

# THagen\_Mod5\_Assignment1

Tabitha Hagen

4/12/2023

## 1 - Load the libraries

## 2 - Download and prep the data

Data set: gutenbergr. The gutenbergr package helps you download and process public domain works from the Project Gutenberg collection.

You can search the Project Gutenberg using the title of the book and then download the text using the gutenbergr package:

*#If you know the book's id in the Project Gutenberg, you can use that id to download the text :*

```
pride_raw<-gutenbergr_download(1342)
```

*#We can create a new column chapter that keeps track of which of the chapters each line of text comes from. We can detect in that text column when it says "Chapter".*

```
pride<-pride_raw %>%  
  mutate(chapter=ifelse(str_detect(text,"Chapter"),text,NA))%>%  
  fill(chapter)%>% #fill down the new column "chapter" until it gets to a new value
```

```
  filter(chapter != "NA") %>% #remove the NA that comes at the end of the last chapter.
```

```
  #set the factor levels by the order you find them in the text  
  mutate(chapter = factor(chapter, levels = unique(chapter)))
```

```
pride %>%
```

```
  count(chapter) #now the stories are in order and n shows how many lines we have in each story
```

```
## # A tibble: 30 × 2
```

```
##   chapter
```

```
n
```

```
##   <fct>
```

```
<int>
```

```
## 1 Heading to Chapter I.
```

```
...
```

```
16
```

```
## 2 Heading to Chapter IV.
```

```
...
```

```
2
```

```
## 3 Heading to Chapter V.
```

```
...
```

```

4
## 4 Tailpiece to Chapter V. ...
2
## 5 Heading to Chapter VI. ...
14
## 6 Heading to Chapter X. ...
6
## 7 Heading to Chapter XII. ...
2
## 8 Heading to Chapter XIII. ...
2
## 9 Heading to Chapter XIV. ...
4
## 10 Heading to Chapter XV. ...
2
## # ... with 20 more rows

```

## Tidy and pre-process text

Transform the text data set to the tidy text dataframe, filter out the stop words, and count the most common words.

### Remove the outlier words (the Names) from the dataset

*#Let's remove the specific repeated names from our dataset.*

```

Names <- c("elizabeth", "bennet", "miss", "lady", "dear", "darcy",
"jane","netherfield",
"bingley", "catherine", "wickham", "kitty", "lydia", "collins", "gardiner",
"charlotte", "lucas", "lizzy", "colonel", "forster", "hurst", "sir",
"hertfordshire",
"william", "eliza", "darcy's", "reynaolds", "fitzwilliam", "longbourn",
"pemberly", "lydia's") # List of names to remove

tidy_pride <- pride %>%
  mutate(line = row_number()) %>% #create a new column to track which line
every word is coming from
  unnest_tokens(word, text) %>% #unnest the text column into the word
column
  anti_join(all_stop_words, by = "word")%>% #filter out the stop words
  filter(!word %in% Names) #filter out the list of names from our list
created above

#tidy_pride

tidy_pride%>%
  count(word,sort = TRUE) #count the words and identify the most common
words

```

```
## # A tibble: 6,063 × 2
##   word      n
##   <chr> <int>
## 1 time    197
## 2 sister  179
## 3 family  151
## 4 day     141
## 5 hope    123
## 6 father  116
## 7 letter  116
## 8 mother  112
## 9 house   107
## 10 friend 106
## # ... with 6,053 more rows
```

### 3 - Implement Topic Modeling (LDA)

#### Cast the tidied dataset to a dtm:

```
pride_dtm<-tidy_pride %>%
  count(chapter, word) %>%
  cast_dtm(chapter, word, n) # create a document-term matrix
```

Now we are ready to use the topicmodels package to create a topic LDA model.

The function LDA() in the topicmodels package performs Latent Dirichlet Allocation Note: it may take a while for the code below to run, since the default setting of the LDA package is to perform a large number of iterations.

The user must specify a value of k, or the number of topics in the corpus, in order to run a topic model (we may need to try a few different values of k).

The control argument is used to pass a random number (1234) to seed the assignment of topics to each word in the corpus. the control argument can be used to specify a number of different options as well, such as the maximum number of iterations that we want our topic model to perform.

#### a - We train a LDA topic model with 6 topics:

```
topic_model<-LDA(pride_dtm, k=6, control = list(seed = 1234)) #Create LDA
topic model
topic_model
```

```
## A LDA_VEM topic model with 6 topics.
```

*#Note: it may take a while for the code below to run, since the default setting of the LDA package is to perform a large number of iterations.*

**\*\* The parameter k is the number of topics that we expect to find in this corpus. We can start by a completely arbitrary random assumption that there are 10 topics in this corpus of news articles, there might be 2 topics, 3, 700! \*\***

**\*\* The control argument is used to pass a random number (321) to seed the assignment of topics to each word in the corpus. the control argument can be used to specify a number of different options as well, such as the maximum number of iterations that we want our topic model to perform. \*\***

**Posterior() determines the posterior probabilities of the topics for each document and of the terms for each topic for a fitted topic model.**

*# Look at the posterior probabilities of the topics for each document and of the terms for each topic for a fitted topic model.*

```
posterior(topic_model)$topics
```

```
##
```

```
1
```

```
## Heading to Chapter I. 1
```

```
0.009357369
```

```
## Heading to Chapter IV. 18
```

```
0.038184379
```

```
## Heading to Chapter V. 22
```

```
0.030381781
```

```
## Tailpiece to Chapter V. 26
```

```
0.051389712
```

```
## Heading to Chapter VI. 27
```

```
0.008798913
```

```
## Heading to Chapter X. 60
```

```
0.018835231
```

```
## Heading to Chapter XII. 75
```

```
0.038184377
```

```
## Heading to Chapter XIII. 78
```

```
0.038184379
```

```
## Heading to Chapter XIV. 84
```

```
0.018833244
```

```
## Heading to Chapter XV. 89
```

```
0.038184380
```

```
## Heading to Chapter XVI. 95
```

```
0.016736871
```

```
## Heading to Chapter XVIII. 113
```

```
0.015030840
```

```
## Heading to Chapter XX. 139
```

```
0.025226825
```

```
## Heading to Chapter XXI. 146
```

```
0.025223549
```

```
## Heading to Chapter XXII. 154
```

```
0.011537295
```

```
## Heading to Chapter XXIV. 168
```

```
0.038184379
```

## Heading to Chapter XXV. 0.009358722	175
## Heading to Chapter XXX. 0.016718664	209
## Heading to Chapter XXXI. 0.038184376	215
## Heading to Chapter XXXII. 0.018834090	221
## Heading to Chapter XXXIV. 0.025247177	235
## Heading to Chapter XXXVI. 0.007118839	253
## Heading to Chapter XL. 0.008309558	278
## Heading to Chapter XLIV. 0.018840585	318
## Heading to Chapter XLVI. 0.010703677	334
## Heading to Chapter XLIX. 0.007117250	368
## Heading to Chapter LVI. 0.008800382	431
## Heading to Chapter LXI. 0.021567849	472
## Chapter I.] 0.026652333	
## Chapter XLVI. 0.170077246	
## 2	
## Heading to Chapter I. 0.0093576969	1
## Heading to Chapter IV. 0.0381851406	18
## Heading to Chapter V. 0.0303835034	22
## Tailpiece to Chapter V. 0.0513897122	26
## Heading to Chapter VI. 0.0087994726	27
## Heading to Chapter X. 0.0188347294	60
## Heading to Chapter XII. 0.0381855188	75
## Heading to Chapter XIII. 0.0381846290	78
## Heading to Chapter XIV. 0.0188325102	84
## Heading to Chapter XV. 0.0381853525	89

## Heading to Chapter XVI. 0.0167408022	95
## Heading to Chapter XVIII. 0.0150324405	113
## Heading to Chapter XX. 0.0252285288	139
## Heading to Chapter XXI. 0.0252353505	146
## Heading to Chapter XXII. 0.0115334171	154
## Heading to Chapter XXIV. 0.0381852256	168
## Heading to Chapter XXV. 0.0093540002	175
## Heading to Chapter XXX. 0.0167154777	209
## Heading to Chapter XXXI. 0.0381854130	215
## Heading to Chapter XXXII. 0.0188314108	221
## Heading to Chapter XXXIV. 0.0252419498	235
## Heading to Chapter XXXVI. 0.0071156305	253
## Heading to Chapter XL. 0.0083076838	278
## Heading to Chapter XLIV. 0.0188321934	318
## Heading to Chapter XLVI. 0.0106998560	334
## Heading to Chapter XLIX. 0.0071147799	368
## Heading to Chapter LVI. 0.0088001616	431
## Heading to Chapter LXI. 0.0215771983	472
## Chapter I.] 0.4071344547	
## Chapter XLVI. 0.0000150249	
## 3	
## Heading to Chapter I. 9.356689e-03	1
## Heading to Chapter IV. 3.818438e-02	18
## Heading to Chapter V. 3.038164e-02	22
## Tailpiece to Chapter V. 5.138971e-02	26

## Heading to Chapter VI. 8.797907e-03	27
## Heading to Chapter X. 1.883539e-02	60
## Heading to Chapter XII. 3.818438e-02	75
## Heading to Chapter XIII. 3.818438e-02	78
## Heading to Chapter XIV. 1.883250e-02	84
## Heading to Chapter XV. 3.818438e-02	89
## Heading to Chapter XVI. 1.671950e-02	95
## Heading to Chapter XVIII. 1.502500e-02	113
## Heading to Chapter XX. 2.523184e-02	139
## Heading to Chapter XXI. 2.523162e-02	146
## Heading to Chapter XXII. 1.153321e-02	154
## Heading to Chapter XXIV. 3.818438e-02	168
## Heading to Chapter XXV. 9.353113e-03	175
## Heading to Chapter XXX. 1.671448e-02	209
## Heading to Chapter XXXI. 3.818438e-02	215
## Heading to Chapter XXXII. 1.883656e-02	221
## Heading to Chapter XXXIV. 2.523221e-02	235
## Heading to Chapter XXXVI. 7.116145e-03	253
## Heading to Chapter XL. 8.308799e-03	278
## Heading to Chapter XLIV. 1.883477e-02	318
## Heading to Chapter XLVI. 1.070120e-02	334
## Heading to Chapter XLIX. 7.115864e-03	368
## Heading to Chapter LVI. 8.802272e-03	431
## Heading to Chapter LXI. 2.156674e-02	472
## Chapter I.] 6.373554e-06	

## Chapter XLVI.	
8.298627e-01	
##	
4	
## Heading to Chapter I.	1
9.356797e-03	
## Heading to Chapter IV.	18
3.818587e-02	
## Heading to Chapter V.	22
3.038070e-02	
## Tailpiece to Chapter V.	26
5.138971e-02	
## Heading to Chapter VI.	27
8.799633e-03	
## Heading to Chapter X.	60
1.883563e-02	
## Heading to Chapter XII.	75
3.818535e-02	
## Heading to Chapter XIII.	78
3.818651e-02	
## Heading to Chapter XIV.	84
1.883232e-02	
## Heading to Chapter XV.	89
3.818544e-02	
## Heading to Chapter XVI.	95
1.672170e-02	
## Heading to Chapter XVIII.	113
1.503024e-02	
## Heading to Chapter XX.	139
2.523645e-02	
## Heading to Chapter XXI.	146
2.522981e-02	
## Heading to Chapter XXII.	154
1.153625e-02	
## Heading to Chapter XXIV.	168
3.818553e-02	
## Heading to Chapter XXV.	175
9.357173e-03	
## Heading to Chapter XXX.	209
1.672190e-02	
## Heading to Chapter XXXI.	215
3.818528e-02	
## Heading to Chapter XXXII.	221
1.883602e-02	
## Heading to Chapter XXXIV.	235
2.523895e-02	
## Heading to Chapter XXXVI.	253
7.116576e-03	
## Heading to Chapter XL.	278
8.307855e-03	



## Heading to Chapter XLIV. 1.885416e-02	318
## Heading to Chapter XLVI. 1.069964e-02	334
## Heading to Chapter XLIX. 7.115420e-03	368
## Heading to Chapter LVI. 8.799736e-03	431
## Heading to Chapter LXI. 2.156669e-02	472
## Chapter I.] 3.366002e-01	
## Chapter XLVI. 1.502798e-05	
## 5	
## Heading to Chapter I. 9.354377e-03	1
## Heading to Chapter IV. 3.818446e-02	18
## Heading to Chapter V. 3.037965e-02	22
## Tailpiece to Chapter V. 5.138971e-02	26
## Heading to Chapter VI. 8.798752e-03	27
## Heading to Chapter X. 1.883604e-02	60
## Heading to Chapter XII. 3.818455e-02	75
## Heading to Chapter XIII. 3.818442e-02	78
## Heading to Chapter XIV. 1.883747e-02	84
## Heading to Chapter XV. 3.818471e-02	89
## Heading to Chapter XVI. 1.672302e-02	95
## Heading to Chapter XVIII. 1.502723e-02	113
## Heading to Chapter XX. 2.523556e-02	139
## Heading to Chapter XXI. 2.523497e-02	146
## Heading to Chapter XXII. 1.153609e-02	154
## Heading to Chapter XXIV. 3.818480e-02	168
## Heading to Chapter XXV. 9.356026e-03	175

## Heading to Chapter XXX. 1.671471e-02	209
## Heading to Chapter XXXI. 3.818484e-02	215
## Heading to Chapter XXXII. 1.883857e-02	221
## Heading to Chapter XXXIV. 2.522816e-02	235
## Heading to Chapter XXXVI. 7.119710e-03	253
## Heading to Chapter XL. 8.308730e-03	278
## Heading to Chapter XLIV. 1.884463e-02	318
## Heading to Chapter XLVI. 1.070004e-02	334
## Heading to Chapter XLIX. 7.117111e-03	368
## Heading to Chapter LVI. 8.801912e-03	431
## Heading to Chapter LXI. 2.157108e-02	472
## Chapter I.] 2.296003e-01	
## Chapter XLVI. 1.503863e-05	
## 6	
## Heading to Chapter I. 9.532171e-01	1
## Heading to Chapter IV. 8.090758e-01	18
## Heading to Chapter V. 8.480927e-01	22
## Tailpiece to Chapter V. 7.430514e-01	26
## Heading to Chapter VI. 9.560053e-01	27
## Heading to Chapter X. 9.058230e-01	60
## Heading to Chapter XII. 8.090758e-01	75
## Heading to Chapter XIII. 8.090757e-01	78
## Heading to Chapter XIV. 9.058320e-01	84
## Heading to Chapter XV. 8.090757e-01	89
## Heading to Chapter XVI. 9.163581e-01	95

```

## Heading to Chapter XVIII. 113
9.248542e-01
## Heading to Chapter XX. 139
8.738408e-01
## Heading to Chapter XXI. 146
8.738447e-01
## Heading to Chapter XXII. 154
9.423237e-01
## Heading to Chapter XXIV. 168
8.090757e-01
## Heading to Chapter XXV. 175
9.532210e-01
## Heading to Chapter XXX. 209
9.164148e-01
## Heading to Chapter XXXI. 215
8.090757e-01
## Heading to Chapter XXXII. 221
9.058234e-01
## Heading to Chapter XXXIV. 235
8.738116e-01
## Heading to Chapter XXXVI. 253
9.644131e-01
## Heading to Chapter XL. 278
9.584574e-01
## Heading to Chapter XLIV. 318
9.057937e-01
## Heading to Chapter XLVI. 334
9.464956e-01
## Heading to Chapter XLIX. 368
9.644196e-01
## Heading to Chapter LVI. 431
9.559955e-01
## Heading to Chapter LXI. 472
8.921504e-01
## Chapter I.]
6.358477e-06
## Chapter XLVI.
1.499131e-05

```

## b - Display top 15 words of each topic

*# Look at the top 15 words of each topic tpo determine patterns in Topics*  
terms(topic\_model, k=15)

```

##      Topic 1      Topic 2      Topic 3      Topic 4      Topic 5
## [1,] "time"      "sister"    "letter"   "family"    "time"
## [2,] "sister"    "friend"   "mother"   "day"       "day"
## [3,] "london"    "house"    "sister"   "ladies"    "father"
## [4,] "family"    "evening"  "father"   "added"     "feelings"
## [5,] "affection" "character" "time"     "hope"      "happy"
## [6,] "home"      "continued" "family"   "love"      "sisters"

```

```

## [7,] "hope"      "opinion"      "hope"      "answer"      "life"
## [8,] "speak"     "manner"       "daughter"  "cried"       "letter"
## [9,] "dare"      "passed"       "happy"     "received"     "aunt"
## [10,] "wife"     "party"        "marriage"  "brother"     "mother"
## [11,] "trouble"  "subject"      "married"   "immediately" "sister"
## [12,] "feelings" "morning"      "uncle"     "agreeable"   "acquaintance"
## [13,] "day"      "acquaintance" "day"       "heard"       "told"
## [14,] "perfectly" "pleasure"     "coming"    "pleasure"    "pleasure"
## [15,] "town"     "attention"    "aunt"      "walk"        "ladies"
##      Topic 6
## [1,] "heading"
## [2,] "ladies"
## [3,] "protested"
## [4,] "accompanied"
## [5,] "read"
## [6,] "speak"
## [7,] "entered"
## [8,] "aunt"
## [9,] "efforts"
## [10,] "obeisance"
## [11,] "date"
## [12,] "crammed"
## [13,] "dawson"
## [14,] "spiteful"
## [15,] "elevation"

```

## c - Interpreting the Results

Topic 1 seems to be about letters to father and mother and the subject seems to relate to how friends and other ladies have opinions on something. Topic 2 seems to be about family members discussion of an expected marriage. Topic 3 seems to be a little random but talks about a day in pleasure. Topic 4 seems to about answering feelings of love and hope. Topic 5 seems to be another letter to family looking forward to a happy day of marriage. Topic 6 seems to be about random things including an aunt, ladies and reading/speaking.

## 4 - Explore word-topic probabilities (The beta $\beta$ matrix)

We can explore and interpret the model using tidying functions from the tidytext and tidyverse packages. We can use the tidy() function from the “broom” package to turn a document-term matrix into a tidy data frame (the broom package by Robinson 2017). We ran our topic model and created an object called topic\_model. We can now plot the betas. Beta is the probability assignments of words to topics.

### a - Calculate $\beta$

**Word-topic probabilities (The beta matrix):** Let’s start with the tidy verb to calculate Beta. Beta is the probability assignments of words to topics. In other words, Beta is the probability that a word contributes to a topic. We get a a combination for each topic and each term: **what is the probability that this term is associated with this topic.**

```
topics <- tidy(topic_model, matrix = "beta") #probability that the term is associated with this topic.
```

```
topics
```

```
## # A tibble: 36,378 × 3
##   topic term      beta
##   <int> <chr>   <dbl>
## 1     1  1 15  7.11e-19
## 2     2  2 15  8.63e-22
## 3     3  3 15  5.53e-23
## 4     4  4 15  8.00e-24
## 5     5  5 15  2.09e-22
## 6     6  6 15  4.13e- 3
## 7     7  1  2  1.74e-20
## 8     8  2  2  2.65e-23
## 9     9  3  2  2.05e-24
## 10    10  4  2  3.74e-25
## # ... with 36,368 more rows
```

*# Notice that this has turned the model into a one-topic-per-term-per-row format.*

### b - Calculate and visualize the 10 terms that has the highest probability of belonging to each topic.

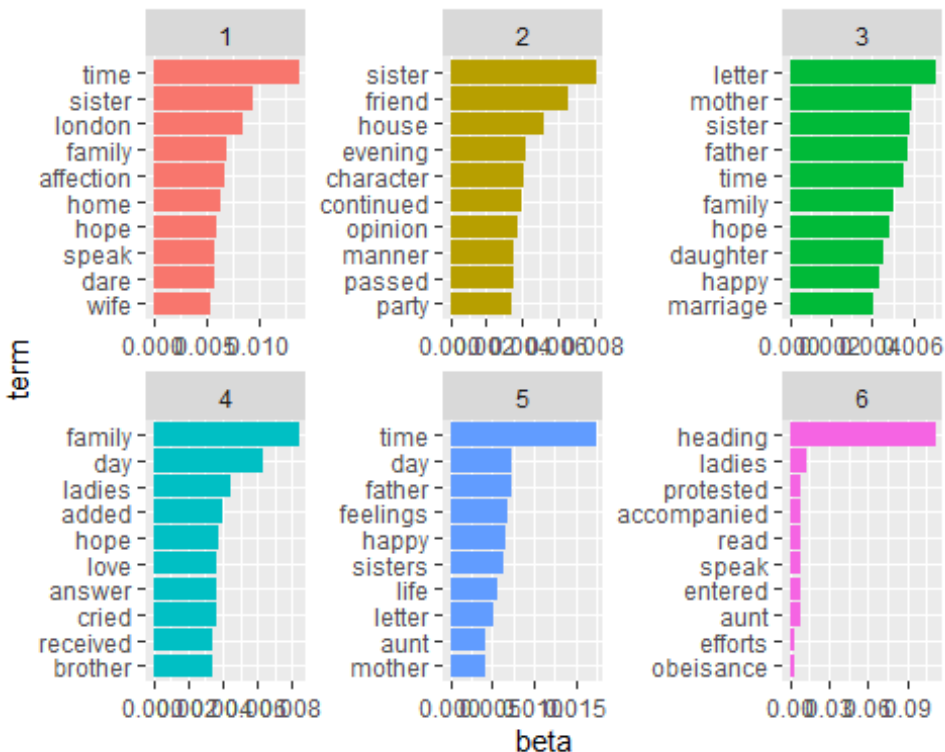
We can use the dplyr’s top\_n() function to find the 10 terms that are most common within each topic ( has the highest probability of belonging to each topic). As a tidy data frame, this lends itself well to a ggplot2 visualization.

```
top_terms <- topics %>%
  group_by(topic) %>% #let's group_by each topic
  top_n(10, beta) %>% #take the top 10 words in each topic
  ungroup() %>% #it's always good to ungroup
  arrange(topic, -beta) #arrange in ascending order, by default
```

```

top_terms %>%
  mutate(term = reorder_within(term, beta, topic)) %>% #reorganize
  ggplot(aes(term, beta, fill = factor(topic))) + # create a plot
  geom_col(show.legend = FALSE) + #create a bar graph
  facet_wrap(~ topic, scales = "free") + # bar graph of topics
  coord_flip() + # switch x and y axis
  scale_x_reordered() # transform the factor levels

```



### c - Interpret the results. What does the beta matrix represent?

The Beta Matrix is a one-topic-per-term-per-row matrix that shows the probability that a specific word contributes to a specific topic. There is some evidence that the words associated with some of the topics cohere into something that could be called a latent theme or topic within the text:

- topic #1 includes words that related to family in a household and a letter which could include friends
- topic #2 includes more family like sister, mother, uncle, daughter with regard to time and marriage
- topic #3 seems like it is about the family take a day of pleasure with an aunt
- topic #4 appears to talk about a sister that is ill and that time has passed since feelings of love and hope have been shared
- topic #5 appears to describe a letter to a father with hope about a time for a happy day for the family

But there is one topic that don't make much sense. Topic #6 has chapter and heading as the words most likely relating to the topic. This is the issue of naively assigning topics.

## 5 - Explore document-topic probability (The gamma $\gamma$ matrix)

LDA also models each document as a mixture of topics. With the `matrix = "gamma"` argument in `tidy()`, you can examine the per-document-per-topic probabilities, called  $\gamma$  ("gamma"). `gamma` is the probability that a document contributes to a topic.

### a - Calculate $\gamma$

```
documents <- tidy(topic_model, matrix = "gamma") # examine the per-document-  
per-topic probabilities  
documents
```

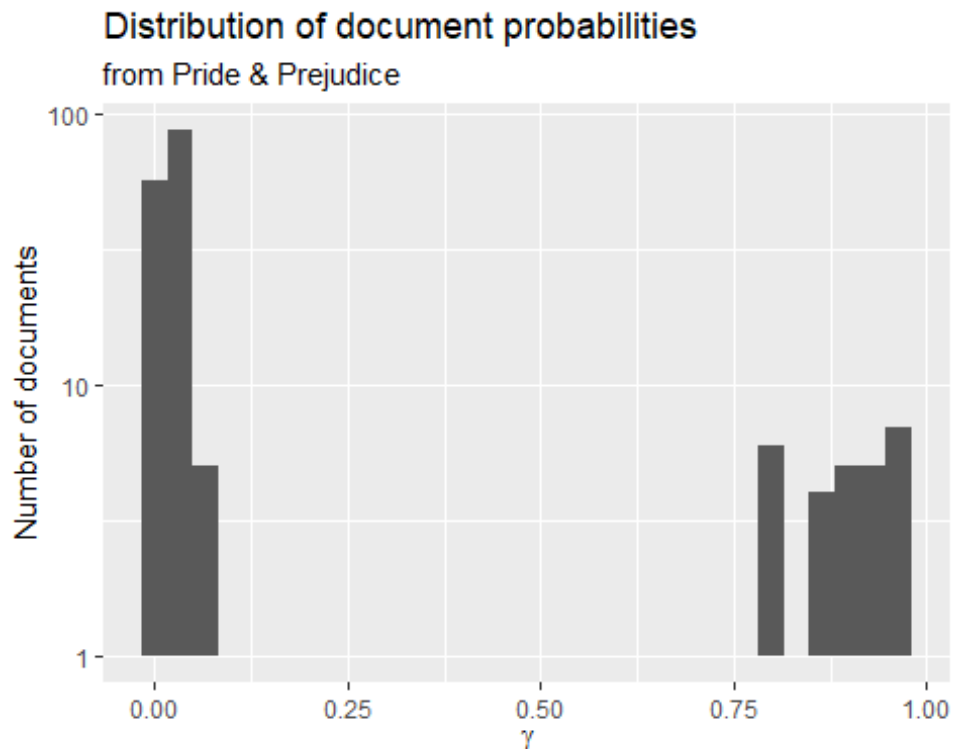
```
## # A tibble: 180 × 3  
##   document                                     topic  
gamma  
##   <chr>                                     <int>  
<dbl>  
## 1 Heading to Chapter I.                     ...    1  
0.00936  
## 2 Heading to Chapter IV.                   ...    1  
0.0382  
## 3 Heading to Chapter V.                   ...    1  
0.0304  
## 4 Tailpiece to Chapter V.                 ...    1  
0.0514  
## 5 Heading to Chapter VI.                  ...    1  
0.00880  
## 6 Heading to Chapter X.                   ...    1  
0.0188  
## 7 Heading to Chapter XII.                 ...    1  
0.0382  
## 8 Heading to Chapter XIII.                ...    1  
0.0382  
## 9 Heading to Chapter XIV.                 ...    1  
0.0188  
## 10 Heading to Chapter XV.                 ...    1  
0.0382  
## # ... with 170 more rows
```

Each of these values is an estimated proportion of words from that document that are generated from that topic. For example, the model estimates that about 83.57% of the words in document "Heading to Chapter XL. 278" on page 11 of the document tibble were generated from topic 4.

## b - Visualize the distribution of document probabilities for each topic

We can plot the results. The y-axis is plotted on a log scale;  $\gamma$  runs from 0 to 1.

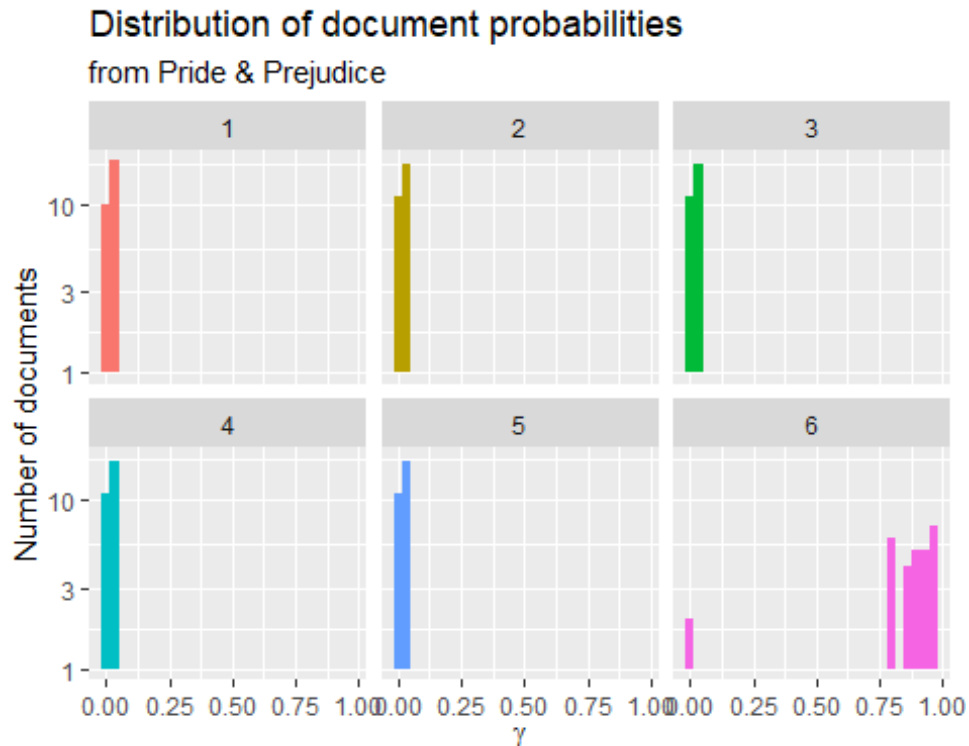
```
ggplot(documents, aes(gamma)) + # create a plot with the document tibble
geom_histogram() + # create a histogram
scale_y_log10() + # make the data more uniform
labs(title = "Distribution of document probabilities",
      subtitle = "from Pride & Prejudice",
      y = "Number of documents", x = expression(gamma))
```



We can also plot the distribution of probability for each topic:

```
ggplot(documents, aes(gamma, fill = as.factor(topic))) + # create plot with
document tibble by factor
geom_histogram(show.legend = FALSE) + # create a histogram without a
Legend
facet_wrap(~ topic, ncol = 3) + # bar graph of topics in 3 columns
scale_y_log10() + # make the data more uniform
labs(title = "Distribution of document probabilities",
      subtitle = "from Pride & Prejudice",
      y = "Number of documents", x = expression(gamma))
```





**c- Interpret the results. What does the gamma matrix represent?**

The graphs show that each document appears in each panel in this plot, and its gamma for that topic tells us that document's probability of belonging in that topic. There are many values near gamma = 0 for both topics 1 through 5; which means there are many documents that do not belong in each topic. There is only topic 6 with values near gamma = 1; these are the documents that do belong in those topics. This gamma distribution shows that documents are being well discriminated as belonging to a topic or not.

**6 - Answer the following questions:**

**a - What is interesting about your topics? Do they match what you expect? If not, what looks peculiar?**

Having read the book, I do think that the topics make sense. There are letters that pass between characters and the family is central to the story which is about marrying off of Mr. Bennet's daughters.

**b - How many topics did you choose? Why?**

I chose 6 topics because it was simple to see when graphed.

**c - Which topics stand out? Which topics seem to be junk?**

The first 5 topics seem to be well grouped and the last topic seemed strange with chapter references.

**d - What are some of the limitation of topic modeling? How can we overcome the limitations of topic models?**

In the LDA modeling, we can just identify the general themes or topics in our corpus. If you estimate the number of topics, then visualize them in beta and gamma distributions, then you can refine the number of topics by increasing or decreasing them.

**e - What new research questions does this assignment inspire? If you had time for a bigger project, what would it be?**

If there was more time, I would include another novel to compare it to. Recently, I read a novel which combines Pride and Prejudice with Frankenstein. It is called "Pride and Prometheus" by John Kessel and is a marvelous addition to the classic.