The Alan Turing Institute

The Turing Way

A handbook for reproducible research

Kirstie Whitaker

Collaborations Workshop Demo, 2 April 2019

Slides at https://doi.org/10.5281/zenodo.2621280

Neurohackweek 2016

Photo credit: Chris Gorgolewski

Founding the Institute

"We will found The Alan Turing Institute to ensure Britain leads the way again in the use of big data and algorithm research"

George Osborne, Chancellor of the Exchequer
Budget Speech, March 2014

The Alan Turing Institute

EPSRC
Engineering and Physical Sciences Research Council

Network of industry, charity, government partners

Network of university members

Strategic government investment

The Institute's partners and collaborators

Our university network

UNIVERSITY OF CAMBRIDGE

THE UNIVERSITY of EDINBURGH

UNIVERSITY OF OXFORD

UCL

WARWICK

University of BRISTOL

UNIVERSITY OF BIRMINGHAM

UNIVERSITY OF EXETER

UNIVERSITY OF LEEDS

MANCHESTER 1824
The University of Manchester

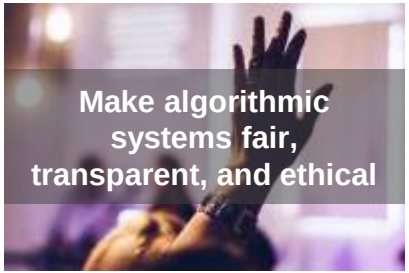
Queen Mary University of London

Newcastle University
UK | Malaysia | Singapore

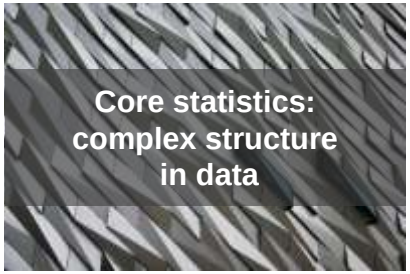
UNIVERSITY OF Southampton

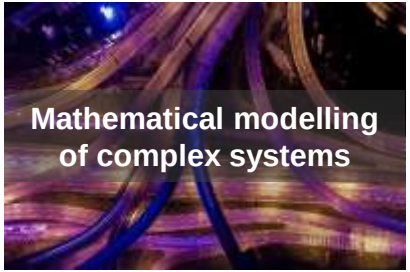
Challenges

Advance data science and artificial intelligence to…



Core capabilities





https://www.turing.ac.uk/news/alan-turing-institute-spearhead-new-cutting-edge-data-science-and-artificial-intelligence

Urban analytics →
Developing data science and AI focused on the process, structure, interactions and evolution of agents, technology and infrastructure within and between cities.

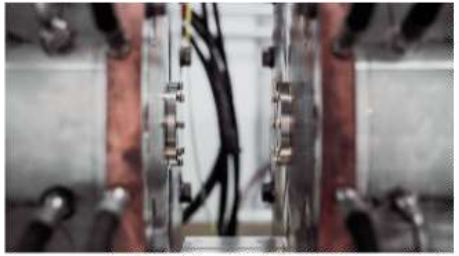
Data-centric engineering →
Bringing together world-leading academic institutions and major industrial partners from across the engineering sector, to address new challenges in data-centric engineering.

Data science for science →
Ensuring that research across science and the humanities can make effective use of state of the art methods in artificial intelligence and data science.

Cross cutting theme: Tools, systems and practices

Health →
Accelerating the scientific understanding of human disease and improving human health through data-driven innovation in AI and statistical science.

Public policy →
Working with policy makers on data-driven public services and innovation to solve policy problems, and developing ethical foundations for data science and AI policy-making.



Research Engineering →
Connecting research to applications, helping create usable and sustainable tools, practices and systems.

https://www.turing.ac.uk/research/ai-science-engineering-health-and-government

Fraud is not our biggest problem

https://doi.org/10.6084/m9.figshare.5558653

Errors in the literature have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

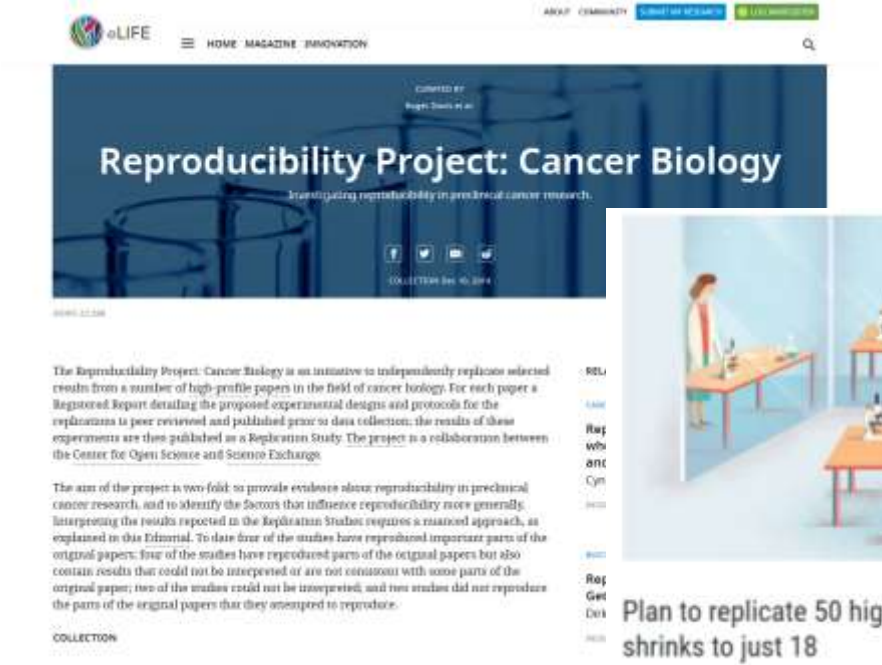
https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there
https://www.bbc.co.uk/news/magazine-22223190

Errors in the literature have real world effects

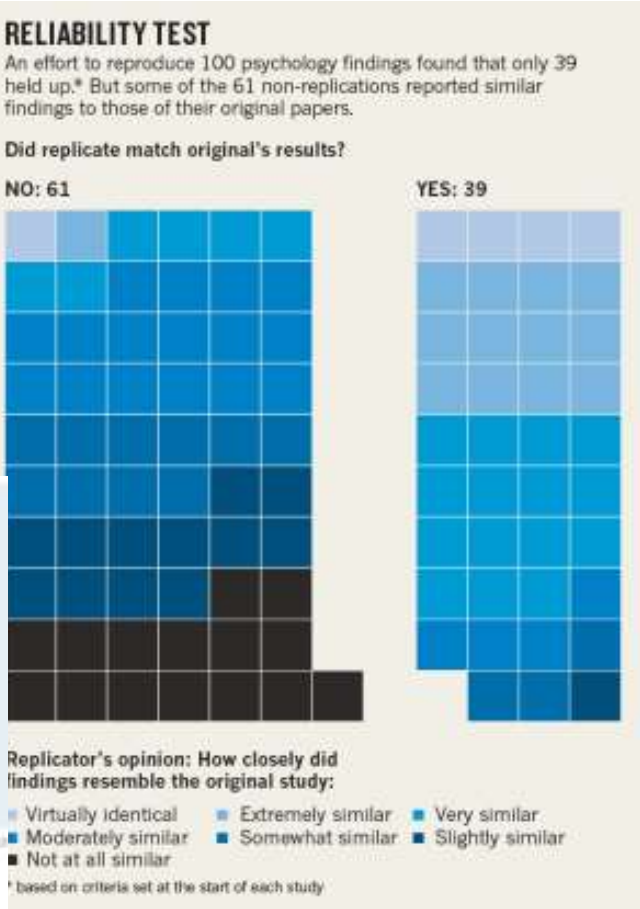
	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.7	3.0	3.5	1.7	5.5
27	Minimum		1.6	0.3	1.3	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.6	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.5	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	=AVERAGE(L30:L44)	

https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there
https://www.bbc.co.uk/news/magazine-22223190

Explicitly replicating research is very (very) hard



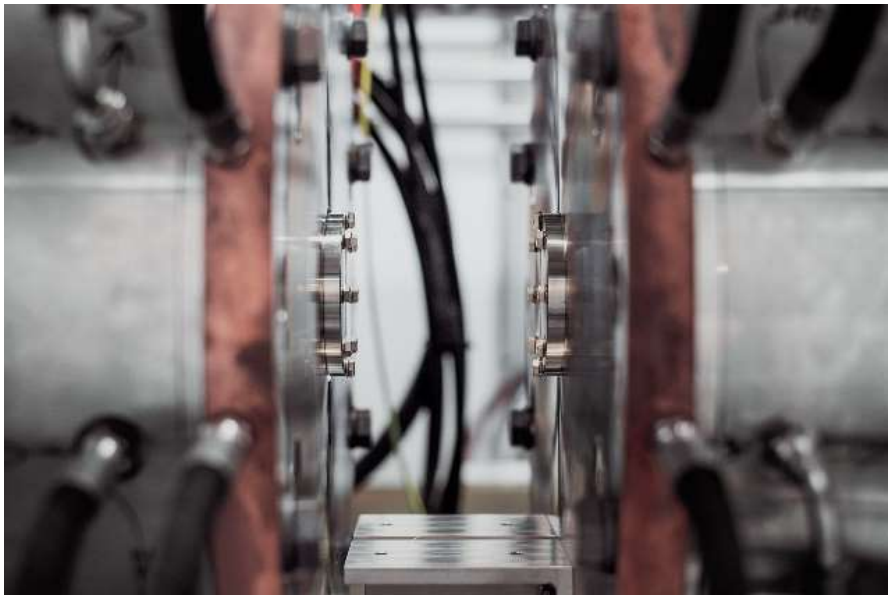
Plan to replicate 50 high-impact cancer papers shrinks to just 18



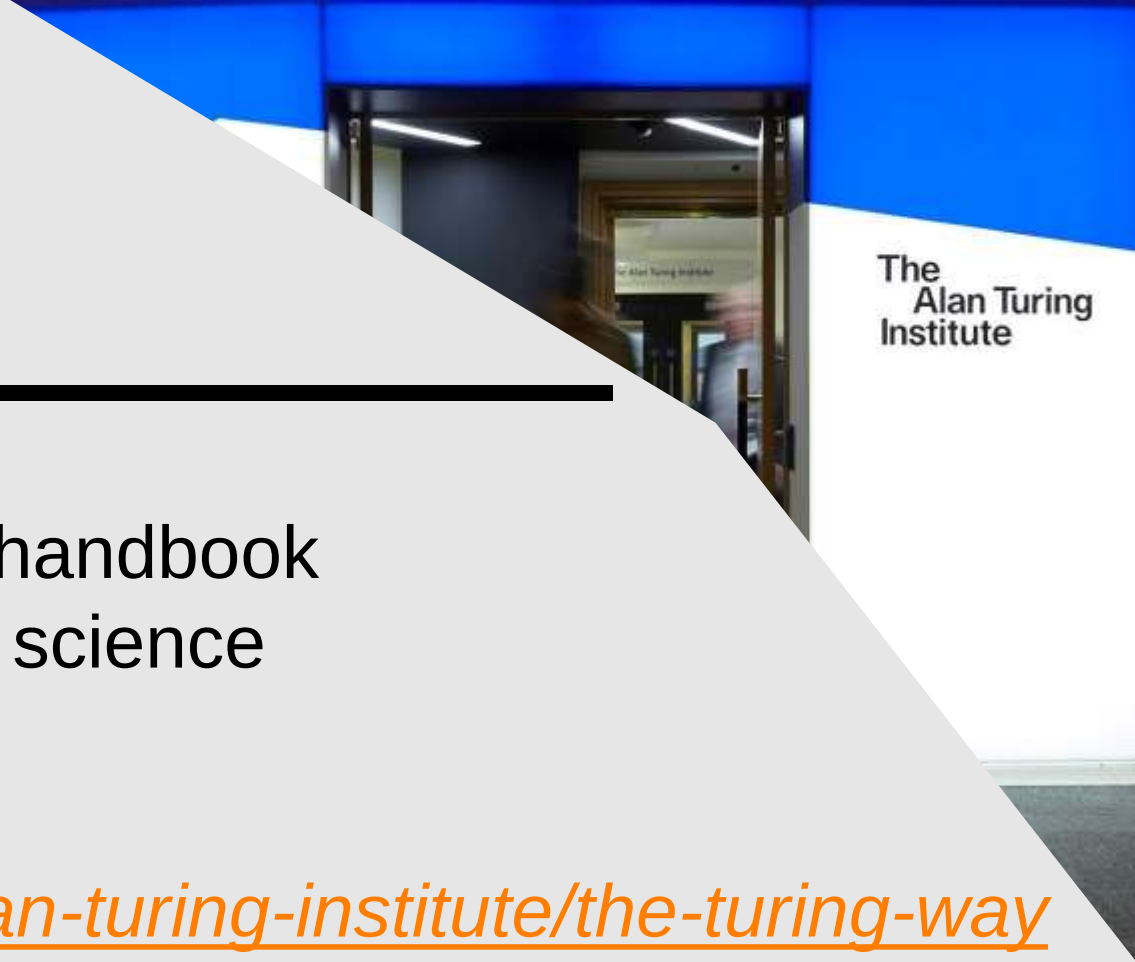
https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology
https://www.sciencemag.org/news/2018/07/plan-replicate-50-high-impact-cancer-papers-shrinks-just-18
https://www.nature.com/news/over-half-of-psychology-studies-fail-reproducibility-test-1.18248

Tools, Practices and Systems

- Focus on real cross-project needs
  - Driven by 'researcher pain points'.
- We will not make things just because we think they're interesting.
  - Usefulness to applied researchers is key.



https://www.turing.ac.uk/research/interest-groups/tools-practices-and-systems



The Turing Way

A lightly opinionated handbook
for reproducible data science

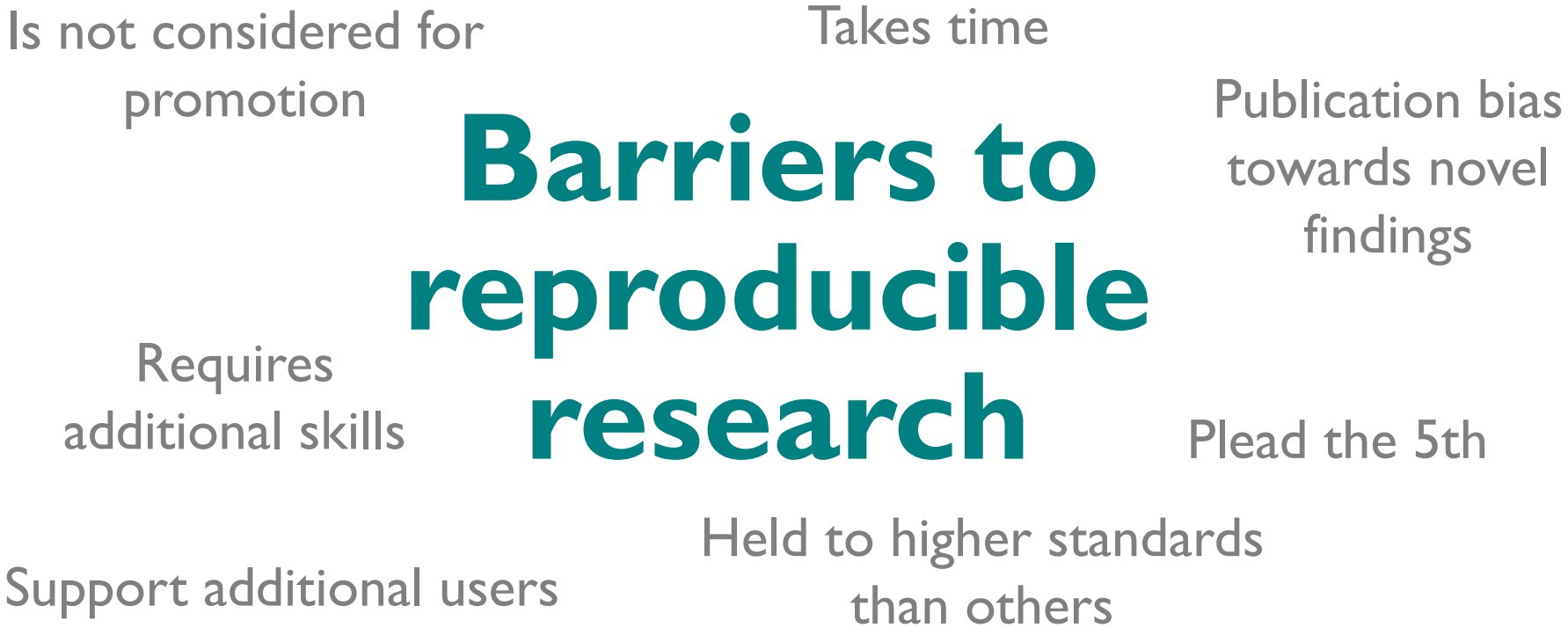
https://github.com/alan-turing-institute/the-turing-way

What does reproducible mean?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

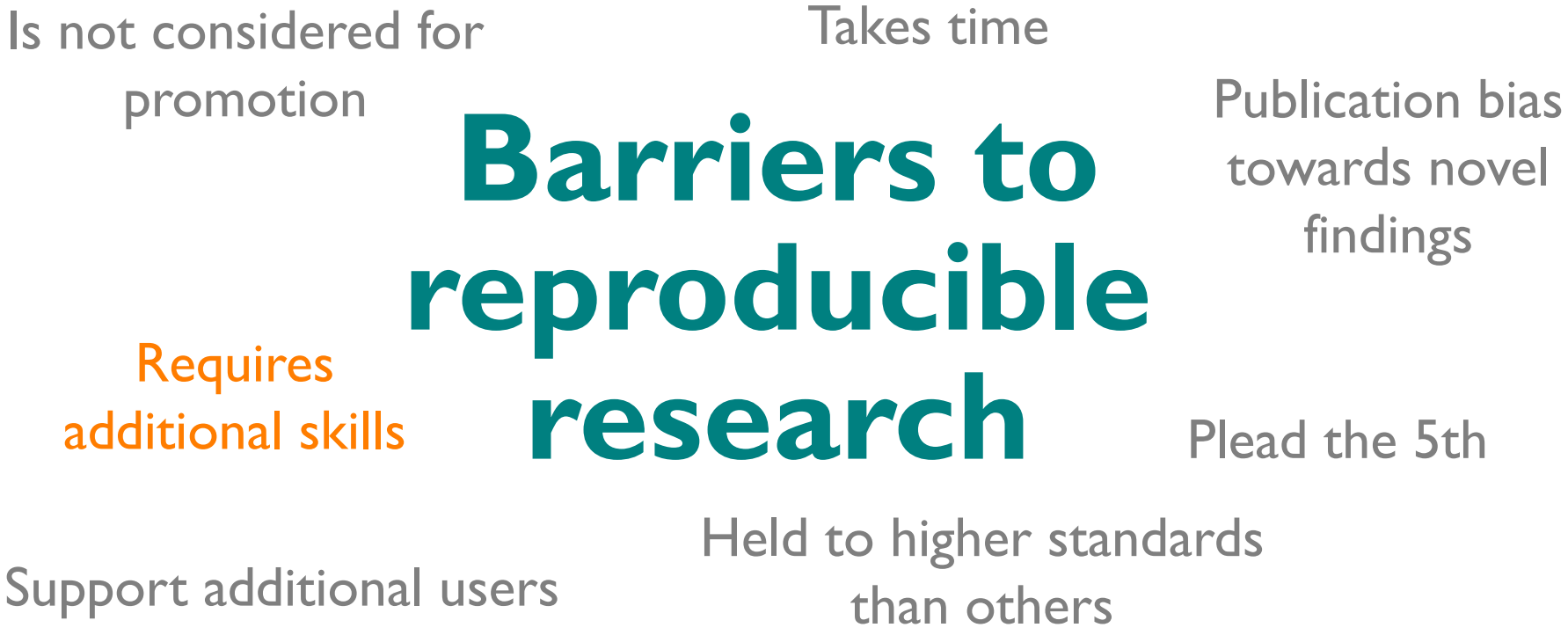
https://dx.doi.org/10.6084/m9.figshare.7140050

Why don’t people do this already?



https://dx.doi.org/10.6084/m9.figshare.7140050

Why don’t people do this already?



https://dx.doi.org/10.6084/m9.figshare.7140050



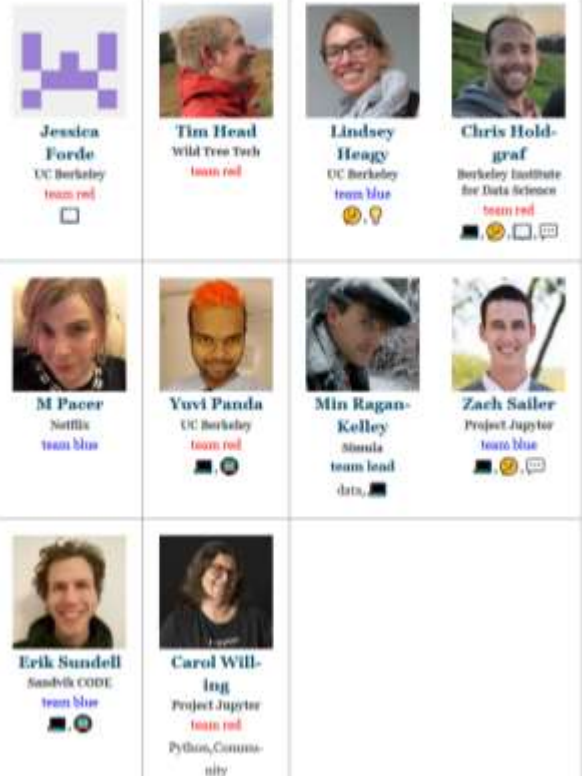

https://github.com/alan-turing-institute/the-turing-way/tree/master/workshops

https://github.com/alan-turing-institute/the-turing-way/tree/master/workshops

Binder Team

Binder's governance and team structure is defined in the Binder Project Governance page. Below we list the current team members of Binder.

(listed alphabetically, with affiliation, and main areas of contribution)



https://github.com/alan-turing-institute/the-turing-way/tree/master/workshops

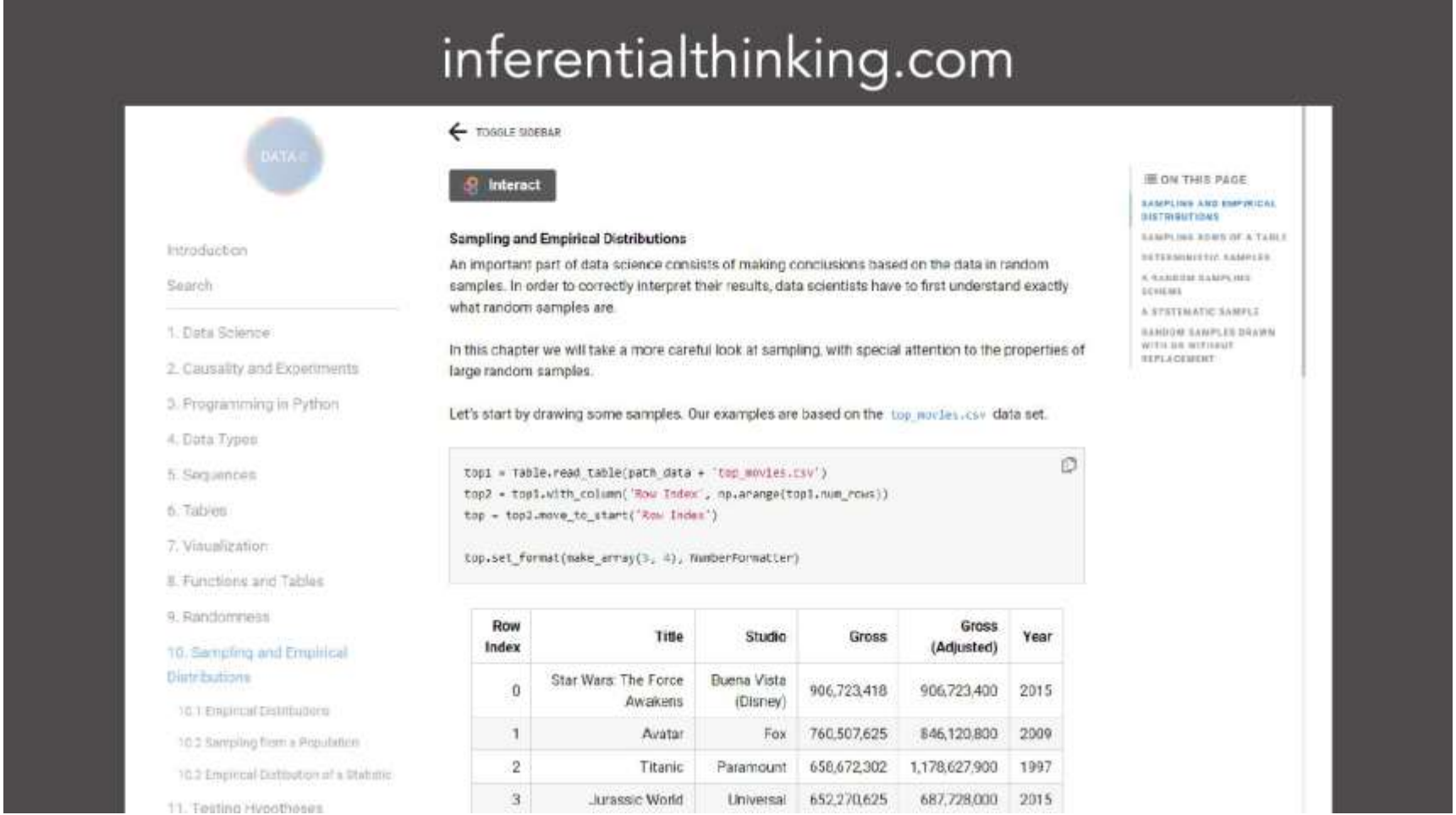
What is Jupyter Book?

Build an online book with
Jupyter Notebooks and Markdown

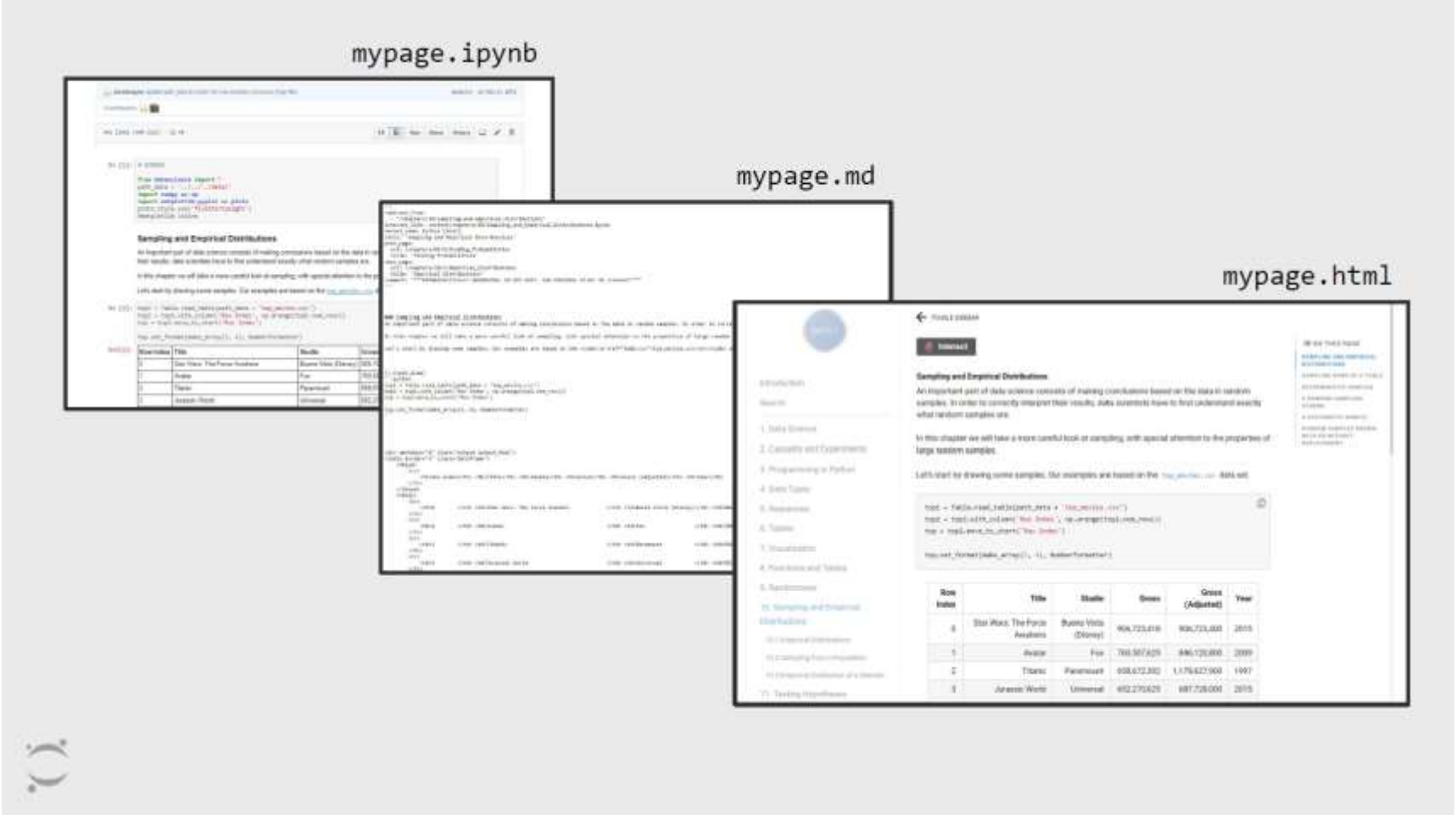
`jupyter.org/jupyter-book`

https://speakerdeck.com/choldgraf/jupyter-book-interactive-books-running-in-the-cloud

inferentialthinking.com



https://speakerdeck.com/choldgraf/jupyter-book-interactive-books-running-in-the-cloud



https://speakerdeck.com/choldgraf/jupyter-book-interactive-books-running-in-the-cloud

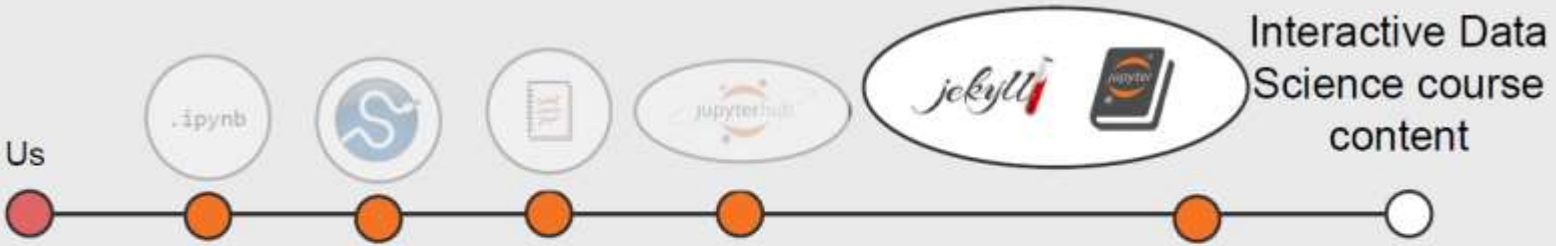
Interactive buttons let readers explore

- A page built from a notebook gets an interact button
- Use JupyterHub/Binder to spawn a Jupyter server
- `git pull` the underlying notebook for the page
- Initialize an interactive environment

https://speakerdeck.com/choldgraf/jupyter-book-interactive-books-running-in-the-cloud

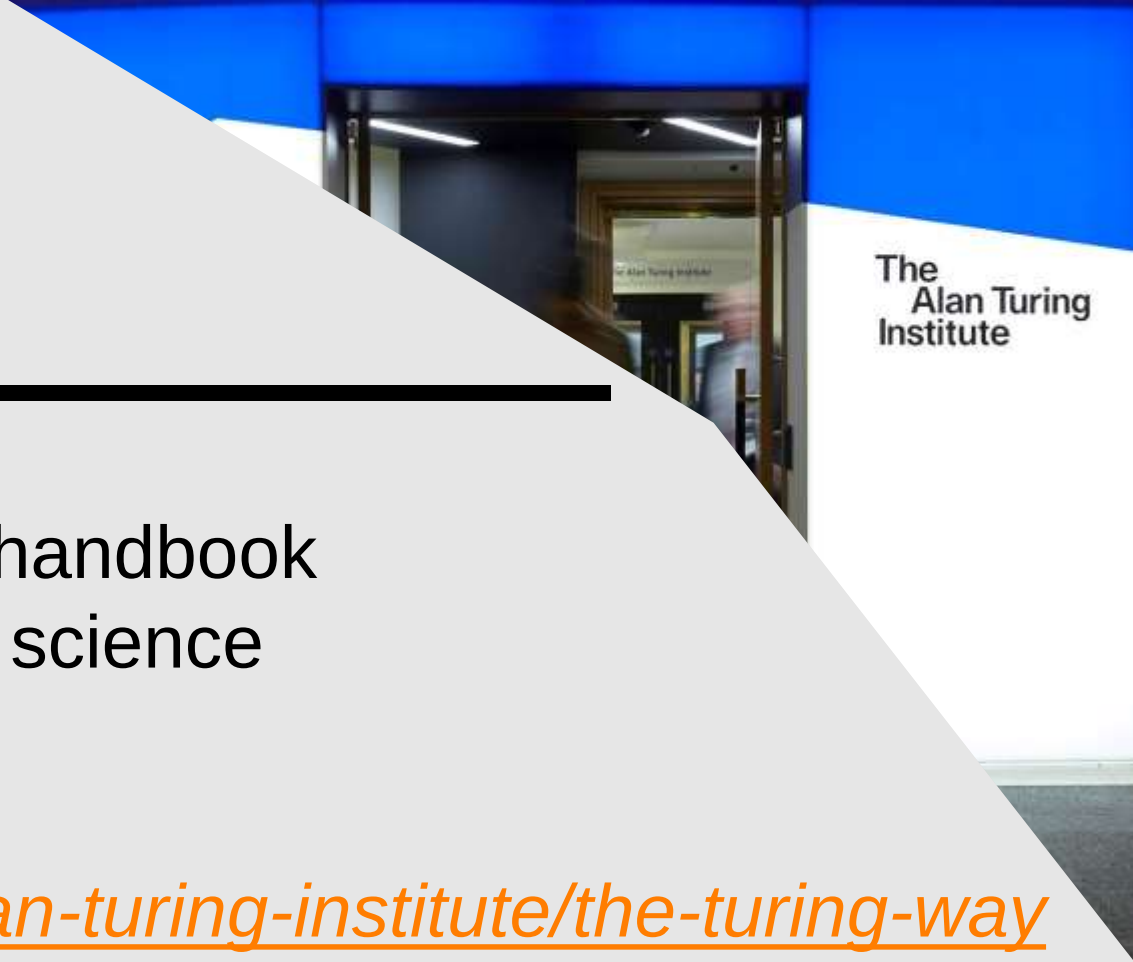
In summary

Jupyter Book builds on tools in the Jupyter ecosystem to create interactive, beautiful books.



jupyter.org/jupyter-book

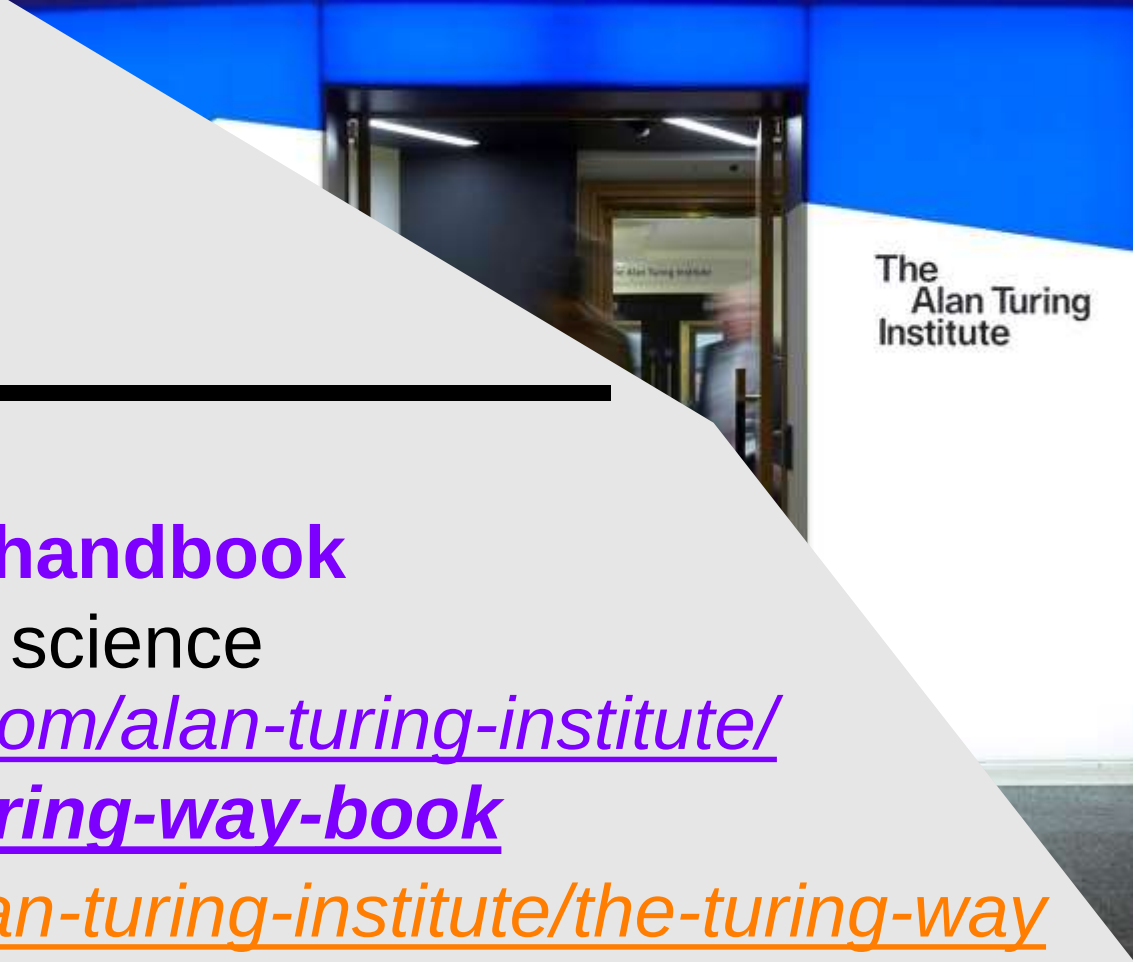
https://speakerdeck.com/choldgraf/jupyter-book-interactive-books-running-in-the-cloud



The Turing Way

A lightly opinionated handbook
for reproducible data science

https://github.com/alan-turing-institute/the-turing-way



The Turing Way

A lightly opinionated **handbook**
for reproducible data science

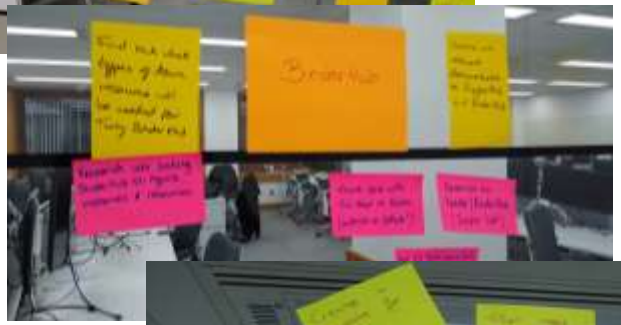
https://github.com/alan-turing-institute/
the-turing-way-book

https://github.com/alan-turing-institute/the-turing-way

Requires additional skills

Chapters will include:

- Research data management
- Open science
- Reproducibility
- Version control with git
- Your working environment (IDE, notebooks etc)
- Capturing your compute environment
- Testing for research
- Continuous integration
- Collaborating through GitHub/GitLab



https://github.com/alan-turing-institute/the-turing-way/blob/master/book_skeleton.md

Requires additional skills

Chapters will include:

- Research data management
- Open research
- Reproducibility
- Version control with git
- Your working environment (IDE, notebooks etc)
- Capturing your compute environment
- Testing for research
- Continuous integration
- Collaborating through GitHub/GitLab

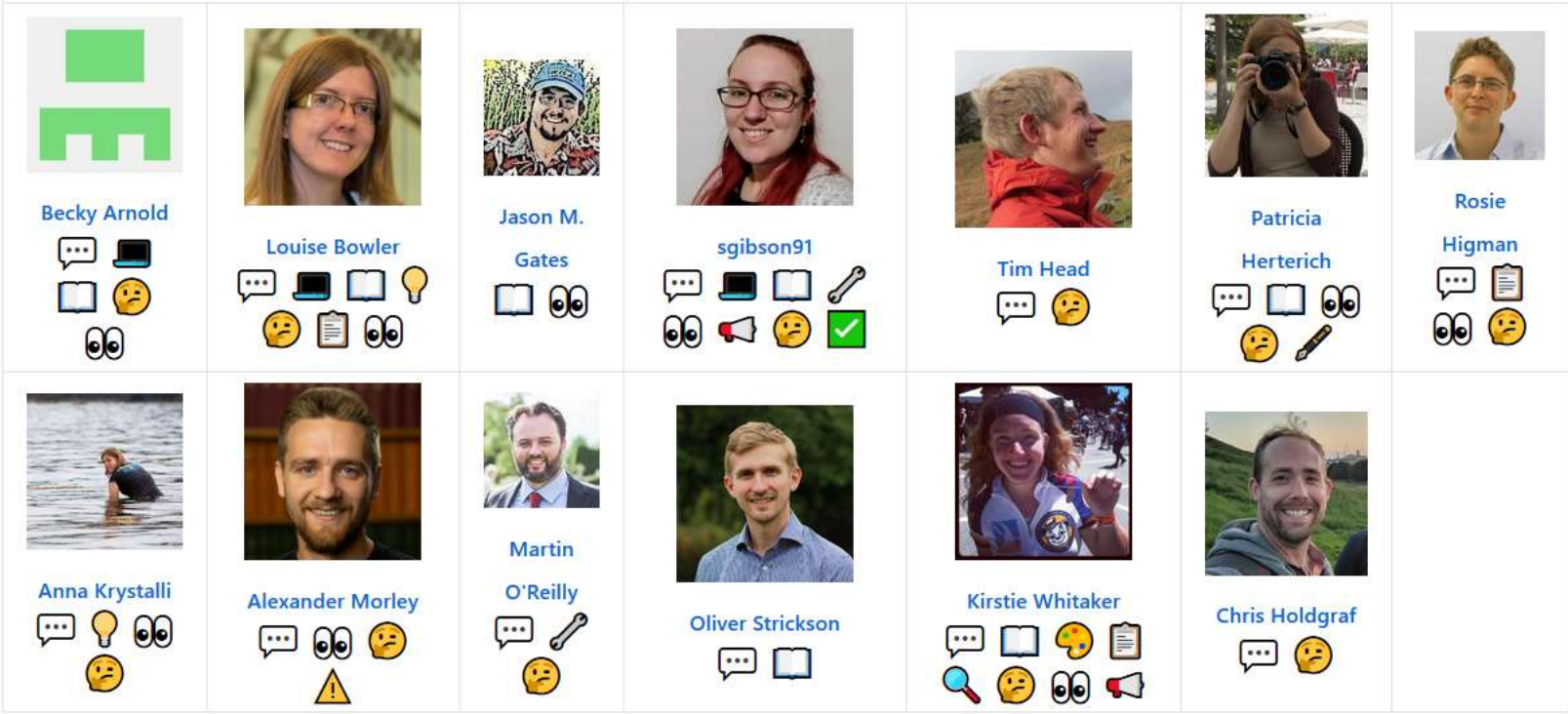
https://github.com/alan-turing-institute/the-turing-way/blob/master/book_skeleton.md

Built by a team….and you!

- Rachael Ainsworth
- Becky Arnold
- Louise Bowler
- Sarah Gibson
- Patricia Herterich
- Rosie Higman
- Anna Krystalli
- Alex Morley
- Martin O’Reilly
- . . .

https://github.com/alan-turing-institute/the-turing-way/blob/master/contributors.md

The emoji key to celebrate our contributors



https://github.com/alan-turing-institute/the-turing-way/blob/master/README.md

Open Leadership Principles

Understanding
You make the work accessible and clear

Sharing
You make the work easy to adapt, reproduce, and spread

Participation & Inclusion
You build shared ownership and agency to make the work inviting and sustainable for all.

Read more
https://mozilla.github.io/olm-whitepaper

@kirstie_j
https://doi.org/10.6084/m9.figshare.7564682

The Alan Turing Institute

moz://a

UNIVERSITY OF CAMBRIDGE

Thank you

Lets discuss how we can work together!

kwhitaker@turing.ac.uk t: @kirstie_j g: @KirstieJane

github.com/alan-turing-institute/the-turing-way
github.com/alan-turing-institute/the-turing-way-book
gitter.im/alan-turing-institute/the-turing-way
tinyletter.com/TuringWay
doi: 10.5281/zenodo.2621280

Checklists for researcher, PI and admin team

- Researcher
  - Version control
  - Capturing compute environment
  - Writing and running the code
- PI
  - Results presented are those from the final run of the analysis
  - Check that another researcher can run the code
- Admin
  - Version control
  - Data and code archive
  - Open access publication

https://github.com/alan-turing-institute/the-turing-way/blob/master/book_skeleton.md

Held to higher standards than others

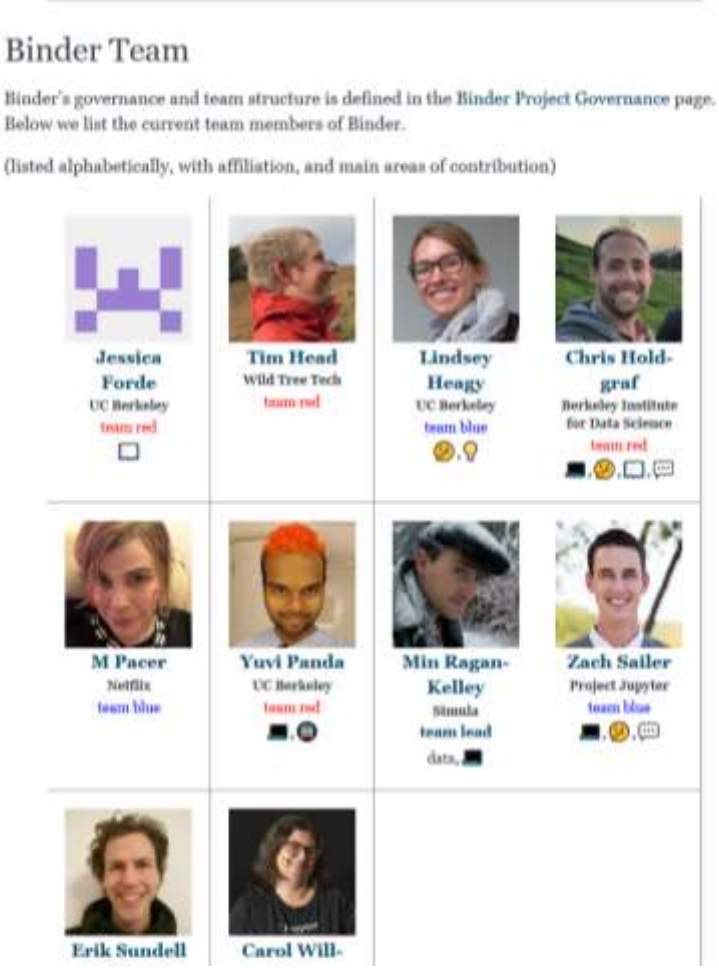
Make reproducibility, "too easy not to do"

Share the responsibility of reproducibility

https://github.com/alan-turing-institute/the-turing-way/blob/master/book_skeleton.md

Interactive checks

- Binder to the rescue!
- Repo2docker: capture the compute environment and builds a container
- Send to cloud resources
- Open a link in a browser and run the code!



Binder Team

Binder's governance and team structure is defined in the Binder Project Governance page. Below we list the current team members of Binder.

(listed alphabetically, with affiliation, and main areas of contribution)

Jessica Forde
UC Berkeley
team red

Tim Head
Wild Tree Tech
team red

Lindsey Heagy
UC Berkeley
team blue

Chris Holdgraf
Berkeley Institute for Data Science
team red

M Pacer
Netflix
team blue

Yuvi Panda
UC Berkeley
team red

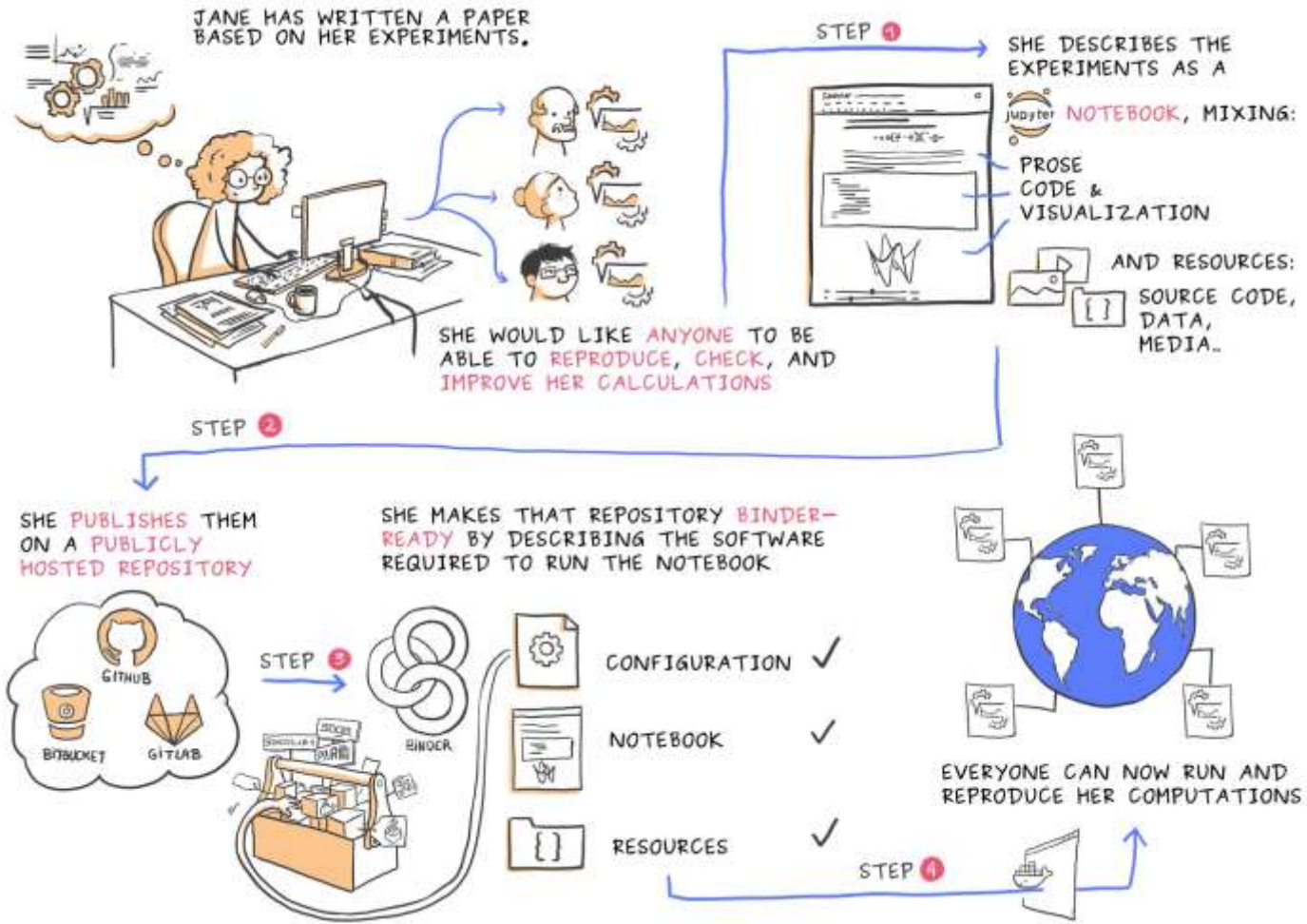
Min Ragan-Kelley
Simula
team lead
data,

Zach Sailer
Project Jupyter
team blue

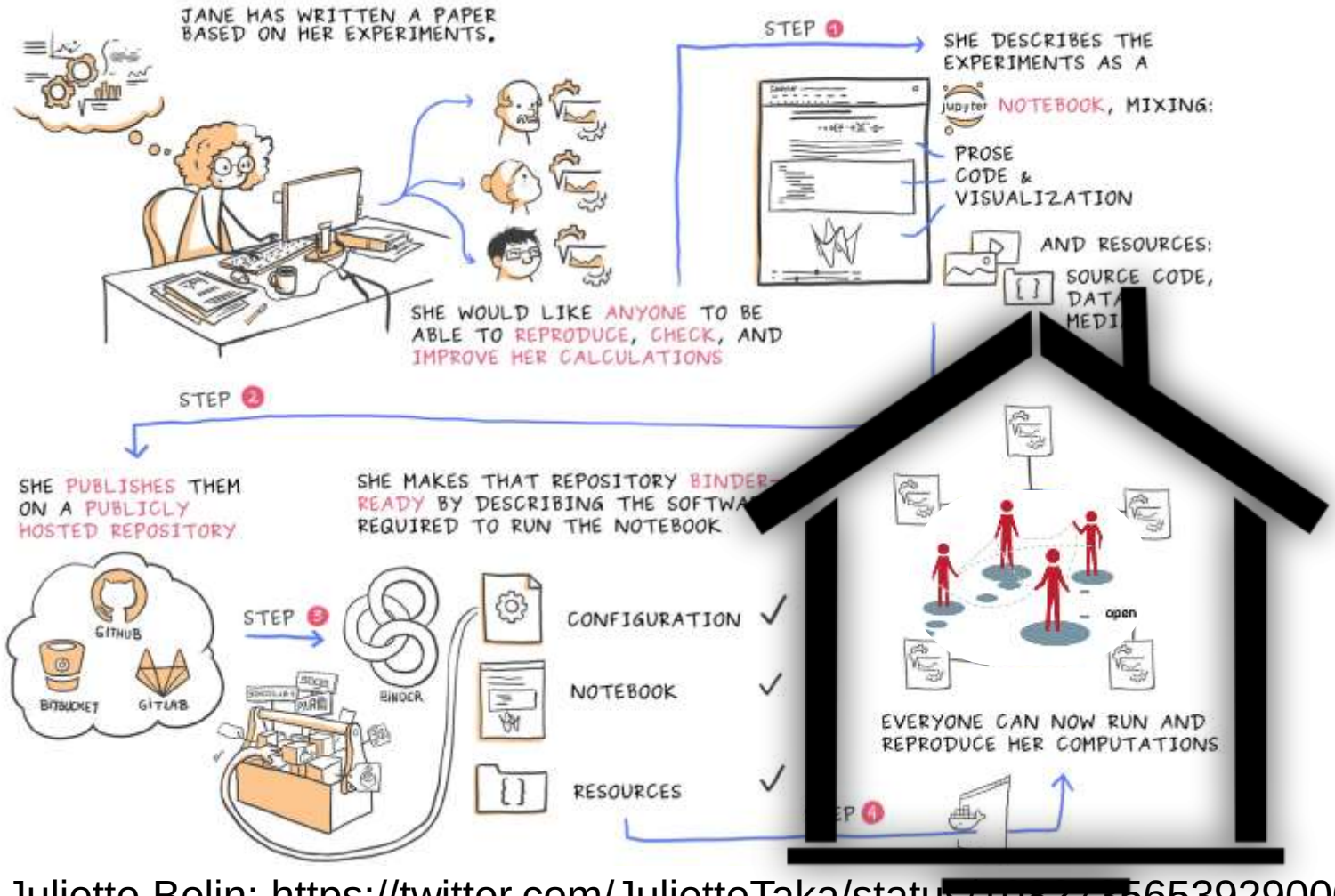
Erik Sundell

Carol Will-

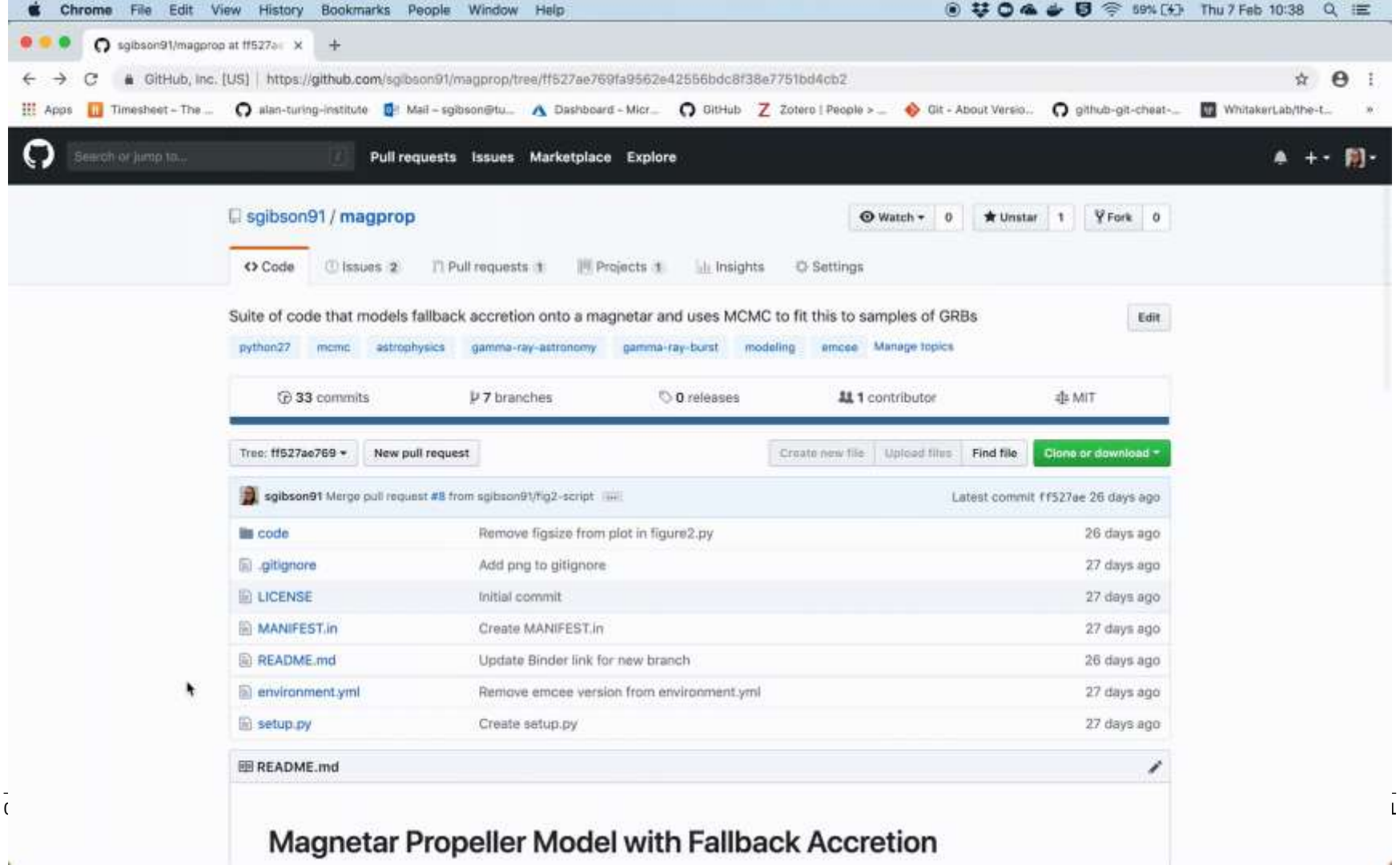
https://jupyterhub-team-compass.readthedocs.io/en/latest/team.html#binder-team



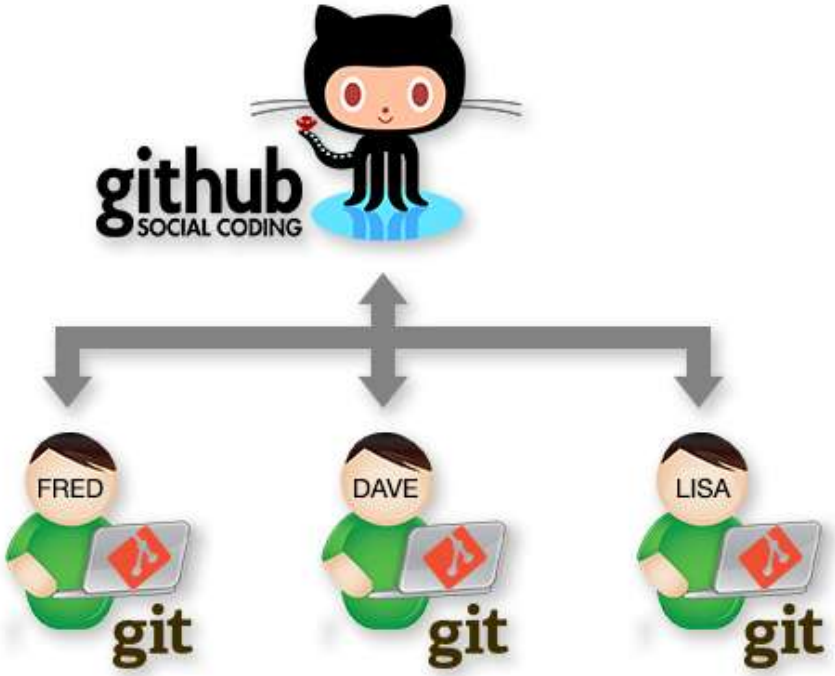
Courtesy of Juliette Belin: https://twitter.com/JulietteTaka/status/1082735653929000960



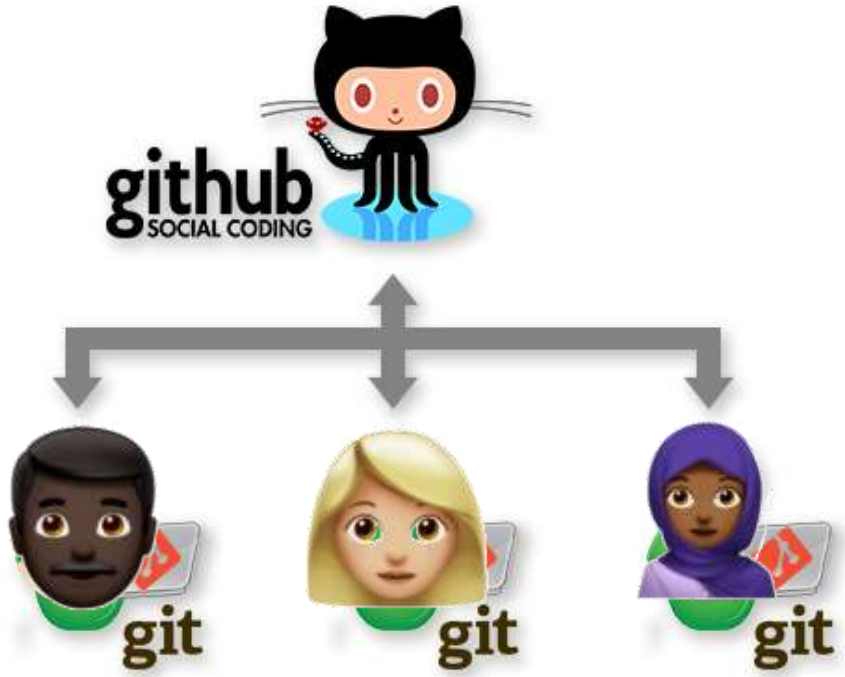
Courtesy of Juliette Belin: https://twitter.com/JulietteTaka/status/1082735653929000960


Chrome File Edit View History Bookmarks People Window Help
Thu 7 Feb 10:38
sgibson91/magprop at ff527ae
GitHub, Inc. [US] https://github.com/sgibson91/magprop/tree/ff527ae769fa9562e42556bdc8f38e7751bd4cb2
Apps Timesheet – The ... alan-turing-institute Mail – sgibson@tu... Dashboard – Micr... GitHub Zotero | People > ... Git – About Versio... github-git-cheat-... WhitakerLab/the-t...
Search or jump to...
Pull requests Issues Marketplace Explore
sgibson91 / magprop
Watch 0 Unstar 1 Fork 0
Code Issues 2 Pull requests 1 Projects 1 Insights Settings
Suite of code that models fallback accretion onto a magnetar and uses MCMC to fit this to samples of GRBs
Edit
python27 mcmc astrophysics gamma-ray-astronomy gamma-ray-burst modeling emcee Manage topics
33 commits 7 branches 0 releases 1 contributor MIT
Tree: ff527ae769 New pull request Create new file Upload files Find file Clone or download
sgibson91 Merge pull request #8 from sgibson91/fig2-script
Latest commit ff527ae 26 days ago
code Remove figsize from plot in figure2.py 26 days ago
.gitignore Add png to gitignore 27 days ago
LICENSE Initial commit 27 days ago
MANIFEST.in Create MANIFEST.in 27 days ago
README.md Update Binder link for new branch 26 days ago
environment.yml Remove emcee version from environment.yml 27 days ago
setup.py Create setup.py 27 days ago
README.md
Magnetar Propeller Model with Fallback Accretion

Version control



Version control



Testing (aka making explicit sanity checks)

Is your code doing what you think it’s doing? Does 2 + 2 = 4?

Testing (aka making explicit sanity checks)

Is your code doing what you think it's doing? Does 2 + 2 = 4?


```
Assert.AreEqual(
    GetTimeOfDay(),
    “Morning” )
```

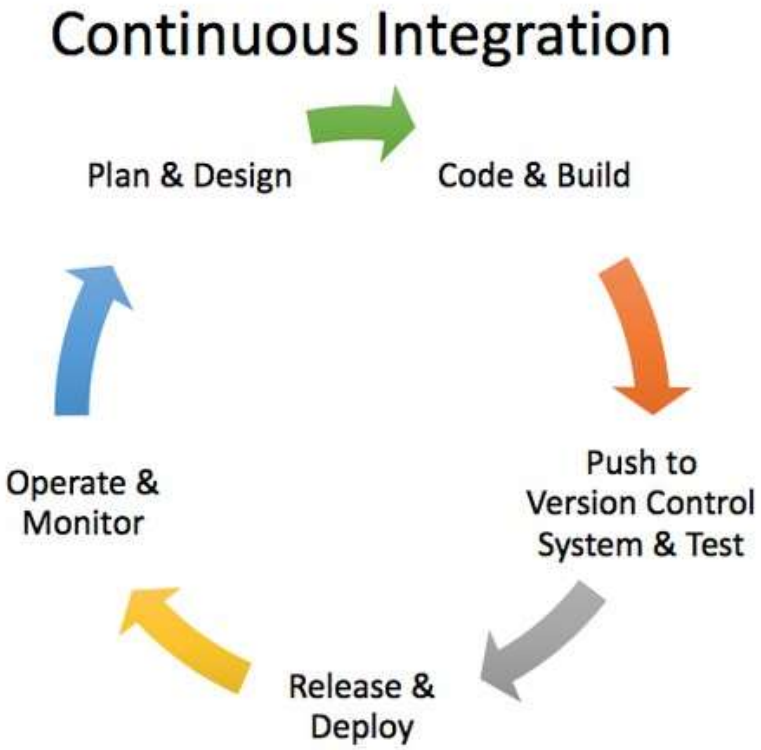
https://www.toptal.com/qa/how-to-write-testable-code-and-why-it-matters

Testing (aka making explicit sanity checks)

Is your code doing what you think it's doing? Does 2 + 2 = 4?

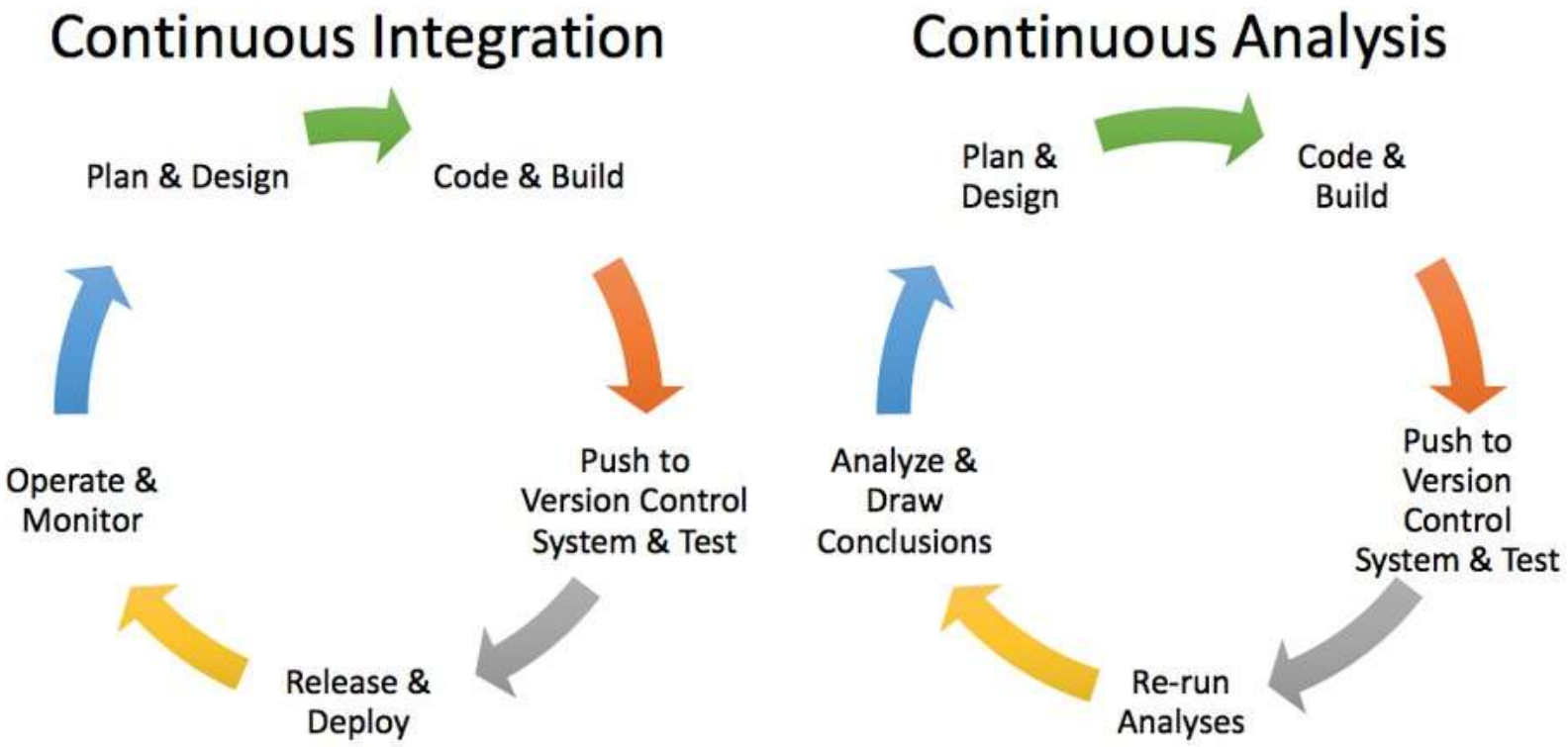
A very simple check: Is total brain volume within an expected range?

Continuous integration for research



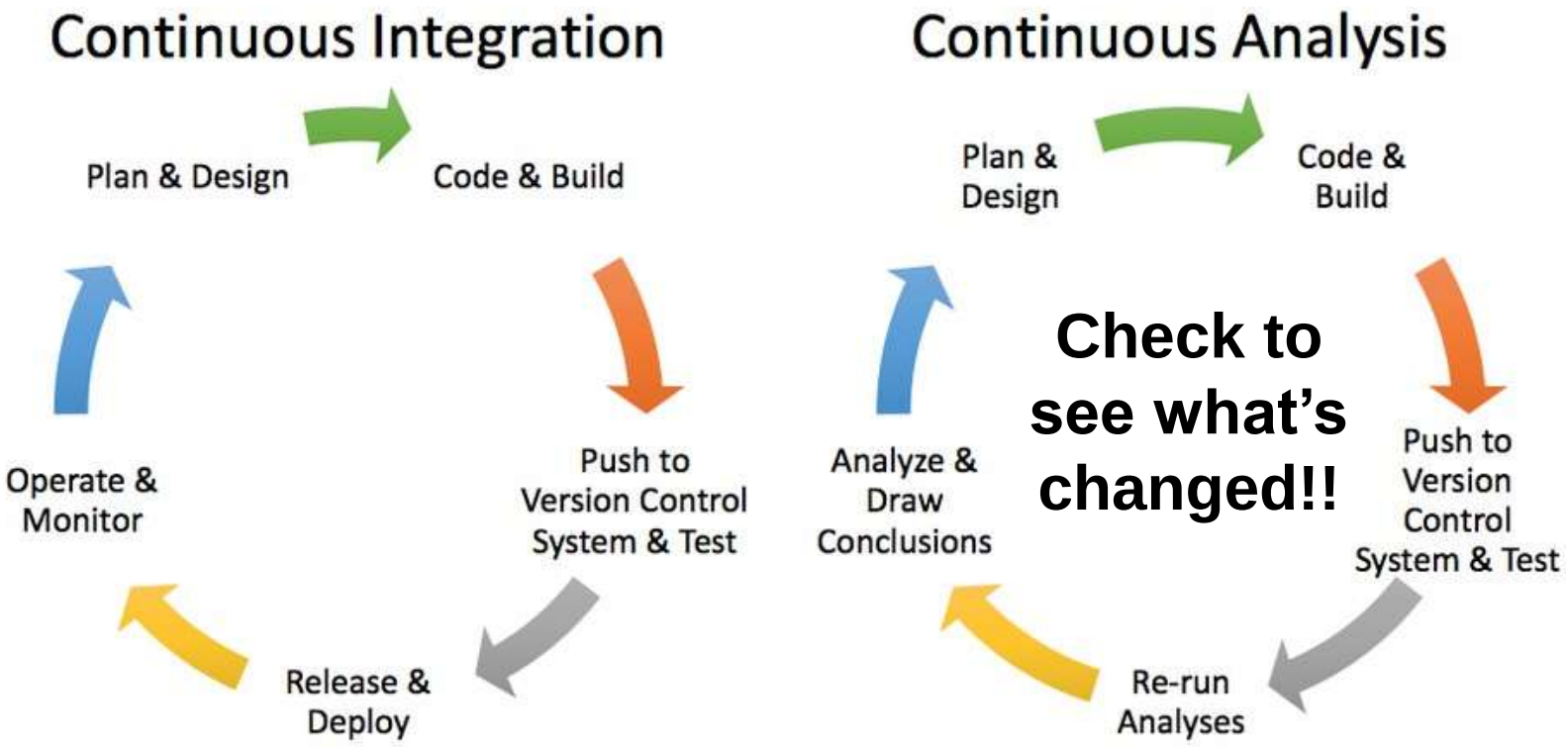
https://elifesciences.org/labs/e623676c/reproducibility-automated

Continuous integration for research



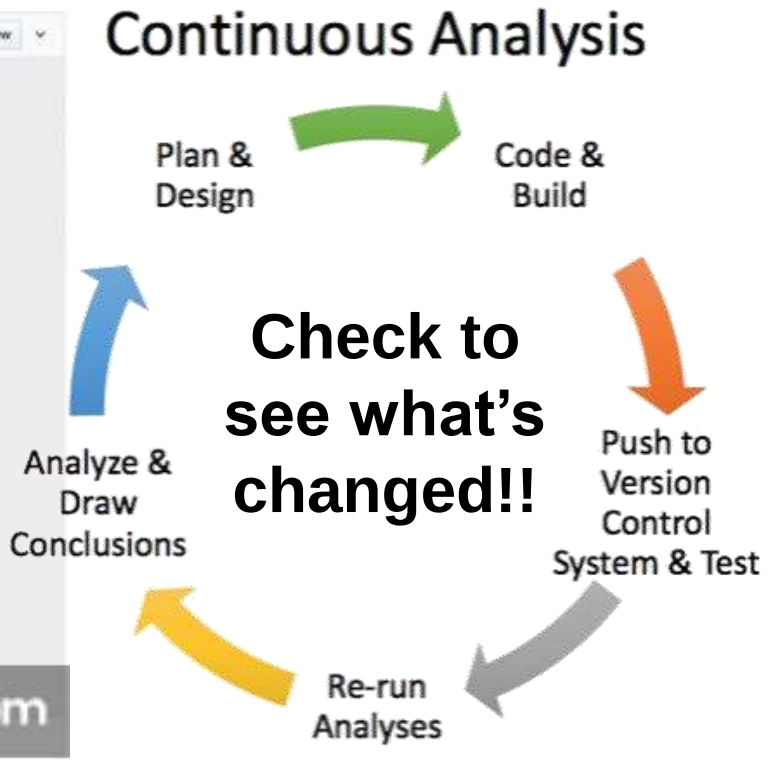
https://elifesciences.org/labs/e623676c/reproducibility-automated

Continuous integration for research



https://elifesciences.org/labs/e623676c/reproducibility-automated

Continuous integration for research

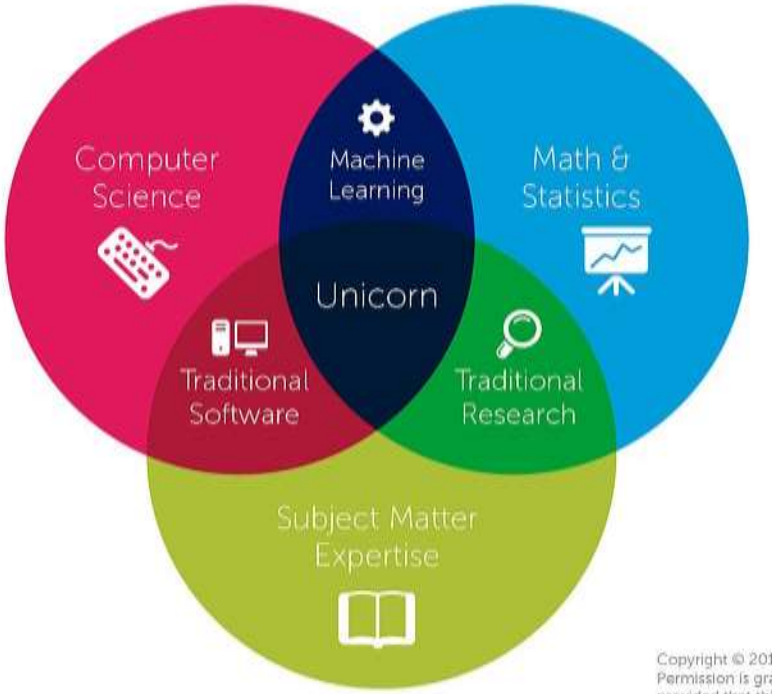


https://elifesciences.org/labs/e623676c/reproducibility-automated

Building a culture of collaborative science

https://github.com/alan-turing-institute/the-turing-way

The Data Science Unicorn



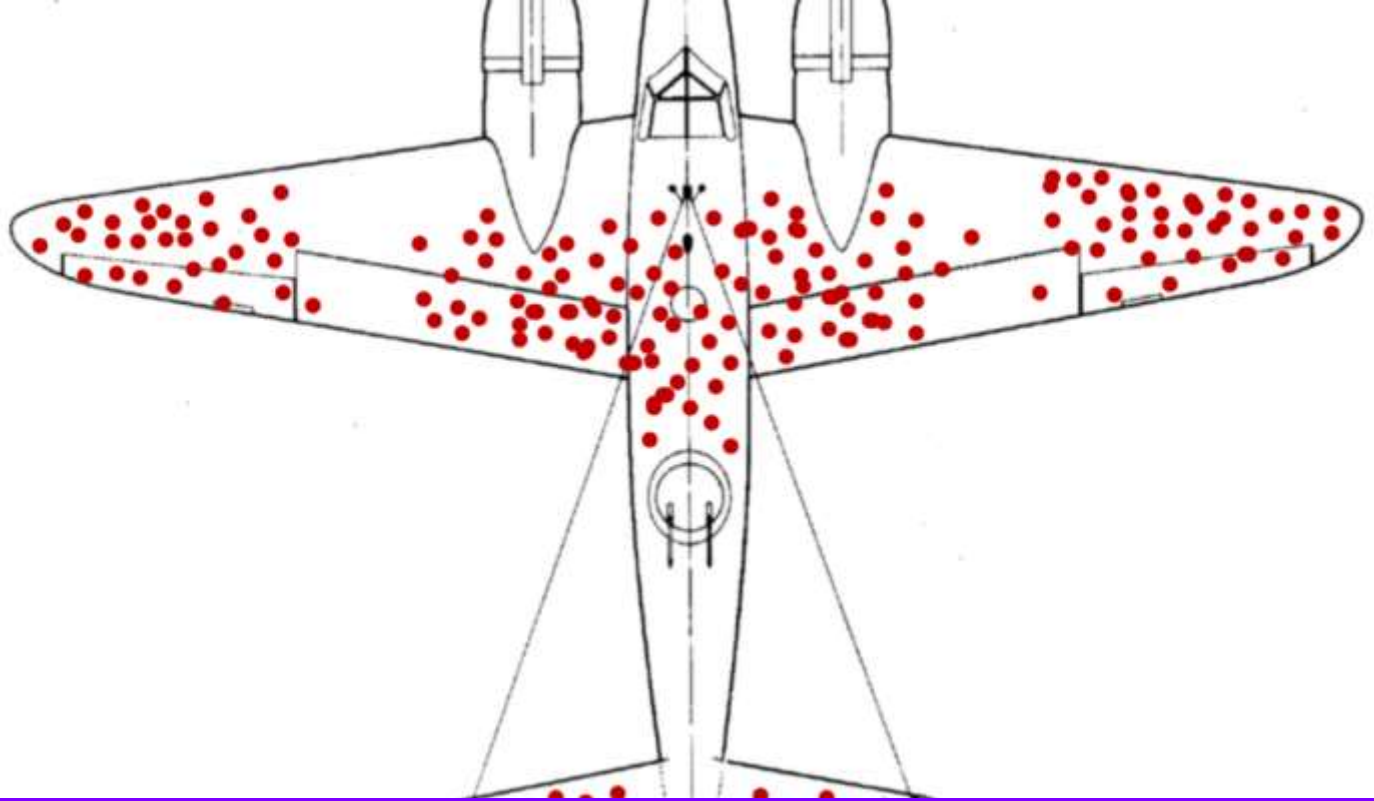
https://www.luther.edu/computer-science/data-science-major/why-study

https://neurohackademy.org/apply

Open is so much more than reproducible

Adapted from: https://www.meetup.com/Berlin-Open-Science-Meetup/

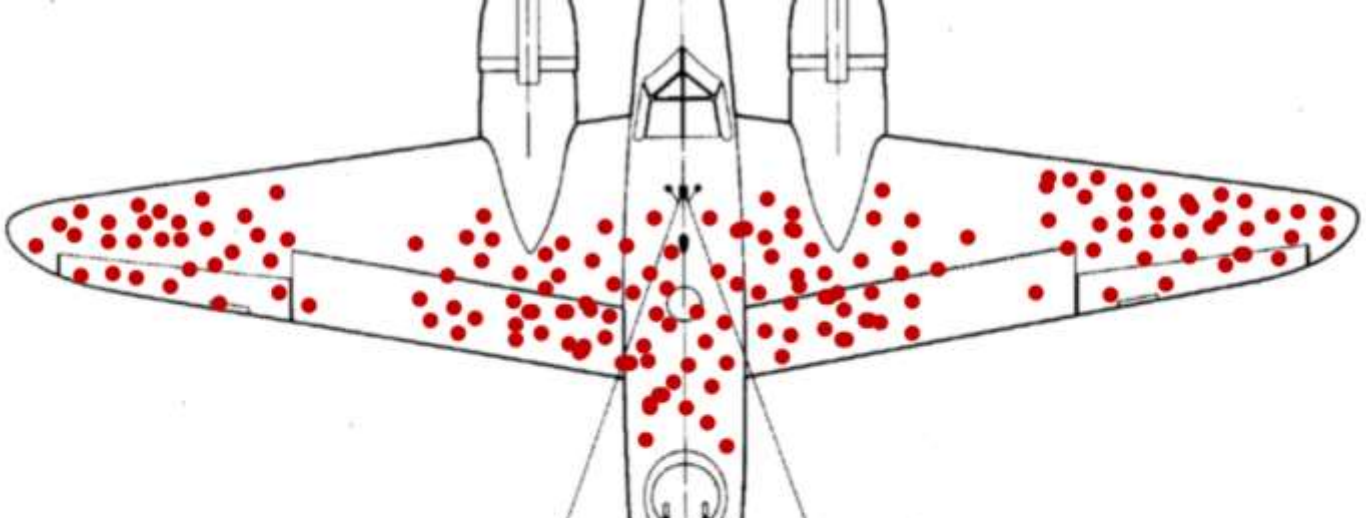
Robin Champieux and Danielle Robinson



https://medium.com/@penguinpress/an-excerpt-from-how-not-to-be-wrong-by-jordan-ellenberg-664e708cfc3d

https://doi.org/10.6084/m9.figshare.7564682

moz://a



The armor, said Wald, doesn't go where the bullet holes are. It goes where the bullet holes aren't: on the engines.

https://medium.com/@penguinpress/an-excerpt-from-how-not-to-be-wrong-by-jordan-ellenberg-664e708cfc3d

https://doi.org/10.6084/m9.figshare.7564682


Lewis Hou
@fiddleBrain
Follow
Privilege to be part of @STEMGamechange & meet so many brilliant folks making #STEM more diverse & inclusive!🎉Lots of actions, reflections & collaborations moving forward - this is just the start!🙌🙏 Thanks to all organisers, our evidence-based #scicomm team & #STEMGamechangers!


INCLUSIVE & INTERSECTIONAL REVOLUTION!


The Alan Turing Institute
Out and About in STEM
Legal information to support global mobility of LGBT+ individuals in STEM
https://stemgamechangers.github.io


Gamechangers for diversity in STEM

Data science at scale

ttps://www.flickr.com/photos/mozfest/22455631157/in/album-72157658649418943

Thank you!

The Alan Turing Institute

UNIVERSITY OF CAMBRIDGE

moz://a

Please come and join us!

github.com/alan-turing-institute/the-turing-way

gitter.im/alan-turing-institute/the-turing-way

@kirstie_j, @whitakerlab

doi: 10.6084/m9.figshare.7649156

https://www.flickr.com/photos/mozfest/22455631157/in/album-72157658649418943