**National Household Education Surveys Program:   
Parent and Family Involvement** **Background**

**Description of the Data Source**

Data about parent involvement in education and school choice from kindergarten to high school are currently collected in the [Parent and Family Involvement in Education (PFI)](https://nces.ed.gov/nhes/parent_involvement.asp). The PFI survey collects data about students who are enrolled in kindergarten through grade 12 and asks questions about school choice and various aspects of parent involvement in education, such as help with homework, family activities, and parent engagement in schools. **Parents (or guardians) are the respondents**.

This National Household Education Survey (NHES) is conducted for the Department of Education by the [US Census Bureau](https://www.census.gov/programs-surveys/nhes.html).

There are different topics covered by NHES. The Fall Data Challenge will use data from the Parent and Family Involvement in Education (PFI) topic. More information on the PFI topic can be found [here](https://nces.ed.gov/nhes/parent_involvement.asp).

**2022 Fall Data Challenge Data**: The focus of the Fall Data Challenge is on students in Public or Private schools. Here is a description of data prep steps regarding whether or not the child is in Public school:

* Those records (rows in the table or responses) for homeschooled children were removed.
* Whether a child is in Public school or Private school is indicated by the variable **EDCPUB**. The original datasets had different types of Private schools, but the Fall Data Challenge is only concerned with Public and Private, regardless of the type of private school.
* The 2016 entries indicating type of school were changed to fit the 2019 variable. So, the **EDCPUB** in the 2016 file is a derived variable, and there won’t be an entry for it in the official codebook found online.

The **required** dataset for the 2022 Fall Data Challenge is the **curated 2019 data**. Teams may use the 2016 data, but this is optional as is the use of external data sources.

**The 2022 Fall Data Challenge dataset is available** [**here**](https://docs.google.com/spreadsheets/d/11rno0SjsOyV1JlevH3q40UMm1s5e6m7g/edit?usp=sharing&ouid=110729637035212744419&rtpof=true&sd=true)**.**

**Overview of the Dataset**

The **BASMID** is a unique identifier. It is not typically used as a variable or characteristic in a model. Also, this is not longitudinal, meaning there is no connection in IDs or respondents from 2016 to 2019.

Number of variables, including the **BASMID**:

* 2019: 75
* 2016: 69

Number of observations/rows.

* 2019: 15,500
* 2016: 13,463

**2022 Fall Data Challenge Prompt:** Imagine you are serving as a consultant to a school district. Your team has been engaged to provide insights to the district on enhancing family involvement. Using the 2022 Fall Data Challenge dataset, how will you advise the district?

**NOTE**: It’s important to note that the number of responses in the official Codebooks will not match what is in the data set because the rows for home-schooled children were removed.

The curated dataset codebook is available[**here**](https://docs.google.com/spreadsheets/d/1X7fDKA_BZjSDwUCIo_3fax1viY1_wcJH/edit?usp=sharing&ouid=108879154306217142702&rtpof=true&sd=true)**.**

You may find it helpful to refer to the full Codebook files for valid entries and definitions.

These websites have some other files some might find useful for a deeper dive into the data fields.

* 2019 Codebook: <https://nces.ed.gov/nhes/data/2019/pfi/cbook_pfi_pu.pdf>
* 2016 Codebook: <https://nces.ed.gov/nhes/data/2016/pfi/cbook_pfi_pu.pdf>

The User Manuals for the survey are not necessary for the analysis. However, the following manuals can be consulted for those who are interested in more details on the entire survey, not just the specific topic area. Topics include the overview/purpose of the survey, survey design, sampling methodology, data collection, and more.

* 2019 User Manual: <https://nces.ed.gov/pubs2021/2021030rev.pdf>
* 2016 User Manual: <https://nces.ed.gov/pubs2018/2018100.pdf>

Several variables in the 2019 data set were not collected in 2016. The list is given below:

* CHLDNT How often does this child use the Internet at home for learning activities?
* EINTNET Child enrolled in online, virtual or cyber courses
* INTNUM How many online, virtual, or cyber courses does this child take?
* MOSTIMPT Of the reasons that this child is enrolled in online, virtual, or cyber courses, which one would you say is the most important to you?
* SCCHOICE Did you feel that you had a choice in what school this child attends?
* SCHLHRSWK About how many hours does this child attend a school each week?

The dataset for the 2022 Fall Data Challenge was curated from **Parent and Family Involvement in Education (PFI)** the original data is available at:

* 2016 files: <https://nces.ed.gov/nhes/dataproducts.asp#2016dp>
* 2019 files: <https://nces.ed.gov/nhes/dataproducts.asp#2019dp>