Data and Code README for

Lockdowns and Innovation: Evidence from the 1918 Flu Pandemic

(to be published at *Review of Economics and Statistics*)

Enrico Berkes
THE Ohio State University

Olivier Deschênes
UC Santa Barbara, IZA, and NBER

Ruben Gaetani
University of Toronto

Jeffrey Lin
Federal Reserve Bank of Philadelphia

Christopher Severen
Federal Reserve Bank of Philadelphia

This README lists the data and code associated with the project "Lockdowns and Innovation: Evidence from the 1918 Flu Pandemic." We have not shared the raw microdata for the Comprehensive Universe of U.S. Patents (CUSP), which has been developed by Enrico Berkes; instead, we provide intermediate data along with contact information for Enrico Berkes.

Organization of contents

In addition to this README file, this folder contains the following folders:

- /data/ contains input data used in the paper. We describe each of these data sources below.
- /codelog/ contains the code used to construct the analysis data file, the code used to run the analysis, and the log files.
    - Also available as a Github repo at https://github.com/cseveren/PandemicInnovation.
- /output/ contains processed datasets, offered as a convenience to those who wish to replicate our results but do not want to reconstruct the data from scratch.
- /results/ contains output from regression results.

See section below entitled "***Replication instructions***" for detailed instructions on how to replicate results and figures in the paper. This is followed by detailed information about all the files included in the folder.

This README has the following sections:

0. Computational requirements
1. Replication instructions

2. Data files (raw)
3. Data files (analysis)
4. Code and related files
5. Results

Obtaining CUSP Microdata

The Comprehensive Universe of U.S. Patents (CUSP) is described in Berkes (2018), which can also be found through Enrico Berkes' website, https://sites.google.com/view/enricoberkes/home. We have provided intermediate data derived from CUSP, collapsed to city-by-month or city-by-month-by-patent class cells for cities and months in our sample.

To obtain the raw microdata for CUSP, interested researchers can contact:

Enrico Berkes
The Ohio State University
410 Arps Hall
1945 N. High St.
T: +1 202-246-8880
E: berkes.8@osu.edu

As an alternative, to obtain raw microdata specifically for this project, contact:

Chris Severen
E: chris.severen@gmail.com

References

Berkes E. Comprehensive universe of US patents (CUSP): data and facts. Unpublished, Ohio State University. 2018. https://assets.pubpub.org/zlwqau96/51606362190691.pdf

0. Computational requirements

Computational Environment: All code was run on an AWS instance running Windows Server 2016 Datacenter, Version 1607.

- Processor: AMD EPYC 7R13 Processer 2.65 GHz
- 64.0 GB of RAM

Required Software:

- STATA v17.0, with additional packages:
    - Carryforward
    - balancetable
    - blindschemes
    - dataout
    - estout
    - parmest
    - reghdfe
    - regsave
    - ftools
    - texdoc

Additional software required for data construction from scratch:

- Python 3.9 with packages:
    - Levenshtein
    - os
    - pandas
    - csv

Execution Time:

    - Replication code from intermediates executes in under and hour
    - Some data processing times can take longer

1. Replication instructions

All output from the last run of the analysis files are saved in the folder /results/. However, if you would like to replicate the results, you have two options:

Replication from intermediate data

This provides the quickest way to replicate data, as it takes intermediate data as given.

1. Open the master.do file (this is in the top folder of the Github repo, or under ./git/ in the data repo).
2. Edit lines 16 and 17 to reflect the appropriate data and code locations on your machine, then select and execute lines 16 through 23.
3. Execute lines 59-74 to run the code and.
    a. Note that the dataout package called in ./code/analysis/cities_npis_expanded.do fails on some Windows machines .

Replication from raw data

If we have additionally provided you with all source files, then

1. Open the master.do file (this is in the top folder of the Github repo, or under ./git/ in the data repo).
2. Edit lines 16 and 17 to reflect the appropriate data and code locations on your machine.
3. Execute lines 16 through 74.
    a. Note that the dataout package called in ./code/analysis/cities_npis_expanded.do fails on some Windows machines .

2. Data files (raw)

The /input/ folders contains the following folders, each listed with contents, provenance, and details.

/input/city_mortality_CLS/

Includes data from Clay, Lewis, and Severnini (2018, JEH) and Clay, Lewis, and Severnini (2019, EHB).

- ./Clay_Lewis_Severnini_2018JEH.zip is the original zippled replication file for Clay, Lewis, and Severnini (2018). Unzipping it reveals several files, two of which are kept unzipped in this folder for ease of access:
    - ./Clay_Lewis_Severnini_2018JEH_data.dta.
    - ./Clay_Lewis_Severnini_2018JEH_Replication_Files.do.
    - See the readme document in their replication files for details.
    - We don't use this data, but find it helpful for interpreting data from EHB_dataset.dta, which is itself from Clay, Lewis, and Severnini (2019, EHB).
    - To download: https://www.openicpsr.org/openicpsr/project/105720/version/V1/view
- ./EHB_dataset.dta is the replication data from Clay, Lewis, and Severnini (2019, EHB). There are two files derived from this dataset also in this folder:
    - ./city_list.dta contains 492 cities from EHB_dataset.dta for which annual mortality is recorded from 1915-1918.
    - ./cls_cities_geotagged.csv contains the full sample of cities in EHB_dataset.dta with two variables manually added to facilitate downstream data merges:
        - npi_id
        - extra_id
    - To download: https://www.openicpsr.org/openicpsr/project/119843/version/V1/view

/input/controls/

Includes data on age and origin used to create robustness checks. This contains several files and folders:

- ./codes/ is folder containing:
    - ./help_data/ is a folder containing:
        - ./cities_list.csv: List of 50 cities in main sample and state two-letter code.
        - ./state_table.csv: More information at the state-level (obtained through NGGIS), used as an input to "get_cities_info_w_germans.py"
    - get_cities_info_w_germans.py: This code queries the census database on the NBER server. It produces city_infos_w_germans_non_imputed_ages.csv (see below)
- ./city_infos_w_germans_non_imputed_ages.csv: This file contains all the city-level controls extracted using the code "get_cities_info_w_germans.py".
- ./citycontrols_1910.csv: Subset of the first file, with only 43 cities, only 1910 controls, and only a subset of the control variables
- ./citycontrols_1910_expanded.csv: Subset of the first file, with 50 cities, only 1910 controls, and only a subset of the control variables
- ./citycontrols_1910_expanded_avg_age.csv: Same as previous file but also includes average age for each city in 1910 (computed using the other variables in the file)

- ./citycontrols_1910_expanded_avg_age_withid.csv: Same as previous file but also includes city identifier

/input/CUSP/

Includes data from the Comprehensive Universe of U.S. Patents (CUSP) (Berkes, 2018). Files marked *[sup]* are suppressed and available by following the instructions at the beginning of the README.

- ./help_files/ is a folder containing:
    - ./unique_ids/ is a folder containing:
        - *[sup]* Disambiguation of inventor names taken from Berkes, Enrico and Peter, 2022, "Nencka Knowledge Access: The Effects of Carnegie Library on Innovation"
    - ./list_of_states.csv List of US States and their abbreviations.
    - ./places_90_info.csv Taken from 1990 Places shape file downloaded from NHGIS. Downloaded September 2020
- ./cities_list.csv This is the list of cities, states in our sample. Compiled by hand.
- *[sup]* ./inv_patents_assigned_cls_cities.csv Patents matched to EHB cities. File generated using the python code codes/assign_patents_to_cls_cities.py that merges CUSP patent data with EHB data.
- *[sup]* ./inv_patents_assigned_npi_cities.csv Patents matched to NPI cities. File generated using the python code codes/assign_patents_to_npi_cities.py that merges CUSP patent data with NPI data.
- *[sup]* ./patent_assignee_name_location_from_1900.csv Patent assignees and their location. Source: CUSP
- *[sup]* ./patent_cpc_categories_from_1900.csv Includes CPC categories for each patent in our sample. Source: CUSP
- *[sup]* ./patent_fyear_iyear_from_1900.csv Includes filing and issue years for each patent in our sample. Source: CUSP
- *[sup]* ./patent_inventor_name_location_from_1900.csv Includes name and place of residence for each inventor in our sample. Source: CUSP

/input/influenza_collins/

Primarily includes data from Collins, Selwyn D., Wade H. Frost, Mary Gover, and Edgar Sydenstricker (1930). "Mortality from Influenza and Pneumonia in 50 Lage Cities of the United States, 1910-1929." *Public Health Reports* 45 (39): 2277-2328, September 26.

- ./available_cities_cz90_cty90_places90.csv is a common file of basic info on the primary 50 cities in our data that we have saved in many folders for easy reference.
- ./influenza_month_collins.csv is data hand transcribed by Enrico Berkes from Collins et al. (1930).
- ./source.txt details the data source and creation history.

/input/influenza_tycho/

Data originally from https://www.tycho.pitt.edu/ but received by Enrico Berkes from Peter Nencka.

- ./flu-penumonia.csv is data on the weekly case and death counts from the flu
- ./npi_tycho_xwalk is a manually created merge document based on available_cities_cz90_cty90_places90.csv.
- ./source.txt details the data source and creation history.

/input/inventors_census_matches/

Data on inventor age and birthplace. Files marked *[sup]* are suppressed and available by following the instructions at the beginning of the README.

- *[sup]* ./inventors_age.csv Age of inventors in our sample (collected from historical censuses). Source: Samuel Bazzi, Enrico Berkes, Martin Fiszbein, Christian Fons-Rosen, 2022, "Individualism and Innovation", mimeo.
- *[sup]* ./inventors_age_bpl.csv Birthplace of inventors in our sample(collected from historical censuses). Source: Samuel Bazzi, Enrico Berkes, Martin Fiszbein, Christian Fons-Rosen, 2022, "Individualism and Innovation", mimeo.

/input/kpst_breakthrough/

Data on "breakthrough" patents from Kelly, B., Papanikolaou, D., Seru, A. and Taddy, M., 2021. Measuring technological innovation over the long run. *American Economic Review: Insights*, *3*(3), pp.303-20. Accessed on Sept 1, 2021, from https://dimitris-papanikolaou.github.io/website/.

- ./KPST_Breakthrough_v201912.dta is the downloaded data file.
- ./readme.txt details its source

/input/Lilley_et_al_data/

Data from Lilley, Andrew, Matthew Lilley, and Gianluca Rinaldi. "Public health interventions and economic growth: Revisiting the Spanish flu evidence." Available at SSRN 3590008 (2020). Accessed on 08/04/2022 from https://almlgr.github.io/.

- ./1918_npi_effects/ is folder containing the downloaded source material, unzipped.
- ./analyze_Lilley_et_al_data.do is the do file that merges the data and produces the regression results. In the process, it produces:
    - ./merged_lilleyetal.csv
    - ./merged_lilleyetal_filled.csv, which is the same as ./merged_lilleyetal.csv, but with the columns "gen_id", "region", and "days_npi" manually added
- ./readme.txt details the source of the Lilley et al data.

/input/mayoral/

Data collected from various sources on mayoral party affiliation.

- ./city_mayors_filledgaps.csv is manually constructed from mayor_list.xlsx to facilitate Stata read in.
- ./mayor_list.xlsx details sources for each mayor, collected by Nassir Holden and Chris Severen.

/input/migration_of_inventors/

The .do files in the folder create the three .csv datasets, as detailed in ./source.txt

- ./generate_stayers_dataset.do and generate_stayers_dataset_1910_1926.do:
    - These files assign the disambiguated inventors from "CUSP\help_files\unique_ids\patents_inventor_name_location_id_90.csv" to the 50 cities in the sample (through the files "CUSP\inv_patents_assigned_to_npi_cities.csv" and "CUSP\inv_patents_assigned_to_cls_cities.csv"). They then generate a panel of patents by city-month, using only patents filed by "stayers", defined either within the baseline sample (1916-1920) and the extended sample (1910-1926), which are contained in the csv output files in the folder.
- ./generate_usborn_dataset.do:
    - This file assigns the disambiguated inventors from "CUSP\help_files\unique_ids\patents_inventor_name_location_id_90.csv"to the 50 cities in the sample (through the files "CUSP\inv_patents_assigned_to_npi_cities.csv" and "CUSP\inv_patents_assigned_to_cls_cities.csv"). It then generates a panel of patents by city-month, using only patents filed by inventors who are identified as being US-born in the file "inventors_census_matches\inventors_age_bpl.csv". The output file "usborn_city_totpat_fyear_fmonth.csv" is contained in the folder.
- ./stayers_city_totpat_fyear_fmonth.csv, generated from above.
- ./stayers_city_totpat_fyear_fmonth_1910_1926.csv, generated from above.
- ./usborn_city_totpat_fyear_fmonth.csv, generated from above.

/input/mayoral/

Data from and derivatives of Markel, H., H. B. Lipman, J. A. Navarro, A. Sloan, J. R. Michalsen, A. M. Stern, and M. S. Cetron, 2007. “Nonpharmaceutical Interventions Implemented by US Cities during the 1918-1919 Influenza Pandemic,” *JAMA*, 298, 644–654.

- ./npi_info_markel_et_al.csv manually transcribed from Markel et al. (2007).
- ./npi_info_markel_et_al_geotagged.csv the above with geolocation data.
- ./npi_info_markel_et_al_plus.csv the above with the seven additional cities in our data.
- ./sevencity_data_calc.xlsx supporting calculations for NPI length for the seven additional cities in our data.
- ./sevencity_plusready.xlsx formatted version of the results from the above file.
- ./source.txt details the data sources.

/input/population/

Data on city-level populations from 1900-1930 for cities in our sample.

- ./source.txt gives details on two files:
    - ./Decades_1900_to_1930_Wiki_Pops.csv
    - ./Wikipedia_1910_1920_43cities.csv
- ./Decades_1900_50_1930_Wiki_Pops_plus2.csv is the same as Decades_1900_to_1930_Wiki_Pops.csv but with two additional cities manually added

/input/prohibition/

Data on prohibition-era county regulations from ICPSR 8343.

- ./ICPSR_08343/ is the original repository, unzipped.
- ./cities_dry.xlsx and ./cities_dry_wdates.xlsx are manually processed to match our geographies.
- ./source.txt details data sources and processing.

/input/universities/

Source material and derived information relating to university locations in 1916, from Andrews M, 2020. "How do institutions of higher education affect local invention? Evidence from the establishment of US colleges," and appendices to this paper.

- The paper and appendices are from Andrew's website:
    - https://sites.google.com/site/michaeljeffreyandrews/research
- ./city_has_uni_in_1916.xlsx interprets this material, according to:
- ./source.txt which details data processing.

3. Data files (analysis)

The /output/ folder contain files that have been pre-processed. These are the files that are used for the analysis steps. We provide these as a convenience so that the data need not be re-constructed from scratch. The source materials for these files are listed in Section 2. The Build portion of Section 4 details how source materials are combined to create these files. The Analysis portion of Section 4 details how these files are used to create output.

- ./output/intermediate/ is a folder that contains the following files, which predominantly used to create other files in ./output/:
    - ../cities_cls_patents_wclass_19001929.dta
    - ../cities_patents_19001929.dta
    - ../cities_patents_list.dta
    - ../citiesall_patents_age.dta
    - ../citiesall_patents_breakthrough.dta
    - ../citiesall_patents_longclass1d_19001929.dta
    - ../citiesall_patents_longclass3d_19001929.dta
    - ../citiesall_patents_ 19001929.dta
    - ../citiesall_patents_list.dta
    - ../cls_excessmortality_exp.dta
    - ../controls_7cities.dta
    - ../controls_43cities.dta
    - ../npis.dta
    - ../pop_main.dta
- ./output/cities_clspatents.dta
- ./output/cities_clspatents_class.dta
- ./output/cities_expandedpatents.dta
- ./output/cities_expandedpatents_breakthrough.dta
- ./output/cities_expandedpatents_cats.dta
- ./output/cities_expandedpatents_longclass1d.dta
- ./output/cities_expandedpatents_longclass3d.dta
- ./output/cities_patents_npis.dta
- ./output/cities_patents_npis_cats.dta
- ./output/cities_prohibition.dta
- ./output/citiesall_patents_ageyoungold.dta
- ./output/citiesall_patents_migration.dta

4. Code and related files

There are two potential locations that contain scripts to (re)create data and replicate results:

- The ./git/ folder
- The git repo at https://github.com/cseveren/PandemicInnovation

The materials in these folders are identical.

The organization of this folder includes the subfolders below. Note that while files are listed in alphabetical order below, they should be executed in the order shown in ./git/master.do

- ./git/master.do is the master file that will execute in sequence all of the results in the paper, see the Replication section of this README for more details.
- ./git/analysis/ contains code to perform the analysis, mostly relying on the intermediate data in /output/
- ./git/build/ contains code to build the data, populating the intermediate data files in /output/. This relies on some files that are suppressed.

Analysis files

/git/code/analysis/analyze_Lilley_et_al_data.do

- Performs comparison the Lilley et al. (2022), as requested by editor. These results do not appear in the paper itself.
- Lines 1-20 prepare the data from Lilley et al. (2022), which are then manually merged with to our id codes. The output of this process is .\input\Lilley_et_al_data\merged_lilleyetal_filled.csv. This can is the input for lines 31-47.

/git/code/analysis/cities_npis_breakthr_migr_age.do

- Inputs (in code order):
  - ./output/citiesall_patents_ageyoungold.dta
  - ./output/citiesall_patents_migration.dta
  - ./output/cities_expandedpatents_breakthrough.dta
- Creates results for Table 2, Panel C.

/git/code/analysis/cities_npis_cutoffs.do

- Input:
  - ./output/cities_expandedpatents.dta
- Creates Figure A.3 showing estimates that vary the binary definition of long NPI.

/git/code/analysis/cities_npis_eventstudy.do

- Input:
  - ./output/cities_expandedpatents.dta
- Creates Figure 2.

/git/code/analysis/cities_npis_expanded.do

- Inputs (in code order):
    - ./output/cities_expandedpatents.dta
    - ./output/intermediate/controls_43cities.dta
    - ./output/intermediate/controls_7cities.dta
- Creates results for Table A.1.
- Creates results for Table A.2.
- Creates results for Table A.3.
- Creates results for Table A.4.
- Creates results for Table A.6.
- Creates results for Table A.7.
- Also creates summary statistics that are referenced in the text.

/git/code/analysis/cities_npis_expanded_cats.do

- Input:
    - ./output/cities_expandedpatents_cat.dta
- Creates results for Table 2, Panels A and B.

/git/code/analysis/cities_npis_jackknife.do

- Input:
    - ./output/cities_expandedpatents.dta
- Creates Figure A.2.
- Also creates jackknife results that are referenced in the text.

/git/code/analysis/cities_npis_main.do

- Inputs (in code order):
    - ./output/cities_expandedpatents.dta
    - ./output/intermediate/controls_43cities.dta
    - ./output/intermediate/controls_7cities.dta
    - ./output/cities_expandedpatents_longclass1d.dta
    - ./output/cities_expandedpatents_longclass3d.dta
- Creates results for Table 1.

/git/code/analysis/cities_npis_mayors.do

- Inputs (in code order):
    - ./input/mayoral/city_mayors_filledgaps.csv
    - ./output/cities_expandedpatents.dta
- Performs additional exercise requested by editor to test for robustness of results to mayoral affiliation as a potential mechanism.
- Results are not reported in main paper.

/git/code/analysis/cities_npis_prohibition.do

- Inputs (in code order):
    - ./output/cities_expandedpatents.dta
    - ./output/cities_prohibition.dta
- Creates results for Table A.5.

/git/code/analysis/graphs_by_npi_group.do

- Input:
    - ./output/cities_expandedpatents.dta
- Creates Figure 1.

/git/code/analysis/graphs_histogram.do

- Input:
    - ./output/cities_expandedpatents.dta
- Creates Figure A.1.

Build files – Preliminary Steps

/git/code/build/make_clsdata.do

- Input:
    - .\input\city_mortality_CLS\EHB_dataset.dta
    - ./output/intermediate/pop_main
    - ./input/city_mortality_CLS/cls_cities_geotagged.csv
- Output:
    - ./output/intermediate/cls_excessmortality_exp
- Read in and prep city-level variables on experience of the 1918 Pandemic.
- This script also generates estimates of city-level excess mortality, using several methods to impute population.
- Those variables were ultimately unused in the final paper, but is included because of its location in the data production process for future researchers.

/git/code/build/make_controls.do

- Input:
    - ./input/controls/citycontrols_1910_expanded_avg_age_withid.csv
- Output:
    - ./output/intermediate/controls_43cities
    - ./output/intermediate/controls_7cities
- Read in and prep city-level control variables.

/git/code/build/make_cusp_breakthroughs.do

- Input:
    - *[sup]* ./input/kpst_breakthrough/KPST_Breakthrough_v201912.dta
    - *[sup]* .\input\CUSP\patents_cpc_categories_from_1900.csv

    - ./output/intermediate/citiescls_patents_list
    - ./output/intermediate/cities_patents_list
- Output:
    - ./output/intermediate/citiesall_patents_breakthrough
- Read in and prep data on breakthrough patents.
- Two input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_cusp_cities.do

- Input:
    - *[sup]* .\input\CUSP\patents_inventor_name_location_from_1900.csv
    - *[sup]* .\input\CUSP\patents_assignee_name_location_from_1900.csv
    - *[sup]* .\input\CUSP\inv_patents_assigned_to_npi_cities.csv
    - *[sup]* .\input\CUSP\patents_fyear_iyear_from_1900.csv
- Output:
    - ./output/intermediate/cities_patents_list
    - ./output/intermediate/cities_patents_19001929
- Read in and prep CUSP patent data for Markel et al cities
- Four input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_cusp_cities_cls.do

- Input:
    - *[sup]* .\input\CUSP\patents_inventor_name_location_from_1900.csv
    - *[sup]* .\input\CUSP\patents_assignee_name_location_from_1900.csv
    - *[sup]* .\input\CUSP\inv_patents_assigned_to_cls_cities.csv
    - *[sup]* .\input\CUSP\patents_fyear_iyear_from_1900.csv
- Output:
    - ./output/intermediate/citiescls_patents_list
    - ./output/intermediate/citiescls_patents_19001929
- Read in and prep CUSP patent data for Clay, Lewis, and Severnini cities
- Four input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_cusp_cities_cls_patcats.do

- Input:
    - *[sup]* .\input\CUSP\patents_cpc_categories_from_1900.csv
    - ./output/intermediate/citiescls_patents_list
- Output:
    - ./output/intermediate/cities_cls_patents_wclass_19001929
- Read in and prep CUSP patent data for Clay, Lewis, and Severnini cities, with patent classes, for wide uses.

- One input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_cusp_cities_patcats_long1d.do

- Input:
  - *[sup]* .\input\CUSP\patents_cpc_categories_from_1900.csv
  - ./output/intermediate/citiescls_patents_list
  - ./output/intermediate/cities_patents_list
- Output:
  - ./output/intermediate/citiesall_patents_longclass1d_19001929
- Read in and prep patent data by 1-digit class, for long uses.
- One input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_cusp_cities_patcats_long3d.do

- Input:
  - *[sup]* .\input\CUSP\patents_cpc_categories_from_1900.csv
  - ./output/intermediate/citiescls_patents_list
  - ./output/intermediate/cities_patents_list
- Output:
  - ./output/intermediate/citiesall_patents_longclass3d_19001929
- Read in and prep patent data by 3-digit class, for long uses.
- One input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_invage.do

- Input:
  - *[sup]* ./input/inventors_census_matches/inventors_age_bpl.csv
  - *[sup]* .\input\CUSP\patents_cpc_categories_from_1900.csv
  - ./output/intermediate/citiescls_patents_list
  - ./output/intermediate/cities_patents_list
- Output:
  - ./output/intermediate/citiesall_patents_age.dta
- Read in and prep data on inventor age.
- Two input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_migration.do

- Input:
  - *[sup]* ./input/migration_of_inventors/usborn_city_totpat_fyear_fmonth.csv
  - *[sup]* ./input/migration_of_inventors/stayers_city_totpat_fyear_fmonth.csv
  - *[sup]* ./input/migration_of_inventors/stayers_city_totpat_fyear_fmonth_1910_1926.csv
- Output:

    - ./output/citiesall_patents_migration.dta
- Read in and prep data on inventor migration and nativity.
- Three input files are not released here, but the output data is included. Suppressed inputs are available by request, see README introduction.

/git/code/build/make_npis.do

- Input:
    - .\input\NPIs_markel\npi_info_markel_et_al_geotagged.csv
- Output:
    - ./output/intermediate/npis
- Read in and prep transcribed NPI data from Markel et al.

/git/code/build/make_pop.do

- Input:
    - .\input\population\wikipedia\Decades_1900_to_1930_Wiki_Pops_plus2.csv
- Output:
    - ./output/intermediate/pop_main
- Read in and prep population data.

/git/code/build/make_prohibition.do

- Input:
    - ./input/prohibition/cities_dry_wdates.xlsx
- Output:
    - ./output/cities_prohibition.dta
- Read in and prep data on pre-Prohibition era status.

Build files – Intermediate Steps

/git/code/build/prep_cities_clspatents.do

- Input:
    - ./output/intermediate/citiescls_patents_19001929
    - ./output/intermediate/cls_excessmortality_exp
- Output:
    - ./output/cities_clspatents
- Process data on cities in Clay, Lewis, and Severnini, but not in Markel.

/git/code/build/prep_cities_clspatents_cats.do

- Input:
    - ./output/intermediate/cities_cls_patents_wclass_19001929
    - ./output/intermediate/cls_excessmortality_exp
- Output:
    - ./output/cities_clspatents_cat

- Process data on cities in Clay, Lewis, and Severnini, but not in Markel, with data broken out by patent classes.

/git/code/build/prep_cities_npispatents.do

- Input:
  - ./output/intermediate/cities_patents_19001929
  - ./output/intermediate/npis
  - .\input\population\wikipedia\wikipedia_1910_1920_43cities.csv
- Output:
  - ./output/cities_patents_npis
- Process data on cities in Markel et al.

/git/code/build/prep_cities_npispatents_cats.do

- Input:
  - ./output/intermediate/cities_patents_wclass_19001929
  - ./output/intermediate/npis
  - .\input\population\wikipedia\wikipedia_1910_1920_43cities.csv
- Output:
  - ./output/cities_patents_npis
- Process data on cities in Markel et al, with data broken out by patent classes.

Build files – Final Steps

/git/code/build/combine_cities_expanded.do

- Input:
  - ./input/NPIs_markel/sevencity_plusready.csv
  - ./output/cities_clspatents
  - ./output/cities_patents_npis
- Output:
  - ./output/cities_expandedpatents
- Main final dataset on patenting by city.

/git/code/build/combine_cities_expanded_ageyoungold.do

- Input:
  - ./output/intermediate/citiesall_patents_age.dta
  - ./output/cities_expandedpatents
- Output:
  - ./output/citiesall_patents_ageyoungold.dta
- Incorporates processed age data with main datasets.

/git/code/build/combine_cities_expanded_breakthrough.do

- Input:

    - ./output/intermediate/citiesall_patents_breakthrough.dta
    - ./output/cities_expandedpatents
- Output:
    - ./output/cities_expandedpatents_breakthrough.dta
- Incorporates processed breakthrough data with main datasets.

/git/code/build/combine_cities_expanded_cats.do

- Input:
    - ./input/NPIs_markel/sevencity_plusready.csv
    - ./output/cities_clspatents_cats
    - ./output/cities_patents_npis_cats
- Output:
    - ./output/cities_expandedpatents_cats
- Wide-format final data on patents by class.

/git/code/build/combine_cities_expanded_cats_long.do

- Input:
    - ./output/intermediate/citiesall_patents_longclass1d_19001929
    - ./output/intermediate/citiesall_patents_longclass3d_19001929
    - ./output/cities_expandedpatents
- Output:
    - ./output/cities_expandedpatents_longclass1d.dta
    - ./output/cities_expandedpatents_longclass3d.dta
- Long-format final data on patents by class.

5. Results

The folder /results/ contains all results presented in the paper. These are not quite in final table format; that step requires copying and pasting entire rows of the tables of results in /results/. However, numbers do not need to be hand copied.

Results are saved under their original table names, under /results/event_study/, or under /results/fortable/, according to the table on the following page, which is replicated in /results/table_key.csv.

TABLE CORRESPONDENCE AND RESULTS LOCATION

Figure/Table	Code File (./git/…)	Code Lines	Results Files	Notes
Figure 1	./code/analysis/graphs_by_npi_group.do	14-47	./results/Figure1.png	
Figure 2	./code/analysis/cities_npis_eventstudy.do	all	./results/event_study/annual_daysnpis.png ./results/event_study/annual_longnpis.png	
Table 1	./code/analysis/cities_npis_main.do	104-145	./results/fortable/main_duringafterpre_daysnpis.tex ./results/fortable/main_duringafterpre_longnpis.tex ./results/fortable/main_post_effect_daysnpis.tex ./results/fortable/main_post_effect_longnpis.tex	
Table 2, Panels A-B	./code/analysis/cities_npis_expanded_cats.do	39-100	./results/fortable/class_post_daysnpis.tex ./results/fortable/class_post_longnpis.tex ./results/fortable/class_duringafterpre_daysnpis.tex ./results/fortable/class_duringafterpre_longnpis.tex	
Table 2, Panel C	./code/analysis/cities_npis_breakthr_migr_age.do	all	./results/fortable/main_post_effect_daysnpis.tex ./results/fortable/main_post_effect_longnpis.tex	
Figure A.1	./code/analysis/graphs_histogram.do	all	./results/FigureAX_histogramNPI.png	
Figure A.2	./code/analysis/cities_npis_jackknife.do	all	./results/FigureAX_leaveoneout.png	Jackknife results in text
Figure A.3	./code/analysis/cities_npis_cutoffs.do	all	./results/FigureAX_longnpi_cutoffs.png	
Table A.1	./code/analysis/cities_npis_expanded.do	47-63	./results/fortable/summary_stats_outcome.tex ./results/fortable/summary_stats_treatment.tex	
Table A.2	./code/analysis/cities_npis_expanded.do	80-95	./results/fortable/50cities.tex	
Table A.3	./code/analysis/cities_npis_expanded.do	101-123	./results/fortable/balance_***.tex	
Table A.4	./code/analysis/cities_npis_expanded.do	129-179	./results/fortable/robust_all.tex	
Table A.5	./code/analysis/cities_npis_prohibition.do	all	./results/fortable/prohibition_posteffect_longnpis.tex ./results/fortable/prohibition_duringafterpre_longnpis.tex	
Table A.6	./code/analysis/cities_npis_expanded.do	186-212	./results/fortable/robustpats_posteffect_longnpis.tex ./results/fortable/robustpats_duringafterpre_longnpis.tex	
Table A.7	./code/analysis/cities_npis_expanded.do	218-321	./results/fortable/poprobust_baseline_longnpis.tex ./results/fortable/poprobust_nopop_longnpis.tex ./results/fortable/poprobust_popcontrol_longnpis.tex ./results/fortable/poprobust_1910pop_longnpis.tex ./results/fortable/poprobust_1910popcontrol_longnpis.tex	
Local Political Affiliation	./code/analysis/cities_npis_mayors.do	all		Response to editor
Comparison to Lilley et al	./code/analysis/analyze_Lilley_et_al_data.do	all		Response to editor