
The Turing Way Book Dash

The #TuringWay team

University of Manchester: 17 May 2019

<http://bit.ly/book-dash-mcr>

Agenda

<http://bit.ly/book-dash-mcr>

- 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

**The
Alan Turing
Institute**

EPSRC
Engineering and Physical Sciences
Research Council

Network of industry,
charity, government
partners

Network of
university members

Strategic
government
investment

The Institute's partners and collaborators



Our university network



The 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 |

Why don't people do this already?

Is not considered for
promotion

Takes time

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional skills

Plead the 5th

Support additional users

Held to higher standards
than others

How can the Turing Way help?

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

*Share the
responsibility
of
reproducibility*

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

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


What is Jupyter Book?

*Build an online book with
Jupyter Notebooks and Markdown*




jupyter.org/jupyter-book





← TOGGLE SIDEBAR

 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

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')

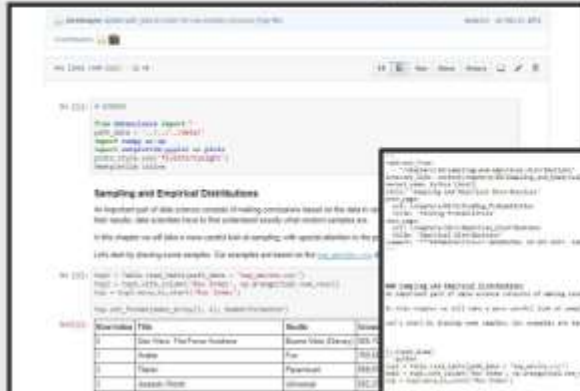
top.set_format(make_array(3, 4), NumberFormatter)
```

| 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 SAMPLING
- K-RANDOM SAMPLING
- SCHEMES
- A STATISTICAL SAMPLE
- RANDOM SAMPLES DRAWN WITH OR WITHOUT REPLACEMENT

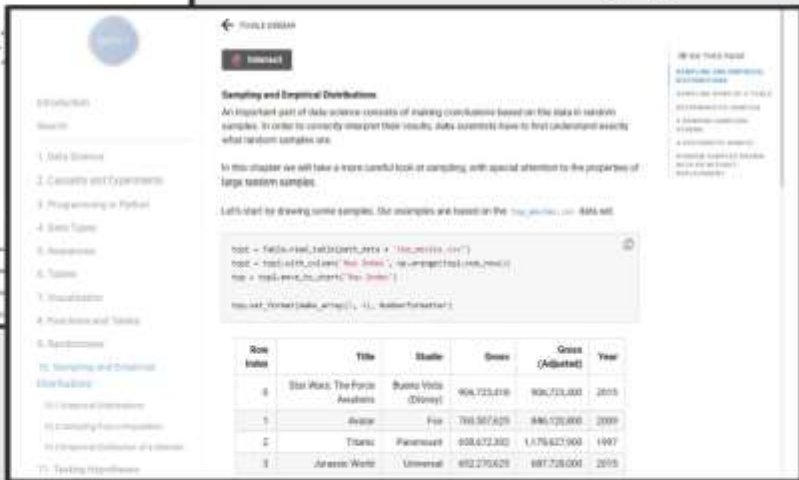
mypage.ipynb



mypage.md



mypage.html



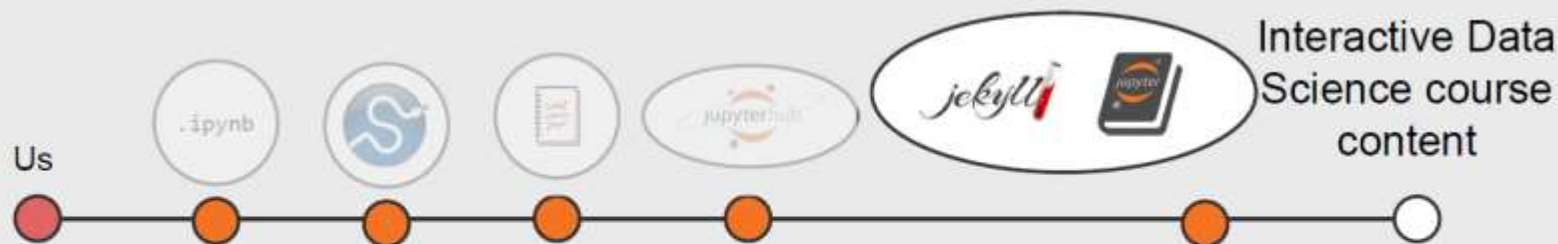
Interactive buttons let readers explore

- 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



Participation & Inclusion

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

@kirstie_j

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

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
- . . .



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

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](#)

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

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 palettes 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](#).

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

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! ✨

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

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

Unwatch 19 Unstar 181 Fork 35

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** [Dismiss](#)

Filters Labels 14 Milestones 0 [New issue](#)

☐ 97 Open ☒ 95 Closed Author Labels Projects Milestones Assignee Sort

- ☐ **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 roselhigman
- ☐ **Check and fix formatting in the RDM chapter** **book** **good first issue**
#459 opened 3 hours ago by roselhigman 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 roselhigman 0 of 9
- ☐ **Turing Way proposal for the Open Science Fair** **commu**
#434 opened 7 days ago by rainworth 3
- ☐ **Ways to track readers of/ interactions with handbook?**
#431 opened 8 days ago by pharierich
- ☐ **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 36

Code Issues 97 Pull requests 12 Actions Projects 2 Wiki Insights

Title

Write Preview

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

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Helpful resources

Contributing

Code of conduct

1. Add descriptive title.
2. Fill in the issue template by replacing the *Lorem ipsum* sections with what needs to be done or fixed.
3. Add assignees and labels where relevant (ignore projects and milestones).
book-dash-mcr
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!



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

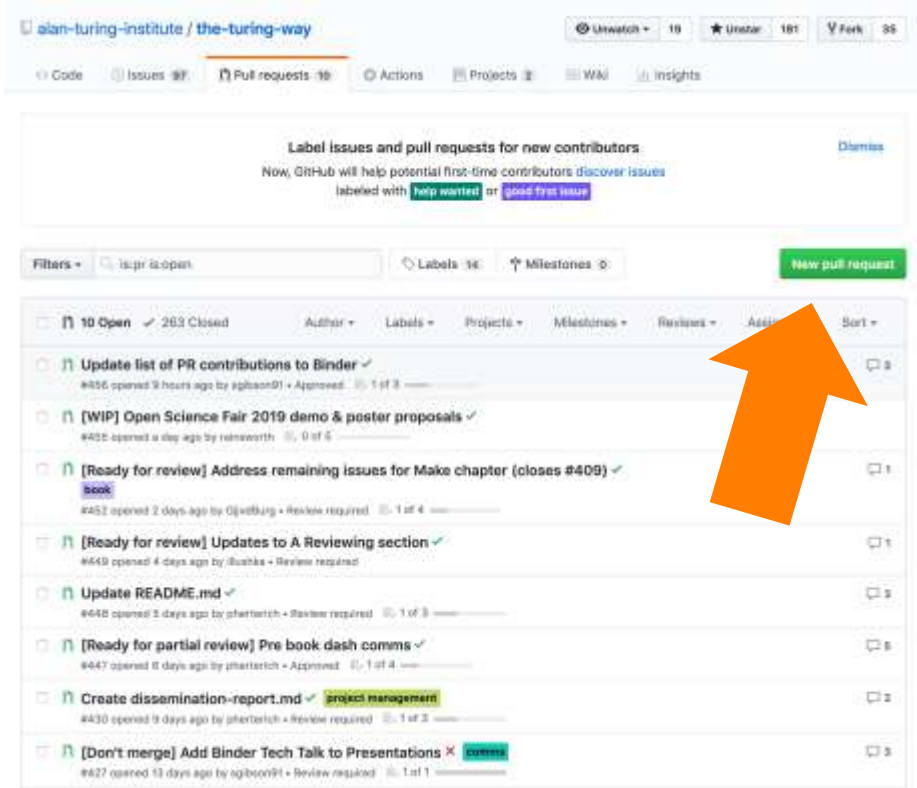
3. Make the changes you've discussed

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.

If you feel tempted to "branch out" then please make a new branch and a new issue to go with it – for example, if writing a new chapter, create a new branch in the Turing Way repository to pull to, named with the chapter title.

4. Submit a pull request



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.

You can also submit pull requests to other contributors' branches!

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



Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also [compare across forks](#).

base repository: alan-turing-institute/the-turing-way base: master

head repository: rainsworth/the-turing-way compare: osf-proposals ✓ Able to merge. These branches can be automatically merged.

[WIP] Open Science Fair proposals

Write Preview

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


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

Allow edits from maintainers. [Learn more](#) Create pull request


1. Select which branches you want to merge: the-turing-way branch ← your branch
2. Add descriptive title with [WIP] or [Ready for review] tag.
3. Fill in the pull request template, replacing the *Lorem ipsum* sections with what has been done & reference relevant issue.
4. Add **book-dash-mcr** label & create pull request!

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










 **Review required** [Show all reviewers](#)

At least 1 approving review is required by reviewers with write access. [Learn more.](#)

 **All checks have passed** [Hide all checks](#)

3 neutral and 3 successful checks

| | |
|---|---|
| ■ |  Header rules Completed in 54s — No header rules processed Details |
| ■ |  Pages changed Completed in 54s — 19 new files uploaded Details |
| ■ |  Redirect rules Completed in 54s — No redirect rules processed Details |
| ✓ |  Mixed content Successful in 54s — No mixed content detected Details |
| ✓ |  continuous-integration/travis-ci/pr — The Travis CI build passed Details |
| ✓ |  denlow/netlify — Denlow preview rebuilt Details |

 **Merging is blocked**

Merging can be performed automatically with 1 approving review.

Merge pull request ▾

You can also [open this in GitHub Desktop](#) or view [command line instructions](#).

A member of the Turing Way team will then review your changes to confirm that they can be merged into the main repository.

You don't need to submit a new pull request when you make a change in response to a review.

You can update your fork of the Turing Way repository and the pull request will automatically update with those changes and the CI will feedback.

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 "reproducibility" they can affect, and what they should do to nudge data science to be 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>.

ON THIS PAGE

A BIT MORE BACKGROUND

THE BOOK ITSELF

THE TURING WAY
COMMUNITY

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

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
- Restructuring the chapter sections (like the reproducibility chapter)
- Let us know if you are struggling with contributing in any way so that we can improve our Contributing Guidelines!

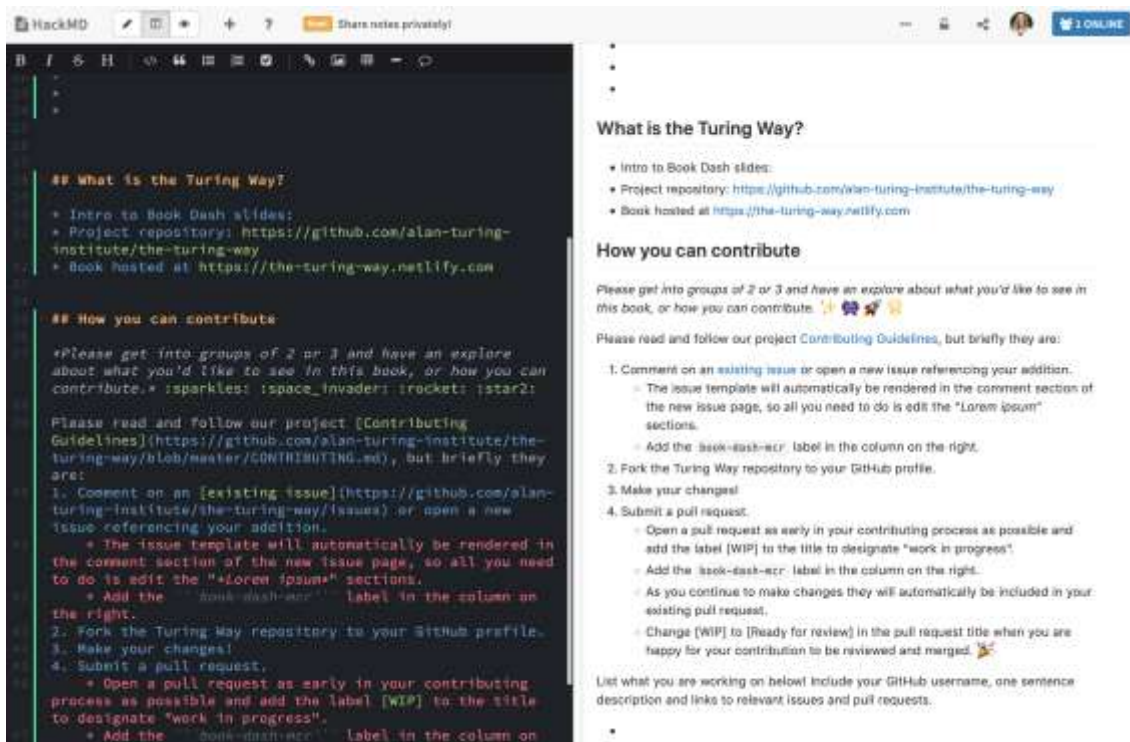
What types of contributions can you make?

- Proofread existing chapters (check for typos, grammar, etc.)
- Add further information to an existing chapter
- Suggest topics for new chapters (particularly in the field of AI ethics!)
- Review [Ready to review] pull requests
- Contribute to discussion pages on the book's issues
- Submit a case study or example of reproducible research with a discussion form (link in README).
- Check for consistency across chapters

NO CONTRIBUTION IS TOO SMALL!

Let us know if you are struggling with contributing in any way so that we can improve our Contributing Guidelines!

Update the HackMD: <http://bit.ly/book-dash-mcr>



The screenshot shows a HackMD document with the following content:

What is the Turing Way?

- Intro to Book Dash slides:
- Project repository: <https://github.com/alan-turing-institute/the-turing-way>
- Book hosted at <https://the-turing-way.netlify.com>

How you can contribute

Please get into groups of 2 or 3 and have an explore about what you'd like to see in this book, or how you can contribute. ✨🚀🚀

Please read and follow our project [Contributing Guidelines](#), but briefly they are:

1. Comment on an [existing issue](#) (<https://github.com/alan-turing-institute/the-turing-way/issues>) or open a new issue referencing your addition.
 - 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.
 - Add the `book-dash-mcr` label in the column on the right.
2. Fork the Turing Way repository to your GitHub profile.
3. Make your changes!
4. Submit a pull request.
 - Open a pull request as early in your contributing process as possible and add the label [WIP] to the title to designate "work in progress".
 - Add the `book-dash-mcr` label in the column on the right.
 - As you continue to make changes they will automatically be included in your existing pull request.
 - Change [WIP] to [Ready for review] in the pull request title when you are happy for your contribution to be reviewed and merged. 🎉

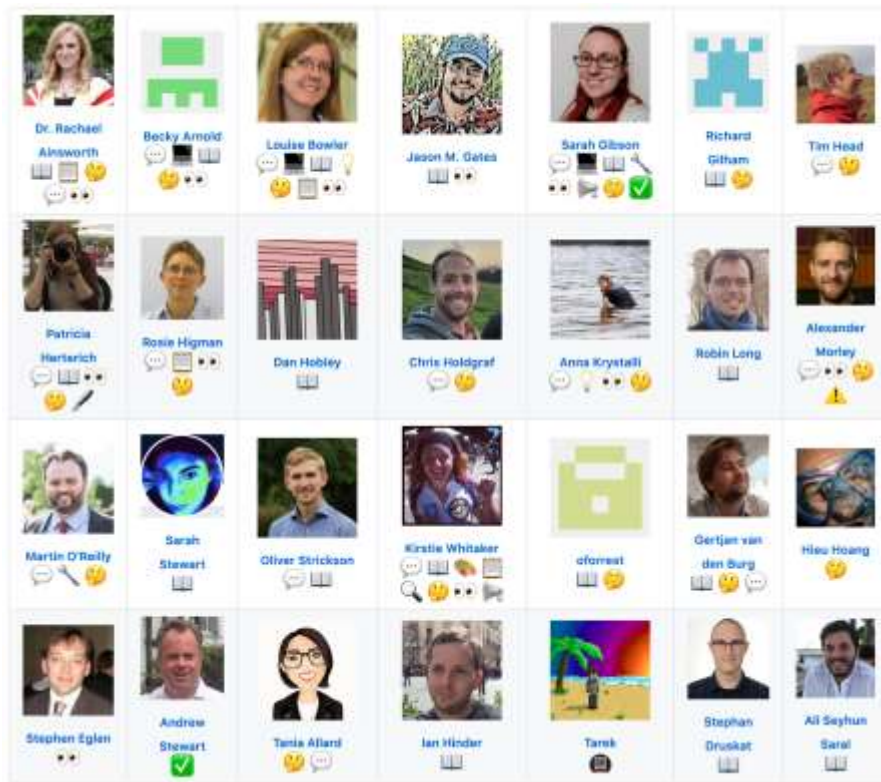
List what you are working on below! Include your GitHub username, one sentence description and links to relevant issues and pull requests.

Add what you're working on, any questions you have or anything you've learned!

Link to your issues and pull requests.

This will help us write up our Book Dash Report.

The emoji key to celebrate our contributors



| 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 |

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

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>

Scriberia



WE HELP OUR CLIENTS TO THINK, WORK AND COMMUNICATE IN A VISUAL WORLD.

Agenda

<http://bit.ly/book-dash-mcr>

- 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

Agenda

<http://bit.ly/book-dash-mcr>

- 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