

MEASURING THE UTILITY OF COLOR RAMPS
IN EARTH SYSTEM SCIENCE DISCIPLINES:
A STUDY OF CONTINUOUS DATA SYMBOLOGY

By

Christy L. Steffke

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Geological Sciences – Master of Science

2015

ABSTRACT

MEASURING THE UTILITY OF COLOR RAMPS IN EARTH SYSTEM SCIENCE DISCIPLINES: A STUDY OF CONTINUOUS DATA SYMBOLOGY

By

Christy L. Steffke

This thesis seeks to determine the efficacy of communicating information in continuous value maps using Earth system science visualizations in the digital environment. The research approach consisted of two parts. First, we investigated commonly-used color ramps employed across many disciplines and devised a method to test color ramp efficacy at conveying information from continuous data maps. A Continuous Data Model (CDM) was developed for use in both studies by manipulating a digital elevation model (DEM), a common geology data construct used in a variety of visualizations and data models for analyzing and displaying topographic data. The resulting CDM was symbolized using four pervasive color ramps and used to derive 16 images from which participants estimated data values. Participants ultimately estimated values at four known map locations on four renditions of the same map. The only variable between participant estimations was the color scheme by which the maps were symbolized. Significant differences in color ramp performance were assessed using participant absolute data estimation differences from known map values as a function of the color ramp used for symbolization. Two methods for data collection were employed to provide information not only on how map readers estimate map values but also how they interact with continuous data maps. The participant-map interaction study included interaction variables that were collected using eye tracking technology and summarized using GIS. Our findings suggest that color ramps commonly used to depict ESS phenomena are not equally effective at communicating continuous map data.

For my family and friends-
Without the foundation you have built, I would not stand where I am today.
Also for Ben and Stella-
You're my rock and my flower. I am grateful for your undying support.

ACKNOWLEDGEMENTS

I would like to express my deep gratitude for my master thesis advisor, Dr. Julie Libarkin. I owe many of my academic achievements and experiences to your generosity. Thanks also to the rest of my guidance committee: Dr. Danita Brandt, Dr. Ashton Shortridge, and Dr. Stephen Thomas. Your dedication and input enhanced my academic experience and furthered my understanding far beyond traditional geological science undertakings. To the lovely office support staff: Jackie Bennett, Heidi Lynde, and Pam Robinson! Thank you for always troubleshooting paperwork and paving the way for us meagre graduate students! We could not do it without you! I am also grateful for the support and the flexibility permitted by Dr. Osvaldo Hernandez, ISP lab coordinator, to his TAs for ISP203L. Your mentoring was meaningful and the freedom to customize lessons for our students had immeasurable benefits.

Special thanks to members of the Department of Forestry who developed and launched the Graduate Certificate in Forest Carbon Science, Policy, and Management during my tenure at MSU. Your program was broadening and inspiring. I am proud to be in the first cohort of successful Certificate holders! I am also grateful for the diverse experiences afforded to me through various extra-curricular mentoring endeavors during my tenure at MSU. Thank you, Stephen Thomas, for fostering the GIS-based mentorship between Don, Jeffery, and I. I am also forever grateful for Zeynep Altinsel to trust me as a Group Facilitator of MSU's Buddy Program and for Melissa McDaniels for her guidance and partnerships in the International Teaching Assistant Orientations. Mark Stephens and John Hesse, your confidence in my GIS skills is inspiring! Finally, I would like to thank Bob Drost, Sheldon Turner, and Nicole LaDue, Terri McElhinny, Sarah Jardeleza, and Carman McCallum for their support in Julie's Lab. Your guidance is unforgettable.

TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES.....	viii
INTRODUCTION.....	1
Amazon Mechanical Turk.....	5
Eye Tracking.....	6
Research Questions.....	8
MATERIALS AND METHODS.....	11
Stimuli.....	11
Study Design.....	14
<i>Study 1: Amazon Mechanical Turk</i>	14
<i>Study 2: Eye Tracking</i>	15
Study Populations.....	16
<i>Study 1: Amazon Mechanical Turk</i>	16
<i>Study 2: Eye Tracking</i>	16
DATA ANALYSIS.....	17
Additional Consideration in Study 2: Map Interaction Variables.....	18
RESULTS.....	20
Study 1: Participant Estimations.....	20
Study 2: Participant Estimations.....	21
Study 2 Participant-Color Ramp Interaction Variables.....	23
<i>Total Gaze Plot Length</i>	23
<i>Filtered Gaze Plot Length</i>	24
<i>Time to Estimate Map Values</i>	25
DISCUSSION AND CONCLUSION.....	27
APPENDICES.....	36
Appendix 1: Tables.....	37
Appendix 2: Figures.....	41
REFERENCES.....	46

LIST OF TABLES

Table 1.	Participant estimation distribution data for Study 1 (Amazon Mechanical Turk, MT) and 2 (Eye Tracking, ET).....	37
Table 2.	Median participant estimation deviation from known map values in Study 1 and Study 2. Negative data point labels represent extreme underestimations of map values while positive data point labels represent extreme overestimations.....	37
Table 3.	Wilcoxon Signed-Rank Test results for median participant estimation deviation data for Study 1 and Study 2.....	37
Table 4.	Kendall’s Coefficient of Concordance (W) for median participant estimation deviation data for Study 1 and Study 2.....	37
Table 5.	Participant Total Gaze Path Length distribution data for Study 2.....	38
Table 6.	Participant median Total Gaze Path Length for Study 2.....	38
Table 7.	Wilcoxon Signed-Rank Test assessing Total Gaze Path Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	38
Table 8.	Kendall’s coefficient of concordance (W) assessing Total Gaze Path Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	38
Table 9.	Participant median filtered Gaze Plot Length distribution data for Study 2.....	39
Table 10.	Participant median filtered Gaze Plot Length for Study 2.....	39
Table 11.	Wilcoxon Signed-Rank Tests assessing Filtered Gaze Plot Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	39
Table 12.	Kendall’s coefficient of concordance (W) assessing Filtered Gaze Plot Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	39
Table 13.	Participant median Time to Estimate Map Values distribution data for Study 2.....	40
Table 14.	Participant median Time to Estimate Map Values for Study 2.....	40

Table 15. Wilcoxon Signed-Rank Test assessing median participant Time to Estimate Map Values mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	40
Table 16. Kendall’s coefficient of concordance (W) assessing median participant Time to Estimate Map Values mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).....	40

LIST OF FIGURES

- Figure 1. The process of developing a continuous data model (CDM): (A) original high resolution digital elevation model of drumlin field in Upstate New York with hydrology and transportation layers for reference; (B) a simple and non-visually complex sub dataset to be used as basis for CDM; and (C) CDM: a unitless and simple continuous dataset with scaled data values between 0 and 500. Note the rotational transformation from original dataset.....41
- Figure 2. Continuous Data Model (CDM) variations symbolized using four ubiquitous color ramps: TempCR, PrecipCR, ElevCR, and WindCR. Four control points of known value: A, B, C, and D, are represented once per color ramp variation. Rotational transformations (in degrees) applied to the CDM are indicated to the upper left of each variation. These sixteen color images were displayed sequentially and in random order to participants during the estimation task. Control point labels were not displayed to test participants but are displayed here for demonstration.....42
- Figure 3. Example of participant-color ramp interaction variables: Total Gaze Plot (A) Length: 26,362.9 pixels and Filtered Gaze Plot (B) Length: 8,855.5 pixels.....43
- Figure 4. Median participant estimation deviation from known map values in Study 1. The zero point of each histogram is emphasized with vertical dashed lines. Known map values at which participants estimated map data is on y-axis. Negative data point labels represent extreme underestimations of map values while positive data point labels represent extreme overestimations.....44
- Figure 5. A subset of participant Total Gaze Plot Lengths on ElevCR, WindCR, PrecipCR, and TempCR at Control Point A (known value = 4). Note the gaze pattern variations across each different color ramp. The differences in participant Time to Estimate Map Values and the accuracy differences may be explained by color ramp complexity influencing participant interactions and estimations.....45

INTRODUCTION

Visualizations are ubiquitous and extensively used to communicate information related to Earth system science (ESS) phenomena. Characteristics of visualizations that allow the communication of a wide variety of information are founded on the perceptual properties of visual variables which, in varying combinations, result in different visualization outcomes. Color, or hue difference, for example, is considered a perceptual variable which allows for the distinction between varying map features but can also be combined with lightness and saturation to imply order or quantity in maps. Combinations of color, saturation, and lightness, often in the form of color ramps, have been widely evaluated for efficacy in symbolizing thematic, or classed, maps. The efficacy of continuous data symbology, however, still lacks empirical evidence (Brewer 1999). Recent developments in mapping software have permitted the customization of color ramps, yet default color schemes are often chosen to symbolize continuous data maps without attention to communication efficacy (Moorland 2009).

The most recent revolution in cartography has promoted the transition from print to online maps, and has resulted in the mass production and dissemination of maps and map data products (Buckley & Frye 2011, McMaster and Thrower 1991, Robinson 1991). This widespread use of readily available spatial data by map makers of varying skill level creates the circumstance in which best practices of data visualization can fall through the cracks. Although guidelines for use of color in data representation exist (Brewer 1999), map makers are known to use color ramps that are ineffective and that misalign with these guidelines (Moreland 2009). In fact, cartographic design processes are known to be based in common conventions that are not aligned with empirical data (Edney 2005). Map readers are thus left to understand map data symbolized through color ramps that may hinder rather than promote communication. The

implications of easy access digital cartography and the consequence of pervasive color ramp-use have not yet been fully realized (Fu and Sun 2010) and warrants further investigation.

Visualizations function as mechanisms for communicating information both within and beyond the scientific community. In disciplines that seek detailed understanding of dynamic processes, visualizations play a valuable role in scientific understanding across a broad range of audiences (Fouh et al 2012). Different types of visualizations serve unique functions relative to the type of data displayed within them (Shneiderman 1996, Behrens 2008, Heer et al 2010, Lima 2011). In Earth system science (ESS) disciplines, spatial data visualizations, or maps, play a leading role in the understanding of dynamic Earth processes. Terrain or elevation maps, for example, are commonly used to identify regions where natural hazards create dangerous circumstances where human health may be compromised (OAS 1991). Similarly, climate projections may take the form of temperature change maps that are disseminated through both academic and non-academic settings and in a variety of forms (US GCRP 2009). Visualizations are a valuable scientific tool

A fundamental feature of scientific visualizations is the use of color to represent continuous data (Moreland 2009). Salient symbologies in cartography (e.g. color, line thickness, highlighting, shading, ruling, geometric pattern, animation) function to enhance a map reader's interaction with visual media (Wright 1942, Tversky 2010). Color, however, is used extensively to convey data because color is aesthetically pleasing and is easy to decipher through visual inspection alone (Brewer 1999; Moreland 2009). At the same time, color can have significant impacts on map reader understanding (Rogowitz and Treinish 1996, 1998).

Color choice in cartography has moved beyond simple aesthetic considerations to include considerations of visual variables that impact understanding, such as hue, saturation and

lightness (Harrower and Brewer 2003). Studies have focused on the least effective color ramps, with ample evidence suggesting that some ubiquitous ramps, such as the rainbow color ramp, are ineffective in many settings (Borland and Taylor 2007, Light and Bartlein 2004, Rogowitz and Treinish 1996, Rogowitz and Treinish 1998, Ware 1988, Ware 2004, Tufte 1997, Pizer and Zimmerman 1983, Rheingans 1992, Healey 1996, Brewer 1999). Despite these studies, color ramp choice is grounded in historical norms and ineffective color ramp use is still quite common (Edney 2005).

The persistent use of common, but potentially ineffective, color ramps has long raised questions about processes used to create specialty and thematic maps (Robinson 1952). Map makers may use ad hoc (Wang and Shen 2011), theoretical (van Wijk 2005), or empirical (Brewer 1999) approaches to generate maps. Ad hoc approaches rely on easy access to map making programs and disciplinary conventions to generate map color schemes (Wang and Shen 2011, Edney 2005, Moreland 2009). Theoretical approaches have utilized, for example, economic models to identify color ramps that are efficient in terms of time required to generate and use them (van Wijk 2005). Other theoretical approaches use perceptually ordered color systems to assign color to data, as used in the Munsell or CIELAB color classifications (Brewer 1999). Finally, empirical approaches match the perceptual dimensions of color (hue, lightness, and saturation) with the organization of data being represented (Brewer 1999) or focus on measuring visual data within an image and optimizing a map-reader's understanding (Chen 2010). The empirical approach to assigning map data is arguably the most advanced because it merges information theory with empirical user data (Brewer 1999, Chen 2008, Chen 2010).

Poor map color choice is concerning because visualization techniques stand as the lenses through which map readers interpret map data. Information lost through inaccurate interpretation

of map color is recognized in the data visualization pipeline that forms the basis of the information theory concept of message transmission (Wang and Shen 2011, Purchase et al 2008). Message transmission consists of multiple stages beginning with the transmission of an encoded message (e.g. temperature data) that passes through a noisy communication channel (e.g. a colored map). The noisy communication channel ultimately influences the decoded message received by the visualization user (e.g., the map reader; Chen and Janicke 2010). Thus, color choice specifically in map symbology can be especially expensive in terms of the accuracy and efficiency by which map readers understand information in maps.

The visualization pipeline is recognized as an important construct for map design (Card et al 1999, Chi 2000, dos Santos and Brodlie 2004, Haber and McNabb 1990, Tominski 2006). The ability of the map-maker to encode visual data and the ability for users to decode map data are dependent upon the transparency of the visual communication channel. In the case of displaying continuous map data, the color ramp used to symbolize information in maps functions as the visual communication channel (Bamford 2003). Thus, choice of color ramp symbology for a map dataset strongly influences map reader understanding.

The Earth system sciences appear to use an ad hoc approach to choose color ramps (Brewer 1999, Edney 2005, Moreland 2009, Wang and Shen 2011). This likely occurs because of a lack of empirical data supporting or refuting the communication usefulness of each color ramp (Brewer 1999) and the ease of access to specific color ramps in mapping software (Borland and Taylor 2007). For example, multi-hue color ramps are the default representation in 8 out of the 9 common visualization programs, and are used in many visualization papers (Borland and Taylor 2007). Very few empirical user studies exist that evaluate the usability of different color ramps, despite multiple calls (Brewer 1999) and a clear need to improve user access to

continuous data sets to enhance public understanding of science and decision making (OAS 1991, Collins et al 2013).

This study addresses the need to better understand how continuous map data symbology impacts estimation of map values by map users. In particular, this work measures general public and college student estimation accuracy and visual interaction with an elevation dataset symbolized using various color ramps, in order to understand how well common color ramps pervasive to ESS disciplines communicate information. Two user studies were conducted and are reported here. The first study utilizes Amazon Mechanical Turk to collect data on participant estimation of maps as a function of color ramp and the second study utilizes eye tracking technology and a common GIS platform to measure participant estimation and interaction with maps as a function of color ramp symbology. Results to these studies provide a gauge on the effectiveness of four commonly used color ramps used to symbolize continuous data within ESS visualizations.

Amazon Mechanical Turk

Amazon Mechanical Turk (MTurk) is an internet crowdsourcing platform that allows researchers to quickly sample a large and diverse workforce (Pontin 2007, Buhrmester et al 2011) for almost unlimited research endeavors. MTurk, as an online labor market, allows researchers or requesters, to develop and launch Human Intelligence Tasks (HITs) designed to survey MTurk workers. Pre-screening MTurk Workers to meet pre-defined qualifications for research subject pre-screening (Paolacci et al 2010) is an added benefit that can quickly and inexpensively satisfy a search for a particular participant pool. Evidence supports that the MTurk population is at least as representative of the US population as traditional subject pools in terms

of race, gender, age, and education (Paolacci et al 2010) and is significantly more representative of the US population than college undergraduate samples (Buhrmester et al 2011).

MTurk is a recognized resource for visualization studies in which visualization viewers' ability to interpret graphics have been successfully measured and validated (Heer and Bostock 2010). Early on, MTurk was used to study user interaction with advertisements, websites, and game design (Mason and Suri 2010). Researchers have used MTurk to identify best practices in symbol positioning (Heer and Bostock 2010), visual cuing (Crump et al 2013), image tagging (Hwang and Grauman 2010), and human perception in social media (Biel et al 2011). MTurk studies specifically related to color also exist, such as research into color and emotion (Volkova et al 2012), color recognition (Branson et al 2010), color choice (Schloss and Palmer 2014), and color impairments (Lin et al 2013). MTurk is a cost effective way to collect a quality dataset on human perception of visual media.

Eye Tracking

Eye tracking (ET) is a method of data collection aimed at deciphering and recording where visual attention is focused. Historically, eye tracking was a crude process in which eye movements were recorded using invasive physical contact with the eye and recording media (Jacob and Karn 2003). Corneal light reflections were then used to measure eye position relative to photographic plates in 1901 (Dodge and Cline), and yet now even more advanced eye tracking techniques are employed that allow for non-stationary participants (Bulling and Gellersen 2013) and no sharp visible lights (Zhu and Ji 2005) or direct physical contact with the eye (Haro et al 2000). Early on, eye tracking was primarily used to gain an understanding of eye movements but is now growing as a method for collecting data on how people interact with visual media

(Schiessl et al 2003). Web page usability is a popular domain in which eye movement and performance data is used to enhance website interfaces based on human perception (Cowen et al 2002). The recording and analysis of eye movement data has facilitated the better understanding of viewer perception with visual media.

In early visual media studies using eye tracking, eye movement data was paired with participant performance data, but lacked the inclusion of a dynamic time element. Eye tracking data, including a temporal component, lends itself well to improving dynamic visual interfaces by providing more specific information on participants' interaction with visual media (Bucher and Schumacher 2006). Visual attention patterns have been assessed using eye movement data, participant preference data, and temporal information from eye tracking experiments to better understand media viewer attention patterns (Bucher and Schumacher 2006). These three types of participant data derived from eye tracking provide a dynamic look at participant interaction with visual media.

As a participant's visual attention is diverted across a sample of visual media, an eye tracker records time and position information related to each eye's movement. A number of metrics are used by eye tracking researchers in their analysis of salient element interactions that are first collected as large point datasets with time components and other information related to pupil dilation (Wang 2009) and latent eye movements (Andersson et al 2010). Each metric can be filtered and assessed separately depending on data validity (Tobii 2011). The raw eye tracking data for even a short session can generate thousands of eye movement recordings which when considered together, can be measured and assessed as a gaze plot through which visual attention can be mapped (Huang and Pashler 2007). Eye track data can further be understood, manipulated, and measured using Geographic Information Science (GIS) technology (Opach and

Nossum 2011). The frequency at which individual eye trackers record eye movements determines the number of data points each gaze plot dataset will contain; the higher frequency the eye tracker, the better suited for analysis using GIS (Dykes et al 2007) that reaches far beyond typical visual analysis of eye tracking data (Raschke et al 2014).

Specific to the current study are interpretations map readers make of the salient, or visually important, elements of visualizations. Specifically, continuous map data estimations are of primary interest in this article as such data provides information directly related to the content conveyed by the continuous data model (CDM) developed for this study and the effort that map readers must expend to extract that map content. Participant estimation accuracy derived from various interactions with the CDM symbolized using competing color schemes, can then be compared in order to provide evidence for map color scheme impact on the map reader's ability to perceive and understand map data as a function of map color symbology. Both accuracy and reaction time are frequently used as measures of success in cognitive experiments (Lloyd and Bunch 2014). Interestingly, eye tracking metrics convey different interpretations depending upon other user data (Poole and Ball 2006). For example, long gaze paths may convey interest (Salvucci and Goldberg 2000), confusion (Poole and Ball 2006), or information pursuit (Iacono, W. and Lykken, D. 2007). As such, eye tracking is a valuable tool that allows for a deeper understand of participant interaction with visual media and facilitates an environment in which improvements can be made to visual media for comparative analyses (Raschke et al 2014).

Research Questions

In this study, the primary goal is to evaluate the performance of ubiquitous color ramps as conveyors of information in continuous value maps. Does the color ramp used to symbolize a

simple continuous dataset impact the accuracy of map value estimation? We hypothesize that estimations will differ. Based on prior studies, we predict that estimations will be most accurate for single-hue color ramps (Tufte 1997) and least accurate for the multi-hue color ramps (Borland and Taylor 2007, Light and Bartlein 2004, Rogowitz and Treinish 1996, Rogowitz and Treinish 1998, Ware 1988, Tufte 1997, Pizer and Zimmerman 1983, Ware 2004, Rheingans 1992, Healey 1996, Brewer 1999). Does color ramp impact participants' visual interaction with maps during the estimation process? We hypothesize that participants will spend more effort engaging with maps symbolized using multi-hue color schemes as the cognitive effort required to understand such outweighs map interaction efficiency as found in research with categorical map data (Lee et al 2013). Taken together, this study seeks to determine whether map viewer effort interacting with continuous value maps align with differences in map value estimation as a function of color ramp used to symbolize continuous value map data.

In Study 1, we launched an Amazon Mechanical Turk (MTurk) Human Intelligence Task (HIT) in which color estimation tasks, a demographics survey including a quality control question, and a color sensitivity test were presented to MTurk Workers across the United States. MTurk Workers estimated map values at four control points within images that presented the same data but varied in terms of the color ramp used to symbolize the images. Average error in MTurk Worker estimations were calculated for each color ramp [participant estimated-true map value] and the significant differences between color ramp performances were assessed using the Wilcoxon Signed-Rank Test.

Finally, in Study 2, an eye tracking study was performed in order to collect data on participant interaction with continuous data maps while they performed the task of estimating values from the same set of images used in Study 1. In addition to participant estimation data, the

eye tracking study resulted in gaze measurements related to participant visualization of the estimation tasks: Total Gaze Plot Length (TGPL), Filtered Gaze Plot Length (FGPL), and Time to Estimate Map Values (TtEMV) which all serve as metrics for participant effort in engaging with and estimating data values from each map. These variables were also compared between color ramps to identify differences in participant interaction with continuous data maps and alignment with map estimation accuracy as a function of color ramp.

MATERIALS AND METHODS

Study 1, a crowdsourcing study, was designed to survey a large number of participants on their ability to estimate values from continuous data maps symbolized by four pervasive color ramps. Differences in performance across color ramps were quantified based on participant ability to estimate values within each map as related to the known map value at each control point. Amazon Mechanical Turk (MTurk) was used as the crowdsourcing platform in study 1 to collect data from a large number of participants. Study 2, an eye tracking study, aimed to determine whether or not participants interacted differently with maps while estimating data values again as a function of color ramp. An eye tracking system was used in Study 2 to record information about participant interaction with map data while they estimated map values. Resulting eye tracking data was compared across color ramps and served to determine the participant effort required to estimate values in each image. Both studies required development of a Continuous Data Model (CDM) from which, known data values and locations were presented to users in estimation tasks. This allowed for the calculation of the difference between participant estimations and known map values as a measure of participant estimation accuracy.

Stimuli

A continuous data set that represents a non-visually complex data range across an area was developed for use as the continuous data model (CDM) for this study. This dataset was derived from a high resolution topographic base map that was downloaded at no cost from the National Map Viewer and Download Platform managed by the U.S. Geological Survey National Geospatial Program (Figure 1a). A region containing a simple range of elevation values was downloaded as a Digital Elevation Model (DEM). This DEM was converted into an ESRI Grid

using ArcGIS for Desktop version 10.2 in order to re-scale the output raster such that the grid cell geometry was square. The resulting grid was cropped again in ArcGIS down to a simple and non-visually complex sub dataset (Figure 1b). The real range of data values for the CDM (130-195 meters) was scaled between 0 and 500 and units were removed in order to minimize the cognitive load for participants completing the value estimation tasks. This base image depicting a unitless and simple continuous data set with scaled data values between 0 and 500 served as the CDM for this study from which all image variations were derived. The CDM was then symbolized using four pervasive color ramps based on the known range of data values within the image.

Four color ramps were chosen to symbolize the CDM for user estimation comparisons in this study. Despite the importance of map color and the large variety of information symbolized in ESS visualizations, a relatively small number of color ramps persist (Romanach et al 2014). Four color ramps are particularly pervasive and will serve as the foci for this article. Each color ramp selected was chosen based on its ubiquitous presence in ESS visualizations. ElevCR is commonly used to symbolize elevation data such as in shaded relief maps published by the United States Geological Survey (USGS 2002). This color ramp is characterized by the highest data values being symbolized using white “for snow-capped peaks, the next brown for treeless areas above the tree line, then light green for sparse vegetation on the upper slopes and a darker green for the verdant valleys” (Buckley 2008). This many hue color ramp can be classified as a rainbow color ramp or a diverging multi-hue color scheme (Brewer 1999). A simpler color ramp, WindCR, is similar to the color ramp used by the National Weather Service (NWS) of the National Oceanic Atmospheric Administration (NOAA) for wind speed and wind direction maps (NOAA 2014a). For example, the National Weather Service (NWS) uses this color ramp to

convey wind speed and wind direction (NOAA 2014a). WindCR is characterized by a single hue which varies in lightness across the data range and can be described as a single hue sequential color scheme (Brewer 1999). A moderately complex color ramp, PrecipCR, is similar to color ramps used to illustrate precipitation amount by the NWS and other meteorological data sources (NOAA 2014b). This two hue color ramp is often used to illustrate precipitation, as used by NWS and other meteorological services (NOAA 2014b). Finally, TempCR, another moderately complex color ramp, is often used to symbolize temperature distribution maps including those published by the Intergovernmental Panel on Climate Change (Collins et al 2013). This three hue color ramp is very familiar as a temperature conveyer, such as those published by the Intergovernmental Panel on Climate Change (Collins et al 2013).

Each color ramp was used to symbolize the CDM to create four variations of the continuous dataset. The resulting set of four base images were fundamental to gauging participant understanding of and interaction with data symbolized using the four pervasive color ramps. The Minimum-Maximum standard histogram stretch in ArcGIS was applied to the datasets in order to normalize any color ramp value groupings. This contrast-stretching enhancement was applied to the rendered screen displays of each image to ensure color ramp symbology comparability across images. Four control points (Cps) of known value were chosen based on their position within the scaled range of the symbolized CDM legend. Control point A (CpA) had a known value of 4 while CpD had a known value of 113, CpC of 317, and CpB of 483. Each color ramp was represented four times in the survey, each containing one of the Cps of known value. To prevent the participants from recognizing that each color ramp variation represented the same dataset, so as to prevent participant ability to game the value estimation tasks, a rotational transformation of 0, 90, 180, or 270 degrees was performed on each image

dataset. A total of 16 ramp-control point combinations were displayed in random order to participants in the studies discussed in this article (Figure 2).

Study Design

Study 1: Amazon Mechanical Turk

In this study, a Human Intelligence Task (HIT) was created which included the 16 color ramp estimation questions, a demographics survey with an included discrete quality control question, and a color sensitivity test consisting of eight Ishihara Color Plates (Ishihara 1954, 1958). The Title of the HIT which was visible to MTurk Workers and described the task to the participants, “Estimate Values in Colorful Images!”, was designed to be enticing and as specific as possible so as to attract MTurk Workers. Further, the purpose of the HIT Description, “View 16 color maps displaying numerical data and estimate the value in a single location in each map”, was to give participants more information on the HIT tasks before they decide to view and pursue the HIT. The following keywords were provided to help Mechanical Turk Workers search for our HIT: estimate, values, image, data, color ramp, raster.

To ensure that data reflected a general audience that could produce reliable responses, we solicited only MTurk Workers who met at least the following qualification requirements: 1) participant HIT Approval Rate for all Requesters’ HITs had to be greater than or equal to 95%, 2) each participant had to have had at least 1000 HITs approved in the past, and 3) participant locations had to be within the United States. MTurk Workers were paid between \$0.40 and \$0.50 to complete each HIT.

Four criteria were used to remove data from the study sample. First, participants were asked to answer a quality control question that identified participants who were responding

without reading the questions. Six participants who failed to answer this quality control question correctly were omitted from this study. Second, estimations that fell outside of the 0-500 range were considered to reflect inattention to the task and were removed (n=13 participants). Third, 30 participants submitted duplicate surveys; all duplicates except for the first submission were removed. Finally, because this study relied heavily on participants' perception of color, a color sensitivity test was presented at the end of each HIT. Participants were prompted to type the number seen in the pattern of dots on each of eight Ishihara Color Plates. Participants (n=8) who answered more than one color plate question incorrectly were omitted. 284 responses were retained in the study.

Study 2: Eye Tracking

Participants in Study 2 completed an eye tracking task similar to Study 1. Participants in this study all successfully completed an eight plate Ishihara color sensitivity test. These participants were instructed to spend as much time needed in estimating the values in the images and to speak aloud their estimation for recording by a researcher. After each estimation, the researcher would then click a key on their keyboard to proceed to the next image. No participant was allowed to return to previous images. Participants also participated in a retrospective interview to provide additional information about their interaction with the task. The sixteen images were presented in the same shuffled sequence as in Study 1. Following the instructions to estimate values within the subsequent images was a Demographics Survey. Thirty two undergraduate students from a large Midwestern university completed the eye tracking study as part of an hour-long session of other eye tracking studies for which, each participant was compensated \$25. Participant eye tracking data that had <50% samples percent as a rough measure of eye tracking recording data quality were removed (n=8 removed). 24 responses were

retained in this study. Participant eye movements were monitored during the estimation tasks with a table-mounted eye tracking system, Tobii T60, at a sample rate of 60 Hz. A PC running Tobii Studio 3.0 was used for data acquisition.

Study Populations

Study 1: Amazon Mechanical Turk

Amazon Mechanical Turk (n=284) were 146 men and 137 women. Nineteen percent of participants were non-white and sixty-one percent had a college degree. Forty-nine percent of participants were Very Confident in their ability to read and understand maps, while forty percent of participants were Somewhat Confident. The remaining eleven percent of participants was either A Little Confident or Not Confident in their ability to read and understand maps. The age of participants ranged from eighteen to sixty- nine years ($M = 34.1$, $m = 31.0$, $SD = 11.7$). Age was non-normally distributed, with skewness of 0.944 ($SE = 0.143$) and kurtosis of 0.019 ($SE = 0.284$).

Study 2: Eye Tracking

Eye tracking study participants (n=24) were thirteen men and eleven women. Forty-two percent of participants were non-white and 58 percent had a college degree. Twenty-nine percent of participants were Very Confident in their ability to read and understand maps, while forty-two percent of participants were Somewhat Confident. The remaining thirteen percent of participants were either A Little Confident or Not Confident in their ability to read and understand maps. The age of participants ranged from eighteen to sixty- one years ($M = 24.3$, $m = 20.0$, $SD = 9.9$). Age was non-normally distributed, with skewness of 2.559 ($SE = 0.481$) and kurtosis of 7.297 ($SE = 0.935$).

DATA ANALYSIS

In both studies, color ramps that had the lowest average errors estimated by participants in each study were deemed the most effective color ramp at communicating continuous data map information. Those color ramps that resulted in participant estimations of greatest significant difference from control point values were deemed the least effective at communication information in continuous data maps. Finally, color ramp performance was compared to the participant interaction variables generated in the eye tracking study, in order to determine the color ramp that resulted in the fastest, least effortful, and accurate estimations. In both studies, Wilcoxon Signed-Rank Tests were conducted to evaluate whether median participant estimations varied significantly as a function of color ramp.

The Wilcoxon Signed-Rank Test is the non-parametric alternative to the dependent samples t-test which compares means of two related groups to determine whether or not significant differences exist between group means. Variations of this statistic can be used for repeated measures studies in which more than one dataset is analyzed for the same participants, such as what we have in this study with participant estimations across the same continuous value map symbolized using four pervasive color ramps. A simple evaluation of the average participant estimation datasets in each study indicates average differences across all four color ramps evaluated in this study. The Wilcoxon Signed-Rank Test was used to evaluate the significance of these differences between color ramp efficacies in conveying continuous value map data. The Wilcoxon Signed-Rank statistic was also employed to determine significant differences in participant-color ramp interaction variables resulting from the Study 2. Kendall's W or Kendall's coefficient of concordance for ranks (W) is a variation of the Friedman statistic and is used to test for significant differences between k-related samples which cannot be assumed to fit a

normal distribution. This statistic was used to verify the significant differences between average participant estimations as a function of color ramp.

Additional Consideration in Study 2: Map Interaction Variables

The resulting spatial data was time stamped and served as the basis for determining participant-color ramp interaction variables: Total Gaze Plot Length (TGPL), Filtered Gaze Plot Length (FGPL), and Time to Estimate Map Values (TtEMV). Raw or unfiltered eye tracking data was imported into ArcGIS in order to be filtered and clustered using custom Python scripts that employ Olsson's interpolation method (2007) and Salvucci and Goldberg's filter recommendations (2000). An unfiltered eye tracking dataset for an individual represents that individual's Total Gaze Plot Length (Figure 3A) which is markedly longer in measurement than the corresponding Filtered Gaze Plot (Figure 3B). Each raw eye tracking dataset contains both invalid data and sporadic gaze measurements called saccades, which demonstrate the sporadic sampling method of the human eye. Excessive saccadic activity may indicate confusion during task-based assignments but is ultimately washed out during the clustering of eye tracking data. Raw eye tracking data must first be filtered to remove data defined by one or more poor validity measurements recorded during instances of blinking or other times when the eye tracker could not properly constrain the eye position. Olsson's (2007) interpolation method was used to interpolate participant gaze where invalid data was removed. The filtered gaze data was then grouped into fixations, or clusters, to eliminate data points that represented instances when the participants were not focusing their gaze. Salvucci and Goldberg's (2000) data filter was used to process the data and is based on spatial dispersion of the eye tracking data. Using this method, a cluster or fixation is defined as being a set of eye tracking points that are in close proximity in

both space and time. The distance threshold of 35 pixels was used for clustering as per default recommendations in common eye tracking software (Tobii 2010). The minimum fixation duration is generally accepted by visual scientists to be 80ms, so with our eye tracking data collected every 1/60th of a second, is 5 sequential points or ~83.33ms. The resulting clustered or Filtered Gaze Plot marks the path which the participant viewed while progressing through the task of estimating map data values. Because eye tracking data is time-stamped, Time to Estimate Map Values (TtEMV) was also recorded and used to indicate participant effort during task-oriented data estimation instances.

RESULTS

Study 1: Participant Estimations

In this study, 341 participants estimated continuous data values as part of a Human Intelligence Task (HIT) developed using the Amazon Mechanical Turk (MTurk) crowdsourcing tool. MTurk Workers who failed to pass the Ishihara Color Sensitivity Test subset were removed from the dataset (n=8 removed) as were those who failed to answer the quality control question properly (n=6 removed). Another method of quality control was to remove MTurk Worker data estimations that fell outside of the 0-500 range. Such inappropriate answers suggested inattention to the task and were removed (n=13 removed). When given the chance to provide comments about the estimation tasks, no participants indicated that they recognized the images as containing the exact same dataset. Because there was no simple way to avoid the same MTurk Worker repeating a posted HIT more than once, we used brute force methods to identify and remove duplicate responses. We found that 12 MTurk Workers took the survey twice, 3 MTurk Workers took it three times, and 1 MTurk Worker took it four times. MTurk Worker submission timestamp information was used to identify subsequent submissions by the same user. Duplicate submissions were isolated and removed. Only the first response from each MTurk Worker was retained in the dataset (n=30 removed). The remaining data (n=284) were used in analysis.

The distributions of participant estimation data in Studies 1 and 2, and participant interaction data in Study 2 were non-normal (Table 1) thus the median was chosen to report results and non-parametric statistics were chosen to analyze the matched map data. Median participant estimation deviation (+/- SD) for each color ramp in Study 1 is displayed in Table 2. Negative data labels represent participant underestimations of map values while positive data labels represent overestimations. A Wilcoxon Signed-Rank Test was applied to the median

participant estimation datasets for each color ramp in Study 1 to evaluate whether or not participant ability to estimate map data values varied significantly as a function of color ramp. The statistic indicated a significant difference in median participant estimations made using ElevCR (median underestimation of 1.0, SD +/-83.0), and WindCR (median underestimation of 17.5, SD +/-89.1), $z = -9.234$, $p < 0.001$. There was no significant difference between ElevCR and PrecipCR (median underestimation of 3.0, SD +/-58.7) performance. There is a significant difference in performance of ElevCR and TempCR (median overestimation of 5.0, SD +/-30.4), $z = -6.483$, $p < 0.001$. There also existed a significant difference in PrecipCR and WindCR, $z = -8.787$, $p < 0.001$. There existed a significant difference in TempCR and WindCR, $z = -10.084$, $p < 0.001$. Finally, a significant difference existed between TempCR and PrecipCR, $z = -5.729$, $p < 0.001$ (Table 3). To further validate the differences in color ramp performance, Kendall's coefficient of concordance (W) was calculated to determine the general disagreement among color ramp performances. In Study 1, the coefficient of 0.220 ($p < 0.001$) indicates a low degree of agreement between color ramp estimation deviation averages (Table 4).

Study 2: Participant Estimations

In this study, thirty-two undergraduate students from a large Midwestern university estimated map values in the same images from Study 1, as part of an hour-long combined eye tracking session. Due to the nature of the researcher-participant interaction during an eye tracking study, no participant datasets needed to be removed due to compromised quality. However, participant eye tracking data that reported <50% samples percent, a rough measure of eye tracking recording data quality, were removed (n=8 removed). 24 responses were retained in this study. Participants were given a color sensitivity test prior to beginning the experiment. No

participants in this portion of the study had color vision impairments and were thus able to complete the estimation tasks. Before beginning each experiment, the eye tracking system was calibrated based on the participant's individual gaze characteristics. The researcher instructed each participant to speak aloud their estimation after viewing each image long enough to determine the value within the image control point. The participant speaking aloud their estimation for each image indicated to the researcher that the participant was ready to move to the next image to begin the subsequent estimation task. Upon hearing this signal, the researcher would then click a key on their keyboard to proceed to the next image which would indicate to the participant that a new estimation task has begun. No participant was allowed to return to previous images. Upon conducting a post-study interview, no participants indicated that they recognized the images as containing the exact same dataset. Participant estimation data (n=24) was tabulated by the researcher and summarized. Median participant estimation deviation data for Study 2 is displayed in Table 2.

A Wilcoxon Signed-Rank Test was applied to the median participant estimation datasets for each color ramp in Study 2 to evaluate whether median participant estimations varied significantly as a function of color ramp (Table 3). The statistic indicated a significant difference in median participant estimations made using ElevCR (median overestimation of 1.5, SD +/- 149.4) and WindCR (median underestimation of 25.0, SD +/- 88.3), $z = -4.029$, $p < 0.001$. There was no significant difference in performance between ElevCR and PrecipCR (median underestimation of 4.0, SD +/- 38.5) or with ElevCR and TempCR, TempCR (median overestimation of 2.0, SD +/- 24.5). There existed a significant difference in PrecipCR and WindCR, $z = -3.272$, $p < 0.01$. There was also a significant difference performance of TempCR and WindCR, $z = -3.972$, $p < 0.001$. Finally, no significant difference existed between TempCR

and PrecipCR (Table 3). For Study 2, Kendall's coefficient of 0.441 ($p < 0.001$) indicates a fair degree of agreement between median estimation datasets collected as a function of color ramp. However, the sample size ($n=24$) is too small to demonstrate significant agreement between average participant estimations as a function of color ramp (Table 4).

Study 2 Participant-Color Ramp Interaction Variables

Total Gaze Plot Length

The Total Gaze Plot Length (TGPL) is derived from unfiltered eye track measurements which are recorded at a sampling rate of sixty per second. Unfiltered eye tracking data includes information on participant overt visual attention but also maintains saccadic eye movement information that could inform other research in visual science. Unfiltered eye tracking data was exported from Tobii Studio 3.0 and imported into ArcGIS for Desktop version 10.0 using custom Python scripts. Once imported, the TGPL was able to be visualized and measured for each participant using the calculate geometry tool of the gaze path lines created to symbolize progression from one gaze measurement to the next (Figure 3A). Each participant estimated values in 16 images derived from the CDM. Sixteen TGPL datasets for each participant were tabulated and summarized using GIS. Participant TGPL distribution data is depicted in Table 5 and median participant TGPL in pixels is displayed in Table 6.

A Wilcoxon Signed-Rank Test was applied to the median TGPL datasets for each color ramp in Study 2 to evaluate whether median TGPL, in pixels, varied significantly as a function of color ramp (Table 7). The statistic indicated a significant difference in median TGPL for ElevCR (median TGPL of 13729.43 pixels, SD +/- 8602.52) and WindCR (median TGPL of 17184.52 pixels, SD +/- 9672.45), $z = -3.314$, $p < 0.01$. There was no significant difference in

performance between ElevCR and PrecipCR (median TGPL of 14725.15 pixels, SD \pm 7934.74) or with ElevCR and TempCR (median TGPL of 15508.88 pixels, SD \pm 10481.28). There existed a significant difference in PrecipCR and WindCR, $z = -2.829$, $p < 0.01$ and also between TempCR and WindCR, $z = -2.457$, $p < 0.05$. Finally, no significant difference existed between TempCR and PrecipCR (Table 7). For the TGPL dataset in Study 2, Kendall's coefficient of 0.238 ($p < 0.001$) indicates a low degree of agreement between median TGPL datasets collected as a function of color ramp (Table 8).

Filtered Gaze Plot Length

Filtered Gaze Plot Length (FGPL) is derived from the Total Gaze Plot Length dataset. Unfiltered eye track measurements are clustered based on adjacency in time and space (Salvucci and Goldberg 2000) and invalid eye tracking data was removed as defined using the data acquisition software's definition of validity. The resulting eye tracking data points that were clustered in both space and time were determined to comprise a single fixation. Each fixation was represented by a single point which had the spatial measurement of the average x,y location of each data point comprising the fixation (Figure 3B) ArcGIS for Desktop version 10.0 was used again to measure the distances between each point using the calculate geometry tool for the connecting gaze path lines. Filtered GPL measurements for each participant were tabulated and summarized using GIS. Participant FGPL distribution data is displayed in Table 9 while median FGPL is displayed in Table 10. Significant differences in participant interactions with continuous dataset as a function of color ramp were determined using Wilcoxon Signed-Rank Tests.

A Wilcoxon Signed-Rank Test was applied to the median FGPL datasets for each color ramp in Study 2 to evaluate whether median FGPL varied significantly as a function of color ramp (Table 11). The statistic indicated a significant difference in median FGPL for ElevCR

(median GPL of 3566.5 pixels, SD +/- 2733.92) and WindCR (median FGPL of 5242.38 pixels, SD +/- 3228.8), $z = -3.943$, $p < 0.001$. There was no significant difference in performance between ElevCR and PrecipCR (median FGPL of 3990.78 pixels, SD +/- 2631.69) or with ElevCR and TempCR (median FGPL of 3886.74 pixels, SD +/- 3877.59). There existed a significant difference in PrecipCR and WindCR, $z = -3.743$, $p < 0.001$ and in TempCR and WindCR, $z = -3.343$, $p < 0.01$. Finally, no significant difference existed between TempCR and PrecipCR (Table 11). For the FGPL dataset in Study 2, Kendall's coefficient of 0.275 ($p < 0.001$) indicates a low degree of agreement between median TGPL datasets collected as a function of color ramp (Table 12).

Time to Estimate Map Values

Time to Estimate Map Values (TtEMV) in milliseconds (ms) can indicate the effort a participant expended completing the estimation task. Using a Tobii T60 eye tracking system, data was recorded at a sampling rate of 60 measurements per second. This information was time-stamped and was used in ArcGIS for Desktop version 10.0 to quantify the time each participant spent estimating each value in each image. The participant median TtEMV distribution data for Study 2 can be found in Table 13. TtEMV median data in seconds for each participant are tabulated and summarized in Table 14. Significant differences in participant TtEMV in continuous datasets as a function of color ramp were determined using Wilcoxon Signed-Rank Tests.

A Wilcoxon Signed-Rank Test was applied to the median TtEMV datasets for each color ramp in Study 2 to evaluate whether median TtEMV varied significantly as a function of color ramp. The statistic indicated a significant difference in median TtEMV for ElevCR (median TtEMV of 6595.0 ms, SD +/- 3983.83) and WindCR (median TtEMV of 7,064.0 ms, SD +/-

3936.6), $z = -2.629$, $p < 0.1$, for ElevCR and PrecipCR (median TtEMV of 5422.0 ms, SD +/- 3065.98), $z = -2.314$, $p < 0.5$, and for ElevCR and TempCR (median TtEMV of 6037.5 ms), $z = -2.629$, $p < 0.1$. There existed a significant difference in PrecipCR and WindCR, $z = -4.229$, $p < 0.001$ and in TempCR and WindCR, $z = -3.457$, $p < 0.01$. Finally, no significant difference existed between TempCR and PrecipCR (Table 15). For the TtEMV dataset in Study 2, Kendall's coefficient of 0.398 ($p < 0.001$) indicates a fair degree of agreement between median TGPL datasets collected as a function of color ramp (Table 16). However, the sample size ($n=24$) is too small to demonstrate significant agreement between average participant TtEMV as a function of color ramp (Table 16).

DISCUSSION AND CONCLUSION

In this study, we investigated the efficacy of four pervasive color ramps used commonly in Earth System Science visualizations to symbolize continuous value data in maps. Participant interaction with and ability to estimate data values from continuous value maps was found to vary as a function of the color ramp used to symbolize map data. ElevCR, which is often charismatically used to symbolize elevation and terrain datasets in Earth system science disciplines, surprisingly tended to induce an approximately accurate estimation of map values, despite the multi-hue nature of its color scheme. ElevCR did not perform statistically differently in Study 1 or Study 2 than PrecipCR, another multi-hue color ramp. Further, participants estimating values in ElevCR tended to expend less effort as measured by Total Gaze Path Length and Filtered Gaze Plot Length while estimating than in any other colored map estimation activity. Finally, on average, it took a significantly longer time for participants to estimate map data values using ElevCR over PrecipCR and TempCR, yet WindCR, the simplest single-hue color ramp, took the most time.

Characteristics of ElevCR that could have played a role in influencing participant estimation and interaction with the data include ElevCR's unique combination of characteristics related to the perceptual dimensions of color (hue, lightness, and saturation). ElevCR shares qualities with that of the infamous rainbow color ramp yet did not induced the expected failed and deviant estimations by the participants included in this paired study. Both participant estimations of map values and interactions with the continuous value maps varied in part due to the unique composition of ElevCR but likely for a variety of reasons not evaluated in this study. This study aimed to evaluate the efficacy of common color ramps for symbolizing continuous

value map data. Further studies will be needed to delve deeper into map performance due to specific color ramp characteristics.

A confounding characteristic of ElevCR that could have influenced participant estimation and interaction with maps in this study is that at either end of the ElevCR scale bar, there exist like-hues which vary similarly in lightness. This characteristic of the ElevCR color scheme of having two light hues at either end of the scale is unique among the color ramps evaluated in this study. PrecipCR and TempCR both tend to have smoother transitions between hues and do not have the unusual lightness variation on either end of the scale bar as does ElevCR. Due to this enigmatic lightness variation (Buckley 2008) on either end of ElevCR, map reader attention may have been diverted with inaccurate estimations induced. In effect, this attention diversion would require more effort expenditure by the map reader to understand the map data and thus possibly inducing a longer time for map value estimation. One could speculate whether or not the two like-hues at either end of ElevCR induced map reader error in both high and low end of the scale by looking at the data from this study.

Figure 4 depicts median participant estimation deviation from known control point values in Study 1, the large Mechanical Turk study in this article. Interestingly, map readers tended to over-estimate while estimating map values on the low end of the scale bar and also tended to under-estimate map values on the scale bar's high end. It is logical that map readers would estimate not beyond the 0-500 range of map data as they were instructed. Over-estimations at the low end of the scale bar would be thus due to the underestimation limit established by the lower range of the study data. Similarly, participants could not possibly overestimate beyond the upper end of the study's map data range which would thus result in a median underestimation of map values at the high end of the scale. A pattern in map reader estimation deviation at the low end of

the scale bar and the high end of the scale can be evaluated further in future research. Further, Figure 4 depicts a possible pattern in median participant estimations using ElevCR that differs somewhat from the other three color ramps addressed in this study. It is not surprising that variations in hue and lightness at either end of the ElevCR scale bar induced a somewhat bimodal distribution of participant estimations with overestimating being prevalent on the low ends of the scale bars while underestimating was prevalent on the high ends.

Humans are most visually sensitive to changes in lightness (Brewer 1992, Slocum et al 2009, Meirelles 2013), so by including such like-colors on either end of the scale bar of ElevCR, confusion could be induced by the map reader. This did not however seem to be the case in the results from this study. Figures 3A and 3B illustrate the overt visual confusion by one participant in Study 2. The repeated attention paid to both the high and low end of the scale bar in ElevCR indicates the presence of confusion by the map viewer. The saccadic activity indicated by an extensive Total Gaze Plot Length in Figure 3A persisted beyond the filtering and clustering stage of the eye tracking data analysis. This resulted in a fixation or cluster on the low end of the scale bar in Figure 3B, showing a significant amount of viewing time by the map reader on the opposite end of the scale bar from where the actual data value could be estimated. Such unique notions made discoverable by eye tracking data are worth further pursuing as such results could help guide more effective color combinations for symbolizing continuous value map data.

WindCR, which is often used in wind speed and wind direction maps, was expected in this study to perform the best at communicating continuous value map data because it was single-hued and varied only in lightness. Maps symbolized using WindCR tended to induce significant underestimations of true map values by participants across both studies. In fact, WindCR performance, as measured by participant ability to estimate map values represented by

this color scheme, was significantly worse from every other color ramp assessed in this study. Further, participants estimating values in WindCR tended to expend significantly more effort in estimating map values as measured by Total Gaze Plot Length and Filtered Gaze Plot Length. Finally, WindCR induced significantly longer Times to Estimate Map Data values than did other color ramps. A color ramp characterized by a single hue and dramatic lightness variation is not the best for representing continuous value map data.

It must be considered that each of the four pervasive color ramps assessed in this study varied greatly in terms of the visual variables that help to define them (Bertin 2010). Both ElevCR and WindCR represent two extremes in terms of commonly-used color ramps and the perceptual dimensions of color. PrecipCR and TempCR, however, both share similarities in their composition of hue, lightness, and saturation. PrecipCR and TempCR performed similarly in that participant estimations varied yet were approximately accurate, while participant-color map interaction variables also did not vary as wildly as those of WindCR. PrecipCR and TempCR required the least time for participants to most accurately estimate data values. These results suggest that color ramps similar in perceptual dimensions of color composition to PrecipCR and TempCR are better for symbolizing continuous value maps than color ramps characterized similarly to ElevCR or WindCR.

Furthermore, another notable characteristic of all the multi-hue color ramps used in this study is the lightness variation within each hue. Each color ramp can be assessed for the hue changes in their legends. Lightness transitions differently through and between hues in the four color ramps assessed in this study. Perhaps the relatively narrower range of data values constrained by each hue of the ElevCR legend aided participants in their estimation tasks. It would be logical then that the color ramps characterized by less hues in their data range would

provide map readers with a less confined data range from which map values could be estimation. This wider range of data not confined by relative map colors in the legend could thus result in lower map estimation accuracy measurements and more confusion by map readers, as measured by map reader time to estimate and gaze path during the estimation task. Thus, the combined effect of visual variable complexity within continuous value maps is what defines the map readers experience and ability to estimate map values. This study aimed to evaluate common color ramps used in Earth system science disciplines. Further research would be required to further evaluate each visual variable's impact on map reader accuracy and interaction with continuous value maps. Taken together, the complexity of each color ramp varied greatly and likely influenced participant estimation and interaction with the continuous value maps during the data estimation tasks.

Figure 5 depicts a subset of participant eye tracking data to illustrate an example of typical map reader interaction variations between color ramps. This participant's Total Gaze Plot Lengths while estimating at Control Point A (known value = 4) during the map estimation task is plotted with the lines depicting participant eye movements during the estimation tasks. Note the gaze pattern variations across each different color ramp. Disregarding the map rotation to disguise the same Control Point being assessed in these four maps, notice the tendency for the gaze of the map reader to return to the legend throughout the estimation tasks. More visits by the map reader's eyes to the legend could suggest confusion or map reader uncertainty during the task. Again, color ramp characteristics would influence participant interactions with and thus estimation of map values during estimation tasks, were captured using eye tracking. The differences in participant Time to Estimate Map Values and the estimation accuracy differences may be explained by color ramp complexity influencing participant interactions and estimations.

This warrants further investigation of color ramps in terms of their composition of hue-lightness combinations and placement along legends in continuous value map visualizations.

Color ramps were chosen for this study based on their ubiquity in scientific visualizations and not for their comparability in terms of their perceptual dimensions of color. Further user studies should evaluate the role that each perceptual dimension plays in participant estimation of continuous value map data. An ideal combination of hues, lightness, and saturation variation could be developed and result in an optimized color ramp for displaying continuous value data so often displayed in scientific visualizations (i.e. elevation, precipitation, wind speed, and temperature). Our data suggests that the color ramp used to symbolize continuous data maps influences not only the way in which people understand map data in a quantifiable manner, but also their understanding of map data as measured by their estimation accuracy.

Our prediction that color ramp would influence participant estimation of continuous map data was supported with our results. However, our hypothesis that maps symbolized using single-hue color ramps would perform better than multi-hue color ramps was refuted, which is in disagreement with the most recent relevant literature (Tufte 1997, Borland and Taylor 2007, Light and Bartlein 2004, Rogowitz and Treinish 1996, Rogowitz and Treinish 1998, Ware 1988, Ware 2004, Tufte 1997, Pizer and Zimmerman 1983, Rheingans 1992, Healey 1996, Brewer 1999). Assumptions that single-hue color ramps are most effective at communication information need to be revised and supported using empirical evidence. Further, the atrophy of the multi-hue, spectral, or “rainbow” color ramp needs to be revisited as the best performing color ramps assessed in this study were multi-hue and outshined single-hue color ramps in terms of efficacy and ease at conveying information.

Participant visual interaction also proved to be an exciting variable to evaluate as our hypothesis that more effort, measured by time, would be required for map readers to understand multi-hue maps than single-hue maps due to the complexity of the color ramp. We hypothesized that the number of hues and variations in lightness, or our version of color ramp complexity, would require higher levels of cognition by the map reader for understanding and thus accurate estimations (Lee et al 2013). It turned out that map readers exerted more effort in trying to decipher continuous value map data symbolized using the single-hue color ramp than those symbolized using any of the multi-hue color ramps. Again, this is in disagreement with the literature as our study supported the opposite: that map readers had an easier time estimating accurately using multi-hue symbology rather than single-hue.

Taken together, both parts of this study provided evidence supporting that map viewer effort aligns with differences in map value estimation as a function of color ramp used to symbolize continuous map data. Color ramp complexity, can be defined using any combination of visual variables to describe a color scheme. Since the focus of this study aimed to evaluate the performance of commonly used color ramps within Earth system science visualizations, a large number of color ramps were not evaluated. Future studies should include a survey or classification in terms of visual variables that evaluates color schemes based on efficacy and ease at communicating information in continuous value maps rather than discipline convention.

Future studies also should evaluate the role of each visual variable in the function of a color ramp in terms of its efficacy at communicating continuous value data and the effort map readers must expend during estimation tasks. Each color ramp used in this study was characterized in terms of visual variables that make each color ramp unique. If an array of color ramp characteristics is evaluated for accuracy and ease of use, then an ideal color scheme or set

of color schemes could be developed for symbolizing continuous data most effectively with least effort.

The Continuous Data Model (CDM) used in this study was developed by manipulating a digital elevation model (DEM) into a single dataset. This base dataset was manipulated or rotated to disguise the fact that it was the same dataset and symbolized using four pervasive color ramps, resulting in 16 images from which participants estimated map values. Participants ultimately estimated values at four known map locations on four renditions of the same map. No participant commented in either part of this study that they recognized the dataset in each image as being the same and it was evident from the data that map readers were not recording the same map values for the same control points on different maps. It was a concern that participants could possibly have recognized the datasets as being rotated and re-symbolized copies, yet no learning effect was evident. In order to wash out the possible influence of participants gaming the estimation tasks, the order of the images presented to the participants was shuffled (Figure 2) and resulting participant estimations were averaged across each image and for each color ramp. A model raster or surface dataset could be generated and used for future studies as then the contrast and variations within the data range could better be controlled. Future studies should more rigorously evaluate a large sample of participant estimations and interactions with a multitude of positions across a similar high contrast CDM. Additionally, a larger eye tracking sample size could provide results for a more generalizable statement on how participants interact with maps as a function of visual variable complexity within each color ramp.

In general, Earth system science discipline visualizations are consumed by different audiences for a great variety of reasons. Continuous value maps may be used to communicate hard data about climatic variations, from which a map reading audience could be expected to

apply past knowledge and make decisions regarding the fate of the planet. On the contrary, continuous value data color may simply be used to enhance the esthetics of a map intended for the enjoyment of a lay audience. Map communication has a wide range of functions across a large map readership. It is increasingly important that maps communicate effectively and in a timely manner.

Based on the results from this study, we strongly recommend abandoning discipline convention (Edney 2005) in symbolizing continuous value map data. We recommend the adherence to data representation guidelines that have been founded on empirical evidence, such as Brewer (1999), in order to more effectively convey continuous value map data. Moreover, continuous value map design instruction should be rigorously aligned with these guidelines to foster the development of map makers who can accurately portray continuous value data for a variety of audiences (US GCRP 2009, OAS 1991, Collins et al 2013).

Maps and other visualizations play a leading role in the understanding of dynamic Earth processes within and beyond the Earth system sciences. The color scheme we use to display such often intricate information has a significant impact on map reader understanding of continuous value data. The implications of the more serious consideration for continuous map color symbology would result in map readers interacting with map symbology designed to promote understanding, which would result in more informative continuous value maps from which a more informed community capable of discerning information would be derived. Educators, decision-makers, policy-makers, citizens, scientists, and most anyone else who engages with visual media would benefit from defined effective color schemes used to portray continuous value data.

APPENDICES

Appendix 1: Tables

	Study 1 (MT)		Study 2 (ET)	
Avg Estimation Differences	Skewness	Kurtosis	Skewness	Kurtosis
ElevCR	3.17	16.896	-0.986	0.562
WindCR	0.988	2.024	-1.564	5.249
PrecipCR	0.743	3.659	-0.92	2.533
TempCR	1.025	3.167	0.15	-0.758

Table 1. Participant estimation distribution data for Study 1 (Amazon Mechanical Turk, MT) and 2 (Eye Tracking, ET).

Avg Estimation Differences	Study 1 (MT)	SD	Study 2 (ET)	SD
ElevCR	-1.0	83.0	1.5	149.4
WindCR	-17.5	89.1	-25.0	88.3
PrecipCR	-3.0	58.7	-4.0	38.5
TempCR	5.0	30.4	2.0	24.5

Table 2. Median participant estimation deviation from known map values in Study 1 and Study 2. Negative data point labels represent extreme underestimations of map values while positive data point labels represent extreme overestimations.







	Avg Estimation Differences	Study 1 (MT)	<i>p</i>	Study 2 (ET)	<i>p</i>
	ElevCR vs WindCR	-9.234	< .001	-4.029	< .001
	ElevCR vs PrecipCR	-0.610	0.542	-0.292	0.770
	ElevCR vs TempCR	-6.483	< .001	-0.852	0.394
	WindCR vs PrecipCR	-8.787	< .001	-3.272	0.001
	WindCR vs TempCR	-10.084	< .001	-3.972	< .001
	PrecipCR vs TempCR	-5.729	< .001	-0.350	0.726

Table 3. Wilcoxon Signed-Rank Test results for median participant estimation deviation data for Study 1 and Study 2.

Kendall's W						
Avg Estimation Differences	Study 1 (MT)	Chi-Sqaure	<i>p</i>	Study 2 (ET)	Chi-Sqaure	<i>p</i>
ElevCR, WindCR, PrecipCR, TempCR	0.220	187.799	< .001	0.441	31.731	< .001

Table 4. Kendall's Coefficient of Concordance (W) for median participant estimation deviation data for Study 1 and Study 2.

	Study 2 (ET)	
Avg TGPL	Skewness	Kurtosis
ElevCR	0.2	-1.097
WindCR	1.045	1.459
PrecipCR	0.039	-0.927
TempCR	0.487	0.033

Table 5. Participant Total Gaze Path Length distribution data for Study 2.

Avg TGPL	Study 2 (ET)	SD
ElevCR	13729.43	8602.52
WindCR	17184.52	9672.45
PrecipCR	14725.15	7934.74
TempCR	15508.88	10481.28

Table 6. Participant median Total Gaze Path Length for Study 2.







	Avg TGPL	Study 2 (ET)	<i>p</i>
	ElevCR vs WindCR	-3.314	< 0.01
	ElevCR vs PrecipCR	-1.314	0.189
	ElevCR vs TempCR	-1.000	0.317
	WindCR vs PrecipCR	-2.829	< 0.01
	WindCR vs TempCR	-2.457	< 0.05
	PrecipCR vs TempCR	-0.343	0.732

Table 7. Wilcoxon Signed-Rank Test assessing Total Gaze Path Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

Kendall's W			
Avg TGPL	Study 2 (ET)	Chi-Sqaure	<i>p</i>
ElevCR, WindCR, PrecipCR, TempCR	0.238	17.15	< .001

Table 8. Kendall's coefficient of concordance (W) assessing Total Gaze Path Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

	Study 2 (ET)	
Avg GPL	Skewness	Kurtosis
ElevCR	1.86	4.528
WindCR	1.089	0.844
PrecipCR	1.039	0.906
TempCR	2.408	6.273

Table 9. Participant median filtered Gaze Plot Length distribution data for Study 2.

Avg GPL	Study 2 (ET)	SD
ElevCR	3566.5	2733.92
WindCR	5242.38	3228.8
PrecipCR	3990.78	2631.69
TempCR	3886.74	3877.59

Table 10. Participant median filtered Gaze Plot Length for Study 2.







	Avg GPL	Study 2 (ET)	<i>p</i>
	ElevCR vs WindCR	-3.943	< .001
	ElevCR vs PrecipCR	-1.314	0.189
	ElevCR vs TempCR	-1.114	0.265
	WindCR vs PrecipCR	-3.743	< .001
	WindCR vs TempCR	-3.343	< .01
	PrecipCR vs TempCR	-0.057	0.954

Table 11. Wilcoxon Signed-Rank Tests assessing Filtered Gaze Plot Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

Kendall's W			
Avg GPL	Study 2 (ET)	Chi-Square	<i>p</i>
ElevCR, WindCR, PrecipCR, TempCR	0.275	19.8	< .001

Table 12. Kendall's coefficient of concordance (W) assessing Filtered Gaze Plot Length mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

	Study 2 (ET)	
Avg Time (ms)	Skewness	Kurtosis
ElevCR	0.756	0.464
WindCR	0.841	0.339
PrecipCR	0.822	0.385
TempCR	2.44	8.481

Table 13. Participant median Time to Estimate Map Values distribution data for Study 2.

Avg Time (ms)	Study 2 (ET)	SD
ElevCR	6595	3983.83
WindCR	7064	3936.6
PrecipCR	5422	3065.98
TempCR	6037.5	5104.25

Table 14. Participant median Time to Estimate Map Values for Study 2.







	Avg Time (ms)	Study 2 (ET)	<i>p</i>
	ElevCR vs WindCR	-2.629	< .01
	ElevCR vs PrecipCR	-2.314	< .05
	ElevCR vs TempCR	-2.629	< .01
	WindCR vs PrecipCR	-4.229	< .001
	WindCR vs TempCR	-3.457	< .01
	PrecipCR vs TempCR	-1.029	0.304

Table 15. Wilcoxon Signed-Rank Test assessing median participant Time to Estimate Map Values mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

Kendall's W			
Avg Time (ms)	Study 2 (ET)	Chi-Sqaure	<i>p</i>
ElevCR, WindCR, PrecipCR, TempCR	0.398	28.65	< .001

Table 16. Kendall's coefficient of concordance (W) assessing median participant Time to Estimate Map Values mean rank differences across ElevCR, WindCR, PrecipCR, and TempCR during Study 2 (Eye Tracking Study).

Appendix 2: Figures

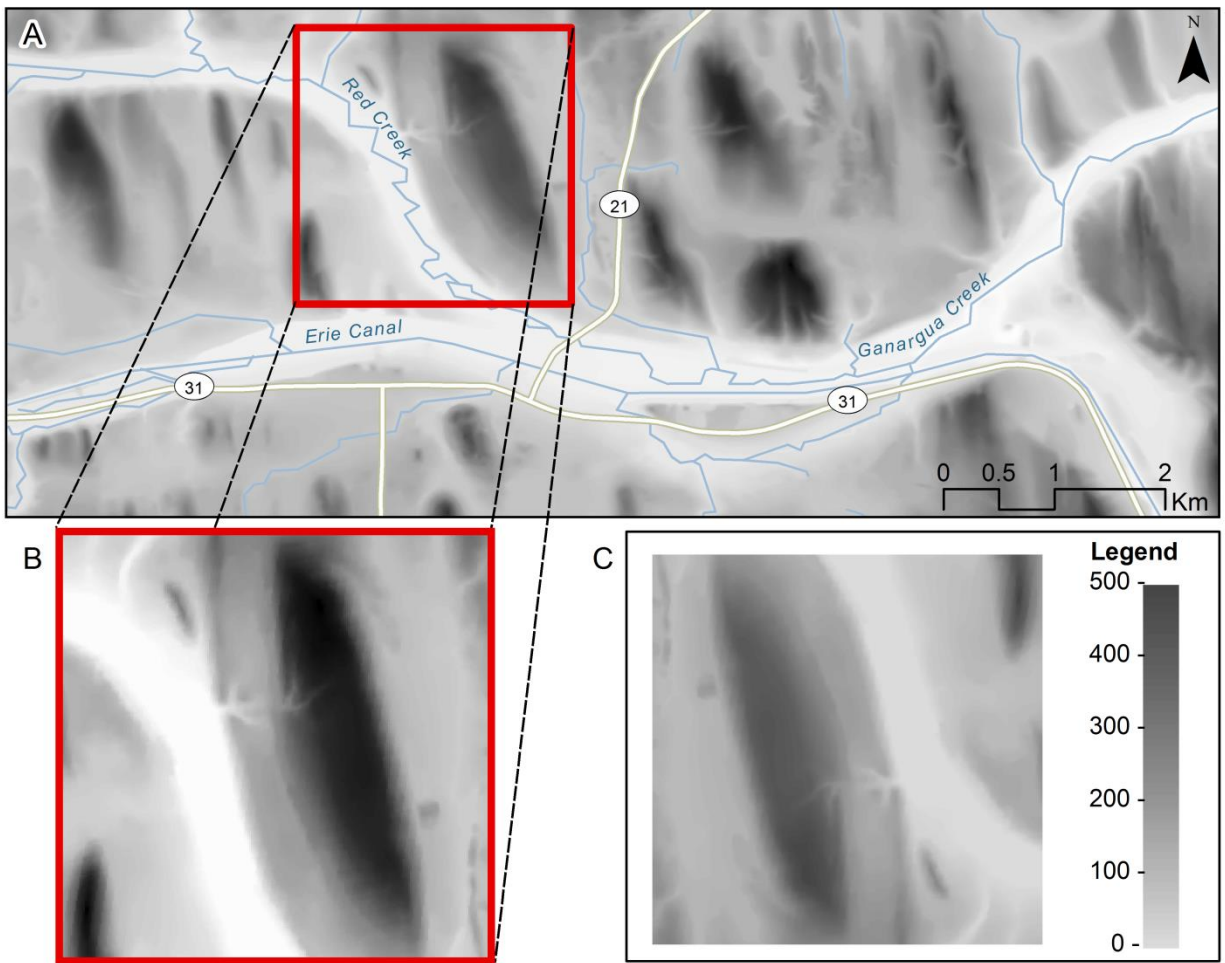


Figure 1. The process of developing a continuous data model (CDM): (A) original high resolution digital elevation model of drumlin field in Upstate New York with hydrology and transportation layers for reference; (B) a simple and non-visually complex sub dataset to be used as basis for CDM; and (C) CDM: a unitless and simple continuous dataset with scaled data values between 0 and 500. Note the rotational transformation from original dataset.

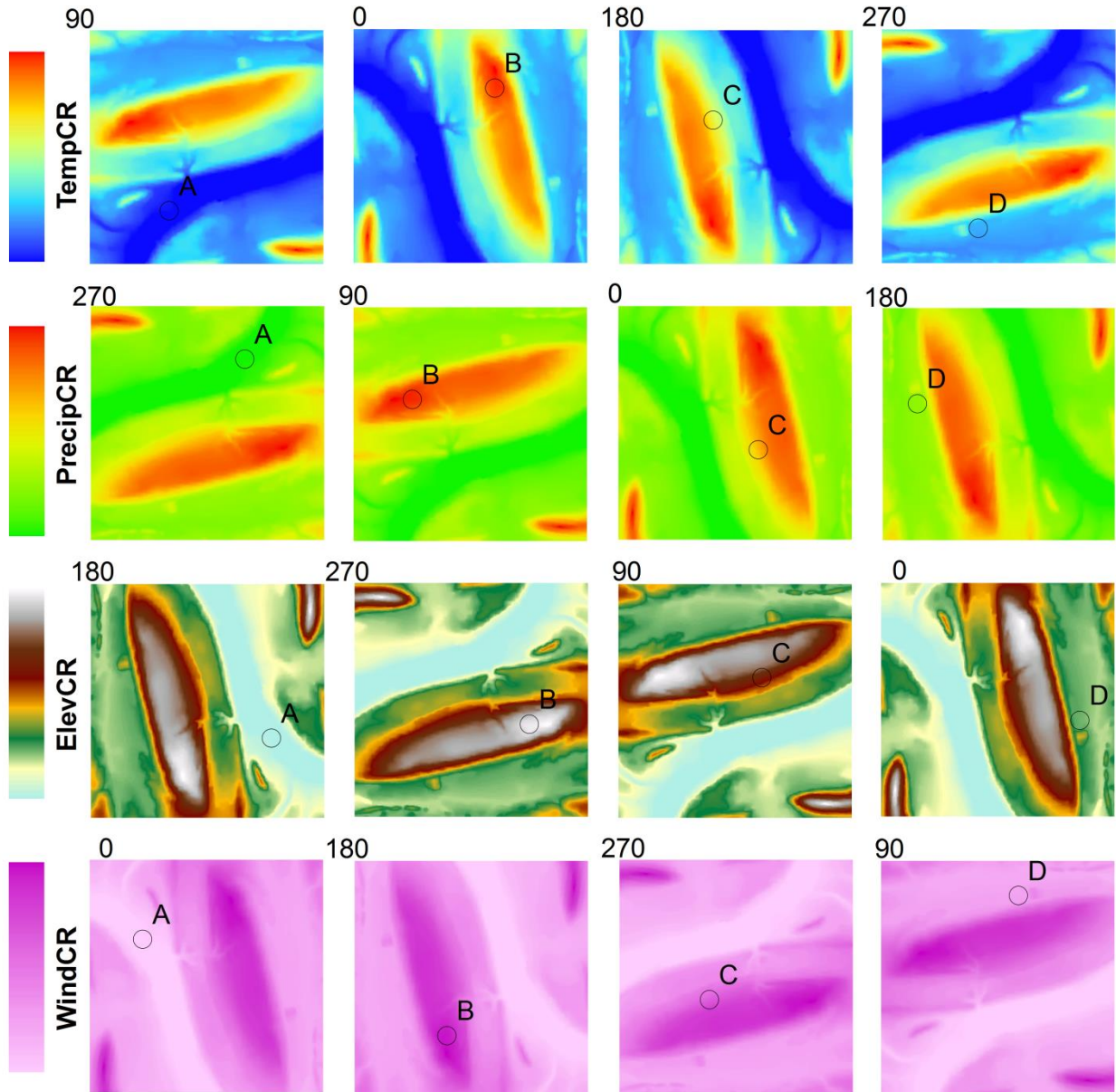


Figure 2. Continuous Data Model (CDM) variations symbolized using four ubiquitous color ramps: TempCR, PrecipCR, ElevCR, and WindCR. Four control points of known value: A, B, C, and D, are represented once per color ramp variation. Rotational transformations (in degrees) applied to the CDM are indicated to the upper left of each variation. These sixteen color images were displayed sequentially and in random order to participants during the estimation task. Control point labels were not displayed to test participants but are displayed here for demonstration.

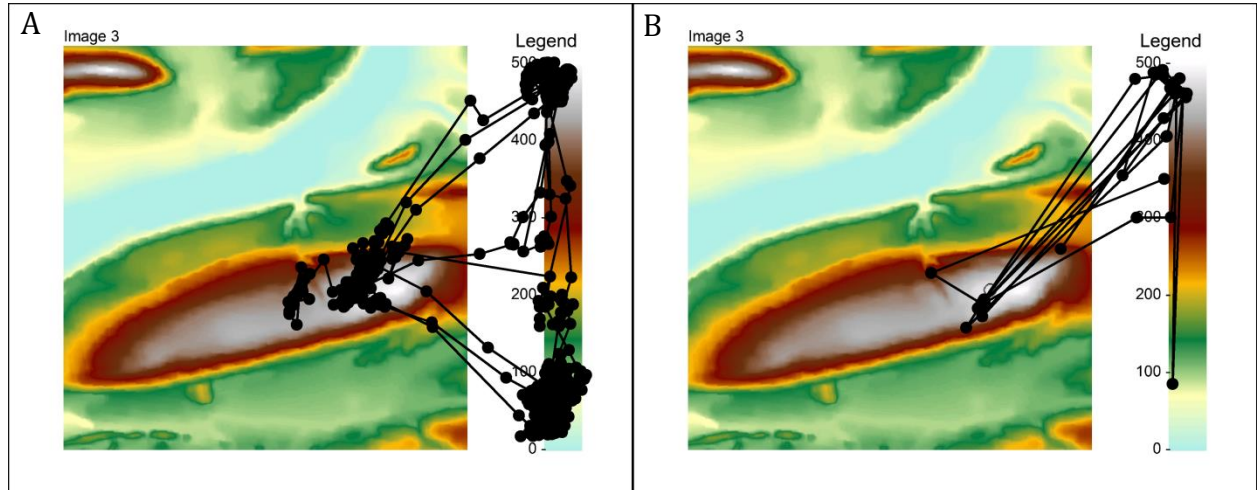


Figure 3. Example of participant-color ramp interaction variables: Total Gaze Plot (A) Length: 26,362.9 pixels and Filtered Gaze Plot (B) Length: 8,855.5 pixels.

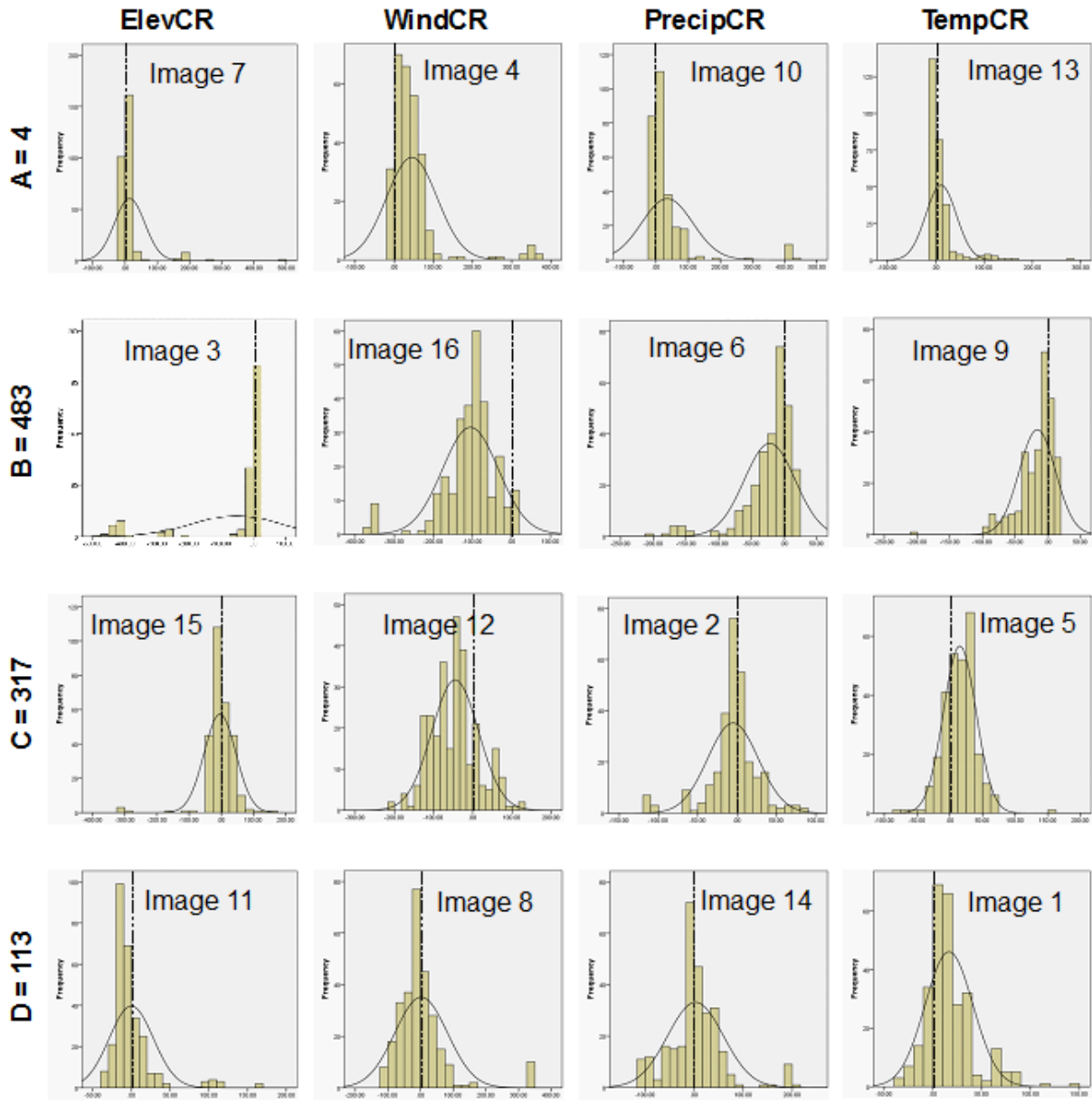


Figure 4. Median participant estimation deviation from known map values in Study 1. The zero point of each histogram is emphasized with vertical dashed lines. Known map values at which participants estimated map data is on y-axis. Negative data point labels represent extreme underestimations of map values while positive data point labels represent extreme overestimations.

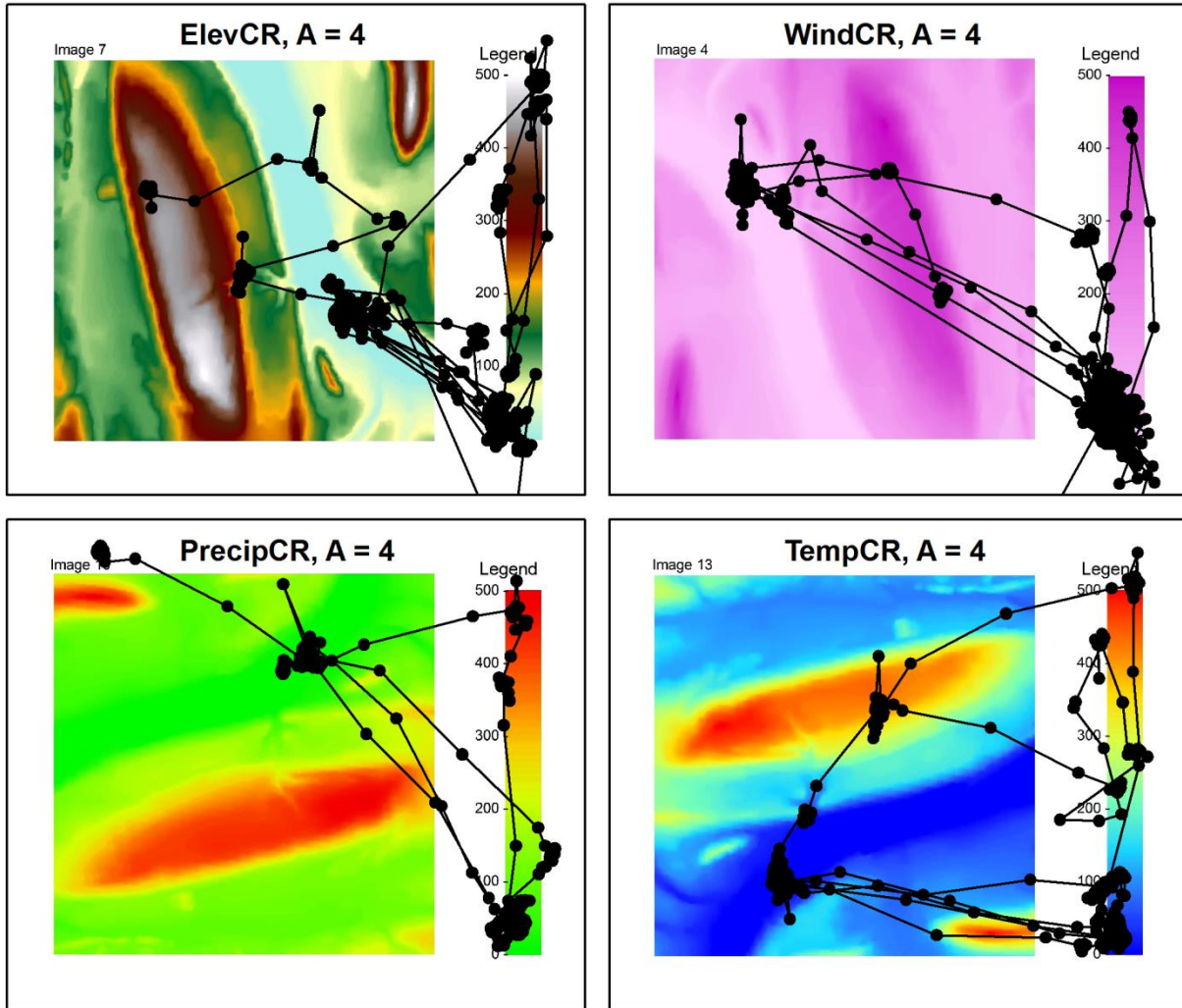


Figure 5. A subset of participant Total Gaze Plot Lengths on ElevCR, WindCR, PrecipCR, and TempCR at Control Point A (known value = 4). Note the gaze pattern variations across each different color ramp. The differences in participant Time to Estimate Map Values and the accuracy differences may be explained by color ramp complexity influencing participant interactions and estimations.

REFERENCES

REFERENCES

- Andersson, R., Nystrom, M., and Holmqvist, K. (2010) Sampling Frequency and Eye-Tracking Measures: How Speed Affects Durations, Latencies, and More. *Journal of Eye Movement Research*.
- Bamford, A. (2003) *The Visual Literacy White Paper*. Adobe Systems Pty Ltd, Australia.
- Behrens, C. (2008) *Design Pattern Repository*. Info Design Patterns.
- Bertin, J. translated by Berg, W. (2010) *Semiology of Graphics: Diagrams, Networks, Maps*. Redlands, CA: Esri Press.
- Biel, J., Aran, O., and Gatica-Perez, D. (2011) You Are Known by How You Vlog: Personality Impressions and Nonverbal Behavior in YouTube. *Association for the Advancement of Artificial Intelligence*
- Borland, D., Taylor II, R.M. (2007) Rainbow color map (still) considered harmful. *IEEE*.
- Branson, S., Wah, C., Schroff, F., Babenko, B., Welinder, P., Perona, P., and Belongie, S. (2010) Visual Recognition with Humans in the Loop. *Computer Vision, ECCV. Lecture Notes in Computer Science Volume 6314*.
- Brewer, C. (1992) *Color Selection for Geographic Data Analysis and Visualization*.
- Brewer, C. A. (1999) *Color Use Guidelines for Data Representation, Proceedings of the Section on Statistical Graphics, American Statistical Association, Alexandria VA*.
- Buckley, A. (2008) *Hypsometric Tinting. ArcGIS Resources. ESRI, Inc., Redlands, California. United States*.
- Buckley, A., Frye, C. (2011) *Web Map Use and Design: Shifting the Cartography Paradigm. ESRI, Inc., Redlands, California. United States*.
- Buhrmester, M., Kwang, T., and Goslong, S. (2007) Amazon's Mechanical Turk: A New Source of Inexpensive yet High-Quality, Data. *Perspectives on Psychological Science*.
- Bulling, A. and Gellersen, H. (2010) *Toward Mobile Eye-Based Human-Computer Interaction. IEEE Pervasive Computing*.
- Bucher, H. and Schumacher, P. (2006) *The Relevance of Attention for Selecting News Content. An Eye-Tracking Study on Attention Patterns in the Reception of Print and Online Media. Communications. Volume 31*.
- Card, S., Mackinlay, K., and Shneiderman, B., ed. (1999) *Readings in Information Visualization: Using Vision to Think*.

- Chen, C. (2008) An information-theoretic view of visual analytics. *IEEE Computer Graphics and Applications*.
- Chen, M., Janicke, H. (2010) An Information-theoretic framework for visualization. *IEEE Transactions of Visualization and Computer Graphics*.
- Chi, E. (2000) A Taxonomy of Visualization Techniques using the Data State Reference Model. *Proceedings of the IEEE Symposium on InfoVis*. IEEE Computer Society Press.
- Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichet, P. Friedlingstein, X. Gao, W.J. Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver and M. Wehner, 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1029–1136, doi:10.1017/CBO9781107415324.024.
- Cowen, L., Ball, L., and Delin, J. (2002) An Eye Movement Analysis of Web Page Usability. *People and Computers*. XVI. Memorable Yet Invisible.
- Crump, M., McDonnell, J., and Gureckis, T. (2013) Evaluating Amazon’s Mechanical Turk as a Tool for Experimental Behavioral Research. *PLoS ONE* 8(3): e57410. doi:10.1371/journal.pone.0057410.
- Dodge, R and Cline, T. (1901). The Angle Velocity of Eye Movements. *Psychological Review*.
- dos Santos, S. and Brodlie, K. (2004) Gaining Understanding of Multivariate and Multidimensional Data through Visualization. *Computers and Graphics*.
- Dykes, W., Slingsby A., and Clarkm K. (2007) Interactive Visual Exploration of a Large spatio-temporal dataset: Reflections on a Geovisualization Mashup. *IEEE Transactions on Visualization and Computer Graphics*.
- Edney, M. (2005) Putting “Cartography” into the History of Cartography: Arthur H. Robinson, David Woodward, and the Creation of a Discipline. *Cartographic Perspectives*. Number 51.
- Fouh, E., Akbar, A., Shaffer, C. (2012) The Role of Visualization in Computer Science Education. Virginia Tech, Blacksburg, Virginia, USA.
- Fu, P. and Sun, J. (2010) *Web GIS: Principles and Applications*. California: Esri Press. Wingert, E.A. 1997. John Clinton Sherman: Academic Cartographer on the Brink of a New Age, *Cartographic Perspectives*.
- Haber, R. and McNabb, D. (1990) *Visualization Idioms: A Conceptual Model for Scientific Visualization Systems*. Visualization in Scientific Computing. IEEE Computer Society Press.

- Haro, A., Essa, I., and Flickner, M. (2000) A Non-Invasive Computer Vision System for Reliable Eye Tracking. Proceedings of Human Factors in Computing Systems.
- Harrower, M., and Brewer, C. (2003) ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal. Volume 40, No. 1.
- Healey, C.G. (1996) Choosing Effective Colors for Data Visualization, Proc. IEEE Visualization, IEEE CS Press, 1996.
- Heer, J., Bostock, M., and Ogievetsky, V. (2010) A Tour through the Visualization Zoo. Communications of the ACM.
- Heer, J. and Bostock, M. (2010) Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI).
- Huang, L. and Pashler, H. (2007) A Boolean Map Theory of Visual Attention. Psychological Review.
- Hwang, S. and Grauman, K. (2010) Reading Between the Lines: Object Localization using Implicit Cues from Image Tags. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Iacono, W. and Lykken, D. (2007) Electro-oculographic Recording and Scoring of Smooth Pursuit and Saccadic Eye Tracking: A Parametric Study using Monozygotic Twins. Psychophysiology. Volume 16.
- Ishihara S. (1954, 1958) Tests for Color Blindness. 11th and 13th Eds. Tokyo, Japan; Shuppan.
- Jacob, R and Karn, K. (2003) Commentary on Section 4. Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises.
- Lee, S., Sips, M., and Seidel, H. (2013) Perceptually Driven Visibility for Categorical Data Visualization. Visualization and Computer Graphics. IEEE Transactions. Volume 19.
- Light, A. and Bartlein, P.J. (2004) The End of the Rainbow? Color Schemes for Improved Data Graphics, EOS Trans. American Geophysical Union, vol. 85, no. 40.
- Lima, M. (2011) Visual Complexity: Mapping Patterns of Information.
- Lin, S., Fortuna, J., Kulkarni, C., Stone, M.m and Heer, J. (2013) Selecting Semantically-Resonant Colors for Data Visualization. Computer Graphics Forum. Volume 32.
- Lloyd, R. and Bunch, R. (2014) Explaining Map-reading Performance Efficiency: Gender, Memory, and Geographic Information. Cartography and Geographic Information Science.

- Mason, W. and Suri, S. (2010) Conducting Behavioral Research on Amazon's Mechanical Turk. Social Science Research Network.
- McMaster, R.B., Thrower, N.J.W. (1991) The Early Years of American Academic Cartography: 1920-45. *Cartography and Geographic Information Systems* 18(3), 151-5.
- Meirelles, I. (2013) *Design for Information: An Introduction to the Histories, Theories, and Best Practices behind Effective Information Visualization*.
- Moreland, K. (2009) Diverging Color Maps for Scientific Visualization. *Proceedings of the 5th International Symposium on Visual Computing*.
- National Oceanic and Atmospheric Administration's (NOAA) National Weather Service. (2014a) WindSpd (Kts) and WindDir. Real-Time Mesoscale Analysis. Graphical Forecasts. <<http://graphical.weather.gov/sectors/conus.php?element=WindSpd>>
- National Oceanic and Atmospheric Administration's (NOAA) National Weather Service. (2014b) 6Hr Precip.Amt (in). National Digital Forecast Database. Graphical Forecasts. <<http://graphical.weather.gov/sectors/conus.php?element=QPF>>
- OAS (Organization of American States). (1991) *Primer on Natural Hazard Management in Integrated Regional Development Planning*. Organization of American States. Department of Regional Development and Environment.
- Olsson, P. (2007) Real-time and offline filters for eye tracking. Master's thesis, Royal Institute of Technology, Sweden.
- Opach, T. and Nossum, A. (2011) Evaluating the Usability of Cartographic Animations with Eye Movement Analysis. *Proceedings of the 25th International Cartographic Conference*.
- Paolacci, G., Chandler, J., and Ipeirotis, P. (2010) Running Experiments on Amazon Mechanical Turk. *Judgment and Decision Making*.
- Pizer, S.M., Zimmerman, J.B. (1983) Color Display in Ultrasonography, *Ultrasound in Medicine and Biology*, vol. 9, no. 4, 1983.
- Pontin, J. (2007) Artificial Intelligence: With Help from the Humans. *The New York Times*. Accessed 8-12-14.
- Poole, A. and Ball, L. (2006) *Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects*.
- Purchase, H.C., Andrienko, N., Jankun-Kelly, T.J., Ward, M. (2008) Theoretical foundations of information visualization. *Information Visualization: Human-centered Issues and Perspectives*.
- Raschke, M., Blascheck, T., and Burch, M. (2014) Visual Analysis of Eye Tracking Data. *Handbook of Human Centric Visualization*. Springer Science and Business Media.

- Rheingans, P. (1992) Color, Change, and Control for Quantitative Data Display, Proc. IEEE Visualization, IEEE CS Press.
- Robinson, A.H. (1952) The Look of Maps.
- Robinson, A.H. (1991) The development of cartography at the University of Wisconsin-Madison. *Cartography and Geographic Information Systems* 18, 156-7.
- Rogowitz, B.E. and Treinish, L.A. (1996) How Not to Lie with Visualization, *Computers in Physics*, vol. 10, no. 3, 1996, pp. 268-273.
- Rogowitz, B.E. and Treinish, L.A. (1998) Data Visualization: The End of the Rainbow, *IEEE Spectrum*, vol. 35, no. 12, 1998, pp. 52-59.
- Romanach, S., McKelvy, M., Suir, K., and Conzelmann, C. (2014) EverVIEW: A Visualization Platform for Hydrologic and Earth Science Gridded Data. *Computers and Geosciences*. Volume 76.
- Salvucci, D. D., and Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the Eye Tracking Research and Applications Symposium*. New York: ACM Press.
- Schiessl, M., Duda, S., Tholke, A., and Fischer, R. (2003) Eye Tracking and its Application in Usability and Media Research. *MMI-Interactive Journal*.
- Schloss, K. and Palmer, S. (2014) The Politics of Color: Preferences for Republican Red Versus Democratic Blue. *Psychonomic Society, Inc. Psychon Bull Rev*.
- Shneiderman, B. (1996) The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings of IEEE Symposium on Visual Languages*.
- Slocum, T., McMaster, R., Kessler, F., Howard, H. (2009) *Thematic Cartography and Geovisualization*. Third Edition.
- Tominski, C. (1990) *Event-Based Visualization for User-Centered Visual Analytics*. Institute for the Computer Science and Electrical Engineering.
- Tobii Eye Tracking (2010) *An Introduction to Eye Tracking and Tobii Eye Trackers*. Tobii Technology.
- Tobii Eye Tracking (2011) *Accuracy and Precision Test Method for Remote Eye Trackers*. Test Specification White Paper.
- Tufte, E. (1997) *Visual Explanations*, Graphics Press, 1997.
- Tversky, B. (2010) Visualizing Thought. *Topics in Cognitive Science* 3 (2011) 499–535
Copyright 2010 Cognitive Science Society, Inc.
- Usher, M.J. (1984) *Information Theory for Information Technologists*. MacMillan.

- United States Geological Survey (USGS) (2002) The National Map – Elevation. Fact Sheet 106-02. November 2002.
- U.S. Global Change Research Program (USGCRP). Climate Change Science Program (2009) Climate Literacy: The Essential Principles of Climate Science. Second Version.
- van Wijk, J.J. (2005) The Value of Visualization. Proceedings of IEEE Visualization, IEEE Computer Society.
- Volkova, S., Dolan, W., and Wilson, T. (2012) CLex: A Lexicon for Exploring Color, Concept and Emotion Associations in Language. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.
- Wang, J. (2009) Pupil Dilation and Eye-Tracking. Handbook of Process-Tracing Methods. Psychology Press.
- Wang, C., Shen, H. (2011) Information Theory in Scientific Visualization. Entropy.
- Ware, C. (1988) Color Sequences for Univariate Maps: Theory, Experiments, and Principles, IEEE Computer Graphics and Applications, vol. 8, no. 5.
- Ware, C. (2004) Information Visualization: Perception for Design, 2nd ed., Morgan Kaufmann.
- Wiener, N. (1948) Cybernetics. John Wiley and Sons
- Wright, J.K. (1942) Map Makers Are Human: Comments on the Subjective in Maps. American Geographical Society. Geographical Review, Vol. 32, No. 4 (Oct., 1942), pp. 527-544
- Zhu, Z. and Ji, Q. (2005) Robust and Real-Time Eye Detection and Tracking Under Variable Lighting Conditions and Various Face Orientations. Computer Vision and Image Understanding.