

Teaching reproducibility

Notes from imposing reproducibility on 200 undergraduate students

ROHAN ALEXANDER, UNIVERSITY OF TORONTO

CANADIAN ECONOMICS ASSOCIATION CONFERENCE 2022

4 JUNE 2022, OTTAWA, CANADA

Plan

Background

Imposing reproducibility on student-work

Student-assessment of reproducibility

Final papers

Student-Led Reproducibility Challenge

Next steps

Background

STA304

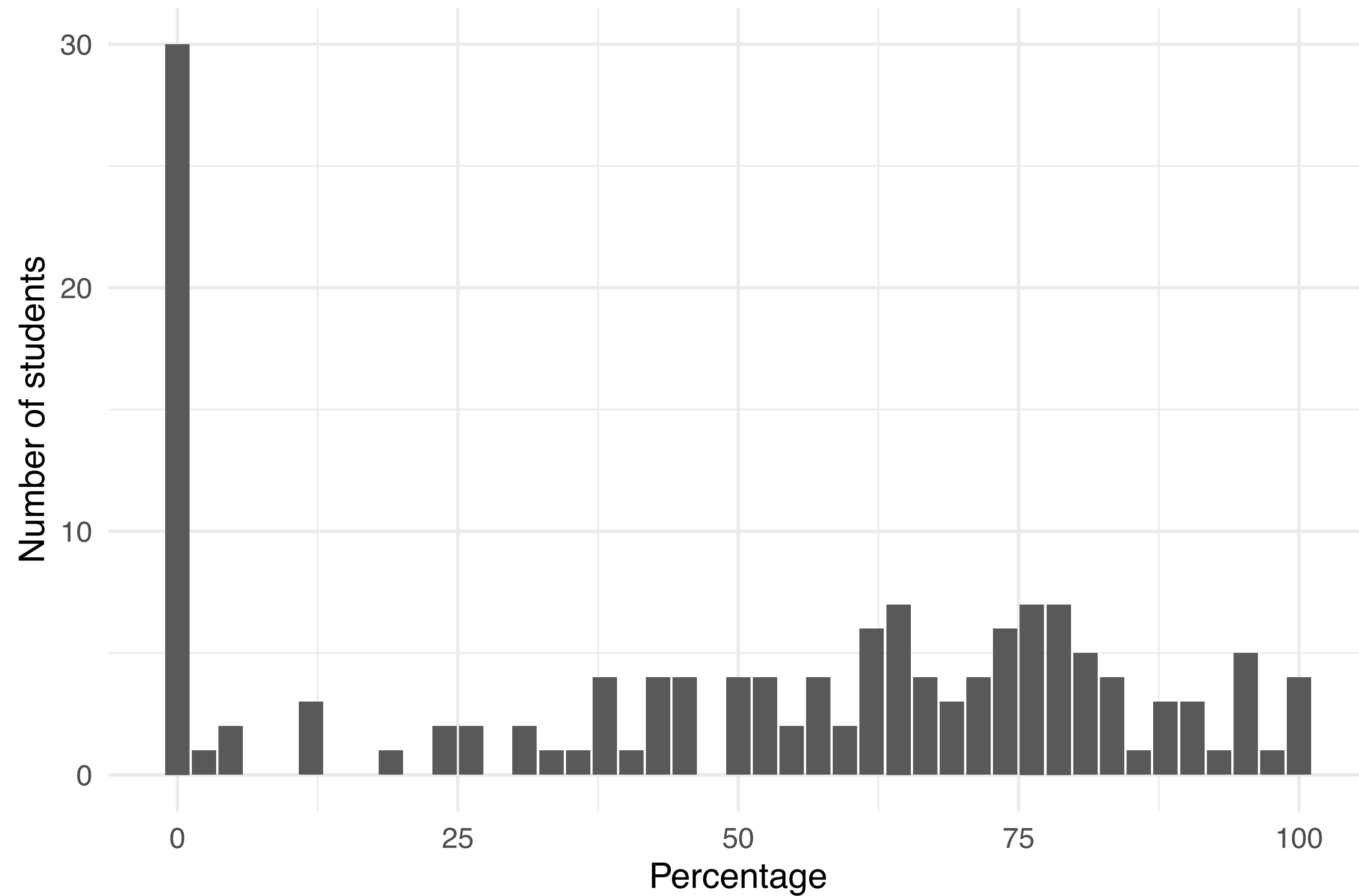
- STA304 ‘Surveys, sampling, and observational data’ upper-level undergraduate course at the University of Toronto’s Department of Statistical Sciences.
- Roughly 200 students enrol.
- No (real) pre-requisite of R, Git, GitHub.
- Assessment involves writing papers, such as:
 - Week 4: Write a short paper using data from Open Data Toronto in a completely reproducible way.
 - Week 7: Pick a paper from an AEA journal published since 2019 and redo a few aspects of it using the SSRP, then use that as a base to do your own research.
 - Final paper: Do original results

Imposing reproducibility on student work

Open Data Toronto paper

- Often this is the first time a student has picked their own interest, gathered their own data, and written a paper.
- Major issues include:
 - Using R
 - Different roles for R and R Markdown/Quarto.
 - Using Git and GitHub.
 - Deciding what to do.
 - Citations.
 - Writing.

Open Data Toronto paper



Git/GitHub and R training

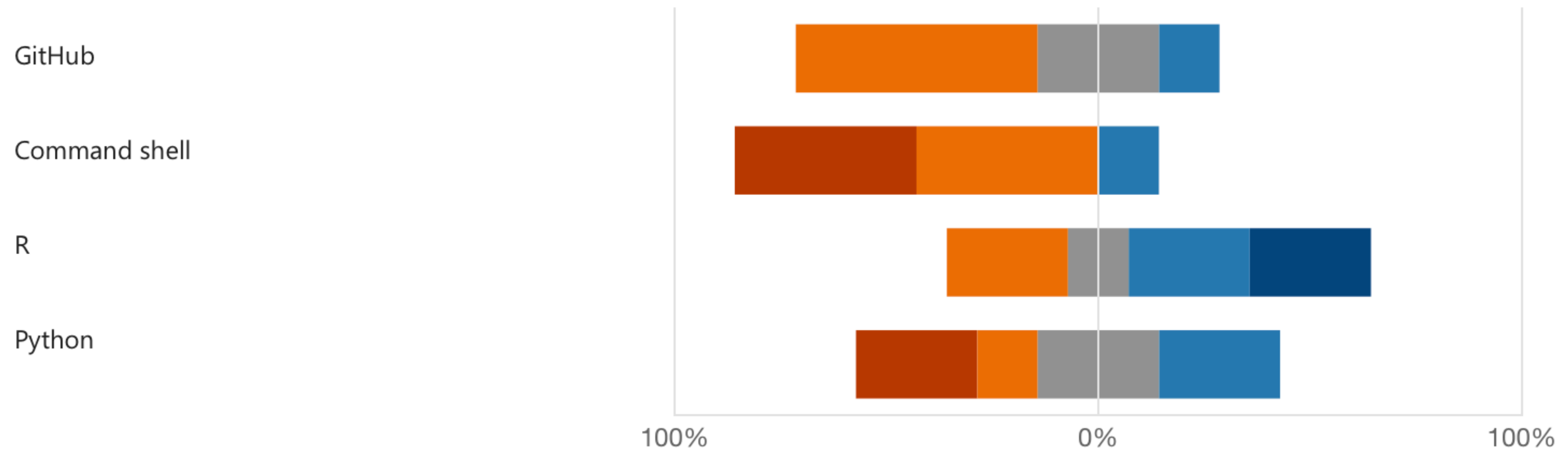
- Two six-hour sessions taught by TAs trained in live-coding and pedagogy by Software Carpentries.
- Cost (to me) \$800 to pay the TAs.
- Optional, but 50 per cent attendance.
- Interested in whether this reduced anxiety:
 - Pre-training survey
 - Post-training survey

Students need help with skills

22. Please rate your confidence current skills level in the following areas:

[More Details](#)

■ Never used ■ Not confident ■ Somewhat confident ■ Confident ■ Very confident



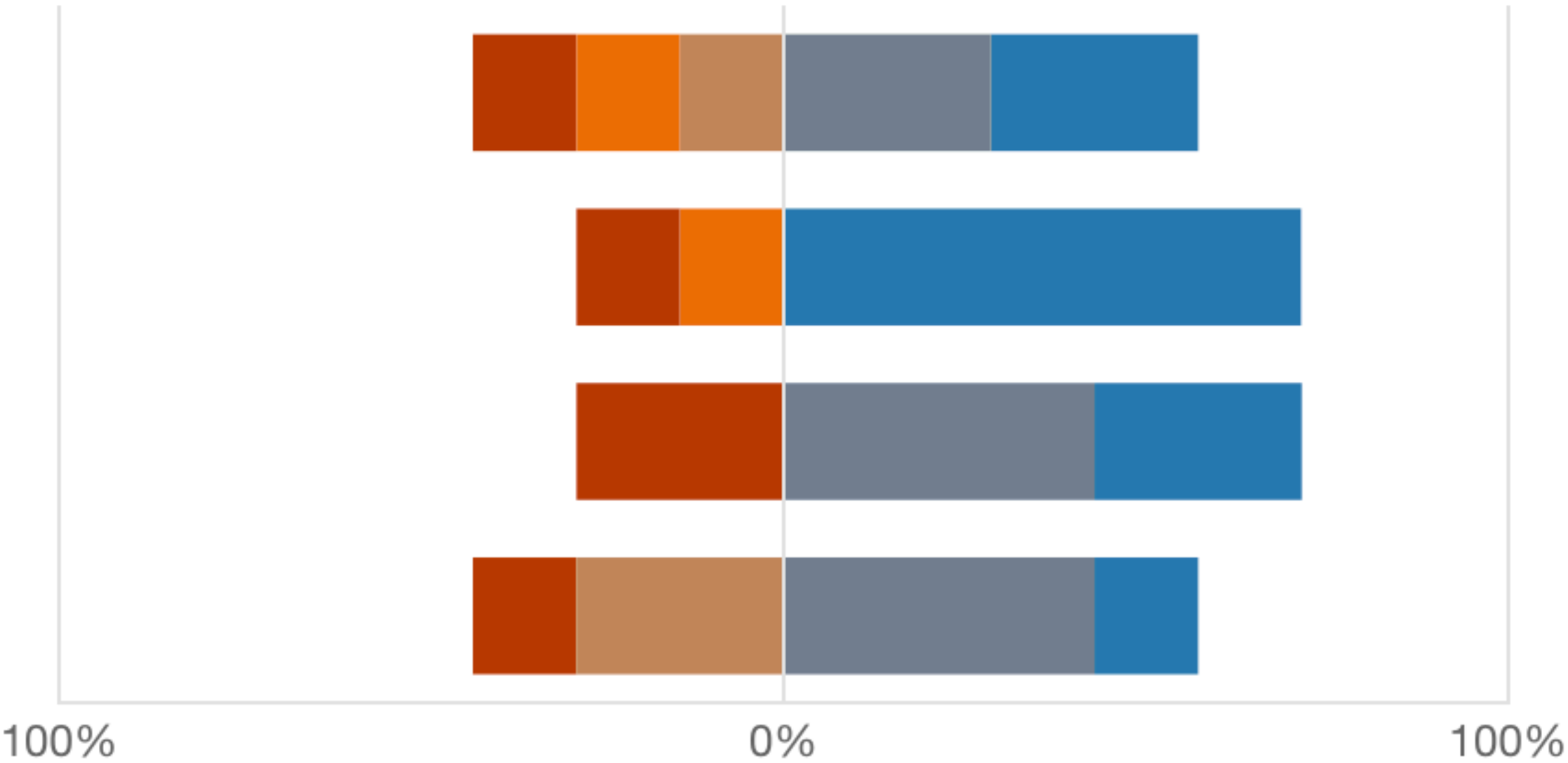
So they can do reproducible research

23. Please rate your confidence with the following general tasks

[More Details](#)

Strongly disagree Disagree Neutral Agree Strongly agree Unsure

- I can write a small program, script, or macro to address a problem in my own work.
- I know how to search for answers to my technical questions online.
- While working on a programming project, if I got stuck, I can find ways of overcoming the problem.
- I am confident in my ability to make use of programming software to work with data.

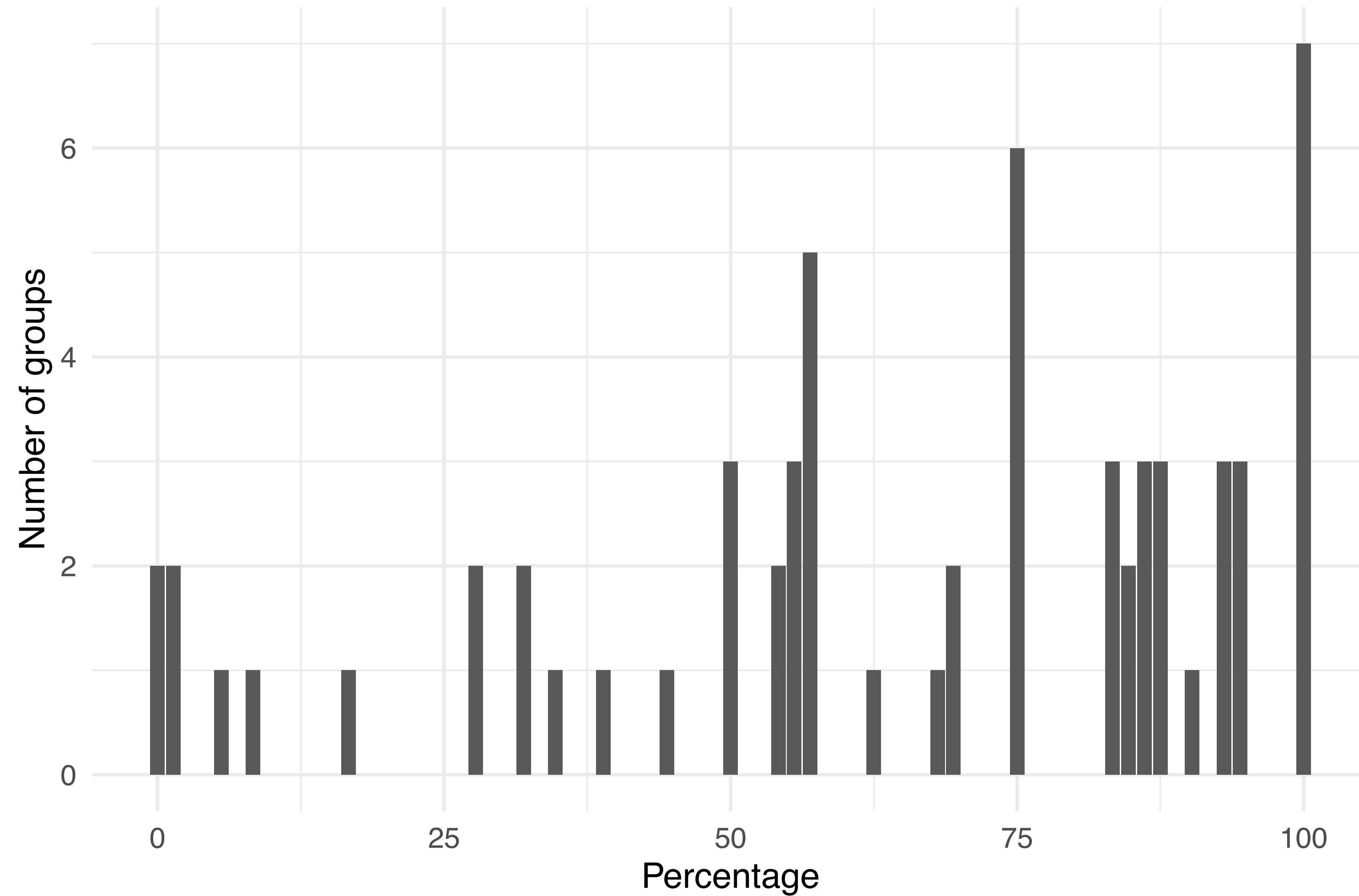


Students reproducing economics papers

AEA journals paper

- Often this is the first time a student has interacted with academic research.
- Major issues include:
 - Picking a paper
 - Using the SSRP
 - Stata

AEA journals paper



AEA journals paper

Western, 2021, 'Inside the Box: Safety, Health, and Isolation in Prison'

Mass incarceration in the US can and should be improved*

The current system results in undesirable outcomes, yet improvements can be made

Alyssa Schleifer

Hudson Yuen

Tamsen Yau

28 February 2022

Abstract

The American prison system has garnered its fair share of criticism, especially regarding safety and rehabilitation. This paper analyzes how prison conditions influence mass incarceration rates in the United States, as well as the extended impact these conditions have on the well-being of prisoners and subsequent recidivism rates in the population. While results suggest that current rehabilitation programs fall short on many fronts, there is still potential for improvement in areas such as social support initiatives and mental health services. Furthermore, minorities are more affected by issues such as violence and crime which points to broader issues of system inequalities that parallel those found in society. Finally, we noted several differences in COVID-19 case numbers between prisons in different states, which indicates how varying public health practices impact the overall safety of the prison system. These findings are worth considering in order to implement more effective initiatives, programs, and policies within the prison system.

AEA journals paper

Kearney, Levine, & Pardue, 2022, 'The Puzzle of Falling US Birth Rates...'

Declining Birth rates in the US: An Analysis of Potential Factors*

Reproduction of 'The Puzzle of Falling US Birth Rates since the Great Recession'
(Kearney, Levine & Pardue, 2022)

Kimlin Chin

28 February 2022

Abstract

Over the past 10 years, US birth rates have been dropping steadily to an all time low of 55.8 per 1,000 women in 2020. Kearney, Levine and Pardue (2022) analyzed this phenomenon by exploring demographic, economic, social and policy factors to little enlightenment of the cause of the decline. However, they did discover a correlation between the decline and the birth cohorts of mothers, and offer the conjecture of shifting priorities as the reasoning. We replicate the results of this study with respect to the demographic and cohort effects, and discuss alternative interpretations of the data considering the impact of the feminist movement and the evolution of technology and media. The results were consistent with the initial study.

AEA journals paper

Kearney, Levine, & Pardue, 2022, 'The Puzzle of Falling US Birth Rates...'

The Decline in U.S Birthrates in Recent Years is Indicative of Cultural and Economic Changes*

Olaedo Okpareke, Arsh Lakhanpal, Swarnadeep Chattopadhyay

27 February 2022

Abstract

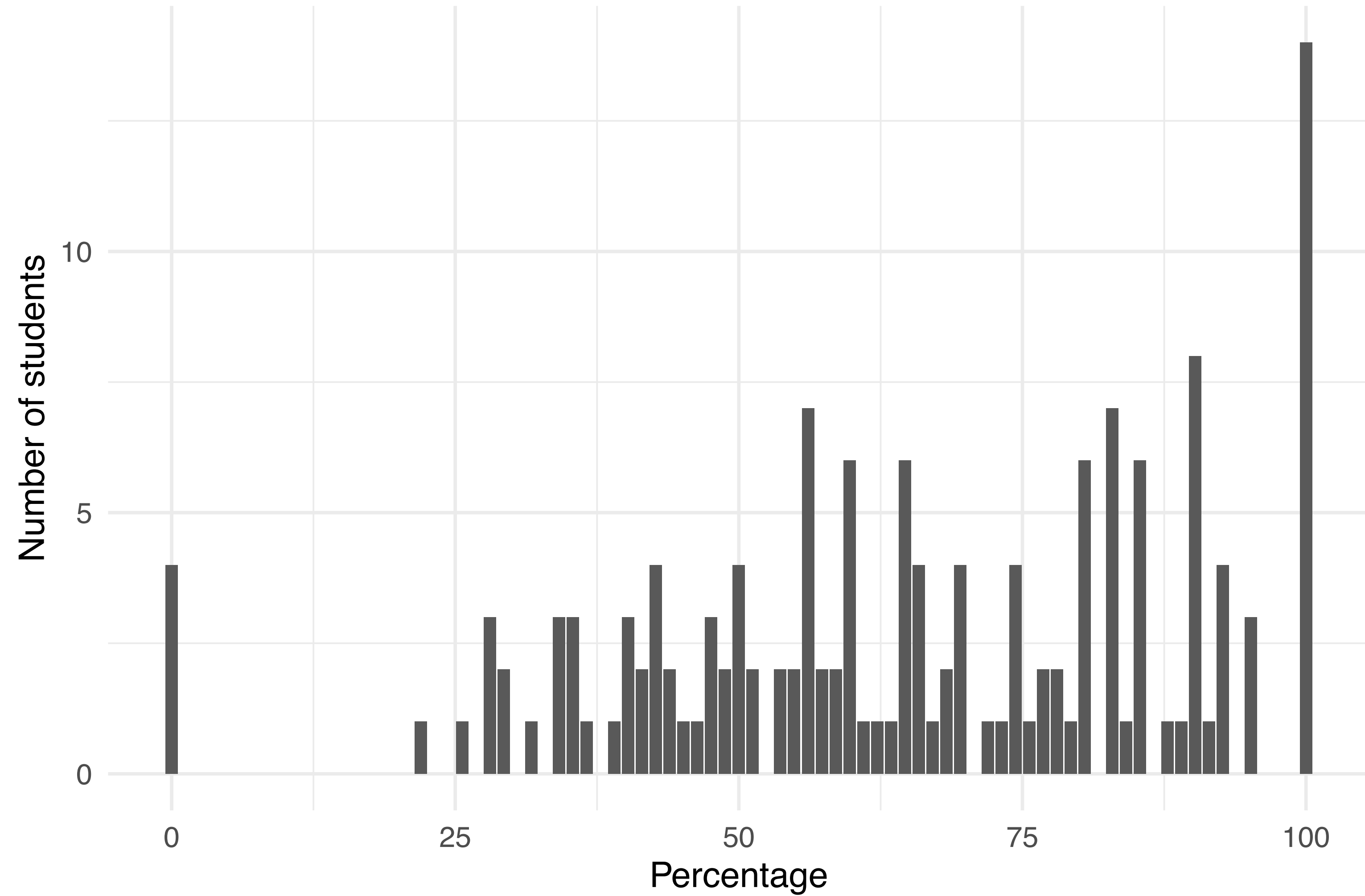
Birth rates provide useful information on population growth, and an above-replacement birth rate is indicative of stability. However, previous research has found that birth rates in industrialized societies such as the U.S have been decreasing in recent years. We use data from the American Economic Association to analyze this by using graphs and tables to observe the trend in birth rates of different demographics, using R (R Core Team 2020) and other packages. We also observe the trends behind changing economic and cultural factors that could affect the birth rates. We find that birth rates of young women of different races have been on a steep decline over the past 20 years, and factors that are known to decrease birth rates have increased. This report shows that the decline in birth rates is the result of cultural and economic changes between two generations of women. While this indicates that the U.S has become more culturally liberated, the declining birth rates showcases systemic effects of the worsening economy and increasing costs on the average American woman.

Final paper

Final Paper

- Task: “Working individually and in an entirely reproducible way please write a paper that involves original work to tell a story with data.”
- Major issues include:
 - Using R
 - ~~Different roles for R and R Markdown/Quarto.~~
 - ~~Using Git and GitHub.~~
 - ~~Deciding what to do.~~
 - Citations.
 - ~~Writing.~~
 - Time.

Final Paper



Final Paper

Assessing Infant Mortality Rate: Problems stemming from Household Living Conditions, Women's Education and Health*

Trends in India from the 1998 Demographic Health Survey

Jacob Yoke Hong Si

27 April 2022

Abstract

What areas can be improved in order to promote the well-being of women in India and hence, reduce the infant mortality rate? Utilizing the data from the 1998-1999 India National Family Health Survey provided by the Demographic and Health Survey (DHS) program, we look to depict the demographics of Indian women and infants in different states of India. We have found that the root causes of poor infant mortality rates stem from having poor living conditions that affect the likelihood of women to attain education and understand the importance of antenatal care and birth delivery assistance. We also explore other factors such as potentially inheritable traits (unhealthy body weight and anaemia disease) as well as an infant's diet. These factors are crucial in the development of an infant and the reduction of the infant mortality rate.

Keywords: demographics, health, living conditions, infant mortality rate, education, antenatal care, body weight, anaemia, diet, birth delivery assistance.

Final Paper

How did universities respond to student mental health throughout COVID-19?*

Sidharth Gupta

03 May 2022

Abstract

Poor mental health is detrimental to learning capabilities, so it is extremely important that universities and colleges provide high-quality and accessible mental health resources. Using a survey dataset with over 10,000 responses from students across 8 universities, we evaluate the relationship between students struggling with their mental health, and how they perceive the quality of their schools mental health resources. We find that there is a weak association between these two features across all schools, and also make connections to how schools supported their students during COVID-19. Our work brings impactful results that university administration can consider to better support student health.

Keywords: mental health, data science, supervised learning, COVID-19

Final Paper

DeFi-ing Cyber Attacks*

A statistical analysis of cybersecurity attacks in decentralized finance

Jack McKay

27 April 2022

Abstract

Decentralized finance (DeFi) describes the emerging ecosystem of financial services, protocols and applications built on and designed for public blockchains such as Ethereum and Algorand. One of the most promising applications of blockchain technology, DeFi has experienced explosive growth over the last few years. But with this growth has also come the attention of bad actors, who aim to exploit flaws in protocol security for personal gain. Using data collected by DEFIYIELD's REKT Database on the 200 most costly attacks on DeFi protocols, this paper will analyze and identify trends in these attacks, with the goal of identifying what types of attacks are most common and most costly to the DeFi sector as a whole, as well as for individual chains, in order to determine where cybersecurity efforts would best be focused.

Final Paper

Determinants of Canadian Politicians' Popularity on Wikipedia.*

Ethan Sansom

27 April 2022

Abstract

Voters internationally are increasingly using the internet as a primary tool to access information on politicians and political parties. Consequently, political organizers and researchers alike are using online platforms, such as Twitter, Facebook, and more recently Wikipedia, to analyze the sentiments of voters. This paper investigates the relationship between Canadian politicians' characteristics - including party affiliation, political position, and age - and politicians' Wikipedia page views in the months prior to the 2021 Federal Election. Wikipedia page views are found to be somewhat greater for politicians in more popular parties, and for Party Leaders, Ministers, and Prime Ministers in particular.

Keywords: Canada elections, election forecasting, Wikipedia, linear regression

Final Paper

Evolution of Income Inequality in Canada*

It's becoming a more serious issue as over time, especially in the richest provinces.

Alicia Yang

27 April 2022

Abstract

People have claimed that the income disparity in Canada has increased alongside with the rapid growth of its economy. To explore whether and how income disparity became a serious issue in Canada, this paper aims to investigate the relationship between time, province, and income inequality in Canada. By utilizing the data from Statistics Canada, we found that income inequality continuously increase over time and is a greater issue in more developed provinces. Particularly, it is closely related to the prosperity of a region. In addition, implications and possible reasons behind this correlation is discussed and possible solutions to the problem of income disparity is proposed.

“I got zero on my first paper and a full mark on my final paper!”

STA304 student

Student-Led Reproducibility Challenge

Reproducibility challenge

- Data Sciences Institute thematic program on reproducibility
- Co-leads (social sciences, life sciences, engineering) all did the same task in their classes.
- Industry presentations: Xero and Bank of Canada
- Results were remarkably consistent
 - Languages are a major hurdle
 - Access to data remains an issue
 - Authors are not providing enough information

Main takeaways and next steps

Cultural change needed

“[I]t might be best for researchers in the area if it was taken down (or modified to not be misleading).”

Random economics professor (not an author on the paper)

Takeaways

- Reproducibility can, and should, be insisted on, at an undergraduate level.
 - Students need short, skills-intensive, boot-camp style courses, on things like Git/GitHub, R, R Markdown/Quarto.
 - Extending (to the extent possible) academic freedom to our students.
 - Need to develop tools for auto-checking.
 - Cultural change needed in economics
-

Rohan Alexander

Assistant Professor

Information and Statistical Sciences

University of Toronto

Assistant Director, Canadian Statistical Sciences Institute (CANSSI) Ontario

Senior Fellow, Massey College

Faculty Affiliate, Schwartz Reisman Institute

Co-lead, DSI Thematic Program in Reproducibility

rohan.alexander@utoronto.ca