


**The
Alan Turing
Institute**

The Turing Way

**Reproducible, Inclusive,
Collaborative Data Science**

Kirstie Whitaker

Pronouns: she/her



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

The Turing Way is:

- a book
- a community
- a global collaboration
- a whole tonne of work



Rachael Ainsworth



Becky Arnold



Louise Bowler



Sarah Gibson



Patricia Herterich



James Hetherington



Rosie Higman



Anna Krystalli



Catherine Lawrence



Alex Morley



Martin O'Reilly

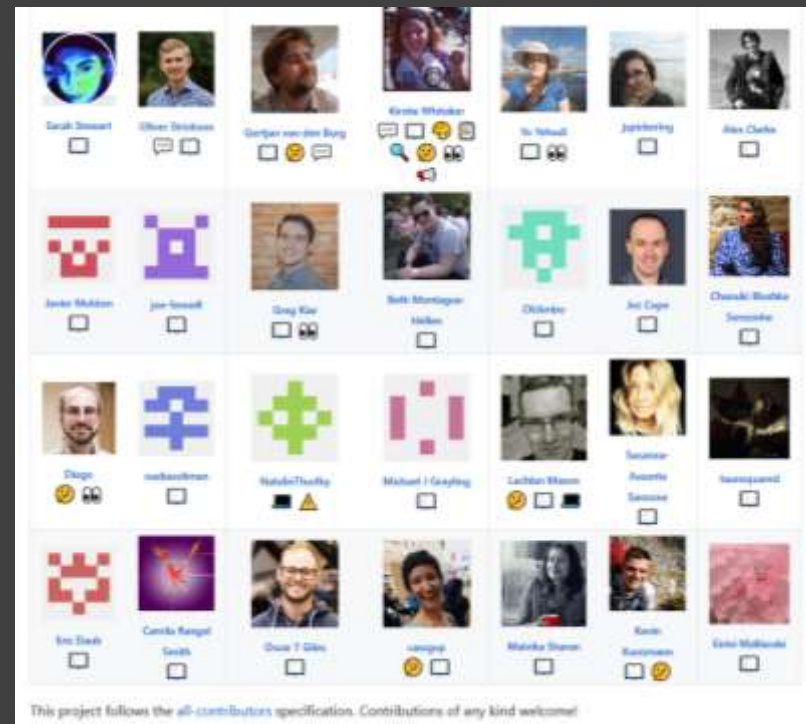


Malvika Sharan

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Thank you to all our contributors




This project follows the [all-contributors](https://allcontributors.org/) specification. Contributions of any kind welcome!

<https://github.com/alan-turing-institute/the-turing-way#contributors>
<https://allcontributors.org/docs/en/emoji-key>

#PyDataCambridge #TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.3543478>

An introduction to me



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



Picture credit: Chris Gorgolewski
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Academic errors 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-rogo-off-i-think-its-best-to-admit-your-errors-and-go-on-from-there>
<https://www.bbc.co.uk/news/magazine-22223190>

#PyDataCambridge #TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.3543478>

Academic errors 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)	

BBC NEWS

Home UK World Business Politics Tech Science Health Family & Education

Magazine

Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Ruth Alexander
BBC News

© 20 April 2013

This week, economists have been astonished to find that a famous academic paper often used to make the case for austerity cuts contains major errors. Another surprise is that the mistakes, by two eminent Harvard professors, were spotted by a student doing his homework.

It's 4 January 2010, the Marriott Hotel in Atlanta. At the annual meeting of the American Economic Association, Professor Carmen Reinhart and the former chief economist of the International Monetary Fund, Ken Rogoff, are presenting a research paper called *Growth in a Time of Crisis*.



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

#PyDataCambridge #TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.3543478>

The humans are the
hardest part of
reproducibility



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional
skills

Plead the 5th

Support additional
users

Takes time

<https://doi.org/10.6084/m9.figshare.5537101>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

The Turing Institute



<https://www.turing.ac.uk/news/enigma-machine-goes-display-alan-turing-institute>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



<https://www.bbc.co.uk/programmes/p0704h04>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



<https://bletchleypark.org.uk>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

University network



THE UNIVERSITY
of EDINBURGH



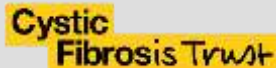
#PyDataCambridge #TuringWay @kirstie_

<https://doi.org/10.5281/zenodo.3543478>

The Institute's partners and collaborators



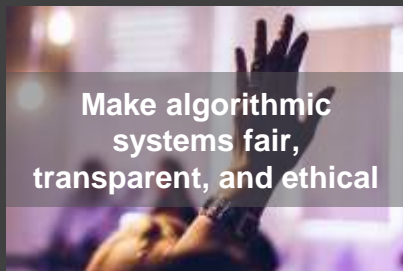
Lloyd's Register
Foundation



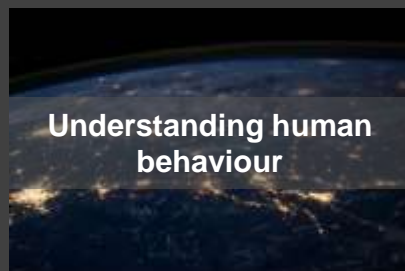
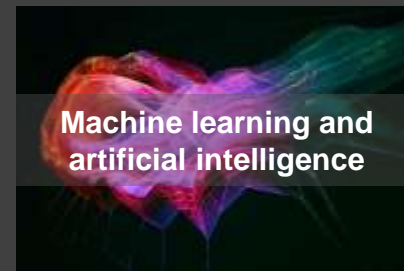
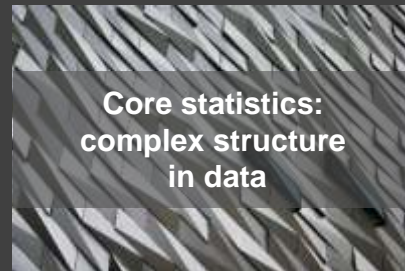
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Challenges

Advance data science and artificial intelligence to...



Core capabilities



Martin O'Reilly

“Make reproducible
research too easy not to
do.”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Martin O'Reilly

“Make reproducible
research too easy not to
do.

Do you need a biscuit?”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Martin O'Reilly

“Make reproducible research too easy not to do.

Do you need a biscuit?

If we can't do it here, we can't do it at all.”



<https://www.turing.ac.uk/people/researchers/martin-oreilly>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

The Turing Way



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment

Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the “responsibility of reproducibility” they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

<https://the-turing-way.netlify.com/introduction/introduction>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment



Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

<https://the-turing-way.netlify.com/introduction/introduction>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment



Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

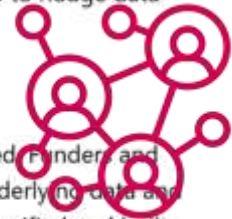
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors



<https://the-turing-way.netlify.com/introduction/introduction>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional
skills

Plead the 5th

Support additional
users

Takes time

<https://doi.org/10.6084/m9.figshare.5537101>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Catherine Lawrence

“We should ensure all our processes for running programmes are FAIR.

- Findable (intranet)
- Accessible (EDI)
- Interoperable across programmes and projects
- Reusable (bus factor)”

<https://www.turing.ac.uk/people/business-team/catherine-lawrence>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>



Testing for research



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Is your code doing what you think its doing?



<https://the-turing-way.netlify.com/testing/testing.html>

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

#PyDataCambridge #TuringWay @kirstie_

<https://doi.org/10.5281/zenodo.3543478>

Is your code doing what you think its doing?



Is your code doing what
you think its doing?

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



Is your code doing what
you think its doing?

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



Louise Bowler

“Add a test before you
change anything.”



<https://www.turing.ac.uk/people/researchers/louise-bowler>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Louise Bowler

“Add a test before you change anything.

Particularly if you’re just going to tidy up your code before sharing it.”

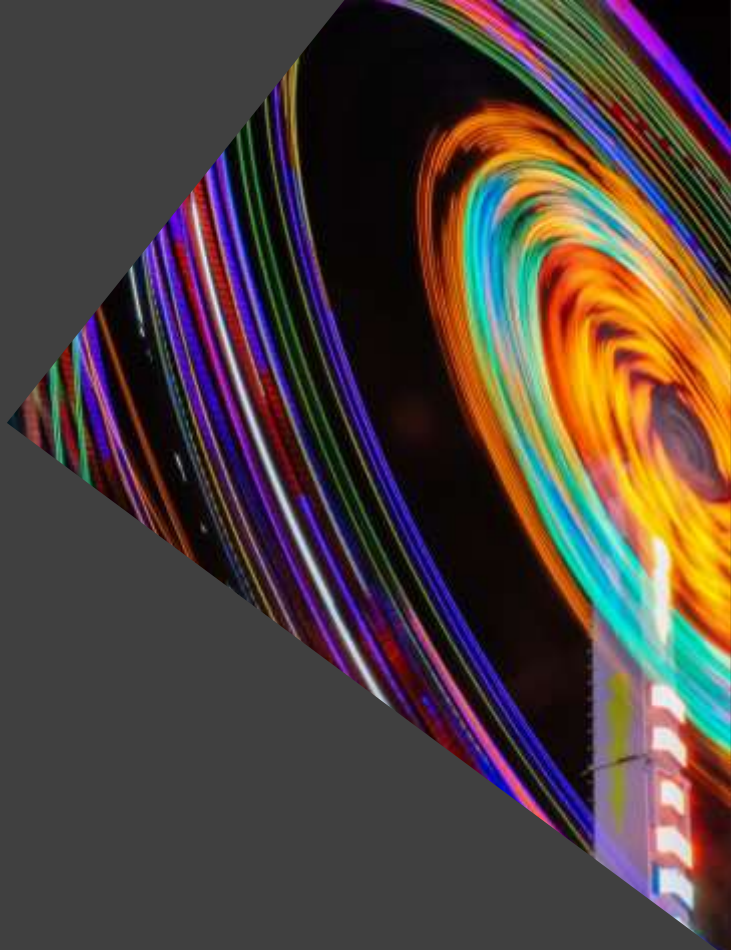


<https://www.turing.ac.uk/people/researchers/louise-bowler>

#PyDataCambridge #TuringWay @kirstie_j

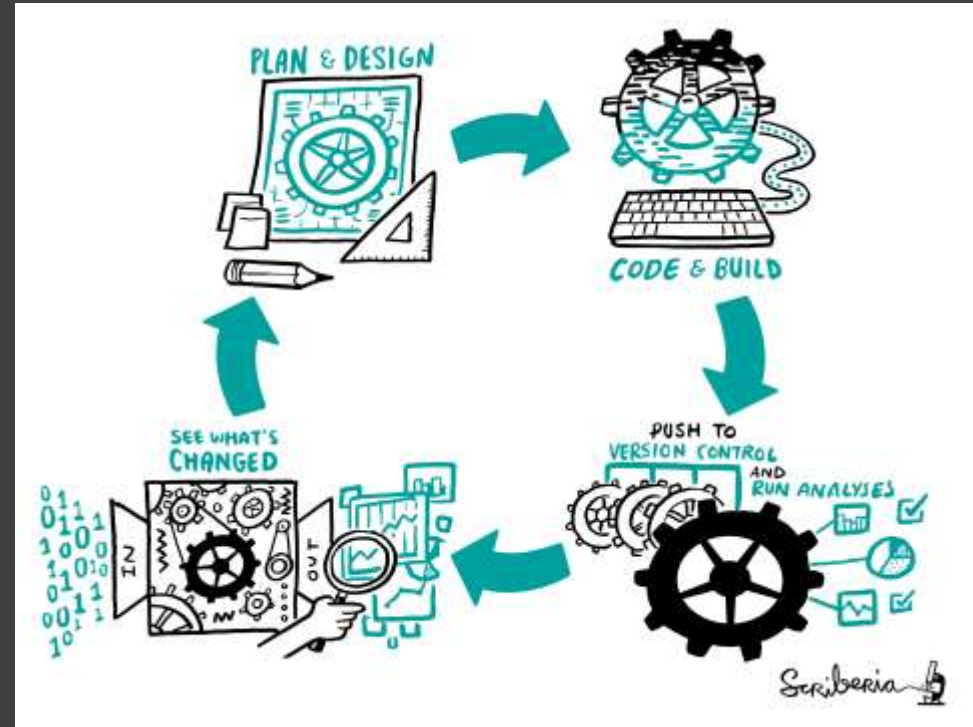
<https://doi.org/10.5281/zenodo.3543478>

Continuous Analysis

An abstract graphic in the top right corner featuring vibrant, multi-colored light trails (red, orange, yellow, green, blue, purple) that swirl and curve against a dark background, resembling a long-exposure photograph of a city at night or a data visualization.

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

- Plan and design your experiment
- Write down those steps in code
- Push to version control and run the analyses
 - Traditionally done on the cloud, but the important part is that all steps are run every time
- Test to see what's changed



The screenshot shows the GitHub interface for the repository 'alan-turing-institute / signatures-psychiatry'. The top navigation bar includes links for Dashboard, Changelog, Documentation, and Help. A search bar on the left allows searching across all repositories. The main content area displays the repository name and a 'build' button. Below this, tabs for 'Current', 'Branches', 'Build History', and 'Pull Requests' are visible. The 'Current' tab is active, showing a build status for 'lab-add-synth-data' with a green checkmark and the message 'Add travis config'. It indicates that 1 test passed, the build ran for 1 minute and 41 seconds, and was completed about 12 hours ago. The commit hash is 023d957, and the branch is 'lab-add-synth-data'. The build was initiated by Louise Bowler using Python 2.7. On the left sidebar, under 'My Repositories', three other repositories are listed: 'alan-turing-institute/Posterior' (failed, 98 tests), 'alan-turing-institute/signature' (passed, 1 test), and 'bids-standard/bids-specificat' (passed, 506 tests).

Search all repositories

alan-turing-institute / signatures-psychiatry

build unknown

Current Branches Build History Pull Requests More options

✓ lab-add-synth-data Add travis config #1 passed Restart build

Commit 023d957

Compare e63a607...023d957

Branch lab-add-synth-data

Louise Bowler

Python: 2.7

Duration: 2 hrs 11 min 35 sec
Finished: about 9 hours ago

✓ alan-turing-institute/signature #1

Duration: 1 min 41 sec
Finished: about 12 hours ago

✓ bids-standard/bids-specificat #506

Duration: 32 sec
Finished: a day ago

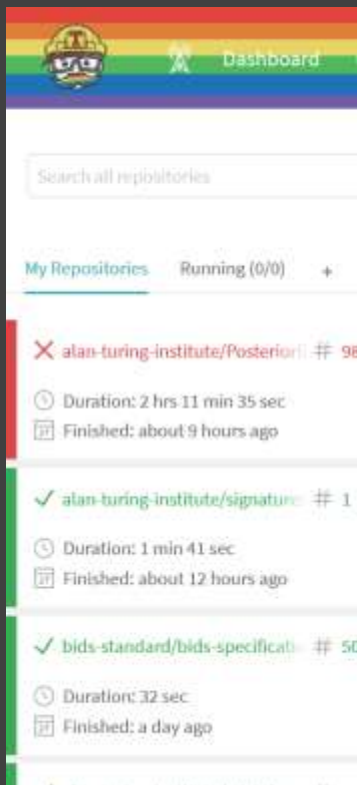
Job log View config

<https://github.com/alan-turing-institute/signatures-psychiatry>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

[https://the-turing-way.netlify.com/
continuous_integration/continuous_integration.html](https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html)



Dashboard

Search all repositories

My Repositories Running (0/0) +

✗ alan-turing-institute/Posterior # 98

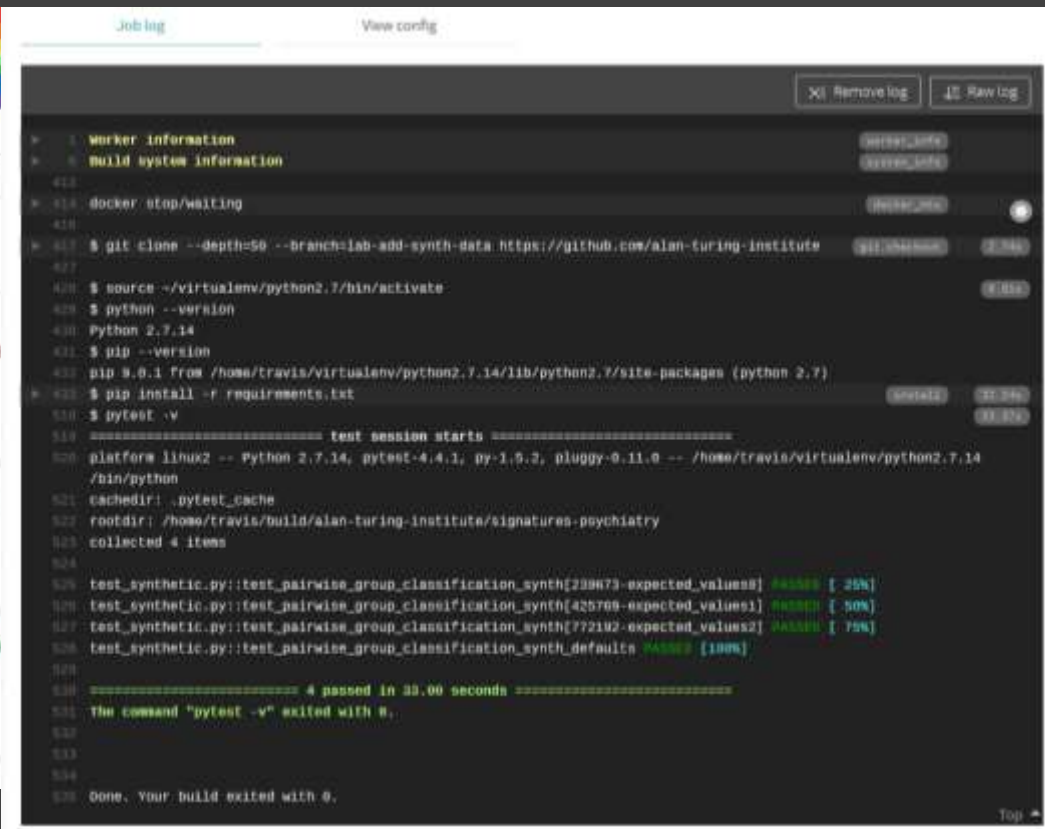
⌚ Duration: 2 hrs 11 min 35 sec
📅 Finished: about 9 hours ago

✓ alan-turing-institute/signature # 1

⌚ Duration: 1 min 41 sec
📅 Finished: about 12 hours ago

✓ bids-standard/bids-specificat # 50

⌚ Duration: 32 sec
📅 Finished: a day ago



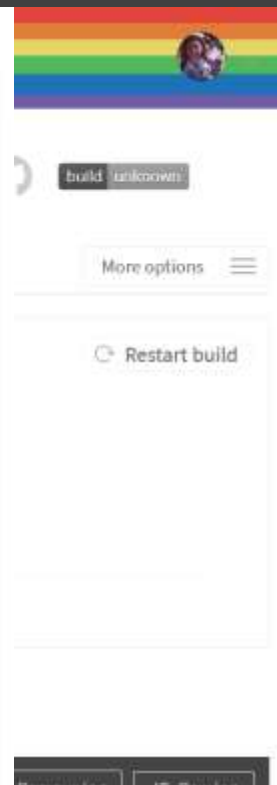
Job log View config

Remove log Raw log

```

> Worker information
> Build system information
412
413 docker stop/waiting
414
415 $ git clone --depth=50 --branch=lab-add-synth-data https://github.com/alan-turing-institute
416
417 $ source ~/virtualenv/python2.7/bin/activate
418 $ python --version
419 Python 2.7.14
420
421 $ pip --version
422 pip 9.0.1 from /home/travis/virtualenv/python2.7.14/lib/python2.7/site-packages (python 2.7)
423 $ pip install -r requirements.txt
424
425 $ pytest -v
426
427 ===== test session starts =====
428 platform linux2 -- Python 2.7.14, pytest-4.4.1, py-1.8.2, pluggy-0.11.0 -- /home/travis/virtualenv/python2.7.14/bin/python
429 cachedir: .pytest_cache
430 rootdir: /home/travis/build/alan-turing-institute/signatures-psychiatry
431 collected 4 items
432
433 test_synthetic.py::test_pairwise_group_classification_synth[239673-expected_values8] PASSED [ 25%]
434
435 test_synthetic.py::test_pairwise_group_classification_synth[425709-expected_values1] PASSED [ 50%]
436
437 test_synthetic.py::test_pairwise_group_classification_synth[772102-expected_values2] PASSED [ 75%]
438
439 test_synthetic.py::test_pairwise_group_classification_synth_defaults PASSED [100%]
440
441 ===== 4 passed in 33.00 seconds =====
442
443 The command "pytest -v" exited with 0.
444
445
446 Done. Your build exited with 0.
  
```

Top



build unknown

More options

Restart build

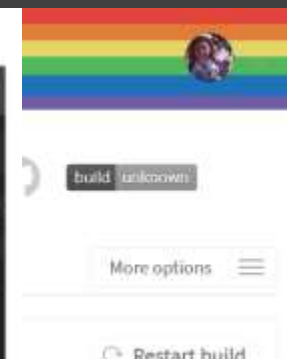
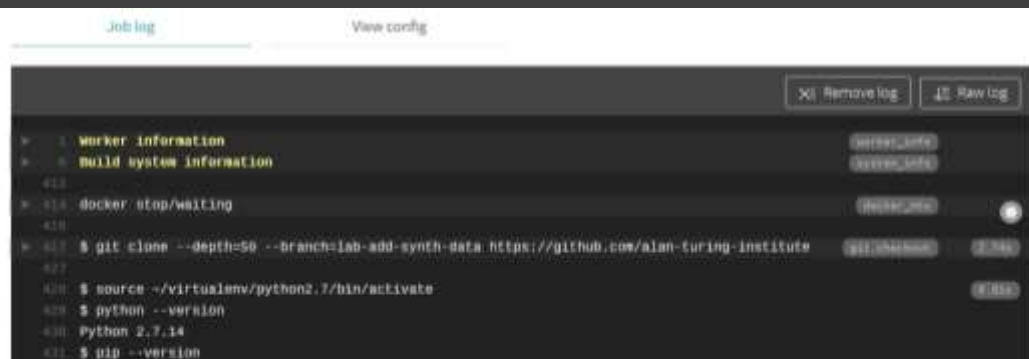
Cancel build Force build

<https://github.com/alan-turing-institute/signatures-psychiatry>

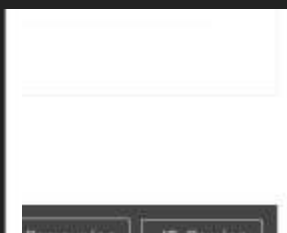
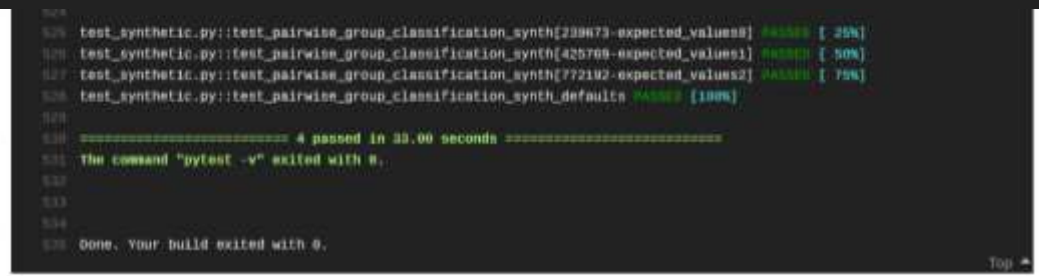
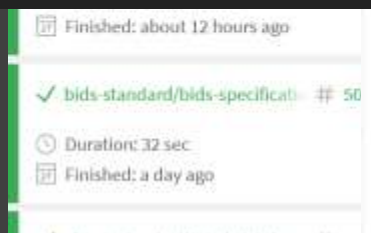
#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html



===== 4 passed in 33.00 seconds =====
 The command "pytest -v" exited with 0.



<https://github.com/alan-turing-institute/signatures-psychiatry>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html

- Run the analysis from start to finish as you work
- Many times tests will fail as expected: you're developing the analysis!
- Sometimes tests will fail unexpectedly
- CI makes you be explicit about what has changed



1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with

Make

12. Risk Assessment

Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors



<https://the-turing-way.netlify.com/introduction/introduction>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Becky Arnold

“There are a lot of things you need to know before you can jump into continuous integration.

Version control is a prerequisite for pretty much everything.”



<https://software.ac.uk/about/fellows/becky-arnold>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with
Make

12. Risk Assessment

Continuous integration

Prerequisite	Importance	Notes
Experience with the command line	Necessary	A tutorial on working via the command line can be found here
Version control	Necessary	See the chapter on this for more information
Testing	Very helpful	See the chapter on this for more information
Reproducible computational environments	Necessary	See the chapter on this for more information, particularly the sections on YAML files and containers

Table of contents

- Summary
- How this will help you/ why this is useful
 - What are continuous delivery and continuous deployment?
- What is Travis and how does it work?
- Setting up continuous integration with Travis
 - Basic steps

https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html

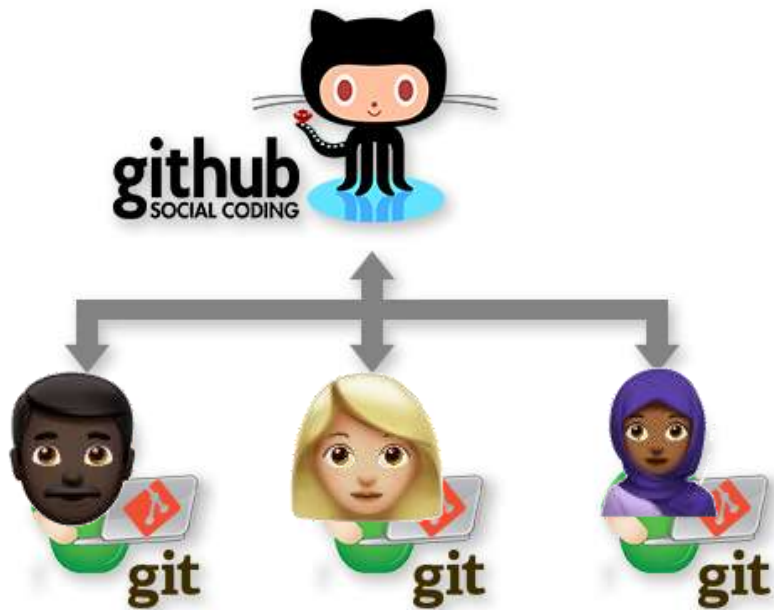
#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Version control



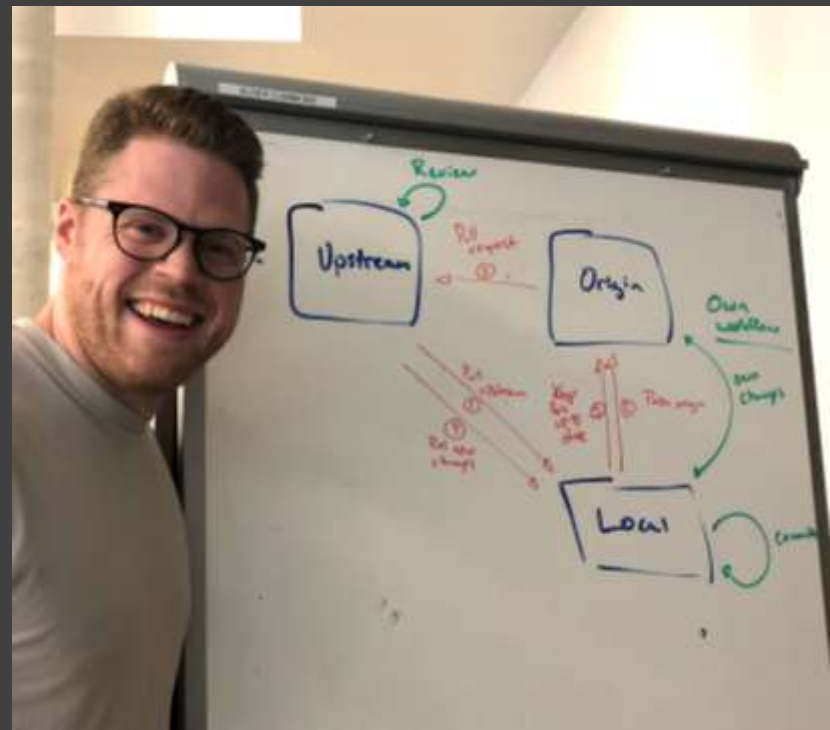
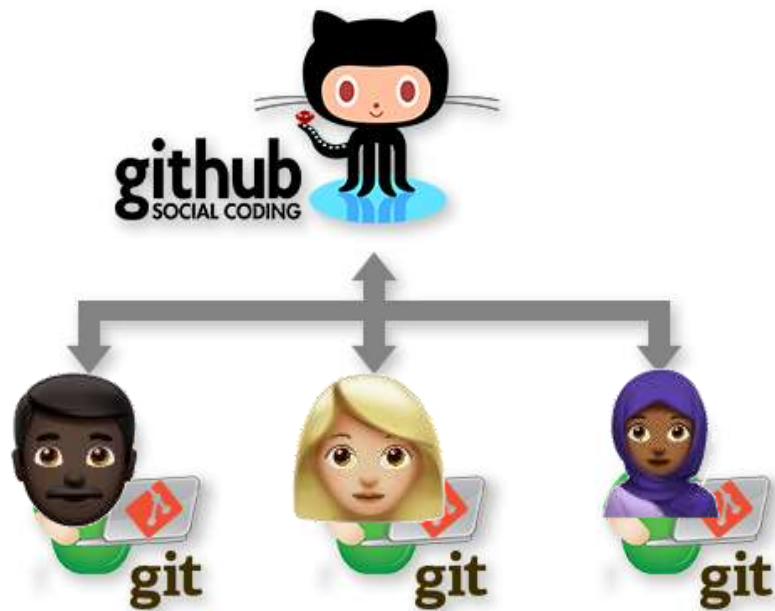
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html

https://the-turing-way.netlify.com/version_control/version_control.html #PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>



https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html

https://the-turing-way.netlify.com/version_control/version_control.html #PyDataCambridge #TuringWay @kirstie_j

<https://neurohackademy.org>

<https://doi.org/10.5281/zenodo.3543478>

Neurohackademy

“Every hackathon should have a gong that you can ring when you complete your first pull request.”



<https://neurohackademy.org>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

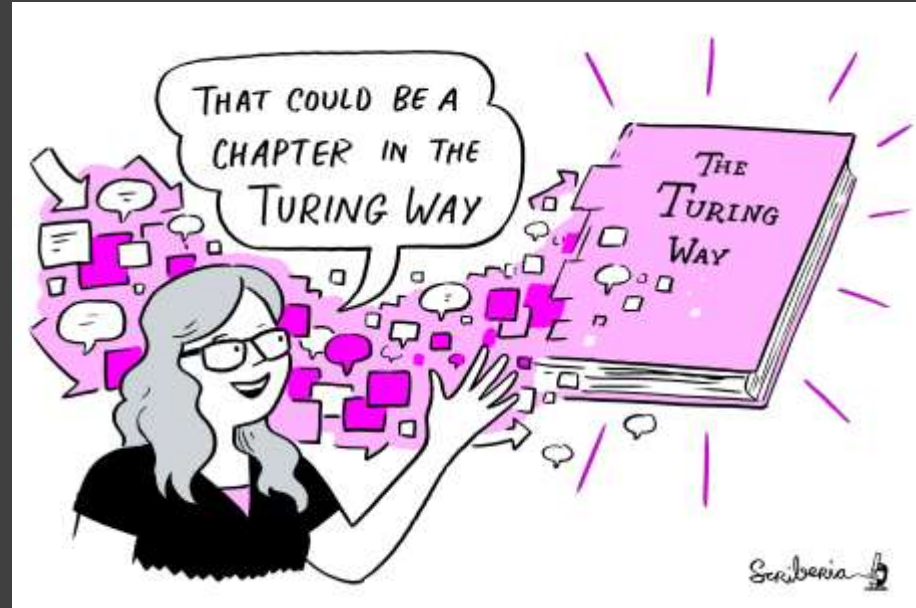
Beyond reproducibility



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

The expansion

- Create a handbook for all data science practices
 - Reproducibility
 - Scoping and designing a data science project
 - Communication and visualisation
 - Ethics
 - Collaborative working



[https://github.com/
alan-turing-institute/the-turing-way/
blob/master/project_management/
tps-funding-application-20190429.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/project_management/tps-funding-application-20190429.md)

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Scoping & designing a data science project



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

An algorithm is a set of instructions



An algorithm is a set of instructions



“AI is whatever hasn't been done yet.”

- No agreed upon definition
- There is a lot of hype
- Setting appropriate expectations in collaborations is really important and hard to do!



Communicating your work



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



Drawing by Juliette Taka: <https://twitter.com/mybinderteam/status/1082556317842264064>

Animation by Pix Video, via Open Dream Kit project:

<https://youtu.be/a5i42ISj-L4>

#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

- Leave PDFs behind
- Share the responsibility of reproducibility with busy PIs
- Requires version control, capturing environment and new build for each change



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Sarah Gibson

“It took me a while to feel like I knew enough to contribute to Binder. But the team are always so excited to have my input. Its really motivating to be part of such a welcoming community.”



<https://www.turing.ac.uk/people/researchers/sarah-gibson>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



<https://chan Zuckerberg.com/rfa/essential-open-source-software-for-science>


<https://medium.com/@cziscience/>

[the-invisible-foundations-of-biomedicine-4ab7f8d4f5dd](#)

#PyDataCambridge #TuringWay @kirstie_j


<https://doi.org/10.5281/zenodo.3543478>

Ethical research



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Open Scholarship



Open Educational
Resources

Equity, Diversity,
Inclusion

Community Science

Open Science

Open Access

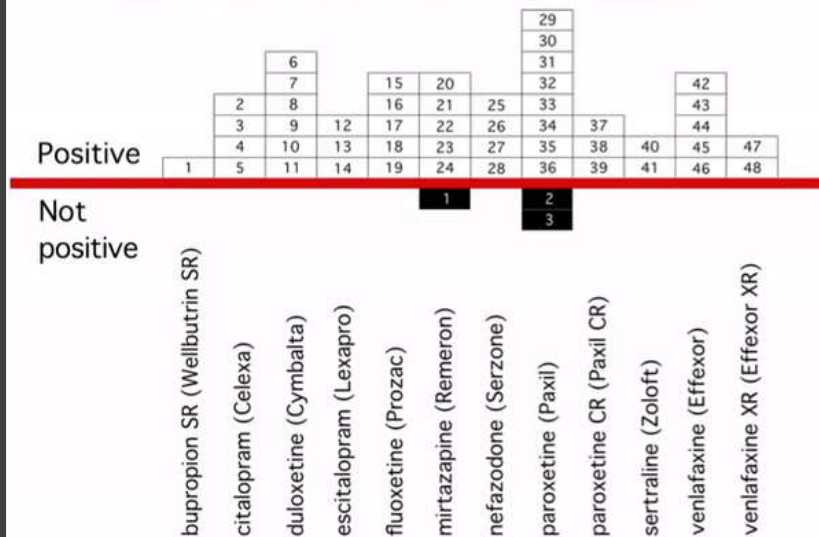
Open Data

Open Source

Ethical data science

– Transparent publishing

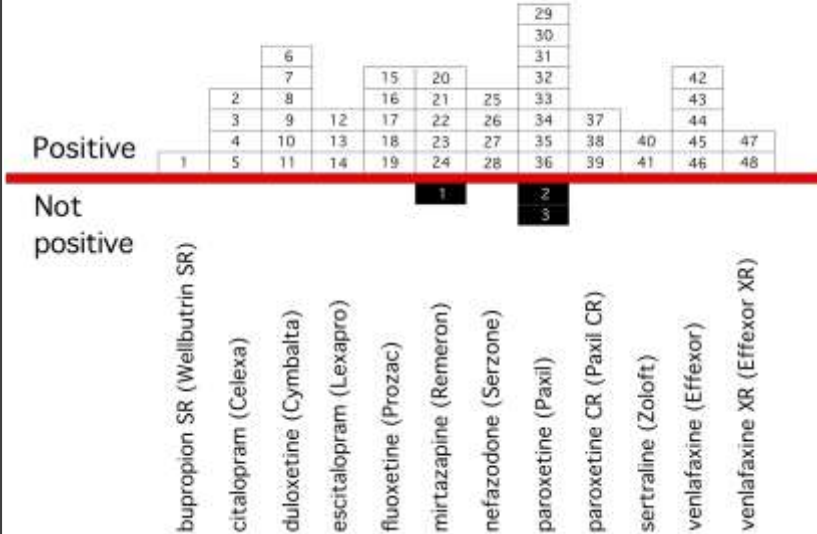
Journal version of antidepressant trials



Ethical data science

– Transparent publishing

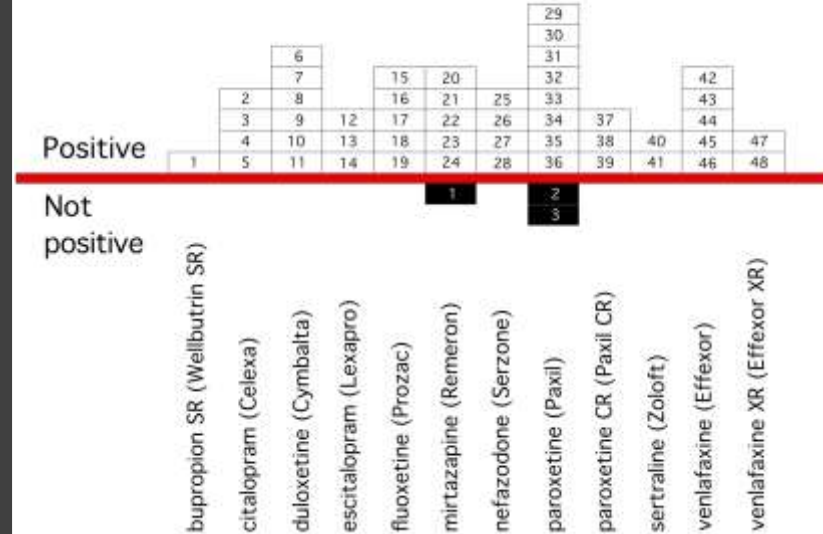
Journal version of antidepressant trials

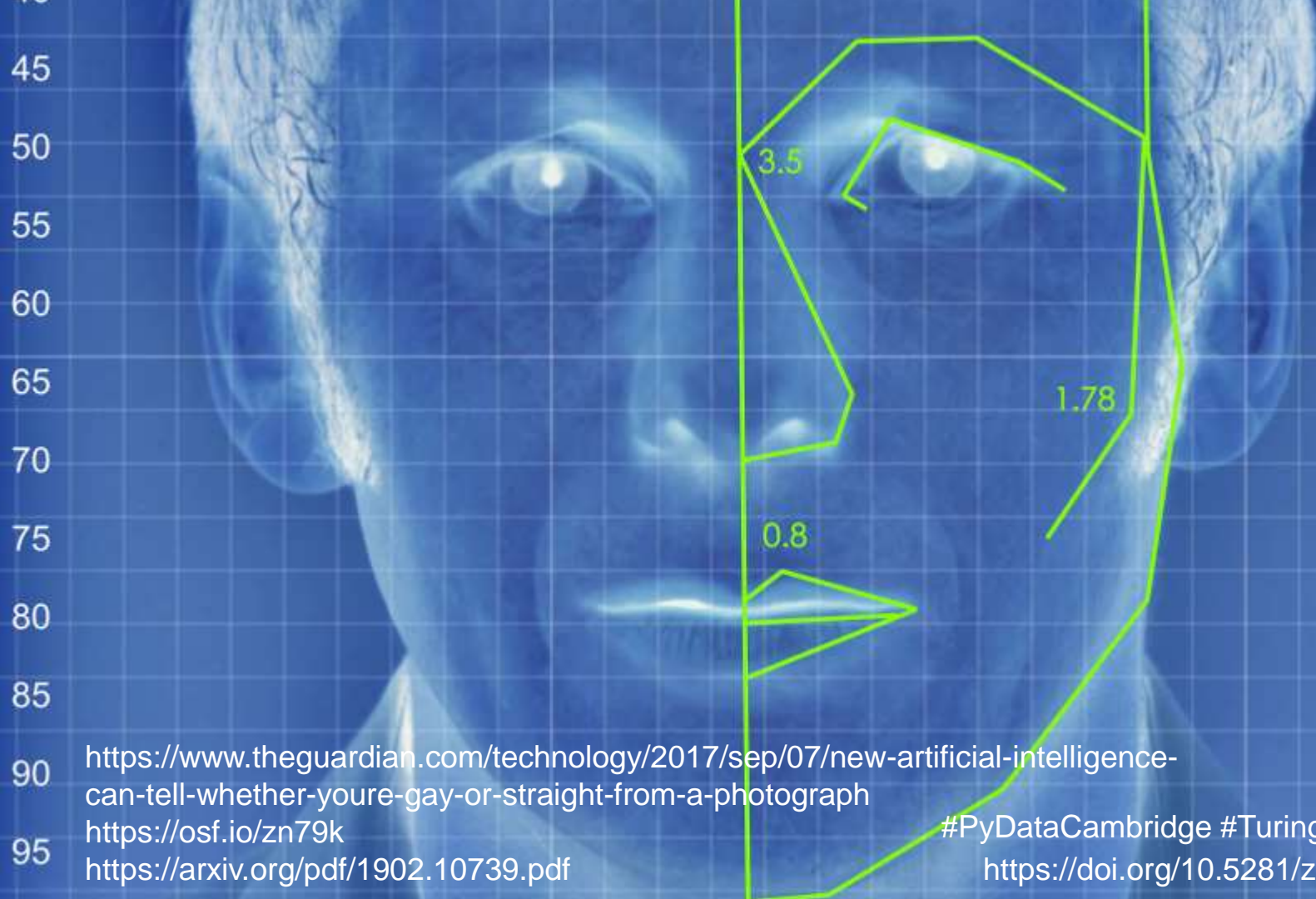


Ethical data science

- Transparent publishing
- Consent (GDPR)
- Privacy (safe haven)
- Societal harm

Journal version of antidepressant trials





<https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph>

<https://osf.io/zn79k>

<https://arxiv.org/pdf/1902.10739.pdf>

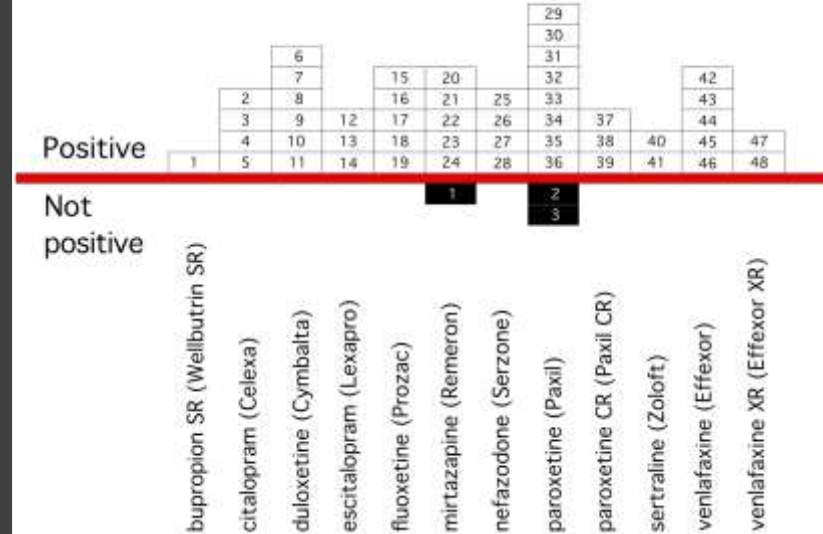
#PyDataCambridge #TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.3543478>

Ethical data science

- Transparent publishing
- Consent (GDPR)
- Privacy (safe haven)
- Societal harm
- Wellbeing of the research team

Journal version of antidepressant trials





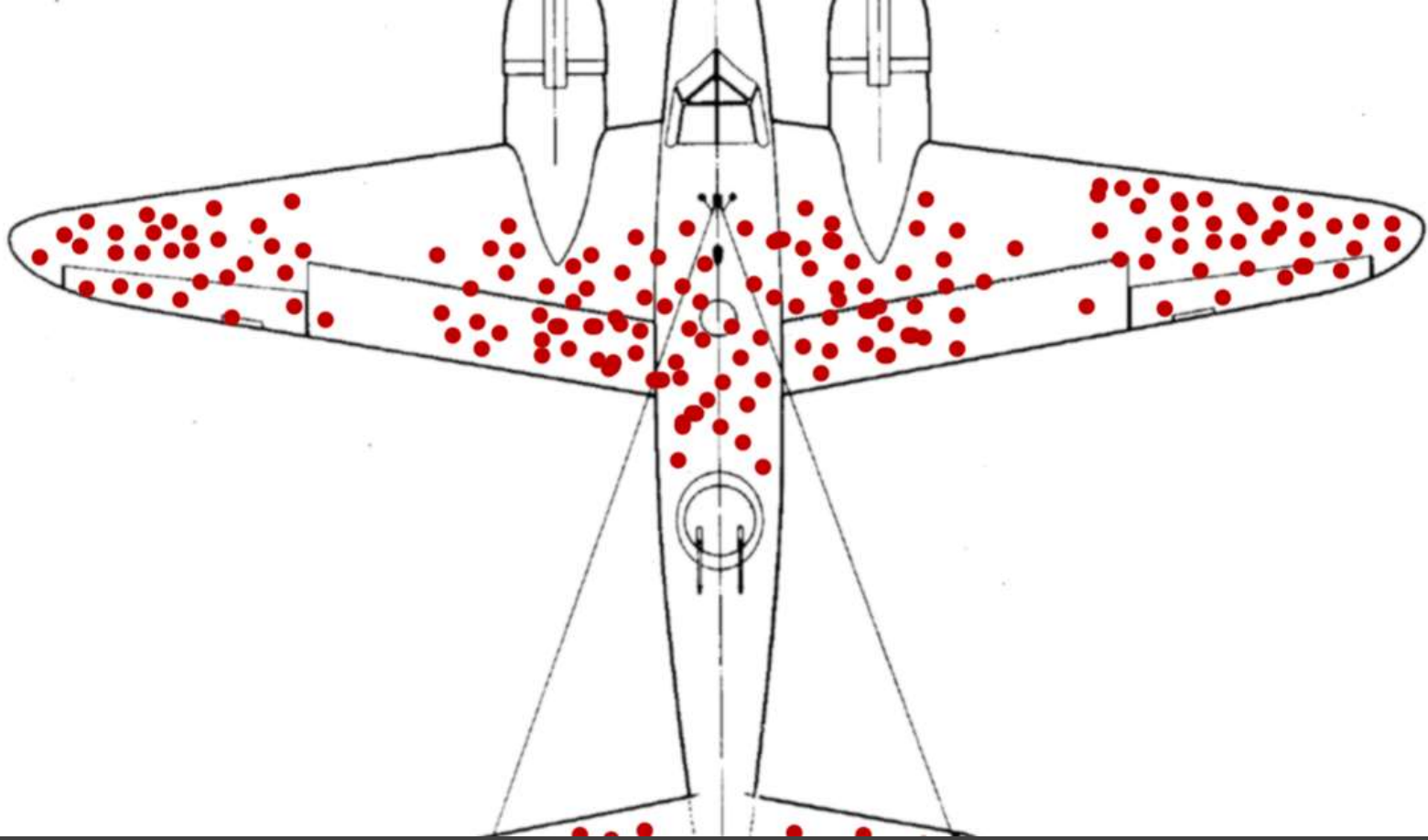
<https://www.theverge.com/2019/2/25/18229714/cognizant-facebook-content-moderator-interviews-trauma-working-conditions-arizona>

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Collaboration

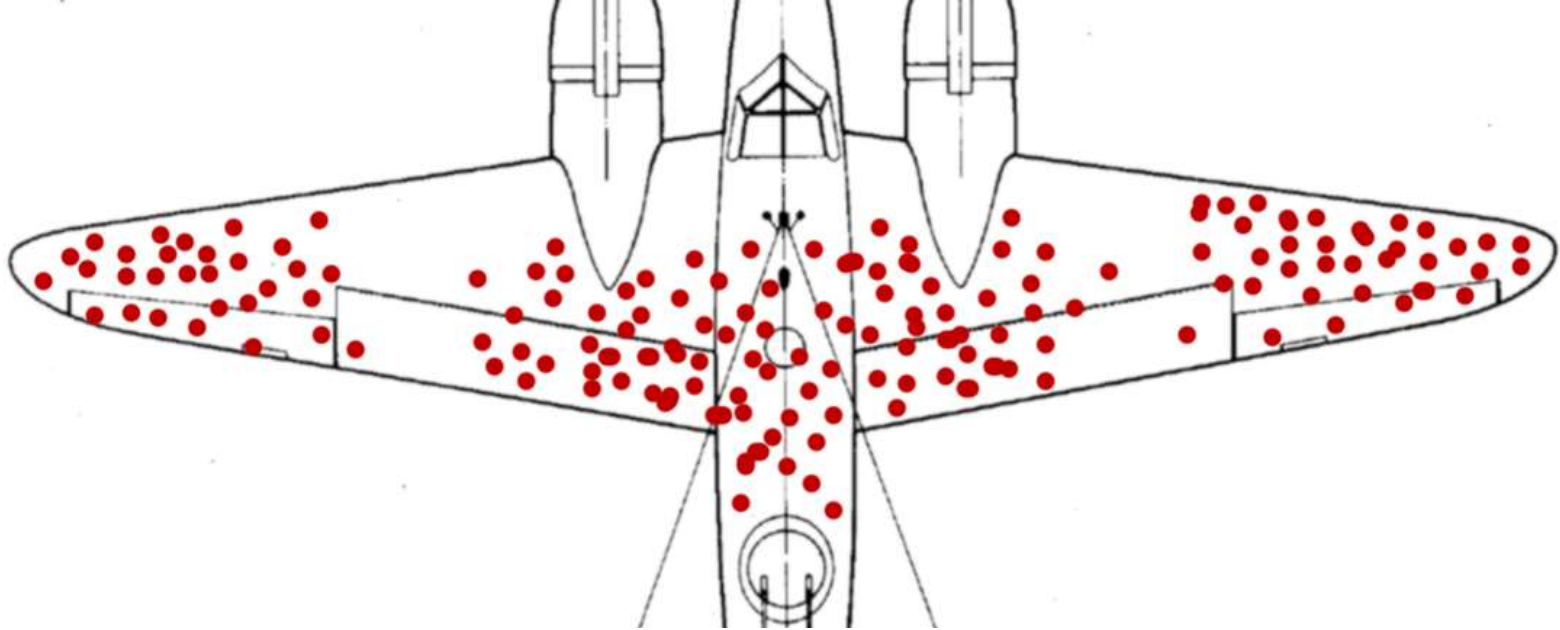


#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



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

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>



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



Whose voices are missing?

- In this room
- In your friend group
- At work and in your open source communities






<https://www.nhm.ac.uk/visit/wpy/gallery/2010/images/eric-hosking-portfolio-award/4372/a-marvel-of-ants.html>

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

A global collaboration



#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Patricia Herterich

“What really sets The Turing Way apart is HOW we’re writing the book. The focus on community, the commitment to transparency and working open right from the beginning is an exciting (and terrifying) new way of working.”



<https://rd-alliance.org/users/patricia-herterich>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

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



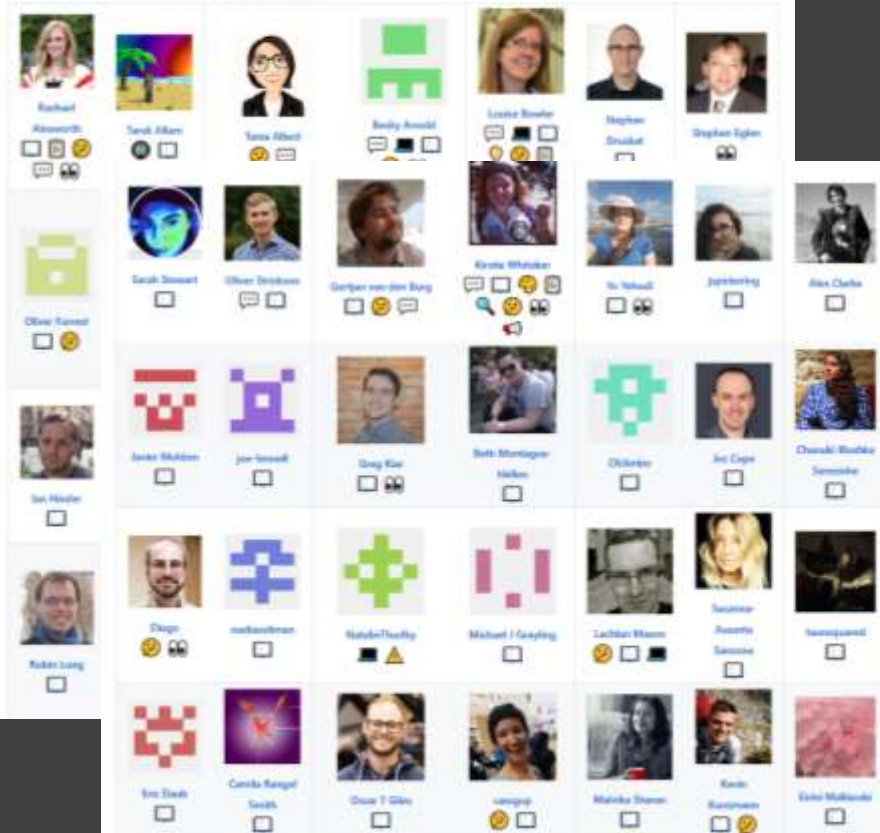
Participation & Inclusion

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




























moz://a

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Thanks goes to these wonderful people [[emoji key](#)]:



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

 question	 eventOrganizing	 review
 bug	 financial	 security
 blog	 fundingFinding	 tool
 business	 ideas	 translation
 code	 infra	 test
 content	 maintenance	 tutorial
 doc	 platform	 talk
 design	 plugin	 userTesting
 examples	 projectManagement	 video

<https://github.com/alan-turing-institute/the-turing-way#contributors>
<https://allcontributors.org/docs/en/emoji-key>

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Rosie Higman

“There’s no point in running events when you’re only preaching to the choir. We need to show researchers the selfish reasons to follow our recommendations.”



<https://rosiehigman.wordpress.com>
#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

Book Dashes

- Manchester and London
- 13 selected people to contribute to the book
- 1:3 support ratio: mentored support to contribute expertise



[https://github.com/
alan-turing-institute/the-turing-way/
blob/master/
workshops/book-dash/
book-dash-\[mcr|ldn\]-report.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/workshops/book-dash/book-dash-[mcr|ldn]-report.md)

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>

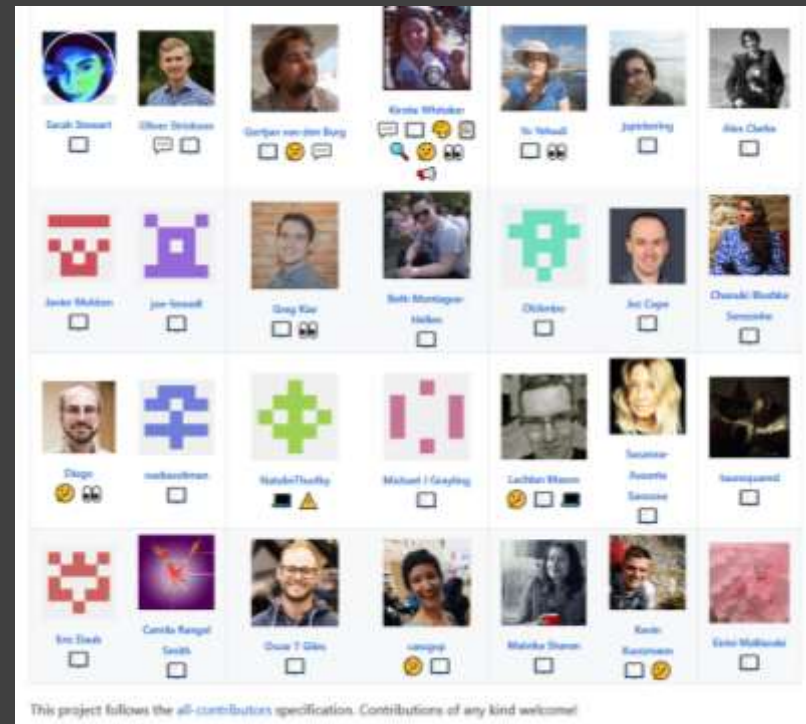
Collaboration cafes

- 1st and 3rd Wednesdays of each month
- All remote participation
 - Zoom call
 - Pomodoro technique
 - Breakout rooms for mentored contributions
- Everyone welcome





Thank you to current (& future) contributors



This project follows the [all-contributors](https://allcontributors.org/) specification. Contributions of any kind welcome!

<https://github.com/alan-turing-institute/the-turing-way#contributors>
<https://allcontributors.org/docs/en/emoji-key>

#PyDataCambridge #TuringWay @kirstie_
<https://doi.org/10.5281/zenodo.3543478>

Thank you

The
Alan Turing
Institute



- Book: <https://the-turing-way.netlify.com>
- Newsletter: <https://tinyletter.com/TuringWay>
- GitHub: <https://github.com/alan-turing-institute/the-turing-way>
- Chat: <https://gitter.im/alan-turing-institute/the-turing-way>
- Next Collaboration Café: 20 November at 7pm UK time
- This work was supported by The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "Tools, Practices and Systems" theme within that grant, and by The Alan Turing Institute under the EPSRC grant EP/N510129/1.
- Unsplash photos by Adolfo Felix, Austin Distel, Chris Ried, Dan Gold, Freddy Castro, James Pond, Kinson Leung, Mateo Vrbnjak, Med Badr, Michael Aleo, Mimi thian, Perry Grone, Rebecca Grant, Wei Ding. Noun Project icons by Luis Prado, Becris, Rose Alice Design, Hyemm.work.
- Original artwork by Scriberia:
<https://doi.org/10.5281/zenodo.3332807>

#PyDataCambridge #TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.3543478>