**IISE Transactions Reproducibility Report**

**Reproducibility Report for An Analysis of COVID-19 Vaccine Hesitancy in the U.S.**

**1. Metadata**

Manuscript Title: An analysis of COVID-19 vaccine hesitancy in the U.S

Manuscript ID (if available):

Authors: Hieu Bui, Sandra Eksioglu, Ruben Proano, and Sarah Nurre Pinkley

**2. Data availability**

\_\_\_\_\_\_\_\_\_\_ A. Either no data are used in this study or all data used are included in the main text or supplemental materials.

\_\_\_\_ x \_\_\_\_ B. The data used in this study is publicly available at the following website\* (please provide the website link).

**Website links**:

* <https://aspe.hhs.gov/reports/vaccine-hesitancy-covid-19-state-county-local-estimates>
* <https://health.google.com/covid-19/open-data/>
* <https://data.cdc.gov/Vaccinations/Vaccine-Hesitancy-for-COVID-19-County-and-local-es/q9mh-h2tw>
* <https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-County/8xkx-amqh>
* <https://data.cdc.gov/Flu-Vaccinations/Influenza-Vaccination-Coverage-for-All-Ages-6-Mont/vh55-3he6>
* <https://data.census.gov/table?q=S2801&g=0100000US$0500000&tid=ACSST1Y2019.S2801>
* <https://data.census.gov/table?t=Income+and+Poverty&g=0100000US$0500000&tid=ACSST1Y2019.S1701>
* <https://www.bls.gov/lau/#cntyaa>
* <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ>
* <https://console.cloud.google.com/bigquery?p=gdelt-bq&d=covid19&t=onlinenewsgeo&page=table&pli=1>
* <https://github.com/grammakov/USA-cities-and-states>
* <https://google-research.github.io/vaccination-search-insights/>?
* <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>
* Tweets are extracted using Twitter API V2.

\_\_\_\_\_\_\_\_\_\_ C. The data used in this study is not yet publicly available but will be made publicly available at the time of paper acceptance\*\* or will be made publicly available subject to an embargo period of \_\_\_\_ years, counting from the time of paper acceptance. If an embargo period is invoked, please explain the reason for embargo. (please provide the website link\*).

\_\_\_\_\_\_\_\_\_\_ D. The data used in this study is not and will not be made publicly available due to the following reason(s). Please present the reason(s).

Note:

\* To share data/code, we recommend the authors to use [Zenodo.org](https://zenodo.org/), a general-purpose open repository operated by Europe’s CERN. The Zenodo website is free, reliable, and easy to use. It assigns a DOI upon submission, making one’s data/code files citable. One could also place the data (and/or code) at his/her research website, on a Github website, or any other publicly accessible websites, if so preferred. We do expect the website holding the datasets/codesets to be stably and reliably accessible over long term. If one desires to place the data (and/or code) through IISE Transactions’ repository service, please visit IISE Transactions website, go to “Instructions for authors,” and then to “Checklist Items 9 and 10”.

\*\* The time of paper acceptance is the time when the Accepted Manuscript Online (AMO) version of the paper is published. This time is documented for every paper in IISE Transactions on the journal’s website.

**3. Data use ethics**

\_\_\_\_\_\_\_\_\_\_ My choice in Section 2 is (A).

\_\_\_\_ x \_\_\_ I certify that the authors have the legitimate access to the data and that nothing in the provisions governing the use of the data prohibits the authors from using the data in this research.

**4. Computer code**# **availability**

\_\_\_\_\_\_\_\_\_\_ A. Either no computer code is used in this study or the settings used in software are fully described in the main text or supplemental materials.   
  
If one uses a programmable language, including but not limited to, MATLAB, R, Python, C/C++, Java, then the second condition above, i.e., “settings in software fully described,” does not apply. Pseudo-code, regardless of how detailed it may be, is not considered as “settings fully described.” In other words, when the authors use code in a programmable language to produce any result in the paper, code-sharing is required, unless explicitly prohibited by a legal document (like a legally binding non-disclosure agreement).

\_\_\_\_x\_\_\_\_\_ B. The computer code used in this study is publicly available at the following website (please provide the website link; see the first note \* in Section 3).

\_\_\_\_\_\_\_\_\_\_ C. The computer used in this study is not publicly yet available but will be made publicly available at the time of paper acceptance or will be made publicly available subject to an embargo period of **1 year**, counting from the time of paper acceptance. **Please describe where to make the data publicly available**. If an embargo period is invoked, please explain the reason for embargo. (please provide the website link; see the first note \* in Section 3)

\_\_\_\_\_\_\_\_\_\_ D. The computer code used in this study is not and will not be made publicly available due to the following reason(s). Please present the reason(s).

**Note:**

# If the authors run their code on a software platform (either commercial or freeware), the authors do not need to provide the software platform, but simply to provide one’s own code.

## The website for making code available and the meaning of “the time of paper acceptance” follow what is explained in Section 2.

**5. Reproducibility**

**5.1 Computer and software environment**

* Please describe the computer hardware conditions and software environment on which the authors produce the results reported in the paper.

**Response**: The code for data analysis and visualization is written Python 3.9. The code can be run on Windows, Linux, or Mac computers. All required Python packages are managed in an Anaconda environment. They can be installed by either Conda or pip.

The scripts to create the figures are simplified so that only a few packages are required. The main Python packages include pandas, geopandas, numpy, shap, scipy, libpysal, matplotlib, and plotly.

“imagemagick” is used to convert figures from PNG to EPS format.

**5.2 Workflow**

* The authors please use the following table to provide the instructions on how to reproduce the results in the figures/tables of the paper.
* The table below is supposed to include each and every figure/table in the paper that is considered as a research output or used to support the research conclusions. Hand-drawn diagrams or flowcharts are excluded from the reproducibility workflow.
* If you choose Option D above, please certify the following statement.

\_\_\_\_ x \_\_\_\_ I certify that the authors have faithfully conducted the reproducibility tasks on their own computational devices and that the following table accurately documented the filenames used, the computational times of execution, and the outcomes.

**Response:** All of the validated figures are located in the Figure folder of the submission file. The format name for each figure follows the structure: ***yyy.png***, where *yyy* is the name of the figure.

We also include a Python script to create figures automatically (i.e., ***create\_all\_figures.py***)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Which results to reproduce** | **Data File** | **Code File** | **Expected output**  **(PNG format)** | **Run time** |
| Figure 1 | Use ***draw.io*** website/app to generate the diagram | fig1\_proposed\_flowchart.draw.io | Figure 1 which is the flowchart in the paper | 0 |
| Figure 2 | fig4\_data.csv and and county\_shapes.geojson | fig2\_code.py | *GVF\_map \_v2* | ~ 11 second |
| Figure 3 | fig3\_data.csv | fig3\_code.py | *hesitant\_state\_v2* | ~ 2 seconds |
| Figure 4 | fig4\_data.csv | fig4\_code.py | *aspe\_delphi\_vhb* | ~ 3 second |
| Figure 5 | fig5\_data\_permut.csv and fig5\_data\_shap.csv | Fig5\_code.py | *feature\_importance* | ~ 1 seconds |
| Figure 6 | fig6\_data.csv | fig6\_code.py | *ranking\_political\_search* | ~ 1 second |
| Figure 7 | fig7\_data.csv | fig7\_code.py | *search\_insights\_v2* | ~ 1 seconds |
| Figure 8 | fig8\_data1.csv and fig8\_data2.csv | fig8\_code.py | *shap\_values\_all\_clusters* | ~ 4 second |
| Figure 9 | fig9\_data.csv | fig9\_code.py | *ranking\_stringency\_v2* | ~ 1 second |
| Figure 10 | fig10\_data.csv | fig10\_code.py | *VH\_c1\_c5* | ~ 1 second |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Which results to reproduce** | **Data File** | **Code File** | **Expected output**  **(PNG format)** | **Run time** |
| Figure S1 | figS1\_data.csv and figS1\_pop.csv | figS1\_code.py | *vaccination\_rate\_completeness* | ~ 4 seconds |
| Figure S2 | figS2\_data.csv | figS2\_code.py | *corr\_heatmap* | ~ 1 second |
| Figure S3 | figS3\_data.csv | figS3\_code.py | *avg\_tweet\_count\_by\_internet\_access\_group* | ~ 1 second |