Who wrote the Federalist papers?

Savaş Dayanık

2024-03-14

Table of contents

Introduction

Wikipedia on The Federalist Papers

> The Federalist Papers is a collection of 85 articles and essays written by Alexander Hamilton, James Madison, and John Jay under the collective pseudonym "Publius" to promote the ratification of the United States Constitution. The collection was commonly known as The Federalist until the name The Federalist Papers emerged in the 20th century. They are published between October 27, 1787 and May 28, 1788.

Wikipedia on Origins

> The Federal Convention (Constitutional Convention) sent the proposed Constitution to the Confederation Congress, which in turn submitted it to the states for ratification at the end of September 1787. On September 27, 1787, "Cato" first appeared in the New York press criticizing the proposition; "Brutus" followed on October 18, 1787. These and other articles and public letters critical of the new Constitution would eventually become known as the "Anti-Federalist Papers". In response, Alexander Hamilton decided to launch a measured defense and extensive

explanation of the proposed Constitution to the people of the state of New York. He wrote in Federalist No. 1 that the series would "endeavor to give a satisfactory answer to all the objections which shall have made their appearance, that may seem to have any claim to your attention."
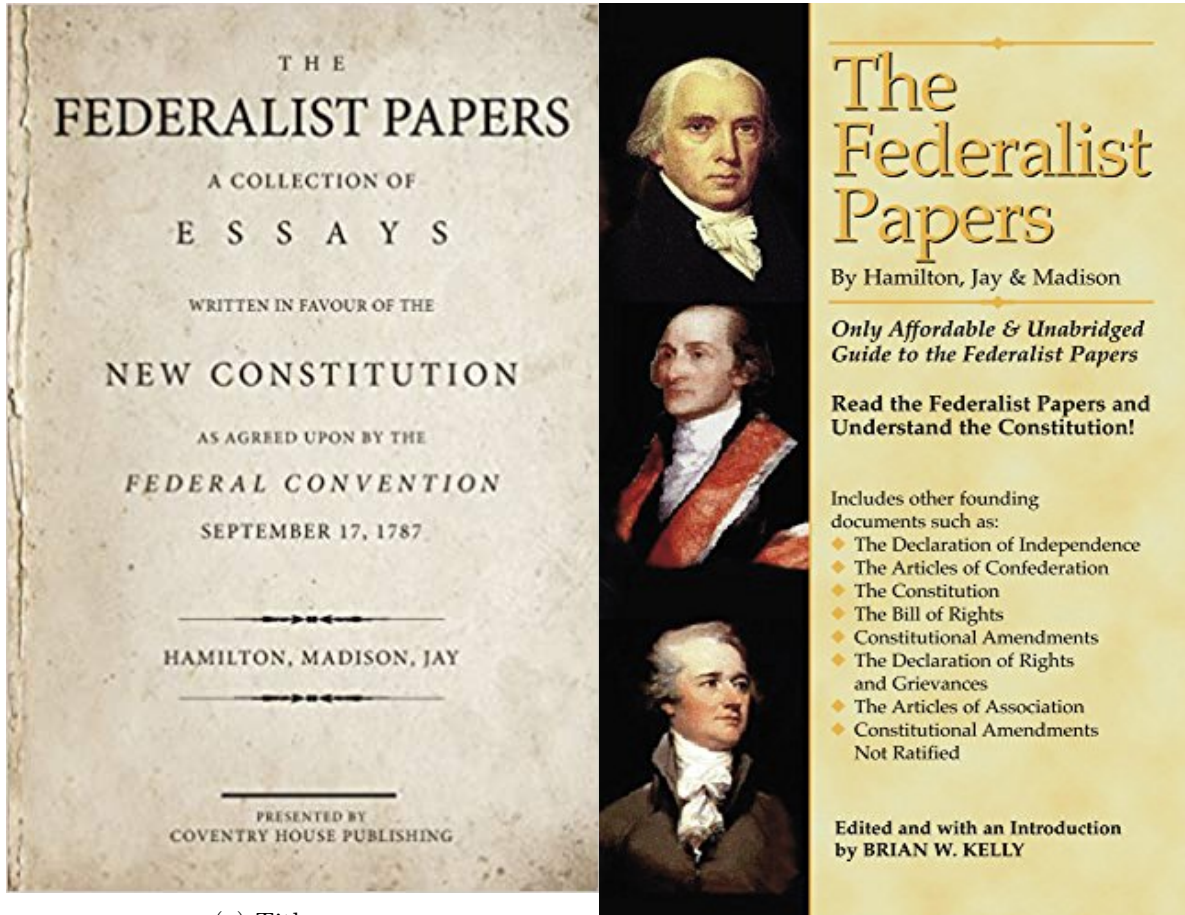
(a) Title page

(b) Back page

Figure 1: Federalist papers in print

Hamilton recruited collaborators for the project. He enlisted John Jay, who after four strong essays (Federalist Nos. 2, 3, 4, and 5), fell ill and contributed only one more essay, Federalist No. 64, to the series. Jay also distilled his case into a pamphlet in the spring of 1788, An Address to the People of the State of New-York; Hamilton cited it approvingly in Federalist No. 85. James Madison, present in New York as a Virginia delegate to the Confederation Congress, was recruited by Hamilton and Jay and became Hamilton's primary collaborator.

Wikipedia on Authorship

Authorship

James Madison, Hamilton's major collaborator, later fourth President of the United States (1809-1817)

At the time of publication, the authors of The Federalist Papers attempted to hide their identities due to Hamilton and Madison having attended the convention. Astute observers, however, correctly discerned the identities of Hamilton, Madison, and Jay. Establishing authorial authenticity of the essays that constitute The Federalist Papers has not always been clear. After Alexander Hamilton died in 1804, a list emerged, claiming that he alone had written two-thirds of The Federalist essays. Some believe that several of these essays were written by James Madison (Nos. 49–58 and 62–63). The scholarly detective work of Douglass Adair in 1944 postulated the following assignments of authorship, corroborated in 1964 by a computer analysis of the text:

- Alexander Hamilton (51 articles: Nos. 1, 6–9, 11–13, 15–17, 21–36, 59–61, and 65–85)
- James Madison (29 articles: Nos. 10, 14, 18–20, 37–58 and 62–63)
- John Jay (5 articles: Nos. 2–5 and 64).

In six months, a total of 85 articles were written by the three men. Hamilton, who had been a leading advocate of national constitutional reform throughout the 1780s and was one of the three representatives for New York at the Constitutional Convention, in 1789 became the first Secretary of the Treasury, a post he held until his resignation in 1795. Madison, who is now acknowledged as the father of the Constitution—despite his repeated rejection of this honor during his lifetime, became a leading member of the U.S. House of Representatives from Virginia (1789–1797), Secretary of State (1801–1809), and ultimately the fourth President of the United States (1809–1817). John Jay, who had been secretary for foreign affairs under the Articles of Confederation from 1784 through their expiration in 1789, became the first Chief Justice of the United States in 1789, stepping down in 1795 to accept election as governor of New York, a post he held for two terms, retiring in 1801.

[Wikipedia on Disputed essays](#) says

Disputed essays

While the authorship of 73 of The Federalist essays is fairly certain, the identities of those who wrote the twelve remaining essays are disputed by some scholars. The modern consensus is that Madison wrote essays Nos. 49–58, with Nos. 18–20 being products of a collaboration between him and Hamilton; No. 64 was by John Jay. The first open designation of which essay belonged to whom was provided by Hamilton who, in the days before his ultimately fatal gun duel with Aaron

Burr, provided his lawyer with a list detailing the author of each number. This list credited Hamilton with a full 63 of the essays (three of those being jointly written with Madison), almost three-quarters of the whole, and was used as the basis for an 1810 printing that was the first to make specific attribution for the essays.

Madison did not immediately dispute Hamilton's list, but provided his own list for the 1818 Gideon edition of The Federalist. Madison claimed 29 essays for himself, and he suggested that the difference between the two lists was "owing doubtless to the hurry in which [Hamilton's] memorandum was made out." A known error in Hamilton's list — Hamilton incorrectly ascribed No. 54 to John Jay, when in fact, Jay wrote No. 64 — provided some evidence for Madison's suggestion.

Statistical analysis has been undertaken on several occasions in attempts to accurately identify the author of each individual essay. After examining word choice and writing style, studies generally agree that the disputed essays were written by James Madison. However, there are notable exceptions maintaining that some of the essays which are now widely attributed to Madison were, in fact, collaborative efforts.

Similar questions arised for the works of Shakespeare, Homer.

```
d <- ProbBayes::federalist_word_study %>%
  mutate_if(is.factor, as.character) %>%
  mutate(Authorship = ifelse(Authorship == "", "Jay", Authorship),
         Authorship = ifelse(Disputed == "yes", "Hamilton or Madison", Authorship))

d %>%
  distinct(Name, Authorship, Disputed) %>%
  count(Authorship, Disputed) %>%
  pander()
```

Authorship	Disputed	n
Hamilton	no	51
Hamilton or Madison	yes	12
Jay	no	5
Madison	no	17

```
panderOptions("table.alignment.default", c("left", "left"))

d %>%
  distinct(Name, Authorship, Disputed) %>%
```

```
mutate(Name = Name %>% str_replace("\\.", "") %>% parse_number()) %>%
arrange(Authorship, Name) %>%
group_by(Authorship) %>%
summarize(Name = paste(Name, collapse = ", ")) %>%
pander()
```

| Authorship	Name
Hamilton	1, 6, 7, 8, 9, 11, 12, 13, 15, 16, 17, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 59, 60, 61, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85
Hamilton or Madison	49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 62, 63
Jay	2, 3, 4, 5, 64
Madison	10, 14, 18, 19, 20, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48

- Mosteller and Wallace (1963) illustrated how to use Bayesian reasoning in solving authorship problem.

- Applied Bayesian and Classical InferenceThe Case of The Federalist Papers
- Gutenberg project
- Shakespeare authorship question
- Homeric Question

Data on word use

- Focus on frequencies of word counts (across different papers, normalized, say in every 1,000 words)
- Focus on "filler" words (also called stopwords)
  - Examples
    * an
    * of
    * upon
    * can
  - Why fillers?
    * Because others will depend on the topic of the current topic as well.
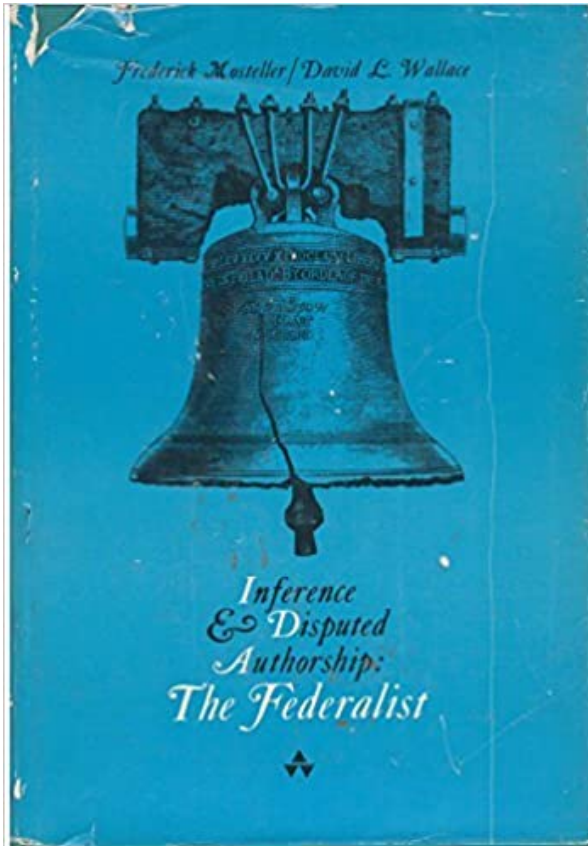
(a) Title page
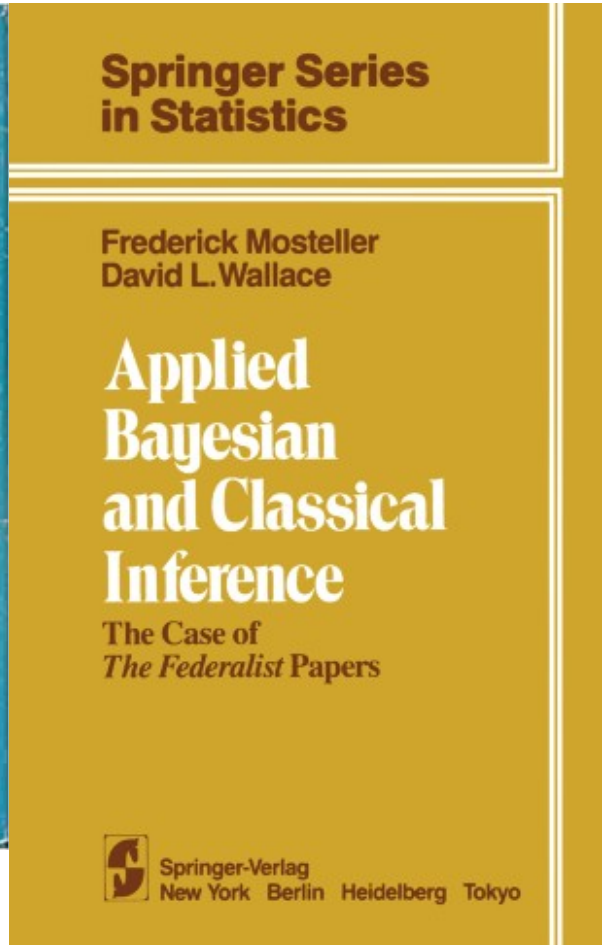
(b) Back page

Figure 2: Mosteller and Wallace's Book on Bayesian Analysis of Federalist Papers

* Because fillers are expected to appear at the same rate across all articles, but the rate will depend on the style of the author. Well at least for some filler words – we hope.

- If you want to learn topics of papers, not the author styles, then remove stopwords and study the co-occurrences of the remaining contextual words.

```
d %>%
  filter(Disputed == "no", Authorship != "Jay") %>%
  filter( word %in% c("the", "of", "to", "and", "be", "that")) %>%
  ggplot(aes(Rate)) +
  geom_density(aes(fill=Authorship), alpha=0.5) +
  facet_wrap(~word, scale = "free_x")
```
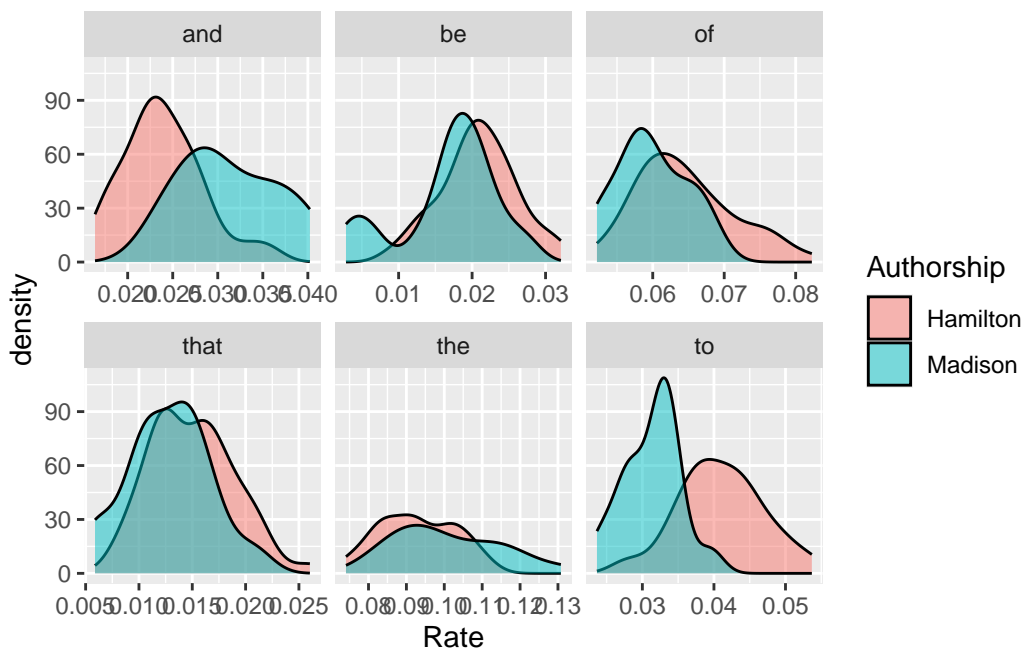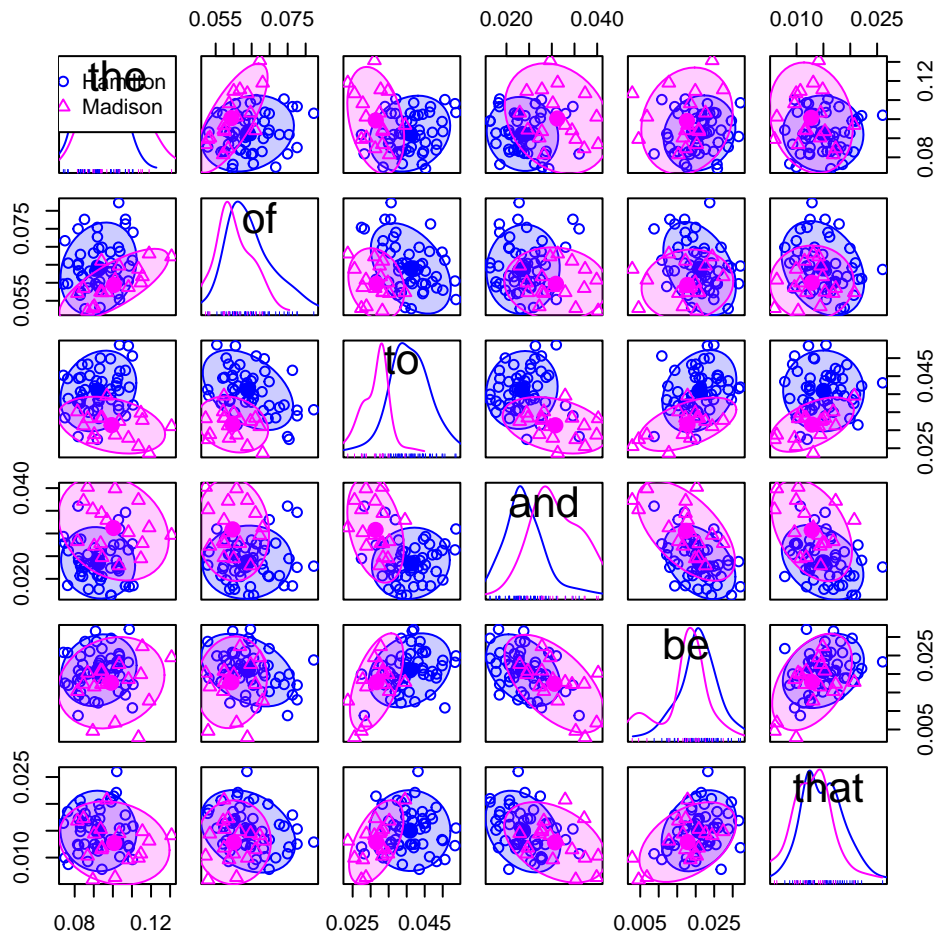


```
pairwise <- d %>%
  filter(Disputed == "no", Authorship != "Jay") %>%
  filter( word %in% c("the", "of", "to", "and", "be", "that")) %>%
  select(Name, word, Rate, Authorship) %>%
  pivot_wider(names_from = word, values_from = "Rate") %>%
  select(-Name)

scatterplotMatrix( ~ the + of + to + and + be + that | Authorship, data = pairwise,
```

Can be used non-Bayesian methods like * Use Fisher discriminant analysis (FDA) * kNN

Classification with kNN

```r
all <- d %>%
  select(Name, word, Rate, Authorship) %>%
  pivot_wider(names_from = word, values_from = Rate) %>%
  mutate_if(is.numeric, ~ifelse(is.na(.), 0, .)) %>%
  select(-Name)
```
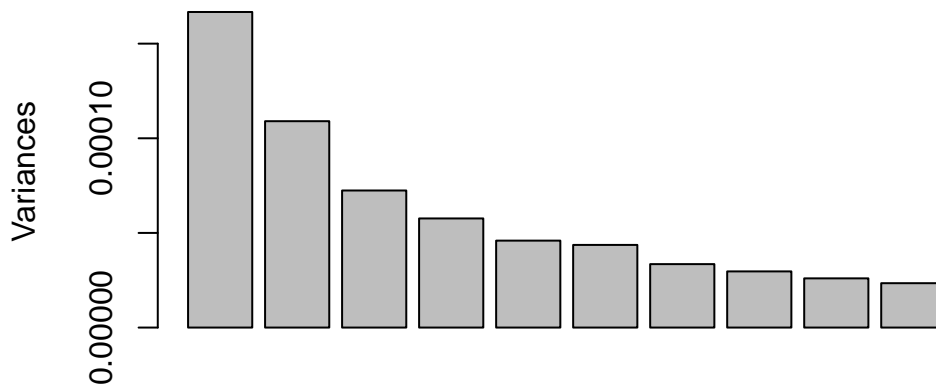
```r
train <- all %>%
  filter(Authorship %in% c("Hamilton", "Madison"))

res_pc <- prcomp(train[,-1])
res_pc %>% plot()
```

.


```r
summary(res_pc)
```

```
Importance of components:
                            PC1     PC2     PC3      PC4      PC5      PC6
Standard deviation      0.01291 0.01044 0.00851 0.007592 0.006777 0.006607
Proportion of Variance  0.15518 0.10148 0.06740 0.053660 0.042750 0.040630
Cumulative Proportion   0.15518 0.25666 0.32406 0.377720 0.420470 0.461100
                            PC7      PC8      PC9     PC10     PC11     PC12
Standard deviation      0.00579 0.005448 0.005098 0.004842 0.004689 0.004584
Proportion of Variance  0.03120 0.027620 0.024190 0.021830 0.020470 0.019550
Cumulative Proportion   0.49230 0.519930 0.544120 0.565940 0.586410 0.605970
                            PC13     PC14    PC15     PC16     PC17     PC18
Standard deviation      0.00444 0.004322 0.00423 0.004041 0.003819 0.003762
Proportion of Variance  0.01835 0.017390 0.01666 0.015200 0.013570 0.013180
Cumulative Proportion   0.62431 0.641700 0.65836 0.673560 0.687140 0.700310
                            PC19    PC20     PC21    PC22     PC23     PC24
Standard deviation      0.003652 0.00357 0.003509 0.00342 0.003375 0.003323
Proportion of Variance  0.012420 0.01186 0.011460 0.01089 0.010600 0.010270
Cumulative Proportion   0.712730 0.72459 0.736050 0.74694 0.757540 0.767810
                            PC25     PC26     PC27    PC28     PC29     PC30
Standard deviation      0.003295 0.003181 0.003048 0.00302 0.002997 0.002919
```

9

```
                                          Proportion of Variance 0.010100 0.009420 0.008650 0.00849 0.008360 0.007930
Cumulative Proportion  0.777920 0.787340 0.795990 0.80448 0.812840 0.820770
                             PC31     PC32     PC33     PC34     PC35     PC36
Standard deviation     0.002909 0.002855 0.002822 0.002764 0.002726 0.002708
Proportion of Variance 0.007880 0.007590 0.007410 0.007110 0.006920 0.006820
Cumulative Proportion  0.828650 0.836230 0.843650 0.850760 0.857680 0.864500
                             PC37     PC38     PC39     PC40     PC41     PC42
Standard deviation     0.002701 0.002595 0.002588 0.002546 0.002489 0.002466
Proportion of Variance 0.006790 0.006270 0.006230 0.006030 0.005770 0.005660
Cumulative Proportion  0.871290 0.877560 0.883790 0.889820 0.895590 0.901250
                             PC43     PC44     PC45     PC46     PC47     PC48
Standard deviation     0.002433 0.002391 0.002361 0.002307 0.002278 0.002248
Proportion of Variance 0.005510 0.005320 0.005190 0.004950 0.004830 0.004700
Cumulative Proportion  0.906760 0.912080 0.917270 0.922230 0.927060 0.931760
                             PC49     PC50     PC51     PC52     PC53     PC54
Standard deviation     0.002215 0.002173 0.002156 0.002125 0.002122 0.002115
Proportion of Variance 0.004570 0.004400 0.004330 0.004200 0.004190 0.004160
Cumulative Proportion  0.936330 0.940720 0.945050 0.949250 0.953440 0.957600
                             PC55     PC56     PC57     PC58     PC59     PC60
Standard deviation      0.00207 0.002029  0.00201 0.001989 0.001956 0.001907
Proportion of Variance  0.00399 0.003830  0.00376 0.003680 0.003560 0.003390
Cumulative Proportion   0.96159 0.965420  0.96919 0.972870 0.976430 0.979810
                             PC61     PC62     PC63     PC64     PC65     PC66
Standard deviation      0.00189 0.001846 0.001811 0.001763 0.001706 0.001702
Proportion of Variance  0.00333 0.003170 0.003050 0.002890 0.002710 0.002700
Cumulative Proportion   0.98314 0.986310 0.989360 0.992260 0.994960 0.997660
                             PC67     PC68
Standard deviation     0.001585 1.112e-17
Proportion of Variance 0.002340 0.000e+00
Cumulative Proportion  1.000000 1.000e+00
```
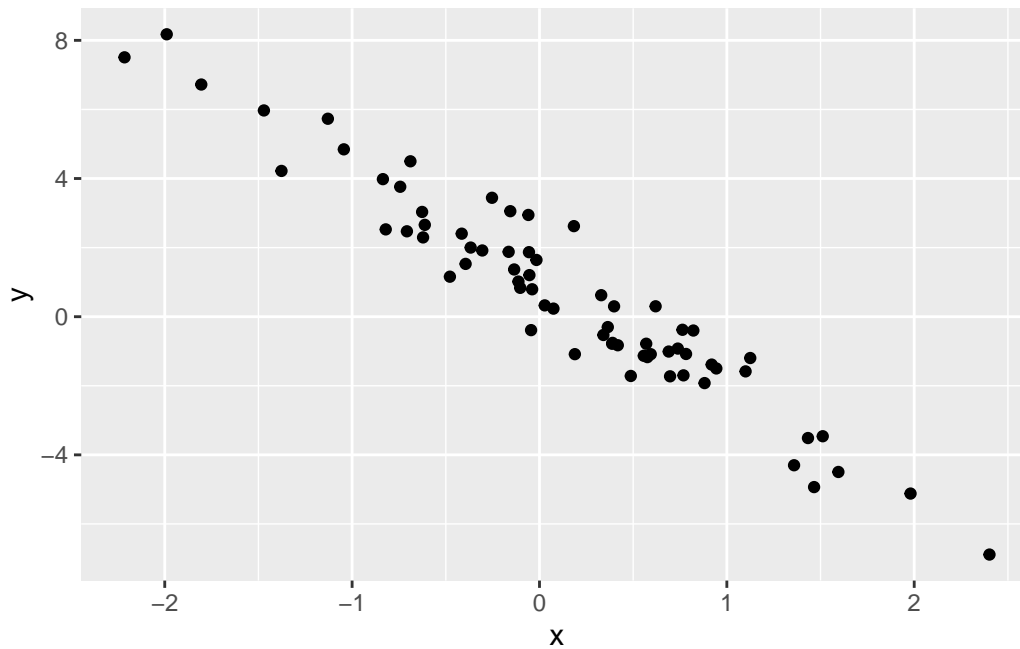
```r
(res_pc$sdev)^2 %>% {./sum(.)} %>% cumsum %>% plot()
```
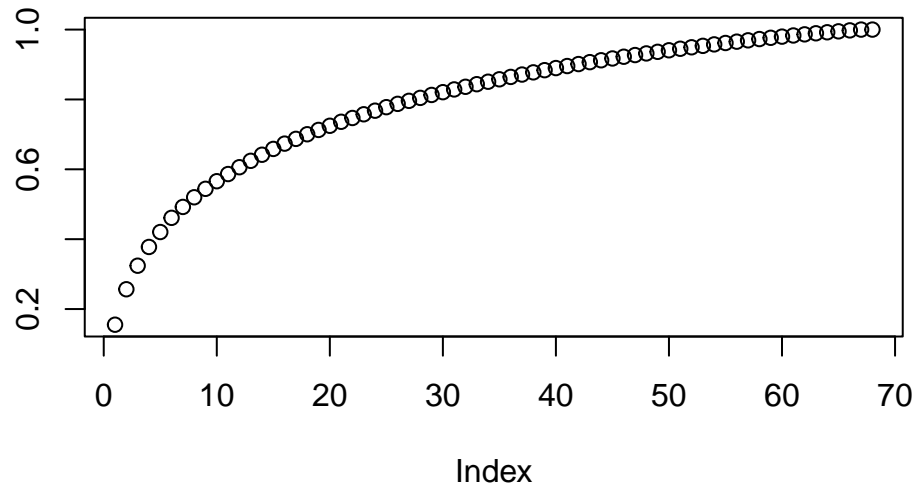
```r
numPC <- 68
xtrain <- res_pc$x[,seq(numPC)]
```

Principal component analysis

```r
set.seed(1)
dt <- tibble(x = rnorm(68), y = -3*x + rnorm(68, 1))
dt %>% ggplot(aes(x, y)) +
  geom_point()
```

```r
res_dt_pc <- prcomp(dt)
res_dt_pc
```

```
Standard deviations (1, .., p=2):
[1] 3.1109263 0.2381988

Rotation (n x k) = (2 x 2):
         PC1         PC2
x  0.2817787 -0.9594794
y -0.9594794 -0.2817787
```
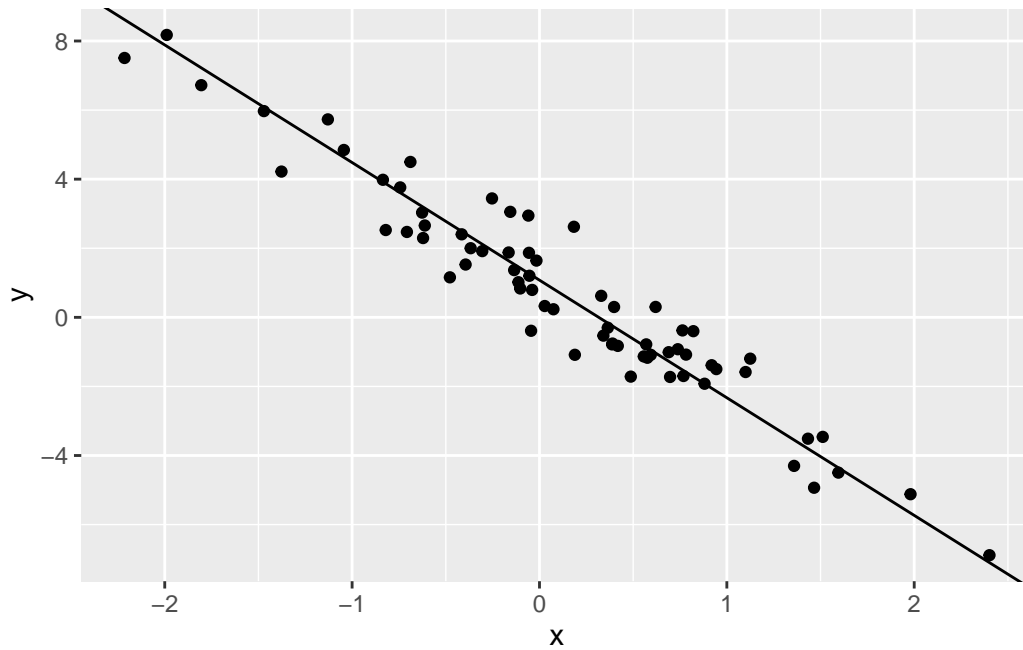
```r
# res_dt_pc$rotation^2 %>% apply(2, sum)
cntr <- dt %>% apply(2, mean)

slope1 <- res_dt_pc$rotation[,"PC1"] %>% {.[2]/.[1]}
intercept1 <- cntr[2] - slope1*cntr[1]
slope2 <- res_dt_pc$rotation[,"PC2"] %>% {.[2]/.[1]}
intercept2 <- cntr[2] - slope2*cntr[1]

dt %>% ggplot(aes(x, y)) +
  geom_point() +
```

```
geom_abline(intercept = intercept1, slope = slope1)
```



```
# +
#   geom_abline(intercept = intercept2, slope = slope2) +
#   coord_fixed()
# +
#   geom_smooth(method = "lm", se = FALSE)
```

```
res_dt_pc %>% summary()
```

```
Importance of components:
                          PC1      PC2
Standard deviation     3.1109  0.23820
Proportion of Variance 0.9942  0.00583
Cumulative Proportion  0.9942  1.00000
```
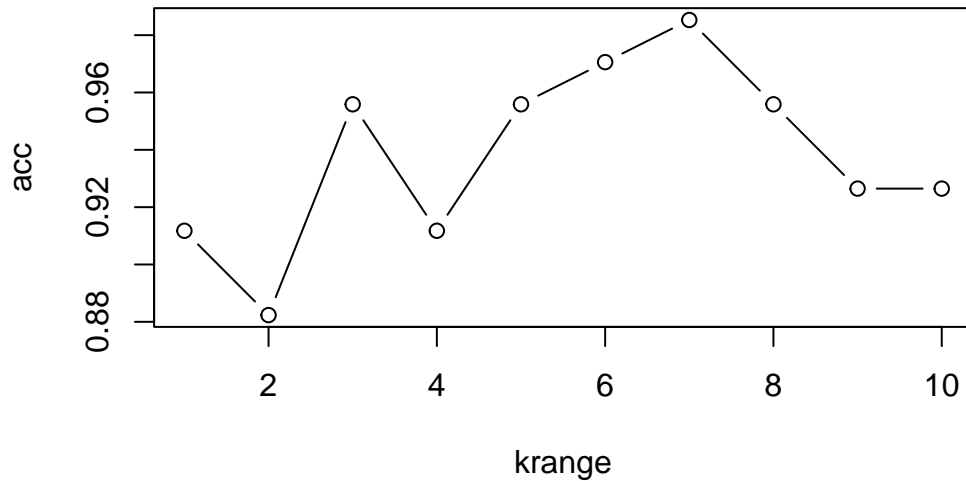
Back to federalist papers.

```
acc <- NULL
krange <- 1:10
```

```
for (i in krange) {
  res_knn <- knn.cv(train = xtrain, cl = train$Authorship, k = i, prob = TRUE)
  acc <- c(acc, table(predicted=res_knn, observed=train$Authorship) %>% prop.table() %>% d
  # cat(i, acc, "\n")
} # if there is tie between the counts of two counts, then tie is broken at random.
plot(krange, acc, t="b")
```



```
test <- all %>%
  filter(Authorship %in% c("Hamilton or Madison"))

xtest <- predict(res_pc, test)[, seq(numPC)]
pred <- knn(train=xtrain, test=xtest, cl = train$Authorship, k = which.max(acc))
table(pred)
```

```
pred
Hamilton  Madison
      7        5
```

```
#knitr::knit_exit()
```

Topic analysis

A simple topic model assumes that

- each document is a bag of words,
- each topic is a probability distribution on the collection of terms (soccer, volleyball, score, coach, income, earnings, salaries, interest, foreign exchange rate, budget, etc.)
- each document is generated by a mixture distribution of topics (sports, finance, politics, history, etc.).

More precisely, let $w_{d,1}, \dots, w_{d,n}$ be the terms appearing in a document d of a corpus D. Then

- each word $w_{d,i}$ is related to some topic $z_{d,i}$, which can be sports, politics, finance, and so on,
- given that $z_{d,i} = t$ is the topic, word w_i can be $v_1, \dots, v_m$ in a fixed dictionary $V = \{v_1, \dots, v_m\}$ with probabilities $beta_{t,1}, \dots, \beta_{t_m}$, respectively. Here $\beta_t = (\beta_{t,i})_{i=1,\dots,m}$ is a discrete **topic-term distribution** identified with topic z and has to be learned from data,
- the topics $z_{d,1}, z_{d,2}, \dots, z_{d,n}$ are independent discrete random variables taking in some topic set $T = \{t_1, \dots, t_k\}$ with probabilities $\gamma_{d,1}, \dots, \gamma_{d,k}$, respectively. Here $\gamma_d = (\gamma_{d,k})_{k=1,\dots,k}$ is called **document-topic distribution** and also has to be learned from data.

These conditional distributions are often presented with a graphical model. Bayesian methods are implemented to learn document-topic distributions γ_d, $d \in D$ and topic-term distributions $\beta_t, t \in T$.

Stop words (is, and, the, can, should, must, etc) are not topical words and are removed from the documents.

```
# ?stop_words
stop_words %>% head()
```

```
# A tibble: 6 x 2
  word      lexicon
  <chr>     <chr>
1 a         SMART
2 a's       SMART
3 able      SMART
4 about     SMART
5 above     SMART
6 according SMART
```

Each document is a bag of words:

```
d_txt <- d %>%
  select(Name, word, N) %>%
```

```
    mutate(line = map2_chr(word, N, ~{rep(.x, .y) %>% paste(collapse=" ")})) %>%
    group_by(Name) %>%
    summarize(text = paste(line, collapse = " "))
  # d_txt$Name
```

Construct the document-term matrix

```
  d_vs <- VectorSource(d_txt$text)
  dtm <- DocumentTermMatrix(Corpus(d_vs),
                            control = list(stemming = FALSE, stopwords = TRUE, minWordLength
                                           removeNumbers = TRUE, removePunctuation = T
  # mm <- dtm %>% as.matrix()
  # dim(dtm)
  # mm %>% apply(1,sum)
```

Fit the latent Dirichlet allocation (LDA) model. Here we fit the model with $k = 5$ topics.

```
  dt <- system.time(res_lda <- LDA(dtm, k = 5, seed = 1234))
  cat("LDA fit time: ", dt["elapsed"], " seconds.", sep="")
```

```
LDA fit time: 6.148 seconds.
```

Extract the topic-term probabilities

```
  topics <- tidy(res_lda, matrix = "beta")

  topics %>%
    filter(topic == 1) %>%
    slice_max(beta, n=10)
```

```
# A tibble: 10 x 3
   topic term            beta
   <int> <chr>          <dbl>
 1     1 will          0.0138
 2     1 government    0.0109
 3     1 states        0.00927
 4     1 may           0.00898
 5     1 state         0.00757
 6     1 union         0.00735
```

```
7       1 can         0.00712
8       1 must        0.00690
9       1 power       0.00683
10      1 upon        0.00640
```

```
topics %>%
  filter(topic == 2) %>%
  arrange(desc(beta)) %>%
  slice_max(beta, n=10)
```

```
# A tibble: 10 x 3
   topic term                  beta
   <int> <chr>                <dbl>
 1     2 will              0.0198
 2     2 states            0.0170
 3     2 government        0.0141
 4     2 state             0.0137
 5     2 people            0.0113
 6     2 may               0.00971
 7     2 federal           0.00919
 8     2 one               0.00706
 9     2 representatives 0.00634
10     2 constitution      0.00599
```

```
topics %>%
  filter(topic == 3) %>%
  arrange(desc(beta)) %>%
  slice_max(beta, n=10)
```

```
# A tibble: 10 x 3
   topic term                  beta
   <int> <chr>                <dbl>
 1     3 will            0.0160
 2     3 state           0.0139
 3     3 may             0.0129
 4     3 states          0.0123
 5     3 power           0.0121
 6     3 constitution 0.00990
 7     3 government      0.00738
```

```
 8      3 courts       0.00618
 9      3 shall        0.00608
10      3 every        0.00586
```

Topics can be described by the most probable ten terms.

```r
topics %>%
  slice_max(beta, n = 10, by = topic) %>%
  group_by(topic) %>%
  summarize(keyword = paste(term, collapse = ", "))
```

```
# A tibble: 5 x 2
  topic keyword
  <int> <chr>
1     1 will, government, states, may, state, union, can, must, power, upon
2     2 will, states, government, state, people, may, federal, one, representat~
3     3 will, state, may, states, power, constitution, government, courts, shal~
4     4 will, government, people, may, one, states, state, must, time, upon
5     5 executive, will, power, may, senate, one, people, government, legislati~
```

1) how president and senate share power to run the state

2) union, nation, people

3) governments, courts are abide by the constitution

4) problems confronted by the union, states, government at war (independence Apr 19, 1775 – Sep 3, 1783)

5) federal states representation in the government

```r
topics %>%
  slice_max(beta, n = 10, by = topic) %>%
  mutate(topic = factor(topic)) %>%
  arrange(topic, desc(beta)) %>%
  group_by(topic) %>%
  mutate(term = paste(sprintf("%02d", seq(n())), term, sep= " ")) %>%
  ggplot(aes(term, beta)) +
  geom_col(aes(fill=topic)) +
  facet_wrap(~topic, scale = "free_y") +
  coord_flip() +
  theme(legend.position = "none")
```
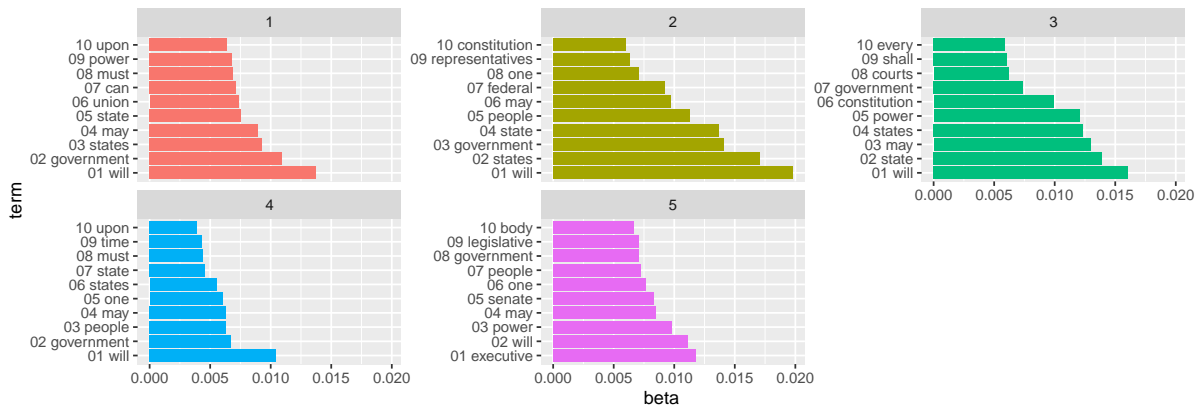
Document-topic probabilities

```
docs <- tidy(res_lda, matrix = "gamma")
docs
```

```
# A tibble: 425 x 3
   document topic     gamma
   <chr>    <int>     <dbl>
 1 1            1 0.402
 2 2            1 0.0000220
 3 3            1 1.00
 4 4            1 0.0444
 5 5            1 1.00
 6 6            1 0.178
 7 7            1 0.421
 8 8            1 1.00
 9 9            1 0.650
10 10           1 0.0000308
# i 415 more rows
```

```
docs %>%
  filter(document %in% sample(unique(document), 20)) %>%
  mutate(topic = factor(topic)) %>%
  ggplot(aes(topic, gamma)) +
  geom_col(aes(fill=topic)) +
  facet_wrap(~document) +
  theme(legend.position = "none")
```
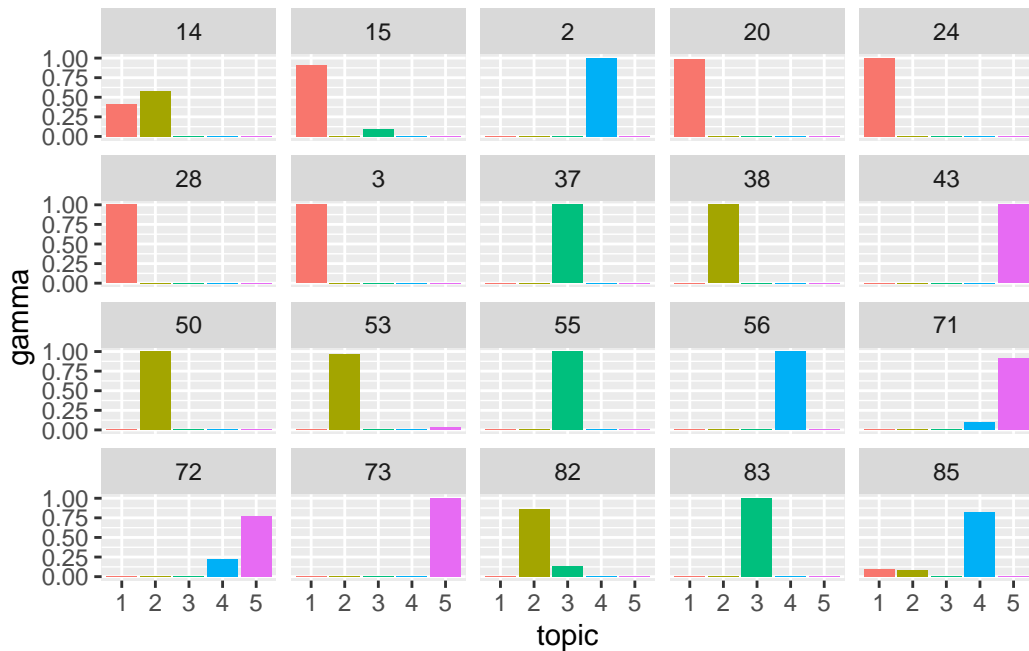
How similar do the topic distributions of articles written by Hamilton and Madison look?

```
# d
docs2 <- docs %>%
  nest(data = -document) %>%
  mutate(Name = d_txt$Name) %>%
  left_join(distinct(d, Name, Authorship), by = "Name") %>%
  unnest(data)

docs2 %>%
  filter(Authorship %in% c("Hamilton", "Madison")) %>%
  nest(data = -c(Authorship, document))  %>%
  slice_sample(n=10, by = Authorship) %>%
  unnest(data) %>%
  arrange(Authorship, document, topic) %>%
  mutate(topic = factor(topic)) %>%
  ggplot(aes(topic, gamma)) +
  geom_col(aes(fill=topic)) +
  facet_wrap(~ Authorship+ document) +
  theme(legend.position = "none")
```
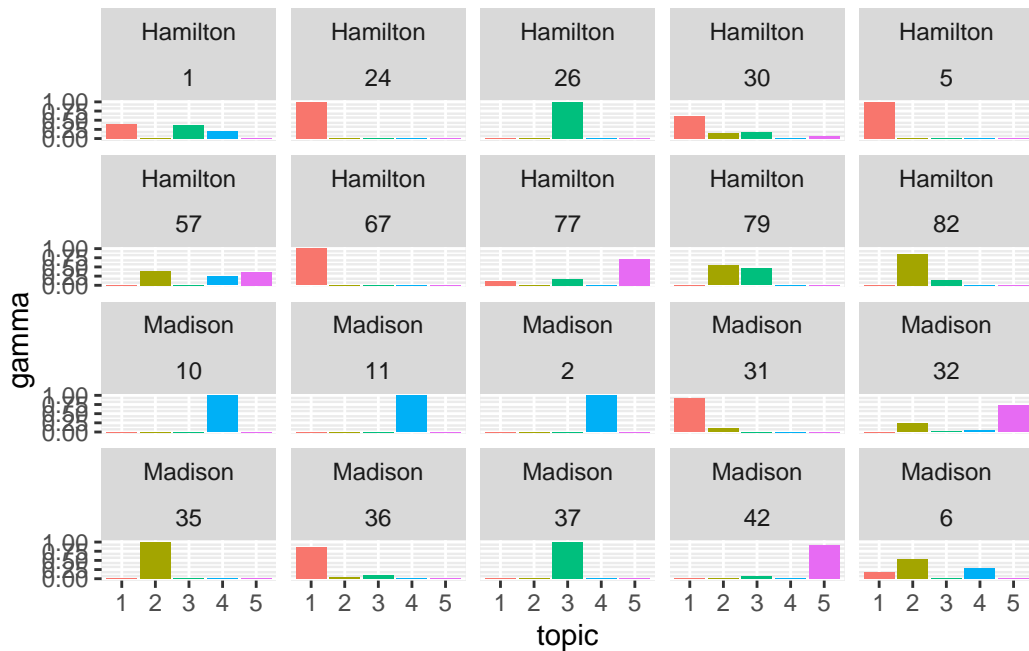
```r
docs3 <- docs2 %>%
  select(document, topic, gamma, Authorship) %>%
  pivot_wider(names_from = topic, values_from = gamma)
```

```r
docs2 %>%
  filter(Authorship %in% c("Hamilton", "Madison")) %>%
  mutate(topic = factor(topic)) %>%
  ggplot(aes(topic, gamma)) +
  geom_boxplot(aes(fill=Authorship)) +
  scale_y_log10()
```
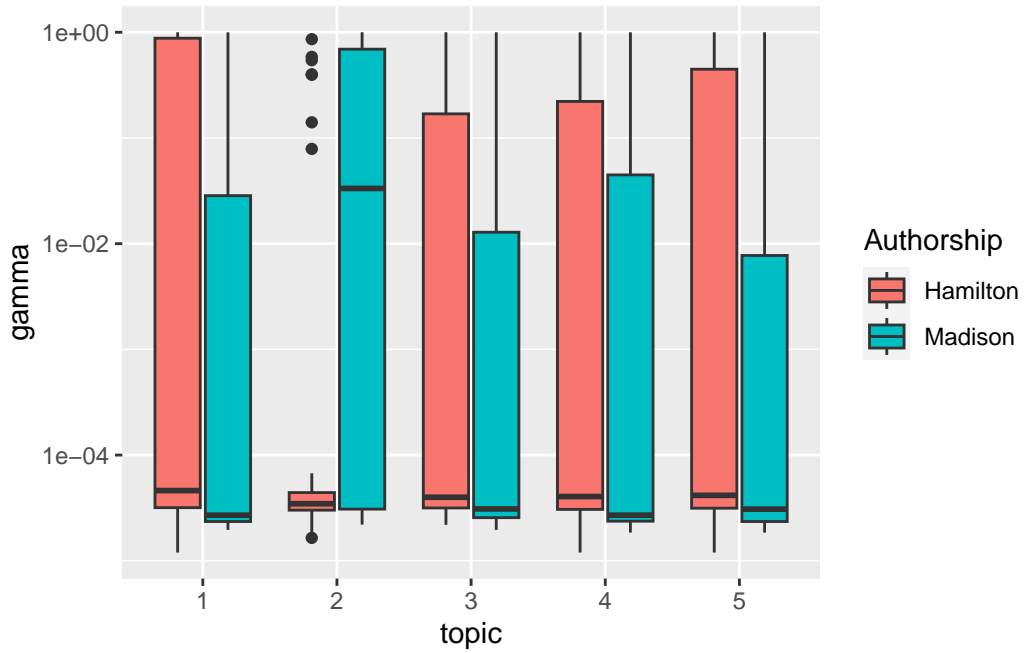
```
scatterplotMatrix(~ `1` +`2` + `3` + `4` + `5`| Authorship,
                  data = docs3 %>% filter(Authorship %in% c("Hamilton", "Madison", "Hamilt
                  log = "xy", smooth = FALSE, regLine = FALSE)
```
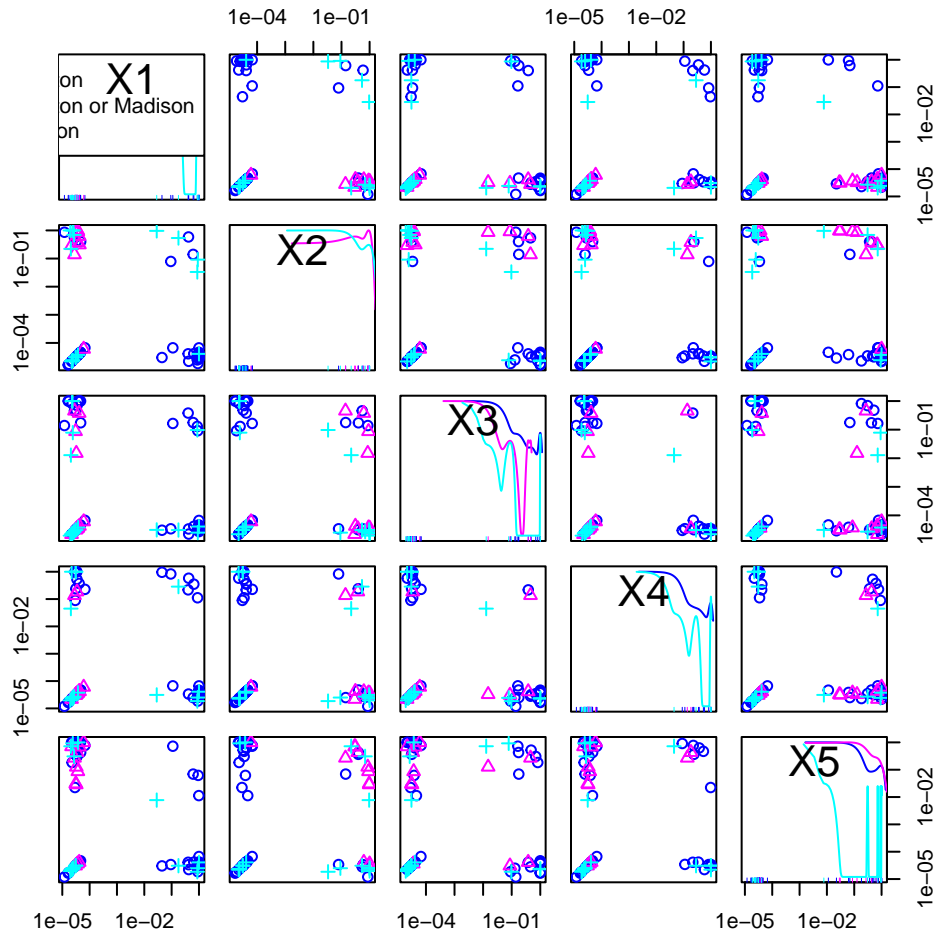
A new kNN model may be fit to the topic distributions of documents to predict the authorships.

An interesting extension of LDA in this problem would be also learn the topic mixtures that each author tends to write about. A new graphical model can be proposed and fit to data. Then for each document one can calculate the probability that the document is written by a given author. Namely, am author-document probability distribution can be estimated.