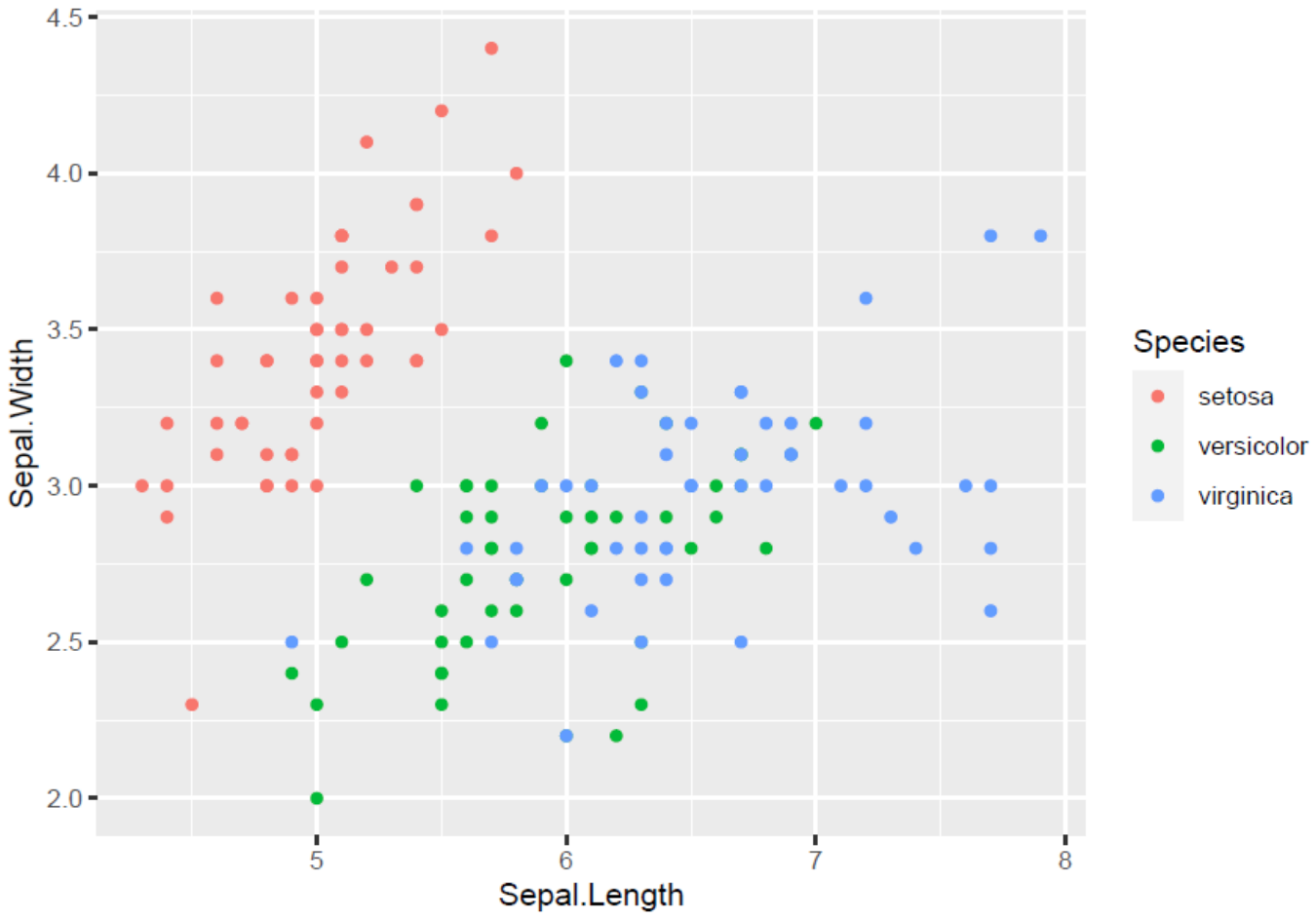
Making Better Graphs

Andrew Muehleisen and Lauren Hallett

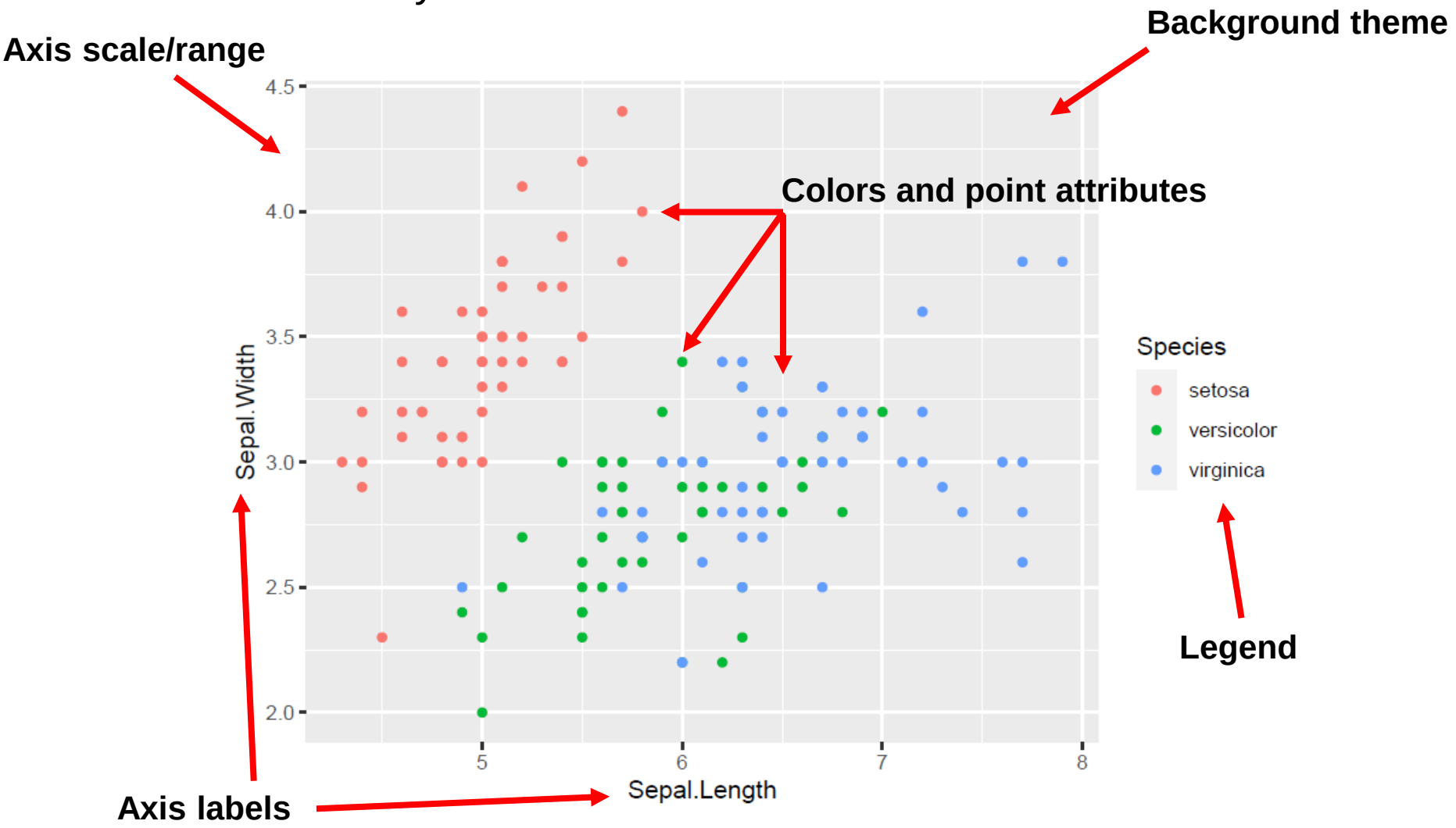
Moving on from defaults

This is the default output of **ggplot** with the ‘iris’ data set. Which elements are set by default?



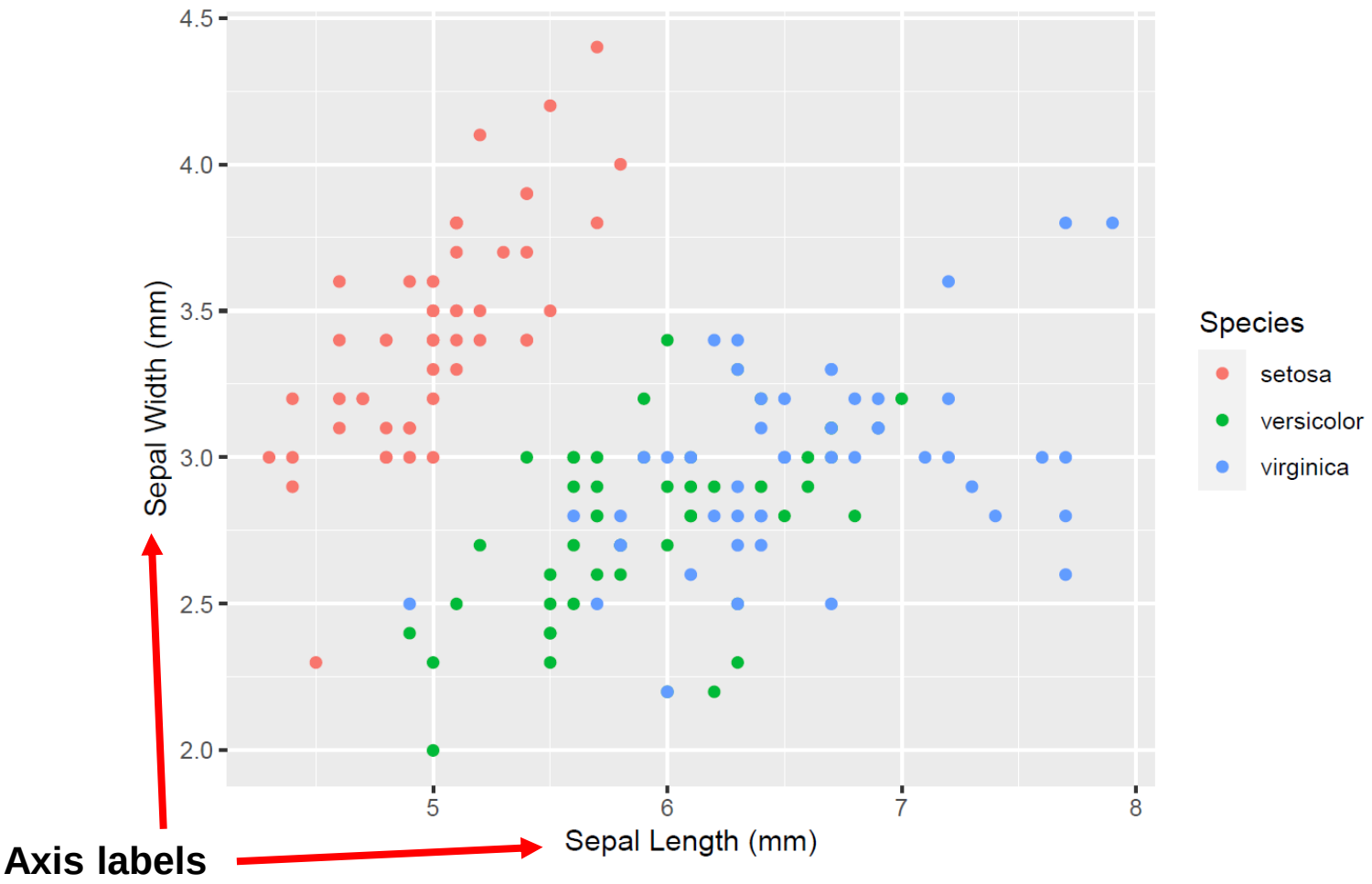
Moving on from defaults

This is the default output of **ggplot** with the 'iris' data set. Which elements are set by default?



Moving on from defaults: **labels**

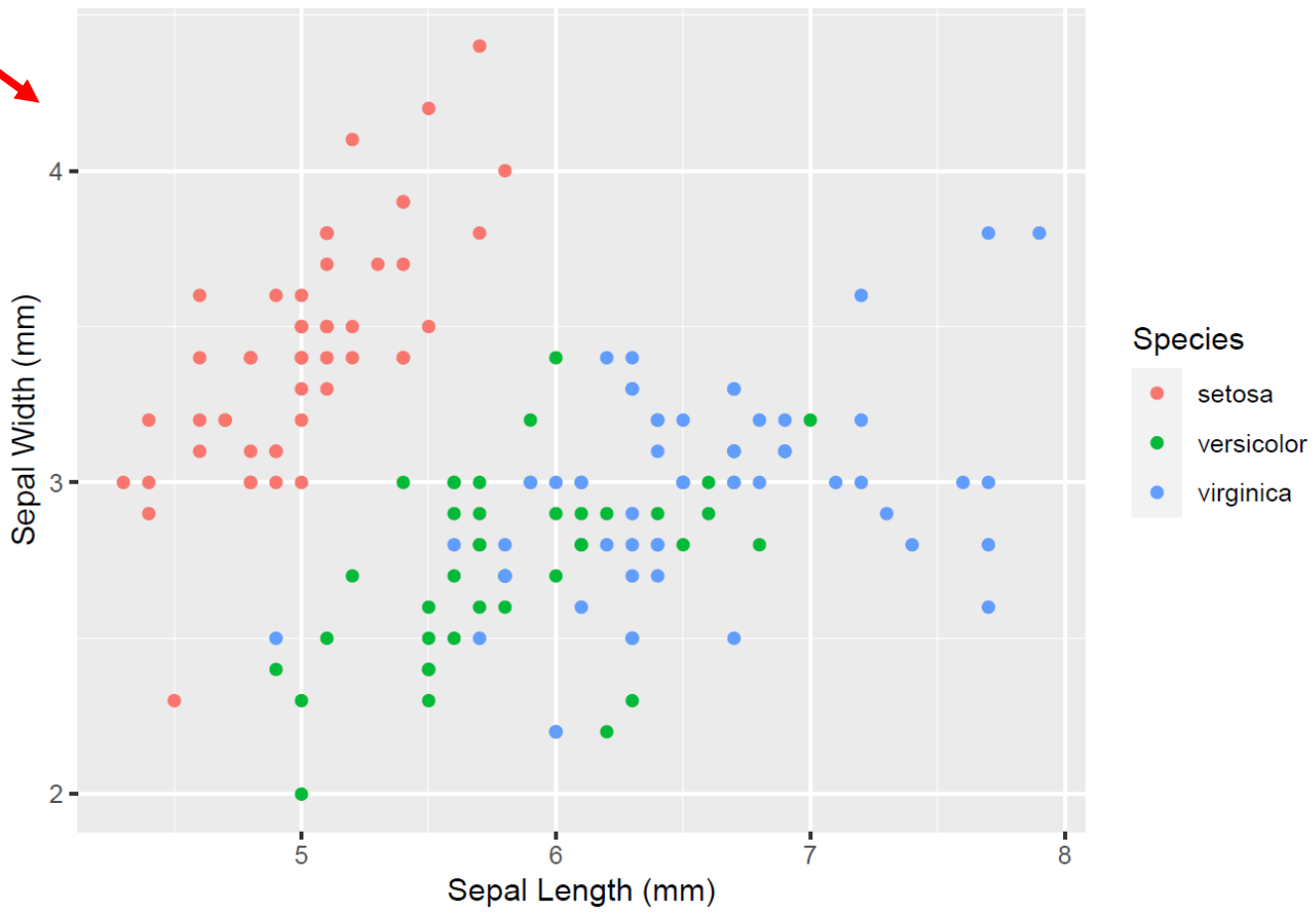
It's always a good idea to use more descriptive axis labels.



Moving on from defaults: **axis properties**

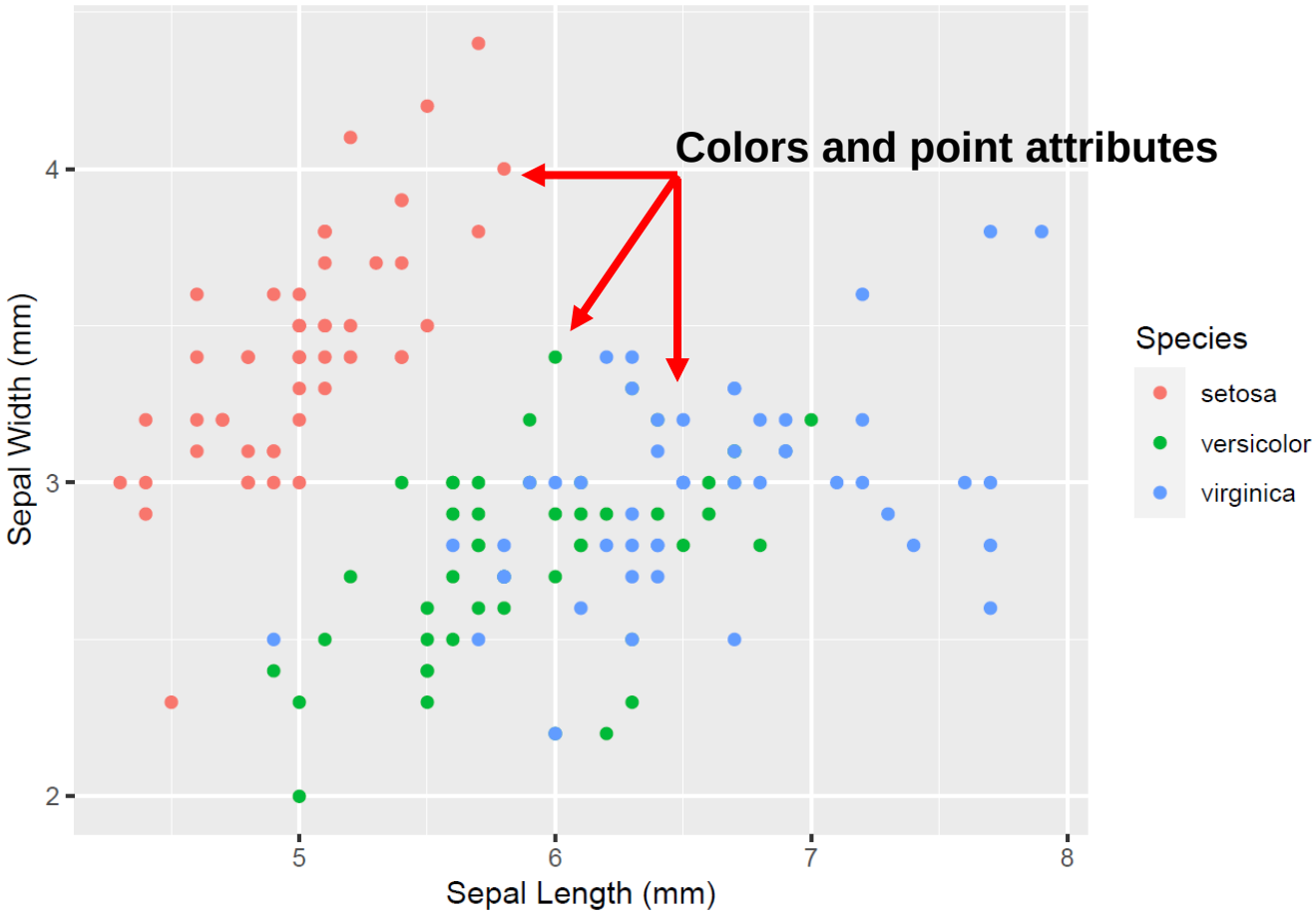
We can also change the scale and tick breaks on our axes. Here we change the y-axis to match the x-axis (both integers).

Axis scale/range



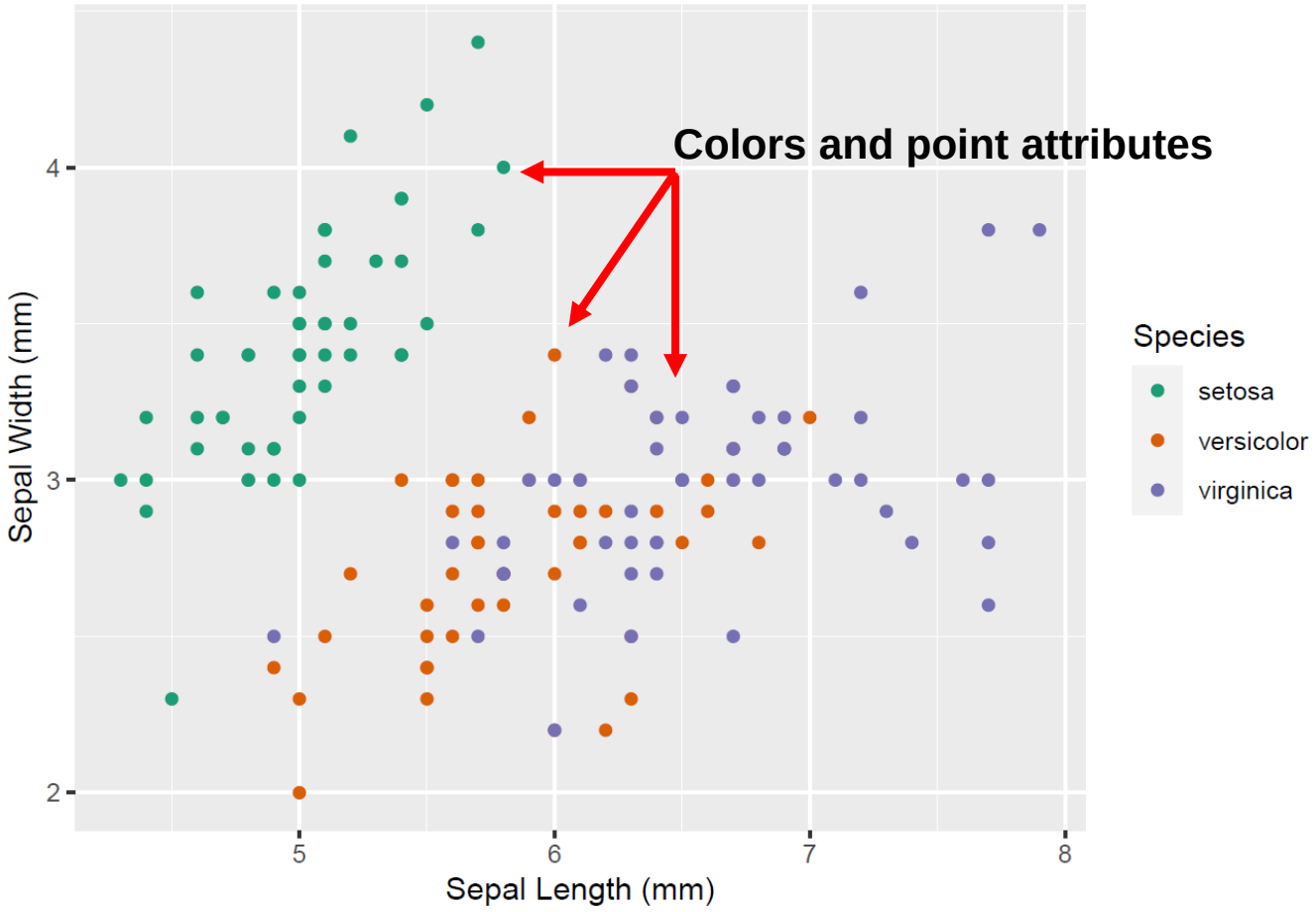
Moving on from defaults: **color**

By default, ggplot chooses colors based around the color wheel.



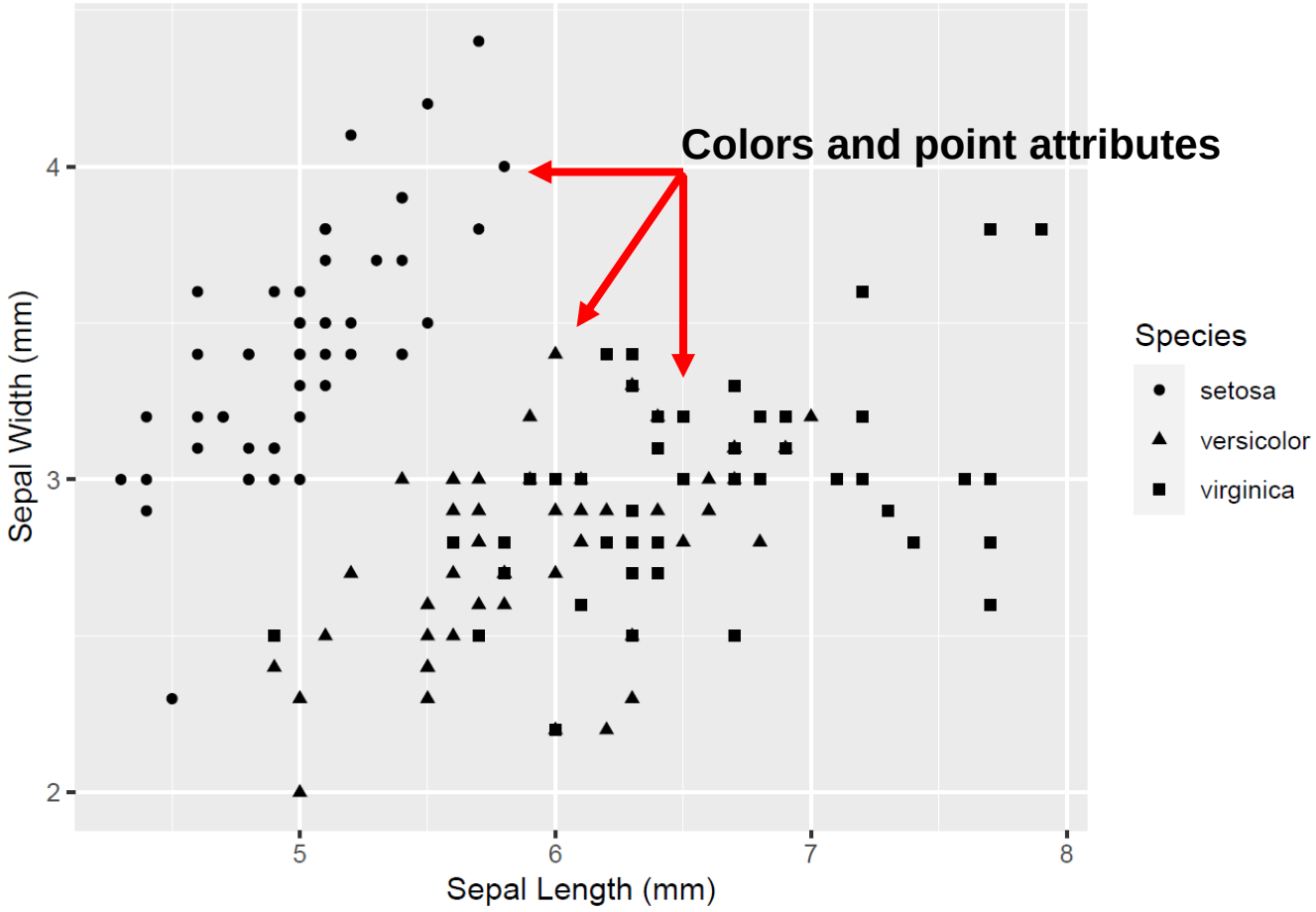
Moving on from defaults: **color**

By default, ggplot chooses colors based around the color wheel. We could instead e.g. use more color-blind friendly colors.



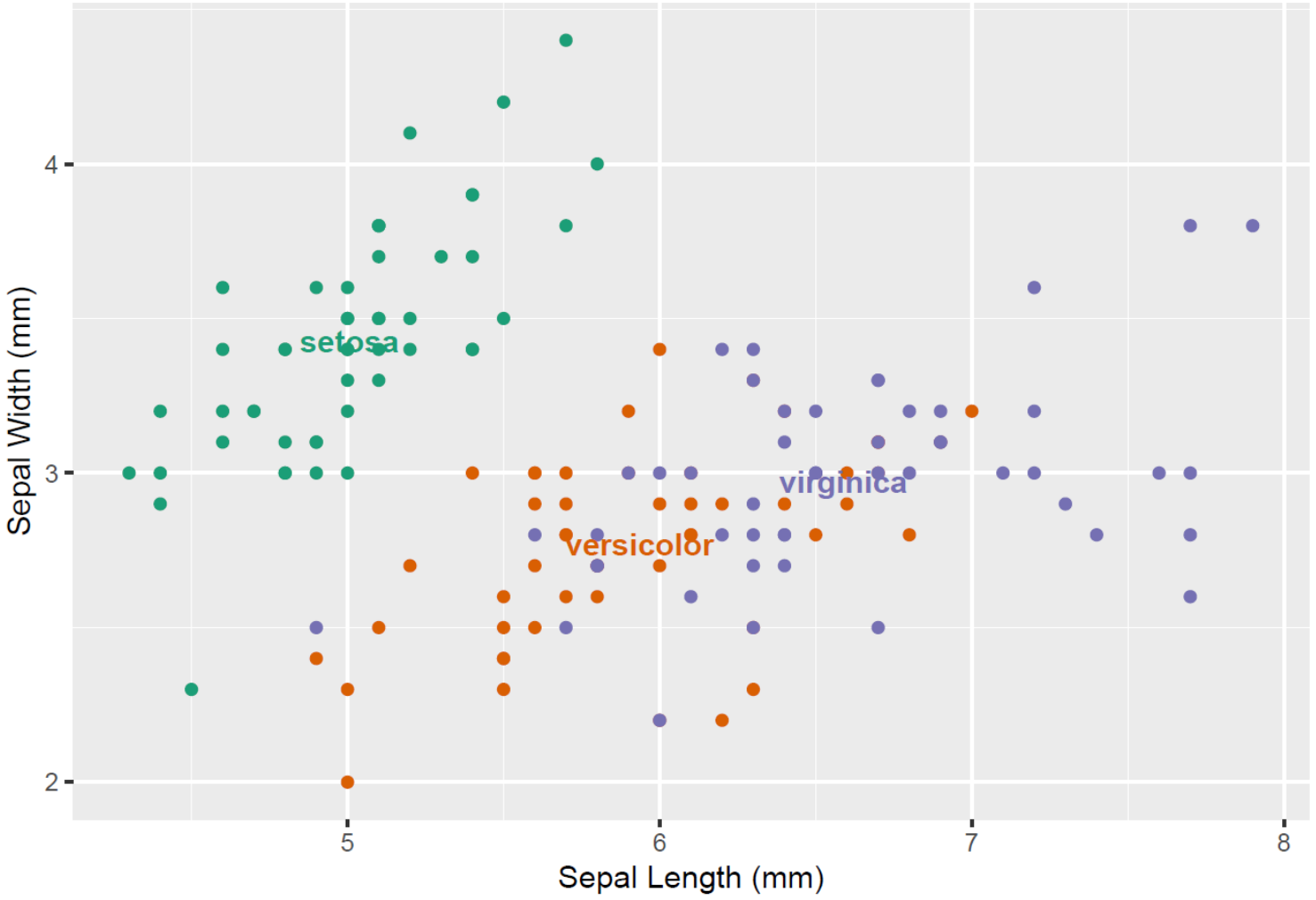
Moving on from defaults: **point type**

While color is a common way to differentiate groups, we could decide that point type does a better job.



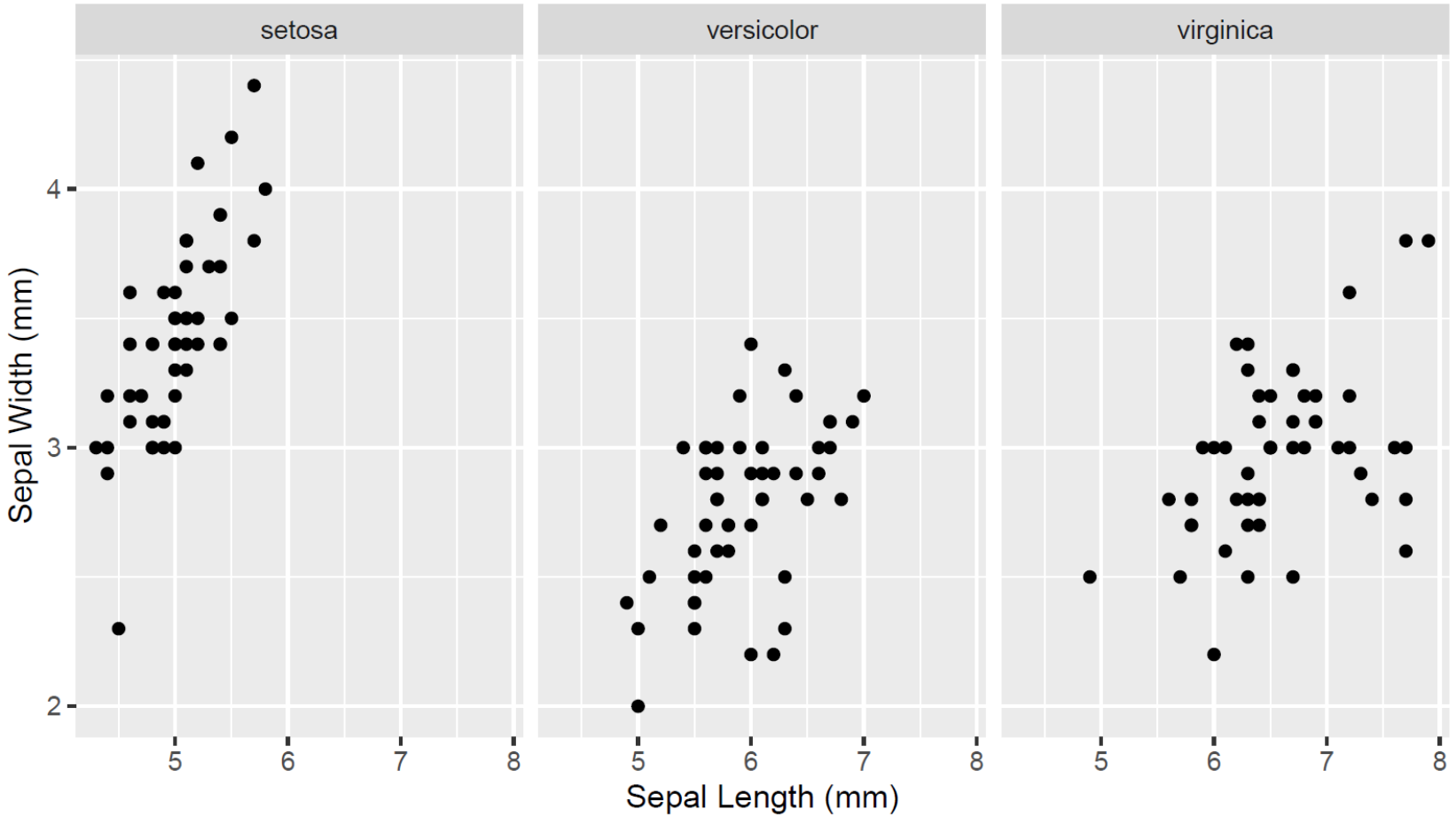
Moving on from defaults: **legends**

There are other options than legends. For example, we could decide to label our points directly in the figure.



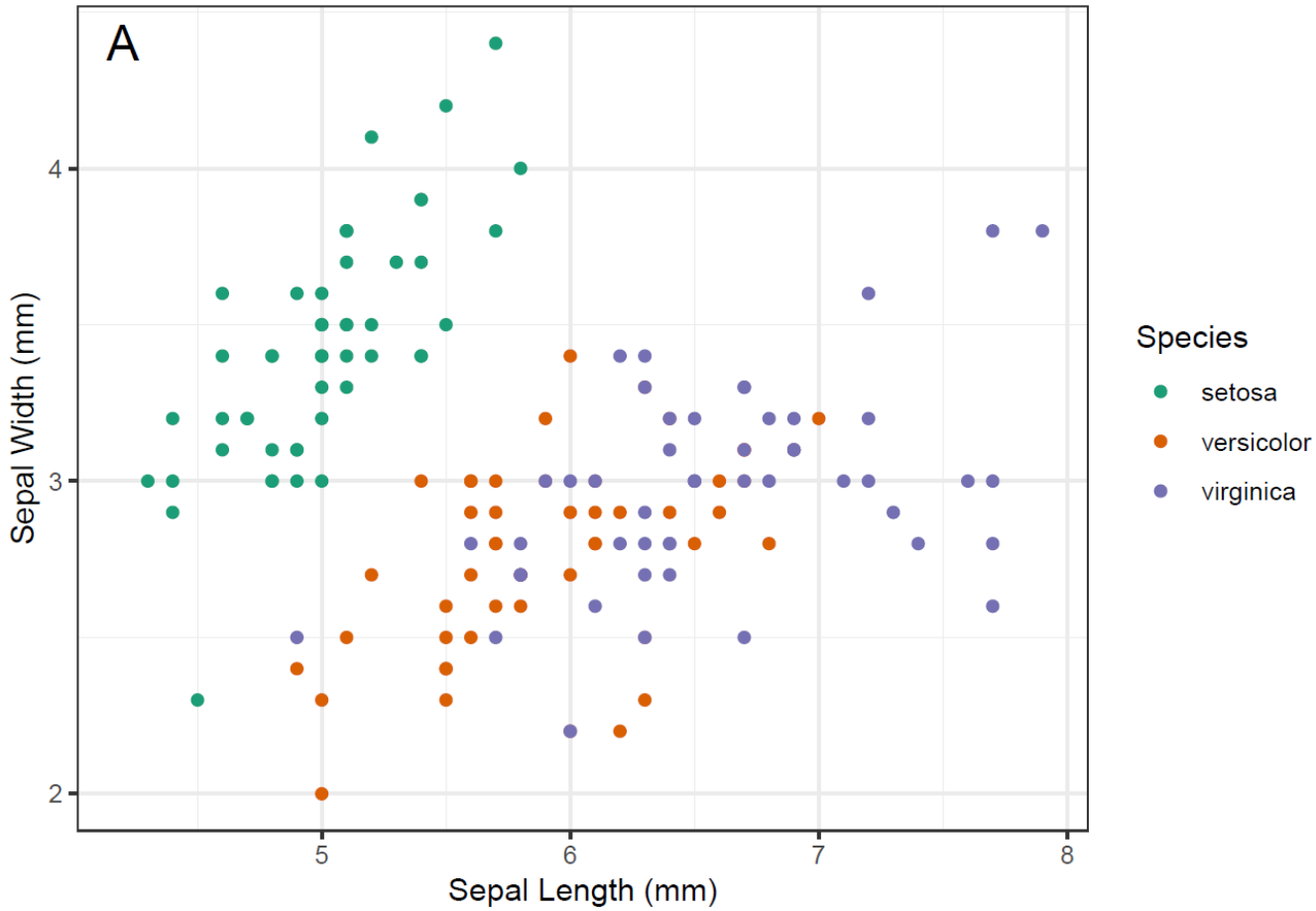
Moving on from defaults: **multiple panels**

We could even separate each species into its own panel.



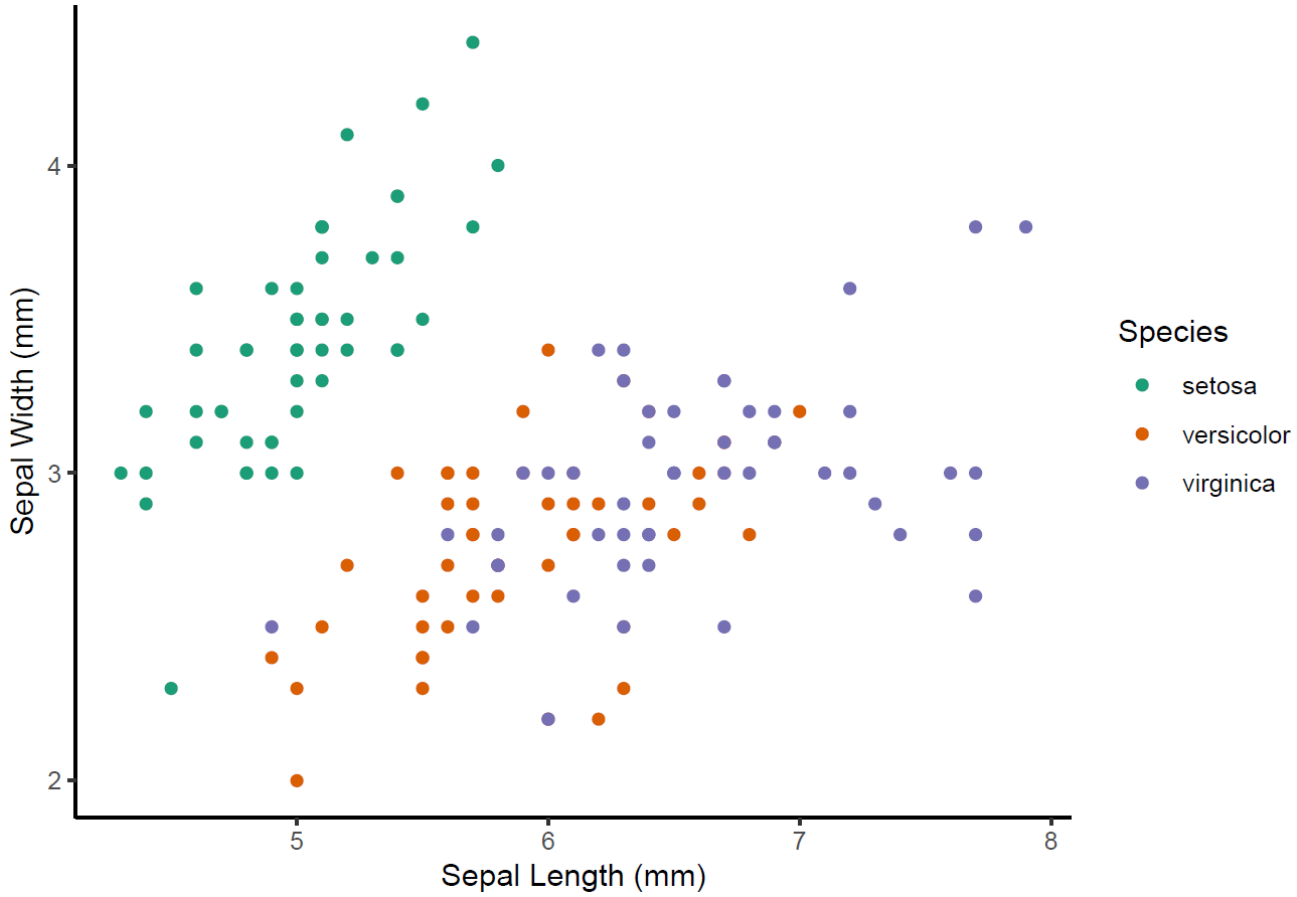
Moving on from defaults: **background themes**

Finally, we could use an entirely new theme. ‘theme_bw’ is a commonly used alternative theme for **ggplot**



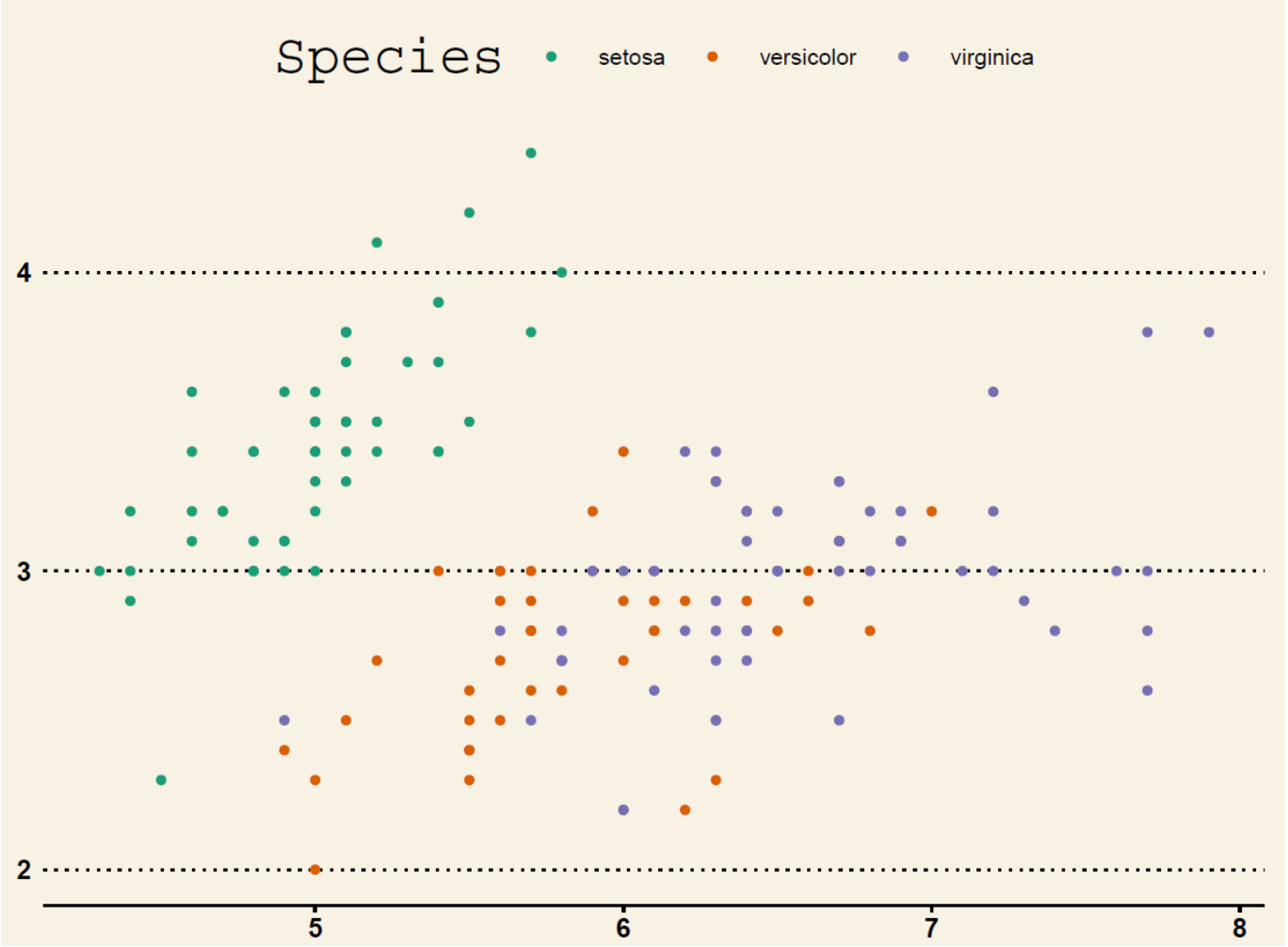
Moving on from defaults: **background themes**

If you don’t like gridlines or the full box, there’s always ‘theme_classic’, which approximates base R figures.



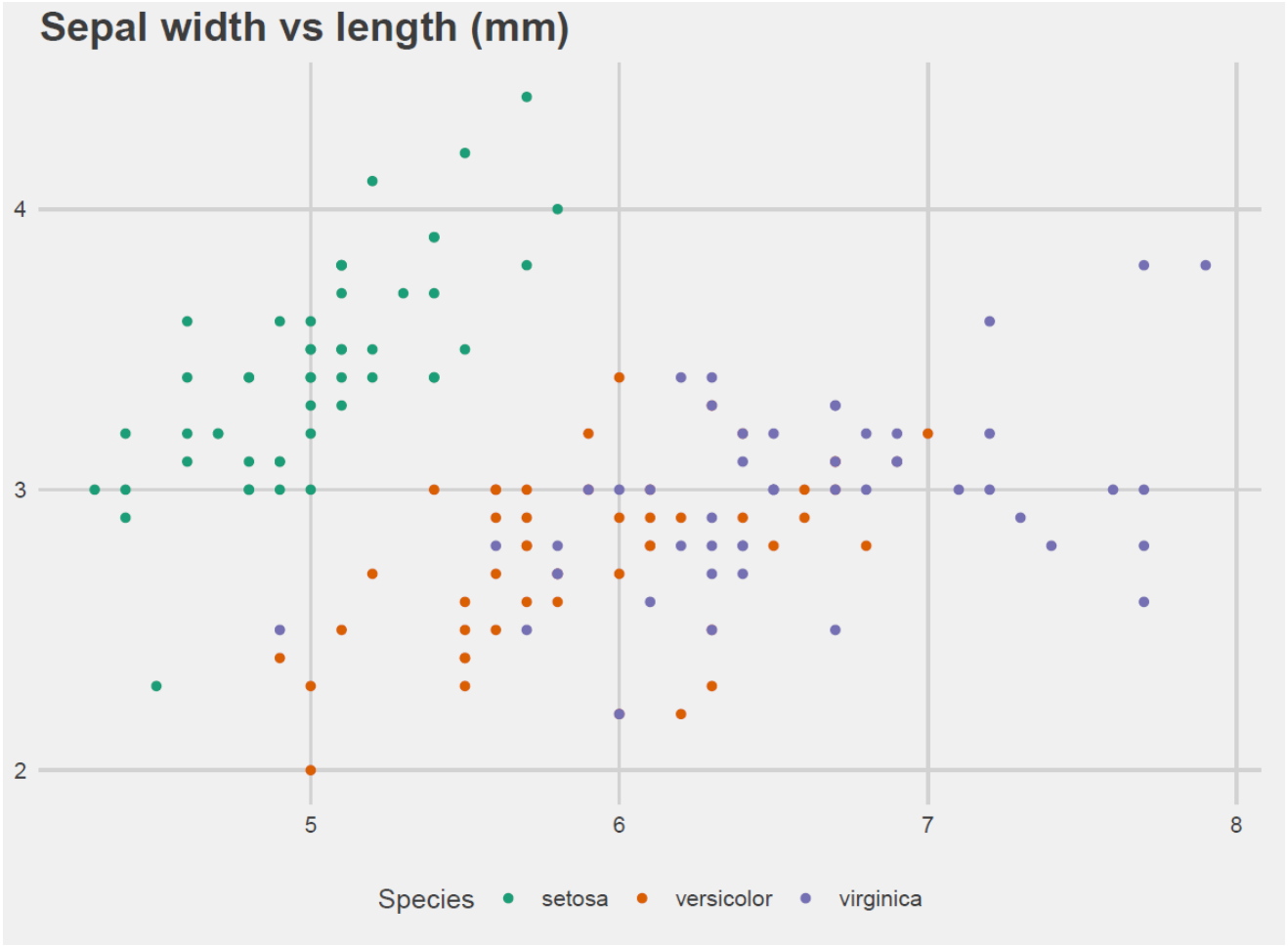
Moving on from defaults: **background themes**

Maybe you just love the Wall Street Journal: ‘theme_wsj’



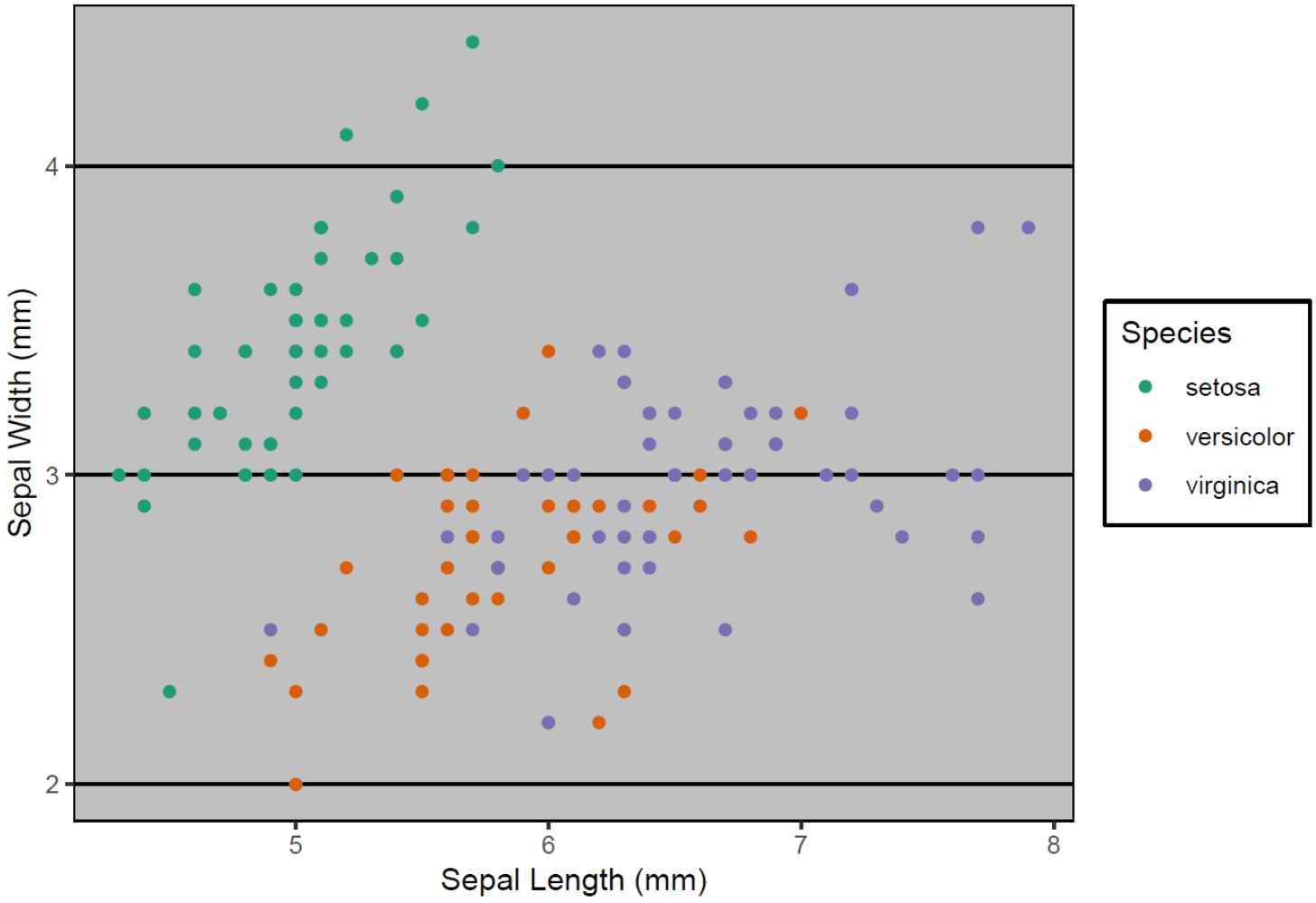
Moving on from defaults: **background themes**

Or 538: ‘theme_fivethirtyeight’



Moving on from defaults: **background themes**

Or maybe you really miss excel: ‘theme_excel’



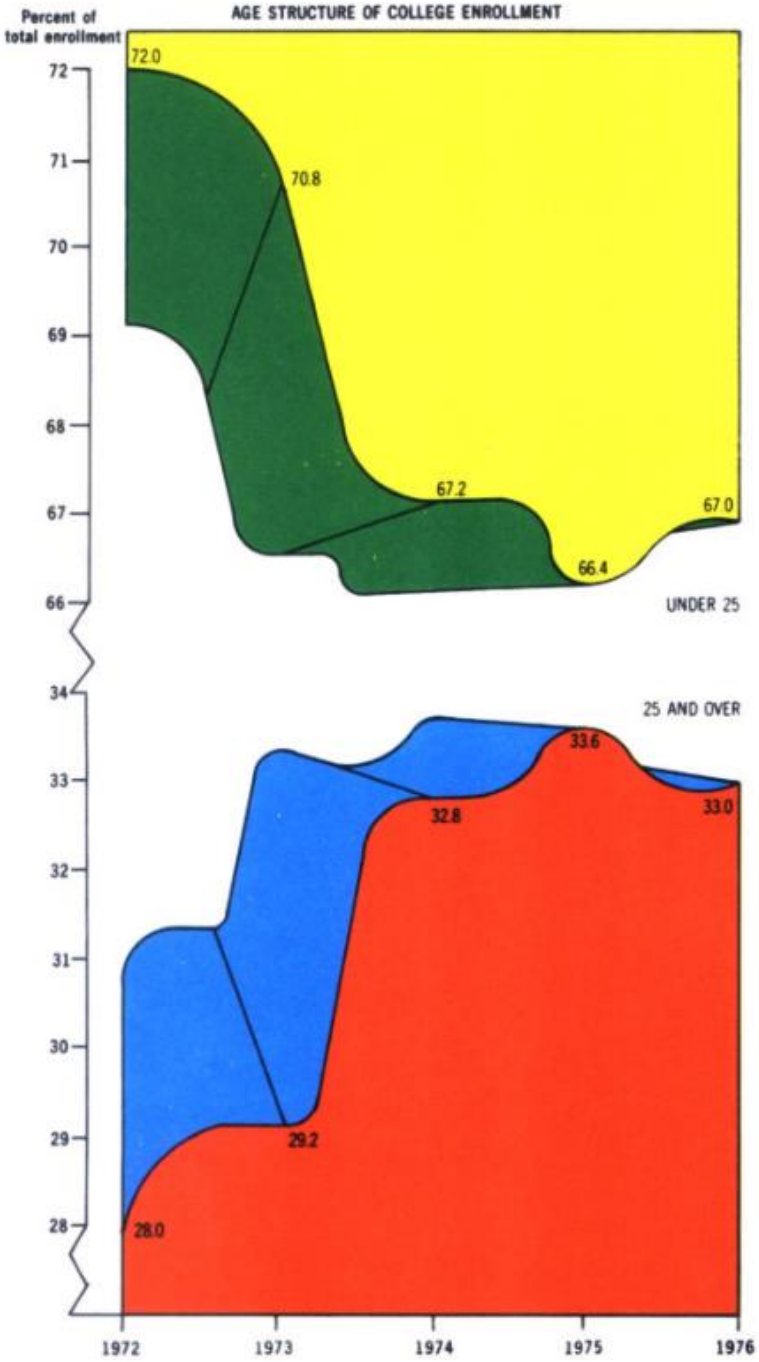
Moving on from the defaults

There are many ways to move beyond plot defaults:

- 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

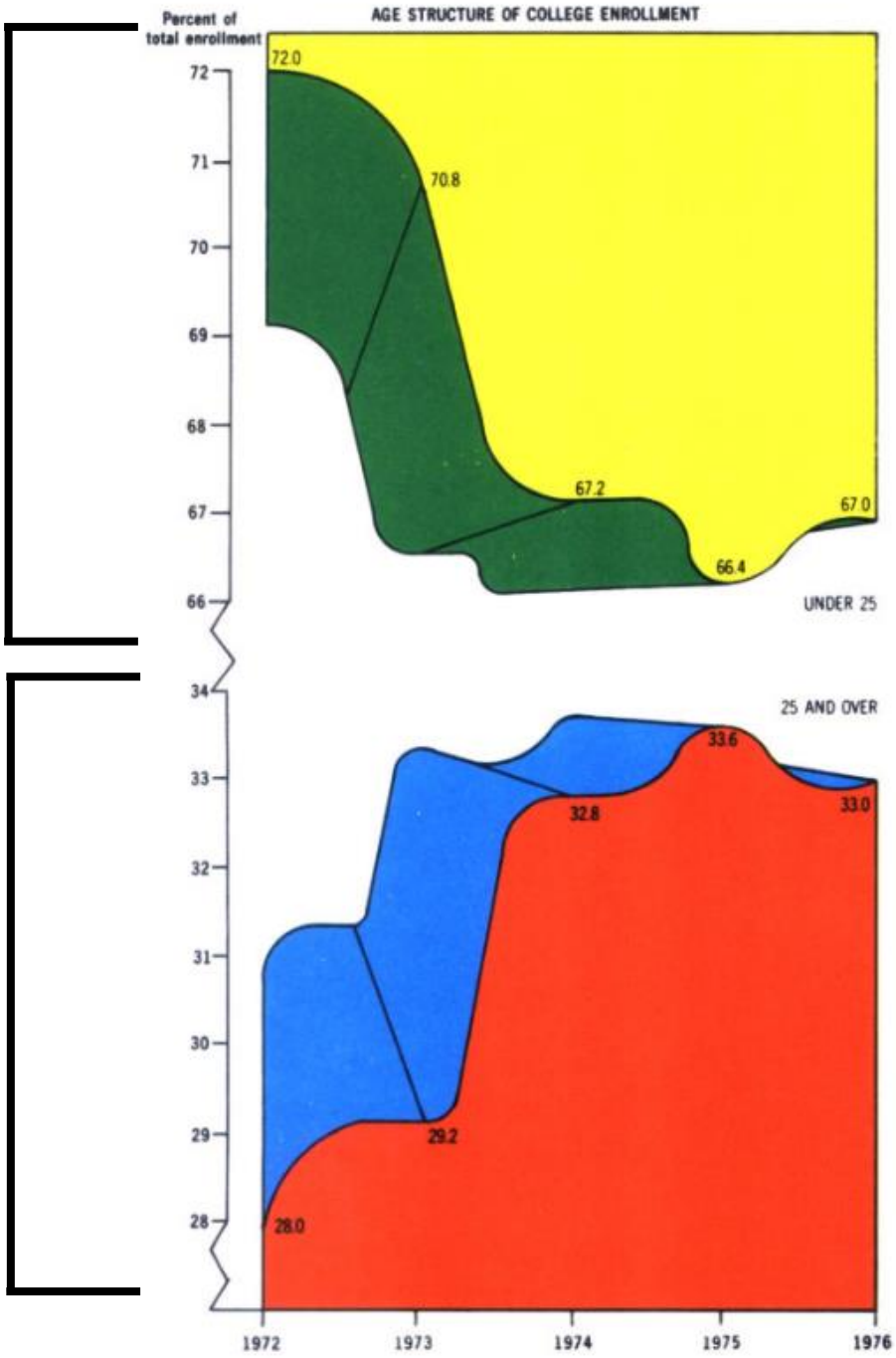
How you deviate from R's defaults is ultimately up to you. However, it is wise to consider some principles of good graphics…

This graph is terrible. Why?



This graph is terrible. Why?

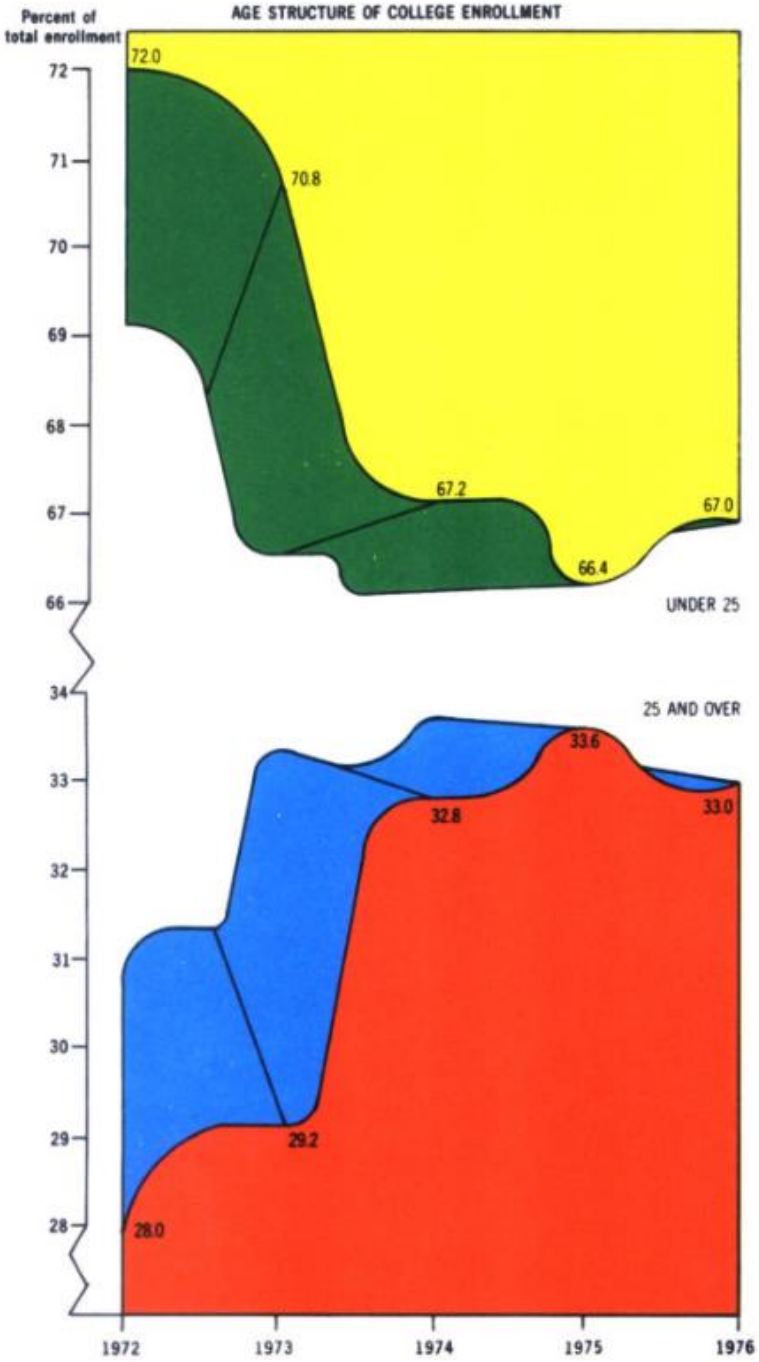
1. Information is duplicated



This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

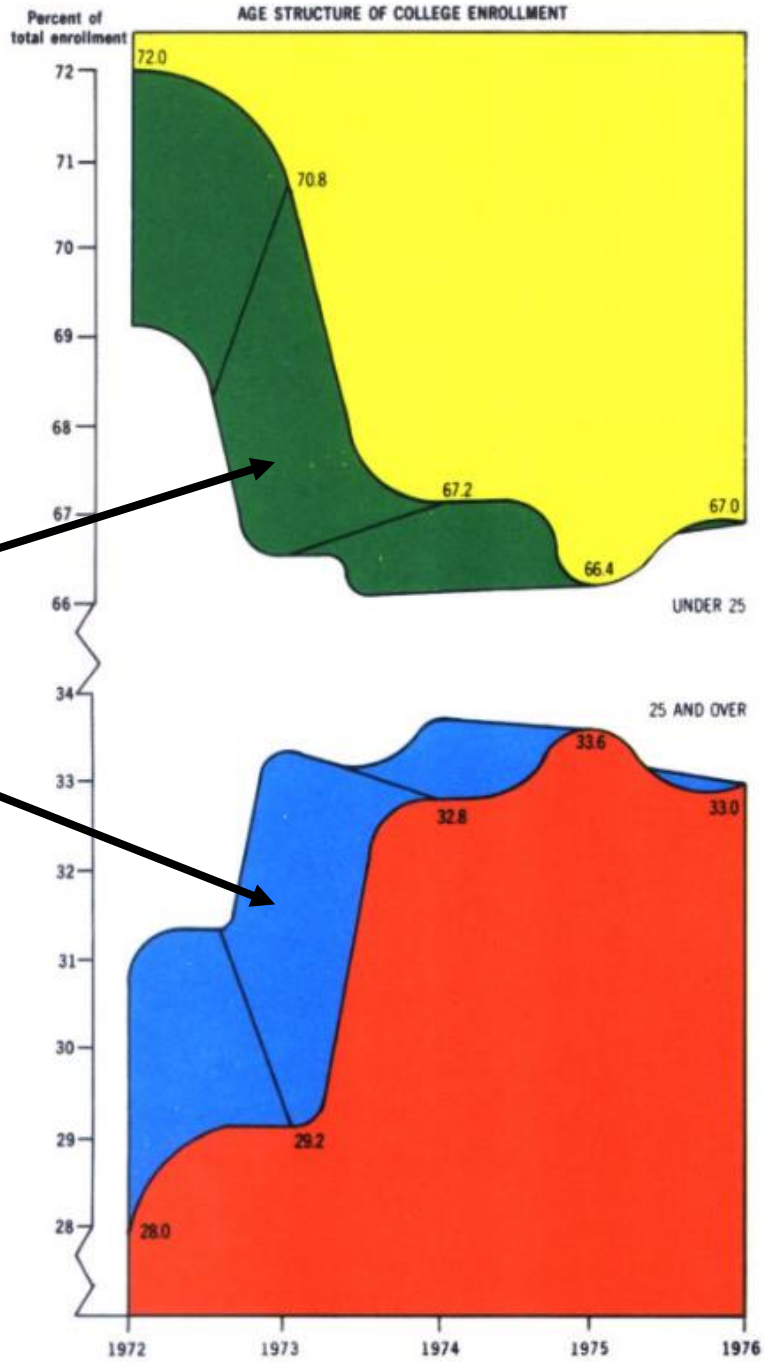


This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

   - 3rd dimension



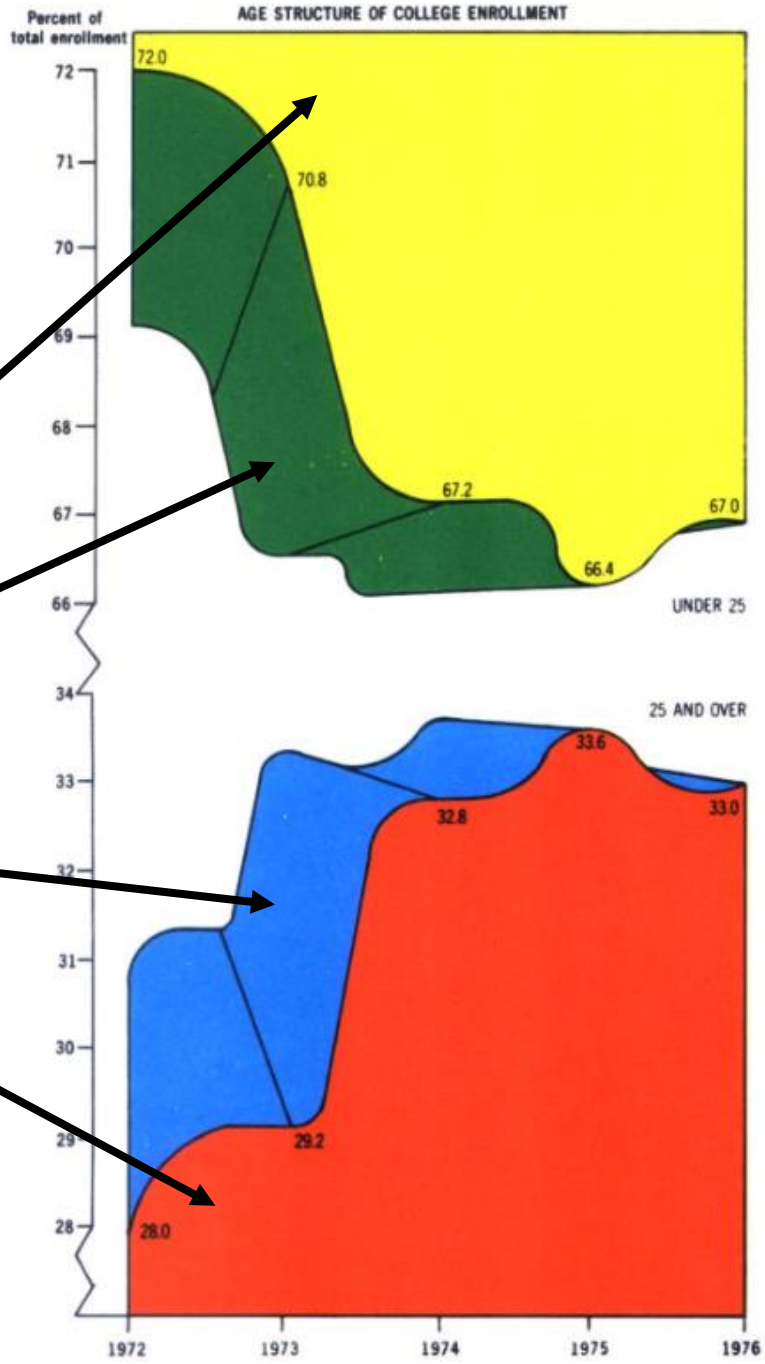
This graph is terrible. Why?

1. Information is duplicated

2. Meaningless details

   - 3rd dimension

   - 4 colors



This graph is terrible. Why?

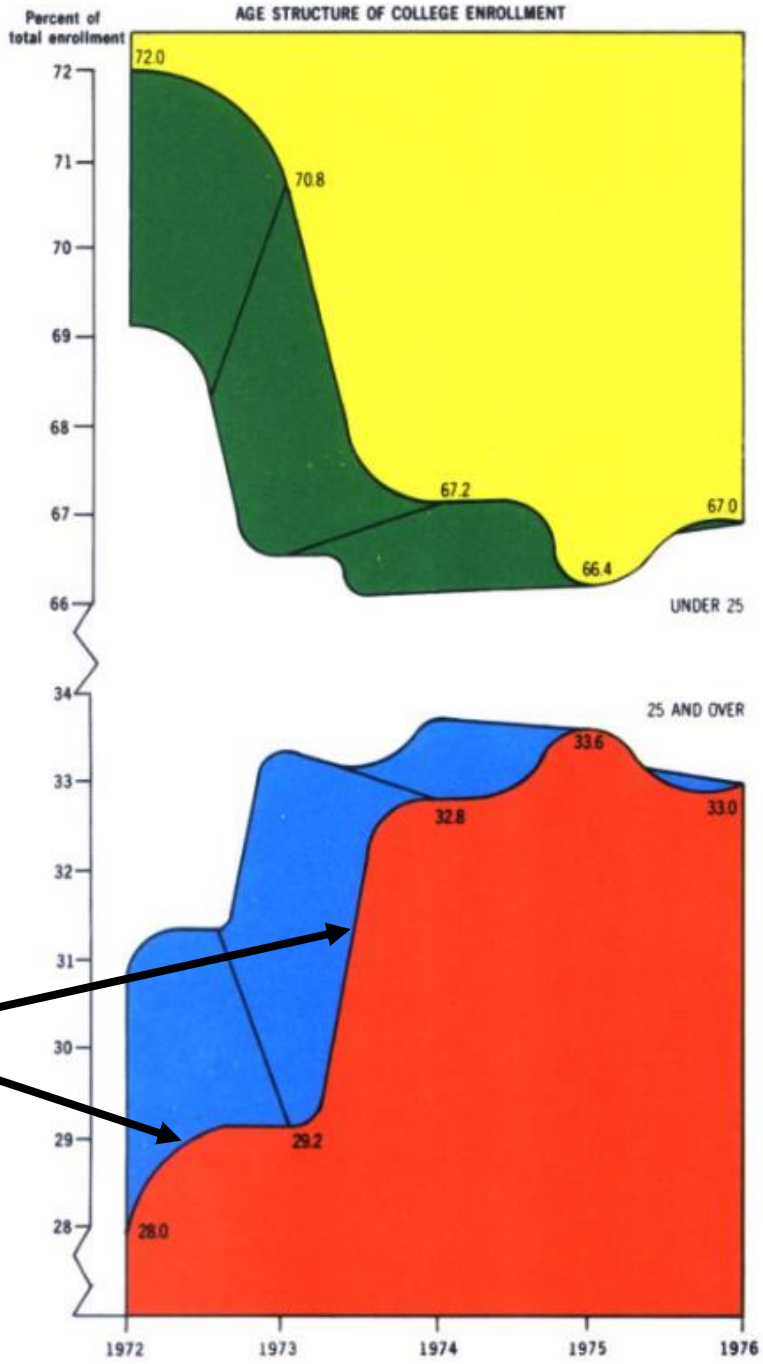
1. Information is duplicated

2. Meaningless details

   - 3rd dimension

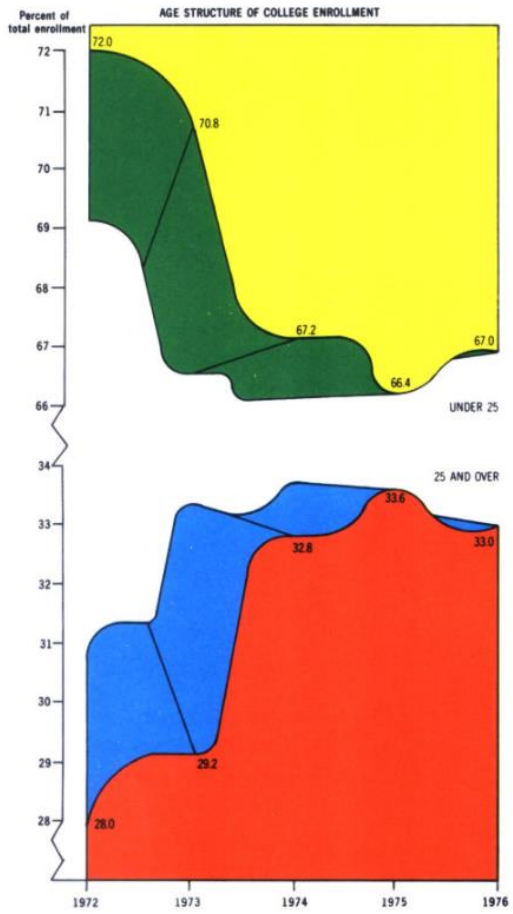
   - 4 colors

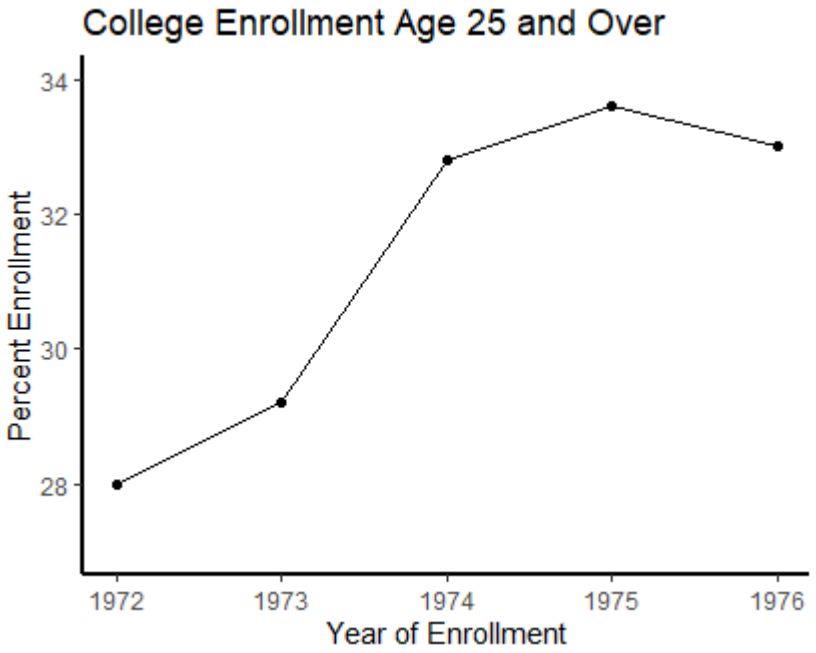
3. Misleading details (what do the curved lines mean?)



Consider: **economy of information**

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

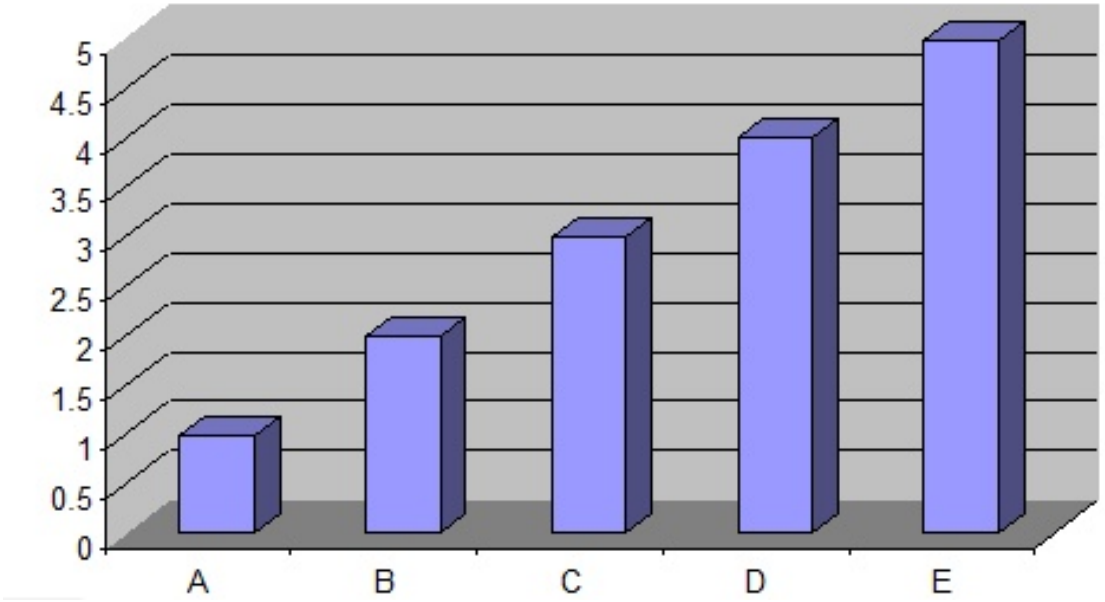




Consider: **avoiding 3-dimensions**

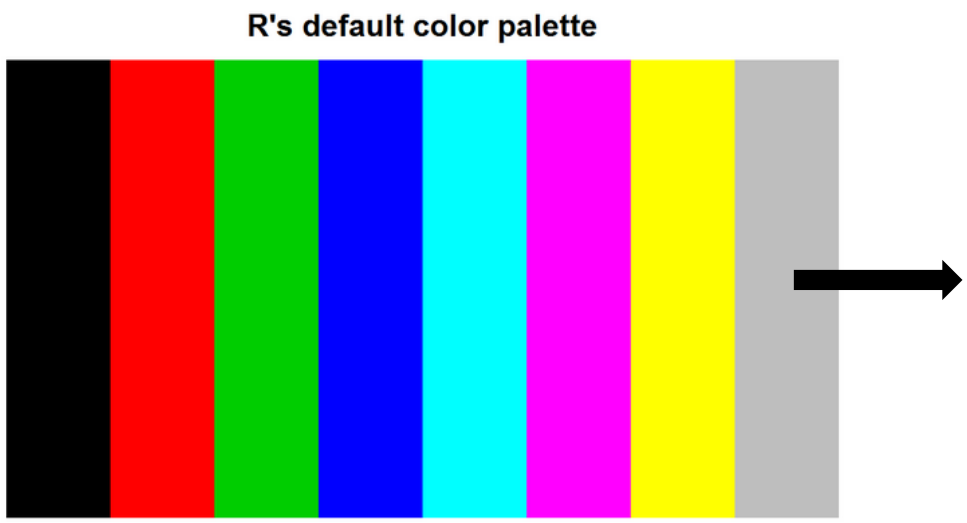
Unless 3 dimensions are serving a specific purpose, like representing an interaction between two variables, keep things 2-dimensional for clarity.

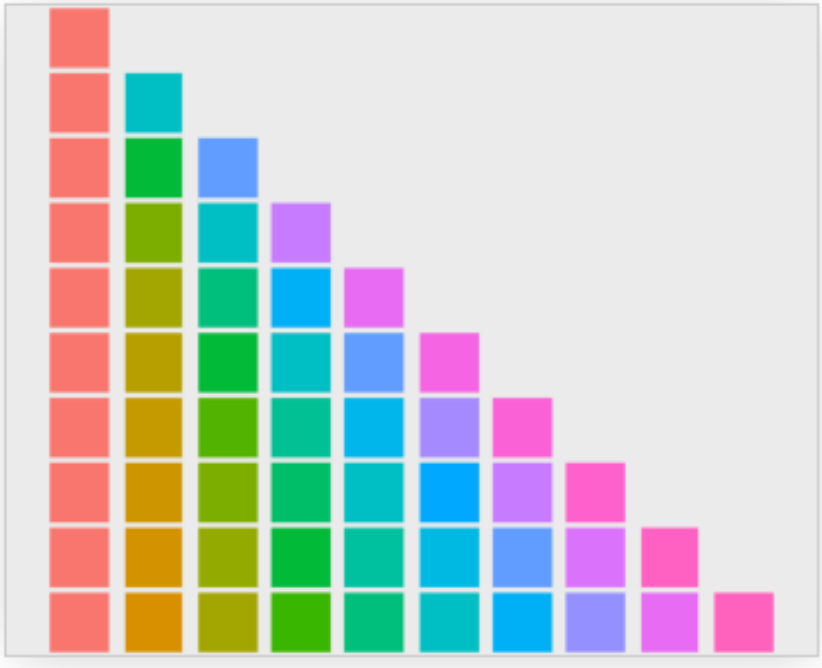
What are the values represented by these bars?



Consider: **color choice**

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

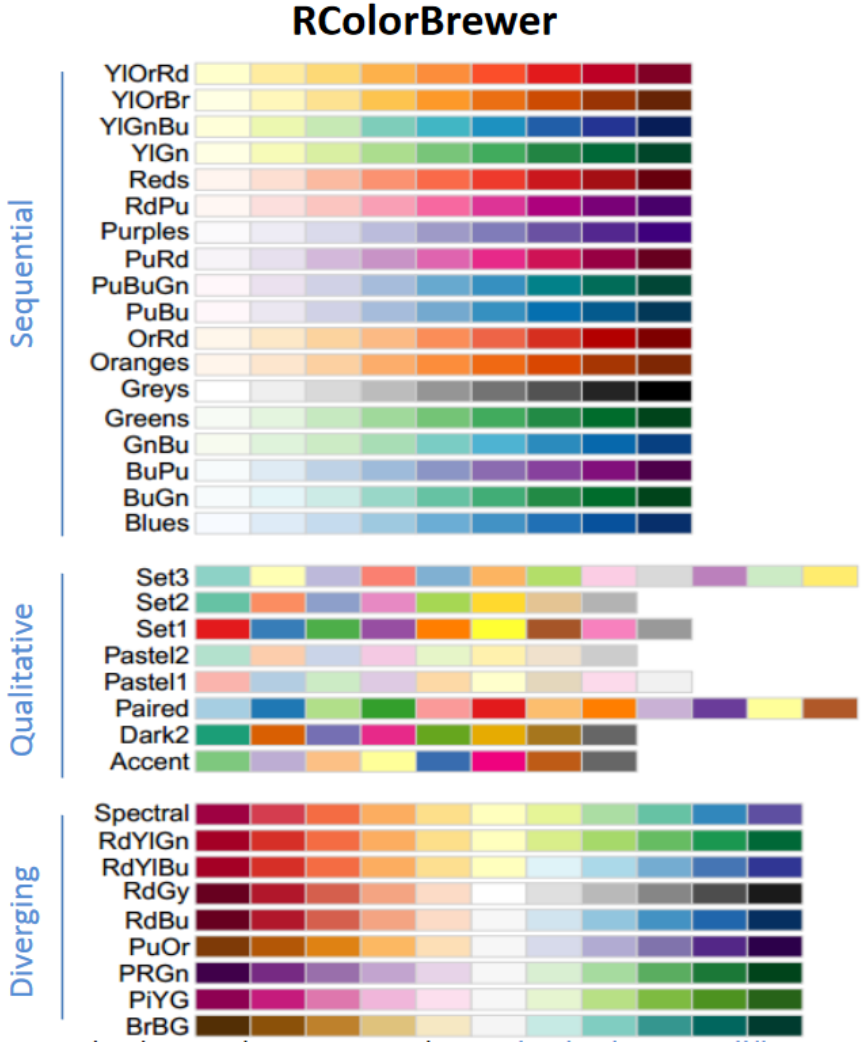


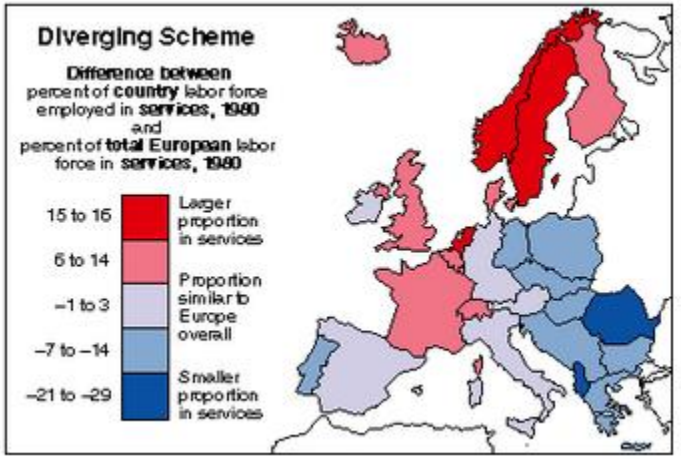


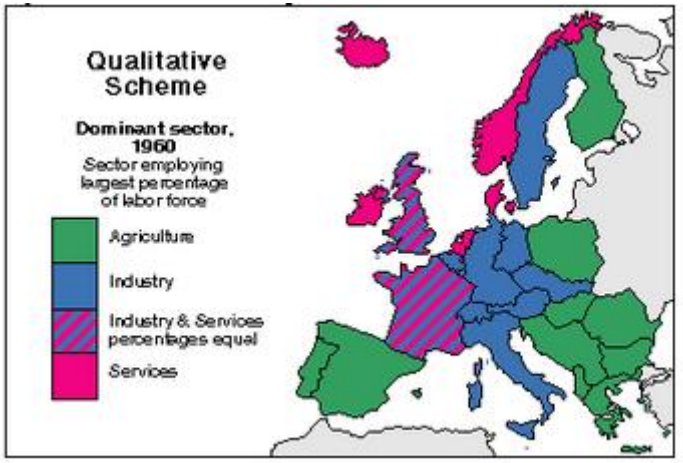
Note: ggplot already uses better looking colors by default

Consider: **color choice**

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



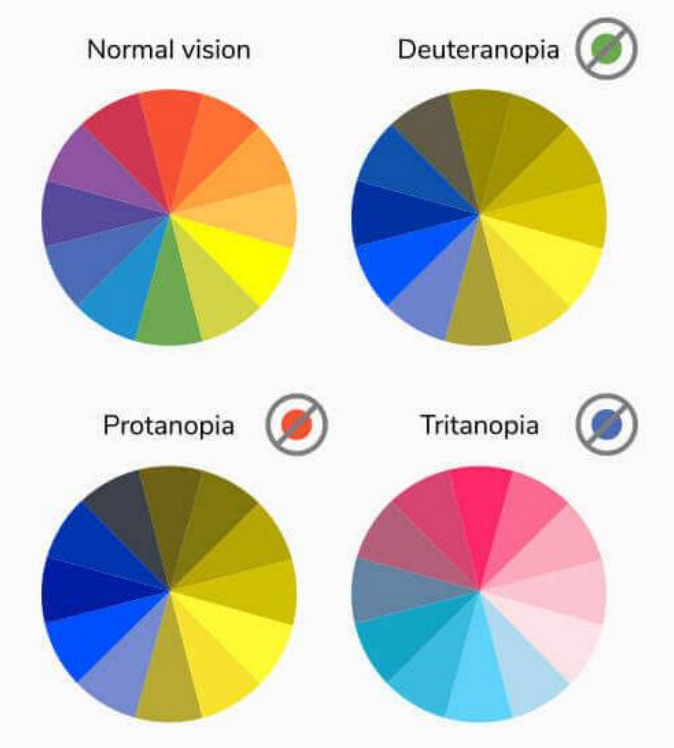


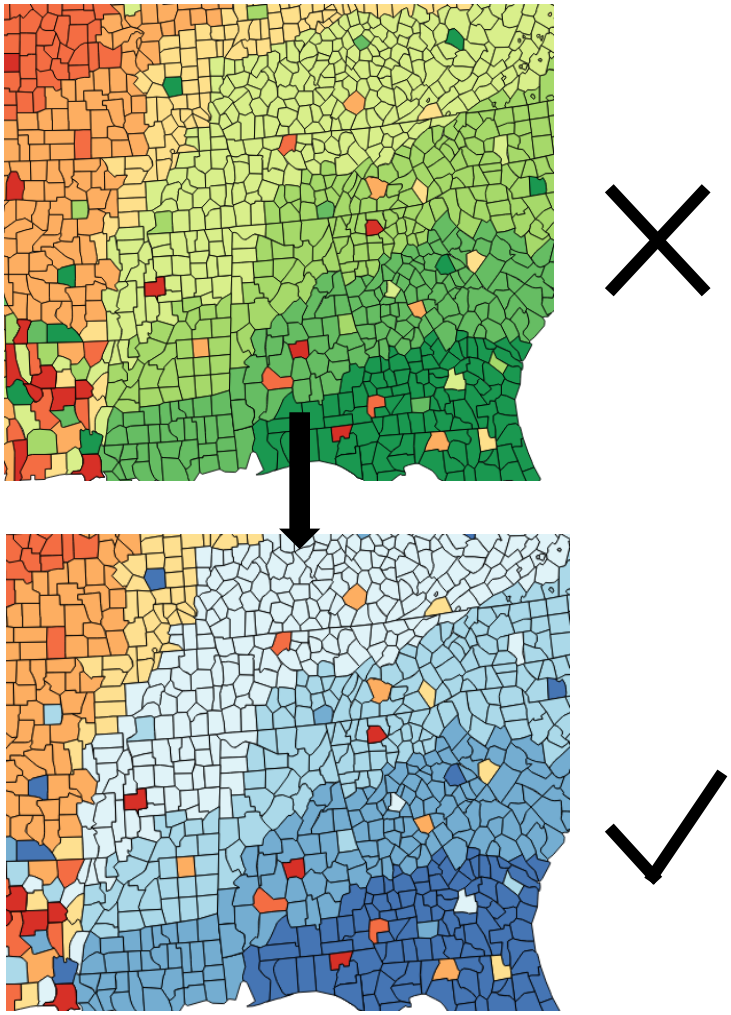


Maps from https://web.natur.cuni.cz/~langhamr/lectures/vtfg1/mapinfo_2/barvy/colors.html

Consider: **color choice**

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

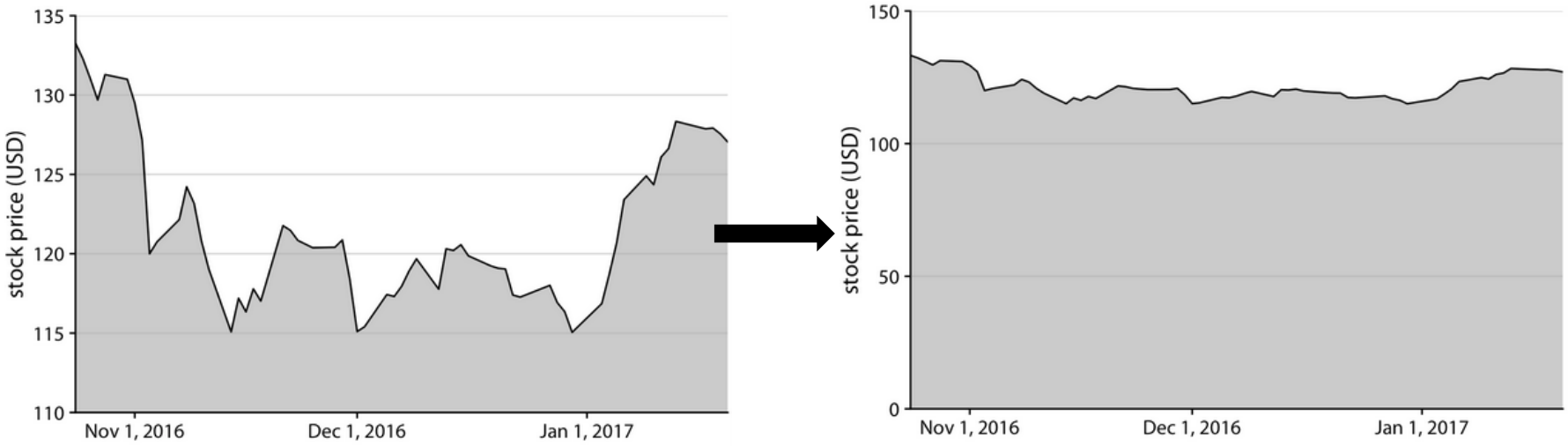




Check out **ColorBrewer2.org** for great color-blind friendly palettes.

Consider: **the number zero**

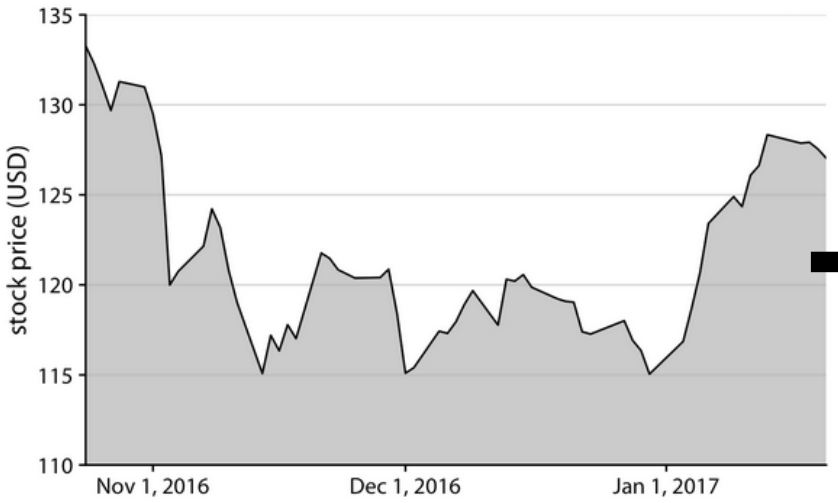
Always keep in mind that differences are distorted by your choice of axis range.

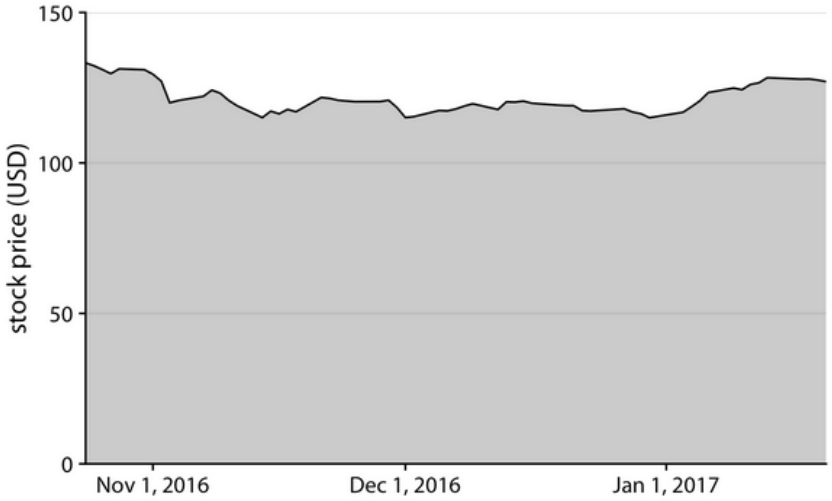


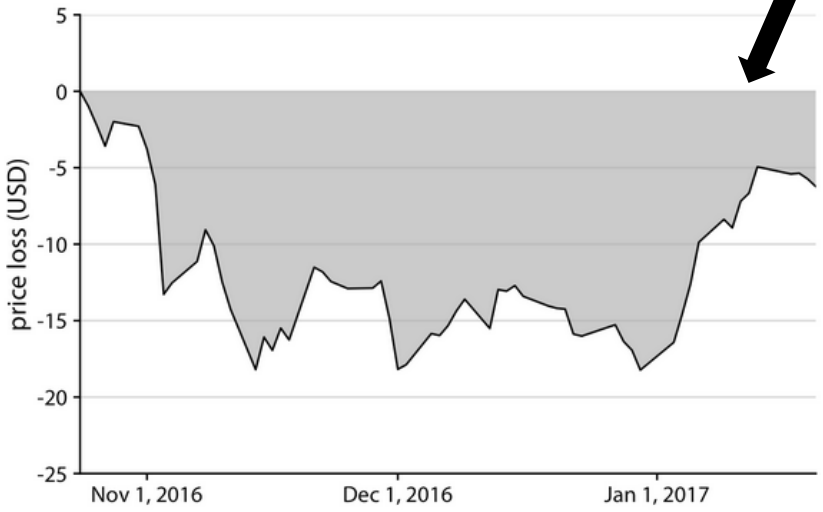
Graphs from *The Fundamentals of Data Visualization*, Claus Wilke

Consider: **the number zero**

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



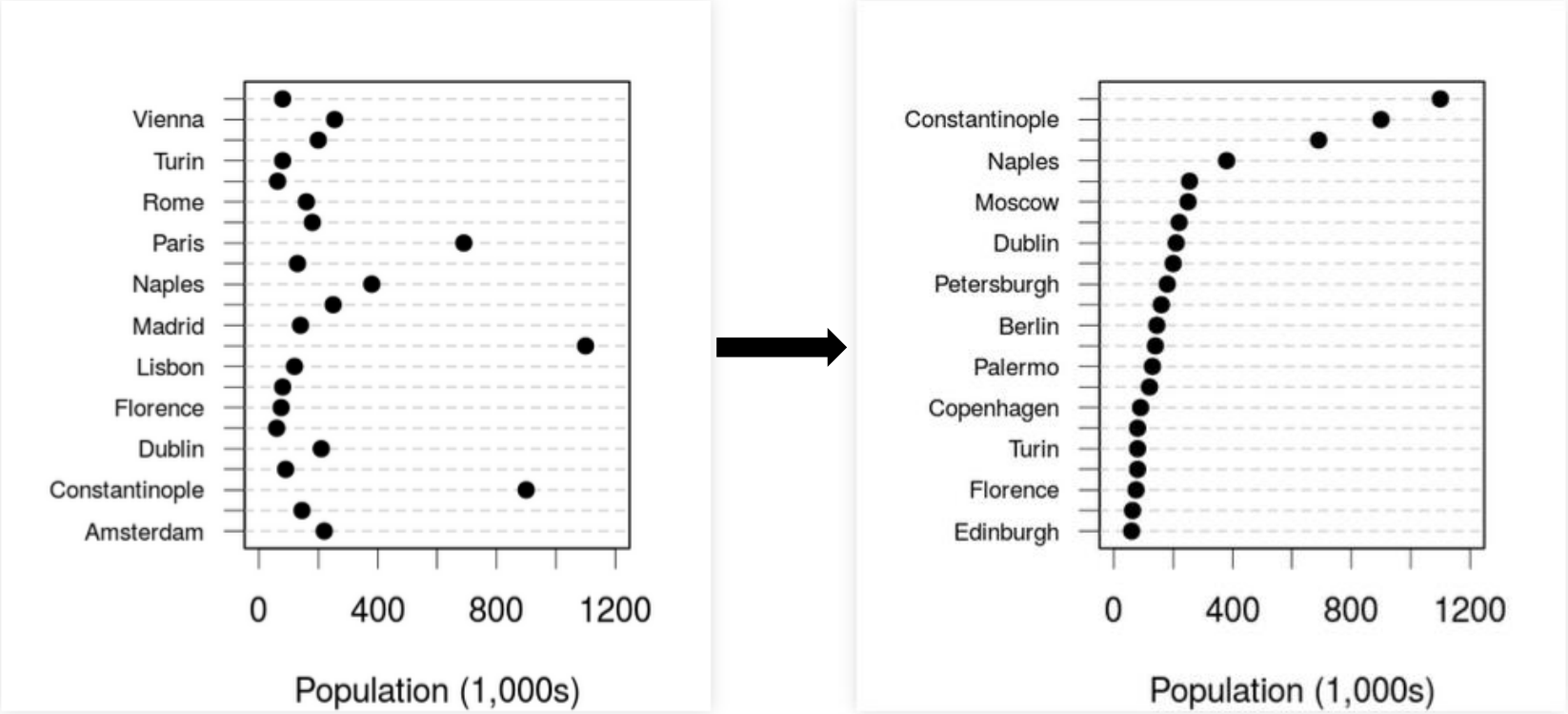




Graphs from *The Fundamentals of Data Visualization*, Claus Wilke

Consider: **aiding comparisons**

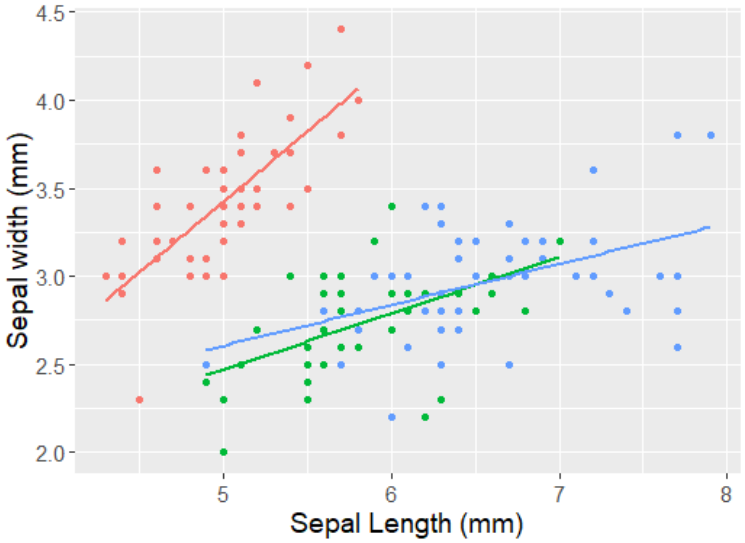
Show your data in a way that facilitates direct comparisons.



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

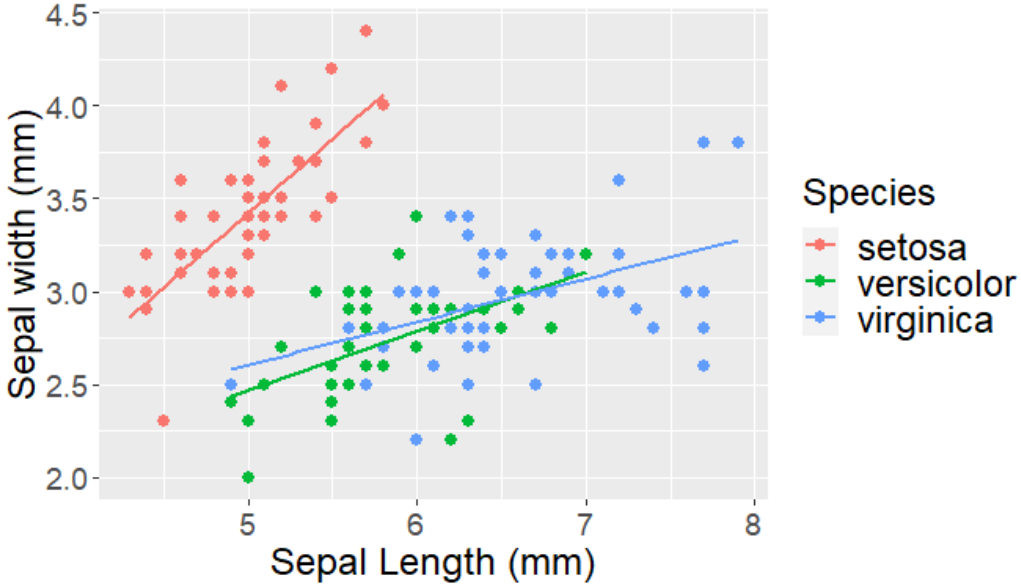
Consider: **your audience**

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



Consider: **whether your graph looks good**

In other words, all of this is very subjective, and there are exceptions to every guideline. Follow your instinct.

Don’t make figures that look like this:

Figure from Data Visualization: A Practical Introduction