

Text Explorer – Describe Unstructured Text Data

The Text Explorer platform can be used to describe the counts and frequencies of words and phrases in unstructured text data, such as text found in comment fields in surveys or incident reports. Combine similar terms, recode text, describe the underlying patterns, or use a word cloud to visualize your text data.

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

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 (≡)).
3. 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.

4. Prepare the text data.

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**.
- 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**.

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.

In this example, the options Centered Layout and Arbitrary Colors were used.

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

Select Columns

▼ 1 Columns

Survey Response

Language

English

Maximum Words per Phrase

4

Maximum Number of Phrases

5000

Minimum Characters per Word

1

Maximum Characters per Word

50

Stemming

No Stemming

Tokenizing

Regex

☐ Customize Regex
 ☐ Treat Numbers as Words

Cast Selected Columns into Roles

Text Columns

Survey Response

optional character

ID

optional

By

optional

OK

Cancel

Remove

Recall

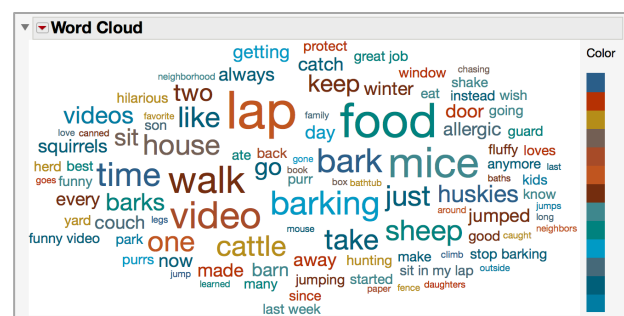
Help

Text Explorer for Survey Response

Number of Terms	Number Cases	Total Tokens	Tokens per Case	Number of Non-empty Cases	Portion Non-empty per Case
372	194	1921	9.90206	150	0.7732

Term and Phrase Lists

Term	Count	Phrase	Count	N
cat	55	video of the cat	5	4
dogs	48	sit in my lap	4	4
dog	47	cat food	4	2
cats	16	dog food	4	2
lap	14	dog barks	3	2
barking	12	duck hunting	3	2
video	11	funny video	3	2
food	10	great job	3	2
sit	10	last week	3	2
walk	10	stop barking	3	2
mice	9	cat in my lap	2	4
just	8	dogs do a great	2	4
like	8	hunting with the dogs	2	4
take	8	sled in the winter	2	4
bark	7	video of a cat	2	4



Notes: For more details on text exploration and analysis, including information on preparing text data, see Chapter 12 in the book *Basic Analysis* (under **Help > Books**) or search for “text explorer” in the JMP Help.