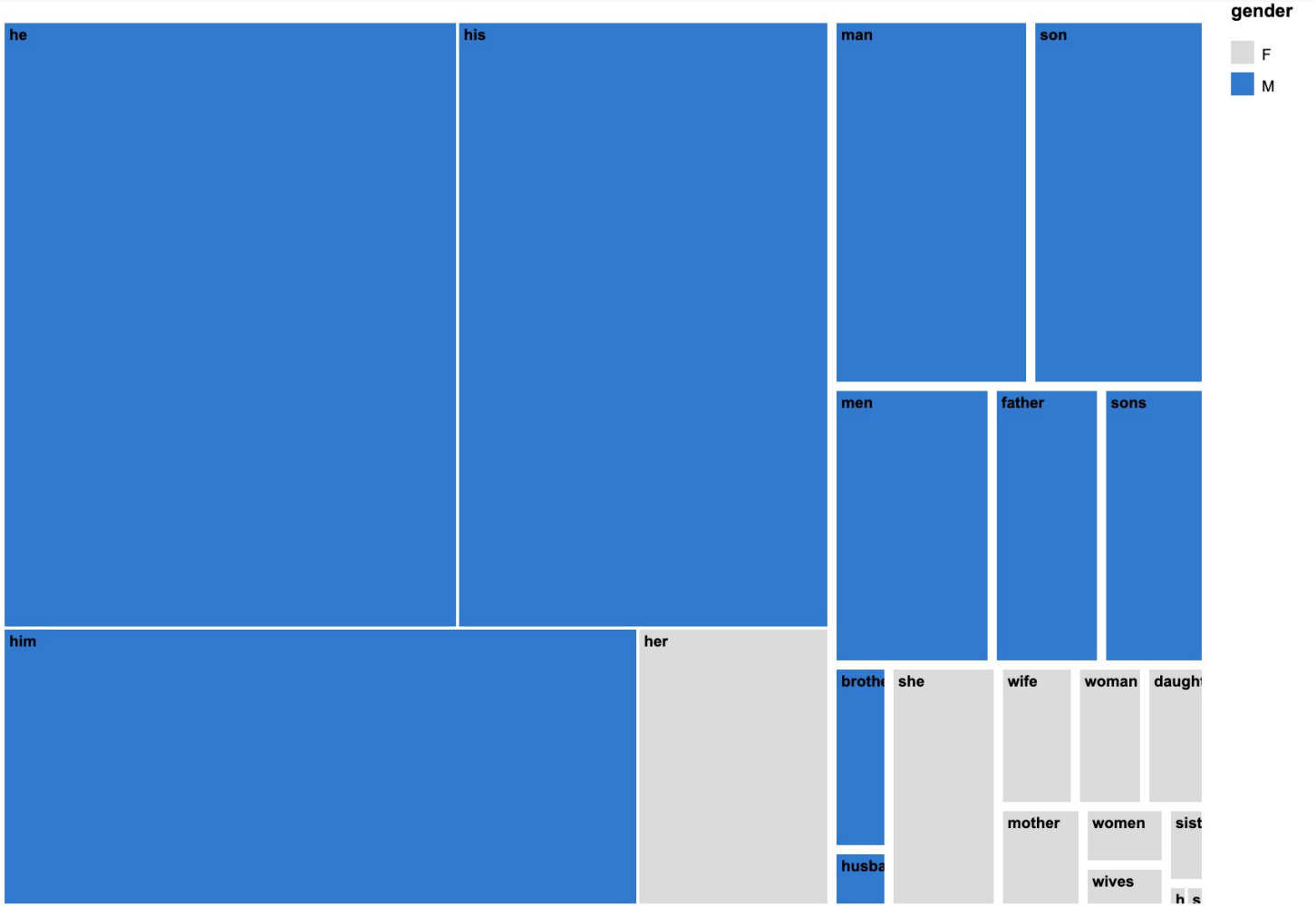
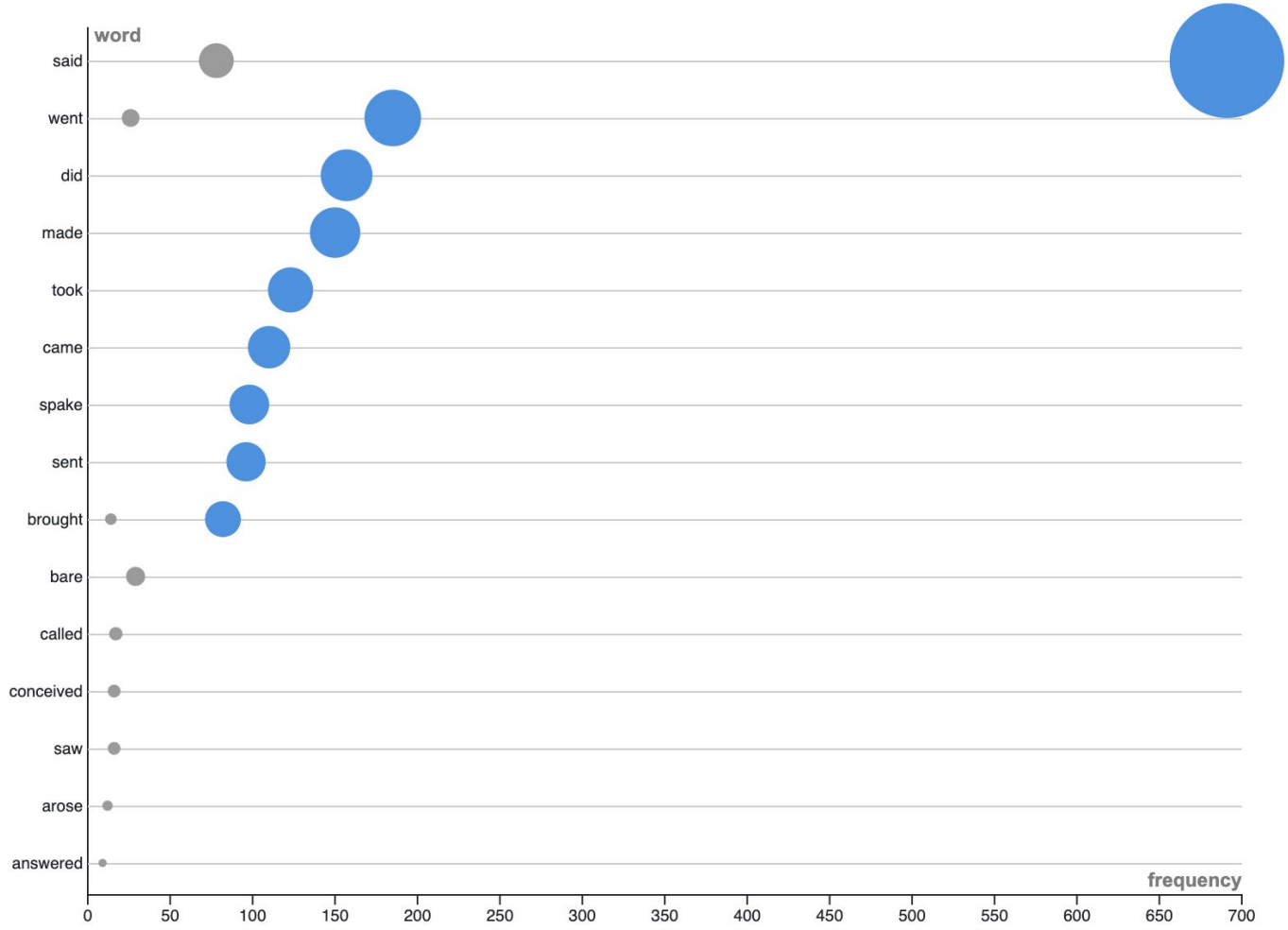
Investigating the representation of men and women in “The King James Version of the Bible” using NLP

Leanne Suen Fa
CART451 Exercise 2

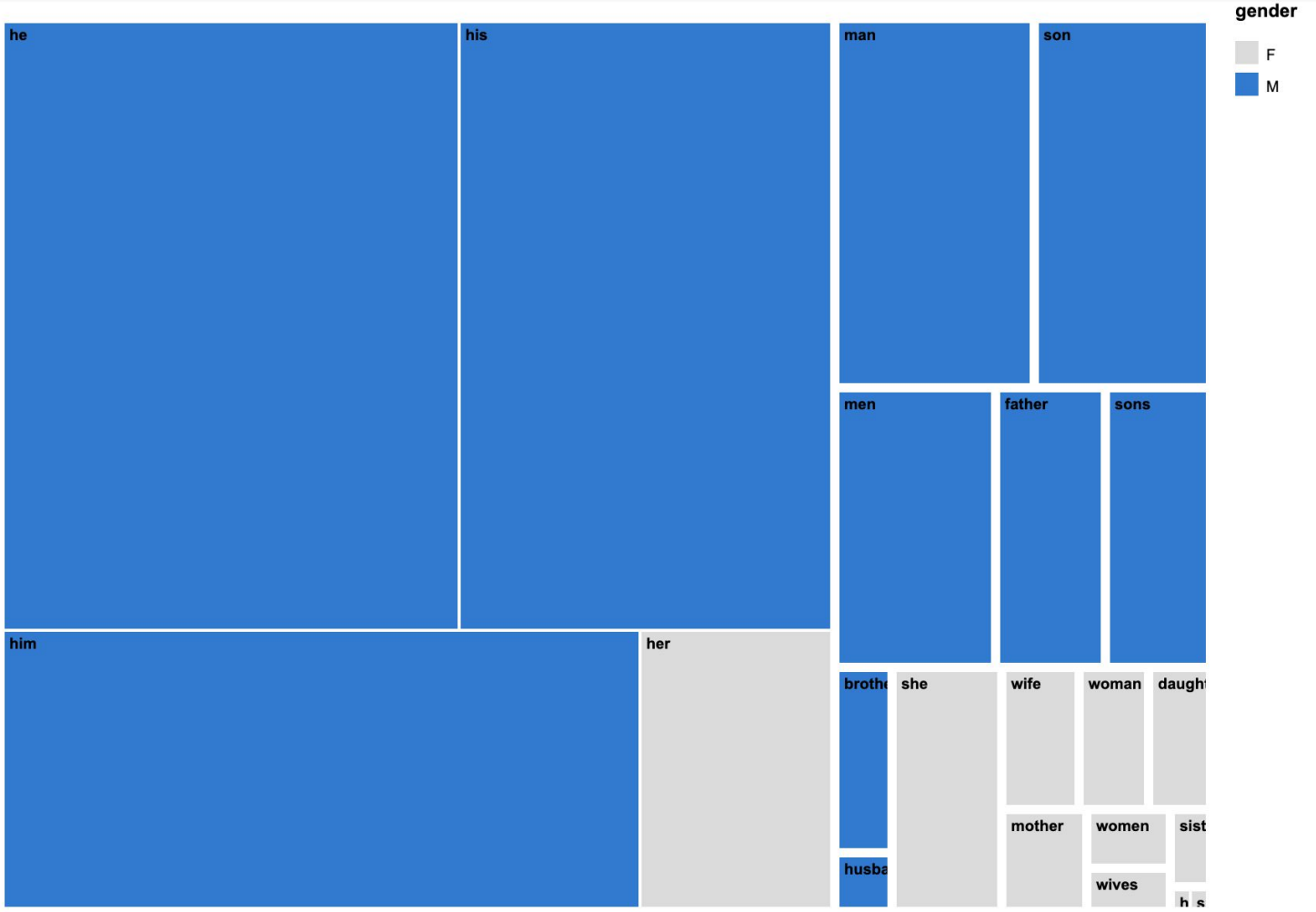

gender
F
M
he
his
man
son
men
father
sons
him
her
brothe
she
wife
woman
daugh
mother
women
sist
husba
wives
h
s


word
said
went
did
made
took
came
spake
sent
brought
bare
called
conceived
saw
arose
answered
frequency
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700

Process

1. Upload text (The King James Bible)
2. Calculate word count to find most commonly appearing words.
    a. Observation : “he” is the 7th most common word determinants and propositions
    b. Compare with frequencies between male and female pronouns and nouns
3. Calculate TFIDF
    a. Compare between male and female pronouns and nouns.
        i. TFIDF of “he” = 0 while TFID of “she” = 301
4. Collect data from terminal into Google Sheet
5. Product Tree Map with Raw Graph to show hierarchies and proportions

	word co	ranl	tfidf	gend
he	10422	6	0.0	M
his	8473	10	0.0	M
lord	7964	11	2443.8	M
him	6659	15	0.0	M
man	2735	43	839.2	M
king	2542	47	780.0	M
son	2392	48	734.0	M
her	1994	61	0.0	F
men	1677	78	514.6	M
father	1127	112	345.8	M
sons	1094	114	335.7	M
she	982	127	301.3	F
wife	407	254	124.9	F
brother	402	257	123.4	M
woman	365	273	112.0	F
kings	340	292	104.3	M
daughter	327	304	100.3	F
mother	320	306	98.2	F
women	181	490	55.5	F
wives	133	613	40.8	F
husband	126	636	38.7	M
sister	116	675	35.6	F


gender
F
M
he
his
man
son
men
father
sons
him
her
brothe
she
wife
woman
daught
mother
women
sist
husba
wives
h
s

Inferences

In general, male nouns and pronouns happen more often than female ones.

Even though both “he” and “she” are both pronouns, there is a huge difference in their TFIDF values (0 and 301 respectively).

- since the TFIDF calculation weigh down very frequent terms to decrease the importance of determiners like “the”, “she” occurs so rarely in the text (10,422 vs 982) that it is deemed to have a higher TFIDF

“He” occurs more often than “him” while “her” occurs more often than “she” which indicating that women are presented as more passive.

Understanding contexts

6.

Tokenise tags and produced n-grams (bi-grams starting with “he” and “she”, then expand up to n-grams of 7 words as well as n-grams of 3 words ending with “him” and “her”

```
let nGramsfound = NGrams.ngrams(file, 3);
let count = 0;

for (let i = 0; i < nGramsfound.length; i++) {
    if (nGramsfound[i][0] == 'she') {
        console.log(nGramsfound[i][1]);
    }
}
```

```
for (let i = 0; i < nGramsfound.length; i++) {
    if (nGramsfound[i][0] == 'she' && nGramsfound[i][1] == 'bare') {
        console.log(nGramsfound[i]);
    }
}
```

trigram starting with he	5-gram starting with he
['he', 'called', 'Night']	['he', 'called', 'Night', 'And', 'the']
['he', 'Seas', 'and']	['he', 'Seas', 'and', 'God', 'saw']
['he', 'made', 'the']	['he', 'made', 'the', 'stars', 'also']
['he', 'him', 'male']	['he', 'him', 'male', 'and', 'female']
['he', 'them', '1']	['he', 'them', '1', '28', 'And']
['he', 'had', 'made']	['he', 'had', 'made', 'and', 'behold']
['he', 'had', 'made']	['he', 'had', 'made', 'and', 'he']
['he', 'rested', 'on']	['he', 'rested', 'on', 'the', 'seventh']
['he', 'had', 'made']	['he', 'had', 'made', '2', '3']
['he', 'had', 'rested']	['he', 'had', 'rested', 'from', 'all']
['he', 'put', 'the']	['he', 'put', 'the', 'man', 'whom']

7. Save terminal results into Google sheets to clean them up

5-gram starting with she	trigram with her in the middle	trigram ending with her	trigram with his in the middle	trigram with him in the middle
['she', 'shall', 'be', 'called', 'Woman']	['brought', 'her', 'unto']	['and', 'brought', 'her']	['after', 'his', 'kind']	['he', 'him', 'male']
['she', 'was', 'taken', 'out', 'of']	['unto', 'her', 'husband']	['also', 'unto', 'her']	['after', 'his', 'kind']	['put', 'him', 'into']
['she', 'took', 'of', 'the', 'fruit']	['with', 'her', 'and']	['husband', 'with', 'her']	['after', 'his', 'kind']	['make', 'him', 'an']
['she', 'gave', 'me', 'of', 'the']	['and', 'her', 'seed']	['seed', 'and', 'her']	['after', 'his', 'kind']	['for', 'him', '2']
['she', 'was', 'the', 'mother', 'of']	['opened', 'her', 'mouth']	['hath', 'opened', 'her']	['after', 'his', 'kind']	['for', 'him', '2']
['she', 'conceived', 'and', 'bare', 'Cain']	['thee', 'her', 'strength']	['unto', 'thee', 'her']	['after', 'his', 'kind']	['unto', 'him', 'Where']
['she', 'again', 'bare', 'his', 'brother']	['of', 'her', 'foot']	['sole', 'of', 'her']	['after', 'his', 'kind']	['sent', 'him', 'forth']
['she', 'conceived', 'and', 'bare', 'Enoch']	['took', 'her', 'and']	['and', 'took', 'her']	['after', 'his', 'kind']	['over', 'him', '4']
['she', 'also', 'bare', 'Tubalcain', 'an']	['pulled', 'her', 'in']	['and', 'pulled', 'her']	['in', 'his', 'own']	['slew', 'him', '4']
['she', 'bare', 'a', 'son', 'and']	['in', 'her', 'mouth']	['lo', 'in', 'her']	['ended', 'his', 'work']	['unto', 'him', 'Therefore']
['she', 'hath', 'appointed', 'me', 'another']	['saw', 'her', 'and']	['Pharaoh', 'saw', 'her']	['all', 'his', 'work']	['on', 'him', 'sevenfold']
['she', 'returned', 'unto', 'him', 'into']	['commended', 'her', 'before']	['and', 'commended', 'her']	['all', 'his', 'work']	['finding', 'him', 'should']
['she', 'had', 'no', 'child', '11']	['for', 'her', 'sake']	['well', 'for', 'her']	['into', 'his', 'nostrils']	['kill', 'him', '4']
['she', 'was', 'very', 'fair', '12']	['taken', 'her', 'to']	['have', 'taken', 'her']	['of', 'his', 'ribs']	['unto', 'him', 'two']
['she', 'asses', 'and', 'camels', '12']	['take', 'her', 'and']	['wife', 'take', 'her']	['leave', 'his', 'father']	['to', 'him', 'also']

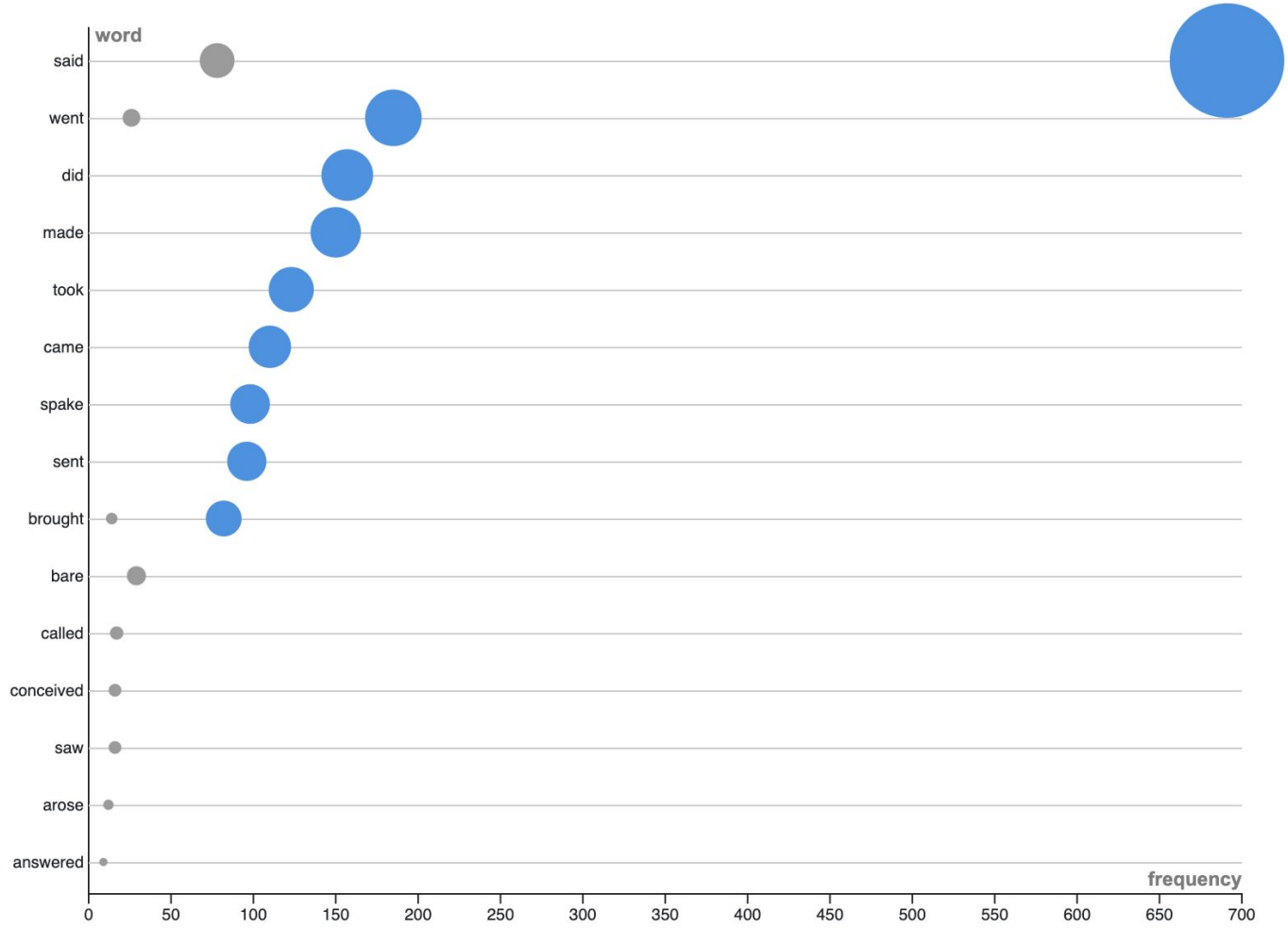
8. create a new text files with the word after he/she in the bigram

9. Count word frequency

10. Create a new Google sheet with 10 most frequent words following he/she and cleaned up the data to remove verbs which are not actions.

11. Feed Data into Raw Graph to produce Beeswarm Plots

Bigram	frequency	gender
said	691	M
went	185	M
did	157	M
made	150	M
took	123	M
came	110	M
spake	98	M
sent	96	M
brought	82	M
said	78	F
bare	29	F
went	26	F
called	17	F
conceived	16	F
saw	16	F
brought	14	F
arose	12	F
answered	9	F


word
said
went
did
made
took
came
spake
sent
brought
bare
called
conceived
saw
arose
answered
frequency
0
50
100
150
200
250
300
350
400
450
500
550
600
650
700

Inferences

Verbs following “he” are more active than verbs following “she”

Most of the verbs following “she” are passive in nature and 20% are related to childbearing.