Text Mining – Analyzing Unstructured Text Data

JMP PRO 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. Additional tools are available in JMP® Pro for further analysis.

The text data must first be prepared for these analyses. See the **Text Explorer – Describing Unstructured Text Data** guide and the **JMP Help** for information on using the Text Explorer platform to do this.

Analyzing Unstructured Text

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

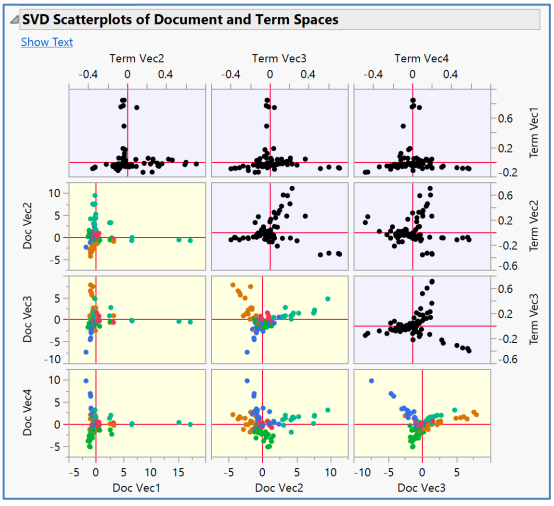
Text analysis involves transforming prepared text data into a **Document Term Matrix (DTM)**. Each row in the DTM corresponds to a document (a cell in the column of text data), and each column in the DTM corresponds to a term. The DTM is then used as an input in all analyses.

Here we selected **Save Document Term Matrix** under the **Red Triangle** and used the default specifications.

A number of analysis options are available from the top red triangle menu in the text explorer analysis report:

- **Latent Class Analysis** groups documents into clusters of similar documents.
- **Latent Semantic Analysis, SVD** is a dimension reduction technique similar to principal components analysis.
- **Topic Analysis** groups words into themes or topics, and is similar to factor analysis.
- **Cluster Terms** or **Cluster Documents** performs hierarchical clustering of the terms or documents.

Here we selected **Latent Semantic Analysis, SVD** using the default settings. Then selected **SVD Scatter Plot Matrix** using the default settings.



Saving Results

Text Explorer analysis results, such as the DTM, can be saved to the data table. The resulting columns can then be used as predictors in other analyses.

- To save information that corresponds to documents, save the Document Term Matrix, Document Singular Vectors or Document Topic Vectors.
- To save information that corresponds to terms, without respect to the specific documents, save the Term Table, Term Singular Vectors or Term Topic Vectors.

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