
The Turing Way Book Dash

The #TuringWay team
University of Manchester: 17 May 2019
Alan Turing Institute: 28 May 2019

Agenda

- 8:30 Arrive at dashing venue
- 9:00 Intro to the book dash
- 9:30 Team forming and brain storming
- 10:30 Coffee break
- 11:00 Group working
- 12:30 Lunch
- 13:30 Group working, coffee available at 15:00
- 16:00 Celebrate successes (group share outs)
- 17:00 Close

Founding the Institute

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

George Osborne, Chancellor of the Exchequer

Budget Speech, March 2014

Network of industry,
charity, government
partners

Network of
university members

Strategic
government
investment



Engineering and Physical Sciences
Research Council

The Institute's partners and collaborators



Lloyd's Register
Foundation



Our university network



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

Tuesday 18 Dec 2018

Learn more ↓

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



Urban analytics



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



Data-centric engineering



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

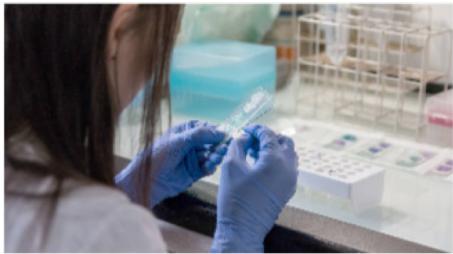


Data science for science



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

Cross cutting theme: Tools, systems and practices



Health



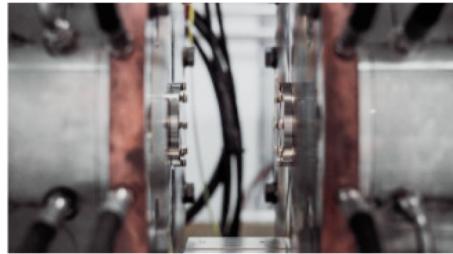
Accelerating the scientific understanding of human disease and improving human health through data-driven innovation in AI and statistical science.



Public policy



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



Research Engineering



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



The Turing Way

A lightly opinionated handbook
for reproducible data science

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

What does reproducible mean?

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

NATURE | NEWS FEATURE

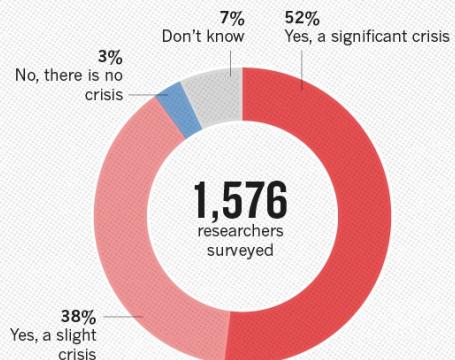
1,500 scientists lift the lid on reproducibility

Survey sheds light on the 'crisis' rocking research.

Monya Baker

25 May 2016 | Corrected: 28 July 2016

IS THERE A REPRODUCIBILITY CRISIS?

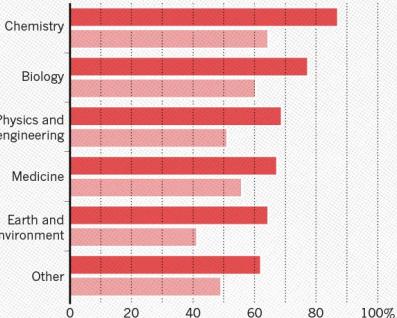


©nature

HAVE YOU FAILED TO REPRODUCE AN EXPERIMENT?

Most scientists have experienced failure to reproduce results.

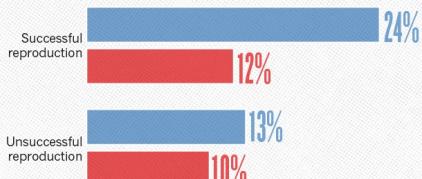
- Someone else's
- My own



HAVE YOU EVER TRIED TO PUBLISH A REPRODUCTION ATTEMPT?

Although only a small proportion of respondents tried to publish replication attempts, many had their papers accepted.

- Published
- Failed to publish

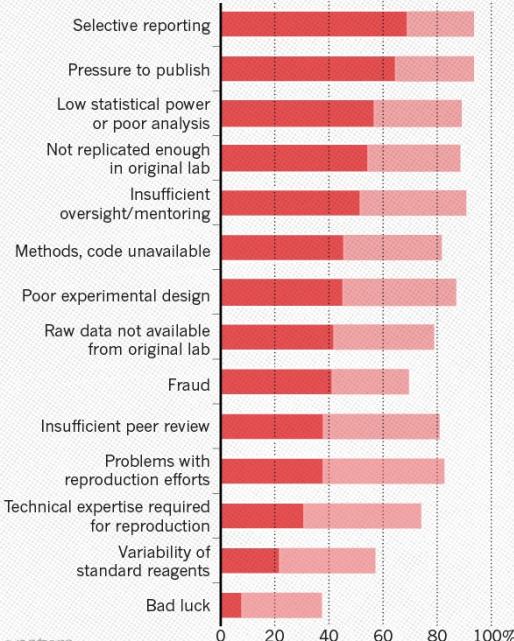


Number of respondents from each discipline:
Biology 703, Chemistry 106, Earth and environmental 95,
Medicine 203, Physics and engineering 236, Other 233

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

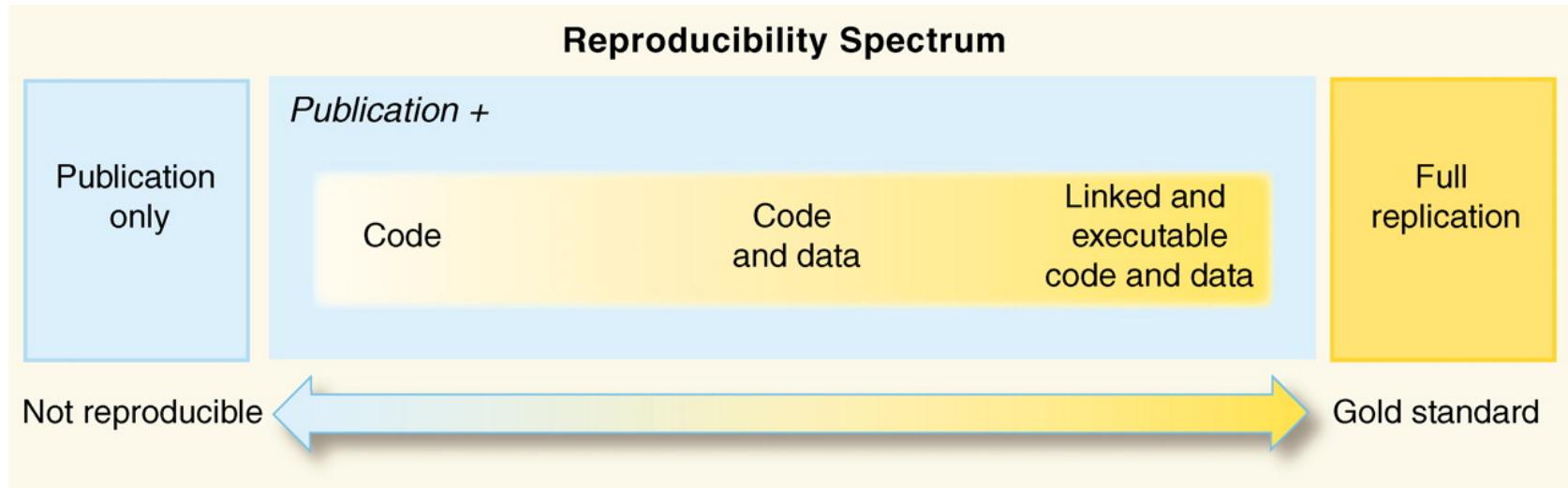
Many top-rated factors relate to intense competition and time pressure.

- Always/often contribute
- Sometimes contribute



©nature

Baker (2016) <https://doi.org/10.1038/533452a>



“Computational science has led to exciting new developments, but the nature of the work has exposed limitations in our ability to evaluate published findings. Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible.”
(Peng 2011; <https://doi.org/10.1126/science.1213847>)

Why don't people do this already?

Is not considered for
promotion

Takes time

Publication bias
towards novel
findings

Requires
additional skills

Barriers to reproducible research

Plead the 5th

Support additional users

Held to higher standards
than others

Why don't people do this already?

Is not considered for
promotion

Takes time

Publication bias
towards novel
findings

Requires
additional skills

Barriers to reproducible research

Plead the 5th

Support additional users

Held to higher standards
than others

How can the Turing Way help?

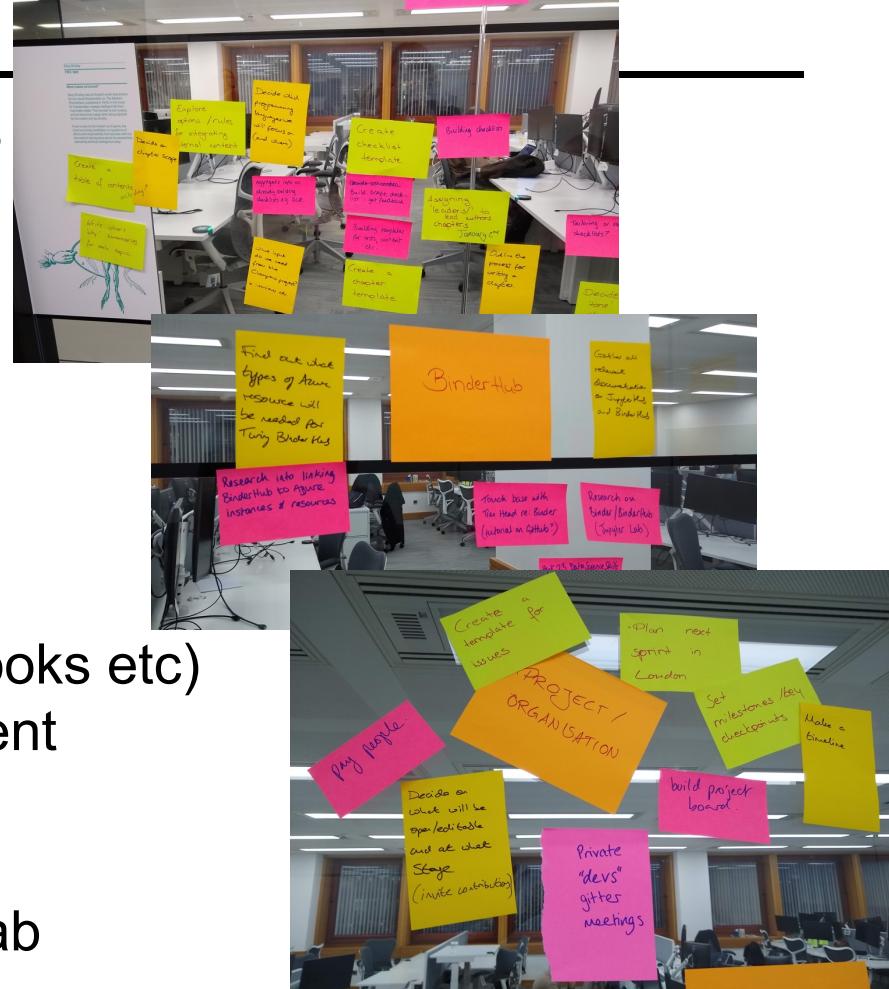
*Make
reproducibility,
“too easy
not to do”*

*Share the
responsibility
of
reproducibility*

Requires additional skills

Chapters will include:

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



Checklists for researcher, PI and admin team

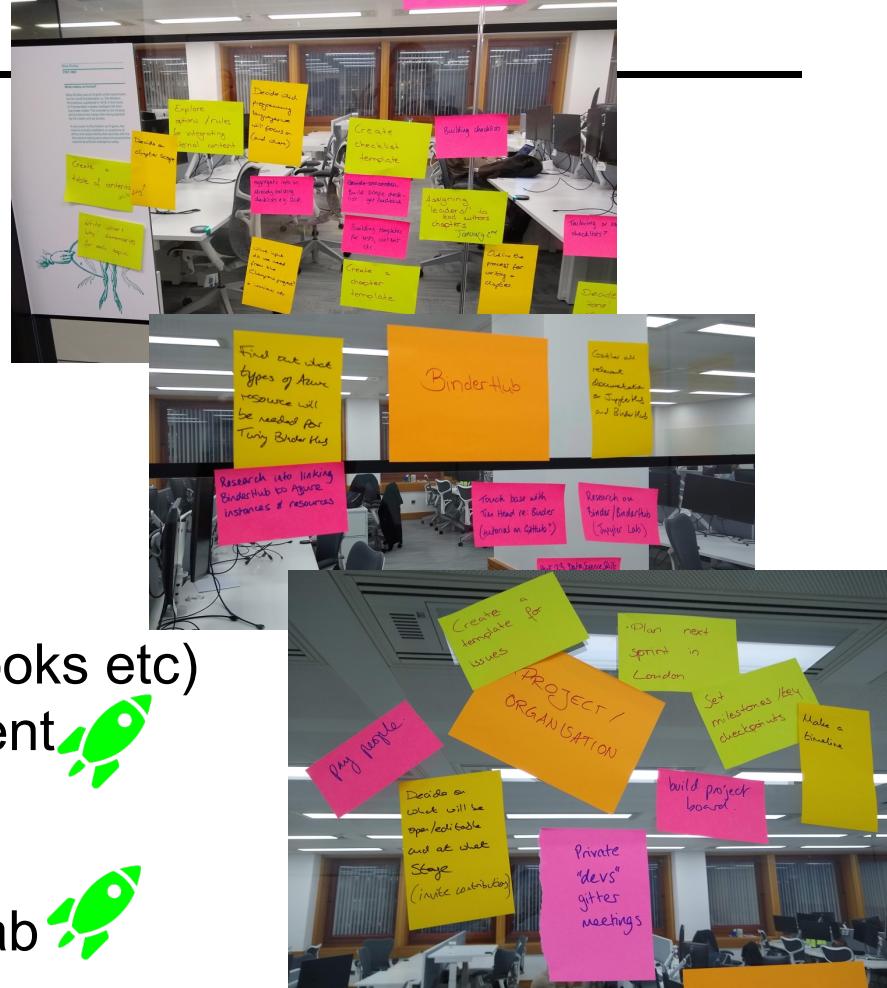


- Researcher
 - Version control
 - Capturing compute environment
 - Writing and running the code
- PI
 - Results presented are those from the final run of the analysis
 - Check that another researcher can run the code
- Admin
 - Version control
 - Data and code archive
 - Open access publication

Requires additional skills

Chapters will include:

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



The Turing Way

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

Powered by [Jupyter Book](#)

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 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>.

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

ON THIS PAGE

A BIT MORE BACKGROUND

THE BOOK ITSELF

THE TURING WAY
COMMUNITY

What is Jupyter Book?

*Build an online book with
Jupyter Notebooks and Markdown*



jupyter.org/jupyter-book

:)

inferentialthinking.com

The screenshot shows a sidebar on the left with a blue circular logo containing 'DATA8'. Below it is a navigation menu:

- Introduction
- Search
- 1. Data Science
- 2. Causality and Experiments
- 3. Programming in Python
- 4. Data Types
- 5. Sequences
- 6. Tables
- 7. Visualization
- 8. Functions and Tables
- 9. Randomness
- 10. Sampling and Empirical Distributions
 - 10.1 Empirical Distributions
 - 10.2 Sampling from a Population
 - 10.3 Empirical Distribution of a Statistic
- 11. Testing Hypotheses

The main content area has a 'TOGGLE SIDEBAR' button and an 'Interact' button with a gear icon.

Sampling and Empirical Distributions

An important part of data science consists of making conclusions based on the data in random samples. In order to correctly interpret their results, data scientists have to first understand exactly what random samples are.

In this chapter we will take a more careful look at sampling, with special attention to the properties of large random samples.

Let's start by drawing some samples. Our examples are based on the `top_movies.csv` data set.

```
top1 = Table.read_table(path_data + 'top_movies.csv')
top2 = top1.with_column('Row Index', np.arange(top1.num_rows))
top = top2.move_to_start('Row Index')

top.set_format(make_array(3, 4), NumberFormatter)
```

Row Index	Title	Studio	Gross	Gross (Adjusted)	Year
0	Star Wars: The Force Awakens	Buena Vista (Disney)	906,723,418	906,723,400	2015
1	Avatar	Fox	760,507,625	846,120,800	2009
2	Titanic	Paramount	658,672,302	1,178,627,900	1997
3	Jurassic World	Universal	652,270,625	687,728,000	2015

ON THIS PAGE

- SAMPLING AND EMPIRICAL DISTRIBUTIONS
- SAMPLING ROWS OF A TABLE
- DETERMINISTIC SAMPLES
- A RANDOM SAMPLING SCHEME
- A SYSTEMATIC SAMPLE
- RANDOM SAMPLES DRAWN WITH OR WITHOUT REPLACEMENT

mypage.ipynb

In [1]:

```
from datascience import *
import numpy as np
import math
np.set_printoptions(suppress=True)
pd.options.display.max_colwidth=1000
```

Sampling and Empirical Distributions

An important part of data science consists of making conclusions based on the data in random samples. In order to correctly interpret their results, data scientists have to first understand exactly what random samples are.

In this chapter we will take a more careful look at sampling, with special attention to the properties of large random samples.

Let's start by drawing some samples. Our examples are based on the [top_movies.csv](#) data set.

In [2]:

```
top1 = Table.read_table(path_data + 'top_movies.csv')
top1 = top1.with_column('Row Index', np.arange(len(top1['Year'])))
top1.set_format(0, NumberFormatter())
out[2]:
```

Row Index	Title	Studio	Gross (Adjusted)
0	Star Wars: The Force Awakens	Buena Vista (Disney)	906,723,418
1	Avatar	Fox	860,507,625
2	Titanic	Paramount	658,672,302
3	Jurassic World	Universal	652,270,625

In [3]:

```
def sample_from(population, size):
    """Return a sample of size n from population.
    population: list or table
    size: int
    """
    if size > len(population):
        raise ValueError("Population too small")
    else:
        return population.sample(size)
```

Sampling and Empirical Distributions

An important part of data science consists of making conclusions based on the data in random samples. In order to correctly interpret their results, data scientists have to first understand exactly what random samples are.

In this chapter we will take a more careful look at sampling, with special attention to the properties of large random samples.

Let's start by drawing some samples. Our examples are based on the [top_movies.csv](#) data set.

mypage.md

[mypage.md](#)

[Sampling and Empirical Distributions](#)

An important part of data science consists of making conclusions based on the data in random samples. In order to correctly interpret their results, data scientists have to first understand exactly what random samples are.

In this chapter we will take a more careful look at sampling, with special attention to the properties of large random samples.

Let's start by drawing some samples. Our examples are based on the [top_movies.csv](#) data set.

mypage.html

TOGGLE SIDEBAR

DATA

Interact

Introduction

Search

1. Data Science

2. Causality and Experiments

3. Programming in Python

4. Data Types

5. Sequences

6. Tables

7. Visualization

8. Functions and Tables

9. Randomness

10. Sampling and Empirical Distributions

10.1 Empirical Distributions

10.2 Sampling from a Population

10.3 Empirical Distribution of a Statistic

11. Testing Hypotheses

ON THIS PAGE

- SAMPLING AND EMPIRICAL DISTRIBUTIONS
- SAMPLING ROWS OF A TABLE
- DETERMINISTIC SAMPLES
- A RANDOM SAMPLING SCHEME
- A SYSTEMATIC SAMPLE
- RANDOM SAMPLES DRAWN WITH OR WITHOUT REPLACEMENT

In this chapter we will take a more careful look at sampling, with special attention to the properties of large random samples.

Let's start by drawing some samples. Our examples are based on the [top_movies.csv](#) data set.

```
top1 = Table.read_table(path_data + 'top_movies.csv')
top2 = top1.with_column('Row Index', np.arange(len(top1['Year'])))
top1 = top2.sample_to_start('Row Index')

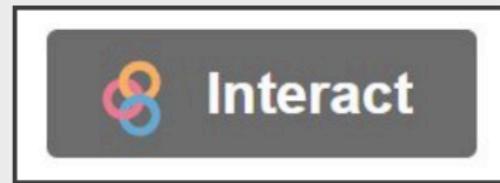
top1.set_format(0, NumberFormatter())
```

Row Index	Title	Studio	Gross (Adjusted)	Year
0	Star Wars: The Force Awakens	Buena Vista (Disney)	906,723,418	2015
1	Avatar	Fox	860,507,625	2009
2	Titanic	Paramount	658,672,302	1997
3	Jurassic World	Universal	652,270,625	2015



Interactive buttons let readers explore

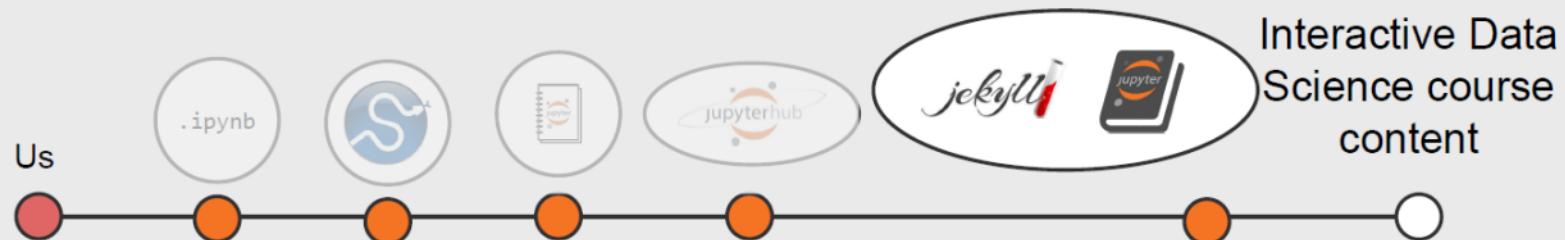
- A page built from a notebook gets an interact button
- Use JupyterHub/Binder to spawn a Jupyter server
- `git pull` the underlying notebook for the page
- Initialize an interactive environment



:)

In summary

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



jupyter.org/jupyter-book



Contributing to The Turing Way

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

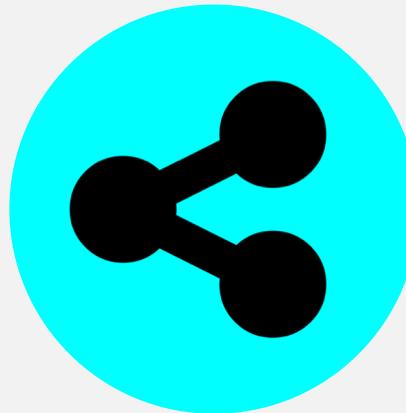
Open Leadership Principles



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

[@kirstie_l](https://doi.org/10.6084/m9.figshare.7564682)



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

Built by a team....and you!

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



Code of Conduct

“The Turing Way team are dedicated to providing a welcoming and supportive environment for all people...we do not tolerate behaviour that is disrespectful to our community members or that excludes, intimidates, or causes discomfort to others.”

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

Code of Conduct

- Be respectful of different viewpoints and experiences.
- Use welcoming and inclusive language.
- Do not harass people.
- Respect the privacy and safety of others – stickers available if you don't want your photo taken
- Be considerate of others' participation.
- Don't be a bystander.

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

Contributing to the Turing Way

🎉🎈🎂 Welcome to the Turing Way repository! 🎉🎈🎂

👋🐥☀️ We're so excited you're here and want to contribute. ☀️🐥👋

The point of this guide is to **welcome new users and contributors to the Turing Way community**. We hope that these guidelines are designed to make it as easy as possible to get involved. Don't let trying to be perfect get in the way of being good - we welcome all contributions and would love it if you could follow these guidelines to make sure your contributions can be easily integrated but exciting 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](#).

Table of contents

Been here before? Already know what you're looking for in this guide? Jump to the following sections:

- [Joining the community](#)
- [Inclusivity](#)
- [Get in touch](#)
- [Contributing through GitHub](#)
- [Writing in Markdown](#)
- [Where to start: issues](#)
- [Making a change with a pull request](#)
- [The process of writing chapters](#)
- [Style Guide](#)
- [Recognising Contributions](#)

Joining the community

The Turing Way is a community-oriented and -led project. We therefore require that all contributions **adhere to our Code of Conduct**.

Inclusivity

This project aims to be inclusive to people from all walks of life and to all research fields. This should be taken into account in contributions.

The following are examples of inclusive actions that we encourage from contributors to the Turing Way:

- Refer to "open research" rather than "open science" so that we do not exclude members of the humanities and social sciences from our community.
- Make sure colour pallettes are accessible to colour-blind readers and contributors. Here's a useful blog post on [tips for designing scientific figures for color blind readers](#) by Luk at [Somersulat 1824](#).

Get in touch

There are many ways to get in touch with the Turing Way team!

- Ping us in our [gitter channel](#).
 - This is our preferred method of open communication and discussion! We'd love for you to swing by to say hello.
- Join the discussion in our [issues](#) and [pull requests](#).
 - Can't find your idea being discussed anywhere? Open a new issue! (See our [Where to start: issues](#) section below.)
- Subscribe to our [mailing list](#) with which we send monthly project updates.
- Check out the [#TuringWay](#) hashtag on Twitter.
- You can contact the PI of the Turing Way project - Kirstie Whitaker - by email at kwhitaker@turing.ac.uk.
- You can also contact members of the Turing Way team through their preferred ways of communication [here](#).

Contributing through GitHub

Git is a really useful tool for version control. GitHub sits on top of Git and supports collaborative and distributed working.

We know that it can be daunting to start using Git and GitHub if you haven't worked with them in the past, but the Turing Way maintainers are here to help you figure out any of the jargon or confusing instructions you encounter! ❤️

In order to contribute via GitHub you'll need to set up a free account and sign in. Here are some [instructions](#) to help you get going. Remember that you can ask us any questions you need to along the way.

Writing in Markdown

GitHub has a helpful page on [getting started with writing and formatting on GitHub](#).

Most of the writing that you'll do will be in [Markdown](#). You can think of Markdown as a few little symbols around your text that will allow GitHub to render the text with a little bit of formatting. For example you could write words as **bold** (****bold****), or in *italics* (**italics**), or as a [link](#) ([link](<https://youtu.be/dQw4w9WgXcQ>)) to another webpage.

Also when writing in Markdown, please start each new sentence on a new line. While this formats in the same way as if the new line wasn't included, it makes the [diffs produced during the pull request](#) review easier to read! ✨

1. Comment on an existing issue or open a new issue referencing your addition

This allows other members of the Turing Way team to confirm that you aren't overlapping with work that's currently underway and that everyone is on the same page with the goal of the work you're going to carry out.

If you open a new issue, please follow the basic guidelines laid out in our issue template. The issue template will automatically be rendered in the comment section of the new issue page so all you need to do is edit the "*Lorem ipsum*" sections.

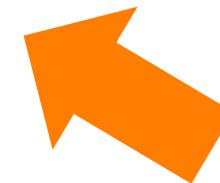
alan-turing-institute / the-turing-way

Code Issues 97 Pull requests 12 Actions Projects 2 Wiki Insights

Label issues and pull requests for new contributors
Now, GitHub will help potential first-time contributors discover issues labeled with **help wanted** or **good first issue**

Filters is:issue is:open Labels 14 Milestones 0 New issue

Open	Closed	Author	Labels	Projects	Milestones	Assignee	Sort
97	95						
Impact Reports for End of Project project management #461 opened 3 hours ago by sgibson91 0 of 1							
Add/link to an intro to the GitHub interface to the Collaborating on GitHub/GitLab chapter book good first issue #460 opened 3 hours ago by rosiehigman							
Check and fix formatting in the RDM chapter book good first issue #459 opened 3 hours ago by rosiehigman 0 of 5							
Promote tools for reproducibility in the book book tools #450 opened 2 days ago by sgibson91 0 of 1							
Add links to prerequisites in each chapter book good first issue #435 opened 7 days ago by rosiehigman 0 of 9							
Turing Way proposal for the Open Science Fair comms #434 opened 7 days ago by rainsworth 3							
Ways to track readers of interactions with handbook? #431 opened 9 days ago by pherterich							
Dissemination report for Turing #429 opened 12 days ago by KirstieJane 2							



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

 alan-turing-institute / the-turing-way

[Unwatch](#) [19](#) [★ Unstar](#) [181](#) [Fork](#) [35](#)

[Code](#) [Issues 97](#) [Pull requests 12](#) [Actions](#) [Projects 2](#) [Wiki](#) [Insights](#)



[Write](#) [Preview](#) [!\[\]\(5fdf67a22d1048a18053c5f2bac7f375_img.jpg\)](#) [!\[\]\(f25daa10f97b691964184f7b675e7463_img.jpg\)](#) [!\[\]\(06c87ff2c9cbec98690b3f9ef52efa0e_img.jpg\)](#) [!\[\]\(d29acc0baab37dc31ac07b04e832bcb3_img.jpg\)](#) [!\[\]\(89b0ce3673b22bd6dabd6f904350c3a1_img.jpg\)](#) [!\[\]\(f86708a570fcb24b61f1e890fae2e9cc_img.jpg\)](#) [!\[\]\(9d679102afd180c5adf1f3afc86237a2_img.jpg\)](#) [!\[\]\(af6c4e96c0ea0569153e7ef8ebe9674b_img.jpg\)](#) [!\[\]\(1495571bd8ebaffb9f96850798141949_img.jpg\)](#) [!\[\]\(f3ba0e4ef7c11ed09d02d20131e94d67_img.jpg\)](#) [!\[\]\(db7bfce78c9322adc4a247769ada8370_img.jpg\)](#)

```
<!--  
Please complete the following sections when you open an issue.  
You are encouraged to keep this top level comment box updated as you  
develop and respond to reviews.  
If you have write access to the repository please also assign the appropriate  
label (or labels) to your issue.  
Note that text within html comment tags will not be rendered.  
-->  
### Summary  
  
<!-- Please provide a detailed description of the change or addition you are proposing, or the  
question you're asking.  
Please provide as much context as possible and link to related issues and/or pull requests.  
-->  
  
*Lorem ipsum dolor sit amet, consectetur adipiscing.*  
  
### What needs to be done?  
  
<!-- We suggest using bullets (indicated by * or -) and filled checkboxes [x] here -->  
  
- [ ] *Lorem ipsum dolor sit amet, consectetur adipiscing.*  
- [ ] *Lorem ipsum dolor sit amet, consectetur adipiscing.*
```

Attach files by dragging & dropping, selecting or pasting them.

 Styling with Markdown is supported

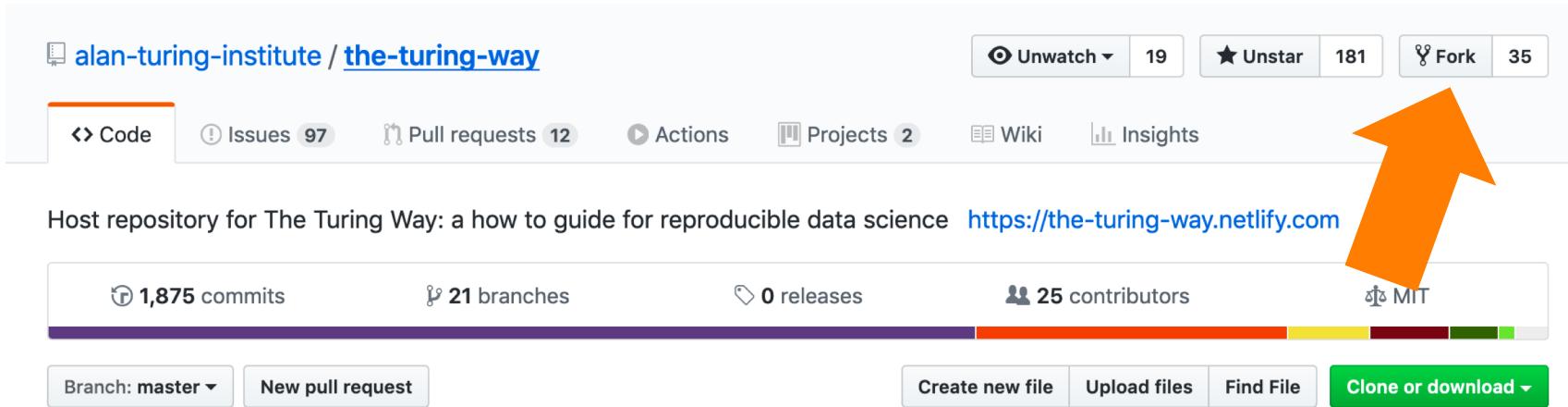
[Submit new issue](#)

1. Add descriptive title.
 2. Fill in the issue template by replacing the *Placeholder text* sections with what needs to be done or fixed.
 3. Add assignees and labels where relevant (ignore projects and milestones).
 4. Submit issue and off you go!

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

2. Fork the Turing Way repository to your profile

This is now your own unique copy of the Turing Way. Changes here won't affect anyone else's work, so it's a safe space to explore edits to the code!



The screenshot shows a GitHub repository page for 'alan-turing-institute / the-turing-way'. The page includes the following elements:

- Repository Header:** Shows the repository name 'alan-turing-institute / the-turing-way' and a fork count of '35'.
- Top Navigation:** Includes links for 'Unwatch' (with 19 notifications), 'Unstar' (with 181 notifications), and a large 'Fork' button.
- Main Navigation:** Includes links for 'Code', 'Issues (97)', 'Pull requests (12)', 'Actions', 'Projects (2)', 'Wiki', and 'Insights'.
- Description:** A brief description: 'Host repository for The Turing Way: a how to guide for reproducible data science' followed by a link: <https://the-turing-way.netlify.com>.
- Statistics:** Shows 1,875 commits, 21 branches, 0 releases, 25 contributors, and an MIT license.
- Branch Selection:** 'Branch: master ▾'.
- Action Buttons:** 'New pull request', 'Create new file', 'Upload files', 'Find File', and a large green 'Clone or download ▾' button.

<https://github.com/alan-turing-institute/the-turing-way/blob/master/CONTRIBUTING.md>

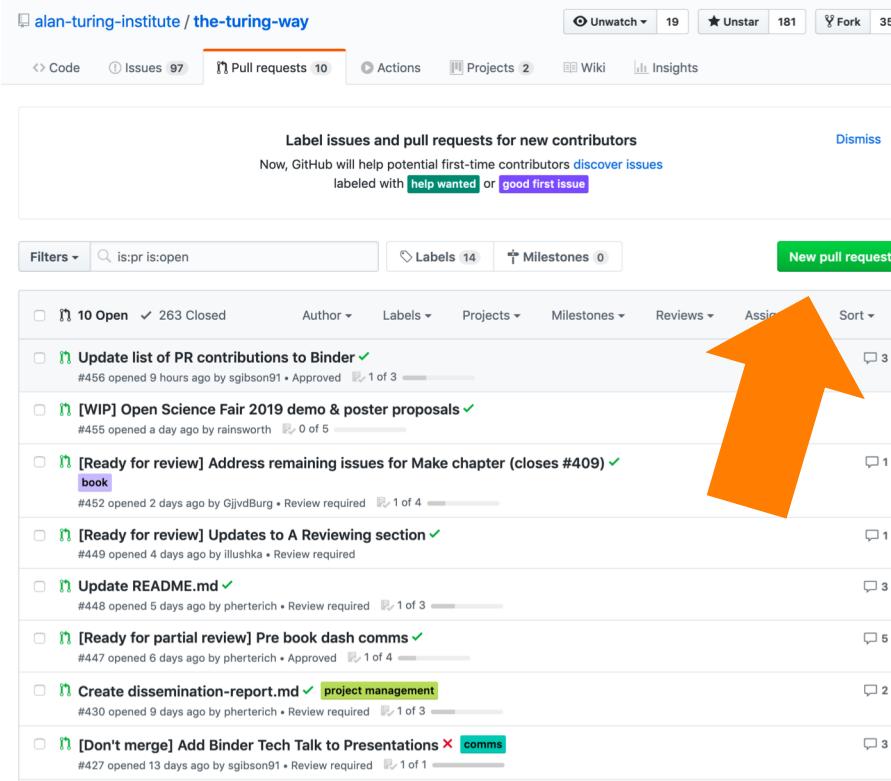
3. Make the changes you've discussed

Try to keep the changes focused. If you submit a large amount of work all in one go it will be much more work for whomever is reviewing your pull request.

While making your changes, commit often and write good, detailed commit messages. It is also perfectly fine to have a lot of commits - including ones that break code. A good rule of thumb is to push up to GitHub when you *do* have passing tests then the continuous integration (CI) has a good chance of passing everything.

If you feel tempted to "branch out" then please make a [new branch](#) and a [new issue](#) to go with it.

4. Submit a pull request



A screenshot of a GitHub repository page showing the pull requests section. At the top, there are buttons for Unwatch (19), Star (181), Fork (35). Below that, a navigation bar includes Code, Issues (97), Pull requests (10), Actions, Projects (2), Wiki, and Insights. A modal window titled "Label issues and pull requests for new contributors" is open, stating "Now, GitHub will help potential first-time contributors discover issues labeled with [help wanted](#) or [good first issue](#)". The main area shows a list of pull requests with various status icons and labels like "WIP", "Ready for review", and "Don't merge". An orange arrow points from the bottom right towards the "New pull request" button at the top right of the list.

Author	Labels	Milestones	Reviews	Assignee	Sort
sgibson91	14	0	1 of 3		3
sgibson91	14	0	1 of 3		1
rainsworth	14	0	0 of 5		3
gjvdburg	14	0	1 of 4		3
illushka	14	0	1 of 4		1
pherterich	14	0	1 of 3		3
pherterich	14	0	1 of 4		5
pherterich	14	0	1 of 3		2
sgibson91	14	0	1 of 1		3

We encourage you to open a pull request as early in your contributing process as possible.

This allows everyone to see what is currently being worked on.

It also provides you, the contributor, feedback in real time from both the community and the continuous integration as you make commits (which will help prevent stuff from breaking).

When you are ready to submit a pull request, you will automatically see the [Pull Request Template](#) contents in the pull request body. It asks you to:

- Describe the problem you're trying to fix in the pull request.
- Reference any related issue and use “Fixes #” to automatically close them.
- List of changes proposed in the pull request.
- Describe what the reviewer should concentrate their feedback on.

Fill out the "*Lorem ipsum*" sections of the pull request template with as much detail as possible to make it really easy for someone to review your contribution!

Add "[WIP]" at the start of the pull request title to indicate "Work in Progress".

When you are happy for it to be merged into the main repository, change the "[WIP]" in the title of the pull request to "[Ready for review]".

The screenshot shows a GitHub repository page for 'alan-turing-institute / the-turing-way'. The top navigation bar includes links for Code, Issues (97), Pull requests (10), Actions, Projects (2), Wiki, and Insights. The repository stats show 19 stars, 181 forks, and 35 open issues. Below the header, a section titled 'Open a pull request' instructs users to 'Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks.' A dropdown menu shows 'base repository: alan-turing-institute/the-turing-way' and 'base: master'. Another dropdown shows 'head repository: rainsworth/the-turing-way' and 'compare: osf-proposals'. A green success message says 'Able to merge. These branches can be automatically merged.' The main area contains a comment box with placeholder text: '<!-- Please complete the following sections when you submit your pull request. You are encouraged to keep this top level comment box updated as you develop and respond to reviews. Note that text within html comment tags will not be rendered. --> ### Summary'. Below the comment box is a file upload section with the instruction 'Attach files by dragging & dropping, selecting or pasting them.' At the bottom are two buttons: 'Allow edits from maintainers. Learn more' and a large green 'Create pull request' button.

1. Select which branches you want to merge (the-turing-way branch ← your branch).
2. Add descriptive title with [WIP] or [Ready for review].
3. Fill in the pull request template by replacing the *Lorem ipsum* sections with what has been done and reference relevant issue(s).
4. Create pull request!

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

A member of the Turing Way team will then review your changes to confirm that they can be merged into the main repository. A [review](#) will probably consist of a few questions to help clarify the work you've done. Keep an eye on your GitHub notifications and be prepared to join in that conversation.

You can update your [fork](#) of the Turing Way [repository](#) and the pull request will automatically update with those changes. You don't need to submit a new pull request when you make a change in response to a review.

You can also submit pull requests to other contributors' branches! Do you see an [open pull request](#) that you find interesting and want to contribute to? Simply make your edits on their files and open a pull request to their branch!

The Turing Way

- 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

Powered by [Jupyter Book](#)

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 “research pipeline” they can affect, and what they should do to nudge data science to be 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 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>.

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

ON THIS PAGE

A BIT MORE BACKGROUND

THE BOOK ITSELF

THE TURING WAY
COMMUNITY

What types of contributions can you make?

- Proofread existing chapters (check for typos, grammar, links, etc.)
- Add further information to an existing chapter
- Suggest topics for new chapters (particularly if you can write them!)
- Review [Ready to review] pull requests
- Contribute to discussion [WIP] pull requests and issues
- Submit a case study or your tips and tricks for reproducible research via our Google submission form (link in README).
- Checklists
- Let us know if you are struggling with contributing in any way so that we can improve our Contributing Guidelines!

What types of contributions can you make?

- Proofread existing chapters (check for typos, grammar, etc.)
- Add further information to an existing chapter
- Suggest topics for new chapters (research areas, tools and issues)
- Review [Ready to review] pull requests
- Contribute to discussions on GitHub and issues
- Submit a contribution to the repository (e.g., tips and tricks for reproducible research)

NO CONTRIBUTION IS TOO SMALL!

If you are struggling with contributing in any way so we can improve our Contributing Guidelines!

The emoji key to celebrate our contributors

Dr. Rachael Ainsworth 	Becky Arnold 	Louise Bowler 	Jason M. Gates 	Sarah Gibson 	Richard Gilham 	Tim Head
Patricia Herterich 	Rosie Higman 	Dan Hobley 	Chris Holdgraf 	Anna Krystalli 	Robin Long 	Alexander Morley
Martin O'Reilly 	Sarah Stewart 	Oliver Strickson 	Kirstie Whitaker 	forrest 	Gertjan van den Burg 	Hieu Hoang
Stephen Eglen 	Andrew Stewart 	Tania Allard 	Ian Hinder 	Tarek 	Stephan Druskat 	Ali Seyhun Saral

Emoji	Represents
	Answering Questions (on gitter, GitHub, or in person)
	Bug reports
	Blogposts
	Code
	Documentation and specification
	Design
	Examples
	Event Organizers
	Financial Support
	Funding/Grant Finders
	Ideas & Planning
	Infrastructure (Hosting, Build-Tools, etc)
	Plugin/utility libraries
	Reviewed Pull Requests
	Tools
	Translation
	Tests
	Tutorials
	Talks
	Videos

Neurohackademy

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



<https://neurohackademy.org>
[#csvconf #TuringWay @kirstie_j](#)
<https://doi.org/10.5281/zenodo.2669548>

Agenda

- 8:30 Arrive at dashing venue
- 9:00 Intro to the book dash
- 9:30 Team forming and brain storming
- 10:30 Coffee break
- 11:00 Group working
- 12:30 Lunch
- 13:30 Group working, coffee available at 15:00
- 16:00 **Celebrate successes (group share outs)**
- 17:00 Close

The Turing Way



#TuringWay



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



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