The Alan Turing Institute

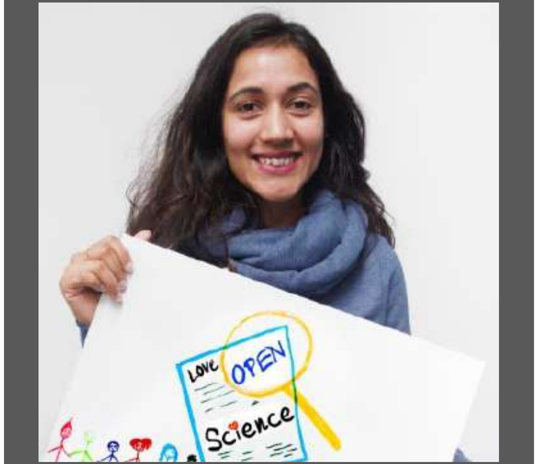
The Turing Way: Open Source and reproducibility aspects of data science

Malvika Sharan

Pronouns: she/her

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway, #TuringWay

About me



- PhD in Bioinformatics
- Open Source and Community Building
- Bio-IT Community Coordinator at EMBL HD (2016-2020)
- The Turing Way (2019 -)

Open Life Science

OPEN BIOINFORMATICS FOUNDATION

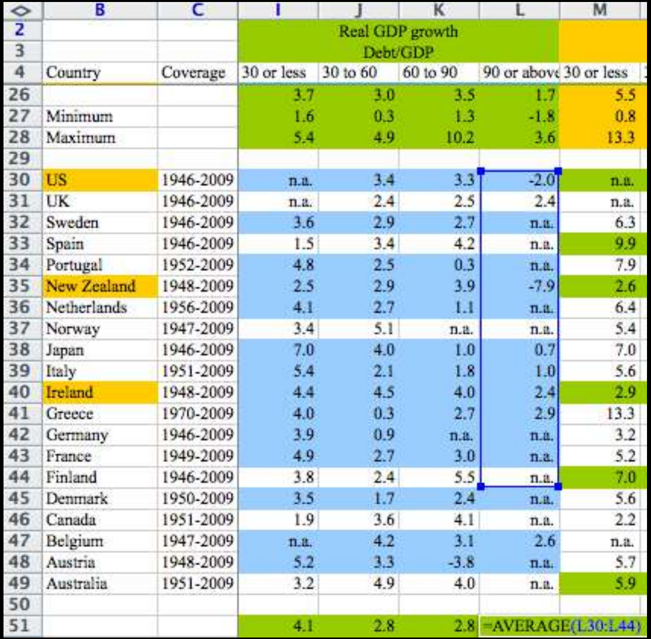
Software Sustainability Institute

THE CARPENTRIES

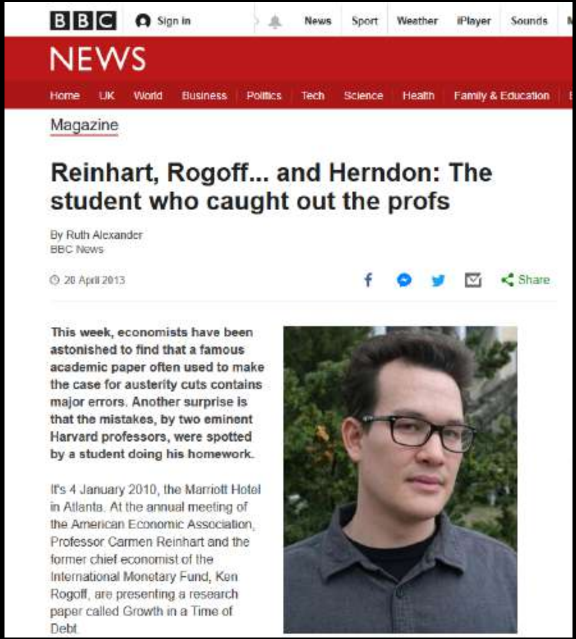
https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway, #TuringWay

1. **About *The Turing Way***
2. guide to computational reproducibility

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)	



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 Debt.

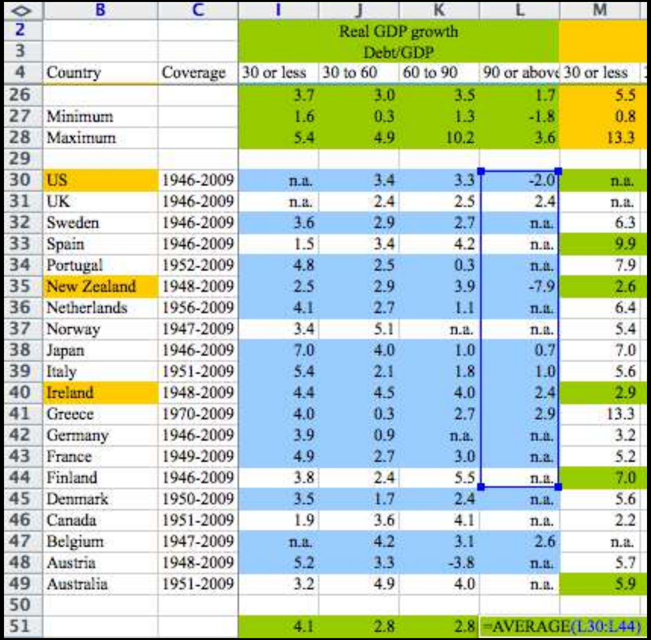
https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there

https://www.bbc.co.uk/news/magazine-22223190

@TuringWay @malvikasharan

Ref: https://doi.org/10.5281/zenodo.4139831

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)	

"…paper often used to make the case for austerity cuts contains major errors."

"… only included 15 of the 20 countries (of average GDP growth in countries with high public debt)."

"We will redouble our efforts to avoid such errors in the future …"

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

@TuringWay @malvikasharan
Ref: https://doi.org/10.5281/zenodo.4139831

Applying best practices in our research requires intention, resources, time and collaboration, which can be overwhelming.

IF I SHARE MY DATA.... PEOPLE MIGHT FIND MISTAKES
NEED TO MAKE IT OK TO BE HUMAN
SPRINGER NATURE

https://www.software.ac.uk/blog/2020-12-17-ten-arguments-against-open-science-you-can-win

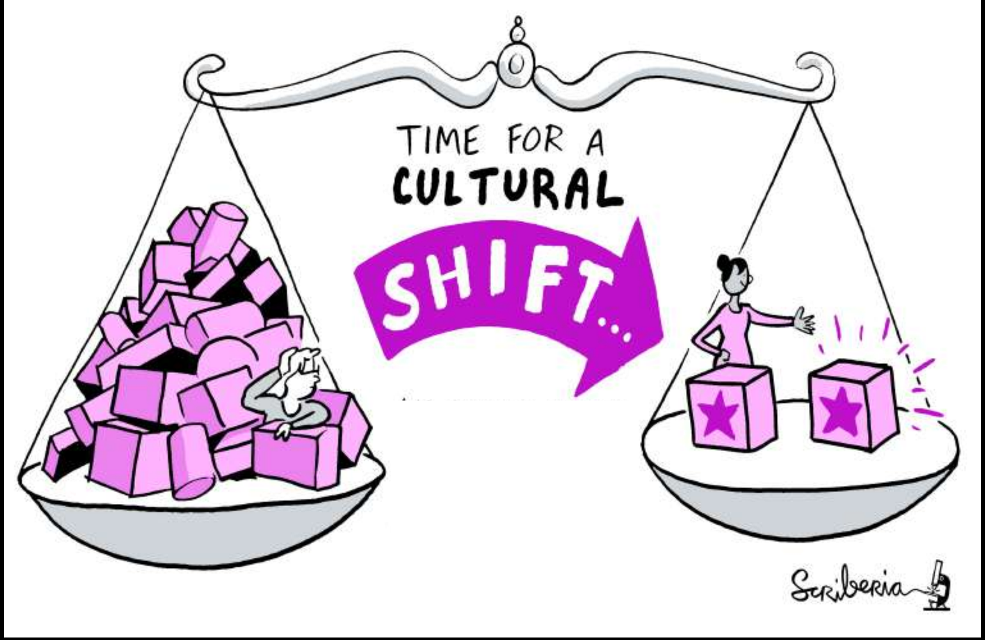
		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway, #TuringWay

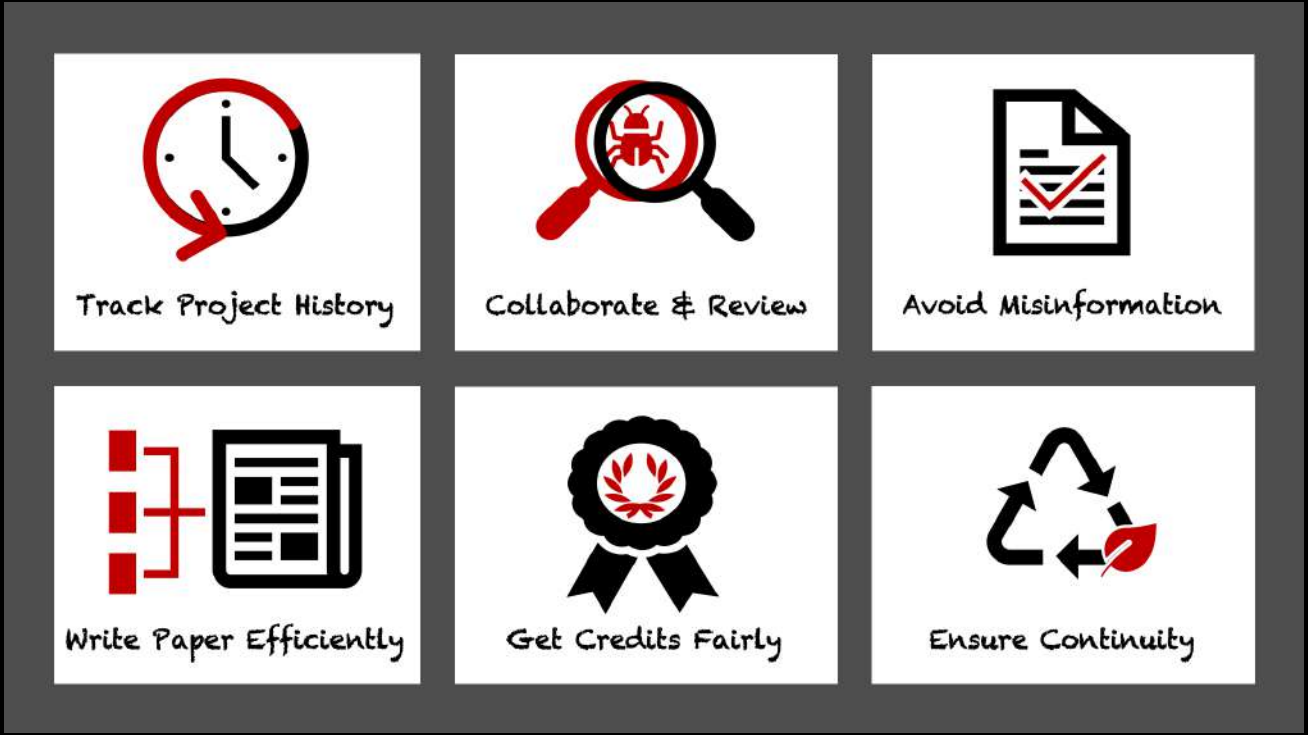
Barriers to Open Research

- Reward system
- Novel findings
- Publication cost
- Training and skill transfer
- Leadership
- Institutions (often)
- Lack of trust



https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway, #TuringWay

Added advantages



https://the-turing-way.netlify.app/reproducible-research/overview/overview-benefit.html, https://doi.org/10.5281/zenodo.4549447

The Turing Way

An **Open Source** project that involves and supports its **diverse community** to make data science is **reproducible**, **ethical,** **collaborative and inclusive** for everyone.

https://github.com/alan-turing-institute/the-turing-way, https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

The Alan Turing Institute

The national institute for data science and artificial intelligence

University network

UNIVERSITY OF CAMBRIDGE

THE UNIVERSITY of EDINBURGH

UNIVERSITY OF OXFORD

UCL

WARWICK
THE UNIVERSITY OF WARWICK

University of BRISTOL

UNIVERSITY OF BIRMINGHAM

UNIVERSITY OF EXETER

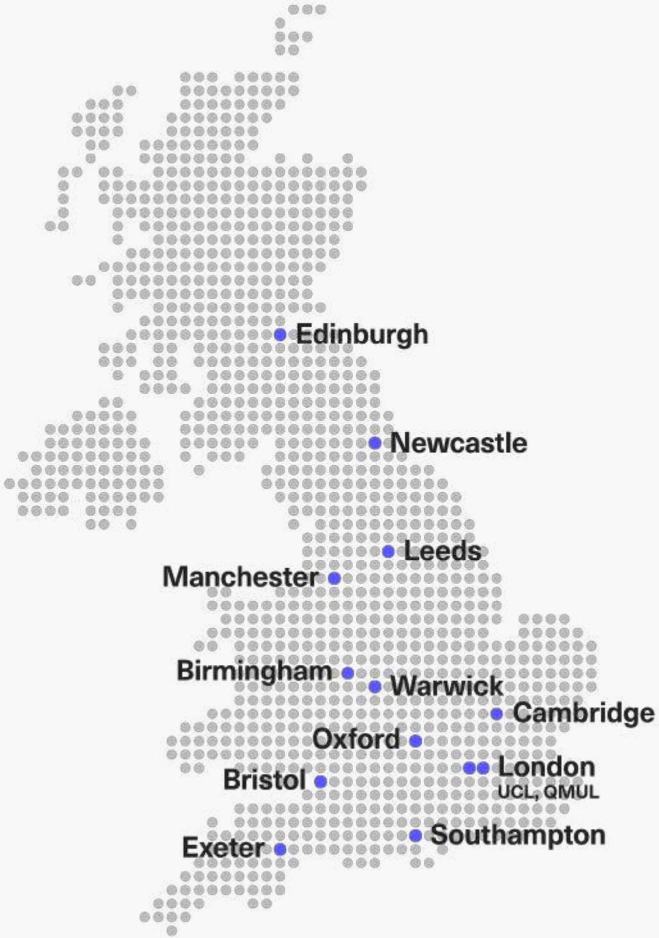
UNIVERSITY OF LEEDS

MANCHESTER 1824
The University of Manchester

Queen Mary
University of London

Newcastle University

UNIVERSITY OF Southampton



https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

The Institute’s partners and collaborators

Lloyd’s Register Foundation

intel

GCHQ

[dstl]

Ministry of Defence

HSBC

uclh

accenture

NATS

British Heart Foundation

Microsoft

TOYOTA mobility FOUNDATION

Office for National Statistics

GREATER LONDON AUTHORITY

CRAY

nationalgrid

SIEMENS

LLOYDS BANKING GROUP

Cystic Fibrosis Trust

ROLLS ROYCE

Scottish Enterprise

Nuffield Foundation

SAMSUNG

BRITISH AIRWAYS

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Challenges

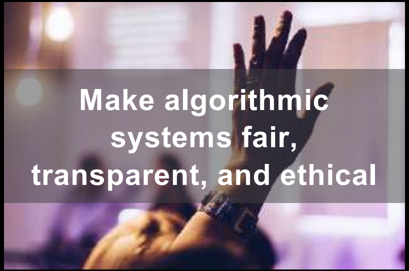
Advance data science and artificial intelligence to…

Revolutionise healthcare

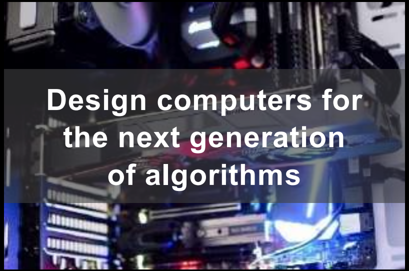
Deliver safer, smarter engineering

Manage security in an insecure world

Shine a light on our economy



Make algorithmic systems fair, transparent, and ethical



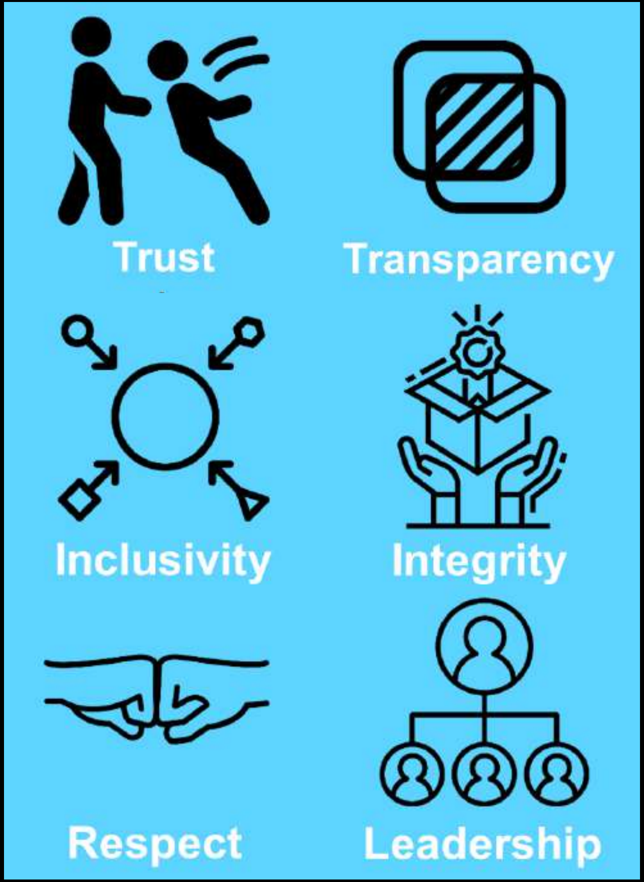
Design computers for the next generation of algorithms

Supercharge research in science and humanities

Foster government innovation

Tools, Practices and Systems

- Trustworthy systems
- Transparent reporting
- Inclusive interoperable design
- Ethical integrity
- Respectful co-creation
- Leadership in open research



https://doi.org/10.5281/zenodo.4549447, @malvikasharan

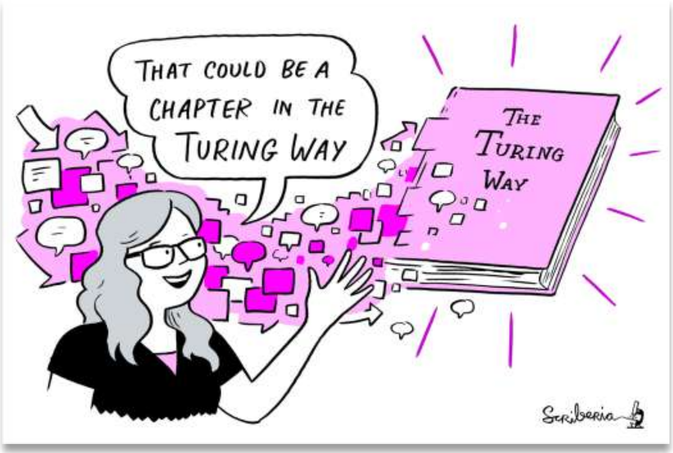
Book:
the-turing-way.netlify.app/

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

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

A Book



A Community

An Open Source Project

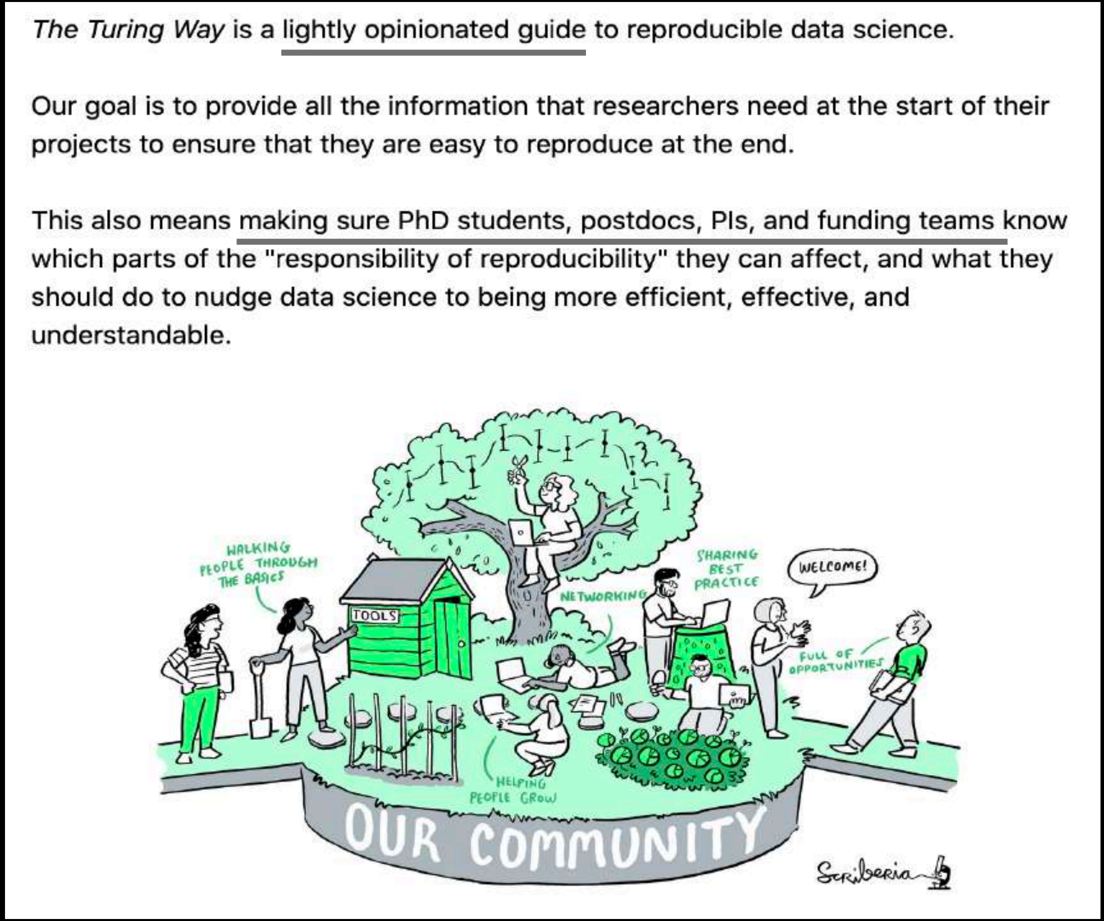
A Culture of Collaboration

THERE'S MORE TO COLLABORATION THAN YOU MIGHT THINK!
DIFFERENT TEAMS
COMMON GOAL
EXCHANGE OF KNOWLEDGE
PRODUCTION
TREAT EACH OTHER KINDLY
BUILD DIVERSE TEAMS
CODE OF CONDUCT
INCLUSIVE WORKSPACE
EXPLICITELY OPEN FOR CONTRIBUTION
Scriberia

The Turing Way Book on Reproducibility

Kirstie Whitaker

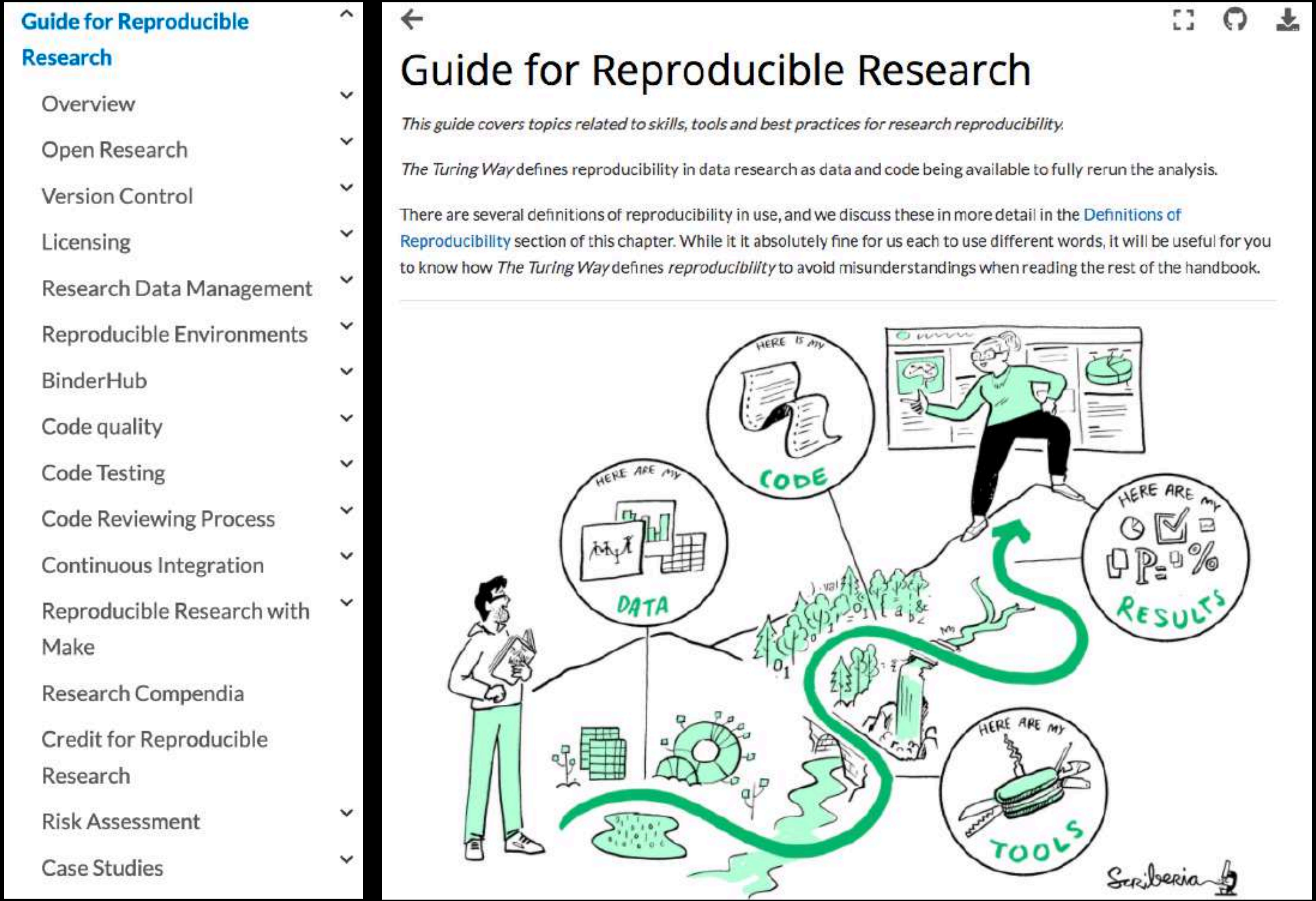
Lead of Tools, Practices & Systems Programme



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.



Guide for Reproducible Research

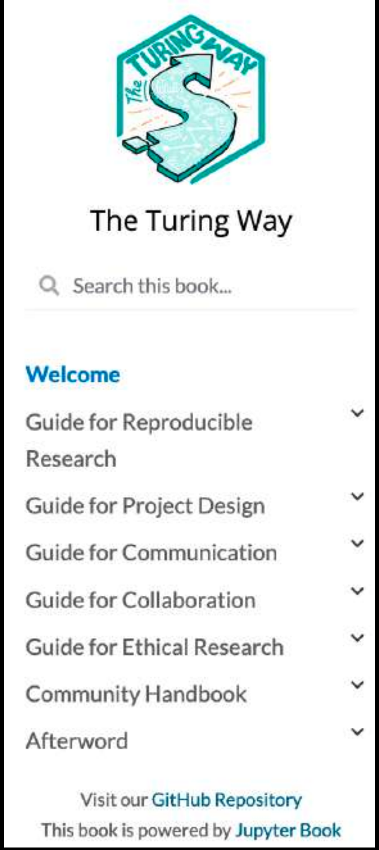
This guide covers topics related to skills, tools and best practices for research reproducibility.

The Turing Way defines reproducibility in data research as data and code being available to fully rerun the analysis.

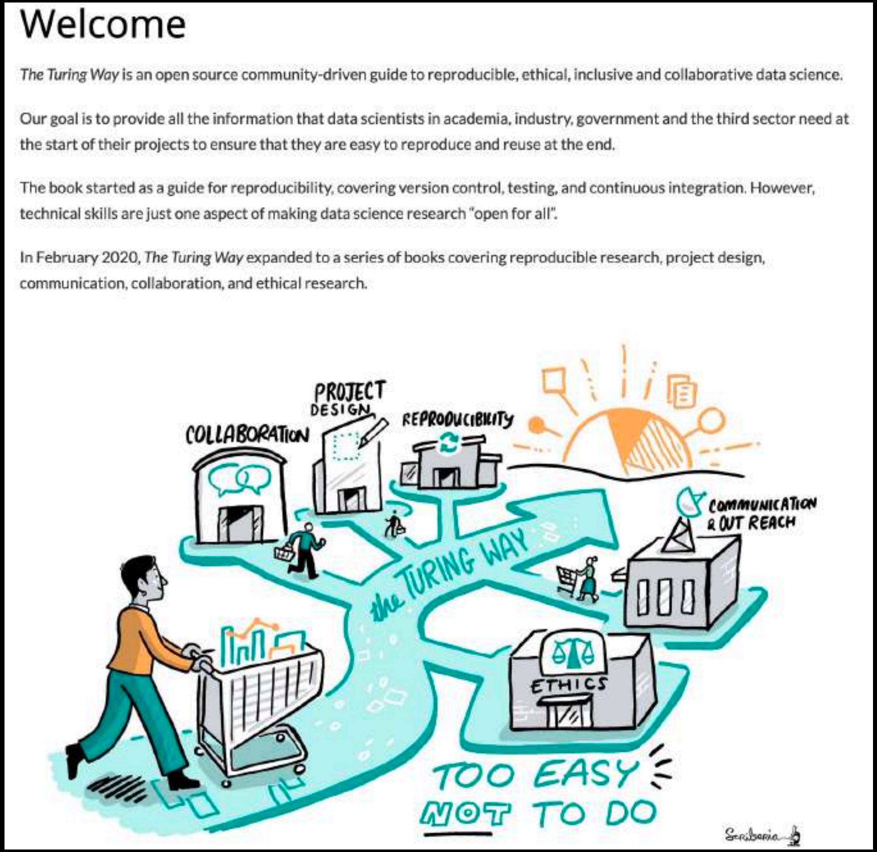
There are several definitions of reproducibility in use, and we discuss these in more detail in the Definitions of Reproducibility section of this chapter. While it it absolutely fine for us each to use different words, it will be useful for you to know how *The Turing Way* defines *reproducibility* to avoid misunderstandings when reading the rest of the handbook.

Book: https://the-turing-way.netlify.app/welcome, https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Moonshot Goal: Reproducibility “too easy not to do”



Visit our GitHub Repository
This book is powered by Jupyter Book



Welcome

The Turing Way is an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

Our goal is to provide all the information that data scientists in academia, industry, government and the third sector need at the start of their projects to ensure that they are easy to reproduce and reuse at the end.

The book started as a guide for reproducibility, covering version control, testing, and continuous integration. However, technical skills are just one aspect of making data science research “open for all”.

In February 2020, *The Turing Way* expanded to a series of books covering reproducible research, project design, communication, collaboration, and ethical research.

Book: https://the-turing-way.netlify.app/welcome, https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

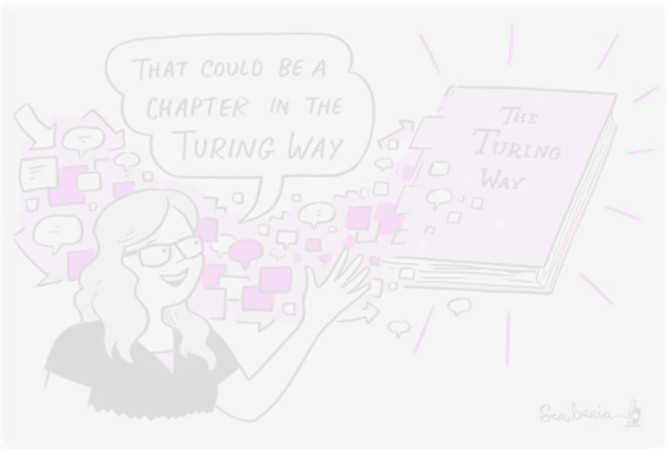
Book:
the-turing-way.netlify.app/

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

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

A Book



A Community

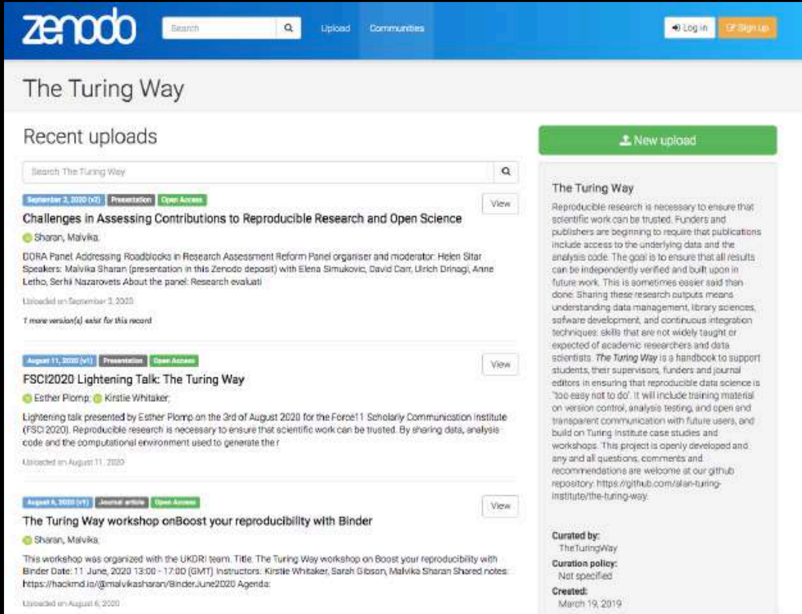
An Open Source Project

PULL REQUEST
CLONE
APPROVED
Scriberia

A Culture of Collaboration

An open source project

- everyone can freely read, reuse, distribute, modify and help develop
- the project belongs to *The Turing Way* community (CC-BY license)



The Turing Way

Recent uploads

September 3, 2020 (v2) Presentation Open Access

Challenges in Assessing Contributions to Reproducible Research and Open Science

Sharan, Malvika;

DORA Panel Addressing Roadblocks in Research Assessment Reform Panel organiser and moderator: Helen Sitar Speakers: Malvika Sharan (presentation in this Zenodo deposit) with Elena Simukovic, David Carr, Ulrich Drinagi, Anne Letho, Serhii Nazarovets About the panel: Research evaluati

Uploaded on September 3, 2020

1 more version(s) exist for this record

August 11, 2020 (v1) Presentation Open Access

FSCI2020 Lightening Talk: The Turing Way

Esther Plomp; Kirstie Whitaker;

Lightening talk presented by Esther Plomp on the 3rd of August 2020 for the Force11 Scholarly Communication Institute (FSCI 2020). Reproducible research is necessary to ensure that scientific work can be trusted. By sharing data, analysis code and the computational environment used to generate the r

Uploaded on August 11, 2020

August 6, 2020 (v1) Journal article Open Access

The Turing Way workshop onBoost your reproducibility with Binder

Sharan, Malvika;

This workshop was organized with the UKDRI team. Title: The Turing Way workshop on Boost your reproducibility with Binder Date: 11 June, 2020 13:00 - 17:00 (GMT) Instructors: Kirstie Whitaker, Sarah Gibson, Malvika Sharan Shared notes: https://hackmd.io/@malvikasharan/BinderJune2020 Agenda:

Uploaded on August 6, 2020

The Turing Way

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 in ensuring that reproducible data science is "too easy not to do". It will include training material on version control, analysis testing, and open and transparent communication with future users, and build on Turing Institute case studies and workshops. This project is openly developed and any and all questions, comments and recommendations are welcome at our github repository: https://github.com/alan-turing-institute/the-turing-way.

Curated by:
TheTuringWay

Curation policy:
Not specified

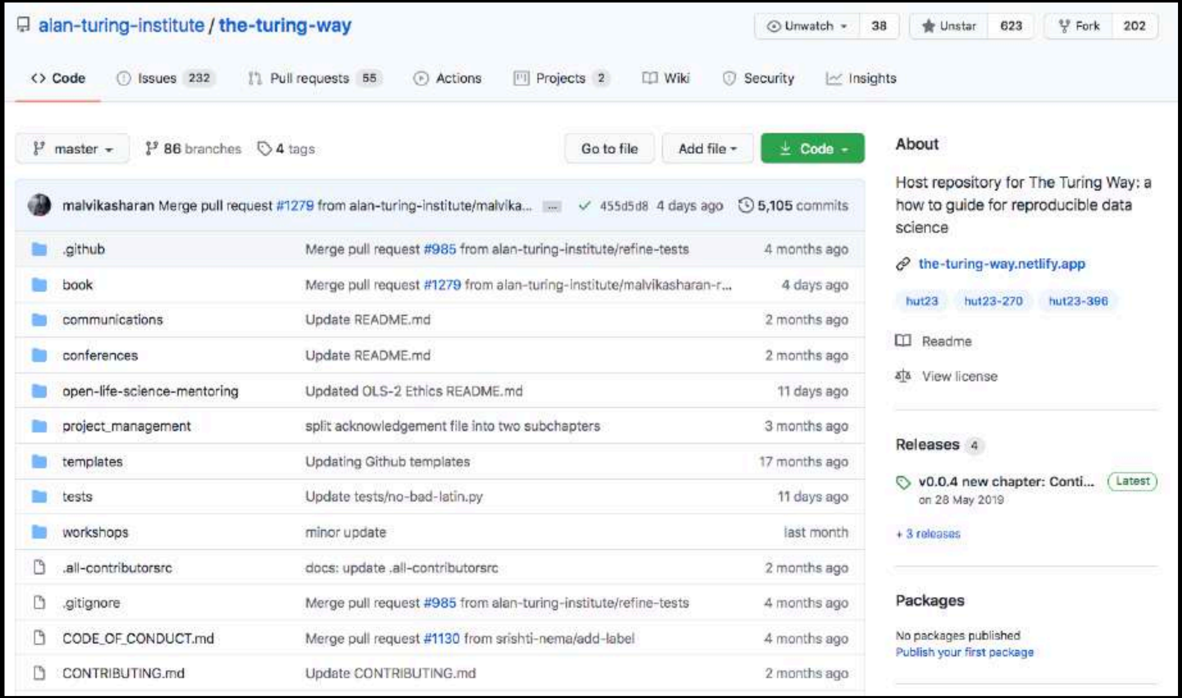
Created:
March 19, 2019

https://zenodo.org/communities/the-turing-way

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

An open source project

Hosted on GitHub



malvikasharan Merge pull request #1279 from alan-turing-institute/malvika... 455d5d8 4 days ago 5,105 commits

.github	Merge pull request #985 from alan-turing-institute/refine-tests	4 months ago
book	Merge pull request #1279 from alan-turing-institute/malvikasharan-r...	4 days ago
communications	Update README.md	2 months ago
conferences	Update README.md	2 months ago
open-life-science-mentoring	Updated OLS-2 Ethics README.md	11 days ago
project_management	split acknowledgement file into two subchapters	3 months ago
templates	Updating Github templates	17 months ago
tests	Update tests/no-bad-latin.py	11 days ago
workshops	minor update	last month
.all-contributorsrc	docs: update .all-contributorsrc	2 months ago
.gitignore	Merge pull request #985 from alan-turing-institute/refine-tests	4 months ago
CODE_OF_CONDUCT.md	Merge pull request #1130 from srishti-nema/add-label	4 months ago
CONTRIBUTING.md	Update CONTRIBUTING.md	2 months ago

About

Host repository for The Turing Way: a how to guide for reproducible data science

the-turing-way.netlify.app

Releases 4

v0.0.4 new chapter: Conti... Latest
on 28 May 2019

+ 3 releases

Packages

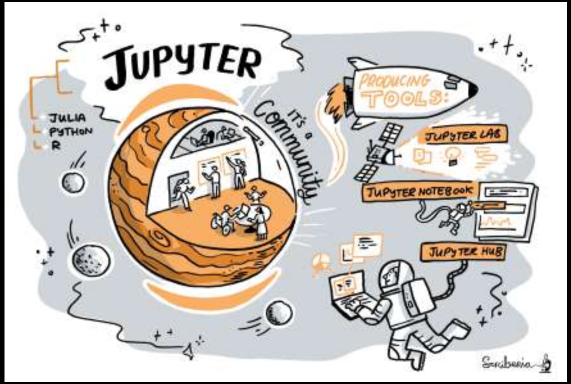
No packages published
Publish your first package

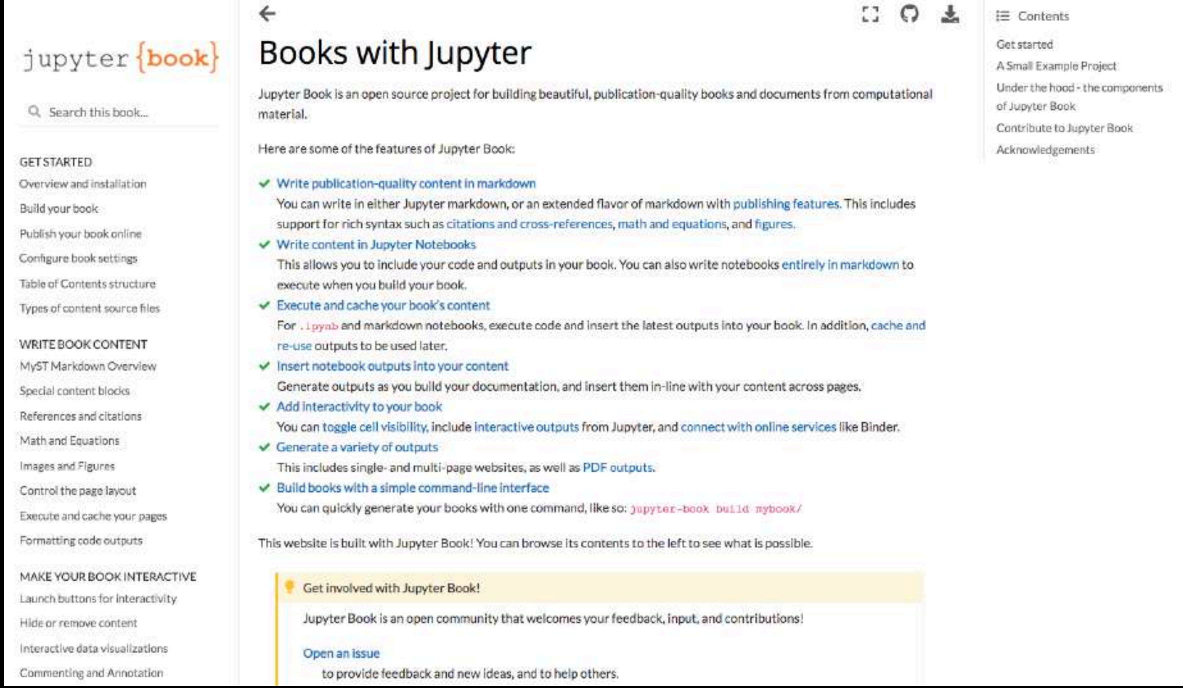
https://github.com/alan-turing-institute/the-turing-way

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

An open source project

built with other Open Source tools





GET STARTED
- Overview and installation
- Build your book
- Publish your book online
- Configure book settings
- Table of Contents structure
- Types of content source files

WRITE BOOK CONTENT
- MyST Markdown Overview
- Special content blocks
- References and citations
- Math and Equations
- Images and Figures
- Control the page layout
- Execute and cache your pages
- Formatting code outputs

MAKE YOUR BOOK INTERACTIVE
- Launch buttons for interactivity
- Hide or remove content
- Interactive data visualizations
- Commenting and Annotation

Books with Jupyter

Jupyter Book is an open source project for building beautiful, publication-quality books and documents from computational material.

Here are some of the features of Jupyter Book:

✓ Write publication-quality content in markdown
You can write in either Jupyter markdown, or an extended flavor of markdown with publishing features. This includes support for rich syntax such as citations and cross-references, math and equations, and figures.

✓ Write content in Jupyter Notebooks
This allows you to include your code and outputs in your book. You can also write notebooks entirely in markdown to execute when you build your book.

✓ Execute and cache your book's content
For `.ipynb` and markdown notebooks, execute code and insert the latest outputs into your book. In addition, cache and re-use outputs to be used later.

✓ Insert notebook outputs into your content
Generate outputs as you build your documentation, and insert them in-line with your content across pages.

✓ Add interactivity to your book
You can toggle cell visibility, include interactive outputs from Jupyter, and connect with online services like Binder.

✓ Generate a variety of outputs
This includes single- and multi-page websites, as well as PDF outputs.

✓ Build books with a simple command-line interface
You can quickly generate your books with one command, like so: `jupyter-book build mybook/`

This website is built with Jupyter Book! You can browse its contents to the left to see what is possible.

Get involved with Jupyter Book!

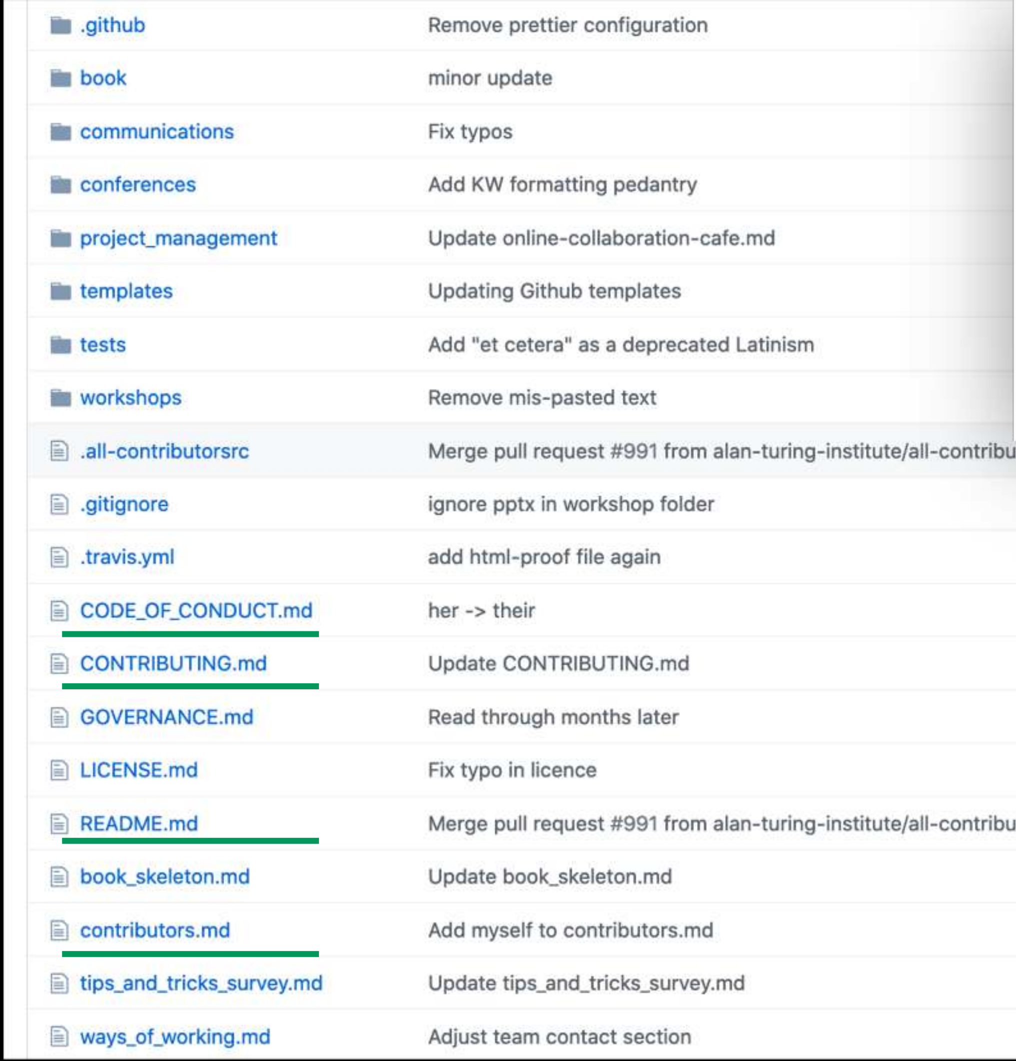
Jupyter Book is an open community that welcomes your feedback, input, and contributions!

Open an issue
to provide feedback and new ideas, and to help others.

Contents
- Get started
- A Small Example Project
- Under the hood - the components of Jupyter Book
- Contribute to Jupyter Book
- Acknowledgements

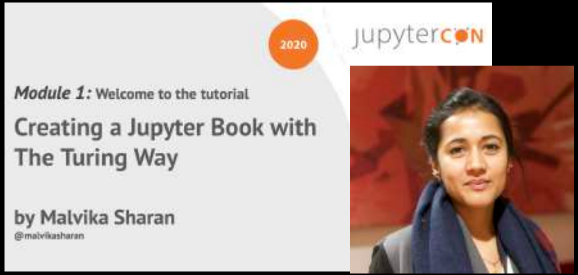
https://jupyterbook.org/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway



.github	Remove prettier configuration	
book	minor update	
communications	Fix typos	
conferences	Add KW formatting pedantry	
project_management	Update online-collaboration-cafe.md	
templates	Updating Github templates	
tests	Add "et cetera" as a deprecated Latinism	
workshops	Remove mis-pasted text	
.all-contributorsrc	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
.gitignore	ignore pptx in workshop folder	9 months ago
.travis.yml	add html-proof file again	last month
CODE_OF_CONDUCT.md	her -> their	6 months ago
CONTRIBUTING.md	Update CONTRIBUTING.md	2 months ago
GOVERNANCE.md	Read through months later	5 months ago
LICENSE.md	Fix typo in licence	2 months ago
README.md	Merge pull request #991 from alan-turing-institute/all-contributors/a...	5 days ago
book_skeleton.md	Update book_skeleton.md	13 months ago
contributors.md	Add myself to contributors.md	11 months ago
tips_and_tricks_survey.md	Update tips_and_tricks_survey.md	14 months ago
ways_of_working.md	Adjust team contact section	5 months ago

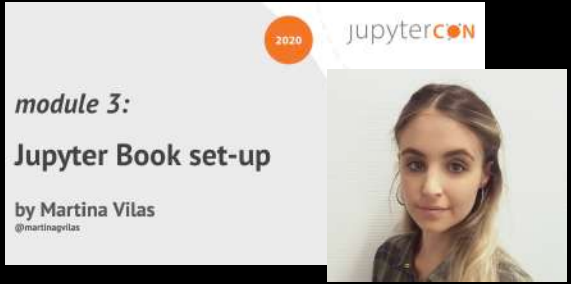
An open source project


2020 jupytercon

Module 1: Welcome to the tutorial

Creating a Jupyter Book with The Turing Way

by Malvika Sharan
@malvikasharan


2020 jupytercon

module 3:

Jupyter Book set-up

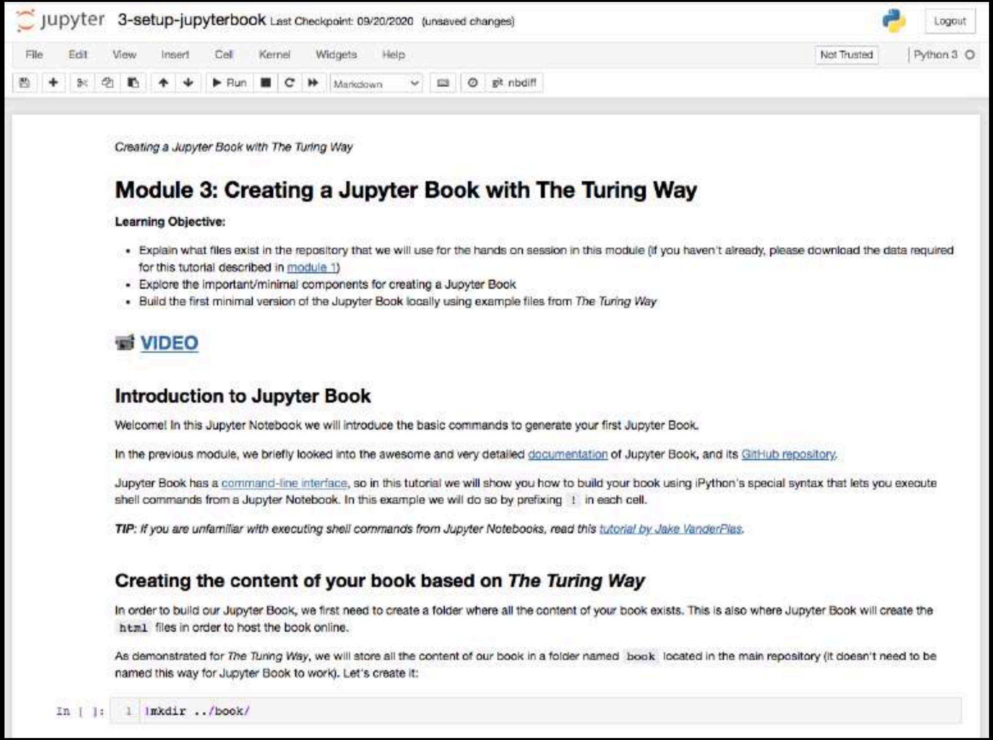
by Martina Vilas
@martinagvilas

2020 jupytercon

module 6:

Jupyter Book CI/CD

by Sarah Gibson
@sgibson91


jupyter 3-setup-jupyterbook Last Checkpoint: 09/20/2020 (unsaved changes)
Logout
File Edit View Insert Cell Kernel Widgets Help
Not Trusted
Python 3
Run
Markdown
nbdiff

Creating a Jupyter Book with The Turing Way

Module 3: Creating a Jupyter Book with The Turing Way

Learning Objective:

- Explain what files exist in the repository that we will use for the hands on session in this module (if you haven't already, please download the data required for this tutorial described in module 1)
- Explore the important/minimal components for creating a Jupyter Book
- Build the first minimal version of the Jupyter Book locally using example files from *The Turing Way*

VIDEO

Introduction to Jupyter Book

Welcome! In this Jupyter Notebook we will introduce the basic commands to generate your first Jupyter Book.

In the previous module, we briefly looked into the awesome and very detailed documentation of Jupyter Book, and its GitHub repository.

Jupyter Book has a command-line interface, so in this tutorial we will show you how to build your book using iPython's special syntax that lets you execute shell commands from a Jupyter Notebook. In this example we will do so by prefixing `!` in each cell.

TIP: *If you are unfamiliar with executing shell commands from Jupyter Notebooks, read this tutorial by Jake VanderPlas.*

Creating the content of your book based on *The Turing Way*

In order to build our Jupyter Book, we first need to create a folder where all the content of your book exists. This is also where Jupyter Book will create the `html` files in order to host the book online.

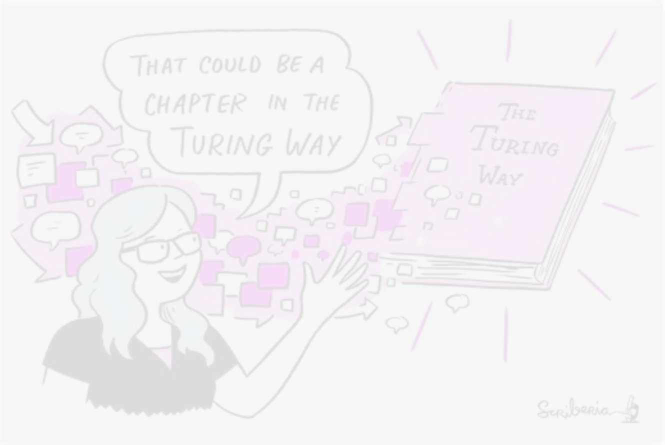
As demonstrated for *The Turing Way*, we will store all the content of our book in a folder named `book` located in the main repository (it doesn't need to be named this way for Jupyter Book to work). Let's create it:

```
In [ ]: !mkdir ../book/
```

https://github.com/martinagvilas/tutorial-jupyterbook-with-turing-way/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

A Book



A Community

An Open Source Project

A Culture of Collaboration

THERE'S MORE TO COLLABORATION THAN YOU MIGHT THINK!
DIFFERENT TEAMS
COMMON GOAL
EXCHANGE OF KNOWLEDGE
PRODUCTION
TREAT EACH OTHER KINDLY
BUILD DIVERSE TEAMS
CODE OF CONDUCT
INCLUSIVE WORKSPACE
EXPLICITELY OPEN FOR CONTRIBUTION
Scriberia

Book:
the-turing-way.netlify.app/

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

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

Pathways for Collaboration

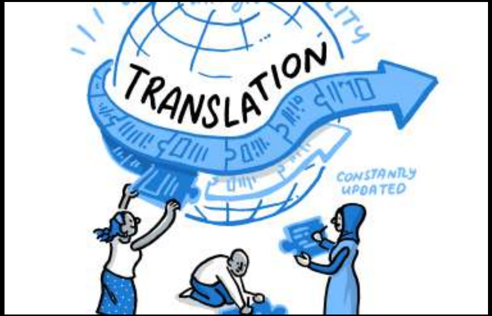
MENTORED CONTRIBUTIONS
REAL TIME INTERACTIONS WITH EXPERIENCED CONTRIBUTORS
EMPOWERING NEW CONTRIBUTORS

Develop & share

Maintain & improve

Share resources

Review and update



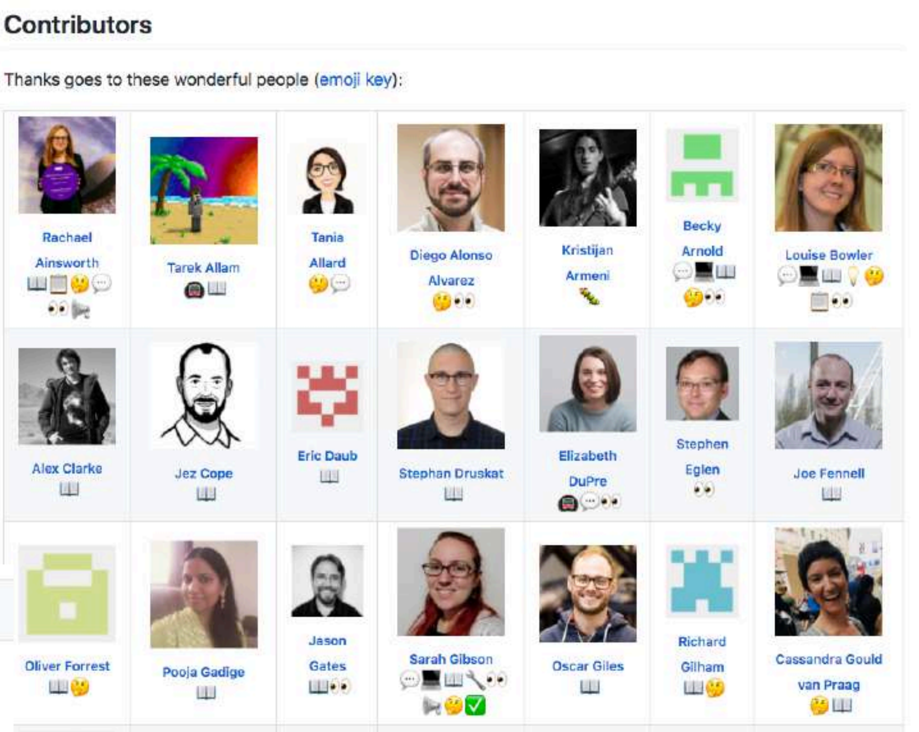
Make it global

Share best practices

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Acknowledgment

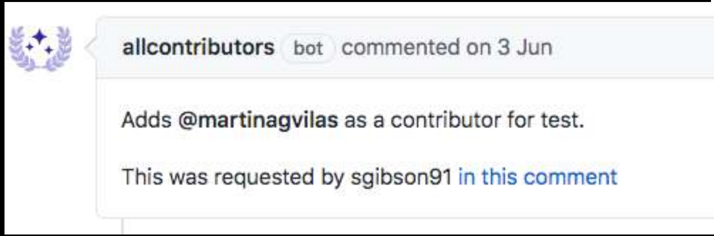
everyone gets acknowledgement for their work



Contributors

Thanks goes to these wonderful people (emoji key):

Rachael Ainsworth 📖📋🤔💬 👀📢	Tarek Allam 🚇📖	Tania Allard 🤔💬	Diego Alonso Alvarez 🤔👀	Kristijan Armeni 🐛	Becky Arnold 💬💻📖 🤔👀	Louise Bowler 💬💻📖💡🤔 📋👀
Alex Clarke 📖	Jez Cope 📖	Eric Daub 📖	Stephan Druskat 📖	Elizabeth DuPre 🚇💬👀	Stephen Eglen 👀	Joe Fennell 📖
Oliver Forrest 📖🤔	Pooja Gadige 📖	Jason Gates 📖👀	Sarah Gibson 💬💻📖🔧👀 📢🤔✅	Oscar Giles 📖	Richard Gilham 📖🤔	Cassandra Gould van Praag 🤔📖



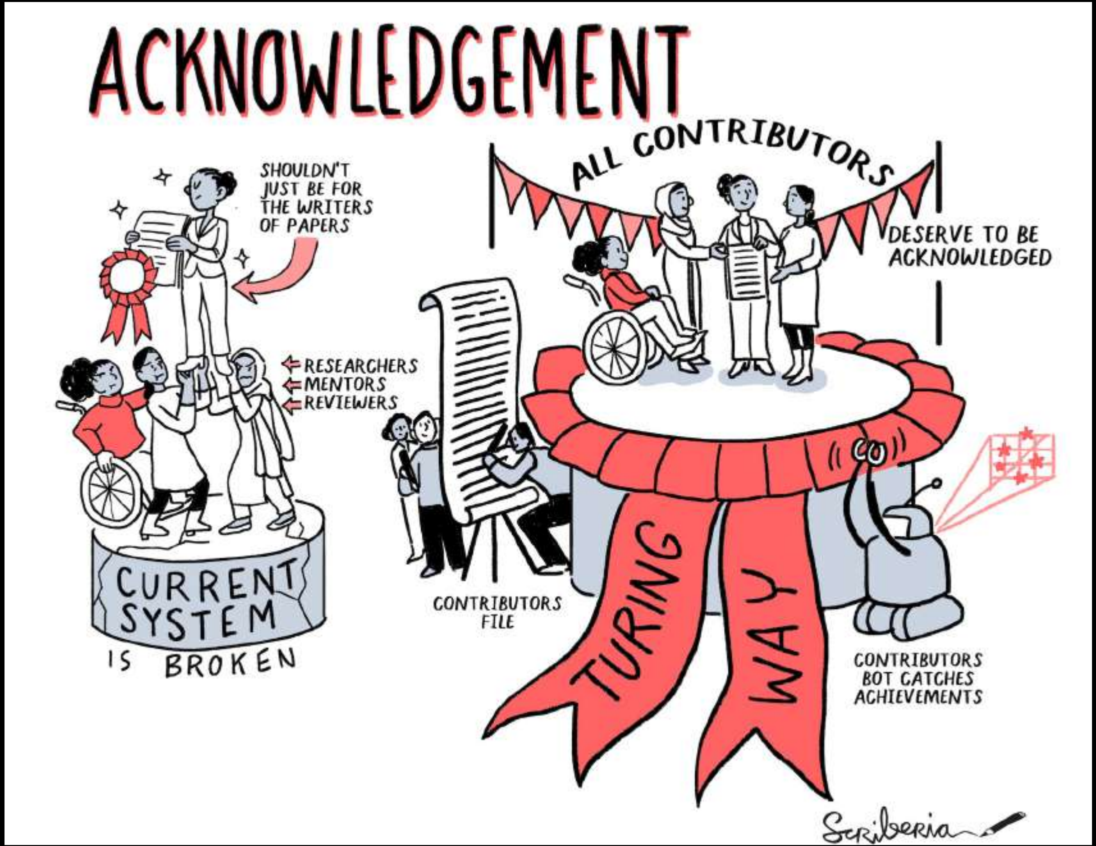
allcontributors bot commented on 3 Jun

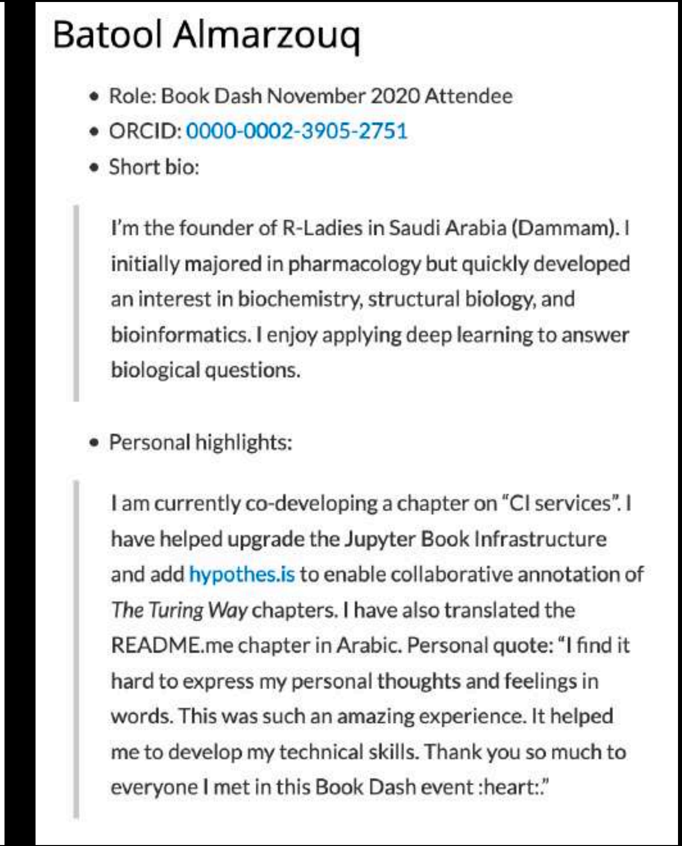
Adds **@martinagvilas** as a contributor for test.

This was requested by sgibson91 in this comment

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Recognise All Contributions





Batool Almarzouq

- Role: Book Dash November 2020 Attendee
- ORCID: 0000-0002-3905-2751
- Short bio:

> I'm the founder of R-Ladies in Saudi Arabia (Dammam). I initially majored in pharmacology but quickly developed an interest in biochemistry, structural biology, and bioinformatics. I enjoy applying deep learning to answer biological questions.

- Personal highlights:

> I am currently co-developing a chapter on "CI services". I have helped upgrade the Jupyter Book Infrastructure and add hypothes.is to enable collaborative annotation of *The Turing Way* chapters. I have also translated the README.me chapter in Arabic. Personal quote: "I find it hard to express my personal thoughts and feelings in words. This was such an amazing experience. It helped me to develop my technical skills. Thank you so much to everyone I met in this Book Dash event :heart:."

CC-BY image by The Turing Way and Scriberia, https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

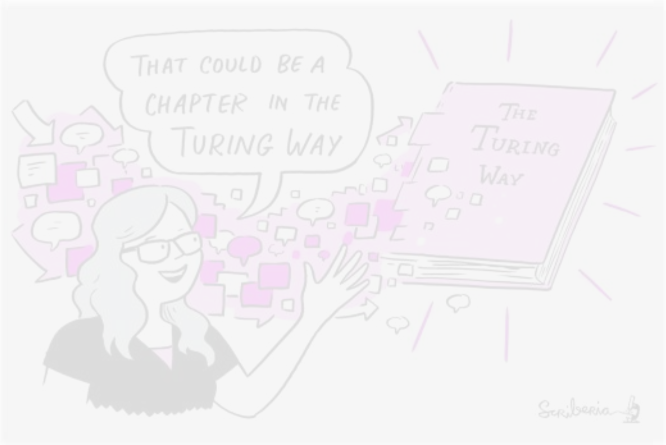
Book:
the-turing-way.netlify.app/

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

Twitter:
twitter.com/turingway

Email:
theturingway@gmail.com

A Book

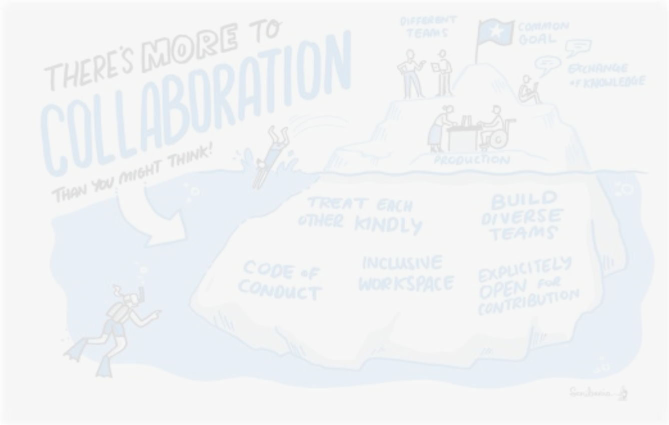


A Community

WALKING PEOPLE THROUGH THE BASICS
TOOLS
NETWORKING
SHARING BEST PRACTICE
WELCOME!
FULL OF OPPORTUNITIES
HELPING PEOPLE GROW
OUR COMMUNITY
Scriberia

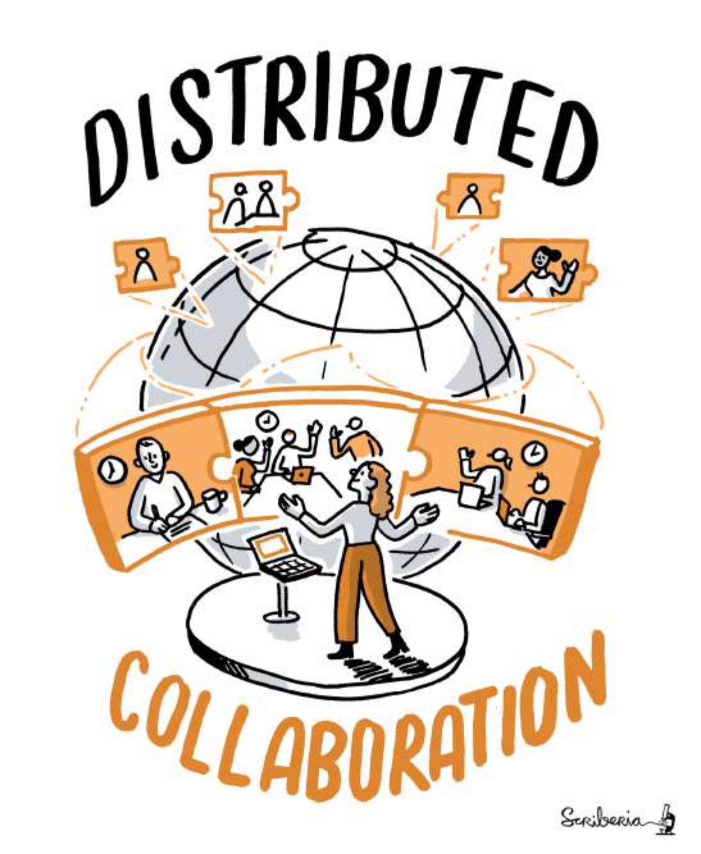
An Open Source Project

A Culture of Collaboration



Open Community

- anyone can join and contribute
- the direction and the goals are determined collaboratively



https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

A Community of Contributors

Collaboration Cafés & Co-working Calls

Book Dash Events

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

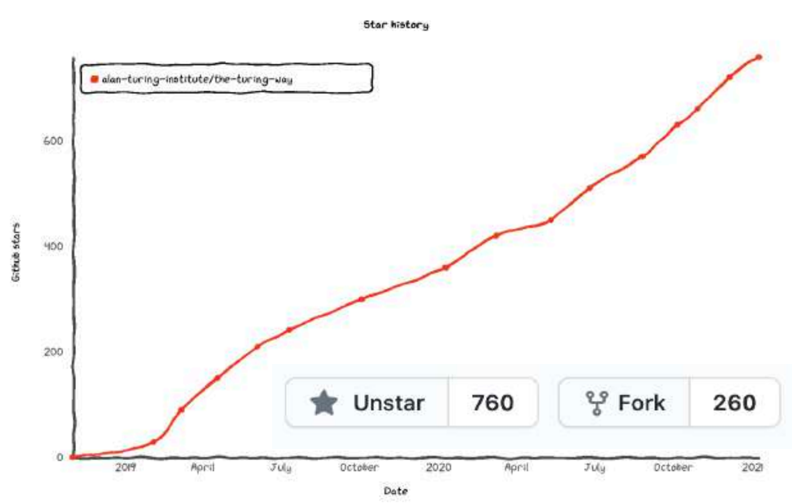
More than a Book!

The co-working hours are friendly for introverts who might be intimidated working with new people. These personal interactions are also crucial for staying motivated!

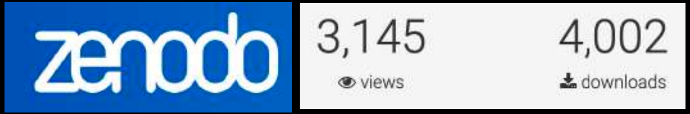
Martina G. Vilas

Project and Community Growth

- 2 years, >134 subchapters
- Community resources, guidance, templates, training
- 250 direct contributors and thousands of users



https://zenodo.org/record/4323154


zenodo

3,145 views

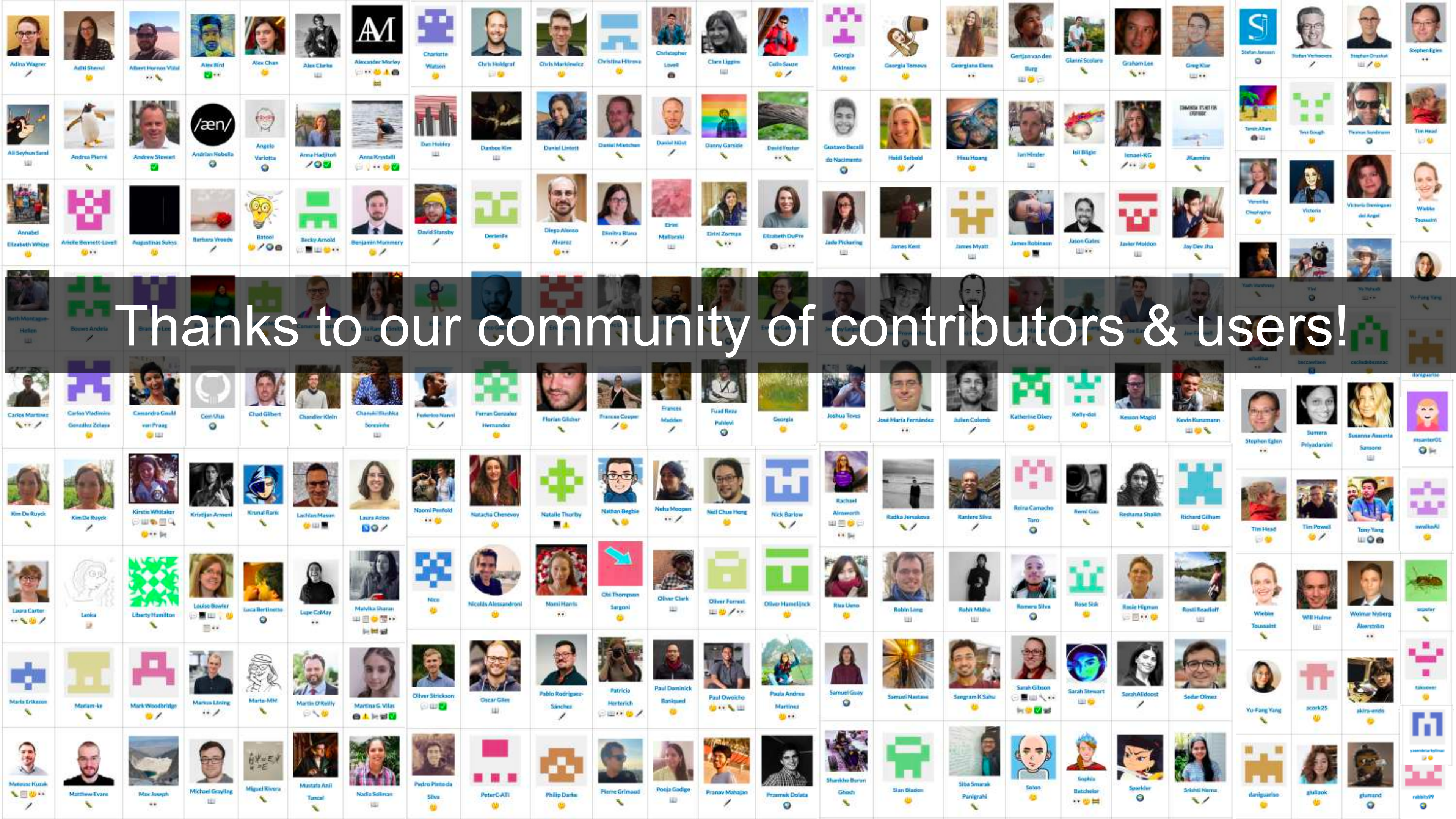
4,002 downloads

https://star-history.t9t.io/#alan-turing-institute/the-turing-way, https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Notable Impacts Beyond the project

Resources are being used by learners, educators, community builders, policy makers and researchers globally

- ➢ Reproducibility of scientific results in the EU 2020 (report)
- ➢ An Emerging Technology Charter for London (policy)
- ➢ Innovation Scholars: UKRI grant 2020 (funding)
- ➢ CodeRefinary and Library Carpentries (training)
- ➢ Cited by multiple peer-reviewed articles



Thanks to our community of contributors & users!

1. what is *The Turing Way*?
2. **guide to computational reproducibility**

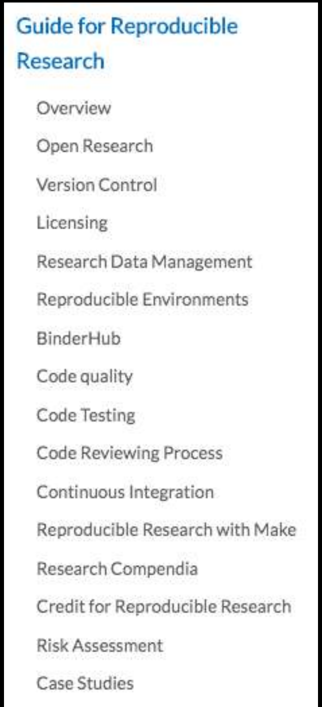
Reproducible Research

same analysis steps on
the same dataset
produces same answer

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

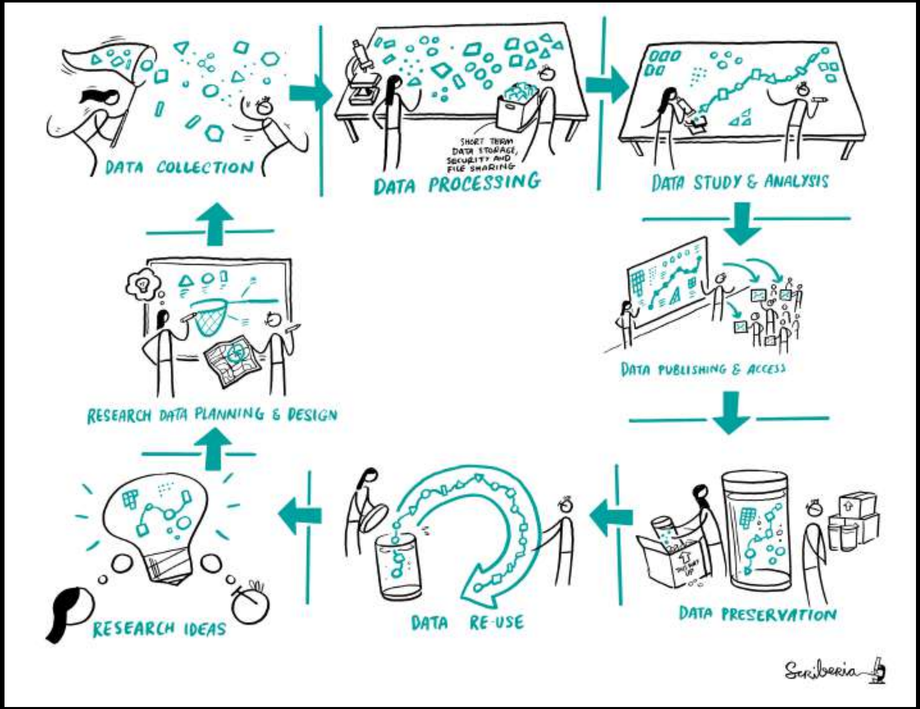
https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html
https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

The Turing Way chapters on reproducibility



Guide for Reproducible Research

- Overview
- Open Research
- Version Control
- Licensing
- Research Data Management
- Reproducible Environments
- BinderHub
- Code quality
- Code Testing
- Code Reviewing Process
- Continuous Integration
- Reproducible Research with Make
- Research Compendia
- Credit for Reproducible Research
- Risk Assessment
- Case Studies



https://the-turing-way.netlify.app

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Computational Reproducibility

- Track changes to your code/resource (version control)
- Write clean, understandable and error free code
- Save and share your computational environment
- Make your code open source

Computational Reproducibility

- → **Track changes to your code (version control)**
- → Write clean, understandable and error free code
- → Save and share your computational environment
- → Make your code open source

Version Control

- records changes to a file or set of files over time
- provides access to any specific version

https://the-turing-way.netlify.app/reproducible-research/vcs.html

Version Control

git

- changes are recorded using **snapshots**
- **distributed** version control system

https://git-scm.com/

Version Control

GitHub

- Web and Desktop App GUI interface
- most Open Source software hosted here

https://git-scm.com/

Computational Reproducibility

→ Track changes to your code (version control)

→ **Write clean, understandable and error free code**

→ Save and share your computational environment

→ Make your code open source

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Style Guide

- set of conventions on how to format code
- e.g.
  - ✓ Indentation
  - ✓ Comments
  - ✓ Imports
  - ✓ Naming

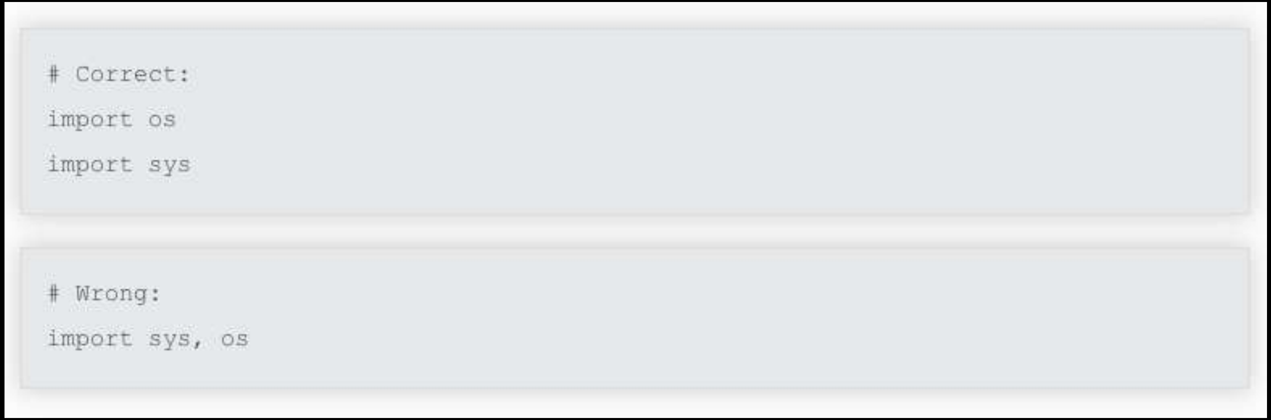
https://the-turing-way.netlify.app/reproducible-research/code-quality/code-quality-style.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Style Guide

PEP 8

e.g. imports



```
# Correct:
import os
import sys
```

```
# Wrong:
import sys, os
```

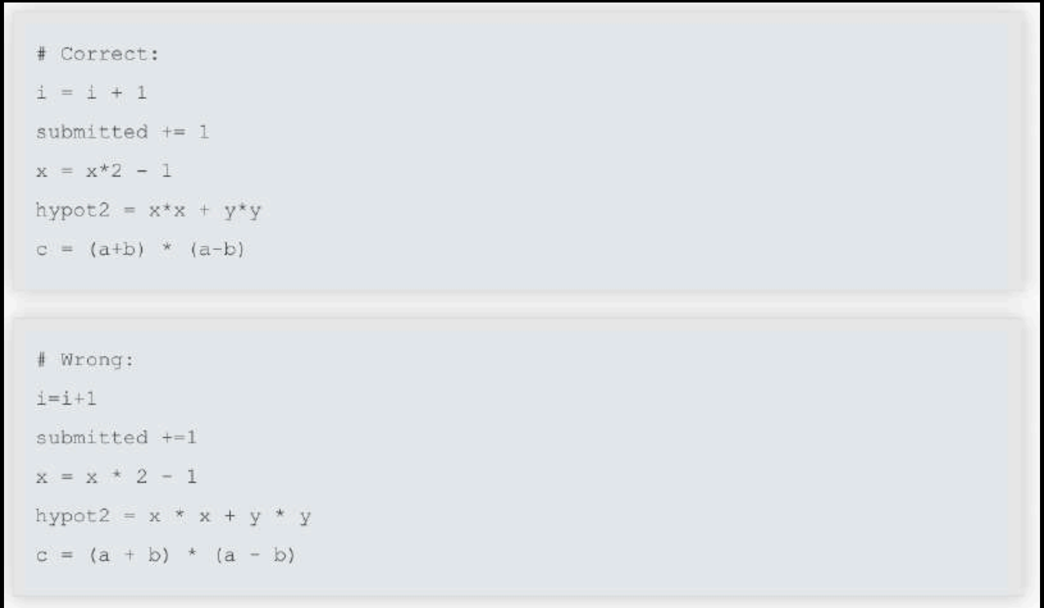
https://www.python.org/dev/peps/pep-0008/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Style Guide

PEP 8

e.g. spaces



```
# Correct:
i = i + 1
submitted += 1
x = x*2 - 1
hypot2 = x*x + y*y
c = (a+b) * (a-b)
```

```
# Wrong:
i=i+1
submitted +=1
x = x * 2 - 1
hypot2 = x * x + y * y
c = (a + b) * (a - b)
```

https://www.python.org/dev/peps/pep-0008/

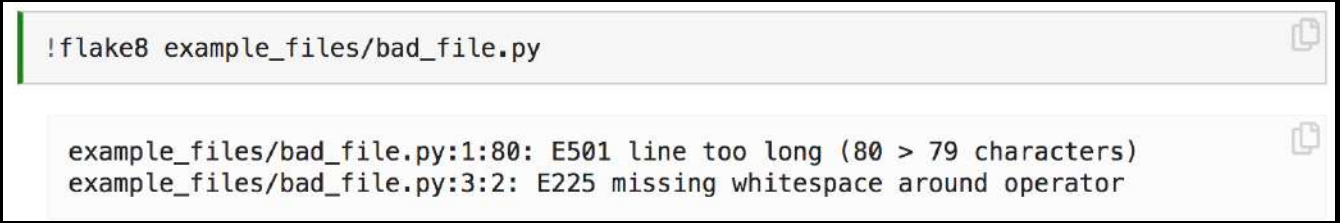
Code Style Guide

Style Guide Enforcement tools

https://flake8.pycqa.org/en/latest/

Code Style Guide

Style Guide Enforcement tools → ***flake8***



```
!flake8 example_files/bad_file.py
```

```
example_files/bad_file.py:1:80: E501 line too long (80 > 79 characters)
example_files/bad_file.py:3:2: E225 missing whitespace around operator
```

https://flake8.pycqa.org/en/latest/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Style Guide

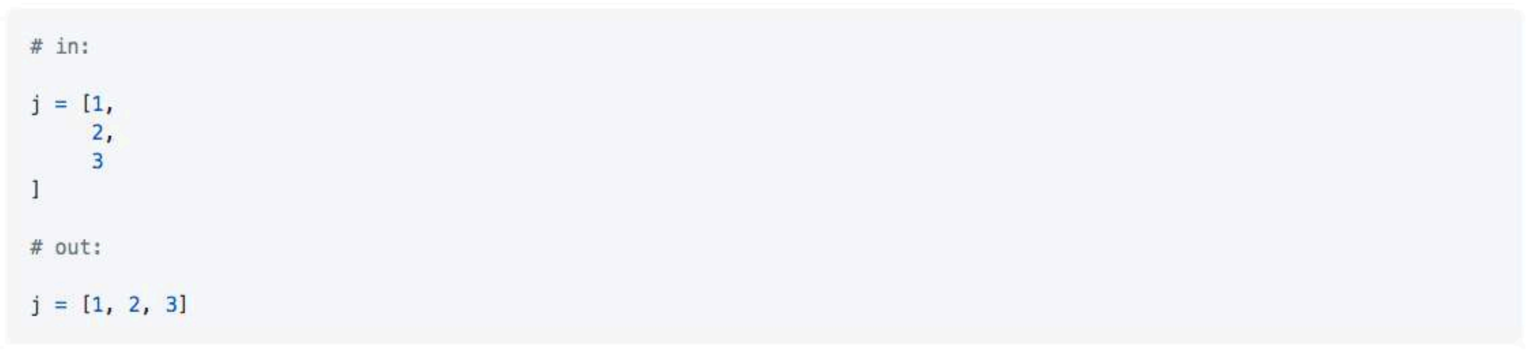
Automatic formatting tools

https://github.com/psf/black

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Style Guide

Automatic formatting tools→ ***Black***



```
# in:

j = [1,
     2,
     3
]

# out:

j = [1, 2, 3]
```

https://github.com/psf/black

Code Review

- have another programmer look over your code and assess it

https://the-turing-way.netlify.app/reproducible-research/reviewing.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Review

types

→ synchronous

- lab meetings presentations

→ asynchronous

- GitHub

https://the-turing-way.netlify.app/reproducible-research/reviewing.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Testing

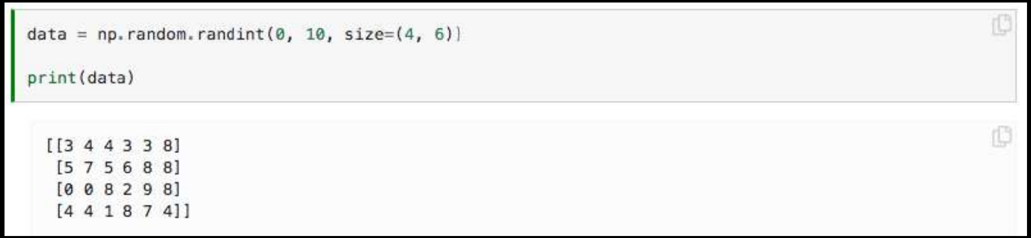
“You should not skip writing tests because you are short on time, you should write tests because you are short on time”

https://the-turing-way.netlify.app/reproducible-research/testing.html
https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code Testing

"You should not skip writing tests because you are short on time, you should write tests because you are short on time"

- you probably are already doing it …



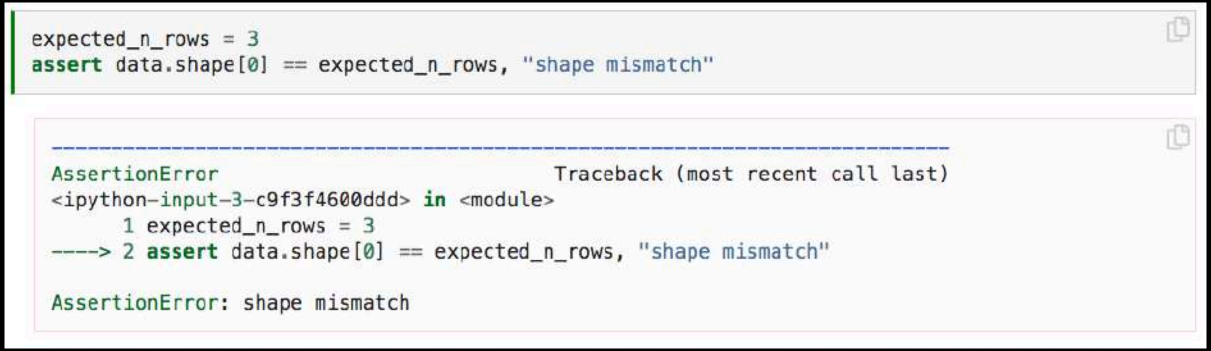
```
data = np.random.randint(0, 10, size=(4, 6))

print(data)
```

```
[[3 4 4 3 3 8]
 [5 7 5 6 8 8]
 [0 0 8 2 9 8]
 [4 4 1 8 7 4]]
```

https://the-turing-way.netlify.app/reproducible-research/testing.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Assert

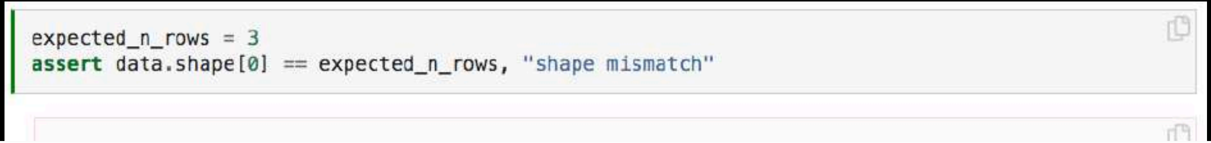


```
expected_n_rows = 3
assert data.shape[0] == expected_n_rows, "shape mismatch"
```

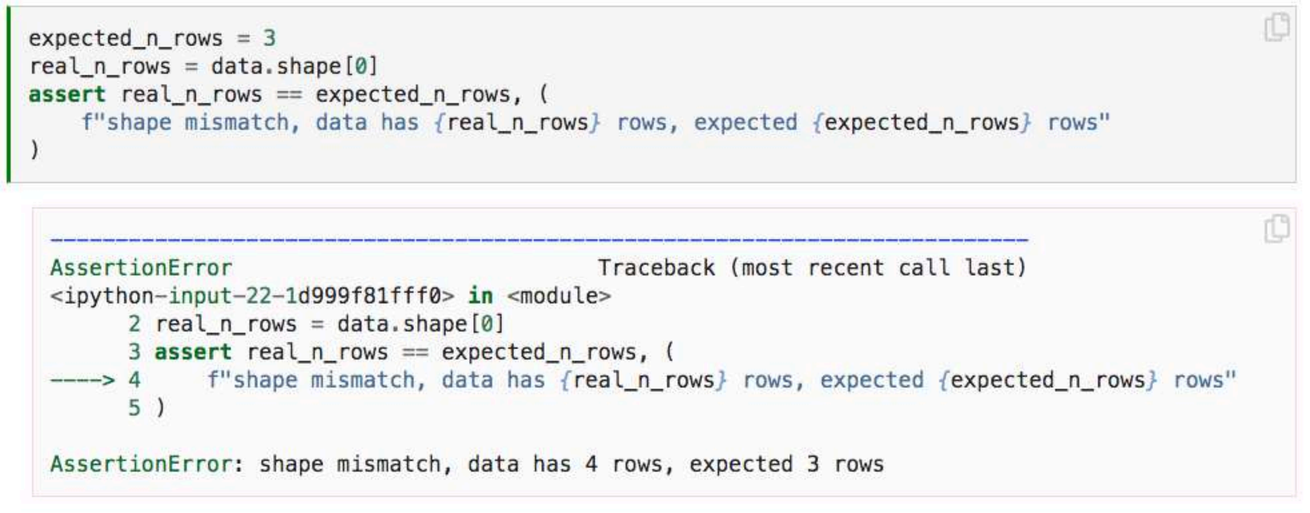
```
---------------------------------------------------------------------------
AssertionError                            Traceback (most recent call last)
<ipython-input-3-c9f3f4600ddd> in <module>
      1 expected_n_rows = 3
----> 2 assert data.shape[0] == expected_n_rows, "shape mismatch"

AssertionError: shape mismatch
```

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Assert



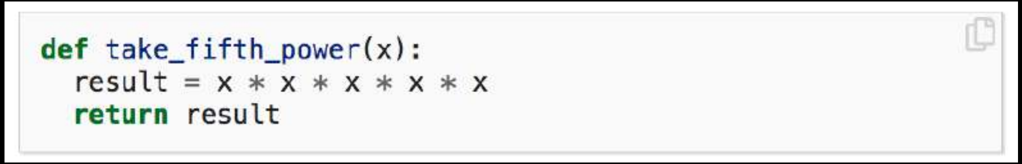
```python
expected_n_rows = 3
assert data.shape[0] == expected_n_rows, "shape mismatch"
```



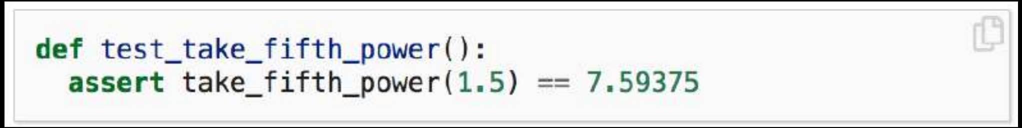
```python
expected_n_rows = 3
real_n_rows = data.shape[0]
assert real_n_rows == expected_n_rows, (
    f"shape mismatch, data has {real_n_rows} rows, expected {expected_n_rows} rows"
)
```

```
---------------------------------------------------------------------------
AssertionError                            Traceback (most recent call last)
<ipython-input-22-1d999f81fff0> in <module>
      2 real_n_rows = data.shape[0]
      3 assert real_n_rows == expected_n_rows, (
----> 4     f"shape mismatch, data has {real_n_rows} rows, expected {expected_n_rows} rows"
      5 )

AssertionError: shape mismatch, data has 4 rows, expected 3 rows
```

Unit Testing



```python
def take_fifth_power(x):
  result = x * x * x * x * x
  return result
```



```python
def test_take_fifth_power():
  assert take_fifth_power(1.5) == 7.59375
```

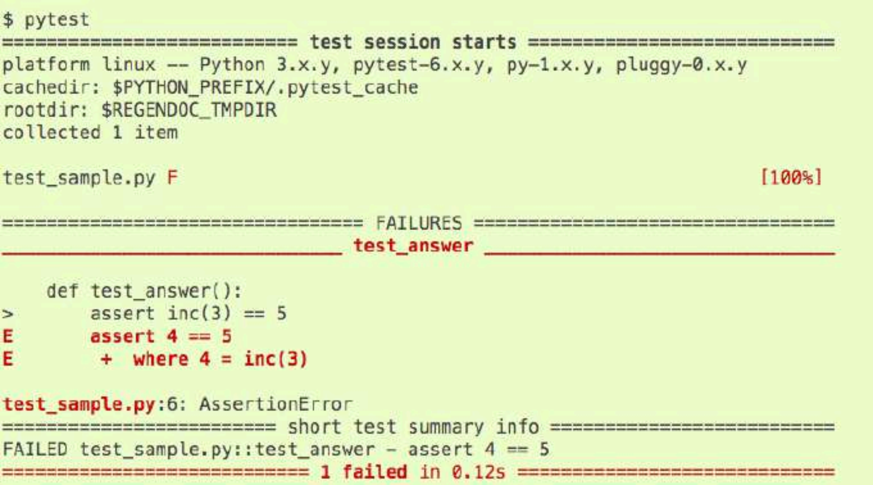
https://the-turing-way.netlify.app/reproducible-research/testing/testing-types.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Pytest

Testing framework

- detailed info of assert statements



```
$ pytest
========================== test session starts ============================
platform linux -- Python 3.x.y, pytest-6.x.y, py-1.x.y, pluggy-0.x.y
cachedir: $PYTHON_PREFIX/.pytest_cache
rootdir: $REGENDOC_TMPDIR
collected 1 item

test_sample.py F                                                    [100%]

================================ FAILURES =================================
______________________________ test_answer ________________________________

    def test_answer():
>       assert inc(3) == 5
E       assert 4 == 5
E        +  where 4 = inc(3)

test_sample.py:6: AssertionError
======================== short test summary info ==========================
FAILED test_sample.py::test_answer - assert 4 == 5
=========================== 1 failed in 0.12s =============================
```

https://docs.pytest.org/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Pytest

Testing framework

- detailed info of assert statements
- auto-discovery of tests

```
setup.py
mypkg/
    __init__.py
    app.py
    view.py
tests/
    test_app.py
    test_view.py
    ...
```

https://docs.pytest.org/

Other Types of Testing

Smoke test

- initial checks designed to ensure very basic functionality

Integration test

- individual units are combined and tested as a group

and more

https://the-turing-way.netlify.app/reproducible-research/testing/testing-types.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Computational Reproducibility

- → Track changes to your code (version control)
- → Write clean, understandable and error free code
- → **Save and share your computational environment**
- → Make your code open source

Reproducible Computational Environments

		Interaction style	
		Graphical	Command line
What is reproduced?	Software and versions	Binder	Conda
	Entire system	Virtual machine	Containers

https://the-turing-way.netlify.app/reproducible-research/renv/renv-options.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Conda

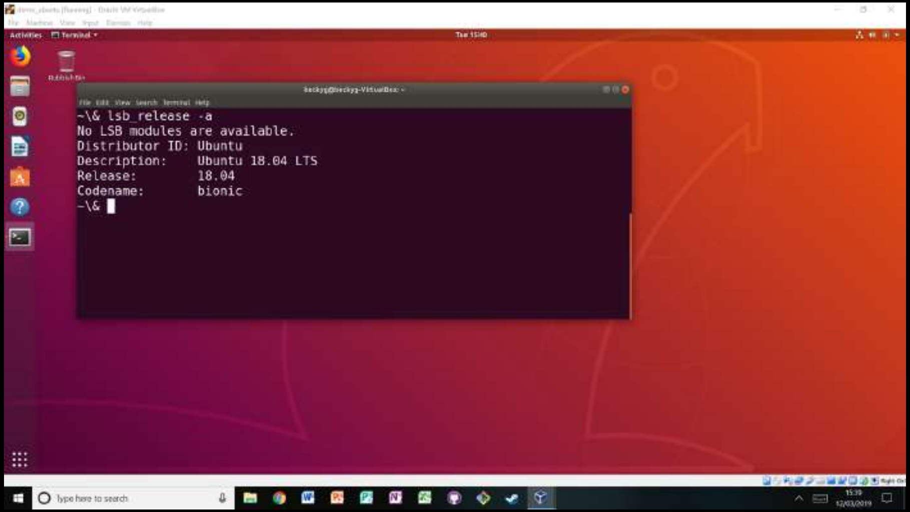
CONDA

- environment management system
- package and version management system

https://the-turing-way.netlify.app/reproducible-research/renv/renv-package.html

Virtual Machines

package a whole computer as an app that can be run


Activities
Terminal
Rubbish Bin
beckyg@beckyg-VirtualBox: ~
File Edit View Search Terminal Help
Type here to search

```
~\& lsb_release -a
No LSB modules are available.
Distributor ID: Ubuntu
Description:    Ubuntu 18.04 LTS
Release:        18.04
Codename:       bionic
~\& 
```

https://the-turing-way.netlify.app/reproducible-research/renv/renv-vm.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

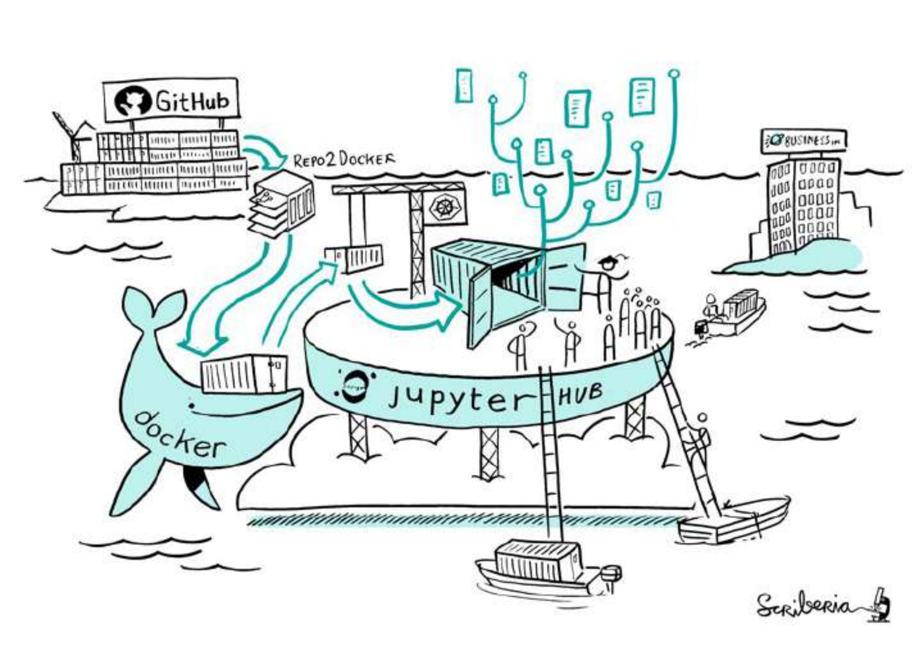
Containers

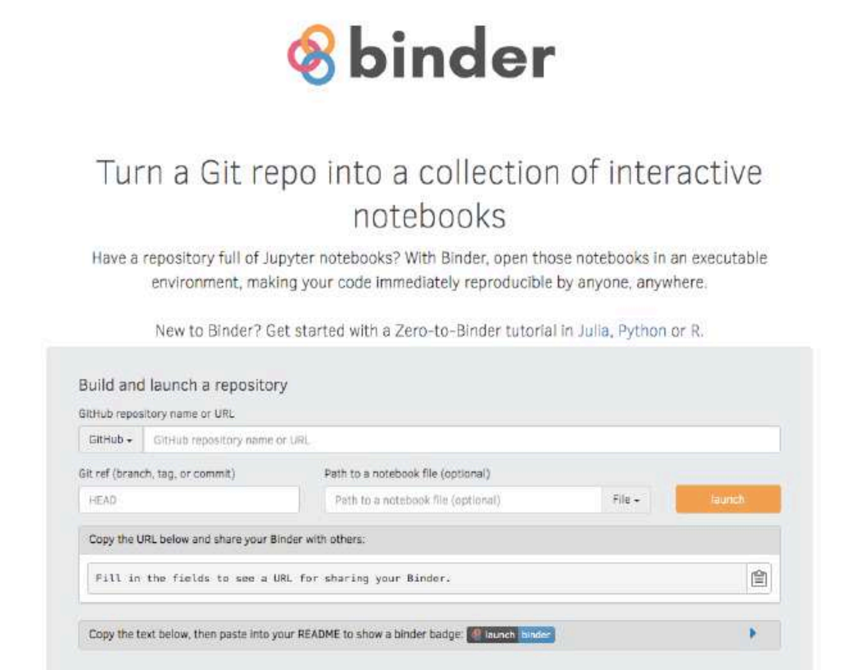
- behaves like a virtual machine
- more lightweight -> only contains the individual components needed to operate the project

https://the-turing-way.netlify.app/reproducible-research/renv/renv-containers.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Binder





Turn a Git repo into a collection of interactive notebooks

Have a repository full of Jupyter notebooks? With Binder, open those notebooks in an executable environment, making your code immediately reproducible by anyone, anywhere.

New to Binder? Get started with a Zero-to-Binder tutorial in Julia, Python or R.

Build and launch a repository

GitHub repository name or URL

GitHub ▾ | GitHub repository name or URL

Git ref (branch, tag, or commit): HEAD

Path to a notebook file (optional): Path to a notebook file (optional) | File ▾

launch

Copy the URL below and share your Binder with others:

```
Fill in the fields to see a URL for sharing your Binder.
```

Copy the text below, then paste into your README to show a binder badge: launch binder

https://the-turing-way.netlify.app/reproducible-research/renv/renv-binder.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Computational Reproducibility

→ Track changes to your code (version control)

→ Write clean, understandable and error free code

→ Save and share your computational environment

→ **Make your code open source**

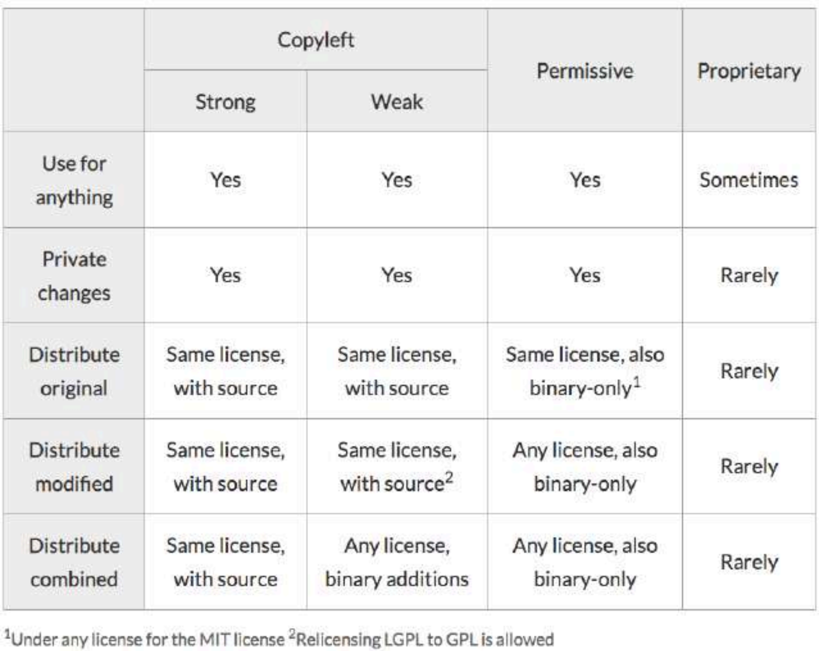
Open Source

- anybody can view, use, modify, and distribute the software for any purpose

https://the-turing-way.netlify.app/reproducible-research/open/open-source.html

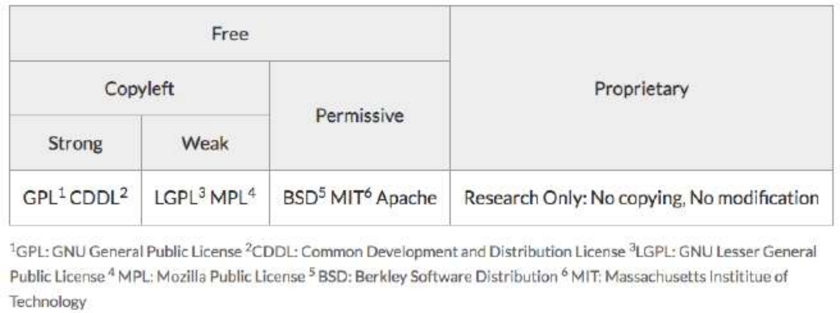
https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Licensing



	Copyleft		Permissive	Proprietary
	Strong	Weak		
Use for anything	Yes	Yes	Yes	Sometimes
Private changes	Yes	Yes	Yes	Rarely
Distribute original	Same license, with source	Same license, with source	Same license, also binary-only[1]	Rarely
Distribute modified	Same license, with source	Same license, with source[2]	Any license, also binary-only	Rarely
Distribute combined	Same license, with source	Any license, binary additions	Any license, also binary-only	Rarely

[1]Under any license for the MIT license [2]Relicensing LGPL to GPL is allowed



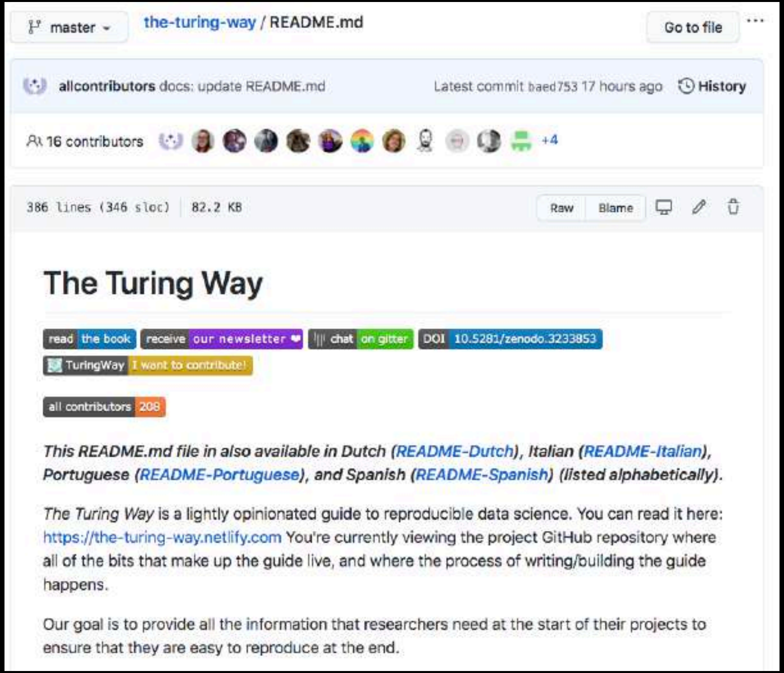
Free			Proprietary
Copyleft		Permissive	
Strong	Weak		
GPL[1] CDDL[2]	LGPL[3] MPL[4]	BSD[5] MIT[6] Apache	Research Only: No copying, No modification

[1]GPL: GNU General Public License [2]CDDL: Common Development and Distribution License [3]LGPL: GNU Lesser General Public License [4] MPL: Mozilla Public License [5] BSD: Berkley Software Distribution [6] MIT: Massachusetts Inistitue of Technology

https://the-turing-way.netlify.app/reproducible-research/licensing/licensing-software.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Community Files



master
the-turing-way / README.md
Go to file
allcontributors docs: update README.md
Latest commit baed753 17 hours ago
History
16 contributors
+4
386 lines (346 sloc) | 82.2 KB
Raw
Blame

The Turing Way

read the book | receive our newsletter | chat on gitter | DOI 10.5281/zenodo.3233853 | TuringWay I want to contribute! | all contributors 208

This README.md file in also available in Dutch (README-Dutch), Italian (README-Italian), Portuguese (README-Portuguese), and Spanish (README-Spanish) (listed alphabetically).

The Turing Way is a lightly opinionated guide to reproducible data science. You can read it here: https://the-turing-way.netlify.com You're currently viewing the project GitHub repository where all of the bits that make up the guide live, and where the process of writing/building the guide happens.

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.

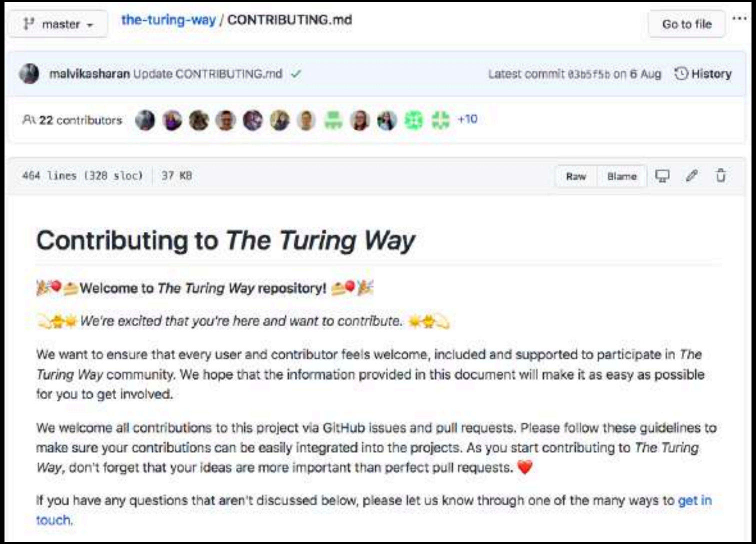
README.md

- project name and main features
- installation instructions
- how to run associated tests
- list of authors/contributors
- contact information
- links to related material

https://the-turing-way.netlify.app/reproducible-research/open/open-source.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Contribution Guideline


master
the-turing-way / CONTRIBUTING.md
Go to file
malvikasharan Update CONTRIBUTING.md ✓
Latest commit 83b5f5b on 6 Aug
History
22 contributors
+10
464 lines (328 sloc) | 37 KB
Raw
Blame

Contributing to *The Turing Way*

🎉🎈🎂 **Welcome to *The Turing Way* repository!** 🎂🎈🎉

💫🌟✨ *We're excited that you're here and want to contribute.* ✨🌟💫

We want to ensure that every user and contributor feels welcome, included and supported to participate in *The Turing Way* community. We hope that the information provided in this document will make it as easy as possible for you to get involved.

We welcome all contributions to this project via GitHub issues and pull requests. Please follow these guidelines to make sure your contributions can be easily integrated into the projects. As you start contributing to *The Turing Way*, don't forget that your ideas are more important than perfect pull requests. ❤️

If you have any questions that aren't discussed below, please let us know through one of the many ways to get in touch.

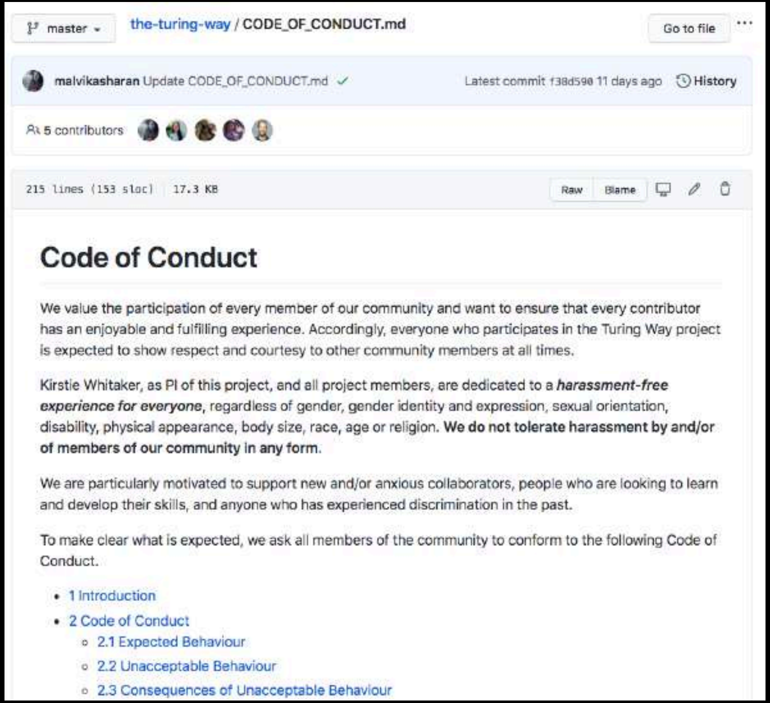
CONTRIBUTING.md

- how to file a bug report
- how to suggest a new feature
- how to contribute changes
- roadmap or vision for the project
- how contributors should (or should not) get in touch with you

https://the-turing-way.netlify.app/reproducible-research/open/open-source.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Code of Conduct


master
the-turing-way / CODE_OF_CONDUCT.md
Go to file
malvikasharan Update CODE_OF_CONDUCT.md ✓
Latest commit f38d590 11 days ago
History
5 contributors
215 lines (153 sloc) | 17.3 KB
Raw
Blame

Code of Conduct

We value the participation of every member of our community and want to ensure that every contributor has an enjoyable and fulfilling experience. Accordingly, everyone who participates in the Turing Way project is expected to show respect and courtesy to other community members at all times.

Kirstie Whitaker, as PI of this project, and all project members, are dedicated to a ***harassment-free experience for everyone***, regardless of gender, gender identity and expression, sexual orientation, disability, physical appearance, body size, race, age or religion. **We do not tolerate harassment by and/or of members of our community in any form.**

We are particularly motivated to support new and/or anxious collaborators, people who are looking to learn and develop their skills, and anyone who has experienced discrimination in the past.

To make clear what is expected, we ask all members of the community to conform to the following Code of Conduct.

CODE_OF_CONDUCT.md

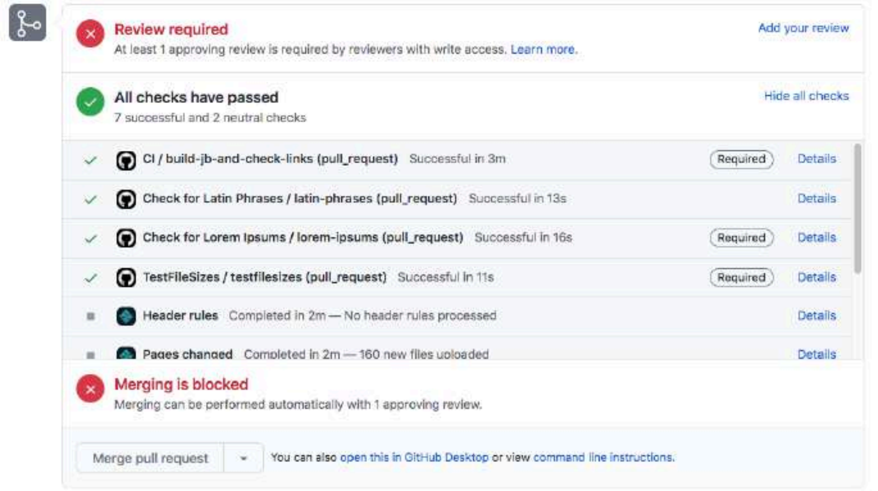
- how you expect participants to behave
- who these expectations apply to
- when they apply
- what to do if a violation occurs

https://the-turing-way.netlify.app/reproducible-research/open/open-source.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Continuous Integration

- practice of integrating changes to a project made by individuals into a main, shared version
- frequently



Review required
At least 1 approving review is required by reviewers with write access. Learn more.

All checks have passed
7 successful and 2 neutral checks

✓	**CI / build-jb-and-check-links (pull_request)** Successful in 3m	Required	Details
✓	**Check for Latin Phrases / latin-phrases (pull_request)** Successful in 13s		Details
✓	**Check for Lorem Ipsums / lorem-ipsums (pull_request)** Successful in 16s	Required	Details
✓	**TestFileSizes / testfilesizes (pull_request)** Successful in 11s	Required	Details
■	**Header rules** Completed in 2m — No header rules processed		Details
■	**Pages changed** Completed in 2m — 160 new files uploaded		Details

Merging is blocked
Merging can be performed automatically with 1 approving review.

You can also open this in GitHub Desktop or view command line instructions.

https://the-turing-way.netlify.app/reproducible-research/ci.html

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Continuous Integration

2020

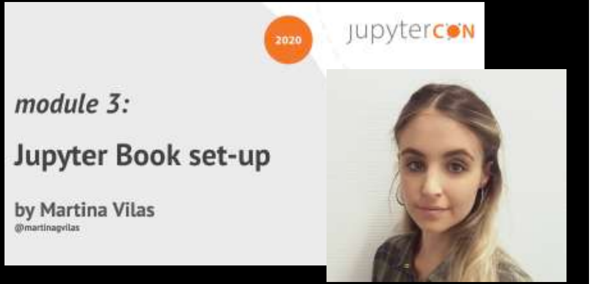
jupytercon

Module 1: Welcome to the tutorial

Creating a Jupyter Book with The Turing Way

by Malvika Sharan

@malvikasharan



2020

jupytercon

module 3:

Jupyter Book set-up

by Martina Vilas

@martinagvilas

2020

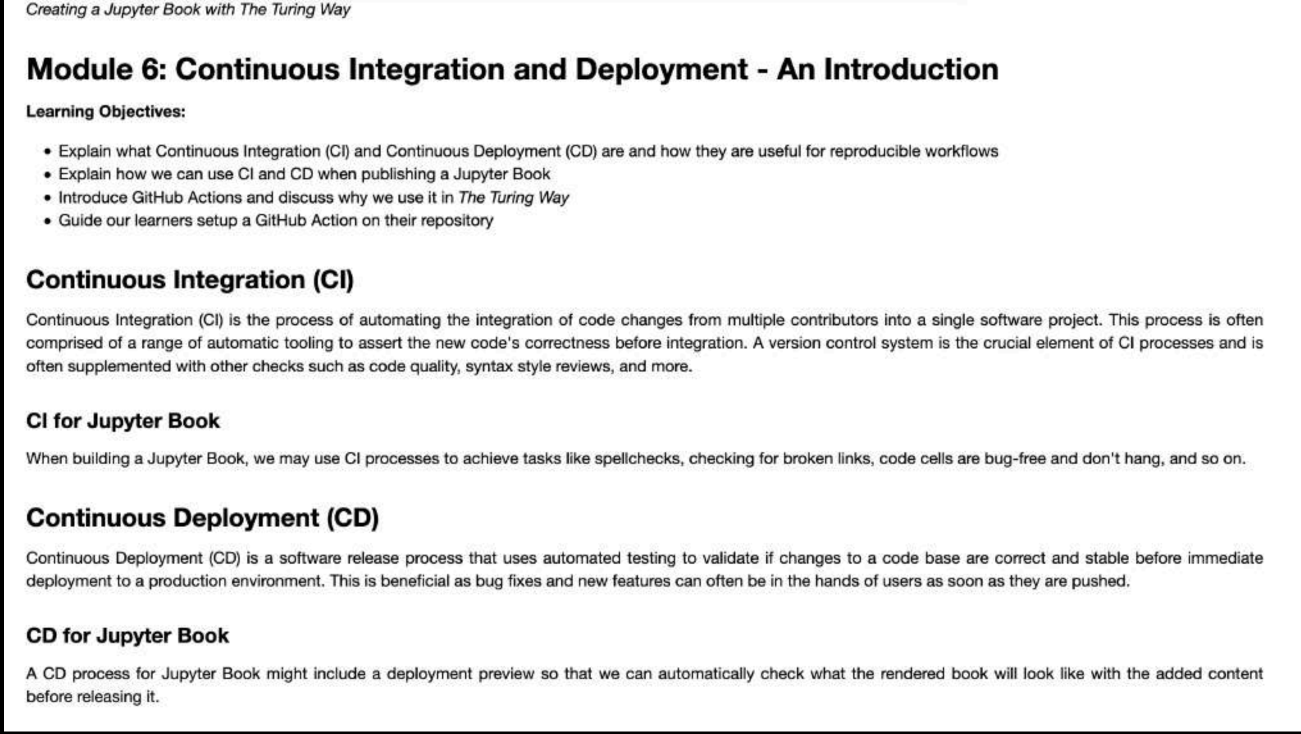
jupytercon

module 6:

Jupyter Book CI/CD

by Sarah Gibson

@sgibson91



Creating a Jupyter Book with The Turing Way

Module 6: Continuous Integration and Deployment - An Introduction

Learning Objectives:

- Explain what Continuous Integration (CI) and Continuous Deployment (CD) are and how they are useful for reproducible workflows
- Explain how we can use CI and CD when publishing a Jupyter Book
- Introduce GitHub Actions and discuss why we use it in *The Turing Way*
- Guide our learners setup a GitHub Action on their repository

Continuous Integration (CI)

Continuous Integration (CI) is the process of automating the integration of code changes from multiple contributors into a single software project. This process is often comprised of a range of automatic tooling to assert the new code's correctness before integration. A version control system is the crucial element of CI processes and is often supplemented with other checks such as code quality, syntax style reviews, and more.

CI for Jupyter Book

When building a Jupyter Book, we may use CI processes to achieve tasks like spellchecks, checking for broken links, code cells are bug-free and don't hang, and so on.

Continuous Deployment (CD)

Continuous Deployment (CD) is a software release process that uses automated testing to validate if changes to a code base are correct and stable before immediate deployment to a production environment. This is beneficial as bug fixes and new features can often be in the hands of users as soon as they are pushed.

CD for Jupyter Book

A CD process for Jupyter Book might include a deployment preview so that we can automatically check what the rendered book will look like with the added content before releasing it.

https://github.com/martinagvilas/tutorial-jupyterbook-with-turing-way/

https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

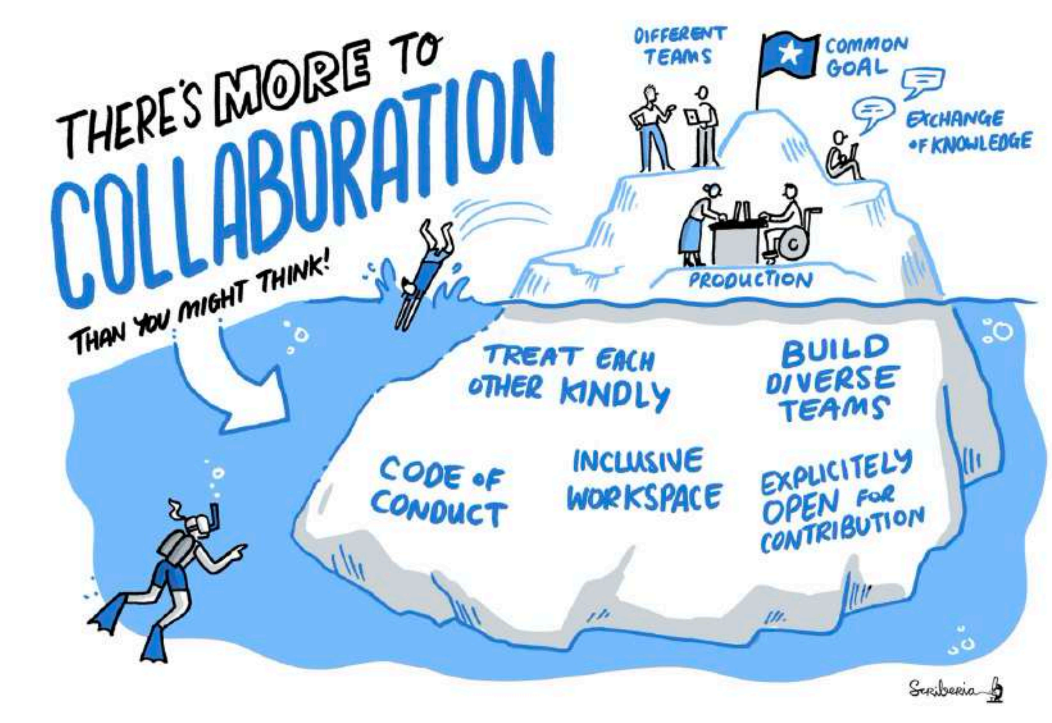
What Else?

- Documentation
- Packaging
- Interactivity
- Collaboration

Check Out other Relevant Resources

- Galaxy pipelines
- Reproducible bioinformatics project: http://www.reproducible-bioinformatics.org/
- Snakemake and Nextflow
- ReproHack: hands-on Reproducibility Hackathon

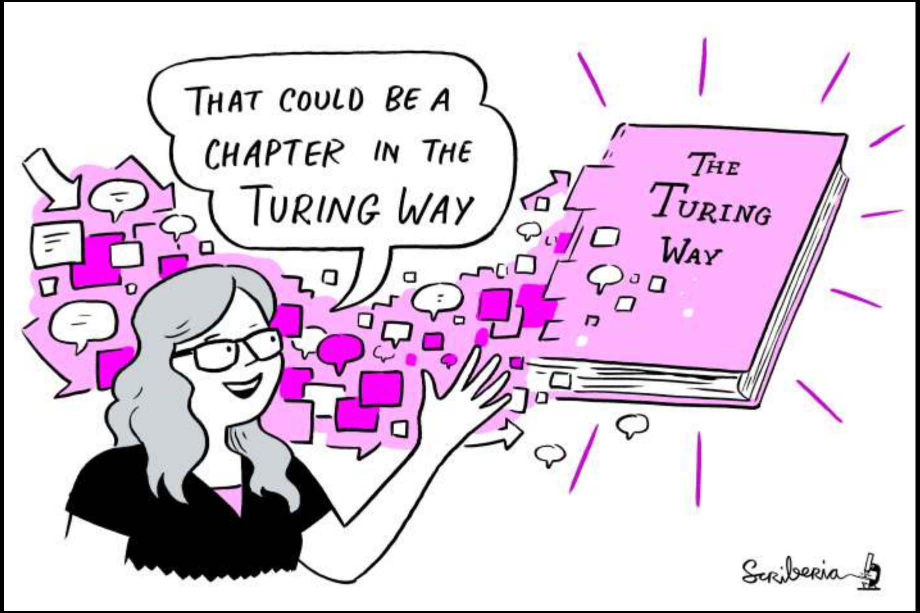
The Turing Way can be beneficial & comprehensible for everyone only when we include diverse perspectives



https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Join us!

- Reading
- Editing
- Writing
- Reviewing
- Linking
- Automating
- Curating
- Translating
- Promoting



https://doi.org/10.5281/zenodo.4549447, @malvikasharan @turingway

Pay it forward!

I realised that the value I left behind in this Open Source project is not in the amount of work I did, but how I enabled other contributors.

Paul Owoicho, GSoD intern

The Alan Turing Institute

Acknowledgements:

- Kirstie Whitaker (@kirstie_j), Project Lead
- Martina G. Vilas (@ martinagvilas), Core Contributor
- *The Turing Way* community, friends & collaborators

Useful links:

- Book: the-turing-way.netlify.com
- Twitter: twitter.com/turingway
- Newsletter: tinyletter.com/TuringWay
- GitHub: github.com/alan-turing-institute/the-turing-way
- Slack: https://tinyurl.com/jointuringwayslack
- Artwork by Scriberia: https://doi.org/10.5281/zenodo.3332808