

STAT 413/613 Homework: Tidy Text

Yunting Chiu

2020-12-10

Contents

Instructions	1
1 Sentiment Analysis	1
2 tf-idf for Mark Twain's books	10
3 Extra Credit Podcasts	14

Instructions

1. Clone this homework repo to your homework directory as a new repo.
2. Rename the starter file under the analysis directory as `hw_01_yourname.Rmd` and use it for your solutions.
3. Modify the “author” field in the YAML header.
4. Stage and Commit R Markdown and HTML files (no PDF files).
5. **Push both .Rmd and HTML files to GitHub.**
 - Make sure you have knitted to HTML prior to staging, committing, and pushing your final submission.
6. **Commit each time you answer a part of question, e.g. 1.1**
7. **Push to GitHub after each major question**
8. When complete, submit a response in Canvas
 - Only include necessary code to answer the questions.
 - Most of the functions you use should be from the tidyverse.
 - Unnecessary Base R or other packages not covered in class will result in point deductions.
 - Use Pull requests and or email to ask me any questions. If you email, please ensure your most recent code is pushed to GitHub.

1 Sentiment Analysis

0. libraries

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
## v ggplot2 3.3.2      v purrr  0.3.4
## v tibble  3.0.3      v dplyr  1.0.2
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(tidytext)
library(gutenbergr)
```

1. Download the following two works from the early 20th century from Project Gutenberg:

- Upton Sinclair: “*The Jungle*” (1906)
- W.E.B. Du Bois: “*The Quest of the Silver Fleece*” (1911)

```
# check each gutenberg ID
#gutenberg_works() %>%
#(title == "The Jungle")
#gutenberg_works() %>%
#filter(str_detect(author, "Bois"))

theJungle <- gutenberg_download(140)
```

```
## Determining mirror for Project Gutenberg from http://www.gutenberg.org/robot/harvest
```

```
## Using mirror http://aleph.gutenberg.org
```

```
theQuest <- gutenberg_download(15265)
```

2. Write a function to take an argument of a downloaded book tibble and return it in tidy text format.

- The function must add line and chapter numbers as variables
- The function must unnest tokens at the word level
- The function must remove any Project Gutenberg formatting so only the words remain
- The function must remove any stop_words and filter out any NAs **8.1.2: - 0 You could have combined the regex into one with an | and not used the if blocks - would have saved lots of duplicate code.**

```
tidyJungleAndQuest <- function(tidyBooks){
  stopifnot(is.data.frame(tidyBooks))
  if(tidyBooks$gutenberg_id[[1]] == 140){
    theJungle %>%
      mutate(linenumber = row_number(),
             chapter = cumsum(str_detect(text, regex( "^chapter [\\divxlc]",
                                                       ignore_case = TRUE)))) %>%
    ungroup() %>%
    unnest_tokens(word, text) %>%
    mutate(word = str_extract(word, "[a-z']+")) %>%
    anti_join(stop_words, by = "word") %>%
    filter(!is.na(word)) -> theJungle
    return(theJungle)
  }else if(tidyBooks$gutenberg_id[[1]] == 15265){
    theQuest %>%
      mutate(linenumber = row_number(),
             text = recode(text, "_Contents_" = "Contents",
```

```

      "_Note_" = "Note"),
      chapter = cumsum(str_detect(text, regex( "(^_)([a-z]+)([-]{0,1})([a-z]+)(_$)",
                                                ignore_case = TRUE)))) %>%

ungroup() %>%
unnest_tokens(word, text) %>%
mutate(word = str_extract(word, "[a-z']+")) %>%
anti_join(stop_words, by = "word") %>%
filter(!is.na(word)) -> theQuest
return(theQuest)
}
}

```

3. Use the function from step 2

- Tidy each book and then add book and author as variables and save each tibble to a new variable.

```

theJungle %>%
  tidyJungleAndQuest() %>%
  mutate(book = "The Jungle",
         author = "Sinclair, Upton") -> theJungle

theQuest %>%
  tidyJungleAndQuest() %>%
  mutate(book = "The Quest of the Silver Fleece: A Novel",
         author = "Du Bois, W. E. B. (William Edward Burghardt)") -> theQuest

```

4. Use a dplyr function to combine the two tibbles into a new tibble.

- It should have 89,513 rows with 6 variables

```

theJungle %>%
  full_join(theQuest,
            by = c("gutenberg_id", "linenumber", "chapter",
                  "word", "book", "author")) -> jungleAndQuest

```

5. Measure the net sentiment using bing for each block of 50 lines

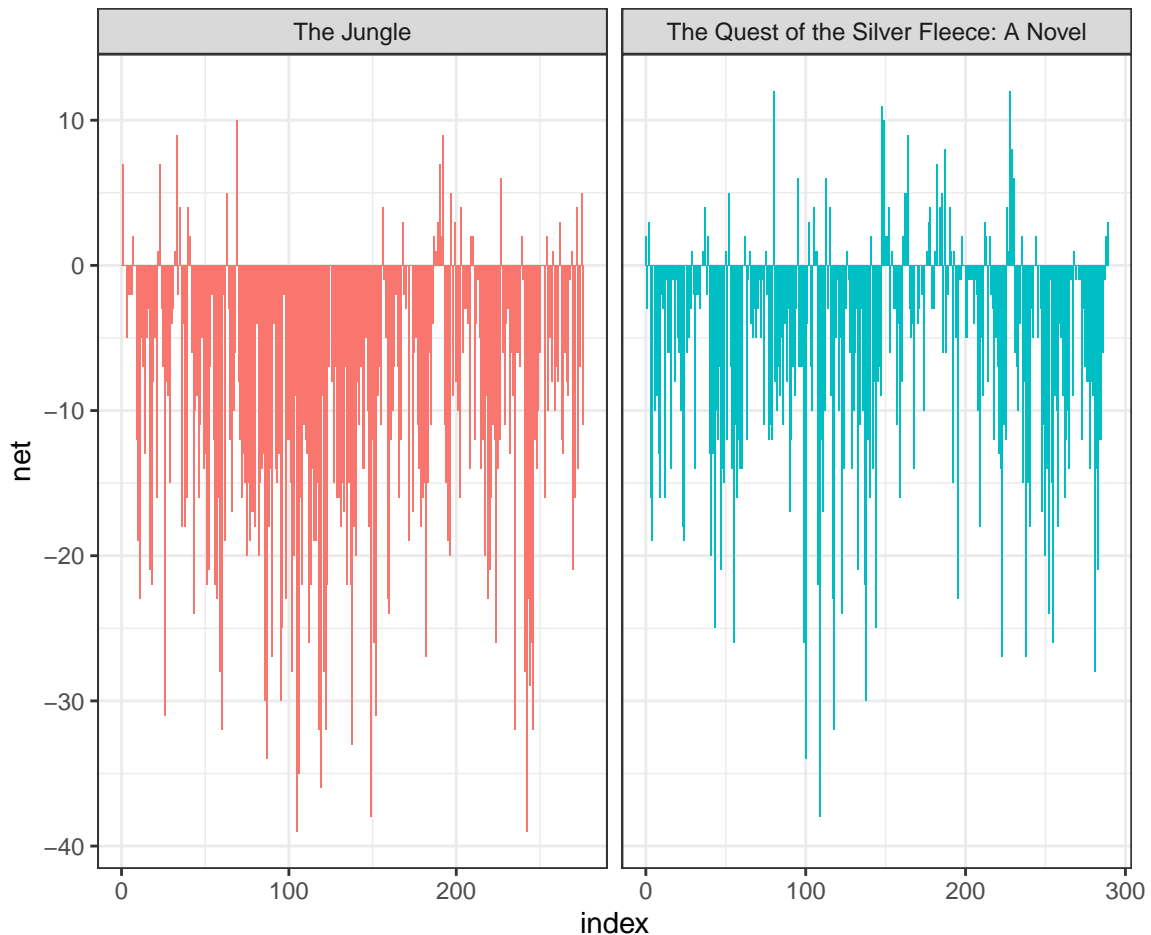
- Plot the sentiment for each book in an appropriate faceted plot - either line or column.
- Be sure to remove the legend.
- Save the plot to a variable
- Interpret the plots for each book and then compare them.
- **Interpretation:** In the following plot, The Quest of the Silver has more negative words than The Jungle. The majority of both contents have a negative sentiment.

```

jungleAndQuest %>%
  inner_join(get_sentiments("bing"), by = "word") %>%
  count(index = linenumber %/% 50, sentiment, book, sort = TRUE) %>%
  pivot_wider(names_from = sentiment, values_from = n,
              values_fill = list(n = 0)) %>%
  mutate(net = positive - negative) -> jungleAndQuestBing50

ggplot(data = jungleAndQuestBing50, aes(x = index, y = net, fill = book)) +
  geom_col(show.legend = FALSE) +
  theme_bw() +
  facet_wrap(~book, ncol = 2, scales = "free_x") -> bing50
plot(bing50)

```

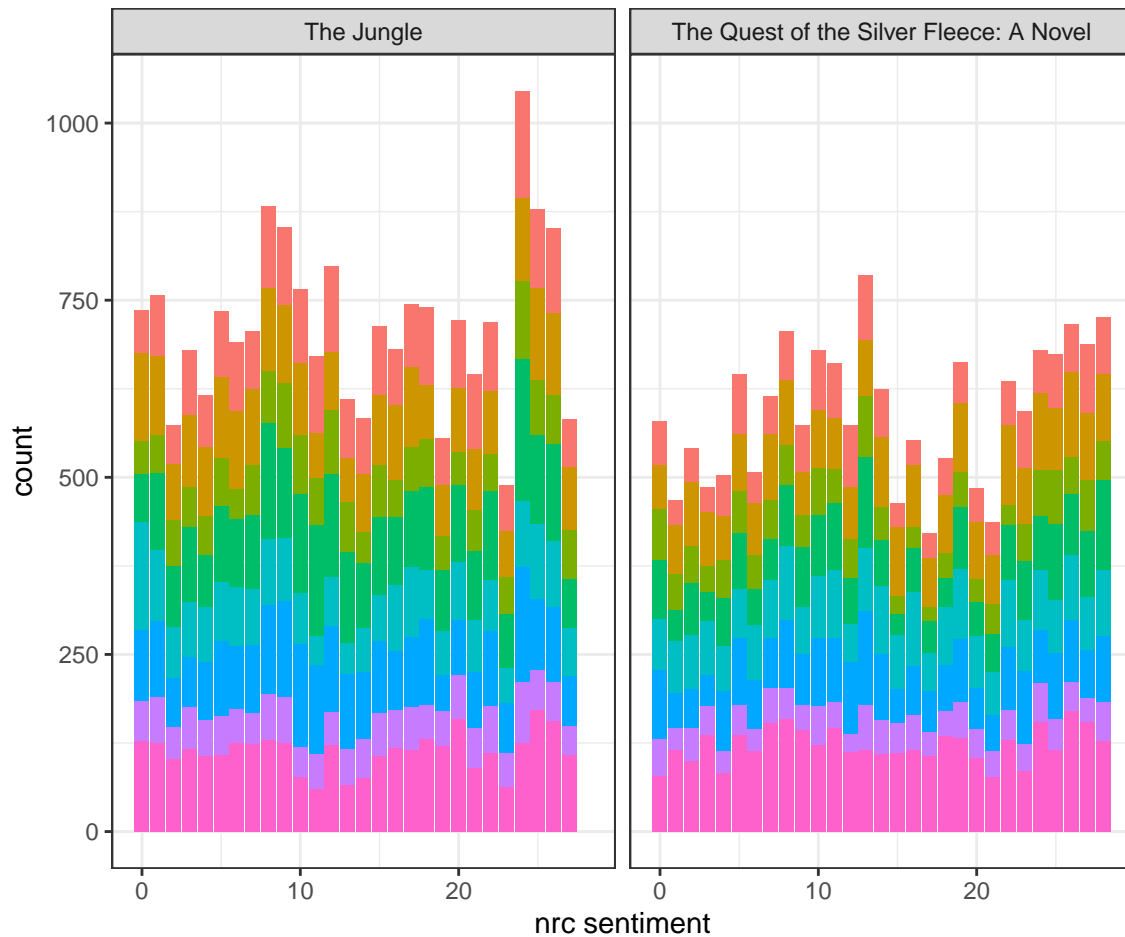


6. Measure the total for each nrc sentiment in each block of 500 lines and then,

- Filter out the “positive” and “negative” and save to a new variable. You should have 456 observations.
- Plot the count of the sentiments for each block in each book in an appropriate faceted plot with the books in two columns and the sentiments in 8 rows.
- Be sure to remove the legend.
- Interpret the plots for each book and then compare them.
- **Interpretation: Interestingly, these two books share the similarities on each block across 500 lines have abundant sentiment of structures. 8.1.6c: - 0.75 Instead of coord flip, should facet_grid sentiment by book**

```
get_sentiments("nrc") %>%
  filter(sentiment != "positive" & sentiment != "negative") %>%
  inner_join(jungleAndQuest, by = "word") %>%
  count(index = linenummer %/% 500, sentiment, book, sort = TRUE) ->
  nrcSentimentsNoPosNeg

ggplot(data = nrcSentimentsNoPosNeg,
       aes(x = index, y = n, fill = sentiment)) +
  geom_col(show.legend = F) +
  theme_bw() +
  facet_grid(~book) +
  labs(x = "nrc sentiment", y = "count")
```



7. Using `bing`, create a new data frame with the counts of the positive and negative sentiment words for each book.

- Show the 20 most frequent words across all book along with their book, sentiment, and count, in descending order by count.

```
jungleAndQuest %>%
  inner_join(get_sentiments("bing"), by = "word") %>%
  count(book, word, sentiment, sort = TRUE) %>%
  ungroup() -> bingSentimentsJungleAndQuest
```

```
head(bingSentimentsJungleAndQuest, 20)
```

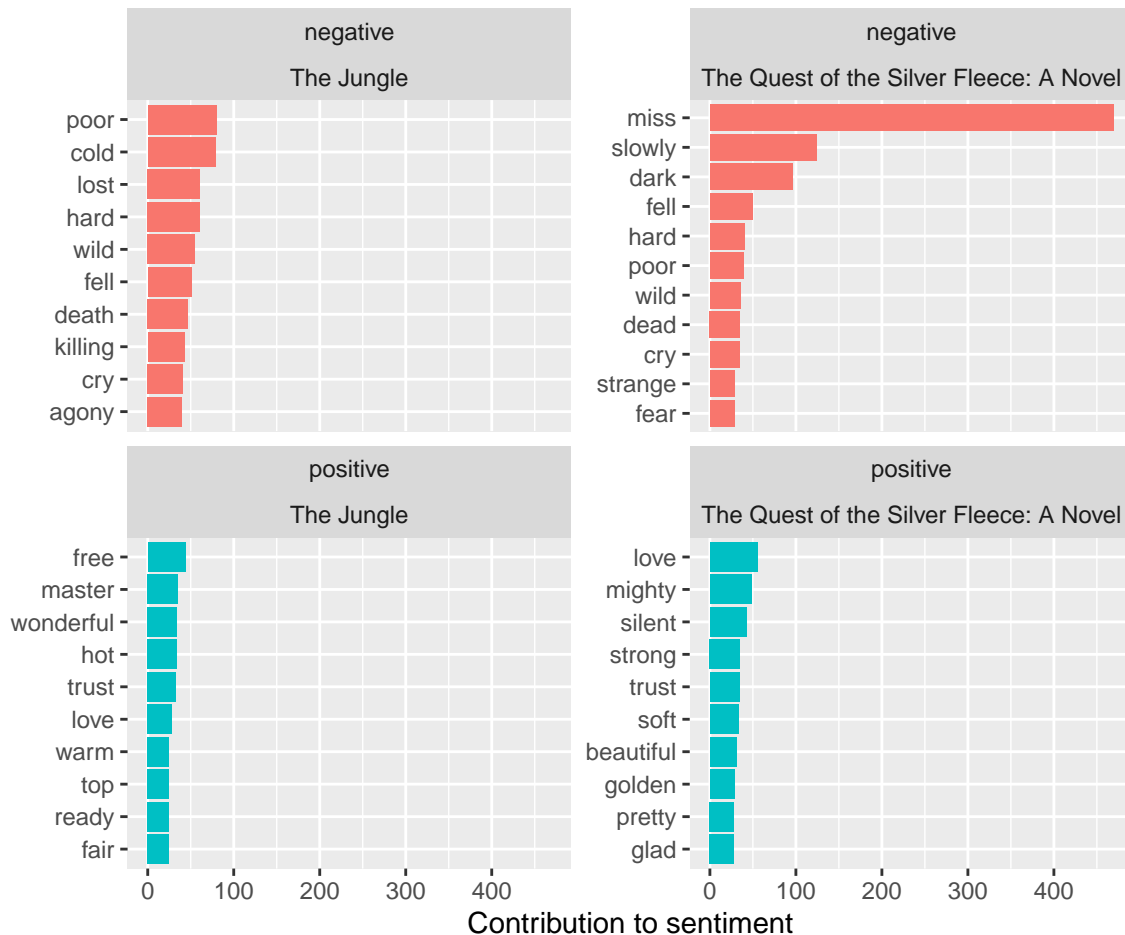
```
## # A tibble: 20 x 4
##   book          word  sentiment    n
##   <chr>         <chr>    <chr>   <int>
## 1 The Quest of the Silver Fleece: A Novel miss    negative   469
## 2 The Quest of the Silver Fleece: A Novel slowly  negative   124
## 3 The Quest of the Silver Fleece: A Novel dark   negative    96
## 4 The Jungle      poor    negative    80
## 5 The Jungle      cold    negative    79
## 6 The Jungle      hard    negative    61
## 7 The Jungle      lost    negative    61
## 8 The Jungle      wild    negative    55
## 9 The Quest of the Silver Fleece: A Novel love    positive    55
```

## 10	The Jungle	fell	negative	51
## 11	The Quest of the Silver Fleece: A Novel	fell	negative	50
## 12	The Quest of the Silver Fleece: A Novel	mighty	positive	48
## 13	The Jungle	death	negative	47
## 14	The Jungle	free	positive	44
## 15	The Jungle	killing	negative	43
## 16	The Quest of the Silver Fleece: A Novel	silent	positive	43
## 17	The Jungle	cry	negative	41
## 18	The Jungle	agony	negative	40
## 19	The Quest of the Silver Fleece: A Novel	hard	negative	40
## 20	The Jungle	dead	negative	39

8. Plot the top ten for each positive and negative sentiment faceting by book.

- Ensure each facet has the words in the proper order for that book.
- Identify any that may be inappropriate and should be excluded from the sentiment analysis.
- **We need to remove the word miss from the bing sentiment lexicon, because miss is not totally a negative word. For instance: I miss you baby. 8.1.8a: - 1 Did not get 10 words for each sentiment, Need to use reorder_within and scale_x_reordered 0 I miss you baby is negative so it is most likely the use as part of someone's name, Miss Jones, Miss Smith**

```
bingSentimentsJungleAndQuest %>%
  group_by(book, sentiment) %>%
  slice_max(order_by = n, n = 10) %>%
  mutate(word = reorder_within(word, n, book)) %>%
  ungroup() %>%
  ggplot(aes(word, n, fill = sentiment)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(sentiment ~ book, scales = "free_y") +
  labs(y = "Contribution to sentiment", x = NULL) +
  coord_flip() +
  scale_x_reordered()
```



```
# theme(strip.background = element_blank(), strip.placement = "outside") #review HW1
```

9. Remove the inappropriate word(s) from the analysis.

- Replot the top 10 for each sentiment per book from step 8.
- Interpret the plots
- **Interpretation:** After removing the word miss, we can detect that The Quest has many negative sentiments. Based on the top ten ranking for each, Negative sentiments are more than positive sentiments in both books.

```
# remove the word "miss" from the bing sentiment lexicon.
```

```
get_sentiments("bing") %>%
  filter(word != "miss") -> bing_no_miss
```

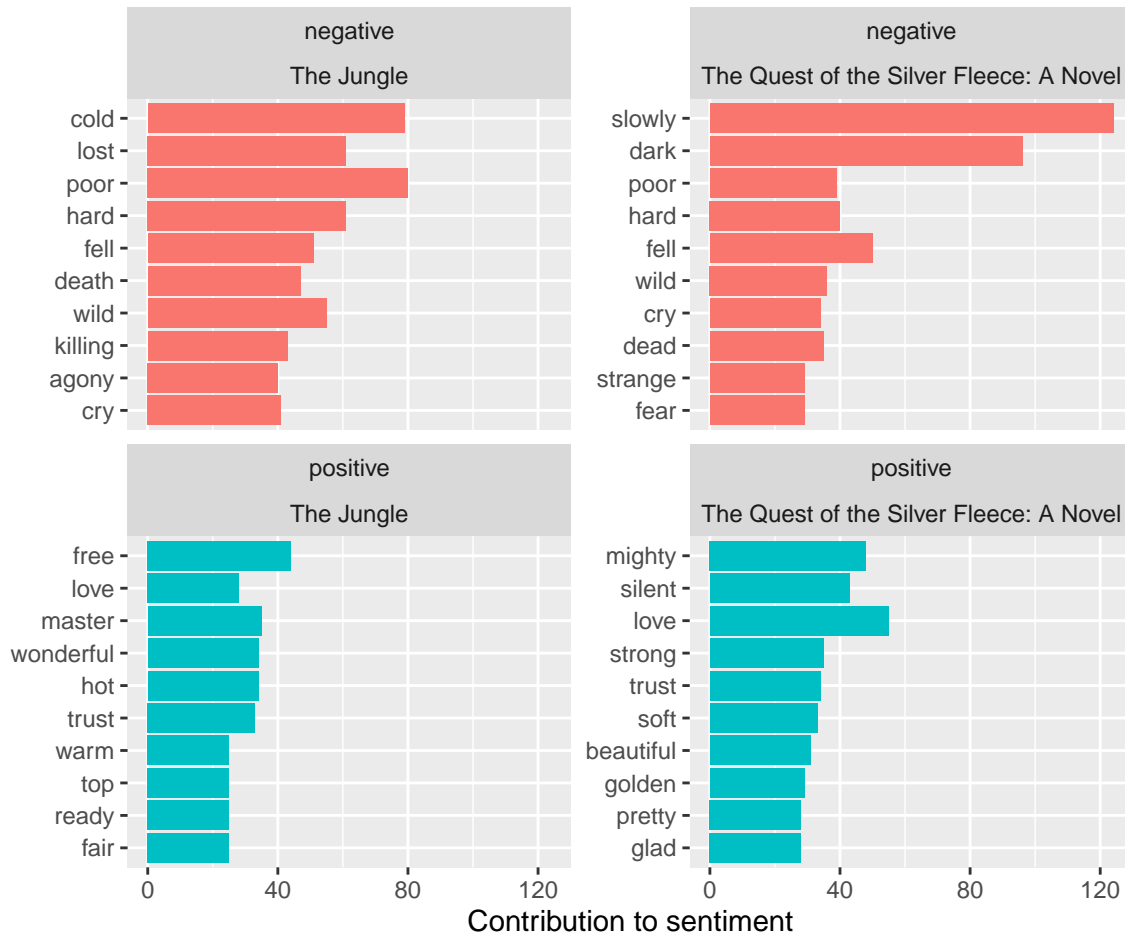
```
# redo the analysis from the beginning
```

```
jungleAndQuest %>%
  inner_join(bing_no_miss, by = "word") %>%
  count(book, word, sentiment, sort = TRUE) %>%
  ungroup() -> bing_word_counts
```

```
# visualize it
```

```
bing_word_counts %>%
  group_by(book, sentiment) %>%
  slice_max(order_by = n, n = 10) %>%
  ungroup() %>%
```

```
mutate(word = reorder(word, n)) %>%
ggplot(aes(x = word, y = n, fill = sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(sentiment ~ book, scales = "free_y") +
labs(y = "Contribution to sentiment", x = NULL) +
coord_flip()
```



10. Extra Credit

- Rerun the analysis from step 5 and recreate the plot with the title “Custom Bing”.
- Show both the original step 5 plot with the new plot in the same output graphic, one on top of the other.
- Interpret the plots
- **Interpretation:** although we filter out miss, the plot shows there is no big change in the analysis.

```
# original code in 1.5
jungleAndQuest %>%
  inner_join(get_sentiments("bing"), by = "word") %>%
  count(index = linenummer %/% 50, sentiment, book, sort = TRUE) %>%
  pivot_wider(names_from = sentiment, values_from = n,
              values_fill = list(n = 0)) %>%
  mutate(net = positive - negative) %>%
  ggplot(aes(x = index, y = net, fill = book)) +
  geom_col(show.legend = FALSE) +
```



```

theme_bw() +
facet_wrap(~book, ncol = 2, scales = "free_x") +
ggtitle("With Miss as Negative") -> p1

# No Miss
jungleAndQuest %>%
  inner_join(bing_no_miss, by = "word") %>%
  count(index = linenumber %/% 50, sentiment, book, sort = TRUE) %>%
  pivot_wider(names_from = sentiment, values_from = n,
              values_fill = list(n = 0)) %>%
  mutate(net = positive - negative) %>%
  ggplot(aes(x = index, y = net, fill = book)) +
  geom_col(show.legend = FALSE) +
  theme_bw() +
  facet_wrap(~book, ncol = 2, scales = "free_x") +
  ggtitle("Without Miss as Negative") -> p2

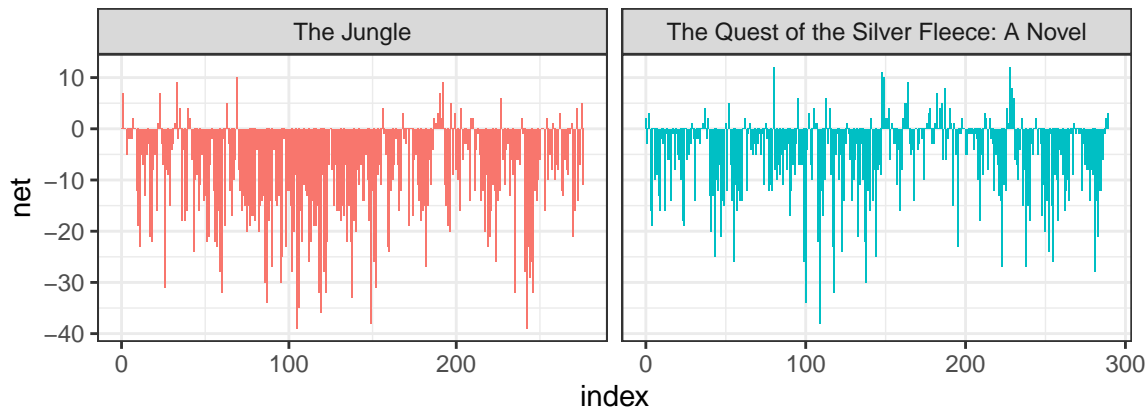
library(gridExtra)

##
## Attaching package: 'gridExtra'
## The following object is masked from 'package:dplyr':
##
##      combine

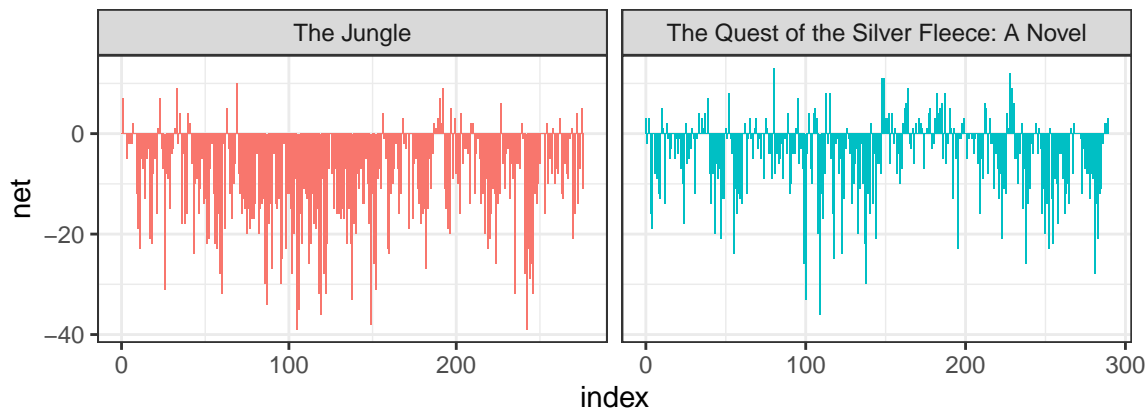
doubleBing50 <- grid.arrange(p1, p2, nrow = 2)

```

With Miss as Negative



Without Miss as Negative



2 tf-idf for Mark Twain's books

1. Download the following books from Author Mark Twain from Project Gutenberg

- Use the `meta_fields` argument to include the Book title as part of the download
- *Huckleberry Finn*, *Tom Sawyer*, *Connecticut Yankee in King Arthur's Court*, *Life on the Mississippi*, *Prince and the Pauper*, and *A Tramp Abroad*

```
gutenberg_works() %>%
  filter(str_detect(author, "Twain")) -> mt

markTwain <- gutenberg_download(c(76, 74, 86, 245, 1837, 119))

adventuresOfHuckleberryFinn <- gutenberg_download(76)
theAdventuresOfTomSawyer <- gutenberg_download(74)
aConnecticutYankeeInKingArthursCourt <- gutenberg_download(86)
lifeOnTheMississippi <- gutenberg_download(245)
thePrinceAndThePauper <- gutenberg_download(1837)
aTrampAbroad <- gutenberg_download(119)

markTwainBooks <- bind_rows(mutate(adventuresOfHuckleberryFinn, book = "Adventures Of Huckleberry Finn"),
  mutate(theAdventuresOfTomSawyer, book = "The Adventures Of Tom Sawyer"),
  mutate(aConnecticutYankeeInKingArthursCourt, book = "A Connecticut Yankee In King Arthurs Court"),
  mutate(lifeOnTheMississippi, book = "Life On The Mississippi"),
```

```
mutate(thePrinceAndThePauper, book = "The Prince And The Pauper"),
mutate(aTrampAbroad, book = "A Tramp Abroad"))
```

2. Modify your earlier function or create a new one to output a tf-idf ready dataframe (**leave the the stop words in the text**)

- Still unnest, remove any formatting, get rid of any NAs and then,
- Add the count for each word by title.
- Use your function to tidy the downloaded texts and save to a variable. It should have 57,130 rows **8.2.2: - 1 Function is too specific. It should take the df as an argument and then operate on it, regardless of which books are in the DF. This is too specific. In fact they have changed the library so the check for an exact number of rows errors out.**

```
readyTfIdf <- function(x){
  stopifnot(is.data.frame(markTwainBooks)) #,nrow(x) == 73326))
  markTwainBooks %>%
  unnest_tokens(word, text) %>%
  mutate(word = str_extract(word, "[a-z']+")) %>%
  filter(!is.na(word)) %>%
  count(book, word, sort = TRUE) -> markTwainBooks

  # calculate how many words in each book = total variable
  markTwainBooks %>%
  group_by(book) %>%
  summarize(total = sum(n), .groups = "keep") -> ttlWords

  markTwainBooks %>%
  left_join(ttlWords, by = "book") -> markTwainBooks
  return(markTwainBooks)
}
```

3. Calculate the tf-idf

- Save back to the data frame.

```
markTwainBooks %>%
  readyTfIdf() %>%
  bind_tf_idf(word, book, n) -> markTwainBooks
```

4. Plot the tf for each book using a faceted graph.

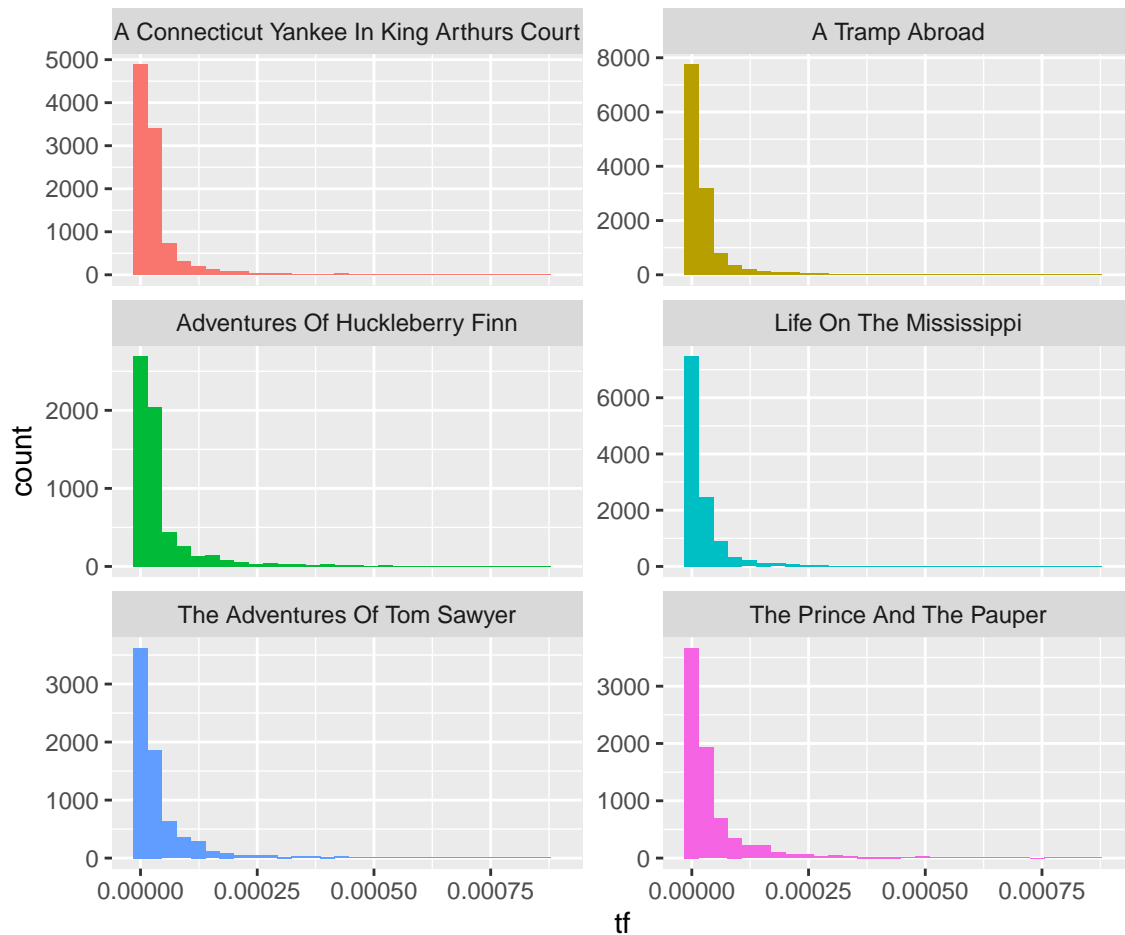
- Facet by book.

```
markTwainBooks %>%
  ggplot(aes(x = tf, fill = book)) +
  geom_histogram(show.legend = FALSE) +
  xlim(NA, 0.0009) +
  facet_wrap(~book, ncol = 2, scales = "free_y")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 852 rows containing non-finite values (stat_bin).
```

```
## Warning: Removed 6 rows containing missing values (geom_bar).
```



5. Show the words with the 15 highest tf-idfs across all books

- Only show those rows.
- How many look like possible names?
- possible names: Hendon, Becky, Huck, Joe, Potter, Launcelot, Merlin, Hugh.

```
markTwainBooks %>%
  arrange(desc(tf_idf)) %>%
  select(book, tf_idf, everything()) %>%
  head(15)
```

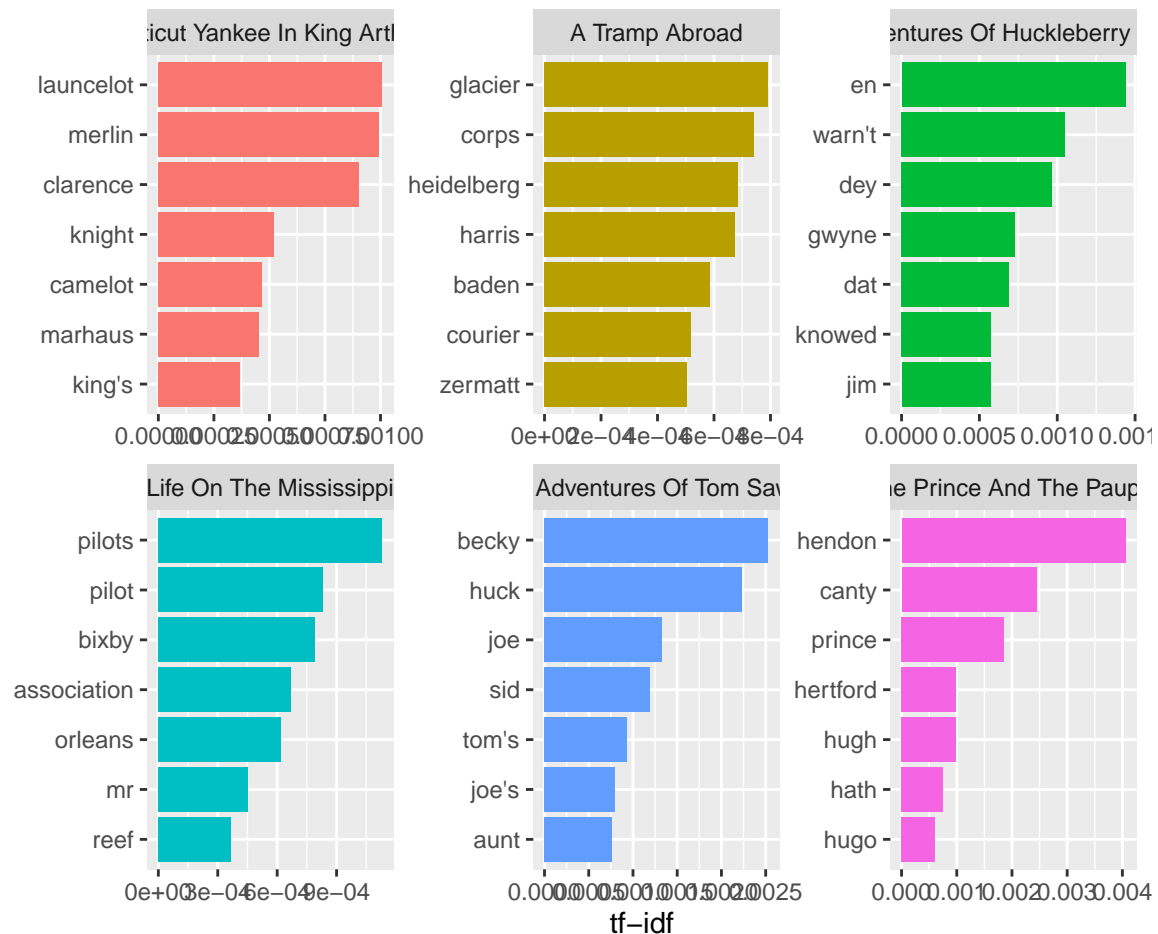
```
## # A tibble: 15 x 7
##   book                                tf_idf word      n total    tf   idf
##   <chr>                                <dbl> <chr>   <int> <int>  <dbl> <dbl>
## 1 The Prince And The Pauper          0.00406 hendon   161  71104 2.26e-3 1.79
## 2 The Adventures Of Tom Sawyer       0.00253 becky   102  72190 1.41e-3 1.79
## 3 The Prince And The Pauper          0.00244 canty    97  71104 1.36e-3 1.79
## 4 The Adventures Of Tom Sawyer       0.00223 huck    232  72190 3.21e-3 0.693
## 5 The Prince And The Pauper          0.00186 prince  191  71104 2.69e-3 0.693
## 6 Adventures Of Huckleberry Finn     0.00144 en      235 113227 2.08e-3 0.693
## 7 The Adventures Of Tom Sawyer       0.00133 joe     138  72190 1.91e-3 0.693
## 8 The Adventures Of Tom Sawyer       0.00119 sid      78  72190 1.08e-3 1.10
## 9 Life On The Mississippi            0.00113 pilots   93 147364 6.31e-4 1.79
## 10 Adventures Of Huckleberry Finn    0.00105 warn't  293 113227 2.59e-3 0.405
## 11 A Connecticut Yankee In King Ar~ 0.00101 launcel~ 67 119087 5.63e-4 1.79
```

```
## 12 A Connecticut Yankee In King Ar~ 0.000993 merlin      66 119087 5.54e-4 1.79
## 13 The Prince And The Pauper        0.000983 hertford    39  71104 5.48e-4 1.79
## 14 The Prince And The Pauper        0.000983 hugh       39  71104 5.48e-4 1.79
## 15 Adventures Of Huckleberry Finn   0.000965 dey        61 113227 5.39e-4 1.79
```

6. Plot the top 7 tf_idf words from each book.

- Sort in descending order of tf_idf
- Interpret the plots.
- **Interpretation:** Many top 7 tf_idf words are the real name or English honorifics, such as aunt and prince. Most of the time the words of speeches are nouns.

```
markTwainBooks %>%
  arrange(desc(tf_idf)) %>%
  mutate(word = parse_factor(word, levels = rev(unique(word)))) %>%
  group_by(book) %>%
  slice_max(order_by = tf_idf, n = 7) %>%
  ungroup() %>%
  ggplot(aes(word, tf_idf, fill = book)) +
  geom_col(show.legend = FALSE) +
  labs(x = NULL, y = "tf-idf") +
  facet_wrap(~book, scales = "free") +
  coord_flip()
```



3 Extra Credit Podcasts

- Choose **One** of the following podcasts and answer the questions below:

a. Sentiment Preserving Fake Reviews

The Original paper

b. Data in Life: Authorship Attribution in Lennon-McCartney Songs

c. Newsha Ajami| Improving Urban Water Systems Through Data Science, Public Policy and Engineering

1. What are some key ideas from this podcast relevant to text sentiment analysis/authorship attribution (1, or 2) or working with large diverse data sets (3)?

- **I'd choose Sentiment Preserving Fake Reviews for answering question.**

- Authorship attribution has been a standard problem within Natural Language Processing (NLP)
- Advanced neural language models (NLMs) are widely used in sequence generation tasks because they are able to produce fluent and meaningful sentences. They can also be used to generate fake reviews, which can then be used to attack online review systems and influence the buying decisions of online shoppers. As Mr. Adelani mentioned, reviews can have great influence on buyers' decisions and it is not only the content but also the volume of the reviews that plays a crucial role.
- Mr. Adelani discussed the idea of generating sentiment-preserving fake online review by using an example of a positive/negative review and generate more fake reviews similar to it. It is relevant to sentiment analysis/authorship attribution because it preserves the sentiment of the writer and it is like paraphrasing the original review.
- It requires experts to train a tailored LM for a specific topic. A low-skilled threat model can be built just by combining publicly available LMs and show that the produced fake reviews can fool both humans and machines.
- Then the GPT-2 NLM is used to generate a large number of high-quality reviews based on a review with the desired sentiment and then using a BERT based text classifier (with accuracy of 96%) to filter out reviews with undesired sentiments.
- Because none of the words in the review are modified, fluent samples like the training data can be generated from the learned distribution. A subjective evaluation with 80 participants demonstrated that this simple method can produce reviews that are as fluent as those written by people.
- Three countermeasures, Grover, GLTR, and OpenAI GPT-2 detector, were found to be difficult to accurately detect fake review.

2. How do you think the ideas discussed may be relevant in your future work?

- As technology constantly evolves, computer comes to compiling and analyzing structured data through Natural Language Processing. For example, as we learned sentiment words data frame this week, we can directly analyze many words and articles by computer, avoid too subjective effects by human beings. So how to organize and optimal words NLP maybe is the most relevant in the future for us.
- Note: the elegant way to tidy book

```
tidydocs <- function(df){
  nums <- c("_One_(?!:)|_Two_(?!:)|_Three_(?!:)|_Four_(?!:)|_Five_(?!:)|_Six_(?!:)|_Seven_(?!:)|_Eight_(?!:)|_Nine_(?!:)|_Ten_(?!:)|_Eleven_(?!:)|_Twelve_(?!:)|_Thirteen_(?!:)|_Fourteen_(?!:)|_Fifteen_(?!:)|_Sixteen_(?!:)|_Seventeen_(?!:)|_Eighteen_(?!:)|_Nineteen_(?!:)|_Twenty_(?!:)|_TwentyOne_(?!:)|_TwentyTwo_(?!:)|_TwentyThree_(?!:)|_TwentyFour_(?!:)|_TwentyFive_(?!:)|_TwentySix_(?!:)|_TwentySeven_(?!:)|_TwentyEight_(?!:)|_TwentyNine_(?!:)|_Thirty_(?!:)|_ThirtyOne_(?!:)|_ThirtyTwo_(?!:)|_ThirtyThree_(?!:)|_ThirtyFour_(?!:)|_ThirtyFive_(?!:)|_ThirtySix_(?!:)|_ThirtySeven_(?!:)|_ThirtyEight_(?!:)|_ThirtyNine_(?!:)|_Forty_(?!:)|_FortyOne_(?!:)|_FortyTwo_(?!:)|_FortyThree_(?!:)|_FortyFour_(?!:)|_FortyFive_(?!:)|_FortySix_(?!:)|_FortySeven_(?!:)|_FortyEight_(?!:)|_FortyNine_(?!:)|_Fifty_(?!:)|_FiftyOne_(?!:)|_FiftyTwo_(?!:)|_FiftyThree_(?!:)|_FiftyFour_(?!:)|_FiftyFive_(?!:)|_FiftySix_(?!:)|_FiftySeven_(?!:)|_FiftyEight_(?!:)|_FiftyNine_(?!:)|_Sixty_(?!:)|_SixtyOne_(?!:)|_SixtyTwo_(?!:)|_SixtyThree_(?!:)|_SixtyFour_(?!:)|_SixtyFive_(?!:)|_SixtySix_(?!:)|_SixtySeven_(?!:)|_SixtyEight_(?!:)|_SixtyNine_(?!:)|_Seventy_(?!:)|_SeventyOne_(?!:)|_SeventyTwo_(?!:)|_SeventyThree_(?!:)|_SeventyFour_(?!:)|_SeventyFive_(?!:)|_SeventySix_(?!:)|_SeventySeven_(?!:)|_SeventyEight_(?!:)|_SeventyNine_(?!:)|_Eighty_(?!:)|_EightyOne_(?!:)|_EightyTwo_(?!:)|_EightyThree_(?!:)|_EightyFour_(?!:)|_EightyFive_(?!:)|_EightySix_(?!:)|_EightySeven_(?!:)|_EightyEight_(?!:)|_EightyNine_(?!:)|_Ninety_(?!:)|_NinetyOne_(?!:)|_NinetyTwo_(?!:)|_NinetyThree_(?!:)|_NinetyFour_(?!:)|_NinetyFive_(?!:)|_NinetySix_(?!:)|_NinetySeven_(?!:)|_NinetyEight_(?!:)|_NinetyNine_(?!:)|_OneHundred_(?!:)|_OneHundredOne_(?!:)|_OneHundredTwo_(?!:)|_OneHundredThree_(?!:)|_OneHundredFour_(?!:)|_OneHundredFive_(?!:)|_OneHundredSix_(?!:)|_OneHundredSeven_(?!:)|_OneHundredEight_(?!:)|_OneHundredNine_(?!:)|_OneHundredTen_(?!:)|_OneHundredEleven_(?!:)|_OneHundredTwelve_(?!:)|_OneHundredThirteen_(?!:)|_OneHundredFourteen_(?!:)|_OneHundredFifteen_(?!:)|_OneHundredSixteen_(?!:)|_OneHundredSeventeen_(?!:)|_OneHundredEighteen_(?!:)|_OneHundredNineteen_(?!:)|_OneHundredTwenty_(?!:)|_OneHundredTwentyOne_(?!:)|_OneHundredTwentyTwo_(?!:)|_OneHundredTwentyThree_(?!:)|_OneHundredTwentyFour_(?!:)|_OneHundredTwentyFive_(?!:)|_OneHundredTwentySix_(?!:)|_OneHundredTwentySeven_(?!:)|_OneHundredTwentyEight_(?!:)|_OneHundredTwentyNine_(?!:)|_OneHundredThirty_(?!:)|_OneHundredThirtyOne_(?!:)|_OneHundredThirtyTwo_(?!:)|_OneHundredThirtyThree_(?!:)|_OneHundredThirtyFour_(?!:)|_OneHundredThirtyFive_(?!:)|_OneHundredThirtySix_(?!:)|_OneHundredThirtySeven_(?!:)|_OneHundredThirtyEight_(?!:)|_OneHundredThirtyNine_(?!:)|_OneHundredForty_(?!:)|_OneHundredFortyOne_(?!:)|_OneHundredFortyTwo_(?!:)|_OneHundredFortyThree_(?!:)|_OneHundredFortyFour_(?!:)|_OneHundredFortyFive_(?!:)|_OneHundredFortySix_(?!:)|_OneHundredFortySeven_(?!:)|_OneHundredFortyEight_(?!:)|_OneHundredFortyNine_(?!:)|_OneHundredFifty_(?!:)|_OneHundredFiftyOne_(?!:)|_OneHundredFiftyTwo_(?!:)|_OneHundredFiftyThree_(?!:)|_OneHundredFiftyFour_(?!:)|_OneHundredFiftyFive_(?!:)|_OneHundredFiftySix_(?!:)|_OneHundredFiftySeven_(?!:)|_OneHundredFiftyEight_(?!:)|_OneHundredFiftyNine_(?!:)|_OneHundredSixty_(?!:)|_OneHundredSixtyOne_(?!:)|_OneHundredSixtyTwo_(?!:)|_OneHundredSixtyThree_(?!:)|_OneHundredSixtyFour_(?!:)|_OneHundredSixtyFive_(?!:)|_OneHundredSixtySix_(?!:)|_OneHundredSixtySeven_(?!:)|_OneHundredSixtyEight_(?!:)|_OneHundredSixtyNine_(?!:)|_OneHundredSeventy_(?!:)|_OneHundredSeventyOne_(?!:)|_OneHundredSeventyTwo_(?!:)|_OneHundredSeventyThree_(?!:)|_OneHundredSeventyFour_(?!:)|_OneHundredSeventyFive_(?!:)|_OneHundredSeventySix_(?!:)|_OneHundredSeventySeven_(?!:)|_OneHundredSeventyEight_(?!:)|_OneHundredSeventyNine_(?!:)|_OneHundredEighty_(?!:)|_OneHundredEightyOne_(?!:)|_OneHundredEightyTwo_(?!:)|_OneHundredEightyThree_(?!:)|_OneHundredEightyFour_(?!:)|_OneHundredEightyFive_(?!:)|_OneHundredEightySix_(?!:)|_OneHundredEightySeven_(?!:)|_OneHundredEightyEight_(?!:)|_OneHundredEightyNine_(?!:)|_OneHundredNinety_(?!:)|_OneHundredNinetyOne_(?!:)|_OneHundredNinetyTwo_(?!:)|_OneHundredNinetyThree_(?!:)|_OneHundredNinetyFour_(?!:)|_OneHundredNinetyFive_(?!:)|_OneHundredNinetySix_(?!:)|_OneHundredNinetySeven_(?!:)|_OneHundredNinetyEight_(?!:)|_OneHundredNinetyNine_(?!:)|_TwoHundred_(?!:)|_TwoHundredOne_(?!:)|_TwoHundredTwo_(?!:)|_TwoHundredThree_(?!:)|_TwoHundredFour_(?!:)|_TwoHundredFive_(?!:)|_TwoHundredSix_(?!:)|_TwoHundredSeven_(?!:)|_TwoHundredEight_(?!:)|_TwoHundredNine_(?!:)|_TwoHundredTen_(?!:)|_TwoHundredEleven_(?!:)|_TwoHundredTwelve_(?!:)|_TwoHundredThirteen_(?!:)|_TwoHundredFourteen_(?!:)|_TwoHundredFifteen_(?!:)|_TwoHundredSixteen_(?!:)|_TwoHundredSeventeen_(?!:)|_TwoHundredEighteen_(?!:)|_TwoHundredNineteen_(?!:)|_TwoHundredTwenty_(?!:)|_TwoHundredTwentyOne_(?!:)|_TwoHundredTwentyTwo_(?!:)|_TwoHundredTwentyThree_(?!:)|_TwoHundredTwentyFour_(?!:)|_TwoHundredTwentyFive_(?!:)|_TwoHundredTwentySix_(?!:)|_TwoHundredTwentySeven_(?!:)|_TwoHundredTwentyEight_(?!:)|_TwoHundredTwentyNine_(?!:)|_TwoHundredThirty_(?!:)|_TwoHundredThirtyOne_(?!:)|_TwoHundredThirtyTwo_(?!:)|_TwoHundredThirtyThree_(?!:)|_TwoHundredThirtyFour_(?!:)|_TwoHundredThirtyFive_(?!:)|_TwoHundredThirtySix_(?!:)|_TwoHundredThirtySeven_(?!:)|_TwoHundredThirtyEight_(?!:)|_TwoHundredThirtyNine_(?!:)|_TwoHundredForty_(?!:)|_TwoHundredFortyOne_(?!:)|_TwoHundredFortyTwo_(?!:)|_TwoHundredFortyThree_(?!:)|_TwoHundredFortyFour_(?!:)|_TwoHundredFortyFive_(?!:)|_TwoHundredFortySix_(?!:)|_TwoHundredFortySeven_(?!:)|_TwoHundredFortyEight_(?!:)|_TwoHundredFortyNine_(?!:)|_TwoHundredFifty_(?!:)|_TwoHundredFiftyOne_(?!:)|_TwoHundredFiftyTwo_(?!:)|_TwoHundredFiftyThree_(?!:)|_TwoHundredFiftyFour_(?!:)|_TwoHundredFiftyFive_(?!:)|_TwoHundredFiftySix_(?!:)|_TwoHundredFiftySeven_(?!:)|_TwoHundredFiftyEight_(?!:)|_TwoHundredFiftyNine_(?!:)|_TwoHundredSixty_(?!:)|_TwoHundredSixtyOne_(?!:)|_TwoHundredSixtyTwo_(?!:)|_TwoHundredSixtyThree_(?!:)|_TwoHundredSixtyFour_(?!:)|_TwoHundredSixtyFive_(?!:)|_TwoHundredSixtySix_(?!:)|_TwoHundredSixtySeven_(?!:)|_TwoHundredSixtyEight_(?!:)|_TwoHundredSixtyNine_(?!:)|_TwoHundredSeventy_(?!:)|_TwoHundredSeventyOne_(?!:)|_TwoHundredSeventyTwo_(?!:)|_TwoHundredSeventyThree_(?!:)|_TwoHundredSeventyFour_(?!:)|_TwoHundredSeventyFive_(?!:)|_TwoHundredSeventySix_(?!:)|_TwoHundredSeventySeven_(?!:)|_TwoHundredSeventyEight_(?!:)|_TwoHundredSeventyNine_(?!:)|_TwoHundredEighty_(?!:)|_TwoHundredEightyOne_(?!:)|_TwoHundredEightyTwo_(?!:)|_TwoHundredEightyThree_(?!:)|_TwoHundredEightyFour_(?!:)|_TwoHundredEightyFive_(?!:)|_TwoHundredEightySix_(?!:)|_TwoHundredEightySeven_(?!:)|_TwoHundredEightyEight_(?!:)|_TwoHundredEightyNine_(?!:)|_TwoHundredNinety_(?!:)|_TwoHundredNinetyOne_(?!:)|_TwoHundredNinetyTwo_(?!:)|_TwoHundredNinetyThree_(?!:)|_TwoHundredNinetyFour_(?!:)|_TwoHundredNinetyFive_(?!:)|_TwoHundredNinetySix_(?!:)|_TwoHundredNinetySeven_(?!:)|_TwoHundredNinetyEight_(?!:)|_TwoHundredNinetyNine_(?!:)|_ThreeHundred_(?!:)|_ThreeHundredOne_(?!:)|_ThreeHundredTwo_(?!:)|_ThreeHundredThree_(?!:)|_ThreeHundredFour_(?!:)|_ThreeHundredFive_(?!:)|_ThreeHundredSix_(?!:)|_ThreeHundredSeven_(?!:)|_ThreeHundredEight_(?!:)|_ThreeHundredNine_(?!:)|_ThreeHundredTen_(?!:)|_ThreeHundredEleven_(?!:)|_ThreeHundredTwelve_(?!:)|_ThreeHundredThirteen_(?!:)|_ThreeHundredFourteen_(?!:)|_ThreeHundredFifteen_(?!:)|_ThreeHundredSixteen_(?!:)|_ThreeHundredSeventeen_(?!:)|_ThreeHundredEighteen_(?!:)|_ThreeHundredNineteen_(?!:)|_ThreeHundredTwenty_(?!:)|_ThreeHundredTwentyOne_(?!:)|_ThreeHundredTwentyTwo_(?!:)|_ThreeHundredTwentyThree_(?!:)|_ThreeHundredTwentyFour_(?!:)|_ThreeHundredTwentyFive_(?!:)|_ThreeHundredTwentySix_(?!:)|_ThreeHundredTwentySeven_(?!:)|_ThreeHundredTwentyEight_(?!:)|_ThreeHundredTwentyNine_(?!:)|_ThreeHundredThirty_(?!:)|_ThreeHundredThirtyOne_(?!:)|_ThreeHundredThirtyTwo_(?!:)|_ThreeHundredThirtyThree_(?!:)|_ThreeHundredThirtyFour_(?!:)|_ThreeHundredThirtyFive_(?!:)|_ThreeHundredThirtySix_(?!:)|_ThreeHundredThirtySeven_(?!:)|_ThreeHundredThirtyEight_(?!:)|_ThreeHundredThirtyNine_(?!:)|_ThreeHundredForty_(?!:)|_ThreeHundredFortyOne_(?!:)|_ThreeHundredFortyTwo_(?!:)|_ThreeHundredFortyThree_(?!:)|_ThreeHundredFortyFour_(?!:)|_ThreeHundredFortyFive_(?!:)|_ThreeHundredFortySix_(?!:)|_ThreeHundredFortySeven_(?!:)|_ThreeHundredFortyEight_(?!:)|_ThreeHundredFortyNine_(?!:)|_ThreeHundredFifty_(?!:)|_ThreeHundredFiftyOne_(?!:)|_ThreeHundredFiftyTwo_(?!:)|_ThreeHundredFiftyThree_(?!:)|_ThreeHundredFiftyFour_(?!:)|_ThreeHundredFiftyFive_(?!:)|_ThreeHundredFiftySix_(?!:)|_ThreeHundredFiftySeven_(?!:)|_ThreeHundredFiftyEight_(?!:)|_ThreeHundredFiftyNine_(?!:)|_ThreeHundredSixty_(?!:)|_ThreeHundredSixtyOne_(?!:)|_ThreeHundredSixtyTwo_(?!:)|_ThreeHundredSixtyThree_(?!:)|_ThreeHundredSixtyFour_(?!:)|_ThreeHundredSixtyFive_(?!:)|_ThreeHundredSixtySix_(?!:)|_ThreeHundredSixtySeven_(?!:)|_ThreeHundredSixtyEight_(?!:)|_ThreeHundredSixtyNine_(?!:)|_ThreeHundredSeventy_(?!:)|_ThreeHundredSeventyOne_(?!:)|_ThreeHundredSeventyTwo_(?!:)|_ThreeHundredSeventyThree_(?!:)|_ThreeHundredSeventyFour_(?!:)|_ThreeHundredSeventyFive_(?!:)|_ThreeHundredSeventySix_(?!:)|_ThreeHundredSeventySeven_(?!:)|_ThreeHundredSeventyEight_(?!:)|_ThreeHundredSeventyNine_(?!:)|_ThreeHundredEighty_(?!:)|_ThreeHundredEightyOne_(?!:)|_ThreeHundredEightyTwo_(?!:)|_ThreeHundredEightyThree_(?!:)|_ThreeHundredEightyFour_(?!:)|_ThreeHundredEightyFive_(?!:)|_ThreeHundredEightySix_(?!:)|_ThreeHundredEightySeven_(?!:)|_ThreeHundredEightyEight_(?!:)|_ThreeHundredEightyNine_(?!:)|_ThreeHundredNinety_(?!:)|_ThreeHundredNinetyOne_(?!:)|_ThreeHundredNinetyTwo_(?!:)|_ThreeHundredNinetyThree_(?!:)|_ThreeHundredNinetyFour_(?!:)|_ThreeHundredNinetyFive_(?!:)|_ThreeHundredNinetySix_(?!:)|_ThreeHundredNinetySeven_(?!:)|_ThreeHundredNinetyEight_(?!:)|_ThreeHundredNinetyNine_(?!:)|_FourHundred_(?!:)|_FourHundredOne_(?!:)|_FourHundredTwo_(?!:)|_FourHundredThree_(?!:)|_FourHundredFour_(?!:)|_FourHundredFive_(?!:)|_FourHundredSix_(?!:)|_FourHundredSeven_(?!:)|_FourHundredEight_(?!:)|_FourHundredNine_(?!:)|_FourHundredTen_(?!:)|_FourHundredEleven_(?!:)|_FourHundredTwelve_(?!:)|_FourHundredThirteen_(?!:)|_FourHundredFourteen_(?!:)|_FourHundredFifteen_(?!:)|_FourHundredSixteen_(?!:)|_FourHundredSeventeen_(?!:)|_FourHundredEighteen_(?!:)|_FourHundredNineteen_(?!:)|_FourHundredTwenty_(?!:)|_FourHundredTwentyOne_(?!:)|_FourHundredTwentyTwo_(?!:)|_FourHundredTwentyThree_(?!:)|_FourHundredTwentyFour_(?!:)|_FourHundredTwentyFive_(?!:)|_FourHundredTwentySix_(?!:)|_FourHundredTwentySeven_(?!:)|_FourHundredTwentyEight_(?!:)|_FourHundredTwentyNine_(?!:)|_FourHundredThirty_(?!:)|_FourHundredThirtyOne_(?!:)|_FourHundredThirtyTwo_(?!:)|_FourHundredThirtyThree_(?!:)|_FourHundredThirtyFour_(?!:)|_FourHundredThirtyFive_(?!:)|_FourHundredThirtySix_(?!:)|_FourHundredThirtySeven_(?!:)|_FourHundredThirtyEight_(?!:)|_FourHundredThirtyNine_(?!:)|_FourHundredForty_(?!:)|_FourHundredFortyOne_(?!:)|_FourHundredFortyTwo_(?!:)|_FourHundredFortyThree_(?!:)|_FourHundredFortyFour_(?!:)|_FourHundredFortyFive_(?!:)|_FourHundredFortySix_(?!:)|_FourHundredFortySeven_(?!:)|_FourHundredFortyEight_(?!:)|_FourHundredFortyNine_(?!:)|_FourHundredFifty_(?!:)|_FourHundredFiftyOne_(?!:)|_FourHundredFiftyTwo_(?!:)|_FourHundredFiftyThree_(?!:)|_FourHundredFiftyFour_(?!:)|_FourHundredFiftyFive_(?!:)|_FourHundredFiftySix_(?!:)|_FourHundredFiftySeven_(?!:)|_FourHundredFiftyEight_(?!:)|_FourHundredFiftyNine_(?!:)|_FourHundredSixty_(?!:)|_FourHundredSixtyOne_(?!:)|_FourHundredSixtyTwo_(?!:)|_FourHundredSixtyThree_(?!:)|_FourHundredSixtyFour_(?!:)|_FourHundredSixtyFive_(?!:)|_FourHundredSixtySix_(?!:)|_FourHundredSixtySeven_(?!:)|_FourHundredSixtyEight_(?!:)|_FourHundredSixtyNine_(?!:)|_FourHundredSeventy_(?!:)|_FourHundredSeventyOne_(?!:)|_FourHundredSeventyTwo_(?!:)|_FourHundredSeventyThree_(?!:)|_FourHundredSeventyFour_(?!:)|_FourHundredSeventyFive_(?!:)|_FourHundredSeventySix_(?!:)|_FourHundredSeventySeven_(?!:)|_FourHundredSeventyEight_(?!:)|_FourHundredSeventyNine_(?!:)|_FourHundredEighty_(?!:)|_FourHundredEightyOne_(?!:)|_FourHundredEightyTwo_(?!:)|_FourHundredEightyThree_(?!:)|_FourHundredEightyFour_(?!:)|_FourHundredEightyFive_(?!:)|_FourHundredEightySix_(?!:)|_FourHundredEightySeven_(?!:)|_FourHundredEightyEight_(?!:)|_FourHundredEightyNine_(?!:)|_FourHundredNinety_(?!:)|_FourHundredNinetyOne_(?!:)|_FourHundredNinetyTwo_(?!:)|_FourHundredNinetyThree_(?!:)|_FourHundredNinetyFour_(?!:)|_FourHundredNinetyFive_(?!:)|_FourHundredNinetySix_(?!:)|_FourHundredNinetySeven_(?!:)|_FourHundredNinetyEight_(?!:)|_FourHundredNinetyNine_(?!:)|_FiveHundred_(?!:)|_FiveHundredOne_(?!:)|_FiveHundredTwo_(?!:)|_FiveHundredThree_(?!:)|_FiveHundredFour_(?!:)|_FiveHundredFive_(?!:)|_FiveHundredSix_(?!:)|_FiveHundredSeven_(?!:)|_FiveHundredEight_(?!:)|_FiveHundredNine_(?!:)|_FiveHundredTen_(?!:)|_FiveHundredEleven_(?!:)|_FiveHundredTwelve_(?!:)|_FiveHundredThirteen_(?!:)|_FiveHundredFourteen_(?!:)|_FiveHundredFifteen_(?!:)|_FiveHundredSixteen_(?!:)|_FiveHundredSeventeen_(?!:)|_FiveHundredEighteen_(?!:)|_FiveHundredNineteen_(?!:)|_FiveHundredTwenty_(?!:)|_FiveHundredTwentyOne_(?!:)|_FiveHundredTwentyTwo_(?!:)|_FiveHundredTwentyThree_(?!:)|_FiveHundredTwentyFour_(?!:)|_FiveHundredTwentyFive_(?!:)|_FiveHundredTwentySix_(?!:)|_FiveHundredTwentySeven_(?!:)|_FiveHundredTwentyEight_(?!:)|_FiveHundredTwentyNine_(?!:)|_FiveHundredThirty_(?!:)|_FiveHundredThirtyOne_(?!:)|_FiveHundredThirtyTwo_(?!:)|_FiveHundredThirtyThree_(?!:)|_FiveHundredThirtyFour_(?!:)|_FiveHundredThirtyFive_(?!:)|_FiveHundredThirtySix_(?!:)|_FiveHundredThirtySeven_(?!:)|_FiveHundredThirtyEight_(?!:)|_FiveHundredThirtyNine_(?!:)|_FiveHundredForty_(?!:)|_FiveHundredFortyOne_(?!:)|_FiveHundredFortyTwo_(?!:)|_FiveHundredFortyThree_(?!:)|_FiveHundredFortyFour_(?!:)|_FiveHundredFortyFive_(?!:)|_FiveHundredFortySix_(?!:)|_FiveHundredFortySeven_(?!:)|_FiveHundredFortyEight_(?!:)|_FiveHundredFortyNine_(?!:)|_FiveHundredFifty_(?!:)|_FiveHundredFiftyOne_(?!:)|_FiveHundredFiftyTwo_(?!:)|_FiveHundredFiftyThree_(?!:)|_FiveHundredFiftyFour_(?!:)|_FiveHundredFiftyFive_(?!:)|_FiveHundredFiftySix_(?!:)|_FiveHundredFiftySeven_(?!:)|_FiveHundredFiftyEight_(?!:)|_FiveHundredFiftyNine_(?!:)|_FiveHundredSixty_(?!:)|_FiveHundredSixtyOne_(?!:)|_FiveHundredSixtyTwo_(?!:)|_FiveHundredSixtyThree_(?!:)|_FiveHundredSixtyFour_(?!:)|_FiveHundredSixtyFive_(?!:)|_FiveHundredSixtySix_(?!:)|_FiveHundredSixtySeven_(?!:)|_FiveHundredSixtyEight_(?!:)|_FiveHundredSixtyNine_(?!:)|_FiveHundredSeventy_(?!:)|_FiveHundredSeventyOne_(?!:)|_FiveHundredSeventyTwo_(?!:)|_FiveHundredSeventyThree_(?!:)|_FiveHundredSeventyFour_(?!:)|_FiveHundredSeventyFive_(?!:)|_FiveHundredSeventySix_(?!:)|_FiveHundredSeventySeven_(?!:)|_FiveHundredSeventyEight_(?!:)|_FiveHundredSeventyNine_(?!:)|_FiveHundredEighty_(?!:)|_FiveHundredEightyOne_(?!:)|_FiveHundredEightyTwo_(?!:)|_FiveHundredEightyThree_(?!:)|_FiveHundredEightyFour_(?!:)|_FiveHundredEightyFive_(?!:)|_FiveHundredEightySix_(?!:)|_FiveHundredEightySeven_(?!:)|_FiveHundredEightyEight_(?!:)|_FiveHundredEightyNine_(?!:)|_FiveHundredNinety_(?!:)|_FiveHundredNinetyOne_(?!:)|_FiveHundredNinetyTwo_(?!:)|_FiveHundredNinetyThree_(?!:)|_FiveHundredNinetyFour_(?!:)|_FiveHundredNinetyFive_(?!:)|_FiveHundredNinetySix_(?!:)|_FiveHundredNinetySeven_(?!:)|_FiveHundredNinetyEight_(?!:)|_FiveHundredNinetyNine_(?!:)|_SixHundred_(?!:)|_SixHundredOne_(?!:)|_SixHundredTwo_(?!:)|_SixHundredThree_(?!:)|_SixHundredFour_(?!:)|_SixHundredFive_(?!:)|_SixHundredSix_(?!:)|_SixHundredSeven_(?!:)|_SixHundredEight_(?!:)|_SixHundredNine_(?!:)|_SixHundredTen_(?!:)|_SixHundredEleven_(?!:)|_SixHundredTwelve_(?!:)|_SixHundredThirteen_(?!:)|_SixHundredFourteen_(?!:)|_SixHundredFifteen_(?!:)|_SixHundredSixteen_(?!:)|_SixHundredSeventeen_(?!:)|_SixHundredEighteen_(?!:)|_SixHundredNineteen_(?!:)|_SixHundredTwenty_(?!:)|_SixHundredTwentyOne_(?!:)|_SixHundredTwentyTwo_(?!:)|_SixHundredTwentyThree_(?!:)|_SixHundredTwentyFour_(?!:)|_SixHundredTwentyFive_(?!:)|_SixHundredTwentySix_(?!:)|_SixHundredTwentySeven_(?!:)|_SixHundredTwentyEight_(?!:)|_SixHundredTwentyNine_(?!:)|_SixHundredThirty_(?!:)|_SixHundredThirtyOne_(?!:)|_SixHundredThirtyTwo_(?!:)|_SixHundredThirtyThree_(?!:)|_SixHundredThirtyFour_(?!:)|_SixHundredThirtyFive_(?!:)|_SixHundredThirtySix_(?!:)|_SixHundredThirtySeven_(?!:)|_SixHundredThirtyEight_(?!:)|_SixHundredThirtyNine_(?!:)|_SixHundredForty_(?!:)|_SixHundredFortyOne_(?!:)|_SixHundredFortyTwo_(?!:)|_SixHundredFortyThree_(?!:)|_SixHundredFortyFour_(?!:)|_SixHundredFortyFive_(?!:)|_SixHundredFortySix_(?!:)|_SixHundredFortySeven_(?!:)|_SixHundredFortyEight_(?!:)|_SixHundredFortyNine_(?!:)|_SixHundredFifty_(?!:)|_SixHundredFiftyOne_(?!:)|_SixHundredFiftyTwo_(?!:)|_SixHundredFiftyThree_(?!:)|_SixHundredFiftyFour_(?!:)|_SixHundredFiftyFive_(?!:)|_SixHundredFiftySix_(?!:)|_SixHundredFiftySeven_(?!:)|_SixHundredFiftyEight_(?!:)|_SixHundredFiftyNine_(?!:)|_SixHundredSixty_(?!:)|_SixHundredSixtyOne_(?!:)|_SixHundredSixtyTwo_(?!:)|_SixHundredSixtyThree_(?!:)|_SixHundredSixtyFour_(?!:)|_SixHundredSixtyFive_(?!:)|_SixHundredSixtySix_(?!:)|_SixHundredSixtySeven_(?!:)|_SixHundredSixtyEight_(?!:)|_SixHundredSixtyNine_(?!:)|_SixHundredSeventy_(?!:)|_SixHundredSeventyOne_(?!:)|_SixHundredSeventyTwo_(?!:)|_SixHundredSeventyThree_(?!:)|_SixHundredSeventyFour_(?!:)|_SixHundredSeventyFive_(?!:)|_SixHundredSeventySix_(?!:)|_SixHundredSeventySeven_(?!:)|_SixHundredSeventyEight_(?!:)|_SixHundredSeventyNine_(?!:)|_SixHundredEighty_(?!:)|_SixHundredEightyOne_(?!:)|_SixHundredEightyTwo_(?!:)|_SixHundredEightyThree_(?!:)|_SixHundredEightyFour_(?!:)|_SixHundredEightyFive_(?!:)|_SixHundredEightySix_(?!:)|_SixHundredEightySeven_(?!:)|_SixHundredEightyEight_(?!:)|_SixHundredEightyNine_(?!:)|_SixHundredNinety_(?!:)|_SixHundredNinetyOne_(?!:)|_SixHundredNinetyTwo_(?!:)|_SixHundredNinetyThree_(?!:)|_SixHundredNinetyFour_(?!:)|_SixHundredNinetyFive_(?!:)|_SixHundredNinetySix_(?!:)|_SixHundredNinetySeven_(?!:)|_SixHundredNinetyEight_(?!:)|_SixHundredNinetyNine_(?!:)|_SevenHundred_(?!:)|_SevenHundredOne_(?!:)|_SevenHundredTwo_(?!:)|_SevenHundredThree_(?!:)|_SevenHundredFour_(?!:)|_SevenHundredFive_(?!:)|_SevenHundredSix_(?!:)|_SevenHundredSeven_(?!:)|_SevenHundredEight_(?!:)|_SevenHundredNine_(?!:)|_SevenHundredTen_(?!:)|_SevenHundredEleven_(?!:)|_SevenHundredTwelve_(?!:)|_SevenHundredThirteen_(?!:)|_SevenHundredFourteen_(?!:)|_SevenHundredFifteen_(?!:)|_SevenHundredSixteen_(?!:)|_SevenHundredSeventeen_(?!:)|_SevenHundredEighteen_(?!:)|_SevenHundredNineteen_(?!:)|_SevenHundredTwenty_(?!:)|_SevenHundredTwentyOne_(?!:)|_SevenHundredTwentyTwo_(?!:)|_SevenHundredTwentyThree_(?!:)|_SevenHundredTwentyFour_(?!:)|_SevenHundredTwentyFive_(?!:)|_SevenHundredTwentySix_(?!:)|_SevenHundredTwentySeven_(?!:)|_SevenHundredTwentyEight_(?!:)|_SevenHundredTwentyNine_(?!:)|_SevenHundredThirty_(?!:)|_SevenHundredThirtyOne_(?!:)|_SevenHundredThirtyTwo_(?!:)|_SevenHundredThirtyThree_(?!:)|_SevenHundredThirtyFour_(?!:)|_SevenHundredThirtyFive_(?!:)|_SevenHundredThirtySix_(?!:)|_SevenHundredThirtySeven_(?!:)|_SevenHundredThirtyEight_(?!:)|_SevenHundredThirtyNine_(?!:)|_SevenHundredForty_(?!:)|_SevenHundredFortyOne_(?!:)|_SevenHundredFortyTwo_(?!:)|_SevenHundredFortyThree_(?!:)|_SevenHundredFortyFour_(?!:)|_SevenHundredFortyFive_(?!:)|_SevenHundredFortySix_(?!:)|_SevenHundredFortySeven_(?!:)|_SevenHundredFortyEight_(?!:)|_SevenHundredFortyNine_(?!:)|_SevenHundredFifty_(?!:)|_SevenHundredFiftyOne_(?!:)|_SevenHundredFiftyTwo_(?!:)|_SevenHundredFiftyThree_(?!:)|_SevenHundredFiftyFour_(?!:)|_SevenHundredFiftyFive_(?!:)|_SevenHundredFiftySix_(?!:)|_SevenHundredFiftySeven_(?!:)|_SevenHundredFiftyEight_(?!:)|_SevenHundredFiftyNine_(?!:)|_SevenHundredSixty_(?!:)|_SevenHundredSixtyOne_(?!:)|_SevenHundredSixtyTwo_(?!:)|_SevenHundredSixtyThree_(?!:)|_SevenHundredSixtyFour_(?!:)|_SevenHundredSixtyFive_(?!:)|_SevenHundredSixtySix_(?!:)|_SevenHundredSixtySeven_(?!:)|_SevenHundredSixtyEight_(?!:)|_SevenHundredSixtyNine_(?!:)|_SevenHundredSeventy_(?!:)|_SevenHundredSeventyOne_(?!:)|_SevenHundredSeventyTwo_(?!:)|_SevenHundredSeventyThree_(?!:)|_SevenHundredSeventyFour_(?!:)|_SevenHundredSeventyFive_(?!:)|_SevenHundredSeventySix_(?!:)|_SevenHundredSeventySeven_(?!:)|_SevenHundredSeventyEight_(?!:)|_SevenHundredSeventyNine_(?!:)|_SevenHundredEighty_(?!:)|_SevenHundredEightyOne_(?!:)|_SevenHundredEightyTwo_(?!:)|_SevenHundredEightyThree_(?!:)|_SevenHundredEightyFour_(?!:)|_SevenHundredEightyFive_(?!:)|_SevenHundredEightySix_(?!:)|_SevenHundredEightySeven_(?!:)|_SevenHundredEightyEight_(?!:)|_SevenHundredEightyNine_(?!:)|_SevenHundredNinety_(?!:)|_SevenHundredNinetyOne_(?!:)|_SevenHundredNinetyTwo_(?!:)|_SevenHundredNinetyThree_(?!:)|_SevenHundredNinetyFour_(?!:)|_SevenHundredNinetyFive_(?!:)|_SevenHundredNinetySix_(?!:)|_SevenHundredNinetySeven_(?!:)|_SevenHundredNinetyEight_(?!:)|_SevenHundredNinetyNine_(?!:)|_EightHundred_(?!:)|_EightHundredOne_(?!:)|_EightHundredTwo_(?!:)|_EightHundredThree_(?!:)|_EightHundredFour_(?!:)|_EightHundredFive_(?!:)|_EightHundredSix_(?!:)|_EightHundredSeven_(?!:)|_EightHundredEight_(?!:)|_EightHundredNine_(?!:)|_EightHundredTen_(?!:)|_EightHundredEleven_(?!:)|_EightHundredTwelve_(?!:)|_EightHundredThirteen_(?!:)|_EightHundredFourteen_(?!:)|_EightHundredFifteen_(?!:)|_EightHundredSixteen_(?!:)|_EightHundredSeventeen_(?!:)|_EightHundredEighteen_(?!:)|_EightHundredNineteen_(?!:)|_EightHundredTwenty_(?!:)|_EightHundredTwentyOne_(?!:)|_EightHundredTwentyTwo_(?!:)|_EightHundredTwentyThree_(?!:)|_EightHundredTwentyFour_(?!:)|_EightHundredTwentyFive_(?!:)|_EightHundredTwentySix_(?!:)|_EightHundredTwentySeven_(?!:)|_EightHundredTwentyEight_(?!:)|_EightHundredTwentyNine_(?!:)|_EightHundredThirty_(?!:)|_EightHundredThirtyOne_(?!:)|_EightHundredThirtyTwo_(?!:)|_EightHundredThirtyThree_(?!:)|_EightHundredThirtyFour_(?!:)|_EightHundredThirtyFive_(?!:)|_EightHundredThirtySix_(?!:)|_EightHundredThirtySeven_(?!:)|_EightHundredThirtyEight_(?!:)|_EightHundredThirtyNine_(?!:)|_EightHundredForty_(?!:)|_EightHundredFortyOne_(?!:)|_EightHundredFortyTwo_(?!:)|_EightHundredFortyThree_(?!:)|_EightHundredFortyFour_(?!:)|_EightHundredFortyFive_(?!:)|_EightHundredFortySix_(?!:)|_EightHundredFortySeven_(?!:)|_EightHundredFortyEight_(?!:)|_EightHundredFortyNine_(?!:)|_EightHundredFifty_(?!:)|_EightHundredFiftyOne_(?!:)|_EightHundredFiftyTwo_(?!:)|_EightHundredFiftyThree_(?!:)|_EightHundredFiftyFour_(?!:)|_EightHundredFiftyFive_(?!:)|_EightHundredFiftySix_(?!:)|_EightHundredFiftySeven_(?!:)|_EightHundredFiftyEight_(?!:)|_EightHundredFiftyNine_(?!:)|_EightHundredSixty_(?!:)|_EightHundredSixtyOne_(?!:)|_EightHundredSixtyTwo_(?!:)|_EightHundredSixtyThree_(?!:)|_EightHundredSixtyFour_(?!:)|_EightHundredSixtyFive_(?!:)|_EightHundredSixtySix_(?!:)|_EightHundredSixtySeven_(?!:)|_EightHundredSixtyEight_(?!:)|_EightHundredSixtyNine_(?!:)|_EightHundredSeventy_(?!:)|_EightHundredSeventyOne_(?!:)|_EightHundredSeventyTwo_(?!:)|_EightHundredSeventyThree_(?!:)|_EightHundredSeventyFour_(?!:)|_EightHundredSeventyFive_(?!:)|_EightHundredSeventySix_(?!:)|_EightHundredSeventySeven_(?!:)|_EightHundredSeventyEight_(?!:)|_EightHundredSeventyNine_(?!:)|_EightHundredEighty_(?!:)|_EightHundredEightyOne_(?!:)|_EightHundredEightyTwo_(?!:)|_EightHundredEightyThree_(?!:)|_EightHundredEightyFour_(?!:)|_EightHundredEightyFive_(?!:)|_EightHundredEightySix_(?!:)|_EightHundredEightySeven_(?!:)|_EightHundredEightyEight_(?!:)|_EightHundredEightyNine_(?!:)|_EightHundredNinety_(?!:)|_EightHundredNinetyOne_(?!:)|_EightHundredNinetyTwo_(?!:)|_EightHundredNinetyThree_(?!:)|_EightHundredNinetyFour_(?!:)|_EightHundredNinetyFive_(?!:)|_EightHundredNinetySix_(?!:)|_EightHundredNinetySeven_(?!:)|_EightHundredNinetyEight_(?!:)|_EightHundredNinetyNine_(?!:)|_NineHundred_(?!:)|_NineHundredOne_(?!:)|_NineHundredTwo_(?!:)|_NineHundredThree_(?!:)|_NineHundredFour_(?!:)|_NineHundredFive_(?!:)|_NineHundredSix_(?!:)|_NineHundredSeven_(?!:)|_NineHundredEight_(?!:)|_NineHundredNine_(?!:)|_NineHundredTen_(?!:)|_NineHundredEleven_(?!:)|_NineHundredTwelve_(?!:)|_NineHundredThirteen_(?!:)|_NineHundredFourteen_(?!:)|_NineHundredFifteen_(?!:)|_NineHundredSixteen_(?!:)|_NineHundredSeventeen_(?!:)|_NineHundredEighteen_(?!:)|_NineHundredNineteen_(?!:)|_NineHundredTwenty_(?!:)|_NineHundredTwentyOne_(?!:)|_NineHundredTwentyTwo_(?!:)|_NineHundredTwentyThree_(?!:)|_NineHundredTwentyFour_(?!:)|_NineHundredTwentyFive_(?!:)|_NineHundredTwentySix_(?!:)|_NineHundredTwentySeven_(?!:)|_NineHundredTwentyEight_(?!:)|_NineHundredTwentyNine_(?!:)|_NineHundredThirty_(?!:)|_NineHundredThirtyOne_(?!:)|_NineHundredThirtyTwo_(?!:)|_NineHundredThirtyThree_(?!:)|_NineHundredThirtyFour_(?!:)|_NineHundredThirtyFive_(?!:)|_NineHundredThirtySix_(?!:)|_NineHundredThirtySeven_(?!:)|_NineHundredThirtyEight_(?!:)|_NineHundredThirtyNine_(?!:)|_NineHundredForty_(?!:)|_NineHundredFortyOne_(?!:)|_NineHundredFortyTwo_(?!:)|_NineHundredFortyThree_(?!:)|_NineHundredFortyFour_(?!:)|_NineHundredFortyFive_(?!:)|_NineHundredFortySix_(?!:)|_NineHundredFortySeven_(?!:)|_NineHundredFortyEight_(?!:)|_NineHundredFortyNine_(?!:)|_NineHundredFifty_(?!:)|_NineHundredFiftyOne_(?!:)|_NineHundredFiftyTwo_(?!:)|_NineHundredFiftyThree_(?!:)|_NineHundredFiftyFour_(?!:)|_NineHundredFiftyFive_(?!:)|_NineHundredFiftySix_(?!:)|_NineHundredFiftySeven_(?!:)|_NineHundredFiftyEight_(?!:)|_NineHundredFiftyNine_(?!:)|_NineHundredSixty_(?!:)|_NineHundredSixtyOne_(?!:)|_NineHundredSixtyTwo_(?!:)|_NineHundredSixtyThree_(?!:)|_NineHundredSixtyFour_(?!:)|_NineHundredSixtyFive_(?!:)|_NineHundredSixtySix_(?!:)|_NineHundredSixtySeven_(?!:)|_NineHundredSixtyEight_(?!:)|_NineHundredSixtyNine_(?!:)|_NineHundredSeventy_(?!:)|_NineHundredSeventyOne_(?!:)|_NineHundredSeventyTwo_(?!:)|_NineHundredSeventyThree_(?!:)|_NineHundredSeventyFour_(?!:)|_NineHundredSeventyFive_(?!:)|_NineHundredSeventySix_(?!:)|_NineHundredSeventySeven_(?!:)|_NineHundredSeventyEight_(?!:)|_NineHundredSeventyNine_(?!:)|_NineHundredEighty_(?!:)|_NineHundredEightyOne_(?!:)|_NineHundredEightyTwo_(?!:)|_NineHundredEightyThree_(?!:)|_NineHundredEightyFour_(?!:)|_NineHundredEightyFive_(?!:)|_NineHundredEightySix_(?!:)|_NineHundredEightySeven_(?!:)|_NineHundredEightyEight_(?!:)|_NineHundredEightyNine_(?!:)|_NineHundredNinety_(?!:)|_NineHundredNinetyOne_(?!:)|_NineHundredNinetyTwo_(?!:)|_NineHundredNinetyThree_(?!:)|_NineHundredNinetyFour_(?!:)|_NineHundredNinetyFive_(?!:)|_NineHundredNinetySix_(?!:)|_NineHundredNinetySeven_(?!:)|_NineHundredNinetyEight_(?!:)|_NineHundredNinetyNine_(?!:)|_TenHundred_(?!:)|_TenHundredOne_(?!:)|_TenHundredTwo_(?!:)|_TenHundredThree_(?!:)|_TenHundredFour_(?!:)|_TenHundredFive_(?!:)|_TenHundredSix_(?!:)|_TenHundredSeven_(?!:)|_TenHundredEight_(?!:)|_TenHundredNine_(?!:)|_TenHundredTen_(?!:)|_TenHundredEleven_(?!:)|_TenHundredTwelve_(?!:)|_TenHundredThirteen_(?!:)|_TenHundredFourteen_(?!:)|_TenHundredFifteen_(?!:)|_TenHundredSixteen_(?!:)|_TenHundredSeventeen_(?!:)|_TenHundredEighteen_(?!:)|_TenHundredNineteen_(?!:)|_TenHundredTwenty_(?!:)|_TenHundredTwentyOne_(?!:)|_TenHundredTwentyTwo_(?!:)|_TenHundredTwentyThree_(?!:)|_TenHundredTwentyFour_(?!:)|_TenHundredTwentyFive_(?!:)|_TenHundredTwentySix_(?!:)|_TenHundredTwentySeven_(?!:)|_TenHundredTwentyEight_(?!:)|_TenHundredTwentyNine_(?!:)|_TenHundredThirty_(?!:)|_TenHundredThirtyOne_(?!:)|_TenHundredThirtyTwo_(?!:)|_TenHundredThirtyThree_(?!:)|_TenHundredThirtyFour_(?!:)|_TenHundredThirtyFive_(?!:)|_TenHundredThirtySix_(?!:)|_TenHundredThirtySeven_(?!:)|_TenHundredThirtyEight_(?!:)|_TenHundredThirtyNine_(?!:)|_TenHundredForty_(?!:)|_TenHundredFortyOne_(?!:)|_TenHundredFortyTwo_(?!:)|_TenHundredFortyThree_(?!:)|_TenHundredFortyFour_(?!:)|_TenHundredFortyFive_(?!:)|_TenHundredFortySix_(?!:)|_TenHundredFortySeven_(?!:)|_TenHundredFortyEight_(?!:)|_TenHundredFortyNine_(?!:)|_TenHundredFifty_(?!:)|_TenHundredFiftyOne_(?!:)|_TenHundredFiftyTwo_(?!:)|_TenHundredFiftyThree_(?!:)|_TenHundredFiftyFour_(?!:)|_TenHundredFiftyFive_(?!:)|_TenHundredFiftySix_(?!:)|_TenHundredFiftySeven_(?!:)|_TenHundredFiftyEight_(?!:)|_TenHundredFiftyNine_(?!:)|_TenHundredSixty_(?!:)|_TenHundredSixtyOne_(?!:)|_TenHundredSixtyTwo_(?!:)|_TenHundredSixtyThree_(?!:)|_TenHundredSixtyFour_(?!:)|_TenHundredSixtyFive_(?!:)|_TenHundredSixtySix_(?!:)|_TenHundredSixtySeven_(?!:)|_TenHundredSixtyEight_(?!:)|_TenHundredSixtyNine_(?!:)|_TenHundredSeventy_(?!:)|_TenHundredSeventyOne_(?!:)|_TenHundredSeventyTwo_(?!:)|_TenHundredSeventyThree_(?!:)|_TenHundredSeventyFour_(?!:)|_TenHundredSeventyFive_(?!:)|_TenHundredSeventySix_(?!:)|_TenHundredSeventySeven_(?!:)|_TenHundredSeventyEight_(?!:)|_TenHundredSeventyNine_(?!:)|_TenHundredEighty_(?!:)|_TenHundredEightyOne_(?!:)|_TenHundredEightyTwo_(?!:)|_TenHundredEightyThree_(?!:)|_TenHundredEightyFour_(?!:)|_TenHundredEightyFive_(?!:)|_TenHundredEightySix_(?!:)|_TenHundredEightySeven_(?!:)|_TenHundredEightyEight_(?!:)|_TenHundredEightyNine_(?!:)|_TenHundredNinety_(?!:)|_TenHundredNinetyOne_(?!:)|_TenHundredNinetyTwo_(?!:)|_TenHundredNinetyThree_(?!:)|_TenHundredNinetyFour_(?!:)|_TenHundredNinetyFive_(?!:)|_TenHundredNinetySix_(?!:)|
```

```

# chapter2 = cumsum(str_detect(text,
#                               "^_(Ni|O|T|S|F|E)[:lower:]+[-]*[:lower:]*_$")),
# chapter=chapter2+chapter1,
# chapter1=NULL,
# chapter2=NULL) %>%
#
#
tidy_df <- df %>%
  # add line and chapter numbers to dataset
  mutate(linenumber = row_number(),
         chapter = cumsum(str_detect(text, c_pattern))) %>%
  # convert all to lower case and put one line per word
  unnest_tokens(word, text) %>%
  # take care of any special formatting characters around words
  mutate(word = str_extract(word, "[a-z']+")) %>%
  # remove stop-words
  anti_join(stop_words) %>%
  # get rid of any NA's
  filter(!is.na(word))
}

```