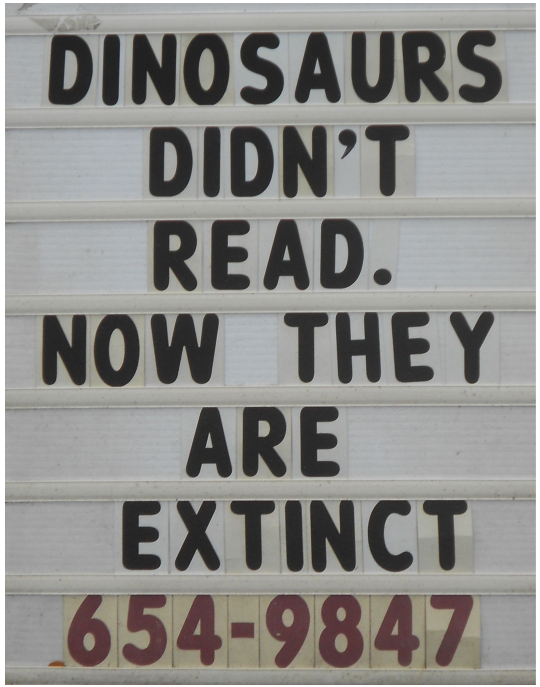
The Impact of the COVID-19 Pandemic on Women's Contribution to Public Code

Annalí Casanueva - Davide Rossi
Stefano Zacchiroli - Théo Zimmermann

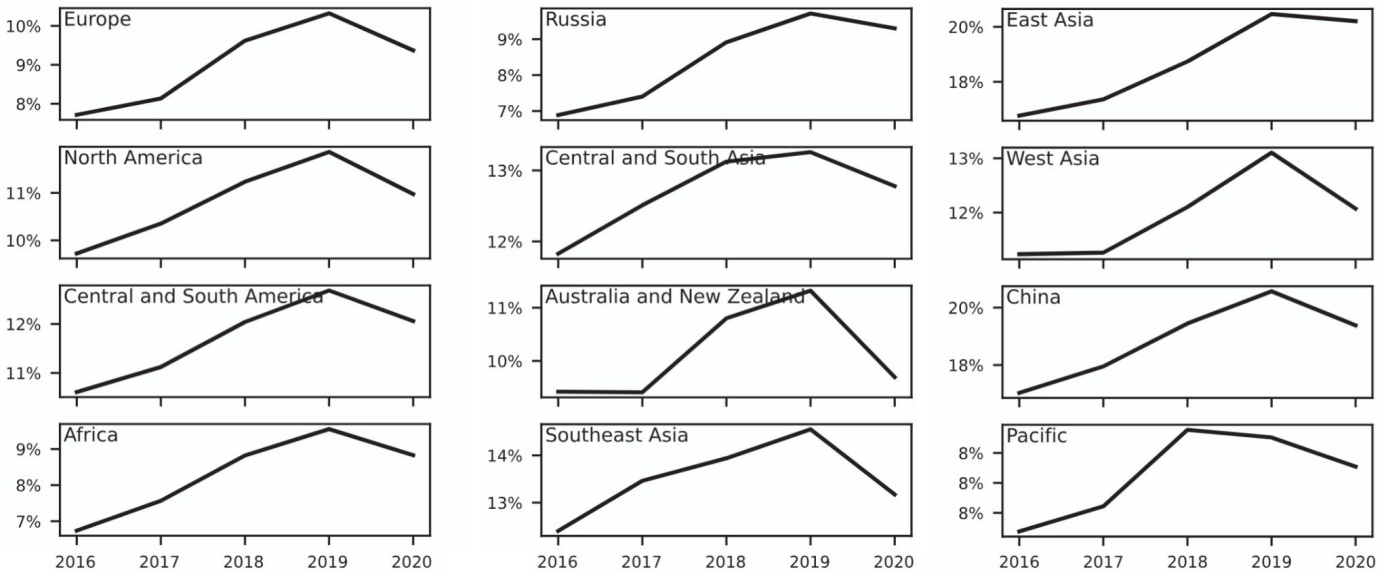
Empir. Softw. Eng. 30(1): 25 (2025). Preprint: https://arxiv.org/pdf/2410.01454

Foreword

Correlation does not imply causation



Background



Ratio of yearly female authors 2016-2020 by world region

Rossi, D. and Zacchiroli, S., 2022. Worldwide gender differences in public code contributions and how they have been affected by the covid-19 pandemic. In Proc. 2022 ACM/IEEE 44th International Conference on Software Engineering.

Questions

- Can we conclude that COVID-19 **caused** reduced women contribution to public source code **at world scale**?

  RQ1: How did the COVID-19 pandemic impact the ability of women (relatively to men) to contribute to public code?

- And, if so, were different groups of women contributors affected in different ways?

  RQ2: Which groups of women contributing to public code were impacted the most (relatively to men in the same groups), in their ability to contribute, by the COVID-19 pandemic?

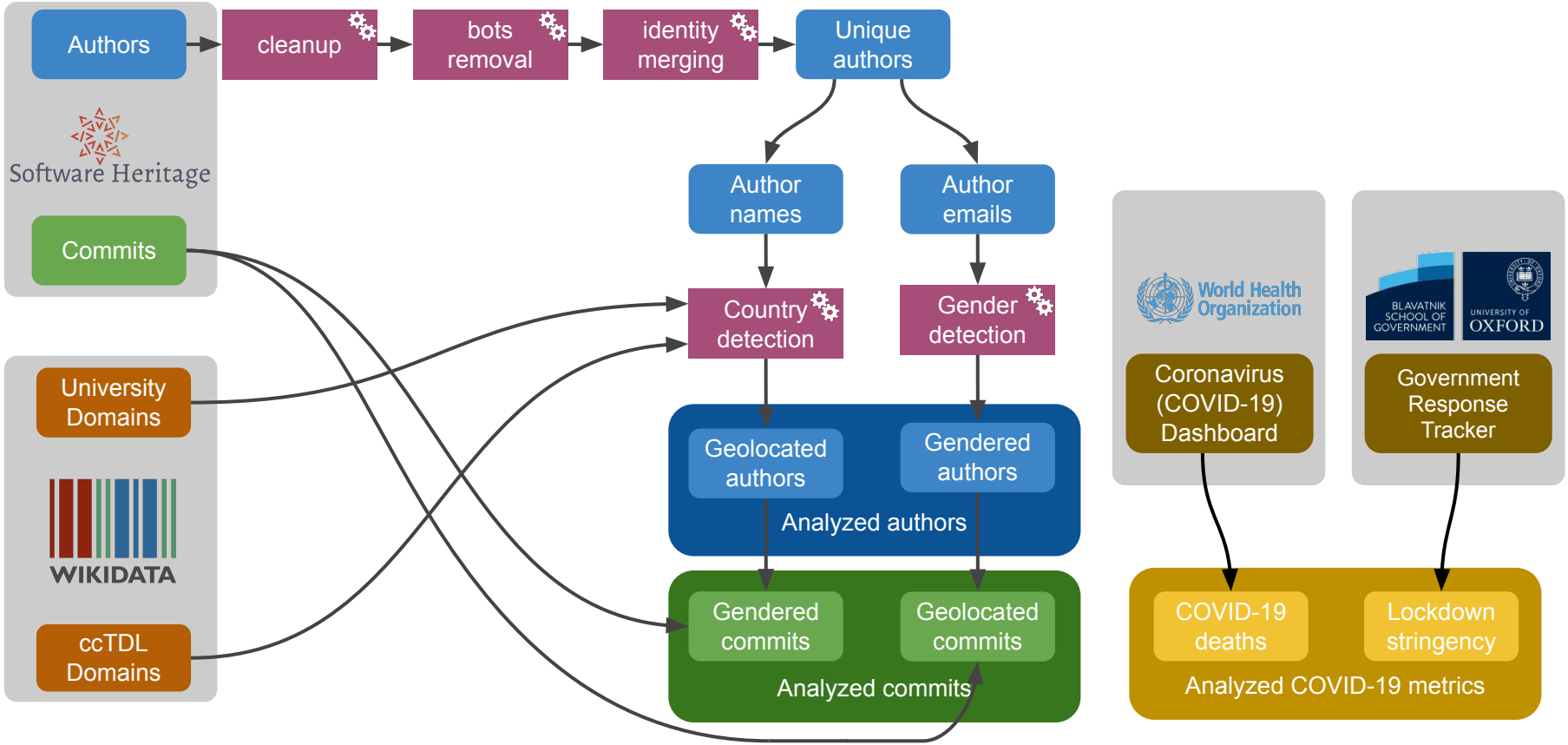
Methodology

- The gold standard for assessing causation is the Randomized Controlled Trial (RCT), which cannot be applied here.
- We adopt **Difference-in-Differences** (**DiD**), a standard econometric technique for causal inference.
- **Core Idea**: Compare the change in outcomes over time between a group that received a treatment and a group that did not.
- **Key Assumption**: Parallel Trends - the treatment group, if untreated, would have behaved *parallel* to the control group.

COVID-19 and DiD

- **Challenge**: every country was eventually “treated” by COVID-19 — there is no “pure” control group.
- **Solution**: We use a **Two-Way Fixed Effects** (TWFE) model, a model that allows for the inclusion of continuous treatment intensity and its staggered implementation.
- **Underlying idea**: leverage variations in the timing and intensity of COVID exposure across countries.

Data pipeline



Regression analysis

Ordinary Least Square (OLS) regression with target:

$$\begin{aligned} Commits_{acw} = \; & \beta_1 Woman_a \times Exposure\ to\ COVID\text{-}19_{cw} \\ & + \beta_2 Exposure\ to\ COVID\text{-}19_{cw} \\ & + \gamma_a + \theta_w + \alpha Time \times Woman_a + \epsilon_{acw} \end{aligned}$$

Regression analysis

Ordinary Least Squ[illegible](OLS) [illegible]ssion [illegible]

The specific impact of COVID-19 on women w.r.t to their men counterparts

Exposure in Country *c* during week *w*

$$Commits_{acw} = \beta_1 Woman_a \times Exposure\ to\ COVID\text{-}19_{cw} + \beta_2 Exposure\ to\ COVID\text{-}19_{cw} + \gamma_a + \theta_w + \alpha Time \times Woman_a + \epsilon_{acw}$$

Regression analysis

Ordinary Least Square (OLS) regression with target:

$$Commits_{acw} = \beta_1 Woman_a \times Exposure\ to\ COVID\text{-}19_{cw} + \beta_2 Exposure\ to\ COVID\text{-}19_{cw} + \gamma_a + \theta_w + \alpha Time \times Woman_a + \epsilon_{acw}$$

Author fixed effect and week fixed effects: account for differences between authors at usual contribution level and non-COVID global "spikes".

Women-specific trends

Results

- **RQ1**: COVID-19 disproportionately affected women's ability to contribute to public code, in a negative way compared to men :-(
- **RQ2 — by experience**: no statistical relevance.
- **RQ2 — by activity level**: no statistical relevance.
- **RQ2 — hobbyists vs professionals**: the more authors were contributing to public code outside of working hours before the pandemic, the more women were negatively impacted by the pandemic in their ability to contribute.

Why it matters?

Our results can be used to:

- Align **public code production with global studies** showing pandemic-induced exacerbation of household and societal gender disparities.
- Call for **systemic policy interventions in open source communities** to address:
  - Pre-existing inequalities
  - Resilience to external shocks
  - Inclusive design of diversity initiatives

Robustness & Sensitivity Analysis

- **Challenge**: The global nature of the pandemic impaired our ability to triangulate findings with alternative causal methods.
- **Our Approach**: we conducted extensive robustness checks to test if our core finding is sensitive to model specification and measurement choices.
  - Alternative COVID intensity proxies: deaths vs stringency index
  - Varying control sets: selectively adding/removing features

Future work

- Bounce back effect?
- Effect by kind/size of project
- Effect by kind of task (development, documentation, testing, QA, …)

The Impact of the COVID-19 Pandemic on Women's Contribution to Public Code

- COVID-19 disproportionately affected women's ability to contribute to public code
- Hobbyists woman contributors have been affected more than those developing as a hobby/service to the community

Learn more

Annalí Casanueva Artís, Davide Rossi, Stefano Zacchiroli, Théo Zimmermann. The impact of the COVID-19 pandemic on women's contribution to public code. Empir. Softw. Eng. 30(1): 25 (2025).

Preprint (open access): https://arxiv.org/pdf/2410.01454