(//www.census.gov)

Post-Secondary Employment Outcomes (PSEO)

Post-Secondary Employment Outcomes (PSEO) are experimental tabulations developed by researchers at the U.S. Census Bureau. PSEO data provide earnings and employment outcomes for college and university graduates by degree level, degree major, and post-secondary institution. These statistics are generated by matching university transcript data with a national database of jobs, using state-of-the-art confidentiality protection mechanisms to protect the underlying data.

The PSEO are made possible through data sharing partnerships between universities, university systems, State Departments of Education, State Labor Market Information offices, and the U.S. Census Bureau. PSEO data are currently only available for post-graduate institutions whose transcript data has been made available to Census Bureau through a data-sharing agreement.

PSE

Download Public-Use Data

Lear
belo

We release two classes of files for each of the tabulations, Graduate Earnings and Employment Flows:

- PSEO
  (//ww
- PSEO
  (/data
  draft/
- PSEO
  (/doc
  (//ww
  Techr
  (/doc
  (//ww

Comprehensive dataset, which includes all institutions and crossings
State datasets, which includes all institutions in a state and is a subset of the above release

Data files are provided in zipped CSV and XLS formats and can be downloaded below. The XLS files have variable labels attached, but do not include all the possible rows from Employment Flows, due to size constraints.

	Graduate Earnings	Employment Flows	Both	Metadata
Graduate Earnings, All Institutions	CSV (/data/pseo/R2019Q2/all/pseoe_all.csv.gz)	CSV (/data/pseo/R2019Q2/all/pseof_all.csv.gz)		TXT (/data/pseo/R2019Q2/all/version_pseo.txt)
Texas	CSV (/data/pseo/R2019Q2/tx/pseoe_tx.csv.gz)	CSV (/data/pseo/R2019Q2/tx/pseof_tx.csv.gz)	XLS (/data/pseo/R2019Q2/tx/pseo_tx.xlsx)	TXT (/data/pseo/R2019Q2/tx/version_pseo.txt)
Colorado	CSV (/data/pseo/R2019Q2/co/pseoe_co.csv.gz)	CSV (/data/pseo/R2019Q2/co/pseof_co.csv.gz)	XLS (/data/pseo/R2019Q2/co/pseo_co.xlsx)	TXT (/data/pseo/R2019Q2/co/version_pseo.txt)
Michigan	CSV (/data/pseo/R2019Q2/mi/pseoe_mi.csv.gz)	CSV (/data/pseo/R2019Q2/mi/pseof_mi.csv.gz)	XLS (/data/pseo/R2019Q2/mi/pseo_mi.xlsx)	TXT (/data/pseo/R2019Q2/mi/version_pseo.txt)
Wisconsin	CSV (/data/pseo/R2019Q2/wi/pseoe_wi.csv.gz)	CSV (/data/pseo/R2019Q2/wi/pseof_wi.csv.gz)	XLS (/data/pseo/R2019Q2/wi/pseo_wi.xlsx)	TXT (/data/pseo/R2019Q2/wi/version_pseo.txt)

Note that in the most recent release, only data from the University of Texas system, public institutions in Colorado, the University of Michigan - Ann Arbor, and the University of Wisconsin - Madison are available. PSEO data will be updated as new cells are able to be published.

PSEO data can also be accessed via the PSEO Explorer (/data/pseo_explorer.html) visualization tool. This light-weight interactive tool allows for comparisons of employment outcomes through dynamic grouped bar charts and employment flows through Sankey diagrams. To browse the PSEO data files in their directory structure or to access them with a FTP program (must be able to access HTTP), go to: lehd.ces.census.gov/data/pseo/ (/data/pseo/). Column definitions and other important information can be found in the schema files located in: lehd.ces.census.gov/data/schema/V4.5.0-draft/lehd_public_use_schema.html (/data/schema/V4.5.0-draft/lehd_public_use_schema.html).

Methodology

Introduction

Post-Secondary Employment Outcomes (PSEO) are experimental tabulations developed by the Longitudinal Employer-Household Dynamics (LEHD) program in collaboration with post-secondary institutions and state agencies. PSEO data provide earnings and employment outcomes for college and university graduates by degree level, degree major, and post-secondary institution. The current PSEO is released as a research data product in "experimental" form.

The PSEO provide data on earnings and employment for recent graduates of partner colleges and universities. Graduate Earnings are available at the 25th, 50th, and 75th percentiles, one, five, and 10 years after graduation, by institution, degree level, degree field, and graduation cohort. Employment Flows provide industry and location of employment for graduates. These statistics are generated by matching university transcript data with a national database of jobs, using state-of-the-art confidentiality protection mechanisms to protect the underlying data.

top

Uses

Economic considerations drive a number of college decisions - whether to attend college, where to attend, and what major to select. Given the resources required to attend college, students want to know whether programs are likely to have a sufficient return to justify their expense. Prospective students would also like to know what labor markets recent graduates are working in and whether or not they are employed in an industry appropriate for their training.

Existing data provide some information on earnings for graduates but have certain limitations. College Scorecard data from the U.S. Department of Education are restricted to federal aid recipients, who may not be representative of the student population. College Scorecard data also include everyone who enrolled in the institution and does not separate out those that received a degree from those who did not. PayScale, a commercial website, publishes earnings by institution and degree, but relies on voluntary self-reported earnings, not generally considered a scientifically valid sampling method. Many states, such as Texas, have matched transcript data to state job records to produce statistics on earnings and employment for graduates. However, state administrative data systems cannot follow students out of state, biasing earnings and employment downward in the matched data.

PSEO statistics have many of the advantages of state-based matching systems (universal coverage of post-secondary graduate population, longitudinal information on earnings and employment after graduation) but with the advantage that LEHD data allow us to measure earnings and employment irrespective of a student's location within the U.S. The PSEO also use cutting-edge differential privacy methods to protect the confidentiality of the underlying data, a protection method developed in computer science to bound the privacy risk to individuals from multiple queries to the same database. Differential privacy methods allow the Census Bureau to release detailed tabulations on student outcomes while minimizing the privacy risk to individuals in the data.

top

Data Sources

The sample frame for the PSEO is persons who received a degree or certificate from an in-scope institution. Institutions securely transmit a graduation file to the Census Bureau, which reports the degree type, degree field, graduation date, and institution for any graduating student. Demographic data on students are also provided.

Transcript data is provided to Census Bureau by higher education systems and individual colleges and universities through data sharing agreements with Census Bureau. In the initial pilot phase of the PSEO, only a handful of institutions are represented, but institutional coverage will expand as the program expands.

top

Post-Graduate Population Coverage

PSEO tabulations include only graduates of in-scope institutions. Students who enroll but do not graduate are omitted from the statistics. Of these graduates, a very small fraction (less than one percent of graduates) are omitted from the published statistics due to poor quality of the personal identifier. A much larger fraction of graduates is omitted from the earnings and employment outcome statistics because of insufficient labor market attachment in the reference year. For example, a graduate with zero earnings for three quarters of the calendar year but positive earnings in a single quarter will not be included in the earnings statistics or employment counts. These graduates are omitted as the PSEO is intended to reflect earnings and employment for graduates who work throughout the year. More specifics on the labor force attachment restrictions are provided in the earnings section.

top

Employment Coverage

The LEHD data at the U.S. Census Bureau is a quarterly database of jobs covering over 96% of employment in the United States. The core jobs data are state unemployment insurance (UI) wage records collected via a voluntary federal-state data sharing partnership. These job records are then supplemented with U.S. Census Bureau surveys and other federal agency administrative records to supply additional information on the characteristics of the workers and firms. This linked employer-employee data for the U.S. is the source data for Census Bureau's Quarterly Workforce Indicators (QWI), LEHD Origin-Destination Employment Statistics (LODES), and Job-to-Job Flows (J2J). More information about the LEHD data is available in Abowd et al. (2009).

Private-industry employment: Covered private-industry employment in the LEHD data includes most corporate officials, all executives, all supervisory personnel, all professionals, all clerical workers, many farmworkers, all wage earners, all piece workers, and all part-time workers. Workers on paid sick leave, paid holiday, paid vacation, and the like are also covered. Workers on the payroll of more than one firm during the period are counted by each employer that is subject to UI, as long as those workers satisfy the preceding definition of employment. Workers have UI wages filed in every quarter they are covered, even though their wages may not be subject to UI tax in the latter months of the year.

Notable exclusions from UI coverage among private sector employers are independent contractors, the unincorporated self-employed, railroad workers covered by the railroad unemployment insurance system, some family employees of family-owned businesses, certain farm workers, students working for universities under certain cooperative programs, salespersons primarily paid on commission, and workers of some non-profits. States have some leeway in designating coverage, for a complete list, see the coverage section of the most recent Comparison of State UI laws (https://oui.doleta.gov/unemploy/statelaws.asp).

State and local government employment: Covered employment in the LEHD data includes most employees of state and local governments with the exception of elected officials, members of a legislative body or members of the judiciary, and some emergency employees.

Federal government employment: Federal government workers are not covered by state UI. LEHD uses data from the Office of Personnel Management (OPM) to generate earnings and employment histories for federal workers. The OPM data covers most federal employees but excludes White House officials, members of Congress, and certain national security agencies, which are excluded for security reasons. Members of the armed forces and the U.S. Postal Service are not covered in OPM data. The OPM data has coverage for 2000-2015.

UI coverage across years: Availability of state UI data in the LEHD system varies by state. LEHD has data for only about ten states in the early 1990s, expanding rapidly to 40 states by the late 1990s, with Massachusetts being the last state entering the data in 2010.

top

Degree, Earnings, and Employment Concepts

Degree, Program, and Institution
Formally, the institution is defined as the 6-digit Office of Post-secondary Education ID (OPEID), and the Degree Level is one of ten values: Associates, Certificate (<1 year, 1-2 years, 2-4 years, post-baccalaureate certificate, post-masters certificate), Bachelors, Masters, Doctoral-Professional Practice and Doctoral-Research/Scholarship. For Graduate Earnings, the Bachelors and Doctoral-Professional Practice degree levels reports the Degree Field at the 4-digit Classification of Instruction Program (CIP) code level, while for all other degree levels, the Degree Field is defined at the 2-digit CIP code level. For the Employment Flows, degree field is at the 2-digit CIP level. For each university system, we process the transcript data to standardize variables and update older CIP codes to the most recent classifications (currently 2010 CIP codes). We consider students who earn multiple degrees in the system to be separate observations.

Year Post-Graduation
For all post-secondary graduates, the first year post-graduation is defined as the first calendar year following their graduation year. So for a student who graduates in May of 2005, year one begins in January of 2006, year five in January 2010, etc.

Earnings
Earnings are total annual earnings for attached workers from all jobs, converted to 2016 dollars using the CPI-U. For the annual earnings tabulations, we impose two labor force attachment restrictions. First, we drop graduates who earn less than the annual equivalent of full-time work at the prevailing federal minimum wage. Additionally, we drop graduates with two or more quarters with no earnings in the reference year. These workers are likely to be either marginally attached to the labor force or employed in non-covered employment.

Employment
While earnings tabulations include earnings from all jobs, for the flows tabulations, we report the graduate's main job for that year only. Main jobs are defined as the job for which graduates had the highest earnings in the reference year. To attach employer characteristics to that job, we assign industry and geography from the highest earnings quarter with that employer in the year. For multi-establishment firms, we use LEHD unit-to-worker imputations to assign workers to establishments, and then assign industry and geography. Workers who do not meet the labor force attachment restrictions outlined above are assigned an employment location 'ZZ' and industry sector 'ZZ' (which we treat as the residual category).

top

Tabulation Levels

Graduate Earnings tables, which summarize earnings outcomes for graduates, are at the Institution (6-digit OPEID), Degree Level, Degree Field, Graduation Cohort, and Year Post-Graduation level.

Degree Level is one of ten values: Associates, Certificate (<1 year, 1-2 years, 2-4 years, post-baccalaureate certificate, post-masters certificate), Bachelors, Masters, Doctoral-Professional Practice and Doctoral-Research/Scholarship. For the Bachelors and Doctoral-Professional Practice degree levels, the Degree Field is defined at the 4-digit CIP code level, while for all other degree levels, the Degree Field is defined at the 2-digit CIP code level. Graduation Cohorts are defined as follows: For the Bachelor's degree level, the graduation cohorts are three-year cohorts, e.g. 2001-2003; 2004-2006; 2007-2009; 2010-2012; 2013-2015. For all other degree levels, the graduation cohorts are five-year cohorts, e.g.: 2001-2005; 2006-2010; 2011-2015.

Employment Flows tables, which summarize earnings outcomes for graduates, are at the institution (6-digit OPEID, Degree Level, Degree Field, Graduation Cohort, Census Division, NAICS Sector level, 1, 5 and 10 years after graduation. We also provide counts for individuals who are in-state, where the state is defined as the state of the institution. Differing from above, degree field is reported at the 2-digit CIP for all degree levels.

top

Comparability to Other Data

The College Scorecard is a data product released by the U.S. Department of Education beginning in 2013 and focuses on entering cohorts of students and their earnings ten years after initial enrollment, although they report longer-term outcomes as well. The Department of Education produces this product by matching federal financial aid data to IRS tax records. They report the 10th, 25th, 50th, 75th, and 90th percentiles of earnings for students. However, the College Scorecard sample frame is only students who received federal financial aid, and these earnings may not reflect those of the entire population of students from that institution.

Additionally, a number of states (including Texas, Colorado, and North Carolina) have released similar tabulations of graduate earnings by matching graduate records to in-state unemployment insurance records. While this match allows them to measure the earnings of graduates that stay within the state, these estimates are biased downwards, as mobility and higher wages are positively correlated.

top

Protection System

The protection system for PSEO must take into account that external parties have access to a large portion of the data. To address these disclosure issues, the PSEO data product uses differential privacy techniques to protect individual confidentiality. We describe the specific method we use for protecting the Graduate Earnings tabulations below. For more information, see the Technical Appendix for PSEO Protection System (/doc/pseo_appendix_tech_doc.pdf) (//www.census.gov/main/www/pdf.html) (225 KB)

Graduate Earnings Tabulations
In the Graduate Earnings tabulations, we release three percentile values (25th, 50th and 75th) and a cell count. To protect the earnings percentiles for a given cell, we categorize the earnings of all individuals into pre-defined histogram bins. We then add noise according to the geometric mechanism to each bin. We use these counts to construct an empirical CDF, from which we calculate the percentiles. We also calculate the protected cell count from the sum of the bin counts. For cells with a protected count of less than 30, we suppress output (due to low quality) and indicate the suppression in the data with a status flag = 5.

Graduate Employment Outcomes Tabulations

We use a geometric noise mechanism to protect the Graduate Employment Outcomes tabulations, since the data are counts.

top

Feedback

Please send questions and comments to CES.PSEO.Feedback@census.gov (mailto:CES.PSEO.Feedback@census.gov).

top

References

[1] John M. Abowd, Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators. In Producer Dynamics: New Evidence from Micro Data, NBER Chapters, pages 149-230. National Bureau of Economic Research, Inc., September 2009.

top