Presenting Science, part II

Presenting data and general tips

Follow up discussion

- We discussed the presentation by Garr Reynolds assigned for viewing in the previous class.

- As a follow up Petra shared a link to another presentation on avoiding death by powerpoint:
  - https://www.youtube.com/watch?v=Iwpi1Lm6dFo

Presenting data

- Life is complicated. So are the data that describes it.
- We can take cues from the positive trends in data journalism to learn about visualising and talking about complex data.
  - See for some commentary on data journalism:

    https://www.oreilly.com/library/view/the-data-journalism/9781449330057/ch01.html
- The data you include needs to be part of the story.
  - The data and the graphics aren't THE story.

Presenting data

- The data gives us hints about how we should present it.
  - Choosing the right chart, summary statistic, or terms is usually decided by the structure and type of the data
- In science we usually have two sets of data: observed and analytical
  - Observed data: the raw data we collect from experiments, surveys, observational studies, or other sources.
  - Analytical data: the data generated through analysis of the data (i.e., by applying some statistical methods).

Presenting data: observed data

Categories of data:

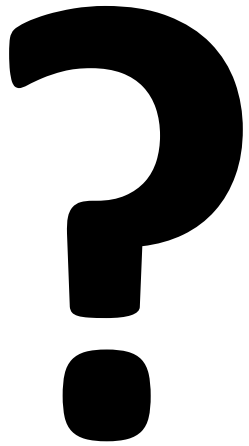
- Binary (yes/no, 0/1, true/false)
- Categorical (yes/no/maybe/don’t know; no order)
- Ordinal (sometimes, usually, often; order matters)
- Binomial (0,1,2,3…n; when limited, e.g., 3 out 10)
- Count (number of coffees purchased in an hour);
- Real-valued (continuous data; distance, MoE, etc.)

Presenting data: observed data

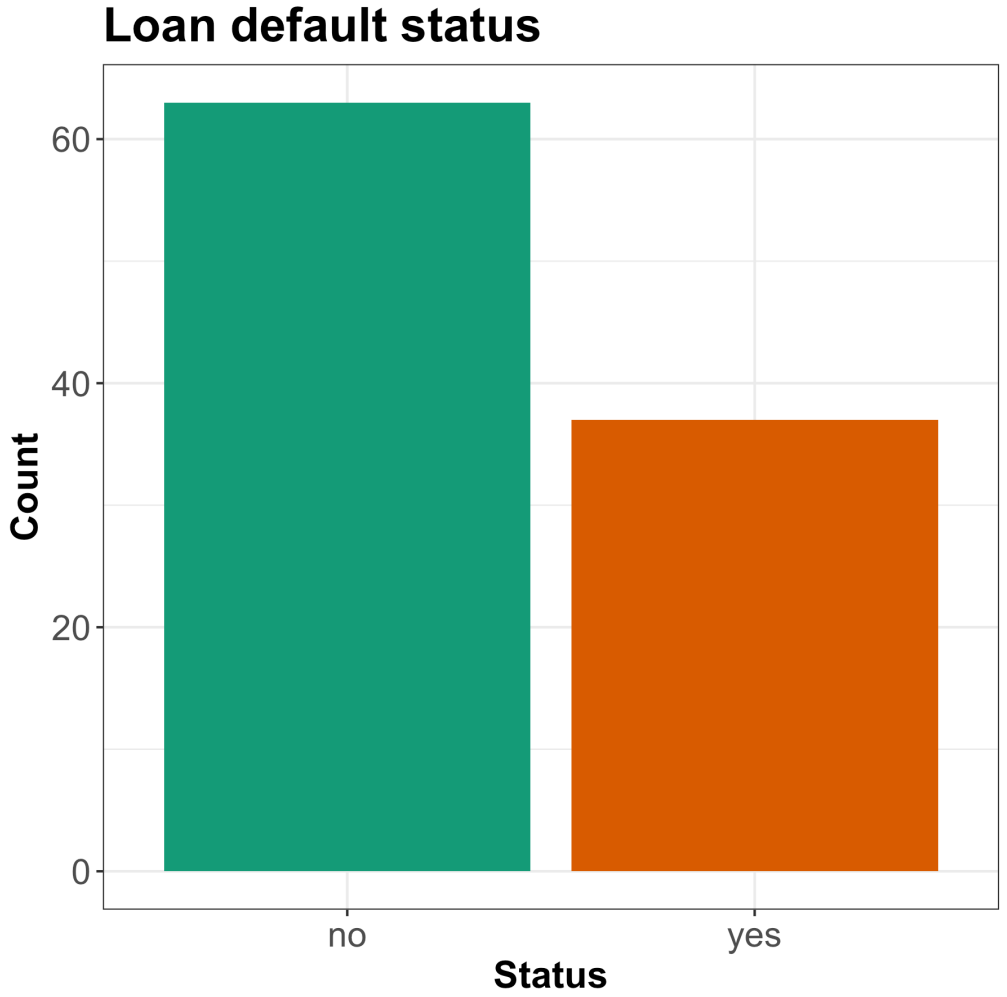
Observed data may also include distributional data:

- Histograms,
- Density plots (Probability density functions),
- Cumulative distribution function (CDF)
- Summary statistics (e.g., point estimates like mean, median and uncertainty like standard deviation, confidence intervals)

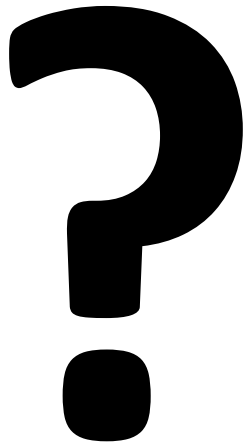
Observed data: binary results



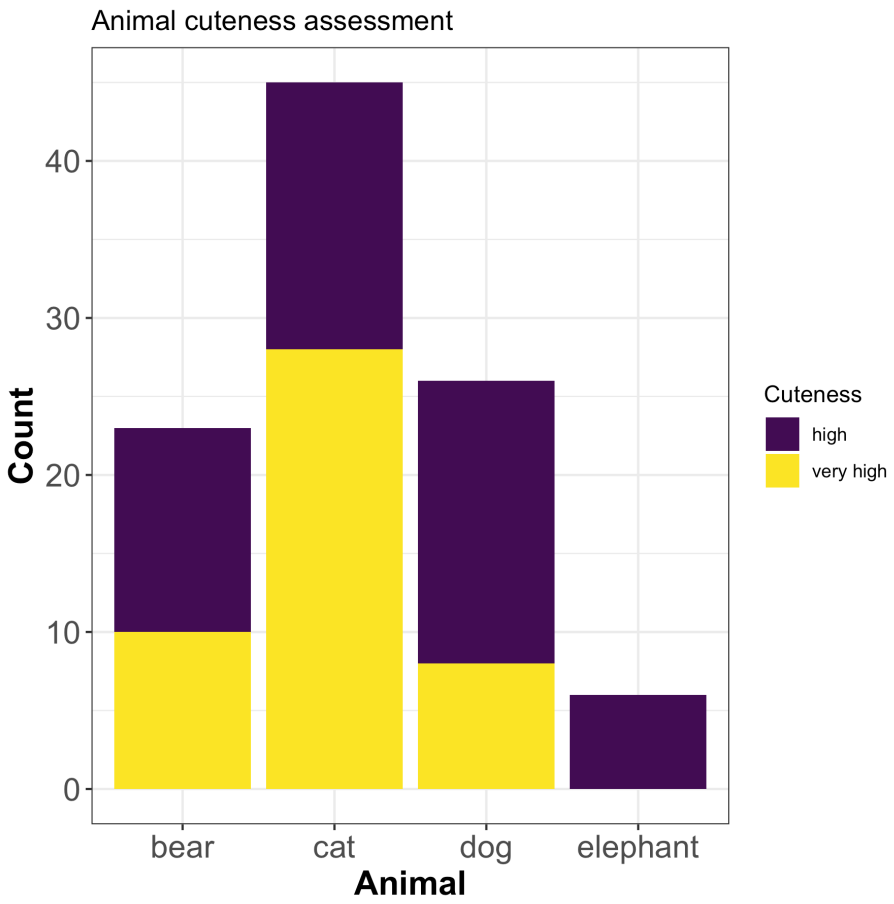
Observed data: binary results

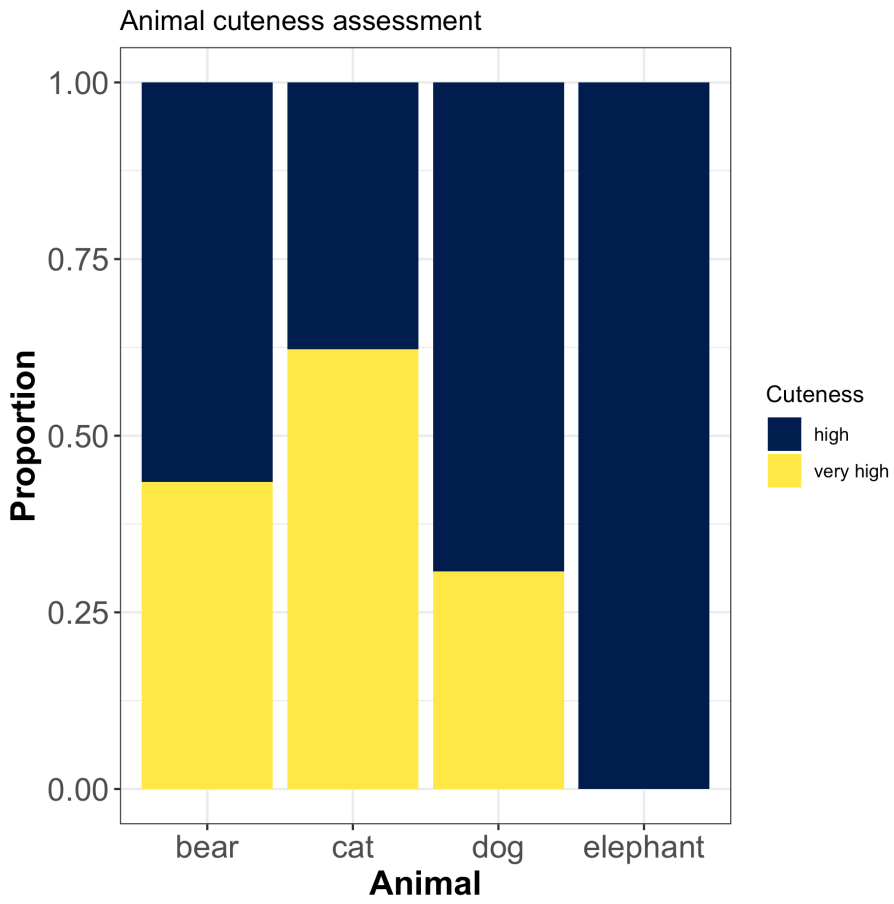


Observed data: categorical results

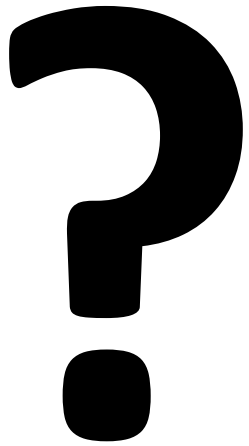


Observed data: categorical results

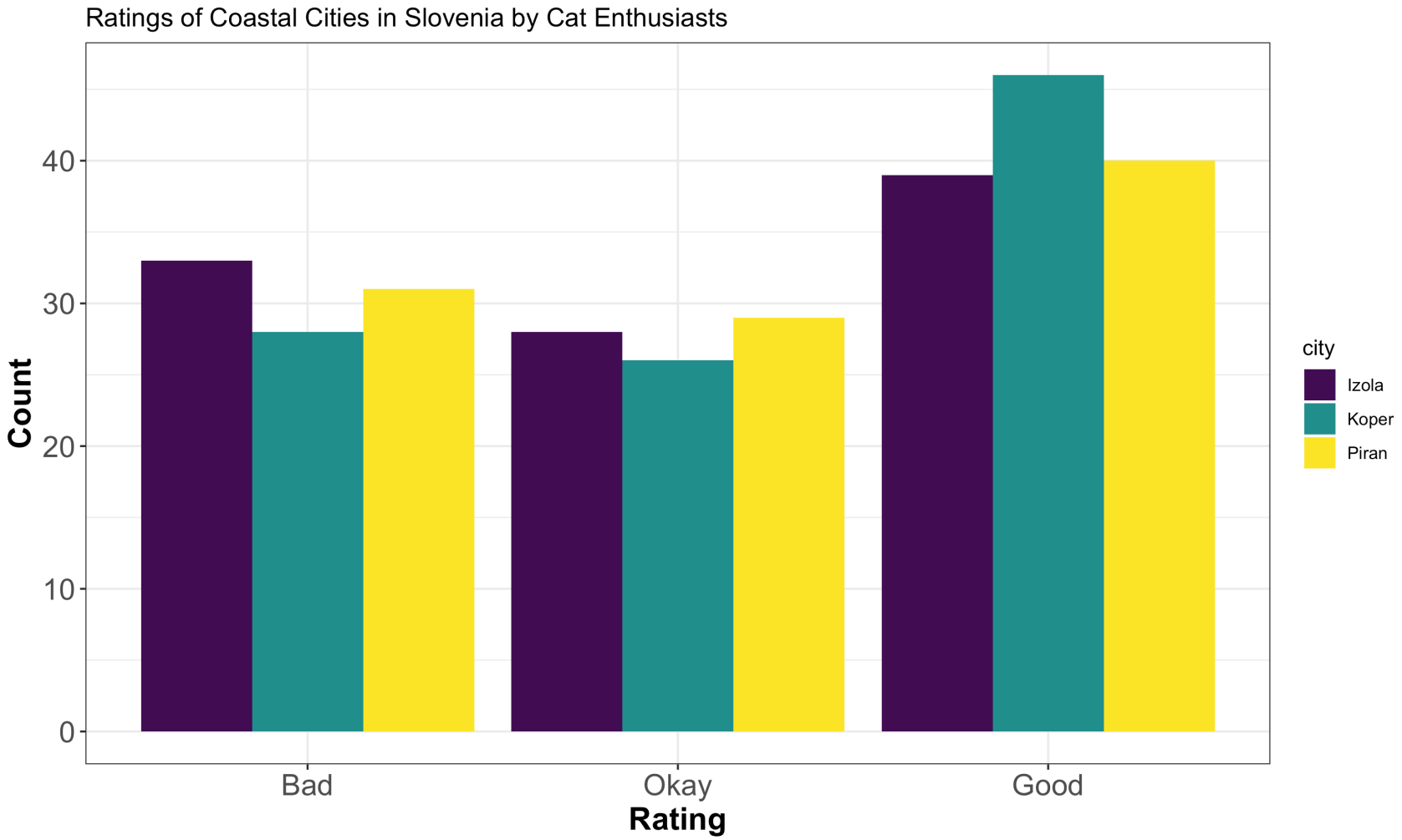




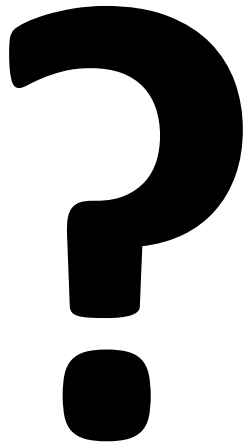
Observed data: ordinal results



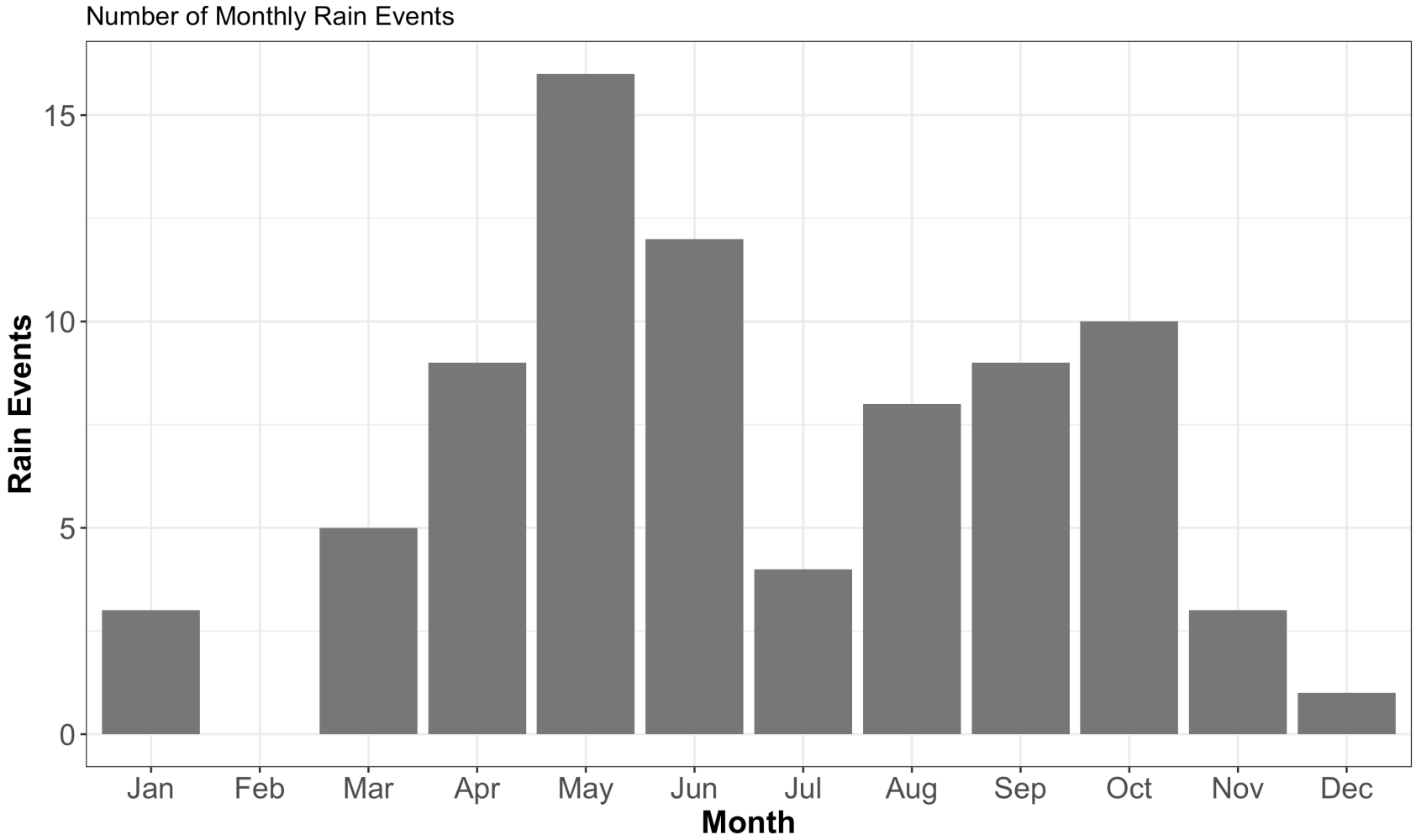
Observed data: ordinal results



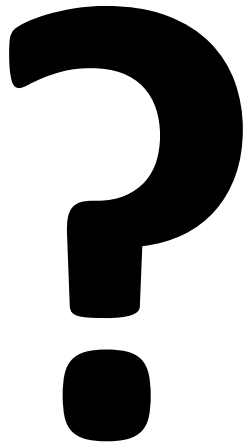
Observed data: binomial results



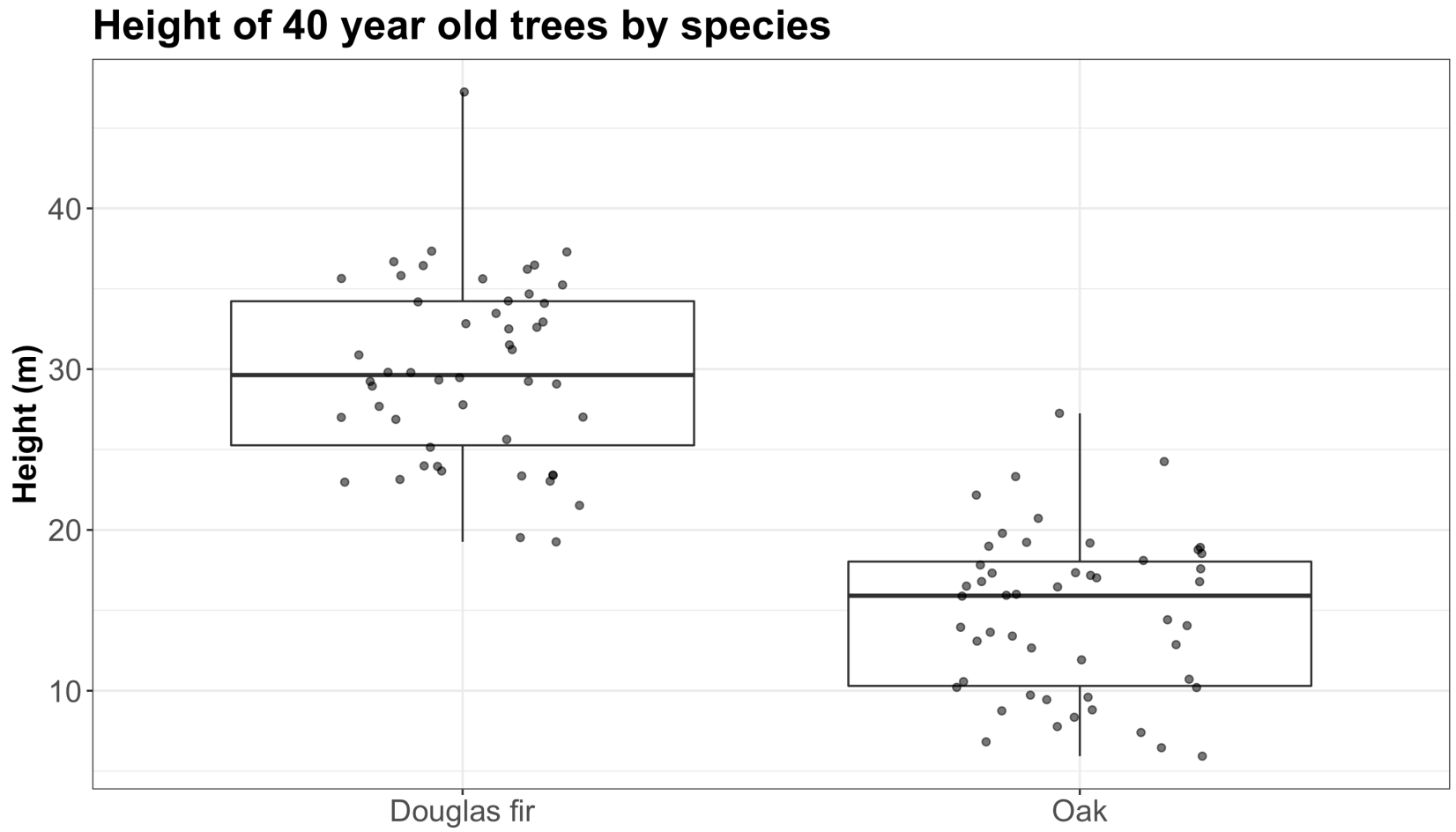
Observed data: binomial results



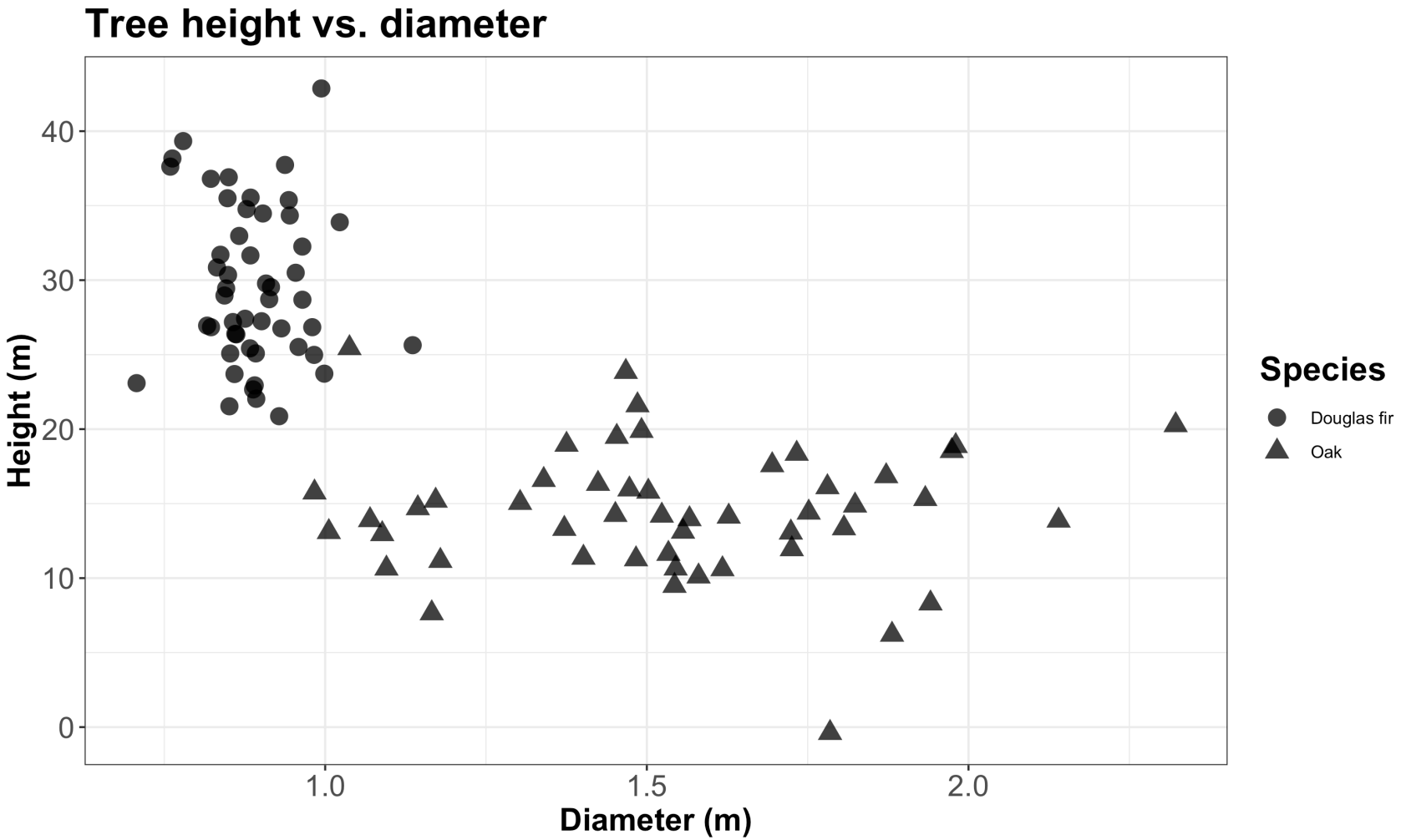
Observed data: Continuous data



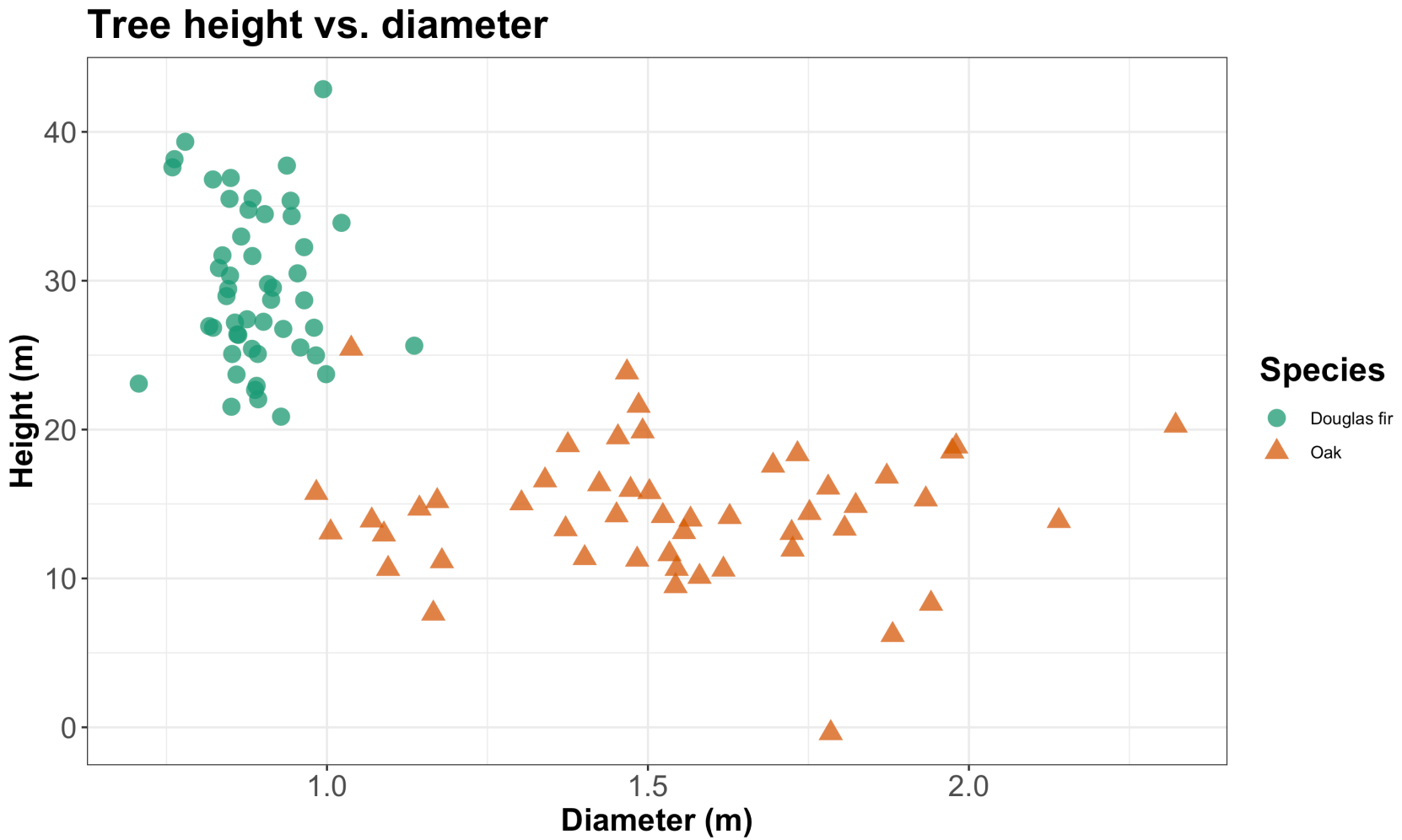
Observed data: Continuous data



Observed data: Continuous data



Observed data: Continuous data



Analytical data

Analytical data includes

- Point estimates (mean, median, mode),
- Uncertainty (standard deviation, variance, confidence intervals)
- Trends (regression lines, time series)
- Effect sizes (ratios, differences)
- …

Elements of a chart: Grammar of Graphics

- Like language, there are some basic rules and a structure to data graphics. Several implementations exist (in R, matlab, python, etc.). Current state of the art is ggplot for R.
- The grammar describes the elements of a plot:
  - Link between data and aesthetics
  - Set(s) of geometric shapes
  - Scales, axes, coordinate system
  - Other elements (titles, annotations, etc.)
  - These are usually combined in layers.
  - (many other things)

Elements of a chart: Grammar of Graphics

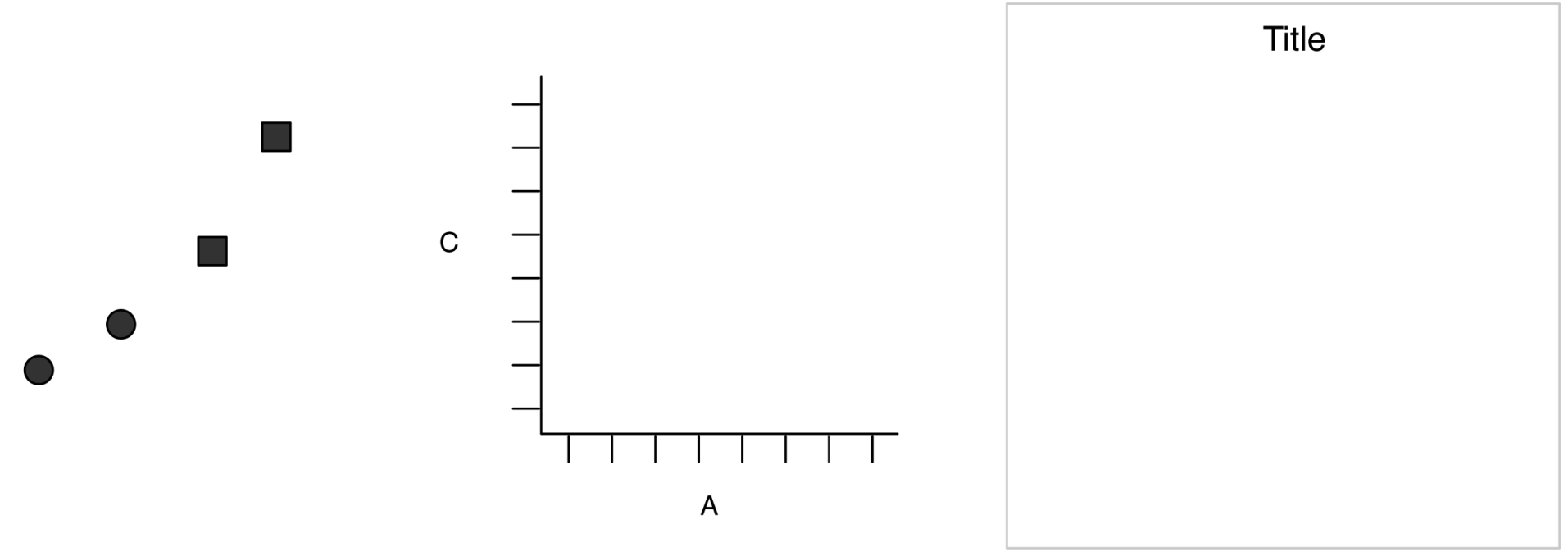


Figure 1. Graphics objects produced by (from left to right): geometric objects, scales and coordinate system, plot annotations.

https://vita.had.co.nz/papers/layered-grammar.pdf

Elements of a chart: Grammar of Graphics

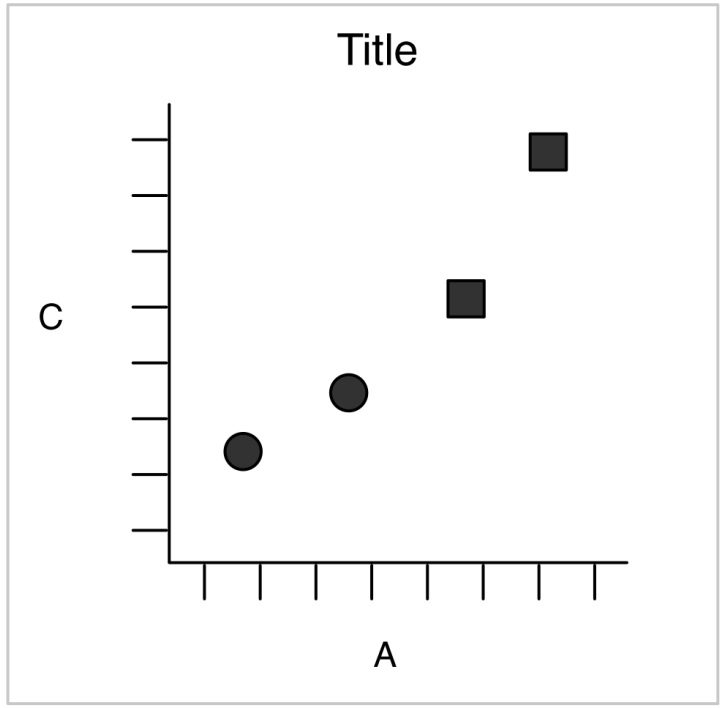


Figure 2. The final graphic, produced by combining the pieces in Figure 1.

https://vita.had.co.nz/papers/layered-grammar.pdf

Data Graphics tools

- The best tools for making charts include data manipulation, analysis, and graphics creation.
  - **R & RStudio**, Python, Matlab($$$)/Orange, (Stat, SPSS, SAS, etc. $$$), Excel(☹), BioVinci ($$), …
- Other tools focus on just the graphics part
  - Adobe Illustrator ($$$), Prism (graphad, student license ~ 114$), chart-studio.plotly.com, PowerPoint(☹), SigmaPlot ($?, not easily available outside of NA), Adobe Spark Charts (https://spark.adobe.com/express-apps/chart/, free but very basic), …
- Some tools focus on specific applications
  - Gephi for networks, ChemSketch for molecules, …

Data Graphics tools

Other tools mentioned in class (thanks Dean):

- https://vnijs.shinyapps.io/radiant/?_ga=2.252052621.504611160.1610445664-896780028.1596441578&SSUID=dfe397c97b
- https://www.jamovi.org

Data Graphics tools

Mike, why don’t you shut up about R & RStudio?

- Free!
- Open Source
- Flexible, Extensible
- Amazing community
- Lots of tutorials, classes, support available

Data Graphics tools

Mike! What if I still want to use Excel.

- Ok… But…
- Throw out the default settings
- Don’t let the chart styles throw you off
- Let the data guide you to the right plot…

Data graphics: colours

Colour matters!

- Colour blindness…
- Colours can be used to differentiate groups (categorical data), show change (continuous data)
- In papers: minimize use of colour, but don’t hide from it (when its free).
- In presentations: use organizational or project theme. If absent, stick with a consistent palette throughout the presentation.

Data graphics: colours

Colour matters! Resources:

- Color Brewer: selection off palettes for cartography, but widely useable. Built into many graphics tools. https://colorbrewer2.org
- Viridis: Colourblind safe palettes & gradients (not only in R):
  - https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
- Adobe Color (Free!): selection of colors based on colour theory and palettes created by others
  - https://color.adobe.com

Data graphics: colours

Colour matters!

- There are many ways of specifying colours.
- Hexidecimal (3 pairs of hexadecimal characters (0-9A-F) for RGB): #FFFFFF = White; #FF0000 = Red; #00FF00 = Green, #0000FF = Blue).
- RGB(a) (3 sets of numeric values: 0-255 for RGB, (0-1 scale for alpha)): 0,0,0,1 = black; 255,255,255 = white)
- CYMK (4 values between 0 and 100 for Cyan, Yellow, Magenta, and Black): 0,0,0,0 = white; 0,0,0,100 = Black

Data graphics: colours

Colour matters!

- HSB (3 numeric values for hue, saturation, and brightness): H is between 0 and 360, S and B between 0 and 100.
- CIELAB (Very important in science! 3 values for Lightness, a* and b*): L is between 0 and 100, a* and b* are between -128 and 127).

… RGB and Hex are most common for screens. CYMK is most common for printing (and what journals request). CIELAB values are mostly used to scientifically identify colour, and especially change in colour.

Data graphics: image formats

Image formats fall into two categories.

- Raster (jpeg, png, tiff, etc.)
  - Best for photos,
  - Widely viewable in many programmes
- Vector (svg, ps, eps, ai, etc.)
  - Best for line graphics (like charts!)
  - Not as widely viewable. Most browsers can deal with svg. Others are more specialized.
- What about PDF? Not an image format, but can be a useful wrapper around vector formats that otherwise aren’t portable.

Data graphics: Type

Typefaces / Fonts are sets of characters for text.

- Two broad categories: serif and sans serif (with and without serifs).
  - Semi-sans exists too
- Monospace vs. variable space is also important.
- MANY other aspects of type / fonts…
- Be consistent. Either sans or serif fonts can be used. Some fonts are designed for print, some for screen.
- Two main types of fonts: open type and true type

Data graphics: Type

Fonts come with open-source licenses!

- Google has some ok open-source fonts.
- Best are here:

  https://www.theleagueofmoveabletype.com

(after all, Lady Gaga uses them!)

- For journals: follow guidelines
- For presentations: Test your presentation on a projector to make sure axes text, legends, annotations are readable.
- Make sure your font has the appropriate characters…

Final Final Assignment

Due 22h Tuesday 2.2.2021

Description to be posted by 15 Jan.