

Promoting Accessibility in Duke Data Visualization

Creating a Color-Blind Friendly Duke Color Package

Aidan Gildea

3/27/23

A detailed explanation of the creation and use of the duke package.

Table of contents

Preface	4
1 Introduction	5
1.1 Personal Motivation	5
1.2 Background	5
1.3 Thesis Goal	6
2 Methods	7
2.1 Color Sources	7
2.2 Color Metrics	9
2.3 Package Development	10
2.3.1 theme_duke()	10
2.3.2 scale_duke_continuous()	11
2.3.3 scale_duke_color_discrete and scale_duke_fill_discrete	11
2.4 Lessons Learned	14
2.5 References	14
3 Function Descriptions	15
3.1 theme_duke()	15
3.2 scale_duke_continuous()	15
3.3 scale_duke_fill_discrete()	15
3.4 scale_duke_color_discrete()	15
4 Discussion	16
4.1 Colorblind Accessibility Best Practices	16
5 About	17
References	18

Preface

This is the primary documentation for the Duke Statistical Science Thesis, “**Promoting Accessibility in Duke Data Visualization: Creating a Color-Blind Friendly Duke Color Package.**” It was developed and published in the Spring of 2023. A special thanks to my wonderful advisor, Dr. Mine Çetinkaya-Rundel, who carefully guided this project’s creation and ensured its success. An additional thank you to the members of my thesis committee, Dr. Colin Rundel and Dr. Merlise Clyde, for offering their extensive experience in statistical computing and R package development. I hope this book successfully details the motivation, process, and results of the *duke* package and how it can effectively empower users to create colorblind friendly Duke visualizations.

For questions and concerns, please reach out to me at aidan.gildea@duke.edu. In addition, I encourage you to post any Issues on [Github](#).

1 Introduction

In nearly all disciplines, there is an increasing push to promote accessibility. Data visualization should be no different. The goal of this book is to detail the motivations, development, and use of the *duke* package. By the end of “Promoting Accessibility in Duke Data Visualization,” you should be sufficiently informed on how to create official Duke visualizations in a colorblind friendly way.

1.1 Personal Motivation

The typical format of an undergraduate thesis was daunting to me. I was not sure that an 80+ page paper with all the academic fixings - literature review, methodology, etc - was of interest to me. This was particularly true because a pertinent research topic did not come to mind. I had previously completed research in national nutritional and food policy, but could not identify a way to translate it into a senior thesis. Therefore, when the Duke Statistics department emphasized their openness to “non-traditional” theses, I was intrigued.

After meeting with my soon-to-be adviser, Dr. Mine Çetinkaya-Rundel, we identified a valuable intersection in data visualization and package development: colorblind accessibility. As our use of data increasingly permeates all aspects of society, it is crucial that we integrate accessibility and inclusivity into our data visualization best practices. We must continuously consider questions like: Who is our audience? How can we make our visualizations more intelligible? Who might we be leaving behind? This thesis aims to empower users right here in the Duke community to more easily answer these questions. By developing an R package that allows for the creation of Duke official visualizations in a colorblind friendly way, we can begin to make accessibility a core tenet of data visualization.

1.2 Background

There are approximately 300 million color blind people across the globe. Approximately 1 in 12 men (8%) and 1 in 200 women (0.5%) are affected. More properly known as Color Vision Deficiency, or CVD, the condition is primarily inherited genetically but can also be acquired due to other chronic conditions such as diabetes. With the data visualization industry expected to nearly double by 2026, its application will further become a mainstay in fields like marketing, research, public health, and journalism. More people will be interacting with data visualization

than ever before – both on the creator and user ends – therefore, it is crucial that it becomes increasingly accessible to those with CVD.

R packages that implement color blind friendly palettes already exist; however, this project aims to construct a package that makes Duke’s color palette more accessible. Duke provides a detailed description of its official [color suite] (<https://brand.duke.edu/colors/>) and stresses its commitment to meeting Web Content Accessibility Guidelines (WCAG). The WCAG are globally recognized as the standard for web accessibility best practices in all fields. Duke details their color combinations that are compliant, but does not offer a simple way to integrate the information into data visualizations.

By establishing a package that allows users to easily integrate Duke’s style guide with color blind friendly adjustments, we can improve the accessibility of Duke related visualizations. The implementation of this package will be an effective step toward accessibility in the Duke community, and hopefully, will prompt future work on making data visualization more inclusive.

1.3 Thesis Goal

Empower Duke users to create color-blind friendly visualizations by constructing an R color package.

Reference

<https://www.w3.org/WAI/standards-guidelines/wcag/>

2 Methods

2.1 Color Sources

The office of University Communications provides an official color guide for the Duke brand. The color palette is intended to both bolster the brand of the University, as well as ensure cohesion among all official content [1]. The primary colors of the official palette are Duke Royal Blue and Duke Navy Blue. It is highly encouraged that one of these colors is present in any Duke official project, as they are reflective of the Duke Brand. The color guide also details an extended palette, which includes an additional 18 secondary and tertiary colors. Colors such as Copper, Eno, Ironweed, Persimmon, and Cast Iron are in the extended palette. University Communications urges the use of this extended palette for, “graphic accents, typography, backgrounds, [and] call-to-action buttons” [2]. For each color in the palette, users can access the following information:

- Color Name
- Pantone Matching System (PMS)
- Hex Code (HEX)
- Cyan, Magenta, Yellow, and Key (CMYK) Color Model
- Red, Green, Blue Color Model

We will primarily interact with the Hex codes for these colors in this project. Hex codes are a hexadecimal (base 16 numbering system) format for representing the amount of RGB in a particular shade [3] [4]. Additionally, Hex codes are recognized as the best color system for HTML and CSS formats; therefore, their use will work well for data visualization in publication and web browser contexts.

Table 2.1: **Duke Color Palette**

Color	Name	Hex Code
	Duke Royal Blue	#00539B
	Duke Navy Blue	#012169
	Copper	#C84E00
	Persimmon	#E89923
	Dandelion	#FFD960

Color	Name	Hex Code
	Piedmont	#A1B70D
	Eno	#339898
	Magnolia	#1D6363
	Prussian Blue	#005587
	Shale Blue	#0577B1
	Ironweed	#993399
	Hatteras	#E2E6ED
	Whisper Gray	#F3F2F1
	Ginger Beer	#FCF7E5
	Dogwood	#988675
	Shackleford	#DAD0C6
	Cast Iron	#262626
	Graphite	#666666
	Granite	#B5B5B5
	Limestone	#E5E5E5

In 2018, Duke added a color accessibility to their brand guide outlining how users could meet Web Content Accessibility Guidelines (WCAG) 2.0 AA guidelines. The WCAG are internationally recognized standards on how to make web content more accessible to people with disability - including those with color vision deficiency. Their accessibility standards are principled on four tenets: Perceivable, Operable, Understandable, and Robust [5]. Duke's accessibility guide illustrates a grid of every text-background color pairing and its respective compliance to the WCAG. There are four compliance categories specified:

- AAA: highest compliance, text-background contrast ratio of 7+
- AA: mid-range compliance, text-background contrast ratio of 4.5+
- AA18: mid-range compliance with large text only (18pt or 14pt bold), text-background contrast ratio of 3+
- DNP: does not pass compliance guidelines, text-background contrast ratio of less than 3

These compliance metrics will be an important resource for identifying color combinations that are accessible and selecting them for use in the *duke* package. The ratios provided by this guideline are intended for background and foreground combinations; however, they also apply to colors near and/or adjacent to one another. Duke's inclusion of an accessibility guide on their branding website is an important step, but it would benefit from a built-in functionality that allows users to interact with its information more easily. The *duke* package aims to empower users to engage with these accessibility recommendations more easily when creating data visualizations.

2.2 Color Metrics

In selecting the optimal color palette, several color difference metrics were considered. Euclidean distance is a commonly used measure; however, it is perceptually uniform, meaning that color pairings that have similar Euclidean distances can appear vastly different perceptually [6]. This is because of irregularity in the standard RGB (sRGB) color space, which causes evenly spaced hues to range in their perceptual similarity or difference [6]. The CIE chromaticity diagram (see Figure X) exhibits the color spectrum according to human vision. Its parabolic nature leads to irregularities in linearly scaled color scales and therefore weakens its efficacy in this context.

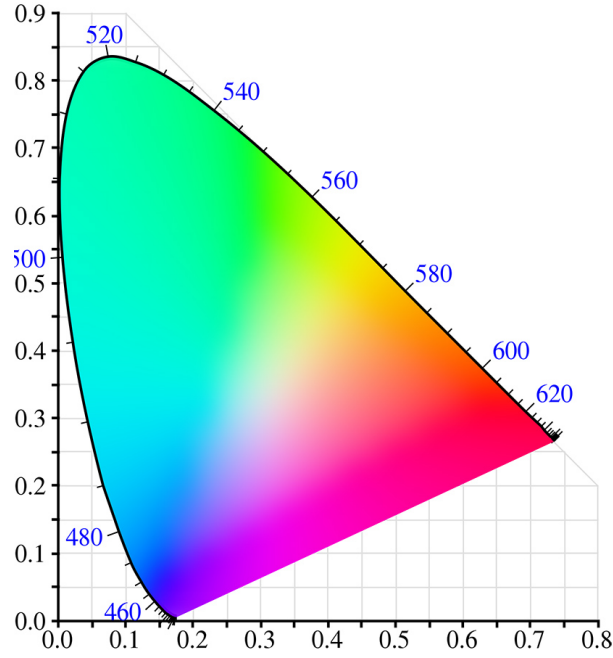


Figure 2.1: Figure X: CIE chromaticity diagram.

To accommodate these irregularities and center our color selection on human perception, we use the color contrast ratio as our primary metric. Color contrast is computed using “relative luminance” [7]. According to WCAG, relative luminance is the “relative brightness of any point in a colorspace, normalized to 0 for darkest black and 1 for lightest white”[8]. The formula for the color contrast ratio is as follows:

$$\frac{(L1 + 0.05)}{(L2 + 0.05)}$$

By dividing the the relative luminance of the lighter color (L1) by that of the darker color (L2), the color contrast ratio can range from a value of 1 (no contrast) to 21 (highest contrast)[9].

In line with the WCAG guidelines, we plan on only including color pairings in the package that meet a compliance level of AA18 or better.

It must be noted that given a total of 22 colors in the Duke official suite, our definition of the “optimal” palette may not be fully reliant on the color contrast ratio. The package is at its core, representative of the Duke brand and should meet certain aesthetic expectations for those with and without color vision deficiency. The development of our color scales, specifically, incorporates a somewhat ad-hoc approach in order to strike a balance between accessibility and recognition of the Duke brand. In finding a color palette successful on both fronts, users can be confident that they are producing visualizations that are identifiably Duke-related and accessible to a greater audience.

2.3 Package Development

As previously mentioned, package development posed a new challenge to the skill set I had learned through the Statistics curriculum. The “R Packages” text book served as an essential guide for learning, understanding, and eventually building the *duke* package. By clearly outlining the key structures of an R package, the functionality of the devtools package, and the importance of testing, I was well equipped to create *duke*.

The *duke* package offers four functions for users to make Duke branded visualizations in a colorblind friendly way.

2.3.1 `theme_duke()`

This function is largely inspired by the `theme()` function and its many derivatives. Its primary focus is to set the aesthetic elements for the visualization’s background and accessory features. These include the background color, grid lines, axis ticks, plot text, etc. With accessibility at the forefront of this project’s goals, it is essential that these choices lead to not only a visually pleasing result, but one that is comprehensible as well. To do so, `theme_duke()` implements a minimalist design with a White background and grid lines removed. This decision reduces excessive noise in the plot, and instead allows audiences to focus on the story the data tells. This design is complemented by uniform font text and coloring across the entirety of the visualization. Titles, axes, captions, and more, are all in Duke Navy Blue because of its strong contrast from the White background. It has a contrast ratio of 14.76 with White, which is the third highest among all colors in the Duke extended palette.

In addition, all text is in Atkinson-Hyperlegible font. Named after the founder of the Braille Institute, J. Robert Atkinson, the font uses “unambiguous, distinctive elements” to improve readability for individuals with vision impairment [9]. Its use - both in the package and all complimentary documentation - is essential to promoting vision accessibility. Font implementation was one of the most difficult challenges throughout the package development process.

RStudio provides a limited suite of fonts, requiring users to download and load additional font families to their own system. On the individual-level, this is not too difficult. However, the process necessitates careful thought when building a package for community use. *How can users integrate quickly access the package's required fonts? What font implementation will require minimal memory? Why is being transparent with users important?* We must closely consider these questions, among others, when adding font specification to **duke**.

TO DO - Will know post meeting with Mine ** Explain font implementation, as well as what happens when users don't have text

2.3.2 `scale_duke_continuous()`

This function integrates Duke's official color palette into geometric objects (geom) that use a continuous scale. Examples of commonly used plots implementing continuous geom objects include scatter plots, heat maps, and jitter plots. The primary challenge of adopting a continuous color scale in a color blind friendly way is the lack of color distinction that naturally occurs. Continuous color scales are defined by a low and high end of a two color gradient, where the data assumes colors within this range based on its respective value. However, continuous plots do not need complete distinction to be successful. It is often most important that readers can identify overall trends in continuous plots.

To optimize `scale_duke_continuous()` for users with impaired color vision, the continuous gradient must sequence from a dark to a light color [10]. This is crucial because the data will be plotted beginning on the darker end of the scale, which has a higher contrast with our White background. Therefore, our gradient scale defines its low end as Duke Royal Blue and its high end as Hatteras. Duke Royal Blue has a color contrast ratio of 7.75 with White, making it sufficiently distinguishable. This, in tandem with Duke's recommendation to include Duke Royal or Navy Blue in all communications, makes it the ideal choice for the low end of the scale. Hatteras is a light grey with subtle violet undertones, making it a pleasing complement to Duke Royal Blue [11]. It is not highly distinct from the White background, but this is typical for the light extreme of continuous scales.

2.3.3 `scale_duke_color_discrete` and `scale_duke_fill_discrete`

To implement a discrete scale from Duke's official color palette, users can add `scale_duke_color_discrete` and `scale_duke_fill_discrete` to their visualizations. These functions provide a discrete palette that can be used for the color and fill parameters of ggplot geometric objects. An appropriate discrete palette is essential to creating user-friendly visualizations because its colors indicate distinct categories that exist in the data. Whereas continuous plots typically exhibit overall trends, discrete plots must be able to delineate effectively between specific groups. A majority of geom objects can incorporate a discrete scale, including bar plots, histograms, and density plots.

As previously mentioned, this project observes color contrast ratio as the primary metric for determining colorblind friendly combinations. To identify visibly distinct colors for our discrete palette, we generate the contrast ratio between all pairing from the Duke color guide, with the addition of White since it is the background color. Using this data, we can generate a heat map to visually explore the contrast of the pairings.

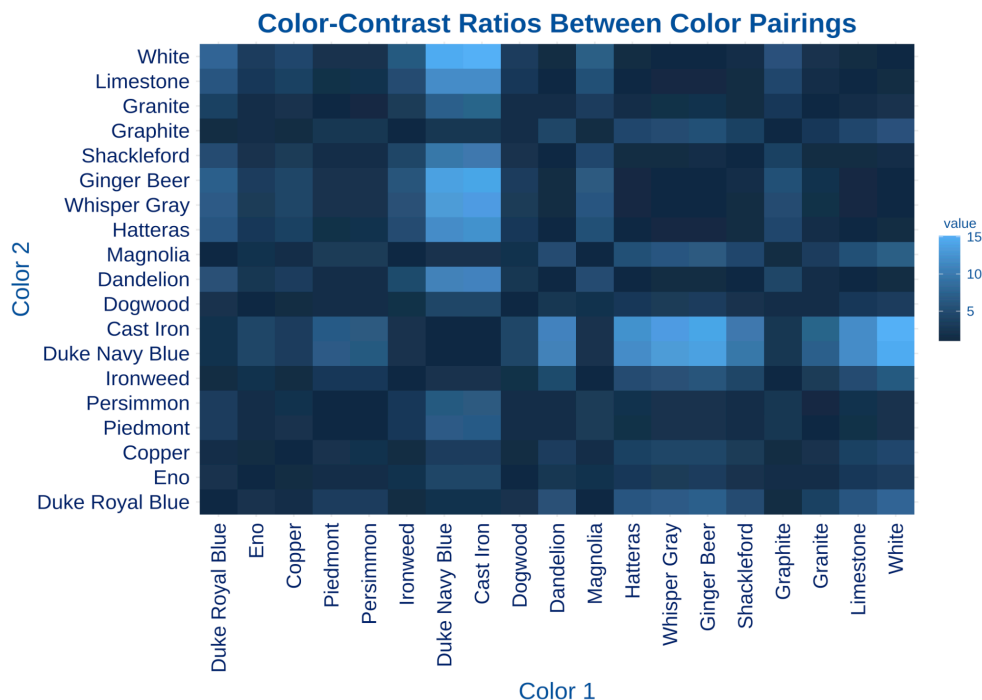


Figure 2.2: Figure X: Heatmap of color contrast ratio between color pairings.

Lighter tiles in the heat map indicate pairings with higher contrast ratios, which are preferred. Using the heatmap and data frame of ratios, we craft a palette that promotes accessibility and the Duke aesthetic. It is quickly evident that selecting eight colors all with pairing ratios compliant with WCAG is not possible. While this is not ideal, visualization best practices recommend users limit their number of discrete categories. This lends to less busy plots that are more easily understandable. Thus, the goal of our discrete palette is to carefully order colors so that those most used (those at beginning of palette) are most compliant. The result is the following palette:

To ensure that this palette is accessible, we use the `colorblindr` package to visualize sample plots from the perspective of individuals with different types of color vision deficiency. There are several categorizations of CVD, yet the most common is red-green color blindness. As the name suggests, it makes it difficult for individuals to distinguish between red and green. Among this population, approximately 62.5% are deuteranomalous - meaning they have green-weak vision that makes green appear more red [12]. Protanopia and deuteranopia prevent



Figure 2.3: Figure X: Discrete palette. Colors (from left to right): Duke Navy Blue, Piedmont, Duke Royal Blue, Eno, Copper, Ironweed, Cast Iron, Persimmon

individuals from distinguishing between red and green at all, yet are far less prevalent at only 12.5% of the red-green colorblind population each [13].

Red-green color blindness is by far the most common type, accounting for approximately 99% of those with CVD. For the purpose of creating a feasible project scope, we focus primarily on accessibility for those with red-green colorblindness. Future work on this project - or one with a similar goal - should investigate the feasibility of visualization accessibility for those with blue-yellow color blindness, monochromacy, etc.

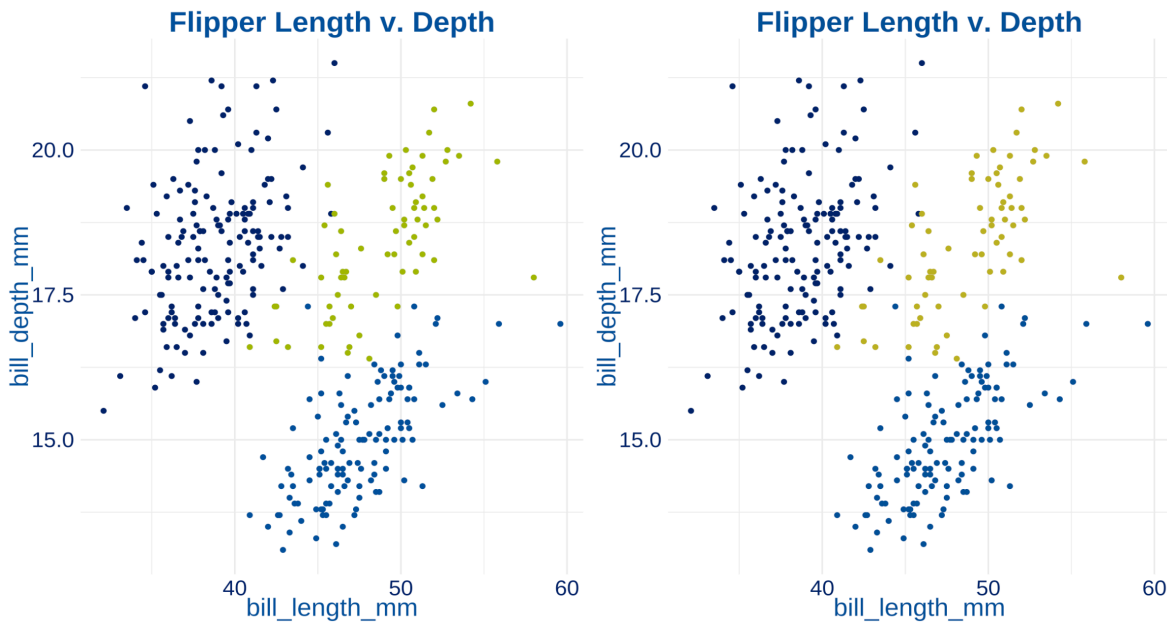


Figure 2.4: Figure X: Sample scatter plot from true perspective (left) and deuteranomalous perspective (right) using colorblindr package.

2.4 Lessons Learned

- Importance of creating test cases (testthat files)
 - Use of cover package
- iterative process --\> identify issues immediately and address them
- engage with existing resources/packages
 - drew a lot from other theme/scale_color functionalities

2.5 References

1. <https://brand.duke.edu/colors/>
2. Ibid
3. <https://www.techopedia.com/definition/29788/color-hex-code>
4. <https://www.sciencedirect.com/topics/engineering/hexadecimal>
5. <https://www.w3.org/TR/UNDERSTANDING-WCAG20/intro.html#introduction-fourprincs-head>
6. <<https://programmingdesignsystems.com/color/perceptually-uniform-color-spaces/>>
7. <https://www.accessibility-developer-guide.com/knowledge/colours-and-contrast/how-to-calcu>
8. <https://www.w3.org/TR/WCAG20/#relativeluminancedef>
9. ibid
9. <https://fonts.google.com/specimen/Atkinson+Hyperlegible/about>
10. <https://clauswilke.com/dataviz/color-pitfalls.html>
11. <https://www.ppgpaints.com/color/color-families/neutrals/hatteras-gray>
12. <https://www.allaboutvision.com/conditions/color-blindness/types-of-color-blindness/>
13. ibid

3 Function Descriptions

3.1 `theme_duke()`

3.2 `scale_duke_continuous()`

3.3 `scale_duke_fill_discrete()`

3.4 `scale_duke_color_discrete()`

4 Discussion

Main idea: What you should do in addition to what this package allows?

4.1 Colorblind Accessibility Best Practices

- importance of text
 - using hyperlegible font
 - ensuring its size is appropriate
 - * scaling it to the rest of the visualization, rather than manually setting so it is appropriate per use
 - use of background colors, lines
 - * clean minimal design allows for things most important to be visible
 - limiting the number of colors to 8 in discrete case
 - * best practice when it comes to readable graphs
 - describe things not necessarily addressed in thesis but other best practices to be aware of
 - * categories
 - * type of data

5 About

References