MAPS

# 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 |

**Note on Institution inclusion / exclusion:** The MAPS project makes a concerted effort to maintain data consistency. To that end, data is selected for inclusion contingent on usability across the three core MAPS dashboards (STEP, Financial Health, Institutional Equity Outcomes). For various reasons, an institution's data may have been excluded, for example:

- Merger and acquisition may have obscured the identity of an institution
- Multi-campus organizations could have compromised the identification process
- Complicated organizational structures can confuse the sector association process
- Omission of data reported to IPEDS, especially with regard to financials, will remove an organization from inclusion

## 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 a changing demographics impact postsecondary education in the United States. As such, the initial metric was one of demographic at a suitable granularity projected out to the year 2030. For this metric, we partnered with the Gardner Policy Institute at the University of Utah to establish projections at the state-level for three scenarios. Sorenson Impact Center also independently examined 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 Integrated Postsecondary Education Data System (IPEDS) indicates that, up to now, prospective students from a state tend to attend schools within the state.

Previous work on the ongoing demographics shift in the US and the likely consequences on postsecondary education focused on the demand-side. Our intent is to extend the scope to encompass supply and demand. Enrollment data from IPEDS, in conjunction with census data, seem to be a good proxy for supply and demand dynamics for higher education.

A potentially significant trend in education is that of distance learning. The prospect of students attending schools remotely, regardless of state lines would drastically change the status quo. Data from IPEDS show that distance education had been steadily trending up (and down too) before the pandemic. For some sectors, distance learning is the main mode of education. The pandemic has had a profound impact on this metric, and the lasting consequences are subjects of future concern.

## Other Metrics

In addition to the primary metrics that were selected for inclusion in the current STEP Dashboard, there were other metrics that were explored but not included at this stage. These are described below for those interested in getting a sense of the directions where MAPS may expand.

For prospective users like Association of Governing Boards, interests extended to financial outcomes and demographics of enrolled students from originating states. While IPEDS has enrollment demographics of students, the data does not differentiate demographically regarding where the students came from - just the number of students from another state. We built models to infer the demographics, but this work was deemed too premature for inclusion. For financial outcomes, we looked to The Department of Education College Scorecard data. In particular, we examined dropout rates at 2 yrs and 6 yrs for the general student-body as well as for low- to mid- income students and average median debt incurred for students who dropped out or graduated. In addition, our intent was to tie the outcomes to demographics. While gender was available, other classifications like Pell-grant

recipients status and household income levels were not. We did begin modeling the distribution of household incomes for each state for 2010-2020 and had preliminary results, but has not been further developed for inclusion in the current STEP Dashboard.

Another group of prospective users were economic development planners for a state. These users had various focus on demographics and types of graduates. One focus group of interest were undergraduate students beyond the typical college going age, the age 25+ category. We compiled data from IPEDS but did not incorporate the data into the STEP Dashboard because of the limited audience.

A third set of users were from admissions offices. While they were interested in the broad demographics of high-school age students within and outside the state (e.g. for placement of satellite recruitment offices), they were also interested in specific groups at a more granular level for recruitment efforts. For instance, students from rural areas or based on household income levels. For these users, we explored correlating census data to regional high school data or first year postsecondary students. Our determination was that high school data was too sporadic and regionally siloed and the time-delayed correlation between census data and first year students needed future study.

---

# 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 a changing United States 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. Having assessed the gamut of needs and interests, 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 useability for the intended audience. Specifically, our intended audience is strategic-minded users of postsecondary education data who are interested in a broad perspective.

## Demographic Projections

Demographics is an established field of study. Since 2015, the Kem C. Gardner Policy Institute has produced long-term demographic and economic planning projections for the state of Utah and its counties. The MAPS project engaged experts at the Kem C. Gardner Policy Institute to provide experimental demographic projections for all 50 states and Washington D.C. for the year 2030. See here for further details. 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 'migration pattern' 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. Note that details of both of these sets of data will be described in subsequent sections. Time will tell if postsecondary education will become borderless. But for now, a sizable majority of undergraduates students in the US postsecondary education system attend public schools. In terms of aggregate numbers then, a study of the available data indicates that demographics at a state-level do matter. For other sectors, the in-state and out-of-state distinction can be less significant.

The main device for showing data is that of a map of the United States with the intent of highlighting regional differences in the quantity of interest. Alternative or augmented views of the data are shown as charts, often plotted against time, to highlight the temporal aspect of the data.

## Enrollment Projections

With respect to models and projections, care must be taken to temper our confidence of their validity. First, projections are guided by models that were derived from historical data. And historically, the data we are working with, in aggregate form, evolves not too widely over the time of observation. Second, the expectation is that the factors driving the dynamics behind the data, in total, will not affect the observed quantity too erratically nor drastically. Our models reflect trends that, left alone to evolve without new or strong change of forces (such as a change in policies at a large school), would continue to behave 'as it were'. Unfortunately, the pandemic that took hold in 2020 was an event on the scale and force that caveats about models and projections should be redoubled.

Enrollment reflects supply and demand, and one of our goals is to help post secondary decision makers maintain a balance in supply and demand as demographics change. The enrollment tab 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 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 planning/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 sectors as the other MAPS dashboards:

- 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 | X |
| Gardner | X | | X | X | X |
| Enrollment | X | X | X | X+ | X |
| Migration | X | X | X | | |
| Distance Learning | X | X | X | | |

US Census estimates and Kem C. 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 Kem C. 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 developed 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’ (FYS). 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. STEP focuses on the former group.

Note that reporting of residency status of FYS is mandatory in even-numbered years and optional in odd-numbered years. Our decision was to include the optional data to highlight this fact. We found this data useful as the mandatory data provided upper-bounds on the ‘flow’ of students into and out of states. Full reporting of this data would be a welcome addition in supporting the construction of models for inferring information that is not otherwise available - such as demographic of FYS from various states.

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

While the data shown on the map in the ‘Migration Patterns By State’ section has a direct correspondence with the filters used to choose the data, the data in ‘Migration Chart (State Level)’ is selected by state. The ‘Inbound Locations’ and ‘Outbound Locations’ tab summarizes the number of FYS opting to relocate from another state or attend an institution in another state, respectively. The ‘Inbound Locations’ tab shows the percentage of FYS attending a postsecondary institution in a state whose residency was in another state (top five.) The ‘Outbound Locations’ tab shows the percentage of FYS who are residents of the selected state and went to another state (top five) for their postsecondary education.

# Distance Learning

A third major theme explored with the STEP Dashboard is the uptake of remote learning at postsecondary institutions in the United States. If geographic location becomes less of a deciding factor on where or how a student would attend school, that would represent a major, systemic shift.

IPEDS collect data on distance learning and have four categories:

- ‘Exclusive Remote’
- ‘Exclusive Remote In-State’
- ‘Exclusive Remote Out-of-State’
- ‘Some Remote’

‘Exclusive Remote’ students do not attend classes in a traditional classroom setting. They acquire their education online. The ‘Exclusive Remote In-State’ and ‘Exclusive Remote Out-of-State’ designations indicate the residency status of the exclusive remote student. This set of data seems to be rather new

and some institutions are still likely working on their reporting and classification capabilities for this type of data. Consequently, reported ‘Exclusive Remote’ is not always the sum of ‘Exclusive Remote In-State’ and ‘Exclusive Remote Out-of-State’, although the difference is typically small. Inconsistencies are most notable with the ‘Some Remote’ classification. The ‘Some Remote’ data was not included in the STEP Dashboard due to not negligible inconsistencies in the ‘Some Remote’ data and the difficulty in describing the nuances of this classification.

Stating that the COVID-19 pandemic had a significant impact on distance learning would not be an overstatement. The lasting effects on distance learning and education in general will likely take some time to assess. The STEP Dashboard attempts to present the trends in online education pre-pandemic and show the magnitude of changes introduced by COVID-19.

The ‘Distance Learning By State’ section depicts a geographic view of data as filtered by the selectors to the left of the map. The ‘Breakdown By Sector (State Level)’ sub-tab highlights and contrasts distance learning trends versus time for different post secondary sectors of a state.