HULT
INTERNATIONAL BUSINESS SCHOOL

Business Insights Report

Text Analytics – Professor Thomas Kurnicki

Created by:
Burcu Dogru

February 10th, 2021

Poetry is resonating with people who are looking for understanding. It is a really good way to explore complex, difficult emotions and uncertainty. Latest increase passion for politics in societies, particularly among teenagers and young millennials, is fueling a dramatic growth in the popularity of poetry, with sales of poetry books hitting an all-time high in 2018.

Well aware of the market opportunity, Poemhunter is a San Francisco based, comprehensive poet and poetry portal which is aiming to set the waves for millennials in the poetry world.

As Poemhunter's target customer segment is millennials on their portal, "interactivity" and "being insightful" is one of the highest priorities.

Poemhunter is offering hashtag based search for poetries, hashtag based commenting/rating options as well as messenger option. Based on the latest survey conducted on the portal, customers rated their favorite poems of all times and there was a discussion board to discuss similarities and differences in the words and poets. Poemhunter decided to run a text analysis and provide real insights for their customers based on the latest hot discussion topic.

In this report, we are going to explain implemented frameworks as well as providing insights for it.

Implemented frameworks for this analysis;
Correlograms, Correlation test, Frequency Histograms, TF-IDF Analysis and Word clouds.

Poems Analyzed for this analysis;
William Blake: The Tyger, The Lamb
William Shakespeare: Shall I Compare Thee To a Summer's Day (Sonnet 18), Sonnet 116
William Wordsworth: Composed Upon Westminster Bridge-September 3-1802, A Slumber Did My Spirit Seal

William Blake was an English poet, painter, and printmaker. Largely unrecognized during his lifetime, Blake is now considered a seminal figure in the history of the poetry and visual arts of the Romantic Age. (1)

William Shakespeare was an English poet and playwright, widely regarded as the greatest writer in the English language and the world's pre-eminent dramatist. He is often called England's national poet and the "Bard of Avon".(2)

William Wordsworth (was an English Romantic poet who, with Samuel Taylor Coleridge, helped to launch the Romantic Age in English literature with their joint publication Lyrical Ballads. (3)

For this analysis, instead of importing pdfs, two most famous poems of each poets were directly pull into R by building separate text objects. The most common used words are removed by using stop words build in the analysis. And the sentiments are assigned to each word.

After creating each data frames for each poet and one additional dataset combining all poets, I started tokenization for all data sets individually.

When I am plotting my correlograms I binded my William Blake, William Shakespare, William Wordsworth dataframes together and used William Blake as my baseline. So for my correlograms I wanted to see the similarity between Blake vs. Shakespeare as well as Blake vs Wordsworth. On Poemhunter portal discussion, Blake and Wordsworth were seen as very similar.
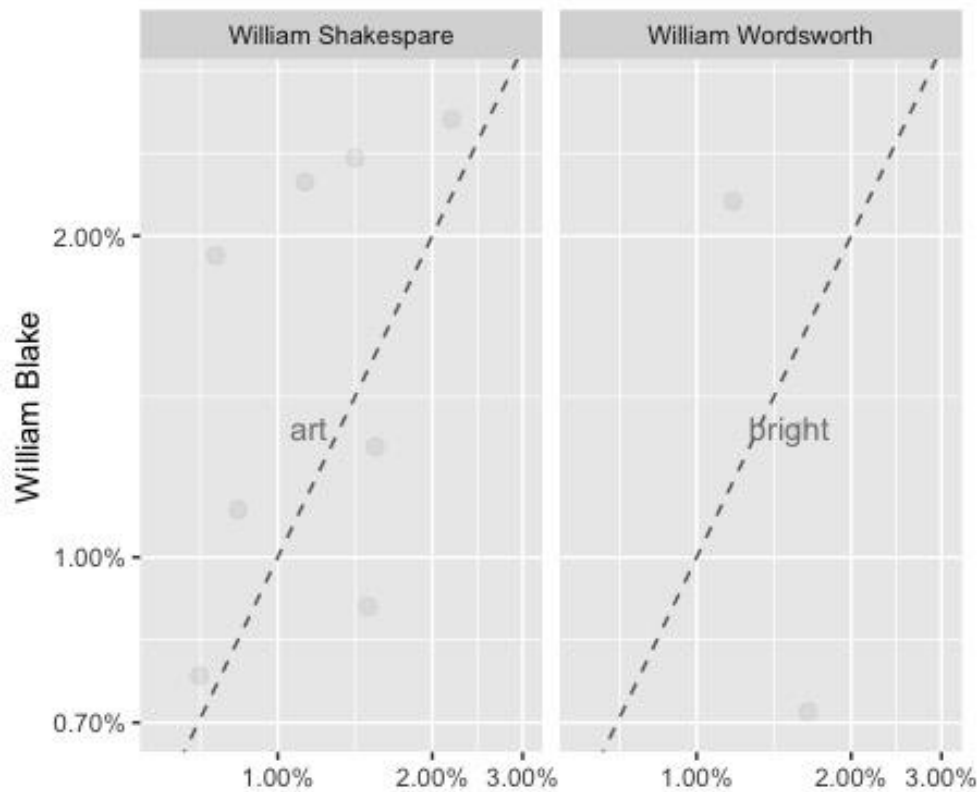

Figure 1: Correlogram

Around rising business question that if William Blake's words any similar to Shakespeare and Wordsworth, our correlogram shows that they both are very dissimilar. When comparing Blake's poem's to Shakespeare the only similar word is art. When comparing Blake's poem's to Shakespeare the only similar word we found is bright.
Following that by correlation test, the standard deviation is found 0, therefore we weren't able to get any number our of correlation test. Both results were NA.

Following histograms are created based on the frequencies of each word found in each poet's text data.
After removing stop words (the most common words which brings no value to our business) we're left with following famous words.
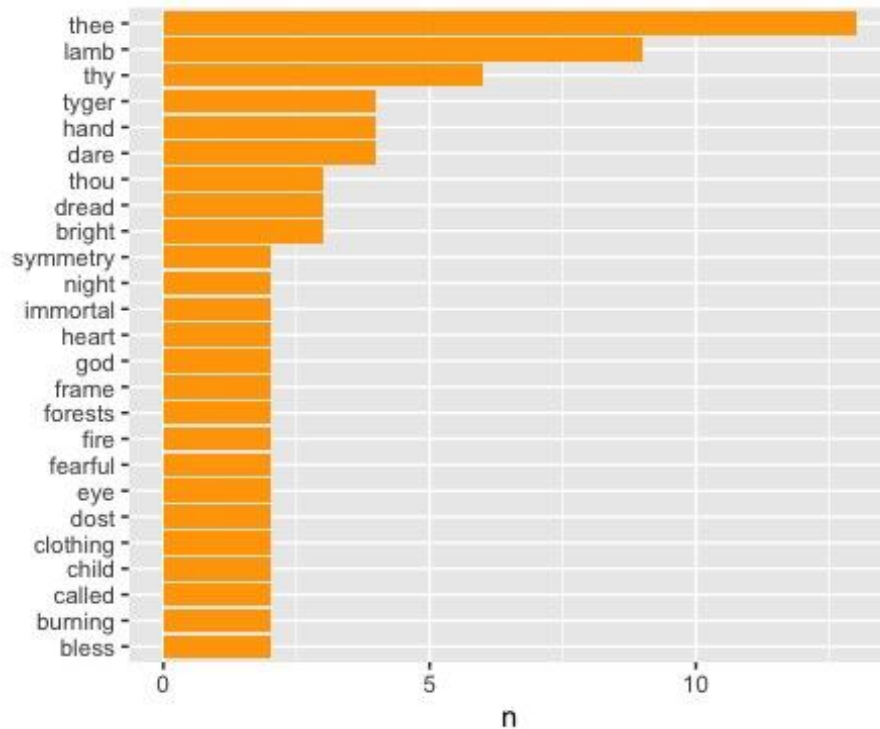
Figure 2: Word Frequencies-William Blake

Most repeated words in Blake's poems found as; thee, thy and thou which is in their times it was used as you/your/yours to especially higher status. Using words lamb, bright, night, forests are the main insights matching with William Blake used in his poems especially nature topics and childish innocent themes.
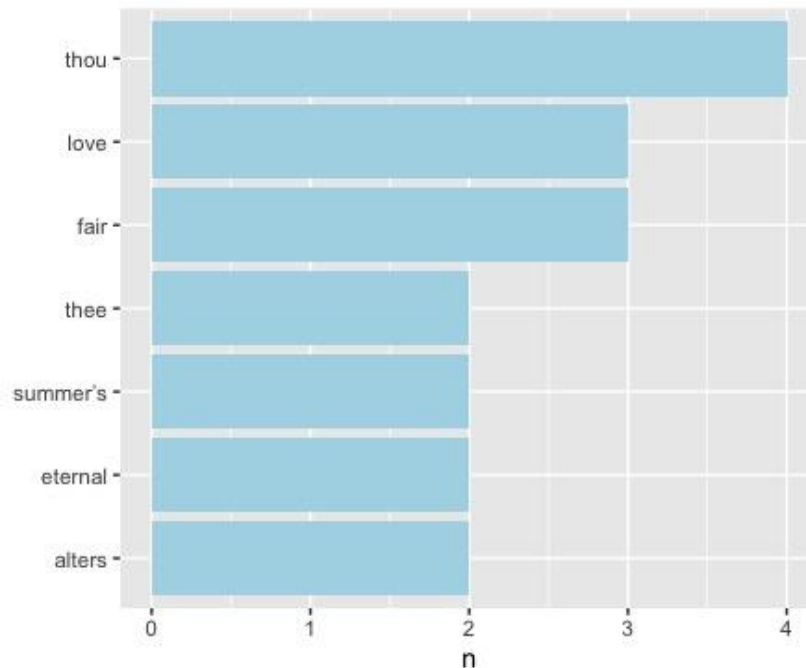

Figure 3: Word Frequencies-William Shakespeare

Unlike William Blake, in his poems Shakespeare uses more words which related to Romance such as love and fairness and eternal.

Figure 4: Word Frequencies-William Wordsworth

When we look at the above chart, it seems like all our frequencies are exclusive. In his two poems William Wordsworth using unrepetitive words, therefore it wasn't able for my code to reduce the number of words presented in the chart.

TF-IDF ANALYSIS:

Tf-idf analysis is a numerical statistic al analysis that is intended to reflect how important a word is to a document or text data.

In the below charts, the left hand side represents the least repetitive words in our document, approaching right hand side that are the words representing repetitive words.

Comparing these three great artists we can say that William Blake has tendency to use more words that are repeating in the poem. William Blake is well known using his rhymes by repeating words.

When it comes to William Wordsworth's word choice, he almost never uses the same words which bring value to our business.
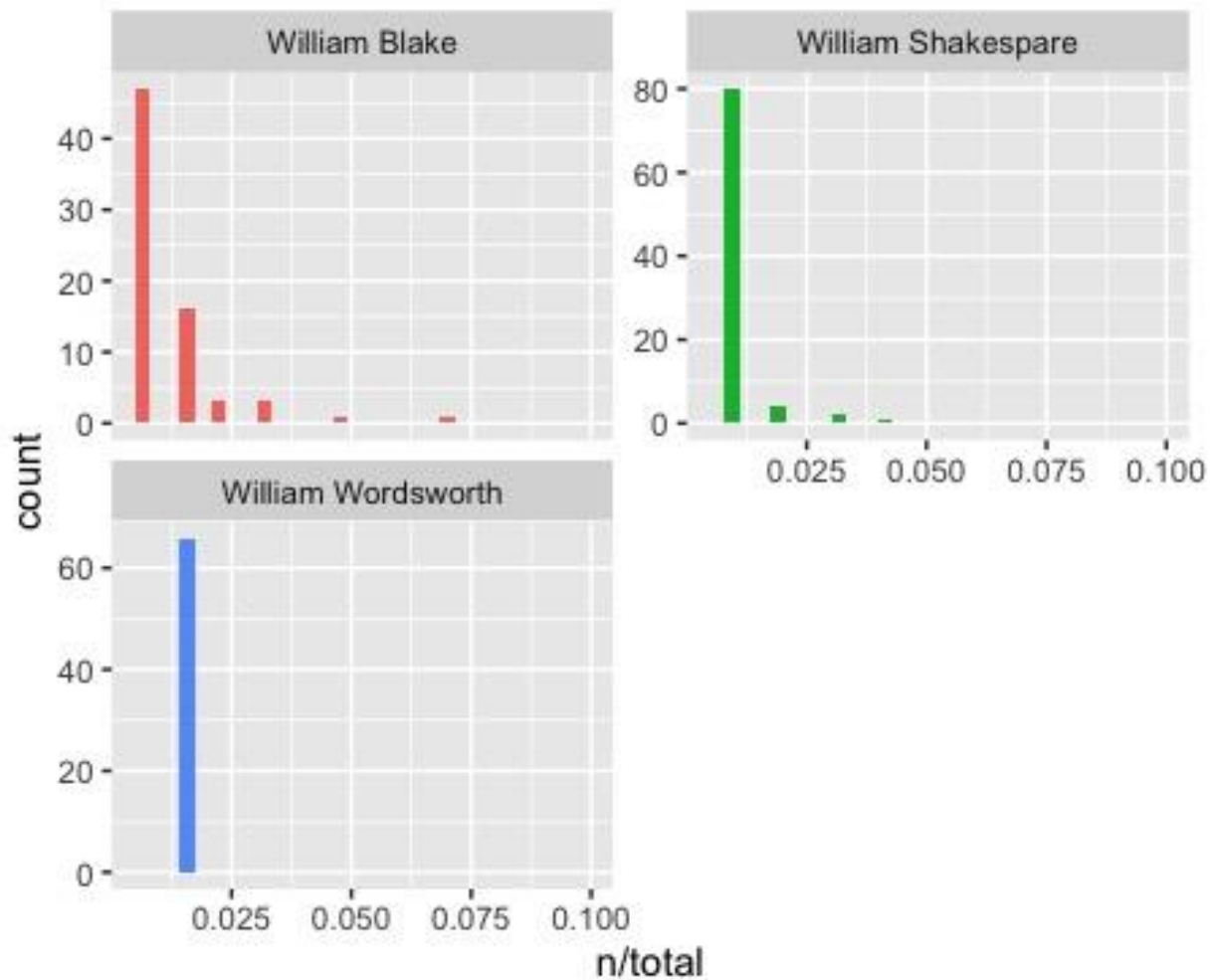
Figure 5: Tf-Idf Analysis

ZIPF'S LAW:

By implying the Zipf's law to our data, we will eliminate the words that are less frequently used but could give us more insights. See the William Blake's plot, we see words such as "lamb", "tyger", "dread" these are indicated more to William Blake's original style and what seperates him from other poets. William Shakespeare on the other hand has using more usual words representing his style. Once again the last plot below indicated that all words that William Wordsworth used in his poems are such unique and has high business insight which we couldn't appropriately rank them.

Figure 6: Zipf's Law

Figure 8: Term Frequency
Above chart shows Wordsworth's word choices shows perfect consistency.

In this last part of our analysis, we compared the "bing" sentiments for each word in our poems. Based on the word clouds below; Blake's and Wordsworth's word choices are more on the positive side while Shakespeare uses positive and negative words equally.
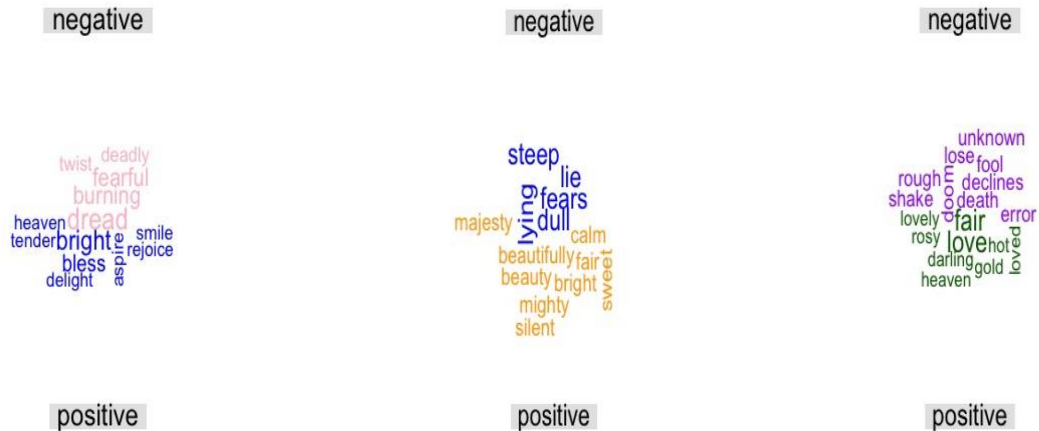

Figure 8: WordClouds-W.Blake(1), W.Wordsworth(2), W.Shakespeare(3)

Based on the conducted methods of text analysis, I would conclude that these three poet's are not quite similar when it comes to word choices. In terms of reflecting emotions using sentiment analysis, W. Wordsworth and W.Blake representing more similar feelings compare to W.Shakespeare. Last but least, the most unique approach for word choice belongs to W.Wordsworth.

```
##########################################################
########## Downloading Packages and Libraries##########
##########################################################
install.packages("textdata")
install.packages("wordcloud")
install.packages(("reshape2"))
library(textdata)
library(dplyr)
library(stringr)
library(tidyr)
library(scales)
library(tidytext)
library(tm)
library(ggplot2)
library(wordcloud)
library(reshape2)

##########################################################
################## Creating Data #######################
##########################################################
William_Blake <- c("Tyger Tyger, burning bright",
            "In the forests of the night",
            "What immortal hand or eye",
            "Could frame thy fearful symmetry?",
            "In what distant deeps or skies.",
            "Burnt the fire of thine eyes?",
            "On what wings dare he aspire?",
            "What the hand, dare seize the fire?",
            "And what shoulder, & what art,",
            "Could twist the sinews of thy heart?",
            "And when thy heart began to beat,",
            "What dread hand? & what dread feet?",
            "What the hammer? what the chain," ,
            "In what furnace was thy brain?",
            "What the anvil? what dread grasp,",
            "Dare its deadly terrors clasp!" ,
            "When the stars threw down their spears",
            "And water'd heaven with their tears:" ,
            "Did he smile his work to see?",
            "Did he who made the Lamb make thee?",
            "Tyger Tyger burning bright" ,
            "In the forests of the night:" ,
            "What immortal hand or eye,",
            "Dare frame thy fearful symmetry?",
            "Little lamb, who made thee?",
```

"Dost thou know who made thee,",
          "Gave thee life, and bid thee feed",
          "By the stream and o'er the mead;",
          "Gave thee clothing of delight,",
          "Softest clothing, woolly, bright;",
          "Gave thee such a tender voice,",
          "Making all the vales rejoice?",
          "Little lamb, who made thee?",
          "Dost thou know who made thee?",
          "Little lamb, I'll tell thee;",
          "Little lamb, I'll tell thee:",
          "He is called by thy name,",
          "For He calls Himself a Lamb.",
          "He is meek, and He is mild,",
          "He became a little child.",
          "I a child, and thou a lamb,",
          "We are called by His name.",
          "Little lamb, God bless thee!",
          "Little lamb, God bless thee!")
##
William_Shakespare <- c("Shall I compare thee to a summer's day?",
                "Thou art more lovely and more temperate",
                "Rough winds do shake the darling buds of May",
                "And summer's lease hath all too short a date",
                "Sometime too hot the eye of heaven shines",
                "And often is his gold complexion dimm'd",
                "And every fair from fair sometime declines",
                "By chance or nature's changing course untrimm'd",
                "But thy eternal summer shall not fade",
                "Nor lose possession of that fair thou ow'st;",
                "Nor shall death brag thou wander'st in his shade",
                "When in eternal lines to time thou grow'st",
                "So long as men can breathe or eyes can see",
                "So long lives this, and this gives life to thee",
                "Let me not to the marriage of true minds",
                "Admit impediments. Love is not love",
                "Which alters when it alteration finds",
                "Or bends with the remover to remove",
                "O no! it is an ever-fixed mark",
                "That looks on tempests and is never shaken;",
                "It is the star to every wand'ring bark,",
                "Whose worth's unknown, although his height be taken.",
                "Love's not Time's fool, though rosy lips and cheeks",
                "Within his bending sickle's compass come",
                "Love alters not with his brief hours and weeks",
                "But bears it out even to the edge of doom",

```
                "If this be error and upon me proved",
                "I never writ, nor no man ever loved")
############################################################################
William_Wordsworth <- c("Earth has not anything to show more fair:",
                "Dull would he be of soul who could pass by",
                "A sight so touching in its majesty:",
                "This City now doth, like a garment, wear",
                "The beauty of the morning; silent, bare,",
                "Ships, towers, domes, theatres, and temples lie",
                "Open unto the fields, and to the sky;",
                "All bright and glittering in the smokeless air.",
                "Never did sun more beautifully steep",
                "In his first splendour, valley, rock, or hill;",
                "Ne'er saw I, never felt, a calm so deep!",
                "The river glideth at his own sweet will:",
                "Dear God! the very houses seem asleep;",
                "And all that mighty heart is lying still!",
                "A slumber did my spirit seal",
                "I had no human fears:",
                "She seemed a thing that could not feel",
                "The touch of earthly years.",
                "No motion has she now, no force;",
                "She neither hears nor sees;",
                "Rolled round in earth's diurnal course,",
                "With rocks, and stones, and trees.")


##########################################################
###########Creating Dataframes#################
##########################################################
William_Blake_df <- data.frame(line=20:63, text=William_Blake)
William_Shakespare_df <- data.frame(line=65:92, text=William_Shakespare)
William_Wordsworth_df <- data.frame(line=94:115, text=William_Wordsworth)

#make sure to run above code separately(line by line) and before the below code (otherwise it
throws an error)
All_poets <- bind_rows(mutate(William_Blake_df, author="William Blake"),
                mutate(William_Shakespare_df, author= "William Shakespare"),
                mutate(William_Wordsworth_df, author="William Wordsworth"))

###########################################################
##################Tokenization################
###########################################################
All_poets_tokens <- All_poets %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
```

```
    count(word, sort=TRUE)
#############################################################
William_Blake_tokens <- William_Blake_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE)
##############################################################
William_Shakespare_tokens <- William_Shakespare_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE)

William_Wordsworth_tokens <- William_Wordsworth_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE)


############################################
############## Correlograms ###############
############################################
correl <- bind_rows(mutate(William_Blake_tokens, author="William Blake"),
              mutate(William_Shakespare_tokens, author= "William Shakespare"),
              mutate(William_Wordsworth_tokens, author="William Wordsworth")
)%>%
  mutate(word=str_extract(word, "[a-z']+")) %>%
  count(author, word) %>%
  group_by(author) %>%
  mutate(proportion = n/sum(n))%>%
  select(-n) %>%
  spread(author, proportion) %>%
  gather(author, proportion,`William Shakespare`,`William Wordsworth`)
####Plotting my correlagram#########################################################
ggplot(correl, aes(x=proportion, y=`William Blake`,
              color = abs(`William Blake`- proportion)))+
  geom_abline(color="grey40", lty=2)+
  geom_jitter(alpha=.1, size=2.5, width=0.3, height=0.3)+
  geom_text(aes(label=word), check_overlap = TRUE, vjust=1.5) +
  scale_x_log10(labels = percent_format())+
  scale_y_log10(labels= percent_format())+
  scale_color_gradient(limits = c(0,0.001), low = "darkslategray4", high = "gray75")+
  facet_wrap(~author, ncol=2)+
  theme(legend.position = "none")+
  labs(y= "William Blake", x=NULL)


############################################
######Correlation Test ###############
```

```r
###############################################
cor.test(data=correl[correl$author == "William Shakespare",],
      ~proportion + `William Blake`)

cor.test(data=correl[correl$author == "William Wordsworth",],
      ~proportion + `William Blake`)

###############################################
######## Frequency Histograms ######
###############################################
freq_hist_bl <- William_Blake_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE) %>%
  top_n(10) %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n))+
  geom_col(fill="orange")+
  xlab(NULL)+  #label
  coord_flip() #Coordinates flipped
print(freq_hist_bl)
###################################################################

freq_hist_sh <- William_Shakespare_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE) %>%
  mutate(word=reorder(word, n)) %>%
  top_n(7) %>%
  ggplot(aes(word, n))+
  geom_col(fill="lightblue")+
  xlab(NULL)+  #label
  coord_flip() #Coordinates flipped
print(freq_hist_sh)
####################################################################
freq_hist_wrd <- William_Wordsworth_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(word, sort=TRUE) %>%
  top_n(1) %>%
  mutate(word=reorder(word, n)) %>%
  ggplot(aes(word, n))+
  geom_col(fill="darkgreen")+
  xlab(NULL)+  #label
  coord_flip() #Coordinates flipped
print(freq_hist_wrd)
```

```r
###############################################################
################# tf_idf Analysis#########################
###############################################################
All_poets_idf <- bind_rows(mutate(William_Blake_df, author="William Blake"),
                mutate(William_Shakespare_df, author= "William Shakespare"),
                mutate(William_Wordsworth_df, author="William Wordsworth")) %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  count(author, word, sort=TRUE) %>%
  ungroup()
  ##########################################
total_words <- All_poets_idf %>%
    group_by(author) %>%
    summarize(total=sum(n))
poets_words <- left_join(All_poets_idf, total_words)

print(poets_words)
 #######################################################
ggplot(poets_words, aes(n/total, fill = author))+
    geom_histogram(show.legend=FALSE)+
    xlim(NA, 0.1) +
    facet_wrap(~author, ncol=2, scales="free_y")

##########################################
########### ZIPF's law ################
##########################################
freq_by_rank <- poets_words %>%
  group_by(author) %>%
  mutate(rank = row_number(),
      `term frequency` = n/total)
freq_by_rank

# plot ZIPF's Law###################################
freq_by_rank %>%
  ggplot(aes(rank, `term frequency`, color=author))+
  geom_abline(intercept=-0.62, slope= -1.1, color='gray50', linetype=2)+
  #let's add a tangent line , the first derivative, and see what the slop is
  geom_line(size= 1.1, alpha = 0.8, show.legend = FALSE)+
  scale_x_log10()+
  scale_y_log10()

poets_words_idf <-poets_words %>%
  bind_tf_idf(word, author, n)

poets_words_idf %>%
```

```
  arrange(desc(tf_idf))
###############################################################
poets_words_idf %>%
  arrange(desc(tf_idf)) %>%
  mutate(word=factor(word, levels=rev(unique(word)))) %>%
  group_by(author) %>%
  filter(n<10) %>%
  top_n(5) %>%
  ungroup %>%
  ggplot(aes(word, tf_idf, fill=author))+
  geom_col(show.legend=FALSE)+
  labs(x=NULL, y="tf-idf")+
  facet_wrap(~author, ncol=2, scales="free")+
  coord_flip()


#######################################
########## WordClouds ###############
#######################################
Cloud_Blake <- William_Blake_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>% #token per sentiment
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("red", "blue"),
             max.words=50, scale=c(1.5, 0.8),
             title.size=1.5)
###############################################################################
Cloud_Shakespeare <- William_Shakespare_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>% #token per sentiment
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("red", "darkgreen"),
             max.words=50, scale=c(1.5, 0.8),
             title.size=1.5)


###############################################################################
Cloud_Wordsworth <- William_Wordsworth_df %>%
  unnest_tokens(word, text) %>%
  anti_join(stop_words) %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort=TRUE) %>% #token per sentiment
  acast(word ~sentiment, value.var="n", fill=0) %>%
  comparison.cloud(colors = c("red", "yellow"),
```

```
            max.words=50, scale=c(1.5, 0.8),
            title.size=1.5)
```

Correlation Output
Pearson's product-moment correlation

data: proportion and William Blake
t = NA, df = 1, p-value = NA
alternative hypothesis: true correlation is not equal to 0
sample estimates:
cor
 NA

Warning message:
In cor(x, y) : the standard deviation is zero

Tf-idf Output for total words

```
print(poets_words)
          author      word  n total
1     William Blake      thee 13  128
2     William Blake      lamb  9  128
3     William Blake       thy  6  128
4     William Blake      dare  4  128
5     William Blake      hand  4  128
6     William Blake     tyger  4  128
7  William Shakespare     thou  4   98
8     William Blake    bright  3  128
9     William Blake     dread  3  128
10     William Blake      thou  3  128
11 William Shakespare     fair  3   98
12 William Shakespare     love  3   98
13     William Blake     bless  2  128
14     William Blake   burning  2  128
15     William Blake    called  2  128
16     William Blake     child  2  128
17     William Blake  clothing  2  128
18     William Blake      dost  2  128
19     William Blake       eye  2  128
20     William Blake   fearful  2  128
21     William Blake      fire  2  128
22     William Blake    forests  2  128
23     William Blake     frame  2  128
24     William Blake       god  2  128
```

25	William Blake	heart	2	128
26	William Blake	immortal	2	128
27	William Blake	night	2	128
28	William Blake	symmetry	2	128
29	William Shakespare	alters	2	98
30	William Shakespare	eternal	2	98
31	William Shakespare	summer's	2	98
32	William Shakespare	thee	2	98
33	William Blake	anvil	1	128
34	William Blake	art	1	128
35	William Blake	aspire	1	128
36	William Blake	beat	1	128
37	William Blake	bid	1	128
38	William Blake	brain	1	128
39	William Blake	burnt	1	128
40	William Blake	calls	1	128
41	William Blake	chain	1	128
42	William Blake	clasp	1	128
43	William Blake	deadly	1	128
44	William Blake	deeps	1	128
45	William Blake	delight	1	128
46	William Blake	distant	1	128
47	William Blake	eyes	1	128
48	William Blake	feed	1	128
49	William Blake	feet	1	128
50	William Blake	furnace	1	128
51	William Blake	grasp	1	128
52	William Blake	hammer	1	128
53	William Blake	heaven	1	128
54	William Blake	life	1	128
55	William Blake	mead	1	128
56	William Blake	meek	1	128
57	William Blake	mild	1	128
58	William Blake	o'er	1	128
59	William Blake	rejoice	1	128
60	William Blake	seize	1	128
61	William Blake	shoulder	1	128
62	William Blake	sinews	1	128
63	William Blake	skies	1	128
64	William Blake	smile	1	128
65	William Blake	softest	1	128
66	William Blake	spears	1	128
67	William Blake	stars	1	128
68	William Blake	stream	1	128
69	William Blake	tears	1	128
70	William Blake	tender	1	128

71	William Blake	terrors	1	128
72	William Blake	thine	1	128
73	William Blake	threw	1	128
74	William Blake	twist	1	128
75	William Blake	vales	1	128
76	William Blake	voice	1	128
77	William Blake	water'd	1	128
78	William Blake	wings	1	128
79	William Blake	woolly	1	128
80	William Shakespare	admit	1	98
81	William Shakespare	alteration	1	98
82	William Shakespare	art	1	98
83	William Shakespare	bark	1	98
84	William Shakespare	bears	1	98
85	William Shakespare	bending	1	98
86	William Shakespare	bends	1	98
87	William Shakespare	brag	1	98
88	William Shakespare	breathe	1	98
89	William Shakespare	buds	1	98
90	William Shakespare	chance	1	98
91	William Shakespare	changing	1	98
92	William Shakespare	cheeks	1	98
93	William Shakespare	compare	1	98
94	William Shakespare	compass	1	98
95	William Shakespare	complexion	1	98
96	William Shakespare	darling	1	98
97	William Shakespare	date	1	98
98	William Shakespare	day	1	98
99	William Shakespare	death	1	98
100	William Shakespare	declines	1	98
101	William Shakespare	dimm'd	1	98
102	William Shakespare	doom	1	98
103	William Shakespare	edge	1	98
104	William Shakespare	error	1	98
105	William Shakespare	eye	1	98
106	William Shakespare	eyes	1	98
107	William Shakespare	fade	1	98
108	William Shakespare	fixed	1	98
109	William Shakespare	fool	1	98
110	William Shakespare	gold	1	98
111	William Shakespare	grow'st	1	98
112	William Shakespare	hath	1	98
113	William Shakespare	heaven	1	98
114	William Shakespare	height	1	98
115	William Shakespare	hot	1	98
116	William Shakespare	hours	1	98

```
117 William Shakespare impediments  1    98
118 William Shakespare        lease  1     98
119 William Shakespare         life  1     98
120 William Shakespare        lines  1     98
121 William Shakespare         lips  1     98
122 William Shakespare        lives  1     98
123 William Shakespare         lose  1     98
124 William Shakespare       love's  1     98
125 William Shakespare        loved  1     98
126 William Shakespare       lovely  1     98
127 William Shakespare         mark  1     98
128 William Shakespare     marriage  1     98
129 William Shakespare        minds  1     98
130 William Shakespare     nature's  1     98
131 William Shakespare       ow'st  1     98
132 William Shakespare   possession  1     98
133 William Shakespare       proved  1     98
134 William Shakespare       remove  1     98
135 William Shakespare      remover  1     98
136 William Shakespare        rosy  1     98
137 William Shakespare       rough  1     98
138 William Shakespare        shade  1     98
139 William Shakespare        shake  1     98
140 William Shakespare       shaken  1     98
141 William Shakespare       shines  1     98
142 William Shakespare        short  1     98
143 William Shakespare      sickle's  1     98
144 William Shakespare         star  1     98
145 William Shakespare       summer  1     98
146 William Shakespare    temperate  1     98
147 William Shakespare     tempests  1     98
148 William Shakespare          thy  1     98
149 William Shakespare         time  1     98
150 William Shakespare       time's  1     98
151 William Shakespare         true  1     98
152 William Shakespare      unknown  1     98
153 William Shakespare     untrimm'd  1     98
154 William Shakespare      wand'ring  1     98
155 William Shakespare      wander'st  1     98
156 William Shakespare        weeks  1     98
157 William Shakespare        winds  1     98
158 William Shakespare      worth's  1     98
159 William Shakespare         writ  1     98
160 William Wordsworth         air  1     66
161 William Wordsworth       asleep  1     66
162 William Wordsworth         bare  1     66
```

163	William Wordsworth	beautifully	1	66
164	William Wordsworth	beauty	1	66
165	William Wordsworth	bright	1	66
166	William Wordsworth	calm	1	66
167	William Wordsworth	city	1	66
168	William Wordsworth	dear	1	66
169	William Wordsworth	deep	1	66
170	William Wordsworth	diurnal	1	66
171	William Wordsworth	domes	1	66
172	William Wordsworth	doth	1	66
173	William Wordsworth	dull	1	66
174	William Wordsworth	earth	1	66
175	William Wordsworth	earth's	1	66
176	William Wordsworth	earthly	1	66
177	William Wordsworth	fair	1	66
178	William Wordsworth	fears	1	66
179	William Wordsworth	feel	1	66
180	William Wordsworth	fields	1	66
181	William Wordsworth	force	1	66
182	William Wordsworth	garment	1	66
183	William Wordsworth	glideth	1	66
184	William Wordsworth	glittering	1	66
185	William Wordsworth	god	1	66
186	William Wordsworth	hears	1	66
187	William Wordsworth	heart	1	66
188	William Wordsworth	hill	1	66
189	William Wordsworth	houses	1	66
190	William Wordsworth	human	1	66
191	William Wordsworth	lie	1	66
192	William Wordsworth	lying	1	66
193	William Wordsworth	majesty	1	66
194	William Wordsworth	mighty	1	66
195	William Wordsworth	morning	1	66
196	William Wordsworth	motion	1	66
197	William Wordsworth	ne'er	1	66
198	William Wordsworth	pass	1	66
199	William Wordsworth	river	1	66
200	William Wordsworth	rock	1	66
201	William Wordsworth	rocks	1	66
202	William Wordsworth	rolled	1	66
203	William Wordsworth	round	1	66
204	William Wordsworth	seal	1	66
205	William Wordsworth	ships	1	66
206	William Wordsworth	sight	1	66
207	William Wordsworth	silent	1	66
208	William Wordsworth	sky	1	66

209 William Wordsworth   slumber 1   66
210 William Wordsworth smokeless 1   66
211 William Wordsworth   soul 1   66
212 William Wordsworth   spirit 1   66
213 William Wordsworth splendour 1   66
214 William Wordsworth   steep 1   66
215 William Wordsworth   stones 1   66
216 William Wordsworth   sun 1   66
217 William Wordsworth   sweet 1   66
218 William Wordsworth   temples 1   66
219 William Wordsworth   theatres 1   66
220 William Wordsworth   touch 1   66
221 William Wordsworth   touching 1   66
222 William Wordsworth   towers 1   66
223 William Wordsworth   trees 1   66
224 William Wordsworth   valley 1   66
225 William Wordsworth   wear 1   66

Zipf's Law Output

freq_by_rank
A tibble: 225 x 6
Groups: author [3]

author	word	n	total	rank	`term frequency`
<chr>	<chr>	<int>	<int>	<int>	<dbl>
1 William Blake	thee	13	128	1	0.102
2 William Blake	lamb	9	128	2	0.0703
3 William Blake	thy	6	128	3	0.0469
4 William Blake	dare	4	128	4	0.0312
5 William Blake	hand	4	128	5	0.0312
6 William Blake	tyger	4	128	6	0.0312
7 William Shakespare	thou	4	98	1	0.0408
8 William Blake	bright	3	128	7	0.0234
9 William Blake	dread	3	128	8	0.0234
10 William Blake	thou	3	128	9	0.0234

… with 215 more rows

TF-Idf Output

poets_words_idf %>%
+   arrange(desc(tf_idf))

	author	word	n	total	tf	idf	tf_idf
1	William Blake	lamb	9	128	0.07031250	1.0986123	0.07724618
2	William Blake	thee	13	128	0.10156250	0.4054651	0.04118005
3	William Blake	dare	4	128	0.03125000	1.0986123	0.03433163
4	William Blake	hand	4	128	0.03125000	1.0986123	0.03433163
5	William Blake	tyger	4	128	0.03125000	1.0986123	0.03433163
6	William Shakespare	love	3	98	0.03061224	1.0986123	0.03363099

7	William Blake	dread	3	128	0.02343750	1.0986123	0.02574873
8	William Shakespare	alters	2	98	0.02040816	1.0986123	0.02242066
9	William Shakespare	eternal	2	98	0.02040816	1.0986123	0.02242066
10	William Shakespare	summer's	2	98	0.02040816	1.0986123	0.02242066
11	William Blake	thy	6	128	0.04687500	0.4054651	0.01900618
12	William Blake	bless	2	128	0.01562500	1.0986123	0.01716582
13	William Blake	burning	2	128	0.01562500	1.0986123	0.01716582
14	William Blake	called	2	128	0.01562500	1.0986123	0.01716582
15	William Blake	child	2	128	0.01562500	1.0986123	0.01716582
16	William Blake	clothing	2	128	0.01562500	1.0986123	0.01716582
17	William Blake	dost	2	128	0.01562500	1.0986123	0.01716582
18	William Blake	fearful	2	128	0.01562500	1.0986123	0.01716582
19	William Blake	fire	2	128	0.01562500	1.0986123	0.01716582
20	William Blake	forests	2	128	0.01562500	1.0986123	0.01716582
21	William Blake	frame	2	128	0.01562500	1.0986123	0.01716582
22	William Blake	immortal	2	128	0.01562500	1.0986123	0.01716582
23	William Blake	night	2	128	0.01562500	1.0986123	0.01716582
24	William Blake	symmetry	2	128	0.01562500	1.0986123	0.01716582
25	William Wordsworth	air	1	66	0.01515152	1.0986123	0.01664564
26	William Wordsworth	asleep	1	66	0.01515152	1.0986123	0.01664564
27	William Wordsworth	bare	1	66	0.01515152	1.0986123	0.01664564
28	William Wordsworth	beautifully	1	66	0.01515152	1.0986123	0.01664564
29	William Wordsworth	beauty	1	66	0.01515152	1.0986123	0.01664564
30	William Wordsworth	calm	1	66	0.01515152	1.0986123	0.01664564
31	William Wordsworth	city	1	66	0.01515152	1.0986123	0.01664564
32	William Wordsworth	dear	1	66	0.01515152	1.0986123	0.01664564
33	William Wordsworth	deep	1	66	0.01515152	1.0986123	0.01664564
34	William Wordsworth	diurnal	1	66	0.01515152	1.0986123	0.01664564
35	William Wordsworth	domes	1	66	0.01515152	1.0986123	0.01664564
36	William Wordsworth	doth	1	66	0.01515152	1.0986123	0.01664564
37	William Wordsworth	dull	1	66	0.01515152	1.0986123	0.01664564
38	William Wordsworth	earth	1	66	0.01515152	1.0986123	0.01664564
39	William Wordsworth	earth's	1	66	0.01515152	1.0986123	0.01664564
40	William Wordsworth	earthly	1	66	0.01515152	1.0986123	0.01664564
41	William Wordsworth	fears	1	66	0.01515152	1.0986123	0.01664564
42	William Wordsworth	feel	1	66	0.01515152	1.0986123	0.01664564
43	William Wordsworth	fields	1	66	0.01515152	1.0986123	0.01664564
44	William Wordsworth	force	1	66	0.01515152	1.0986123	0.01664564
45	William Wordsworth	garment	1	66	0.01515152	1.0986123	0.01664564
46	William Wordsworth	glideth	1	66	0.01515152	1.0986123	0.01664564
47	William Wordsworth	glittering	1	66	0.01515152	1.0986123	0.01664564
48	William Wordsworth	hears	1	66	0.01515152	1.0986123	0.01664564
49	William Wordsworth	hill	1	66	0.01515152	1.0986123	0.01664564
50	William Wordsworth	houses	1	66	0.01515152	1.0986123	0.01664564
51	William Wordsworth	human	1	66	0.01515152	1.0986123	0.01664564
52	William Wordsworth	lie	1	66	0.01515152	1.0986123	0.01664564

53	William Wordsworth	lying	1	66	0.01515152	1.0986123	0.01664564
54	William Wordsworth	majesty	1	66	0.01515152	1.0986123	0.01664564
55	William Wordsworth	mighty	1	66	0.01515152	1.0986123	0.01664564
56	William Wordsworth	morning	1	66	0.01515152	1.0986123	0.01664564
57	William Wordsworth	motion	1	66	0.01515152	1.0986123	0.01664564
58	William Wordsworth	ne'er	1	66	0.01515152	1.0986123	0.01664564
59	William Wordsworth	pass	1	66	0.01515152	1.0986123	0.01664564
60	William Wordsworth	river	1	66	0.01515152	1.0986123	0.01664564
61	William Wordsworth	rock	1	66	0.01515152	1.0986123	0.01664564
62	William Wordsworth	rocks	1	66	0.01515152	1.0986123	0.01664564
63	William Wordsworth	rolled	1	66	0.01515152	1.0986123	0.01664564
64	William Wordsworth	round	1	66	0.01515152	1.0986123	0.01664564
65	William Wordsworth	seal	1	66	0.01515152	1.0986123	0.01664564
66	William Wordsworth	ships	1	66	0.01515152	1.0986123	0.01664564
67	William Wordsworth	sight	1	66	0.01515152	1.0986123	0.01664564
68	William Wordsworth	silent	1	66	0.01515152	1.0986123	0.01664564
69	William Wordsworth	sky	1	66	0.01515152	1.0986123	0.01664564
70	William Wordsworth	slumber	1	66	0.01515152	1.0986123	0.01664564
71	William Wordsworth	smokeless	1	66	0.01515152	1.0986123	0.01664564
72	William Wordsworth	soul	1	66	0.01515152	1.0986123	0.01664564
73	William Wordsworth	spirit	1	66	0.01515152	1.0986123	0.01664564
74	William Wordsworth	splendour	1	66	0.01515152	1.0986123	0.01664564
75	William Wordsworth	steep	1	66	0.01515152	1.0986123	0.01664564
76	William Wordsworth	stones	1	66	0.01515152	1.0986123	0.01664564
77	William Wordsworth	sun	1	66	0.01515152	1.0986123	0.01664564
78	William Wordsworth	sweet	1	66	0.01515152	1.0986123	0.01664564
79	William Wordsworth	temples	1	66	0.01515152	1.0986123	0.01664564
80	William Wordsworth	theatres	1	66	0.01515152	1.0986123	0.01664564
81	William Wordsworth	touch	1	66	0.01515152	1.0986123	0.01664564
82	William Wordsworth	touching	1	66	0.01515152	1.0986123	0.01664564
83	William Wordsworth	towers	1	66	0.01515152	1.0986123	0.01664564
84	William Wordsworth	trees	1	66	0.01515152	1.0986123	0.01664564
85	William Wordsworth	valley	1	66	0.01515152	1.0986123	0.01664564
86	William Wordsworth	wear	1	66	0.01515152	1.0986123	0.01664564
87	William Shakespare	thou	4	98	0.04081633	0.4054651	0.01654960
88	William Shakespare	fair	3	98	0.03061224	0.4054651	0.01241220
89	William Shakespare	admit	1	98	0.01020408	1.0986123	0.01121033
90	William Shakespare	alteration	1	98	0.01020408	1.0986123	0.01121033
91	William Shakespare	bark	1	98	0.01020408	1.0986123	0.01121033
92	William Shakespare	bears	1	98	0.01020408	1.0986123	0.01121033
93	William Shakespare	bending	1	98	0.01020408	1.0986123	0.01121033
94	William Shakespare	bends	1	98	0.01020408	1.0986123	0.01121033
95	William Shakespare	brag	1	98	0.01020408	1.0986123	0.01121033
96	William Shakespare	breathe	1	98	0.01020408	1.0986123	0.01121033
97	William Shakespare	buds	1	98	0.01020408	1.0986123	0.01121033
98	William Shakespare	chance	1	98	0.01020408	1.0986123	0.01121033

```
99  William Shakespare    changing  1     98 0.01020408 1.0986123 0.01121033
100 William Shakespare      cheeks  1     98 0.01020408 1.0986123 0.01121033
101 William Shakespare     compare  1     98 0.01020408 1.0986123 0.01121033
102 William Shakespare     compass  1     98 0.01020408 1.0986123 0.01121033
103 William Shakespare   complexion  1     98 0.01020408 1.0986123 0.01121033
104 William Shakespare      darling  1     98 0.01020408 1.0986123 0.01121033
105 William Shakespare        date  1     98 0.01020408 1.0986123 0.01121033
106 William Shakespare         day  1     98 0.01020408 1.0986123 0.01121033
107 William Shakespare       death  1     98 0.01020408 1.0986123 0.01121033
108 William Shakespare     declines  1     98 0.01020408 1.0986123 0.01121033
109 William Shakespare      dimm'd  1     98 0.01020408 1.0986123 0.01121033
110 William Shakespare        doom  1     98 0.01020408 1.0986123 0.01121033
111 William Shakespare        edge  1     98 0.01020408 1.0986123 0.01121033
112 William Shakespare       error  1     98 0.01020408 1.0986123 0.01121033
113 William Shakespare        fade  1     98 0.01020408 1.0986123 0.01121033
114 William Shakespare       fixed  1     98 0.01020408 1.0986123 0.01121033
115 William Shakespare        fool  1     98 0.01020408 1.0986123 0.01121033
116 William Shakespare        gold  1     98 0.01020408 1.0986123 0.01121033
117 William Shakespare      grow'st  1     98 0.01020408 1.0986123 0.01121033
118 William Shakespare        hath  1     98 0.01020408 1.0986123 0.01121033
119 William Shakespare       height  1     98 0.01020408 1.0986123 0.01121033
120 William Shakespare         hot  1     98 0.01020408 1.0986123 0.01121033
121 William Shakespare       hours  1     98 0.01020408 1.0986123 0.01121033
122 William Shakespare impediments  1     98 0.01020408 1.0986123 0.01121033
123 William Shakespare       lease  1     98 0.01020408 1.0986123 0.01121033
124 William Shakespare       lines  1     98 0.01020408 1.0986123 0.01121033
125 William Shakespare        lips  1     98 0.01020408 1.0986123 0.01121033
126 William Shakespare       lives  1     98 0.01020408 1.0986123 0.01121033
127 William Shakespare        lose  1     98 0.01020408 1.0986123 0.01121033
128 William Shakespare       love's  1     98 0.01020408 1.0986123 0.01121033
129 William Shakespare       loved  1     98 0.01020408 1.0986123 0.01121033
130 William Shakespare       lovely  1     98 0.01020408 1.0986123 0.01121033
131 William Shakespare        mark  1     98 0.01020408 1.0986123 0.01121033
132 William Shakespare     marriage  1     98 0.01020408 1.0986123 0.01121033
133 William Shakespare       minds  1     98 0.01020408 1.0986123 0.01121033
134 William Shakespare     nature's  1     98 0.01020408 1.0986123 0.01121033
135 William Shakespare       ow'st  1     98 0.01020408 1.0986123 0.01121033
136 William Shakespare   possession  1     98 0.01020408 1.0986123 0.01121033
137 William Shakespare       proved  1     98 0.01020408 1.0986123 0.01121033
138 William Shakespare      remove  1     98 0.01020408 1.0986123 0.01121033
139 William Shakespare      remover  1     98 0.01020408 1.0986123 0.01121033
140 William Shakespare        rosy  1     98 0.01020408 1.0986123 0.01121033
141 William Shakespare       rough  1     98 0.01020408 1.0986123 0.01121033
142 William Shakespare       shade  1     98 0.01020408 1.0986123 0.01121033
[ reached 'max' / getOption("max.print") -- omitted 83 rows ]
```

REFERENCES:

(1) https://en.wikipedia.org/wiki/William_Blake
(2) https://en.wikipedia.org/wiki/William_Shakespeare
(3) https://en.wikipedia.org/wiki/William_Wordsworth
(4) https://www.poemhunter.com/