Lesson 20: Data Storytelling

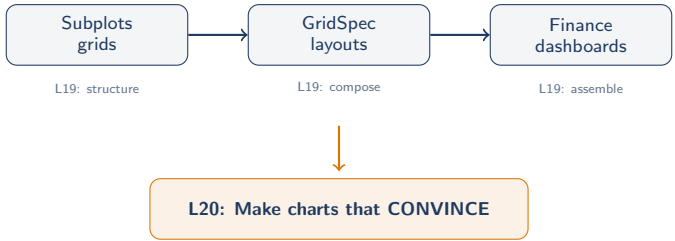
Data Science with Python – BSc Course

Data Science Program

BSc Course

45 Minutes

Previously on Data Science...



L19 built: multi-panel dashboards **L20 asks:** how do we make them *tell a story*?

A journalist needs headline, evidence, conclusion – so does your chart

Learning Objectives

After this lesson, you will be able to:

1. Apply narrative arc structure to data visualizations
2. Select the right chart type for any communication goal
3. Use color strategically to guide viewer attention
4. Transform bad charts into effective ones (before/after)
5. Build executive summary dashboards for stakeholders

Module 3 capstone: from raw numbers to compelling visual narratives

The Chart That Tells the Wrong Story

A beautiful chart that tells the wrong story is worse than an ugly chart that tells the right one.

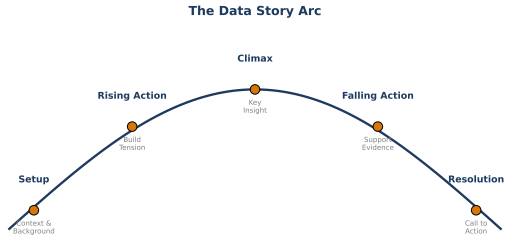
- A 3D pie chart hides the real proportions
- A dual-axis line chart implies a relationship that may not exist
- Rainbow colors distract instead of informing

Think like a journalist:

- **Headline:** what is the ONE takeaway?
- **Evidence:** which data supports it?
- **Conclusion:** what should the reader do?

Data storytelling is where analysis meets persuasion

The Narrative Arc for Data

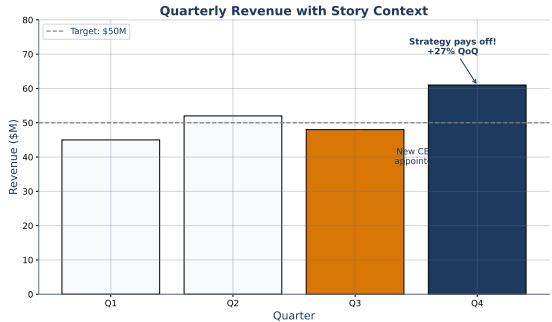


Every data story follows: Context → Rising Action → Climax → Resolution

---

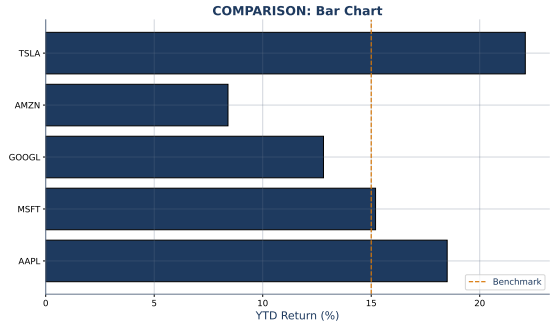
Start with what the audience knows, build tension, reveal the insight

Annotations Tell the Story



Annotations guide the viewer to key insights – don't make them search

Chart Selection: Comparison



"How do sectors compare?" → Bar chart

Sorted bars make the ranking instantly clear

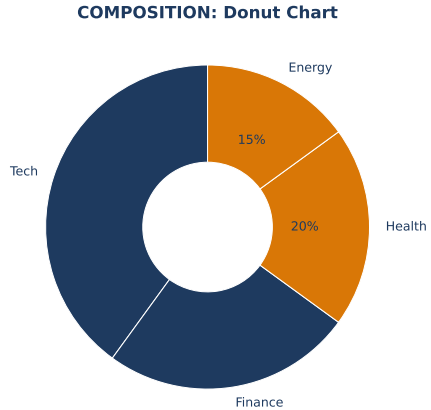
Bar charts excel at categorical comparisons – always sort meaningfully

Chart Selection: Trend Over Time

"How did performance evolve?" → Line chart (time flows left to right)

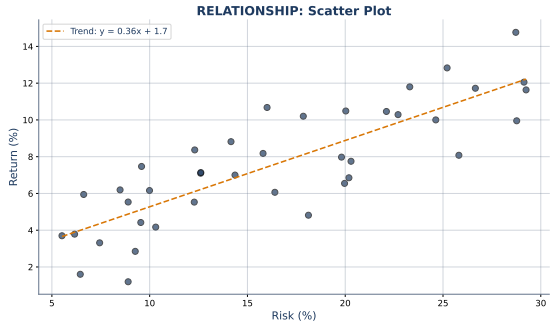
Line charts show trends – the x-axis MUST be time or ordered sequence

Chart Selection: Parts of a Whole



Donut/pie for composition – limit to 5–7 slices, use only when parts sum to 100%

Chart Selection: Relationship



"Are risk and return related?" → Scatter plot

Add regression line to quantify the trend

Scatter plots reveal relationships – add transparency for overplotting

The Complete Chart Selection Guide

Chart Selection Matrix: Quick Reference

Chart Type	Best For	Avoid When	Finance Example
Bar Chart	Comparing categories Ranking values	More than 10 categories Time series data	*Sector performance*
Line Chart	Trends over time Continuous data	Categorical X-axis Few data points	*Stock price history*
Histogram	Distribution shape Frequency analysis	Comparing groups Small samples	*Return distribution*
Pie/Donut	Part-to-whole Simple proportions	Precise comparison More than 5 slices	*Portfolio allocation*
Scatter Plot	Relationships Correlation patterns	Categorical variables Too many points	*Risk vs return*
Heatmap	Matrix data Correlations	Non-matrix data Few variables	*Asset correlations*

Match your MESSAGE to the chart type – not the other way around

Checkpoint: The CEO Asks...

CEO: "How did we perform this year?"

A: Pie chart of sector weights	B: Line chart portfolio vs benchmark	C: Histogram of daily returns

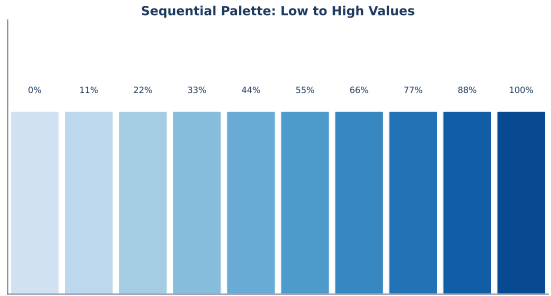
Think: the question is about TREND over TIME

A shows composition, C shows distribution

Only B answers "how did we perform?"

Answer: B – a line chart showing cumulative returns vs benchmark over time

Color Strategy: Sequential Palette

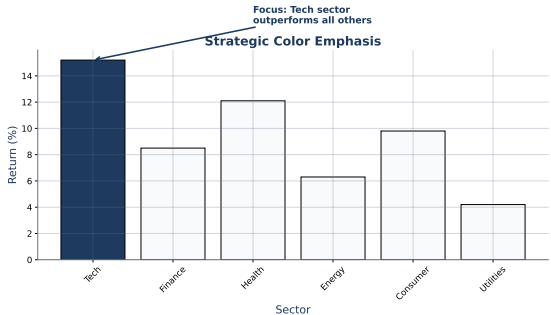


Rule: light-to-dark for ordered data (low to high, cold to hot)

Use viridis, Blues, or Oranges for continuous scales

Sequential palettes encode magnitude – darker means more

Color Strategy: Emphasis

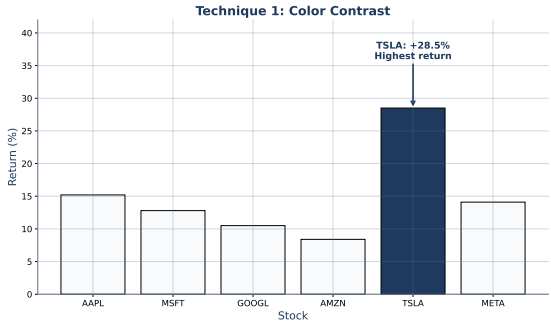


Rule: one bright color for the hero, gray for everything else

Don't let supporting data compete with your main message

Pre-attentive processing: the eye finds the colored bar in milliseconds

Three Ways to Emphasize



Color: bright vs gray **Size:** large vs small **Position:** top-left draws the eye first

Combine for maximum impact – but never use all three on different elements

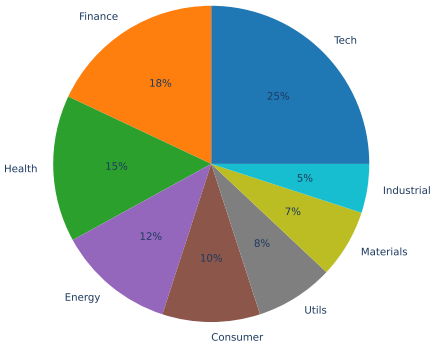
---

Emphasis directs attention – use it to answer "what should I look at first?"

Before/After: Pie Chart Makeover

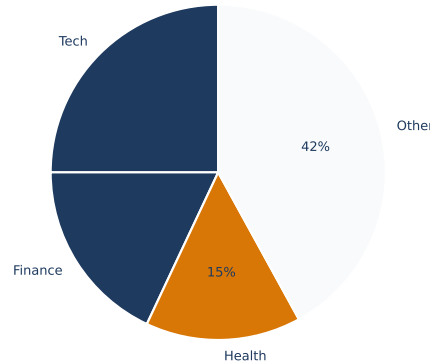
Before

BEFORE: Cluttered - Too Many Segments



After

AFTER: Focused - Grouped Small Segments



Better colors, direct labels, emphasis on key slice – small changes, big impact

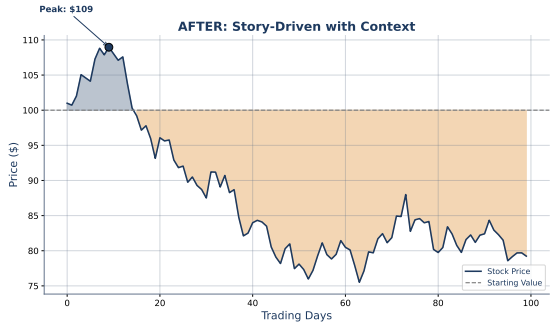
Before: Cluttered Line Chart

Problems: too many lines, no emphasis, rainbow colors, no story

---

When everything is highlighted, nothing is highlighted

After: Focused Line Chart

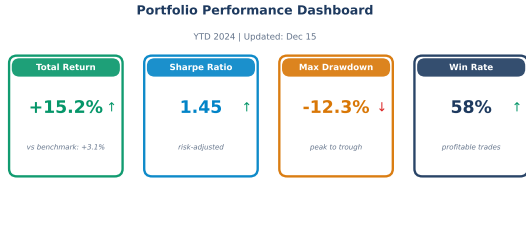


Fixes: one hero line in color, others in gray, clear annotation, title as insight

---

The after version answers a question; the before version just shows data

Executive Summary: KPI Cards



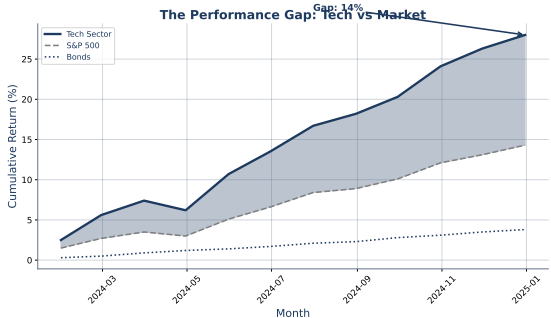
Lead with the numbers: Total Return, Sharpe Ratio, Max Drawdown, Win Rate

Stakeholders read the KPI cards first, then look at the charts

---

KPI cards answer "how are we doing?" before the audience even asks

Finance Story: The Performance Gap



The narrative: "We underperformed by 3.2% – here's why and what we're changing"

A chart without a story is just decoration

Every finance chart should answer: so what? what next?

Hands-On: Chart Makeover Challenge

Task: Transform a bad chart into a compelling visual story.

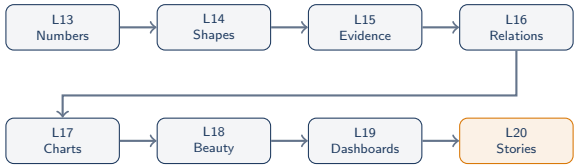
The bad chart: 8-line rainbow spaghetti plot of sector returns

Your makeover:

1. Pick the ONE sector with the best story (biggest gain or loss)
2. Color that sector in amber, everything else in light gray
3. Add a title that states the insight (not "Sector Returns")
4. Add one annotation at the key moment
5. Add a `bottomnote` with the "so what?"

A good chart title is a headline: "Tech led all sectors with 42% gains"

Module 3 Complete: The Data Spoke



"The data spoke.
Here's what
it said."

Your journey: descriptive stats → distributions → hypothesis tests → correlation → matplotlib → seaborn → dashboards → **stories**

You can now turn any dataset into a visual argument – that's data science

Key Takeaways

What you learned today:

1. **Narrative arc:** context → tension → insight → action
2. **Chart selection:** match chart type to your *message*, not your data
3. **Color strategy:** one hero color, gray for context, sequential for magnitude
4. **Before/after:** small changes create dramatic improvements
5. **Executive summary:** lead with KPI cards, support with focused charts

The golden rule: every chart should answer one question clearly

---

If you can't state the insight in one sentence, the chart needs a redesign

Next: Module 4 – Supervised Learning: Regression

Module 3 gave you the tools to SEE data.
Module 4 teaches you to PREDICT with it.

Coming up:

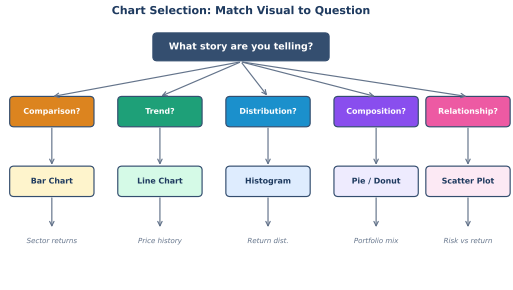
- L21: Linear Regression – fitting a line to data
- L22: Regularization – preventing overfitting
- L23: Regression Metrics – measuring prediction quality
- L24: Factor Models – multi-factor finance models (CAPM, Fama-French)

The big shift: from describing what happened to predicting what will happen

---

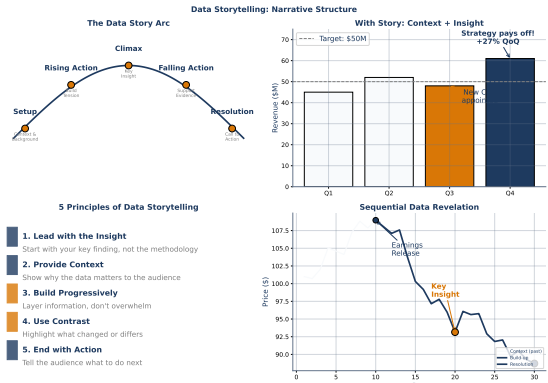
Module 4 begins the machine learning journey – bring your visualization skills!

Self-Study: Chart Selection Flowchart



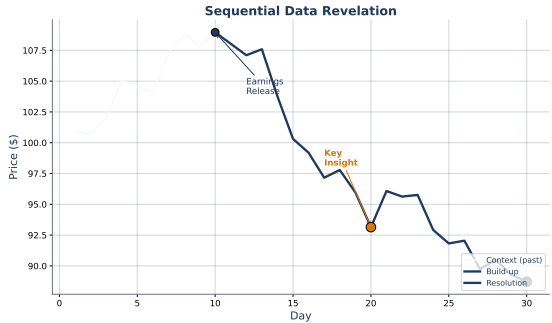
Decision tree for choosing the right visualization type

Self-Study: Storytelling Flow



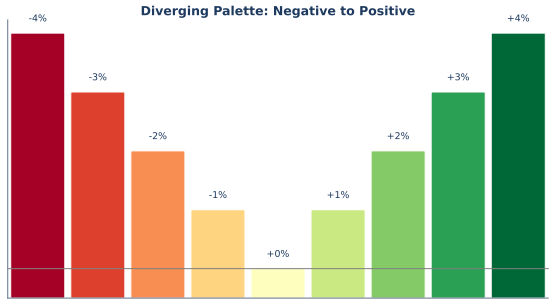
The complete storytelling workflow from data to presentation

Self-Study: Sequential Revelation



Build up complexity progressively – don't show everything at once

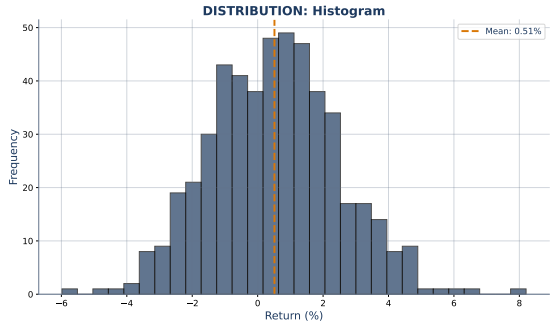
Self-Study: Diverging Palette



Use for: Returns, Changes, Deviations from baseline

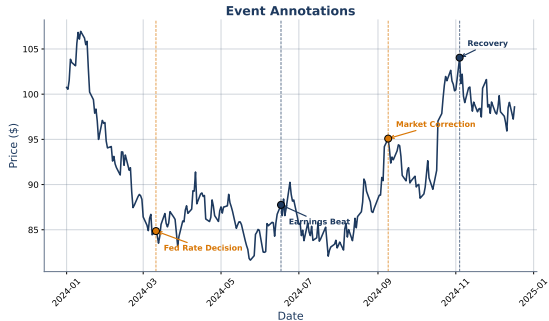
Use diverging colors when data has a meaningful midpoint (e.g., 0% return)

Self-Study: Distribution Histogram



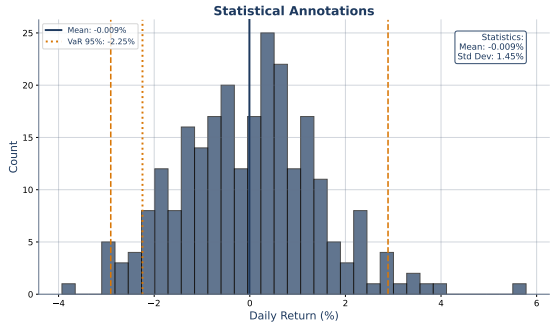
Histograms reveal the shape of data – skewness, outliers, modes

Self-Study: Event Annotations



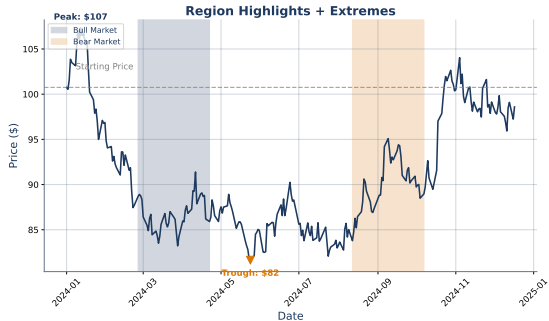
Mark important events with vertical lines and text labels

Self-Study: Statistical Annotations



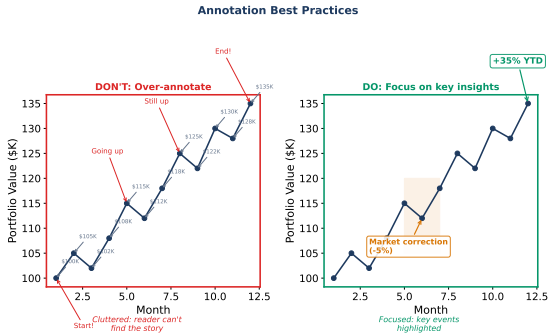
Add mean, median, thresholds as reference lines for context

Self-Study: Region Highlights



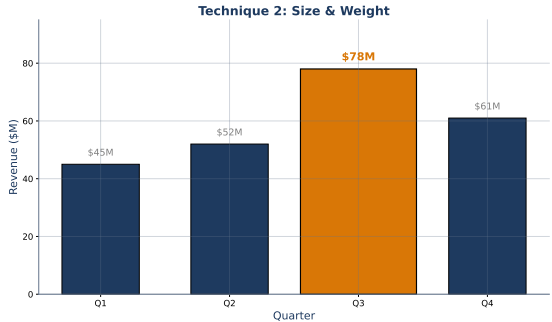
Use axvspan() to shade regions of interest (recessions, events)

Self-Study: Annotation Dos and Don'ts



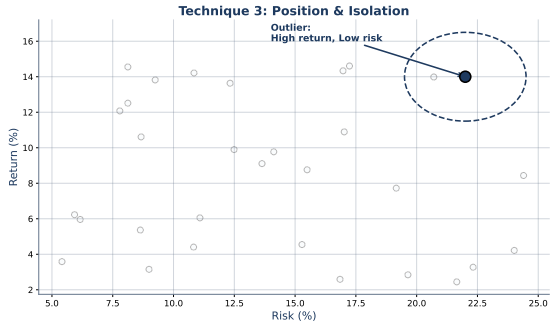
Annotate insights, not data points – less is more

Self-Study: Size Emphasis



Larger elements receive more attention – use for key data points

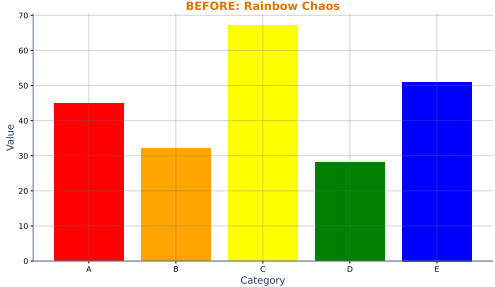
Self-Study: Position Emphasis



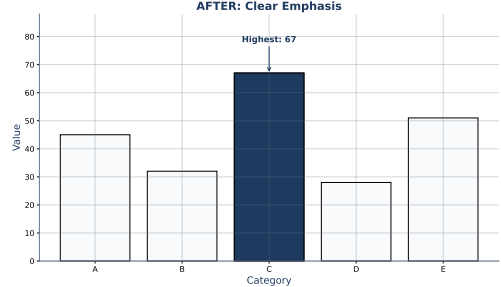
Top-left and center positions draw the eye first (Gutenberg diagram)

Self-Study: Bar Chart Before/After

Before



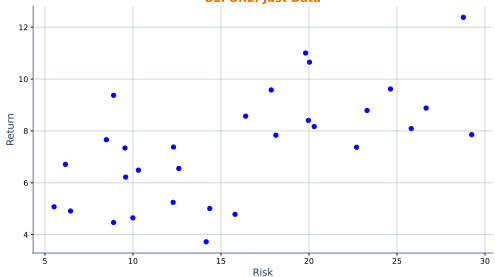
After



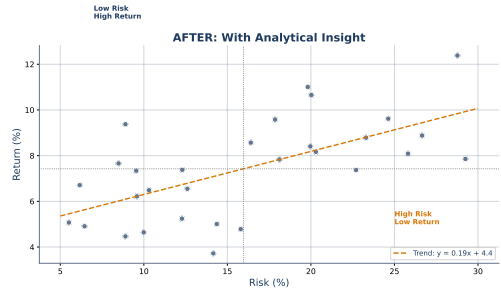
Remove 3D effects, sort bars, add data labels, use consistent colors

Self-Study: Scatter Plot Before/After

Before

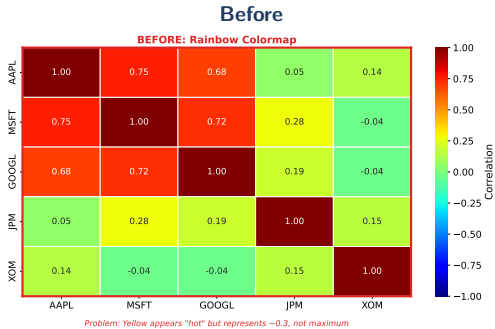


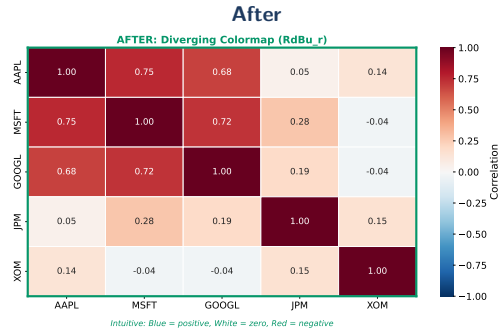
After



Add transparency for overplotting, trendline for pattern, annotations for outliers

Self-Study: Heatmap Before/After

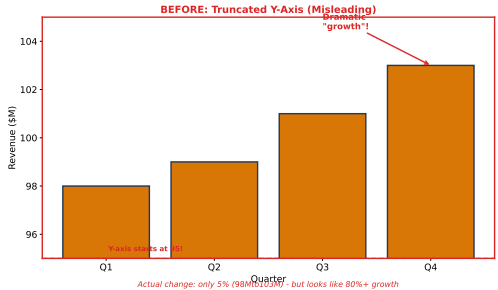




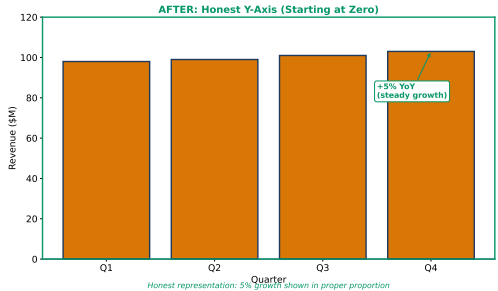
Use diverging colormap centered at zero, add annotations, clean labels

Self-Study: Axis Manipulation Before/After

Before



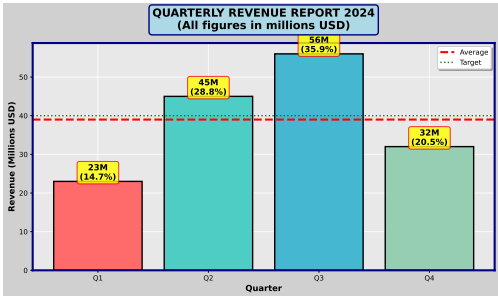
After



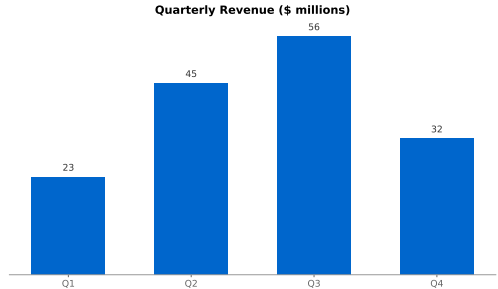
Truncated axes exaggerate differences – always start bar charts at zero

Self-Study: Data-Ink Ratio

Before

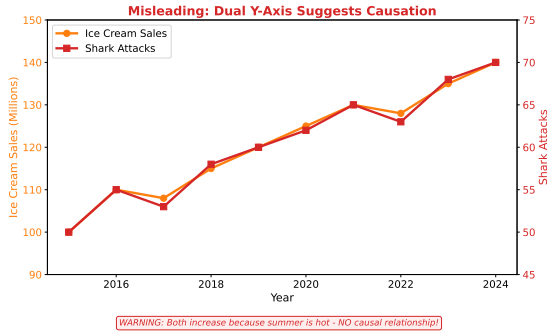


After



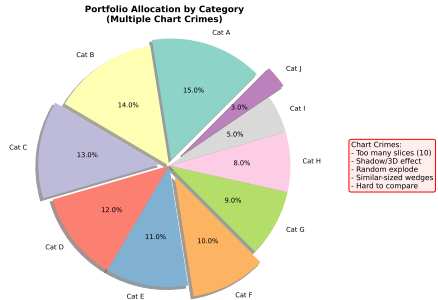
Tufte's data-ink ratio: maximize data, minimize non-data ink

Self-Study: Dual Axis Manipulation



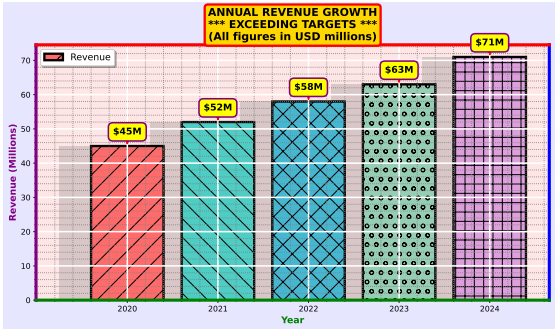
Dual axes can imply false correlations – use with extreme caution

Self-Study: Pie Chart Crimes



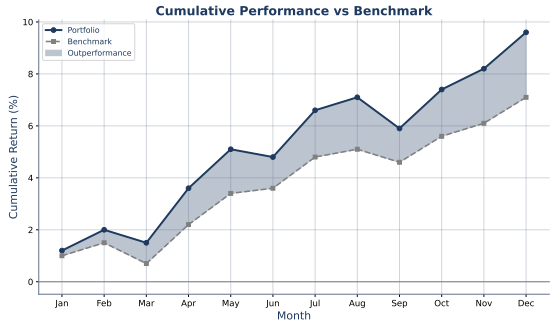
3D pie charts distort area perception – never use them

Self-Study: Chartjunk Example



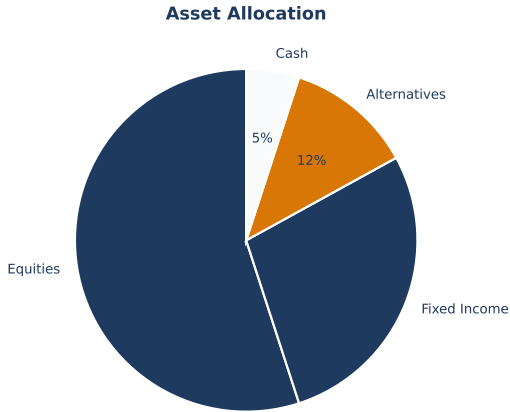
Gratuitous decoration that adds no information – remove it all

Self-Study: Performance Chart



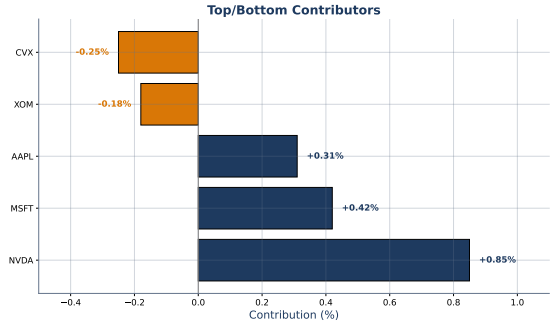
Show trend vs target or benchmark for executive reporting

Self-Study: Allocation Pie



Portfolio composition visualization for stakeholder reports

Self-Study: Top/Bottom Contributors

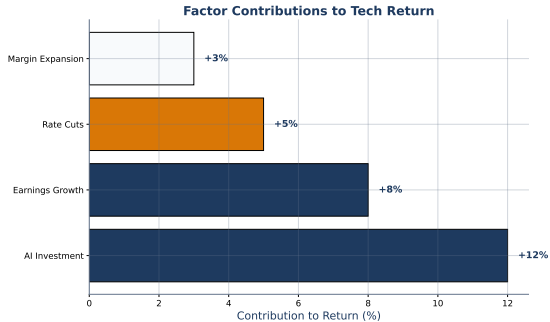


Highlight winners and losers – horizontal diverging bar chart

Self-Study: Risk Metrics Dashboard

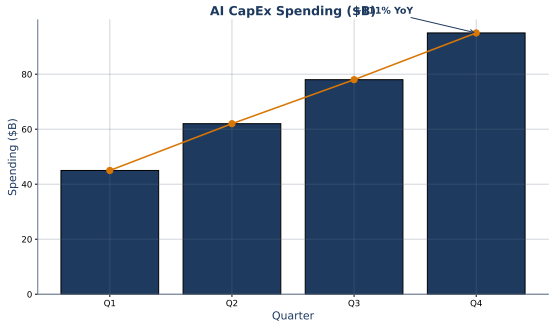
Risk dashboard combining VaR, volatility, drawdown for stakeholder review

Self-Study: Factor Contributions



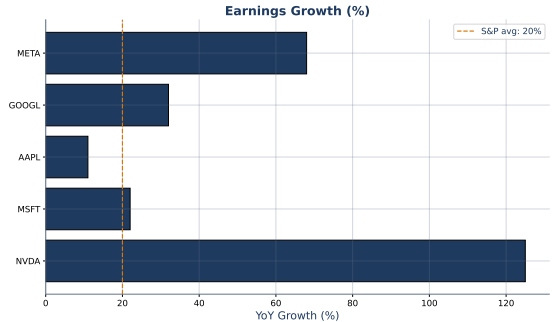
Attribution analysis explains sources of return – waterfall chart

Self-Study: AI CapEx Trends



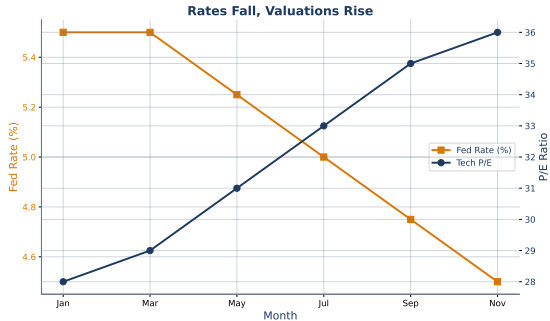
Tell a story with industry data – annotate the inflection points

Self-Study: Earnings Growth



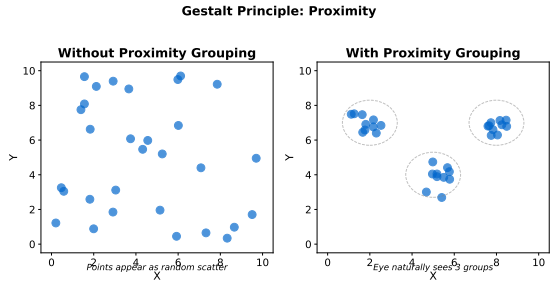
Narrative: growth drivers and outlook – combine actual and forecast

Self-Study: Rates and Valuations



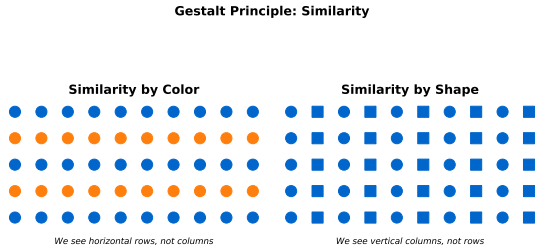
Connect macro factors to market valuations – the big picture story

Self-Study: Gestalt – Proximity



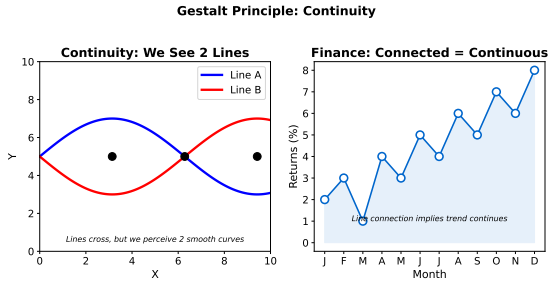
Items placed close together are perceived as a group

Self-Study: Gestalt – Similarity



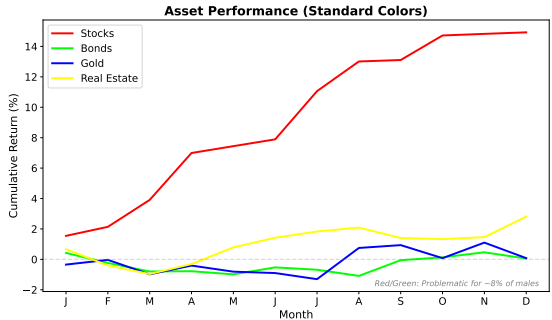
Similar items (color, shape, size) are perceived as belonging together

Self-Study: Gestalt – Continuity



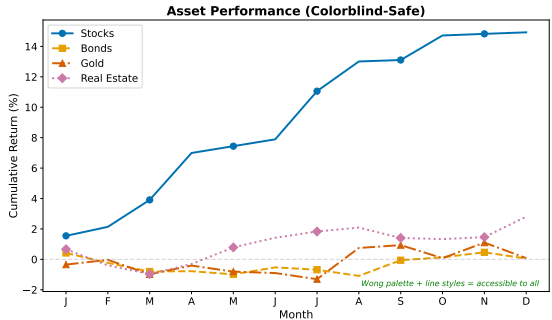
The eye follows smooth lines and curves – use for flow and direction

Self-Study: Colorblind Original



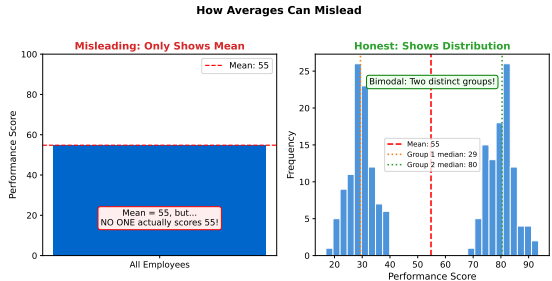
This chart uses red-green – problematic for 8% of males

Self-Study: Colorblind Simulation



Same chart simulated for deuteranopia – use viridis or add patterns

Self-Study: Misleading Statistics



Same data, different stories – always disclose your design choices

Self-Study: Narrative Sequencing

Order your charts to build a logical argument – setup, evidence, conclusion