Data Management Challenge: Select All That Apply: JMP® to the Rescue

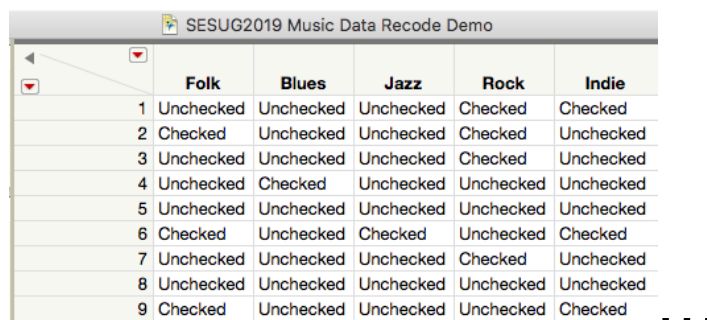Mira Shapiro, MSc., Analytic Designers LLC

ABSTRACT

Online surveys often provide questions with a variety of choices and instruct the user to "Select All That Apply". Depending upon the backend of the system, the desired analytics and output, and available tools, these sort of responses can sometimes create data management challenges. Hidden under JMP Col→Utilities are several tools that allow the user to recode the individual responses and then combine them into one column characterizing the user's complete response to a particular question. In this discussion we will show step by step how to use these techniques and demonstrate how the results can be used for analytics.

INTRODUCTION

Surveys of all kinds often have instructions for some questions that include the phrase "check all that apply." Very useful for healthcare-related questions like "Which symptoms are you experiencing?." Often researchers are interested in the counts of the individual choice, number of symptoms overall, which symptoms are most likely to occur together, and the various associated descriptive statistics. Furthermore, the recording of the choices in the data is often not ideal or ready for analytics. In this discussion, the answer to the question, "Which music genres do you enjoy?" and the associated "check all that apply" responses are used to demonstrate the features in JMP that make the data transformation and analysis of these type of responses a quick and easy task.

RAW DATA CLEAN-UP

The raw data used for this discussion came in the format as shown in Display 1. The platform collecting the responses to the online survey, recorded them as "Checked" or "Unchecked", allowed them to be downloaded as a CSV file, and here they are loaded into JMP using the point-and-click interface. As shown in Display 2. blank rows toward the end of the dataset were loaded into JMP with missing entries designated as blanks. Prior to data transformation, it is easiest to delete the blank rows since they contain no information. As shown in Display 3. Rows→Delete Rows was used to remove the empty rows.


	Folk	Blues	Jazz	Rock	Indie
1	Unchecked	Unchecked	Unchecked	Checked	Checked
2	Checked	Unchecked	Unchecked	Checked	Unchecked
3	Unchecked	Unchecked	Unchecked	Checked	Unchecked
4	Unchecked	Checked	Unchecked	Unchecked	Unchecked
5	Unchecked	Unchecked	Unchecked	Unchecked	Unchecked
6	Checked	Unchecked	Checked	Unchecked	Checked
7	Unchecked	Unchecked	Unchecked	Checked	Unchecked
8	Unchecked	Unchecked	Unchecked	Unchecked	Unchecked
9	Checked	Unchecked	Unchecked	Unchecked	Checked

Display 1. Raw data read from a CSV file into JMP.

		Folk	Blues	Jazz	Rock	Indie
	108	Checked	Unchecked	Unchecked	Unchecked	Unchecked
	109	Unchecked	Unchecked	Checked	Unchecked	Unchecked
	110					
	111					
	112					
	113					
	114					
	115					
	116					
	117					
	118					
	119					

Display 2. Empty rows read into JMP. These rows were deleted because they contained no information. If they contained missing values indicating a valid survey entry, they would have been retained and used in the analysis.
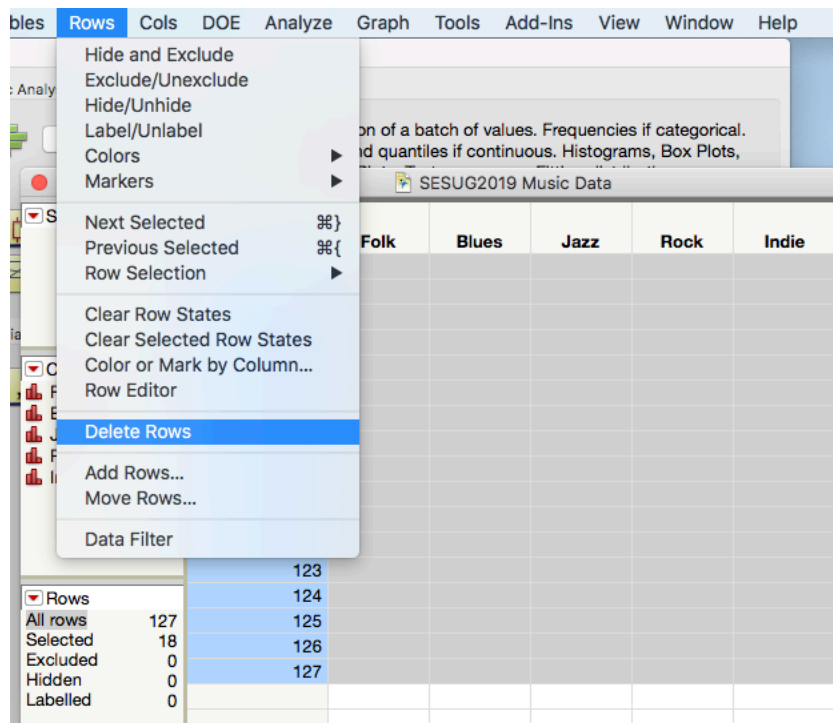
Display 3. The rows were eliminated using Rows→Delete Rows,

DATA TRANSFORMATION

With data generated from a "Check all that apply" question, usually several columns need to be transformed the same way. Rather than working with each column independently, JMP provides the "Standardize Attributes" operation under the "Cols" tab. As shown in Display 4, first select the columns and then select the operations of interest.
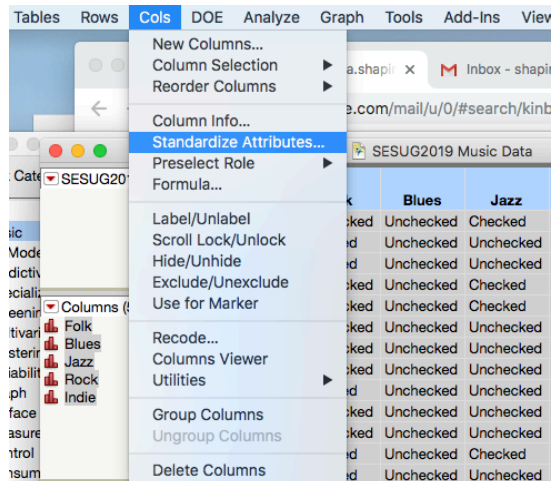
Display 4. Select the columns of interest and then Cols→Standardize Attributes…

As shown in Display 5., the next step is to recode "Unchecked" to 0 and "Checked" to 1. Always remember to click the "Recode" button in the active window to complete the operation.
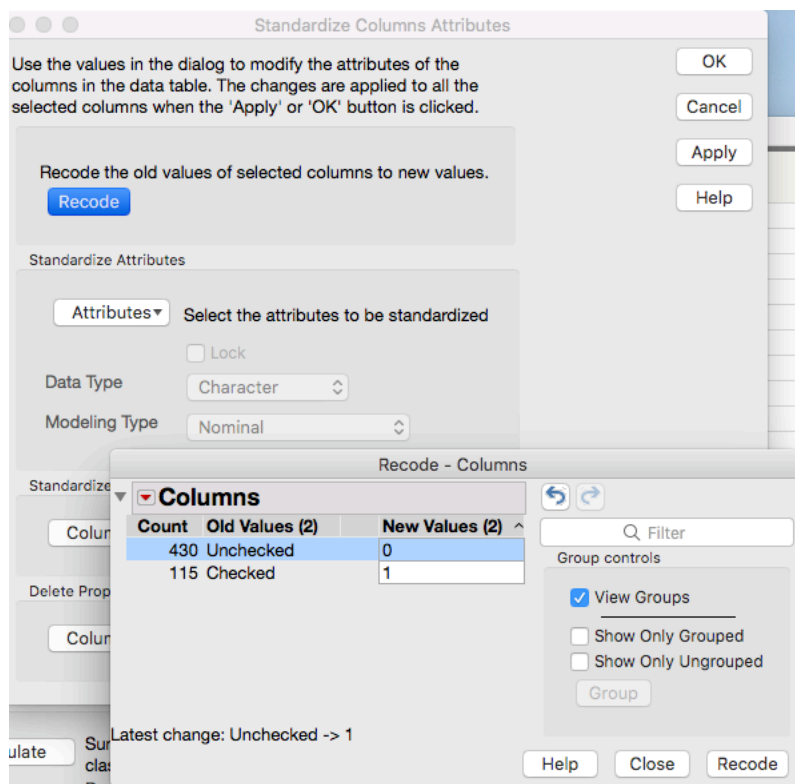
Display 5. Recode both to recode the characters to indicator values. Click the "Recode" button.

In addition, in order to use the indicator variable in numeric operations, it is necessary to change the Data Type from "Character" to "Numeric". As always in JMP, remember to click the "Apply" and then "OK" buttons to execute the operations.
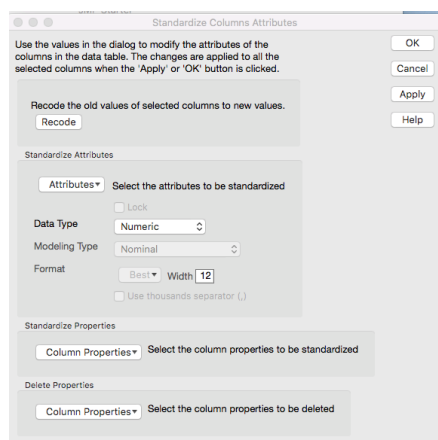
Display 6. Change the Data Type to Numeric. Again, remember to click "Apply" and "OK" .

The data is now in a format that allows for use of the JMP "Utilities" to report on the combinations that result on the "Check all that apply" instruction. Notice in Display 7. that the columns are now coded as ones and zeroes. It is usually recommended that in the "Column Properties" section that "Value Labels" be created that reflect the meaning of the binary indicator variables. In that section of "Column Properties" in JMP, the labels can be used or not. In this circumstance the meaning is very clear, and it is useful to display the numeric values.

Display 7 Data recoded to zeroes and ones.

Display 8. through Display 16., and their associated captions, provide step-by-step instructions to create a summary column indicating the names of the columns that each respondent selected, and a column with a count of how many they selected. The former is useful for reporting on combinations of responses, and the count column, although can it be calculated easily, provides a shortcut for multiple reporting operations.
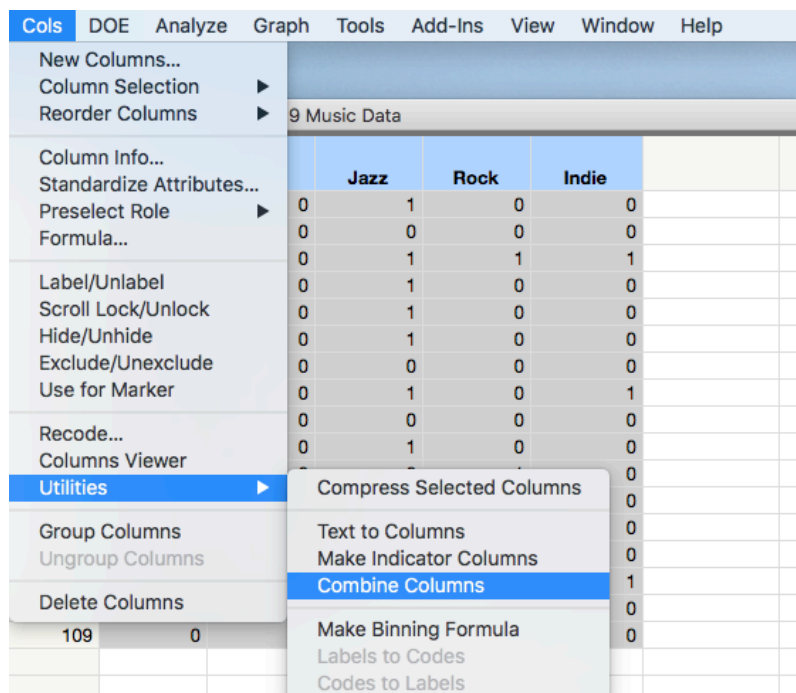

Display 8. Use Cols→Utilities→Combine Columns to create a column that names those columns selected by each respondent.
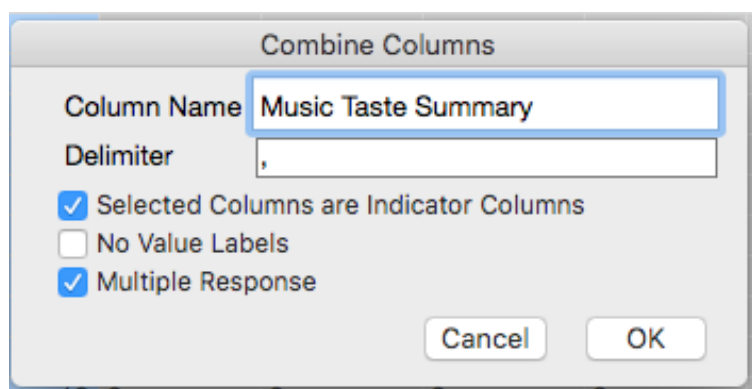

Display 9. Indicate the "Delimiter", in this case "," and select desired attribute of the new combined column. In this example, the columns are "Indicator Columns" and this does constitute "Multiple Response" selections.

|	Music Taste Summary
17	
18	
19	Folk
20	Folk,Blues,Jazz,Ro…
21	
22	Folk,Blues
23	
24	
25	Folk
26	Jazz,Rock
27	Folk
28	
29	Folk,Jazz
30	
31	Blues,Jazz
32	Folk
33	
34	Blues,Jazz,Rock
35	Jazz

Display 10. The entries in the newly created summary column contain the column headings of the indicator columns that were selected by each respondent. A blank column entry indicates that the respondent filled out the survey entry fields but did not respond to this question.
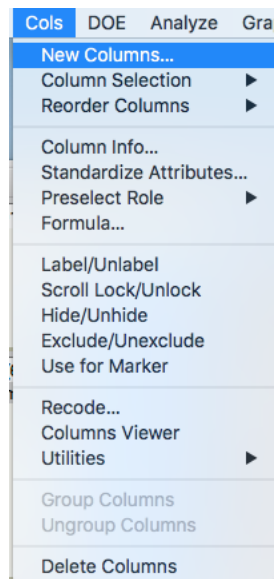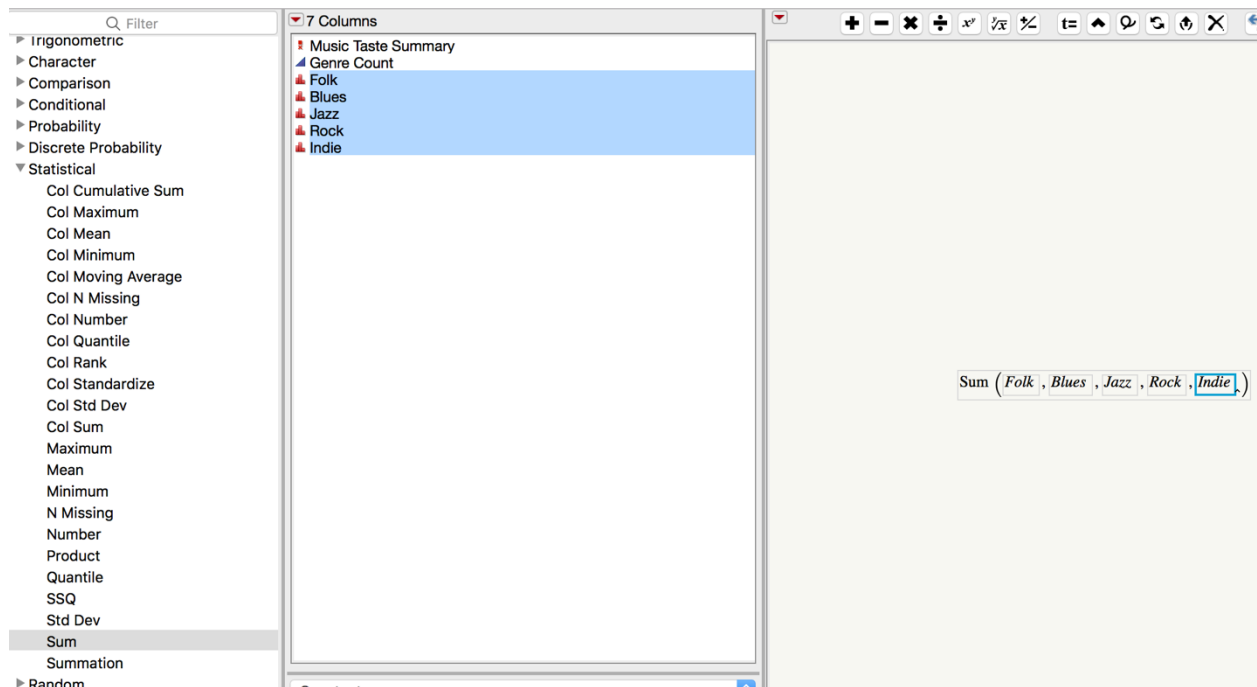

Display 11. Create a new column using Cols→New Columns… which will house the count of the selections for each respondent.
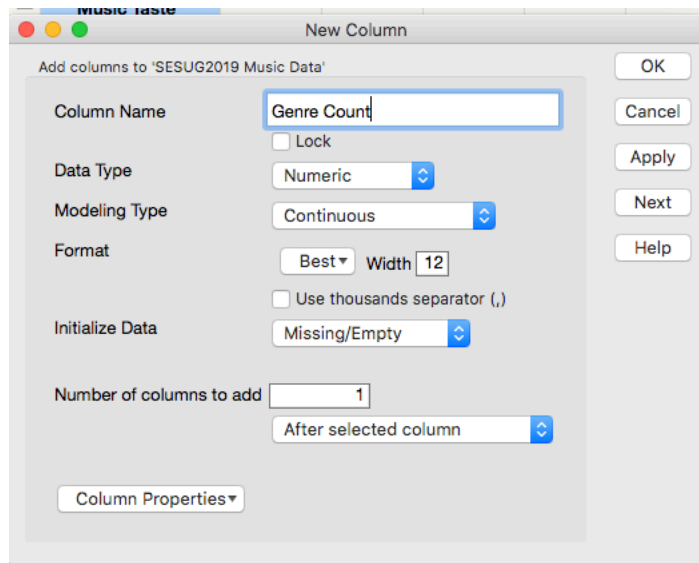
Display 12. Choose a name for the column, here "Genre Count", and make the Data Type Numeric. If desired, it is possible at this stage to create multiple columns.
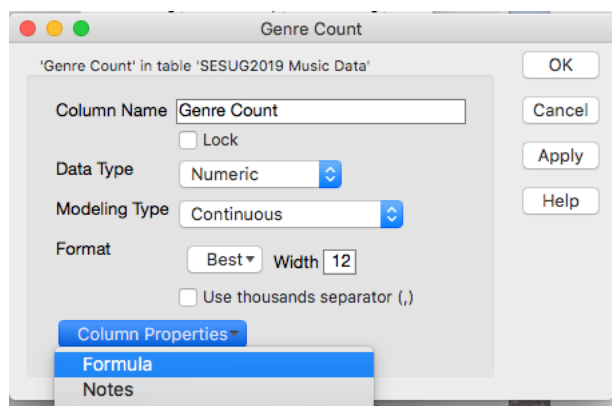
Display 13. Select Column Properties→Formula to reveal the formula window.

Display 14. Under functions, select Statistical→ Sum and then drag the indicator columns into the function to add up the total number of responses per individual.
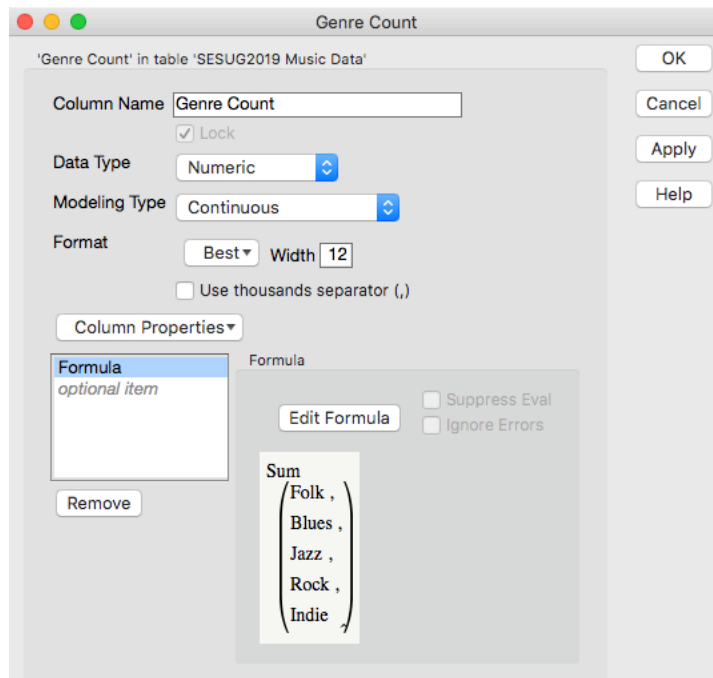
Display 15. Click "Apply" and "OK" to complete the column creation.

22	Folk,Blues	2	1	1	0
23		0	0	0	0
24		0	0	0	0
25	Folk	1	1	0	0
26	Jazz,Rock	2	0	0	1
27	Folk	1	1	0	0
28		0	0	0	0
29	Folk,Jazz	2	1	0	1
30		0	0	0	0
31	Blues,Jazz	2	0	1	1
32	Folk	1	1	0	0
33		0	0	0	0
34	Blues,Jazz,Rock	3	0	1	1
35	Jazz	1	0	0	1
36	Jazz	1	0	0	1

Display 16. The data is fully transformed and ready for analysis. The leftmost column houses the choices separated by commas; the second column contains the count of the choices.

Now that the data is transformed, the count and summary columns are available for use in analysis. The challenge with "all that apply" responses is capturing the summary without double counting and not properly counting the number of responses per individual. This coding system takes the guesswork out of these operations and facilitates reporting.

DESCRIPTIVE STATISTICS

Even with this simple sample data set, there are numerous types of reports and statistics that can be generated with JMP. The following examples only scratch the surface of the possibilities of how the data can be evaluated. Each Display and Output are labeled with the instructions for how to create the report.

Display 17. Select Tables→Summary to create a new Data Table that contains the Music Taste Summary column and a column indicating the number of rows that contain that combination of selections.
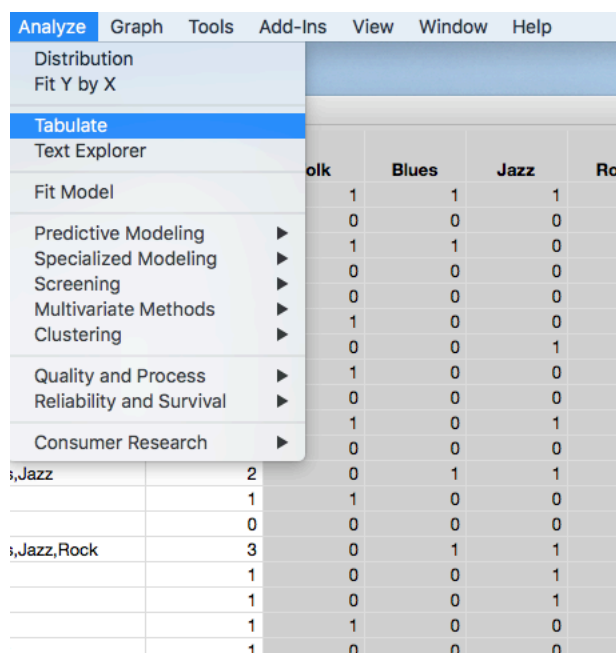
Display 18. Drag and drop Music Taste Summary to the Group and then click "OK".

Output 1. This Data Table created as described in Display 17. and Display 18., shows combination of responses and a count of the number of times that each combination occurred.

Display 19. Use the Analyze→Tabulate to create a report showing the individual music genre selections.

	0		1	
	N	**% of Total**	**N**	**% of Total**
Folk	67	61.47%	42	38.53%
Blues	97	88.99%	12	11.01%
Jazz	74	67.89%	35	32.11%
Rock	92	84.40%	17	15.60%
Indie	100	91.74%	9	8.26%

OUTPUT 2. This report shows the number of times that each individual type of music was selected.

CONCLUSION

What first appears to be a simple task, evaluating "Select all that apply" question responses require some necessary data transformations before embarking on an analytics exercise. Under JMP's menus there are often some "hidden gems" that might go unnoticed. JMP Customer Support, when I was approaching this challenge, thankfully pointed me to the Cols→Standardize Attributes and Cols→Utilities menus and how to use them. Using this approach saved countless hours of programming and QC time in data transformation. Although the example data in about music, the actual project is one that will have an impact on the health and outcomes for children with serious health conditions. Employing this technique with healthcare data provides a very efficient and accurate approach for the "Check all that apply" questions that are essential in healthcare and other surveys.

REFERENCES

https://www.jmp.com/en_us/support/jmp-documentation.html

ACKNOWLEDGMENTS

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Mira Shapiro, MSc.
Analytic Designers LLC
mira.shapiro AT analyticdesigners.com