Text Mining – Describe Unstructured Text Data

The Text Explorer platform is used to explore frequently used words and phrases in unstructured text data such as text found in product reviews, social media posts, comment fields in surveys, incident reports, etc. Results are summarzied through frequency tables and word clouds. Tools such as recoding, combining terms, creating stop words, among others are available to clean the data and help target the anaysis efforts.

Note: Additional text analysis tools are available in JMP Pro. See the **Text Explorer – Analyzing Unstructured Text Data** guide for more information.

Example: Pet Survey.jmp (Help > Sample Data Folder)

Summarizing Text Data

1. From an open JMP® data table, select **Analyze > Text Explorer**.
2. Click on a categorical or unstructured text variable from **Select Columns**, and click **Text Columns** (categorical variables have red or green bars and unstructured text variables have grey lines (≡). Click **OK**.

   By default, JMP will provide the following results:

   - A summary of the number of terms and cases.
   - Term and Phrase Lists, providing a summary of the most commonly occurring terms and phrases.

Cleaning and Targeting

Many text preparation options, including stemming and parsing, are available under the red triangle. Additional options are available from the Term and Phrase Lists.

- To remove terms or phrases, select the terms (or phrases), right-click, and select **Add Stop Word**.
  Here, we selected 'cat' 'cats' 'dog' and 'dogs' as **Stop Words**.
- To add phrases to the term list, select the phrases, right-click, and select **Add Phrase**.
- To combine terms or change the values, select the terms, right-click, and select **Recode**.

These steps are very helpful in targeting analysis efforts to better find and quantify the importance of key themes.

Visualizing Text Data with Word Clouds

To generate a word cloud for your prepared text data, select **Display Options > Show Word Cloud** from the red triangle menu.

Options for customizing the word cloud are available from the red triangle next to Word Cloud such as coloring the words based on the value of a variable. In this example, the options Centered Layout and Arbitrary Colors were used.

Visit **Basic Analysis > Text Explorer** in **JMP Help** to learn more.