

README: Federal Research Funding and STEM Education

Replication Package Documentation

1. Overview

This document describes the data and code needed to reproduce all figures and tables in "Federal Research Funding and STEM Education." The analysis uses Stata 18.0 MP on a Windows PC.

2. Statement of Rights

The authors certify they have legitimate access to and permission to use all data in this manuscript.

3. Software Requirements

- Stata 18.0 MP (Windows, 64-bit). Earlier versions may work but are not tested.
- Required user-written packages: `outreg2`, `reghdfe`, `ftools`, `coefplot`, `xlincom`, `statstates`, `missings` (install via `ssc install`).

4. Steps to Reproduce Results

1. Set the working directory (global root) in each `.do` file to match the local path of the replication package.
2. Download all raw data listed in Section 6 and place them in the folder paths indicated.
3. Run `do`-files in the following order:

- a. Build intermediate datasets by running all the codes in folder:

- `Codes/Construct_Admissions`
- `Codes/Construct_HEGIS`
- `Codes/Construct_IPEDS`

Note: Please note that the code in `Codes/Construct_HEGIS` cannot be executed using only the provided materials, as the underlying HEGIS microdata are restricted-use and excluded from this replication package. Users must obtain raw data independently. Please refer to `HEGISDataDownloadGuide.pdf` provided herein for guidance.

- b. `01_ProgramsDataFile.do`
- c. `02_FSSSupportSurvey.do`
- d. `03_CombineACTSAT.do`
- e. `04_ConstructDatasetforRegressions.do`

- f. 05a_Regressions.do
- g. 05b_Regressions.do
- h. 06_CalculateNumbers.do.

5. Folder Structure

- Codes/
 - Construct_Admissions/ —codes to build College Handbook/Admission Data
 - Construct_AdmissionData —build CH_CPALL.dta
 - Construct_HEGIS/ —codes to build HEGIS Dataset (cannot be run; see note in Step 3)
 - Construct_HEGIS_Degrees —build degrees.dta
 - Construct_HEGIS_Finance —build HEGIS_finance_limit.dta
 - Construct_HEGIS_Instchar —build HEGIS_InstChar.dta
 - Construct_IPEDS/ —codes to build IPEDS Dataset
 - Construct_AllIPEDSData —build AllIPEDSData.dta and AllIPEDSData22.dta
 - 01_ProgramsDataFile.do
 - 02_FSSSupportSurvey.do
 - 03_CombineACTSAT.do
 - 04_ConstructDatasetforRegressions.do
 - 05a_Regressions.do
 - 05b_Regressions.do
 - 06_CalculateNumbers.do
 - nameclean.ado
- Data/
 - Raw/ — unmodified raw data files
 - FSS_Support_Survey/ — NSF FSS microdata and pre-1971 Excel files
 - HEGIS_Degrees/ — Raw HEGIS degree completions annual files. Download links are provided in Section 6; files are not included in this package as redistribution is prohibited under ICPSR terms of use.
 - HEGIS_Finance/ — Raw HEGIS finance annual files. Download links are provided in Section 6; files are not included in this package as redistribution is prohibited under ICPSR terms of use.
 - HEGIS_InstitutionalChars/ — Raw HEGIS institutional characteristics annual files. Download links are provided in Section 6; files are not included in this package as redistribution is prohibited under ICPSR terms of use.
 - IPEDS_Degrees/ — Raw IPEDS degree completions annual files
 - IPEDS_InstitutionalChars/ — Raw IPEDS Institutional Characters annual files
 - IPEDS_Finance/ — Raw IPEDS Finance annual files
 - Processed/ — Intermediate cleaned/transformed files

- Temp/ — Temporary files (safe to delete after running code)
- Final/ — Final datasets used as direct inputs to regressions
- Figures/ — All figures presented in the paper
- Tables/ — All tables presented in the paper

6. Data Availability and Sources

The table below lists all raw input datasets with full bibliographic information. For datasets downloaded from URLs, the link points to the exact page or file used. Access conditions are noted under License.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
Institution-year admissions data compiled from the College Board's College Handbook, covering 24 academic years. Each sheet in the Excel file corresponds to one year and contains institution-level SAT and ACT score distributions. This file was assembled by the authors from annual editions of the College Handbook. Location: Data/Raw All Admissions Data.xlsx (Sheets 3–26, one sheet per year)	College Board, College Handbook, various years	-	-	College Board. The College Handbook, various editions. New York: College Board. Data digitized by the authors from print volumes.	This file was manually digitized by the authors from print editions of the College Board's College Handbook. The original print volumes are not available in digital form.
CIP_HEGIS_Crosswalk_1981/03135-0001-Data.txt A fixed-width text file downloaded from ICPSR that maps old HEGIS taxonomy codes to their corresponding CIP codes and program titles. Used in Construct_HEGIS_Degrees.do to construct a crosswalk between HEGIS subject codes and CIP codes for the 1983 and 1984 survey years, which were the transitional years when HEGIS switched from its own taxonomy to the CIP classification system.	Inter-university Consortium for Political and Social Research (ICPSR), Higher Education General Information Survey (HEGIS)	https://www.icpsr.umich.edu/web/ICPSR/studies/3135	May 2017	United States Department of Education. National Center for Education Statistics. Crosswalk Between CIP and HEGIS Taxonomy, 1981. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2003-12-02. https://doi.org/10.3886/ICPSR03135.v1	ICPSR Terms of Use; open access download, not open redistribution. We provide download links for all raw datasets but are unable to distribute the files directly, as redistribution is prohibited under the applicable terms of use.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
Location: Data/Raw/HEGIS_Degrees/CIP_HEGIS_Crosswalk_1981/ 03135-0001-Data.txt					
Raw annual HEGIS microdata files downloaded from ICPSR, one per survey year. Each file contains one record per institution–field–degree level combination, reporting the number of degrees conferred broken down by sex (male/female) across bachelor's, master's, doctoral, and first-professional degrees. Location: Data/Raw/HEGIS_Degrees/{year}_ICPSR_{dataset no} 02081-0001-Data.txt (HEGIS I, 1965-66) 02082-0001-Data.txt (HEGIS II, 1966-67) 02083-0001-Data.txt (HEGIS III, 1967-68) 02084-0001-Data.txt (HEGIS IV, 1968-69) 02138-0001-Data.txt (HEGIS VI, 1970-71) 02139-0001-Data.dta (HEGIS VII, 1971-72) 02140-0001-Data.dta (HEGIS VIII, 1972-73) 02141-0001-Data.txt (HEGIS IX, 1973-74) 02143-0001-Data.txt (HEGIS XI, 1975-76) 07651-0001-Data.txt (HEGIS, 1976-77) 02145-0001-Data.txt (HEGIS, 1977-78) 02146-0001-Data.txt (HEGIS XIV, 1978-79) 02147-0001-Data.txt (HEGIS, 1979-80) 02148-0001-Data.txt (HEGIS, 1980-81) 08287-0001-Data.txt (HEGIS, 1981-82) 02157-0001-Data.txt (HEGIS, 1983-84) 02158-0001-Data.txt (HEGIS, 1984-85)	Inter-university Consortium for Political and Social Research (ICPSR), Higher Education General Information Survey (HEGIS)	https://www.icpsr.umich.edu/web/ICPSR/studies/2081/version/s/V1 https://www.icpsr.umich.edu/web/ICPSR/studies/2082/version/s/V1 https://www.icpsr.umich.edu/web/ICPSR/studies/2083/variables https://www.icpsr.umich.edu/web/ICPSR/studies/02084/publications https://www.icpsr.umich.edu/web/ICPSR/studies/2138/publications https://www.icpsr.umich.edu/web/ICPSR/studies/2139/summary https://www.icpsr.umich.edu/web/ICPSR/studies/2140/summary https://www.icpsr.umich.edu/web/ICPSR/	May 2017	United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) I: Degrees and Other Formal Awards Conferred Between July 1, 1965, and June 30, 1966 . Inter-university Consortium for Political and Social Research [distributor], 2004-09-16. https://doi.org/10.3886/ICPSR02081.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) II: Degrees and Other Formal Awards Conferred Between July 1, 1966, and June 30, 1967 . Inter-university Consortium for Political and Social Research [distributor], 2004-07-26. https://doi.org/10.3886/ICPSR02082.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) III, Degrees and Other Formal Awards Conferred Between July 1, 1967 and June 30, 1968. Inter-university Consortium for Political and Social Research [distributor], 2005-09-30. https://doi.org/10.3886/ICPSR02083.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) IV: Degrees and Other Formal Awards Conferred Between July 1, 1968 and June 30, 1969. Inter-university Consortium for Political and Social	ICPSR Terms of Use; open access download, not open redistribution. We provide download links for all raw datasets but are unable to distribute the files directly, as redistribution is prohibited under the applicable terms of use.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
		studies/2141/summary https://www.icpsr.umich.edu/web/ICPSR/studies/2143/summary https://www.icpsr.umich.edu/web/ICPSR/studies/7651/summary https://www.icpsr.umich.edu/web/ICPSR/studies/2145/summary https://www.icpsr.umich.edu/web/ICPSR/studies/2146/summary https://www.icpsr.umich.edu/web/ICPSR/studies/2147 https://www.icpsr.umich.edu/web/ICPSR/studies/2148 https://www.icpsr.umich.edu/web/ICPSR/studies/8287 https://www.icpsr.umich.edu/web/ICPSR/studies/2157 https://www.icpsr.umich.edu/web/ICPSR/studies/2158/publications		Research [distributor], 2005-11-04. https://doi.org/10.3886/ICPSR02084.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) VI: Degrees and Other Formal Awards Conferred Between July 1, 1970, and June 30, 1971. Inter-university Consortium for Political and Social Research [distributor], 2005-11-04. https://doi.org/10.3886/ICPSR02138.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) VII: Degrees and Other Formal Awards Conferred Between July 1, 1971, and June 30, 1972. Inter-university Consortium for Political and Social Research [distributor], 2005-11-14. https://doi.org/10.3886/ICPSR02139.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) VIII: Degrees and Other Formal Awards Conferred Between July 1, 1972, and June 30, 1973. Inter-university Consortium for Political and Social Research [distributor], 2005-12-06. https://doi.org/10.3886/ICPSR02140.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) IX: Degrees and Other Formal Awards Conferred Between July 1, 1973, and June 30, 1974. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004-07-09. https://doi.org/10.3886/ICPSR02141.v1	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XI: Degrees and Other Formal Awards Conferred Between July 1, 1975, and June 30, 1976. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004-08-12. https://doi.org/10.3886/ICPSR02143.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1976-1977: Earned Degrees . Inter-university Consortium for Political and Social Research [distributor], 2000-02-01. https://doi.org/10.3886/ICPSR07651.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1977-1978: Earned Degrees . Inter-university Consortium for Political and Social Research [distributor], 1999-11-19. https://doi.org/10.3886/ICPSR02145.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XIV: Degrees and Other Formal Awards Conferred Between July 1, 1978, and June 30, 1979. Inter-university Consortium for Political and Social Research [distributor], 2004-08-12. https://doi.org/10.3886/ICPSR02146.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1979-1980: Earned Degrees . Inter-university Consortium for	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				Political and Social Research [distributor], 1999-02-25. https://doi.org/10.3886/ICPSR02147.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1980-1981: Earned Degrees . Inter-university Consortium for Political and Social Research [distributor], 1999-02-25. https://doi.org/10.3886/ICPSR02148.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1981-1982: Degrees Conferred. Inter-university Consortium for Political and Social Research [distributor], 1992-02-16. https://doi.org/10.3886/ICPSR08287.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1983-1984: Earned Degrees . Inter-university Consortium for Political and Social Research [distributor], 1999-02-25. https://doi.org/10.3886/ICPSR02157.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1984-1985: Earned Degrees . Inter-university Consortium for Political and Social Research [distributor], 1999-03-18. https://doi.org/10.3886/ICPSR02158.v1	
Author-constructed lookup tables mapping HEGIS subject codes (ITEMNUM) to field names. Each file covers a distinct range of survey years, reflecting changes in HEGIS's	-	-	-	-	Created by the authors; no upstream license restrictions apply.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
subject coding taxonomy over time. Used to label degree records before HEGIS switched to CIP codes in 1983. Location: Data/Raw/HEGIS_Degrees lookupfields_65_67.dta lookupfields_68.dta lookupfields_70_81.dta					
Author-constructed crosswalk files mapping state identifier codes to state names, used to standardise geographic identifiers across HEGIS survey years. Location: Data/Raw/HEGIS_Degrees states.dta maps the OESTATE variable (used in 1965–1967) states2.dta maps the GEOGCODE variable (used from 1973 onward).	-	-	-	-	Created by the authors; no upstream license restrictions apply.
Raw annual HEGIS Finance microdata files downloaded from ICPSR, one per survey year. Each file contains institution-level data on revenues and expenditures. Location: Data/Raw/HEGIS_Finance/{year}_Finance_{dataset no} 02094-0001-Data.txt (1968-69) 02095-0001-Data.txt (1969-70) 02096-0001-Data.txt (1970-71) 02097-0001-Data.txt (1971-72) 02098-0001-Data.txt (1972-73) 02099-0001-Data.txt (1973-74) 02100-0001-Data.txt (1974-75) 07649-0001-Data.txt (1975-76) 02102-0001-Data.dta (1976-77)	Inter-university Consortium for Political and Social Research (ICPSR), Higher Education General Information Survey (HEGIS)	https://www.icpsr.umich.edu/web/ICPSR/studies/2094 https://www.icpsr.umich.edu/web/ICPSR/studies/2095 https://www.icpsr.umich.edu/web/ICPSR/studies/2096 https://www.icpsr.umich.edu/web/ICPSR/studies/2097 https://www.icpsr.umich.edu/web/ICPSR/studies/2098	June 2017	United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1968-1969: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02094.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1969-1970: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02095.v1	ICPSR Terms of Use; open access download, not open redistribution. We provide download links for all raw datasets but are unable to distribute the files directly, as redistribution is prohibited under the applicable terms of use.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
02103-0001-Data.dta (1977-78) 02104-0001-Data.dta (1978-79) 02105-0001-Data.dta (1979-80) 02106-0001-Data.dta (1980-81) 08290-0001-Data.dta (1981-82) 02107-0001-Data.dta (1982-83) 08514-0001-Data.dta (1983-84) 02109-0001-Data.txt (1984-85) 02110-0001-Data.txt (1985-86)		https://www.icpsr.umich.edu/web/ICPSR/studies/2099 https://www.icpsr.umich.edu/web/ICPSR/studies/2100 https://www.icpsr.umich.edu/web/ICPSR/studies/7649 https://www.icpsr.umich.edu/web/ICPSR/studies/2102 https://www.icpsr.umich.edu/web/ICPSR/studies/2103 https://www.icpsr.umich.edu/web/ICPSR/studies/2104 https://www.icpsr.umich.edu/web/ICPSR/studies/2105 https://www.icpsr.umich.edu/web/ICPSR/studies/2106 https://www.icpsr.umich.edu/web/ICPSR/studies/8290 https://www.icpsr.umich.edu/web/ICPSR/studies/2107		United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1970-1971: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02096.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1971-1972: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02097.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1972-1973: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02098.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS), 1973-1974: Financial Statistics. Inter-university Consortium for Political and Social Research [distributor], 1998-12-17. https://doi.org/10.3886/ICPSR02099.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) X: Financial Statistics of Institutions of Higher Education for Fiscal Year Ending 1975 . Inter-university Consortium for Political and Social	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
		https://www.icpsr.umich.edu/web/ICPSR/studies/8514 https://www.icpsr.umich.edu/web/ICPSR/studies/2109 https://www.icpsr.umich.edu/web/ICPSR/studies/2110		Research [distributor], 2004-10-18. https://doi.org/10.3886/ICPSR02100.v1 United States Department of Education. National Center for Education Statistics. Higher Education Finance, 1976. Inter-university Consortium for Political and Social Research [distributor], 1998-12-23. https://doi.org/10.3886/ICPSR07649.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XII: Financial Statistics of Institutions of Higher Education for Fiscal Year Ending 1977. Inter-university Consortium for Political and Social Research [distributor], 2007-05-02. https://doi.org/10.3886/ICPSR02102.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XIII: Financial Statistics of Institutions of Higher Education for Fiscal Year Ending 1978. Inter-university Consortium for Political and Social Research [distributor], 2007-06-21. https://doi.org/10.3886/ICPSR02103.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XIV: Financial Statistics of Institutions of Higher Education for Fiscal Year 1979 . Inter-university Consortium for Political and Social Research [distributor], 2005-02-03. https://doi.org/10.3886/ICPSR02104.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XV: Financial Statistics	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				of Institutions of Higher Education for Fiscal Year 1980 . Inter-university Consortium for Political and Social Research [distributor], 2005-03-04. https://doi.org/10.3886/ICPSR02105.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XVI: Financial Statistics of Institutions of Higher Education for Fiscal Year 1981 . Inter-university Consortium for Political and Social Research [distributor], 2005-03-04. https://doi.org/10.3886/ICPSR02106.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XVIII: Financial Statistics of Institutions of Higher Education for Fiscal Year 1983 . Inter-university Consortium for Political and Social Research [distributor], 2005-04-15. https://doi.org/10.3886/ICPSR02107.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XIX: Financial Statistics of Institutions of Higher Education for Fiscal Year 1984 . Inter-university Consortium for Political and Social Research [distributor], 2005-02-03. https://doi.org/10.3886/ICPSR08514.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XX: Financial Statistics of Institutions of Higher Education for Fiscal Year 1985 . Inter-university Consortium for Political and Social Research	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				[distributor], 2004-10-29. https://doi.org/10.3886/ICPSR02109.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XXI: Financial Statistics of Institutions of Higher Education for Fiscal Year 1986. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2004-11-24. https://doi.org/10.3886/ICPSR02110.v1	
Raw annual HEGIS Institutional Characteristics microdata files downloaded from ICPSR, one per survey year. Each file contains institution-level data on tuition, fees, room and board, control type, and affiliation. Note that only selected years are used in the analysis (1970, 1972–1974, 1977–1978, 1983). Location: Data/Raw/HEGIS_InstitutionalChars/{year}_InstChar_ICPSR_{dataset no} 02111-0001-Data.dta (1970) 02042-0001-Data.dta (1972) 02043-0001-Data.dta (1973) 02044-0001-Data.dta (1974) 07647-0001-Data.dta (1977) 06903-0001-Data.dta (1978) 08291-0001-Data.dta (1983)	Inter-university Consortium for Political and Social Research (ICPSR), Higher Education General Information Survey (HEGIS)	https://www.icpsr.umich.edu/web/ICPSR/studies/2111 https://www.icpsr.umich.edu/web/ICPSR/studies/2042 https://www.icpsr.umich.edu/web/ICPSR/studies/2043 https://www.icpsr.umich.edu/web/ICPSR/studies/2044 https://www.icpsr.umich.edu/web/ICPSR/studies/7647 https://www.icpsr.umich.edu/web/ICPSR/studies/6903 https://www.icpsr.umich.edu/web/ICPSR/studies/8291	June 2017	United States Department of Education. Institute of Education Sciences. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) V: Institutional Characteristics of Colleges and Universities, 1970-1971. Inter-university Consortium for Political and Social Research [distributor], 2005-09-12. https://doi.org/10.3886/ICPSR02111.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) VII: Institutional Characteristics of Colleges and Universities, 1972-1973. Inter-university Consortium for Political and Social Research [distributor], 2005-12-15. https://doi.org/10.3886/ICPSR02042.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) VIII: Institutional Characteristics of Colleges and Universities, 1973-1974. Inter-university Consortium for Political and Social Research [distributor], 2005-09-22. https://doi.org/10.3886/ICPSR02043.v2	ICPSR Terms of Use; open access download, not open redistribution. We provide download links for all raw datasets but are unable to distribute the files directly, as redistribution is prohibited under the applicable terms of use.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) IX: Institutional Characteristics, 1974-1975. Inter-university Consortium for Political and Social Research [distributor], 2005-10-26. https://doi.org/10.3886/ICPSR02044.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XII: Institutional Characteristics of Colleges and Universities, 1977-1978. Inter-university Consortium for Political and Social Research [distributor], 2005-11-14. https://doi.org/10.3886/ICPSR07647.v2 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XIII: Institutional Characteristics of Colleges and Universities, 1978-1979. Inter-university Consortium for Political and Social Research [distributor], 2006-01-06. https://doi.org/10.3886/ICPSR06903.v1 United States Department of Education. National Center for Education Statistics. Higher Education General Information Survey (HEGIS) XVIII: Institutional Characteristics of Colleges and Universities, 1983-1984. Inter-university Consortium for Political and Social Research [distributor], 2005-11-22. https://doi.org/10.3886/ICPSR08291.v2	
Raw annual IPEDS Completions Survey microdata files downloaded from the NCES IPEDS Data Center, one per survey year. Each file contains one record per institution–CIP code–degree level combination,	NCES, Integrated Postsecondary Education Data System (IPEDS)	Go to https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx	June 2017	United States Department of Education. National Center for Education Statistics. Integrated Postsecondary Education Data System (IPEDS), Completions Survey, 1987–2020. U.S. Department of Education [distributor].	Open access

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
reporting the number of degrees conferred broken down by sex. Location: Data/Raw/IPEDS_Degrees dct_c1987.dta dct_c1988.dta dct_c1989.dta dct_c1990.dta dct_c1991.dta dct_c1992.dta dct_c1993.dta dct_c1994.dta dct_c1995.dta dct_c1996.dta dct_c1997.dta dct_c1998.dta dct_c1999.dta dct_c2000.dta dct_c2001.dta dct_c2002.dta dct_c2003.dta dct_c2004.dta dct_c2005.dta dct_c2006.dta dct_c2007.dta dct_c2008.dta dct_c2009.dta dct_c2010.dta dct_c2011.dta dct_c2012.dta dct_c2013.dta dct_c2014.dta dct_c2015.dta dct_c2016.dta dct_c2017.dta dct_c2018.dta dct_c2019.dta dct_c2020.dta		Under the year dropdown, select the desired year (repeat for each year from 1987 to 2020) Download the Stata dataset for the Completions survey. Select "Awards/degrees conferred by program (6-digit CIP code), award level, race/ethnicity, and gender" Place each downloaded file into Data/Raw/IPEDS_Degrees/ and rename it to dct_c{year}.dta (e.g. dct_c1987.dta for the 1987 file)		https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx	
Raw annual IPEDS Finance Survey microdata files downloaded from the NCES	NCES, Integrated	Go to https://nces.ed.gov/i	June 2017	United States Department of Education. National Center for Education Statistics.	Open access

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
IPEDS Data Center, one per survey year (1987–2022). Each file contains institution-level data on revenues and expenditures. Note that the file naming convention changed across periods and some years are split across multiple part files that are merged or appended in the code. Location: Data/Raw/IPEDS_Finance dct_f1987_a, dct_f1987_b, dct_f1987_e (1987, three parts) dct_f1988 (1988, single file) dct_f{year}_a, dct_f{year}_b, dct_f{year}_e (1989–1996, three parts each) dct_f9697_f1, dct_f9697_f2 (1997, two parts) dct_f9798_f1, dct_f9798_f2 (1998, two parts) dct_f9899_f1, dct_f9899_f2 (1999, two parts) dct_f9900_f1, dct_f9900f2 (2000, two parts) dct_f0001_f1, dct_f0001_f2 (2001, two parts) dct_f0102_f1, dct_f0102_f2, dct_f0102_f1a (2002, three parts) dct_f0203_f1, dct_f0203_f2, dct_f0203_f1a (2003, three parts) dct_f0304_f1a, dct_f0304_f2 (2004, two parts)	Postsecondary Education Data System (IPEDS)	peds/datacenter/DataFiles.aspx Under the year dropdown, select the desired year (repeat for each year from 1987 to 2022) Download all Stata dataset files for the Finance survey component. Note that depending on the year, the Finance survey may be split into multiple part files (e.g. _f1, _f2, _f1a). Download all parts for each year Place all downloaded files into Data/Raw/IPEDS_Finance/ and rename them for simplicity		Integrated Postsecondary Education Data System (IPEDS), Finance Survey, 1987–2022. U.S. Department of Education [distributor]. https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx	

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
dct_f0405_f1a, dct_f0405_f2 (2005, two parts) dct_f0506_f1a, dct_f0506_f2 (2006, two parts) dct_f0607_f1a, dct_f0607_f2 (2007, two parts) dct_f0708_f1a, dct_f0708_f2 (2008, two parts) dct_f0809_f1a, dct_f0809_f2 (2009, two parts) dct_f0910_f1a, dct_f0910_f2 (2010, two parts) dct_f1011_f1a, dct_f1011_f2 through dct_f2223_f1a, dct_f2223_f2 (2011–2022, two parts each)					
Raw annual IPEDS Institutional Characteristics Survey microdata files downloaded from the NCES IPEDS Data Center, one per survey year (1987–2023). Each file contains institution-level data on tuition, fees, control type, and other institutional attributes. Files are split across multiple parts per year and the naming convention changed substantially over time. Location: Data/Raw/IPEDS_InstitutionalChars dct_ic1987_a, dct_ic1987_b (1987, two parts) dct_ic1988_a, dct_ic1988_b (1988, two parts) dct_ic1989_a, dct_ic1989_b (1989, two parts)	NCES, Integrated Postsecondary Education Data System (IPEDS)	Go to https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx Under the year dropdown, select the desired year (repeat for each year from 1987 to 2023) Download all Stata dataset files for the following survey components: Institutional Characteristics: all available part files (the number of parts varies by year, ranging from two to seven)	June 2017	United States Department of Education. National Center for Education Statistics. Integrated Postsecondary Education Data System (IPEDS), Institutional Characteristics Survey, 1987–2023. U.S. Department of Education [distributor]. https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx	Open access

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
dct_ic1990_hdr, dct_ic90abce, dct_ic90d (1990, three parts) dct_ic1991_hdr, dct_ic1991_ab, dct_ic1991_c, dct_ic1991_d, dct_ic1991_e, dct_ic1991_f, dct_ic1991_other (1991, seven parts) dct_ic1992_a, dct_ic1992_b (1992, two parts) dct_ic1993_a, dct_ic1993_b (1993, two parts) dct_ic1994_a, dct_ic1994_b (1994, two parts) dct_1995hd, dct_ic9596_b (1995, two parts) dct_ic9697_a, dct_ic9697_b, dct_ic9697_c (1996, three parts) dct_ic9798_hdr, dct_ic9798_ab, dct_ic9798_c, dct_ic9798_d, dct_ic9798_e, dct_ic9798_f (1997, six parts) dct_ic98hdac, dct_ic98_ab, dct_ic98_c, dct_ic98_d, dct_ic98_e, dct_ic98_f (1998, six parts) dct_ic99_hd, dct_ic99abcf, dct_ic99_d, dct_ic99_e, dct_ic99_actot (1999, five parts) dct_hd2000, dct_ic2000, dct_ic2000_ay, dct_ic2000_py, dct_ic2000_actot (2000, five parts) dct_hd{year}, dct_ic{year}, dct_ic{year}_ay, dct_ic{year}_py (2001–2013, four parts each)		Admissions: for years 2014–2023 only Place all downloaded files into Data/Raw/IPEDS_InstitutionalChars/ and rename them for simplicity			

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
dct_hd{year}, dct_ic{year}, dct_ic{year}_ay, dct_ic{year}_py, dct_adm{year} (2014–2023, five parts each)					
All CIP.xls A single Excel file containing all CIP code lists and crosswalks for the 1985, 1990, and 2000 taxonomies, downloaded from the NCES CIP resources page and renamed from CIP.xls to All CIP.xls. Only the sheet Crosswalk_CIP85toCIP90 is used in the analysis, to harmonize IPEDS completions data for years before 1992 which were coded under the 1985 CIP taxonomy. Location: Data/Raw	NCES CIP Code Resources	Go to https://nces.ed.gov/ipeds/cipcode/resources.aspx?y=55 Click "MS Excel version of all CIP 1985-2000 tables (686 KB)" Extract the downloaded cip.zip file Rename CIP.xls to All CIP.xls and place it in Data/Raw/	September 2024	National Center for Education Statistics, CIP Code Resources Citation: National Center for Education Statistics. Classification of Instructional Programs (CIP): All CIP 1985–2000 Tables. U.S. Department of Education. https://nces.ed.gov/ipeds/cipcode/resources.aspx?y=55	Open access
2000aligncip.dta Author-constructed crosswalk mapping 2-digit CIP codes (2000 taxonomy) to the broad major group classification used in the analysis. Assigns each CIP to one of 7 broad categories: Arts and Architecture, Business and Communications, Education, Engineering/Math/Sciences, Health, Social Science and Humanities, and Vocational/Other. Used throughout the IPEDS completions processing (1987–2020) to consistently group degrees into the broad fields used in the regressions. Location: Data/Raw	-	-	-	-	Created by the authors; no upstream license restrictions apply.
Microdata files from the Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions (NSF FSS), containing records of federal research funding obligations by institution and agency (1971–present).	National Science Foundation, National Center for Science and Engineering	-	September 2025	National Science Foundation, National Center for Science and Engineering Statistics. Survey of Federal Science and Engineering Support to Universities, Colleges, and Nonprofit Institutions,	Obtained directly from NSF staff. No data use agreement was required.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
data_obligation.dta contains the main funding records, c_agency.dta contains agency codes and labels, and c_inst_codebook.dta contains institution identifiers including name, FICE code, state, and institution type. Location: Data/Raw/FSS_Support_Survey/Microdata_stata_export data_obligation.dta c_agency.dta c_inst_codebook.dta	Statistics (NCSES)			microdata, 1971–present. Data provided by NSF staff.	
Author-digitized Excel files containing institution-level total federal research funding data for fiscal years 1963–1970, manually transcribed from scanned NSF annual reports. The original print reports are publicly available through HathiTrust and ERIC but no digital data version exists. Location: Data/Raw/FSS_Support_Survey/FSS_support_preperiod Federal Support to Universities and Colleges_1963-1966.xlsx (FY 1963–1966) Federal Support to Universities and Colleges Fiscal Year 1967.xlsx (FY 1967) Federal Support to Universities and Colleges Fiscal Year 1968.xlsx (FY 1968) 1969 Totals _William_2-16-22.xlsx (FY 1969) 1970 Totals _William_2-16-22.xlsx (FY 1970)	National Science Foundation — Federal Support to Universities and Colleges scanned reports (digitized by authors)	https://catalog.hathitrust.org/Record/101711191 https://eric.ed.gov/?id=ED029601 https://babel.hathitrust.org/cgi/pt?id=osu.32435023876337 https://babel.hathitrust.org/cgi/pt?id=mdp.39015024488895 https://babel.hathitrust.org/cgi/pt?id=mdp.39015075988447	July 2024	National Science Foundation. Federal Support to Universities and Colleges, Fiscal Years 1963–66. https://catalog.hathitrust.org/Record/101711191 National Science Foundation. Federal Support to Universities and Colleges, Fiscal Year 1967. https://eric.ed.gov/?id=ED029601 National Science Foundation. Federal Support to Universities and Colleges, Fiscal Year 1968. https://babel.hathitrust.org/cgi/pt?id=osu.32435023876337 National Science Foundation. Federal Support to Universities, Colleges, and Selected Nonprofit Institutions, Fiscal Year 1969. https://babel.hathitrust.org/cgi/pt?id=mdp.39015024488895 National Science Foundation. Federal Support to Universities, Colleges, and Selected Nonprofit Institutions, Fiscal Year 1970.	Original reports are public domain government documents. Digitized Excel files created by the authors; no upstream license restrictions apply.

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
				https://babel.hathitrust.org/cgi/pt?id=mdp.39015075988447	
Institution-level data from the 1994 edition of the Carnegie Classification of Institutions of Higher Education, used to identify and filter the analysis sample to doctoral/research universities. Contains institution identifiers including FICE code, UNITID, institution name, and state. Location: Data/Raw Carnegie_1994_edition_data.xls Carnegie_1994_edition_data.csv	Carnegie Foundation for the Advancement of Teaching. A Classification of Institutions of Higher Education, 1994 Edition.	https://carnegieclassifications.acenet.edu/ Note: 1994 version is no longer available online	July 2021	Carnegie Foundation for the Advancement of Teaching. A Classification of Institutions of Higher Education, 1994 Edition. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching, 1994.	Open access
Annual CPI multiplier files used to deflate nominal dollar values to real dollars, constructed from Minneapolis Fed source. Location: Data/Raw CPI_long_2019.dta: deflates to real 2019 dollars. Used in Construct_HEGIS_Instchar.do to deflate HEGIS-era tuition, room, and board variables CPI_long_2022.dta: deflates to real 2022 dollars. Used in Construct_AllIPEDSData.do and 04_ConstructDatasetforRegressions.do to deflate IPEDS finance variables and FSS federal funding.	Federal Reserve Bank of Minneapolis / U.S. Bureau of Labor Statistics	Go to https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913- Retrieved the CPI table. It contains annual CPI values from 1913 onward Compute the multiplier as: multiplier = CPI_baseyear / CPI_year where the base year is 2019 or 2022. This converts a value in nominal year t dollars to real base-year dollars when multiplied Save as a Stata .dta file with variables	February 2020 May 2024	Federal Reserve Bank of Minneapolis. Consumer Price Index, 1913–present. Retrieved from https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1913-	Open access

Description and Filename(s)	Source	URL/Access Instruction	Access Date	Citation	License
		year and multiplier2019 (or multiplier 2022), one observation per year			
nameclean.ado Author-written Stata program used across multiple scripts (02_FSSSupportSurvey.do, 01_ProgramsDataFile.do, and others) to standardize institution names for merging across datasets. It takes a dataset containing instnm (raw institution name) and college_state (state abbreviation) as inputs and produces a cleaned variable instnm_clean. It works by (1) converting all names to uppercase and removing special characters; (2) standardizing common institutional title abbreviations; (3) replacing full state names with two-letter abbreviations; and (4) applying a large set of manually coded institution-specific corrections to handle name changes, mergers, campus consolidations, and historical naming inconsistencies across HEGIS, IPEDS, FSS, and College Handbook data sources. Location: Codes/nameclean.ado	-	-	-	-	Created by the authors; no upstream license restrictions apply.

7. Description of Do-files

The table below documents every .do file in the package, including the intermediate construction scripts. Files must be run in the order listed.

Do-file	Description	Key Inputs	Key Outputs	Notes
Construct_Admissions/Construct_AdmissionData.do	Constructs CH_CPALL.dta by (1) importing each of the 24 year-specific sheets (Sheet3–Sheet26) from the raw Excel file All Admissions Data.xlsx; (2) destringging all numeric variables	Data/Raw/All Admissions Data.xlsx (Sheets 3–26)	Intermediate outputs: Data/Processed/Admissions/CH_CP3.dta through CH_CP26.dta Final output:	Must be run before 03_CombineACTSAT.do, which uses CH_CPALL.dta as its main input.

Do-file	Description	Key Inputs	Key Outputs	Notes
	and dropping empty columns; and (3) appending all sheets into a single institution-year panel.		Data/Processed/CH_CPALL.dta	
Construct_HEGIS/Construct_HEGIS_Degrees.do (cannot be run — see Section 4)	Constructs degrees.dta by (1) reading in raw HEGIS degree microdata files for each available survey year (1965–1984, excluding 1969, 1974, and 1982); (2) merging each year with field label lookup files (lookupfields_65_67.dta, lookupfields_68.dta, lookupfields_70_81.dta) that translate numeric subject codes to field names, and with state crosswalk files; (3) constructing a crosswalk from HEGIS taxonomy codes to CIP codes for 1983 and 1984; and (4) appending all years into a single panel.	Data/Raw/HEGIS_Degrees/ (annual HEGIS degree microdata files, ICPSR 2081–2158) Data/Raw/HEGIS_Degrees/lookupfields_65_67.dta Data/Raw/HEGIS_Degrees/lookupfields_68.dta Data/Raw/HEGIS_Degrees/lookupfields_70_81.dta Data/Raw/HEGIS_Degrees/state s.dta Data/Raw/HEGIS_Degrees/state s2.dta Data/Raw/HEGIS_Degrees/CIP_HEGIS_Crosswalk_1981/03135-0001-Data.txt	Intermediate outputs: Data/Processed/HEGIS_Degrees/degrees_{year}.dta (one file per year) Data/Processed/HEGIS_Degrees/crosswalk.dta, cipdesc.dta Final output: Data/Processed/degrees.dta	This script cannot be run as the underlying HEGIS degree microdata files are restricted-use and not included in this replication package due to ICPSR terms of use. The processed output degrees.dta is provided directly in Data/Processed/.
Construct_HEGIS/Construct_HEGIS_Finance.do (cannot be run — see Section 4)	Constructs HEGIS_finance_limit.dta by reading in raw annual HEGIS Finance microdata files (1968–1985), harmonizing revenue and expenditure variables across years, and collapsing to an institution-year panel of total revenues.	Data/Raw/HEGIS_Finance/ (annual HEGIS finance microdata files, ICPSR 2094–2110)	Final output: Data/Processed/HEGIS_finance_limit.dta	This script cannot be run as the underlying HEGIS finance microdata files are restricted-use and not included in this replication package due to ICPSR terms of use. The processed output HEGIS_finance_limit.dta is provided directly in Data/Processed/.
Construct_HEGIS/Construct_HEGIS_Instchar.do (cannot be run — see Section 4)	Constructs HEGIS_InstChar.dta by reading in raw HEGIS Institutional Characteristics microdata files for selected years (1970, 1972–1974, 1977–1978, 1983), harmonizing tuition and fee	Data/Raw/HEGIS_InstitutionalChars/ (ICPSR 2111, 2042–2044, 7647, 6903, 8291) Data/Raw/CPI_long_2019.dta	Intermediate outputs: Data/Processed/HEGIS_InstitutionalChars/{year}_InstChar.dta (one file per year)	This script cannot be run as the underlying HEGIS institutional characteristics microdata files are restricted-use and not

Do-file	Description	Key Inputs	Key Outputs	Notes
	variable names to match IPEDS conventions, appending into a single institution-year panel, and deflating tuition, room, and board to real 2019 dollars using CPI.		Final output: Data/Processed/HEGIS_InstChar.dta	included in this replication package due to ICPSR terms of use. The processed output HEGIS_InstChar.dta is provided directly in Data/Processed.
Construct_IPEDS/Construct_AllIPEDSData.do	Constructs AllIPEDSData.dta and AllIPEDSData22.dta by (1) building an institutional characteristics panel from raw annual IPEDS Institutional Characters files (1987–2023), merging multiple part files per year and harmonizing variables across the changing file naming conventions; (2) building a finance panel from raw annual IPEDS Finance files (1987–2022), similarly merging multiple part files per year; and (3) merging the two panels to produce the final outputs.	Data/Raw/IPEDS_InstitutionalChars/ (annual dct_ic* files) Data/Raw/IPEDS_Finance/ (annual dct_f* files) Data/Raw/CPI_long_2022.dta	Intermediate outputs: Data/Processed/IPEDS_InstitutionalChars/instchar{year}.dta Data/Processed/IPEDS_InstitutionalChars/instchar.dta Data/Processed/IPEDS_Finance/fin{year}.dta Data/Processed/IPEDS_Finance/finance.dta Final outputs: Data/Processed/AllIPEDSData.dta Data/Processed/AllIPEDSData22.dta	Must be run before 01_ProgramsDataFile.do. If you do not wish to re-run this script, the processed outputs AllIPEDSData.dta and AllIPEDSData22.dta are provided directly in Data/Processed/ and can be used as-is.
01_ProgramsDataFile.do	Constructs the master institution-year program dataset cs_progs.dta by (1) cleaning raw HEGIS degree records from degrees.dta, mapping subject codes to broad major groups, generating degree counts by major group, degree level, and sex, and collapsing to an institution-year panel; (2) merging the HEGIS panel with finance (HEGIS_finance_limit.dta) and institutional characteristics (HEGIS_InstChar.dta) data; (3) building a consistent IPEDS institution-year completions panel (1987–2020) by mapping CIP codes to broad major groups	Data/Processed/degrees.dta Data/Processed/HEGIS_finance_limit.dta Data/Processed/HEGIS_InstChar.dta Data/Processed/AllIPEDSData.dta Data/Processed/AllIPEDSData22.dta Data/Raw/All CIP.xls	Intermediate outputs: Data/Processed/cs_HEGIS.dta Data/Processed/IPEDS_char.dta Data/Temp/8590crosswalk.dta Data/Processed/cs_{year}.dta (one file per IPEDS year, 1987–2020, erased after appending) Data/Processed/cs_IPEDS.dta Data/Processed/majors_countbyyear.dta	Depends on Construct_HEGIS outputs (degrees.dta, HEGIS_finance_limit.dta, HEGIS_InstChar.dta) being available in Data/Processed/, and on Construct_AllIPEDSData.do having been run first.

Do-file	Description	Key Inputs	Key Outputs	Notes
	across the 1985/1990/2000 classification changes using All CIP.xls and 2000aligncip.dta, collapsing to institution-year counts by major group, degree level, and sex; and (4) appending the HEGIS and IPEDS panels, merging with finance and admissions characteristics from AllIPEDSData.dta and AllIPEDSData22.dta, cleaning institution identifiers and FICE/unitid links, adding state FIPS codes, constructing aggregate degree totals, gender ratios, and CS-offering indicators.	Data/Raw/2000aligncip.dta Data/Raw/IPEDS_Degrees/dct_c 1987.dta through dct_c2020.dta	Final output: Data/Final/cs_progs.dta	
02_FSSSupportSurvey.do	Constructs institution-year panels of federal research funding from the NSF Federal Support to Universities and Colleges (FSS) survey by (1) processing post-1971 FSS microdata (data_obligation.dta) into an institution-year panel of total federal funding disaggregated by agency (DoD, NASA, HHS), harmonizing units to thousands of dollars, and converting from federal fiscal years to academic years; and (2) importing and harmonizing five pre-1971 digitized Excel files (fiscal years 1963–1970), cleaning institution names and state identifiers using nameclean.ado, and appending into a single institution-year panel converted to academic years.	Data/Raw/FSS_Support_Survey/Microdata_stata_export/data_obligation.dta Data/Raw/FSS_Support_Survey/Microdata_stata_export/c_agency.dta Data/Raw/FSS_Support_Survey/Microdata_stata_export/c_inst_codebook.dta Data/Raw/FSS_Support_Survey/FSS_support_preperiod/Federal Support to Universities and Colleges_1963-1966.xlsx Data/Raw/FSS_Support_Survey/FSS_support_preperiod/Federal Support to Universities and Colleges Fiscal Year 1967.xlsx Data/Raw/FSS_Support_Survey/FSS_support_preperiod/Federal Support to Universities and Colleges Fiscal Year 1968.xlsx	Intermediate outputs: Data/Processed/FSS_support_preperiod/FSS1966.dta through FSS1970.dta (5 files) Final outputs: Data/Processed/FSSsupport.dta (institution-year panel, 1971–present) Data/Processed/FSSpre1971.dta (institution-year panel, 1963–1970)	Must be run before 04_ConstructDatasetfor Regressions.do. Self-contained, no dependency on other construction scripts.

Do-file	Description	Key Inputs	Key Outputs	Notes
		Data/Raw/FSS_Support_Survey/ FSS_support_preperiod/1969 Totals_William_2-16-22.xlsx Data/Raw/FSS_Support_Survey/ FSS_support_preperiod/1970 Totals_William_2-16-22.xlsx		
03_CombineACTSAT.do	Cleans and harmonizes SAT and ACT score data from the College Handbook dataset (CH_CPALL.dta) by (1) constructing institution-year average ACT composite and SAT combined scores from available score ranges (low and high); (2) manually correcting bad data points for specific institutions and years; and (3) collapsing to a single observation per institution per year.	Data/Processed/CH_CPALL.dta	Intermediate outputs: None Final output: Data/Final/oldactdata.dta	Depends on Construct_AdmissionData.do having been run first to produce CH_CPALL.dta. Output is used in 04_ConstructDatasetforRegressions.do.
04_ConstructDatasetforRegressions.do	Builds the main analysis and regression-ready datasets by (1) filtering to Carnegie research universities using the 1994 Carnegie Classification, applying sample restrictions (dropping graduate-only, single-field, and anomalous institutions), and saving an institution ID crosswalk (idfile.dta); (2) merging FSS pre- and post-1971 funding data to institution IDs, deflating all funding variables to real 2022 dollars in units of 10 millions, and generating up to 7 leads and lags of total, DoD, NASA, and HHS funding; (3) merging in HEGIS/IPEDS degree and revenue data, constructing deflated total revenue variables and revenue per student; (4) constructing cleaned degree and CIP classification variables by field and degree level, merging Education and Vocational into a	Data/Raw/Carnegie_1994_edition_data.xls Data/Raw/CPI_long_2022.dta Data/Final/cs_progs.dta Data/Final/oldactdata.dta Data/Processed/FSSsupport.dta Data/Processed/FSSpre1971.dta	Intermediate outputs: Data/Processed/idfile.dta Final outputs: Data/Final/FSSsupport.dta Data/Final/FSSpre1971.dta Data/Final/analysis_aera.dta Data/Final/didregression.dta	Depends on 01_ProgramsDataFile.do, 02_FSSSupportSurvey.do, and 03_CombineACTSAT.do having been run first. Produces the two core datasets used in all regressions.

Do-file	Description	Key Inputs	Key Outputs	Notes
	single category, and generating field shares as percentages of total degrees; (5) merging SAT/ACT selectivity data from oldactdata.dta, interpolating missing values, standardizing ACT and SAT scores by year, and constructing a combined admissions control variable that favors ACT or SAT based on institutional reporting patterns; (6) generating 1971 baseline controls for institution size and degrees by field and degree level, and constructing cumulative funding variables; and (7) constructing STEM and non-STEM degree and CIP grouping variables by field and degree level for use in the reshape and regression analysis.			
05a_Regressions.do	Produces all main tables and figures for the paper by running triple difference-in-differences (DiD) and two-way fixed effects (TWFE) regressions to estimate the effect of federal science funding on STEM degree completion and STEM degree programs concentration across Carnegie research universities, reported separately for bachelor's, master's, and doctoral degrees, by (1) running the main triple DiD specification with three-way fixed effects (university × STEM, STEM × year, university × year) across all STEM subfields and generating Figure 1 and the regression tables for Tables A1 and A2; (2) running robustness checks that vary the non-STEM control group one field at a time and generating Figure 2 and Tables A3 and A4; (3) running two-way fixed effects regressions for each broad major group separately, generating	Data/Final/didregression.dta (Sections 1, 2, 5, 6) Data/Final/analysis_aera.dta (Sections 3, 4)	Intermediate outputs: None Final outputs: Tables/tableappendix/ (Tables A1–A5) Tables/table1and2/summary.xlsx (Tables 1–2) Tables/table3/ (Table 3) Tables/twfe/ (two-way FE regression tables) Figures/figure1/ (Figure 1) Figures/figure2/ (Figure 2) Figures/figure3andb1/ (Figure 3 and Figure B1)	Depends on 04_ConstructDatasetforRegressions.do having been run first.

Do-file	Description	Key Inputs	Key Outputs	Notes
	placebo plots for non-STEM fields and STEM subfield plots for Figure 3 and Figure B1; (4) computing summary statistics for degree programs, degrees awarded, and university characteristics at 1975, 1995, and 2015 and exporting Tables 1 and 2; (5) running the main triple DiD specification augmented with 3-year leads and exporting Table 3; and (6) running the same specification with 6-year leads and exporting Table A5.			
05b_Regressions.do	Produces additional tables for the paper by running triple DiD and two-way fixed effects regressions, by (7) running the triple DiD specification within sub-periods (pre-1985, 1985–1989, 1990–1999, 2000–2015) and by above/below median baseline enrollment and federal grants to examine heterogeneity, and exporting Table 4; (8) running the STEM total two-way fixed effects regression under five progressively richer sets of controls (no controls, baseline enrollment, adding gender ratio, adding tuition, adding admissions selectivity) and exporting Table A11; (9) re-running the main triple DiD specification with 3-year leads included alongside the main lag structure and exporting Tables A6 and A7; (10) running the same with 6-year leads and exporting Tables A8 and A9; and (11) exporting the CIP classification crosswalk to Table A10.	Data/Final/didregression.dta (Sections 7, 9, 10) Data/Final/analysis_aera.dta (Section 8) Data/Raw/2000aligncip.dta (Section 11)	Intermediate outputs: None Final outputs: Tables/tableappendix/ (Tables A6–A11) Tables/table4/ (Table 4) Tables/CIPTable.xlsx (Table A10)	Depends on 04_ConstructDatasetforRegressions.do having been run first.
06_CalculateNumbers.do	Computes back-of-envelope calculations for the implied effects of federal research funding on STEM degrees awarded and	Data/Final/didregression.dta Data/Final/analysis_aera.dta	Intermediate outputs: None	Depends on 04_ConstructDatasetforRegressions.do and 05a_Regressions.do

Do-file	Description	Key Inputs	Key Outputs	Notes
	degree programs at the university and national level, using coefficient estimates from the main regressions in 05a_Regressions.do. Regression coefficients are hardcoded as locals; all other summary statistics are computed directly from the data. Calculations cover (1) all STEM fields combined, reported as counts and percentages of average STEM-only and combined degree totals at both the university and national level; (2) Biology degrees and programs separately; (3) Engineering degrees and programs separately; and (4) heterogeneity by above/below median cumulative funding as of 1971, reporting implied effects at both the university and national level for each subgroup and as weighted averages.		Final outputs: None (all results displayed via di commands in the Stata output window and not saved to any files)	having been run first, as the regression coefficients hardcoded in this script come from the main regressions in 05a_Regressions.do.

8. List of Figures and Tables

The table below maps each exhibit in the paper to the do-file that produces it and the output location.

Exhibit	Description	Program	Output File(s)
Table 1	Summary statistics for degree programs, degrees awarded, and university characteristics at 1975, 1995, and 2015	05a_Regressions	Tables/table1and2/summary.xlsx
Table 2	Summary statistics (continued)	05a_Regressions	Tables/table1and2/summary.xlsx
Table 3	Triple DiD estimates with 3-year leads for degrees awarded and degree programs	05a_Regressions	Tables/table3/
Table 4	Heterogeneity analysis by sub-period and above/below median baseline enrollment and federal grants	05b_Regressions	Tables/table4/

Exhibit	Description	Program	Output File(s)
Table A1	Main triple DiD: STEM–funding effects on degrees awarded, by subfield and degree level	05a_Regressions	Tables/tableappendix/
Table A2	Main triple DiD: STEM–funding effects on degree programs, by subfield and degree level	05a_Regressions	Tables/tableappendix/
Table A3	Robustness: varying non-STEM control group (degrees awarded)	05a_Regressions	Tables/tableappendix/
Table A4	Robustness: varying non-STEM control group (degree programs)	05a_Regressions	Tables/tableappendix/
Table A5	Triple DiD with 6-year leads for degrees awarded and degree programs	05a_Regressions	Tables/tableappendix/
Table A6	Triple DiD on degrees awarded: 3-year leads + main lag structure	05b_Regressions	Tables/tableappendix/
Table A7	Triple DiD on degree programs: 3-year leads + main lag structure	05b_Regressions	Tables/tableappendix/
Table A8	Triple DiD on degrees awarded: 6-year leads + main lag structure	05b_Regressions	Tables/tableappendix/
Table A9	Triple DiD on degree programs: 6-year leads + main lag structure	05b_Regressions	Tables/tableappendix/
Table A10	2-digit CIP classification used in the analysis	05b_Regressions	Tables/CIPTable.xlsx
Table A11	Sensitivity of two-way FE estimates to progressively richer controls	05b_Regressions	Tables/tableappendix/
Figure 1	Triple DiD estimates of federal funding effects on STEM degrees awarded and degree programs, by subfield and degree level	05a_Regressions	Figures/figure1/
Figure 2	Robustness of triple DiD estimates across alternative non-STEM control groups	05a_Regressions	Figures/figure2/
Figure 3	Two-way FE estimates of federal funding effects by degree level	05a_Regressions	Figures/figure3andb1/
Figure B1	Placebo: two-way FE estimates for non-STEM fields	05a_Regressions	Figures/figure3andb1/