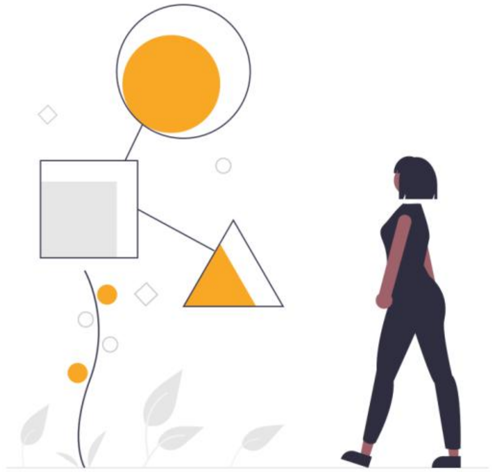
1. Introduction to Data Analytics

Part 1

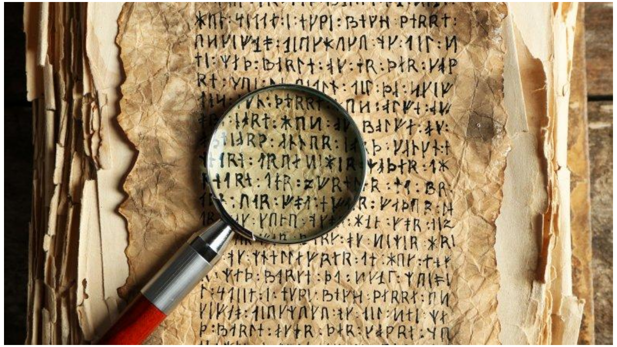


Khalil Israfilzada, PhD
Faculty of Economics and Management
Vytautas Magnus University
Kaunas, 2025

Why Data is Important?

Have you ever been fascinated with ancient languages, perhaps those now known as “**dead**” languages? The complexity of these languages can be mesmerizing, and the best part about them is the extent to which ancient peoples went to preserve them.

Using **ink** made from burned **wood**, **water**, and oil they **copied** the **text** to **papyrus paper**. Some used tools to chisel the text into pottery or stone.



What is the commonality between dead languages and business analytics?

Why Data is Important?

Over 5,000 years ago, the ancient Mesopotamians started to record **quantities on clay tablets**. They partitioned the tablet into **rows** and **columns**. Within each cell, they drew a picture of the type of item and made holes indicating the quantity of it. Each type of item had its own standard pictographic representation, making this ledger language one of the earliest form of human writing we've discovered. It's called **"Proto-cuneiform**" because it later evolved into a complete written language called "Cuneiform".

In other words, the ancients invented Excel before Word!

Why Data is Important?

When it comes to business, product and market data can provide an edge over the competition. That makes this data worth its **weight in gold** (maybe oil?). Important data can include weather, trends, customer tendencies, historical events, outliers, products, and anything else relevant to an aspect of business. What is different about today is how data can be stored. It no longer has to be hand-copied to papyrus or chiselled into stone. It is an automatic process that requires very little human involvement and can be done on a massive scale.

Why Data is Important?

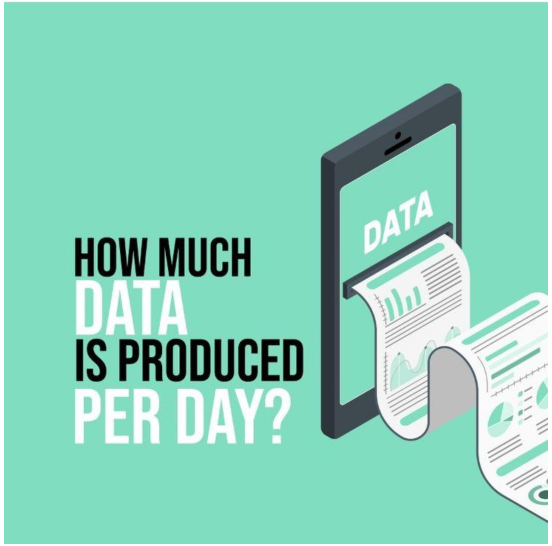
Today, gathering data to help you better understand your customers and business is relatively easy. In fact, it's become so easy there's **the danger of having too much data to deal with.**

In a recent article, data and analytics guru Bernard Marr said:

"While the average small business has less self-generated data than big players. . .this doesn't mean big data is off limits. In fact, in many ways, big data is more suited to small businesses, because **they're generally more agile and able to act more quickly on data-driven insights**."

Big Data Overview

Data is created constantly, and at an ever-increasing rate. **Mobile phones**, **social media**, **imaging technologies** to determine a **medical diagnosis**—all these and more **create new data**, and that **must be stored** somewhere for **some purpose**. Devices and sensors automatically generate diagnostic information that needs to be stored and processed in real time.



Big Data Overview

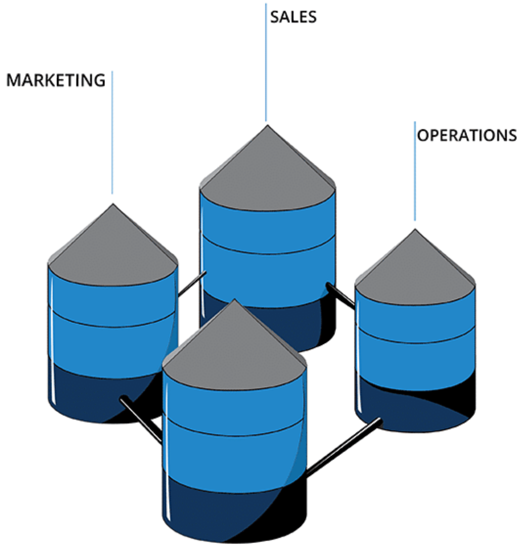
- Gartner is an independent analysis firm that reports on the technology sector. They phrased one need for the data in several comments on their website including the need for members of an organization ***to speak the same language.***

Imagine an organization where the marketing department speaks French, the product designers speak German, the analytics team speaks Spanish and **no one speaks a second language**… That's essentially how **a data-driven business functions** when **there is no data literacy.**

Big Data Overview

- We have spent hundreds of billions of dollars collecting data, but most of it sits in silos. Silos of data never analyzed, never touched again. Not only a **sunk cost** to acquire but also the cost to maintain, backup or archive – the cost is tremendous. Everyone has talked about monetizing this data, but few are very successful.

Data Silos = More Team Members Taking Longer to Achieve Less



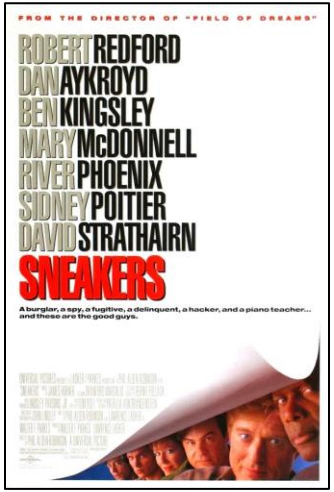
To be clear, it is not technology that is hampering progress.

It is lack of vision, human capital, and execution.

Big Data Overview

Cosmo: **The world isn't run by weapons anymore, or energy, or money. It's run by little ones and zeroes, little bits of data. It's all just electrons.**

Sneakers (1992), movie



Data costs include the **cost to acquire it**, the cost to **store** it, the **legal liability** of **keeping** it, the potential **risk** of a **data breach**. We normally think that data is cheap, but when you consider the **total cost of ownership**, it is **really quite expensive.**

Big Data Overview

- **Technology Keeps Raging, but We Need More Than Technology to Be Successful**

Technology is raging and has been for several years and **data is the new oil.** Much of the growth in the last few years could be loosely described as **“creating value from data.”** Value could mean increasing sales revenue, reducing avoidable costs, improving patient satisfaction, targeting high-value customers and prospects, creating policy for the broadest social good and much, much more.

Big Data Overview

- There is a HUGE overlap in the areas where value is created from data. This is not by accident as we will explain. But, for now take a look at the dizzying list of overlapping subject areas:

- Business Intelligence (BI)
- Visual BI, Analytics
- Visual Analytics
- Business Analytics
- Data Analytics
- Predictive Analytics
- Prescriptive Analytics
- Advanced Analytics
- Big Data
- Data Science

- Text Analytics
- Graph Analytics
- Social Analytics
- Network Analytics
- Modern Analytics
- Directed Acyclic Graph (DAG) Analytics
- Statistics
- Optimization
- Data Mining
- Data Modeling
- ML

- Decision Science
- Enterprise or Business Decision Management
- Business Process Management
- Data Engineering
- AI
- Computational Intelligence
- Management Science
- Linear and Mathematical Programming
- Deep Learning, Informatics
- Decision Science
- Many others

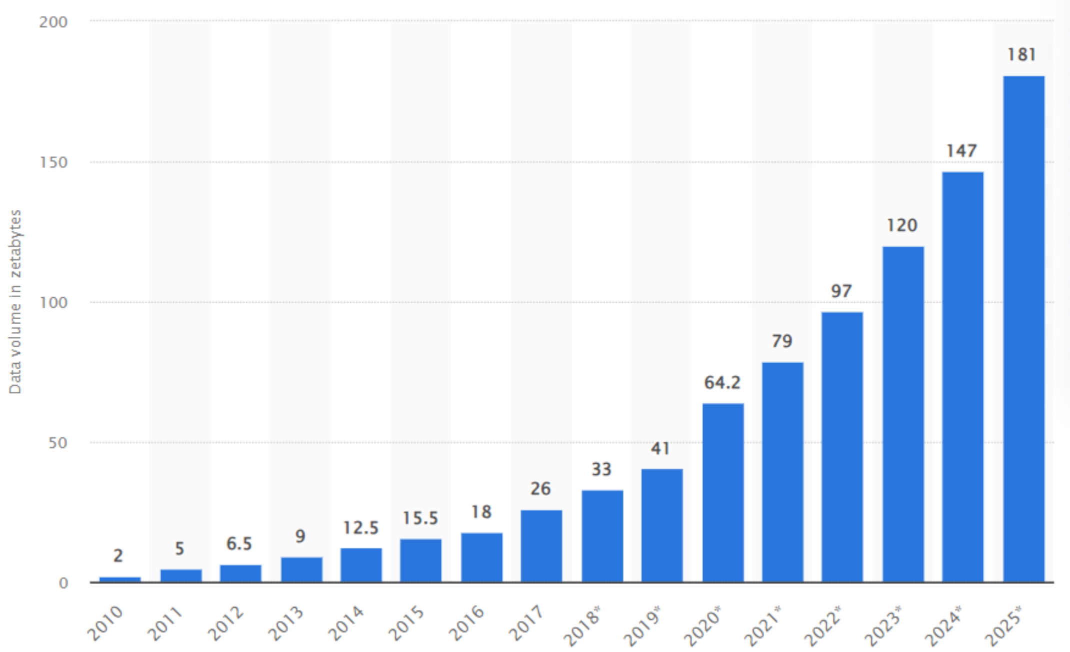
Big Data Overview

Therefore, there is a need for simplification, for **"One Global Term"** that applies to each of these. It is believed ***analytics*** is the appropriate umbrella term that captures the spirit of all these methods; the term ***analytics*** will be used frequently when it is not focusing on a specific form.

! Analytics is the process of *extracting* and *creating* information from raw data by *filtering*, *processing*, *categorizing*, *condensing* and *contextualizing* the data.

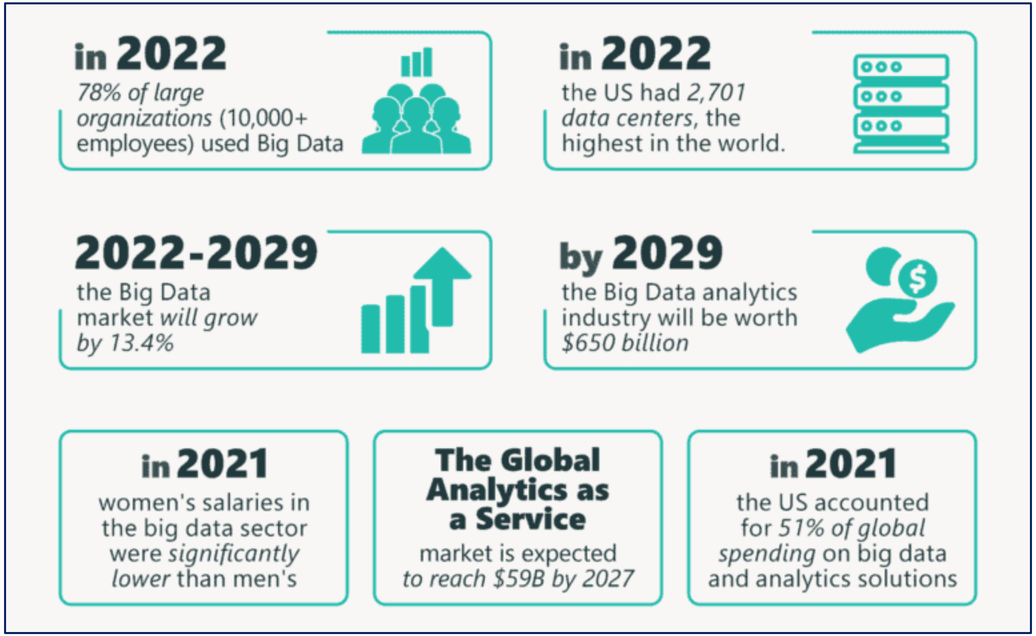
Data and Analytics Explosion

The volume of data collected is growing exponentially with no end in sight. In November 2018 there were 5 billion consumers that interacted with data, but by 2025 it will be 6 billion or 75% of the world's population. In 2025 there will be 150 billion devices creating data in real time.

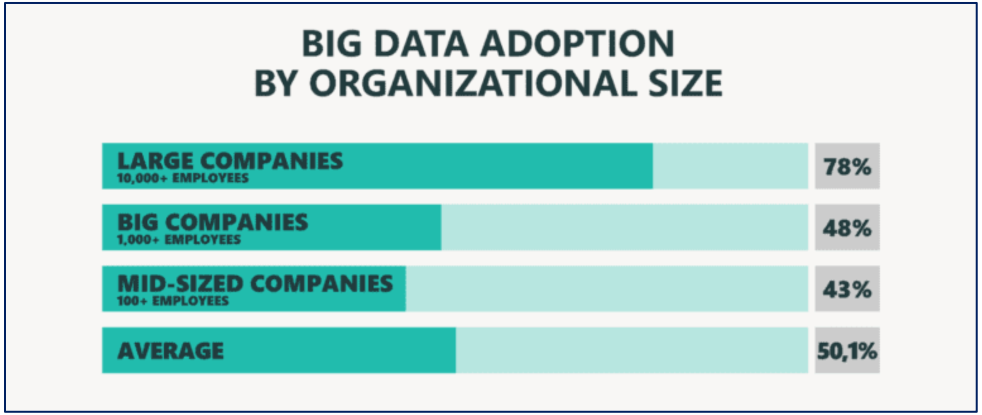


Data and Analytics Explosion

The market was worth $240 billion in 2021 and is projected to grow considerably over the next few years to around **$650 billion in 2029**. From 2020 to 2024, the ratio of unique data to duplicated data is predicted to decrease gradually from 1:9 to 1:10.



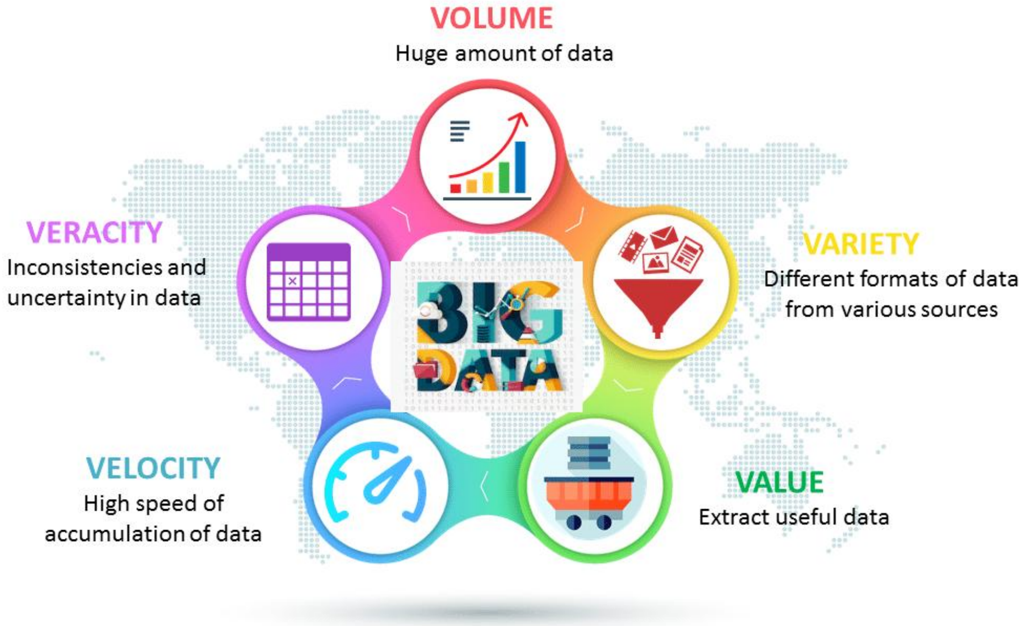
Data and Analytics Explosion



FUN FACT! According to these numbers, there is a probability that **YOU** will end up working for a company that has implemented big data analytics. The probability is approximately **FIFTY** per cent.

Characteristics of Big Data

As with anything huge, we need to make proper categorizations in order to improve our understanding. As a result, features of big data can be characterized by **five Vs.: volume, variety, velocity, value,** and **veracity.**



Characteristics of Big Data

Volume. Big data is a form of data whose volume is so large that it would not fit on a single machine therefore specialized tools and frameworks are required to store process and analyze such data.

Velocity. Velocity of data refers to how fast the data is generated. Data generated by certain sources can arrive at very high velocities, for example, social media data or sensor data.

Variety. Variety refers to heterogeneous sources and the nature of data, both structured and unstructured.

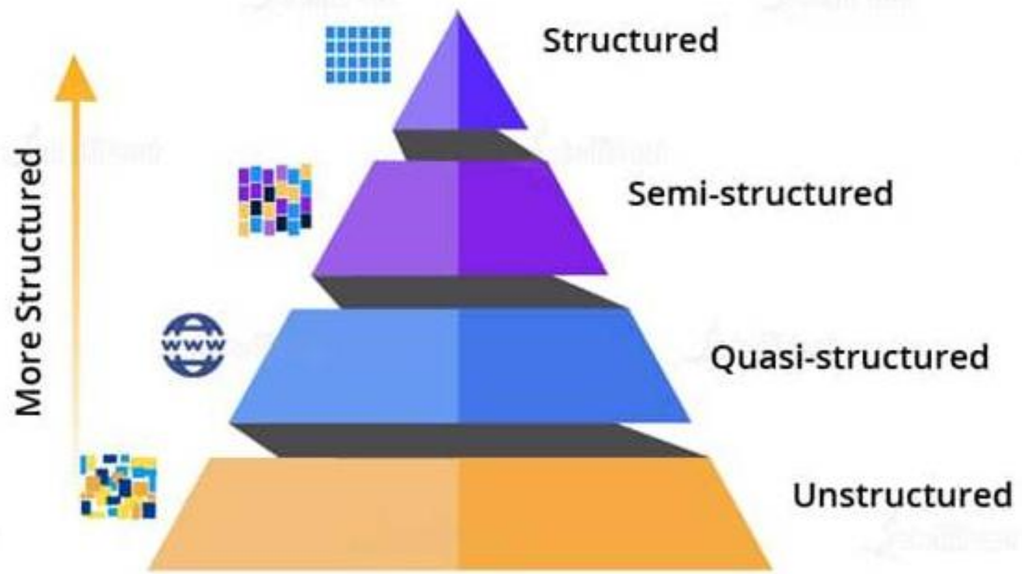
Characteristics of Big Data

Veracity. Veracity refers to the trustworthiness and quality of the data. If the data is not trustworthy and/or reliable, then the value of Big Data remains unquestionable. Veracity refers to how accurate is the data. To extract value from the data, the data needs to be cleaned to remove noise.

Value. Value of data refers to the usefulness of data for the intended purpose. The end goal of any big data analytics system is to extract value from the data.

Data Structures

Big Data comes in many forms; it can be Structured, Semi-Structured, Quasi-Structured or Unstructured as in figure shows the Data structure types and the data growth accordingly.



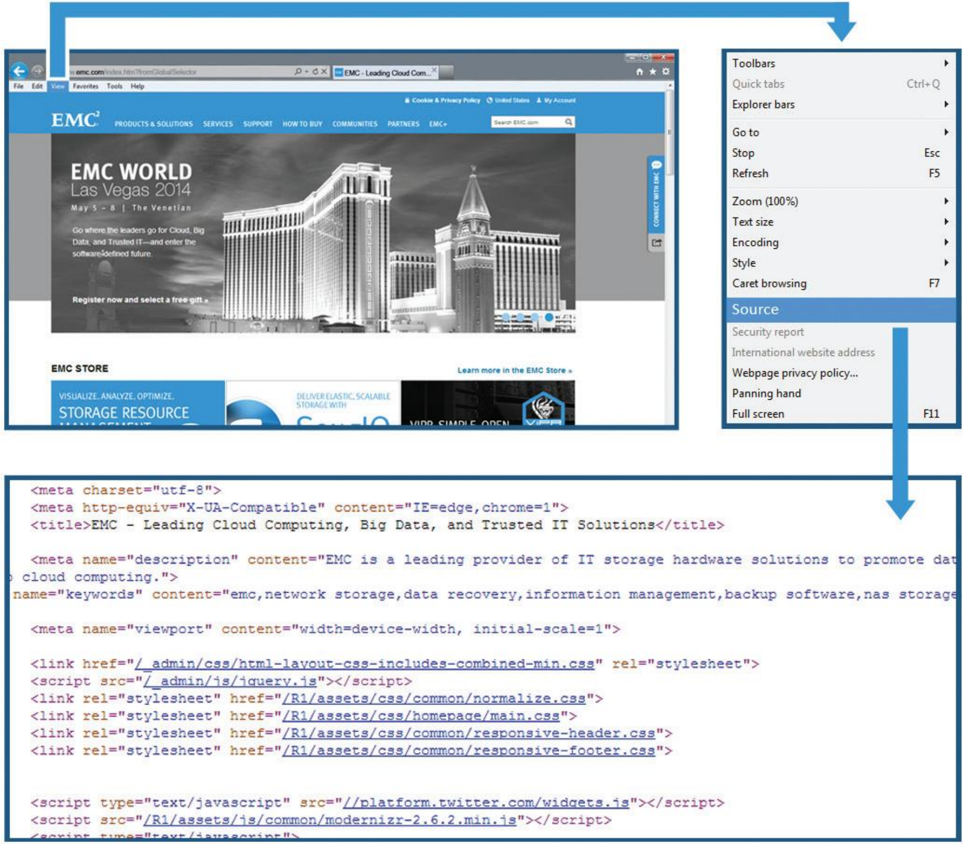
Data Structures

1. Structured Data. Structured data is organized and formatted in a way that it's easily searchable in databases, generally in rows and columns, facilitating straightforward analysis.

Marketing teams can harness structured data to segment their audience and orchestrate targeted campaigns based on various customer attributes like **age**, **location**, and **purchasing patterns**, thus enabling **data-driven decision-making and strategy planning**.

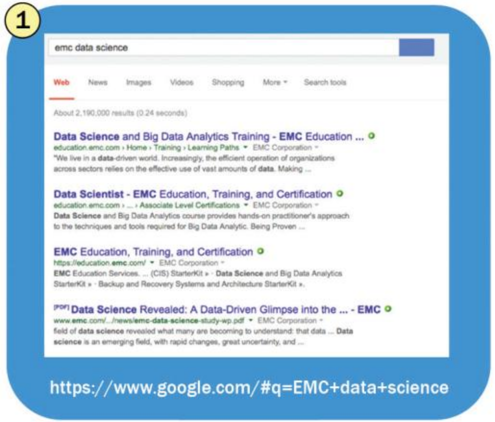
SUMMER FOOD SERVICE PROGRAM 1] (Data as of August 01, 2011)				
Fiscal Year	Number of Sites	Peak (July) Participation	Meals Served	Total Federal Expenditures 2]
	------------Thousands------------		--Mil.--	---Million $---
1969	1.2	99	2.2	0.3
1970	1.9	227	8.2	1.8
1971	3.2	569	29.0	8.2
1972	6.5	1,080	73.5	21.9
1973	11.2	1,437	65.4	26.6
1974	10.6	1,403	63.6	33.6
1975	12.0	1,785	84.3	50.3
1976	16.0	2,453	104.8	73.4
TQ 3]	22.4	3,455	198.0	88.9
1977	23.7	2,791	170.4	114.4
1978	22.4	2,333	120.3	100.3
1979	23.0	2,126	121.8	108.6
1980	21.6	1,922	108.2	110.1
1981	20.6	1,726	90.3	105.9
1982	14.4	1,397	68.2	87.1
1983	14.9	1,401	71.3	93.4
1984	15.1	1,422	73.8	96.2
1985	16.0	1,462	77.2	111.5
1986	16.1	1,509	77.1	114.7
1987	16.9	1,560	79.9	129.3
1988	17.2	1,577	80.3	133.3
1989	18.5	1,652	86.0	143.8
1990	19.2	1,692	91.2	163.3

Data Structures

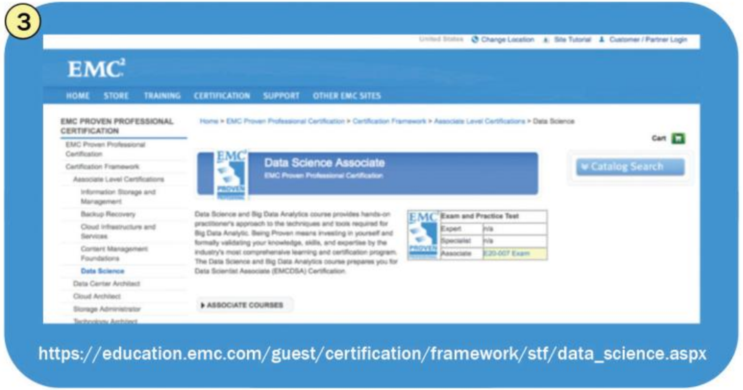


2. Semi-Structured Data. Semi-structured data, although not organized in rows and columns, has some elements of structure, such as tags and hierarchies, which facilitate its analysis to a certain extent.

Data Structures



3. Quasi-Structured Data. Quasi-structured data, a type of data that doesn't follow a fixed format or structure but exhibits some levels of organization or patterns that can be extracted and analyzed with specific tools.



Data Structures

4. Unstructured Data. Unstructured data is information that doesn't adhere to a specific form or structure, encompassing a variety of data types such as text, images, and videos, which require advanced tools for effective analysis.

Properties and scales of measurement

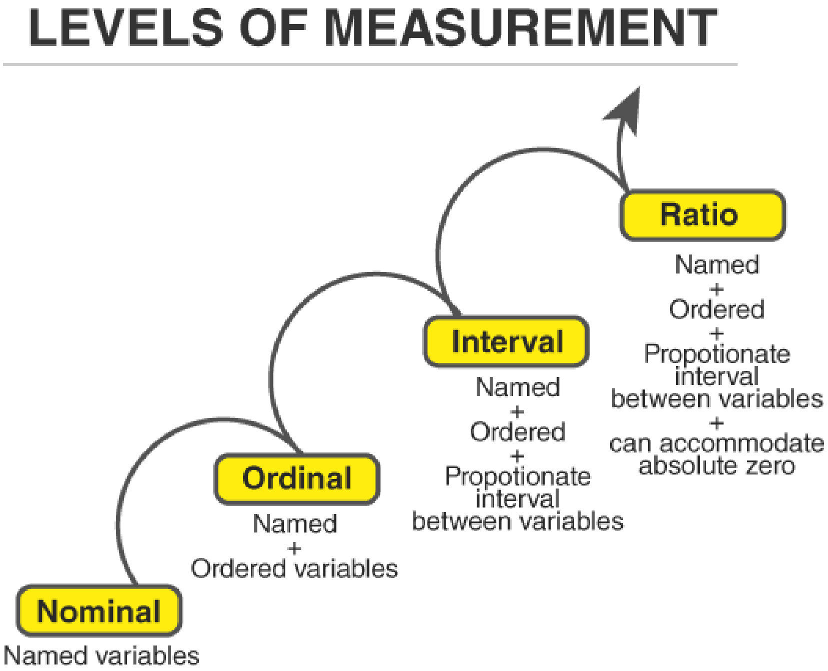
Scales of measurement is how variables are defined and categorised. Psychologist **Stanley Stevens** developed the four common scales of measurement: **nominal**, **ordinal**, **interval** and **ratio**. Each scale of measurement has properties that determine how to properly analyse the data. The properties evaluated are **identity**, **magnitude**, **equal intervals** and a **minimum value of zero**.

THE FOUR LEVELS OF MEASUREMENT:

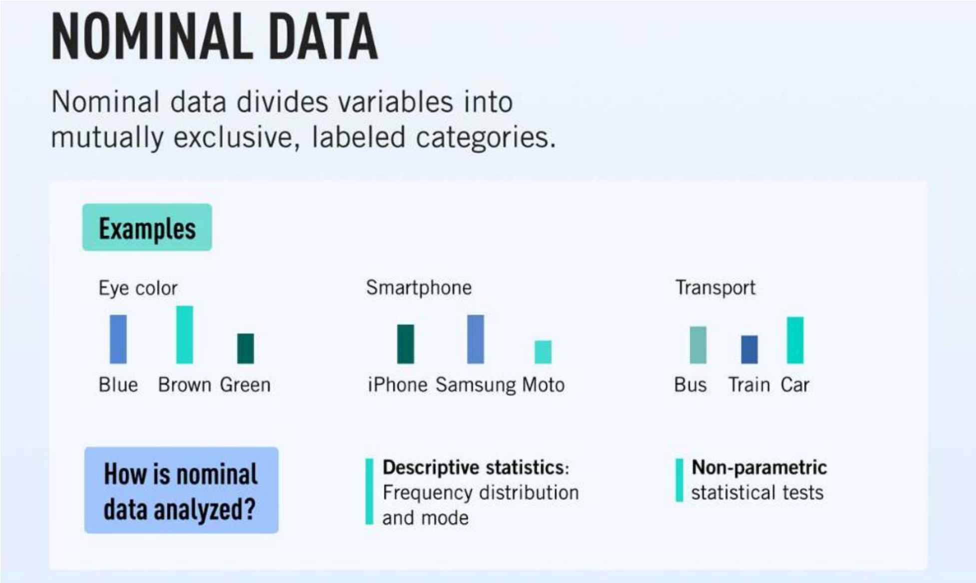
	Nominal	Ordinal	Interval	Ratio
Categorizes and labels variables	✓	✓	✓	✓
Ranks categories in order		✓	✓	✓
Has known, equal intervals			✓	✓
Has a true or meaningful zero				✓

Properties and scales of measurement

Levels of Measurements. There are four different scales of measurement. The data can be defined as being one of the four scales. The four types of scales are:

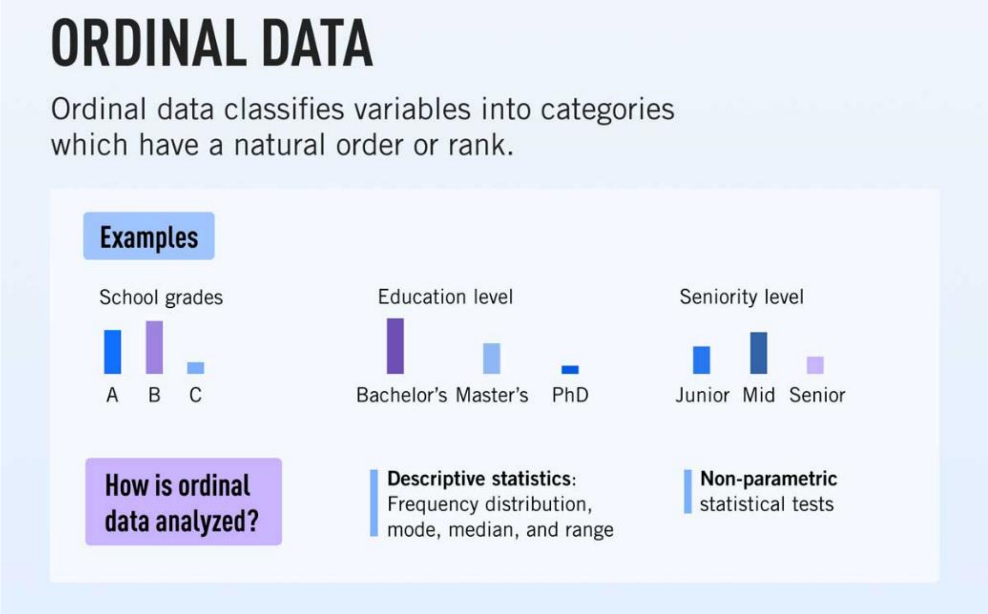


Properties and scales of measurement



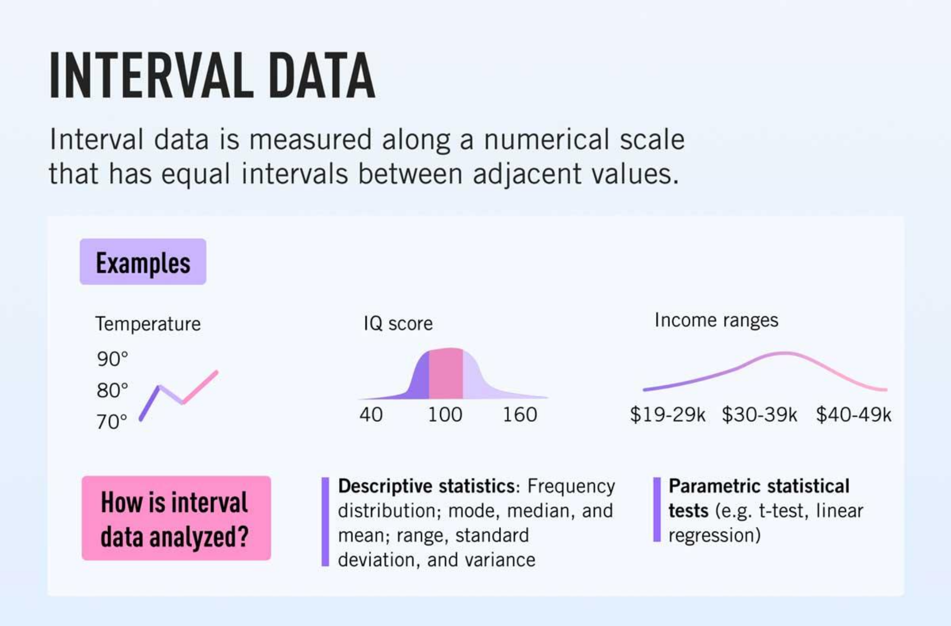
The nominal scale, also known as the categorical scale, is one of the simplest scales of measurement used in statistics and data analytics. A nominal scale is the 1st level of measurement scale in which the numbers serve as "**tags**" or "**labels**" to classify or identify the objects.

Properties and scales of measurement



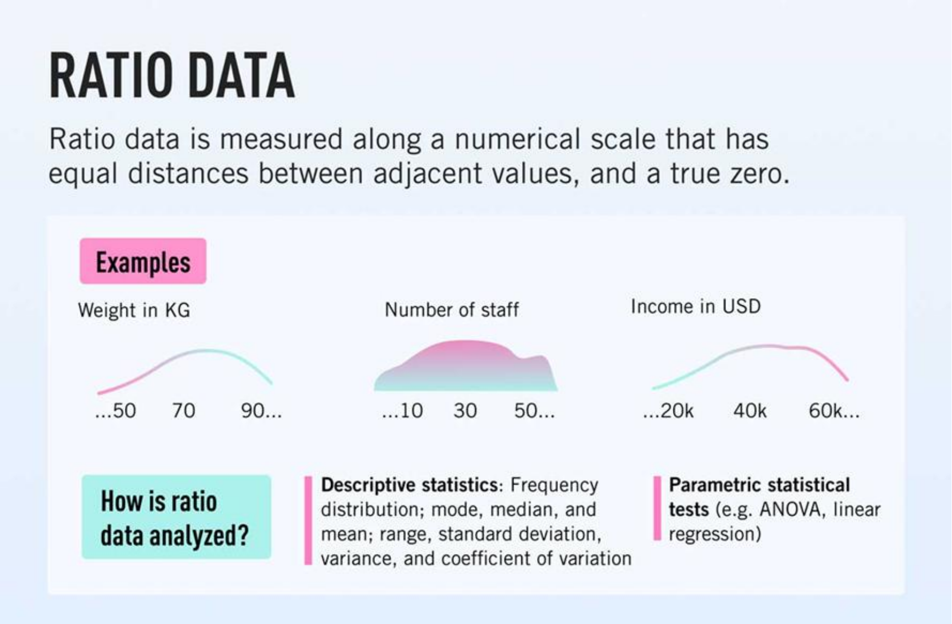
The ordinal scale of measurement is used to categorize data and indicate the relative order of the items being measured. Unlike nominal scales, ordinal scales provide clear directions regarding the hierarchy or order of categories.

Properties and scales of measurement



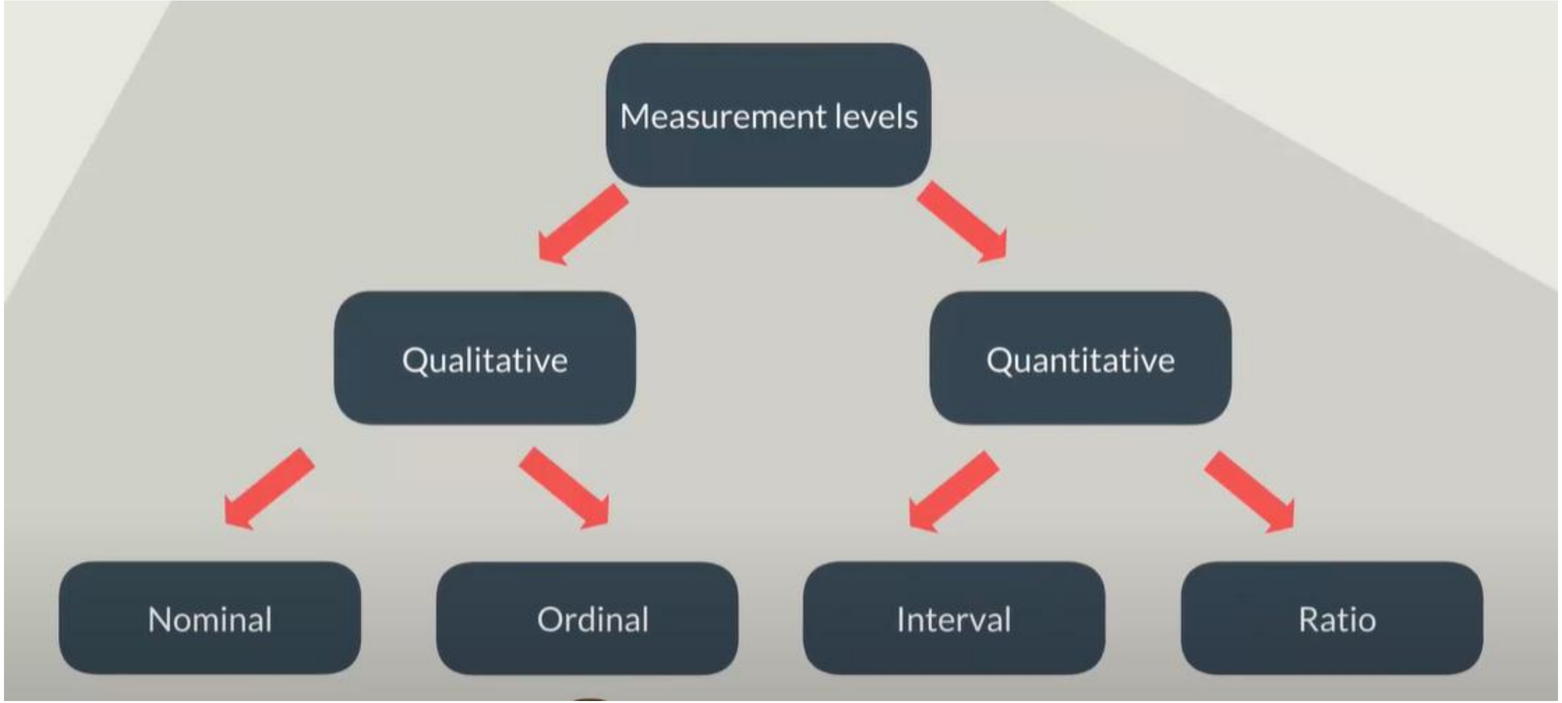
The interval scale of measurement represents a step further in complexity from the ordinal scale, incorporating a standardized scale of measurement that allows for the determination of the exact distances between scale points.

Properties and scales of measurement



The ratio scale of measurement is the highest level of measurement, which shares all the characteristics of the interval scale and also includes a true zero point. This allows for a wide range of statistical analyses to be conducted, including the calculation of ratios, which is not possible with interval data.

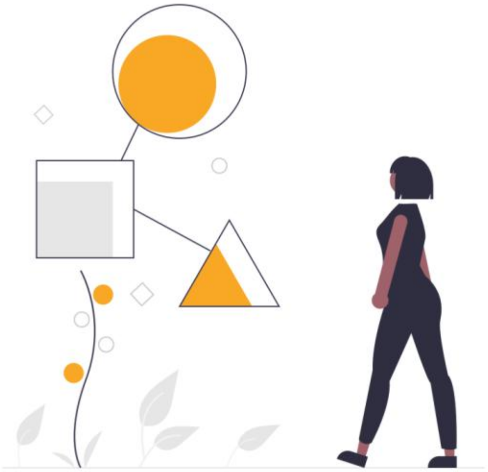
Data at the highest level: qualitative and quantitative



ANY QUESTIONS?

1. Introduction to Data Analytics

Part 2



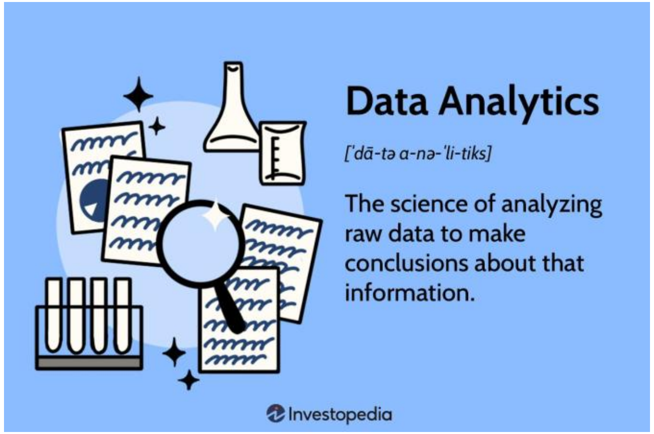
Khalil Israfilzada, PhD
Faculty of Economics and Management
Vytautas Magnus University
Kaunas, 2025

WHAT ARE DATA ANALYTICS IN BUSINESS?

Data analytics is a systematic approach to harnessing data to make informed business decisions, strategize, and identify new opportunities.

Process: Involves extracting meaningful insights from raw data through various processes and technologies.

Tools: Microsoft Excel, Power BI, Google Charts, Data Wrapper, Infogram, Tableau, Zoho Analytics, etc.



Who needs data analytics and why?

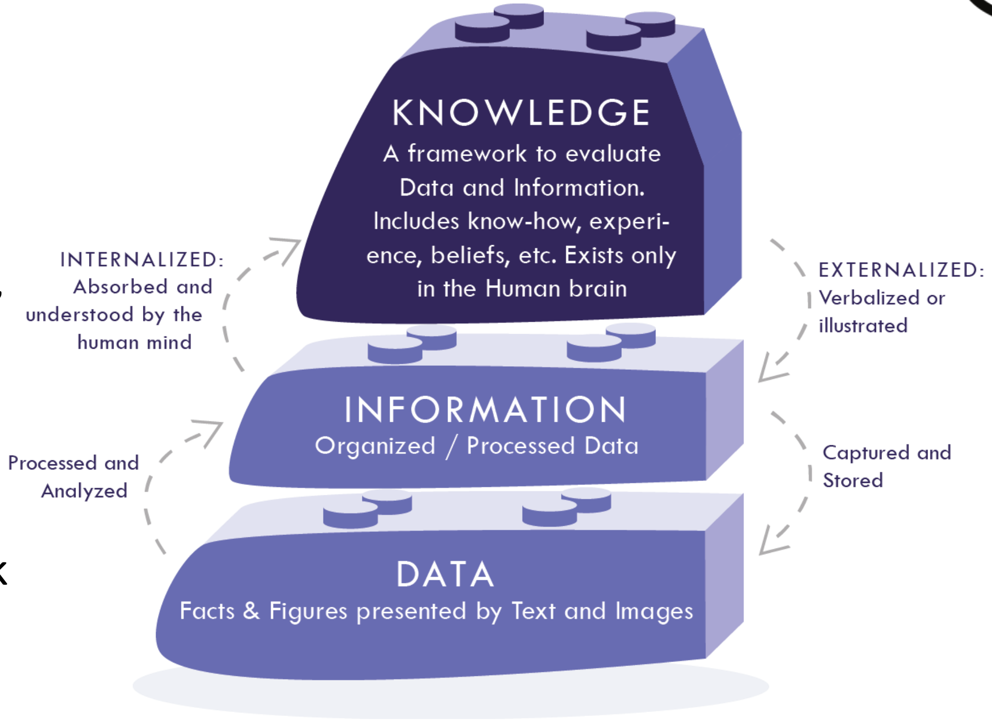
Necessity: Essential for professionals in the modern business environment.

Applications:

- **Marketers**: For crafting effective marketing strategies.
- **Product Managers**: To enhance product offerings based on market trends and user data.
- **Finance Professionals**: For accurate forecasts and financial planning.
- **HR Professionals**: To foster an inclusive and harmonious work environment.

Data Vs Information Vs Knowledge

- **Data** 📊
  Raw facts and figures without context.
  Example: “100, 200, 300”
- **Information** 📑
  Processed data that has meaning.
  Example: “Sales this week were 100, 200, and 300 units.”
- **Knowledge** 💡
  Insights gained from interpreting information to make decisions.
  Example: “Sales are increasing each week → our new marketing campaign is working.”



Data Quality: Why "Garbage In = Garbage Out"

One of the most important principles in data analytics is **"Garbage In = Garbage Out (GIGO)."** This phrase means that if the data we put into a system is inaccurate, incomplete, or misleading, then the results of the analysis will also be flawed, no matter how advanced the tools or models used.

Data Quality: Why “Garbage In = Garbage Out”

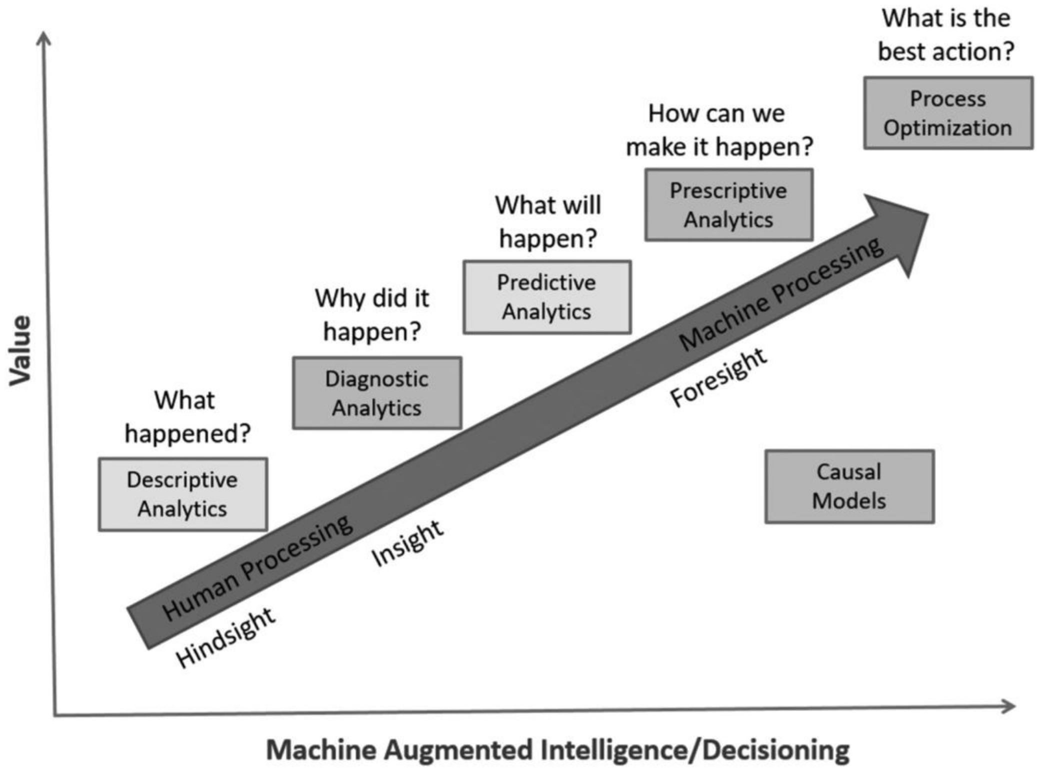
Key dimensions of data quality include:

- **Accuracy** – Are the values correct?
- **Completeness** – Is any important data missing?
- **Consistency** – Do values match across different sources?
- **Timeliness** – Is the data up-to-date?

- **Example (Business).** If a retail company collects incorrect sales figures (e.g., duplicate entries or missing transactions), the analysis might show false profit trends. Decisions made based on this faulty analysis—like increasing stock of the wrong product—could lead to financial losses.

Breaking it Down, Categorizing Analytics

This is very difficult as there are methodological types of analytics, functional types of analytics and analytics that address certain types of data.

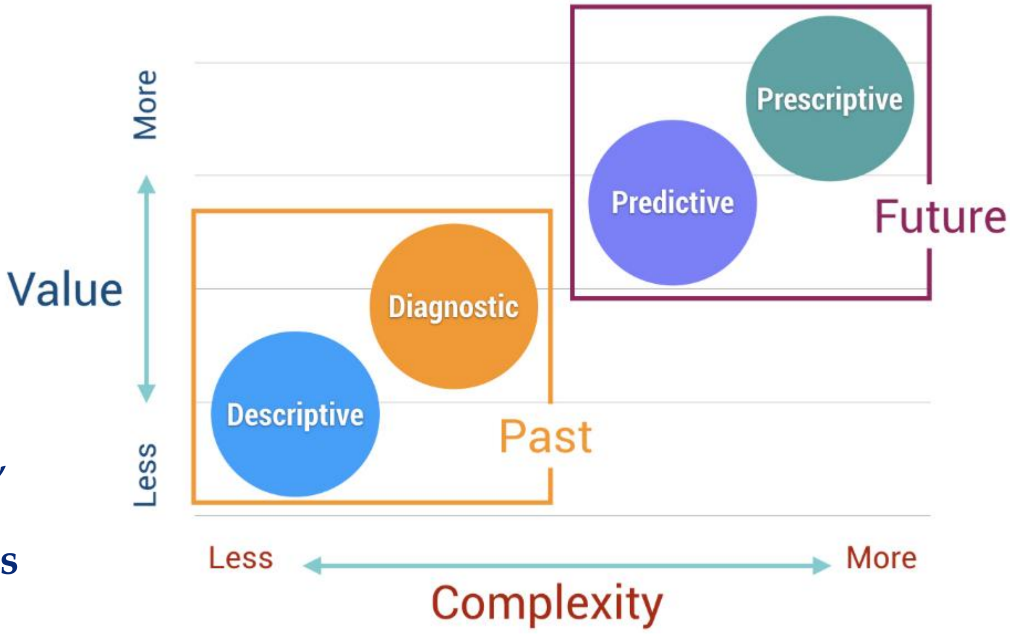


4 KEY TYPES OF DATA ANALYTICS

Breaking it Down, Categorizing Analytics

- **Descriptive Analytics**: Analyzes past events to explain "What happened?“
- **Diagnostic Analytics**: Investigates the reasons behind past events, answering "Why did this happen?“
- **Predictive Analytics**: Forecasts potential future outcomes, answering "What might happen in the future?“
- **Prescriptive Analytics**: Suggests actions to achieve desired results, answering "What should we do next?"

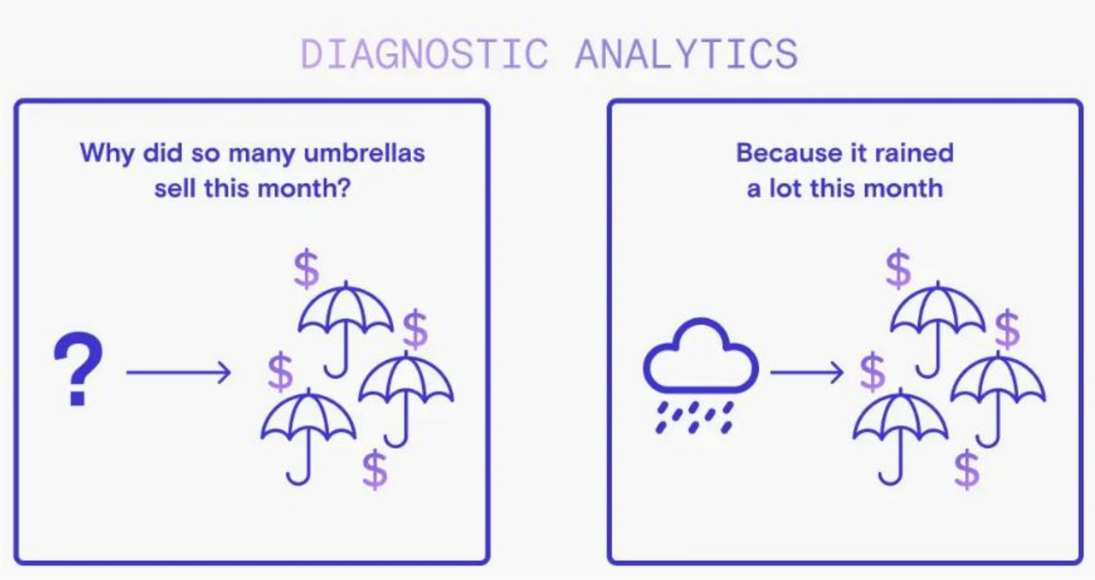
Breaking it Down, Categorizing Analytics



The first two types concern themselves with analyzing **past events**. In contrast, the **latter two types are forward-looking**, focusing on **future predictions and prescriptions**

WHAT IS DIAGNOSTIC ANALYTICS?

Diagnostic analytics is the process of using data to determine the causes of trends and correlations between variables. It can be viewed as a logical next step after using descriptive analytics to identify trends.



WHAT IS DIAGNOSTIC ANALYTICS?

There several concepts to understand before diving into diagnostic analytics:

- ❑ Hypothesis Testing.
- ❑ Correlation vs. Causation.
- ❑ Diagnostic Regression Analysis.
- ❑ Data mining.
- ❑ Data drilling.

What are the advantages of diagnostic analysis?

- ❑ Offers insights into tiny data aspects
- ❑ Compares input and output data for correlation or cause-and-effect relationships
- ❑ Offers concrete evidence for testing hypotheses
- ❑ Identifies anomalies and outliers to establish significance and accuracy
- ❑ Analyses prior occurrences to prevent mistakes and duplicate beneficial results
- ❑ Fosters data-driven decision-making culture and improves data collection, analysis, and diagnosis across the enterprise.

What are the disadvantages of diagnostics analysis?

- limits the ability to predict future events
- requires additional sources to supplement your analysis, including third-party historical and real-time data
- takes more time and requires higher-level skills than descriptive analytics, but there are new platforms that mitigate this challenge
- leads analysts to mistake correlation for causation, which can easily create costly consequences

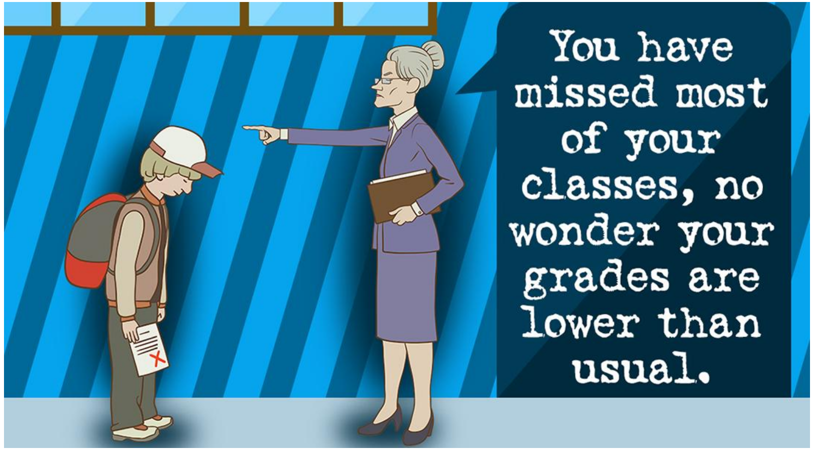
What is the process of performing diagnostic data analytics?

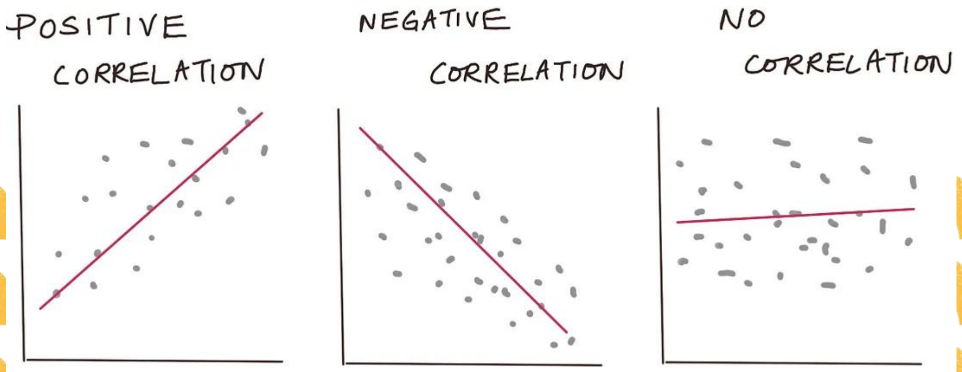
CAUSATION VS CORRELATION

If you've ever argued with someone who has taken Statistics 101, you might have heard them say with pride, ***"correlation does not imply causation."*** This mantra is repeatedly applied when people erroneously assume that two variables bear a cause-and-effect relationship rather than merely displaying a similar pattern of occurrences. Although the rooster's crow happens every morning as the sun rises, it does not cause the sun to rise.

CAUSATION VS CORRELATION

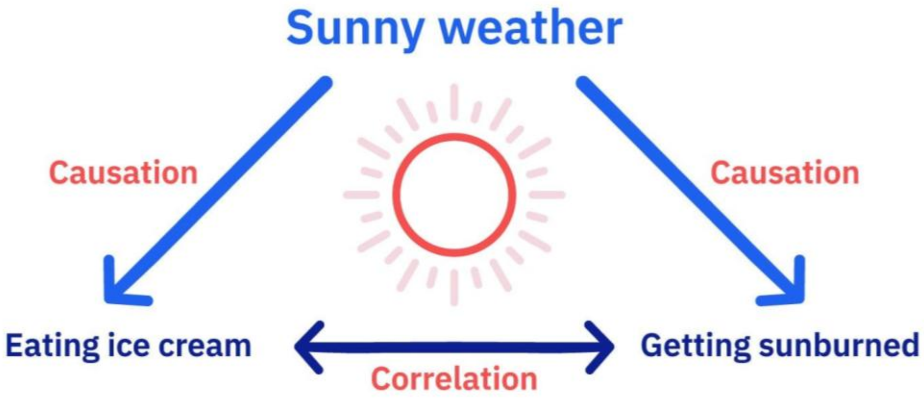
When two variables tend to run together, we say they are associated, they co-occur, they are **correlated**. When one category of a variable is present, a second variable category is often present. Or, in the case of numerical attributes, when one variable increases the second variable increases (*positive relationship*) or decreases (*negative relationship*).





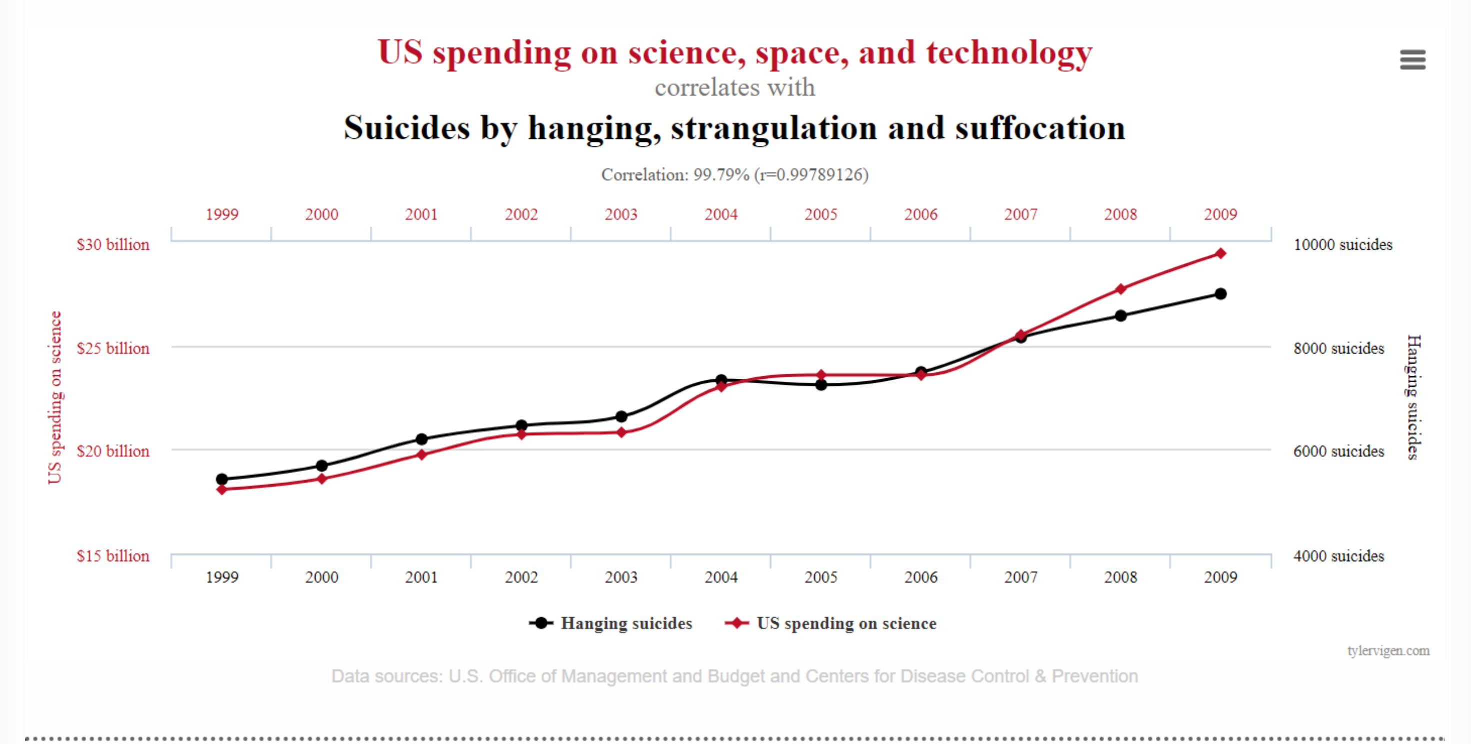
CAUSATION VS CORRELATION

While causation and correlation can exist simultaneously, correlation does not imply causation. **Causation means one thing causes another—in other words, action A *causes* outcome B**. On the other hand, **correlation is simply a relationship where action A *relates to* action B**—but one event doesn't necessarily cause the other event to happen.



CAUSATION VS CORRELATION

If you want some fun, we recommend that you check out Tyler Vigen's website at **http://tylervigen.com/spurious-correlations**. You will find a lot of interesting spurious correlations! We recreated a sample of this in the spirit of Tyler's website.



CAUSATION VS CORRELATION

In big data analytics for marketing, it's essential to:

1. **Understand Customer Behavior**
2. **Avoid Misinterpretations**
3. **Predictive Analytics**
4. **Resource Allocation**
5. **Experiment Design**
6. **Data-Driven Decision Making**

Examples of diagnostic analytic data

- Market Fluctuation Analysis
- Customer Churn Analysis
- Supply Chain Disruptions
- Product Performance Analysis
- Health care
- Human resources
- Manufacturing

What is Prescriptive Analytics?

- **Prescriptive analytics** has been called **“the future of data analytics,”** and for good reason. This type of analysis goes beyond explanations and predictions to recommend the best course of action moving forward. It’s especially useful in driving data-informed decision-making.

- Prescriptive analytics is the process of using data to determine an optimal course of action. By considering all relevant factors, this type of analysis yields recommendations for next steps.

What is Prescriptive Analytics?

- Machine-learning algorithms are often used in prescriptive analytics to parse through large amounts of data faster—and often more efficiently—**than humans can**. Using “**if**” and “**else**” statements, algorithms comb through data and make recommendations based on a specific combination of requirements. For instance, if at least 50 percent of customers in a dataset selected that they were “very unsatisfied” with your customer service team, the algorithm may recommend additional training.

What is Prescriptive Analytics?

- It's important to note: While algorithms can provide data-informed recommendations, they can't replace human discernment. Prescriptive analytics is a tool to inform decisions and strategies and should be treated as such. Your judgment is valuable and necessary to provide context and guard rails to algorithmic outputs.

Benefits of using prescriptive algorithms

Company executives are constantly looking for ways to optimize business efficiency, and these data give them the ability to do so. Here are several of the benefits:

- ❑ **Providing a roadmap for success.**
- ❑ **Reducing human error or bias.**
- ❑ **Freeing up valuable time for other tasks**
- ❑ **Decreasing exposure to risk**

Prescriptive Analytics in Marketing

Marketers can use prescriptive analytics to stay ahead of consumer trends. Using past trends and past performance can give internal and external marketing departments a competitive edge.

By employing prescriptive analytics, marketers can come up with effective campaigns that target specific customers at specific times like, say, advertising for a certain demographic during the Superbowl. Corporations can also identify how to engage different customers and how to effectively price and discount their products and services.

Real-world Implementation of Prescriptive Analytics

1. **UPS**: The logistics giant uses prescriptive analytics to optimize its delivery routes and reduce fuel consumption. Using sophisticated algorithms, UPS determines the most efficient ways for its delivery trucks, considering factors such as traffic patterns, weather conditions, and package weight.

2. **Hilton Worldwide**: Hilton uses prescriptive analytics to optimize its hotel room pricing strategy. By analyzing historical data on room occupancy rates, customer preferences, and competitive pricing, Hilton can adjust room rates in real-time to maximize revenue and occupancy.

Real-world Implementation of Prescriptive Analytics

1. **The Weather Company**: The Weather Company uses prescriptive analytics to predict the weather and help businesses prepare for severe weather events. Using data from sensors, satellites, and other sources, the company can provide accurate weather forecasts and recommend actions to minimize severe weather's impact.

2. **Procter & Gamble:** P&G uses prescriptive analytics to optimize its supply chain operations. By analyzing demand, inventory levels, and production capacity data, P&G can make real-time decisions on allocating inventory and production resources to maximize efficiency and reduce costs.

Tools Used for Prescriptive Analytics

Tool Type	Examples
Optimization tools	IBM ILOG CPLEX, Gurobi, FICO Xpress
Simulation tools	AnyLogic, Simul8, Arena
Machine learning algorithms	Neural networks, decision trees, regression analysis
Business intelligence tools	Tableau, QlikView, Power BI
Natural language processing (NLP) tools	Google Cloud Natural Language API, Microsoft Azure Cognitive Services, IBM Watson NLU
Prescriptive analytics software	FICO Decision Management Suite, IBM Decision Optimization, SAP Analytics Cloud

ANY QUESTIONS?

2. Descriptive Analytics

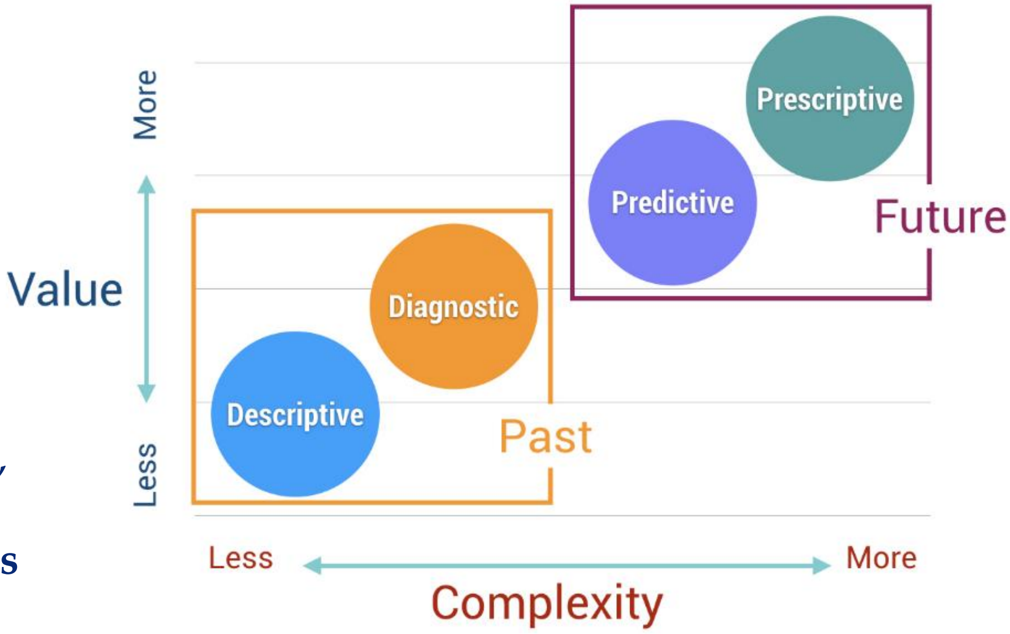
Khalil Israfilzada, PhD
Faculty of Economics and Management
Vytautas Magnus University
Kaunas, 2025

WHAT IS DESCRIPTIVE ANALYTICS?

Descriptive analytics is the process of using current and historical data to identify trends and relationships. It's sometimes called the simplest form of data analysis because it describes trends and relationships but doesn't dig deeper.

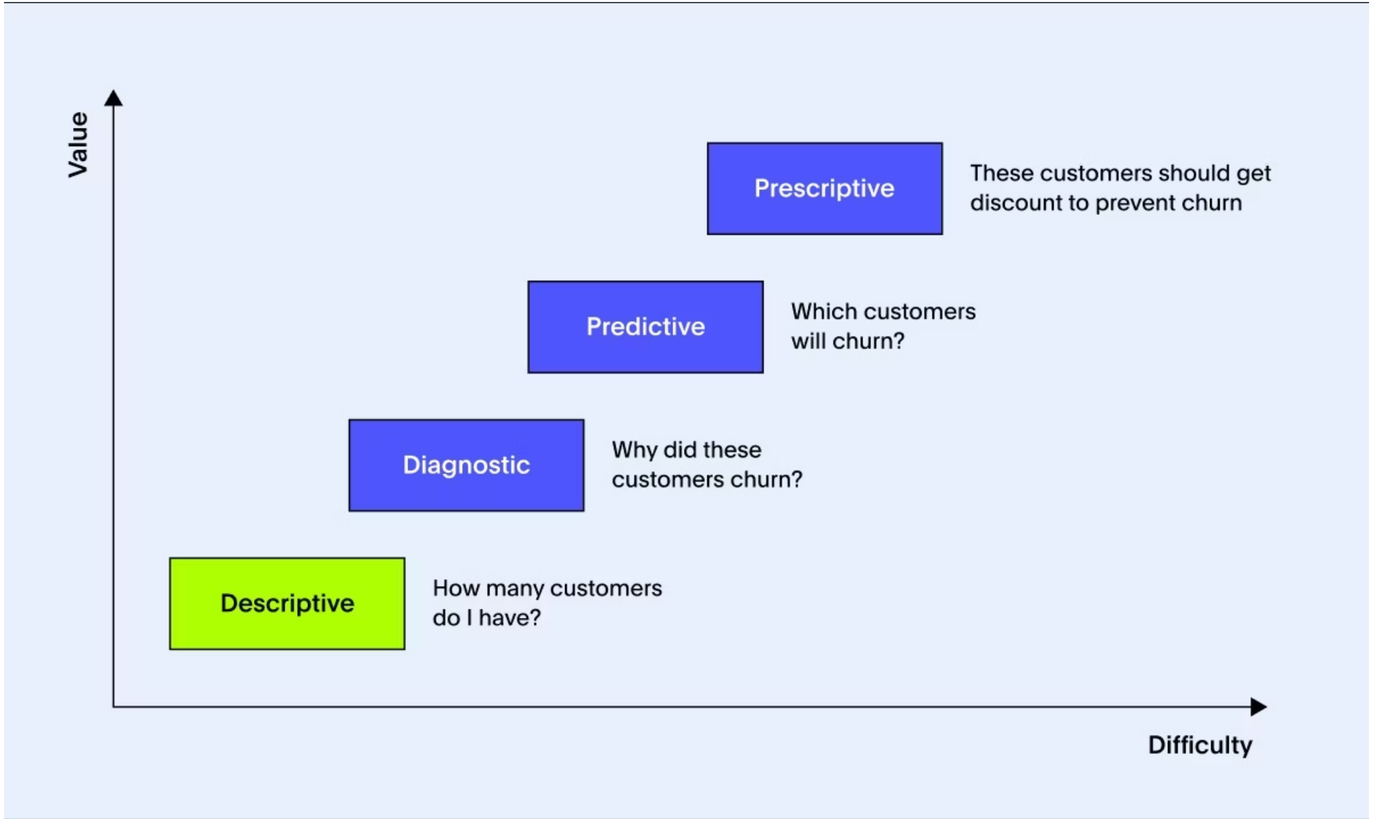
1. How much revenue did we generate last quarter?
2. Are the revenue numbers growing or falling?
3. How many new customers did we acquire?
4. How many existing customers did we lose? Etc.

Breaking it Down, Categorizing Analytics



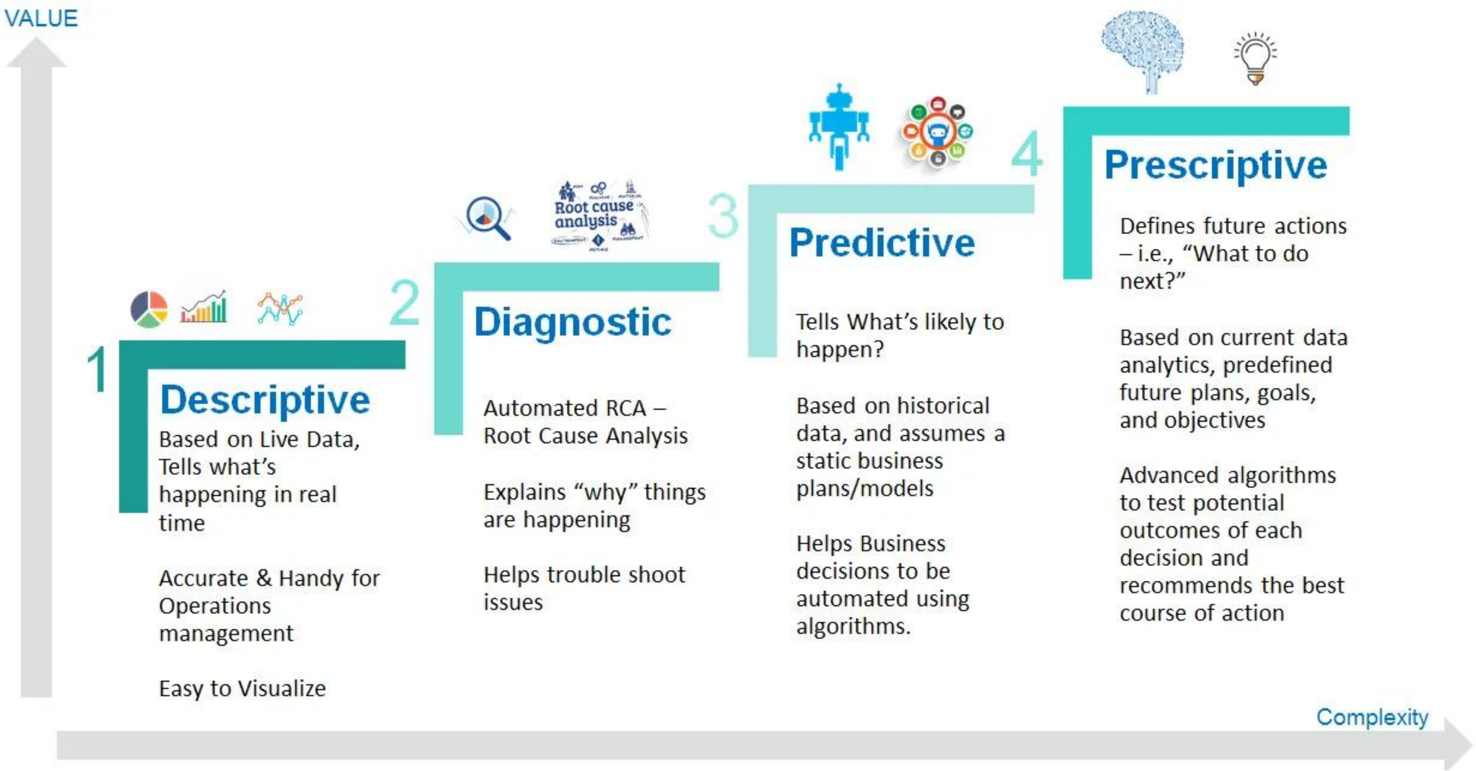
The first two types concern themselves with analyzing **past events**. In contrast, the **latter two types are forward-looking**, focusing on **future predictions and prescriptions**

DESCRIPTIVE ANALYTICS?



DESCRIPTIVE ANALYTICS?

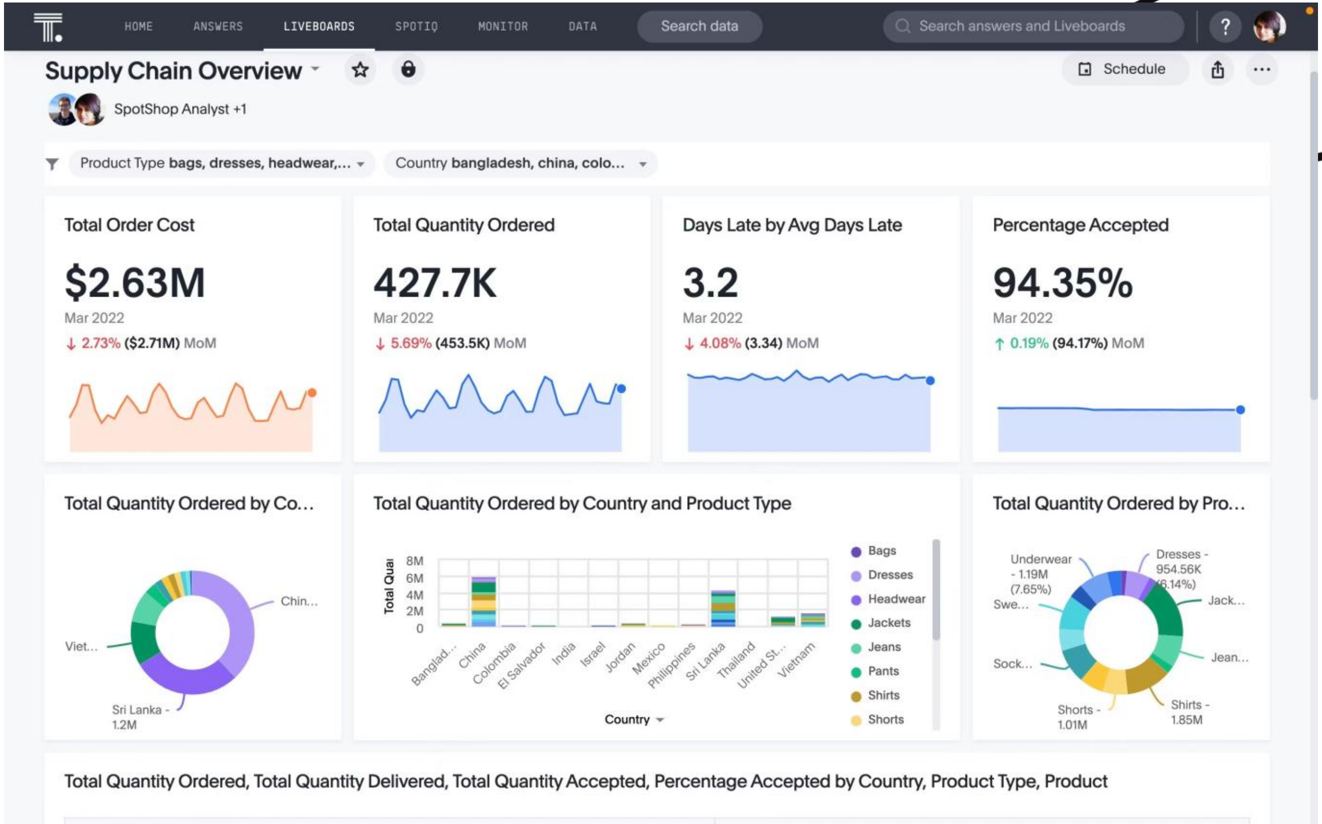
Usually, descriptive analytics is the foundation upon which all other analytic branches are built. **Descriptive analytics has a very long historical foundation**.



Descriptive analytics examples

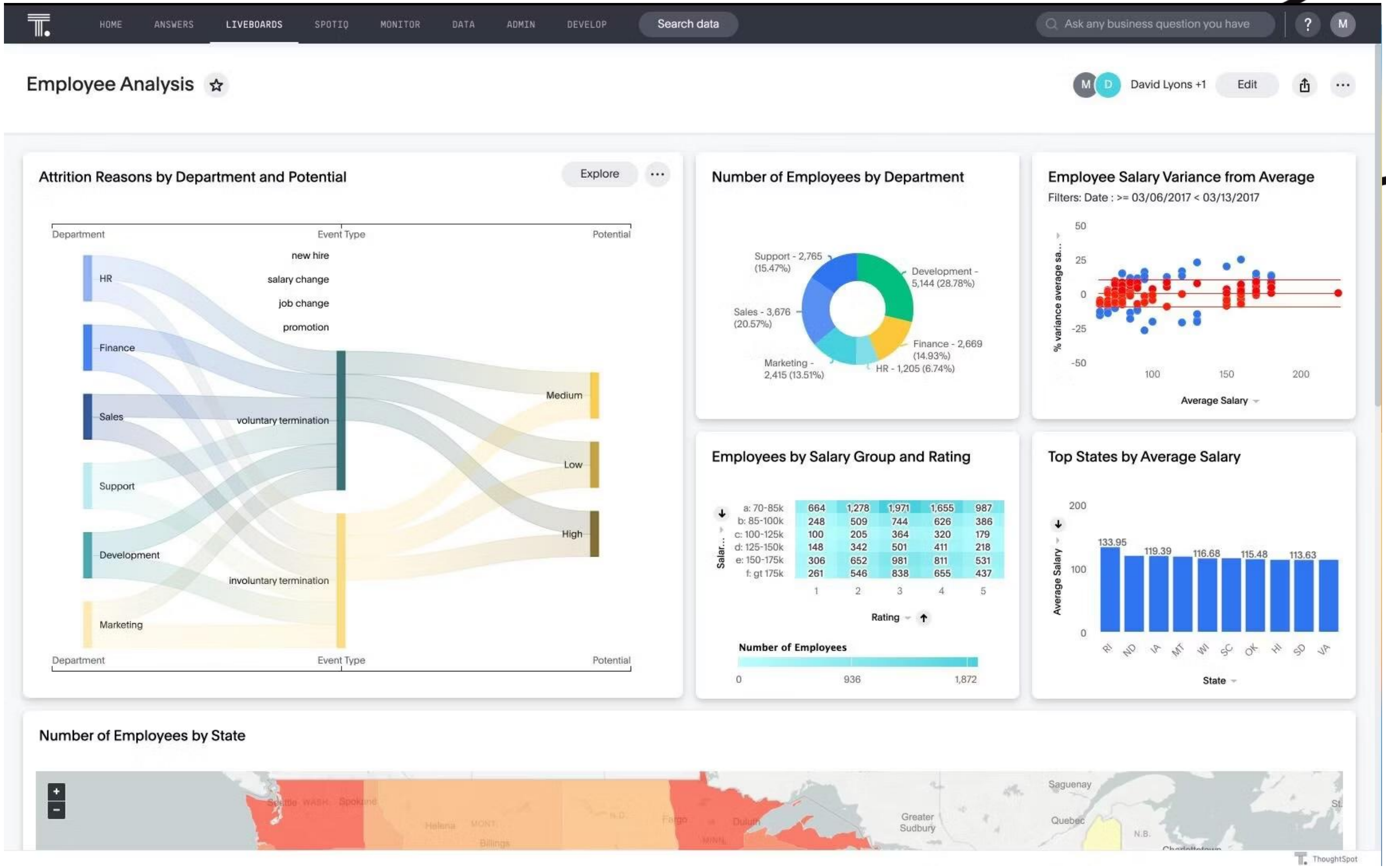
Although digital marketing is well known for providing advertisers with a huge volume of reporting data, it can often be difficult to determine (and easily access) the metrics that are most important to you. A digital marketing dashboard provides a single view of the KPIs Digital Marketing Managers need to see, in order to understand their digital marketing campaigns, and take action to optimize performance.

Descriptive analytics examples



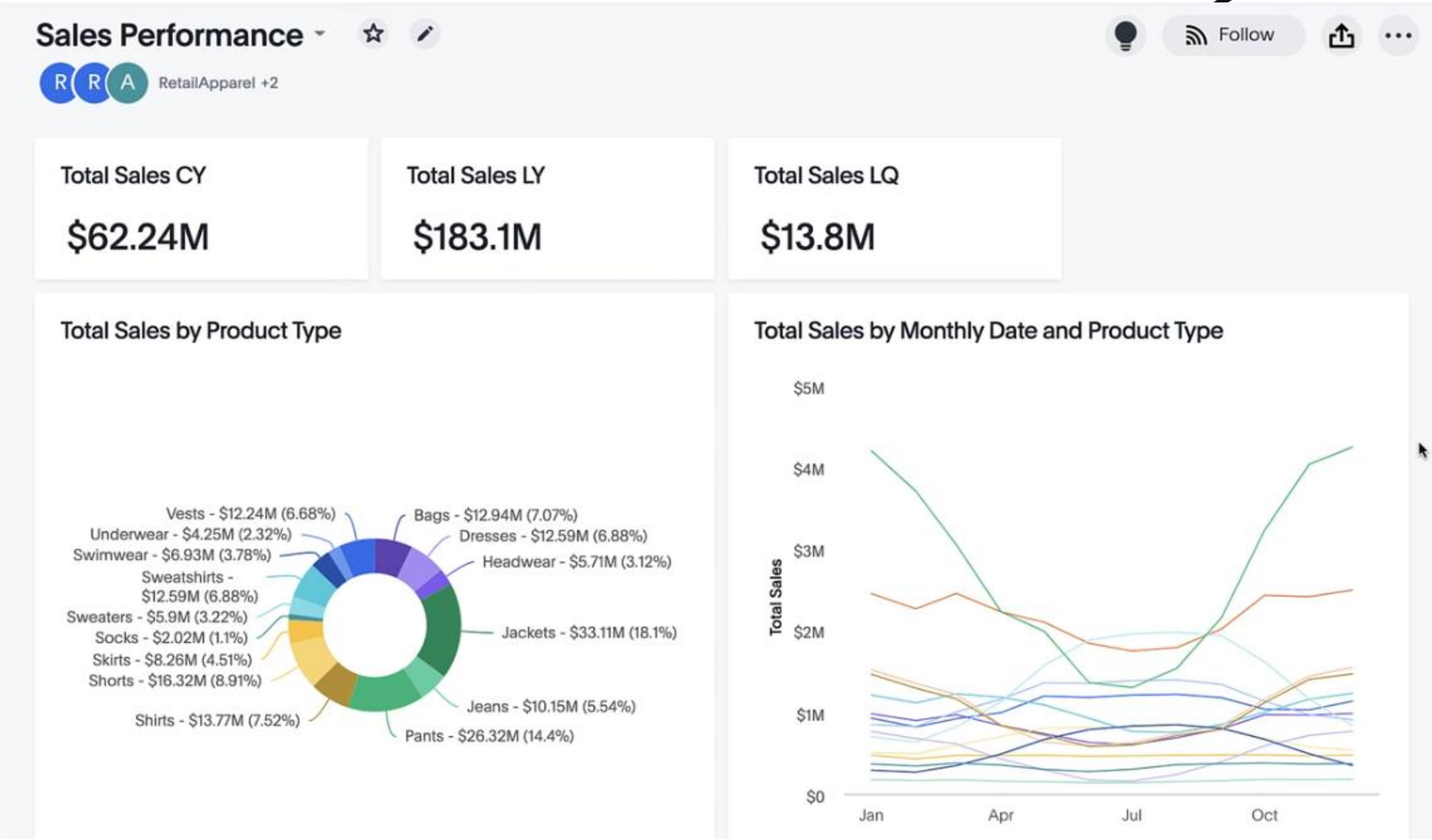
Companies lose millions due to poor inventory planning. In fact, research by Oracle shows that holding or carrying costs related to inventory planning can add up to 20–30% of your overall business cost. This creates an immense need for operational leaders to drill into data, keep track of what products or services are in stock, and see where they have shortages or overstock.

Descriptive analytics examples



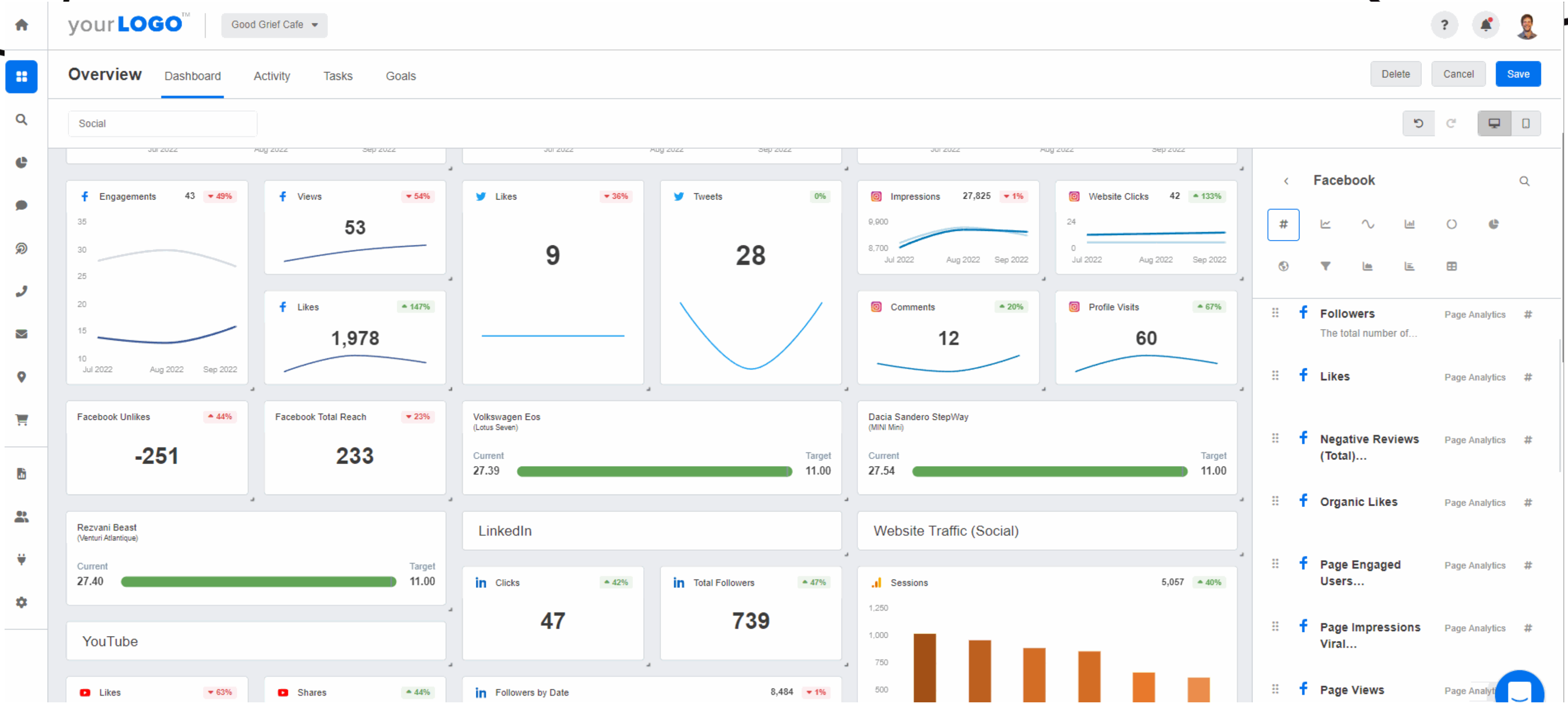
With descriptive analytics, you can dig deep into [HR metrics]() and past workforce data to understand what's affecting employee performance. Depending on the insights you find, you can create targeted retention strategies, identify departments with the highest attrition rate, or take measures to improve overall employee satisfaction.

Descriptive analytics examples



When used strategically, descriptive analytics helps sales managers gain visibility into their pipeline. You can unlock valuable trends in your sales data, including revenue, conversion rates, and sales growth over time.

Descriptive analytics examples



Importance of Descriptive Analytics

- Helps evaluate performance, understand customers, and guide decisions
- Three key roles:
  1. Understanding past performance
  2. Identifying customer patterns
  3. Supporting decisions

Importance of Descriptive Analytics

➢ Understanding Past Performance in Marketing

- Provides a clear view of past outcomes
- Evaluates ROI & KPIs in campaigns
- Key uses:
  - **Channel Performance** → Which platforms work best
  - **Campaign Analysis** → CTR, engagement, conversion
  - **Website Insights** → Page views, bounce rates, user flow

Importance of Descriptive Analytics

➢ Understanding Past Performance in Sales & Commerce

- Converts transactions into meaningful reports
- Key uses:
  - **Financial Reporting** → Revenue, margins, AOV
  - **Product Performance** → Top-sellers vs. weak products
  - **Regional & Demographic Trends** → Growth areas & market penetration

Importance of Descriptive Analytics

➢ Identifying Customer Patterns & Trends

- Transforms raw data into meaningful insights
- Key uses:
  - **Seasonality** → Demand spikes (e.g., fitness in January)
  - **Associative Patterns** → Products bought together (cross-sell)
  - **Behavioural Segmentation** → High-value, discount seekers, dormant users
- Enables **anticipation, personalization, and optimization**

Importance of Descriptive Analytics

➢ Supporting Strategic and Operational Decisions

❑ Supporting Strategic Decisions

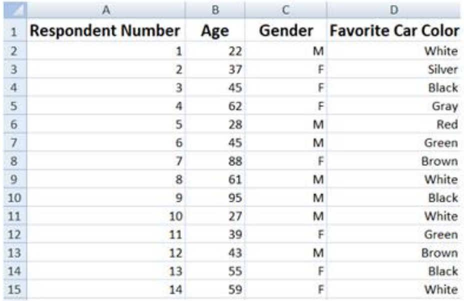
- **Long-term, high-impact choices**
- Example: Expansion into new markets → Use data from traffic, inquiries, early sales
- Descriptive analytics reduces risks by showing real demand

❑ Supporting Operational Decisions

- **Day-to-day, short-term choices**
- Example: Retail store forecasting holiday traffic → Staffing, inventory, cashier allocation
- Moves companies from **data-rich** → **data-driven**

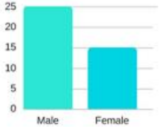
Descriptive Statistics and Inferential Statistics

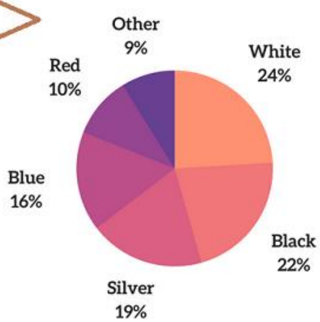
Descriptive Statistics is a branch of statistics concerned with describing characteristics of the population under study. Thus, descriptive statistics supports and is part of analytics. These characteristics / descriptive statistics are retrospective in nature. Most people are familiar with averages, minimums, maximums, etc., which people refer to as "statistics."



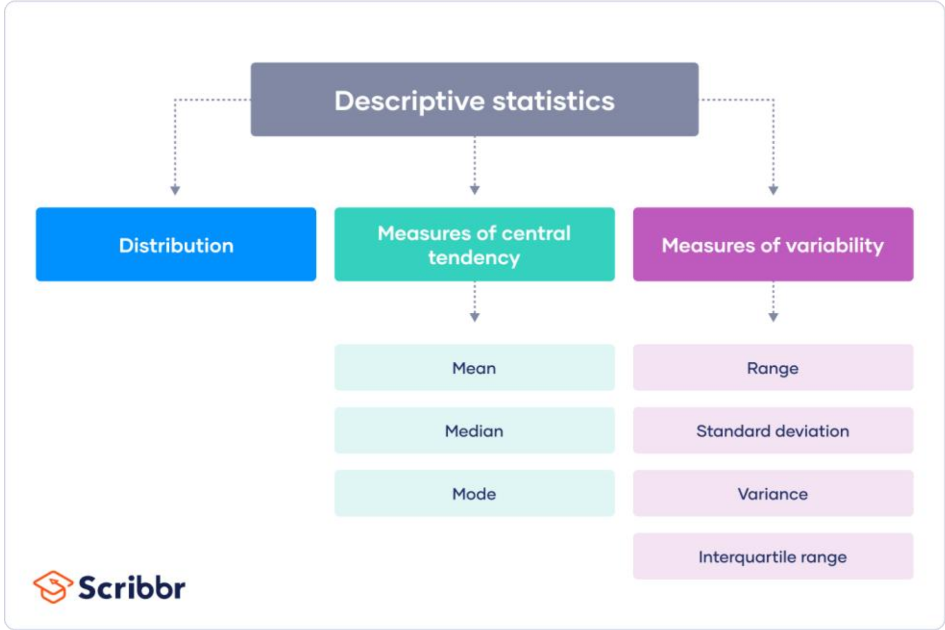
	A	B	C	D
1	Respondent Number	Age	Gender	Favorite Car Color
2	1	22	M	White
3	2	37	F	Silver
4	3	45	F	Black
5	4	62	F	Gray
6	5	28	M	Red
7	6	45	M	Green
8	7	88	F	Brown
9	8	61	M	White
10	9	95	M	Black
11	10	27	M	White
12	11	39	F	Green
13	12	43	M	Brown
14	13	55	F	Black
15	14	59	F	White

RAW DATA



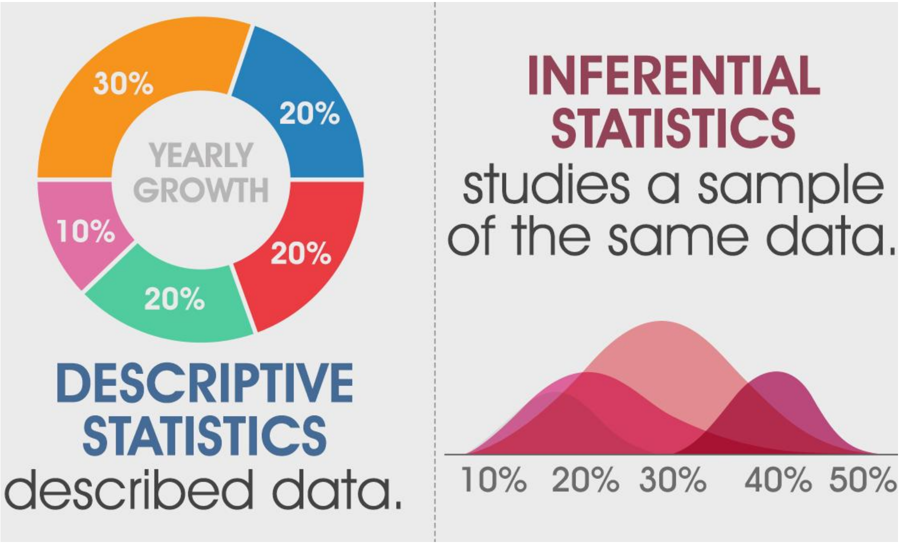


Descriptive Statistics



Descriptive Statistics and Inferential Statistics

Inferential Statistics is a type of statistics that focuses on drawing conclusions about the population, based on sample analysis. This conclusion about the population may extend beyond the data currently available, meaning processes drive (future) data such that the entire population is not currently available to sample. *Inferential statistics typically requires assumptions*.



What techniques are used in descriptive analytics?

1. Data Aggregation

1. **Summation**: The simplest form of aggregation is summation, where you're adding up specific metrics, such as the total number of sales, total revenue, or total customers.
2. **Averages**: Aggregated data can be averaged to provide a more balanced view of a particular metric over a certain time period.
3. **Grouping or Segmentation**: Data can be grouped by various categories like geographical locations, age groups, time periods, or product types to provide segmented views. This is often done using SQL queries or pivot tables in spreadsheet programs.
4. **Time-series Analysis**: Aggregating data over time intervals (hour, day, month, etc.) can help in identifying trends and seasonal variations.

What techniques are used in descriptive analytics?

2. Descriptive Statistics

1. **Measures of Central Tendency**: Includes the mean, median, and mode. These provide a central point for the data distribution.
2. **Measures of Dispersion**: Range, variance, and standard deviation help in understanding how spread out the data is.
3. **Percentiles and Quartiles**: These offer more granular insights into data distribution.
4. **Correlation Coefficients**: Though not always classified under descriptive analytics, understanding the relationship between two variables can be helpful.
5. **Frequency and Relative Frequency**: Particularly useful for categorical data.
6. **Cross-tabulation**: Often used for understanding the relationship between two or more categorical variables.

What techniques are used in descriptive analytics?

3. Data Visualization

Importance of Process

Understanding the basics of descriptive analytics seems simple enough, but applying it in real life can be challenging. There are several steps that an organization needs to follow to apply descriptive analytics to their business.

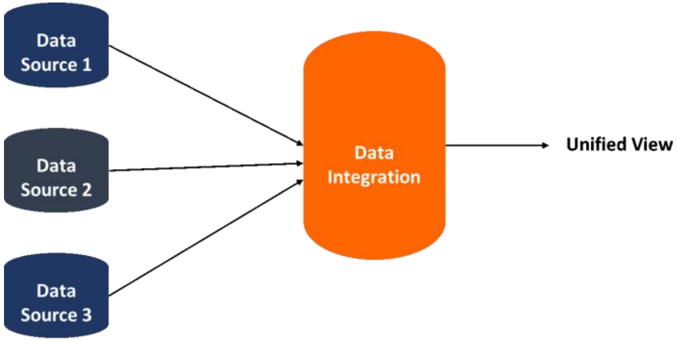
Data Collection

1. **Reliability**
2. **Timeliness**
3. **Scope**

Importance of Process

Data Cleaning

1. Identify Missing Values
2. Noise Reduction
3. Data Transformation



Data Integration

1. Normalization
2. De-duplication
3. Enrichment

Importance of Process

KPIs	Metrics
All KPIs are Metrics	All Metrics are not KPIs
KPIs give a holistic view of the performance of different functions in your organization	Metrics give you a picture of how different individual activities rolled out within the functions are progressing
KPIs tell you where exactly your teams stand with respect to the overall business goals	Individual Metrics do not give any insights on their own
Examples: Pre-sales KPIs, Email Marketing KPIs, Customer Success KPIs	Examples: Open Rate, Conversations in the last 2 weeks, Deals lost last quarter

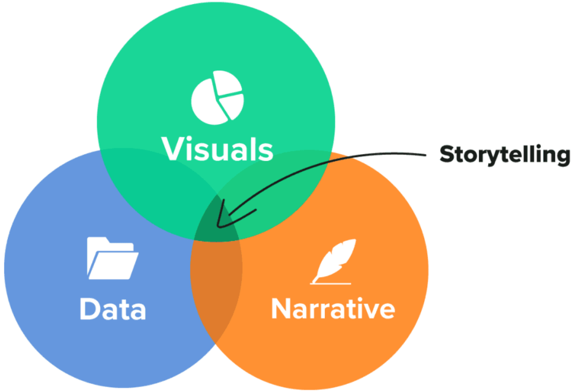
Analysis

1. Choosing Metrics
2. Method Selection
3. Iterative Approach

Importance of Process

Interpretation

1. Contextual Understanding
2. Critical Evaluation



Communication

1. Simplification
2. Visualization
3. Storytelling

Advantages of Descriptive Analytics

1. Simplicity and Ease of Interpretation
2. Basis for Further Analysis
3. Cost-effective
4. Real-time Insights
5. Broad Applications

Disadvantages of Descriptive Analytics

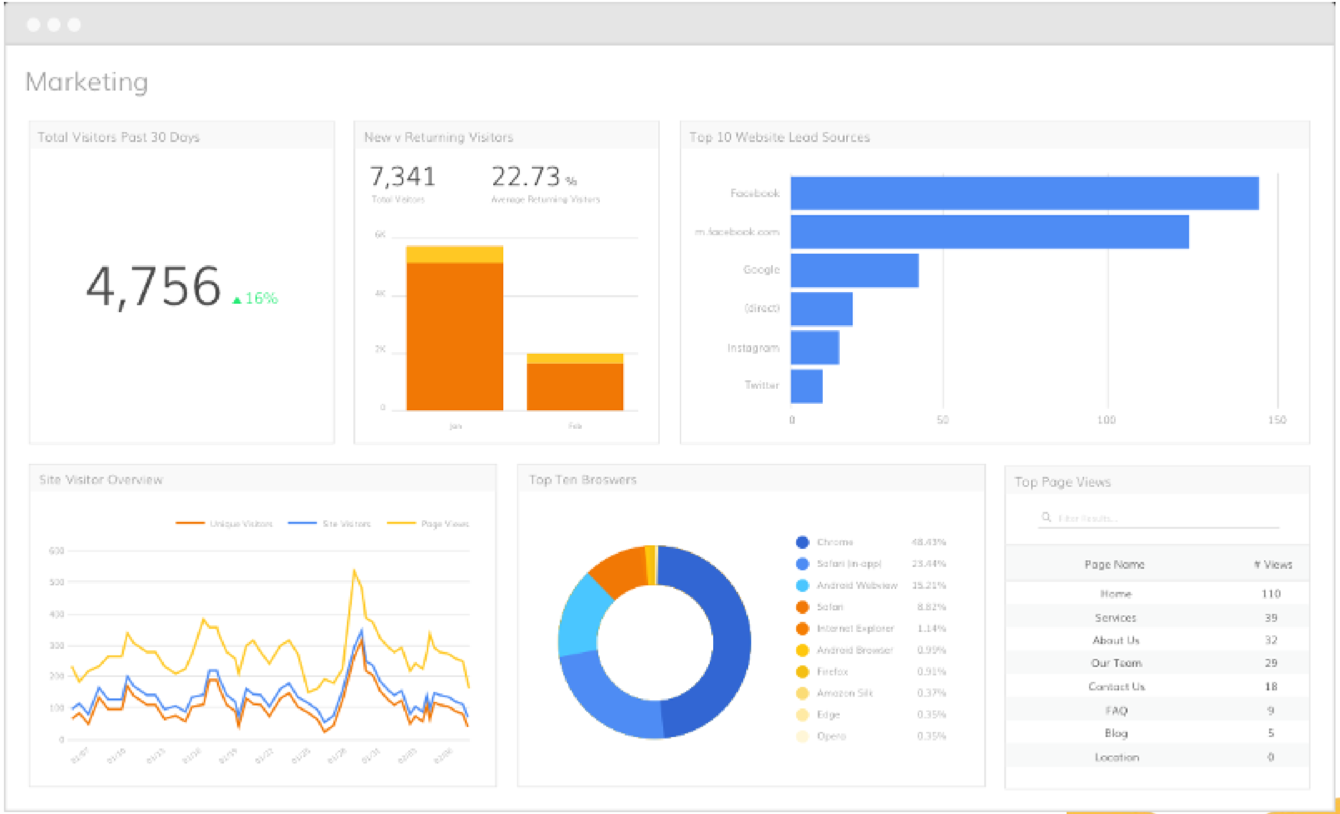
1. Limited Depth
2. Potential for Misinterpretation
3. Data Quality Dependency
4. Not Forward-Looking
5. Lack of Competitive Advantage

Use cases of descriptive analytics

- Healthcare
- Retail
- Manufacturing
- Banking and Finance
- Sports
- Government and Public Policy
- Media and Entertainment
- Education

What is Descriptive Analytics in Marketing?

- ❑ Campaign Performance.
- ❑ Customer Segmentation.
- ❑ Market Share and Competitor Analysis.
- ❑ Sales Trends.
- ❑ Customer Behavior.
- ❑ Budget Utilization



The Role of Descriptive Analytics in Future Data Analysis

In summary, the future of analytics is indeed a blend of descriptive, predictive, and prescriptive analytics, each feeding into and enhancing the other. Descriptive analytics serves not merely as a starting point but continues to play a role in refining, validating, and enhancing more advanced analytics endeavors.

The Bottom Line. Descriptive analytics can be a great way for companies to begin analyzing their performance metrics. That's because it's one of the easiest forms of data analysis. It's a straightforward approach to provide management, investors, and analysts with a direct comparison to similar metrics, such as quarter-over-quarter revenue. Using past performance can help key stakeholders better understand what happened so they make better, more informed decisions for the future.

ANY QUESTIONS?

3. Data visualization

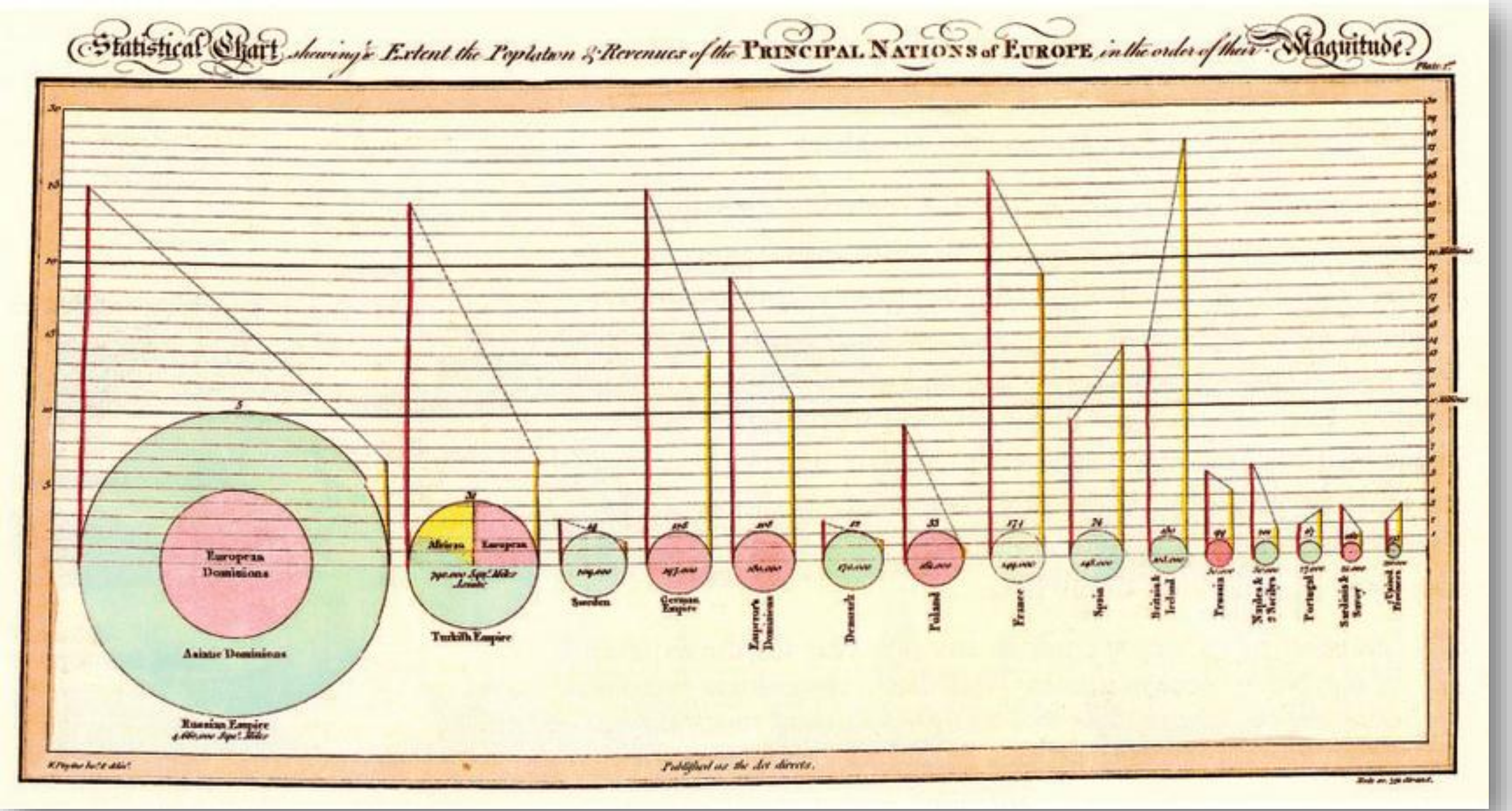
Khalil Israfilzada, PhD

Faculty of Economics and Management
Vytautas Magnus University
Kaunas, 2024

Becoming Visual

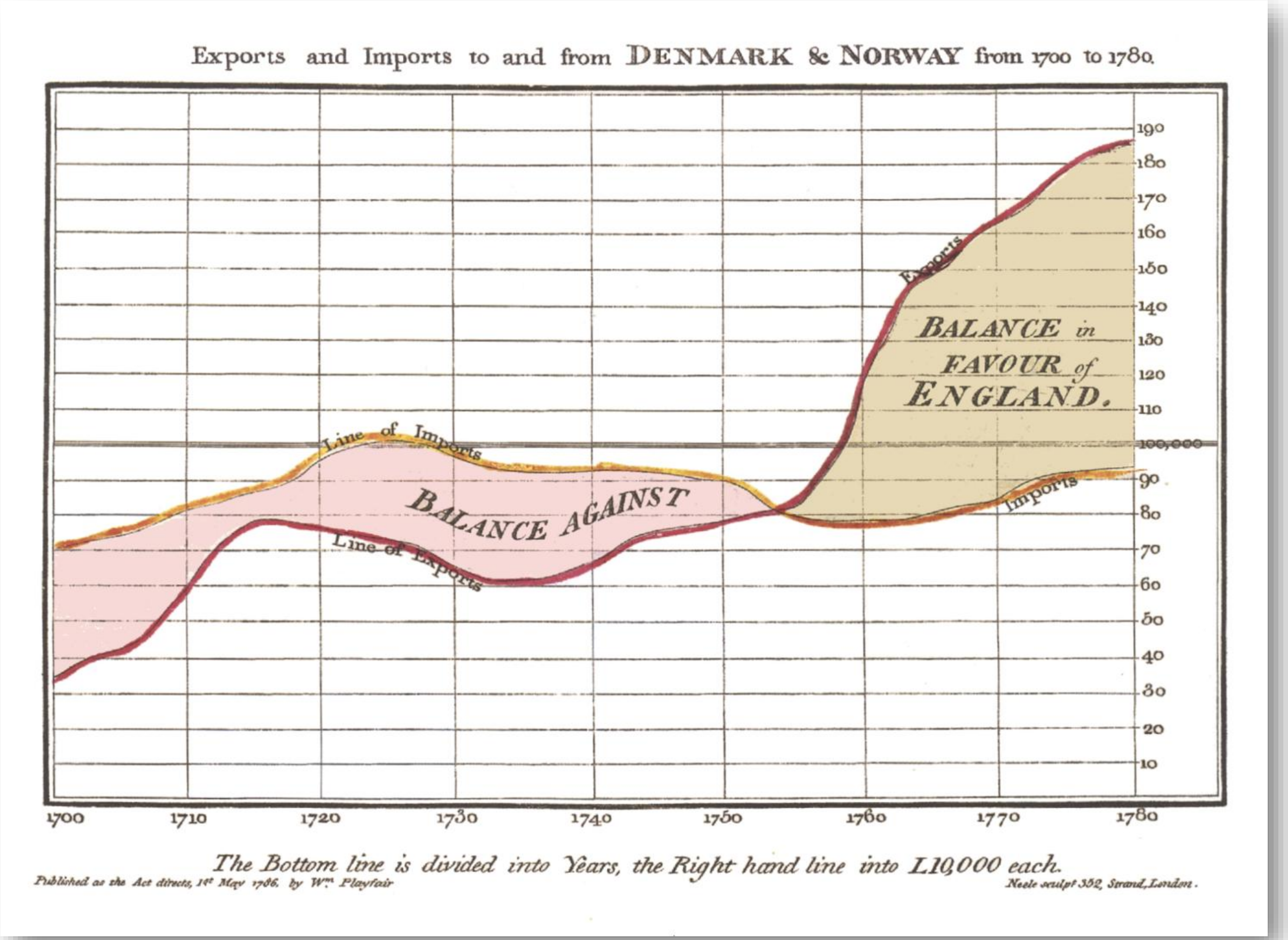
Beautiful (and not so beautiful) charts and graphs are everywhere. Visualization of information is a human practice dating back to the Chauvet cave drawings, over 32,000 years ago.

Becoming Visual



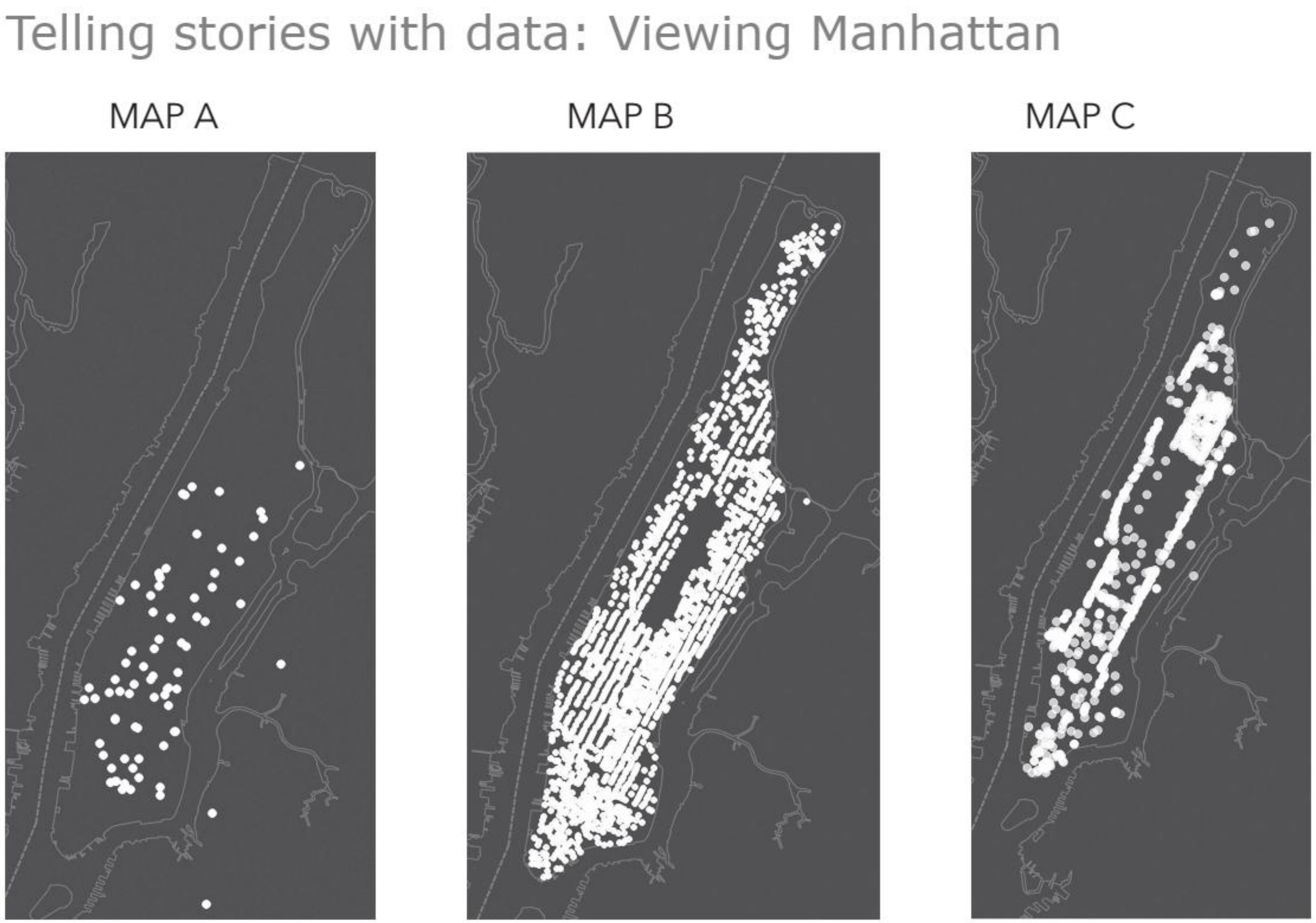
William Playfair (1786) is credited as the pioneer who showed economic data using bar charts. Playfair (1786) also invented the line graph. Playfair's work in the 1700s is paramount to the field of data visualization; it provided the foundation for future statistical data displays.

Becoming Visual



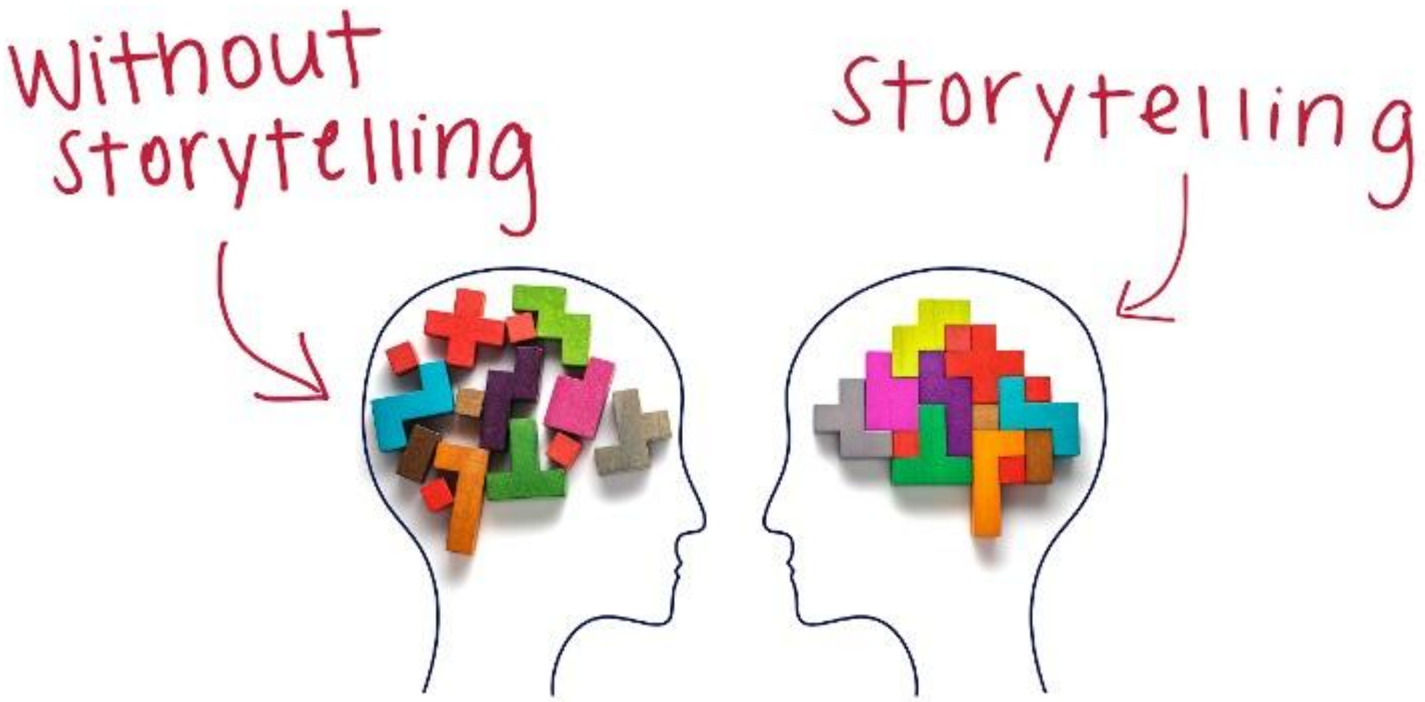
Playfair was experimenting with data visualization long before his invention of the pie chart. He also came up with the more truthful bar chart, history's first example of which appeared in his Commercial and Political Atlas of 1786.

Trends in Data Visualization—Storytelling.



- **Highlighting behaviors** > Who's hailing a cab when the clocks strike midnight on New Year's Eve? Map A shows the location of taxi cab customer pickups at 12:00am on January 1, 2016.
- **Revealing similarities and differences** > Where do the most motor vehicle accidents occur in Manhattan? Map B is a point map that shows the locations of each accident during the month of January 2016.
- **Displaying locations** > Where can I pick up free Wi-Fi? Map C shows the location of each Wi-Fi hotspot in Manhattan.

Trends in Data Visualization—Storytelling.



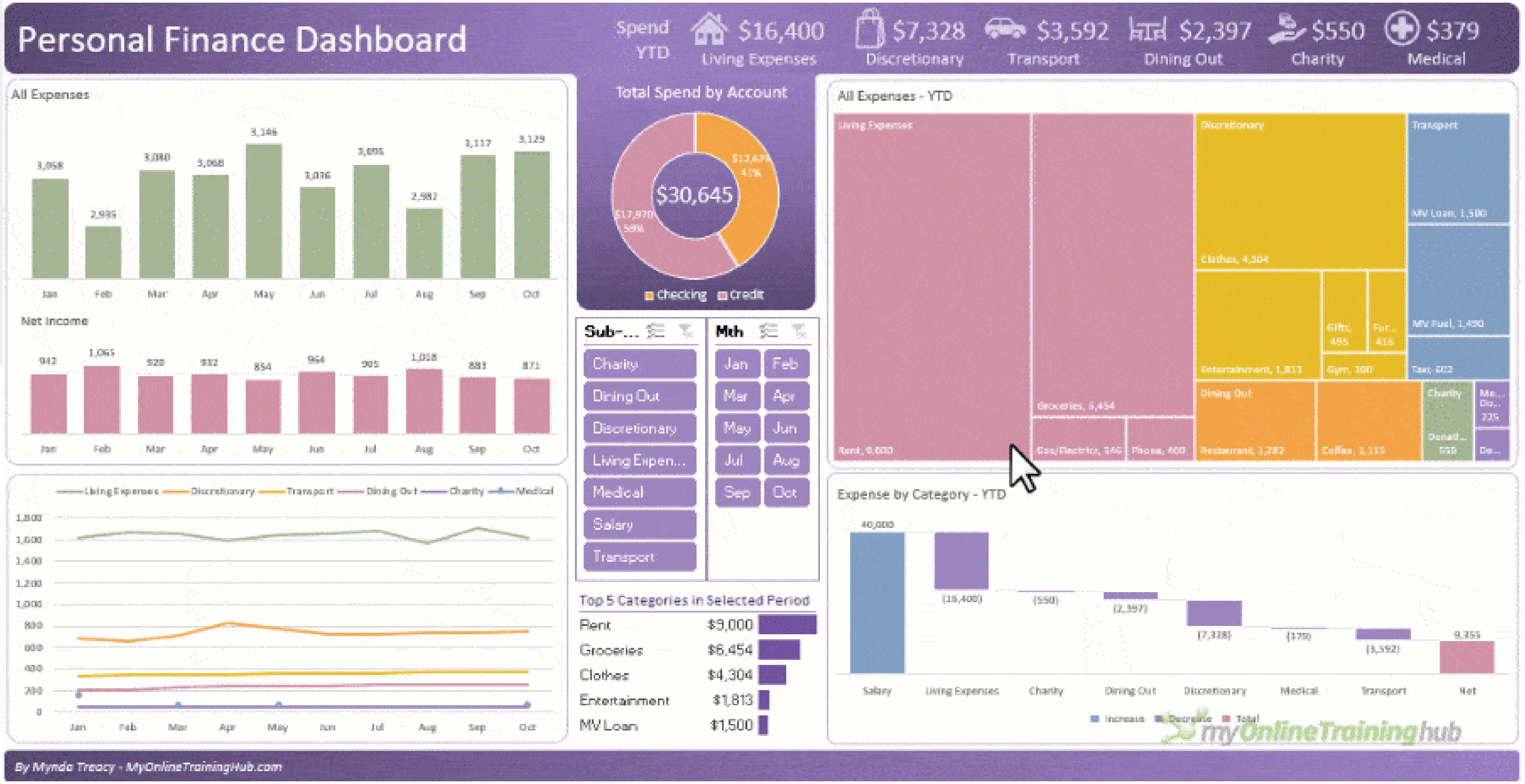
There are three key components to data storytelling:

1. **Data**: Thorough analysis of accurate, complete data s
Analyzing data using descriptive, diagnostic, predictive, and prescriptive analysis can enable you to understand its full picture.

2. **Narrative**: A verbal or written narrative, also called a storyline, is used to communicate insights gleaned from data, the context surrounding it, and actions you recommend and aim to inspire in your audience.

3. **Visualizations**: Visual representations of your data and narrative can be useful for communicating its story clearly and memorably. These can be charts, graphs, diagrams, pictures, or videos.

Trends in Data Visualization—Interactive Graphics.



What Is Data Visualization?

➢ They say a picture is worth a thousand words, and this is especially true for data analytics.

❑ data visualization as a process used to create data graphics.

Gestalt principles of visual perception

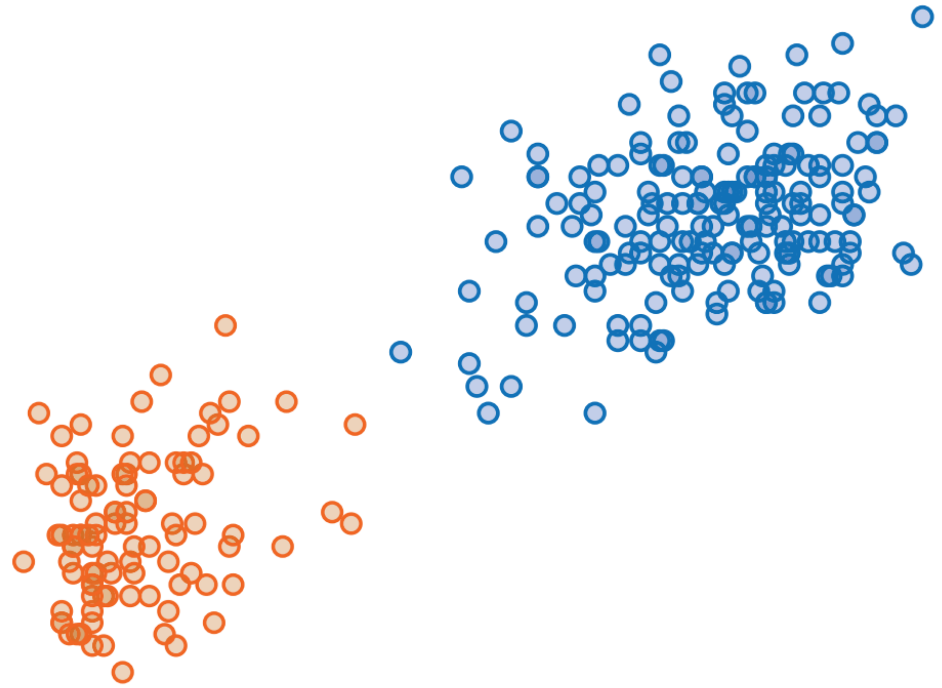
Proximity

We tend to think of objects that are physically close together as belonging to part of a group.

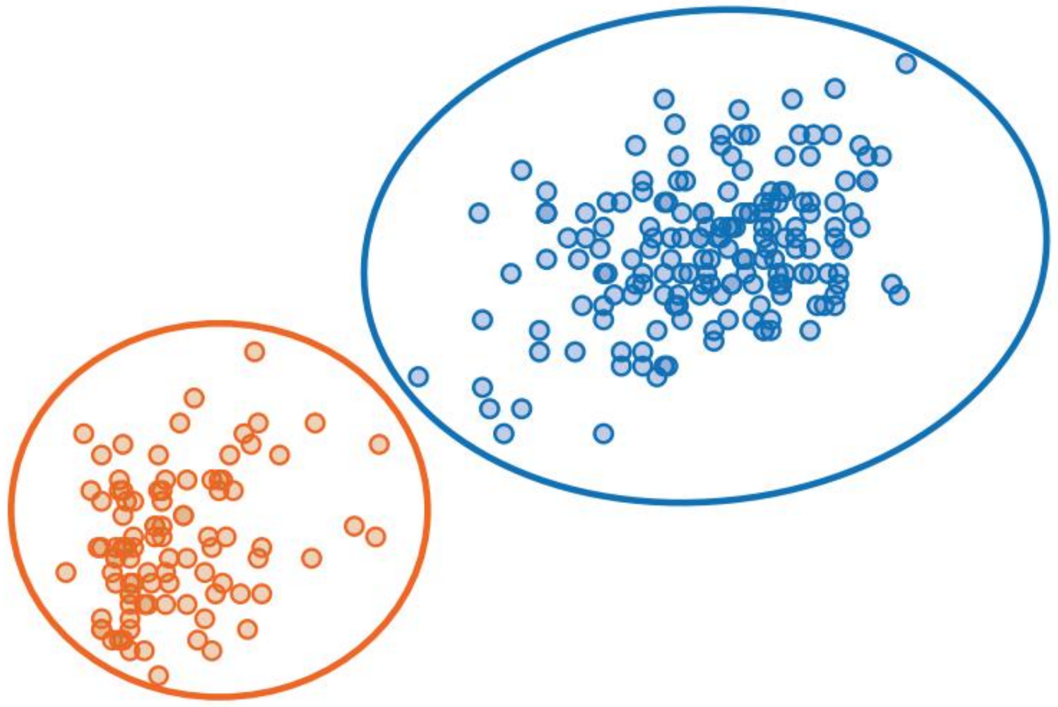
Similarity

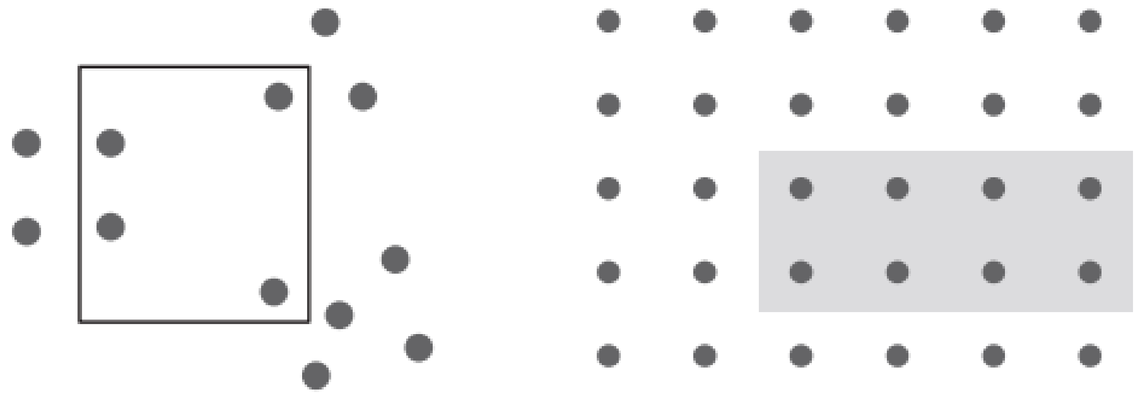
Objects that are of similar color, shape, size, or orientation are perceived as related or belonging to part of a group. Our brains group objects that share the same color, shape, or direction.

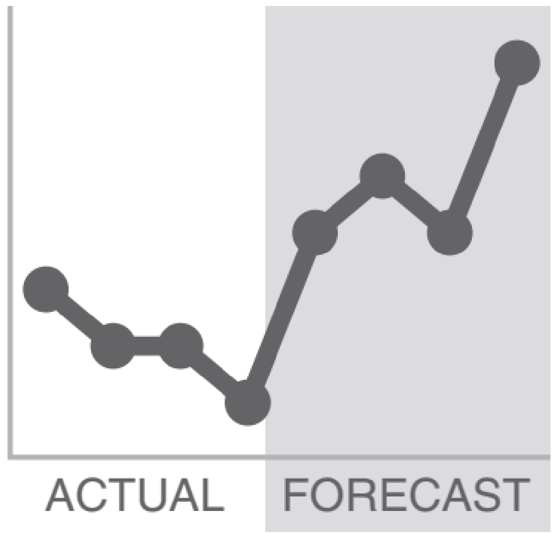
Adding color to the above scatterplot reinforces the two groups.



Enclosure



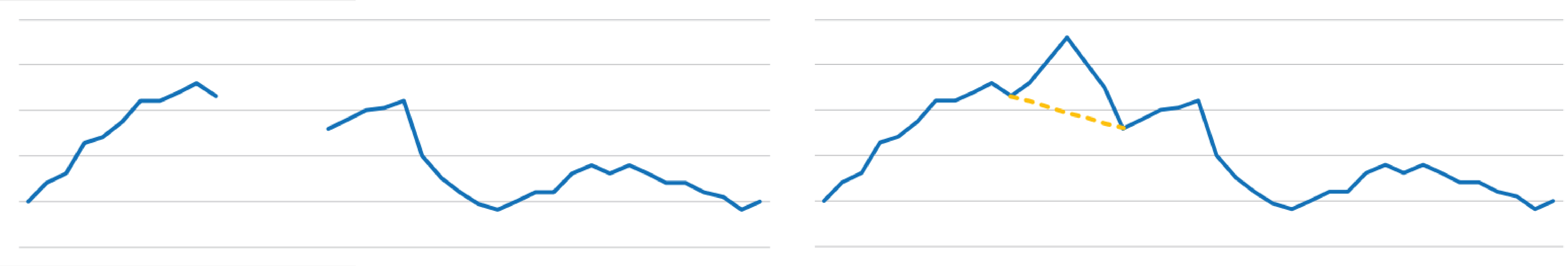




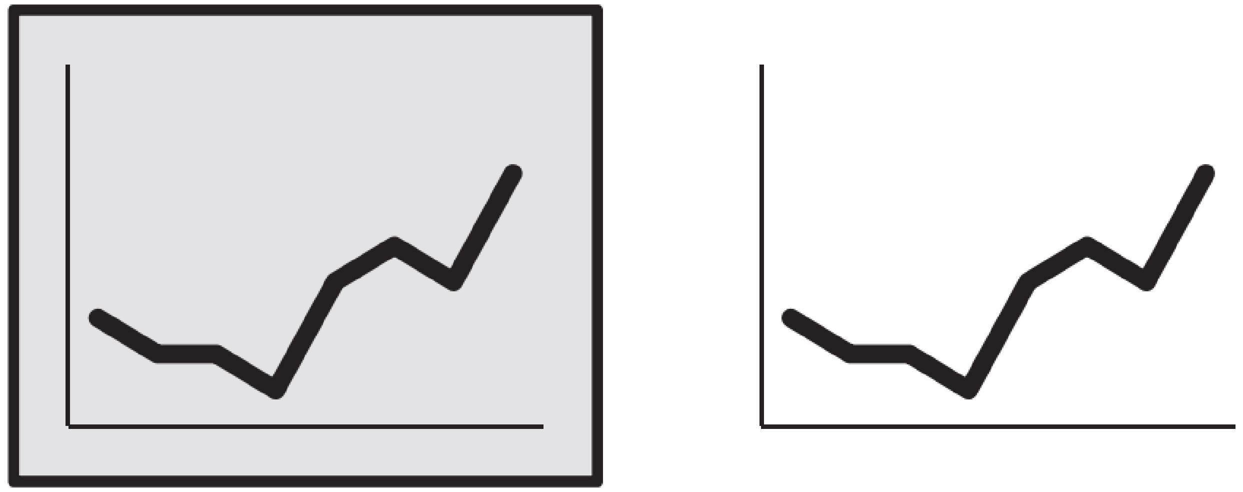
We think of objects that are physically enclosed together as belonging to part of a group. Bounded objects are perceived as a group.

Here, in addition to using color, we can enclose the two groups with circles or other shapes.

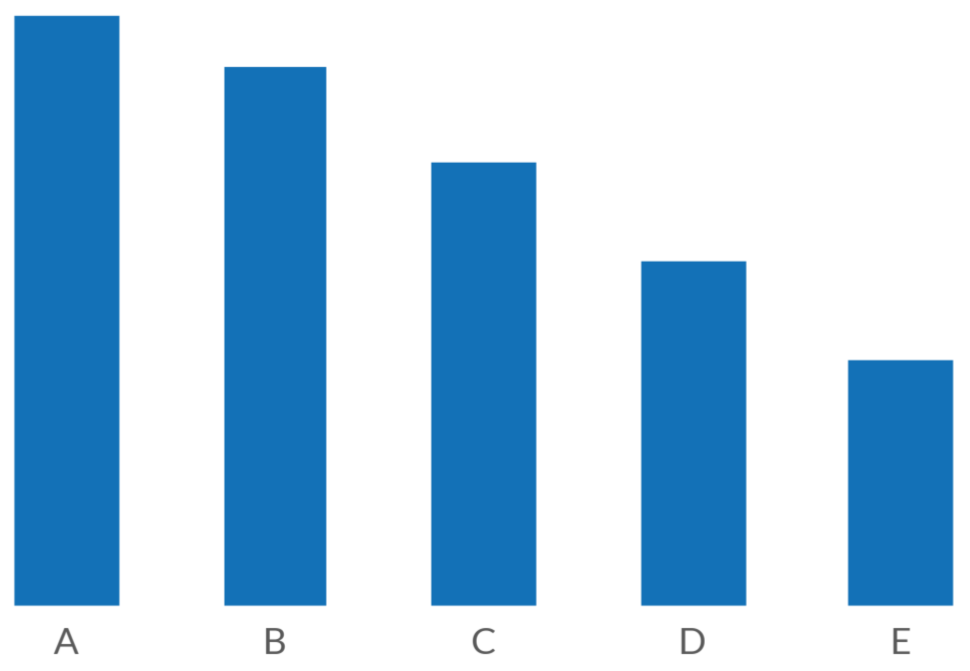
Closure

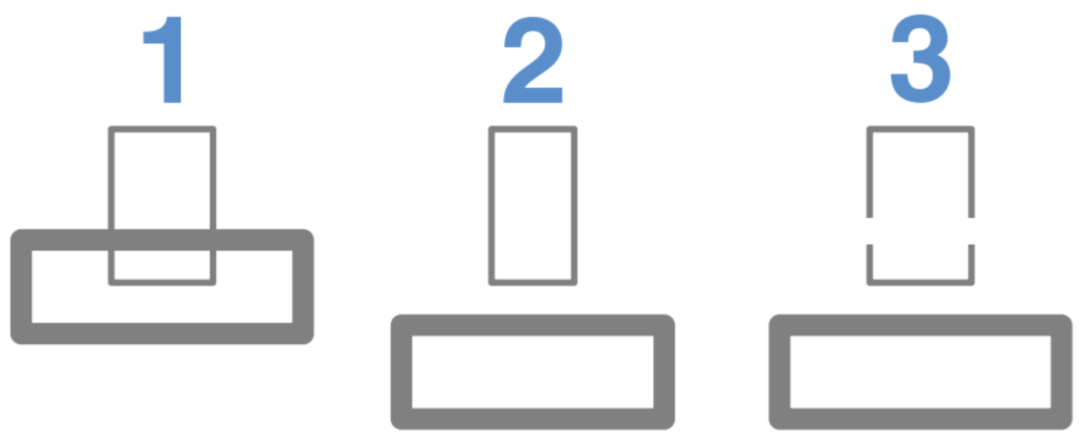


The closure concept says that people like things to be simple and to fit in the constructs that are already in our heads. Because of this, people tend to perceive a set of individual elements as a single, recognizable shape when they can—when parts of a whole are missing, our eyes fill in the gap.



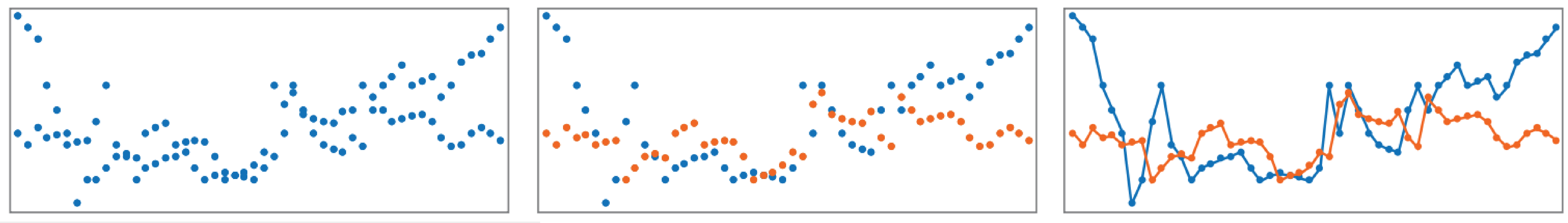
Continuity

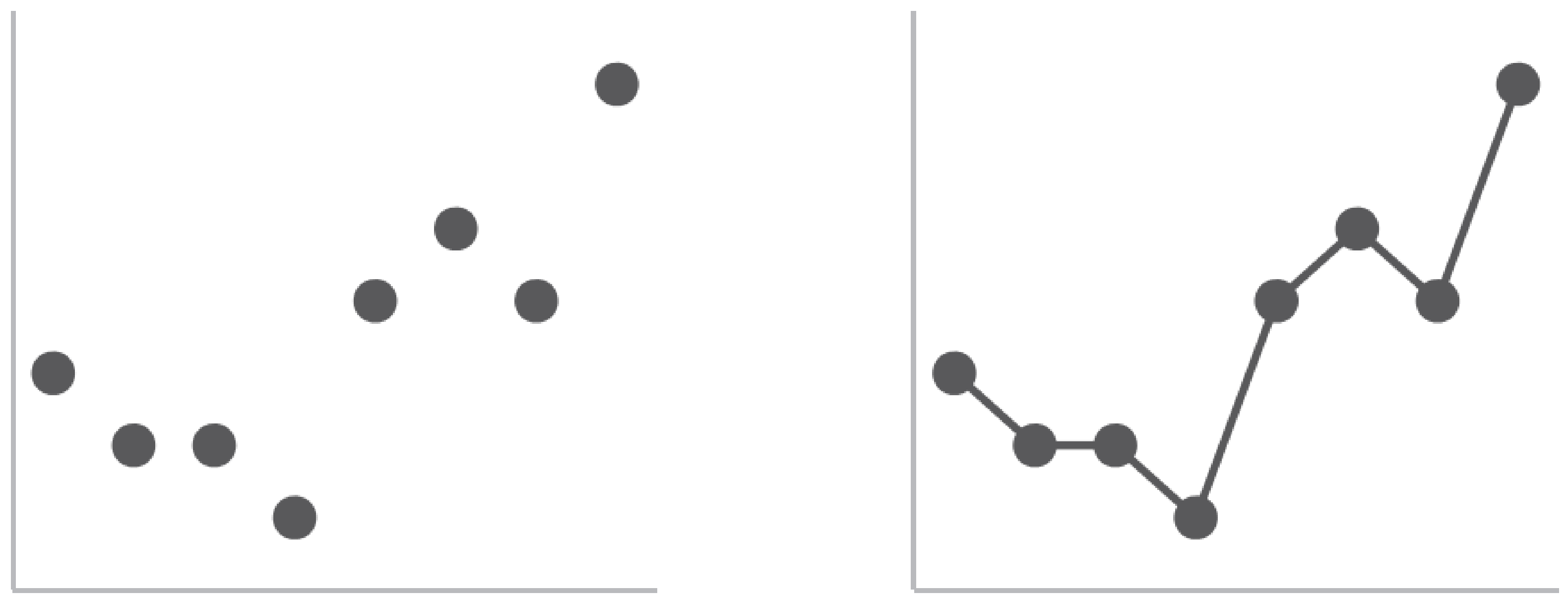




The principle of continuity is similar to closure: when looking at objects, our eyes seek the smoothest path and naturally create continuity in what we see even where it may not explicitly exist.

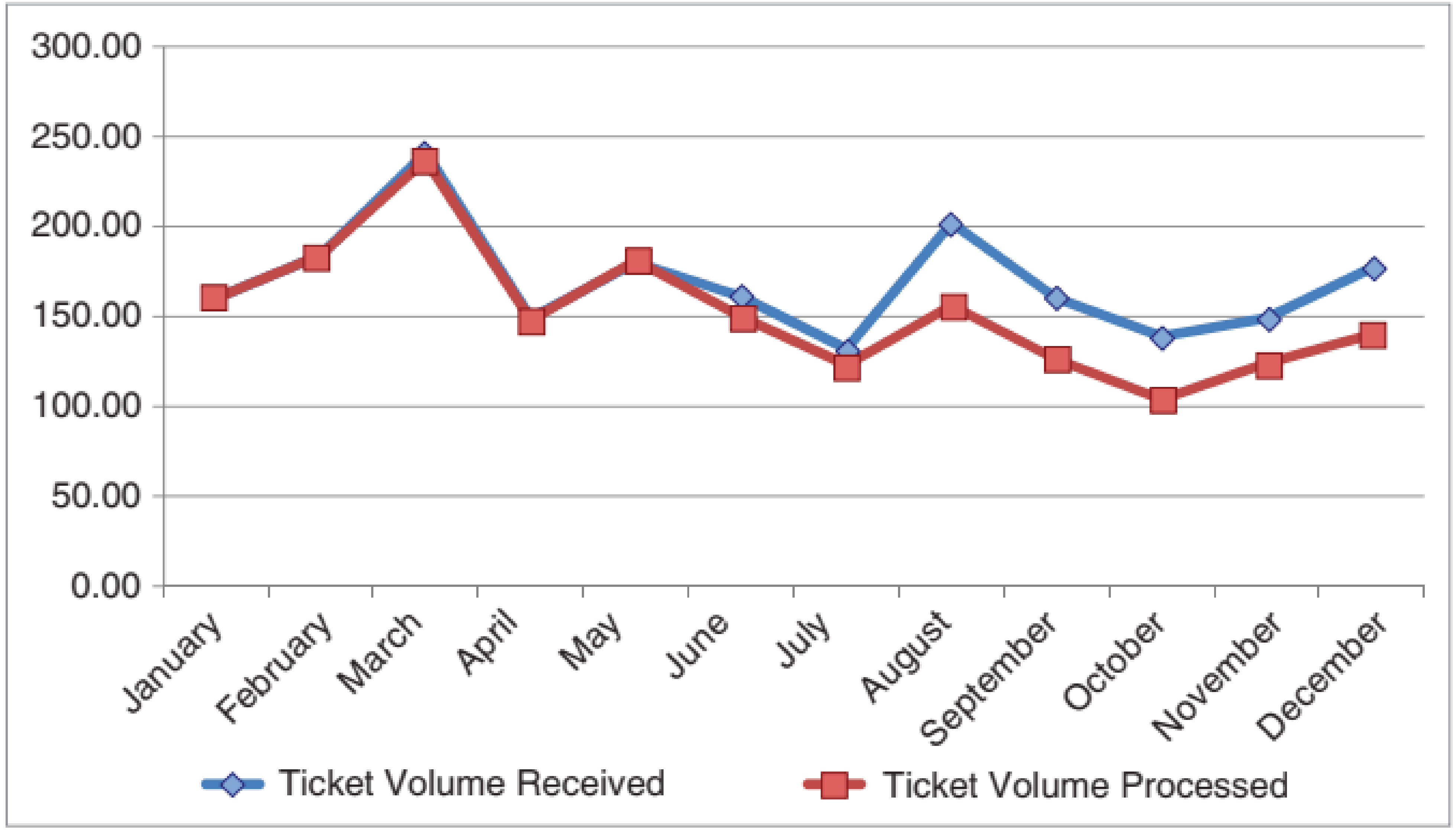
Connection





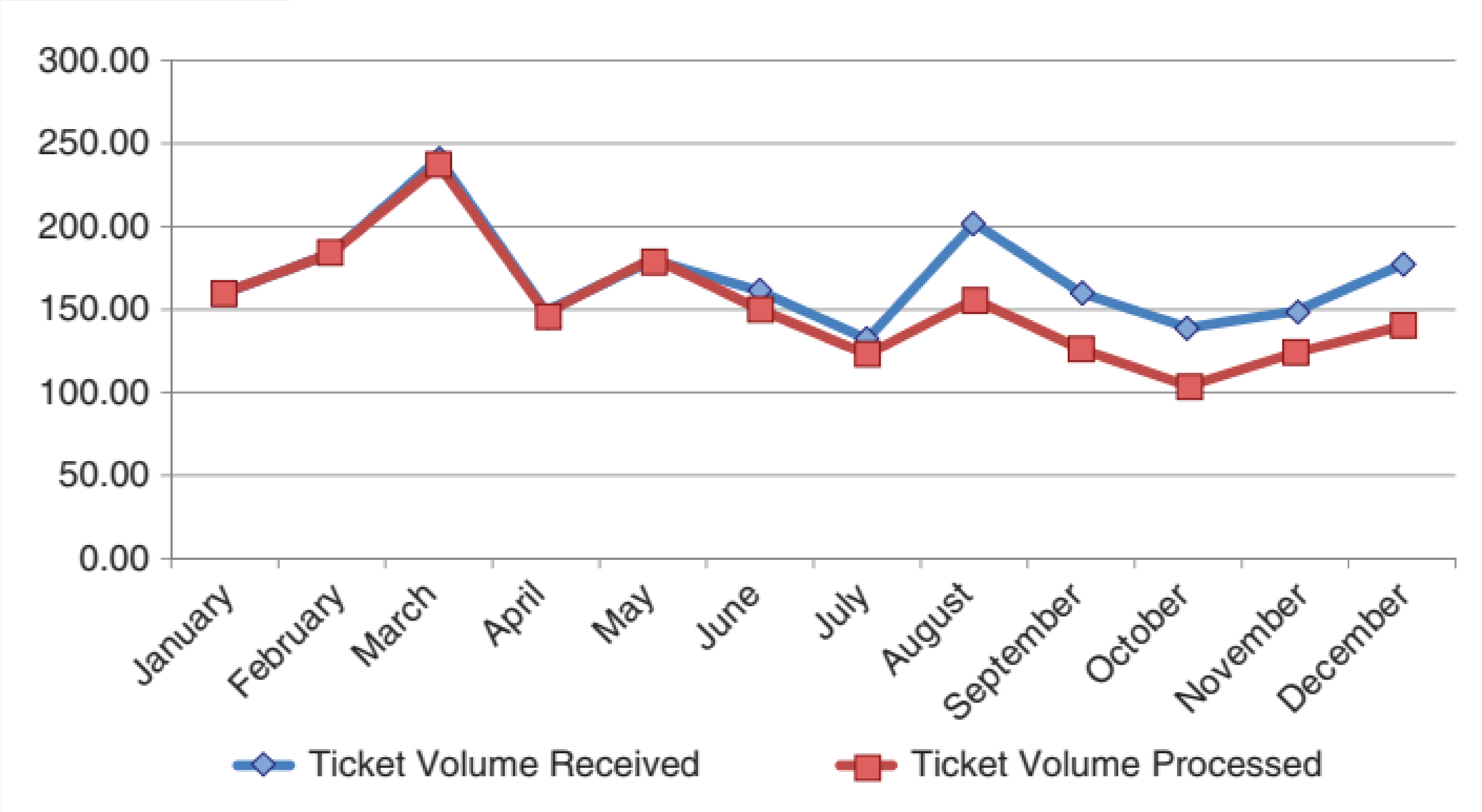
The final Gestalt principle we'll focus on is connection. We tend to think of objects that are physically connected as part of a group. The connective property typically has a stronger associative value than similar color, size, or shape.

Decluttering: step-by-step



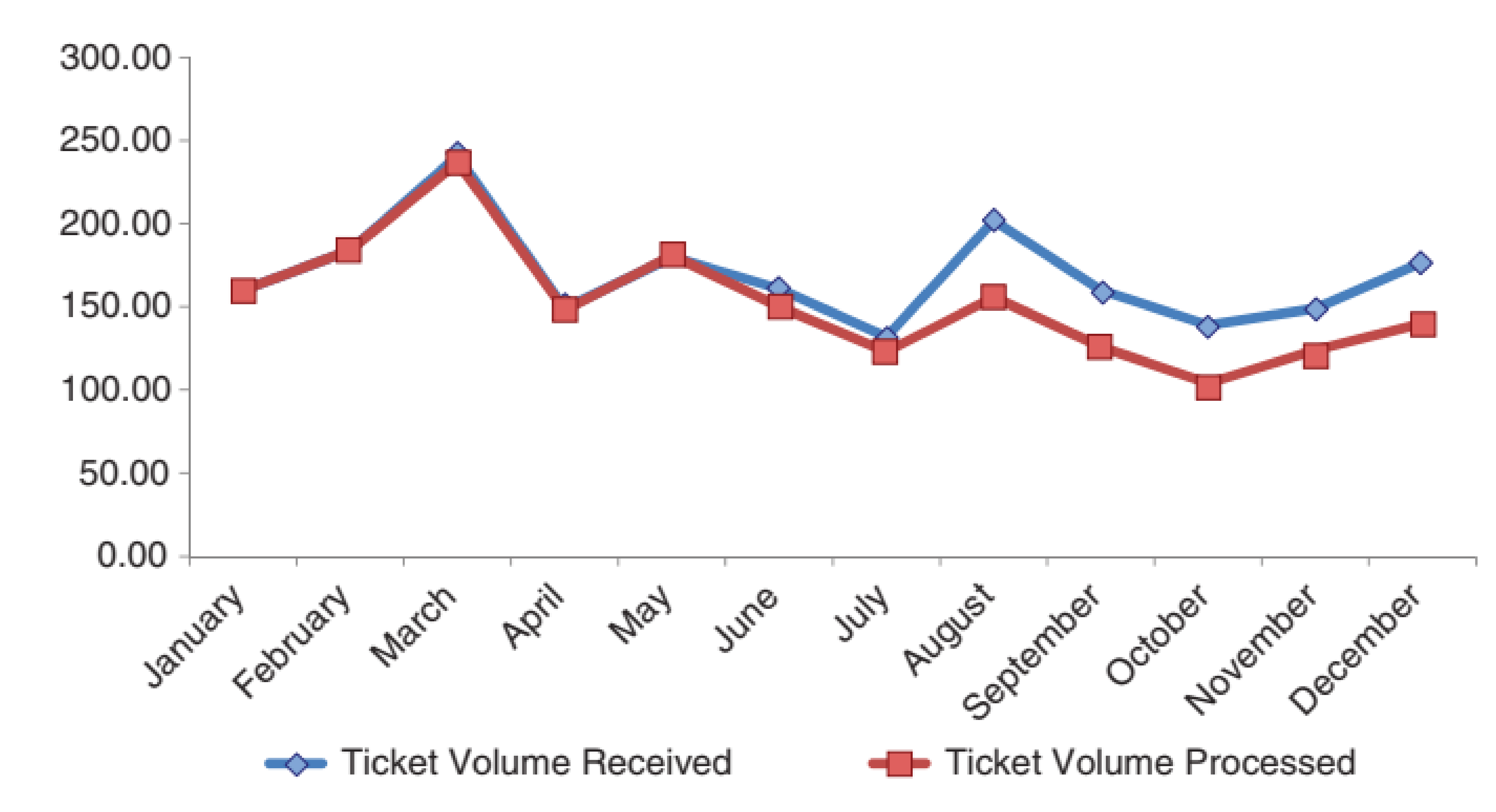
Scenario: Imagine that you manage an information technology (IT) team. Your team receives tickets, or technical issues, from employees. In the past year, you've had a couple of people leave and decided at the time not to replace them.

Decluttering: step-by-step



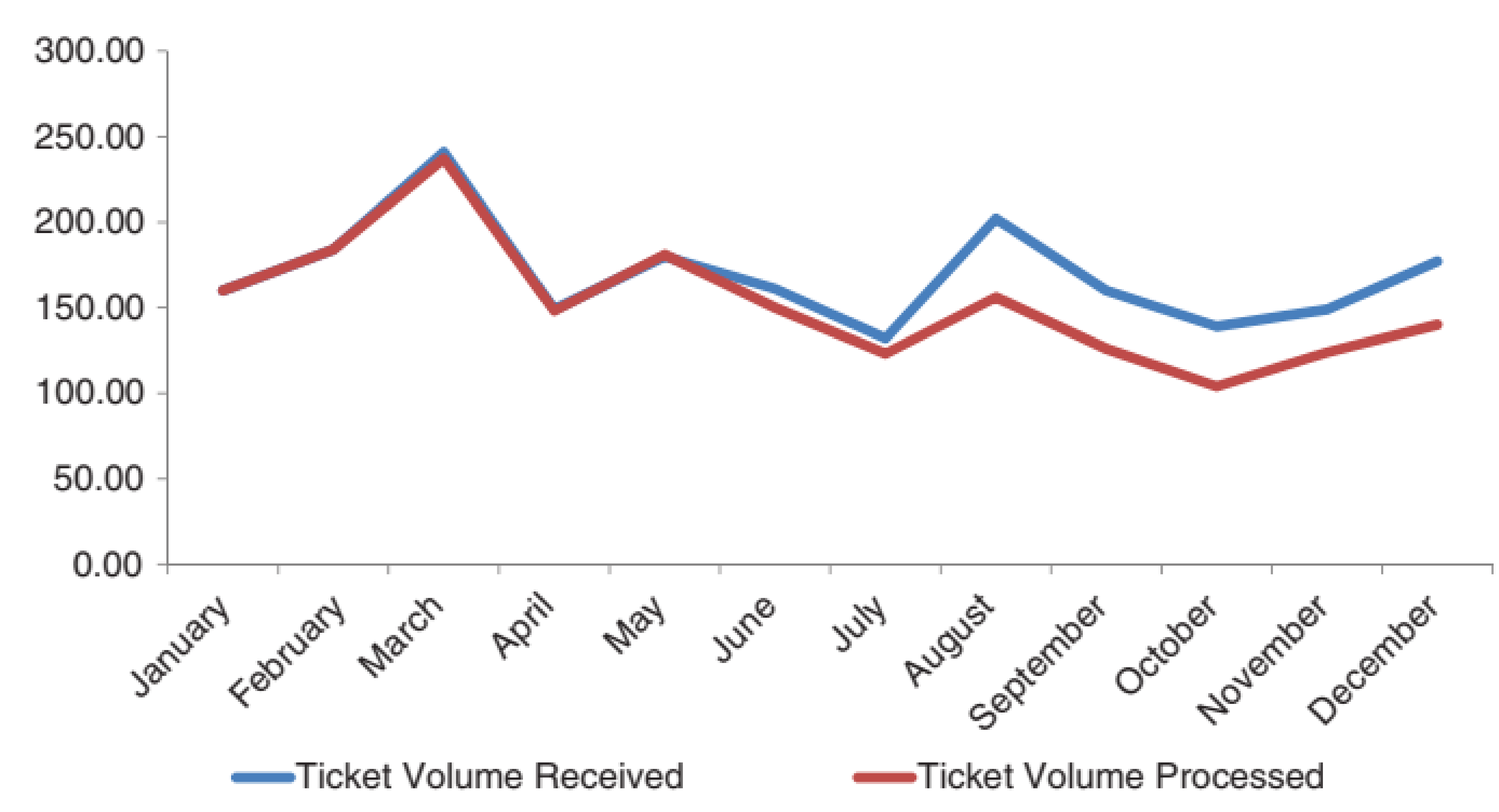
1. Remove chart border. Chart borders are usually unnecessary, as we covered in our discussion of the Gestalt principle of closure. Instead, think about using white space to differentiate the visual from other elements on the page as needed.

Decluttering: step-by-step



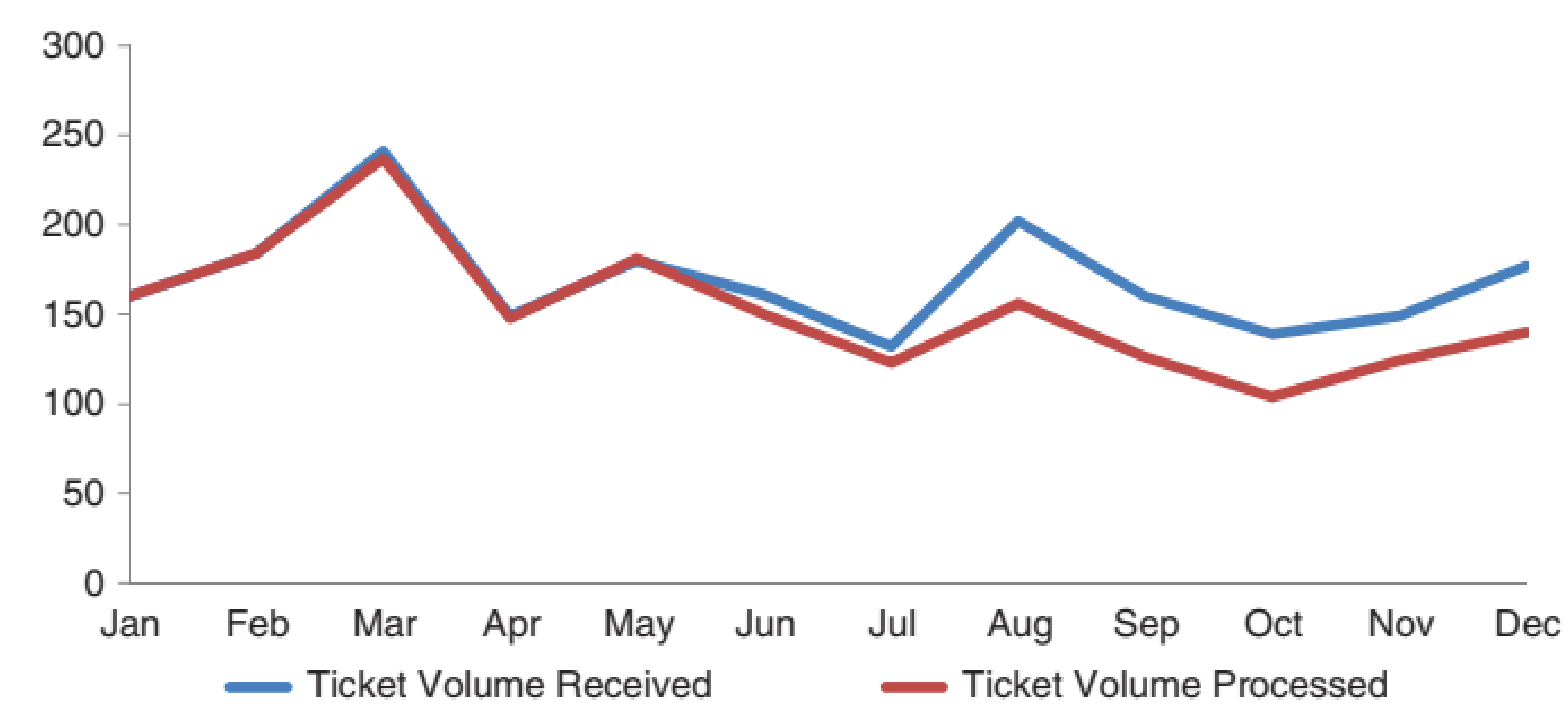
2. Remove gridlines. If you think it will be helpful for your audience to trace their finger from the data to the axis, or you feel that your data will be more effectively processed, you can leave the gridlines.

Decluttering: step-by-step



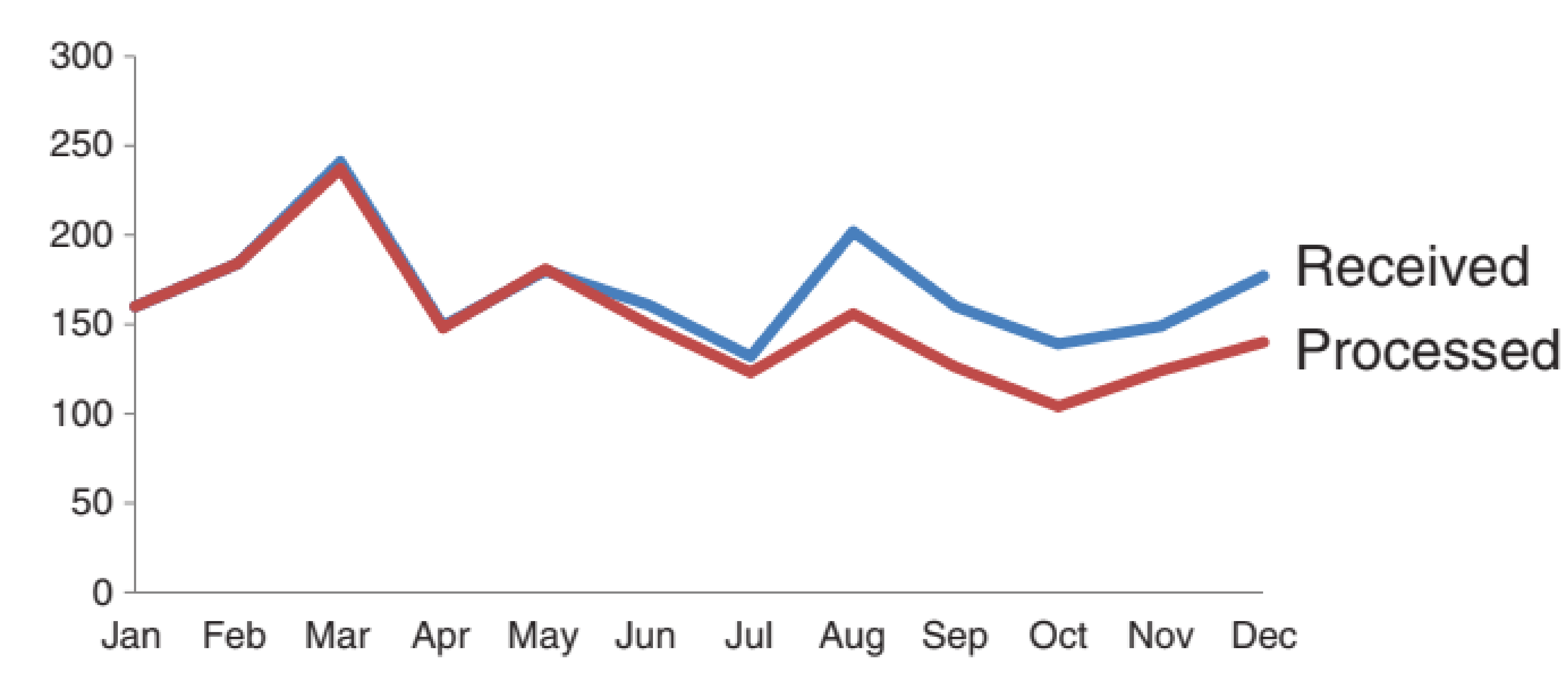
3. Remove data markers. Remember, every single element adds cognitive load on the part of your audience. Here, we're adding cognitive load to process data that is already depicted visually with the lines.

Decluttering: step-by-step



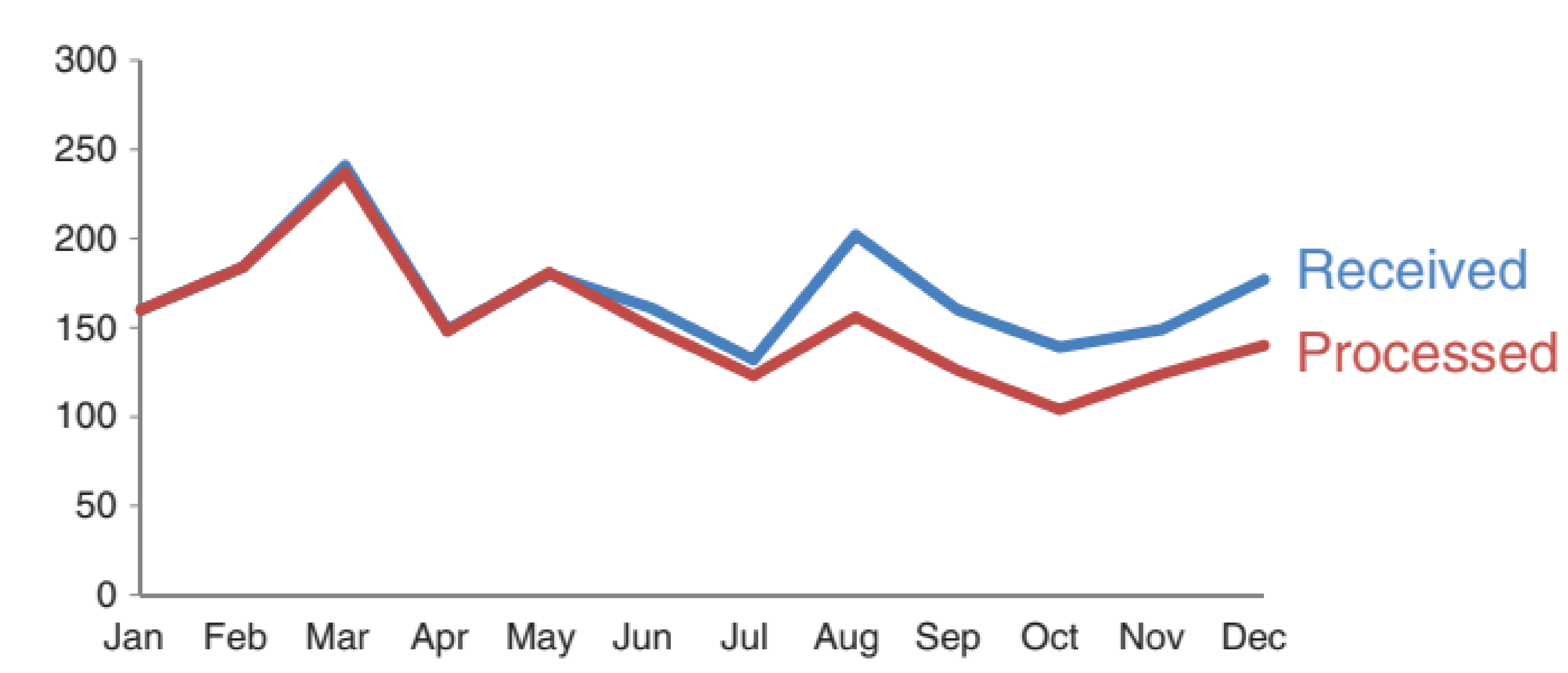
4. Clean up axis labels. One of my biggest pet peeves is trailing zeros on *y*-axis labels: they carry no informative value, and yet make the numbers look more complicated than they are! Get rid of them, reducing their unnecessary burden on the audience's cognitive load.

Decluttering: step-by-step



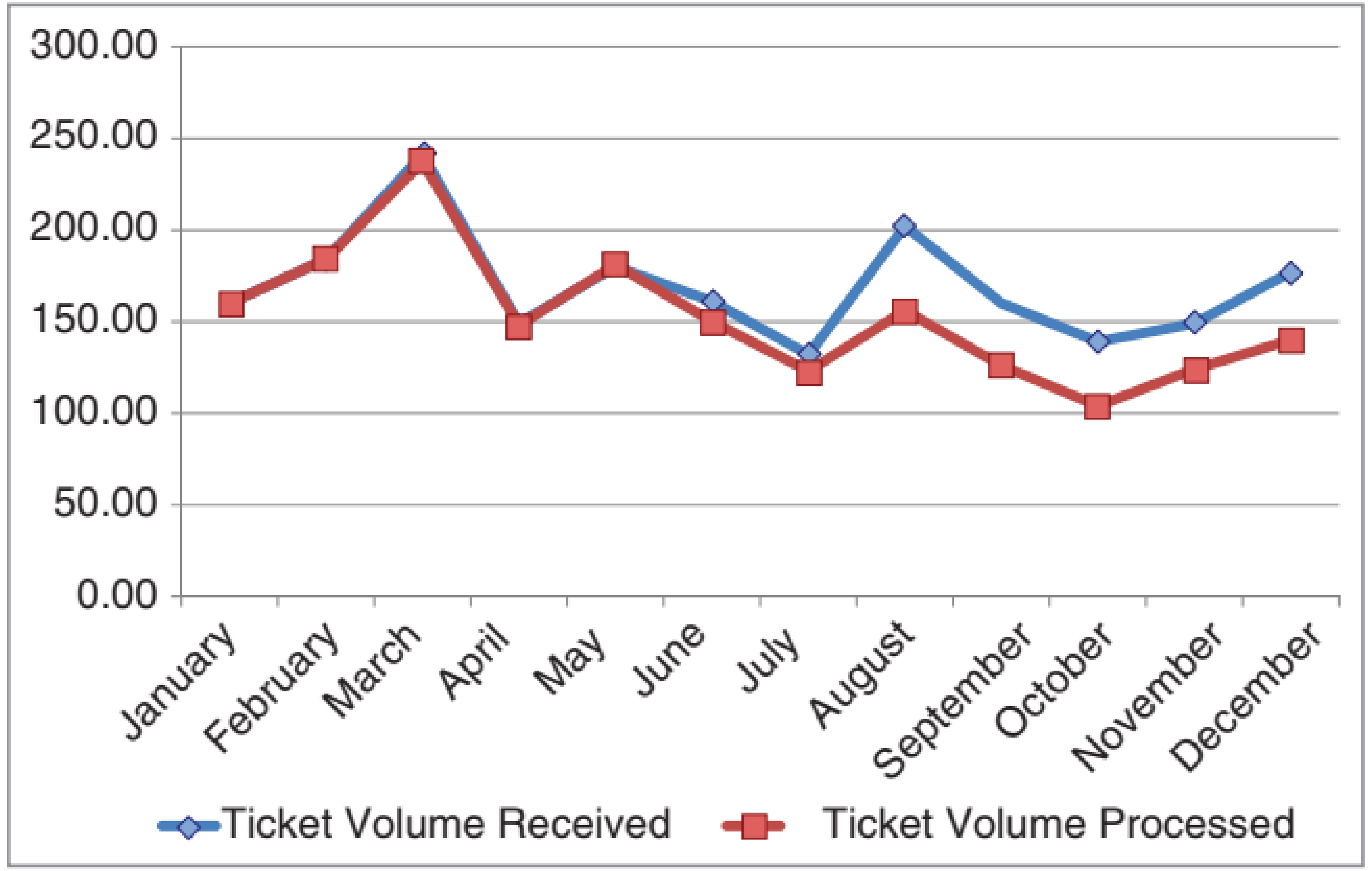
5. Label data directly. Now that we have eliminated much of the extraneous cognitive load, the work of going back and forth between the legend and the data is even more evident.

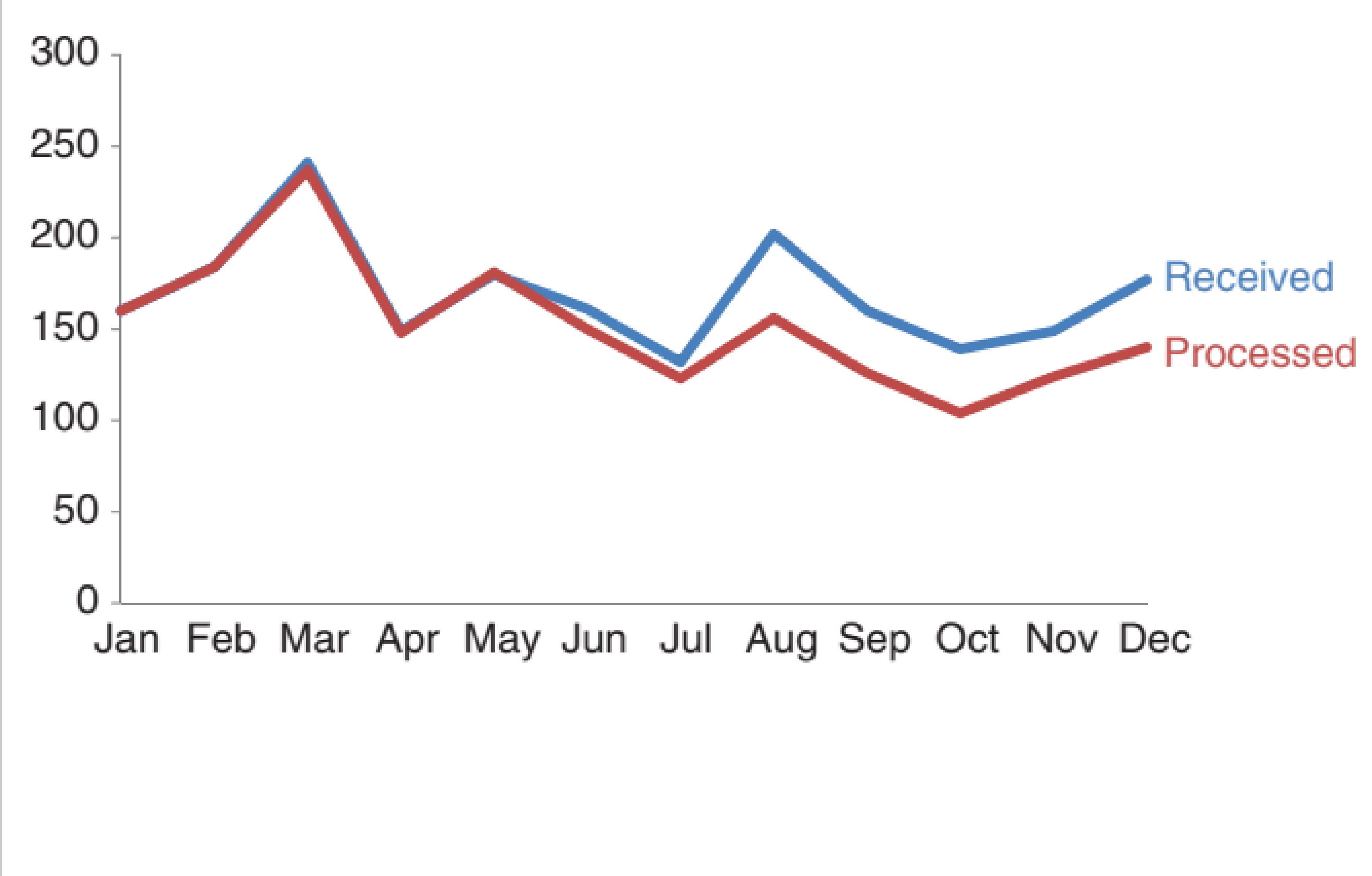
Decluttering: step-by-step



6. Leverage consistent color. While we leveraged the Gestalt principle of proximity in the prior step, let's also think about leveraging the Gestalt principle of similarity and make the data labels the same *color* as the data they describe.

Decluttering: step-by-step





This visual is not yet complete. But identifying and eliminating the clutter has brought us a long way in terms of reducing cognitive load and improving accessibility. Take a look at the before-and-after shown.

PREATTENTIVE PROCESSING

	Q1	Q2	Q3	Q4
Bob	26	35	72	84
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	95	13	64
Jon	83	62	46	48
Karen	30	65	98	82
Ken	38	28	45	71
Lauren	98	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	**$400**	**$475**	**$500**	**$600**

The concept of “preattentive processing” is a subset of Gestalt theory, and it is the visual process that consider most when creating data visualizations.

As we just saw, because our eyes can detect a limited set of visual characteristics, we combine various features of an object and unconsciously perceive them as a single image.

PREATTENTIVE PROCESSING

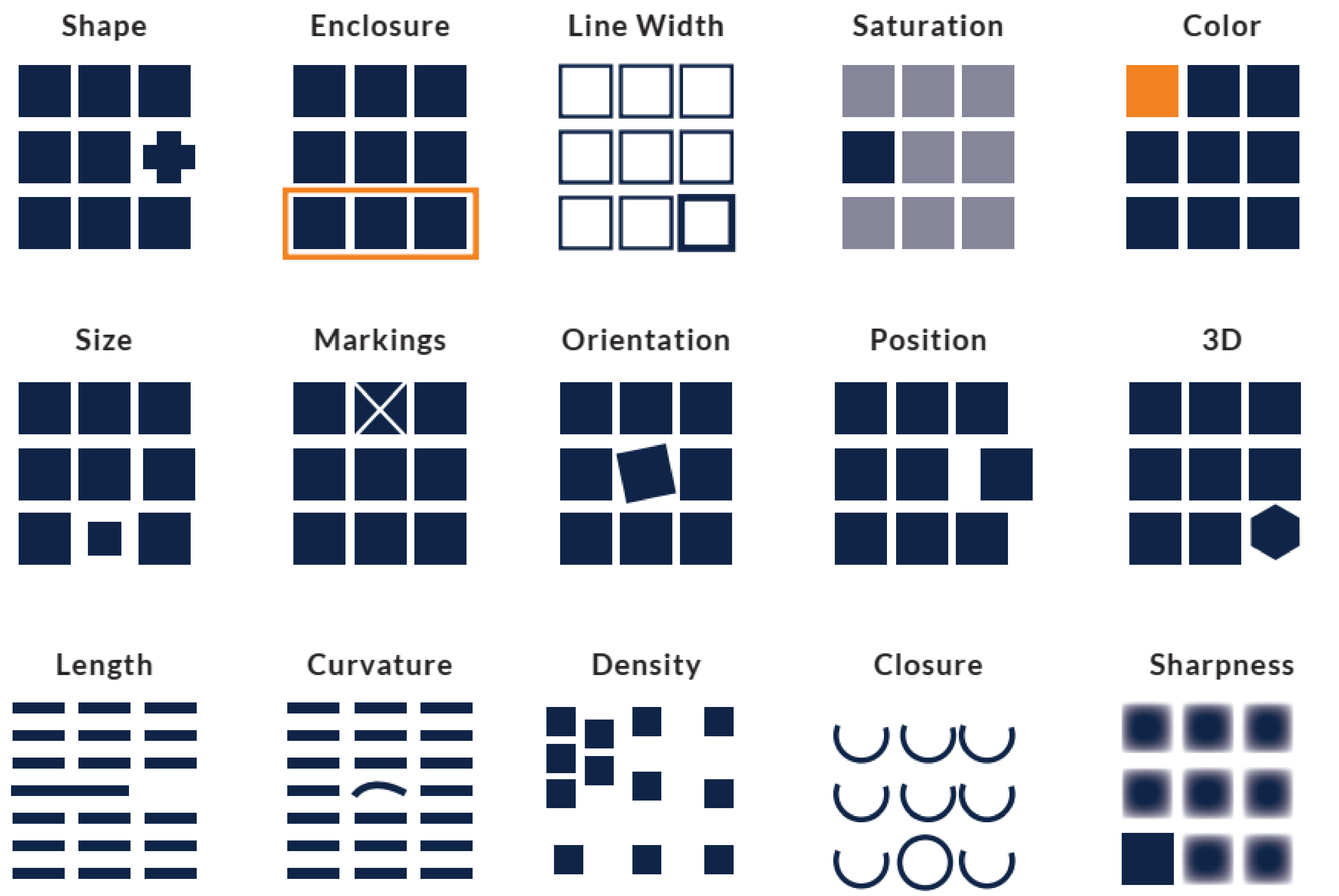
	Q1	Q2	Q3	Q4
Bob	26	35	72	84
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	95	13	64
Jon	83	62	46	48
Karen	30	65	98	82
Ken	38	28	45	71
Lauren	98	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	**$400**	**$475**	**$500**	**$600**

	Q1	Q2	Q3	Q4
Bob	26	35	72	**84**
Ellie	22	15	61	35
Gerrie	19	20	71	55
Jack	22	**95**	13	64
Jon	83	62	46	48
Karen	30	65	**98**	82
Ken	38	28	45	71
Lauren	**98**	81	41	63
Steve	16	50	23	41
Valerie	46	24	30	57
Total	**$400**	**$475**	**$500**	**$600**

Preattentive attributes here direct our attention to the large numbers immediately.

Hard to do, right? Now try it with these versions that use color (on the left) and intensity (on the right) to highlight those four numbers.

PREATTENTIVE PROCESSING



Examples of preattentive attributes that we can use in our visualizations to direct our reader's attention.

It's easier to find the numbers in these two tables than the first because the numbers are encoded using *preattentive attributes*: color and weight. Each distinction helps us effortlessly identify the key number.

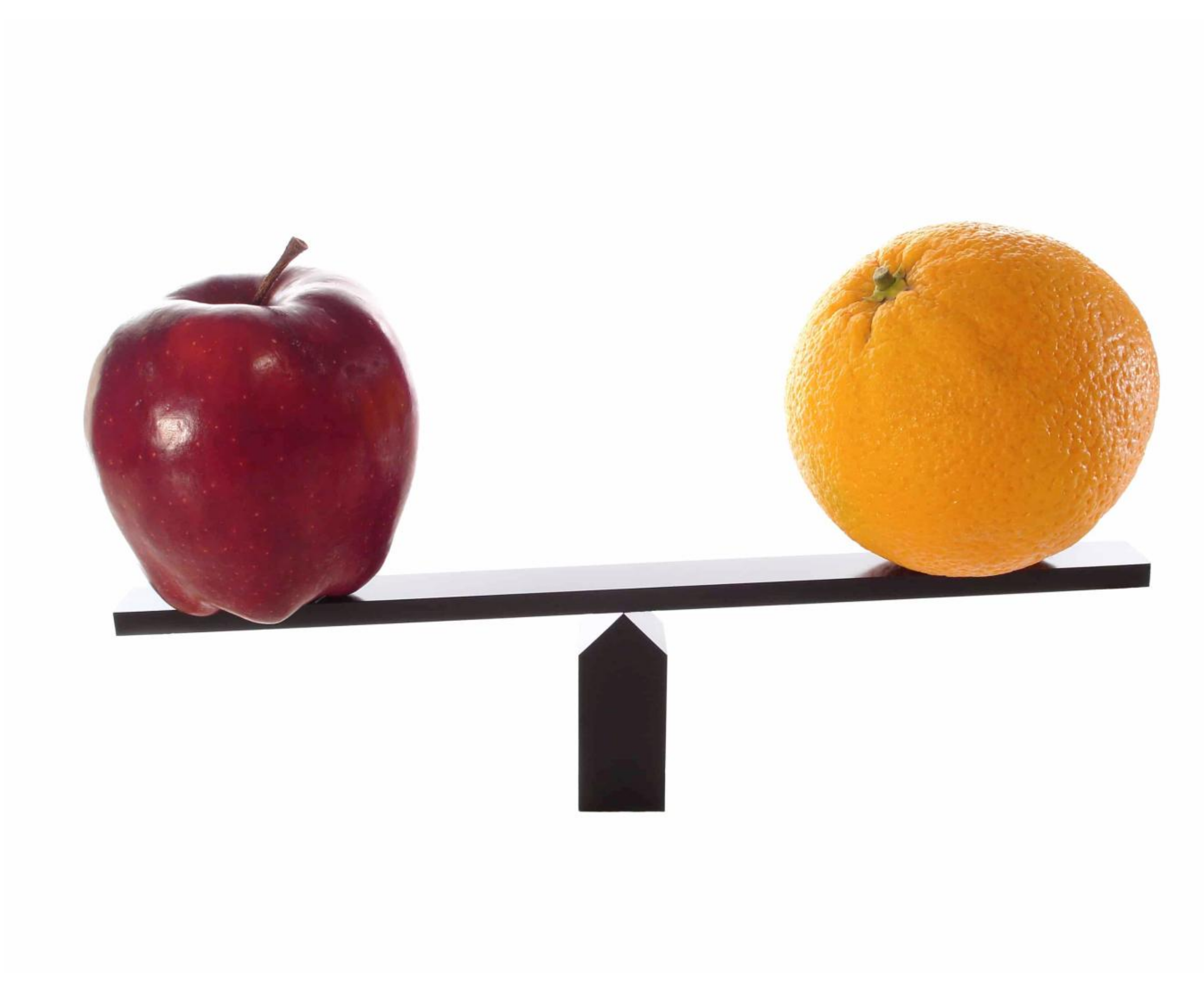
PREATTENTIVE PROCESSING

Notice how your eye gravitates toward the four tomatoes in the top-right part of the image on the left. The image on the right is balanced, so your eye doesn't immediately focus on any particular area. Photos by NordWood Themes (left) and Tim Gouw (right) on Unsplash.

Choosing an effective visual

There are many different graphs and other types of visual displays of information, but a handful will work for the majority of your needs.

Comparisons of Categories and Time



Questions:

1. What's the best? What's the worst? Compared to what?
2. Who's ranked the highest? The lowest?
3. How does performance compare to the target or goal? For example, did total sales exceed the forecast?

Insight: use comparisons to illustrate the similarities and differences among categories. This includes the minimum value, maximum value, rank, performance, sum, totals, counts, and quantities.

Comparisons of Categories and Time

Vertical bar

Bars are arranged vertically on the x-axis. Each bar represents a category or sub-category. The bar height measures the quantity (count) or sum.

- Keep bars the same color and shade when they measure the same variable (Wong, 2010).
- Use a zero baseline for the y-axis.
- Show negative values below the baseline.
- Keep the width of the bar about twice the width of the space between the bars (Wong, 2010).

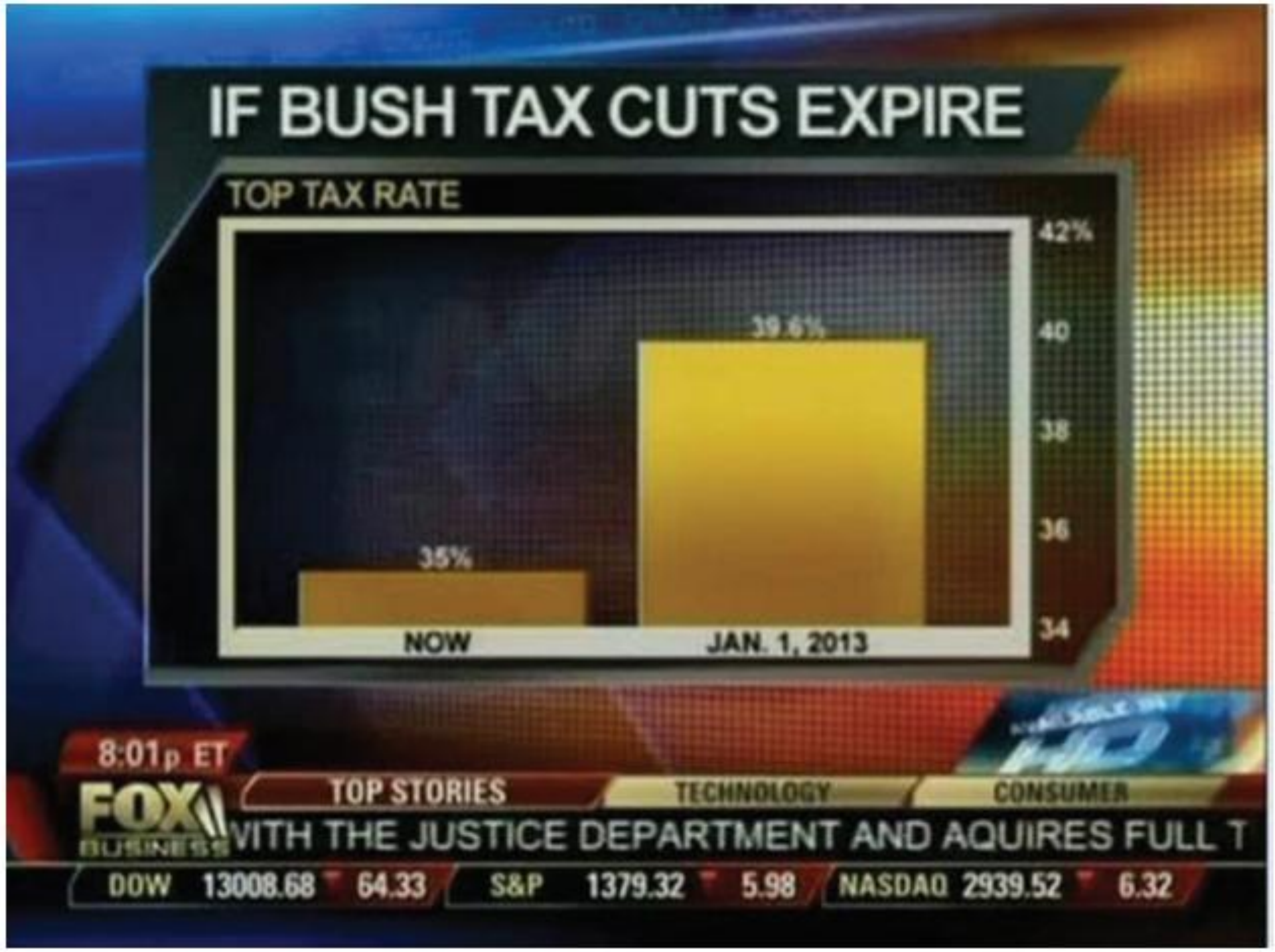
Column bar

Column bar charts present two series for each category.

- Use different color shading for each series.
- Shade bars from lightest to darkest (Wong, 2010).

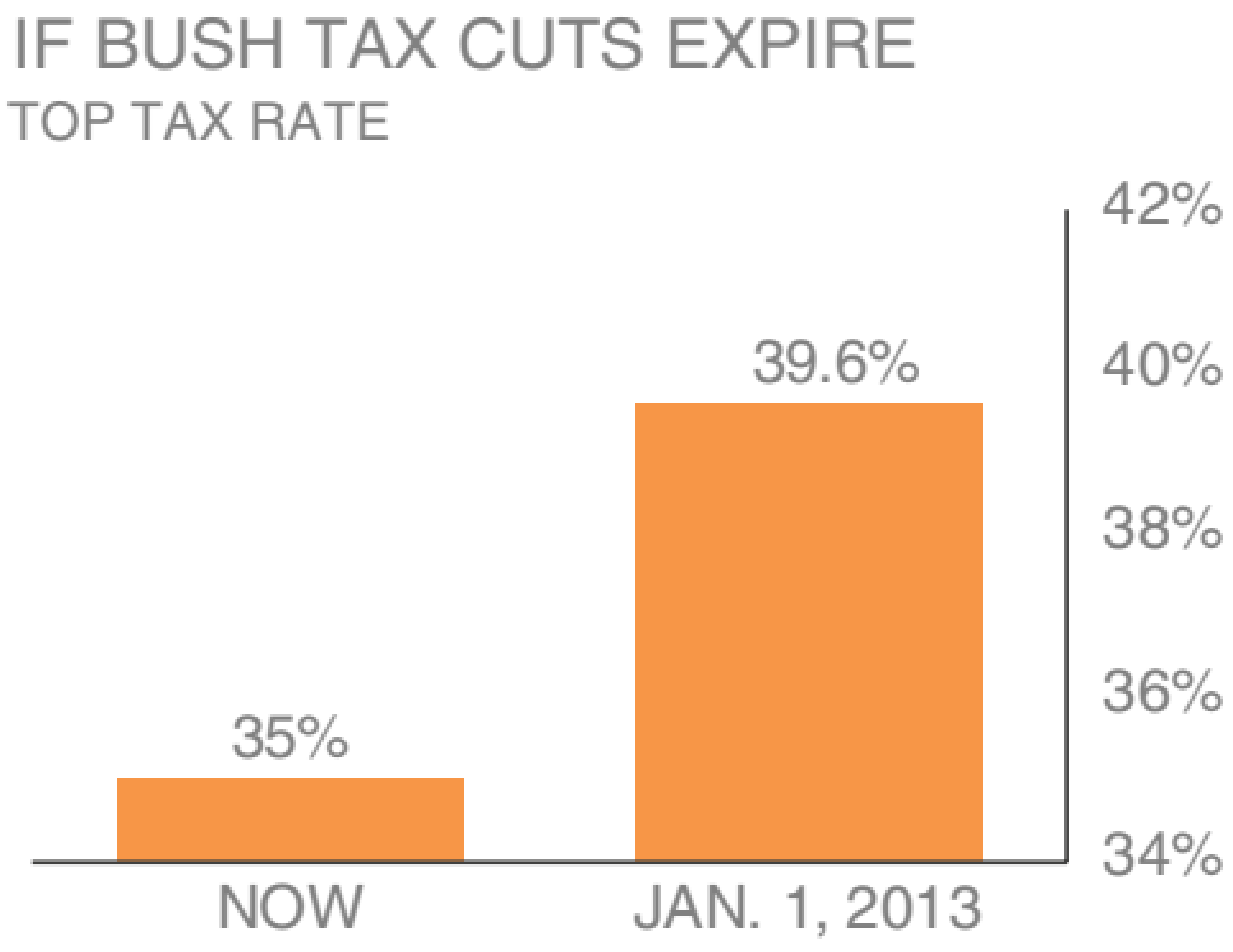
Comparisons of Categories and Time

For this example, let's imagine we are back in the fall of 2012. We are wondering what will happen if the Bush tax cuts expire. On the left-hand side, we have what the top tax rate is currently, 35%, and on the right-hand side what it will be as of January 1, at 39.6%.

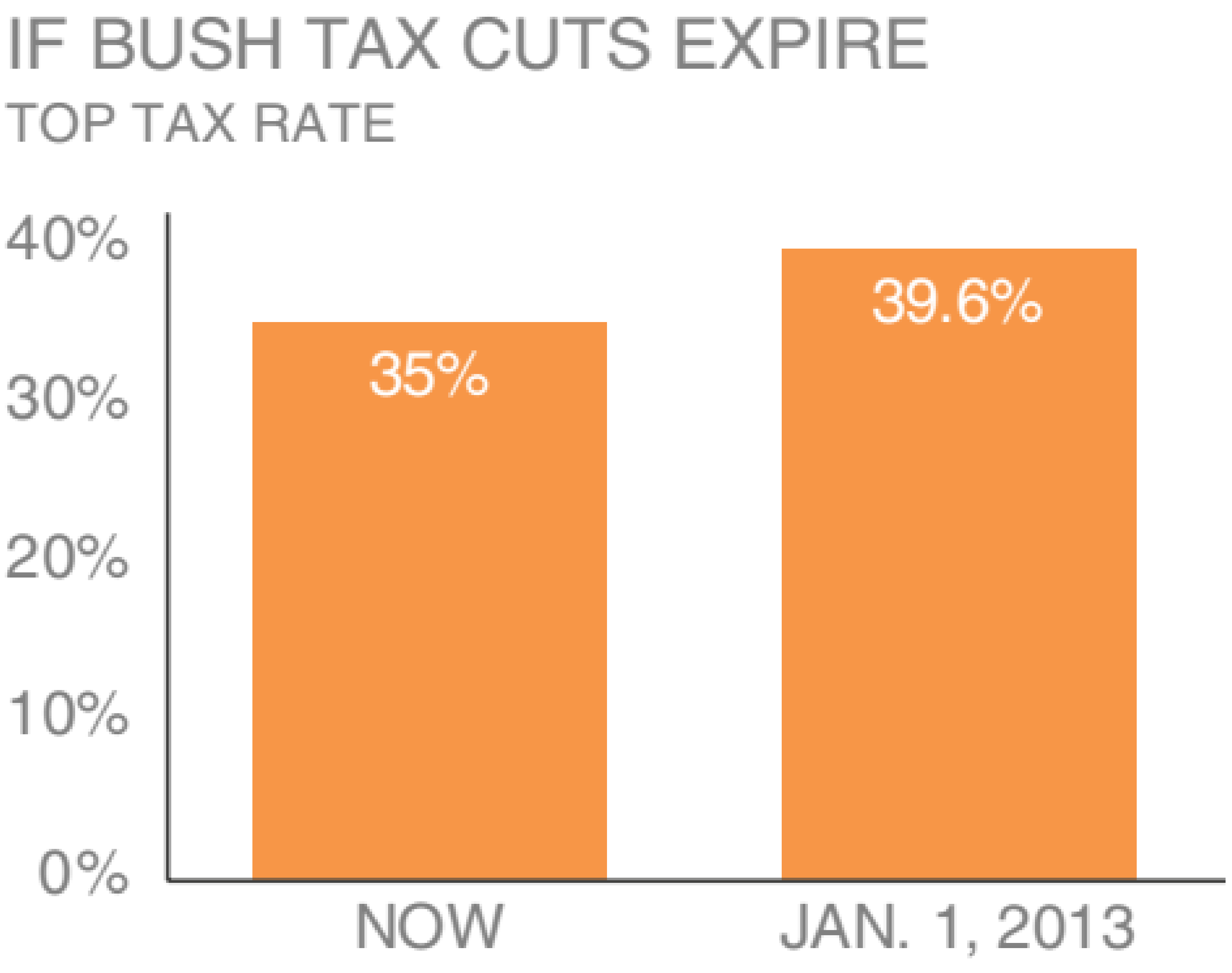


Comparisons of Categories and Time

Non-zero baseline: as originally graphed



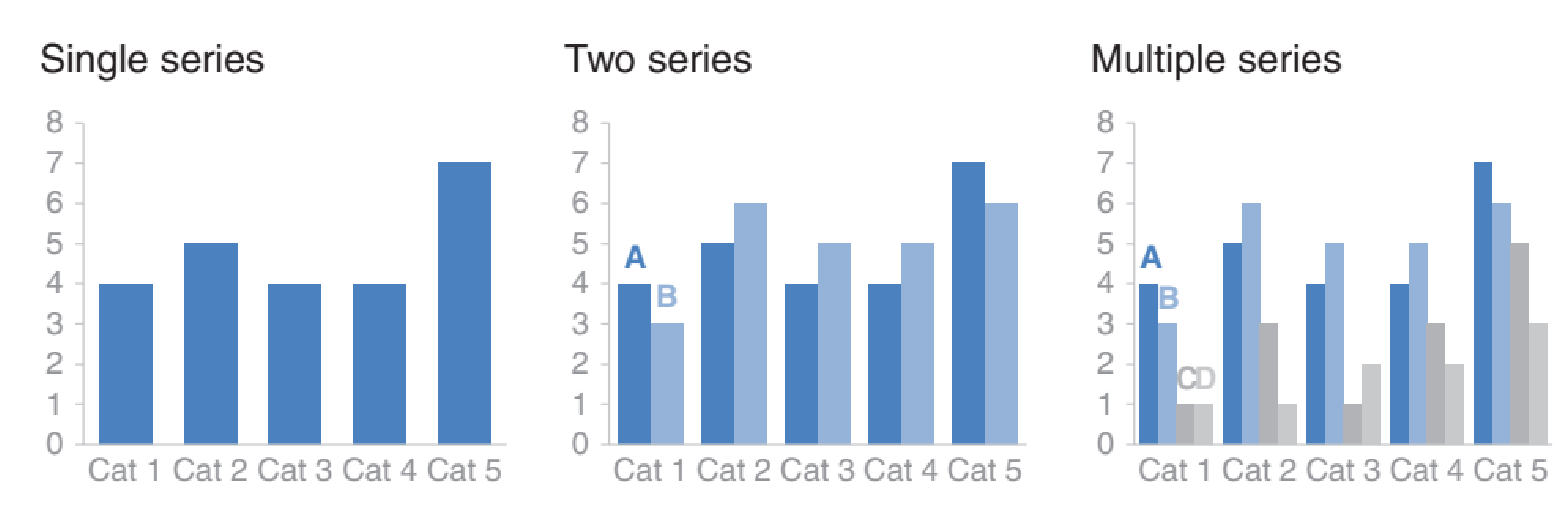
Zero baseline: as it should be graphed



Comparisons of Categories and Time

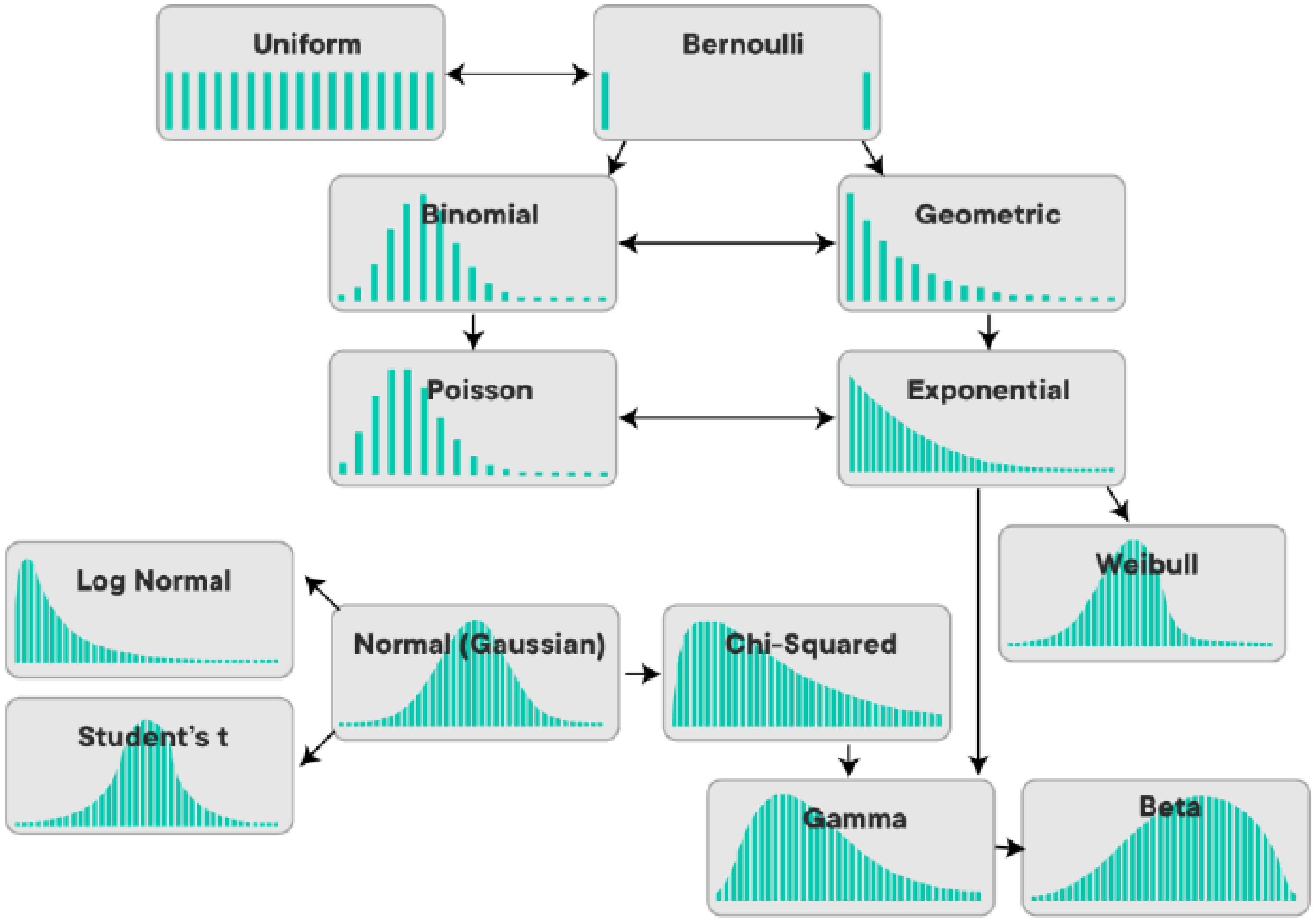
Be aware also that there is visual grouping that happens as a result of the spacing in bar charts having more than one data series. This makes the relative order of the categorization important. Consider what you want your audience

Comparisons of Categories and Time



Be aware also that there is visual grouping that happens as a result of the spacing in bar charts having more than one data series. This makes the relative order of the categorization important. Consider what you want your audience to be able to compare, and structure your categorization hierarchy to make that as easy as possible.

Distributions



Questions:

1. What are the highest, middle, and lowest values?

2. Does one thing stand out from the rest?

3. What does the shape of the data look like?

Insight: use to distributions charts reveal outliers, the shape of the distribution, frequencies, range of values, minimum value, maximum value, and the median.

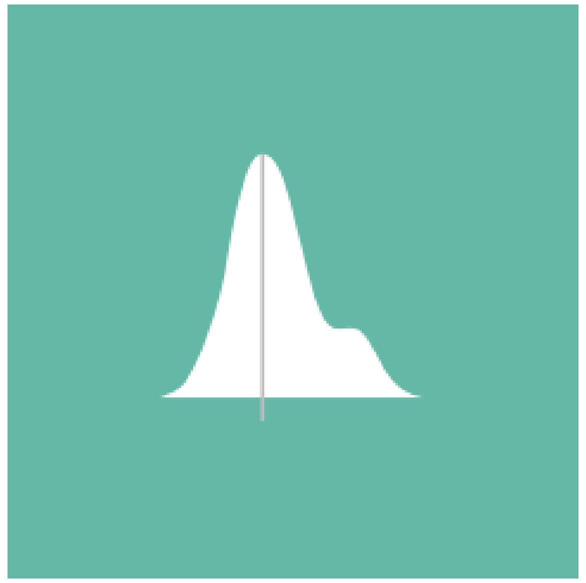
Distributions

Histogram

Histograms show frequencies of a single variable grouped into bins or frequency ranges on the x-axis. The y-axis of the histogram shows the frequency count or percentage.

- A large bin size can obscure the data.
- Adjust the size of the bins to best reveal the shape of the frequency distribution.

Density plot

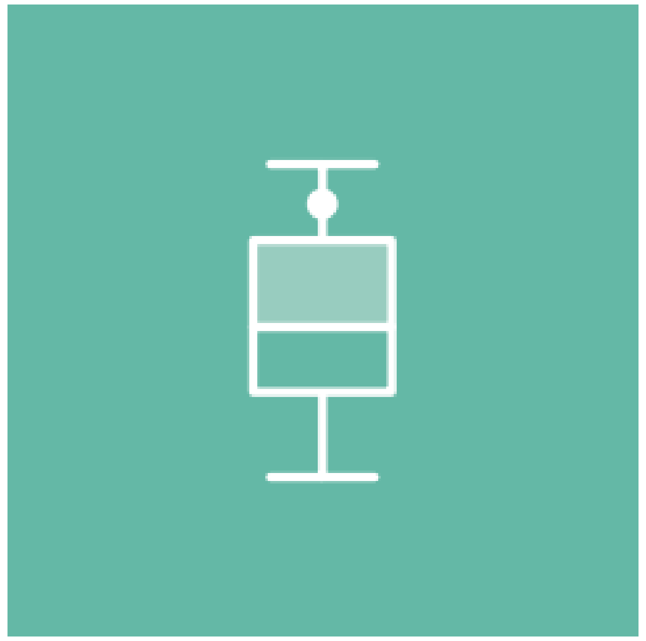


Density plots show probability densities and the distribution of a single variable. The area under the curve emphasizes the shape of the distribution of data.

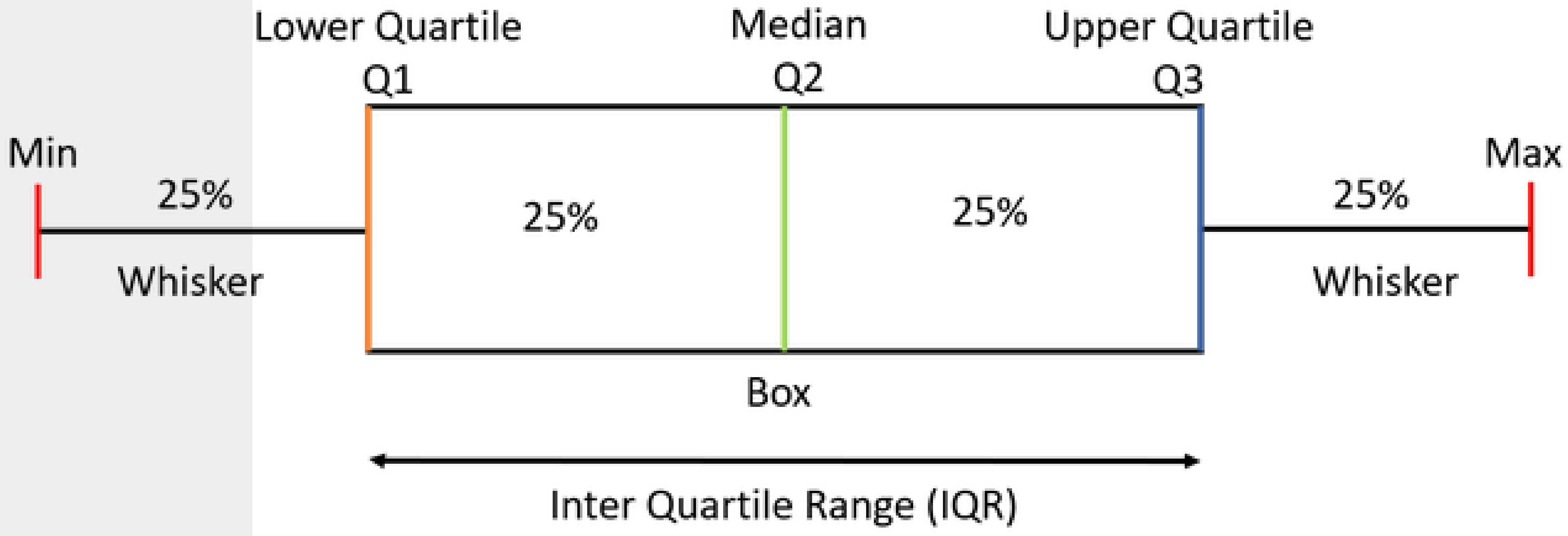
Annotate the mean to draw attention to the center of the distribution.

Distributions

Boxplot



Boxplots show the range of a single variable including the minimum, 25th percentile, 50th percentile, median (not the average), 75th percentile, and the maximum value. Boxplots are helpful to spot outliers.



Proportions

Questions:

1. What are the parts that make up the whole?
2. What part is the largest or smallest?
3. What parts are similar or dissimilar?

Insight: use to show summaries, similarities, anomalies, percentage related to the whole (by category, subcategory, and over time).

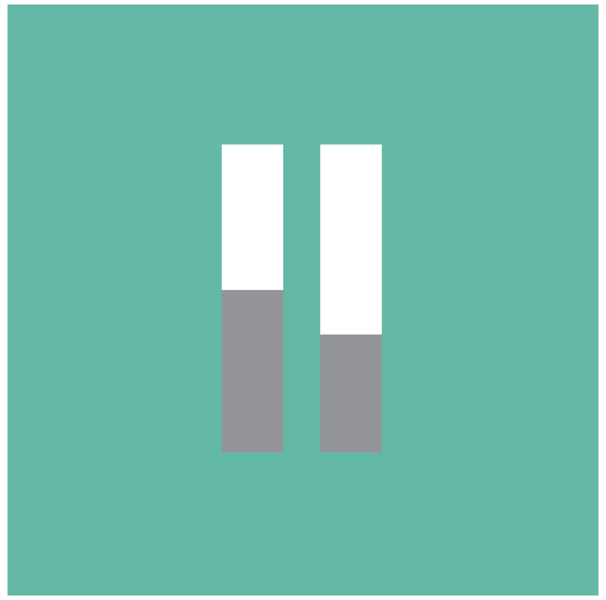
Proportions

Pie

Pie charts show proportions within a whole. The slices are subcategories of a single category. Slices add up to 100% or 1.

- Avoid using pie charts if all the slices are similar in size.
- Limit pie charts to eight slices or less (Wong, 2010).
- Label directly on the pie slices, rather than using a legend.
- Keep pie slices the same color. Use the whitespace between slices to differentiate the slices.

Stacked bar



Stacked bar charts show proportions and quantities within a whole category. They show absolute and relative differences.

- Limit the number of subcategories to four or less.
- Use stacked bars that add up to 100% to show the relative differences between quantities within each group.

Proportions

Tree map

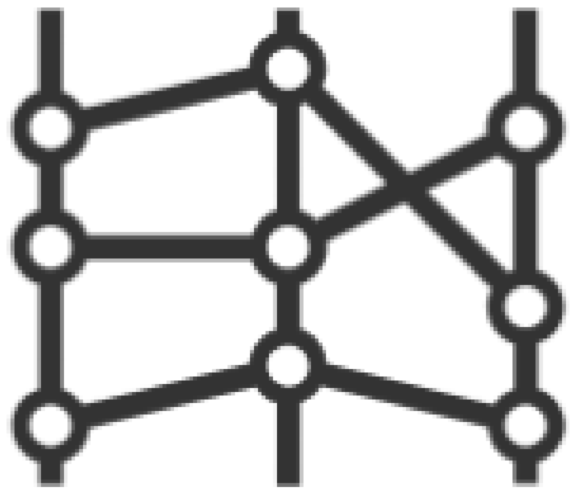
Tree maps show parts of the whole by using nested rectangles. Each rectangle is designated a size and a shade of a color. This enables you to emphasize both the importance (usually shown by size) and urgency (usually represented by color) of a data point.

- Used often for portfolio analysis to highlight similarities and anomalies.
- Usually require interactivity such as mouse-over, to read the subcategory labels for the smallest rectangles.
- This chart type is best used for analysis and exploration rather than presentation.

Doughnut

Doughnut charts present proportions of a whole through slices of a doughnut shaped graphic It is just a pie chart with the center missing. This type of chart can contain multiple series, represented as doughnuts arranged inside one another.

Relationships



Questions:
1. Is the relationship positive, negative, or neither?
2. How are x and y related to each other?
3. What makes one group or cluster different from another?

Insight: use to show outliers, correlations, positive, and negative relationships among two or more variables.

Relationships

Scatterplot

Scatterplots show relationships between two variables. For example, they show the change in x given y.

- Use to show positive or negative correlations, or linear and nonlinear relationships between variables.
- Labeling of every data point reduces readability but increases interpretation.

Scatterplot matrix

Scatterplot matrices help identify a correlation between multiple variables. It makes it easy to observe the relationship between pairs of variables in one set of plots.

This chart type is best reserved for exploration versus presentation.

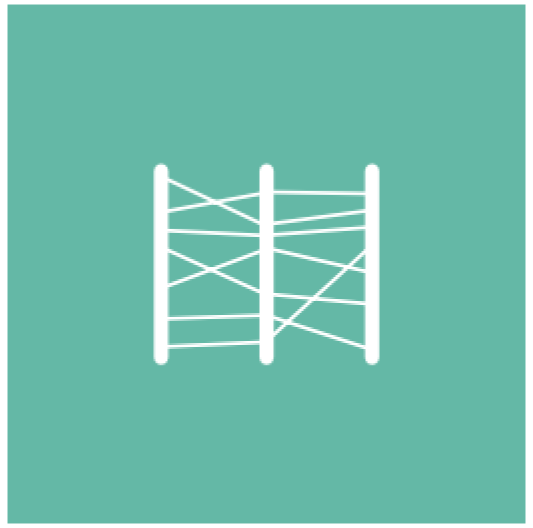
Relationships

Bubble chart

A bubble chart is a scatterplot that shows relationships between three or four variables. The position of the bubble shows the relationship between the x and y variables.

- The bubble size is based upon a numerical variable, such as population, or sales.
- The bubble color is best reserved for categorical data, such as region.
- Bubble charts are best when the bubble sizes vary significantly.

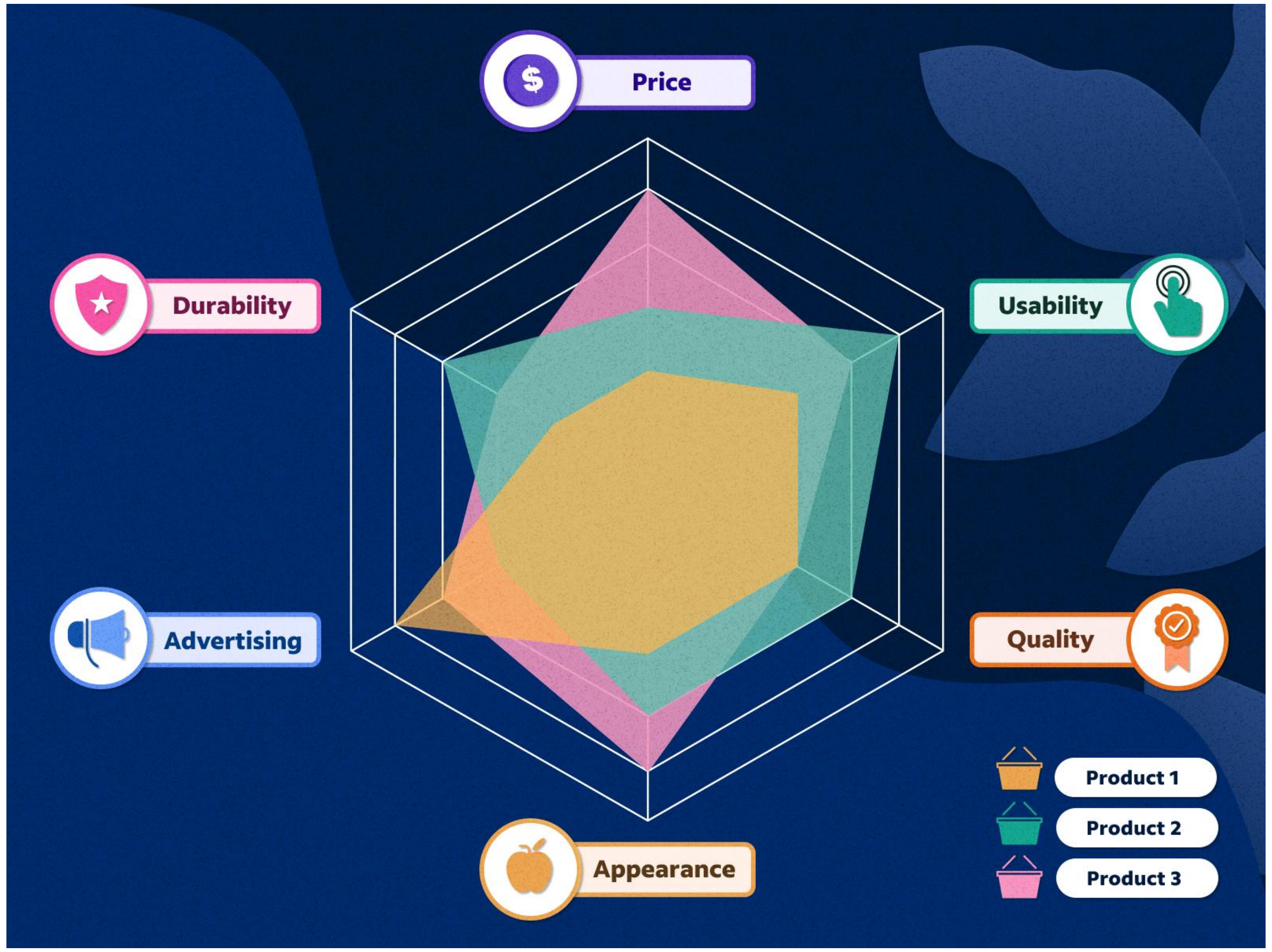
Parallel coordinates



Parallel coordinates map each column in a data table as a vertical parallel line with its own axis. Each observation (row) is represented by a point on the parallel line. That point is then connected to the next point on the next parallel line by a horizontal line.

- Use the technique of highlighting the lines that touch any number of values in either of the categories, called brushing, to provide data context while focusing on select series.
- This chart type is best reserved for exploration over presentation.

Relationships



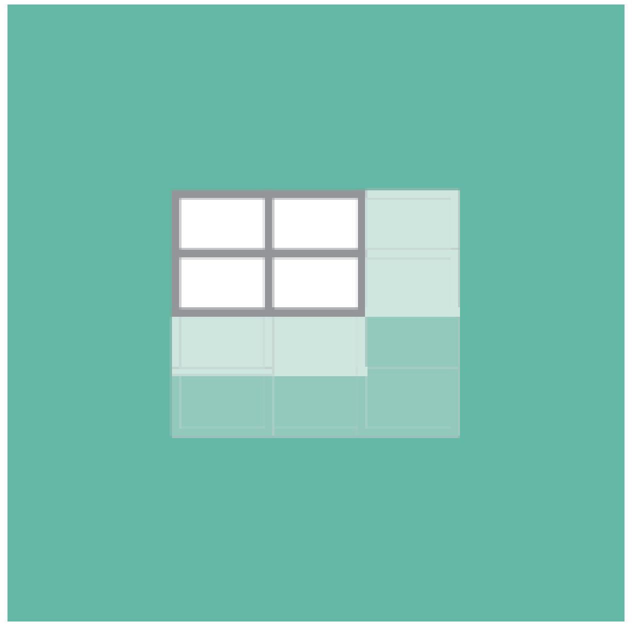
Radar

Radar charts compare multiple numerical variables. They show which variables have similar values, and to spot outliers, high values, and low values. Each variable is provided its own individual axis, but the axes are arranged radially. Every observation connects to form a shaded polygon.

- Limit the number of variables to reduce the number of axes to increase readability.
- Scaling is affected when variables have dissimilar minimum and maximum ranges.

Relationships

Heat map



A heat map is a graphical representation of a table of data. The individual values are arranged in a table/ matrix and represented by colors. Use grayscale or gradient for coloring. Sorting of the variables changes the color pattern.

Table

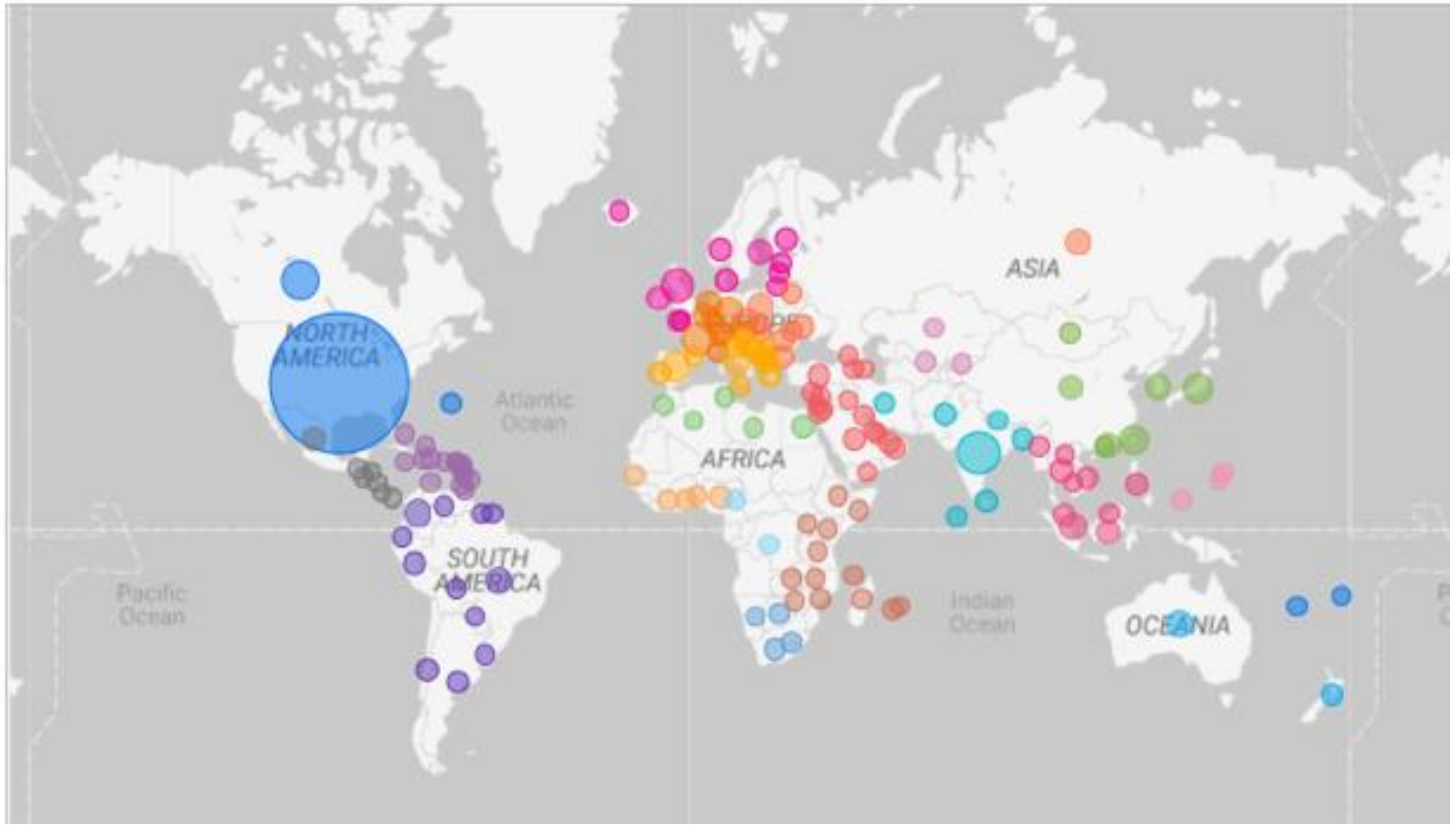
	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Heatmap

LOW-**HIGH**

	A	B	C
Category 1	15%	22%	42%
Category 2	40%	36%	20%
Category 3	35%	17%	34%
Category 4	30%	29%	26%
Category 5	55%	30%	58%
Category 6	11%	25%	49%

Locations



Questions:
1. Where can the most or least be found?
2. How does one area compare to another?
3. What is the distance from one place to another?
4. How does a variable change by location?

Insight: use to demonstrate similarities and differences by location, density, distance, and counts (such as population).

Locations

Point map

Point maps show a specific location. These dots can vary in size, form, or color.

Point maps illustrate density when the individual locations are easily distinguishable. Too many points can obscure the location. Consider the size of the points and the labeling of the points.

Symbol or bubble map

Symbol maps are point maps that use different sized bubbles or shapes to mark a location. These symbols are sized by a certain variable.

Too many or too large bubbles can obscure the locations referenced.

Locations

Connection or path maps

Connection maps graph a line from one or more points to another. Use to show distances or pathways between one or more locations.

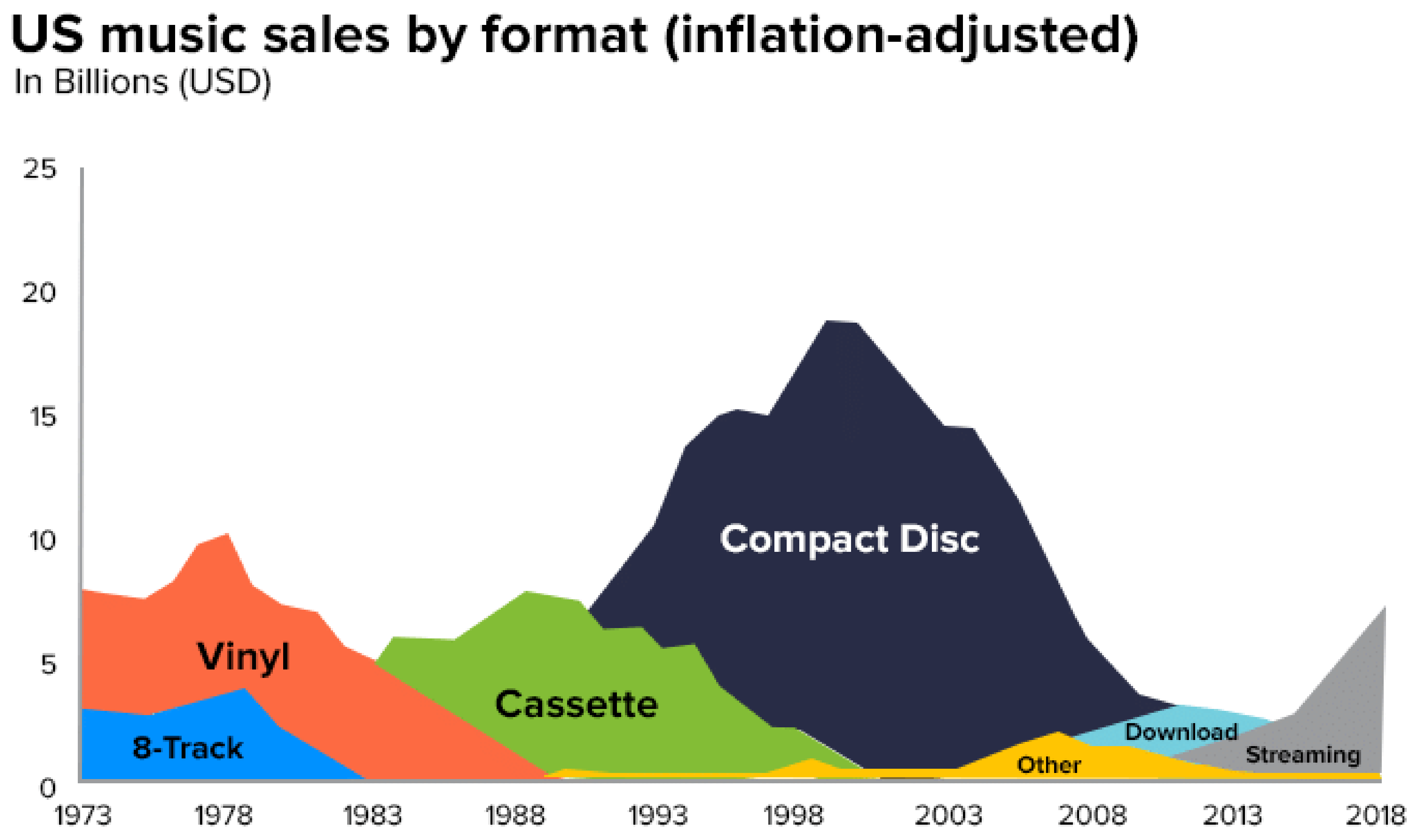
Use high contrasting colors for the map projection and the lines that connect the points. Avoid too many overlapping lines.

Geographic heat map (Isopleth)

Isopleth maps show gradual change over geography. This technique uses a color value (lightness/darkness) and hue to show density. The color value is not constrained by boundary lines (e.g., such as zip code).

Use for events that are continuous and unbounded (e.g., such as temperature).

Trends—Showing Comparisons Over Time or Composition Over Time



Questions:
- What changed today from yesterday?
- How does time of year affect sales, results, outcomes, etc.?
- What times are the most popular? Least popular?

Insight: change over time, cycles, or comparisons over time.

Trends—Showing Comparisons Over Time or Composition Over Time

Line chart

Line charts show the change over time for one or more series (sales per hour). The line connects each data point in the series (shown or not). The y-axis baseline should be equal or less than the minimum value in the data.

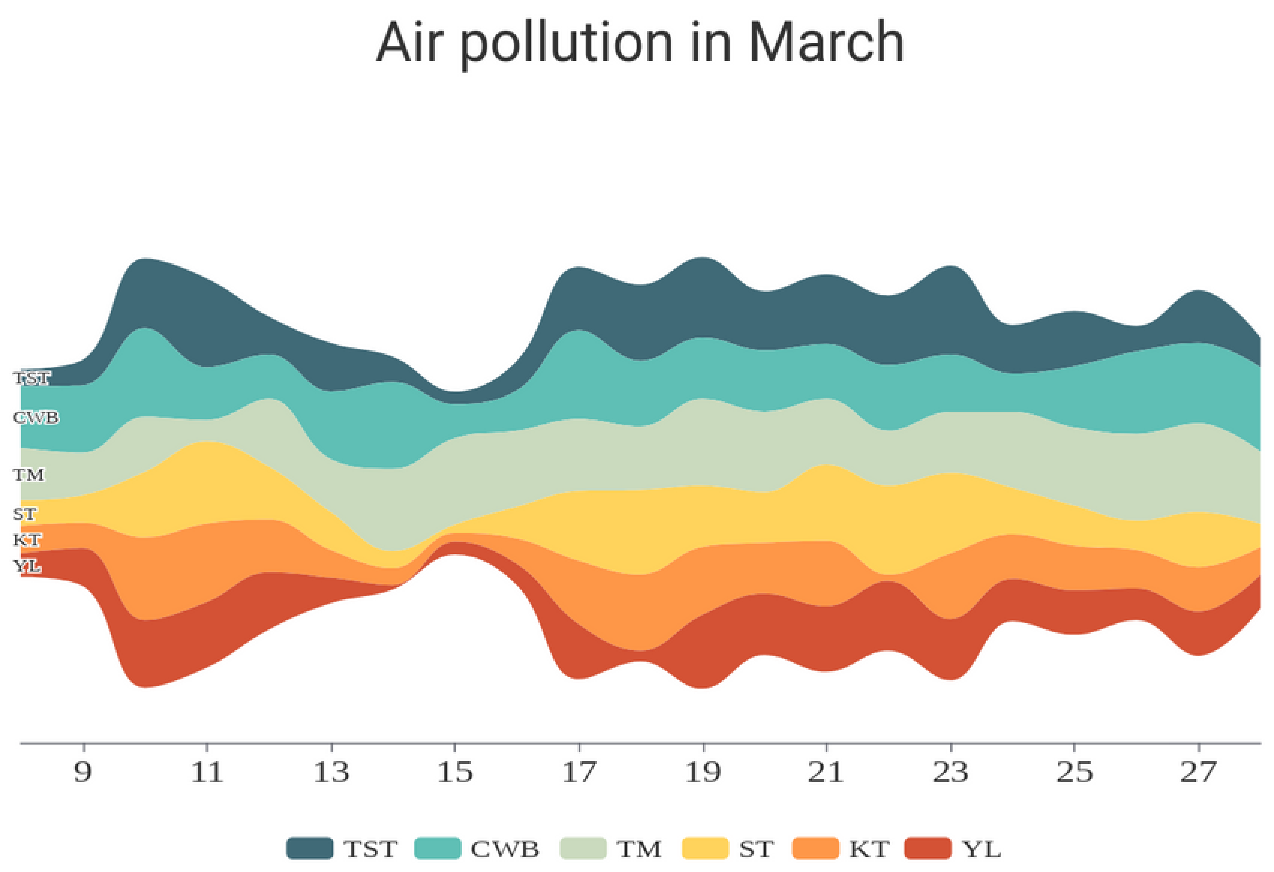
- Show four or fewer series of lines on a line chart (Wong, 2010).
- Label each series directly or use an ordered legend.

Sparkline

A sparkline is a line chart without axes or much detail. It is a small graphic designed to give a quick representation of change over time.

- Not intended to provide the quantitative precision of a normal line graph.
- Label the last data point to provide additional information.

Trends—Showing Comparisons Over Time or Composition Over Time



Stream graph

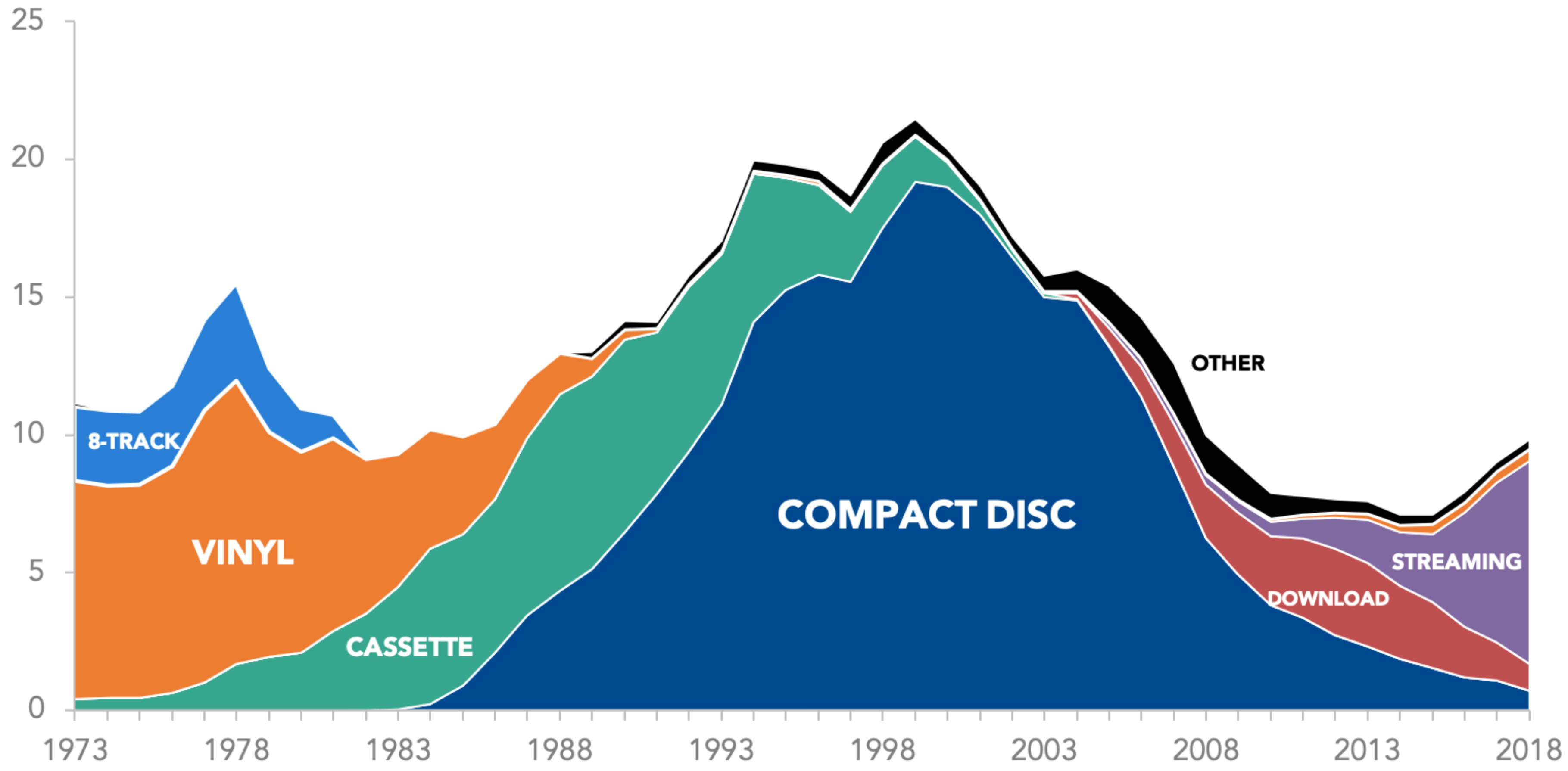
Stream graphs show changes over time for different data series. Color is used to distinguish the categories. Each stream represents a single category proportional change over time. Stream graphs are used to provide a general overview, not when accuracy is important.

Use for large time series data sets with five or fewer categories.

Trends—Showing Comparisons Over Time or Composition Over Time

US music sales by format (inflation-adjusted)

IN BILLIONS (USD)



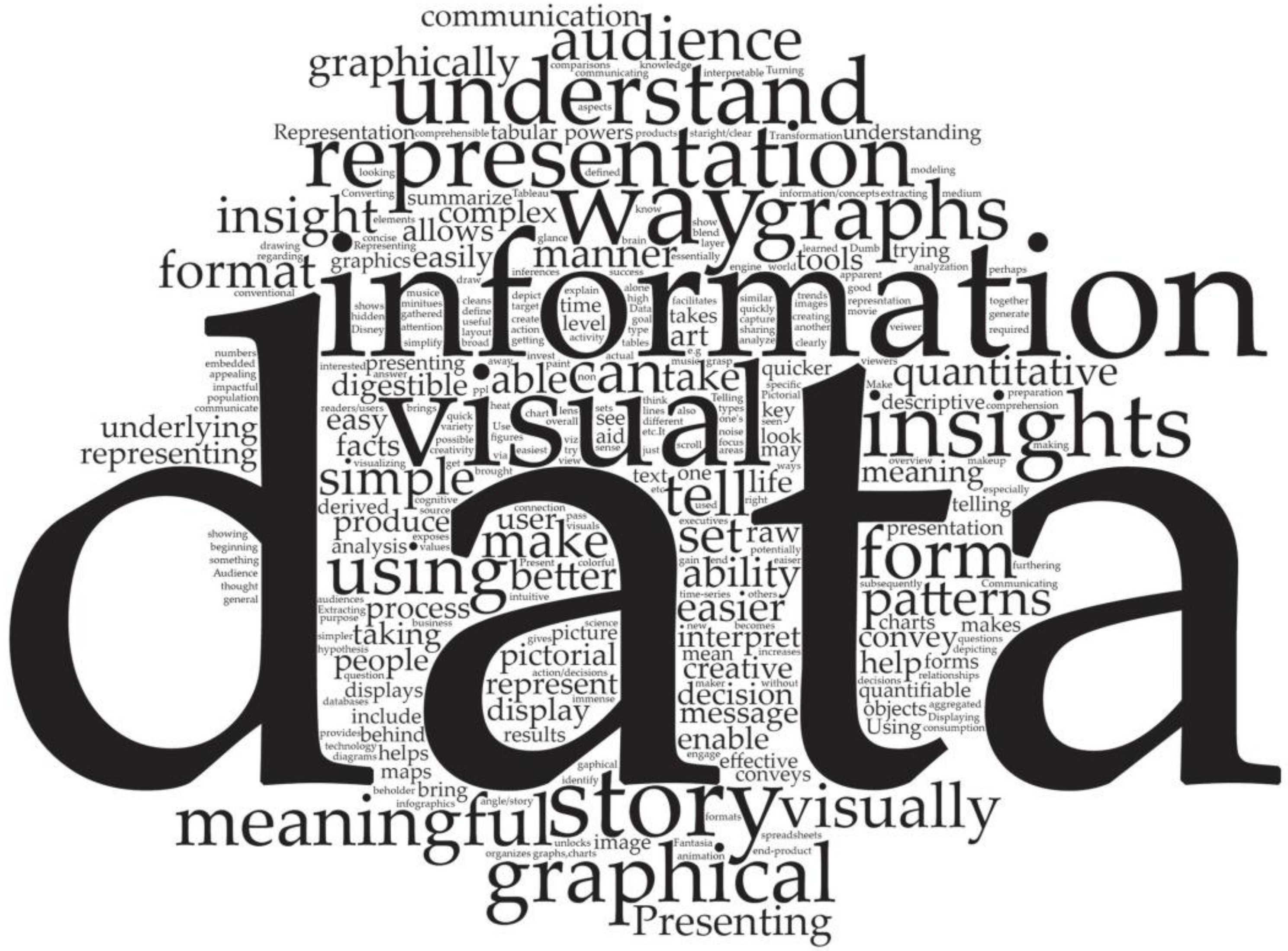
SOURCE: Recording Industry Association of America

Area graph

Area graphs are line charts with the area below the line filled in with color. They can show a single series or multiple time series using stacked areas.

Use the same color for the line and the area beneath it. See Table 3.4 for use of stacked areas to show proportional change over time.

Word Frequency and Sentiment



Questions:

- How many times does a given word or phrase appear?
- What words or phrases appear most often? Least often?
- What words appear together?
- Are most words or phrases positive or negative?

Insight: frequency or counts of words and phrases. The count of the positive or negative direction of the sentiment of the words or phrases

Word Frequency and Sentiment

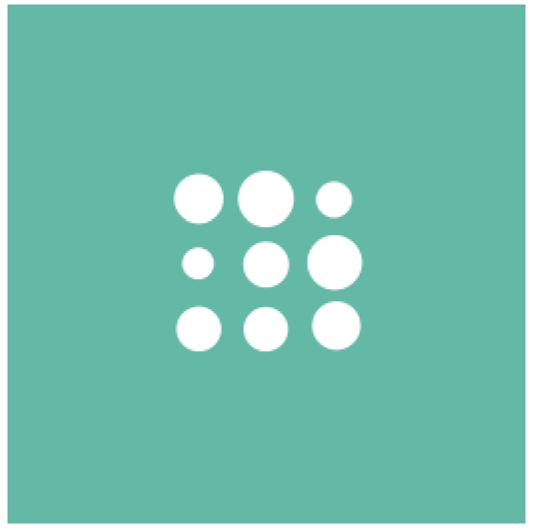
Word cloud

Words are arranged in a cluster or cloud of words. Words can be arranged in any format: horizontal lines, columns, or within a shape.

Color is used to categorize words by sentiment, or another categorical variable.

Word Frequency and Sentiment

Proportional bubble area chart



Words are ranked by their frequency. The frequency is represented by sized bubbles or squares. The bubbles /squares are arranged in a grid with words on the x-axis and observation on y.

Works well for the top 10 words (difficult to view beyond that).

Simple text

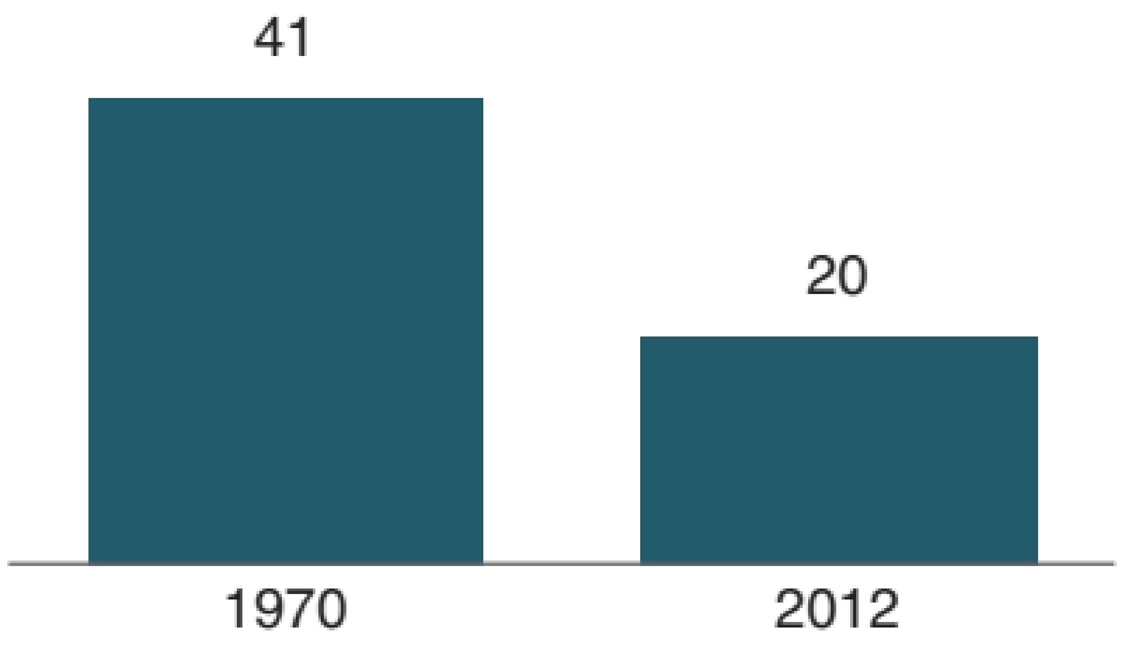
20%

of children had a **traditional stay-at-home mom** in 2012, compared to 41% in 1970

When you have just a number or two to share, simple text can be a great way to communicate. Think about solely using the number— making it as prominent as possible—and a few supporting words to clearly make your point.

Children with a "Traditional" Stay-at-Home Mother

% of children with a married stay-at-home mother with a working husband



Note: Based on children younger than 18. Their mothers are categorized based on employment status in 1970 and 2012.

Tables

Heavy borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Light borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

Minimal borders

Group	Metric A	Metric B	Metric C
Group 1	$X.X	Y%	Z,ZZZ
Group 2	$X.X	Y%	Z,ZZZ
Group 3	$X.X	Y%	Z,ZZZ
Group 4	$X.X	Y%	Z,ZZZ
Group 5	$X.X	Y%	Z,ZZZ

- ❑ **Tables interact with our verbal system, which means that we read them.**
- ❑ **Tables are great for just that communicating to a mixed audience whose members will each look for their particular row of interest. If you need to communicate multiple different units of measure, this is also typically easier with a table than a graph.**

Borders should be used to improve the legibility of your table. Think about pushing them to the background by making them grey, or getting rid of them altogether. The data should be what stands out, not the borders.

ANY QUESTIONS?

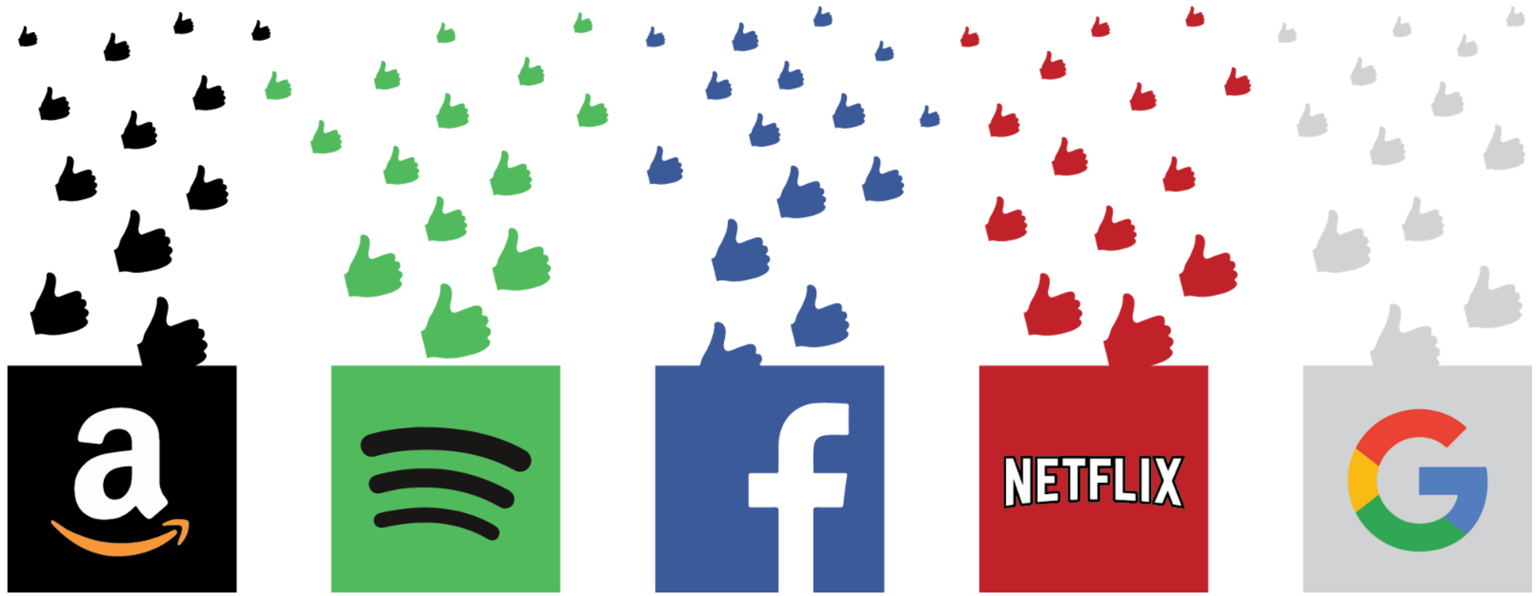
4. Recommendation Systems

Khalil Israfilzada, PhD

Faculty of Economics and Management
Vytautas Magnus University
Kaunas, 2025

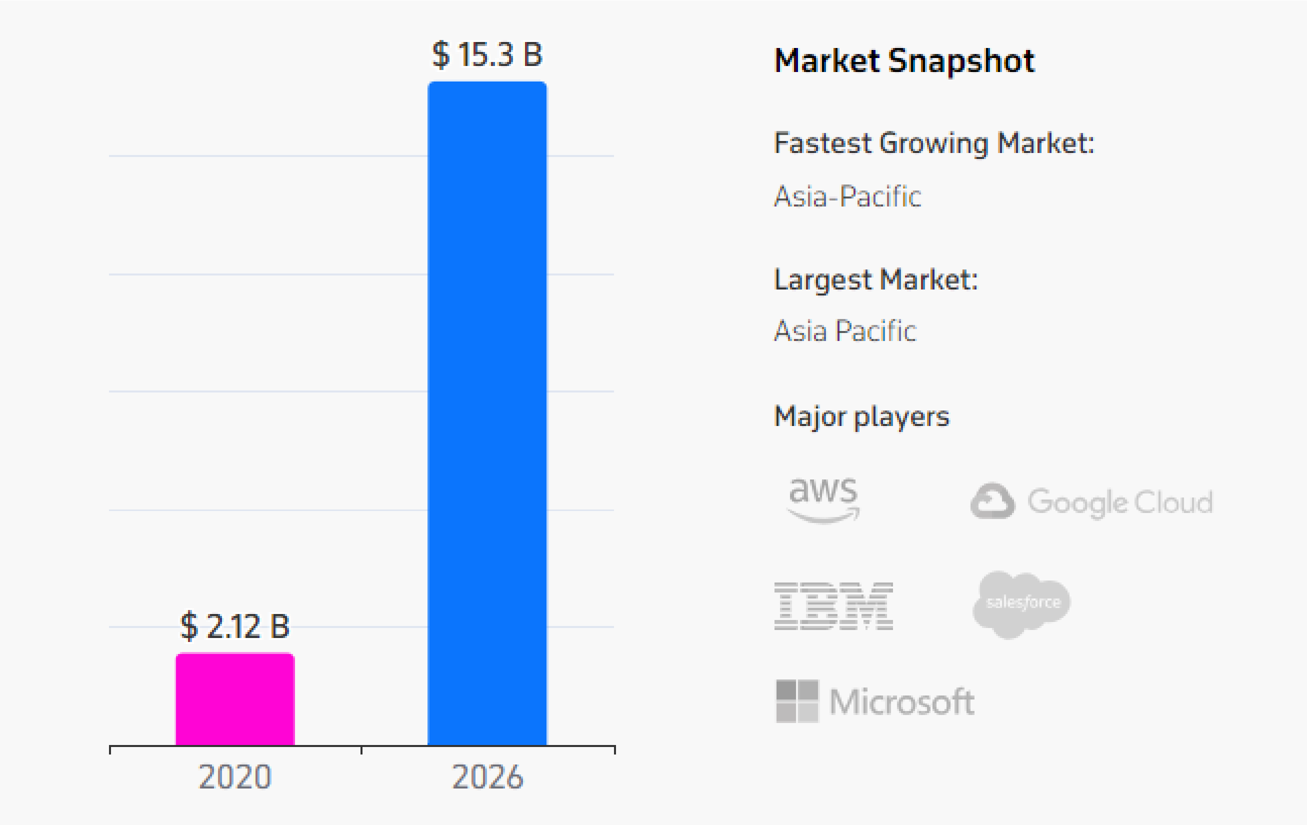
Introduction

Definition. Recommender systems are advanced tools developed with the aim of suggesting items or content relevant to users. They leverage powerful algorithms to enhance the online experience on a wide range of platforms, from shopping websites and movie streaming services to social networking sites.



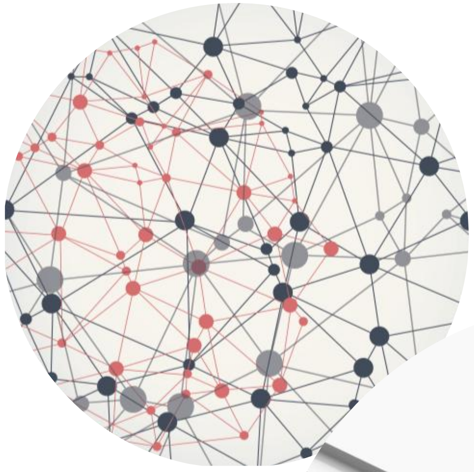
Introduction

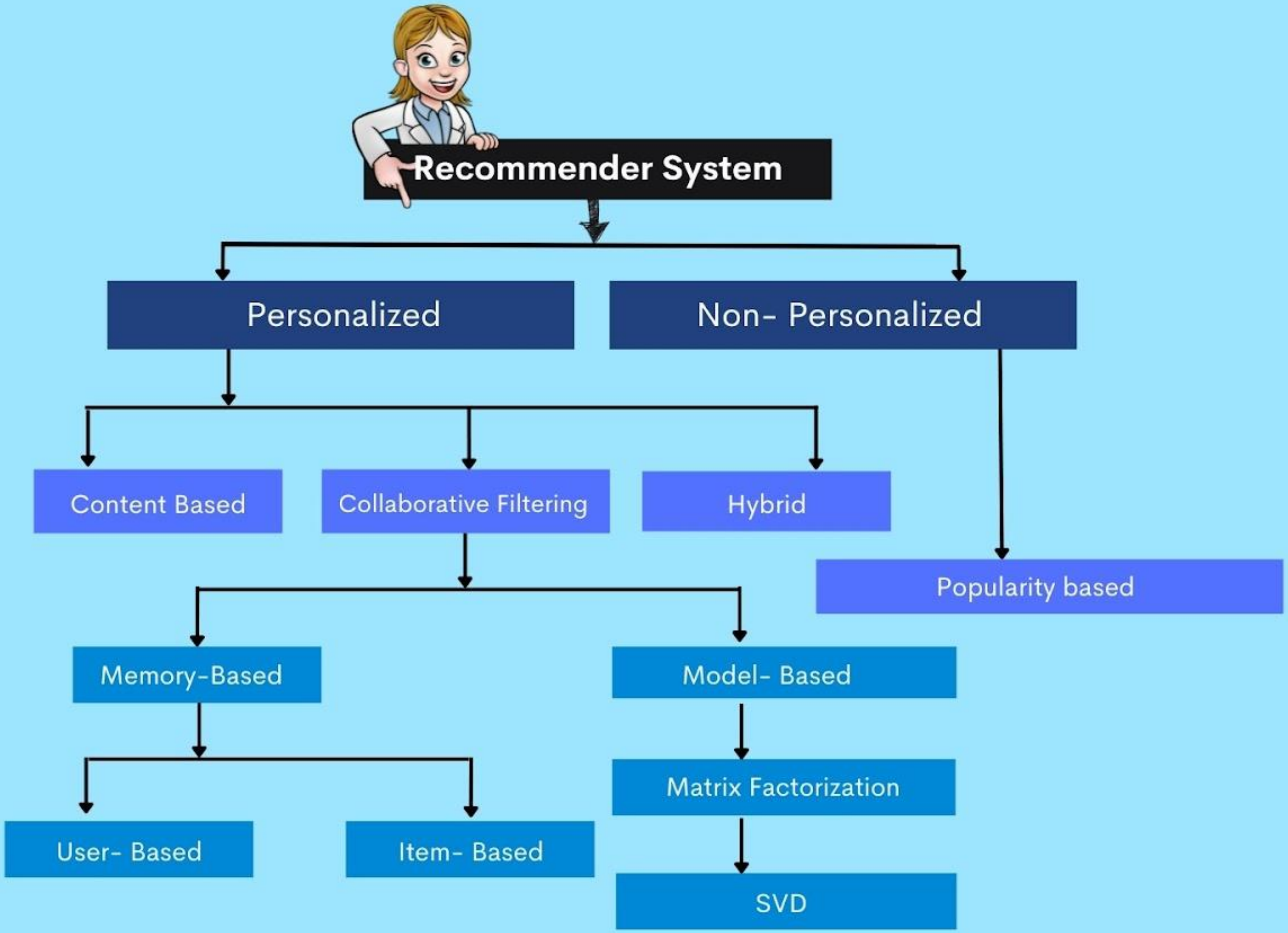
The global recommendation engine market size was valued at USD 2.12 billion in 2020 and is expected to expand at a compound annual growth rate (CAGR) of 33.0% from 2021 to 2026. The increasing need to enhance customer experience is fuelling the demand for recommendation engines.



How Do They Enhance Marketing?

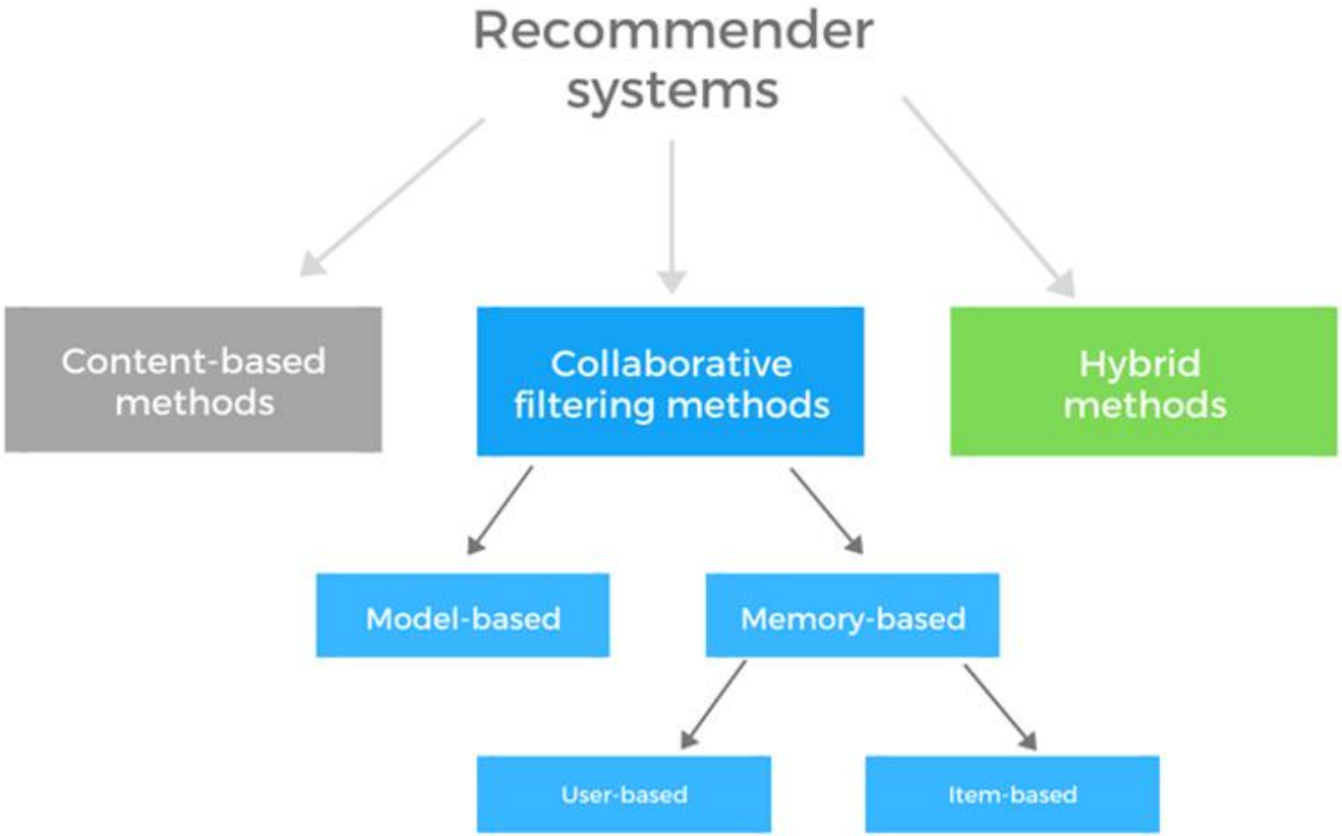
1. Gathering Data
2. Data Management
3. Algorithm Selection
4. Crafting User Profiles
5. Item Profiling
6. Generating Recommendations
7. Final Touches - Ranking and Showcasing



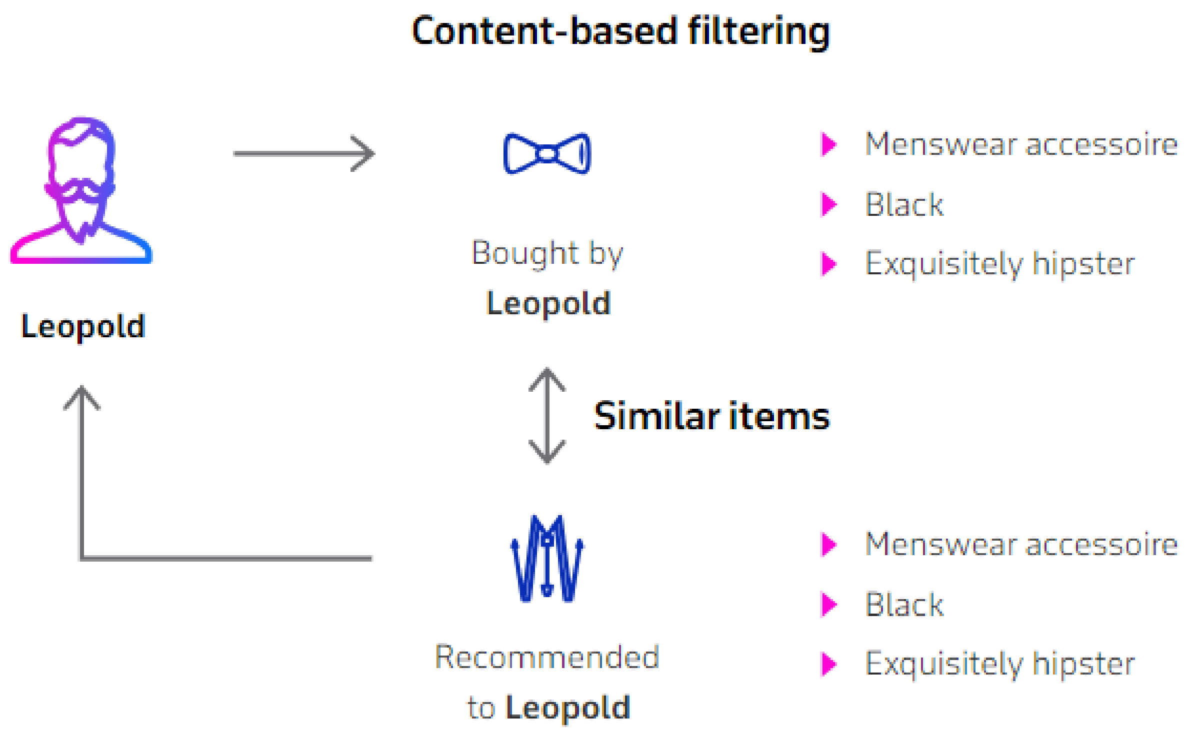

Recommender System
Personalized
Non- Personalized
Content Based
Collaborative Filtering
Hybrid
Popularity based
Memory-Based
Model- Based
Matrix Factorization
User- Based
Item- Based
SVD

PERSONALIZED RECOMMENDATION SYSTEMS

Personalized recommendation systems are designed to provide tailored recommendations to individual users based on their past behavior, preferences, and demographic information. Based on the user's data such as purchases or ratings, personalized recommenders try to understand and predict what items or content a specific user is likely to be interested in.



CONTENT-BASED FILTERING



Content-based filtering is a technique used in recommendation systems where items are matched with users based on the features of the items and a profile of the user's preferences. The idea is to recommend new items to users by comparing the content of the items and a user's profile, with content being described in terms of several descriptors or terms that are inherent to the item.

CONTENT-BASED FILTERING

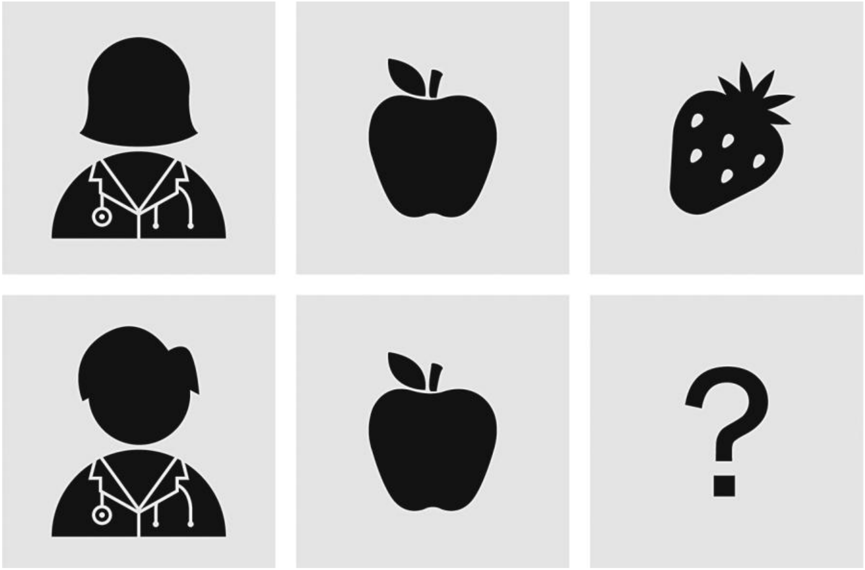
Pros of the content-based approach. The content-based approach is one of the common techniques used in personalized recommendation systems.

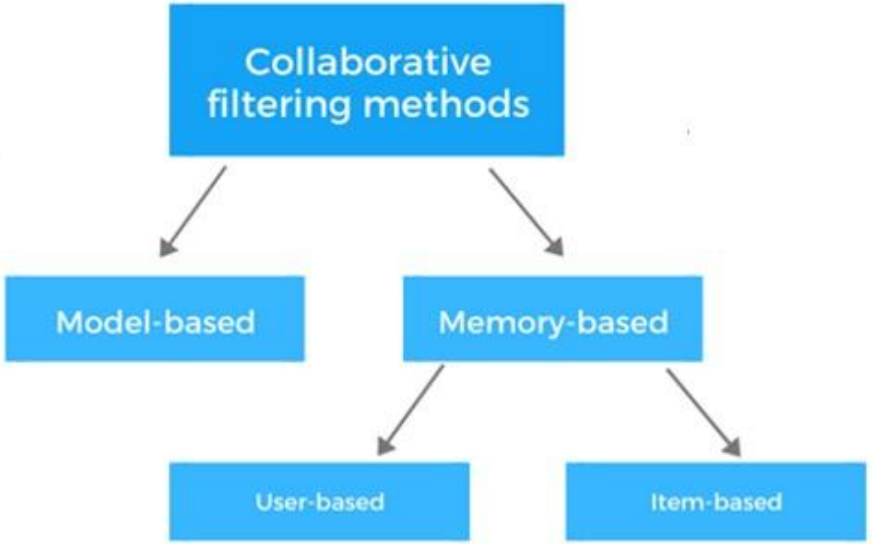
- **Less cold-start problem**
- **Transparency**
- **Diversity**
- **Reduced data privacy concerns**

Cons of the content-based approach. On the other hand, the content-based approach can come with a few disadvantages, too. These can include:

- **The "Filter bubble"**
- **Limited serendipity**
- **Over-specialization**

COLLABORATIVE FILTERING



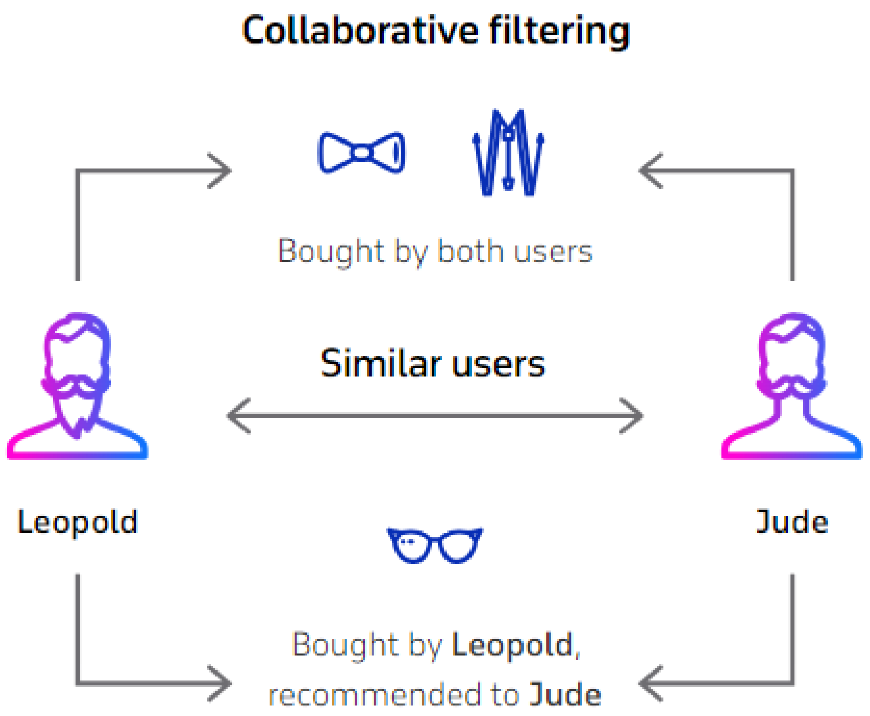


Collaborative filtering is akin to asking a group of people for book recommendations, assuming those who have similar reading tastes to yours in the past will likely suggest books you'll enjoy in the future. In the digital realm, this method recommends items by analyzing patterns of user behavior and leveraging the choices and preferences of many users.

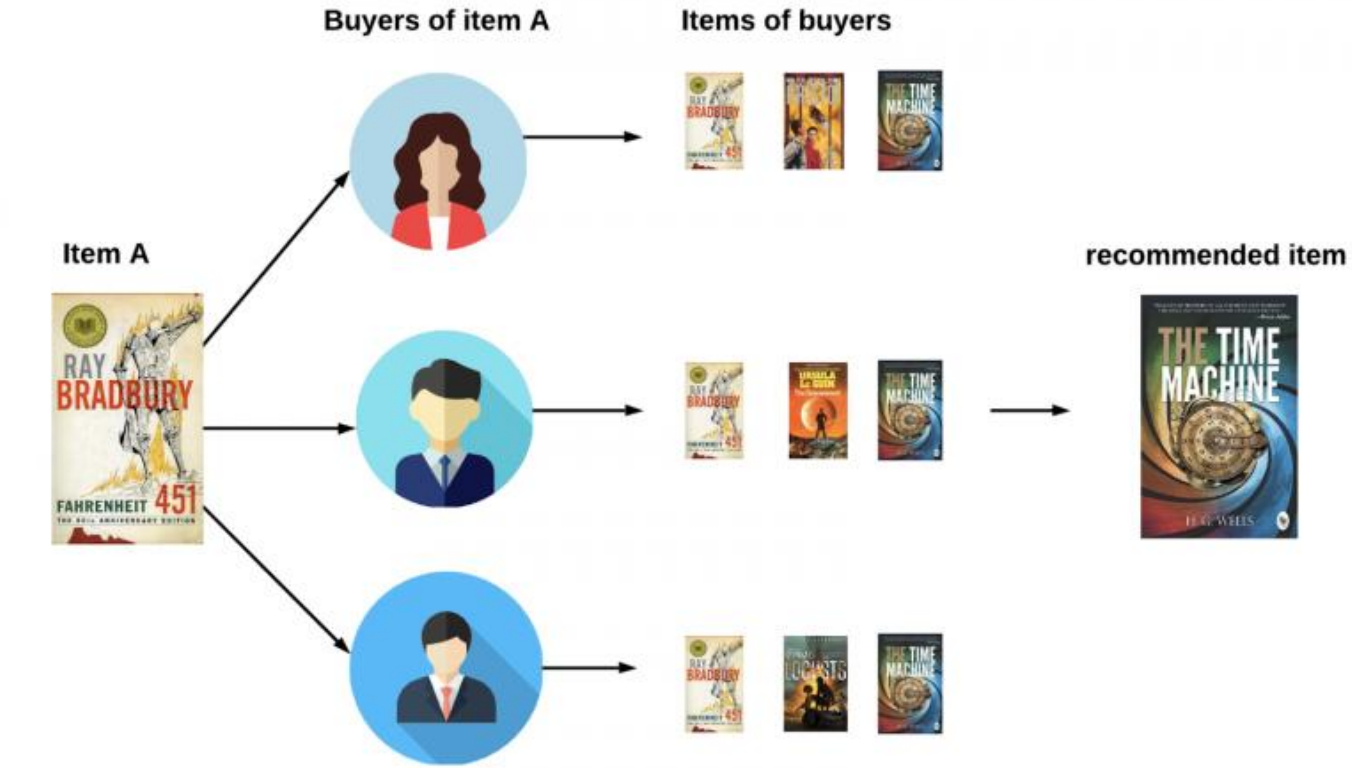
COLLABORATIVE FILTERING

1. Memory-Based Recommenders. Memory-based recommenders rely on the direct similarity between users or items to make recommendations. Usually, these systems use raw, historical user interaction data, such as user-item ratings or purchase histories, to identify similarities between users or items and generate personalized recommendations. Memory-based recommenders can be categorized into two main types **user-based** and **item-based** collaborative filtering.

User-Based Collaborative Filtering.
Recommendations are crafted by identifying users who have shown tastes similar to the target user.



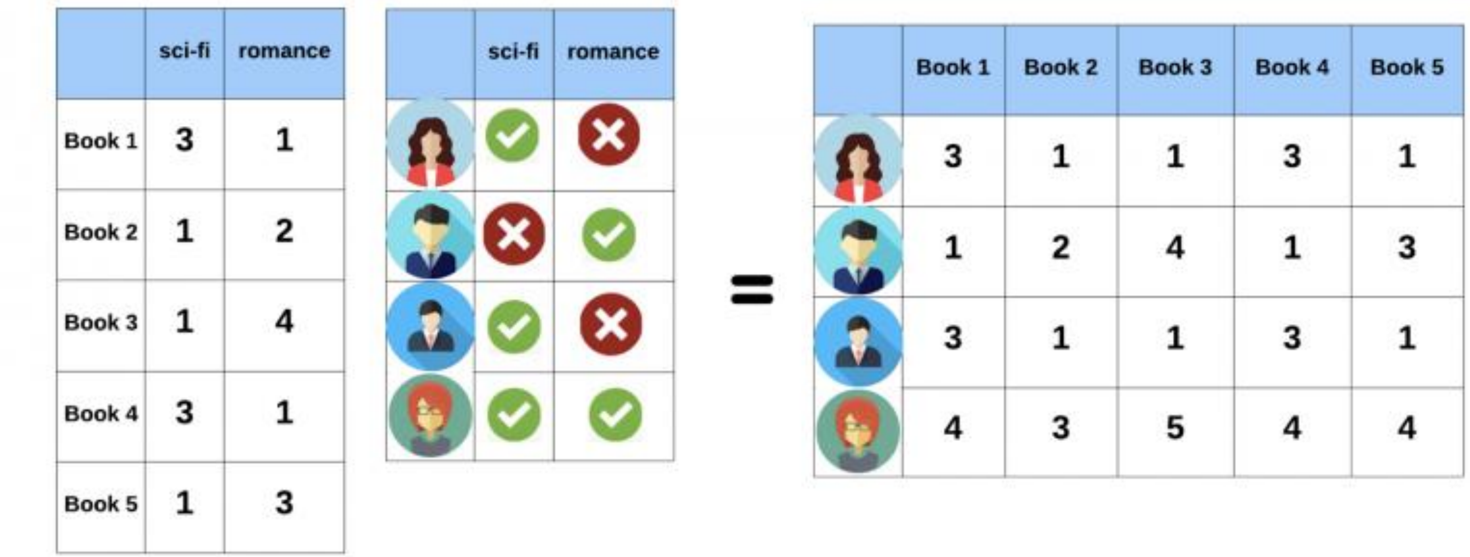
COLLABORATIVE FILTERING



Item-Based Collaborative Filtering. The focus here is on item similarities based on user interactions. In item-based collaborative filtering, recommendations are made by identifying items that are similar to the ones the target user has already interacted with.

The idea is to find items that share similar user interactions and recommend those items to the target user. This can include "users who liked this item also liked…" type of recommendations.

COLLABORATIVE FILTERING



2. Model-Based Recommenders. **What are Model-Based Recommenders?** Model-based recommenders are a subset of recommendation systems that use machine learning to identify patterns in user-item interactions. They're designed to discern correlations and behaviors from historical data to predict future preferences. There are several methodologies employed in model-based recommenders. These include **matrix factorization, Singular Value Decomposition (SVD), neural networks**, and more.

COLLABORATIVE FILTERING

Collaborative Filtering, Pros and Cons. Let's dissect its pros and cons from a marketer's standpoint.

Pros of Collaborative Filtering:

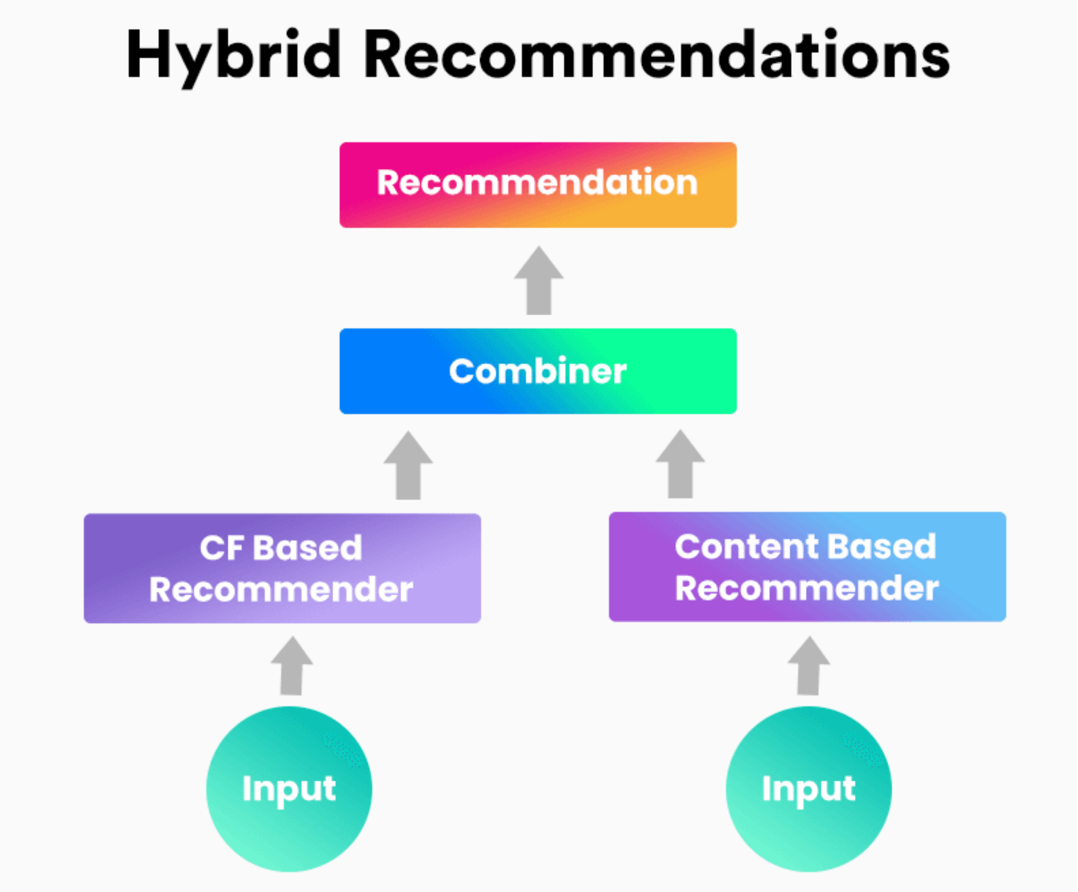
1. **Effective Personalization**
2. **Independence from Item Attributes**
3. **Serendipitous Discoveries**

Cons of Collaborative Filtering:

1. **The "Cold-Start" Problem**
2. **Sensitivity to Sparse Data**
3. **Potential for Popularity Bias**

HYBRID RECOMMENDERS

Hybrid recommendation systems combine multiple recommendation techniques or approaches to provide more accurate, diverse, and effective personalized recommendations. They are particularly valuable in **real-world recommendation scenarios** because they can provide more robust, accurate, and adaptable recommendations. The choice of which hybrid approach to use depends on the specific requirements and constraints of the recommendation system and the nature of the available data.



HYBRID RECOMMENDERS

Advantages of Hybrid Recommenders:

1. Enhanced Recommendation Quality
2. Adaptability & Flexibility
3. Overcoming Recommendation Limitations.

Challenges with Hybrid Recommenders:

1. Increased Complexity
2. Resource Intensiveness
3. Tuning Sensitivity

NON-PERSONALIZED RECOMMENDATION SYSTEMS

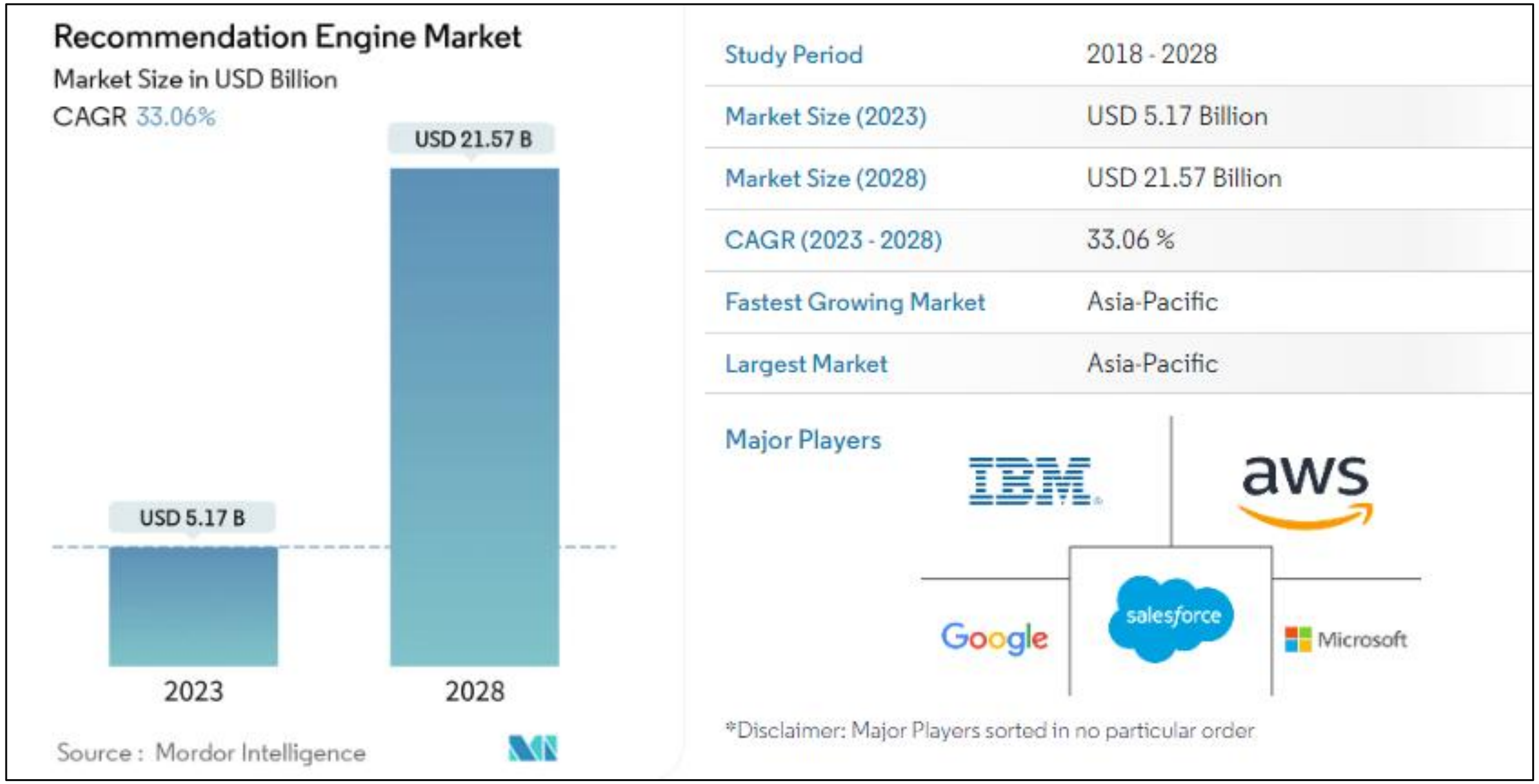
Non-personalized recommendation systems provide recommendations to users without taking into account their individual preferences or behavior. These systems make recommendations based on the characteristics of items or content themselves rather than relying on user-specific data.

A popular non-personalized recommender is the popularity-based recommender which recommends the most popular items to the users, for instance:

- Top-10 movies,
- Top 5 trending products,
- New products.

Global Recommendation Engine Market Overview

The Recommendation Engine Market size is expected to grow from USD 5.17 billion in 2023 to USD 21.57 billion by 2028, at a CAGR of 33.06% during the forecast period (2023-2028).



Global Recommendation Engine Market Overview

Recommendation Engine Market Trends. Increasing Demand for Customization of Digital Commerce Experience Across Mobile and Web Drives the Market's Growth.

- **Digital Shift:** Personalized online experiences differentiate brands.
- **Global Comparisons:** Consumers weigh global online offerings, not just local.
- **AI & ML Edge:** Recommendation engines give businesses a competitive advantage.
- **Omnichannel Importance:** Strong strategies lead to growth and high retention.
- **Unified Branding:** Consistency across digital channels is crucial.

Any questions?