PROBLEM
The Bible is an epic account of humanity that spans 1600 years of writing and over 40 authors. Its daily readership numbers hundreds of millions, and its profound influence and cultural significance has made it a subject of research in linguistics, history, and religion among many other fields.

However, the Bible as a physical document is unwieldy: with 66 chapters and over 700,000 words, it is difficult—even for experts—to traverse methodically.

How can we stitch together the content of the Bible, its immense contextual metadata, and interactive visualization techniques to facilitate exploration and illuminate understanding?

MOTIVATION
Without a higher-level view of the Bible, trends and themes are hidden behind pages of text. Organizational metadata such as books and key events are similarly tethered to pages, and are not leveraged to their fullest potential.

We can illustrate the problem with a task: locating a theme in the Bible, e.g. which passages discuss love? With just the text and no cognitive aids, this is very difficult to do. First, we need a high-level view to see the hot spots where love occurs throughout the Bible. Among these, we need to be able to compare local use cases to see which spots contribute an appropriate treatment of love. But how do we define local? How can we empower the user to systemically partition the Bible, or give them the freedom to define scope? Finally, we need to link back to the text.

The challenge becomes clear; we need a single interface that supports multiple zoom-levels, granularities, and comparisons across text, all while providing a contextual information trail for rapid exploration.

APPROACH

Text Processing
Initially, we stripped the Bible text of stop words and punctuation. We then clustered the remaining words into contiguous groupings of 300 words. Finally, we associated each grouping of words with relevant metadata such as the start and end verses of the grouping and the frequency breakdown by verse of each word within the grouping.

In addition, we sliced the Bible text into a format that allowed for easy and fast lookup of the text of a verse by its name, so that we could complement the user’s exploration by efficiently displaying different contexts of a word’s usage.

Visualization Techniques

Bird’s eye
- Animated bars: quickly locate regions of interest, rapid search and iteration
- Small multiples: compare search query snapshots
- Annotations: x-axis labeled with books and key events
- Book segmentation: where in the Bible are we?

Middle ground
- Verse tooltips: relating bars to verse occurrence
- Contextual brushing: verse range and aggregate statistics
- Multi-brushing: compare results across different ranges
- Book-brushing: filter results by book
- Concordances: view local context of word in selected range (either book or user-defined), or comparison across multi-brushes

Zeroing in
- Verse linking: bottoming out at the source text, linking from bars and concordance lists

Technology
Python: text processing, producing JSON structures
NLTK: filtering stop words, removing punctuation
D3: charting, brushing, small multiples, book segmentation
jQuery: multi-brushing, concordances, tooltip

FUTURE DIRECTIONS
- Interactive techniques targeting direct comparison of multiple words (co-occurrence)
- Search for phrases with regular expression, wildcard
- Edition and language support, approximation across languages (equivalences)
- Further NLP analysis within selected regions, comparison metrics