Mind the Gap: JMP® on the Text Explorer Express

Sept 21, 2016

W. Heath Rushing, Adsurgo LLC
Andrew Karl, Adsurgo LLC
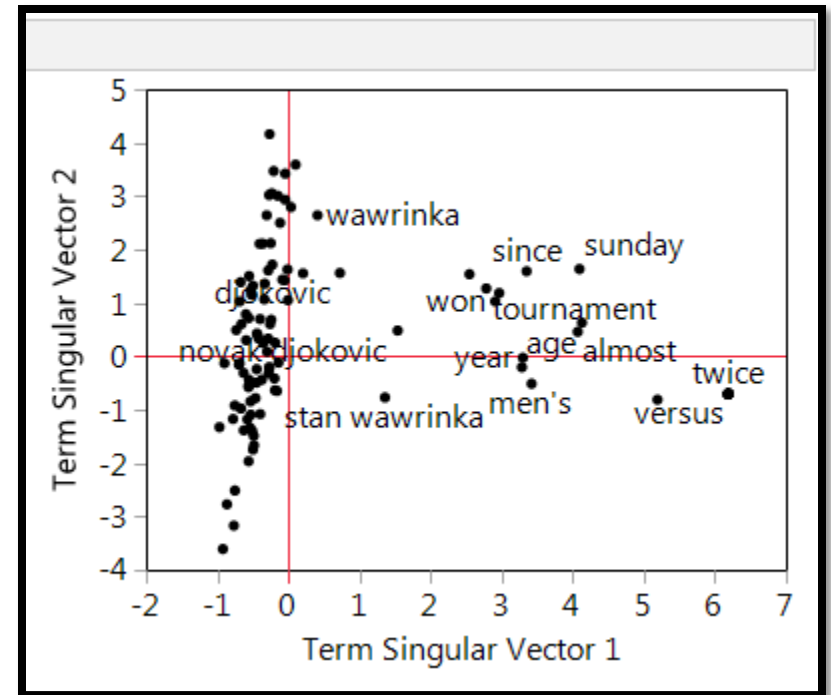James Wisnowski, Adsurgo LLC

Warm up

- Recent news article

- Word frequencies

- Scatterplot showing connections between words

- Word cloud

Use these basic JMP 13 Text Explorer tools to determine what happened.

Warm up

Term	Count
wawrinka	21
djokovic	17
match	10
time	9
first	7
grand slam	7
win	6
us open	6
stan wawrinka	6
novak djokovic	6
almost	5
final	5
hours	5
one	5
player	5
set	5
sunday	5
third	5
big	4
champion	4
four	4
just	4
mentally	4
point	4
think	4
two	4

Warm up

Warm up

http://www.espn.com/tennis/usopen16/story/_/id/17525397/2016-us-open-stan-wawrinka-overcomes-novak-djokovic-4-sets-win-us-open

Outline

- Text Exploration in JMP® 13

- Demonstration – Workers' compensation

- Demonstration – National Science Foundation (NSF) awards: 2010 and 2015

- Outline best practices

Text Exploration – JMP® 13

- Import text data into JMP from multiple sources: database, text files, Excel.
- Determine most frequent terms and phrases.
- Quickly add phrases to the analysis.
- Use Recode to correct text errors and exploit synonyms.
- Employ JMP's dimensionality reduction, multivariate, and graphical capabilities to examine relationships between words and documents.
- Cluster documents and utilize topic analysis to classify similar themes into categories.
- Employ other JMP platforms (Partition, Graph Builder, regression, logistic regression, Cluster Variables) to determine if any frequent terms/phrases or underlying themes are related to structured data in the corpus.
- Use JMP's capabilities to (By, Data Filter) to subset the analysis by any category.

Quickly, turn data into information.

Text Exploration: Workers Compensation

- Data: 3,037 workers' compensation insurance claims.

- <u>Objective</u>: Use text exploration and document clustering techniques to prevalent themes in fraud cases.

- JMP® 13 Text Explorer:
  - Use of Show Text for contextual meaning.
  - String processing: stop words, synonyms, terms and phrases lists.
  - SVD plots.
  - Document clustering.
  - Use of Partition with the Document-Term-Matrix (DTM).

Reference: *Text Analytics Using SAS® Text Miner.* **SAS Institute: Cary, 2011.**

Text Exploration: NSF Awards

- Data: All 25,886 NSF grant awards from 2010 and 2015.

- <u>Objective</u>: Use text exploration and topic analysis to determine:
  - The top themes in these proposals.
  - If the top themes have any relationship to US geography or changed between 2010 and 2015.
  - If any of these top themes are awarded more money than others.
  - If there are any natural clusters of these themes.

- JMP® 13 Text Explorer:
  - Analysis of Document Topic Vectors via Graph Builder and the Fit Model and Cluster Variables platforms.
  - Use of Show Text for contextual meaning.

Text exploration – Best practices

- Do not limit yourself to structured data analysis. If you are collecting unstructured (text) data, do something with it! Text exploration is a great source of hypothesis generation.

- Use the Terms and Phrases Lists in JMP to clean text as well as filter words and phrases.
  - Correct misspellings and synonyms.
  - Create stop word list.
  - Identify multi-words (proposed research) and multiple word phrases of interest (motor vehicle accident).

- Utilize Show Text…too much.

- Analyze structured and unstructured data together.

- Understand analysis is continuous.

- Communicate the results effectively using JMP's graphical capabilities.

Summary

- Text Exploration in JMP® 13

- Demonstrations
  - Workers' compensation
  - National Science Foundation (NSF) awards: 2010 and 2015

- Best practices