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.

www.gse.harvard.edu/sdp/toolkit

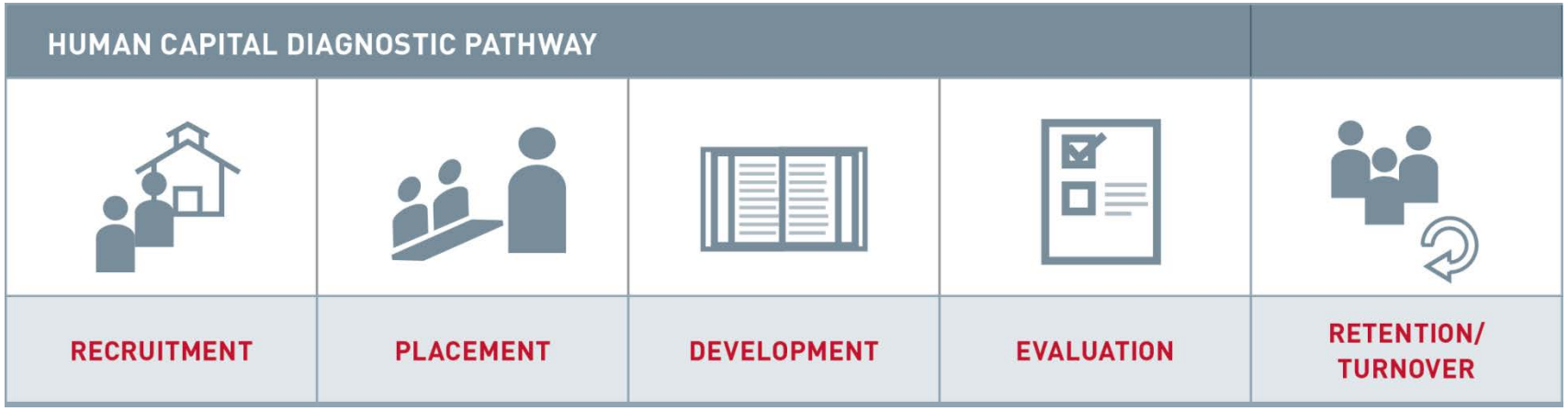
Patty Diaz-Andrade

Director of Education and Outreach

What is the toolkit?

- Highly technical document aimed at supporting analysts' work in education agencies
- Five-part resource guide for collecting, cleaning, merging, and analyzing data
- Two versions: human-capital and college-going

The SDP Human Capital Diagnostic Pathway

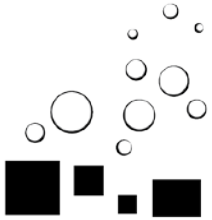


SDP TOOLKIT

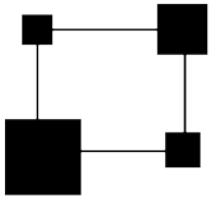
FOR EFFECTIVE DATA USE

A GUIDE FOR CONDUCTING DATA ANALYSIS IN EDUCATION AGENCIES

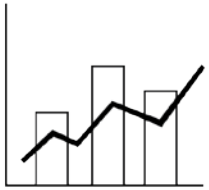
1. Identify essential data elements



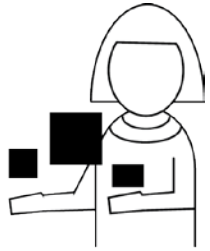
2. Clean check, and build variables for your datasets



3. Connect relevant datasets from different sources



4. Analyze your datasets



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

Toolkit Documents

An Introduction to the SDP Toolkit for Effective Data Use

Identify: Data Specification Guide

Clean: Data Building Guide for College-Going
Clean: Data Building Guide for Human Capital BETA

Connect: Data Linking Guide for College-Going
Connect: Data Linking Guide for Human Capital BETA

Analyze: College-Going Success Analysis Guide
Analyze: Human Capital Analysis Guide BETA

Adopt: Coding Style Guide

SDP Stata Glossary

VERSION: 1.2
Last Modified: September 2, 2013

| Authored by Todd Kawakita and the SDP Research Team

1. Identify

Data Specification Guide

Successful data analysis begins with proper identification of data elements necessary to answer key questions of interest.

1. **Identify**: Data Specification Guide

Identify essential data elements for analysis from your organization;

Identify: Data Specification Guide is a resource to identify data elements required to analyze student achievement, postsecondary attainment, and teacher effectiveness data. To address these different areas, we organize data elements into **research files** that contain important information at the student-, school-, and teacher- levels. These research files comprise the elements needed to **Clean**, **Connect**, and then **Analyze** your data.

These columns indicate files necessary to answer questions about college-going success or human capital.

STUDENT DATA FILES

		College-going	Human Capital	page
Student_Attributes	Demographic, cohort, and graduation data for students.	■	■	7
Student_School_Year	School year and attendance data for students.	■	■	8
Student_School_Enrollment	School enrollment/withdrawal data for students.	■	■	9
Student_Class_Enrollment	Class enrollment, grades, and credits earned data for students.	■	■	10
Student_Test_Scores	Standardized test data for students (state standardized tests, advanced placement, SAT, ACT, etc). Every attempt at a test by a student should be recorded.	■	■	11
Student_NSC_Enrollment	The National Student Clearinghouse (NSC) Student Tracker student-level data report that provides information on postsecondary outcomes.	■		12

SCHOOL DATA FILES

School	Location and classification of schools.	■	■	13
Class	Class level scheduling data.		■	14

STAFF DATA FILES

Staff_Attributes	Demographic and recruitment data of staff.		■	15
Staff_School_Year	Pay, experience, school placement, and job codes of staff.		■	16
Staff_Degrees	Educational achievement of staff. Each degree a staff member received should be recorded once.		■	17
Staff_Certifications	Teaching certifications received by staff.		■	18

Student_Attributes

Demographic, cohort, and graduation data for students.

Identifies unique observation: **sid**

Variable Name	Values or Data Type	Definition	Importance	Notes
sid	numeric	Student identifier unique to each student. This identification number is typically assigned to students upon enrollment in your agency. State agencies may have different identification numbers than district agencies for the same student.	5 Cannot Be Missing	
male	0 = female 1 = male	Student gender.	4 Absolutely Necessary	
race_ethnicity	1 = African American 2 = Asian American 3 = Hispanic 4 = American Indian 5 = White, not Hispanic 6 = Other 7 = Multiple	Student race and ethnicity. For systems where race and ethnicity are treated as a combined variable.	4 Absolutely Necessary	Use either the race_ ethnicity combined variable, or separate ethnicity and race variables If the system allows the indication of multiple categories simultaneously (e.g., African American and white) report "multiple"
race	1 = African American 2 = Asian American 4 = American Indian 5 = White 6 = Other 7 = Multiple	Student race. For systems or school years within systems where race and ethnicity are treated as separate variables.	4 Absolutely Necessary	Use either the race_ ethnicity combined variable, or separate ethnicity and race variables If the system allows for the indication of multiple categories simultaneously (e.g., African American and white) report "multiple"
ethnicity	0 = not Hispanic 1 = Hispanic	Student ethnicity. For systems where race and ethnicity are treated as separate variables and Hispanic or Latino origin is asked as a separate question.	4 Absolutely Necessary	Use either the race_ ethnicity combined variable, or separate ethnicity and race variables
birth_date	date format (yyyy-mm-dd)	Student birth date.	2 Good to Have	
first_9th_school_ year_reported	spring calendar year	The school year the student was a 9th grader for the first time. For this variable, report what the system recorded for 9th grade school year. Not all systems will record this information.	1 Not Essential	
hs_diploma	0 = no high school diploma 1 = has high school diploma	Indicator variable equal to 1 if the student received a high school diploma from the system.	4 Absolutely Necessary	
hs_diploma_type	use local values	Any locally defined description of diploma the student received. Include instances when more than one type of diploma is awarded, (i.e. Honors diploma, College Prep diploma, or General Education Diploma (GED) diploma.)	4 Absolutely Necessary	Needed when multiple types of diplomas are issued
hs_diploma_date	date format (yyyy-mm-dd)	The date on which the student received a high school diploma. If only a month and year, or only a school year is known report the partial information.	4 Absolutely Necessary	Can also be graduation date
zip_code	xxxxx or xxxxx-yyyy	The zip code of the student's home address.	1 Not Essential	

Staff_Attributes

Time invariant demographic and recruitment data related to staff.

Identifies unique observation: **tid**

Variable Name	Values or Data Type	Definition	Importance	Notes
tid	numeric	Unique staff or teacher identifier. State agencies may have different identification numbers than district agencies for the same staff/teacher.	5 Cannot Be Missing	Staff includes teachers but also other employees in the agency
male	0 = female 1 = male	Staff gender.	2 Good to Have	
race_ethnicity	1 = African American 2 = Asian American 3 = Hispanic 4 = American Indian 5 = White, not Hispanic 6 = Other 7 = Multiple	For systems where race and ethnicity are treated as a combined variable. If the system allows multiple categories (e.g., African American and white) report "multiple."	2 Good to Have	Use either the race_ethnicity combined variable, or separate ethnicity and race variables
race	1 = African American, not Hispanic 2 = Asian American 4 = American Indian 5 = White 6 = Other 7 = Multiple	For systems where race and ethnicity are treated as separate variables. If the system allows for multiple categories (e.g., African American and white) report "multiple."	2 Good to Have	Use either the race_ethnicity combined variable, or separate ethnicity and race variables
ethnicity	0 = not Hispanic 1 = Hispanic	For systems where race and ethnicity are separate and Hispanic or Latino origin is asked separately.	2 Good to Have	Use either the race_ethnicity combined variable, or separate ethnicity and race variables
birth_date	date format (yyyy-mm-dd)	Staff birth date.	1 Not Essential	
zip_code	xxxxx or xxxxx-yyyy	Zip code of the staff member's home address.	1 Not Essential	
offer_date_first	date format (yyyy-mm-dd)	The date the staff member was offered a job.	2 Good to Have	
offer_date_most_recent	date format (yyyy-mm-dd)	If the staff member left and was re-hired, the most recent date they were offered a job to work in the system.	2 Good to Have	
hire_date_first	date format (yyyy-mm-dd)	The first date the staff member was hired to work.	2 Good to Have	Can be substituted with start date or first paycheck. Used to determine late hires
hire_date_most_recent	date format (yyyy-mm-dd)	If the staff member left and was re-hired, the most recent date they were hired to work in the system.	2 Good to Have	
termination_date_first	date format (yyyy-mm-dd)	The first date staff members terminated employment.	1 Not Essential	
termination_date_most_recent	date format (yyyy-mm-dd)	If the staff member left, was re-hired and then left again, the most recent date they left the system.	1 Not Essential	
certification_path	use local values	The teacher's certification pathway. For example, "university/college" , "alternative", or "uncertified."	2 Good to Have	Where data exists, TFA or NYC Teaching Fellows can be included

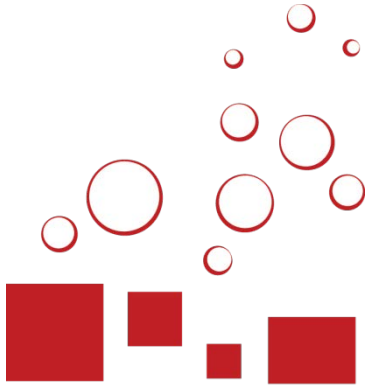
1. Identify

Help you to:

- Take stock of the data available in your agency
- Understand the data elements you will need to conduct analytics in the areas of teacher recruitment, placement, effectiveness and retention.
- Begin the process of establishing a common language for data elements across your agency

Upon completing, you will:

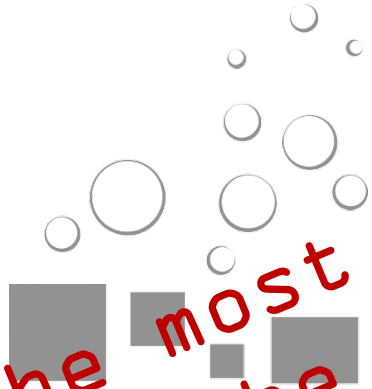
- Pull together the data you will need to successfully complete this Toolkit
- Have a good feel for your agency's data availability and perhaps begin processes for collecting data that may be missing



2. Clean

Data Building Tasks

Upon collecting essential data elements, ensure that the data can be reliably used in future analyses.



2. Clean

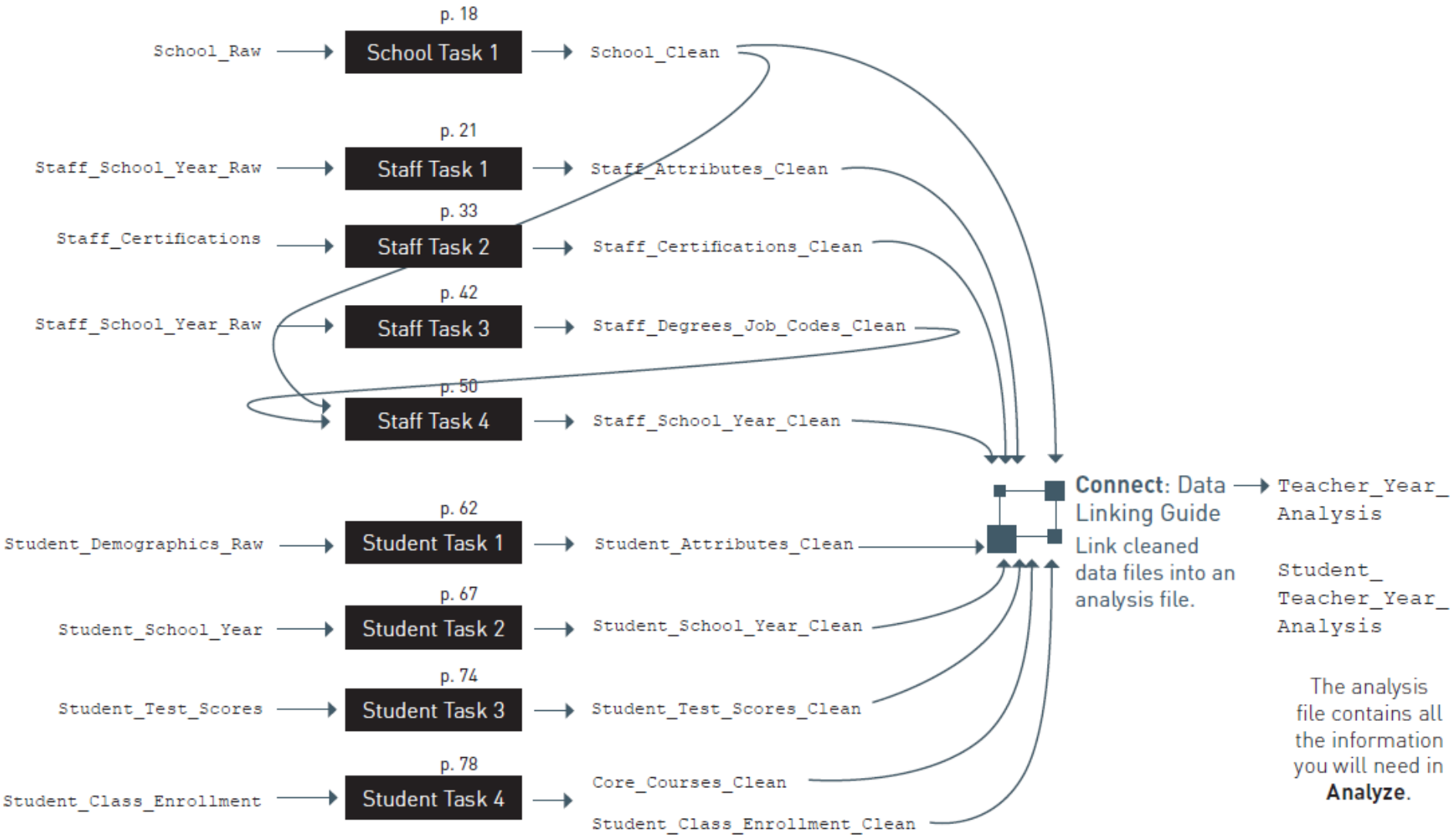
Data Building Tasks

Upon collecting essential data elements, ensure that the data can be reliably used in future analyses.

Probably the most tedious work BUT the most important work! This will take 80% of your time but its value is absolutely essential!

Task Map

This map summarizes the inputs and outputs for each task, and shows how the outputs are used in **Connect** to produce the college-going analysis and college-going analysis on-track file. The map also serves as a linked Table of Contents.



DECISION RULES GLOSSARY

This decision rules glossary provides a list of common data problems and SDP's decision rules for addressing them. It does not explain how to implement the rules step-by-step as the tasks provide this. The glossary can be a quick reference guide for recalling the decision rules and may be particularly useful for users of different analytic software.

Student_Teacher_Link

Class enrollment for students, class, teacher and grades information. Identifies unique observation: **sid + cid**

Field Name	Values or Data Type	Definition
cid	numeric	Course identifier.
sid	numeric	Student identifier unique to each student. This identification number is typically assigned to students upon enrollment in your agency. State agencies may have different identification numbers than district agencies for the same student.
final_grade_mark	string	The final grade or mark the student received in the class ("final" means last, cumulative grade assigned). Grades can range from "Alpha Plus" (A+) through F.
final_grade_mark_numeric	numeric	The final grade or mark the student received in the class ("final" means last, cumulative grade assigned). Grades can range from 0 to 5.
school_year	spring calendar year	Academic school year from fall to spring, denoted here as the spring calendar year.
tid	numeric	Unique staff or teacher identifier. State agencies may have different identification numbers than district agencies for the same staff/teacher.
school_code	numeric	Local identifier of schools.
math	0 = not math 1 = math	Indicates if the course is a math course.
ela	0 = not ELA 1 = ELA	Indicates if the course is an ELA course.
core	0 = not core 1 = core	Indicates if the course is a core course.
section_code	numeric	Section code for the course.
section_code_desc	string	Description of the section.
course_code	numeric	Course number.

// 13. Generate a variable that assigns the highest value of `temp _ cert _ last` (the only non-missing value) to all observations for a teacher. Call this variable `cert _ last`.

```
egen ...
```

// 14. Replace `certificaton _ path` with `cert _ last` if `cert _ mode` is missing (a teacher has multiple modes for `certification _ path`).

```
replace ...
```

// 15. Drop the temporary variables you created.

```
drop ...
```

/* Step 3. Create one consistent value for race_ethnicity for each teacher across years. This process is similar to that of gender and certification_pathway, so data snapshots are not provided in this section. ***/**

To be consistent with federal guidelines, teachers who have more than one value for `race _ ethnicity` will be reported as multiracial, unless one of the values for their `race _ ethnicity` is Hispanic, in which case they will be reported Hispanic.

Some districts allow the indication of multiple categories simultaneously, so you will first create consistent `race _ ethnicity` values by school year for each teacher based on the guidelines above, and then create a consistent value for each teacher across years.

// 1. Tabulate the `race _ ethnicity` variable to see its values and check if any are missing.

```
tab race_ethnicity, mi
```

// 2. Create a numeric variable that has consistent values for each race/ethnicity. Use a `for` loop to standardize values for African American and Hispanic, which have several different spelling variations.

```
gen race_num = .
foreach afam in "AFAM" "AFRICAN AMERICAN" "BLACK" "afam" "african
american" "black" {
      replace race_num = 1 if race_ethnicity=="`afam'"
}
replace race_num = 2 if race_ethnicity=="ASIAN" | race_
ethnicity=="asian"
foreach hisp in "HISP" "HISPANIC" "hisp" "hispanic" {
      replace race_num = 3 if race_ethnicity=="`hisp'"
}
replace race_num = 4 if race_ethnicity=="NATIVE AMERICAN" | race_
ethnicity=="native american"
replace race_num = 5 if race_ethnicity=="WHITE" | race_
ethnicity=="white"
replace race_num = 6 if race_ethnicity=="6"
```

// 3. Destring the `race _ num` variable.

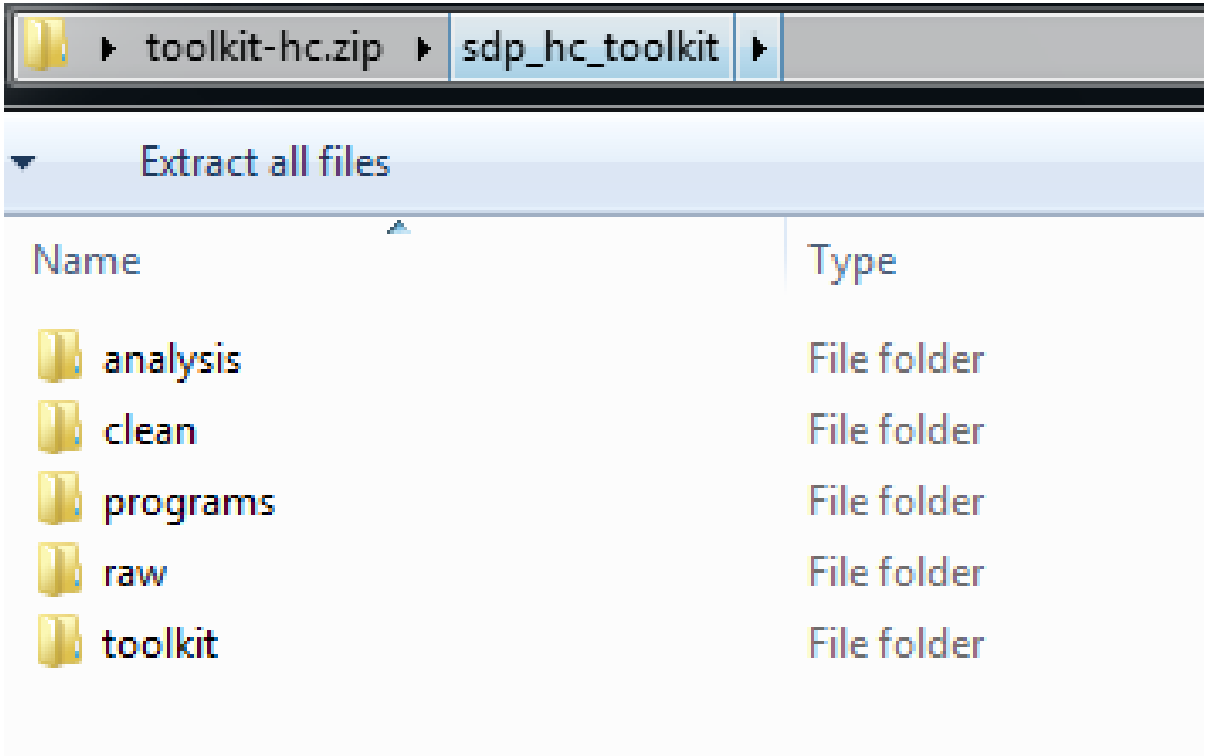
```
destring race_num, replace
```

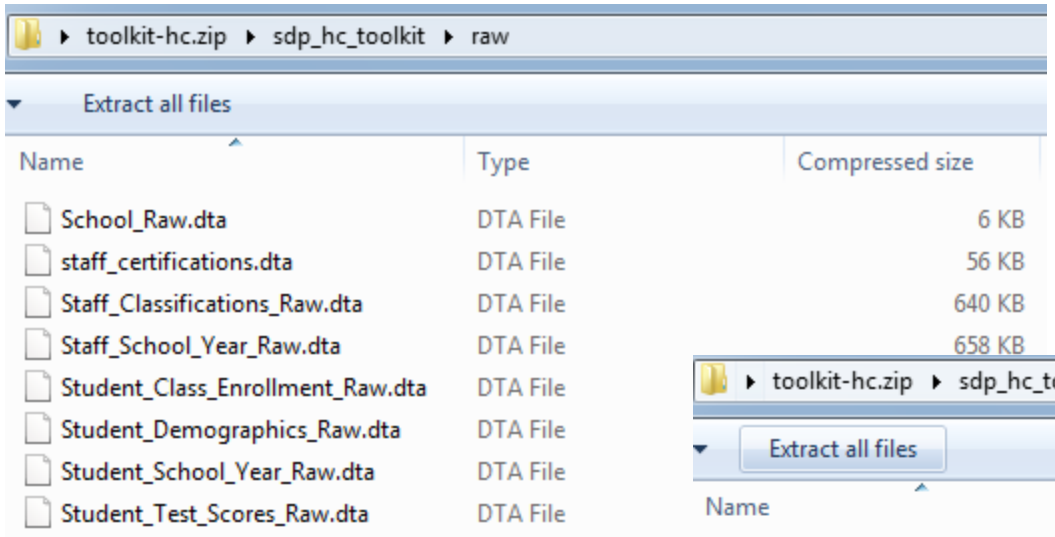
// 4. Label the values of `race _ num`.

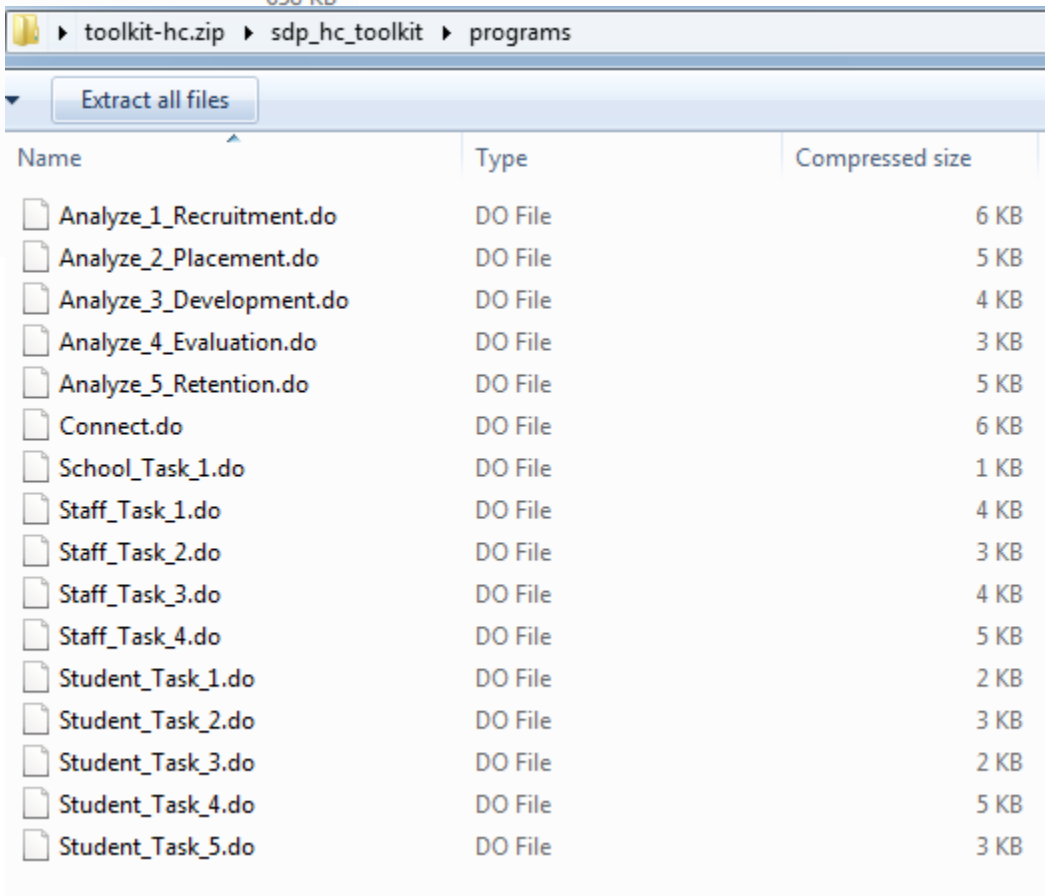
```
label define race 1 "Black" 2 "Asian" 3 "Hispanic" 4 "Native
American" 5 "White" 6 "Multiple/Other"
label values race_num race
```

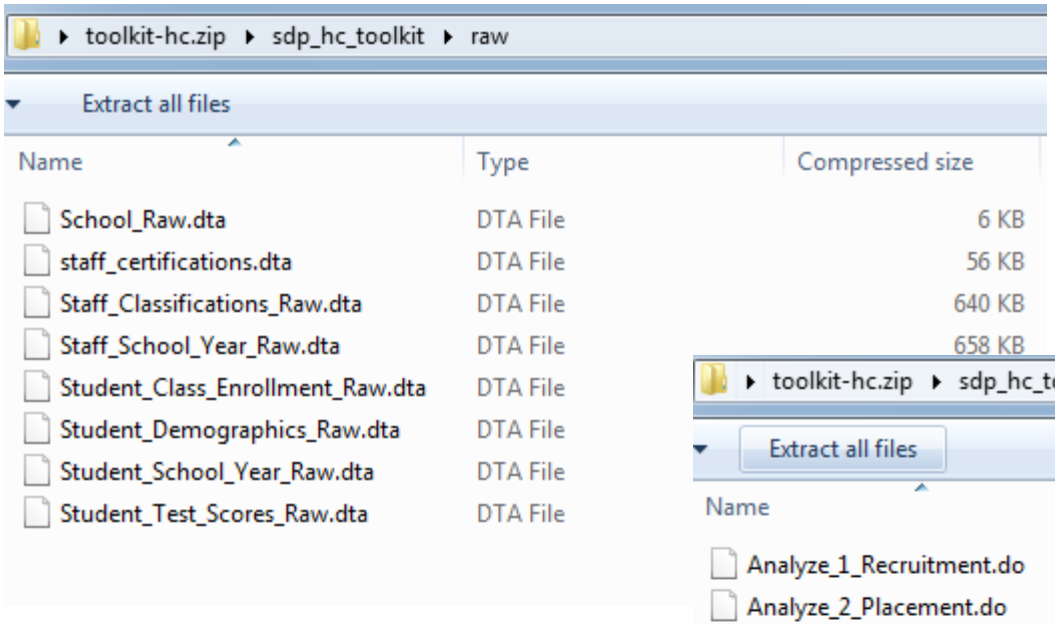
Download the entire toolkit with template files, synthetic data, and folder structure as a single zip file RECOMMENDED:

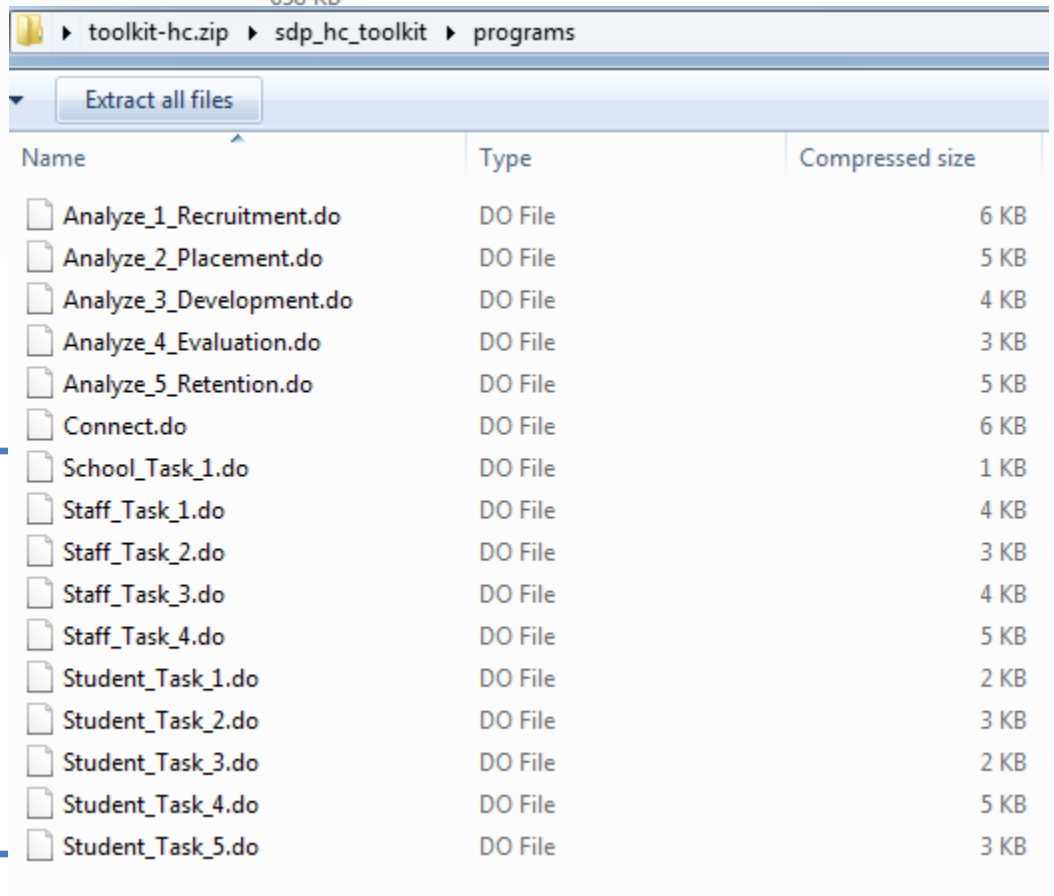
DOWNLOAD




toolkit-hc.zip ▸ sdp_hc_toolkit ▸ raw
Extract all files
Name
Type
Compressed size
School_Raw.dta
DTA File
6 KB
staff_certifications.dta
DTA File
56 KB
Staff_Classifications_Raw.dta
DTA File
640 KB
Staff_School_Year_Raw.dta
DTA File
658 KB
Student_Class_Enrollment_Raw.dta
DTA File
Student_Demographics_Raw.dta
DTA File
Student_School_Year_Raw.dta
DTA File
Student_Test_Scores_Raw.dta
DTA File


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ programs
Extract all files
Name
Type
Compressed size
Analyze_1_Recruitment.do
DO File
6 KB
Analyze_2_Placement.do
DO File
5 KB
Analyze_3_Development.do
DO File
4 KB
Analyze_4_Evaluation.do
DO File
3 KB
Analyze_5_Retention.do
DO File
5 KB
Connect.do
DO File
6 KB
School_Task_1.do
DO File
1 KB
Staff_Task_1.do
DO File
4 KB
Staff_Task_2.do
DO File
3 KB
Staff_Task_3.do
DO File
4 KB
Staff_Task_4.do
DO File
5 KB
Student_Task_1.do
DO File
2 KB
Student_Task_2.do
DO File
3 KB
Student_Task_3.do
DO File
2 KB
Student_Task_4.do
DO File
5 KB
Student_Task_5.do
DO File
3 KB


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ raw
Extract all files
Name
Type
Compressed size
School_Raw.dta
DTA File
6 KB
staff_certifications.dta
DTA File
56 KB
Staff_Classifications_Raw.dta
DTA File
640 KB
Staff_School_Year_Raw.dta
DTA File
658 KB
Student_Class_Enrollment_Raw.dta
DTA File
Student_Demographics_Raw.dta
DTA File
Student_School_Year_Raw.dta
DTA File
Student_Test_Scores_Raw.dta
DTA File


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ programs
Extract all files
Name
Type
Compressed size
Analyze_1_Recruitment.do
DO File
6 KB
Analyze_2_Placement.do
DO File
5 KB
Analyze_3_Development.do
DO File
4 KB
Analyze_4_Evaluation.do
DO File
3 KB
Analyze_5_Retention.do
DO File
5 KB
Connect.do
DO File
6 KB
School_Task_1.do
DO File
1 KB
Staff_Task_1.do
DO File
4 KB
Staff_Task_2.do
DO File
3 KB
Staff_Task_3.do
DO File
4 KB
Staff_Task_4.do
DO File
5 KB
Student_Task_1.do
DO File
2 KB
Student_Task_2.do
DO File
3 KB
Student_Task_3.do
DO File
2 KB
Student_Task_4.do
DO File
5 KB
Student_Task_5.do
DO File
3 KB

For Clean

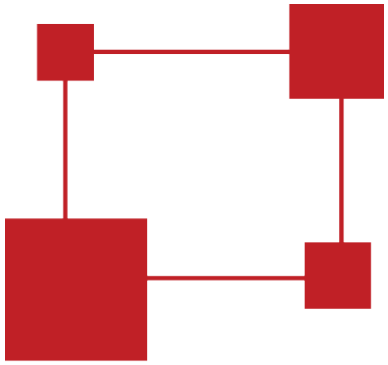
2. Clean

Help you to:

- Understand the importance of decision rules when defining data elements
- Identify data inconsistencies, including duplicate and missing records
- Develop a set of robust datasets that are accurate and reliable

Upon completing, you will:

- Have a set of datasets that include clean, reliable data to take you into the next step of the toolkit
- Have established (or begun to establish) decision rules and shared them with others who will use the same data elements to conduct other analyses



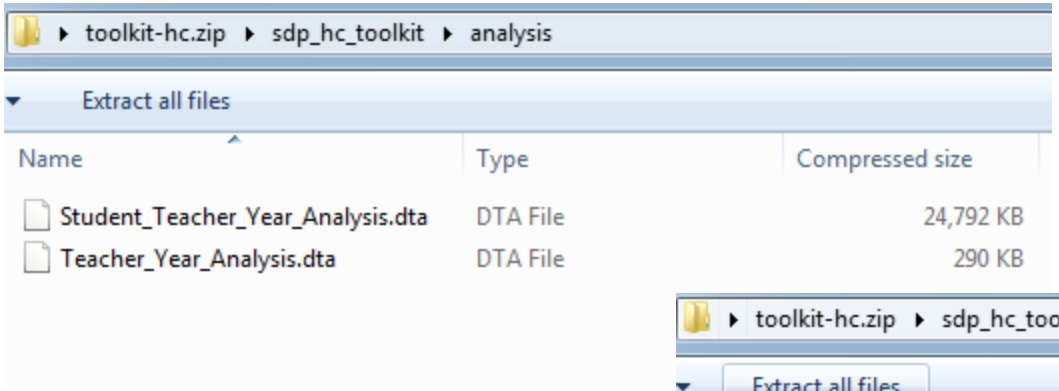
3. Connect

Data Linking Guide

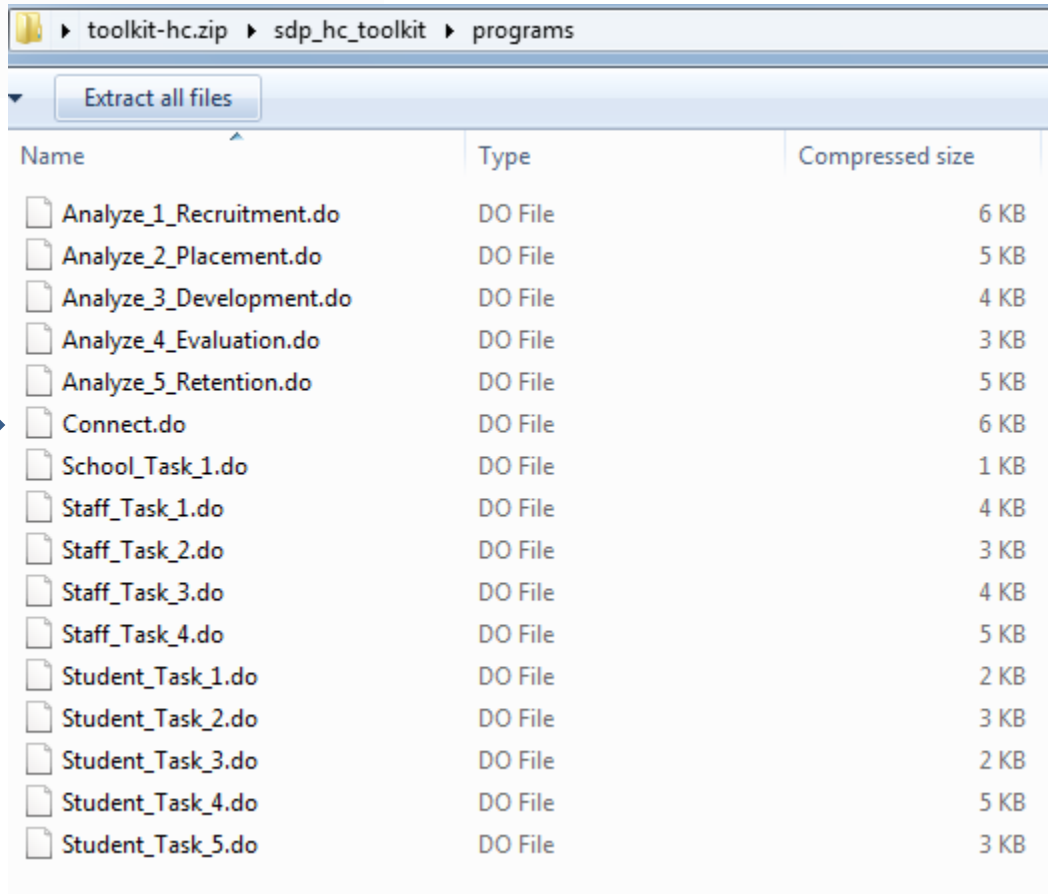
Now that you have collected and cleaned your data, you will want to merge the files together to create an analysis file.

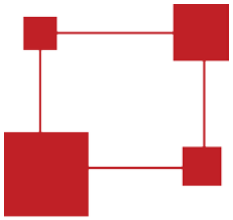
Student_Teacher_Year_Analysis			
sid	tre_math	qrt_ela	_CLmean_s_fulllunch_math
school_year	current_tre_math	qrt_ela_tm1	_CLmean_s_reducedlunch_math
grade_level	current_tre_ela	qrt_math	_CLmean_s_freelunch_math
cid_ela	curr2year_tre_math	qrt_math_tm1	_CLmean_s_misslunch_math
cid_math	curr2year_tre_ela	language_version_ela	_CLstd_scaled_score_math_tm1
school_code	tre_ela_2008	language_version_math	_CLstd_scaled_score_math_tm1_sd
school_name	tre_ela_2009	language_version_ela_tm1	_CLnumber_students_math
school_poverty_quartile	tre_ela_2010	language_version_math_tm1	_CLpct_missing_std_math_tm1
elementary	tre_ela_2011	_CLmean_s_ever_iep_ela	_COmean_s_ever_iep
middle	tre_math_2008	_CLmean_s_ever_gifted_ela	_COmean_s_ever_gifted
high	tre_math_2009	_CLmean_s_ell_ela	_COmean_s_ell
alternative	tre_math_2010	_CLmean_s_retained_ela	_COmean_s_retained
school_lvl	tre_math_2011	_CLmean_s_move_nonstruct_ela	_COmean_s_move_nonstruct
tid_ela	tre_ela_2008and2009	_CLmean_s_move_struct_ela	_COmean_s_move_struct
tid_math	tre_ela_2009and2010	_CLmean_s_absence_high_ela	_COmean_s_absence_high
t_birth_date	tre_ela_2010and2011	_CLmean_s_absence_miss_ela	_COmean_s_absence_miss
t_race_ethnicity	tre_math_2008and2009	_CLmean_s_male_ela	_COmean_s_male
t_afam	tre_math_2009and2010	_CLmean_s_afam_ela	_COmean_s_afam
t_asian	tre_math_2010and2011	_CLmean_s_asian_ela	_COmean_s_asian
t_hisp	s_iep	_CLmean_s_hisp_ela	_COmean_s_hisp
t_naam	s_gifted	_CLmean_s_naam_ela	_COmean_s_naam
t_white	s_ell	_CLmean_s_white_ela	_COmean_s_white
t_mult	s_retained	_CLmean_s_mult_ela	_COmean_s_mult
t_racemiss	s_move_nonstruct	_CLmean_s_racemiss_ela	_COmean_s_racemiss
t_is_teacher	s_move_struct	_CLmean_s_fulllunch_ela	_COmean_s_fulllunch
t_job_code	s_absence_high	_CLmean_s_reducedlunch_ela	_COmean_s_reducedlunch
t_hire_date	s_absence_miss	_CLmean_s_freelunch_ela	_COmean_s_freelunch
t_termination_date	s_frpl	_CLmean_s_misslunch_ela	_COmean_s_misslunch
t_experience	s_male	_CLstd_scaled_score_ela_tm1	_COstd_scaled_score_ela_tm1

p.5 of Data Linking Guide *(partial view)*


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ analysis
Extract all files
Name
Type
Compressed size
Student_Teacher_Year_Analysis.dta
DTA File
24,792 KB
Teacher_Year_Analysis.dta
DTA File
290 KB

For Connect


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ programs
Extract all files
Name
Type
Compressed size
Analyze_1_Recruitment.do
DO File
6 KB
Analyze_2_Placement.do
DO File
5 KB
Analyze_3_Development.do
DO File
4 KB
Analyze_4_Evaluation.do
DO File
3 KB
Analyze_5_Retention.do
DO File
5 KB
Connect.do
DO File
6 KB
School_Task_1.do
DO File
1 KB
Staff_Task_1.do
DO File
4 KB
Staff_Task_2.do
DO File
3 KB
Staff_Task_3.do
DO File
4 KB
Staff_Task_4.do
DO File
5 KB
Student_Task_1.do
DO File
2 KB
Student_Task_2.do
DO File
3 KB
Student_Task_3.do
DO File
2 KB
Student_Task_4.do
DO File
5 KB
Student_Task_5.do
DO File
3 KB



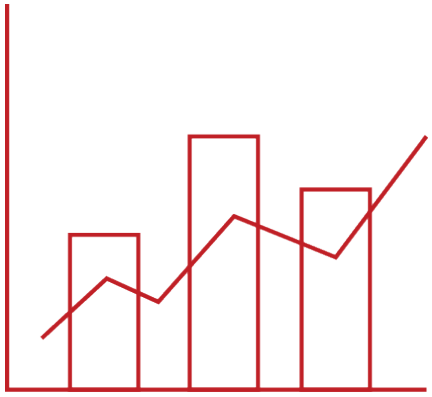
3. Connect

Help you to:

- Merge your cleaned data files
- Connect complex datasets through a unique student and teacher identifier

Upon completing, you will:

- Have a complete, robust, and comprehensive dataset that will allow you to conduct analyses in Analyze.
- Have a dataset you can share with other analysts in your department or agency



4. Analyze

Human Capital Analysis Guide

The final step in the SDP Toolkit for Effective Data Use is to analyze the data you have identified, cleaned, and connected.

Illustrative Guiding Questions

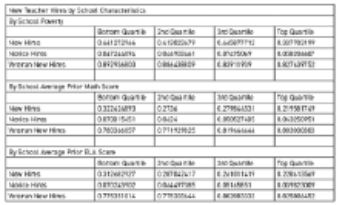
- **Recruitment:** What share of the agency's teaching force is made up of novice teachers?
- **Placement:** In which schools do novice teachers teach? Which students are taught by inexperienced teachers?
- **Development:** How effective are novice teachers? How does teacher effectiveness change with experience, and how does this relate to teacher salary scales?
- **Evaluation:** To what extent does prior teacher effectiveness predict future effectiveness? How do measures of effectiveness relate to teacher tenure decisions?
- **Retention:** For how long do novice teachers remain in the same school or district? How does teacher retention relate to teacher effectiveness?

Map of Analyses

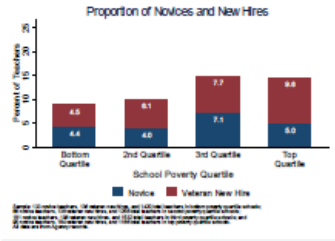
p.7-8 in Human Capital Analysis Section

A. Recruitment

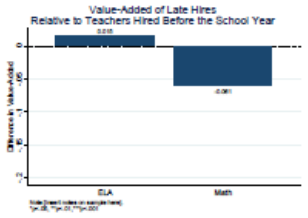
An examination of the kinds of teachers the agency hires, how many teachers it hires, what their preparation for teaching has been, and when they are hired.



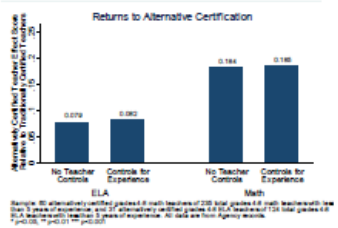
1. Table of Descriptive Information on Key Recruitment Practices (p. 10)



2. New Hires by Poverty Quartile (p. 13)



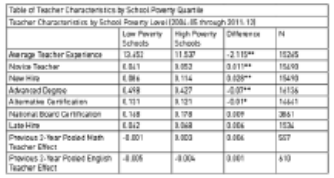
3. Value-added of Late Hires Relative to Teachers Hired Before the School Year Begins (p. 16)



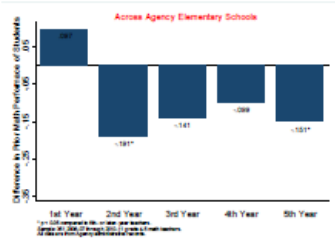
4. Value-added by Certification Pathway (p. 21)

B. Placement

An examination of the patterns in student assignment to teachers across and within schools to identify places where efforts to reform placement policies could positively impact students and teachers.

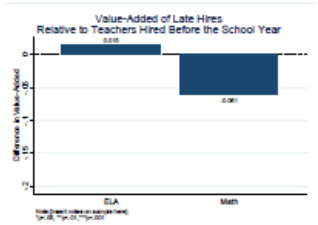


1. Table of Teacher Characteristics by School Poverty Quartile (p. 26)



2. Prior Achievement of Students Placed with Teachers by Teacher Experience (p. 29)

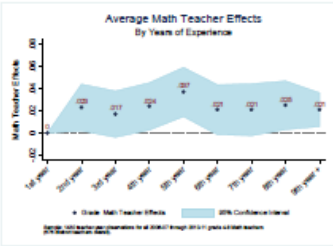
Strategic Performance Indicator



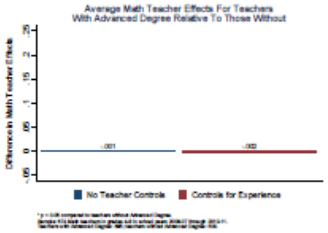
3. Prior Achievement of Students Placed with Teachers by Prior Teacher Effect Quartile (p. 34)

C. Development

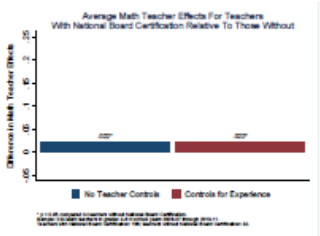
An examination of the ways teachers develop during their careers and an exploration of whether agency incentives are aligned with gains in teacher effectiveness.



1. Returns to Teaching Experience (p. 40)



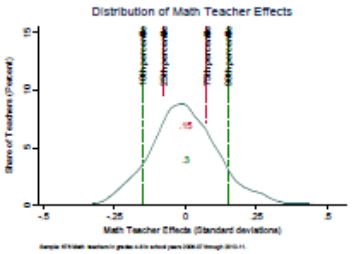
2. Returns to Advanced Degrees (p. 44)



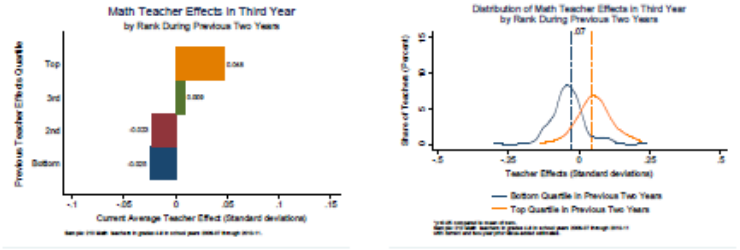
3. Returns to National Board Certification (p. 48)

D. Evaluation

A good measure of teacher effectiveness will be spread out enough to distinguish exemplary teachers from developing ones in addition to being well correlated over time. The Evaluation section of the diagnostic examines the extent to which value-added estimates meet these criteria.



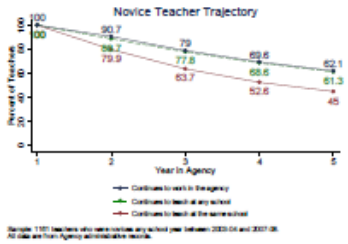
1. Distribution of Teachers by Value-Added Teacher Effect Estimates (p. 51)



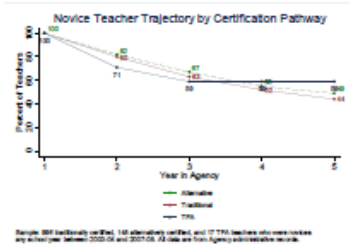
2. Predictive Power of Value-Added in Future Years Based on Prior Effectiveness Estimates (p. 54)

E. Retention

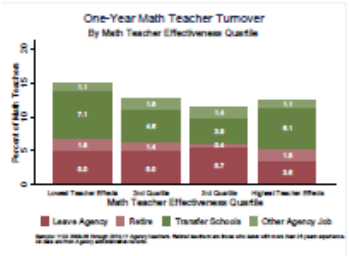
An examination of the types of teachers who transfer schools within the system, take nonteaching positions, and leave teaching in the agency altogether. This section examines how patterns vary across school characteristics, and among teachers with different teacher effectiveness estimates.



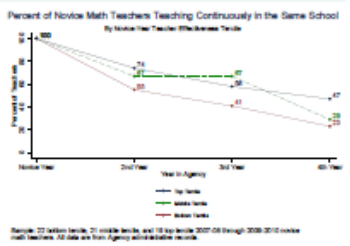
1. Four-Year Trajectory of Novice Teachers
(p. 60)



2. Four-Year Trajectory of Novice Teachers by Certification Pathway
(p. 64)

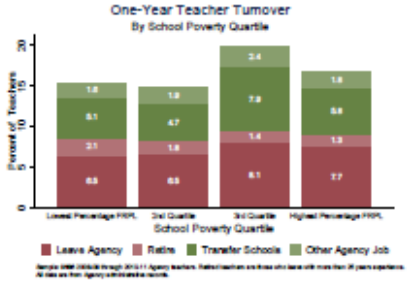


3. Retention by Teacher Effect Quartile
(p. 68)



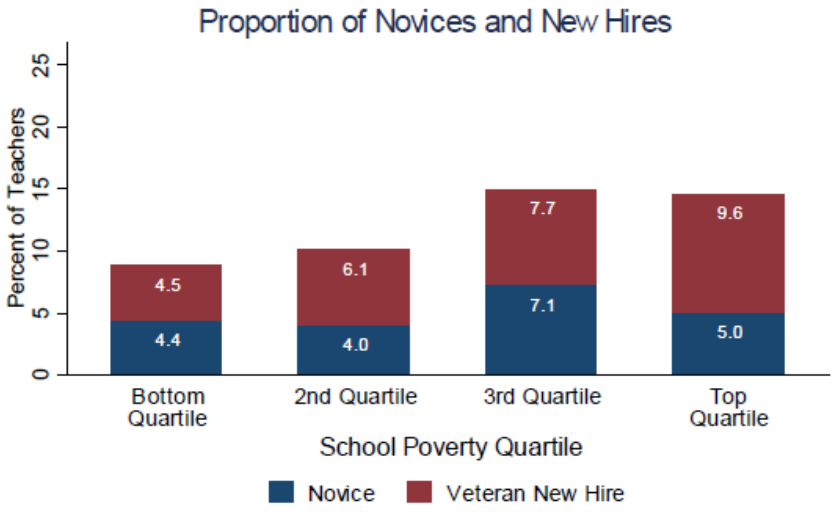
4. Four-Year Trajectory of Novice Teachers by Teacher Effectiveness Tercile
(p. 71)

Strategic Performance Indicator



5. Retention by School Poverty Quartile
(p. 75)

2. NEW HIRES BY SCHOOL POVERTY QUARTILE



Sample: 133 novice teachers, 136 veteran new hires, and 1420 total teachers in bottom poverty quartile schools; 88 novice teachers, 136 veteran new hires, and 1258 total teachers in second poverty quartile schools; 181 novice teachers, 196 veteran new hires, and 1532 total teachers in third poverty quartile schools; and 95 novice teachers, 184 veteran new hires, and 1186 total teachers in top poverty quartile schools. All data are from Agency records.

Purpose:
Examine the extent to which new hires are distributed unevenly across the agency according to school characteristics.

Required analysis file variables:

```
sch_pov_qrt
t_novice
t_veteran_newhire
t_is_teacher
```

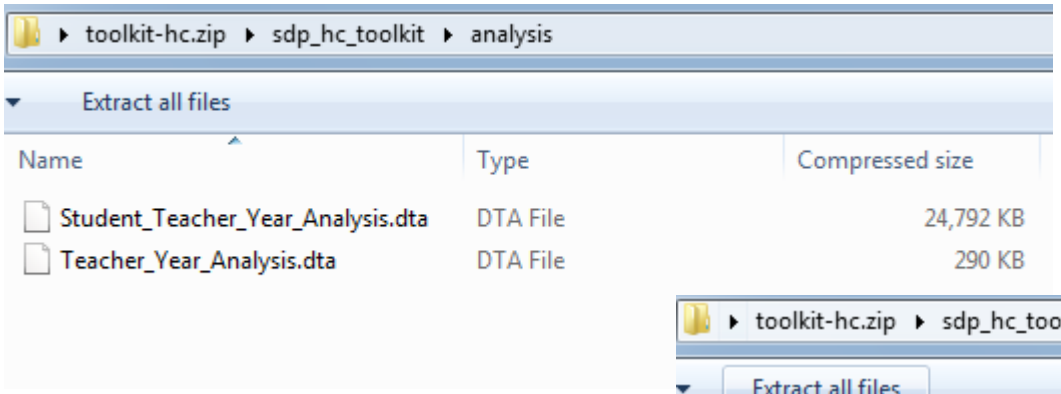
Analysis-specific sample restrictions:

- Keep only employees whose job code is "teacher."

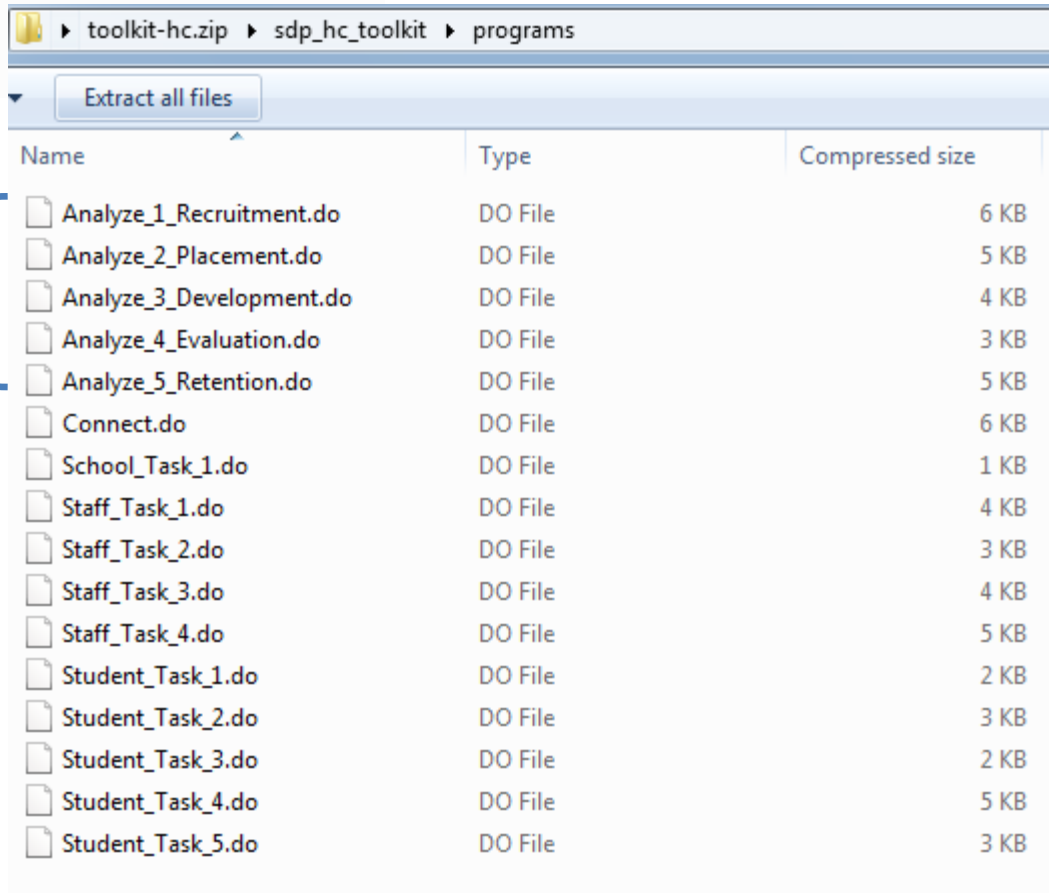
Ask yourself:

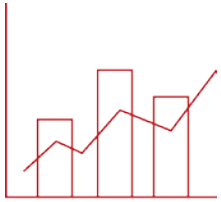
- How do hiring patterns differ between high- and low-poverty schools?
- Are the shares of novice and veteran new hires distributed equitably and strategically across school poverty quartiles?

Potential further analyses:
This graph is easily replicable to explore how the distribution of new hires varies across other school characteristics (e.g., AYP status, zone, school level, etc.).


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ analysis
Extract all files
Name
Type
Compressed size
Student_Teacher_Year_Analysis.dta
DTA File
24,792 KB
Teacher_Year_Analysis.dta
DTA File
290 KB

For Analyze


toolkit-hc.zip ▸ sdp_hc_toolkit ▸ programs
Extract all files
Name
Type
Compressed size
Analyze_1_Recruitment.do
DO File
6 KB
Analyze_2_Placement.do
DO File
5 KB
Analyze_3_Development.do
DO File
4 KB
Analyze_4_Evaluation.do
DO File
3 KB
Analyze_5_Retention.do
DO File
5 KB
Connect.do
DO File
6 KB
School_Task_1.do
DO File
1 KB
Staff_Task_1.do
DO File
4 KB
Staff_Task_2.do
DO File
3 KB
Staff_Task_3.do
DO File
4 KB
Staff_Task_4.do
DO File
5 KB
Student_Task_1.do
DO File
2 KB
Student_Task_2.do
DO File
3 KB
Student_Task_3.do
DO File
2 KB
Student_Task_4.do
DO File
5 KB
Student_Task_5.do
DO File
3 KB



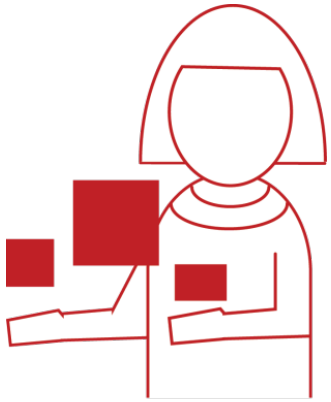
4. Analyze

Help you to:

- Take the robust dataset you've created and jump into the world of analytics
- Replicate the SDP Human Capital Diagnostic
- Identify trends, across your agency and across schools, in teacher recruitment, placement, effectiveness, and retention.

Upon completing, you will:

- Understand teacher placement (within and across your agency) by effectiveness
- Understand teacher recruitment patterns in your agency
- Have a set of analyses (graphs) that you can share with senior leadership
- Have produced 2 human capital Strategic Performance Indicators (SPIs) for your agency
- Have earned an SDP badge for toolkit completion



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.

NAMING CONVENTIONS

File Naming

Files should be named using "compositional identifiers" that allow an individual to understand the contents of a file at a high level without having to open the file. This is especially important for program files (e.g. Stata .do files or SQL scripts) and graphs. The Compositional identifiers file name should descend in order of importance so the files group together in an intuitive order when sorted by name (default in Windows Explorer). For example, files should be named with the following compositional identifiers in the following order of importance:

- Project Name,
- Component of process - for instance student demographic data or survey data,
- Date (in YYYYMMDD format), version number, or state of file (i.e. temp, test, review, final, etc)

So, for example:

- School_Report_Student_Attributes_20110601_DRAFT.do
- School_Report_Student_Attributes_20110601_REVIEW.do
- School_Report_Student_Attributes_20110601_FINAL.do

Even though folder structure may imply the contents of a file and the above guidelines may seem redundant, files can be shared across departments in your organization or with other external entities and therefore names should convey the same meaning outside of folder structure.

Additionally, file names should be as consistent as possible, especially output files (graphs, logs) related to a program file. For example, a graph output of the above Stata .do file may be named School_Report_Student_Attributes_20110601_FINAL_ethnicity.gph.

Variable Naming

The number of characters used to name variables is limited. For example, Stata variable names may contain up to 32 characters. Database columns may be limited to 30 characters depending on platform. Additionally, many Stata commands only print 12 characters by default. Keep this in mind when you name variables. Try to be both specific and concise in your variable names.

All variables in a Stata dataset should be labeled (as should database columns). For commonly used variables with existing definitions, consider reading in labels from a common external file rather than entering labels manually or by copy/paste. Alternatively, call upon a separate .do file in your main code that contains standard labels rather than including labeling code in your primary .do file.

```
// label variables in standard student file
do "$programs/dcps_student_labeling.do"

// label variables not in standard labeling do file
label stu_struc_move "student had structural move, moving schools"
```

COMMENTING AND READABILITY

The following is an example of well indented code:

```
if $teacher == 1 {

    local numyrs = 4

    // define empty matrix of Yr x Subj
    mat out = J(`numyrs',2,.)
    local row = 1
    local col = 1

    foreach subj in math read {
        use "$data/student_teacher_`subj'_vam.dta", clear

        forval yr = 2(1)`numyrs' {
            gen late_exp_`yr' = ever_late_hire*t_exp`yr'
        }
    } // end of loop on subject
} // end of teacher processing
```

The following is an example of poorly indented code:

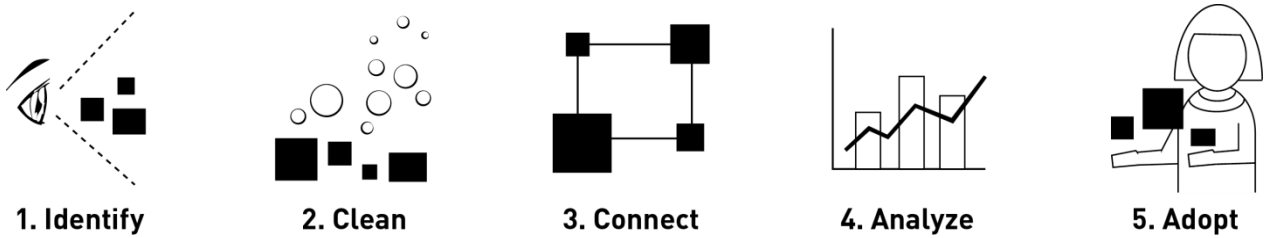
```
if $teacher == 1 {
local numyrs = 4
// define empty matrix of Yr x Subj
mat out = J(`numyrs',2,.)
local row = 1
local col = 1

foreach subj in math read {
    use "$data/student_teacher_`subj'_vam.dta", clear
    forval yr = 2(1)`numyrs' {gen late_exp_`yr' = ever_late_hire*t_exp`yr'}
}
}
```

Do not indent braces following a condition.

```
 if x > 0
    {
    dis "x is positive"
    }
 else
    {
    dis "x is negative"
    }
```

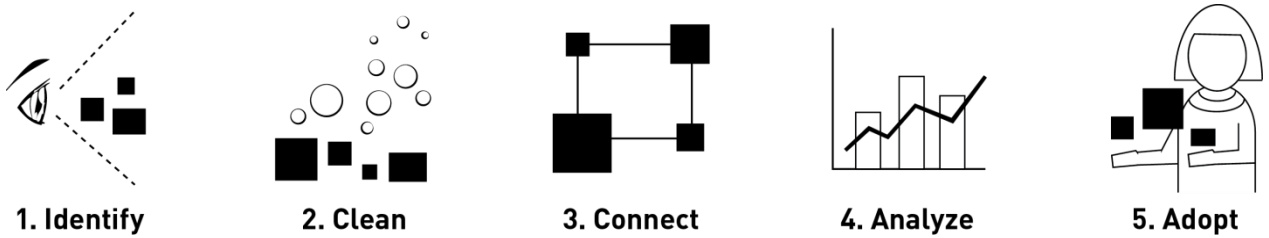
Also, do not double or triple indent when a single indent is sufficient.

```
 if x > 0 {
        dis "x is positive"
    }
else {
        dis "x is negative"
    }
```



What makes you a prime user for the Toolkit?

- Access to large datasets with information on student achievement
- Interest, in your agency, to understand and explore either high school completion and college enrollment trends, or teacher recruitment, placement, evaluation, and retention
- Knowledge of some statistical programming language

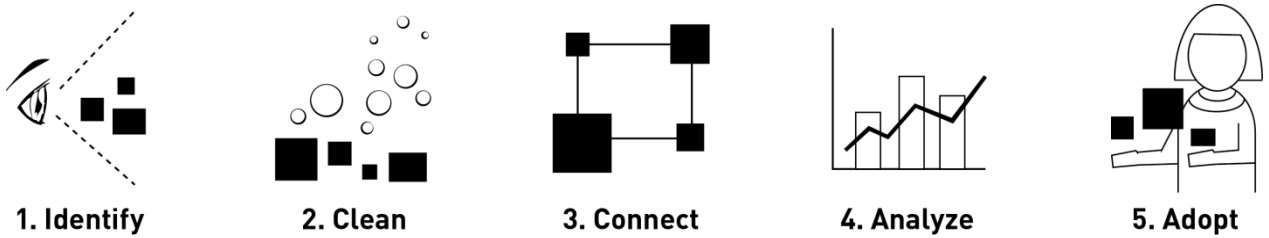


What makes you a prime user for the Toolkit?

- Access to large datasets with information on student achievement
- Interest, in your agency, to understand and explore either high school completion and college enrollment trends, or teacher recruitment, placement, evaluation, and retention
- Knowledge of some statistical programming language

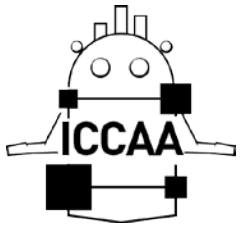
Beyond the analytic questions found in ANALYZE, the Toolkit will help you to answer structural questions like:

- Are we missing key data elements that will help us better understand teacher recruitment, placement, effectiveness, and retention?
- Is there a set of decision rules that all analysts in our agency understand and embrace?
- Do datasets from disparate warehouses easily match-up, or do we find vast inconsistencies in the way that data are collected in our agency?



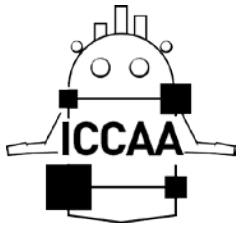
What to do if you're serious about digging into the Toolkit:

- Get into the mental mode and psych yourself out. This is a 300+ page document and no easy task. Are you up for it?
- Devote time in your calendar, perhaps 1-2 hours every other day to digging into the toolkit. This means that you are consciously putting all of your other demands on hold for a few hours a week!
- Find a toolkit completion buddy – either in your agency or across the SDP network.
- Decide on what program you will use – Stata, SPSS, R, perhaps Excel?
- Practice with our datasets first, then start to use your own; but, create duplicate datasets of all originals.
- Keep SDP in the loop – let us know where you run into bumps or when you've completed an analysis, particularly the SPIs!



Q&A

Thank You



Check www.gse.harvard.edu/sdp/toolkit for the most recent toolkit version.

Please contact us at sdp@gse.harvard.edu if you have any questions about the toolkit.