MAPS STUDENT TRENDS AND ENROLLMENT PROJECTIONS DASHBOARD METHODOLOGY

SORENSON IMPACT CENTER
DAVID ECCLES SCHOOL OF BUSINESS

OVERVIEW

The MAPS Student Trends and Enrollment Projections (STEP) Dashboard visualizes the relationship between demographic and enrollment data, student migration patterns, and the potential impacts of trends in distance learning. The dashboard provides higher education leaders with an understanding of how and where undergraduate students are choosing to attend college, how that has changed over time, and what enrollment could be like in the future. These actionable insights can help inform recruitment efforts, strategic planning, resource allocation, and innovative investment opportunities.

This dashboard was developed by a team of data scientists and social impact consultants at the Sorenson Impact Center at the University of Utah. The dashboard displays IPEDS data at the institution level and US Census Bureau data at the state level. Data are disaggregated by race, Pell recipient status, and other characteristics when available from these two data sources.

This work is a part of the MAPS Project (Model, Analyze, Prototype, Share) created by the Sorenson Impact Center at the University of Utah and supported by the Bill and Melinda Gates Foundation.

What the Model Provides

The MAPS Landscape Model utilizes both state-level and sector-level data from several sources and disaggregates by student characteristics where possible.

Created in partnership with the Kem C. Gardner Policy Institute at the University of Utah, this tool displays state-level population projections out to 2030 by age, sex, and racial/ethnic categories using historical data from the following sources:

- US Census Bureau National Population Projections
- University of Virginia Weldon Cooper Center National Population Projections

The tool also includes data from:

- Integrated Postsecondary Education Data System (IPEDS)
- US Department of Education College Scorecard

Population and Enrollment Trends

The dashboard shows US Census population data on young adults between the ages of 18-24 by race/ethnicity and gender. State-level population estimates are projected out to 2030 using various regression methods. Alternative projections based on high, mid, and low migration scenarios that may impact future state populations of this particular age group are also incorporated. Enrollment data by race/ethnicity and gender are shown as total undergraduates enrolled with no age range stipulation. Various methods are applied to historical enrollment data to project future enrollment in higher education institutions out to 2030. Enrollment projections are also combined with population projections to establish a ratio between the number of available students to predicted enrollment.

Student Migration Trends

The dashboard shows how students are enrolling at higher education institutions either inside or outside of their home states. This dashboard visualizes answers to the following questions:

- Of the first-year, first-time, full-time students currently enrolled at institutions within a state, what percentage are from in-state? From out of state?
- What percentage of first-year, first-time, full-time students enroll in their home state? How many enroll in out-of-state schools?
- Where do students leaving their home state go most often?
- Which states are contributing most to the out-of-state student population of a particular state?

Distance Learning Trends

Distance learning data is shown as the total of exclusively remote undergraduate enrollment in a selected state broken down by students from in-state vs. out-of-state. The dashboard shows how exclusively remote learning has changed over the past several years.

METRICS

Metrics across the following categories in the STEP Dashboard are from these sources:

Metrics	Source
State-level demographics	US Census
School sector and state classifications	IPEDS
Enrollment data by demographics	IPEDS
First-year student migration	IPEDS
Distance learning	IPEDS
State-level demographic projections for 2030	Kem C. Gardner Policy Institute

Primary Metrics

The metrics included in the STEP Dashboard were selected based on the intended audience and the applicable scope and presentability of a metric, given the audience. STEP's intended audience is those interested in a broad, demographic view — along the lines defined by the US government — of postsecondary education.

A fundamental issue being addressed by the STEP Dashboard is: To what extent will changing demographics impact postsecondary education in the United States? As such, the initial metric

selection was one of demographics at a suitable granularity projected out to the year 2030. For these projections, we partnered with the Kem C. Gardner Policy Institute at the University of Utah to establish projections at the state-level for three scenarios. Sorenson Impact Center also independently examined US Census demographics data at the state and city levels.

Next, the central concern was whether demographics at the state-level are relevant for the US postsecondary landscape. For most students and institutions, the answer was yes. First year student migration data collected by the Integrated Postsecondary Education Data System (IPEDS) indicates that, up to now, prospective students from a state tend to attend schools within the state.

Another significant trend in education is that of distance learning. Data from IPEDS show that distance education had been steadily trending up even before the pandemic. For some sectors, distance learning is the main mode of education.

METHODOLOGY

Disclaimer

The MAPS Project represents a concerted effort to improve equitable student outcomes in postsecondary education. The applications were developed with the assumption that any organization or individual who uses them has a baseline access to data and tools. These applications cannot serve the specialized needs of every individual or group and do not prescribe specific solutions. Where available, please defer to additional sources of data and consult specialists on the use of data and models for your needs.

General Approach

Starting with the thesis that shifting demographics may affect the future operational model of postsecondary education, the approach was to explore what information was available around that theme and then consult users of postsecondary data on attention areas. Effort was then taken to focus on a target audience and pare down the scope of the application. A major criteria was on clarity of presentation and usability for the intended audience.

Demographic Projections

We engaged experts at the Kem C. Gardner Policy Institute at the University of Utah to provide experimental demographic projections for all 50 states and Washington D.C. out to the year 2030. Independently, the data science team at the Sorenson Impact Center examined population estimates from the United States Census Bureau in their relations to postsecondary data from IPEDS.

Relevance of Demographics

An important question to address was the relevance of demographics at the state-level to postsecondary education as a whole. For this purpose, the demographics of enrolled undergraduate students as well as the state-to-state migration patterns of first-year students from IPEDS data were analyzed. Both sets of data were needed; for instance, the aggregate enrolled undergraduates at a state could have come equally from every available source. At the same time, the demographics of enrolled undergraduate students were needed because migration data were not broken down by demographics.

Enrollment Projections

These projections are guided by models that were derived from historical data. Our models reflect trends that, left alone to evolve without new or strong change of forces like a change in policies at a large school, would continue.

The enrollment tab in the dashboard attempts to align demographics with enrollment and projects plausible relationships out to 2030. As previously mentioned, demographics projections for 2030 were done independently by the Kem C. Gardner Policy Institute. The Sorenson Impact Center retrieved historical data from the US Census and IPEDS, and aligned the data along demographic categories, which will be referred to as OMB7, as defined by the Office of Management and Budget (OMB). Also, enrollment data classification of race/ethnicity extends beyond that defined by the OMB — Nonresident Alien and Race Unknown are included. Note on IPEDS' adoption of the OMB categorization: "Reporting using the new categories is mandatory for the Fall Enrollment component in 2010-11 IPEDS." Furthermore, the OMB is considering revisions to the current approach.

A plausible connection between demographics and postsecondary enrollment is that young adults of a certain age are the primary consumers of postsecondary education. Following IPEDS' age categorization of undergraduates to age 25 and above and those under the age of 25, we grouped census demographics data and projections by gender, ages 18-24, and the OMB7 categorization established by the OMB. The lower bound on age of 18 was selected based on the idea that young adults of that age are serious candidates for applying for postsecondary schools. Most undergraduate students have acquired a degree or have made a decision on pursuing a degree by age 25.

For enrollment data, prospective users emphasized that grouping of educational institutions into different 'sectors' was important for their use cases. The STEP Dashboard uses the same sector as the other MAPS applications:

- Private, 2-year
- Private, 4-year or above
- Proprietary, 2-year
- Proprietary, 4-year or above
- Public, 2-year
- Public, 4-year or above

A table showing the alignment of data is as follows:

Source	Year	Sector	State	OMB7	Gender
Census	X		X	X	
Gardner	X		X	X	
Enrollment	X	X	X	X	
Migration	X	X	X		
Distance Learning	X	X	X		

US Census estimates and Gardner Policy Institute projections can be viewed as the pool from which a vast majority of undergraduates at postsecondary institutions are enrolled from. Enrollment into various sectors reflect the supply and demand dynamics amongst various participants.

A quantity that aids the analysis of the propensity for various demographic groups to enroll at different types of postsecondary institutions in a state is the ratio of enrollment-to-census projections. This ratio shows trends that can be difficult to determine from a visual inspection of census or enrollment data alone. For example, if the population of a certain demographic group is increasing but at a faster rate than enrollment, then this ratio slopes downward. This ratio increasing or decreasing versus time reflects preferences of prospective students and postsecondary institutions for different types of education.

The sheer number of combinations of groupings, large differences in quantities amongst the groups, and the limited number of years of aggregated data points steered modeling to using the simple, tried, and trusted: different varieties of statistical regression for each group. Visually, each regression model suggests plausible paths given historical data - and projections for census data. Each model incorporates some form of prior information or characteristics of the historical data. By default, census data are modeled with the ‘MED’ projections provided by the Gardner Policy Institute. These scenarios provide an ‘anchor’ point for the year 2030. Users can choose one of ‘LOW’, ‘MED’, ‘HIGH’ scenarios for the ‘anchor’. The multiple models are averaged to provide a single curve that aims to be a good estimate of a likely outcome. Note that the data points do not all fall on the curve. Statistical models assume that the model is being fitted to data points with errors.

Enrollment models and projections for each group were done using historical data only. The average of enrollment models, census models, and the ratio of enrollment-to-census can all be found on the ‘Overview’ section of the ‘Student Enrollment’ tab. At times, the ‘curves’ can look ‘out of place’ for those who have a better understanding of the data for their specific region or target group. In such cases, opinions of specialized experts should be preferred.

Details of the models averaged to obtain the single curves are in the ‘Breakdown’ section. Specific models can be plotted against data by selecting them in the legend. Variations between models can

give an indication of approximate ‘uncertainty’ of projections. Tightly bunched curves suggest that the averaged values are likely a good approximation. Whereas different models showing a wide range of values suggest sizeable uncertainty. A brief description of the component models follows:

- Linear regression - data fluctuates around a linear trend.
- Cubic regression splines - data fluctuates around smoothly joined piecewise cubic polynomials. The last three data points have the most influence on projections.
- Penalized b splines - data fitted using B-splines where fit is determined by data and a penalty function that favors smoothness.
- NPREG smoothing splines - more flexible approach to modeling where the algorithm is given more discretion in choosing coefficients and emphasizing certain data points.
- GAM loess - locally linear smoothing of data that is more adept at handling outliers.
- REML method - focuses more on accurately portraying the fluctuations of data points.

Favoring a specific model for projections may be more appropriate than using the ensemble (average.) Users may have more insight or knowledge about the data than can be modeled with available data.

Student Migration

IPEDS record two categories of first-year students. Those who are first-time degree/certificate-seeking undergraduate students, and those who are first-time degree/certificate-seeking undergraduate students that graduated from high schools in the past 12 months. The STEP Dashboard focuses on the former group.

The primary means to determine the in-state and out-of-state designations of an applicant is by the residency address associated with the application at the time they apply. IPEDS have procedures for imputing residency where appropriate. Only reported or imputed entries were used in dashboard.