Data Visualization in Python

QB Bootcamp, Day 3
Friday, 5 September 2023
11:30am - 12:00pm

Matplotlib is the foundation for most graphics in Python

matplotlib
Version 3.1.1

Fork me on GitHub

home | examples | tutorials | API | contents » modules | index

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter notebook, web application servers, and four graphical user interface toolkits.

Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, errorcharts, scatterplots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the `pyplot` module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Installation

Visit the Matplotlib installation instructions.

Documentation

This is the documentation for Matplotlib version 3.1.1.

To get started, read the User's Guide.

Other versions are available:

- 3.1.1 Stable version.
- 2.2.4 LTS LTS version.
- 3.2.x Latest git master (unstable)
- 3.0.3 Older stable version.

Trying to learn how to do a particular kind of plot? Check out the examples gallery or the list of plotting commands.

Other learning resources

There are many external learning resources available including printed material, videos and tutorials.

Quick search

Go

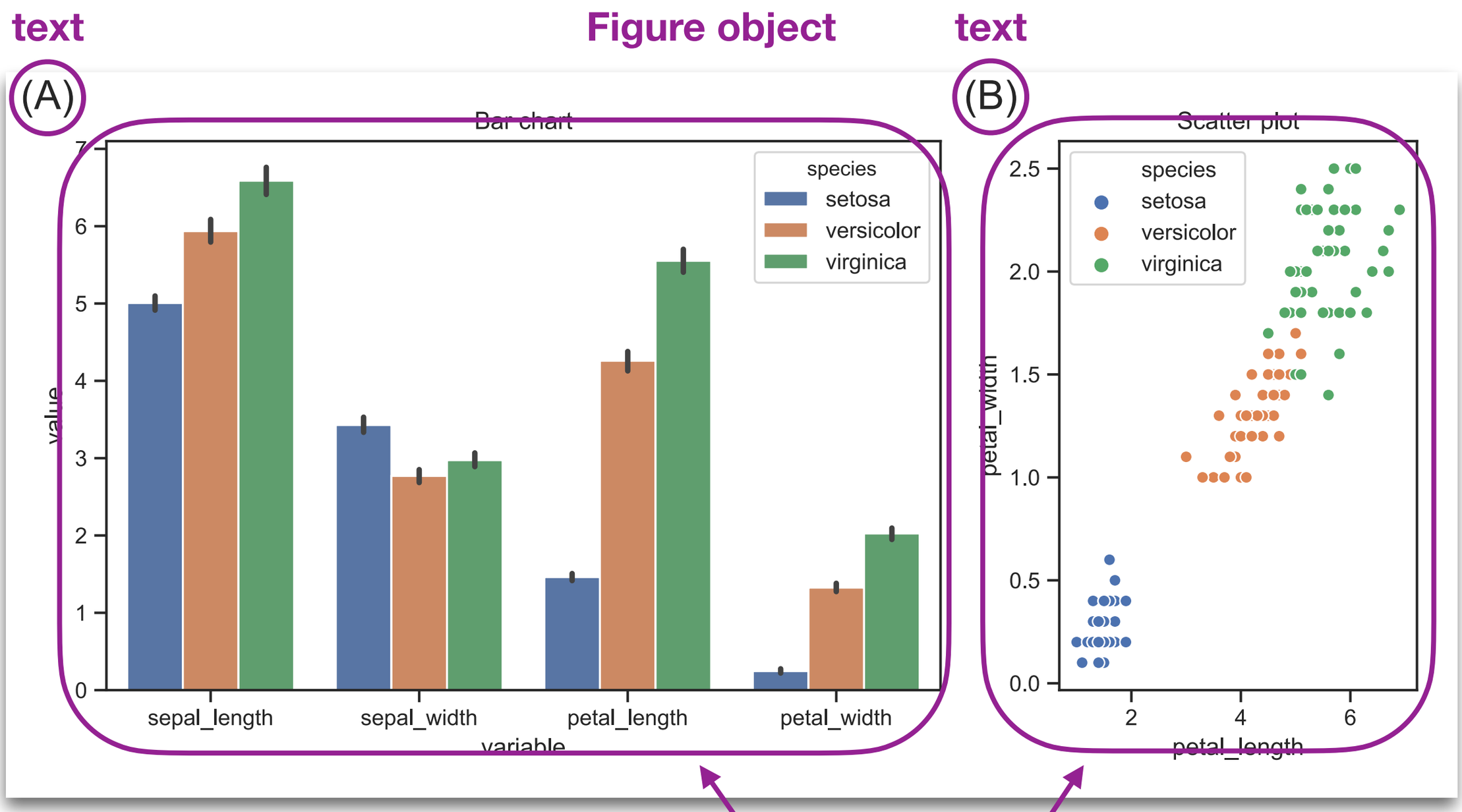
Matplotlib 3.0 is Python 3 only.

For Python 2 support, Matplotlib 2.2.x will be continued as a LTS release and updated with bugfixes until January 1, 2020.

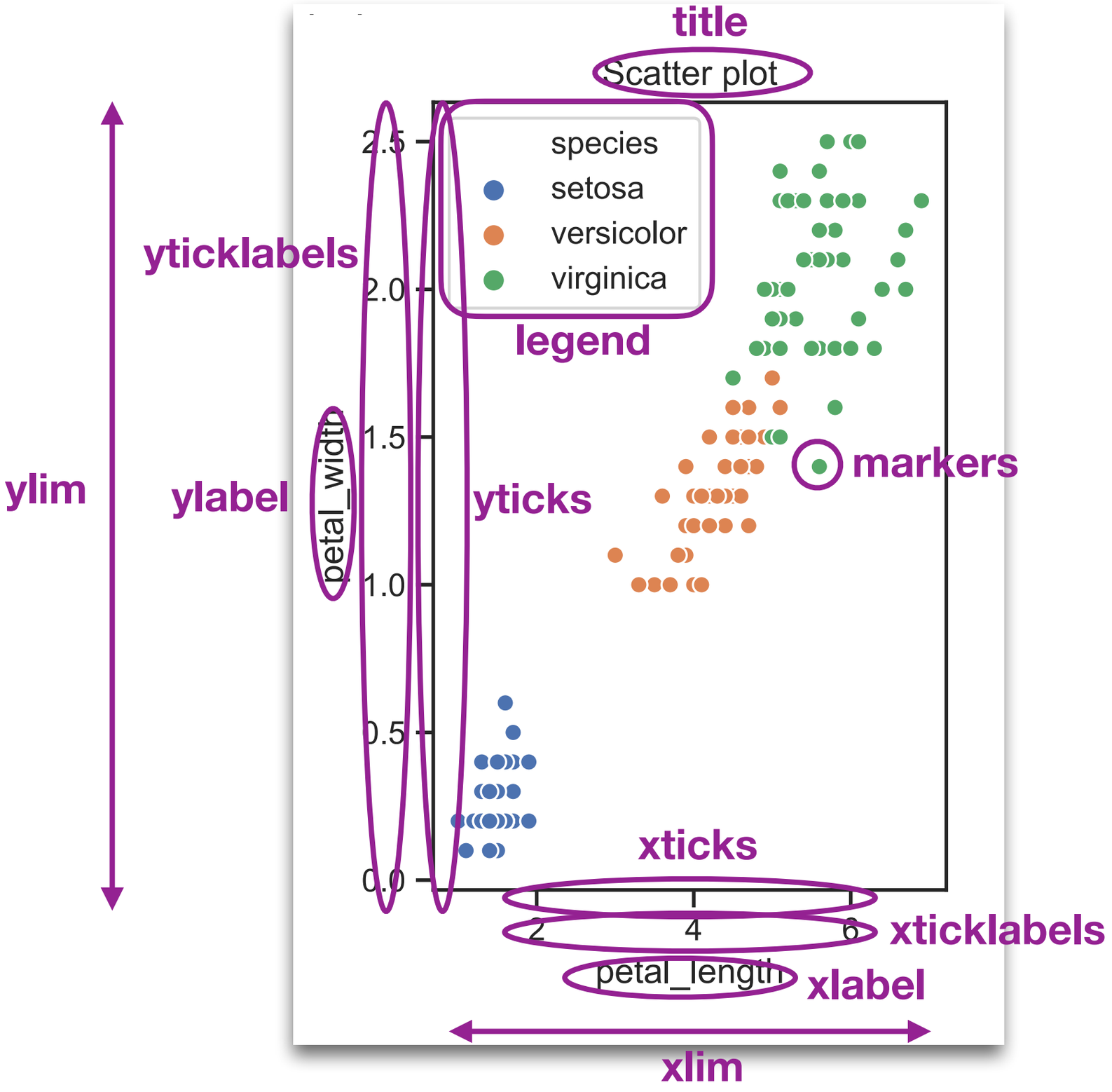
Support Matplotlib

matplotlib.org

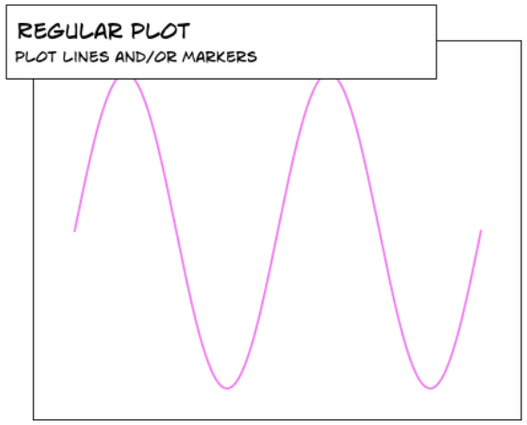
Matplotlib figures contain one or more Axes objets

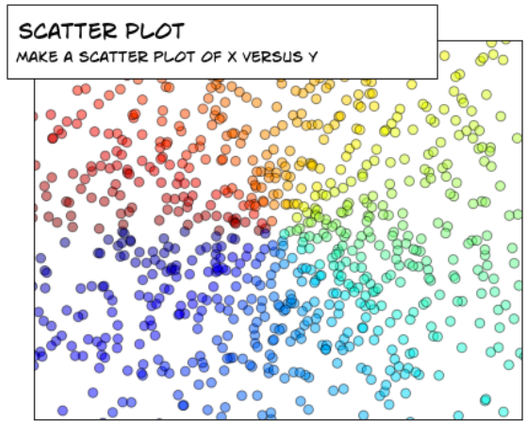


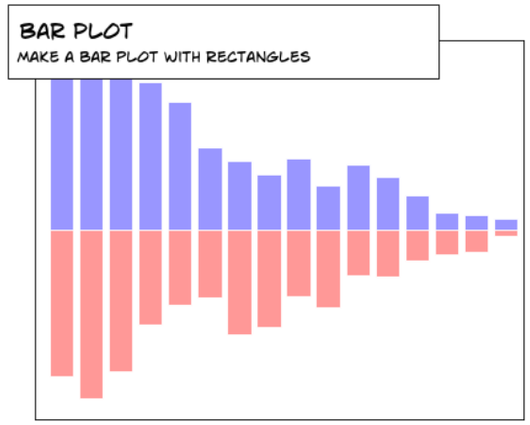
Each Axes object contains many individual elements that can be adjusted

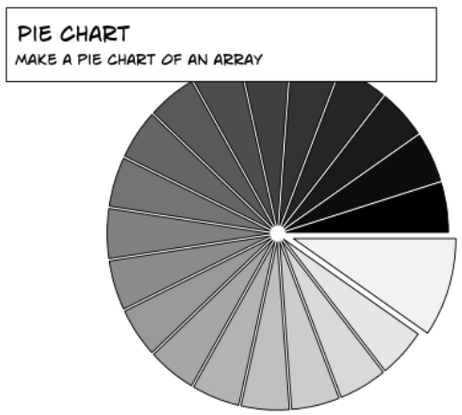


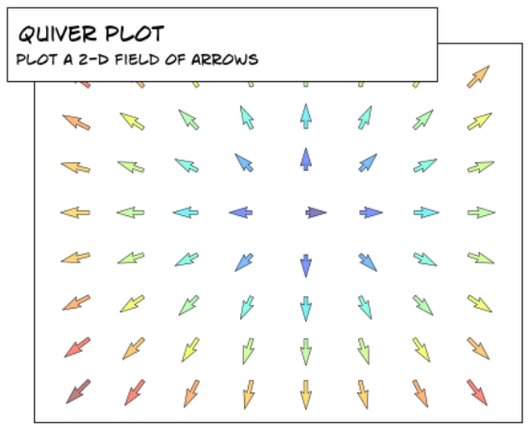
There are many, many types of plots that one can make in Matplotlib

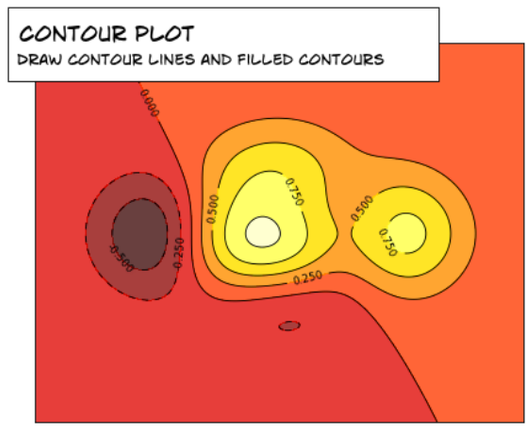


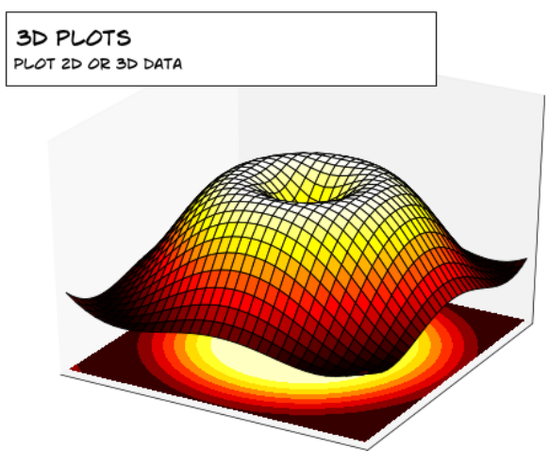


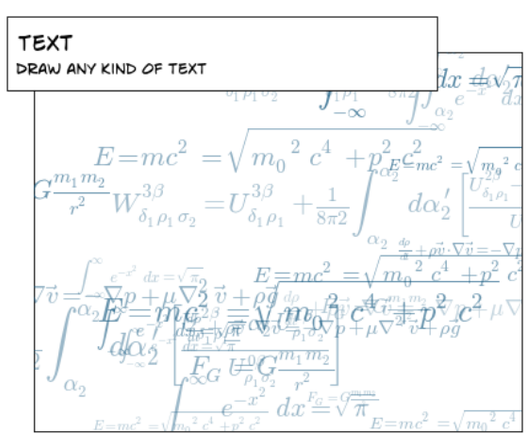


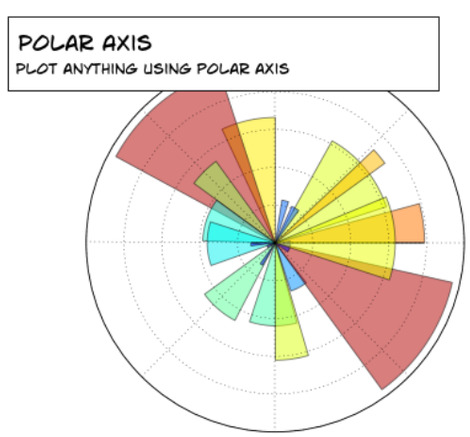








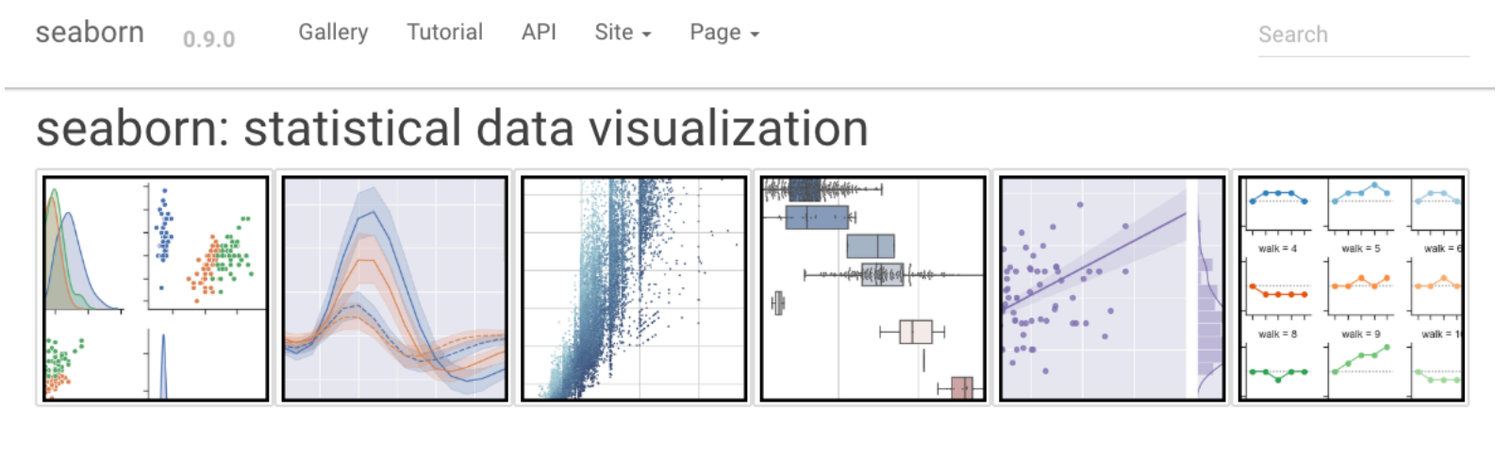




https://github.com/rougier/matplotlib-tutorial

Seaborn facilitates rapid data visualization

Seaborn is a *wrapper* for Matplotlib.
It provides methods for rapidly generating a wide range of decent-looking plots from data stored in Pandas dataframes.



Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

For a brief introduction to the ideas behind the library, you can read the introductory notes. Visit the installation page to see how you can download the package. You can browse the example gallery to see what you can do with seaborn, and then check out the tutorial and API reference to find out how.

To see the code or report a bug, please visit the github repository. General support issues are most at home on stackoverflow, where there is a seaborn tag.

Contents

- Introduction
- Release notes
- Installing
- Example gallery
- Tutorial
- API reference

Features

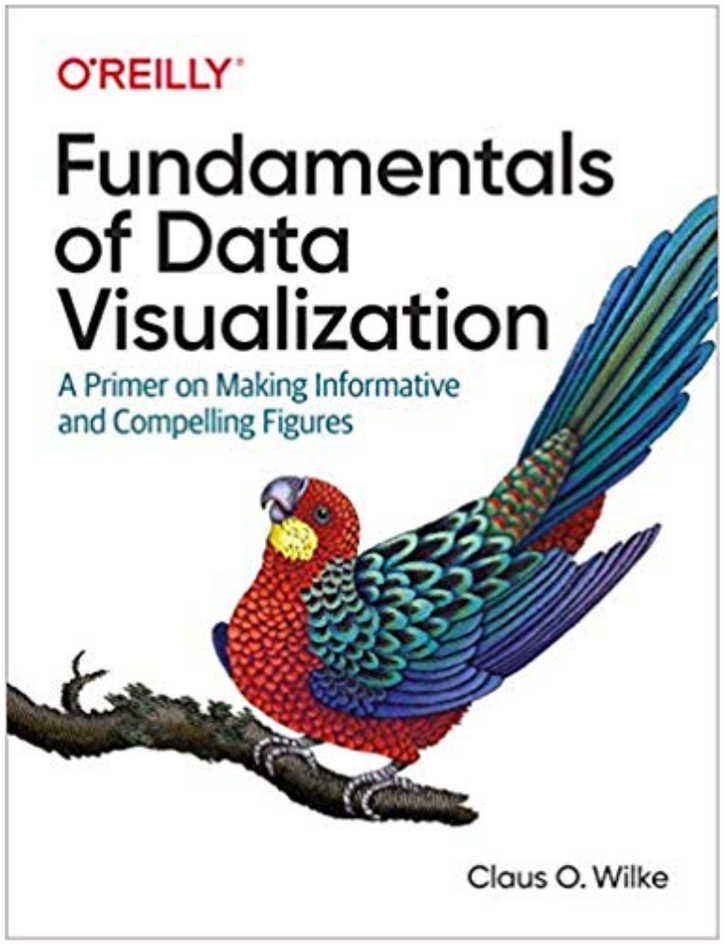
- Relational: API | Tutorial
- Categorical: API | Tutorial
- Distributions: API | Tutorial
- Regressions: API | Tutorial
- Multiples: API | Tutorial
- Style: API | Tutorial
- Color: API | Tutorial

Back to top

seaborn.pydata.org

Data visualization resources

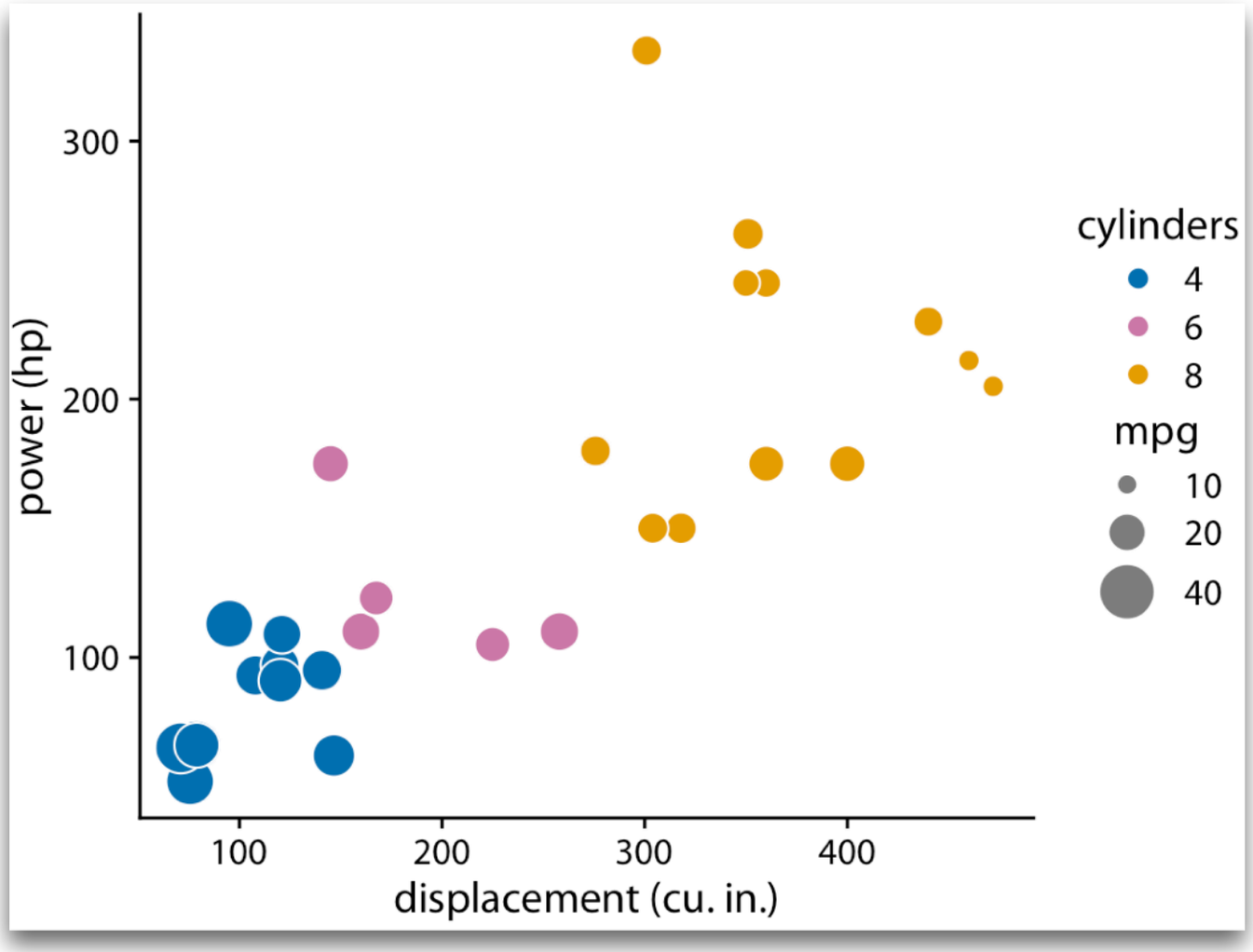
Claus Wilke gives excellent guidance on data visualization do's and don'ts

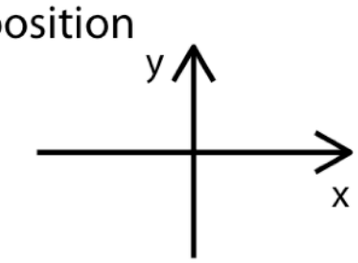


Fundamentals of Data Visualization
Wilke, 2019

All data visualization uses “aesthetics” to encode quantitative information

fuel efficiency data (4D)

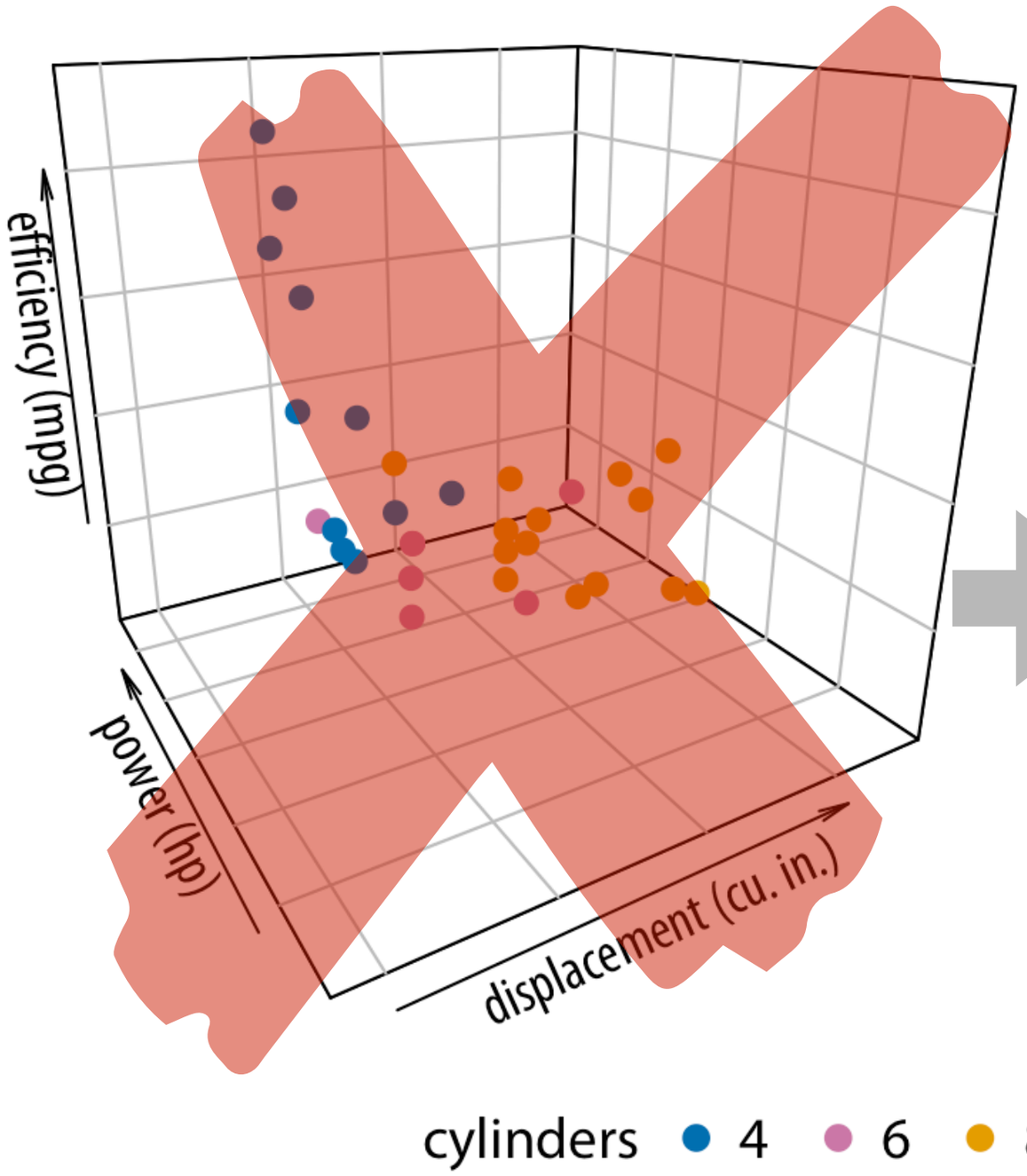


aesthetic	quantity
position 	**x: displacement** **y: power**
size 	**mpg**
color 	**# cylinders**

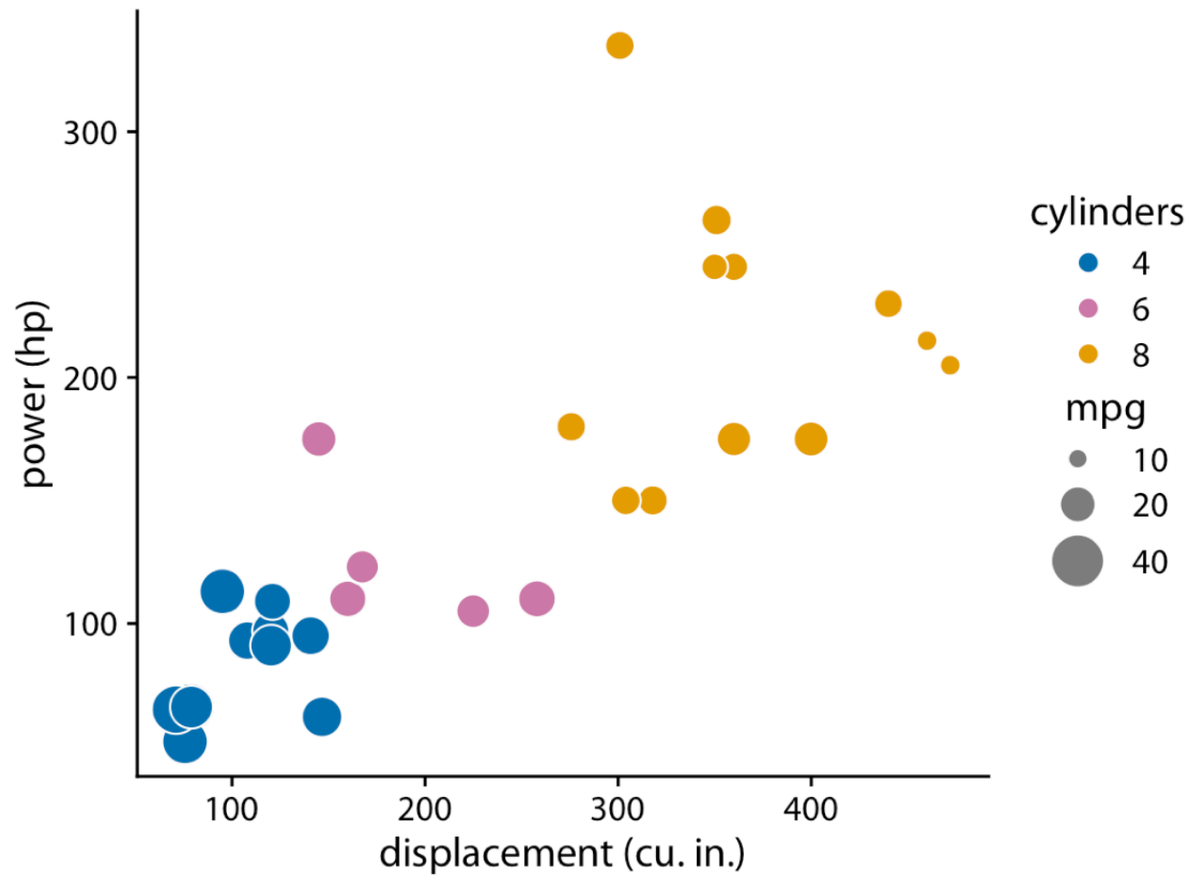
Wilke, 2019

Please don’t plot data in 3D if you can help it

In a 3D figure, it’s virtually impossible to figure out the coordinates of each point by eye.



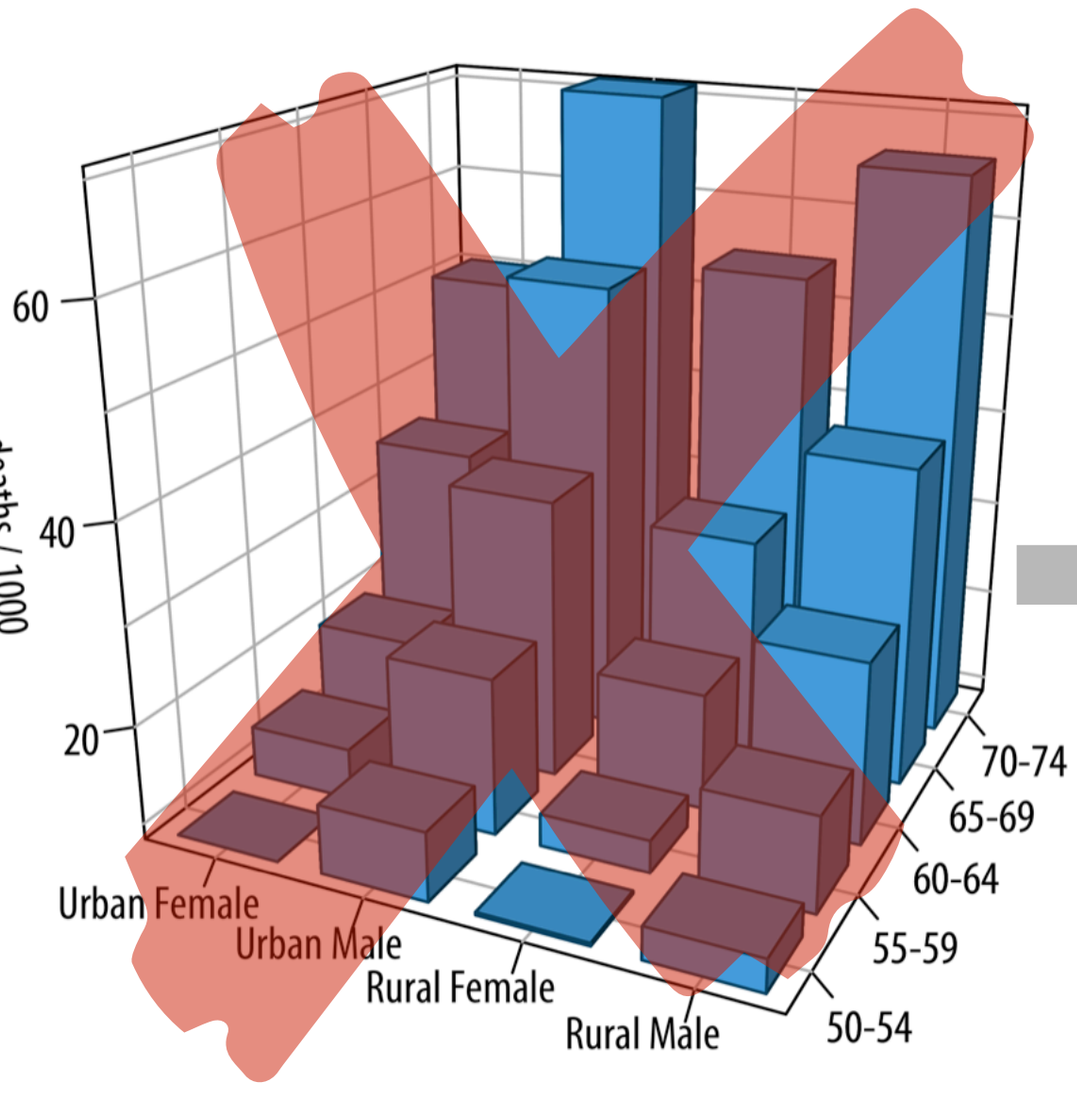
Plot in 2D using multiple aesthetics instead



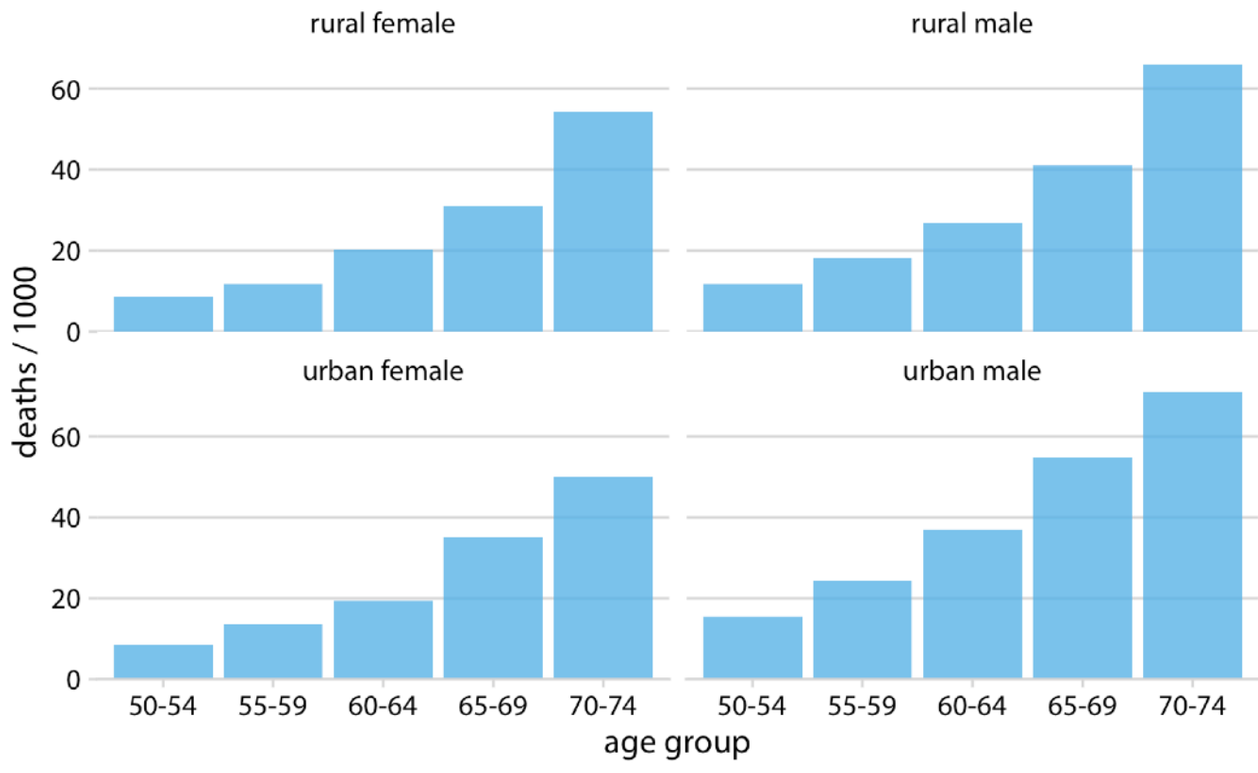
Wilke, 2019

Please don’t plot data in 3D if you can help it

In a 3D figure, it’s virtually impossible to figure out the coordinates of each point by eye.



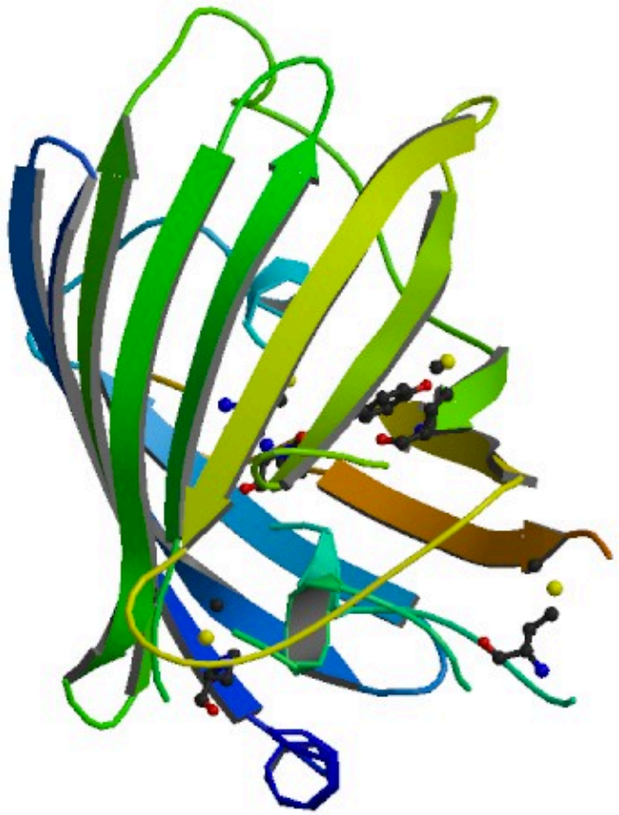
Use “small multiples” of 2D plots



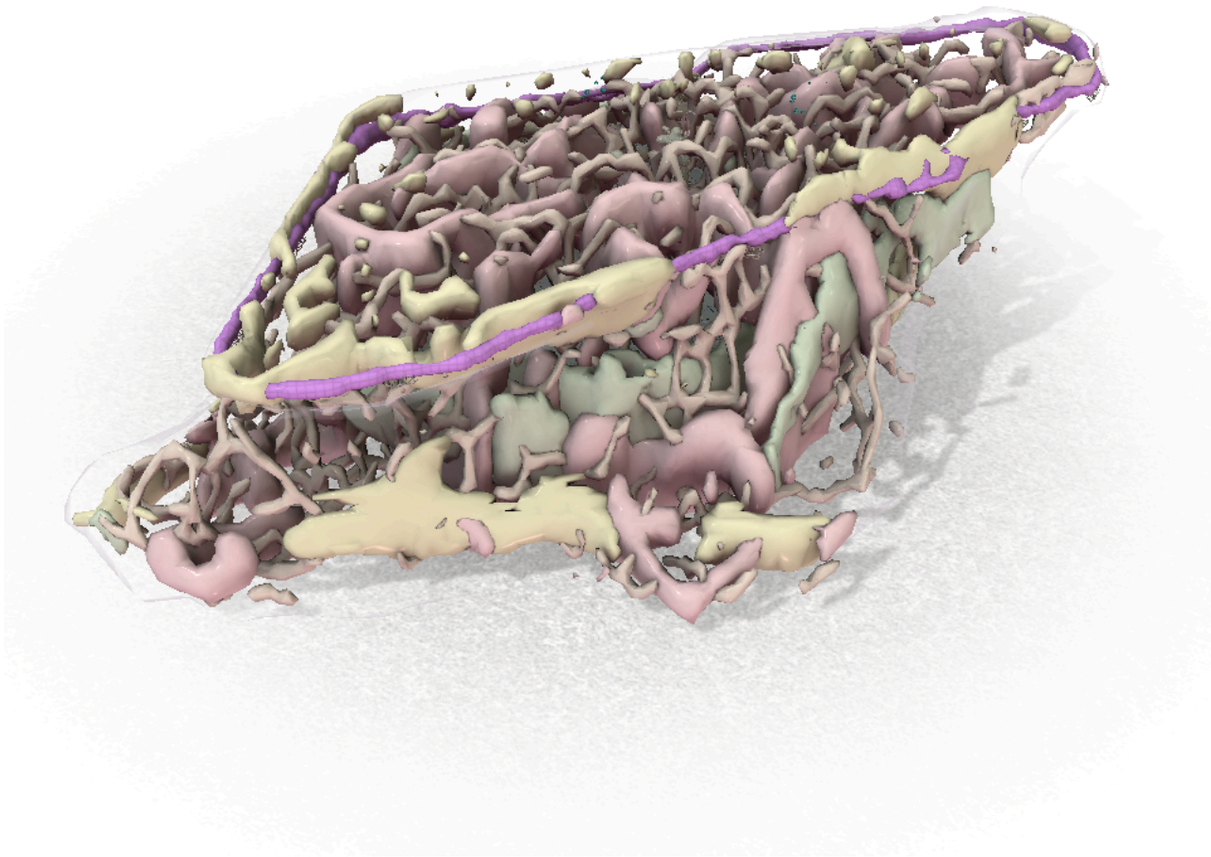
Wilke, 2019

Only use 3D if you're illustrating some inherently 3D object

Structure of GFP
(PDB:1EMA)

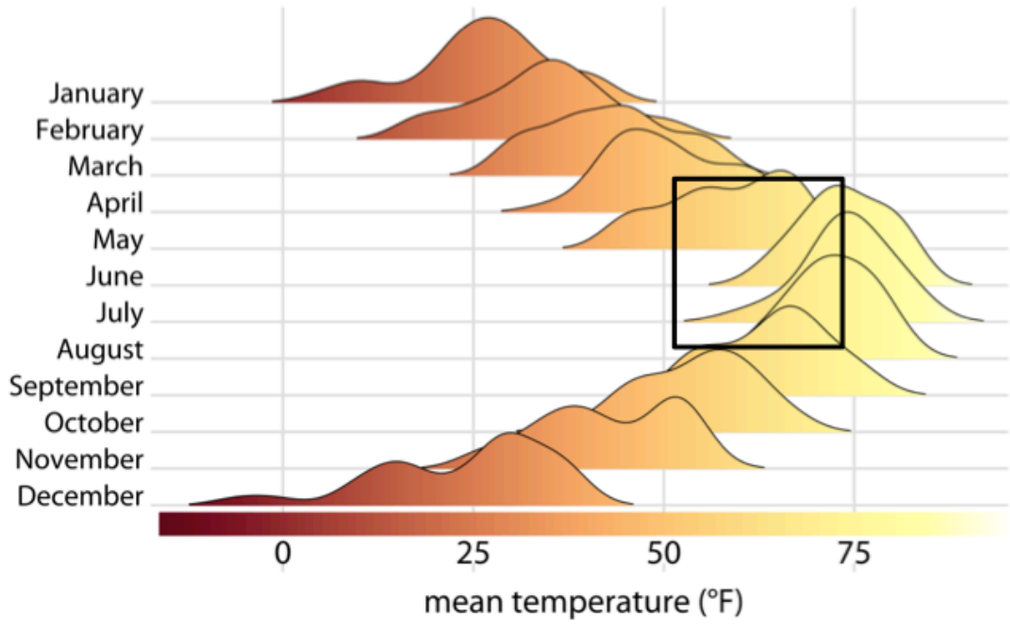


cellular structures
(Allen Cell Explorer)

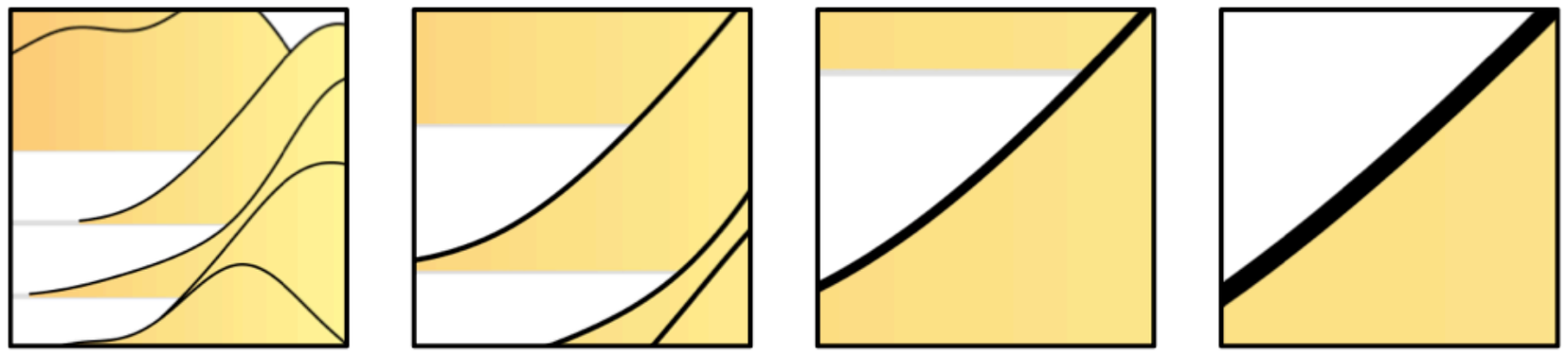


https://www.rcsb.org/structure/1ema
https://www.allencell.org/visual-guide-to-human-cells.html

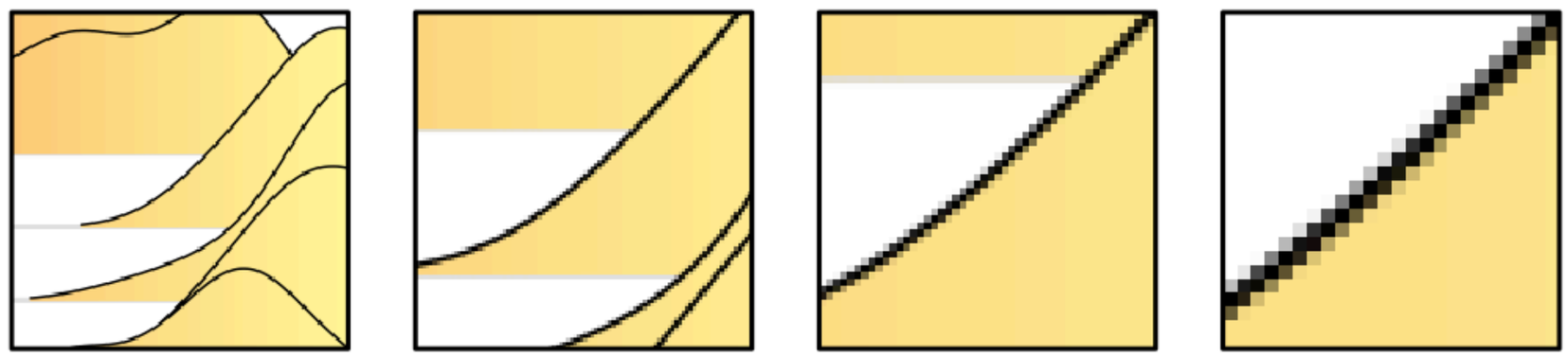
Keep in mind the difference between vector graphics and bitmaps



vector format
(PDF, SVG)

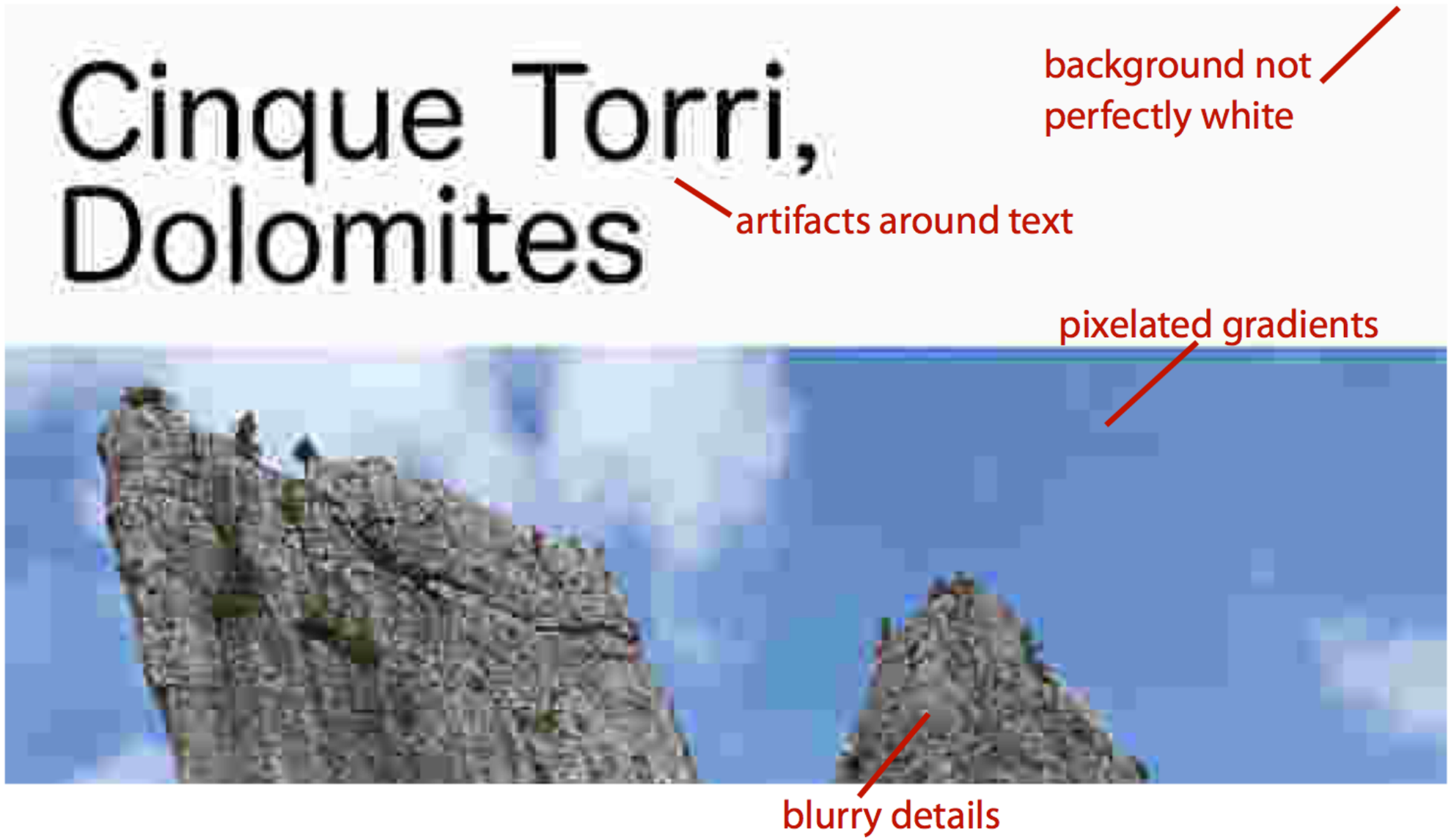


bitmap format
(TIFF, PNG, JPEG)



Wilke, 2019

JPEG is often good for compressing images, but can produce strange artifacts



Wilke, 2019

Quick guide to graphics file formats

PDF

- Standard vector graphics format
- Easily imported and edited in Illustrator
- Bad for plotting many datapoints

PNG

- Lossless bitmap compression
- Standard bitmap format for plots
- Doesn’t compress natural images well

JPEG

- Lossy image compression
- Standard bitmap format for images
- Can produce strange artifacts in plots

Tips on making figures

- Be a perfectionist.
- Draw your figures so that they look like figures in your target journal.
- Use Adobe Illustrator.
- Make all figures the correct size to begin with (use size guidelines for target journal).
- Make all figures in a single .ai file.
- Computationally generate plots so that they can be “placed” within the .ai file behind any artwork, and without requiring any modification.
- Write scripts to automate the generating of publication-ready plots.
- Separate your computationally expensive analysis scripts from your figure-generating scripts.

Dropbox/20_mavenn/manuscri | fig6 - Jupyter Notebook

localhost:8888/notebooks/Dropbox/20_mavenn/manuscript/v9_nat_meth/scripts_m1/Figure_6_script_M1/fig6.ipynb

Apps | Toggl | URP Course | Nature Methods | DropFiles | VSTMS github | Papers | MAVE-NN (readth... | MAVE-NN (local) | Poll Worker Dashb... | Reading List

jupyter fig6 Last Checkpoint: 06/22/2021 (unsaved changes) Logout

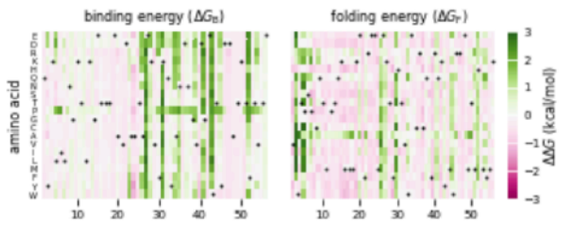
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

Run Code Notify: Disabled auto-scroll threshold: 100 (default)

```
ax.set_title('   folding energy ($\Delta G_\mathrm{F}$)',fontsize=fontsize)
ax.tick_params(labelsize=5, rotation=0, axis='y')
ax.set_yticks([])
ax.tick_params(rotation=0, axis='x')
ax.set_xticks(range(10,60,10))

#fig.subplots_adjust(left=3, bottom=1, right=5, top=2, wspace=5, hspace=None)
# Fix up plot
fig.tight_layout(h_pad=1,w_pad=2)
fig.savefig('fig_panels/panel_b.png', bbox_inches='tight', dpi=300, facecolor='white')
plt.show()
```

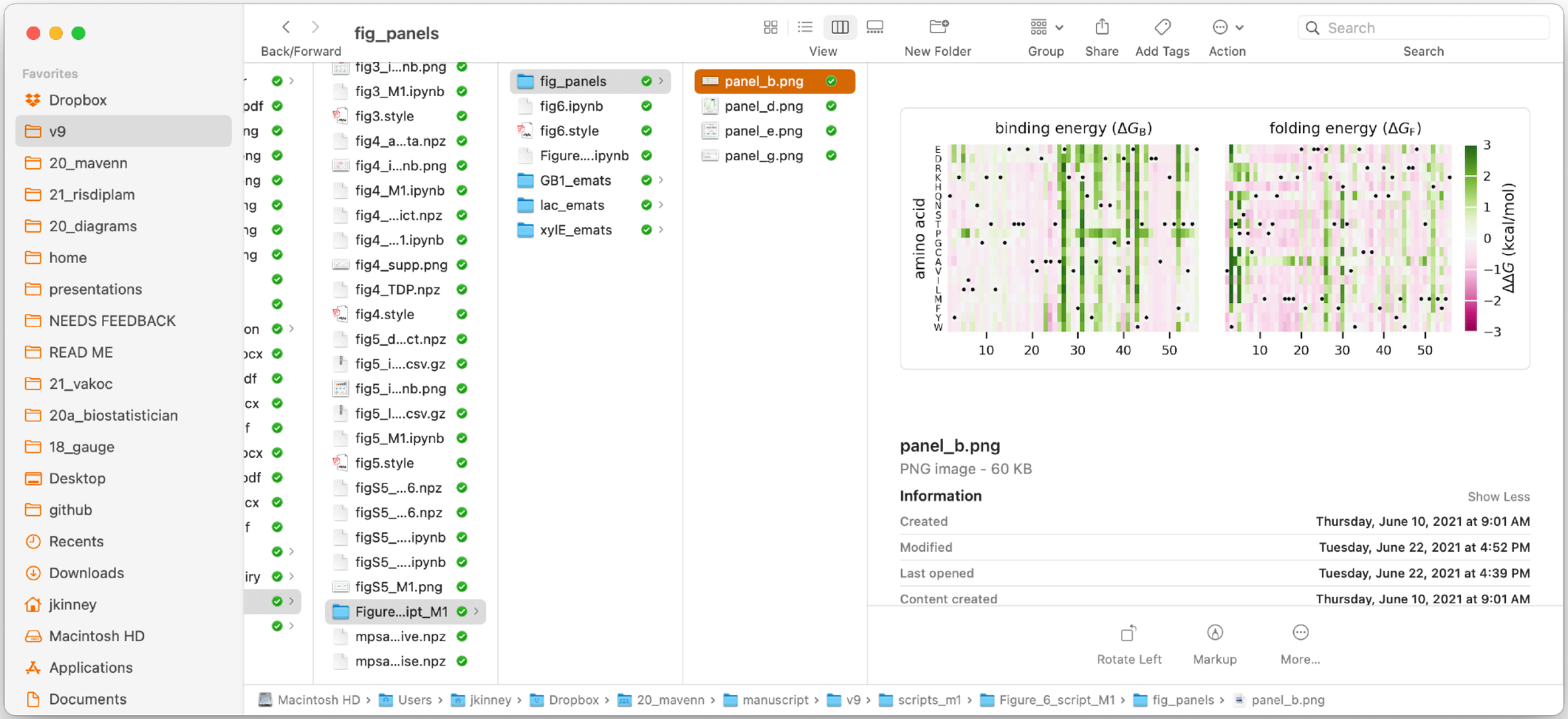
executed in 546ms, finished 16:52:56 2021-06-22

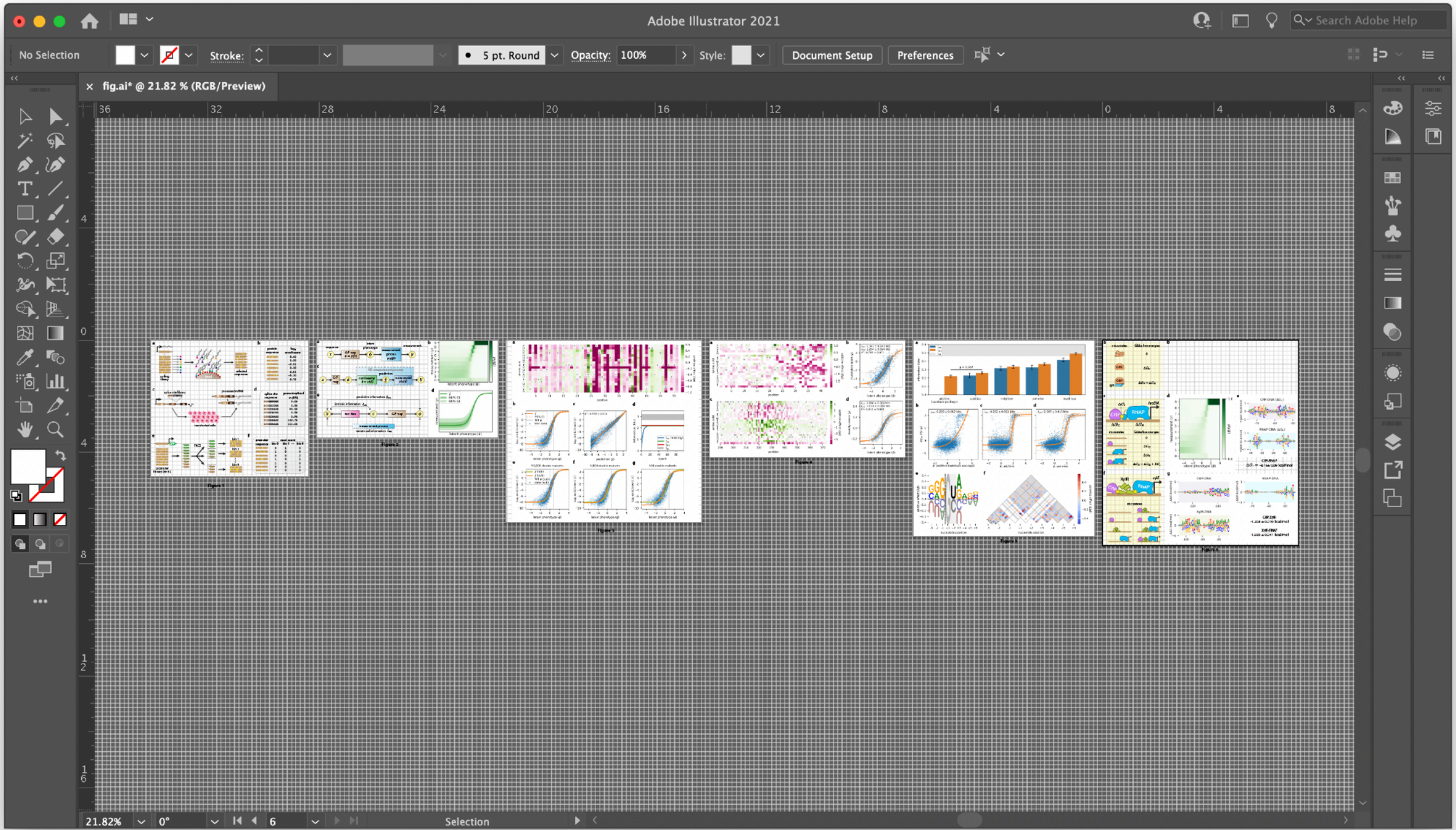


Load xylE data

In [7]:

```
style_file_name = 'fig6.style'
s = """
axes.linewidth:     0.5     # edge linewidth
font.size:          7.0
axes.labelsize:     7.0  # fontsize of the x any y labels
axes.titlesize:     7.0
xtick.labelsize:    7.0  # fontsize of the tick labels
ytick.labelsize:    7.0  # fontsize of the tick labels
legend.fontsize:      8.0
legend.borderpad:     0.2  # border whitespace
legend.labelspacing:  0.2  # the vertical space between the legend entries
legend.borderaxespad: 0.2  # the border between the axes and legend edge
legend.framealpha:    1.0
"""
```


Favorites
Dropbox
v9
20_mavenn
21_risdiplam
20_diagrams
home
presentations
NEEDS FEEDBACK
READ ME
21_vakoc
20a_biostatistician
18_gauge
Desktop
github
Recents
Downloads
jkinney
Macintosh HD
Applications
Documents
Back/Forward
fig_panels
View
New Folder
Group
Share
Add Tags
Action
Search
fig3_i...nb.png
fig3_M1.ipynb
fig3.style
fig4_a...ta.npz
fig4_i...nb.png
fig4_M1.ipynb
fig4_...ict.npz
fig4_...1.ipynb
fig4_supp.png
fig4_TDP.npz
fig4.style
fig5_d...ct.npz
fig5_i....csv.gz
fig5_i...nb.png
fig5_l....csv.gz
fig5_M1.ipynb
fig5.style
figS5_...6.npz
figS5_...6.npz
figS5_....ipynb
figS5_....ipynb
figS5_M1.png
Figure...ipt_M1
mpsa...ive.npz
mpsa...ise.npz
fig_panels
fig6.ipynb
fig6.style
Figure....ipynb
GB1_emats
lac_emats
xylE_emats
panel_b.png
panel_d.png
panel_e.png
panel_g.png
binding energy (ΔG_B)
folding energy (ΔG_F)
amino acid
E D R K H Q N S T P G C A V I L M F Y W
10 20 30 40 50
3 2 1 0 −1 −2 −3
ΔΔG (kcal/mol)
panel_b.png
PNG image - 60 KB
Information
Show Less
Created
Thursday, June 10, 2021 at 9:01 AM
Modified
Tuesday, June 22, 2021 at 4:52 PM
Last opened
Tuesday, June 22, 2021 at 4:39 PM
Content created
Thursday, June 10, 2021 at 9:01 AM
Rotate Left
Markup
More...
Macintosh HD › Users › jkinney › Dropbox › 20_mavenn › manuscript › v9 › scripts_m1 › Figure_6_script_M1 › fig_panels › panel_b.png


Adobe Illustrator 2021
Search Adobe Help
No Selection
Stroke:
5 pt. Round
Opacity:
100%
Style:
Document Setup
Preferences
fig.ai* @ 21.82 % (RGB/Preview)
21.82%
0°
6
Selection

Illustrator File Edit Object Type Select Effect View Window Help — Wed Jun 30 7:57:02

Adobe Illustrator 2021

No Selection | 5 pt. Round | Opacity: 100% | Style: | Document Setup | Preferences | Search Adobe Help

fig.ai*

File menu:
- New... ⌘N
- New from Template... ⇧⌘N
- Open... ⌘O
- Open Recent Files
- Browse in Bridge... ⌥⌘O
- Close ⌘W
- Close All ⌥⌘W
- Save ⌘S
- Save As... ⇧⌘S
- Save a Copy... ⌥⌘S
- Save Selected Slices...
- Version History
- Save as Template...
- Revert F12
- Search Adobe Stock...
- Place... ⇧⌘P
- Export
- Export Selection...
- Package... ⌥⇧⌘P
- Scripts
- Document Setup... ⌥⌘P
- Document Color Mode
- File Info... ⌥⇧⌘I
- Print... ⌘P

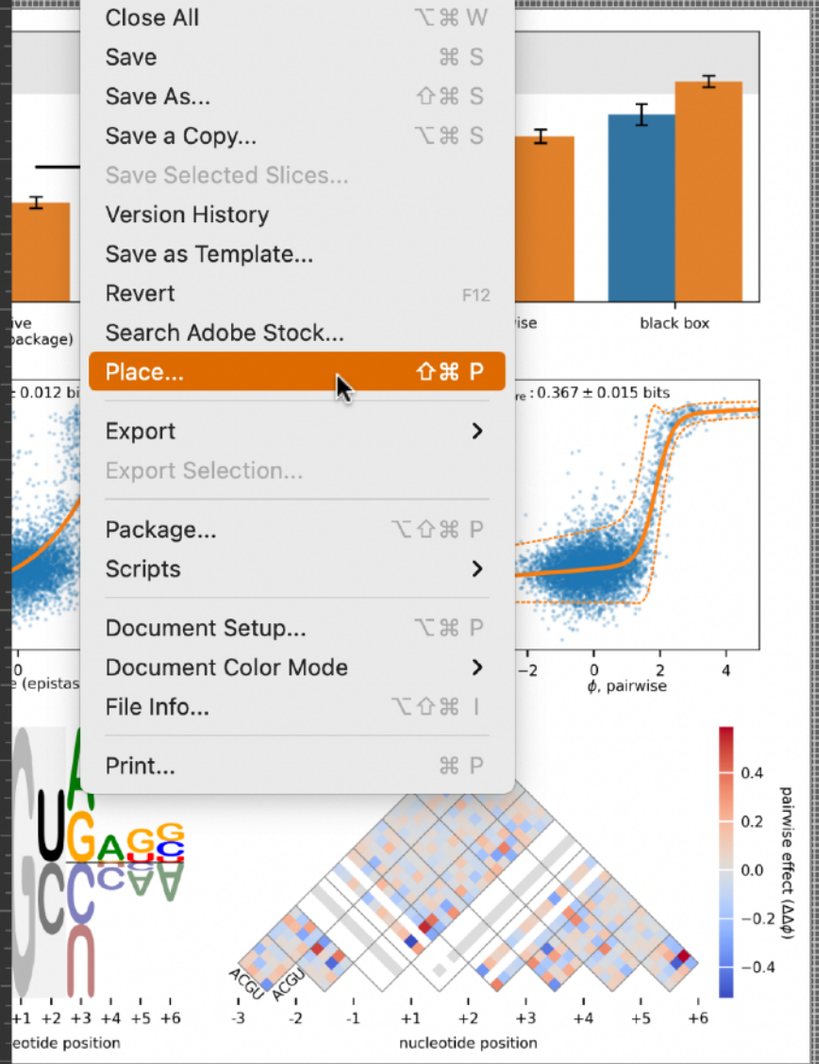


Figure 5

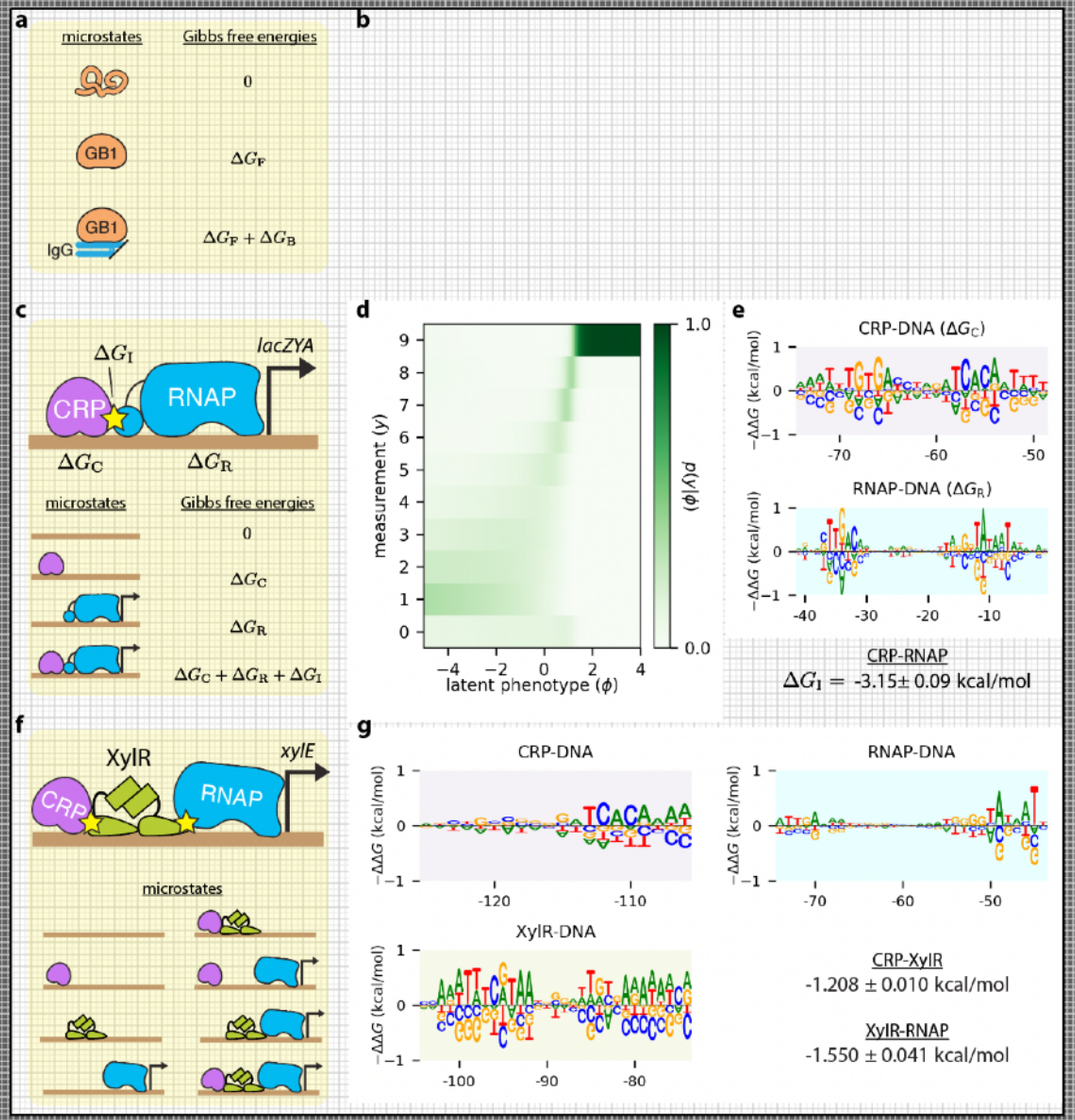


Figure 6

65.92% | 0° | 6 | Selection

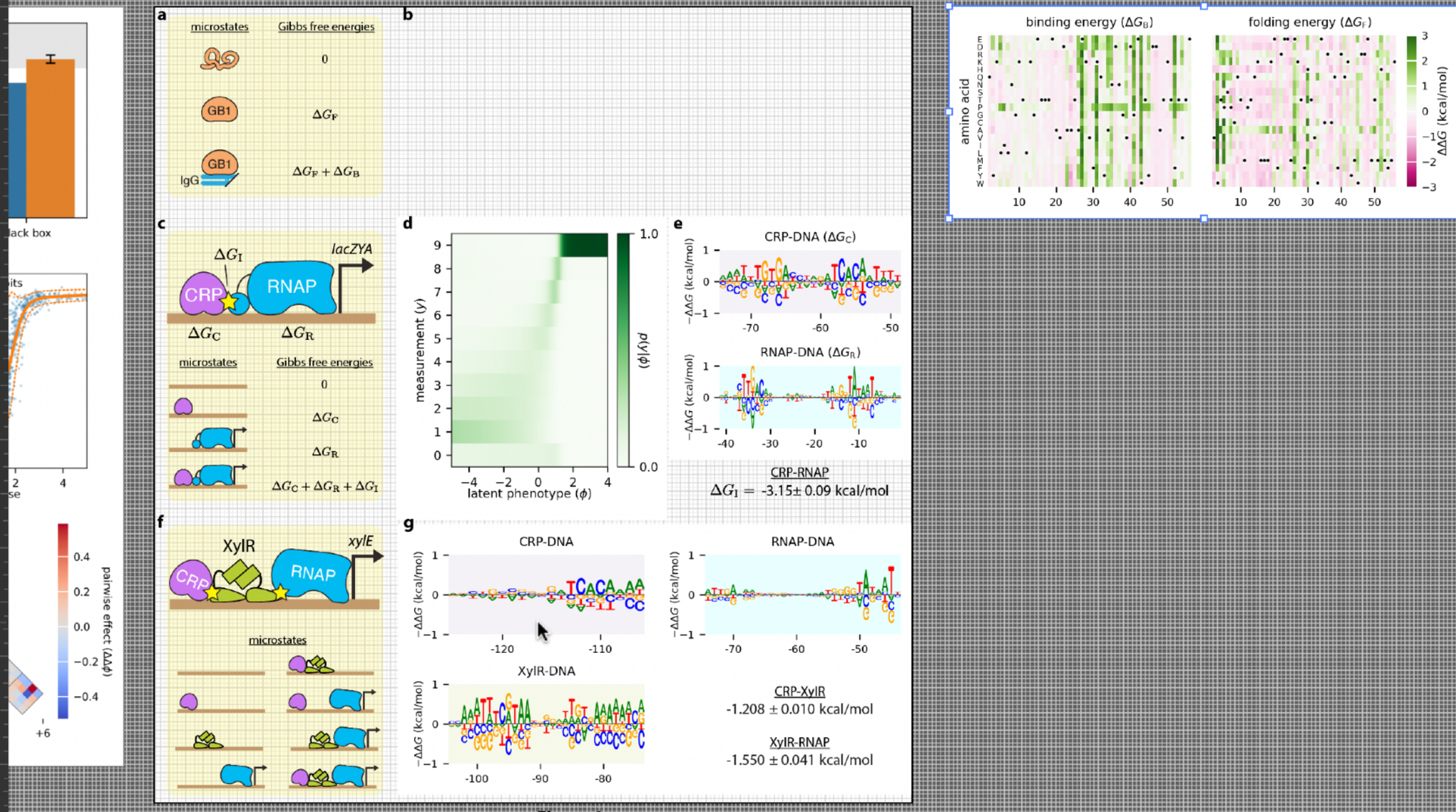

a
microstates
Gibbs free energies
0
GB1
ΔG_F
GB1
IgG
$\Delta G_F + \Delta G_B$
b
binding energy (ΔG_B)
folding energy (ΔG_F)
amino acid
$\Delta\Delta G$ (kcal/mol)
c
ΔG_I
lacZYA
CRP
RNAP
ΔG_C
ΔG_R
microstates
Gibbs free energies
0
ΔG_C
ΔG_R
$\Delta G_C + \Delta G_R + \Delta G_I$
d
measurement (y)
latent phenotype (ϕ)
$p(y|\phi)d$
e
CRP-DNA (ΔG_C)
RNAP-DNA (ΔG_R)
$-\Delta\Delta G$ (kcal/mol)
CRP-RNAP
ΔG_I = -3.15± 0.09 kcal/mol
f
XylR
xylE
CRP
RNAP
microstates
g
CRP-DNA
RNAP-DNA
XylR-DNA
$-\Delta\Delta G$ (kcal/mol)
CRP-XylR
-1.208 ± 0.010 kcal/mol
XylR-RNAP
-1.550 ± 0.041 kcal/mol
pairwise effect ($\Delta\Delta\phi$)
+6

Figure 6

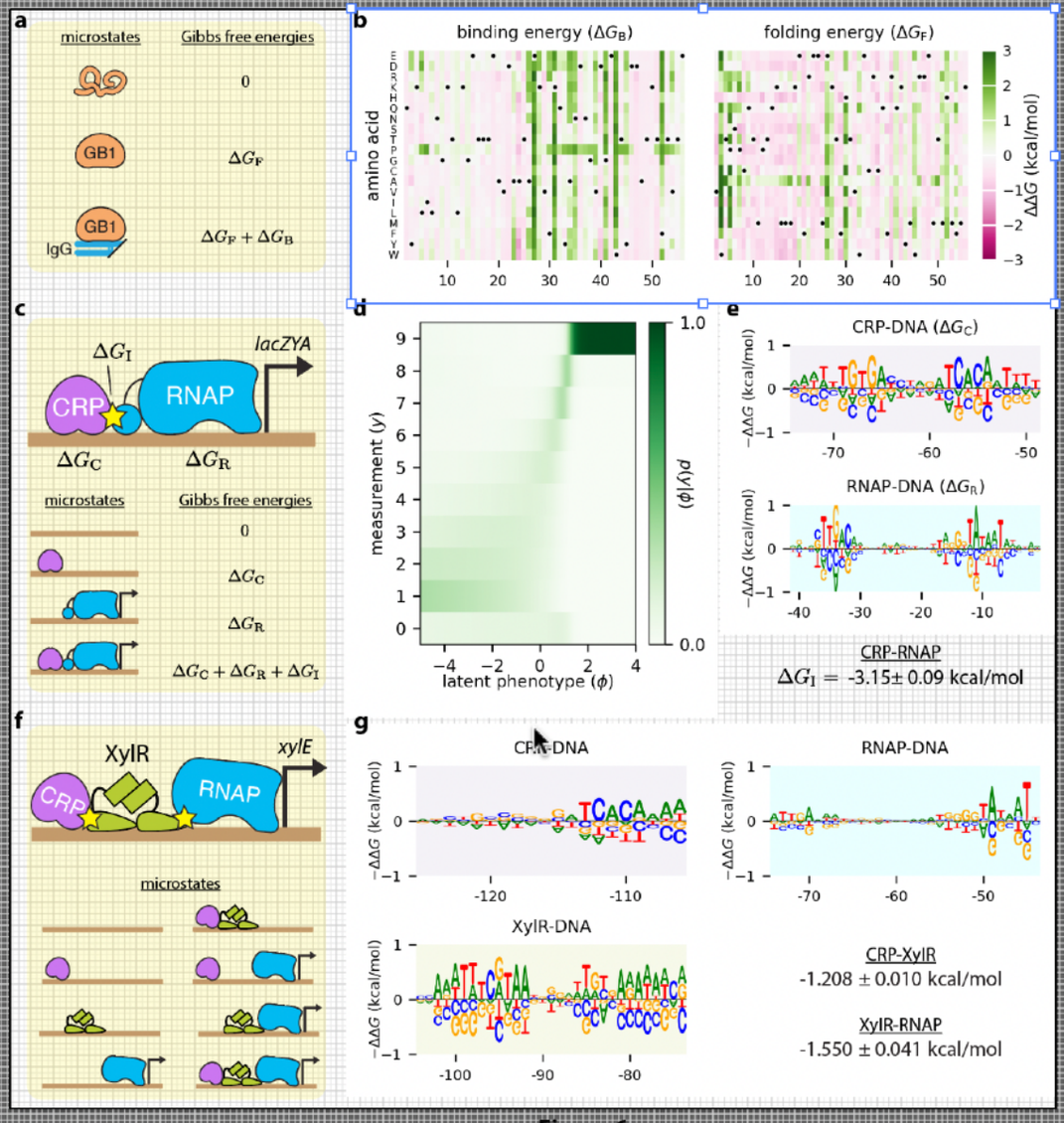

a
microstates
Gibbs free energies
GB1
IgG
ΔG_F
$\Delta G_F + \Delta G_B$
b
binding energy (ΔG_B)
folding energy (ΔG_F)
amino acid
$\Delta\Delta G$ (kcal/mol)
c
ΔG_I
lacZYA
CRP
RNAP
ΔG_C
ΔG_R
$\Delta G_C + \Delta G_R + \Delta G_I$
d
measurement (y)
latent phenotype (ϕ)
$p(y|\phi)$
e
CRP-DNA (ΔG_C)
RNAP-DNA (ΔG_R)
$-\Delta\Delta G$ (kcal/mol)
CRP-RNAP
ΔG_I = -3.15± 0.09 kcal/mol
f
XylR
xylE
g
CRP-DNA
RNAP-DNA
XylR-DNA
CRP-XylR
-1.208 ± 0.010 kcal/mol
XylR-RNAP
-1.550 ± 0.041 kcal/mol

Figure 6

Illustrator File Edit Object Type Select Effect View Window Help — Wed Jun 30 7:58:55

Adobe Illustrator 2021

Image | panel_b.png Transparent RGB PPI: 300 | Unembed | Edit Original | Image Trace | Mask | Crop Image | Opacity: 100% | Transform

fig.ai* @ 65.92 % (RGB/Preview)

Export

Save As: fig.png

Tags:

v9

Search

Favorites
- Dropbox
- v9
- 20_mavenn
- 21_risdipl...
- 20_diagr...
- home
- presentat...
- NEEDS F...
- READ ME
- 21_vakoc
- 20a_bios...
- 18_gauge

Enrich-0.2.tar.gz
GB1_models
laub
logo
manuscript

v3
v4
v5_biorxiv
v6
v9

cover_letter
cover_letter.pdf
fig_1_exp...ments.png
fig_2_models.png
fig_3_gb1.png

Format: PNG (png)

Use Artboards | All | Range: 6

New Folder | Cancel | Export

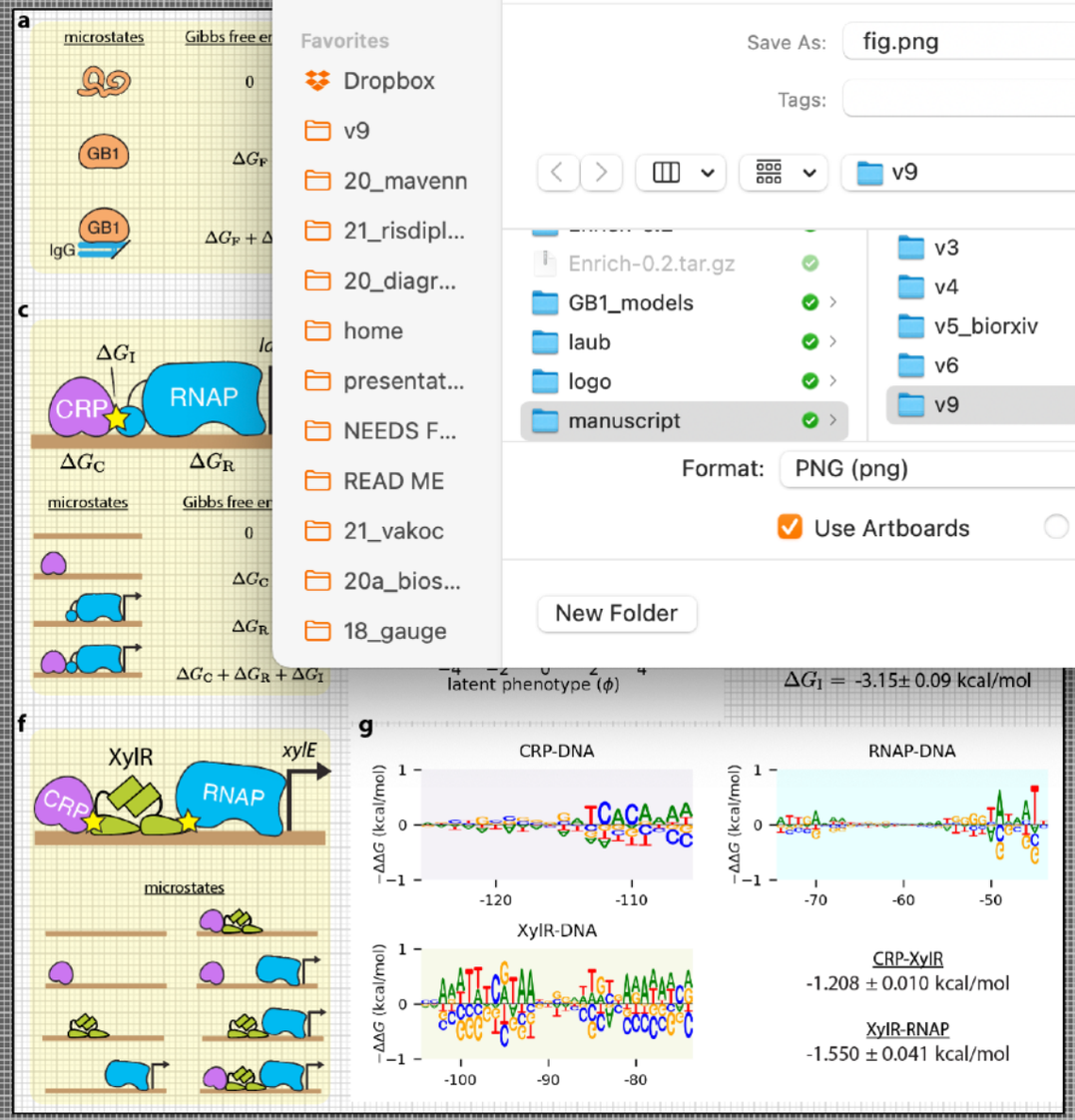


Figure 6

65.92% | 0° | 6 | Selection

fig_6_thermo.png