Lisbon School of Economics & Management
Universidade de Lisboa

LISBOA
UNIVERSIDADE DE LISBOA

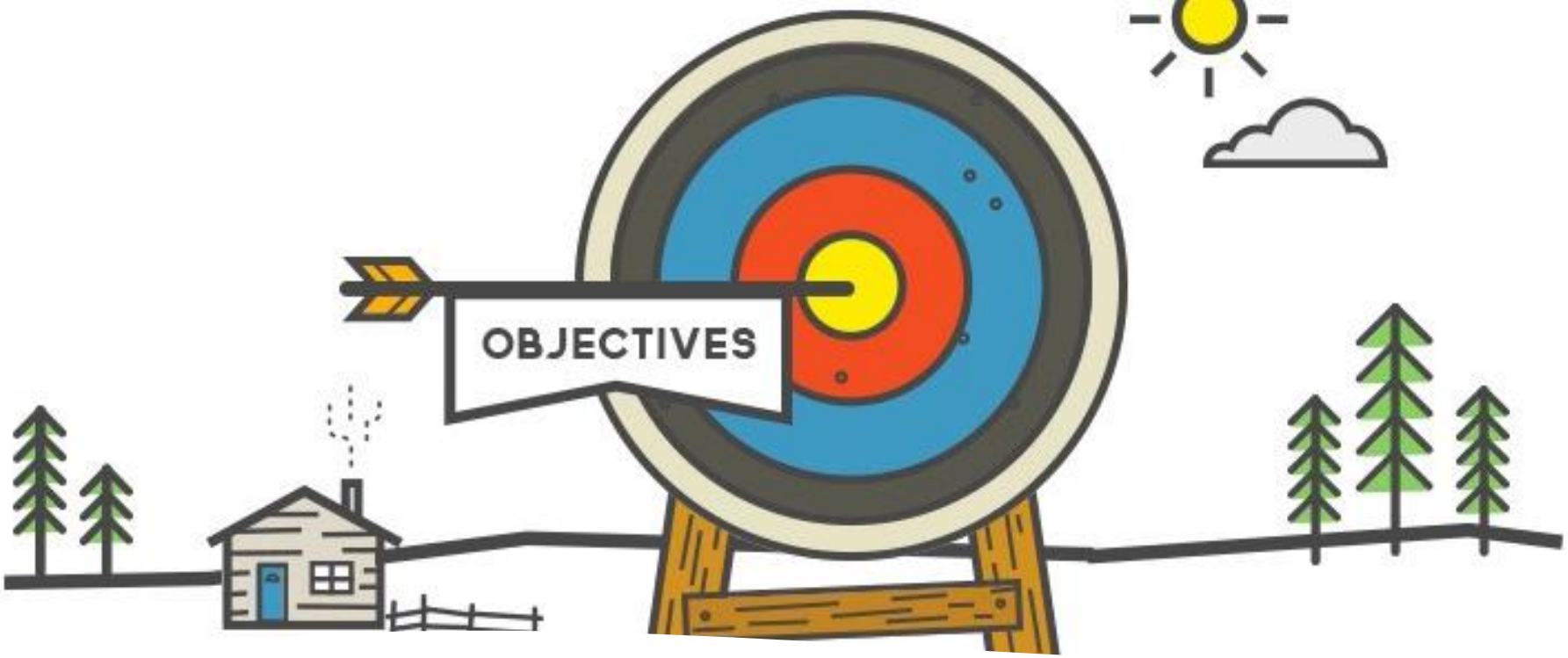
Types of Data and Visualization Techniques

Prof. Carlos J. Costa, PhD

Saeed Angorani, DBA

Session Overview

- Data types and why they matter
- Core visualization techniques
- Choosing the right chart
- Design principles (Data-to-Ink)
- Hands-on Excel lab



Learning Goals

By the end of this session, you will be able to:

1. Data types and why they matter
2. Core visualization techniques
3. Choosing the right chart
4. Design principles (Data-to-Ink)
5. Hands-on Excel lab

Why Data Type Matters

- The same data can tell different stories depending on visualization.
- Wrong chart = wrong decision.

- Example: Using pie chart for trends hides time patterns.

What is Data?

- Data = raw observations without interpretation.
- Visualization = turning data into meaning.

- Your role: translator between numbers and decisions.

Your default chart tool?

Poll: Which tool do you use most for charts?

A) Excel

B) Google Sheets

C) Power BI / Tableau

D) Python

E) Other

Warm-up Question: Classify the data

- **Question:** For each column, identify: (1) qualitative or quantitative, (2) structured or unstructured.

Column	Example value
Region	“Lisbon”
Revenue	12,450
Customer Review	“Delivery was late”
Order ID	104982

Overview of Data Types

Big Picture: Data Type Drives Chart Type

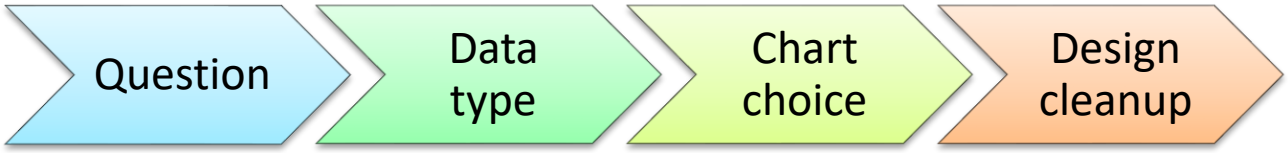
Visualization is a translation:

- Data type limits what comparisons are valid.

- The question determines what pattern you want the chart to reveal.

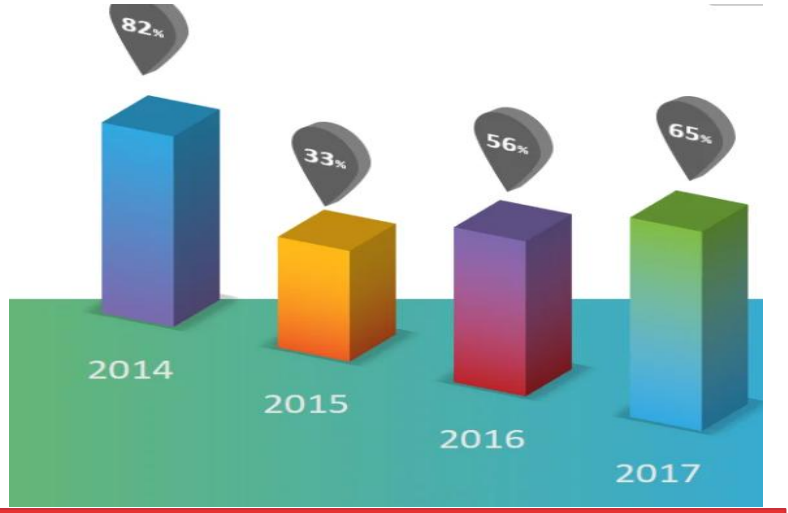
- The chart must align with human perception: some encodings are easier to read than others.

Rule: Start with the question, not your favorite chart.



Qualitative (Categorical) Data

- Qualitative data represents categories/labels, not magnitudes.
- Examples: Region, Channel, Product Category, Customer Segment.
- What you can do: counts, proportions, comparisons across categories.
- What you cannot assume: arithmetic meaning unless categories have numeric structure.

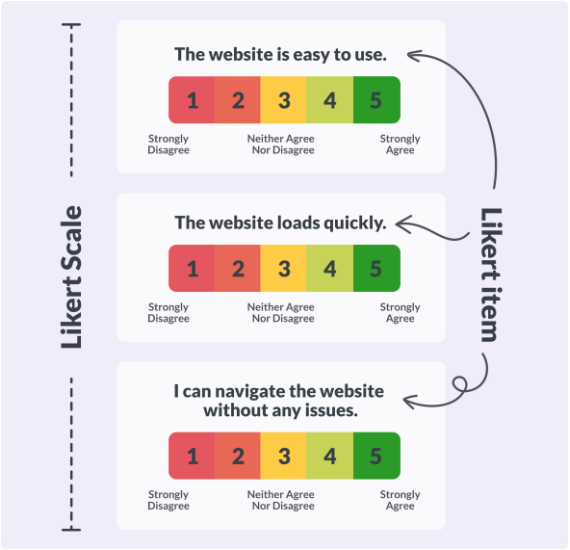


Nominal vs Ordinal Categories

Nominal: no natural order (e.g., Region, Payment method).

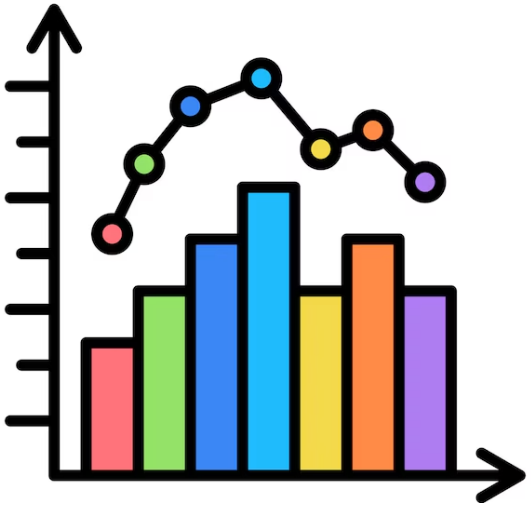
Ordinal: meaningful order (e.g., Satisfaction: Low–Medium–High).

Ordinal data can be ranked; nominal data should not be "averaged" as numbers.



Quantitative Data

- Quantitative data represents measurable values.
- Examples: Revenue (€), Profit (%), Units sold, Delivery time (days).
- Quantitative data supports: distribution, trends, relationships, and comparisons.
- But you must also know if it’s discrete or continuous.



Discrete vs Continuous (Practical Implications)

Discrete: counts (0,1,2,…) — Units, Orders, Visits.

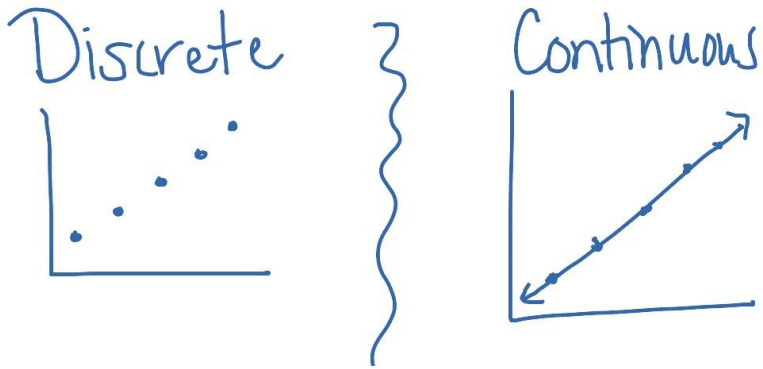
Continuous: measurable range — Revenue, Time, Temperature.

Implications for charts:

- Discrete counts with categories → bar charts often fit.

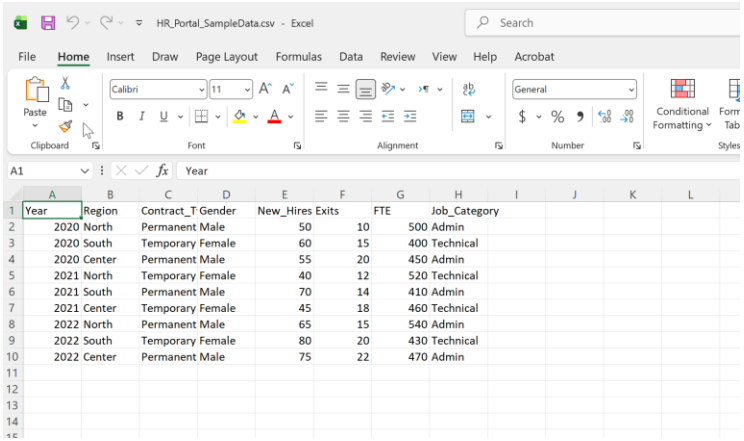
- Continuous measures → histograms/boxplots help show distribution.

- Continuous over time → line charts.



Structured vs Unstructured Data

- Structured data: rows/columns, clear schema (Excel, databases).
- Unstructured data: text, images, audio, logs.
- To visualize unstructured data, you usually need preprocessing:
- keyword extraction, sentiment coding, categorization, or aggregation.



Year	Region	Contract_T	Gender	New_Hires	Exits	FTE	Job_Category
2020	North	Permanent	Male	50	10	500	Admin
2020	South	Temporary	Female	60	15	400	Technical
2020	Center	Permanent	Male	55	20	450	Admin
2021	North	Temporary	Female	40	12	520	Technical
2021	South	Permanent	Male	70	14	410	Admin
2021	Center	Temporary	Female	45	18	460	Technical
2022	North	Permanent	Male	65	15	540	Admin
2022	South	Temporary	Female	80	20	430	Technical
2022	Center	Permanent	Male	75	22	470	Admin

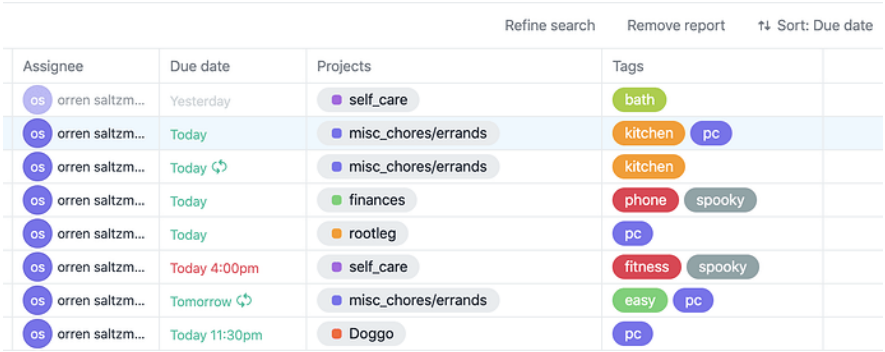
Common Traps in Data Types (Business Reality)

Three frequent traps:

1) **ID fields** look numeric but are often categorical labels.

2) Dates stored as text → wrong sorting (e.g., “2026-2” after “2026-10”).

3) Codes (e.g., 1=Low,2=Med,3=High) treated as real numeric scales.



Refine search Remove report ↑↓ Sort: Due date

Assignee	Due date	Projects	Tags
os orren saltzm...	Yesterday	self_care	bath
os orren saltzm...	Today	misc_chores/errands	kitchen, pc
os orren saltzm...	Today	misc_chores/errands	kitchen
os orren saltzm...	Today	finances	phone, spooky
os orren saltzm...	Today	rootleg	pc
os orren saltzm...	Today 4:00pm	self_care	fitness, spooky
os orren saltzm...	Tomorrow	misc_chores/errands	easy, pc
os orren saltzm...	Today 11:30pm	Doggo	pc

Question: What operations are meaningful?

Question: Which operations are meaningful? (Choose all that apply.)

A) Average Region

B) Count Orders

C) Average Satisfaction (Low/Med/High coded 1/2/3)

D) Sum Revenue

E) Compare Revenue across Channels

Group Activity 1 (8 minutes): Data Type Card Sort

- **Group Activity (teams of 3–5):**
  You get a list of dataset columns. For each column, label:
  1) Qualitative / Quantitative
  2) Structured / Unstructured
  3) Suggested chart type (if any)
- **Columns:** Region, Store, Channel, Category, Product, Date, Units, Revenue, Profit, Discount, Review_Text.

Debrief: The Rules You'll Use All Semester

1) Type first: label columns before charting.

2) Question first: compare, trend, distribution, relationship, composition?

3) Aggregate responsibly: sums/means must match variable meaning.

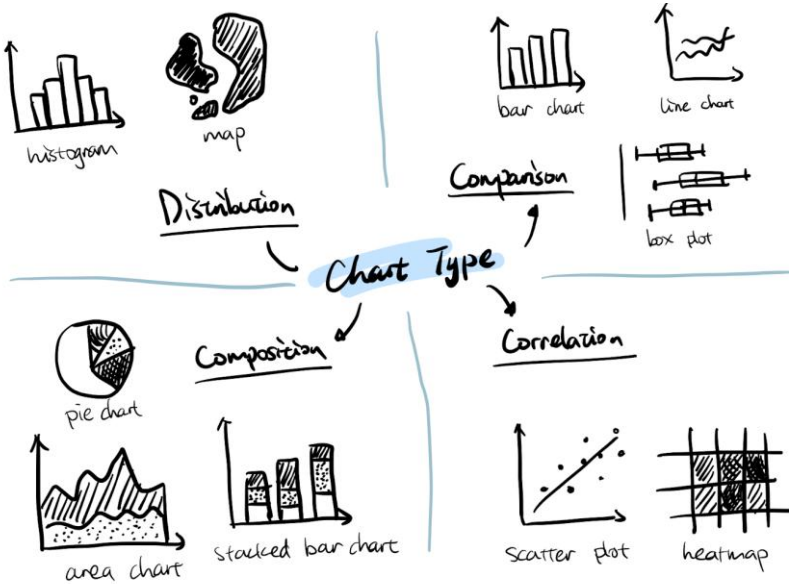
4) Design for reading: clarity beats decoration.

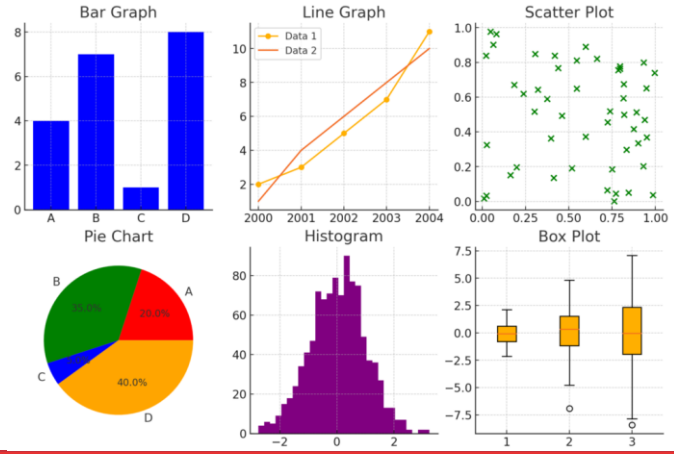
Types of Visualizations: Charts, Graphs, Maps, Infographics

The Four Families We Must Know

Today's taxonomy:

- Charts: common statistical displays (bar, line, pie, histogram).
- Graphs: relationship structures (scatter, networks).
- Maps: geographic reasoning (where patterns
- Infographics: narrative + design for broad audiences.





Why Some Charts “Feel Easier” to Read

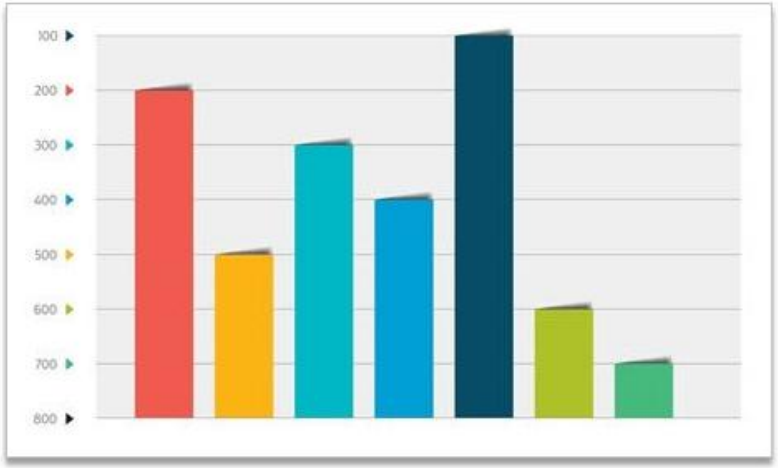
Human perception reads some encodings more accurately than others:

- Position and length are typically easier than angle/area.

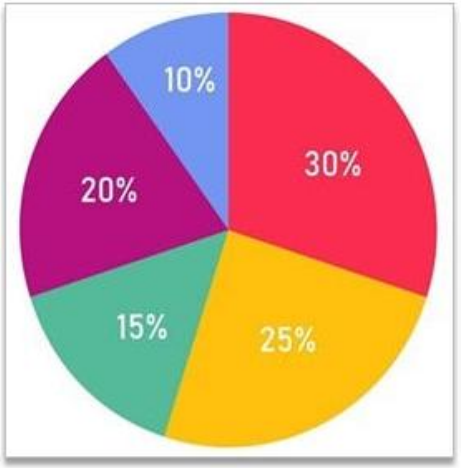
Practical translation:

- For precise comparisons, **bar charts often outperform pie charts**.

Source: Cleveland & McGill (Graphic perception research).



VS.

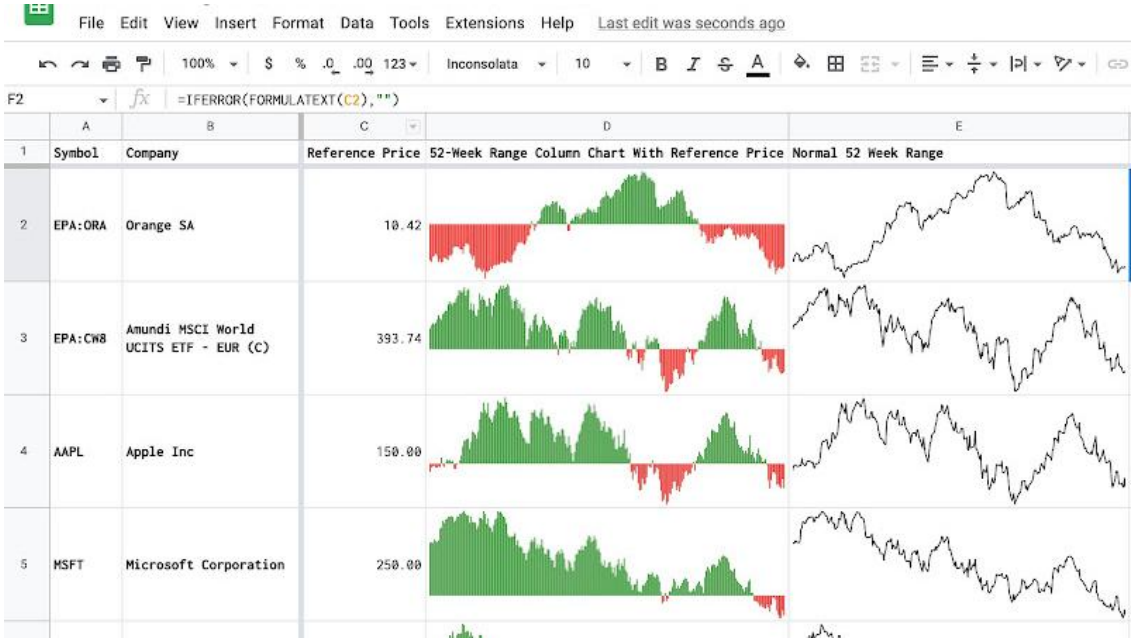


Charts: Line Chart (Trends Over Time)

Use lines when your goal is: **show change over time**.
Line charts emphasize: direction, volatility, seasonality.

Example (Monthly revenue €):

Jan 80k → Feb 90k → Mar 70k → Apr 110k

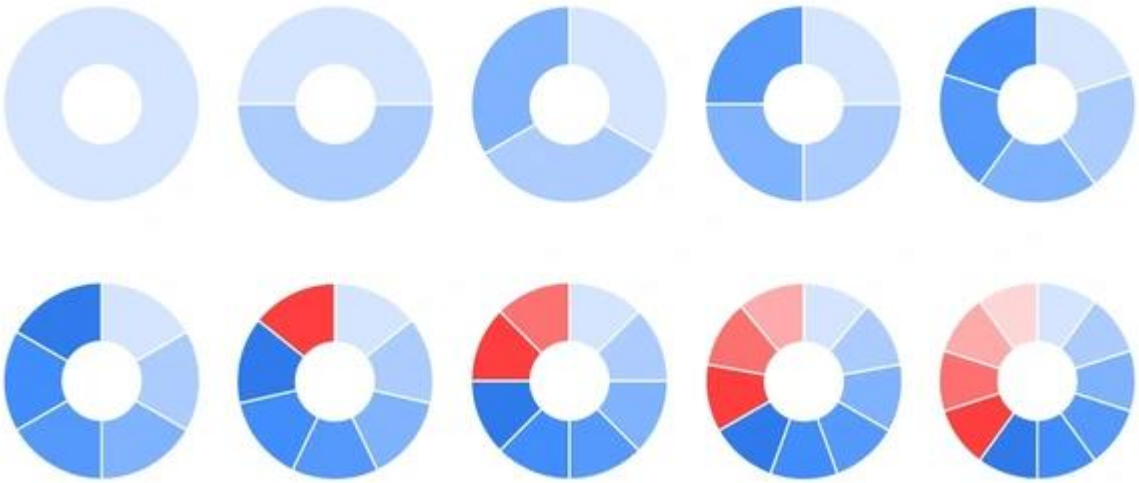


Charts: Pie/Donut (Composition) — Use Carefully

Use pie/donut only when:

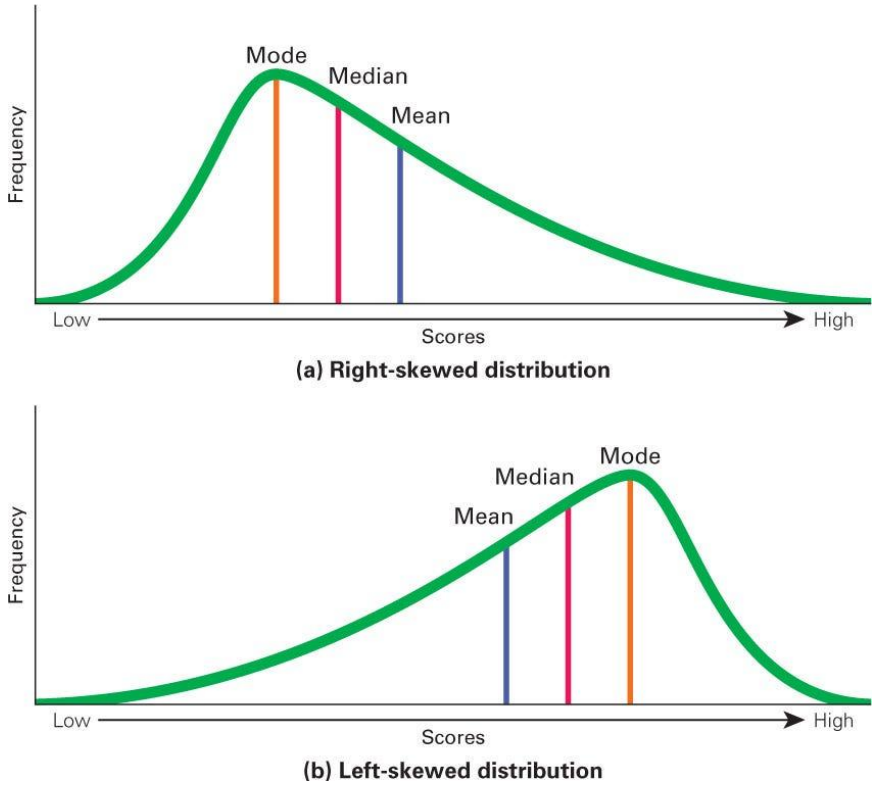
- Part-to-whole is the main question, AND
- Few categories (ideally 2–4), AND
- Precision is not the goal.

If many categories: prefer a sorted bar chart or 100% stacked bar.



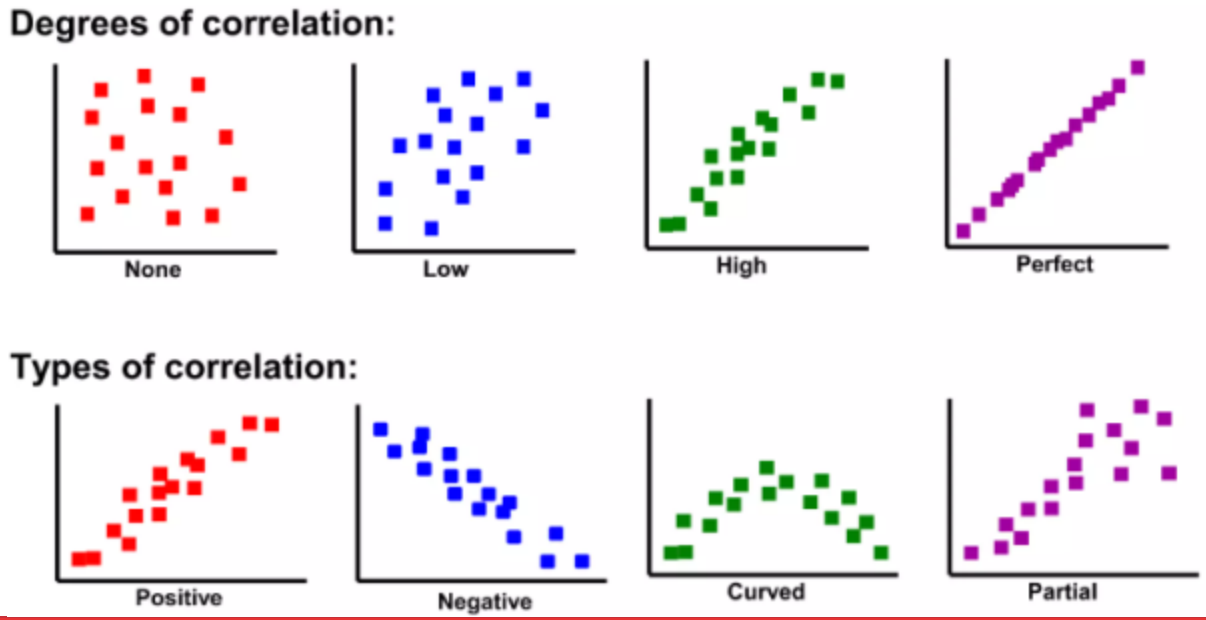
Charts: Histogram (Distribution)

- Histograms answer: “How are values distributed?”
- They reveal: skewness, clusters, outliers.
- Business example: distribution of order value (€).
- Averages hide shape; histograms show shape.



Graphs: Scatter Plot (Relationships)

- Scatter plots answer: “Do two variables move together?”
- Example: Discount (%) vs Profit (€).
- A scatter can reveal: correlation, non-linearity, clusters.
- Reminder: correlation ≠ causation.



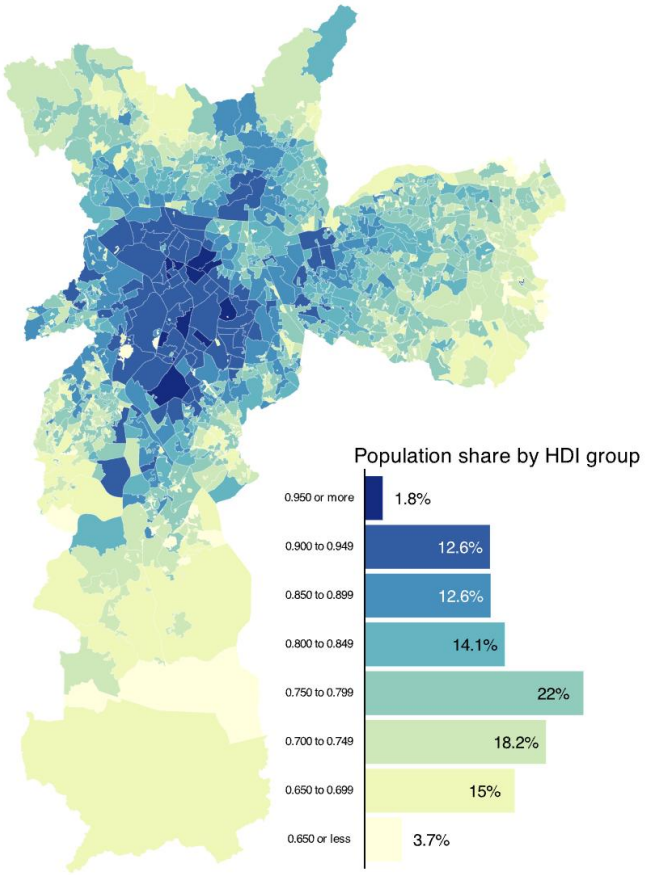
Maps: When Geography Is the Question

Use maps when: location is essential to the insight.

Examples:

- Revenue by district
- Delivery delays by region

Avoid maps when: you only need comparison and geography adds no value.



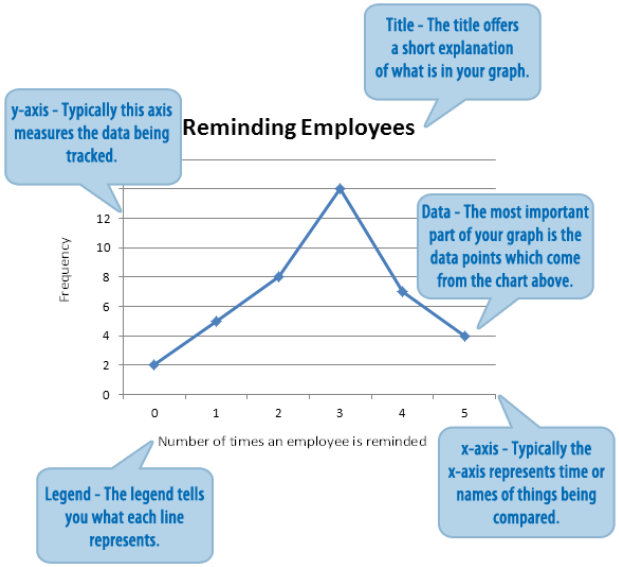
Infographics: Visual Story for Non-Technical Audiences

Infographics combine data + design + narrative.

They work best when you need:

- A single message
- A broad audience
- Context and explanation with minimal analysis steps

But they must still preserve accuracy: scales and numbers matter.



Question: CEO has 30 seconds—what do you show?

If you have 30 seconds with a CEO, which output is best?

A) Table of numbers
B) One clean chart
C) A map
D) A mini infographic
E) A dashboard screenshot

Question: Match question → visualization family

- Match each question to the best family:

Chart / Graph / Map / Infographic

1) “Which region is underperforming?”
2) “Is higher discount associated with lower profit?”
3) “How did revenue change monthly?”
4) “We need a one-page story for investors.”

Group Activity 2 (10 minutes): Choose the best visualization (and defend it)

Scenario A: Compare revenue across 12 categories.

Scenario B: Show revenue trend over 24 months.

Scenario C: Show relationship between discount and profit.

Scenario D: Show where delivery delays occur.

Task: For each scenario, pick a visualization and write a one-sentence justification.

Choosing the Right Visualization for Different Data Types

Start with the Question (Not the Chart)

1) What is the question? (compare / trend / distribution / relationship / composition)

2) What type of data supports it?

3) What chart reveals that pattern most clearly?

4) What design choices reduce errors and noise?

A Practical Decision Framework (4 steps)

Step 1 — Goal: Compare? Trend? Distribution? Relationship? Composition?

Step 2 — Data: Categorical vs numerical; time series; geographic.

Step 3 — Audience: technical team vs executive vs public.

Step 4 — Constraints: time, screen size, printing, LMS upload.

Chart-Selection Decision Table (Recommended + Anti-patterns)

Goal	Data type	Recommended	Avoid (anti-pattern)
Compare categories	Qual + Quant	Sorted bar, dot plot	Pie with many slices, 3D bars
Trend over time	Time + Quant	Line, area (careful)	Bars for long time series
Distribution	Quant	Histogram, boxplot	Pie, line chart
Relationship	Quant + Quant	Scatter, bubble (careful)	Dual axis “Franken-charts”
Composition	Parts of whole	100% stacked bar, pie (few cats)	Pie with >5 categories
Geographic pattern	Geo + measure	Choropleth, point map	Map for non-geo comparison

Anti-patterns Snapshot: Six Ways to Mislead Fast

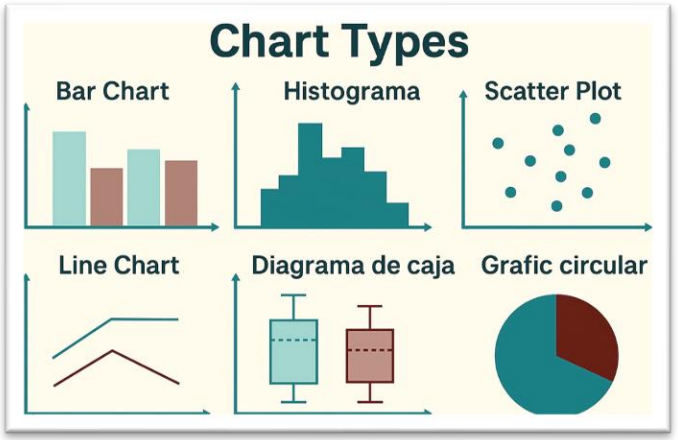
Common traps:

1) Truncated y-axis (bar chart exaggeration)
2) 3D effects (distortion)
3) Too many categories (unreadable)
4) Dual-axis comparisons (false relationships)
5) Color abuse (rainbow, low contrast)
6) Maps with raw counts (area bias; missing normalization)

In-class Question: Choose the chart (fast)

Pick the best chart (A/B/C/D) for each:

1) “Revenue by channel (Online/Store/Partner)”
2) “Monthly revenue for 2 years”
3) “Order value distribution”
4) “Discount vs Profit relationship”



Group Activity 3 (8 minutes): Write the headline first

Choose one scenario from Activity 2 and write:

1) A one-sentence insight headline (not just topic).

2) The chart that best supports that headline.

Example headline format:

“Online channel drove +18% growth in Q4 despite flat store sales.”

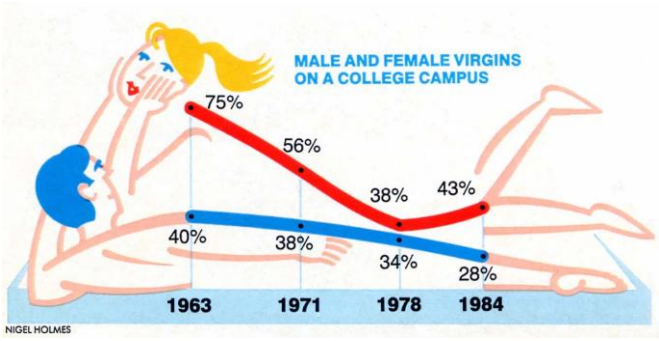
Data-to-Ink Ratio, Simplicity, and Visual Aesthetics

Data-to-ink asks: what fraction of visual ink represents *actual data*?

Replace decoration with information.

Practical rule: if it does not help interpret values, remove it.

Source: Tufte (Data-ink / chartjunk principle).



Excel Defaults That Create Chartjunk (and what to do)

5-second test: Can someone understand the main message in 5 seconds?

If not, simplify:

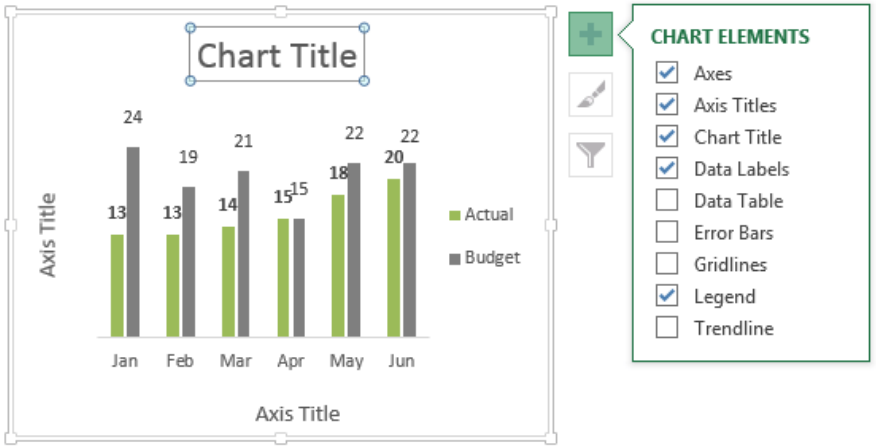
- Reduce categories
- Sort bars
- Highlight only what matters
- Use one clear takeaway in the title

Visual Aesthetics: Hierarchy, Typography, Alignment

Professional visuals follow a hierarchy:

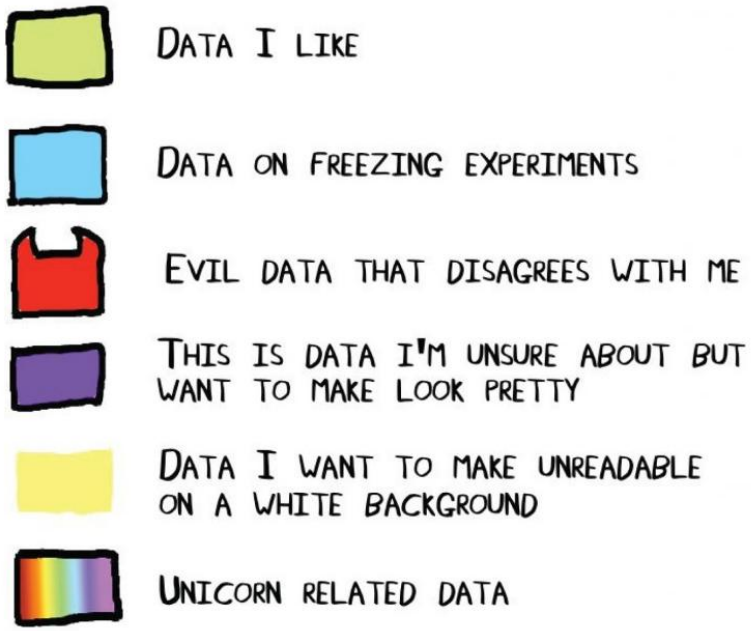
1) Insight title (sentence)
2) Chart (supports the claim)
3) Labels + units
4) Small source note if needed

Typography rules: consistent font, consistent sizes, avoid all-caps, maintain whitespace.



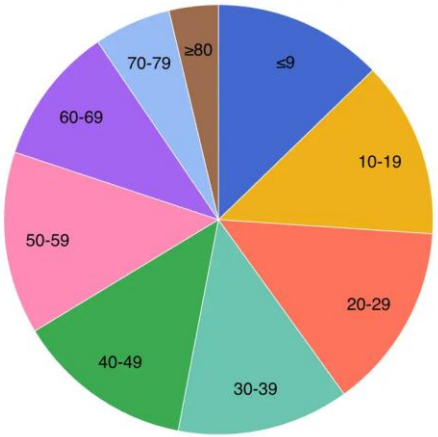
Color & Accessibility (Contrast Matters)

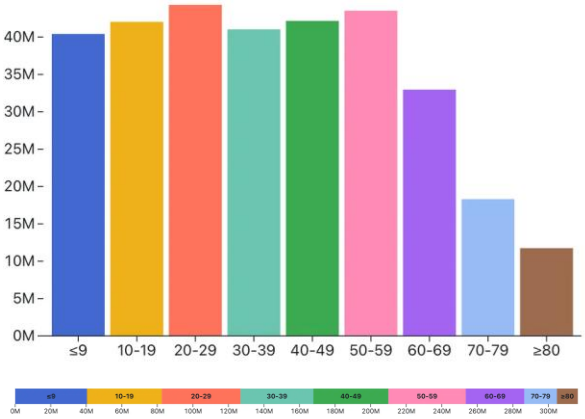
- Color should encode meaning, not decorate.
- Avoid: rainbow palettes, low-contrast text, meaning by color only.
- Do: use limited palette + one highlight color + direct labels.
- Accessibility note: some students or audiences may have color-vision deficiency.

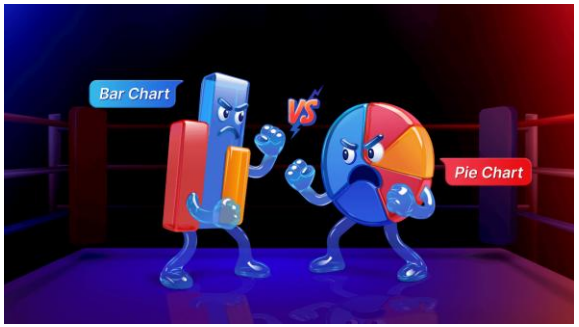


Good vs Bad #1: 3D Pie vs Sorted Bar

Bad: 3D pie, many slices, hard comparison.
Good: sorted bar chart with clear labels.
Critique (why bad): angle/area comparison + 3D distortion + clutter.



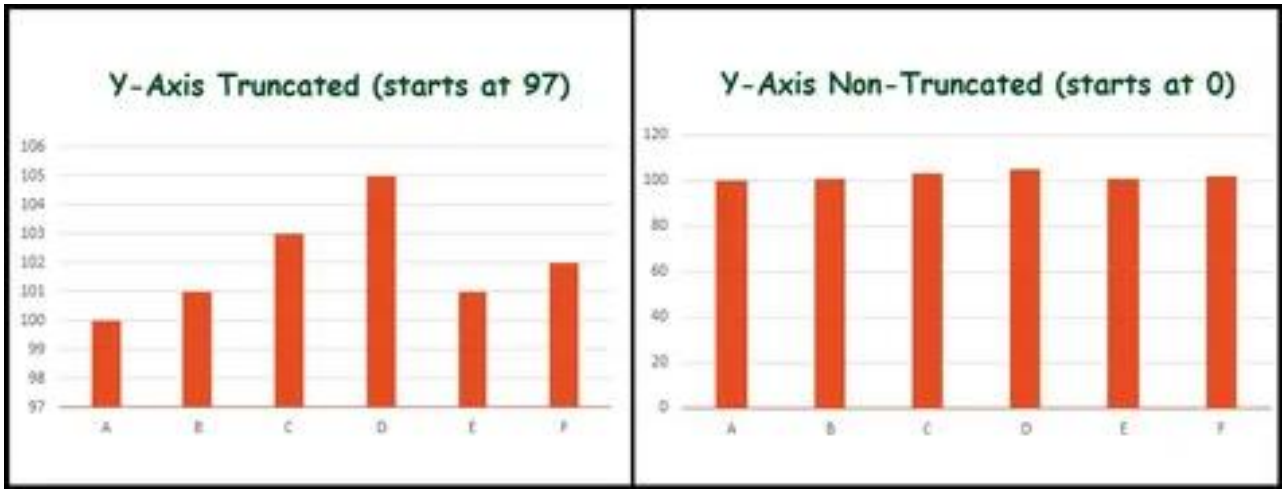




The breakdown of the U.S.'s 2015 population into ten-year age groups.

Good vs Bad #2: Truncated Y-axis (Exaggeration)

Bad: bar chart with y-axis starting at 97 instead of 0 → exaggerates differences.
Good: bar chart starting at 0 + labels.
Critique: truncated baseline changes perceived magnitude.



Good vs Bad #3: Dual Axis “Franken-chart”

Bad: two y-axes → false correlation impressions.

Good: small multiples or separate charts aligned by time.

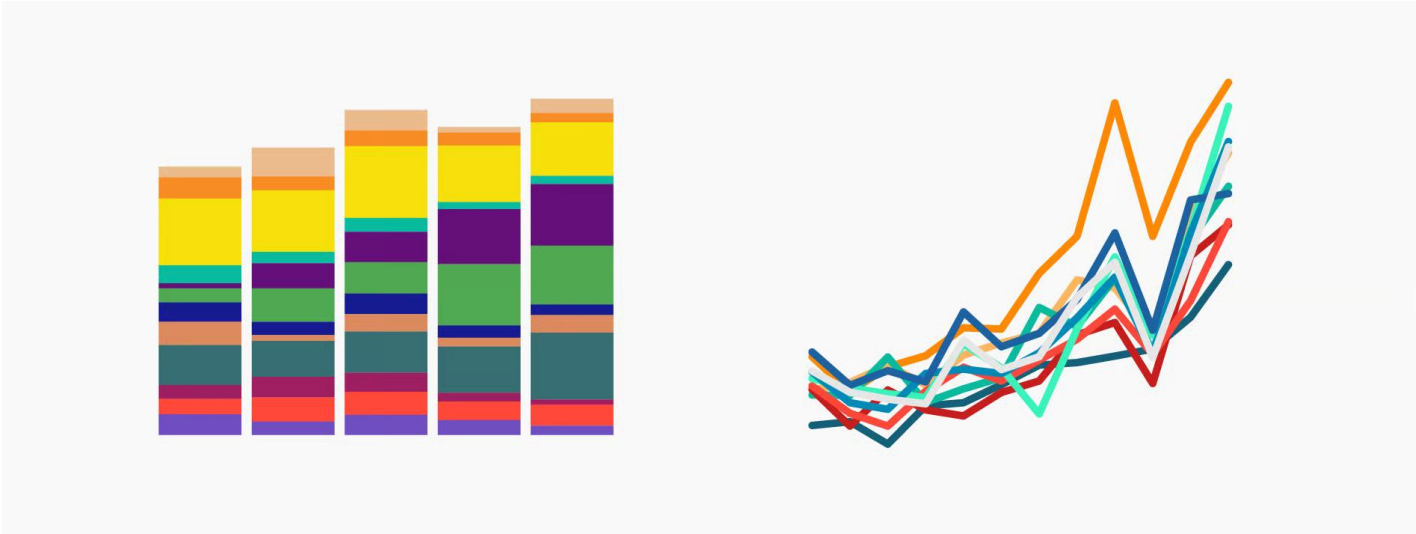
Critique: viewer confuses scale alignment with relationship.

Good vs Bad #4: Rainbow Color Abuse

Bad: too many colors with no meaning.

Good: neutral color for most categories + highlight one focus category.

Critique: color should guide attention, not create noise.



Good vs Bad #5: Too Many Categories + No Sorting

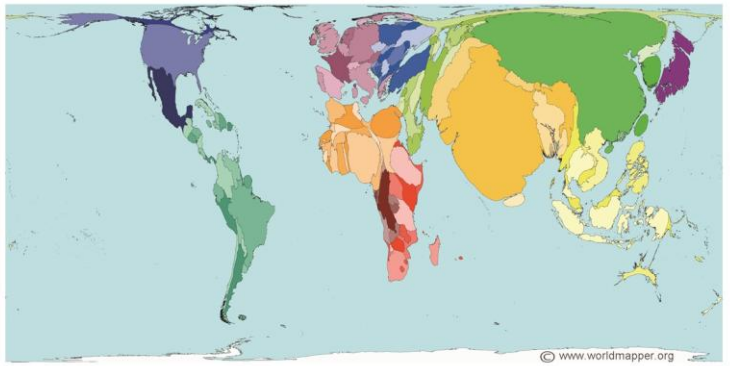
Bad: 25 categories unsorted, rotated labels.
Good: top 10 sorted + “Other” grouped, readable labels.
Critique: readability collapses when the chart exceeds human scanning capacity.

Good vs Bad #6: Map Misuse (Counts vs Rates)

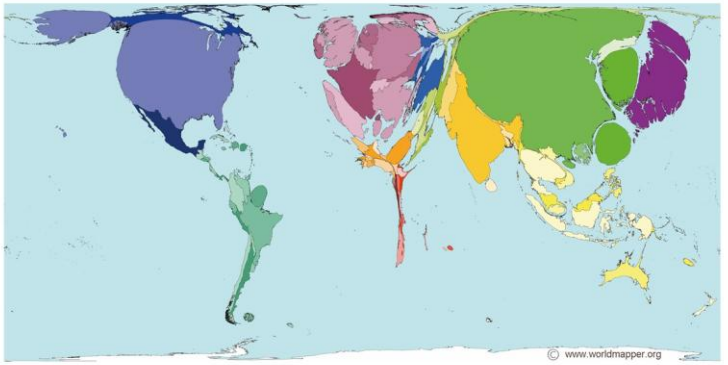
Bad: choropleth of raw counts → large areas "look" important.
Good: normalized rate (e.g., revenue per store, per capita) or a sorted bar chart.

Critique: maps require careful normalization to avoid area bias.

Population by Country (2015)



Wealth by Country (2015)



Questions

- Thank you!
- Questions and discussion.