The Effect of Color Scales on Climate Scientists' Objective and Subjective Performance in Spatial Data Analysis Tasks

Arita Dasgupta, Jorge Poco, Bernice Rogowitz, Kyungsik Han, Enrico Bertini, Claudio T. Silva.

IEEE Transactions on Visualization and Computer Graphics.
Oct 17 2018. doi: 10.1109/TVCG.2018.2876539. [Epub ahead of print]

Introduction

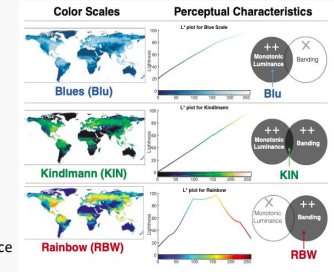
- Geographical maps encoded with rainbow color scales are widely used by climate scientists
  - De facto standard
- Evidence from literature show many shortcomings of rainbow color scale
- This study:
  - Explains potential reasons for the mismatch between theory and practice
  - Compares the effect of various color scales on performance accuracy
  - Investigates how user confidence with the rainbow scale influences performance accuracy and subjective impressions

Background Information

- Researchers focused exclusively on color scales for 2D scalar fields
- Two critical perceptual characteristics of a good color scale:
  - Luminance Monotonicity: increments in luminance should be monotonic with increase in value
  - Banding: the perception of bands of hues
- Researchers used 3 color scales in this study:
  - Blues (Blu)
  - Kindlmann (KIN)
  - Rainbow (RBW)

Color Scales and Their Properties

- Blues:
  - Single hue (blue)
  - Monotonic
  - No banding
- Kindlmann:
  - Monotonic
  - Banding
  - Alternative to RBW
- Rainbow:
  - Non-monotonically varying luminance
  - Banding

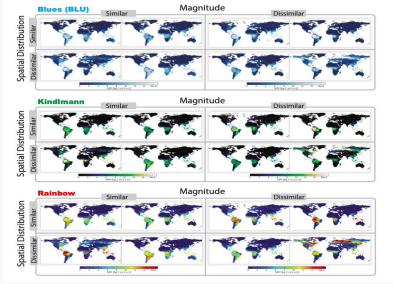


Study Design

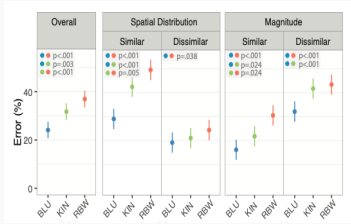
- 39 climate scientist participants
- Each was asked to complete 3 tasks on 4 map pairs using 3 color scales
  - Task 1: judge magnitude similarity in overall Gross Primary Productivity (GPP) between map pairs
  - Task 2: judge GPP spatial distribution similarity between map pairs
  - Task 3: judge maximum GPP spatial distribution difference between map pairs
- Participants also rated familiarity, preference, perceived accuracy, and comfort

Selection of Stimuli

- Map pairs differ in magnitude and spatial distribution
- Grouped based on difference in magnitude and spatial distribution
- Co-varied rows and columns

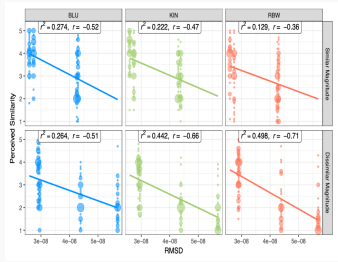


Results (Task 1) – GPP Magnitude Difference



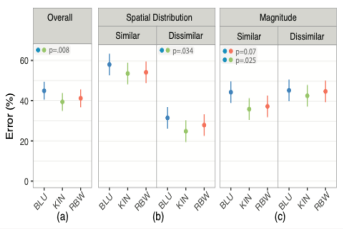
- % error in judging GPP magnitude plotted for each color scale is shown
- Overall effect: BLU < KIN < RBW
- Spatial distribution: BLU < KIN < RBW
- Magnitude: BLU < KIN < RBW
- Summary: monotonic luminance has positive effect, hue banding has negative effect on magnitude comparison

Results (Task 2) – Degree of Similarity



- Perceived similarity (rated on a Likert scale) is shown for each scale
- Y-axis: perceived similarity
- X-axis: computed dissimilarity
- Similar magnitude: BLU > KIN > RBW
- Dissimilar magnitude: RBW > KIN > BLU
- Summary: on average all 3 color scales provided equal benefit

Results (Task 3) – Maximum Difference



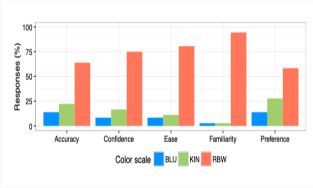
- % error in identifying the most dissimilar region is shown
- Overall: KIN < RBW < BLU
- Spatial distribution: KIN < RBW < BLU
- Magnitude: KIN < RBW < BLU
- Summary: hue banding enables more accurate judgements of differences in spatial distribution

Analyzing Subjective Performance Measures

- Another goal of the study was to compare the perceived accuracy and confidence of the scientists with objective performance:
  - confidence
  - perceived accuracy
  - ease of use of the color scales
  - familiarity
  - preference

Results

- Overall usage confidence:
  - RBW > KIN > BLU
- Overall perceived accuracy:
  - RBW > KIN > BLU
- Objective performance (average % of error):
  - RBW → 37.4%
  - KIN → 31.4%
  - BLU → 24.7%



Analysis Summary

- Who:
  - 39 anonymously-chosen climate scientists
- What:
  - Spatial data
  - Measure effectiveness of different color scales in climate modeling tasks
  - Subjective performance measures
- How:
  - counterbalanced user-study design
  - 3 different tasks of map pairs
  - 3 color scales co-varying in luminance monotonicity and hue banding

Critique

- Strengths:
  - Straightforward and well thought-out design study
  - Good selection of color scales that covaried in luminance monotonicity and hue banding
  - Nicely explores the limitations of rainbow color scale despite its popularity
- Weaknesses/Limitations:
  - Extremely small sample size (n=39, 3 color-blind participants)
  - Use of a *de facto* RBW scale with BLU or KIN scales
  - Confounding variable in RBW/KIN scale: pop-out effect?