

# Codebook for Statistics on the College Pipeline

Updated as of 08.03.2023

---

## Online Data Table 1: Pipeline Analysis by College and Parental Income Bin

This table reports our pipeline analysis numbers by college and parental income bin for students in the US who took the SAT or ACT in 2011, 2013, or 2015 which we can link to parent incomes. The table is sorted alphabetically based on colleges' name. Our primary measure of parental income is total household-level pre-tax income (as described in detail in Section 2.5 of Chetty, Deming, Friedman 2023).

This table uses the Pipeline Analysis sample, which we construct by starting from the raw income tax data and retaining the subset of individuals who (1) have a valid Social Security Number (SSN) or Individual Taxpayer Identification Number (ITIN), (2) can be linked to parents, and (3) appear in either the SAT or ACT data in 2011, 2013, or 2015 (as described in detail in Section 2.1). As in other pipeline analysis graphs and tables, we rank then group parental income into 13 bins for pooled national statistics, as well as a 14th bin that pools students in the p99-99.9 and top 0.1% bins into a top 1% bin (for in-state and out-of-state statistics and all test score band statistics we provide only the pooled top 1% bin.) All parental income percentiles are based on households' percentile in the national income distribution for households with children born in a given year.

Following established disclosure standards, we report estimates rather than exact values of the statistics for each college in this table. The estimates are quite accurate: the estimation error is comparable to the fluctuation in the true statistics across years for a typical college that arises due to sampling error. Appendix E of the paper describes the procedure used to construct these estimates. In places where the resulting value after adding the noise is negative, we replace the observation with a missing value.

We construct three measures of attendance, application, and attendance conditional on application. First, we construct simple attendance and application rates for students with test scores in a specific narrow band. Second, we construct a more comprehensive statistic that averages together the score-specific attendance and application rates using the shares of attending students at each college as the weights. Third, we construct simple (unweighted) attendance and application rates. We calculate each of these six statistics and then add noise; we then calculate the attendance conditional on application rate as the ratio of the post-privacy attendance rate to the post-privacy application rate.

We also construct average attendance and application rates for each college, calculated as the attendance (application) rate for each college that would prevail if all students attended (applied) at the same average rate as all other students with the same test score. We also construct simple (unweighted) averages of these rates for each college. We add noise to these college-specific statistics as well and then report relative

attendance (application) rates as the ratio of the absolute attendance (application) rate to the college average. The relative attendance (application) rate can be interpreted as the ratio of the number of students currently attending (applying to) a given school from a given parental income background to the number of students that would attend (apply to) the school if all students did so at the same average rate as other students (irrespective of parent income) who share the same test score. The unweighted relative rates can be interpreted as the ratio of the number of students currently attending (applying to) a given school from a given parental income background to the number of students that would attend (apply to) the school if all students did so at the same rate. We further calculate the relative attendance rate conditional on application as the ratio of relative attendance rate to relative application rate.

We report statistics for 139 selective colleges in the U.S., comprised of the following groups:

- Ivy-Plus colleges: the eight Ivy League colleges, Stanford, Duke, MIT, and Chicago
- Other highly selective private colleges: 9 of the highest ranked private colleges (excluding the Ivy-plus) according to the 2022-2023 U.S. News and World Report for National Universities for which we have data
- 9 highly selective public flagship colleges for which we have data
- All remaining private colleges in Barron's top selectivity category, other than the Ivy-plus
- New England Small College Athletic Conference (NESCAC) colleges
- All other private and public colleges that are in the 100 highest ranked schools according to the 2022-2023 U.S. News and World Report for National Universities
- All other flagship public colleges

Note that we omit some colleges from the categories above because we are unable to distinguish between specific campuses of multi-campus state universities or have insufficient coverage in the administrative data we use for our analysis.

For each college, we release the following variables and their standard errors.

<b>Variable Name</b>	<b>Description</b>
super_opeid	Institution OPEID / Cluster ID when combining multiple OPEIDs
name	Name of college (or college group)
par_income_bin (par_income_lab)	Parent household income group (and the label) based on percentile in the income distribution

rel_apply	Test-score-reweighted relative application rate: Calculated using adjusted score-sending rates, the relative fraction of all standardized test takers who send test scores to a given college.
attend rel_attend	Test-score-reweighted absolute and relative attendance rate: Calculated as the fraction of students attending that college among all test-takers within a parent income bin in the Pipeline Analysis Sample. Relative attendance rates are reported as a proportion of the mean attendance rate across all parent income bins for each college.
rel_att_cond_app	Calculated as the ratio of rel_attend to rel_apply.
rel_apply_unwgt	Unweighted relative application rate: Calculated using adjusted score-sending rates, the relative fraction of all standardized test takers who send test scores to a given college.
attend_unwgt rel_attend_unwgt	Unweighted absolute and relative attendance rate: Calculated as the fraction of students attending that college among all test-takers within a parent income bin in the Pipeline Analysis Sample. Relative attendance rates are reported as a proportion of the mean attendance rate across all parent income bins for each college.
rel_att_cond_app_unwgt	Calculated as the ratio of rel_attend_unwgt to rel_apply_unwgt.

rel_apply_sat	Relative application rate for specific test score band based on school tier/category. Selected test score band is the 50-point band that had the most attendees in each school tier/category. Below is the selected range Ivy Plus: SAT 1460-1510 Elite Public: SAT 1180-1230 Top Private: SAT 1410-1460 NESCAC: SAT 1370-1420 Tier 2 Private: SAT 1290-1340 Top 100 Private: SAT 1170-1220 Top 100 Public: SAT 1110-1160 Other Flagship: SAT 1070-1120
attend_sat rel_attend_sat	Absolute and relative attendance rate for specific test score band based on school tier/category
rel_att_cond_app_sat	Relative attendance rate, conditional on application, for specific test score band based on school tier/category
rel_apply_instate rel_apply_oostate	Test-score-reweighted relative application rate for in-state (out-of-state) students. In-state status is measured using the students' address when they take a standardized test. Only available for public schools.
attend_instate rel_attend_instate attend_oostate rel_attend_oostate	Test-score-reweighted absolute and relative attendance rate for in-state (out-of-state) students. Only available for public schools.
rel_att_cond_app_instate rel_att_cond_app_oostate	Test-score-reweighted relative attendance rate, conditional on application, for in-state (out-of-state) students. Only available for public schools.
[variable]_unwgt_instate rel_[variable]_unwgt_instate [variable]_unwgt_oostate rel_[variable]_unwgt_oostate	Unweighted absolute and relative estimates for in-state (out-of-state) students. Only available for public schools. Absolute estimates only available for attend.
[variable]_instate_sat rel_[variable]_instate_sat [variable]_oostate_sat rel_[variable]_oostate_sat	Absolute and relative estimates on a specific test-score for in-state (out-of-state) students. Only available for public schools. Absolute estimates only available for attend.
stderr_[variable]	Standard error for a given variable described above.

[variable]_level	The school average estimates reweighting on test score. Divide the test-score-reweighted absolute variables by this average to calculate the test-score-reweighted relative variables.
[variable]_unwgt_level	The unweighted school average estimates. Divide the unweighted absolute variables by this average to calculate the unweighted relative variables.
tier (tier_name)	<p>Selectivity and type combination:</p> <ul style="list-style-type: none"> <li>1 = Ivy-Plus (Ivy League colleges plus Stanford, Chicago, Duke, and MIT)</li> <li>2 = Other elite college (Barron's top selectivity category, other than the Ivy-plus, both public and private combined)</li> <li>3 = Highly selective public college (Barron's 2nd selectivity group)</li> <li>4 = Highly selective private college (Barron's 2nd selectivity group)</li> <li>5 = Selective public college (Barron's 3rd, 4th, and 5th selectivity groups)</li> <li>6 = Selective private college (Barron's 3rd, 4th, and 5th selectivity groups)</li> </ul> <p>See Chetty, Friedman, Saez, Turner, and Yagan (2020) for more information on how the tier is defined</p>
flagship	Indicator for public flagship universities (defined using the College Board Annual Survey of Colleges, 2016)
public	Indicator for public universities
test_band_tier	School group for the test-score band statistics