



STEP BY STEP GUIDE

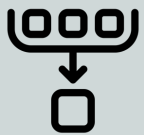
SDP PDP Extension

Using Your PDP Data to Analyze Student Success

What is the SDP PDP Extension?



An extension of the **SDP CTE Diagnostic** using the **Postsecondary Data Partnership (PDP) Analysis Ready (AR) Files**



It **simplifies the data preparation process** required for the CTE Diagnostic by leveraging the standardization of the AR files



It **adds new dimensions to the analysis** included in the CTE Diagnostic by looking at student outcomes specific to the AR files

What's included in this Extension?



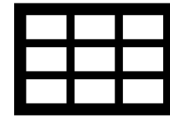
Detailed Documentation

A README document and this presentation with step by step explanations on how to find and use the PDP AR files, and how to transform them to generate the CTE diagnostic



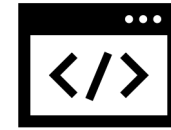
Code Scripts to Transform Data

Stata .do files for analysts to either run directly or to translate to their programming language of choice, which transform the AR files into relevant data tables for the CTE diagnostic



Data Entry Templates

Data table templates to enter and label pathway data for students, which is the only piece of student information used in the CTE Diagnostic that is not already included in the AR files



Code Scripts to Analyze Data

Stata .do files for analysts to either run directly or to translate to their programming language of choice, which use the transformed datasets and generate graphs and analyses



Example Report

An example report (generated in Markdown and pdf formats) with data visualization and analyses generated with the code files included in this extension

Contents of this guide

Part I. Reminder : The CTE Diagnostic

Part II. The Postsecondary Data Partnership (PDP) and its Analysis Ready (AR) Files

Part III. Getting Started with this Tool

Reminder: The CTE Diagnostic

Part I

What is the CTE Diagnostic?



Goal: To provide both strategic insights as well as guide inquiry to support decision-making by institutional leadership regarding their CTE programming and how to better support students' progression through their chosen pathway



Four sections, ordered to identify challenges and then dig deeper into potential explanations to point institutions to areas of focus and evaluation



Descriptive analyses that can be conducted using institutional or state data

CTE Diagnostic- Key Guiding Questions



Are pathways supporting students to complete or transfer?



What enrollment choices do students make over time? Might churn between pathways slow students' progress?



Do certain courses impede student progress towards completion/transfer?



Are students able to gain momentum in accumulating credits? If students aren't completing, is it an issue of not attempting or not attaining credits?

What's included in the Diagnostic?



Framework for Exploration

Research questions, relevant research, and suggestions for how to use analyses to drive practice



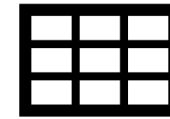
Example Visuals

Data visualization mock-ups with interpretation keys



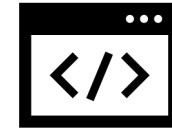
Analysis Summaries

High-level description of analyses, including purpose and analytic approach



Data Specifications

Detailed descriptions of the data needed for the analyses, including key variables and data formatting needs



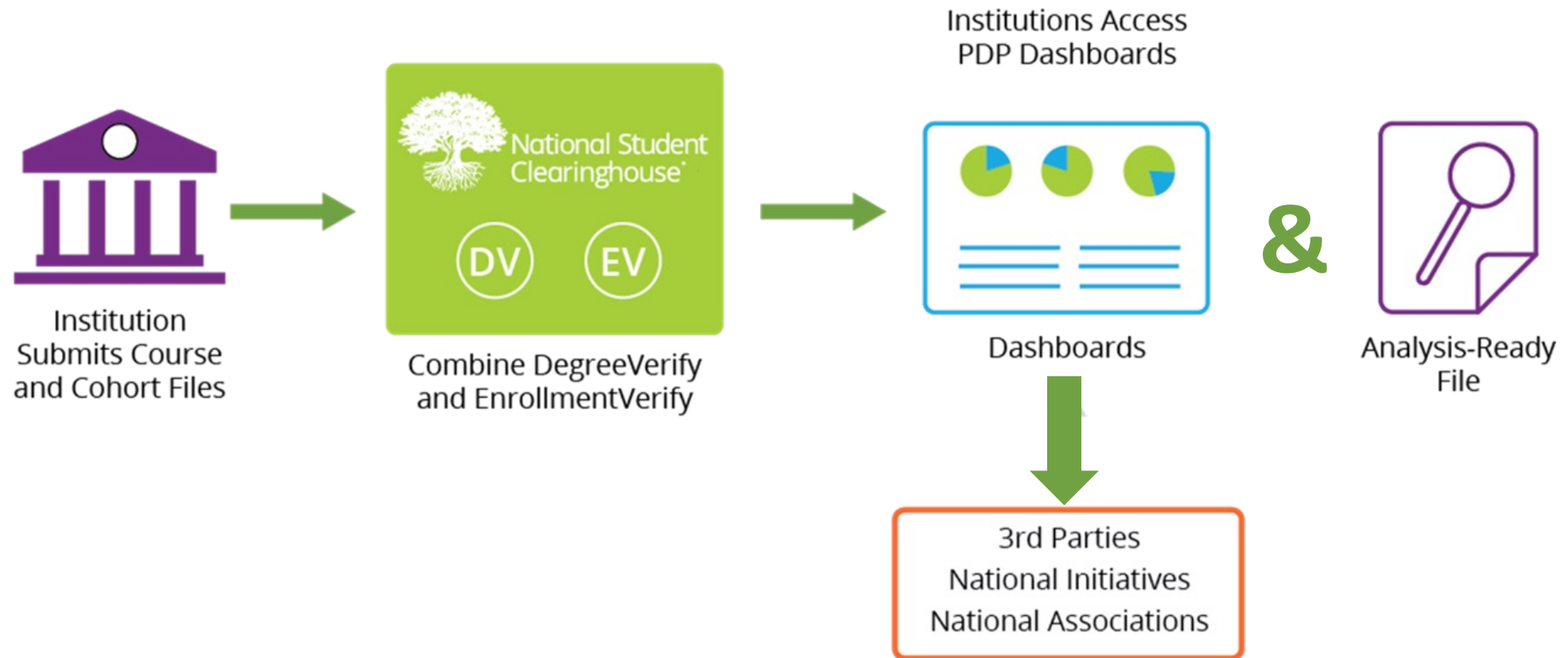
Code Files

Stata .do files for analysts to either run directly or to translate to their programming language of choice

The Postsecondary Data Partnership (PDP) and its Analysis Ready (AR) Files

Part II

What is the Postsecondary Data Partnership? (PDP)



What is the Analysis Ready File? (AR file)

- An excel file containing student level data
- The file is created from the course file, cohort file, and other NSC data sources
- There are three analysis ready files- cohort, course, and financial aid
- Cohort level file: One row per student with 50+ columns of data elements and calculated outcomes
- Course level file: One row per student per course with 40 columns of data elements and calculated outcomes
- Financial aid file: Optional, includes information to help calculate unmet need

Sample Cohort Analysis Ready File

First Name	Middle Name	Last Name	SSN	Date of Birth	Student ID	Institution ID	Cohort	Cohort Term	Student Age	Enrollment Type	Enrollment intensity first term
Hiroko	Randall	Lee	555-55-5555	19910524	555-55-5555	00675000	2015-16	Fall	Older than 24	First-time	Full-time
Hillary	Ocean	Ashley	555-55-5556	19920619	555-55-5556	00675000	2012-13	Winter	20 and younger	Transfer-in	Part-time
Anika	Jayne	Mays	555-55-5557	19930720	555-55-5557	00675000	2013-14	Spring	>20-24	First-time	Full-time
Branden	Wang	Michael	555-55-5558	20100819	555-55-5558	00675000	2016-17	Summer	Older than 24	Transfer-in	Part-time
Emma	Illiana	Strong	555-55-5559	20090615	555-55-5559	00675000	2017-18	Fall	20 and younger	First-time	Full-time
Mark	Gemma	Reilly	555-55-5560	20080612	555-55-5560	00675000	2014-15	Fall	>20-24	Transfer-in	Part-time
Brian	Hunter	Lee	555-55-5561	19910524	555-55-5561	00675000	2015-16	Fall	Older than 24	First-time	Full-time
Derek	Kyra	Coffey	555-55-5562	19920619	555-55-5562	00675000	2012-13	Winter	20 and younger	Transfer-in	Part-time
Nora	Sheila	Mccarthy	555-55-5563	19930720	555-55-5563	00675000	2013-14	Spring	Older than 24	First-time	Full-time
Kelly	Tucker	Flowers	555-55-5564	20100819	555-55-5564	00675000	2016-17	Summer	Older than 24	Transfer-in	Part-time
Cynthia	Luke	Kane	555-55-5565	20090615	555-55-5565	00675000	2017-18	Fall	Older than 24	First-time	Full-time
Quemby	Anastasia	Brady	555-55-5566	20080612	555-55-5566	00675000	2014-15	Fall	20 and younger	Transfer-in	Part-time
Chaim	Kadeem	Pace	555-55-5567	19910524	555-55-5567	00675000	2015-16	Fall	20 and younger	First-time	Full-time
Grady	Cameran	Evans	555-55-5568	19920619	555-55-5568	00675000	2012-13	Fall	Older than 24	Transfer-in	Part-time
Colin	Hedy	Santiago	555-55-5569	19930720	555-55-5569	00675000	2013-14	Fall	Older than 24	First-time	Full-time
Christen	Maris	Perry	555-55-5570	20100819	555-55-5570	00675000	2016-17	Fall	20 and younger	Transfer-in	Part-time
Brody	Kelly	Mayer	555-55-5571	20090615	555-55-5571	00675000	2017-18	Fall	20 and younger	First-time	Full-time
Alvin	Lilah	Emerson	555-55-5572	20080612	555-55-5572	00675000	2014-15	Fall	20 and younger	Transfer-in	Part-time
Pamela	Callie	Valencia	555-55-5573	19910524	555-55-5573	00675000	2015-16	Fall	>20-24	First-time	Full-time
Colorado	Sylvester	Joyner	555-55-5574	19920619	555-55-5574	00675000	2012-13	Fall	>20-24	Transfer-in	Part-time
Elmo	Henry	Marsh	555-55-5575	19930720	555-55-5575	00675000	2013-14	Fall	>20-24	First-time	Full-time
Sophia	Simon	Ellison	555-55-5576	20100819	555-55-5576	00675000	2016-17	Fall	Older than 24	Transfer-in	Part-time
Myra	Meredith	Dotson	555-55-5577	20090615	555-55-5577	00675000	2017-18	Fall	Older than 24	First-time	Full-time
Margaret	Xenos	Ramsey	555-55-5578	20080612	555-55-5578	00675000	2014-15	Fall	>20-24	Transfer-in	Part-time
Abdul	Len	Ball	555-55-5579	20100819	555-55-5579	00675000	2015-16	Fall	Older than 24	First-time	Full-time
India	Moses	Puckett	555-55-5580	20090615	555-55-5580	00675000	2012-13	Fall	Older than 24	Transfer-in	Part-time
Leonard	Stella	Conley	555-55-5581	20080612	555-55-5581	00675000	2013-14	Winter	20 and younger	First-time	Full-time
Clinton	Madeson	Bond	555-55-5582	19910524	555-55-5582	00675000	2016-17	Spring	20 and younger	Transfer-in	Part-time

MOCK DATA

What kind of elements & outcomes are in the file?

➤ Student Demographic & Academic Information

Name, DOB, Student ID/SSN, Cohort information, Race, Ethnicity, Gender, First Gen status, Pell Status First Year, Enrollment information

➤ Student Progression (“Early Momentum”)

GPA, number of attempted and completed credits, Gateway and dev ed information (placement, attempted/completed), retention/persistence

➤ Long Term Outcomes

Year earned bachelor, associates, or certificate (at your institution or another), year of last enrollment, time to credential, transfer information

How can I access my analysis ready files?

- Identify your PDP user administrator on campus
- Campus user administrator can scan the QR code or go to the link below for instructions on accessing the file
- If you are still having trouble email pdpSERVICE@nationalstudentclearinghouse.com
- <https://tinyurl.com/ARFILEPLEASE>



How can I use the analysis ready files?

- Create cohorts to help answer questions unique to your campus
- Identify opportunities for immediate intervention
- Analyze past trends
- Merge/append other datasets
- Link with your Tableau Dashboards for deeper insights
- **...with this new tool!**

Getting Started with this Tool

Part III

How does the PDP Extension work ?

1. **Download the PDP Extension folder** to your computer and save it locally
2. **Read the README file**
3. **Add your PDP Analysis Ready Files** to your local folder (in the subfolder 1_data-pdp)
4. **Add student pathway information** to your local folder (in the subfolder 2_data-toolkit)
5. **Define the parameters and options** for your analysis
6. **Run the data transformation .do files** to transform your data into the appropriate data frame
7. **Run the analytical .do files** to generate graphs, tables and analyses to put into a report

How to access the PDP Extension ?

- Open any web browser and go to https://github.com/Research-Ed/sdp_cte_pdp_extension
- This takes you to the **github repository** (or **repo**) where the **tool** is stored
- What is **github**?
 - Github is a website that makes code sharing easy, and where a lot of projects that use code (no matter the language : Stata, R, Python, JavaScript, etc.) are shared publicly
- What is a **repo**?
 - A repo or repository is the name used on github to call a project
- What is a **tool** or **toolkit**?
 - A tool or toolkit can take many forms, but in this case it takes the form of a series of code scripts that you can download and run with your own data to generate analyses of interest
 - For this toolkit, all the scripts are in Stata, and we refer to them as scripts or .do files interchangeably
 - The toolkit also contains a README document, which is a document that you have to read (like its name suggest) when you first use the toolkit, as well as an example Report document that showcases the type of analyses produced by this toolkit

What am I looking at ?

The screenshot shows the GitHub interface for the repository 'researched-pdp-toolkit' by user 'BeaLeydier'. The repository is public and has 4 issues, 0 pull requests, 0 actions, 0 projects, 0 wiki pages, 0 security alerts, 0 insights, and 0 settings. The repository is currently on the 'main' branch, which has 1 branch and 0 tags. The repository is pinned, unwatched by 1 user, forked by 0 users, and starred by 0 users.

The repository contains the following files and folders:

File/Folder	Description	Last Commit
0_scripts	Prototype section 1 (#10)	last month
1_data-pdp	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
2_data-toolkit	Prototype section 1 (#10)	last month
3_data-diagnostics	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
4_output	Prototype section 1 (#10)	last month
.gitignore	Label outputs (#2)	5 months ago
0.Set-Up.do	Initialize instructions	3 weeks ago
1.Define-Parameters.do	Prototype section 1 (#10)	last month
2.Enter-Data.do	Initialize instructions	3 weeks ago
3.Make-Data.do	Initialize instructions	3 weeks ago
4.Make-Report.do	Initialize instructions	3 weeks ago
README.md	Initialize instructions	3 weeks ago
ReportDemo.md	Prototype section 1 (#10)	last month

The repository is titled 'researched-pdp-toolkit' and is described as 'ResearchEd Post-Secondary Data Partnership Toolkit'. It has 0 stars, 1 watching, and 0 forks. The repository is currently on the 'main' branch, which has 1 branch and 0 tags. The repository is pinned, unwatched by 1 user, forked by 0 users, and starred by 0 users.

The repository is currently on the 'main' branch, which has 1 branch and 0 tags. The repository is pinned, unwatched by 1 user, forked by 0 users, and starred by 0 users.

The repository is currently on the 'main' branch, which has 1 branch and 0 tags. The repository is pinned, unwatched by 1 user, forked by 0 users, and starred by 0 users.

You have landed at the root of the repository, which contains files. Ignore everything else!

Bealeydiier / researched-pdp-toolkit

researched-pdp-toolkit Public

main 1 Branch 0 Tags

Go to file Add file <> Code

File	Description	Last Commit
0_scripts	Prototype section 1 (#10)	last month
1_data-pdp	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
2_data-toolkit	Prototype section 1 (#10)	last month
3_data-diagnostics	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
4_output	Prototype section 1 (#10)	last month
.gitignore	Label outputs (#2)	5 months ago
0.Set-Up.do	Initialize instructions	3 weeks ago
1.Define-Parameters.do	Prototype section 1 (#10)	last month
2.Enter-Data.do	Initialize instructions	3 weeks ago
3.Make-Data.do	Initialize instructions	3 weeks ago
4.Make-Report.do	Initialize instructions	3 weeks ago
README.md	Initialize instructions	3 weeks ago
ReportDemo.md	Prototype section 1 (#10)	last month

About

ResearchEd Post-Secondary Data Partnership Toolkit

Readme Activity 0 stars 1 watching 0 forks

Releases

No releases published
[Create a new release](#)

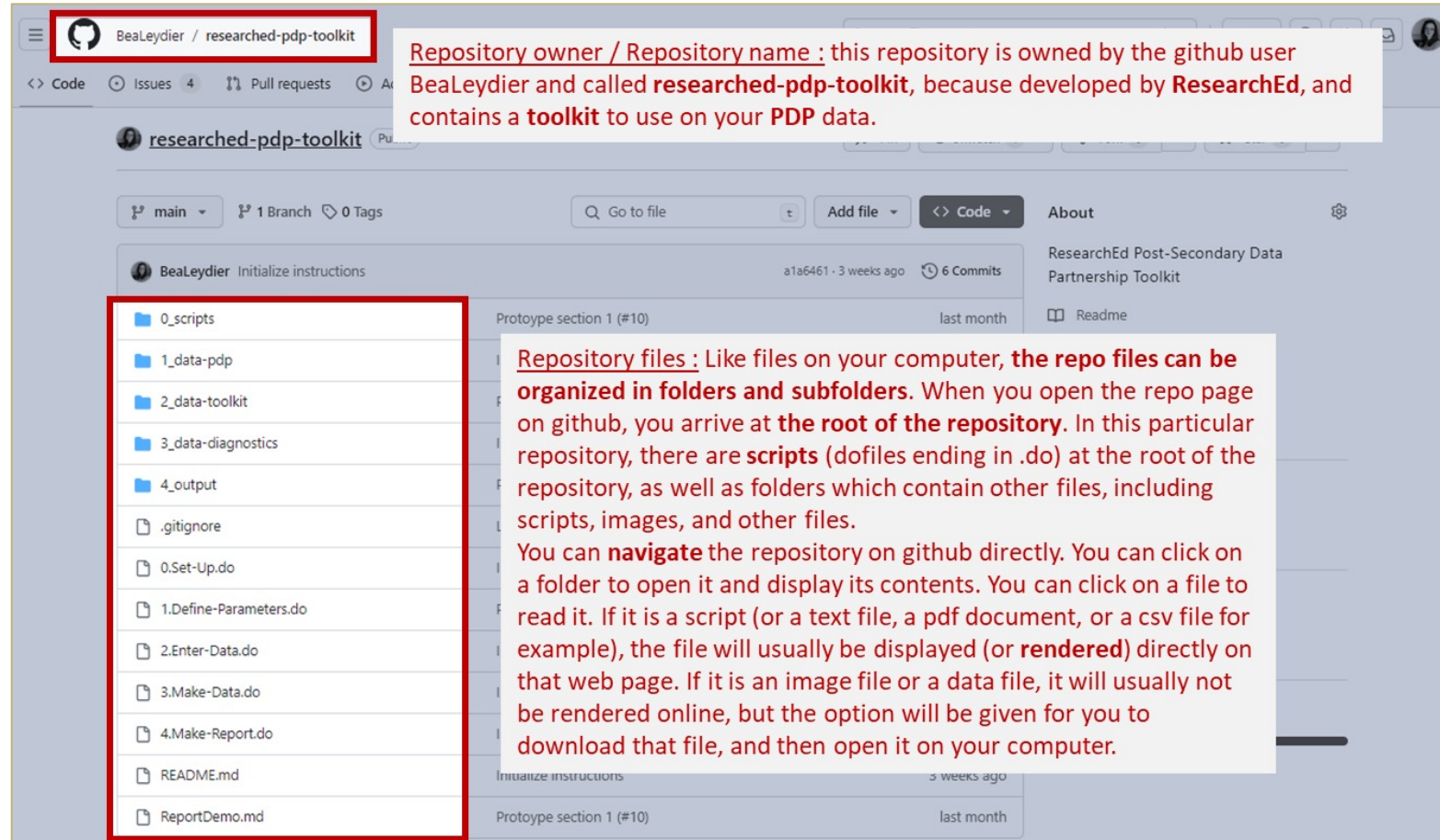
Packages

No packages published
[Publish your first package](#)

Languages

Stata 100.0%

You have landed at the root of the repository, which contains files. Ignore everything else!



Repository owner / Repository name : this repository is owned by the github user BealeyDier and called **researched-pdp-toolkit**, because developed by **ResearchEd**, and contains a **toolkit** to use on your **PDP** data.

Repository files : Like files on your computer, **the repo files can be organized in folders and subfolders**. When you open the repo page on github, you arrive at **the root of the repository**. In this particular repository, there are **scripts** (dofiles ending in .do) at the root of the repository, as well as folders which contain other files, including scripts, images, and other files.

You can **navigate** the repository on github directly. You can click on a folder to open it and display its contents. You can click on a file to read it. If it is a script (or a text file, a pdf document, or a csv file for example), the file will usually be displayed (or **rendered**) directly on that web page. If it is an image file or a data file, it will usually not be rendered online, but the option will be given for you to download that file, and then open it on your computer.

- 0_scripts
- 1_data-pdp
- 2_data-toolkit
- 3_data-diagnostics
- 4_output
- .gitignore
- 0.Set-Up.do
- 1.Define-Parameters.do
- 2.Enter-Data.do
- 3.Make-Data.do
- 4.Make-Report.do
- README.md
- ReportDemo.md

Ok, there are a few more things you can look at on the repo landing page.

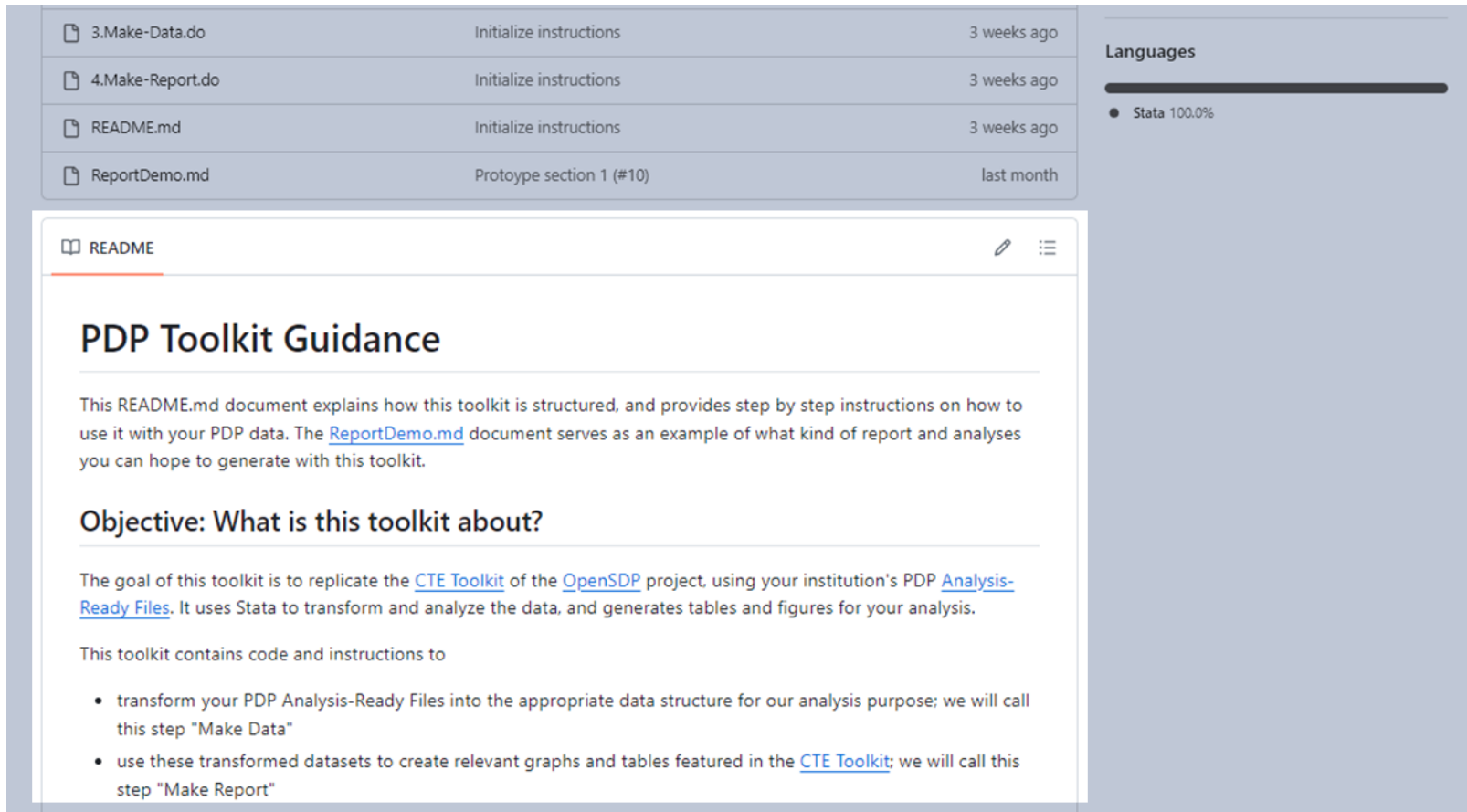
About: tells you what the repo is about! (unsurprisingly)

About
ResearchEd Post-Secondary Data Partnership Toolkit

Languages: tells you what languages are used in the repo, taking into account all the script files of the repo. This repository is 100% Stata.

Languages
Stata 100.0%

On most github repos, there will be a README document that is rendered automatically at the root of the repo, and displayed on the repo landing page on github. You can scroll down and know what the repo is about from the README.



The screenshot displays a GitHub repository interface. At the top, a file list shows four items: '3.Make-Data.do', '4.Make-Report.do', 'README.md', and 'ReportDemo.md'. Each file is associated with 'Initialize instructions' and a timestamp ('3 weeks ago' for the first three, and 'last month' for the last). Below the file list, the 'README' tab is selected, showing the rendered content. The README is titled 'PDP Toolkit Guidance' and contains a paragraph explaining the document's purpose, a section titled 'Objective: What is this toolkit about?', and a bulleted list of tasks. To the right of the README content, a 'Languages' section shows a progress bar for 'Stata' at 100.0%.

File Name	Instructions	Timestamp
3.Make-Data.do	Initialize instructions	3 weeks ago
4.Make-Report.do	Initialize instructions	3 weeks ago
README.md	Initialize instructions	3 weeks ago
ReportDemo.md	Prototype section 1 (#10)	last month

PDP Toolkit Guidance

This README.md document explains how this toolkit is structured, and provides step by step instructions on how to use it with your PDP data. The [ReportDemo.md](#) document serves as an example of what kind of report and analyses you can hope to generate with this toolkit.

Objective: What is this toolkit about?

The goal of this toolkit is to replicate the [CTE Toolkit](#) of the [OpenSDP](#) project, using your institution's PDP [Analysis-Ready Files](#). It uses Stata to transform and analyze the data, and generates tables and figures for your analysis.

This toolkit contains code and instructions to

- transform your PDP Analysis-Ready Files into the appropriate data structure for our analysis purpose; we will call this step "Make Data"
- use these transformed datasets to create relevant graphs and tables featured in the [CTE Toolkit](#); we will call this step "Make Report"

Languages

- Stata 100.0%

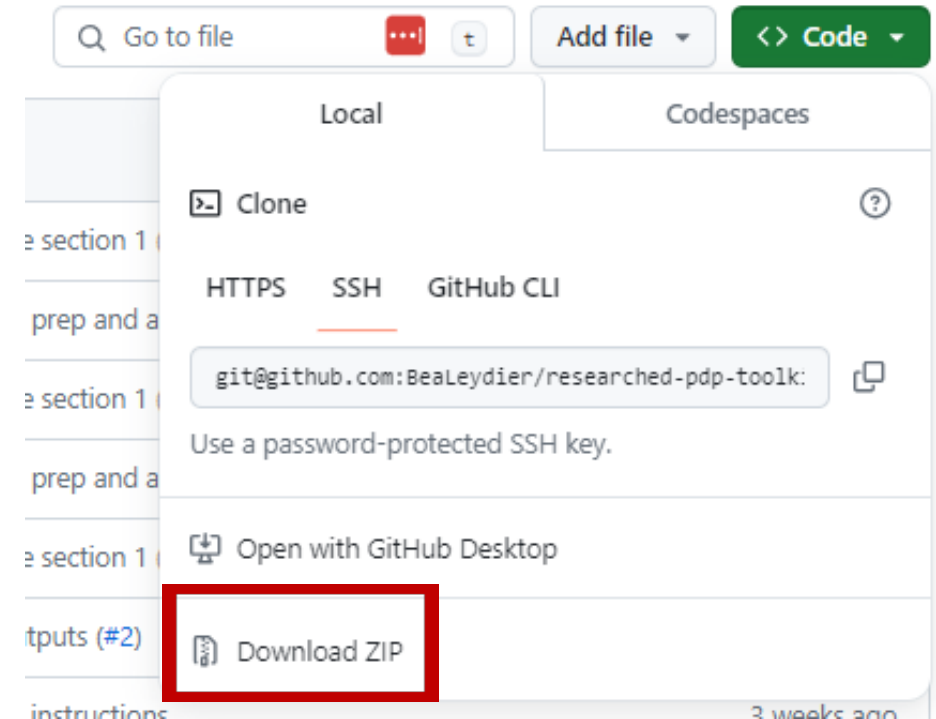
How to get these files on your computer?

The screenshot shows the GitHub repository page for 'researched-pdp-toolkit' by user 'Bealeydiier'. The repository is public and has 1 branch and 0 tags. A red box highlights the green '<> Code' button, with a text overlay saying 'Click on the green <> Code button'. The repository contains a table of files and folders, including '0_scripts', '1_data-pdp', '2_data-toolkit', '3_data-diagnostics', '4_output', '.gitignore', '0.Set-Up.do', '1.Define-Parameters.do', '2.Enter-Data.do', '3.Make-Data.do', '4.Make-Report.do', 'README.md', and 'ReportDemo.md'. The right sidebar shows the 'About' section with the description 'ResearchEd Post-Secondary Data Partnership Toolkit', 'Releases' section with 'No releases published', and 'Languages' section with 'Stata 100.0%'.

File/Folder	Description	Last Commit
0_scripts	Prototype section 1 (#10)	last month
1_data-pdp	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
2_data-toolkit	Prototype section 1 (#10)	last month
3_data-diagnostics	Init data prep and analysis for sections 1 and 2 (#1)	5 months ago
4_output	Prototype section 1 (#10)	last month
.gitignore	Label outputs (#2)	5 months ago
0.Set-Up.do	Initialize instructions	3 weeks ago
1.Define-Parameters.do	Prototype section 1 (#10)	last month
2.Enter-Data.do	Initialize instructions	3 weeks ago
3.Make-Data.do	Initialize instructions	3 weeks ago
4.Make-Report.do	Initialize instructions	3 weeks ago
README.md	Initialize instructions	3 weeks ago
ReportDemo.md	Prototype section 1 (#10)	last month

Option 1 : Download

- From the Code Button, Select **Download ZIP**
- You will download the files onto your computer, and then never have to visit github ever again
- You will have used github like a cloud you download files from



Option 1 will download a .zip file in your chosen location on your own computer

- A .zip file is a compressed folder
 - It is a folder, meaning that it is an object on your computer that contains other files and subfolders
 - It is compressed, meaning that it takes less space on your computer
- You need to unzip or extract or uncompress it
 - Copy, Right click and Extract..., Un-Zip
 - WikiHow based on your OS: <https://www.wikihow.com/Open-a-Zip-File>
- This creates a regular folder on your computer
 - That folder is now your toolkit! It contains all the files you saw on the github page, with the same folder structure (some scripts at the root, some subfolders and files in them)

Option 2 : Fork and Clone

- This is only if you already know github and want to keep using github to work on this project
 - Fork the repo so you have writing access
 - Clone it on your local machine – remember, not in a cloud synced folder
 - Note : .gitignore
 - The repository contains a gitignore that is set to ignore every data file like the AR file you will add to the toolkit
 - This ensures privacy of your data even if you use github to collaborate on the toolkit
 - Data files are always local, never tracked with git, and never pushed to github

Who should have access to the toolkit files?

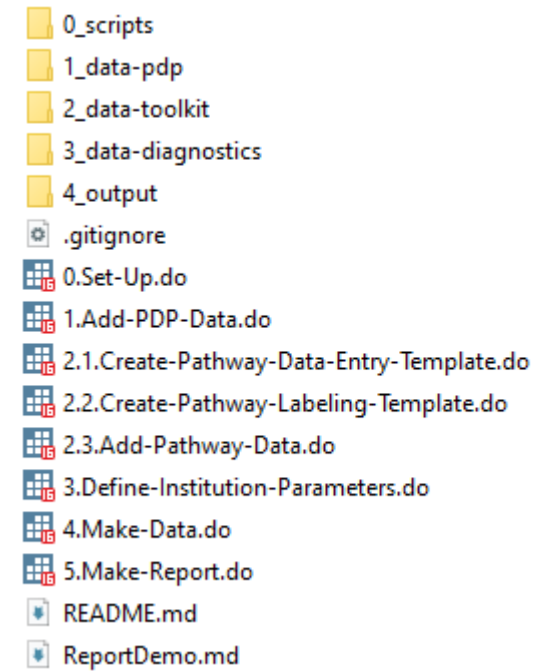
- Anyone who wants to work with your PDP AR data using this toolkit
- Inside the toolkit folder, you will add your PDP Analysis Ready files
 - Make sure your toolkit folder isn't saved somewhere your PDP AR files shouldn't be
 - For example, if you use an online cloud system for your files, you shouldn't save your toolkit folder inside a shared folder that is also shared with people who aren't allowed to access the PDP AR files
- How to work and collaborate on the toolkit
 - **Each person who works on the toolkit can download the files from github to their own machine**, copy the AR files in the dedicated toolkit subfolder (1_data-pdp) and run the toolkit this way
 - Or : You can use a cloud system (Google Drive, Box, Dropbox, OneDrive, etc) to have one toolkit folder on your cloud shared with everyone who will work on the toolkit
 - One person downloads the files from github, saves them in your cloud system, adds the AR files to the dedicated toolkit subfolder (1_data-pdp) and shares the cloud folder with relevant colleagues
 - Or : you can use the forked github repo and collaborate on it
 - Each person cloning the repo will need to also add the AR files locally
 - The gitignore ensures that the AR files are never uploaded onto github
 - Remember: never clone a repo in a cloud-syncing folder, always clone a repo in a local folder only

Will my PDP AR files be shared with anyone if I use this toolkit?

- Short answer : **No**
- Whether you download or clone the toolkit from github, the toolkit code **runs locally** on your own machine, and none of the contents of the **1_data-pdp**, **2_data-toolkit** and **3_data-diagnostic** subfolders are shared with anyone or anything
 - Even if you clone the repo and work on it with github, the contents of these folders are not tracked by git and not pushed to github (thanks to the .gitignore)
- For the toolkit to run, **you need to copy your AR files in the 1_data-pdp subfolder** of the toolkit on your machine, wherever you saved your toolkit
 - The toolkit itself doesn't read any file outside of what's in the toolkit folder
 - The toolkit doesn't share any file that is under the **1_data-pdp**, **2_data-toolkit** and **3_data-diagnostic** subfolders of the toolkit
 - It is important for you to save the toolkit in a location where it is safe to add the AR files (e.g. not on a public cloud folder)

From here onward, follow the README

- The README takes you step by step to everything you need to do to set up the toolkit, get the data ready and run the analyses
 - Set up Stata
 - Install user written commands (once)
 - Define your local file path (at the top of each dofile)
 - Add your PDP AR files
 - Add your pathway data
 - Set parameters
 - Adapt and run scripts for data transformation
 - Adapt and run scripts for data analysis



How do I change my file path?

```
/* Note : In order to define your own file path, enter your machine
username where it says "INSERT-MACHINE-USERNAME" and enter the file
path of your local toolkit folder where it says "INSERT MACHINE SPECIFIC
FILEPATH".
```

```
If you do not know what is your machine username, you can run the
following command into Stata:
```

```
dis ``c(username)''
```

```
What is displayed in response is your machine username. To see all the
other computer and system parameters stored by Stata, you can run
creturn list
```

```
* INSTRUCTIONS: Define machine-specific file path
```

```
if c(username)=="bl517" {
    global root "C:/Users/bl517/Documents/Github/researched-pdp-toolkit"
}
1 → else if c(username)=="INSERT-MACHINE-USERNAME" {
2 → global root "INSERT MACHINE-SPECIFIC FILEPATH"
}
else {
    di as err "Please enter machine-specific path information"
    exit
}
```


Why are file paths defined inside an “if”

The lines of code below store the root of the local folder where this tool is saved in a global called `root`. As a reminder on globals in Stata, they allow you to store once something that the code refers to frequently. The global works as a placeholder in the rest of the code. To call a global in Stata, you use the global name preceded by `$`. In this case, we are using the global `root` to refer to your local machine filepath where this tool is saved. In the toolkit code, each time we are calling a file in the tool, it will be called from `$root`, which will be automatically replaced by your own local filepath. This ensures you only need to define the filepath once at the top of each dofile, and not each time a file is read or exported.

The local filepath typically changes from one user to the next. The if condition below ensures that this file can be run on multiple machines at once, which is particularly useful if you are collaborating on this tool with multiple people, for example using github or a shared cloud storage like Dropbox. In Stata, when we use an “if” condition followed by brackets, the code inside the brackets is run only if the “if” condition is true. Otherwise, that code is ignored, and Stata moves on to the next lines of code after the brackets. The “else if” works the same way : if the condition is true, it runs the code inside the brackets that follow it, if not, it ignores it.

Here, the first “if” condition will be true if the machine you are running this file from has `bl517` as its username (FYI this is the username of the developer of this tool), and if that is the case, the contents of the code inside the brackets that follow the “if” will be run. In this case, this code defines the global `root`, which is the placeholder for the filepath of the root of the code folder (this placeholder is used in all of this tool's code). Here, it is defined as the filepath the user `bl517` defined for their own machine. When you are running the file on your own machine, that first “if” condition will be false and that code ignored.

You can use the first “else if” condition to add your own username in the if condition (which will then return true when the file is run from your own machine), and define your own local filepath as the global `root`. If you are collaborating on this tool with other people, you can add another else if block of code (with subsequent brackets) for them to add their username and define their own filepath in the global `root`.

The final else condition returns an error message in red and exits the script (i.e stops the execution of the code) if none of the previous conditions returned true. That is, as long as you haven't added your machine-specific username in the “else if” command, the script will return an error. This functions as a reminder to do it, given none of the subsequent code will work if you haven't defined the global `root` as you local machine-specific filepath.

```
* =====
* PART 1. - File path
* =====

* Stata set up
set more off

* INSTRUCTIONS: Define machine-specific file path

if c(username)=="bl517" {
    global root "C:/Users/bl517/Documents/Github/researched-pdp-toolkit"
}
else if c(username)=="INSERT-MACHINE-USERNAME" {
    global root "INSERT MACHINE-SPECIFIC FILEPATH"
}
else {
    di as err "Please enter machine-specific path information"
    exit
}
```


How do I add my PDP data?

1

2

1.Add-PDP-Data.do

```
1
2  /*****
3
4      ADD PDP DATA
5
6  *****/
7
8  * =====
9  * PART 1. - Add your PDP Analysis Ready Files to the 1_data-pdp subolder
10 * =====
11
12  /* INSTRUCTIONS:
13     Find your PDP Analysis Ready Files, copy and save them under
14     the 1_data-pdp subfolder of this toolkit.
15
16     For this toolkit, we are only using the AR Cohort file and the AR
17     Course file, so you only need to copy these two in that folder.
18  */
19
20 * =====
21 * PART 2. - Add the AR File Names
22 * =====
23
24  /* INSTRUCTIONS:
25     Replace the file names in quotes with the names of the PDP Analysis
26     Ready files you just saved under 1_data-pdp.
27
28     Do NOT change the names of the globals (arcohortfile and arcooursefile).
29
30     Globals in Stata allow you to store once something that the code refers
31     to frequently. The global works as a placeholder in the rest of the code.
32     In this case, we are using the global arcohortfile to refer to your
33     AR Cohort File, and the global arcooursefile to refer to your AR Course
34     File. In the toolkit code, everytime we are calling one of these files,
35     we are calling the corresponding global name (preceded by $), and Stata
36     will automatically replace it with the name you defined here. This
37     avoids you having to change the filename in every line of code where
38     the file is being opened or referenced. You only have to change
39     it once here, and everywhere else in the code, these files are
40     referred to from the name of the globals directly.
41  */
42
43  * Name of your PDP AR files (including the file extension)
44  global arcohortfile "add-your-file-name-here.xlsx"
45  global arcooursefile "add-your-file-name-here.csv"
46
```

How do I add my student pathway data?



StudentID	FirstName	MiddleName	LastName	Cohort	CohortTerm	AcademicYear	ProgramofStudyYear_input
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2013-14	540101
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2014-15	540101
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2015-16	513899
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2016-17	513899
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2015-16	321624
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2016-17	321624
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2017-18	321624
14891478	Derek	Kyra	Coffey	2018-19	SPRING	2018-19	540101
14891478	Derek	Kyra	Coffey	2018-19	SPRING	2019-20	540101

Template generated by the tool, listing all students present in your PDP data for each year they are listed, and adding a column for you to fill their program of study for each year (highlighted in yellow in the picture, which contains mock data). Even if you don't use this template, you will need to add a student pathway file which contains the 3 columns in red, named and formatted the same way (**StudentID**, **AcademicYear**, **ProgramofStudyYear_input**). The other variables are added just for information. The **ProgramofStudyYear_input** variable should contain the program code of the pathway.

Template generated by the tool, listing all program of study codes present in your PDP data and your student pathway data combined, adding a unique index (starting from 1) to identify them (**ProgramofStudy_ID**), and a column for you to add the pathway name or label (in yellow in the picture with mock data). Even if you don't use this template, you will need to add a pathway label file with these 3 columns formatted the same way.

ProgramofStudy_ID	ProgramofStudy	ProgramofStudy_Label
1	230101	Engineer Tech
2	234501	Literature
3	261504	Economics
4	321624	Mech Repair
5	422814	Health
6	513801	Mech Engineer
7	513899	CS
8	540101	IT Tech



How do I change the institution parameters?

By default, the institution type is set as 4year (for 4 year institutions or programs). If you want to change it to 2year (for 2 year institutions or programs), remove the asterix in front of line 20 where the institution type “2year” is defined, to un-comment it, and add an asterix in front of line 21 where the institution type “4year” is define, to comment it out.



```
3.Define-Institution-Parameters.do  Untitled
1
2  /*****
3
4      PARAMETERS
5
6      This is where you define the parameters for your analysis, like whether
7      you are a 2-year or 4-year college.
8
9  *****/
10
11  *  =====
12  *  PART 1. - Institution Type
13  *  =====
14
15  /* INSTRUCTIONS:
16     Comment out the institution type that you are NOT, by placing an
17     asterix in front of it.
18  */
19
20  *global institutiontype "2year"
21  global institutiontype "4year"
22
```



```
19
20  global institutiontype "2year"
21  *global institutiontype "4year"
22
```