Data Visualization

Chapter(6)

Evaluating Visualization Techniques and Issues

Lecturer: Er. Saroj Ghimire
Qualification: Msc.CSIT, BE(COMPUTER)

Overview

1. User and data characteristics
2. Visualization characteristics
3. Structures for evaluating visualizations
4. Visualization bench marking
5. Issues of data, Issues of cognition, Perception and reasoning
6. Issues of hardware and software.

User and data characteristics

- ✓The term **"user"** refers to the individuals or stakeholders who interact with visualized data.
- ✓Users can include data analysts, decision-makers, researchers, or general audiences seeking insights from the visual representations.
- ✓Understanding the needs, preferences, and cognitive abilities of users is crucial in designing effective and user-friendly visualizations that cater to their specific requirements.
- ✓The term **"data"** in data visualization refers to the information or raw material that is being represented visually.
- ✓Data can come in various forms, such as numerical values, text, images, or spatial information.
- ✓It is the foundation of data visualization, providing the content and substance that is transformed into graphical representations.
- ✓Data quality, accuracy, relevance, and organization are key considerations in creating meaningful and informative visualizations.

User Characteristics:

1. **Domain Knowledge:**
   1. Consider the level of expertise and familiarity that users have with the subject matter or domain related to the data. For example, visualizations created for financial analysts may require a deeper understanding of financial concepts than visualizations aimed at a general audience.
   2. Example: When designing a visualization to analyze stock market trends, it is essential to consider that the intended audience consists of experienced traders who are well-versed in financial terminology.
2. **Cognitive Abilities:**
   1. Take into account the cognitive abilities and limitations of the users. Consider factors such as attention span, memory capacity, and visual perception capabilities. Design the visualization to match the cognitive abilities of the target audience.
   2. Example: When creating a visualization for children, it is crucial to use simple and intuitive visual elements that are easy for them to understand and remember.
3. **Familiarity with Visualization Techniques:**
   1. Determine the level of familiarity that users have with different visualization techniques. Avoid using complex or unfamiliar visualization types if the audience is not well-versed in their interpretation.
   2. Example: In an educational setting, when presenting data to students who are new to data analysis, it is advisable to start with basic visualizations such as bar charts and gradually introduce more advanced techniques

Data Characteristics:

1. **Data Type:**
   - Consider the nature of the data being visualized, such as categorical, numerical, temporal, or textual. Different visualization techniques are suitable for different data types.
   - Example: If visualizing the performance of different car models, a bar chart could be used to compare sales figures (categorical data), while a line chart may be employed to show the trend in fuel efficiency over time (temporal data).
2. **Data Structure:**
   - Analyze the structure of the data, whether it is tabular, hierarchical, network-based, or spatial. The visualization technique should align with the underlying data structure for effective representation.
   - Example: When visualizing a hierarchical organization chart, a tree diagram or sunburst chart can effectively represent the hierarchical relationships between departments and employees.
3. **Data Dimensionality:**
   - Consider the dimensionality of the data, whether it is low-dimensional or high-dimensional. High-dimensional data may require dimensionality reduction techniques or interactive visualization methods to convey meaningful insights.
   - Example: When visualizing customer data with multiple attributes (e.g., age, gender, location, purchase history), dimensionality reduction techniques like t-SNE or PCA can be used to reduce the data to two or three dimensions for visualization purposes.

Visualization characteristics

- **Visual Encodings:**
  - Position: The spatial position of visual elements can represent data values. For example, the height of bars in a bar chart represents the magnitude of a variable.

    Example: In a scatter plot, the x and y coordinates of the data points encode two different numerical variables, such as age and income.
  - Length, Size, or Area: The length, size, or area of visual elements can encode quantitative values. For instance, the size of bubbles in a bubble chart can represent population size.

    Example: In a bubble chart representing countries, the size of the bubbles can be proportional to the country's GDP, allowing viewers to compare the economic strength of different nations.
  - Color: Different colors or shades can represent distinct categories or numerical values. For example, a heat map can use color gradients to represent temperature variations.

    Example: In a choropleth map, different shades of color can indicate the population density of regions, with darker shades representing higher density.
  - Shape or Symbol: Different shapes or symbols can be used to differentiate categories or represent specific data points. For instance, different shapes in a scatter plot can indicate different classes of data points.

    Example: In a scatter plot representing flower species, circles can represent one species, triangles another, and squares a third.

Visualization characteristics

➢ Layout and Composition:

- Grids and Axes: Gridlines and axes provide a frame of reference and aid in interpreting the data. They create a visual structure for the visualization and facilitate accurate data comparison.
- Example: A line chart typically includes vertical and horizontal axes with labeled tick marks to indicate specific time points or numerical values.

- Legends and Labels: Legends and labels provide explanations or descriptions for the visual elements used in the visualization. They help viewers understand the meaning behind the visual encodings.
- Example: A bar chart may include a legend that associates different colors with specific categories or labels for each bar, indicating the corresponding data values.

Visualization characteristics

➢ Color Schemes:

- Color Palette: A color palette defines a set of colors that are used consistently throughout the visualization. It ensures visual coherence and aids in differentiating between categories or representing variations in numerical values.
- Example: A color palette with shades of blue can be used to represent different political parties in a map visualization, allowing viewers to easily distinguish between them.

- Color Contrast: Color contrast ensures that visual elements are easily distinguishable and readable. It is essential for viewers with color vision deficiencies and enhances overall accessibility.
- Example: When using color to represent different data points, such as in a line chart, ensure that the chosen colors have sufficient contrast to allow users to distinguish between them, even for those with color vision impairments.

Visualization characteristics

➢ **Interactivity:**

- Tooltips: Tooltips provide additional information when hovering over or clicking on specific data points or elements. They allow users to explore details without cluttering the main visualization.
- Example: A scatter plot may display a tooltip with the precise numerical values of a data point when the user hovers the cursor over it.

- Zooming and panning: Zooming in or out and panning across a visualization allows users to examine specific regions or areas of interest in greater detail.
- Example: In a geographical map visualization, users can zoom in to explore a specific region or zoom out to view a larger area.

Structures for Evaluating Visualizations

- Structures for evaluating visualizations refer to systematic frameworks or methodologies used to assess and analyze the effectiveness, accuracy, usability, and other relevant aspects of visualization techniques.
- These structures provide a structured approach to evaluate and compare different visualizations, enabling researchers, designers, and analysts to make informed decisions about selecting, improving, or comparing visualization techniques.
- These evaluation structures typically consist of a set of criteria, metrics, or guidelines that are applied to assess various aspects of visualizations.
- By using these evaluation structures, the strengths and weaknesses of visualizations can be identified, leading to iterative improvements and informed decision-making.
- The evaluation structures may encompass different dimensions, such as visual effectiveness, accuracy, usability, task support, and user satisfaction. They may involve qualitative assessments through user feedback, interviews, or observations, as well as quantitative measures using metrics or performance indicators.

Structures for Evaluating Visualizations

➢ Task-Driven Evaluation:

- Identify the specific tasks or goals that the visualization aims to support. This could include tasks like trend analysis, outlier detection, or pattern recognition.

- Evaluate how well the visualization technique enables users to accomplish these tasks. Assess factors such as the clarity of the visual representation, the ease of identifying relevant information, and the efficiency of completing the task.

- Example: If the goal is to identify patterns in customer purchasing behavior, evaluate how effectively a scatter plot visualizes the relationship between variables such as age and purchase frequency.

Structures for Evaluating Visualizations

➢ Accuracy and Perceptual Effectiveness:

- Assess the accuracy of the visualization in representing the underlying data. Compare the visualization against the raw data or a known ground truth to ensure the accuracy of the visual encoding.

- Evaluate the effectiveness of the chosen visual encodings in conveying the intended information. Consider factors such as the perceptual tasks required by the user, the clarity of the visual cues, and the ease of interpretation.

- Example: When visualizing the distribution of a numerical variable, compare the accuracy and perceptual effectiveness of a histogram, box plot, and kernel density plot in capturing the shape and characteristics of the distribution.

Structures for Evaluating Visualizations

- **User Feedback and Testing:**
  - Collect user feedback through interviews, surveys, or usability testing to understand users' experience and comprehension of the visualization. Gather insights on their preferences, comprehension difficulties, and suggestions for improvement.
  - Conduct user testing to observe how users interact with the visualization, identify usability issues, and assess their overall satisfaction and understanding of the displayed information.
  - Example: Conducting a usability test where users are asked to perform specific tasks on a line chart, such as identifying peaks or valleys, can provide valuable insights into the effectiveness and usability of the visualization.

Structures for Evaluating Visualizations

- **Comparison to Baselines:**
  - Establish baseline visualizations or existing visualization techniques that are commonly used for the same purpose. Compare the performance and effectiveness of the chosen visualization technique against these baselines.
  - Evaluate factors such as the accuracy, efficiency, user preference, and interpretability of the visualization technique in comparison to the baseline methods.
  - Example: Compare the effectiveness of a scatter plot against a parallel coordinates plot in visualizing multivariate data and determining correlations or patterns among variables.

Visualization Bench Marking

- A visualization benchmark would act as a centralized point for discussion of new and relevant evaluation measures, as well as provide a clear and well-documented set of measures for evaluating visualization system performance.
- It involves defining metrics and criteria for assessment and conducting experiments or tests to measure how well each visualization technique meets those criteria.

Visualization Bench Marking

1. Define Evaluation Criteria: Draw a box labeled "Evaluation Criteria" and list the criteria such as clarity, interpretability, aesthetics, interactivity, and time required.
2. Select Visualization Techniques: Draw a box labeled "Visualization Techniques" and list the chosen techniques such as bar charts, line graphs, scatter plots, and tree maps.
3. Create Visualizations: Draw a box labeled "Create Visualizations" and connect it to the "Visualization Techniques" box. Inside this box, depict the process of creating visualizations using each technique and the dataset.
4. Assess Visualizations: Draw a box labeled "Assessment" and connect it to the "Create Visualizations" box. Within this box, indicate the evaluation of visualizations against the defined criteria. This can involve checking for clarity, interpretability, aesthetics, and other factors.
5. Gather Feedback: Draw a box labeled "Feedback" and connect it to the "Assessment" box. Inside this box, illustrate the process of collecting feedback from the target audience about their preferences and perceptions of each visualization.
6. Rank/Rate Techniques: Draw a box labeled "Ranking/Rating" and connect it to the "Assessment" box and "Feedback" box. Within this box, indicate the process of ranking or rating each visualization technique based on the evaluation criteria and feedback received.
7. Identify Effective Technique: Draw a box labeled "Effective Technique" and connect it to the "Ranking/Rating" box. This box represents the identification of the most suitable visualization technique based on the benchmarking process.

Visualization Bench Marking

Here's an example to illustrate visualization benchmarking:

- Let's say you are working on a data analysis project and need to present your findings to a non-technical audience. You have a dataset with various variables and want to create visualizations that effectively communicate the insights you've gained.
- To benchmark different visualization techniques, you would first define criteria for evaluation. This could include factors such as clarity, interpretability, aesthetics, interactivity, and the ability to highlight key trends or patterns. You might also consider factors like the ease of use and the time required to create each visualization.
- Next, you would select a set of visualization techniques to compare. For example, you could choose bar charts, line graphs, scatter plots, and tree maps as potential options.
- Then, you would create visualizations using each technique based on your dataset. You would ensure that the same dataset is used for all techniques to maintain consistency.
- Once the visualizations are ready, you would assess them against the predefined criteria.
  - For clarity, you might evaluate how well each visualization conveys the main message and whether it is easy to understand.
  - For aesthetics, you could assess the visual appeal and whether the colors and design choices enhance or distract from the information being presented.
- You might also gather feedback from your target audience to determine their preferences and perceptions of each visualization.
- Based on the assessment and feedback, you would rank or rate each technique according to the criteria. This would help you identify the most effective visualization technique for your specific purpose.

Issues of Data

- ✓ Issues of data refer to challenges or problems that can arise in the collection, storage, processing, analysis, interpretation, and use of data.
- ✓ These issues can include data quality and accuracy, data privacy and security, data integration and interoperability, data bias and discrimination, data governance and management, and ethical considerations in data handling.
- ✓ Resolving these issues is crucial to ensure reliable, trustworthy, and meaningful data that can be effectively utilized for decision-making, research, and various applications.

1. Data Quality: Ensuring the accuracy, completeness, and reliability of the data used for visualization. Data errors, missing values, and inconsistencies can affect the validity of the visualizations.
2. Data Preprocessing: Dealing with data cleaning, transformation, and integration tasks to prepare the data for visualization. This includes handling outliers, normalizing data, and resolving data format issues.
3. Data Privacy and Ethics: Addressing concerns related to data privacy, security, and ethical considerations. Sensitive data must be protected, and visualization practices should adhere to ethical guidelines and regulations.

Issues of Cognition

- Issues of cognition pertain to challenges or limitations related to the cognitive processes involved in interpreting and understanding visualized data.
- These issues can encompass difficulties in processing complex visual information, comprehending data relationships, identifying patterns or outliers, and making accurate judgments or decisions based on the visualized data.
- Cognitive issues in data visualization can arise from factors such as information overload, cognitive biases, lack of domain knowledge, or cognitive impairments that affect attention, memory, or problem-solving abilities.
- Cognitive Load: Designing visualizations that manage the cognitive load on viewers' mental resources.
- Complex or cluttered visualizations can overwhelm users, hindering their ability to understand and extract insights from the data.

Issues of Perception

- Issues of perception refer to challenges or biases that can arise in the way individuals perceive and interpret visual representations of data.
- These issues can include misleading or ambiguous visual cues, misinterpretation of scale or proportions, cognitive biases affecting how information is processed, and limitations in visual perception such as color blindness or visual acuity.
- These perceptual issues can potentially lead to misinterpretation or misrepresentation of the data, affecting the accuracy and effectiveness of data-driven insights and decision-making.
    1. Visual Encoding: Selecting appropriate visual variables (e.g., position, size, color) to represent data attributes accurately and effectively. Choosing visual encodings that align with the intended message and prevent misleading interpretations.
    2. Color Perception: Considering color choices and the potential impact on perception. Understanding color blindness and ensuring color schemes are accessible to all viewers.
    3. Gestalt Principles: Applying principles such as proximity, similarity, and continuity to facilitate the organization and grouping of visual elements. Leveraging these principles to support the perception of patterns and relationships in the data

Issues of Reasoning

- Issues of reasoning in data visualization refer to challenges or limitations in the logical and analytical processes involved in making inferences, drawing conclusions, and deriving meaningful insights from visualized data.
- These issues can include errors in logical reasoning, misinterpretation of visual cues, failure to consider alternative explanations or hypotheses, and biases in the interpretation of data patterns.
- Flawed reasoning can lead to incorrect conclusions, flawed decision-making, or the misrepresentation of data-driven insights.
- Addressing these reasoning issues is essential for ensuring accurate and reliable data analysis and decision-making based on visualized data
  - Cognitive Biases: Recognizing and mitigating cognitive biases that can influence reasoning when interpreting visualizations. Biases like confirmation bias, anchoring bias, and framing effects can impact decision-making based on visual insights.
  - Interpretation and Inference: Enabling users to make accurate interpretations and meaningful inferences from visualizations. Providing appropriate context, clear labels, and concise explanations to guide users' reasoning processes.
  - Uncertainty and Complexity: Addressing the challenges posed by uncertain or complex data. Visualizations should appropriately represent uncertainty, handle multivariate data, and support comprehension of intricate relationships.

Issues of hardware and software.

- Issues of hardware and software in data visualization refer to challenges and problems that can arise in the technical infrastructure and tools used to create, process, and display visualized data.
- Hardware issues may include limitations in computational power, insufficient memory or storage capacity, compatibility issues, or hardware failures that can impact the performance and usability of data visualization systems.
- Software issues may involve bugs, glitches, or inadequacies in data visualization software, limitations in functionality, poor user interface design, or issues related to data integration and compatibility with different file formats.
- Addressing these hardware and software issues is crucial to ensure smooth and efficient data visualization processes and to optimize the accuracy and usability of visualized data.

Issues of hardware and software.

- Hardware Issue: Insufficient computational power - Suppose a data visualization project involves processing and visualizing a large dataset with complex calculations. However, the hardware used for the task has limited processing capabilities or insufficient memory. As a result, the visualization process becomes slow, and the system may struggle to handle the data efficiently, leading to delays and potential performance issues.
- Software Issue: Compatibility problems - Imagine a scenario where a data visualization software package is unable to read or properly interpret a specific file format used for the dataset. The software may not support the required data structure or have compatibility issues with the operating system. As a consequence, the data cannot be loaded into the visualization tool, hindering the visualization process and making it challenging to analyze and present the data effectively.

Questions

1. What are some key characteristics of data that are relevant to data visualization?
2. How do visualization characteristics contribute to the effectiveness of data representations?
3. What are some common structures or frameworks used for evaluating visualizations?
4. How can visualization benchmarking help assess the performance and quality of visualizations?
5. What are some issues related to data that can impact data visualization processes?
6. How do issues of cognition affect the interpretation and understanding of visualized data?
7. What role does perception play in data visualization, and what are some related issues that can arise?
8. How do reasoning issues influence the accuracy and reliability of insights derived from visualized data?
9. What are some examples of hardware issues that can affect the performance of data visualization systems?
10. What are software-related challenges that can arise in data visualization and how do they impact the usability of visualization tools?

Questions

11. If all data visualizations are based on data, can a visualization exist without any underlying data?
12. If a data visualization is ineffective in conveying information, can it still be considered successful?
13. If two data visualizations have the same accuracy but differ in simplicity, which one would be preferred?
14. If a visualization receives a high benchmark score, does that guarantee its usefulness in decision-making?
15. If an individual has a cognitive impairment, can they still accurately interpret and understand visualized data?
16. If perception issues are not addressed in data visualization, can it lead to incorrect conclusions or judgments?
17. If reasoning errors occur in the interpretation of visualized data, how might they impact decision-making?
18. If a data visualization tool requires high computational power, can it run smoothly on low-end hardware?
19. If software compatibility issues prevent the loading of a dataset into a visualization tool, can alternative solutions be used?
20. If a data visualization addresses hardware and software issues but lacks data quality, can it be considered reliable?