The Social Science Reproduction Platform

RT2

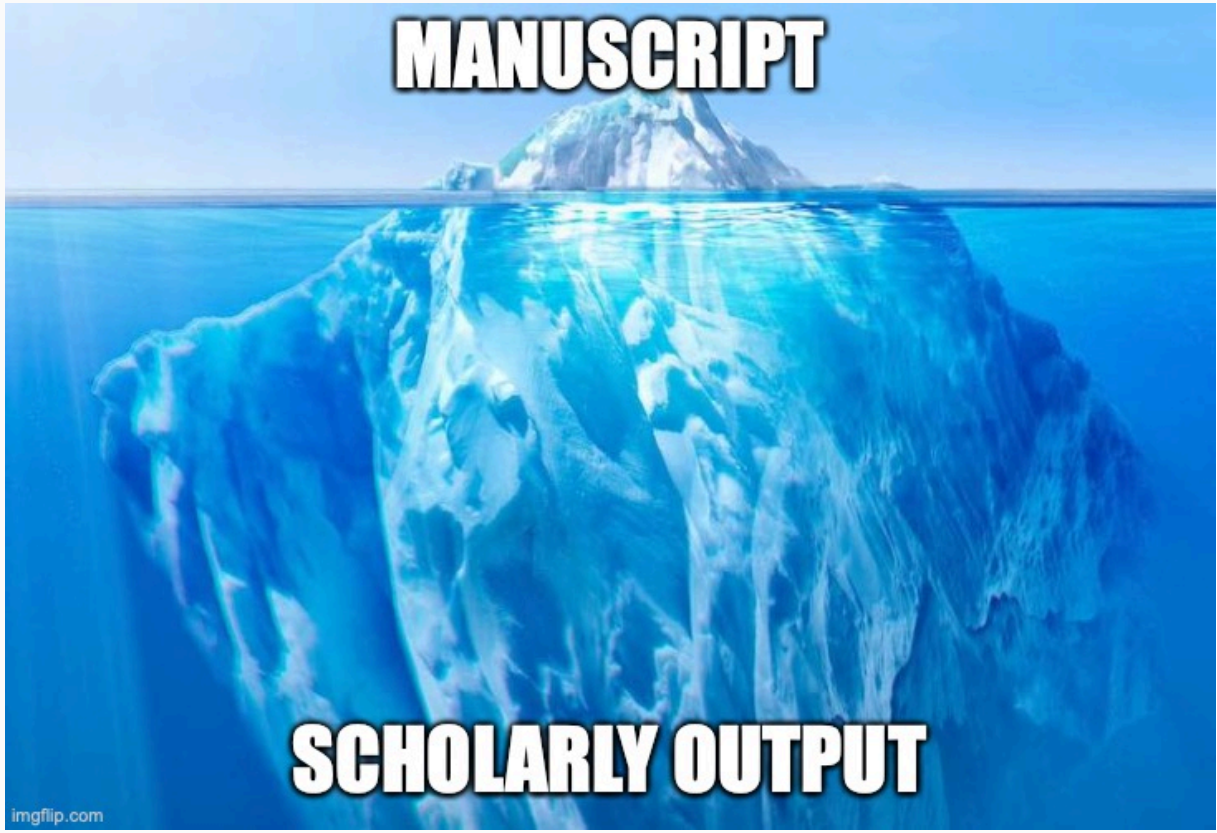
Fernando Hoces de la Guardia, BITSS
06 June 2024 Slides

Motivation I: Computational Reproducibility

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)



Motivation II: 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 (select and verify)	Data and code exist?
Assess	Degree of reproducibility for specific part of the paper
Improve	E.g. fixed paths, libraries, added missing files, etc.
Test robustness	Results are robust to additional specifications

Key challenge: Standardization

Motivation II: Prevent Loss of Knowledge

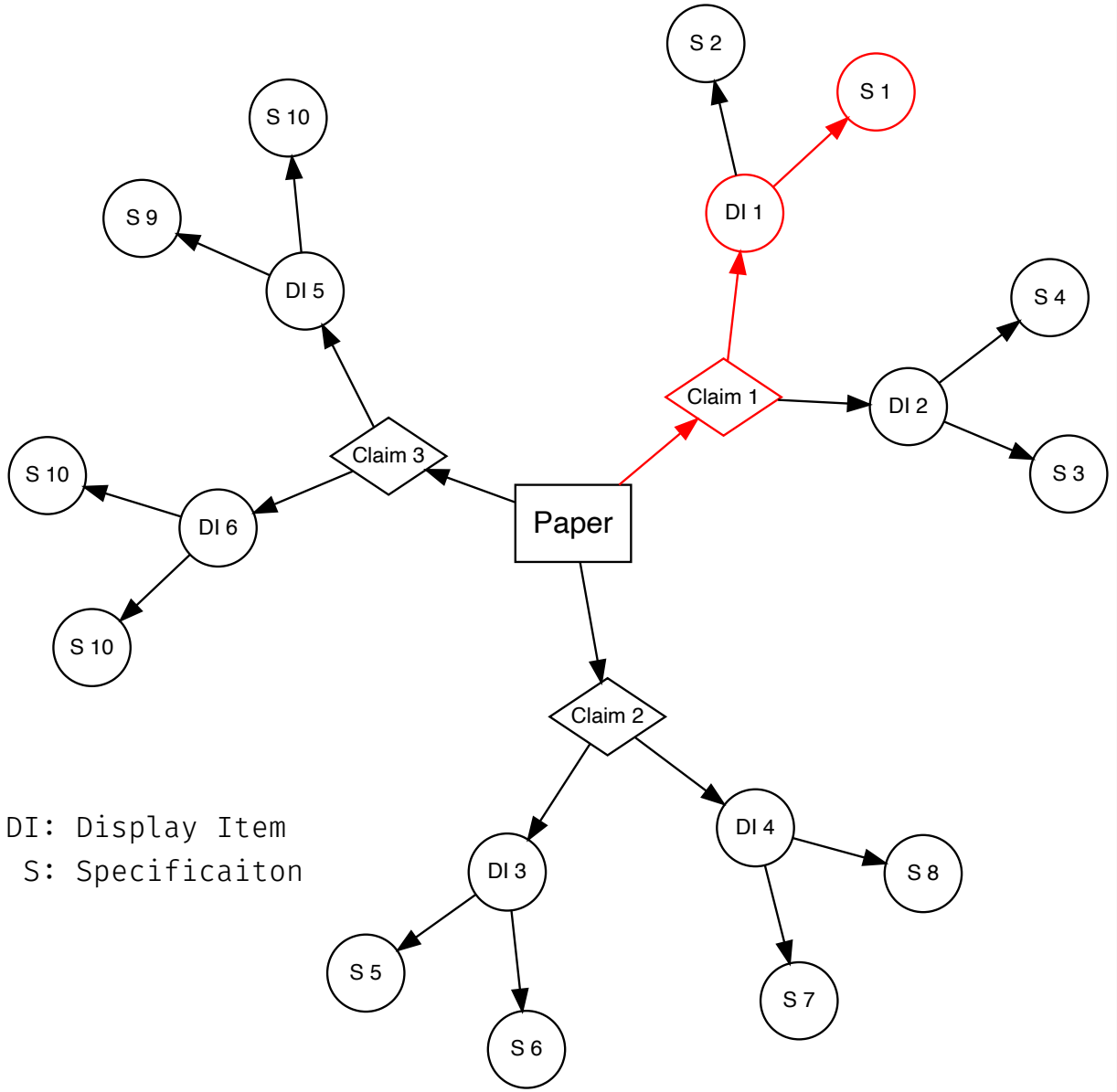
Standarized Scope of a Reproduction

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

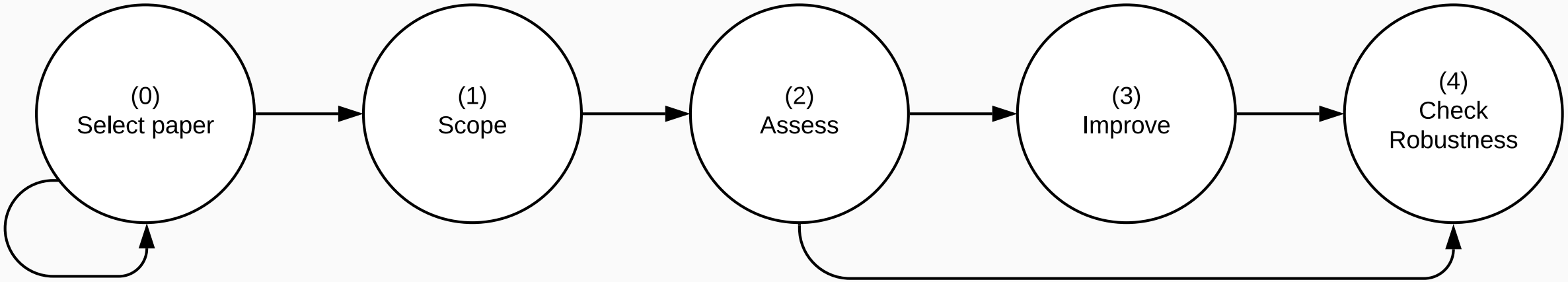
One paper can contain several claims.

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.



Stages

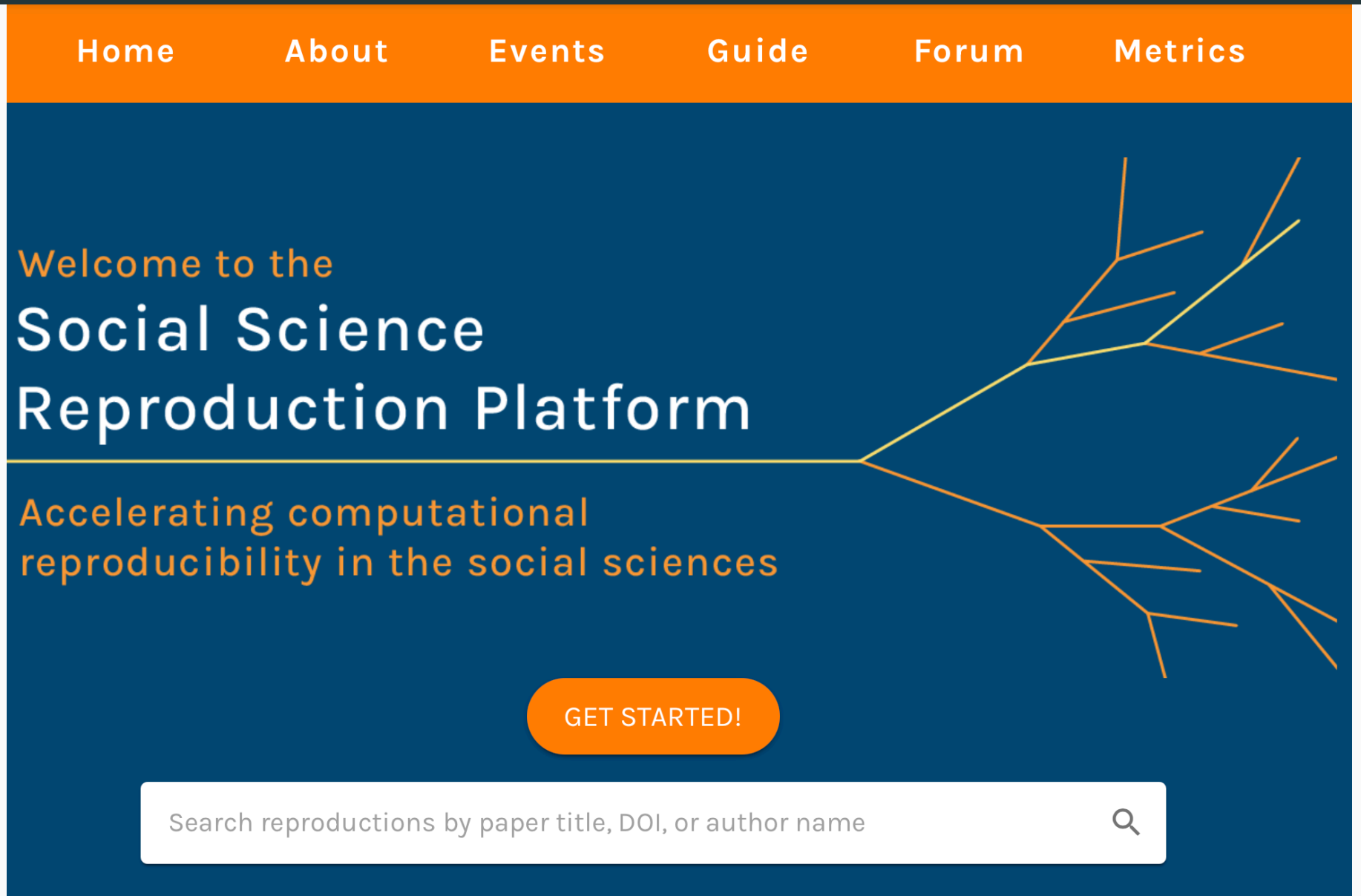


Standzarized Assessment

	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	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

[a] **Computationally Reproducible from Analytic data (CRA):** The output can be reproduced with

Demo: socialsciencereproduction.org



Scope

Go to page 19

C. *Results*

Table 2 presents ordinary least squares (OLS) estimates of equation (12). In
umn 1 I estimate the effect of the lowest-cost route effective distance on trade
when the relative costs of each mode (α) are set to observed historical rel
freight rate estimates. I use the relative per unit distance freight rates describe
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 der
strates that the elasticity of trade costs with respect to the lowest-cost route effe
distance, calculated at observed freight rates, is 0.088, and this is statistically si
icant at the 5 percent level.

However, as argued in Section IB, it is possible that these observed rel
freight rates do not capture the full benefits (such as increased certainty or
savings) of railroad transport relative to alternative modes of transportation.
this reason the NLS specification in column 2 estimates the relative freight
(i.e., the parameters α) that minimize the sum of squared residuals in equation (12).
Column 2 is my preferred specification. When the mode-wise distance costs (i.e., α) are not restricted to be equal to the observed freight rates, the estimated elasticity of trade costs with respect to effective distance (i.e., δ) 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.

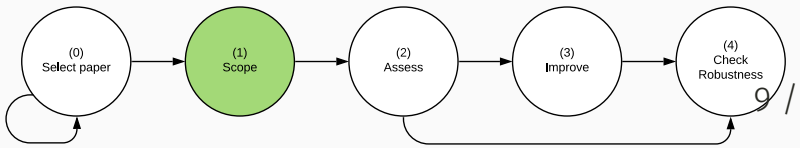
Col 1: OLS with alpha at historical rates. 1% inc in effc dist (in km) -> 0.088% inc in trade costs

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

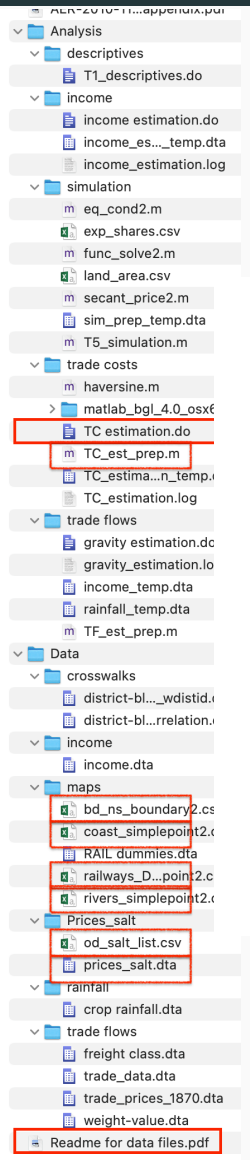
TABLE 2—RAILROADS AND TRADE COSTS: STEP 1

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

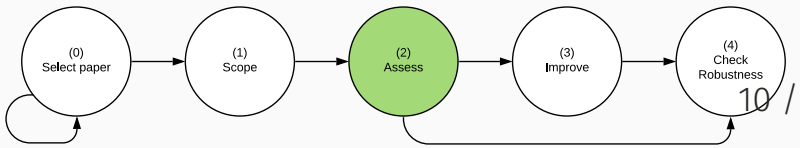
Notes: Regressions estimating equation (12) using data on 6 types of salt (listed in online Appendix A), from 133 districts in Northern India, annually from 1861 to 1930. Column 1 and column 2 estimated by OLS and NLS respectively; both include salt type × year and salt type × destination fixed effects.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



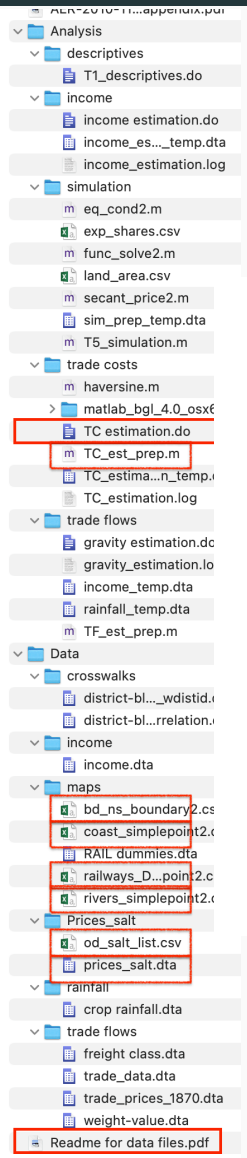
Standzarized Assessment




```
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
```



Standzarized Assessment




```
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
```

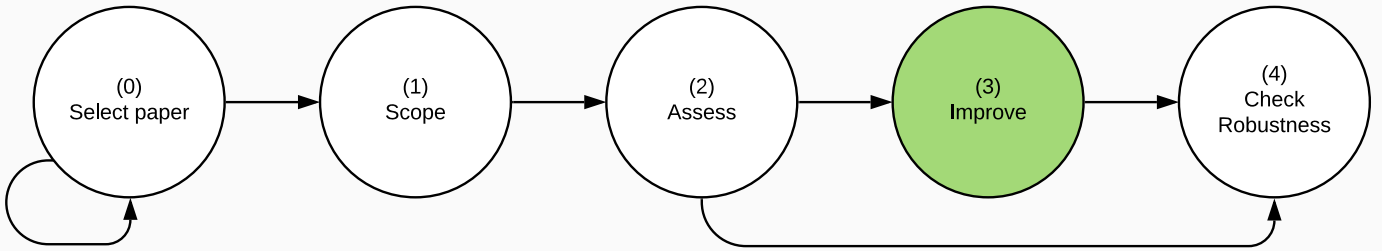
Table 3.5: Levels of Computational Reproducibility
(P denotes "partial," C denotes "complete")

	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	✔	✔	✔	✔	✔	✔	✔	✔	✔	✔

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

[b] **Computationally Reproducible from Raw data (CRR):** The output can be reproduced with minimal effort from the *raw* datasets.

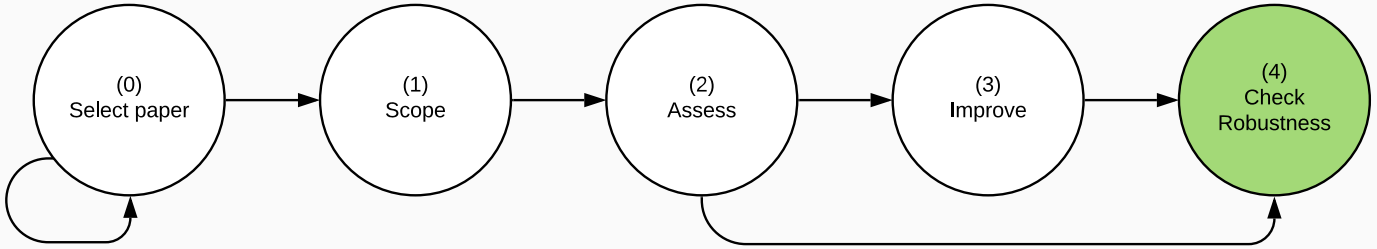
Standarized Improvements



Three types of improvements:

1. Improvements at the paper level (e.g., improve documentation following TIER's protocols)
2. Improvements at the display-item level
3. Specific future improvements

Robustness Checks



Two main parts for robustness:

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

Completed Reproduction: Generate Knowledge and Get Credit

Reproduction of: Railroads of the Raj: Estim
Transportation Infrastructure *American Ecor*
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

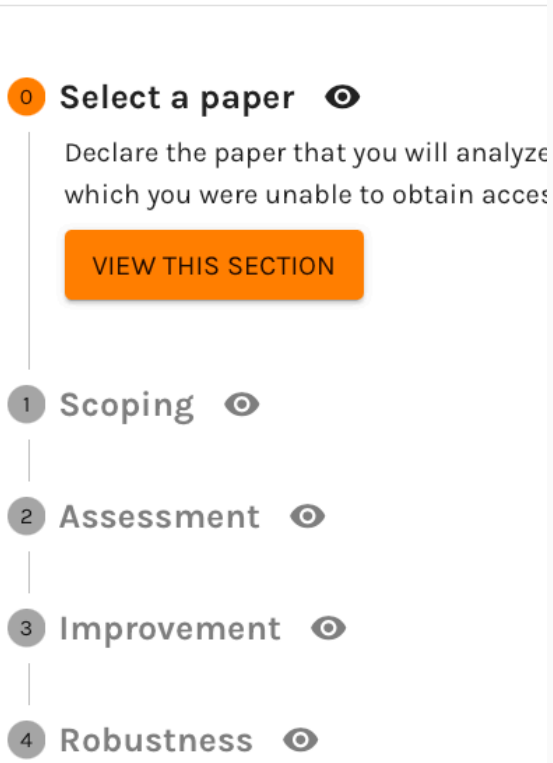
Claims identified by reproducer:

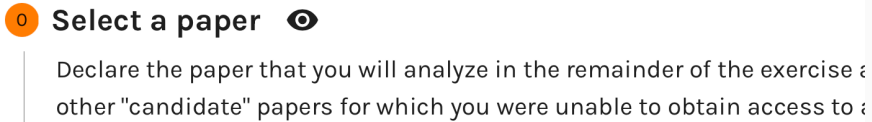
1. "The paper estimated the value of the extent to whi
within colonial India using OLS and NLS methods. U
costs of different modes of transportation are estir
reduction of LCRED (lowest-cost route effective dis
trading cost. This estimate has a 95% confidence ir

Reproducibility of Display Items:

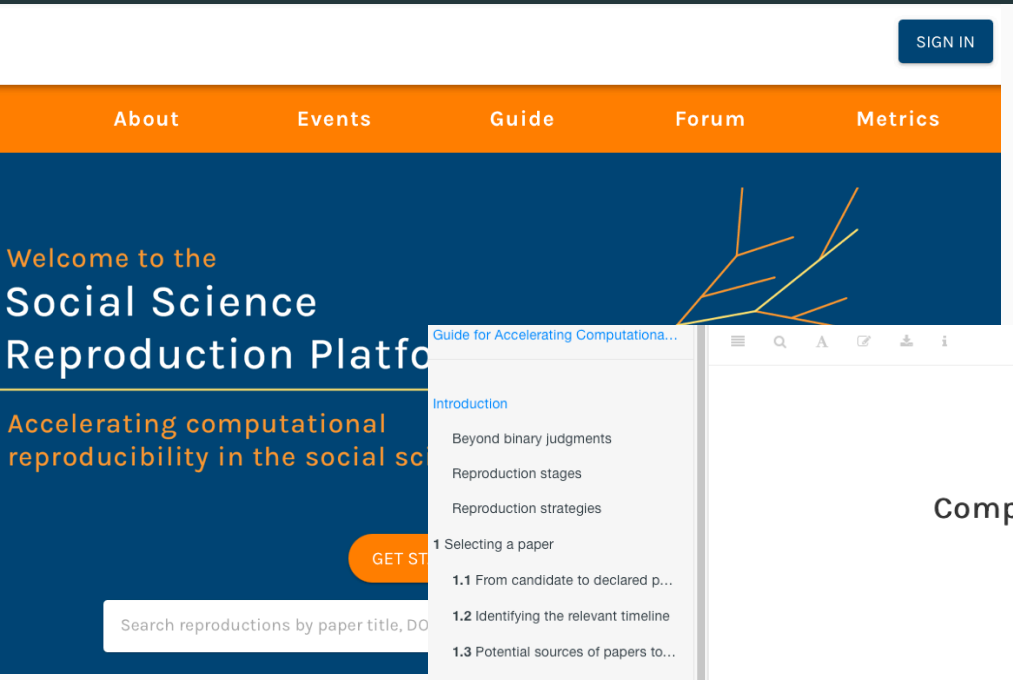
1. Table 2 - Table 2 -- Railroads and Trade Costs: Step 1...

Original reproduction package: https://www.aeaweb.org/articles?id=10.1257/aer.201
Revised reproduction package: https://github.com/em-ng21/railroads-of-the-raj
Original authors' availability for further inquiries: Not sure



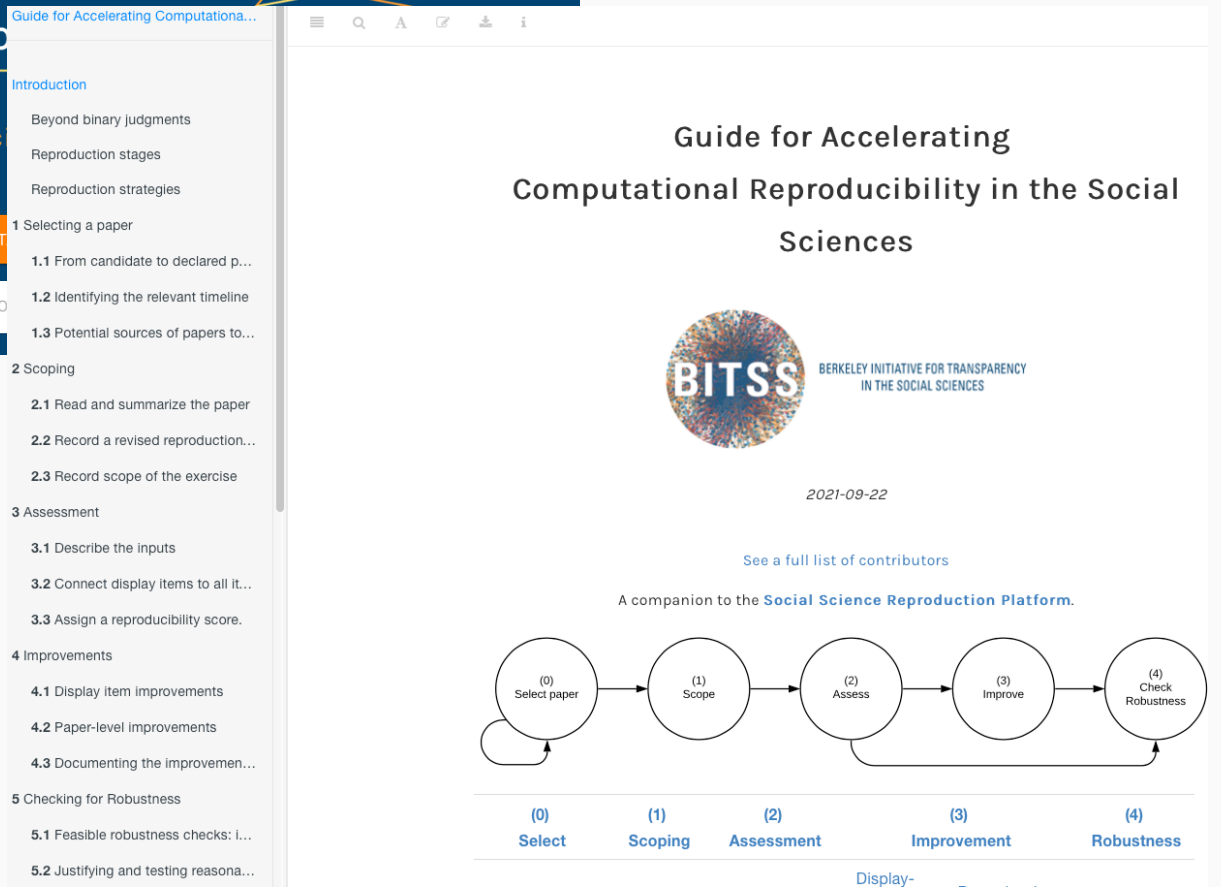


How Do I Get Started?

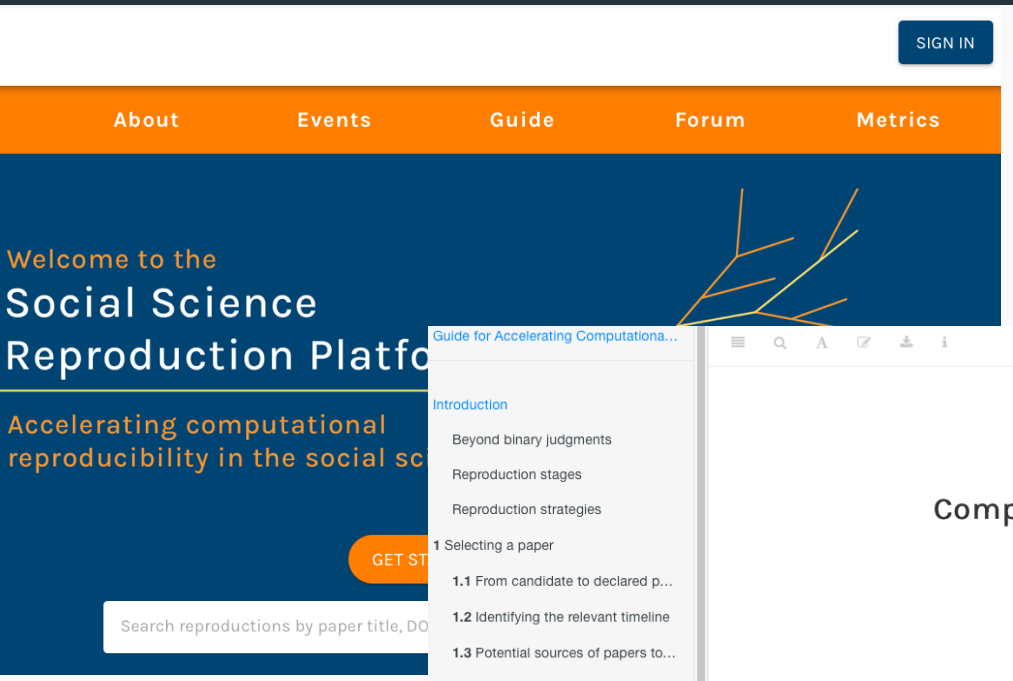


Use it for your class or independent project

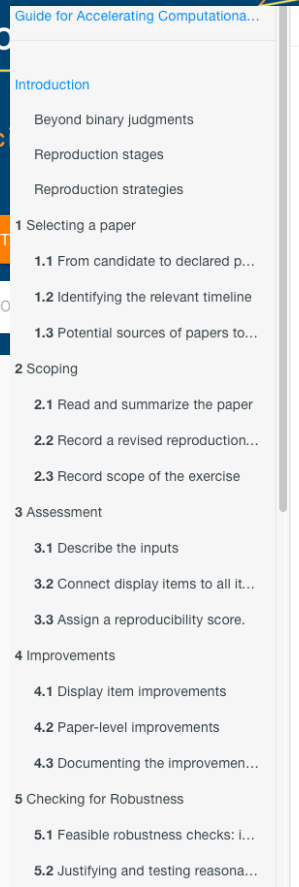
Consult the supporting **guide**

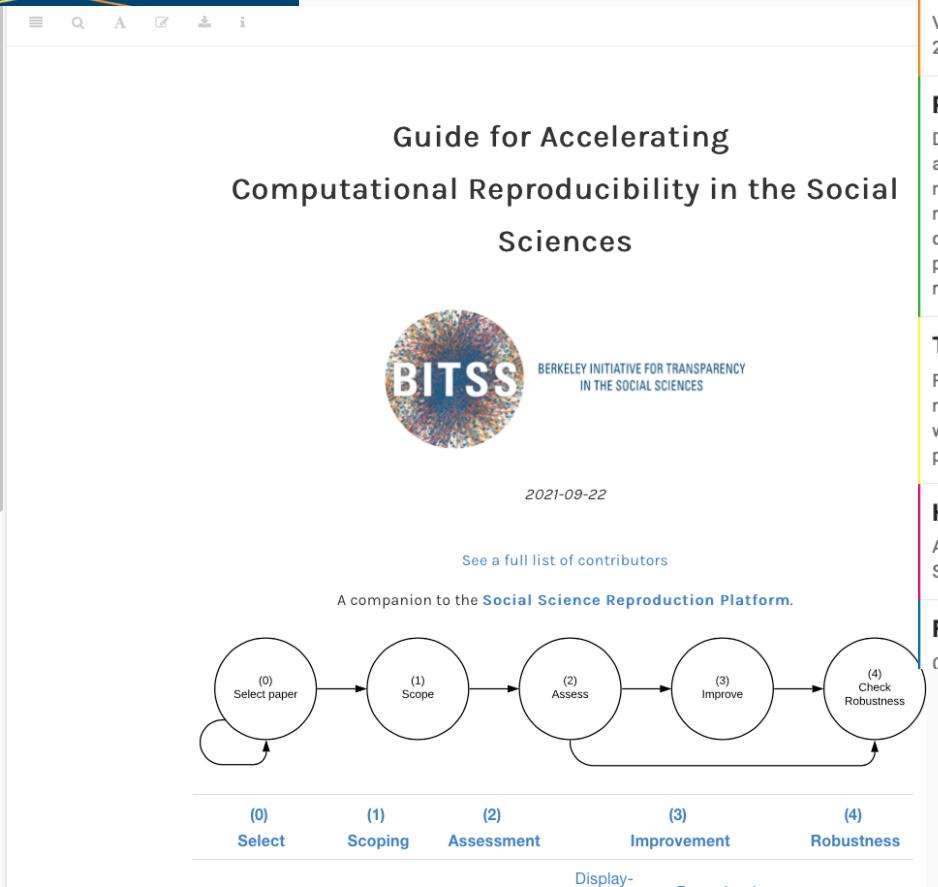


How Do I Get Started?

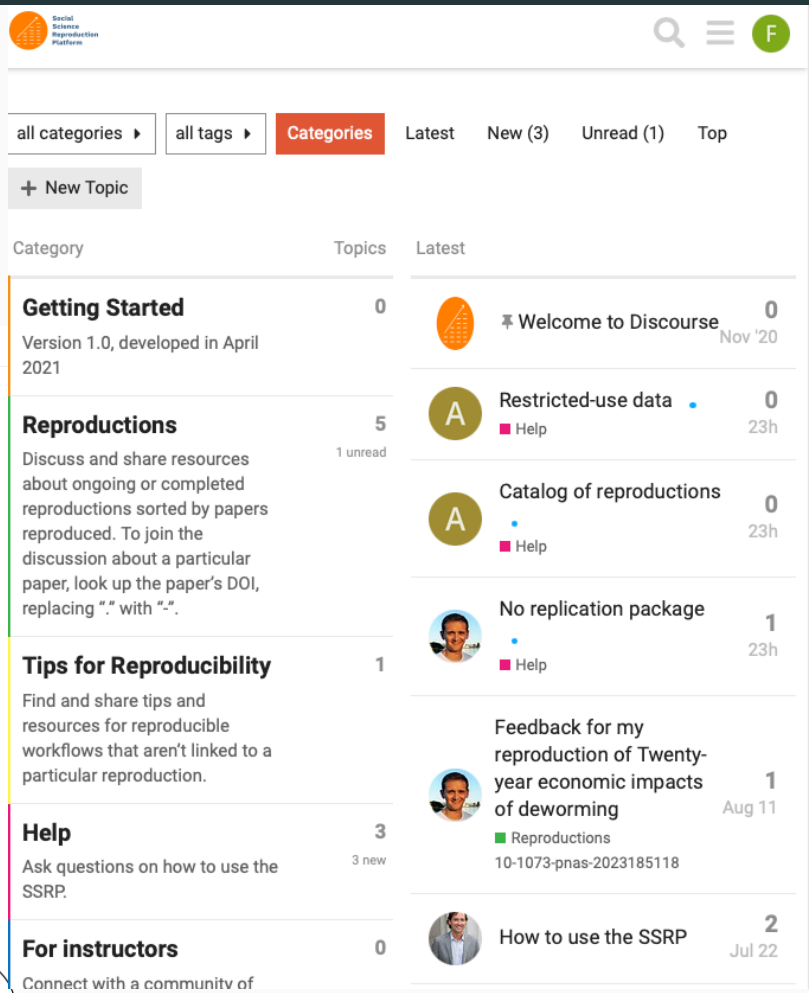


Use it for your class or independent project





Consult the supporting **guide**



Ask questions

First Insights Comming Out of the SSRP

Standardizing and Crowdsourcing Analysis to Assess Reproducibility in Economics

Abel Brodeur, University of Ottawa

Fernando Hoces de la Guardia, University of California, Berkeley

Edward Miguel, University of California, Berkeley and NBER

Seung Yong Sung, University of California, Berkeley

Lars Vilhuber, Cornell University

April 2024

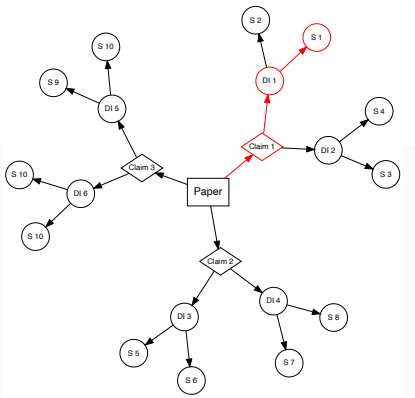
Abstract

This paper presents a framework to standardize computational reproductions in economics, and discusses a platform that implements this framework to crowdsource reproductions from around

First Insights: Some Utilization

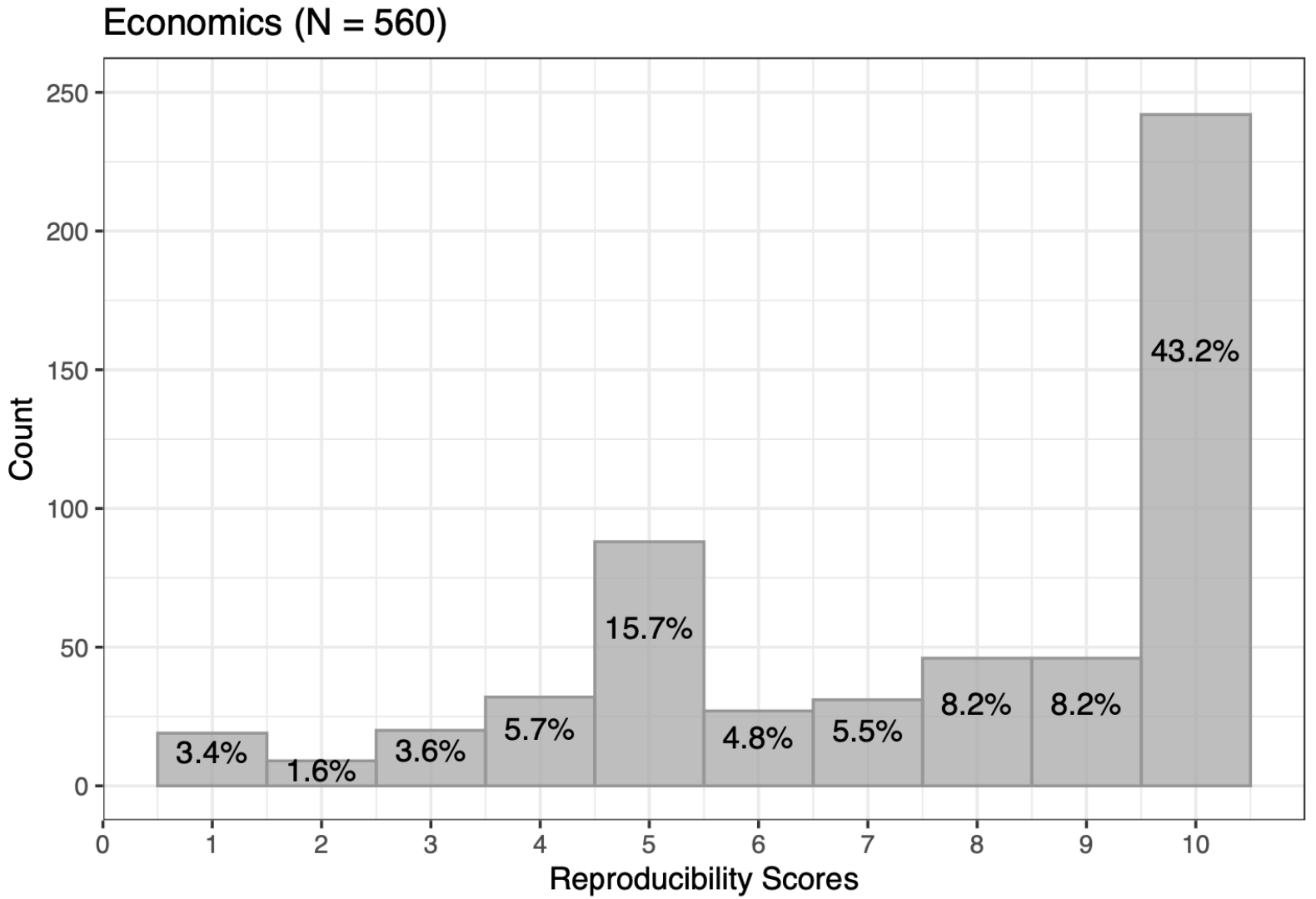
Table 2: Number of Items Evaluated on the Social Science Reproduction Platform (SSRP)

Reproduced Item	Count	Econ Only
Paper	186	145
Reproductions	260	206
Claims	510	384
Display Item	724	560
Claim+Display Item	840	587
Abandoned	10	10



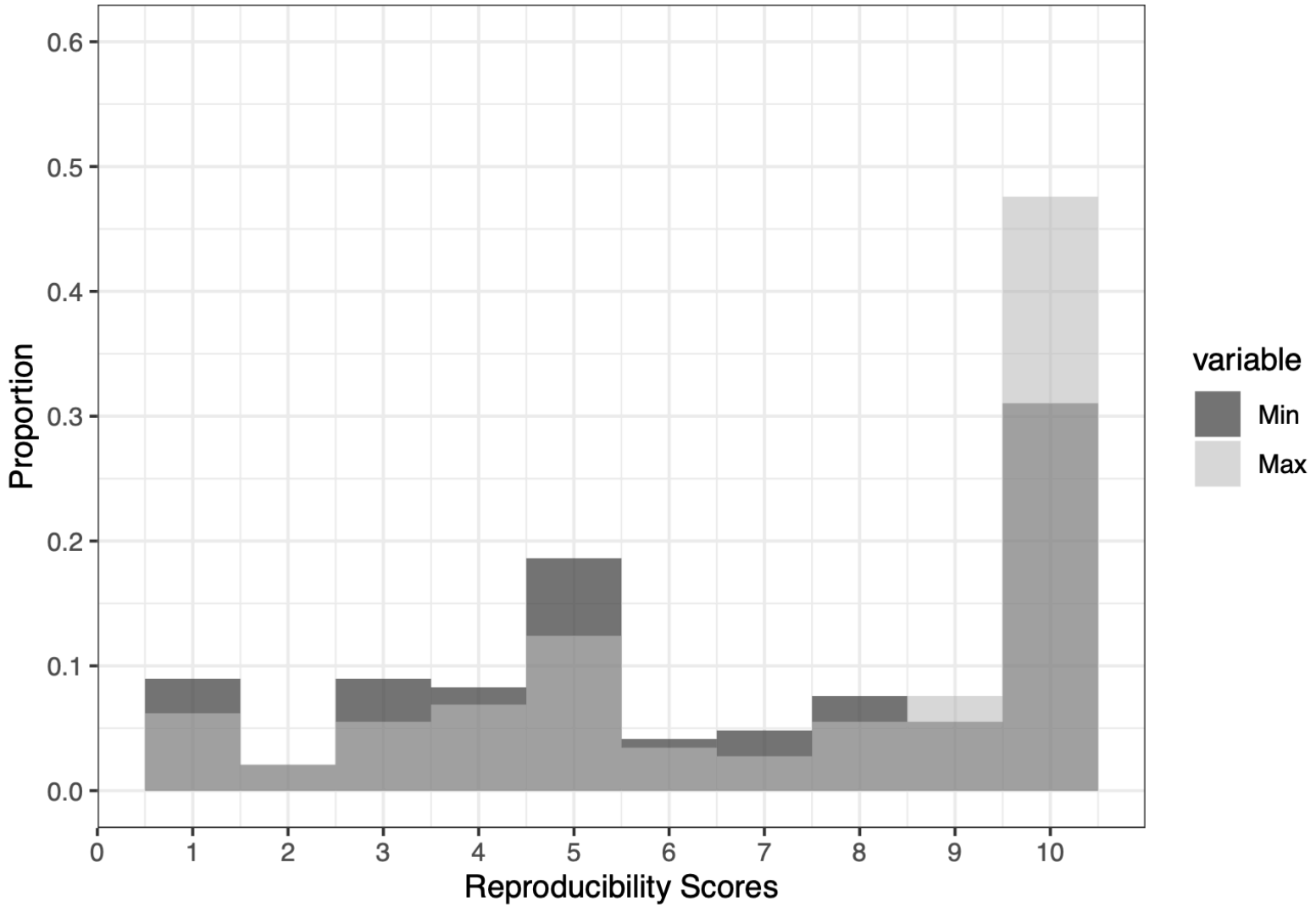
First Insights: Distribution of Scores for Display Items

Figure 2: Dist. of Display Item Reproducibility Scores in Economics (N = 560)



Can Aggregate at Paper Level In Different Ways

Figure 5: Dist. of Paper Level Reproducibility Scores in Economics (N = 145)



Thank You

acre@berkeley.edu
fhoces@berkeley.edu