

## Introduction

This document describes my workflow to create an Excel file from the 2020 Privacy Protected Microdata Files (PPMF). It accompanies an example Excel file

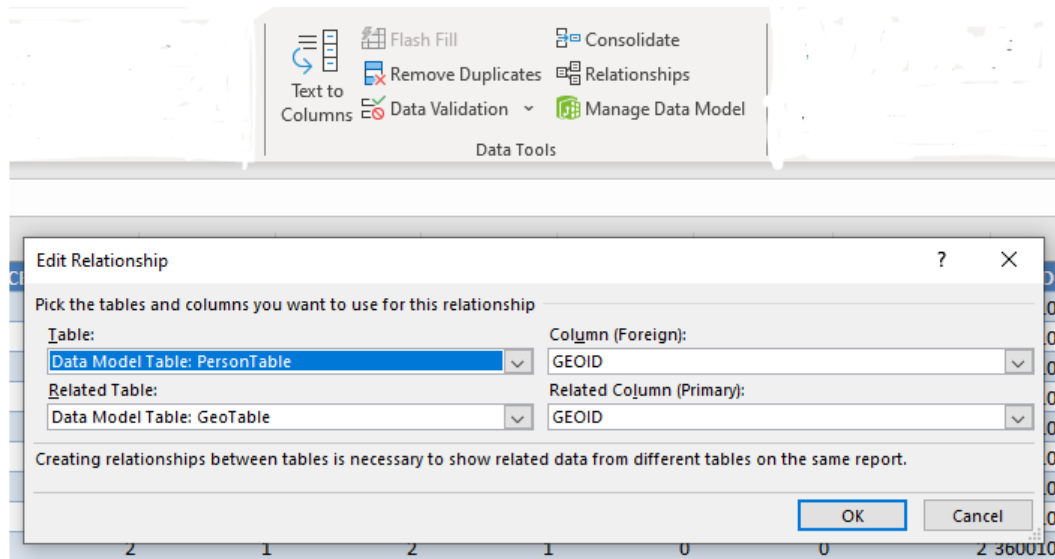
[https://pad.human.cornell.edu/census2020/downloads/PPMF2020\\_36001\\_Per\\_geo.xlsx](https://pad.human.cornell.edu/census2020/downloads/PPMF2020_36001_Per_geo.xlsx) that can be used to experiment with the PPMF data.

## Creation of PPMF Excel file

1. Download the file from <https://www2.census.gov/programs-surveys/decennial/2020/data/privacy-protected-microdata-file>
2. Unzip the person file (120GB)
3. **SAS**
  - a. Import the .csv file, the file has headers in the first row and I used them to define the columns
  - b. I added a GEOID file by concatenating State, County, Tract and Block codes. My import process read those geographic codes as integers, so in the concatenating I added leading zeroes to get a two digit state, a three digit county, 6 digit tract and 4 digit block code
  - c. The PPMF file contains all block assignment information for each individual and thus repeating the same block assignment information to all persons in a block. To reduce the size of the output, I put the person statistics plus the GEOID in one data table, and the block assignment data in another. The GEOID is the field they have in common. I unduplicated block assignment table.
  - d. I exported the person table and the block assignment table to two different Excel sheets.

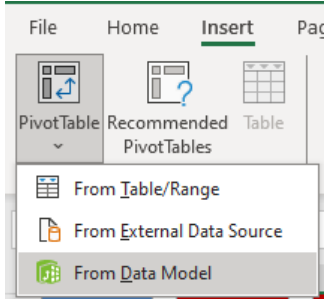
#### 4. Excel

- a. Format the data as tables [“Format as Table”] and give them better names under [“Table Design”]
- b. Create a relationship between the tables. This option can be found under [“Data Tools”] section of the [“Data”] menu



This creates a one-to-many relation between the geotable and the person table

- c. I inserted a few columns in my Excel person table. The column [Count] is used to ease counting the number of persons of certain characteristics. The column [Count04] is an example of how to count only people with age  $\leq 4$ . If you define the formula for the first row, it automatically propagates down the whole column. The [RaceRecode] columns is an example of recoding the [CENRACE] variable and giving a code 7 to all multi-race codes.
- d. Creating a pivot table by inserting a pivot table from Data Model



This allows to select PivotTable fields from both tables

## 5. Pivot Table

- a. Example Age/Sex distribution of people living alone in the City of Albany, NY

The screenshot shows a PivotTable configuration interface with four main sections: Filters, Columns, Rows, and Values. In the Filters section, 'COUSUBFP' and 'LIVE\_ALONE' are selected. In the Columns section, 'QSEX' is selected. In the Rows section, 'QAGE' is selected. In the Values section, 'Sum of Count' is selected.

I added a County Subdivisions filter (Cities are places and also county subdivisions in New York) from the GeoTable and a Live\_Alone filter from the person table. Sex will be displayed in the columns, age in the rows. The values in the table is the sum of my added column [Count] which is equal to one for every person.

- b. I know that the City of Albany has FIPS code 01000, but in the process leading zeroes have gone missing, so I select 1000 as my filter. From the included codebook I see that persons that live alone have code 1.
- c. After setting the filters, my result is:

	A	B	C	D
1	COUSUBFP	1000		
2	LIVE_ALONE	1		
3				
4	Sum of Count	QSEX		
5	QAGE	1	2	Grand Total
6	17	2	2	4
7	18	2	11	13
8	19	28	25	53
9	20	47	63	110
10	21	68	120	188
11	22	103	113	216
12	23	131	178	309
13	24	191	210	401
14	25	184	198	382
15	26	217	133	350
16	27	228	196	424
17	28	190	215	405
18	29	112	174	286

and I see that there were two 17 yr old males and two females living alone in the city.

Tip: different layouts under ["Design" -> "Report Layout"] can make your table more readable.