

Example 1 - Joining datasets by a common variable: Creating a single table using multiple datasets

Other features illustrated: Aggregate data multi-variable recode, computational calculation

Background: Some datasets may have variables whose value sets are defined alike (comparable). Examples include demographic variables (e.g. gender, race, marital status) and geographic variables (e.g. state code, county code). Co-tabulation allows you to tabulate information from the different datasets side by side using the comparable variable as a hinge.

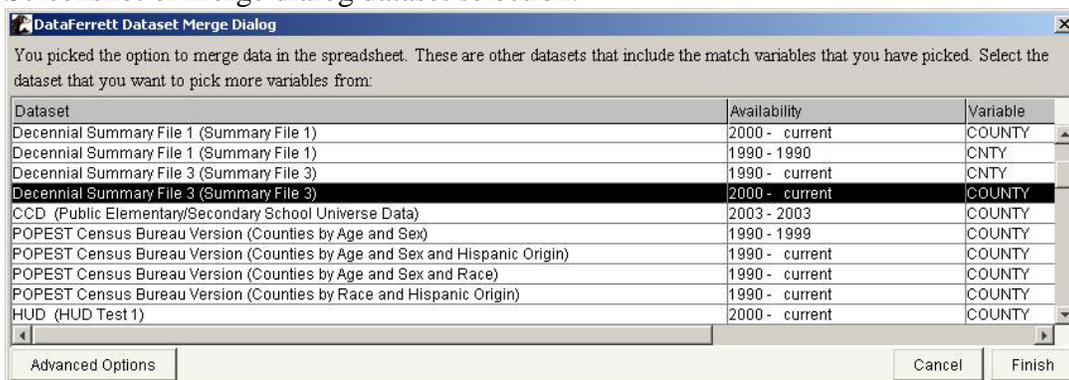
Goals: 1) Estimate the number of people per gas station in each county in Maryland. 2) Add the number of households without vehicles to the table.

Steps:

Goal 1 Estimate the number of people per gas station in each county in Maryland.

1. Select first dataset - County Business Patterns, County Level, 2003.
2. Select variables - Selectable Geographies (all counties in MD), Establishments (est), and industry (naics=447///).
3. Select second dataset - Decennial Census, Summary File 3, 2000.
 - o Go to Step 2: DataBasket/Download/Make a table tab.
 - o Highlight the county variable.
 - o Press the "Merge Datasets" button on the right-hand side.
 - o The merge dialog appears with information that you will only be able to create tables and not extract data. Press the "Next" button. This will bring back a list of datasets with a comparable variable and may take a while to populate.
 - o After the results return, enlarge the window by dragging the right side farther to the right.
 - o Scroll down the list looking for the name of the second dataset named in step 3. Highlight this dataset and press the "Finish" button. You will get a warning that you will be taken back to step 1 where you will select variables from the second dataset.

Screenshot of merge dialog dataset selection:



4. Select the variable from Summary File 3 - P1. Total Population (P001001).
5. Go to Step 2 tab and press the "Make a Table" button. This brings up the DataFerrett spreadsheet window.

6. Click and drag the first "MERGED-200 Three digit FIPS County Number" variable into C1,R1 to define the counties in the rows. This is the comparable dimension and all other variables will be dropped in the columns.
7. Click and drag the "est" variable into C2,R1.
8. Click and drag the "P001001" variable into C3,R1.
9. Create a formula calculating people per gas station in column 4 (C4)
 - o Highlight column 4 by clicking in the gray column header.
 - o In the formula bar above the column headers enter the formula - **=comp(c3/c2)** and press the Enter key.
 - o Add a description to the column header: "People per gas station."

Screenshot of co-tabulation table and a formula:

| | C1 | C2 | C3 | C4 |
|-----|----------------------------------|--------------------------------|------------------|------------------------|
| R1 | | Total Number of Establishments | Total population | People per gas station |
| R2 | Total MERGED-202 | ? | ? | ? |
| R3 | Allegany County, Maryland | ? | ? | ? |
| R4 | Anne Arundel County, Maryland | ? | ? | ? |
| R5 | Baltimore County, Maryland | ? | ? | ? |
| R6 | Calvert County, Maryland | ? | ? | ? |
| R7 | Caroline County, Maryland | ? | ? | ? |
| R8 | Carroll County, Maryland | ? | ? | ? |
| R9 | Cecil County, Maryland | ? | ? | ? |
| R10 | Charles County, Maryland | ? | ? | ? |
| R11 | Dorchester County, Maryland | ? | ? | ? |
| R12 | Frederick County, Maryland | ? | ? | ? |
| R13 | Garrett County, Maryland | ? | ? | ? |
| R14 | Harford County, Maryland | ? | ? | ? |
| R15 | Howard County, Maryland | ? | ? | ? |
| R16 | Kent County, Maryland | ? | ? | ? |
| R17 | Montgomery County, Maryland | ? | ? | ? |
| R18 | Prince George's County, Maryland | ? | ? | ? |
| R19 | Queen Anne's County, Maryland | ? | ? | ? |
| R20 | St. Mary's County, Maryland | ? | ? | ? |
| R21 | Somerset County, Maryland | ? | ? | ? |
| R22 | Talbot County, Maryland | ? | ? | ? |
| R23 | Washington County, Maryland | ? | ? | ? |
| R24 | Wicomico County, Maryland | ? | ? | ? |
| R25 | Worcester County, Maryland | ? | ? | ? |
| R26 | Baltimore city, Maryland | ? | ? | ? |
| R27 | | | | |

10. Press the green "Go Get Data" button in the toolbar.

Goal 2 Add the number of households without vehicles to the table.

At this point, we can explore how many households have no vehicle within each county.

1. DO NOT CLOSE the spreadsheet window, but go back to your main DataFerrett window using your Windows taskbar.
2. Go to the Step 1 tab so that you can add an additional variable from SF3.
3. Scroll down the dataset list under SF3 until you find table "H44. Tenure by vehicles available", and view all of the variables in this table.
4. Select the "D_H044_1 - Total Housing Units: No Vehicle Available" variable. This is a pre-defined recode that sums together the H044003 and H044010 variables.
5. Now go back to the tabulation window from your taskbar.
6. The added variable is there, now drag it into column 5 (C5,R1) and press the "Go Get Data" button.
7. If this summed variable had not already been defined, we could have created it ourselves.

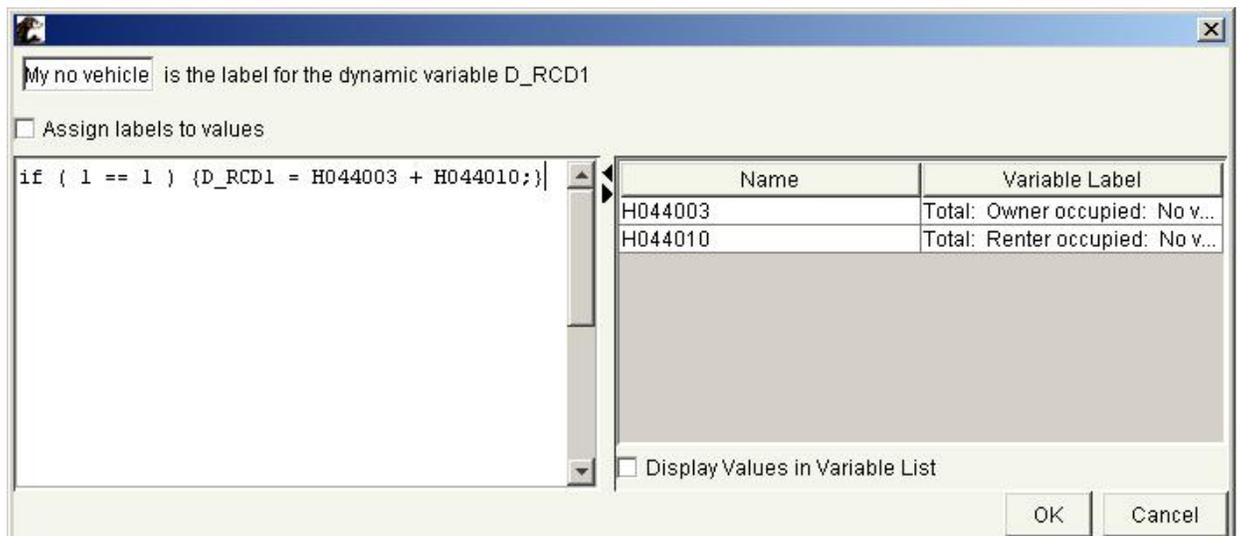
- o DO NOT CLOSE the spreadsheet window, but go to your main DataFerrett window using your Windows taskbar.
- o Select the H044003 and H044010 variables.
- o Go to the Step 2 tab and highlight the H044003 and H044010 variables, then pre the "Create Multi-variable Data Step" button on the right-hand side. This brings up the multi-variable recode dialog window.
- o Make sure you give your new variable a good label in the text field at the upper left.
- o Enter the formula `if (1 == 1) {D_RCD1 = H044003 + H044010;}` in the text area as shown below.

Note: the new variable name in the formula, D_RCD1, is the dynamic variable name shown at the top of the window and your formula must use the name shown.

The "if (1 == 1)" syntax is used to force the condition to be true in all cases.

The syntax of the formula is like programmatic code and must follow this: `if (condition) { action ;}`

Screenshot of multi-variable recode of aggregate variables:



- o Press the "OK" button to save the created variable.
8. Now go back to the tabulation window from your taskbar.
 9. The newly created variable is there, now drag it into column 6 (C6,R1) and press the "Go Get Data" button.
- Note: The numbers in columns 6 and 5 should be identical.