DATA 227

Using Color in Data Visualizations

2022-10-20

Numeric Variables

- Discrete
  - A variable that takes only values with jumps or steps (0, 1, 2, etc.)
  - Something about which you might ask, "How many?"
  - Number of cars in a household, number of full siblings a person has, etc.
- Continuous
  - A variable that can take any value (even if restricted to a particular range)
  - Something about which you might ask, "How much?"
  - Height, price, etc.

Categorical Variables

- Nominal
  - A variable that can be sorted into categories that have no relation to each other
  - Eye color, home state, etc.
- Ordinal
  - A variable that can be sorted into categories that have a certain order
  - 1-5 star ratings, Likert scale data, hurricane scale

Types of Visualizations 1

- Bar charts
  - Used to describe the distribution of categorical data
  - Discuss largest, smallest categories
- Histograms
  - Used to describe the distribution of numerical data
  - Discuss center (mean $\bar{x}$ or median), shape (number of modes, symmetry), spread (sd s, IQR, range), unusual values

Types of Visualizations 2

- Box plots
  - Used to describe the distribution of a numerical variable over multiple categories
  - Same as histograms, but compare across categories (distribution of category A has a higher center than distribution of category B)
- Scatter plots
  - Used to describe the relationship between two numerical variables.
  - Discuss form (linear, exponential, etc.), direction (positive or negative), strength (strong, moderate, weak–or use correlation r), unusual values

Bar Charts

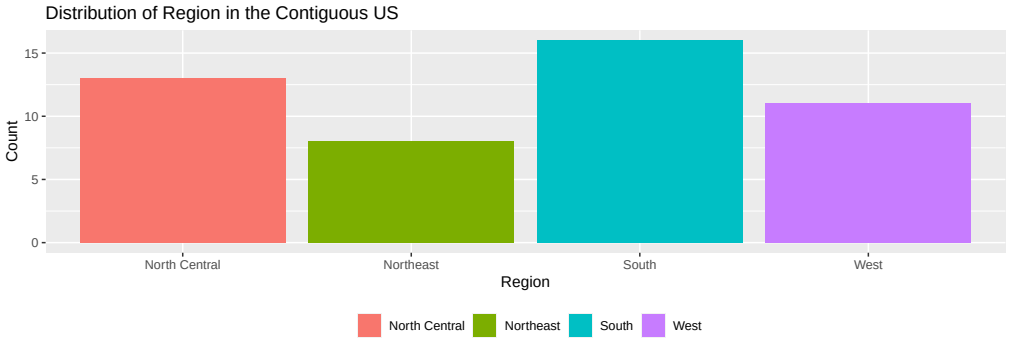


Figure 1: This bar chart shows the distribution of "region" across the 48 contiguous states in the US. The region with the most states is the South, and the region with the least states is the Northeast.

Histograms

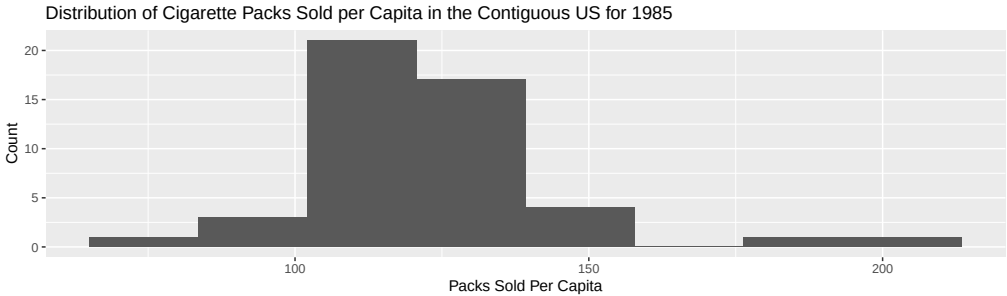


Figure 2: This histogram shows the distribution of packs sold per capita across the 48 contiguous states in the US for 1985. The distribution is centered at approximately 120, and ranges from 20 to 220. It is a unimodal, slightly right-skewed distribution. There are two potential unusual values near 200.

Box Plots

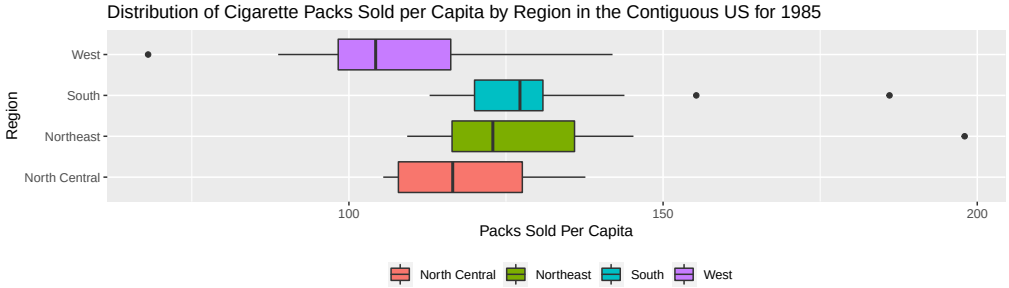


Figure 3: These boxplots shows the distribution of packs sold per capita by region across the 48 contiguous states in the US for 1985. Packs sold per capita appear to be the lowest in the West, and highest in the South. There are four potential unusual values–one on the low end of the distribution in the West, two on the high end of the distribution in the South, and one on the high end of the distribution in the North East.

Scatter Plots

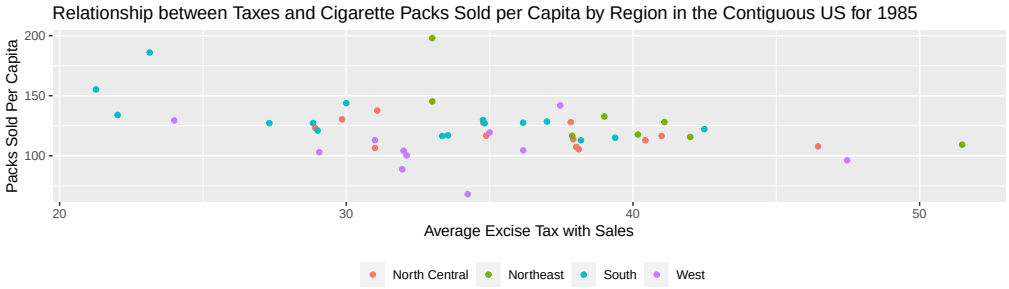
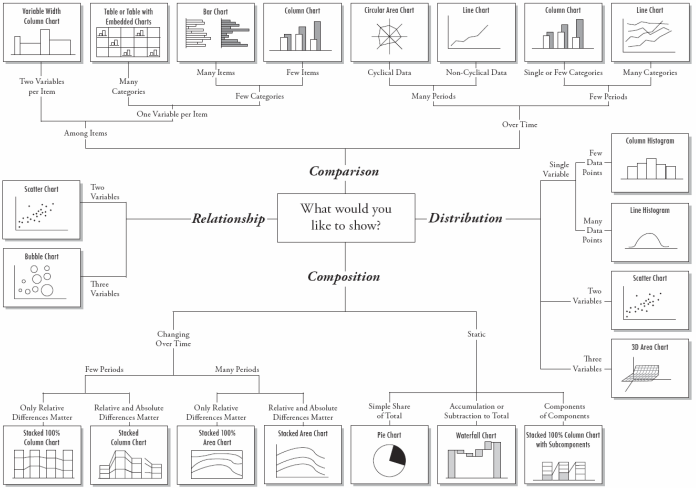


Figure 4: This scatterplot displays the relationship between packs sold per capita and average excise taxes for the across the 48 contiguous states in the US for 1985. It could be described as a negative linear relationship, but the relationship is not very strong (correlation -0.37). There is one unusual value from the Northeast around (33, 200).

Choosing a Graph Type 1

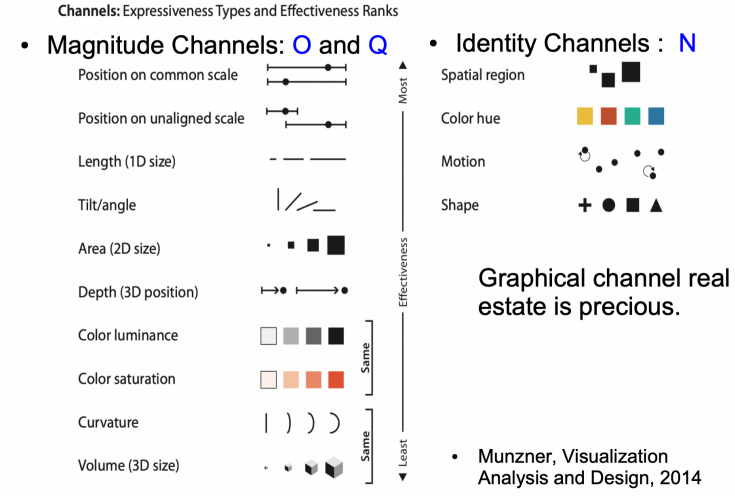


Choosing a Graph Type 2

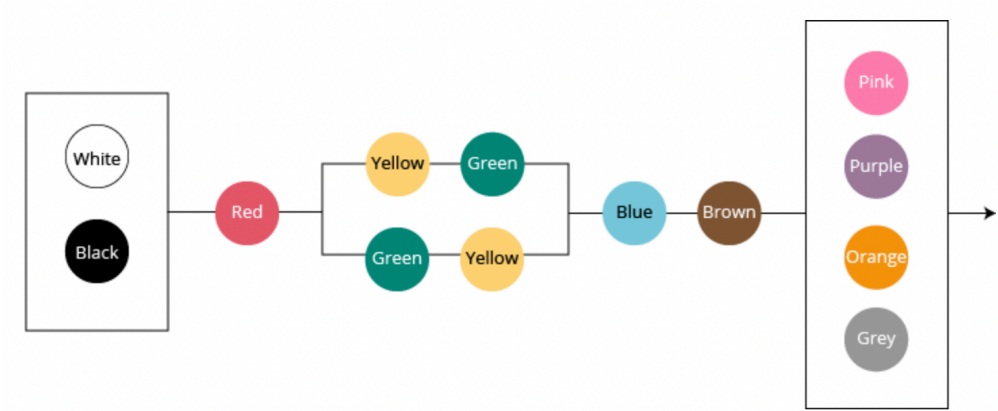
Other resources:

- Choosing a Graph Type
- Data Visualization Catalog
- One Chart at a Time

Encoding Channels



Universal Color Names

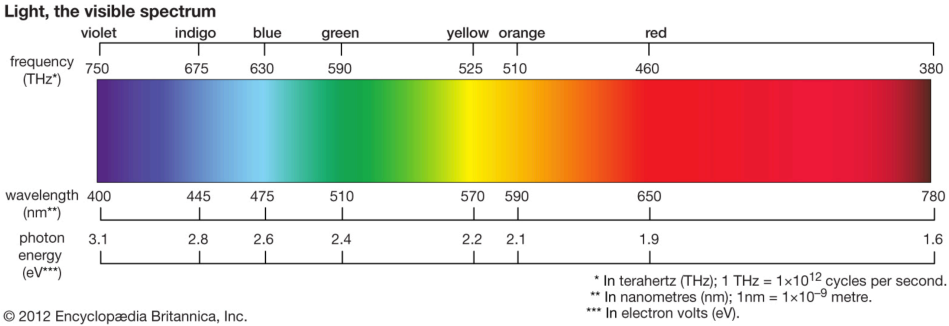


Berlin and Kay (1969)

Color Theory 1

Color Spectrum Image

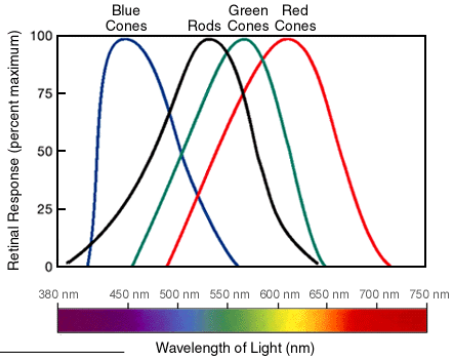
- "At the physical level, colour is produced by a mixture of wavelengths of light." [1]



[1] https://www.britannica.com/science/color

Color Theory 2

"Fortunately for us the human eye only has three different colour receptors, and so we can summarise the perception of any colour with just three numbers. To characterise a colour completely, we need to know the complete mixture of wavelengths." [2]



[2] https://ggplot2-book.org/scale-colour.html

Color Models

- RGB
  - "You may be familiar with the RGB encoding of colour space, which defines a colour by the intensities of red, green and blue light needed to produce it." [3]
  - Hex: Hexadecimal way of representing RGB color codes, used frequently in HTML.
- CYM
  - Cyan, magenta, and yellow are added together to reproduce different colors (mostly in printing, dyes).
- HCL Color Model:
  - Colors are defined in terms of their hue, chroma, and luminance.

[3] https://ggplot2-book.org/scale-colour.html

RGB vs. HCL

- "One problem with [the RGB] space is that it is not perceptually uniform: **the two colours that are one unit apart may look similar or very different depending on where they are in the colour space.** This makes it difficult to create a mapping from a continuous variable to a set of colours. There have been many attempts to come up with colours spaces that are more perceptually uniform. We'll use a modern attempt called the HCL colour space, which has three components of hue, chroma and luminance." [4]

[4] https://ggplot2-book.org/scale-colour.html

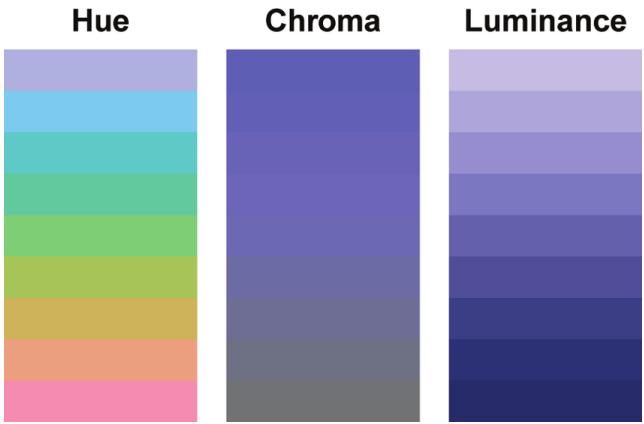
Color Definitions 2

HCL Color Model:

- "**Hue** ranges from 0 to 360 (an angle) and gives the "colour" of the colour (blue, red, orange, etc)."[5]
- "**Chroma** is the "purity" of a colour, ranging from 0 (grey) to a maximum that varies with luminance."
- "**Luminance** is the lightness of the colour, ranging from 0 (black) to 1 (white)."

[5] https://ggplot2-book.org/scale-colour.html

Visualizing HCL



Somewhere Over the Rainbow: How to Make Effective Use of Colors in Meteorological Visualizations

Using Color

- **Apply colors intentionally to data visualizations to highlight the focus of the data visualization. [6] Don't give into the temptation to decorate a graph in a way that undermines its ability to present data clearly. [7]**

We have three main tasks for color:

1. Marking nominal categories (legend).
2. Painting quantitative values (color map).
3. Highlighting key patterns.

[6] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts
[7] Practical Rules for Using Color in Charts

Nominal Categories (Task 1)

- Use different colors to represent different categories.
  - Hue on encoding channels chart.
- These colors should be as far apart from each other as possible in your color space–a.k.a., as distinct from one another as possible.
- No one color should stand out from the others!

Distinguish categories (qualitative)

Quantitative Values (Task 2) 1

- If you want color to show a numerical value, use a range that goes from light to dark in one of the universal color categories (Luminance and Saturation on encoding channels chart).

- If you need to represent diverging numeric values (e.g., from hot to cold), use two colors. Two colors may also help to provide more contrast.

Quantitative Values (Task 2) 2

- The sequence of colors should clearly indicate which colors correspond to small values, and which correspond to large values. * It should also indicate how distant two specific values are from one another.
  - Often, this implies that color changes uniformly over the scale, but not always.

Quantitative Values (Task 2) 3

- Note that by using a color gradient, we may be providing extra information about the variable.
- Gradients take two (or more) colors and linearly interpolate between them–in this way, we can give the impression that many values are possible between the two ends of the scale.
  - "For a two-point gradient scale, you want to convey the perceptual impression that the values are sequentially ordered, so you want to keep hue constant, and vary chroma and luminance."[8]
  - "Three-point gradient scales have slightly different design criteria. Typically the goal in such a scale is to convey the perceptual impression that there is a natural midpoint (often a zero value) from which the other values diverge."

[8] https://ggplot2-book.org/scale-colour.html

Ordinal Variables (Task 2)

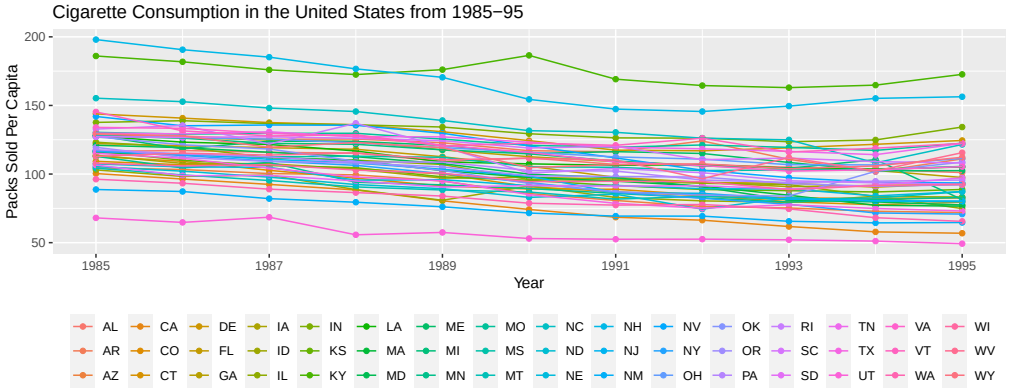
- Ordinal variables (and even to some extent, discrete variables) behave a bit differently.
- With ordinal variables, you want to preserve the natural ordering–but you don't necessarily want to use a gradient, since not all values are possible.
- With ordinal variables, you might want to use a sequential or divergent scale with binned colors–that way, humans can detect the inherent "separate-ness" of the values, but also the order.

Highlights (Task 3)

- When we want to emphasize a particular piece of data, we might use a different color to represent it in a graph.
- Contrary to using color to distinguish between categories, when we use color to highlight, we do want the colors to stand out!
- One way to do so is to use a saturated version of the color to highlight, and a more muted color for the rest of the data.

Tips for Using Color 1

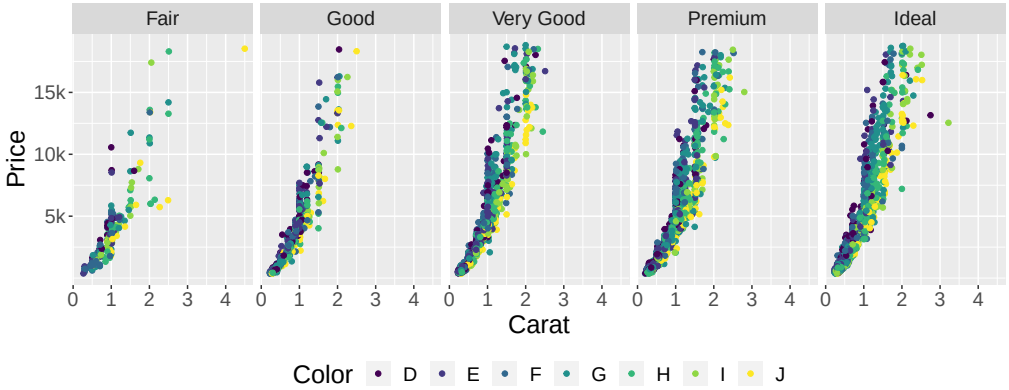
- Don't use more than (about) six colors![9]



[9] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Tips for Using Color 2

- If using small graphs for multiple categories, choose the same colors for all graphs.[10]



[10] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Tips for Using Color 3

- **Remove the colors if not necessary**. [11]
  - Possible exceptions: branding.

Figure 5: U of C Primary Scale

Figure 6: U of C Secondary Scale

[11] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

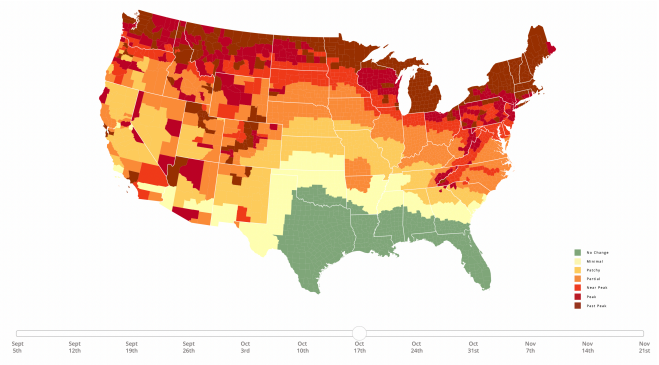
Tips for Using Color 4

- Make sure your color scheme is intentional–it should represent your data well![12]
- Don't choose a color scheme with meanings that don't match the story you are telling.
  - Hot and cold: Red and Blue
  - Republican and Democrat: Red and Blue
  - Hospital Codes: Pink (child abduction), Orange (hazardous materials or spills), Violet (violent or combative individuals), Yellow (disaster), Red (fire), Blue (cardiac arrest)
- We tend to think of dark as "more" and light as "less"–this is generally a good rule to stick to.

[12] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Tips for Using Color 5

On the flip side, if your plot does have meaning, you can choose the colors you use intentionally!



Fall Foliage Map

Tips for Using Color 6

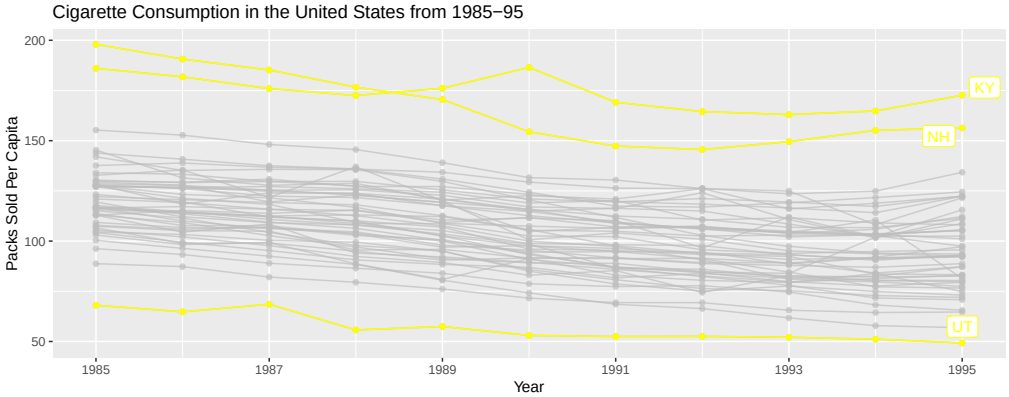
- Make sure that the color is legible when printed in black and white.
- The `viridis` package, available in `ggplot2` and `matplotlib`, and its companion package `viridisLite` [provide a series of color maps that are] perceptually-uniform, both in regular form and also when converted to black-and-white for printing. [13]
- These color maps are designed to be:
  - Colorful, spanning as wide a palette as possible so as to make differences easy to see,
  - Perceptually uniform, meaning that values close to each other have similar-appearing colors and values far away from each other have more different-appearing colors, consistently across the range of values
  - Robust to colorblindness, so that the above properties hold true for people with common forms of colorblindness, as well as in grey scale printing.

[13] Introduction to the viridis color maps

Viridis Color Scales

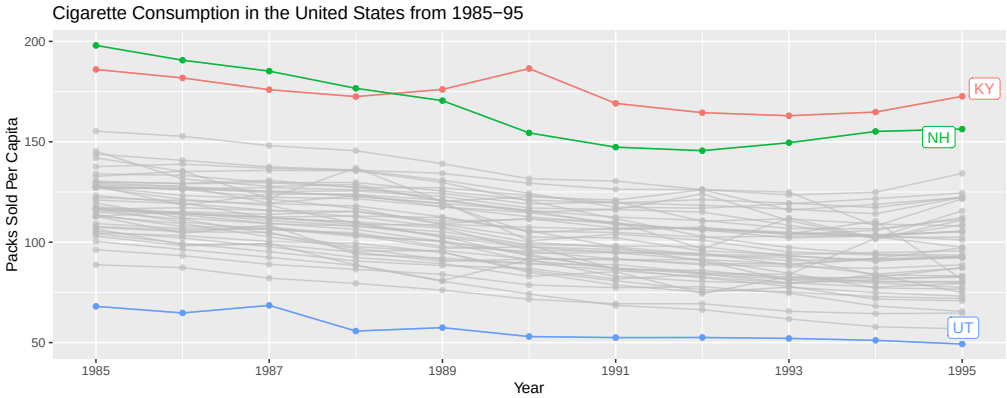
Tips for Using Color 7

- Use text that sufficiently contrasts background. [14]



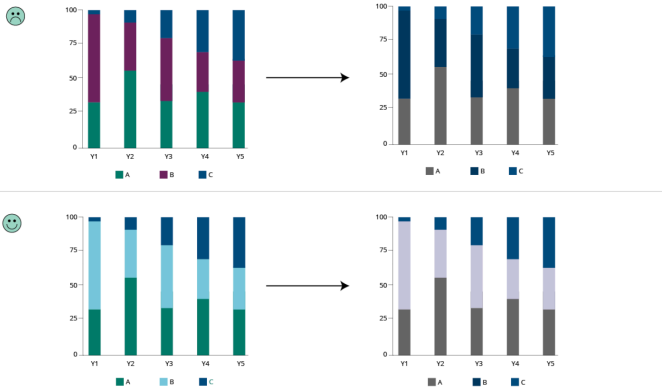
[14] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Tips for Using Color 8



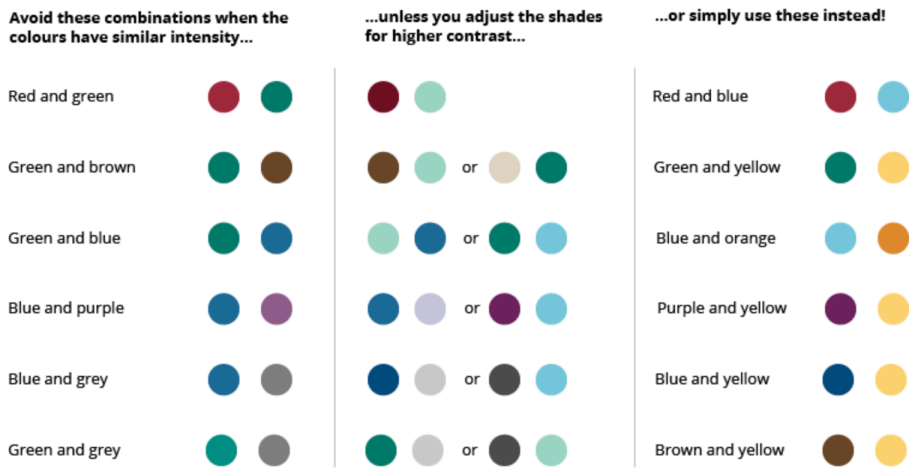
Accessibility 1

- Make your graph easy to read–that means easy to read for EVERYONE. Be aware of color blindness! [15]



[15] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Accessibility 2



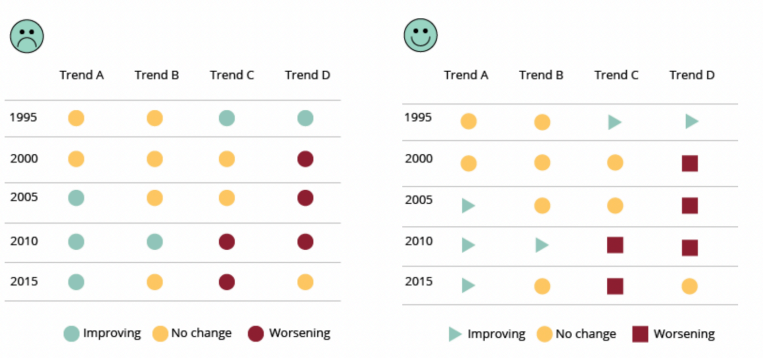
Accessibility 3

- `viridis` and `viridisLite` "provide a series of color maps that are designed to improve graph readability for readers with common forms of color blindness and/or color vision deficiency." [16]
  - These packages can solve two problems (color blindness and grey scale printing) at once!

[16] Introduction to the viridis color maps

Accessibility 4

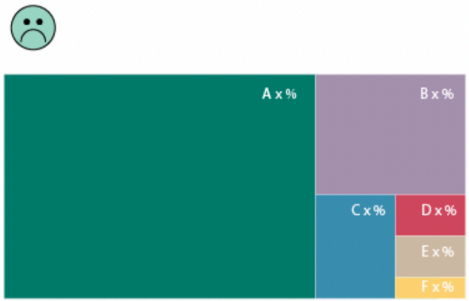
- Do not rely just on color to communicate–different shapes, patterns and textures can help distinguish between different design elements in your visualization.[17]



[17] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Accessibility 6

- Do use a monochromatic palette in different shades.[18]



[18] https://www.eea.europa.eu/data-and-maps/daviz/learn-more/chart-dos-and-donts

Sources

- Dos and don'ts of data visualisation — European Environment Agency
  - Color Blindness Part 1, Part 2, Part 3
  - Which Color Scale to Use
- `ggplot2`: Elegant Graphics for Data Analysis
- Introduction to the viridis color maps