

TheBookofMormon.csv

Maya Rowen

We took an extensive sample of the terms indicating positive themes and a similar sample of terms indicating negative themes and counted the occurrences of these terms throughout the *Book of Mormon*.

Positive:

[1]	"love"	"loving"	"loves"	"loved"	"joy"
[6]	"joyous"	"joyful"	"joyfully"	"triumph"	"triumphant"
[11]	"triumphs"	"triumphed"	"happy"	"happiness"	"happily"
[16]	"kind"	"kindness"	"kindly"	"beauty"	"beautiful"
[21]	"beautifully"	"good"	"goodly"	"goodness"	"heaven"
[26]	"heavenly"	"heavens"	"laugh"	"laughter"	"laugh"
[31]	"fun"	"adore"	"adores"	"adored"	"adoring"
[36]	"holy"	"holiness"	"friend"	"friends"	"friendship"
[41]	"hope"	"greatness"	"gratefully"	"grateful"	"gratitude"

Negative:

[1]	"hell"	"hellish"	"satan"	"devil"	"devilish"
[6]	"wicked"	"wickedly"	"wickedness"	"evil"	"evilly"
[11]	"evilness"	"evil"	"bad"	"badness"	"enemy"
[16]	"enemies"	"nemesis"	"nemesis"	"kill"	"killer"
[21]	"killers"	"killed"	"killing"	"hurt"	"die"
[26]	"dying"	"died"	"death"	"deathly"	"illness"
[31]	"ill"	"sick"	"sickly"	"sickness"	"ruin"
[36]	"ruins"	"ruined"	"disease"	"diseases"	"diseased"
[41]	"mutilate"	"mutilated"	"mutilates"	"distress"	"distressed"
[46]	"distresses"	"destroy"	"destroys"	"destroyed"	"destruction"
[51]	"destruction"	"slay"	"slays"	"slain"	"slew"
[56]	"slayer"	"slayers"	"murder"	"murders"	"murdered"
[61]	"murderers"	"injured"	"injure"	"injures"	"fear"
[66]	"fears"	"feared"	"hatred"	"hate"	"hating"
[71]	"hates"	"hated"	"hateful"	"anger"	"angry"
[76]	"mad"	"madness"	"hopeless"	"hopelessly"	"violent"
[81]	"violence"	"violently"			

Reducing our samples to the top 20 most frequent terms in each category, we plotted the number of incidences of terms in each category per book.

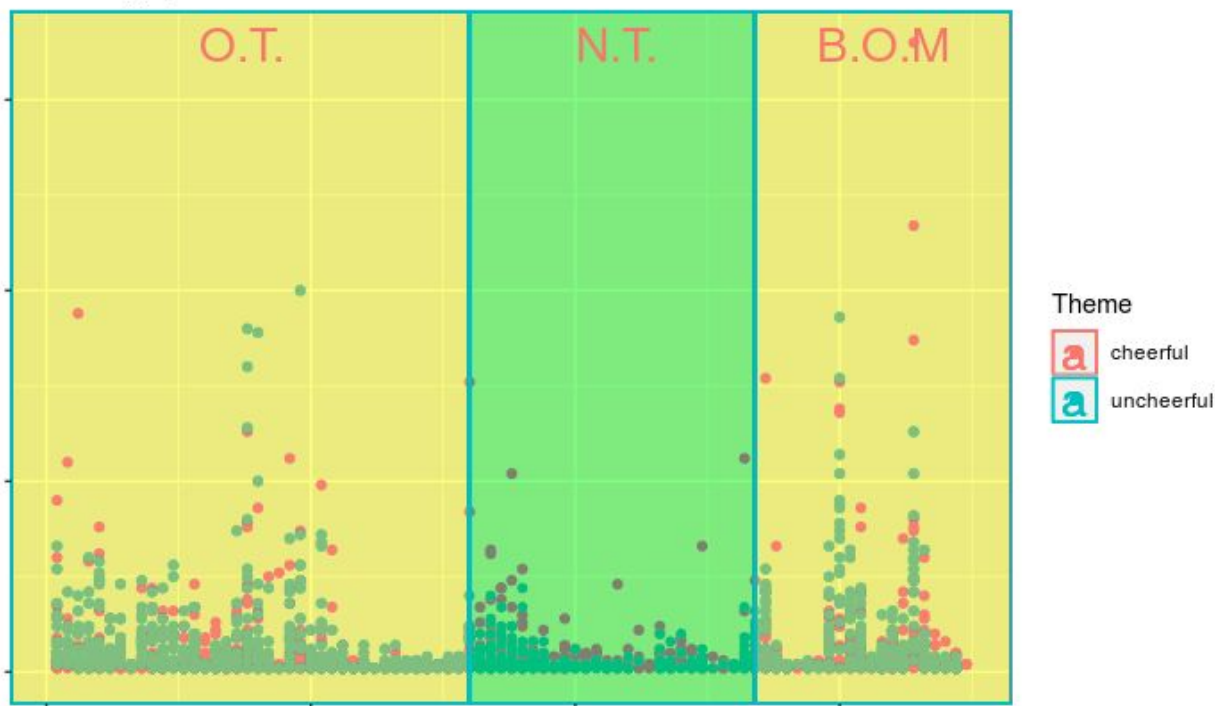
Positive:

[1] holy good heaven love joy heaven's hope loved
 [9] kind friends goodness holiness greatness friend beauty
 kindness
 [17] heavenly beautiful goodly happy

Negative:

[1] evil death fear wicked destroy enemies
 [7] die anger slain wickedness destroyed died
 [13] destruction slay slew devil enemy satan
 [19] kill hell

How Happy is the Bible



The image above suggests that “The Old Testament” and “The Book of Mormon” alternate between optimistic and pessimistic verses, whereas “The New Testament” tends to be more upbeat overall. Prior to this analysis, we would have predicted that “New Testament” and “Book of Mormon” would be more cheerful in nature, while “the Old Testament” would be a bit darker, considering the plagues and brothers killing brothers and snakes in the garden.

This study primarily examined trends in verse structure and word frequency in the Book of Mormon. The Book of Mormon was converted to csv format, with each row designating a unique verse, with columns indicating which chapter, book, and testament a given verse resides in. Notably, the three ‘testaments’ in the *Book of Mormon* are the standard Old Testament and New Testament from the traditional Christian bible as well as the addition of the “Book of Mormon”, which was allegedly translated by the late Joseph Smith from some golden tablets he discovered in his lovely garden in Rochester, New York, in 1820 C.E. Per Smith, the text on the tablets was writing of God himself, and the prophet Mormon buried them circa 385 C.E. They lay underground for

The variable “scripture_text” exists as the entire written verse, in string format (in English). Each Additional columns indicate the chapter, book, and volume (testament) that the scripture resides in. In a physical copy of the bible, the verses, chapters, and books are numbered such that each book begins with Chapter 1: verse 1, and each book is indexed by a number within it’s testament but more commonly referred to by a unique name. For instance “Exodus” is the second book of the Old Testament.

Verses, chapters, books, and volumes are each also indexed by “verse_id”, “chapter_id”, “book_id”, and “volume_id”, respectively. Each of these columns gives a natural number based on the order the element appear in the *Book of Mormon*. So calling a verse_id of one, we have

```
``{r}
verse_id_1<-filter(lds1, verse_id==1)
verse_id_1$scripture_text
``
```

```
[1] IN the beginning God created the heaven and the earth.
41659 Levels: "By which also he went and preached unto the spirits in prison; ...
```

A summary of the “id” columns indicates the number of each category (volume, book, chapter, and verse):

volume_id	book_id	chapter_id	verse_id
Min. :1.000	Min. : 1.0	Min. : 1.0	Min. : 1
1st Qu.:1.000	1st Qu.:13.0	1st Qu.: 344.0	1st Qu.:10500
Median :1.000	Median :26.0	Median : 824.0	Median :20998
Mean :1.826	Mean :36.7	Mean : 790.3	Mean :20998
3rd Qu.:3.000	3rd Qu.:67.0	3rd Qu.:1204.0	3rd Qu.:31496
Max. :5.000	Max. :87.0	Max. :1582.0	Max. :41995

Meanwhile, commanding a summary of the “name” columns, which refer to elements as they would be in a hard copy of the *Book of Mormon* (volume_title, book_title, chapter_number, and verse_number) yields the following result:

volume_title		book_title	chapter_number	verse_number
Book of Mormon	: 6604	Doctrine and Covenants:	3654	Min. : 1.00
Min. :	1.0			
Doctrine and Covenants:	3654	Psalms	: 2461	1st Qu.: 7.00
1st Qu.:	7.0			
New Testament	: 7957	Alma	: 1975	Median : 15.00
Median :	15.0			
Old Testament	:23145	Genesis	: 1533	Mean : 24.63
Mean :	18.2			
Pearl of Great Price	: 635	Jeremiah	: 1364	3rd Qu.: 29.00
3rd Qu.:	24.5			
		Isaiah	: 1292	Max. :150.00
Max. :	176.0			
		(Other)	:29716	

Both systems are sufficient for analyzing the text.

Here’s one way to call the fourth verse in the second chapter of “Exodus”:

```
``{r}
Exodus.1.4<-filter(lds1, book_title=="Exodus",chapter_number==2,verse_number==4)
script1<-Exodus.1.4$scripture_text
```
```

And here’s another way to call the same verse:

```
``{r}
last_verse<-nrow(filter(lds1,
book_title=="Genesis"|book_title=="Exodus"&(chapter_number==1|chapter_number==2&(verse
_number<4))))
is_verse_id<-last_verse+1
is_row<-filter(lds, verse_id==is_verse_id)
script2<-is_row$scripture_text
```
```

Testing the sameness of the objects summoned above we have

```
``{r}
script1
```

```
script2
identical(script1,script2)
'''
```

And the output is:

```
[1] And his sister stood afar off, to wit what would be done to him.
41659 Levels: "By which also he went and preached unto the spirits in prison; ...
[1] And his sister stood afar off, to wit what would be done to him.
41659 Levels: "By which also he went and preached unto the spirits in prison; ...
[1] TRUE
```

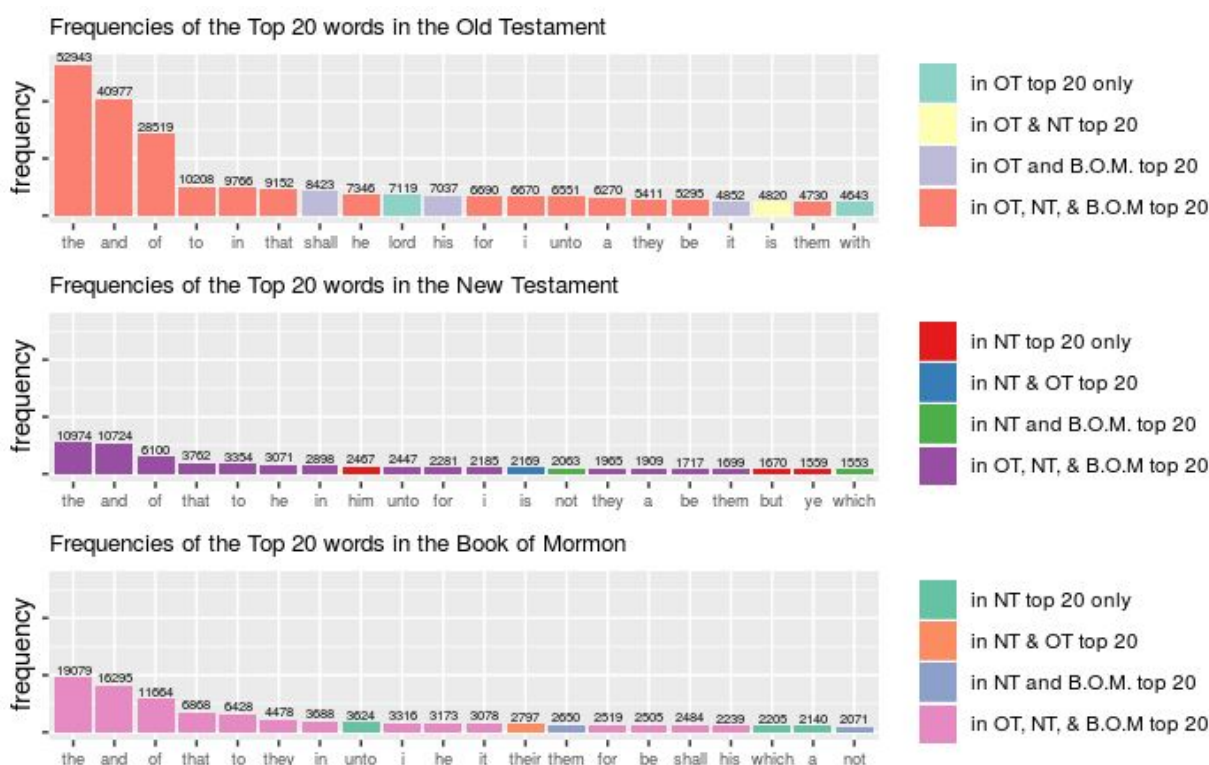
Despite the fact that both methods work and give identical results, the “id” system is more efficient and especially favorable in the context of building ggplots.

The primary focuses of our visual analysis of *The Book of Mormon* are word counts, word frequencies, and trends in word use over the course of the text. Although it is often sold (or more probably, given away coercively by over-zealous missionaries) as a single, unified religious text, *The Book of Mormon* is particularly interesting because the three testaments were written in three radically different eras, by multiple groups of people who interacted in no way besides the contemporaries studying the written work of their formers. Indeed, the “Old Testament” dates back as far as 1200 B.C.E, and serves as sacred scriptures in the Jewish Faith as well as part of the bible in most sects of Christianity. *The New Testament*, which is the second portion of the bible and an important text for most Christians was written in the first 100 years of the common era. *The Book of Mormon*: “Another Testament of Jesus Christ”, was either first conceived of by Joseph Smith in the 19th century C.E. or else written by God a long while beforehand and recorded on golden tablets by some helpful angels. Regardless, the variations in the contexts in which the three portions of the *Book of Mormon* were written inspire curiosity regarding nuances and differences in the word use and diversity as well as character presence and broader themes. across segments of the text. We begin our analysis with a look at word counts and frequencies at the verse level.

To detect and count words, we use the “tokenizers” package available in R. The “tokenize_words” function separates entries in the “scripture_text” column into lists of words, and allows for repetitions. Thus, running the function in the entire column produces a list of lists of words: a nested list, as such:

```
[1] "in"      "the"      "beginning" "god"      "created"  "the"
     "heaven" "and"      "the"      "earth"
```

To create a tokenized vector of all words in *The Book of Mormon*, we unlist our nested list. By doing so, we sacrifice the property that the words are sorted by verse, but we are able to investigate the overall frequency of words in each testament as well as the overlap in the most commonly used word across testaments.



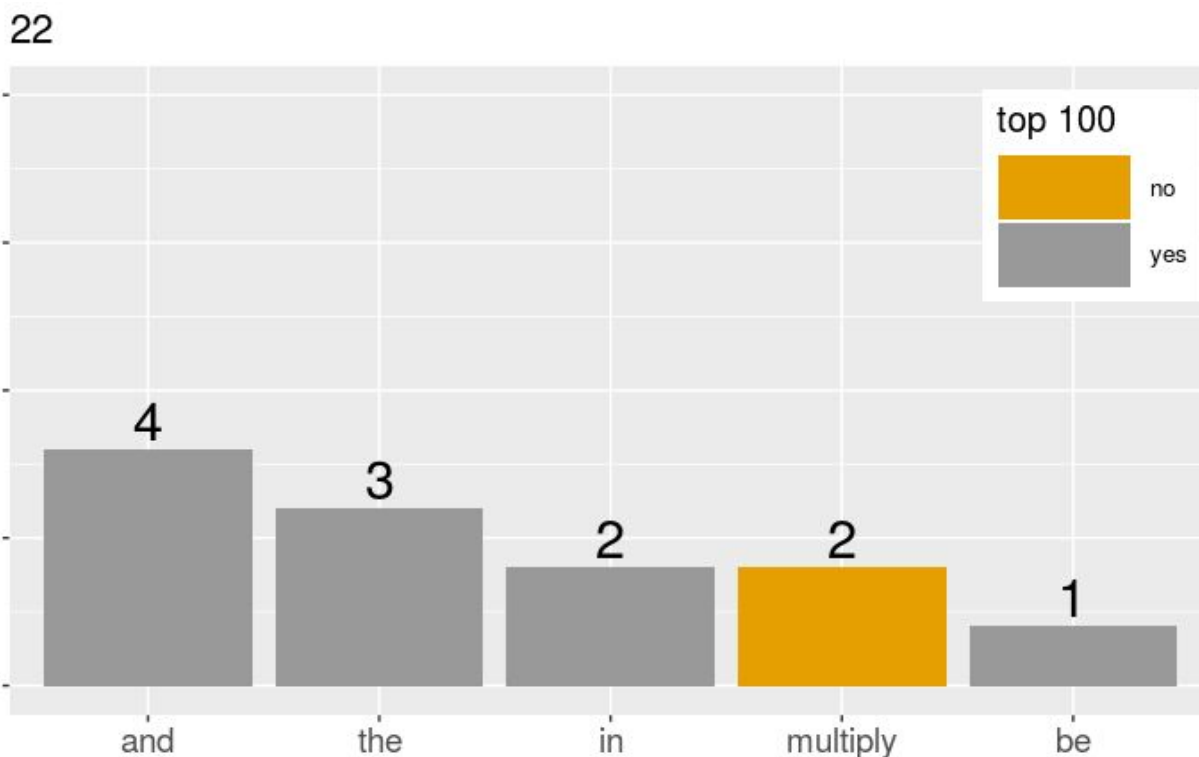
From the column charts above, we observe that articles and prepositions such as “the”, “and”, “of”, “that” and “a” are the most used words in each testament. Notably, “The Old Testament” has a larger concentration of its words falling into the top three columns (“the”, “and”, and “of”). This is a logical result, since language expands overtime, so the writers of “The Old Testament” were most limited in their choice of articles and prepositions.

This graph also suggests the differences in lengths and word counts between the testaments; “the Old Testament” is by far the longest of the three, followed by the “Book of Mormon”

Another interesting insight from the graphic above lies in the unique words, i.e., the words in the categories “in OT top 20 only”, “in NT top 20 only”, and “in B.O.M. top 20 only”. For instance,

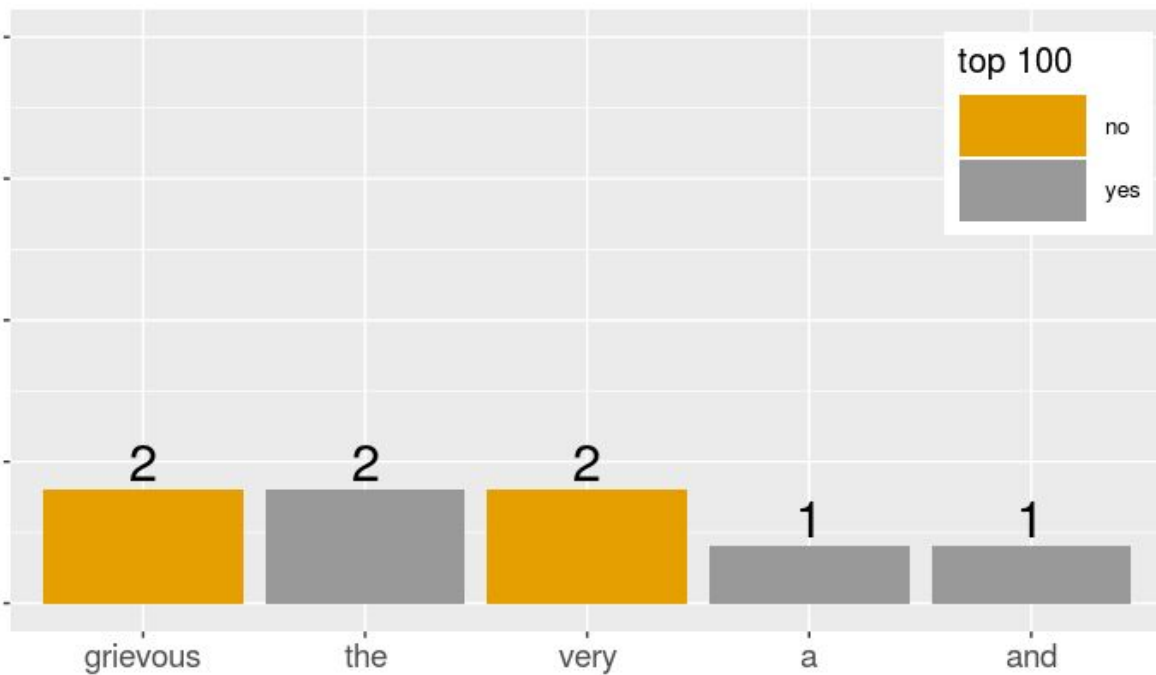
the word “ye” makes the top 20 in only “the New Testament”. In today’s terms, “ye” is antiquated, but it appears that it peaked in use sometime after “the Old Testament” was translated to English, and began to taper before “The Book of Mormon” took America by storm in 1823. The only top 20 word in any testament that is not an article, preposition, pronoun, or very common verb is “lord”, which ranks ninth in “The Old Testament” only.

We can also use our nested list of tokenized words to look at the distribution of common words in any verse:



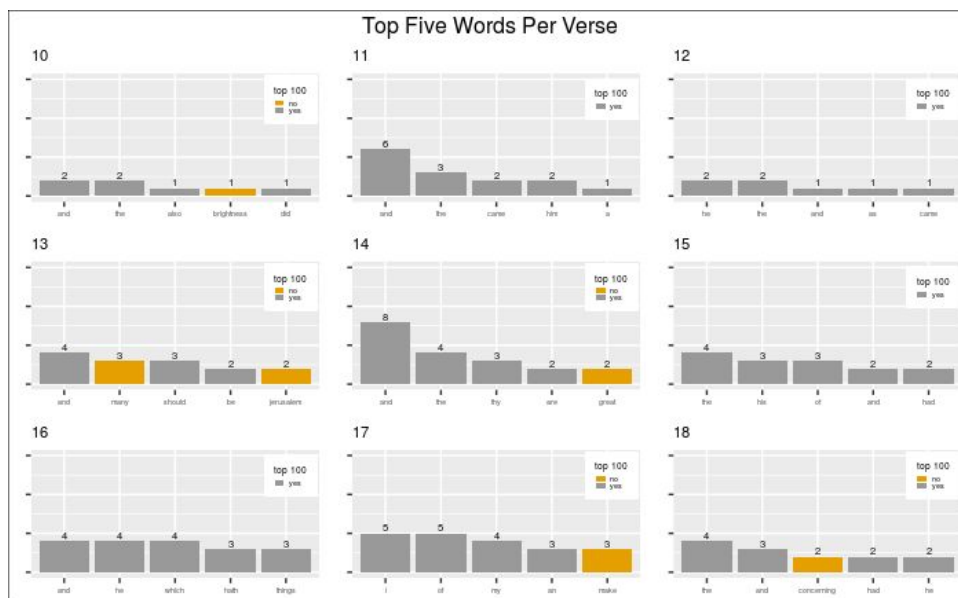
Constructing a column chart of the top five words and respective word counts in Verse 22, we observe that the first, second, third, and fifth words in the verse are among the top 100 most common words in the entire *Book of Mormon*. Number four, however, “multiply”, is a more rare word that makes the top five in verse 22.....think “*be fruitful and multiply*”–God.

40050

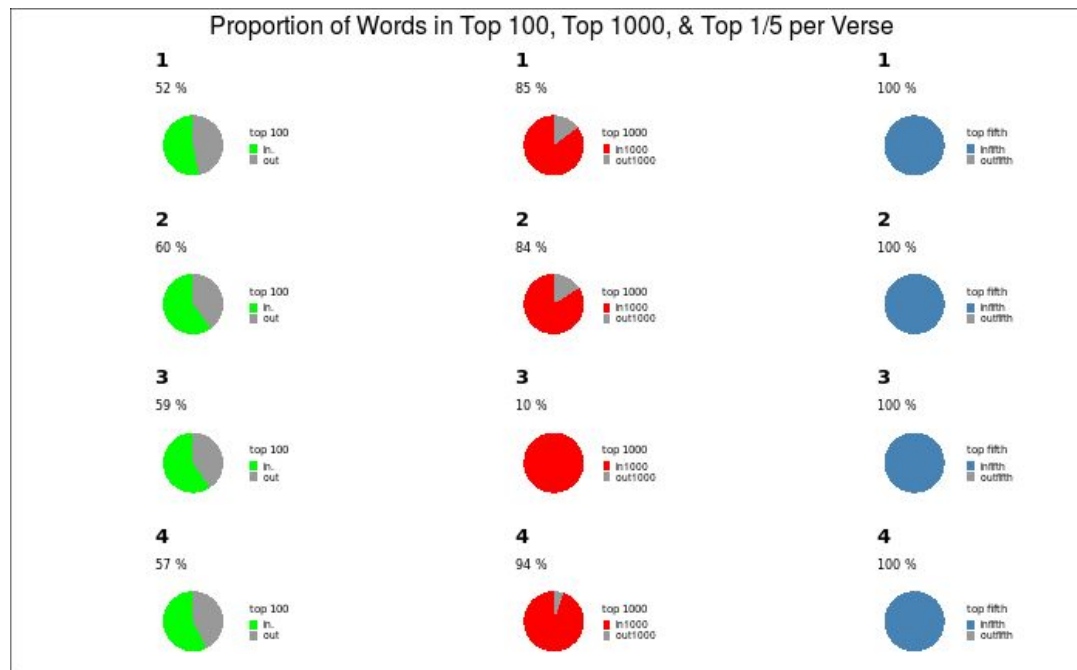


Verse 40050 deviates a bit more from our most favored word selection. Here, we see that the verse used the word “grievous” twice and “very” twice. Something very grievous must be happening in verse 40050.

We can use compare the top five words in a sequence of verses as well:



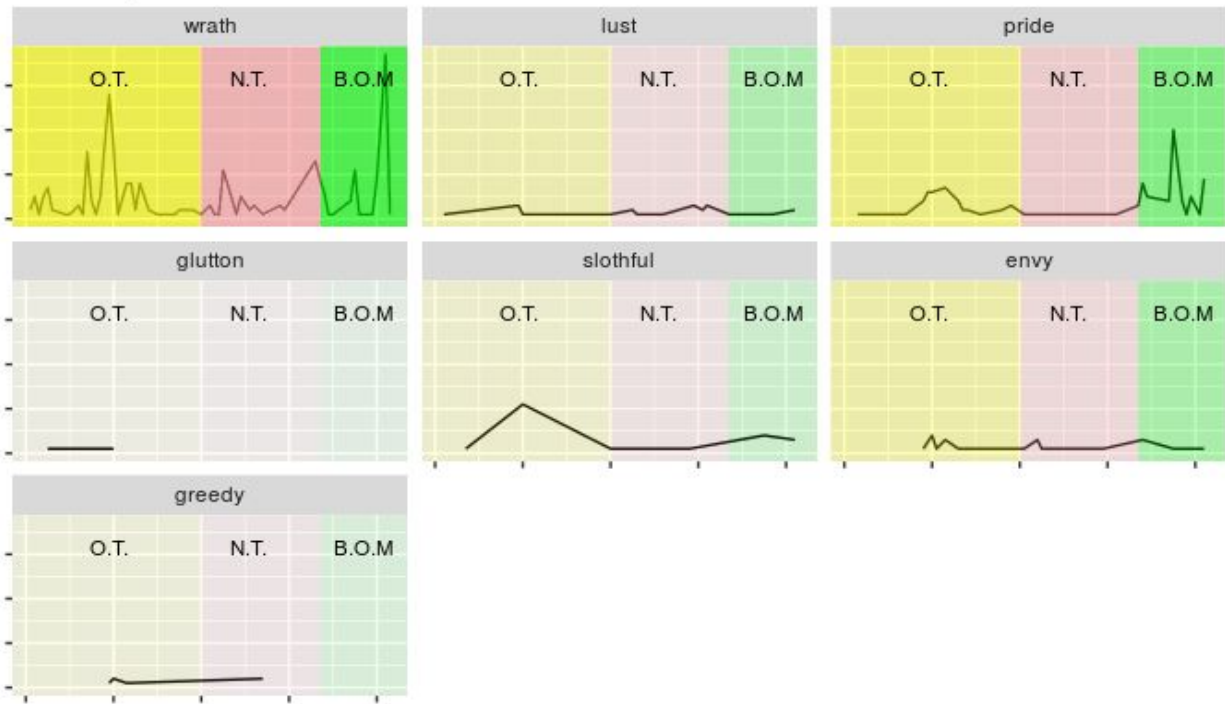
We use pie charts to investigate the proportions of words in each that fall into the most common 100, and 1000, and ~8,000 (top 1/5) words in the *Book of Mormon*.



In the first four verses, all words qualify as top 20 percenters, and nearly all words meet that top 1,000 criteria. Between $\frac{1}{2}$ and $\frac{2}{3}$ of words are in the top 100. That seems like a pretty large overall proportion. Since the first four verses are all members of “the old testament, this supports our hypothesis that “the Old Testament” was written, or translated at least, under a condition of a limited English vocabulary.

Another topic to consider is the journey of characters and theme throughout the text. For example, we use a line graph to track mentions of the seven deadly sins across the three testaments.

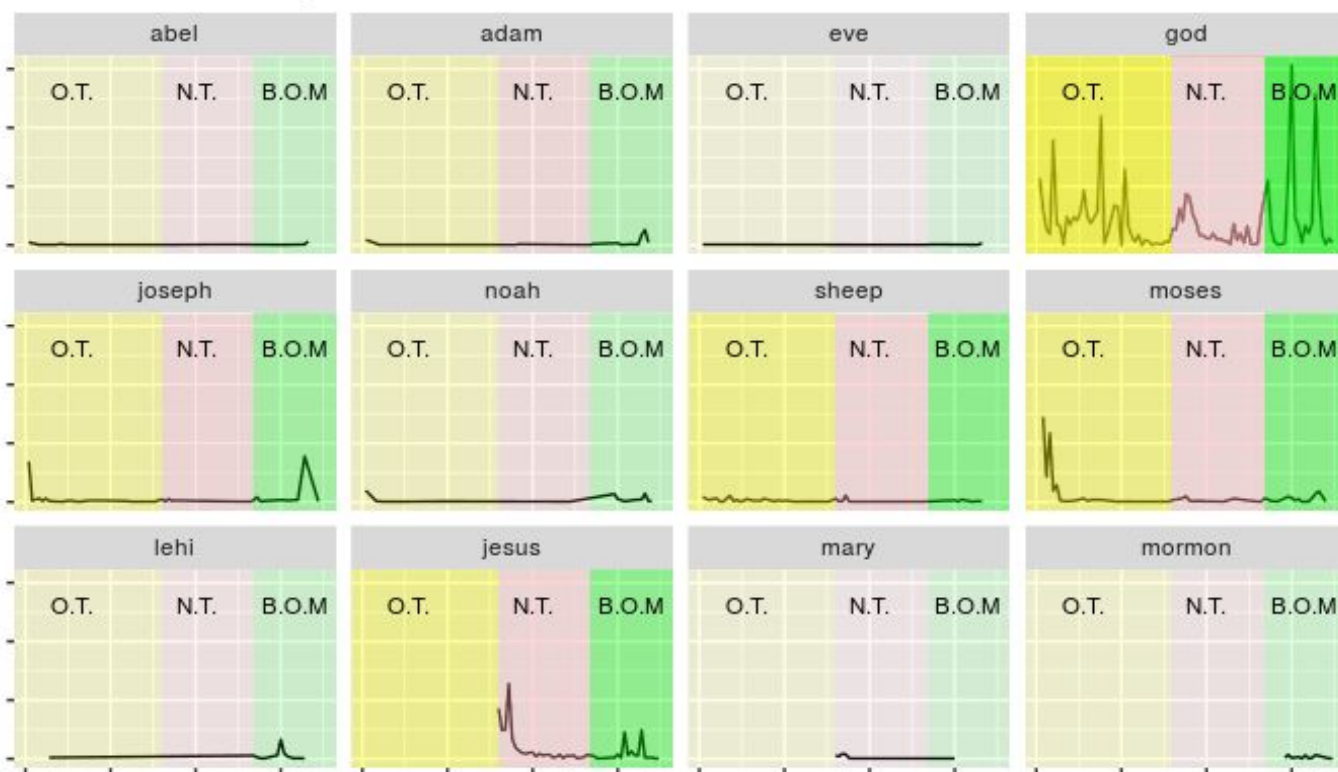
7 Deadly Sins



Wrath is clearly the most mentioned sin in all three testaments. Other sins, such as pride and sloth, are more topical in single testament, whereas lust is more consistently relevant. Lust is a consistent problem.

We run a similar analysis on famous biblical characters such as Jesus, God, Adam, and Mormon across all three testaments:

Character Development



Some characters, such as Mary and Mormon, are entirely isolated to one testament. Others, such as Moses and Abel have a spike in whichever Testament they are most relevant, but also receive occasional mentions throughout the text.

Joseph and Adam are New and Old testament characters that happen to share names “Book of Mormon characters. Another fascinating cause of names reappearing in “The Book of Mormon” may be as follows:

“The Book of Mormon” takes the bible to America, and it’s time frame falls before, during, and in between Old and New Testaments. It recounts a history of Ancient people (Adam and Eve, etc.) originating in the Garden of Eden (present Jackson County, Missouri). Overtime, however, their tribes abandoned America for Israel and the Middleeast. It was there that the rest of the old testament stories took place, and that Jesus lived and died. At some point, however, Ancient Israelites travelled back to America on homemade kayaks and settled there.

When Jesus was resurrected, he appeared before these ‘Native Americans’, and taught them his holy ways. Unfortunately, however, the Native Americans were a divided people, and the Lamanites killed all of the Nephites. As time wore on, the wicked Lamanites forgot the blessed

teachings of Jesus and his true intentions were buried underground in upstate New York until the prophet Joseph Smith dug up the golden tablets and rekindled faith in the Latter Day Saints of Jesus Christ, or as most people know it today, Mormonism.

And that's why Noah and Lehi are mentioned throughout the text, with observable spikes in "The Book of Mormon".

Sheep are around pretty consistently too.

References:

<https://programminghistorian.org/en/lessons/basic-text-processing-in-r>
http://www.cookbook-r.com/Manipulating_data/Renaming_levels_of_a_factor/
<https://www.datanovia.com/en/blog/ggplot-legend-title-position-and-labels/#change-legend-title>
https://ggplot2.tidyverse.org/reference/aes_colour_fill_alpha.html
<https://stackoverflow.com/questions/12187891/how-to-get-top-n-companies-from-a-data-frame-in-decreasing-order>
http://www.cookbook-r.com/Graphs/Shapes_and_line_types/
http://ggplot.yhathq.com/docs/geom_line.html
<http://zevross.com/blog/2014/08/04/beautiful-plotting-in-r-a-ggplot2-cheatsheet-3/#add-x-and-y-axis-labels-labs-xlab>
<https://www.datamentor.io/r-programming/list/>

The Data:

<http://docs.nephi.org/scriptures/downloads/lds-scriptures.csv.zip>