The Disputed Federalist Papers
John Fields
August 1, 2019

Introduction

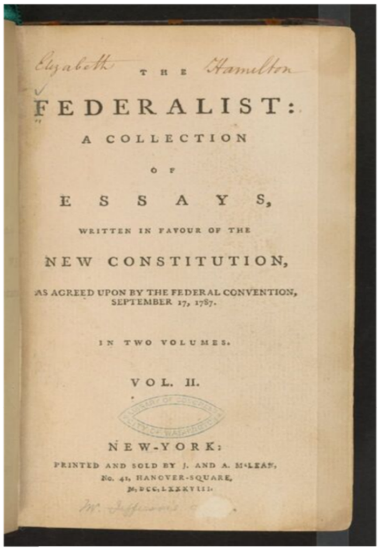
THE FEDERALIST: A COLLECTION OF ESSAYS, WRITTEN IN FAVOUR OF THE NEW CONSTITUTION, AS AGREED UPON BY THE FEDERAL CONVENTION, SEPTEMBER 17, 1787. IN TWO VOLUMES. VOL. II. NEW-YORK: PRINTED AND SOLD BY J. AND A. M'LEAN, No. 41, HANOVER-SQUARE, M,DCC,LXXXVIII.

Figure 1 - Federalist Papers - Volume II Title Page

The Federalist Papers were first published anonymously under the pen name Publius in New York newspapers in 1787 and 1788. Later, it was revealed that these 85 essays were written by Alexander Hamilton, James Madison and John Jay. These documents are now considered to be one of the most important sources for interpreting the United States Constitution.[i] Since publication, there has been a debate around who authored which documents and it is generally agreed that 51 were authored by Hamilton, 15 by Madison, 3 by Hamilton and Madison, and 5 by Jay. The remaining 11 documents were believed to be authored by either Hamilton or Madison and both claimed authorship. The more recent discussion over authorship includes a 1944 paper by Douglass Adair which explores the historical theories about authorship. [ii] Later, in 1963, a statistical analysis of the texts was conducted by Frederick Mosteller and David Wallace which provided a framework for how to analyze the essays using more innovative statistical techniques.

Today, modern data science techniques (e.g. text mining and clustering) are available that can utilize the work of Mosteller and Wallace's to develop computer-aided statistical based conclusions about the authorship of the last 11 documents.

Analysis and Models

<u>About the Data</u>

The data for this analysis includes the Library of Congress transcribed text from the 85 essays in the Federalist Papers. Three of the essays were eliminated that were co-written by Hamilton and Madison. These essays would have created challenges when trying to interpret the results since the writings of Hamilton and Madison would be co-mingled in these essays. An additional five essays written by John Jay were removed since the disputed 11 essays are claimed to be written by Hamilton or Madison. Finally, 15 essays from Hamilton (29%) and 4 essays from Madison (27%) were excluded to create a "test" dataset. The "test" datasets were not used for this analysis and will be utilized for future research using supervised learning (e.g. decision trees).

The following table will be a useful reference for the file numbers used throughout this paper. Consecutive file numbers were used in lieu of the essay numbers since it is easier to see the patterns in groups of numbers when performing the clustering analysis.

Author	File Number	Essay Number
Unknown	1-11	49-57,62,63
Alexander Hamilton	12-47 (48-62 removed for testing)	1,6,7-9,11-13,15-17,21-36,59-61,65-70 (71-85 removed for testing)
John Jay	63-67 (removed)	2-5,64 (removed)
James Madison	68-78 (79-82 removed for testing)	10,14,37-45 (46-48,58 removed for testing)
James Madison/Alexander Hamilton	83-85 (removed)	18-20 (removed)

Table 1 - Author and essay number cross-reference

After importing the data, the (FedCorpus) command was executed to view the number of documents. The number of files was 59 and a hidden 'icon' file was discovered in the folder with the text files. This was removed and the count is now 58 as shown in Figure 2.

```
<<TermDocumentMatrix (terms: 7449, documents: 58)>>
Non-/sparse entries: 36189/395853
Sparsity           : 92%
Maximal term length: 19
Weighting          : term frequency (tf)
```

Figure 2- FedCorpus summary

Once the essay text was loaded, the cleaning of the data followed these steps:

1. Remove punctuation, numbers, "stop" words like "and", "the", "or" and whitespace
2. Convert all case to lowercase
3. Remove additional words - like, can, I, also, lot, publius, federalist, nbsp). The characters nbsp are non-breaking space which is created by some software programs to add spaces.
4. After cleaning, the process of lemmatization was applied to normalize the text by comparing variations on words (e.g. sing, sings, singing, singer, etc.). Figure 1 below shows the resulting text example with and without lemmatization.

05.txt

subject continued house representatives new york packet tuesday february author alexander hamilton james madison people state new york shall perhaps reminded current observation annual elections end tyranny begins true often remarked sayings become proverbial generally founded reason less true established often applied cases reason extend need look proof beyond case us reason proverbial observation founded man subject ridicule pretending natural connection subsists sun seasons period within human virtue bear temptations power happily mankind liberty respect confined single point time lies within extremes afford sufficient latitude variations may required various situations circumstances civil society

05.txt (with lemmatization)

subject continu hous repres new york packet tuesday februari author alexand hamilton jame madison peopl state new york shall perhap remind current observ annual elect end tyranni begin true often remark say becom proverbi general found reason less true establish often appli case reason extend need look proof beyond case us reason proverbi observ found man subject ridicul pretend natur connect subsist sun season period within human virtu bear temptat power happili mankind liberti respect confin singl point time lie within extrem afford suffici latitud variat may requir various situat circumst civil societi elect magistr might found expedi instanc actual daili

Figure 3 - Sample of partial text from 05.txt with and without lemmatization

The final step in the data processing is to put the Term Document Matrix in a dataframe. Figure 4 below shows the view after this step and it is concerning that several files are missing any text (e.g. 71.txt and 74.txt)

71.txt	
72.txt	conform plan republican principl independ journal au…
73.txt	power convent form mix govern examin sustain new …
74.txt	

Table 1 - Sample of view showing missing data

However, when the dataframe was exported to a CSV file, the text is included.

71.txt	subject continu incoher object new plan expo
72.txt	conform plan republican principl independ jou
73.txt	power convent form mix govern examin susta
74.txt	general view power confer constitut indepen

Table 2 - Sample of CSV export showing text included

The missing information was investigated and determined to be an issue with the view command in R Studio. During the modeling phase, troubleshooting will continue to ensure the data from all files is included.

Prior to testing different models, it is also useful to understand the types of words and frequency by document as shown in the examples on this page.

	Docs									
Terms	11.txt	20.txt	24.txt	46.txt	47.txt	68.txt	71.txt	74.txt	76.txt	77.txt
author	5	8	10	15	9	1	7	7	24	15
constitut	8	8	6	10	13	3	16	20	24	31
govern	16	16	14	7	14	17	21	13	28	12
may	18	7	20	10	16	16	9	23	25	6
must	12	9	9	5	9	10	4	15	8	6
nation	9	13	14	11	5	4	1	14	2	3
one	9	1	10	11	10	8	13	8	13	5
peopl	42	5	10	4	6	5	6	9	3	6
power	8	13	18	23	9	2	24	40	11	44
state	11	18	34	26	10	6	21	17	71	52

Table 3 - Word sample and count from selected documents

Words occurring more than 100 times are shows in Figure 4.

```
 [1] "appear"     "appoint"    "author"     "bodi"       "case"       "circumst"   "common"     "consid"     "consider"
[10] "constitut"  "convent"    "danger"     "differ"     "either"     "elect"      "equal"      "establish"  "even"
[19] "everi"      "execut"     "exist"      "far"        "first"      "forc"       "form"       "found"      "general"
[28] "govern"     "great"      "interest"   "latter"     "law"        "legisl"     "legislatur" "less"       "liberti"
[37] "made"       "may"        "mean"       "member"     "men"        "might"      "much"       "must"       "nation"
[46] "natur"      "necessari"  "new"        "number"     "object"     "one"        "part"       "peopl"      "person"
[55] "place"      "plan"       "polit"      "power"      "principl"   "proper"     "propos"     "provis"     "public"
[64] "reason"     "regul"      "respect"    "right"      "secur"      "sever"      "shall"      "state"      "subject"
[73] "time"       "two"        "view"       "well"       "without"    "york"       "import"     "particular" "upon"
[82] "citizen"    "countri"    "feder"      "necess"     "requir"     "union"      "unit"       "articl"     "repres"
[91] "senat"      "confeder"   "make"
```

Figure 4 - Words occurring more than 100 times

The frequency of words can also be displayed visually in a "word cloud" with the size of font and color showing words with greater frequency.

Figure 5 - Word cloud for Federalist Papers corpus

Model 1 - Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering (AHC) is a technique which "start with the points as individual clusters and, at each step, merge the closest pair of clusters".[iii] Preliminary AHC modeling can be done to visualize the relationships between the different words as shown in Figure 6. The settings used for this first model were 4 clusters, Euclidean distance and the Ward clustering approach. Other popular methods that could be used for clustering are Single Link, Complete Link, Group Average and Centroid. Other methods will be tested during modeling to determine which provides the best clustering for this data.

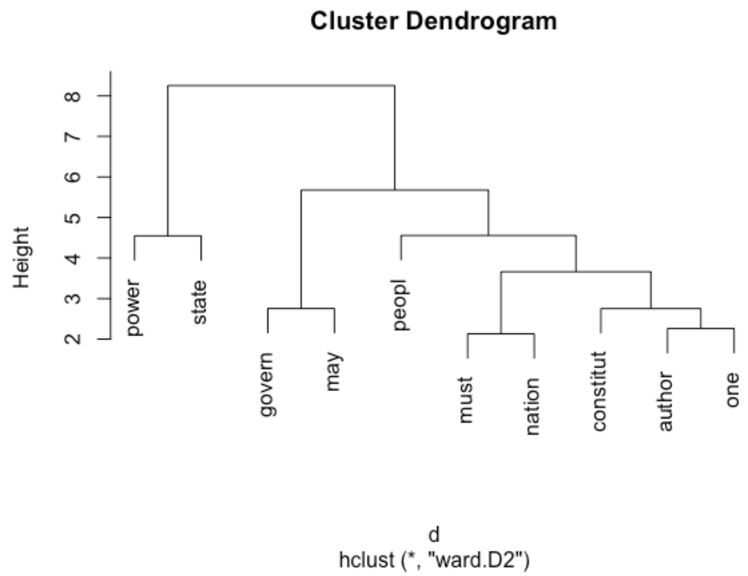

Figure 6 - Hierarchical Clustering Before Normalization

Although most of the text documents are similar in size, normalization was done to ensure that equivalent weighting is given to documents of varying size. The resulting Term Document Matrix and Document Term Matrix examples are shown in Figures 7 and 8.

```
<<TermDocumentMatrix (terms: 4383, documents: 58)>>
Non-/sparse entries: 30785/223429
Sparsity           : 88%
Maximal term length: 18
Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
Sample             :
         Docs
Terms          13.txt      14.txt      17.txt      18.txt       26.txt      31.txt      44.txt      69.txt       70.txt
  armi     0.000000000 0.000000000 0.000000000 0.001809134 0.0183958514 0.016976476 0.000000000 0.000000000 0.0000000000
  claus    0.000000000 0.000000000 0.000000000 0.000000000 0.0062881870 0.000000000 0.044309783 0.000000000 0.0000000000
  elect    0.000000000 0.000000000 0.000000000 0.000000000 0.0008476547 0.000000000 0.000000000 0.000000000 0.0006169080
  militia 0.000000000 0.000000000 0.000000000 0.000000000 0.0052773364 0.065747039 0.000000000 0.000000000 0.0000000000
  presid   0.001823397 0.001671877 0.000000000 0.000000000 0.0000000000 0.000000000 0.022874912 0.000000000 0.0000000000
  repres   0.001335793 0.000000000 0.000000000 0.000000000 0.0007134517 0.001316807 0.000000000 0.003990939 0.0000000000
  senat    0.000000000 0.000000000 0.000000000 0.000000000 0.0000000000 0.000000000 0.034331315 0.000000000 0.0000000000
  taxat    0.000000000 0.000000000 0.000000000 0.003993430 0.0000000000 0.000000000 0.000000000 0.000000000 0.0000000000
  upon     0.002061499 0.005198035 0.002636149 0.003410908 0.0038536886 0.005080497 0.003879295 0.000000000 0.0004006637
  year     0.000000000 0.000000000 0.000000000 0.000000000 0.0016898272 0.001559445 0.001984564 0.001575439 0.0000000000
         Docs
Terms           71.txt
  armi     0.0024756576
  claus    0.0000000000
  elect    0.0010266731
  militia 0.0000000000
  presid   0.0035386772
  repres   0.0008641274
  senat    0.0046587254
  taxat    0.0013661733
  upon     0.0013335882
  year     0.0020467061
```

Figure 7 - Term Document Matrix

```
<<DocumentTermMatrix (documents: 58, terms: 4383)>>
Non-/sparse entries: 30785/223429
Sparsity           : 88%
Maximal term length: 18
Weighting          : term frequency - inverse document frequency (normalized) (tf-idf)
Sample             :
       Terms
Docs          armi       claus        elect    militia      presid       repres      senat       taxat         upon       year
  13.txt 0.000000000 0.000000000 0.0000000000 0.000000000 0.001823397 0.0013357933 0.000000000 0.000000000 0.0020614994 0.000000000
  14.txt 0.000000000 0.000000000 0.0000000000 0.000000000 0.001671877 0.0000000000 0.000000000 0.000000000 0.0051980348 0.000000000
  17.txt 0.000000000 0.000000000 0.0000000000 0.000000000 0.000000000 0.0000000000 0.000000000 0.000000000 0.0026361492 0.000000000
  18.txt 0.001809134 0.000000000 0.0000000000 0.000000000 0.000000000 0.0000000000 0.000000000 0.003993430 0.0034109084 0.000000000
  26.txt 0.018395851 0.006288187 0.0008476547 0.005277336 0.000000000 0.0007134517 0.000000000 0.000000000 0.0038536886 0.001689827
  31.txt 0.016976476 0.000000000 0.0000000000 0.065747039 0.000000000 0.0013168074 0.000000000 0.000000000 0.0050804974 0.001559445
  44.txt 0.000000000 0.044309783 0.0000000000 0.000000000 0.022874912 0.0000000000 0.034331315 0.000000000 0.0038792945 0.001984564
  69.txt 0.000000000 0.000000000 0.0000000000 0.000000000 0.000000000 0.0039909393 0.000000000 0.000000000 0.0000000000 0.001575439
  70.txt 0.000000000 0.000000000 0.0006169080 0.000000000 0.000000000 0.0000000000 0.000000000 0.000000000 0.0004006637 0.000000000
  71.txt 0.002475658 0.000000000 0.0010266731 0.000000000 0.003538677 0.0008641274 0.004658725 0.001366173 0.0013335882 0.002046706
```

Figure 8 - Document Term Matrix

The dendrogram results in Figures 9 and 10 are now more interesting with the clustering of the terms and documents after normalization. The number of words was limited to 10 to make the term document matrix diagram readable (left). The document term matrix (right) includes all of the documents in the corpus.

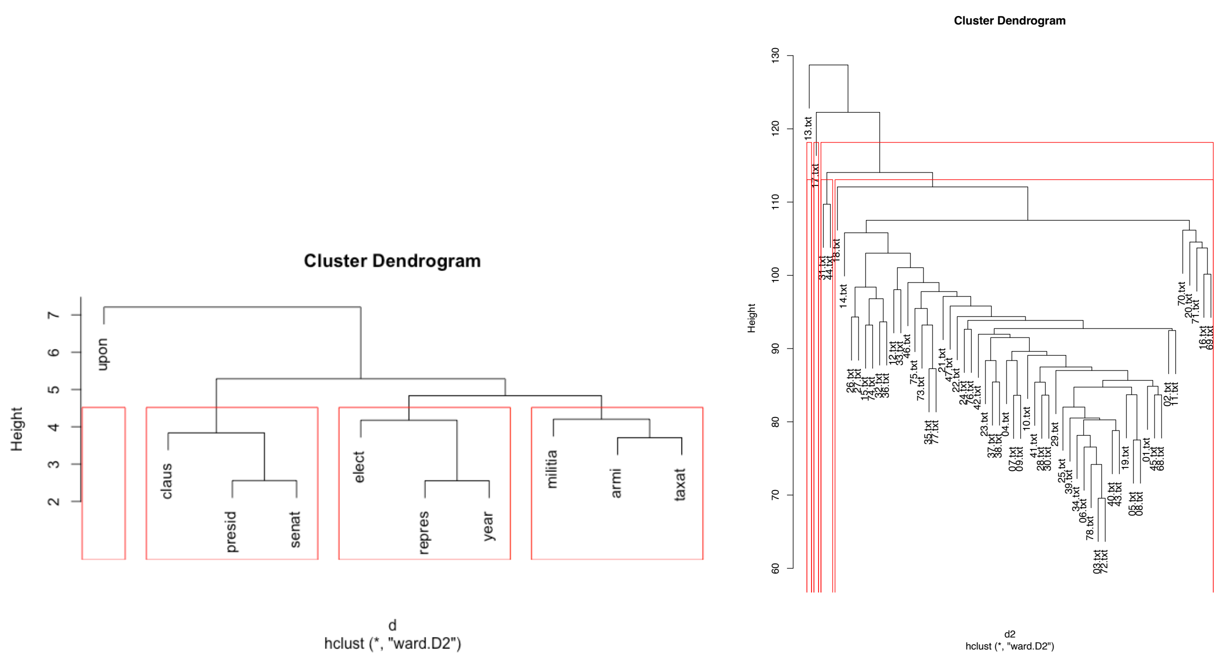

Figure 9 - Clustering of selected terms (left) and documents (right) using AHC (Euclidean/ Ward/4 clusters) after normalization

Several other methods were tested for the document term matrix with cluster sizes from 2-4, different distance measures (manhattan and maximum), and the clustering model was tested with single, complete, average and median. These all delivered similar results as the right graph shown in Figure 9 and observations will be described in more detail in the Results section.

Model 2 - K-Means

The second clustering model that will be used to analyze the Federalist Papers is K-Means. Stanford Assistant Professor Chris Piech defines K-Means as:

> "One of the most popular 'clustering' algorithms. K-means stores centroids that it uses to define clusters. A point is considered to be in a particular cluster if it is closer to that cluster's centroid than any other centroid."[iv]

The K-means algorithm was run first on the data using 3 clusters which generated the results shown in Figure 10. Subsequent tests were run with normalized data and 3 clusters (Figure 11) and normalized data with 2 clusters (Figure 12).

```
01.txt 02.txt 03.txt 04.txt 05.txt 06.txt 07.txt 08.txt 09.txt 10.txt 11.txt 12.txt 13.txt 14.txt 15.txt 16.txt 17.txt 18.txt
     2      2      2      1      1      1      1      1      1      1      1      2      2      2      2      2      2      2
19.txt 20.txt 21.txt 22.txt 23.txt 24.txt 25.txt 26.txt 27.txt 28.txt 29.txt 30.txt 31.txt 32.txt 33.txt 34.txt 35.txt 36.txt
     2      2      2      2      2      3      2      2      2      2      2      2      2      2      2      2      2      2
37.txt 38.txt 39.txt 40.txt 41.txt 42.txt 43.txt 44.txt 45.txt 46.txt 47.txt 68.txt 69.txt 70.txt 71.txt 72.txt 73.txt 74.txt
     2      3      1      1      1      2      2      2      2      3      2      1      2      2      3      3      3      3
75.txt 76.txt 77.txt 78.txt
     3      3      3      3
```

Figure 10 - K-means results with three clusters

```
01.txt 02.txt 03.txt 04.txt 05.txt 06.txt 07.txt 08.txt 09.txt 10.txt 11.txt 12.txt 13.txt 14.txt 15.txt
     1      1      2      2      2      2      2      2      2      2      2      1      1      1      1
16.txt 17.txt 18.txt 19.txt 20.txt 21.txt 22.txt 23.txt 24.txt 25.txt 26.txt 27.txt 28.txt 29.txt 30.txt
     1      1      1      1      1      1      1      1      3      1      1      1      1      1      1
31.txt 32.txt 33.txt 34.txt 35.txt 36.txt 37.txt 38.txt 39.txt 40.txt 41.txt 42.txt 43.txt 44.txt 45.txt
     1      1      1      1      1      1      1      3      1      1      1      1      1      1      1
46.txt 47.txt 68.txt 69.txt 70.txt 71.txt 72.txt 73.txt 74.txt 75.txt 76.txt 77.txt 78.txt
     3      2      2      1      2      3      3      3      3      3      3      3      3
```

Figure 11 - K-means results with three clusters on normalized data

```
01.txt 02.txt 03.txt 04.txt 05.txt 06.txt 07.txt 08.txt 09.txt 10.txt 11.txt 12.txt 13.txt 14.txt 15.txt 16.txt 17.txt 18.txt
     2      1      2      1      1      2      1      1      1      1      1      2      2      2      2      2      2      2
19.txt 20.txt 21.txt 22.txt 23.txt 24.txt 25.txt 26.txt 27.txt 28.txt 29.txt 30.txt 31.txt 32.txt 33.txt 34.txt 35.txt 36.txt
     2      2      2      2      2      2      2      2      2      2      2      2      2      2      2      2      2      2
37.txt 38.txt 39.txt 40.txt 41.txt 42.txt 43.txt 44.txt 45.txt 46.txt 47.txt 68.txt 69.txt 70.txt 71.txt 72.txt 73.txt 74.txt
     2      2      1      2      1      1      1      1      1      1      2      2      2      2      2      1      2      2
75.txt 76.txt 77.txt 78.txt
     2      2      2      2
```

Figure 12 - K-means results with two clusters on normalized data

After running the K-means test for 2 and 3 clusters, another function was used to calculate the optimum number of clusters as shown in Figure 13.

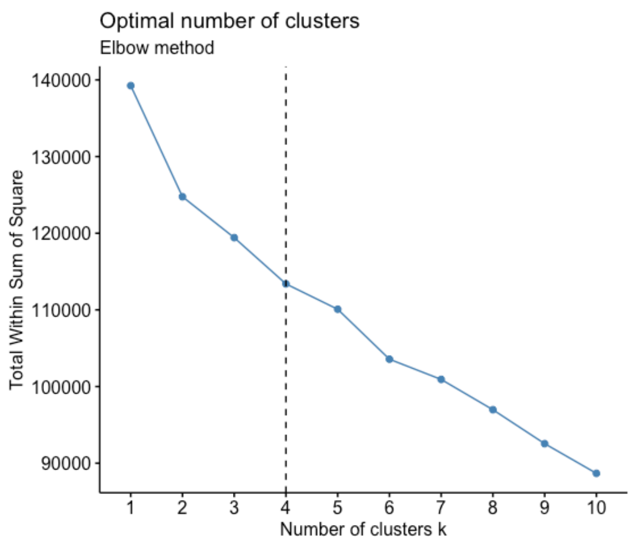

Figure 13 - Optimal number of clusters for K-means

The K-means was run again with 4 clusters and the following visualization shows the grouping of the documents into each cluster.

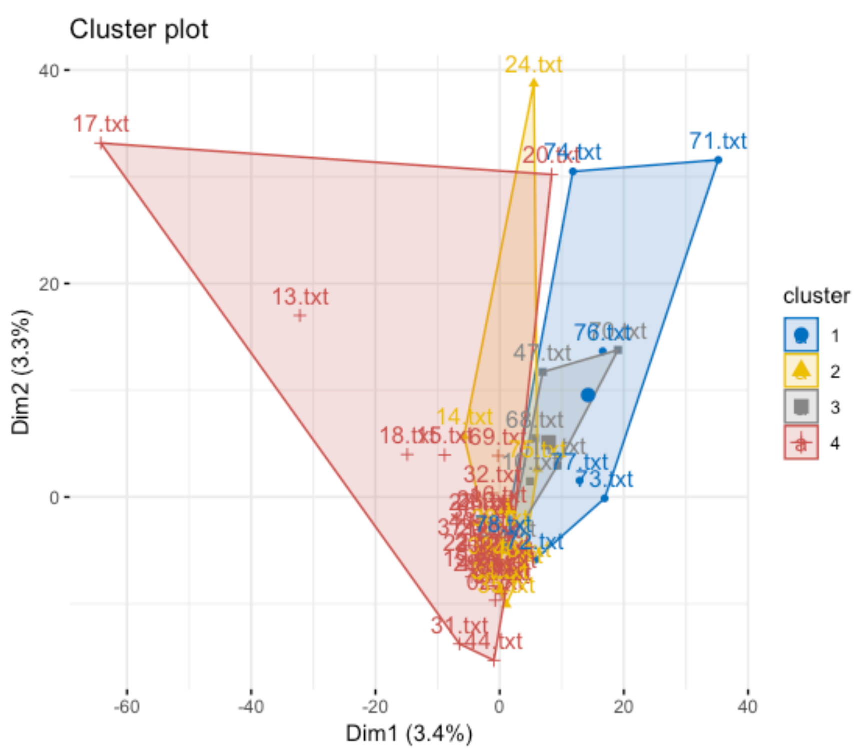

Figure 14 - Document clustering with K-means and 4 clusters

The final step in the analysis of this corpus was to perform a cosine similarity test to determine which documents are "similar" to each other. This test was performed and then a graph was generated to help visualize the results.

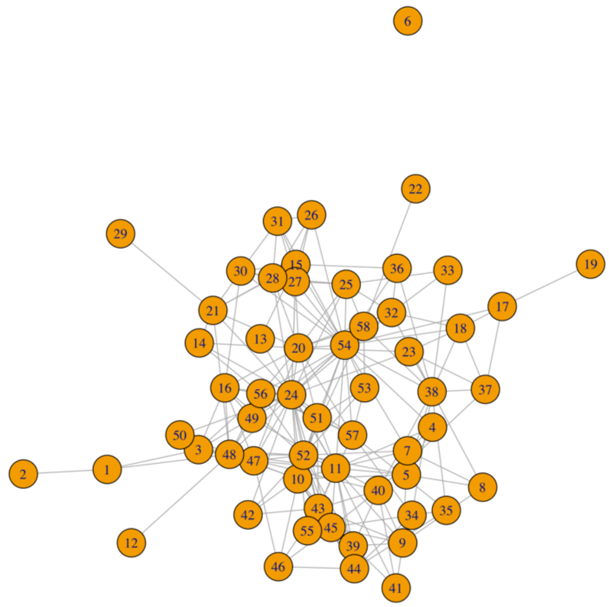

Figure 15 - Graph of cosine similarity between essays

Results

After cleansing the data and loading into a corpus, two different clustering algorithms were performed: Agglomerative Hierarchical Clustering (AHC) and K-Means as described in the analysis section.

The AHC was run initially on the normalized term document matrix data and below is the resulting graph. One interesting feature of this graph is the word "upon". In Frederick Mosteller and David L. Wallace's 1963 paper, Inference in an Authorship Problem, they commented, "The single discriminator we have ever discovered is *upon*, whose rate is about 3 per thousand for Hamilton and about 1/6 per thousand for Madison." This word was shown to be differentiated in this small sample of words but unfortunately the algorithm did not appear to utilize this to differentiate clearly between the document clusters as described later in this section.

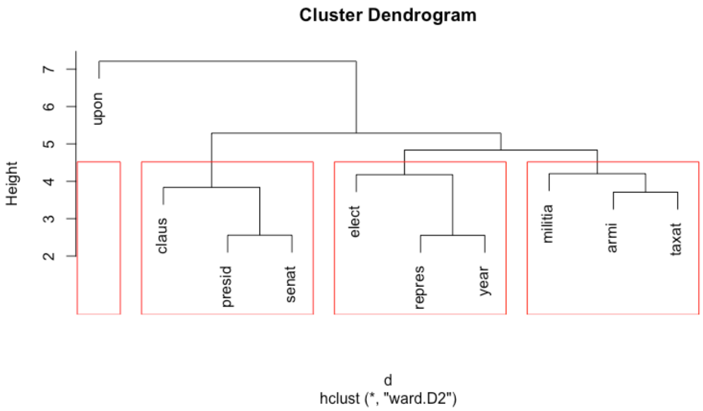

Figure 16 - AHC for a sample of words (Euclidean, Ward, 4 clusters)

The AHC was then run on the normalized document term matrix data and grouped 53 of 58 essays in one large cluster. Unfortunately, this did not provide any assistance in determining the authorship of the disputed papers even after trying different clusters (2,3,4) and different distance measures as described in the analysis section.

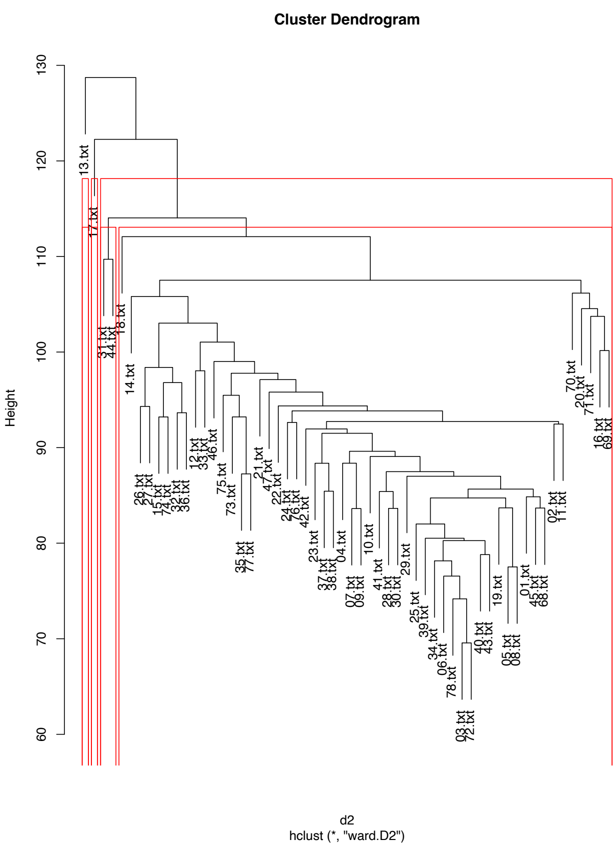

Figure 17 - AHC for 58 documents (Euclidean, Ward, 4 clusters)

The second model used was K-Means and this provided better results and did a better job of dividing the documents when using normalized data. Table 4 shows the grouping by author for each of the 3 clusters. K-Means did a reasonable job of grouping these with 82% of Disputed in Cluster 2, 91% of Hamilton in Cluster 1, and 73% of Madison in Cluster 3. Figure 16 also provides a visual example of how the three clusters were grouped with Clusters 1 and 3 overlapping Cluster 2.

Note: The optimum calculated number of clusters shown in Figure 13 was 4. However, 3 clusters will be used in this section since 4 clusters did not provide any additional clarity and 3 clusters fits better with the logical groupings.

	Cluster 1	Cluster 2	Cluster 3
Disputed (1-11)	2	9	0
Hamilton (12-47)	32	0	3
Madison (68-78)	1	2	8

Table 4 - Comparison of K-means for clusters and authors

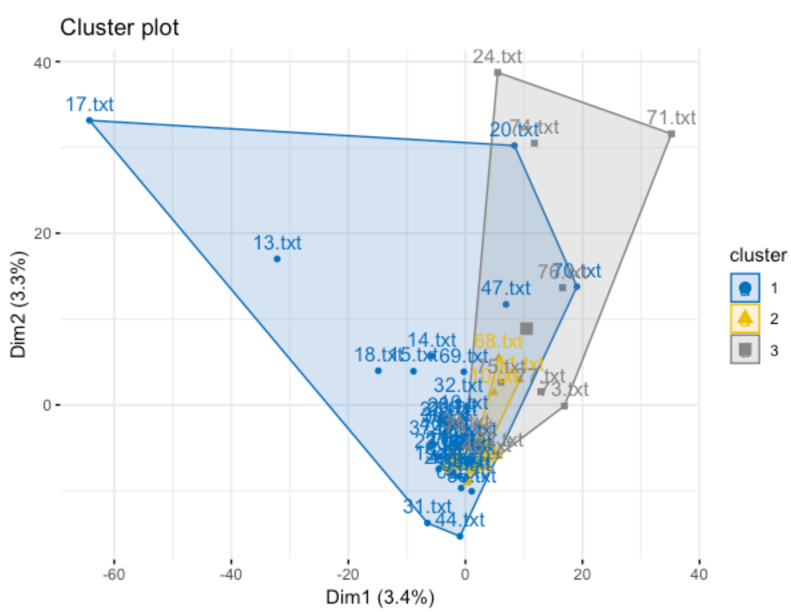

Figure 18 - Graph of K-Means for 3 clusters

However, this did not help to answer the question of whether Hamilton or Madison authored the disputed papers, so the K-Means was also run with k=2 to force these into one of two groups. Nine of eleven disputed papers were grouped with majority Hamilton papers (yellow in Figure 17 below) but these were very close to an "edge" of Cluster 1 shown in blue.

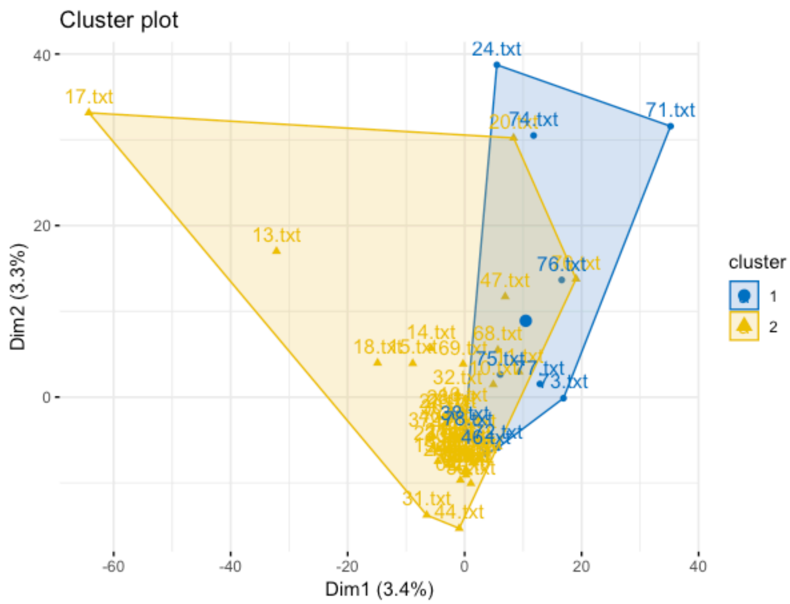

Figure 19 - Graph of K-means for 2 clusters

Another result of the analysis was the output of the cosine similarity. The table below shows the disputed documents in each column and the most familiar document and % match in the rows. Nine of eleven (82%) of the disputed documents have the closest match with another disputed document (1-11). This could be an indication that the disputed documents appear to be a separate group from the Hamilton or Madison documents.

	01.txt	02.txt	03.txt	04.txt	05.txt	06.txt	07.txt	08.txt	09.txt	10.txt	11.txt
Most similar	03.txt	01.txt	68.txt	05.txt	08.txt	40.txt	09.txt	05.txt	07.txt	11.txt	10.txt
%	.24	.21	.30	.26	.32	.14	.31	.32	.31	.29	.29

Table 2 - Cosine similarity comparison of disputed documents with most similar other document

Conclusion

Since the Federalist Papers were first published in the late 18th century, there has been an on-going debate about the authorship of some essays. In the 1963 paper from Mosteller and Wallace, they conclude that Madison was the author of all of the disputed papers (essays 49-57,62,63 or files 1-11 in this analysis) except for possibly one (essay 55, file 7).

The results of the analysis in this paper do not agree with Mosteller/Wallace's conclusion that Madison was the author of the disputed essays. There are three discoveries that point to a different theory:

1. The clustering of the majority of the disputed documents (82%) did not group with either Madison or Hamilton when the option was given for a third group (cluster).
2. When forced to group the essays in either Madison or Hamilton, the majority of the disputed documents went to Hamilton, but these were on the "edge" of the boundary with Madison and there was significant overlap between the clusters.
3. Looking at the similarity of the essays revealed that the 11 disputed documents were most similar to other disputed documents. This was the case in 9 of 11 essays or 82%.

Therefore, the conclusion from this analysis is that neither Madison nor Hamilton wrote the disputed papers independently. The evidence points to a new theory where they collaborated on the disputed essays. Although controversial, this conclusion is also supported by the 2012 paper by Joseph Rudman, "The Twelve Disputed 'Federalist' Papers: A Case for Collaboration."[v] In this paper, Rudman makes the case for the flaws in Mosteller/Wallace's work which utilized outside texts, relied on manual human counts and did not have the benefit of modern data analytics methods and tools.

To further validate the theory that Madison and Hamilton collaborated on the disputed essays, additional analysis could be done by including the Madison/Hamilton collaboration essays (18-20) with the disputed essays to examine how closely these compare with the disputed texts. The collaboration theory will be explored further in a future paper using other data analytics techniques with the goal of providing new insights to the 232-year debate about who authored the disputed Federalist Papers.

[i] Hamilton, A., Madison, J., Jay, J., Chase, S., Hamilton, E. S., Church, A. S. [...] John Davis Batchelder Collection. (1788) *The federalist: a collection of essays, written in favour of the new Constitution, as agreed upon by the Federal Convention,: in two volumes*. [New-York: Printed and sold by J. and A. M'Lean ..., MDCCLXXXVIII] [Pdf] Retrieved from the Library of Congress, https://www.loc.gov/item/09021562/.

[ii] Adair, Douglass. "The Authorship of the Disputed Federalist Papers." *The William and Mary Quarterly* 1, no. 2 (1944): 98-122. doi:10.2307/1921883.

[iii] Tan, Pang-Ning, et al. *Introduction to Data Mining*. Pearson Education, Inc., 2019, p.554.

[iv] Piech, Chris. “K Means .” *CS221*, 2013, stanford.edu/~cpiech/cs221/handouts/kmeans.html.

[v] Rudman, Joseph. “The Twelve Disputed 'Federalist' Papers: A Case for Collaboration.” *Digital Humanities 2012*, 2012, www.dh2012.uni-hamburg.de/conference/programme/abstracts/the-twelve-disputed-federalist-papers-a-case-for-collaboration.1.html.