Data Visualization (DSC 530/CIS 602-02)

Color

Dr. David Koop
Flattening the Sphere?

Central Meridian  
(selected by mapmaker)

Great distortion at high latitudes
Examples of two rhumb lines (direction true between any two points)
Equator touches cylinder if cylinder is tangent
Reasonably true shapes and distances within 15 degrees of Equator

[USGS Map Projections]
Choropleth (Two Hues)

[M. Ericson, New York Times]
Size Encoding

[M. Ericson, New York Times]
Cartograms

[Election Results by Population, M. Newman, 2012]
Project Proposal

• Due Tuesday, March 21
• Identify dataset
  - Potential data sources: https://github.com/caesar0301/awesome-public-datasets
• Understand domain
• Decide on tasks
• Start brainstorming on visualization and interaction design
Assignment 3

• Networks and Maps
• Data: Soccer Players
• Tasks:
  - Country of Origin (Spatial Patterns?)
  - Teammates (Link Patterns?)
Color
Color
Color and Light

• Color is a **perceptive** property: color depends on the eyes and brain

• Visible light is a small portion of the **electromagnetic spectrum** which is composed of waves that at various frequencies (wavelengths), all traveling at the speed of light
Electromagnetic Spectrum

Penetrates Earth's Atmosphere?

Radiation Type
- Radio
- Microwave
- Infrared
- Visible
- Ultraviolet
- X-ray
- Gamma ray

Wavelength (m)
- Radio: $10^3$
- Microwave: $10^{-2}$
- Infrared: $10^{-5}$
- Visible: $0.5 \times 10^{-6}$
- Ultraviolet: $10^{-8}$
- X-ray: $10^{-10}$
- Gamma ray: $10^{-12}$

Approximate Scale of Wavelength
- Buildings
- Humans
- Butterflies
- Needle Point
- Protozoans
- Molecules
- Atoms
- Atomic Nuclei

Frequency (Hz)
- Radio: $10^4$
- Microwave: $10^8$
- Infrared: $10^{12}$
- Visible: $10^{15}$
- Ultraviolet: $10^{16}$
- X-ray: $10^{18}$
- Gamma ray: $10^{20}$

Temperature of objects at which this radiation is the most intense wavelength emitted
- 1 K: $-272 \, ^{\circ}C$
- 100 K: $-173 \, ^{\circ}C$
- 10,000 K: 9,727 \, ^{\circ}C
- 10,000,000 K: $\sim 10,000,000 \, ^{\circ}C$

[WikiMedia, NASA]
Light Reflection & Absorption

[Image of fruits with coloring indicating wavelengths]
Color != Wavelength

[Diagram showing relative energy density vs. wavelength (nm) with peaks at yellow and brown colors.]
Human Color Perception

• Humans are **trichromatic**—we have three different types of cones
  - S (430nm): blue
  - M (540nm): green
  - L (570nm): "red"
• Note that the response curves **overlap**
• Spectra of visible light are "covered" by these responses
• Three numbers -> color

[Vanessaezekowitz at en.wikipedia]
Human Color Perception

- Humans do not detect individual wavelengths of light
- Use rods and cones to detect light
  - rods capture intensity
  - cones capture color

Human Color Perception

Metamerism: same three responses == same color

[via M. Meyer]
Metamerism

[via M. Meyer]
Color

• Cones respond to different areas of the visible light spectrum
• Cover all wavelengths but certain wavelengths generate greater responses
• Color is determined by calculations based on the responses from the different cones
• Opponent Process Theory: three "opponent" channels
  - Light/Dark
  - Blue/Yellow
  - Red/Green
• Opposite colors are not perceived together
Opponent Process Theory

[Machado et. al, 2009]
Color Blindness

[Ishihara (Plate 9) via Wikipedia]
Color Blindness

• Sex-linked: 8% of males and 0.4% of females of N. European ancestry
• Abnormal distribution of cones (e.g. missing the S, M, or L types)
• Either dichromatic (only two types of cones) or anomalous trichromatic (one type of cones has a defect)
  - Protanopia (L missing), Protanomaly (L defect)
  - Deuteranopia (M missing), Deuteranomaly (M defect) [Most Common]
  - Tritanopia (S missing), Tritanomaly (S defect) [Rare]
• Dichromacy is rarer than anomalous trichromacy
• Opponent process model explains why colors cannot be differentiated
Color Blindness

Normal

700 650 600 550 500 450 400

Prota-nopia

700 650 600 550 500 450 400

Deuteranopia

700 650 600 550 500 450 400

Tritanopia

700 650 600 550 500 450 400

[via M. Meyer]
Simulating Color Blindness

[Images of flowers under different conditions: Empty, Photop. Subst., Scale Ratio, 0.96*Ratio, Brettel]

Protopanopia

Deuteranopia

[Machado et. al, 2009]
Simulating Color Blindness

[Machado et. al, 2009]
Simulating Color Blindness

[Machado et. al, 2009]
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 Spaces and Gamuts

[http://dot-color.com/2012/08/14/color-space-confusion/]
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 on in the 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 \( \neq \) the gamut of someone else's \( \neq \) the gamut of a printer
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

L from HSL

All the same

Luminance

$L^*$

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

- $L^*a^*b^*$ allows perceptually accurate comparison and calculations of colors

[J. Rus, CC-BY-SA (changed to horizontal layout)]
Luminance Perception (Spatial Adaption)

Edward H. Adelson

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

[Edward H. Adelson, 1995]
Simultaneous Contrast
Simultaneous Contrast
Simultaneous Contrast
Simultaneous Contrast
Colormap

• A colormap specifies a mapping between colors and data values
• Colormap should follow the expressiveness principle
• Types of colormaps:

**Binary**

-1 0 +1

**Diverging**

-1 0 +1

**Categorical**

T F A

**Sequential**

3 2 1
Categorical vs. Ordered

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

[Luminance]

[Saturation]

[Hue]

[Munzner (ill. Maguire), 2014]
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 to 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

[Reference: Bergman et al., 1995]
Evaluating the Impact of Binning 2D Scalar Fields
Lace Padilla, P. Samuel Quinan, Miriah Meyer, and Sarah H. Creem-Regehr

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

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

Elevation (m)
-510
-470
-430
-390
-350
-310
-270
-230
-190
-150
-110
-75

[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

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 car•ography 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

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

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

[Padilla et al., 2017]
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!

[Padilla et al., 2017]