
The Turing Way

Building a culture of collaborative science

Kirstie Whitaker

Cambridge Psychiatry PI Day, January 2019

Slides at <https://doi.org/10.6084/m9.figshare.7649156>





Neurohackweek 2016

Photo credit: Chris Gorgolewski

- **BSc Physics**
- **MSc Medical Physics**
- **PhD Neuroscience**
- **Postdoc Dept Psychiatry, Cambridge**
- **Mozilla Fellow for Science**
- **Research fellow Alan Turing Institute & senior research associate Dept Psychiatry**



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



THE UNIVERSITY
of EDINBURGH

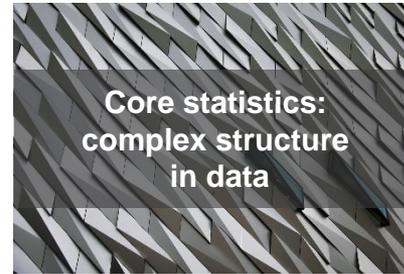
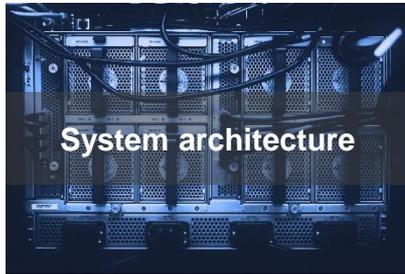


Challenges

Advance data science and artificial intelligence to...



Core capabilities



The Alan Turing Institute to spearhead new cutting-edge data science and AI research after £48 million government funding boost

Tuesday 18 Dec 2018

Learn more ↓

<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.



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.



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.



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.



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.



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.



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.



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.



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.



Health



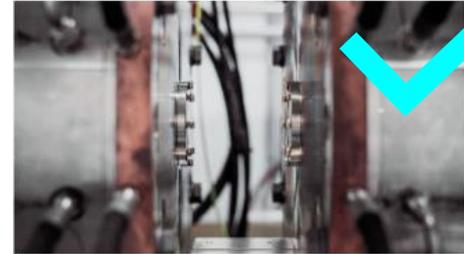
Accelerating the scientific understanding of human disease and improving human health through data-driven innovation in AI and statistical science.



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.



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.



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.

Cross cutting theme: Tools, systems and practices

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

Why don't people do this already?

Is not considered for
promotion

Takes time

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional skills

Plead the 5th

Support additional users

Held to higher standards
than others

Why don't people do this already?

Is not considered for
promotion

Takes time

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional skills

Plead the 5th

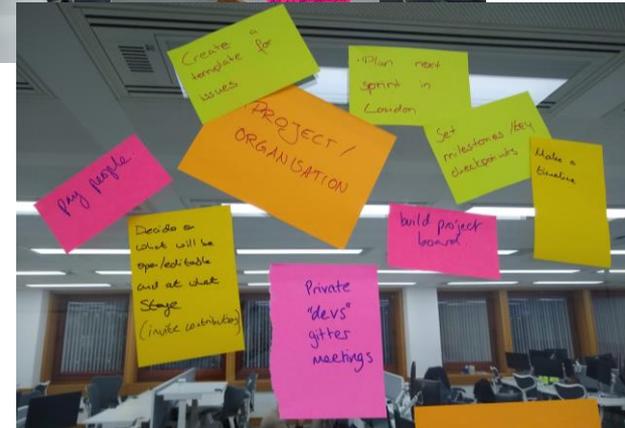
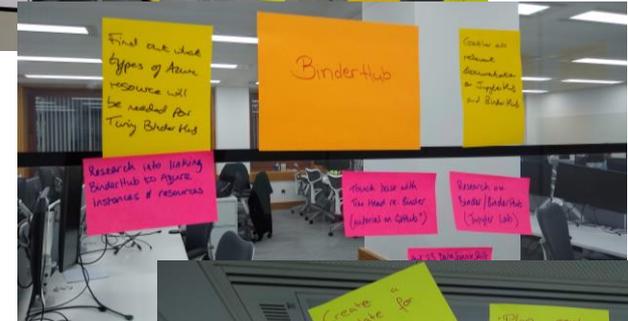
Support additional users

Held to higher standards
than others

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



Built by a team....and you!

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



Open Leadership Principles

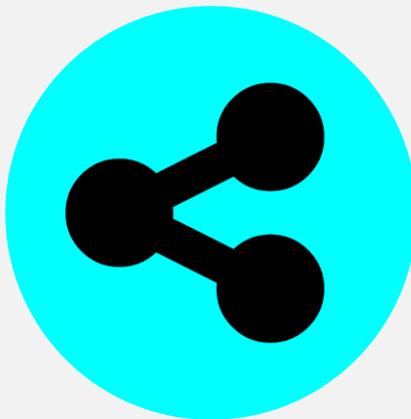


Understanding

You make the work accessible and clear

Read more

<https://mozilla.github.io/olm-whitepaper>



Sharing

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



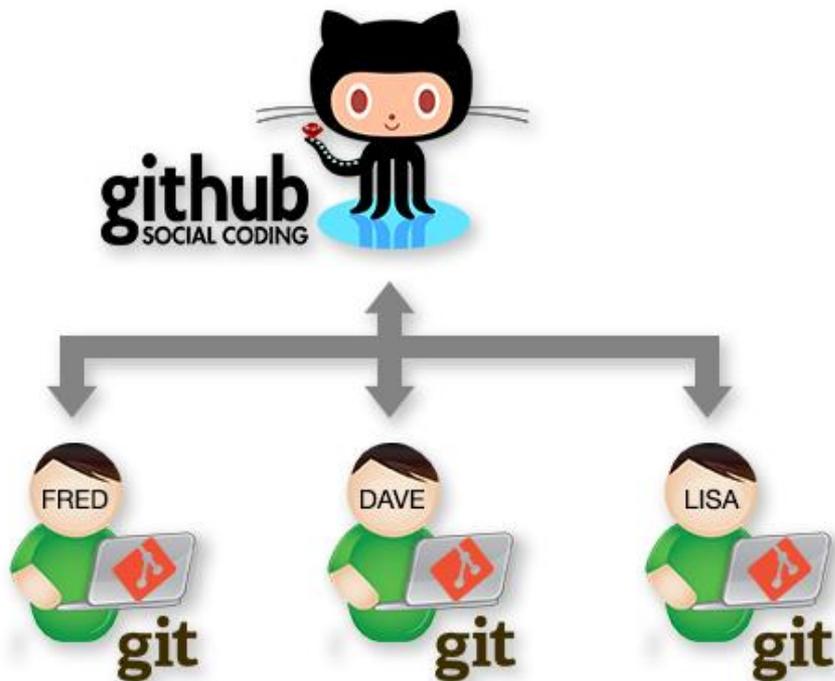
Participation & Inclusion

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

[@kirstie_j](#)

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

Version control



"FINAL".doc



FINAL.doc!



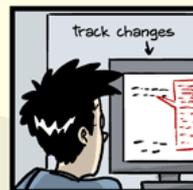
FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



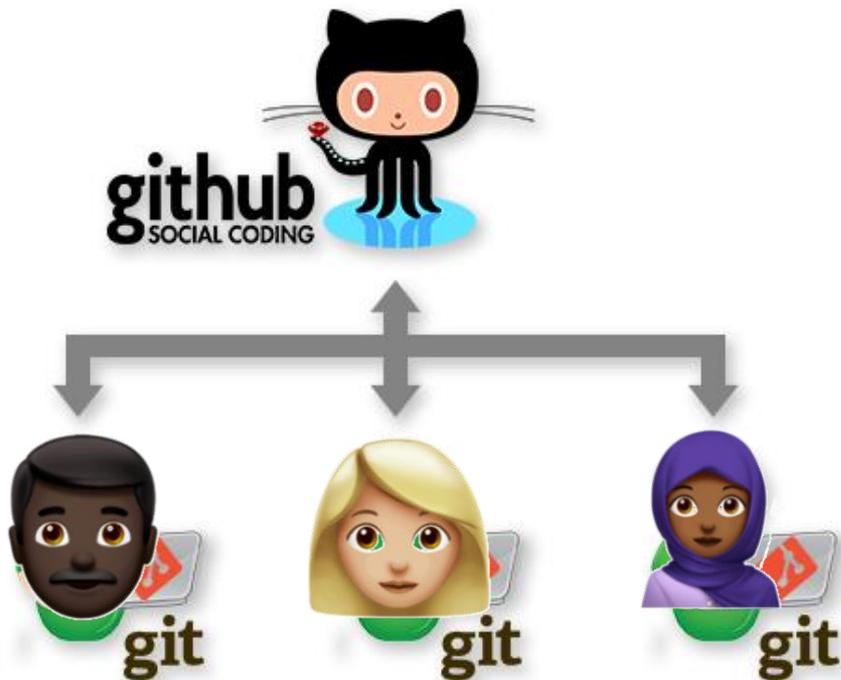
FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSSCHOOL?????..do



Version control



"FINAL".doc



FINAL.doc!



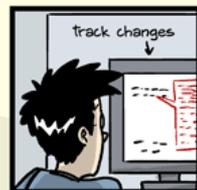
FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSCHOOL?????.doc

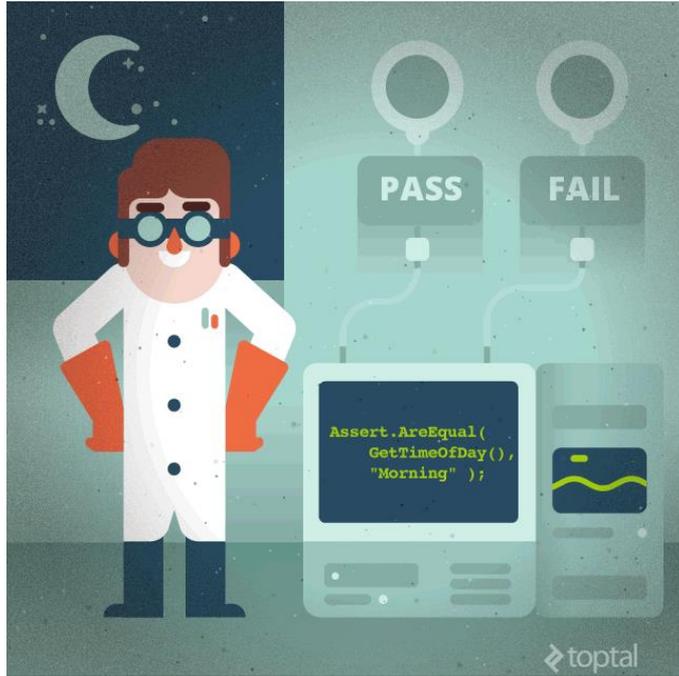
JORGE CHAN © 2012

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)

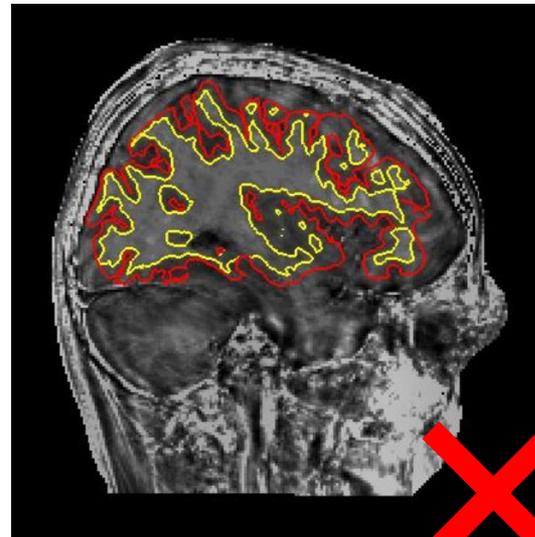
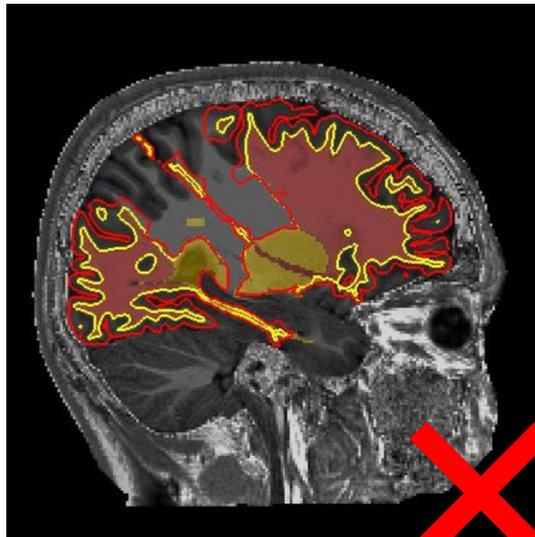
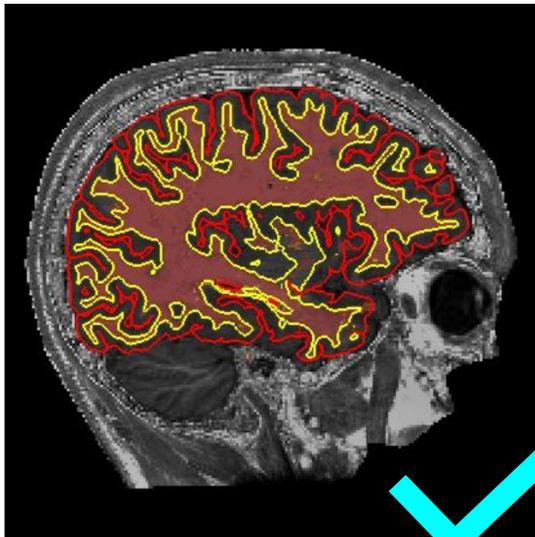
Is your code doing what you think it's doing? Does $2 + 2 = 4$?



```
Assert.AreEqual (
    GetTimeOfDay () ,
    "Morning" )
```

Testing (aka making explicit sanity checks)

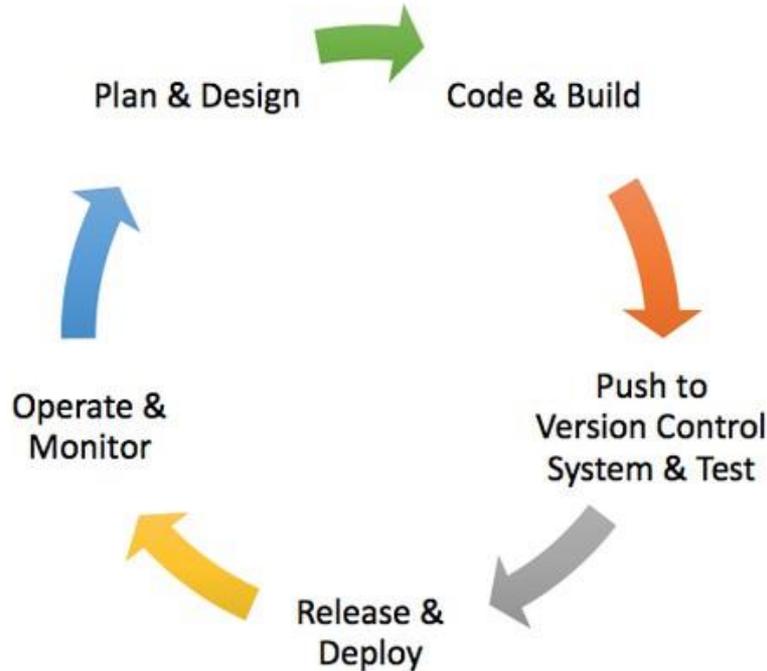
Is your code doing what you think it's doing? Does $2 + 2 = 4$?



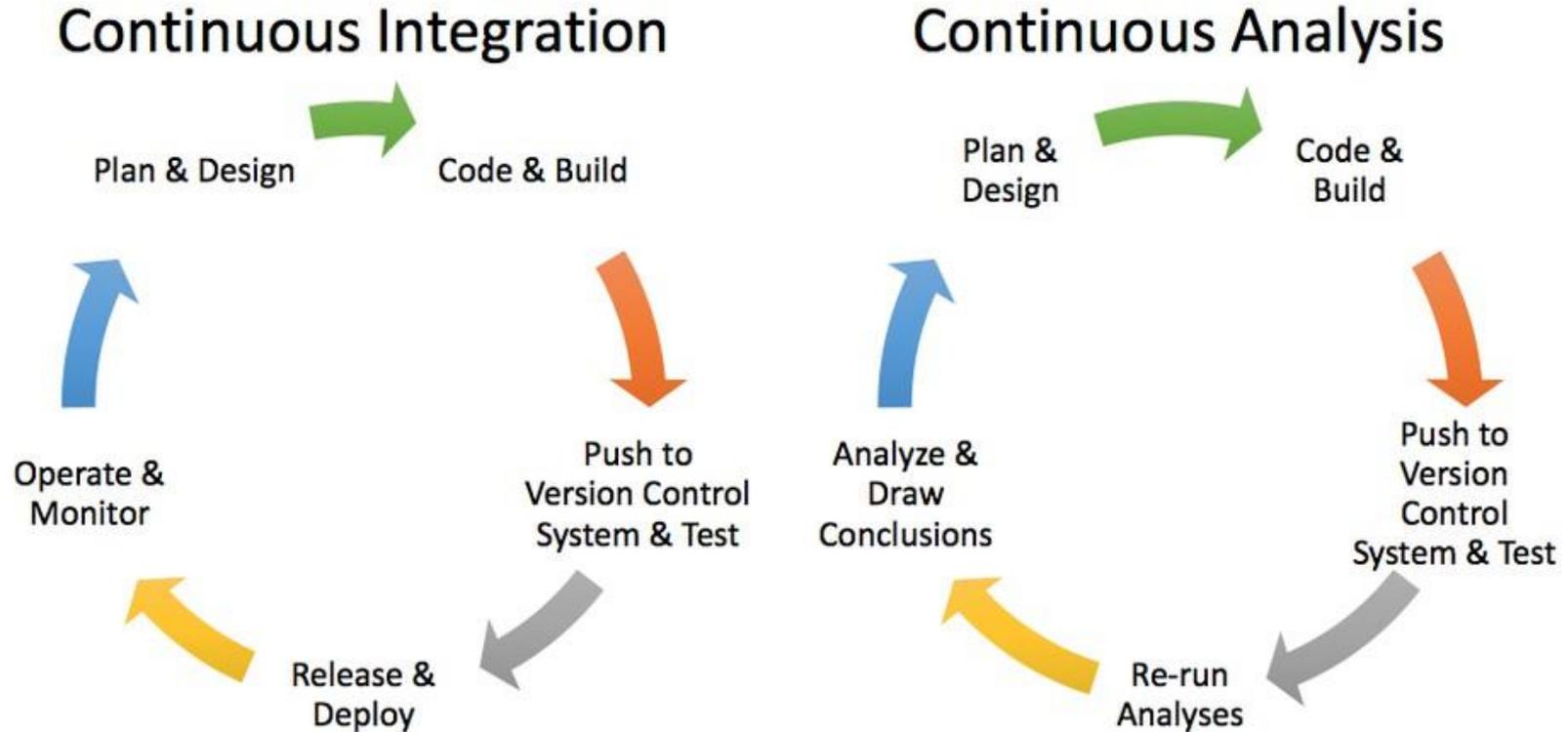
A very simple check: Is total brain volume within an expected range?

Continuous integration for research

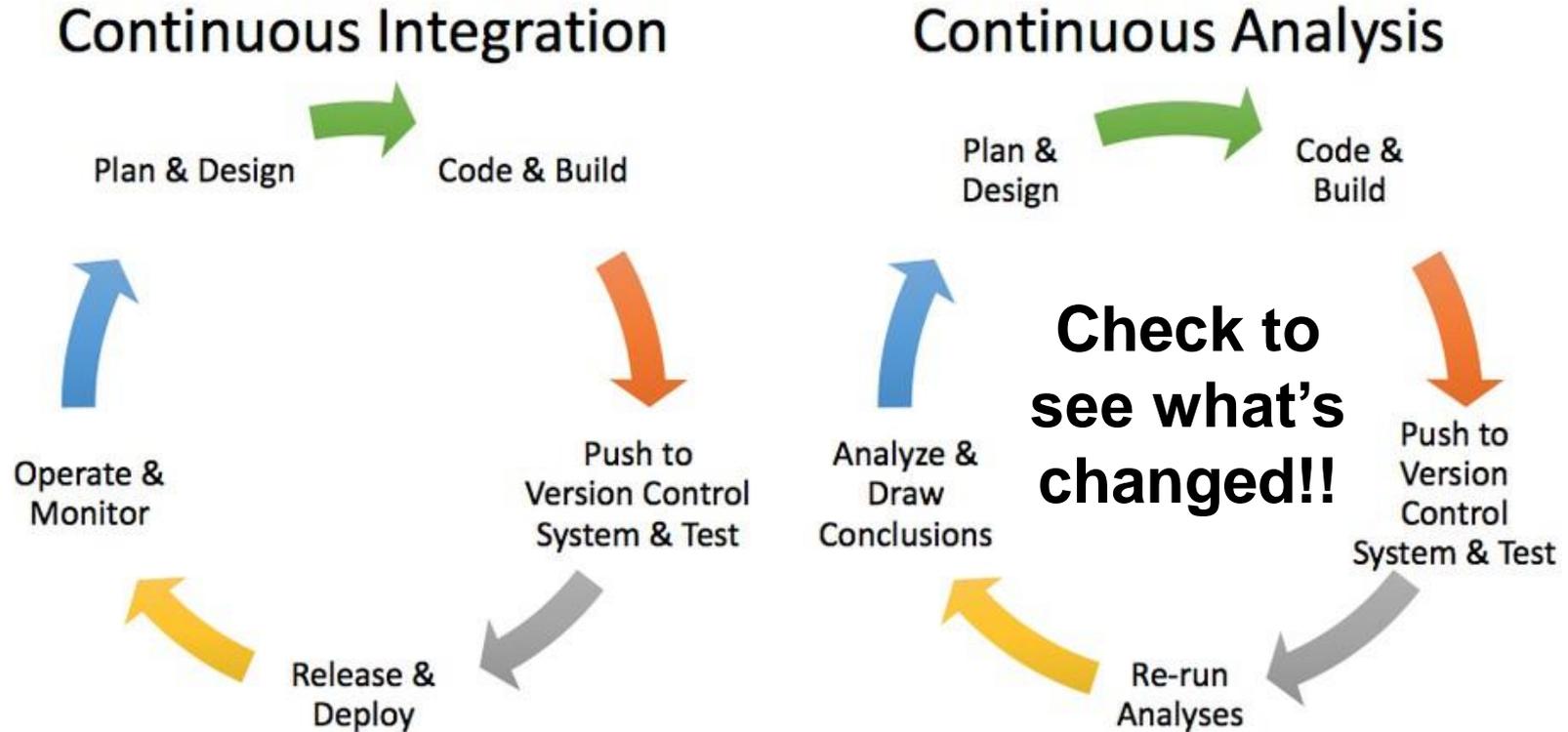
Continuous Integration



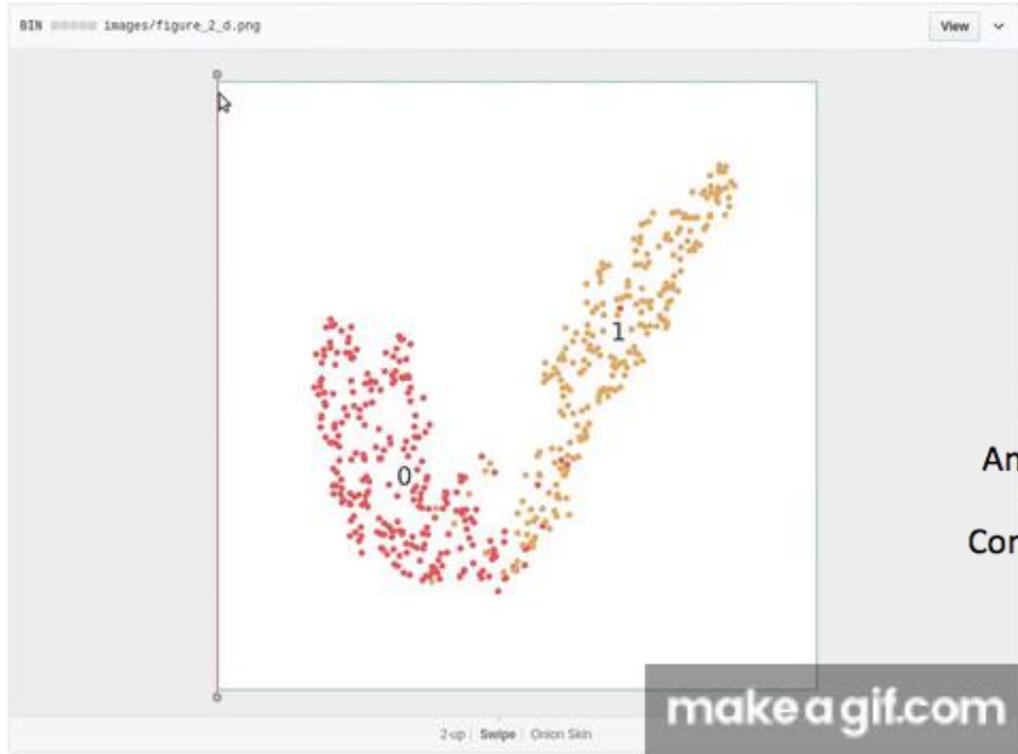
Continuous integration for research



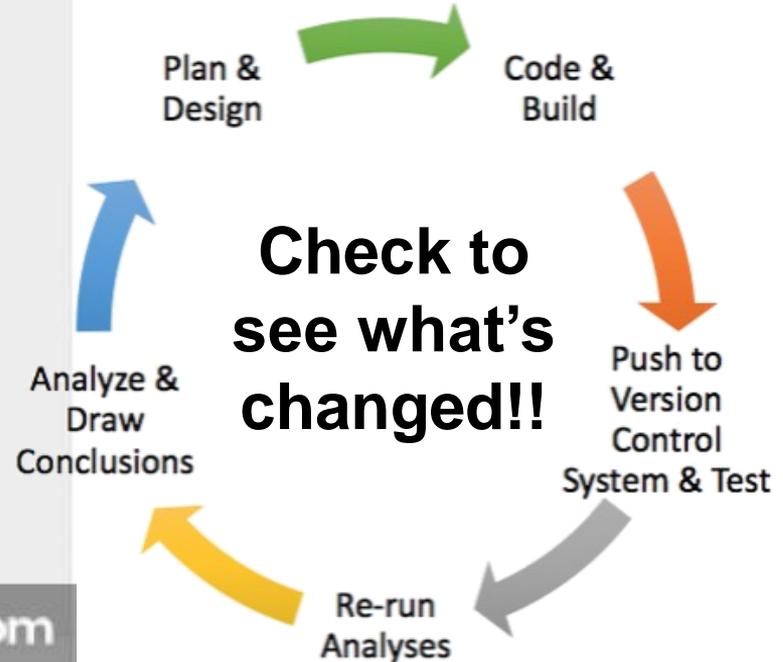
Continuous integration for research



Continuous integration for research



Continuous Analysis



Held to higher standards than others

*Make
reproducibility,
“too easy
not to do”*

*Share the
responsibility
of
reproducibility*

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

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!

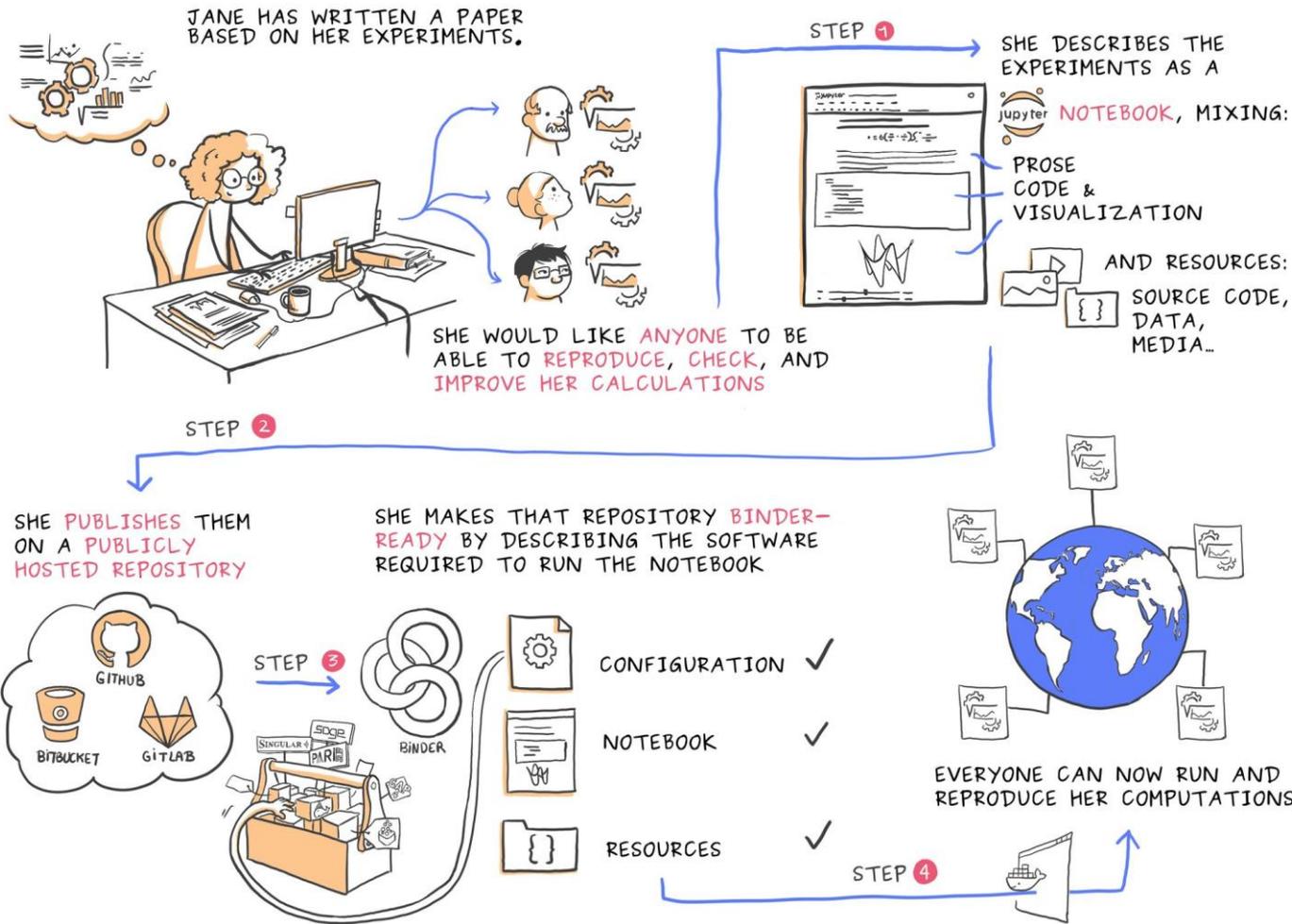
Binder Team

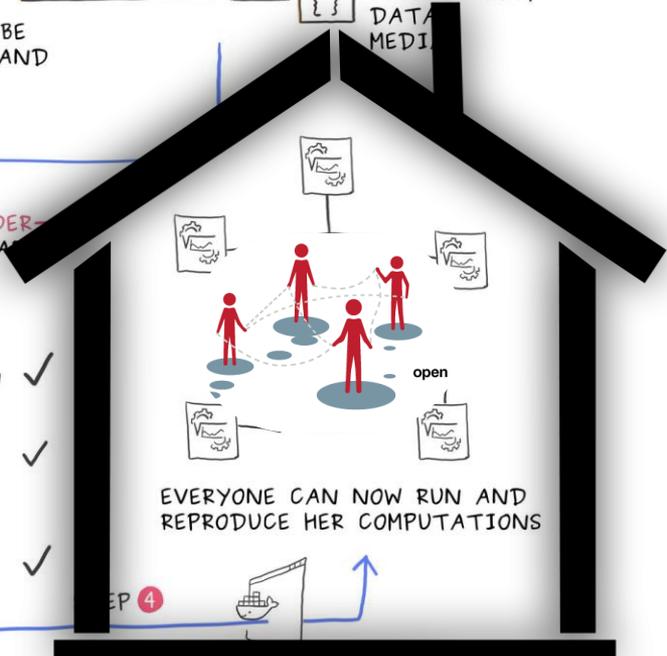
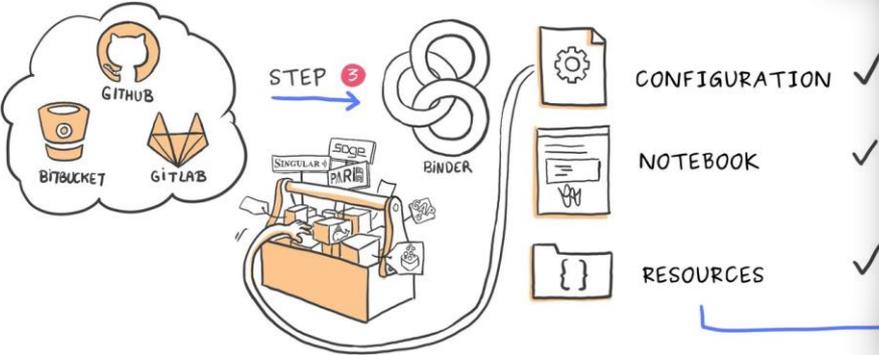
Binder's governance and team structure is defined in the [Binder Project Governance](#) page. Below we list the current team members of Binder.

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

 <p>Jessica Forde UC Berkeley team red</p> 	 <p>Tim Head Wild Tree Tech team red</p>	 <p>Lindsey Heagy UC Berkeley team blue</p> 	 <p>Chris Holdgraf Berkeley Institute for Data Science team red</p> 
 <p>M Pacer Netflix team blue</p>	 <p>Yuvi Panda UC Berkeley team red</p> 	 <p>Min Ragan-Kelley Simula team lead</p> <p>data, </p>	 <p>Zach Sailer Project Jupyter team blue</p> 
 <p>Erik Sundell</p>	 <p>Carol Will-</p>		

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



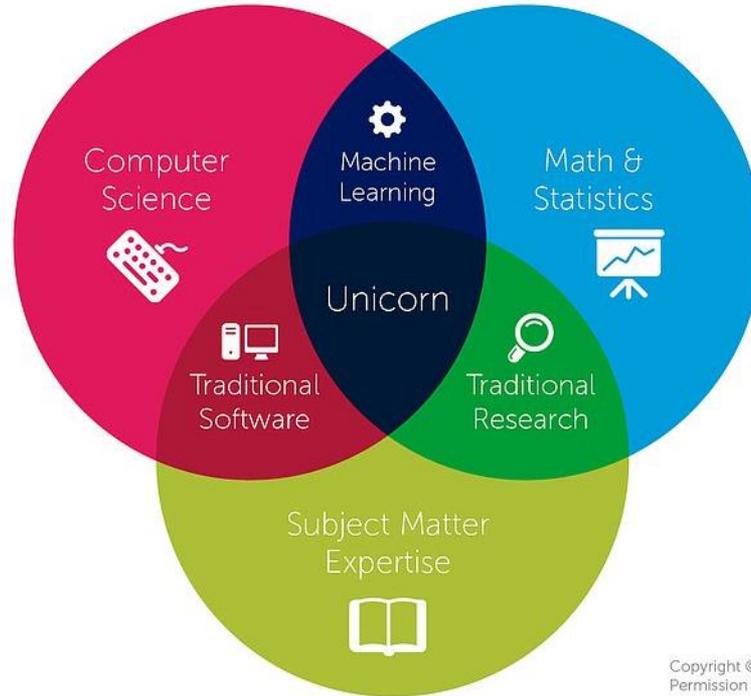




Building a culture of collaborative science

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

The Data Science Unicorn



Copyright © 2014 by Steven Geringer Raleigh, NC.
Permission is granted to use, distribute, or modify this image,
provided that this copyright notice remains intact.

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

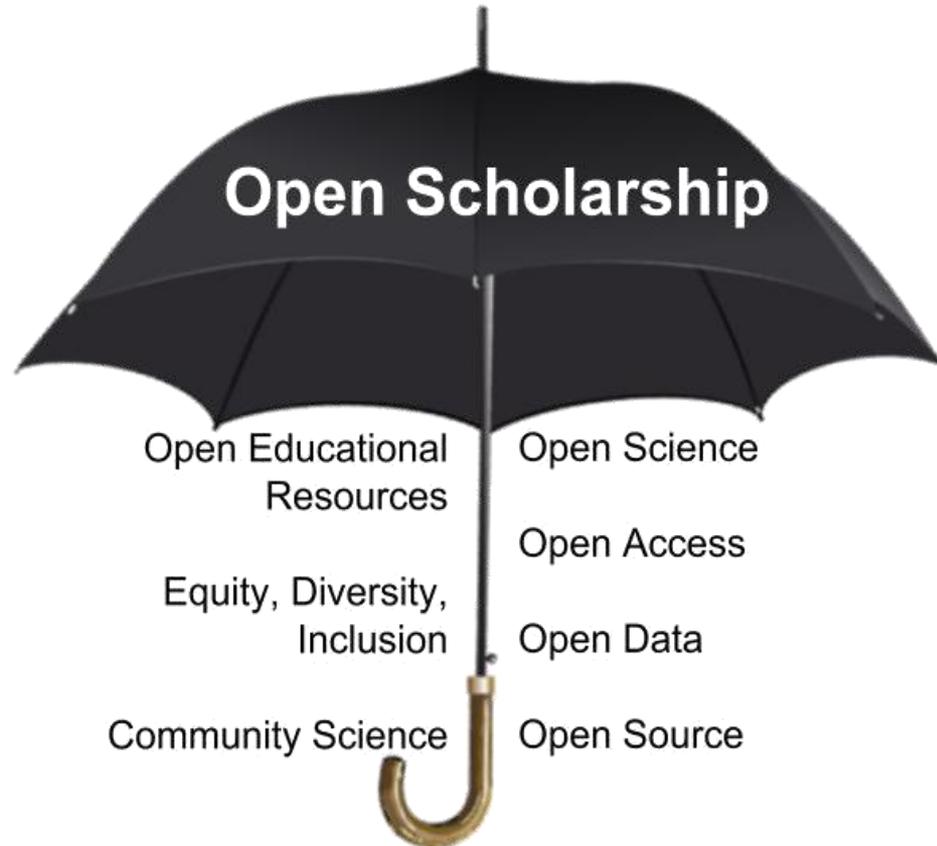


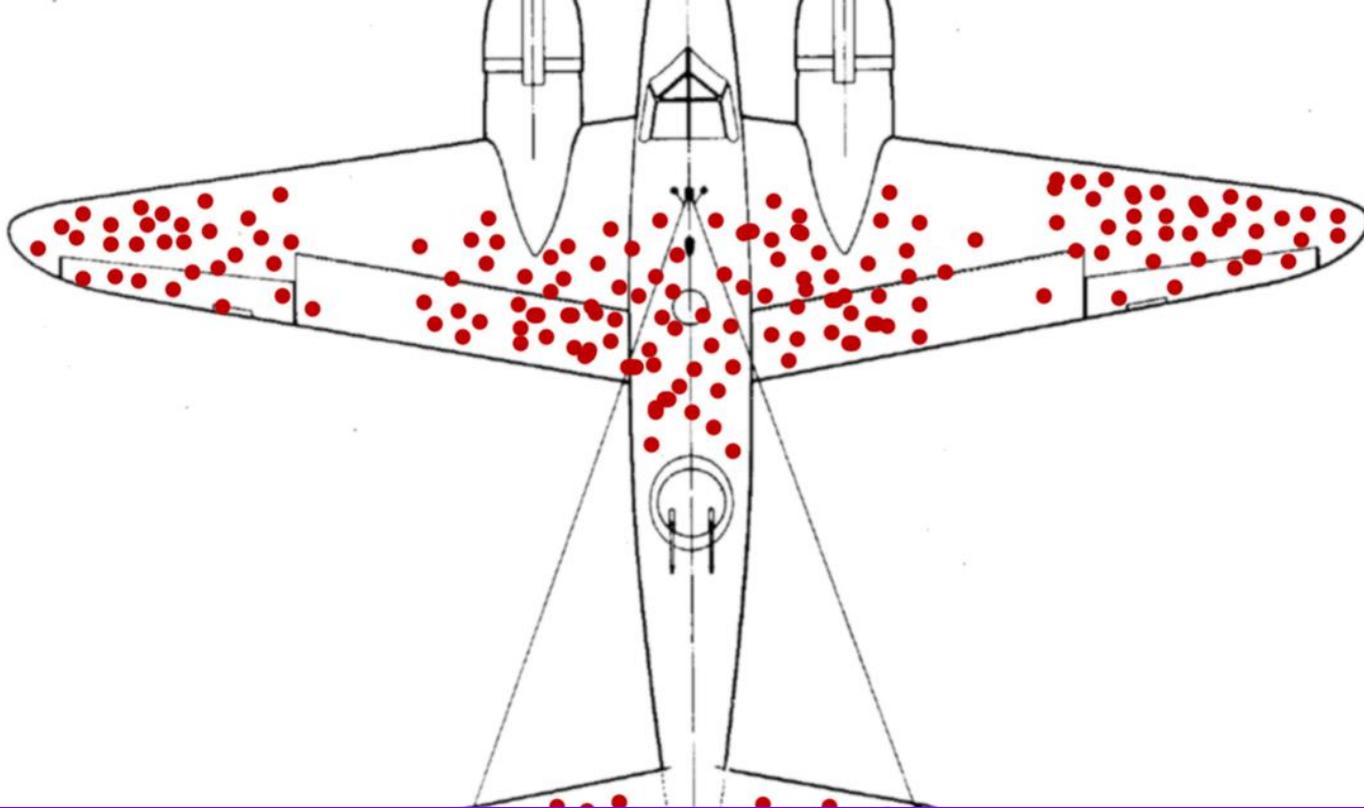
*How can we incentivise
team science?*

<https://neurohackademy.org/apply>

30/01/2019

Open is so much more than reproducible

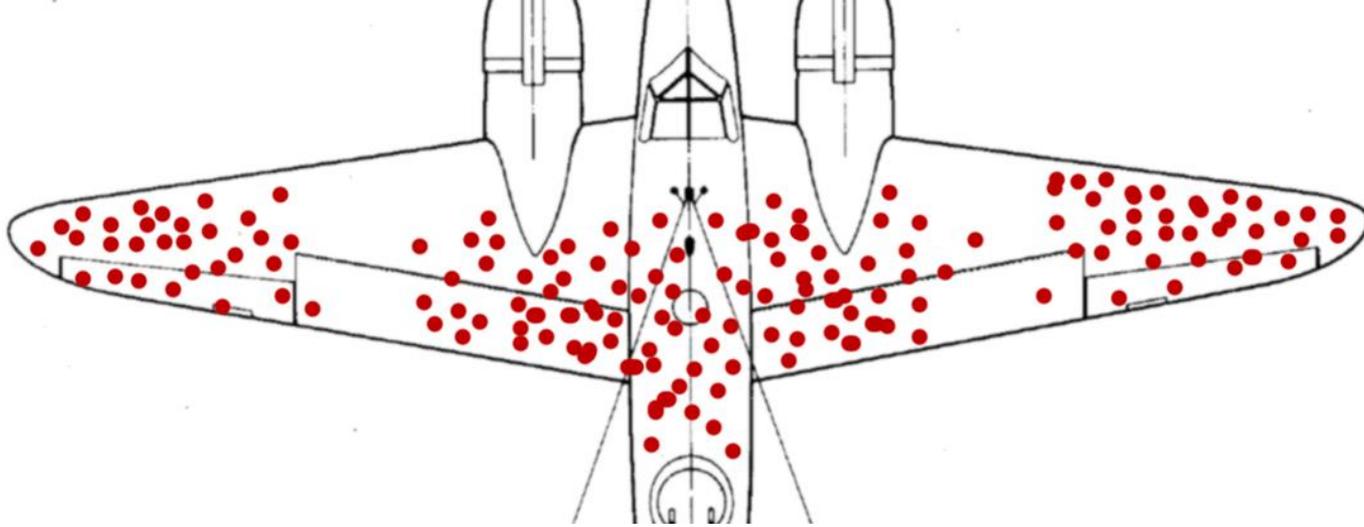




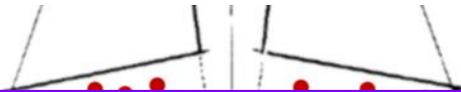
<https://medium.com/@penguinpress/an-excerpt-from-how-not-to-be-wrong-by-jordan-ellenberg-664e708cfc3d>

<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>

moz://a



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!



The Alan Turing Institute

Out and About in STEM

Legal information to support global mobility of LGBT+ individuals in STEM



<https://stemgamechangers.github.io>

Data science at scale



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://doi.org/10.6084/m9.figshare.7649156)