



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

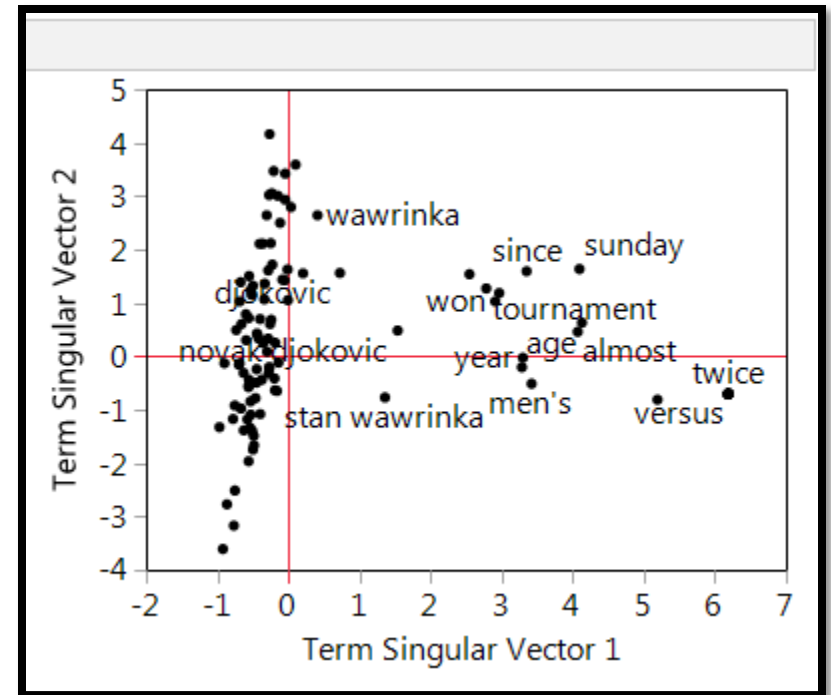
Warm up

- 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

The screenshot shows the ESPN website interface. At the top, there's a navigation bar with the ESPN logo and various sports categories: NFL, NBA, MLB, NCAAF, Soccer, NHL, Tennis, and a menu icon. Below this is a secondary navigation bar with 'Tennis' selected, and links for Home, Scores, Schedule, Rankings, Players, Grand Slam History, Shop, Tickets, and USOpen.org. The main content area features a large article titled 'Stan Wawrinka tops Novak Djokovic in US Open final for 3rd major title'. To the left of the article is a sidebar with several news snippets, including one about a Jaguar F-Pace car and several sports-related headlines. Below the article title is a video player showing Wawrinka celebrating with his trophy. A score overlay on the video shows the match results: Novak Djokovic (7-4-5-3) vs Stan Wawrinka (6-6-7-6). Below the video are social media sharing icons for Facebook, Twitter, and Pinterest, along with a date stamp 'Sep 12, 2016'.

ESPN NFL NBA MLB NCAAF Soccer NHL Tennis ... Watch Listen Fantasy

Tennis Home Scores Schedule Rankings Players Grand Slam History Shop Tickets USOpen.org

THE ALL-NEW F-PACE

#MANNERUP JAGUAR

EXPLORE F-PACE

Stan Wawrinka tops Novak Djokovic in US Open final for 3rd major title

US OPEN MEN'S FINAL

Novak Djokovic	7	4	5	3
Stan Wawrinka	6	6	7	6

Wawrinka: 3-0 in grand slam finals vs world No. 1

ESPN

Master Sgt. Cedric King on Colin Kaepernick: 'Kneeling will not stop the injustice'

3h - David Newton

WR Cruz jabs Norman for not covering Brown

NEW YORK GIANTS 13m - Jordan Raanan

ACC moving neutral-site title games from N.C.

WAKE FOREST DEMON DEACONS 8m - Andrea Adelson

Big Ben: Steelers don't want injury payback

PITTSBURGH STEELERS 2h - Jeremy Fowler

CFP chaos theories: Worst-case scenarios for each conference

ALABAMA CRIMSON TIDE 7h - Heather Dinich

1 Related

Uwe von Schamann and the kick that beat Ohio State

OKLAHOMA SOONERS 8h - Jake Trotter

Zimmer might not tell Vikings who'll start

ESPN.com news services Sep 12, 2016

NEW YORK -- [Stan Wawrinka](#) is the first to acknowledge he hasn't always been the most consistent player -- or the strongest mentally. That's why when he shows his mettle during a match he likes to point his right index finger to his temple.

That signature gesture got a lot of use in the US Open final Sunday, when Wawrinka surprisingly managed to wear down [Novak Djokovic](#) and beat the defending champion 6-7 (1), 6-4, 7-5, 6-3 for his first US Open title and third

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.
- Objective: 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).

Text Exploration: NSF Awards

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