STRATEGIC DATA PROJECT

ResearchEd

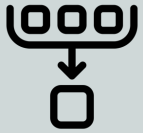
STEP BY STEP GUIDE

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

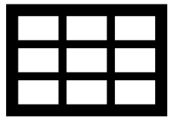
What's included in this Extension?

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

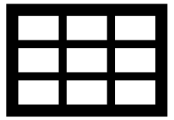
Research questions, relevant research, and suggestions for how to use analyses to drive practice

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

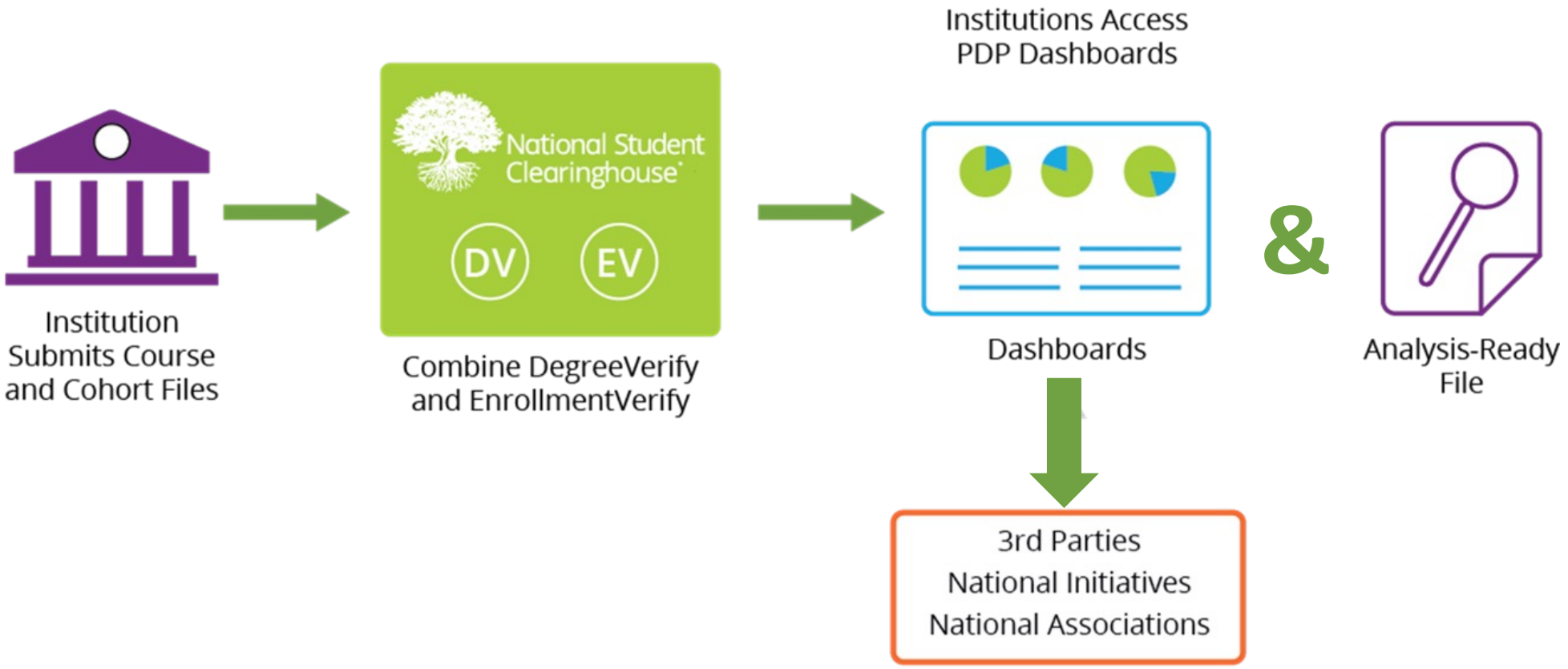
Code Files

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? (A/R 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	Jayme	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	ll	>20-24	Transfer-in	Part-time
Brian	Hunter	Lee	555-55-5561	19910524	555-55-5561	00675000	2		Older than 24	First-time	Full-time
Derek	Kyra	Coffey	555-55-5562	19920619	555-55-5562	00675000			20 and younger	Transfer-in	Part-time
Nora	Sheila	Mccarthy	555-55-5563	19930720	555-55-5563	00675			Older than 24	First-time	Full-time
Kelly	Tucker	Flowers	555-55-5564	20100819	555-55-5564	0		ll	Older than 24	Transfer-in	Part-time
Cynthia	Luke	Kane	555-55-5565	20090615	555-55-5			Fall	Older than 24	First-time	Full-time
Quemby	Anastasia	Brady	555-55-5566	20080612	555		14-15	Fall	20 and younger	Transfer-in	Part-time
Chaim	Kadeem	Pace	555-55-5567	19910524			2015-16	Fall	20 and younger	First-time	Full-time
Grady	Cameran	Evans	555-55-5568	1992		00	2012-13	Fall	Older than 24	Transfer-in	Part-time
Colin	Hedy	Santiago	555-55-5569			675000	2013-14	Fall	Older than 24	First-time	Full-time
Christen	Maris	Perry	555-55-5570		7(	00675000	2016-17	Fall	20 and younger	Transfer-in	Part-time
Brody	Kelly	Mayer	555-55-		5-5571	00675000	2017-18	Fall	20 and younger	First-time	Full-time
Alvin	Lilah	Emerson	555		555-55-5572	00675000	2014-15	Fall	20 and younger	Transfer-in	Part-time
Pamela	Callie	Valencia	55	524	555-55-5573	00675000	2015-16	Fall	>20-24	First-time	Full-time
Colorado	Sylvester	Joyner	555-	9920619	555-55-5574	00675000	2012-13	Fall	>20-24	Transfer-in	Part-time
Elmo	Henry	Marsh	555-55	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

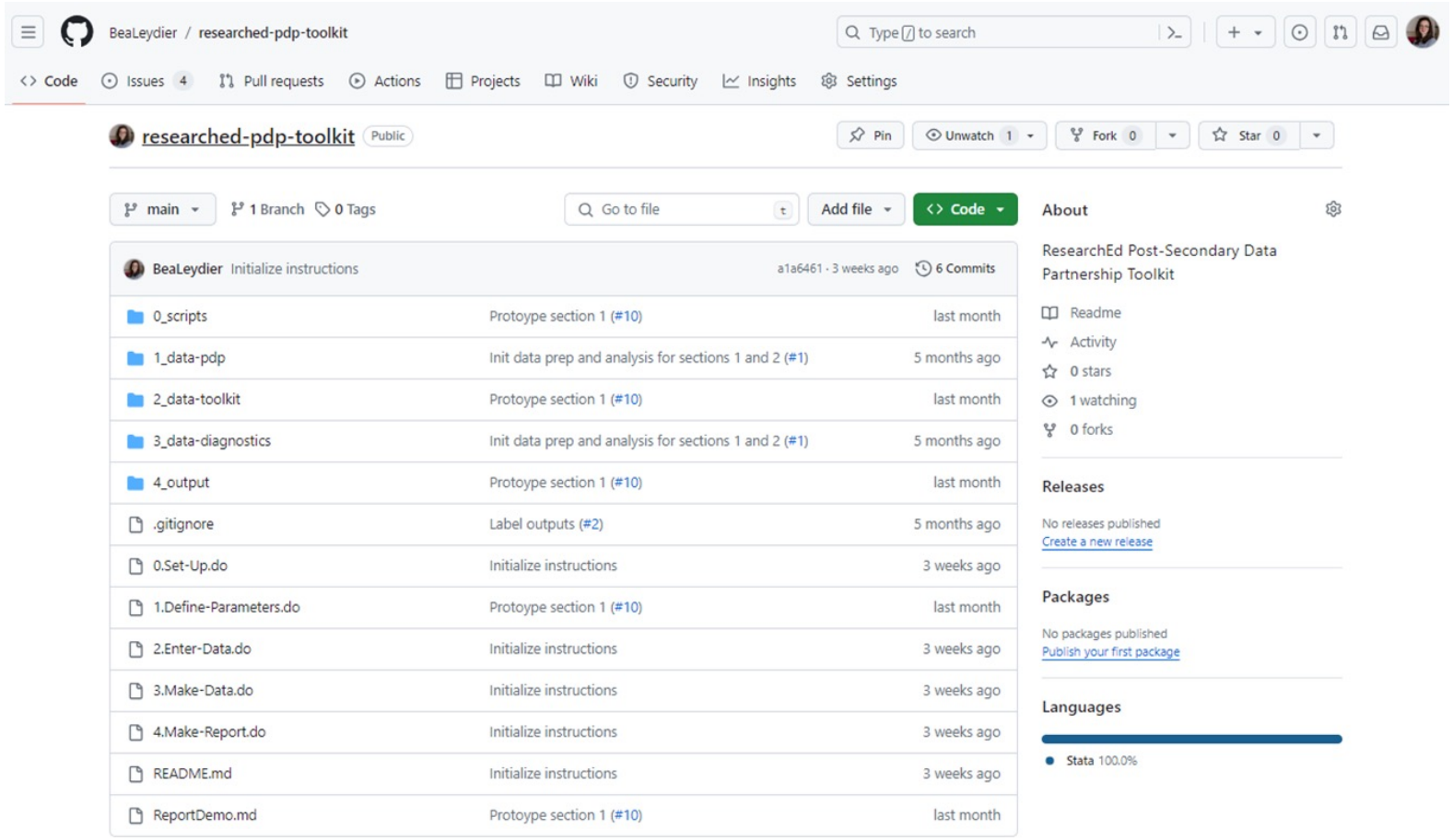
How does the PDP Extension work ?

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

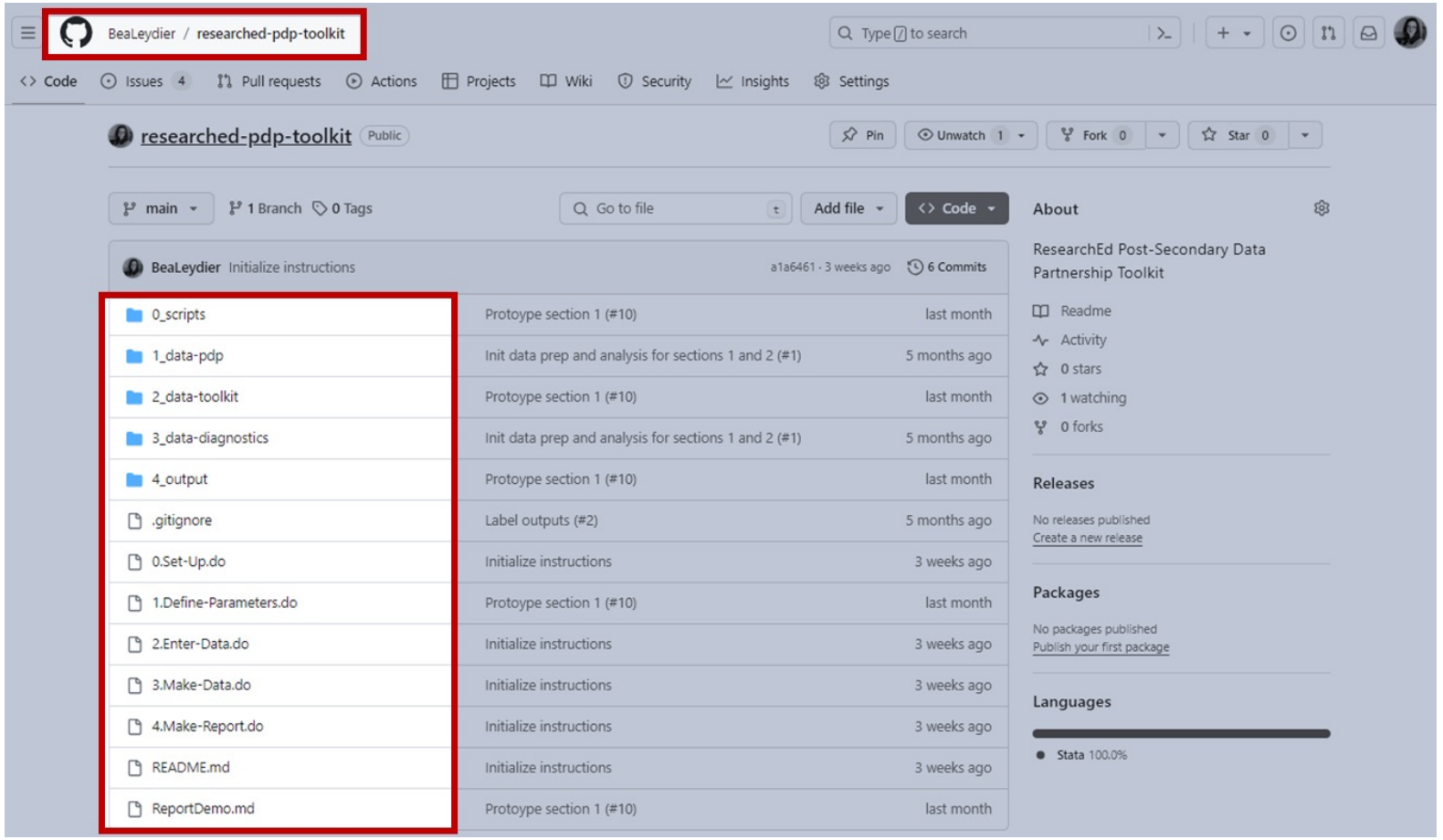
How to access the PDP Extension ?

- Open any web browser and go to https://github.com/BeaLeydier/researched-pdp-toolkit
- 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

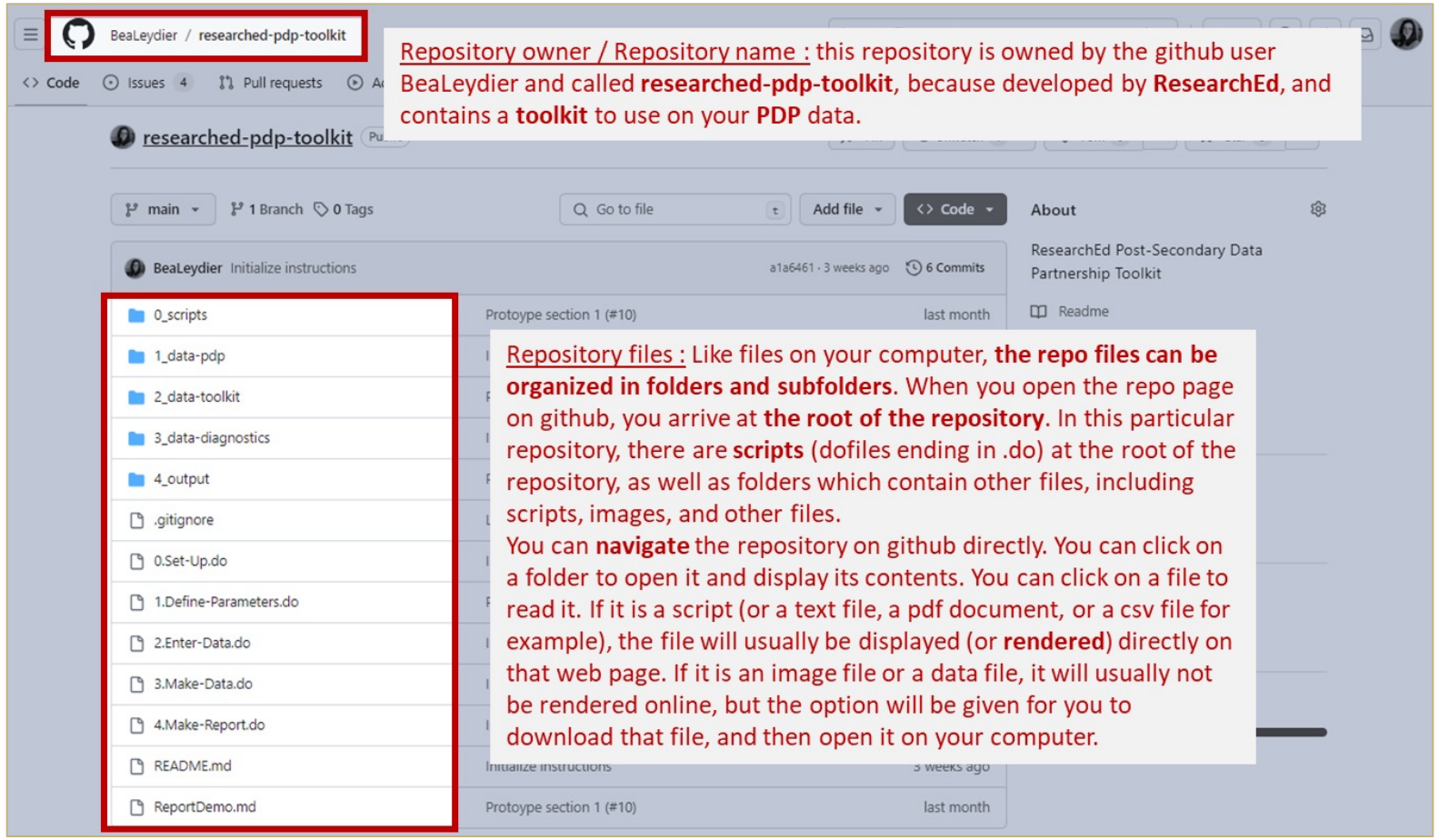
What am I looking at ?



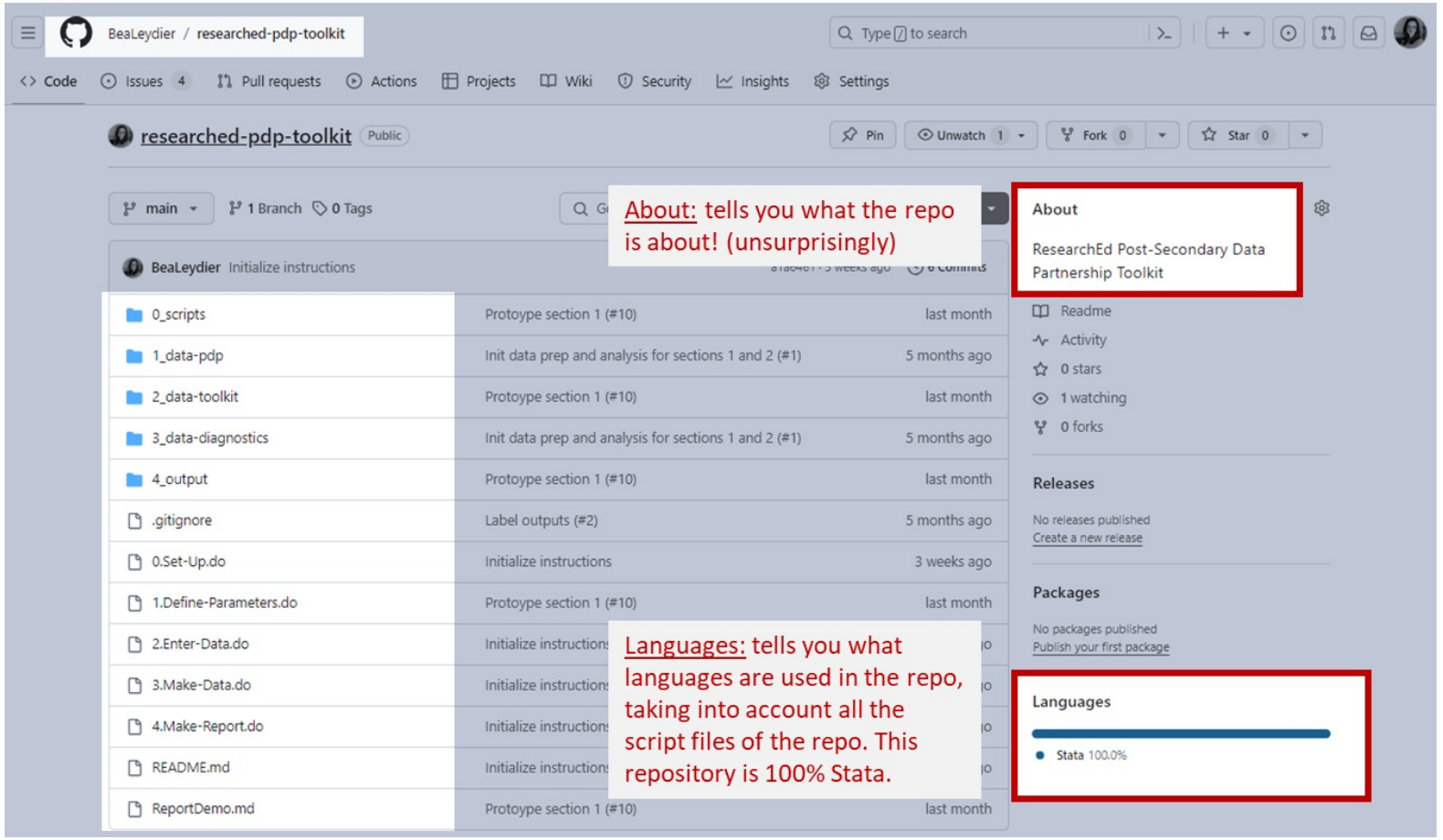
You have landed at the root of the repository, which contains files. Ignore everything else!



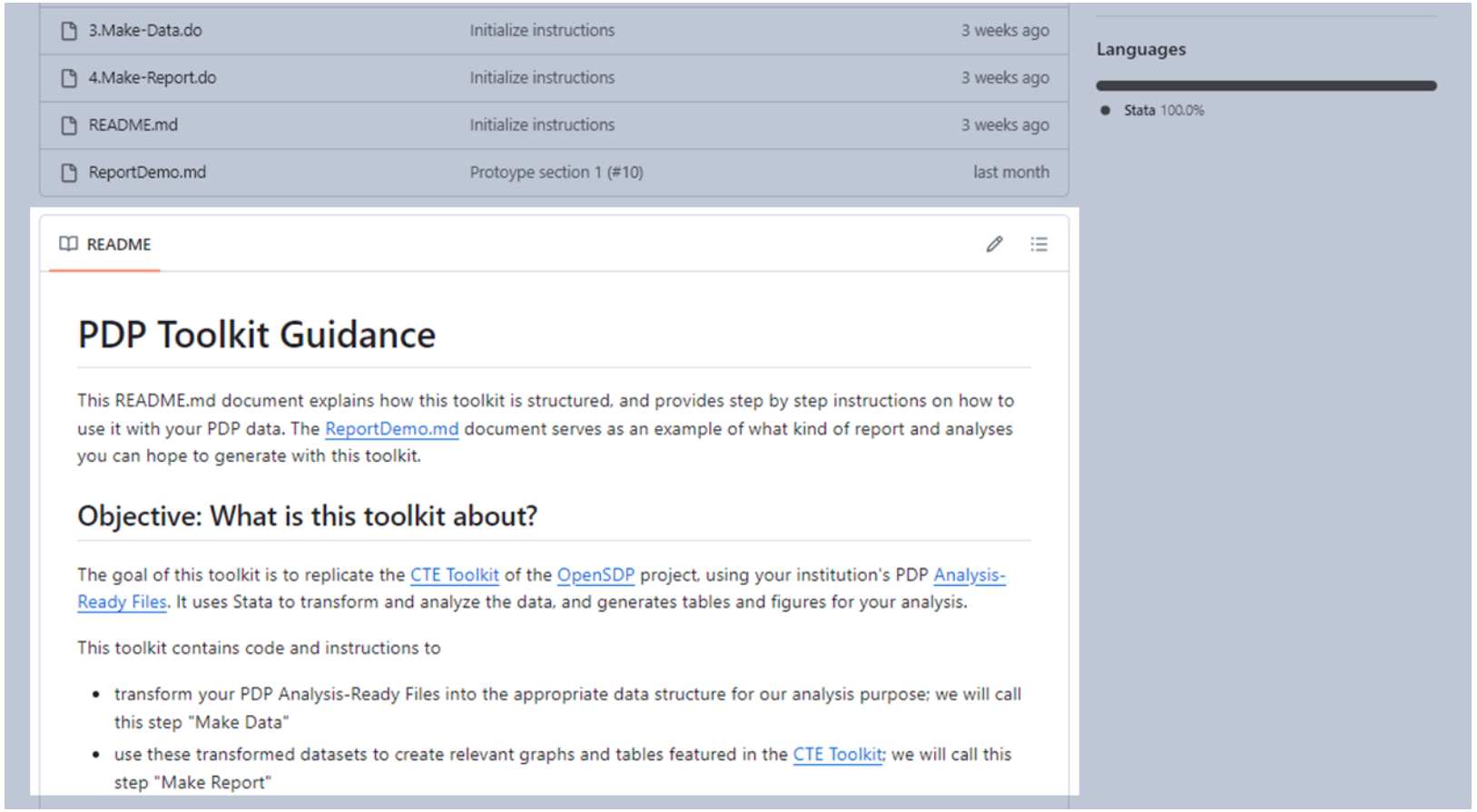
You have landed at the root of the repository, which contains files. Ignore everything else!



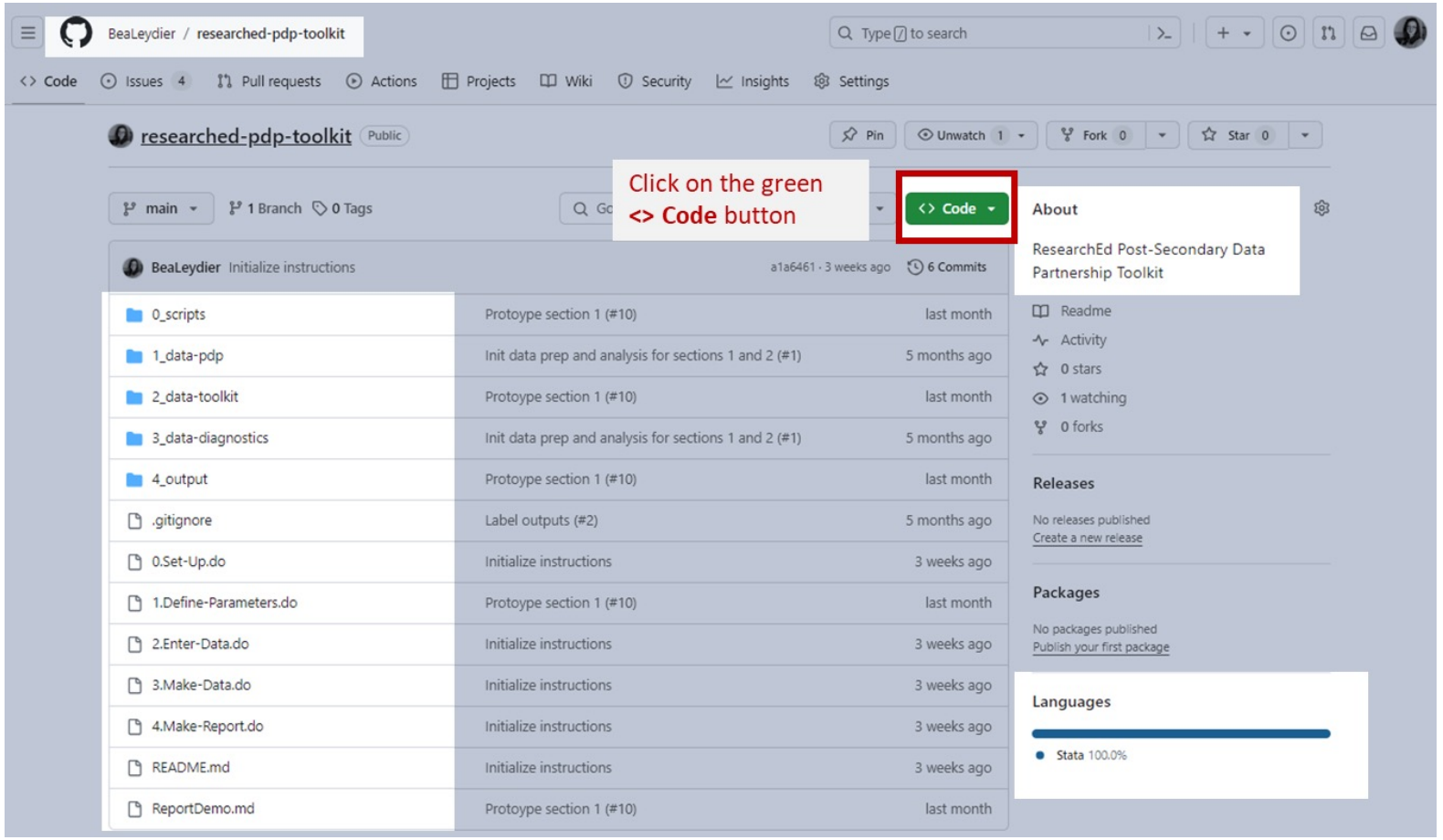
Ok, there are a few more things you can look at on the repo landing page.



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.

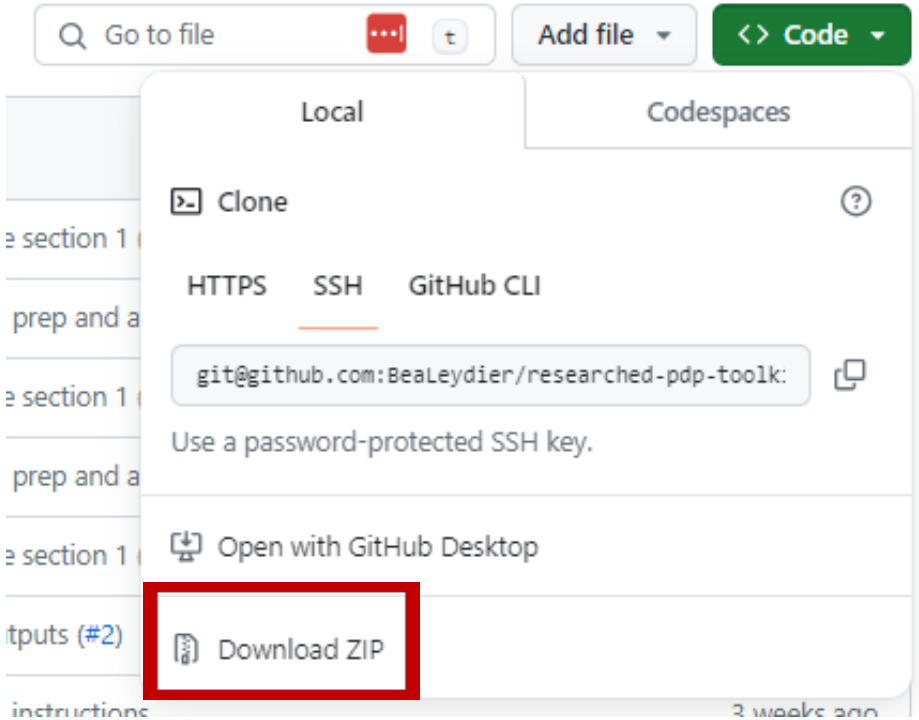


How to get these files on your computer?



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?

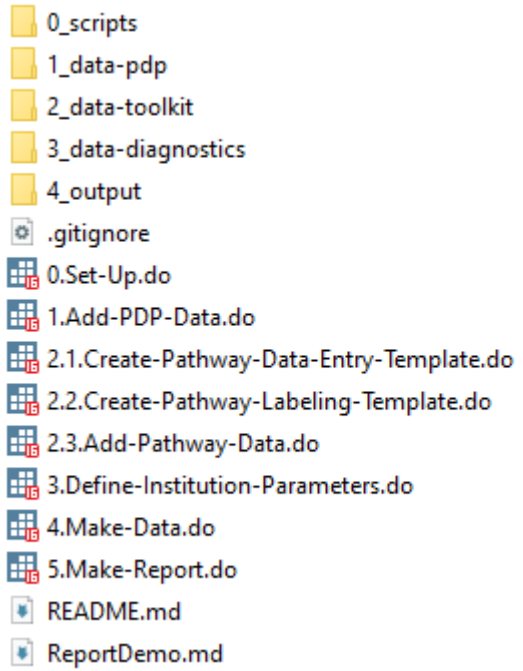
- Anyone who wants to work with your PDP AR data using this toolkit
- Inside the toolkit folder, you will add your PDP Analysis Ready files
  - Make sure your toolkit folder isn't saved somewhere your PDP AR files shouldn't be
- How to work and collaborate on the toolkit
  - **Each person who works on 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
    - 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 can't access the PDP AR files

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
  - 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 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.Add-PDP-Data.do

```
/*******************************************************************************

        ADD PDP DATA

*******************************************************************************/

*   ========================================
*   PART 1. - Add your PDP Analysis Ready Files to the 1_data-pdp subolder
*   ========================================

    /* INSTRUCTIONS:
        Find your PDP Analysis Ready Files, copy and save them under
        the 1_data-pdp subfolder of this toolkit.

        For this toolkit, we are only using the AR Cohort file and the AR
        Course file, so you only need to copy these two in that folder.
    */

*   ========================================
*   PART 2. - Add the AR File Names
*   ========================================

    /* INSTRUCTIONS:
        Replace the file names in quotes with the names of the PDP Analysis
        Ready files you just saved under 1_data-pdp.

        Do NOT change the names of the globals (arcohortfile and arcoursefile).

        Globals in Stata allow you to store once something that the code refers
        to frequently. The global works as a placeholder in the rest of the code.
        In this case, we are using the global arcohortfile to refer to your
        AR Cohort File, and the global arcoursefile to refer to your AR Course
        File. In the toolkit code, everytime we are calling one of these files,
        we are calling the corresponding global name (preceded by $), and Stata
        will automatically replace it with the name you defined here. This
        avoids you having to change the filename in every line of code where
        the file is being opened or referenced. You only have to change
        it once here, and everywhere else in the code, these files are
        referred to from the name of the globals directly.
    */

* Name of your PDP AR files (including the file extension)
global arcohortfile "add-your-file-name-here.xlsx"
global arcoursefile "add-your-file-name-here.csv"

```

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	
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2014-15	
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2015-16	
12345678	Hiroko	Rendall	Lee	2013-14	SPRING	2016-17	
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2015-16	
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2016-17	
14448096	Hillary	Ocean	Ashley	2015-16	FALL	2017-18	
14891478	Derek	Kyra	Coffey	2018-19	SPRING	2018-19	
14891478	Derek	Kyra	Coffey	2018-19	SPRING	2019-20	

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

How do I define my institution's parameters?

3.Define-Institution-Paramete...

```
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
23
```