

How to use Microsoft Access to extract data from the 2020 Census P.L. 94-171 Legacy Format Summary Files

This document provides a step by step example of how to use the Census Bureau provided Microsoft Access database shell to import the 2020 Census P.L. 94-171 Summary File data and then extract data from the files for use.

Notes and Assumptions:

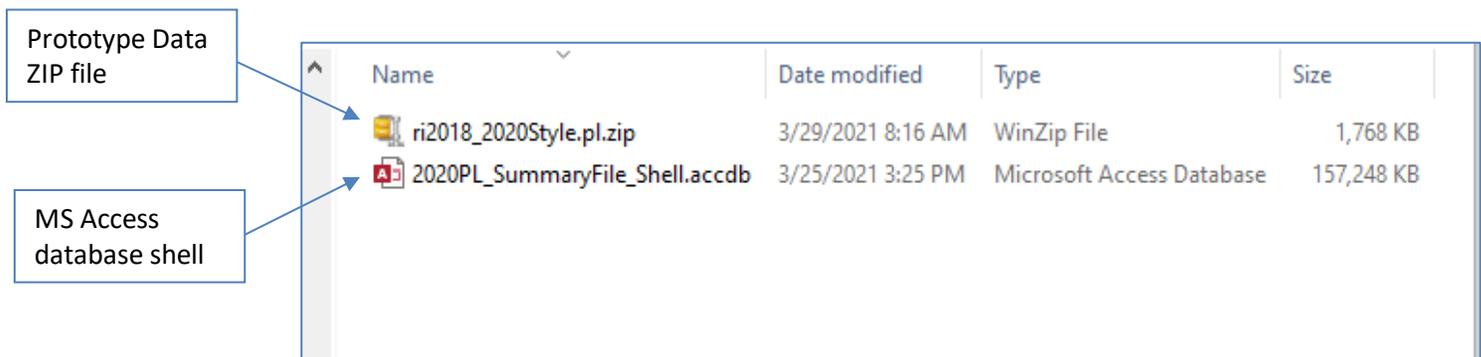
- This example uses a Microsoft Access [2020PL Summary File database shell](#).
- This example uses the [Prototype redistricting data](#) produced from the 2018 End-to-End Census Test, but the same process will work for any state's dataset once they are made available.
- The legacy format redistricting data summary files consist of four relational tables: the geoheader, data segment 01 (contains P1 & P2 tables), data segment 02 (contains P3, P4 & H1 tables), and data segment 03 (contains P5 table)
- The Microsoft Access shell includes examples of extract queries for blocks, counties, and county subdivisions. To identify the summary level number for other geographies, please refer to the technical documentation in [English](#) or [Spanish](#).
- When working with a large state like **California** or **Texas**, all four files cannot be loaded into a single Microsoft Access database due to a Microsoft Access file size limitation. This limitation can be overcome by loading both the geoheader and each data segment into separate databases, repeating for each data segment to generate three databases, each one containing the geoheader and a single data segment. Once this is done the procedure is the same for extracting data from the data segment in each database.

IMPORTING THE DATA

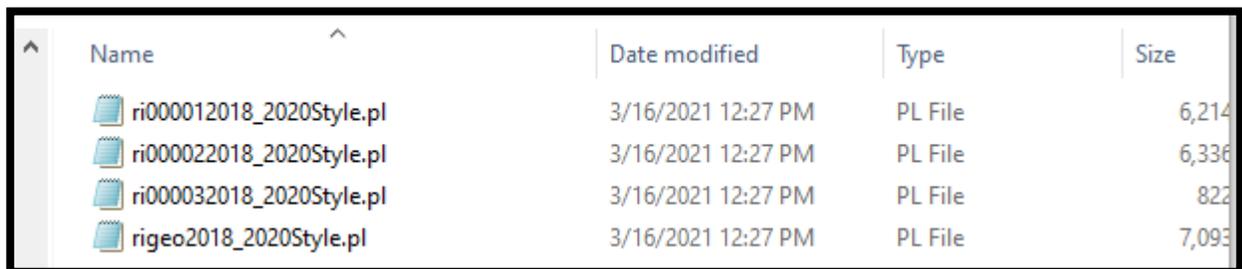
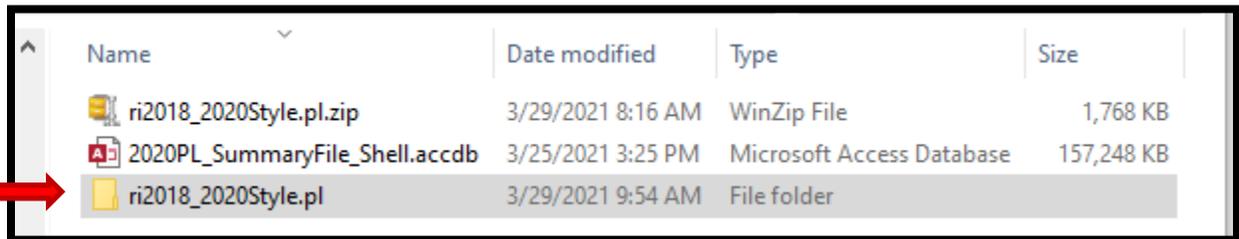
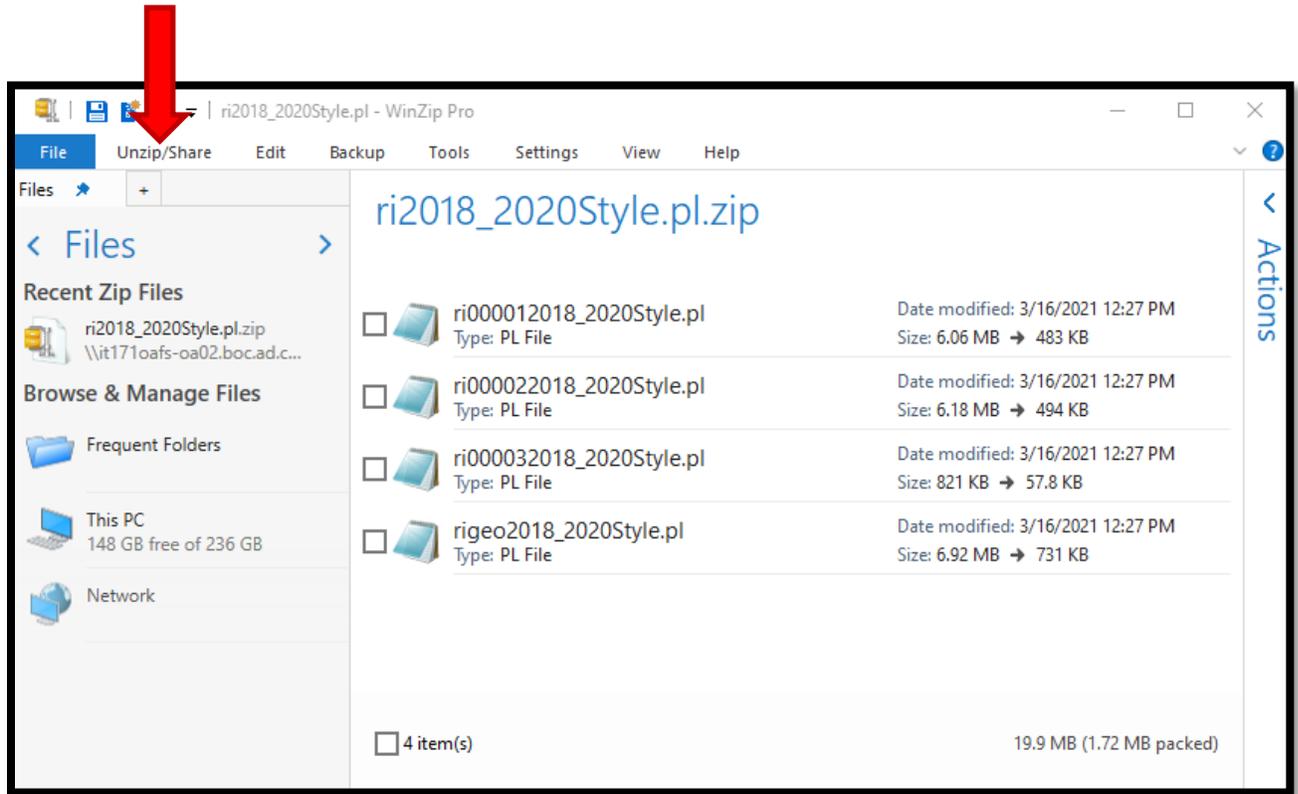
A.) Start by downloading the MS Access shell and the data files you want to use. The Access shell and the prototype data files can be found at:

<https://www.census.gov/programs-surveys/decennial-census/about/rdo/program-management.html#P3>

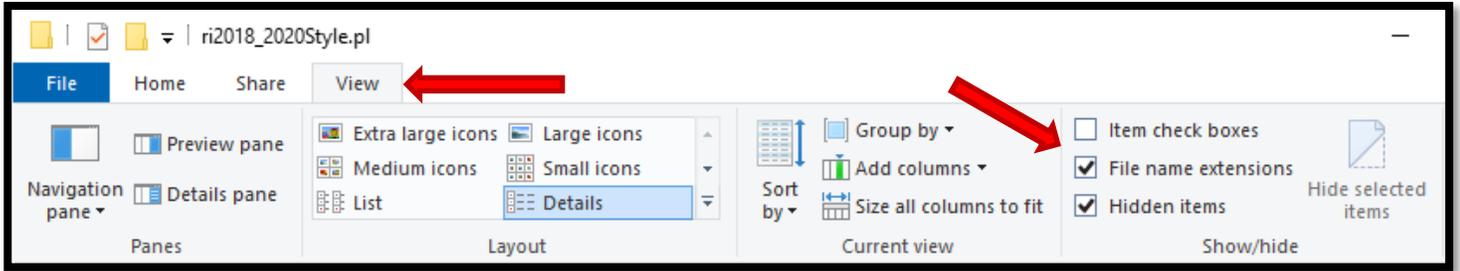
Links to the official data will be posted here (<https://www.census.gov/programs-surveys/decennial-census/about/rdo/summary-files.html#P1>) once those files are published.



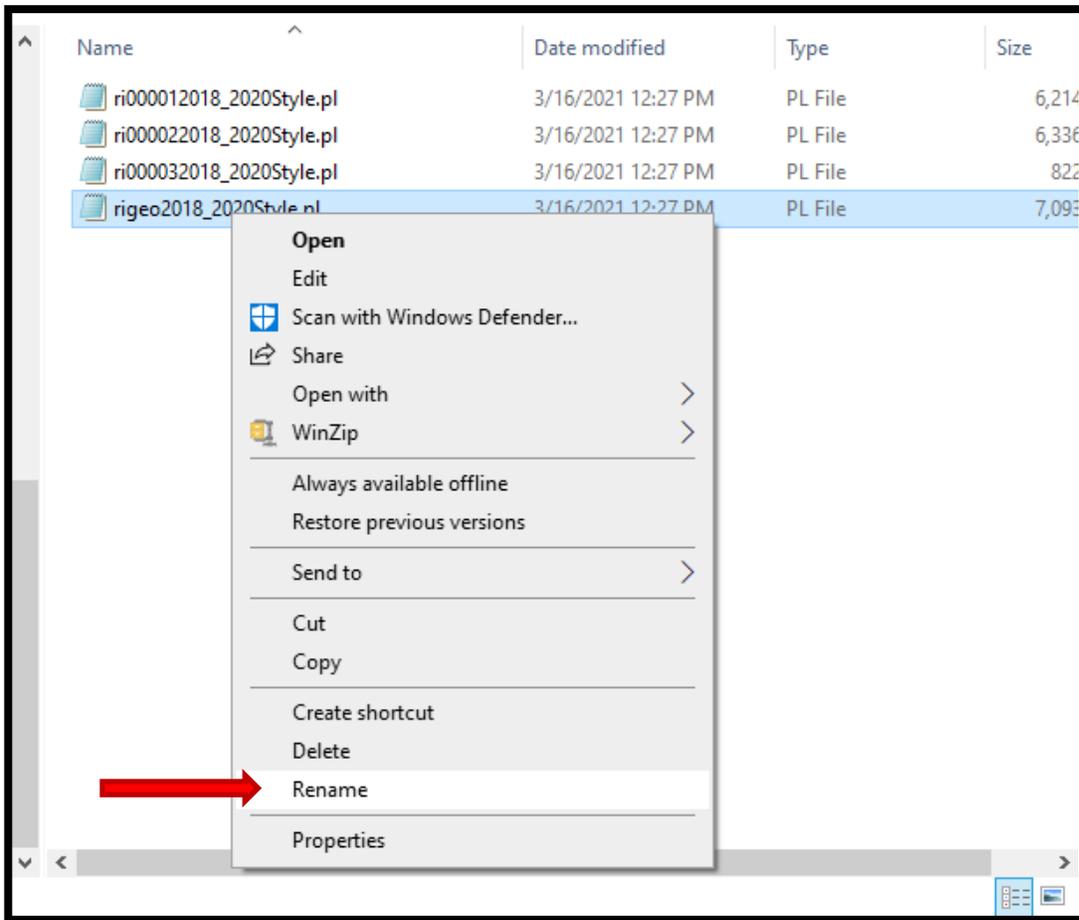
B.) Open the .zip file and extract the files found inside. This example uses WinZip, you may use another software package or Microsoft Windows to unzip these files. The files must be unzipped before beginning the import process. Once unzipped, navigate into the folder that contains the unzipped files.



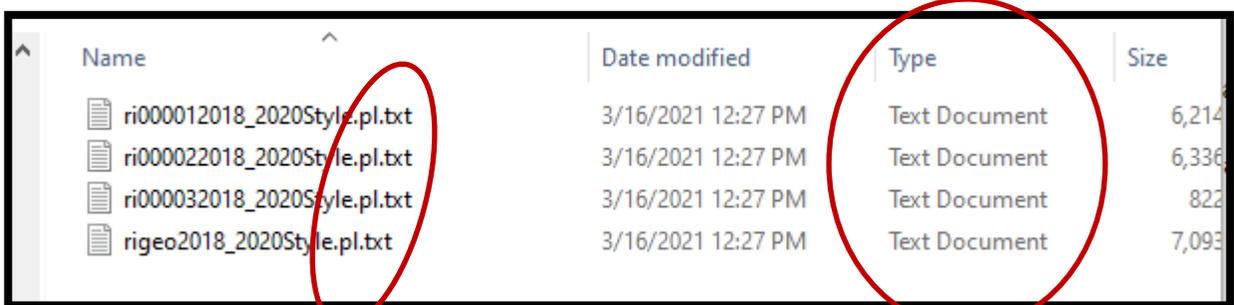
C.) Your Windows environment must be set to show file name extensions. This can be set through the View function of Windows Explorer. Make sure the “File name extensions” box is checked.



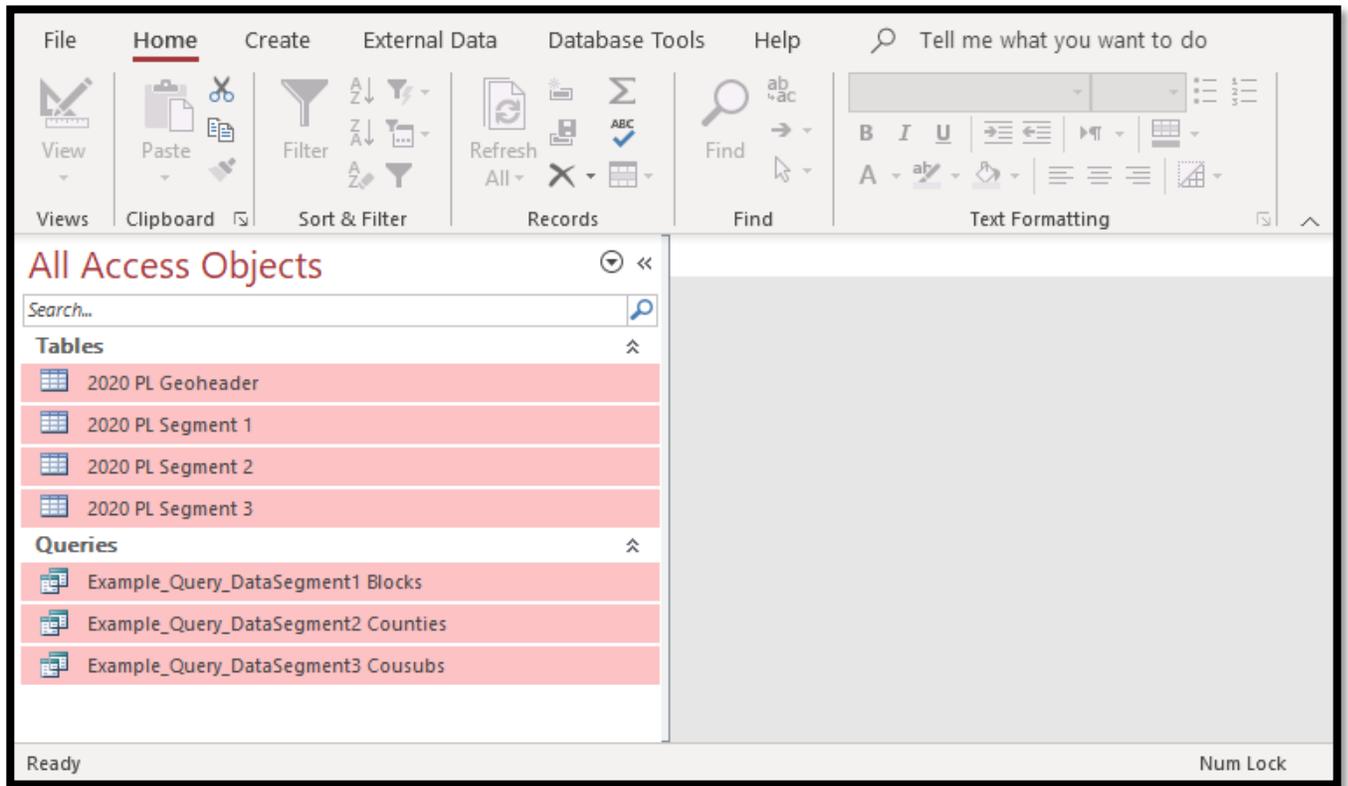
D.) All files with a .pl extension must be changed to .txt files. Right click on the first file that has a .pl extension. Choose “Rename” and add .txt to the end of the filename. Hit Enter. Repeat for each file that has a .pl extension.



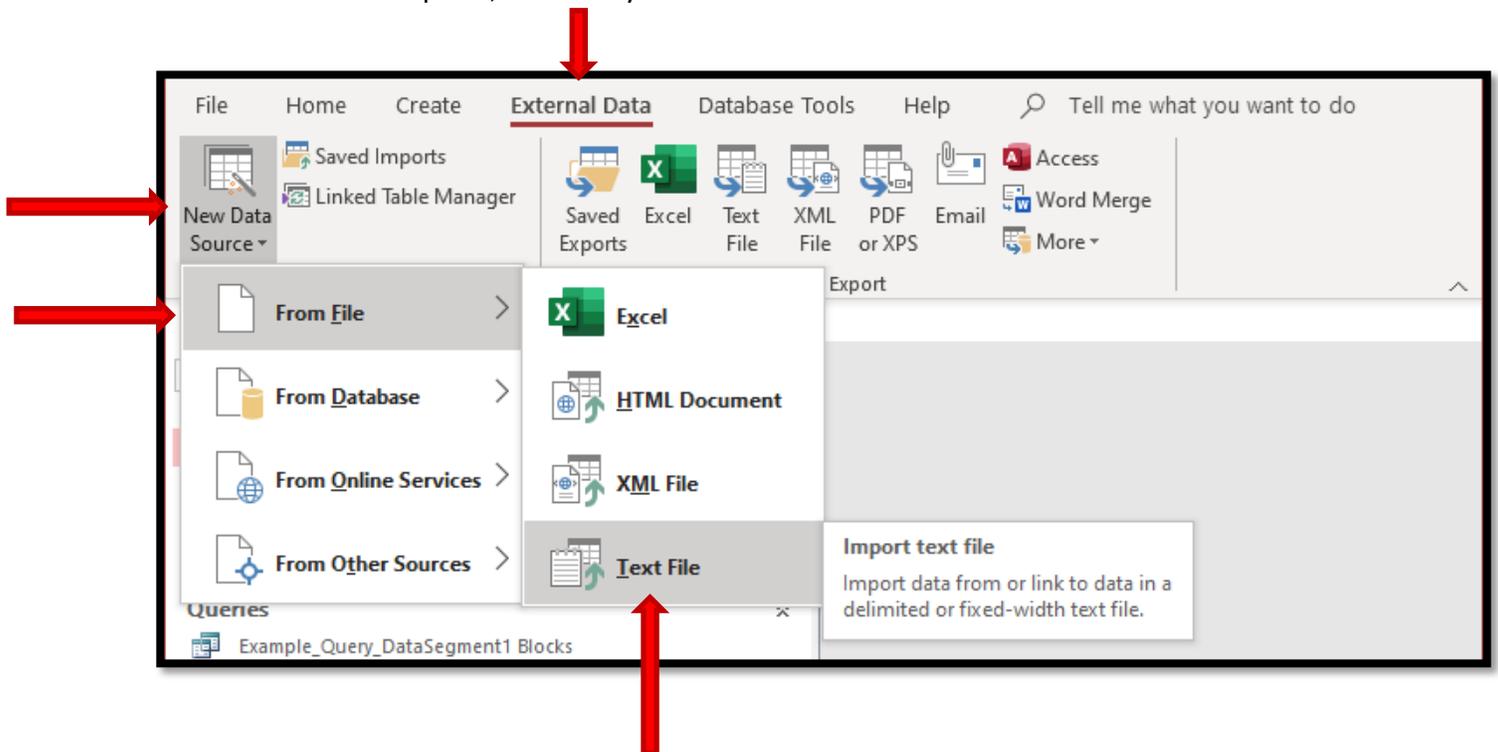
Note: The file “Type” should change from “PL File” to “Text File” after you add the .txt file extension to the name and hit enter.



E.) Open the 2020PL_SummaryFile_Shell using Microsoft Access. You'll see the placeholder tables for the geoheader (2020 PL Geoheader) and 3 data segments (2020 PL Segment 1; 2020 PL Segment 2; 2020 PL Segment 3). You will also see three example data queries, one for blocks, one for counties and one for county subdivisions.



F.) Click on the “External Data” menu, select the “New Data Source” option, select the “From File” option, and finally select “Text File”.



G.) Use the “Browse” button to navigate to the files’ location and select which file you want to bring into the database.

Get External Data - Text File

Select the source and destination of the data

Specify the source of the definition of the objects.

File name: Products\MSAccessShell\GuideWorkingFolder\ri2018_2020Style\ri2018_2020Style.pl.txt Browse...

Specify how and where you want to store the data in the current database.

We will not import table relationships, calculated columns, validation rules, default values, and columns of certain legacy data types such as OLE Object.

Search for "Import" in Microsoft Access Help for more information.

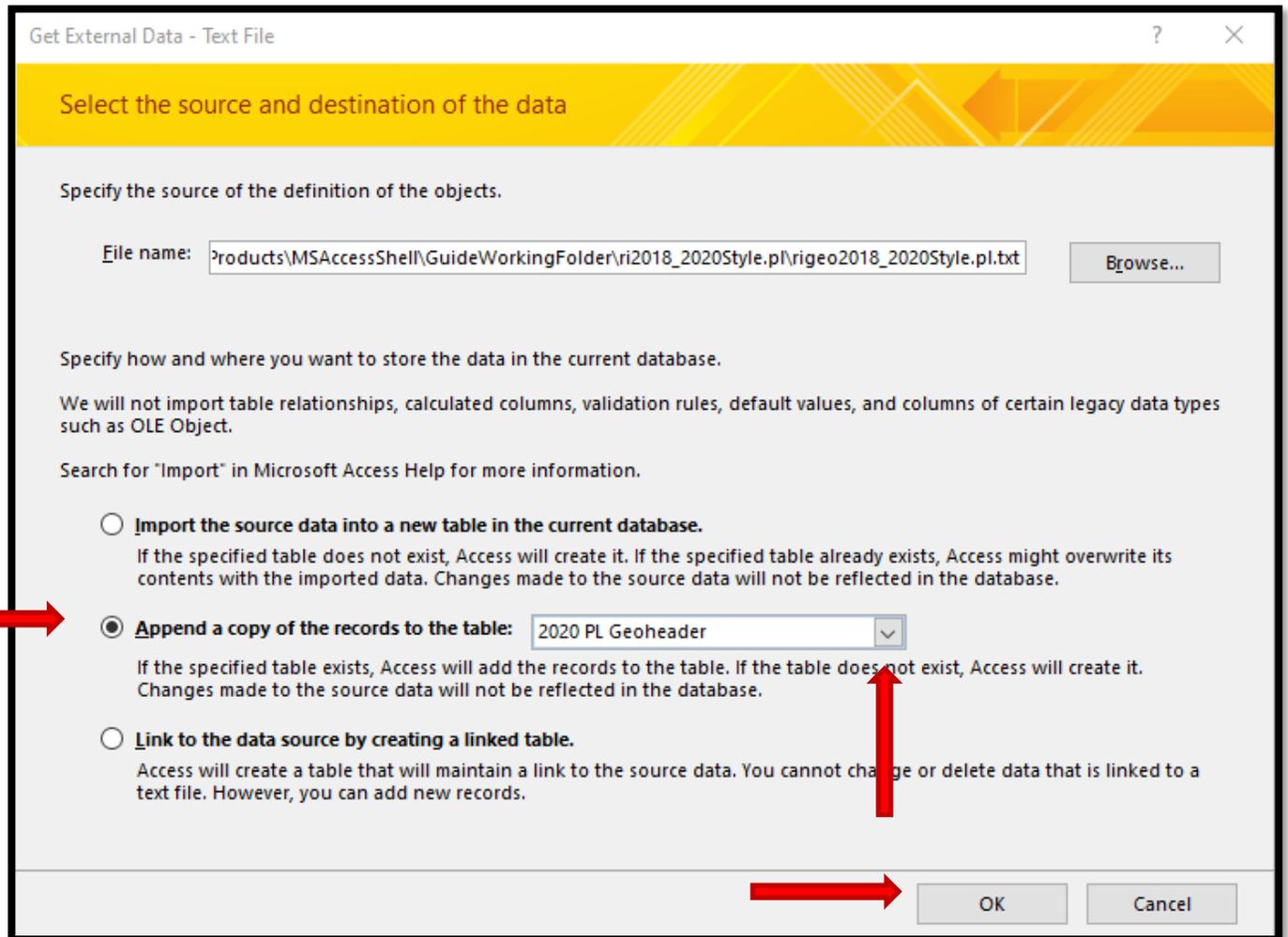
Import the source data into a new table in the current database.
If the specified table does not exist, Access will create it. If the specified table already exists, Access might overwrite its contents with the imported data. Changes made to the source data will not be reflected in the database.

Append a copy of the records to the table: 2020 PL Geoheader
If the specified table exists, Access will add the records to the table. If the table does not exist, Access will create it. Changes made to the source data will not be reflected in the database.

Link to the data source by creating a linked table.
Access will create a table that will maintain a link to the source data. You cannot change or delete data that is linked to a text file. However, you can add new records.

OK Cancel

I.) Click on the radio button next to “Append a copy of the records to the table.” Select the table in the pull-down menu that corresponds with the file that you are importing. Once everything is selected, hit the “OK” button



Note: Make sure to import the correct text file into the correct Microsoft access table shell. For the prototype data from the 2018 End-to-End Test conducted in Rhode Island, you can identify each in the file name. The correct table shells and text file imports are listed in the table below.

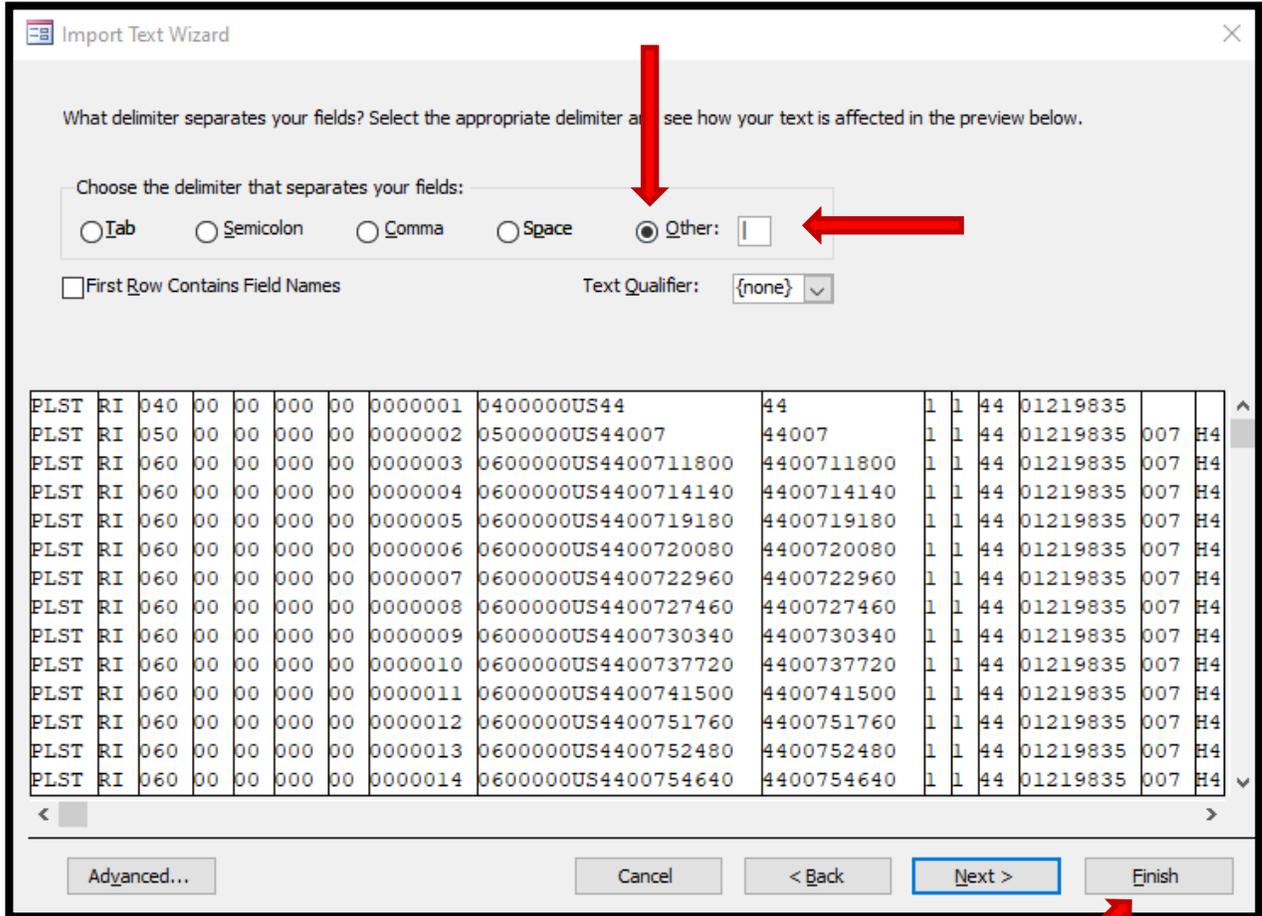
Microsoft Access Shell Table Name	Text Document Name
2020 PL Geoheader Fields	rigeo2018_2020Style.pl.txt
2020 PL Segment 1 Fields	ri00012018_2020Style.pl.txt
2020 PL Segment 2 Fields	ri00022018_2020Style.pl.txt
2020 PL Segment 3 Fields	ri00032018_2020Style.pl.txt

Two Character State Abbreviation

Data Segment number

Year

K.) Set the radio button selection to “Other” and type a pipe character “|” into the adjoining text box. NOTE: To type a pipe character on a standard U.S. keyboard, hold down the “Shift” key and click on the backslash “\” key. Then, select the “Finish” button.



Use these keys together to get a pipe “|”.

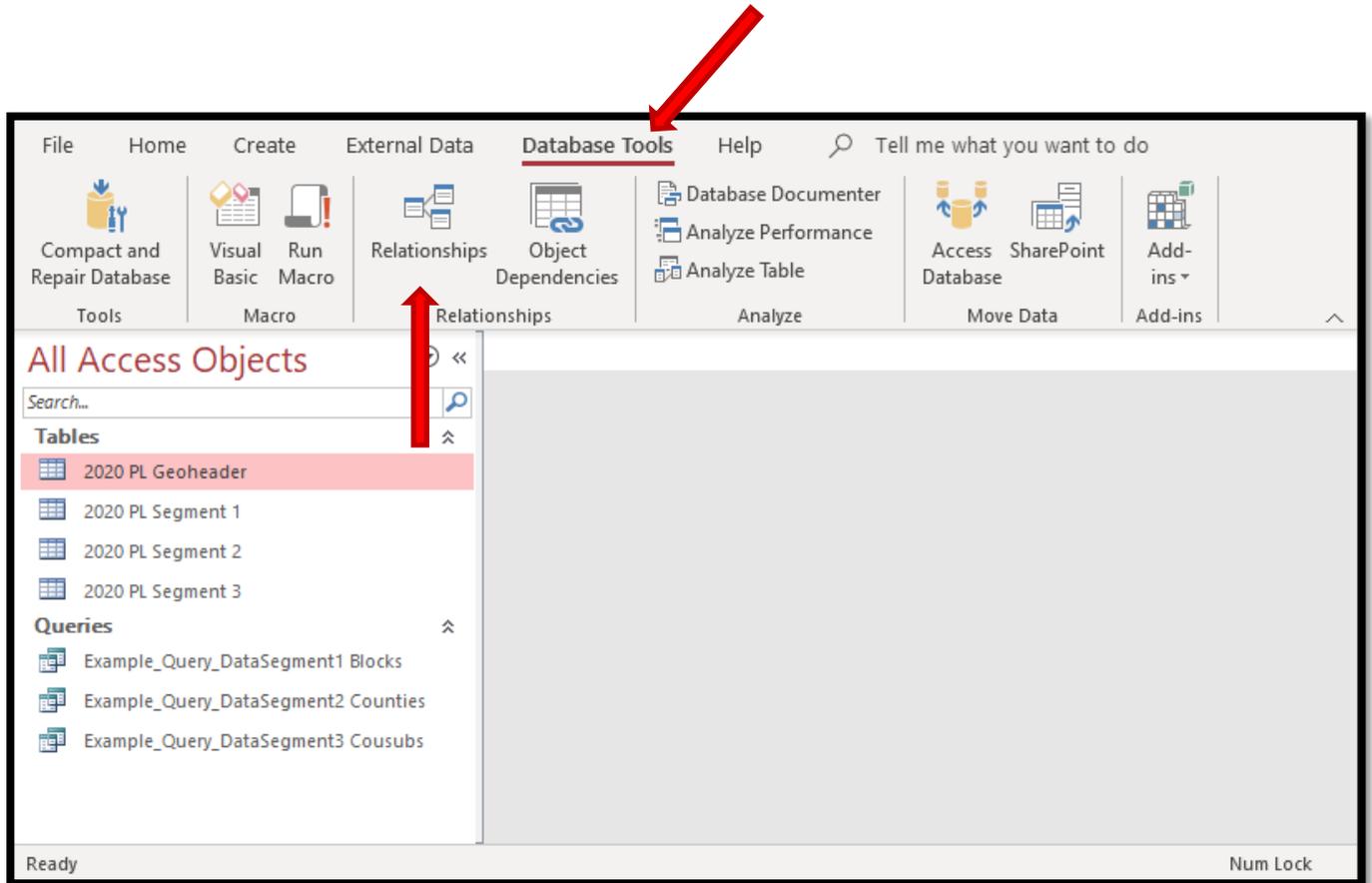
L.) When the import completes, click the “Close” button to close the Import Wizard. Users can select to save their import steps at this stage, but it is not required. NOTE: You can verify that the data was imported by opening the table in the MS Access database. To open the table, double click the table name in the table of contents.

M.) Repeat steps F through L for each file until the Geoheader and all of the data segments are loaded. (Please remember that when working with a large state like **California** or **Texas**, all four files cannot be loaded into a single Microsoft Access database due to a Microsoft Access file size limitation. This limitation can be overcome by loading both the geoheader and each data segment into separate databases, repeating for each data segment to generate three databases, each one containing the geoheader and a single data segment.)

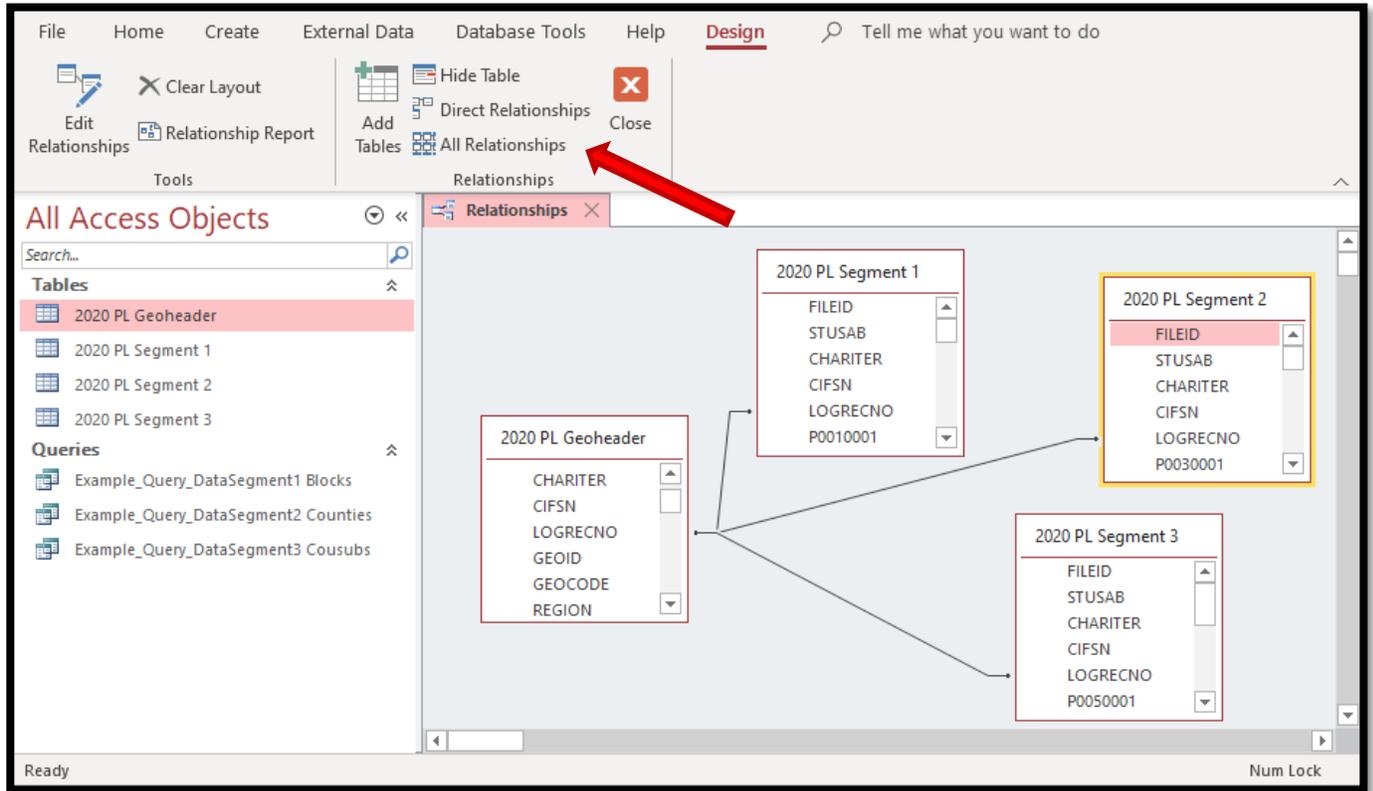
Extracting the Data

The 2020PL_SummaryFile_Shell has the relationship between the tables needed for extracting data already established. In addition, there are three example queries, one for blocks, one for counties, and one for county subdivisions (COUSUBs).

A.) To check these relationships are indeed established, click on the “Database Tools” menu and then the “Relationships” button.



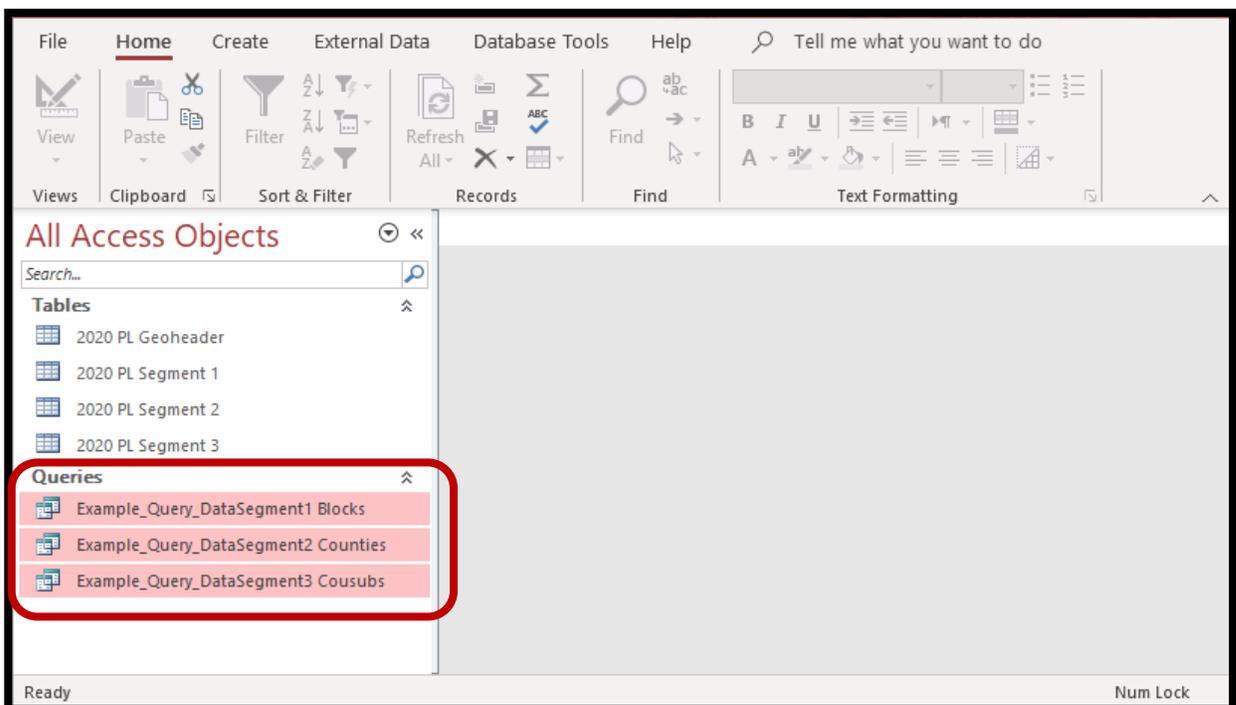
B.) If the relationships don't show right away, click the "All Relationships" button. You should see the tables with connections depicted by black lines between the common field LOGRECNO.



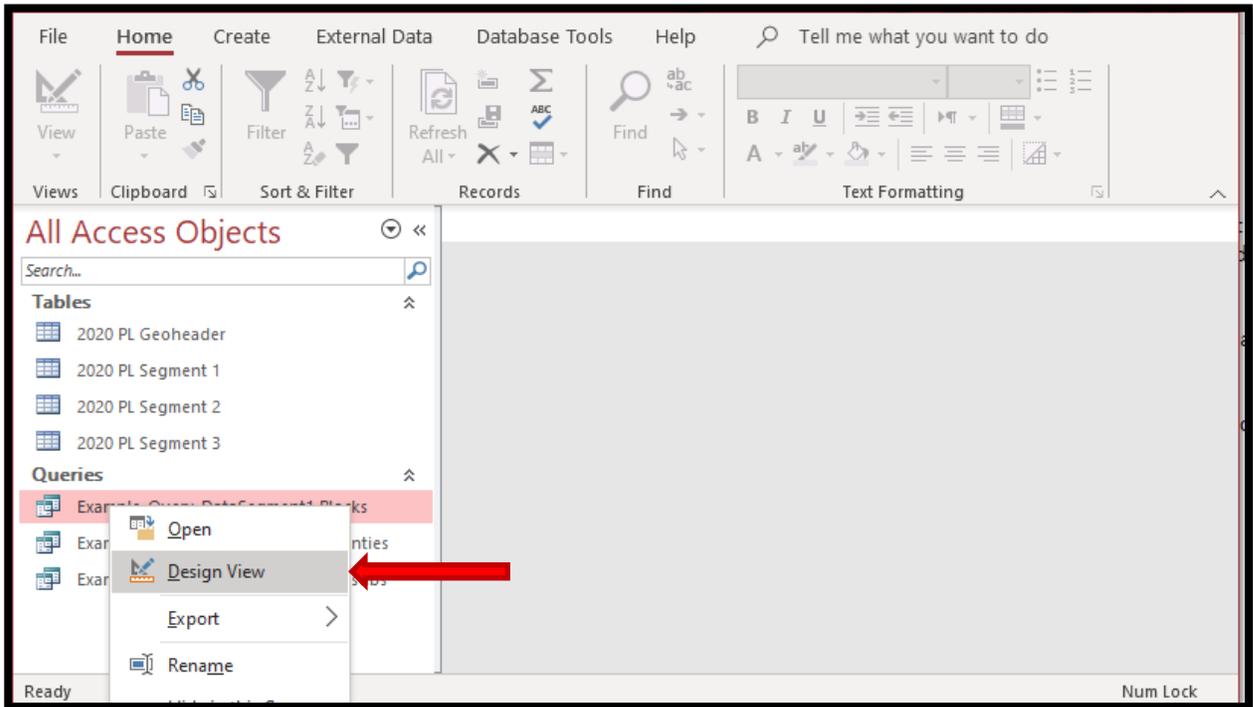
C.) To pull data from the database you just built, it is necessary to construct a query.

The MS Access database shell has three example queries built in.

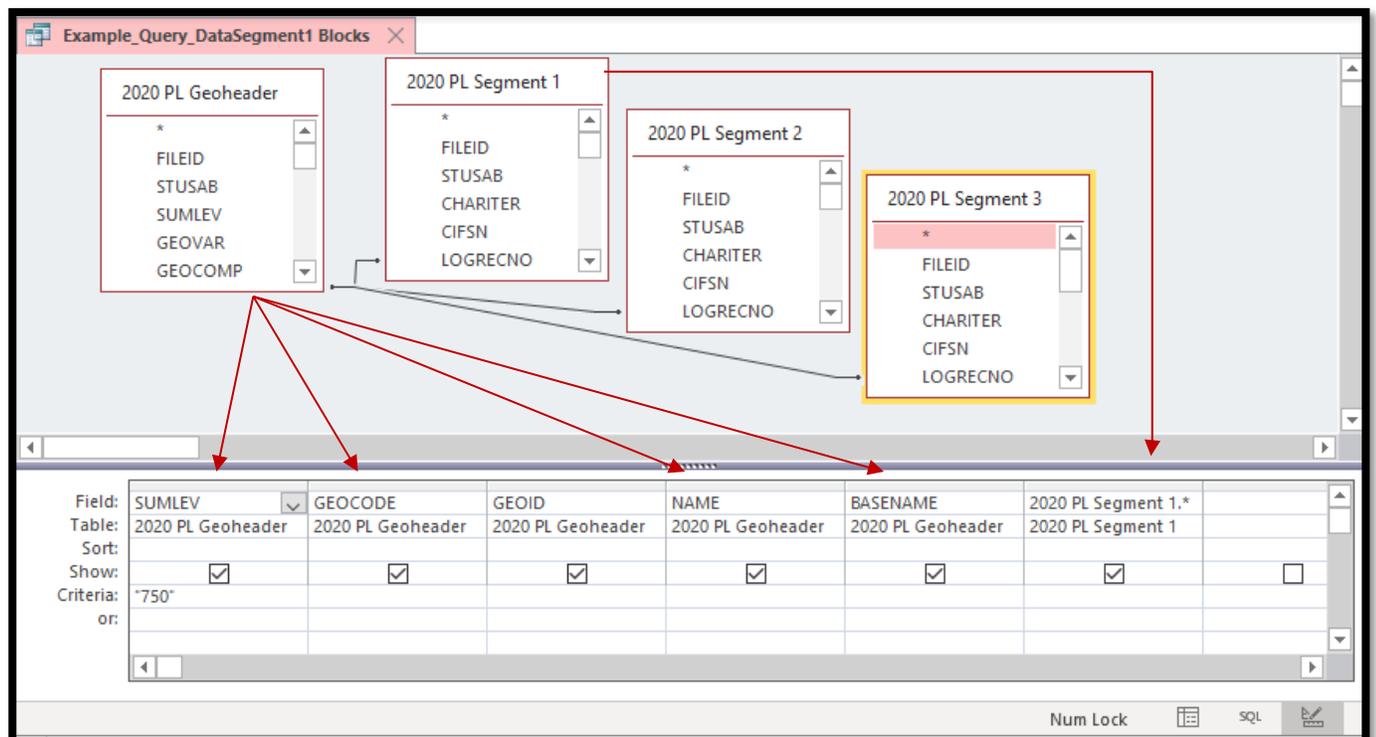
- The first pulls block data from data segment one
- The second pulls county data from data segment two
- The third pulls county subdivision data from data segment three



D.) To modify one of these queries, right click on the query and select Design View. Here you can change the summary level (SUMLEV) to change the geography and/or add additional data fields you want to see in the final extracted table. This example will look at the block query.



E.) In the Design View window, you will see the tables in the area at the top and then the selected fields from those tables in the fields below. To add fields to the query, double click the desired field in the table in the top window.



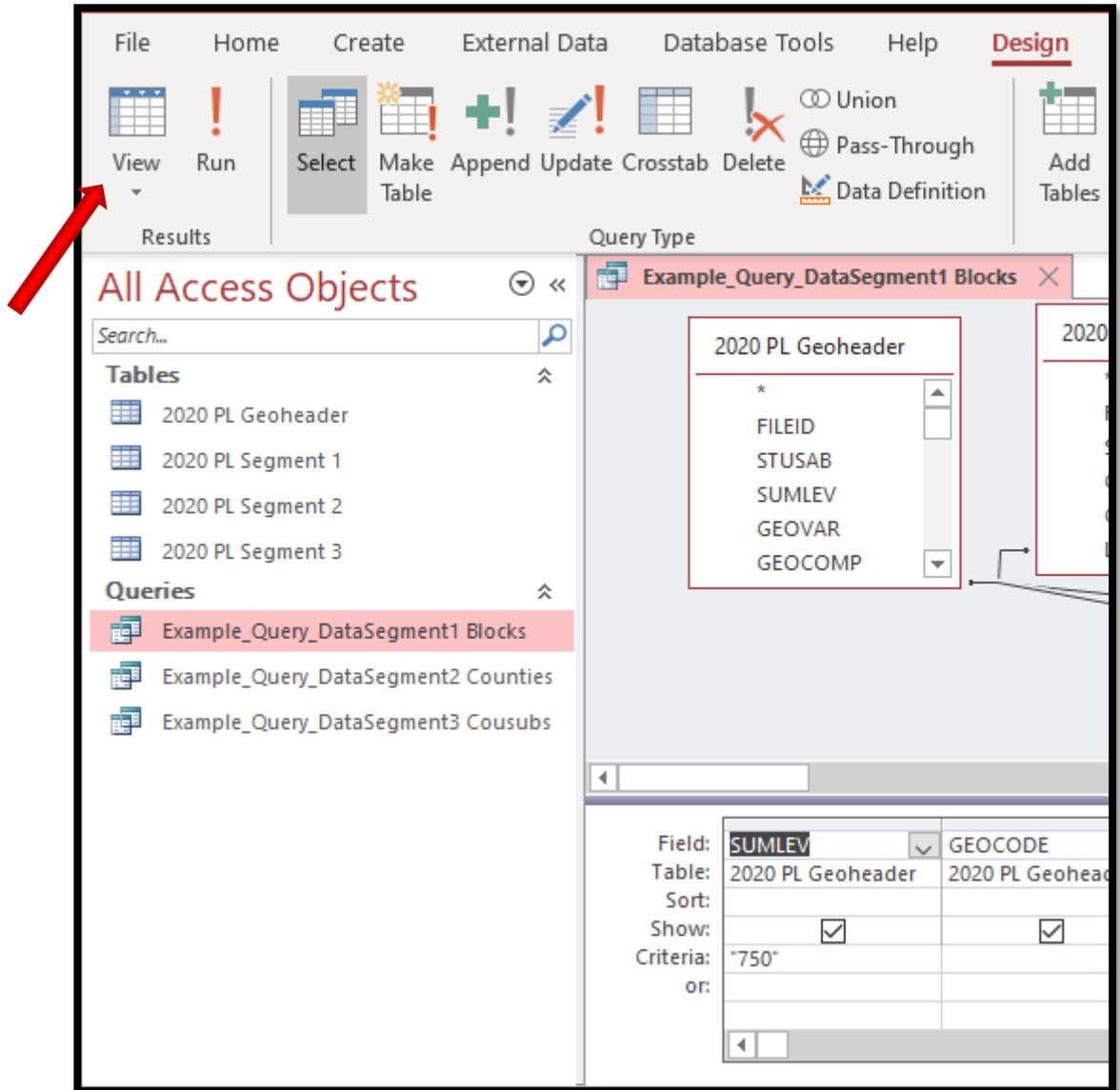
F.) In this query we selected block level data by selecting the summary level field, "SUMLEV", from the geoheader and then setting its value to 750, the code for blocks. These summary level codes can be pulled from the summary level sequence chart in the technical documentation.

The screenshot shows a query builder interface with the following components:

- Tables:** 2020 PL Geoheader, 2020 PL Segment 1, 2020 PL Segment 2, and 2020 PL Segment 3.
- Fields:** FILEID, STUSAB, SUMLEV, GEOVAR, GEOCOMP (from Geoheader); FILEID, STUSAB, CHARITER, CIFSN, LOGRECNO (from Segments 1, 2, and 3).
- Criteria:** SUMLEV is set to "750".

Field:	2020 PL Geoheader	2020 PL Segment 1	2020 PL Segment 2	2020 PL Segment 3
Table:	2020 PL Geoheader	2020 PL Geoheader	2020 PL Geoheader	2020 PL Geoheader
Sort:				
Show:	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Criteria:	"750"			
or:				

G.) To view the results of your constructed query, click the “View” button.



H.) You should now see a table with all the fields you had added in the design view.

SUMLEV	GEOCODE	GEOID	NAME	BASENAME	FILEID	STUSAB	CHARITER	C	LOGRECNO	P0010001
750	440070001011000	7500000US440070001011000	Block 1000	1000	PLST	RI	000	01	6727	0
750	440070001011001	7500000US440070001011001	Block 1001	1001	PLST	RI	000	01	6728	0
750	440070001011002	7500000US440070001011002	Block 1002	1002	PLST	RI	000	01	6729	0
750	440070001011003	7500000US440070001011003	Block 1003	1003	PLST	RI	000	01	6730	50
750	440070001011004	7500000US440070001011004	Block 1004	1004	PLST	RI	000	01	6731	0
750	440070001011005	7500000US440070001011005	Block 1005	1005	PLST	RI	000	01	6732	0
750	440070001011006	7500000US440070001011006	Block 1006	1006	PLST	RI	000	01	6733	18
750	440070001011007	7500000US440070001011007	Block 1007	1007	PLST	RI	000	01	6734	0
750	440070001011008	7500000US440070001011008	Block 1008	1008	PLST	RI	000	01	6735	86
750	440070001011009	7500000US440070001011009	Block 1009	1009	PLST	RI	000	01	6736	19
750	440070001011010	7500000US440070001011010	Block 1010	1010	PLST	RI	000	01	6737	11
750	440070001011011	7500000US440070001011011	Block 1011	1011	PLST	RI	000	01	6738	194
750	440070001011012	7500000US440070001011012	Block 1012	1012	PLST	RI	000	01	6739	38
750	440070001011013	7500000US440070001011013	Block 1013	1013	PLST	RI	000	01	6740	78
750	440070001011014	7500000US440070001011014	Block 1014	1014	PLST	RI	000	01	6741	214
750	440070001011015	7500000US440070001011015	Block 1015	1015	PLST	RI	000	01	6742	19
750	440070001011016	7500000US440070001011016	Block 1016	1016	PLST	RI	000	01	6743	34
750	440070001011017	7500000US440070001011017	Block 1017	1017	PLST	RI	000	01	6744	8
750	440070001011018	7500000US440070001011018	Block 1018	1018	PLST	RI	000	01	6745	513

I.) You can now save and close your query for future use or go back into Design View to adjust what fields and summary level you want in your extract.

J.) If you want to export the table you created, right click on name of the query you saved and select "Export". This will provide many format options for your exported table. Note: Some formats may not be compatible with the size of the table you are exporting.

