Data Visualization

Chapter(1)

Introduction

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

Overview

- Introduction of visual perception
- Visual representation of data
- Data abstraction,
- Visual encodings, Use of color, Perceptual issues, Information overloads

Introduction of visual perception

- Visual perception is the process by which the human brain interprets visual information from the environment.
- It involves a complex interplay between the physical properties of light and the neural mechanisms that process this information in the brain.
- Visual perception plays a crucial role in our everyday lives, as it allows us to make sense of the world around us and to interact with it effectively.
- An example of visual perception in action can be seen when we look at a simple object, such as a ball. When we see a ball, our eyes detect the light that is reflected off its surface and send this information to the brain. The brain then processes this information to create a visual representation of the ball, including its size, shape, color, and texture.

Introduction of visual perception

- Visual perception also allows us to perceive depth and distance. For example, when we look at a mountain range, our brain uses cues such as perspective, shadow, and interposition to determine which peaks are closer and which are farther away.
- This information is used to create a 3D representation of the scene, which allows us to navigate and interact with the environment effectively.
- In addition to allowing us to perceive and interact with the physical environment, visual perception also plays a crucial role in social communication.
- Facial expressions, body language, and other visual cues are used to convey emotions, intentions, and social status, allowing us to communicate and interact with others effectively.

Visual representation of data

- Visual representation of data involves the use of graphical elements to present complex data sets in a way that is easy to understand and interpret.
- This can include various types of charts, graphs, diagrams, and other visualizations that allow users to quickly grasp key insights from large amounts of data.
- Here are some examples of different types of visual representation of data:
  - Bar Charts: A bar chart is a simple and effective way to compare the values of different categories of data. For example, a bar chart can be used to visualize the sales figures of different products, where each product is represented by a vertical bar with the height of the bar corresponding to the sales value.
  - Line Graphs: A line graph is a type of chart that shows the relationship between two variables over time. For example, a line graph can be used to visualize the trend of stock prices over a period of time, where the x-axis represents time and the y-axis represents the price.
  - Scatter Plots: A scatter plot is a type of chart that shows the relationship between two variables. For example, a scatter plot can be used to visualize the relationship between a person's age and their income, where each data point represents an individual and the x-axis represents age and the y-axis represents income.

Visual representation of data

- Pie Charts: A pie chart is a circular chart that is divided into slices to represent the proportion of each category of data. For example, a pie chart can be used to visualize the market share of different companies in a particular industry, where each slice represents a company and the size of the slice represents the proportion of the market share.
- Heat Maps: A heat map is a type of chart that uses color to represent the intensity of a variable. For example, a heat map can be used to visualize the temperature distribution of a region, where the color of each cell represents the temperature value.
- Tree maps: A tree map is a type of chart that represents hierarchical data using nested rectangles. For example, a tree map can be used to visualize the organizational structure of a company, where each rectangle represents a department and the size of the rectangle represents the number of employees in the department.
- Visual representation of data can be a powerful tool for communicating complex information to users in a clear and concise way. By choosing the right type of visualization for the data and effectively using visual encodings such as color, size, and shape, designers can create visual displays that are easy to interpret and provide valuable insights.

Data abstraction

- Data abstraction is the process of simplifying complex data by focusing on the most important features or characteristics, while ignoring or hiding irrelevant details. In the context of data visualization, data abstraction involves reducing the amount of data presented to the user to provide a clearer, more concise view of the underlying patterns or trends. Here are some

1. Aggregation:
   - Aggregation involves combining multiple data points into a single summary value. This is often done to provide a higher-level view of the data, or to make it easier to compare values across different groups.
   - Example: A bar chart that shows the sales figures for different products can be aggregated by product category to provide a summary of the total sales for each category.
2. Sampling:
   - Sampling involves selecting a subset of the data for visualization. This is often done to provide a representative view of the data, or to reduce the amount of data being presented.
   - Example: A scatter plot that shows the relationship between a person's age and their income can be sampled to include only a subset of the population, such as individuals within a certain age range.
3. Filtering:
   - Filtering involves selecting a subset of the data based on certain criteria. This is often done to focus on a specific aspect of the data, or to remove outliers or irrelevant data points.
   - Example: A heat map that shows the temperature distribution of a region can be filtered to show only temperatures above a certain threshold, or to exclude data points from areas with inconsistent or unreliable temperature readings.

Data abstraction

3. Clustering:
   - Clustering involves grouping similar data points together based on their characteristics. This is often done to reduce complexity and provide a more focused view of the data.
   - Example: A tree map that shows the organizational structure of a company can be clustered to group departments with similar functions or responsibilities, such as grouping all marketing-related departments together.
4. Dimensionality Reduction:
   - Dimensionality reduction involves reducing the number of variables used to represent the data. This is often done to simplify the data and improve visual clarity.
   - Example: A line graph that shows the trend of stock prices over time can be simplified by only showing the closing price of the stock at the end of each day, rather than including all of the intermediate price values throughout the day.

- Overall, data abstraction techniques are important in data visualization because they help to simplify complex data and provide a more focused view of the most important aspects of the data. By using techniques such as aggregation, sampling, filtering, clustering, and dimensionality reduction, designers can create visualizations that are easier to understand and interpret, while still conveying the most important insights from the data
- Data abstraction is an important technique in data visualization because it allows designers to focus on the most important aspects of the data while reducing clutter and improving the user's ability to identify patterns or trends. However, it is important to strike a balance between data abstraction and data fidelity, as removing too much data can lead to loss of important insights

Visual Encodings

- Visual encodings are the visual cues or symbols used to represent data values in a visual display.
- This can include different types of shapes, colors, sizes, and textures that are used to represent different types of data.
- Choosing appropriate visual encodings is important for creating effective visualizations, as it can help users to quickly and accurately interpret the data.
- For example, in a scatter plot, the x and y axes may be encoded with different colors or symbols to represent different types of variables.

Visual Encodings

- Here are some examples of visual encodings in data visualization:

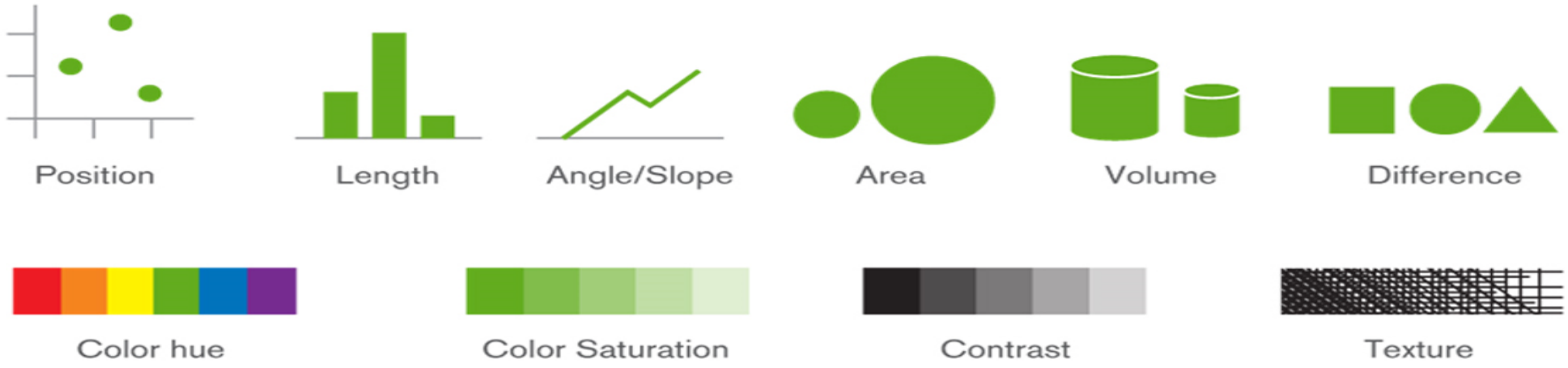
1. Position:
   - Position is one of the most basic visual encodings, and involves mapping data to the x and y coordinates of a visual element. This encoding is often used in scatter plots, line graphs, and bar charts to show the relationship between two variables.
   - Example: A scatter plot that shows the relationship between a person's age and their income uses position encoding to map each data point to a specific location on the x and y axes.
2. Size:
   - Size encoding involves mapping data to the size of a visual element, such as the area of a circle or the width of a bar. This encoding can be used to highlight differences in magnitude or to emphasize important data points.
   - Example: A bubble chart that shows the population and GDP of different countries uses size encoding to map each bubble to the total population of the country, with larger bubbles representing larger populations.
3. Color:
   - Color encoding involves mapping data to the hue, saturation, or brightness of a visual element. Color encoding can be used to highlight differences in categorical data, or to convey information about the magnitude or value of a quantitative variable.
   - Example: A heat map that shows the temperature distribution of a region uses color encoding to map each data point to a specific color on a gradient, with hotter temperatures represented by warmer colors like red and orange.

Visual Encodings

4. Shape:
   - Shape encoding involves mapping data to the shape of a visual element, such as using different symbols or icons to represent different categories or data points. Shape encoding can be useful for highlighting differences in categorical data or for creating distinctive visual patterns.
   - Example: A scatter plot that shows the relationship between a person's height and weight can use different symbols or shapes to represent different genders, such as circles for females and squares for males.
5. Texture:
   - Texture encoding involves mapping data to the texture or pattern of a visual element, such as using diagonal lines or dots to represent different categories or data points. Texture encoding can be useful for creating visual interest and for highlighting differences in categorical data.
   - Example: A choropleth map that shows the distribution of different types of vegetation in a region can use different patterns or textures to represent different vegetation types, such as diagonal lines for grasslands and dots for forests.

- visual encodings are an important aspect of data visualization because they allow designers to effectively communicate information about the data through visual means.
- By choosing the appropriate visual encoding for the data, designers can create visualizations that are more engaging, informative, and memorable for the viewer.

Visual Encodings



Visual Encodings

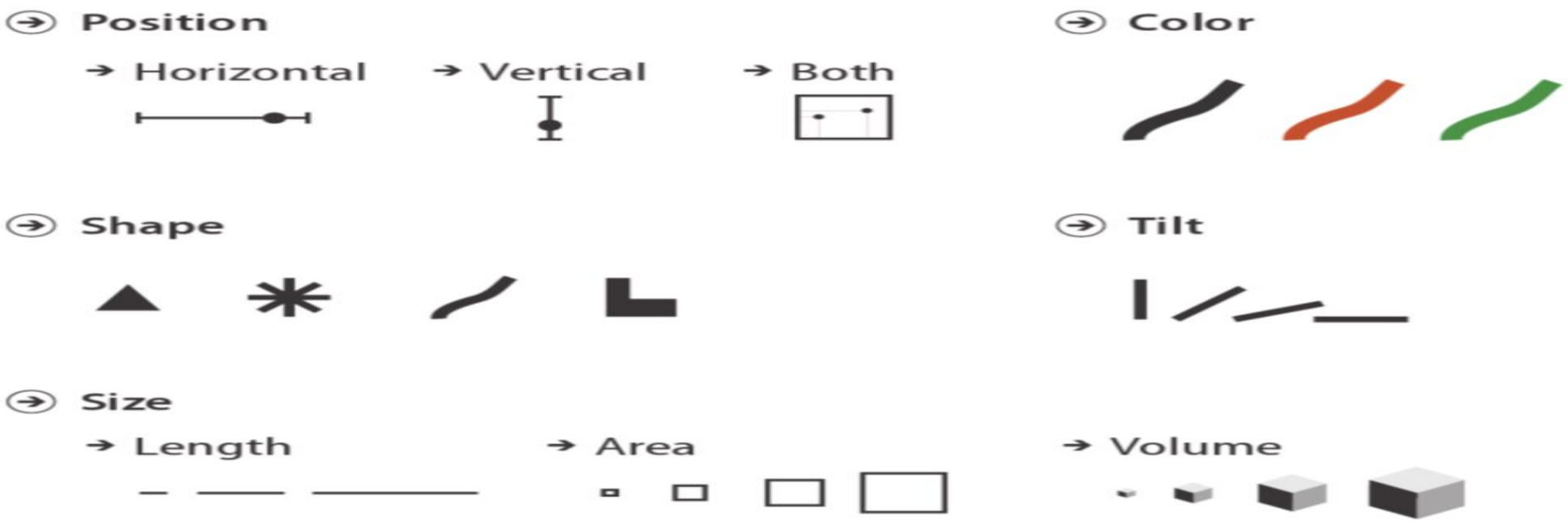


Figure 5.3. Visual channels control the appearance of marks.

Use of color

Color is a powerful tool for visual representation of data, as it can be used to highlight important information and to differentiate between different types of data.

However, the use of color can also introduce perceptual issues, such as color blindness or visual fatigue.

When using color in visualizations, it is important to choose colors that are easy to distinguish and that are appropriate for the target audience.

For example, a heat map may use a range of colors to represent different levels of data, but the colors chosen should be easily distinguishable by users with different types of color vision

Use of color

Color is a powerful tool for data visualization as it can be used to convey different types of information, such as highlighting specific data points, indicating the magnitude of a variable, or distinguishing between different categories. However, the use of color in data visualization requires careful consideration, as inappropriate or inconsistent use of color can lead to confusion and misinterpretation of the data. Here are some examples of how color can be used in data visualization:

- Highlighting specific data points:
  - Using a bright or contrasting color to highlight specific data points can draw attention to important information and make it easier for the viewer to identify patterns or outliers.
  - Example: In a scatter plot that shows the relationship between two variables, using a red color to highlight the data points that fall outside of a certain range can help to identify outliers or anomalies.
- Indicating the magnitude of a variable:
  - Using a color scale to represent the magnitude of a variable can make it easier for the viewer to understand the data and identify patterns or trends.
  - Example: In a heat map that shows the distribution of a variable across a geographic region, using a color scale to represent the magnitude of the variable can help to identify areas with high or low values.
- Distinguishing between different categories:
  - Using different colors to represent different categories can make it easier for the viewer to distinguish between different types of data and understand the relationships between them.
  - Example: In a bar chart that shows the sales of different products, using different colors to represent each product can make it easier to compare the sales of each product and identify the most popular or profitable items.

Use of color

When using color in data visualization, there are some important considerations to keep in mind:

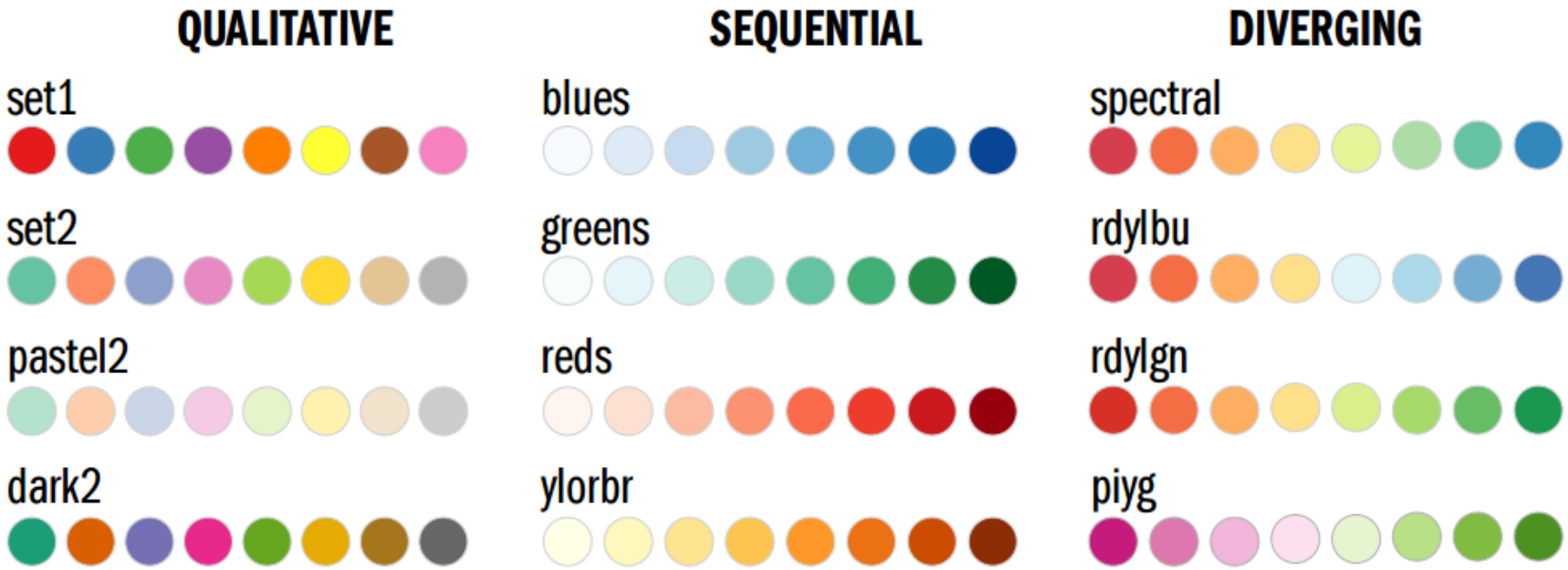
- Color blindness:
  - Approximately 8% of men and 0.5% of women have some form of color blindness, which can affect their ability to distinguish between certain colors.
  - To ensure that colorblind viewers can still understand the data, it is important to use color schemes that are accessible and avoid using colors that are easily confused with one another.
- Consistency:
  - Using consistent color schemes throughout a data visualization can make it easier for the viewer to understand the data and identify patterns or relationships.
  - Using too many colors or changing color schemes can create confusion and make it more difficult for the viewer to understand the data.
- Context:
  - The use of color should always be considered in the context of the data being presented and the goals of the visualization.
  - For example, using bright, attention-grabbing colors may be appropriate for highlighting important data points, but may be distracting or overwhelming in other contexts.

Overall, color can be a powerful tool for data visualization, but it should be used thoughtfully and consistently to ensure that

Use of color


COLOR THEORY
RED
DANGER / LOVE
ORANGE
JOYFUL / CREATIVE
YELLOW
ENERGY / CAUTION
GREEN
MONEY / GROWTH
BLUE
TRUST / DETACHED
PURPLE
ROYAL / MAGICAL
PINK
CALMNESS / OPTIMISM
BLACK
MYSTERY / DEATH

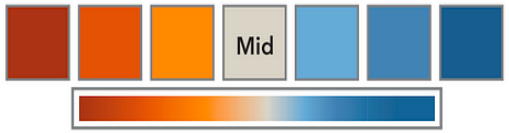
Use of color


QUALITATIVE
set1
set2
pastel2
dark2
SEQUENTIAL
blues
greens
reds
ylorbr
DIVERGING
spectral
rdylbu
rdylgn
piyg

Use of color

SEQUENTIAL
color is ordered from low to high

DIVERGING
two sequential colors with a neutral midpoint



CATEGORICAL
contrasting colors for individual comparison

HIGHLIGHT
color used to highlight something

ALERT
color used to alert or warn reader

- **Sequential color** is the use of a single color from light to dark. An example is encoding the total amount of sales by the state in blue, where the darker blue shows higher sales and a lighter blue shows lower sales.
- **Diverging color** is used to show a range of diverging from a midpoint. It can also be used to show the weather, with blue showing the cooler temperatures and red showing hotter temperatures. The midpoint can be the average, the target, or zero in cases where there are positive and negative numbers..
- **Categorical color** uses different color hues to distinguish between different categories.
- **Highlight color** is used when there is something that needs to stand out to the reader, but not alert or alarm them. For instance as in highlighting a certain data point, text in a table, or a specific bar in a bar chart.
- **Alerting color** is used when there is a need to draw attention to something for the reader. It's often best to use bright, alarming colors.

Perceptual issues

- Perceptual issues can arise in visual representation of data when the visual cues used to represent data values are not aligned with the way the human brain processes visual information. For example, certain types of visual encodings may be more difficult to interpret than others, or the use of certain colors may be difficult for some users to distinguish. To address perceptual issues, designers can use techniques such as data scaling, labeling, and filtering to help users interpret the data more effectively.

Perceptual issues

Perceptual issues refer to potential problems that can arise in data visualization due to the limitations of human perception. Even the most well-designed visualization can be rendered ineffective if the human visual system cannot process it accurately. Here are some examples of perceptual issues in data visualization:

- Size perception:
  - Humans tend to perceive differences in size more easily than differences in other visual properties, such as color or shape. However, our ability to accurately judge the relative size of objects can be influenced by a variety of factors, such as distance, angle, and orientation.
  - Example: In a bar chart, using bars of different heights to represent the magnitude of a variable may be more effective than using different colors or shapes to convey the same information, as size is a more easily perceptible variable.
- Color perception:
  - While color can be a powerful tool for data visualization, our ability to accurately perceive differences in color can be influenced by a variety of factors, such as lighting, background color, and color blindness.
  - Example: Using a color scale to represent the magnitude of a variable may be effective in some contexts, but may be less effective if the viewer is colorblind or if the color scale is not clearly visible against the background.

Perceptual issues

- Pattern perception:
  - Humans tend to perceive patterns and textures more easily than individual objects or shapes, and these patterns can be used to convey information in data visualization. However, our ability to accurately perceive patterns can be influenced by factors such as size, complexity, and orientation.
  - Example: Using a heat map to represent the distribution of a variable may be effective if the viewer can easily perceive the pattern, but may be less effective if the map is too small or if the pattern is too complex to be easily perceived.
- Grouping and proximity perception:
  - Humans tend to perceive objects that are close together as belonging to the same group or category, and this grouping can be used to convey information in data visualization. However, our ability to accurately group objects can be influenced by factors such as distance, size, and orientation.
  - Example: Using a scatter plot to represent the relationship between two variables may be effective if the data points are clearly grouped based on their values, but may be less effective if the points are too close together or if there are outliers that do not fit the pattern.

In general, when designing data visualizations, it is important to consider the limitations of human perception and to use visual encodings that are easy to perceive and interpret. This may involve using multiple visual encodings, such as size, color, and shape, to convey the same information, or using interactive visualizations that allow the viewer to explore the data in different ways. By taking these perceptual issues into account, data visualizations can be designed to accurately convey complex information and facilitate better decision making.

Information overload

- Information overload occurs when a visual display presents too much information at once, making it difficult for users to effectively process and interpret the data. This can occur when the visual display is too cluttered or when too many variables are presented at once.
- To avoid information overload, designers can use techniques such as data abstraction, filtering, and hierarchical organization to help users focus on the most important information.
- For example, a financial dashboard may use a tree structure to present financial data, with the most important variables at the top level and more detailed information at lower levels

Information overload

Information overload can occur in any type of visual display, from simple charts and graphs to complex dashboards and reports.

When a visual display presents too much information at once, it can be difficult for users to identify patterns and relationships, make informed decisions, or draw meaningful conclusions.

Here are some examples of techniques that designers can use to avoid information overload:

- Data abstraction:
  - Data abstraction involves reducing the complexity of a dataset by removing irrelevant or redundant information, while preserving the most important aspects of the data. This can help users focus on the most relevant information and avoid being overwhelmed by too much data.
  - Example: A stock market dashboard may use data abstraction to show only the most important stocks and their key metrics, such as price, volume, and change, rather than displaying data for every stock in the market.
- Filtering:
  - Filtering involves selectively displaying data based on user-defined criteria, such as time period, location, or category. This can help users focus on specific subsets of the data and avoid being overwhelmed by too much information.
  - Example: An e-commerce website may allow users to filter products by price, brand, or customer rating, to help them find the products that best meet their needs.

Information overload

- Hierarchical organization:
  - Hierarchical organization involves organizing data into a nested structure, with the most important information at the top level and more detailed information at lower levels. This can help users navigate through large amounts of data and focus on the most relevant information.
  - Example: A marketing dashboard may use hierarchical organization to show overall campaign performance at the top level, with more detailed metrics such as click-through rates and conversion rates at lower levels.
- Interactive visualization:
  - Interactive visualization involves allowing users to interact with the data display, such as by zooming in and out, panning, or drilling down into the data. This can help users explore the data in more detail and focus on the most relevant information.
  - Example: A geographic information system (GIS) may allow users to interactively explore maps and data layers, such as by zooming in and out, selecting different data layers, or displaying data for specific geographic areas.
- In summary, avoiding information overload in data visualization involves using a variety of techniques to help users focus on the most relevant information and avoid being overwhelmed by too much data. These techniques can include data abstraction, filtering, hierarchical organization, and interactive visualization, among others. By carefully designing data visualizations that are tailored to the needs of users, designers can help ensure that the data is effectively communicated and interpreted, leading to better decision making and improved outcomes.

Assignment

1. How does the brain interpret and process visual information?
2. What are the different types of visual representations of data, and how can they be used to convey information effectively?
3. How can data abstraction techniques be used to simplify complex datasets and make them more easily understandable?
4. What are some common visual encodings used in data visualization, and how do they impact the clarity and accuracy of the information presented?
5. How can the use of color in data visualization impact the viewer's perception of the information?
6. What are some common perceptual issues that can arise in data visualization, and how can they be avoided or addressed?
7. How can information overload be prevented in data visualization, and what techniques can be used to help viewers focus on the most important information?

Assignment (class)

Here is an example of weather data that you can use to create a data visualization of hot temperatures:

Date	Location	Temperature (°C)
01/01/2023	New York City	7
01/01/2023	Los Angeles	25
01/01/2023	Miami	28
01/02/2023	New York City	10
01/02/2023	Los Angeles	27
01/02/2023	Miami	30
01/03/2023	New York City	9
01/03/2023	Los Angeles	29
01/03/2023	Miami	32
01/04/2023	New York City	12
01/04/2023	Los Angeles	30
01/04/2023	Miami	33

You can use this data to create a data visualization that shows the hottest temperatures across different locations over time. One way to do this is to create a line chart with the dates on the x-axis and the temperature on the y-axis, with different lines representing each location. Another way to visualize this data is to create a heat map that shows the hottest temperatures in different colors, with each location represented as a column and the dates as rows. The choice of visualization depends on what story you want to tell with the data and what insights you want to highlight. Use different way of data visualization techniques to represent above data.