Paper 3889-2019

Using the JMP® Match Function to Analyze Survey Skip Patterns

Mira Shapiro MSc., Analytic Designers LLC

ABSTRACT

JMP® has a robust set of functions that facilitate data cleaning and transformation. This discussion focuses on using the Match function in a formula to create new columns. Branched survey question responses from the MyLymeData patient registry are used to demonstrate the power of the Match function to quickly and accurately create a Likert scale describing patient reported health status following antibiotic treatment and to further create subgroups based on their responses.

INTRODUCTION

Patient registry data is an important source of information in understanding disease and treatment. Like most surveys, often patients are fielded different questions that are branched based on their response to an earlier question. Special care must be taken in analysis of data in these circumstances to make sure that unanswered questions due to branching are not counted as missing. Using patient responses to three questions from the MyLymeData patient registry developed by LymeDisease.org, we will demonstrate the power of the JMP Match function for handling these circumstances. The Match function makes this process easier by replacing a series of If-then-else statements, and by self-populating the fields with the responses recorded in the column under evaluation. We looked at responses to survey questions where patients reported themselves as unwell and reported the degree to which their condition had changed following antibiotic treatment by using a widely-used global rate of change scale. The scale asks whether patients report their health condition is better, worse, or unchanged since treatment. (Kamper, 2009).

Those who respond better or worse are asked a follow up question where they report the degree to which their condition either improved or worsened. In this analysis responses are then associated with a value from -7 to +7. The result is a 15-point Likert scale, ranging from -7 to +7, with the unchanged midpoint at 0. For example, compared with no change (0), those responding worse are better select a magnitude ranging from (0) "almost the same", (1) "hardly any better at all," (2) "a little better," (3) "somewhat better," (4) "moderately better," (5) "a good deal better," (6) "a great deal better," and (7) "a very great deal better." Those responding worse, choose among similar responses, ranked (-1 to -7). (Wang, 2011)

This transformation to the Likert scale from the answers to the three questions was done efficiently and cleanly using the JMP Pro 13.1 Match function. This discussion focuses on the strengths and ease of use of the Match function as it was used to take the answers to the patient questions to create a numeric scale and subgroup classification for analysis.

DATA AND METHODS

MyLymeData is a patient registry developed by LymeDisease.org that enables pooling of longitudinal healthcare data. Participation in the registry is voluntary, and all respondent identities have remained strictly confidential. The MyLymeData registry and its surveys have been approved by the Chesapeake Institutional Review Board.

Since its launch in November 2015, over 11,000 patients have enrolled. Phase 1 of the registry data collection was completed in November 2016. We began this analysis with a sample of 3,595 self-identified unwell participants from Phase 1 and excluded 164 participants that we could not identify as residents of a US state, who were not diagnosed with Lyme disease by a healthcare provider, or who did not identify as being "unwell". The 3,431 participants (Table 1) used in this analysis included US residents clinically diagnosed with Lyme disease who completed the Phase 1 surveys and identified their health status as "unwell" when completing the survey.

	Unwell
Completed baseline	3,595
Not US resident	-164
Ending sample size	3,431

Table 1 Final Phase 1 Sample of Unwell used for this analysis

CREATION OF A LIKERT SCALE VARIABLE FROM RESPONSES

The participants were asked a series of questions. For this analysis, we used the response to the question: "In general overall, I would say that with antibiotic therapy, my Lyme symptoms are worse, unchanged, or better". Based on their response, they were asked a question concerning how much worse or better they felt following treatment. The first 18 rows of the JMP Data Table are shown in Figure 1.

	Better Worse ...	Worse Degree	Better Degree
1	Unchanged		
2	Unchanged		
3	Better		A good deal better
4	Better		
5	Worse	Almost the same	
6	Unchanged		
7	Worse	A good deal worse	
8	Better		A great deal better
9	Unchanged		
10	Unchanged		
11	Unchanged		
12	Better		A great deal better
13	Better		A very great deal better
14	Better		A great deal better
15	Unchanged		
16	Better		A little better
17	Better		A great deal better
18	Worse	Moderately worse	

Figure 1 First 18 rows of the JMP Data Table with question responses

The first step in the process is to examine columns containing the responses to the three questions we are using in the analysis. This sample dataset has been cleaned and verified prior to this analysis. At this stage, we will use the *Value Ordering* column property to arrange the values so that the output of any analysis with JMP will be in the desired order. For example, we would like the "Better Worse Unchanged" column to appear in results as Better, Unchanged and Worse which corresponds to the numeric scale we are creating from +7 down to -7. Likewise, as shown in Figure 2, the Better Degree column will be ordered from "A very great deal better" down to "Almost the same". By using the "Move Up", "Move Down" and "Reverse" buttons, we ordered the "Better Degree" column so the resulting output and interim *Match* function displays will be ordered in a way to make it easier to assign and use the Likert scale values. As with all JMP operations, be sure to click the Apply button and then the OK button to apply the changes. The same process was used to order the "Worse Degree" column.

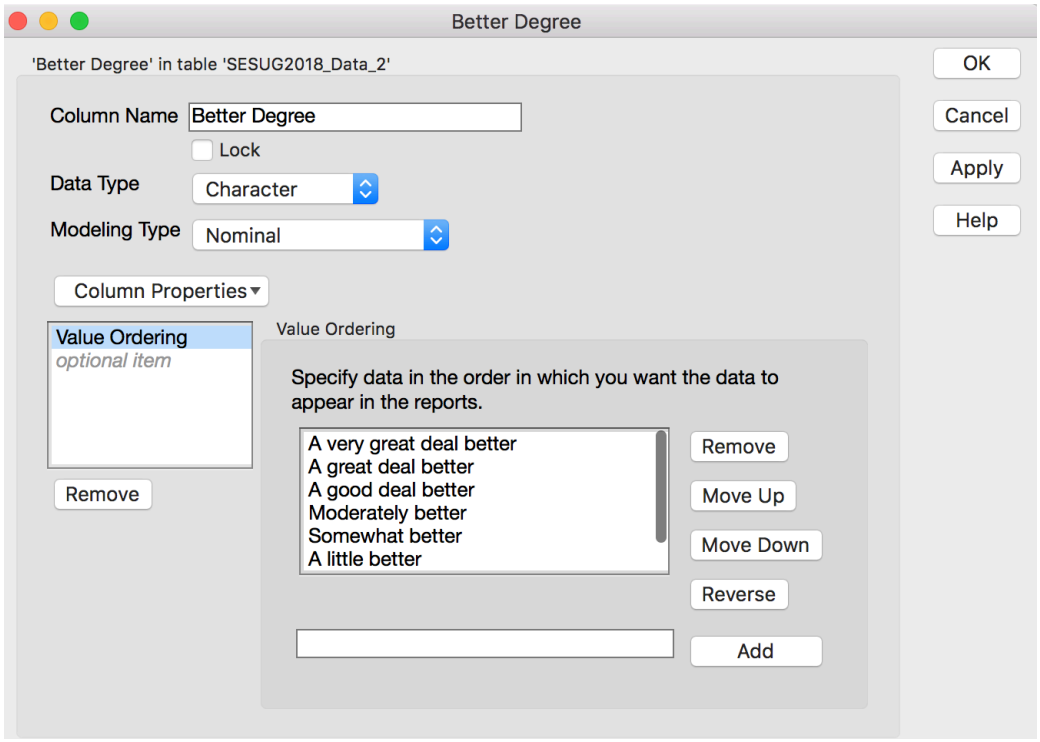


Figure 2 Use the *Value Ordering* column property to prepare the data

The next step is to create a new column to house the Likert scale values. As shown in Figure 3, The new column dialog allows for assigning the Column Name and other column properties.

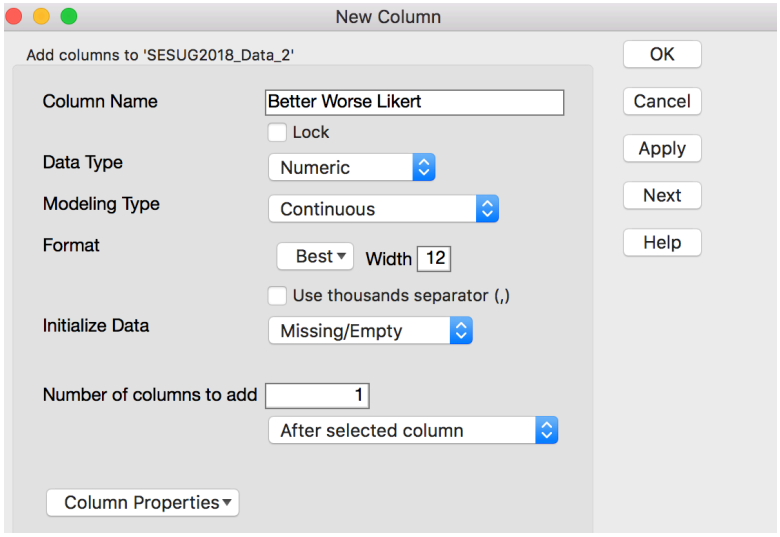


Figure 3 Create a new column named Better Worse Likert

As shown in Figure 4, Under Column Properties in the New Column window we will navigate to the Formula property so we can use the *Match* function populate the column.

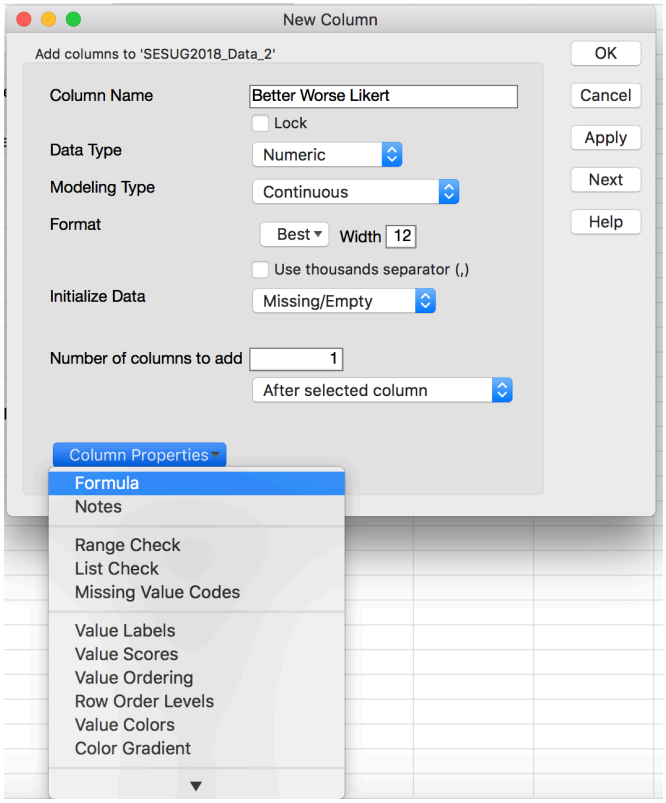
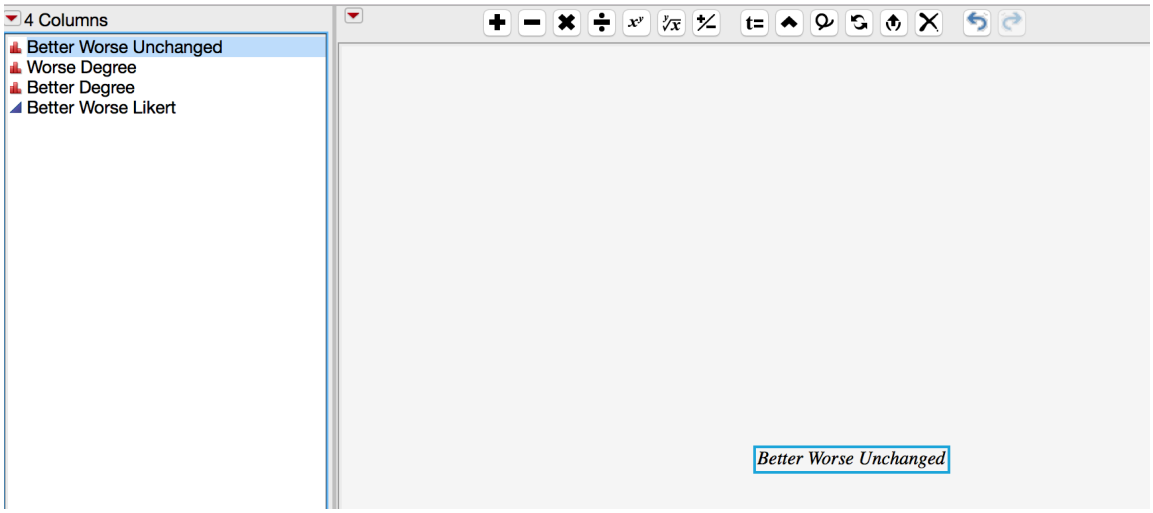


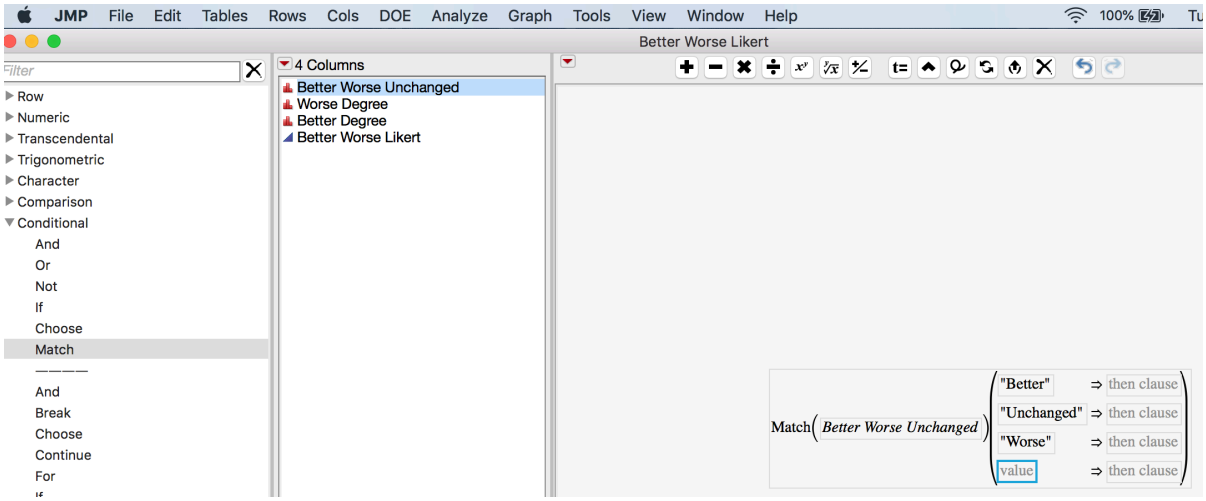
Figure 4 Select the Formula column property

Using the Formula window, use the JMP drag and drop capability and the *Match* function multiple times to create the desired Likert scale.

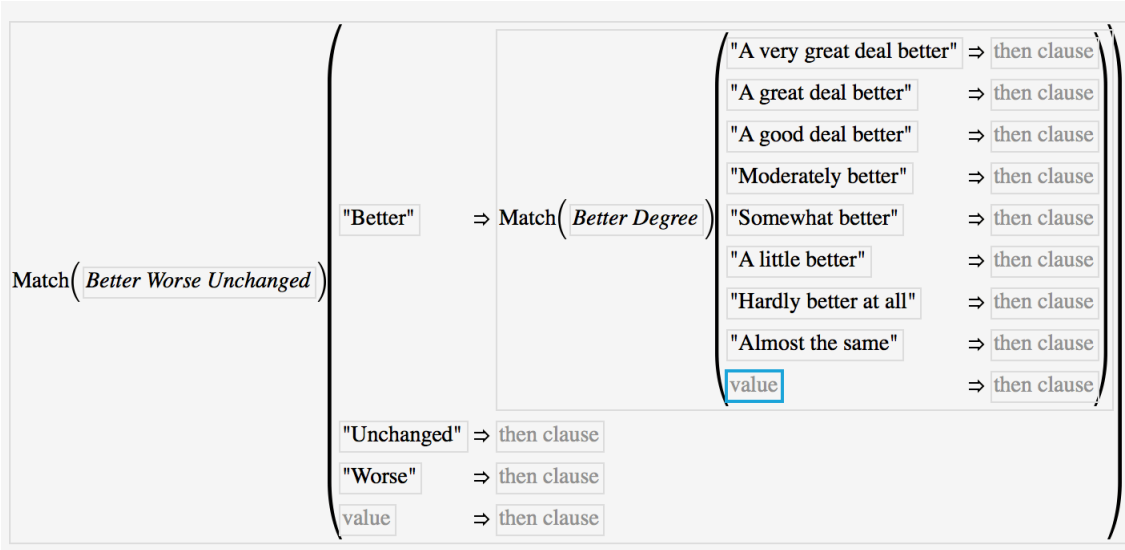
Step 1: Drag to Better Worse Unchanged variable to the formula box.



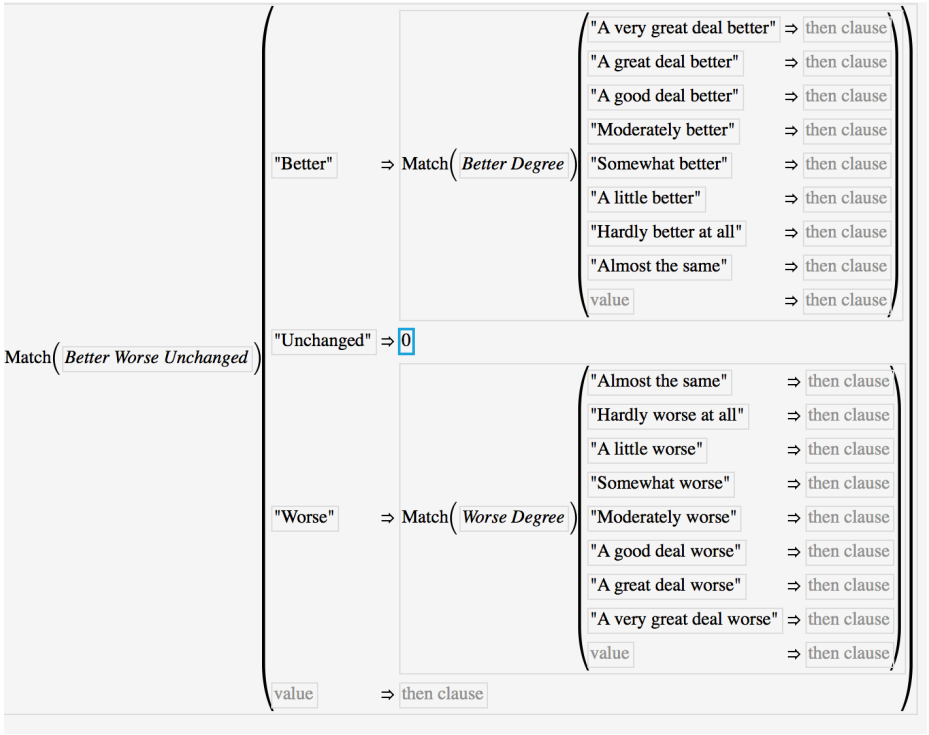
Step 2: Apply the *Match* function that appears in the Conditional function list to the Better Worse Unchanged Column values. Hold the shift key when clicking *Match* and JMP will self-populate the function window with the values contained in the column.



Step 3: Drag and drop the Better Degree column into the *then clause* area associated with "Better". Then repeat the process described in Step 2 by holding the shift key when clicking *Match* and JMP will self-populate the values from the Better Degree column.

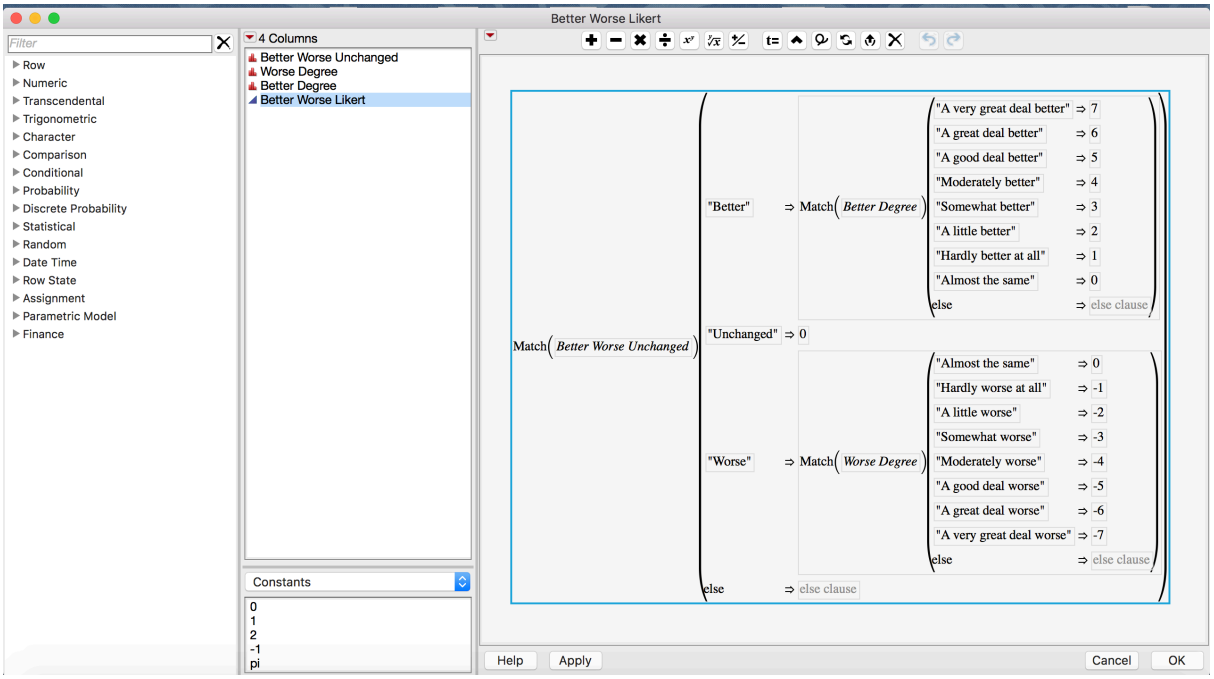


Step 4: Drag and drop the Worse Degree column into the *then clause* area associated with "Worse". Then repeat the process described in Step 2 by holding the shift key when clicking *Match* and JMP will self-populate the values from the Worse Degree column. Enter a "0" in the then clause area associated with "Unchanged".



Step 5: Enter the numeric values associated with the values ranging from 7 down to -7. This step is easy to complete since earlier we used the Value Ordering property to arrange each column so that values are displayed in the proper order.

Step 6: Once the values are entered, click the Apply button and then the OK button to populate the new column.

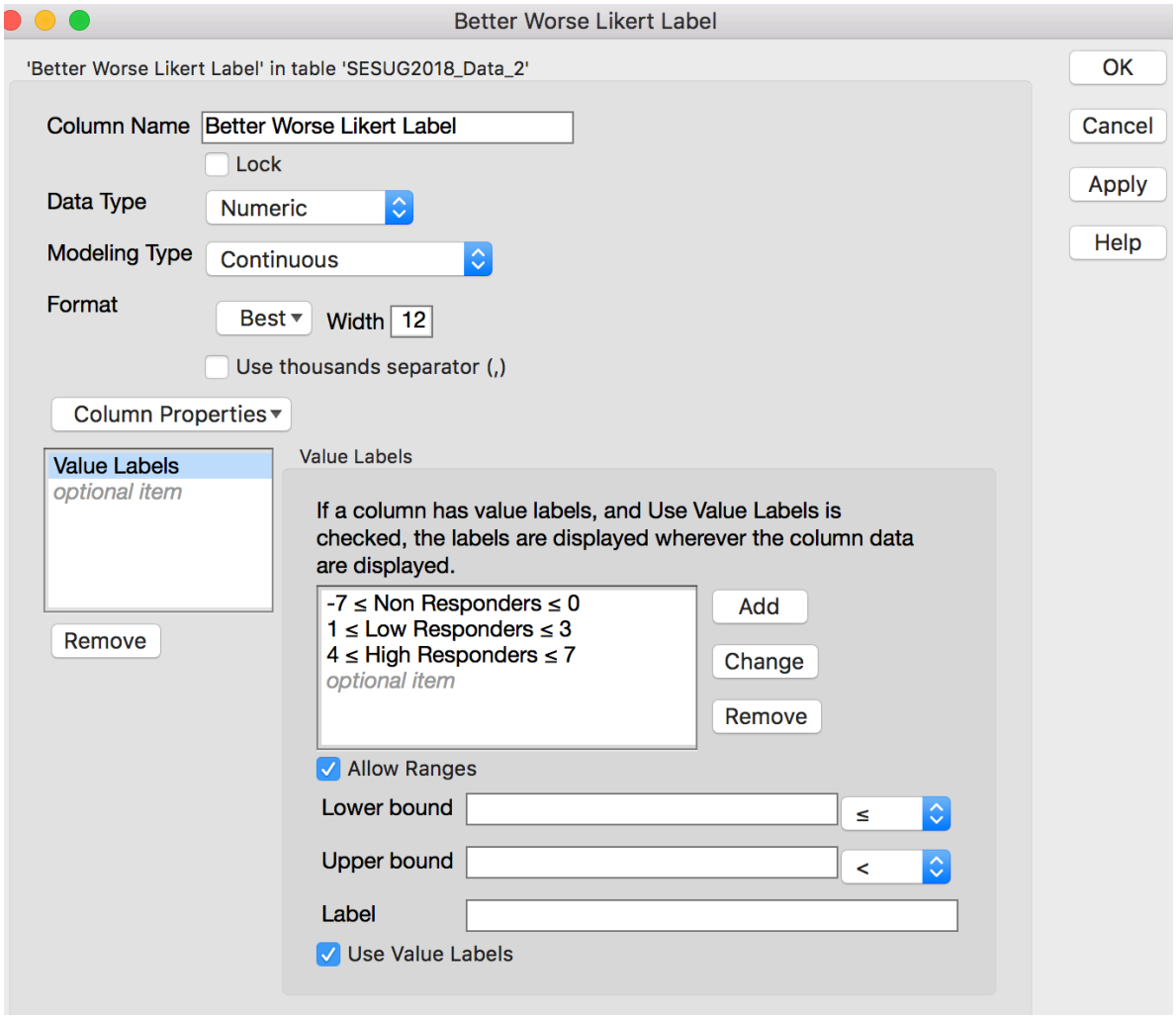


As shown in Figure 5, the new column, Better Worse Likert is populated and ready for use.

Better Worse Likert
0
0
5
•
0
0
-5
6
0
0
0
6
7
6
0
2
6
-4

Figure 5 The first few rows of the Better Worse Likert column.

Step 7: Create subgroups Non Responders (-7 to 0), Low Responders (1 to 3), High Responders (4 to 7). This can be accomplished multiple ways in JMP. The first way would be to create a new column and use the *Match* function to assign categories as previously shown when the Likert scale was created. Alternatively, labels can be assigned to the Better Worse Likert column. For this analysis, we chose to create a new column that contained the labeled Likert values. By using the Value Labels feature under the Column Properties dialog, enter the Lower bound, Upper bound and Label for each group and click the Add button for each. Make sure to choose the proper symbol if the range should be inclusive. When the three categories have been defined, click the Apply button and then click OK. Note there is an option to deselect Use Value Labels to display the numeric values in the column.



The resulting Column Better Worse Likert Label can be displayed as shown in Figure 5 by deselecting Use Value Labels or as shown in Figure 6.

Better Worse Likert Label
Non Responders
Non Responders
High Responders
•
Non Responders
Non Responders
Non Responders
High Responders
Non Responders
Non Responders
Non Responders
High Responders
High Responders
High Responders

Figure 6 The first few rows of the Better Worse Likert Label column.

BASIC VISUAL ANALYSIS IN JMP

One of JMP's strengths is the capability to quickly create data visualizations. As shown in Figure 7, one of the easiest ways to get a quick look at the contents of a column in JMP is to go to the Menu Bar and Select Analyze->Distribution and then drag and drop the column or columns of interest into the Y_1 Columns section of the window and then click OK.

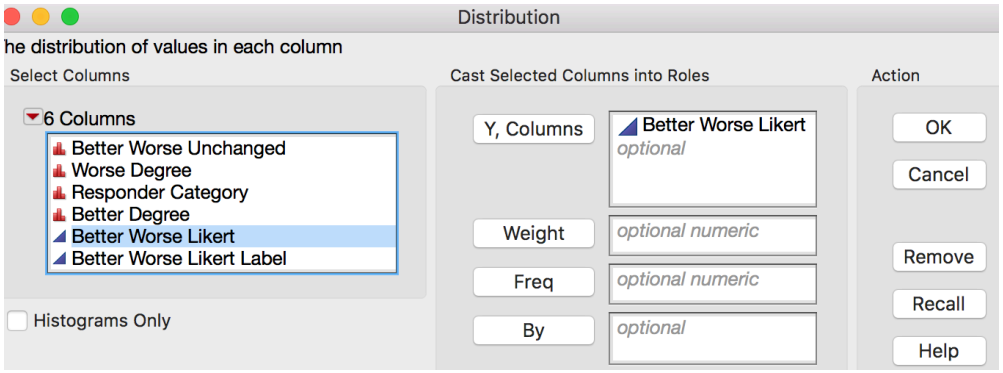


Figure 7 Analyze->Distribution window.

The results, as shown in Figure 8 are vertically displayed but can easily be changed to horizontal by clicking the red triangle to the left of the column name->Display Options->Horizontal Output. For most purposes, horizontal output (Figure 9) is desirable. There are multiple options for refining and embellishing this quick analysis of the column.

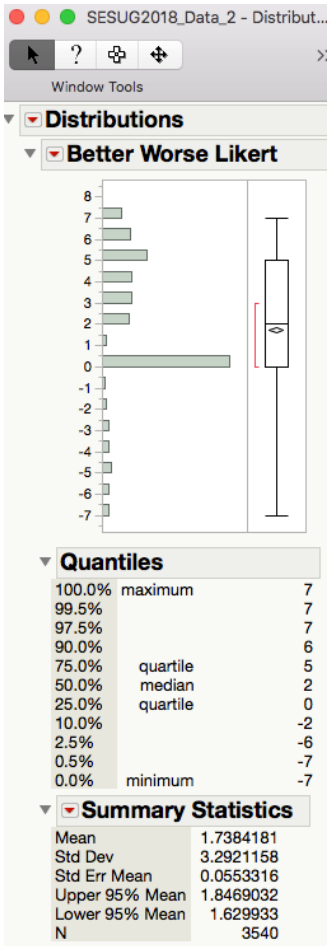


Figure 8 Default results of Analyze->Distribution.

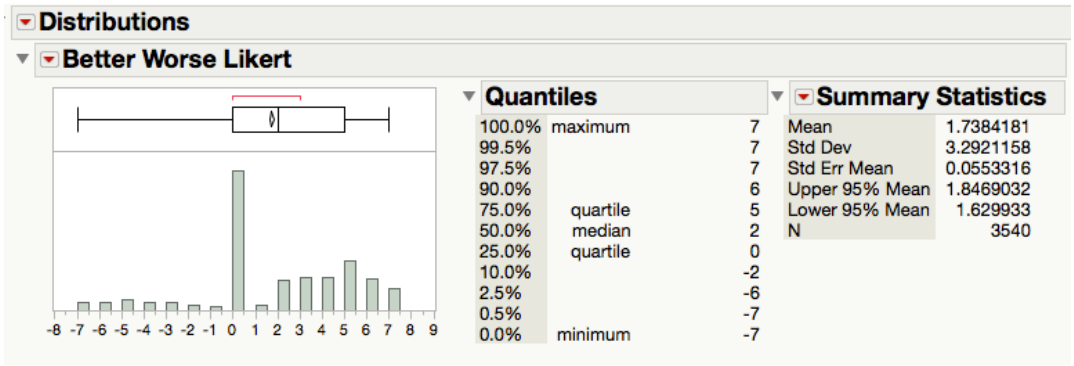


Figure 9 Preferred horizontal layout of Analyze->Distribution.

To take a quick look at the subgroups we created, the Analyze->Tabulate (Figure 10) and Graph->Graph Builder (Figure 11) operations were used. By dragging and dropping the Responder Category variable into these tools, simple visualizations are created.

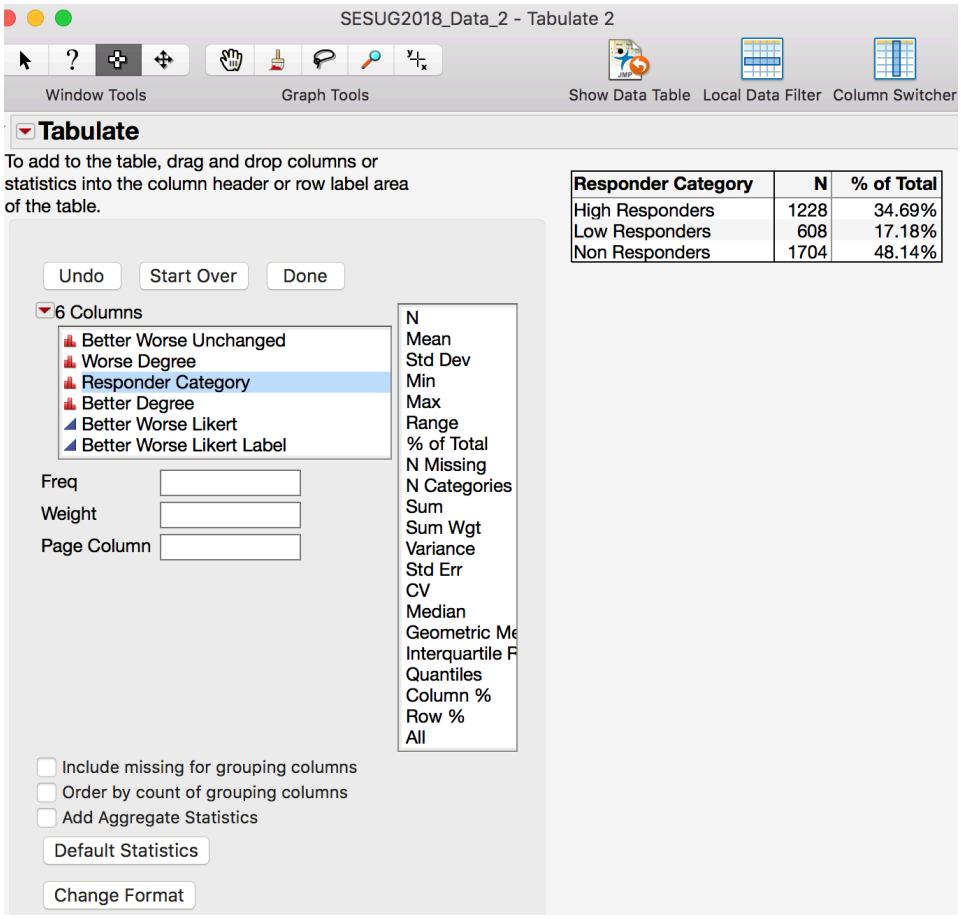


Figure 10 Analyze->Tabulate to examine the Responder Category column

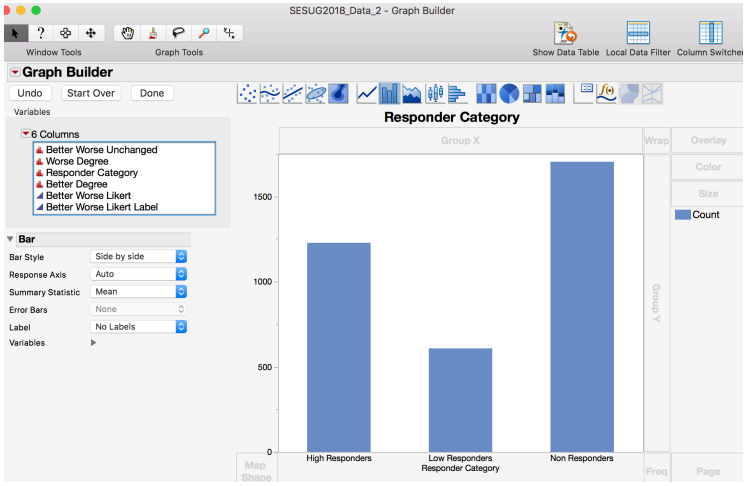


Figure 11 Graph->Graph Builder to create a Bar Chart of Responder Category column

CONCLUSION

We have not even scratched the surface of the data transformation, analysis and visualization and customization capabilities of JMP. This discussion was focused on the use of the JMP *Match* function to take branching survey questions and easily transform them into a numeric scale for analysis. The ability to drag and drop column names into the formula window and use the shift key in invoking the *Match* function to display column values, is an easy-to-use self-documenting approach to what can often be a complex programming exercise. Although this technique is simple to use and can be used by experienced programmers and non-programmers alike, it is important to note that the user should fully understand the data, question branching, and survey design to create accurate and meaningful results.

REFERENCES

Kamper SJ, Maher CG, Mackay G. Global rating of change scales: a review of strengths and weaknesses and considerations for design. J Man Manip Ther. 2009;17(3):163-70.

Wang YC, Hart DL, Stratford PW, Mioduski JE. Baseline dependency of minimal clinically important improvement. Phys Ther. 2011 May;91(5):675-88.

ACKNOWLEDGMENTS

Special thanks to the CEO of LymeDisease.org Lorraine Johnson, JD, MBA. We also thank the patients who participate in MyLymeData and Dr. Raphael Stricker, the Co-Principal Investigator on the project for his thoughtful comments on the concept.

Thank you to MaryAnne DePesquo the SAS Global Forum Conference Chair 2019 and her team of volunteers.

RECOMMENDED READING

- *LymeDisease.org*
- **Removing the Mask of Average Treatment Effects in Chronic Lyme Disease Research Using Big Data and Subgroup Analysis** Lorraine Johnson JD MBA, Mira Shapiro MSc., Jennifer Mankoff PhD, *Healthcare* **2018**, *6*(4), 124; https://doi.org/10.3390/healthcare6040124

CONTACT INFORMATION <HEADING 1>

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

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