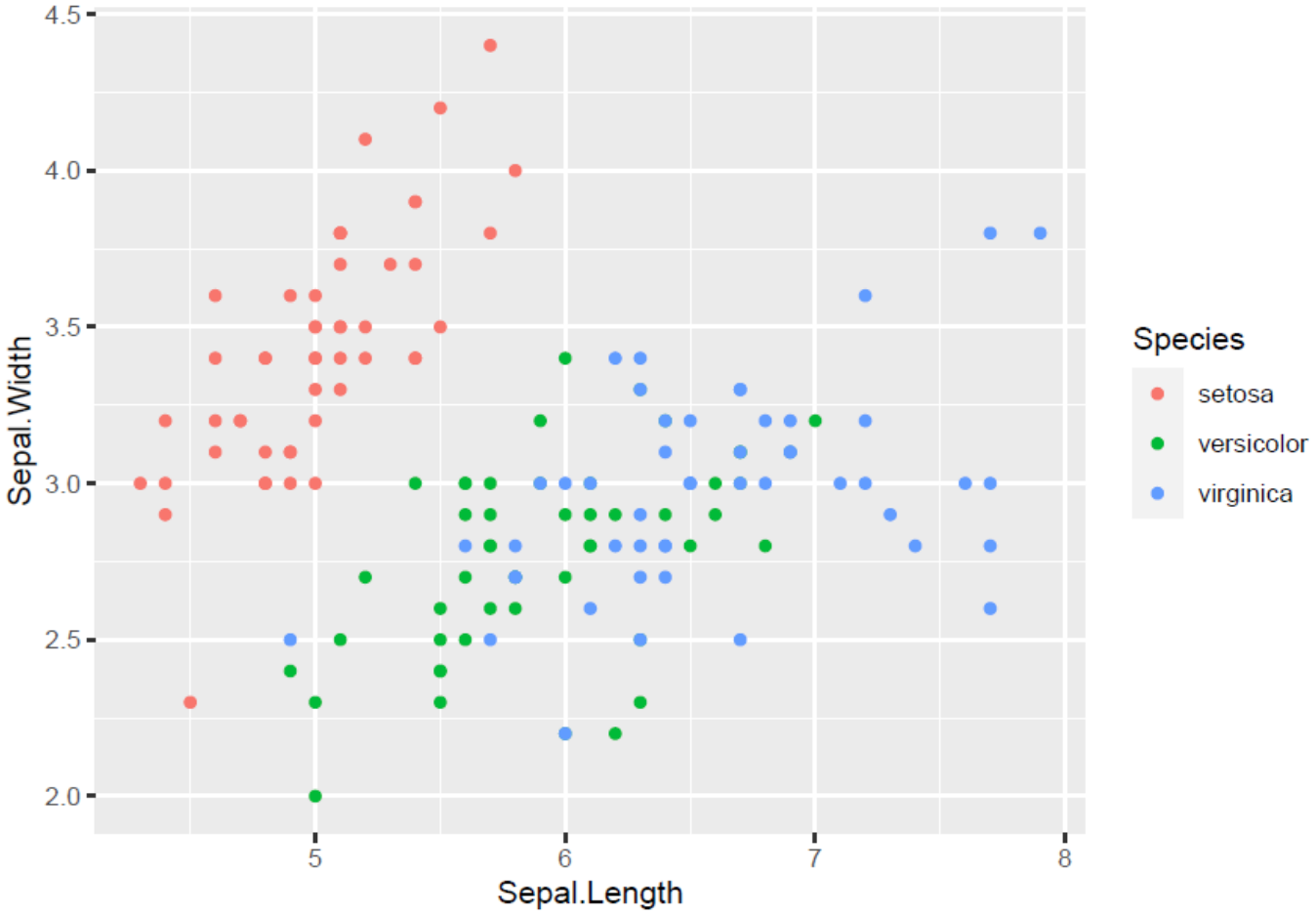
Making Better Graphs

Andrew Muehleisen and Lauren Hallett

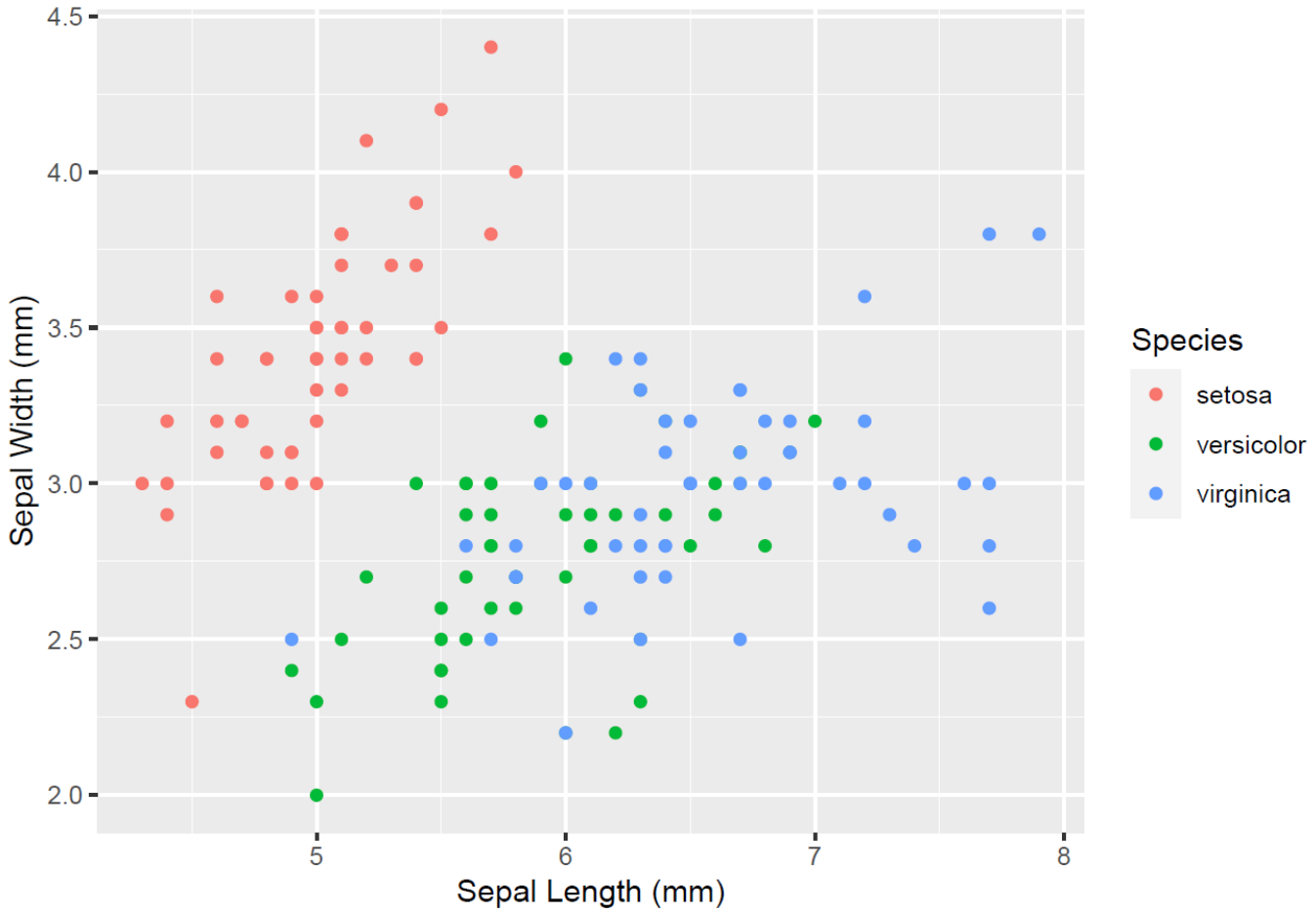
Moving on from the defaults

We've been making figures all term – now let's learn how to adjust default settings. To begin, here is a default graph plotted from the 'iris' dataset.



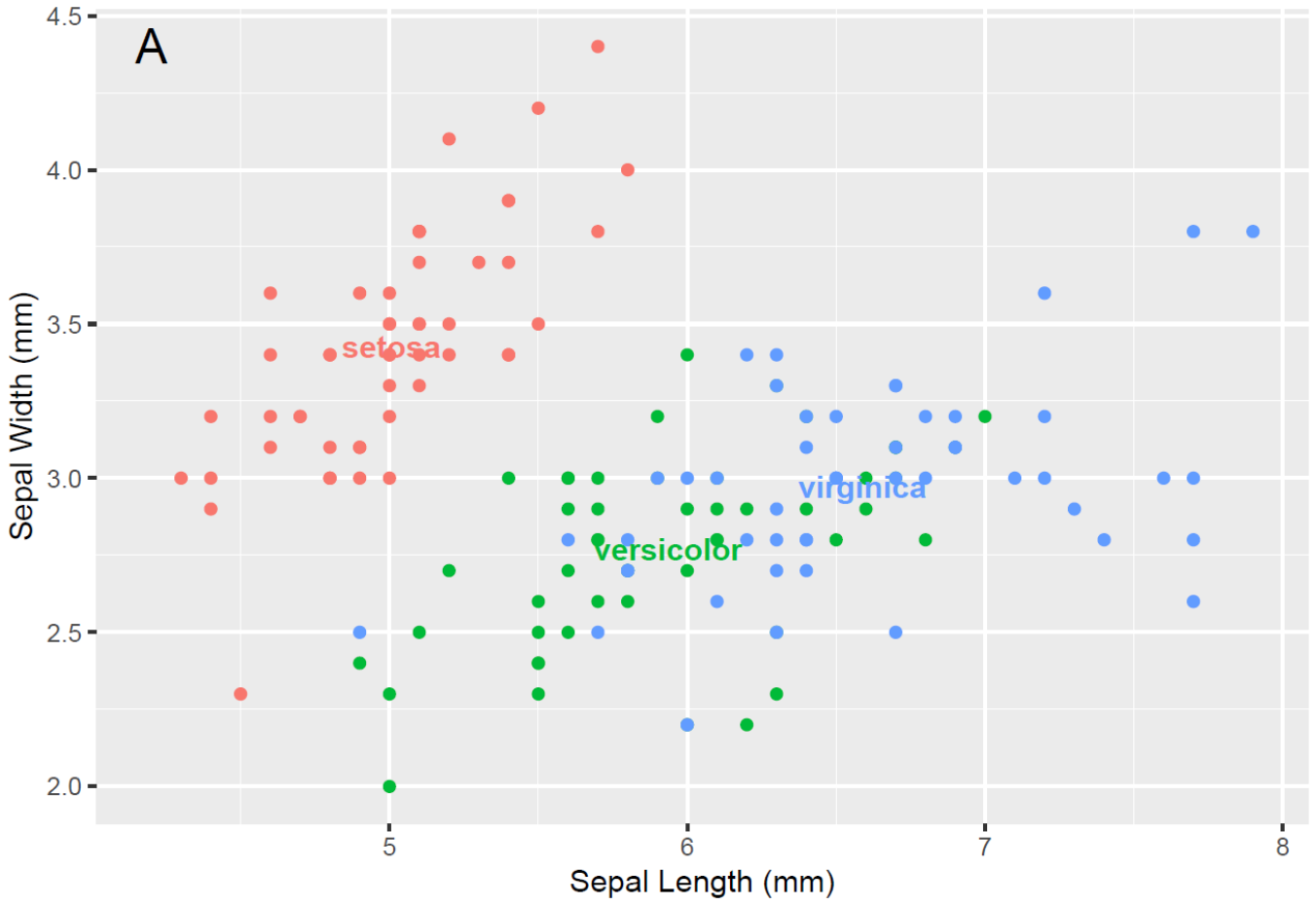
Moving on from the defaults

By default, variable names from the data set are included as our axis labels. Let's use more descriptive labels instead.



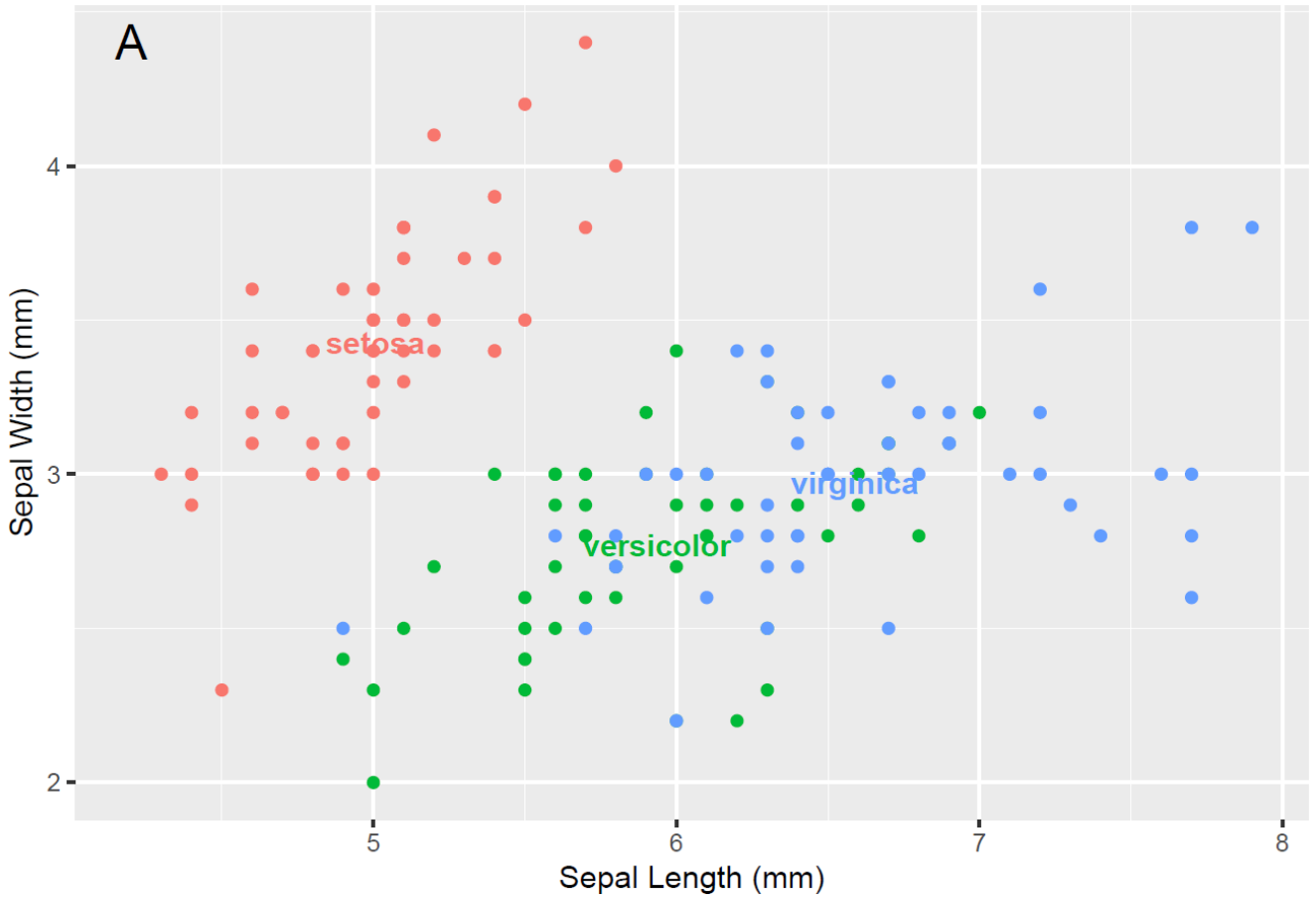
Moving on from the defaults

Rather than use a legend, we could label our points directly in the figure. Let's add a plot label in the upper left corner as well.



Moving on from the defaults

We can also change the scale and tick breaks on our axes. Here we change the y-axis to match the x-axis.

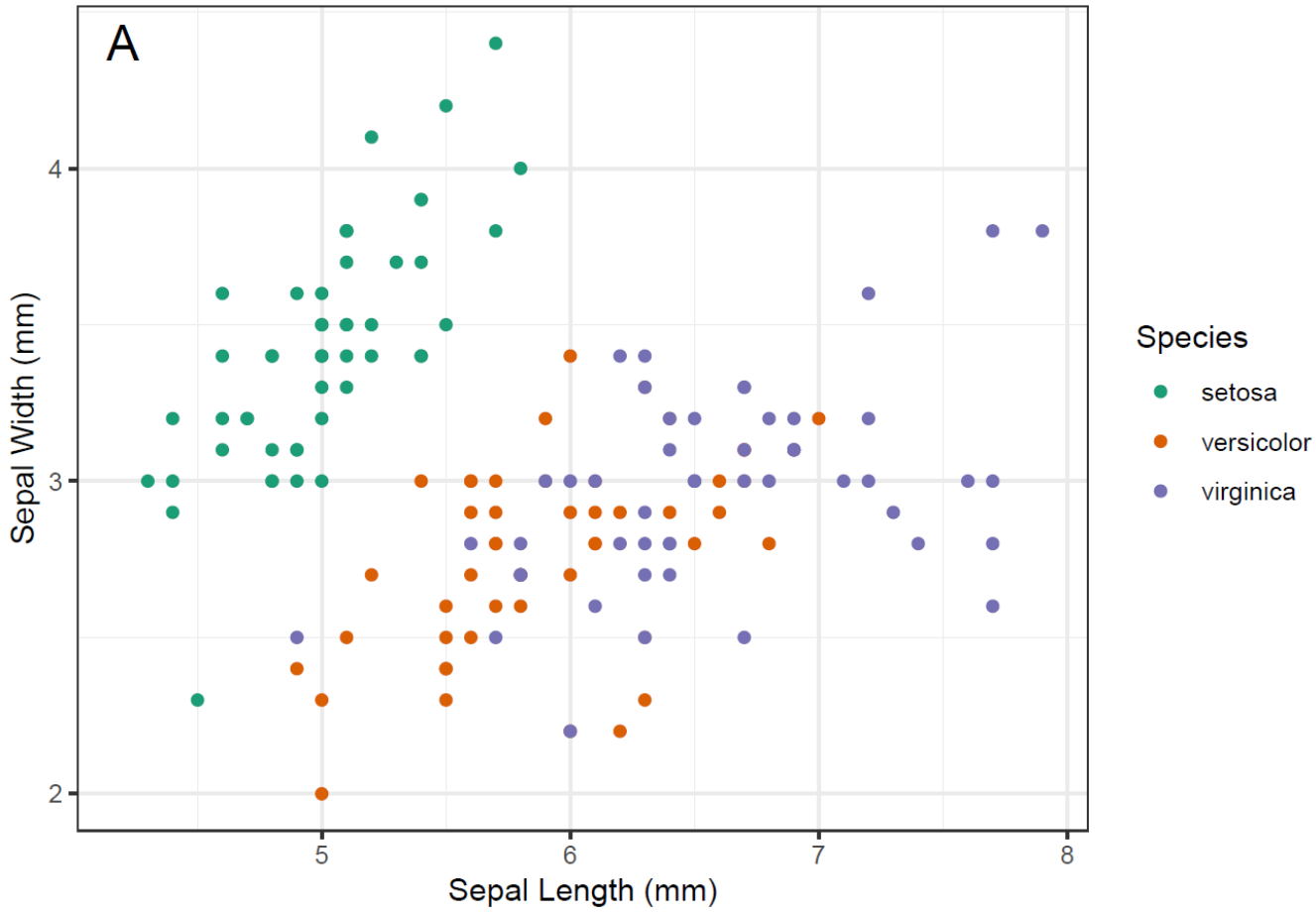


Moving on from the defaults

By default, ggplot chooses colors based around the color wheel. Lets instead use a more color-blind friendly palette.

Moving on from the defaults

Or, we could use an entirely new theme, such as ‘theme_bw’ from ggplot.



Moving on from the defaults

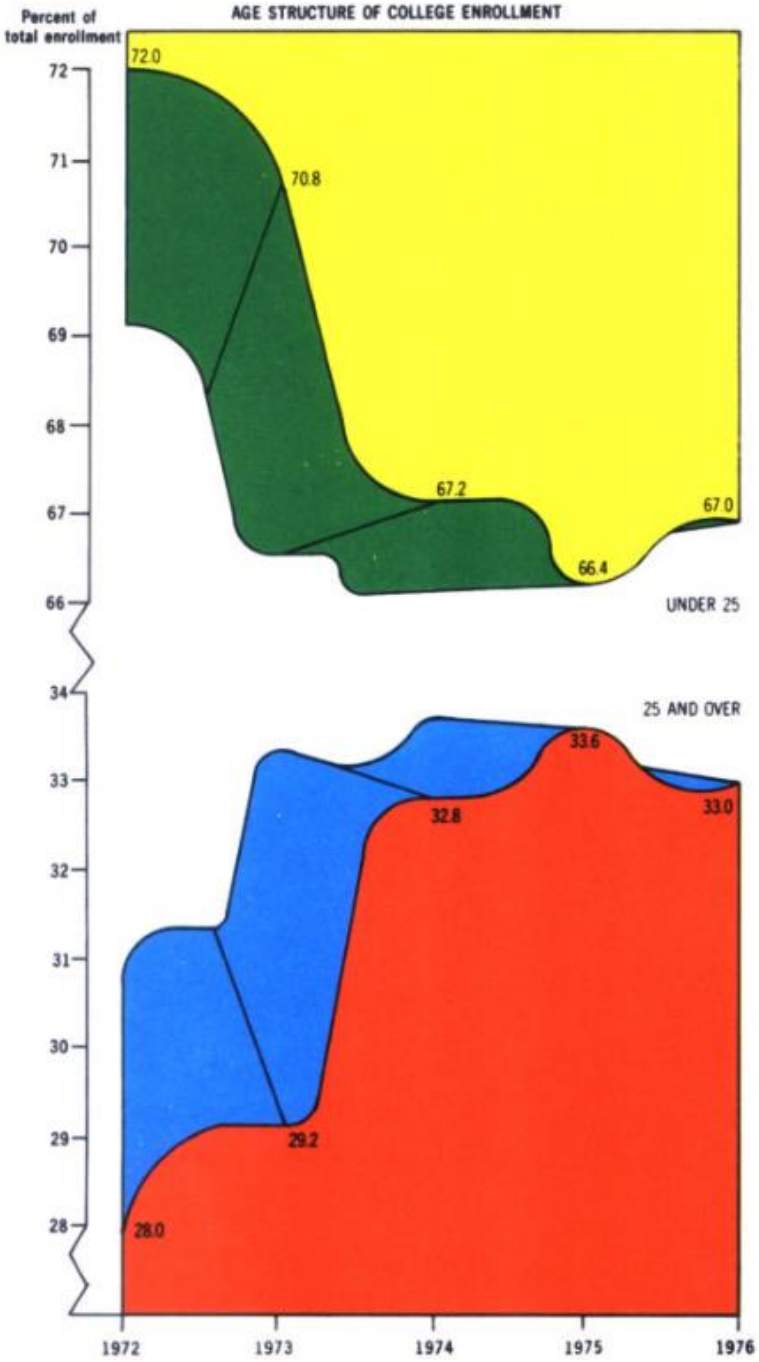
There are many ways to move beyond defaults in graphs, including:

- Labels and title
- Color palette, transparency
- Axis scale
- Point size, type
- Legends and annotations
- Background theme, grid, etc…
- Extra geometries, e.g. regression lines
- Multiple panels

How you deviate from R’s defaults is ultimately up to your discretion.

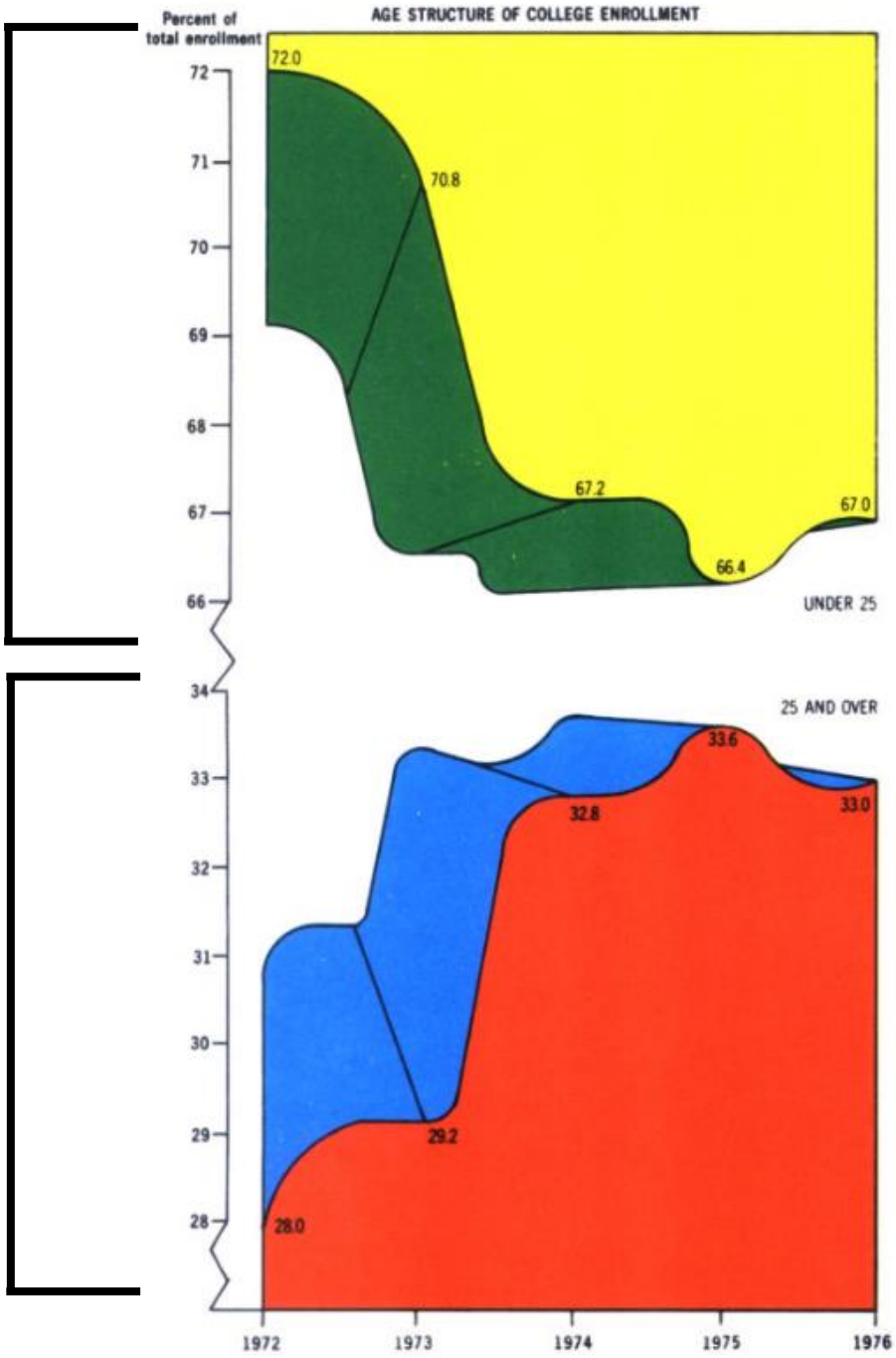
However, it is wise to consider some best practices…

This graph is terrible. Why?



This graph is terrible. Why?

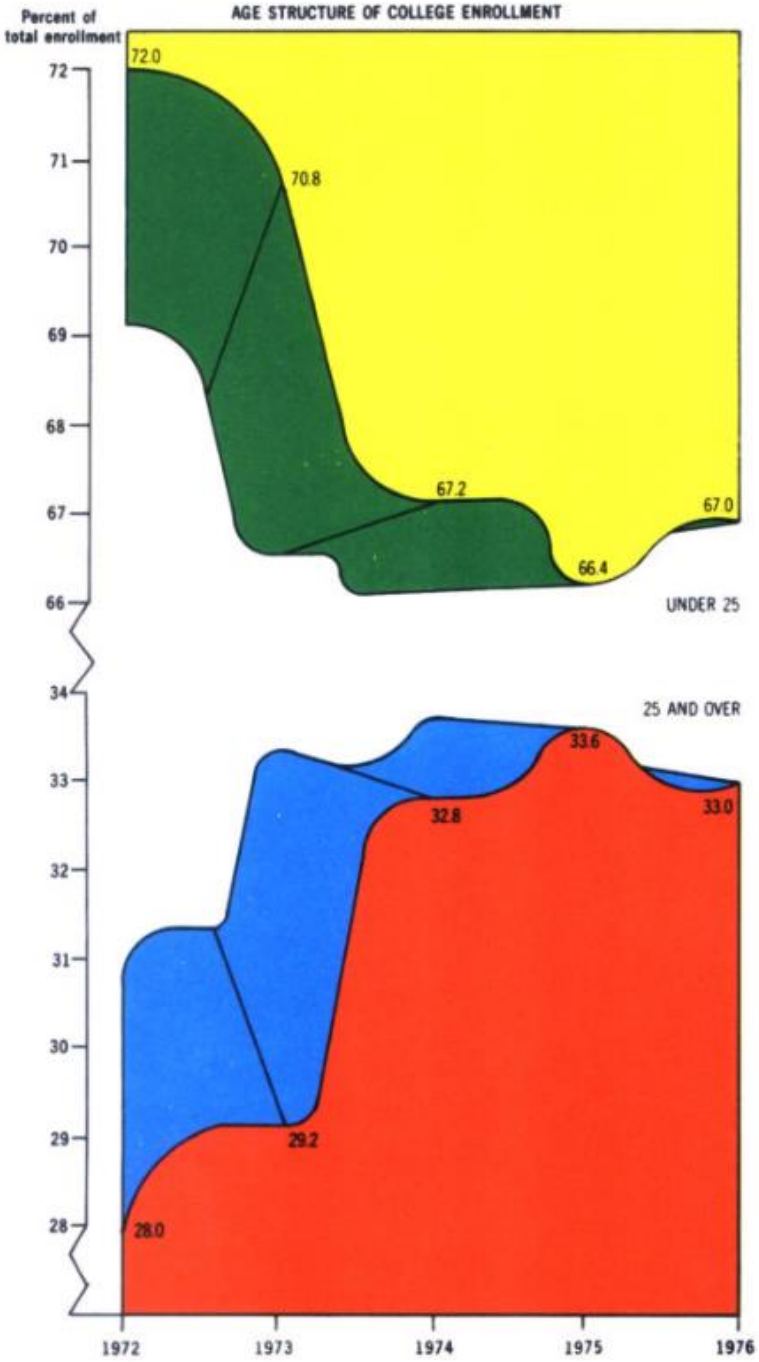
1. Information is duplicated



This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

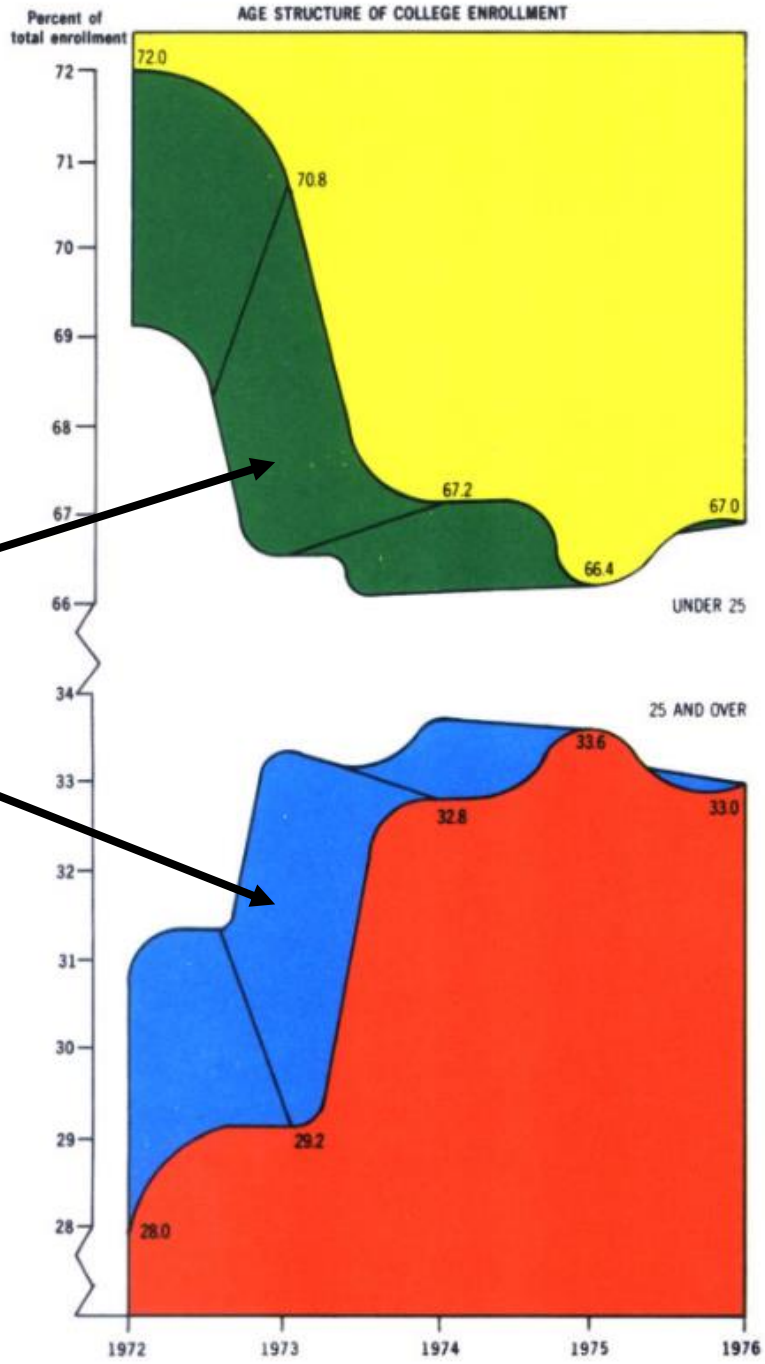


This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

   - 3rd dimension



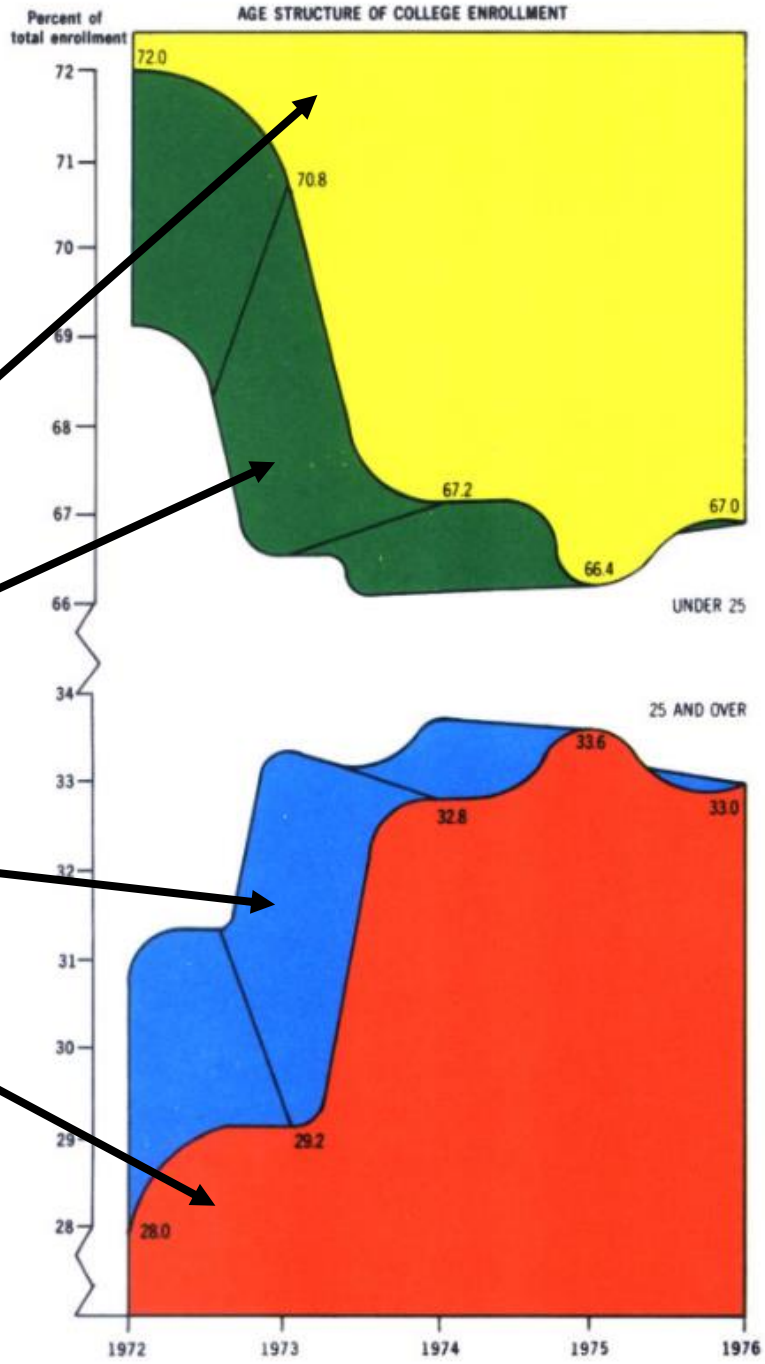
This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

   - 3rd dimension

   - 4 colors



This graph is terrible. Why?

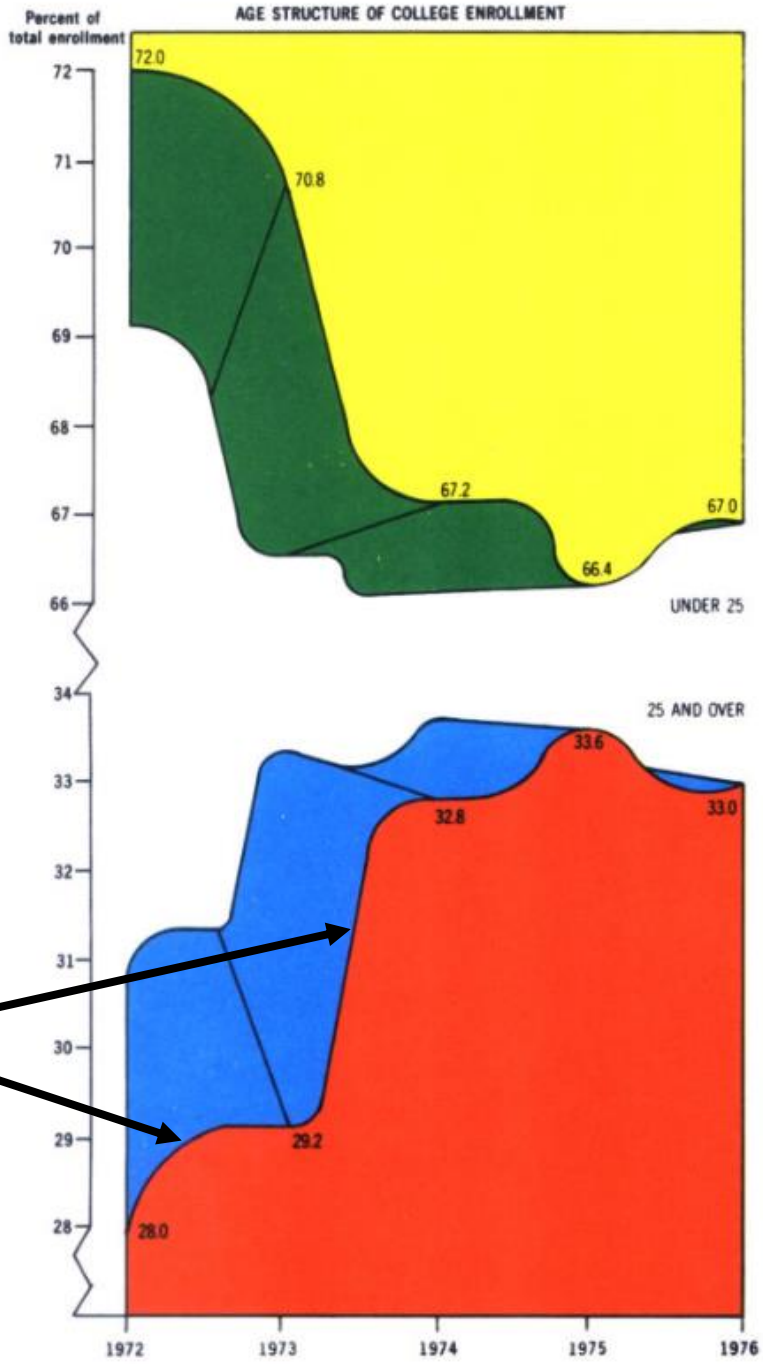
1. Information is duplicated

2. Meaningless details

   - 3rd dimension

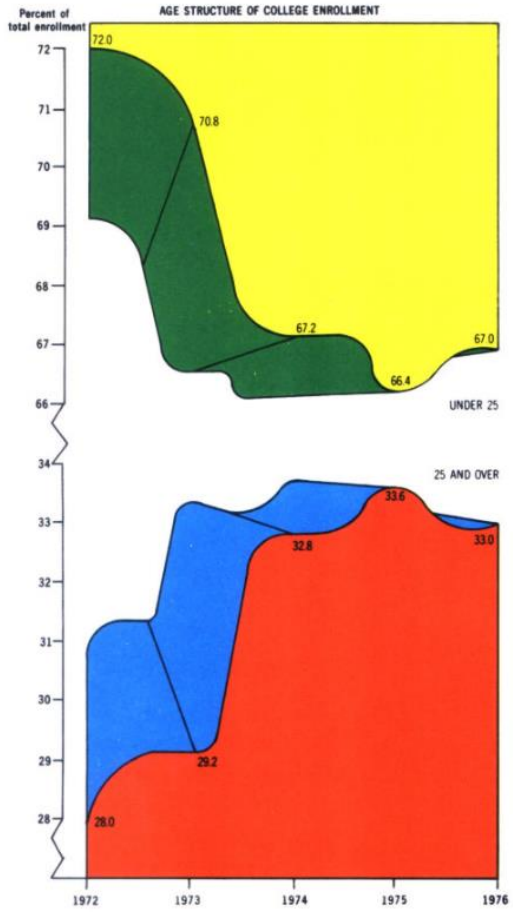
   - 4 colors

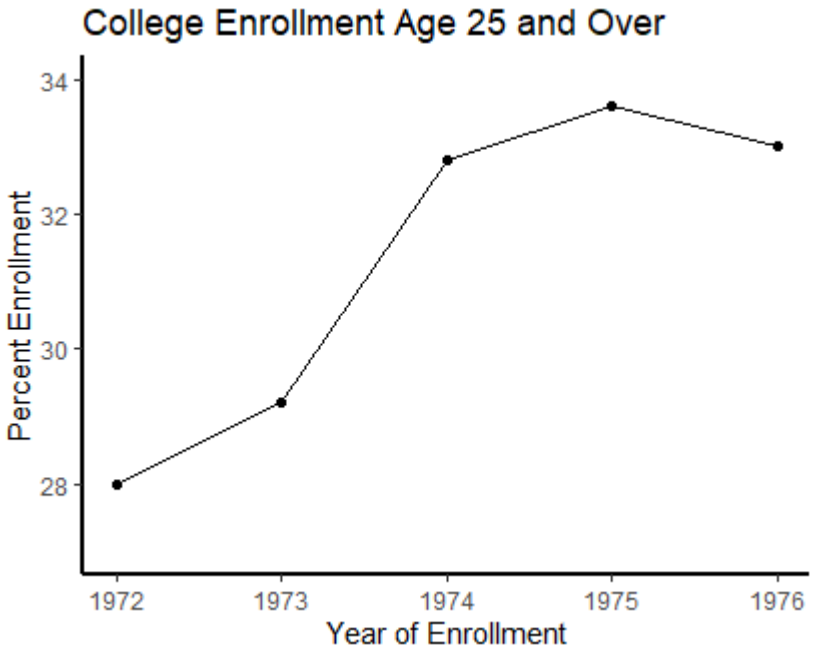
3. Misleading details (what do the curved lines mean?)



Consider: **Economy of Information**

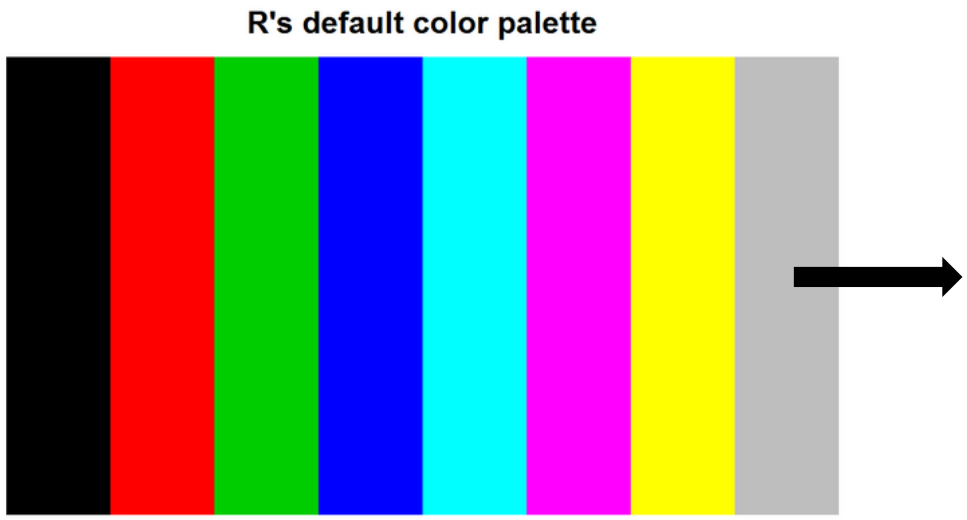
Don’t provide more detail than you need, particularly if it could undermine comprehension.

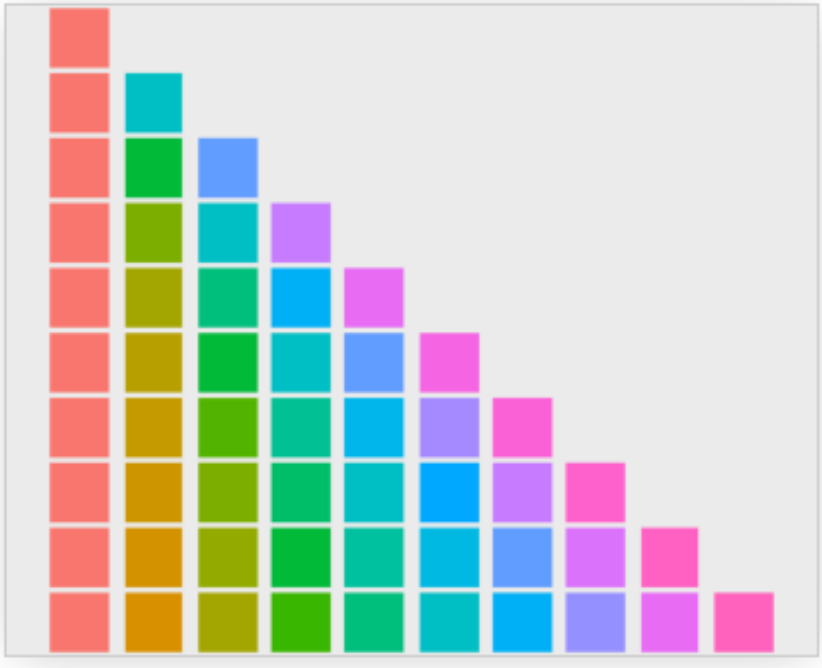




Consider: **Color Choice**

Use **softer color palettes**, choose color contrasts that compliment the nature of your data, and choose color-blind friendly palettes.

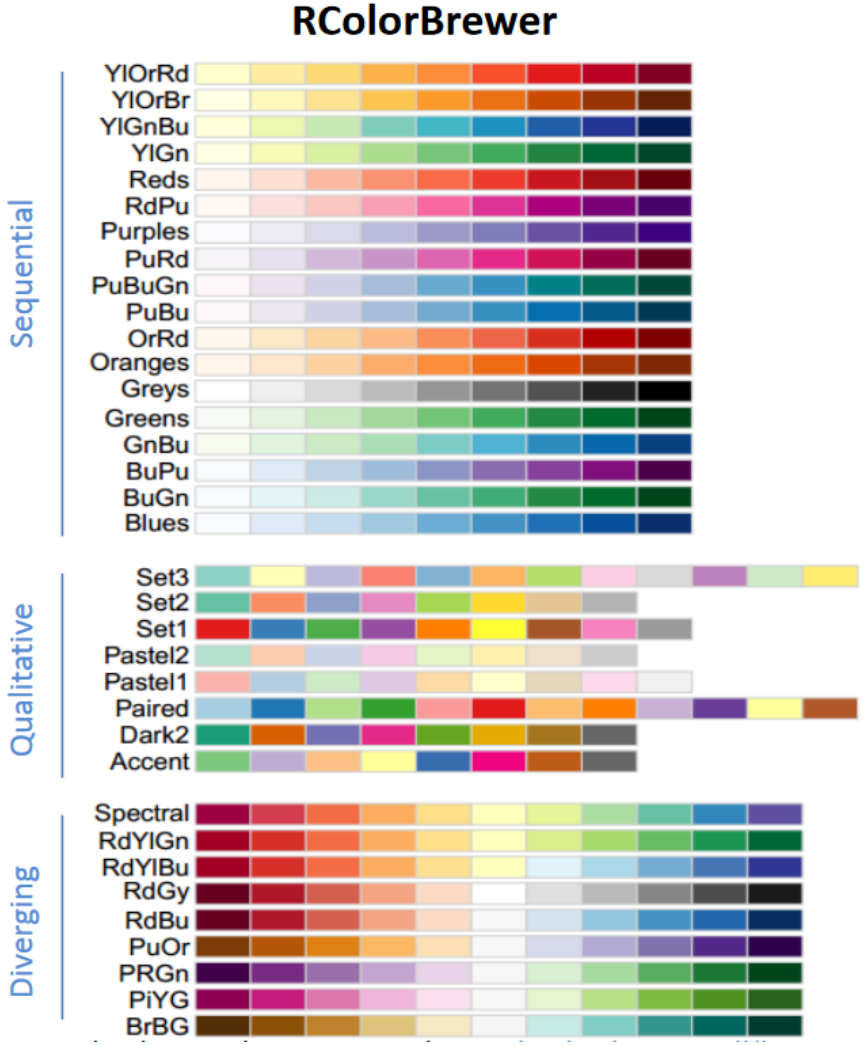


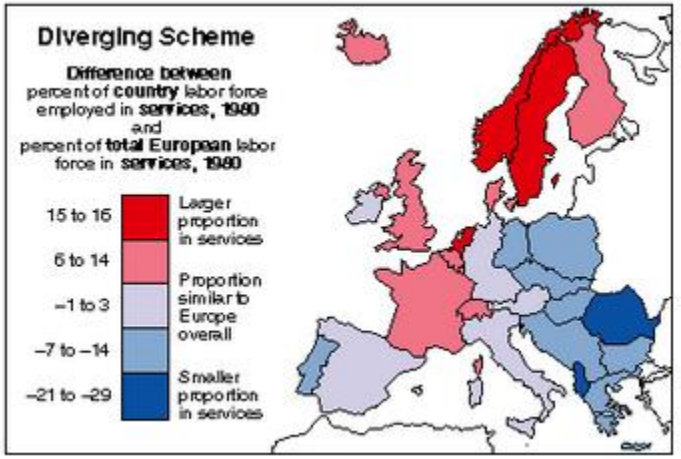


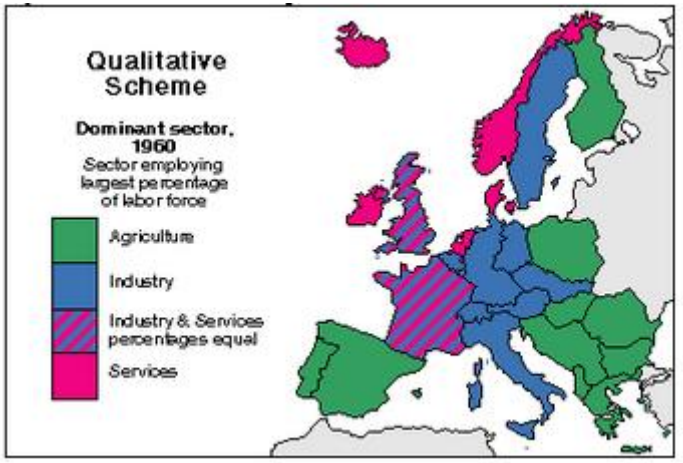
Note: ggplot already uses better looking colors by default

Consider: **Color Choice**

Use softer color palettes, **choose color contrasts that compliment the nature of your data**, and choose color-blind friendly palettes.



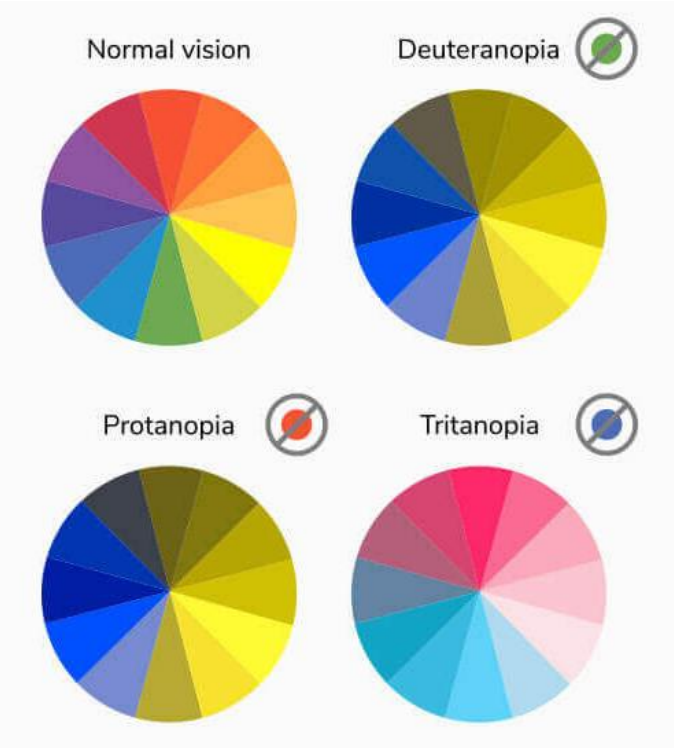


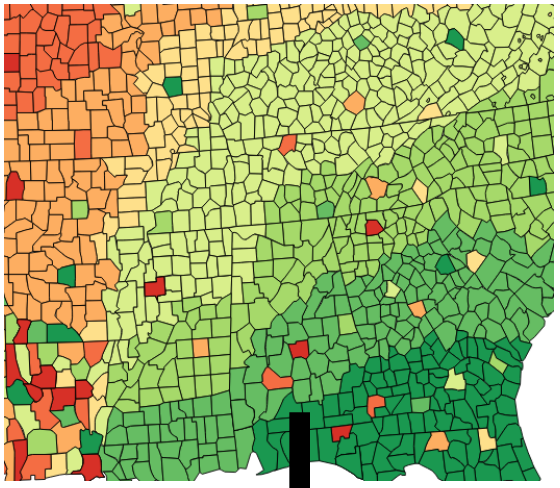


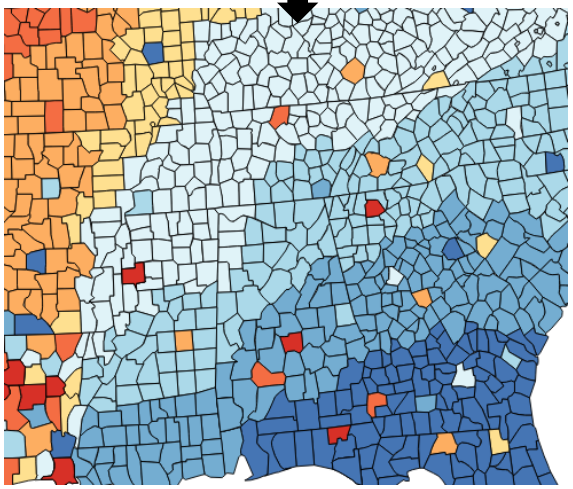
Maps from https://web.natur.cuni.cz/~langhamr/lectures/vtfg1/mapinfo_2/barvy/colors.html

Consider: **Color Choice**

Use softer color palettes, choose color contrasts that compliment the nature of your data, and **choose color-blind friendly palettes**.



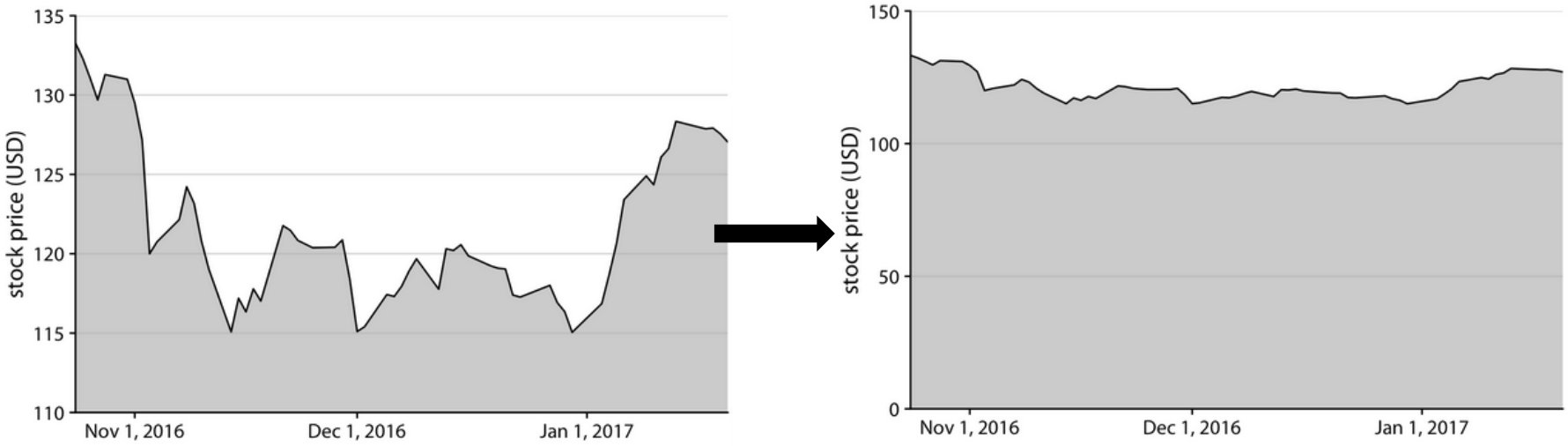




Check out **ColorBrewer2.org** for great color-blind friendly palettes.

Consider: **The Number Zero**

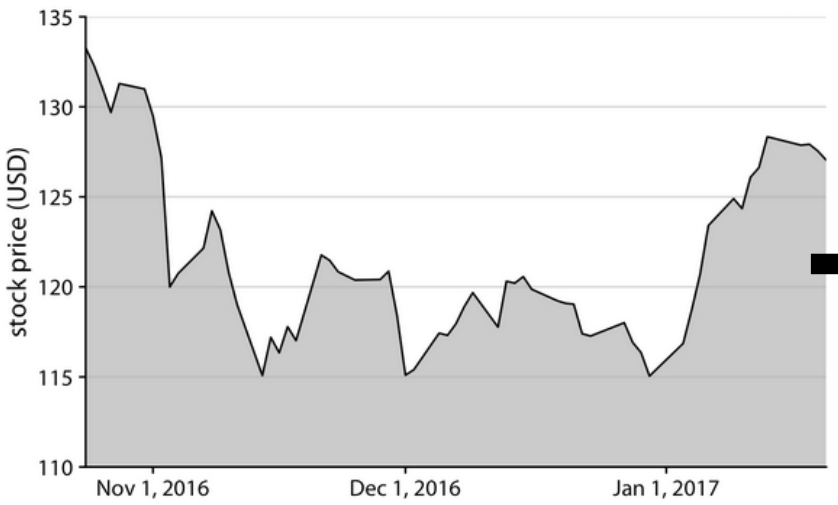
Always keep in mind that differences are distorted by your choice of axis range.

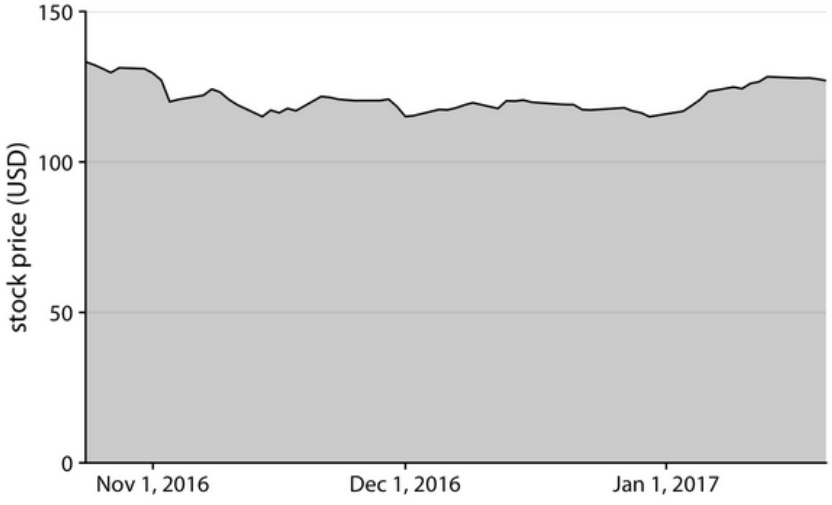


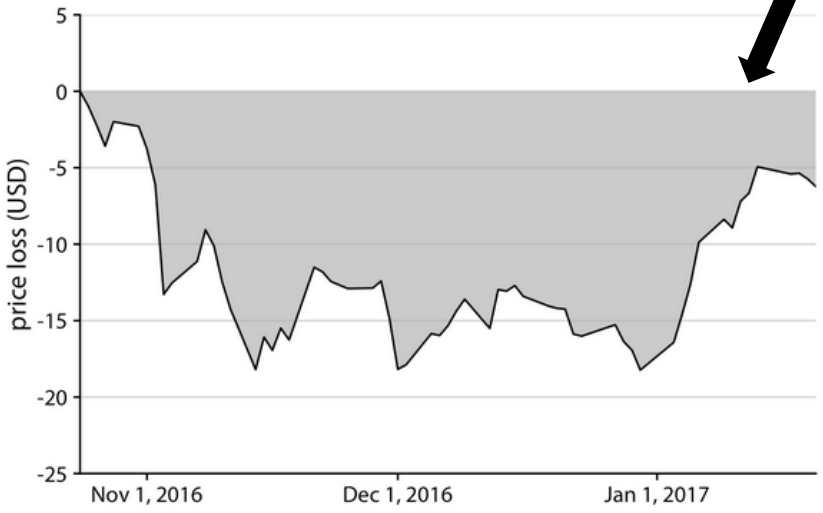
Graphs from *The Fundamentals of Data Visualization*, Claus Wilke

Consider: **The Number Zero**

Always keep in mind that differences are distorted by your choice of axis range.



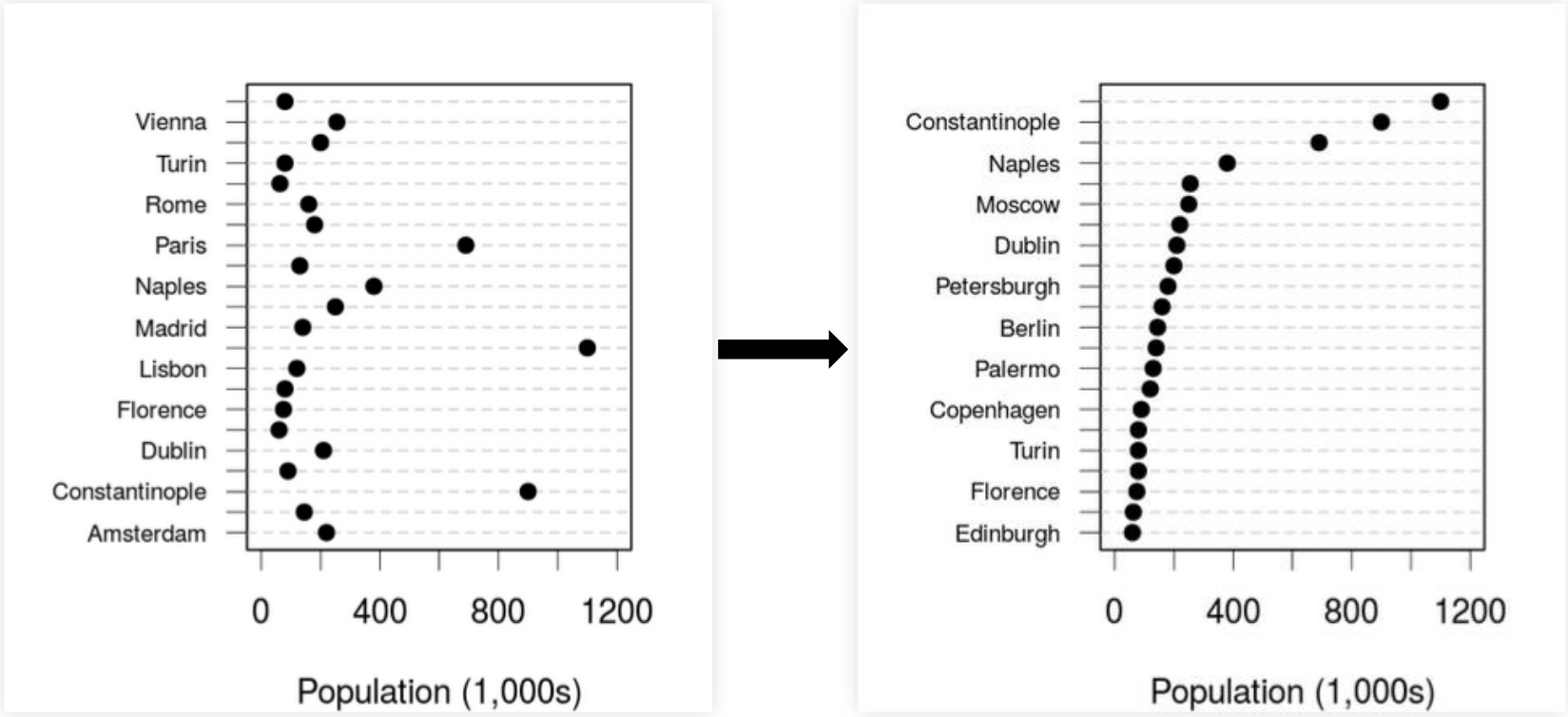




Graphs from *The Fundamentals of Data Visualization*, Claus Wilke

Consider: **Aiding Comparisons**

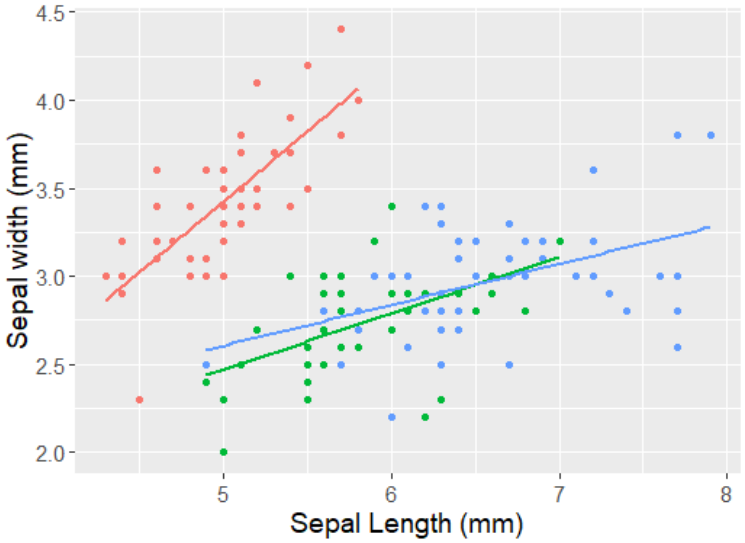
Show your data in a way that facilitates direct comparisons.



Figures courtesy of https://www.intro2r.info/unit5/principles-of-graphics.html

Consider: **Your Audience**

Your choice and size of points, fonts, etc… will depend on your audience. Is this a graph for a presentation or a publication?



Better for a paper

Better for a presentation