How to make research **reproducible?**

Agata Bochynska, PhD
Open Research and Digital Scholarship Center
University of Oslo Library

@AgataBochynska
agata.bochynska@ub.uio.no

UNIVERSITY
OF OSLO

Bochynska, A (2023). *How to make research reproducible?*

Open and reproducible research courses

Sep 8th – 20th 2023

Time and place: Sep. 8, 2023 10:00 AM – 11:00 AM, Zoom

Open and reproducible research: An overview

Learn about what open research is and how to make your own research more transparent and reproducible.

Time and place: Sep. 13, 2023 9:00 AM – 11:00 AM, Zoom

How to preregister research studies?

Learn about what preregistration is and how to preregister your own studies.

Time and place: Sep. 15, 2023 10:00 AM – 11:00 AM, Zoom

How to make research reproducible?

Learn about tools and practices for more reproducible and effective research.

Time and place: Sep. 18, 2023 10:00 AM – 11:30 AM, Zoom

How to publish openly?

Learn about preprints, peer-review process, Open Access and how can you choose the best way to publish your results openly.

Time and place: Sep. 20, 2023 10:00 AM – 11:30 AM, Zoom

How to make research more visible?

Learn about different tools, platforms and services to share your research and other contributions, and how you utilise them to make yourself and your work more visible to the academic community and the society at large.

https://www.ub.uio.no/english/courses-events/courses/other/research-reproducibility/

Roadmap

- Some definitions
- Open research and reproducibility
- Reproducible data acquisition, processing, analyses and reports/publications (with some useful tools)
- Take-aways
- Q&A time!

		DATA	
		Same	Different
ANALYSIS	Same	**Reproduced**	**Replicated**
	Different	**Robust**	**Generalized**

UNIVERSITY
OF OSLO

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

Reproduced

results are consistent when following the same method and analysis steps with the **same input data**

https://www.nap.edu/catalog/25303

Quantitative studies

computational reproducibility

Qualitative studies

process transparency

Quantitative studies
computational reproducibility

“obtaining consistent computational results using the same input data, computational steps, methods, code, and conditions of analysis”

https://www.nap.edu/catalog/25303

Quantitative studies
computational reproducibility

Re-running analyses/code with the same data

Qualitative studies
process transparency

Arriving at similar (consistent) interpretation by following the same analysis process

Qualitative studies
process transparency

Following the step-by-step reasoning and interpretation process

https://blog.ukdataservice.ac.uk/show-me-the-data/

Reproducibility is strongly associated with **transparency**

“**Open Science** has the potential of making the scientific process more **transparent**, **inclusive** and **democratic**. It is (…) a true game changer in bridging the science, technology and innovation gaps and fulfilling the **human right to science**.”

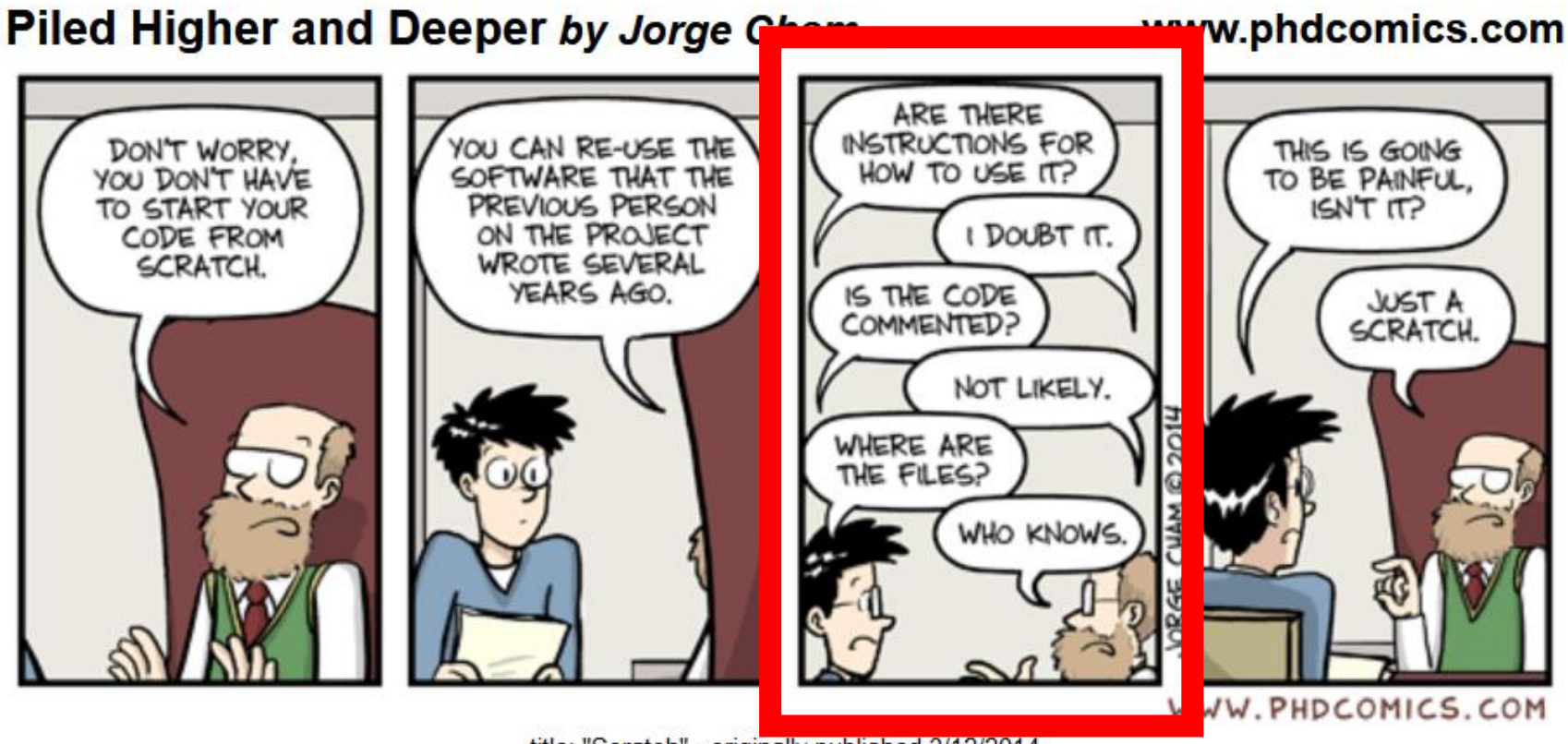
UNESCO Recommendation on Open Science

https://youtu.be/I3Wkvx_ZaFo
https://www.unesco.org/en/natural-sciences/open-science

Reproducible does not (have to) mean fully **open**

As open as possible, as closed as necessary

Open but not usable



title: "Scratch" - originally published 3/12/2014

Reasons for irreproducibility:

- Unavailablility of materials, data and/or analyses
- Poor data management
- Unclear analysis specification
- Lack of documentation
- Errors in reporting numbers
- Lack of quality checking procedures
- Insufficient peer review

Reproducible research workflows

Data acquisition and processing
Data analyses
Data reports (manuscripts)

Data acquisition and processing

Data organization
Data documentation
Version control

Organized data

```
project_name/
├── README.md             # overview of the project
├── data/                 # data files used in the project
│   ├── README.md         # describes where data came from
│   └── sub-folder/       # may contain subdirectories
├── processed_data/       # intermediate files from the analysis
├── manuscript/           # manuscript describing the results
├── results/              # results of the analysis (data, tables, figures)
├── src/                  # contains all code in the project
│   ├── LICENSE           # license for your code
│   ├── requirements.txt  # software requirements and dependencies
│   └── ...
└── doc/                  # documentation for your project
    ├── index.rst
    └── ...
```

Research project with a proper data file structure. Image taken from CodeRefinery, Lesson on Reproducible Research. Shared under CC-BY 4.0.

https://coderefinery.github.io/reproducible-research/organizing-projects/

Versioned data

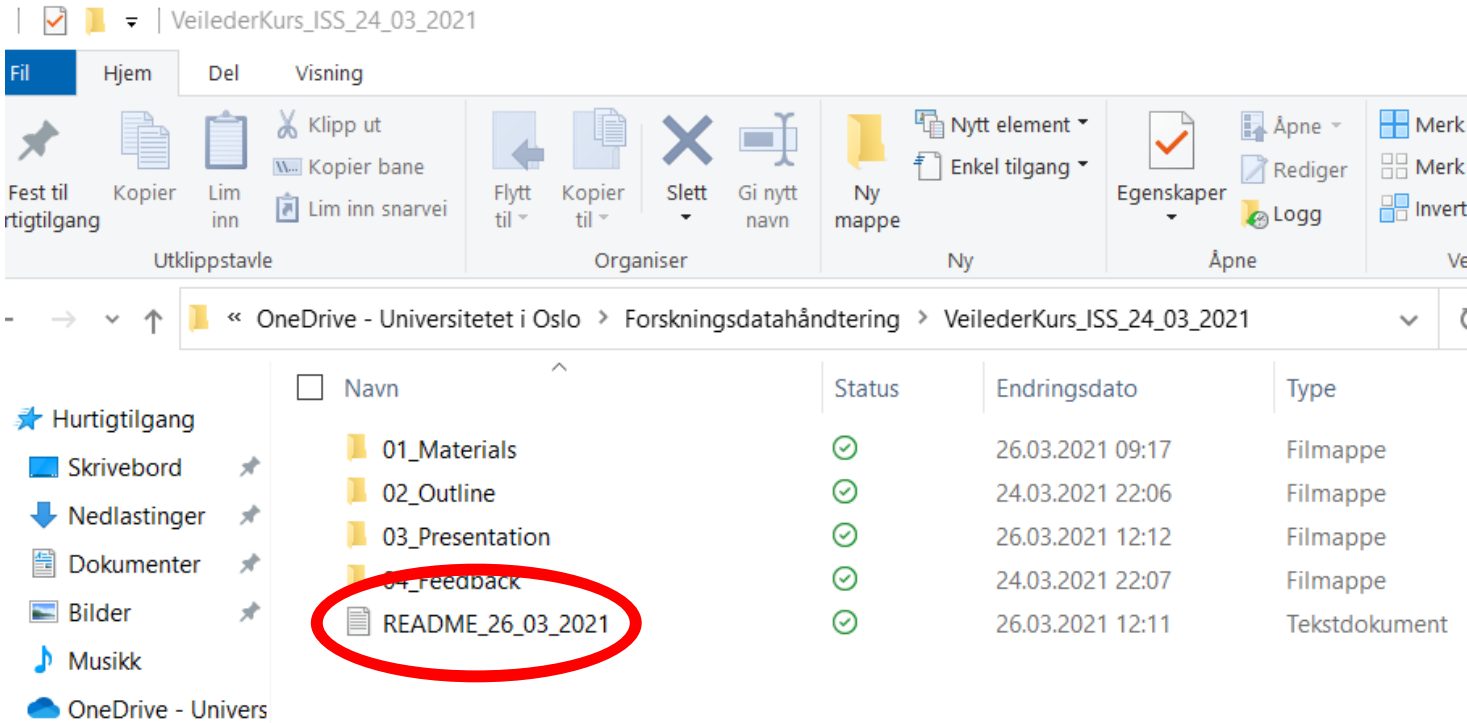
Versioning refers to saving **new copies** of your files when you make changes so that later you can go back and **retrieve** specific **versions** of your files

- DataFileName_1.0 = original document
- DataFileName_1.1 = original document with minor revisions
- DataFileName_2.0 = document with substantial revisions

https://pitt.libguides.com/managedata/organizingdata#s-lg-box-13287647

Documented data: README-files

- The first file to open
- Map for navigating and exploring files and their content
- One README.txt file per folder



Documented data: CODEBOOK

- Explains all variables and their codes in the dataset
- It typically contains:

- variable names, variable labels, variable codes, variable formats, missing data (in quantitative research)

- codes, code definitions, examples of what to include with a given code (in qualitative research)

- Can be also called **Data Dictionary**

Example

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214
27	10026	1	1	1	2	2	332
28	10027	1	1	0	1	2	348
29	10028	1	1	1	1	2	324
30	10029	0	2	9999	0	0	293

Example

Sex: biological sex self-reported by the participant (0 – Female, 1 – Male)

Gender: self-reported gender (0 – Female, 1 – Male, 2 – Other)

Age: age in months at the date of test (automatically calculated in the system as the difference between the date of test and the date of birth)

	A	B	C	D	E	F	G
1	subjid	sex	gender	recruitment	condition	group	age
2	10001	1		2	9999	9999	272
3	10002	1	1	1	2	2	187
4	10003	1		1	9999	9999	291
5	10004	1	1	1	1	1	300
6	10005	1	1	1	1	1	288
7	10006	1		6	0	9999	254
8	10007	1	1	4	1	2	233
9	10008	0	2	9999	0	0	266
10	10009	1	1	2	1	1	195
11	10010	1	2	2	1	1	207
12	10011	1	1	1	1	2	234
13	10012	1		2	1	2	231
14	10013	1	1	0	1	2	257
15	10014	0	2	0	0	0	309
16	10015	0	2	1	2	2	254
17	10016	0	2	5	0	9999	161
18	10017	1		6	0	9999	
19	10018	1	1	1	1	2	314
20	10019	1	1	5	0	0	188
21	10020	1		0	1	9999	
22	10021	1	1	4	9999	9999	
23	10022	1		1	1	2	294
24	10023	1	1	0	9999	9999	
25	10024	1	1	1	1	1	291
26	10025	1	1	2	1	1	214
27	10026	1	1	1	2	2	332
28	10027	1	1	0	1	2	348
29	10028	1	1	1	1	2	324
30	10029	0	2	9999	0	0	293

Tools that help: Templates

- Cornell University template and guide to README.txt-files:

  https://data.research.cornell.edu/content/readme

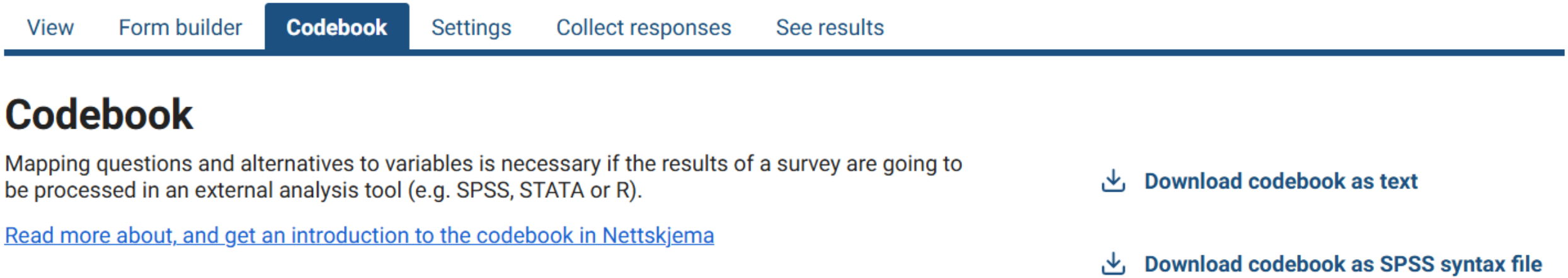
- README.txt-files from DataverseNO:

  General template

  Example for social sciences

  Example for life sciences

Tools that help: Nettskjema codebook



https://www.uio.no/english/services/it/research/sensitive-data/help/create-nettskjema.html
https://www.uio.no/tjenester/it/adm-app/nettskjema/hjelp/kodebok.html

Tools that help: ELN

Electronic Lab Notebooks

help document research, experiments and procedures performed in laboratories

eLabJournal

UNIVERSITY
OF OSLO

https://www.uio.no/english/services/it/adm-services/elabjournal/

Courses in research data management and sharing

November 6th – 27th

Upcoming Courses

Data management planning

Nov. 6, 9:00 AM, Zoom

Data organization, metadata, and documentation

Nov. 8, 9:00 AM, Zoom

Data classification and storage selection

Nov. 10, 9:00 AM, Zoom

Copyright and licensing

Nov. 13, 9:00 AM, Zoom

Sharing and archiving research data

Nov. 16, 9:00 AM, Zoom

Finding and reusing research data

Nov. 17, 9:00 AM, Zoom

Upcoming Workshops

Workshop on Data Management Plans

Nov. 20, 12:00 PM, Georg Sverdrups hus: DSC-Oasen

Workshop on Data Documentation

Nov. 23, 9:00 AM, Georg Sverdrups hus: DSC-Oasen

Workshop on Archiving in DataverseNO

Nov. 27, 12:00 PM, Georg Sverdrups hus: DSC-Oasen

https://www.ub.uio.no/english/courses-events/courses/other/research-data/

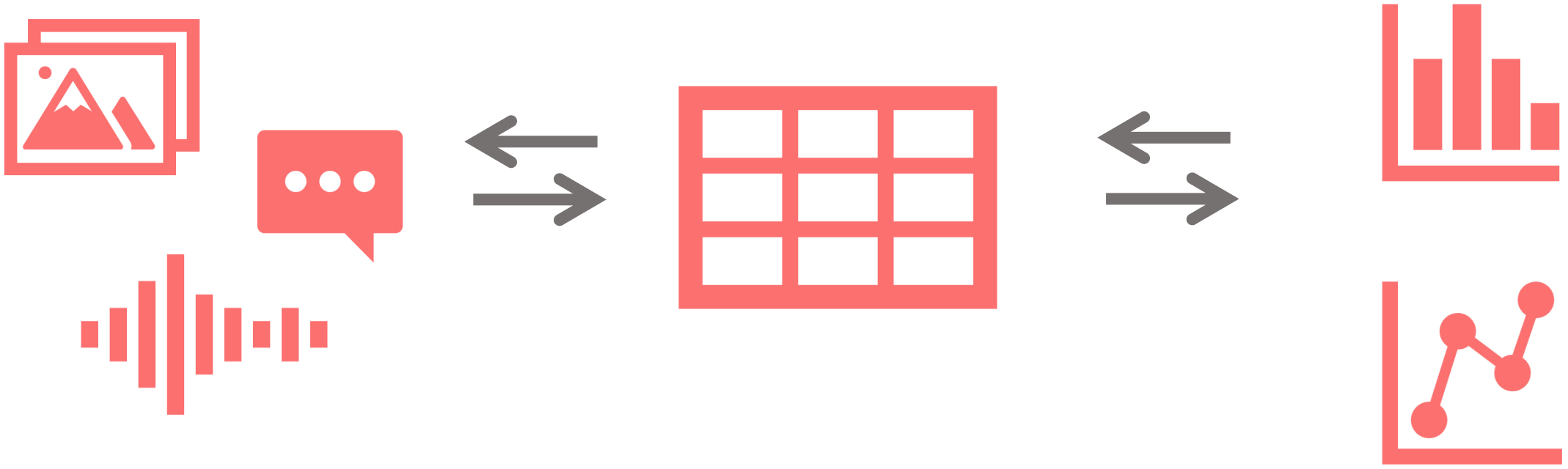
Data analysis

Step-by-step documentation
Version control
Cloud computing and/or containers

Raw data

Processed data

Analyzed data



Qualitative studies
process transparency

Following the step-by-step reasoning and interpretation process

https://blog.ukdataservice.ac.uk/show-me-the-data/

Tools that help: Annotations

Use annotations to comment on selected parts of a source or node

Like scribbled notes in the margin, annotations let you record comments, reminders or observations about specific content in a source or node.

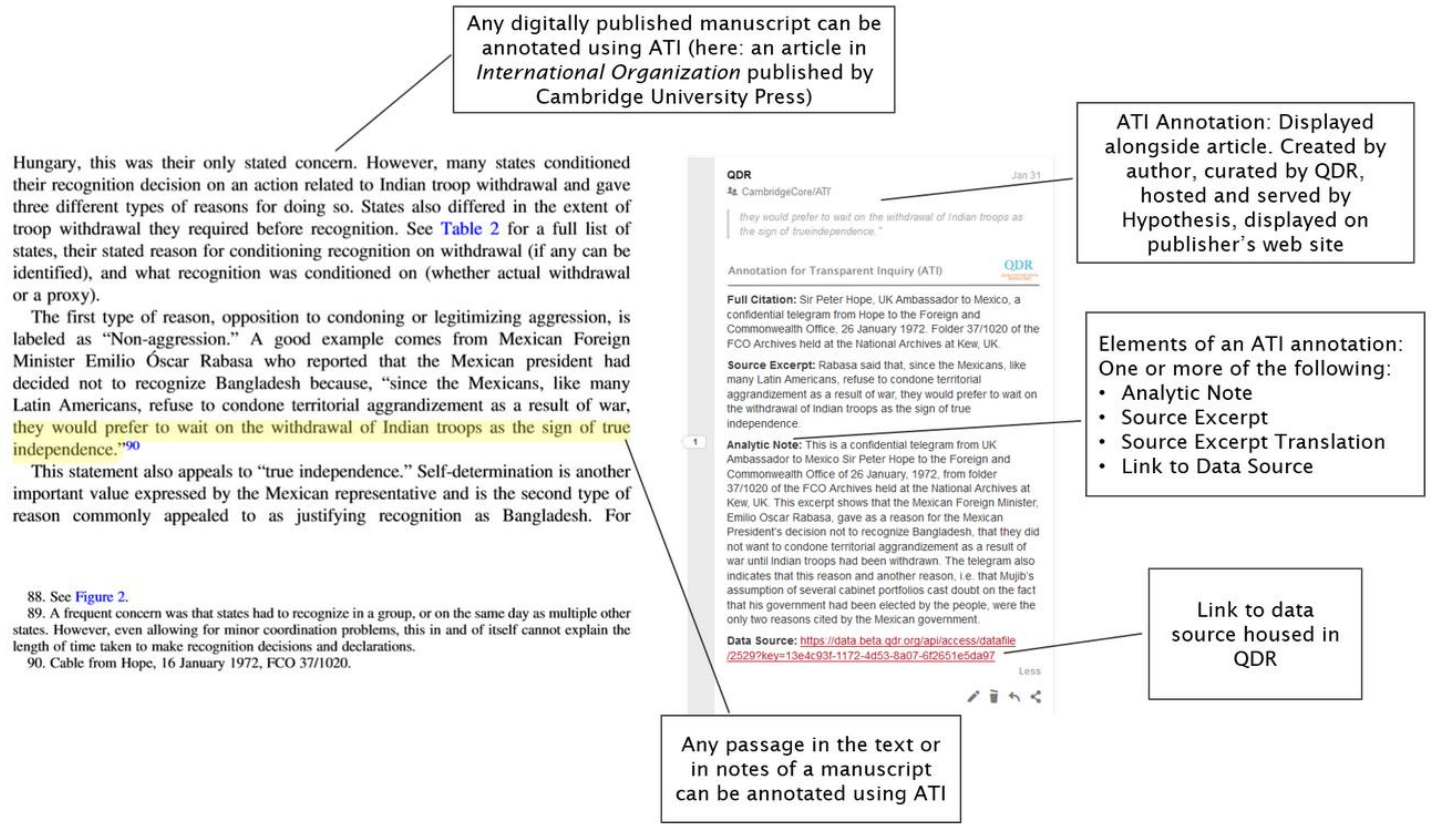
Annotated content is highlighted in blue and the text of the annotation is displayed in the **Annotations** tab at the bottom of the window.

UNIVERSITETET
I OSLO

https://help-nv11.qsrinternational.com/desktop/concepts/about_annotations.htm

Annotation for Transparent Inquiry (ATI) at a Glance

Annotation for Transparent Inquiry (ATI) facilitates transparency in qualitative research by allowing scholars to "annotate" specific passages in an article. Annotations amplify the text and, when possible, include a link to one or more data sources underlying a claim; data sources are housed in a repository.



https://qdr.syr.edu/ati

SYMPOSIUM

Active Citation: A Precondition for Replicable Qualitative Research

Andrew Moravcsik, *Princeton University*

https://www.princeton.edu/~amoravcs/library/ps.pdf

Quantitative studies
computational reproducibility

Re-running analyses/code with the same data

Tools that help: analysis via code

AutoSave Off | Book1 - Excel | Search | Agata Bochynska AB

File Home Insert Draw Page Layout Formulas Data Review View Help | Share | Comments

Get Data | Refresh All | Queries & Connections | Properties | Edit Links | Get & Transform Data | Queries & Connections

Stocks (En... | Geography... | Data Types

Sort | Filter | Clear | Reapply | Advanced | Sort & Filter

Text to Columns | Data Tools

What-If Analysis | Forecast Sheet | Forecast

Outline

A1 | 1

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	1	A	0,657792															
2	2	A	0,443001															
3	3	A	0,490799															
4	4	A	0,315724															
5	5	A	0,86144															
6	6	A	0,481306															
7	7	A	0,337931															
8	8	A	0,126269															
9	9	A	0,571799															
10	10	A	0,317623															
11	11	B	0,285825															
12	12	B	0,931719															
13	13	B	0,615969															
14	14	B	0,068008															
15	15	B	0,918656															
16	16	B	0,531989															
17	17	B	0,695157															
18	18	B	0,810593															
19	19	B	0,989223															
20	20	B	0,54528															
21	21	B	0,4301															
22	22	B	0,130967															
23	23	B	0,384371															

Sheet1

```{r}
# Main Analysis Data

#load wide format data and preview
sum_data <- read.csv("Data/Experiment2_SumData.csv")
head(sum_data)

#check summary statistics for the dataset
describe(sum_data)
```

```{r}
# Main Analysis

#ttest on the total proportion looking to shape change against chance (0.5)
t.test(sum_data$ShapeProportion, mu=0.5)
sd(sum_data$ShapeProportion)
se <- sd(sum_data$ShapeProportion)/sqrt(length(sum_data$ShapeProportion))
se

#compute the effect size (Cohen's D)
cohensD(sum_data$ShapeProportion, mu=0.5)


# Bayesian ttest on the total proportion looking to shape change against chance (0.5)
testMain <- ttestBF(sum_data$ShapeProportion, mu=0.5)
testMain
sd(sum_data$ShapeProportion)
se <- sd(sum_data$ShapeProportion)/sqrt(length(sum_data$ShapeProportion))
se
```

Tools that help: version control

- Git: Free and open source version control system

- GitHub: is an internet hosting service for software development and version control using Git

https://git-scm.com/
https://github.com/
https://youtu.be/gY2JwRfin1M

Carpentry@UiO

Carpentry@UiO is a community of people who are passionate about learning, teaching, and sharing best practices and digital skills for making the research process more reproducible and effective. If you want to get involved, or join one of our workshops, check us out!

UNIVERSITETET I OSLO | THE CARPENTRIES | DATA CARPENTRY | software carpentry | library carpentry

Upcoming workshops

Carpentry: Version Control with Git - September 19

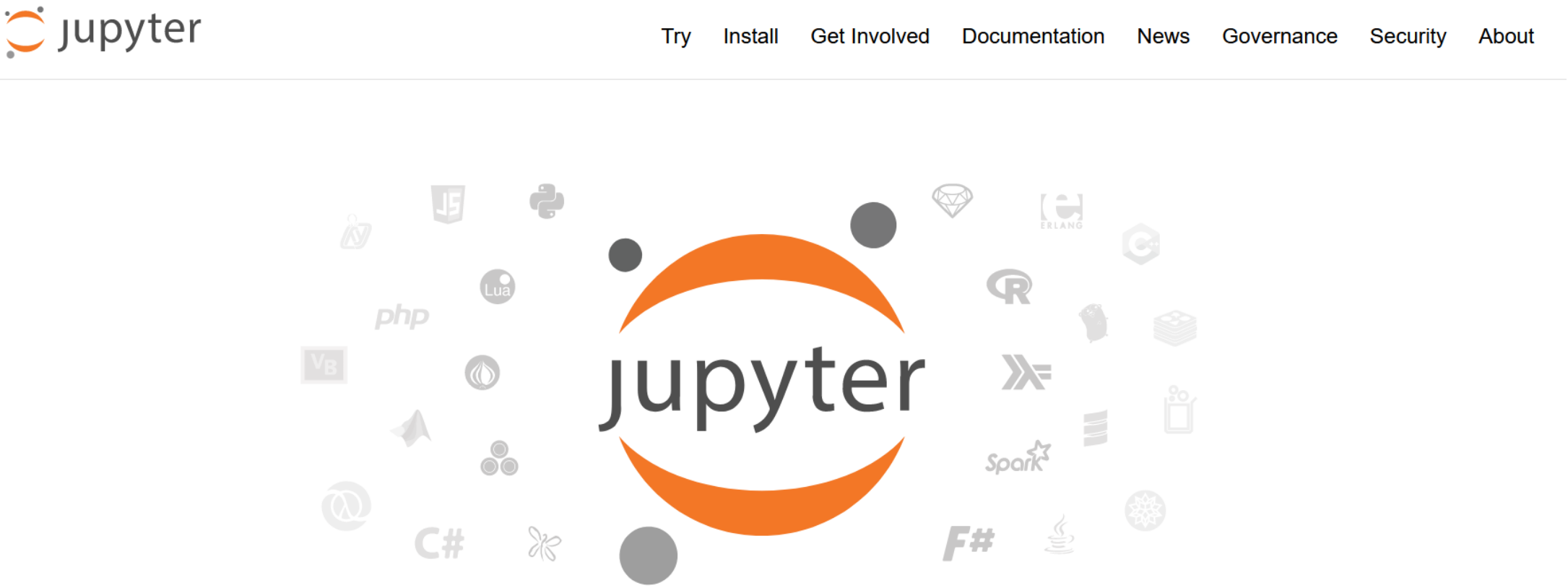
Sep. 19, 2023 9:00 AM – 3:00 PM, GSH: LINKEN

Carpentry: Unix Shell - October 12

Oct. 12, 2023 9:00 AM – 4:00 PM, GSH: LINKEN

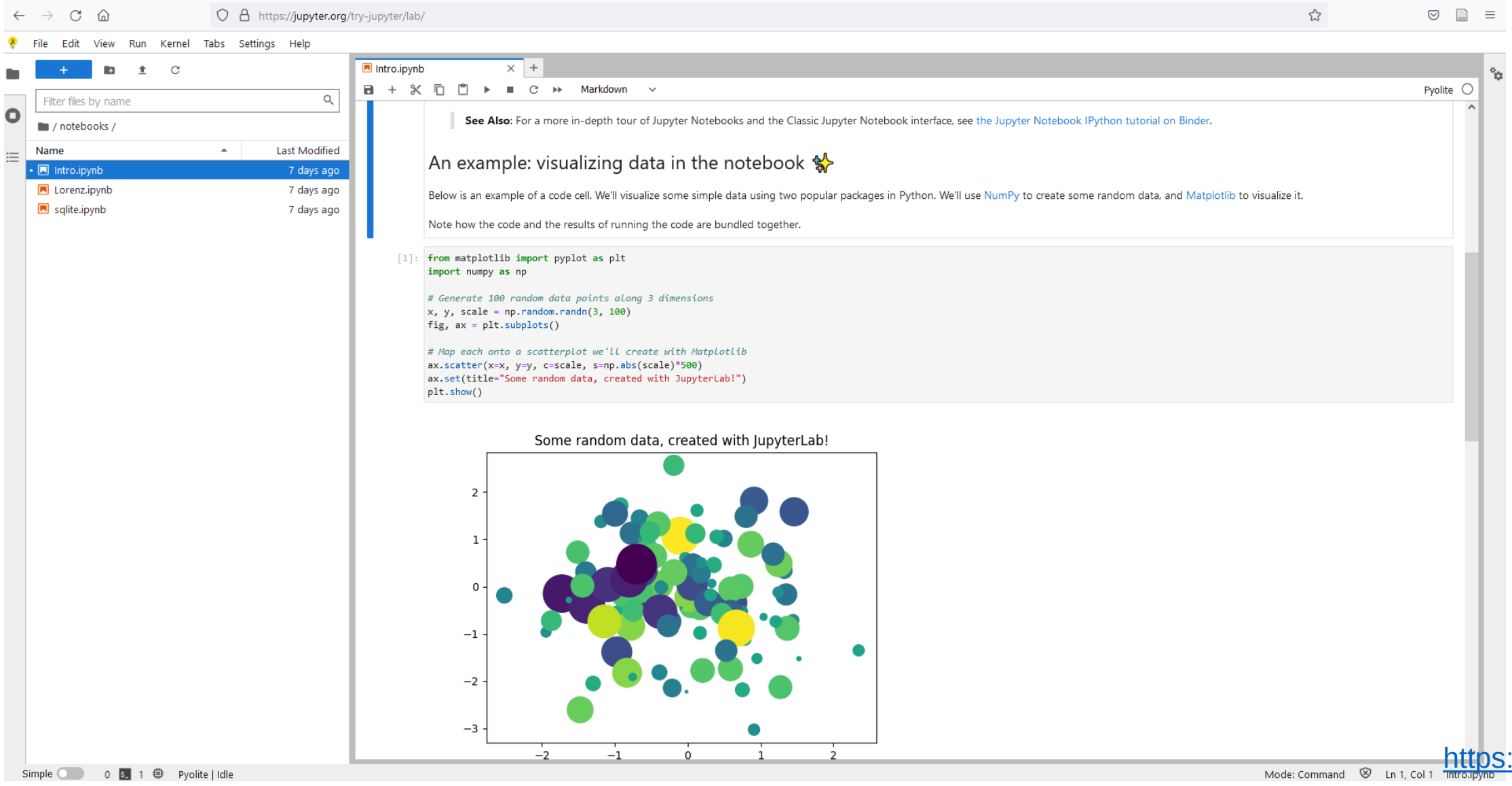
https://www.ub.uio.no/english/courses-events/courses/other/Carpentry/

Tools that help: shared notebooks



https://jupyter.org/

Tools that help: shared notebooks



https://jupyter.org/

Tools that help: Quality check

Language	Static code analysis tool
Python	Pylint, prospector
R	lintr
Shell/Bash	shellcheck

Language	Formatter Tool
Python	Black, yapf
R	formatR
Shell/Bash	ShellIndent
HTML	Tidy

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

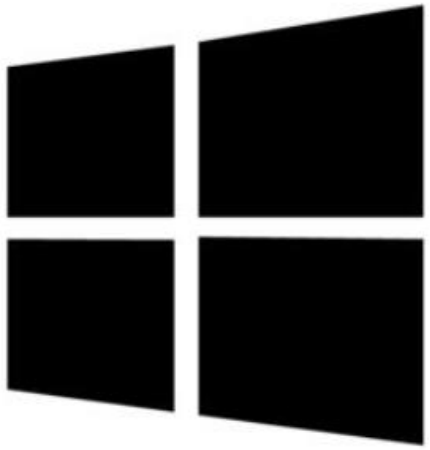
Tools that help: Code review

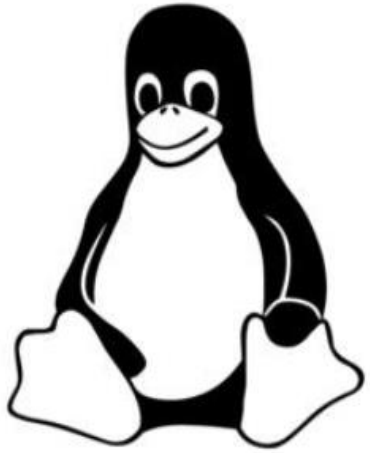
Catching bugs. *The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807.

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

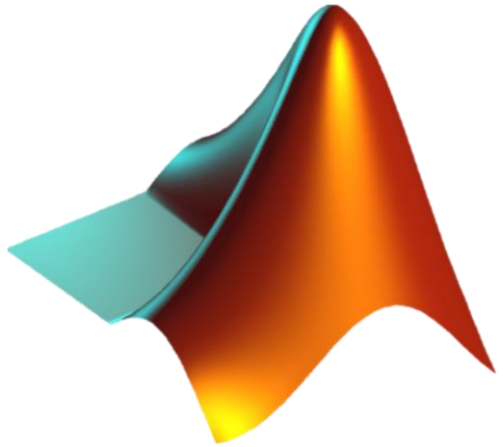
However…

Different operating systems



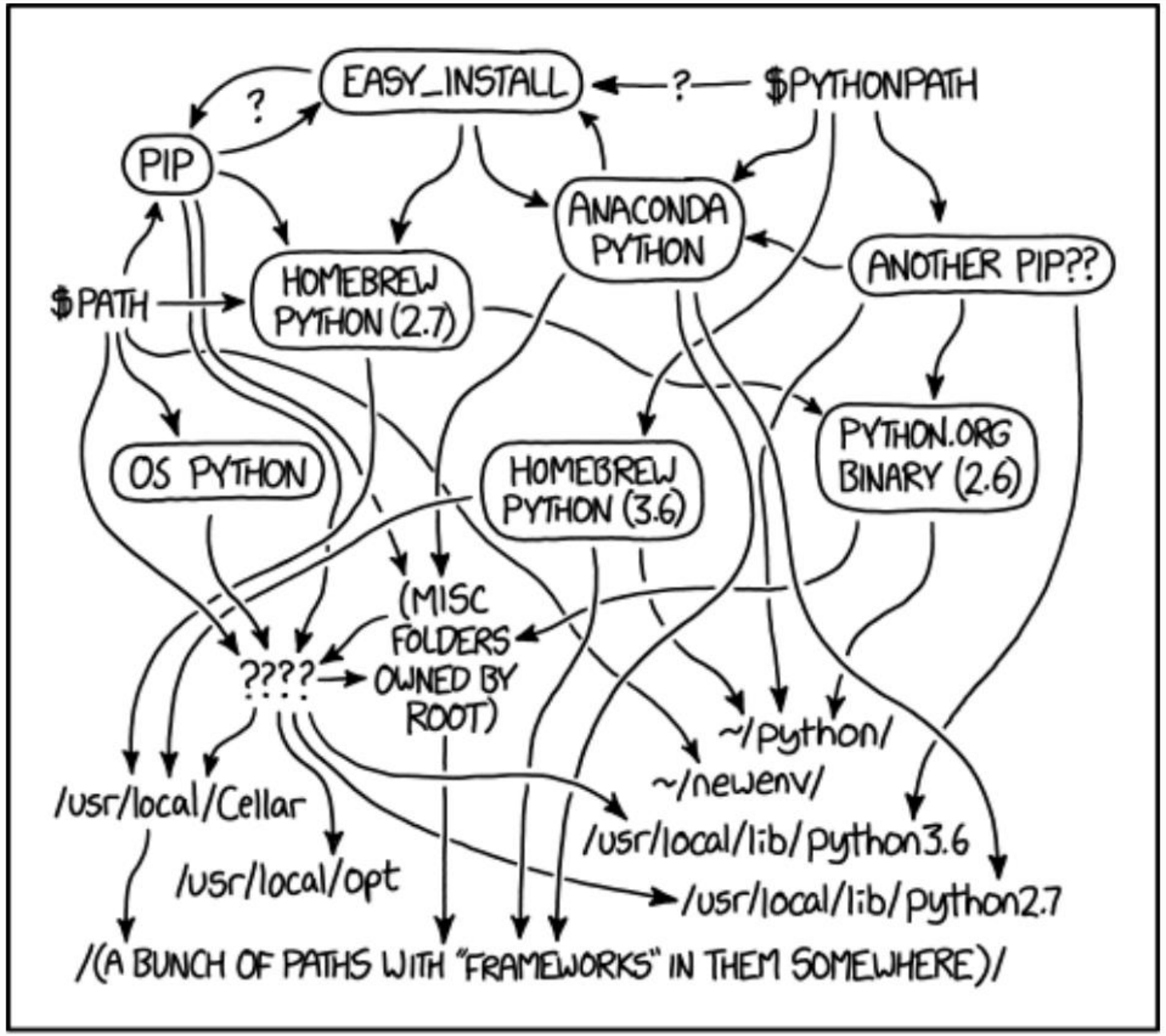


Different programs/languages



Different packages

https://jiayiwu.me/blog/2021/05/13/how-to-make-a-r-package-in-rstudio.html
https://nazca-design.org/wp-content/uploads/2017/04/python-modules.png



MY PYTHON ENVIRONMENT HAS BECOME SO DEGRADED THAT MY LAPTOP HAS BEEN DECLARED A SUPERFUND SITE.

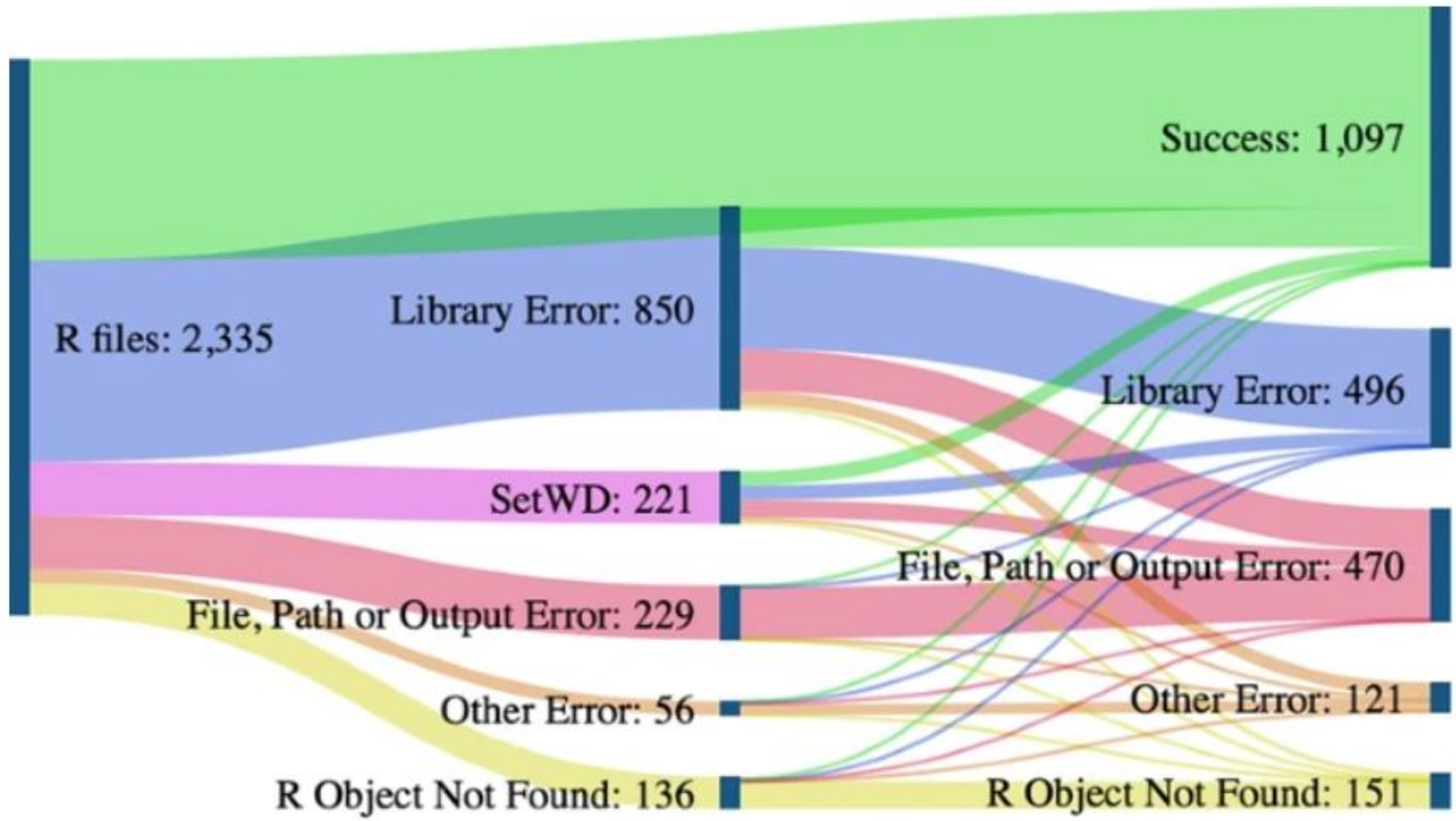
https://xkcd.com/

Analysis | Open Access | Published: 21 February 2022

A large-scale study on research code quality and execution

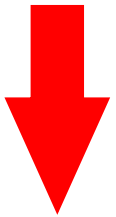
Ana Trisovic, Matthew K. Lau, Thomas Pasquier & Mercè Crosas

Scientific Data **9**, Article number: 60 (2022)



https://www.nature.com/articles/s41597-022-01143-6

Computational environment



Container

Tools that help: Containers

What is reproduced? \ Interaction style	Graphical	Command line
Software & versions	binder	CONDA
Entire system		docker

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

Private Untitled Capsule Oct 19, 2022 17:16

Capsule File Help

master Share

Files App Panel Tabs

Core Files

- metadata 62 B
  - metadata.yml 62 B
- environment 195 B
  - Dockerfile 195 B
- code 0 B
- data Manage Datasets 0 B

Results

- results

Other Files

Upload

or

Start with Sample Files

Environment | Metadata | metadata.yml | Dockerfile

Editing the Dockerfile will permanently disable the Environment Editor. Please commit before editing.

Use Environment Editor Unlock

```
1  # hash:sha256:f137bdfbb7795cd46da870b7886b0ebcfe5d14a2070513da9f40f979e584cc81
2  FROM registry.codeocean.com/codeocean/r-studio:2022.07.0-548-r4.2.1-ubuntu18.04
3
4  ARG DEBIAN_FRONTEND=noninteractive
5
6  RUN Rscript -e 'remotes::install_version("ggplot")' \
7      && Rscript -e 'remotes::install_version("tidyverse")'
8
```

Reproducible Run

or launch a cloud workstation

lab RStudio jupyter Shiny

Timeline

Submit for publication...

What happens once I publish?

Select filter...

Agata Bochynska committed
Oct 19, 2022

Added packages

Oct 19, 2022
Created Capsule

Reproducibility

CODE OCEAN

The Digital Lab for Computational Scientists

Start faster. Reproduce reliably. Focus on science.

https://codeocean.com/

Data reports (manuscripts)

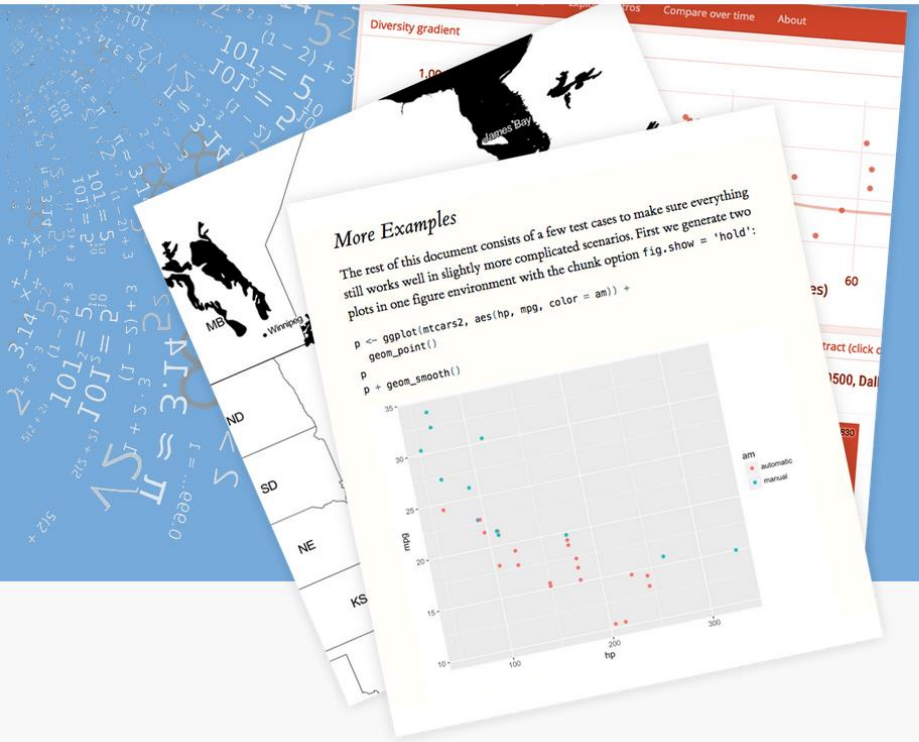
Linked tables and analyses
Version control
Collaboration

Tools that help: R Markdown

R Markdown

from RStudio

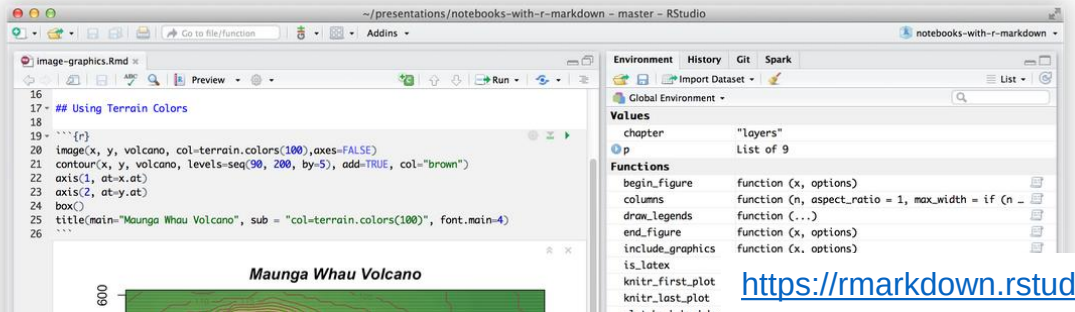
Get Started Gallery Formats Articles Book References



Analyze. Share. Reproduce.

Your data tells a story. Tell it with R Markdown. Turn your analyses into high quality documents, reports, presentations and dashboards.

R Markdown documents are fully reproducible. Use a productive notebook interface to weave together narrative text and code to produce



https://rmarkdown.rstudio.com/

Tools that help: R Markdown

Efficient

- everything in one place
- re-do analyses and automatically update throughout all text
- decrease the possibility of making mistakes

Flexible

- output to various formats, e.g., HTML, LaTeX, PDF, Word

Open access/source

- use for free
- create documents accessible to anyone (with a computer and internet connection)

Adapted from: https://resulumit.com/teaching/rmd_workshop.html

Tools that help: Quarto

https://quarto.org/

Other tools

- Overleaf (collaborative LaTeX editor)
- HackMD (a realtime web-based collaborative Markdown editor)
- Manuscripts.io (a collaborative authoring tool that support scientific content and reproducibility)
- Rrtools (instructions, templates, and functions for making a basic compendium suitable for writing a reproducible journal article or report with R)
- Jupyter Notebooks (can be used for supplementary material with journal articles.

https://coderefinery.github.io/reproducible-research/organizing-projects/

Reproducible research workflows

Data acquisition and processing
Data analyses
Data reports (manuscripts)

Take-aways

- Be **transparent** about your full research workflow: research questions, methods, data, step-by-step procedures and analyses
- Make sure you have good **documentation** for all outputs and all stages of your research process
- Keep track of **versions** and do a solid **quality check** of your methods, data and analyses
- **Verify** your own work: try to reproduce your own results and/or have others do it
- Make your methods, data and analyses **open** (if you can)

UNIVERSITETET
I OSLO

ReproducibiliTea

Journal Club

JOIN IN AND DISCUSS WITH FELLOW STUDENTS AND RESEARCHERS

OPEN RESEARCH, REPRODUCIBILITY and RESEARCH IMPROVEMENT

Join us

Everyone is welcome to join us - whether you are an enthusiast of open and reproducible research, a skeptic, or a cautious explorer. Currently, all meetings are hybrid with the possibility of joining on-site at Blindern or via Zoom. Grab a cup of tea (coffee?) and join us!

Subscribe to our mailing list →

https://www.ub.uio.no/english/libraries/dsc/open-repro-research/reproducibilitea/

UNIVERSITETET
I OSLO

ReproducibiliTea

Journal Club

JOIN IN AND DISCUSS WITH FELLOW STUDENTS AND RESEARCHERS

OPEN RESEARCH, REPRODUCIBILITY and RESEARCH IMPROVEMENT

Upcoming Monday!

Sep. 18th at 14.00

A Reproducible Data Analysis Workflow With R Markdown, Git, Make, and Docker

Aaron Peikert [1,2] , Andreas M. Brandmaier [1,3]

https://www.ub.uio.no/english/libraries/dsc/open-repro-research/reproducibilitea/

NORRN

OUR MISSION | TERMS OF REFERENCE | CODE OF CONDUCT | MEMBERS | CONTACT

Welcome to Norway's Reproducibility Network

Towards open & reproducible science

JOIN US

Our Mission

The Norwegian Reproducibility Network (NORRN) is a peer-led network that aims **to promote and enable rigorous, robust and transparent research practices in Norway**. We attempt to achieve this goal by establishing appropriate training activities, designing, and evaluating research improvement efforts, disseminating best practices, and working with stakeholders to ensure coordination of efforts across the sector. NORRN's activities span multiple levels, including researchers, librarians, institutions, and other stakeholders (e.g., funders and public authorities).

https://www.norrn.no/

https://www.ub.uio.no/english/libraries/dsc/

Digital Scholarship Centre

At the Digital Scholarship Centre (DSC) you get guidance on how you can make the best possible use of digital tools and methods in your research and communication activities.

Open Access →

Information about open access publishing, publisher agreements, self-archiving, requirements, and guidelines.

Open and reproducible research →

Make your research more transparent and reproducible.

Research Data Management →

Managing your data both during and after a research project.

Data Managers Network →

An informal network that aims to enable information exchange, initiate action, and create new interdisciplinary collaborations

Text Mining →

Information about digital tools for searching, mining, and analysing textual data.

Systematic search →

Information about systematic literature searching, how to get started, and how to get help.

Visualization →

Use of visual methods to explore, communicate and understand data.

Carpentry@UiO →

Offers workshops in foundational digital skills such as coding and data management.

Reference management →

Styles, tools, and information on reference management.

Digiforsk blog →

A blog about work and resources at UiO where digital research methods are involved. Posts in both Norwegian and English.

Open and reproducible research

Learn about how to make your research more open and reproducible and get involved in initiatives and communities that are interested in sharing and improving research at UiO.

More and more researchers and students across disciplines are implementing open research practices, preregistering their hypotheses, methods, and analysis plans and sharing research materials, data and analysis scripts. Digital Scholarship Center can help you learn about and implement these practices in your own research as well as advise on the policies and requirements from funders.

Open Science Lunch →

Every last Thursday of the month we meet at noon to discuss topics related to open research.

ReproducibiliTea@UiO →

Join us for a Journal Club where we read and discuss papers on open research and meta-science.

Norwegian Reproducibility Network →

Join a broader community that aims to promote and enable rigorous, robust and transparent research practices in Norway

Courses and workshops →

Click here for the list of upcoming and previous courses and workshops on open and reproducible research at UiO.

https://www.ub.uio.no/english/libraries/dsc/open-repro-research/

Menu

← Libraries and centres

https://www.ub.uio.no/english/libraries/dsc/

Norwegian version of this page

Digital Scholarship Centre

At the Digital Scholarship Centre (DSC) you get guidance on how you can make the best possible use of digital tools and methods in your research and communication activities.

https://sympa.uio.no/ub.uio.no/subscribe/dsc-news/subscribe

More resources:

The Turing Way: Guide for Reproducible Research
https://the-turing-way.netlify.app/reproducible-research/reproducible-research.html

CodeRefinery: Reproducible Research
https://coderefinery.github.io/reproducible-research/motivation/

CodeRefinery workshop September 19-21 and 26-28, 2023

Thank you!

- Be **transparent** about the full research workflow: questions, methods, data, step-by-step procedures and analyses
- Make sure you have good **documentation** for all outputs and all stages of your research process
- Keep track of **versions** and do a solid **quality check** of your methods, data and analyses
- **Verify** your own work: try to reproduce your own results and/or have others do it
- Make your methods, data and/or analyses **open** (if you can)

Agata Bochynska, PhD
Open Research and Digital Scholarship Center
University of Oslo Library

@AgataBochynska
agata.bochynska@ub.uio.no

UNIVERSITY
OF OSLO