

Joining Data from Two Tables in Excel

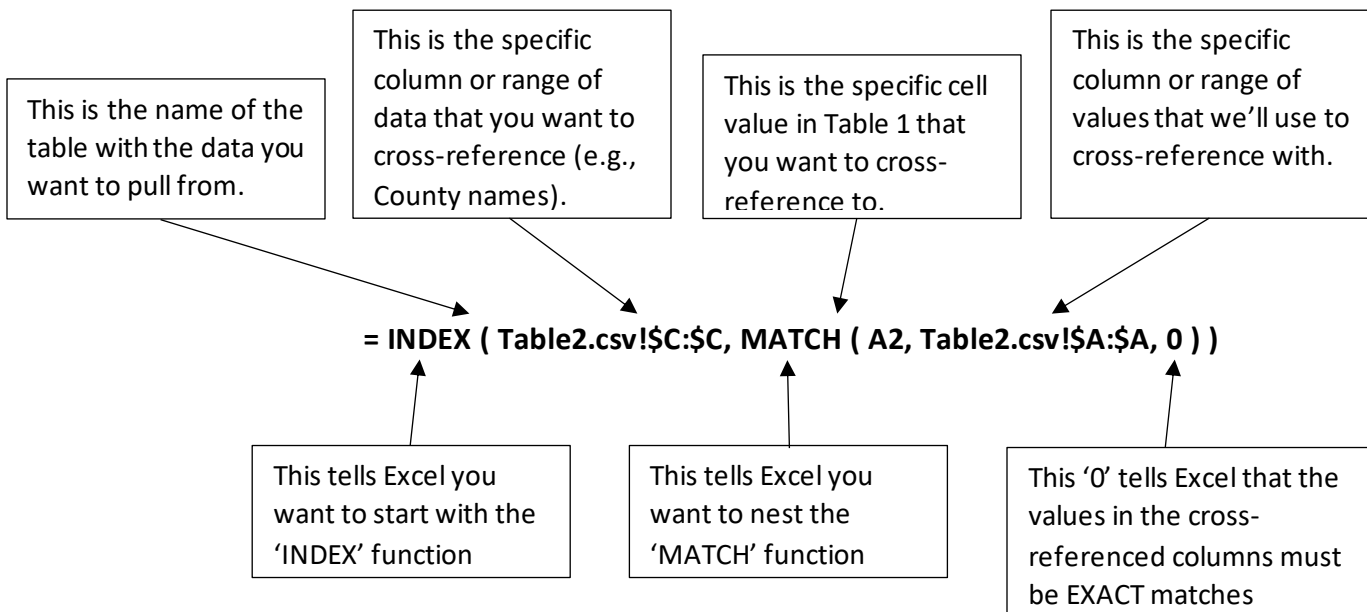
Sometimes you have data in a pair of independent tables that you would like to cross-reference. For example, perhaps you have a table of data of standardized testing scores for school children, and another table that lists where each student lives along with some basic household data on income, parents' level of education, etc. If the names of the students are listed in both tables, you might want to cross-reference the household data to the test scores data, to do some analysis on household factors that may influence testing scores. . .you get the idea. If you just want to pull a few items (i.e., columns) from one table and cross-reference them to the other, it is sometimes faster to use a nested INDEX + MATCH function in Excel than to go through a true data management program like R. Here's how to do it in Excel.

For this exercise, you were given 2 tables (.csv files). Table1 lists a variety of socio-economic variables for counties throughout the southeastern US. Each county is identified by its "FIPS" (Federal Information Processing Standard) code. Table2 lists all US counties by State and County Name. It also includes latitude and longitude coordinates for the county centroids (center points), as well as the FIPS code.

To quickly cross-reference a column from Table2 to Table1, we will use the Excel INDEX + MATCH function as follows:

First, insert a new "CountyName" column next to the "FIPS" column of Table1.

Next, select the first blank entry of the new "CountyName" column (cell B2) and type in the following:



Duplicate letters separated by a colon (e.g., "C:C") in Excel indicate that you want to select the entire column. Finally, use the FILL-DOWN function to fill the cross-referenced CountyName values.

NOTE: [To do this same table join in R](#), load Table1 and Table2, then run the following command:

Table3 <- merge (x = Table1, y = Table2, by = 'FIPS')