

Data Sources

ERIN KEITH



Goals

1. Return to Design
2. Consider “real-world” Data

Process Overview

- Define software system goal
 - Find examples
- Group data by dynamics
 - Static / infrequent changes
 - inserted into table at creation, rarely updated
 - Frequent changes
 - inserted, modified, or deleted as the system is running
- Identify data sources
 - Collect, examine, and clean initial data
 - Download, scrape, etc.
- Design database schema
- Create database and tables
 - Import static data

College Rankings

- Define software system goal
- Find examples
- <https://www.usnews.com/education/best-global-universities/neuroscience-behavior>
- <https://www.bestcolleges.com/psychology/graduate-schools-for-clinical-psychology/>

College Rankings

- Group data by dynamics
- Static / infrequent changes
 - <https://collegescorecard.ed.gov/>
 - <https://nces.ed.gov/ipeds>

IES

NCES

National Center for Education Statistics

MENU

Search

IPEDS

Integrated Postsecondary Education Data System

Data Tools

Help Desk 1 866-558-0658

Start over

Save session

Help

MAIN MENU

Complete Data Files

Data Release Info

Years & Surveys

2022

All Surveys

Continue

Data files are available in ZIP format.

Year	Survey	Title	Data File	Stata Data File	Programs	Dictionary
2022	Institutional Characteristics	Directory information	HD2022	HD2022_STATA	SPSS , SAS , STATA	Dictionary
2022	Institutional Characteristics	Educational offerings, organization, services and athletic associations	IC2022	IC2022_STATA	SPSS , SAS , STATA	Dictionary
2022	Institutional Characteristics	Student charges for academic year programs	IC2022_AY	IC2022_AY_STATA	SPSS , SAS , STATA	Dictionary
2022	Institutional Characteristics	Student charges by program (vocational programs)	IC2022_PY	IC2022_PY_STATA	SPSS , SAS , STATA	Dictionary
2022	Institutional Characteristics	Response status for all survey components	FLAGS2022	FLAGS2022_STATA	SPSS , SAS , STATA	Dictionary
2022	12-Month Enrollment	12-month unduplicated headcount by race/ethnicity, gender and level of student: 2021-22	EFFY2022	EFFY2022_STATA	SPSS , SAS , STATA	Dictionary
2022	12-Month Enrollment	12-month unduplicated headcount by distance education status and student level: 2021-22	EFFY2022_DIST	EFFY2022_DIST_STATA	SPSS , SAS , STATA	Dictionary
2022	12-Month Enrollment	12-month instructional activity: 2021-22	EFIA2022	EFIA2022_STATA	SPSS , SAS , STATA	Dictionary
2022	12-Month Enrollment	Response status all survey components	FLAGS2022	FLAGS2022_STATA	SPSS , SAS , STATA	Dictionary

- Identify data sources
- **Collect**, examine, and clean initial data

hd2022

College Rankings

- Identify data sources
- Collect, **examine**, and clean initial data

varnumber	varname	DataType	Fieldwidth	format	imputationvar	varTitle
1	UNITID	N	6	Cont		Unique identification number of the institution
10001	INSTNM	A	120	Alpha		Institution (entity) name
16081	IALIAS	A	2000	Alpha		Institution name alias
10006	ADDR	A	100	Alpha		Street address or post office box
10011	CITY	A	30	Alpha		City location of institution
10016	STABBR	A	2	Disc		State abbreviation
10021	ZIP	A	10	Alpha		ZIP code
10026	FIPS	N	2	Disc		FIPS state code
10031	OBEREG	N	2	Disc		Bureau of Economic Analysis (BEA) regions
10036	CHFNM	A	50	Alpha		Name of chief administrator
10041	CHFTITLE	A	50	Alpha		Title of chief administrator
10046	GENTELE	A	15	Alpha		General information telephone number
10061	EIN	A	9	Cont		Employer Identification Number
10069	UEIS	A	2000	Cont		Unique Entity Identifier (UEI) Numbers
10071	OPEID	A	8	Cont		Office of Postsecondary Education (OPE) ID Number
10076	OPEFLAG	N	1	Disc		OPE Title IV eligibility indicator code
10081	WEBADDR	A	150	Alpha		Institution's internet website address
10058	ADMINURL	A	200	Alpha		Admissions office web address
10053	FAIDURL	A	200	Alpha		Financial aid office web address
10059	APPLURL	A	200	Alpha		Online application web address
10083	NPRICURL	A	200	Alpha		Net price calculator web address
16131	VETURL	A	200	Alpha		Veterans and Military Servicemembers tuition policies web address
80081	ATHURL	A	150	Alpha		Student-Right-to-Know student athlete graduation rate web address
10060	DISAURL	A	200	Alpha		Disability Services Web Address
10086	SECTOR	N	2	Disc		Sector of institution
10091	ICLEVEL	N	2	Disc		Level of institution
10096	CONTROL	N	2	Disc		Control of institution
10101	HLOFFER	N	2	Disc		Highest level of offering
10106	UGOFFER	N	2	Disc		Undergraduate offering
10111	GROFFER	N	2	Disc		Graduate offering
10122	HDEGOFR1	N	2	Disc		Highest degree offered
10126	DEGGRANT	N	2	Disc		Degree-granting status
10151	HBCU	N	2	Disc		Historically Black College or University
10156	HOSPITAL	N	2	Disc		Institution has hospital
10161	MEDICAL	N	2	Disc		Institution grants a medical degree
10166	TRIBAL	N	2	Disc		Tribal college

College Rankings

- Identify data sources
- Collect, examine, and **clean** initial data

UNITID	INSTNM	ALIAS	ADDR	CITY	STABBR	ZIP	FIPS	OBREG	CHFNAM	CHFTITLE	GENTELE	EIN	UEIS	OPEID	OPEFLAG	WEBADDR	ADMINUR	FAIDURL	APPLURL	NPRICURL	VETURL	ATHURL	DISAURL	SE
100654	Alabama A&M		4900 Meri Normal	AL		35762	1	5	Dr. Daniel	President	2.56E+09	6.36E+08	JDVGS67M	100200	1	www.aam	https://w	https://w	https://w	https://w	www.aam		https://w	
100663	University UAB		Administr Birmingha	AL		35294-011	1	5	Ray L. Wal	President	2.06E+09	6.36E+08	YND4PLM	105200	1	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
100690	Amridge U Southern		1200 Taylc Montgom	AL		36117-355	1	5	Michael C	President	3.34E+13	2.37E+08	RB27R4GL	2503400	1	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
100706	University UAH Univ		301 Sparki Huntsville	AL		35899	1	5	Chuck Kar	President	2.57E+09	6.31E+08	HB6KNGV	105500	1	www.uah	https://w	finaid.uah	register.u	finaid.uah	www.uah	www.uah	www.uah	www.uah
100724	Alabama S		915 S Jacki Montgom	AL		36104-027	1	5	Quinton T	President	3.34E+09	6.36E+08	DLJWLSM	100500	1	www.alas	www.alas	www.alas	www.alas	www.alas	www.alas	https://w	www.alas	
100733	University		500 Unive Tuscaloos	AL		35401	1	5	Finis T. Jc	Chancellor	2.05E+09	6.36E+08		800400	2	www.uas						www.uas		
100751	The Unive		739 Unive Tuscaloos	AL		35487-010	1	5	Dr. Stuart	President	2.05E+09	6.36E+08	RCNJEH28	105100	1	www.ua.e	gobama.u	financiala	apply.ua.e	financiala	https://ve	registrar.u	ods.ua.ed	
100760	Central Al		1675 Cher Alexande	AL		35010	1	5	Jeff Lynn	President	2.56E+09	6.31E+08	G1C3X4AM	100700	1	www.cacc	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
100812	Athens St.		300 N Bea Athens	AL		35611	1	5	Dr. Cather	Interim Pr	2.56E+09	6.31E+08	GFZJMKW	100800	1	www.athe	www.athe	www.athe	www.athe				www.athe	
100830	Auburn Un AUM Au		7440 East Montgom	AL		36117-359	1	5	Carl A. Stc	Chancellor	3.34E+09	6.36E+08	S2Y7SMV2	831000	1	www.aum	www.aum	www.aum	www.aum	www.aum	www.aum	www.aum	www.aum	www.aum
100858	Auburn Un		Auburn	AL		36849	1	5	Chris Rob	President	3.35E+09	6.36E+08	DMQNDJC	100900	1	www.aub	www.aub	www.aub	www.aub	https://w	veterans.e	https://au	cws.aubur	
100937	Birmingha BSC		900 Arkad Birmingha	AL		35254	1	5	Mr. Danie	President	2.05E+09	6.31E+08	L9NKL7V8	101200	1	www.bsc	www.bsc	https://w	https://w	www.bsc	https://w	https://w	https://w	https://w
101028	Chattahoc CVCC		2602 Colle Phenix Cit	AL		36869	1	5	Jacquelin	President	3.34E+09	6.31E+08	MEKHNSC	1218200	1	https://w	https://w	https://w	https://w	https://ch	https://ex	https://w	https://w	https://w
101116	South Uni		5355 Vaug Montgom	AL		36116	1	5	Kandis R.	'Dean of A	3.34E+09	6.3E+08	YJAH2LJ7	1303906	1	www.sou	www.sou	www.sou	www.sou	www.sou	www.sou	www.sou	www.sou	www.sou
101143	Enterprise Enterprise		600 Plaza Enterprise	AL		36330-130	1	5	Danny Lor	President	3.34E+09	6.31E+08	CDVSG3N	101500	1	www.escc	https://w	www.escc	https://w	https://w	https://w	www.escc	www.escc	www.escc
101161	Coastal Al		1900 U S H Bay Minet	AL		36507-269	1	5	Craig Pour	President	2.52E+09	6.31E+08	JK46KBHM	106000	1	www.coas	www.coas	www.coas	www.coas	www.coas	www.coas	www.coas	www.coas	www.coas
101189	Faulkner Faulkner		5345 Atlar Montgom	AL		36109-339	1	5	Mitch Her	President	3.34E+09	6.3E+08	X4N7WED	100300	1	www.faul	www.faul	www.faul	applynow	www.faul	www.faul	https://w	https://w	https://w
101240	Gadsden GSCC Ga		1001 Geor Gadsden	AL		35903	1	5	Dr. Kathy	President	2.57E+09	6.31E+08	F52PB51Z	101700	1	www.gad	www.gad	www.gad	https://w	www.gad	www.gad	https://w	www.gad	https://w
101277	New Begi		421 Martli Albertvill	AL		35951	1	5	Amanda B	DIRECTOR	2.57E+09	2.03E+08		4187200	1	www.nbc				nbcosme			www.nbc	
101286	George C WALLACE		1141 Walli Dothan	AL		36303-923	1	5	Linda C Yo	President	3.35E+09	6.31E+08	RSMGDHC	101800	1	www.wall	www.wall	www.wall	www.wall	www.wall	www.wall	www.wall	www.wall	www.wall
101295	George C Wallace S		801 Main Hanceville	AL		35077-200	1	5	Vicki P. Ka	President	2.56E+09	6.3E+08	WM9JND4	787100	1	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
101301	George C Wallace C		3000 Earl Selma	AL		36702-280	1	5	Dr. James	President	3.35E+09	6.31E+08	JEYJLCR37	569900	1	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
101365	Herzing Un		280 West Birmingha	AL		35209	1	5	Sarah Hig	President	2.06E+09	2.72E+08	JDF5MPY3	962107	1	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w	https://w
101435	Huntingdo		1500 East Montgom	AL		36106-214	1	5	J. Camero	President	3.35E+09	6.3E+08	MACKQAF	101900	1	www.hun	www.hun	www.hun	www.hun	www.hun	www.hun	www.hun	www.hun	www.hun
101453	Heritage C		3625 Helct Florence	AL		35630-997	1	5	W. Kirk Br	President	2.57E+09	2.37E+08	MYJFHUU	2199700	1	www.hcu	www.hcu	www.hcu	www.hcu	www.hcu	www.hcu	www.hcu	www.hcu	www.hcu
101462	J. F. Drake Drake Stat		3421 Meri Huntsville	AL		35811	1	5	Dr. Patrici	President	2.57E+09	6.3E+08	Y6RMRVL	526000	1	https://w	https://w	https://w	https://w	https://dr	https://w	https://w	https://w	https://w
101471	J F Ingram ISTC		5375 Ingra Deatsville	AL		36022	1	5	Annette F	President	3.34E+13	6.31E+08		3002500	1	www.istc				istc.edu/n			www.istc	
101480	Jacksonvil		700 Pelhai Jacksonvil	AL		36265	1	5	Don Killin	President	2.57E+09	6.36E+08	D7N5GMC	102000	1	www.jsu	www.jsu	www.jsu	www.jsu	www.jsu	www.jsu	www.jsu	www.jsu	www.jsu

College Rankings

- Group data by dynamics
- Frequent changes

#1 The Best Graduate Schools for Clinical Psychology Programs

[Brenau University](#)

📍 Gainesville, GA • 4 years • Online + Campus

Programmatic Accreditation: [Masters in Psychology and Counseling Accreditation Council...](#)

▼ More Details

Average Undergrad Tuition	
In-state	\$30,000
Out-of-state	\$30,000

Average Graduate Tuition	
In-state	\$13,356
Out-of-state	\$13,356

★Sponsored 📍 California, USA • 4 year • Online + Campus

[Psychology@Pepperdine](#)

Pepperdine University's online psychology master's program prepares students for meaningful careers in psychology, human services, and related fields through a values-based curriculum. Prepare for counseling licensure in as few as 27 months wit...

▼ More Details

Explore Program ↗

#2 The Best Graduate Schools for Clinical Psychology Programs

[Indiana University-Purdue University-Indianapolis](#)

📍 Indianapolis, IN • 4 years • Online + Campus

Programmatic Accreditation: [American Psychological Association](#)

...

Average Undergrad Tuition	
In-state	\$8,795
Out-of-state	\$30,477

Average Graduate Tuition	
In-state	\$9,090
Out-of-state	\$24,864

College Rankings

- Group data by dynamics
- Static / infrequent changes
 - <https://collegescorecard.ed.gov/>
 - <https://nces.ed.gov/ipeds>

2022	12-Month Enrollment	12-month unduplicated headcount by race/ethnicity, gender and level of student: 2021-22	EFFY2022	EFFY2022_STATA	SPSS, SAS, STATA	Dictionary
2022	12-Month Enrollment	12-month unduplicated headcount by distance education status and student level: 2021-22	EFFY2022_DIST	EFFY2022_DIST_STATA	SPSS, SAS, STATA	Dictionary
2022	12-Month Enrollment	12-month instructional activity: 2021-22	EFIA2022	EFIA2022_STATA	SPSS, SAS, STATA	Dictionary
2022	12-Month Enrollment	Response status all survey components	FLAGS2022	FLAGS2022_STATA	SPSS, SAS, STATA	Dictionary
2022	Completions	Awards/degrees conferred by program (6-digit CIP code), award level, race/ethnicity, and gender: July 1, 2021 to June 30, 2022	C2022_A	C2022_A_STATA	SPSS, SAS, STATA	Dictionary
2022	Completions	Number of students receiving awards/degrees, by race/ethnicity and gender: July 1, 2021 to June 30, 2022	C2022_B	C2022_B_STATA	SPSS, SAS, STATA	Dictionary
2022	Completions	Number of students receiving awards/degrees, by award level and by gender, race/ethnicity and age categories: July 1, 2021 to June 30, 2022	C2022_C	C2022_C_STATA	SPSS, SAS, STATA	Dictionary
2022	Completions	Number of programs offered and number of programs offered via distance education, by award level: July 1, 2021 to June 30, 2022	C2022DEP	C2022DEP_STATA	SPSS, SAS, STATA	Dictionary
2022	Completions	Response status for all survey components	FLAGS2022	FLAGS2022_STATA	SPSS, SAS, STATA	Dictionary

College Rankings

- Identify data sources
 - Collect, examine, and clean initial data

CIPFamily	CIPCode	Action	TextChange	CIPTitle	CIPDefinition	CrossRef	Examples										
01	01	No substa	yes	AGRICULT	Instructional programs that focus on agriculture, animal, plant, veterinary, and related sciences and that prepare students for employment in these fields.												
01	01.00	No substa	no	Agricultur	Instructional content is defined in code 01.0000.												
01	01.0000	No substa	no	Agricultur	A program	14.0301 - Agricultural Engineering.											
01	01.01	No substa	no	Agricultur	Instructional content for this group of programs is defined in codes 01.0101 - 01.0199.												
01	01.0101	No substa	no	Agricultur	A general program that focuses on modern business and economic principles involved in the organization, operation, and management of agricultural businesses.												
01	01.0102	No substa	no	Agribusin	A program that prepares individuals to manage agricultural businesses and agriculturally related operations.												
01	01.0103	No substa	no	Agricultur	A program	03.0204 - Examples: - Agroecomics											
01	01.0104	No substa	no	Farm/Farr	A program that prepares individuals to manage farms, ranches, and similar enterprises. Includes instruction in farm management, farm machinery, and farm safety.												
01	01.0105	No substa	no	Agricultur	A program	52.1803 - Retailing and Retail Operations., 52.1901 - Auctioneering., 52.0202 - Purchasing, Procurement, and Distribution.											
01	01.0106	No substa	yes	Agricultur	A program that prepares individuals to perform specialized support functions related to agricultural businesses.												
01	01.0199	No substa	no	Agricultur	Any instructional program in agricultural business and management not listed above.												
01	01.02	No substa	no	Agricultur	Instructional content for this group of programs is defined in codes 01.0201 - 01.0299.												
01	01.0201	No substa	no	Agricultur	A program that generally prepares individuals to sell, select, and service agricultural or agribusiness technical equipment.												
01	01.0204	No substa	no	Agricultur	A program	49.0202 - Examples: - Farm Equipment Operator											
01	01.0205	No substa	yes	Agricultur	A program that prep	Examples: - Agricultural Equipment Mechanic, - Agricultural Equipment Technician, - Agricultural Equipment Repairer.											
01	01.0207	New	no	Irrigation	A program that prepares individuals for employment in the field of agricultural technology and irrigation. Includes instruction in irrigation system design, installation, and maintenance.												
01	01.0299	No substa	no	Agricultur	Any instructional program in agricultural mechanization not listed above.												
01	01.03	No substa	no	Agricultur	Instructional content for this group of programs is defined in codes 01.0301 - 01.0399.												
01	01.0301	No substa	no	Agricultur	A program that focuses on the general planning, economics, and use of facilities, natural resources, equipment, and personnel in the management of agricultural operations.												
01	01.0302	No substa	no	Animal/Li	A program that prepares individuals to select, breed, care for, process, and market livestock and small farm animals.												

College Rankings

- Identify data sources

- Collect, examine, and clean initial data

UNITID	CIPCODE	MAJORNU	AWLEVEL	XCTOTAL	CTOTAL	XCTOTALM	CTOTALM	XCTOT
100654	1.0999	1	5 R		9 R		3 R	
100654	1.1001	1	5 R		7 R		3 R	
100654	1.1001	1	7 R		7 R		1 R	
100654	1.1001	1	17 R		3 R		1 R	
100654	1.9999	1	5 R		1 R		1 Z	
100654	1.9999	1	7 R		8 R		2 R	
100654	1.9999	1	17 R		3 R		2 R	
100654	3.0599	1	5 R		8 R		6 R	
100654	4.0301	1	5 R		3 R		2 R	
100654	4.0301	1	7 R		12 R		4 R	
100654	9.0199	1	7 R		11 R		3 R	
100654	10.0202	1	5 R		17 R		10 R	
100654	11.0101	1	5 R		41 R		23 R	
100654	11.0101	1	7 R		16 R		9 R	
100654	13.0101	1	8 R		0 R		0 Z	
100654	13.0401	1	7 R		1 R		1 Z	
100654	13.1001	1	5 R		3 R		2 R	
100654	13.1001	1	7 R		1 Z		0 R	
100654	13.1012	1	5 R		8 Z		0 R	
100654	13.1202	1	5 R		6 Z		0 R	
100654	13.1202	1	7 R		1 R		1 R	
100654	13.1205	1	7 R		4 R		1 R	
100654	13.121	1	5 R		1 Z		0 R	

College Rankings

- Design database schema
 - Entity Relationship Diagram (ERD)
- Create database and tables
 - Import static data

Importing to PostgreSQL

<https://www.postgresqltutorial.com/postgresql-tutorial/import-csv-file-into-postgresql-table/>

Split into groups (3 – 5)

Consider a system for reviewing dorms

- Define software system goal
 - Find 2+ examples
- Group data by dynamics
 - Static / infrequent changes
 - Frequent changes
- Identify data sources
 - Find 2+ sources
 - Examine the data
 - Do you need to change anything?

Next Class

Module:

Week 9: Data Sources

Topic:

APIs and ETLs

