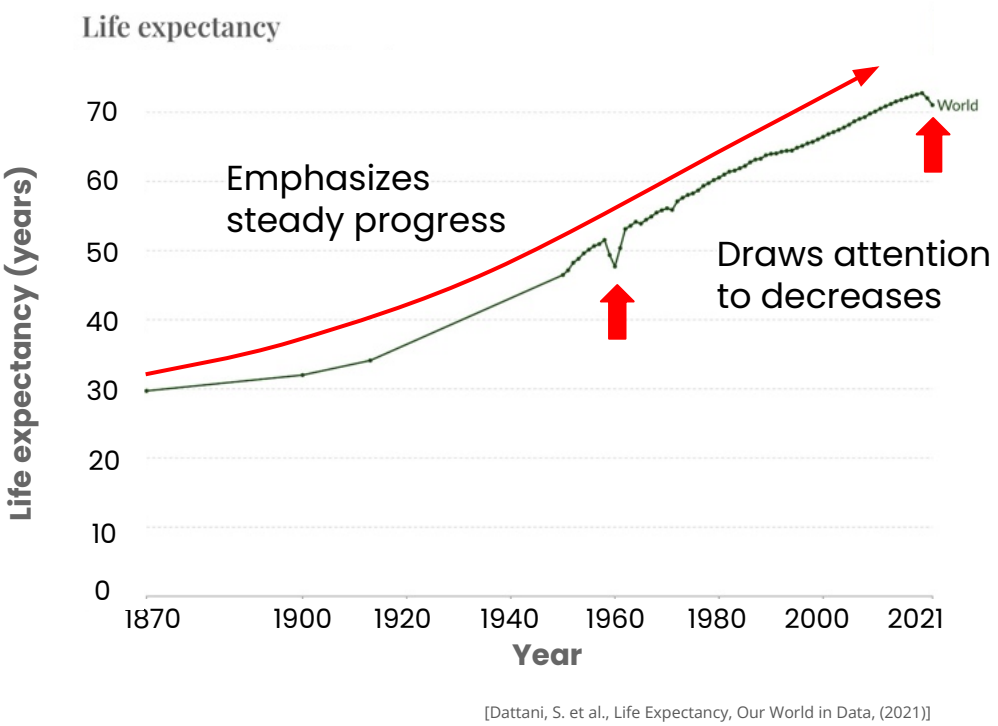
Combining descriptive statistics with visualizations

Example 1️⃣

In the last 150 years, life expectancy has more than doubled from

32 to **71**

in 2021



[Dattani, S. et al., Life Expectancy, Our World in Data, (2021)]

This combination:

- Provides context
- Helps audience grasp key insights

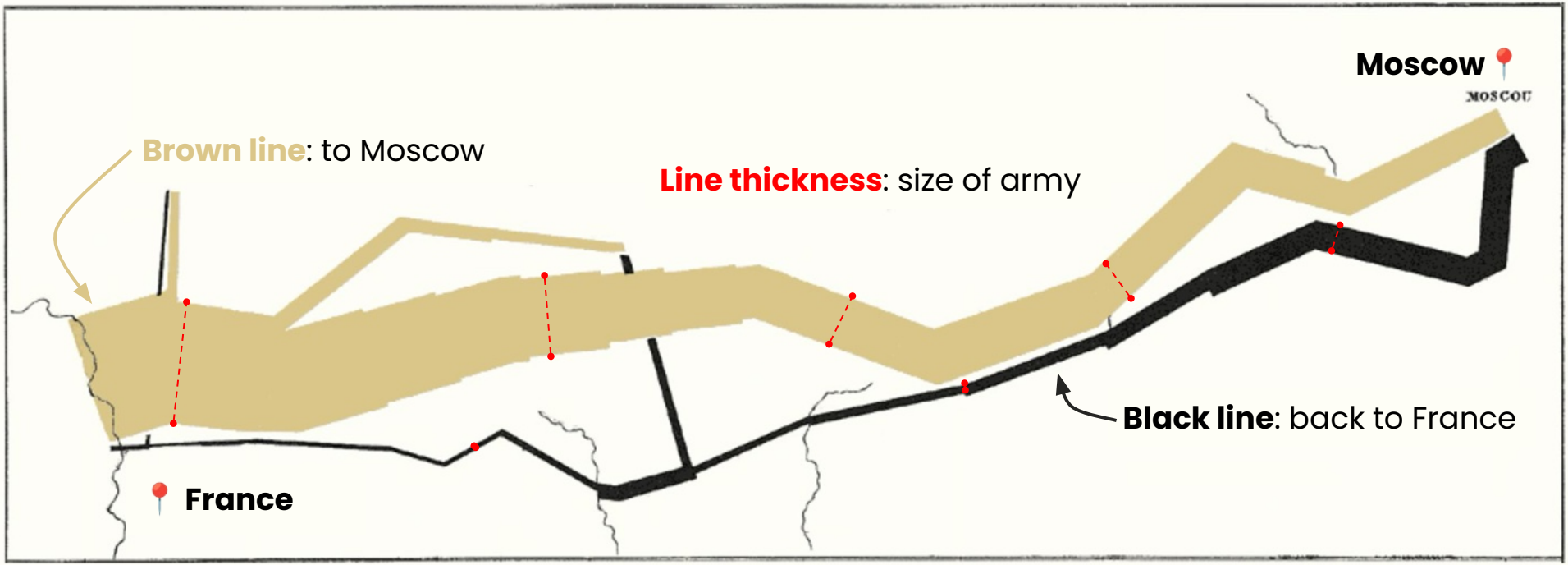
Communicating to an audience

Audience	Level of polish
External audiences	Polished visualizations
Internal audiences	Rougher, exploratory visualizations

In this course, we'll focus primarily on:

📖 Storytelling aspect

🎨 Design elements for data narratives



🎯 **Goal**: To tell the story of Napoleon's Russian campaign during the War of 1812

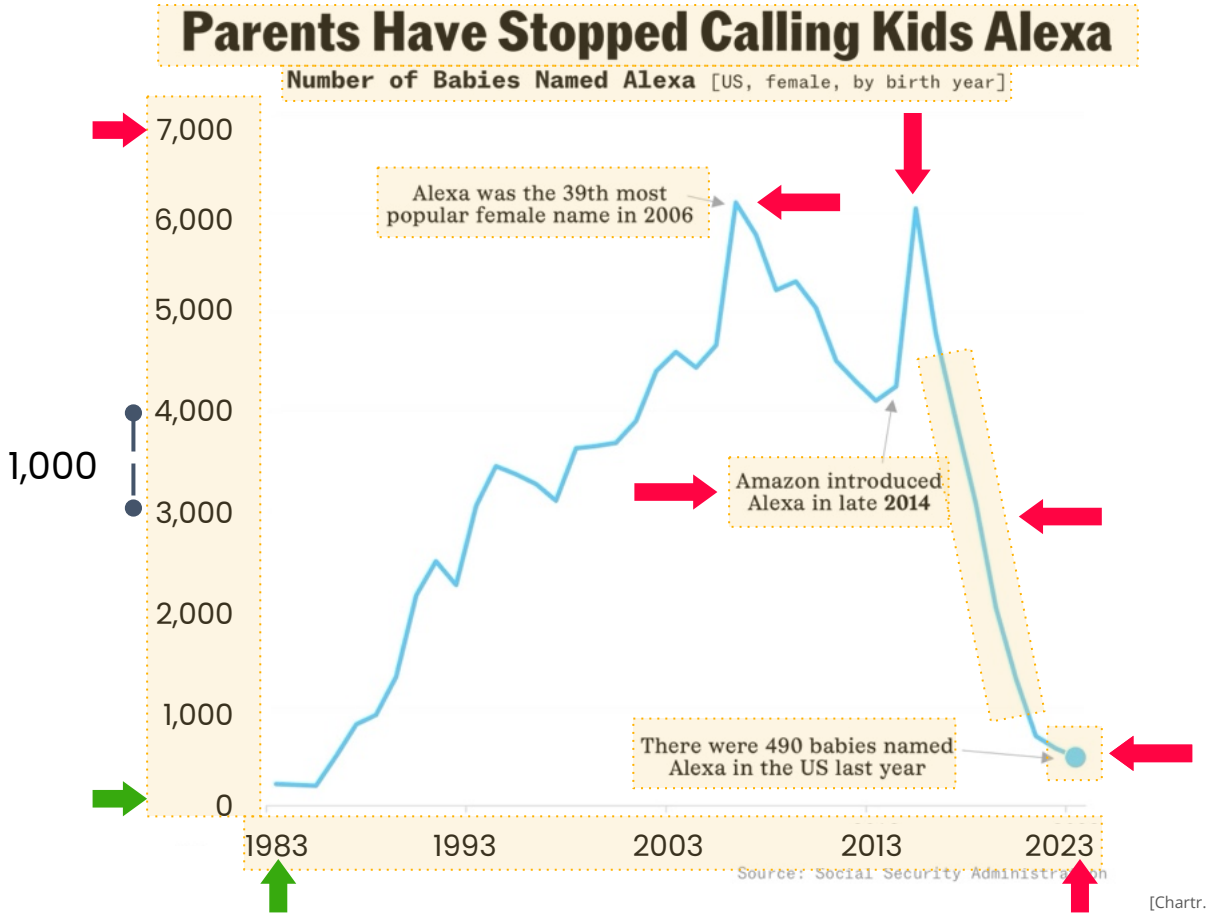
- Map
- Line chart

[Minard, C, Carte figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie (1812–1813)]

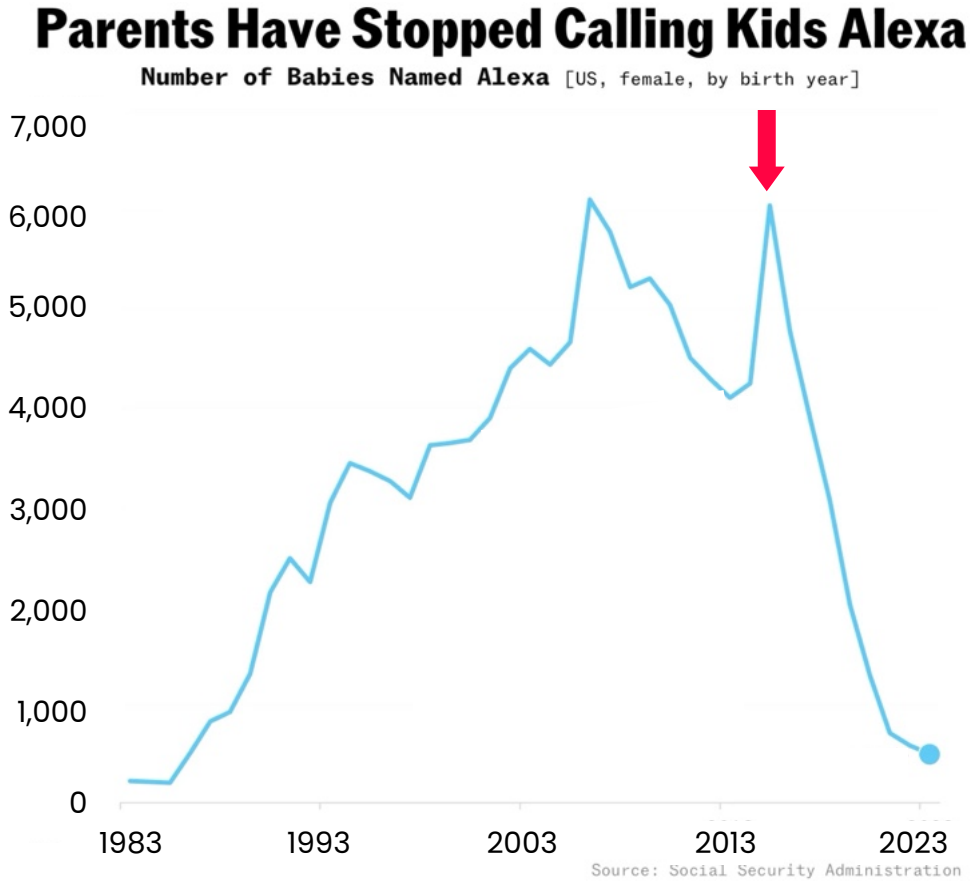
Data visualization

The language of
data visualizations

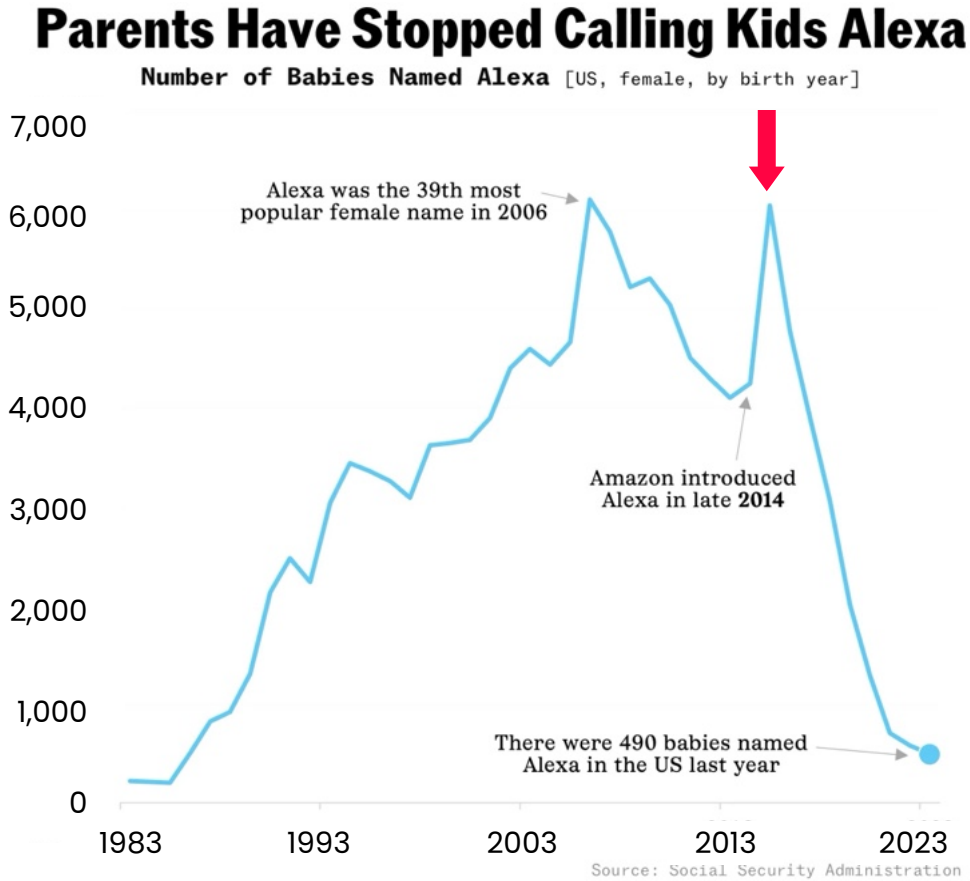
DeepLearning.AI



[Chartr.co., The rise and fall of the name Alexa (2024)]

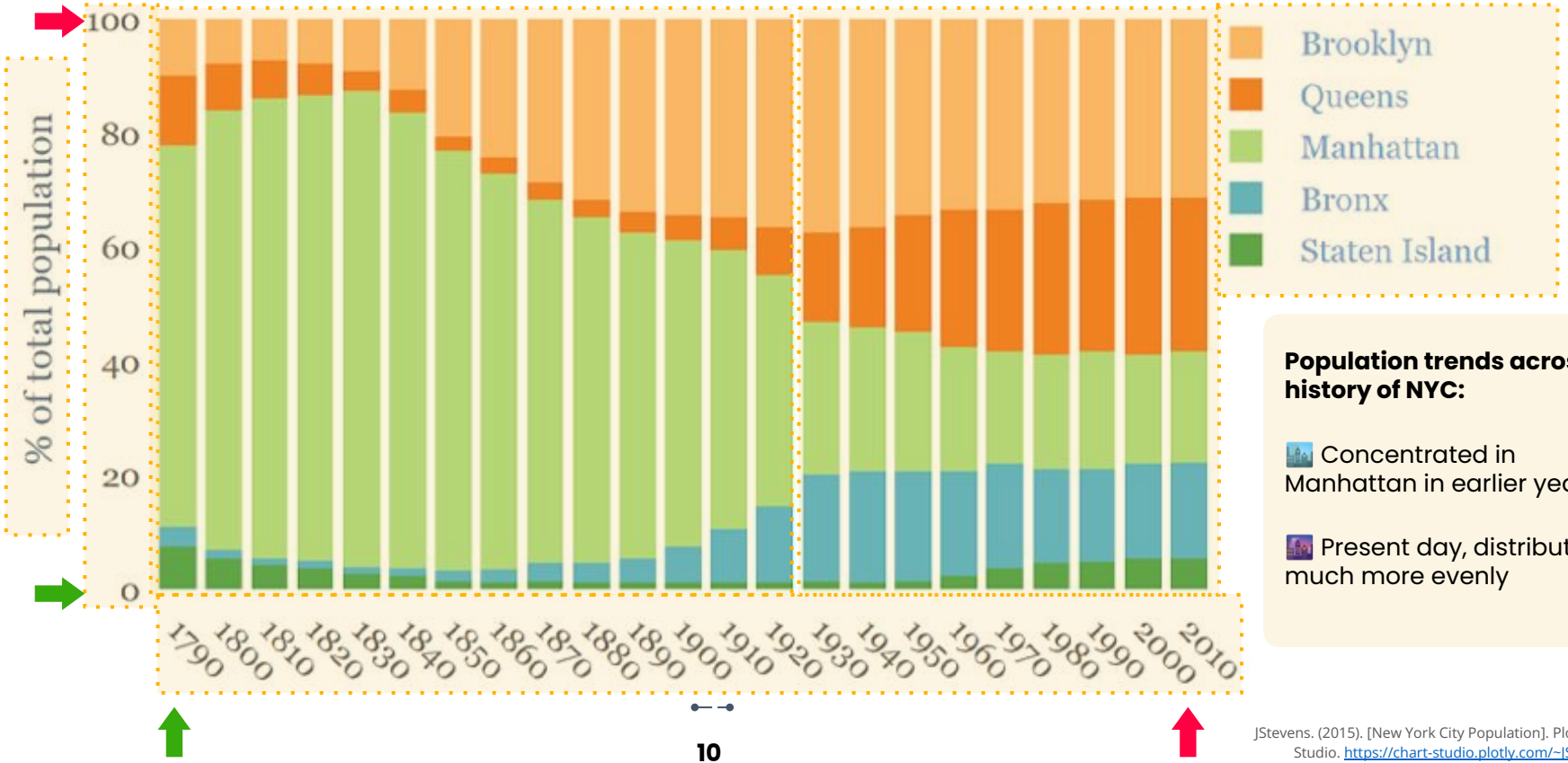


[Chartr.co., The rise and fall of the name Alexa (2024)]



[Chartr.co., The rise and fall of the name Alexa (2024)]

City of New York & Boroughs Population



Population trends across history of NYC:

🏙️ Concentrated in Manhattan in earlier years

🌆 Present day, distributed much more evenly

JStevens. (2015). [New York City Population]. Plotly Chart Studio. https://chart-studio.plotly.com/~JStevens/1

Data encoding

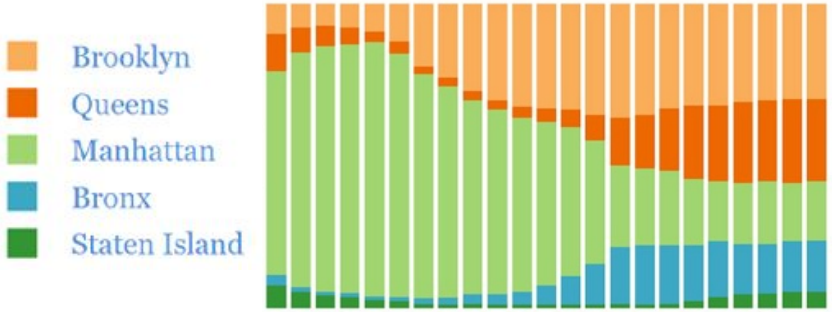
Color

Size

Markers

Encoding means translating data into visual properties.

Examples



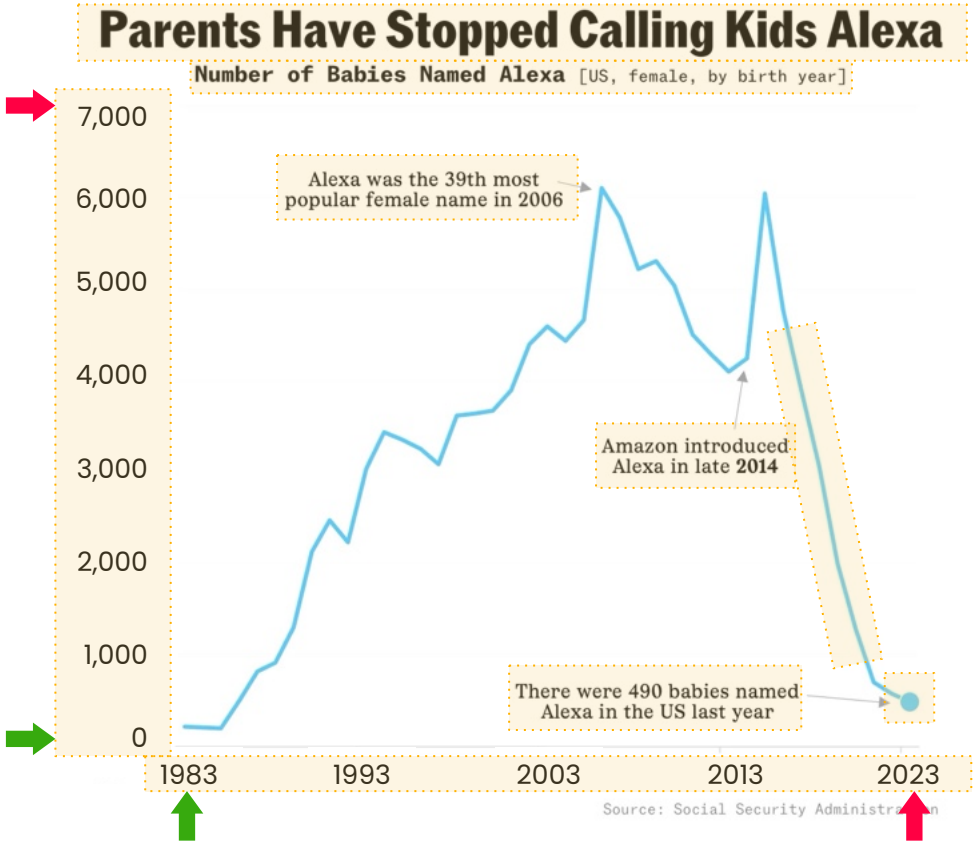
▲ iPhone — *Encoded* category

■ Android

Legends communicate how data has been *encoded*

A structured approach

1. Title
2. Axes: ↔ x-axis, ↕ y-axis
3. Encoded categories
4. Annotations
5. Big picture



Sean Barnes

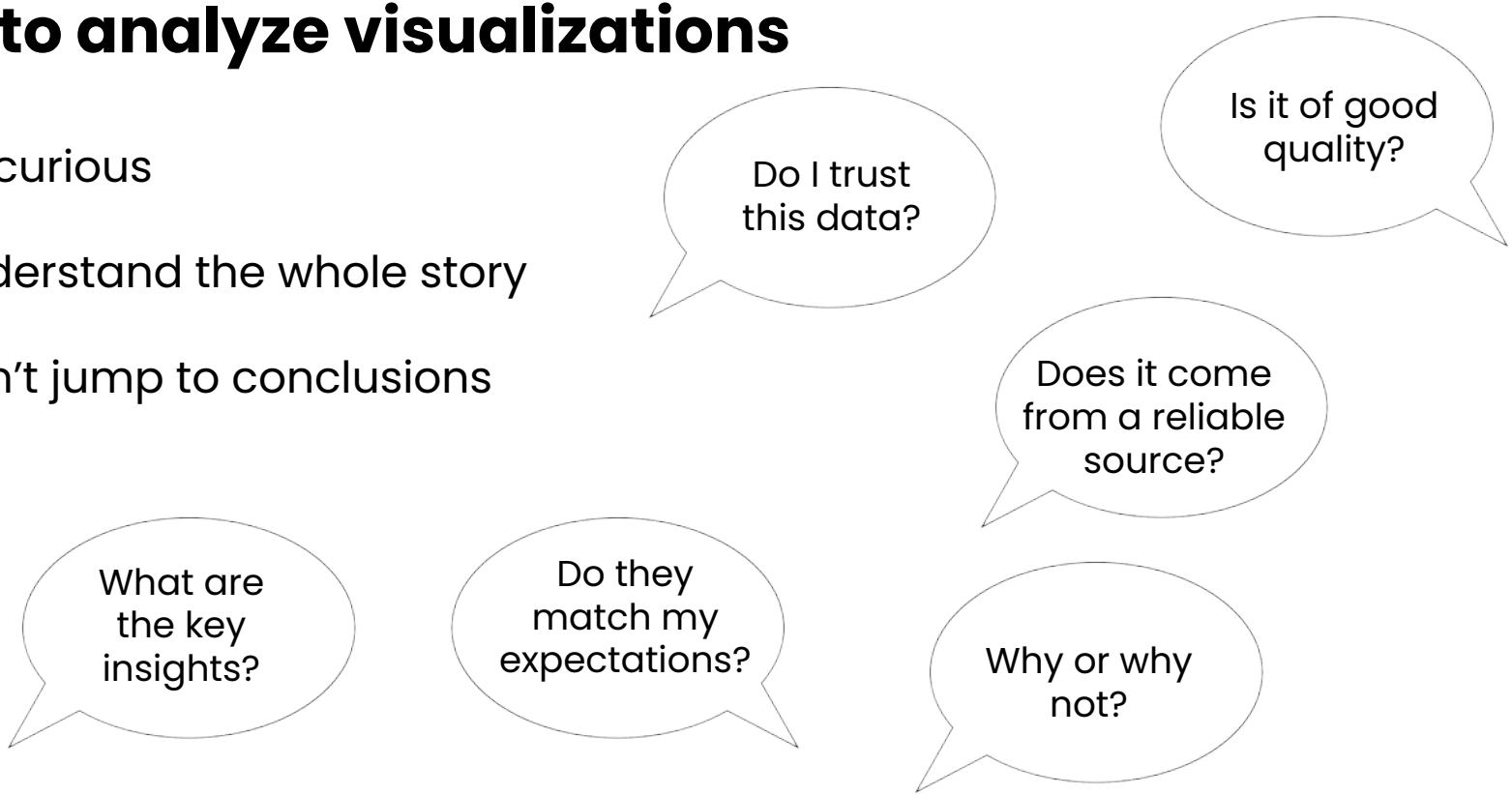
Data visualization

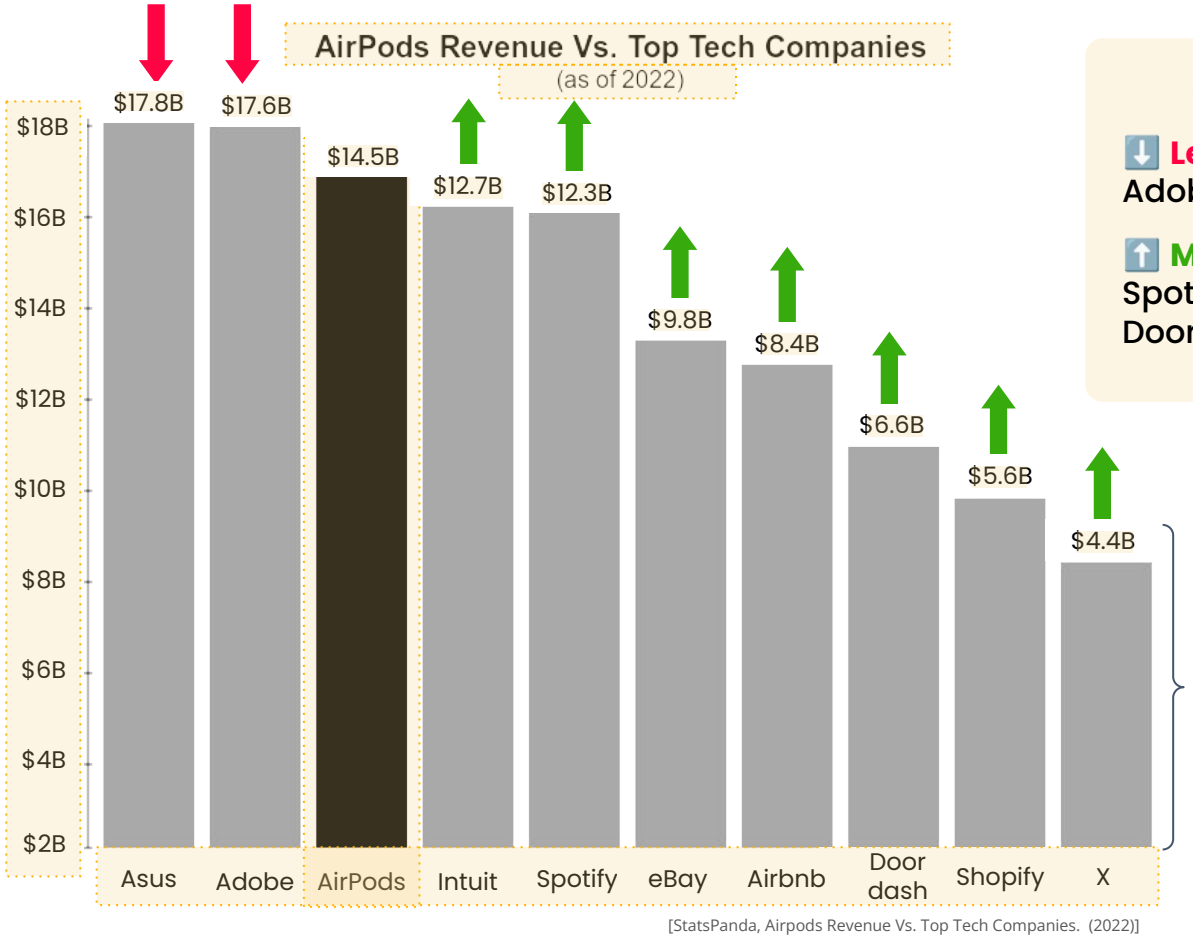
Analyzing visualizations

DeepLearning.AI

How to analyze visualizations

- ✅ Be curious
- ✅ Understand the whole story
- ⛔ Don’t jump to conclusions





[StatsPanda, Airpods Revenue Vs. Top Tech Companies. (2022)]

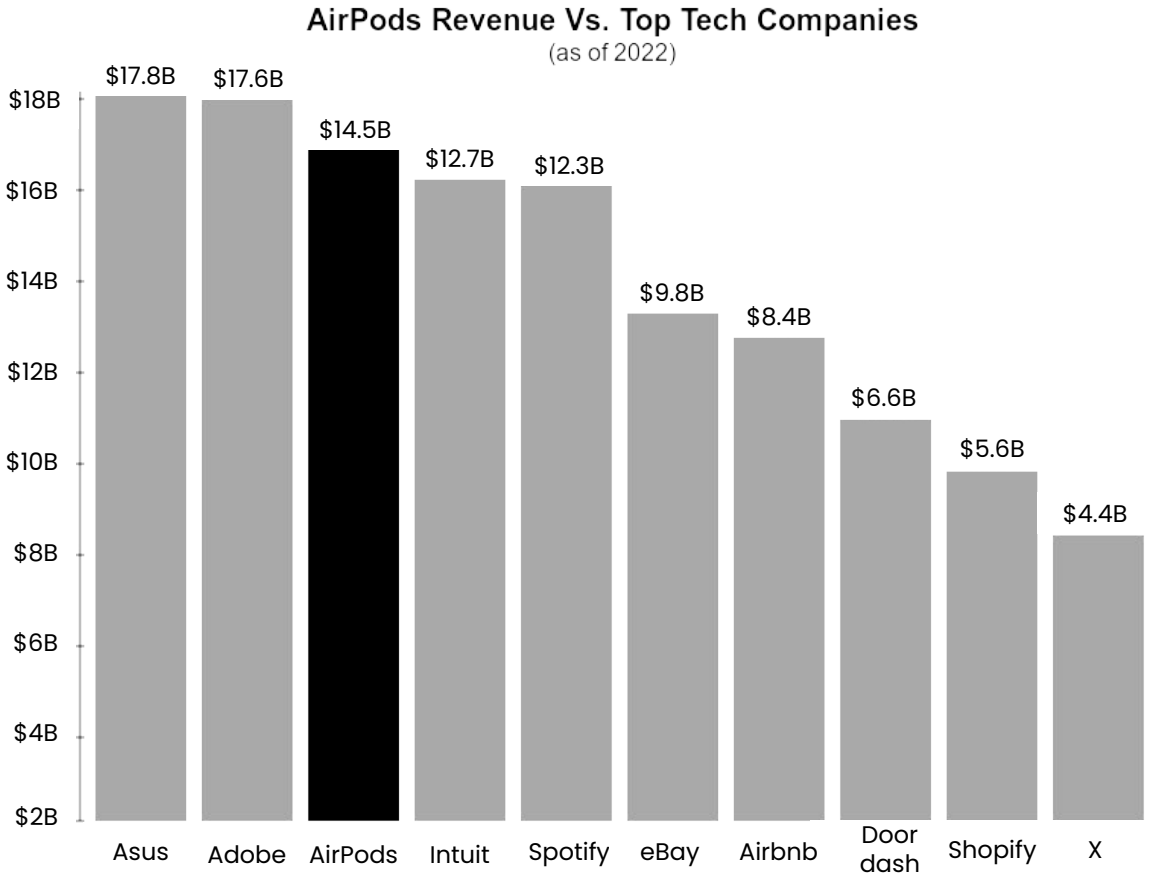
Airpods generated:

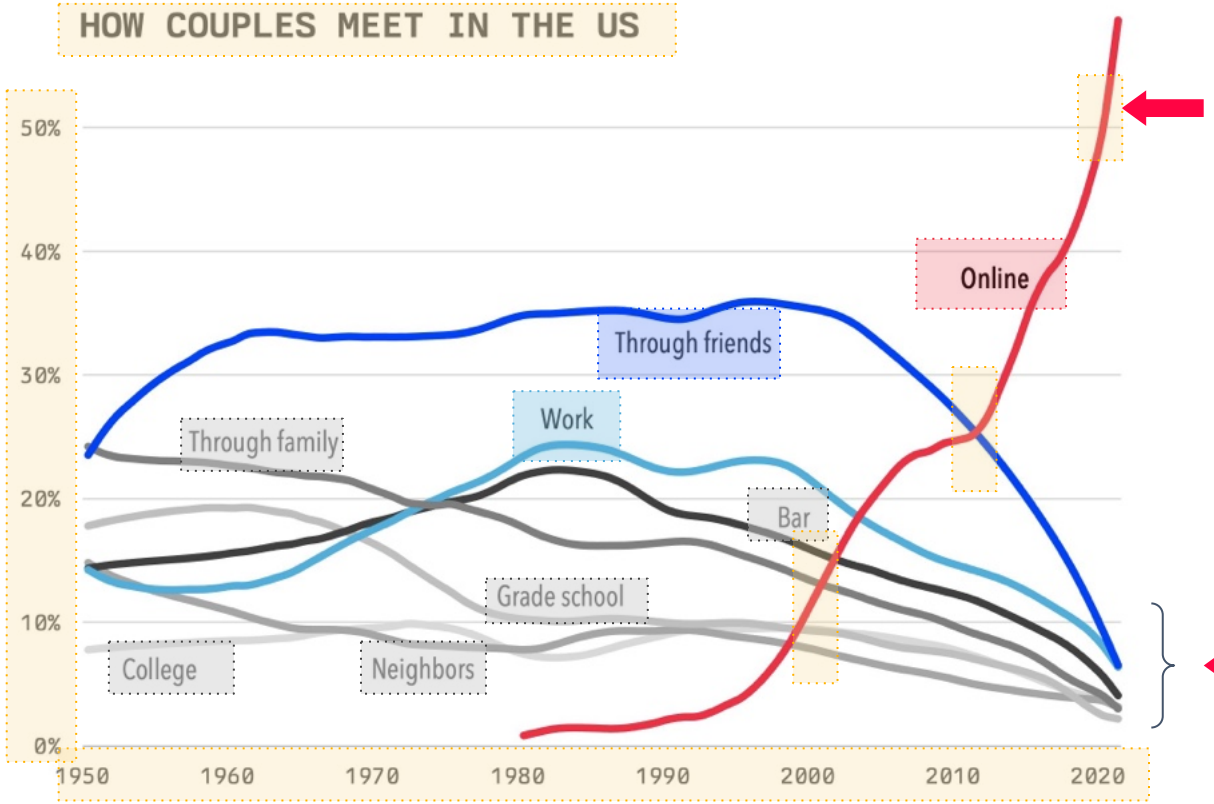
⬇️ **Less** than Asus and Adobe

⬆️ **More** than Intuit, Spotify, eBay, Airbnb, Doordash, Shopify, and X.

Double encoding

Bar height and **label** tell you the revenue


AirPods Revenue Vs. Top Tech Companies
(as of 2022)
$18B
$16B
$14B
$12B
$10B
$8B
$6B
$4B
$2B
$17.8B
$17.6B
$14.5B
$12.7B
$12.3B
$9.8B
$8.4B
$6.6B
$5.6B
$4.4B
Asus
Adobe
AirPods
Intuit
Spotify
eBay
Airbnb
Door dash
Shopify
X



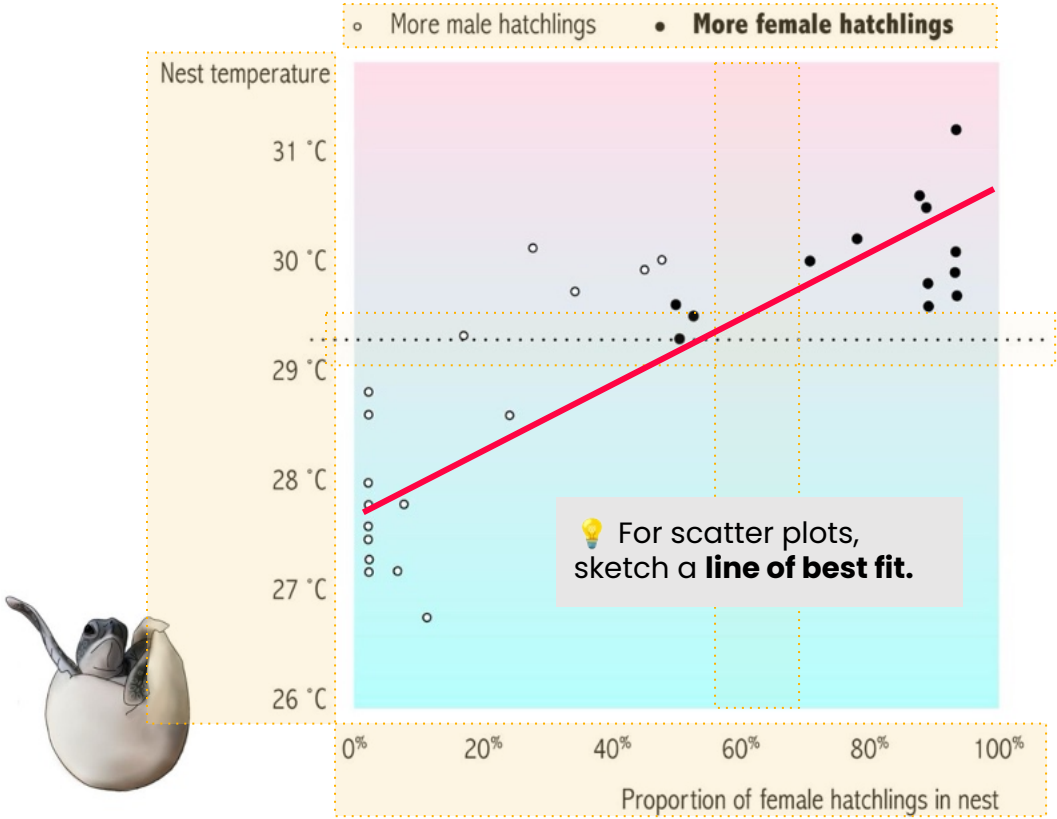
Source: "How Couples Meet and Stay Together": a longitudinal study of social life in the US by M. J. Rosenfeld, Reuben J. Thomas, and Sonia Hausen. Analysis of original survey data (n=6,519); "bars & restaurants" category cleaned to not double count couples who first met online.

MARRIAGE PACT

U.S. couples meet:

⬇️ Less often using methods aside from online

⬆️ More often online



[Storytelling with data. (2018), https://www.storytellingwithdata.com/blog/2018/10/23/scores-of-scatterplots]

Nests above the pivotal temperature produce more female baby green turtles.

The pivotal temperature for green turtles is **29.3˚C**

⬆️ As temperature increases, proportion of females increases.

⬇️ Below pivotal temperature, no nests with >30% females.

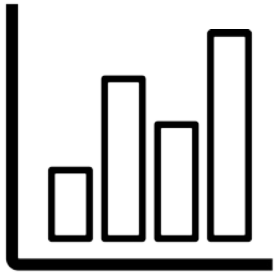
⬆️ Above pivotal temperature, a lot more nests with mostly females.

Data visualization

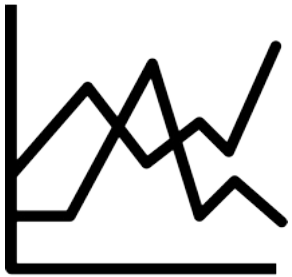
The right chart for
the right insight

DeepLearning.AI

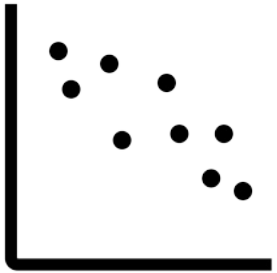
Data visualization types



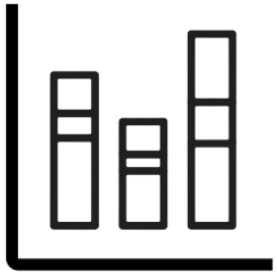
Bar/column charts



Line charts



Scatter plots

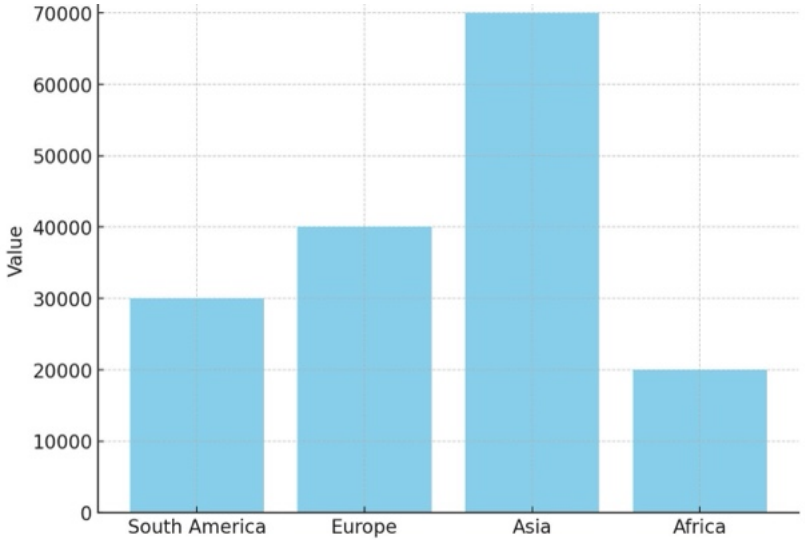


Stacked or grouped bar/column charts

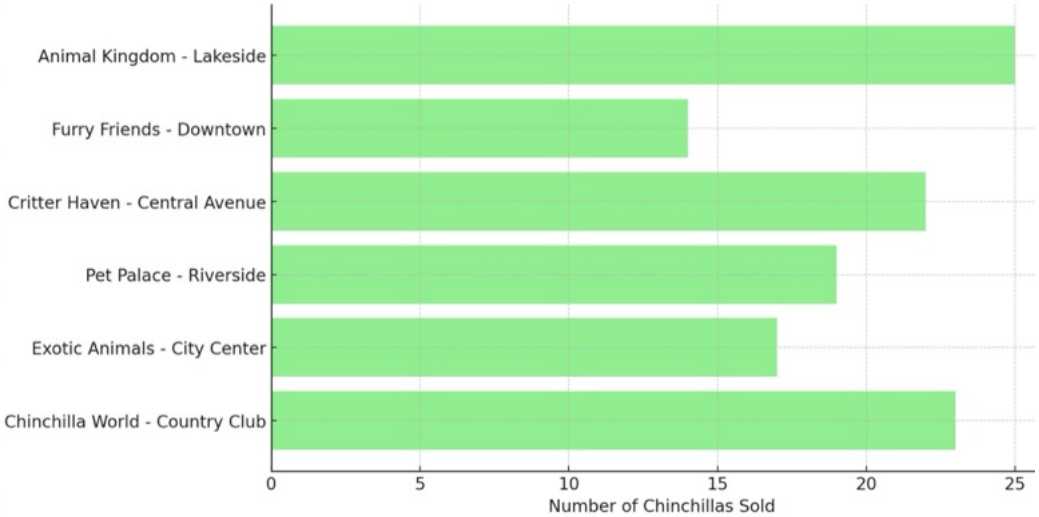
Bar and column charts

Purpose: To compare a **numerical feature** across a **categorical feature**

Album sales per region



Number of chinchillas per store



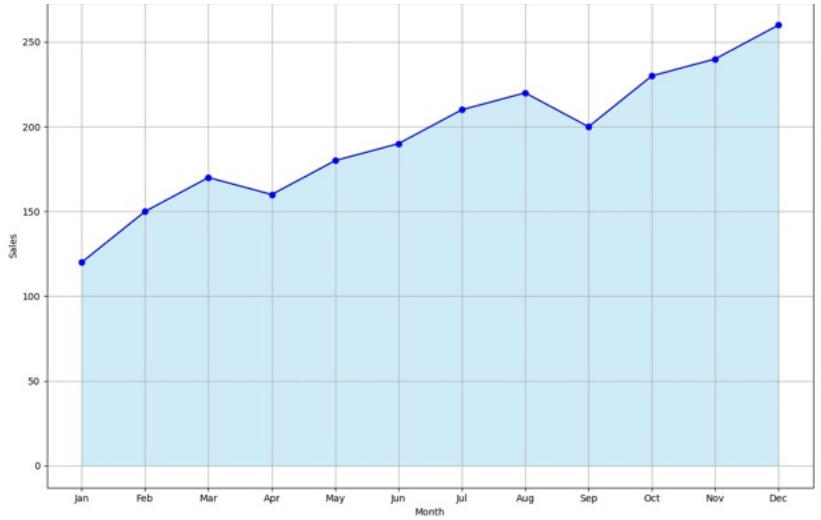
Line charts

Purpose: To show trends in a **numerical feature** over **time**

Line chart

- Easier to see **how sharply** sales increase or decrease

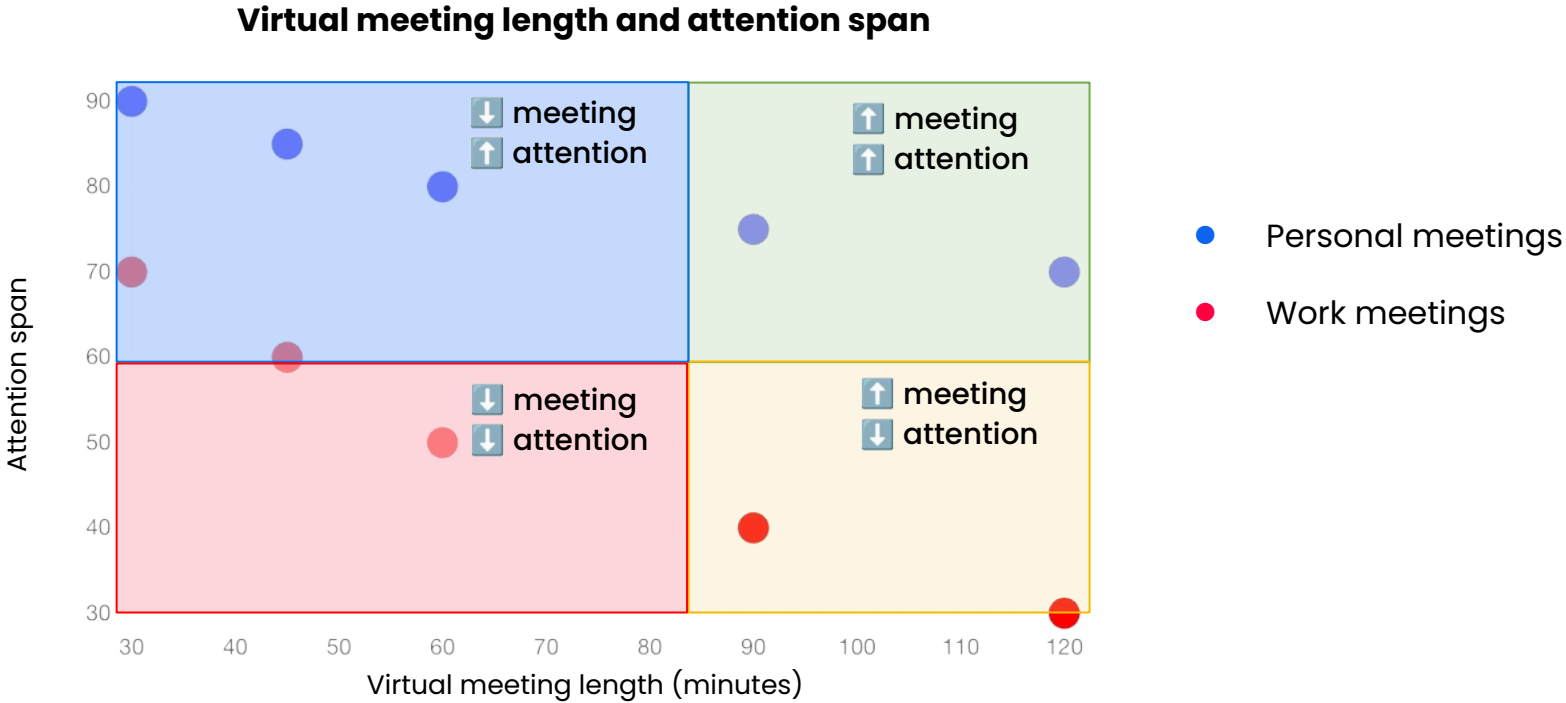
Area chart



- Common variation of line chart
- Emphasizes volume of data over time

Scatterplots

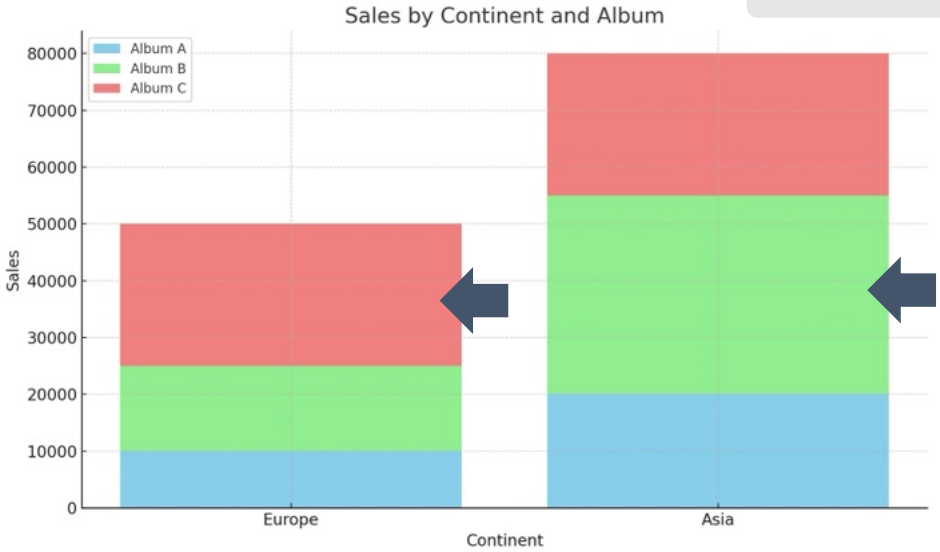
Purpose: To compare **two numerical features**



Stacked and grouped column

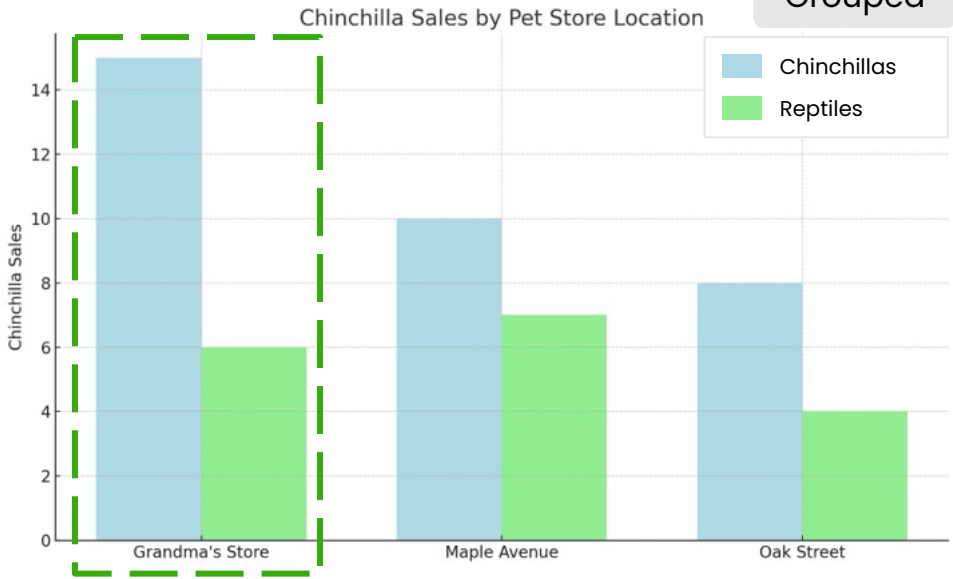
Purpose: Compare **a numerical feature** across **multiple categorical features**

Stacked



- Shows a part-to-whole relationship

Grouped

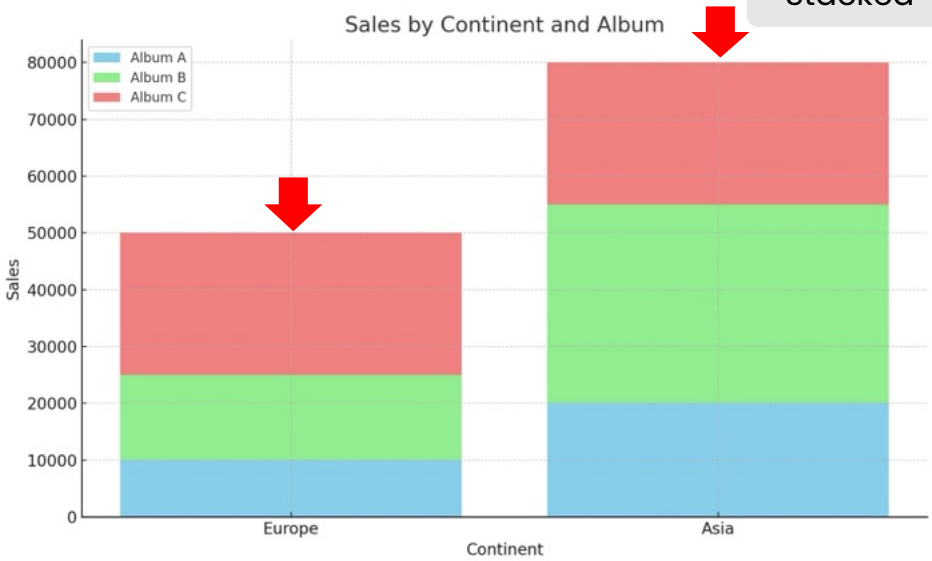


- Better for direct comparison between categories

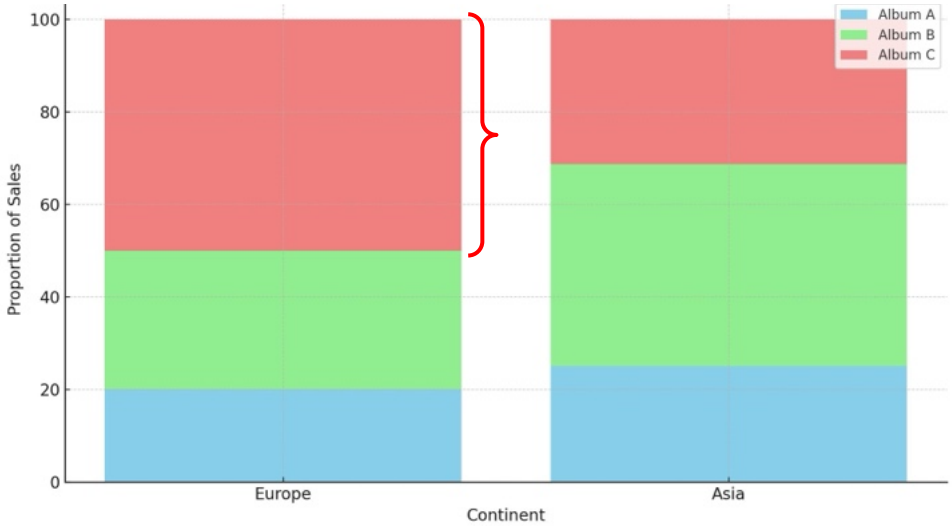
Stacked and grouped column

Variation: relative charts showing the **proportion** of each feature combination.

Stacked



Relative



- Make comparison across groups easier when the total size of each group is different.

Choosing the right visualization

1️⃣

Data

- What types of data are you working with?
- How many features are involved?
- What's the primary outcome of interest?

2️⃣

Primary message

- What's the primary message?
- Who is going to look at the visualization?
- Can they easily grasp what you're showing?

3️⃣

Relationships

- Comparing categories?
- Showing changes over time?
- Displaying relationships between features?

Cheat sheet

Data type	Recommended chart type
Time series data	Line charts
Comparisons between categories	Bar or column charts
Relationships between two numerical features	Scatterplots
Comparing parts of a whole or multiple categories over time	Stacked or grouped bar/column charts

Number of James Bond movies with each of the 7 different James Bond actors

- Bar or column chart
- Line chart
- Scatter plot
- Stacked or grouped bar chart

Global coffee consumption by country over the last 50 years

Bar or column chart

Line chart

Scatter plot

Stacked or grouped bar chart

Proportion of five different pizza toppings ordered in New York vs. Chicago

Bar or column chart

Line chart

Scatter plot

Stacked or grouped bar chart

Correlation between a country's chocolate consumption and Nobel Prize winners

Bar or column chart

Line chart

Scatter plot

Stacked or grouped bar chart

Data visualization

Demo: Bar & column charts

DeepLearning.AI

Data visualization

Demo: Customizing charts

DeepLearning.AI

Data visualization

Demo: Scatter plots

DeepLearning.AI

Data visualization

Demo: Grouped bar & column charts

DeepLearning.AI

Data visualization

Demo: Stacked bar & column charts

DeepLearning.AI

Data visualization

Demo: Line charts

DeepLearning.AI

Data visualization

Strategies for effective data visualization

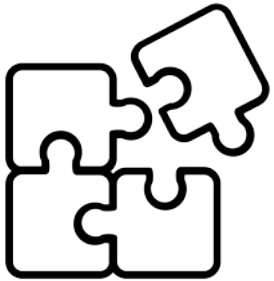
DeepLearning.AI

Process of creating an effective visualization

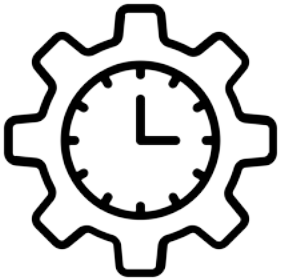
You might **iterate** through these steps **multiple times**

Principles of evaluating visualizations

Clarity



Context



Efficiency

Clarity

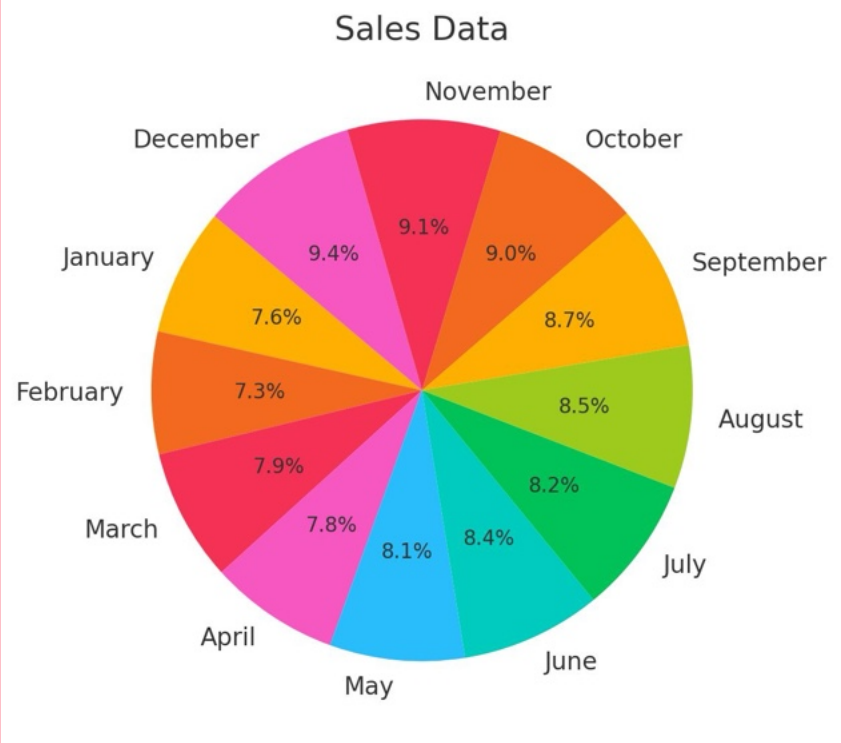
💡 **Purpose**: Ensuring that audience interprets visualization as intended

Is your chart clear?

- ⬜ Choose the appropriate chart
- ⬜ Avoid unnecessary complexity
- ⬜ Use clear labels, titles, and annotations
- ⬜ Use consistent color schemes, fonts, and scales
- ⬜ Share your work with others

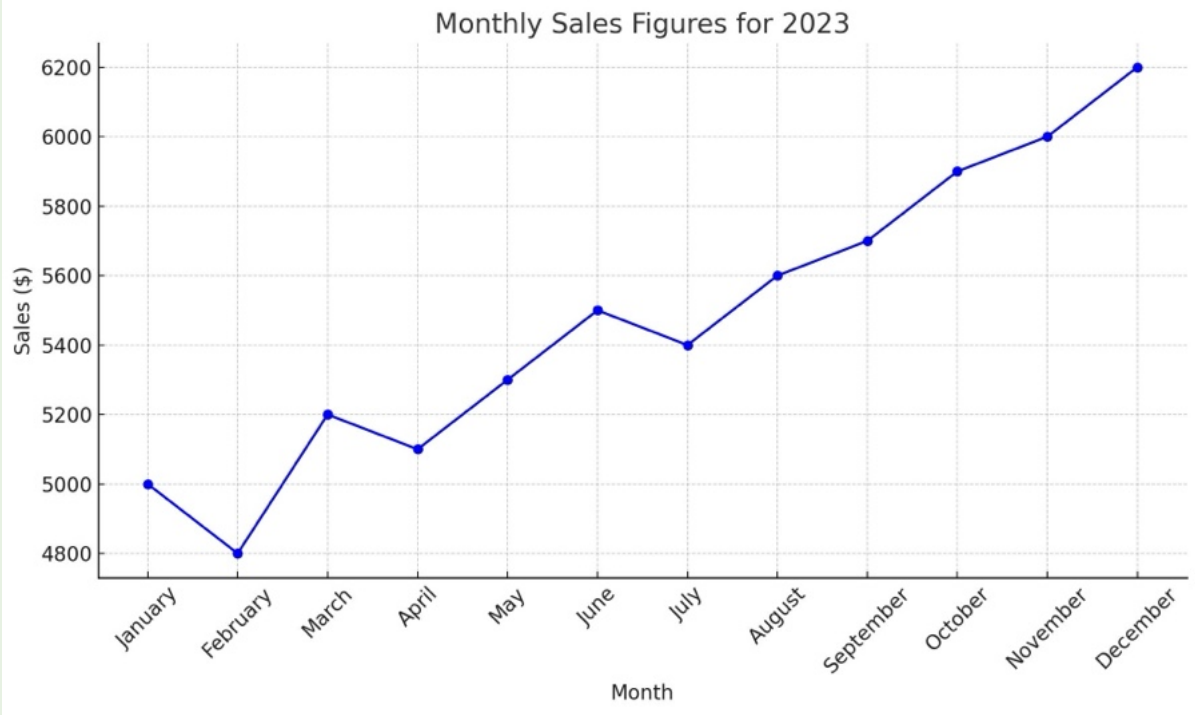
Clarity

👎🏽 Unclear visualization



Clarity

👍🏽 Clear visualization



Clarity

💡 **Purpose**: Ensuring that audience interprets visualization as intended

Is your chart clear?

- ✅ Choose the appropriate chart
- ✅ Avoid unnecessary complexity
- ✅ Use clear labels, titles, and annotations
- ✅ Use consistent color schemes, fonts, and scales
- ✅ Share your work with others

👍🏽 Clear visualization

Efficiency

💡 **Purpose**: Including only elements that serve a purpose

Data-ink ratio: proportion of ink used to show actual data compared with decorations

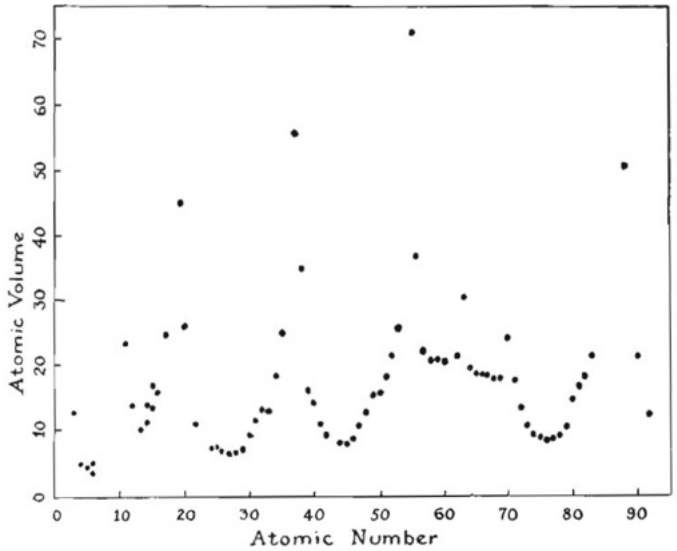
"Chartjunk"

Includes:

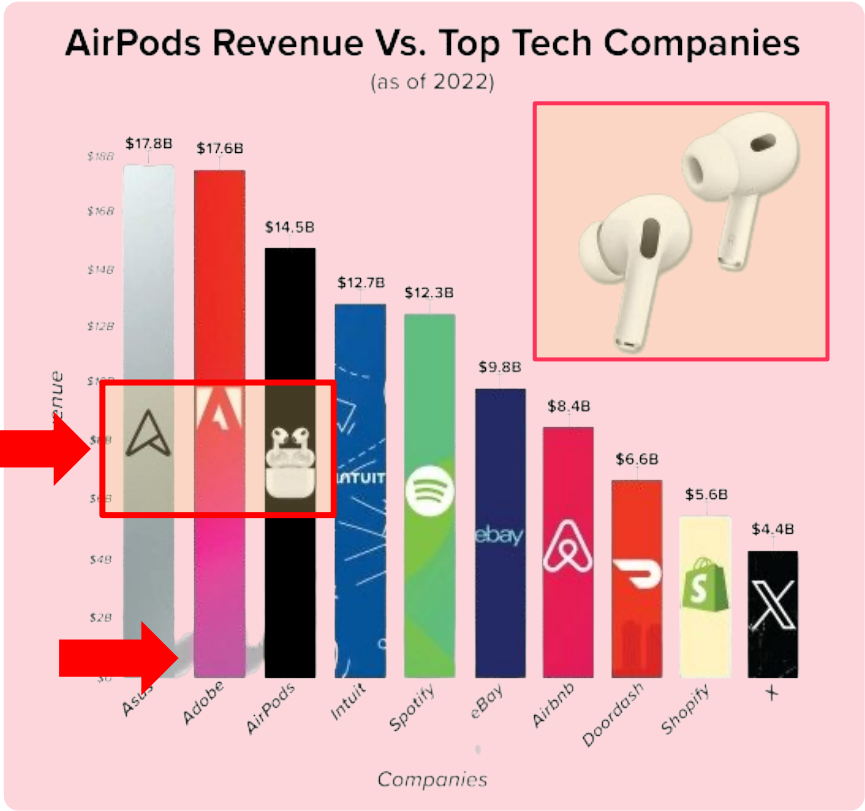
- ✅ Bars
- ✅ Markers
- ✅ The line in a line chart
- ✅ Axis labels
- ✅ Concise annotations
- ✅ Data labels

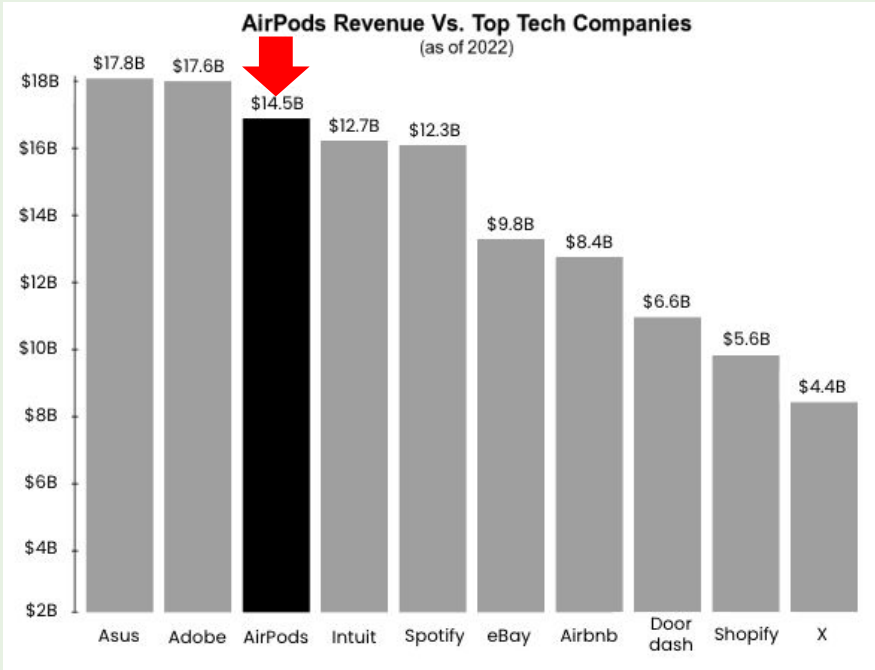
Does not include:

- ⛔ 3D effects
- ⛔ Heavy borders
- ⛔ Shadows
- ⛔ Excessive gridlines
- ⛔ Overly descriptive annotations



[Tufte, E, The Visual Display of Quantitative Information, modified from original by Roger Hayward, published in Pauling's General Chemistry (1947)].

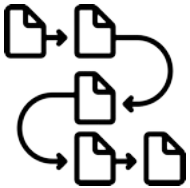




[StatsPanda, Airpods Revenue Vs. Top Tech Companies. (2022)]

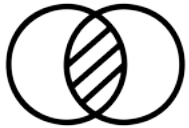
Context

💡 **Purpose**: Grounding your audience's understanding



Create narrative structure

Provide background



Compare with familiar concepts

Define jargon

Explain significance of the data

⚠️ Err on the side of including more context.

Data visualization

Data encoding

DeepLearning.AI

Data encoding vs. chart elements

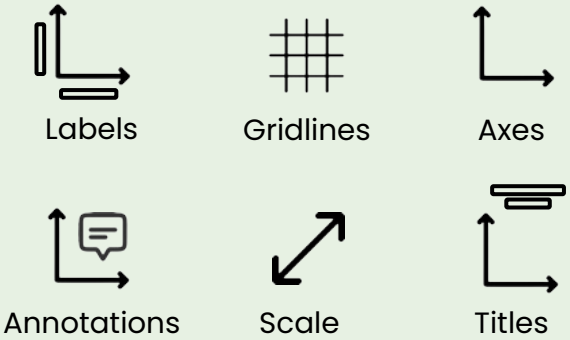
Data encoding

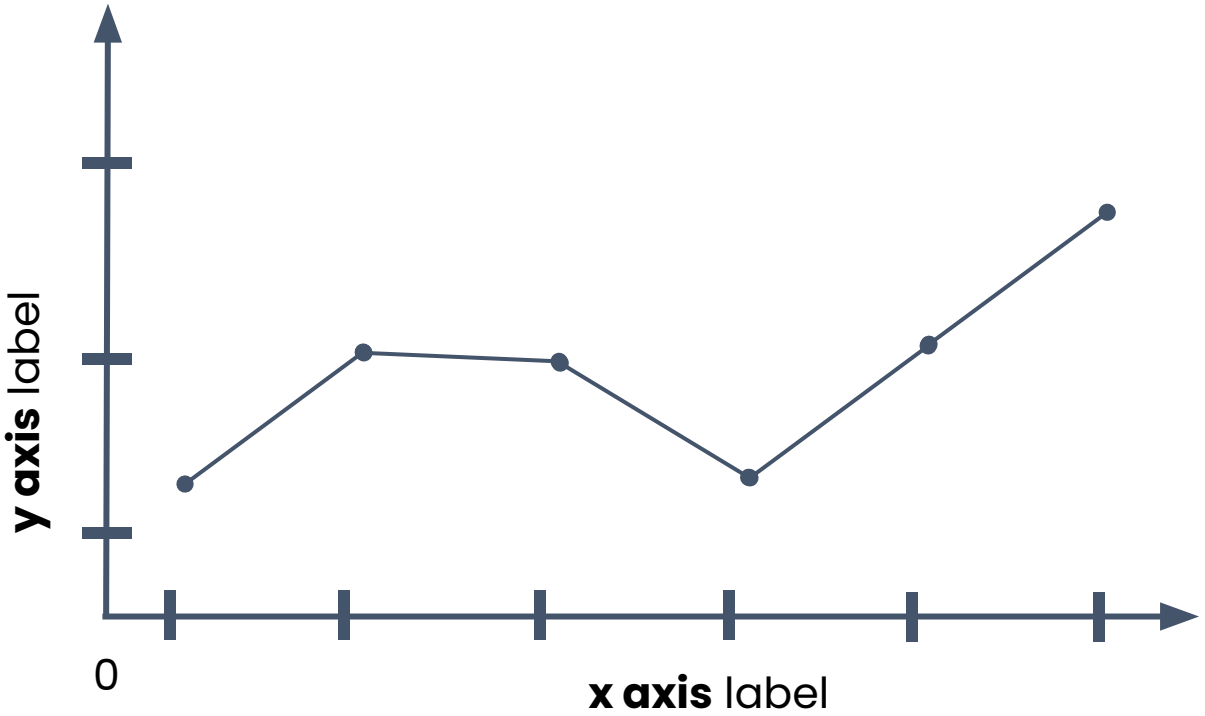
- How the data is visually represented

- Subset of data-ink that directly shows the data

Chart elements

- Additional tools to improve clarity and context





Make sure axes are:

- ✅ Easy to read
- ✅ Labeled
- ✅ Intuitive

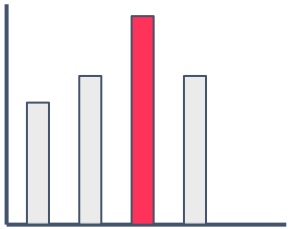
Avoid:

- ⛔ Exaggerating or compressing data

Color

One of your most powerful tools for creating clarity and context

Use cases:



Highlight key insights

Provide context

Be aware!

🎨 4.5% of people have some form of colorblindness

➕ Use double encoding by combining color with another element

💡 Additional clarity helps everyone

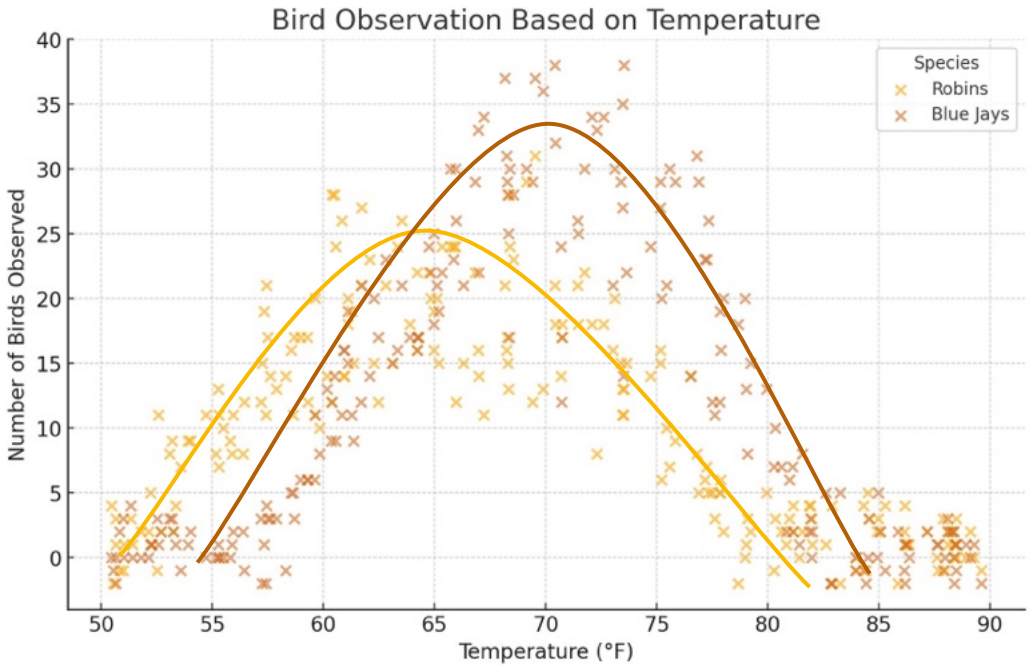
Number of dimensions

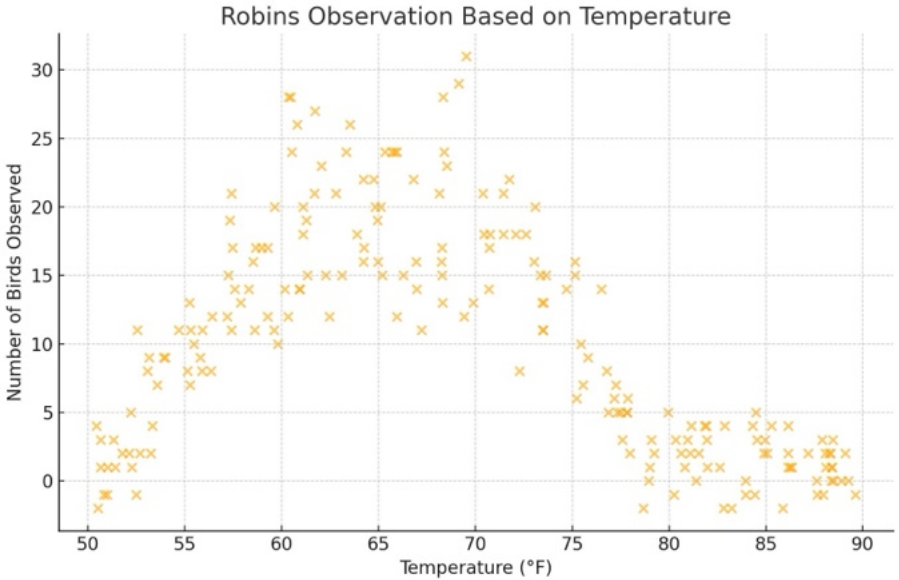
- Keep data to **two** dimensions: **x** and **y**
- For **three or more,** try using multiple plots next to each other

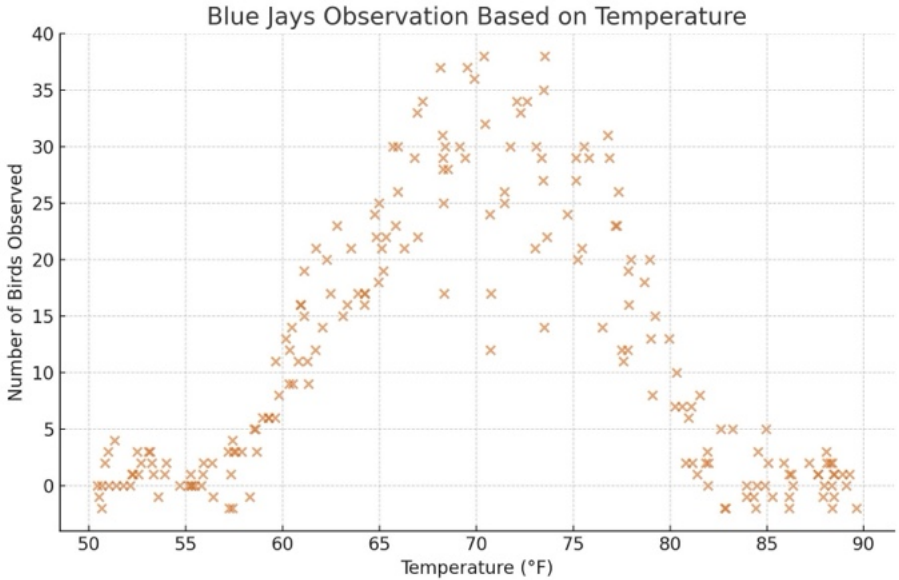
---

Scenario: Number of birds observed each day based on temperature

- Temperature
- Number of birds observed
- Species




Robins Observation Based on Temperature
Number of Birds Observed
Temperature (°F)
0
5
10
15
20
25
30
50
55
60
65
70
75
80
85
90


Blue Jays Observation Based on Temperature
Number of Birds Observed
Temperature (°F)
0
5
10
15
20
25
30
35
40
50
55
60
65
70
75
80
85
90

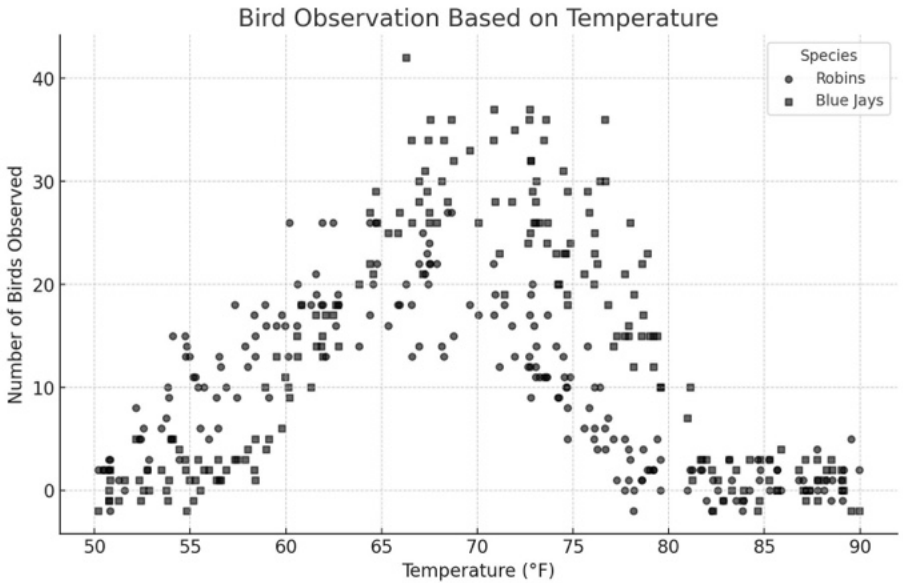
Markers

Data encoding element used to add a third dimension

✅ Useful if comparison is clear

⛔ If using **more than two types,** rethink your approach:

- Using color
- Separating data into multiple charts



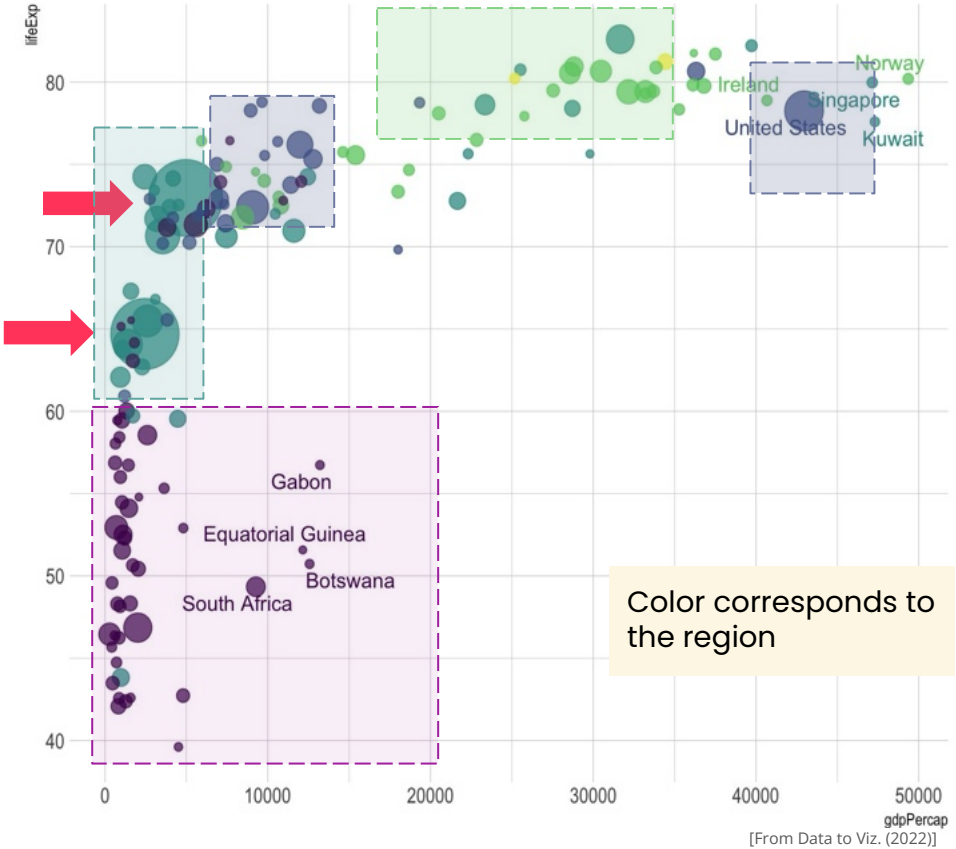
Size variations

Work well when there is a natural analogy to the size:

- 🌍 Population size
- 💵 Dollar amounts

⛔ Don’t overdo it with visual elements

✅ Each addition should enhance understanding



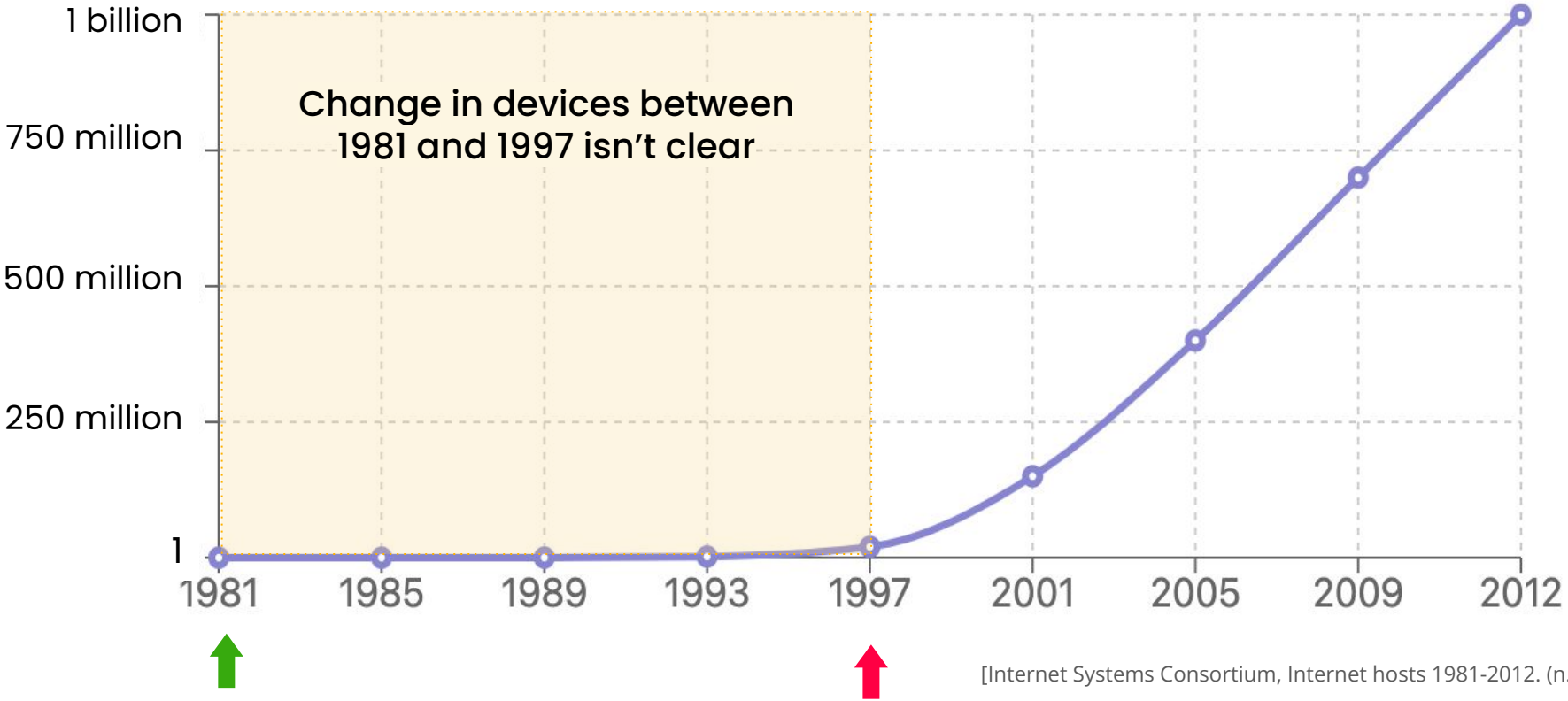
[From Data to Viz. (2022)]

Data visualization

Chart elements

DeepLearning.AI

Internet hosts 1981-2012



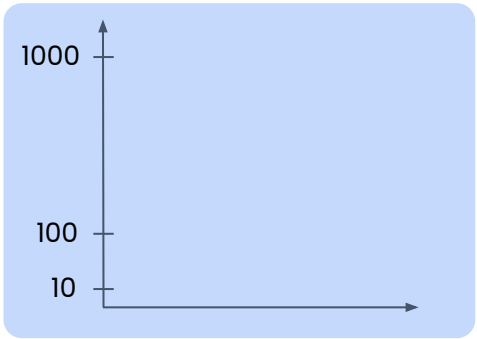
[Internet Systems Consortium, Internet hosts 1981-2012. (n.d.)]

Logarithmic scale

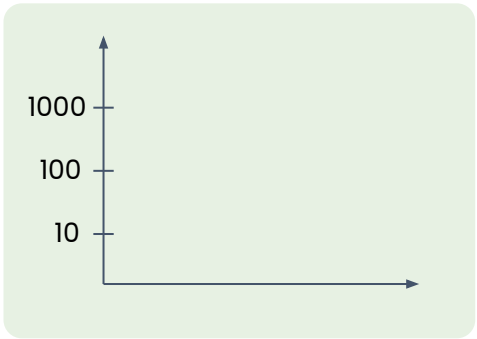
Changes the distances between values on the **y-axis**:

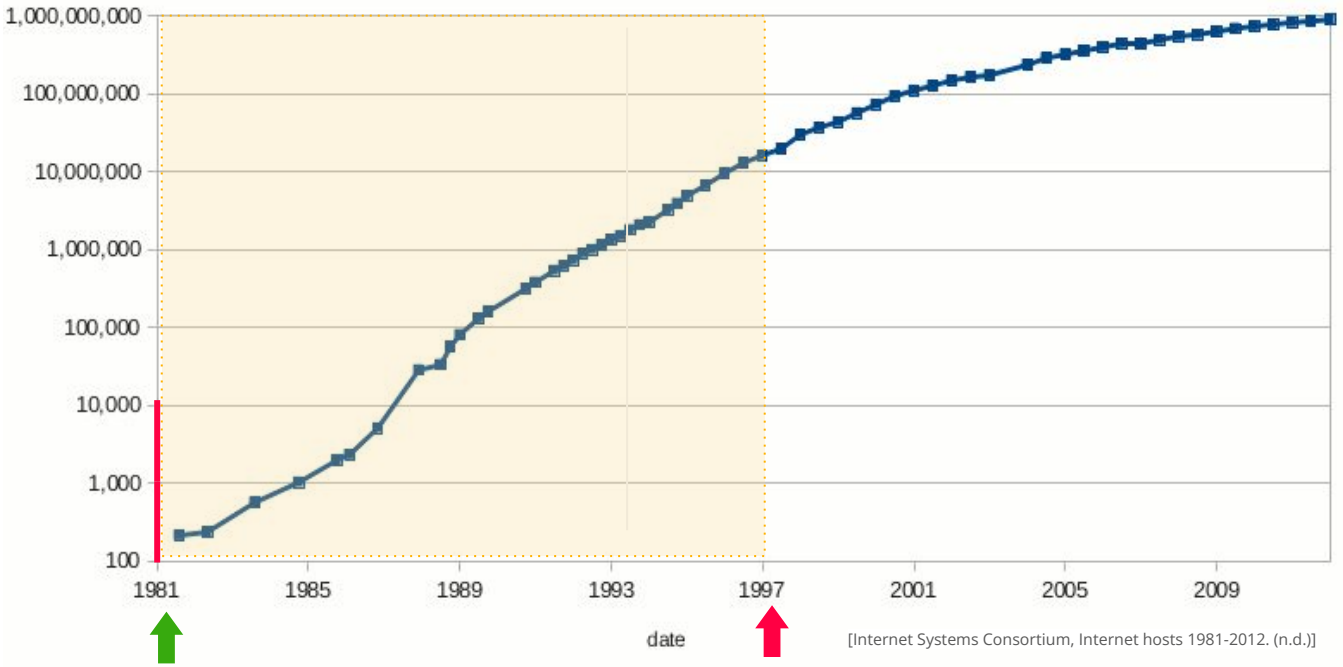
- Spreading out smaller values
- Compressing larger values
- Making patterns across the **lower range** more visible

Linear scale



Logarithmic scale





Consider log scale to:

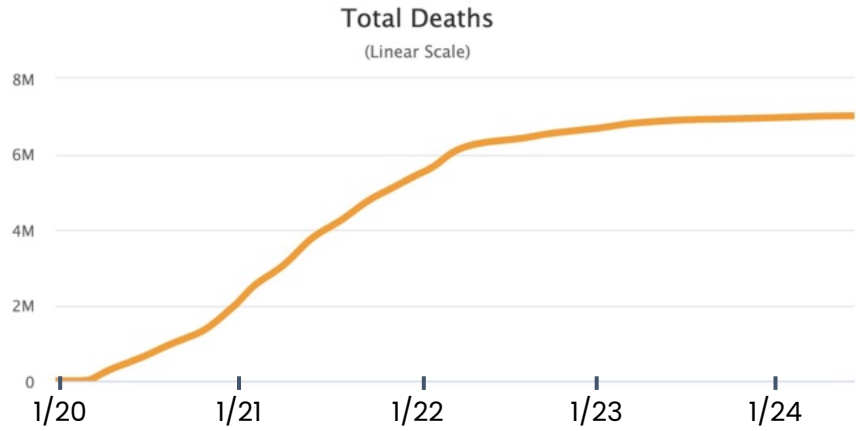
- ✅ Cover a large range of data
- ✅ Emphasize proportional changes
- ✅ Spread out clustered data points for visibility

⛔ Log scale cannot be used with negative or 0 values

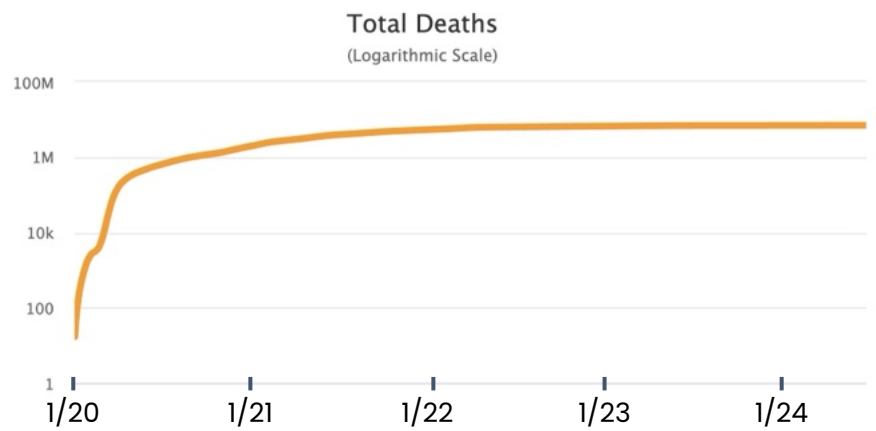
Be careful about:

- How well the audience can interpret what they're seeing
- Consider whether it's worth the complexity

Linear scale



Logarithmic scale



[Worldometers.info/coronivrus, (2024)]

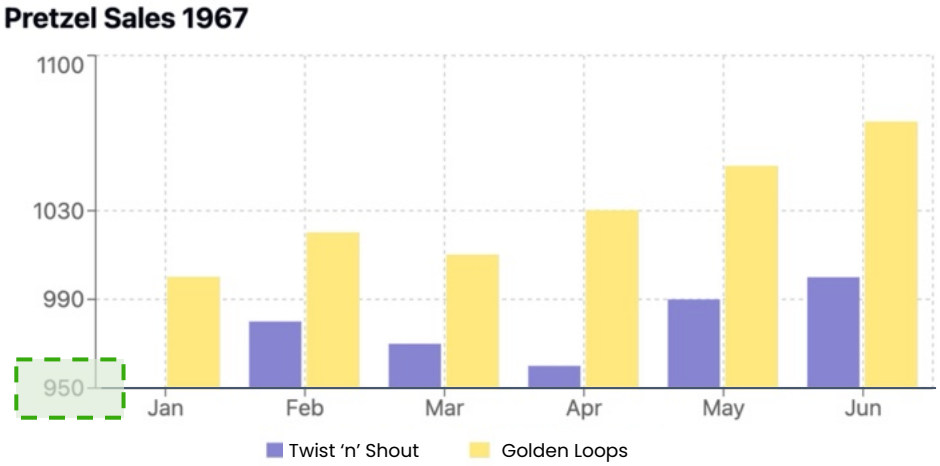
💡 **Rule of thumb**: If your data is bunched up in one area, consider a log scale

Significantly lower comprehension when asked to:

- compare deaths across weeks
- predict the number of deaths in a future week

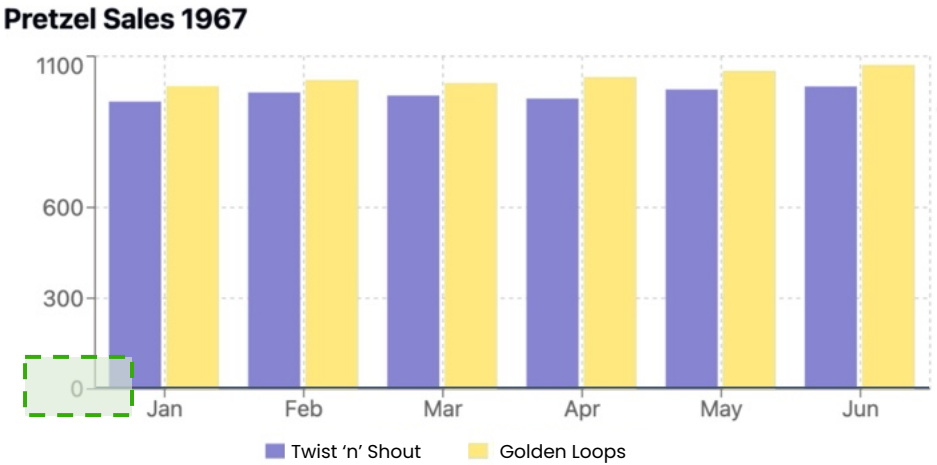
Axis scale: zero

✅ Excluding zero can be helpful to emphasize small differences



Zoom in on differences between the two brands

✅ Helps communicate the magnitude of your data, in particular absolute value



Assess the most popular brand

Annotations

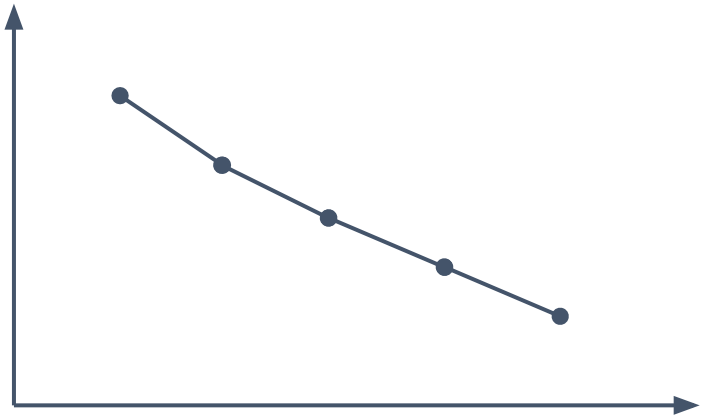
- Tool for guiding audience attention
- Without annotations, eyes wander all over the chart
- Lock in focus on most important elements
- Choose one to three key points to highlight

Viewing context	Recommendation
Presenting in person	• Additional callouts • Fewer annotations needed
Viewed independently	• Add a caption to explain key points

Chart title

Crime Decreasing in Berlin This Year

~~Crime Data in Berlin~~

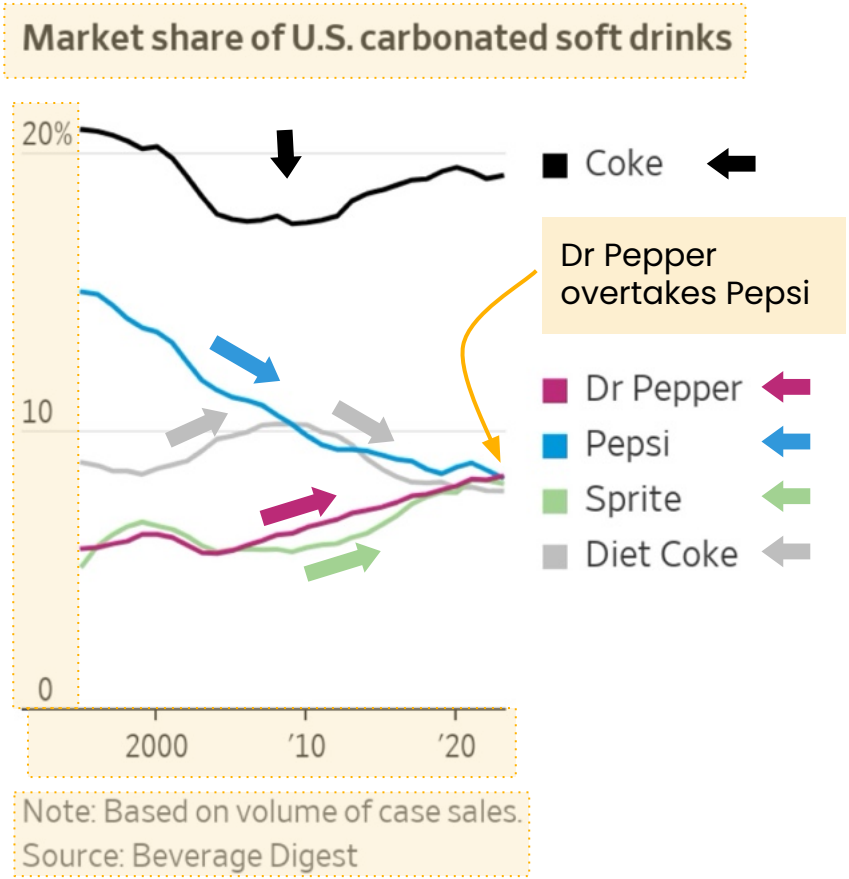


- ✅ Draws attention to your main point
- ✅ Helps prevent misinterpretation
- ✅ Provides crucial context
- ✅ Helps audience understand quickly

Data visualization

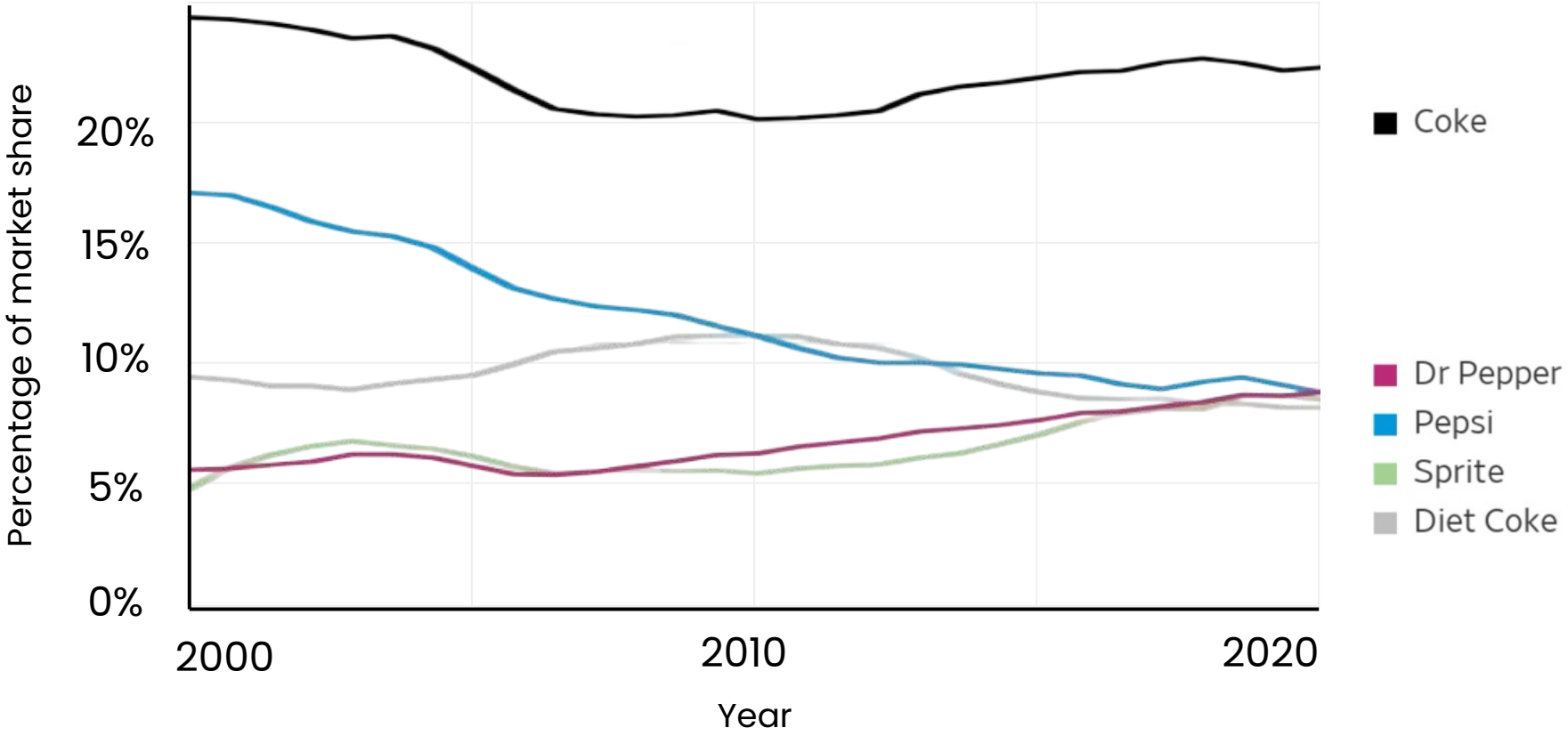
Data visualization:
the good and the better

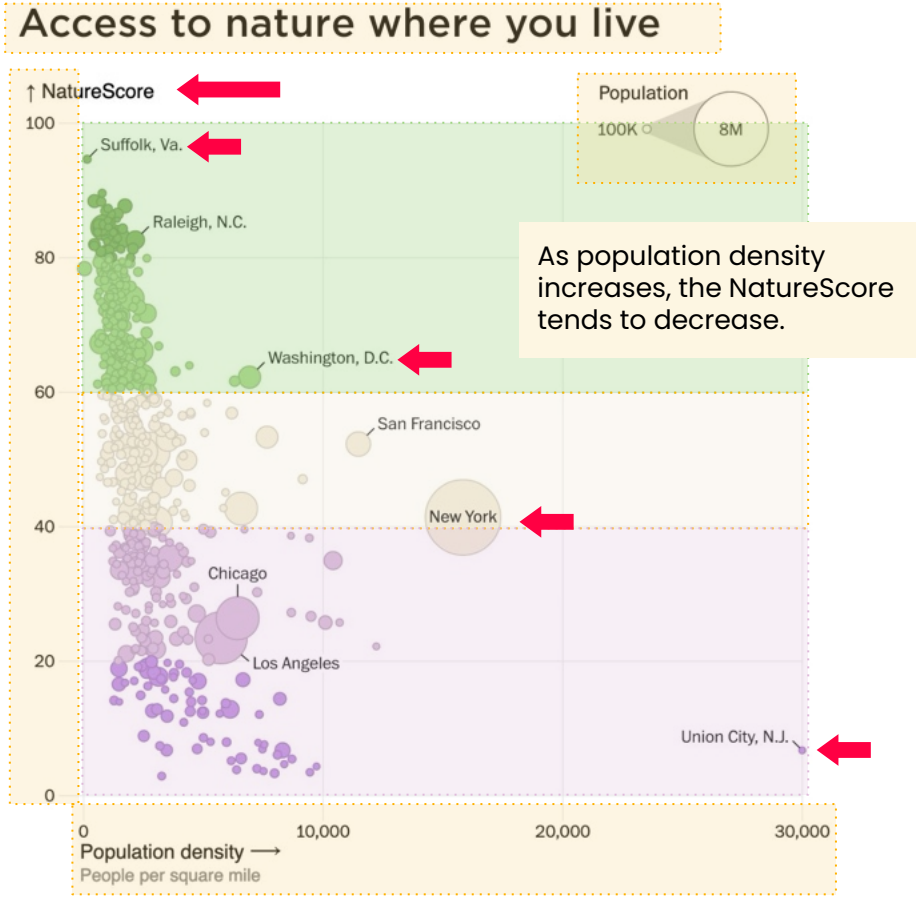
DeepLearning.AI



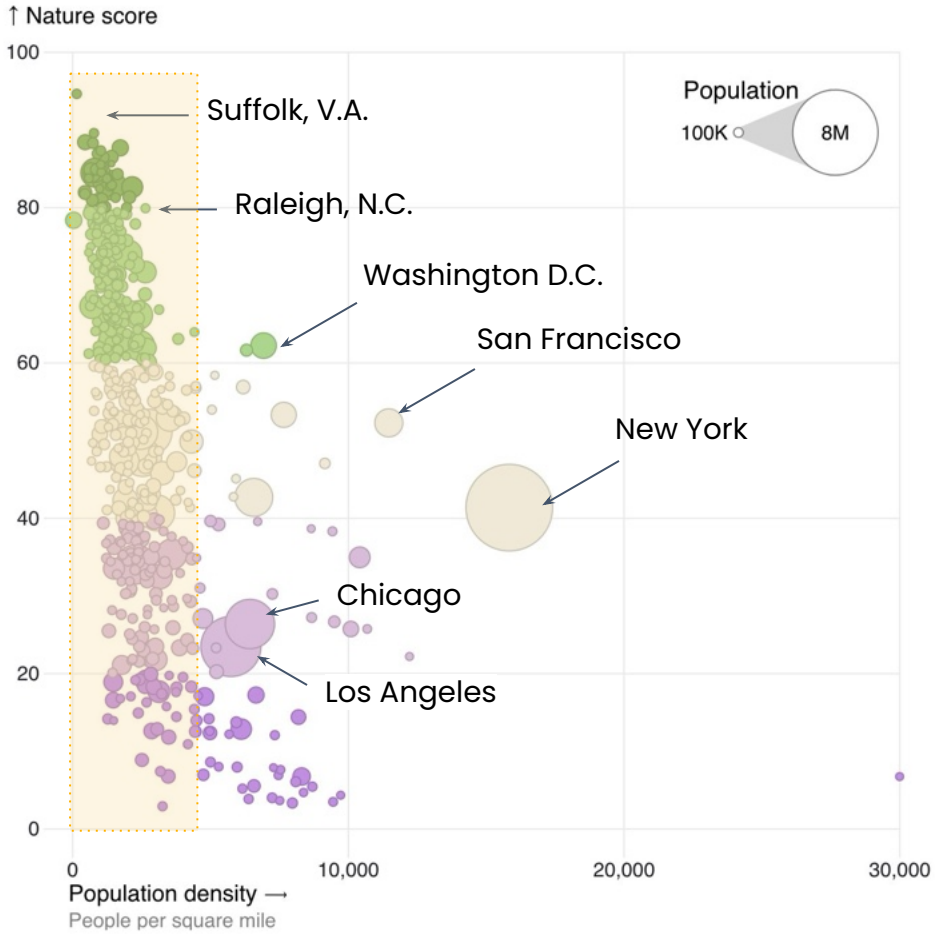
[Beverage Digest. Carbonated Soft Drink Dollars up +13.2% in First-Half 2023. Volume Down -3.5%. (2023)]

Dr Pepper tops Pepsi for second most popular US soft drink





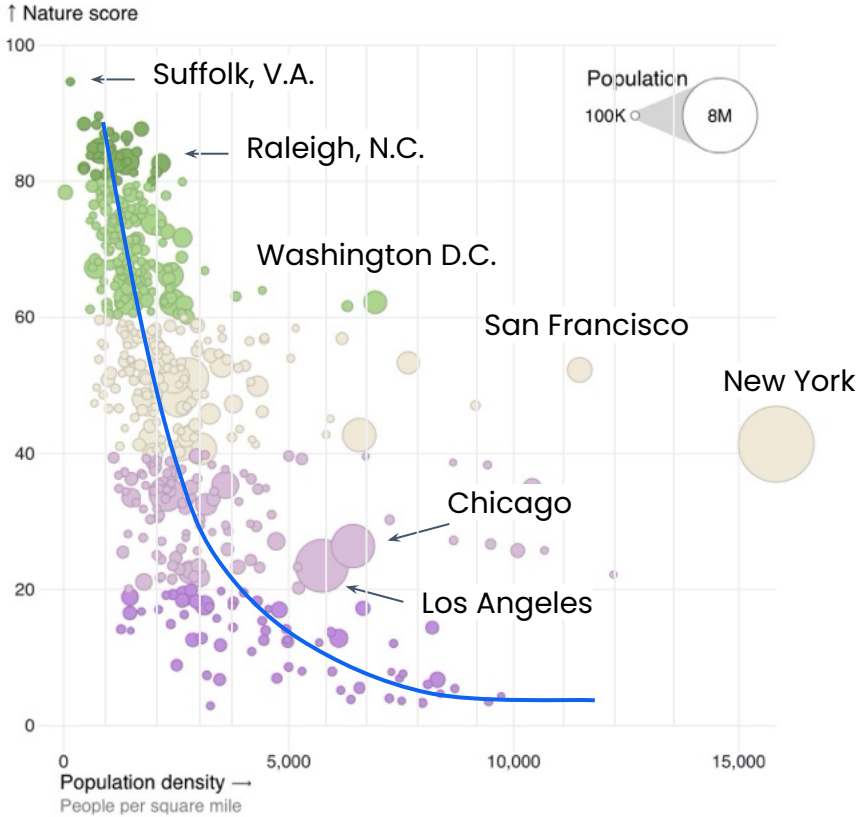
[Stevens, H., Mapping America's access to nature, neighborhood by neighborhood, The Washington Post. (2024)]



To improve this chart:

- Increase the font size
- Use using log scale to spread out lower values

Cities with lower population density offer better access to nature



To improve this chart:

- Increase the font size
- Use using log scale to spread out lower values
- Drop Union City, NJ
- Add more gridlines to help with fine-grained comparison
- Adding a trendline to make the overall relationship clearer

Data visualization

Demo: Interpreting data visualizations with LLMs

DeepLearning.AI

Data visualization

Demo: Creating data visualizations with LLMs

DeepLearning.AI