

LECTURE 6:

Effective Data Storytelling

FACULTY OF COMPUTING & INFORMATICS
MULTIMEDIA UNIVERSITY
CYBERJAYA, MALAYSIA

Objectives of this subject

Towards the end of this trimester, you are able to

- ① define data storytelling
- ② name different steps and criteria for effective data storytelling

Data Storytelling...



Data Storytelling...

The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades.

Hal Varian
Chief Economist at Google

A story is a fact, wrapped in an emotion that compels us to take an action that transforms our world.

Richard Maxwell and Robert Dickman

The Fact is...

Data stories combine visualizations with narrative flow. This combination can breach the barriers between people and data, engaging the former and delving deeper into the latter.

James Richardson
Research Director at Gartner

The Fact is...

- ① The potential value of a discovery is directly proportional to the level of resistance it will face.
- ② While we may want to believe insights are harmless gifts, they can have subtle-to-significant repercussions that may be difficult for people to accept.
- ③ Generally, the bigger an insight is, the more disruptive it will be to the status quo. People can struggle with giving up what's routine and familiar.
- ④ When a new insight isn't well understood and doesn't sound compelling, it will have no chance of overcoming resistance to change.

THE THREE DATA STORY ELEMENTS COMPLEMENT EACH OTHER



Data storytelling involves the skillful combination of three key elements: *data*, *narrative*, and *visuals*.

Explain

Data Storytelling

When narrative is coupled with data, it helps to explain to your audience what's happening in the data and why a particular insight is important. The narrative element adds structure to the data and helps to guide the audience through the meaning of what's being shared.

When visuals are applied to data, they can enlighten the audience to insights that they wouldn't see without charts or graphs. Many interesting patterns and outliers in the data would remain hidden in the rows and columns of data tables without the help of data visualizations.

Engage

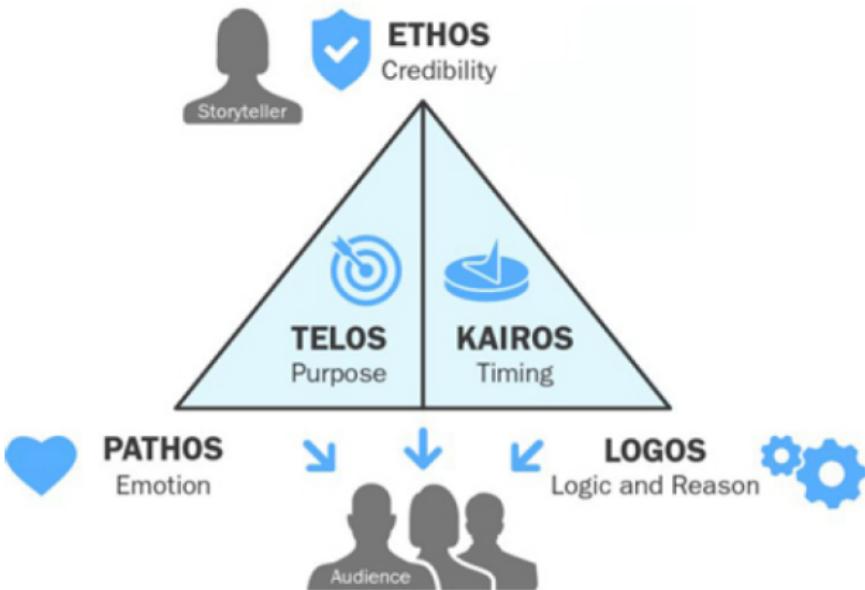
Data Storytelling

Finally, when narrative and visuals are merged together, they can engage or even entertain an audience. A good story is what holds our attention and transports us to other places and perspectives

EFFECTIVE DATA STORIES CAN DRIVE CHANGE



ARISTOTLE'S RHETORICAL TRIANGLE



Effective Data Stories

First, from an ethos perspective, the success of your data story will be shaped by your own credibility and the trustworthiness of your data.

Second, because your data story is based on facts and figures, the logos appeal will be integral to your message.

Third, as you weave the data into a convincing narrative, the pathos or emotional appeal makes your message more engaging.

Fourth, having a visualized insight at the core of your message adds the telos appeal, as it sharpens the focus and purpose of your communication.

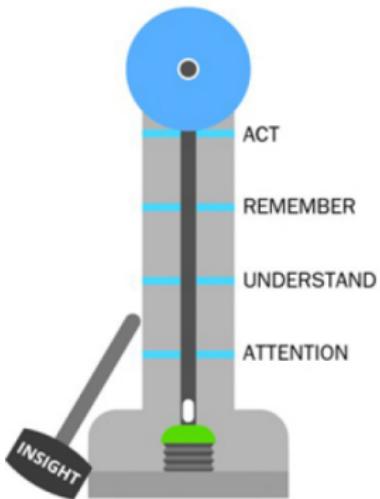
Fifth, when you share a relevant data story with the right audience at the right time (kairos), your message can be a powerful catalyst for change.

Effective Data Stories

The strength of how you communicate an insight can be measured by what effect it has on your audience. Driving an action is the ultimate goal, as your insight will then have the potential to create value.

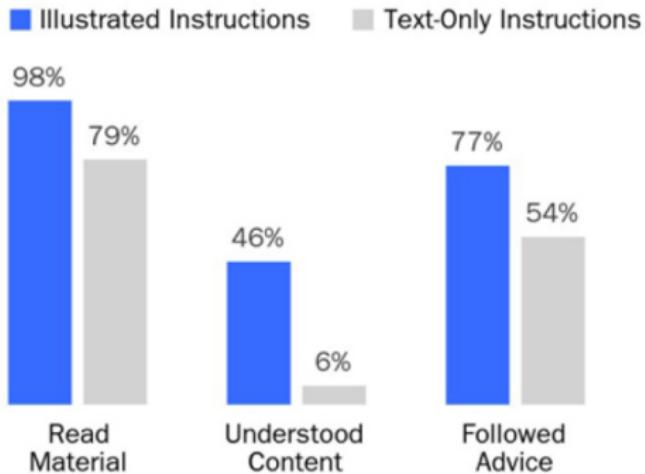
[click here to watch](#)

HOW STRONG IS YOUR DATA COMMUNICATION?



Why Your Insights Need Narrative and Visuals

ILLUSTRATED INFORMATION OUTPERFORMS TEXT-ONLY INFORMATION



The combination of narrative and visuals can help fetch facts into your audience's memory and encourage them to act on your ideas

Data Storytelling Continuum

THE DATA STORYTELLING CONTINUUM



If a data communication bears more attributes from the right side of the data storytelling continuum, it will likely be a better fit for telling a data story. If it has more attributes from the left side, then it may not be as well-suited for data storytelling.

Informative versus insightful

Data Storytelling Continuum

- ① Informative is defined as providing interesting or useful information. However, insightful goes beyond just being informative and is defined as exhibiting a clear, deep perception or understanding.
- ② Your data communication can be very informative— packed with lots of fascinating or helpful information—but that doesn't mean it will yield any specific insights. With informative content, you're choosing breadth over depth.
- ③ when you have a clear, distinct insight, your message can be more focused and more easily formed into an engaging data story.

Exploratory versus explanatory

Data Storytelling Continuum

- ➊ When the end users control how they view the data, you can't anticipate what specific insights they'll discover.
- ➋ The audience may get lost in exploration.
- ➌ when you have a particular insight in mind, it's much easier to explain—in the form of a data story—what your insight is and why it's important.

Abstract versus concrete

Data Storytelling Continuum

- ① When data is at abstract level, it frees up the data to be interpreted in a variety of ways. However, by leaving the possibility of having multiple interpretations, you forego the ability to tell a specific data story
- ② when the insights in your data composition are more concrete and specific, it is much easier to build a coherent data story because you are highlighting a particular view of the data

Continuous versus finite

Data Storytelling Continuum

- ① The data visualizations are constantly shifting to reflect the latest trends. As a result, interesting results can come and go. The transient nature of the data makes it difficult to tell stories
- ② By capturing these fixed moments, you're able to break down what's happening and examine an insight at a much deeper level.

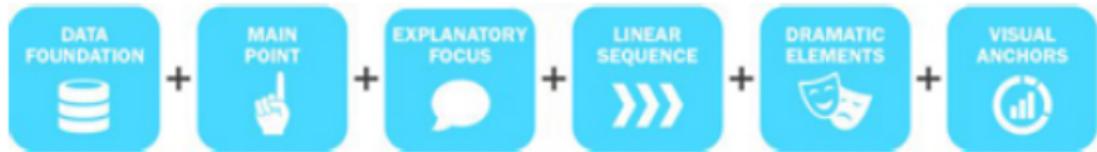
Automated versus curated

Data Storytelling Continuum

- ① Automated dissemination of information can struggle with the identification and communication of truly meaningful insights.
- ② Skilled human intervention is often required to assemble and weave together meaningful visual narratives for important insights.

To **curate** means to “select, organize, and present information or content, typically using professional or expert knowledge” (Oxford 2019).

Six Essential Elements to a Data Story



Data Foundation

Six Essential Elements to a Data Story

The building blocks of every data story are quantitative or qualitative data, which are frequently the results of an analysis or insightful observation

Main Point

Six Essential Elements to a Data Story

A data story must have a central insight or idea—it must have a main point. While you can certainly share different facts in a data story, they should all support an overarching insight. By focusing on a main point, you ensure your data story has a clear purpose

Explanatory Focus

Six Essential Elements to a Data Story

To describe: To represent or give an account of in words or pictures.

To explain: To make plain or clear; render understandable or intelligible.

When you describe something, you provide details of its features or characteristics, especially related to who, what, when, and where. However, when you explain something, you go a step further to clarify the insight and ensure it is understood by your audience.

An explanatory focus will often mean helping the audience to interpret the data by drilling into aspects such as how and why.

Linear Sequence

Six Essential Elements to a Data Story

Every data story follows a linear sequence in which supporting data points build on each other until a main point or conclusion is reached. The general definition of a story is “an account of a causally related or connected series of events.”

Dramatic Elements

Six Essential Elements to a Data Story

You must also provide sufficient contextual details in order for your audience to properly grasp your insights. To set up a data story, you may need the audience to understand the time frame, data sources, past performance, and other details for context.

For example, \$2 million in sales in the last quarter may be bad or great depending on what the sales results were like in the same quarter last year.

Visual Anchors

Six Essential Elements to a Data Story

Charts are frequently used in the analysis process to discover the insights. For example, icons can be used to create mental shortcuts for your audience, or photos can add emotional emphasis to key data points.

Seven Essential Principles for Better Visual Storytelling

PART 1—THE SETUP

- 1  RIGHT DATA

- 2  RIGHT VISUALIZATIONS

- 3  RIGHT CONFIGURATION

PART 2—THE POLISH

- 4  REMOVE NOISE

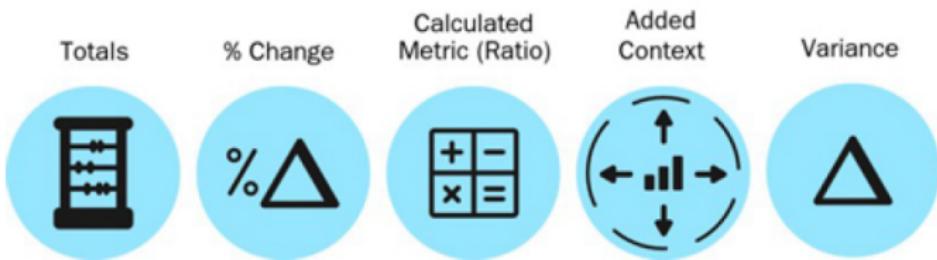
- 5  FOCUS ATTENTION

- 6  MAKE APPROACHABLE

- 7  INSTILL TRUST

The seven key principles of visual storytelling are divided into two major sections or parts: the Setup and the Polish.

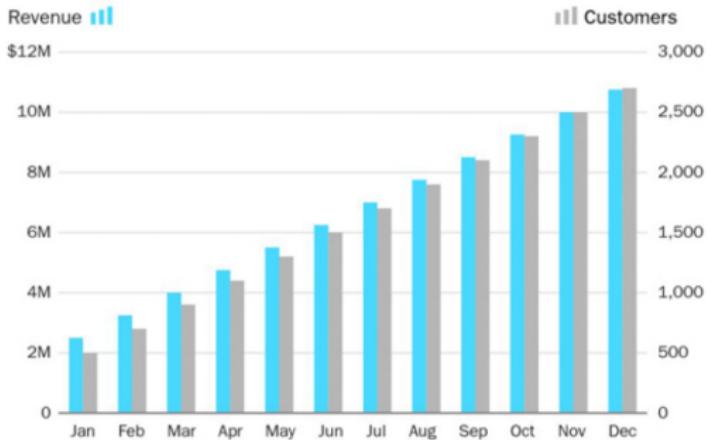
Principle 1: Visualize the Right Data



Don't always deal with *counts* and *sums*. Sometimes, adjusting the underlying data may better convey your key insight to the audience

Principle 1: Visualize the Right Data

**TOTAL VALUES MAY NOT COMMUNICATE
YOUR POINTS AS EFFECTIVELY**



Take note of the chart above, it has 2 y-axis. It's difficult to see that the monthly revenue isn't expanding as rapidly as the customer base because each metric has a different scale.

Principle 1: Visualize the Right Data

PERCENT CHANGE PUTS DIFFERENT METRICS ON THE SAME % AXIS



A better version of comparison. The chart is now having the same y-axis.

Principle 1: Visualize the Right Data

**CALCULATED METRICS MAY HELP
CLARIFY A PROBLEM**



By trending the revenue per customer on the second axis, it's easy to see the company is acquiring more customers that spend less with their organization.

Principle 1: Visualize the Right Data

VARIANCE: DO THE MATH FOR YOUR AUDIENCE

Safety Incidents by Plant

■ 2017 ■ 2018



Year-to-Year Variance in Safety

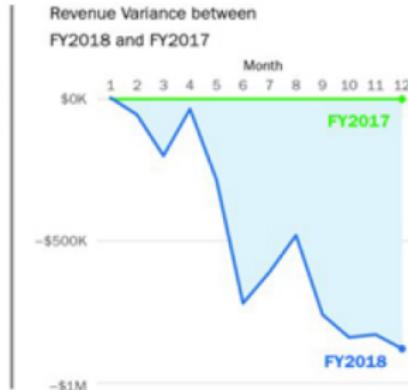
Incidents by Plant (2017–2018)



The left chart requires the audience to compare the differences between the two years for each plant. The right chart does the math for the audience and focuses on the deviation between the two years by plant.

Principle 1: Visualize the Right Data

SHOWING THE VARIANCE CAN ADD EMPHASIS TO YOUR KEY POINT



The left chart reveals FY2018 is underperforming the previous year. However, the right chart, with its focus on revenue variance, highlights more emphatically how far FY2018 has deviated from the previous year's sales results.

Principle 2: Choose the Right Visualizations

The important criterion for a graph is not simply how fast we can see a result; rather it is whether through the use of the graph we can see something that would have been harder to see otherwise or that could not have been seen at all.

William Cleveland
statistician and author

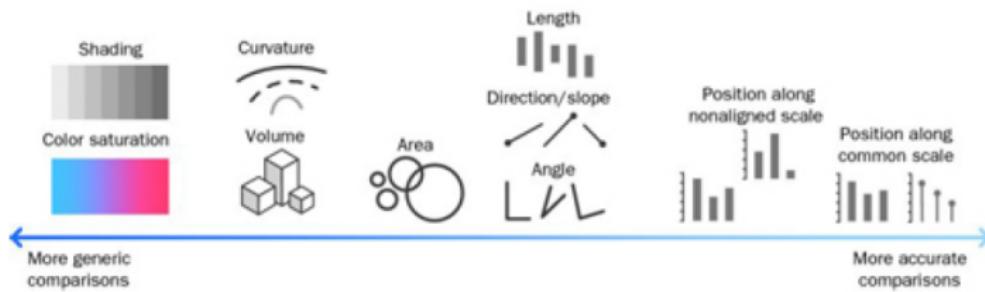
Principle 2: Choose the Right Visualizations

MAJOR CHART TYPE CATEGORIES FOR BUSINESS PROFESSIONALS

Comparison	Trend	Composition	Relationship	Distribution	Spatial	Flow
Bar & Column	Line	Waffle	Scatterplot	Histogram	Choropleth	Funnel
Slopegraph	Small Multiple (Line)	100% Stacked Column	Bubble	Pyramid	Dot Density	Waterfall
Dumbbell	Column and Stacked Column	Pie and Donut		Box plot	Proportional Symbol	Sankey
Table	Stacked Area	Treemap	XY Heatmap	Cloud	Heatmap	Chord

Principle 2: Choose the Right Visualizations

CLEVELAND AND MCGILL'S GRAPHICAL PERCEPTION MODEL



Cleveland and McGill found data visualizations that align more with the perceptual tasks on the right side of the comparison axis supported more detailed, accurate comparisons, while those that aligned more with the perceptual tasks on the left facilitated more high-level, generic comparisons.

Principle 2: Choose the Right Visualizations

PERCEPTUAL TASKS CAN DIFFER IN THE SAME CHART



Only the bottom values in this stacked column can be compared by position with a common scale. The rest of the values must be compared by perceiving their length, which is harder for making precise comparisons.

Principle 2: Choose the Right Visualizations

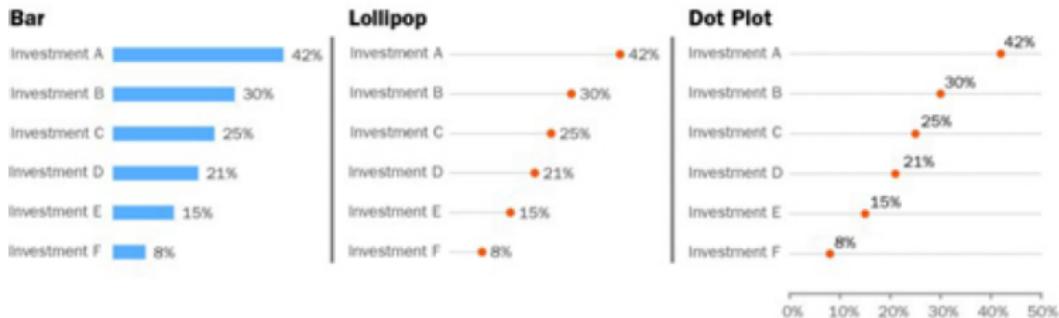
PIE CHART VERSUS BAR CHART



If you want to show the proportion of sales coming from different industry verticals, you may visualize the data in a pie chart, but without labels for the actual values, the size differences are difficult to determine. On the other hand, a bar chart offers more precise comparisons even without labeling.

Principle 2: Choose the Right Visualizations

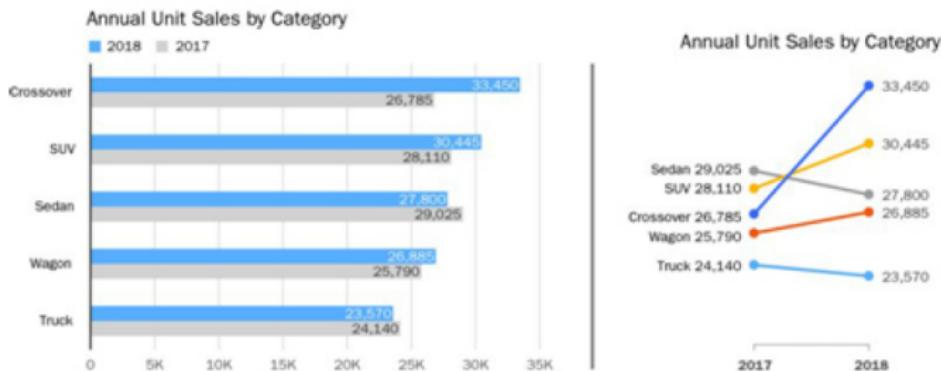
BAR CHART ALTERNATIVES: LOLLIPOP AND DOT PLOT CHARTS



Lollipop and dot plot charts are two alternatives to the bar chart.

Principle 2: Choose the Right Visualizations

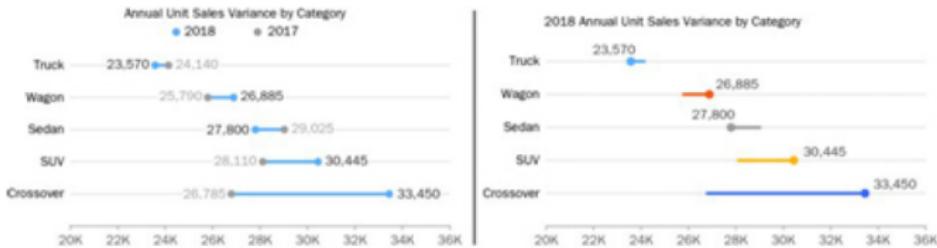
BAR CHART VERSUS SLOPEGRAPH



We often use bar and column charts to compare data for two different categories or time periods. Instead, a slopegraph can be used to represent the differences or changes between the paired values.

Principle 2: Choose the Right Visualizations

VARIANCE ALTERNATIVES: DUMBBELL AND TADPOLE CHART



The dumbbell chart on the left shows the difference in unit sales for each automotive category. The tadpole chart on the right shows the same data but emphasizes the more recent year (2018).

Principle 3: Calibrate the Visuals to Your Message



After you have determined you have the right data visualization, you need to calibrate your chart to the message you want to convey. This is to ensure that the visuals match the words or message.

Principle 3: Calibrate the Visuals to Your Message

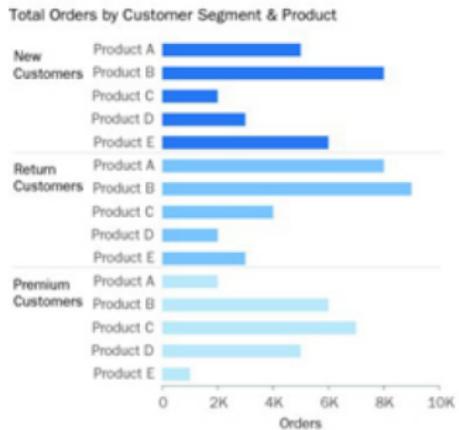
Key areas you can focus on to keep your message and visuals are properly calibrated:

- ① Keep comparisons in close proximity.
- ② Provide a common baseline for comparisons.
- ③ Ensure charts are consistent for comparisons.

Principle 3: Calibrate the Visuals to Your Message

Keep comparisons in close proximity

MAKE IT EASY FOR YOUR AUDIENCE TO MAKE COMPARISONS



The left side focuses on comparing products while the right one comparing customer segments.

Principle 3: Calibrate the Visuals to Your Message

Provide a common baseline for comparisons

STACKED BAR CHART VERSUS PANEL BAR CHARTS

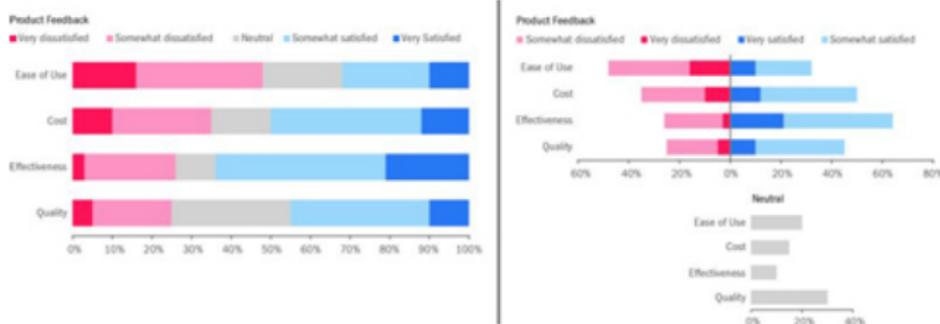


A stacked bar chart is difficult to compare the stacked values that are not aligned with the baseline along the y-axis. A panel bar chart would give each quarter its own baseline for easier comparisons.

Principle 3: Calibrate the Visuals to Your Message

Provide a common baseline for comparisons

100% STACKED BAR CHART VERSUS DIVERGING BAR CHART



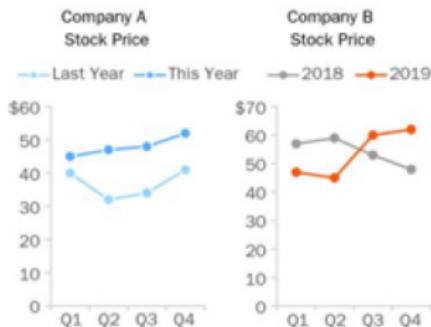
In the 100% on the ends, but the values in the middle don't have a common baseline. The diverging bar chart can alleviate some of these problems if the neutral values (gray) are removed.

Principle 3: Calibrate the Visuals to Your Message

Ensure charts are consistent for comparisons

CONSISTENCY FACILITATES COMPARISONS

INCONSISTENT



CONSISTENT



On the left, subtle differences between the two line charts can interfere with making comparisons. On the right, the visuals are consistent so the audience can focus on interpreting the data differences.

Principle 4: Remove Unnecessary Noise

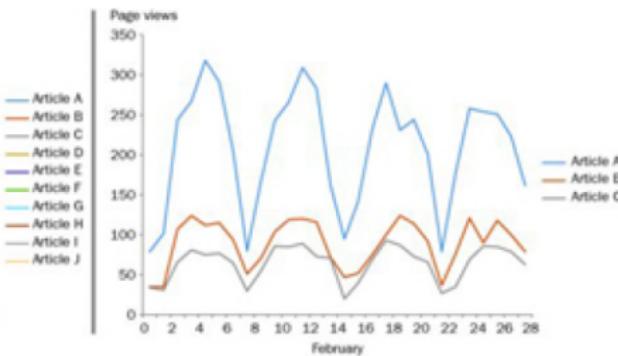
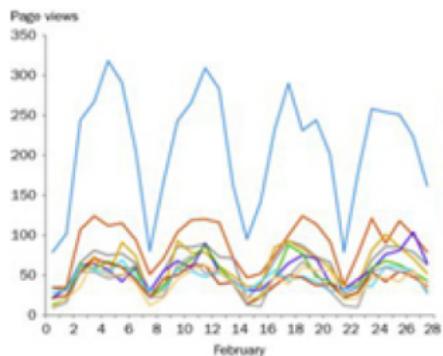
We could reduce noise by:

- ① Remove surplus data
- ② Aggregate less important data
- ③ Separate overlapping data

Principle 4: Remove Unnecessary Noise

Remove surplus data

SIMPLIFY A “SPAGHETTI CHART” TO REDUCE THE NOISE

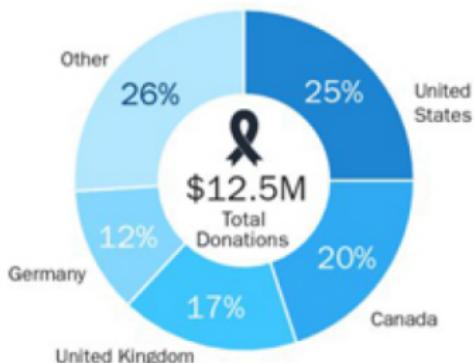
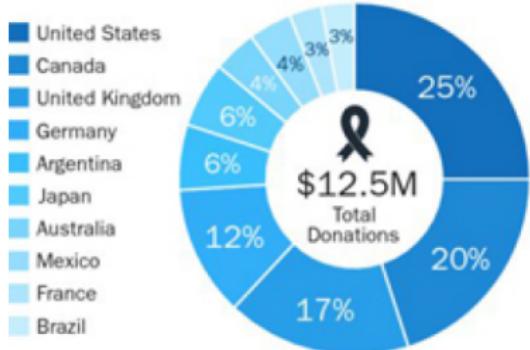


Focus only on the aim of comparison. Remove unnecessary extra data points.

Principle 4: Remove Unnecessary Noise

Aggregate less important data

COMBINE LESS IMPORTANT VALUES INTO “OTHER” GROUPING

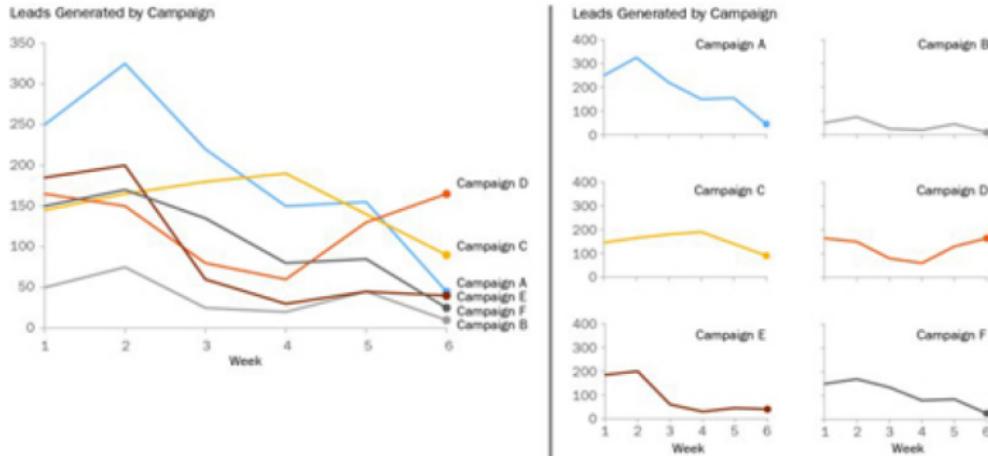


In order to simplify the donut chart on the left, the lesser slices are combined into an “Other” slice in the version on the right

Principle 4: Remove Unnecessary Noise

Separate overlapping data

USE A PANEL CHART TO SEPARATE LAYERS OF NOISE



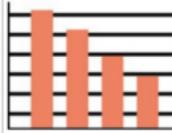
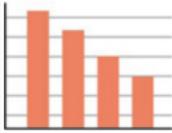
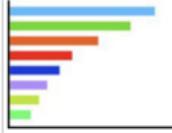
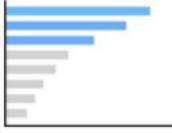
The separation of the busy line chart on the left into facets or panel line charts makes it easier for people to make comparisons

Principle 4: Remove Unnecessary Noise

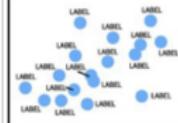
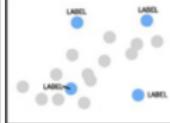
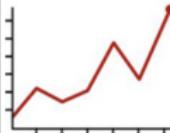
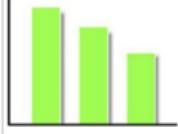
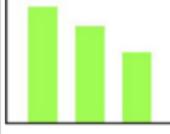
In 1983, Tufte coined the term chartjunk, which refers to all the nonessential visual elements in a chart that aren't required to understand its information and may even detract from its perception or comprehension (Tufte 1983)

Principle 4: Remove Unnecessary Noise

chartjunk

	More Chartjunk	Less Chartjunk
3D effects: Adding a three-dimensional perspective to a chart can distort its information and make it harder to interpret. If you're trying to facilitate comparisons, 3D charts should be avoided at all costs.		
Dark gridlines: Dark or thick gridlines can compete with the information in the foreground. Gridlines can serve a valid purpose and assist people in evaluating data. However, they should be subtle (thin, light color) so they don't overpower the core information.		
Nonstrategic use of color: Color is a powerful tool in your data storytelling toolbox, but too often it is used carelessly or randomly in charts. Instead, color should be applied purposefully to convey key points within your chart (see Principle #5).		

Principle 4: Remove Unnecessary Noise

	More Chartjunk	Less Chartjunk
Overuse of labeling: While it may be necessary to attach labels to data points, systematic labeling can add a significant amount of textual noise. Instead, it is better to focus labeling strategically on key values that matter to your story.	 A scatter plot with 15 blue circular data points. Each point is labeled with the word "LABEL" in capital letters. The labels are placed directly next to their respective points.  A scatter plot with 15 blue circular data points. Only two points are labeled with the word "LABEL" in capital letters. These labels are placed further away from the points, and the points themselves are surrounded by a cluster of smaller gray circles.	
Granular scales: Rather than your data being too complex, the scale of a vertical or horizontal axis may be too detailed for what's needed. It may be helpful to simplify the scale so it doesn't overwhelm your audience with unnecessary detail.	 A line graph showing a red line connecting six data points. The x-axis has 10 major tick marks, but only 5 data points, creating a very granular and cluttered appearance.  A line graph showing a red line connecting six data points. The x-axis has 5 major tick marks, which corresponds to the number of data points, creating a clean and clear appearance.	
Artistic effects: Different design effects such as shading, beveling, or gradients are sometimes used to give charts more “visual impact.” However, these effects should be avoided if they’re distracting and make comparisons more difficult.	 A bar chart with four light green bars. Each bar has a different shade of green, creating a gradient effect. The bars represent data values that are visually difficult to compare due to the shading.  A bar chart with four light green bars. All bars have the same uniform color and height, making it easy to compare the data values.	

Principle 5: Focus Attention on What's Important

Colour Contrast

COLOR CAN BE BOTH SIGNAL AND NOISE

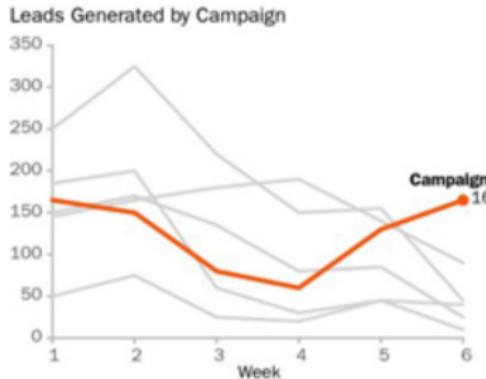
1094839875	1094839875	1094839875
8930431716	8930431716	8930431716
2394851204	2394851204	2394851204
1158902859	1158902859	1158902859
9387284016	9387284016	9387284016

The color contrast between the blue and light gray numbers in the middle set of numbers makes it easier to notice and count the number of eights.

Principle 5: Focus Attention on What's Important

Colour Contrast

CHOOSE WHAT DATA IS IN THE FOREGROUND AND BACKGROUND



Leads Generated by Campaign in Sixth Week

A bar chart titled "Leads Generated by Campaign in Sixth Week" showing leads for each campaign. Campaign D is the clear leader with 165 leads, followed by Campaign C with 90 leads. The other campaigns have significantly fewer leads.

Campaign	Leads
Campaign D	165
Campaign C	90
Campaign A	45
Campaign E	40
Campaign F	25
Campaign B	10

By highlighting the main insight with color and using grayscale for the less-important information, you can establish what is in the foreground and background of your visuals.

Principle 5: Focus Attention on What's Important

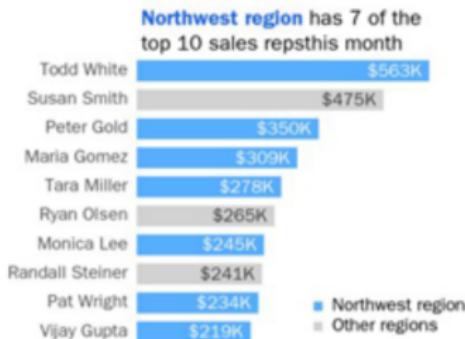
Text

EXPLANATORY TITLES CAN HELP TELL THE STORY

DESCRIPTIVE



EXPLANATORY



While you could figure out the point of the left chart, the explanatory title helps the audience to quickly orient themselves to the intended takeaway from the visual. Because “Northwest region” is colored blue, the legend is optional

Principle 5: Focus Attention on What's Important

Text

USE ANNOTATIONS TO GUIDE PEOPLE THROUGH THE DATA



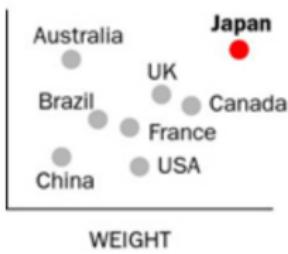
In this area chart, a mix of observational (194%), additive (competitor X, new campaign), and hybrid (232% increase after product update) annotations guide the audience through the most salient parts of the data.

Principle 5: Focus Attention on What's Important

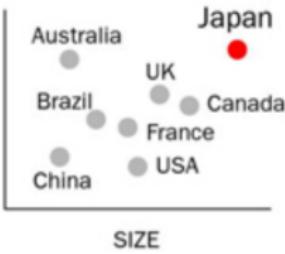
Typography

USE TYPOGRAPHIC ELEMENTS TO FOCUS ATTENTION

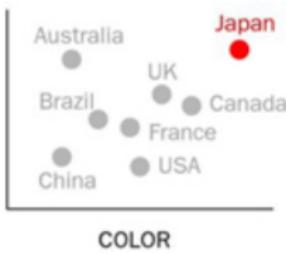
Japan Is 35% higher than the next highest country



Japan is 35% higher than the next highest country



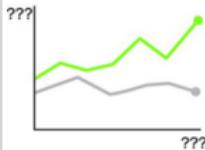
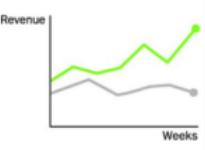
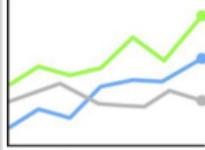
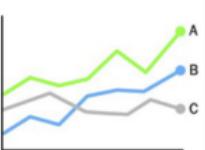
Japan is 35% higher than the next highest country



Typographic elements such as weight, size, and color can be used to focus attention on specific information within a data visualization

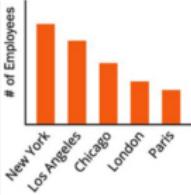
Principle 6: Make Your Data Approachable and Engaging

Tips to Streamline Readability of Charts

	More Challenging	Less Challenging
Axis labels: In order to make sense of the data in a chart, it's critical that all axes are clearly labeled. If an axis or its scale is ambiguous, then it will only slow your audience's ability to understand the information in your chart. An exception to this rule might be date units such as months or years that are straightforward and easy to understand without labels.		
Direct labels: Whenever possible, it's better to directly label values and avoid forcing the audience to reference a legend that is further removed from the data. Although looking back and forth between data values and a legend is less efficient, in certain situations, it may be the only option (grouped column charts).		

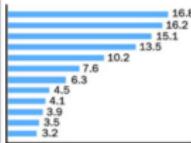
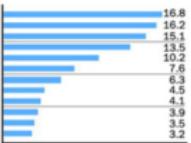
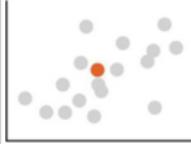
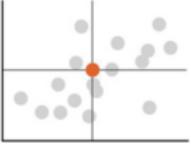
Principle 6: Make Your Data Approachable and Engaging

Tips to Streamline Readability of Charts

	More Challenging	Less Challenging																														
Legible text: It is harder to read text that is at a vertical or diagonal angle than text that is horizontal. For categories or values that require longer labels, you may consider using a bar chart instead of a column chart because it is easier to accommodate longer labels in a horizontal format.	 <p>A bar chart titled '# of Employees' showing the number of employees for five cities. The bars are orange and have black outlines. The x-axis labels are rotated vertically: New York, Los Angeles, Chicago, London, and Paris. The y-axis is labeled '# of Employees'. The approximate data is: New York (~10), Los Angeles (~9), Chicago (~7), London (~5), and Paris (~4).</p> <table border="1"><thead><tr><th>City</th><th># of Employees</th></tr></thead><tbody><tr><td>New York</td><td>~10</td></tr><tr><td>Los Angeles</td><td>~9</td></tr><tr><td>Chicago</td><td>~7</td></tr><tr><td>London</td><td>~5</td></tr><tr><td>Paris</td><td>~4</td></tr></tbody></table>	City	# of Employees	New York	~10	Los Angeles	~9	Chicago	~7	London	~5	Paris	~4	 <p>A horizontal bar chart titled '# of Employees' showing the number of employees for five cities. The bars are orange and have black outlines. The x-axis labels are horizontal: New York, Los Angeles, Chicago, London, and Paris. The y-axis is labeled '# of Employees'. The approximate data is: New York (~10), Los Angeles (~9), Chicago (~7), London (~5), and Paris (~4).</p> <table border="1"><thead><tr><th>City</th><th># of Employees</th></tr></thead><tbody><tr><td>New York</td><td>~10</td></tr><tr><td>Los Angeles</td><td>~9</td></tr><tr><td>Chicago</td><td>~7</td></tr><tr><td>London</td><td>~5</td></tr><tr><td>Paris</td><td>~4</td></tr></tbody></table>	City	# of Employees	New York	~10	Los Angeles	~9	Chicago	~7	London	~5	Paris	~4						
City	# of Employees																															
New York	~10																															
Los Angeles	~9																															
Chicago	~7																															
London	~5																															
Paris	~4																															
City	# of Employees																															
New York	~10																															
Los Angeles	~9																															
Chicago	~7																															
London	~5																															
Paris	~4																															
Simple increments: You want to use scale increments that are natural and easy for the audience to process. For example, increments of 1, 2, 5, or 10 are less awkward to navigate than increments such as 3, 6, or 12.	 <p>A line graph with two y-axes. The left y-axis ranges from 0 to 18 with major ticks at 3, 6, 9, 12, 15, and 18. The right y-axis ranges from 0 to 20 with major ticks at 5, 10, 15, and 20. The x-axis has four data points marked by blue dots connected by a blue line. The approximate data is: (1, 6), (2, 9), (3, 12), and (4, 16).</p> <table border="1"><thead><tr><th>Point</th><th>Left Y-axis</th><th>Right Y-axis</th></tr></thead><tbody><tr><td>1</td><td>6</td><td>6</td></tr><tr><td>2</td><td>9</td><td>9</td></tr><tr><td>3</td><td>12</td><td>12</td></tr><tr><td>4</td><td>16</td><td>16</td></tr></tbody></table>	Point	Left Y-axis	Right Y-axis	1	6	6	2	9	9	3	12	12	4	16	16	 <p>A line graph with two y-axes. The left y-axis ranges from 0 to 18 with major ticks at 3, 6, 9, 12, 15, and 18. The right y-axis ranges from 0 to 20 with major ticks at 5, 10, 15, and 20. The x-axis has four data points marked by blue dots connected by a blue line. The approximate data is: (1, 5), (2, 8), (3, 11), and (4, 15).</p> <table border="1"><thead><tr><th>Point</th><th>Left Y-axis</th><th>Right Y-axis</th></tr></thead><tbody><tr><td>1</td><td>5</td><td>5</td></tr><tr><td>2</td><td>8</td><td>8</td></tr><tr><td>3</td><td>11</td><td>11</td></tr><tr><td>4</td><td>15</td><td>15</td></tr></tbody></table>	Point	Left Y-axis	Right Y-axis	1	5	5	2	8	8	3	11	11	4	15	15
Point	Left Y-axis	Right Y-axis																														
1	6	6																														
2	9	9																														
3	12	12																														
4	16	16																														
Point	Left Y-axis	Right Y-axis																														
1	5	5																														
2	8	8																														
3	11	11																														
4	15	15																														

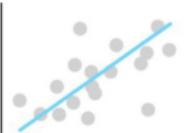
Principle 6: Make Your Data Approachable and Engaging

Tips for Using Reference Lines

	More Challenging	Less Challenging
Guidelines: When you have a long list of values in a bar chart or table, your audience may struggle with comparing individual items. By adding thin guidelines between groups of three to five items, you make it easier for the audience to locate and compare different items, especially when the value labels are justified to the same spot.		
Anchor lines: When you highlight a point in a scatter plot, you may want to add thin lines for the point's x and y positions. These horizontal and vertical lines can better support comparisons between key data points. Anchor lines can also be used in bar, column, and line charts to provide more context in a visual (mean value, target).		

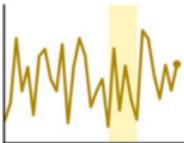
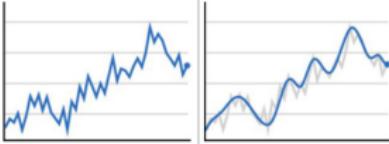
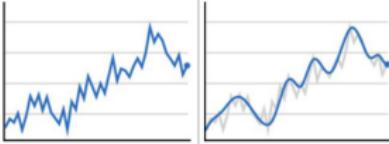
Principle 6: Make Your Data Approachable and Engaging

Tips for Using Reference Lines

	More Challenging	Less Challenging
Trendlines: With line and scatter plot charts, you may consider adding a “line of best fit” that displays the trend of the relationship between the two variables. It also helps to communicate the positive or negative relationship through the slope of the line.		
Shaded regions: In some cases, you may want to highlight a certain range or band of values within a chart. For example, you may use a shaded area to indicate the negative territory or a targeted outcome. The regions can help the audience more quickly interpret the data by having additional visual context.		

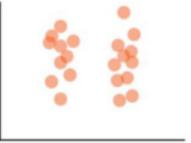
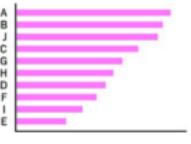
Principle 6: Make Your Data Approachable and Engaging

Tips for Formatting

	More Challenging	Less Challenging
Aspect ratio: In order for your data to communicate effectively, you may need to adjust the width-to-height ratio so the audience can more easily discern the key trends or patterns in the data. With a little more breathing room between the data points, a detailed line chart may be easier to examine and understand.		
Smoothing: When you provide granular data, the jaggedness caused by short-term fluctuations in the data can make it harder to assess overall patterns in column or line charts. If you apply a weighted average to smooth the data, the resulting curved lines will make the general movements in the data more visible.		

Principle 6: Make Your Data Approachable and Engaging

Tips for Formatting

	More Challenging	Less Challenging
Transparency and jittering: In “overplotting” situations where multiple values are plotted on top of each other, it can be hard to ascertain the density of the data points. Making the dot points more transparent can help rectify this problem. For some dot plots, you may want to “jitter” the data (randomly assign them a horizontal position) so the points are more spread apart to further clarify the density.	 A dot plot with two categories, A and B, on the x-axis. Each category contains several overlapping orange dots. The dots are closely packed, making it difficult to determine the exact count or density of points.  A dot plot with two categories, A and B, on the x-axis. Each category contains several orange dots that have been randomly assigned horizontal positions, creating a more spread-out appearance than the first plot.	
Sorting: Arranging the items by rank from largest to smallest can make your data easier to read and follow. However, in some cases, it still may be more important to maintain a specific order (alphabetical) to facilitate looking up certain values.	 A horizontal bar chart showing items A through J. The bars are pink and decrease in length from left to right, indicating a ranking from highest to lowest. The items are listed vertically on the left: A, B, C, D, E, F, G, H, I, J.  A horizontal bar chart showing items A through E. The bars are pink and decrease in length from left to right, indicating a ranking from highest to lowest. The items are listed vertically on the left: A, B, C, D, E.	

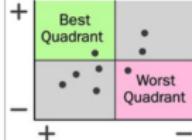
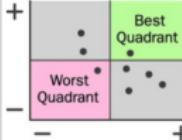
Principle 6: Make Your Data Approachable and Engaging

Tips for Convention Adherence

	More Challenging	Less Challenging
Color polarity: If your data shows a positive or negative result, you may consider assigning an appropriate color to it (green = good, red = bad). If you assign the wrong color, your audience will be confused by your chart. To avoid color blindness issues, you can use dark and light hues or redundant symbols to clarify the differences.		
Color association: Some items may already be associated with certain colors (countries, companies, political parties, etc.). Rather than assigning random colors to them, you can strengthen the encoded information with familiar, corresponding colors.		
Independent variable on x-axis: A common practice is to put the independent variable (“cause” variable) on the x-axis and the dependent variable (“effect” variable) on the y-axis. For example, it’s customary to measure time along the horizontal x-axis since it is not changed by the other factors you’re measuring.		

Principle 6: Make Your Data Approachable and Engaging

Tips for Convention Adherence

	More Challenging	Less Challenging
Starting position: In the case of radial charts (pie, donut, radar), the 12 o'clock, or zero-degree, position is a natural starting point for these types of data visualizations. Generally, the largest slice is located at this 12 o'clock position, and then the rest of the slices follow it in clockwise fashion, sorted by size.		
Direction: Typically, we associate the left and downward directions with negative values and the right and upward directions with positive ones. In a quadrant analysis, the top-right quadrant is frequently seen as the most desirable location.		

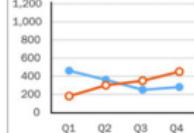
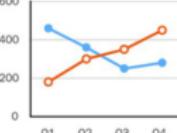
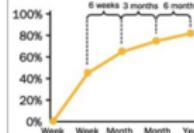
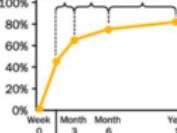
Principle 7: Instill Trust in Your Numbers

Deceptive Practices to Avoid

	Potentially Misleading	Not Misleading																																
Truncated axis (column/bar): Whenever you truncate the y-axis of a column or the x-axis of a bar chart, the lengths of the bars are no longer representative of the actual values. This approach can be misleading, as it can exaggerate small differences between values.	<p>A bar chart comparing two data series. The left chart has a y-axis ranging from 75 to 95, while the right chart has a y-axis ranging from 0 to 100. Both charts show three bars with heights of approximately 78, 85, and 92 respectively. The right chart's y-axis is truncated at 100, making the bars appear taller than they actually are relative to the scale.</p> <table border="1"><thead><tr><th>Month</th><th>Series A</th><th>Series B</th><th>Series C</th></tr></thead><tbody><tr><td>May</td><td>78</td><td>78</td><td>78</td></tr><tr><td>Jun</td><td>85</td><td>78</td><td>78</td></tr><tr><td>Jul</td><td>92</td><td>85</td><td>85</td></tr></tbody></table>	Month	Series A	Series B	Series C	May	78	78	78	Jun	85	78	78	Jul	92	85	85	<p>A bar chart comparing two data series. The left chart has a y-axis ranging from 75 to 95, while the right chart has a y-axis ranging from 0 to 100. Both charts show three bars with heights of approximately 78, 85, and 92 respectively. The right chart's y-axis is truncated at 100, making the bars appear taller than they actually are relative to the scale.</p> <table border="1"><thead><tr><th>Month</th><th>Series A</th><th>Series B</th><th>Series C</th></tr></thead><tbody><tr><td>May</td><td>78</td><td>78</td><td>78</td></tr><tr><td>Jun</td><td>85</td><td>78</td><td>78</td></tr><tr><td>Jul</td><td>92</td><td>85</td><td>85</td></tr></tbody></table>	Month	Series A	Series B	Series C	May	78	78	78	Jun	85	78	78	Jul	92	85	85
Month	Series A	Series B	Series C																															
May	78	78	78																															
Jun	85	78	78																															
Jul	92	85	85																															
Month	Series A	Series B	Series C																															
May	78	78	78																															
Jun	85	78	78																															
Jul	92	85	85																															
Truncated axis (area): Similar to column charts, the shaded region of an area chart shouldn't have a truncated y-axis, as it distorts what is displayed. However, it is generally accepted that line charts don't require a zero baseline in the same way that column and area charts do.	<p>An area chart showing monthly data from May to September. The y-axis ranges from 20K to 26K. The area is shaded light blue. The chart shows a peak in July followed by a decline. The y-axis is truncated at 26K, which makes the area appear larger than it would if the full scale were shown.</p> <table border="1"><thead><tr><th>Month</th><th>Value</th></tr></thead><tbody><tr><td>May</td><td>20K</td></tr><tr><td>Jun</td><td>21K</td></tr><tr><td>Jul</td><td>25K</td></tr><tr><td>Aug</td><td>23K</td></tr><tr><td>Sep</td><td>21K</td></tr></tbody></table>	Month	Value	May	20K	Jun	21K	Jul	25K	Aug	23K	Sep	21K	<p>A line chart showing monthly data from May to September. The y-axis ranges from 20K to 26K. The line starts at 20K in May, rises to a peak of about 25K in July, and then gradually declines. The y-axis is truncated at 26K, which makes the peak appear higher than it would if the full scale were shown.</p> <table border="1"><thead><tr><th>Month</th><th>Value</th></tr></thead><tbody><tr><td>May</td><td>20K</td></tr><tr><td>Jun</td><td>21K</td></tr><tr><td>Jul</td><td>25K</td></tr><tr><td>Aug</td><td>23K</td></tr><tr><td>Sep</td><td>21K</td></tr></tbody></table>	Month	Value	May	20K	Jun	21K	Jul	25K	Aug	23K	Sep	21K								
Month	Value																																	
May	20K																																	
Jun	21K																																	
Jul	25K																																	
Aug	23K																																	
Sep	21K																																	
Month	Value																																	
May	20K																																	
Jun	21K																																	
Jul	25K																																	
Aug	23K																																	
Sep	21K																																	

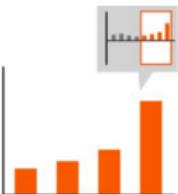
Principle 7: Instill Trust in Your Numbers

Deceptive Practices to Avoid

	Potentially Misleading	Not Misleading
Overstated axis scale or aspect ratio: To manipulate the data in a chart, the axis scale can be inflated, or the aspect ratio (width-to-height ratio) can be stretched disproportionately. In general, your chart should be wider than taller, and the data should occupy about two-thirds of the chart's scale range.	 A line chart with four data points corresponding to Q1, Q2, Q3, and Q4. The blue line starts at approximately 450 for Q1 and decreases to about 250 for Q4. The orange line starts at approximately 150 for Q1 and increases to about 400 for Q4. The y-axis ranges from 0 to 1,200 with major grid lines every 200 units. The x-axis labels are Q1, Q2, Q3, and Q4.  The same data as the first chart, but with a y-axis ranging from 0 to 600 with major grid lines every 200 units. The x-axis labels are Q1, Q2, Q3, and Q4.	
Inconsistent date intervals: If the dates in a time series do not follow regular intervals, then plotting them at equal intervals can be misleading. Some audience members may question why the dates are irregular and not consistent. You should use a proportional scale on the x-axis rather than one with equal intervals.	 A line chart showing data over a year. The x-axis has labels for Week, 6 weeks, Month, 3 months, Month, 6 months, and Year. The y-axis ranges from 0% to 100% with major grid lines every 20%. The blue line starts at 0% for Week 0, rises to ~45% at Week 6, ~65% at Month 3, ~75% at Month 6, and reaches ~80% at Year 1.  The same data as the first chart, but with a x-axis that uses consistent weekly intervals (labeled Week 0, Week 3, Week 6, Month, Month, 6 months, Year). The y-axis ranges from 0% to 100% with major grid lines every 20%. The blue line follows the same path as the first chart, starting at 0% for Week 0 and reaching ~80% at Year 1.	

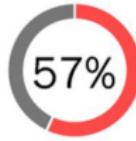
Principle 7: Instill Trust in Your Numbers

Deceptive Practices to Avoid

	Potentially Misleading	Not Misleading
Limited date ranges: Generally, it's not a good idea to "cherry pick" a timeframe so that you can mask results that could undermine a desired narrative. While it's fine to limit the amount of data you show, you shouldn't hide relevant context, especially if it could materially influence how people perceive and interpret your data. In the left example, big shifts in prior time periods are material. In the right example, they aren't as substantive, so they could be left out.	 	
Irregular binning: How you aggregate or bin the data can shape what a data visualization communicates—whether it's a map, bar chart, or histogram. Unusual or unexpected binning approaches may be viewed skeptically by your audience as attempts to skew the results.	 	

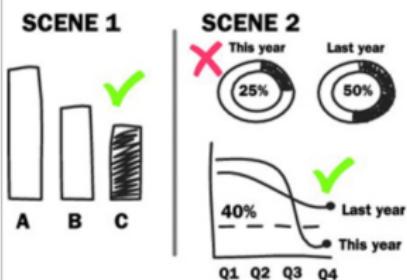
Principle 7: Instill Trust in Your Numbers

Deceptive Practices to Avoid

	Potentially Misleading	Not Misleading
Missing sources: Anyone can make up a statistic or tout a data point from a questionable source. Whenever a key statistic doesn't list a clear source, it can come across as evasive and deceptive. However, once you've stated the source, it doesn't need to be highlighted in every subsequent chart unless the source changes or you anticipate individual visuals being shared independently from the rest of the data story.	 <p>of teens fear a shooting could happen at their school</p> <p>Source: PEW RESEARCH Survey of US teens ages 13 to 17 conducted March 7-April 10, 2018</p>	 <p>57% of teens fear a shooting could happen at their school</p>
Erroneous proportions: Some analytics tools give you the option to create bubbles that are proportional by diameter or radius length rather than by area. While the differences are more dramatic, they are also inaccurate and misrepresentative of the values.		

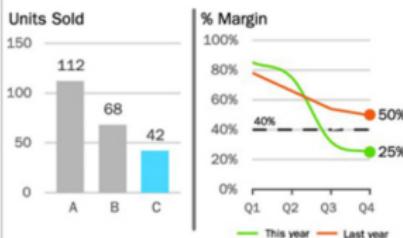
Three-step Process to Plan and Design Your Data Story

Sketch. Before you start creating any charts in an analytics tool, you might find it helpful to sketch a mock-up for each story point, especially if your story has multiple scenes. You can use a notepad, sticky notes, or a whiteboard to create a rough vision for how you're going to visualize the information. Because you're only investing a small amount of time to create each sketch, you're freed up to iterate quickly until you identify the best visual approach for each chart. The act of sketching your entire data story also gives you an opportunity to step back and ensure that your visuals are not too repetitive and that you have the right flow. You can also walk a colleague through your story to gather feedback and make modifications before you spend any time building charts.



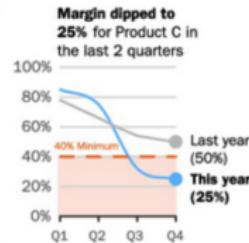
Three-step Process to Plan and Design Your Data Story

Design. Once you're happy with the rough visual outline of your data story, you can start designing the data visualizations. At this stage, you're focused on making sure the charts are aligned to your story points and convey the information in an effective manner. As you work with the actual data, you may discover you need to modify your charts from what you planned in the sketch phase. It's important to be flexible and defer to whatever approach will best communicate your insights.



Three-step Process to Plan and Design Your Data Story

Refine. After you've created the initial charts for your data story, you should evaluate how well each one supports your storyline. You may find small edits to your charts can dramatically enhance their overall effectiveness. For example, you may add direct labels to make a chart easier to read or realign the colors used in a chart to be more consistent. These kinds of simple but important refinements can add polish to your final scenes and help them resonate with your audience.



A Data-Driven Change Agent



To become a data-driven change agent, you need to be sufficiently data literate so you can understand and interpret the data correctly. Next, you need to be curious and free to explore the data to find meaningful insights. Last, you need to acquire data storytelling skills so you can communicate your insights in an effective manner.

Declaration & Acknowledgment

The contents presented in this slide are meant for teaching purpose only. No commercialized component exist. The texts are partially copied from the textbooks and images are downloaded from the internet.