

Distr.  
GENERAL

CES/SEM.40/24  
22 October 1998

ENGLISH ONLY

STATISTICAL COMMISSION and  
ECONOMIC COMMISSION FOR EUROPE

STATISTICAL OFFICE OF THE  
EUROPEAN COMMUNITIES (EUROSTAT)

CONFERENCE OF EUROPEAN STATISTICIANS

Joint ECE-EUROSTAT work session  
on Population and Housing Censuses  
(Dublin, Ireland, 9-11 November 1998)

Study topic 2

#### IMPS CROSSTAB AND MAP VIEWER FOR DATA DISSEMINATION

Supporting paper submitted by the U.S. Bureau of the Census<sup>2</sup>

#### I. CrossTab

1. Figure 1, below, shows the CrossTab main screen. The user has opened a data dictionary named "popstan.ddw." This is a description of a census data file for an imaginary country, including the names and valid values of all the variables. The screen shows that the user has specified an Age by Sex table for married people, to be produced by province.  
Figure 1.

---

<sup>1</sup> The papers which are prepared for this work session will be treated in the same manner, as papers that are prepared for seminars.

<sup>2</sup> Prepared by Glenn Ferri. This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more-limited review than official Census Bureau publications. This report is released to inform interested parties of research and to encourage discussion.

### Tabulating variables

2. Note the box labeled "Tabulate." This is where the user specifies the row items, column items, value and weight. In the example in Figure 1, the user has selected "Age in 5 year groups" as the row item and "Sex" as the column item. The user may specify one or two items (variables) to appear in the table rows, and one or two items to appear in the columns. If two items are selected for either dimension, their values are nested. For example, if

IMPS 4.1 Cross Tabulation - [Spec file = Untitled / Data dict = popstan.ddw]

File Edit View Options Tools Window Help

Add Insert Delete Table 1. Age in 5 year groups by Sex for Marital Status = Married

Title  
Table 1. Age in 5 year groups by Sex for Marital Status = Married ☐ Lock

Universe Marital Status = Married

Area Province

Tabulate

Row Items	Column Items
Age in 5 year groups <input type="button" value="Item"/>	Sex <input type="button" value="Item"/>
<input type="text"/> <input type="button" value="Item"/>	<input type="text"/> <input type="button" value="Item"/>
Value: <input type="text"/> <input type="button" value="Item"/>	Weight: <input type="text"/> <input type="button" value="Item"/>

Percent  
☐ None ☐ Total ☐ Row ☒ Column ☐ Percents only
 ☒ Include undefined row and column

Ready

the row items are "Sex" and "Urban/Rural," the table would show urban males, urban females, rural males, and rural females, in addition to totals. If the user selects only one row item and no column items, the system will produce a frequency distribution, including cumulatives.

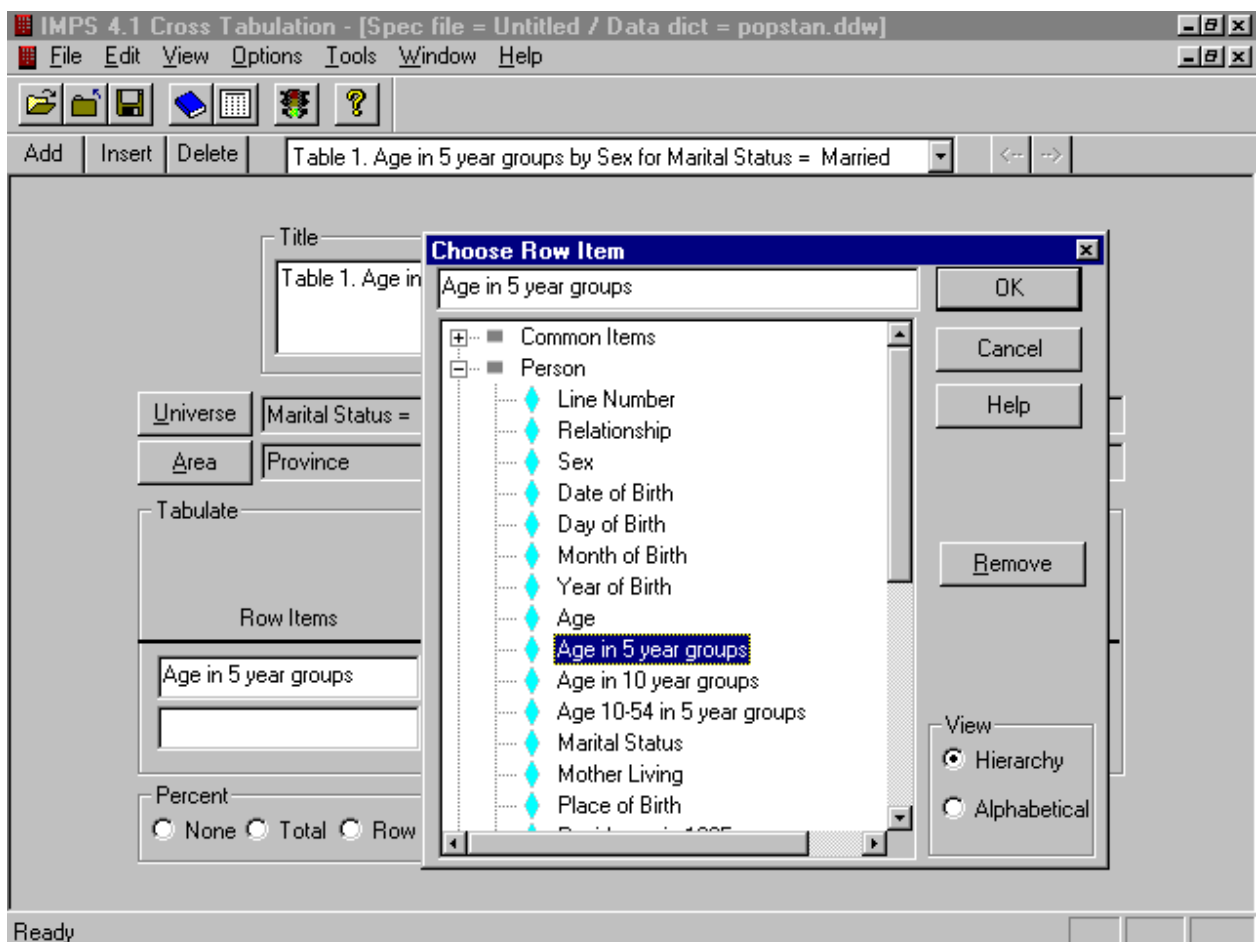
3. The "Value" item determines what value is tabulated into the corresponding row and column for a given data record. If no value item is specified, the system will increment the proper table cell by one each time, producing simple counts. If a value item is specified, the system will increment the table cell by the magnitude of the value item. For example, a value item might be "income," if the user wants to tabulate total incomes for each table cell. The "Weight" item works similarly. If there is a weight item specified, the system will multiple the value to tally by the weight each time.

4. To specify a row item, column item, value or weight, the user simply clicks on the button marked "Item" next to the appropriate slot on the screen. Figure 2, below, shows the dialog box which appears when the user clicks on the row item button.

Figure 2.

5. The hierarchical tree in the dialog box shows the variables in the data file, grouped by record type. The user may also see the variables in alphabetical order, or may simply type the first few letters of the variable name into the slot at the top of the dialog box. The row and column items of the cross tabulation may be variables from the same record type or from different record types.

*Percents*



6. In Figure 1, the box labeled "Percent," under the "Tabulate" box, allows the user to select among several methods of computing percents. In each case (except "None"), the system will add an extra percent column next to each existing column in the table. In the case of "Total," the percentages will be calculated using the grand total of the table as the denominator. In the case of "Row," the percentages will total to 100

horizontally; in the case of "Column," the percentages will total to 100 vertically. If the "Percents only" box is checked, only the percent columns will appear in the table.

7. When the "Include undefined row and column" box is checked, the system will add a row at the bottom and a column at the right for values found in the data which do not match the valid values for that variable in the data dictionary. For example, any values in the data for the "Sex" field not corresponding to "male" or "female" would appear in the "Undefined" column.

#### *Table title*

8. Toward the top of the screen in Figure 1 lies the "Title" box. CrossTab automatically generates a title for the table based on the user selections. The user may change this title and check the "Lock" box to stop the system from generating a new title in the event of a change to the table specification.

#### *Universe*

9. A very powerful yet simple feature of CrossTab is the ability to define a universe (filter). A universe is a set of conditions that restricts the tabulations. Figure 3 shows the dialog box for defining the universe. In this example, only married people are included in the tabulation. Other examples of possible universes are "women between the ages of 15 and 49," "households with electricity," and "household income > 10,000." The Universe dialog box leads the user through the definition one step at a time. In the example, the user first saw the Items (variables) dialog box, as in Figure 2, and chose "Marital Status;" then chose "=" from among the relations =, <>, >, >=, <, and <=, then chose "Married" from among the possible values for "Marital Status." The user may also type in the explicit values rather than choose among the predefined values. In this example, values "2,3" would mean "divorced or separated." Typing in "0:14,65:99" for age values would mean "between 0 and 14 or between 65 and 99."

10. A universe may be a complex expression. "Women between the ages of 15 and 49" would be expressed as two lines in the Universe dialog box: "sex = female" and "age = 15:49." By using "And," "Or," and the parentheses, the user may specify a universe of virtually unlimited complexity.

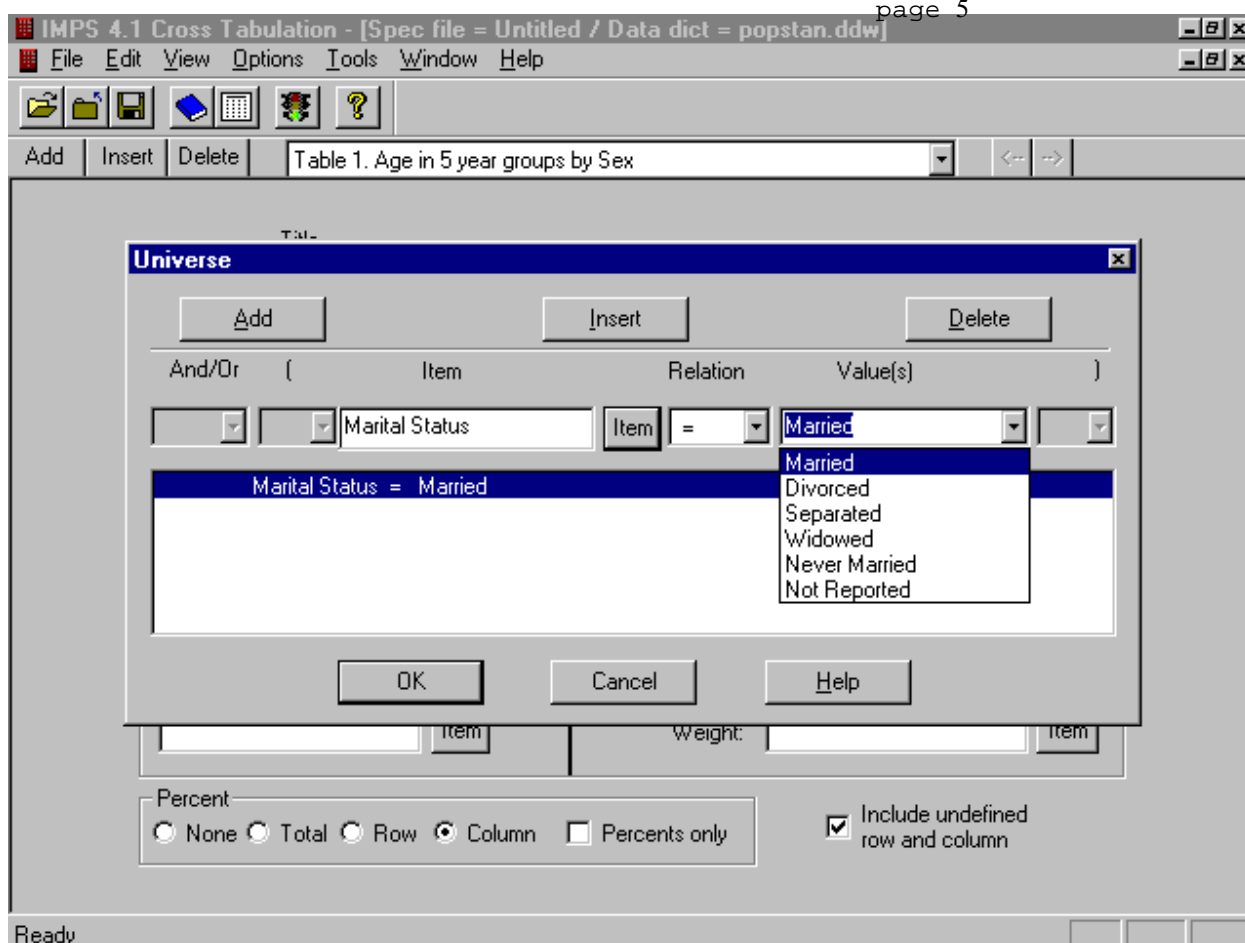


Figure 3.

### *Area Processing*

11. CrossTab offers a very powerful area processing capability. This means a given table can be produced at lower levels of geography and "rolled up" automatically to higher levels. The "Area" button allows the user to select at which geographic level he would like to produce the specified table. In our example, the country is divided into provinces, which are divided into districts. The user has chosen "Province" as the area at which to tabulate. This means that a separate age by sex table will be produced for each province in the data, as well as for the entire (national) data set.

### *Multiple tables*

12. CrossTab allows the user to produce a virtually unlimited number of tables in a single pass through the data file. Below the tool bar in Figure 1 are three buttons to Add, Delete, and Insert tables. To the right of these buttons is a drop box containing the titles of the tables that have been defined. The user may select the table to work on from this list, or use the arrow keys to move to the previous and next tables. While each table has its own Universe specification, the area processing specification applies to all tables.

# *Running, viewing and printing the tables*

13. Once the tables have been specified, the user simply clicks on the green traffic light "Go" button to run the tabulations. A dialog box appears prompting for the name of the data file to tabulate. The user may select several data files and the system will treat them as if they were a single concatenated file. CrossTab tabulates very quickly. The system has been carefully tuned to maximize speed for raw ASCII files. When the tabulations are complete, the Table Viewer shows the results, as in Figure 4, below.

The screenshot shows the IMPS 4.1 Table Viewer window. The title bar reads "IMPS 4.1 Table Viewer - [Cross01.tbw (popstan.dat)]". The menu bar includes File, Edit, View, Options, Window, and Help. The toolbar contains icons for back, forward, home, save, print, zoom, and help. The title field shows "Table 1. Age in 5 year groups by Sex for Marital Status = Married". The table content is as follows:

Age in 5 year groups	Total		Male		Female		Undefined	
	Total	Percent	Total	Percent	Total	Percent	Total	Percent
Popstan								
Total	6,985	100.0	3,439	100.0	3,546	100.0	-	-
0 to 4 years	-	-	-	-	-	-	-	-
5 to 9 years	-	-	-	-	-	-	-	-
10 to 14 years	12	0.2	5	0.1	7	0.2	-	-
15 to 19 years	258	3.7	41	1.2	217	6.1	-	-
20 to 24 years	836	12.0	279	8.1	557	15.7	-	-
25 to 29 years	1,089	15.6	476	13.8	613	17.3	-	-
30 to 34 years	1,135	16.2	588	17.1	547	15.4	-	-
35 to 39 years	1,042	14.9	513	14.9	529	14.9	-	-
40 to 44 years	774	11.1	431	12.5	343	9.7	-	-
45 to 49 years	509	7.3	278	8.1	231	6.5	-	-
50 to 54 years	439	6.3	263	7.6	176	5.0	-	-
55 to 59 years	332	4.8	193	5.6	139	3.9	-	-

At the bottom of the window, it says "For Help, press F1".

Figure 4.

14. Table Viewer scrolls, if necessary, while keeping the table headings and stubs frozen. The user may choose any font to apply to the table. If more than one table was specified, the user may select which table to view from the drop-box under the tool bar. Table Viewer can print tables and automatically paginates intelligently if the table does not fit on one piece of paper. It includes a Print Preview facility to see how the tables will look on paper before actually printing.

*Using tables in other software packages*

15. The user can save a table, or set of tables, to a disk file. The file is ASCII, but in a format native to IMPS. This means only the Table Viewer can retrieve the tables. However, Table Viewer has very powerful clipboard functions (copy and paste) to move all or part of a table to other packages.

16. The user can select (highlight) an entire table, or a subset of it, such as a group of rows, columns or cells. When the user clicks on File/Copy (or Ctrl-C), the table piece goes to the clipboard in both tab-delimited format and Rich Text Format (RTF). Tab-delimited format means one can paste the table piece into any commercial spreadsheet package, and the table cells will become spreadsheet cells. Rich Text Format means one can paste the table into a word processor and take advantage of any "table" facility it may have. (Both Microsoft Word and WordPerfect have such facilities.)

17. The user can also see selected table cells as a thematic map in the Map Viewer with a couple of mouse clicks. This feature is described in the section "Making thematic maps from tables," presented after a discussion of the Map Viewer features.

## II. Map Viewer

Figure 5, below, shows the Map Viewer main screen. The user has opened a map data file named "Madag.mdf." This is a collection of variables whose data points can be made into thematic maps.

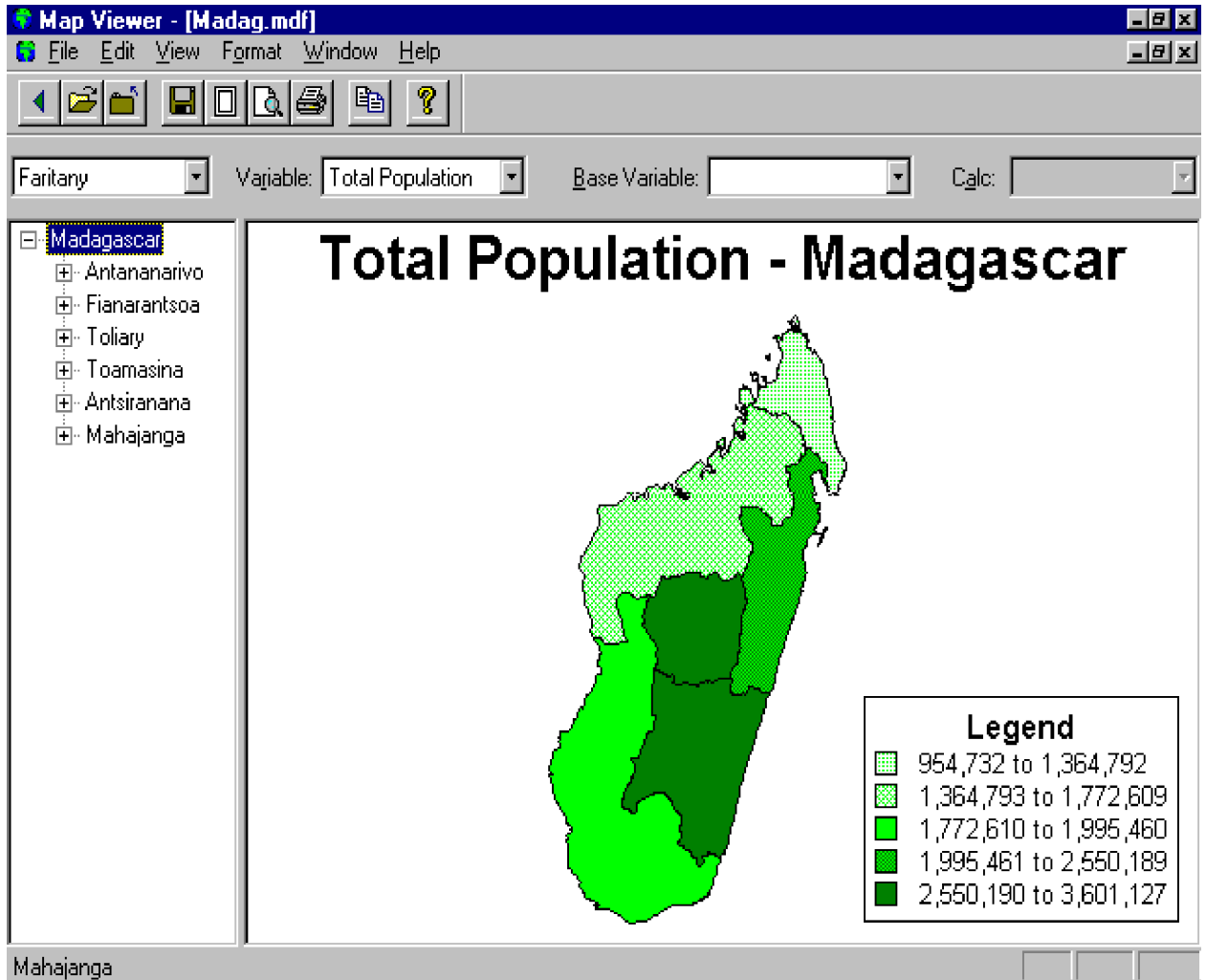


Figure 5.

### *Mapping variables*

18. The example above shows a thematic map of the "Total Population" variable. A map data file (.MDF) may contain a virtually unlimited number of variables. These are typically summary data. The drop box labeled "Variable:" above the map contains a list of the variables available in the map data file.

19. The user may perform some simple calculations on any two variables in the set and show a thematic map of the result. This is an easy way to show



changes over time periods or to show certain ratios, for example. To do so, the user also selects from the "Base Variable:" and "Calc" drop box. There are four possible calculations:

Percent change	= (variable - base variable) / base variable
Percent	= (variable / base variable) * 100
Ratio	= (variable / base variable)
Difference	= (variable - base variable)

20. The user may see the data value for any particular geographic area by simply right-clicking on the map itself.

#### *Levels of geography*

21. Map Viewer can contain several layers of geography, depending on available memory and speed of the microcomputer. It maintains the layers in the form of a hierarchy tree, as shown on the left side of the screen in Figure 5. The user can open and close nodes on the tree to see the entire hierarchy, and click on any name to see a map of the current variable (or calculated variable) for that geographic unit. The user can also simply click on an area of the map itself to "drill down" to a thematic map of just that geographical unit, with appropriate legend.

22. By default, the thematic maps show different colors for polygons at the level of geography just below the level of the bounding map. For example, if a country has provinces composed of districts, the map of the country would show the provinces and a map of any province would show its districts. However, Map Viewer allows the user to show the polygons for any lower level of geography, so that in the above example the user could choose to show either provinces or districts when viewing the national map. This is accomplished by choosing the level from the drop box just above the hierarchical tree on the left.

#### *Map objects - polygons, title and legend*

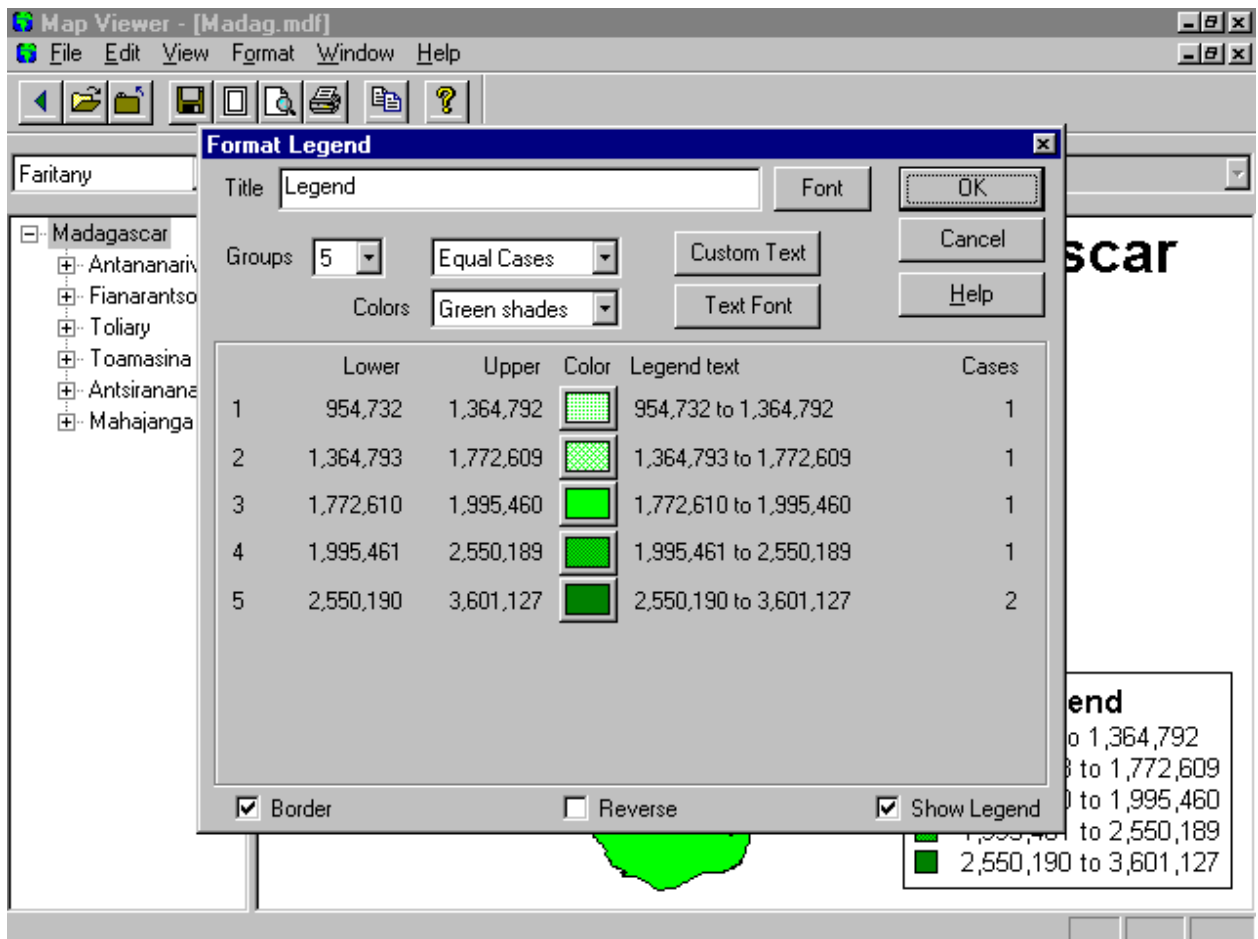
23. A thematic map consists of the map (colored polygons), the title and the legend. The user can reposition these elements. By default, the title appears at the top and the legend appears at the bottom right, as in Figure 5. The user can move the title or legend simply by clicking on it, dragging, and dropping it into the desired location.

24. The user can right click on the title to change it. One can then modify the text itself and/or modify the title's font by way of the standard Windows Font dialog box.

25. The user can modify the legend specification by right-clicking on the legend box, as described in the next section.

### Changing the legend format

26. The user has a lot of control over the way the map looks, as shown in the Format Legend dialog box in Figure 6, below.  
Figure 6.



27. The user can choose the number of intervals (colors) in the map by clicking on the "Groups" drop box. There can be as few as only two intervals, or as many as seven.

28. There are three ways to define the intervals. The default option is "Equal Cases." This means Map Viewer will find reasonable lower and upper limits for each interval so that the data points distribute themselves equally among them. In Figure 6, the six data points are spread across the five intervals so that there is one in each and two in the last interval. The second way to define the intervals is "Equal Intervals." This means Map Viewer will calculate the upper and lower limits based on the smallest and largest data point so that each interval is the same size. The third option is "Custom" intervals. This means Map Viewer lets the user define the limits by typing them in. This is commonly used to establish "round numbers" for the intervals. Using "Custom" intervals also freezes the interval

boundaries, so that they do not change when the user selects a different geographic unit to map. When the method is "Equal Cases" or "Equal Intervals," Map Viewer recalculates the interval boundaries each time the user chooses a new geographic unit.

29. The user controls the colors on the map. Map Viewer supplies several color schemes ready to use: blues, greens, reds, yellows, grays and "cold to warm." The user may customize any particular color by way of the standard Windows Color dialog box. The "Reverse" check box, at the bottom center of the dialog box, allows the users to invert the colors. In Figure 6, light shades represent small numbers and dark shades represent large numbers. Checking the Reverse box would make the colors go from dark to light with increasing values.

30. The user also has control over the text and fonts which appear in the Legend. Map Viewer generates a default text for each interval, and for the top line of the legend box, which the user can override.

#### *Using thematic maps in other software packages*

31. Map Viewer has powerful clipboard functions. When the user selects "Edit/Copy" from the main menu, the map image goes to the clipboard in .BMP format and as a Windows Metafile. This means the map can then be pasted into just about any software package that handles graphics. These include word processors, desktop publishing systems and presentation software.

32. Users can easily post a thematic map on the Internet. Map Viewer can save a map in .GIF format, so that it can be included in any HTML page.

#### *Making thematic maps from tables*

33. The user can generate thematic maps directly from the Table Viewer, if the table has been created with area processing. To do so, the user simply clicks on a cell to highlight it, then clicks on the Map toolbar button. This will invoke the Map Viewer, showing a thematic map. In Figure 4 the user has highlighted the cell containing 3,543, which is the total number of females for Popstan. The generated thematic map would show the "Total number of females" as the mapped variable.

34. The user can keep adding variables to the map data file by going back and forth between the Table Viewer and Map Viewer. For example, the user can click on the total number of males, click the Map button and the "Total number of males" would be shown as a thematic map and added to the list of variables in Map Viewer. The user can make a whole new tabulation and add more variables to the list as long as they share the same geography. The user can save the map data file at any time, and even come back to it later to add yet more variables.

### **III. Creating Data Dissemination Products**

35. The IMPS CrossTab and MapViewer modules have been designed to make it as easy as possible for statistical data producers to create data dissemination products. Typically, such products would be in the form of a CD-ROM or on an Internet Web site. The U.S. Bureau of the Census has included CrossTab on several CD-ROM data dissemination products, including the 1990 Census Public Use Microdata and various iterations of the Current Population Survey.

36. A tabulation product would require (1) the data file or files, (2) an IMPS data dictionary and (3) the CrossTab and Data Dictionary modules (.EXEs and .DLLs, totaling less than three megabytes). The CrossTab module has been described in detail above. The data dictionary is a text file which describes the records, variables and values in the data file. The Data Dictionary program (.EXE) is the tool one uses to create a data dictionary. This Windows program is as easy to use as CrossTab or Map Viewer. Creating the data dictionary for the data files in the product is a one-time task. However, the Data Dictionary program is generally included with the dissemination product so that end users can create their own value sets for particular values. For example, the pre-made data dictionary may include the Age variable by single years and also in 5 year brackets. The user may wish to tabulate by 10 year brackets, and would use the Data Dictionary program to do so.

37. A thematic mapping product would require (1) a file containing polygons and geographic codes (map file), (2) a file containing data points corresponding to the geographic codes (map data file) and (3) the Map Viewer module (.EXEs and DLLs, totaling less than two megabytes).

38. The map file need only be created once for a given country or region. Just about any map in electronic form can be imported into Map Viewer. IPC has a utility to convert ESRI shape files to Map Viewer format. Most GIS packages are capable of exporting such shape files. The conversion utility will be released as part of the Map Viewer software package in the near future. In the meantime, IPC will perform conversions for any statistical office which can supply shape files. IPC has also produced a collection of map files for many countries and regions which are available for download on the World Wide Web.

39. A map data file is a straightforward tab-delimited text file. It can be created from summary data using standard spreadsheet packages, such as Microsoft Excel, or even entered using text editors, such as Notepad. The format of the map data file is fully documented in Map Viewer's on-line help system. Map data files can also be constructed directly from CrossTab output as described in the section above named "Making thematic maps from tables."

40. Once the map file is prepared (or downloaded), one can create many corresponding map data files. A given map data file may include related variables from any particular census, survey or data base.