Selver Softic

Data visualization

Fundamentals and Applications

Selver Softic

DATA VISUALIZATION
Fundamentals and Applications

CITATION

This book may be cited in academic and professional work using the following recommended citation format:

Softic, S. (2026). Data Visualization: Fundamentals and Applications. ----, ----. ISBN: [Insert ISBN].

If citing a specific chapter or figure, please reference the chapter number or figure title accordingly.

All citations of external sources used within this book are provided according to APA 7th Edition standards.

INSTRUCTIONS FOR USE

This book is intended for educational, academic, and professional purposes.
Readers may:
- Use the material for personal study or teaching
- Quote short passages with proper citation
- Refer to diagrams and examples in academic work, as long as attribution is provided
Readers may not:
- Reproduce the book in whole or in part without permission
- Distribute digital or printed copies
- Modify or republish the content without express written consent from the author

For permission requests, please contact the author directly.

NOTICE OF RESPONSIBILITY

The author is solely responsible for the content of this book. All analyses, interpretations, and conclusions expressed herein reflect the author's professional judgement and do not necessarily represent the views of any institution or organization.

All sources have been cited with due diligence. Any errors or omissions remain the responsibility of the author.

GRAPHICS / IMAGE RIGHTS NOTES

John Snow's "Cholera Map of Soho" (1854) is in the public domain. No permission is required for reproduction.

Florence Nightingale's "Diagram of the Causes of Mortality in the Army in the East" (1858) is in the public domain. No permission is required for reproduction.

Charles Joseph Minard's "Carte figurative et analytique de la campagne de Russie 1812–1813" (1869) is in the public domain. No permission is required for reproduction.

Selver Softic

DATA VISUALIZATION

Fundamentals and Applications

Scientific reviewers:

Assoc. Prof. Dr. Cristina Fleşeriu
Babeş-Bolyai University, Cluj-Napoca

Assoc. Prof. Dr. Vlad Bocăneţ
Technical University of Cluj-Napoca

Data Visualization
Fundamentals and Applications
FH-Prof. DI Dr. techn. Selver Softic, BSc
CAMPUS 02 University of Applied Science
Graz, Austria

Contents

Preface

This book provides a comprehensive introduction to the fundamentals, methods, and applications of data visualization. It is based on carefully selected materials and case studies to offer readers a more profound understanding of this important topic. Additionally, practical examples and recommendations are provided, making it valuable for both beginners and experienced professionals.

Mastering these integrated principles, understanding the data, applying the science of visual design, and structuring insights into a compelling narrative—is the key to transforming data into knowledge and influence.

Author,
Selver Softic

Dedication

To my sons,

May this work remind you that there is nothing in life you cannot achieve.
With courage, persistence, and belief in yourselves, every path is possible and every dream is within reach.

Introduction to Data Visualization

Why Data Visualization?

In an era where vast amounts of data are generated every second, data visualization has become an indispensable tool for making sense of complex information. Raw data, often stored in spreadsheets, databases, or unstructured text formats, can be overwhelming and difficult to interpret. However, by transforming this data into visual formats such as charts, graphs, heatmaps, and dashboards, patterns emerge that might otherwise remain hidden.

Data visualization is valuable because it offloads part of the thinking process from working memory onto the visual system. When data are encoded into position, length, or color, the viewer can compare values and notice patterns more quickly than by reading a table row by row. This is why visualization is not only "presentation" but also a practical cognitive tool for exploration and decision-making.

Data visualization helps to bridge the gap between data and understanding, allowing individuals and organizations to quickly extract meaning, identify trends, and make well-informed decisions. Whether in business, science, healthcare, or government, clear and effective visual representation of data enhances communication and drives better outcomes.

The ability to translate vast datasets into clear, comprehensible insights is no longer a niche skill but a fundamental competency for professionals across all fields. Data visualization serves as a universal form of communication, bridging the gap between raw data and human understanding. It allows us to make the invisible visible, revealing patterns, trends, and connections that might otherwise remain obscured in tables of numbers.

As Hal Varian, Google's Chief Economist, aptly states (Varian, 2009):

"The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that's going to be a hugely important skill in the next decades, ... because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it."

At its core, data visualization is a computer-supported, cognitive process that transforms symbolic data into geometric forms, using inter-active visual representations to amplify human cognition. This discipline, first formally defined by researchers at Xerox PARC, involves a journey from abstract data to the formation of a mental model in the viewer's mind. It is both a science and an art, grounded in the systematic mapping of data to visual attributes to facilitate insight (McCormick et al., 1987; Card, Mackinlay, & Shneiderman, 1999; R. Mazza, 2009).

One of the most powerful aspects of data visualization is its ability to distill complex datasets into intuitive and easily interpretable visuals. This process enables:

Pattern Recognition – Trends, correlations, and anomalies can be spotted at a glance.

Data Exploration – Interactive dashboards allow users to drill down into details and uncover deeper insights.

Quick Decision-Making – Decision-makers can react to insights more rapidly than if they were relying solely on raw data analysis.

For instance, a well-designed heatmap can instantly reveal areas of high customer engagement on a website, while a time-series graph can highlight seasonal sales trends in an e-commerce business.

Example in Figure 1 illustrates brake events and vibrations along the route of an e-scooter. More intensive events are represented in red, while less intensive ones are shown in other colors within the visible spectrum, primarily violet or blue when no significant events occur.
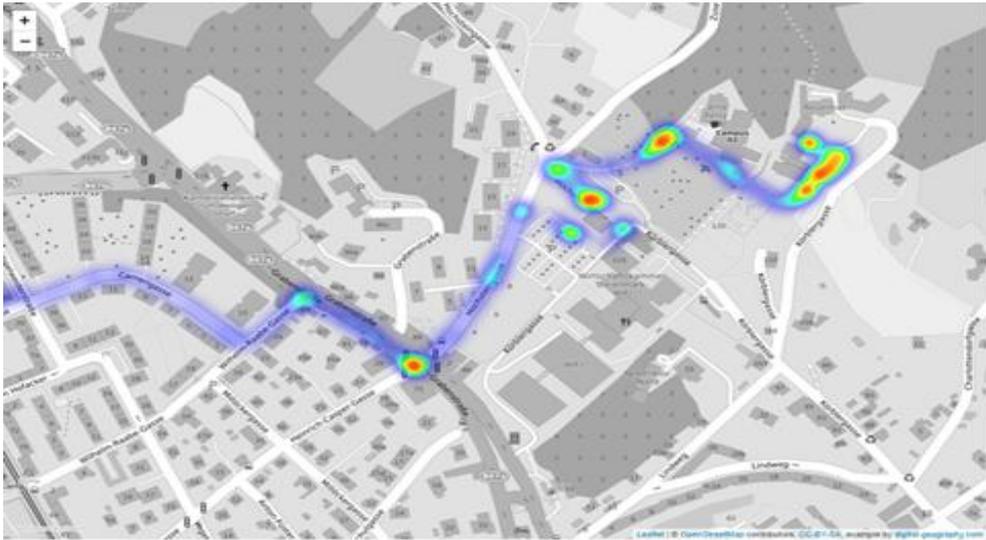
Figure 1. *Heatmap visualization of moving patterns*

Visualizing data helps to simplify relationships between variables, making them easier to interpret. Whether it's sales performance across different regions or disease outbreaks over time, visual tools like bar charts, scatter plots, and network diagrams reveal patterns that may not be evident in spreadsheets or textual reports.

For example, a line chart comparing stock market trends across different years can quickly show whether the market is following a similar cycle to previous years, allowing investors to make more informed trading decisions.

Data visualization removes ambiguity and makes information more accessible, particularly for audiences who may not have expertise in data analysis. Well-crafted visualizations can help different stakeholders, whether executives, engineers, marketers, or policymakers—grasp critical insights quickly without needing to sift through raw data.

For example:
  – A sales dashboard can display KPIs (Key Performance Indicators) such as revenue growth, customer retention rates, and profit margins at a glance.
  – A supply chain heatmap can show potential disruptions in real-time, allowing logistics teams to proactively manage issues.

- A COVID-19 spread map can help policymakers decide where to allocate medical resources most effectively.

In a fast-paced business environment, speed is crucial. Decision-makers cannot afford to spend hours analyzing spreadsheets or deciphering large datasets. By using real-time dashboards, visual reports, and automated data analytics, organizations can accelerate their decision-making processes.

For instance:

- Finance departments use interactive financial reports to track revenue, expenses, and profitability in real-time.
- Marketing teams leverage data visualizations of customer engagement metrics to refine their advertising strategies.
- Healthcare professionals analyze patient data through dashboards to improve diagnoses and treatment outcomes.

By reducing cognitive load and allowing quick comparisons, visualized data ensures that leaders and analysts can make confident, evidence-based decisions.

Data visualization plays a critical role across a wide range of industries, including:

- Business and Finance – Used for revenue forecasting, market analysis, risk assessment, and fraud detection.
- Healthcare and Medicine – Helps track disease outbreaks, analyze patient data, and visualize clinical trial results.
- Science and Research – Facilitates the interpretation of experimental data, genetic analysis, and climate studies.
- Marketing and Customer Analytics – Aids in understanding consumer behavior, campaign performance, and user engagement.
- Government and Policy-Making – Supports decision-making in urban planning, public health, and national security.
- For example, in healthcare, data visualization tools were crucial in tracking the global spread of COVID-19. Interactive dashboards developed by organizations like Johns Hopkins University helped the world monitor case numbers, vaccination rates, and death tolls in real-time.

Data visualization is more than just a way to make data look appealing—it is a critical tool for uncovering insights, improving

communication, and making informed decisions quickly. In today's data-driven world, organizations and individuals who leverage effective visual analytics gain a competitive edge by being able to identify opportunities, mitigate risks, and respond to challenges efficiently.

By harnessing the power of clear, intuitive, and interactive visual representations, we can transform raw data into meaningful insights that drive innovation and positive change across industries.

Definitions

Definitions of data visualization typically differ in how strongly they emphasize representation versus interaction. Representation-focused definitions stress the mapping from symbolic values (numbers, categories) into geometric marks (points, bars, lines). Interaction-focused definitions highlight that users can filter, zoom, and reconfigure views to test ideas and refine questions iteratively.

These definitions highlight different perspectives on data visualization, emphasizing either geometric transformation or cognitive amplification:

"Transformation of the symbolic into the geometric" (McCormick et al., 1987)
  – Focuses on the conversion of abstract, symbolic data into geometric forms (e.g., charts, graphs, and maps).
  – This definition emphasizes visual clarity—turning complex, intangible data into something that can be easily interpreted visually.
  – Example: A bar chart translating numerical values into bar heights.

"The use of computer-generated, interactive, visual representations of abstract data to amplify cognition" (Card, Mackinlay, & Shneiderman, 1999)
  – Highlights the interactive nature of modern data visualization.
  – Stresses how visualization enhances human cognition—helping users explore, interpret, and understand data more effectively.
  – Example: A dynamic dashboard where users can filter, zoom, or interact with data.

Both definitions underscore the importance of visual encoding but differ in scope—one focusing on representation, the other on interaction and cognitive benefits.

For teaching, it helps to keep a short working definition: data visualization is the intentional encoding of data into visual form to support understanding, analysis, and communication for a specific audience and purpose. The words "intentional," "encoding," and "audience/purpose" are important because they move the discussion from "what looks nice" to "what works".

Data Vizualisation – an Encoding Process?

Data visualization is the graphical representation of data and information. It transforms raw data into visual formats such as charts, graphs, maps, and infographics, making it easier to analyze, interpret, and communicate insights.

By leveraging visual elements, data visualization helps identify trends, patterns, and outliers that might not be apparent in raw data. It plays a crucial role in various fields, including business, science, engineering, and social research, enabling informed decision-making and effective data-driven storytelling.

Data visualization is an interdisciplinary field that involves the graphical representation of data using shapes, colors, and other visual elements. Its purpose is to present information in a way that allows for quick and efficient interpretation. It is also a cognitive process in which external visual representations are used to construct mental models of the world.

The process of data visualization begins with data collection and digitization, followed by machine-based cleaning and transformation. The goal is to present the data within a specific context that is relevant to the target audience.

Data Visualization Process (Final Optimized Version)
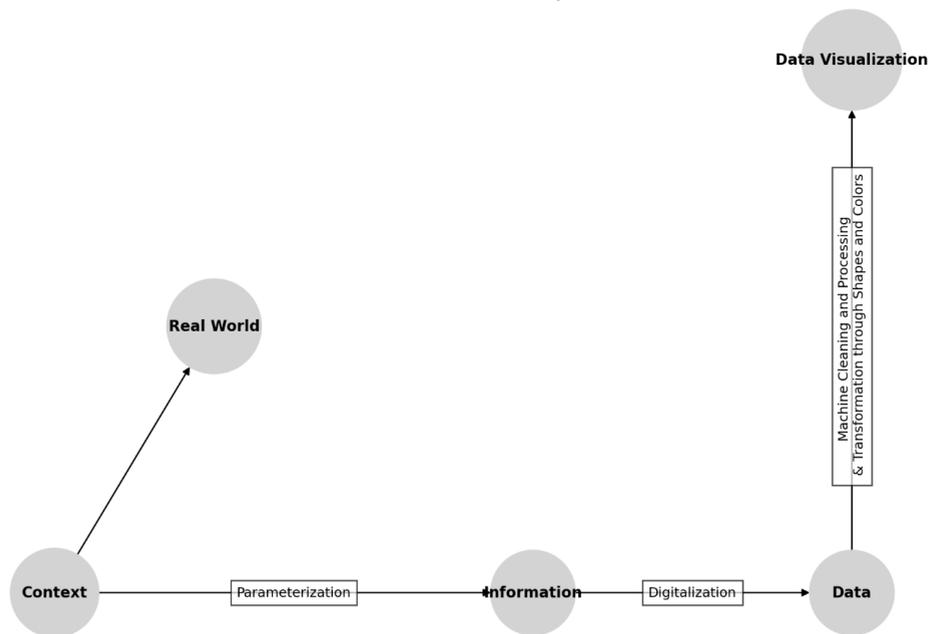
Figure 2. *Data Visualization Process*

Thinking of visualization as an encoding process clarifies what design choices actually do. The designer chooses marks (e.g., bars, points, lines) and visual channels (e.g., position, length, color, shape) and thereby creates a "visual code" that the viewer decodes. If the encoding is poorly matched to the task—such as using 3D volume to compare small differences—the decoding becomes error-prone and slow.

A practical implication is that you can debug a visualization like software: identify the message, identify the encodings used, and check whether those encodings are accurate and easy to compare. This mindset is especially helpful in dashboard contexts, where many charts compete for attention and encoding choices must be consistent across views.

Historical Development

The history of data visualization dates back centuries. Even in ancient times, simple maps and charts were used to convey information. Over the

years, techniques and tools evolved, leading to the development of modern visualization methods.

Figure 3. Cholera map showing deaths and pump locations during the 1854 Soho outbreak. From J. Snow, 1855, On the Mode of Communication of Cholera (Public domain)

A remarkable example from the 19th century is John Snow's cholera map (Tufte, 2006), which played a pivotal role in the development of modern epidemiology. In 1854, during a severe cholera outbreak in London, Snow meticulously plotted cholera cases on a map of the Soho district. His analysis revealed a concentration of cases around a specific water pump on Broad Street. By removing the pump handle, Snow was able to halt the spread of the disease, providing strong evidence that cholera was waterborne rather than airborne, as commonly believed at the time.

This groundbreaking approach not only revolutionized public health practices but also demonstrated the power of data visualization in identifying patterns and informing decision-making. Snow's work laid the foundation for spatial analysis and epidemiological mapping, techniques that continue to be essential in disease control and public health planning today.

Another significant example is Florence Nightingale's diagram on mortality analysis during the Crimean War (1853–1856) (Cook & Spear, 2019). Nightingale, a pioneering nurse and statistician, used data visualization to advocate for healthcare reforms, demonstrating how preventable diseases were the primary cause of death among soldiers rather than battle wounds.
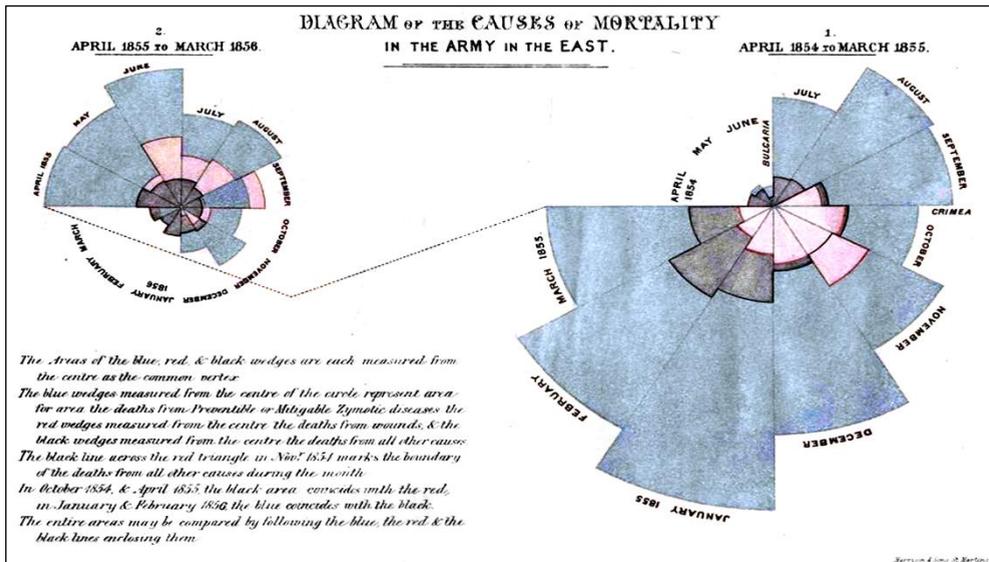
Figure 4. *Nightingale, F. (1858). Diagram of the causes of mortality in the army in the East (Public domain)*

To communicate her findings effectively, Nightingale created a polar area diagram, commonly referred to as the "Coxcomb" diagram. Unlike traditional bar or line charts, this innovative visualization displayed mortality rates in a circular format, making it easy to understand for policymakers and the general public.

Her diagram divided each year into twelve segments, with each sector representing a month. The area of each segment was proportional to the number of deaths that occurred in that month. She used different colors to differentiate between causes of death:

- Blue represented deaths from preventable diseases such as typhus, dysentery, and cholera.
- Red indicated deaths from wounds sustained in battle.
- Black denoted deaths from other causes.

The striking visual impact of the diagram made it clear that the majority of deaths were due to unsanitary conditions rather than combat injuries. Nightingale's work provided undeniable evidence that poor hygiene, lack of ventilation, and overcrowding in military hospitals were responsible for high mortality rates.

As a result of her findings, Nightingale successfully persuaded the British government to improve sanitation standards in military hospitals.

Simple interventions such as proper waste disposal, clean water supply, and better hospital ventilation drastically reduced mortality rates. Following her recommendations, the mortality rate among soldiers dropped from 42% to just 2%.

Her work did not stop with military hospitals; Nightingale's advocacy extended to civilian hospitals across Britain and beyond. She helped establish the first professional nursing schools and promoted hygiene reforms that became standard practice in hospitals worldwide.

Florence Nightingale's use of data visualization was groundbreaking. At a time when statistical analysis was not widely applied to healthcare, she used compelling visual evidence to drive policy changes. Her Coxcomb diagram is one of the earliest examples of data storytelling, demonstrating how well-designed visualizations can influence decision-makers and save lives.

Today, her approach continues to inspire modern data-driven healthcare policies, epidemiological studies, and public health campaigns. The principles she introduced—using data to highlight social issues and advocate for change—are still fundamental in fields such as medical research, global health, and disease prevention.
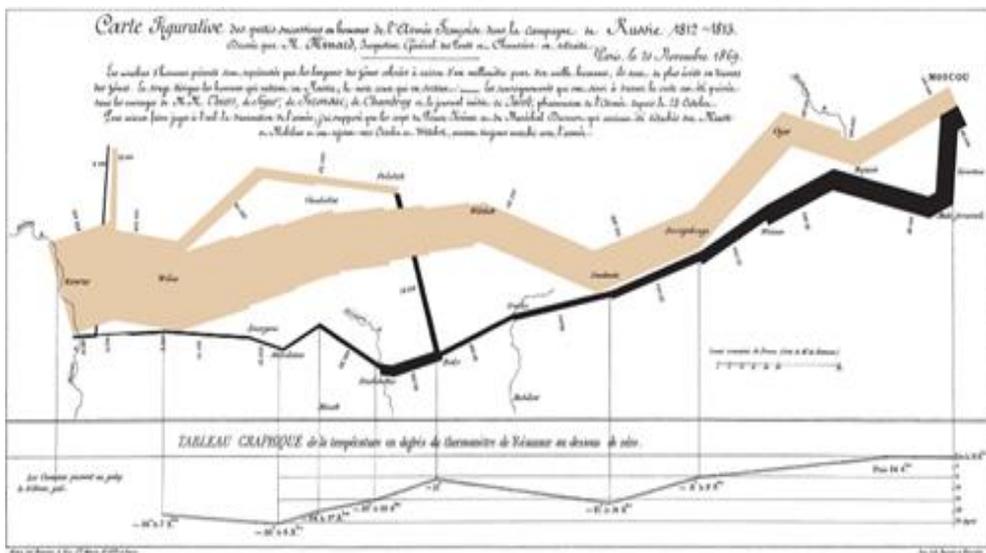
Figure 5. *Visual depiction of troop losses during Napoleon's 1812 Russian campaign. From C. J. Minard, Carte figurative et analytique de la campagne de Russie 1812–1813, 1869 (Public domain)*

Charles Joseph Minard's 1869 flow map depicting Napoleon's Russian campaign of 1812 is widely regarded as one of the greatest visualizations in history. His diagram masterfully conveys the devastating losses suffered by Napoleon's army during the ill-fated invasion of Russia, integrating multiple data points to tell a compelling story of military disaster.

Minard's diagram is exceptional because it combines six different variables into a single, easy-to-interpret visual representation :

1. Army Size – The thickness of the line represents the number of soldiers at each point in the campaign. The width shrinks dramatically as the army moves deeper into Russia and then retreats.

2. Geographical Position – The path of the army is plotted on a map, showing major cities along the route, including Smolensk, Moscow, and the Berezina River.

3. Direction of Movement – The tan-colored line represents the advance towards Moscow, while the black line illustrates the retreat back to France.

4. Temperature – A crucial aspect of the diagram, a separate temperature scale at the bottom of the map records extreme cold during the retreat.

5. Time Progression – The diagram is chronologically structured, helping viewers see how the campaign unfolded over time.

6. Deaths Over Distance – By linking troop numbers with specific locations and temperature data, Minard graphically depicts how many soldiers perished at each stage of the journey.

Napoleon started his invasion with 422,000 soldiers, confidently advancing toward Moscow. However, as the campaign progressed, Minard's diagram shows a steady decline in troop strength, highlighting the disastrous consequences of starvation, disease, exhaustion, and combat. By the time the army reached Moscow, only 100,000 troops remained.

The black line, representing the retreat, shrinks even more rapidly, reflecting the brutal toll of the Russian winter and continued attacks from enemy forces. By the time Napoleon's army crossed the Berezina River and exited Russia, only 10,000 troops survived—a staggering 98% loss of the original force.

The temperature readings at the bottom of the chart are particularly striking. The extreme cold—plummeting to -30°C (-22°F)—coincides with significant drops in the width of the retreating army's line, visually correlating freezing temperatures with mass casualties.

Minard's work remains one of the most celebrated visualizations in history. It has inspired modern data scientists, statisticians, and information designers, including Edward Tufte, who praised it as "the best statistical graphic ever drawn." (Tufte, 2006)

The diagram is still widely studied today in fields such as:
– Data visualization and information design – As an early example of combining multiple variables effectively.
– History and military strategy – Providing an unparalleled visual summary of one of history's greatest military disasters.
– Decision-making and risk analysis – Serving as a cautionary tale about the dangers of logistical failures and extreme environmental conditions.

Minard's Napoleon's Russian Campaign diagram is far more than just a historical chart—it is a masterpiece of data visualization (Tufte, 1983). It tells a tragic story through data, allowing viewers to grasp the sheer scale of the disaster in an instant. By skillfully integrating multiple data layers, Minard set a precedent for modern data visualization techniques, proving that a well-designed chart can be more powerful than pages of written analysis.

Historical examples remain relevant because they illustrate timeless principles: show data in context, connect evidence to action, and design for the decision-maker. Snow's cholera map shows how spatial encoding can reveal a causal mechanism; Nightingale shows how a persuasive visual argument can change policy; Minard shows multi-variable integration without losing the story. These examples also remind students that visualization quality is judged by what it enables people to do, not only by aesthetics.

To connect history to modern practice, you can explicitly ask: What is the core question in each historical graphic, what encodings answer it, and what decision did it influence? This turns "history" into an analytical skill that we can reuse for contemporary cases such as product analytics, manufacturing quality, or epidemiological dashboards.

Data Visualization Goals

Data visualization serves several key purposes, each aimed at making data more accessible, interpretable, and actionable. These goals ensure that data is not only analyzed effectively but also communicated in a way that supports better decision-making and understanding.

Visualization goals usually fall into three overlapping categories: exploration (find what you don't know yet), explanation (communicate what you learned), and monitoring (track a process against targets). Confusion arises when a single graphic tries to serve all three goals simultaneously — for example, using an exploratory, dense chart in an executive presentation. Aligning goal, audience, and context prevents over-complex designs and improves interpretability.

Some of the main goals can be defined as:

Analysis – Gaining Insights from Data

One of the primary objectives of data visualization is to extract meaningful insights from raw data that might not be immediately apparent. By transforming numbers into visual formats such as charts, graphs, and heatmaps, users can detect patterns, trends, and correlations.

Example:
- A scatter plot showing customer age vs. spending habits can reveal relationships between demographics and purchasing behavior.
- A time-series graph of website traffic can highlight peak usage hours, helping businesses optimize their content strategy.

Communication – Conveying Complex Information Clearly

A well-designed visualization should make complex information easy to understand for different audiences, including non-experts. The goal is to eliminate ambiguity and enhance clarity so that key messages are effectively conveyed.

Example:
- A COVID-19 case map can instantly communicate the severity of outbreaks in different regions, helping the public and policymakers take appropriate action.

– A dashboard summarizing company performance allows executives to quickly assess revenue, customer growth, and operational efficiency.

Decision Support – Enabling Data-Driven Actions

Effective data visualization helps decision-makers act based on facts rather than intuition. By presenting relevant metrics in an intuitive way, businesses, governments, and organizations can make more informed choices.

Example:
– A real-time financial dashboard displaying market trends helps investors decide when to buy or sell stocks.
– A logistics heatmap showing delivery times can assist in optimizing supply chain efficiency.

Illustration – Presenting Data in an Engaging and Understandable Way

A strong visualization should not only be informative but also visually appealing. Cluttered, hard-to-read graphs can obscure insights, whereas clear, well-structured visualizations improve engagement and understanding.

Example:
– A well-designed pie chart showing budget distribution makes it easier to understand how funds are allocated compared to a table of numbers.
– A network diagram of social media interactions can visually highlight key influencers in a community.

Good Vs. Bad Visualizations

A good visualization is one where the reader can (1) identify what is being compared, (2) see the main pattern quickly, and (3) trust that the encoding is not misleading. Bad visualizations often fail for predictable reasons: unclear chart purpose, poor scale choices, inappropriate use of area/3D, clutter, or color choices that confuse categories or hide variation. Many "bad" charts still look professional, which makes teaching critique skills essential.

An example of poor visualization is the incorrect representation of data, which can lead to misunderstandings. A good visualization, on the other hand, corrects such errors and improves clarity.

The two images in figures 6 and 7 represent different visualizations of the U.S. smartphone market share.
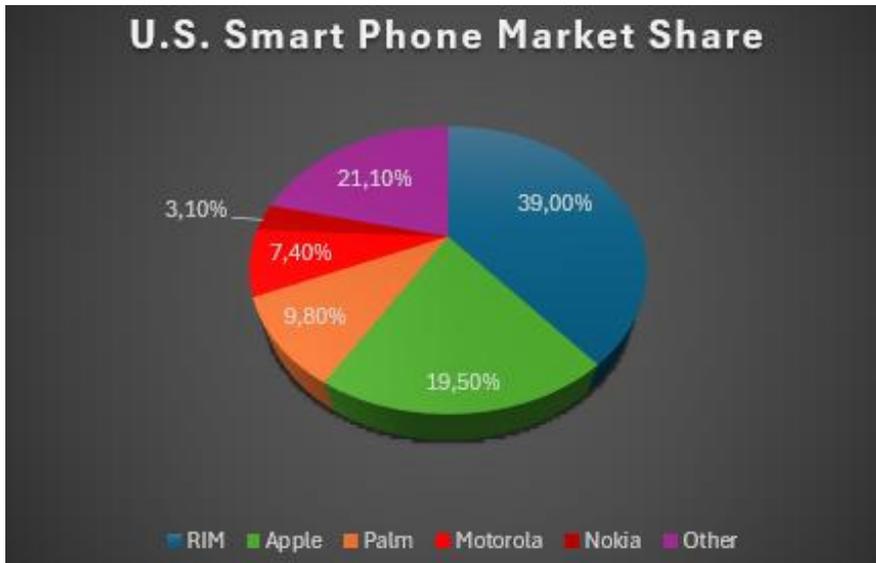
Figure 6. *Pie chart representation of market share*

Figure 7. *Same representation of market share as bar chart*

Here is an analysis of both:

Pie Chart (Figure 6)

The first image displays a pie chart that visualizes the market share of various smartphone manufacturers. Each segment is color-coded and labeled with its respective percentage.

Advantages:
– Provides a quick visual representation of the market distribution.
– Large segments stand out clearly, making it easy to see dominant players.

Disadvantages:
– Pie charts make it difficult to compare similar values accurately.
– The 3D effect can distort the perception of segment sizes, making some portions appear larger or smaller than they actually are.

Bar Chart (Figure 7)

The second image presents the same data in a bar chart format. The bar heights represent the market share percentages of different smartphone brands.

Advantages:
– Easier to compare different values directly.
– More precise for reading exact values compared to a pie chart.

Potential Manipulation:
– If the Y-axis does not start at zero, differences between companies might appear exaggerated.
– Unusual formatting, such as disproportionate bar widths or colors, can emphasize certain categories more than others.

By focusing on clarity, accuracy, and usability, good visualizations eliminate confusion and ensure data-driven decisions are based on reliable information.

Let's look concluding this section as another example of manipulation through visualization. The images you see in figure 8 depict two different visual representations of Iraq's conflict-related deaths over time. They

illustrate how data visualization choices—such as color, scale, orientation, and framing—can influence perception.

Figure 8. Manipulative Representation of Data in a Bar Chart

Analysis of the Two Visuals (Figure 8.):

Left Chart – "Raising Deadly Violence"
- Uses an inverted bar chart with red coloring, emphasizing violence and bloodshed.
- The downward bars create an impression of unstoppable destruction.
- The choice of color and aggressive design makes the deaths feel more severe and overwhelming.

Right Chart – "Delining Deaths in Iraq"
- Uses a conventional upright bar chart with blue coloring, creating a calmer, more analytical feel.
- The title frames the story positively, focusing on the decline rather than total deaths.
- Includes contextual elements (maps, line charts, and explanatory text) that soften the impact.

Manipulative Visualization Tactics:
− **Color Psychology**: Red is associated with danger and violence, while blue conveys stability and decline.
− **Chart Orientation**: The left chart appears more alarming due to the downward spikes, evoking a "falling" or "bleeding" effect.
− **Framing & Narrative**: The left chart emphasizes the scale of deaths, while the right focuses on improvement over time.

Takeaway:

This comparison highlights how the same data can be framed differently to influence perception. It underscores the need for **critical evaluation** of visualizations, particularly in news and policy discussions.

Conclusion

Data visualization is an indispensable tool in a data-driven world. It helps to effectively analyze information, communicate insights, and support decision-making. From its historical beginnings to modern interactive dashboards, it remains a dynamic field that continues to evolve.

Data visualizations are not neutral representations of reality but structured arguments that influence how audiences interpret information. As Cairo (2016) emphasizes, every chart, map, or graphic makes implicit claims through design choices such as scale, color, and data selection, which places a strong ethical responsibility on those who create visualizations to ensure accuracy, transparency, and honesty in communication.

Before moving on, emphasize that visualization is a pipeline discipline: weak data preparation, unclear intent, or misunderstood perception principles will surface later as "design problems." The next chapters therefore treat visualization as an end-to-end practice: prepare data, choose encodings, design for perception, and communicate insight as a coherent narrative.

Processing an Understanding Data

Data in visualization should be treated as a model of reality, not reality itself. Every dataset reflects measurement choices, sampling decisions, and operational definitions (e.g., what counts as a "customer," what qualifies as an "incident"). Making these choices explicit helps prevent false certainty and improves the credibility of conclusions drawn from visuals.

In summary, the process of data visualization is an iterative approach in which data is continuously examined, transformed, and visualized to ensure the best possible representation and interpretation .

Data in Visualisations

Data visualization is a central element of modern data analysis and decision-making processes. The process begins with data collection, which is carried out from various sources such as APIs, databases, or sensors. Today, data availability is no longer a problem—rather, the challenge lies in handling and analyzing it effectively.

After collecting, the next step is data cleaning. This is a crucial process to ensure that the data is error-free, consistent, and complete. Inadequate data cleaning can lead to inaccurate or misleading analytical results.

The subsequent step is data transformation, where the data is converted into a format suitable for visualization. This may include converting raw data into tabular or hierarchical structures. Additionally, missing values may be filled in, and data points can be normalized.

Visualization itself is the core part of the process. Here, the processed data is represented in visual forms such as bar charts, line charts, or heatmaps. Choosing the right type of visualization is critical to maximizing

the informational value of the data and effectively reaching the target audience.

Interactive features such as zooming, filtering, or drilldowns allow users to explore the data in greater depth and gain new insights. This interactivity makes data visualization a dynamic and adaptable tool.

Data, Information and Knowledge

The DIKW (Data – Information- Knowledge, -Wisdom) framing becomes practical when tied to decisions. Data are recorded values; information is data organized to answer a question; knowledge is the validated understanding that can guide action in new situations. Visualization mainly operates at the boundary between information and knowledge, because it helps humans build mental models and test explanations efficiently.

The terms data, information, and knowledge represent different stages of processing.

- Data are raw materials without immediate meaning, such as numbers, characters, or symbols.
- Information arises when data are placed in context. For example, the number "25" becomes information when interpreted as "temperature in degrees Celsius." This context gives data meaning and makes it understandable.
- Knowledge is the next level, emerging from the combination of information with experience and insights. For instance, the information "25°C" can generate the knowledge that this is a comfortable temperature for outdoor activities.

The relationship between data, information, and knowledge is often illustrated as a knowledge pyramid (Rowley, 2007) (see Figure 9.).

It highlights how data form the foundation, upon which information and ultimately knowledge are built. This structure underscores the idea that the quality of data determines the quality of knowledge.
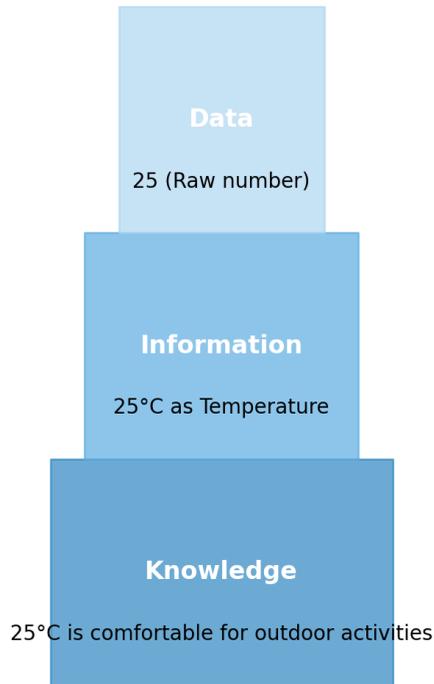
Figure 9. Illustration of knowledge pyramid

In the domain of data visualization, the transformation from raw data to meaningful insight requires more than graphical representation; it depends on how information is structured, contextualized, and interpreted by users. Ackoff's (1989) hierarchy from data to information, knowledge, and wisdom highlights that data alone lack meaning unless they are organized and embedded within a conceptual framework. Applied to data visualization, this perspective emphasizes that charts and dashboards should not merely display data, but actively support sense-making, interpretation, and decision-making by providing context, relationships, and explanatory cues that guide users toward knowledge and informed judgment.

The ability to effectively transform data into information and knowledge is essential for businesses and organizations. Well-interpreted data and information lead to better decision-making, while misunderstandings or poor data quality can result in significant errors.

Collecting and Processing Data

Collecting and processing data is a fundamental part of the data visualization process. It lays the groundwork for all subsequent steps and significantly determines the quality of the resulting visualizations.
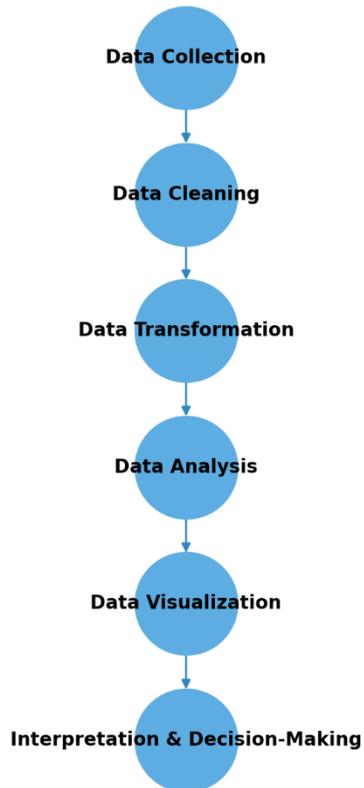
Figure 10. Illustration of Data Lifecycle in the Visualization Process

Here, in Figure 10 is a Data Lifecycle in the Visualization Process diagram. It illustrates the sequential stages involved:

– **Data Collection** – Gathering raw data from various sources.
– **Data Cleaning** – Identifying and correcting errors, inconsistencies, and missing values.
– **Data Transformation** – Structuring, aggregating, and converting data for analysis.

- **Data Analysis** – Extracting patterns and insights from the processed data.
- **Data Visualization** – Representing the findings visually through charts, graphs, and dashboards.

Interpretation & Decision-Making – Using visualized data for insights and informed decision-making.

Data Collection Methods

Data can originate from a wide variety of sources, each offering unique advantages depending on the application context. One important category is web scraping, which refers to the automated extraction of information from websites. Specialized tools and libraries—such as BeautifulSoup[1], Scrapy[2], or Selenium[3]—parse HTML content and convert it into structured datasets. Web scraping is particularly valuable when no official programming interface is available and information is only presented visually on web pages.

A closely related approach is the use of API interfaces, which provide structured, standardized access to data through well-defined endpoints. APIs typically return information in JSON or XML format and are designed for reliable, real-time data retrieval. They are widely used in modern systems—for example, to access social media data, weather information, or financial market feeds. Because APIs follow explicit documentation and technical standards, they ensure a consistent and legally compliant method of obtaining data.

Data stored within organizations is frequently retrieved through database querying. Relational databases such as MySQL, PostgreSQL, or Oracle rely on SQL, whereas NoSQL systems like MongoDB or Elasticsearch provide their own query mechanisms. Querying allows precise filtering, aggregation, and extraction of records from large operational systems or

---

[1] https://beautiful-soup-4.readthedocs.io/en/latest/

[2] https://www.scrapy.org/

[3] https://www.selenium.dev/

data warehouses. This method is central to enterprise analytics, transaction processing, and business intelligence applications.

In many cases, data is neither fully structured nor entirely unstructured but belongs to the category of semi-structured sources. Formats such as XML, JSON, CSV, email headers, or server logs require parsing, a process that identifies tags, patterns, or delimiters to extract meaningful attributes. Parsing converts heterogeneous data fragments into structured representations suitable for further analysis. It is commonly used to process machine logs, document metadata, or exported system reports.

Beyond digital repositories, data is also collected directly from the physical world. Sensors and measurements play a key role in fields such as the Internet of Things (IoT), industrial automation, environmental monitoring, and predictive maintenance. Devices can capture temperature, humidity, pressure, vibration, motion, or geolocation in real time. Such data sources enable continuous monitoring and are essential for applications that require high temporal resolution or immediate feedback.

Not all information can be obtained automatically. Manual input remains an important method for collecting subjective or qualitative data, often through surveys, questionnaires, interviews, or web forms. This approach is indispensable when human judgment, personal preferences, or expert assessments are required. While more labor-intensive and potentially prone to bias, manually collected data provides insights that automated systems cannot capture.

Finally, large-scale digital ecosystems often rely on web crawling, a technique where automated programs—known as crawlers or spiders—systematically navigate the web by following hyperlinks. Web crawling is used by search engines to index content but also serves research purposes, such as collecting corpora of news articles or scientific documents. Crawlers explore entire website structures, enabling comprehensive data acquisition across interconnected pages.

Together, these diverse methods provide a rich foundation for modern data-driven applications. They cover the spectrum from structured digital repositories and real-time sensor streams to human-generated content and

large-scale web ecosystems, ensuring that data analysts and researchers can obtain information suited to virtually any analytical task.

Data Cleaning

In the context of data visualization, the reliability and interpretability of visual representations fundamentally depend on the quality of the underlying data. Rahm and Do (2000) emphasize that data cleaning—including the detection and correction of errors, inconsistencies, duplicates, and missing values—is a critical prerequisite for any data-driven analysis. Applied to data visualization, this perspective highlights that misleading or noisy data can propagate through visual encodings and result in incorrect interpretations, whereas systematic data cleaning ensures that visualizations accurately reflect reality and support trustworthy insight and decision-making.

Once data has been collected, a crucial next step is data cleaning, which ensures that the dataset is accurate, consistent, and ready for analysis. Raw data—regardless of its source—often contains imperfections such as errors, inconsistencies, duplicates, or missing values. If left unaddressed, these issues can distort analytical results and significantly reduce the reliability of any models or insights built upon the data. Therefore, data cleaning represents one of the most important preparatory phases in the data processing pipeline.

A fundamental aspect of data cleaning is the removal of duplicates. Duplicate records may occur through repeated measurements, multiple submissions, system errors, or inconsistencies in data integration. Removing them helps prevent biased results, especially in statistical analysis or machine learning tasks where repeated entries could skew distributions or artificially inflate the apparent frequency of events.

Another core activity involves handling missing values, which may arise from sensor failures, incomplete survey responses, system downtime, or formatting errors. Multiple strategies exist for addressing missing data. Simple methods include deletion of incomplete rows or columns, while more sophisticated approaches use imputation techniques—such as mean, median, mode replacement, regression-based estimation, or advanced

algorithms like k-nearest neighbors or multiple imputation—to fill in gaps in a statistically meaningful way. The choice of method depends on the context, the amount of missing data, and the analytical goals.

In addition, data often requires standardization of formats, units, and representations. Inconsistent date formats, different measurement units (e.g., Celsius vs. Fahrenheit, meters vs. inches), or divergent categorical labels can create significant analytical obstacles. Standardization ensures uniformity across the dataset, allowing comparisons, aggregations, and computations to be performed reliably. This stage may involve converting units, harmonizing naming conventions, normalizing text fields, or applying encoding schemes to categorical attributes.

Overall, data cleaning is a meticulous but indispensable process that transforms raw, heterogeneous data into a coherent and trustworthy foundation for subsequent analysis. Without systematic cleaning, even the most advanced analytical techniques cannot deliver valid or meaningful results.

Data Transformation

After data cleaning, an equally important phase is data transformation, which reshapes and converts raw information into forms suitable for analysis and visualization. Transformation ensures that data is consistent, comparable, and computationally efficient, enabling meaningful statistical or machine-learning workflows. It bridges the gap between raw inputs and analytical readiness, and plays a key role in producing accurate insights.

One central transformation technique is aggregation, which summarizes detailed records into higher-level representations. Aggregation helps reduce noise, reveal patterns, and make large datasets manageable.

For example, instead of analyzing millions of individual webshop transactions, a data analyst may aggregate the data to daily or weekly revenue totals, allowing clearer identification of trends such as seasonal demand or sales cycles. Similarly, sensor data that records temperature every second might be aggregated into hourly averages to detect long-term environmental patterns without being overwhelmed by short-term fluctuations.

A second major aspect of transformation is the normalization of units and scales. Data from different equipment, locations, or systems often uses inconsistent measurement units, which must be harmonized to ensure comparability.

For example, product weights collected from various suppliers might be listed in kilograms, pounds, and ounces. Converting all values to one unit—such as kilograms—ensures meaningful comparisons. In addition, normalization techniques like min-max scaling or z-score standardization may be applied to bring numerical features onto a comparable scale. This is particularly important in machine learning: for instance, when predicting house prices, the feature square meters may range into the hundreds while number of rooms ranges only into single digits. Normalization prevents larger-scaled variables from dominating the model.

A third important category of transformation involves converting textual or categorical information into numerical formats so that algorithms can process them. Most statistical procedures and machine-learning models operate exclusively on numbers, not text strings.

For example, a column containing product categories like "Electronics," "Clothing," and "Furniture" may be transformed using one-hot encoding, where each category becomes a binary variable. For more complex text, such as customer reviews, techniques like TF-IDF transform words into numerical representations that capture term frequency and importance. More advanced approaches—such as word embeddings like Word2Vec or BERT—convert entire sentences or documents into dense numerical vectors that preserve semantic meaning and can be used for tasks such as sentiment analysis or clustering.

In essence, data transformation ensures that datasets are coherent, standardized, and ready for sophisticated analysis. Through aggregation, normalization, and numerical encoding, raw data becomes structured and meaningful, allowing analysts, researchers, and AI models to produce accurate and interpretable results. Without these transformations, even well-cleaned data may remain unsuitable for higher-level modeling, visualization, or decision-making.

Data Preparation Tools

A wide range of software tools can support and streamline the processes of data cleaning, transformation, and preparation. These tools reduce manual effort, ensure reproducibility, and offer features that make complex operations more accessible even to non-experts. Depending on the scale of the data and the user's technical background, both graphical tools and programming-based solutions are available.

One widely used tool is OpenRefine[4], designed specifically for cleaning and structuring messy datasets. Formerly known as Google Refine, it allows users to explore data quality issues interactively, identify inconsistencies, cluster similar values, and apply transformations across entire columns. For example, OpenRefine can automatically detect variations of categorical entries such as "USA," "U.S.A.," and "United States," and standardize them into a single, consistent label. Its powerful "facet" and "clustering" functions make it particularly valuable for preparing data collected from semi-structured or heterogeneous sources.

Another important tool is Tableau Prep[5], which focuses on intuitive data transformation and workflow automation. It offers a visual, drag-and-drop interface that allows users to merge tables, filter records, create calculated fields, reshape datasets, and apply data cleaning steps without writing any code. Tableau Prep also records every action in a reproducible workflow, enabling analysts to update data pipelines easily when new datasets arrive. This makes it ideal for business environments where recurring data preparation tasks must be performed efficiently and consistently.

For users who require more control and programmability, Python libraries such as Pandas provide a highly flexible and powerful environment for data processing. Pandas offers functionalities for reading structured formats (CSV, Excel, SQL databases, JSON), handling missing values, filtering and grouping data, applying complex transformations, and integrating with machine-learning libraries like scikit-learn. For example, a data scientist can

---

[4] https://openrefine.org/

[5] https://www.tableau.com/de-de/products/prep

use Pandas to merge multiple datasets from different sources, compute aggregated metrics, normalize values, and prepare feature matrices for predictive models—all within a few lines of code. This makes Python the tool of choice in many scientific, industrial, and academic contexts.

Together, these tools offer both automated processes and user-friendly interfaces that considerably improve the efficiency and reliability of data preparation. Whether through visual workflows or programmatic control, they help ensure that data is clean, consistent, and ready for advanced analysis and visualization.

Transformation prepares data for analysis and for the chosen visual encodings. Aggregation (daily → weekly), reshaping (wide ↔ long), binning, and feature creation (rates, per-capita measures, rolling averages) directly determine what patterns will be visible. Tools such as OpenRefine, Tableau Prep, and pandas support reproducible transformations; reproducibility matters because the same chart should be regenerable when data updates.

Visual Analytics

Before advanced analytical methods can be applied, data must first be collected, cleaned, and transformed into a reliable and structured form. These earlier steps—covering data acquisition, data cleaning, data transformation, and the use of specialized tools such as OpenRefine, Tableau Prep, and Python's Pandas[6]—lay the foundation for meaningful analysis. High-quality visual and analytical insights are only possible when the underlying data is accurate, consistent, and well-prepared. In this sense, data preparation is a prerequisite for all subsequent stages of Visual Analytics.

Within data visualization and visual analytics, effective insight generation depends on structured and iterative knowledge discovery processes rather than isolated analytical steps. Kurgan and Musilek (2006) survey and compare major Knowledge Discovery in Data Mining (KDD) process models and emphasize that knowledge extraction is an iterative, human-centered cycle involving data understanding, preprocessing,

---

[6] https://pandas.pydata.org/

modeling, evaluation, and interpretation. In the context of data visualization, this highlights the role of visual representations as integral components of the KDD process, supporting exploration, hypothesis refinement, and validation across multiple stages rather than serving as mere end-point presentations of results.

Visual Analytics (Keim et al., 2008) is an innovative approach that combines visual and analytical methods to provide decision-makers with deeper insights into complex data. This discipline merges the strengths of humans and machines to analyze data in ways that would otherwise not be possible.

A typical Visual Analytics process includes:

– **Data Integration** – Merging data from various sources.
– **Transformation** – Adapting data to meet the specific requirements of the analysis.
– **Visual Exploration** – Using interactive dashboards to discover patterns and anomalies.
– **Feedback Loops** – Continuously refining visualizations based on new insights.

User interaction plays a central role in the Visual Analytics process. Features such as zooming, filtering, and drilldowns allow users to dive deeper into the data and test hypotheses. This interactivity makes Visual Analytics a dynamic tool that adapts to the user's needs.

Another major advantage of Visual Analytics is the ability to combine human creativity with computational power. While machines can efficiently analyze large datasets, humans provide contextual knowledge and creative problem-solving approaches.

The key advantage is that interaction (filtering, drill-down, linking views) supports hypothesis refinement and rapid error detection—outliers may be data quality issues, segment definitions may be wrong, or an apparent pattern may vanish under a different normalization.
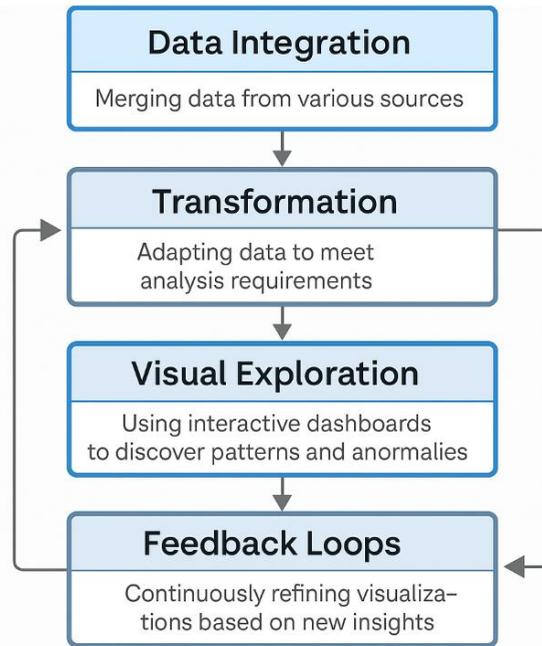
Figure 11. *Illustration of the Visual Analytics process*

Summary

The processing and understanding of data form the essential foundation for creating meaningful and accurate data visualizations. As this chapter has shown, a structured approach—beginning with data collection, followed by cleaning, transformation, and preparation—ensures that visual representations are built on reliable and consistent information. Visual Analytics expands this process by integrating computational methods with human visual perception, enabling deeper insights into complex datasets and supporting informed and effective decision-making.

Taken together, these steps demonstrate that data visualization is far more than the act of producing charts or dashboards. It is a comprehensive, iterative process that demands technical expertise, methodological rigor, and interdisciplinary collaboration. When supported by appropriate tools, robust workflows, and analytical thinking, organizations can fully leverage their data to reveal patterns, uncover relationships, and derive strategic value.

With this solid understanding of the data preparation and Visual Analytics foundations, the next chapter will focus on the methods and core elements of data visualization. It will introduce fundamental visualization types, explain visual encoding principles, discuss perceptual factors, and explore how design choices influence clarity, accuracy, and storytelling. This progression marks the transition from preparing data to effectively communicating insights—an essential skill for analysts, researchers, and decision-makers alike.

Methods and Elements of Data Visualization

Methods and elements of data visualization serve as the building blocks that enable the clear representation of complex data. They not only help convey information effectively but also reveal patterns and relationships that may remain hidden in raw data.

Successful data visualization requires careful selection of methods and elements. This choice depends on several factors, including:

- **Target audience**: Expert viewers vs. general audiences.
- **Purpose**: Is visualization meant to inform, persuade, or analyze?
- **Type of data**: Are the data numerical, categorical, or time-based?

This chapter introduces the most common visualization types and techniques for representing categories, proportions, temporal patterns, and relationships. Additionally, best practices will be discussed to maximize the effectiveness of visualizations.

Common Types of Visualisations

Selecting an appropriate visualization type is a fundamental step in ensuring that data is communicated clearly and effectively. The choice depends not only on the nature of the data—such as whether it is categorical, numerical, or temporal—but also on the analytical goals and the intended audience. In many cases, simple visualizations can be more powerful than complex ones, provided they align well with the message being conveyed. Among the most widely used and versatile formats are plain text summaries, tables, and heatmaps.

Plain text representations are highly effective when the goal is to communicate a small number of key values. Rather than using elaborate graphics, plain text emphasizes essential figures directly and without distraction. This approach is particularly suitable for executive summaries, dashboards, and headlines where clarity and immediacy are crucial.

For example, stating that the proportion of women in R&D positions in Luxembourg increased from 17.6% in 2005 to 27.3% in 2013 delivers a concise and powerful message. Such textual highlights are often more effective than charts when dealing with one or two central metrics or when the numeric values themselves carry the primary meaning.

Tables, by contrast, are best suited for displaying multiple metrics or dimensions in a structured and comparable format. They allow the reader to inspect precise values, compare categories, and understand data that may span different units or currencies. Tables are particularly valuable for heterogeneous audiences—such as stakeholders in finance, management, or engineering—who may each focus on different aspects of the data.

A common use case is found in financial reporting, where revenue, costs, margins, and profits are often presented in tabular form across multiple time periods and currencies. Tables prioritize accuracy, detail, and completeness, making them indispensable when exact values matter more than visual patterns.

Heatmaps offer a more visually intuitive alternative to traditional tables by using color intensities or gradients to encode numerical values. This allows users to identify patterns and outliers at a glance, even within large datasets. Heatmaps are especially useful when the structure resembles a matrix, where relationships between two categorical dimensions can be visualized simultaneously.

In fields such as bioinformatics, heatmaps are routinely used to examine gene expression levels across different samples or experimental conditions. Similarly, in clustering and machine learning applications, heatmaps can reveal groupings, correlations, and anomalies that may be less apparent in a simple table.

Together, these visualization forms demonstrate the importance of selecting the appropriate method based on the context, the complexity of the dataset, and the message that needs to be communicated. Whether highlighting a single key figure, comparing precise values, or uncovering structural patterns, choosing the right visualization type is essential for enabling accurate interpretation and informed decision-making.
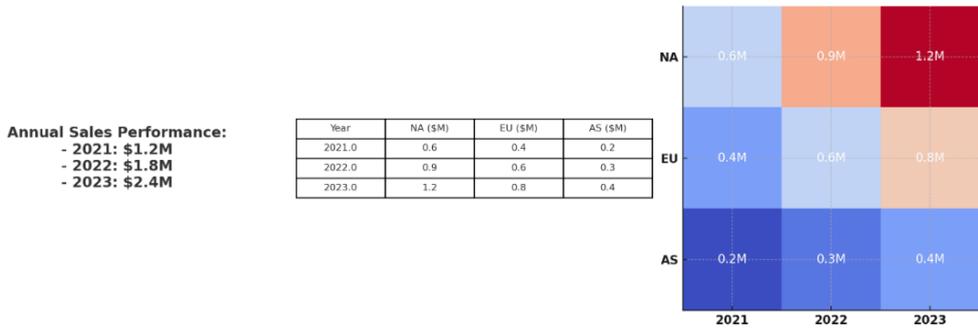
Annual Sales Performance:
- 2021: $1.2M
- 2022: $1.8M
- 2023: $2.4M

Year	NA ($M)	EU ($M)	AS ($M)
2021.0	0.6	0.4	0.2
2022.0	0.9	0.6	0.3
2023.0	1.2	0.8	0.4

Figure 12. *Text, Table and Heatmap as visualization types*

Chart Types and Their Applications

Charts form the core of modern data visualization. They translate numerical and categorical information into visual structures that can be interpreted quickly, intuitively, and with minimal cognitive effort. Selecting the appropriate chart type is essential, as different visual forms emphasize different aspects of the data—relationships, comparisons, trends, or proportions. The effectiveness of a chart depends on both the underlying data structure and the specific message the visualization aims to convey.

Among the most fundamental chart types are **scatter plots**, which are widely used to explore the relationship between two continuous variables. Each point represents an observation, and the spatial distribution of these points reveals patterns such as correlations, clusters, or anomalies. Scatter plots are particularly powerful for identifying whether one variable predicts or influences another.

For example, in fleet management, a scatter plot can illustrate the relationship between the distance traveled by commercial vehicles and their corresponding fuel costs. Patterns in the chart may highlight unusually high fuel consumption for certain routes, indicate inefficiencies, or reveal maintenance-related issues.

Building on scatter plots, **bubble charts** add a third dimension by encoding an additional variable using the size of each bubble or even fourth suing the color of the bubble. This makes it possible to analyze more complex datasets while maintaining interpretability. Bubble charts are often used in

economic and social analyses, for instance, comparing countries by GDP (x-axis), life expectancy (y-axis), and population size (bubble size) while for instance the color represents the continent where the country is situated. While visually appealing and informative, they should be used with care, as size-based encoding can be misinterpreted if not scaled appropriately.

Line charts are indispensable for visualizing developments over time. They connect individual data points to form a continuous progression, making them ideal for detecting trends, seasonal patterns, or long-term changes.

Common applications include stock market evolution, daily temperature changes, website traffic patterns, or monthly production volumes in a factory. Because time-series datasets are ubiquitous across industries, line charts remain one of the most frequently used visualization types. They support forecasting, anomaly detection, and comparative analysis across multiple time lines.

Bar and column charts are among the most versatile visualizations for comparing discrete categories. They encode values through bar length, enabling straightforward comparisons across groups. Bar charts can represent frequencies, totals, proportions, or averages, and they work equally well with nominal and ordinal data.

Examples include comparing sales across product categories, analyzing customer satisfaction ratings, or visualizing regional differences in employment. Bar charts also support grouping and stacking, which makes it possible to breakdown comparisons across subcategories or display cumulative effects.

Pie and donut charts specialize in illustrating the relative proportions of categories within a whole. Although they have limitations—such as difficulty in accurately comparing slices of similar sizes they remain widely used in business and communication contexts due to their immediate intuitive appeal.

They are ideal when the goal is to emphasize part-to-whole relationships, such as the percentage distribution of market shares, budget allocations, or demographic segments. Donut charts, which add a central

empty area, can further improve readability and allow placement of labels or summary metrics.

Taken together, these chart types represent the foundation of visual communication. By choosing the right type of chart, analysts can highlight relationships, reveal patterns, illustrate change, or emphasize proportions. Effective use of charts is therefore essential for conveying insights clearly and enabling data-driven decision-making.
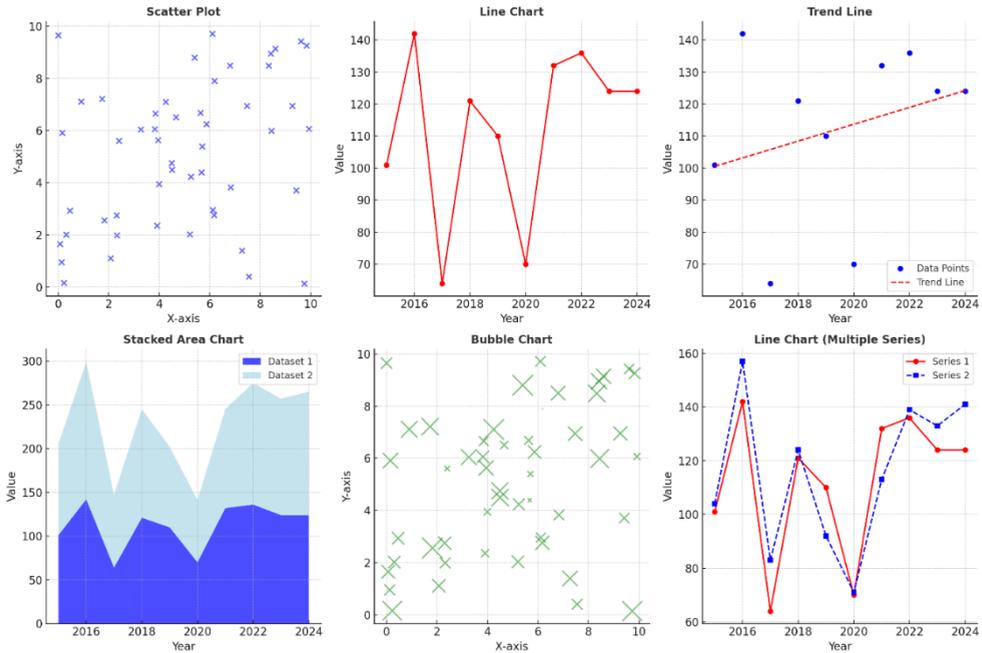

Figure 13. *Scatter and line chart representations*

Visualizing Categories, Proportions and Temporal Patterns

Different types of data require specific visualization techniques to convey insights effectively. Common approaches include:

– **Categories**: Bar charts, heatmaps, treemaps.
– **Proportions**: Pie charts, donut charts, stacked column charts.
– **Temporal Patterns**: Line charts, slope charts, sparklines.
– **Relationships**: Scatter plots, bubble charts.

A pie chart, for instance, is an effective method for showing market shares of different products. This method helps illustrate relative proportions within a dataset.

Heatmaps are an excellent choice for highlighting categories and proportions within tables. A common example is the visualization of 7-day COVID-19 incidence rates across different age groups.

For temporal patterns, line charts are often the best choice. They provide a clear depiction of increases, decreases, and fluctuations over time. A variation of this is the slope chart, which is particularly useful for comparing two specific time points directly.
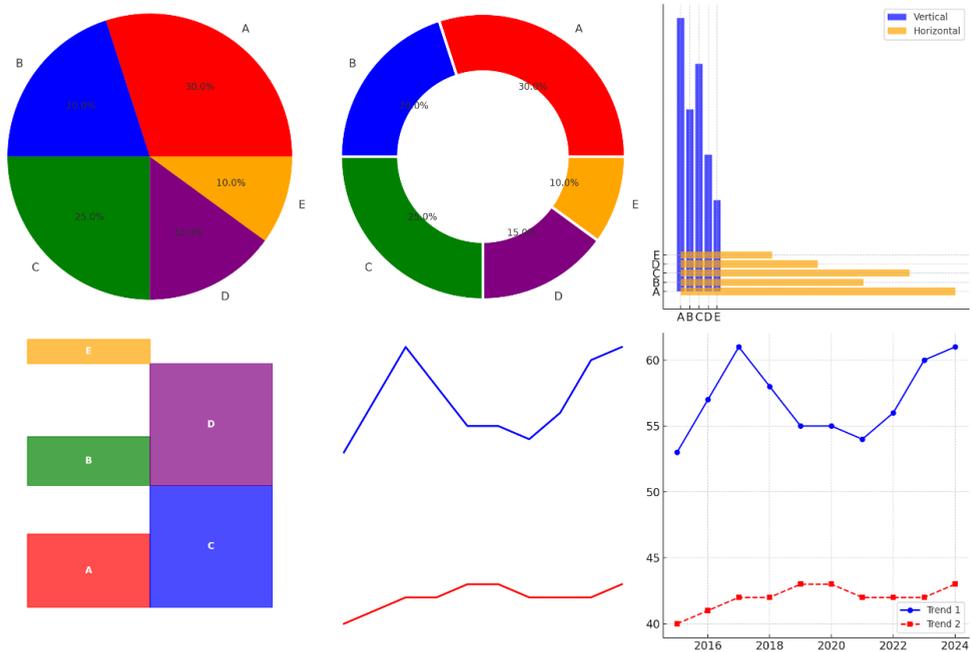

Figure 14. Category, proportions and temporal pattern visualizations

Summary and Key Take Aways

Selecting the right visualization method is crucial for maximizing the impact of data. It is essential to consider both the target audience and the intended purpose of the visualization.

A fundamental principle to keep in mind is:

"Context is king!"

The choice of visualization type should match the specific requirements of the data and its intended message. Inappropriate chart selection can distort insights and lead to misinterpretations.

Special caution should be taken when using 3D charts. These should only be used when a third dimension is truly necessary; otherwise, they may create more confusion than clarity.

Overall, the wide range of visualization methods and elements offers numerous opportunities to effectively present data. The careful selection and application of these techniques are key to creating successful visualizations.

Aspects of Visual Perception

Visual perception is not a passive "camera" that records what is in front of us. It is an active interpretation process: the brain selects signals, groups them into meaningful units, and fills gaps using expectations and prior knowledge. In visualization practice this means two things: first, design choices determine which patterns become visible; second, viewers can confidently misread a chart if the encodings conflict with human perceptual habits.

Introduction to Visual Perception

Effective data visualization is not merely a matter of technical skill; it is deeply rooted in the principles of human psychology and cognitive science. Visualizations are not just *seen* by the eyes—they are *processed* by the brain. To create clear and impactful graphics, we must understand how the brain takes in sensory information, processes it under the constraints of its memory systems, and manages cognitive load.

Visual perception is a complex process in which the human brain processes, interprets, and transforms visual stimuli into meaningful information. It plays a crucial role in data visualization as it determines how effectively a visualization conveys its message.

The human visual system is designed to quickly recognize patterns, colors, shapes, and movements. This process is accelerated by preattentive processing, where certain visual features such as color or size are automatically perceived without the need for conscious attention.

Visualizations that leverage the principles of visual perception can present complex data in an intuitively understandable way. For example, highlighting a trend in a chart using a high-contrast line color can immediately draw the viewer's attention to the relevant infomation.

The goal of this chapter is to explain the fundamentals of visual perception and demonstrate how they can be intentionally applied in data visualization.

In the context of data visualization, effective charts and interactive graphics can be understood and improved by modeling the cognitive processes of users rather than relying solely on visual intuition. Card, Moran, and Newell (1983) demonstrate that human interaction with information systems follows predictable psychological principles, and their frameworks—such as the Model Human Processor—provide a basis for estimating perceptual effort, cognitive load, and interaction efficiency. Applied to data visualization, these models help explain how users perceive visual encodings, process patterns, and interact with visual interfaces, thereby supporting the design of visualizations that are not only aesthetically appealing but cognitively efficient and easier to understand.

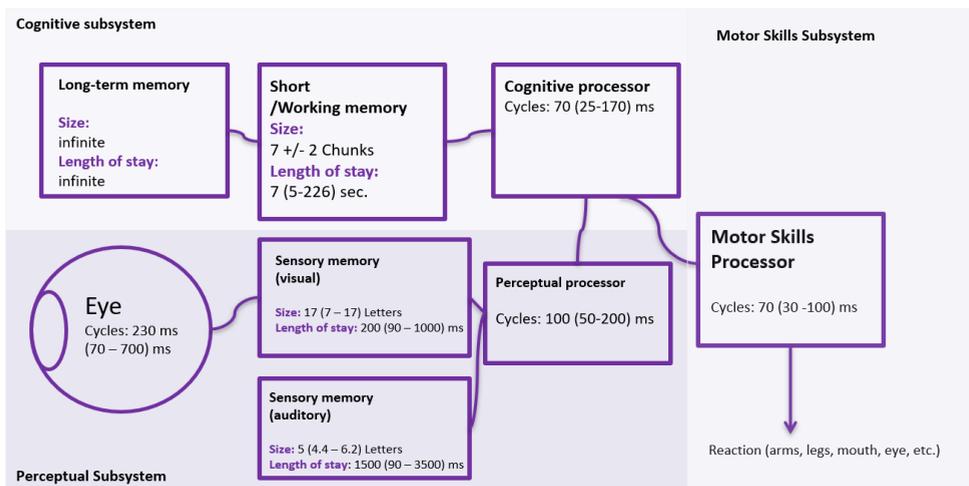## Visual Processing in HPM (Card S.K., 1981)

Figure 15. Illustration of the human visual system, HPM by Card, redrawn by author

The Human Information Processor Model (HPM) (figure 15) provides a powerful analogy for understanding how we process information by comparing the human mind to a computer's architecture. This model outlines three critical stages:

1. **Sensory Input (Perception):** Visual stimuli are first received by the eyes.
2. **Working Memory (Short-Term Memory):** This is where conscious processing, analysis, and comparison occur. Its capacity is famously limited.
3. **Long-Term Memory**: Relevant patterns, concepts, and knowledge are stored for future recall.

Much like a computer relies on RAM for active tasks, our brain uses a highly limited Working Memory for conscious processing. The goal of effective design is to present information in a way that respects this bottleneck, preventing the mental equivalent of a system crash.

Table 1. Memory types overview

Stage	Application in Visualization	Key Takeaway
Sensory Input	Preattentive attributes (like color, size, and shape) are used to draw the viewer's attention to key elements before conscious thought begins.	Direct the viewer's gaze purposefully to the most important parts of the visualization.
Working Memory	Information must be presented in a clear, simple, and grouped manner to avoid overwhelming this limited resource. It can hold approximately **7 +/- 2 chunks** of information at a time.	Less is more. Use grouping strategies (chunking) to package information into digestible units.
Long-Term Memory	Familiar patterns, symbols, and metaphors are leveraged to make new information easier to understand and remember.	Avoid unnecessary cognitive effort by using conventional chart types and symbols that the audience already knows.

Perception and Memory

Memory plays a central role in the perception and processing of data. It influences how quickly and effectively we can absorb and interpret information from visualizations. Memory can be divided into three main categories:

- **Sensory memory:** Stores visual information for only a few milliseconds. It enables rapid recognition of basic features such as colors or shapes.
- **Short-term memory (working memory):** Has a limited capacity (7±2 elements) and is essential for processing and analyzing information in real time.
- **Long-term memory:** Stores experiences and knowledge that can be used to interpret new information.

Example: When looking at a bar chart, your sensory memory instantly registers the colors and lengths of the bars. Your short-term memory allows you to compare these values, while your long-term memory helps you understand the meaning of the visualization based on prior knowledge.
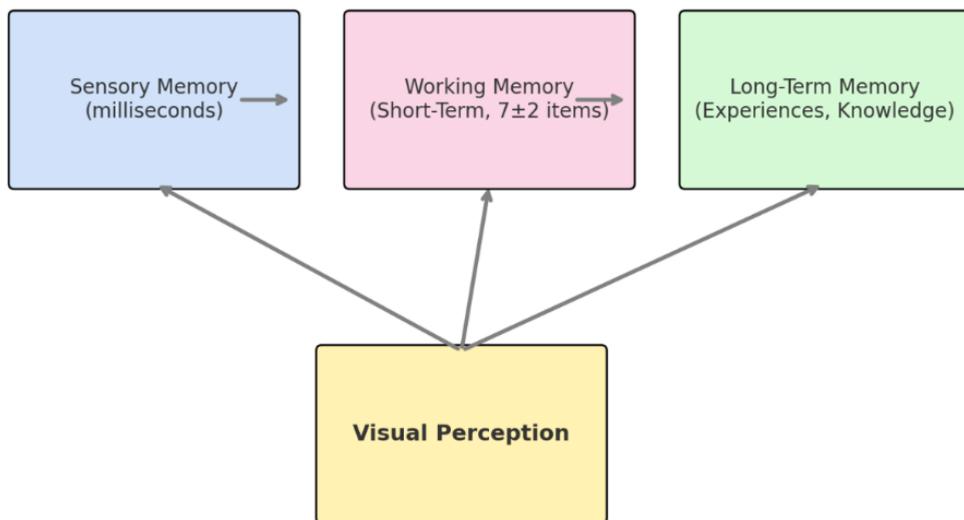

Figure 16. Diagram showing memory types and their influence on perception

Understanding the limited capacity of working memory is essential when designing visualizations. Overloaded graphics overwhelm the memory and hinder information processing. In contrast, a clear and focused design facilitates interpretation.

In a visualization, sensory memory supports rapid detection of basic features (contrast, orientation, large differences), but it fades almost immediately. Working memory supports conscious comparisons, but it is

limited; when a chart forces the viewer to remember one value while searching for another (e.g., legends far away, small multiples without alignment), comprehension slows down and errors increase. Long-term memory helps when visual conventions are familiar (time on the x-axis, bars for comparison, red meaning "attention"), so conventions are not "boring"— they are cognitive shortcuts.

One of the most useful teaching rules here is: if a viewer must repeatedly look back and forth between chart and legend, you are consuming working memory on navigation rather than on insight. Direct labeling, consistent ordering, and stable encodings allow the viewer to focus cognitive effort on reasoning, not on decoding.

Cognitive Load and Visual Hierarchies

The effort required to process information is known as Cognitive Load. **Cognitive load** refers to the mental effort required to process information. In data visualization, this means that overly complex or cluttered visualizations can hinder comprehension.

An example of high cognitive load is an infographic with numerous colors, shapes, and text elements. The volume of information can feel overwhelming and make it difficult to grasp the main message. One way to avoid this is by establishing a visual hierarchy.

Visual hierarchies use preattentive attributes such as color, size, position, and contrast to guide the viewer's attention. A well-designed hierarchy emphasizes the most important information and leads the viewer through the visualization.

Example: In a dashboard, key performance indicators (KPIs) can be highlighted using larger font sizes and high-contrast colors. Secondary information can be displayed in smaller and more neutral tones.
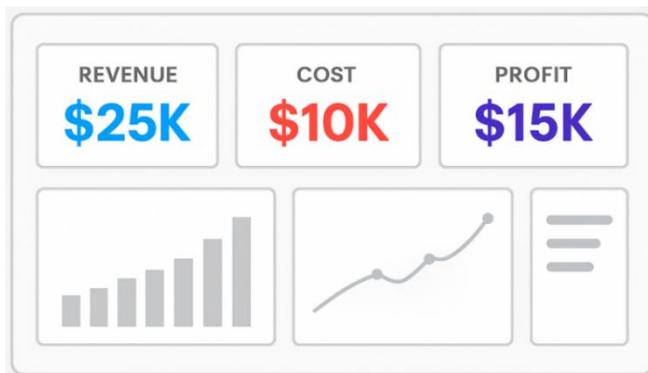
Figure 17. Visual hierarchy in a dashboard

From a data visualization perspective, the effectiveness of visual representations is strongly influenced by the cognitive load they impose on users during information processing and problem solving. Sweller's (1988) Cognitive Load Theory distinguishes between intrinsic, extraneous, and germane cognitive load, emphasizing that learning and reasoning are hindered when unnecessary mental effort is introduced.

1. **Intrinsic Cognitive Load**: The inherent difficulty of the subject matter itself.
2. **Extraneous Cognitive Load**: The load generated by the way information is presented. This is the only type the designer can control.
3. **Germane Cognitive Load**: The effort dedicated to processing information and constructing schemas (mental models) in long-term memory.

Applied to data visualization, this implies that poorly designed charts, cluttered layouts, or unintuitive visual encodings increase extraneous cognitive load, whereas well-structured visualizations reduce mental effort and support pattern recognition, reasoning, and insight generation.

Cognitive Load Theory becomes actionable when you translate it into design control points. You cannot change intrinsic load (a genuinely complex topic), but you can reduce extraneous load by removing non-essential marks, avoiding confusing encodings, and presenting information in a sequence that matches the question being asked. Germane load is the "good effort": the viewer's mental work spent forming a correct understanding of the data, and good design should protect that effort from being wasted on distractions.

A visual hierarchy is the main instrument for reducing extraneous load. Hierarchy answers: What should the viewer notice first, second, and third? In practice, hierarchy is achieved through position (top/left priority), size, contrast, whitespace, and restraint with color. If everything is highlighted, nothing is highlighted; the result is a noisy, flat design that forces the viewer to hunt for the message.

The primary goal of effective visualization design is to minimize Extraneous Cognitive Load. Visual elements that create excessive extraneous load are collectively known as Clutter. Clutter consists of any component that is not essential to the message and ultimately degrades the viewer's ability to process the information. As communications expert Jean-Luc Doumont noted, "Communication is never neutral. It is always either signal or noise/distraction."

Preattentive Attributes and Visual Cues

Preattentive attributes are visual features that the human brain processes almost instantly—within 200 milliseconds—before conscious attention is even engaged. These attributes allow viewers to spot patterns, outliers, or trends at a glance, making them powerful tools for effective data visualization.

Common Preattentive Attributes:
- **Color** (hue, intensity)
- **Orientation** (horizontal, vertical, diagonal)
- **Size** (length, area, volume)
- **Shape** (circle, square, triangle, etc.)
- **Position** (spatial location in a layout)
- **Motion** (for dynamic visualizations)
- **Grouping** (connectedness, enclosure, proximity)

These attributes form the building blocks of **visual cues—signals** that guide the viewer's eye and cognition to the most relevant parts of a visualization.

Preattentive attributes are powerful because they let the brain "compute" differences before conscious attention is engaged. This makes

them ideal for tasks such as spotting outliers, detecting clusters, and noticing changes. However, preattentive cues should be used with discipline: they work best when only one or two cues dominate at a time, otherwise cues compete and the display becomes visually "loud".

Why Are Preattentive Attributes Important?

Preattentive cues are important because they change the cost of reading a chart. When a chart supports "scan first, read later," the viewer can form an overview quickly and then verify details. Without effective cues, the viewer is forced into "read everything," which is slow and demotivating—especially in dashboards where attention is limited and decisions are time-sensitive.

Visualizations compete for attention in environments rich with information. By using preattentive attributes effectively:

- You can **highlight important data** (e.g. trends, anomalies) without requiring users to read every label.
- You can **reduce cognitive load**, allowing viewers to extract meaning faster.
- You help users scan and interpret data efficiently—even in complex visual layouts.
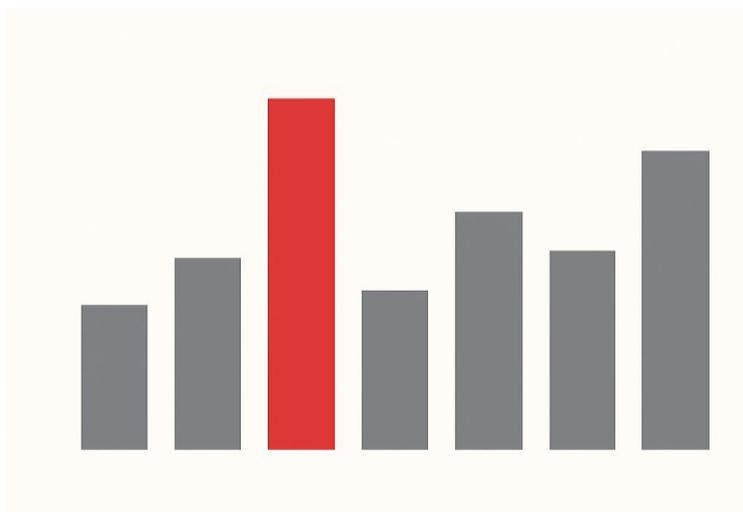
Example 1: Highlighting an Outlier Using Color

Figure 18. *Color as preattentive attribute*

In this bar chart, all bars are gray except one red bar. The red bar immediately draws attention—this is a use of **color as a preattentive attribute** to highlight an outlier. The viewer doesn't need to read axis labels to notice something important is happening there.

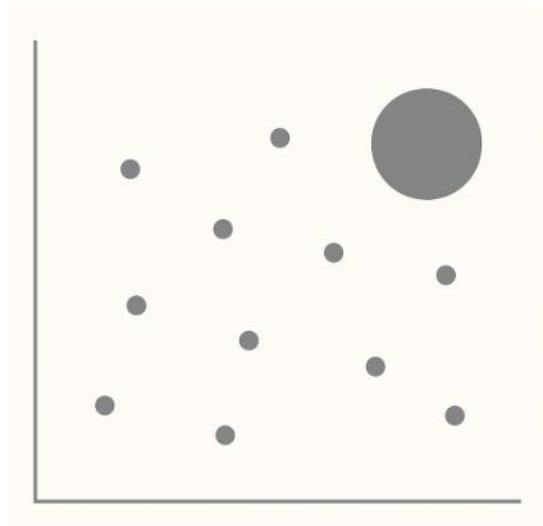Example 2: Emphasizing Size Differences in a Scatterplot

Figure 19. Size as preattentive attribute

Here, all data points are circles of similar size, except for one which is significantly larger. The difference in size serves as a visual cue, signaling the importance or high value of that specific point. The viewer's eye is naturally drawn to it, without requiring textual explanation.

Gestalt Principles and Their Application

Gestalt principles, as introduced by Wertheimer (1923), provide fundamental guidance for designing visualizations that align with human perceptual organization.

The Gestalt principles are fundamental rules of visual perception that describe how we organize visual stimuli into coherent patterns. These principles are particularly useful in designing effective visualizations.

Key Gestalt principles include:

– **Similarity**: Elements with similar attributes (e.g., color, shape) are perceived as belonging together.
– **Proximity**: Elements that are close to each other are perceived as a group.
– **Continuity**: Lines and curves that form a logical path are seen as a unit.
– **Closure**: Our brain tends to complete incomplete shapes.
– **Figure-ground separation**: We distinguish between the main object (figure) and the background.

Concepts such as proximity, similarity, continuity, and closure explain how viewers naturally group visual elements, perceive patterns, and extract meaningful structures from complex data. Applied to data visualization, these principles help designers create charts and graphs that are immediately interpretable, reduce perceptual ambiguity, and enhance the recognition of trends, clusters, and relationships within data sets.

Gestalt principles explain why grouping is often more important than decoration. Proximity, similarity, enclosure, and connection determine what viewers perceive as belonging together. If these grouping signals are inconsistent—for example, if color implies one grouping while layout implies another—viewers experience ambiguity and may infer relationships that are not in the data.

Example: In a stacked bar chart, the principle of similarity is applied by using different colors to clearly distinguish data categories.

Another **example** is a line chart where the principle of continuity guides the viewer along a timeline.
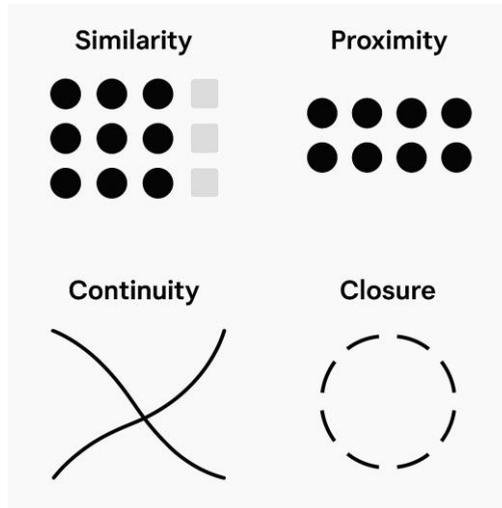
Figure 20. *Visualizations illustrating the Gestalt principles*

Summary and Recommendations

The principles of visual perception form the foundation for designing effective data visualizations. Understanding these principles helps create visualizations that are both intuitively understandable and aesthetically pleasing.

Key design guidelines include:

– **Reduce cognitive load** through clear and simple designs.
– **Use preattentive attributes** to draw attention to key information.
– Apply Gestalt principles to emphasize patterns and relationships in the data.
– **Test visualizations with the target audience** to ensure they are interpreted as intended.

In summary, integrating the principles of visual perception into the design process is essential for creating visualizations that are both functional and visually compelling.

The Gestalt principles provide a theoretical foundation for organizing visual elements in data visualization. They help present information in a way that is intuitively understood.

The principles of visual perception and Gestalt psychology offer valuable insights for designing effective data visualizations. A deeper understanding of these principles enables designers to create visualizations that are both aesthetically pleasing and functional.

Example: The principle of similarity is applied in a line chart where each line is represented with a specific color and style, making it easier for viewers to associate each line with a category or variable.

Another **example** is the principle of proximity: In a stacked bar chart, bars placed close to each other are perceived as related. This principle is often used to group categories or time periods.

In conclusion, integrating these principles into the design process is essential to create visualizations that effectively communicate their message and strategically capture the viewer's attention.

Effective Data Visualization and Design Strategy

In data visualization, effective communication relies on design principles that align with fundamental characteristics of human perception and cognition. Lidwell, Holden, and Butler (2010) compile universal design principles—such as simplicity, consistency, hierarchy, and alignment—that explain why certain visual structures are easier to perceive and understand than others.

Effective data visualization requires a systematic framework that integrates data characteristics, tasks, and visual encoding principles to support accurate interpretation and decision-making. Munzner (2014) presents a comprehensive analysis and design model that emphasizes the importance of understanding the type of data, the goals of the user, and the perceptual effectiveness of visual encodings.

Ware (2013) emphasizes that visual perception principles—such as pre-attentive processing, visual grouping, and pattern recognition—play a critical role in guiding the design of charts, graphs, and interactive interfaces.

Applied to data visualization, these principles support the creation of charts and dashboards that guide attention, reduce perceptual ambiguity, and enhance interpretability, ensuring that visual representations communicate data clearly, efficiently, and with minimal cognitive effort.

Basic Consideration

Effective data visualization is not only a question of aesthetics, but above all of communication. The main goal of any visualization should be to present complex data in a way that can be intuitively understood.

Several basic considerations come into play here:

– **Clarity over complexity**: Simplicity is the key to effective communication. It's better to get a message across clearly than to combine several complicated statements into a single visualization.

- **Data-centricity**: The visualization should put the data first. Superfluous decorative elements that distract attention from the content should be avoided.
- **Target group orientation**: The visualization must be tailored to the needs of the target group. A professional audience needs different visualizations than a broad audience.

For example, a presentation of sales data for a management meeting should focus on the most important trends instead of showing each individual transaction.

A simple time series history could be much more effective here than a detailed graph with hundreds of data points.

Another important aspect is **context**. The data should always be presented in the right context so that viewers can fully understand the meaning of the information.

A practical way to translate "clarity over complexity" into concrete decisions is to separate analysis views from communication views. Analysis views can be dense and interactive because the analyst is willing to invest attention. Communication views must be readable in seconds and should emphasize the conclusion, not the entire dataset. Many visualization failures come from showing an analysis view to a communication audience without adaptation.

Context is part of clarity. A number without a baseline, target, or historical comparison forces the viewer to guess whether performance is good or bad. Wherever possible, add reference information: targets, last period, long-term average, or benchmark groups. This turns "data" into "meaning" with minimal extra ink.

Preattentive Attributes and Their Application

Preattentive attributes are visual characteristics that our brain automatically perceives without us having to actively focus on them. These attributes include color, size, shape, orientation, and movement. These features are invaluable in data visualization because they help highlight important information.

For example, in a scatter plot, critical data points can be highlighted by a high-contrast **color**, such as red. This immediately draws the viewer's attention to the relevant points without the need for additional labels.

Size can also be used as a preattentive attribute. For example, in a bubble chart, where larger bubbles can represent important data. This is especially effective when size differences are applied clearly and consistently.

A <u>wise application of preattentive attributes</u> can significantly **reduce cognitive load** and facilitate the absorption of complex information.

Color Theory in Data Visualization

Color is one of the most versatile tools in data visualization. It can be used to categorize data, highlight patterns, and draw attention. However, the effective use of color requires an understanding of color theory and its principles.

Color is most effective when its meaning is stable and limited. Use categorical palettes for categories (distinct hues), sequential palettes for ordered magnitude (light → dark), and diverging palettes for deviations around a meaningful midpoint (e.g., target, zero). Avoid using many hues for ordered data because the viewer cannot infer order reliably from arbitrary color

The HSV (Hue, Saturation, Value) color model is especially useful for selecting color palettes. It allows the separation of hue, saturation and brightness, which makes it easier to create high-contrast and harmonious visualizations.

Accessibility is part of correctness. If a visualization depends on red/green distinction, a portion of the audience may not decode it reliably; therefore use colorblind-friendly palettes, add redundant cues (labels, patterns, position), and keep contrast strong. Treat color as a data channel, not decoration: every colored pixel should either encode information or guide attention.

A common challenge is taking color blindness into account. About 8% of men and 0.5% of women are colorblind, mostly for red-green

combinations. To ensure that visualizations are accessible to everyone, color schemes such as ColorBrewer's (Brewer, Hatchard, & Harrower (2003).) should be used, which are colorblind friendly.

For example, in a visualization that represents the global temperature change, a color scale from blue (colder) to red (warmer) could be used. This intuitive representation uses the psychological connection between color and temperature.

Design Strategies and De-Cluttering

A good design is characterized by clarity, consistency and accessibility. One of the most effective strategies for improving visualizations is de-cluttering – removing unnecessary elements.

Common elements that can be removed or collapsed include excessive gridlines, duplicate labels, decorative backgrounds, and redundant legends. The aim is to draw attention to the essential data.

An example of successful de-cluttering is the direct labeling of data points in a graph instead of a separate legend. This reduces cognitive load and makes it easier to access information quickly.

Other design strategies include:- Using negative spaces to improve readability.- Applying visual hierarchies to highlight the most important information.- Using consistent fonts and colors to ensure a harmonious appearance.

De-cluttering is not about making charts empty; it is about increasing signal-to-noise. Remove anything that does not support the viewer's task: heavy gridlines, redundant borders, unnecessary 3D, repeated legends, excessive tick marks, and decorative icons. Then reinvest the saved attention into the message: annotate the key point, add a reference line, or label the critical segment directly.

A helpful heuristic is: "Can you explain the chart in one sentence, and can the viewer find evidence for that sentence in under five seconds?" If not, either the story is unclear or the design does not guide attention to the relevant evidence.

Case Studies and Good Practices

Case studies provide practical insights into the application of design principles. A typical case study might include the before-and-after comparison of an overloaded and an optimized visualization

Best practices also include avoiding pie charts in favor of more accurate alternatives such as bar charts, ensuring accessibility through color gamuts for the colorblind, and testing visualizations with the target audience.

A data visualization makeover shows how targeted changes can turn an ineffective visualization into a clear, understandable, and engaging presentation.

Scenario:
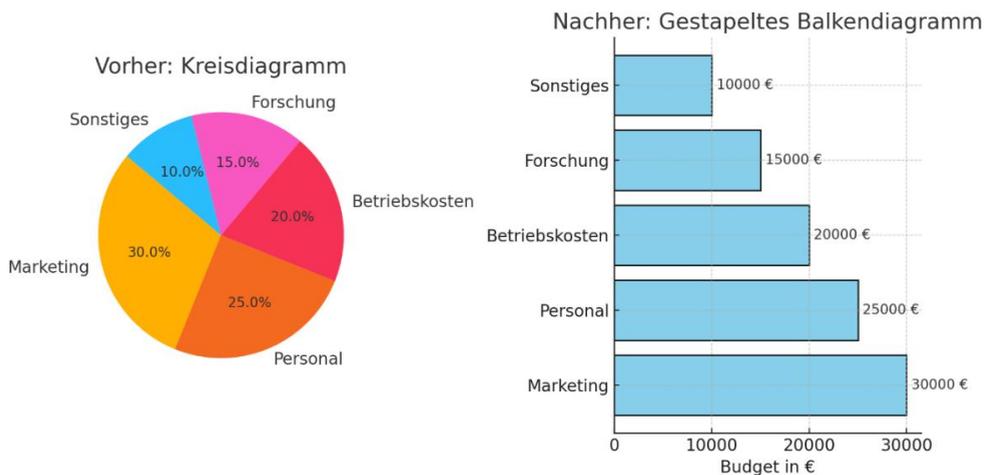
Let's make an example with a budget report:

Figure 21. Example of Makeover

Starting point (before):

A pie chart that shows the budget breakdown for different categories (e.g., marketing, staff, operating expenses):

Problem:

The categories are difficult to compare, especially with similar proportions.

Excessive colors make visualization confusing. There is no direct number in the visualization.

Makeover (after):

A horizontally stacked bar chart:

Clear arrangement of the categories, which allows the viewer to easily compare the values.

The bar lengths directly represent the size of the budget, which is more intuitive to interpret.

Numerical values are embedded directly into the bars, eliminating the need for a separate legend.

A consistent color palette with subtle hues is used to ensure a harmonious display.

An example of **de-cluttering** shows how the reduction of superfluous elements transforms an overloaded visualization into a clearly understandable presentation.

Scenario:

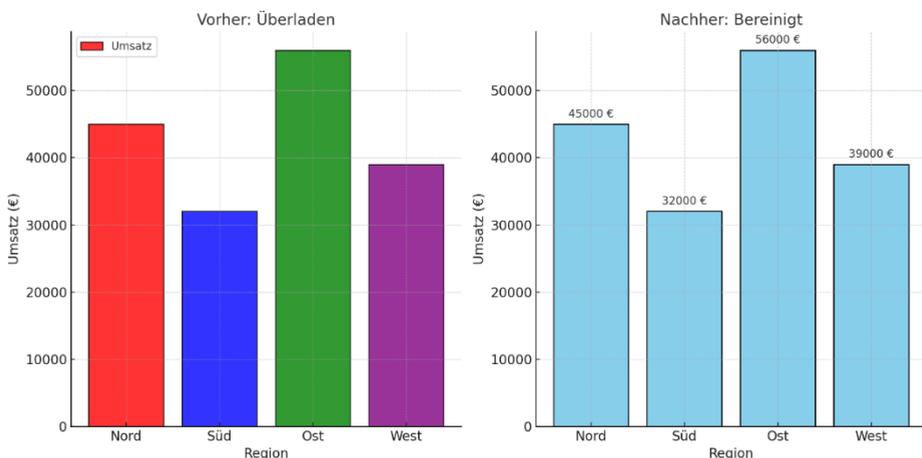A chart shows sales figures of companies in different regions.

Figure 22. Example of De-Cluttering

Before (overloaded):

Grid lines in two directions (horizontal and vertical).

Labels of the axes are redundant and not clearly placed.

A separate legend, although colors could be labeled intuitively.

Unnecessary colors and shadow effects.

After (adjusted):

Remove the vertical grid lines.

Direct labeling of the bars for better readability.

Simplified color palette without distracting decorative elements.

Removal of the separate legend by labeling directly in the visualization.

Scenario:

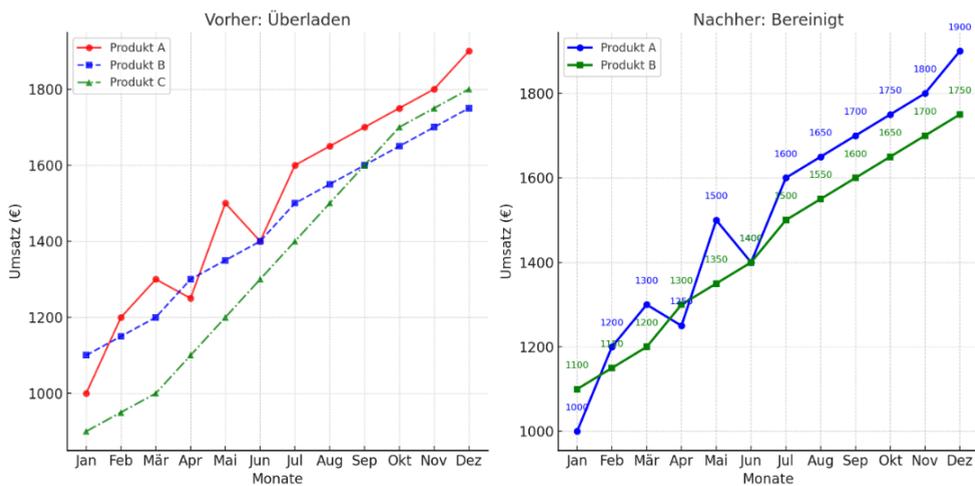The line chart shows the sales history of a company over several months.

Figure 23. Example of De-Cluttering

Before (overloaded):

Too many lines at once, without a clear distinction.

Excessive grid lines that interfere with the display.

Unnecessary shadow effects and a separate legend.

After (adjusted):

Selection of two to three central lines for better focusing.

Direct labeling of the lines to avoid the legend.

Reduction of grid lines and removal of decorative effects.

Clear color palette to improve readability.

Communication Through Visualization

Effective data visualization is not only about accurate representation of data but also about communicating a compelling story that guides the audience to insight. Duarte (2010) emphasizes that visual storytelling transforms audiences by structuring information in a narrative arc, highlighting key messages, and connecting data to meaningful context. Similarly, Knaflic (2015) focuses on the practical application of storytelling principles to data, showing how deliberate design choices—such as ordering, emphasis, and simplification—can make complex datasets understandable and memorable. Together, these works highlight that integrating narrative techniques into data visualization enhances comprehension, engagement, and decision-making by ensuring that users not only see the data but also understand the story it tells.

Introduction to Communication and Data Visualization

Communication through data visualization is much more than just presenting numbers and charts. It is a means of presenting data in such a way that it can be understood, interpreted and used. Effective visualizations are the result of a careful process that combines data analysis, design, and storytelling.

Communication through visualization begins with intent: what you want the audience to think, feel, and do next. Unlike exploratory analysis, communication has a "preferred reading"—a path you want the viewer to follow. Your job as designer is to reduce ambiguity: make the main message easy to see, provide enough context to interpret it, and remove distractions that encourage alternative (and often wrong) interpretations

The goal is to translate complex data into simple and meaningful messages. This requires an understanding of the audience, the context of the data, and the desired impact of the visualization. Communication through

visualization is especially effective when it is based on storytelling principles.

Storytelling provides a narrative structure that organizes information and puts it in a logical flow. Stories can be used to present data not only informatively, but also emotionally appealing, which increases its impact and memorability.

The Role of Storytelling

Storytelling combines data, visualizations, and narrative structures to make a message clear, compelling, and memorable. A well-told story brings data to life and gives it meaning beyond the mere numbers.

Example: A company that wants to showcase the success of a new product might start with an introduction that describes the market challenges. After that, sales figures are shown in line charts that prove the positive development, accompanied by comments on the measures that led to this success.

Storytelling is most useful when the audience needs help connecting evidence to implication. If the purpose is decision support, the story should make the decision obvious: what changed, why it matters, and what action is recommended. A story is not "making up a narrative"; it is structuring true information so the audience can follow the reasoning without getting lost.

The research of Kosara and Mackinlay (2013) shows that narrative visualizations are particularly effective due to their structure and focus on central points. Stories capture attention and make information more memorable.

Elements of an Effective Story

A practical way to keep stories disciplined is to start with the message and work backwards. First write the one-sentence takeaway, then decide which evidence is necessary to justify it, and only then choose the visuals. This prevents the common pattern where the author begins with a dataset, produces many charts, and hopes a story emerges afterward.

A successful data story consists of several key elements:

- **Message**: The central point you want the audience to understand.
- **Narrative structure**: A clear beginning, a challenge or conflict, a climax, and a solution.
- **Visualization**: Data visualizations that support and highlight the key messages.
- **Emotional reference**: Stories that evoke emotions are remembered longer.

One example is the visualization of a country's investment in education. Bar charts show how investments have increased while student performance is increasing. A narrative explanation connects the data to the positive impact on society.

Each of these elements helps ensure that data stories are not only informative, but also compelling and memorable.

Structure of a Data Story

The structure of a data story is crucial for its effectiveness. A well-designed structure should tell the story in logical steps:

- **Set-Up:** An introduction that captures the audience's attention and sets the context.
- **Supporting Facts:** Data and visualizations that support the main message.
- **Main Insight**: The central insight or message of the story.
- **The Solution:** A clear call to action or a solution based on the insights presented.

Example: A data story about climate change could start with a surprising statistic on global temperature (Set-Up), show diagrams of emission developments and their effects (Supporting Facts), point out the importance of immediate action (Main Insight) and finally provide concrete recommendations for action (The Solution).

Techniques for Effective Storytelling

Annotation is often more effective than adding complexity. A single callout ("policy change here", "new product launch", "outlier due to outage") can transform a chart from descriptive to explanatory. Sequencing is the second major technique: reveal information in steps (small multiples, progressive disclosure, or a set of slides) so working memory is not overloaded. Effective storytelling in data visualization requires specific techniques:

– **Big Idea**: Summarize the core message in a single sentence to make it clear and concise.
– **3-minute story**: Conveying the key points of the story in three minutes or less.
– **Storyboarding**: Planning the story in visually structured steps to ensure narrative flow.

An **example** of the Big Idea technique: 'With targeted measures, we were able to increase our customer loyalty by 20%, which resulted in a 15% growth in sales.' This concise statement provides both the context and the central message.

Case Studies of Various Data Visualizations and Their Analysis

Introduction to Visualization Analysis

Data visualization is not only a tool for representing data, but also a method for analysis and communication. Choosing the right visualization is critical to effectively identifying and communicating patterns, relationships, and trends in the data.

This chapter examines different types of data visualizations using case studies. Advantages and disadvantages, possible applications and potential implementation challenges are analysed.

Compare and Display Multiple Variables

Comparing and representing multiple variables is a key challenge in data visualization. Commonly used techniques include:

- **Parallel coordinates**: Ideal for high-dimensional data. Each dimension is represented by an axis, while lines connect data points (Figure 23).
- **Chernoff faces**: A creative method in which variables are represented by features of human faces (Figure 24).
- **Net diagram (also called radar chart, spider chart, or web chart**): compares and displays multiple variables simultaneously by arranging them around a circle and connecting their values**. (Figure 25).

For example, a company analyzes the performance of its products in different markets using parallel coordinates to identify correlations and patterns.
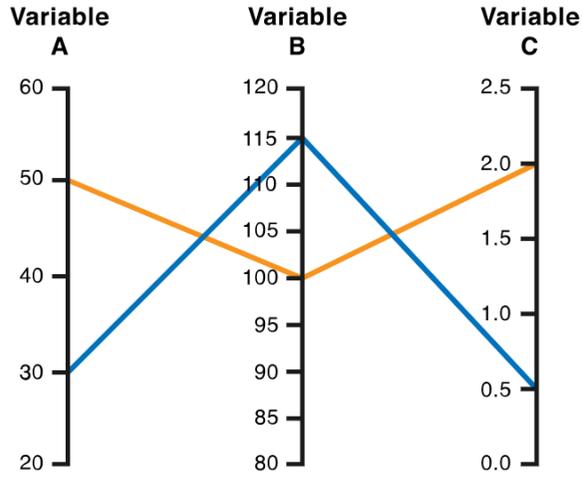
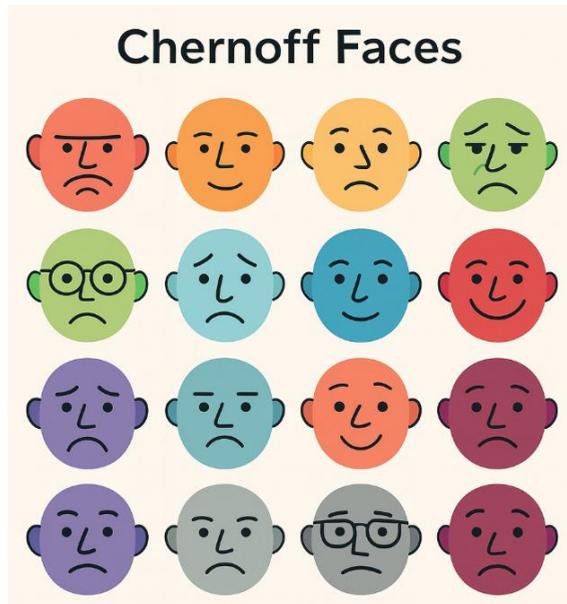| Data			
	Variable A	Variable B	Variable C
Item 1	50	100	2.0
Item 2	30	115	0.5

Figure 24. *Generic example of parallel coordinates*

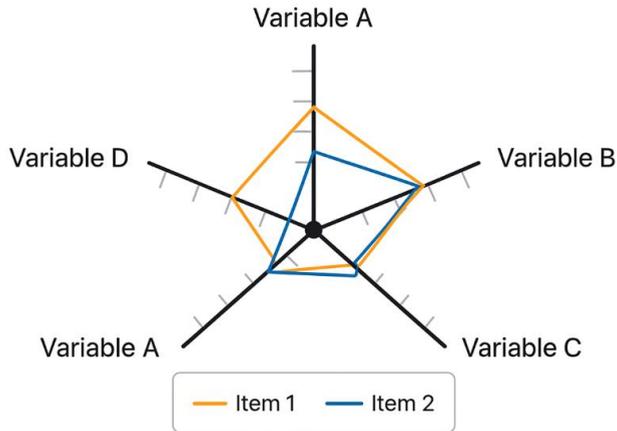**Figure 25.** *An example of Chernoff's faces*

Figure 26. An example of a net diagram

Representing Relationsships

Visualizations such as arc charts and chord charts are especially useful for showing relationships between entities. These techniques help to make networks and connections in data understandable.
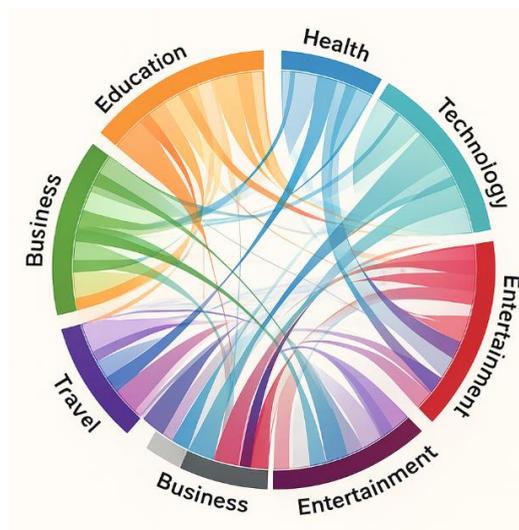

Figure 27. Chord-Diagramm[7]

---

[7] http://bl.ocks.org/nbremer/raw/75c76f4be60fce435aba/

One example is the visualization of exchanges between branches. Here, a chord chart shows how branches correspond one with another.

Advantages of such diagrams are their ability to clearly represent complex relationships. However, a disadvantage is that if there are too many connections, the diagram quickly becomes confusing.

Visualizing Volume Flows

Sankey charts and parallel sets are effective tools for visualizing quantity flows. They are often used to represent flows of materials, energy, or money. A typical example is the visualization of voter movements between parties during an election. A Sankey chart can show how voters switched from one party to another. Challenges in creating such diagrams include accurately coding the flows and avoiding overlaps that could affect readability.
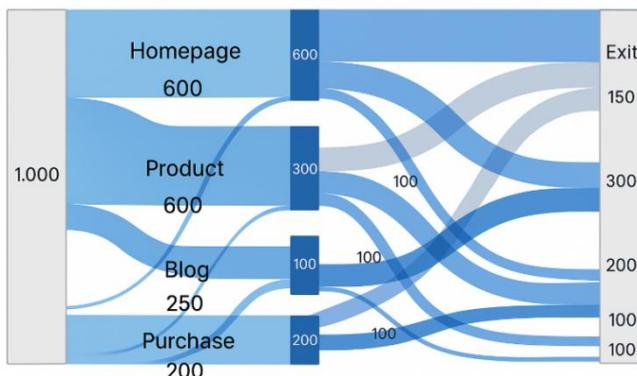

Figure 28. Sankey-Diagram

Displaying Hierarchies and Categories

Hierarchies and categories can be visualized through tree diagrams, sunburst diagrams, or tree maps. These techniques are particularly useful for representing the structure and relationships within data

Example: A company visualizes the organizational structure with a tree diagram that shows the relationships between departments and employees.

Treemaps are particularly effective when data volumes need to be represented by area size, e.g. sales figures per category.

Figure 29. *Treemap Diagram*

Recap and Recommendations

Choosing the right visualization depends on the specific requirements and the target audience. Every visualization technique has its strengths and weaknesses. It is important to thoroughly analyze the data and the intended message before making a decision.

Key recommendations:
- Use parallel coordinates for high-dimensional data.
- Use Sankey diagrams for flow analysis.
- Select treemaps or sunburst charts for hierarchical data.

Dashboards: Visualizing Performance at a Glance

In modern, data-driven business management, dashboards have become an indispensable tool. They are much more than just a collection of diagrams; they act as central hubs for performance management. By translating complex data into an understandable, visual form, they enable decision-makers to quickly grasp the state of a business or process and make informed decisions.

A dashboard is a contract between data and action: it exists to support monitoring and timely intervention. This implies constraints: limited space, consistent encodings, and priority ordering of metrics. If the dashboard does not change what users do (alerts, prioritization, escalation), it is likely a report masquerading as a dashboard.

Dashboards are powerful tools for conveying complex data at a glance, but their effectiveness depends on both clear visual design and alignment with user goals. Few (2012) emphasizes that tables and graphs should be designed to enlighten rather than merely display numbers, highlighting important trends and facilitating rapid comprehension. Complementing this perspective, Yigitbasioglu and Velcu (2012) review dashboard frameworks in accounting and management contexts, showing that dashboards are most effective when metrics are carefully selected, visual encodings are intuitive, and interactive features support exploration and decision-making. Together, these works underline that well-designed dashboards integrate **data accuracy, perceptual principles, and user-centered design** to provide actionable insight quickly and efficiently.

Definition

Speaking of dashboards, data visualization expert Few (2006) provided a concise definition that is still valid today:

"A dashboard is a visual display of the most important information to achieve one or more goals; consolidated and arranged on a single screen, so that the information can be monitored at a glance."

This definition highlights the four key components of an effective dashboard:

1. **Visual display**: It uses graphical elements to make data quickly understandable.
2. **Key information**: It focuses on the key metrics (KPIs) that are relevant to success.
3. **Goal Achievement**: It is directly linked to specific, well-defined business goals.
4. **Single screen**: It presents all relevant information on a single page without the need for scrolling to ensure an instant overview.

From a business perspective, the primary purpose of a dashboard is to provide an ongoing overview of the current situation. It answers key operational questions such as:

– "How's it going right now?"
– "Are we on track?"
– "Are there any problems that require immediate attention?"
– "Do we have to intervene somewhere?"

The focus is clearly on the operational business and the monitoring of defined targets, comparable to the dashboard in a car or airplane, which provides the pilot with the most important information about the controls.

Demarcation: What a dashboard is not

To fully understand the purpose of a dashboard, it's equally important to know what it's not designed for.

The dashboard is:

– **No detailed report**: Traditional reports are better for reading exact numbers, offering more detail and deeper granularity.
– **No analysis tool**: For deeper knowledge and analysis of data, there are specialized data exploration and analysis tools that enable interactive and complex queries.

The following table highlights the key differences between these three tools:

Table 2. Difference between dashboard, report and exploration

Feature	Dashboard	Report	Data Exploration & Analysis
Objective	Ongoing monitoring of the current situation	In-depth investigation of issues	Discovering new, unknown insights
Data Volume	Highly aggregated (KPIs)	Aggregated and detailed	Large, heterogeneous, raw data
Frequency	High (Real-time, Hourly, Daily)	Lower (Weekly, Monthly, Quarterly)	Ad-hoc, as needed
Usage Time	Brief (1-5 minutes)	Longer (15-60 minutes)	Extended (Hours, Days, Weeks)

It is crucial to distinguish dashboards from other common data tools. Each serves a different purpose and is designed for a different mode of interaction.

A practical way to pick a dashboard type is to start from the user role and time horizon. Executives typically need trend + goal tracking (weeks/months) and a small number of high-leverage KPIs. Operational teams need short feedback cycles (hours/days), alerts, and segmentation by region/product/system to locate issues quickly. Analysts need "diagnostic dashboards" with filters and linked views, but those should not be forced onto executive users.

Categories and Applications of Dashboards

Dashboards are not a one-size-fits-all solution. Their design and functionality depend largely on the strategic purpose, the target group and the underlying data context. Clear categorization is therefore the first step in designing an effective dashboard that meets the needs of its users.

Dashboards can be systematized based on various characteristics, ranging from strategic orientation to technical interactivity:

Role:
– **Strategic:** Used by top management to monitor long-term goals and KPIs.
– **Analytical:** Allows subject matter experts to explore trends and gain deeper insights through interaction.
– **Operational:** Helps day-to-day employees monitor real-time processes and respond quickly to deviations.

Type of data:
– **Quantitative:** Numerical data such as sales figures or production volumes.
– **Non-Quantitative:** Qualitative information that provides context.

Domain of data:
– Sales, finance, marketing, production, human resources etc.

Type of measured values:
– **Balanced Scorecard:** Includes key figures from different

Business perspectives (finance, customers, processes, learning).
– **Six Sigma:** Focuses on key figures for process quality and error reduction.
– **Non-Performance:** Shows non-performance data.

Data range:

– Company-wide, cross-departmental or individual.

Update frequency:

– From monthly to hourly or real-time.

Interactivity:

– **Static:** A pure display with no interaction options.

– **Interactive:** Provides features such as drill-down or filters for more detailed analysis.

Presentation mechanism:

– Mainly graphic, mainly text-based, or an integration of both.

Portal functionality:

– Provides redirects to additional data or does not have a portal function.

Primary application areas

In practice, these features can be grouped into three primary application areas, each with specific design requirements:

1. **(Real-time) monitoring:**

– **Goal:** To provide a quick, up-to-date overview of dynamic processes.

– **Design requirement:** The design must find a compromise between analytical potential and efficient summarization. The information must be prepared in such a way that it supports decision-making, even if the topicality changes at short intervals (seconds, minutes or hours).

2. **Analytical tasks:**

– **Objective:** To enable the user to interactively explore data to find root causes and test hypotheses.

– **Design requirement:** Usability must be intuitive, as interaction (e.g., filtering, zooming) is required. Such dashboards often require specific prior knowledge on the part of the user in order to be able to use the tools offered sensibly.

3. **Executive Summary:**
   – **Objective:** To provide management with a highly condensed but meaningful summary of the company's performance.
   – **Design requirement:** Transparency is the top priority here. The level of detail must be skillfully weighed so that the overview is sufficiently informative to encourage decisions without overloading with details.

Typical Key Figures by Business Unit

The practical application of these dashboard types is concretized by the selection of relevant key performance indicators (KPIs). Here are some typical examples of different areas of business:

Sales
– Bookings, billing, sales pipeline (expected sales), number of orders, order quantities, sales prices.

Marketing
– Market share, campaign success, customer demographics.

Finance
– Revenues, costs, profits.

Technical support
– Number of support calls/tickets, resolved cases, customer satisfaction, call length.

Fullfillment
– Number of days to shipment, backorder, stock levels.

Production
– Number of units produced, manufacturing times, number of defects.

Regardless of the type and use case, the effectiveness of a dashboard depends on fundamental design principles. These principles ensure that the visual ads are not only engaging, but more importantly, clear, under-standable, and action-oriented.

Design Principles for Maximum Clarity and Impact

Effective dashboard design is a discipline that applies principles of visual perception and information hierarchy to minimize cognitive load on the viewer and speed up decision-making. It's about organizing and presenting data in such a way that the most important findings immediately catch the eye.

Five fundamental aspects are crucial for a coherent and effective dashboard design:

1. **Metrics:** Choosing the right metrics.
2. **Structure:** The logical structure and arrangement of the elements.
3. **Flow:** The visual guidance of the viewer through the information.
4. **Relationships:** The representation of relationships between the data.
5. **Grouping:** The meaningful summary of related information.

A dashboard should always reflect the nature of the underlying data and tell a story that is relevant to the user.

Choosing the right metrics

The selection of metrics is the foundation for the relevance of a dashboard. A common fallacy is the choice of key figures that are easy to measure but not very meaningful. Therefore, avoid the following pitfalls:

– **Don't be too simple:** Metrics need to be meaningful enough to provide relevant insights.
– **Don't be too complex:** You need to remain understandable without in-depth analysis.
– **Don't be too numerous:** An overload of KPIs distracts from the essentials.
– **Don't be too historical:** The focus should be on current, actionable data.
– **Don't be too vain ("Vanity Metrics"):** They should reflect real performance and not just impress superficially.

Visual design rules for maximum impact

Once the right metrics are selected, the visual design determines success. The following rules are central to this:

1. **Limit to one screen page:** A dashboard should be fully visible without scrolling. Human visual memory is limited. This limitation supports natural visual flow and allows the brain to perceive relationships between different metrics without the cognitive effort of remembering off-screen content.

2. **Set focus:** The most important information must be highlighted most visually. This can be achieved through size, position, or color to create a clear visual **structure** and direct the **flow** of attention in a targeted manner.

3. **Eliminate unnecessary visual elements (de-cluttering):** Every pixel should serve a purpose. Remove redundant data pixels and eliminate any non-data pixels such as redundant grid lines, borders, or decorative effects. This process reduces cognitive load and strengthens the perception of meaningful **relationships** in noise.

4. **Avoid gimmicky or space-intensive graphs:** Visual effects and inefficient chart types often distract from the message. So-called "gauges" (speedometer displays) have a low data density and consume a lot of space. More efficient alternatives such as "bullet graphs" convey the same information in a more compact way and support a clear **structure**.

5. **Avoid 3D and area diagrams:** Our perception is optimized for a 2D world and has difficulty correctly comparing areas and volumes in pseudo-3D representations. This bias undermines the ability to see precise **relationships** between data points. Area diagrams also carry the risk of obscuring important data through overlaps.

6. **Enable quick interpretation:** Data must always be presented in the right context (e.g., through benchmarks, benchmarks, or time series) so that the viewer can grasp its meaning immediately. Only the context establishes meaningful **relationships** and turns pure numbers into actionable information.

Features of an Excellent Dashboard

An excellent dashboard is characterized by the disciplined application of these principles. The analysis of good examples shows recurring characteristics:

- **Economical use of color:** Color is used specifically to encode information or draw attention, not for decoration.
- **Focused main area:** The most important data takes up the most prominent space on the screen.
- **Concise display media:** Compact and information-dense visualizations are used to present a lot of data in a small space.
- **Combination of graphics and text:** Important key figures are often displayed both graphically and as an exact number.
- **Sensible use of white space:** White space is deliberately used to group content and create a clear visual structure.

The best way to understand the concrete application or disregard of these principles is through practical examples that illustrate the differences between a good, a bad and a non-usable design.

In conclusion, an effective dashboard supports decisions and provides a clear overview to achieve business goals. Success does not depend on the complexity of the technology or the number of metrics shown, but on the disciplined application of fundamental design principles that focus on clarity, focus, and comprehensibility.

Tables

Figures

Bibliography

Ackoff, R. L. (1989). From data to wisdom. Journal of Applied Systems Analysis, 16, 3–9.

Brewer, C. A., Hatchard, G. W., & Harrower, M. A. (2003). ColorBrewer in print. Cartography and Geographic Information Science, 30(1), 5–32. https://doi.org/10.1559/152304003100011090

Cairo, A. (2016). The truthful art: Data, charts, and maps for communication. New Riders.

Card, S. K., Mackinlay, J. D., & Shneiderman, B. (1999). Readings in information visualization: Using vision to think. Morgan Kaufmann.

Card, S. K., Moran, T. P., & Newell, A. (1983). The psychology of human–computer interaction. Lawrence Erlbaum.

Cook, M., & Spear, M. (2019). The remarkable story of Florence Nightingale's polar area diagram. Significance, 16(2), 14–19. https://doi.org/10.1111/j.1740-9713.2019.01252.x

Duarte, N. (2010). Resonate: Present visual stories that transform audiences. Wiley.

Few, S. (2006). Information dashboard design. O'Reilly.

Few, S. (2012). Show me the numbers: Designing tables and graphs to enlighten. Analytics Press.

Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., & Melançon, G. (2008). Visual analytics: Definition, process, and challenges. In Information Visualization (pp. 154–175). Springer. https://doi.org/10.1007/978-3-540-70956-5_7

Knaflic, C. N. (2015). Storytelling with data. Wiley.

Kosara, R., & Mackinlay, J. (2013). Storytelling: The next step for visualization. IEEE Computer Graphics and Applications, 33(3), 44–53. https://doi.org/10.1109/MCG.2013.36

Kurgan, L. A., & Musilek, P. (2006). A survey of Knowledge Discovery and Data Mining process models. The Knowledge Engineering Review, 21(1), 1–24. https://doi.org/10.1017/S0269888906000737

Lidwell, W., Holden, K., & Butler, J. (2010). Universal principles of design. Rockport.

Mazza, R. (2009). Introduction to information visualization. Springer. https://doi.org/10.1007/978-1-84800-906-0

McCormick, B. H., DeFanti, T. A., & Brown, M. D. (1987). Visualization in scientific computing. NSF.

Munzner, T. (2014). Visualization analysis & design. CRC Press.

Rahm, E., & Do, H. H. (2000). Data cleaning. IEEE Data Engineering Bulletin, 23(4), 3–13.

Rowley, J. (2007). The wisdom hierarchy. Journal of Information Science, 33(2), 163–180. https://doi.org/10.1177/0165551506070706

Shneiderman, B. (1992). Tree visualization with treemaps. ACM Transactions on Graphics, 11(1), 92–99. https://doi.org/10.1145/102377.115768

Sweller, J. (1988). Cognitive load during problem solving. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tufte, E. R. (1983). The visual display of quantitative information. Graphics Press.

Tufte, E. R. (2006). Beautiful evidence. Graphics Press.

Varian, H. R. (2009). How the Web challenges managers. McKinsey Quarterly.

Ware, C. (2013). Information visualization: Perception for design (3rd ed.). Morgan Kaufmann.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt II. Psychologische Forschung, 4, 301–350. https://doi.org/10.1007/BF00410640

Yigitbasioglu, O. M., & Velcu, O. (2012). A review of dashboards. International Journal of Accounting Information Systems, 13(1), 41–59. https://doi.org/10.1016/j.accinf.2011.08.002

Glossary

Aggregation – Combining individual data points into summarized values.

Preattentive Attributes – Visual properties such as color, size or shape used to encode data.

Cognitive Load – The mental effort required to process information.

Dashboard – A visual interface displaying key performance indicators on a single screen.

Data Encoding – The mapping of data values to visual attributes.

Data Lifecycle – The sequence from data collection to cleaning, transformation, analysis and visualization.

DIKW Pyramid – A hierarchy describing the relationship between Data, Information, Knowledge, and Wisdom.

Gestalt Principles – Perceptual rules that explain how humans group and interpret visual elements.

Heatmap – A visualization using color intensity to represent magnitude.

Interactivity – Ability for users to filter, zoom, hover, or explore data dynamically.

Knowledge Extraction – Process of generating insights from structured or unstructured data.

Preattentive Processing – Immediate, unconscious perception of visual features.

Sankey Diagram – A flow diagram where width encodes volume or quantity.

Treemap – A hierarchical visualization using nested rectangular areas.

Visual Analytics – Integration of automated data analysis and interactive visualization.

Visual Hierarchy – Arrangement of elements that guides attention from most to least important.

Closing Remarks

This book brings together the essential foundations, cognitive principles, and practical strategies of data visualization into a coherent and application-oriented guide. In an age where data shapes nearly every industry, the ability to communicate information clearly and persuasively has become indispensable.

My aim in writing this book was not only to present methods and tools, but also to demonstrate how data visualization can transform complexity into clarity, and how visual storytelling can turn raw numbers into meaningful insights.

I would like to express my gratitude to colleagues, research partners, and academic teams who contributed through discussions, feedback, and collaborative work across projects and institutions. Their expertise and commitment have enriched the depth of this work.

I hope this book supports you—students, practitioners, researchers, and professionals—in creating visualizations that are accurate, expressive, and powerful. May it inspire you to approach data not merely as information, but as an opportunity to understand, communicate, and create impact.

Graz 2026
Selver Softic

Data Visualization:
Fundamentals and Applications

By Selver Softic

In a world driven by data, the ability to transform information into clear, meaningful, and visually compelling insights has become a core competency across industries. Data Visualization: Fundamentals and Applications offers a comprehensive and practice-oriented introduction to the science, psychology, and craft of visualizing data.

This book unites foundational theory with modern techniques, guiding readers through every step of the visualization journey—from data preparation and cognitive principles, to design strategies, storytelling, and dashboard development. Through vivid explanations, historical case studies, and practical examples, readers gain a deep understanding of how visualizations work and how they can be designed to maximize clarity, accuracy, and impact.

Inside this book you will discover:
–   The historical evolution of visualization and its roots in science and design
–   Essential concepts such as data encoding, perception, memory, and cognitive load
–   The most important chart types, visualization methods, and when to use them
–   Techniques for visual storytelling that turn data into compelling narratives
–   Design strategies for clarity, de-cluttering, color theory, and accessibility
–   Practical guidance on dashboards, KPIs, analytical reasoning, and visual analytics

Written for students, educators, analysts, engineers, designers, and decision-makers, this book bridges academic foundations with real-world application. Whether you are new to data visualization or seeking to refine your craft, this book provides the tools to transform data into knowledge, and knowledge into action.

About the Author

Selver Softic is a Professor of Information Systems at CAMPUS 02 University of Applied Sciences in Graz, Austria, specializing in data visualization, AI-driven analytics, semantic technologies, and digital transformation. With extensive academic and industry experience, he has contributed to international research projects, taught at leading universities, and published widely in the fields of data science, visualization, and intelligent systems