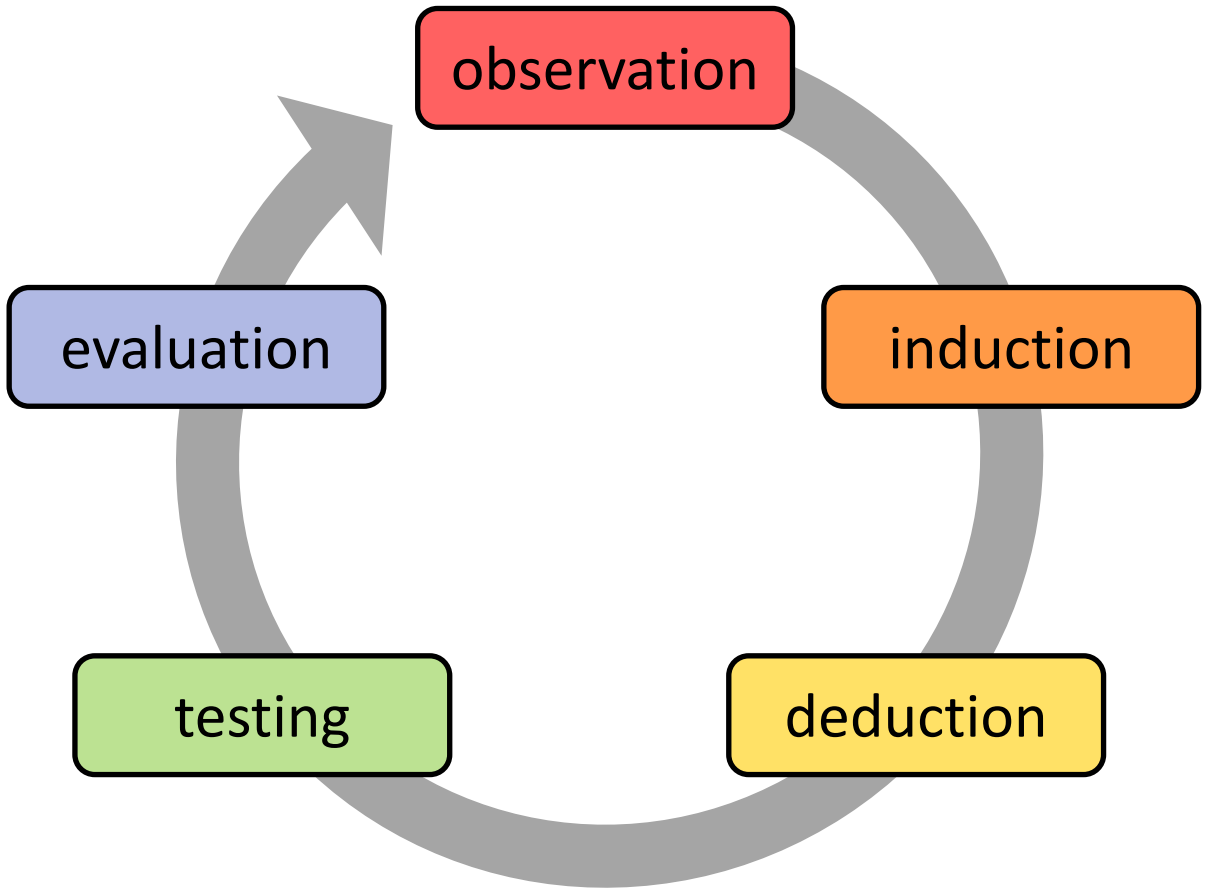
Open Science

The Open Empirical Cycle

Hanne Oberman

Materials by Herbert Hoijtink, Bianca Kramer & Jeroen Bosman; re-use license: CC BY

The Empirical Cycle



The Empirical Cycle in Practice

observation	Step 1: Preparation
induction	Step 2: Formulate research hypotheses
deduction	Step 3: Planning of data collection and analysis; write Data Management Plan
	Step 4: Get approval from the (medical) ethical testing committee
	Step 5: Pre-registration (or registered report)
testing	Step 6: Execution of data collection and analyses
evaluation	Step 7: Write report; publish pre-print
	Step 8: Publish (meta-)data and analyses
	Step 9: Having your report reviewed
	Step 10: Publish open access

observation

Step 1: Preparation

Reviewing:

- Review literature
- Gaps in the literature
- Quality of the literature
- Variables that are not covered in the literature

Writing:

- Writing it up sharpens thoughts
- Write down you research questions:
    1. New questions
    2. Replication studies
    3. Exploratory studies

Step 2: Formulate Research Hypotheses

induction

A research hypothesis is
a verbal representation
of the expected relations
between the variables
resulting from Step 1.

Step 3: Planning of Data Collection and Analysis

deduction

Design:
- Describe research design
- Exploration?
- Replication?

Data:
- Describe population
- Describe variables
- Derived variables?

Analyses:
- Statistical model
- Formal hypotheses
- Power analyses or updating
- Missing data?
- Data exclusion?

→ Data Management Plan

Step 4: Get (M)ETC Approval

deduction

- Much of the previous three steps
- Data management plan (e.g., data storage)
- Informed consent forms
- Privacy regulations (e.g., GDPR)

Step 5: Pre-registration or Registered Report

deduction

This is the
pre-data-*collection*
(or pre-data-*access*)
account
of all that has been covered
in the previous four steps.

Step 6: Execution of Data Collection and Analysis

testing

Execute steps in pre-registration
after data *collection* (or *access*).

Keep note of
any necessary deviations
from the pre-registration.

Step 7: Write Report

Write a reproducible report.

Include a link to your
pre-registration (Step 5).

Include a link to your
data-analyses repository (Step 8).

Publish pre-print.

evaluation

Step 8: Publish Data and Analyses

Data and analyses should be accessible to all interested parties, or **FAIR**:
Findable,
Accessible,
Interoperable,
Reusable.

evaluation

Step 8: Publish Data and Analyses

- **Findable:** Place your data and analyses in a public repository
- **Accessible:** Make certain your data come with a codebook, and your analyses with annotations
- **Interoperable:** Ensure that data and analyses can be opened on different types of computers
- **Reusable:** Include a license, that is, make clear what others are (not) allowed to do with your data

evaluation

Step 8: Publish Data and Analyses

With open data & analyses, others can:
- Inspect your data
- Reproduce your analyses
- Read your interpretation of the analyses
- Data are available for meta-analyses and “null-findings” become accessible

Being open will add to the trust in your research, that of the group to which you belong, and science in general

evaluation

Step 9: Having your Report Reviewed

May lead to
changes in your report
and possibly
deviations from your pre-registration.

These deviations can be highlighted
in your report using footnotes.

evaluation

Step 10: Publish Open Access

Publishing open access is an important feature of open science.

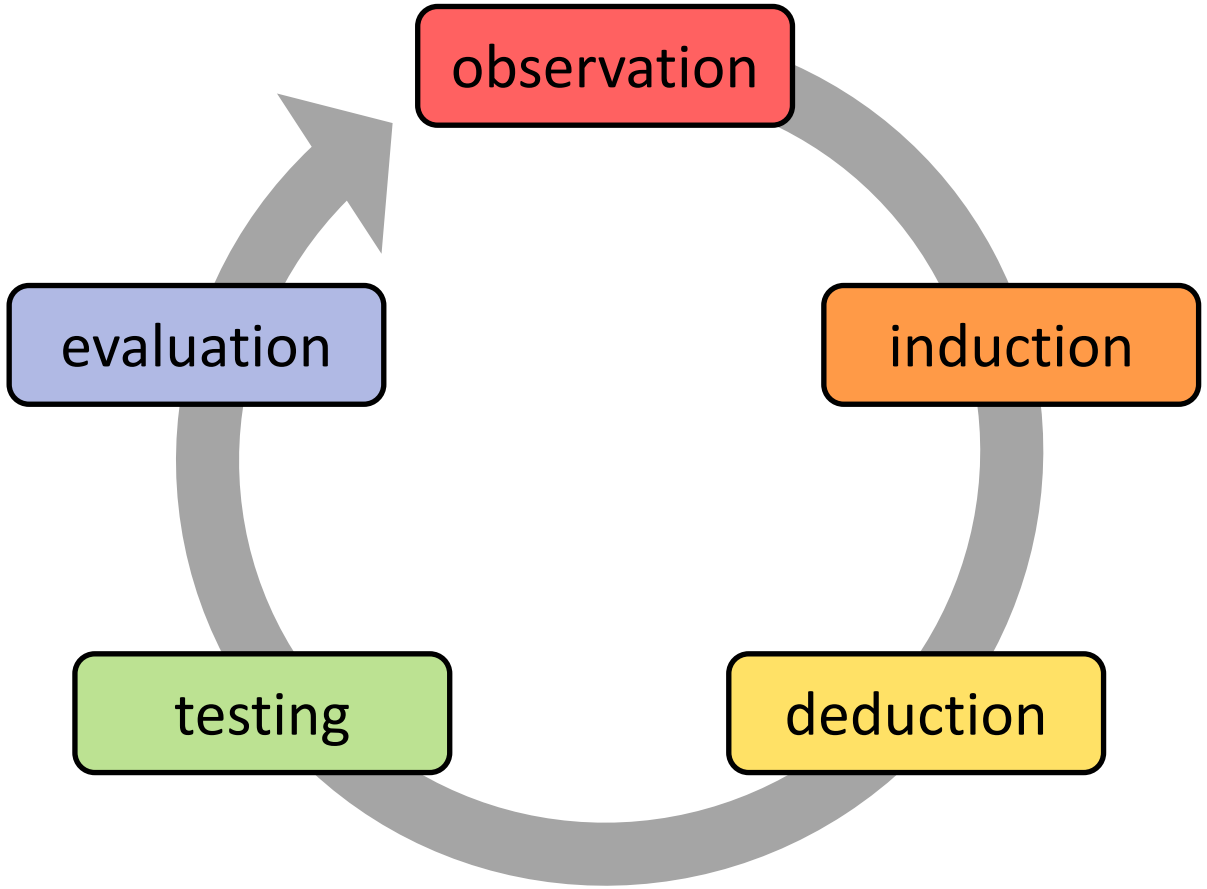
Being able to unobstructedly obtain everything (also the report) related to a research project will enable anyone to:

- benefit from your research
- reuse (parts) of your research
- engage in a fully informed discussion about your research.

This should both increase the impact of your research and increase the trust in your research and the trust in science in general.

evaluation

The Open Empirical Cycle



Links to …

JASP

A Crash Course into JASP

How to use JASP

REPOSITORIES

Open Science Foundation

Yoda (Utrecht University)

LICENCES

Creative Commons

OpenAIRE

ANONIMIZED AND FAIR DATA

How to Make you Data FAIR

Handling Personal Data

FAIR Cheatsheets

General Data Protection Regulation

OPEN SCIENCE

Open Science Community Utrecht

Summer school ‘Open Science Bootcamp’

Utrecht University

Links to …

THE WORKSHOP MATERIALS

Open your Course/Bachelor Thesis

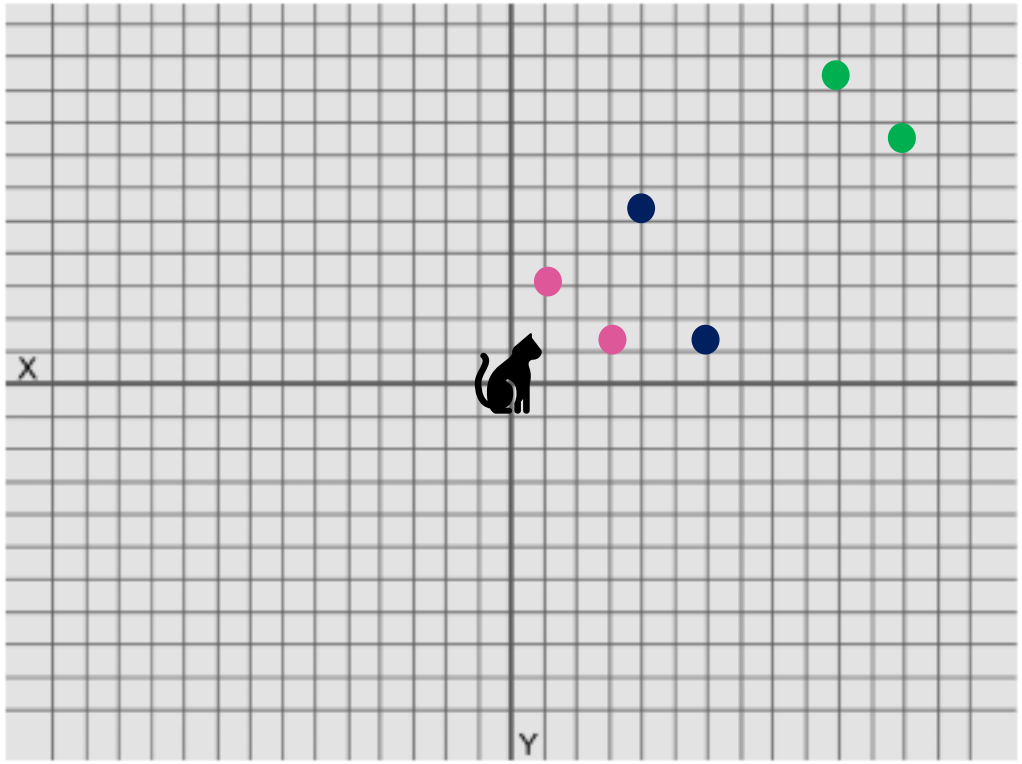
Utrecht University

How to Open your Data & Analyses

An Experiment and its Replication

An experiment with three conditions:

- The “close” condition
- The “intermediate” condition
- The “distant” condition



Participants Rated:

Attachment to

- Siblings
- Parents
- Home-town

on a
1 (not at all strong) – 7 (extremely strong)
Likert scale

which are averaged to obtain
the dependent variable
attachment

The description given here is a modification of and inspired by the actual experiment executed by Williams, L.E. and Bargh, J.A. (2008). Keeping One’s Distance. The Influence of Spatial Distance Cues on Affect and Evaluation. *Psychological Science, 19,* 302-308.

Williams and Bargh (2008) tested:

H_0: $\mu_{close} = \mu_{intermediate} = \mu_{distant}$,
that is, the three means are equal

rendering

p-value = .01, that is, smaller than .05,
that is, the means are significantly different

with

$m_{close} = 5.61$, $m_{intermediate}$ =  , $m_{distant} = 4.86$

and

$\eta^2 = .11$,
that is, the three conditions explain
11% of the variation in attachment,
which is a medium to strong effect of condition

The replication by Joy-Gaba, Clay, and Cleary (2016) rendered

p-value = .79

$\eta^2 = .00$

Joy-Gaba, J., Clay, R., and Cleary, H. (2016). Replication of keeping one's distance: The influence of spatial distance cues on affect and evaluation by Williams L.E. and Bargh J.A. (2008) *Psychological Science, 19,* 302-308). Retrieved from https://osf.io/a78bm/

The Replication Crisis

This is only one of 100 psychological experiments of which only about 33% were successfully replicated (OSC, 2015).

This resulted in a reduced trust in science by scientists and society: The replication crisis was born.

Scientists are alerted:

- Estimating the reproducibility of psychological science (OSC, 2015)
- An open investigation of the reproducibility of cancer biology research (Errington et al., 2014)

"Society" is alerted:

- Is psychology a real science? (Is psychologie wel een echte wetenschap, Volkskrant, 12-8-2016)
- Public Trust in Science (Rathenau Instituut, August 28, 2018)

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, 6251. https://osf.io/ezcuj/

Errington, T.M., Iorns, E., Gunn, W., Tan, F.E., Lomax, J., and Nosek, B.A. (2014). An open investigation of the reproducibilty of cancer biology research. *eLIFE, 3*, e04333. https://elifesciences.org/collections/9b1e83d1/reproducibility-project-cancer-biology

Volkskrant (2016). https://www.volkskrant.nl/columns-opinie/is-psychologie-wel-een-echte-wetenschap~b9978e6c

Rathenau Instituut (2018). Public Trust in Science. https://www.rathenau.nl/en/science-figures/impact/trust-science/public-trust-science

FAIR Data Analyses of your Thesis

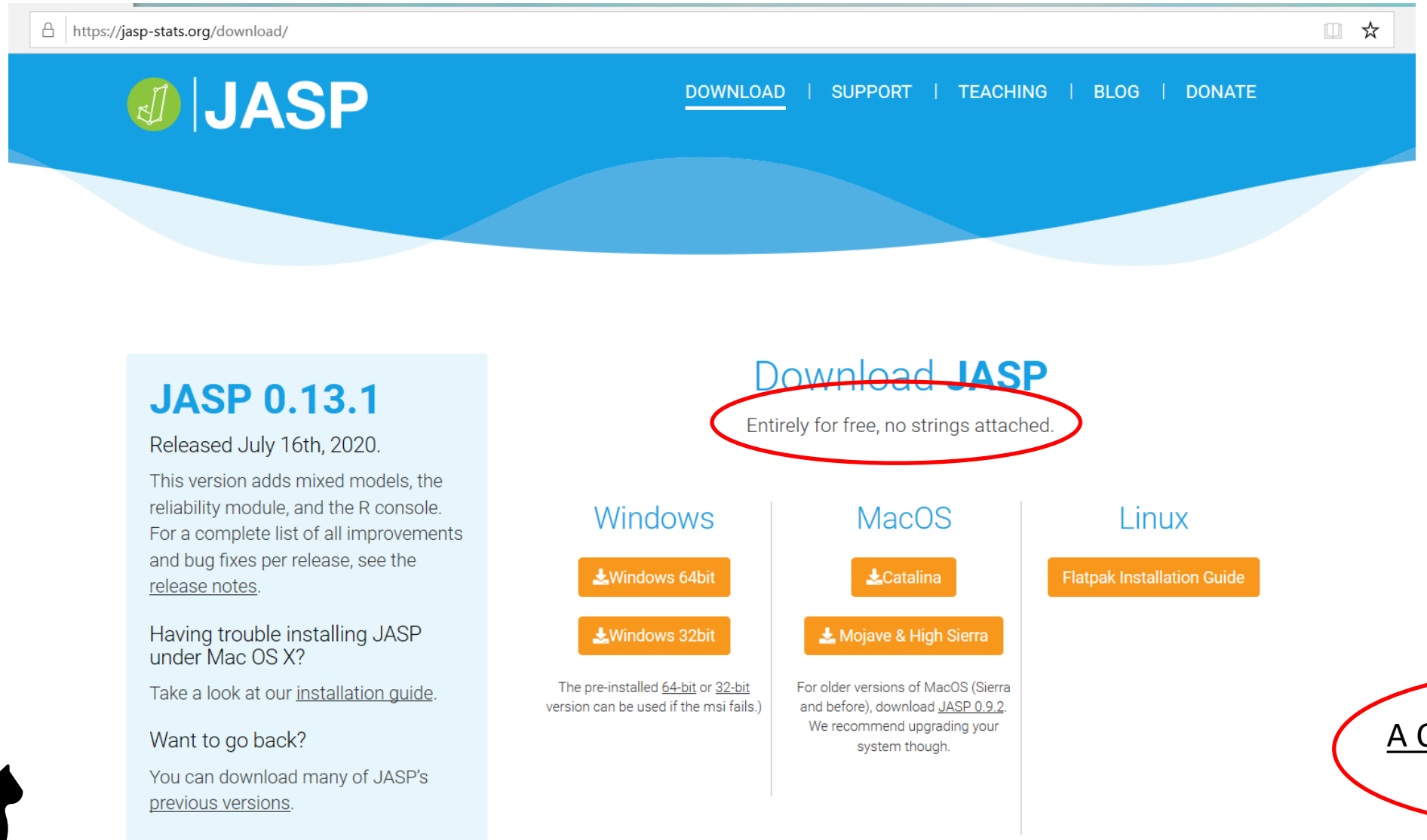
… can easily be done using:

… applies to:

All research based on quantitative data

Any file type related to your research (e.g., dataset, pre-registration, tutorial, video, etc.)

Interoperable



A Crash Course into JASP

How to use JASP

Reusable

Fully Open Data Analyses: Include the CC0 1.0 Universal License With Your Data

The license can be included in your mydata.jasp file.

The CC0 1.0 Universal Public Domain Dedication:

- Is truly open, that is, anybody can use your data for whatever purpose
- If you include a reference to a paper or your contact information in the mydata.jasp file, anybody using your data can refer to you

The CC0 1.0 Universal Public Domain Dedication: https://creativecommons.org/publicdomain/zero/1.0/

Accessible

Joy-Gaba, Clay, and Cleary (2016) replicated Willams and Bargh (2008)

The replication data and analyses are contained in openJoyGaba.jasp

The Data Collected by Joy-Gaba, Clay, and Cleary (2016)

openJoyGaba (C:\Users\hoijt101\Desktop\JASP-UU\KENNISCLIP Open Your Course)

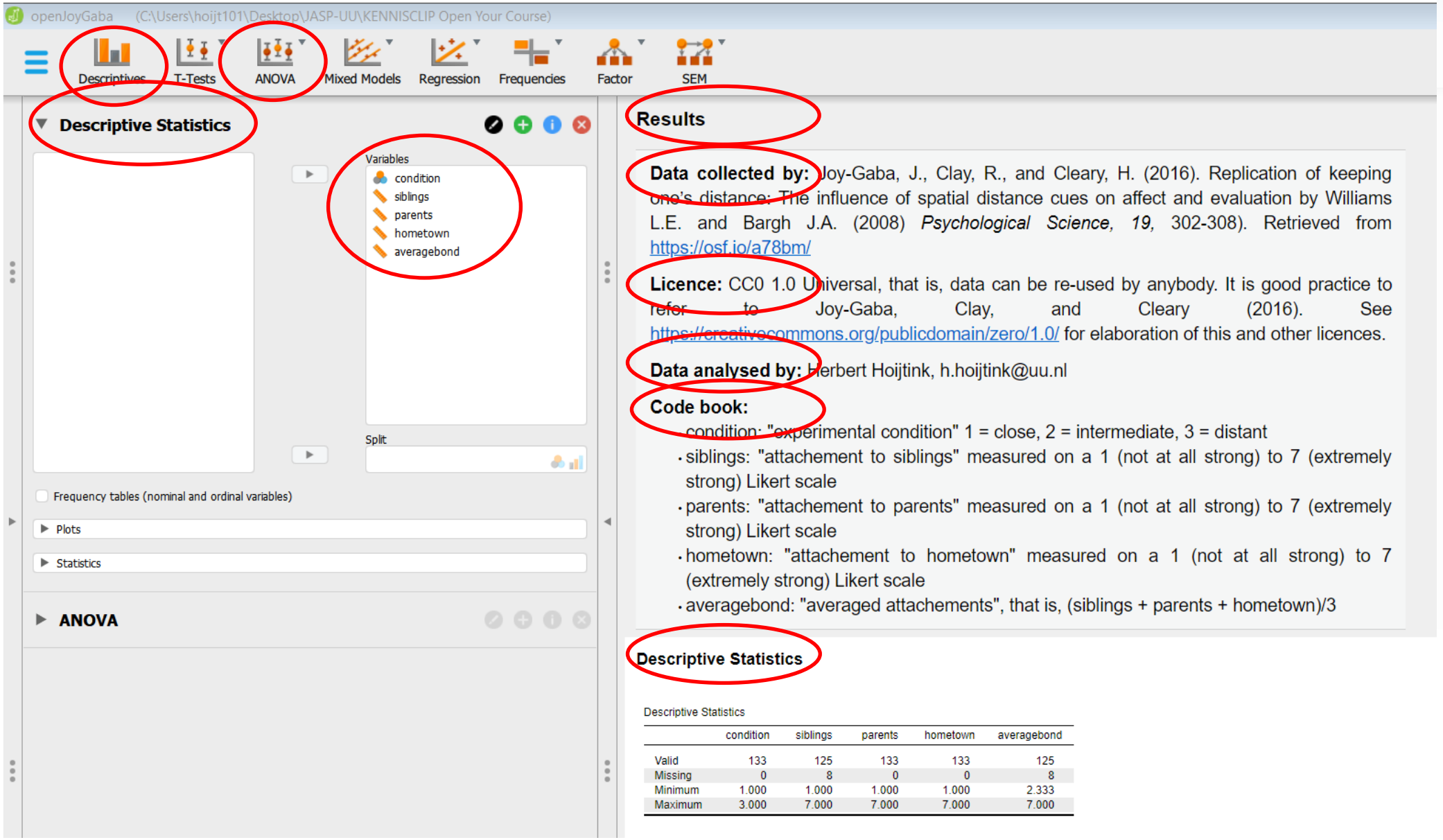
Descriptives | T-Tests | ANOVA | Mixed Models | Regression | Frequencies | Factor | SEM

	condition	siblings	parents	hometown	averagebond
1	close	6	7	5	6
2	intermediate	6	6	3	5
3	distant	5	6	3	4.66667
4	close	5	5	4	4.66667
5	intermediate	6	5	4	5
6	distant	5	6	3	4.66667
7	intermediate	6	6	6	6
8	close	5	4	6	5
9	close	7	5	5	5.66667
10	close	7	5	6	6
11	close	7	6	5	6
12	intermediate		7	7	
13	close	7	7	6	6.66667
14	intermediate	6	5	4	5
15	distant	6	5	7	6
16	close	6	6	5	5.66667
17	distant	5	6	2	4.33333
18	close	5	6	5	5.33333
19	intermediate	6	6	5	5.66667
20	intermediate	6	5	3	4.66667
21	intermediate	7	7	5	6.33333
22	close	6	7	6	6.33333
23	distant	6	7	6	6.33333
24	intermediate	7	7	3	5.66667
25	distant	3	5	2	3.33333
26	distant	7	7	5	6.33333
27	intermediate	5	6	6	5.66667
28	distant	5	6	3	4.66667
29	close	5	5	6	5.33333
30	distant	6	7	5	6

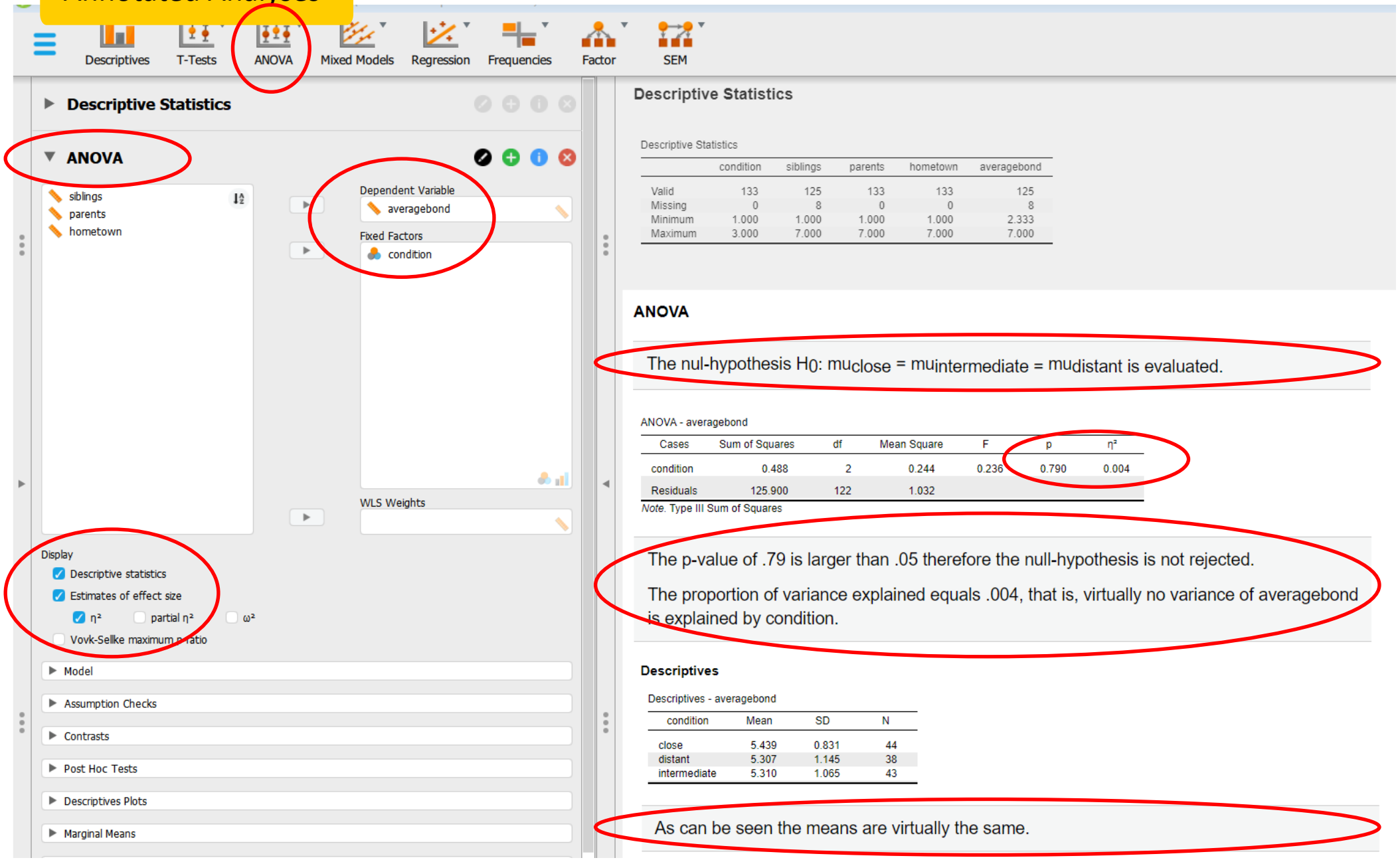
Include a Codebook & Annotations of your Analyses

Explain using annotations which analyses were executed and what your interpretation of the outcomes was

Data collected by, License, Data analyzed by, Code book, and Descriptive Statistics



Annotated Analyses



	condition	siblings	parents	hometown	averagebond
Valid	133	125	133	133	125
Missing	0	8	0	0	8
Minimum	1.000	1.000	1.000	1.000	2.333
Maximum	3.000	7.000	7.000	7.000	7.000

ANOVA

The nul-hypothesis H_0: $mu_{close} = mu_{intermediate} = mu_{distant}$ is evaluated.

ANOVA - averagebond

Cases	Sum of Squares	df	Mean Square	F	p	η²
condition	0.488	2	0.244	0.236	0.790	0.004
Residuals	125.900	122	1.032			

Note. Type III Sum of Squares

The p-value of .79 is larger than .05 therefore the null-hypothesis is not rejected.

The proportion of variance explained equals .004, that is, virtually no variance of averagebond is explained by condition.

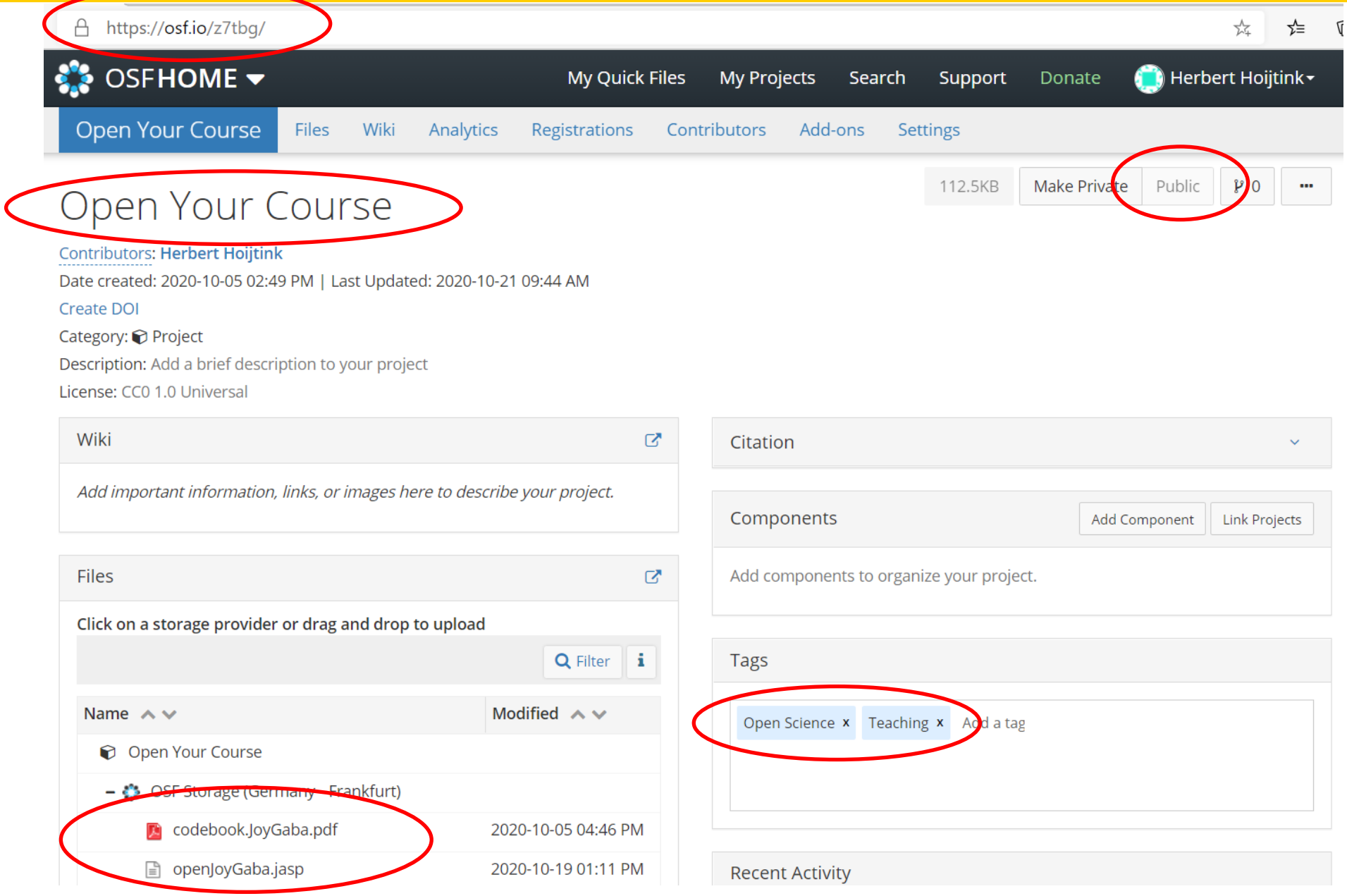
Descriptives

Descriptives - averagebond

condition	Mean	SD	N
close	5.439	0.831	44
distant	5.307	1.145	38
intermediate	5.310	1.065	43

As can be seen the means are virtually the same.

Findable


https://osf.io/z7tbg/
OSFHOME
My Quick Files
My Projects
Search
Support
Donate
Herbert Hoijtink
Open Your Course
Files
Wiki
Analytics
Registrations
Contributors
Add-ons
Settings
112.5KB
Make Private
Public
0
Open Your Course
Contributors: Herbert Hoijtink
Date created: 2020-10-05 02:49 PM | Last Updated: 2020-10-21 09:44 AM
Create DOI
Category: Project
Description: Add a brief description to your project
License: CC0 1.0 Universal
Wiki
Add important information, links, or images here to describe your project.
Files
Click on a storage provider or drag and drop to upload
Filter
Name
Modified
Open Your Course
OSF Storage (Germany - Frankfurt)
codebook.JoyGaba.pdf
2020-10-05 04:46 PM
openJoyGaba.jasp
2020-10-19 01:11 PM
Citation
Components
Add Component
Link Projects
Add components to organize your project.
Tags
Open Science x
Teaching x
Add a tag
Recent Activity

How To …

Use one of the example data sets that downloaded with this presentation: data.sav, data.txt, or data.csv.

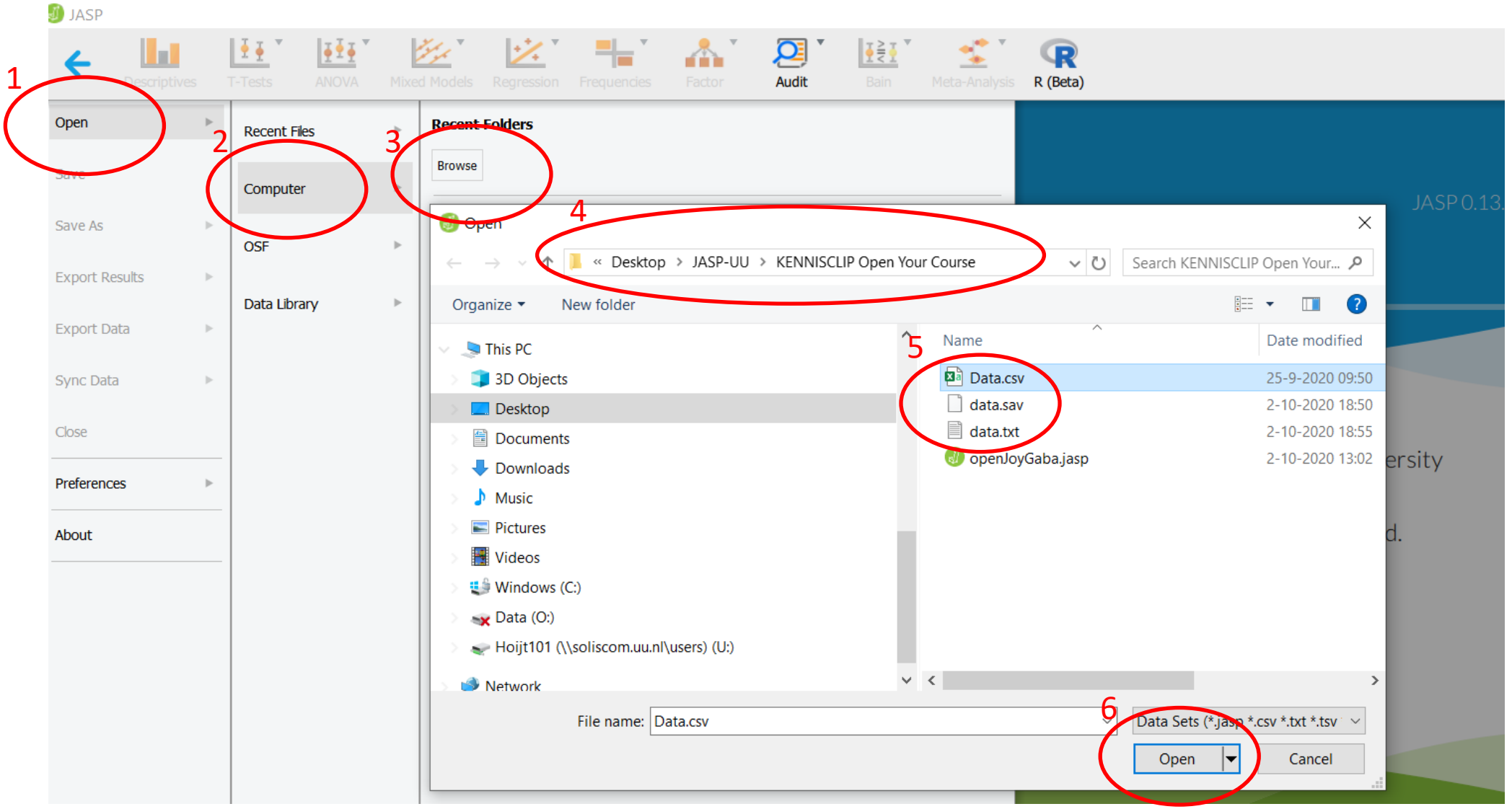
Execute each of the steps in the following slides.

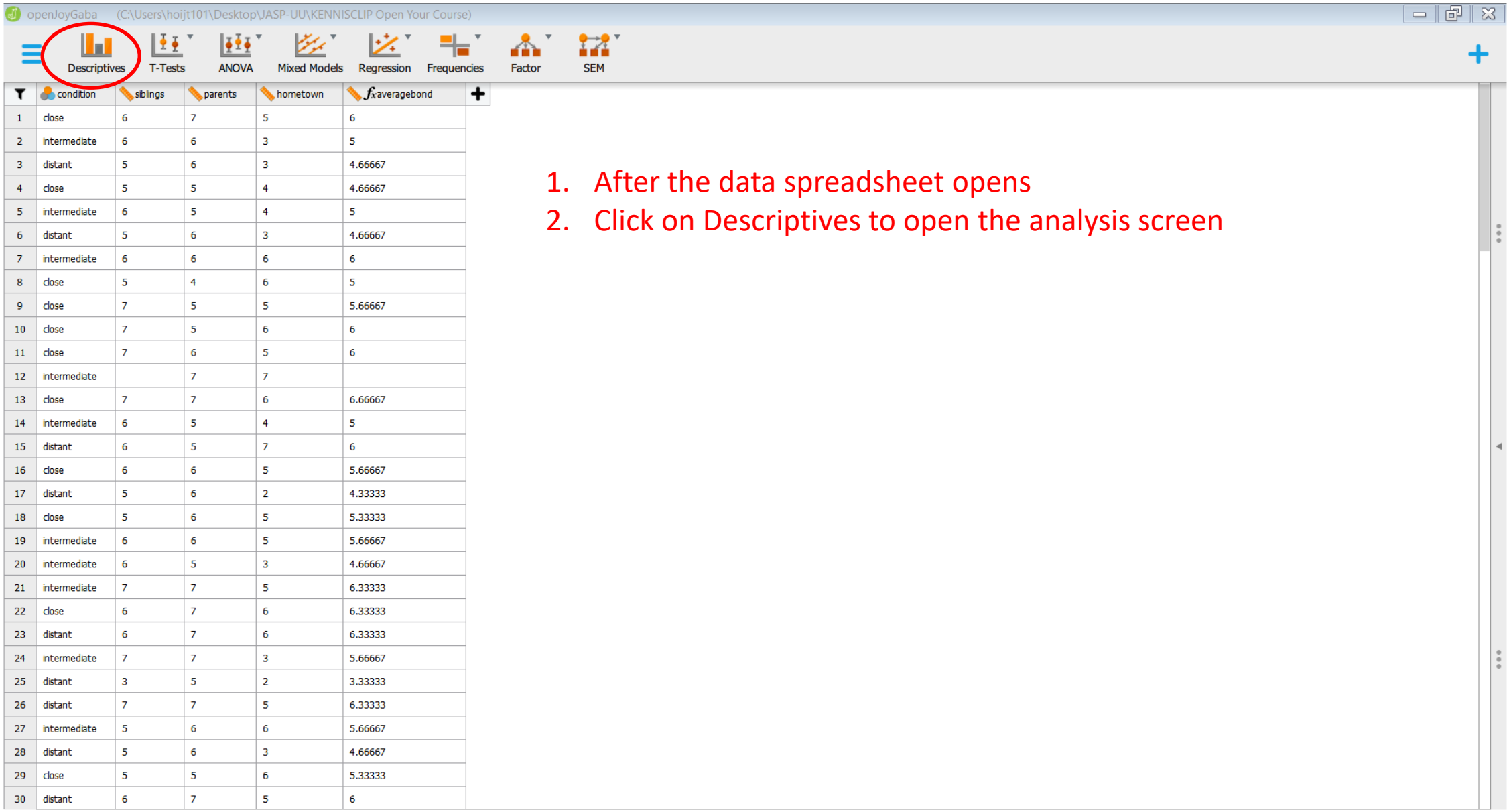
How to Create a mydata.jasp File

Make certain that JASP is installed on your computer. If not, install it from https://jasp-stats.org/download/

If JASP is installed, start the program.

Open a .csv, .sav, or .txt file containing your data

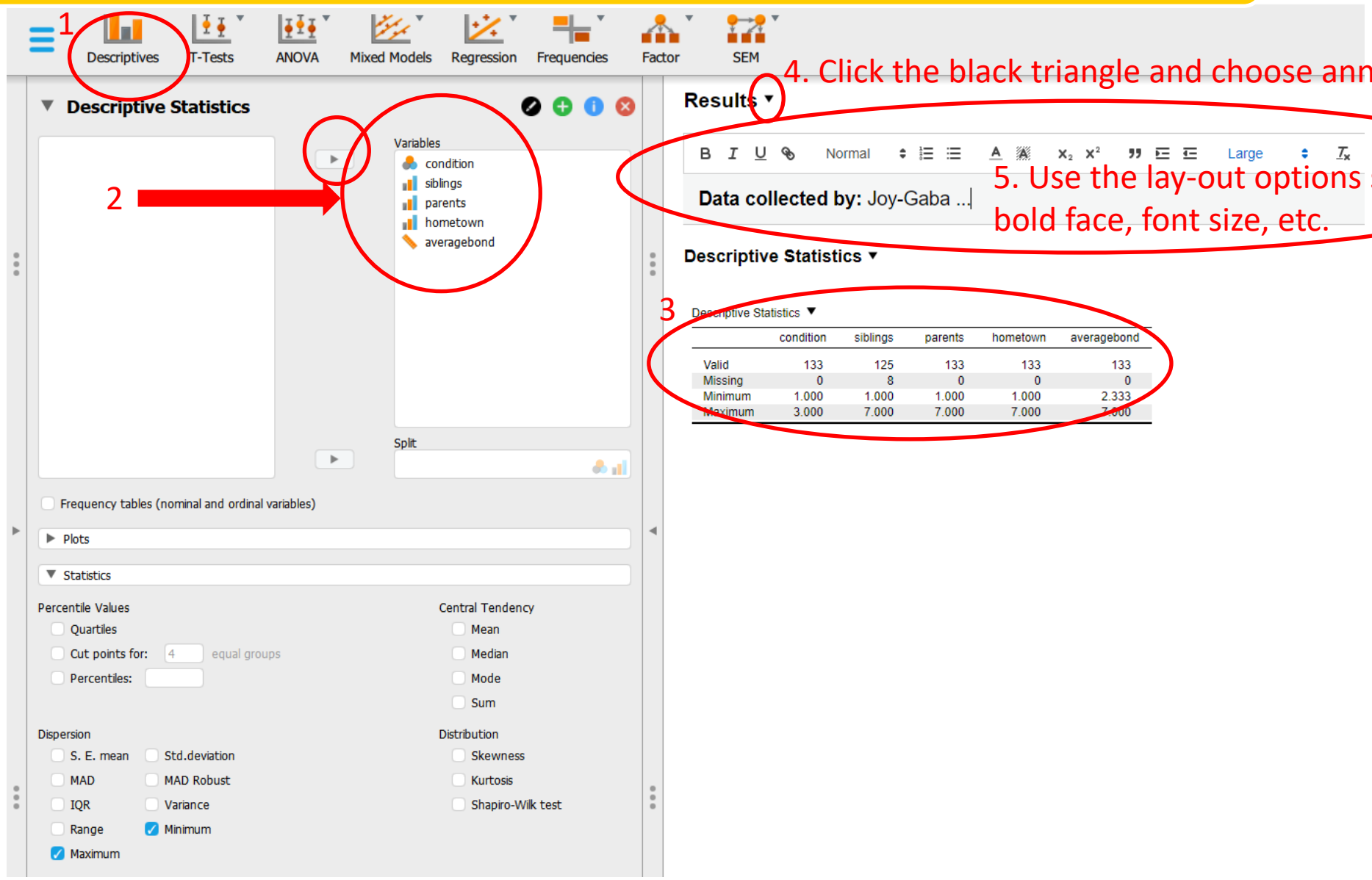




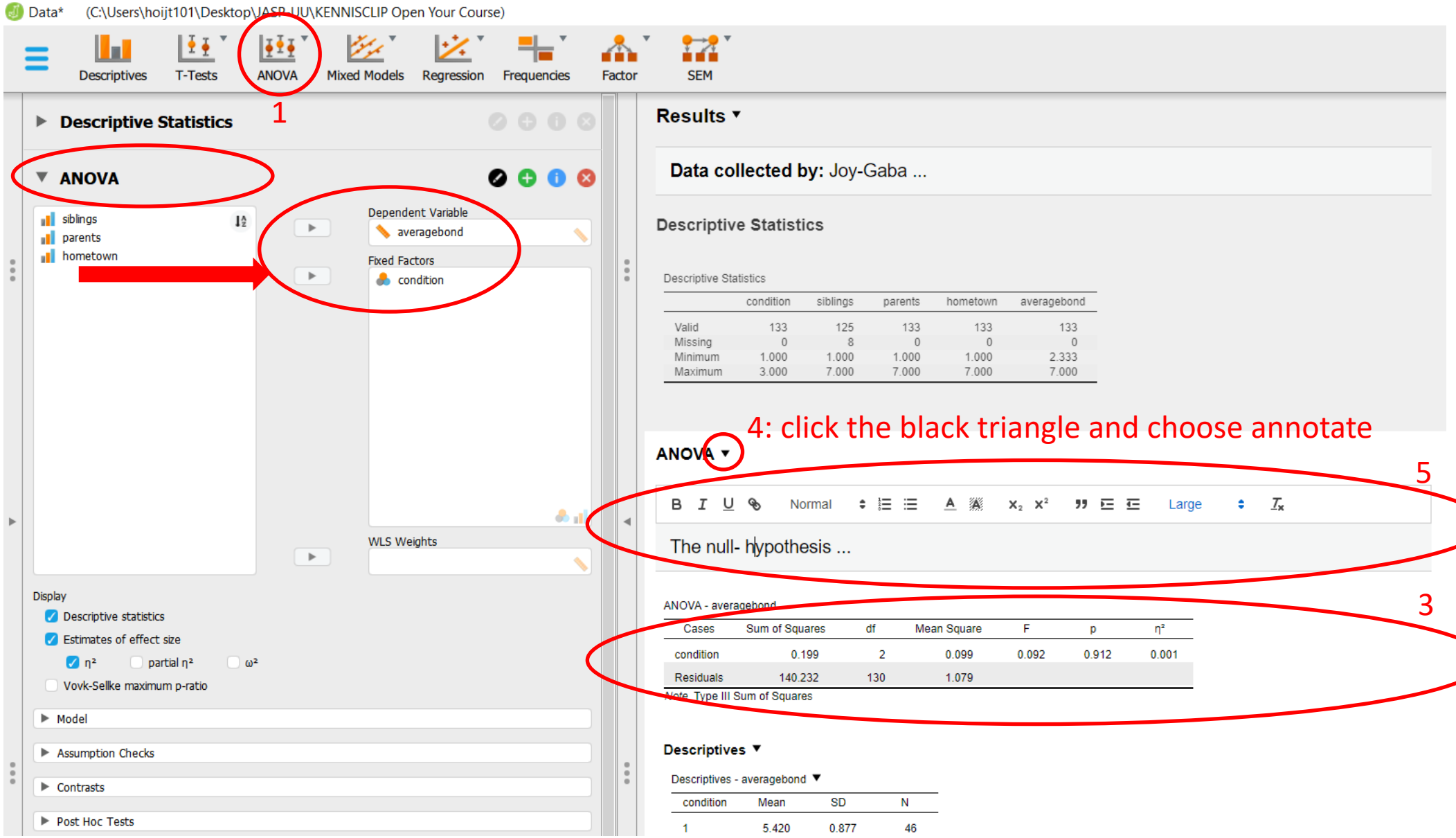
	condition	siblings	parents	hometown	averagebond
1	close	6	7	5	6
2	intermediate	6	6	3	5
3	distant	5	6	3	4.66667
4	close	5	5	4	4.66667
5	intermediate	6	5	4	5
6	distant	5	6	3	4.66667
7	intermediate	6	6	6	6
8	close	5	4	6	5
9	close	7	5	5	5.66667
10	close	7	5	6	6
11	close	7	6	5	6
12	intermediate		7	7	
13	close	7	7	6	6.66667
14	intermediate	6	5	4	5
15	distant	6	5	7	6
16	close	6	6	5	5.66667
17	distant	5	6	2	4.33333
18	close	5	6	5	5.33333
19	intermediate	6	6	5	5.66667
20	intermediate	6	5	3	4.66667
21	intermediate	7	7	5	6.33333
22	close	6	7	6	6.33333
23	distant	6	7	6	6.33333
24	intermediate	7	7	3	5.66667
25	distant	3	5	2	3.33333
26	distant	7	7	5	6.33333
27	intermediate	5	6	6	5.66667
28	distant	5	6	3	4.66667
29	close	5	5	6	5.33333
30	distant	6	7	5	6

1. After the data spreadsheet opens
2. Click on Descriptives to open the analysis screen

Execute your Analyses – Start with Simple Descriptives – and Add Notes to your Results



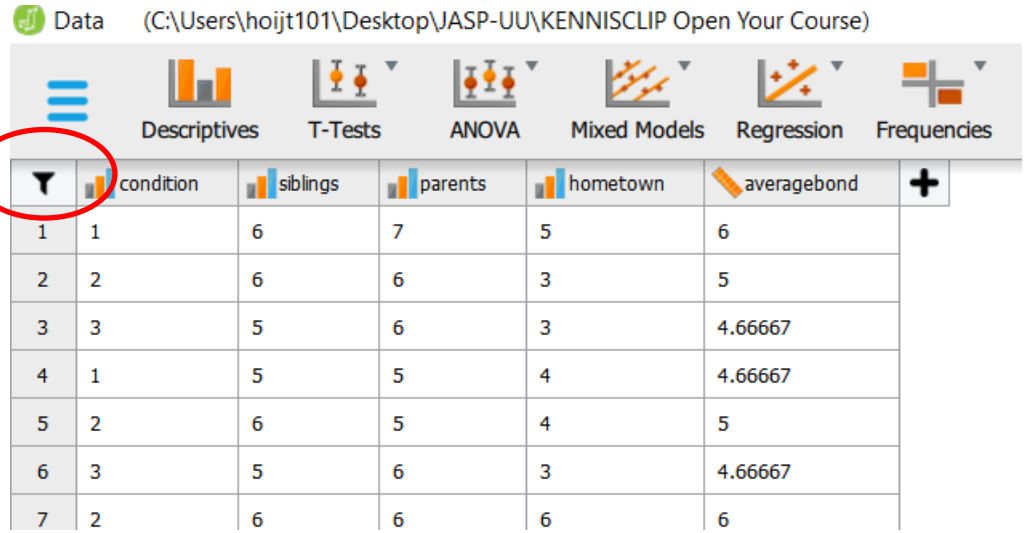
Continue your Analyses with an ANOVA and Add Notes to your Results



Please note that …

Currently you have to make a different mydata.jasp file for each selection of cases you make, e.g., before filtering (marked by the red ellipse), and before deleting or adding cases. Otherwise, all your analyses will be based on the latest selection of cases created.

This will be remedied in one of the future JASP releases.

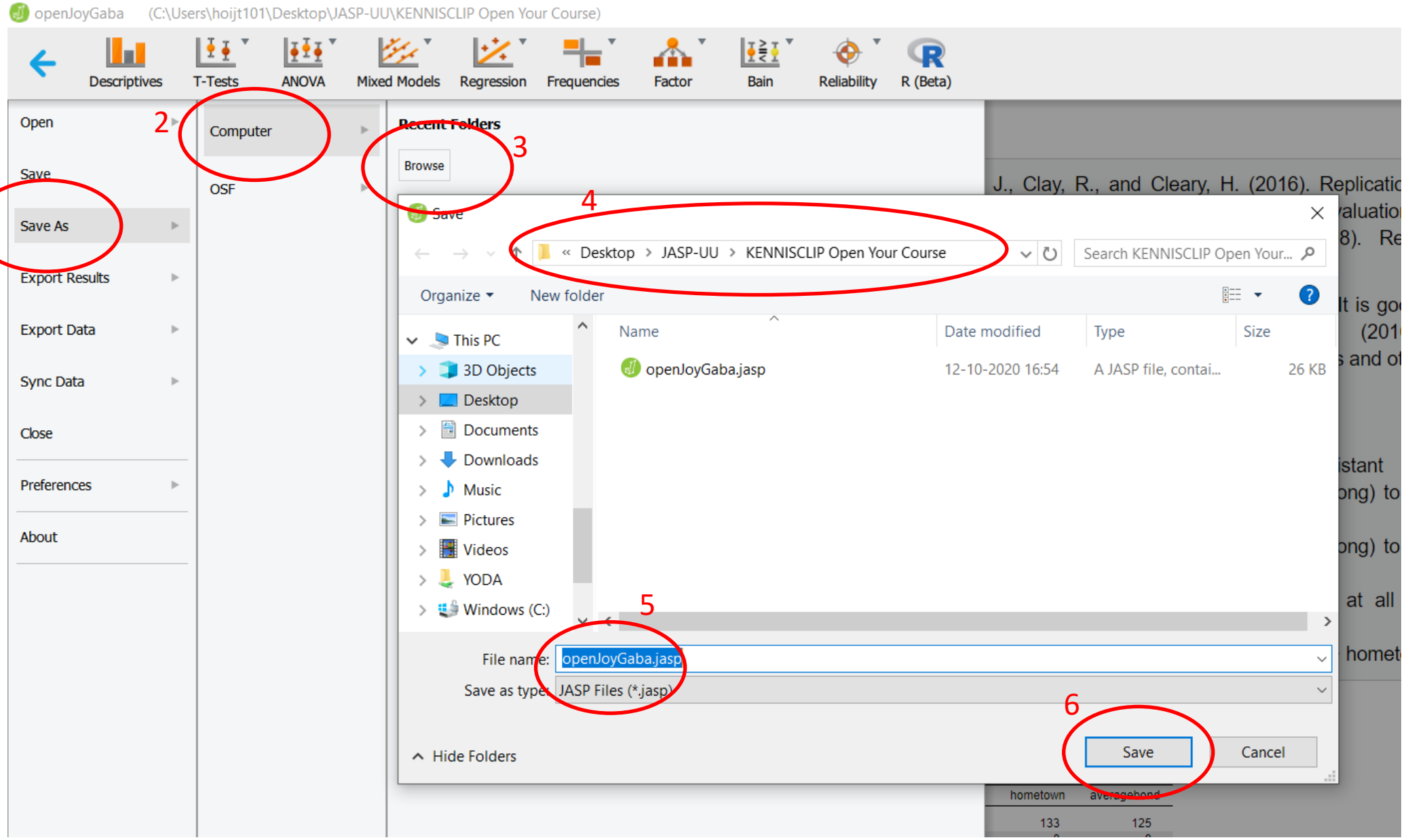


	condition	siblings	parents	hometown	averagebond
1	1	6	7	5	6
2	2	6	6	3	5
3	3	5	6	3	4.66667
4	1	5	5	4	4.66667
5	2	6	5	4	5
6	3	5	6	3	4.66667
7	2	6	6	6	6

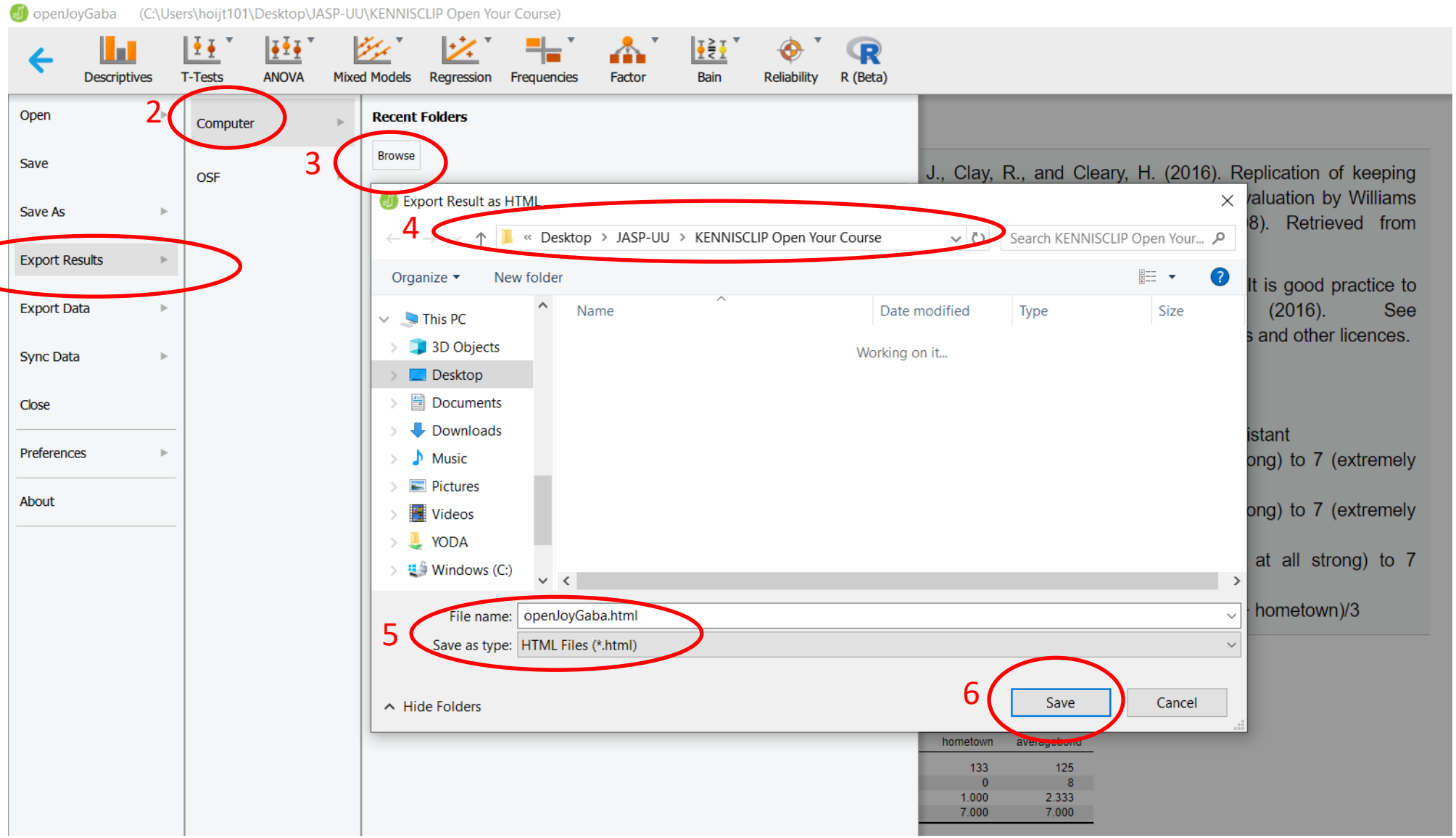
A knowledge clip explaining the main features of JASP (selecting cases, computing variables, recoding, opening a data file, saving a mydata.jasp file, executing analyses, getting help, etc.) can be found at https://osf.io/z7tbg/

A virtually complete elaboration of all the features of JASP can be found at https://jasp-stats.org/how-to-use-jasp/

Save as a mydata.jasp File on your Computer – contains data, analyses input and annotated analyses results

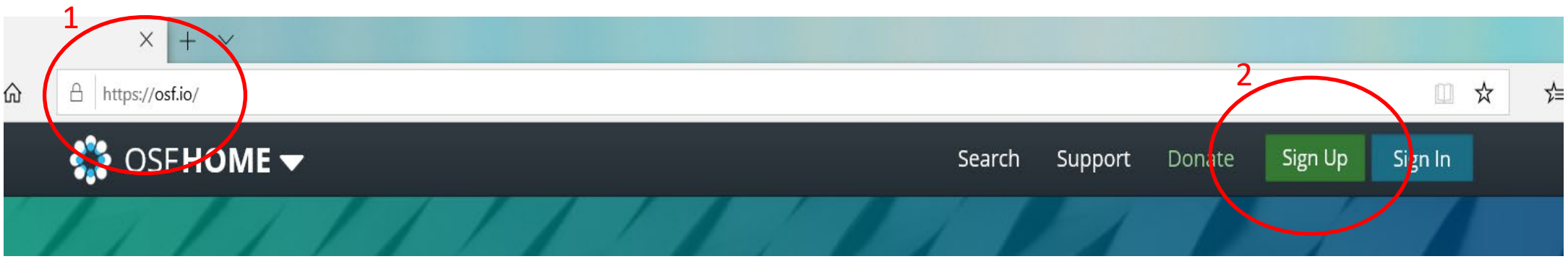


Save a mydata.jasp file as a myresults.html File on your Computer – only contains the annotated results

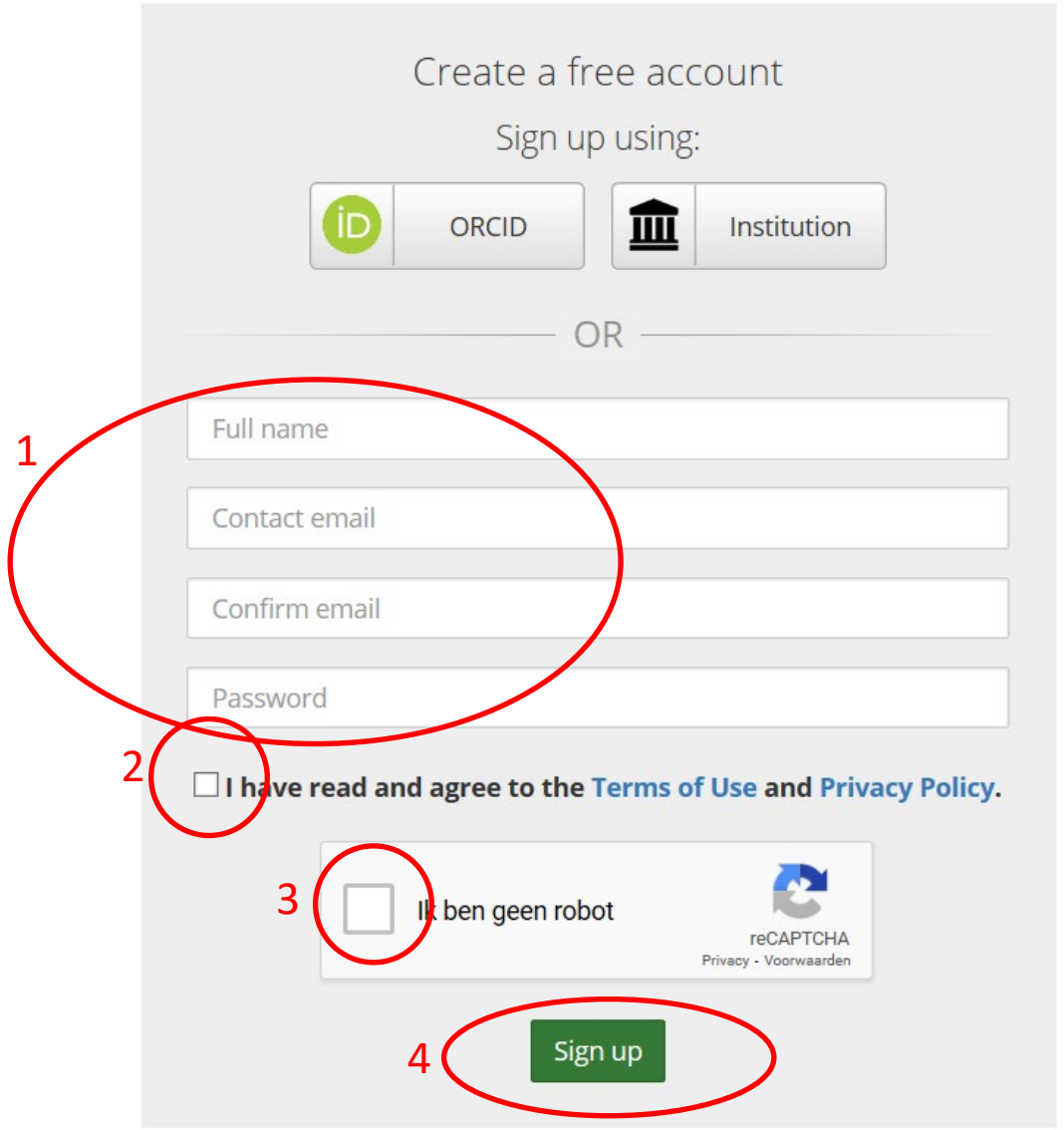


How to Make your Data Analyses Findable Using the Open Science Foundation

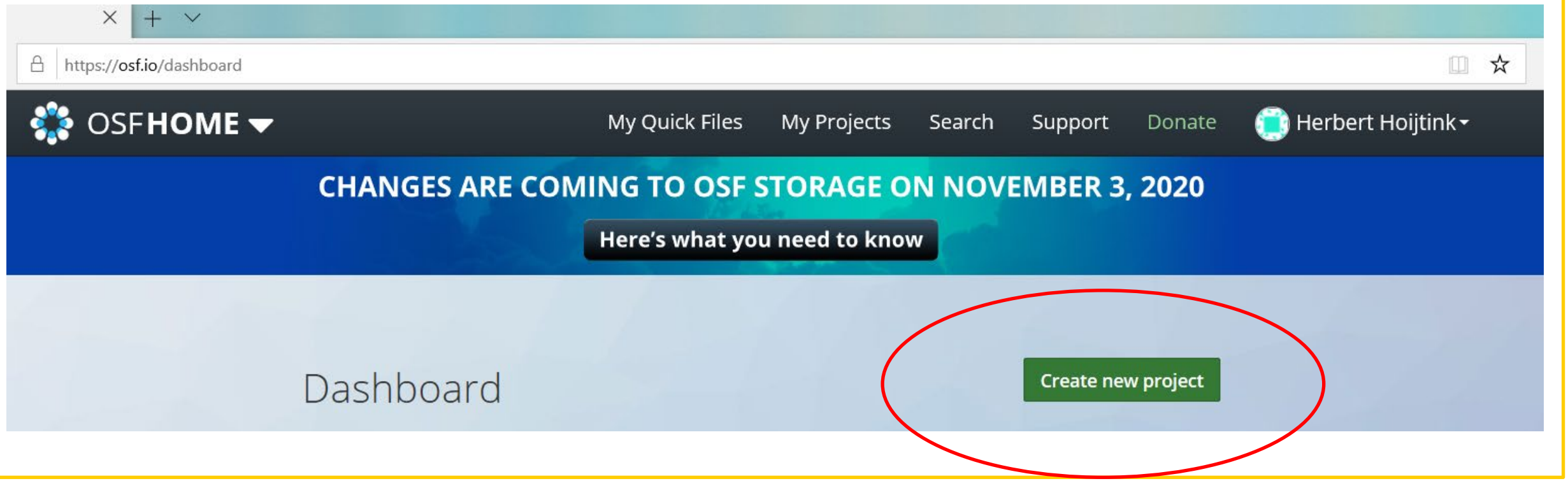
Create a Project to Store your Data at the Open Science Foundation in Four Simple Steps



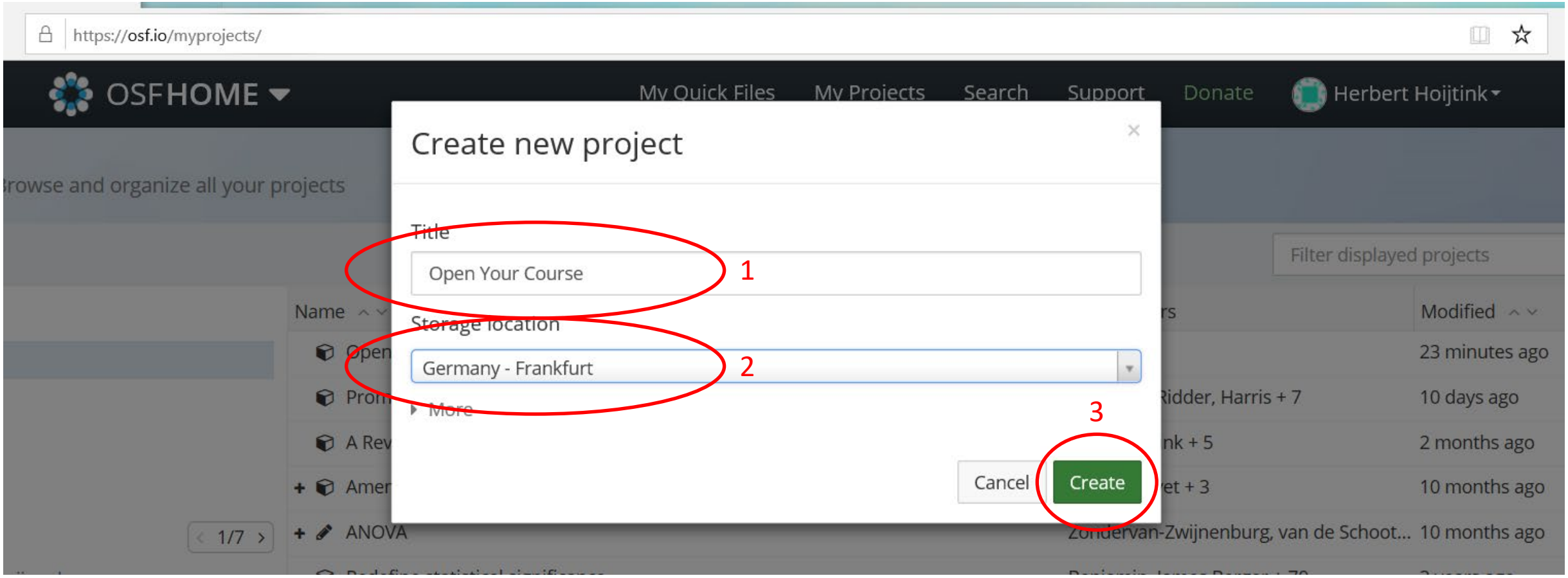
Step 1: Create a Free Account



Step 2: Create a New Project



Step 2: Create a New Project (continued)



Step 3: Make your Project Publicly Available

https://osf.io/z7tbg/ 1: anybody can surf here to access your data analyses if …

OSFHOME | My Quick Files | My Projects | Search | Support | Donate | Herbert Hoijtink

Open Your Course | Files | Wiki | Analytics | Registrations | Contributors | Add-ons | Settings

112.5KB | Make Private | Public | 0 | ···

2: … you make your project publicly available

Open Your Course

Contributors: **Herbert Hoijtink**

Date created: 2020-10-05 02:49 PM | Last Updated: 2020-10-21 09:44 AM

Create DOI

Category: Project

Description: Add a brief description to your project

License: CC0 1.0 Universal 3

Wiki

Add important information, links, or images here to describe your project.

Citation

Components | Add Component | Link Projects

Add components to organize your project.

4

Files

Click on a storage provider or drag and drop to upload

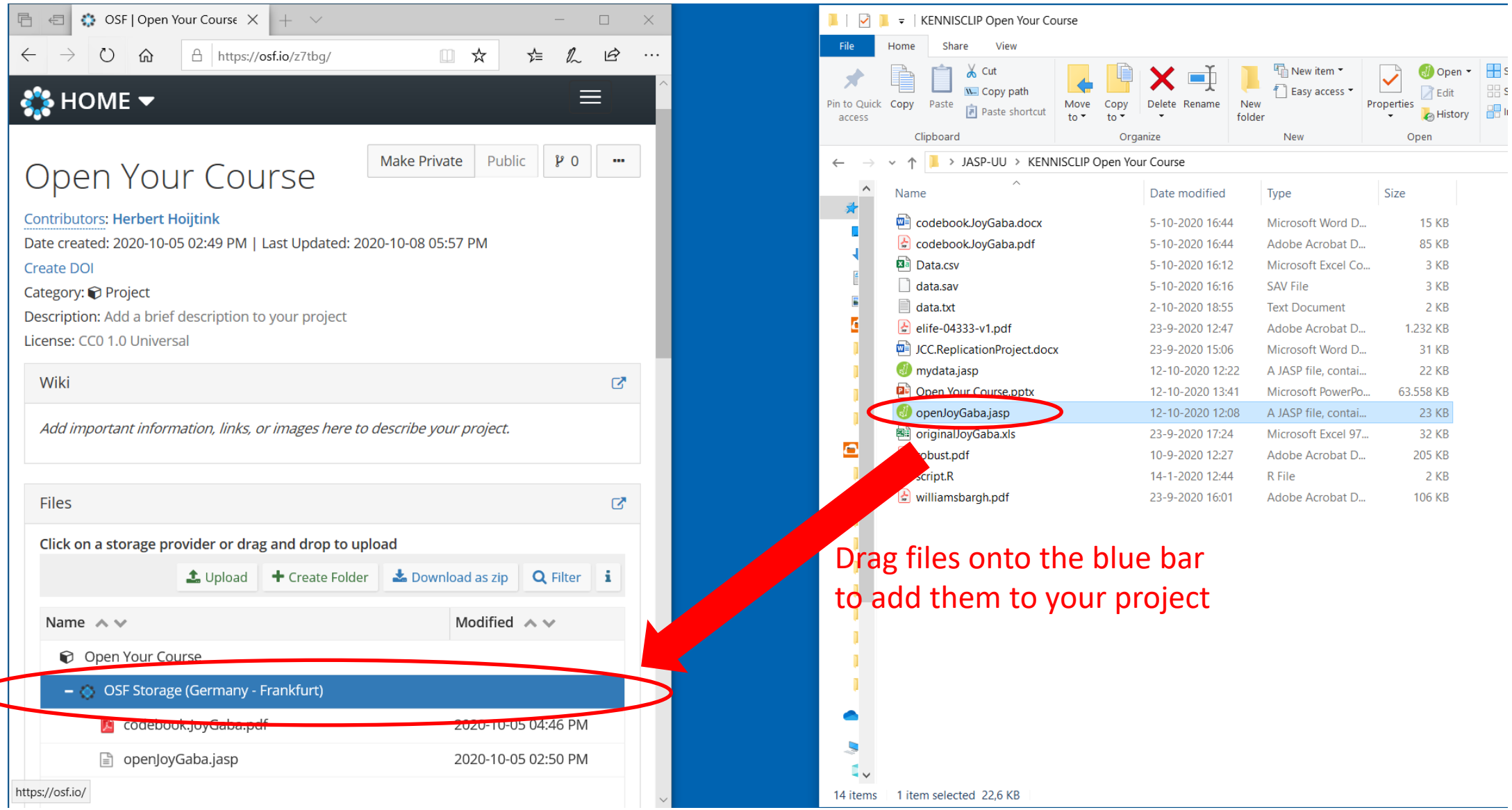
Filter

Name	Modified
Open Your Course	

Tags

Open Science x | Teaching x | Add a tag

Drag your Data and other Relevant Files into the Project you Created



The Conflict between Fully Open Data Analyses and Privacy Regulations

Personal data can only be published if privacy is guaranteed. One way to achieve this is, if data are truly anonymized (General Data Protection Regulation: https://gdpr-info.eu/).

- Truly anonymized data is no longer personal data and thus no longer subject to the GDPR
- It involves the complete and utter removal of all personal identifiers in a database
- Anonymized data can no longer be attributed to any particular individual by any means

If anonymization cannot be achieved, you can still “Open your Analyses” by publishing only your code book, results, references, license and annotated analyses (these are often called meta data). This can be done using a myresults.html file that can be created based on your mydata.jasp file.

This page was constructed using quotes and input from Handling Personal Data in Research (Utrecht University)

The Conflict between Fully Open Data Analyses and your Future Research Plans

- Therefore, you may want to impose restrictions on the use of your data, e.g., “can only be used to reproduce the analyses you executed”
- However, this cannot be arranged via the application of a license, data are often considered to be facts and facts cannot be copyrighted (OpenAIRE)
- In such cases you can consider publishing only a part of your data, e.g., only the data and analyses that are used in a specific research report (but do provide complete meta data, most importantly, the code book of your complete data set)
- Then the unused data are only available for you and thereby you avoid being scooped out of your next paper

OpenAIRE: https://www.openaire.eu/how-do-i-license-my-research-data

Open Science

Open Science is more than having publications and data openly available …

Open Science implies openly creating, sharing and assessing research, wherever viable

Open Science

Open to participation

- No barriers based on race, gender, income, status, language, ableness, age
- Involvement of societal partners in research priority setting
- Evaluations that include societal relevance
- Citizen science & co-creation
- Broadly considering all knowledge (including local knowledge)
- Error friendly environment

Open to (re)use

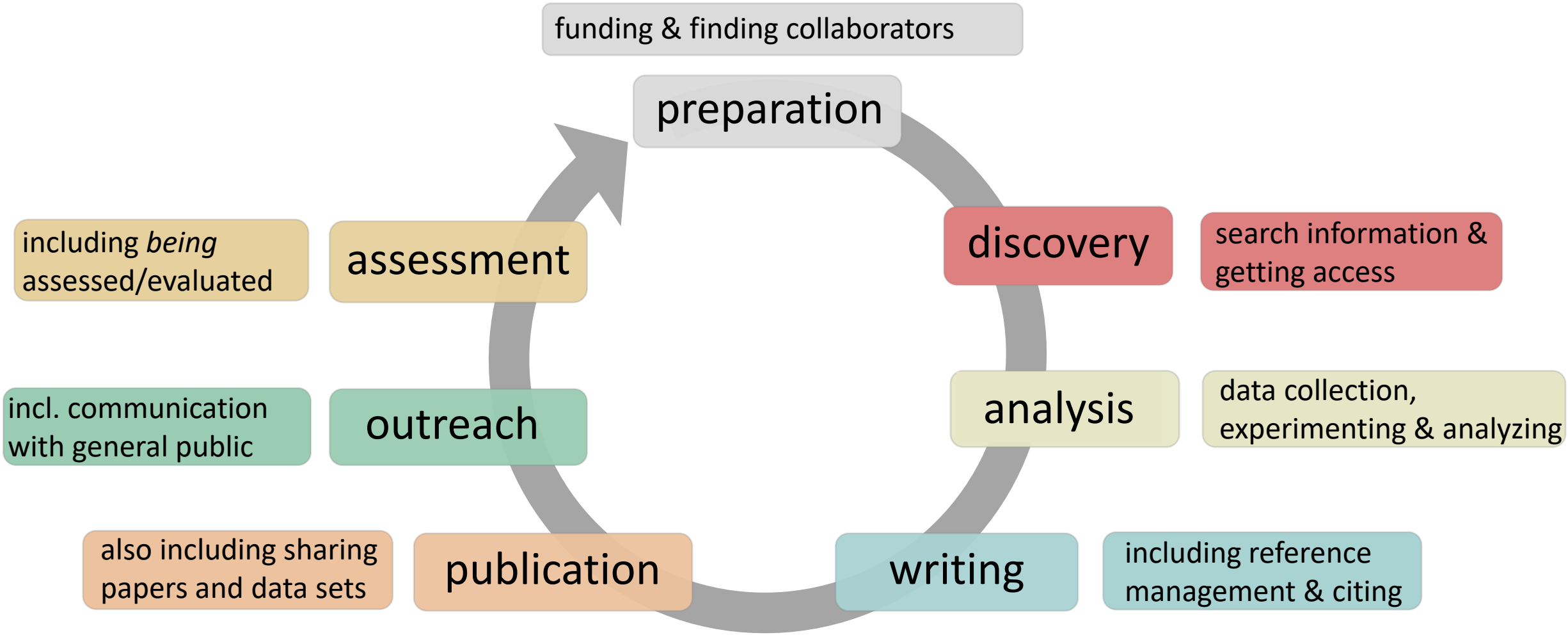
- Open Access, f. people/machines, to:
  - Proposals and applications
  - Preregistrations
  - Data
  - Code
  - Preprints, working papers
  - Papers and books
  - Reviews and comments
  - Posters and presentations
- Open, non-proprietary standards
- Open licences
- Full documentation of process, including negative results

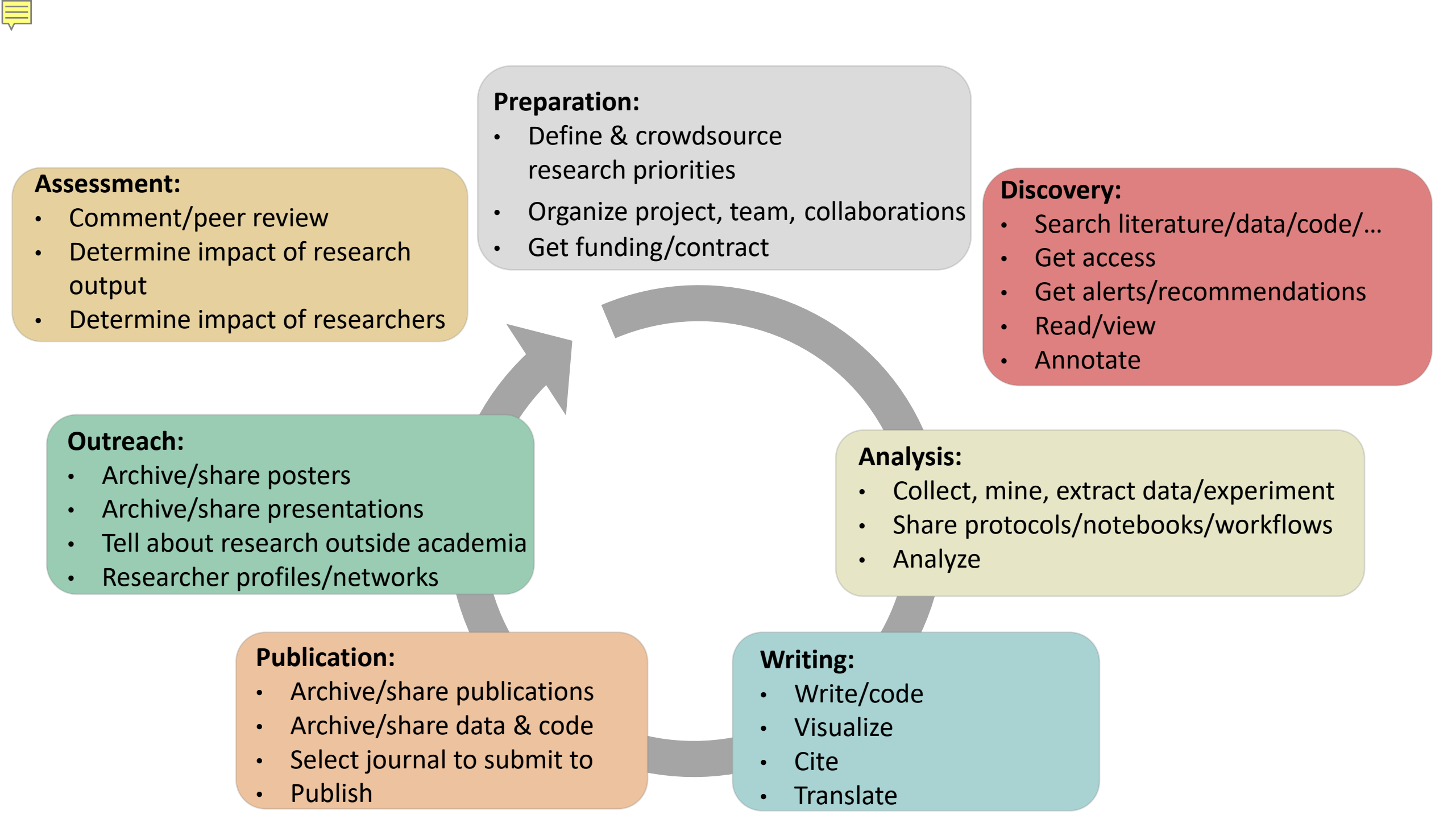
Open to the world

- Translations
- Plain language explanations
- Outreach beyond academia
- Open to questions from outside academia
- Curation and annotation of non-scholarly information
- Participation in public debate

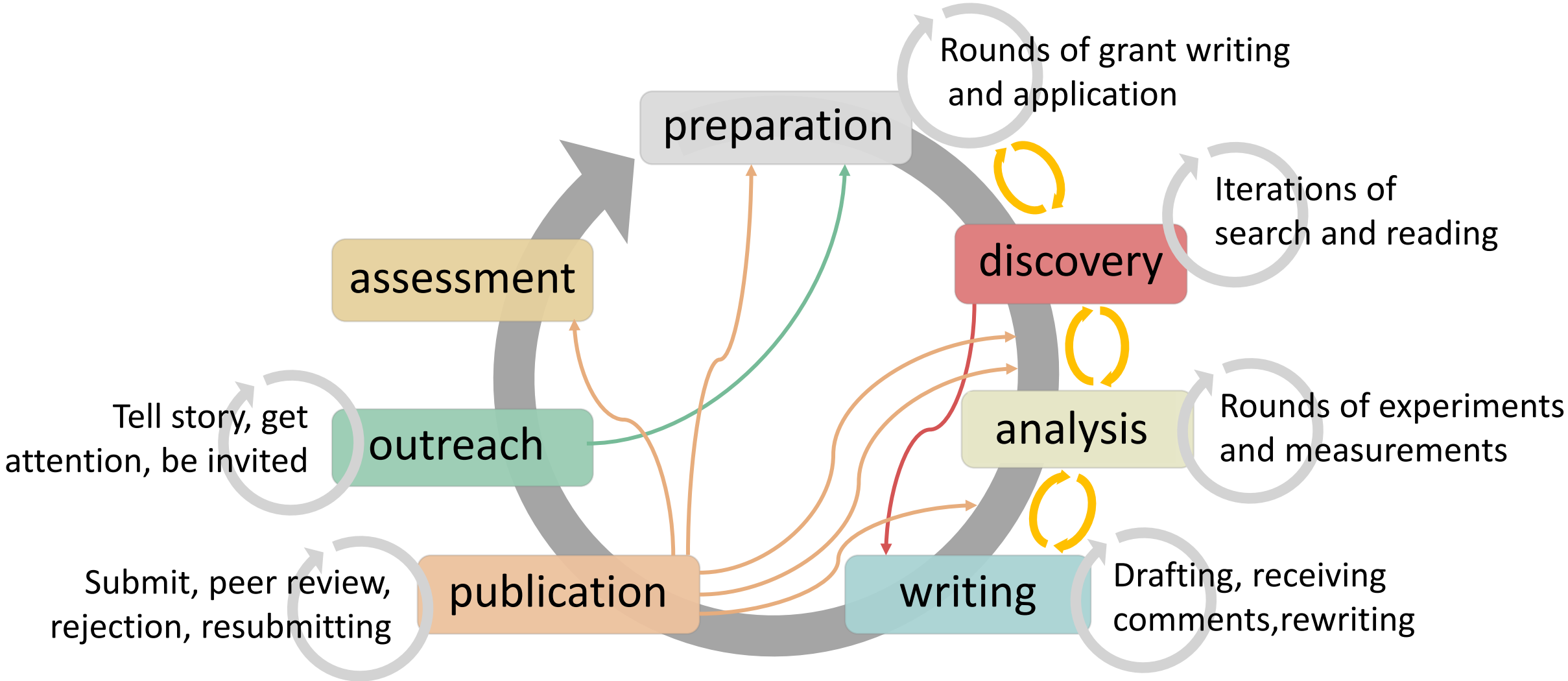
From: Bosman & Kramer (2017) Defining open science definitions

The Open Research Workflow

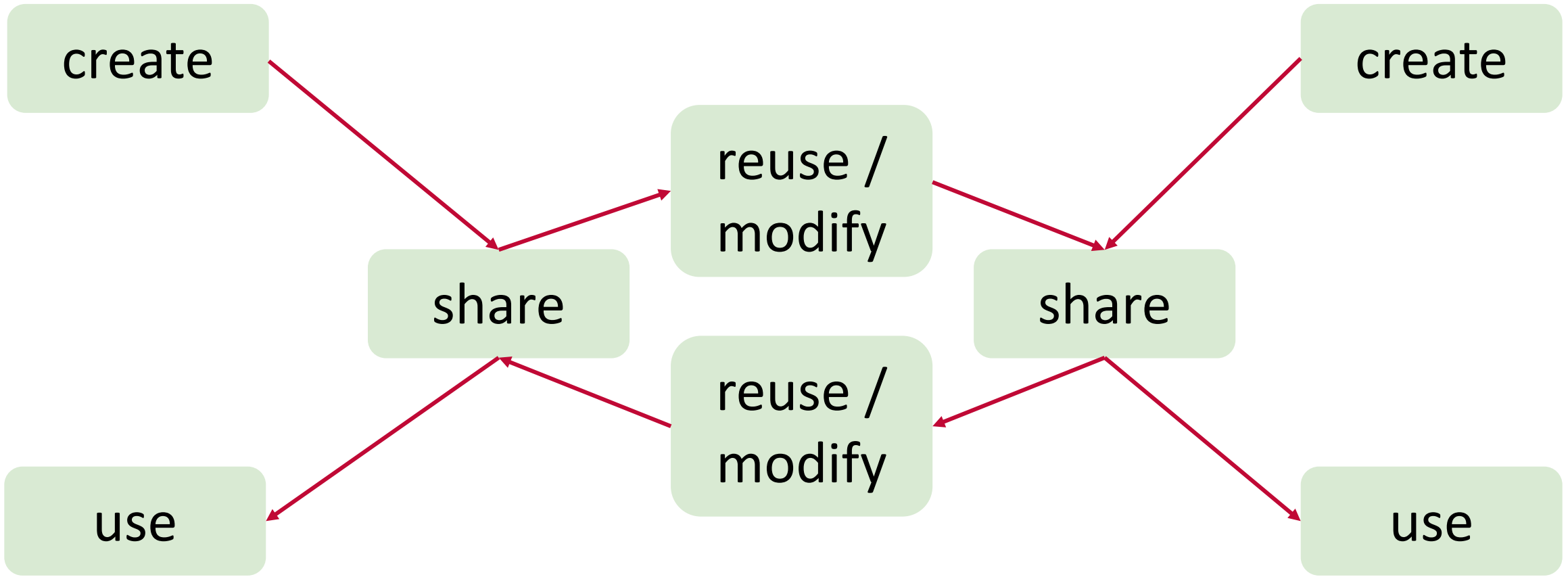



Preparation:
• Define & crowdsource research priorities
• Organize project, team, collaborations
• Get funding/contract
Discovery:
• Search literature/data/code/…
• Get access
• Get alerts/recommendations
• Read/view
• Annotate
Analysis:
• Collect, mine, extract data/experiment
• Share protocols/notebooks/workflows
• Analyze
Writing:
• Write/code
• Visualize
• Cite
• Translate
Publication:
• Archive/share publications
• Archive/share data & code
• Select journal to submit to
• Publish
Outreach:
• Archive/share posters
• Archive/share presentations
• Tell about research outside academia
• Researcher profiles/networks
Assessment:
• Comment/peer review
• Determine impact of research output
• Determine impact of researchers

The Open Research Workflow



Open Science as a mindset



Why Open Science?

- Transparency, accountability
- Efficiency
- Reproducibility & verifiability
- Relevance & stakeholder involvement