Reproducibility as A Goal in Research

```
library(“dramatic_entrance_music”)

# Leafia Ye
# PhD Candidate in Sociology

presentation() +
theme_data_nerd()
```

Acknowledgement

I drew ideas and examples from the following sources:

- Gelman, A., & Loken, E., 2013. “The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time.” link
- Alston and Rick, 2020. “A Beginner’s Guide to Conducting Reproducible Research” link
- Goldman & Andrasfay, 2022. “Life expectancy loss among Native Americans during the COVID-19 pandemic” link

Outline

1. What is reproducibility?
2. Steps to quantitative research that should be reproducible
3. What does reproducible work look like?
4. Five reasons why reproducible research is a goal
   - From the most selfish reasons to the least selfish ones

What is reproducibility?

- Obtaining consistent results using the same data and same code
  - If someone else downloads your data, your code, and follows your instructions, they should be able to:
    1. Get all figures and tables in your paper by clicking “Run”
    2. Understand what’s going on in the file

- Reproducibility is different from **replicability** – the latter means obtaining consistent results across studies aimed at answering the same question - but different data and code

Steps to quantitative research using secondary data

1. Have a research question

   "How large is the loss in life expectancy in 2020 relative to 2019 by race & gender?"

2. Download + clean the data

   CDC WONDER (Wide-ranging Online Data for Epidemiologic Research) – race, gender, age, and year-specific mortality rates

3. Run the analysis

   Calculate life expectancies at birth and at age 65 by race & gender in 2020 and 2019

4. Stylize the output

   Visualize life expectancy loss by race & gender in a figure

5. Write the paper

   Type type type eat potato chips type type

Steps to quantitative research using secondary data

1. Have a research question

2. Download + clean the data
3. Run the analysis
4. Stylize the output

} Steps 2-4 need to be reproducible

5. Write the paper

What does reproducible work look like?

Demographic Research: Volume 47, Article 9
Descriptive Finding

Life expectancy loss among Native Americans during the COVID-19 pandemic

Noreen Goldman[1]
Theresa Andrasfay[2]

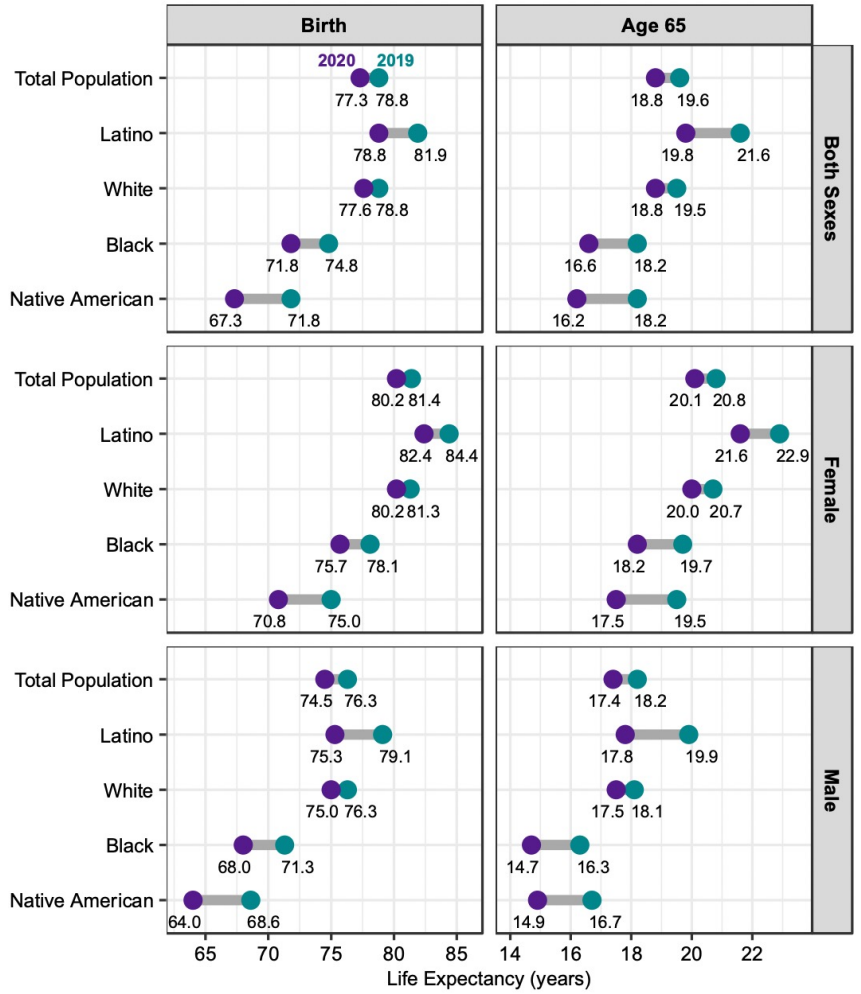
Abstract

BACKGROUND
There has been little systematic research on the mortality impact of COVID-19 in the Native American population.

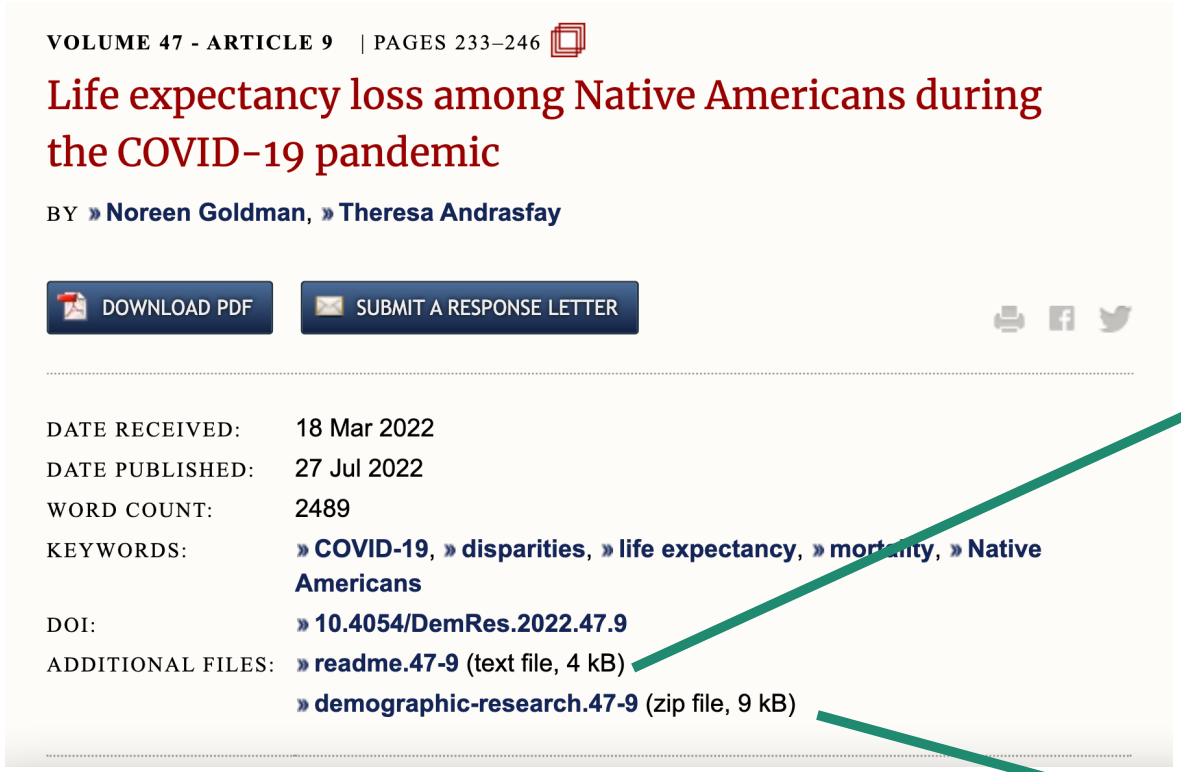
OBJECTIVE
We provide estimates of loss of life expectancy in 2020 and 2021 relative to 2019 for the Native American population.

https://dx.doi.org/10.4054/DemRes.2022.47.9

Figure 2: **Life expectancy at birth and at age 65 by race/ethnicity, estimates for 2020 compared to 2019**



What does reproducible work look like?



VOLUME 47 - ARTICLE 9 | PAGES 233–246

Life expectancy loss among Native Americans during the COVID-19 pandemic

BY » Noreen Goldman, » Theresa Andrasfay

DOWNLOAD PDF SUBMIT A RESPONSE LETTER

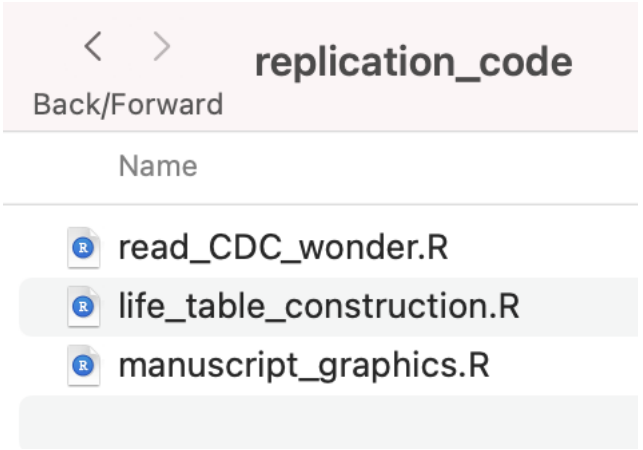
DATE RECEIVED:	18 Mar 2022
DATE PUBLISHED:	27 Jul 2022
WORD COUNT:	2489
KEYWORDS:	» COVID-19, » disparities, » life expectancy, » mortality, » Native Americans
DOI:	» 10.4054/DemRes.2022.47.9
ADDITIONAL FILES:	» readme.47-9 (text file, 4 kB)
	» demographic-research.47-9 (zip file, 9 kB)

```
Replication Files

1) read_CDC_wonder.R
   This file reads in the txt files obtained from CDC WONDER containing
the AI/AN mortality data for 2019-2021
   Details on the CDC WONDER queries are provided in this file for those
wishing to download the data themselves.

2) life_table_construction.R
   This file takes the output of file (1) and creates lifetables for
2019-2021 for the AI/AN population.

3) manuscript_graphics.R
   This file takes the output of file (2) and creates manuscript
graphics.
```

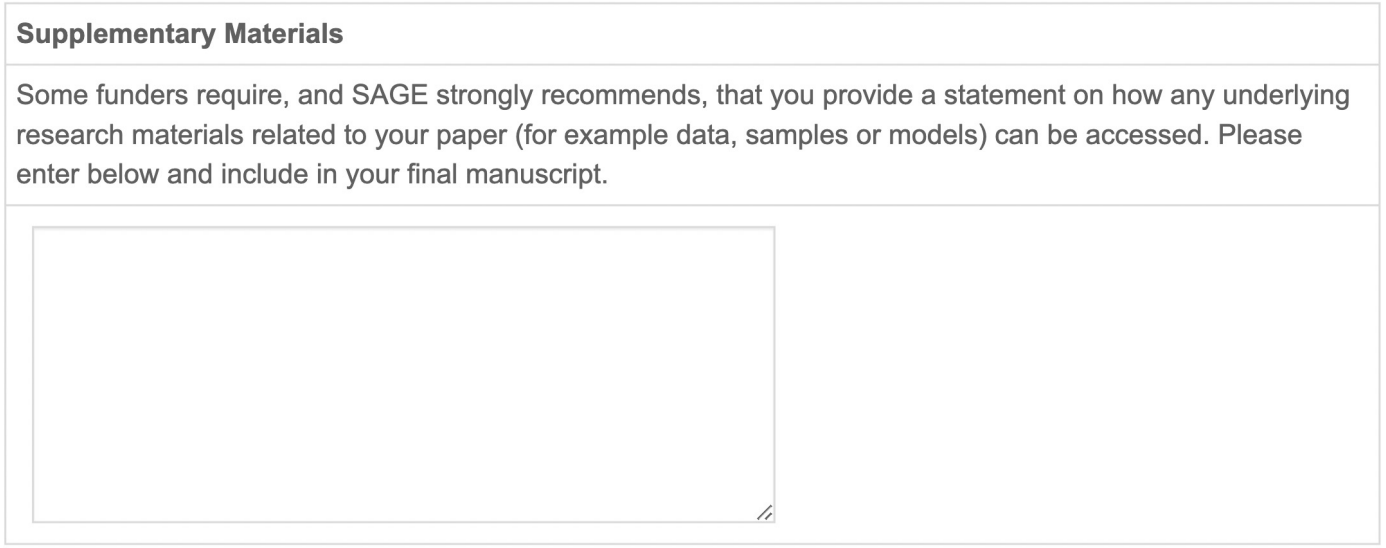


Reproducible research makes our lives easier

- Research takes a long time. Reproducible work helps **future you** pick up what you did
  - It's like writing down your favorite recipe in detail so you can make it again five months later!
- Allows your collaborators to read code easily and add their contribution

Reproducible research is sometimes required

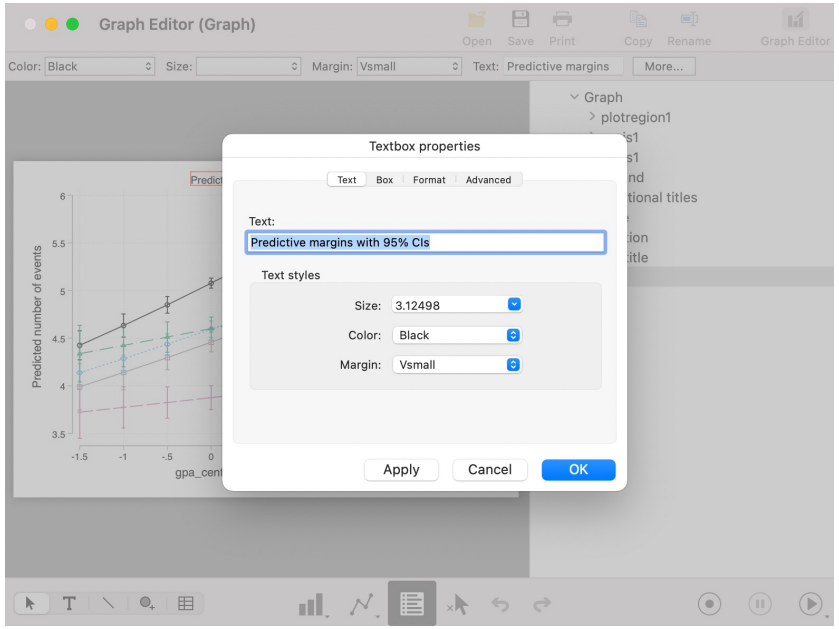
- Some journals will ask you to submit materials for reproducing your work



From the submission portal of *Socius: Sociological Research for a Dynamic World*

Reproducible research makes us better coders

- Pushes you to write code for every step, instead of manually pasting and calculating values / editing the aesthetics of a graph



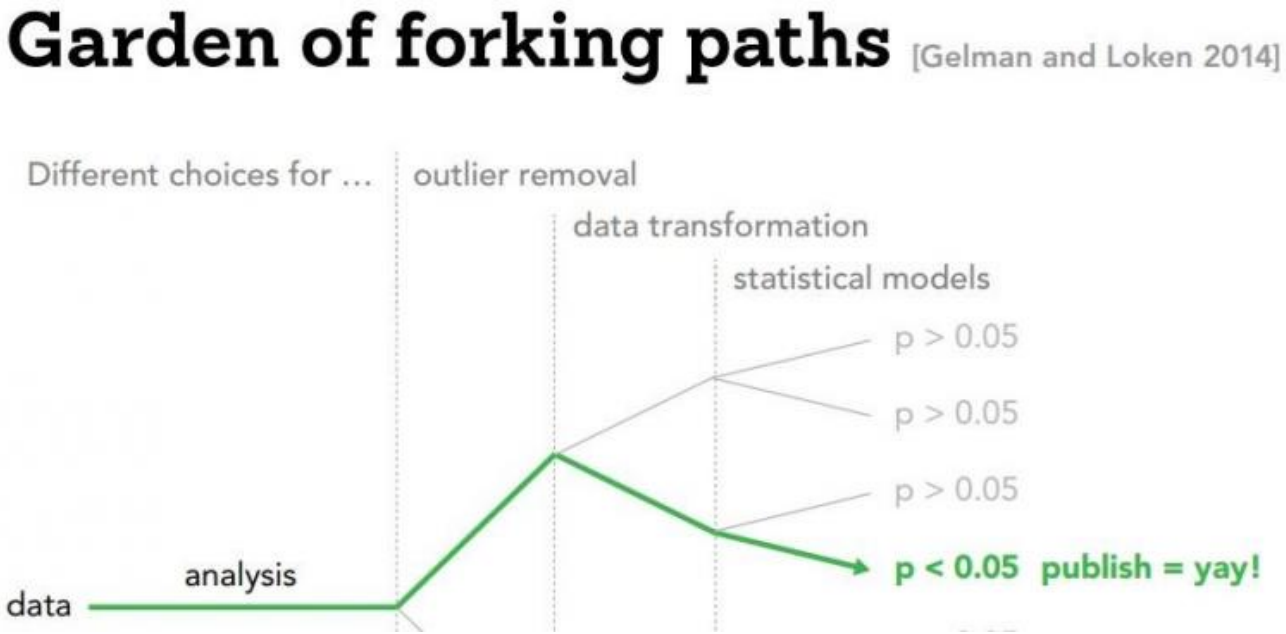
Decisions made using graph editor are *not* saved; if anything changes, need to remake graph and start all over

```
margins, over(raceimm5) at (gpa_centered=(-1.5(.5)1.5))
marginsplot, ytitle("Predicted level of popularity") ///
xtitle("Grade Point Average") ///
title(Predicted level of popularity by race/immigrant status) ///
xsc(r(-1.5 1.5)) xlabel(-1.5"1" -0.5"2" 0.5"3" 1.5"4") ///
xsize(6)
graph export Output200720/Robustness_Figure2.png, as(png) replace
```

All decisions made are saved; if anything changes, can just rerun code

Reproducible research makes us better scholars

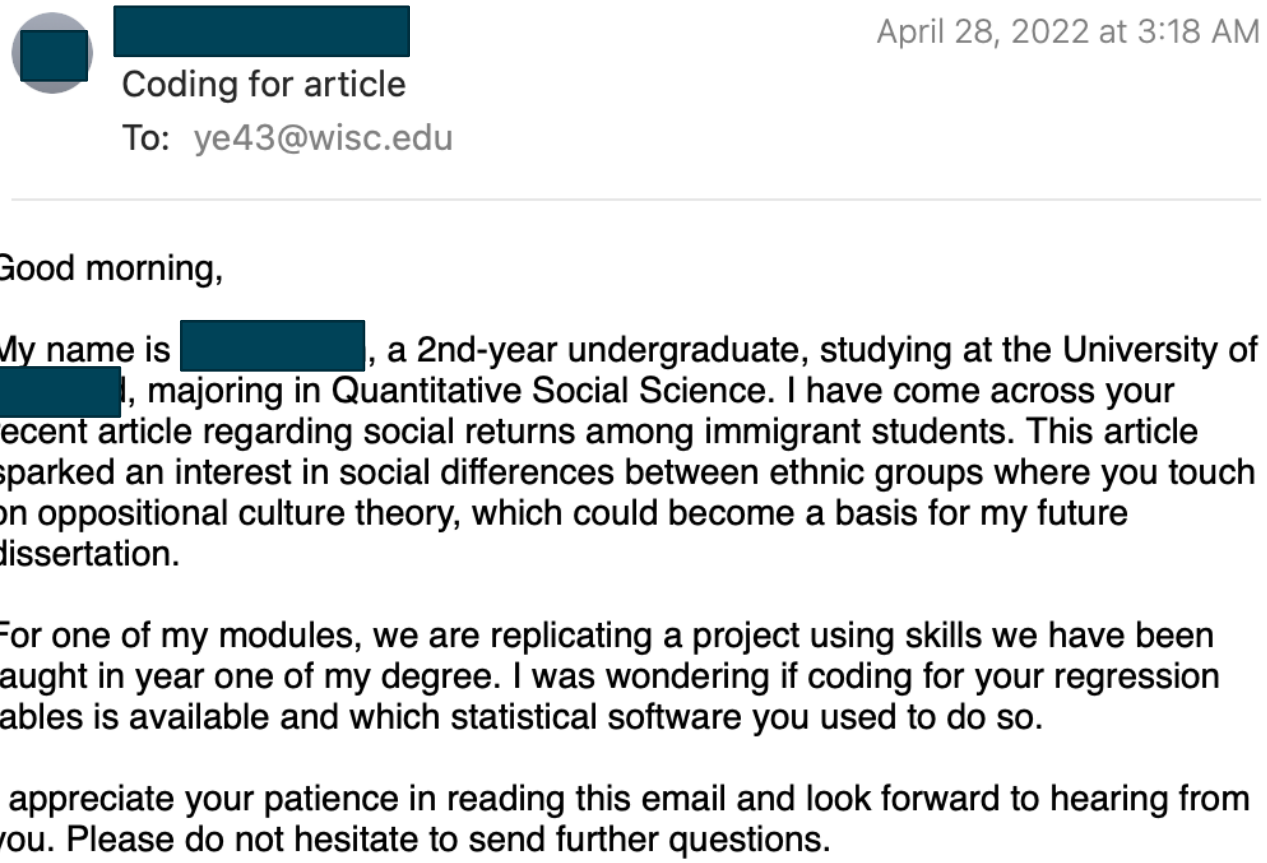
- Pushes you to think carefully about what each step is doing & whether your decision is justified → higher quality research



Replication packages of papers = a clear statement of decisions made before one reached their conclusion

Reproducible research is desired by others

- Makes the community feel more confident in research findings
- Allows others to learn from your work



[redacted]

April 28, 2022 at 3:18 AM

Coding for article

To: ye43@wisc.edu

Good morning,

My name is [redacted], a 2nd-year undergraduate, studying at the University of [redacted], majoring in Quantitative Social Science. I have come across your recent article regarding social returns among immigrant students. This article sparked an interest in social differences between ethnic groups where you touch on oppositional culture theory, which could become a basis for my future dissertation.

For one of my modules, we are replicating a project using skills we have been taught in year one of my degree. I was wondering if coding for your regression tables is available and which statistical software you used to do so.

I appreciate your patience in reading this email and look forward to hearing from you. Please do not hesitate to send further questions.

```
### Thank you! ###


# Leafia Ye
# leafiaye@ssc.wisc.edu

library("dramatic_exit_music")
```