

Data Transformations: Scale, Units, and Interpretation



- Raw numbers mislead without context, scale, and comparability assumptions



- The transformation triad: scale change, unit conversion, interpretive boundaries



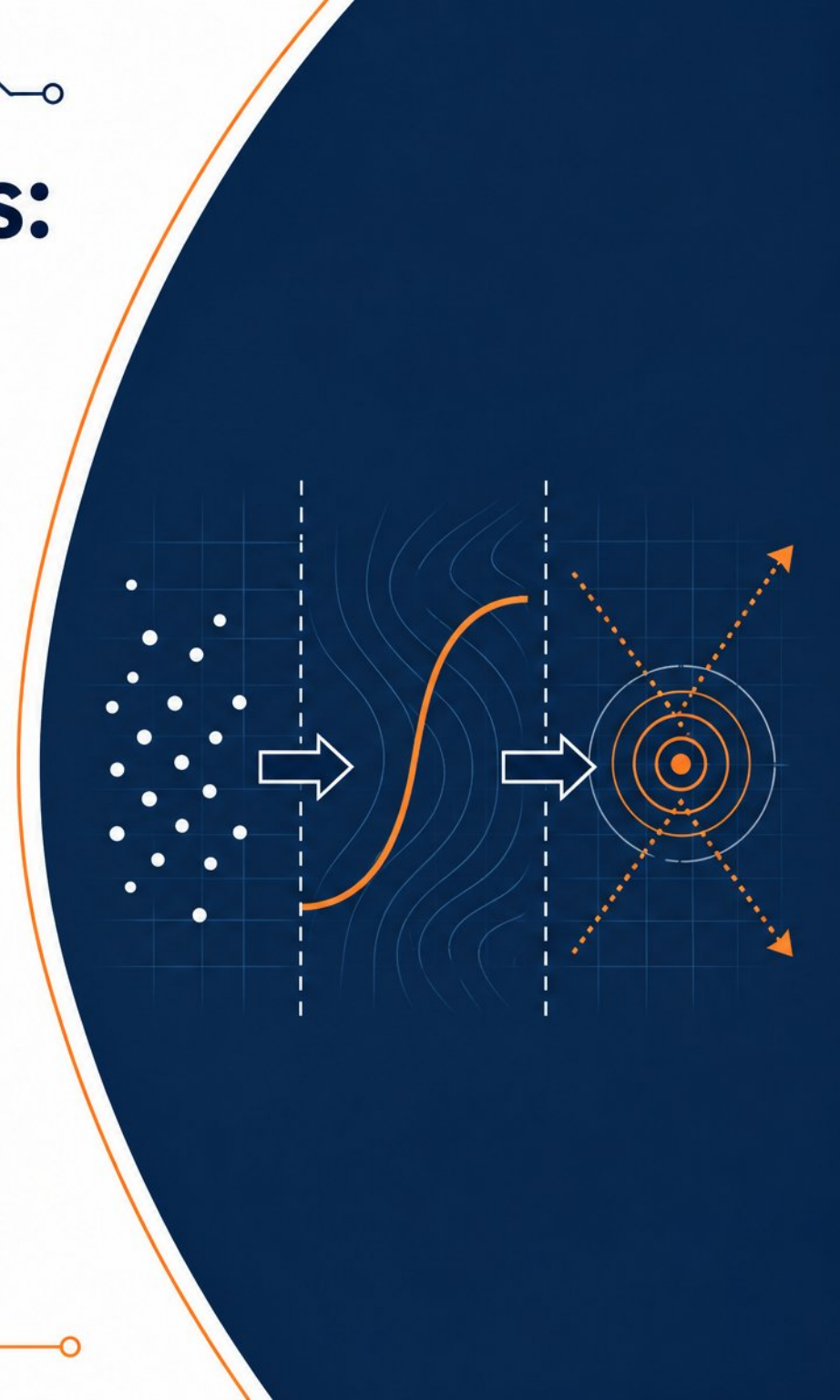
- Real examples where transformation choices reversed outcomes



- Taxonomy: linear scaling, nonlinear transforms, unit conversions, indexing



- Roadmap: core principles through hands-on application



Why Transform? Goals, Principles, and the Decision to Intervene



- Enable fair comparison and stabilize variance



- Balance accuracy gains against loss of direct meaning



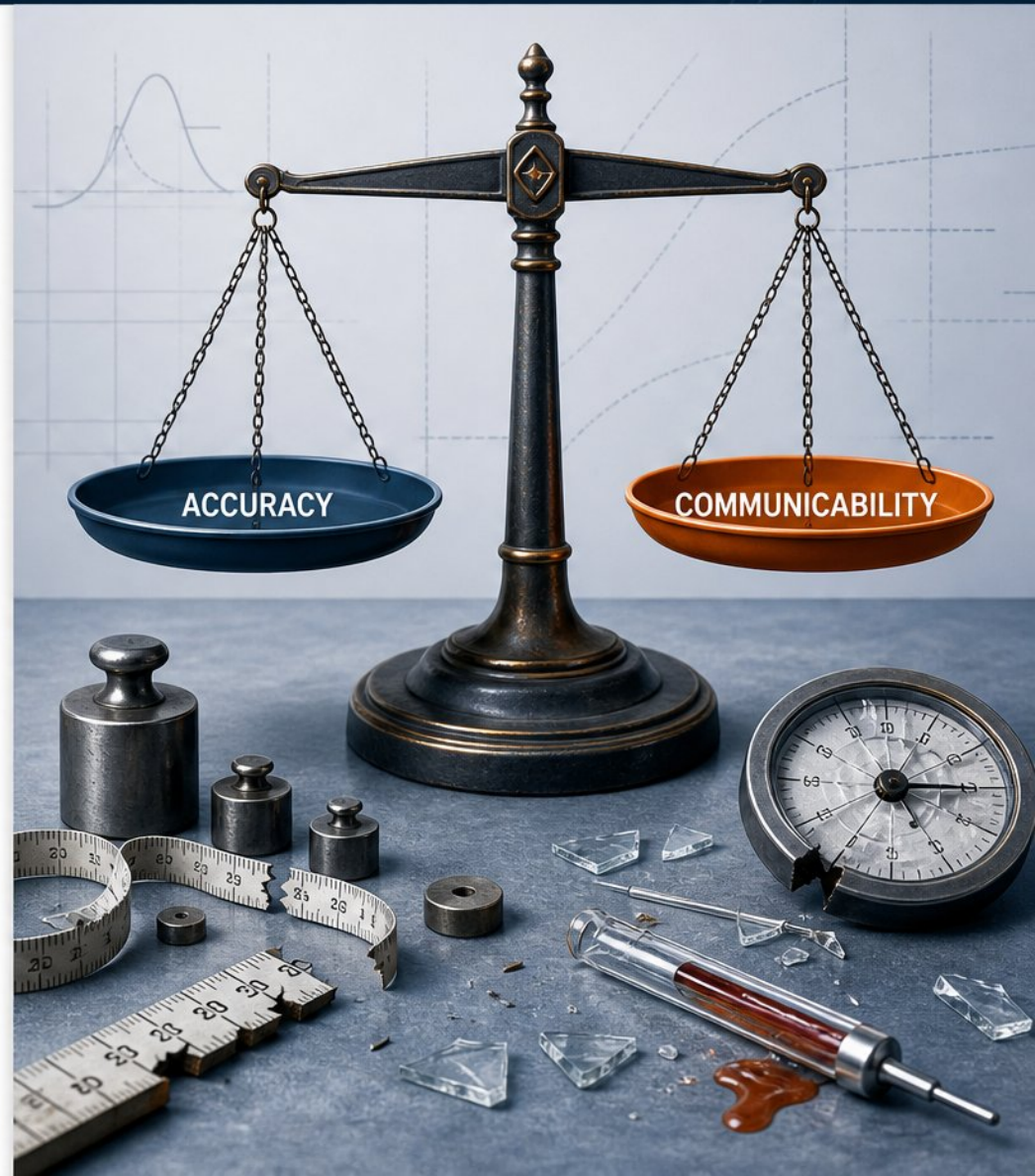
- Essential for heteroscedasticity, skew, multiplicative data



- Avoid when scales are interpretable or bias is severe

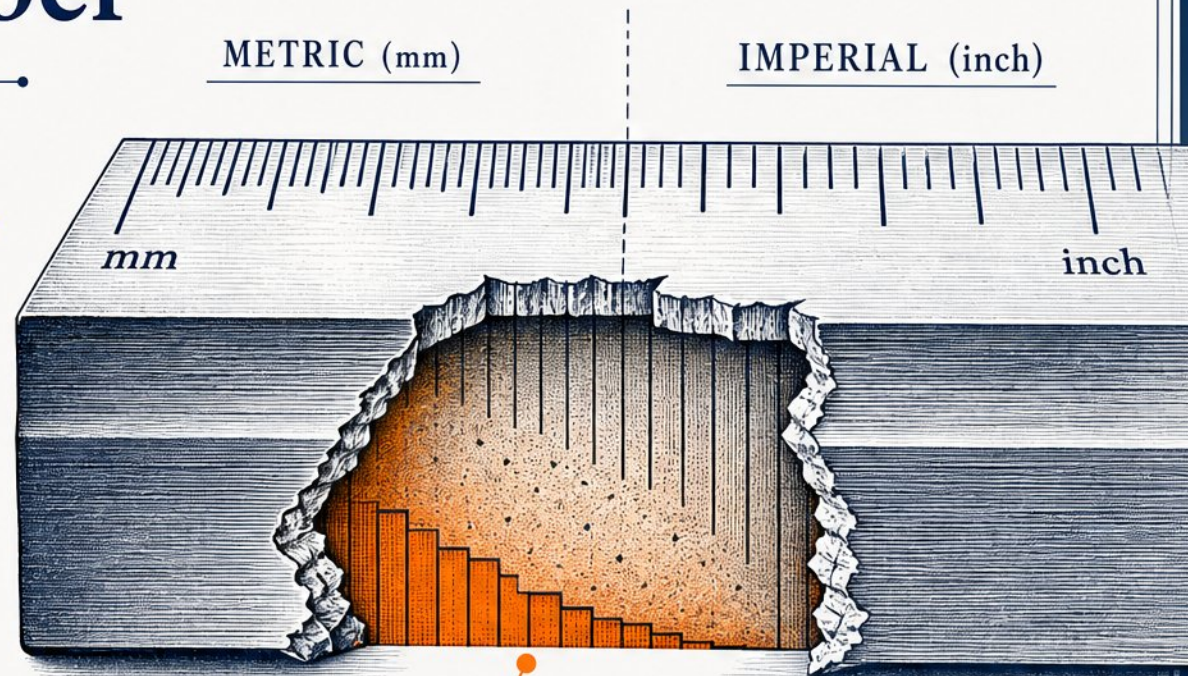


- Anchor every transformation to the core decision



Units of Measurement: The Hidden Framing of Every Number

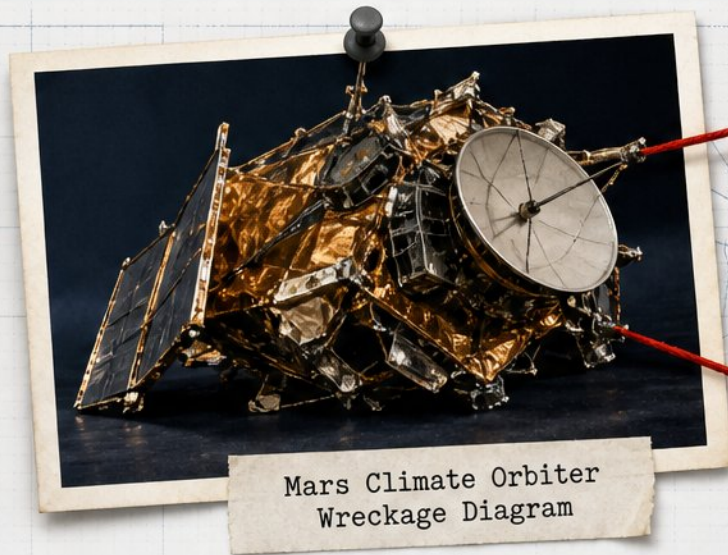
- Absolute vs. relative units shape what comparisons are even possible
- Prefixes (kilo, mega, micro) create magnitude intuition—or mislead it
- Conversion risks: rounding errors can cascade into costly failure
- Derived units: per capita, rates, indexes embed hidden assumptions
- Nominal vs. real values: inflation adjustments that rewrite the story



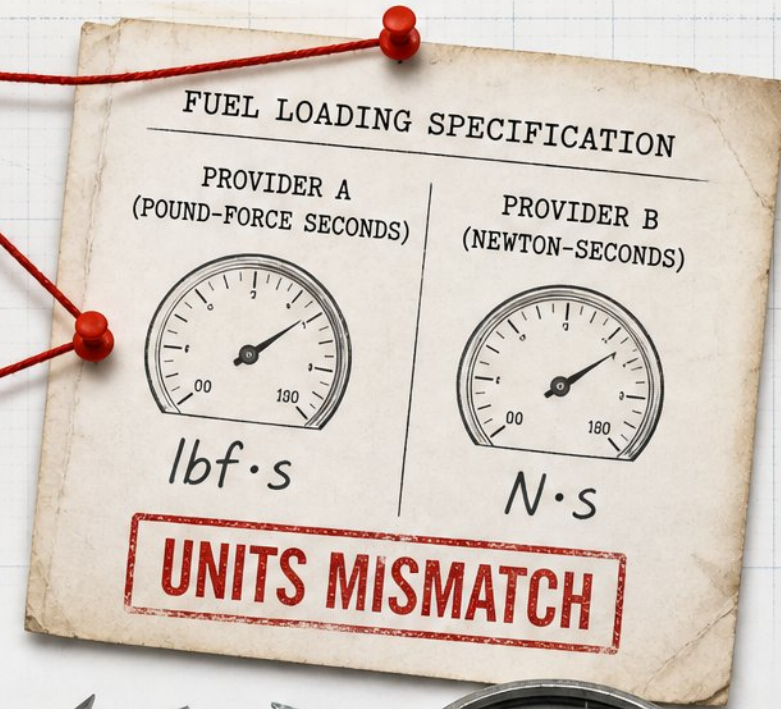
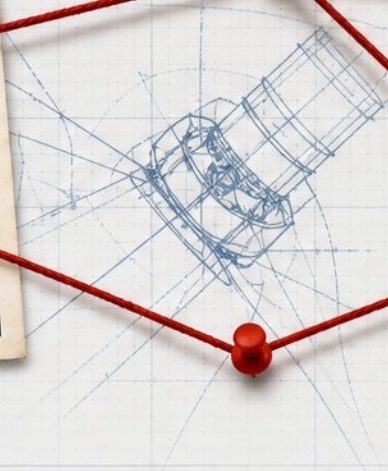
HIDDEN PRECISION LOSS
*Rounding and truncation
remove information*



Unit Disasters and Everyday Lessons: Case Studies



Mars Climate Orbiter
Wreckage Diagram



Mars Climate Orbiter: pound-force seconds vs. newton-seconds cost \$327.6M



Gimli Glider: pounds vs. kilograms in fuel loading nearly caused catastrophe



2026 lab error: suppressed alarm code caused a 100-fold reporting mistake



Common failures: silent defaults, dropped metadata, hand-off meaning loss



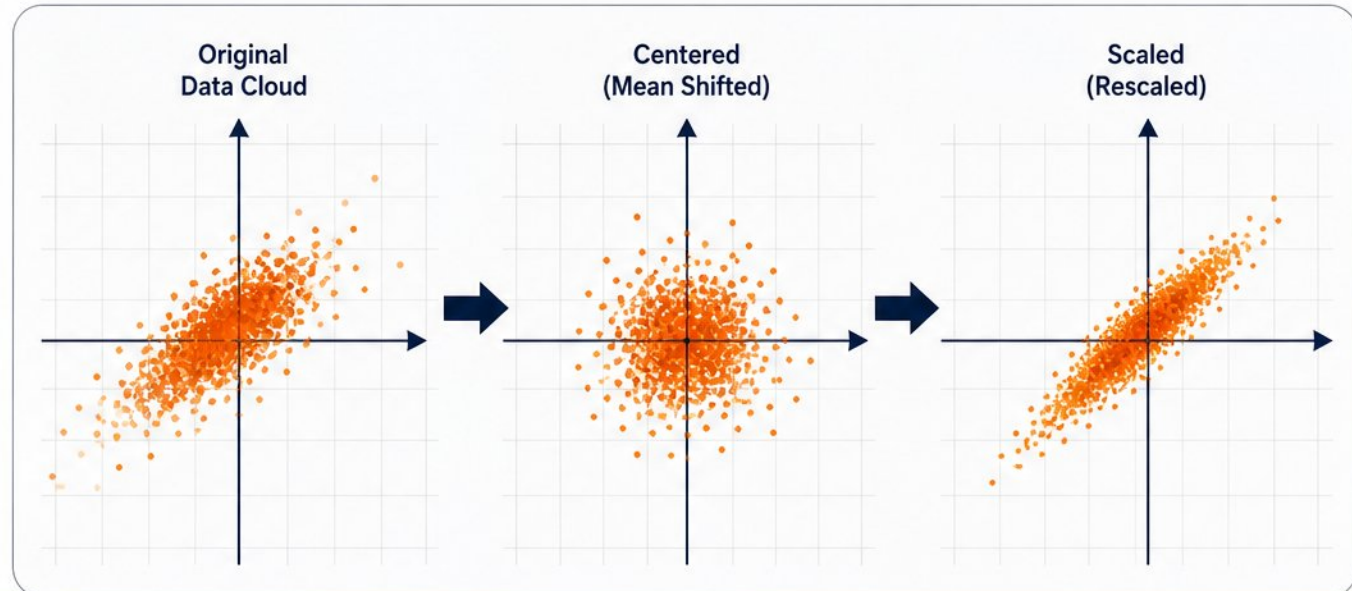
Modern risks: ms vs. seconds, GB vs. GiB, mismatched risk scales



Linear Transformations: Centering, Scaling, and Normalization

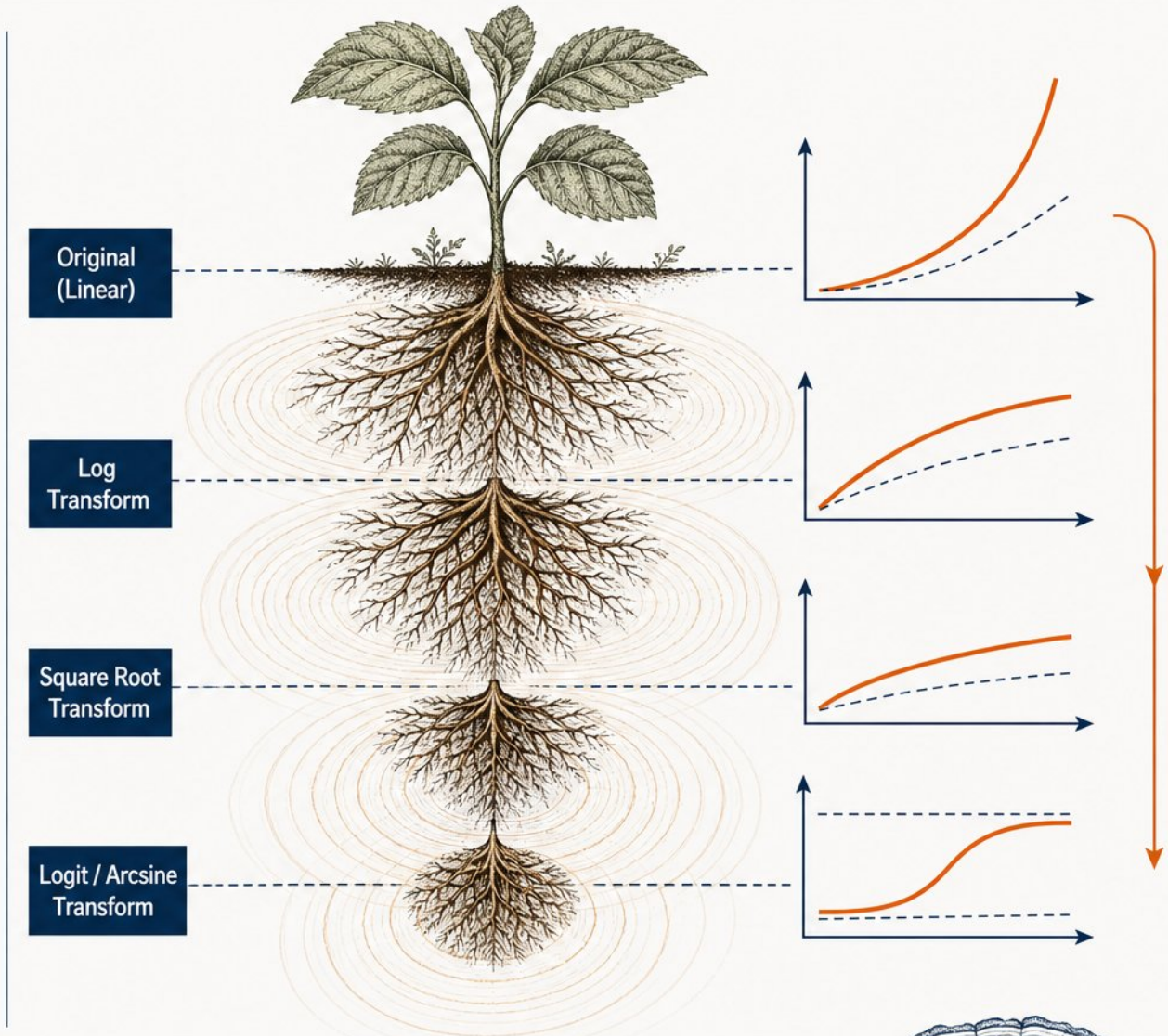


- **Z-score:** subtract mean, divide by std; sensitive to extreme outliers
- **Min-max scaling:** compresses everything into $[0,1]$; one outlier flattens the rest
- **Mean normalization & unit-norm rescaling** reshape distance metrics and clustering
- **Linear transforms** alter Euclidean distance but preserve cosine similarity
- **Interpretation trap:** a z-score of 1.5 loses connection to original units



Nonlinear Transformations: Changing Relationships and Shape

- Log transform for data spanning orders of magnitude or multiplicative processes
- Box-Cox and Yeo-Johnson: systematic power selection for variance stabilization
- Logit and arcsine transforms for bounded proportions and (0,1) data
- Interpretation shifts from linear effects to semi-elastic percentage changes
- Back-transformation bias: geometric mean $\neq$ arithmetic mean; Duan's smearing corrects



The Logarithmic Scale: Powerful, Misunderstood, and Easy to Weaponize



- Log axes turn exponential growth into a straight line, hiding absolute magnitude



- Randomized trial: log vs. linear charts shifted COVID-19 comprehension by 43 points

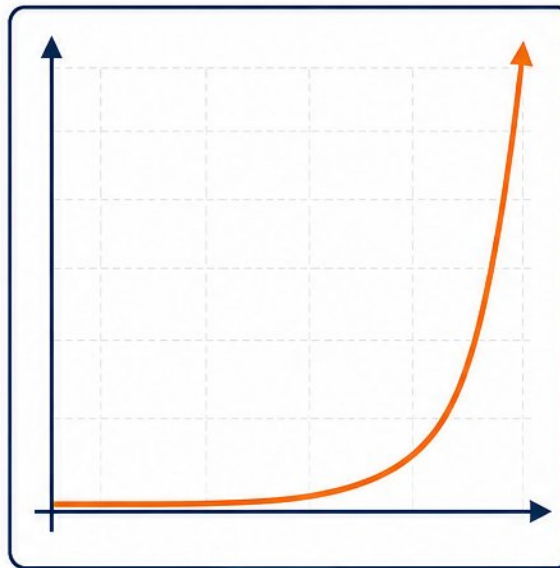


- Labeling log values with original units makes slopes dangerously misleading

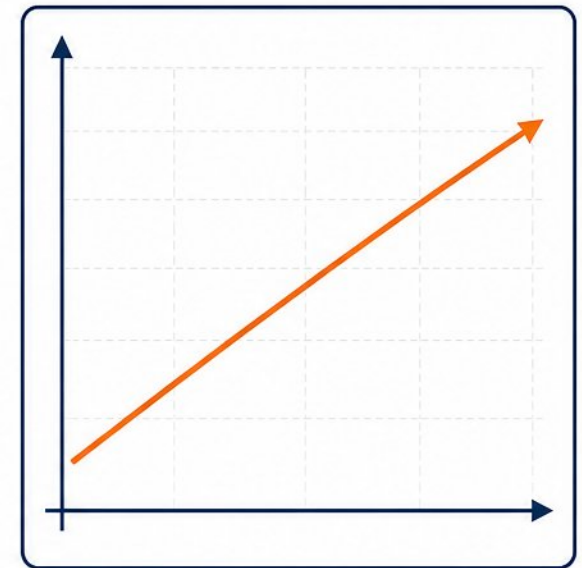


- Best practices: label scale prominently, use power-of-base ticks, show both views

LINEAR SCALE



LOG SCALE



Indexing and Rebasing: Creating and Questioning Relative Narratives

Base Year A



Base Year B



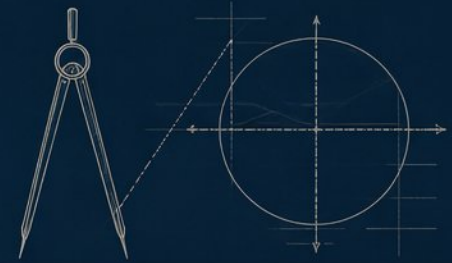
Base Year C



- Choosing a base year shapes every comparison and can be manipulated
- Laspeyres, Paasche, Fisher: different assumptions, different answers
- Chained vs. fixed-base indices alter CPI inflation narratives (2024–2026)
- Composite indices (HDI, CPI) conceal weighting and inclusion choices
- Truncated axes and cherry-picked base years distort visual perception



Scale Types and What You Are Allowed to Do With Them



- Nominal, ordinal, interval, ratio: each permits different math
- Computing ratios requires a true zero (ratio scale)
- Meaningless means: the Likert-scale debate
- Scale type constrains charts, color, and algorithm choice
- Downgrading loses information; upgrading adds assumptions



Diagnostics: Detecting When a Transformation Is Needed



- Histogram and Q-Q plots reveal skewness, tails, and non-normality patterns



- Residual vs. fitted plots diagnose heteroscedasticity and nonlinear structure



- Shapiro-Wilk and Breusch-Pagan tests provide formal evidence for intervention



- Domain cues: multiplicative processes, bounded proportions, and count data signal need



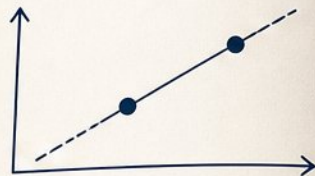
- High VIF values indicate multicollinearity that scaling may help mitigate

Communicating Transformed Data to Stakeholders

Log Scale

DATA LANGUAGE

$$\log(y_2) - \log(y_1) = \log(y_2 / y_1)$$



Differences on the log scale are additive.

Plain Language

TRANSLATION

$\approx$ % change

Think in percentages when changes are small.



Footnote:

Approximation breaks down when changes exceed ~10%.



- Log effects $\approx$ % change; breaks down when changes exceed ~10%



- Report geometric means for logs, arithmetic for original scales



- Annotate axes, footnotes, and guides to prevent misreading

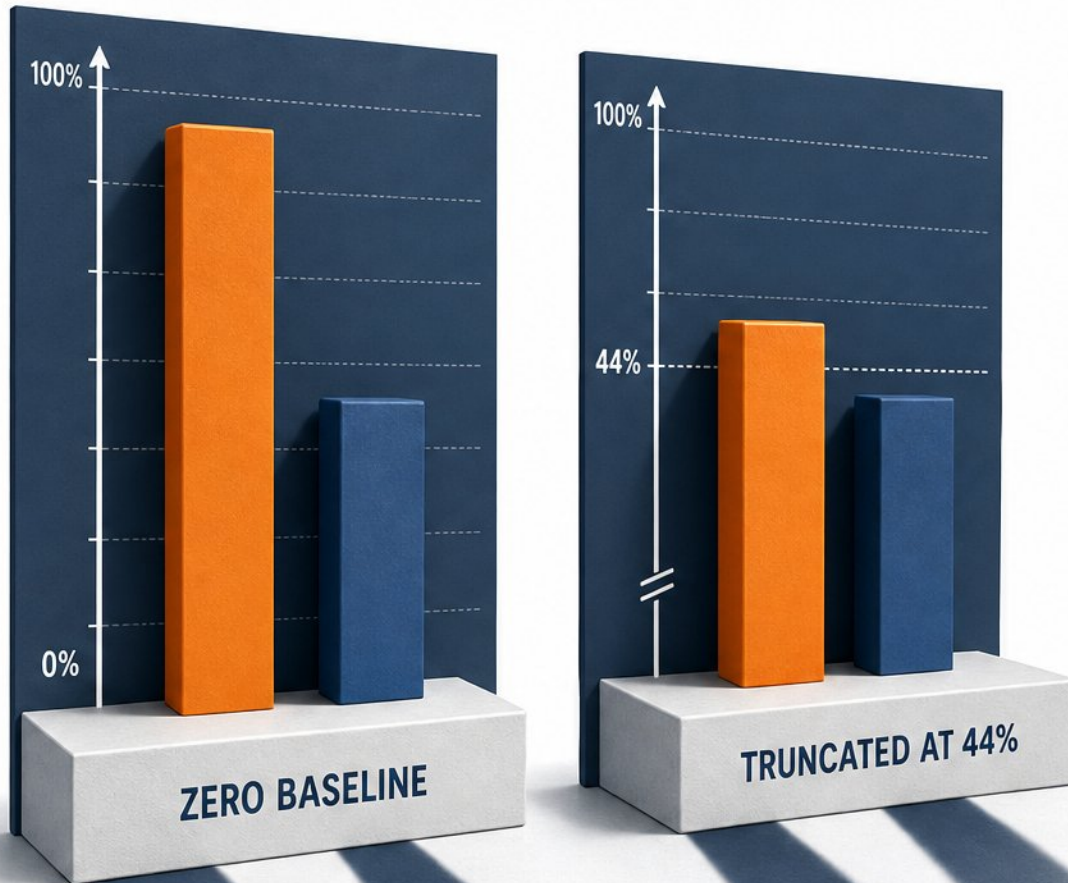


- Dual-axis charts imply false correlations; use small multiples instead



- Summarize insight first; transformation details stay in appendix

When Visualizations Mislead: Patterns, Pitfalls, and Fixes



- Truncated axes: bars starting at 44% create false 3-to-1 impressions



- Log-axis errors: 29% of biology articles distort fold change by 4.9×



- Cherry-picked time ranges, 3D charts, overloaded pie charts mislead



- Dual-axis illusion: independent scaling synchronizes unrelated trends



- Fixes: zero baselines, explicit labels, small multiples over dual axes

Integration Exercise: Choose, Apply, and Defend a Transformation



- **Hands-on scenario:** select transformations for given raw data and a business question



- **Document pipeline:** justify each step in comparability, interpretability, and communicability



- **Peer critique:** identify alternative transformations and trade-offs introduced

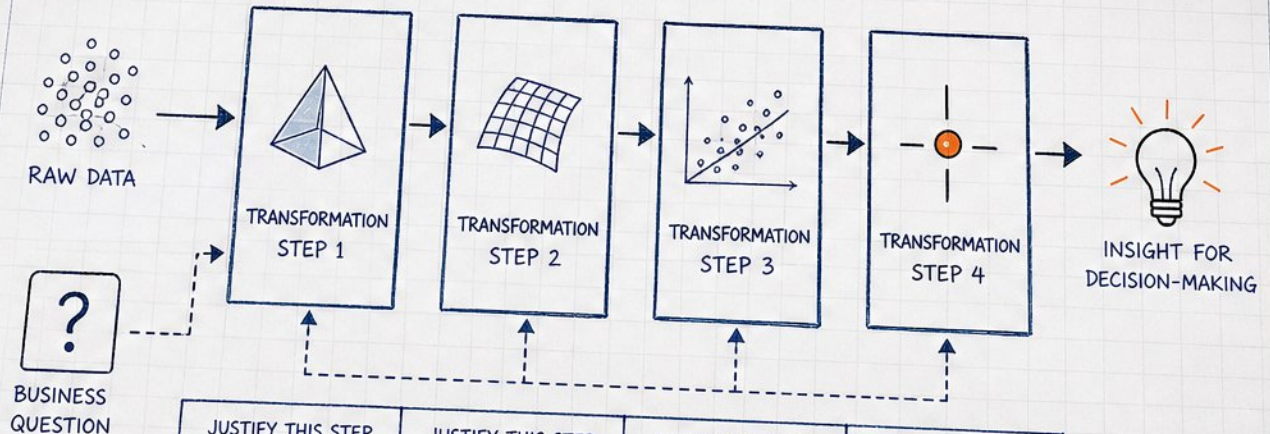


- **Navigate ambiguity:** discuss edge cases where no single transformation is clearly best



- **Key takeaways,** common mistakes, and curated resources for continued learning

PIPELINE BLUEPRINT



JUSTIFY THIS STEP

- Comparability
- Interpretability
- Communicability

JUSTIFY THIS STEP

- Comparability
- Interpretability
- Communicability

JUSTIFY THIS STEP

- Comparability
- Interpretability
- Communicability

JUSTIFY THIS STEP

- Comparability
- Interpretability
- Communicability

PersonWise.

PersonWise • AI digital-human course platform



Scan the code or click anywhere to watch this course live, with voice Q&A

personwise.ai/t/3cfe208b/public