SAID in Graphics, Contest Data Description
Data Point – Parent and Student Expectations of Highest Education Level (NCES 2019015)

This Data Point report is available at: https://nces.ed.gov/pubs2019/2019015.pdf

The full High School Longitudinal Study of 2009 (HSLS:09) data set is restricted. However, many of the variables are in the publicly available version, including those used for this report. The data set in your team DropBox is derived from these, but in some cases has had observations with missing variables removed, categories collapsed, and contains only a single weight variable (called Weight). In particular, it only contains students who responded to the 2016 follow-up and whose parents responded in 2012.

This data set is not to be used for any research projects as it is missing some information to be used correctly in those circumstances. It is provided here for demonstration purposes.

Figure 1 in the Data Point is based on variables P1EdExp and P2EdExp.

Figure 2 in the Data Point is based on variables S1EdExp, S2EdExp, and S4EdExp.

As your data set (including the weights used in constructing the graphs) have been modified, you will not be able to reproduce the values in the graphs exactly, but will be close. There are also additional variables provided in your data set that may be used, as described below.

The Variables

The dataset in your DropBox is derived from the publicly released data from the High School Longitudinal Study of 2009 (HSLS:09). The initial participants were in 9th grade in 2009, with follow-ups in 2012 and 2016. Additional information on this survey and its data can be found at: https://nces.ed.gov/surveys/hsls09/ . The particular variables included in your portion of the data are described below.

WEIGHT In order to be representative of the target population, weights are used for each observation. The total of the weights (4,133,580) is the estimated number of students who were 9th graders in 2009.

X2RaceEth is a collapsed student race and ethnicity variable based on X2RACE. The categories are:

Black/African-American, Non-Hispanic
Hispanic
Other, Non-Hispanic
White, Non-Hispanic

X2SEX is the student sex. The categories are

Male
Female

X2DUALLANG is the student's first language. The categories are

First language is English Only
First language is a non-English language only
First language is English and non-English equally

P1SLD is whether a doctor or the school has told parent the 9th grader has a learning disability. The categories are

Yes
No

P1HONORS is the parental response to whether the 9th grader is currently enrolled in any honors classes. The categories are

Yes
No

X1TXMTSCOR is a standardized T score for a mathematics test that gives a norm-referenced measure of achievement relative to fall 2009 9th graders in 2009 (mean 50 and standard deviation 10 for the whole population). The sample values range from 24.58 to 82.19.

X2TXMTSCOR is a standardized T score for a mathematics test that gives a norm-referenced measure of achievement relative to students in 2012 who were 9th graders in 2009 (mean 50 and standard deviation 10 for the whole population). The sample values range from 26.66 to 84.91.

X1POVERTY, X1POVERTY130, and X1POVERTY185 X1POVERTY is whether the student's family was at/above or below the 2008 poverty threshold. X1POVERTY130 is for 130% of the poverty level, and X1POVERTY185 is for 185% of the poverty level. The response categories for the respective questions are:

At or above poverty threshold
Below poverty threshold

At or above 130% poverty threshold
Below 130% poverty threshold

and

At or above 185% poverty threshold
Below 185% poverty threshold

X1REGION is the geographic region of the school attended in 2009. The categories are:

Northeast
Midwest
South
West

X3CREDAPIB is the total Carnegie credits (one-year, five day a week course) in AP and IB courses, captured from the transcript after the senior year.   The values range from 0 to 13.

S1EdExp, P1EdExp, S2EdExp, P2EdExp, and S4EdExp are derived from the variables S1EDUEXPECT, P1EDUEXPECT, S2EDUEXP, P2EDUEXP, and S4EDUEXP involving how far in school the student thinks they will get (S1, S2, and S4) and how far in school the parent thinks they will get (P1, P2). They were asked in 2009 (P1, S1), 2012 (P2, S2), and 2016 (S4).   The response categories for each are

Less than high school
High, school diploma or GED
Occupational certificate or associate's degree
Bachelor's degree
Master's degree
Professional degree
Don't know

Professional degree includes a Ph.D., M.D., law degree, or other high level professional degree. Starting a degree but not finishing was counted as the category below. [The base data file appears to have two categories collapsed for P2EDUEXP, and your data set goes with the higher of the two as appears to have been done in the Data Point].

X2ParEdu is derived from X2PAREDU, the highest level of education achieved by either parent. The educational levels are:

Less than high school
High school diploma or GED
Occupational certificate of associate's degree
Bacehlor's degree
Master's degree
Professional Degree