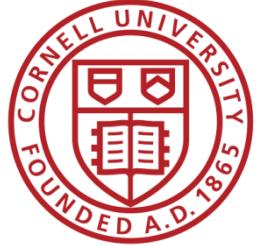


Implementing Increased Transparency, and Reproducibility in Economics

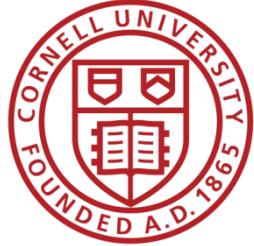
Lars Vilhuber
Cornell University

The opinions expressed in this talk are solely the authors, and do not represent the views of the U.S. Census Bureau, the American Economic Association, or any of the funding agencies.



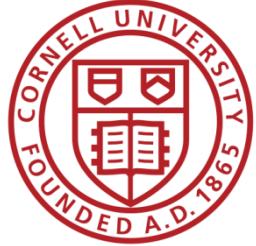
Efficiency of scholarly discourse?

- Early publications (20th century) contained **tables of data**, and the **math** was simple (maybe)
 - **Data** became electronic, was no longer **included** or **cited**
 - **Math** was transcribed to **code**, and was no longer **included**



SEASONAL VARIATIONS IN THE NEW YORK MONEY MARKET, 1890-1908												CIRCULATION OF DEPOSIT CURRENCY ^b		EXCHANGE RATES IN CHICAGO ON NEW YORK, ^c 1899-1908		NET INTERIOR MOVEMENT OF CASHS OUT OF AND INTO N. Y. CITY BANKS, ^c 1899-1908		STERLING EXCHANGE, DEMAND DRAFFTS ^f		EXPORTATION AND IMPORTATION OF GOLD, U. S., 1890-1908 (IN FIGURES) ^e	
CALL INTEREST RATES ON STOCK EXCHANGE ^b	INTEREST RATES ON 60-90 DAY, 2 NAME COMMERCIAL PAPER ^b	PERCENTAGE OF RESERVES TO DEPOSITS, N. Y. ASSOCIATED BANKS ^b	CIRCULATION OF DEPOSIT CURRENCY ^b	EXCHANGE RATES IN CHICAGO ON NEW YORK, ^c 1899-1908	NET INTERIOR MOVEMENT OF CASHS OUT OF AND INTO N. Y. CITY BANKS, ^c 1899-1908	STERLING EXCHANGE, DEMAND DRAFFTS ^f	EXPORTATION AND IMPORTATION OF GOLD, U. S., 1890-1908 (IN FIGURES) ^e														
AVERAGE RATE	SEASONAL INDEX NUMBER	AVERAGE RATE	SEASONAL INDEX NUMBER	AVERAGE PERCENTAGE	SEASONAL INDEX NUMBER	AVERAGE CLEARINGS (\$000,000)	SEASONAL INDEX NUMBER	AVERAGE RATE (PREMIUM OR DISCOUNT)	SEASONAL INDEX NUMBER	AVERAGE AMOUNT OUT OF 000	INTO 000	SEASONAL INDEX NUMBER	AVERAGE RATE	SEASONAL INDEX NUMBER	TOTAL EXPORTS 000	TOTAL IMPORTS 000					
6.4	43.4	5.0	53.1	28.6	44.3	\$1,937.5	60.8	2.5 P	64.7	86,684	87.2	4,8606	49.7	Jan.	832,747						
3.6	23.8	4.7	41.5	29.1	78.8	*1,233.6	*59.6	5 P	67.4	6,621	84.9	4,8657	54.7								
2.8	14.9	4.5	31.2	29.9	96.9	*1,234.7	*54.4	5 P	67.7	7,773	90.7	4,8679	59.4								
2.5	11.9	4.3	22.7	30.3	77.8	*1,140.0	*44.0	10 P	72.1	6,895	87.6	4,8697	64.1								
2.5	11.1	4.3	22.9	29.9	65.4	*1,190.5	*52.5	2 P	63.0	4,749	64.1	4,8695	64.1	Feb.	13,408						
2.4	10.1	4.3	22.1	29.2	58.1	*1,084.1	*38.4	6 D	54.8	2,376	77.0	4,8696	64.8								
2.5	9.8	4.3	22.2	28.8	53.6	*1,004.8	*32.1	9 D	50.7	1,436	63.7	4,8708	66.9								
2.7	13.4	4.4	26.5	28.5	45.5	*944.0	*22.6	20 D	38.8	1,157	52.5	4,8697	65.4								
3.0	15.1	4.6	32.6	28.1	43.1	*1,165.7	*51.5	29.5 D	28.1	1,679	88.5	4,8692	65.7	March	\$ 43,233						
3.6	19.7	*4.6	*34.3	27.9	41.1	*1,067.9	*38.2	23 D	35.0	604	30.5	4,8676	62.0								
3.9	22.4	4.8	40.0	27.7	37.0	*1,119.7	*42.7	13 D	45.9	716	49.8	4,8665	59.1								
3.2	19.2	4.8	39.6	27.9	39.9	1,042.3	33.1	14.5 D	43.5	1,533	54.4	4,8681	61.6								
3.6	22.0	4.8	38.1	28.0	40.5	1,051.4	35.5	5 D	53.9	999	53.5	4,8704	65.9	April	25,888						
4.0	23.8	4.7	36.7	27.8	35.7	1,135.4	48.0	14 D	44.5	868	53.9	4,8711	67.4								
3.8	23.1	4.6	33.4	27.9	39.9	1,119.0	42.9	7.5 D	52.9	1,903	59.0	4,8714	68.2								
3.0	17.5	4.5	31.9	28.4	50.9	1,123.5	46.7	4 P	66.3	2,085	62.1	4,8734	73.6								
2.9	15.4	4.4	27.5	28.6	54.4	1,107.6	43.3	9 D	48.4	1,379	61.6	4,8743	78.1	May							
3.4	19.3	4.4	26.9	28.3	48.3	1,283.3	67.3	3.5 D	55.9	594	56.5	4,8739	76.3								
3.5	19.5	4.4	24.5	28.4	49.0	1,175.4	52.7	2.5 P	62.0	9,952	63.0	4,8734	74.2								
2.6	13.9	4.3	22.7	28.6	51.6	1,123.4	48.0	16 P	76.7	4,306	74.5	4,8739	75.5								
2.4	11.2	4.2	19.9	29.0	60.3	1,011.8	34.1	16 P	77.3	4,329	74.7	4,8752	79.1	June							
2.3	9.6	4.1	17.1	28.8	57.2	908.1	21.4	10 P	71.1	3,862	60.9	4,8760	80.9								
2.3	8.0	4.1	15.8	28.7	56.1	1,039.4	37.9	5 P	64.6	3,229	68.6	4,8757	81.1								
2.4	7.7	4.1	15.3	28.7	56.7	967.8	31.1	4 P	63.6	3,354	66.7	4,8756	81.0								
2.5	8.0	4.3	18.4	28.7	57.5	938.7	25.8	10.5 P	72.8	3,897	68.5	4,8742	79.0	July							
3.6	16.4	4.5	22.0	28.4	53.5	1,013.9	35.4	11.5 P	73.6	2,158	58.3	4,8721	74.6								
3.4	13.6	4.5	25.0	27.9	45.0	991.5	33.1	16.5 D	40.3	1,441	53.1	4,8715	72.9								
2.9	9.6	4.6	26.9	28.4	56.3	1,034.6	35.6	7.5 D	50.6	3,456	68.0	4,8717	72.6								
2.3	5.3	4.6	31.1	28.7	63.3	970.2	26.6	8 D	52.6	3,692	69.3	4,8717	72.6								
2.4	5.6	4.6	33.5	28.7	65.4	924.6	21.1	10.5 D	50.0	4,735	73.1	4,8720	73.2								
2.5	6.0	4.6	35.2	28.3	60.8	969.7	27.9	11 D	48.7	2,955	63.4	4,8702	69.6	August	44,300						
2.5	6.3	4.8	40.5	28.0	54.3	910.6	20.8	17.5 D	41.8	1,395	57.3	4,8693	68.0								
2.6	7.4	4.9	43.7	27.8	49.3	948.0	25.9	19 D	40.1	9,517	49.4	4,8669	61.3								
3.7	13.6	5.3	49.5	27.7	47.7	931.1	23.9	34.5 D	92.7	8249	45.5	4,8651	56.9								
3.0	12.3	5.3	51.8	27.6	42.6	956.8	29.0	37.5 D	18.8	1,477	33.7	4,8626	50.4								
4.1	20.7	5.3	55.4	27.2	32.8	880.7	19.2	36.5 D	19.1	2,690	99.9	4,8601	43.7								
4.2	23.4	5.1	57.5	27.0	29.8	1,033.6	38.6	25 D	34.7	2,589	30.3	4,8584	35.2								
4.3	30.6	5.3	64.7	27.1	31.9	1,058.7	44.3	26 D	33.5	3,434	94.8	4,8552	32.0								
4.2	29.6	5.3	63.2	27.5	37.4	1,066.1	36.9	33 D	26.1	3,489	97.0	4,8557	31.9	Oct.							
4.5	27.9	*6.2	*61.7	27.3	33.0	1,135.2	59.0	32 D	27.2	3,883	39.0	4,8538	27.3								
4.0	24.4	*5.1	*61.5	27.3	33.0	1,094.1	46.4	29.5 D	29.0	32.5	30.3	4,8540	29.7								
3.6	19.4	*4.9	*53.2	27.5	34.1	1,139.3	49.6	27.5 D	30.8	3,014	34.7	4,8549	33.9								
6.5	29.3	*4.9	*51.4	27.6	36.4	1,144.0	50.1	31 D	24.2	3,685	29.6	4,8576	41.5								
7.1	32.9	*4.9	*48.9	27.2	27.5	1,140.7	54.3	29 D	27.5	2,700	34.7	4,8567	39.7	Nov.	96,743						
5.4	30.3	*4.9	*51.3	27.1	29.7	1,077.6	45.3	20 D	36.9	2,666	37.1	4,8554	38.8								
4.8	26.1	*5.0	*53.5	27.4	29.4	1,983.9	65.7	4.5 D	33.4	1,530	43.6	4,8554	44.1								
4.2	26.1	*4.7	*46.0	27.8	36.1	1,107.7	48.1	2.5 D	56.3	49.0	4,8623	49.5									
4.0	26.8	4.8	48.6	27.6	32.3	1,191.3	65.2	11.5 D	47.3	836	44.3	4,8615	49.3								
4.9	30.3	*4.7	*47.8	27.2	24.9	1,202.4	63.5	5 P	64.7	615	52.4	4,8596	45.6								
5.5	39.2	*4.8	*51.6	27.4	29.4	1,201.1	60.8	3.5 P	65.1	60	47.8	4,8604	47.0								
6.6	46.1	*4.8	*49.3	27.5	32.8	1,015.3	35.9	3.5 P	65.1	9,188	61.7	4,8611	49.0								
7.4	49.3	*4.9	*52.9	27.7	35.3					4,8592	45.0										

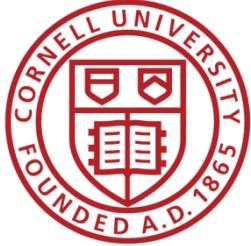
From @sdellavi
AER 1911



Efficiency of scholarly discourse!

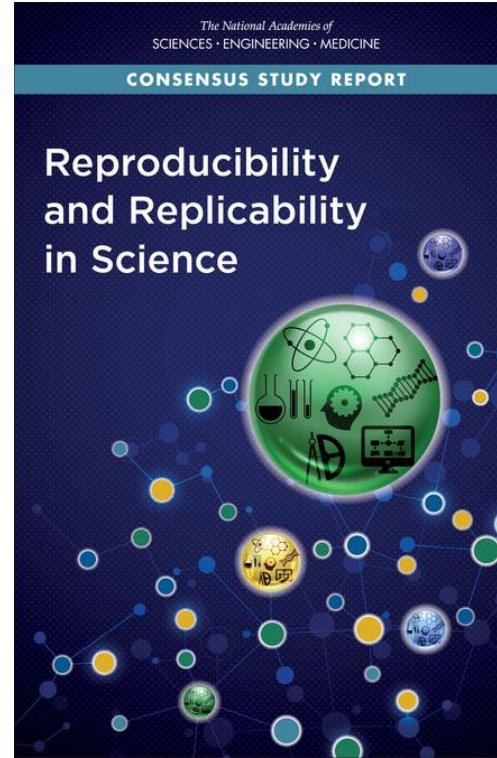
**Modern publications thus need
the same transparency and completeness
as in the old days
to facilitate replicability**

Replication?



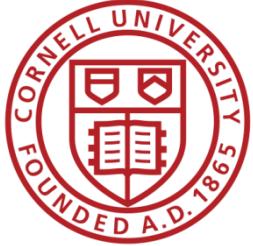
Replication continuum

<https://doi.org/10.17226/25303>

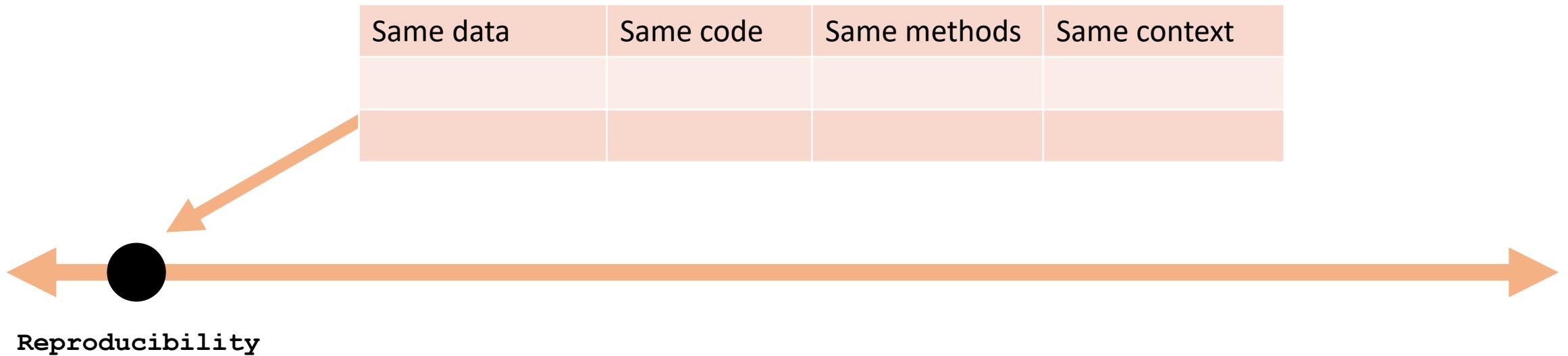


Reproducibility

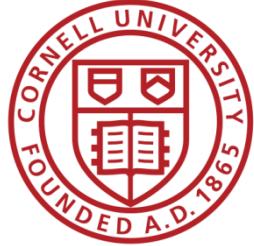
- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)



Replication continuum



- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)



Replication continuum

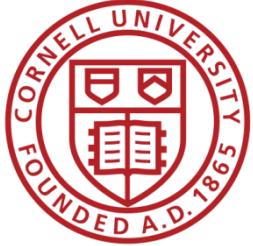


Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)



Replication continuum

Same data	Different code or software	Same methods	Same context

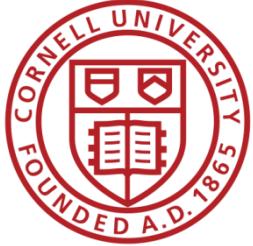


Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)



Replication continuum

New data collection	Same code	Same methods	Same context

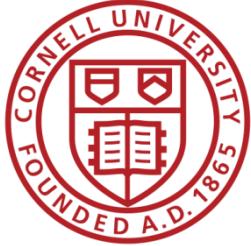


Reproducibility

- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)



Replication continuum



Reproducibility

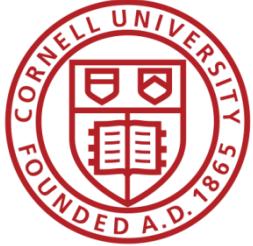
- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

Replicability

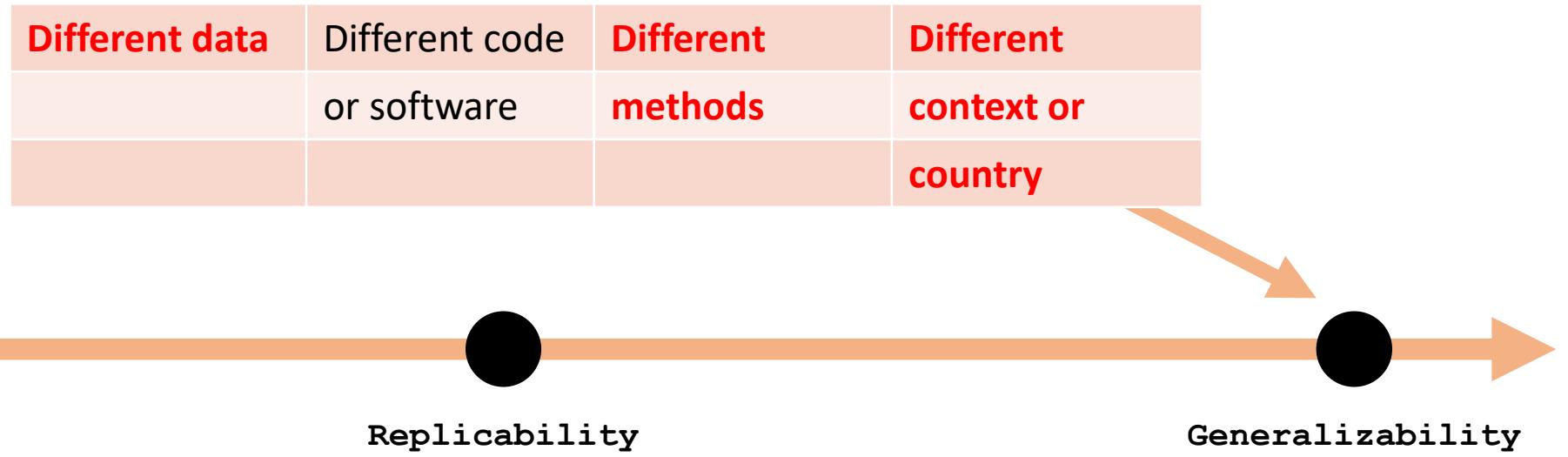
- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

Generalizability

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)



Replication continuum

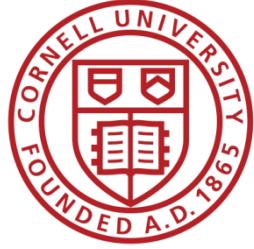


- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)

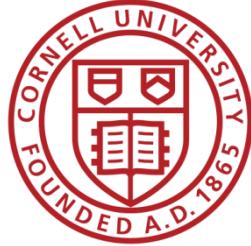
Progress



Progress

- Replication archives and Data (Code) Availability policies





Progress

- Replication archives and Data (Code) Availability policies
- Shared open source software



Statistical Software Components

From [Boston College Department of Economics](#)
Boston College, 140 Commonwealth Avenue, Chestnut Hill MA 02467 U:
Contact information at [EDIRC](#).
Bibliographic data for series maintained by Christopher F Baum (baum@bc.edu)

[Access Statistics](#) for this software series.

Track citations for all items by [RSS feed](#)

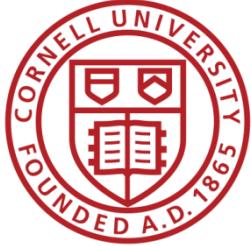
Is something missing from the series or not right? See the RePEc data [series](#).

[GAPPORT: Stata module to calculates seats in party-list representation](#) [downloads](#)

Ulrich Kohler

[GCLSORT: Stata module to sort a single variable via ege](#)
Philippe Van Kerm

[GPROD: Stata module to extend egen for product of obs](#)
Philip Ryan

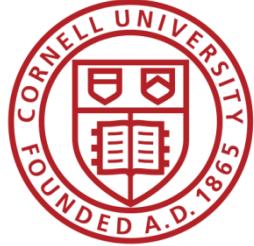


Progress

- Replication archives and Data (Code) Availability policies
- Shared open source software
- Better public-use and shared data



INSTITUT FÜR ARBEITSMARKT- UND
BERUFSFORSCHUNG
Die Forschungseinrichtung der Bundesagentur für Arbeit



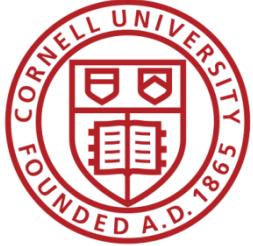
Progress

- Replication archives and Data (Code) Availability policies
- Shared open source software
- Better public-use and shared data
- Better ways of accessing preprints/ grey literature

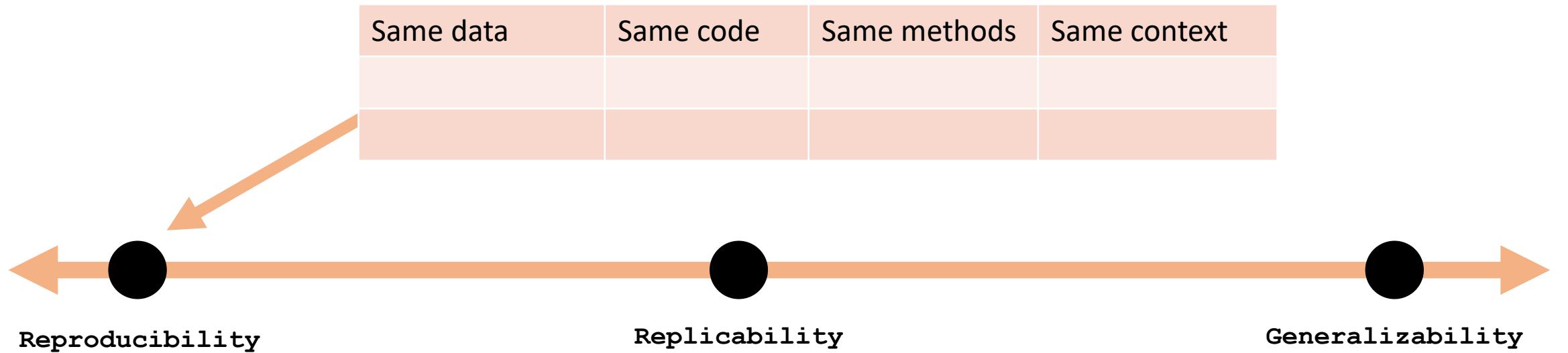
RePEc



Issues



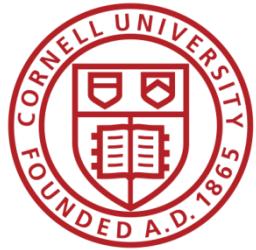
Replication continuum



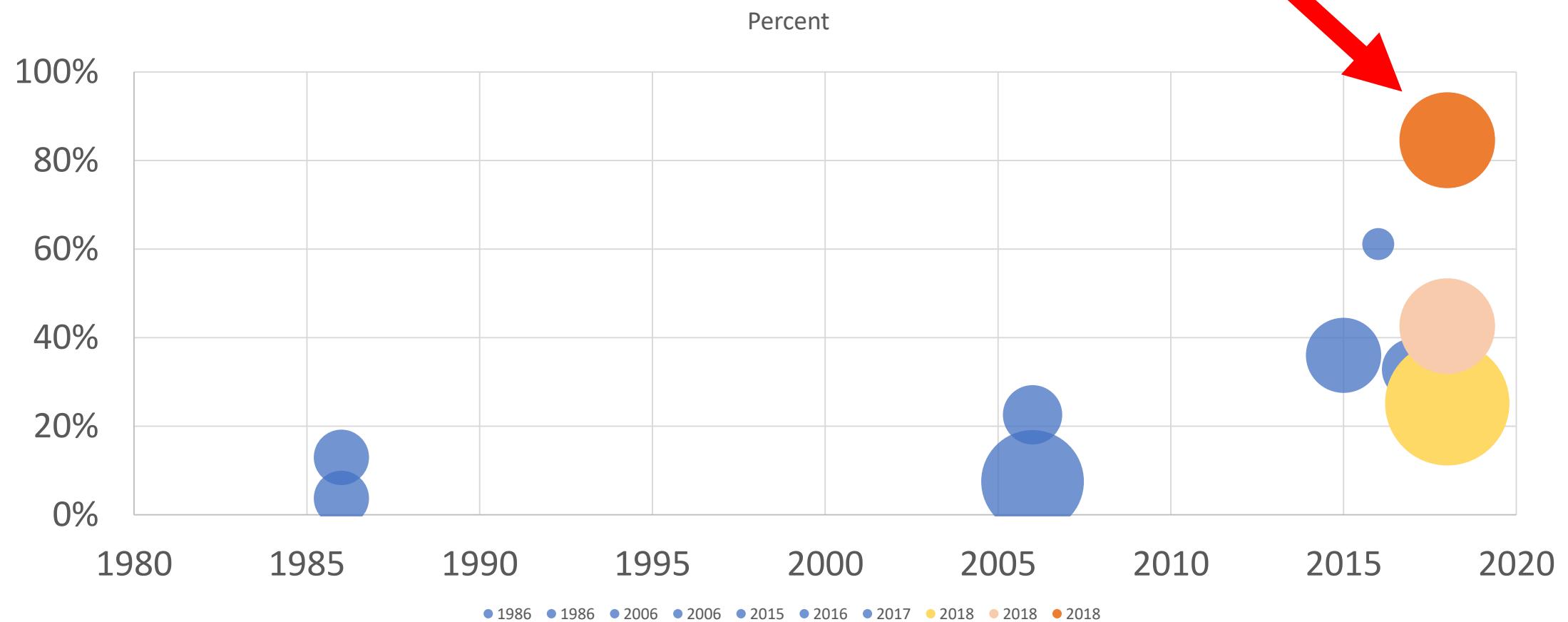
- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)

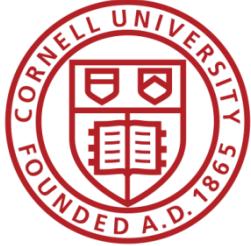
- Wide Replication (Pesaran 2003)
- Statistical Replication (Hamermesh 2007)
- Reproduction/Reanalysis (Clemens 2015)

- Wider Replication (Pesaran 2003)
- Scientific Replication (Hamermesh 2007)
- Reanalysis/Robustness (Clemens 2015)



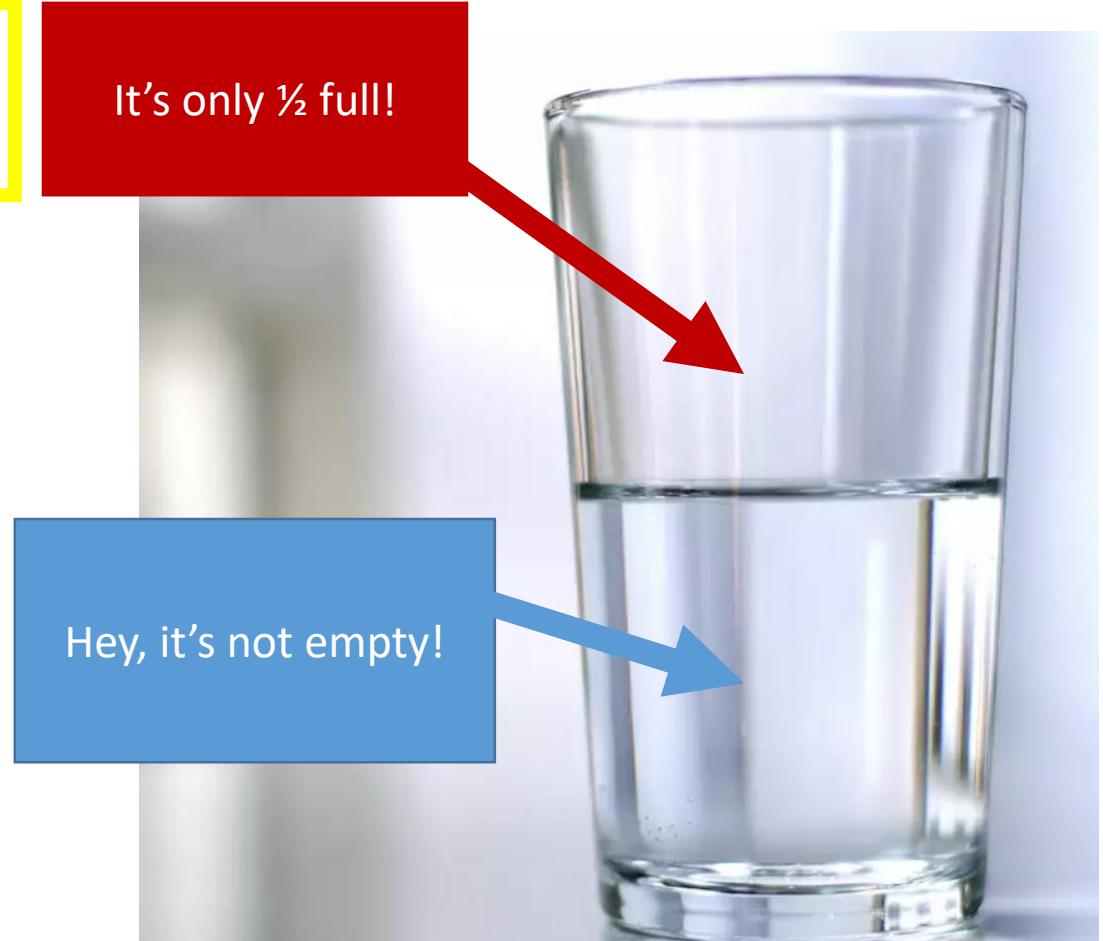
Results?

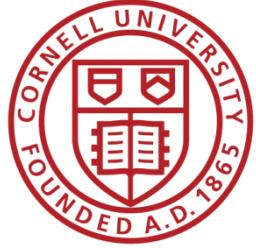




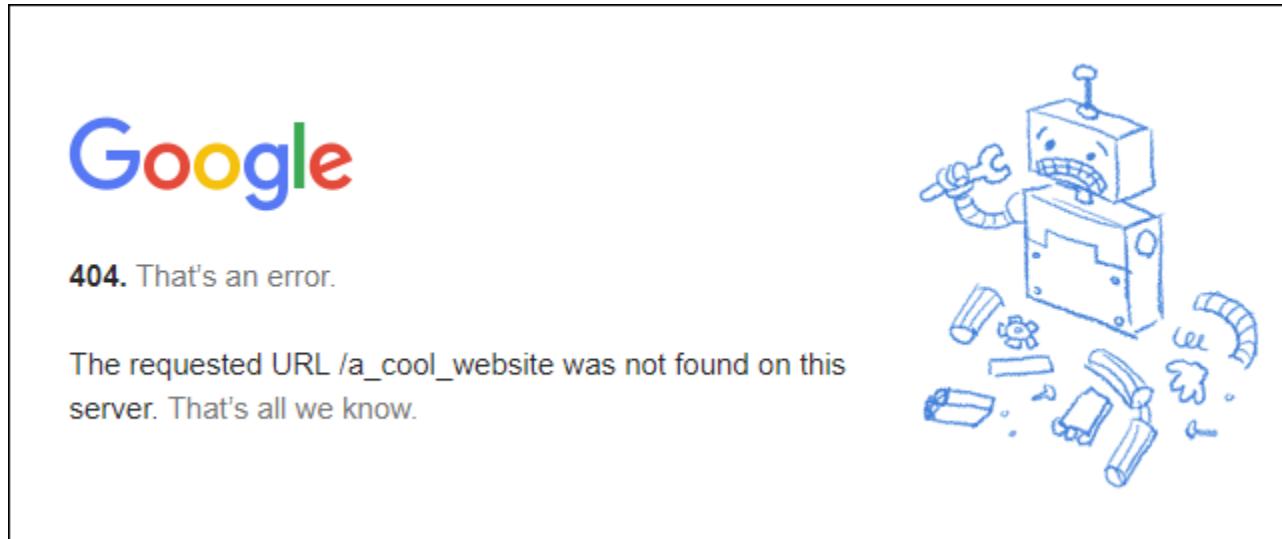
In a nutshell

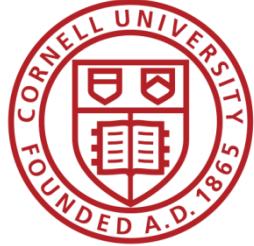
- **40%** use restricted-access data
- **25%** use public-use data and are mostly or completely reproducible
- **25%** use public-use data and are only partially reproducible
- **10%** fail to yield useful results





Failure to curate





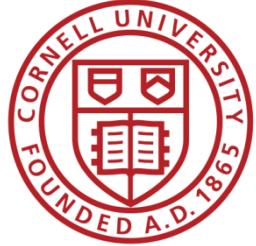
Poor citation practices

- **Macrodata:**

“We use data downloaded from
the Bureau of Economic Analysis...”

- **Microdata:**

“... this paper uses data from
the Current Population Survey...”

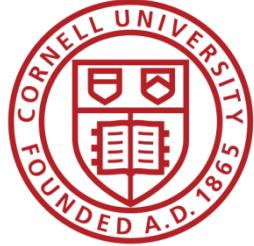


Problems describing RELIABLE archives

Many datasets

- Are imperfectly described
 - Very few data citations
- Are badly documented
- Have no (permanent) location defined
 - Even for data from high-profile organizations!
- All of the above

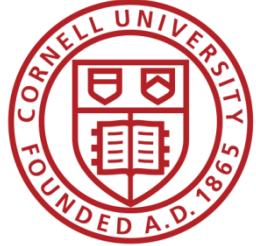
What to
do?



Second round (2012-)

- Greater enforcement of data (and code) availability
 - 2015, AJ Political Science
 - 2016, Data Editor for ASA Software Section
 - 2016, Statistical review added Science
 - 2017: AEA appoints Data Editor, with mandate to do similar activities (also EJ, Restud)

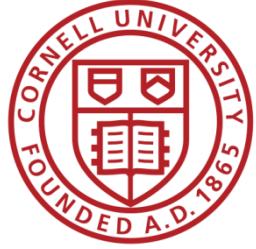
Verifying
reproducibility



Current Data Availability Policies are Broken

- If the Data is
not open-access,

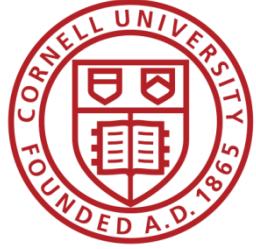
**no systematic information is
collected
("exemption")**



We asked for “deposits”...

If you used files at
the National Archives,

would we ask you to
“deposit” them?

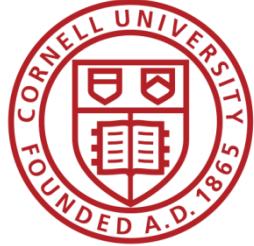


We asked for “deposits”...

If you used files at
the National Archives,

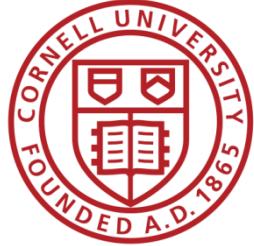
**you should describe
where they are!**

Greater
transparency of
data/code



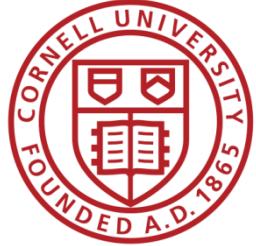
Why do journals like “supplemental ZIP files” and affiliated repositories?

- They can ensure **longevity/ persistence**
- They can ensure **access**
- They can ensure **availability**



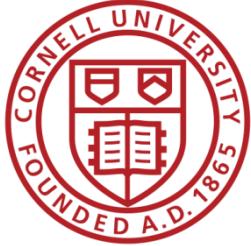
What are the characteristics of trusted repositories (data archives)?

- They DO ensure **longevity/ persistence**
- They DO ensure **access**
- They DO ensure **availability**



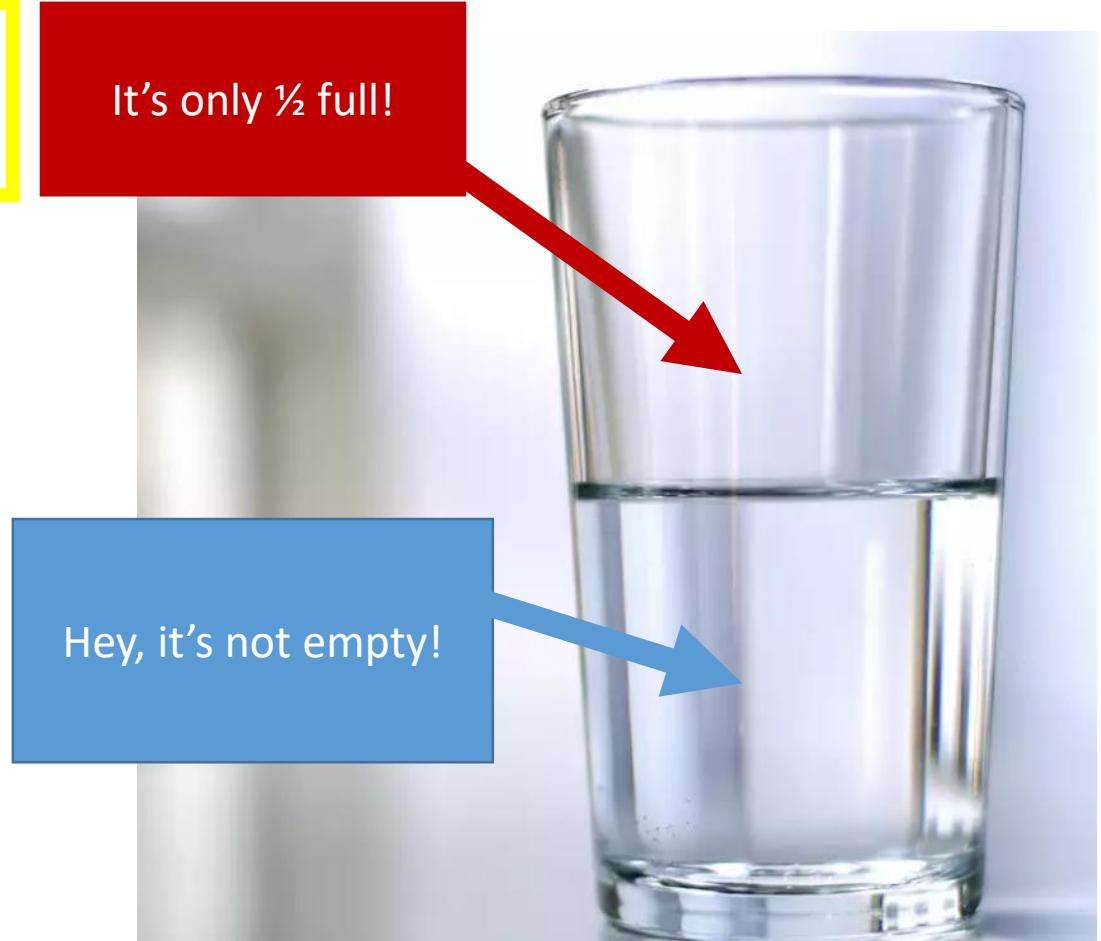
Evolving Journal and Data Infrastructure

- More self-deposit repositories in the social sciences
 - Dataverse
 - Figshare
 - (open)ICPSR
 - Zenodo
 - Qualitative Data Repository (QDR)
 - Others...



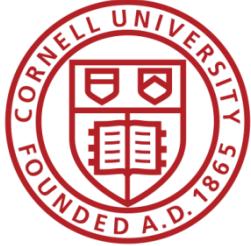
In a nutshell

- **40%** use restricted-access data
- **25%** use public-use data and are mostly or completely reproducible
- **25%** use public-use data and are only partially reproducible
- **10%** fail to yield useful results



It's only $\frac{1}{2}$ full!

Hey, it's not empty!

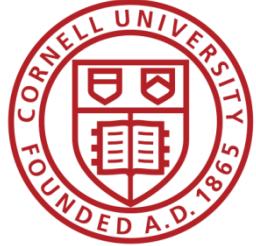


Evolving Journal and Data Infrastructure

- More self-deposit repositories in the social sciences
 - Dataverse
 - Figshare
 - (open)ICPSR
 - Zenodo
 - Qualitative Data Repository (QDR)
 - Others...

- CASD
 - IAB
 - Norway
 - US Federal Statistical RDC
 -

Use trusted
repositories

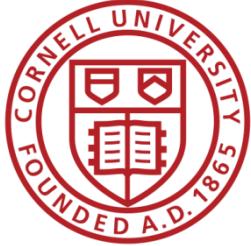


Evolving Journal and Data Infrastructure

Goal: Use any
repository!

(subject to conditions)

Problems
with that?



Here are the problems...



Failure to curate



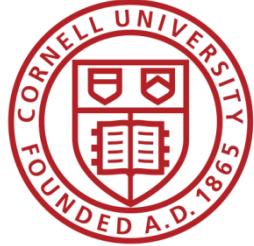
Economics makes wide use of public-use data

- **Macrodata:**

“We use data downloaded from the Bureau of Economic Analysis...”

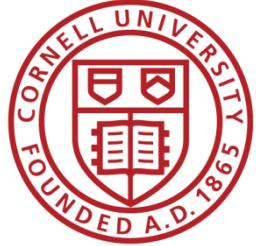
- **Microdata:**

“... this paper uses data from the Current Population Survey...”



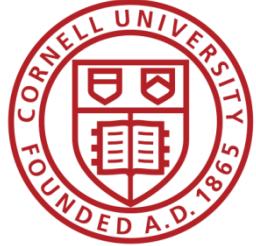
Verifying Data and Code Deposits

- Not every data repository is created equal
 - **Github, Dropbox, etc. are not data or code repositories**
 - Is the institutional repository at the University of Southern Venezuela a reliable repository?
 - Is the institutional repository at Cornell University a reliable repository?
 - Is the institutional repository at Harvard University (Dataverse!) a reliable repository?
 - **Are the National Archives a reliable repository?**



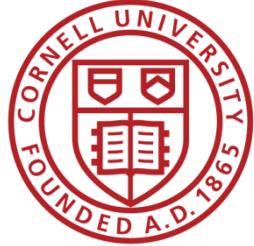
Verifying Data and Code Deposits

- Not every restricted-access repository is created equal
 - The **Second Bank of Third City credit card data** is not a data/code repository
 - Is the School Board of Third City a reliable repository?
 - Is the JPMC Institute a reliable repository?
 - Is the **US Census Bureau** a reliable repository?
 - **Are any restricted-access repositories reliable archives?**



Evolving Journal and Data Infrastructure

So: Describe them!
(cite them!)



Action: Data citations and metadata

What is FAIR?

- Findable,
- Accessible,
- Interoperable, and
- Re-usable

The FORCE11 logo features a blue circular icon with a white target-like pattern next to the word "FORCE11". Below it is the tagline "The Future of Research Communications and e-Scholarship". A navigation bar below the logo includes "ABOUT", "COMMUNITY", and "CODE OF CON".

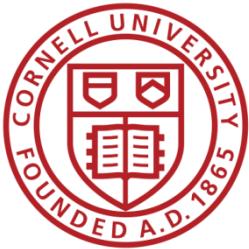
FORCE11 » Groups » The FAIR Data Principles

THE FAIR DATA PRINCIPLES

JOIN IN THE DISCUSSION - LEARN
FAIR Data Principles

Preamble

One of the grand challenges of data-intensiv



perceived criteria of importance.

1. Importance

Data should be considered legitimate, citable products of research. Data should be accorded the same importance in the scholarly record as citat research objects, such as publications[1].



Data Citation Principles

2. Credit and Attribution

Data citations should facilitate giving scholarly credit and normative and le attribution to all contributors to the data, recognizing that a single style or of attribution may not be applicable to all data[2].

3. Evidence

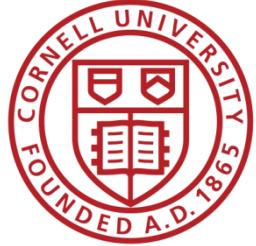
In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited[3].

4. Unique Identification

A data citation should include a persistent method for identification that i actionable, globally unique, and widely used by a community[4].

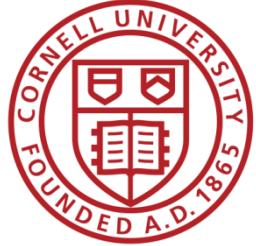
5. Access

Data citations should facilitate access to the data themselves and to such metadata, documentation, code, and other materials as are necessary for



Evolving Journal and Data Infrastructure

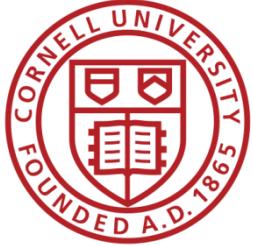
Data Citations are not
enough!



Why are data citations not enough?

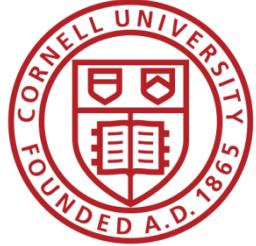
- They tell you “where”
- But most do not
 - “who can access”
 - “for how long”
 - “under what conditions”

(Though in theory, these are covered by
the Data Citation Principles)



Data Availability Statements (DAS)

- A statement about **where data** supporting the results reported in a published article can be found
 - including unique identifiers linking to publicly archived datasets analyzed or generated during the study.
- DASs can **increase transparency** by providing a reason why data cannot be made (immediately) available
 - **need for registration**, ethical or legal restrictions, or because of an embargo period



Data Availability Statements

- A statement about **how long** data will be **available** (policy)
 - DOI assignments implies **long-term curation**
 - But long-term curation **does not require DOI!**
- A statement about **usage rights**
 - Not every dataset is in the public domain
 - Not everybody knows that U.S. Government data are usually in the public domain

Improve
provenance
documentation



American Economic Review



The *American Economic Review* is a general-interest economics journal. Established in 1911, the AER is among the nation's oldest and most respected scholarly journals in economics.

Journal of Economic Literature



The *Journal of Economic Literature* (JEL), first published in 1969, is designed to help economists keep abreast of and synthesize the vast flow of literature.

American Economic Journal: Applied Economics



American Economic Journal: Applied Economics publishes papers covering a range of topics in applied economics, with a focus on empirical microeconomic issues.

American Economic Journal: Macroeconomics



American Economic Journal: Macroeconomics focuses on studies of aggregate fluctuations and growth, and the role of policy in that context.

AMERICAN ECONOMIC ASSOCIATION

American Economic Review: Insights



AER: Insights is designed to be a top-tier, general-interest economics journal publishing papers of the same quality and importance as those in the AER, but devoted to publishing papers with important insights that can be conveyed succinctly.

Journal of Economic Perspectives



The *Journal of Economic Perspectives* (JEP) fills the gap between the general interest press and academic economics journals.

American Economic Journal: Economic Policy

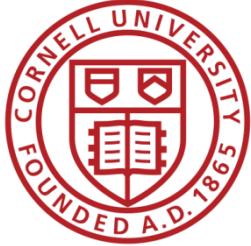


American Economic Journal: Economic Policy publishes papers covering a range of topics, the common theme being the role of economic policy in economic outcomes.

American Economic Journal: Microeconomics

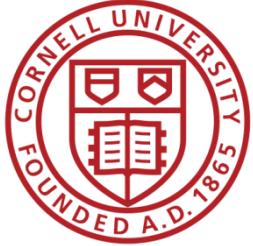


American Economic Journal: Microeconomics publishes papers focusing on microeconomic theory; industrial organization; and the microeconomic aspects of international trade, political economy, and finance.



Current efforts at the AEA

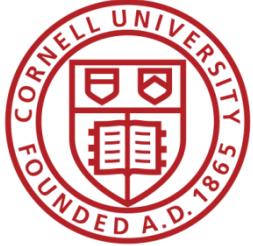
- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information



AEA Data Availability Policy (2018)

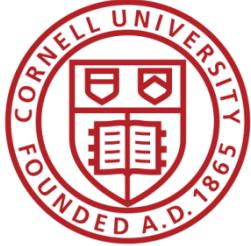
- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely** documented and are **readily available** to any researcher for purposes of replication.
- Authors of accepted papers that contain empirical work, simulations, or experimental work must **provide, prior to publication**, the data, programs, and other details of the computations **sufficient to permit replication**. These will be **posted on the AEA website**. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.

July 16, 2019



AEA Data & Code Availability Policy (2019)

- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely documented** and **access to the data and code is clearly and precisely documented and is non-exclusive to the authors.**
- Authors of accepted papers that contain empirical work, simulations, or experimental work must **provide, prior to acceptance**, the data, programs, and other details of the computations **sufficient to permit replication**, as well as **information about access to data and programs**.

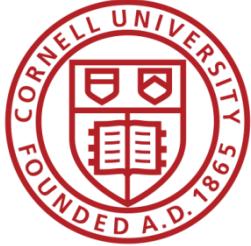


AEA DCAP (2018→2019)

- These will be **posted on the AEA website**. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the requirements above cannot be met.

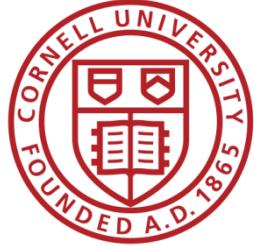


Data and programs should **be archived in the AEA Data and Code Repository**. Authors will **provide access** to editors and reviewers, if requested, **to both data and programs prior to acceptance**. The Editor should be notified at the time of submission if access to the data used in a paper is restricted or limited, or if, for some other reason, the requirements above cannot be met. **The AEA Data Editor will assess compliance with this policy, and will verify the accuracy of the information prior to acceptance by the Editor.**



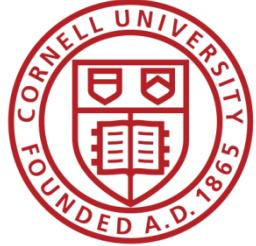
Current efforts at the AEA

- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information

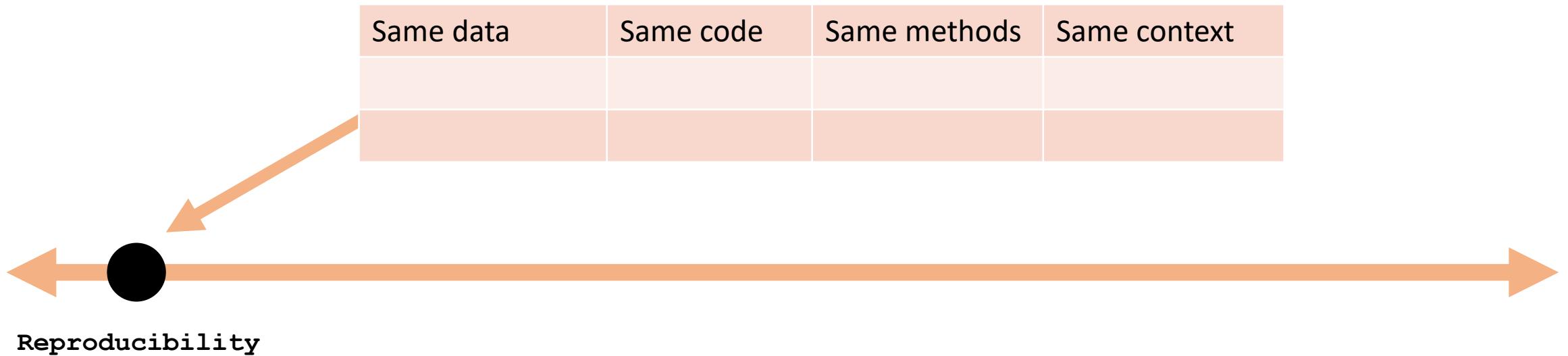


AEA Pre-Publication Verification

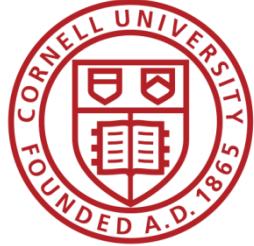
- Every paper that receives a “conditional acceptance” is verified
 - *Data citations*
 - *Quality of README*
 - *Quality of code*
 - *Reproducibility of code*
 - *Quality of metadata in the repository*



Replication continuum



- Narrow Replication (Pesaran 2003)
- Pure Replication (Hamermesh 2007)
- Verification (Clemens 2015)



Action: Reproducibility Check



Data and Code Guidance by Data Editors

Guidance for authors wishing to create data and code supplements, and for replicators.

Verification guidance

On this page:

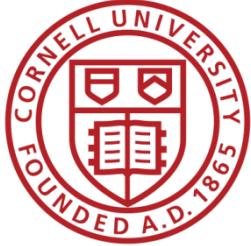
- Overview
- Review the README file
- For each listed data source
- For each listed table, figure, in-text number
- Conduct a code verification, if data is available
- Examples

Overview

This document describes

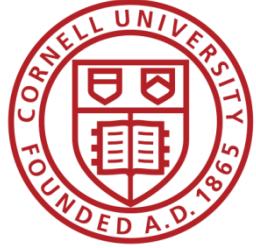
- what authors should check before providing data and code to journals
- what verifier teams should check for in the data and code provided to them for the purpose of verification



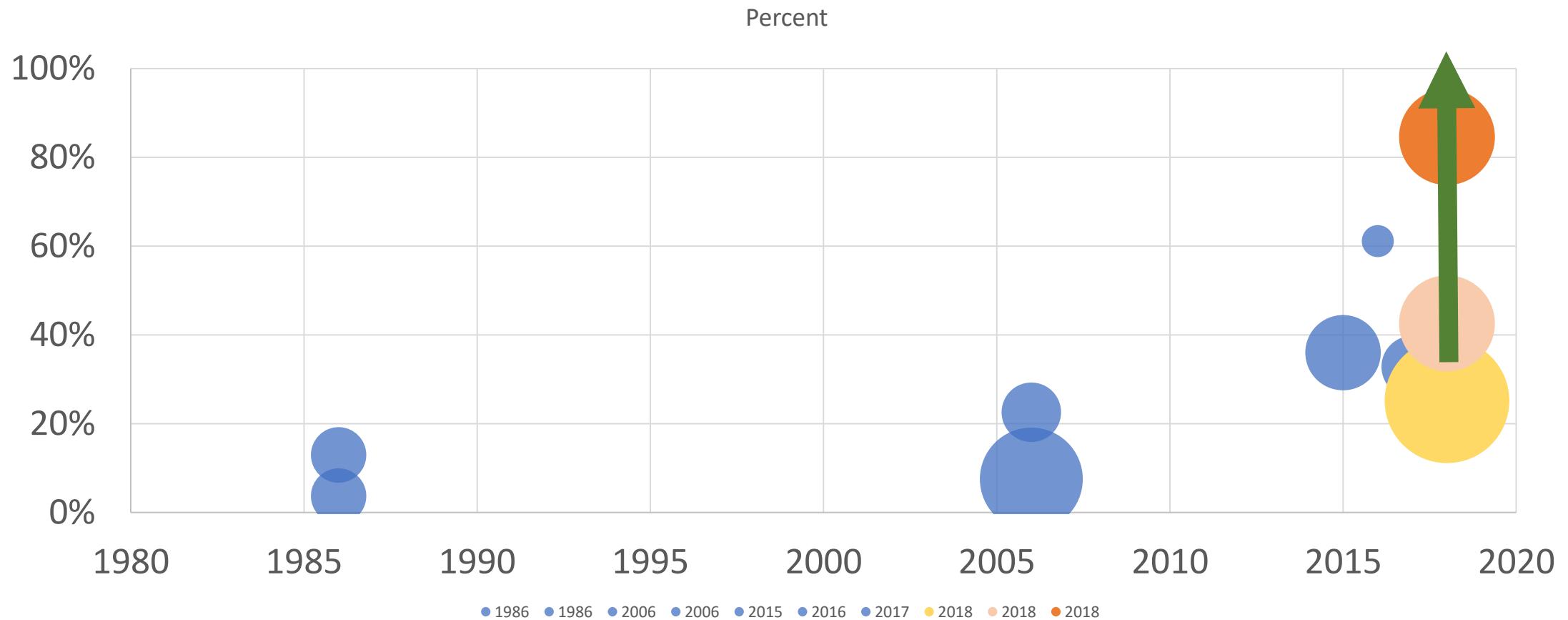


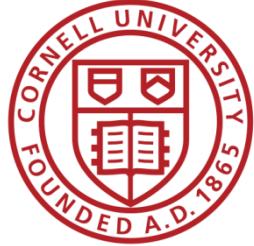
Who is doing that?

- Earlier reproducibility work: **Flavio Stanchi** (now at AirBnb), **Sylvie Herbert** (on the market), **Hautahi Kingi** (Impaq)
- Current lead graduate students: **David Wasser** (until Dec 2019), **Meredith Welch** (since Jan 2020)
- Current and past undergraduate students: Alexia Ge, Anthony Peraza, Craig Schulman, Elijah B. Ruiz, Gabriel Bond, Jason S. Katz, **Jeong Hyun Lee**, **Jiayin Song**, John Park, **Joshua Passel**, Kirubeal T. Wondimu, Linchen Zhang, **Louis Liu**, **Luis Lopez Cabrera**, Luke O'Leary, **Mary-Jo Ajiduah**, **Naomi Li**, Nicholas Swan, Nishat Peuly, **Ryan Ali**, Samuel Frey, Siyang (Elaine) Yu, **Steve Yeh**, **Weilun Shi**, William Hernandez, **Yanyun (Iris) Chen**, Yuan-Hsuan (Sharon) Lin, **Zebang Xu**, Xing Su, Jiazen Tan , Xueshi Su, Vendela Norman, Anderson Park, **Nehedin Juarez**, **Rubal Mistry**, **Syon Verma**, **William Silverman**, **Zechariah Karsana**
- Other graduate students: Aviv Caspi, Leah Kim



Goal: Improve reproducibility





Verifying Data and Code Deposits

- Check README
 - Legible? Intelligible? Complete?
- Check Code
 - Where is Table 1? Figure 1? Could this work?
- Check Access Rights
 - Can the author provide us with data?
 - Does the data access as described work?

AEA Data and Code Guidance



**AMERICAN
ECONOMIC
ASSOCIATION**

Guidance for authors wishing to create data and code supplements, and for replicators.

Unofficial guidance on various topics by the AEA Data Editor

These web pages provide unofficial and developing guidance on the implementation of the American Economic Association (AEA)'s Data and Code Availability Policy. We also provide links to [generic guidance](#) being developed by a loose collective ("guild") of data editors and people in a similar role at various social science journals.

Follow @aeadata

Order in which AEA authors should read these resources:

1. Start with the [official Data and Code Availability Policy](#)
2. Look for general guidance at the [Social Science Data Editors](#) pages
3. Read the [AEA's FAQ](#)
4. Look for any guidance specific to the AEA at the [Unofficial AEA Data and Code Guidance](#)
5. Last but not least, have a look at the [draft FAQ on this site](#)

Comments are welcome, please file them as [issues](#) in our Github repo.

Guidance on creating replicable data and program archives

How should researchers create replicable data and program archives? How

Code

Issues 1

Pull requests 0

Actions

Wiki

Security

Insights

Settings

Branch: master ▾

replication-template / REPLICATION.md

[Find file](#) [Copy path](#)

larsvilhuber Minor edits to the report - clarifications about data preparation pro...

e9ad1f8 10 days ago

1 contributor

195 lines (135 sloc) | 10.3 KB

[Raw](#)[Blame](#)[History](#)

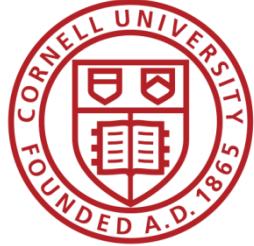
[MC number] [Manuscript Title] Validation and Replication results

INSTRUCTIONS: Once you've read these instructions, DELETE THESE AND SIMILAR LINES. In the above title, replace [Manuscript Title] with the actual title of the paper, and [MC number] with the Manuscript Central number (e.g., AEJPol-2017-0097) Go through the steps to download and attempt a replication. Document your steps here, the errors generated, and the steps you took to alleviate those errors.

You may want to consult [Unofficial Verification Guidance](#) for additional tips and criteria.

SUMMARY

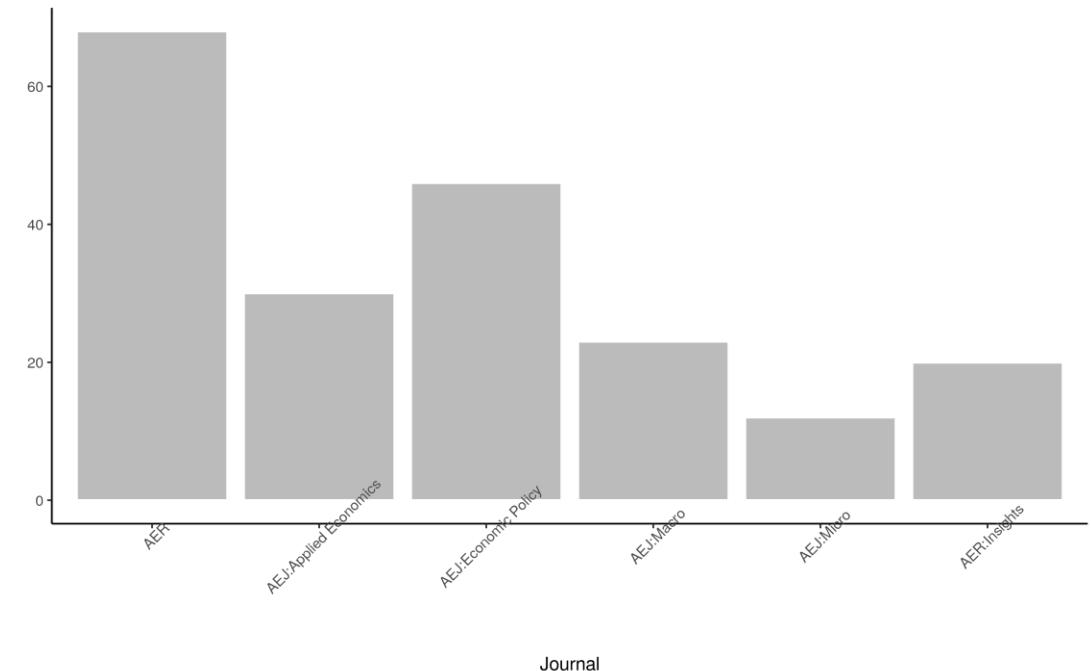
INSTRUCTION: The Data Editor will fill this part out. It will be based on any [REQUIRED] and [SUGGESTED] action

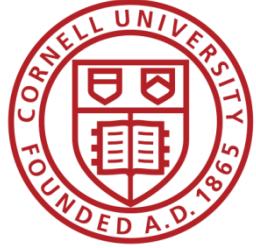


Stats on reproduced articles

Between July 16, 2019, and November 28, 2019 (4.5 mths), the AEA Data Editor team conducted

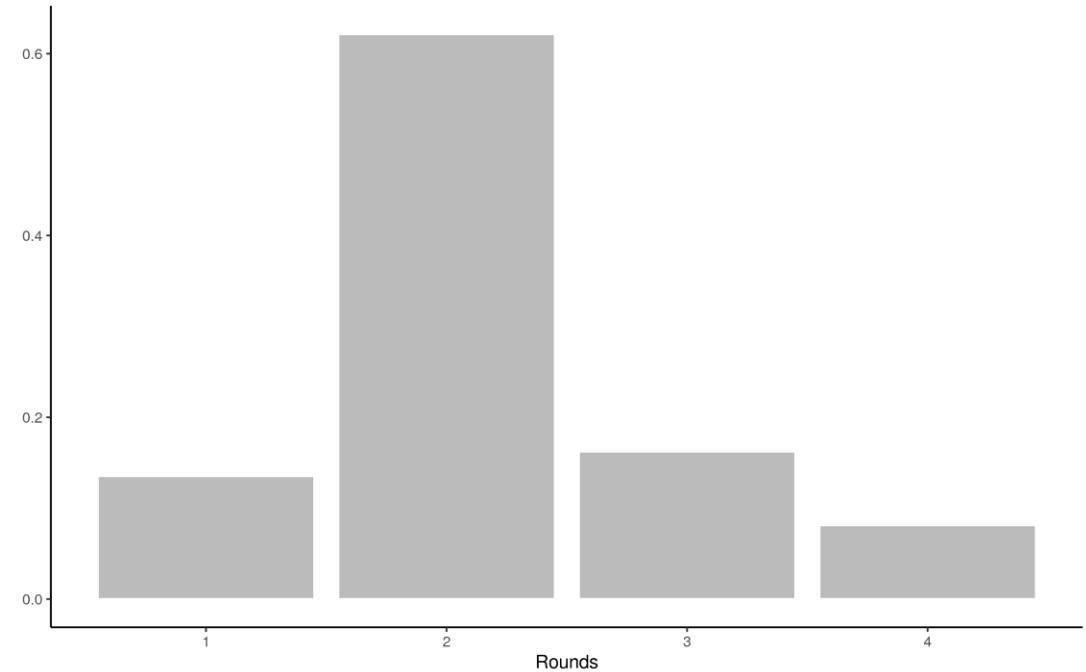
- **216 assessments**
- for **138 manuscripts**.
- (as of today, approx. 400)

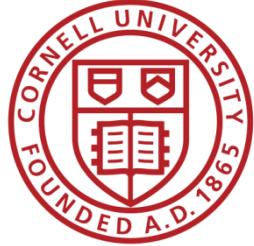




Stats on reproduced articles

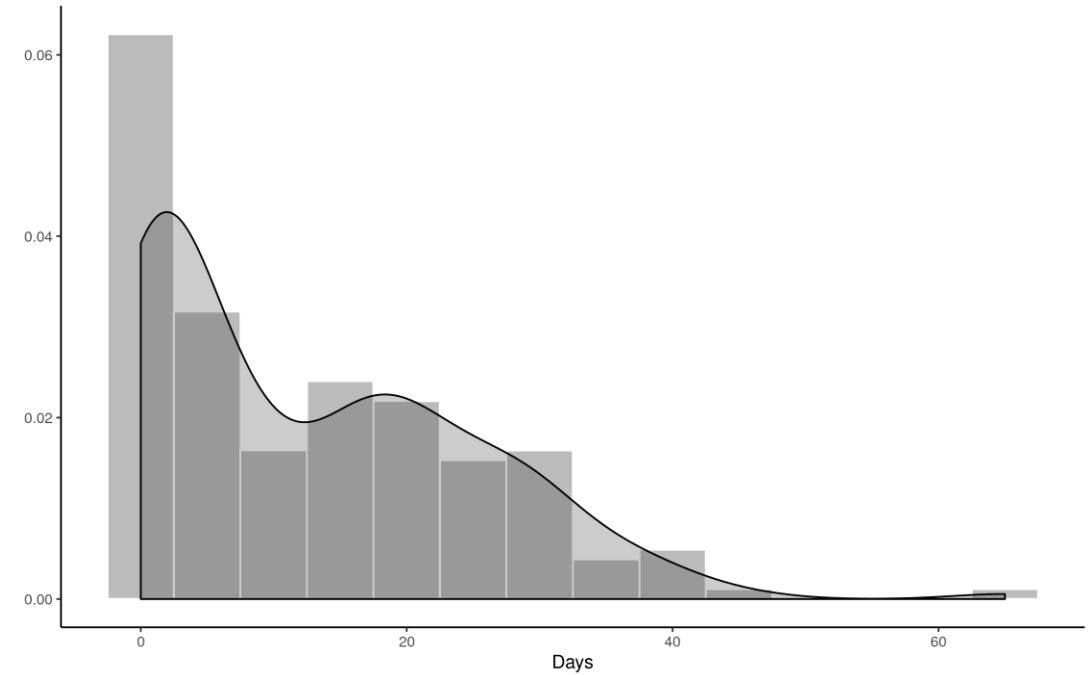
- The typical article goes through **at least two rounds** of assessment (none were perfect)

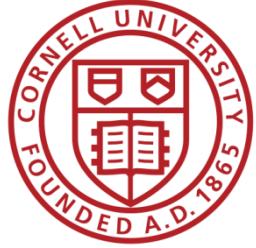




Stats on reproduced articles

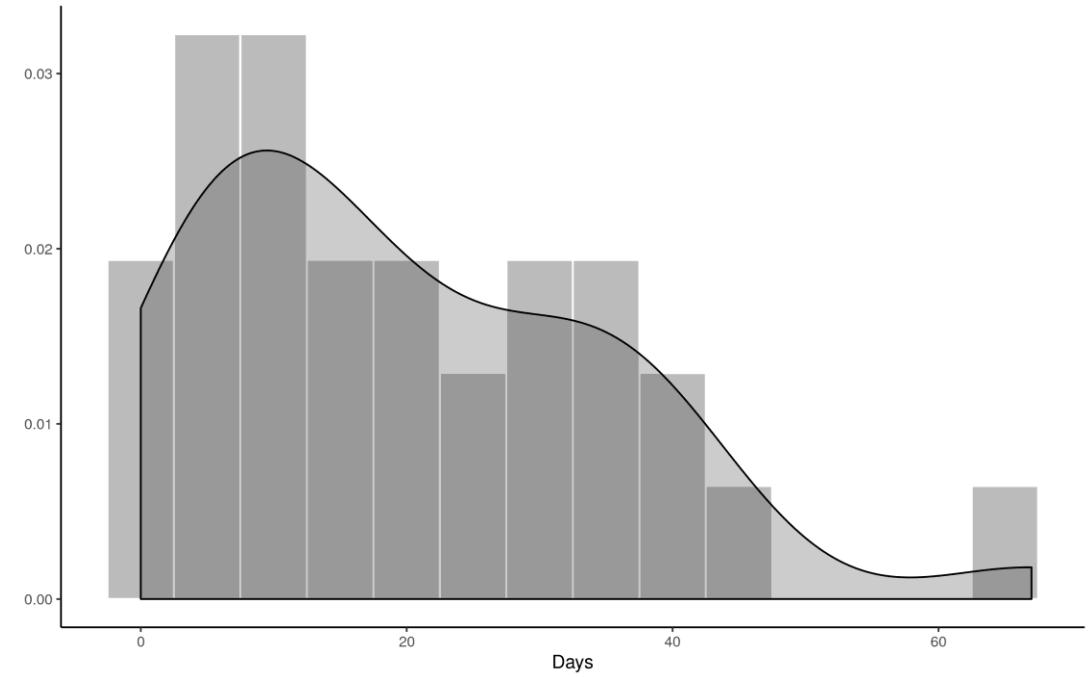
- Goal is turnaround of **two weeks**
- Currently still **too long**

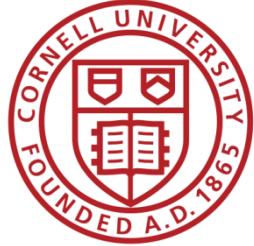




Stats on reproduced articles

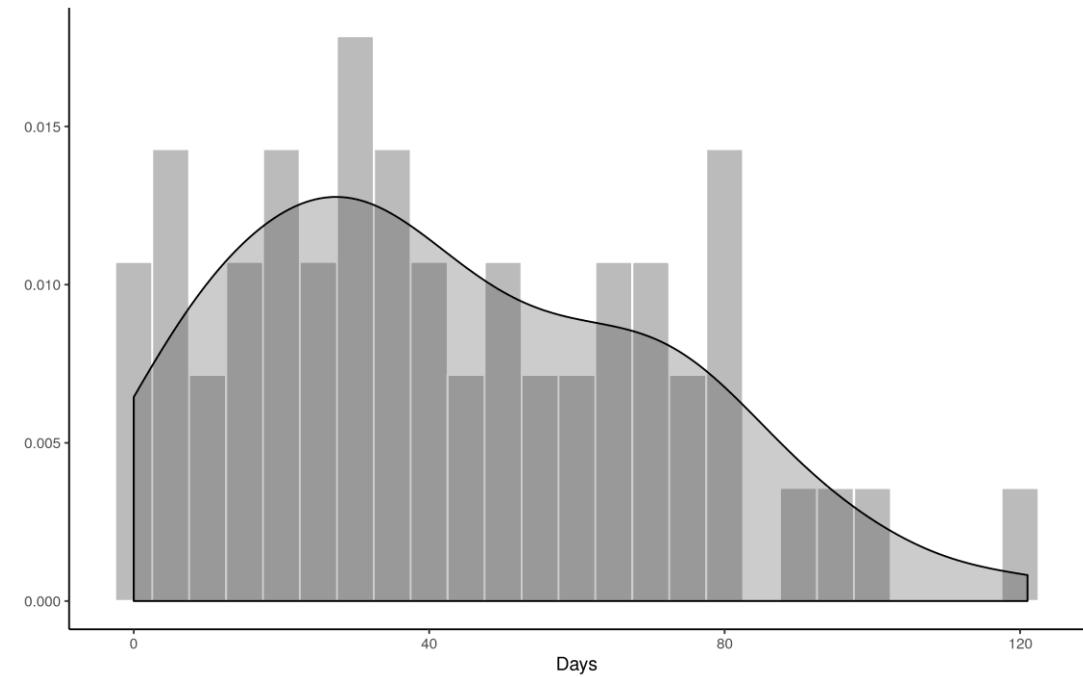
- But author response time is also a contributor

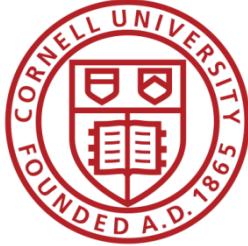




Stats on reproduced articles

- Total time (from first submission to final signoff) is **too long**





Increasing team size

- Grown from **7 undergrads** + 1 graduate assistant
- To **14 undergraduates**, **5 trainees**, + 1 graduate assistant (+ 1 volunteer)
- And:


cascad
the first certification agency for scientific code & data

A cascad certification allows researchers to signal the reproducibility nature of their research to their peers

CISER CORNELL INSTITUTE for Social and Economic Research



RESULTS REPRODUCTION (R-SQUARED)

Results Reproduction (R-Squared) is a service that computationally reproduces the Reproducibility and Transparency – think of it as *enhanced proofreading for your paper*.

HOME / ABOUT / NEWS /

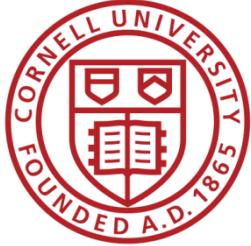
Announcing the Alexander and Diviya Magaro Peer Pre-Review Program at IQSS

January 10, 2019

The Institute for Quantitative Social Science is excited to announce the Alexander and Diviya Magaro Peer Pre-Review Program (PPR). PPR is designed to help IQSS-affiliated faculty improve scholarship before it becomes public, speed scientific discovery and publication, and reduce substantial inefficiencies for individual researchers.

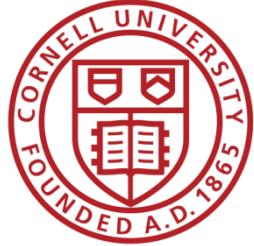
The process of turning a draft paper into a journal publication may take months or years through multiple rounds of often peer review.





Current efforts at the AEA

- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information



Full-featured repository

OPEN ICPSR [Find Data](#) [Share Data](#) [openICPSR Repositories ▾](#) [GO](#) [Sign Up](#) [Sign In](#)

 **AMERICAN
ECONOMIC
ASSOCIATION**

[AEA Deposit Instructions](#) [Browse AEA Deposits](#) [Contact](#)

Depositing Data in the AEA Data and Code Repository

The *American Economic Association journals* require authors to deposit data and materials with a community-recognized or general repositories. The *AEA Data and Code Repository at ICPSR* serves that purpose. Please see the AEA's [Data and Code Availability Policy](#) and data citation guidance at the [Sample References](#) page for more details. **Authors are required to include a citation pointing to the deposit in the reference section of the final version of the article sent to the AEA.** The openICPSR repository automatically generates a citation when the data are "published."

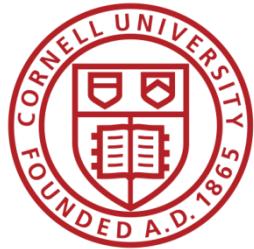
Deposits should include all data, annotated program code, command files, and documentation that is needed to replicate the findings from the authors' submitted article.

- **Data** should be comprehensively documented (see ICPSR's [Guide to Social Science Data Preparation and Archiving, 5th Edition](#) for guidance). The **author** is responsible for removing identifying information from the data to protect [confidentiality](#). Neither the AEA nor ICPSR review submissions for disclosure risk.
- **Program** code and command files should be annotated to facilitate replication and ensure clear correspondence between code and figures, tables, and analyses in the published article.
- Authors retain ownership and copyright to the data and code. Authors are required to affirm that they have the right to publish and redistribute the material. However,
 - ICPSR requires a license for distribution of data.
 - An **open license** is required by the AEA, in order to allow others to re-use the data and code, in particular for replication. Authors can select from several license options, including CC-BY 4.0 for data and Modified BSD for software and code. If an author would like to use multiple licenses or create a customized license, she should select the "Other" license option and upload a LICENSE file alongside the data and documentation.

By depositing in the AEA Data and Code Repository, the depositors allow the AEA staff to add keywords and other metadata which are important for proper indexing in linking. Any other changes are subject to the license chosen for the materials.

[View more extensive \(unofficial\) guidance.](#)

[Start Your Deposit](#)



FAIR data principles rely on metadata

— Scope of Project

Subject Terms ?

Do not copy/paste multiple terms into this field. Terms must be entered individually.

[✖ Russia](#) [✖ Industry](#) [✖ Factories](#) [✖ Russian Empire](#) [✖ Corporations](#)

JEL Classification ?

[✖ L20 General](#) [✖ N63 Europe: Pre-1913](#) [✖ O43 Institutions and Growth](#)

Manuscript Number ?

AER-2015-1656.R3 [edit](#) [remove](#)

Geographic Coverage ? [+ add value](#)

European Russia (Russian Empire) [edit](#) [remove](#)

Time Period(s) ? [+ add value](#)

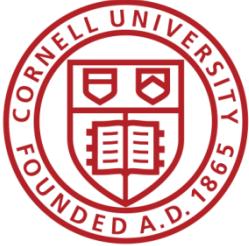
1894 – 1908 (Three years: 1894, 1900, and 1908) [edit](#) [remove](#)

Collection Date(s) ? [+ add value](#)

Universe ?

Manufacturing establishments in the European part of the Russian Empire. [edit](#) [remove](#)

Data Type(s) ?

[Find Data](#) / [Imperial Russian Factory Database, 1894-1908](#)

Imperial Russian Factory Database, 1894-1908

Principal Investigator(s): Amanda Gregg, Middlebury College

Version: V1



Name	File Type	Last Modified
1894MicroData.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	4.5 MB 08/08/2019 11:01:AM

Project Citation:

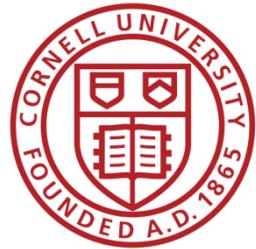
Gregg, Amanda. Imperial Russian Factory Database, 1894-1908. Nashville, TN: American Economic Association [publisher], 2020. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-01-29. <https://doi.org/10.3886/E110681V1>

AG_Corp_CleaningandDatabaseCompiler.do	text/x-stata-syntax	23.4 KB	08/08/2019 11:02:AM
--	---------------------	---------	---------------------

Related Publications

The following publications are supplemented by the data in this project.

- Gregg, Amanda. "Factory Productivity and the Concession System of Incorporation in Late Imperial Russia, 1894-1908." *American Economic Review* 110, no. 2 (February 2020): 401-27. <https://doi.org/10.1257/aer.20151656>.

[Find Data](#) / [Imperial Russian Factory Database, 1894-1908](#)

Imperial Russian Factory Database, 1894-1908

Principal Investigator(s): Amanda Gregg, Middlebury College

Version: V1

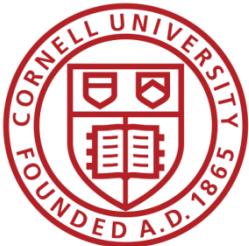


```
<meta name="DC.identifier" content="10.3886/E110681V1" />
<meta name="DC.title" content="Imperial Russian Factory Database, 1894-1908" />

<meta name="DC.creator" content="Amanda Gregg, Middlebury College" />

<meta name="DC.publisher" content="Inter-university Consortium for Political and Social Research (ICPSR)" />
<meta name="DC.date" content="2020-01-29" />
<meta name="DC.type" content="Dataset" />
```

			MB	
	officedocument.spreadsheetml.sheet			08:53:AM
	1908MicroData.xlsx	application/vnd.openxmlformats-officedocument.spreadsheetml.sheet	2.3 MB	08/07/2019 11:06:AM
	AG_Corp_CleaningandDatabaseCompiler.do	text/x-stata-syntax	23.4 KB	08/08/2019 11:02:AM
	AG_Corp_Prod_AppendixCode.do	text/x-stata-syntax	42.2 KB	12/09/2019 09:19:AM
	AG_Corp_Prod_Code.do	text/x-stata-syntax	26.6 KB	12/12/2019 03:01:AM
	AG_Corp_Prod_Database.dta	application/x-stata	11 MB	08/07/2019 08:55:AM
		application/x-stata	11.9	10/08/2014



[Find Data](#) / Imperial Russian Factory Database, 1894-1908

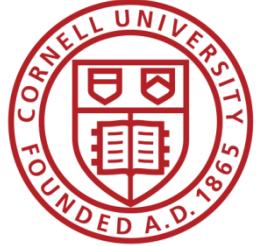
Imperial Russian Factory Database, 1894-1908

Principal Investigator(s):  Amanda Gregg, Middlebury College



```
<script type="application/ld+json">
  {"name": "Imperial Russian Factory Database, 1894-1908", "identifier": "http://doi.org/10.3886/E110681V1", "description": "This database digitizes manufacturing censuses. For each factory, the database includes industry, province, enterprise form, total workers, total revenue, and identifiers that span 908 years also include information on the factory's total machine power. The dataset was constructed to study why some Russian firms chose to become part of a consuming concession system. Note that the final analysis files exclude factories located outside of European Russia and, in the main data files, factories.", "url": "http://doi.org/10.3886/E110681V1", "version": "V1", "keywords": ["Russia", "Industry", "Factories", "Russian Empire", "Corporations"], "spatial": "Russia", "temporalCoverage": ["1894-01-01--1908-12-31 (Three years: 1894, 1900, and 1908)"], "creator": [{"name": "Amanda Gregg", "affiliation": ["Middlebury College"]}], "publisher": "openICPSR Self-Deposit Archive", "url": "http://www.openicpsr.org/", "@type": "DataCatalog"}, "funder": [{"name": "Economic History Association", "url": "http://www.eh.net/", "@type": "Organization"}, {"name": "Yale University Department of Sociology, Anthropology, and Anthropology", "url": "http://soc.yale.edu/", "@type": "Organization"}, {"name": "Yale Economic Growth Center", "url": "http://yaleegc.yale.edu/", "@type": "Organization"}, {"name": "Yale MacMillan Center", "url": "http://macmillan.yale.edu/", "@type": "Organization"}], "fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_Prod_Database.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/download/terms?path=/openicpsr/110681/fcr:versions/V1/AG_Corp_RuscorpMasterFile_Cleaned.dta&type=application/x-stata", "encodingFormat": "application/zip"}, {"fileFormat": "stata", "contentURL": "https://www.openicpsr.org/openicpsr/project/110681/version/V1/AG_Corp_Prod_Database_withAktsiz.dta&type=application/x-stata", "encodingFormat": "application/zip"}], "license": "https://creativecommons.org/licenses/by/4.0", "@context": "http://schema.org", "@type": "Dataset"}
</script>
```

	AG_Corp_CleaningandDatabaseCompiler.do		KB	11:02:AM
	AG_Corp_Prod_AppendixCode.do	text/x-stata-syntax	42.2 KB	12/09/2019 09:19:AM
	AG_Corp_Prod_Code.do	text/x-stata-syntax	26.6 KB	12/12/2019 03:01:AM
	AG_Corp_Prod_Database.dta	application/x-stata	11 MB	08/07/2019 08:55:AM
		application/x-stata	11.9 KB	10/08/2014 09:45:AM



... and findability relies on metadata

Google



Imperial Russian Factory



▼ Updated Date

▼ Download Format

▼ Usage Rights

Free

2 datasets found



Imperial Russian Factory

Database, 1894-1908

www.openicpsr.org

www.da-ra.de

Updated Jan 29, 2020



Middlebury

Imperial Russian Factory Database, 1894-1908

[Explore at openICPSR Self-Deposit Archive](#)

[Explore at www.da-ra.de](#)

Unique identifier

<https://doi.org/10.3886/E110681V1>

Dataset updated Jan 29, 2020

Dataset provided by

[Middlebury College](#)

Authors

Amanda Gregg

License

[Attribution 4.0 \(CC BY 4.0\)](#)

License information was derived automatically

Area covered

European Russia (Russian Empire)

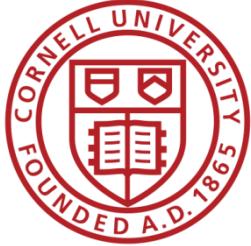


Data from: Антиосманские
выступления болгар и русско-
турецкие войны второй...
explore.openaire.eu
Updated 24 нояб. 2015 г.



Not seeing a result you expected?
[Learn](#) how you can add new
datasets to our index.

Oct 11, 2019



Migrating archives

- Between Oct 11 and Oct 13, 2019, the staff at openICPSR ingested **2,552 historical AEA supplements**, increasing the size of the openICPSR repository **by a factor of 3**.
- The migrated archives are now available through the **openICPSR search interface**, the **general ICPSR search interface**, as well as through a variety of federated search interfaces such as **Google Dataset Search**.



Filters

Recency	
New/updated this week	0
New/updated this month	0
New/updated this quarter	1
New/updated this year	16

Access Level	
Public Use	16

Search Results

Showing 1 - 16 of 16 results.

esther duflo

GO

VIEW ALL

[search tips ▾](#)

Summaries: Hidden

Sort by: Study Relevance ▾

Study Title/Investigator

1. [Replication data for: \(Dis\)organization and Success in an Economics MOOC](#)
Banerjee, Abhijit V.; Duflo, Esther
2. [Replication data for: Bundling Health Insurance and Microfinance in India: There Cannot Be Adverse Selection If There Is No Demand](#)
Banerjee, Abhijit; Duflo, Esther; Hornbeck, Richard
3. [Replication data for: Education, HIV, and Early Fertility: Experimental Evidence from Kenya](#)
Duflo, Esther; Dupas, Pascaline; Kremer, Michael
4. [Replication data for: Incentives Work: Getting Teachers to Come to School](#)
Duflo, Esther; Hanna, Rema; Ryan, Stephen P.
5. [Replication data for: Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya](#)
Duflo, Esther; Kremer, Michael; Robinson, Jonathan
6. [Replication data for: Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya](#)
Duflo, Esther; Dupas, Pascaline; Kremer, Michael
7. [Replication data for: Up in Smoke: The Influence of Household Behavior on the Long-Run Impact of Improved Cooking Stoves](#)
Hanna, Rema; Duflo, Esther; Greenstone, Michael
8. [Replication data for: Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco](#)
Crépon, Bruno; Devoto, Florencia; Duflo, Esther; Parienté, William

FIND DATA

SEARCH/COMPARE VARIABLES

DATA-RELATED PUBLICATIONS

RESOURCES FOR STUDENTS

HELP

Filters

Subject Terms ▾

- child health (1)
- community health (1)
- education (1)
- employment (1)
- families (1)

[view all](#)

Restriction Type ▾

Data Format ▾

Collection Method ▾

Data Type ▾

Time Period ▾

Recent Releases ▾

Funding Agency ▾

Search Results

Showing 1 - 17 of 17 results.

esther duflo

GO

VIEW ALL

[search tips ▾](#)[Studies \(17\)](#)[Variables \(0\)](#)[Series \(1\)](#)[Data-related Publications \(1\)](#)[ICPSR Website \(0\)](#)Summaries: Hidden

Sort by: Study Relevance ▾

Study Title/Investigator

Released/Updated

1. [Replication data for: \(Dis\)organization and Success in an Economics MOOC ↗](#) OPEN
ICPSR 2019-10-11

Banerjee, Abhijit V.; Duflo, Esther

2. [Replication data for: Bundling Health Insurance and Microfinance in India: There Cannot Be Adverse Selection If There Is No Demand ↗](#) OPEN
ICPSR 2019-10-11

Banerjee, Abhijit; Duflo, Esther; Hornbeck, Richard

3. [Replication data for: Education, HIV, and Early Fertility: Experimental Evidence from Kenya ↗](#) OPEN
ICPSR 2019-10-12

Duflo, Esther; Dupas, Pascaline; Kremer, Michael

4. [Replication data for: Incentives Work: Getting Teachers to Come to School ↗](#) OPEN
ICPSR 2019-10-11

Duflo, Esther; Hanna, Rema; Ryan, Stephen P.

5. [Replication data for: Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya ↗](#) OPEN
ICPSR 2019-10-11

Duflo, Esther; Kremer, Michael; Robinson, Jonathan

6. [Replication data for: Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya ↗](#) OPEN
ICPSR 2019-10-11

[▼ Updated Date](#)[▼ Download Format](#)[▼ Usage Rights](#)[Free](#)

54 datasets found



Women Empowerment and Child Health in Developing Countries - Continuation 2

www.icpsr.umich.edu

Updated May 14, 2015



Women Empowerment and Child Health in Developing Countries - Continuation 2

[Explore at www.icpsr.umich.edu](http://www.icpsr.umich.edu)

Unique identifier

<https://doi.org/10.3886/ICPSR35890.v1>**Dataset updated** May 14, 2015

Dataset provided by

[Inter-university Consortium for Political and Social Research](#)

Authors

Duflo, Esther

License

<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/35890/terms>

Area covered

Ghana, Kenya



Replication data for: Marry for What? Caste and Mate Selection in Modern...

[www.openicpsr.org
search.datacite.org](http://www.openicpsr.org/search.datacite.org)
+1more

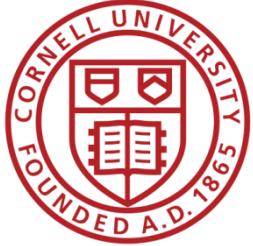
Updated Aug 30, 2012



Replication data for: Education, HIV, and Early Fertility: Experimental...

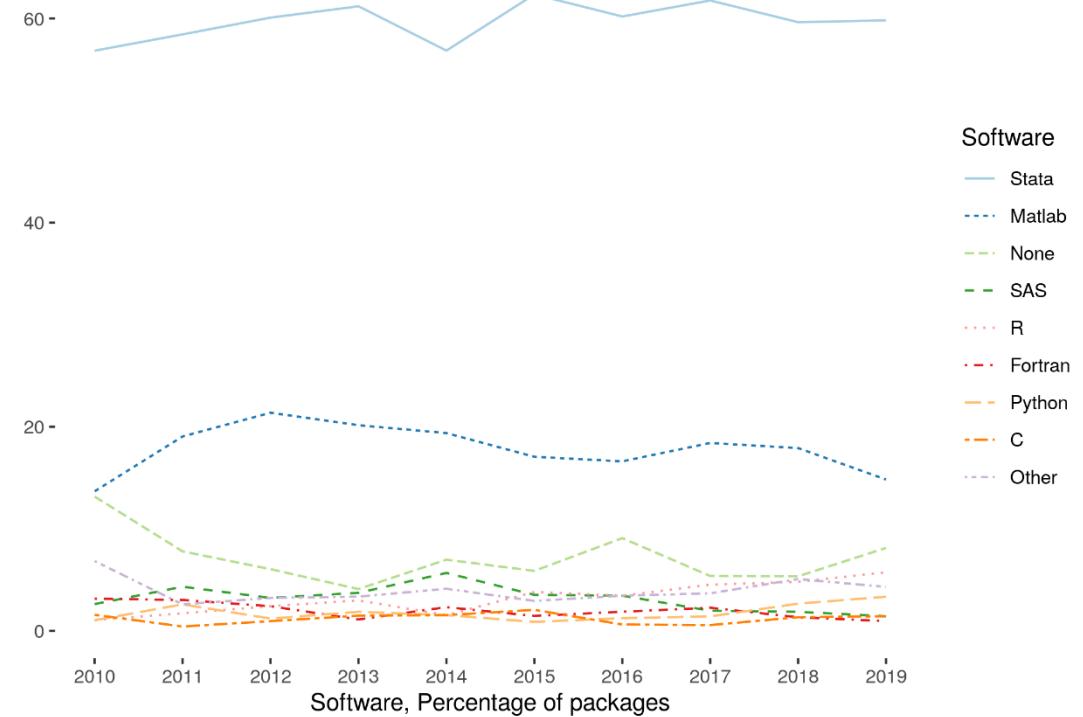
www.openicpsr.org

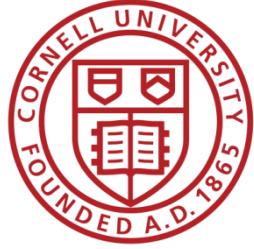
Updated Oct 12, 2019



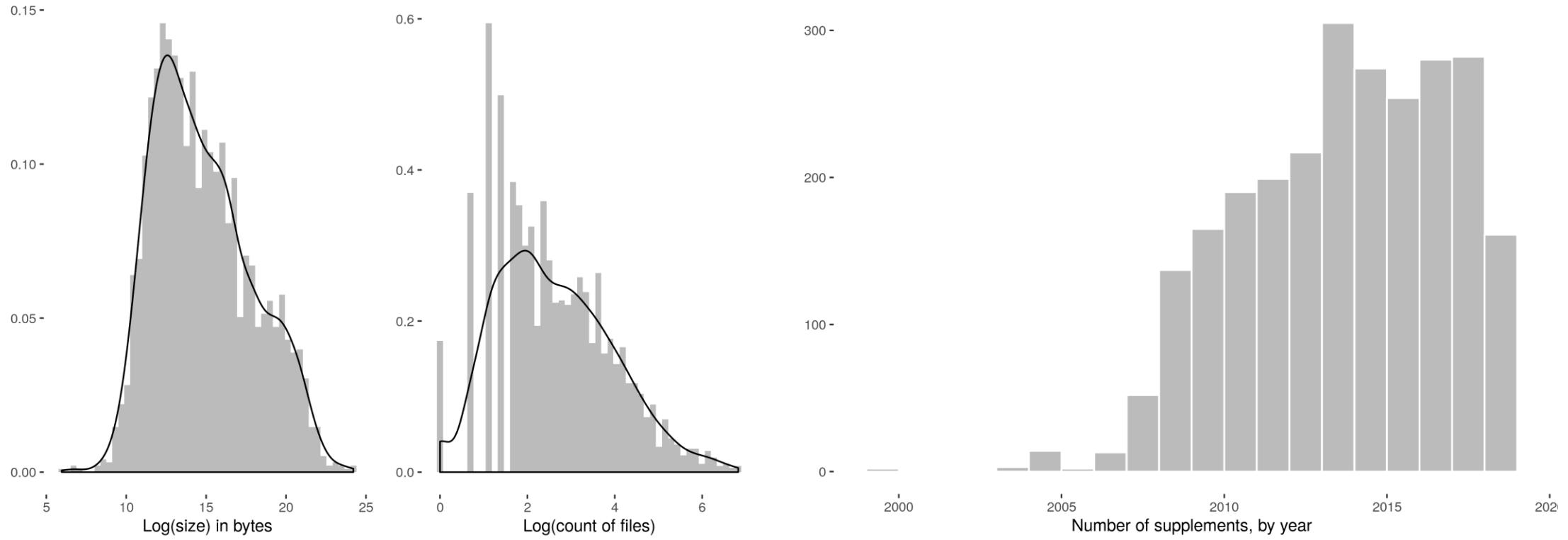
Some statistics

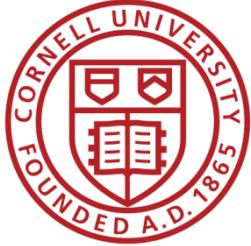
- **Stata** is the most popular statistical software in the journals of the AEA
(72.96% of all supplements)
- followed by **Matlab** (**22.45%**)





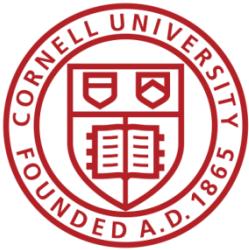
More stats





Current efforts at the AEA

- **Pre-emptively improve code archives**
 - By conducting reproducibility checks when we can
 - By working with groups that conduct reproducibility checks when we cannot
- **Better archives**
 - Greater transparency of the code and data archives
- **Better provenance tracking**
 - Leave code where it is when appropriate
 - Leave data where it is almost always
 - Display that information



perceived criteria of importance.

1. Importance

Data should be considered legitimate, citable products of research. Data should be accorded the same importance in the scholarly record as citation research objects, such as publications[1].



Data Citation Principles

2. Credit and Attribution

1 | **Bureau of Labor Statistics.** 2000–2010. “Current Employment Statistics: Colorado, Total Nonfarm, Seasonally adjusted - SMS080000000000000001.” United States Department of Labor. <http://data.bls.gov/cgi-bin/surveymost?sm+08> (accessed February 9, 2011).

corresponding data should be cited[3].

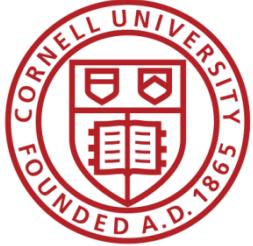
4. Unique Identification

A data citation should include a persistent method for identification that is actionable, globally unique, and widely used by a community[4].

5. Access

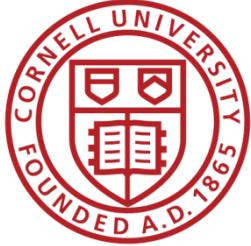
Data citations should facilitate access to the data themselves and to such metadata, documentation, code, and other materials as are necessary for

Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [<https://www.force11.org/group/joint-declaration-data-citation-principles-final>].



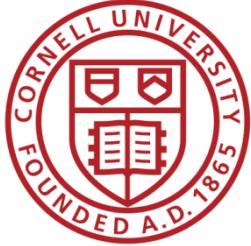
AEA “Data Availability Policy” (2019)

- It is the policy of the American Economic Association to publish papers only if the data used in the analysis are **clearly and precisely documented and access to the data and code is clearly and precisely documented and is non-exclusive to the authors.**
- Authors of accepted papers that contain empirical work, simulations, or experimental work must **provide, prior to acceptance**, the data, programs, and other details of the computations **sufficient to permit replication**, as well as **information about access to data and programs.**



Every manuscript is checked

- What datasets are used
- Are they cited?
 - → in Article?
 - → in Online Appendix?
 - → in README?

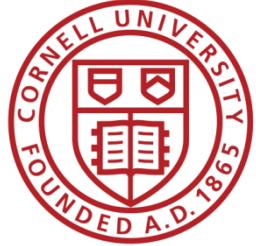


Every manuscript is checked

- What datasets are used
- Are they cited?
- Is there additional info
access?
 - → URL leads to exact d
 - → URL leads to applica
 - → other access proced



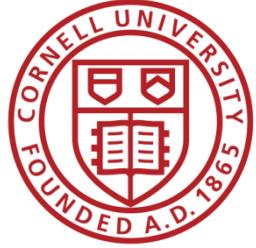
Le labyrinthe des juges, par [Guillaume Beck](#)



Every manuscript is checked

- What datasets are used
- Are they cited?
- Is there additional information on access?
- Is there license/ data use information?
 - → Should the author provide the data?
 - → Is the author allowed to provide data?

Proposed:
Explicit DAS or
Incorporate into README

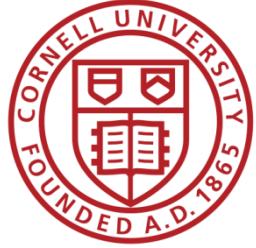


Evolving Journal and Data Infrastructure

Authors struggle with
this!

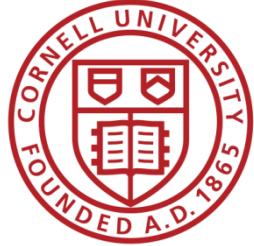
The author contacted the data provider, and received an answer similar to this one:

Hi John Doe, The simple answer is no, we do not want this data to be released publicly. Doing so would require revisiting the data use agreement and another round of Privacy Impact Assessment for internal approval. Regards,
Your Data Provider



Evolving Journal and Data Infrastructure

Data Publishers struggle with this!



Example 1: OECD

URL does not always change!

The screenshot shows the OECD Stat website at <https://stats.oecd.org/Index.aspx>. A blue arrow points from the text "URL does not always change!" to the browser's address bar.

1. Gross domestic product (GDP) ⓘ

Please refer to the dataset **Gross domestic product (GDP), 2019 archive** to access longer time series based on the new reference year.

Please note that **OECD reference year from 2010 to 2015 changed on Tuesday 3rd of December, 2019.**

Transaction

B1_GA: Gross domestic product (output approach)

B1G_P119: Gross value added at basic prices, excluding FISIM

B1G: Gross value added at basic prices, total activity

B1GVA: Agriculture, forestry and fishing (ISIC rev4)

B1GVB_E: Industry, including energy (ISIC rev4)

B1GVB_E: Industry, including energy (ISIC rev4)

B1GVC: of which construction (ISIC rev4)

B1GVF: Construction (ISIC rev4)

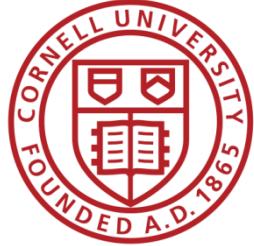
B1GVG_I: Distributive trade, repairs; transport (ISIC rev4)

B1GVJ: Information and communication (ISIC rev4)

B1GVK: Financial and insurance activities (ISIC rev4)

B1GVL: Real estate activities (ISIC rev4)

B1GVM_N: Prof., scientific, techn.; admin., sup rev4)

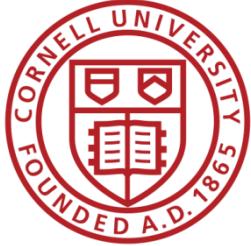


Example 1: OECD

URL does not always change!
(but sometimes it does)

The screenshot shows the OECD.Stat website interface. At the top, there is a navigation bar with back, forward, refresh, and home icons. The URL in the address bar is <https://stats.oecd.org/Index.aspx?DatasetCode=STLABOUR>. Below the address bar, the OECD logo is displayed. On the left side, there is a sidebar with a search bar containing "our labour" and a dropdown menu. The menu includes sections like "Data by theme", "Popular queries", "in Themes", and "Reset". Under "Labour Force Statistics", there are "Short-Term Statistics" and "Short-Term Labour Market Statistics". The "Short-Term Labour Market Statistics" section is highlighted with a yellow background. The main content area is titled "Short-Term Labour Market Statistics". It features filters for "Subject" (Employment rate, Aged 15-64, All persons), "Measure" (Level, rate or quantity series, s.a.), "Frequency" (Quarterly), "Unit" (Q4-2016, Q1-2017, Q2-2017, Q3-2017), and "Time" (Q4-2016, Q1-2017, Q2-2017, Q3-2017). The "Country" section lists several countries with their corresponding employment rates: Australia (72.3, 72.4, 72.8), Austria (71.7, 71.8, 72.2), Belgium (63.3, 62.4, 62.9), Canada (72.9, 73.3, 73.4), Chile (62.4, 62.4, 62.6), Czech Republic (72.7, 73.1, 73.3), Denmark (72.0, 72.0, 72.0), France (72.0, 72.0, 72.0), Germany (72.0, 72.0, 72.0), Greece (72.0, 72.0, 72.0), Hungary (72.0, 72.0, 72.0), Ireland (72.0, 72.0, 72.0), Italy (72.0, 72.0, 72.0), Japan (72.0, 72.0, 72.0), Korea (72.0, 72.0, 72.0), Luxembourg (72.0, 72.0, 72.0), Mexico (72.0, 72.0, 72.0), Netherlands (72.0, 72.0, 72.0), Norway (72.0, 72.0, 72.0), Portugal (72.0, 72.0, 72.0), Spain (72.0, 72.0, 72.0), Sweden (72.0, 72.0, 72.0), Switzerland (72.0, 72.0, 72.0), Turkey (72.0, 72.0, 72.0), United Kingdom (72.0, 72.0, 72.0), and United States (72.0, 72.0, 72.0).

Country	Q4-2016	Q1-2017	Q2-2017	Q3-2017
Australia	72.3	72.4	72.8	
Austria	71.7	71.8	72.2	
Belgium	63.3	62.4	62.9	
Canada	72.9	73.3	73.4	
Chile	62.4	62.4	62.6	
Czech Republic	72.7	73.1	73.3	
Denmark	72.0	72.0	72.0	
France	72.0	72.0	72.0	
Germany	72.0	72.0	72.0	
Greece	72.0	72.0	72.0	
Hungary	72.0	72.0	72.0	
Ireland	72.0	72.0	72.0	
Italy	72.0	72.0	72.0	
Japan	72.0	72.0	72.0	
Korea	72.0	72.0	72.0	
Luxembourg	72.0	72.0	72.0	
Mexico	72.0	72.0	72.0	
Netherlands	72.0	72.0	72.0	
Norway	72.0	72.0	72.0	
Portugal	72.0	72.0	72.0	
Spain	72.0	72.0	72.0	
Sweden	72.0	72.0	72.0	
Switzerland	72.0	72.0	72.0	
Turkey	72.0	72.0	72.0	
United Kingdom	72.0	72.0	72.0	
United States	72.0	72.0	72.0	

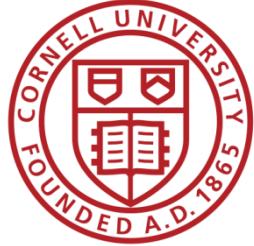


Example 1: OECD

URL does not always change!
(and then it doesn't...)

The screenshot shows the OECD.Stat website at the URL <https://stats.oecd.org/Index.aspx?DatasetCode=STLABOUR>. The page displays 'Short-Term Labour Market Statistics' for 'Unemployed Population'. The search interface on the left shows the query 'Unemployed Population' has been entered. The results table lists data for various countries across different time periods (Q2-2017, Q3-2017, Q4-2017, Q1-2018) and includes columns for Subject, Measure, Frequency, Unit, and Time.

Country	Q2-2017	Q3-2017	Q4-2017	Q1-2018
Australia	721	718	717	
Austria	249	245	242	
Belgium	363	355	319	
Canada	1 275	1 215	1 185	
Chile	595	581	609	
Czech Republic	164	147	132	
Denmark	174	177	160	
Estonia	48	38	38	
Iceland				



Example 2: Academic data publisher

 **ECONOMIC POLICY UNCERTAINTY**

Home Methodology Media Research & Applications About Us

EPU Indices

[All Country-Level Data](#)

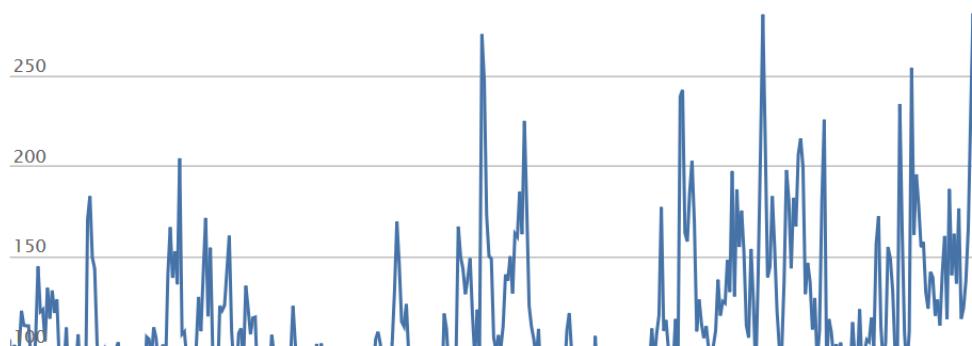
Global	USA
Australia	Brazil
Canada	Chile
China	Colombia
Croatia New	France
Germany	Greece
Hong Kong	India
Ireland	Italy
Japan	South Korea

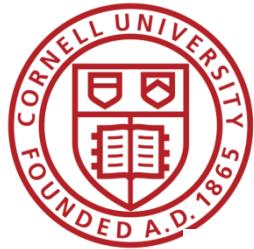
Economic Policy Uncertainty Index

We develop indices of economic policy uncertainty for countries around the world.

Monthly US Economic Policy Uncertainty Index

Zoom [1m](#) [3m](#) [6m](#) [1y](#) [7y](#) [All](#)





Example 2: Academic data publisher

https://www.policyuncertainty.com/index.html SEP DEC JAN
103 captures 14 2018 2019 2020
18 Aug 2012 - 14 Dec 2019

ECONOMIC POLICY UNCERTAINTY [Home](#) [Methodology](#) [Media](#) [Research & Applications](#) [About Us](#)

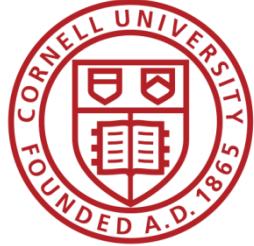
EPU Indices **Economic Policy Uncertainty Index**

All Country-Level Data We develop indices of economic policy uncertainty for countries around the world.

Globe Australia Canada China Croatia France Germany Greece Hong Kong India Ireland Italy Japan South Korea

© 2012-2018 by Economic Policy Uncertainty

The chart displays a highly volatile line graph representing the Economic Policy Uncertainty Index. The y-axis ranges from 100 to 200, and the x-axis spans from approximately 2012 to 2018. The index fluctuates significantly, with major peaks around 2013, 2015, and 2017, and troughs around 2014, 2016, and 2018.



Example 2: Academic data publisher-new!

 **ECONOMIC POLICY UNCERTAINTY**

[Home](#) [Methodology](#) [Media](#) [Research & Applications](#) [About Us](#)

[EPU Indices](#)

All Country-Level Data

Global [USA](#)

[Monthly US Economic Policy Uncertainty Index](#)

We develop indices of economic policy uncertainty for countries around the world.

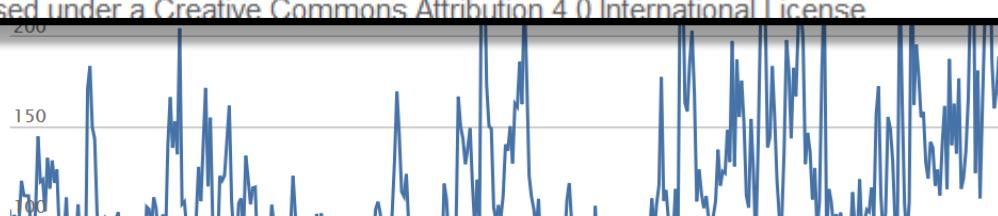
This work is licensed under a Creative Commons Attribution 4.0 International License

Germany [Greece](#)

[Hong Kong](#) [India](#)

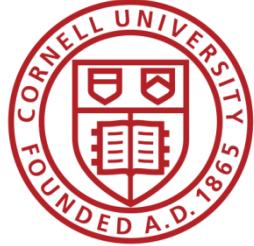
[Ireland](#) [Italy](#)

[Japan](#) [South Korea](#)



Response

Thanks, but I'll stick with what I've been doing for at least 20 years. At some point I might figure out the right license, but it's been working so far. And your inference is correct, the authors can use the data but not redistribute it. In this specific case, there is no reason for them to do so because the data are freely available to everyone.



Example 3: FRED (St. Louis Fed)

FRED ECONOMIC RESEARCH
FEDERAL RESERVE BANK OF ST. LOUIS

MY ACCOUNT

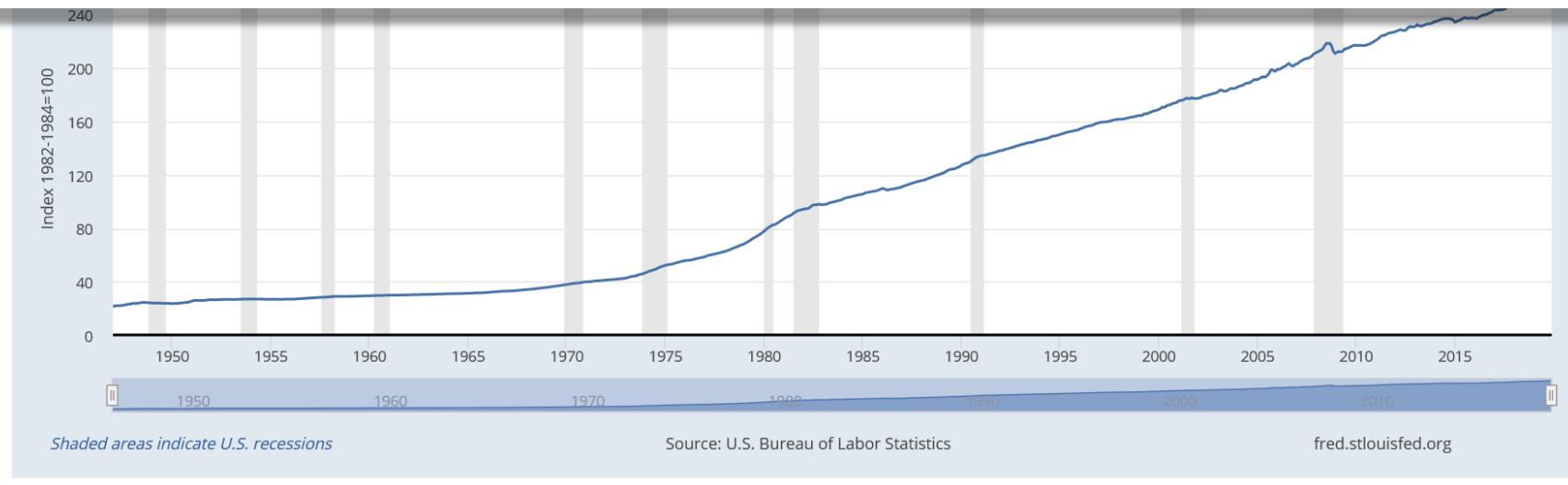
Search FRED

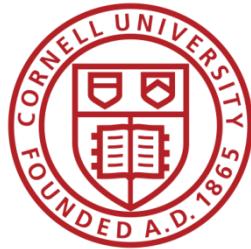
FRED® Economic Data Information Services Publications Working Papers Economists About St. Louis Fed Home

Categories > Prices > Consumer Price Indexes (CPI and PCE)

Suggested Citation:

U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCSL], retrieved from FRED, Federal Reserve Bank of St. Louis; <https://fred.stlouisfed.org/series/CPIAUCSL>, February 3, 2020.





Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13
BHP 7517 v1	10.5164/IAB.BHP7517.de.en.v1	2018-12-12
BHP 7516 v1	10.5164/IAB.BHP7516.de.en.v1	2018-04-11

External data

Data Archive

Data Access

Campus Files

Publications

Events

Projects of FDZ users

FDZ Projects

Complaint point of the
RatSWD

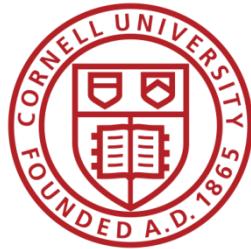
Figures of the FDZ

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13



Example 4: German Restricted-access



RESEARCH DATA CENTRE (FDZ)
of the German Federal Employment Agency (BA)
at the Institute for Employment Research (IAB)

[Home](#) | [Newsletter](#) | [Jobs](#) | [Contact](#) | [Data Privacy](#) | [Imprint](#)



Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13
BHP 7517 v1	10.5164/IAB.BHP7517.de.en.v1	2018-12-12
BHP 7516 v1	10.5164/IAB.BHP7516.de.en.v1	2018-04-11

External data

Data Archive

Data Access

Campus Files

Publications

Events

Projects of FDZ users

FDZ Projects

Complaint point of the
RatSWD

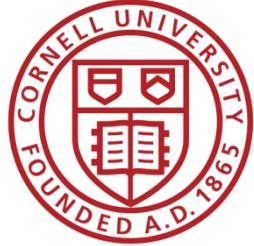
Figures of the FDZ

employees, both in total and broken down by gender, age, occupational status, qualification and nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Data Versions

Old versions are only available for replication studies and only in justified exceptional cases for new Projects.

Data Version	DOI (Link to Description of Data Version)	Availability (yyyy-mm-dd)
BHP 7518 v1 (current)	10.5164/IAB.BHP7518.de.en.v1	2020-01-13



Example 4: German Restricted-access

Establishment History Panel (BHP) – Version 7518 v1

DOI: 10.5164/IAB.BHP7518.de.en.v1

Summary

Data source:

Data Access

The IAB Establishment Panel is available via the following ways of access:

- On-site use at the FDZ. Further information on Applying for [on-site use](#).
- Remote data Access. Further information on Applying for [remote data access](#).

nationality. Means and medians of wages for full-time employees are given, too. Additional datasets providing information about (gross) worker flows and about foundations and closures of establishments are available on request.

Dataset Descriptions and Frequencies

German

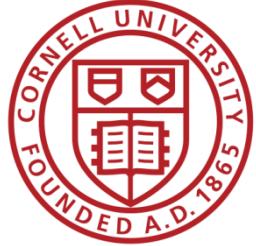
- DOI: [10.5164/IAB.FDZD.2001.de.v1](https://doi.org/10.5164/IAB.FDZD.2001.de.v1)
- [FDZ-Datenreport 01/2020](#)
- [Fallzahlen und Labels](#)

English

- DOI: [10.5164/IAB.FDZD.2001.en.v1](https://doi.org/10.5164/IAB.FDZD.2001.en.v1)

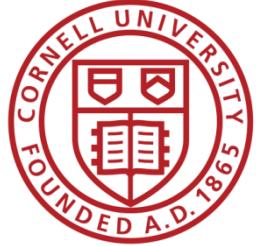
Some Suggestions

For authors



Action: Encourage Best Practices

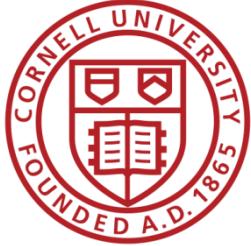
- **Follow robust coding**
 - Ensure that code reliably produces results
(possibly automated)
 - Before you finish the manuscript, run all analysis code again
(if not too onerous)



Reproducibility is harder than it should be

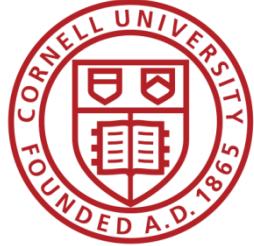
- Often done piecemeal
 - At different times
 - By different people
- Software versions
 - Stata 9? 15? 42?
 - rdrobust 2014? 2016? 2018 bug fix?
- Compilers and exotic software

For institutions



Better support for researchers

- Training in methods (with various centers, institutions, etc.)
 - For current researchers (Carpentries, custom, etc.)
 - For integration into curriculums
- Tools to streamline the process
 - A few technical things (not described here)
 - Coordinate among journals (no duplicate effort)
- Awareness
 - Consider badges/ certification
 - Address issues with confidential data



Verification services

 **cascad**
*the first certification
agency for scientific
code & data*

A cascad certification allows researchers to signal the reproducibility nature of their research to their peers

The screenshot shows the homepage of the CASD (Secure Data Hub) website. At the top, there is a dark header bar with the CASD logo on the left, a menu icon (three horizontal lines), and two navigation links: "PROJETS" and "DONNÉES" on the right. Below the header, the page has a dark purple background with white text. It features the title "Secure Data Hub" next to a network-like icon. Below this, there are four main categories: "Travail, Emploi / 189 projets", "Société, Justice, Éducation / 113 projets", "Économie, Entreprises, Finance / 267 projets", and "Environnement, Agriculture / 187 projets". At the bottom, it says "Santé / 244 projets".

= **CASD** • PROJETS DONNÉES

Secure Data Hub

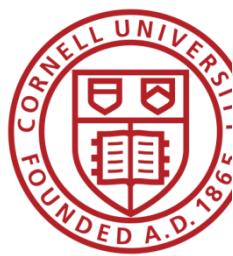
Travail, Emploi / 189 projets

Société, Justice, Éducation / 113 projets

Économie, Entreprises, Finance / 267 projets

Environnement, Agriculture / 187 projets

Santé / 244 projets



Cornell University

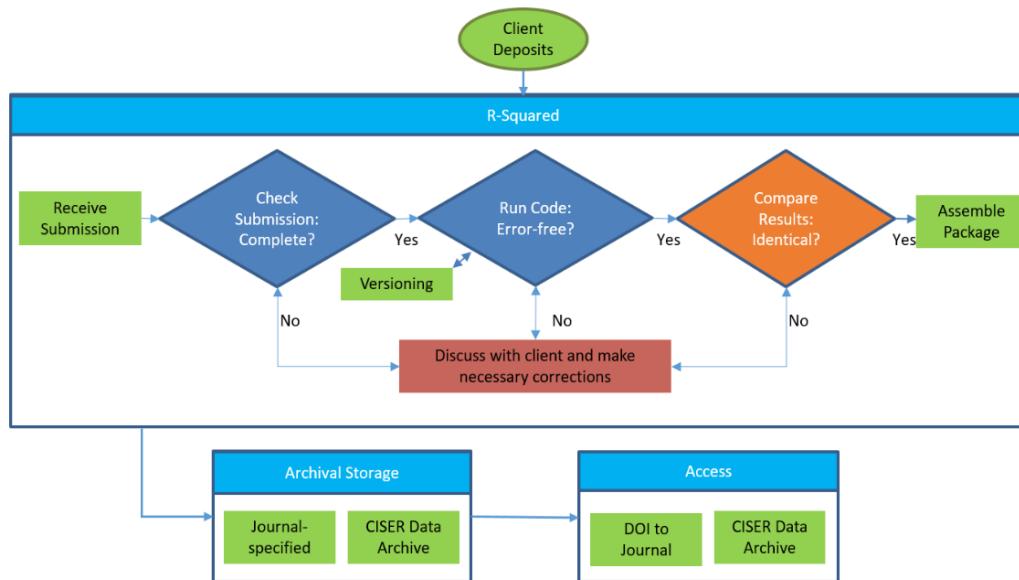
CISER

CORNELL INSTITUTE for
Social and Economic Research

Home > Research > **Results Reproduction (R-squared)**

RESULTS REPRODUCTION (R-SQUARED)

Results Reproduction (R-Squared) is a service that computationally reproduces the results of your research to ensure Reproducibility and Transparency – think of it as *enhanced proofreading for your Data and Code*.



HARVARD UNIVERSITY

People

IQSS

The Institute for Quantitative Social Science

About ▾ Programs & Products ▾ Research

HOME / ABOUT / NEWS /

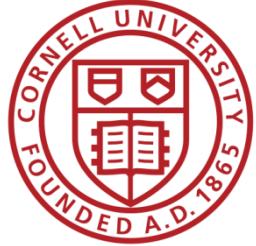
Announcing the Alexander and Diviya Magaro Peer Pre-Review Program at IQSS

January 10, 2019

The Institute for Quantitative Social Science is excited to announce the Alexander and Diviya Magaro Peer Pre-Review Program (PPR). PPR is designed to help IQSS-affiliated faculty improve scholarship before it becomes public, speed scientific discovery and publication, and reduce substantial inefficiencies for individual researchers.



For journals



Predation, Protection, and Productivity: A Firm-Level Perspective.



Abstract



References



Online appendix



Supplementary
materials



Notes

Supplementary materials

- Code and Data

Besley, Timothy, and Hannes Mueller. 2018. "Replication data for: Predation, Protection, and Productivity: A Firm-Level Perspective." American Economic Association [publisher] DOI: 10.1257/mac.20160120.data

cite!

- Data is freely accessible under CC BY-NC 4.0 at [10.1257/mac.20160120.data](https://doi.org/10.1257/mac.20160120.data).

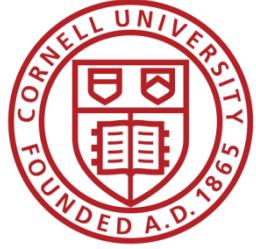
- Data

Statistics Norway. 2015. "Firm-level statistics 1975-2013 [dataset]" Norwegian Data Archive [curator], v2. DOI: 10.7654/nda::7643A::34

cite!

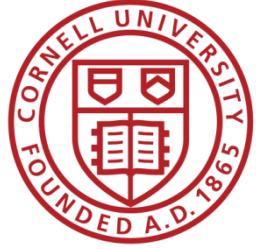
- Data restricted-access, under Norwegian Data Access license (has residency requirement, has citizenship requirement), accessible at [Norwegian Data Archive in Oslo, Norway](#)

Collaboration
and
Standardization



Goal: Transportability

Any standards, tools, methods: must be transportable across journals (no custom solutions)

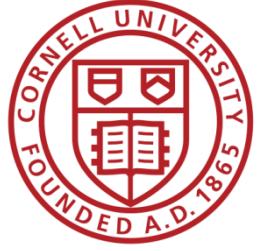


Social science “guild”

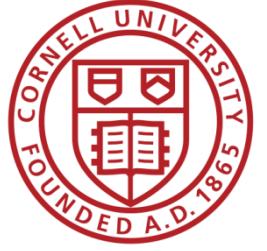


[https://
social-science
-data-editors.
github.io/
guidance/](https://social-science-data-editors.github.io/guidance/)

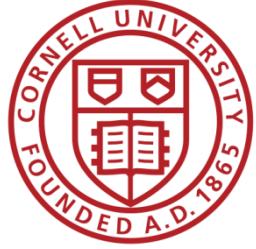
Challenges?



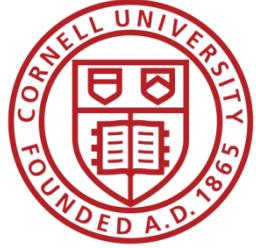
You...



Me...



Change ingrained habits...



File Edit View Repository Branch Help

Current repository **replicability-presentation2019**

Current branch **master** Last fetched just now

Changes History 0 changed files

No local changes

You have no uncommitted changes in your repository! Here are some friendly suggestions for what to do next.

Pull 1 commit from the origin remote
The current branch (`master`) has a commit on GitHub that does not exist on your machine.
Pull origin
Always available in the toolbar when there are remote changes or `Ctrl Shift P`

Open the repository in your external editor
Configure which editor you wish to use in [options](#)
Repository menu or `Ctrl Shift A` **Open in Atom**

View the files in your repository in Explorer
Repository menu or `Ctrl Shift F` **Show in Explorer**

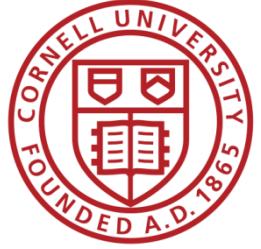
Open the repository page on GitHub in your browser
Repository menu or `Ctrl Shift G` **View on GitHub**

Summary (required)

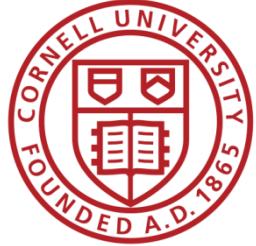
Description

Commit to **master**

A screenshot of a GitHub desktop application window. The window title is "replicability-presentation2019". The main area displays a message: "No local changes" with a sub-message: "You have no uncommitted changes in your repository! Here are some friendly suggestions for what to do next." Below this are three suggestions: "Pull 1 commit from the origin remote", "Open the repository in your external editor", and "View the files in your repository in Explorer". At the bottom of the window, there is a summary section with a "Description" field and a "Commit to master" button.



New skills to learn...



Shape Your D

Join us at DockerCon 2019,
things Kubernetes, microser

Turn a Git repo into a collection of
notebooks

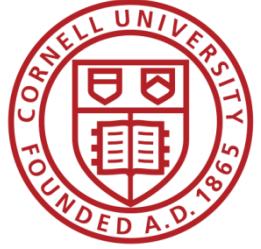
The collage consists of four distinct sections:

- Docker:** A blue banner with the Docker logo and the text "Shape Your D". Below it, a snippet of text reads "Join us at DockerCon 2019, things Kubernetes, microser".
- Why Docker:** A screenshot showing a "Why Docker" section with a "Private Untitled Capsule" interface. It displays a file tree with "environment", "code", and "data Manage Datasets". Below the tree are icons for "Upload" and "Start with Sample Files".
- R Markdown:** A screenshot of an R Markdown document titled "R Markdown" from R Studio. The document content includes "from R Studio", a preview of a plot titled "Diversity gradient", and the Python logo.
- Python in R Studio:** A screenshot of a Python code editor in R Studio. The code is a Fibonacci series generator:

```
n = 3: Fibonacci series up to n
fib(n):
    a, b = 0, 1
    while a < n:
        print(a, end=' ')
        a, b = b, a+b
```

Below the collage, two large blue boxes contain promotional text:

- Welcome to RStudio**
Do, share, teach and learn da
- (The second box is partially cut off)



New methods to use ...

Search



UQAM > Archipel

Delete

Répertoires Facultés Bibliothèques

Université de Montréal

New upload

Instructions: (i) Upload minimum one file or fill-in the form below.

Files

import requests

```
>>> r = requests.get("https://zenodo.org/api/deposit/depositions")  
>>> r.status_code
```

401

```
>>> r.json()
```

{

```
    "message": "The server could not verify that you are authorized to access the URL on this server."
```

Accueil

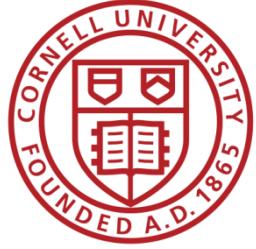
Recherche

Add a file

MON COMPTE

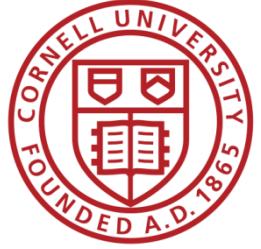
Ouvrir une session

Nouvel utilisateur?



We!

- Ingrained habits
- New skills to learn
- New methods to use

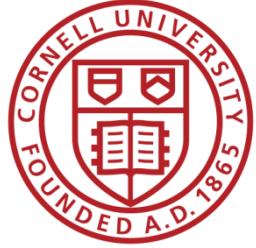


Push for better support...

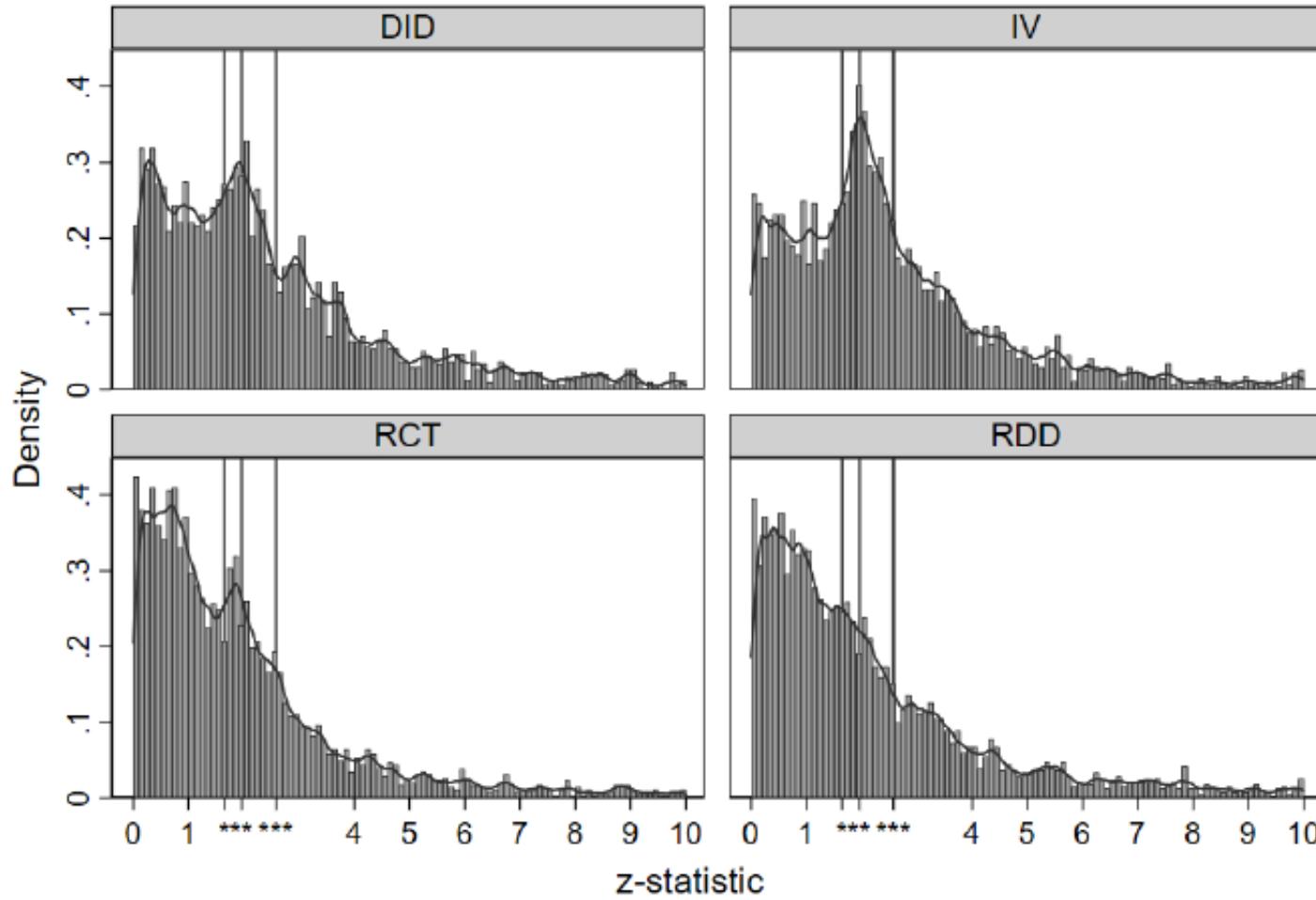
Thank you!

DOI: to come

Extra slides

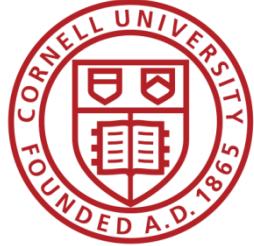


Other challenges remain: Publication bias



Graphs by Method

Abel Brodeur (2018),
<https://osf.io/hg9an/>



Correcting for publication bias

