README file for

“Social Media and Mental Health”

By Luca Braghieri, Ro’ee Levy, and Alexey Makarin

openicpsr-175582

This readme file contains instructions on how to replicate the results of our paper. We provide instructions for accessing the data sets used in the project on the next page. We also provide the cleaning code that can be used to transform the raw data for analysis as well as all of our analysis code.

Project Folders

The following diagram shows the levels of the project tree structure at a high level.

Project_Folder/

- README.pdf
- Code/
  - ado/
  - Data_Preparation/
  - Analysis/
  - Master_file.do
  - pathnames.do
- Data/
  - Input/
  - Intermediate/
- Exhibits/
- Temp/

Before running the code, it is necessary to update the global paths in the pathnames.do file for the relevant file structure on your computer. If you set up your directories as in the above diagram, then it is only necessary to update the global $PROJ with the location of the project folder.

Within each sub-directory in the cleaning directory, there is a single code file that calls all of the other do-files in that directory (if relevant). In order to run all of the data preparation code in order, use the do-file “Master_file_data_prep.do” in the “Code/Data_Preparation” directory.

In order to run all of the analysis code, use the do-file “Master_file_analysis.do” in the “Code/Analysis” directory. In order to run all code at once, use the do-file “Master_file.do” in the “Code” directory. Our code also uses a few custom ado files, which can be found in the “Code/ado” directory. See the below tables for details on each of the individual do-files we use for cleaning and analysis.

The analysis output is all placed in the “Code/Exhibits” directory, where each of the tables and figures is named based on their name in the paper (e.g., “Table 1.tex”).

Software and Computational Requirements

This section describes the software and computational requirements for running the relevant analysis code, as well as the relevant packages that must be installed. The Stata code involves interaction with R, and for this purpose the needed R packages described below also must be installed.

- Stata version 17
    - rscript (as of 2022-07-01)
    - estwrite (as of 2022-07-01)
    - coefplot (as of 2022-07-01)
    - ftools (as of 2022-07-01)
    - event_plot (as of 2022-07-01)
    - did_imputation (as of 2022-07-01)
    - csdid (as of 2022-07-01)
    - drdid (as of 2022-07-01)
    - did_multiplegt (as of 2022-07-01)
    - eventstudyinteract (as of 2022-07-01)
    - avar (as of 2022-07-01)
    - reghdfe (as of 2022-07-01)
    - distinct (as of 2022-07-01)
    - boottest (as of 2022-07-01)
    - ietoolkit (as of 2022-07-01)
    - lassopack (as of 2022-07-01)
    - icw_index (as of 2022-07-01)
    - estout (as of 2022-07-01)

All of the packages mentioned above can be installed using the “*ssc install [PACKAGE NAME]*” command.

- R version 4.2.1
    - remotes (as of 2022-07-01)
    - dplyr (as of 2022-07-01)
    - data.table (as of 2022-07-01)
    - ggplot2 (as of 2022-07-01)

- kableExtra (as of 2022-07-01)
- glmnet (as of 2022-07-01)
- haven (as of 2022-07-01)
- qualtRics (as of 2022-07-01)

All of the packages mentioned above can be installed using the "install.packages("*[PACKAGE NAME]*")" command.

Memory and Runtime Requirements

The code was last run on an AMD EPYC 7642 48-Core Processor 2.30 GHz (2 processors) with 256 GB of RAM and 10 TB of free disk space. The total run time for the analysis code is approximately 2.5 hours. The total run time for the data preparation code is approximately 1 hour, and the total run time for the analysis code is approximately 1.5 hours.

Data Sources

Name of data	**Description of data**	**Instructions for access**
Facebook Expansion Dates Data	Data on the first 100 colleges that received Facebook access: college name, number of students, and Facebook introduction dates. Data are included in Table A.32.	Data from previous studies are publicly available: see Traud et al., (2012) and Jacobs et al. (2015). The file "NetworkDataTraudEtAl.csv" is based on Table A.1 available at https://arxiv.org/pdf/1102.2166v1.pdf .
Facebook Expansion Dates Data	Data on 675 colleges additional that received Facebook: college name, and Facebook introduction dates. Data are included in Table A.32.	Hand collected using Wayback Machine. See Section 3.1 of the manuscript for instructions.
National College Health Assessment Data (NCHA)	Data on more than 430,000 responses to the NCHA survey. It inquires about demographics, physical health, mental health, alcohol and drug use, sexual behaviors, and perceptions of these behaviors among one's peers, and includes questions about symptoms of mental	The data are not publicly available and are not provided in the replication package. In order to obtain them, one can contact Valerie Hartman, Senior Research Coordinator at the ACHA (vhartman@acha.org), and complete the application form. Additional information can be found on the website

	illness and questions about the take-up of mental healthcare services. The data we use is administered between the spring of 2000 and the spring of 2008.	https://www.acha.org/NCHA/ . ACHA provided us with a customized dataset that includes a variable indicating the semester in which Facebook was rolled out at each college (see Section 3.1 of the manuscript for the details).
Integrated Postsecondary Education Data System data (IPEDS)	Data at the college level, including the number of full-time undergraduate students.	The data are publicly available on the IPEDS website. One can access the file "CSV_1182021-579.csv" at https://nces.ed.gov/ipeds/use-the-data by clicking "Custom data files" (under Survey Data). Select "By Groups" -> "EZ Group" -> "Data Collection [change year]" – choose 2005 -> choose "U.S. only" -> "Search" -> "Continue"; "Available Year(s)": choose 2005 -> "Fall enrollment" -> "Gender, attendance status, and level of student: Fall 2005"; Under "Select Qualifying Variable(s)" – "Level of student" – choose the following: "All students total", "Undergraduate total", "Degree/certificate-seeking total", "Non-degree/certificate-seeking", "Graduate and first professional" -> "Save"; Under "Select from the List of Variables", choose the following: "Grand total" and "Full time total". Choose "Continue" and download as CSV. Definitions can be found here: https://surveys.nces.ed.gov/ipeds/public/glossary . One can access the file "hd2005.csv" at https://nces.ed.gov/ipeds/datacenter/data/HD2005.zip .
National Survey of Drug Use and Health, 2002-2019	Data on national estimates of substance use, mental health, and other health-related issues in the United States.	The data are publicly available on the SAMHSA website. Specifically, the file "NSDUH_2002_2019.dta" can be accessed at: https://www.datafiles.samhsa.gov/dataset/nsduh-2002-2019-ds0001-nsduh-2002-2019-ds0001 .

U.S. Census Bureau data	Data on Census Regions and Divisions of the United States.	The data are publicly available at https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf

Statement about Rights

The authors of the manuscript certify that they have legitimate access to and permission to use the data used in this manuscript. The authors also certify that they have documented permission to redistribute/publish the data contained within this replication package.