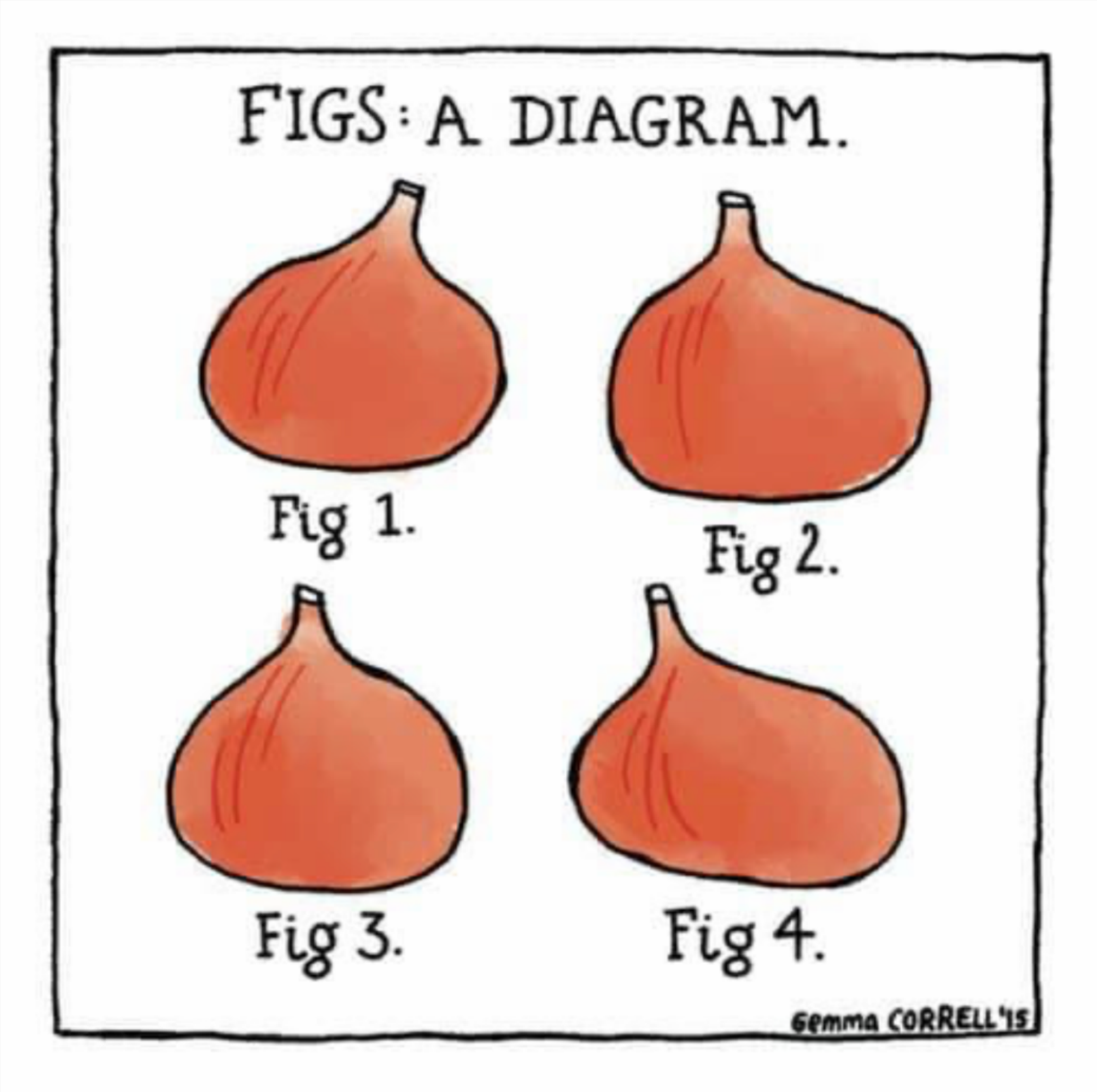
Academic Communication
in (Astro)Physics

Lecture 8: Figures & Tables

FIGURES & TABLES



General guidelines

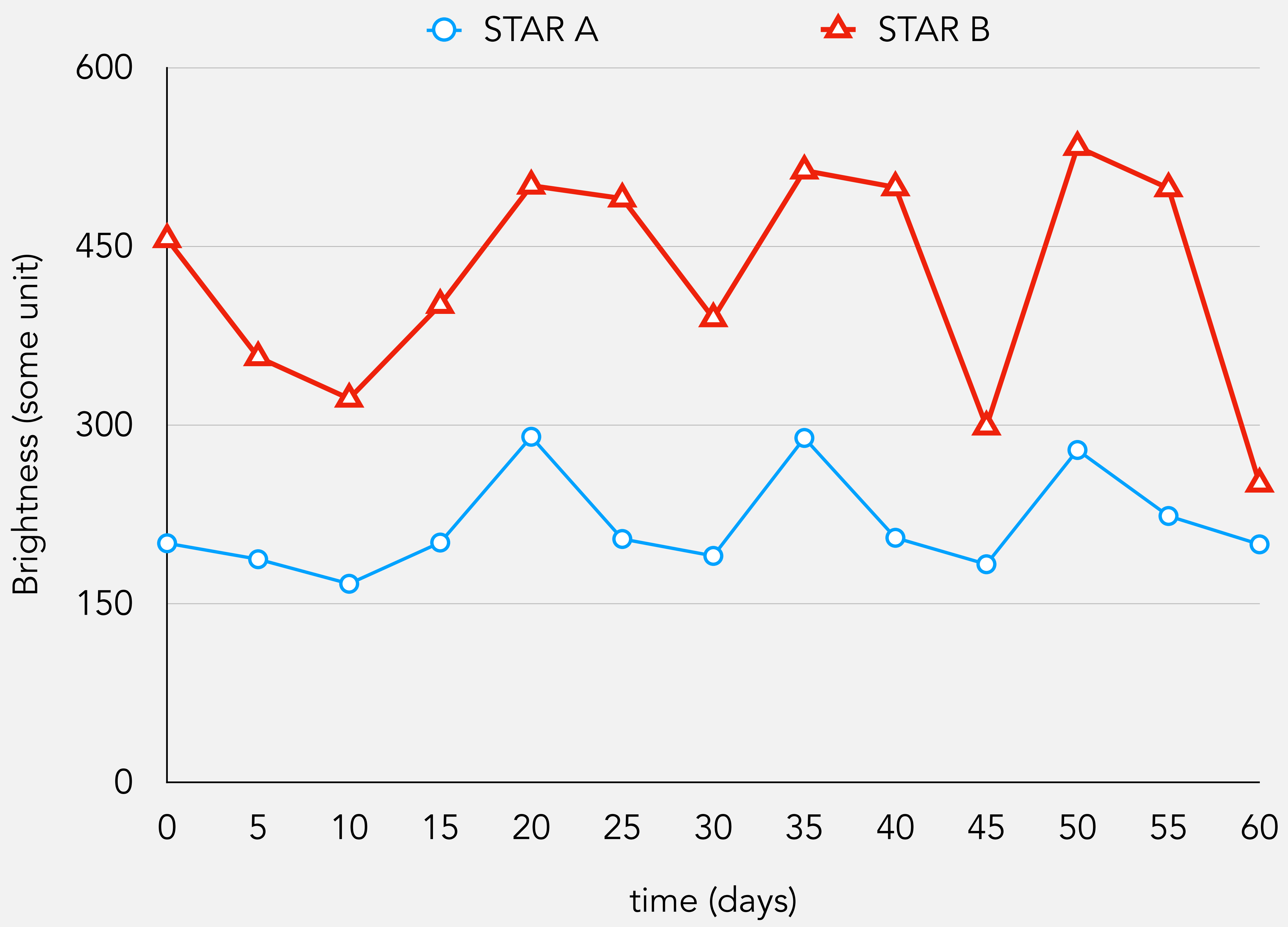
A PICTURE IS WORTH A THOUSAND WORDS.

- Decide whether to present data in a table, a figure, or in the text.
- Use the fewest figures and tables you need to tell a story (don't overwhelm the reader!)
- Design tables and figures to have a strong visual impact – should be immediately clear to the reader.
- Use them because they carry the point better than just text, not just because you can!
- Figures and tables should stand on their own (without referring to the text).
- Prepare with the reader in mind – place information where the reader expects to find it.
- It's a good idea to prepare figures early when writing a paper: 1) make figures, 2) decide which ones to keep and whether you need others, 3) draft captions.

Figure or Table?

- show trends/relationships
- capture complex datasets

- if exact values are important
- if there aren't enough data for a plot



TIME (DAYS)	STAR A	STAR B
0	200.5	455.8
5	187.1	356.7
10	166.5	321.9
15	201.1	400.6
20	289.9	500.7
25	204.1	489.9
30	189.9	389.4
35	288.9	513.4
40	205.1	499.3
45	182.9	298.5
50	278.8	533.2
55	223.4	498.5
60	199.6	250.6

Professional figures

- Follow the instructions for authors of your target journal (e.g., AAS journals, https://journals.aas.org/manuscript-preparation/#figures)

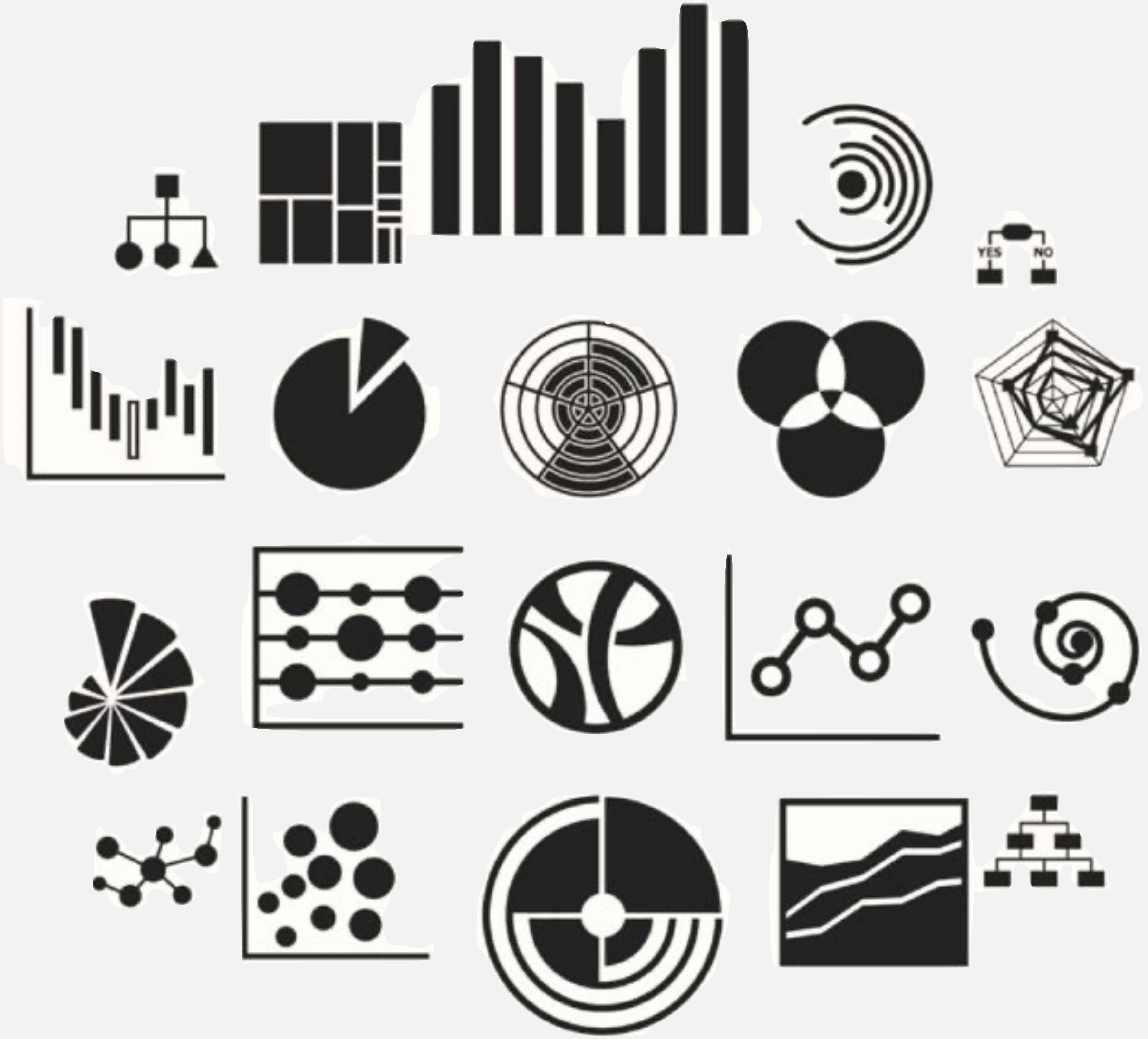
- Use the right software (e.g., matplotlib, IDL, etc)
  - Excel, Topcat: great to quickly look at data while doing your research, but they don't produce high enough quality figures for professional journals: avoid.
  - sometimes Adobe Illustrator can be useful for retouching colours, making diagrams etc.

- Design carefully:
  - use vector graphics when possible
  - good contrast against (plain white) background
  - symbols and annotations must be clear; label and identify all components
  - include scale bars in images
  - don't include grid lines
  - avoid clutter
  - should be aesthetically pleasing (or at least not the opposite)

Be objective and honest when making plots!

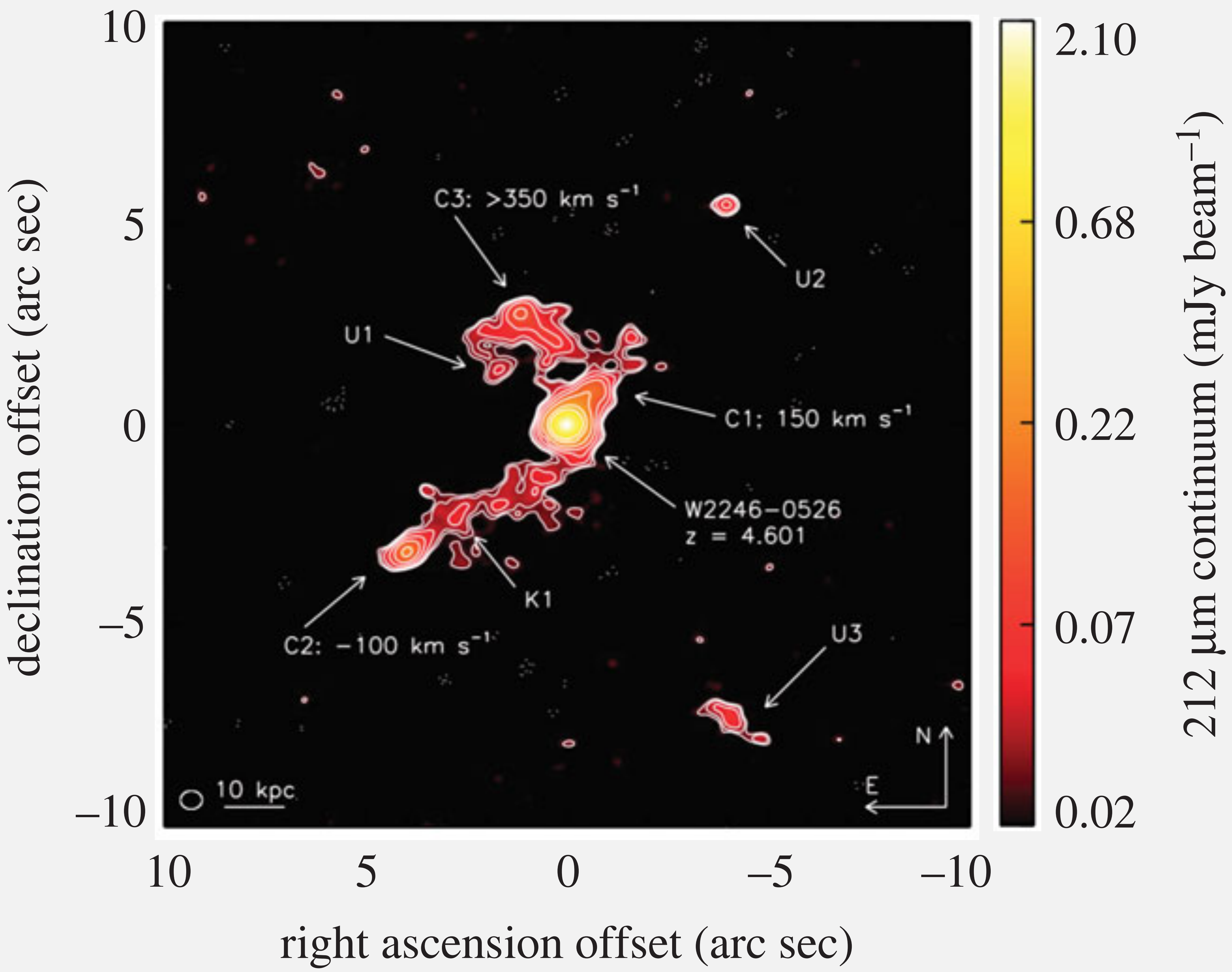
The data you present may be interpreted in more than one way. You can highlight, exaggerate, or even misrepresent a set of data depending on the way you chose to display it. Thus, making a plot is more than just plotting points: you need to understand the rationale behind the quantitative methods, and **be as objective and honest as possible.**

- do not mislead the reader! (e.g., deleting data points, massaging line fits)
- provide statistical information in figures and legends to lend validity to your data
- plot error bars (if appropriate)
- show variability of data points (e.g., scatter plots or box plots)
- indicate statistical tests used
- choose scales carefully, and mark them clearly (e.g., ensure identical scales to allow readers to compare data directly, arrange plots next to each other)



Types of Figures

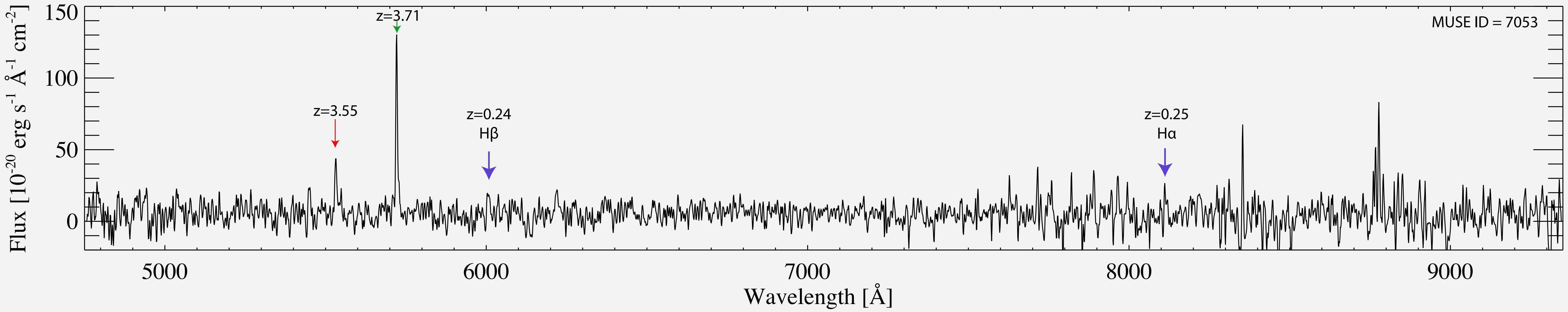
Images and spectra



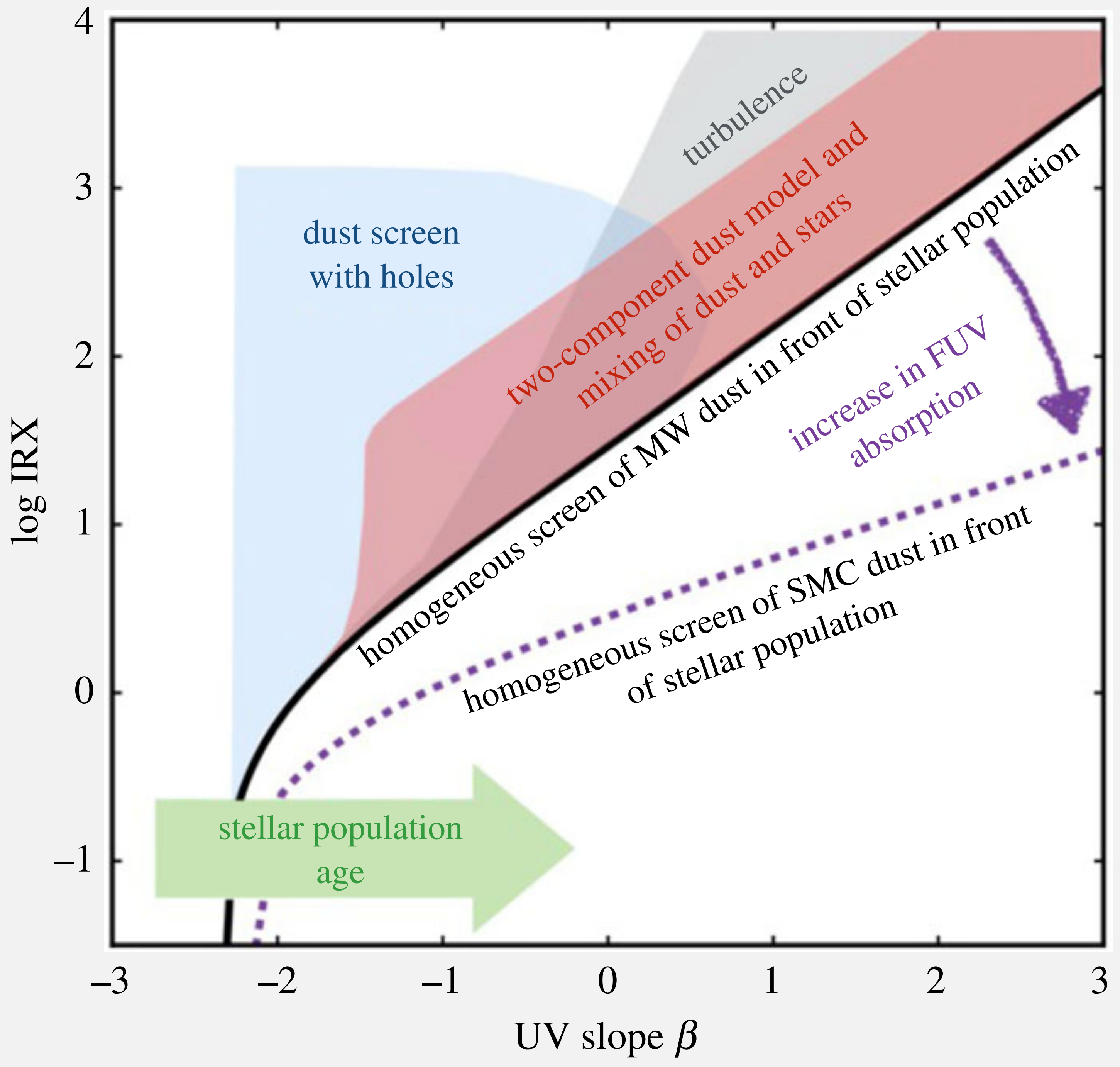
Diaz-Santos et al., 2018

- show actual data
- make it best quality possible: sharply focused, good contrast etc
- label everything that is relevant
- always include scale bars (or state scale very clearly in the caption)

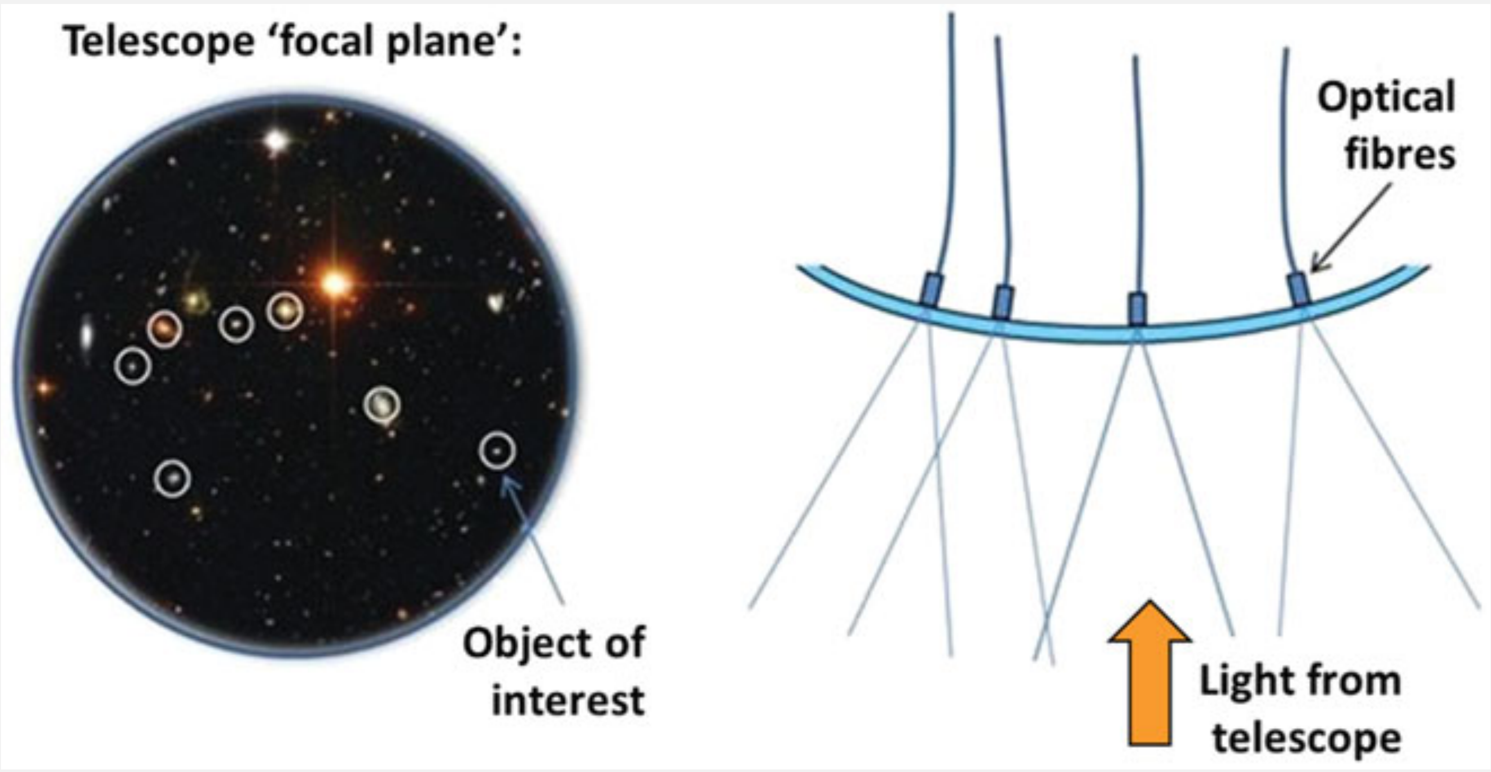
Brinchmann et al., 2017



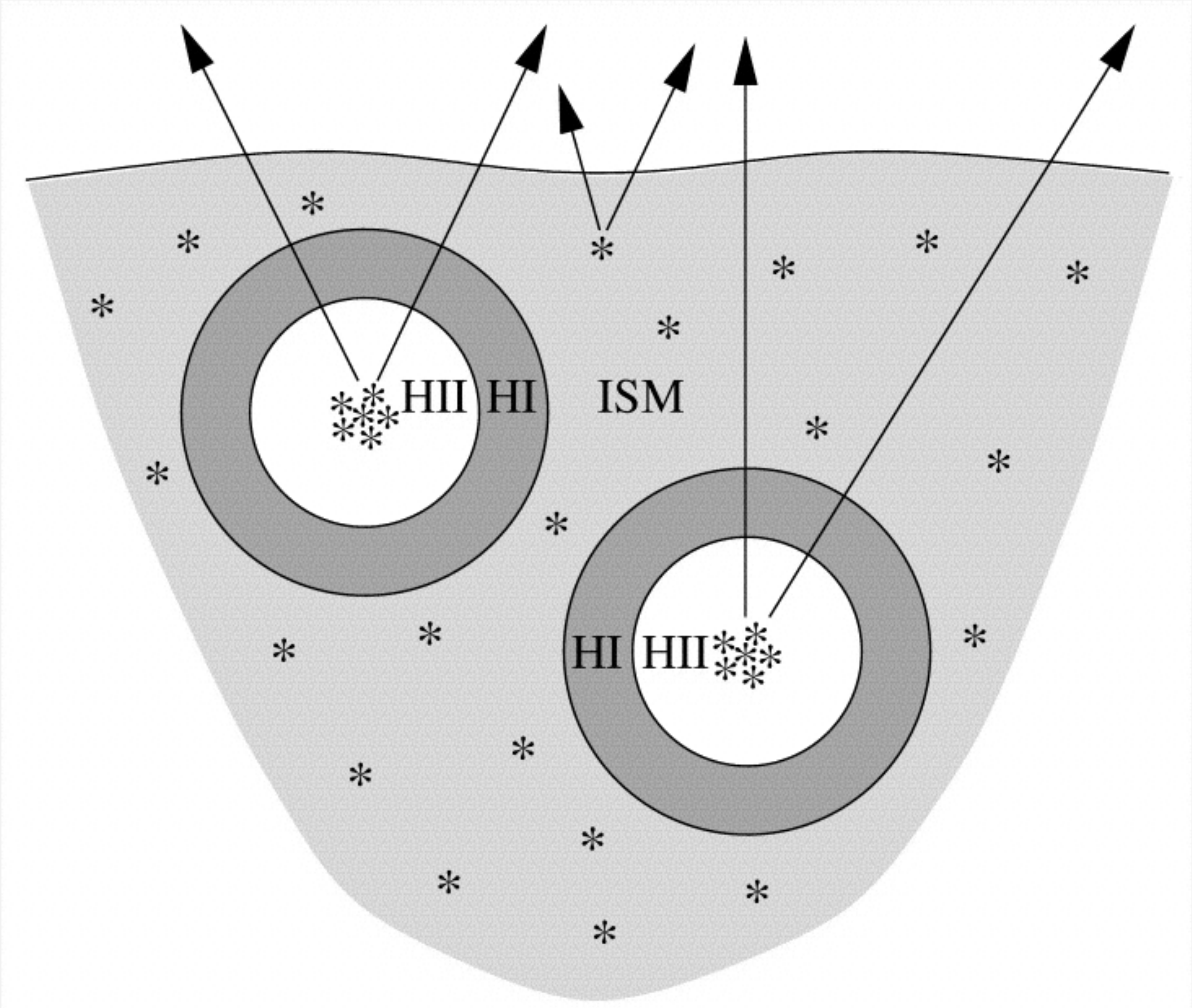
Drawings and diagrams



Popping et al. (2017)



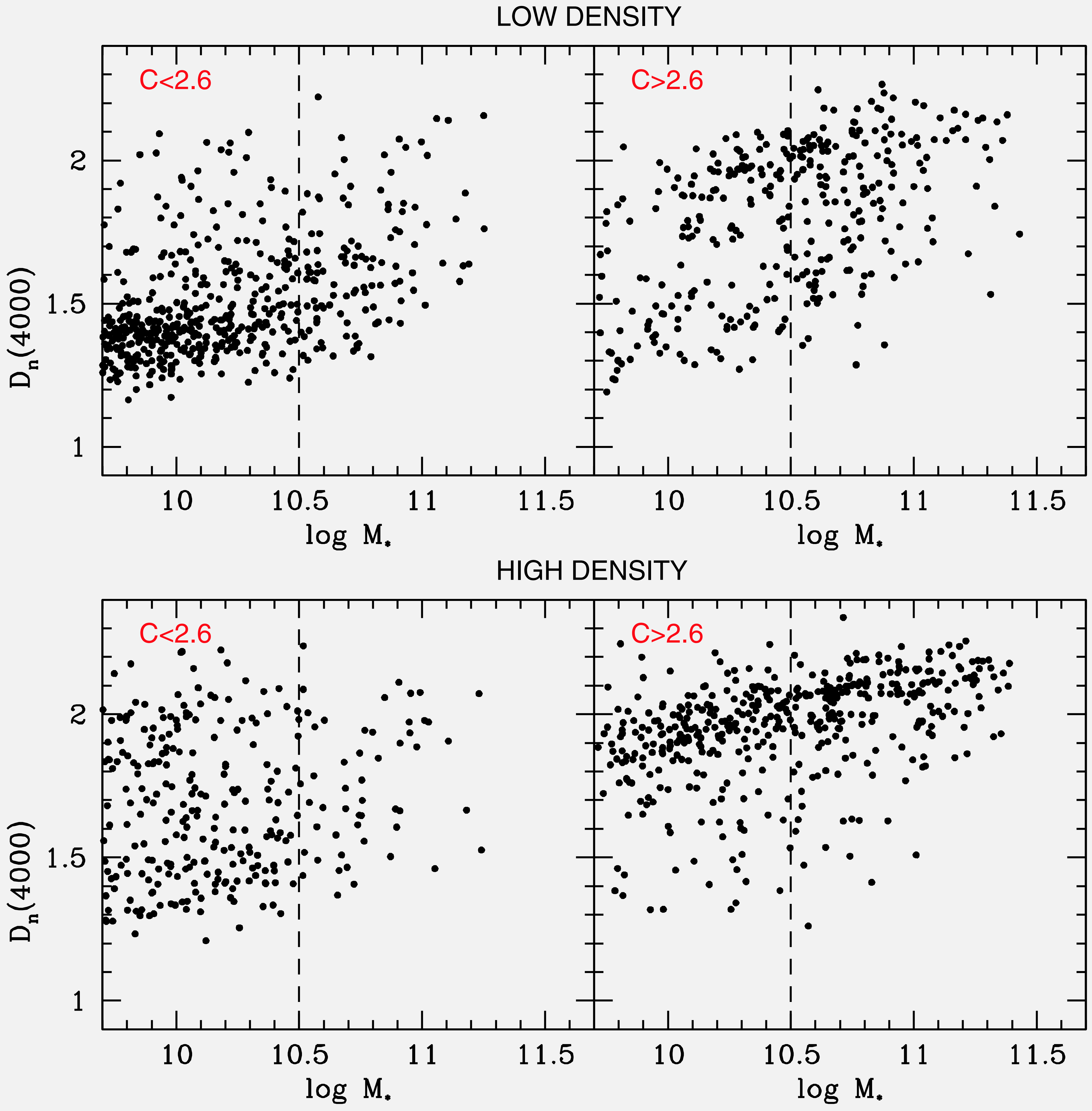
da Cunha et al. (2017)



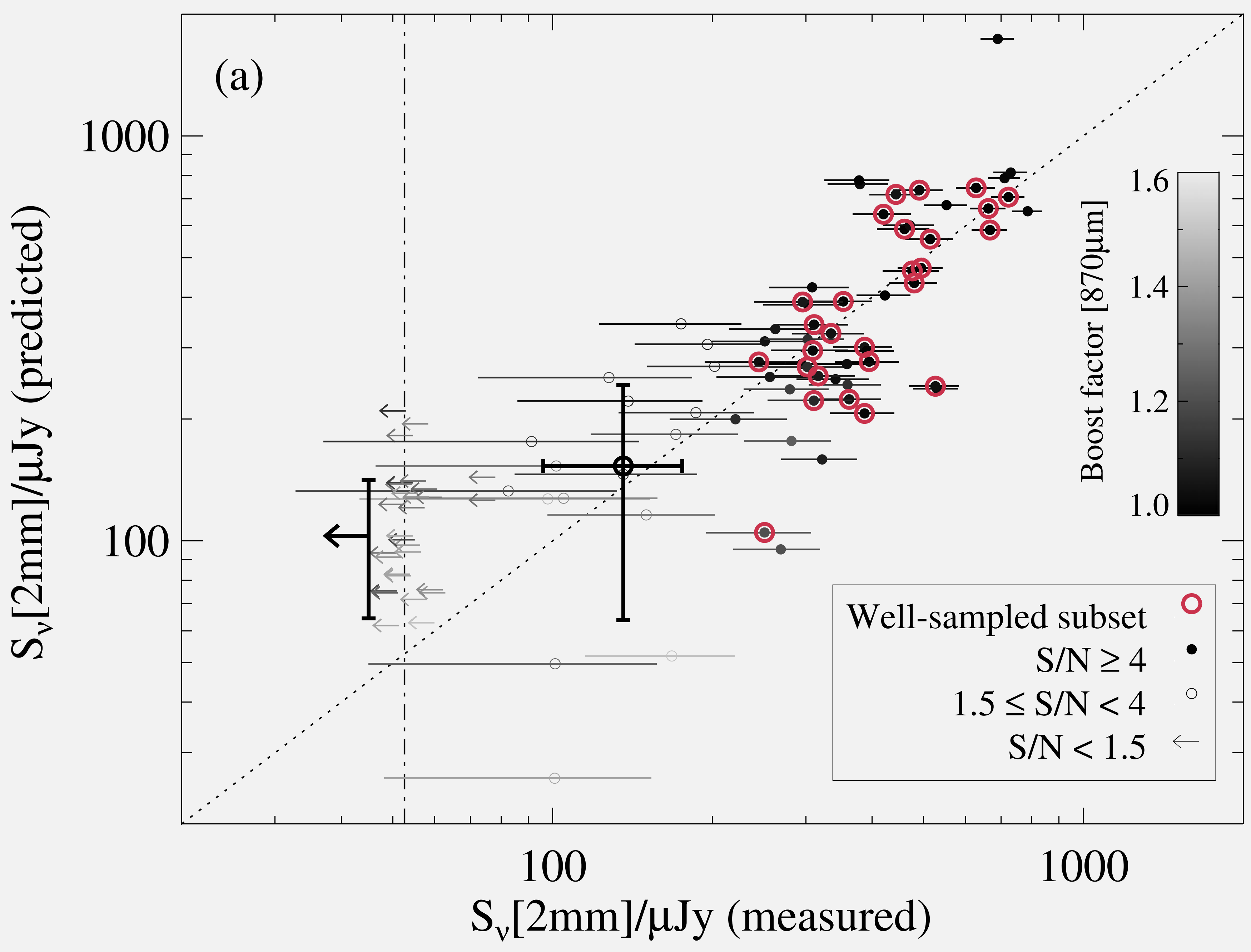
Charlot & Fall (2000)

- illustrate basic principles or otherwise explain text material
- includes flow charts, diagrams, line art, etc
- make them look as polished and professional as possible (Adobe Illustrator skills might come in handy)
- advantage: present unusual perspectives while controlling the amount of detail
- ensure key features are immediately visible
- keep them as simple as possible

Scatter plots



Kauffmann et al. (2004)

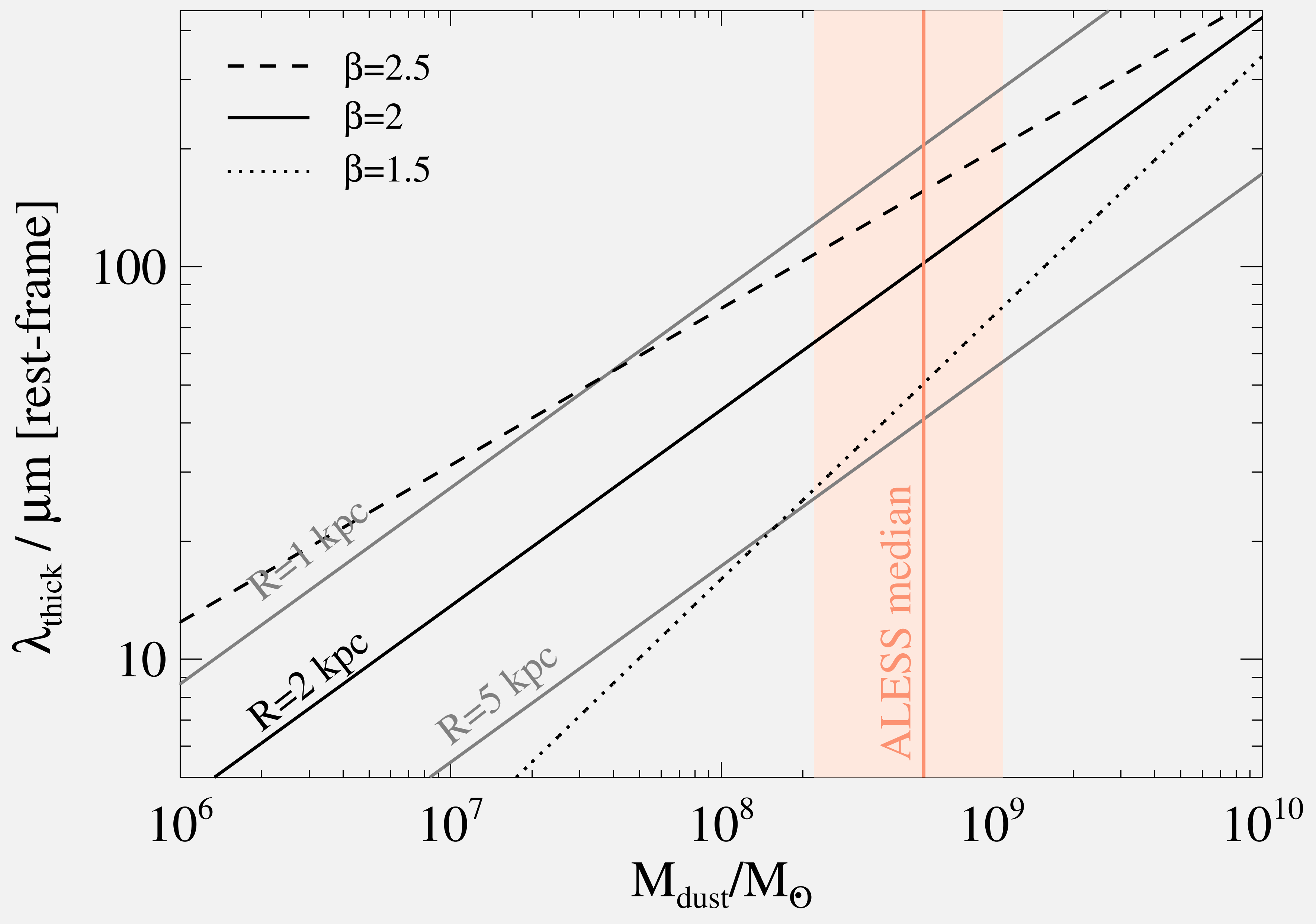


da Cunha et al. (2021)

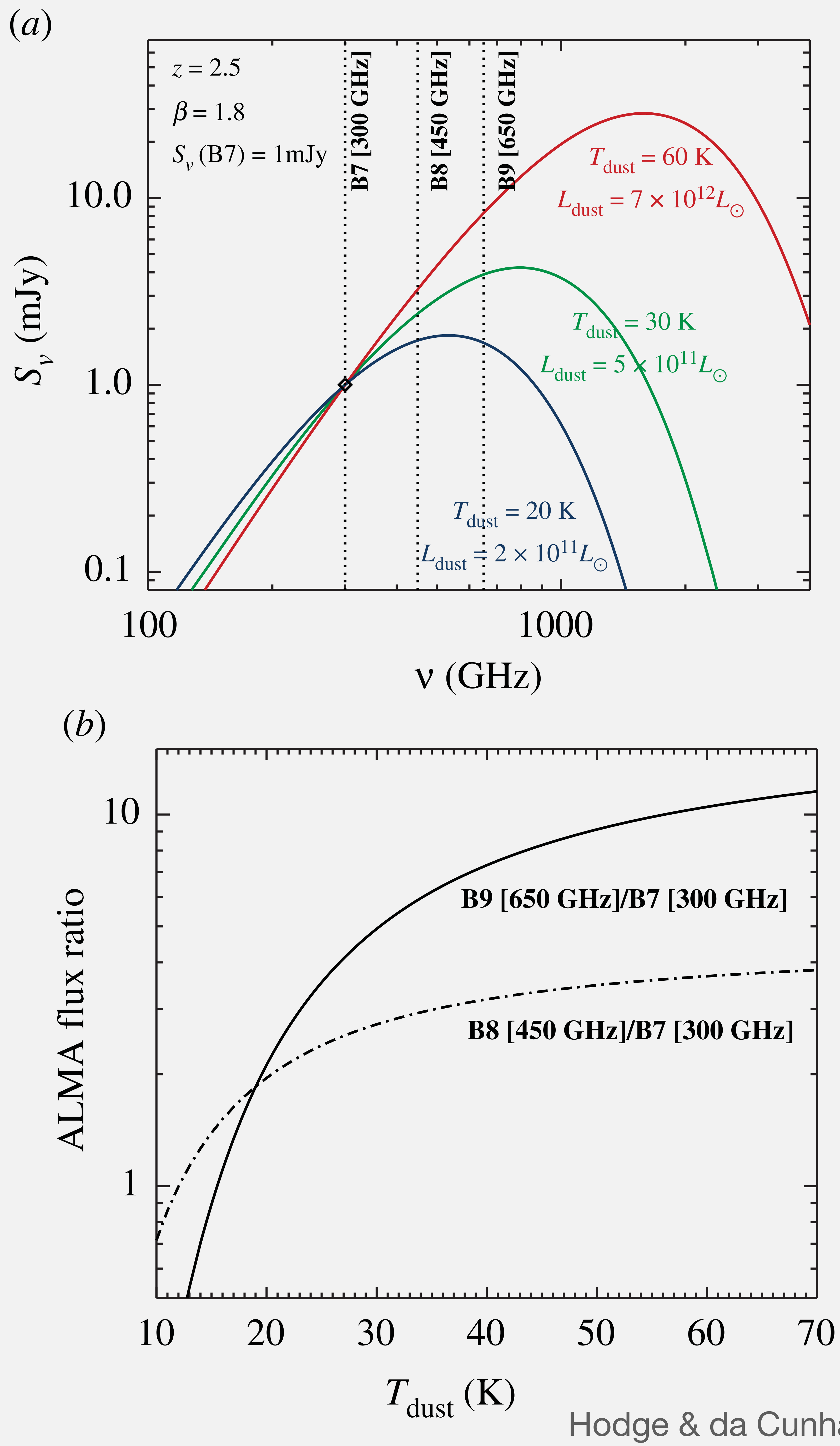
- show/find how variables are correlated
- line fit to data if appropriate
- can be produced in 2 or 3 dimensions
- don't forget the error bars!

Line plots

- compare different functions etc
- do not cram too much into one figure (ideally 3-4 lines max)
- use different line colours, patterns, thicknesses to distinguish different lines
- use lines to connect data points or to show fits to the points.



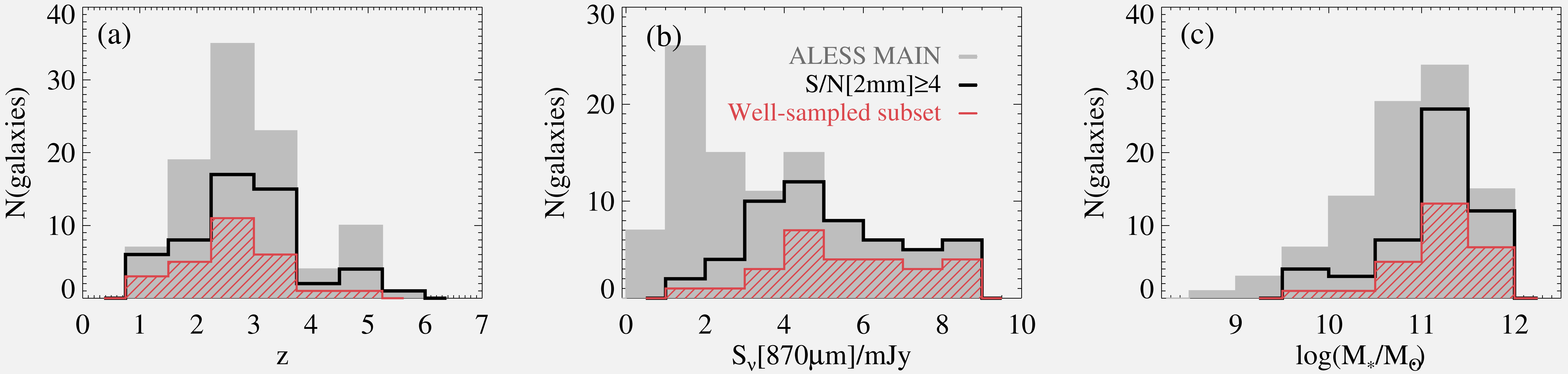
da Cunha et al. (2021)



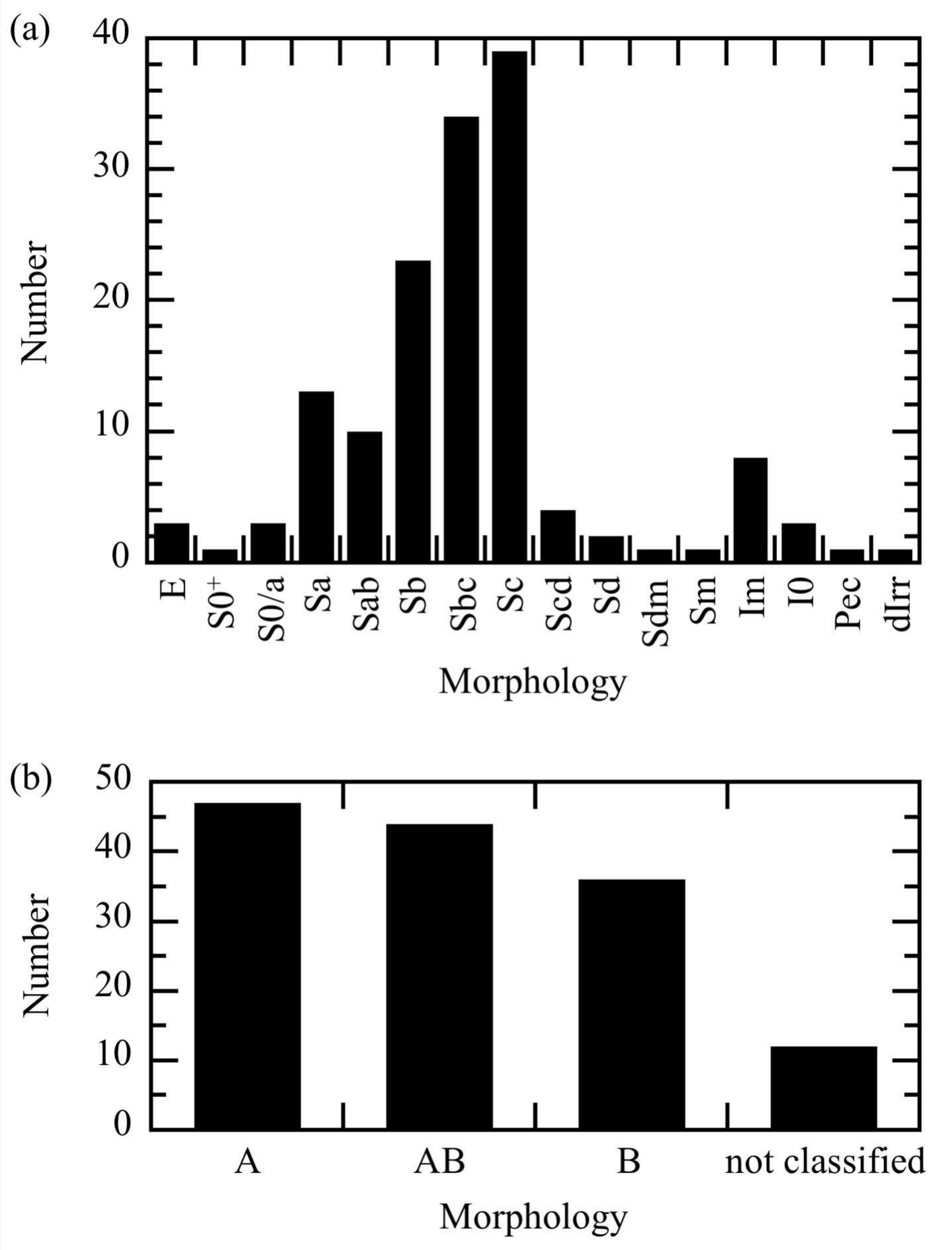
Hodge & da Cunha (2020)

Bar graphs / histograms

- when findings can be subdivided and compared
- usually more effective with general audiences
- use vertical rather than horizontal
- optimize number of bins: too many – histogram is noisy; too few – information is obscured.



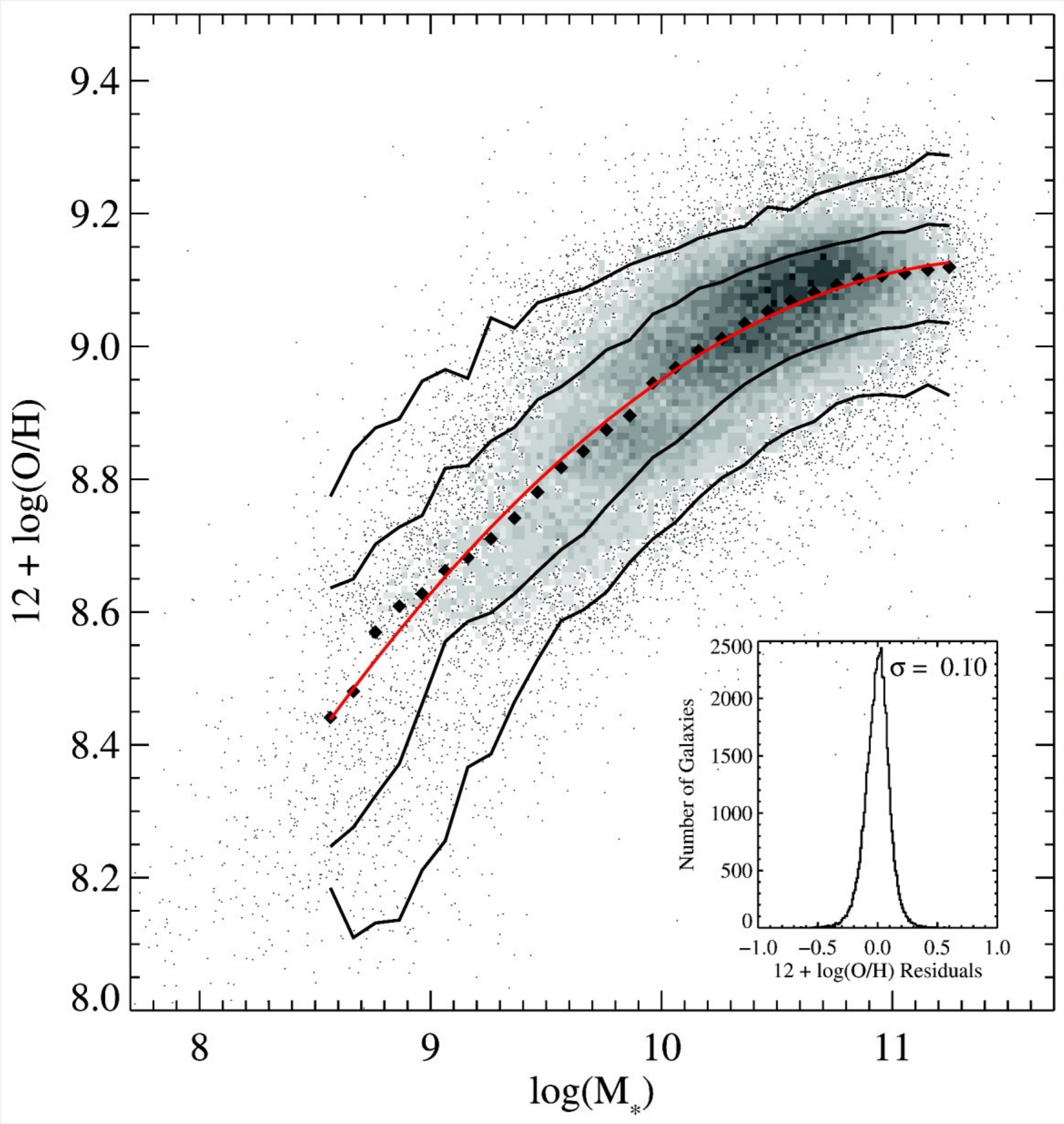
da Cunha et al. (2021)



Sorai et al. (2019)

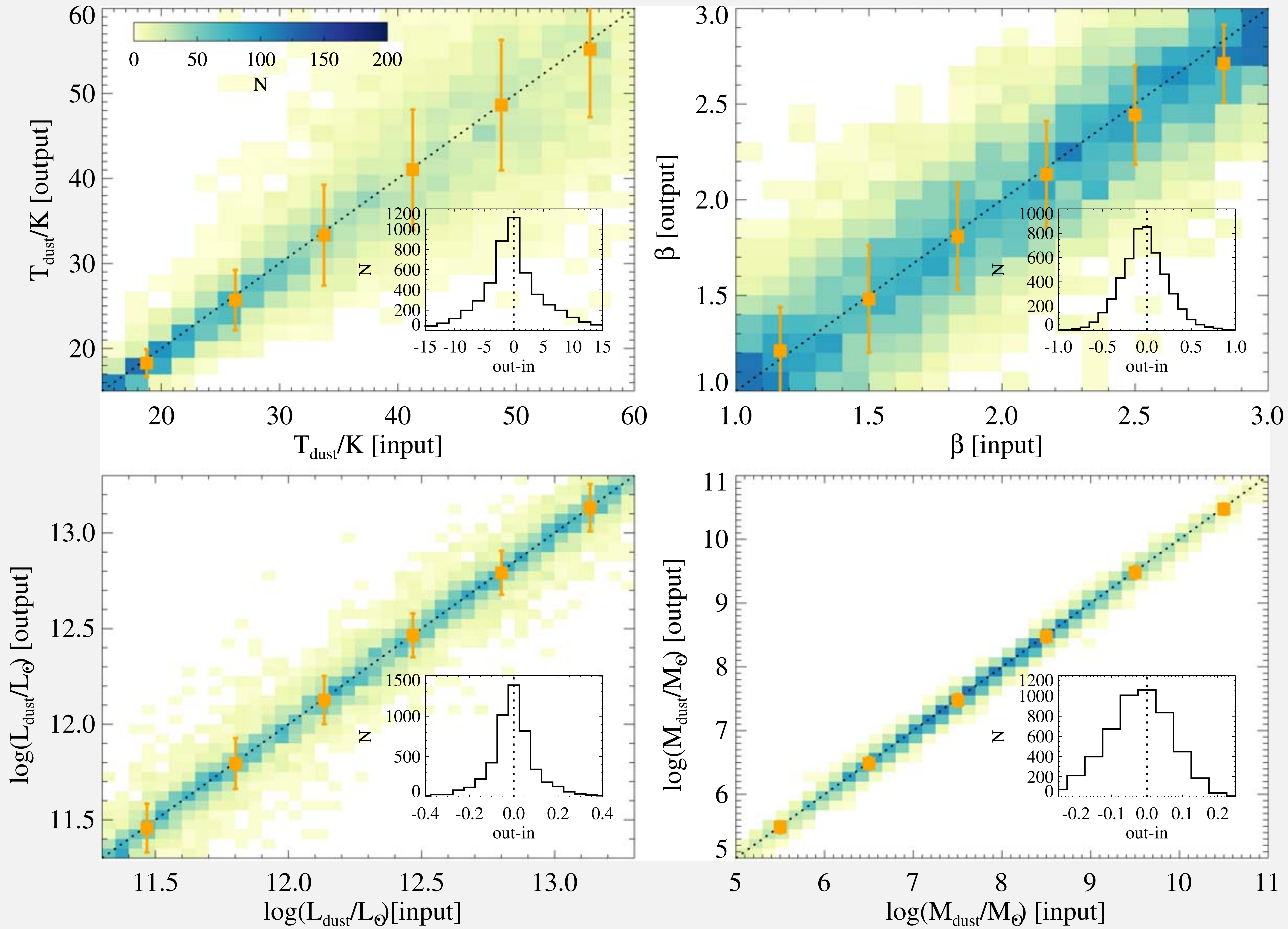
2D histograms/contours

- show the 3rd dimension of data point density
- nice way to plot large datasets



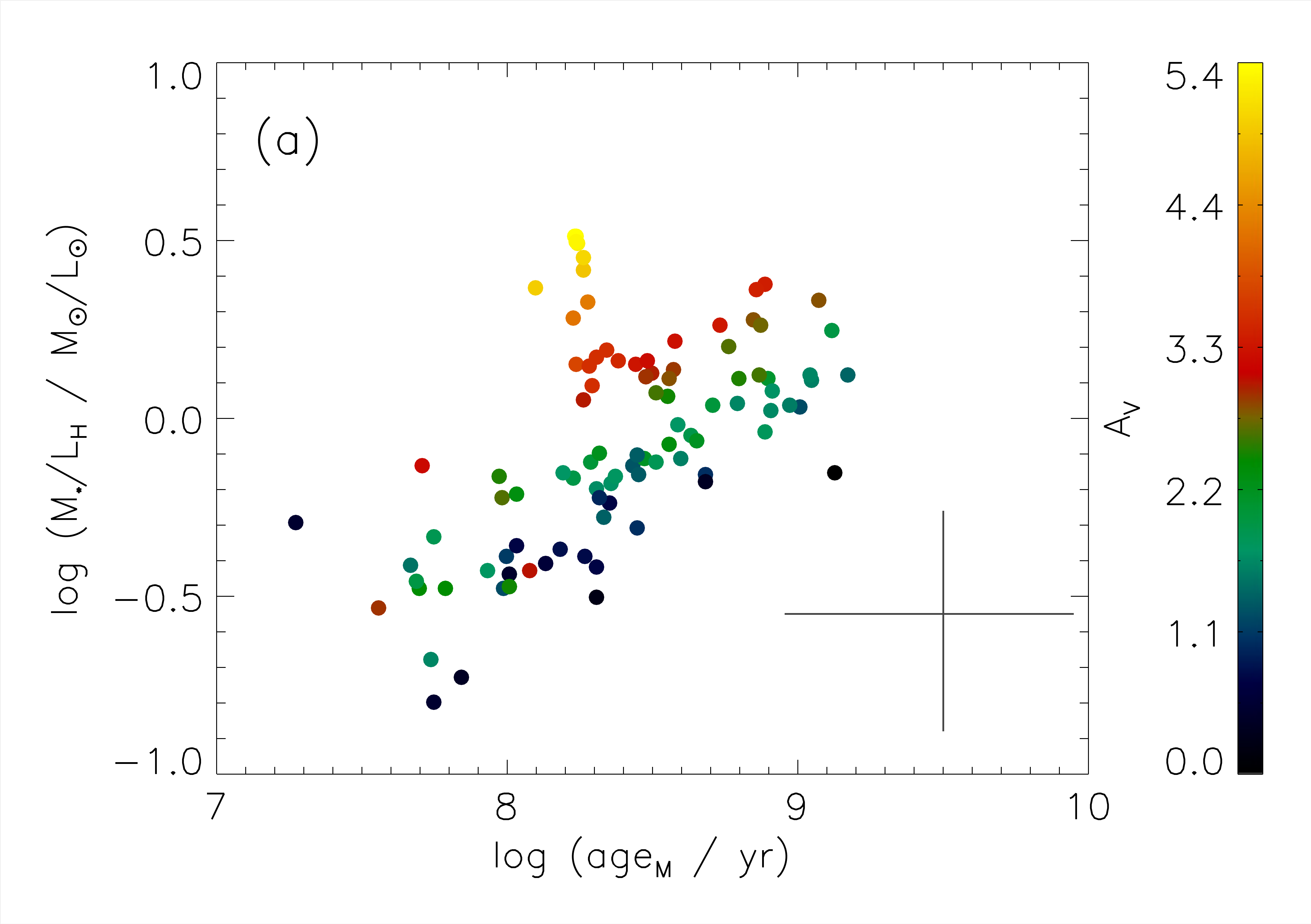
Tremonti et al. (2004)

da Cunha et al. (2021)

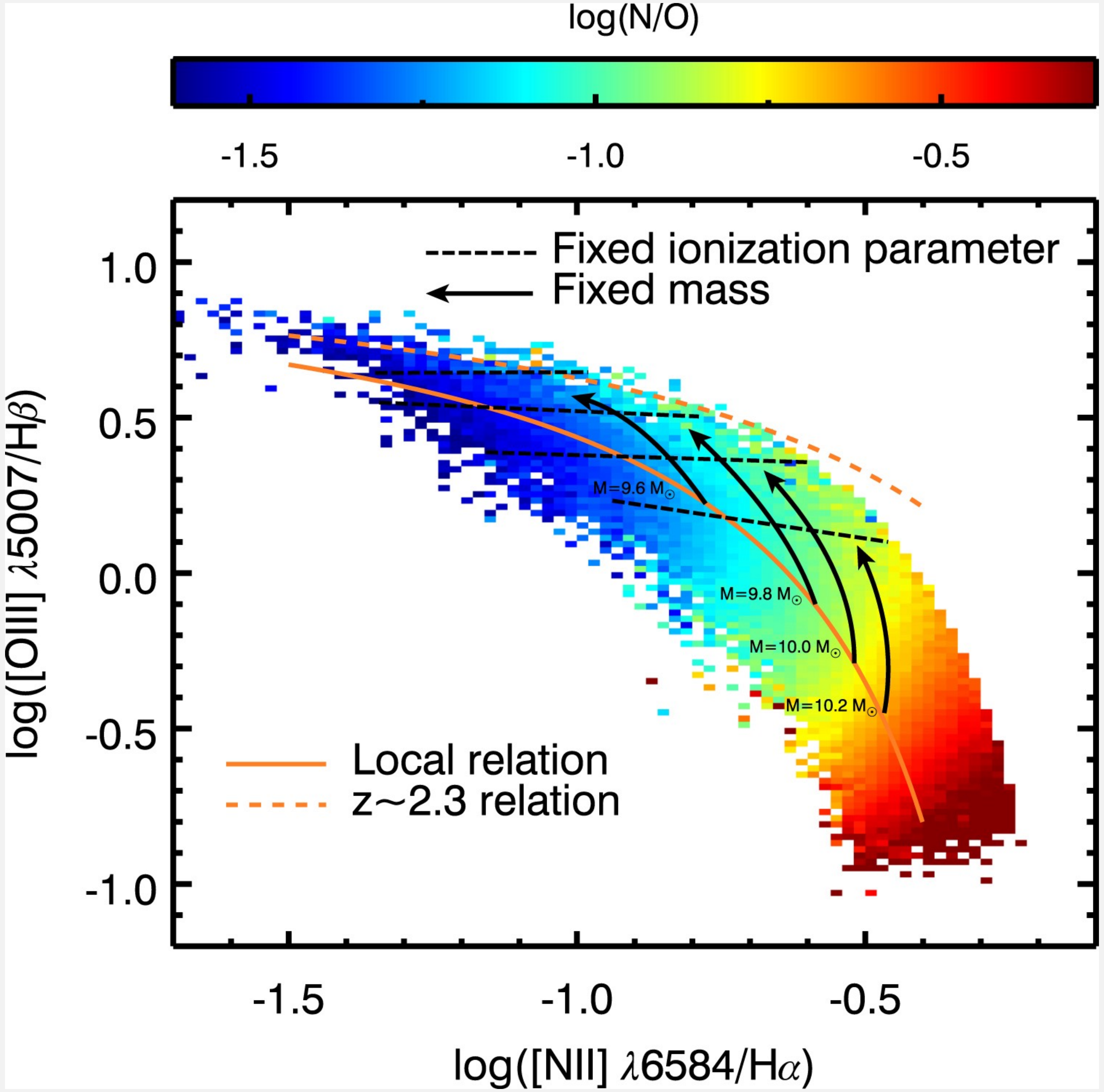


Colour maps

- indicate a 3rd variable on a 2D plot
- choose your colour map well



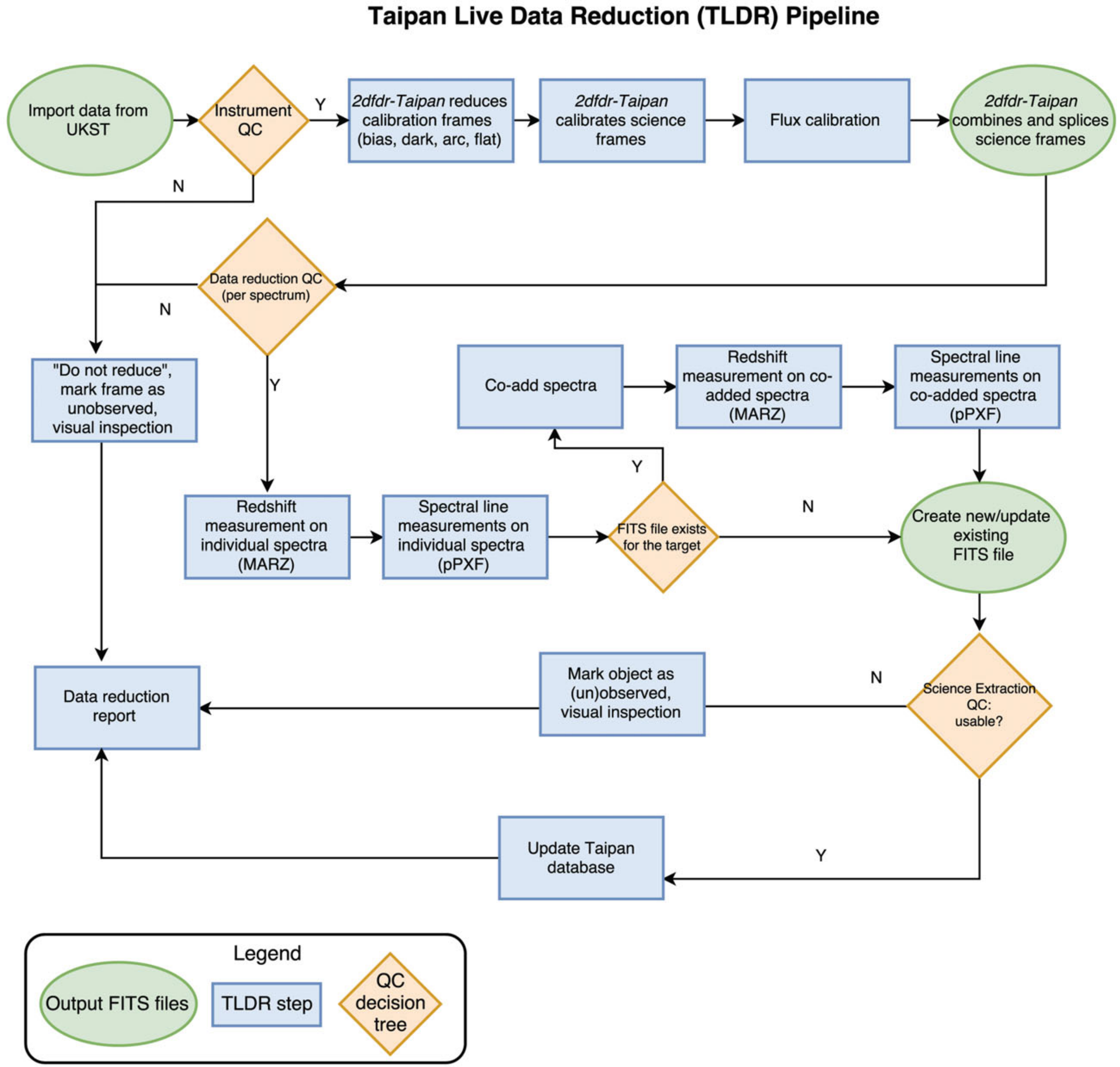
da Cunha et al. (2015)



Masters et al. (2016)

Flowcharts

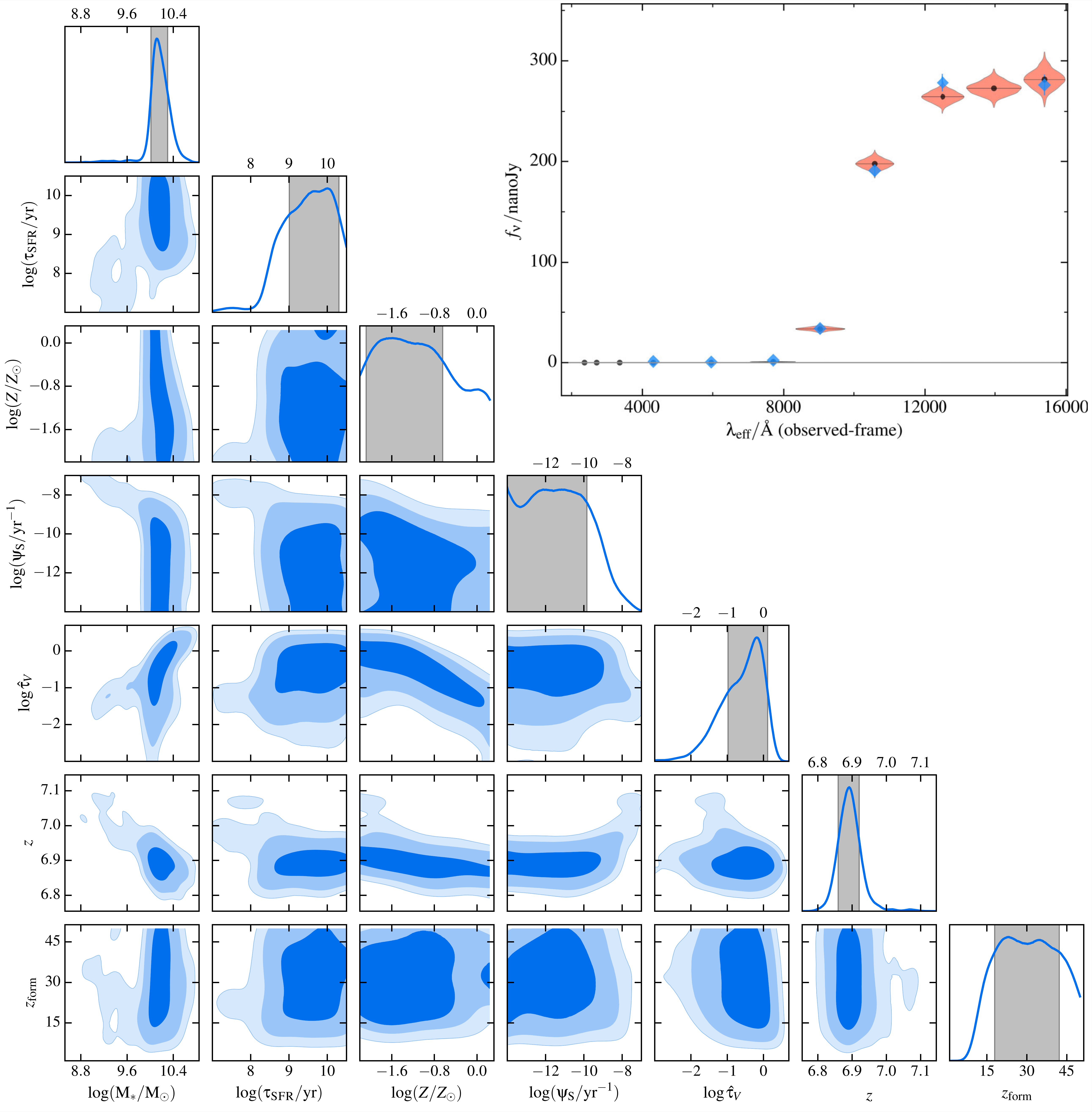
- very good for describing algorithms and other methods



da Cunha et al. (2017)

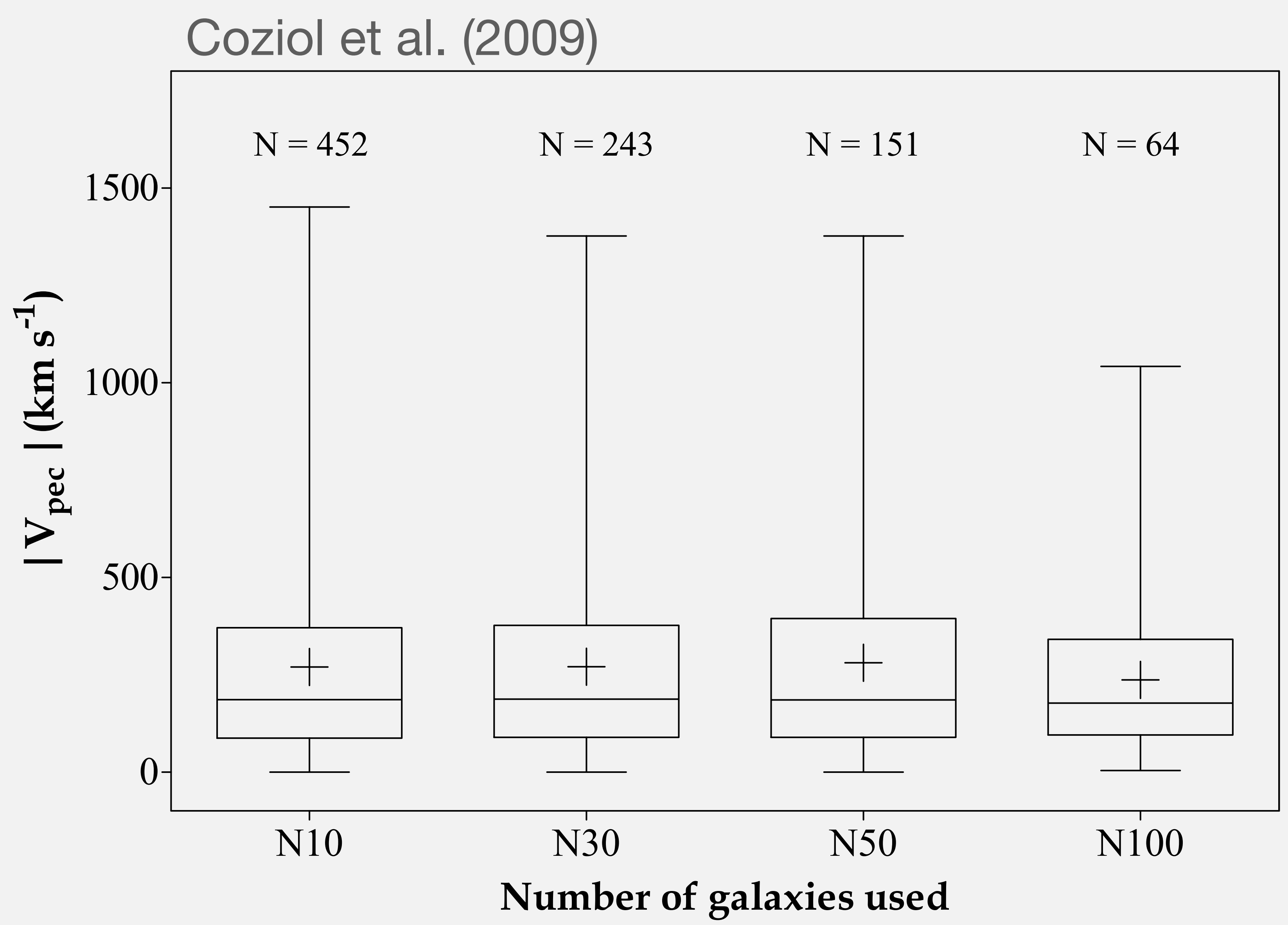
Corner plots

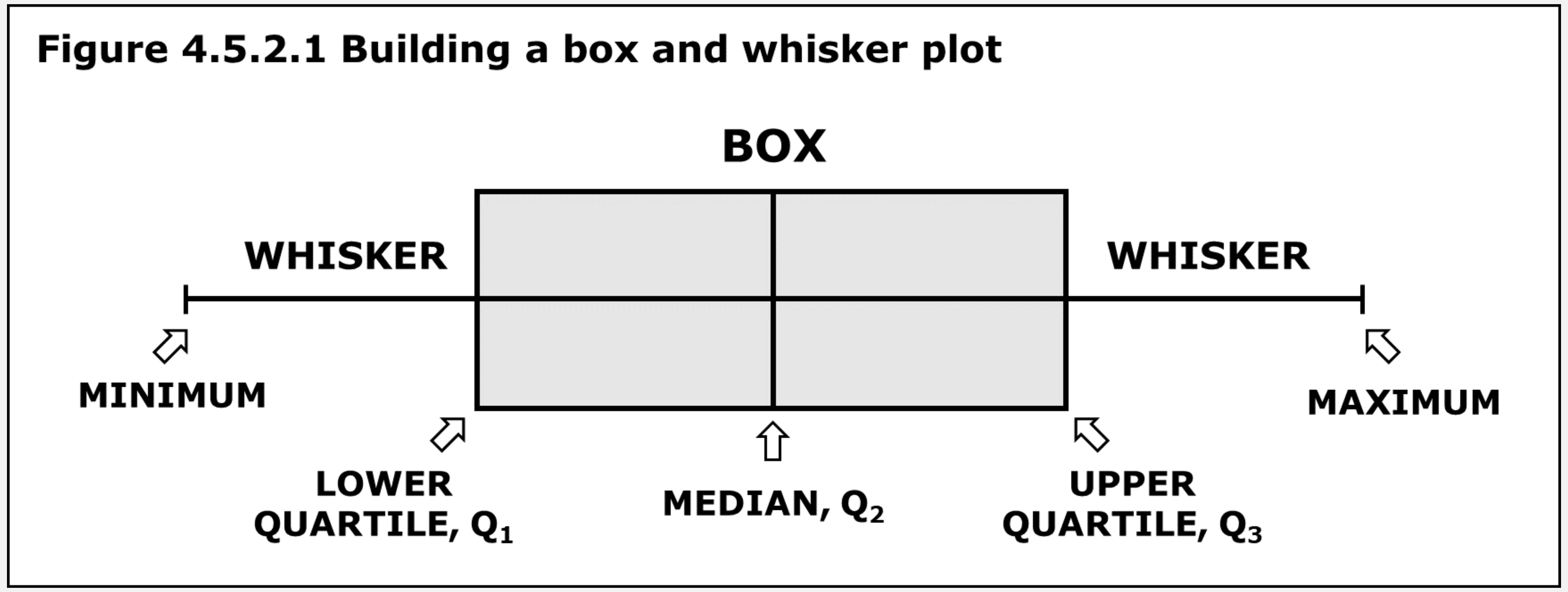
- good to look at likelihood distributions
- highlight co-variance between pairs of parameters

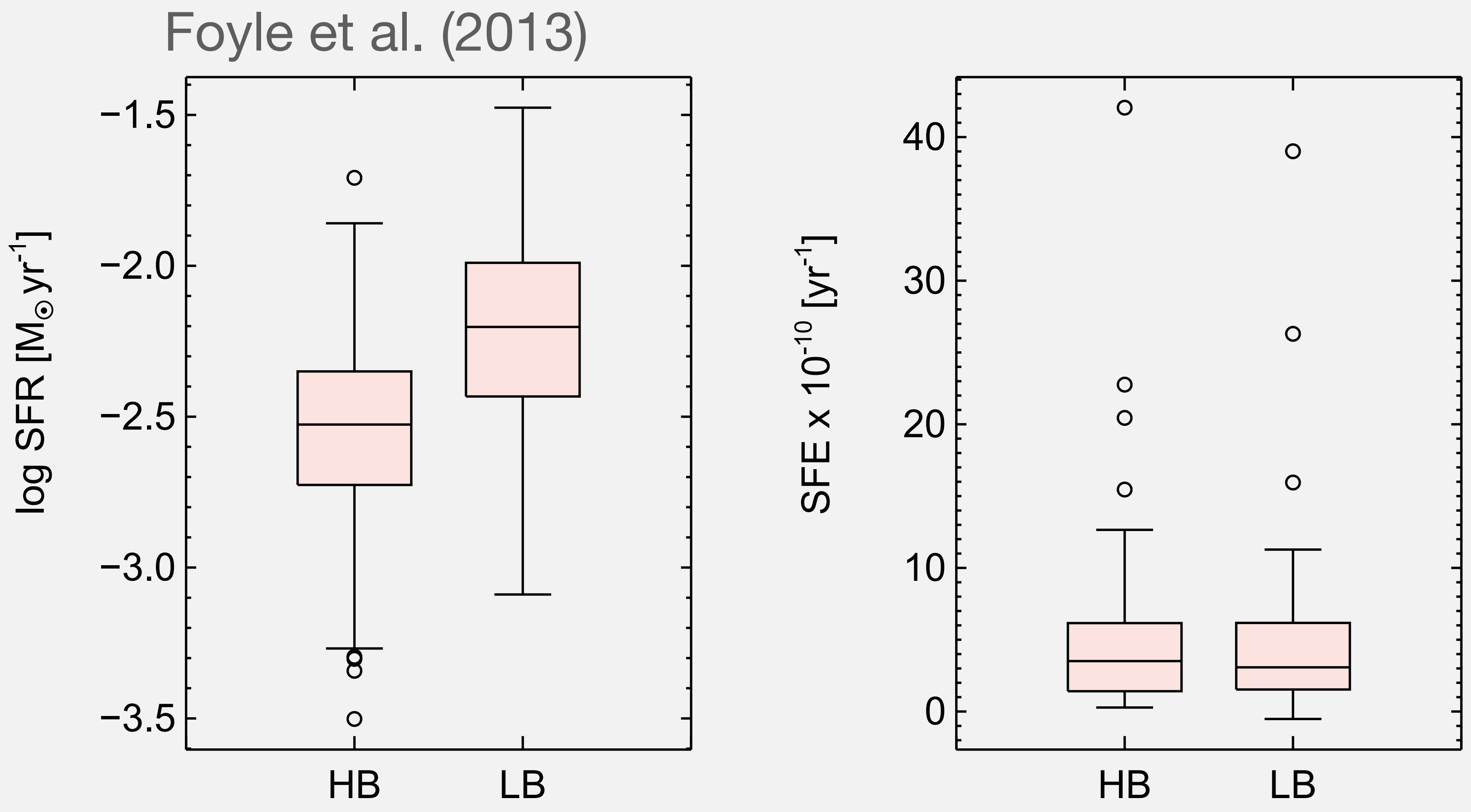


Chevallard & Charlot (2016)

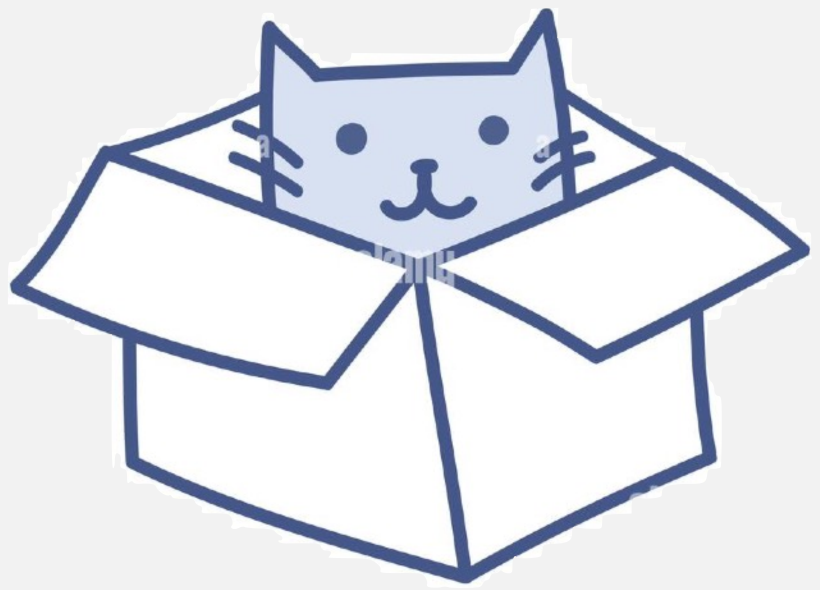
Box & whisker plots







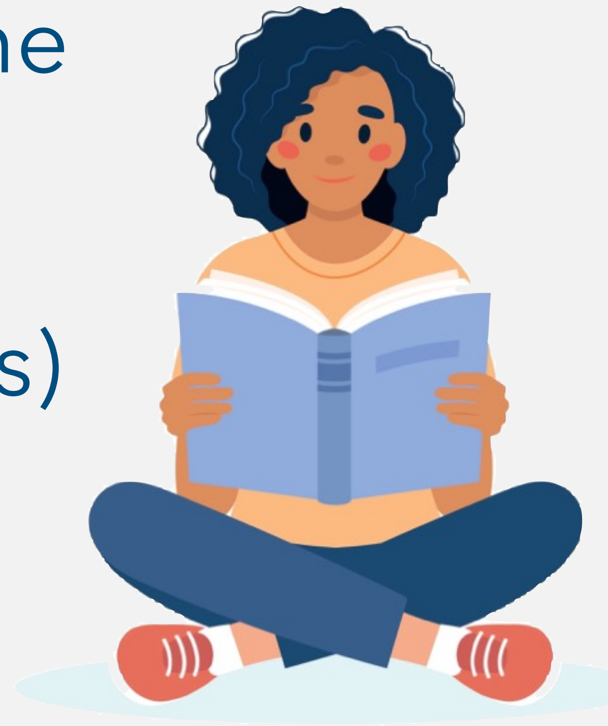
- good to display differences between data sets, especially in descriptive statistics
- divides data into quartiles, visually depicts central value, variability, and range of the data set
- can be vertical or horizontal



Format graphs with the reader in mind

1) y versus x

- Place the independent variable at the x-axis and the dependent variable at the y-axis
- Make key information immediately obvious; highlight important information using different colours, line weights, arrows, call-outs, font size, or labels.
- De-emphasize less important information such as axes, axis labels, keys, titles, and figure labels.
- Make plotted points stand out well.
- Draft short but informative descriptions for the axes.
- Be consistent: use the same symbols when the same entities occur in several figures, and use the same coordinates for different figures if values in them are to be compared.
- convention is to say "y label vs x label" or "dependent vs independent" (e.g., metallicity vs stellar mass)



Format graphs with the reader in mind

2) easy to read

- lettering size should be big enough to read after the graph is reduced (no smaller than 8pt after reduction) – check legibility by printing the figure in publication size.
- symbols should be ~3–5x the width of curves; avoid "x", "+", "0", or "*" as these symbols aren't distinctive enough – circles, triangles, and squares are good. If possible, keep similarly shaped symbols and colours separate.
- differentiate curves by using different symbols for points joined by the same type of line, or by using different types of lines.
- do not clutter: if there's no room for labels or a key, define in a separate legend. Avoid grid lines.
- make the lines of error bars thinner than other lines in the plot, and do not let them overlap.
- rather than using a key, label curves directly when possible; keep labels well away from lines and curves, and position them horizontally.

Format graphs with the reader in mind

3) axes and scales

- label axes as briefly and simply as possible; lower case better than all caps
- always give units! (my preference is as a fraction rather than parentheses)
- choose scales carefully
- mark tick marks clearly; use only as many as necessary for clarity
- use a log scale when your data cover a large range over many powers of 10 (often the case in Astronomy!)– creates a more uniform distribution of data points, easier to see the data. Think about the dynamic range, but also about the convention for certain quantities
- multi-panel plots: consider if vertical/horizontal scales should be the same (to make comparisons easier)

Figure captions

- always include a caption, except in slides
- caption should objectively describe the content of the figure
- the figure + caption should be able to stand alone
- always include:
    - **a title**
    - **description of figure contents**
    - **explanation of symbols and abbreviations (written in telegraph style)**
    - **statistical and experimental details as appropriate and required**
- the title should be brief; often, state important results depicted by the figure
- description of content: avoid "see text for explanation"; use consistent and parallel wording for similar illustrations

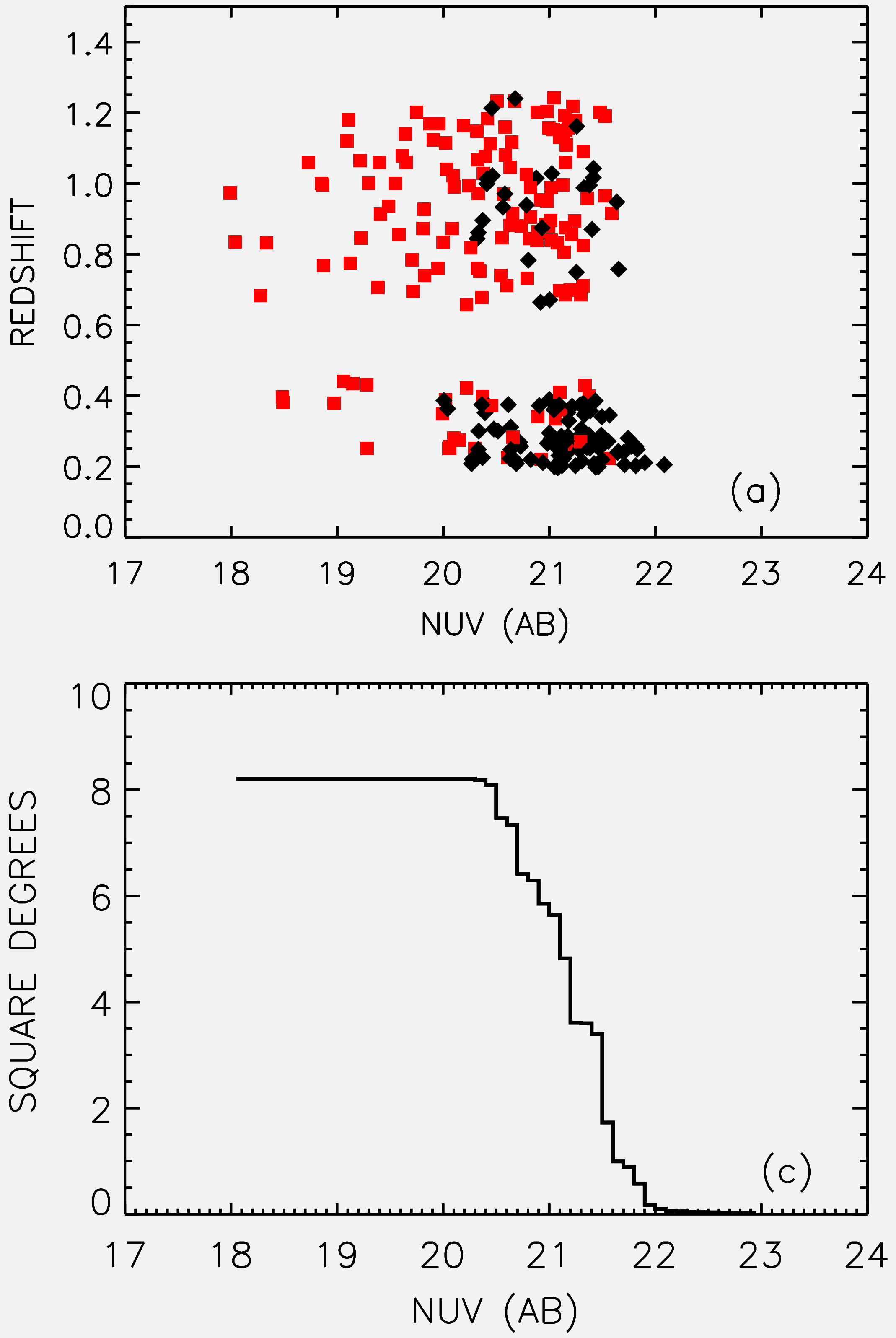


Figure 4. (a) Redshift vs. NUV magnitude for our Lyα selected sample from all nine *GALEX* fields. Sources classified as AGNs based on high-excitation lines in the spectra (otherwise) are shown as red squares (black diamonds). The depths of the fields range from NUV(AB) = 20.5 in HDFN 00 to 21.8 in GROTH 00. (b) Redshift vs. FUV magnitude with symbols as in (a). Here sources fainter than 24.7 are shown at a nominal magnitude of 24.7. (c) Total observed area for the nine fields vs. NUV magnitude.

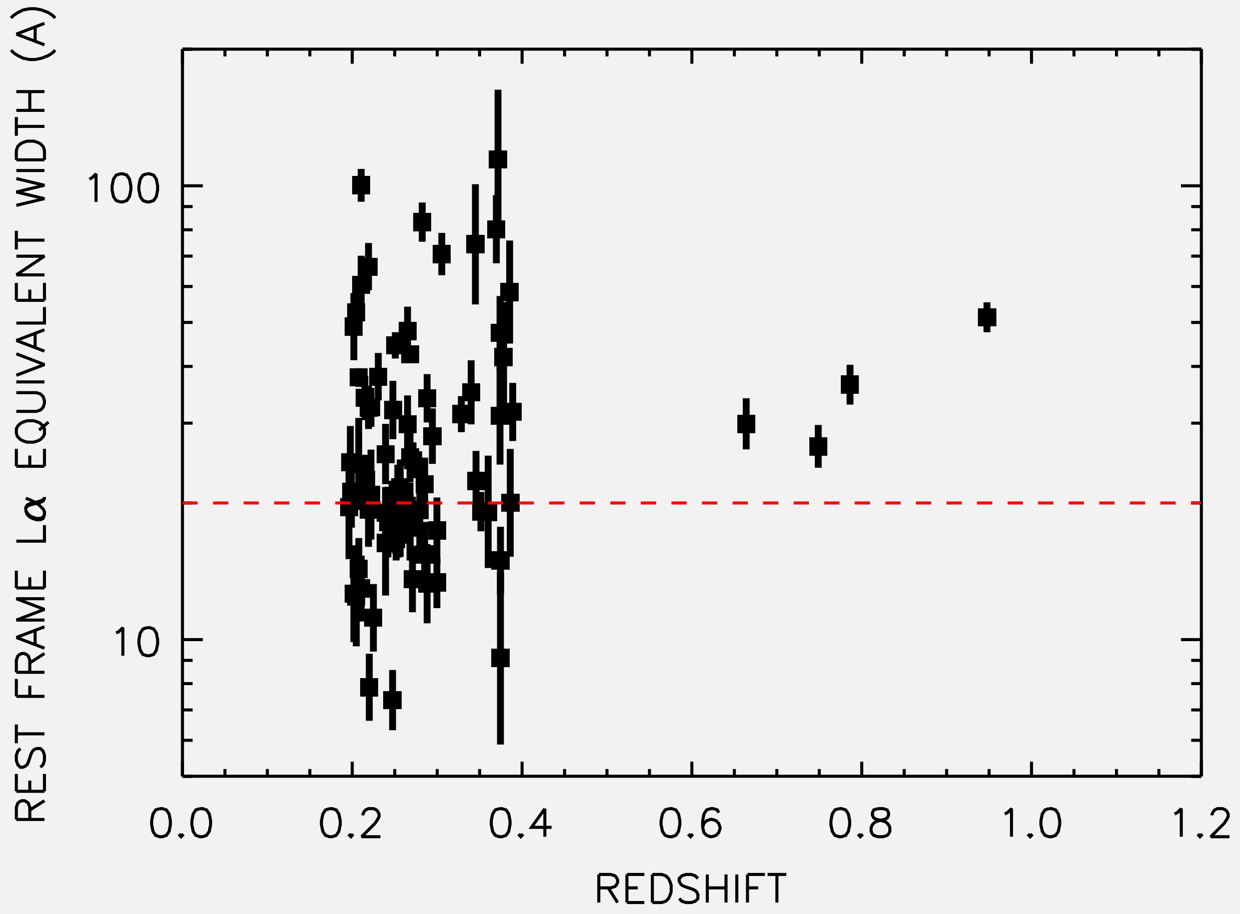


Figure 12. Rest-frame EW(Lyα) for the candidate Lyα Galaxies vs. redshift. This consists of objects with no high-excitation lines in the UV and with widths less than 15 Å in the low-redshift interval and less than 30 Å in the moderate-redshift interval. Objects with FUV − NUV < 1.8 are also excluded in the moderate-redshift interval. The final LAE sample is taken to be those sources with a rest-frame EW(Lyα) > 20 Å (red dashed line). The error bars show the $\pm 1\sigma$ errors from the formal statistical fit of the Gaussian and baseline. As discussed in the text, there may be comparable systematic errors resulting from the choice of fitting procedure.

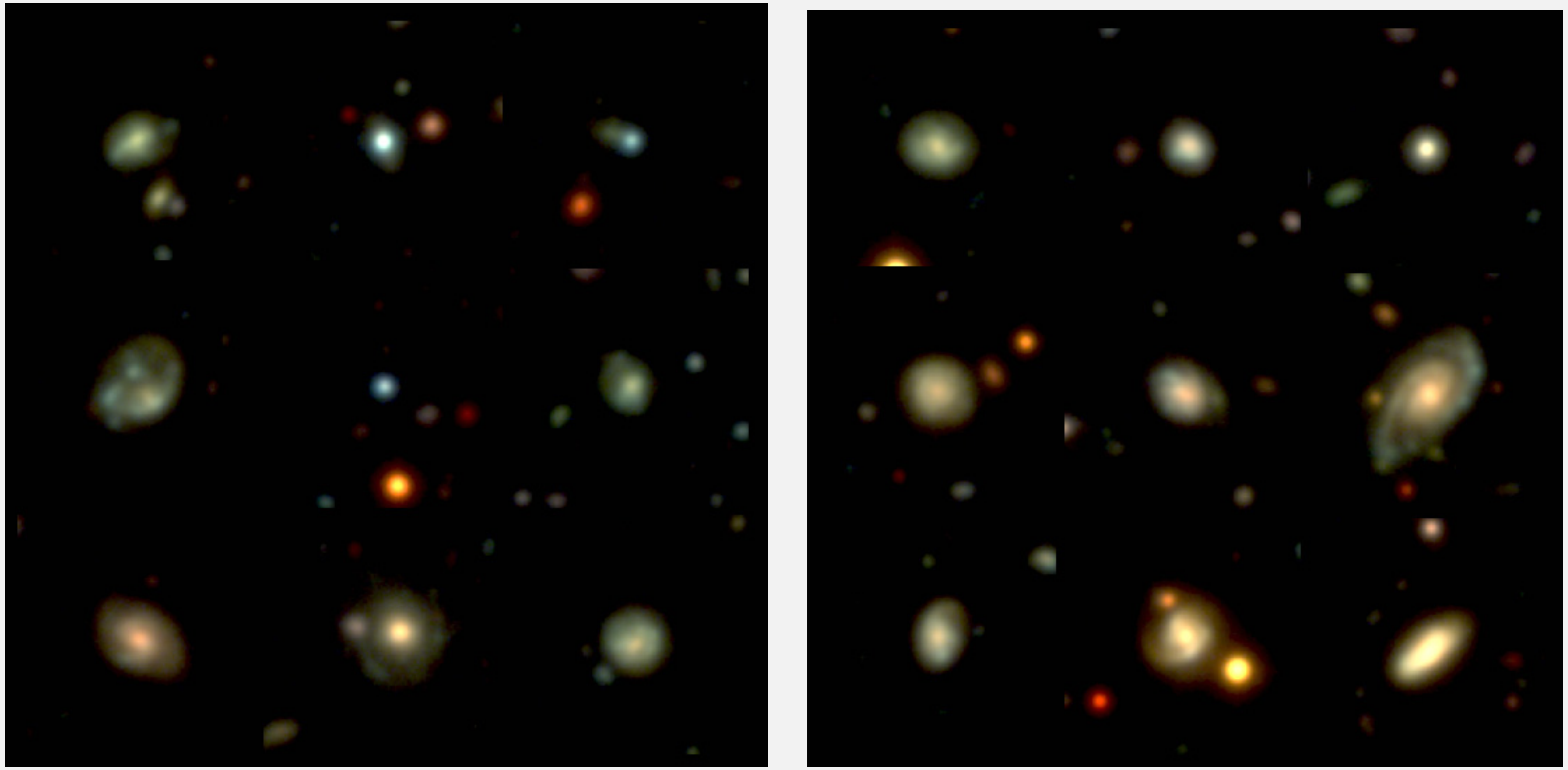


Figure 27. Morphologies of some of the *GALEX* sources in the GROTH 00 field that appear to be star formers based on their optical spectra and that lie in the redshift interval $z = 0.195$–0.44. These include sources in the optically confirmed Lyα Galaxy sample (left panel) and sources in the optically confirmed NUV-continuum selected galaxy sample (right panel). The blue, green, and red colors correspond to the u^*-, g'-, and i'-band images from the CFHT Legacy Survey deep observations of this field. In the right panel nearly all of the galaxies are spirals. We classify only the two right-most galaxies in the left row as compact. Some of the optically confirmed Lyα Galaxies in the left panel are also spirals (the bottom three galaxies), but a much larger fraction are mergers (the leftmost galaxy in the second row) or compacts (the remaining galaxies).

Cowie et al. (2010)

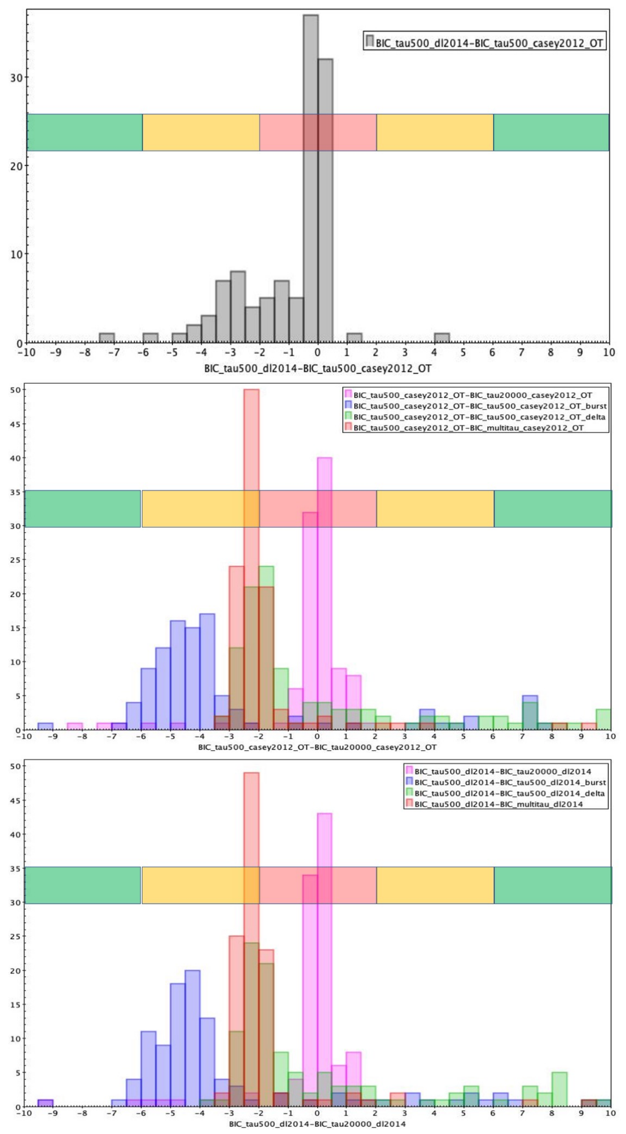


Fig. 6. Result of the ΔBIC test on our sample of Hiz-SFGs. A delayed SFH without a burst and $\tau_{\rm main}$ = 500 Myr is labeled as "tau500". With an additional burst at the end of the SFH, it is labeled as "burst". A constant SFH without a burst and $\tau_{\rm main}$ = 20 Gyrs is labeled as "tau20000". And when several $\tau_{\rm main}$ could be selected in the SED fitting, it is labeled as "multitau" in the legend. *Top*: ΔBIC test that compares the influence of the DL2014 model and the PL+OT_MBB dust emissions in building the IR template. *Center*: ΔBIC test on the SFH assuming the PL+OT_MBB for the IR template. *Bottom*: ΔBIC test on the SFH assuming the DL2014 model for the IR template. The color band allows one to interpret the results of the evidence (ΔBIC) against the model with the higher BIC: red means "faint evidence", orange means "positive evidence", and green mean "strong evidence". We do not see any strong evidence that DL2014 or PL+OT_MBB are better for fitting the data. An SFH that includes a burst is positively ruled out, while a delayed SFH with τ = 500 Myr is weekly favored.

4. Analysis of the results

From the previous analysis, we derived a set of physical parameters for each of the galaxies in our sample. These parameters allowed us to define and build diagnostic diagrams that permitted us to characterize these galaxies and, more specifically, their SFH and their dust properties. We analyzed the locations of our sample in the IRX versus $\beta_{\rm UV}$ diagram, the $A_{\rm FUV}$ versus $M_{\rm star}$ diagram, and in the specific dust mass versus specific star formation rate: $sM_{\rm dust}$ versus sSFR $= M_{\rm dust}/M_{\rm star}$ versus SFR/$M_{\rm star}$ (dust formation rate diagram or DFRD).

4.1. The SFR versus M_{star} diagram

Figure 7 presents the SFR versus $M_{\rm star}$ diagram[3]. The points corresponding to this work are found in the expected range when compared to Pearson et al. (2018) who also used CIGALE:

$$\log_{10}({\rm SFR}) = (1.00 \pm 0.22)\ (\log_{10}(M_{\rm star}) - 10.6) + (1.92 \pm 0.21)$$

and compared to Speagle et al. (2014) with their "mixed" (preferred fit)":

$$\log_{10}({\rm SFR}) = [(0.73 \pm 0.02) - (0.027 \pm 0.006) \times t[{\rm Gyr}]] \log_{10}(M_{\rm star}) - [(5.42 \pm 0.22) + (0.42 \pm 0.07) \times t[{\rm Gyr}]]$$

function evaluated at z = 5.0 (i.e., $t_{\rm universe}$ = 1.186 Gyr), which was converted to a Chabrier IMF by subtracting 0.03 to $\log_{10}$(SFR) and to $\log_{10}(M_{\rm star})$. The first result is that regardless of the dust emission used (DL2014 or PL+OT_MBB), we found about the same main sequence. Our data are in good agreement with Faisst et al. (2020a) and Khusanova et al. (2021) and also they generally follow the relations derived by Speagle et al. (2014) and Pearson et al. (2018). Most of the detections are found at a relatively large stellar mass ($\log_{10}(M_{\rm star}) \sim 10$). We confirm the evolution of the main sequence to $z = 4.5 - 5.5$ with our sample of galaxies (mainly objects not detected in dust continuum), which extend to $\log_{10}(M_{\rm star}) \sim 8.5$.

4.2. The $A_{\rm FUV}$ versus $M_{\rm star}$ diagram

The relation between the dust attenuation ($A_{\rm FUV}$ or its proxy, IRX $= L_{\rm dust}/L_{\rm FUV}$) and the stellar mass ($M_{\rm star}$) is another way to estimate the dust attenuation in galaxies without far-IR data. This relation between the stellar mass and dust attenuation has been the focus of numerous studies (as early as Xu et al. 2007; Buat et al. 2009, and references therein). However, even if this relation could be very useful in addition to the IRX versus $\beta_{\rm FUV}$ one ($f_\lambda \propto \lambda^{\beta_{\rm FUV}}$), the link between the far-UV (FUV) dust attenuation and the stellar mass ($A_{\rm FUV}$ versus $M_{\rm star}$) is not well established at all redshifts.

Bouwens et al. (2016) define what a consensus relation could be: $\log_{10}({\rm IRX}) = \log_{10}(M_{\rm star}/M_\odot) - 9.17$ assuming the dust temperature evolves with the redshift. This relation is linear in the plane $\log_{10}$(IRX) versus $\log_{10}(M_{\rm star}/M_\odot)$, and in the range $9.0 < \log_{10}(M_{\rm star}/M_\odot) < 11.0$. In a recent paper, Carvajal et al. (2020) used the stacking method for 1582 UV LBGs with photometric redshifts in the range $z \sim$ 2–8 to reach down to $\log_{10}(M_{\rm star}/M_\odot) = 6.0$. However, the constraints from this stacking are only upper limits, which are less useful than detections (see their Fig. 16). Fudamoto et al. (2020) made use of the ALPINE data and show that the $\log_{10}$(IRX) versus $M_{\rm star}$ relation derived from their observations is inconsistent with the previously determined relations at $z \leq 4$. They found a fast decrease in IRX at $z \sim 4$ in massive galaxies which suggests an evolution of the average amount of dust attenuation in star forming galaxies. Bernhard et al. (2014) assume an evolving normalization of the $\log_{10}$(IRX) versus $\log_{10}(M_{\rm star})$ relation in the

[3] We only present the results using the reference SFH derived in the present section, i.e., delayed SFR(t) $= t/\tau^2_{\rm main}\ \exp(t/\tau)$ with $\tau_{\rm main}$ = 500 Myr, but with the following two options for dust emission: Draine et al. (2014) and a power law in a mid-IR and optically thin blackbody.

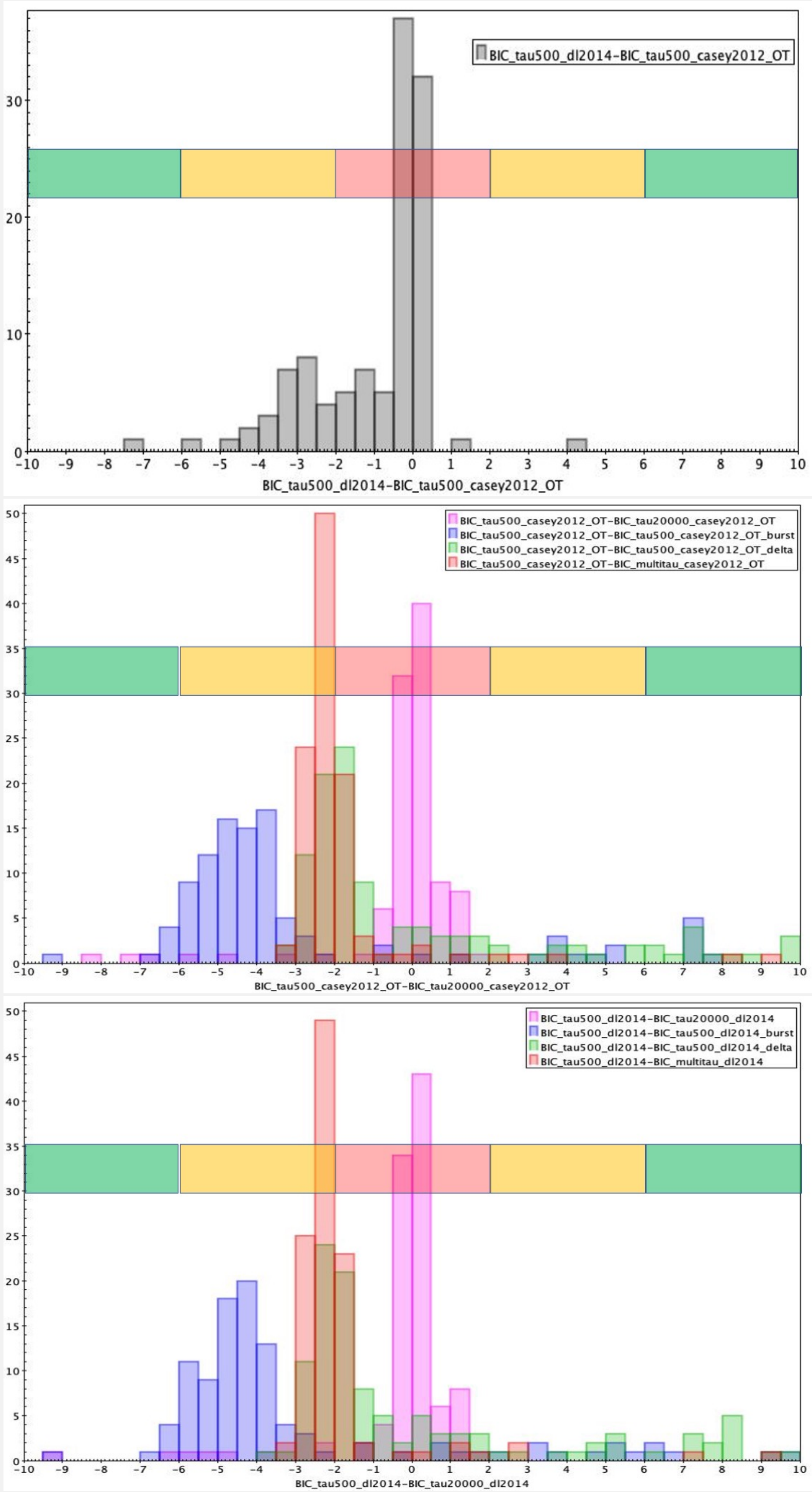


Fig. 6. Result of the ΔBIC test on our sample of Hiz-SFGs. A delayed SFH without a burst and $\tau_{\rm main}$ = 500 Myr is labeled as "tau500". With an additional burst at the end of the SFH, it is labeled as "burst". A constant SFH without a burst and $\tau_{\rm main}$ = 20 Gyrs is labeled as "tau20000". And when several $\tau_{\rm main}$ could be selected in the SED fitting, it is labeled as "multitau" in the legend. *Top*: ΔBIC test that compares the influence of the DL2014 model and the PL+OT_MBB dust emissions in building the IR template. *Center*: ΔBIC test on the SFH assuming the PL+OT_MBB for the IR template. *Bottom*: ΔBIC test on the SFH assuming the DL2014 model for the IR template. The color band allows one to interpret the results of the evidence (ΔBIC) against the model with the higher BIC: red means "faint evidence", orange means "positive evidence", and green mean "strong evidence". We do not see any strong evidence that DL2014 or PL+OT_MBB are better for fitting the data. An SFH that includes a burst is positively ruled out, while a delayed SFH with τ = 500 Myr is weekly favored.

Burgarella et al. (2022)

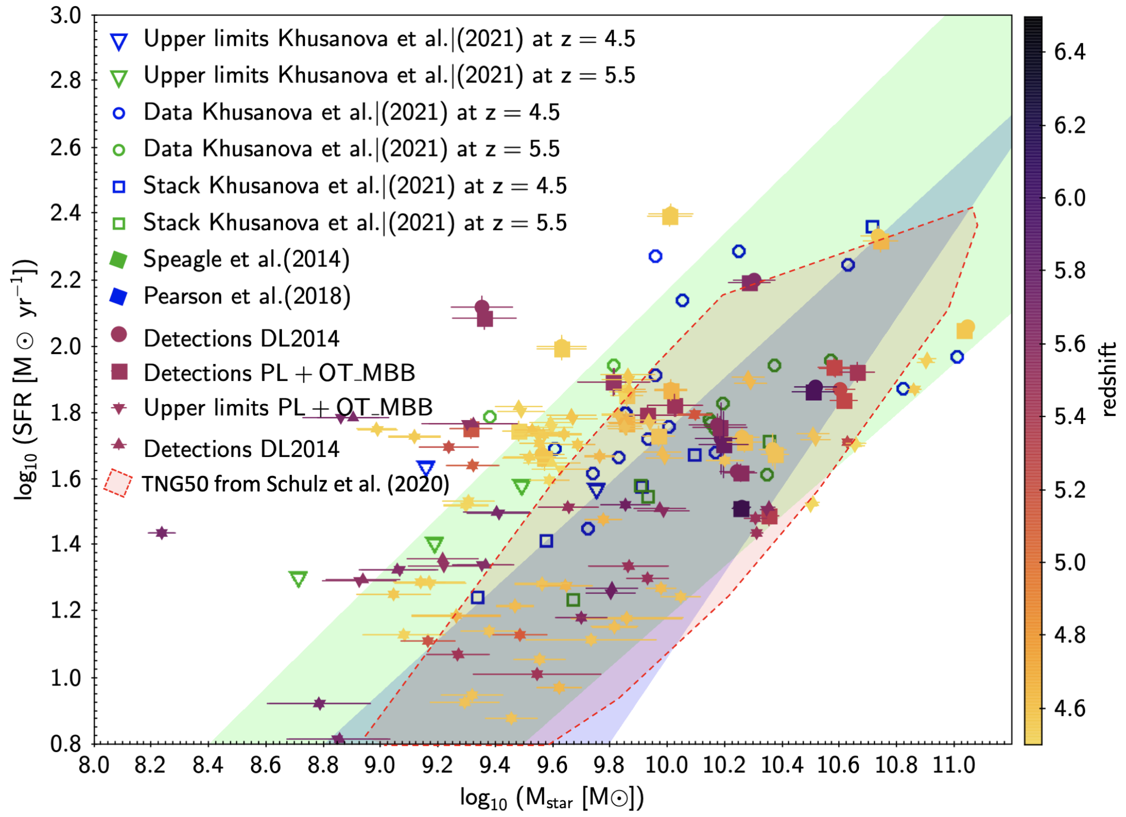


Fig. 7. In the $\log_{10}$ (SFR) vs. $\log_{10}$ ($M_{\rm star}$) diagram, the main part of the sample is found within the limits for the fits of the main sequence at $z = 5.0$ by Speagle et al. (2014; green shading), by Pearson et al. (2018) at $z \sim 5.2$ (purple shading), and by Faisst et al. (2020a), who found that the galaxies are in agreement with Speagle et al. (2014). Is is important to note, however, that some of our objects are at redshifts larger than the 4.5–5.6 ALPINE sample (see color code of the markers). For Speagle et al. (2014), we used the "mixed" (preferred fit)" function (as defined by Speagle et al. 2014). The results from the two fits with DL2014 and PL+OT_MBB are presented. The two types of dust emission do not significantly modify the location of the points in the diagram. Detections are shown as dots and boxes and upper limits are shown as downward- and upward-pointing triangles. The uncertainty range for upper limits extends to the bottom of the plot. In addition to having mainly upper limits, at $\log_{10}$ ($M_{\rm star}$) < 9.5, the sample is very likely incomplete which means that it is difficult to estimate a trend from these data over the entire mass range. We also added the objects and stacks from Khusanova et al. (2021) with open markers. Finally, the selection of TNG50 galaxies used in Schulz et al. (2020) is also provided (red-shaded area).

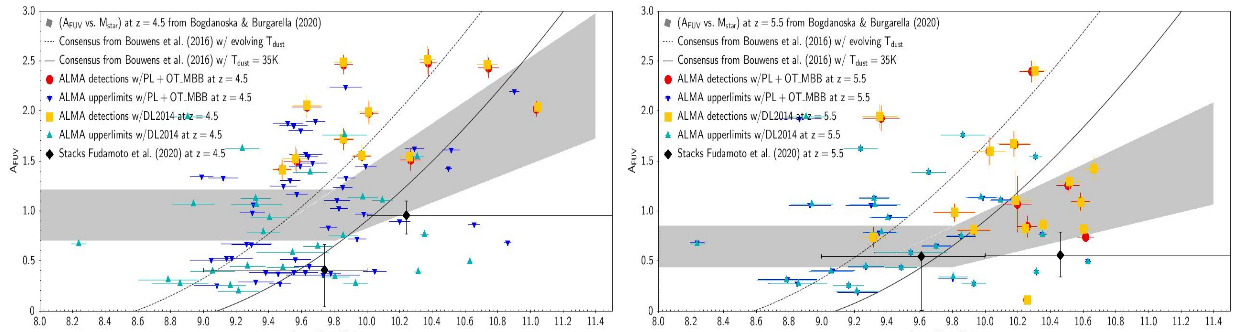


Fig. 8. $A_{\rm FUV} - M_{\rm star}$ diagram at $z \sim 4.5$ (*left*) and $z \sim 5.5$ (*right*). The gray areas correspond to the expected relation at $z = 4.5$ and 5.5 from Bogdanoska & Burgarella (2020). This relation was formed by a broken line, which is flat at $\log_{10} M_{\rm star} \leq 9.8$ and rises at $\log_{10}$ ($M_{\rm star}$) > 9.8. The ALPINE data are very dispersed at $z \sim 4.5$, while this flatness is supported by the data at $z \sim 5.5$. The conversion from IRX to $A_{\rm FUV}$ is from Burgarella et al. (2005): $A_{\rm FUV} = -0.028[\log_{10}(\rm IRX)]^3 + 0.392\,[\log_{10}(\rm IRX)]^2 + 1.094\,[\log_{10}(\rm IRX)] + 0.5$.

assigned to this galaxy. The dust density distribution is derived from the TNG50 gas density distribution with assumptions on the dust-to-metal ratio, the gas metallicity, the gas temperature, and the instantaneous SFR. Finally, SKIRT (Baes et al. 2011) is used to model the emission of the galaxies. Their fiducial model is a multicomponent dust mix, which models a composition of graphite, silicate, and polycyclic aromatic hydrocarbon (PAH) grains, with various grain size bins for each grain type which reproduce the properties of Milky Way (MW), large Magellanic cloud (LMC), and SMC type dust. In SKIRT, the dust of the molecular birth clouds mentioned is treated separately from the diffuse ISM dust.

Schulz et al. (2020) used the output of the TNG50 simulation and suggest a redshift-dependent systematic shift toward lower

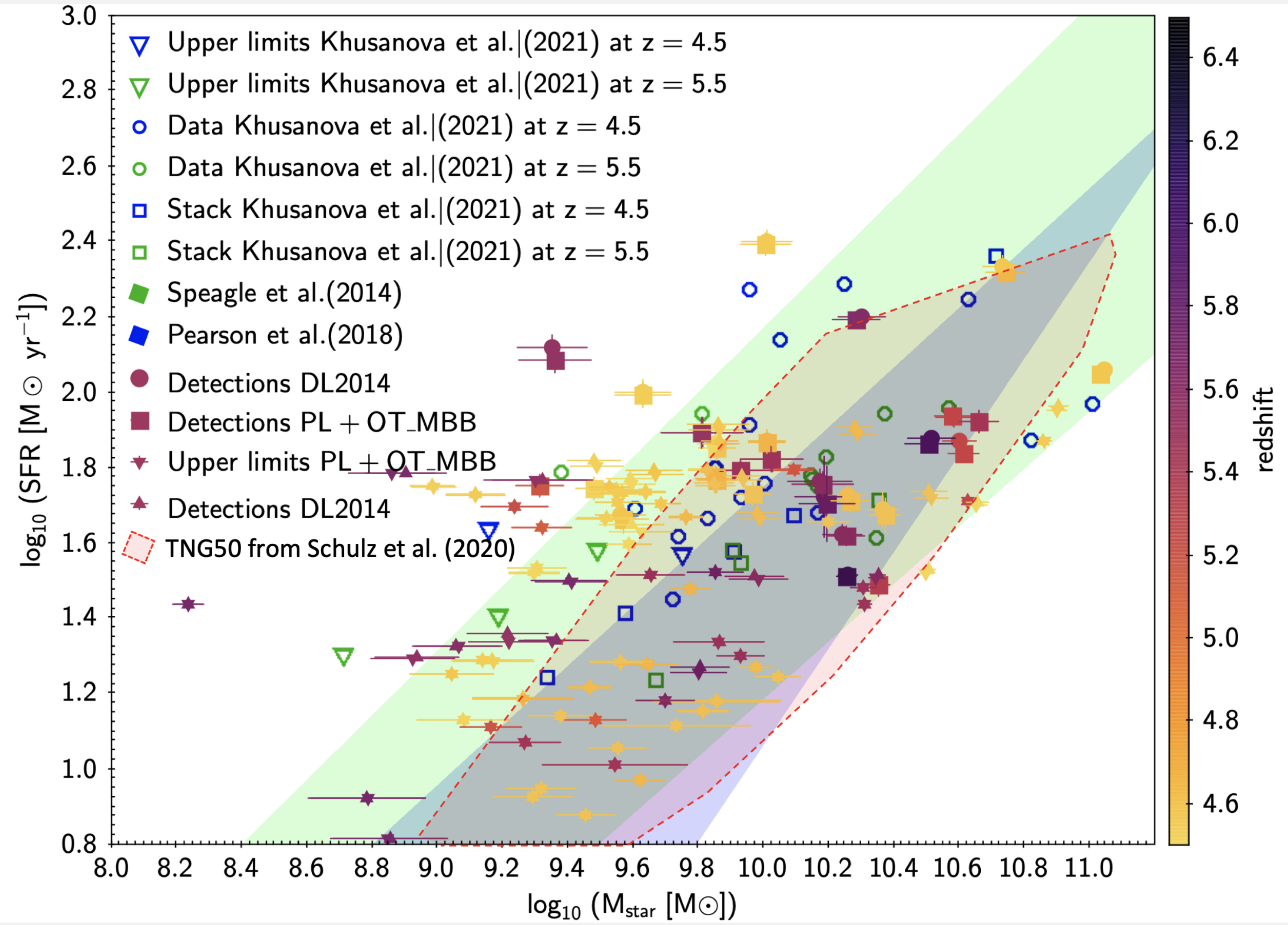


Fig. 7. In the $\log_{10}$ (SFR) vs. $\log_{10}$ ($M_{\rm star}$) diagram, the main part of the sample is found within the limits for the fits of the main sequence at $z = 5.0$ by Speagle et al. (2014; green shading), by Pearson et al. (2018) at $z \sim 5.2$ (purple shading), and by Faisst et al. (2020a), who found that the galaxies are in agreement with Speagle et al. (2014). Is is important to note, however, that some of our objects are at redshifts larger than the 4.5–5.6 ALPINE sample (see color code of the markers). For Speagle et al. (2014), we used the "mixed" (preferred fit)" function (as defined by Speagle et al. 2014). The results from the two fits with DL2014 and PL+OT_MBB are presented. The two types of dust emission do not significantly modify the location of the points in the diagram. Detections are shown as dots and boxes and upper limits are shown as downward- and upward-pointing triangles. The uncertainty range for upper limits extends to the bottom of the plot. In addition to having mainly upper limits, at $\log_{10}$ ($M_{\rm star}$) < 9.5, the sample is very likely incomplete which means that it is difficult to estimate a trend from these data over the entire mass range. We also added the objects and stacks from Khusanova et al. (2021) with open markers. Finally, the selection of TNG50 galaxies used in Schulz et al. (2020) is also provided (red-shaded area).

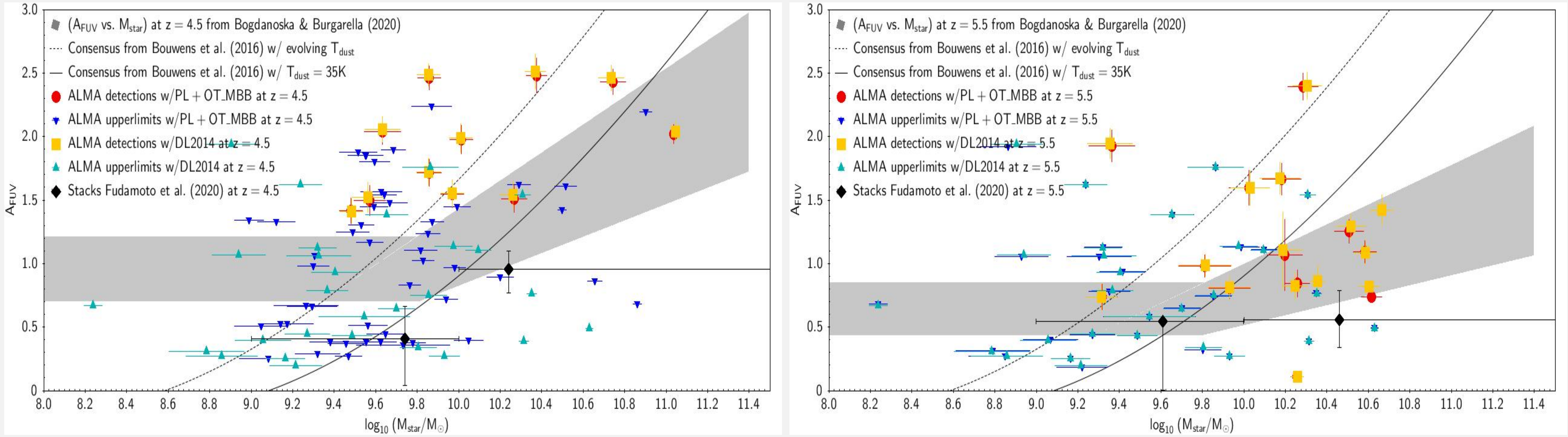


Fig. 8. $A_{\rm FUV} - M_{\rm star}$ diagram at $z \sim 4.5$ (*left*) and $z \sim 5.5$ (*right*). The gray areas correspond to the expected relation at $z = 4.5$ and 5.5 from Bogdanoska & Burgarella (2020). This relation was formed by a broken line, which is flat at $\log_{10} M_{\rm star} \leq 9.8$ and rises at $\log_{10}$ ($M_{\rm star}$) > 9.8. The ALPINE data are very dispersed at $z \sim 4.5$, while this flatness is supported by the data at $z \sim 5.5$. The conversion from IRX to $A_{\rm FUV}$ is from Burgarella et al. (2005): $A_{\rm FUV} = -0.028[\log_{10}(\rm IRX)]^3 + 0.392\,[\log_{10}(\rm IRX)]^2 + 1.094\,[\log_{10}(\rm IRX)] + 0.5$.

Burgarella et al. (2022)

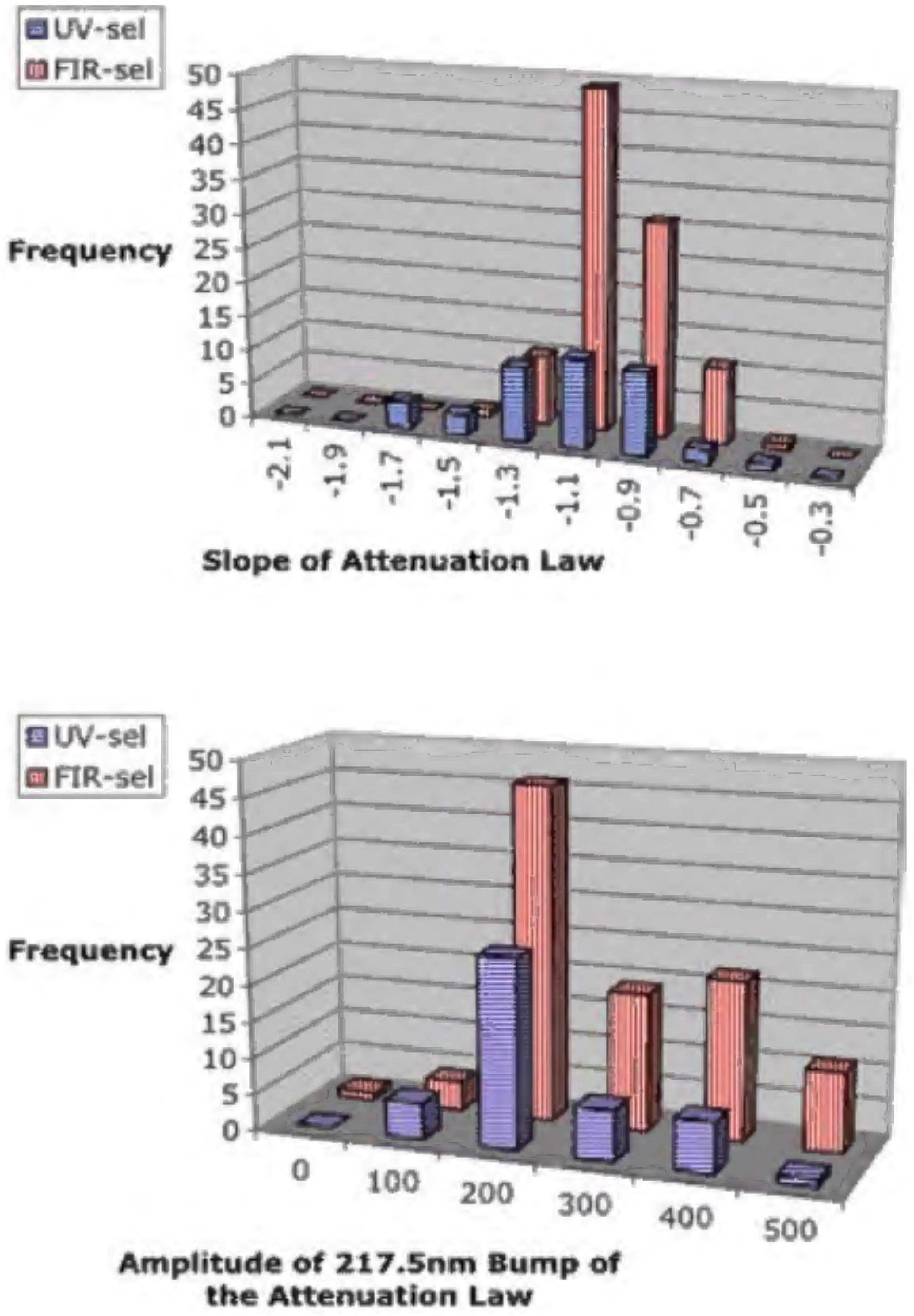


Figure 6. (a) Histogram of the slope α of the dust attenuation law for our UV-selected (blue) and FIR-selected (red) samples and (b) Histogram of the amplitude A_{bump} of the dust attenuation law for our UV-selected (blue) and FIR-selected (red) samples.

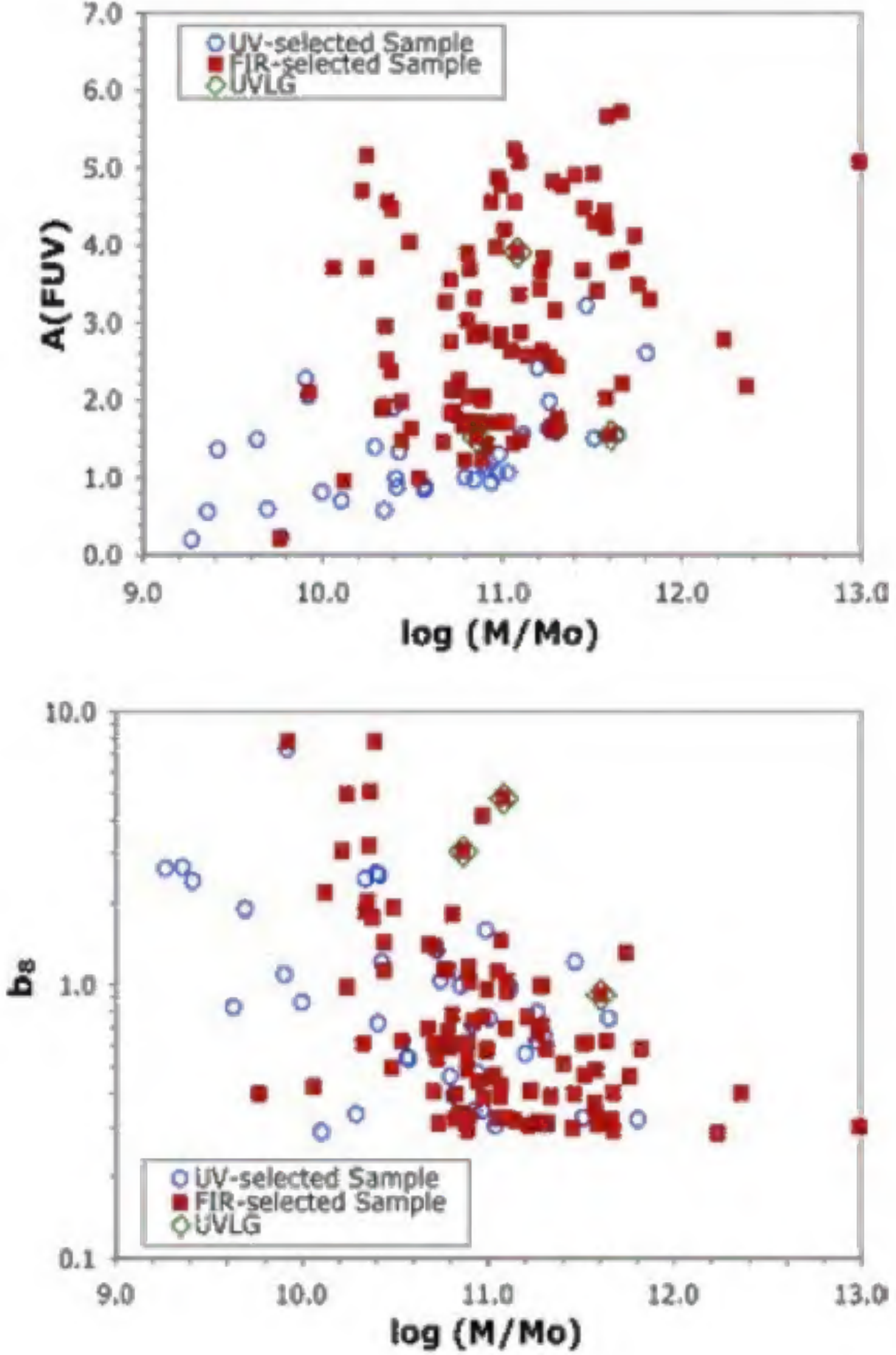


Figure 12. (a) The most massive galaxies are the most attenuated ones while the less massive ones appear to be lighter and (b) the most massive galaxies exhibit the lowest present (within 100 Myr) to past SFR: they are the oldest ones of the sample while the youngest ones are mainly UV selected and less massive. The three UVLGs detected in our sample are also plotted in these diagrams as diamonds.

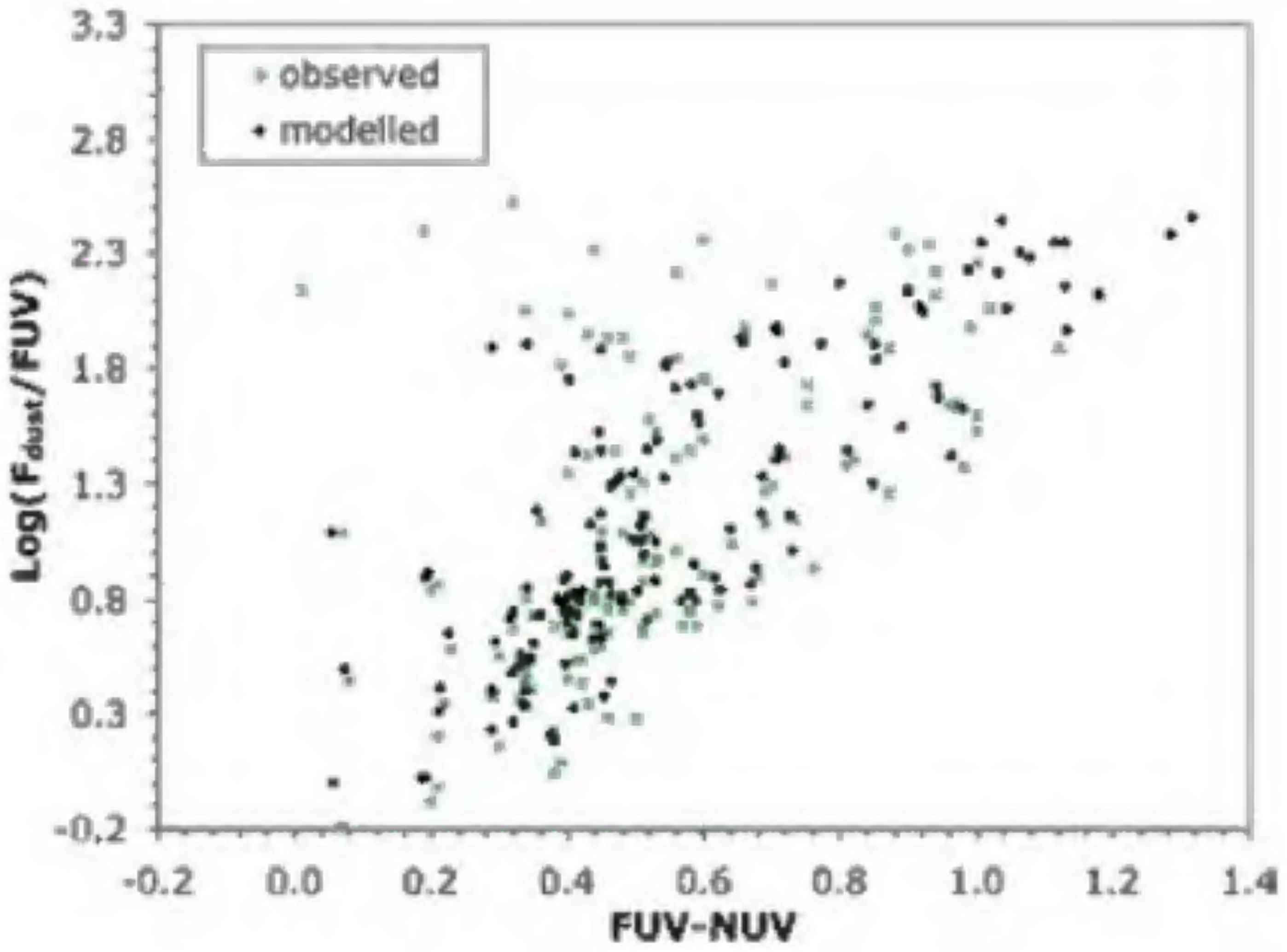


Figure 14. The comparison of the observed (crosses) and modelled (dots) diagram is rather good for all parts of the diagram, suggesting that our physical parameters are enough for the process to find a good solution both in UV and in FIR. However, the program fails to find solution for galaxies which are, at the same time rather blue (i.e. FUV – NUV < 0.5 and with extreme attenuations ($A_{\mathrm{FUV}} > 2.0$).

Burgarella et al. (2005)

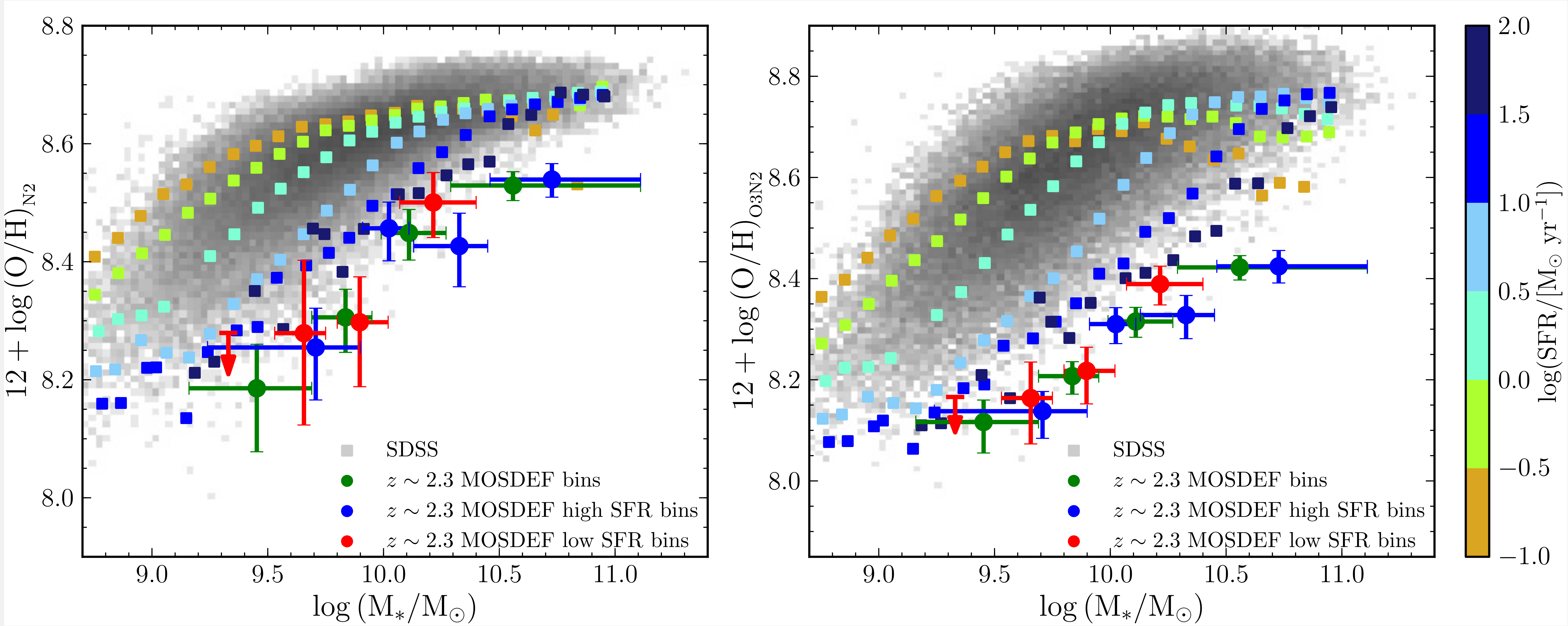


Figure 5. Comparison of M_*, metallicity, and SFR between $z \sim 2.3$ and local galaxies. The $z \sim 2.3$ high- and low-SFR subsamples are separated into four stellar mass bins, with metallicities determined using the N2 (left) and O3N2 (right) indicators. The blue and red points and error bars indicate bins of the high- and low-SFR subsamples, respectively. The green points with error bars indicate stellar mass bins from the full sample. Error bars for all binned points are the same as in Figure 2. The gray two-dimensional histogram shows the density of local SDSS galaxies in this parameter space. Colored squares are M_*-SFR bins of local SDSS star-forming galaxies from Andrews & Martini (2013), with the color indicating the range of SFRs in a bin (see colorbar). Red MOSDEF $z \sim 2.3$ bins are comparable to SDSS bins with $\log(\mathrm{SFR}) = 1.0$–1.5 (medium blue), while blue MOSDEF bins are comparable to those with $\log(\mathrm{SFR}) = 1.5$–2.0 (dark blue).

Sanders et al. (2015)

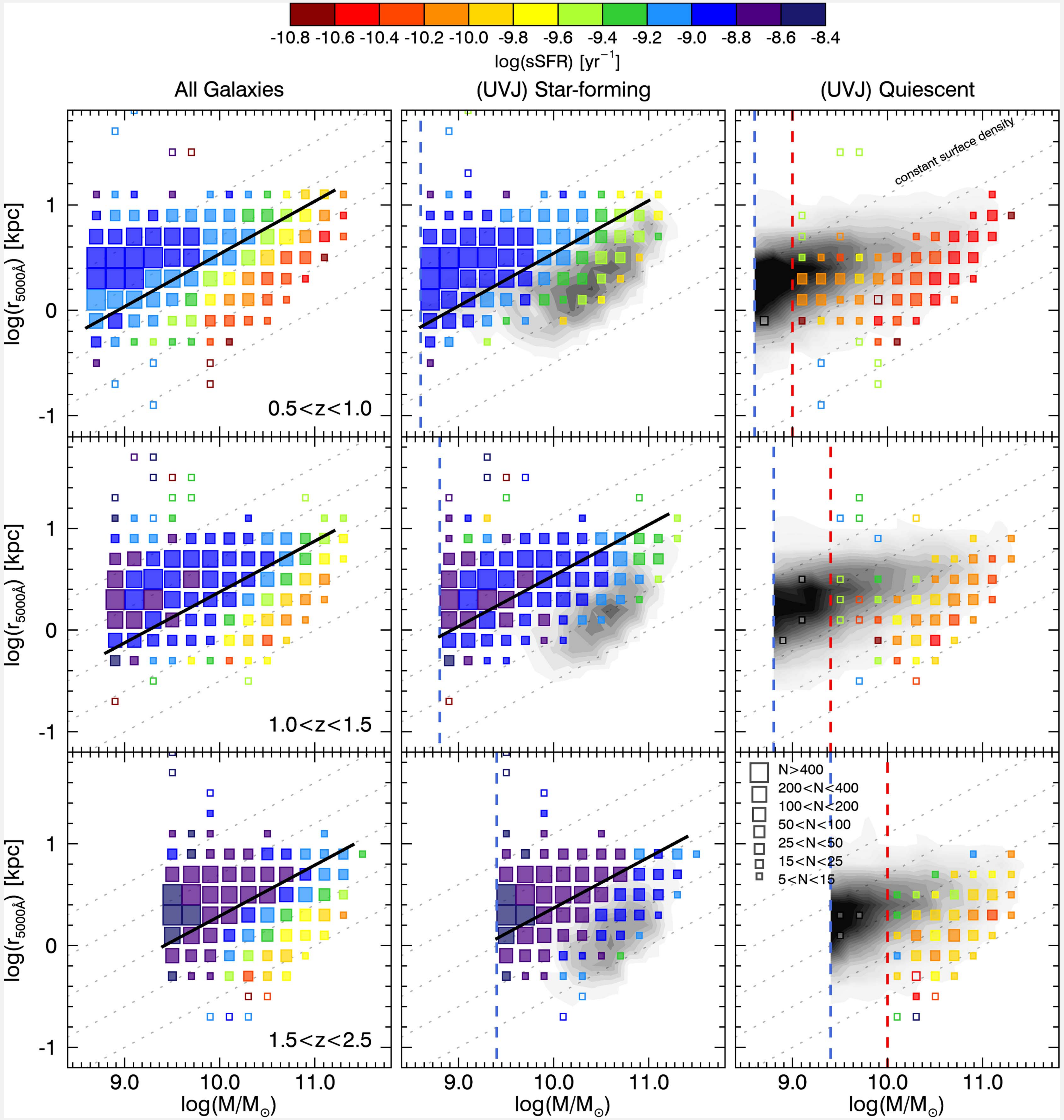


Figure 1. Rest-frame 5000 Å size of galaxies as a function of stellar mass, color-coded by the sSFRs derived from UV+IR median stacking analyses in 0.2 dex bins of $\log M_\star/M_\odot$ and $\log r_{5000\,Å}$. The size of the symbol depends on the number of galaxies that enter each bin. The vertical dashed lines correspond to the stellar mass limits down to which the structural parameters can be trusted for star-forming (blue) and quiescent (red) populations. The black dotted lines correspond to lines of constant surface density, stellar mass per unit area, with the solid line corresponding to the characteristic central density measured in Figure 9.

Whitaker et al. (2017)

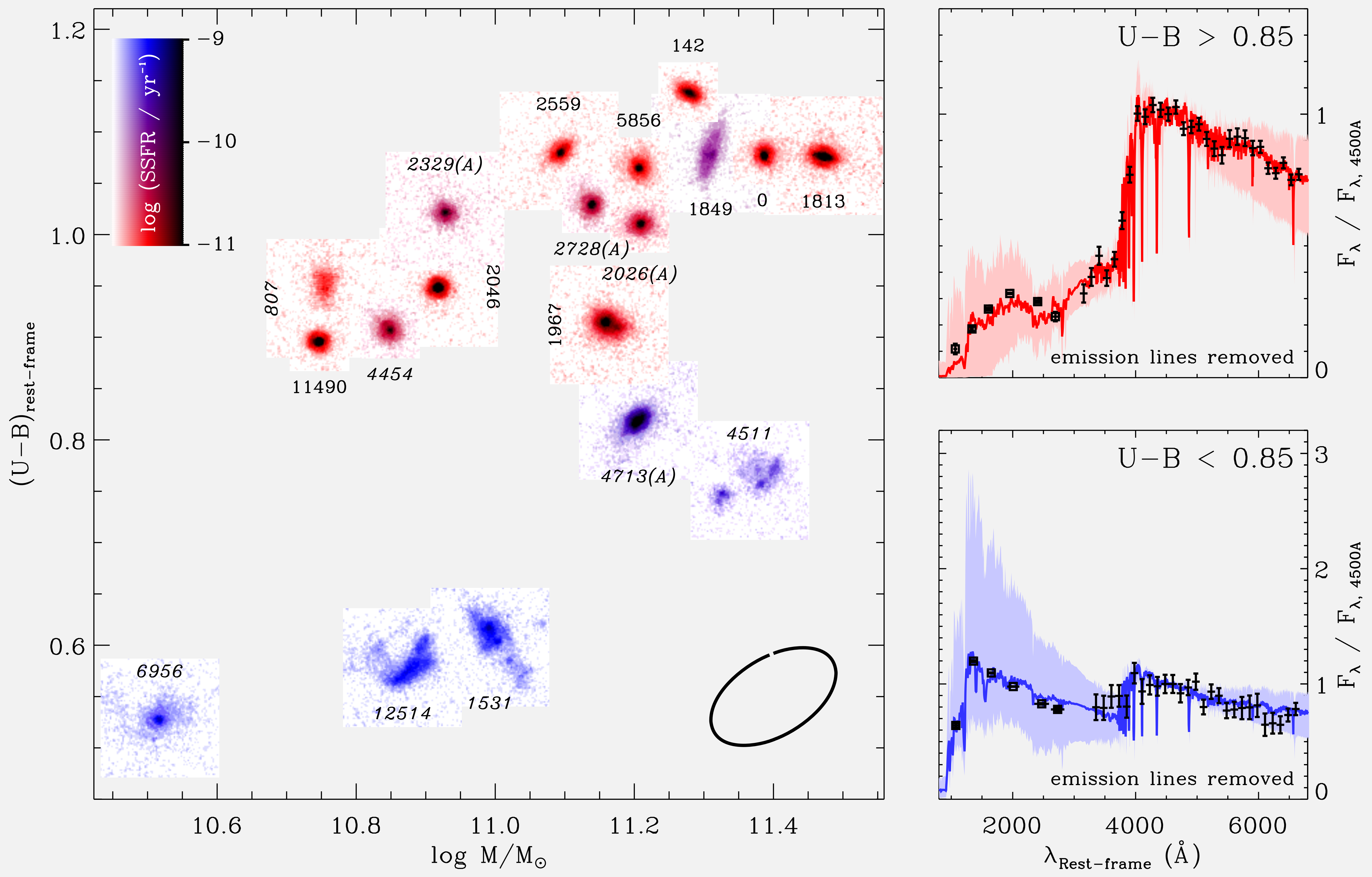


Figure 3. Left: rest-frame $U - B$ color vs. stellar mass for a massive galaxy sample at $z \sim 2.3$ with rest-frame optical spectroscopy. We use the *HST* NIC2 images as symbols. The color coding reflects the specific SFR of the galaxy. The emission-line galaxies can be recognized by their *italic* ID numbers, and *A* indicates the AGNs. The galaxies clearly separate into two classes: the large (irregular) star-forming galaxies in the blue cloud, and the compact, quiescent galaxies on the red sequence. We do caution that this sample is small and not complete. The ellipse represents the average 1σ confidence interval. Right: stacked SEDs, composed of the rest-frame UV photometry and rest-frame optical spectra of all blue (bottom panel) and red galaxies (top panel) at $2 < z < 3$ in our spectroscopic sample. We also show the stack and full range of best-fit stellar population synthesis (SPS) models. The SPS models do not have emission lines and thus they are correspondingly removed from the stacks.

Kriek et al. (2009)

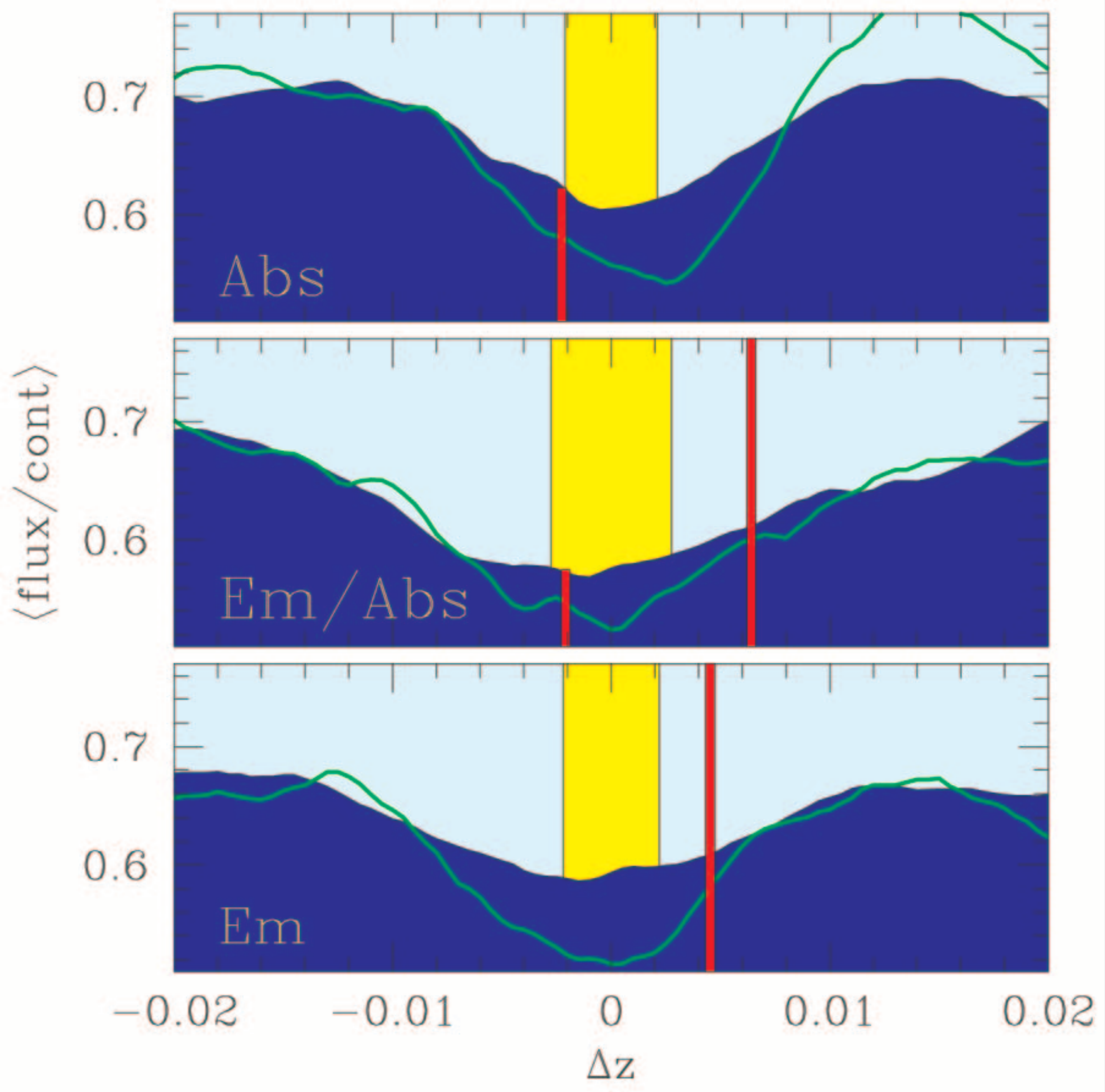


FIG. 4.—Mean transmissivity of all parts of the QSO spectra that lie within 4 h^{-1} (*solid line*) or 6 h^{-1} (*shaded region*) comoving Mpc of an LBG. The height of these curves at $\Delta z = 0$ shows our best estimate from the redshift assignment described in § 2.2.2; the height of the curves at other values of Δz shows the mean transmissivity we would have measured had we applied an additional redshift adjustment of Δz to each galaxy. Large systematic errors in our redshift estimates would show up as asymmetries about $\Delta z = 0$ in this plot, but in fact $\Delta z = 0$ lies near the minimum of each curve. The sample was divided into classes defined by the type of features we could detect in each object's spectrum. The mean redshift of the detected features relative to our assigned redshift is marked with vertical bars, short for absorption lines and tall for emission lines. The size of our estimated uncertainty ($\pm 1\ \sigma$) in a single galaxy's redshift is indicated by the rectangular strip surrounding $\Delta z = 0$.

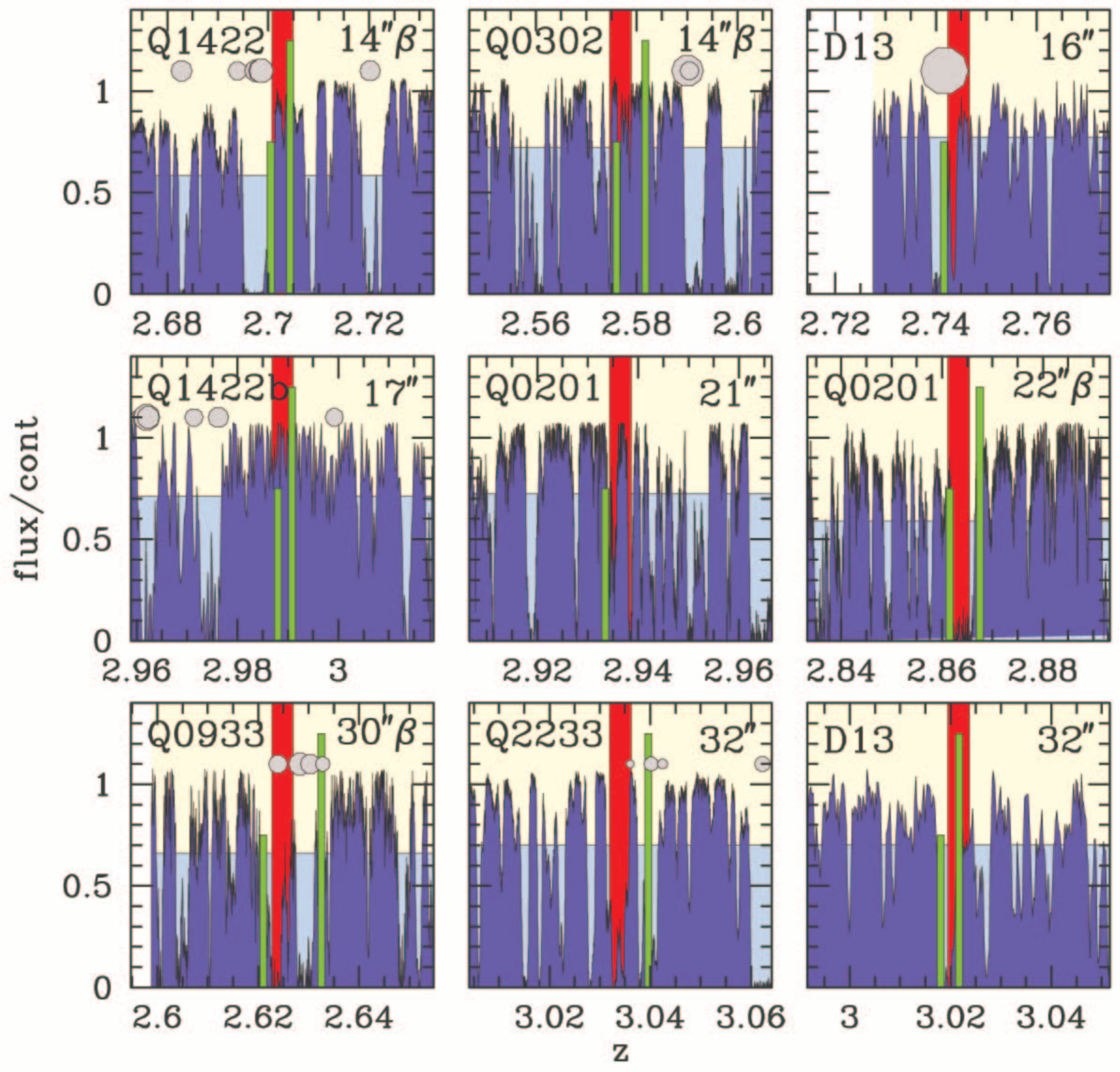


FIG. 10.—Distribution of neutral hydrogen and C IV absorption along the segments of the QSO spectra that pass closest to an LBG. The shaded curve shows the Lyα forest. The horizontal line marks the mean transmissivity at this redshift in the QSO's spectrum. Circles mark the redshifts of detectable C IV absorption. Larger circles correspond to larger C IV column densities; a tripling in the circle's area corresponds to a factor of 10 increase in column density. Numerical values for the H I and C IV column densities of the absorbing gas near to one of these galaxies can be found in the last entry of Table 2. The wide vertical region shows our estimated redshift ($\pm 1\ \sigma$) for each galaxy. Narrower vertical bars mark the redshifts of Lyα (*tall bar*) or interstellar absorption (*short bar*) in each galaxy's spectrum. The distance from each galaxy to the QSO sight line is indicated; 10″ corresponds to roughly 200 h^{-1} comoving kpc. A "β" next to the distance indicates that the Lyα forest at this redshift may be contaminated by Lyβ (or higher) absorption from gas at larger redshifts.

Adelberger et al. (2003)

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

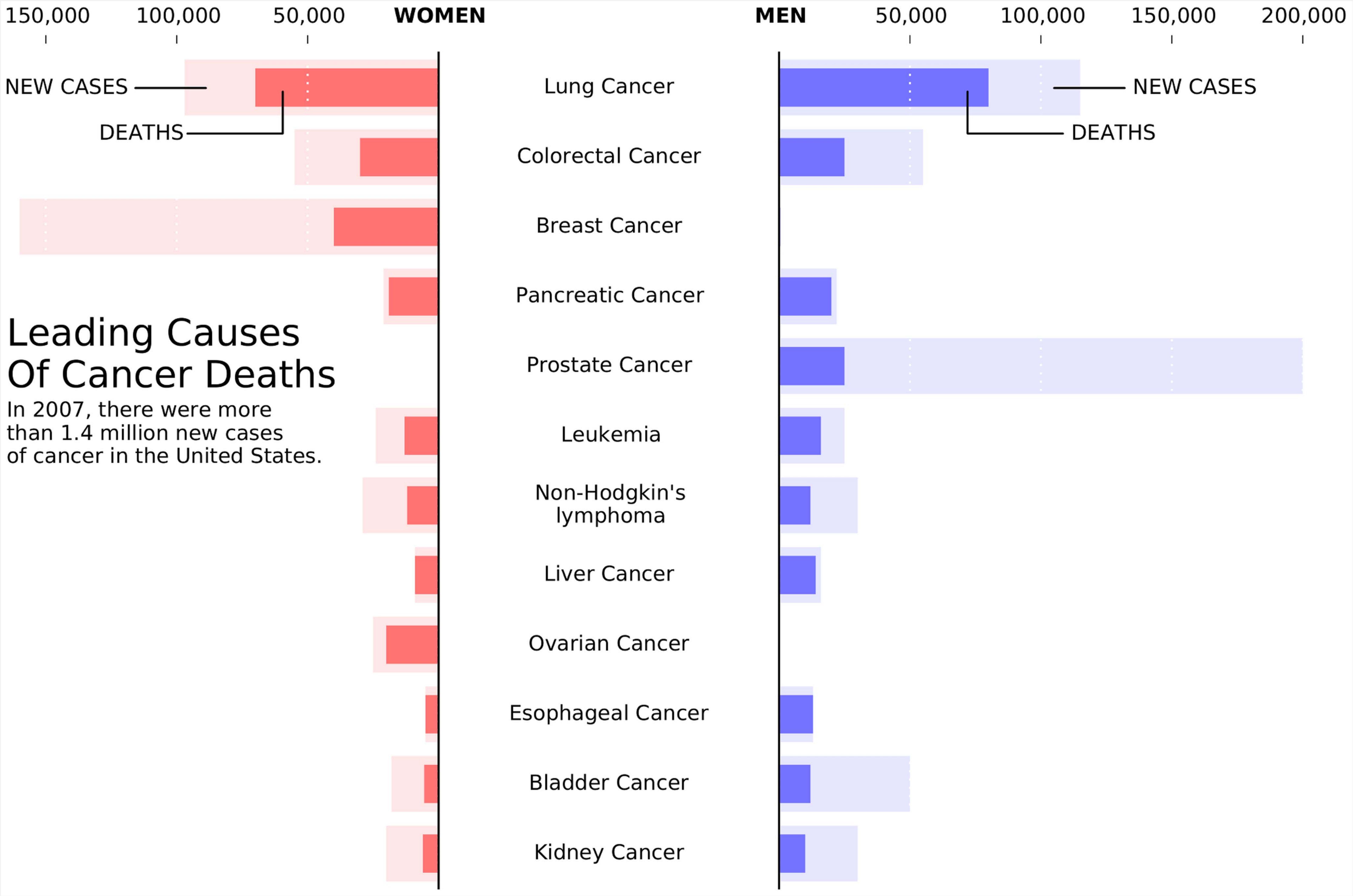


Figure 1. Know your audience. This is a remake of a figure that was originally published in the New York Times (NYT) in 2007. This new figure was made with matplotlib using approximated data. The data is made of four series (men deaths/cases, women deaths/cases) that could have been displayed using classical double column (deaths/cases) bar plots. However, the layout used here is better for the intended audience. It exploits the fact that the number of new cases is always greater than the corresponding number of deaths to mix the two values. It also takes advantage of the reading direction (English [left-to-right] for NYT) in order to ease comparison between men and women while the central labels give an immediate access to the main message of the figure (cancer). This is a self-contained figure that delivers a clear message on cancer deaths. However, it is not precise. The chosen layout makes it actually difficult to estimate the number of kidney cancer deaths because of its bottom position and the location of the labelled ticks at the top. While this is acceptable for a general-audience publication, it would not be acceptable in a scientific publication if actual numerical values were not given elsewhere in the article.

doi:10.1371/journal.pcbi.1003833.g001

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

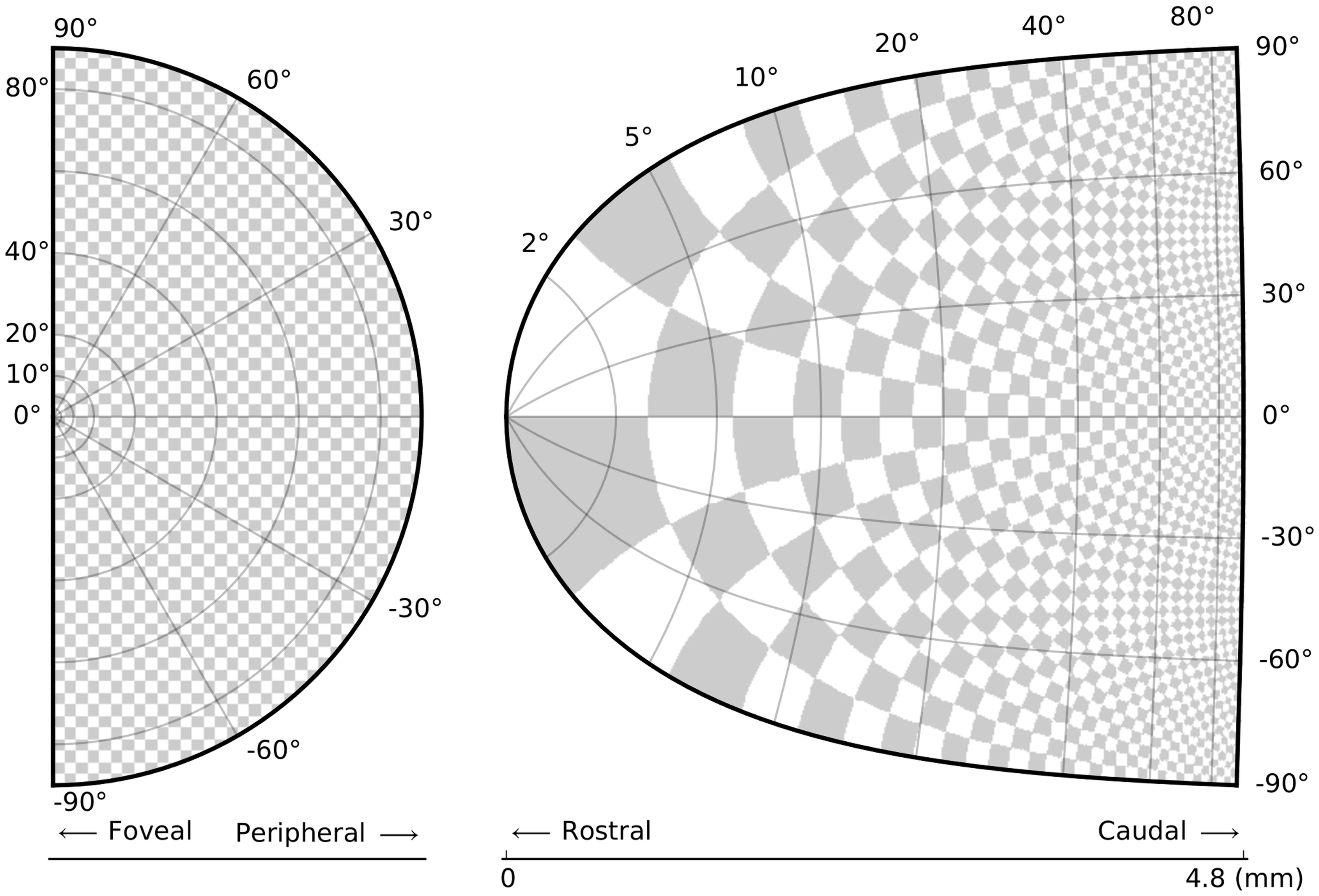


Figure 2. Identify your message. The superior colliculus (SC) is a brainstem structure at the crossroads of multiple functional pathways. Several neurophysiological studies suggest that the population of active neurons in the SC encodes the location of a visual target that induces saccadic eye movement. The projection from the retina surface (on the left) to the collicular surface (on the right) is based on a standard and quantitative model in which a logarithmic mapping function ensures the projection from retinal coordinates to collicular coordinates. This logarithmic mapping plays a major role in saccade decision. To better illustrate this role, an artificial checkerboard pattern has been used, even though such a pattern is not used during experiments. This checkerboard pattern clearly demonstrates the extreme magnification of the foveal region, which is the main message of the figure.
doi:10.1371/journal.pcbi.1003833.g002

Ten simple rules for better figures

1. **Know your audience**
2. **Identify your message**
3. **Adapt the figure to the support medium**
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

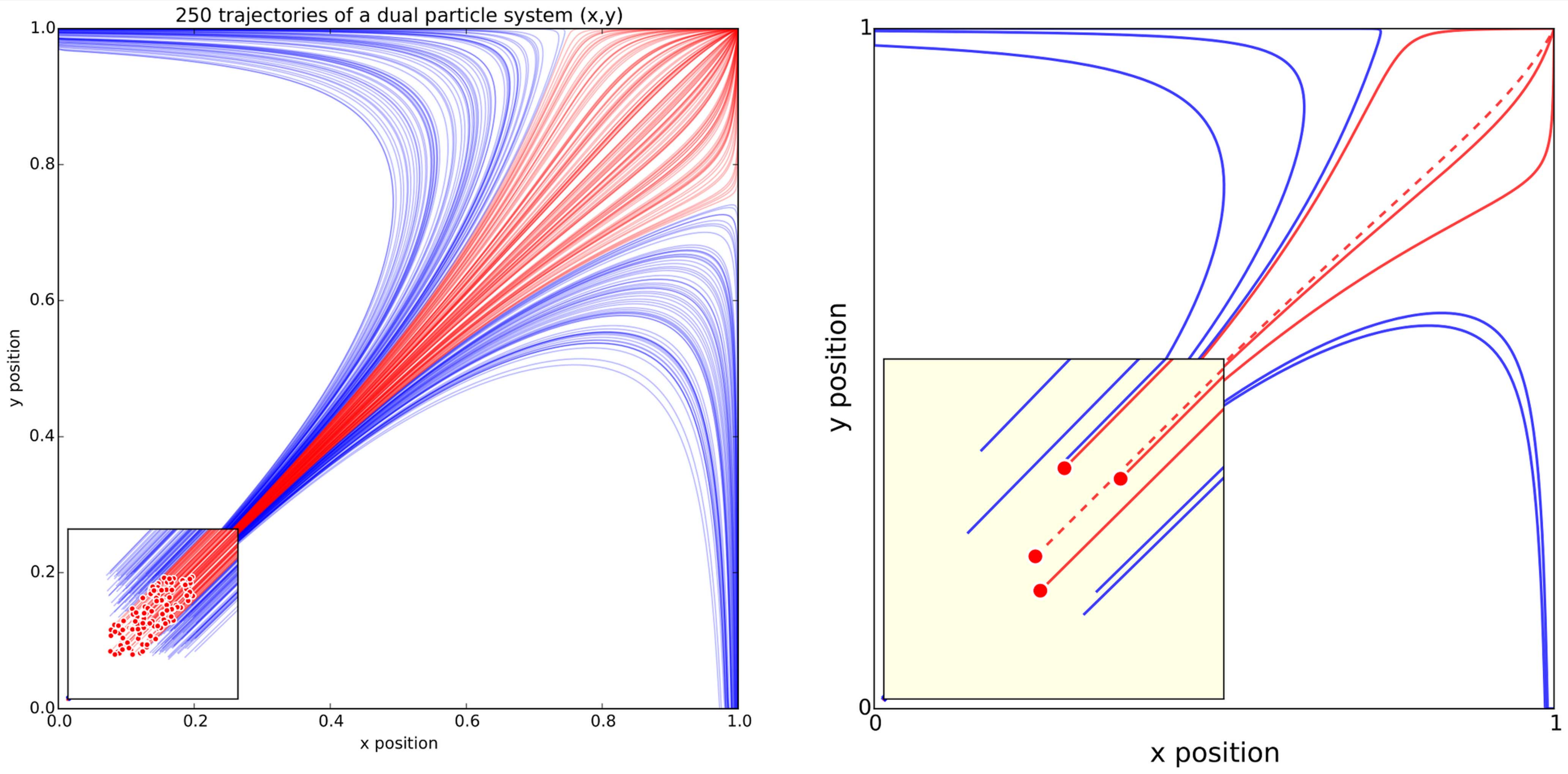


Figure 3. Adapt the figure to the support medium. These two figures represent the same simulation of the trajectories of a dual-particle system ($\frac{dx}{dt} = (^1\!/_4 + (x-y))(1-x),\ x \geq 0,\ \frac{dy}{dt} = (^1\!/_4 + (y-x))(1-y),\ y \geq 0$) where each particle interacts with the other. Depending on the initial conditions, the system may end up in three different states. The left figure has been prepared for a journal article where the reader is free to look at every detail. The red color has been used consistently to indicate both initial conditions (red dots in the zoomed panel) and trajectories (red lines). Line transparency has been increased in order to highlight regions where trajectories overlap (high color density). The right figure has been prepared for an oral presentation. Many details have been removed (reduced number of trajectories, no overlapping trajectories, reduced number of ticks, bigger axis and tick labels, no title, thicker lines) because the time-limited display of this figure would not allow for the audience to scrutinize every detail. Furthermore, since the figure will be described during the oral presentation, some parts have been modified to make them easier to reference (e.g., the yellow box, the red dashed line).
doi:10.1371/journal.pcbi.1003833.g003

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

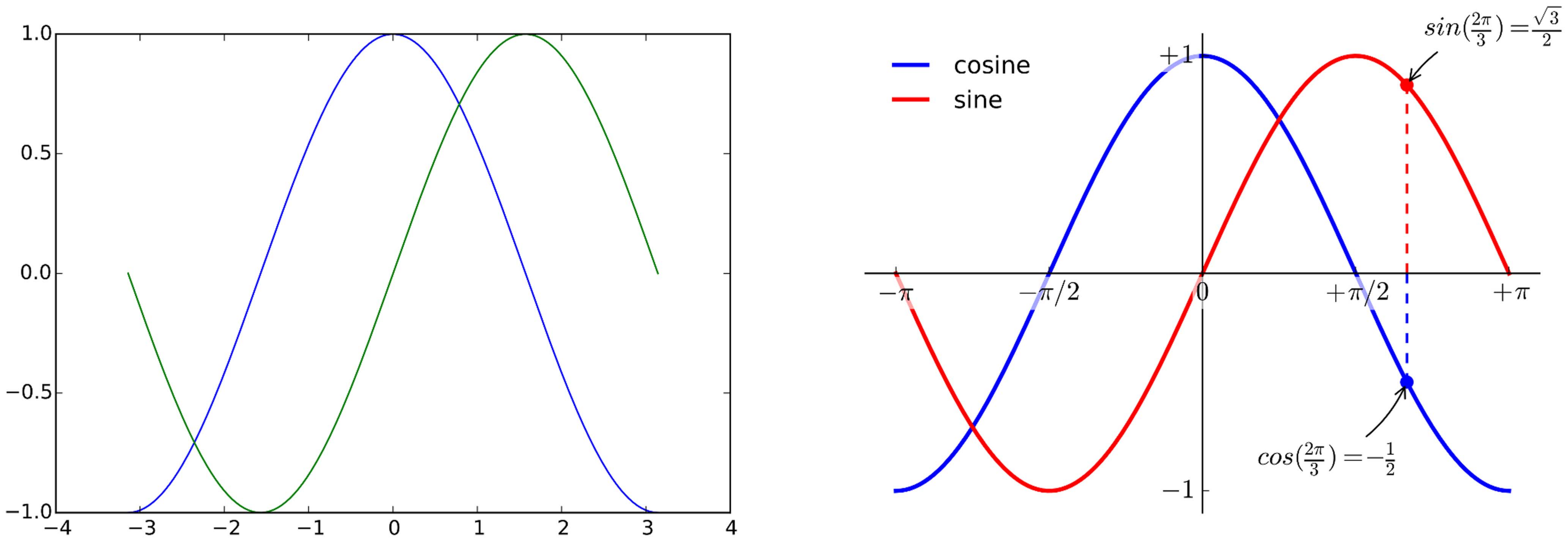


Figure 4. Do not trust the defaults. The left panel shows the sine and cosine functions as rendered by matplotlib using default settings. While this figure is clear enough, it can be visually improved by tweaking the various available settings, as shown on the right panel.
doi:10.1371/journal.pcbi.1003833.g004

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

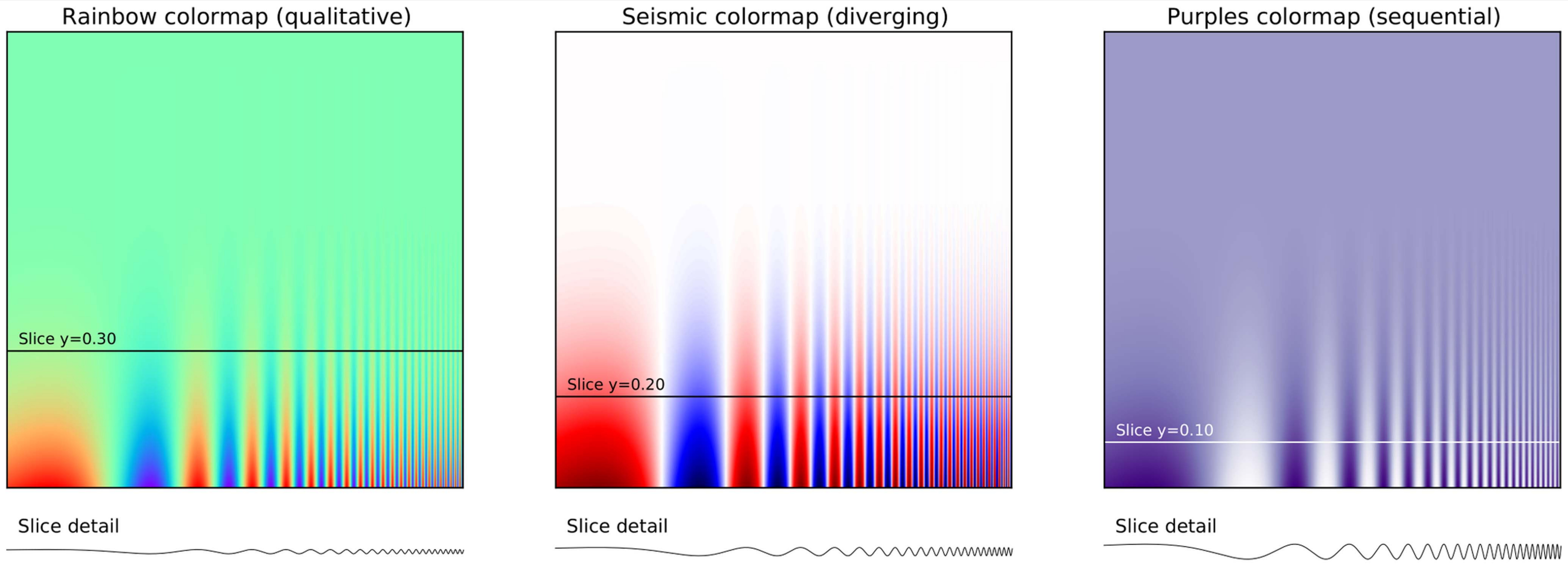
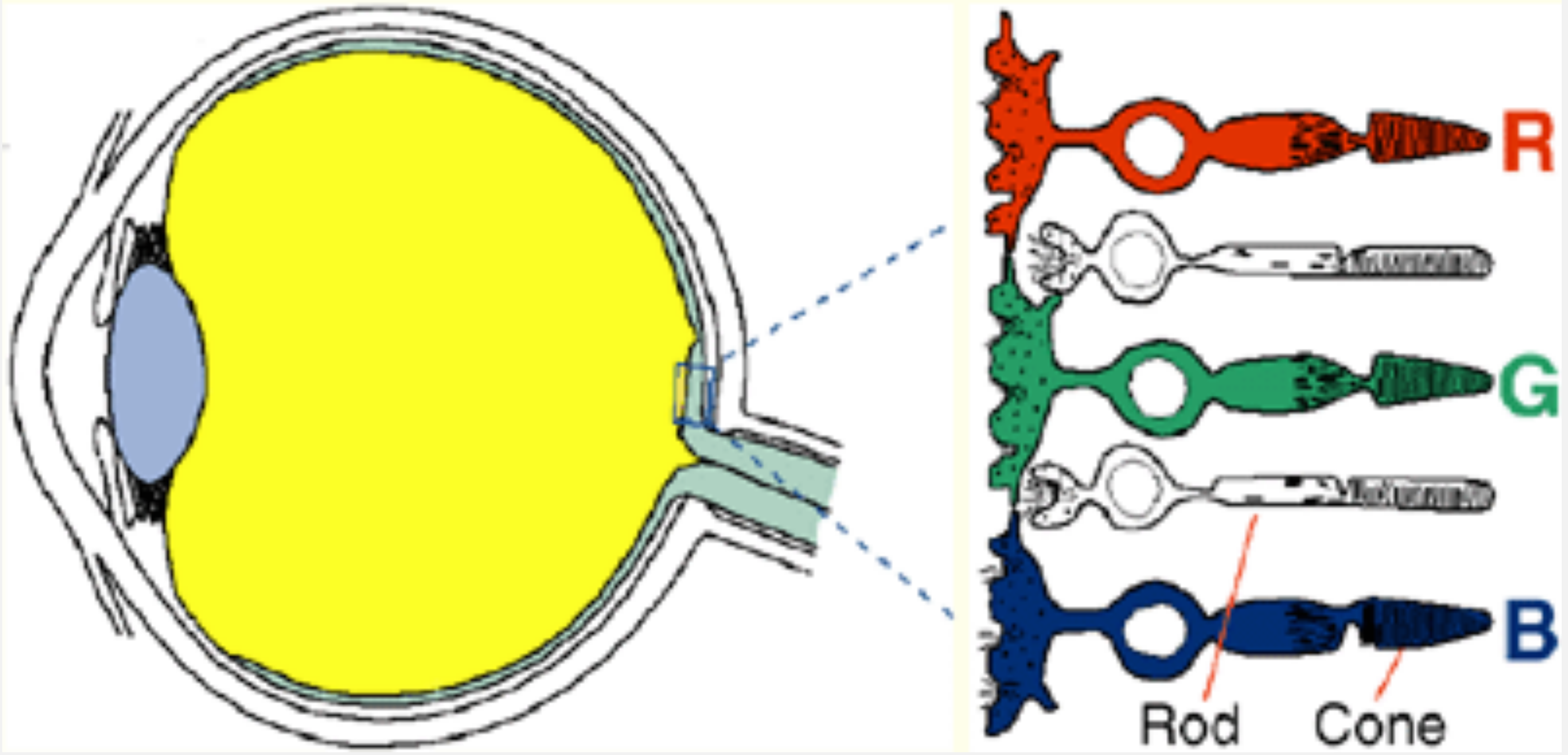
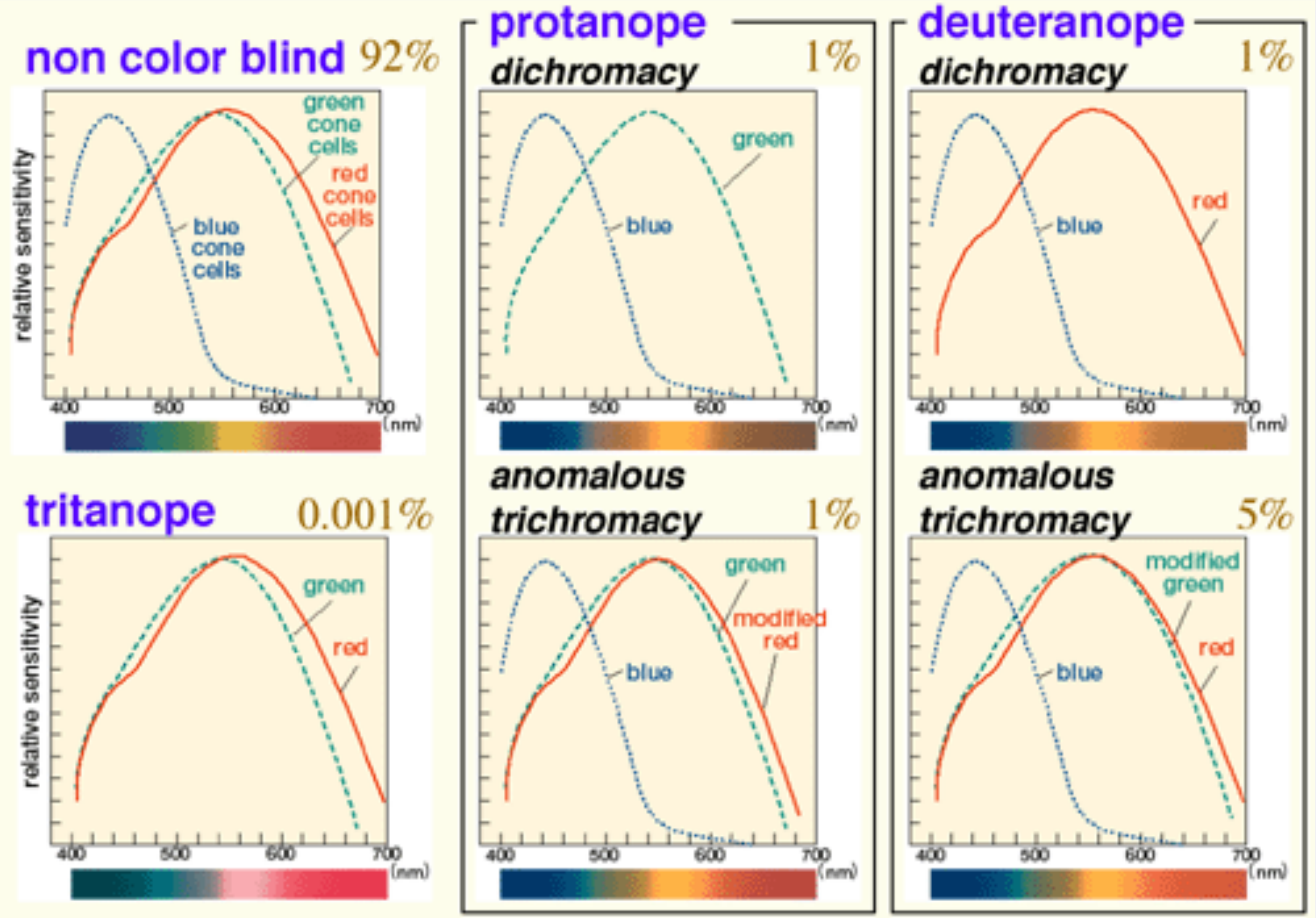


Figure 5. Use color effectively. This figure represents the same signal, whose frequency increases to the right and intensity increases towards the bottom, using three different colormaps. The rainbow colormap (qualitative) and the seismic colormap (diverging) are equally bad for such a signal because they tend to hide details in the high frequency domain (bottom-right part). Using a sequential colormap such as the purple one, it is easier to see details in the high frequency domain. Adapted from [5].
doi:10.1371/journal.pcbi.1003833.g005

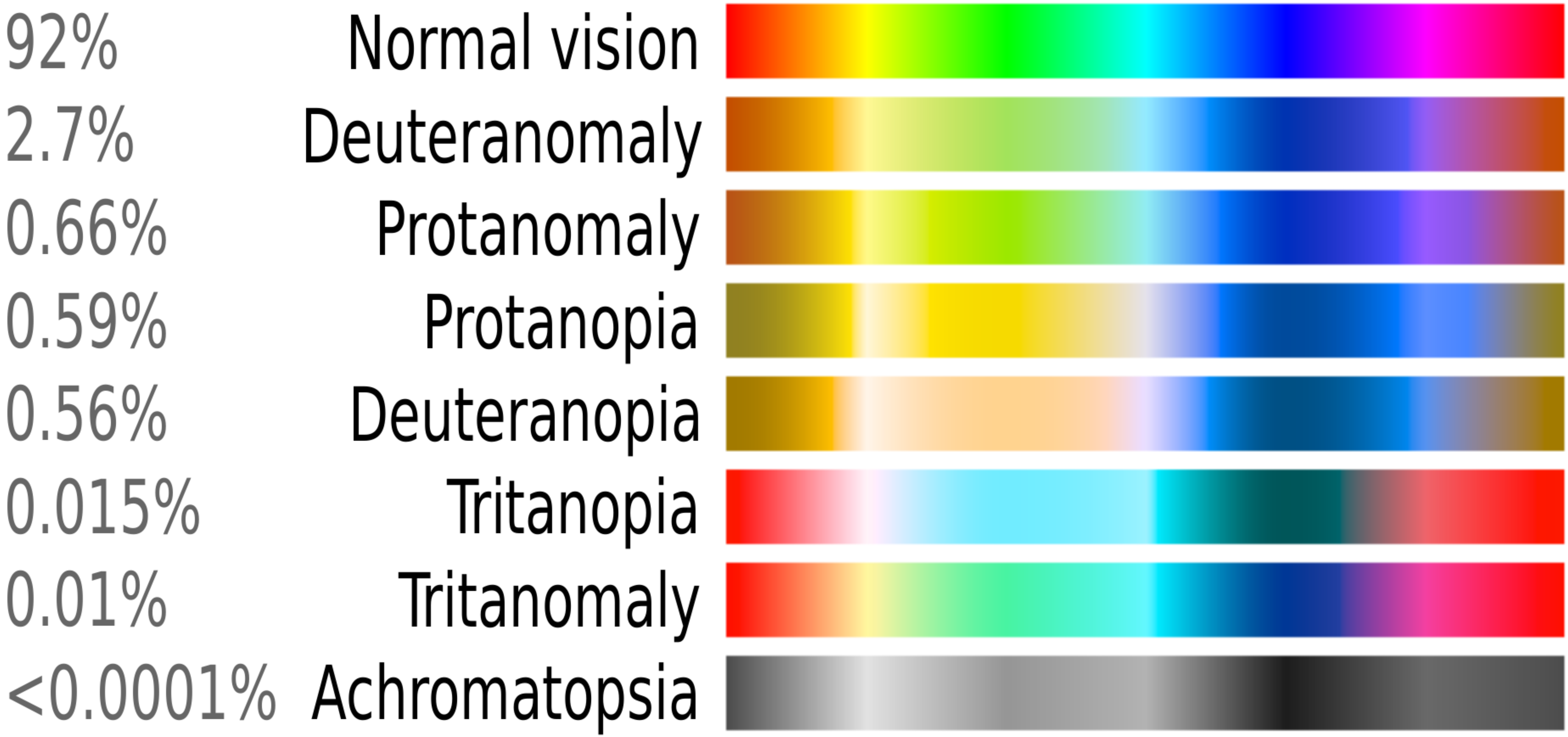
Colour blindness (also called colour vision deficiency)





- about 4.5% of the population
- 1 in 12 men; 1 in 200 women
- most common: red/green colour blindness

Colour blindness (also called colour vision deficiency)



Normal Vision

Protanopia

Deuteranopia

Tritanopia

colourblindnessawareness.org

Pronatopes are more likely to confuse:

- black with many shades of red
- dark brown with dark green, dark orange dark red, dark blue/purple and black
- some blues with some reds, purples and dark pinks
- mid-greens with some oranges

Deuteranopes are more likely to confuse:

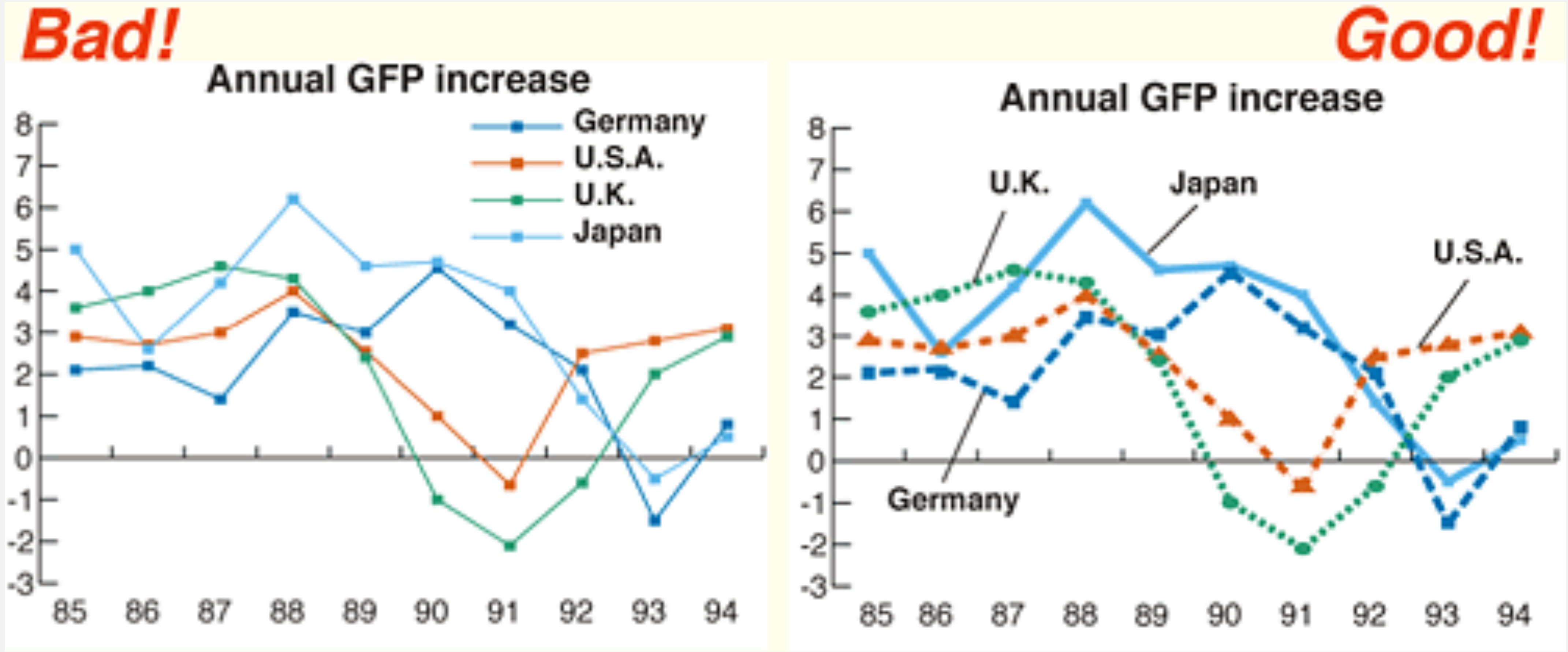
- mid-reds with mid-greens
- blue-greens with grey and mid-pinks
- bright greens with yellows
- pale pinks with light grey/white
- mid-reds with mid-brown
- light blues with lilac

Trinatopes are more likely to confuse:

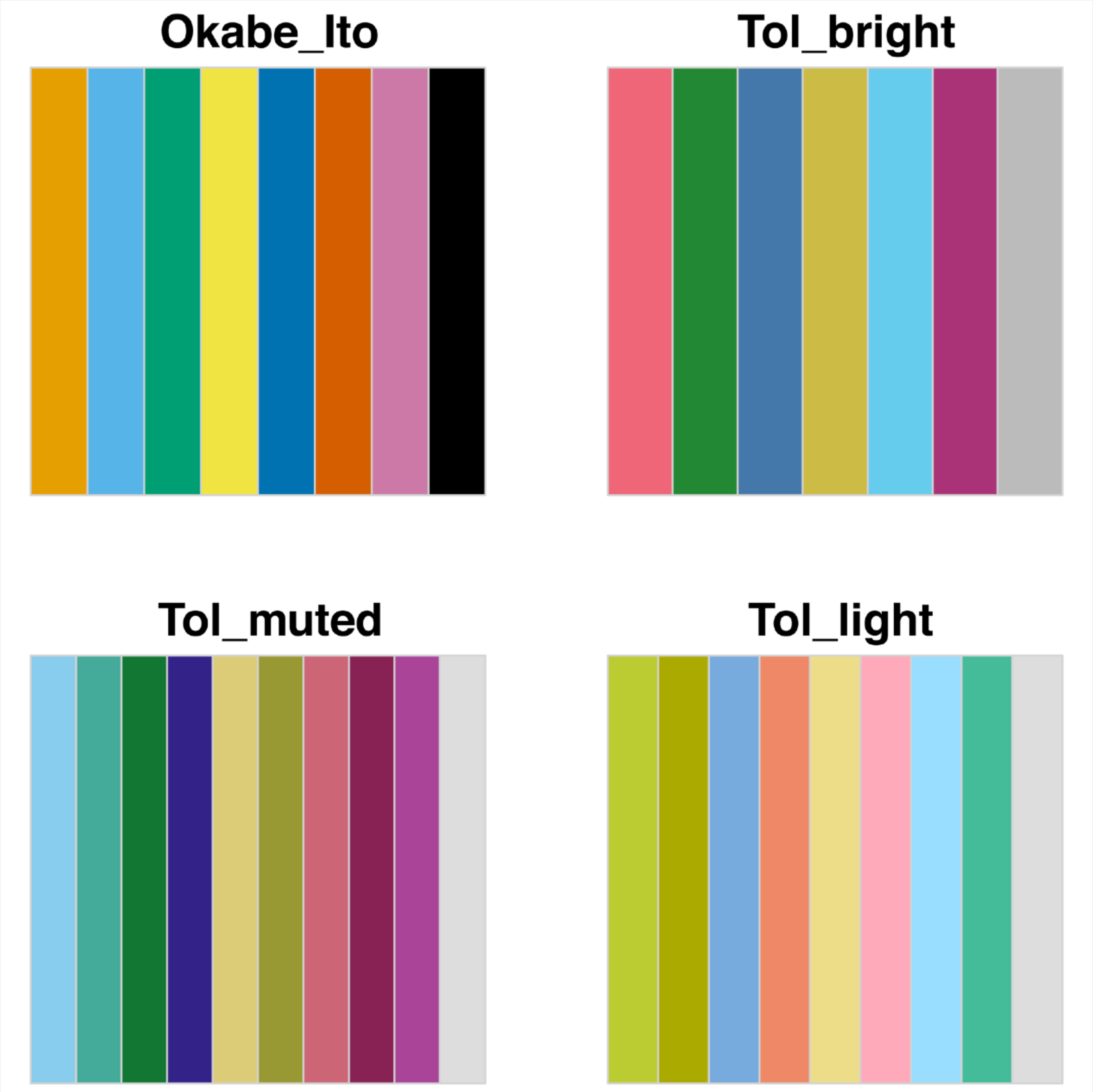
- light blues with greys
- dark purples with black
- mid-greens with blues
- oranges with reds

How to make your graphics colourblind-friendly

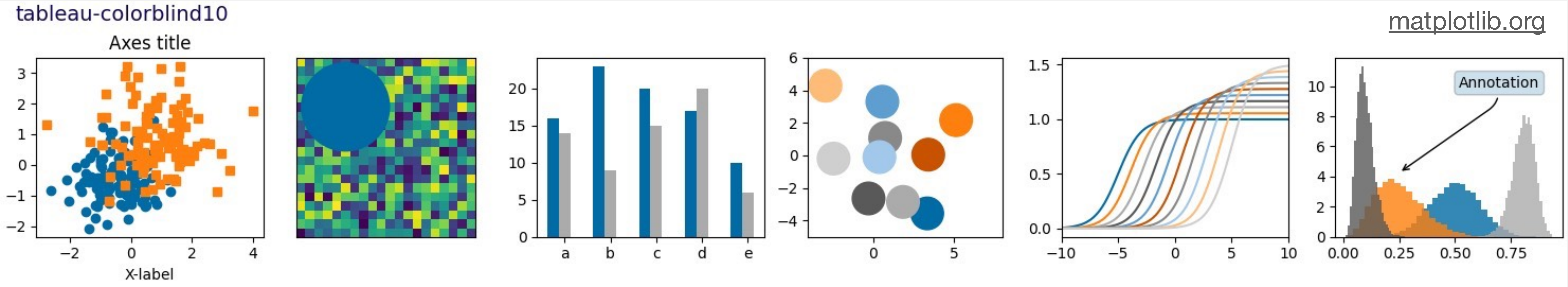
- do not convey information only with colour (use also different symbols, line styles, etc)
- keep the number of colours to a minimum
- keep contrast not only in hue but also in brightness
- communicate without using colour names (i.e. use a legend)
- consider using a colourblind-friendly palette or a colourblindness simulator (see also matplotlib resources)



https://jfly.uni-koeln.de/color/index.html



https://thenode.biologists.com/data-visualization-with-flying-colors/research/



Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

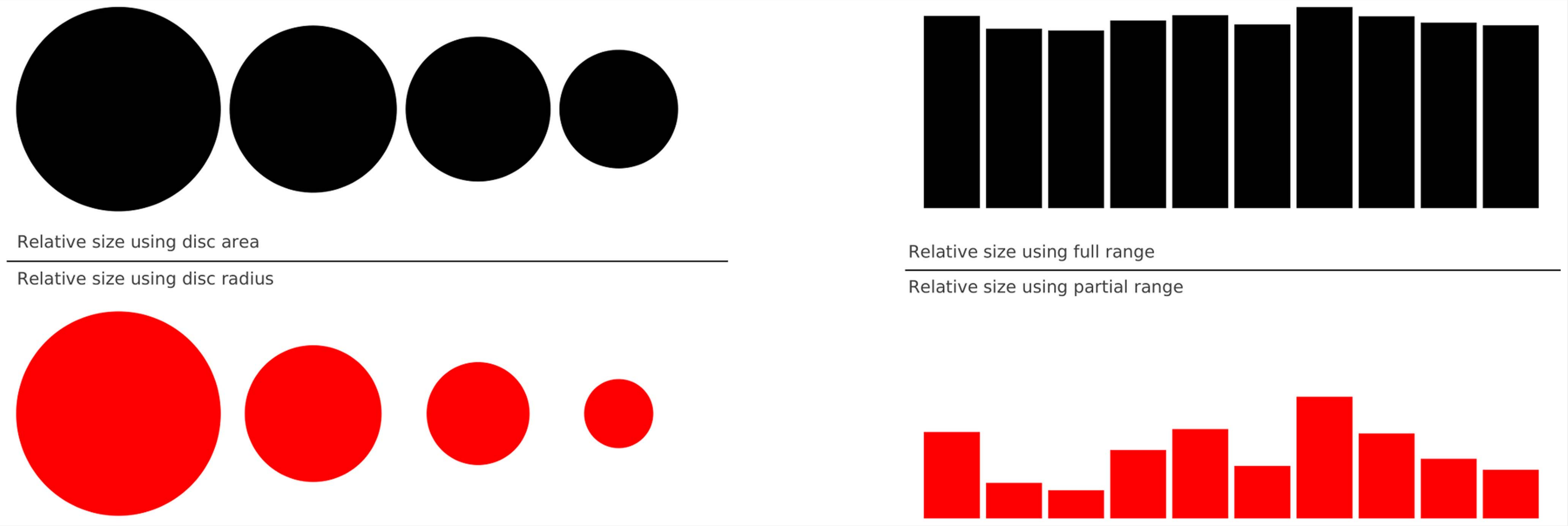


Figure 6. Do not mislead the reader. On the left part of the figure, we represented a series of four values: 30, 20, 15, 10. On the upper left part, we used the disc area to represent the value, while in the bottom part we used the disc radius. Results are visually very different. In the latter case (red circles), the last value (10) appears very small compared to the first one (30), while the ratio between the two values is only 3:1. This situation is actually very frequent in the literature because the command (or interface) used to produce circles or scatter plots (with varying point sizes) offers to use the radius as default to specify the disc size. It thus appears logical to use the value for the radius, but this is misleading. On the right part of the figure, we display a series of ten values using the full range for values on the top part (y axis goes from 0 to 100) or a partial range in the bottom part (y axis goes from 80 to 100), and we explicitly did not label the y-axis to enhance the confusion. The visual perception of the two series is totally different. In the top part (black series), we tend to interpret the values as very similar, while in the bottom part, we tend to believe there are significant differences. Even if we had used labels to indicate the actual range, the effect would persist because the bars are the most salient information on these figures.

doi:10.1371/journal.pcbi.1003833.g006

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tool

From Rougier et al., 2014, PLOS Computational Biology

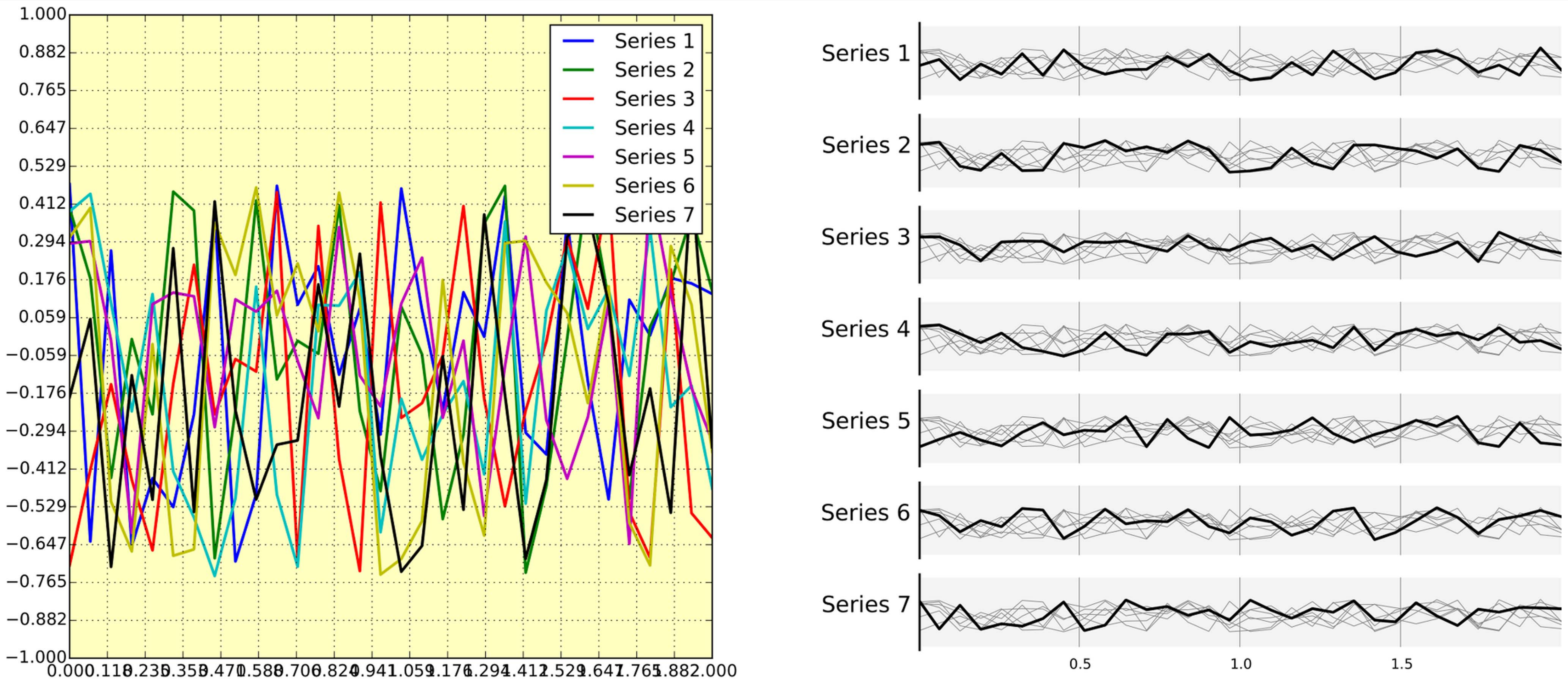


Figure 7. Avoid chartjunk. We have seven series of samples that are equally important, and we would like to show them all in order to visually compare them (exact signal values are supposed to be given elsewhere). The left figure demonstrates what is certainly one of the worst possible designs. All the curves cover each other and the different colors (that have been badly and automatically chosen by the software) do not help to distinguish them. The legend box overlaps part of the graphic, making it impossible to check if there is any interesting information in this area. There are far too many ticks: x labels overlap each other, making them unreadable, and the three-digit precision does not seem to carry any significant information. Finally, the grid does not help because (among other criticisms) it is not aligned with the signal, which can be considered discrete given the small number of sample points. The right figure adopts a radically different layout while using the same area on the sheet of paper. Series have been split into seven plots, each of them showing one series, while other series are drawn very lightly behind the main one. Series labels have been put on the left of each plot, avoiding the use of colors and a legend box. The number of x ticks has been reduced to three, and a thin line indicates these three values for all plots. Finally, y ticks have been completely removed and the height of the gray background boxes indicate the [−1,+1] range (this should also be indicated in the figure caption if it were to be used in an article).
doi:10.1371/journal.pcbi.1003833.g007

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tools

From Rougier et al., 2014, PLOS Computational Biology

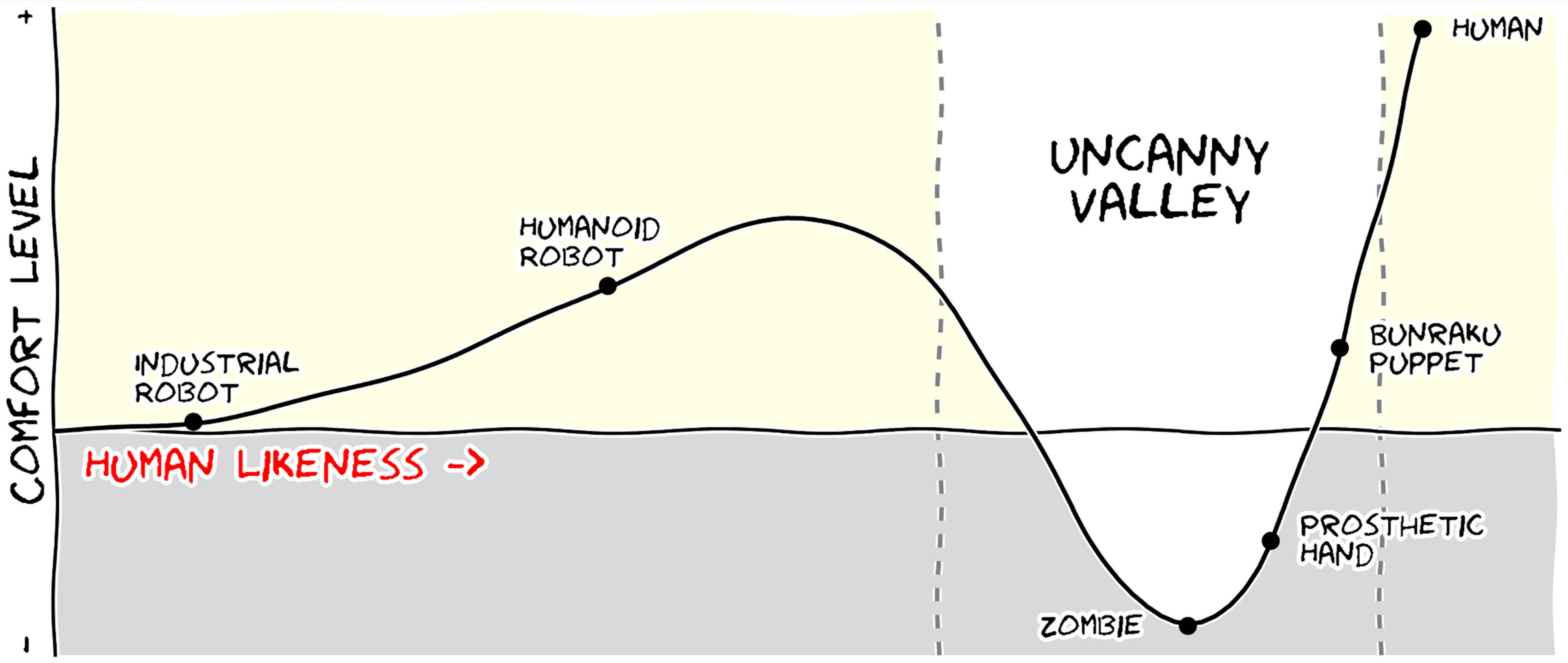


Figure 8. Message trumps beauty. This figure is an extreme case where the message is particularly clear even if the aesthetic of the figure is questionable. The uncanny valley is a well-known hypothesis in the field of robotics that correlates our comfort level with the human-likeness of a robot. To express this hypothetical nature, hypothetical data were used ($y = x^2 - 5e^{-5(x-2)^2}$) and the figure was given a sketched look (xkcd filter on matplotlib) associated with a cartoonish font that enhances the overall effect. Tick labels were also removed since only the overall shape of the curve matters. Using a sketch style conveys to the viewer that the data is approximate, and that it is the higher-level concepts rather than low-level details that are important [10].

doi:10.1371/journal.pcbi.1003833.g008

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the support medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid “chartjunk”
9. Message trumps beauty
10. Get the right tools

From Rougier et al., 2014, PLOS Computational Biology

Guidelines for Tables

- Prepare tables rather than plots when it is important to give precise numbers
- Design separate tables for separate topics
- Limit the number of tables in a paper (don't use them to show off how much data you have!) ; consider including as online-only supplementary material
- Arrange the data to allow for easy interpretation (familiar context on the left, important information on the right)
- Keep structure as simple as possible (look at similar papers for formatting ideas)
- label dependent variables in the column headings and independent variables in row headings
- Always include units!
- Set the headings apart by using different font, boldface, or capitalization (check journal guidelines)

Formulas & equations

- **Treat equations and formulas as part of the text.**
  - explain all the details of an equation
  - state any assumptions you had to make
  - explain why and how you arrived at solution
  - treat grammatically as if it were part of the text – **punctuation rules apply!**

The fraction of stellar radiation absorbed by dust in the stellar birth clouds and in the ambient ISM is reradiated in the infrared. We write the total luminosity absorbed and reradiated by dust as the sum

$$L_{\mathrm{d}}^{\mathrm{tot}}(t) = L_{\mathrm{d}}^{\mathrm{BC}}(t) + L_{\mathrm{d}}^{\mathrm{ISM}}(t), \tag{5}$$

where

$$L_{\mathrm{d}}^{\mathrm{BC}}(t) = \int_0^\infty \mathrm{d}\lambda \left(1 - \mathrm{e}^{-\hat{\tau}_\lambda^{\mathrm{BC}}}\right) \int_0^{t_0} \mathrm{d}t'\,\psi(t - t')S_\lambda(t') \tag{6}$$

is the total infrared luminosity contributed by dust in the birth clouds, and

$$L_{\mathrm{d}}^{\mathrm{ISM}}(t) = \int_0^\infty \mathrm{d}\lambda \left(1 - \mathrm{e}^{-\hat{\tau}_\lambda^{\mathrm{ISM}}}\right) \int_{t_0}^{t} \mathrm{d}t'\,\psi(t - t')S_\lambda(t') \tag{7}$$

is the total infrared luminosity contributed by dust in the ambient ISM. For some purposes, it is also convenient to define the fraction of the total infrared luminosity contributed by the dust in the ambient ISM:

$$f_\mu(t) \equiv L_{\mathrm{d}}^{\mathrm{ISM}}(t)/L_{\mathrm{d}}^{\mathrm{tot}}(t). \tag{8}$$

This depends on the total effective V-band absorption optical depth of the dust, $\hat{\tau}_V$, the fraction μ of this contributed by dust in the ambient ISM, and the star formation history (and IMF) determining the relative proportion of young and old stars in the galaxy.

da Cunha et al. (2008)