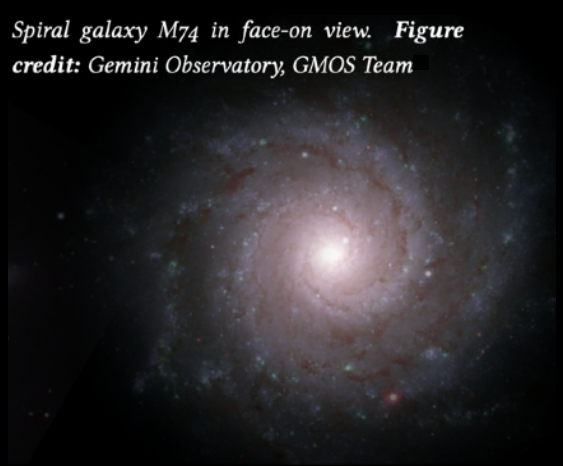
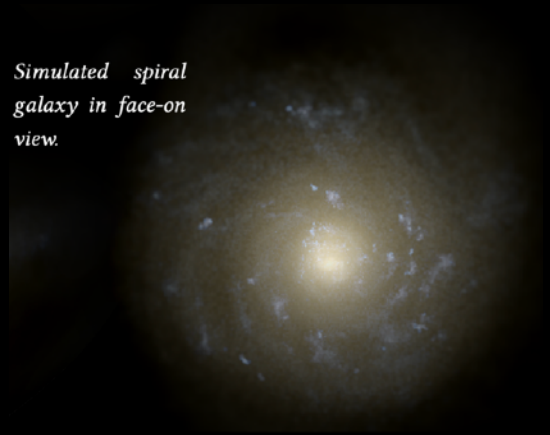
ASTR4004

COMPUTATIONAL ASTRONOMY

Week 4: https://github.com/svenbuder/astr4004_2025_week4



Spiral galaxy M74 in face-on view. ***Figure credit:*** *Gemini Observatory, GMOS Team*



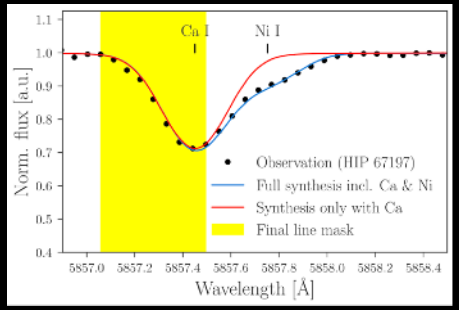
Simulated spiral galaxy in face-on view.

Figure credit: *Tobias Buck*

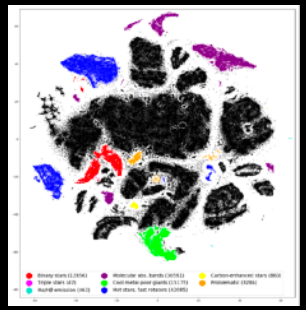
DR SVEN BUDER | RESEARCH SCHOOL OF ASTRONOMY AND ASTROPHYSICS

Australian National University

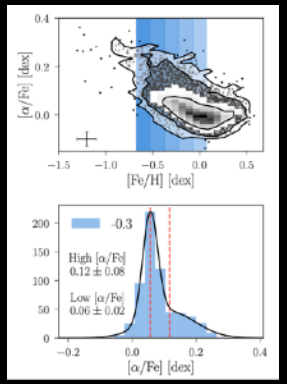
How I got into Computational Astronomy

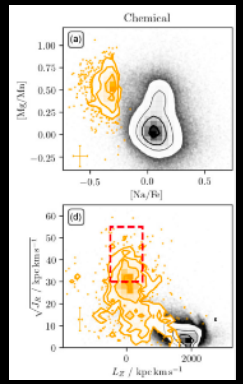


1 Mio. Spectrum analyses
Physics-& Data-driven



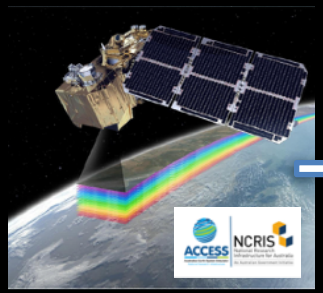
Dimensionality Reduction
(Sorting Spectra)

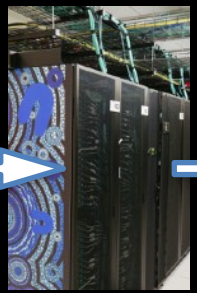


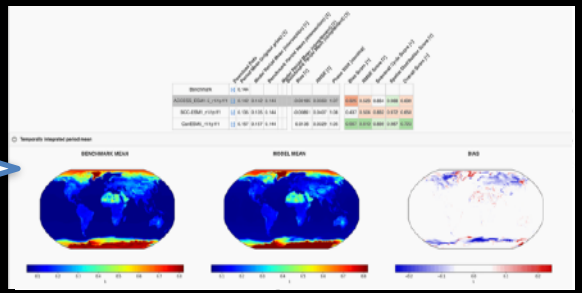


(Gaussian) Mixture Models of 2 or more components in N dimensions

Secondment to Climate Modellers: git







My idea for this week:

Week	Summary	What I actually plan to talk about
4	Data Processing	git, csv/FITS Files, ADQL/SQL, joining & cleaning catalogues, …
	Statistics	How to calculate and report uncertainties
	Plot Clinic	How to plot well & better - with my (and hopefully your) examples, …

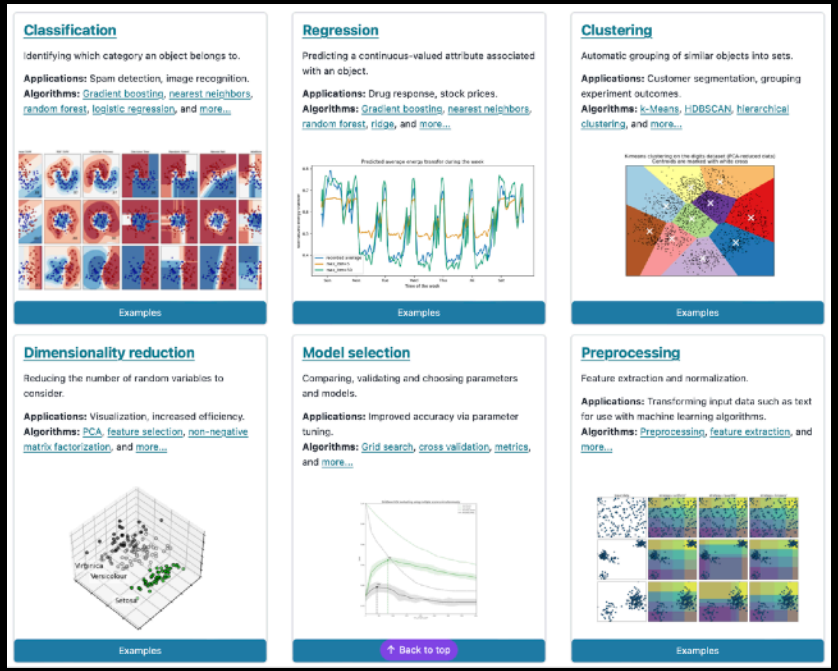
What I will also teach you later in the semester:
How to fit data *properly*,
How to reduce dimensionality of data
How to identify clustering in data
How to judge different models and confirm/reject hypotheses
How to apply what you learned outside of astronomy (other science or "industry")

Disclaimer: I can only cover the basics

There is a lot of great introductions to computational astronomy out there - including great tutorials!

I have selected what I found most useful in my personal experience.

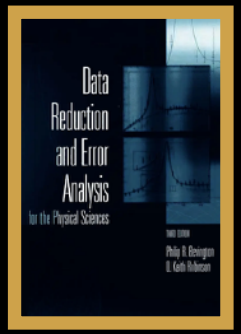
I am more than happy to adjust the course and go more into depth for specific parts if you want to.



https://scikit-learn.org/stable/index.html

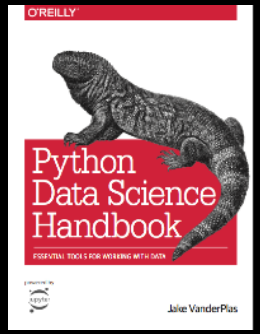
Some useful online references

Bevington & Robinson: Data Reduction and Error Analysis for Physical Sciences



Bishop: Pattern Recognition and Machine Learning

VanderPlas: Python Data Science Handbook



https://scikit-learn.org/
Preproessing, Model selection, Regression, Classification, Clustering, Dimensionality Reduction

https://www.astropy.org/

https://docs.scipy.org:
integrate, optimize, interpolate,
Fast Fourier Transforms: fftsignal
Linear algebra: linalg

GitHub tutorials such as big-data-tutorial
Chat GPT is also helpful (if you reflect on answers critically!!)

What I want you to get out of the course

A better understand of how to think critically

Something useful for your research + your career

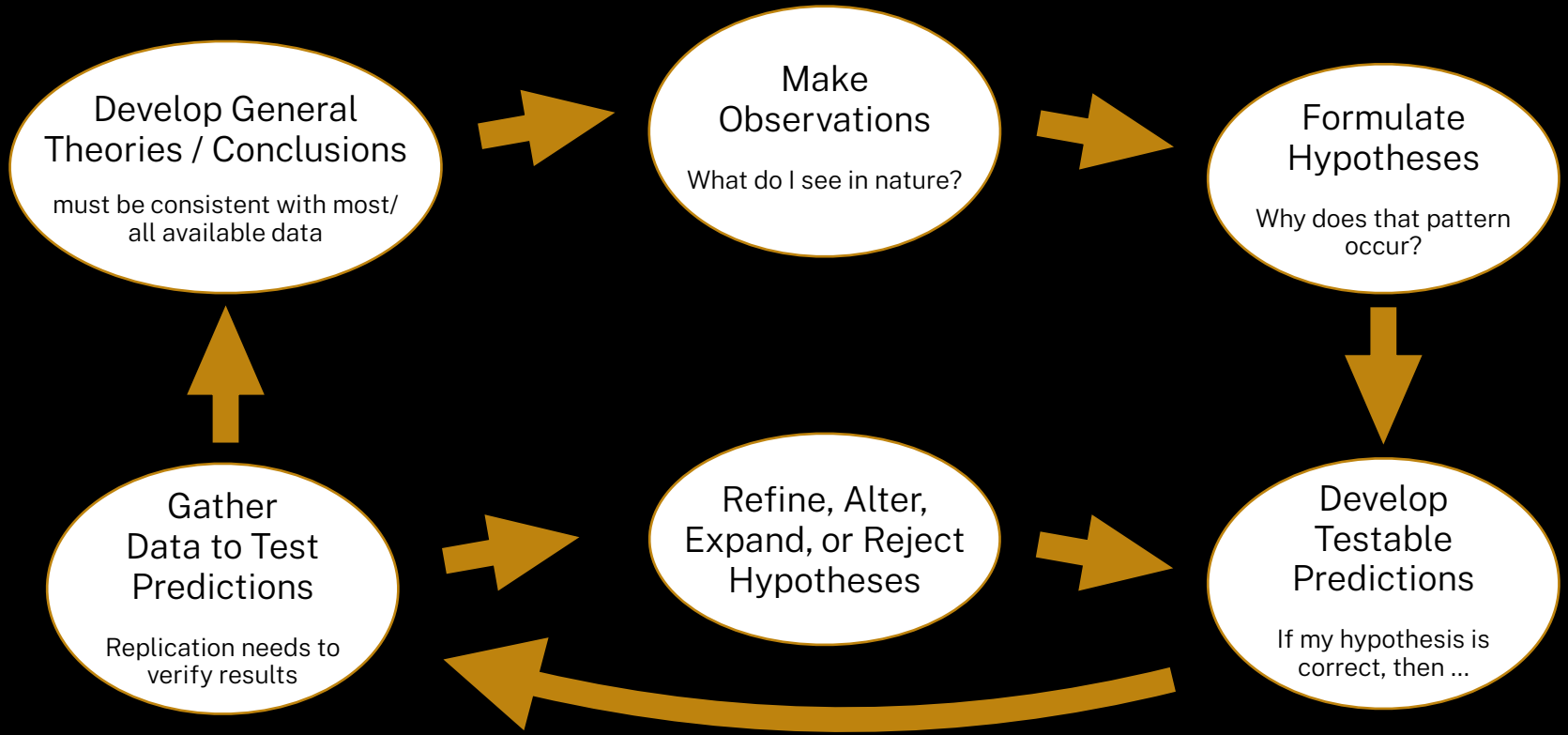
How can we use computing to enable (better) science?

A toolset for computing
in and outside of astronomy

BEFORE YOU START CODING...

Australian National University

There is a method for how to do science!



Responsible Research Practices: Three Rs

Repeatability: same results when repeated under same conditions by same team using same software using same method: consistency within original setting

Reproducibility: same results when repeated under same conditions by different team using different software with same method: sound & transparent methodology

Replicability: consistent results when repeated under different conditions in a different context: generalisability of the methodology

Rationality, (Un)Certainty, and Open Science

Claims of scientific truth can be opposed by:

- falsifying them
- questioning their certainty
- asserting the claim to be incoherent

Problems that might get in the way of doing exactly that:

- You do not publish your data (you may have your reasons) - or nobody can actually open your data files: Preprocessing
- You are not calculating or reporting your uncertainties: Statistics
- Nobody understands what you are showing or claiming: Plot Clinic
- Nobody can reproduce your results: Codesharing on GitHub

In practice: Your "marking guideline" in journals

Is the subject appropriate for the journal?	○ Yes	○ No	○ Unsure
Is the content sufficiently significant to warrant publication in MNRAS? Please comment on the significance and originality of the paper in the comments box below.	○ Yes	○ No	○ Unsure
Are the methods and results set out clearly?	○ Yes	○ No	○ Unsure
Are all the necessary and appropriate references given?	○ Yes	○ No	○ Unsure
Is the title appropriate and sufficiently informative?	○ Yes	○ No	○ Unsure
Is the abstract sufficiently informative?	○ Yes	○ No	○ Unsure
Can the paper be shortened without loss of clarity?	○ Yes	○ No	○ Unsure
Are all the figures and tables necessary and adequate?	○ Yes	○ No	○ Unsure
Is the paper in reasonable English?	○ Yes	○ No	○ Unsure
If the paper contains long tables, finding charts etc., can these be published as supplementary material?	○ Yes	○ No	○ Unsure

Referee questionnaire from MNRAS

Please address the following questions in your report to the Scientific Editor

1. *Why should this paper be published?*
2. *Are the assumptions spelled out clearly?*
3. *Are the methods fully described?*
4. *Are the new results adequately emphasized?*
5. *Are all the figures and tables necessary and properly laid out?*
6. *Which material (sections, tables, figures) should be published in electronic form only?*
7. *Is the designation of objects according to IAU rules?*

Referee questionnaire from A&A:
Bertout & Schneider (2004), A&A, 420, E1

"The best reviews are those that are thorough, identify strengths as well as weaknesses, are critical yet constructive, and are respectful to the authors."

Schultz (2022), Monthly Weather Review, 150, 6

CODESHARING

Australian
National
University

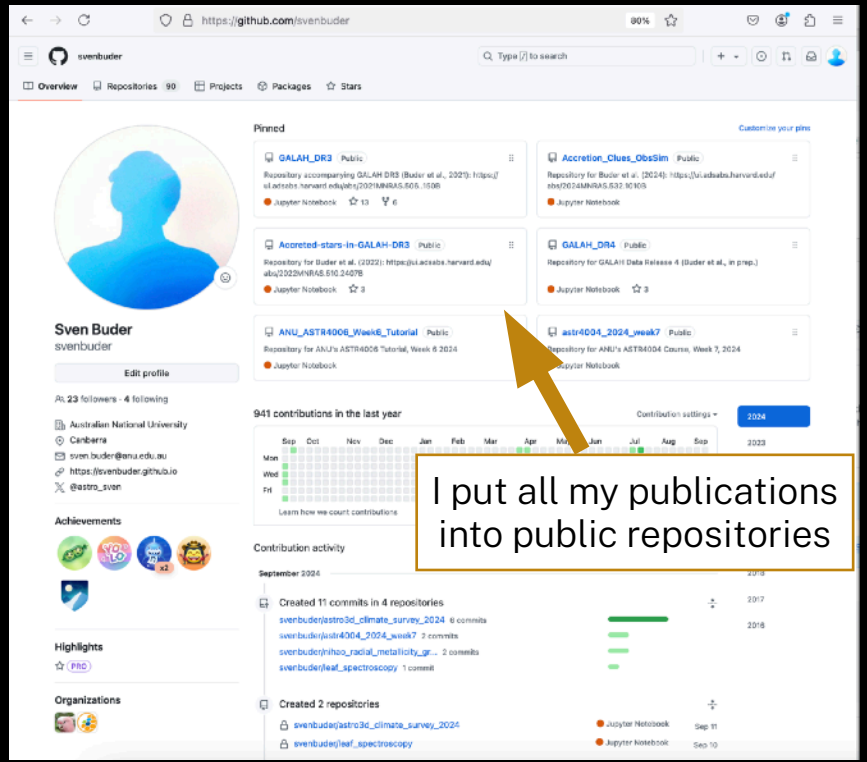
github.com: upload your code and share it

Redundancy & Flexibility:

- Large data files (>100MB) should be stored on servers
- As much of your data & code as possible should be stored in multiple places (redundancy)
- using git allows you to save multiple versions (in parallel) via branches.

A very real example: Your computer is stolen or breaks… what now?

Another reason: Easily test the same code on your Apple computer & the Linux supercomputer!



github.com: upload your code and share it

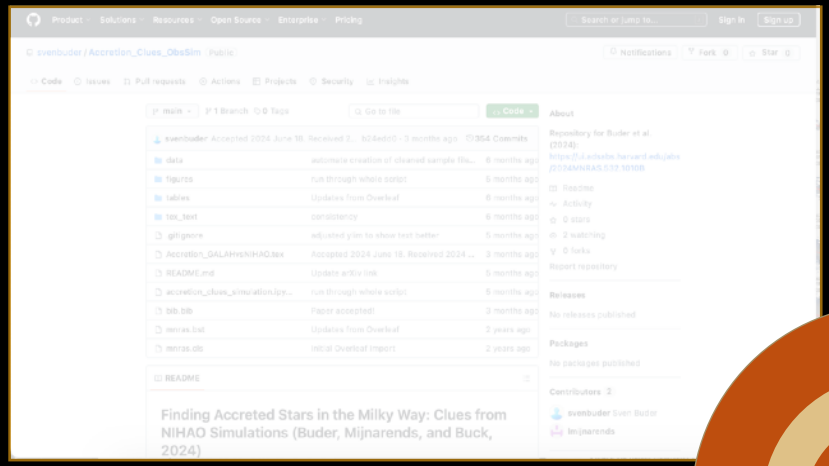
If you have not signed up to GitHub
or have never used it on any computer

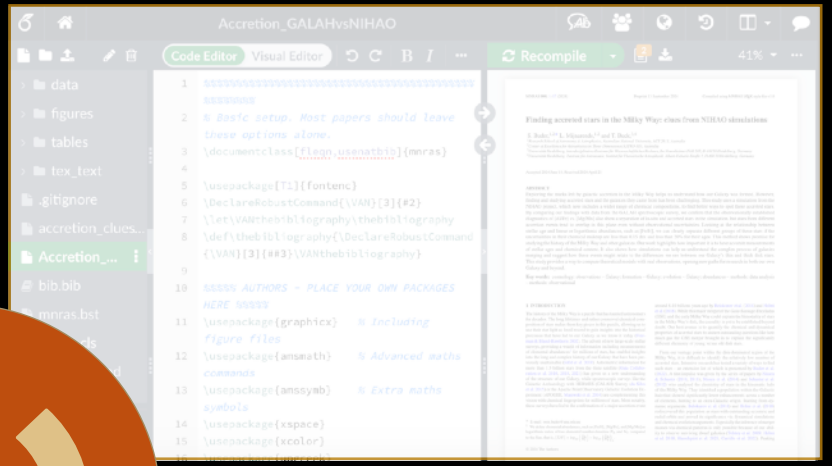
https://swcarpentry.github.io/git-novice/

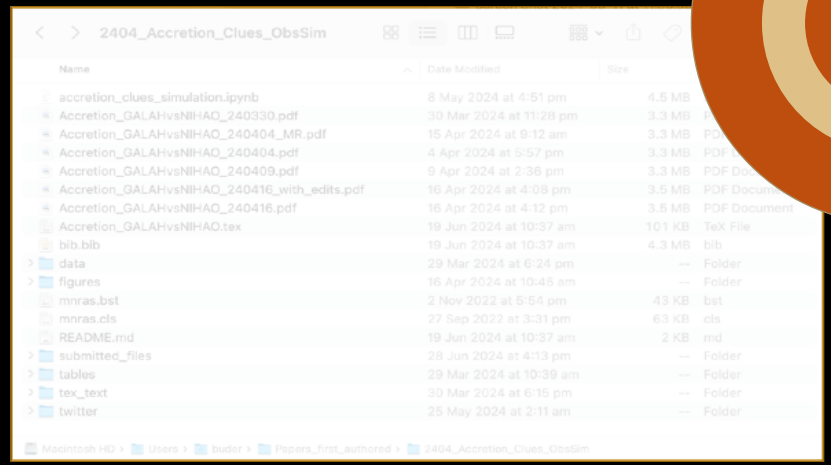
(intro to git)

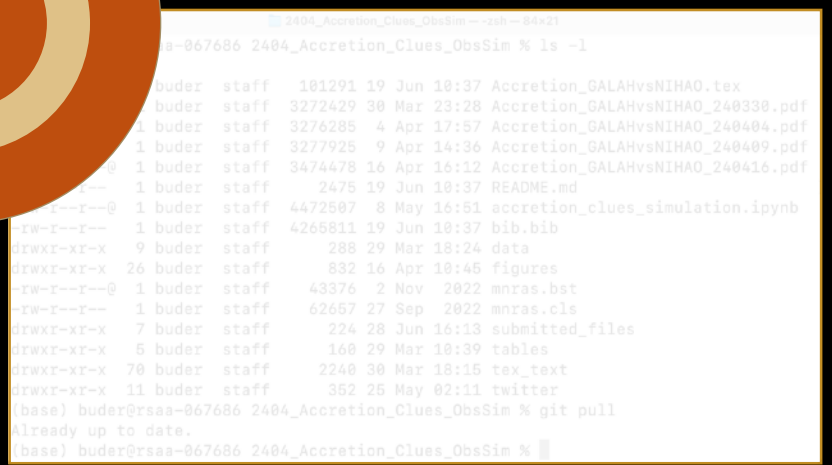
https://ivastar.github.io/advanced-git/

(advanced git developed by StSci)

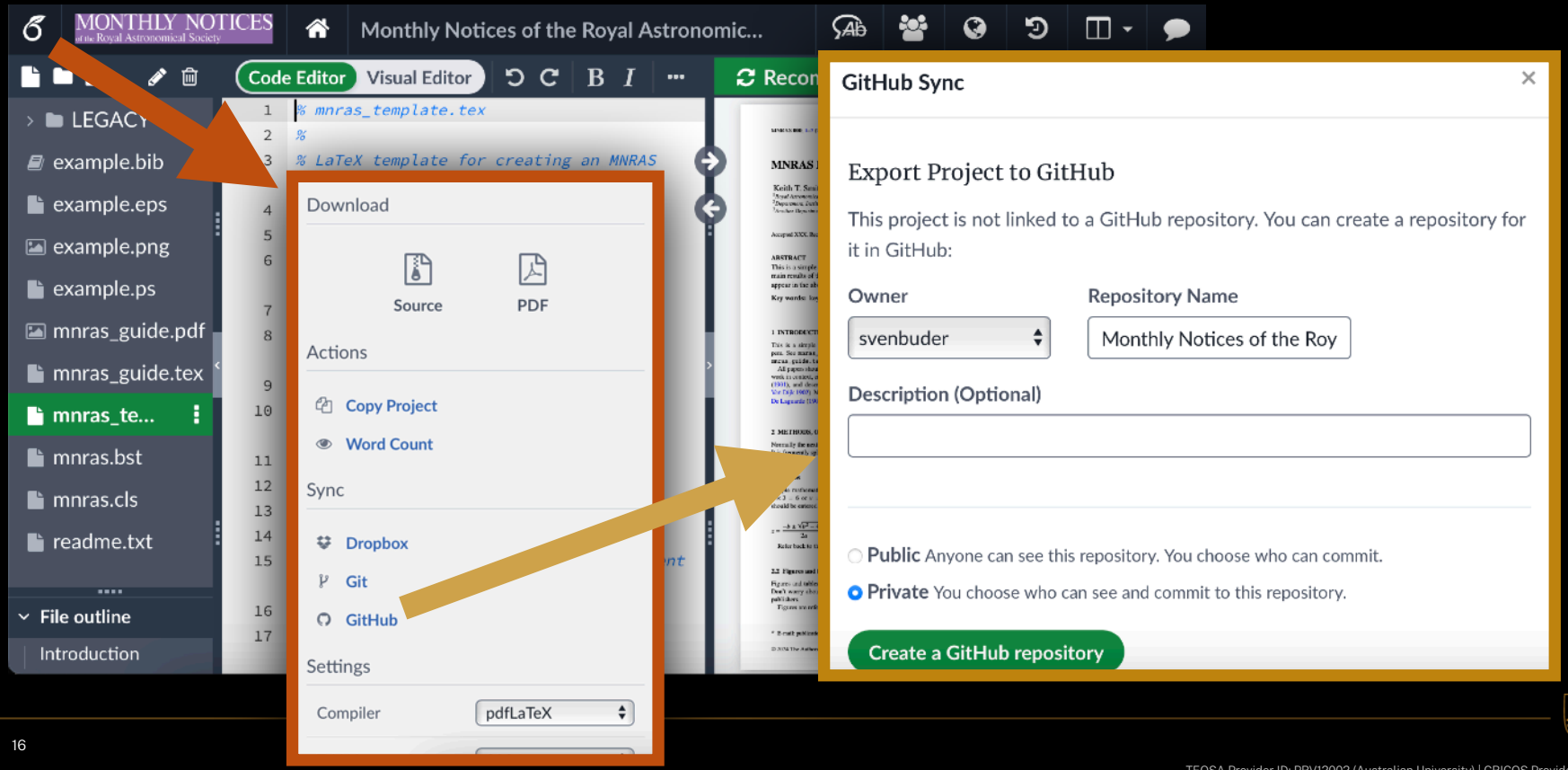








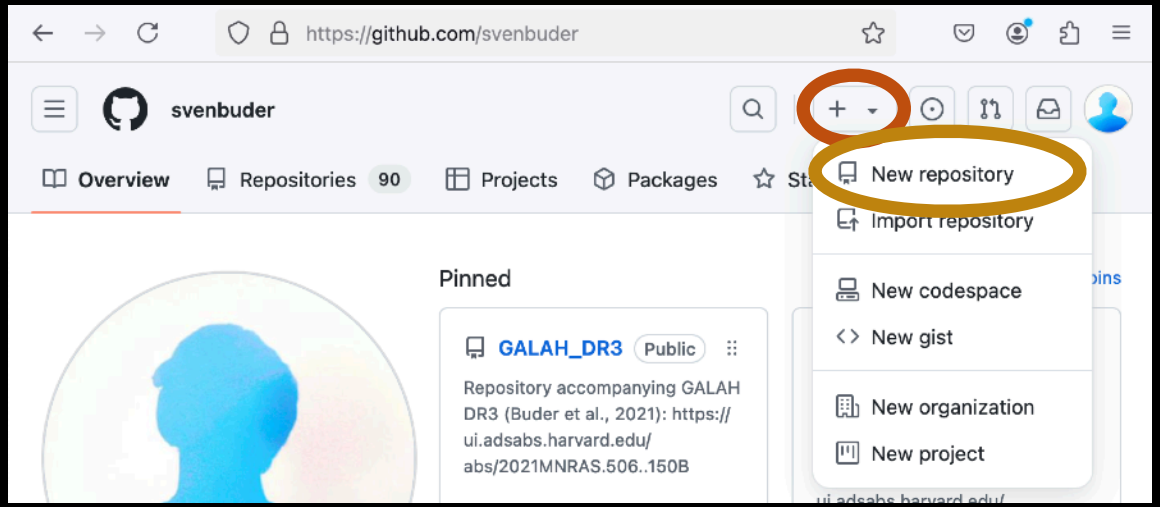
Connect git (coding) & overleaf (writing)



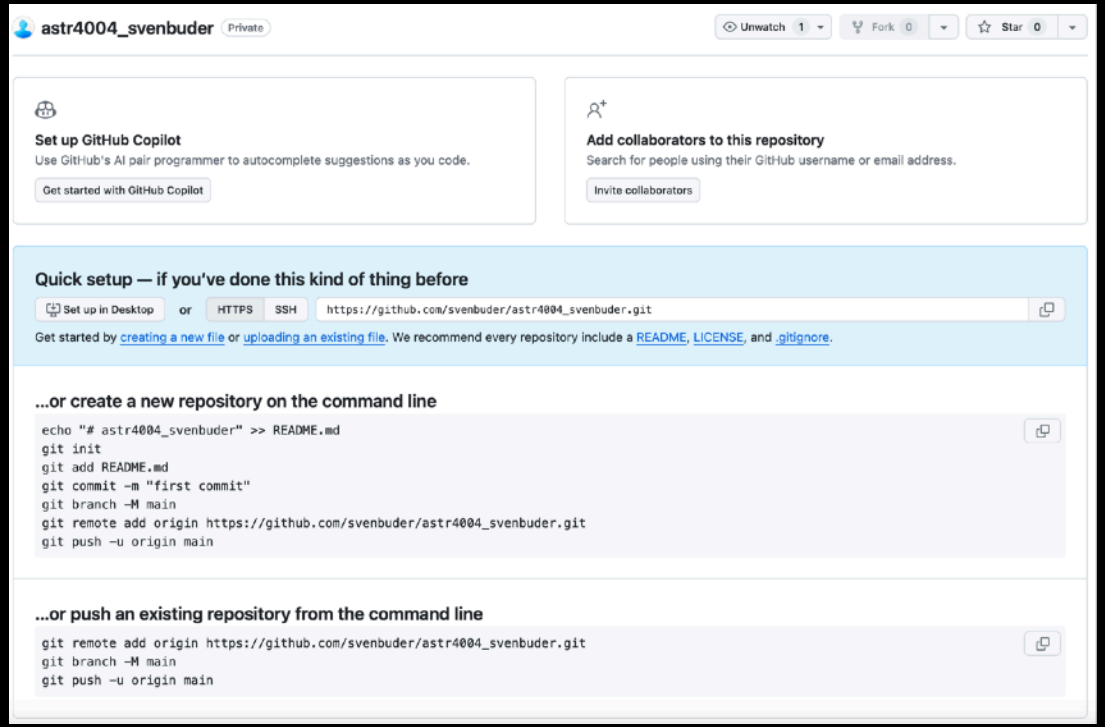
Creating a new repository

If starting a PhD: setup a journal template in overleaf once
then clone it for new projects + link it to GitHub

If just starting a new coding project: create a private GitHub Repo



Creating a new repository



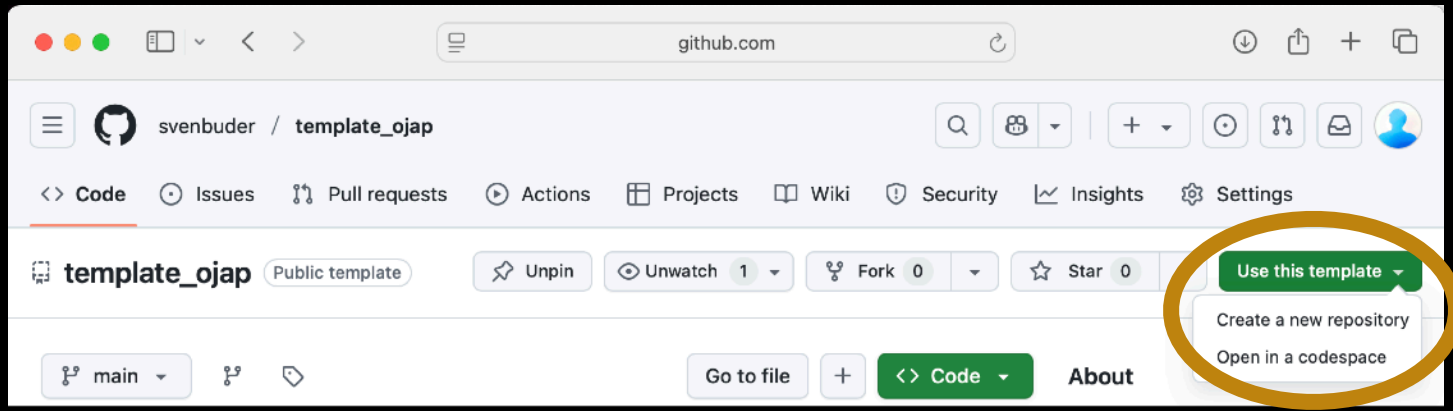
There are a lot of tutorials out there on how to set your repository up nicely.

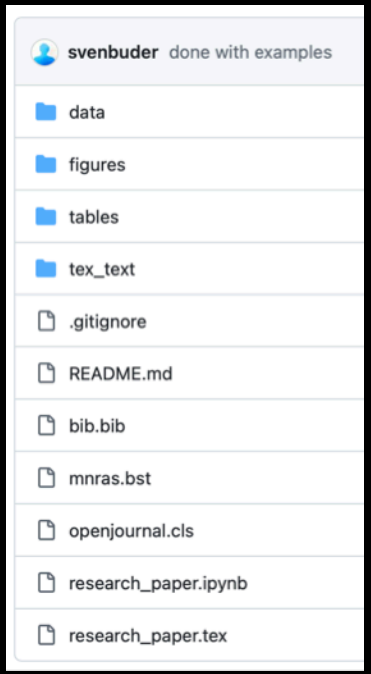
create .gitignore with everything want git to ignore:
*.file_ending
directory/
specific_file

and then use
git add .

Creating a new repository

Use someone else's template, e.g. https://github.com/svenbuder/template_ojap



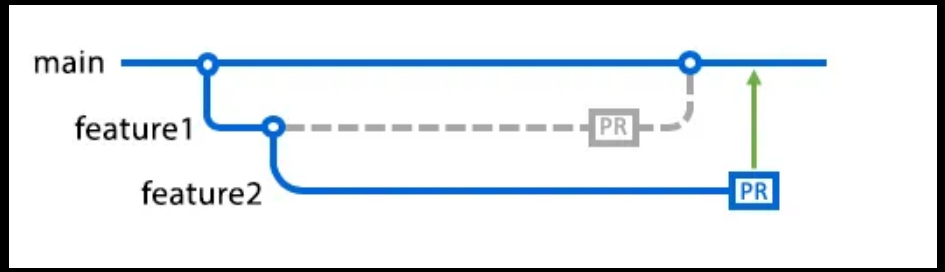


Step-by-step guide on how to set up a research paper and data structure

Creating a new repository branch

This can be extremely important for collaborative coding!

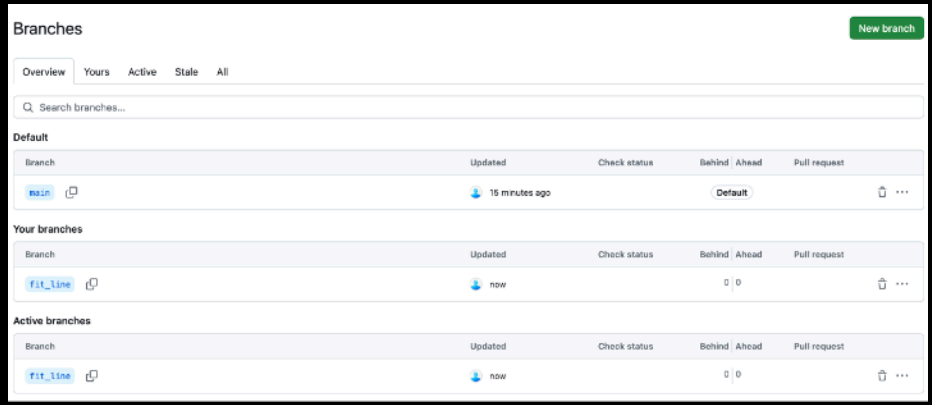
main branch:
the one that is "live" for example a website or a computer ~~game~~ program



development branch, e.g. feature1 or feature2:
the one you actively code on that can break without causing disaster

Creating a new repository branch

You can create branches online:



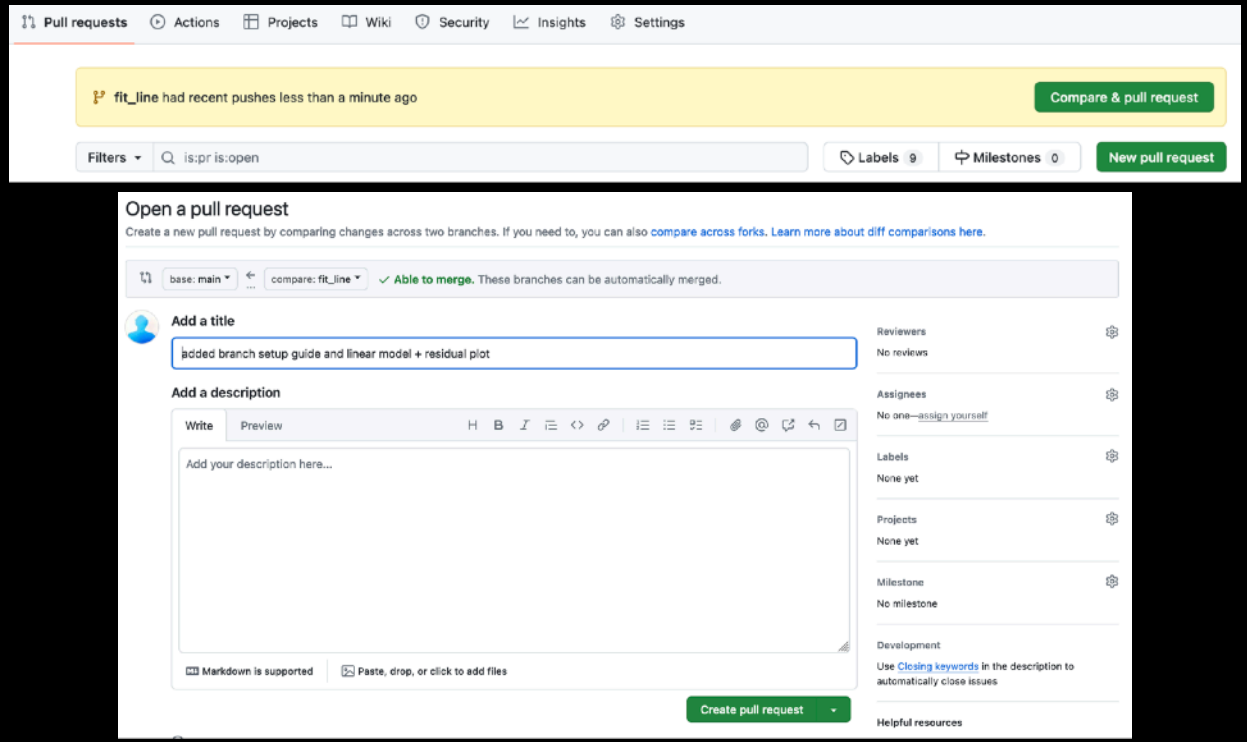
But you can also do it in bash:

```
git checkout -b fit_line
git push --set-upstream origin fit_line
git pull origin main
```

Check with branch you're on locally, see all local+remote branches, and switch to another branch:

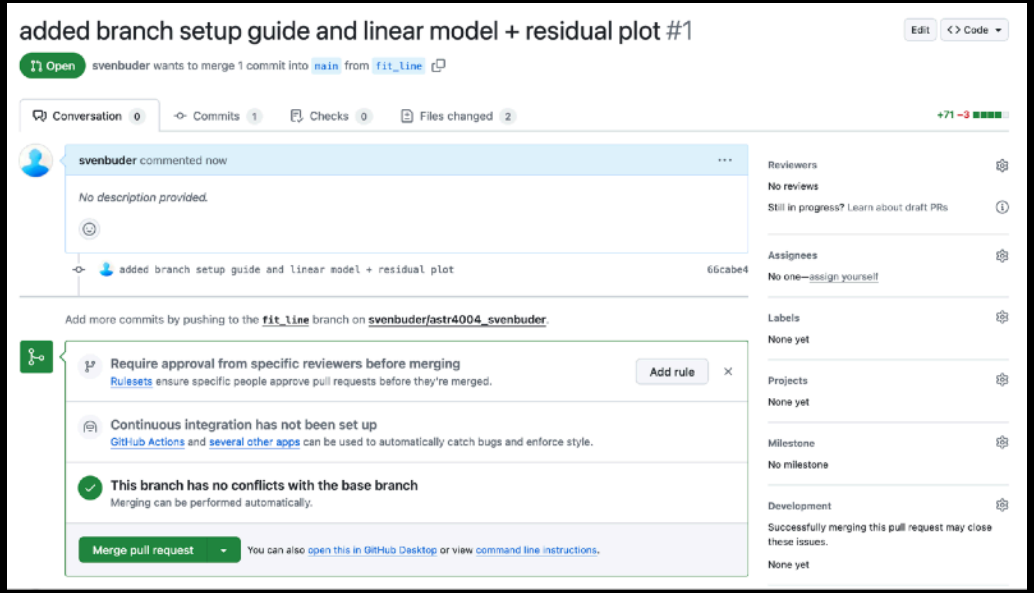
```
git status
git branch -a
git checkout my-branch-name
```

Merge into another branch



For large projects, you can add a description of what you actually changed

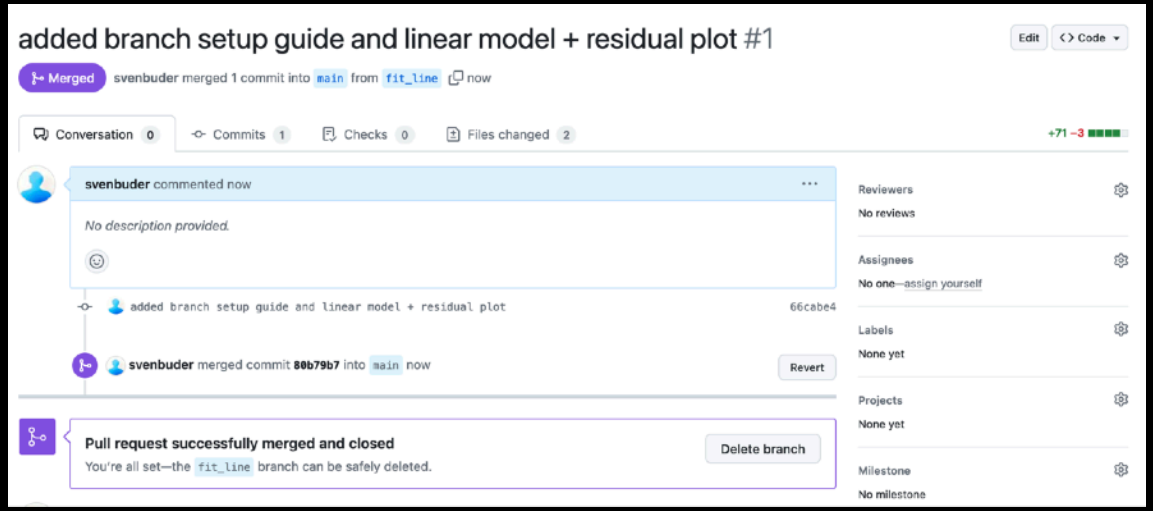
Merge into another branch



For professional coding projects, teams often work with a peer review

We will not discuss it here, but you can also include "actions"
(like github running code online if you open a merge request or push content)

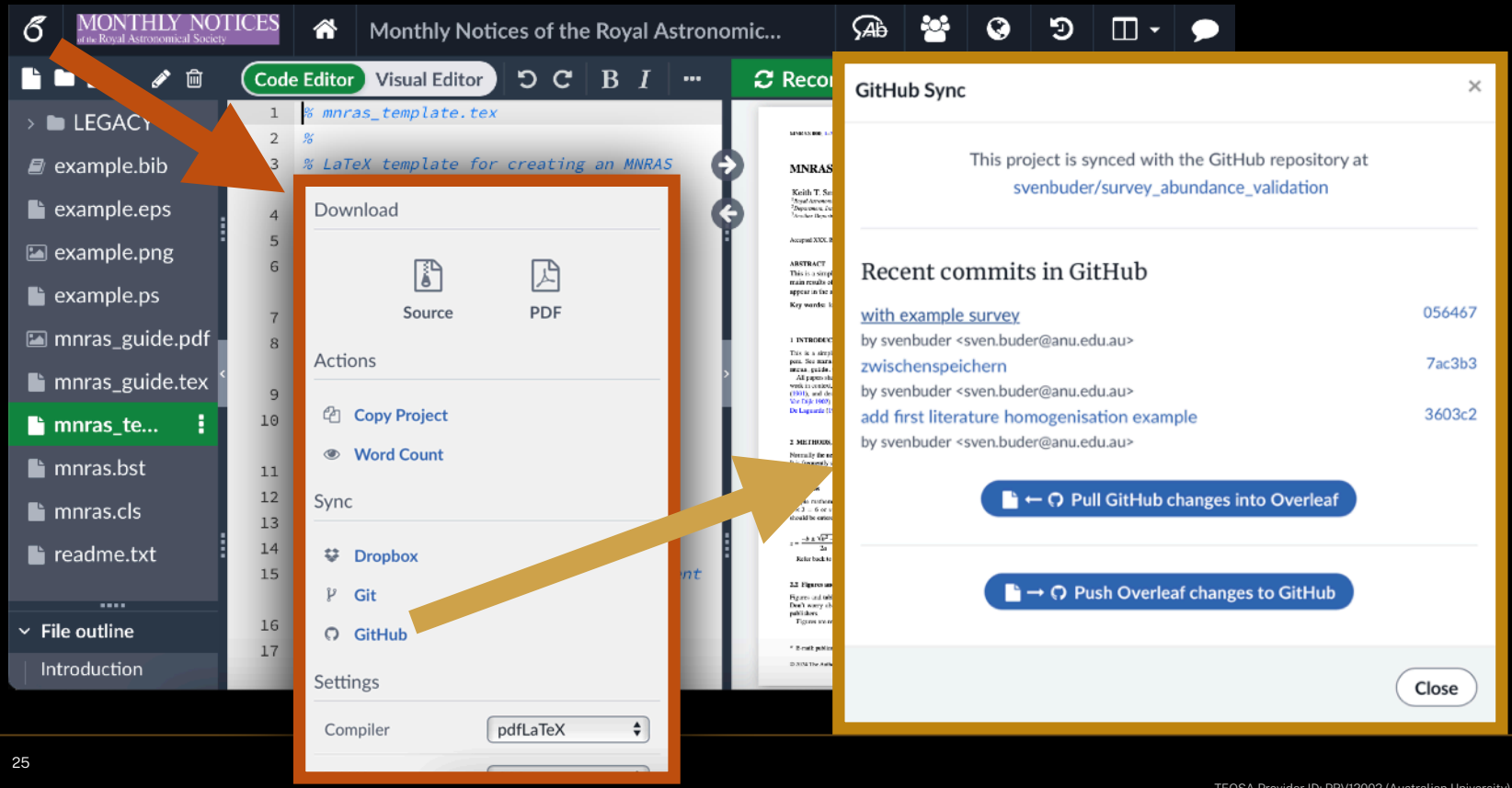
Merge into another branch



This is exactly what I did with a team of software engineers when I was working for "industry", that is, a climate modelling institute

git's version control also allows you to publish packages and releases

For science: don’t forget to synchronise git and overleaf



DATA PROCESSING

1) Writing/Reading files and catalogues:
   Excel, CSV, FITS, HDF5, Xarray
2) Working with catalogues:
   Download with ADQL, joining & cleaning catalogues
3) Interpolation

For this part, we will turn to the jupyter notebook
anu_astr4004_2025_week4a_preprocessing.ipynb

Australian National University

STATISTICS

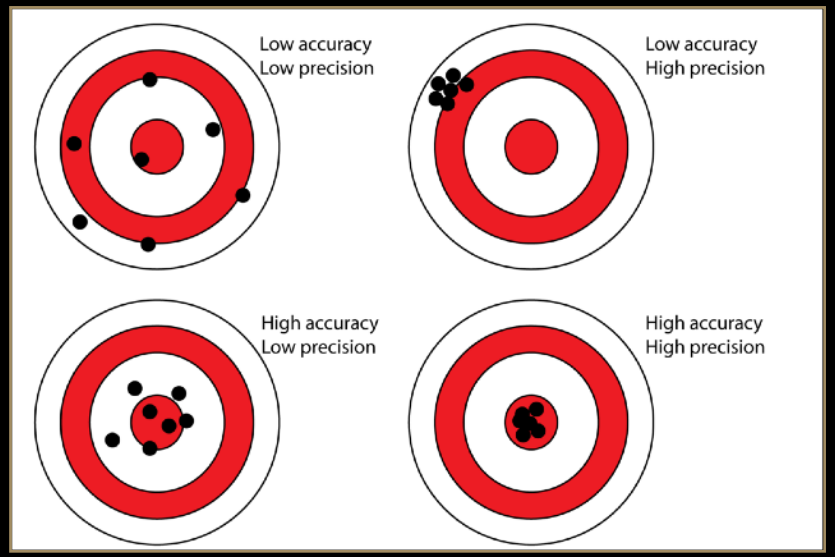
Australian
National
University

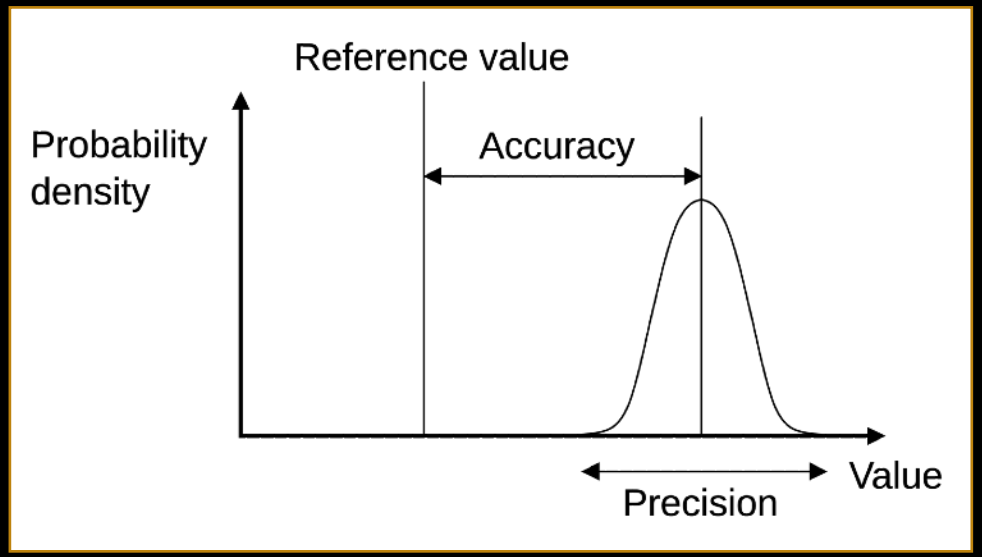
TEQSA Provider ID: PRV12002 (Australian University)
CRICOS Provider Code: 00120C

Measurement Uncertainties: Accuracy & Precision

Accuracy is a description of *systematic uncertainty*, a measure of bias.

Precision is a description of *random uncertainties*, a measure of variability.





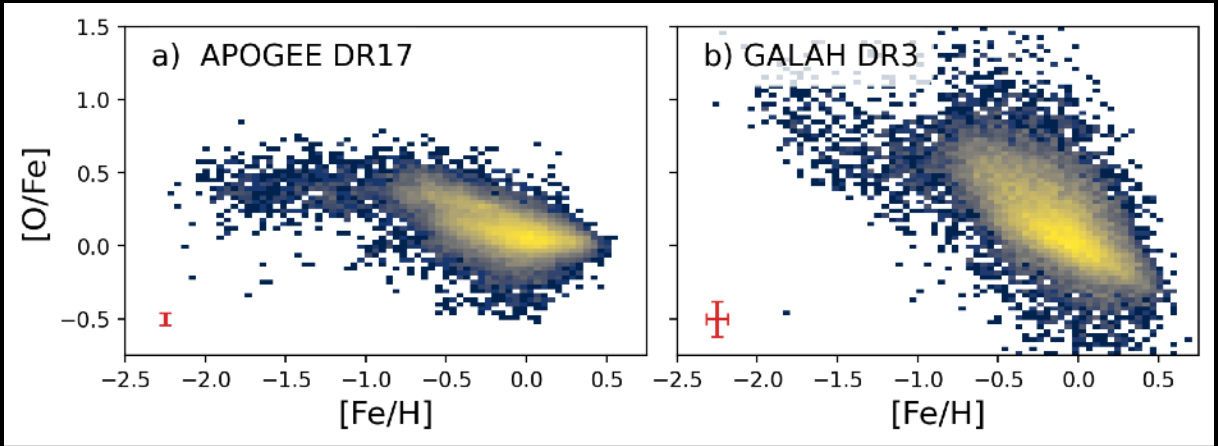
Measurement Uncertainties: Accuracy & Precision

An example: the amount of oxygen and iron in the same 26,000 stars

Error bars indicate the measurement precision (APOGEE is more precise)

But even with more precise measurements, the trends would not agree!

One of the two measurements (or both) are inaccurate!



Measurement Uncertainties: Accuracy

Estimating the accuracy of measurements can often be quite hard.
You need a reference that you can compare yourself with.

We do not "measure" oxygen, but we measure the influence from oxygen on light of stars (absorption lines in spectra) and then use models to estimate what amount of oxygen would cause such a shape.

Also note: one person’s analysis output
can be another one person’s analysis input

Measurement Uncertainties: Accuracy

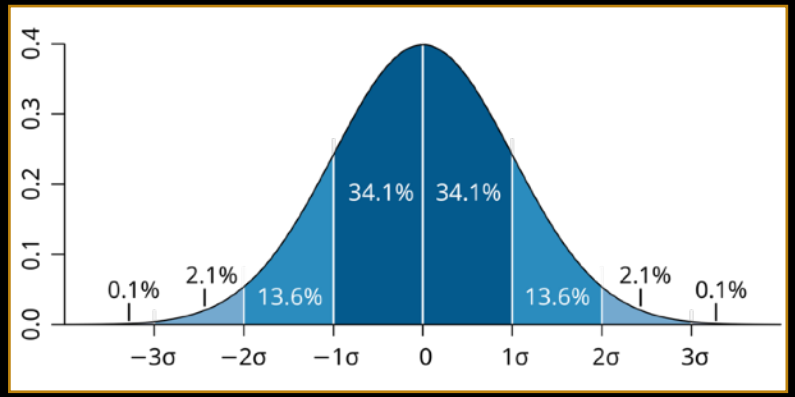
Also note: one person's analysis output
can be another one person's analysis input

Chemical evolution modellers then use these "measurements" as their input to create models that predict how oxygen is created.

To be able to make any progress, we sometimes neglect measurement inaccuracies (but need to always keep them in mind and discuss them)

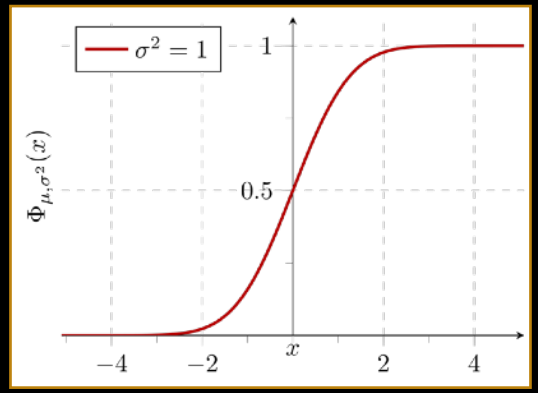
Measurement Uncertainties: Precision

For simplicity: assume we have perfect accuracy around true value 0
The only uncertainty we have is a Gaussian precision uncertainty



Distribution Function:
68.3% within ± 1 σ, 13.7% outside

2 σ : 94.4% / 5.6%; 3 σ : 99.7% / 0.3%



Cumulative
Distribution Function (CDF)

Statistics: Significance

Extraordinary claims require extraordinary evidence!

3 sigma for most measurements

5 or even 7 sigma for planet detections or cosmology

You can also turn this around with upper limits: I would have detected it, if the signal would have been *that* strong