Dos and Don’ts for Less Ugly Visualizations

Juraj Medzihorsky

Durham University

July 10, 2023

Abstract

This is a guide for making less ugly data visualizations in dissertations. The dos and don’ts apply to many other formats, such as academic articles and texts in popular media. Unless otherwise noted, the example visualizations are by me. They are not chosen for excellence, but simply to illustrate the points.

Contents

1 Captioning **2**

2 Image file formats **3**
- 2.1 Raster images: Lossy and lossless compression . . . 3
- 2.2 Raster images: Resolution . . . 5

3 Text in visualizations **7**
- 3.1 Fonts and typefaces . . . 7
- 3.2 Scaling . . . 8
- 3.3 Direction . . . 8

4 Color palettes **9**

5 Design principles **14**
- 5.1 Ink-to-information ratio . . . 14
- 5.2 Small multiples . . . 14
- 5.3 Positive and negative space . . . 14
- 5.4 Aspect ratios . . . 14
- 5.5 Show the uncertainty! . . . 14
- 5.6 Don’t mislead . . . 14

6 Common dataviz formats **14**
- 6.1 Univariate . . . 14

1 Captioning

In academic writing, including dissertations, all figures should be captioned and numbered. Sometimes, it may help to label figures with the letters of the alphabet instead of numbers, or combine numbers and letters.

Don't *just drop visualizations into text like this:*

Do *add informative captions and include sources where appropriate:*

Figure 1: A Brawndo can. Source: *WIRED*, `https://www.wired.com/2007/12/brawndo-mutilat/`. Cropped and color balance adjusted.

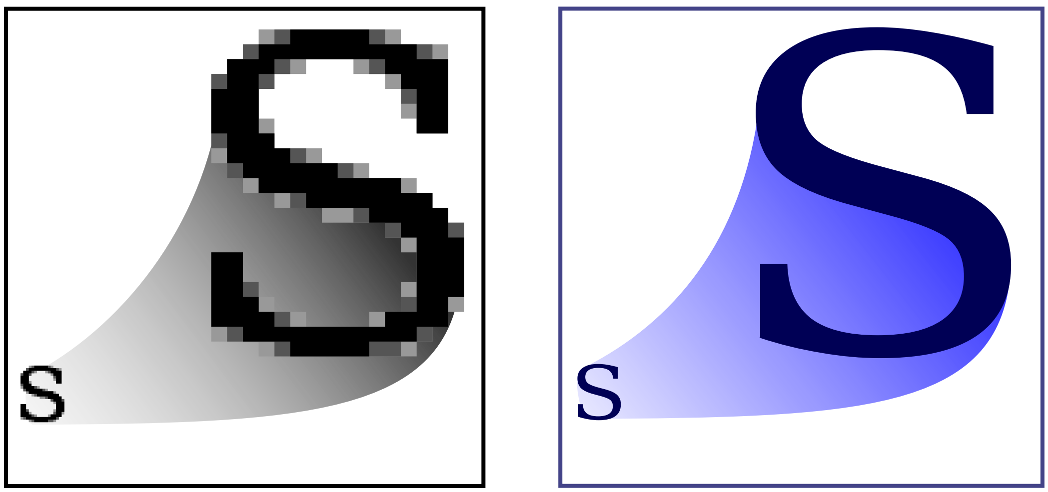
Refer to figures like this:

Figure 1 shows the iconic can design adapted from the 2006 sci-fi comedy Idiocracy.

2 Image file formats

There are two kinds of image file formats: raster and vector, illustrated in Figure 2.

- In raster formats, the image is recorded as a table in which each cell represents a pixel, and the value of the cell corresponds to the color of the pixel. As a result, raster images may look worse if displayed at higher or smaller resolution. Blurring and smudging of edges is a common issue. Common raster formats include `PNG`, `BMP`, or `JPEG`.
- In vector formats, elements are recorded as shapes. Thus, vector images can be displayed at many resolutions without issues, preserving sharp edges. Common vector image format is `SVG`, and vector images can be also saved in `PDF`, `WMF`, `EPS` and some other file formats.



Raster | Vector
GIF, JPEG, PNG | SVG

Figure 2: Raster vs. vector image formats. Source: *Wikimedia Commons* `commons.wikimedia.org/wiki/File:Bitmap_VS_SVG.svg`

> ***Don't*** *use raster formats unless you absolutely have to.*

> ***Do*** *use vector formats if you can.*

2.1 Raster images: Lossy and lossless compression

Raster image formats usually compress the table that represents the pixels. Such compression can be either lossless or lossy:

- Under lossless compression the original table can be perfectly recovered from the image file.
- Under lossy compression, some information on the original table gets lost in the process, and cannot be perfectly recovered afterwards.

A common uncompressed image format is RAW. Common lossless formats are PNG and BMP, and TIFF can be also used in this way. Common lossy format is JPEG.

Since lossy compression algorithms are optimized for images where sharp high-contrast edges are rare. This is an issue in data visualizations where such edges are the rule rather than the exception.

Don't *use lossy compression unless you absolutely have to.*

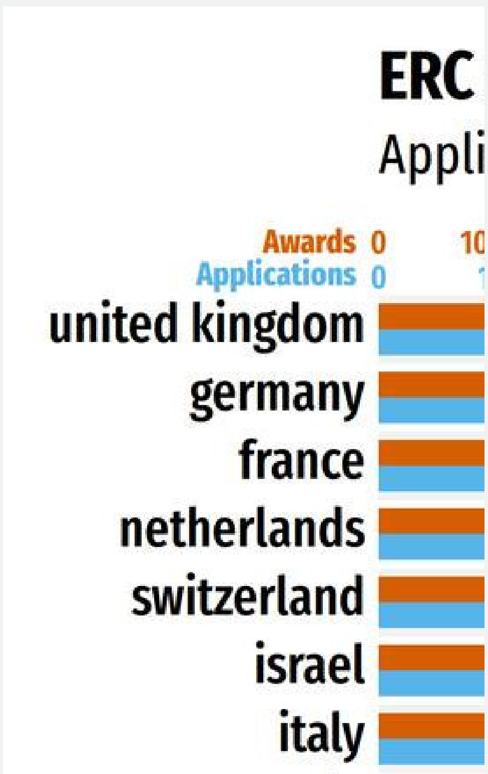


Figure 3: 577 × 934 pixel detail of a JPEG image (259kB).

Do *use lossless formats if you can.*

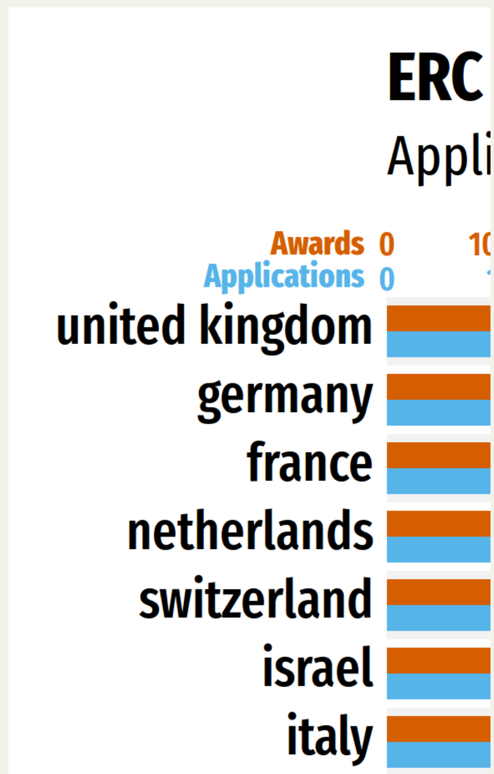


Figure 4: 577 × 934 pixel detail of a PNG image (493kB).

2.2 Raster images: Resolution

Resolution in raster image formats is defined by the number of horizontal and vertical pixels, and by the number of pixels per inch. Figure 5 illustrates different PPI values. It is usually better to save images in a higher resolution than that available on the devices on which it will be viewed. This is because raster images usually degrade quickly by 'zooming in', but much less so by 'zooming out'.

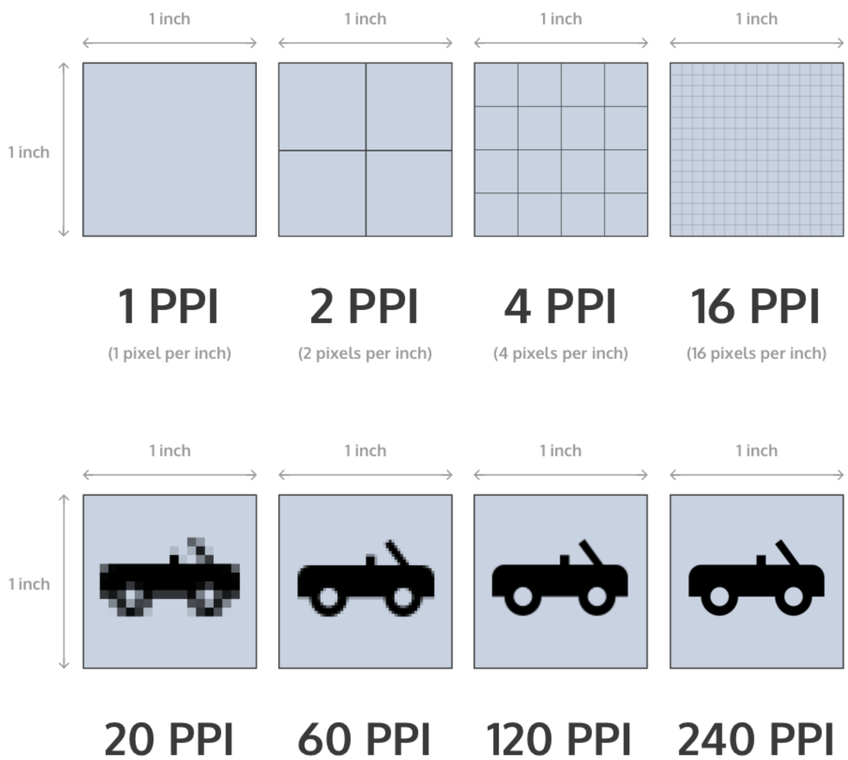


Figure 5: Illustration of increasing points per inch (PPI). Source: `developerlife.com/2018/07/21/designing-for-multiple-screen-densities-on-android/`

Don't use low resolutions and pixel densities.

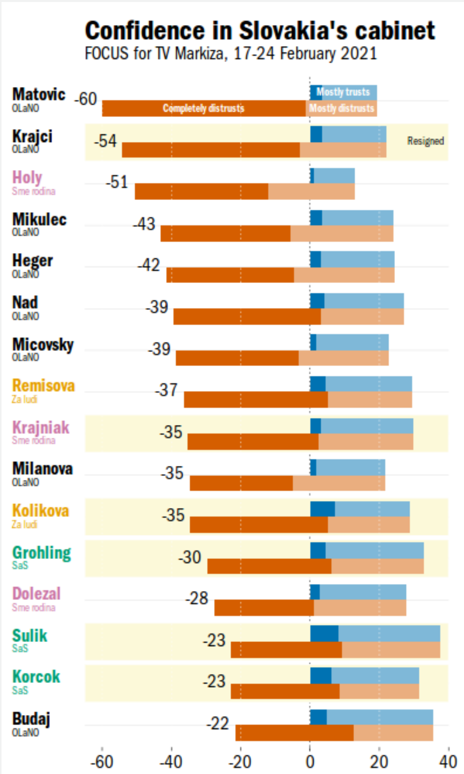


Figure 6: 400 × 667 PNG image.

Do use high resolutions and pixel densities.

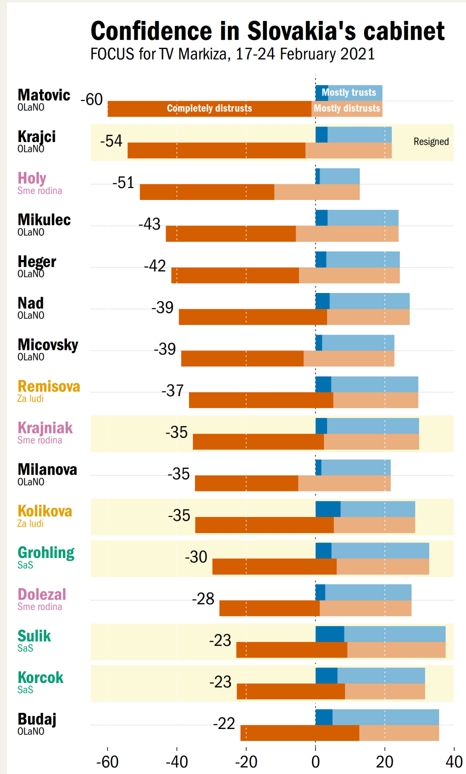


Figure 7: 1800 × 3000 PNG image.

3 Text in visualizations

Most visualizations will have some text in them, some a lot of it. Making sure the text is as easy to read as possible helps.

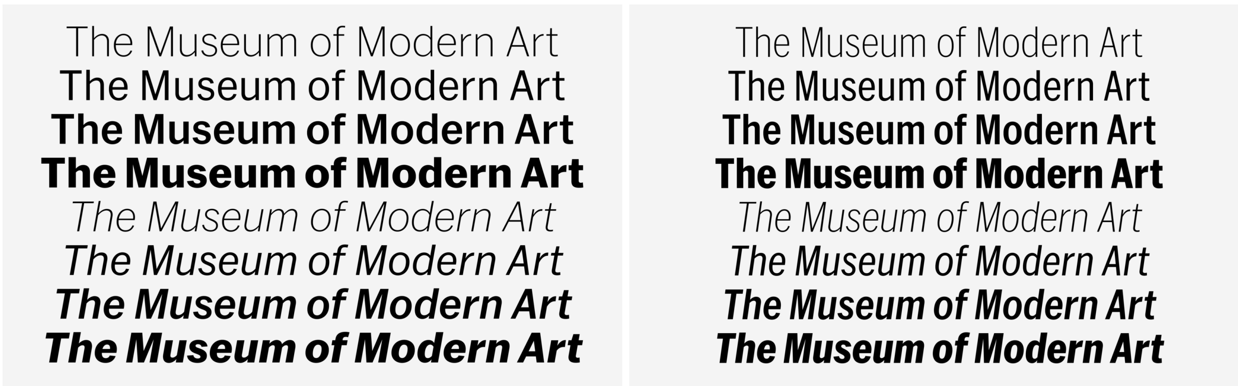


Figure 8: Illustration of the MoMA Sans typeface designed for the Museum of Modern Art in NYC, USA. Source: `commercialtype.com/news/moma_sans_for_the_museum_of_modern_art`

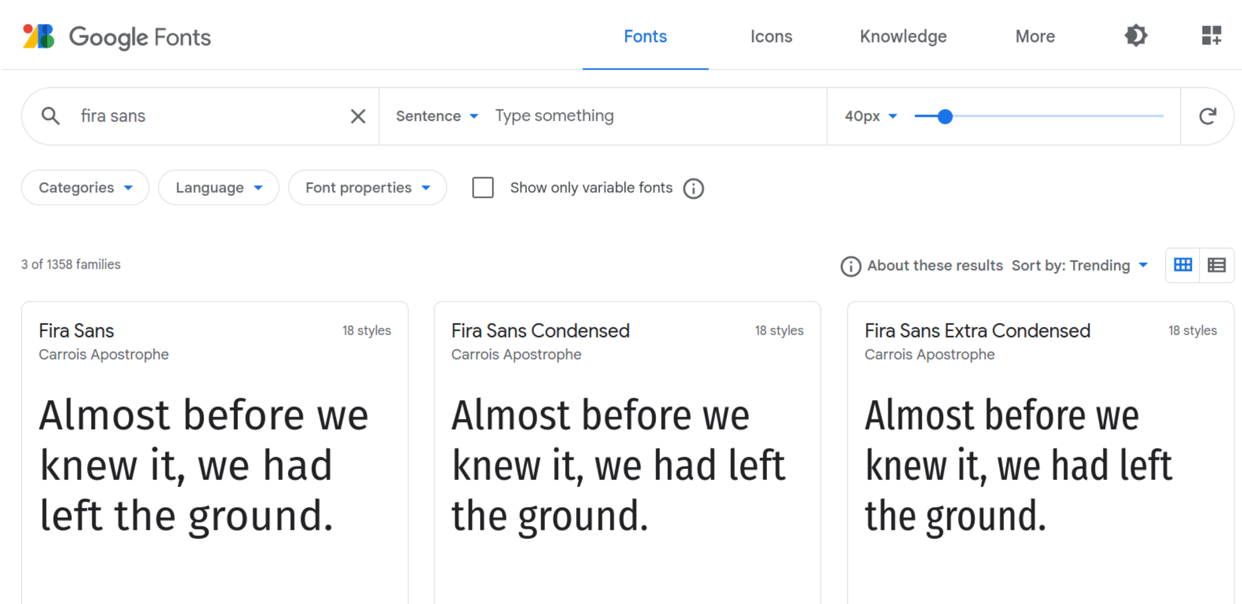


Figure 9: The Fira Sans typeface. Source: `fonts.google.com`

3.1 Fonts and typefaces

Choosing a good typeface for the visualization goes a long way towards making it more readable, and may improve its impact by eliciting more favorable aesthetic judgment as well. Most typefaces come in at least three forms: normal, **bold**, and *italics*. Many come in more than two weights, that is degrees of boldness. Some also come in different widths.

Figure 8 illustrates this on the MoMA Sans typeface, showing its four weights and two widths. While MoMA Sans is exclusively used by the Museum of Modern Art, there are many freely available typefaces. The Google Fonts service (`fonts.google.com`) provides access to a large and ever growing collection of such fonts. One of the most successful, and well-suited for data visualizations is the Fira Sans typeface, shown in Figure 9. It comes in 9 weights and 3 widths, and support for many special characters.

There are many different kinds of typefaces, shown in Figure 10. Contemporary academic writing routinely uses at least three of them: Serif, Sans Serif, and `monospaced`. Serif fonts are typically used for the main text, Sans Serif fonts for visualizations and/or headings, and `monospaced` fonts for computer code, where having all letters and symbols equally wide enhances readability. Sometimes, more ornamental typefaces are used in math.

Figure 10: Different typeface formats. Source: *Wikimedia Commons*

If you inspect the document properties of this document, you will find at least three typefaces: Cochineal, Cabin, and `inconsolata`.

3.2 Scaling

3.3 Direction

4 Color palettes

When choosing the color palette for a visualization, consider the following points:

- Are the colors easy enough to distinguish from each other? Even for readers with color-impaired vision?
- Will the mapping between the qualities or quantities represented by the colors of the palette easily understood by the audience?
- If the visualization will be viewed printed in greyscale, will the palette still work?

Not all humans perceive colors in the same way. Due to physiological differences, some people see fewer colors. This is also known as color-impaired vision or color blindness. Figure 11 illustrates three common types of color-deficient vision.

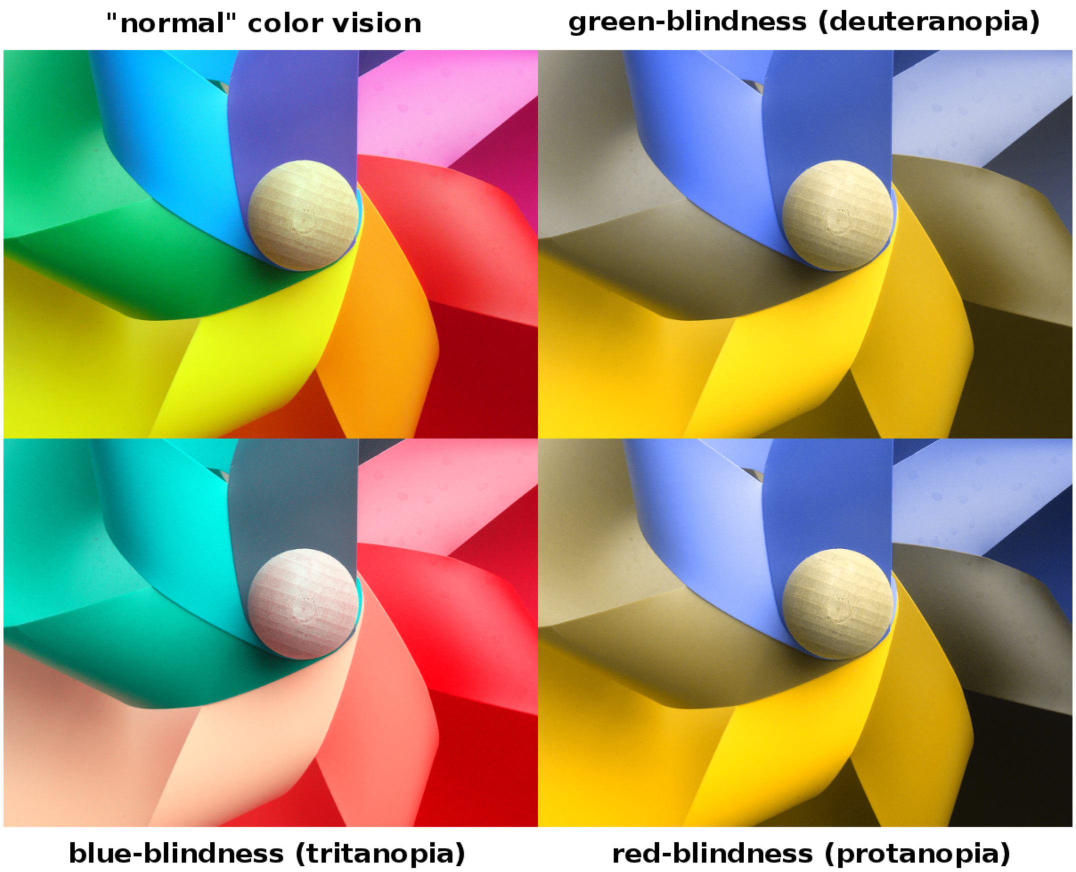


Figure 11: Illustration of different types of color-deficient vision. Source: *Wikimedia Commons*.

Don't *use color palettes with colors that are difficult to distinguish from each other.*

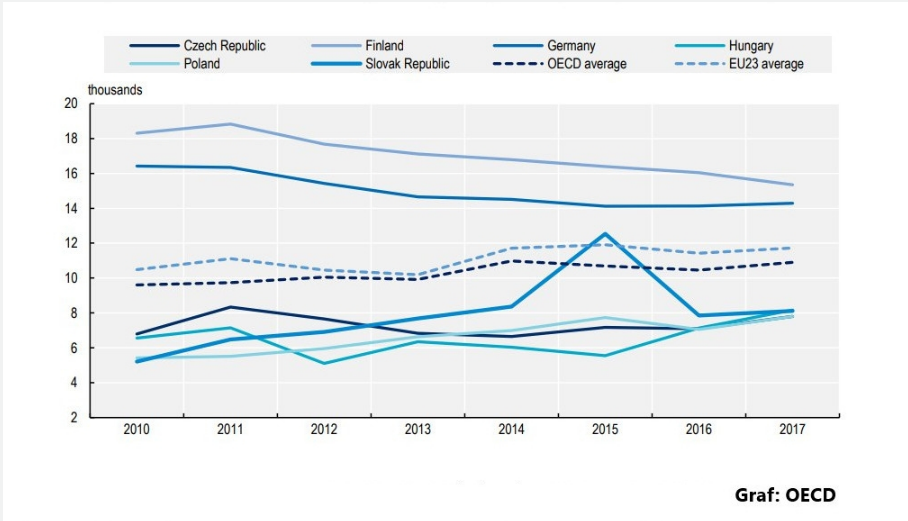


Figure 12: An illustration of a poorly legible color palette. To make things worse, it's a low resolution lossily compressed JPEG. Source: OECD.

Don't *use color palettes just because they are set as defaults in your plotting software.*

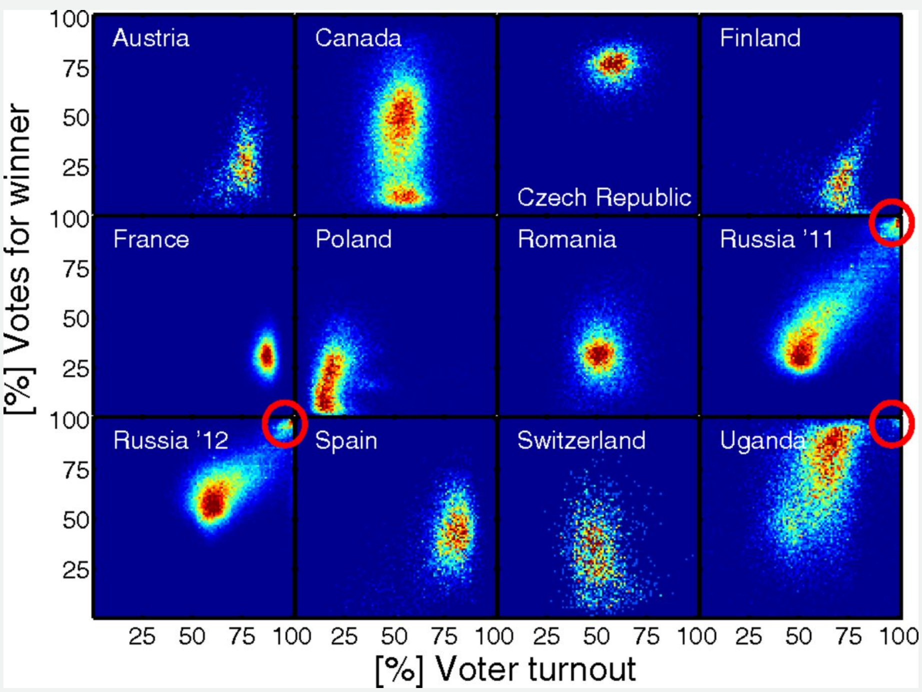


Figure 13: An illustration of software default color palette. To make things worse, it's a low resolution lossily compressed JPEG. Source: Klimek et al. (2012) `doi.org/10.1073/pnas.1210722109`

***Do** use accessible color palettes.*

Color	Color name	RGB (1–255)	CMYK (%)	P	D
	Black	0, 0, 0	0, 0, 0, 100		
	Orange	230, 159, 0	0, 50, 100, 0		
	Sky blue	86, 180, 233	80, 0, 0, 0		
	Bluish green	0, 158, 115	97, 0, 75, 0		
	Yellow	240, 228, 66	10, 5, 90, 0		
	Blue	0, 114, 178	100, 50, 0, 0		
	Vermillion	213, 94, 0	0, 80, 100, 0		
	Reddish purple	204, 121, 167	10, 70, 0, 0		

Figure 14: An accessible color palette by Wong (2011). `www.nature.com/articles/nmeth.1618`

Figure 14 shows the colors of Wong's accessible palette in two coordinate systems: Red, Green, and Blue, and Cyan, Magenta, Yellow, and Key. Figure 15 shows how the two coordinate systems work. RGB is usually used for electronic devices (mixing lights) and CMYK for printing (mixing inks).

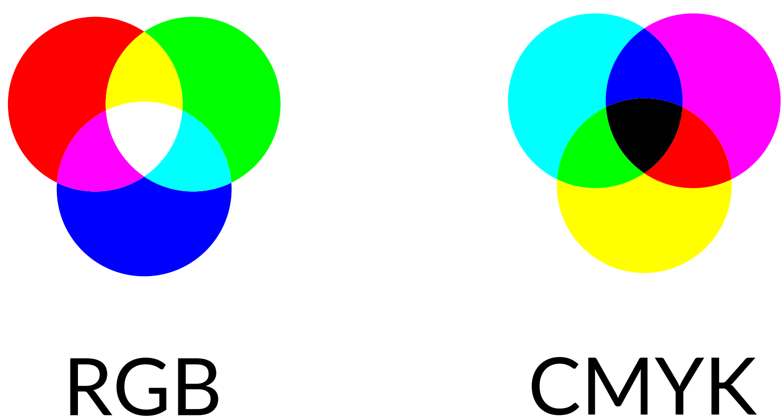


Figure 15: RGB and CMYK color coordinate systems. Source:`openclipart.org/detail/247535/rgb-cmyk`

***Don't** use palettes optimized for color screens without checking how they print in greyscale if you expect the audience to read them in that way.*

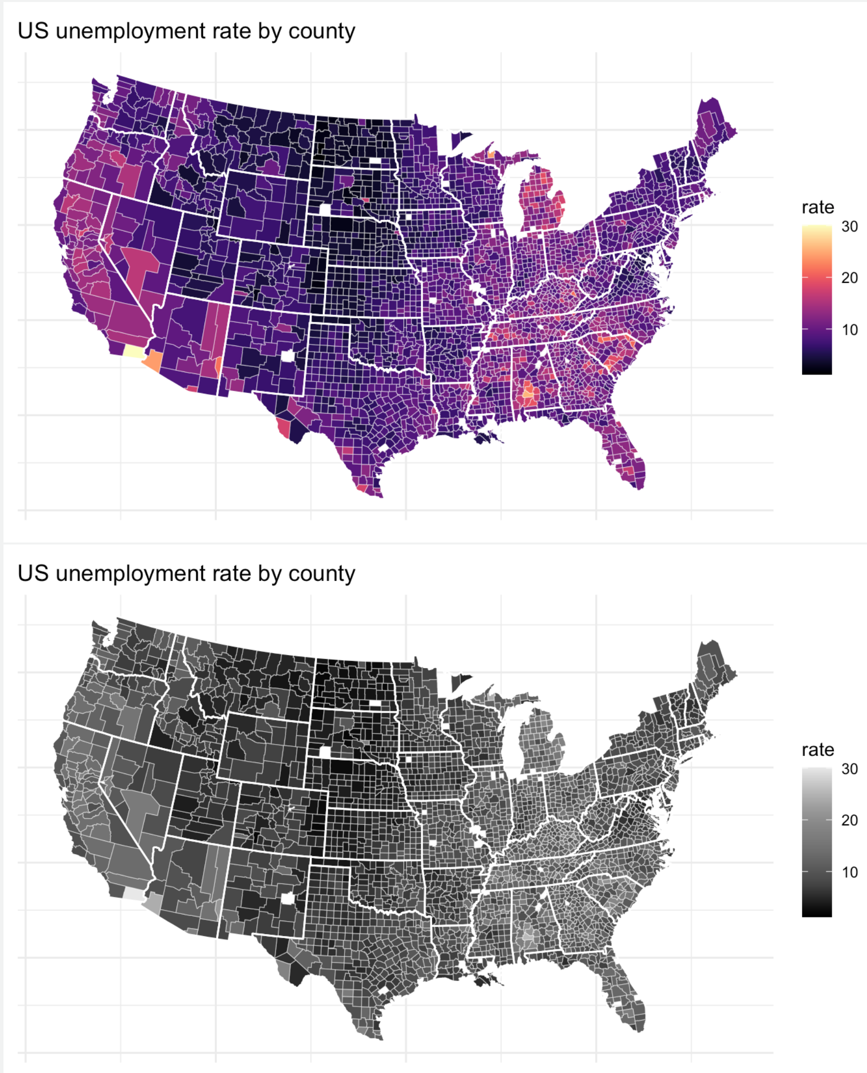


Figure 16: Top: A chloropleth using the highly popular `viridis` palette. Source: `cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html` Bottom: The same image projected onto greyscale.

***Do** use color palettes that build on well-known color vocabularies, unless it diminishes accessibility.*

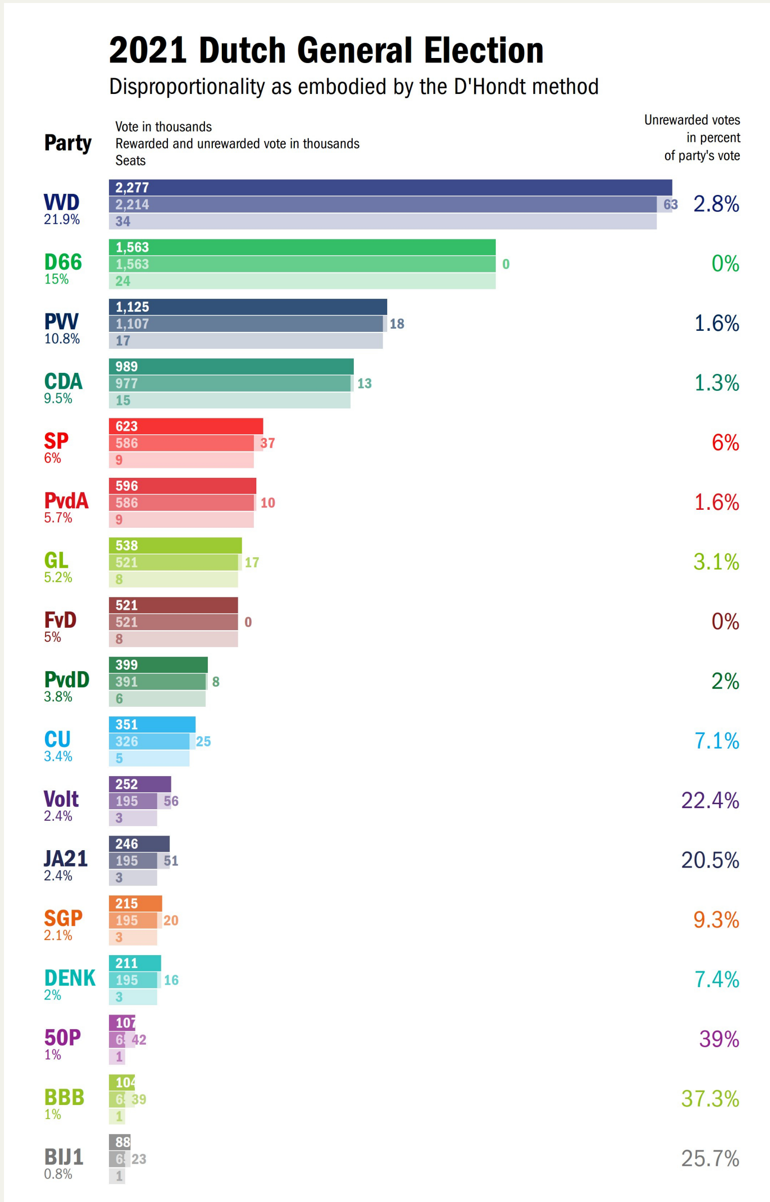


Figure 17: Election visualization using the parties' official colors. Hex codes for the colors are accessible by inspecting the HTML code of the Wikipedia entries for the parties.

5 Design principles

5.1 Ink-to-information ratio

5.2 Small multiples

5.3 Positive and negative space

5.4 Aspect ratios

5.5 Show the uncertainty!

5.6 Don't mislead

6 Common dataviz formats

6.1 Univariate