How-to-Reprohack

Kristina Hettne | Daniela Gawehns | Leiden University

LCRDM – Network Meeting 2022

29 March 2022

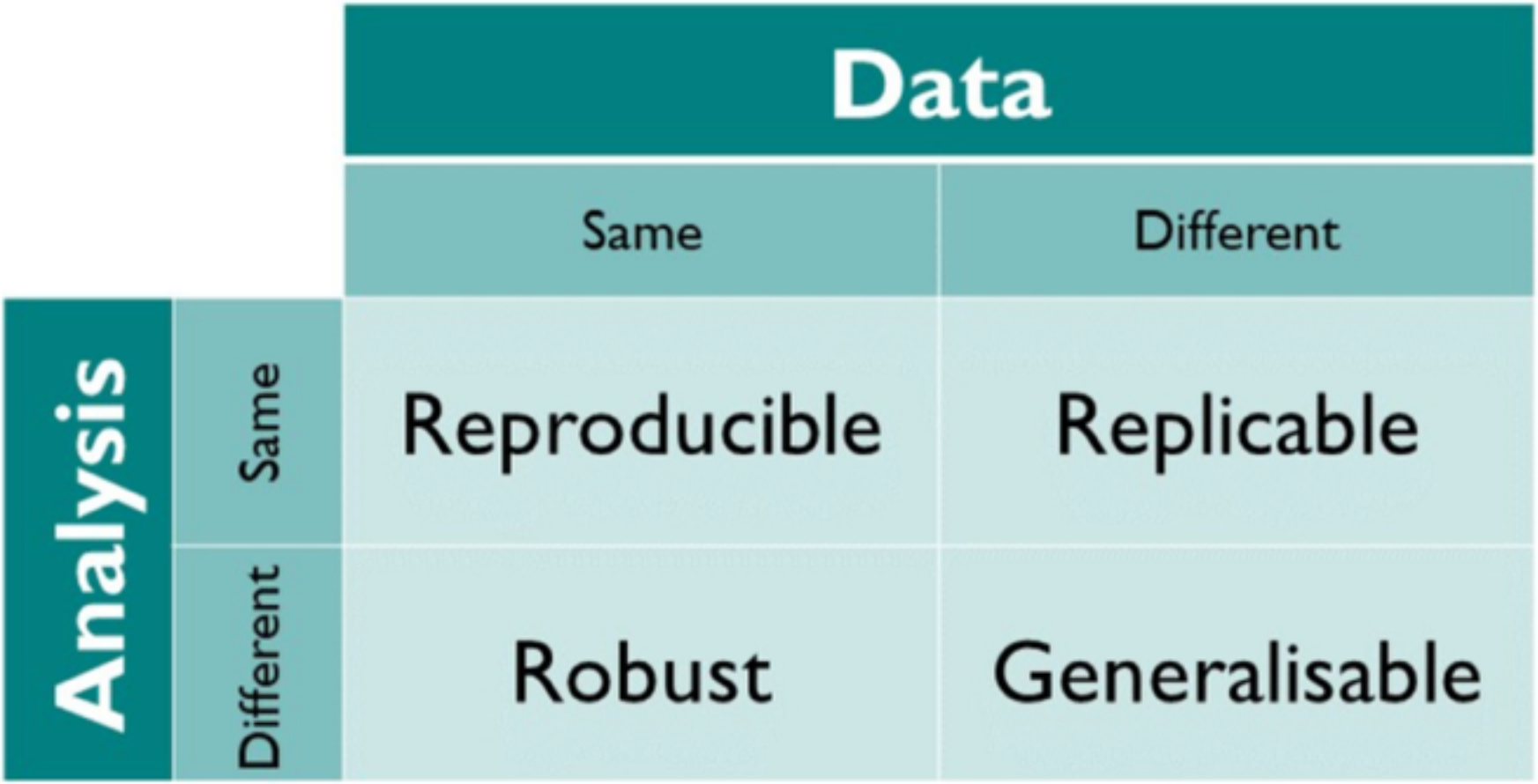
Universiteit Leiden
The Netherlands

Who are we?

Daniela Gawehns, PhD candidate
Leiden University

Kristina Hettne, Digital Scholarship Librarian
Leiden University Library

What is computational reproducibility?



The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, … Kirstie Whitaker. (2019, March 25). The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4). Zenodo. http://doi.org/10.5281/zenodo.3233986

How does Reprohack work and why is it awesome?

ReproHack events provide a low-pressure sandbox environment for practicing reproducible research practices!

Participant – learn/practice/try

Furthering Science – increase skills for reviewing and producing reproducible research

Author– get feedback/acknowledgment for efforts

What will you do during this workshop?

- **Learn about the Reprohack Hub**
- **Plan your own Reprohack (Group Assignment)**
- **Get all the contact info you need**

The Reprohack Hub

Participants: access papers, upload feedback

Organizers: event platform, materials

Authors: access feedback, upload papers

Group Assignment

Plan your own reprohack in a few steps:

- **1. Use the "constraints" or decide as a group what kind of reprohack you want to work on:**
  - **A research group one**
  - **A faculty-wide one**
  - **One for new PI's at your institute/university/faculty**
- **2. Go through the worksheet and fill in as many details as possible together**
- **3. if you have time: Come up with 5 communication channels to attract participants and 5 channels to attract paper donations**

Where can you find help/inspiration/support?

- Hub and organizer guidelines/info: https://www.reprohack.org
- Twitter: @ReproHack
- Slack Invite: https://reprohack-autoinvite.herokuapp.com/
- Email address Reprohack: team@reprohack.org

- Reading material: Paper in IASSIST Quarterly: https://doi.org/10.29173/iq977

Open questions/wishes/feedback!

Daniela Gawehns
Email: d.gawehns@liacs.leidenuniv.nl
Twitter: @dgawehns

Kristina Hettne
Email: k.m.hettne@library.leidenuniv.nl
Twitter: @KristinaHettne

Thank you

Definitions

- **Reproducible:** A result is reproducible when the same analysis steps performed on the same dataset consistently produces the same answer.
- **Replicable:** A result is replicable when the same analysis performed on different datasets produces qualitatively similar answers.
- **Robust:** A result is robust when the same dataset is subjected to different analysis workflows to answer the same research question (for example one pipeline written in R and another written in Python) and a qualitatively similar or identical answer is produced. Robust results show that the work is not dependent on the specificities of the programming language chosen to perform the analysis.
- **Generalisable:** Combining replicable and robust findings allow us to form generalisable results. Note that running an analysis on a different software implementation and with a different dataset does not provide generalised results. There will be many more steps to know how well the work applies to all the different aspects of the research question. Generalisation is an important step towards understanding that the result is not dependent on a particular dataset nor a particular version of the analysis pipeline.

The Turing Way Community, Becky Arnold, Louise Bowler, Sarah Gibson, Patricia Herterich, Rosie Higman, … Kirstie Whitaker. (2019, March 25). The Turing Way: A Handbook for Reproducible Data Science (Version v0.0.4). Zenodo. http://doi.org/10.5281/zenodo.3233986