Data Visualization (CSCI 627/490)

Colormaps

Dr. David Koop
Human Color Perception

- Humans **do not** detect individual wavelengths of light
- Use **rods** and **cones** to detect light
  - rods capture intensity
  - cones capture color
Color != Wavelength

But rather, a combination of wavelengths and energy.
Human Color Perception

![Graph showing the process of human color perception]

[via M. Meyer]

D. Koop, CSCI 627/490, Fall 2022
Simulating Deuteranopia (Colormaps)

Simulation of green deficient colour blindness (deuteranopia) at 0%

- RdYIGn
- Dark2
- jet
- Spectral
- Purples
- RdYIBu
- YIGnBu
- linearL
- parula

@neilrkaye

[@neilrkaye, reddit]
Simulating Deuteranopia (Colormaps)

Simulation of green deficient colour blindness (deuteranopia) at 0%

- RdYIGn
- Dark2
- jet
- Spectral
- Purples
- RdYIBu
- YIGnBu
- linearL
- parula

@neilrkaye
[@neilrkaye, reddit]
Assignment 3

• Same visualization
• Different tools
  - Tableau (Public or Desktop)
  - Observable Plot
  - D3
Midterm

• Thursday, October 13
• Covers material through this week
• Format:
  - In Person, Pen(-cil) & Paper
  - Multiple Choice
  - Free Response (often multi-part)
  - CS 627 students will have extra questions related to the research papers discussed
Project

• Two Possibilities:
  - Create an interactive visualization
  - Work on a research project

• Dataset Choices
  - NFL Data
  - Colorado River Data
  - Prescription Drug Cost Data
  - Others?
Primary Colors?

• Red, Green, and Blue
• Red, Yellow, and Blue
• Orange, Green, and Violet
• Cyan, Magenta, and Yellow
Primary Colors?

- Red, Green, and Blue
- Red, Yellow, and Blue
- Orange, Green, and Violet
- Cyan, Magenta, and Yellow
- All of the above!
Color Addition and Subtraction

- Color Addition: RGB (Red, Green, Blue)
- Color Subtraction: CMY (Cyan, Magenta, Yellow)
Color Spaces and Gamuts

- **Color space**: the organization of all colors in space
  - Often human-specific, what we can see (e.g. CIELAB)

- **Color gamut**: a subset of colors
  - Defined by corners of color space
  - What can be produced on a monitor (e.g. using RGB)
  - What can be produced on a printer (e.g. using CMYK)
  - The gamut of your monitor != the gamut of someone else's or a printer

[Anatomy of a CIE Chromaticity Diagram]
Color Models

- **A color model** is a representation of color using some basis
- RGB uses three numbers (red, blue, green) to represent color
- Color space ~ color model, but there can be many color models used in the same color space (e.g. OGV)
- Hue-Saturation-Lightness (HSL) is more intuitive and useful
  - Hue captures pure colors
  - Saturation captures the amount of white mixed with the color
  - Lightness captures the amount of black mixed with a color
  - HSL color pickers are often circular
- Hue-Saturation-Value (HSV) is similar (swap black with gray for the final value), linearly related
Luminance

- HSL does not truly reflect the way we perceive color
- Even though colors have the same lightness, we perceive their luminance differently
- Our perception ($L^*$) is **nonlinear**

![Corners of the RGB color cube](image1)

![L from HSL](image2)

![All the same](image3)

![Luminance](image4)

![$L^*$](image5)

[Munzner (ill. Maguire), 2014 (based on Stone, 2006)]
Perceptually Uniform Color Spaces

- L*a*b* allows perceptually accurate comparison and calculations of colors
Luminance Perception (Spatial Adaption)

[Edward H. Adelson, 1995]
Luminance Perception (Spatial Adaption)

[Image of a checkerboard with a cylinder and bars, labeled A and B.]

Edward H. Adelson

[E. H. Adelson, 1995]
Simultaneous Contrast
Simultaneous Contrast
Simultaneous Contrast
Simultaneous Contrast
What colors?
What colors?

Red, yellow, blue

Purple, orange
do not exist!

[A. Kitaoka]
What does this mean for visualization?
What does this mean for visualization?

- We need to be aware of colorblindness when encoding via color.
- Our brains may misinterpret color (surrounding colors matter!) even if we aren't colorblind.
- Be careful! Don't assume that adding color always works the way you intended.
- Use known colormaps when possible.
Violations of CIELAB Assumptions

• CIELAB:
  - Approximately perceptually linear
  - 1 unit of Euclidean distance = 1 Just Noticeable Difference (JND)
  - JND: people detect change at least 50% of the time

• Assumptions CIELAB makes:
  - Simple world
  - Isolation
  - Geometric
Simple World Assumption

[Image of a person looking at a computer monitor]

[D. Szafir, 2017]
Problems with Simple World Assumption

[Diagram showing viewing distance, environmental surround, ambient illumination, direct illumination, viewing population, and gamma, whitepoint, resolution, peak color outputs]

Simple World Assumption
- Isolation Assumption
- Geometric Assumption

Visualizations violate three CIELAB assumptions.

Crowdsourced Sampling
- Szafir, Stone, & Gleicher, 2014
- Reinecke, Flatla, & Brooks, 2016

Problems with Simple World Assumption

[D. Szafir, 2017]
Isolation Assumption
Problems with Isolation Assumption

[D. Szafir, 2017]
Geometric Assumption
Size Problem with Geometric Assumption

[Size Problem with Geometric Assumption Diagram]

[D. Szafir, 2017]
Shape Problem with Geometric Assumption

[D. Szafir, 2017]
Types of Geometry

Diagonally Symmetric Marks

Elongated Marks

Asymmetric Marks

Area Marks

[D. Szafir, 2017]
Run the tests!
Color Study

6 (diameters, within) × 6 (color differences, within) × 3 (color axis, between)

81 participants on Mechanical Turk (5,668 trials)

25 pixels 1.0°

37 pixels 1.5°

50 pixels 2.0°

[D. Szafir, 2017]
Point Size: consistent with previous results

50% JND for Scatterplot Points

$\Delta b^*$
$\Delta a^*$
$\Delta L^*$

Stone et al.

$\Delta E = 1.0$

D. Koop, CSCI 627/490, Fall 2022
Bar Thickness and Length: longer bars help

50% JND for Bars

Bar Length encoded as point size

JND in CIELAB

0 5 10 15 20 25
0 0.5 1.0 1.5 2.0

Bar Thickness

Visual Angle

Δb* for points

Δa* for points

ΔL* for points

Bar Thickness encoded as point size

Accounting for Gains:

Elongation

Area = longest edge / shortest edge

Longest edge

Shortest edge

D. Szafir, 2017
Line Thickness: better than points

50% JND for Lines

- $\Delta b^*$ for points
- $\Delta a^*$ for points
- $\Delta L^*$ for points

$R^2 = .97$

$R^2 = .93$

$R^2 = .90$

[Line Thickness: better than points]
Color perception in real-world visualizations is complicated
Akiyoshi Kitaoka's Illusion pages
Colormaps
Colormap

- A colormap specifies a mapping between colors and data values
- Colormap should follow the expressiveness principle
- Types of colormaps:
  
  **Binary**
  
  - y
  - n

  **Diverging**
  
  - -1
  - 0
  - +1

  **Categorical**
  
  - T
  - F
  - A

  **Sequential**
  
  - 3
  - 2
  - 1

[Munzner (ill. Maguire), 2014]
Categorical vs. Ordered

- Hue has no implicit ordering: use for categorical data
- Saturation and luminance do: use for ordered data
Categorical Colormap Guidelines

- Don't use too many colors (~12)
- Remember your background has a color, too
- Nameable colors help
- Be aware of luminance (e.g. difference between blue and yellow)
- Think about other marks you might wish to use in the visualization
Categorical Colormaps

[link to colorbrewer2.org]
Categorical Colormaps

[link: colorbrewer2.org]
Number of distinguishable colors?

[Sinha & Meller, 2007]
Number of distinguishable colors?

[Sinha & Meller, 2007]
Discriminability

- Often, fewer colors are better
- Don't let viewers combine colors because they can't tell the difference
- Make the combinations yourself
- Also, can use the "Other" category to reduce the number of colors
Ordered Colormaps

- Used for ordinal or quantitative attributes
- \([0, N]\): Sequential
- \([-N, 0, N]\): Diverging (has some meaningful midpoint)
- Can use hue, saturation, and luminance
- Remember hue is not a magnitude channel so be careful
- Can be **continuous** (smooth) or **segmented** (sharp boundaries)
  - Segmented matches with ordinal attributes
  - Can be used with quantitative data, too.
Continuous Colormap

US EPA Regional Oxidant Model -- Midwest
Ozone (ppbv): June 26, 1987, 18:00

[Bergman et al., 1995]
Segmented Colormap

US EPA Regional Oxidant Model -- Midwest
Ozone (ppbv): June 26, 1987, 18:00

[Bergman et al., 1995]
Is continuous better than segmented?
Evaluating the Impact of Binning 2D Scalar Fields

Lace Padilla, P. Samuel Quinan, Miriah Meyer, and Sarah H. Creem-Regehr

Fig. 1: Experimental stimuli for five binning conditions: A. Continuous, B. 10m binning, C. 20m binning, D. 30m binning, E. 40m binning

Abstract
— The expressiveness principle for visualization design asserts that a visualization should encode all of the available data, and only the available data, implying that continuous data types should be visualized with a continuous encoding channel. And yet, in many domains binning continuous data is not only pervasive, but it is accepted as standard practice. Prior work provides no clear guidance for when encoding continuous data continuously is preferable to employing binning techniques or how this choice affects data interpretation and decision making. In this paper, we present a study aimed at better understanding the conditions in which the expressiveness principle can or should be violated for visualizing continuous data. We provided participants with visualizations employing either continuous or binned greyscale encodings of geospatial elevation data and compared participants' ability to complete a wide variety of tasks. For various tasks, the results indicate significant differences in decision making, confidence in responses, and task completion time between continuous and binned encodings of the data. In general, participants with continuous encodings were faster to complete many of the tasks, but never outperformed those with binned encodings, while performance accuracy with binned encodings was superior to continuous encodings in some tasks. These findings suggest that strict adherence to the expressiveness principle is not always advisable. We discuss both the implications and limitations of our results and outline various avenues for potential work needed to further improve guidelines for using continuous versus binned encodings for continuous data types.

Index Terms — Geographic/Geospatial Visualization, Qualitative Evaluation, Color Perception, Perceptual Cognition

1 INTRODUCTION
A foundational design principle in visualization is the expressiveness principle, which states that a visual encoding should express all of the relationships in the data, and only the relationships in the data [24, 35]. For a continuous data type, this implies that a continuous encoding channel is a good choice. In practice, however, domains such as cartography [43] and meteorology [36] have strong conventions that visualize continuous data with a discrete encoding. These domains rely on visual channels, such as color and saturation to encode a continuous function defined over two-dimensional space, known as a 2D scalar field. They commonly do so by employing discrete colormaps or contour lines, also called isarithmic maps [43].

Existing literature provides little guidance about encoding continuous, 2D scalar fields with binned colormaps, or how this design decision affects data interpretation and decision making. Research into properties of colormaps for encoding continuous data types has largely focused on continuous colormaps [2, 28, 38, 48]. This line of research provides guidance on how to capture properties of the data, such as divergence around a center point [48] or emphasis on one end of the data range [2]. These papers go so far as proposing corresponding binned colormaps, but do not make claims, or even discuss, their efficacy for continuous data. Work on transfer function design has also proposed methods for binning colors, but with a focus on volumetric scalar fields, with the underlying goal of classifying materials or features [12], as opposed to directly understanding the continuous nature of the data.

Continuous [Padilla et al., 2017]
Evaluating the Impact of Binning 2D Scalar Fields

Lace Padilla, P. Samuel Quinan, Miriah Meyer, and Sarah H. Creem-Regehr

Fig. 1: Experimental stimuli for five binning conditions: A. Continuous, B. 10m binning, C. 20m binning, D. 30m binning, E. 40m binning

Abstract

The expressiveness principle for visualization design asserts that a visualization should encode all of the available data, and only the available data, implying that continuous data types should be visualized with a continuous encoding channel. And yet, in many domains binning continuous data is not only pervasive, but it is accepted as standard practice. Prior work provides no clear guidance for when encoding continuous data continuously is preferable to employing binning techniques or how this choice affects data interpretation and decision making. In this paper, we present a study aimed at better understanding the conditions in which the expressiveness principle can or should be violated for visualizing continuous data. We provided participants with visualizations employing either continuous or binned greyscale encodings of geospatial elevation data and compared participants' ability to complete a wide variety of tasks. For various tasks, the results indicate significant differences in decision making, confidence in responses, and task completion time between continuous and binned encodings of the data. In general, participants with continuous encodings were faster to complete many of the tasks, but never outperformed those with binned encodings, while performance accuracy with binned encodings was superior to continuous encodings in some tasks. These findings suggest that strict adherence to the expressiveness principle is not always advisable. We discuss both the implications and limitations of our results and outline various avenues for potential work needed to further improve guidelines for using continuous versus binned encodings for continuous data types.

Index Terms

Geographic/Geospatial Visualization, Qualitative Evaluation, Color Perception, Perceptual Cognition

1 INTRODUCTION

A foundational design principle in visualization is the expressiveness principle, which states that a visual encoding should express all of the relationships in the data, and only the relationships in the data. For a continuous data type, this implies that a continuous encoding channel is a good choice. In practice, however, domains such as cartography and meteorology have strong conventions that visualize continuous data with a discrete encoding. These domains rely on visual channels, such as color and saturation to encode a continuous function defined over two-dimensional space, known as a 2D scalar field. They commonly do so by employing discrete colormaps or contour lines, also called isarithmic maps.

Existing literature provides little guidance about encoding continuous, 2D scalar fields with binned colormaps, or how this design decision affects data interpretation and decision making. Research into properties of colormaps for encoding continuous data types has largely focused on continuous colormaps. This line of research provides guidance on how to capture properties of the data, such as divergence around a center point or emphasis on one end of the data range. These papers go so far as proposing corresponding binned colormaps, but do not make claims, or even discuss, their efficacy for continuous data. Work on transfer function design has also proposed methods for binning colors, but with a focus on volumetric scalar fields, with the underlying goal of classifying materials or features, as opposed to directly understanding the continuous nature of the data.
Evaluating the Impact of Binning 2D Scalar Fields

Lace Padilla, P. Samuel Quinan, Miriah Meyer, and Sarah H. Creem-Regehr

Fig. 1: Experimental stimuli for five binning conditions: A. Continuous, B. 10m binning, C. 20m binning, D. 30m binning, E. 40m binning

Abstract

The expressiveness principle for visualization design asserts that a visualization should encode all of the available data, and only the available data, implying that continuous data types should be visualized with a continuous encoding channel. And yet, in many domains binning continuous data is not only pervasive, but it is accepted as standard practice. Prior work provides no clear guidance for when encoding continuous data continuously is preferable to employing binning techniques or how this choice affects data interpretation and decision making. In this paper, we present a study aimed at better understanding the conditions in which the expressiveness principle can or should be violated for visualizing continuous data. We provided participants with visualizations employing either continuous or binned greyscale encodings of geospatial elevation data and compared participants' ability to complete a wide variety of tasks. For various tasks, the results indicate significant differences in decision making, confidence in responses, and task completion time between continuous and binned encodings of the data. In general, participants with continuous encodings were faster to complete many of the tasks, but never outperformed those with binned encodings, while performance accuracy with binned encodings was superior to continuous encodings in some tasks. These findings suggest that strict adherence to the expressiveness principle is not always advisable. We discuss both the implications and limitations of our results and outline various avenues for potential work needed to further improve guidelines for using continuous versus binned encodings for continuous data types.

Index Terms

Geographic/Geospatial Visualization, Qualitative Evaluation, Color Perception, Perceptual Cognition

1INTRODUCTION

A foundational design principle in visualization is the expressiveness principle, which states that a visual encoding should express all of the relationships in the data, and only the relationships in the data [24, 35]. For a continuous data type, this implies that a continuous encoding channel is a good choice. In practice, however, domains such as cartography [43] and meteorology [36] have strong conventions that visualize continuous data with a discrete encoding. These domains rely on visual channels, such as color and saturation to encode a continuous function defined over two-dimensional space, known as a 2D scalar field. They commonly do so by employing discrete colormaps or contour lines, also called isarithmic maps [43].

Existing literature provides little guidance about encoding continuous, 2D scalar fields with binned colormaps, or how this design decision affects data interpretation and decision making. Research into properties of colormaps for encoding continuous data types has largely focused on continuous colormaps [2, 28, 38, 48]. This line of research provides guidance on how to capture properties of the data, such as divergence around a center point [48] or emphasis on one end of the data range [2]. These papers go so far as proposing corresponding binned colormaps, but do not make claims, or even discuss, their efficacy for continuous data. Work on transfer function design has also proposed methods for binning colors, but with a focus on volumetric scalar fields, with the underlying goal of classifying materials or features [12], as opposed to directly understanding the continuous nature of the data.

Fewer Segments

[Padilla et al., 2017]
Types of Tasks

- Locate/Explore & Identify: Highest Point (Global, In Region), 275m
- Locate/Explore & Compare: Height Compare/Rank
- Explore & Identify: Steepest
- Lookup & Identify: Lookup
- Explore & Compare: Steepness Compare/Rank
- Browse & Summarize: Average Height
- Browse & Compare: Compare Average Height
- Combination: Steepest at 355m
Results

- "[C]ontrary to the expressiveness principle, no cases were found in which a continuous encoding of 2D scalar field data was advantageous for task accuracy, and for some tasks, specific binned encodings facilitated accuracy."
- "[S]upport for the counterintuitive finding that decisions with binned encoding were slower than those made with continuous encoding"
- Word of caution: single image!
Don't Use Rainbow Colormaps

Which has a discontinuity?
Other Colormaps Work Better

Which has a discontinuity?

[M. Bussonnier]
Not only does the rainbow color map confuse viewers through its lack of perceptual ordering and obscure data through its inability to present small details, but it actively misleads the viewer by introducing artifacts to the visualization. The rainbow color map appears as if it’s separated into bands of almost constant hue, with sharp transitions between hues. Viewers perceive these sharp transitions as sharp transitions in the data, even when this is not the case (see Figure 3). When combined with the lack of perceptual ordering, viewers face a daunting task when trying to correctly interpret the data via the rainbow color map. The goal of visualization is to present data so that viewers can quickly and accurately learn about the underlying data. The rainbow color map does a great deal to hinder this learning process by introducing confusing artifacts in some locations and reducing detail in others.

Prevalence of the rainbow color map

Although researchers have well documented these deficiencies, the visualization community still widely uses the rainbow color map. We present the findings of two surveys illustrating this prevalence. The first survey looks at papers in the IEEE Visualization Conference proceedings; the second considers visualization toolkits.

IEEE Visualization proceedings

We searched the IEEE Visualization conference proceedings from 2001 through 2005 for papers that displayed data using a pseudocolor map. We included visualizations in which the rainbow color map was applied to surfaces, such as isosurfaces and streamlines. We excluded volume renderings as the literature does not address the relative merits of the rainbow color map when used for a color transfer function (although it seems clear that the same objections would apply). We did not count visualizations that used a banded version of the rainbow color map because explicit banding can be a useful visualization technique. We only included scalar data visualizations, excluding techniques such as mapping vector components to RGB—which is common with diffusion tensor MRI images. Such visualizations can appear at first glance to use the rainbow color map, but they are in fact using a different technique (see Rheingans for a discussion of the hazards of encoding multiple values into a pseudocolor map).

Results

Table 1 (next page) presents statistics from the 2001 through 2005 IEEE Visualization Conference proceedings. The table gives percentages of papers implementing pseudocoloring to display data using the rainbow color map. We’ve included all papers that include at least one use of the rainbow color map. The results are alarming: Each year between 40 and 59 percent of all papers using pseudocoloring used a rainbow color map.

[Borland & Taylor, 2007]
Rainbow Colormap

[Bergman et al., 1995]
Artifacts from Rainbow Colormaps

[Borland & Taylor, 2007]
Artifacts from Rainbow Colormaps

[Borland & Taylor, 2007]
Two-Hue Colormap

[Bergman et al., 1995]
"Get It Right in Black and White" - M. Stone

jet colormap

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

jet colormap

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

parula colormap

[S. Eddins (Matlab Blog), 2014]
"Get It Right in Black and White" - M. Stone

parula colormap

[S. Eddins (Matlab Blog), 2014]
Isoluminant Rainbow Colormap

\[
\begin{align*}
\text{Original} & \\
\text{Isoluminant} & \\
\end{align*}
\]
Turbo Colormap (August 2019)

Jet

Turbo
Turbo: More Detail in Disparity Maps?
Turbo: Lightness Profiles

Jet

Viridis

Turbo

[A. Mikhailov]
Turbo Discussion

- Turbo is an improvement over jet
- Some fields (e.g. meteorology) have long used rainbow-like colormaps
- Argument is that segments are more easily located
- Turbo post claims that hue is prioritized in attention, but this seems to misinterpret the study…
- Brightness and saturation are more important than hue in attracting attention [Camgöz et al., 2004 h/t J. Stevens]
More Guidelines

• Nice set of articles by Lisa Charlotte Rost:
  - https://blog.datawrapper.de/colorguide/
  - https://blog.datawrapper.de/beautifulcolors/

• Her guidelines on choosing colors:
  1. Copy from others
  2. Use Tools
  3. ...