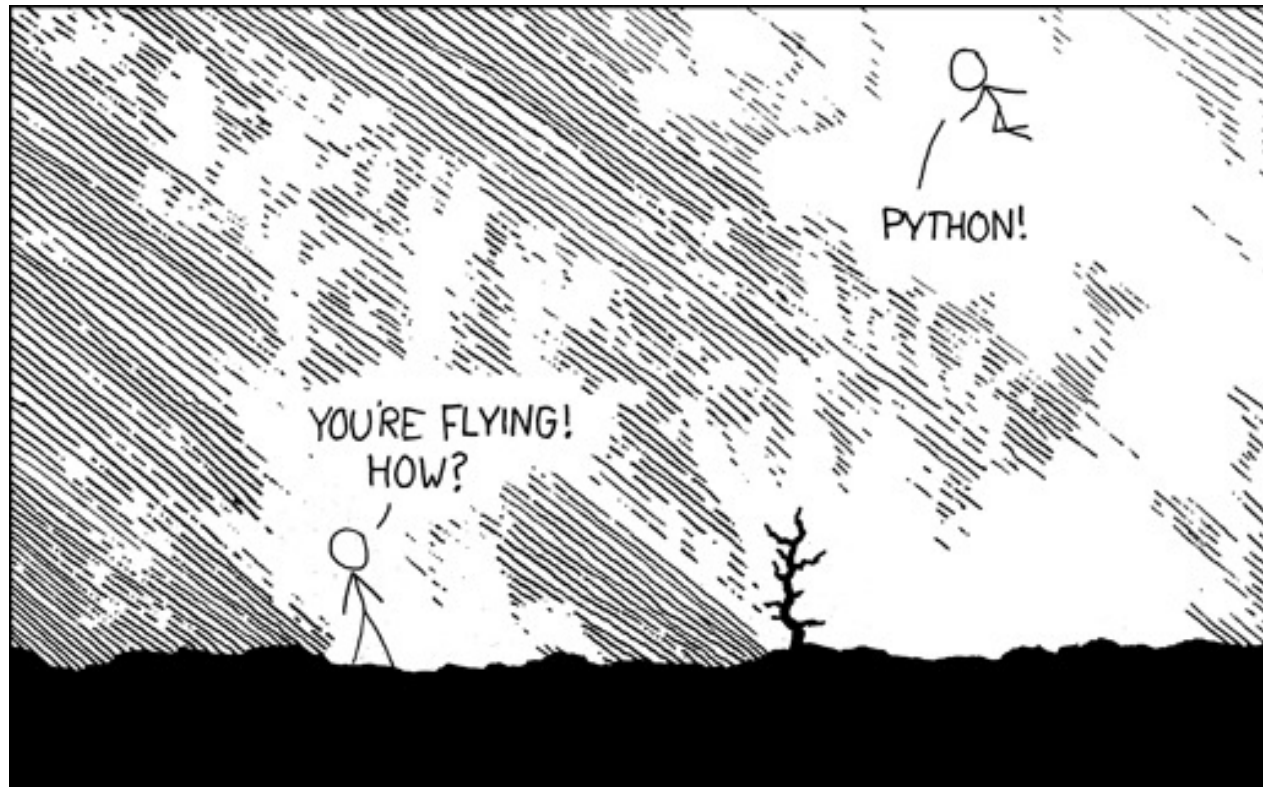# Python gives you wings!
How a little bit of code can improve the quality and efficiency of your research
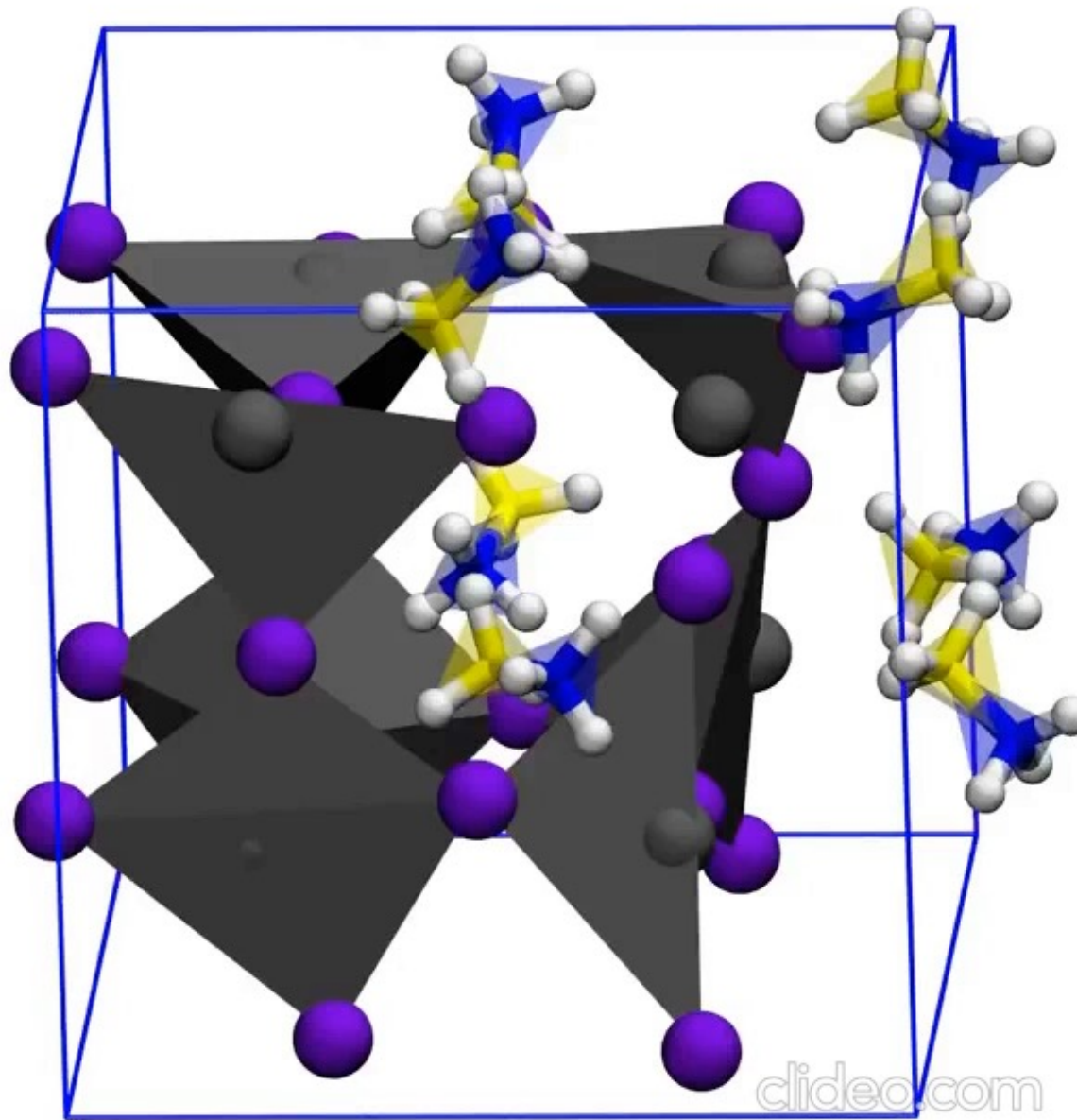

+ a little bit on High Performance Computing

Hello!! 👋

I'm Lucy (*Dr Lucy Whalley*)

- **Computational materials scientist**:
  solid state physics +
  quantum chemistry +
  high-performance-computing +
  software engineering

- **Fellow of the Software Sustainability Institute** (*better software, better research*)

- **Topic Editor for the Journal of Open Source Software**

Animation courtesy of Dr Jarvist Frost

Writing and sharing research code can transform the way you do science

It can make your research more:

💪 Robust   reduce "silly" mistakes
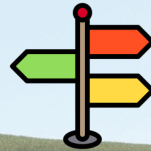
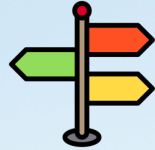🌟 Efficient   automate repetitive tasks

🌱 Reproducible   people can repeat your work

🚀 Impactful   people can use and extend your work

Learning to code is a journey

1. Research Robustness

Reduce silly mistakes

"you cannot just throw together some scripts that use excel files when you have national, significant data"
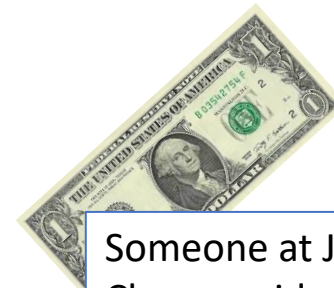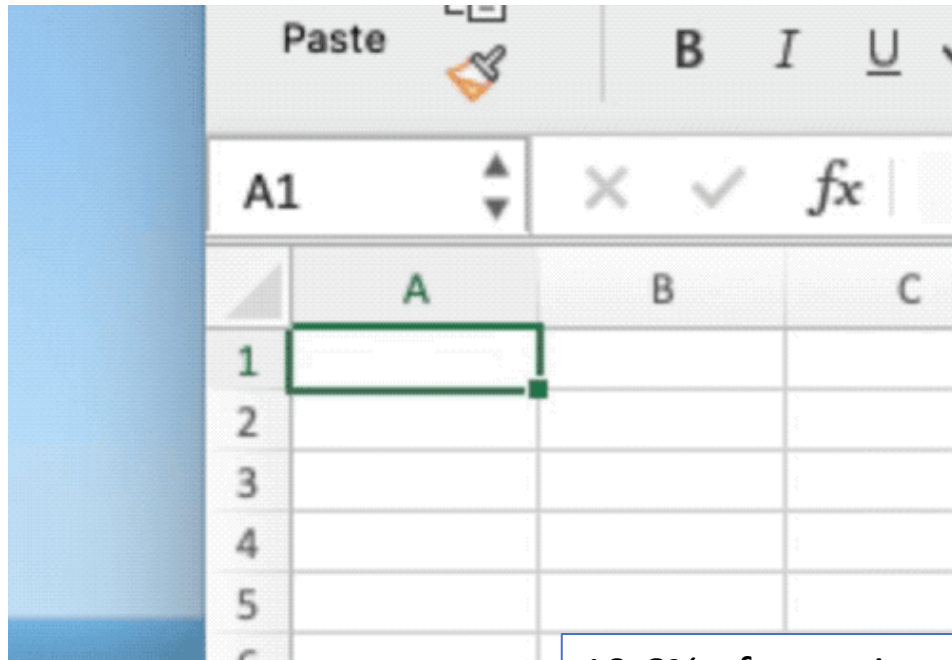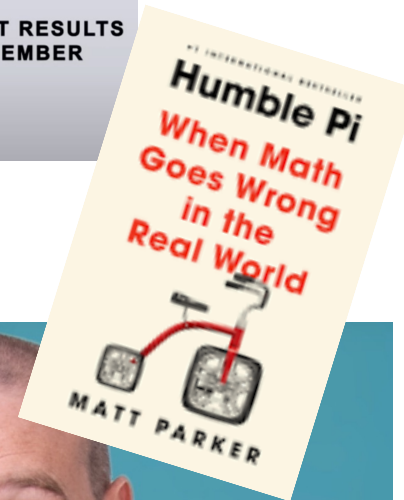
THE UK GOVERNMENT TEMPORARILY LOST

15,841

POSITIVE COVID TEST RESULTS BETWEEN 25TH SEPTEMBER AND 2ND OCTOBER

Youtube: stand up maths

Humble Pi
When Math Goes Wrong in the Real World

MATT PARKER

Someone at JPMorgan Chase accidentally added two cells in a spreadsheet instead of taking their average and lost the company $2 billion USD.

Matt Parker

From The Verge

19.6% of genetic research crunched in excel contains errors

1. Research Robustness

🛠️ Key tool: Python

72,000 entries

Average total rated capacity of an installation in 2018

US Wind turbine database

```
[1] import pandas as pd

[2] df = pd.read_csv('./uswtdb_v5_3_20230113.csv')
[3] df_2018 = df[df['p_year'] == 2018.0]
[4] df_2018['p_cap'].mean()
```

Import Python libraries

Read in data

Select 2018 data

Calculate mean

198.6597295514512

Total rated capacity in MW

2. Research Efficiency

Automating common tasks

72,000 entries

US Wind turbine database

2,647 entries

UK Wind turbine database

25,200 entries

European Wind turbine database

```python
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('./uswtdb_v5_3_20230113.csv')

mean_cap = []
for year in pd.unique(df['p_year'].values):
    df_year = df[df['p_year'] == year]
    mean_cap.append(df_year['p_cap'].mean())

plt.bar(pd.unique(df['p_year'].values),mean_cap)
```
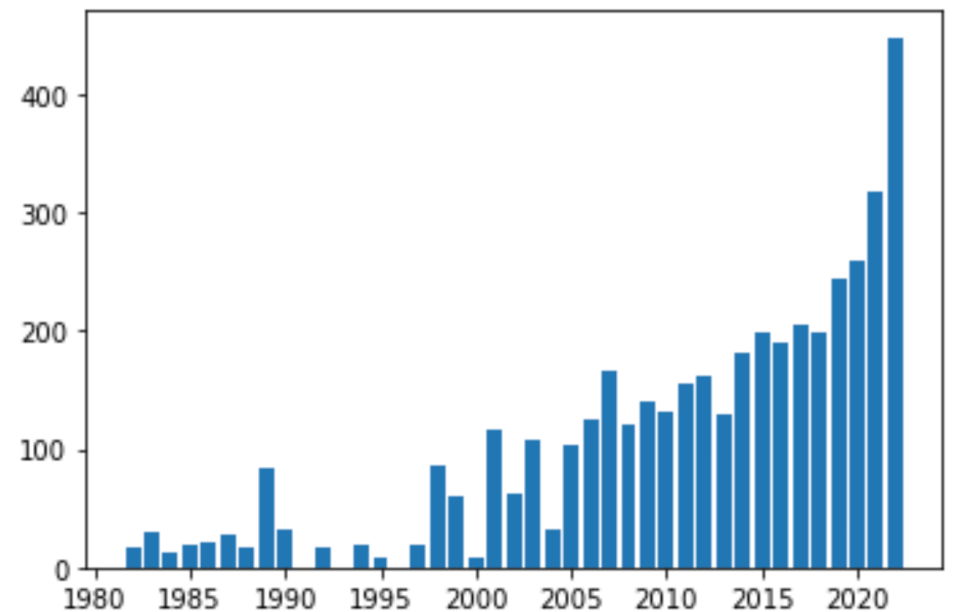
Plot the average capacity of an installation by year

2. Research Efficiency

🛠️ Key tool: Community

Our Core Lessons in English

Lesson	Site	Repository	Reference	Instructor Notes
The Unix Shell	🗔	🐱	👁	⊕
Version Control with Git	🗔	🐱	👁	⊕
Programming with Python	🗔	🐱	👁	⊕
Plotting and Programming in Python	🗔	🐱	👁	⊕
Programming with R	🗔	🐱	👁	⊕
R for Reproducible Scientific Analysis	🗔	🐱	👁	⊕

Upcoming Workshops

Click on an individual event to learn more about that event, including contact information and registration instructions.

🔵 🇬🇧 **University of Oxford, School of Geography and the Environment** Mar 14 - Mar 15, 2023
Instructors: Tom Russell, Fred Thomas
Helpers: Sebastian Engelstaedter, Samira Barzin

🔵 🇺🇸🌐 **Michigan Technological University** Mar 20 - Mar 24, 2023
Instructors: Pradeep Eranti, Christian Ayala

🔵 🇩🇪 **University of Stuttgart, Germany** Mar 28 - Mar 31, 2023
Instructors: Matthias Braun, Björn Schembera, Jan Range
Helpers: Anett Seeland, Dorothea Iglezakis

🔵 🇩🇪 **Max Planck Institute for Evolutionary Biology** Apr 4 - Apr 6, 2023
Instructors: Carsten Fortmann-Grote, Nikoleta Glynatsi
Helpers: Beate Gericke

🔵 🇫🇮 **Skidaway Institute of Oceanography** May 19 - May 20, 2023
Instructors: Lucy Quirk, Claire Zwiers Cook
Helpers: Mallory Mintz

2. Research Efficiency

🛠️ Key tool: Community

Code mentorship programme
Deadline: 20th March

Research Software Camp: FAIR software

The next Research Software Camp will be held from 19th to 30th June 2023. Over the course of two weeks, we will host a panel discussion, basic skills workshops (R, Python, visualisation, etc.) and a mentorship programme.

Research Software Camp: Supporting Mental Health

Our autumn Research Software Camp will address approaches to wellbeing and mental health in research software from 7 to 18 November 2022.

Research Software Camp: Next steps in coding

Our Research Software Camp: Next steps in coding ran from from 16th to 27th May 2022. It focused on improving computational and training skills and exploring existing resources out there.

Research Software Camp: Beyond the Spreadsheet

The Research Software Camp: Beyond the Spreadsheet ran from 1st to 12th November 2021 and focussed on the uses of spreadsheets in research and first steps into further use of software in research.

Research Software Camp: Research Accessibility

The first Research Software Camp on Research Accessibility took place from 22 February to 5 March 2021. The topic covered how to make research tools, techniques, datasets, software, protocols, etc. accessible and how to get new or more collaborators.

3. Research Reproducibility

Others can repeat your work

Reproducibility

"Reproducibility means that when you repeat an experiment or investigate something again, the result should be the same. Reproducibility is a cornerstone of the scientific process."
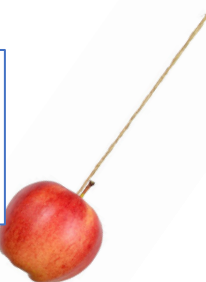
Gustav Nilsonne

Is science r**the lid on reproducibility**eproducibility crisis,

Is science in trouble? An insight into the reproducibility crisis

Yan-Yi Lee talks about the reproducibility crisis as well as the recent collective efforts that scientists have shown to address it.
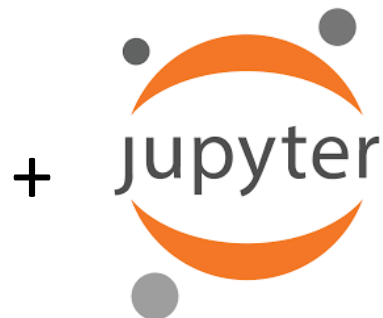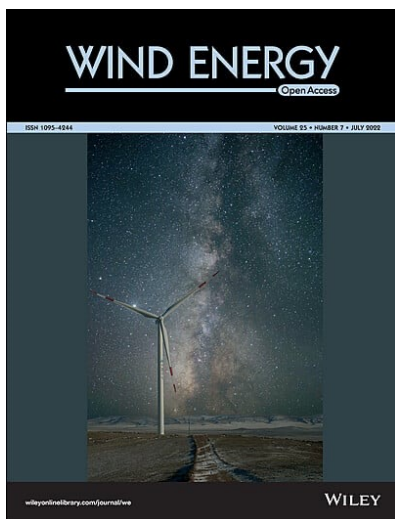
1,500

Monya b

Relatively low hanging fruit: *share your code and data*

3. Research Reproducibility
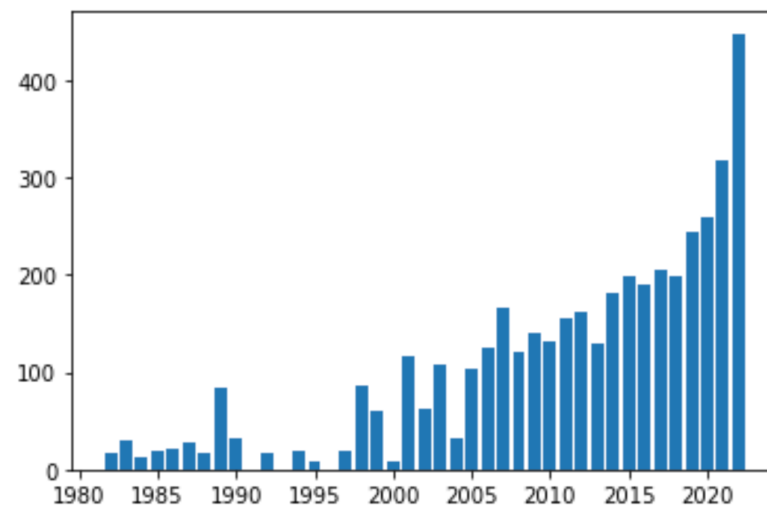
🛠️ Key tool:

Jupyter Notebook

+

---

jupyter **Windfarm_analysis** Last Checkpoint: an hour ago   (unsaved changes)

File   Edit   View   Insert   Cell   Kernel   Widgets   Help

Analysis of average installation capacity

```
In [26]:  1  import pandas as pd
          2  import matplotlib.pyplot as plt
          3
          4  df = pd.read_csv('./uswtdb_v5_3_20230113.csv')
          5
          6  mean_cap = []
          7  for year in pd.unique(df['p_year'].values):
          8      df_year = df[df['p_year'] == year]
          9      mean_cap.append(df_year['p_cap'].mean())
         10
         11  plt.bar(pd.unique(df['p_year'].values),mean_cap)
         12
```
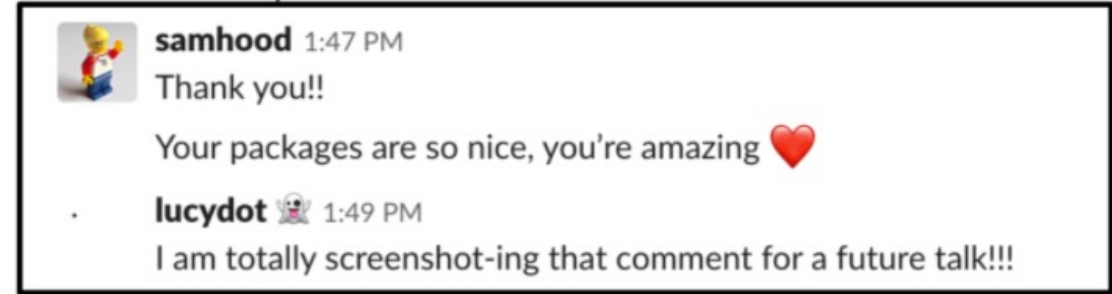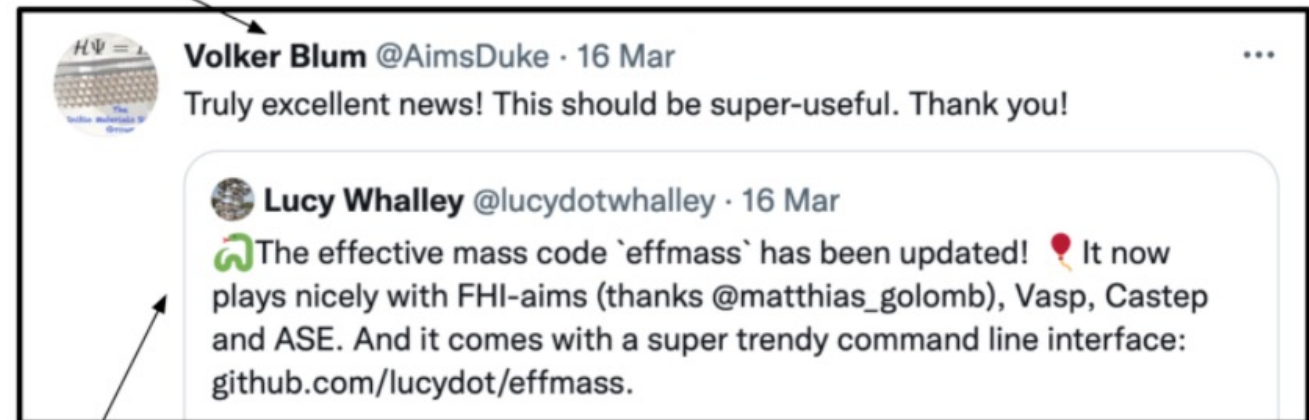
Out[26]: <BarContainer object of 41 artists>

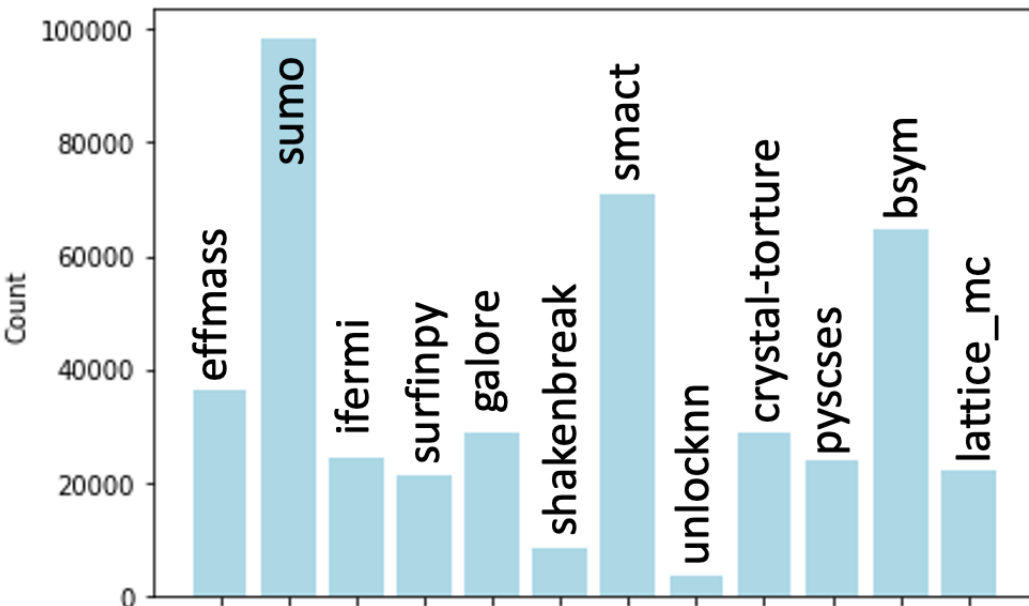# 4. Research Impact

Others can use and extend your work

4. Research Impact
🛠️ Key tool: Github / Gitlab

github.com/NU-CEM/ThermoPot

Digital Object Identifier

Run notebooks without download

Publish your code in a journal

☰ README.md   ✏️

ThermoPot: An ab-initio thermodynamic modelling package `DOI 10.5281/zenodo.7351356`

⚠️⚠️⚠️ This repository is still in development, and some functionality is untested. Use with caution, or wait until we fully test it! ⚠️⚠️⚠️

`build-docs passing` `run-tests passing` `lint-code passing`

Software for ab-initio thermodynamic modelling of material formation and decomposition.

- 📚 The documentation is here.
- 🔄 This code is made available under the GNU General Public Licence (GPL) v3. See the LICENSE file for the full text.
- 🖊️ If you use this package for your research please cite accordingly.

Please note

This work adapts and extends a previous repository developed by Adam Jackson: Thermodynamic model of CZTS.
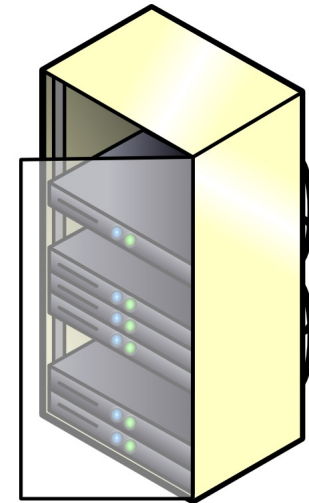
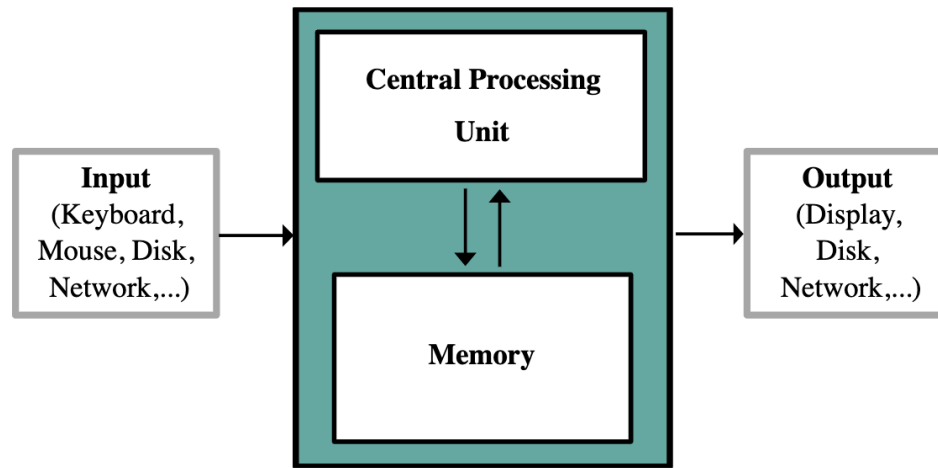# There is a whole world out there

High Performance Computing

A standard laptop for a standard task

input → computation → output

When tasks take too long.....

When one server is not enough....

Input
(Keyboard, Mouse, Disk, Network,...)

Central Processing Unit

Memory

Output
(Display, Disk, Network,...)

Server

cluster / supercomputer

https://rse.shef.ac.uk/hpc-intro-tuos-citc/

High Performance Computing

https://www.archer2.ac.uk/training/

Oswald User Guide

Cores / nodes / memory

SSH keys

Terminal / Bash

Internet
or
network

Oswald

Laptop

Cluster

https://rse.shef.ac.uk/hpc-intro-tuos-citc/

🌟 # Thank You

Code mentorship programme
Deadline: 20th March

Questions / discussion

- Do you write code for your research?

- Are there particular research tasks you could automate?

- Do you feel that you *just can't code*?

- Would you like to attend coding-related events at Northumbria?

- What code-related barriers do you face?

Lucy Whalley - @lucydot (github) – lucydot.github.io – l.whalley@northumbria.ac.uk