

# Accelerating Computational Reproducibility with the Social Science Reproduction Platform

Conference on Teaching & Research in Economic Education (CTREE),  
Chicago

---

Fernando Hoces de la Guardia, BITSS

June 2022 | [slides](#)

The Berkeley Initiative for Transparency in the Social Sciences works to improve the credibility of science by advancing transparency, reproducibility, rigor, and ethics in research and policy analysis.

## Core ACRe

Aleks Bogdanoski

Albert Chae

Fernando Hoces de la Guardia

Edward Miguel

Lars Vilhuber



BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

We are part of the Center for Effective Global Action (CEGA).



## Many Others

CEGA staff

Undergrad and Graduate RAs

Catalysts

Outside Collaborators  
(Researchers, Programmers)

# M2: More Inclusive Concept Scholarly Output

Clarebout Principle:

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

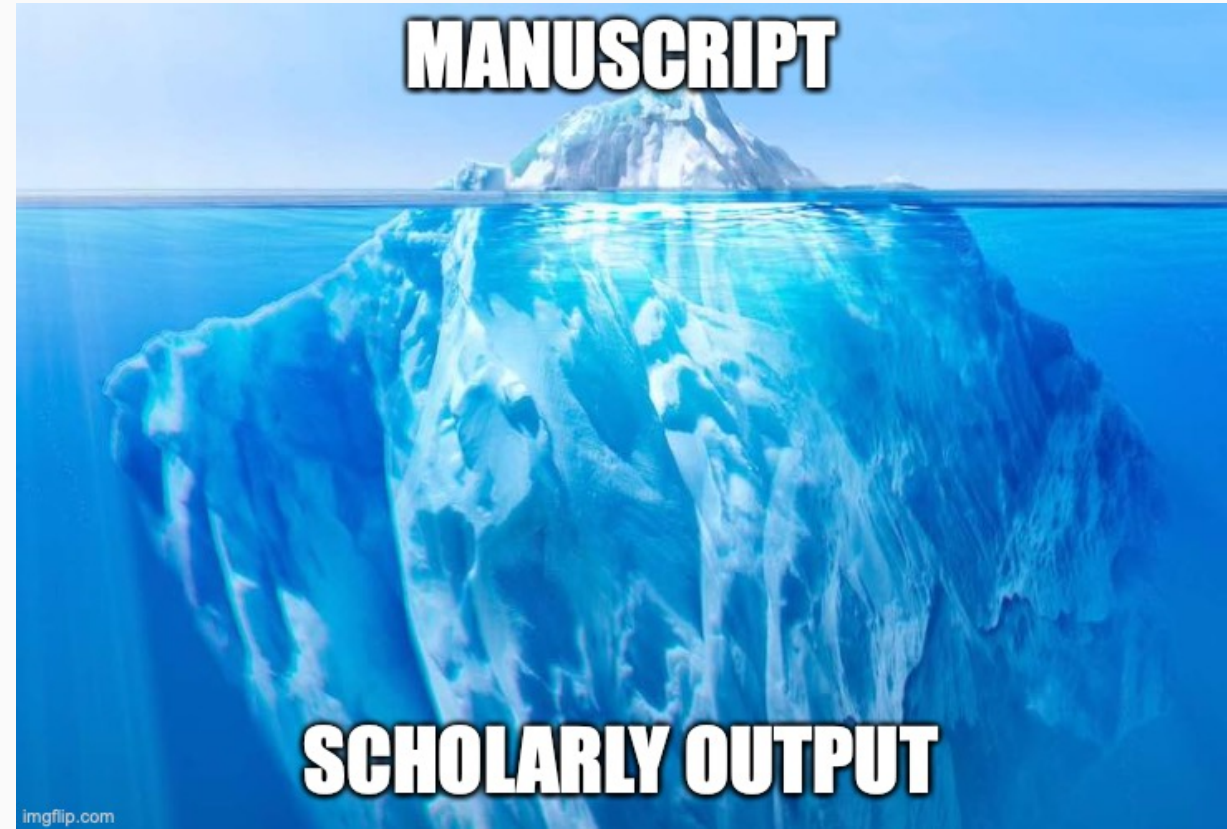
*Buckheit and D.L. Donoho (1995, 2009)*

# M2: More Inclusive Concept Scholarly Output

Clarebout Principle:

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

*Buckheit and D.L. Donoho (1995, 2009)*



# M2: More Inclusive Concept Scholarly Output

Potential benefits of following the Clarebout Principle

Well discussed potential positive effects on:

- Pedagogy
- Incremental generation of knowledge

Under discussed:

- Possible positive effect on diversity, equity and inclusion: no connections or language skills ("appropriate politeness") required to obtain materials

# M3: Prevent Loss of Knowledge

Every semester, graduate students around **the world** take an Empirical/Applied [ ... ] course (e.g., Labor Economics, Social Psychology). A typical assignment consists of reproducing the results of a paper and, possibly, testing the robustness of its results.

| Stage                        | New Knowledge   |
|------------------------------|---|
| Scope<br>(select and verify) | Data and code exist?  |
| Assess                       | Degree of reproducibility for<br>specific part of the paper |
| Improve                      | E.g. fixed paths, libraries,<br>added missing files, etc.   |
| Test robustness              | Results are robust to additional<br>specifications          |

# M3: Prevent Loss of Knowledge

- In each of the previous steps there was valuable new knowledge
- This knowledge is not usually disseminated
  - End of semester report, or presentation that is not published.
  - Language across courses is not standard (reproduction? replication?)
  - Current attempts to record this are closer to a full paper format (see [Replication Wiki](#))
- As a result there is a large amount of missed knowledge.
  - Each new semester, new students repeat similar excesses and have to reinvent the wheel.
  - Lost opportunity for building on top of previous exercises.
  - Lost opportunity for aggregating the new knowledge.

# Context for ACRE

- American Economics Association (AEA) creates first data policy in 2006.
  - Must publish some data (waivers available)
- AEA updates policy in 2019.
  - Must post all data and code. Publication is conditional on verifying reproducibility (if confidential: must document extensively)
  - A new requirement is to post all cleaning code, even for data that is not public
- We should expect high levels of computational reproducibility after 2019 (AEA).
- We should not demand 100% reproducibility before, but we could identify the gaps and try to improve some.



# Accelerating Computational Reproducibility: Framework

## Beyond Binary Judgments

Reproductions can easily gravitate towards adversarial exchanges.

- Early career researchers (ECR) have incentives to emphasize unsuccessful reproductions
- Original authors have a more senior position and can use it to deter in-depth reproductions from ECRs.
- The media also focuses on eye-catching headlines

## Our approach:

We do not want to say

"Paper X is (ir)reproducible"

We do want to say

"Result Y in paper X has a high/low **level** of reproducibility according to **several** reproduction attempts. Moreover, **improvements** have been made to the original reproduction package, **increasing** its reproducibility to a higher level"

# ACRe Framework

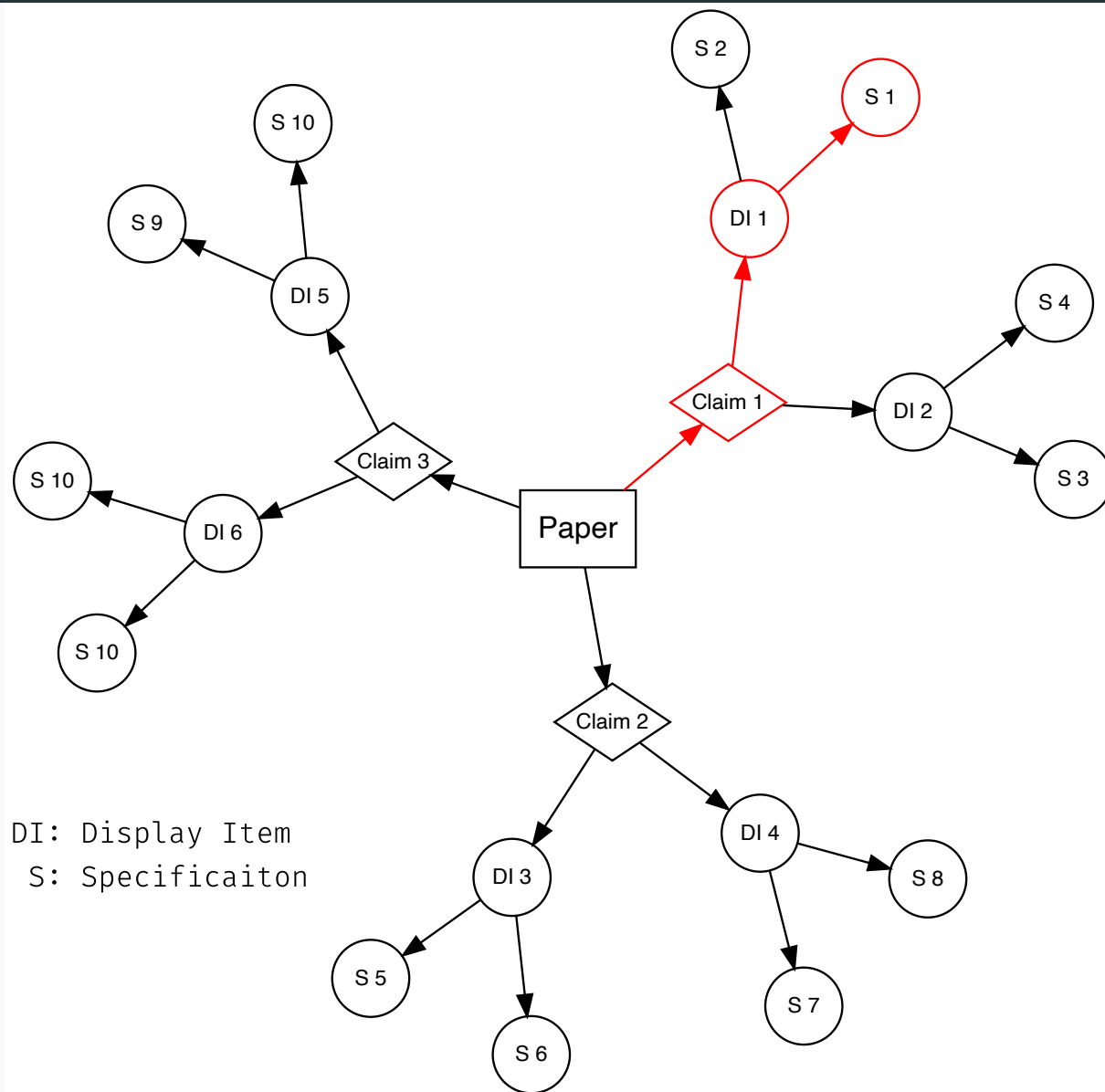
Each **reproduction attempt** is centered around scientific **claims** (following **SCORE**).

One paper can contain several claims.

Each claim may be supported by various **display items**: tables, figures & inline results.

A reproduction attempt is at the claim level, and reproducers must record their **specifications** of interest.

Key challenge: **standardization** of concepts and formats.



# Large part of this exercise is about standardization

- **Computational Reproduction (or Reproduction)**
- **Replication (will not mention this term again!)**
- Reproduction attempt
- **Reproduction package**
- Claim
- Display item
- Specification
- Preferred specification
- **Raw data**
- **Analysis data**
- **Candidate paper**
- **Declared paper**
- Reproduction tree
- Complete Workflow
- Computationally Reproducible from Analytic data (CRA)
- Computationally Reproducible from Raw data (CRR)
- Reasonable test
- Feasible test
- Minimal effort

# Demo: socialsciencereproduction.org

[Home](#)[About](#)[Events](#)[Guide](#)[Forum](#)[Metrics](#)

Welcome to the

## Social Science Reproduction Platform

Accelerating computational  
reproducibility in the social sciences

GET STARTED!

Search reproductions by paper title, DOI, or author name



# Select a paper

## Article Information

### Abstract

How large are the benefits of transportation infrastructure projects, and what explains these benefits? This paper uses archival data from colonial India to investigate the impact of India's vast railroad network. Guided by four results from a general equilibrium trade model, I find that railroads: (1) decreased trade costs and interregional price gaps; (2) increased interregional and international trade; (3) increased real income levels; and (4) that a sufficient statistic for the effect of railroads on welfare in the model accounts well for the observed reduced-form impact of railroads on real income in the data.

### Citation

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*. 108 (4-5): 899-934.

DOI: 10.1257/aer.20101199

Choose Format:

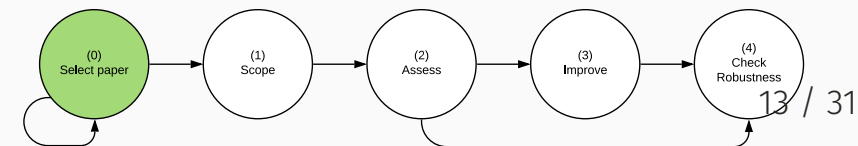


### Additional Materials

Data Set (17.80 MB)

Online Appendix (247.97 KB)

Author Disclosure Statement(s) (63.46 KB)



C. Results

Table 2 presents ordinary least squares (OLS) estimates of equation (12). In column 1 I estimate the effect of the lowest-cost route effective distance on trade when the relative costs of each mode ( $\alpha$ ) are set to observed historical relative freight rate estimates. I use the relative per unit distance freight rates described in Section IB (at their midpoints):  $\alpha^{road} = 4.5$ ,  $\alpha^{river} = 3.0$ , and  $\alpha^{coast} = 2.25$  relative to the freight rate of railroad transport, normalized to 1). Column 1 demonstrates that the elasticity of trade costs with respect to the lowest-cost route effective distance, calculated at observed freight rates, is 0.088, and this is statistically significant at the 5 percent level.

However, as argued in Section IB, it is possible that these observed relative freight rates do not capture the full benefits (such as increased certainty or savings) of railroad transport relative to alternative modes of transportation. For this reason the NLS specification in column 2 estimates the relative freight elasticity (i.e., the parameters  $\alpha$ ) that minimize the sum of squared residuals in equation (12).

Column 2 is my preferred specification. When the mode-wise distance costs (i.e.,  $\alpha$ ) are not restricted to be equal to the observed freight rates, the estimated elasticity of trade costs with respect to effective distance (i.e.,  $\delta$ ) rises to 0.169. Even when controlling for all unobserved, time-constant determinants of trade costs between all salt sources and destinations, as well as unrestricted shocks to the source price of each salt type, reductions in trade costs along lowest-cost routes (estimated from railroad-driven time variation in these routes alone) have a large effect on reducing salt price gaps over space.

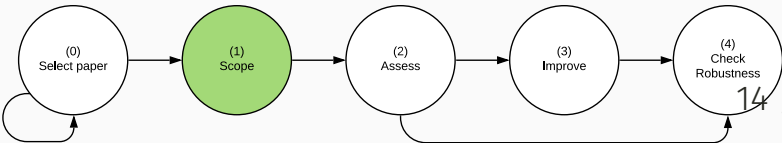
TABLE 2—RAILROADS AND TRADE COSTS: STEP 1

| Dependent variable: log salt price at destination  | (1)              | (2)                      |
|--|------------------|--------------------------|
| log effective distance to source, along lowest-cost route<br>(at historical freight rates) | 0.088<br>(0.028) |                          |
| log effective distance to source, along lowest-cost route<br>(at estimated mode costs)     |                  | 0.169<br>[0.062, 0.296]  |
| Estimated mode costs per unit distance:<br>Railroad (normalized to 1)                      |                  | 1<br>N/A                 |
| Road   |                  | 2.375<br>[1.750, 10.000] |
| River  |                  | 2.250<br>[1.500, 6.250]  |
| Coast  |                  | 6.188<br>[5.875, 10.000] |
| Observations   | 7,345            | 7,345                    |
| R <sup>2</sup>   | 0.946            | 0.946                    |

Notes: Regressions estimating equation (12) using data on 6 types of salt (listed in online Appendix A), from 133 districts in Northern India, annually from 1861 to 1930. Column 1 and column 2 estimated by OLS and NLS respectively; both include salt type  $\times$  year and salt type  $\times$  destination fixed effects. “Effective distance to source.”

Col 1: OLS with alpha at historical rates. 1% inc in effc dist (in km) -> 0.088% inc in trade costs

Col 2: NLS with alpha chosen to minimize SSR of col1. effc dist (in km) -> 0.169% inc in trade costs



# Assess

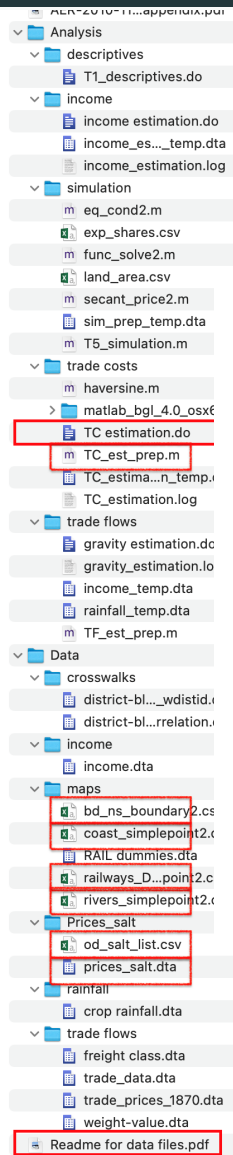
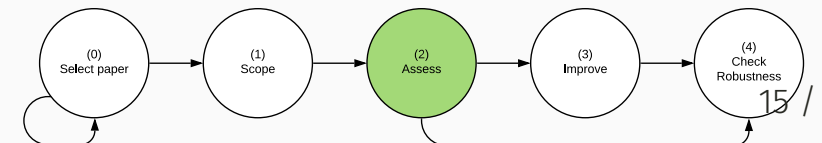


Table 2

|  |  |
|--|--|
| TC estimation.do                       |  |
| LCRED_RoX_CoY_RiZ.csv                  |  |
| TC_est_prep.m                          |  |
| railways_Dissolve_Simplify2_point2.csv |  |
| bd_ns_boundary2.csv                    |  |
| rivers_simplepoint2.csv                |  |
| rivers_simplepoint2.csv                |  |
| coast_simplepoint2.csv                 |  |
| od_salt_list.csv                       |  |
| prices_salt.dta                        |  |
| railways_Dissolve_Simplify2_point2.csv |  |
| bd_ns_boundary2.csv                    |  |
| rivers_simplepoint2.csv                |  |
| coast_simplepoint2.csv                 |  |
| od_salt_list.csv                       |  |



# Assess

Analysis

descriptives

T1\_descriptives.do

income

income estimation.do

income\_es...\_temp.dta

income\_estimation.log

simulation

eq\_cond2.m

exp\_shares.csv

func\_solve2.m

land\_area.csv

secant\_price2.m

sim\_prep\_temp.dta

T5\_simulation.m

trade costs

haversine.m

matlab\_bg1\_4.0\_osxf

TC estimation.do

TC\_est\_prep.m

TC\_estima...n\_temp.dta

TC\_estimation.log

trade flows

gravity estimation.do

gravity\_estimation.lo

income\_temp.dta

rainfall\_temp.dta

TF\_est\_prep.m

Data

crosswalks

district-bl...wdistid.dta

district-bl...relation.dta

income

income.dta

maps

bd\_ns\_boundary2.csv

coast\_simplepoint2.csv

RAIL\_dummies.dta

railways\_D...point2.csv

rivers\_simplepoint2.csv

Prices\_salt

od\_salt\_list.csv

prices\_salt.dta

rainfall

crop rainfall.dta

trade flows

freight class.dta

trade\_data.dta

trade\_prices\_1870.dta

weight-value.dta

Readme for data files.pdf

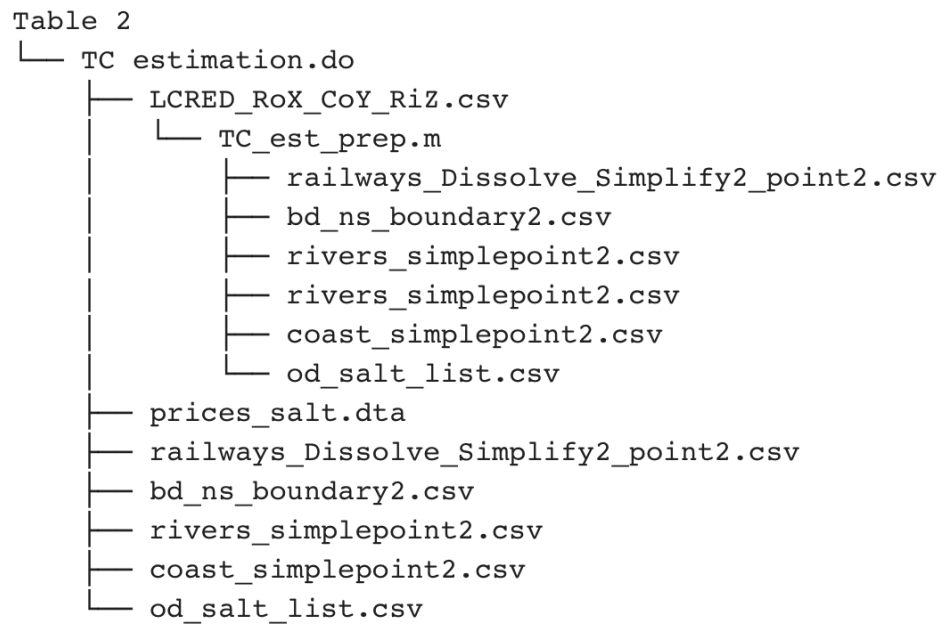


Table 3.5: Levels of Computational Reproducibility  
(P denotes “partial,” C denotes “complete”)

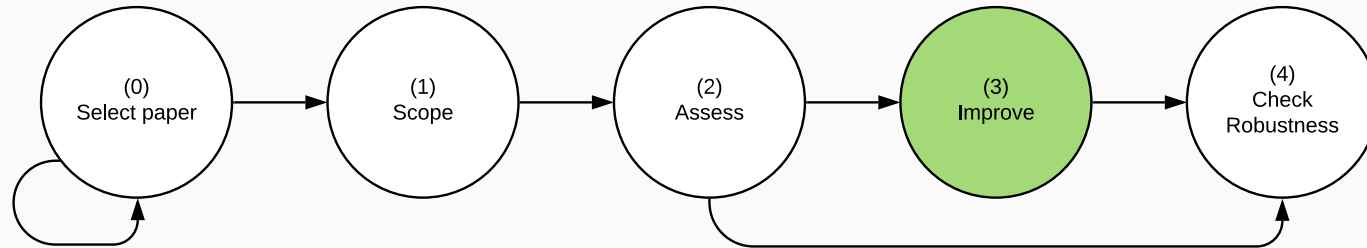
| Availability of materials, and reproducibility |               |   |               |   |     |               |   |          |   |     |
|--|---------------|---|---------------|---|-----|---------------|---|----------|---|-----|
|  | Analysis Code |   | Analysis Data |   | CRA | Cleaning Code |   | Raw Data |   | CRR |
|  | P             | C | P             | C |     | P             | C | P        | C |     |
| L1: No materials                               | –             | – | –             | – | –   | –             | – | –        | – | –   |
| L2: Only code                                  | ✓             | ✓ | –             | – | –   | –             | – | –        | – | –   |
| L3: Partial analysis data & code               | ✓             | ✓ | ✓             | – | –   | –             | – | –        | – | –   |
| L4: All analysis data & code                   | ✓             | ✓ | ✓             | ✓ | –   | –             | – | –        | – | –   |
| L5: Reproducible from analysis                 | ✓             | ✓ | ✓             | ✓ | ✓   | –             | – | –        | – | –   |
| L6: All cleaning code                          | ✓             | ✓ | ✓             | ✓ | –   | ✓             | ✓ | –        | – | –   |
| L7: Some raw data                              | ✓             | ✓ | ✓             | ✓ | –   | ✓             | ✓ | ✓        | – | –   |
| L8: All raw data                               | ✓             | ✓ | ✓             | ✓ | –   | ✓             | ✓ | ✓        | ✓ | –   |
| L9: All raw data + CRA                         | ✓             | ✓ | ✓             | ✓ | ✓   | ✓             | ✓ | ✓        | ✓ | –   |
| L10: Reproducible from raw data                | ✓             | ✓ | ✓             | ✓ | ✓   | ✓             | ✓ | ✓        | ✓ | ✓   |

<sup>a</sup> **Computationally Reproducible from Analytic data (CRA):** The output can be reproduced with minimal effort starting from the *analytic* datasets.

<sup>b</sup> **Computationally Reproducible from Raw data (CRR):** The output can be reproduced with minimal effort from the *raw* datasets.



# Improvements



## Three types of improvements:

1. Improvements at the paper level
2. Improvements at the display-item level
3. Specific future improvements

# Improvements: Paper-level

- Use version control software (Git/Github).
- Improve documentation: comments, indentations, object names, etc.
- Re-organize the reproduction package into a set of folders and sub-folders that follow **standardized best practices**, and add a master script that executes all the code in order, with no further modifications. **See AEA's reproduction template.**
- Literate programming environment (e.g., Jupyter notebooks, RMarkdown)
- Re-write code using a different statistical software (ideally open source, like R, Python, or Julia).
- Set up a computing capsule (e.g., **Binder** and **Code Ocean**).



# Improvements: Display item-level

- Adding missing raw data: files or meta-data
  - Example: "Add raw temperature and relative humidity data"
- Adding missing analytic data files
  - Example: "Copy the row files from Data folder into new `Analysis\trade cost\Input`"
- Adding missing analysis or cleaning code
  - Example: "Replaced broken Wald bootstrap code with updated code/command"
- Debugging code
  - Example: "was counting each group 4 times in round 1, so fixed that"



# Improvements: future possible

We ask reproducer to leave concise and actionable tasks for other reproducers in the future.

Example 1:

"Revise the .aml and .bat code scripts to reflect reorganized structure"

Example 2:

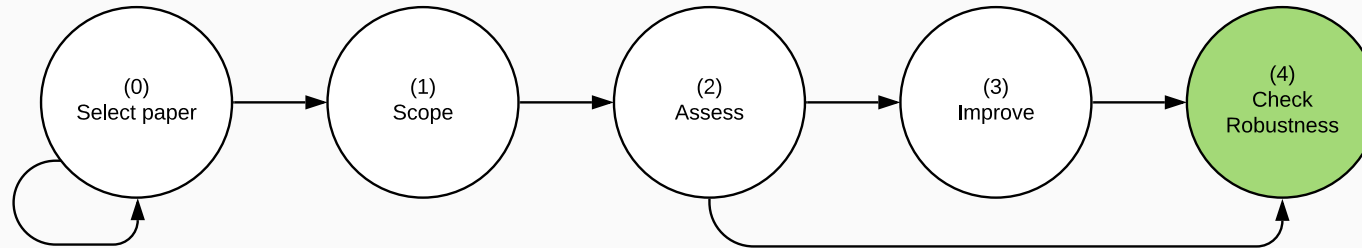
"Provide data and codes generating the other two figures in the paper, which are not given in the replication file."

Example 3:

"Table 3 can be reproduced identically from the [...] analytic data files. I was not able to reproduce the analytic data files due to lack of access to ArcGIS software, but the code scripts and raw data files [...] are included in the reproduction package."



# Robustness Checks



## Two main parts for robustness:

1. Increase the number of robustness checks
2. Justify the appropriateness of a specific test

# Robustness

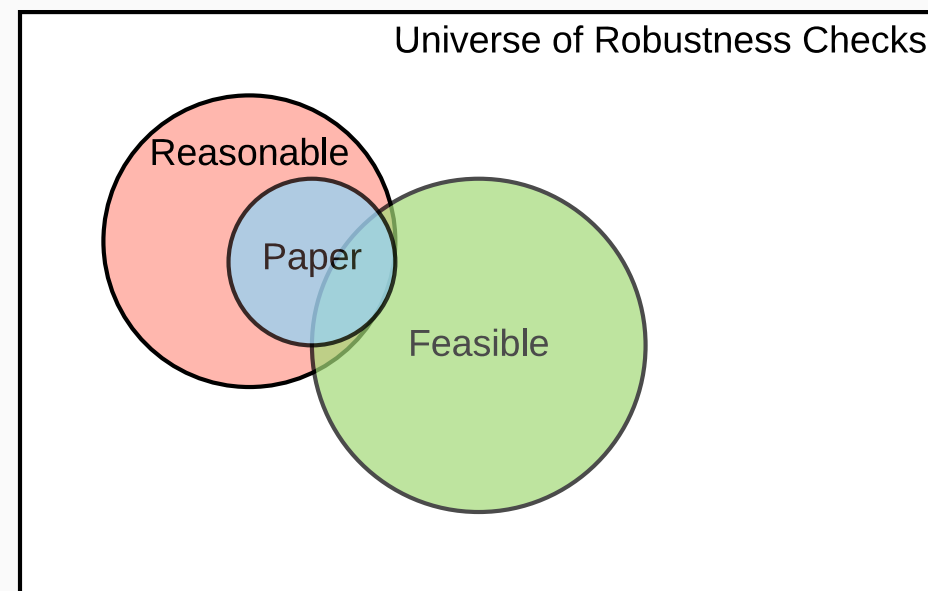
**Robustness checks:** any possible change in a computational choice, both in data analysis and data cleaning

**Reasonable specifications** (Simonsohn et. al., 2018):

1. Sensible tests of the research question
2. Expected to be statistically valid, and
3. Not redundant with other specifications in the set.

Reproducers will be able to record two types of contributions:

- Mapping the universe of robustness checks
- Proposing a specific robustness check



# Completed Reproduction: Example

**Reproduction of: Railroads of the Raj: Estimating the Value of Transportation Infrastructure**  
American Economic Review  
<https://doi.org/10.1257/aer.20101199>

Reproducer: Emma Ng

Date created: November 19 2020

Date submitted: August 03 2021

Claims assessed: 1

Display Items (DI) assessed: 1

Improvements: 3

Robustness tests:

- Feasible choices added: 0
- Specifications justified: 0

Claims identified by reproducer:

1. "The paper estimated the value of the extent to which the costs of different modes of transportation are estimated using the reduction of LCRED (lowest-cost route effective distance) trading cost. This estimate has a 95% confidence interval."

Reproducibility of Display Items:

1. Table 2 - Table 2 -- Railroads and Trade Costs: Step 1. ...

Original reproduction package: <https://www.aeaweb.org/articles?id=10.1257/aer.20101199>

Revised reproduction package: <https://github.com/em-ng21/railroads-of-the-raj>

Original authors' availability for further inquiries: Not sure

## 0 Select a paper

Declare the paper that you will analyze in the remainder of the exercise and other "candidate" papers for which you were unable to obtain access to

[VIEW THIS SECTION](#)

## 1 Scoping

## 2 Assessment

## 3 Improvement

## 4 Robustness

## 0 Select a paper

Declare the paper that you will analyze in the remainder of the exercise and other "candidate" papers for which you were unable to obtain access to

# Completed Reproduction: Example

Reproduction of: Railroads of the Raj: Estimating the Impact of Transportation Infrastructure on American Economic Growth  
<https://doi.org/10.1257/aer.20101199>

Reproducer: Emma Ng

Date created: November 19 2020

Date submitted: August 03 2021

Claims assessed: 1

Display Items (DI) assessed: 1

Improvements: 3

Robustness tests:

- Feasible choices added: 0
- Specifications justified: 0

Claims identified by reproducer:

1. "The paper estimated the value of the extent to which the costs of different modes of transportation are estimated to reduce the reduction of LCRED (lowest-cost route effective distance trading cost. This estimate has a 95% confidence interval."

Reproducibility of Display Items:

1. Table 2 - Table 2 -- Railroads and Trade Costs: Step 1. ...

Original reproduction package: <https://www.aeaweb.org/articles?id=10.1257/aer.20101199>

Revised reproduction package: <https://github.com/em-ng21/railroads-of-the-raj>

Original authors' availability for further inquiries: Not sure

## 0 Select a paper

Declare the paper that you will analyze in the remainder of the exercise and record other "candidate" papers for which you were unable to obtain access to a reproduction package. See detailed guidance [here](#).

## 0 Select a paper

Declare the paper that you will analyze in the remainder of the exercise and record other "candidate" papers for which you were unable to obtain access to a reproduction package. See detailed guidance [here](#).

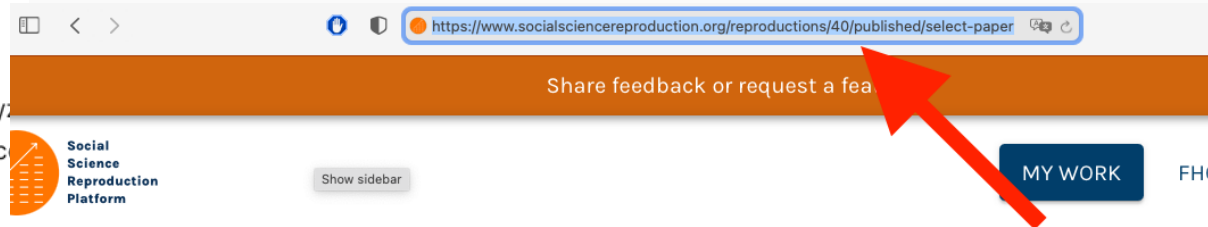
[VIEW THIS SECTION](#)

## 1 Scoping

## 2 Assessment

## 3 Improvement

## 4 Robustness



## Select a paper

Declare the paper that you will analyze in the remainder of the exercise and record other "candidate" papers for which you were unable to obtain access to a reproduction package. See detailed guidance [here](#).

[RETURN TO STAGES OVERVIEW](#)

### Basic information

- **Share it:** Original authors, instructors, other researchers, add it to your CV
- **Discuss it:** Discourse forum for reproductions
- **Cite it:** Reproduction DOI coming soon



# Promoting a Constructive Exchange

- 1 - Contacting the original author(s) when there is no reproduction package
  - 2 - Contacting the original author(s) to request specific missing items of a reproduction package
  - 3 - Asking for additional guidance when some materials have been shared
  - 4 - Response when the original author has refused to share due to *undisclosed reasons*
  - 5 - Response when the original author has refused to share due to legal or ethical restrictions of the data
  - 6 - Contacting the original author to share the results of your reproduction exercise
  - 7 - Responding to hostile responses from original authors
- Under development: sample responses from authors to reproducers

# Example 1: There is no reproduction package

**Subject:** Reproduction package for [“Title of the paper”]

Dear Dr. [Lastname of Corresponding Author],

I am contacting you to request a reproduction package for your paper titled [Title] which was published in [Reference]. A reproduction package may contain (raw and/or analytic) data, code, and other documentation that makes it possible to reproduce paper. Would you be able to share any of these items?

I am a [position] at [Institution], and I would like to reproduce the results, tables, and other figures using the reproduction materials mentioned above. I have chosen this paper because [add context ...]. **Unfortunately, I was not able to locate any of these materials on the journal website, Dataverse [or other data and code repositories], or in your website.**

**I will record the result of my reproduction attempt on ACRE [...]. With your permission, I will also record the materials you share with me, which would allow access for other reproducers and avoid repeated requests directed to you. Please let me know if there are any legal or ethical restrictions that apply to all or parts of the reproduction materials so that I can take that into consideration during this exercise.**

In addition to your response above, would you be available to respond to future (non-repetitive) inquiries from me or other reproducers conducting an ACRE exercise? **Though your cooperation with my and/or any future request would be extremely helpful, please note that you are *not required to comply*.**

Since I am required to complete this project by [date], I would appreciate your response by [deadline].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [Name (email)] for further details on this exercise. Thank you in advance for your help!

Best regards,

[Reproducer]

# Example 1: Following up on additional materials

## Template email:

**Subject:** Clarification for reproduction materials for ["Title of the paper"]

Dear Dr. [Lastname of Corresponding Author],

Thank you for sharing the materials. They have been immensely helpful for my work.

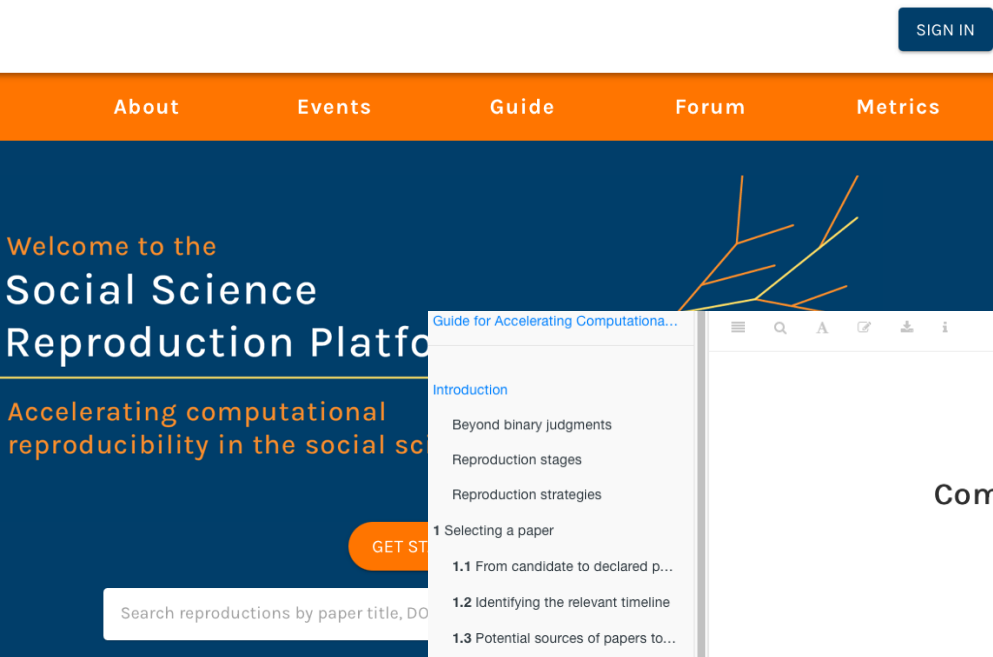
Unfortunately, I ran into a few issues as I delved into the reproduction exercise, and I think your guidance would be helpful in resolving them. **[Describe the issues and how you have tried to resolve them. Describe whatever files or parts of the data or code are missing. Refer to examples 1 and 2 below for more details]**.

Thank you in advance for your help.

Best regards,

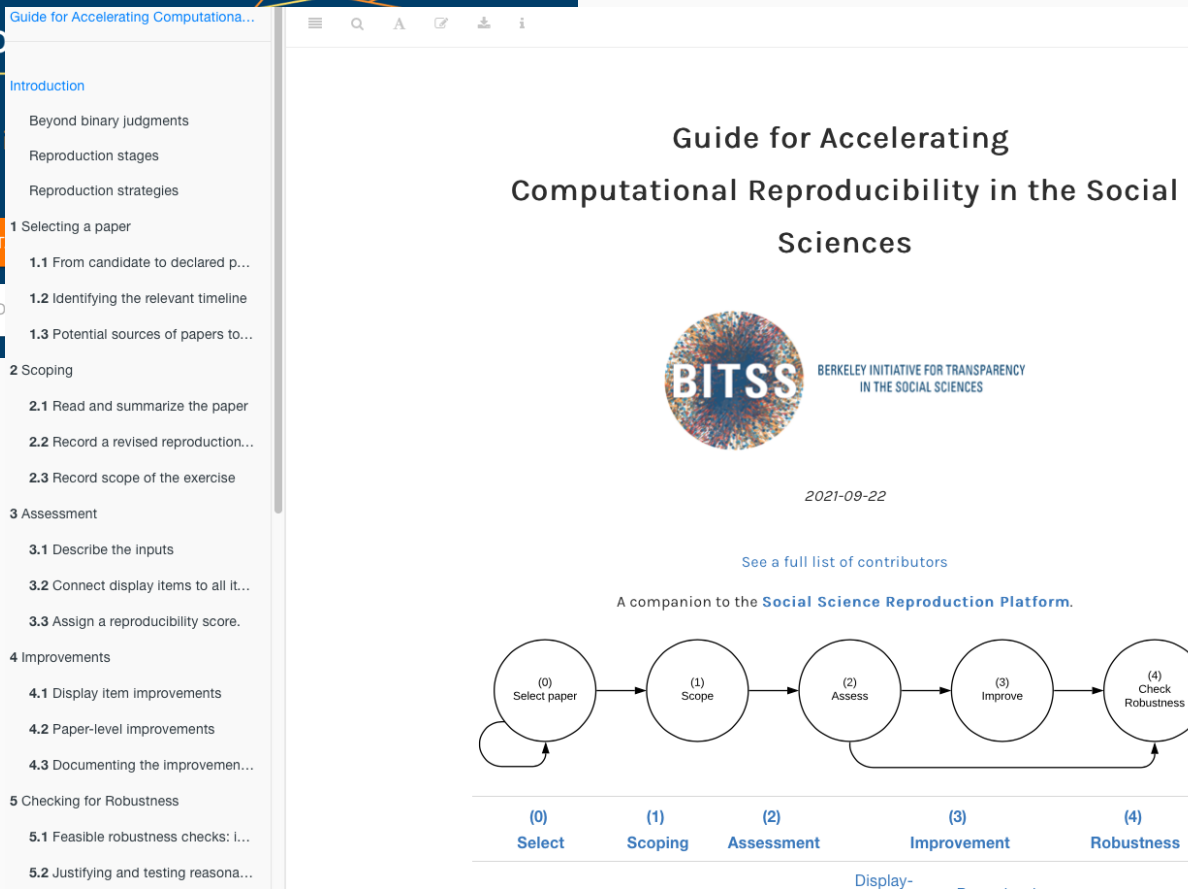
[Reproducer]

# How Do I Get Started?



Consult the  
supporting  
**guide**

**Use it** for your class  
or independent project



# How Do I Get Started?

About

Events

Guide

Forum

Metrics

WELCOME TO THE

SOCIAL SCIENCE

REPRODUCTION PLATFORM

ACCELERATING COMPUTATIONAL

REPRODUCIBILITY IN THE SOCIAL SCIENCES

GET STARTED

Search reproductions by paper title, DOI, or keyword

Guide for Accelerating Computational Reproducibility in the Social Sciences

Introduction

Beyond binary judgments

Reproduction stages

Reproduction strategies

1 Selecting a paper

1.1 From candidate to declared paper

1.2 Identifying the relevant timeline

1.3 Potential sources of papers to consider

2 Scoping

2.1 Read and summarize the paper

2.2 Record a revised reproduction

2.3 Record scope of the exercise

3 Assessment

3.1 Describe the inputs

3.2 Connect display items to all items

3.3 Assign a reproducibility score

4 Improvements

4.1 Display item improvements

4.2 Paper-level improvements

4.3 Documenting the improvements

5 Checking for Robustness

5.1 Feasible robustness checks: I...

5.2 Justifying and testing reasona...

Use it for your class  
or independent project

## Guide for Accelerating Computational Reproducibility in the Social Sciences

BERKELEY INITIATIVE FOR TRANSPARENCY  
IN THE SOCIAL SCIENCES

2021-09-22

[See a full list of contributors](#)

A companion to the [Social Science Reproduction Platform](#).

```
graph LR; A((0) Select paper) --> B((1) Scope); B --> C((2) Assess); C --> D((3) Improve); D --> E((4) Check Robustness); E --> A;
```

(0) Select (1) Scoping (2) Assessment (3) Improvement (4) Robustness

all categories ▾

all tags ▾

Categories

Latest New (3) Unread (1) Top

+ New Topic

Category

Topics

Latest

Getting Started

0

Version 1.0, developed in April 2021

Reproductions

5

1 unread

Discuss and share resources about ongoing or completed reproductions sorted by papers reproduced. To join the discussion about a particular paper, look up the paper's DOI, replacing "." with "-".

Tips for Reproducibility

1

Find and share tips and resources for reproducible workflows that aren't linked to a particular reproduction.

Help

3

3 new

Ask questions on how to use the SSRP.

For instructors

0

Connect with a community of

Welcome to Discourse

0

Nov '20

Restricted-use data

0

23h

Help

Catalog of reproductions

0

23h

Help

No replication package

1

23h

Help

Feedback for my reproduction of Twenty-year economic impacts of deworming

1

Aug 11

Reproductions

10-1073-pnas-2023185118

How to use the SSRP

2

Jul 22

Consult the  
supporting  
guide

Ask questions

29 / 31

# Who Is Using The SSRP?

## Reproducibility Dashboard

125

Reproductions



91

Papers



206

Claims



Levels of Reproducibility

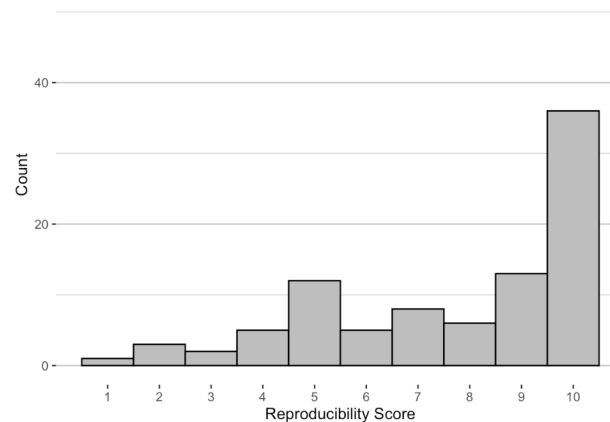
Abandoned Included

Max within paper

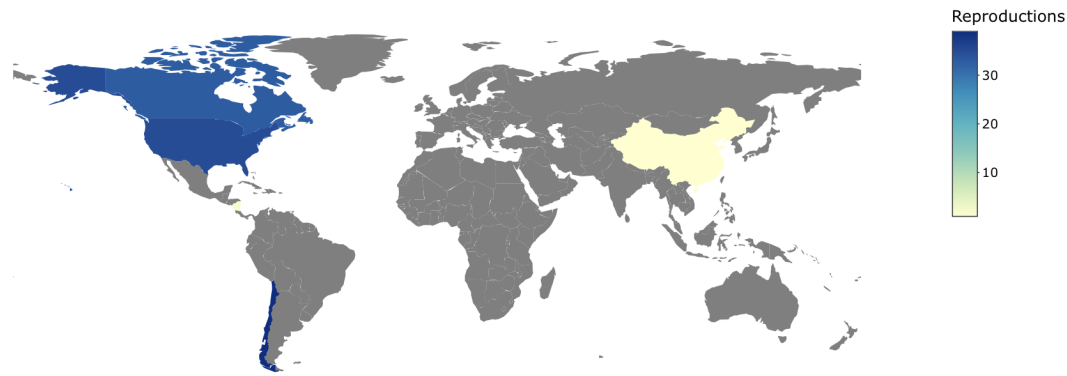
Min within paper

Summary of Levels

### Distribution of Reproduction Scores: Paper Level



### Reproducers Around the World



# Thank You

---

acre@berkeley.edu  
fhoces@berkeley.edu