

Reproducible Research: What, Why, and How?

Thomas J. Leeper

Department of Political Science and Government
Aarhus University

October 28, 2014

- Reinhart and Rogoff

- Reinhart and Rogoff
- Psychology's "replication crisis"

- Reinhart and Rogoff
- Psychology's "replication crisis"
- "Most published research findings are false"

- Reinhart and Rogoff
- Psychology's "replication crisis"
- "Most published research findings are false"
- Diedrick Stapel

1 Why?

2 What?

3 How?

1 Why?

2 What?

3 How?

Why reproducible research?

- External reasons
- Internal reasons

External Reasons

External Reasons

- Philosophical perspective

External Reasons

- Philosophical perspective
- Journal requirements

External Reasons

- Philosophical perspective
- Journal requirements
- Funding agency requirements

External Reasons

- Philosophical perspective
- Journal requirements
- Funding agency requirements
- The coming revolution

Internal Reasons

Internal Reasons

- Confidence in your own work

Internal Reasons

- Confidence in your own work
- Easier workflow

Internal Reasons

- Confidence in your own work
- Easier workflow
- Easier collaboration

So what does that mean?

So what does that mean?



Open Science
@openscience



Following

"Reproducibility is collaboration with people you don't know, incl. yourself next week." –
[@philipbstark](#) [#openscience](#)



So what does that mean?

- 1 Do it for *yourself* first!
- 2 Do it for *science* second.

Why is research still irreproducible?

Why is research still irreproducible?

Barriers to Data and Code Sharing in Computational Science

Survey of Machine Learning Community, NIPS (Stodden, 2010):

Code		Data
77%	Time to document and clean up	54%
52%	Dealing with questions from users	34%
44%	Not receiving attribution	42%
40%	Possibility of patents	-
34%	Legal Barriers (ie. copyright)	41%
-	Time to verify release with admin	38%
30%	Potential loss of future publications	35%
30%	Competitors may get an advantage	33%
20%	Web/disk space limitations	29%

Why is research still irreproducible?

Why is research still irreproducible?

1 Technology



Why is research still irreproducible?

1 Technology

Why is research still irreproducible?

- 1 Technology
- 2 Individual actions

Why is research still irreproducible?

- 1 Technology
- 2 Individual actions
- 3 Collective behavior and norms

1 Why?

2 What?

3 How?

So what is reproducible research?

So what is reproducible research?

- Evolving standards and technology

So what is reproducible research?

- Evolving standards and technology
- Discipline-specific meaning

So what is reproducible research?

- Evolving standards and technology
- Discipline-specific meaning
- Hard to define

American Association for Public Opinion Research “Disclosure Standards”

American Political Science Association “A Guide to Professional Ethics in Political Science”

American Psychological Association “Ethical Principles of Psychologists and Code of Conduct”

Association for Psychological Science “Submission Guidelines”

American Anthropological Association “Code of Ethics”

CONSORT Group “CONSORT Statement”

European Research Council “Open Access Guidelines for researchers funded by the ERC”

PLoS “Editorial and Publishing Policies”

Irreproducibility

- Fabrication

Irreproducibility

- Fabrication
- Human error

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats
- Unavailable data

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable
- “Available from the author ”

Irreproducibility

- Fabrication
- Human error
- Lack of methodological transparency
- Ambiguous data citations
- Proprietary data and file formats
- Unavailable data
- Analysis uses proprietary software/hardware
- Analysis unavailable
- “Available from the author (now deceased)”

**Kaitlin Thaney**

@kaythaney



Following

"'Reproducible research' is a redundant term. 'Irreproducible research' just used to be known as 'bullshit'." - [@fperez_org](#)
::slow clap::



RETWEETS

122

FAVORITES

61



6:11 PM - 8 May 2014

Distinguish from other concepts

Distinguish from other concepts

- *Reproducible* versus *Replicable*

Distinguish from other concepts

- *Reproducible* versus *Replicable*
- *Reproducible* versus *Automated*

Distinguish from other concepts

- *Reproducible* versus *Replicable*
- *Reproducible* versus *Automated*
- *Reproducible* versus *True*

Arrive at a definition

Stanford University's David Donoho:

“An article about computational science in a scientific publication is not the scholarship itself, it is merely advertising of the scholarship. The actual scholarship is the complete software development environment and the complete set of instructions which generated the figures.”

Reproducible research
enumerates a complete set
of physical actions needed
to transform transparent
inputs into outputs.

1 Why?

2 What?

3 How?

What makes up the ideal reproducible research product?

Past

- Data and method description
- Closed data and analysis
- Use of proprietary software
- Paywalled publications

Present

- Detailed or full protocols
- Data and analysis sharing (on request)
- Mix of proprietary and open software
- “Green” open access

Future

- Study preregistration and “outcome-blind” review
- Open lab notebooks
- Persistent, archived, open-licensed data
- Open source software
- Open access publication
- Literate, reproducible output

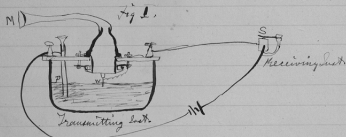
How do you make your work more reproducible?

How do you make your work more reproducible?

**Always think about your
future self!**

(1) Write Everything Down

March 10th 1876



1. The improved instrument shown in Fig. 1 was constructed this morning and tried this evening. P is a brass pipe and W the platinum wire M the mouth piece and S the armature of the Receiving Instrument.

Mr. Watson was stationed in one room with the Receiving Instrument. He pressed one ear closely against S and closed his other ear with his hand. The Transmitting Instrument was placed in another room and the doors of both rooms were closed.

I then shouted into M the following sentence: "Mr. Watson - Come here - I want to

see you." To my delight he came and declared that he had heard and understood what I said.

I asked him to repeat the words - ~~He said~~. He answered "You said 'Mr. Watson - come here - I want to see you'." He then changed places and I listened at S while Mr. Watson read a few passages from a book into the mouth piece M.

It was certainly the case that articulate sounds proceeded from S. The effect was loud but indistinct and muffled.

If I had read beforehand the passage given by Mr. Watson I should have recognized every word. As it was I could not make out the sense - but on occasional word here and there ~~was~~ quite distinct. I made out "to" and "out" and "further"; and finally the sentence "Mr. Bell do you understand what I say? Do - You - under - stand - what - I - say" came quite clearly and intelligibly. No sound was audible when the armature S was removed.

(1) Write Everything Down

(1) Write Everything Down

- 1 Mark up your analysis files

(1) Write Everything Down

- 1 Mark up your analysis files
- 2 Write (and maintain) your research protocols

(1) Write Everything Down

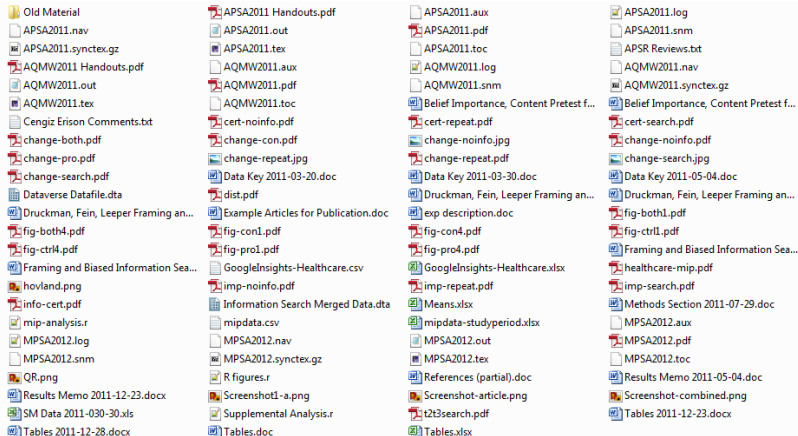
- 1 Mark up your analysis files
- 2 Write (and maintain) your research protocols
- 3 Keep codebooks, questionnaires, and stimulus materials

(1) Write Everything Down

- 1 Mark up your analysis files
- 2 Write (and maintain) your research protocols
- 3 Keep codebooks, questionnaires, and stimulus materials
- 4 Try version control

(2) Get Organized

My dissertation folder



(2) Get Organized

- 1 Use a folder structure than can be shared

Project Directory Structure

- Data
- Analysis
- Figures
- Tables
- Paper
- Presentation
- Materials
- README

Project Directory Structure

- Data
 - RawData.csv
 - CleanData.csv
 - Codebook.txt
- Analysis
- Figures
- Tables
- Paper
- Presentation
- Materials
- README

Project Directory Structure

- Data
- Analysis
 - GatherAndMerge.R
 - DataCleaning.R
 - Descriptives.R
 - Regression.R
 - Figures.R
- Figures
- Tables
- Paper
- Presentation
- Materials
- README

Project Directory Structure

- Data
- Analysis
- Figures
 - Distributions.png
 - MarginalEffects.png
 - PredictedValues.png
- Tables
- Paper
- Presentation
- Materials
- README

Project Directory Structure

- Data
- Analysis
- Figures
- Tables
 - Descriptives.tex
 - Regression.tex
 - MarginalEffects.tex
- Paper
- Presentation
- Materials
- README

Project Directory Structure

- Data
- Analysis
- Figures
- Tables
- Paper
 - Draft.tex
 - References.bib
- Presentation
- Materials
- README

Project Directory Structure

- Data
- Analysis
- Figures
- Tables
- Paper
- Presentation
 - Slides.tex
- Materials
- README

Project Directory Structure

- Data
- Analysis
- Figures
- Tables
- Paper
- Presentation
- Materials
 - Protocol.tex
 - StimulusMaterials.pdf
 - Questionnaire.txt
- README

Cataloging Information



DATA & ANALYSIS

Comments (0)







Versions

i Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.





☐ Select all files
 [Download Selected Files](#)
Total Number of Files: **10**Total Downloads: **48**Downloads of Files in This Version: **44****Codebook**

- | | | |
|---|--|------------------|
| <input type="checkbox"/> Study 1 (Lab) Key.docx
application/octet-stream - 21 KB - 7 downloads |  Download | Study 1 data key |
| <input type="checkbox"/> Study 2 (Exit Poll) Key.docx
application/octet-stream - 16 KB - 3 downloads |  Download | Study 2 data key |



Data

- | | | |
|---|--|--|
| <input type="checkbox"/> Study 1 (Lab) Data.tab
Tab Delimited - 54 KB - 18 downloads + analyses |  Download as... | Study 1 Data |
| TABULAR DATA 647 Cases 25 Variables |  Access Analysis + Subsetting |  View Data Citation [+] |
| <input type="checkbox"/> Study 2 (Exit Poll) Data.tab
Tab Delimited - 42 KB - 4 downloads + analyses |  Download as... | Study 2 Data |
| TABULAR DATA 765 Cases 22 Variables |  Access Analysis + Subsetting |  View Data Citation [+] |

Experimental Materials

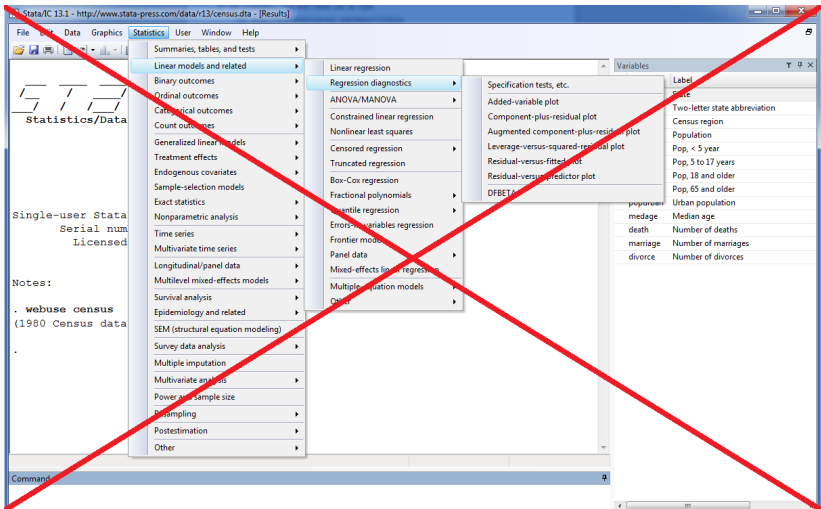
- | | | |
|--|--|-------------------------------|
| <input type="checkbox"/> Con Pretreatment Articles.doc
MS Word - 28 KB - 1 download |  Download | Con Pretreatment Articles |
| <input type="checkbox"/> Control Pretreatment Articles.doc
MS Word - 23 KB - 1 download |  Download | Control Pretreatment Articles |
| <input type="checkbox"/> Pro Pretreatment Articles.doc
MS Word - 28 KB - 2 downloads |  Download | Pro Pretreatment Articles |
| <input type="checkbox"/> Questionnaire.doc
MS Word - 97 KB - 4 downloads |  Download | Questionnaire |

Replication code

- | | | |
|---|--|--------------------------|
| <input type="checkbox"/> Study 1 (Lab).do
Stata Syntax - 5 KB - 3 downloads |  Download | Study 1 Replication code |
| <input type="checkbox"/> Study 2 (Exit Poll).do
Stata Syntax - 2 KB - 1 download |  Download | Study 2 Replication code |

(2) Get Organized

- 1 Use a folder structure than can be shared
- 2 Never use absolute file paths in code



(3) Abandon Point-and-Click

- 1 Don't clean data by hand
- 2 Use scripts rather than menus for graphics
- 3 Record your OS and software (and their versions)

(4) Publicly Archive Your Research

- 1 Use persistent, public archives, not your website or “on request”

Where do you archive your research?

- Dataverse Network
- Data Dryad
- figshare

(4) Publicly Archive Your Research

- 1 Use persistent, public archives, not your website or “on request”
- 2 Use Simple, Structured, and Semantic open file formats

File	Edit	Format	View	Help		
1	25	11331112144243343311	111121221112111362	12141234234341434313	1 1 13524	3422332322
2	23	313211212444424431	11111122122112213613	1351414224444214414	1 1 13125	4243232221
3	20	111211211321111111	11 1111212121214313314	14511144245433333341	2 1 1123	1141131311
4	20	11112121223123321	21 11221112121221222	12522234332442333411	1 1 3335	3133331232
5	20	422211112121111112	11111121111111234114	144111121141112221	1 1 32234	1111131111
6	19	22211111212212212	11 2111112111113324411	13111131344244423421	2 2 11412	1312121313
7	19	22121111222112111	11 3 11213 1221112422	1352212224244223244	4 4 1123	2412222211
8	20	31232212442244432	1121112121212123412	1443244434444333444	1 3 112331	2322132422
9	21	42222111221222221	21211111212211212342	1351124224443213323	2 1 13385	2333232223
10	21	11222121233223331	22 121212212311231132	1351234324443344443	1 3 3185	4242222343
11	21	31211112323233332	3 3 11123 1111331232	2 4112222413343441	2 1 11131	3332222432
12	25	2221112132223232	12 2222221222212224614	14511444124441424441	1 1 1183	2222254423
13	21	4213222242244311	21112122112332432462	2 52224424442223244	2 2 13335	3432333233
14	18	22212111422144313211	111111121211442222	2 312232143334433423	1 1 11234	2222222311
15	17	42121121222233221	11211111212123233122	12222424442243444	1 2 23155	3232232323
16	20	42221111212314221	111212121211133452	124122242444424444	1 1 13183	2443342434
17	22	22121111321223321	22 3 11112123112222313125	11124224443233333	3 1 11232	3443222434
18	22	31122111224222441	1111222121111421242	2 22122124134242431	2 2 23135	4132332223
19	18	22331121441244413321	11212111131215246121352	1244243442224442	4 4 1135	4144441344
20	17	42121122323224431	4 21221111221133233131354	144414444244444	2 2 13133	3422231133
21	18	11121212321142311	11 111121111321332342	1152122424344232441	1 2 3145	3222231223
22	19	31111112234244231	12211111211212542262	1213224424434333444	2 2 21145	3132343222
23	22	1121122224442422	12 21111121122142235122	53244444442444443	1 4 1183	3223212343
24	24	22122111342332411	11 1111222121221231111212224	24344122242434 1 1 1133	4414244222	4414244222
26	23	42121112121221411	111112212111243431313441	1444144431434341	1 1 33155	4414241424
28	20	22111121321333311	12 2122222131134142122	511244144441422331	3 1 3354	3223121122
29	24	31211111121341232	212111211212131125142	51111111444234242	3 2 2114231	2111121411
30	18	11111221444244422	11 21122212122244113122	521443244432234213	1 1 1335	4134412242
31	20	11121111111141211	11 1122212111111123132	414152134433422421	1 3 113234	2222132321
33	20	2211222322243221	22 21121221131132142132	511443144443244443	3 1 1143	2133221242

Conclusion

[illegible]

1996 Codebook.pdf - Adobe Reader

File Edit View Window Help

Open [Icons] 23 / 38 101% [Icons]

Tools Fill & Sign Comment

you describe your opinion of (INSERT ITEM: ROTATE ITEMS a.-i. AND g.-i.) as very favorable, mostly favorable, mostly UNfavorable, or very unfavorable? (INTERVIEWERS: PROBE TO DISTINGUISH BETWEEN "NEVER HEARD OF" AND "CAN'T RATE")

		Very Favorable	Mostly Favorable	Mostly Unfavorable	Very Unfavorable	Never Heard of	Can't Rate
(115)	a. Network television news (1-96)	1	2	3	4	5	6
(116)	b. Local TV news (1-96)	1	2	3	4	5	5
(117)	c. The daily newspaper you are most familiar with (1-96)	1	2	3	4	5	6
(118)	d. Congress (1-96)	1	2	3	4	5	6
(119)	e. Tobacco companies (7-94)	1	2	3	4	5	6
(120)	f. Labor unions (2-96)	1	2	3	4	5	6
(121)	g. Bill Clinton (2-96)	1	2	3	4	5	6
(122)	h. Hillary Clinton (2-96)	1	2	3	4	5	5
(123)	i. Bob Dole (2-96)	1	2	3	4	5	6

(4) Publicly Archive Your Research

- 1 Use persistent, public archives, not your website or “on request”
- 2 Use Simple, Structured, and Semantic open file formats
- 3 Be explicit about data licensing

How to license data?



Attribution
CC BY



Attribution-NonCommercial
CC BY-NC



Attribution-NonCommercial-ShareAlike
CC BY-NC-SA



Attribution-ShareAlike
CC BY-SA



Attribution-NonCommercial
CC BY-NC



Attribution-NonCommercial-NoDerivs
CC BY-NC-ND

(4) Publicly Archive Your Research

- 1 Use persistent, public archives, not your website or “on request”
- 2 Use Simple, Structured, and Semantic open file formats
- 3 Be explicit about data licensing
- 4 Create useful metadata

(5) Learn Literate Programming



Learn
to knit
after
lunch!

Where to go next?

- rOpenSci
- “Challenges in Irreproducible Research”
- Karl Broman’s resources
- 2011 “Reproducible Research” conference slides
- “Six steps to a Better Relationship with Your Future Self.”
- “Ten Simple Rules for Reproducible Computational Research.”
- *Reproducible Research with R and RStudio.*
- Software Carpentry
- Johns Hopkins Data Science Certificate on Coursera

People to follow?

- @victoriastodden
- @carlystrasser
- @I_peer
- @OSFramework and @BrianNosek
- @RetractionWatch
- @UCBITSS
- @OpenScience

Reproducibility isn't everything

- Data archiving and data citation
- Open protocols and materials
- Methodological transparency
- Free and open-source software (FOSS)
- Open access

1 Why?

2 What?

3 How?

In the end...

- Be reproducible *for you*
- Science will benefit as a result

