

MTH 208 Exploratory Data Analysis

Lesson 05: Color Theory & Accessibility in Visualization

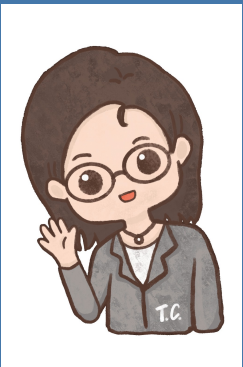
Ying-Ju Tessa Chen, PhD

Associate Professor
Department of Mathematics
University of Dayton

 @ying-ju

 ying-ju

 ychen4@udayton.edu



Learning Objectives

In this lesson, we will talk about the fundamental principles of color theory and how to design visualizations that are accessible to a broader audience, including individuals with color vision deficiencies.

- Basics of color theory
- Designing for color-blindness and accessibility

Basics of color theory

We will learn the fundamental concepts of color theory and its application in data visualization.

Color Wheel and Color Relationships

Definition and History: The color wheel is a fundamental tool in the study of color theory, serving as a visual representation of colors arranged according to their chromatic relationship. Its origins can be traced back to the work of Sir Isaac Newton in the late 17th century. Newton conducted experiments with a prism and discovered that when light passes through it, a spectrum of colors emerges, forming what he called the "color circle." This was a groundbreaking realization that colors could be organized in a circle based on their wavelength, laying the groundwork for what would become the modern color wheel.

The History of The Color Wheel



Color Wheel and Color Relationships (Continued)

Structure of the Color Wheel: The color wheel is structured into three main types of colors: primary, secondary, and tertiary.

- **Primary Colors:** Red, blue, and yellow are considered primary colors because they cannot be created by mixing other colors together. They are the foundation from which all other colors are derived.
- **Secondary Colors:** When two primary colors are mixed together in equal parts, they produce secondary colors. The secondary colors are green (blue + yellow), orange (red + yellow), and purple (red + blue). These colors are positioned between the primary colors on the color wheel.
- **Tertiary Colors:** Tertiary colors are created by mixing a primary color with a secondary color, resulting in a blend that is named by combining the names of the primary and secondary colors (e.g., red-orange, yellow-green). This results in six tertiary colors, each of which is located between a primary and a secondary color on the wheel.

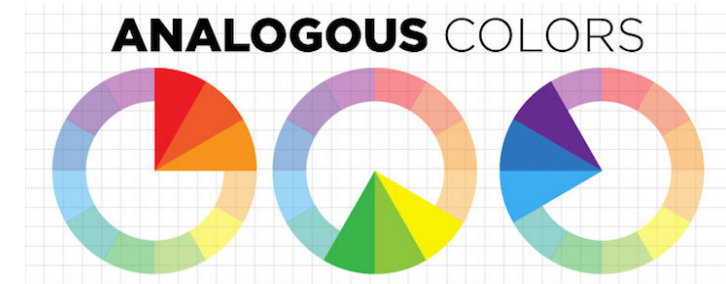
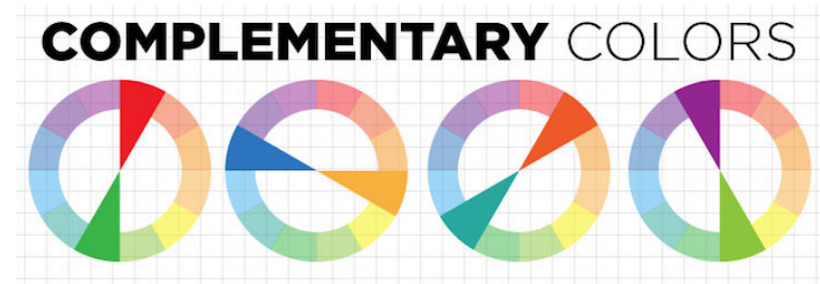


Color Wheel and Color Relationships (Continued)

Color relationships: The arrangement of colors on the wheel allows us to understand their relationships more deeply:

- **Complementary Colors** are directly opposite each other on the color wheel. This relationship is useful for creating contrast and highlighting elements within visualizations. However, when used in close proximity (e.g., text on a background), it can cause visual strain.
- **Analogous Colors** are next to each other on the wheel. This scheme provides a harmonious and cohesive look, often found in nature, making it suitable for creating calm and unified visualizations.
- **Triadic Colors** consist of three colors that are evenly spaced around the color wheel. This scheme is vibrant, even if the colors are unsaturated, and offers a balanced level of contrast suitable for drawing attention to key elements without overwhelming the viewer.

Understanding the color wheel and these color relationships is crucial for anyone involved in data visualization.



Color Properties

Common Color Properties: Hue, Saturation, and Brightness

Understanding the fundamental properties of color is essential for effective visualization design. These properties — hue, saturation, and brightness — determine how we perceive and interpret color in visualizations.

- **Hue:** Hue refers to the color itself, the aspect of color that is determined by the specific wavelength of light. It is what we most commonly think of as "color" — red, blue, green, and so on. In data visualization, different hues are used to distinguish categories, highlight specific data points, or encode information.
- **Saturation:** Saturation describes the intensity or purity of a color. A highly saturated hue appears vivid and bright, while a less saturated hue appears more muted or gray. Adjusting saturation can enhance clarity or reduce visual fatigue, depending on the amount of data and the complexity of the visualization. High saturation can draw attention to areas of importance, but overuse can make a visualization overwhelming.
- **Brightness:** Brightness (or value) refers to how light or dark a color appears. It can influence the readability and emphasis within a visualization. Brighter colors tend to advance in visual perception, while darker colors recede. This property can be manipulated to create depth or focus in a visualization.

Color Properties (Example)

Hue



Saturation



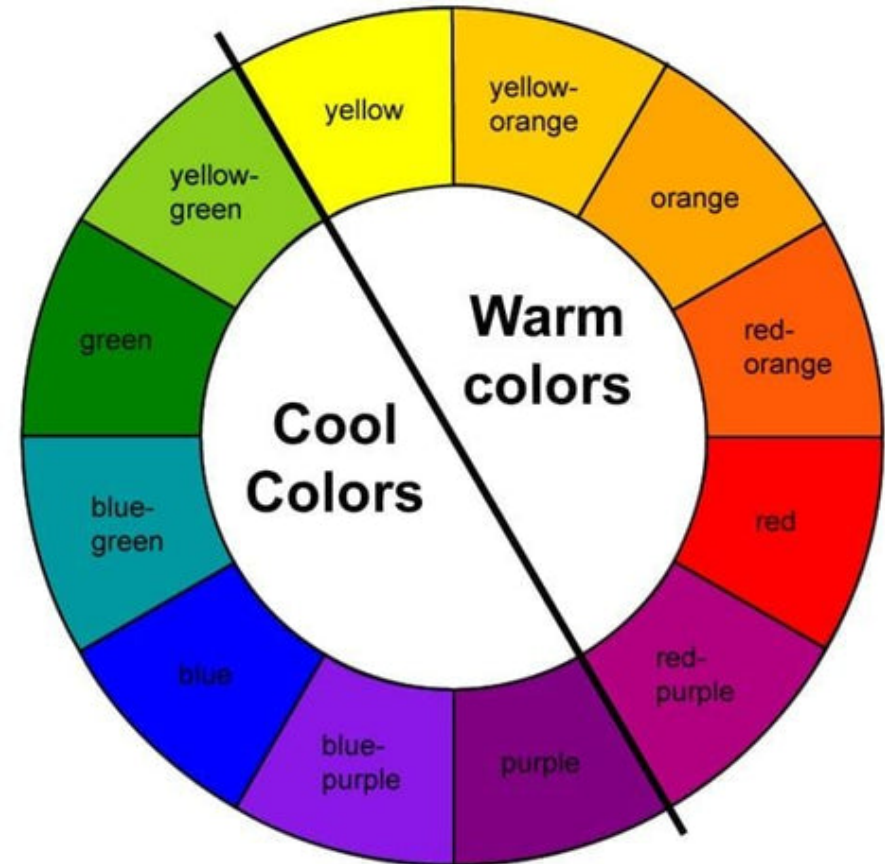
Brightness



Color Properties (Continued)

Warm vs. Cool Colors: The color spectrum is also divided into warm and cool colors, each evoking different psychological responses and interpretations.

- **Warm Colors:** Warm colors include reds, oranges, and yellows. These hues are associated with energy, passion, and action. They can be used in visualizations to draw attention, signify caution, or stimulate viewer engagement. Warm colors are often used to represent positive values or trends, such as growth or increase.
- **Cool Colors:** Cool colors encompass blues, greens, and purples. They are perceived as calming and soothing and are often associated with trust, stability, and peace. In visualizations, cool colors are useful for background elements or to represent negative values or trends, such as decline or loss.



Color Properties (Continued)

- **Psychological Effects and Interpretation:** The choice between warm and cool colors can significantly affect a viewer's interpretation of the data. Warm colors might be used to highlight areas of concern or interest, while cool colors can provide a calming counterpoint or denote areas of less concern. Understanding the psychological impact of color can help designers create more intuitive and impactful visualizations.

Incorporating knowledge of hue, saturation, brightness, and the psychological effects of warm vs. cool colors allows data visualization practitioners to craft visualizations that are not only visually appealing but also enhance the viewer's ability to understand and interpret the data presented.

Color Schemes in Visualization

In data visualization, the choice of color scheme plays a critical role in effectively communicating information. The type of data you're visualizing — whether nominal, ordinal, interval, or ratio — should guide your selection of color schemes. Here, we'll explore three primary types of color schemes: sequential, diverging, and qualitative, and discuss best practices for applying these schemes to different data types.

Sequential Color Schemes

- **Description:** Sequential color schemes are characterized by a gradual transition from light to dark hues of the same color or from one color to another closely related color. They are ideal for displaying data that has ordering.
- **Best for:** Ordinal, interval, and ratio data where the magnitude of the data values matters, such as temperature ranges, age groups, or income levels.
- **Best Practices:**
 - Use lighter shades for lower values and darker shades for higher values to naturally indicate an increase in magnitude.
 - Ensure there is enough contrast across the range to distinguish between different data points clearly.

Color Schemes in Visualization (Continued)

Diverging Color Schemes

- **Description:** Diverging color schemes start with two contrasting colors at the extremes and meet in a neutral middle with lighter shades. They are best used when data values deviate around a median, zero, or any other significant midpoint.
- **Best for:** Interval and ratio data where emphasizing the deviation from a critical midpoint is essential, such as changes in temperature from a baseline, stock market gains and losses, or survey responses ranging from positive to negative.
- **Best Practices:**
 - Choose colors that contrast well to highlight differences effectively.
 - Use the midpoint color to represent the median or zero value, ensuring it's neutral and distinct from the extremes.
 - Make sure the transition between colors is smooth and maintains the readability of the data.

Color Schemes in Visualization (Continued)

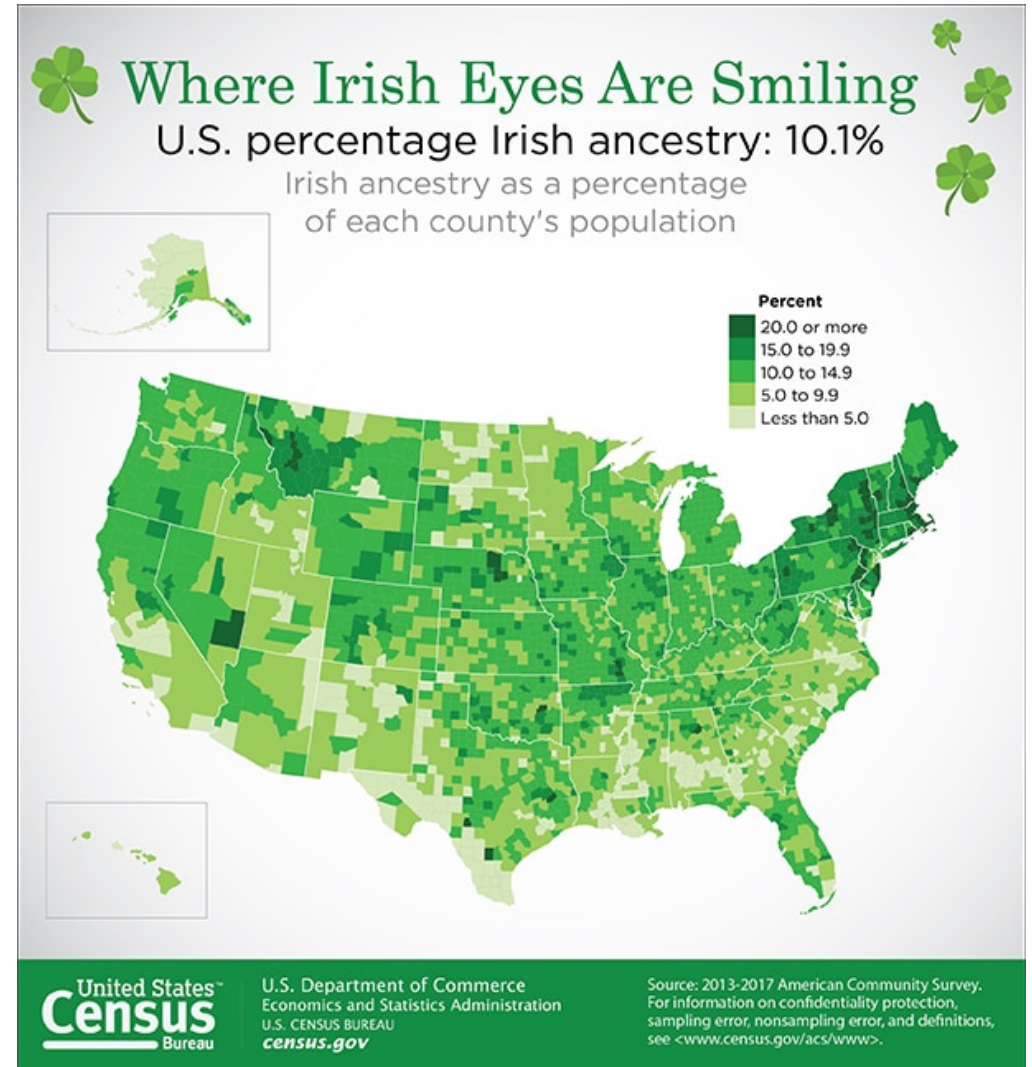
Qualitative Color Schemes

- **Description:** Qualitative (or categorical) color schemes consist of distinct colors that are used to represent different categories without implying order or magnitude. The colors are chosen to be visually distinct from each other to differentiate between categories clearly.
- **Best for:** Nominal data where the focus is on distinguishing between categories, such as types of fruits, departments in a company, or different species.
- **Best Practices:**
 - Select colors that are easily distinguishable from one another to avoid confusion.
 - Avoid using too many categories in a single visualization to prevent cognitive overload. If many categories are necessary, consider using other methods in conjunction with color to differentiate between them (e.g., patterns or labels).
 - Be mindful of color vision deficiencies and choose colors that are distinguishable to all viewers, employing tools like Color Oracle or Coblis for testing.

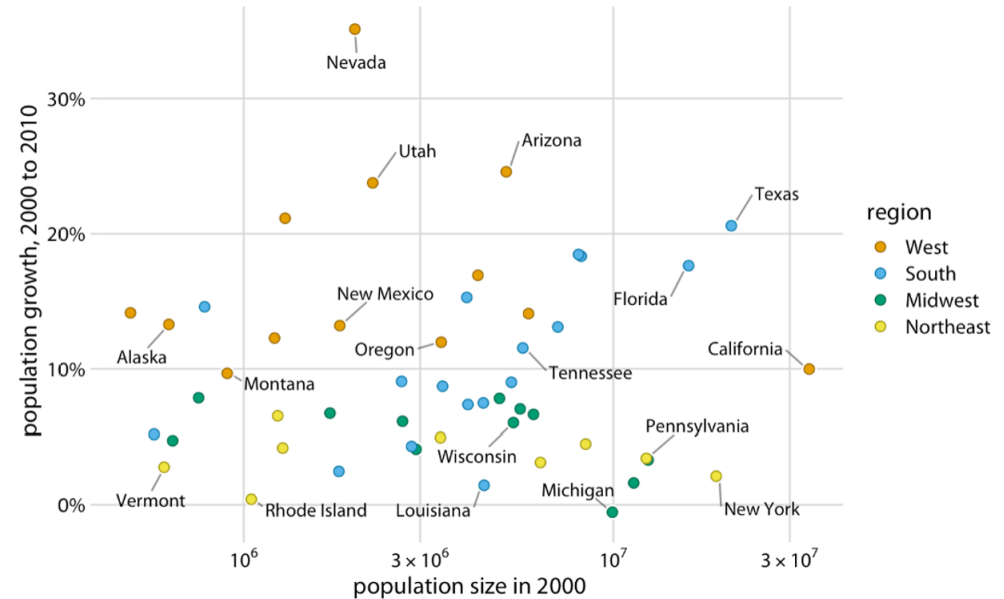
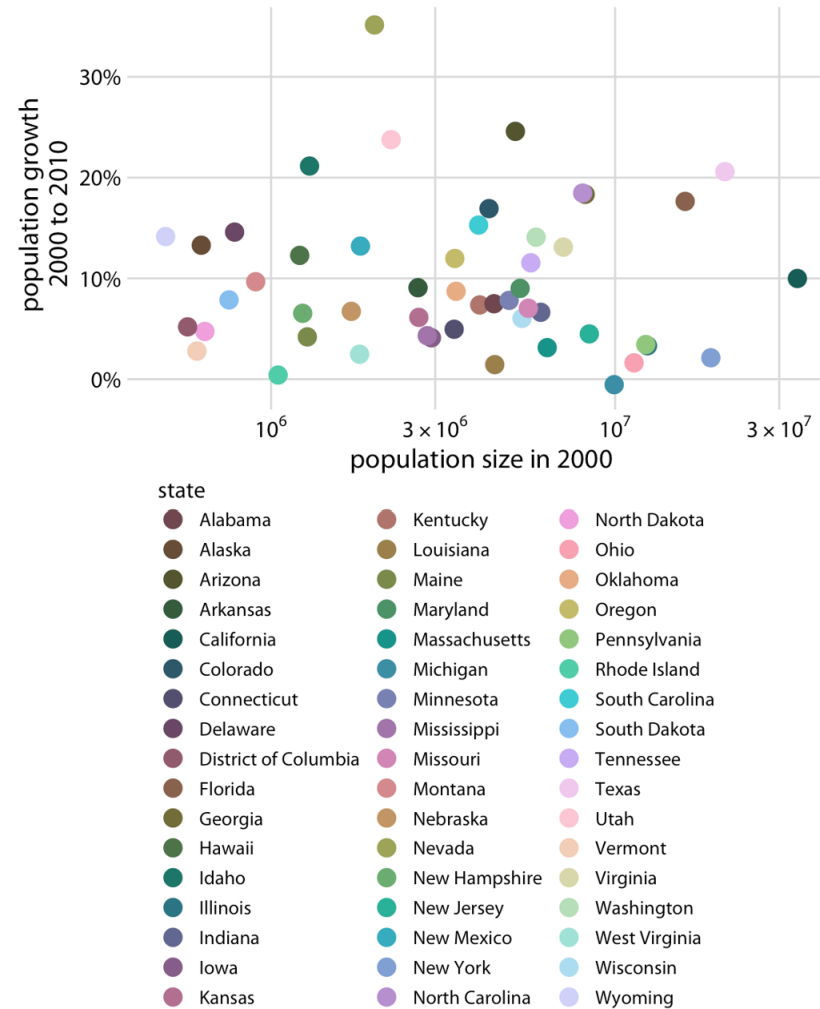
By thoughtfully applying these color schemes based on the data type and the message we want to convey, we can enhance the interpretability and aesthetic appeal of our visualizations. **Remember, the goal of color usage in data visualization is not just to make the visual attractive but to support the viewer in understanding the data more effectively.**

Case Study - 1

- **Analysis:** For each example, analyze the color scheme based on its appropriateness for the data type, effectiveness in conveying the intended message, and accessibility considerations.
- **Critique:** Identify what works well and what does not in each color scheme application. Discuss how the ineffective elements could be improved while maintaining or enhancing the visualization's overall impact.
- **Recommendations:** Propose alternative color schemes for the examples identified as having bad practices, explaining why your choices would be more effective.



Case Study - 2



Designing for Color-blindness and Accessibility

We will discuss the importance of designing visualizations that are accessible to individuals with color vision deficiencies and provide strategies to achieve this.

Understanding Color-blindness

Color vision deficiencies (CVD) affect a significant portion of the population, influencing how individuals perceive colors. Understanding these conditions is crucial for designing accessible visualizations.

Here are the most common types of Color Vision Deficiencies: the most common types: Deuteranomaly (green-weak: reduced sensitivity to green light), Protanomaly (red-weak: involves a diminished sensitivity to red light), Tritanomaly (blue-weak: a decreased sensitivity to blue light), and their more severe forms, Deuteranopia (Green-Blind), Protanopia (Red-Blind), and Tritanopia (Blue-Blind).

Understanding these color vision deficiencies is essential for creating visualizations that are accessible and interpretable by everyone, including those with CVD. By considering the impacts of these conditions on color perception, designers can select color schemes and visualization techniques that communicate effectively to a broader audience.

Accessible Color Palettes

Creating data visualizations that are accessible to a broad audience, including those with color vision deficiencies (CVD), requires careful selection and testing of color palettes. Here's how to approach this task to ensure your visualizations are clear and understandable to everyone.

How to Select and Test Color Palettes for Accessibility

- Understand Common CVD Types
- Use CVD-Simulating Tools
- Limit the Number of Colors
- Prefer High Contrast: Ensure high contrast not just in color hues but also in lightness and saturation levels. High contrast helps individuals with and without CVD distinguish between different elements in a visualization.
- Avoid Problematic Colors: Some color combinations, such as red-green, are particularly problematic for people with CVD. Avoid these combinations or use them in conjunction with other differentiation methods, like patterns or shapes.

Accessible Color Palettes (Continued)

We can Utilize tools to Choose color-blind friendly palettes. We list some useful tools below.

- **ColorBrewer**: An online tool designed for scientists and statisticians that offers a selection of color palettes optimized for data visualization, including options specifically tailored for color-blind viewers. ColorBrewer palettes are categorized into sequential, diverging, and qualitative types, with indications of CVD-friendly options.
- **Adobe Color**: This tool provides a color wheel that can simulate various forms of color blindness, helping designers see how their palettes appear to those with different CVD types. It's useful for creating custom palettes with accessibility in mind.
- **WebAIM (Web Accessibility In Mind)**: Provides a contrast checker that evaluates text color against background color, ensuring that your visualizations are legible to viewers with vision impairments, including CVD.

Design Strategies for Accessibility:

Beyond color: Using patterns, textures, and shapes to differentiate data points.

Ensuring sufficient contrast and readability in visualizations. Annotation and labeling techniques to complement color coding.

References

The lectures of this course are based on the ideas from the following references.

- Exploratory Data Analysis by John W. Tukey
- A Course in Exploratory Data Analysis by Jim Albert
- The Visual Display of Quantitative Information by Edward R. Tufte
- Data Science for Business: what you need to know about data mining and data-analytic thinking by Foster Provost and Tom Fawcett
- Storytelling with Data: A Data Visualization Guide for Business Professionals by Cole Nussbaumer Knaflic