AST1501 - Introduction to Research

Jo Bovy

Next few weeks

How to make a good plot

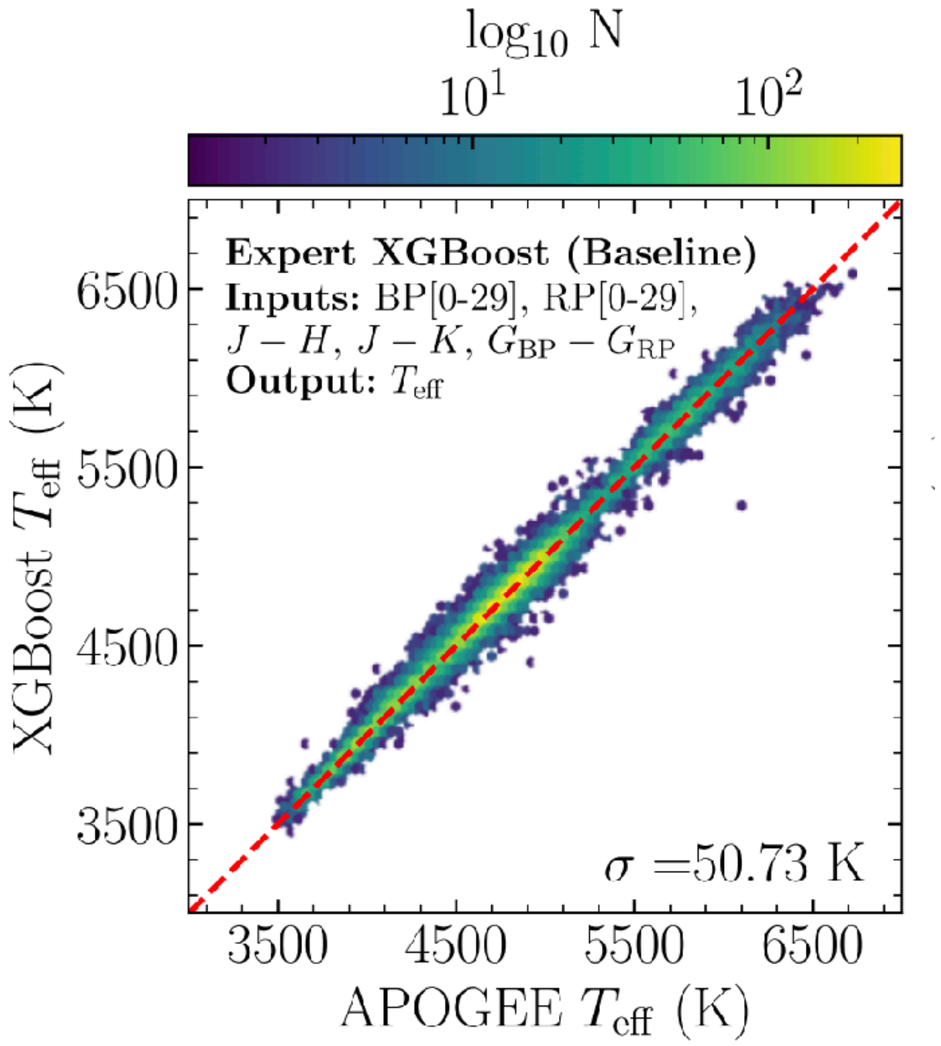
Focus on figure in scientific paper

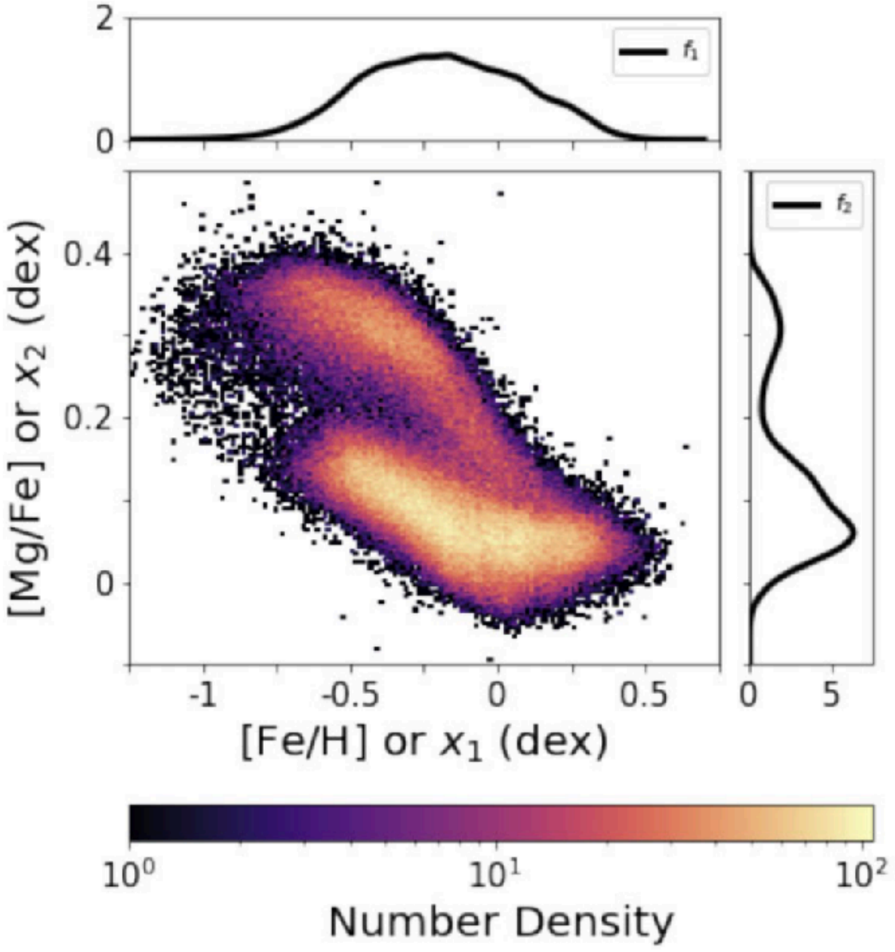
- A good plot is:
  - Informative, but not too information-dense
  - Quantitative: allows the reader to get quantitative information
  - Easy to read and interpret
  - Attractive to look at
  - Accessible

Quantitative plots

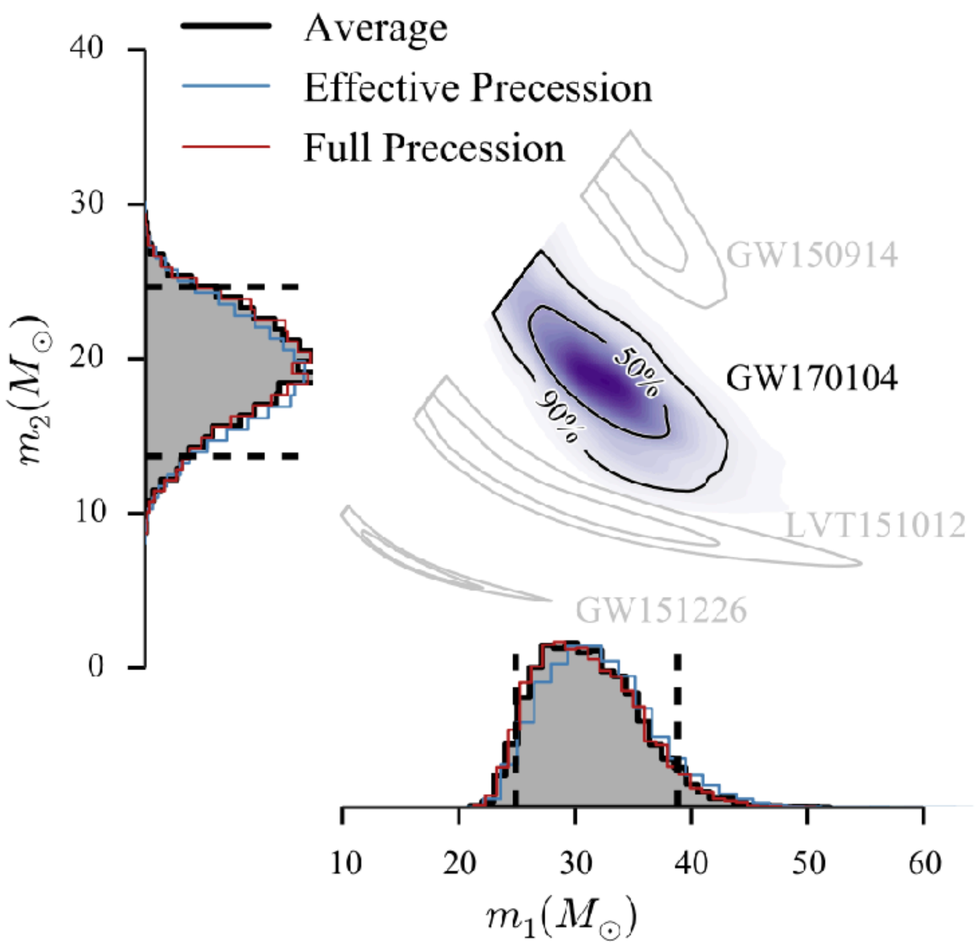
- Plots should allow the reader to gain quantitative information
    - Axes should be included (!) and have their scale clearly labeled
    - Box plots make it easier to read values off a plot
    - Don’t crowd too many lines together

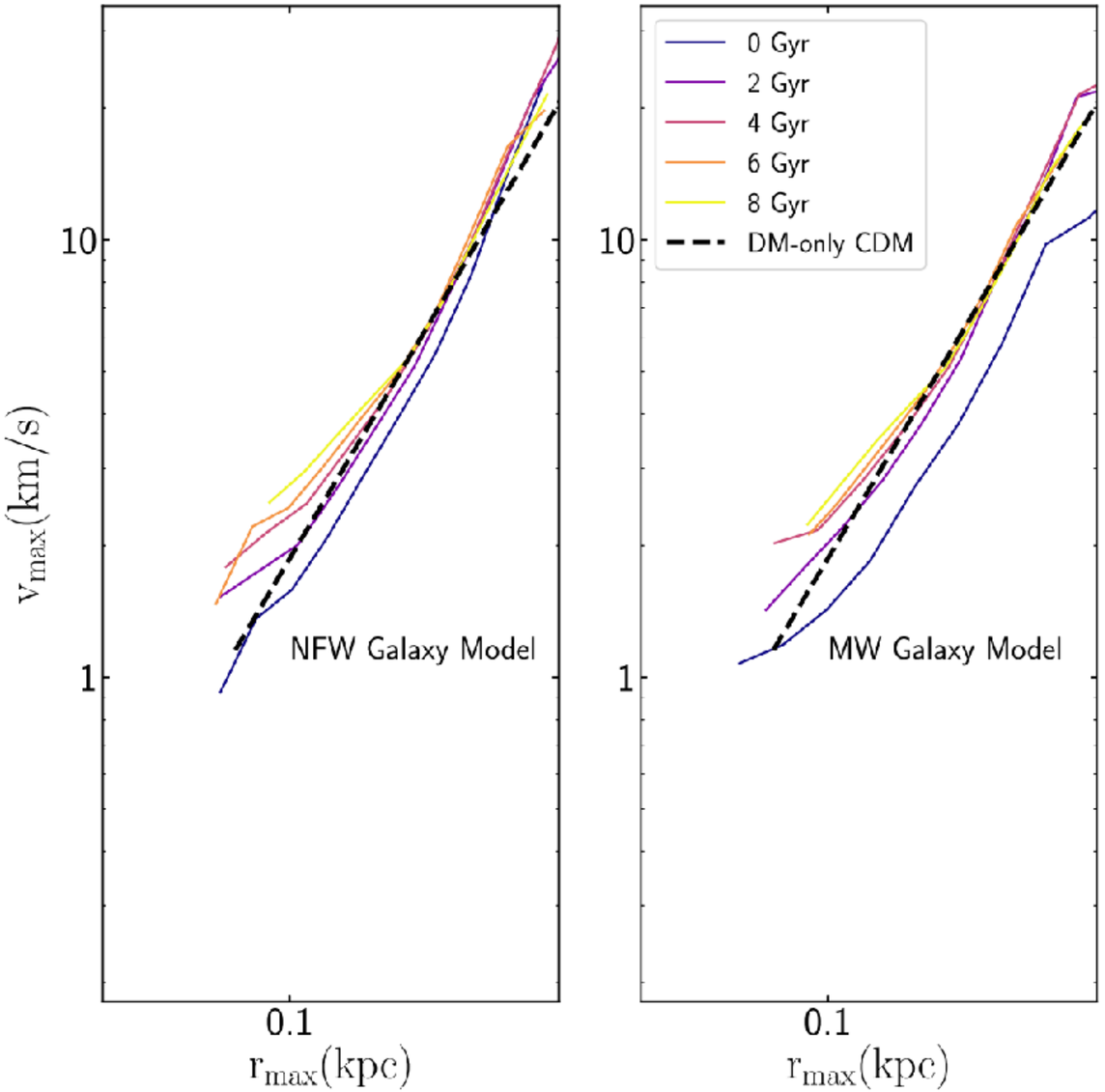
Quantitative plots





Quantitative plots





Abbott et al. (2017; GW170104)

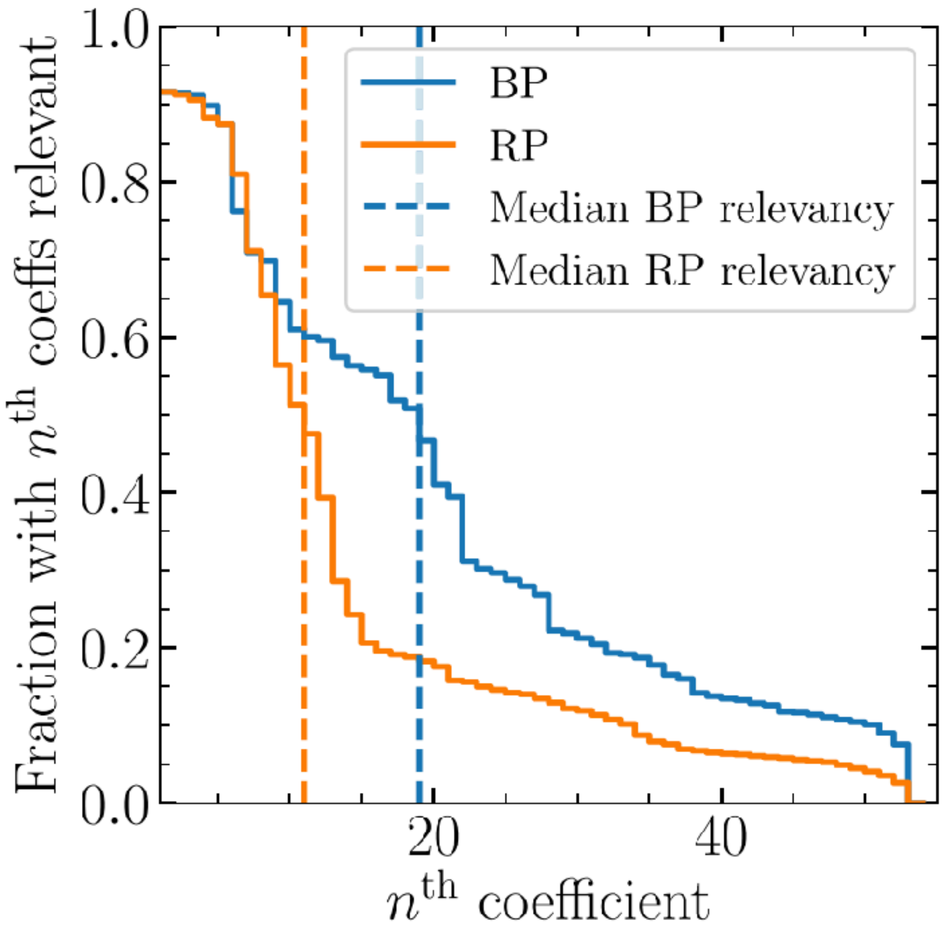
How to make a plot easy to read and interpret

How to make a plot easy to read and interpret

- Some basics:
    - All text should be the same size or larger than the text in the paper itself (requirement for most journals —> they will ask you to change your plot if this is not the case)
    - If you show multiple things, they should be able to be uniquely identified, either in the plot itself or in the caption
        - Generally, multiple things should *not* only be identified in the caption
    - Don’t include too many different things (e.g., lines) —> if there are many things you need to show, spread over multiple plots or re-think how to visualize

How to label lines

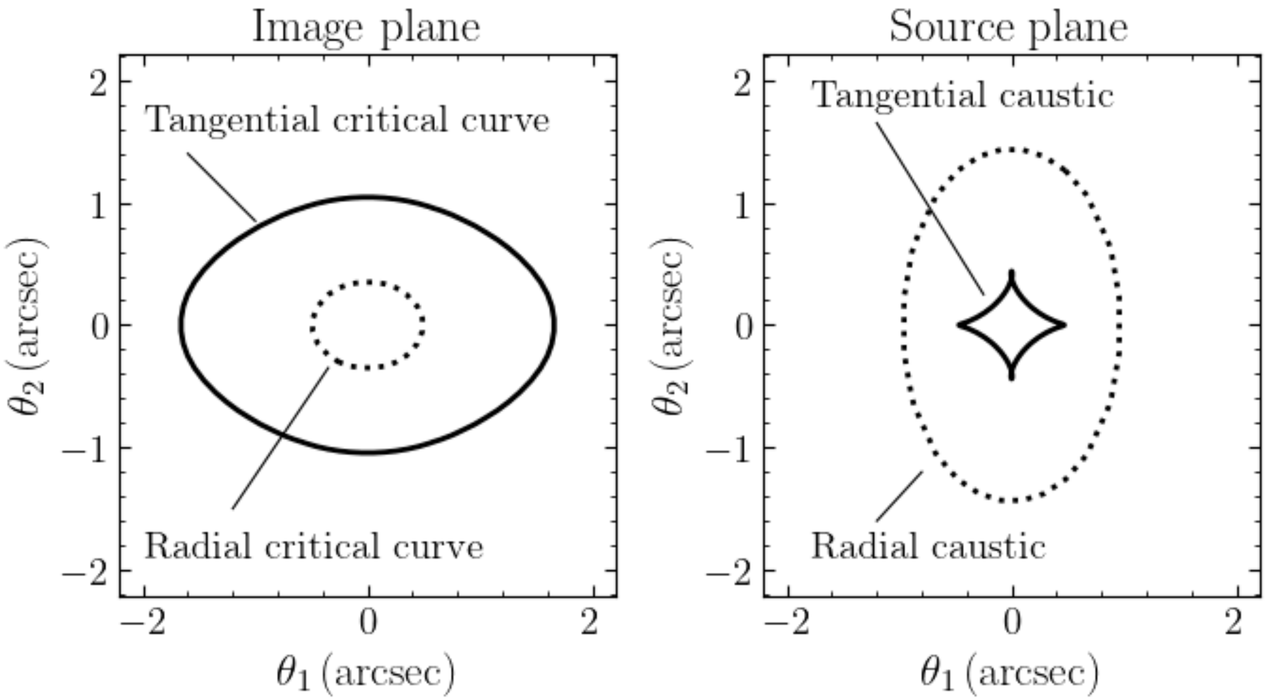
- Default and easiest way to label lines is to use a *legend*
- Really only works if you have only a few lines
- With too many lines, your eyes need to do too much work going back and forth to identify lines



- Would have been better to have two legends for qualitatively different lines

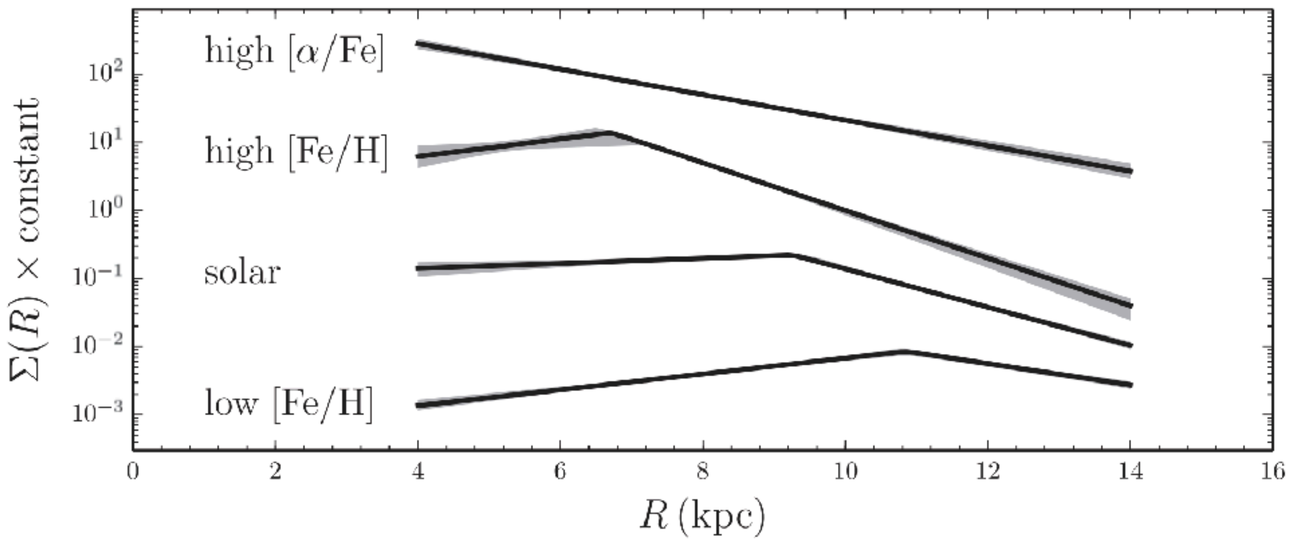
How to label lines

- A better, but more labor-intensive way to label lines is to directly label them
- Unfortunately no automated way to cleanly do this
- But can be worth the effort for important figures or for figures where using a legend would really be too confusing



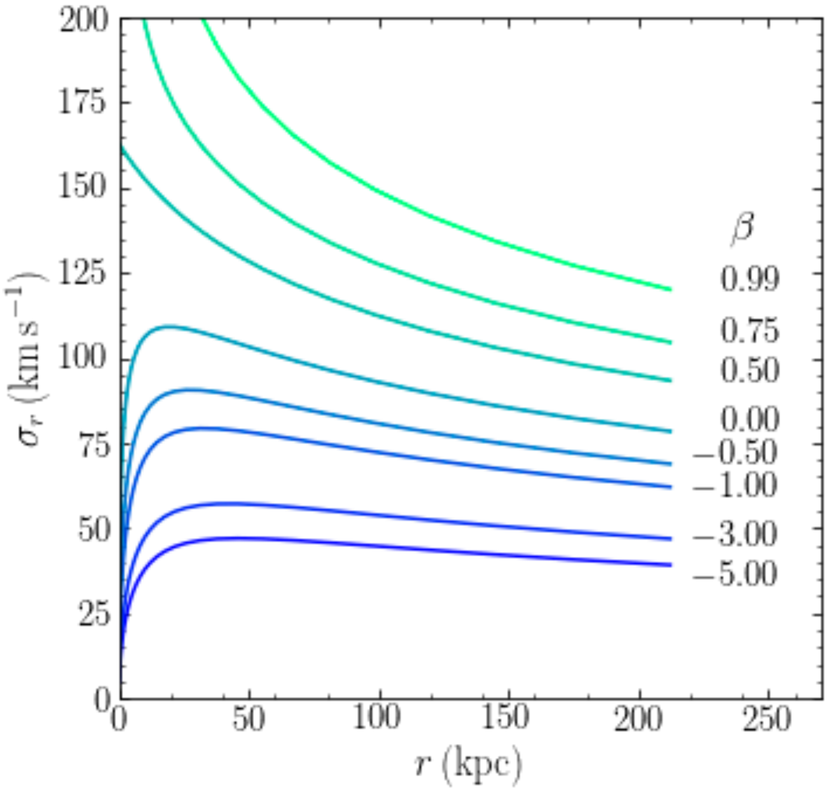
How to label lines

- A better, but more labor-intensive way to label lines is to directly label them
- Unfortunately no automated way to cleanly do this
- But can be worth the effort for important figures or for figures where using a legend would really be too confusing



How to label lines

- A better, but more labor-intensive way to label lines is to directly label them
- Unfortunately no automated way to cleanly do this
- But can be worth the effort for important figures or for figures where using a legend would really be too confusing

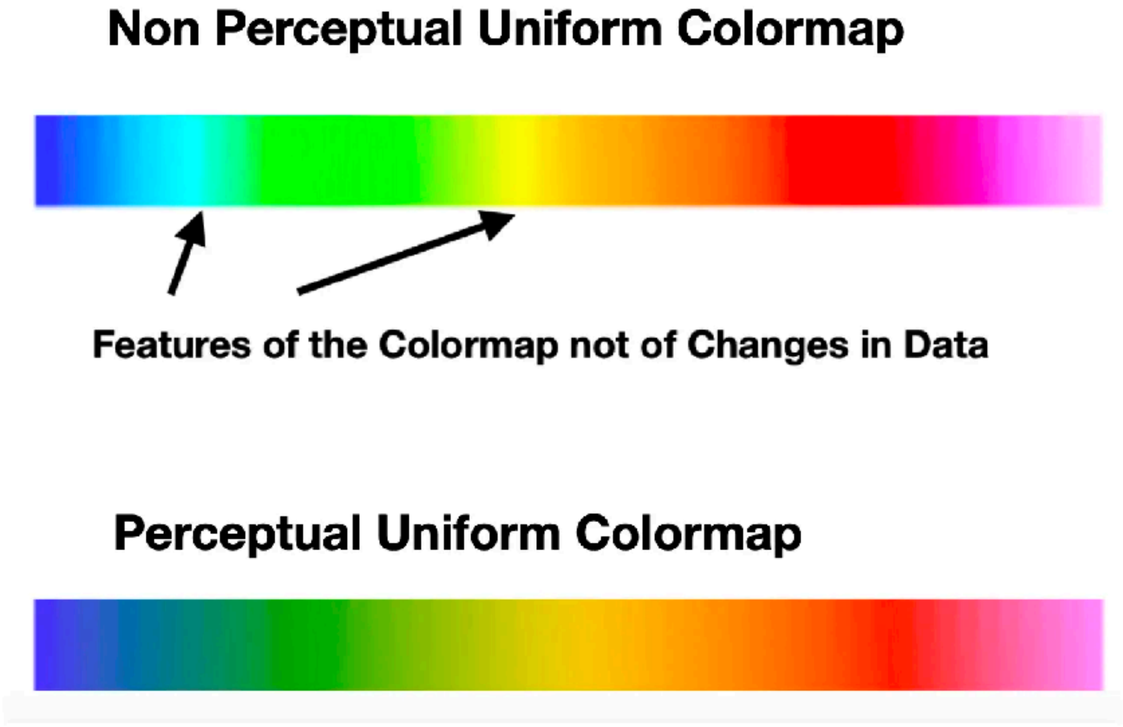


More on lines

- People rarely print papers these days, so use color liberally
- Nevertheless, still useful to use non-color ways of distinguishing lines:
    - Grayscale
    - Line thickness
    - Linestyle: solid, dashed, dotted, dash-dotted, … (note that you can make your own patterns in `matplotlib`)
- For example, if lines fall along two axes, use color along one and linestyle for the other
- Rule of thumb for grayscale, linestyle: conserve ‘ink’: make lighter lines thicker, dotted lines thicker

Colorbars

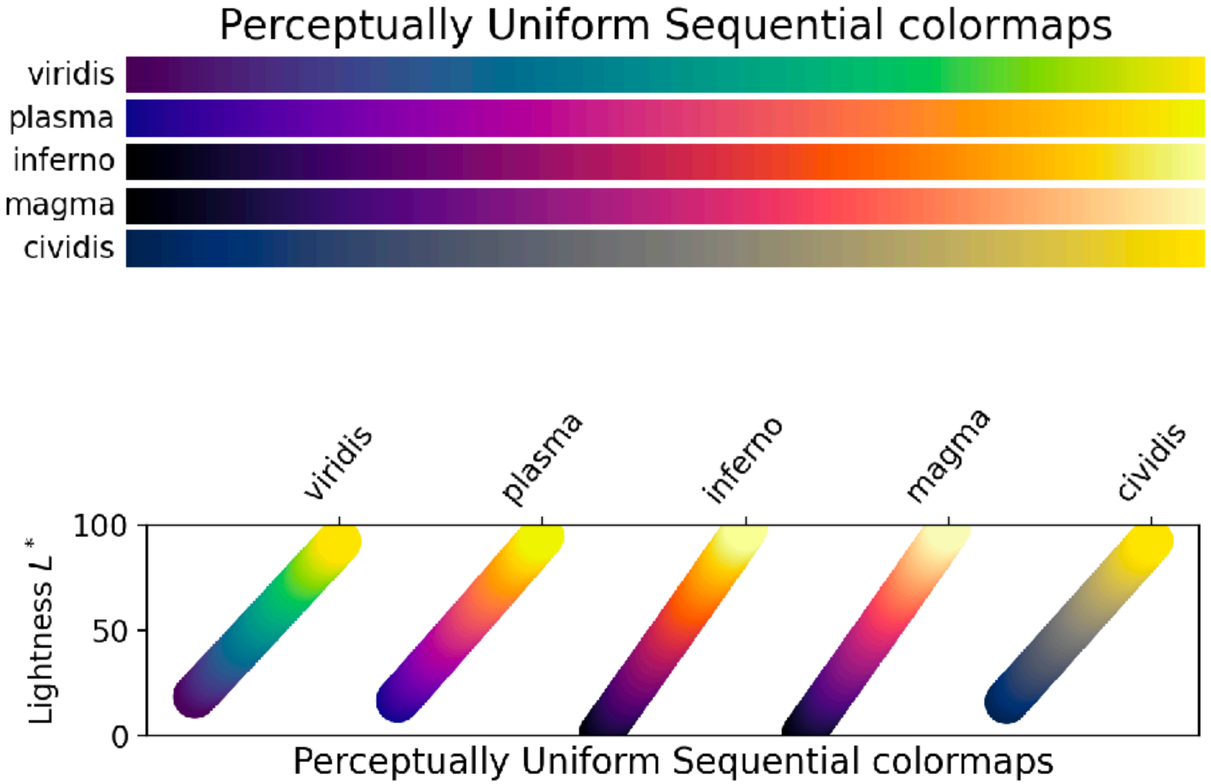
- Use colorbars for 2D plots of functions of two variables, line-colours for a sequence of lines, etc.
- Basically two qualitative options: sequential vs. divergent (see https://matplotlib.org/stable/users/explain/colors/colormaps.html for a good overview)
- Sequential: use perceptually uniform colormap



From: https://medium.com/nightingale/color-in-a-perceptual-uniform-way-1eebd4bf2692

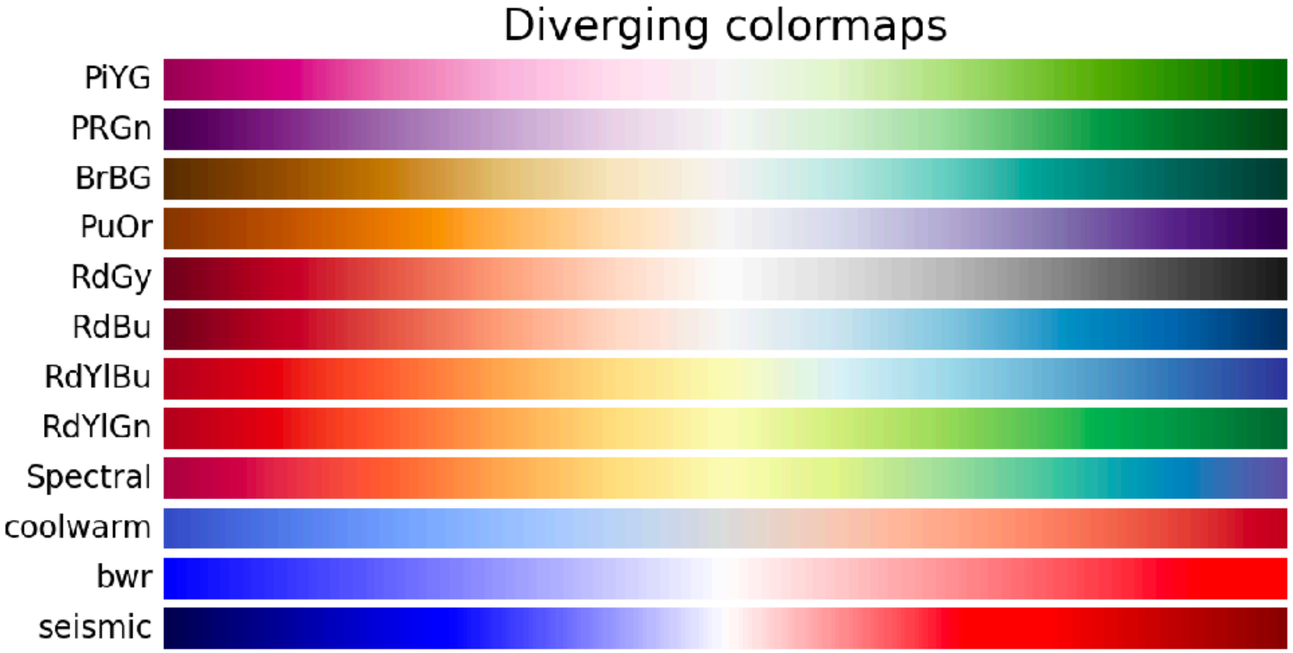
Colorbars: sequential

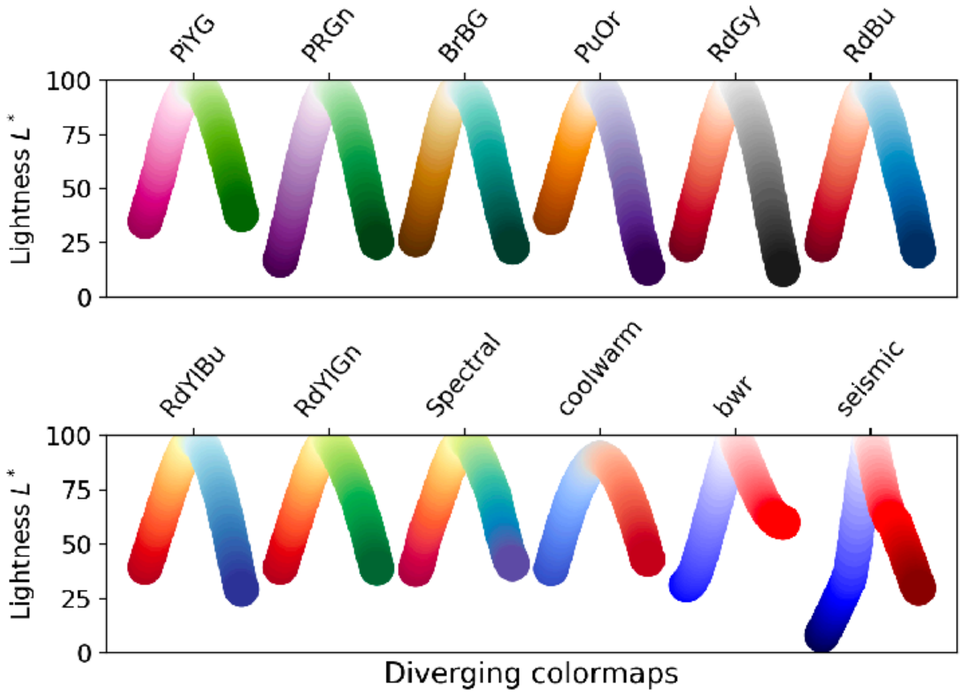
- Use for ordered data without a special value in the range
- Best to use one of matplotlib's perceptually-uniform colormaps (or equivalent in other plotting programs)
- Viridis standard to some degree, but sometimes another one can look better



Colorbars: divergent

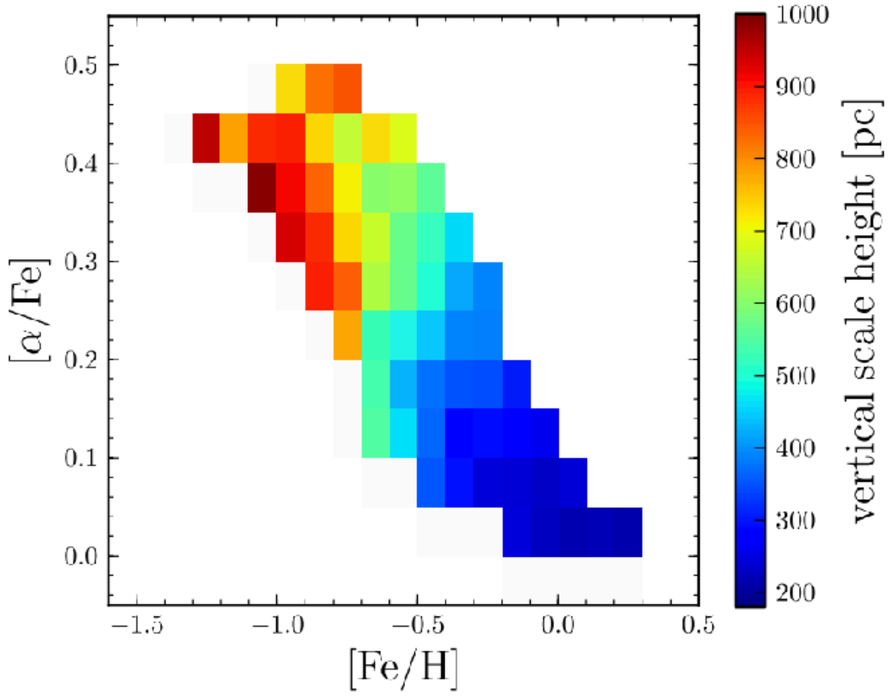
- Use when you want to show deviation from standard value, often zero (residuals etc.)
- People generally use a red-blue type

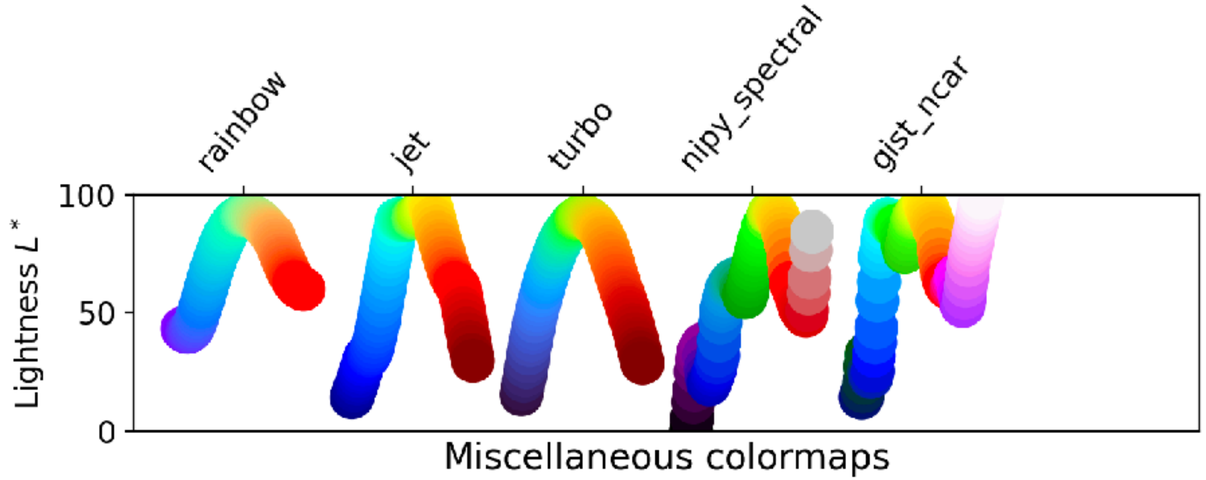




Colorbars: don’t use jet!

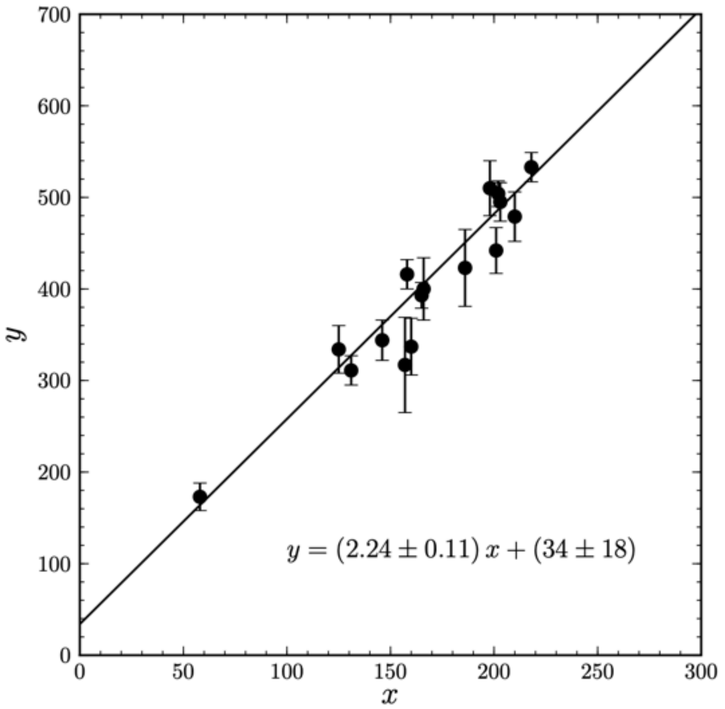
- Jet used to be the standard matplotlib colorbar (and standard in many other languages)
- Bad because:
  - Not perceptually uniform —> creates features in the plot that aren’t in the data
  - Bad for colorblindness





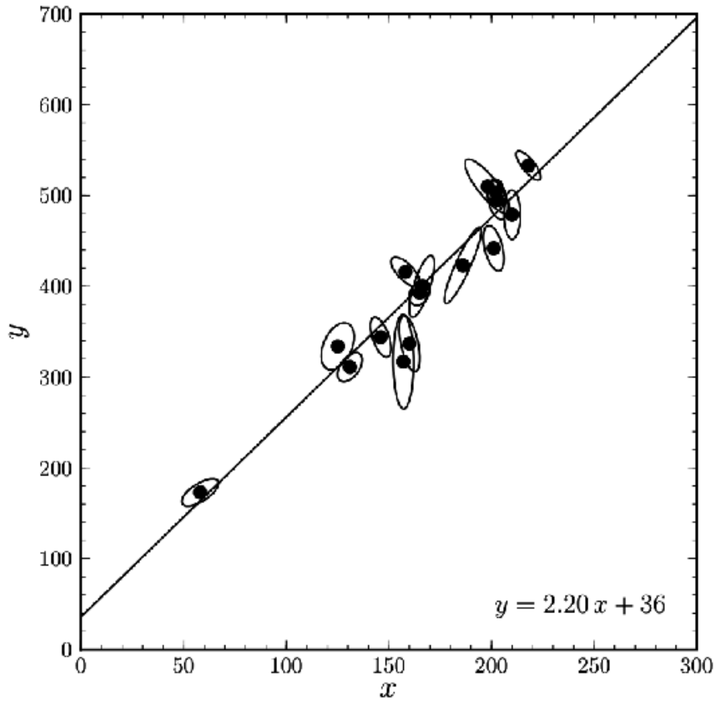
Data

- Data: points (x,y) potentially with errors (x_err, y_err)
- Standard way to plot these is as a point + error bar
- When showing multiple sets of data, use same considerations as for lines (use color and marker-style in this case)
    - But even more important to not put too many different types in one plot
- For large number of points with errors, consider not showing individual error bars but instead summary



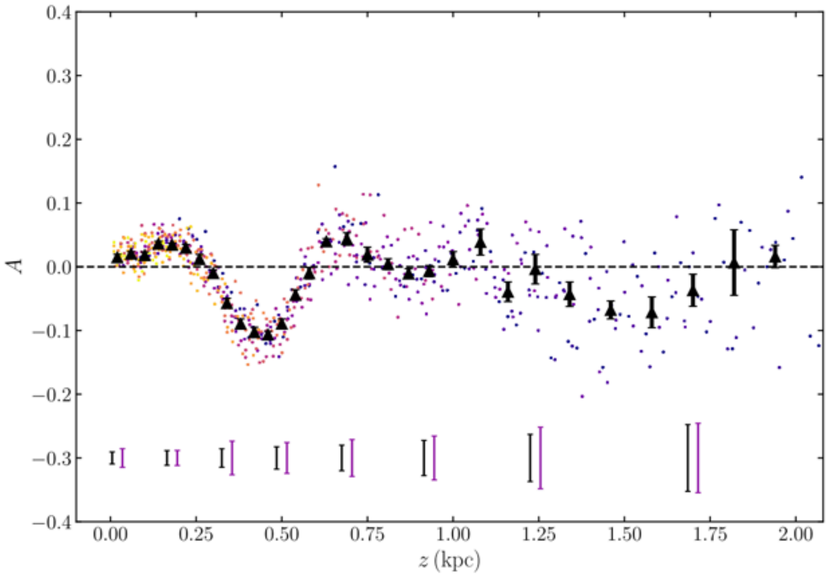
Data

- Data: points (x,y) potentially with errors (x_err, y_err)
- Standard way to plot these is as a point + error bar
- When showing multiple sets of data, use same considerations as for lines (use color and marker-style in this case)
    - But even more important to not put too many different types in one plot
- For large number of points with errors, consider not showing individual error bars but instead summary



Data

- Data: points (x,y) potentially with errors (x_err, y_err)
- Standard way to plot these is as a point + error bar
- When showing multiple sets of data, use same considerations as for lines (use color and marker-style in this case)
    - But even more important to not put too many different types in one plot
- For large number of points with errors, consider not showing individual error bars but instead summary



Other considerations

- Don’t plot too many data points, because PDF will save each point —> takes a long time to load
    - Solution: use `rasterized=True` in matplotlib

Accessibility

Accessibility considerations

- Main ones are different kinds of colour-blindness (red-green etc.)
- Wikipedia: “The most common form is caused by a genetic condition called congenital red–green color blindness (including protan and deutan types), which affects up to 1 in 12 males (8%) and 1 in 200 females (0.5%).”
- Can use programs such as Sim-Daltonism on Macs to check for Color-blindness issues
- Other considerations are font-size and useful captions for visually-impaired people