

STRATEGIC **DATA** PROJECT

# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

---

**A GUIDE FOR CONDUCTING DATA  
ANALYSIS IN EDUCATION AGENCIES**

Please mute your computer speakers and phone microphone.



**3. Connect**

[www.gse.harvard.edu/sdp/tools](http://www.gse.harvard.edu/sdp/tools)



**STRATEGIC DATA PROJECT**

# **MISSION**

*Transform the use of data in education to improve student achievement.*



## **Patty Diaz**

Senior Program Manager, Fellows

## **Todd Kawakita**

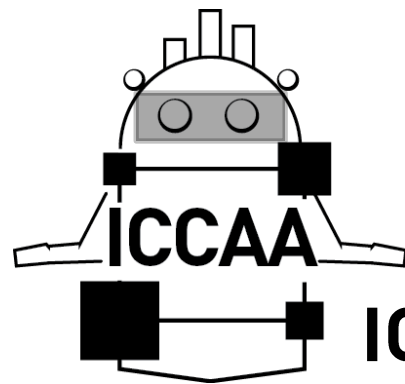
Manager of Product Development





## **Patty Diaz**

Senior Program Manager, Fellows



**ICCAA**

Toolkit Mascot



## **Todd Kawakita**

Manager of Product Development

# CONNECT WEBINAR:

1

## SDP TOOLKIT

FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA  
ANALYSIS IN EDUCATION AGENCIES

2



3. Connect

3

Q & A

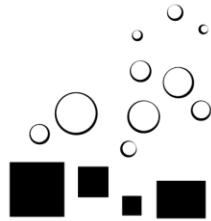
# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

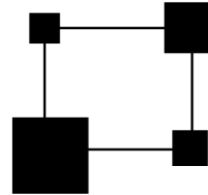
A GUIDE FOR CONDUCTING DATA  
ANALYSIS IN EDUCATION AGENCIES



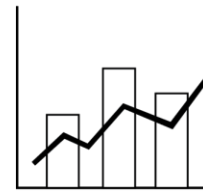
**1. Identify**  
essential data  
elements



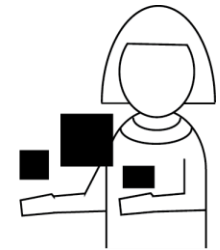
**2. Clean,**  
check, and  
build variables  
for your  
datasets



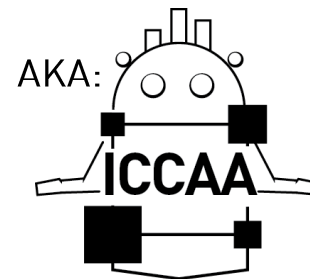
**3. Connect**  
relevant  
datasets from  
different  
sources



**4. Analyze**  
your datasets



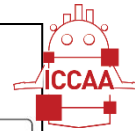
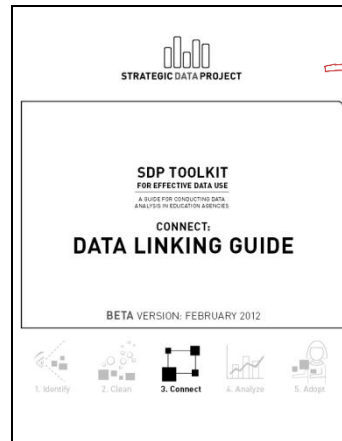
**5. Adopt**  
best practices  
to facilitate  
shared and  
replicable data  
analysis



# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES



I



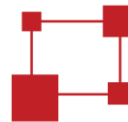
**Identify:** Data Specification Guide

C



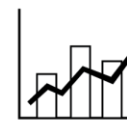
**Clean:** Data Building Tasks

C



**Connect:** Data Linking Guide

A

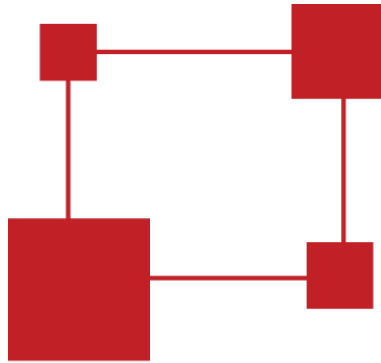


**Analyze:** Diagnostic Analyses Guide

A



**Adopt:** Coding Style Guide

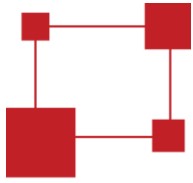


## 3. Connect

### Data Linking Guide

Now that you have collected and cleaned your data, you will want to merge the files together to create an analysis file.

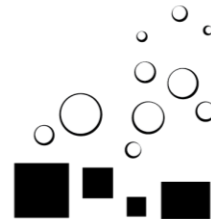




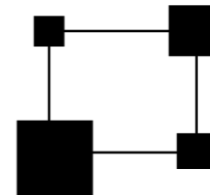
### 3. Connect

## Learning Objectives:

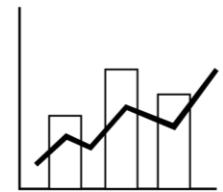
1. Produce comprehensive student level-files that track high school completion and graduation, using **Clean** and **Identify** data files;
2. Link post-secondary college enrollment and persistence data, available through the National Student Clearinghouse (NSC) to your agency's student achievement records;
3. Merge disparate research files to create a single, robust analysis file that will support your work in the **Analyze** stage.



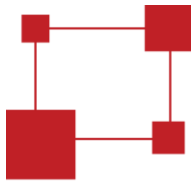
### 2. Clean



### 3. Connect

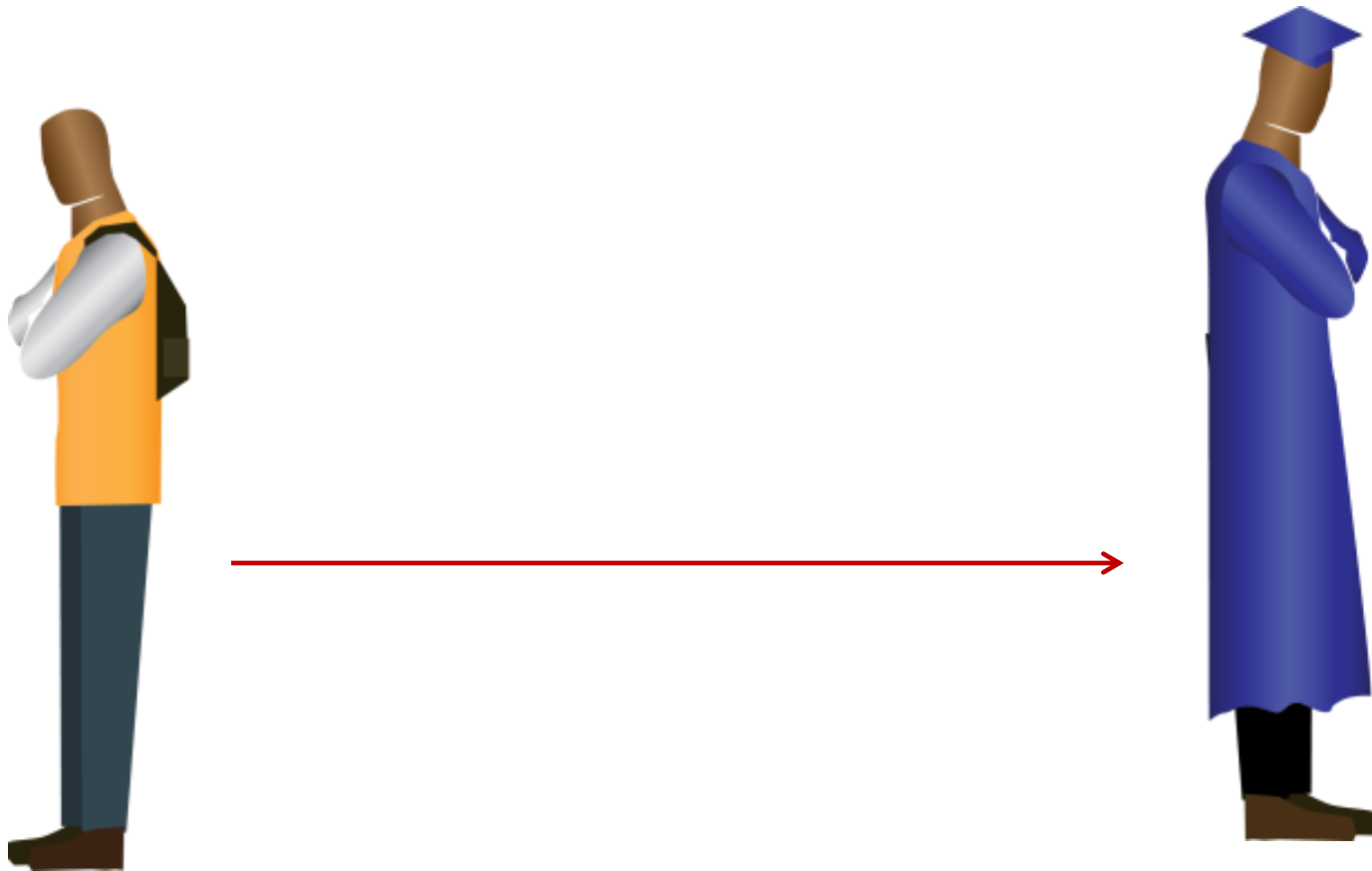


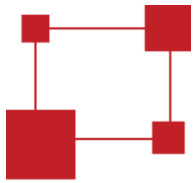
### 4. Analyze



### 3. Connect

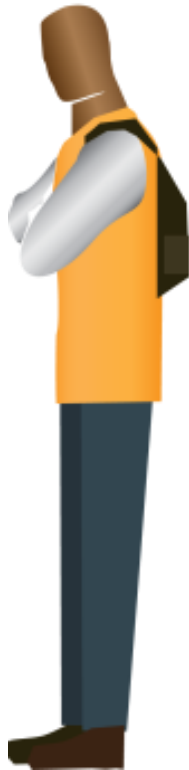
Q: How do we track student progression from high school through college?



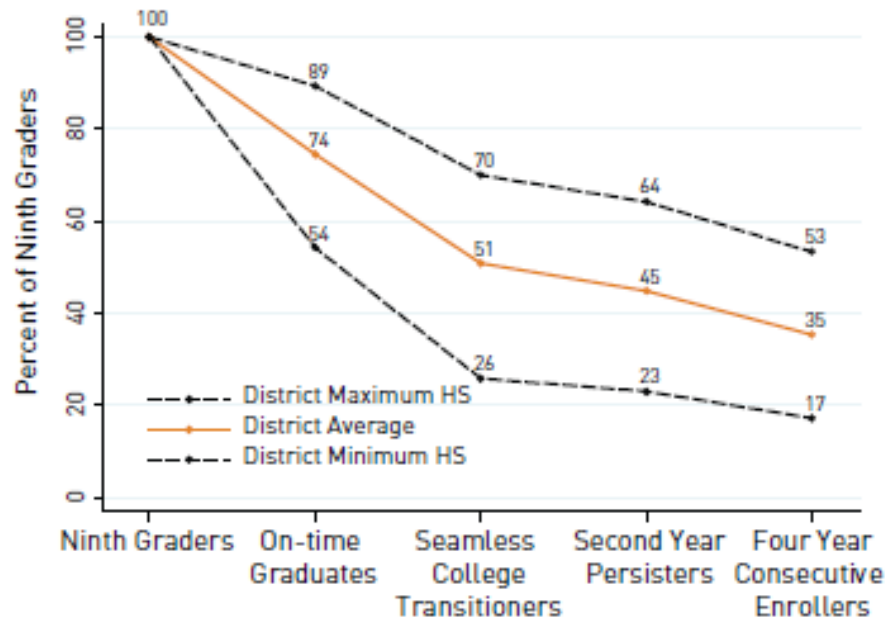


### 3. Connect

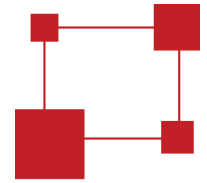
Q: What information do we need to understand high school progression, graduation and college enrollment?



Student Progression from Ninth Grade through College



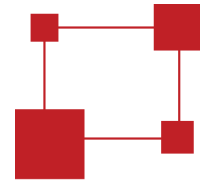
# Final File: Analysis File



## 3. Connect

CG_Analysis_File									
1	sid	12	first_hs_id	21	ontime_grad	28	test_math_8	37-39	first_college_opeid_*
2	male	13	last_hs_id	22	late_grad	29	test_ela_8	40-42	first_college_name_*
3	race_ethnicity	14	longest_hs_id	23	stil_enrl	30	test_composite_8	43-45	enrl_1oct_grad_yr1_*
4	hs_diploma	15	first_hs_name	24	transferout	31	test_math_8_std	46-48	enrl_1oct_grad_yr2_*
5	hs_diploma_date	16	last_hs_name	25	dropout	32	test_ela_8_std	49-51	enrl_ever_w2_grad_*
6	frpl_ever	17	longest_hs_name	26	ged	33	test_composite_8_std	52-54	enrl_grad_persist_*
7	iep_ever	18	lastwithcode_group	27	disappear	34	qrt_8_math	55-57	enrl_grad_all4_*
8	ell_ever	19	chrt_ninth			35	qrt_8_ela	58-60	enrl_1oct_ninth_yr1_*
9	frpl_ever_hs	20	chrt_grad			36	qrt_8_composite	61-63	enrl_1oct_ninth_yr2_*
10	iep_ever_hs							64-66	enrl_ever_w2_ninth_*
11	ell_ever_hs							67-69	enrl_ninth_persist_*
								70-72	enrl_ninth_all4_*

# Final File: Analysis File



## 3. Connect

CG_Analysis_File									
1	sid	12	first_hs_id	21	ontime_grad	28	test_math_8	37-39	first_college_opeid_*
2	male	13	last_hs_id	22	late_grad	29	test_ela_8	40-42	first_college_name_*
3	race_ethnicity	14	longest_hs_id	23	stil_enrl	30	test_composite_8	43-45	enrl_1oct_grad_yr1_*
4	hs_diploma	15	first_hs_name	24	transferout	31	test_math_8_std	46-48	enrl_1oct_grad_yr2_*
5	hs_diploma_date	16	last_hs_name	25	dropout	32	test_ela_8_std	49-51	enrl_ever_w2_grad_*
6	frpl_ever	17	longest_hs_name	26	ged	33	test_composite_8_std	52-54	enrl_grad_persist_*
7	iep_ever	18	lastwithcode_group	27	disappear	34	qrt_8_math	55-57	enrl_grad_all4_*
8	ell_ever	19	chrt_ninth			35	qrt_8_ela	58-60	enrl_1oct_ninth_yr1_*
9	frpl_ever_hs	20	chrt_grad			36	qrt_8_composite	61-63	enrl_1oct_ninth_yr2_*
10	iep_ever_hs							64-66	enrl_ever_w2_ninth_*
11	ell_ever_hs							67-69	enrl_ninth_persist_*
								70-72	enrl_ninth_all4_*



Student Attributes  
+ Student School  
Year



School  
Enrollment



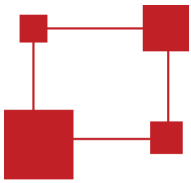
HS Indicators  
+ Outcomes



Prior  
Achievement



NSC Data



### 3. Connect

Q: But wait, what's up with the NSC file?

www.studentclearinghouse.org

Harvard Graduate Sch... Home 5 Harvard Grad... HGSE Printers HGSE Xdrive

**NATIONAL STUDENT CLEARINGHOUSE®**  
The nation's trusted source for student degree and enrollment verification,™

**USER LOGIN**  
User Name  
Password

Colleges & Universities  
High Schools & Districts  
Students & Alumni  
Guarantors, Lenders & Servicers  
Degree & Enrollment Verifiers

Forgot password? ▶ Audit Resource Center ▶ FTP Login ▶ About Us ▶ Careers ▶ Contact Us Search

**Order or Transcript**

**Verify a Degree, Past Attendance or Certification**

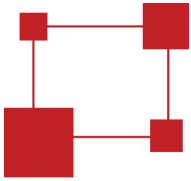
- Participating Schools
- Certification Providers
- Join & Save

**Verify Current Enrollment**

- Participating Schools
- Join & Save

**Coming Soon:  
The New Clearinghouse Web Site**

Visit [www.studentclearinghouse.org/preview](http://www.studentclearinghouse.org/preview) for a look at what's in store when we launch



### 3. Connect

# Student Tracker

The screenshot shows the website [www.studentclearinghouse.org/highschools/default.htm](http://www.studentclearinghouse.org/highschools/default.htm). The page features a navigation bar with links for FTP Login, About Us, Contact Us, and Home. A search bar is also present. The main content area is divided into three columns. The left column contains a sidebar with a menu for 'StudentTracker for High Schools' including links for Outcomes Report, Case Studies, Testimonials, How to Join, Getting Started, Electronic Transcript Exchange, and How to Join. A red arrow points from the 'Student Tracker' text to the 'How to Join' link. Below the menu is a box that says 'COMING SOON: The next generation of StudentTracker for High Schools! Read more now.' The middle column has a heading 'Measure the College Success of Your High School Graduates' and a list of key questions. The right column contains a 'Resource Center' with various links and a call to action for students and alumni.

**USER LOGIN**

User Name  
Password

**National Student Clearinghouse**

High Schools & Districts    FTP Login    About Us    Contact Us    Home    Search

▶ **StudentTracker for High Schools**

- Outcomes Report
- Case Studies
- Testimonials
- How to Join
- Getting Started
- Electronic Transcript Exchange
- How to Join

**COMING SOON: The next generation of StudentTracker for High Schools! Read more now.**

### Measure the College Success of Your High School Graduates

StudentTracker for High Schools is a unique program designed to help high school districts more accurately gauge the college success of their graduates by answering such key questions as:

- ▶ How many of my high school graduates enroll in college?
- ▶ Do they persist and graduate from college?
- ▶ How long does it take for them to get their degree?
- ▶ Do they go in-state or out-of-state and do they attend a 4-year or 2-year school?
- ▶ Which colleges do they most commonly attend?
- ▶ Did any students go on to college who started 9th grade in my high school but did not graduate?

### Benefits of Using StudentTracker

- ▶ Tracks your students' college attendance nationwide
- ▶ Provides turnkey professional-quality reports at an affordable price
- ▶ Covers 93% of all U.S. college enrollment
- ▶ Includes all types of institutions: 2-/4-year, graduate, public/private, trade, vocational, etc.
- ▶ Helps you accurately assess your college preparation efforts
- ▶ Measures your graduates' persistence in college
- ▶ Complies with FERPA
- ▶ Sign up for StudentTracker for High Schools and get FREE [Electronic Transcript Exchange](#)

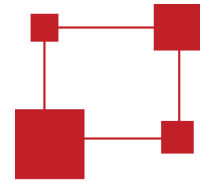
To sign up or learn more, contact [highschools@studentclearinghouse.org](mailto:highschools@studentclearinghouse.org).

**Resource Center**

- Clearinghouse Colleges
- StudentTracker Help
- Secure FTP Access
- Sample Report
- Detail Report Guide
- Flyer & FAQs
- FERPA Compliance
- Research Services Web Site
- Contact Us

**Are you a student or alumni? Visit [MyStudentCenter.org](http://MyStudentCenter.org).**

# 10 Steps (p.4 of Connect)



## 3. Connect

### Prepping, some merging

1. Prior Achievement: Part I
2. School Crosswalk
3. Student Attributes
4. Student School Year
5. Student School Enrollment

### High School Outcomes, more merging

4. High School Indicators & Outcomes
5. Prior Achievement: Part II
6. Examining the Analysis File: Part I

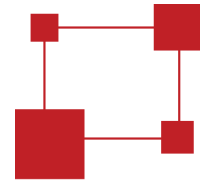
### College Enrollment and Persistence, final merge

8. National Student Clearinghouse Data
9. Examining the Analysis File: Part II



# Structure of 10 Steps

Step 4 Example



**3. Connect**

Purpose

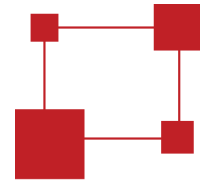
Files Needed  
(Input)

At the End ...  
(Output)



# Structure of 10 Steps

## Step 4 Example



3. Connect

Purpose



Files Needed  
(Input)

Merge on Student School Year

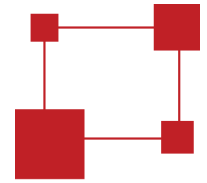
STUDENT ATTRIBUTES	
sid	
male	
race_ethnicity	
hs_diploma	
hs_diploma_type	
hs_diploma_date	

STUDENT SCHOOL YEAR	
sid	
school_year	
grade_level	
first_9th_school_year_reported	
frpl	
iep	
ell	

At the End ...  
(Output)

# Structure of 10 Steps

## Step 4 Example



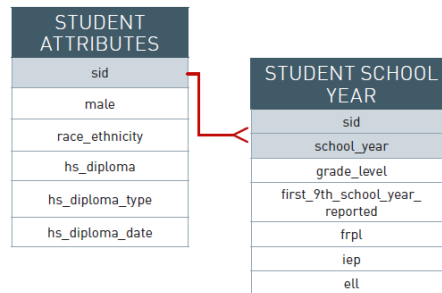
3. Connect

Purpose



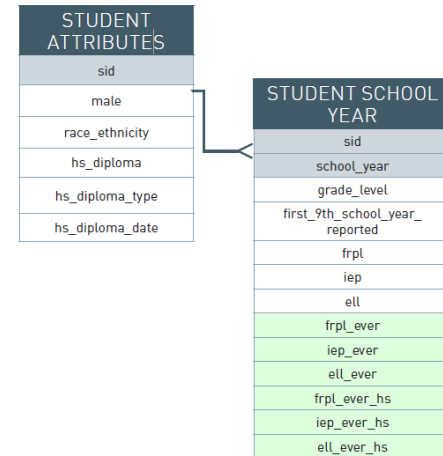
Files Needed  
(Input)

Merge on Student School Year

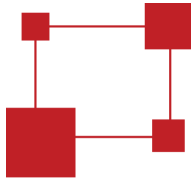


At the End ...  
(Output)

Current State of the Analysis File



variables you just generated:  
program participation status  
variables for ever qualified and  
qualified in high school



### 3. Connect

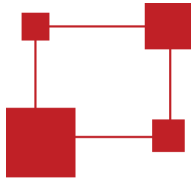
Prepping the Files, Some Merging

Step 2 Example

School

X

Walk



### 3. Connect

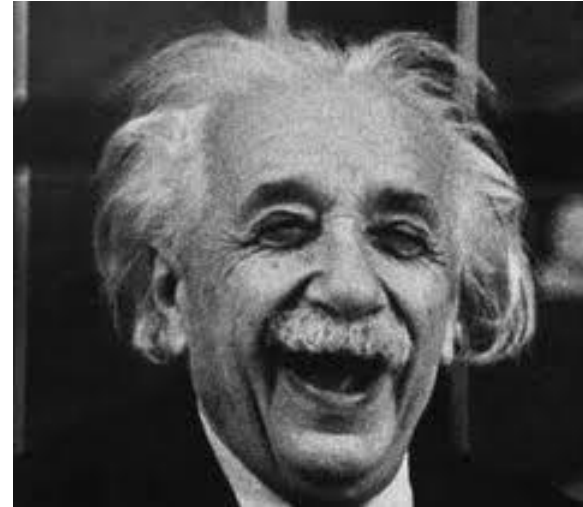
Prepping the Files, Some Merging

Step 2 Example

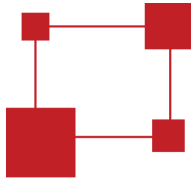
School

X

Walk



Patty's Alma Mater



### 3. Connect

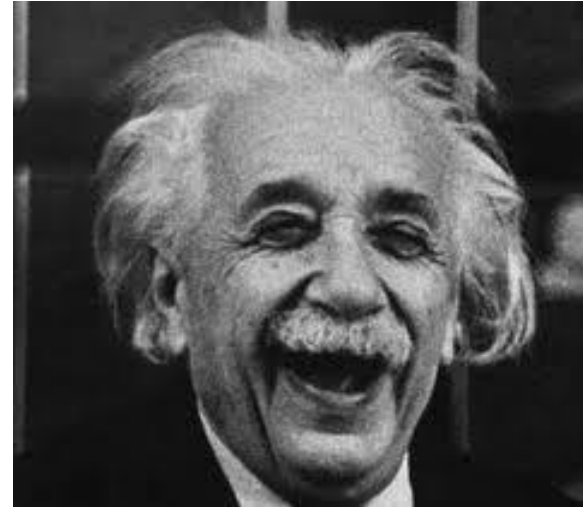
Prepping the Files, Some Merging

Step 2 Example

School

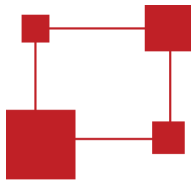
X

Walk



Patty's Alma Mater

- A. Einstein
- Albert Einstein
- Alert Einst
- AS



### 3. Connect

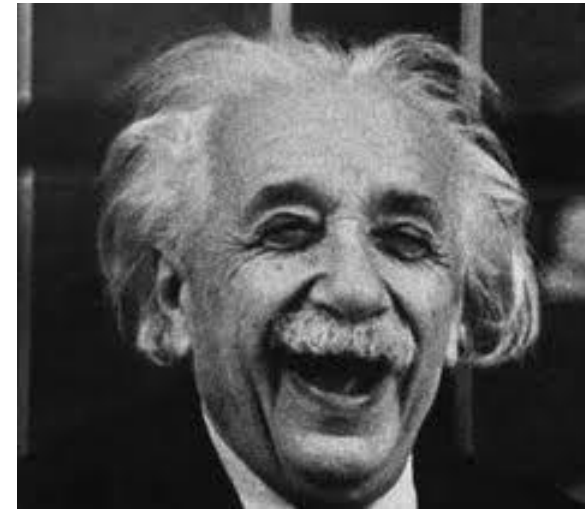
## Prepping the Files, Some Merging

Step 2 Example

# School

# X

# Walk



Patty's Alma Mater

#### Load the File and Check Uniqueness

SCHOOL
school_code
school_name
grade_span

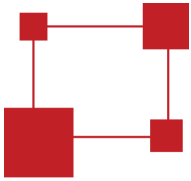
```
// load School
use School.dta, clear

// keep only schools that contain high school grades
keep if grade_span == "9-12"
drop grade_span
```

A crosswalk will help you to ensure that the final file is unique by school\_code and that a single school\_code maps to only one school\_name. For example in an uncleaned file, Albert Einstein High School might be spelled three different ways, "A. Einstein HS," "Einstein High School," or "A.E. HS," but have the same school\_code. Alternatively, Jones High School might show up as 153 and 154. Fix these issues before moving onward.

```
// check that the file is unique by school_code
isid school_code
```

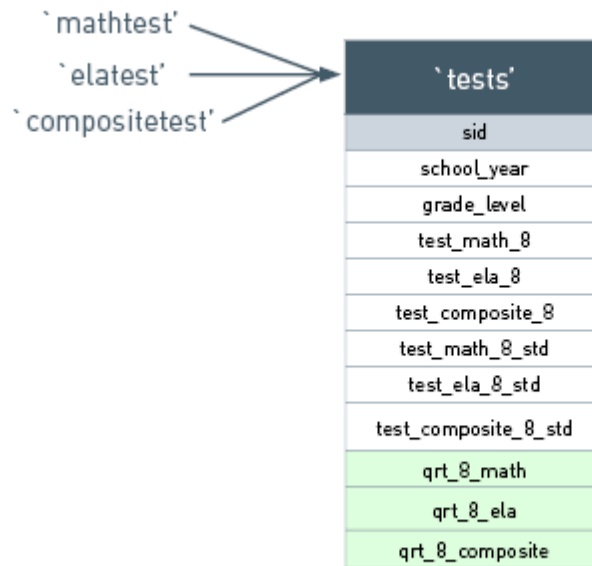
- A. Einstein
- Albert Einstein
- Alert Einst
- AS



### 3. Connect

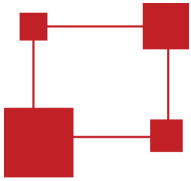
## Prepping the Files, Some Merging

### Prior Achievement



Pull assessment data;  
Create quartiles based on test scores  
Tell it to “  
hang tight”





## High School Outcomes, More Merging

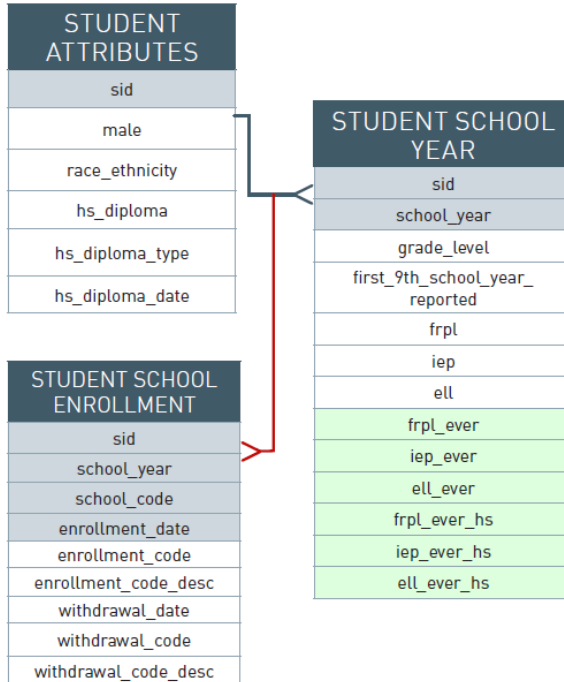
### 3. Connect

## Student Attributes, School Year & Enrollment

narrative



### Merge on Student School Enrollment

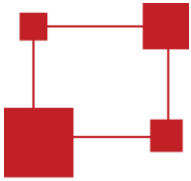


Merge the Student School Enrollment file onto the analysis file in memory. This will allow you to identify the high schools students were enrolled in at different points in time.

This is a 1:m type on sid + school\_year, as the data file resulting from the previous two steps is unique by sid + school\_year and the Student School Enrollment file is unique by sid + school\_year + enrollment\_date.

```
// merge on Student School Enrollment from Task 4
merge 1:m sid school_year using Student_School_
Enrollment_Output.dta
```

Note that before conducting the merge, the Student School Enrollment file should be unique by sid, school\_year, school\_code, and enrollment\_date. This means that each row is an enrollment instance for a student in a school\_year at a different school with a different enrollment\_date, and that there should be no overlapping enrollment instances at the same school.



## Merging

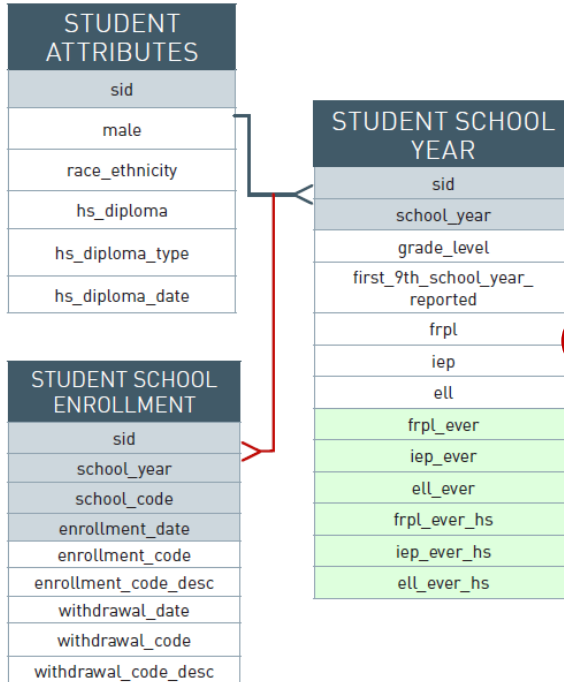
### 3. Connect

## Student Attributes, School Year & Enrollment

narrative



### Merge on Student School Enrollment



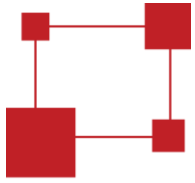
Merge the Student School Enrollment file onto the analysis file in memory. This will allow you to identify the high schools students were enrolled in at different points in time.

This is a 1:m type on sid + school\_year, as the data file resulting from the previous two steps is unique by sid + school\_year and the Student School Enrollment file is unique by sid + school\_year + enrollment\_date.

**code**

```
// merge on Student School Enrollment from Task 4
merge 1:m sid school_year using Student_School_
Enrollment_Output.dta
```

Note that before conducting the merge, the Student School Enrollment file should be unique by sid, school\_year, school\_code, and enrollment\_date. This means that each row is an enrollment instance for a student in a school\_year at a different school with a different enrollment\_date, and that there should be no overlapping enrollment instances at the same school.

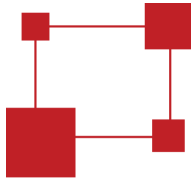


## High School Outcomes, More Merging

### 3. Connect

## High School Outcomes

First, last and longest high school  
Ninth grade cohort  
Graduation cohort



## High School Outcomes, More Merging

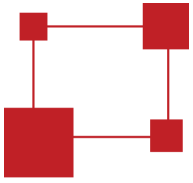
### 3. Connect

## High School Outcomes

First, last and longest high school  
Ninth grade cohort  
Graduation cohort



On-time Graduate  
Late Graduate

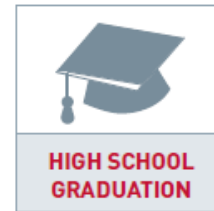


## High School Outcomes, More Merging

### 3. Connect

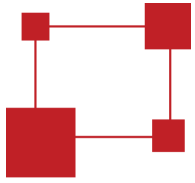
## High School Outcomes

First, last and longest high school  
Ninth grade cohort  
Graduation cohort



On-time Graduate  
Late Graduate

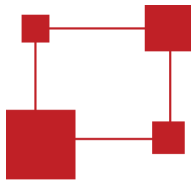
Still Enrolled  
Transferred Out  
Drop Out  
GED/Adult Ed./Certificate Program  
Disappear



### 3. Connect

High School Outcomes, More Merging

Prior Achievement, Part II  
(p.20 of Connect)



### 3. Connect

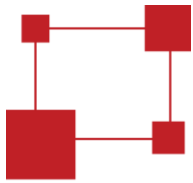
## Final Merge

Student_CollegeGoing							
1	sid	12	first_hs_id	21	ontime_grad	28	test_math_8
2	male	13	last_hs_id	22	late_grad	29	test_ela_8
3	race_ethnicity	14	longest_hs_id	23	sti_enrt	30	test_composite_8
4	hs_diploma	15	first_hs_name	24	transferout	31	test_math_8_std
5	hs_diploma_date	16	last_hs_name	25	dropout	32	test_ela_8_std
6	frpl_ever	17	longest_hs_name	26	ged	33	test_composite_8_std
7	iep_ever	18	lastwithcode_group	27	disappear	34	qrt_8_math
8	ell_ever	19	chrt_ninth			35	qrt_8_ela
9	frpl_ever_hs	20	chrt_grad			36	qrt_8_composite
10	iep_ever_hs						
11	ell_ever_hs						

## Examine the Intermediate Analysis File: Check Your Work

### Connect, Page 21:

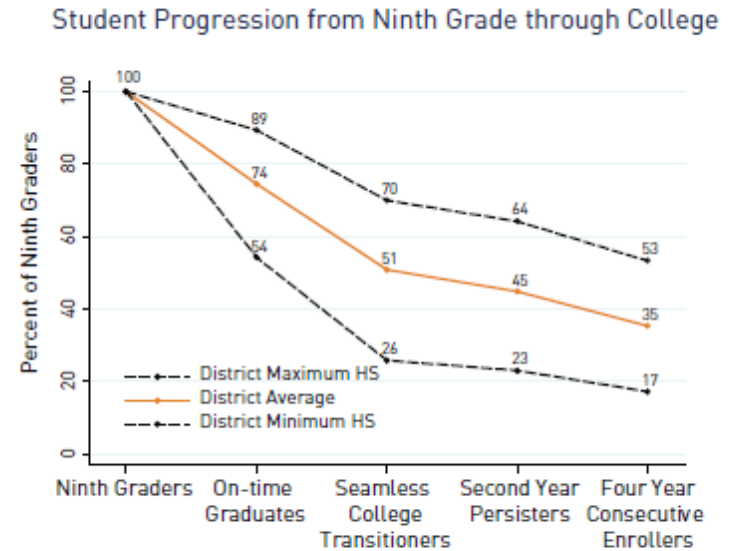
1. (1-5): Which research file are these variables from?
2. (6-11): Which research file are these variables derived from?
3. (12-14): How are first, last, and longest high school codes identified?
4. (15-17): From what research file are the first, last, and longest high school names obtained?
5. (18): What does the **lastwithcode\_group** variable describe?
6. (19-20): How are the ninth grade and graduation cohorts defined?
7. (21-27): How are graduation and high school enrollment outcomes defined?
8. (28-36): Which research files are these variables from?



Final Merge

### 3. Connect

## National Student Clearinghouse Data



You're interested in 2-year, 4-year, and any college enrollment

- Year-by-year enrollment
- Enrollment within two years of graduation

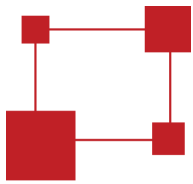
Also interested in persistence

- Persistence to the second year of college
- Continuous persistence

### Some Cautions:

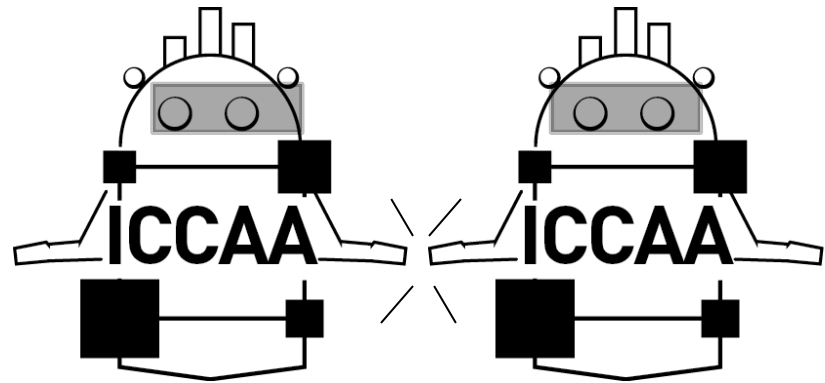
- Not your data
- Some students will get dropped when you merge the files.  
*Can anyone tell me why?*
- "Cut-off" dates matter here





Final Merge

3. Connect



Bring it all Together!

CG_Analysis_File									
1	sid	12	first_hs_id	21	ontime_grad	28	test_math_8	37-39	first_college_opeid_*
2	male	13	last_hs_id	22	late_grad	29	test_ela_8	40-42	first_college_name_*
3	race_ethnicity	14	longest_hs_id	23	sti_enrl	30	test_composite_8	43-45	enrl_1oct_grad_yr1_*
4	hs_diploma	15	first_hs_name	24	transferout	31	test_math_8_std	46-48	enrl_1oct_grad_yr2_*
5	hs_diploma_date	16	last_hs_name	25	dropout	32	test_ela_8_std	49-51	enrl_ever_w2_grad_*
6	frpl_ever	17	longest_hs_name	26	ged	33	test_composite_8_std	52-54	enrl_grad_persist_*
7	iep_ever	18	lastwithcode_group	27	disappear	34	qrt_8_math	55-57	enrl_grad_all4_*
8	ell_ever	19	chrt_ninth	End of high school outcomes for high school graduates and non-graduates	35	qrt_8_ela	58-60	enrl_1oct_ninth_yr1_*	
9	frpl_ever_hs	20	chrt_grad		36	qrt_8_composite	61-63	enrl_1oct_ninth_yr2_*	
10	iep_ever_hs	First, last, and longest hs; ninth grade and graduation cohorts			Prior Achievement		64-66	enrl_ever_w2_ninth_*	
11	ell_ever_hs			67-69			enrl_ninth_persist_*		
						70-72	enrl_ninth_all4_*		

Student Attributes + Student School Year

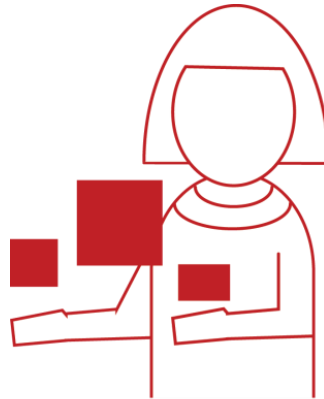
First, last, and longest hs; ninth grade and graduation cohorts

End of high school outcomes for high school graduates and non-graduates

Prior Achievement

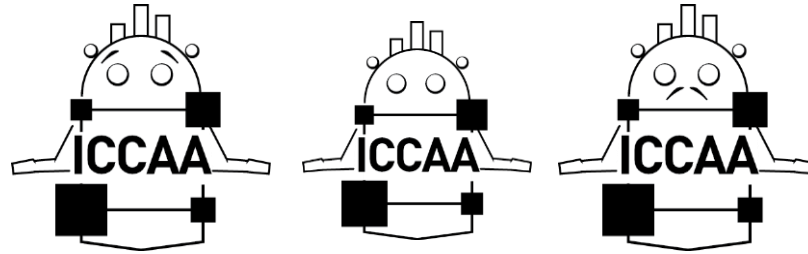
College Enrollment and Persistence Outcomes

\* consists of a 2-yr, 4-yr, and any college type version of the variable

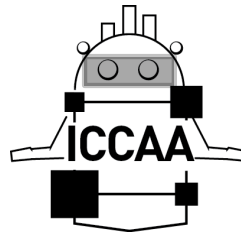


## **5. Adopt** Coding Style Guide

To ensure that statistical code is easily shared across a team and is replicable by future users, SDP and the Center for Education Policy Research (CEPR) recommends that you follow best coding, programming, and data management practices.



# Q & A



# SDP TOOLKIT

## FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES



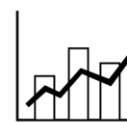
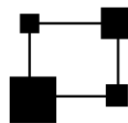
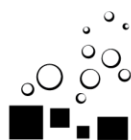
I

C

C

A

A



**Identify:** Data Specification Guide

**Clean:** Data Building Tasks

**Connect:** Data Linking Guide

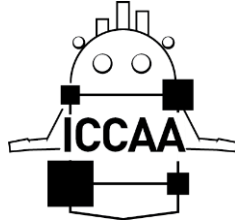
**Analyze:** Diagnostic Analyses Guide

**Adopt:** Coding Style Guide

Will Be Released Prior to Webinar On:

Thursday, February 23

# Thank You



*The toolkit is currently in **BETA**.*

*Please send us your feedback at [goo.gl/AAvdF](http://goo.gl/AAvdF).*

*Check [www.gse.harvard.edu/sdp/tools](http://www.gse.harvard.edu/sdp/tools) for the most recent toolkit version.*

*Please contact us at [sdp@gse.harvard.edu](mailto:sdp@gse.harvard.edu) if you have any questions about the toolkit.*