Precision in Data Visualization

Common Pitfalls and Best Practices

Introduction

- **Objective:** Understand common color pitfalls in data visualization.
- **Overview:** Explore best practices for using color effectively, considering readability and accessibility.

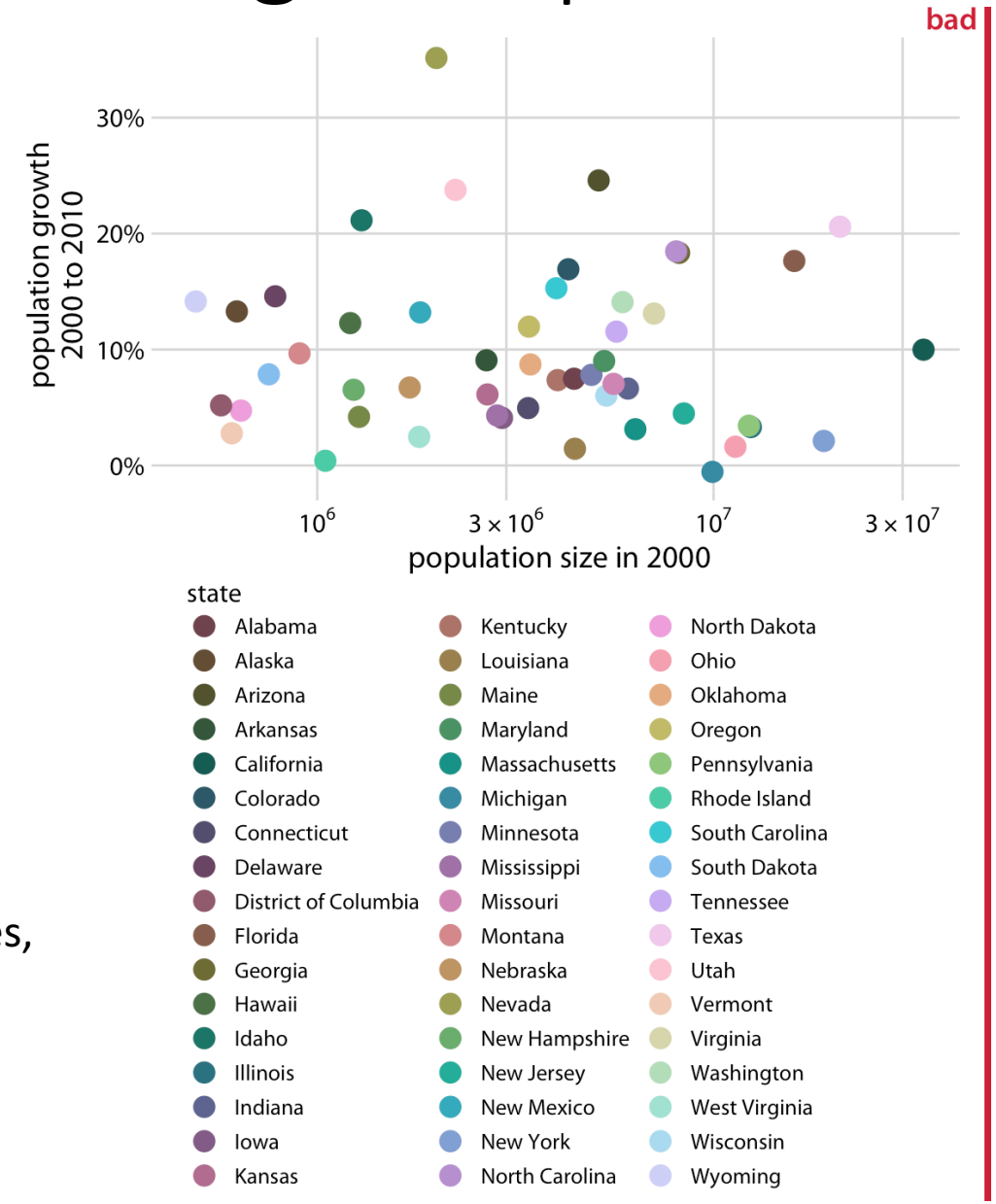
Power and Pitfalls of Color

- **Effective Use of Color:**
  - Color enhances data clarity when used with purpose.
  - Poor choices obscure data insights.
- **Goal:** Make data clearer, not create visual puzzles.

Pitfall 1: Encoding Too Much Information with Color

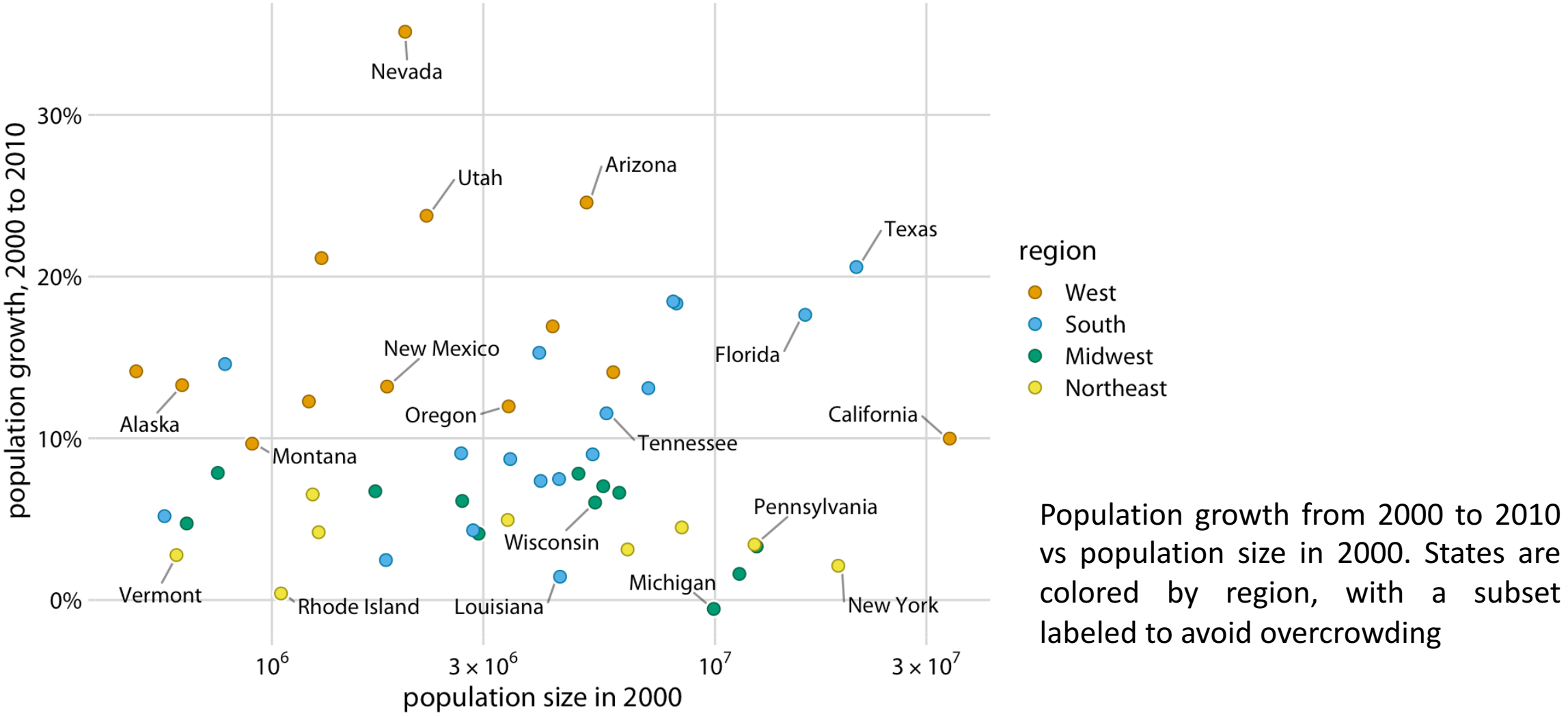
- **Problem:** Using too many colors makes it hard to read and analyze.
- **Example:** U.S. States population growth – too many colors for individual states.
- **Better Approach:**
  - Use color for larger categories (e.g., geographic regions).
  - Use direct labeling for individual states.

Challenging Color Matching in Population Growth Data



Population growth vs. size (2000-2010) for U.S. states, with each state marked in a different color, making color matching challenging.

Population Growth and Size by Region

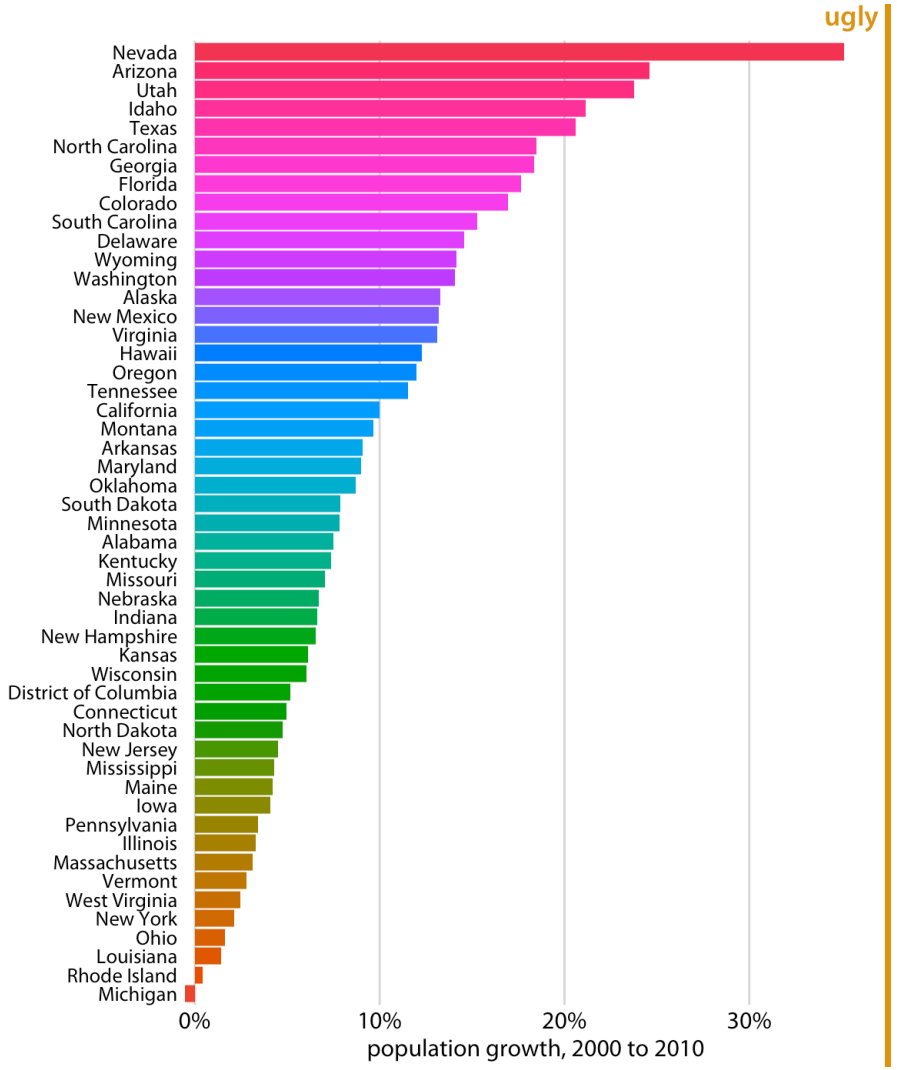


Population growth from 2000 to 2010 vs population size in 2000. States are colored by region, with a subset labeled to avoid overcrowding

Pitfall 2: Coloring for Aesthetic Purposes Only

- **Problem:** Random colors create visual noise without adding insights.
- **Example:** U.S. population growth, rainbow-colored bars.
- **Recommendation:** Use color meaningfully, avoid aesthetic-only choices.

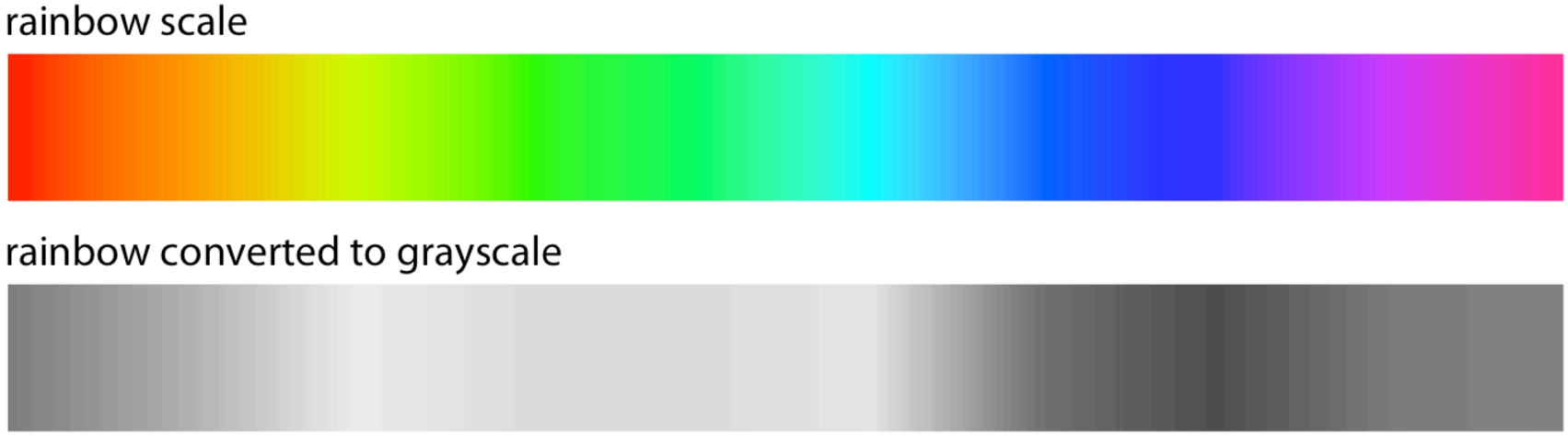
Misleading Rainbow Color Usage



Population growth in the U.S. from 2000 to 2010. The rainbow coloring of states serves no purpose and is distracting. Furthermore, the colors are overly saturated.

Pitfall 3: Non-Monotonic Color Scales

- **Problem:** Non-monotonic color scales confuse data interpretation.
- **Example:** Rainbow color scale, where color changes are inconsistent.
- **Solution:** Use sequential color scales for continuous data.

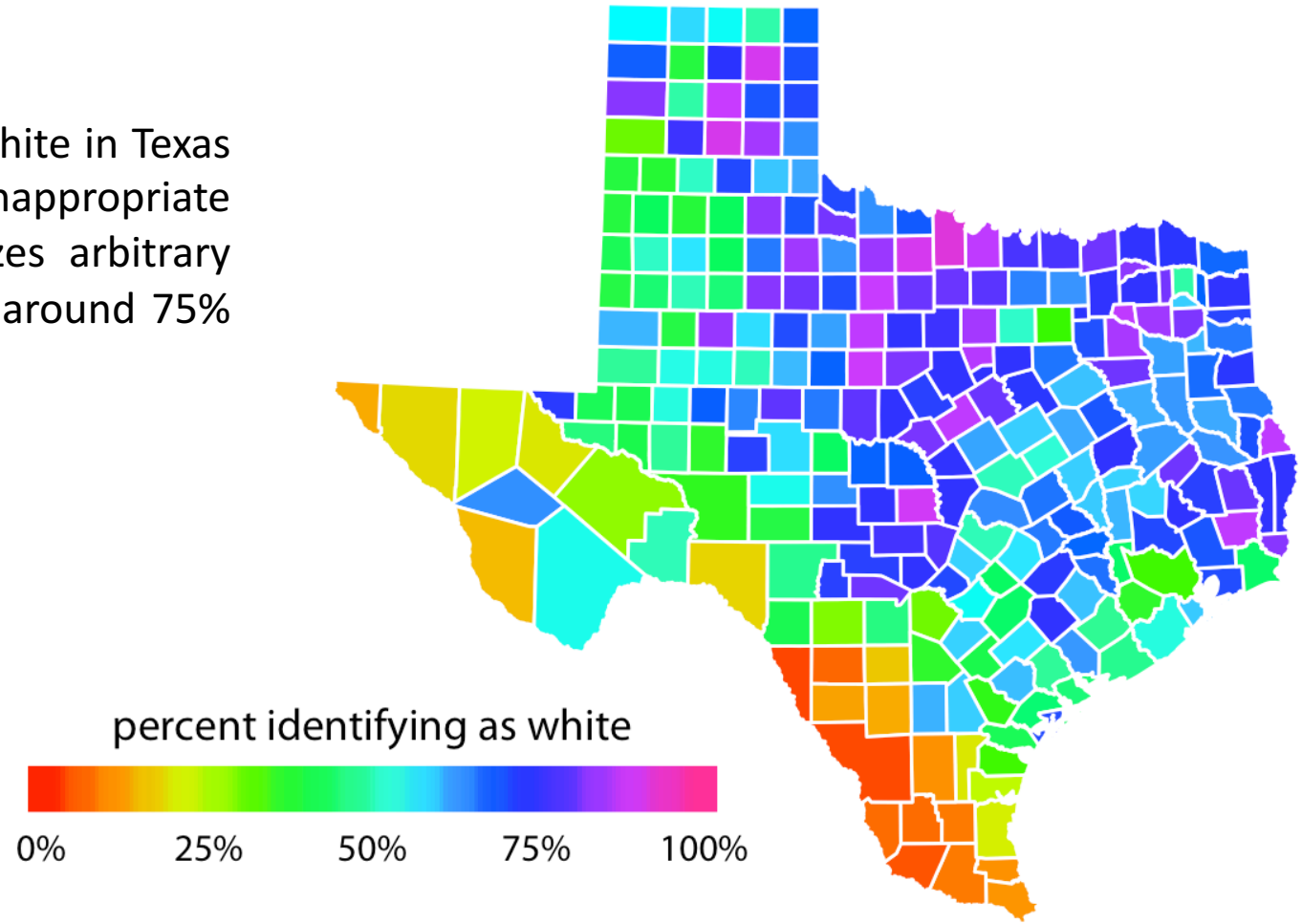


The rainbow colorscale is non-monotonic, with lightness changing non-uniformly. Converting to gray values shows the scale going from dark to light, back to dark, with the lightest section (yellow, light green, cyan) occupying nearly a third, while the darkest part (dark blue) is concentrated in a narrow region.

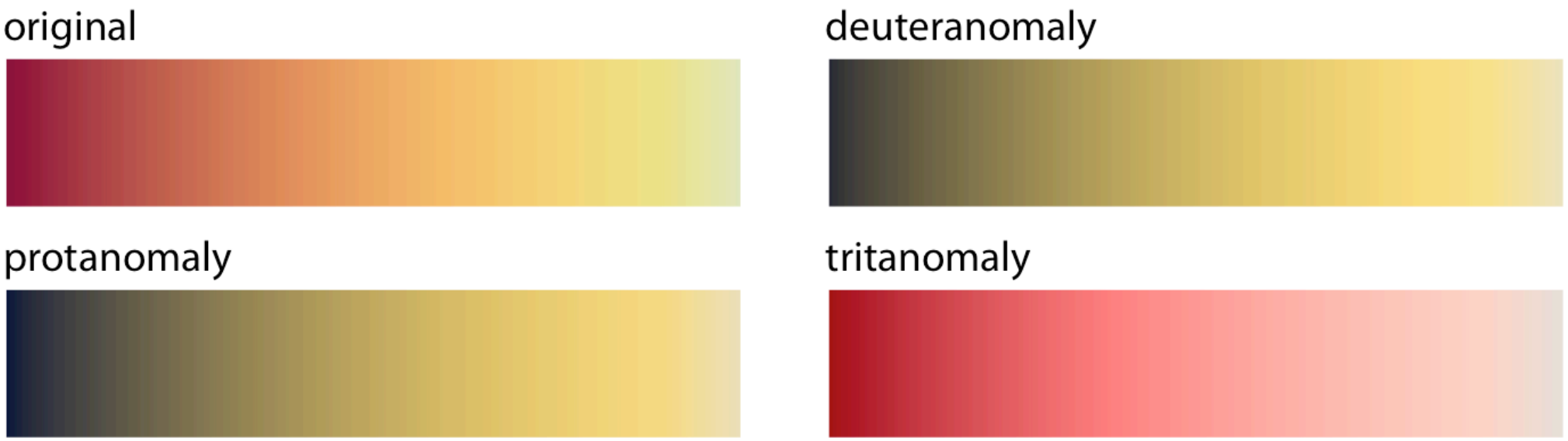
White Population Percentage in Texas Counties

bad

Percentage of people identifying as white in Texas counties. The rainbow color scale is inappropriate for continuous data, as it emphasizes arbitrary features, particularly counties where around 75% of the population identifies as white.



CVD Simulation of the Heat Color Scale

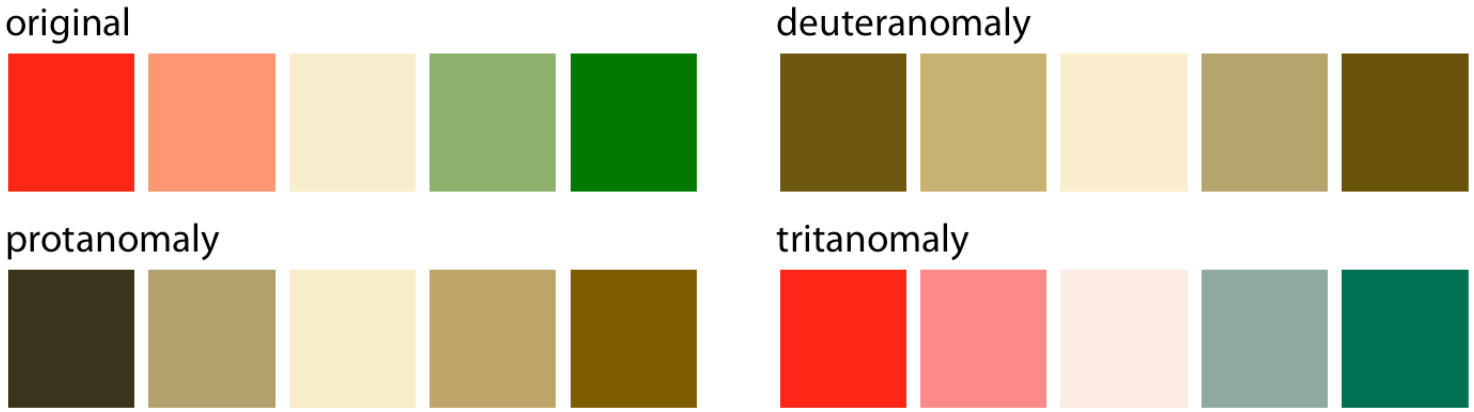


CVD simulation of the Heat color scale, running from dark red to light yellow. From left to right and top to bottom, we see the original scale and its appearance under deuteranomaly, protanomaly, and tritanomaly simulations. Despite color differences, a clear gradient from dark to light is visible in each case, making this color scale safe for CVD.

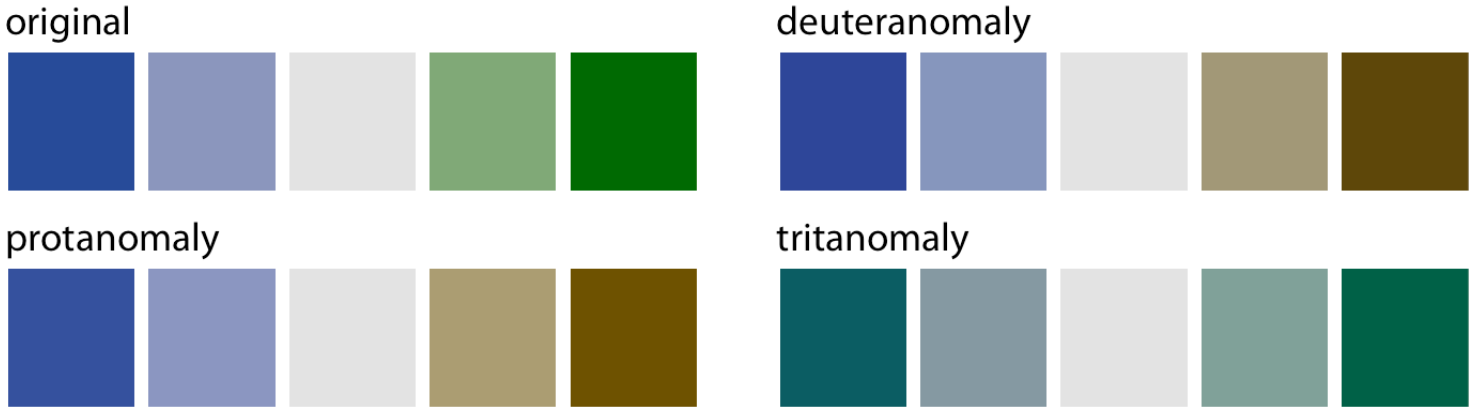
Pitfall 4: Not Designing for Color-Vision Deficiency

- **Problem:** Color choices may not be distinguishable for colorblind individuals.
- **Example:** Red-green and blue-green contrasts.
- **Solution:** Use color scales designed for color-vision deficiency.

CVD Issues with Diverging Scales and Color Contrasts

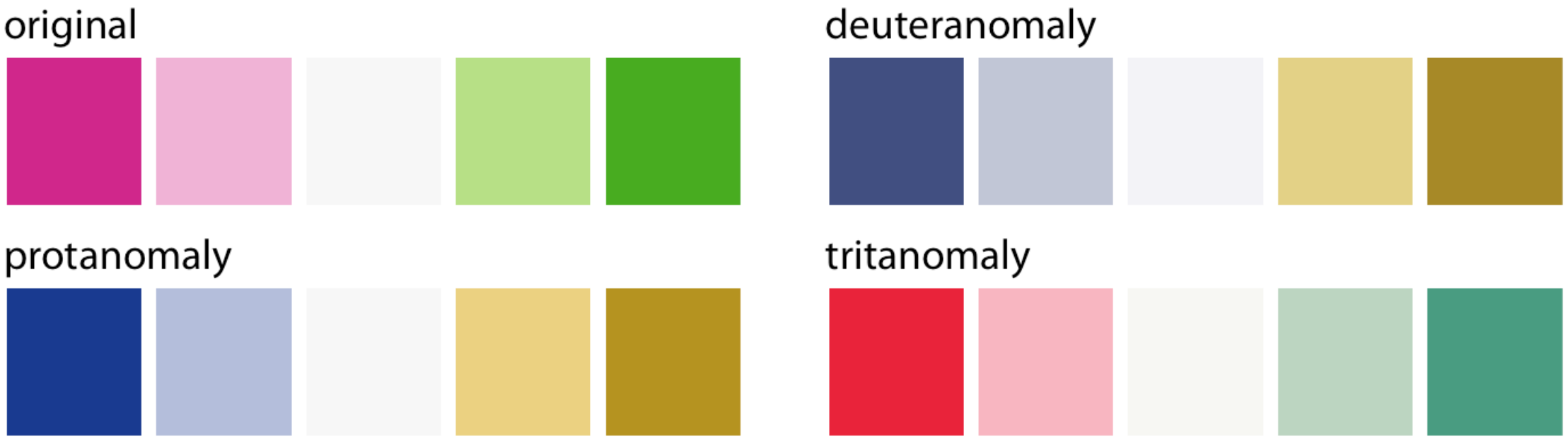


A red–green contrast becomes indistinguishable under red–green cvd (deuteranomaly or protanomaly).



A blue–green contrast becomes indistinguishable under blue–yellow cvd (tritanomaly).

ColorBrewer PiYG: CVD-Friendly Contrast



The ColorBrewer PiYG scale (pink to yellow-green) appears as red-green to regular vision but remains distinguishable for all types of color-vision deficiency. This works because the pink hue contains blue, and the greenish hue contains yellow, ensuring contrast across different forms of cvd.

Practical Guidelines for Color Use

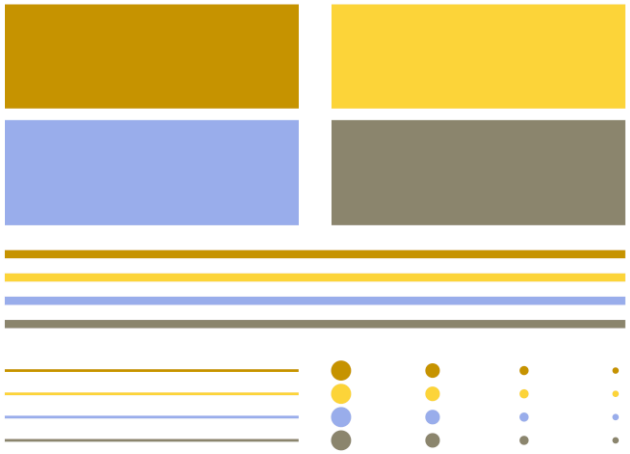
- **Guidelines:**
  - Limit the number of colors to avoid confusion.
  - Use direct labels instead of too many colors.
  - Ensure color choices are distinguishable for colorblind users.
  - Test with color-vision deficiency simulations.

Challenges with Small Colored Elements

original

deuteranomaly

protanomaly



tritanomaly

Colored elements become harder to distinguish at smaller sizes. The top left panel (labeled “original”) shows four rectangles, thick lines, thin lines, and groups of points, all in the same four colors. As the elements shrink, distinguishing the colors becomes more difficult, and this issue is worsened in cvd simulations.

Conclusion

- **Key Takeaways:**
  - Color should clarify, not confuse.
  - Simplicity and accessibility are essential for effective visualizations.
  - Always consider the potential impact on users with color-vision deficiencies.