DEPARTMENT *of* COMPUTING

College of Business & Technology

EAST TENNESSEE STATE UNIVERSITY

CSCI 1100
Using Information Technology

Data Literacy

Professor Ryan Haas

What is data literacy?

Data literacy:

The ability to read, understand, create, and communicate data.

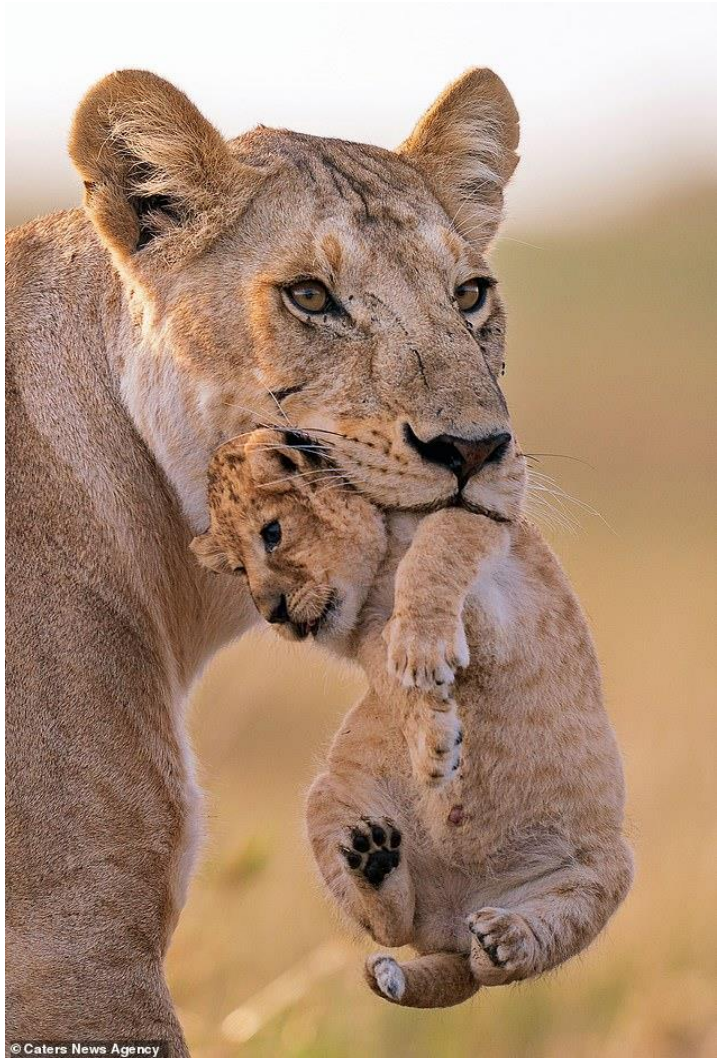
- Enables **informed decision-making** based on data analysis.

Key concerns:

- Understanding data sources and data types.
- Understanding the context in which data is used.
- Accurately interpreting trends and behaviors and making informed inferences.

Lion moms sometimes eat their cubs when prey is scarce

... The context we use to understand and present data matters



So. Much. Data…

Every day on the Internet…

34 million videos are posted on TikTok. [1]

1.3 billion photos are shared on Instagram. [2]

An estimated 402.74 million terabytes of data are created. [3]

1. Sendshort.ai. TikTok Statistics
2. Omnicoreagency.com. Instagram Statistics
3. Explodingtopics.com. How Much Data Is Generated Per Day?

Defining Data

Defining data

Data Definitions:

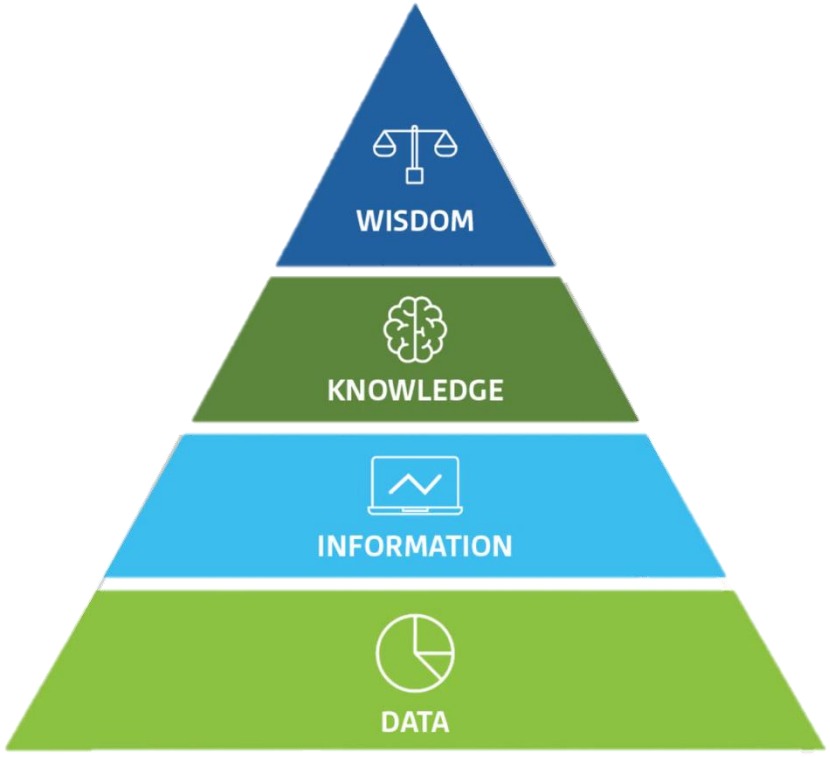
1. Values representing a specific concept or concepts. [1]

2. “raw figures, numbers, or text that serve as the starting point of analysis.” [2]

- Data becomes *information* when we **analyze** it and possibly **combine** it with other data to derive **meaning** and **context**. [1]

[1] Data.gov. Glossary: Data vs. information.

[2] Kalé Nitin and Nancy Jones (2020). *Practical Analytics* (2nd ed.).



Above: The DIKW model, which we will explore in more detail.

Defining data: Qualitative vs. Quantitative

Qualitative data

Descriptive data that cannot be measured numerically, categories.

Ex.: Opinions, observations, experiences, months, demographic groups.

Quantitative data

Numerical data that can be measured and analyzed statistically.

Ex.: Sales figures, test scores, experiment results.

Defining data: Qualitative, Quantitative – What's the use?

Qualitative

Allows for rich, in-depth insights into behaviors, attitudes and experiences.

Quantitative

Allows for statistical analysis, trend identification, and data-driven decision-making.

Combining both!?

Allows for deeper, more comprehensive understanding of a topic.

Defining data: Structured vs. unstructured

Structured data

Typically **quantitative**—highly organized (standard formats).

Ex.: Dates, names, addresses, credit card numbers.

Unstructured data

Typically **qualitative**—lacks a predefined format.

Ex.: text and social media posts, Internet of Things (IoT) sensor data.

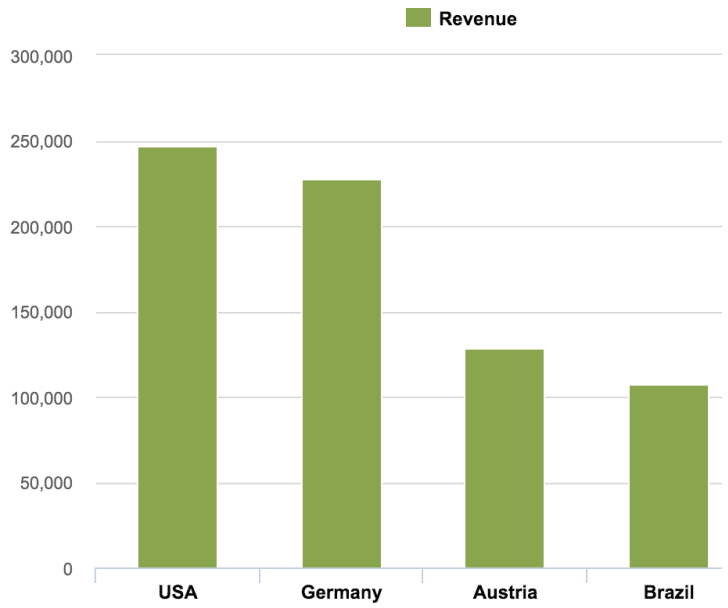
IBM (2021). Structured vs unstructured data.

Visualizing Data

Common chart types

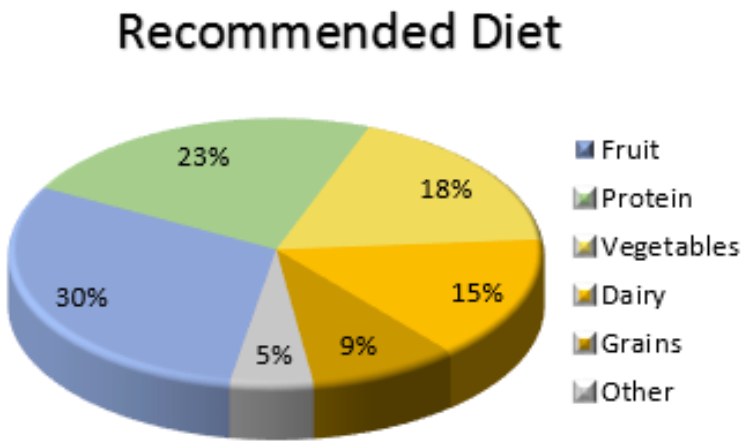
Column/bar Chart

Compare values for different categories or compare value changes over a period of time for a single category.



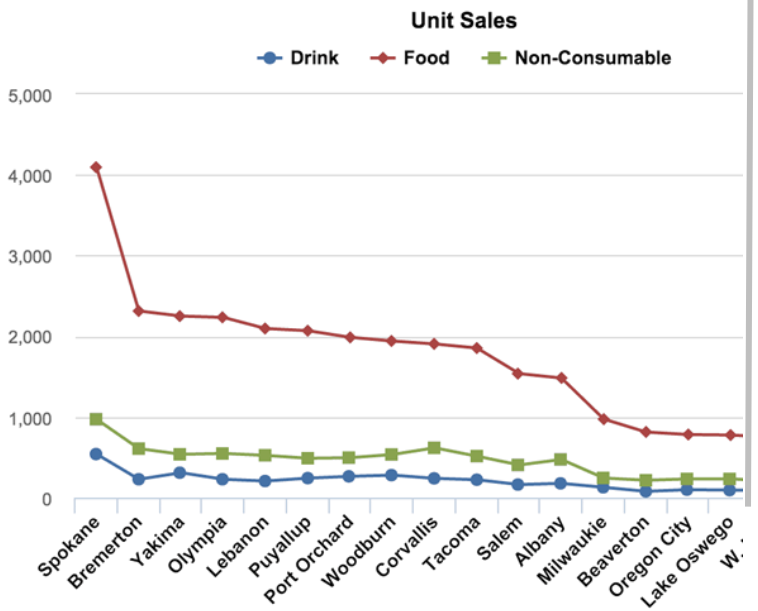
Pie Chart

Visualize parts of a whole (percentages), when comparing exact values is not the goal.



Line Chart

Emphasize the continuation/flow of values (trends) over time (time series data).



When to use different chart types

Column/bar Chart

Best for: Comparing values across categories.

Dos and Don’ts:

- Start the numerical axis at zero! Our eyes are sensitive to height.
- Don’t use for more than 5-7 categories.
- If you represent time, always put it on the x-axis. Progress from left to right.

Pie Chart

Best for: Showing proportional relationship of parts of a whole.

Dos and Don’ts:

- Ensure the sum of the segments is 100 percent.
- Don’t use for more than 5 categories, ideally, only two categories E.g., men compared to women; market share and compared to whole market.
- Don’t use a pie chart if category values are very similar or completely unrelated.

Line Chart

Best for: Showing trends over time.

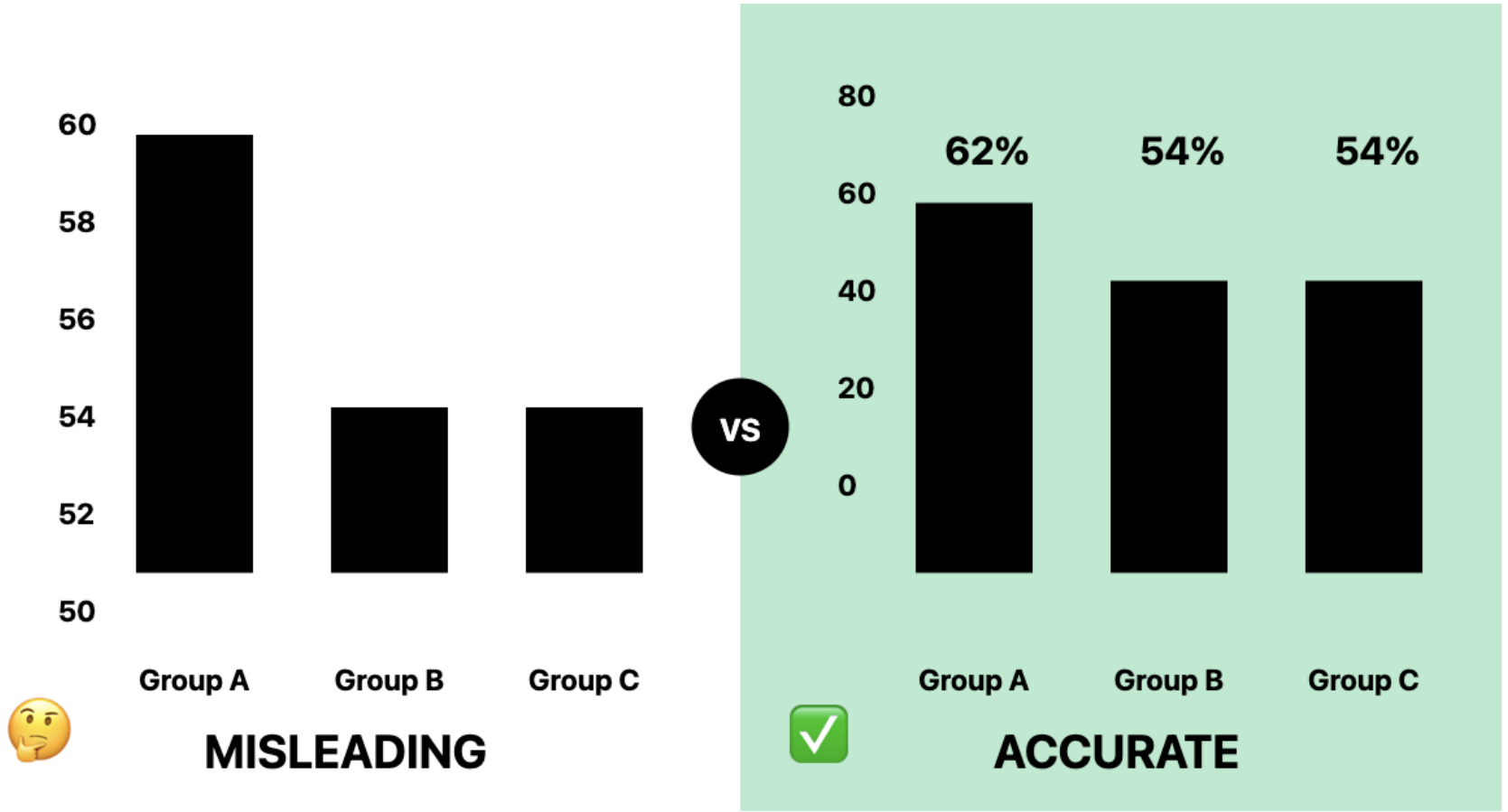
Dos and Don’ts:

- Typically, start the axis at zero.
- Don’t skip values for consistent data intervals that present trend information. E.g., don’t skip days with zero values.
- Once again, time goes on the x-axis and should progress from left to right.

(Mis) Interpreting Charts

Misleading charts: Omitting the Baseline

- Starting with a vertical axis at 50 stretches the trend making the difference among the groups seem massive.
- We are only looking at an 8% difference!

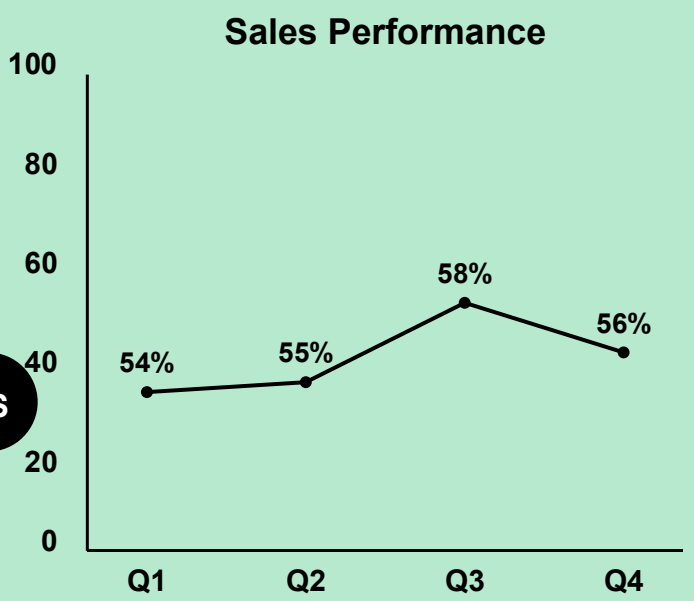


Misleading charts: Manipulating the y-axis

- This time, the trick is not removing the baseline but using the scale to make the performance look steeper.
- So we could exaggerate a spike if or flatten it, depending on what we were trying to “sell”.

🤔 **MISLEADING**

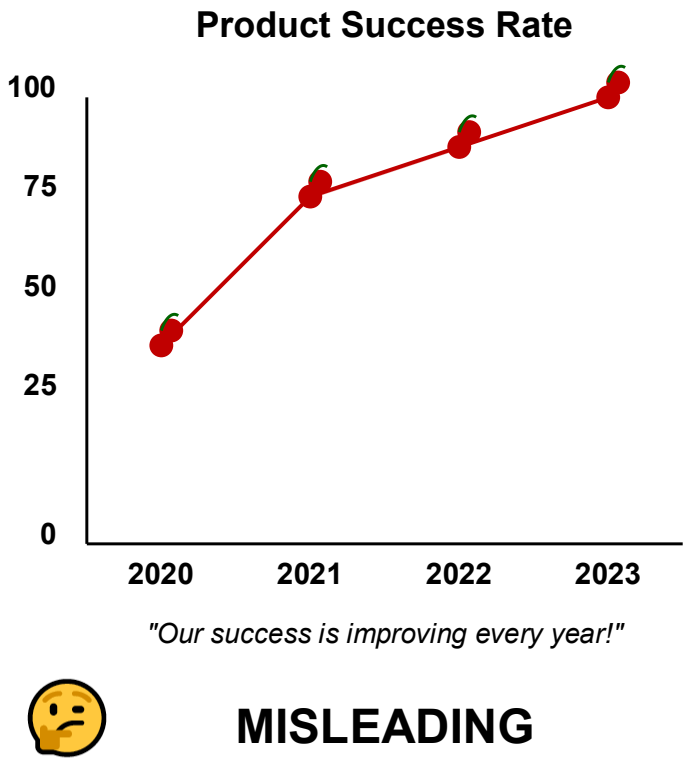
VS



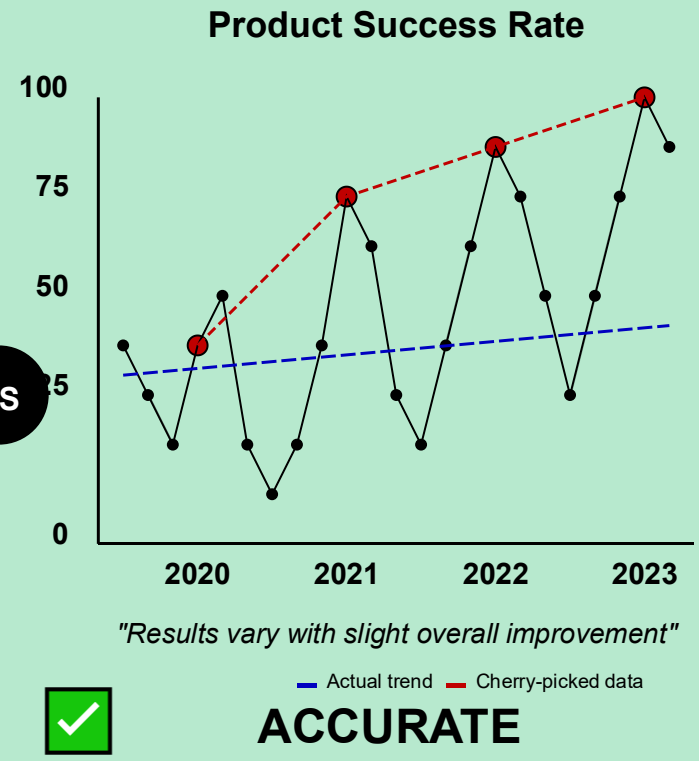
✅ **ACCURATE**

Misleading charts: Cherry-picking

- 4 carefully selected data points: strong upward trend.
- Doesn't quite resemble the average shown by our linear regression (blue) trendline, does it?

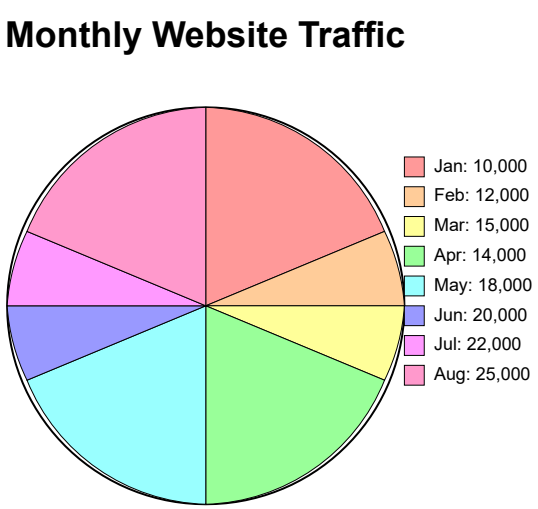


VS



Misleading charts: Wrong chart type!

- August does account for most traffic, but you wouldn't know it by looking at the pie chart on the left.
- A trend over time should be represented with a line chart!



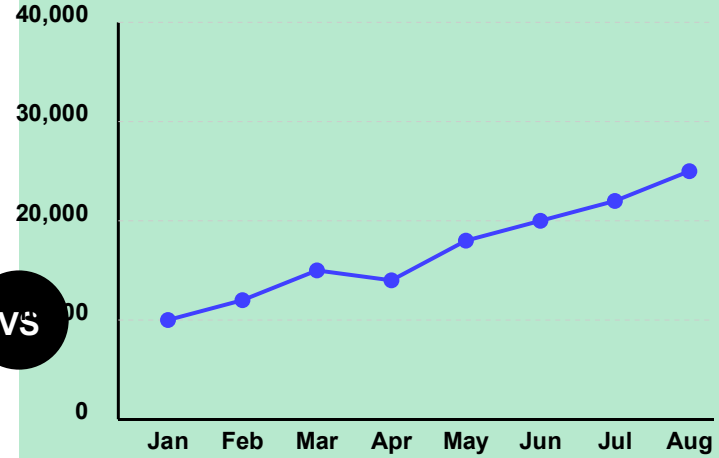
"August accounts for most traffic!"

Pie charts show parts of a whole, not trends over time!

🤔

MISLEADING

VS



"Traffic has consistently increased over time"

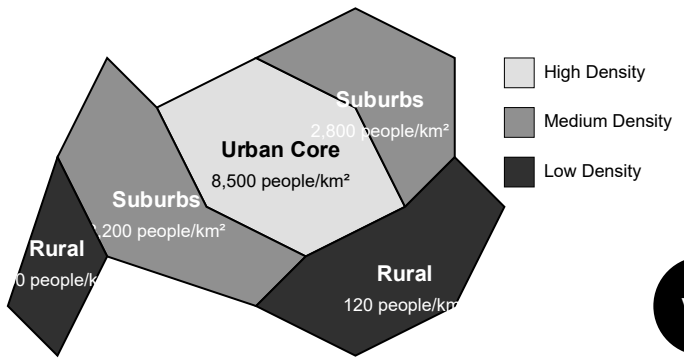
Line charts clearly show trends and changes in values over a time period

✅

ACCURATE

Misleading charts: Wrong chart type!

- Convention:
  - lower value = greater density
  - higher value = lesser density
- Can’t help but perceive the rural areas as most dense, urban areas as least.
- We love when you’re creative, just not with your charts 😉

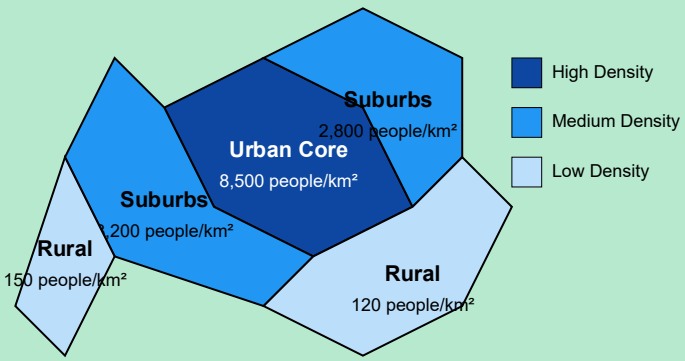


"Rural areas appear more prominent!"

Lighter colors for higher density breaks convention
Dark areas draw more attention to less populated regions

🤔 **MISLEADING**

VS



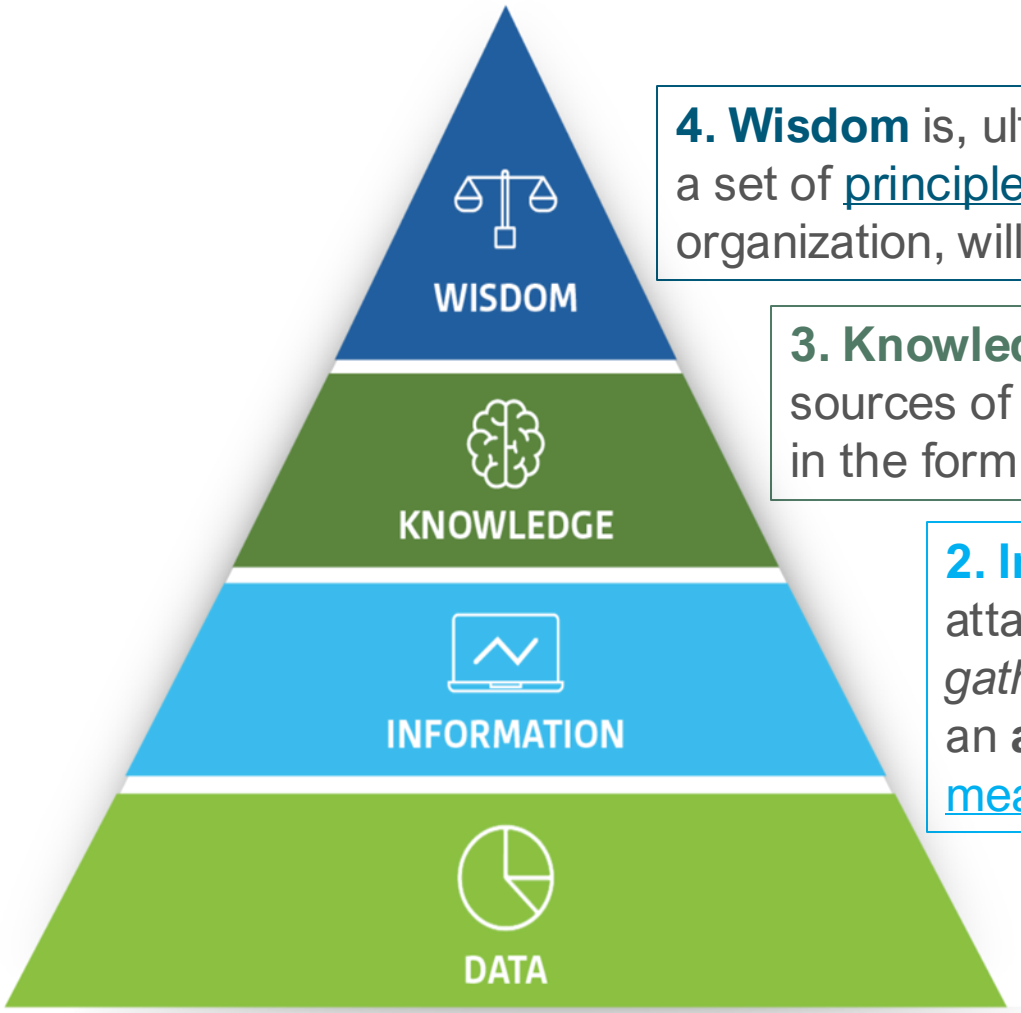
"Urban density is immediately visible"

Darker colors for higher density follows convention
Focus naturally drawn to population centers

✅ **ACCURATE**

The Big Picture

The big picture: Why we care about data



4. Wisdom is, ultimately, a state of awareness. It's a set of principles by which a person, or organization, will act.

3. Knowledge represents the **synthesis** of multiple sources of information over time. Knowledge often comes in the form of frameworks and theories.

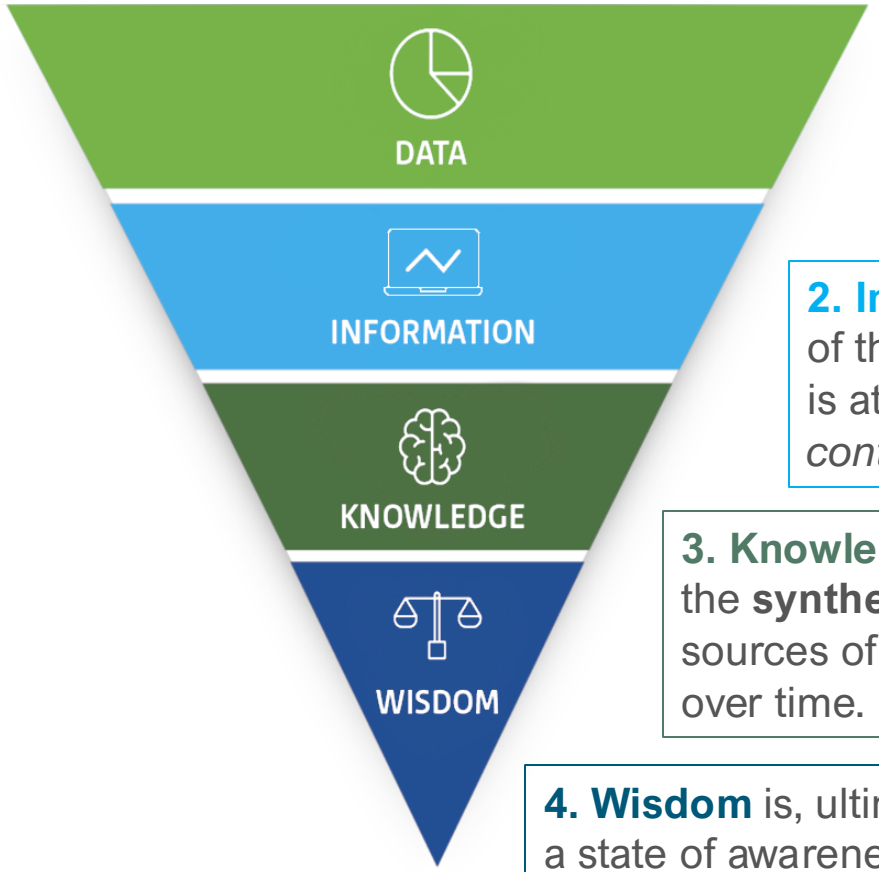
2. Information sits on top of the data foundation and is attained by providing *context* to data. Information is *gathered* via an **analytics platform** or *deciphered* by an **analytics team** which begins to give the data meaning and purpose.

1. Data are the raw figures, numbers, or text that serve as the starting point of analysis.

Above: The DIKW Model

Kalé Nitin and Nancy Jones (2020). *Practical Analytics* (2nd ed.).

The big picture: A process of synthesis



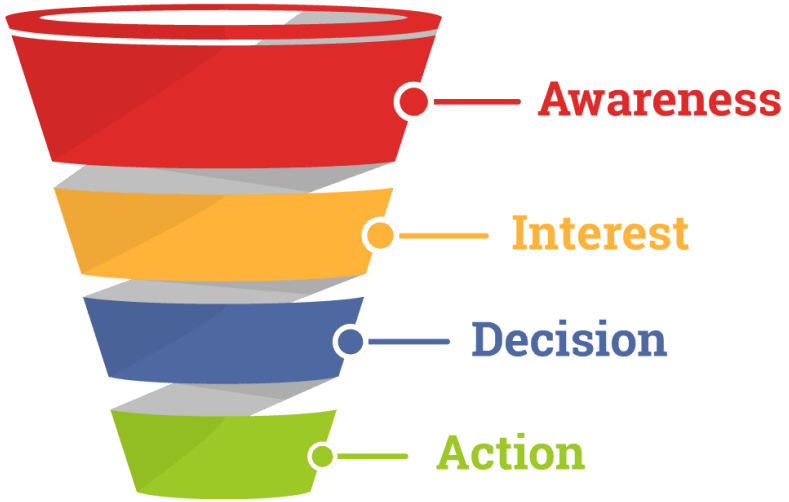
1. Data are the raw figures, numbers, or text that serve as the starting point of analysis.

2. Information sits on top of the data foundation and is attained by providing *context* to data.

3. Knowledge represents the **synthesis** of multiple sources of information over time.

4. Wisdom is, ultimately, a state of awareness. It's a set of principles.

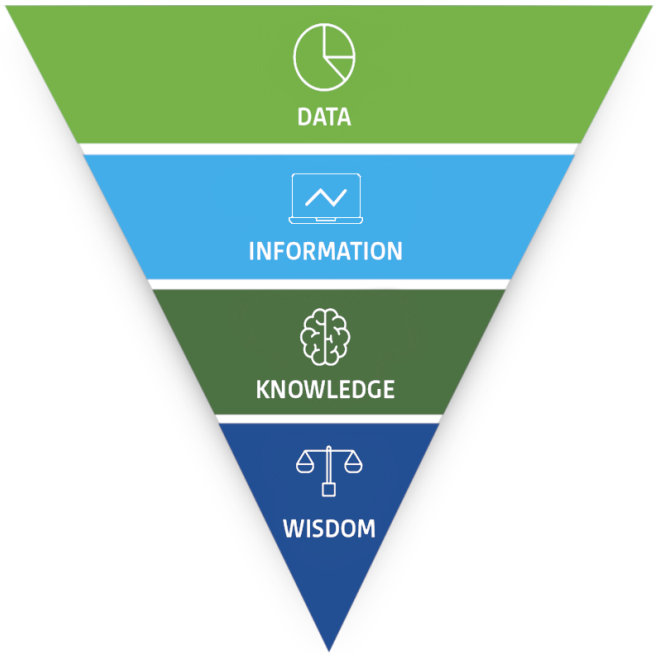
Kalé Nitin and Nancy Jones (2020). *Practical Analytics* (2nd ed.).



Above: A sales funnel, sometimes divided into more stages.

Similar to how the volume of people diminishes at each stage in sales toward a customer's action, each level in DIKW is a further abstraction from the original data.

DIKW scenario: Eggs and milk



When a customer comes to the store to buy eggs and milk, they're seeing the result of this entire DIKW process.

The store's ability to have the right products available at the right time is the culmination of:

- **data** collection,
- processing that data into useful **information**,
- applying **knowledge** to predict demand and manage stock, and…
- using **wisdom** to make strategic decisions that satisfy customer needs while optimizing operations.

DIKW scenario: Data

Raw facts: The grocery store collects data such as:

- The number of eggs and milk cartons sold each day.
- Dates and times of sales.
- Inventory levels.
- Delivery schedules.
- Customer preferences (e.g., organic vs. conventional products).
- External factors like weather reports that might impact shopping habits.

DIKW scenario: Information

<u>Processed data → information:</u> The store processes the raw data to generate:

- **Sales reports**: Summarizing the daily, weekly, or monthly sales of eggs and milk.
- **Inventory status**: Current stock levels and the rate at which they are depleting.
- **Customer trends**: Identifying patterns, such as a higher demand for eggs on weekends or an increased purchase of milk during colder months.
- **Supplier performance**: Tracking the delivery times and quality of shipments from suppliers.

DIKW scenario: Knowledge

<u>Understanding and insights:</u> The store applies knowledge by:

- **Forecasting demand**: Using historical sales data to predict future needs, ensuring that enough eggs and milk are stocked without overstocking.
- **Stock management**: Knowing when to reorder products based on the rate of sale and lead time from suppliers.
- **Customer behavior**: Understanding that promotions or seasonal events (like baking during holidays) might increase demand for eggs and milk.

Examples: Understanding Customer Behavior Matters!

Market Basket Analysis

- Optimized store layout + targeted promotions = 💰

Customer Segmentation

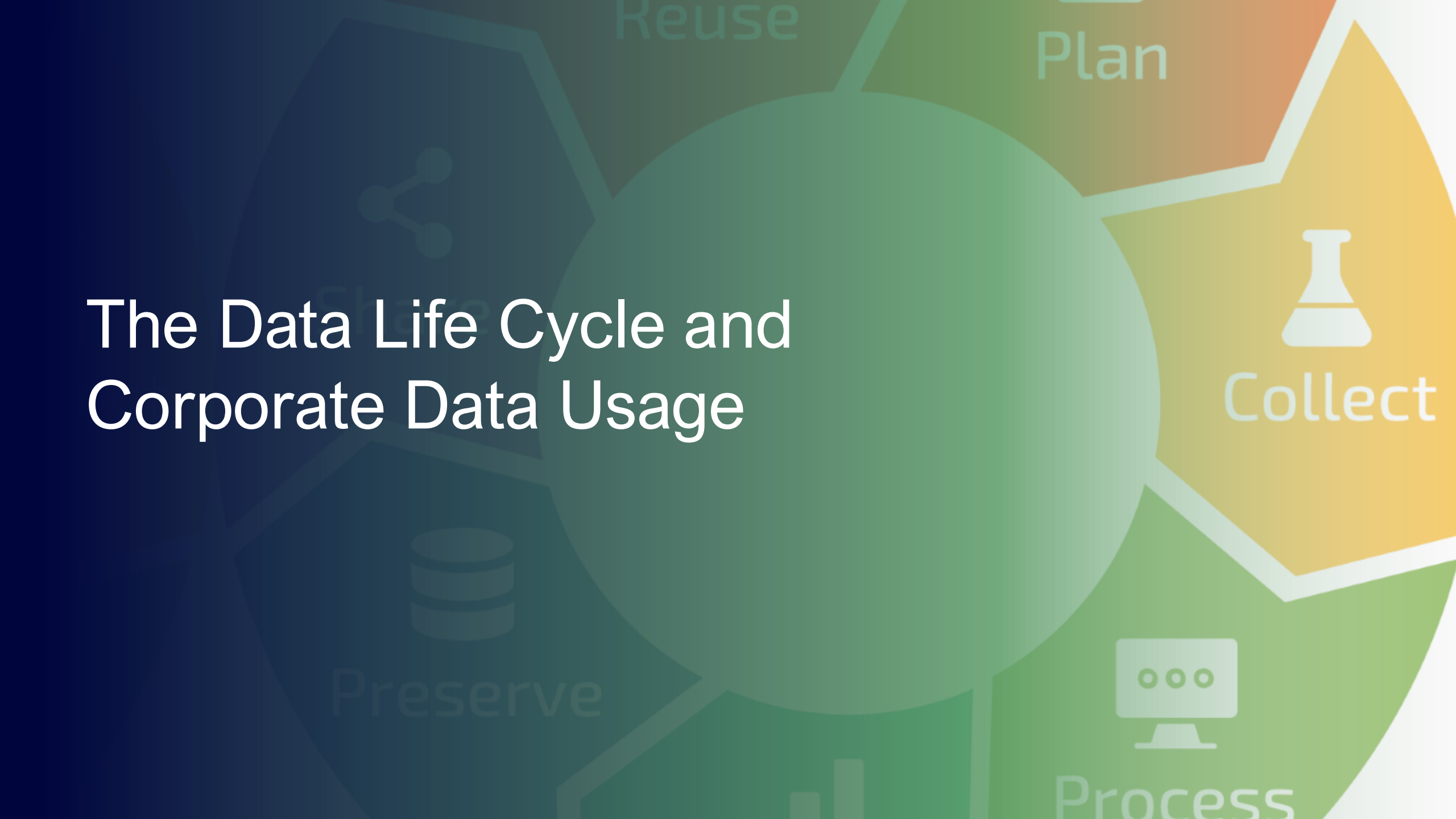
- Targeted promotions + customer loyalty = 💰

DIKW scenario: Wisdom

Decision-making and strategy: The store uses wisdom to make informed decisions, such as:

- **Optimizing inventory**: Deciding to stock more of certain types of milk (e.g., lactose-free) if customer preferences shift, or adjusting order quantities based on upcoming holidays.
- **Supplier relationships**: Choosing to partner with suppliers who provide the best balance of cost, quality, and reliability to meet customer demands efficiently.
- **Sustainability practices**: Implementing strategies to reduce waste, like using data to track expiration dates and optimize the supply chain, minimizing spoilage.

The Data Life Cycle and Corporate Data Usage

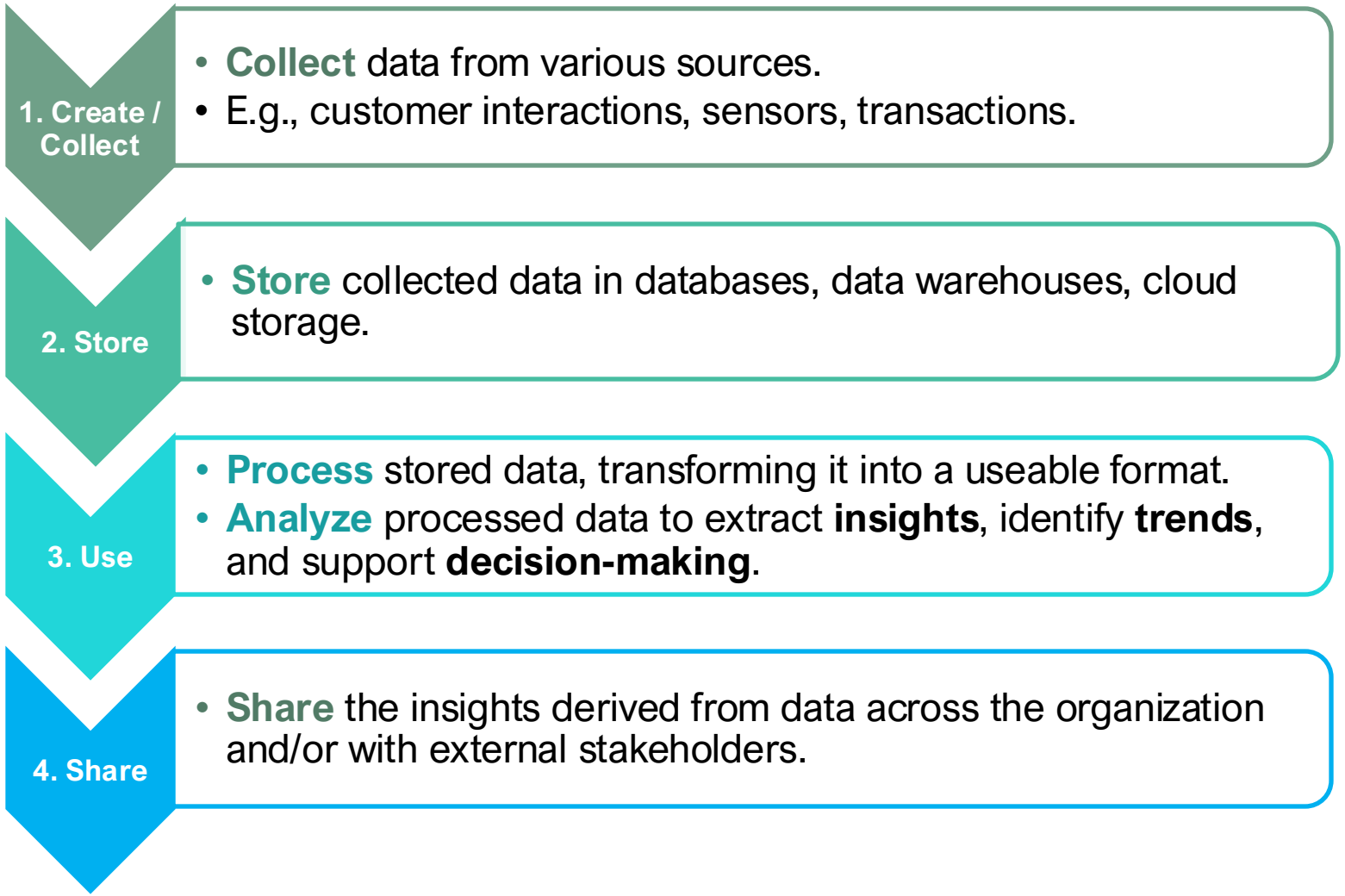


What is the Data Life Cycle?

Data Life Cycle (DLC): The stages that data goes through from its initial creation or collection to its eventual disposal.

- It's a systematic process that ensures data is properly managed, used, and protected throughout its life.
- Understanding the DLC is crucial for organizations to harness data effectively and comply with regulations.

Stages of the Data Life Cycle



Stages of the Data Life Cycle

5. Archive

- **Store** data that is no longer actively used for the long term, typically for legal, historical, and recovery purposes.

6. Destroy

- **Destroy** or delete data that is no longer needed. Data should be securely destroyed to protect sensitive information.

How do you think the DIKW model aligns with these stages?

- Where is information, knowledge, and wisdom created in this life cycle?

Why we care about the DLC

1. Efficiency:

Proper management of data throughout its life cycle ensures that data is always available when needed, reducing delays in **decision-making**.

2. Compliance:

Adhering to the data life cycle helps organizations comply with legal and regulatory requirements, such as **data protection** laws.

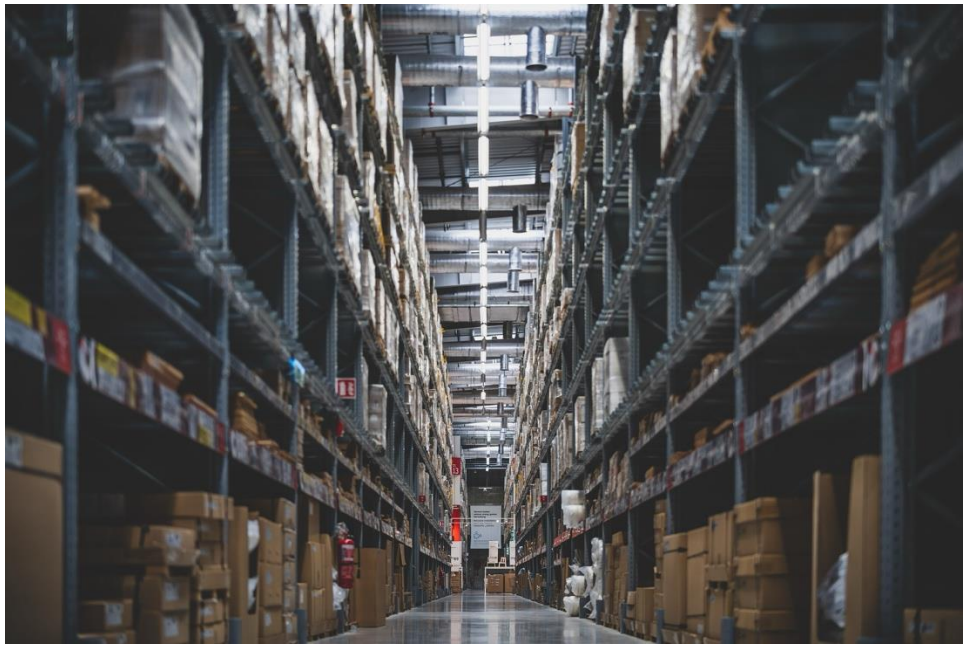
3. Cost management:

Effective data management **reduces costs** associated with storage, processing, and data breaches.

4. Security:

Managing data through its life cycle ensures that sensitive data is **protected** from unauthorized access and misuse.

Case study: Data-driven decision-making



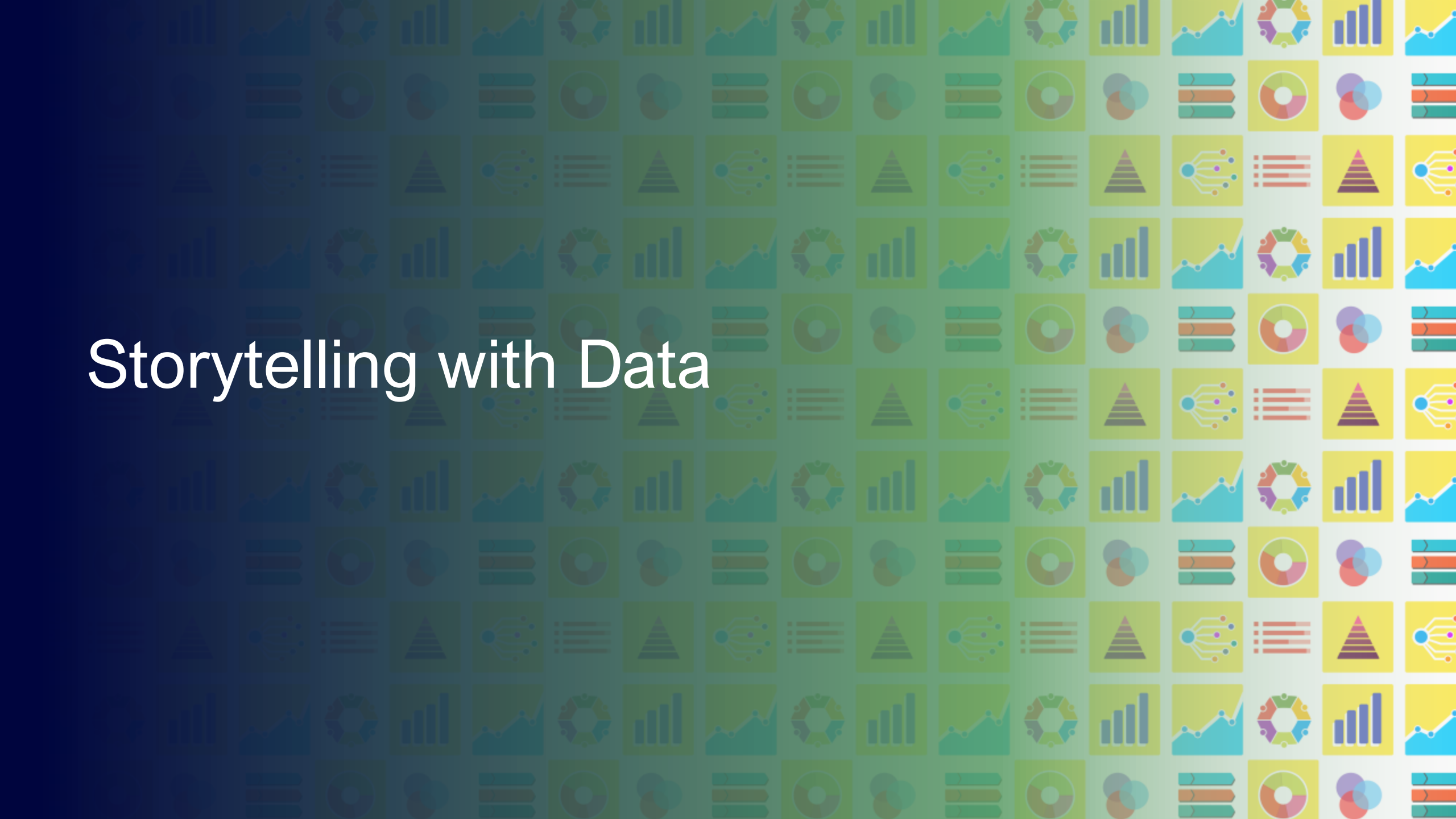
Amazon uses vast amounts of data to optimize its supply chain, personalize customer recommendations, and improve delivery efficiency.

- **Example**: Amazon's predictive analytics allow it to anticipate customer orders and stock items in warehouses close to the customers *before* they even place the order!
- **Impact**: Reduced delivery times, greater customer satisfaction, and competitive edge in the retail market.

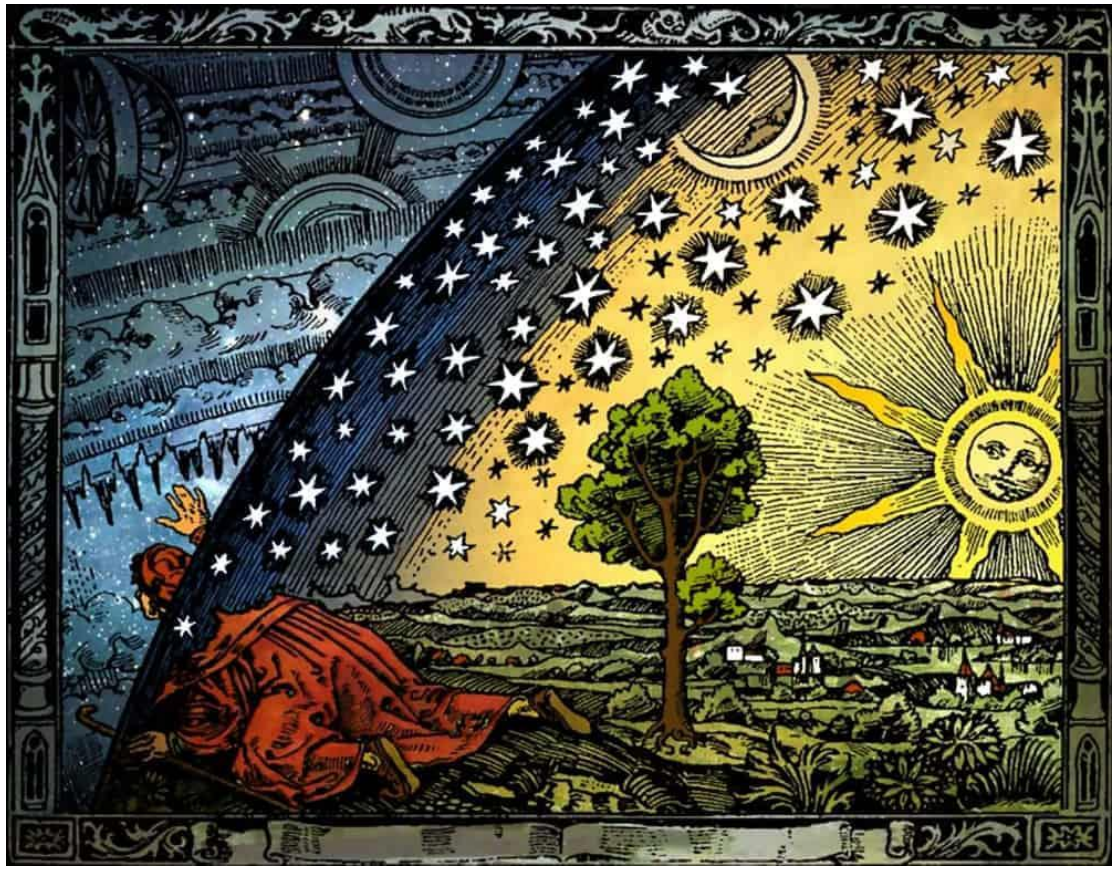
Amulyaa Vaisshnavi Kalvakuntla (2023). How Amazon Uses Data Analytics to Deliver Packages Faster.

Greg Bensinger (2014). Wall Street Journal. Amazon Wants to Ship Your Package Before You Buy It.

Storytelling with Data



Storytelling with data

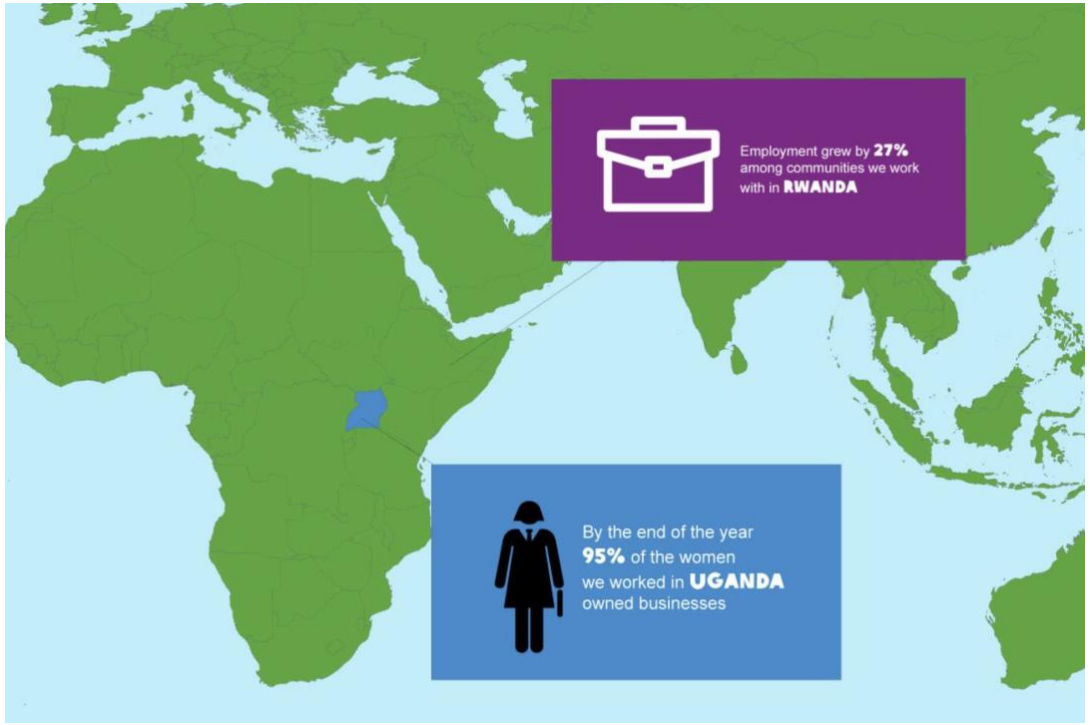


Incorporating data visualizations into stories can make data more engaging, understandable, and memorable.

- Narratives simplify complex trends, making them easier to understand and interpret.
- Stories are persuasive!

Storytelling with data: Elements of a data story

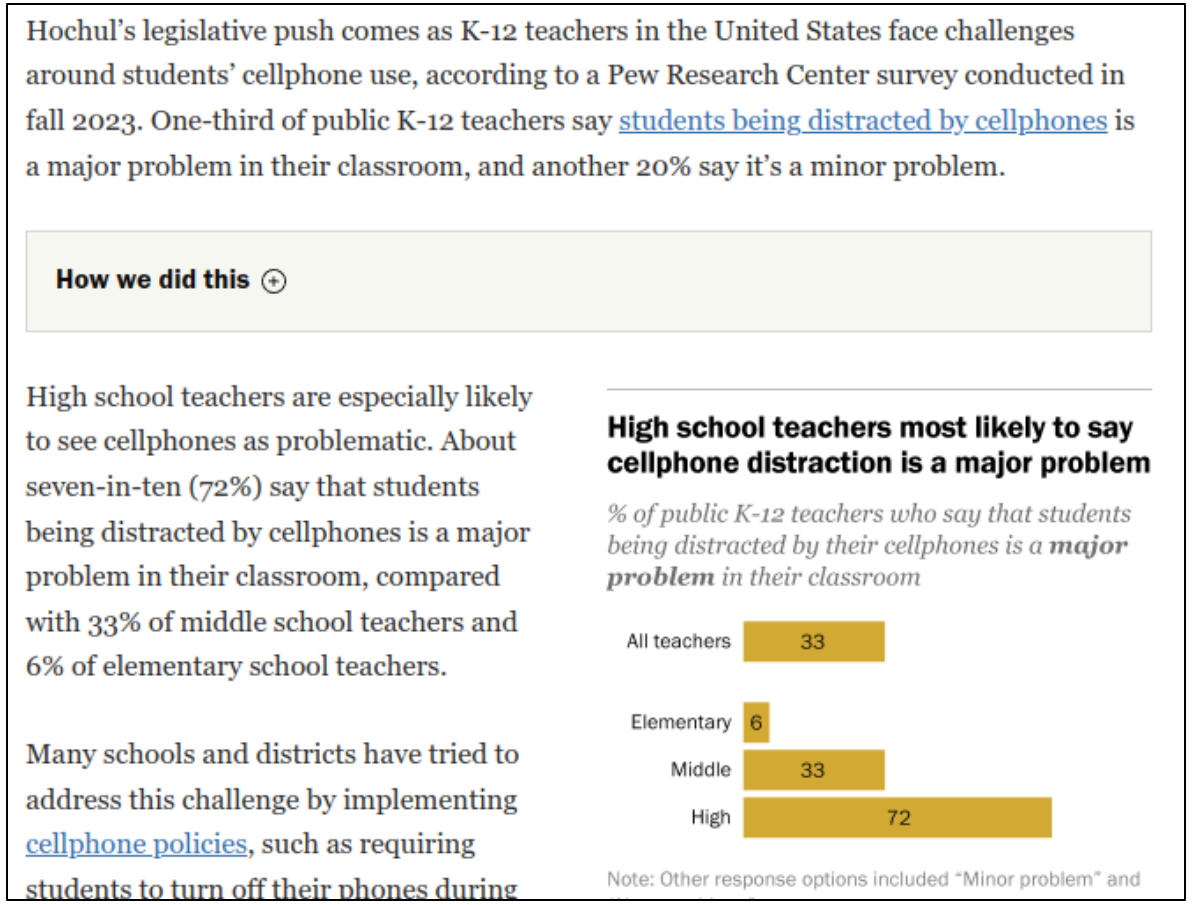
- A **clear storyline** guides the audience through the data.
- **Visualizations** support the narrative, highlighting key insights.
- A **call to action** compels the audience to act on the information.



Above: NGO Oxfam Ireland track and report on their goals using an interactive annual impact report to tell their year's story with data.

Storytelling with data: Example

- The article tells the story of K-12 teachers “fac[ing] challenges” concerning student cellphone use.
- The charts help us see why certain cellphone policies may be implemented.



Hochul’s legislative push comes as K-12 teachers in the United States face challenges around students’ cellphone use, according to a Pew Research Center survey conducted in fall 2023. One-third of public K-12 teachers say students being distracted by cellphones is a major problem in their classroom, and another 20% say it’s a minor problem.

How we did this ⊕

High school teachers are especially likely to see cellphones as problematic. About seven-in-ten (72%) say that students being distracted by cellphones is a major problem in their classroom, compared with 33% of middle school teachers and 6% of elementary school teachers.

Many schools and districts have tried to address this challenge by implementing cellphone policies, such as requiring students to turn off their phones during

High school teachers most likely to say cellphone distraction is a major problem

*% of public K-12 teachers who say that students being distracted by their cellphones is a **major problem** in their classroom*

Note: Other response options included “Minor problem” and

Above: A June 2024 Pew Research article tells the story of how smartphone use is perceived by teachers and students in public schools amidst K-12 teachers’ challenges with student cellphone use.

Storytelling with data: Best practices

Keep it simple

Avoid overwhelming your audience with too much data; focus on key insights.

Use consistent visuals

Ensure your visualizations are clear and aligned with the narrative.

Focus on storyline

Make sure the data supports the story, not the other way around.