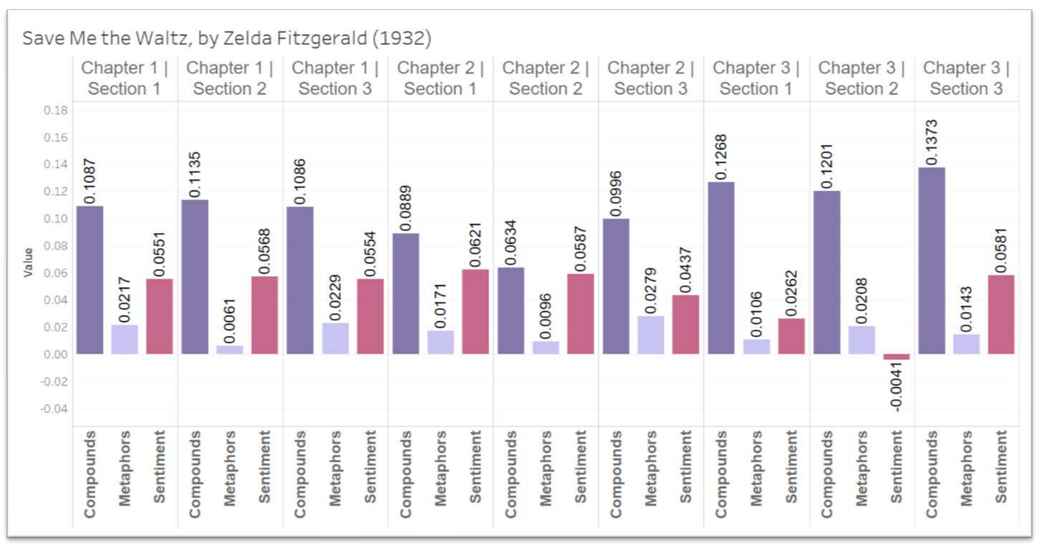
Another Hammer in the Humanities Librarians' Toolkit: **Text Analysis/Mining**

Working with faculty members on text analysis/mining projects has enhanced the status of humanities librarians by giving them entrée to more collaborative projects.

Josh Been – Eileen Bentsen – Bill Hair

BAYLOR University Libraries

Use of Language and Sentiment in Save me the Waltz

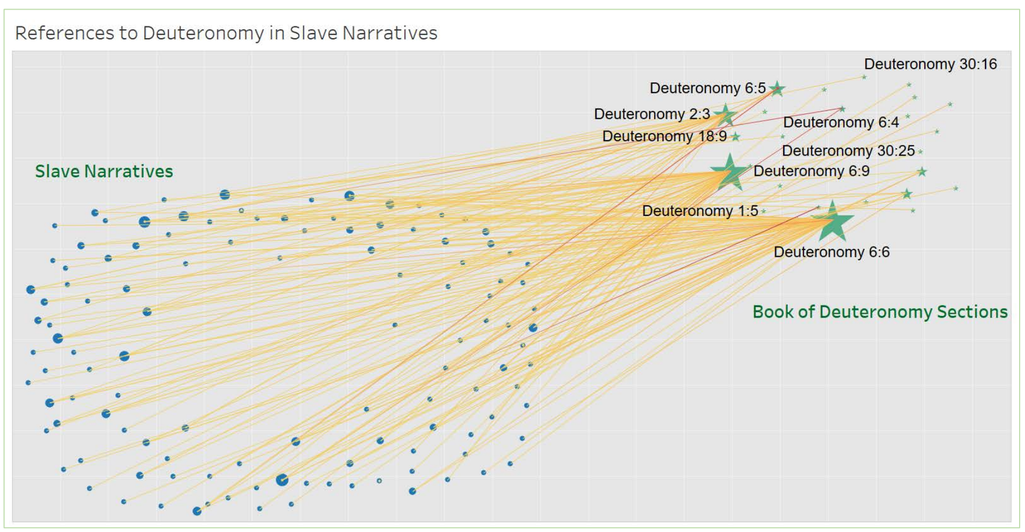


Identifying compound sentences, metaphor, and sentiment for each sentence.

Source: HTML Online
Calculation Method: (1) Compound: FANBOYS,
(2) Metaphor: Two noun definition similarity,
(3) Sentiment: VADERSentiment (code) (tabular results)

First text analysis project we undertook! - Introduced English faculty to the library's digital humanities services and gave humanities librarian the confidence to promote these services.

Fugitive and Former Slave Narrative References to Deuteronomy

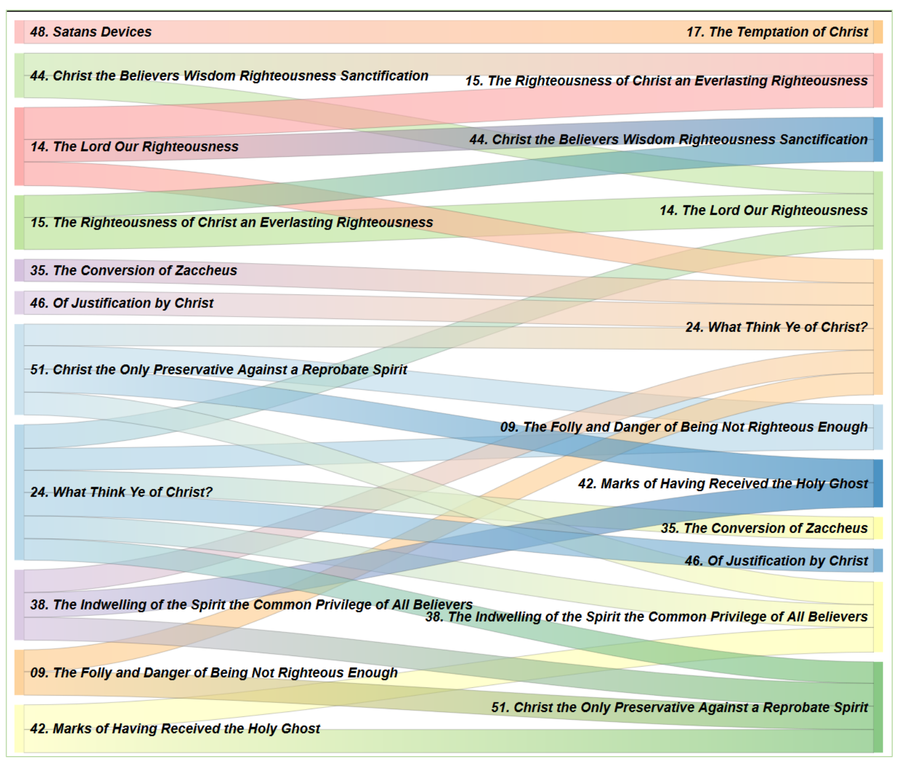


Identify biblical references from transcribed fugitive and former slave narratives

Source: UNC: Documenting the American South – North American Slave Narratives
Calculation Method: Fuzzy N-gram similarities (code) (tabular results)
Visualization Method: Circular network created using NodeXL and visualized using Tableau.

Opened doors to more collaborative projects with Seminary faculty, including a department-wide digital humanities demonstration and additional faculty research partnerships around text analysis.

Similarities Between George Whitfield's Sermons



Measure the similarities between 57 of George Whitfield's sermons with each other

Source: CRTA (Center for Reformed Theology and Apologetics)
Calculation Method: TF-IDF Cosine Difference (code) (tabular results)
Visualization Method: Sankey Chart using Google Charts JavaScript Classes. All similarity scores .35 or higher (35% similar) displayed.

This project was done in partnership with a music graduate student, and this is opening doors to topic modeling, similarity analysis, and tf-idf keyword generation across music lyrics.