POLI 175 Challenge: Hiliary Cliton Email Text Analysis

Grant Kim

Ayan Gupta

6 June 2020

Executive Summary

The first exploratory analysis on the entire trigram dataset using the Principal Components Analysis resulted in two general findings. First, there are subsets of emails that form unique clusters. Second, the variables “house_select_benghazi,” "re_turkey_armenia," "Huma Abedin" account for substantial variation in the entire dataset. We further delve into the data by utilizing topic models with 27 topics, conduct a sentiment analysis based on topics, finding that sentiments for some topics did not quite match our preconceived notions given the United States’ relationship with certain countries, and that topics 5,14, and 18 correspond to the three variables identified above in the PCA analysis on the entire dataset. In turn, we do the PCA analysis for each topic models 5, 14, and 18, and find three essential findings. First, there are identifiable clusters of emails that discuss the "Turkey-Armenia" relationship with former prime minister of Turkey Ahmet Davutoglu involved. Second, emails about Lona Valmoro and Huma Abedin can be sub-categorized and viewed separately. Lastly, not only the words “redactions", "benghazi", "sensitive" are closely related but also they serve to cluster some of the emails that were also identifiable as clusters in the entire PCA analysis, meaning that they are distinct from other emails and must be further studied.

Principal Components Analysis

We first use the Principal Components Analysis (PCA) to find a low dimensional representation of the data that captures most variation in the original data with too many features/words. For PCA, the trigram’s principal components explain variation in the data better than that of the unigram. The first ten principal components (PCs) suffice to explain most variation in the data for the trigram (Figure 1), while more than 20 PCs are needed for the unigram (Figure 2).

Unfortunately, the PCA biplot is hard to interpret because the numbers of variables and rows in the data are too large. However, analyzing the biplot (Figure 3), PCA Score Vector Graph (Figure 4), and the sorted PC score table (Figure 5) all together, shows that there are certain identifiable clusters of emails based on PC1 and PC2 that explains around 70% of the variation: email 2348, 43, 215, 98 score extremely low for PC1, email 98, 109, 176, 107, 182, 117, 89, 87 score medium for PC1 and high for PC2, email 5789 and 5490 score extremely score medium for PC1 and low for PC2. While, the majority of the emails are located on the right side of the other plot, with high values for PC1 and medium for PC2. These clusters may be interesting to take a closer look.

Also, we use the loading scores to determine the variables that have the largest influence for the first ten PCs (Figure 6). For PC1 and PC2, the variables with the largest influence are almost identical, interestingly, with words that are related to one of the widely discussed topics regarding Hillary Clinton's emails "house_select_benghazi" and "agreement_on_sensitive." Furthermore, in other PCs which accounts for 30% of the variation, "Huma Abedin," a political figure who was investigated by the FBI in regards to Clinton's emails (Zapotosky 2018), "re_turkey_armenia," a refugee related topic in regards to Clinton's emails (BBC 2016). Conclusively, these different clusters of emails and words, can be a starting point for investigators. Next, we move onto the topic model for further analysis, as the biplot analysis for the PCA model was hard to interpret.

Structural Topic Modeling

We wanted to get a better sense of the various topics within the emails so we used topic modeling to split the emails into distinct categories. We implemented structural topic modeling using the stm package in R. A key decision when using this package is determining how many categories we want to split the emails into (finding the right K value). So we ran several different iterations where K = 5, 15, 20, 25, 30, and 40. After running these iterations we made the observation that very small values of K did not give us specific enough categories to make meaningful observations, while values of K that were too large gave us redundant categories that also did not add useful information. We believe that K = 30 was just a bit too large as we started to see redundant categories in this iterations while K = 25 was a little too small as some informative categories that were in the K = 30 iteration were left out so we ran stm for K values between 25 and 30 and settled on K = 27 as being the optimal value for K in this dataset.

Sentiment Analysis for All 27 Topics

We now had 27 distinct topics as created by R, but we needed to assess these topics qualitatively so the similar aspects of each topic can be defined in a way that makes analysis useful. Figure 7 shows how R divided the emails into topics and the most distinct and frequent words for each topic. Using this data we read through each topic and determined a specific label that would fit most of the emails in a category given the top words as reported by R. In Figure 8 you can see how we have defined each topic and what we considered the “topic defining words” that led us to define the topic in that way. After creating 27 distinct topic labels for all the topics we wanted to perform sentiment analysis to determine how positively or negatively these topics were discussed within emails. Using positive and negative word dictionaries we analyzed the most likely words that would show up in each topic and given this set of words what was the ratio of positive words to all positive and negative words. This percentage of positive words is reported in Figure 8. An interesting observation that can be gleaned from this table is when speaking about the US government and Senate the words found are 75% positive which is surprising given we’d expect more negativity when speaking about the other party. Another interesting observation was that we see 16% positivity for emails regarding North Korea and 33% positivity for emails about the embassy in Kabul, but 62.5% positivity for emails about Pakistan, Iran, and Afghanistan. Given the tension between the United States and these countries, it is puzzling why one topic is so positive while the others are quite negative. As expected emails pertaining to the situation in Benghazi are relatively negative on average as well. For the most part, the sensitivity analysis confirmed many of the ideas we had about Clinton’s opinion of various topics but did have some interesting outliers that went against our expectations.

PCA based on Topic Models

As previously mentioned, the initial PCA model was hard to interpret, but resulted in intriguing words: “house_select_benghazi”, “Huma Abedin,” and “re_turkey_armenia.” Interestingly, these keywords were each labeled as topic labels for topic models 5, 14, and 18. In turn, we do a further PCA analysis on these topic models only. First, for biplot interpretability, we significantly reduce the number of emails and words. For each email, we choose the highest theta (topic weight) as the corresponding topic for that email. Then, we only include emails that have a theta value higher than .5 or 50%. For words, instead of using the most probable words defined by theta which contains uninteresting words such as “subject”, we arbitrarily chose words that are more defining qualitatively from the topic model output.

For topic 5 that relates to Armenia-Turkey, we chose Armenia, Turkey, davutoglu, kaidanow, and huma, as variables of interest. As seen in the biplot in figure 9, most of the variation in the data is explained by PC 1, which represents words davutoglu, turkey, and armenia strongly. PC 2 minimally explains the variation in the data (figure 10). Since Ahmet Davutoglu is the former Prime Minister of Turkey, the two words seem likely to be related. However, the word Armenia is also closely related to PC1, meaning that emails with higher values of PC1 are likely to be a discussion related to the Turkey-Armenia relationship.

For topic model 14 that relates to Huma Abedin, we use the words “huma,” “Abedin,” “lona,” and “valmoro.” The biplot (Figure 11) for topic model 14, tells us that PC1 is strongly related to Huma Abedin and PC2 is strongly related to Lona Valmoro, a senior advisor to Clinton during her terms in the Senate and her campaign for presidential election (George Washington University). An interesting insight is that although the topic model identified these terms as similar, the PCA tells us that these emails can be subcategorized; Emails with higher values of PC1 are related to Huma Abedin and emails with higher values of PC2 are related to Lona Valmoro.

For topic 18, we choose words: "benghazi," "libya," "sensitive," "redactions," "aqim," "juwali," and "qaddafi.” As seen in Figure 13, the biplot for topic 18 shows that words redactions, sensitive, and benghazi are strongly related with PC1 while other words are related with PC2. Most of the variations are explained by the first two PCs (Figure 14). It is essential to notice that there is a distinguishable cluster of emails as identified strongly by PC1 on the right side of the biplot. Using the sorted PC Score table (Figure 15), we find that these are emails 109, 107, 176, 182, 117, and 87. Surprisingly, these emails are also identified as unique clusters by the PCA analysis we did on the entire dataset, confirming that these emails are essential to the Hillary Clinton email analysis.

Appendix

Figure 1: Scree Plot Trigram PCA

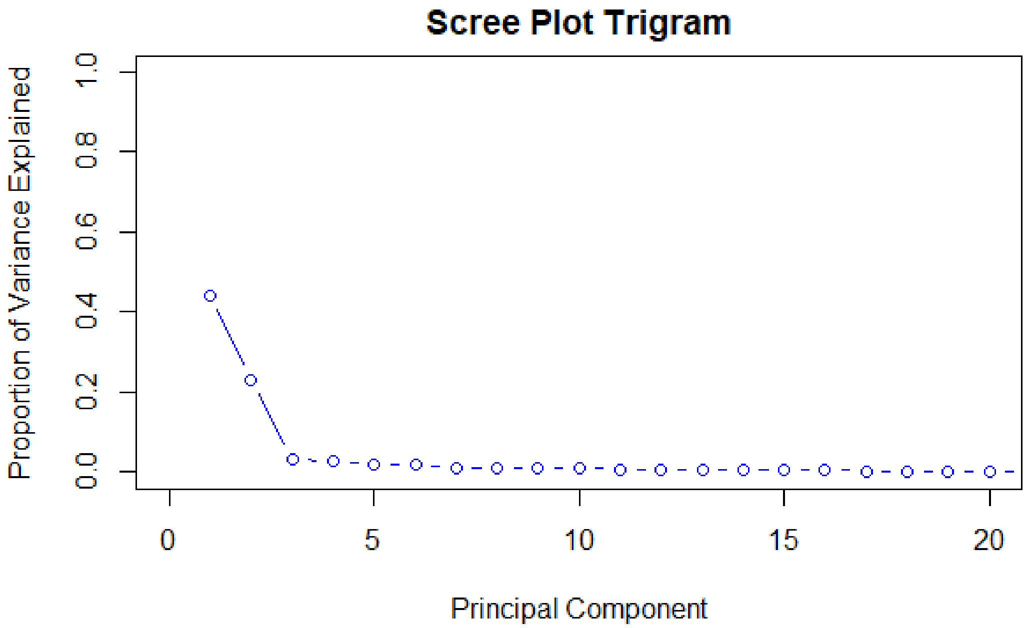

Figure 2: Scree Plot Unigram

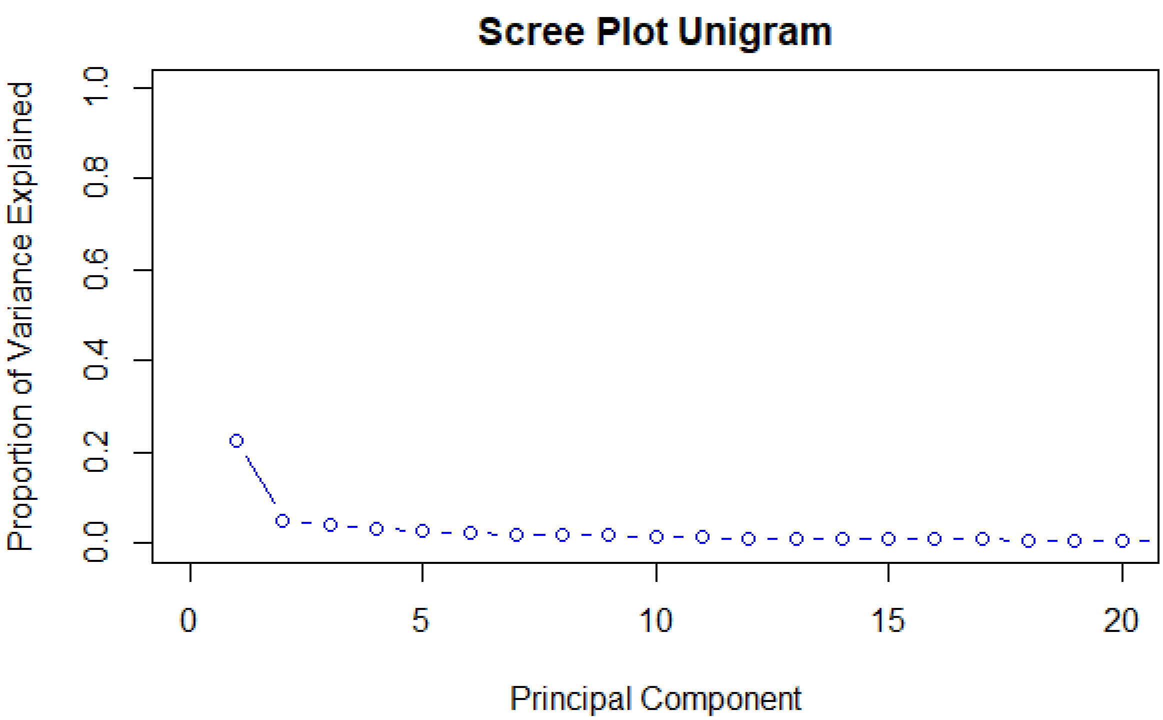

Figure 3: Biplot PCA Trigram

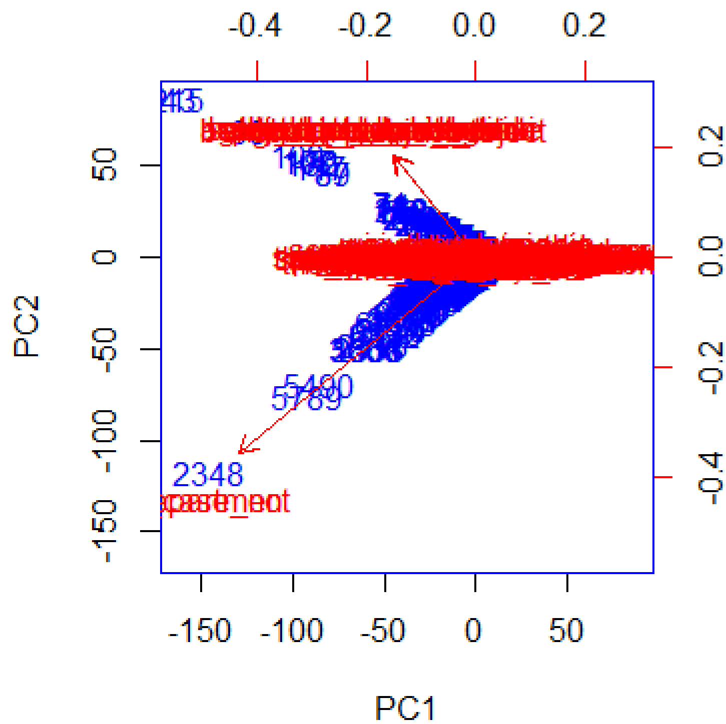

Figure 4: PCA Graph

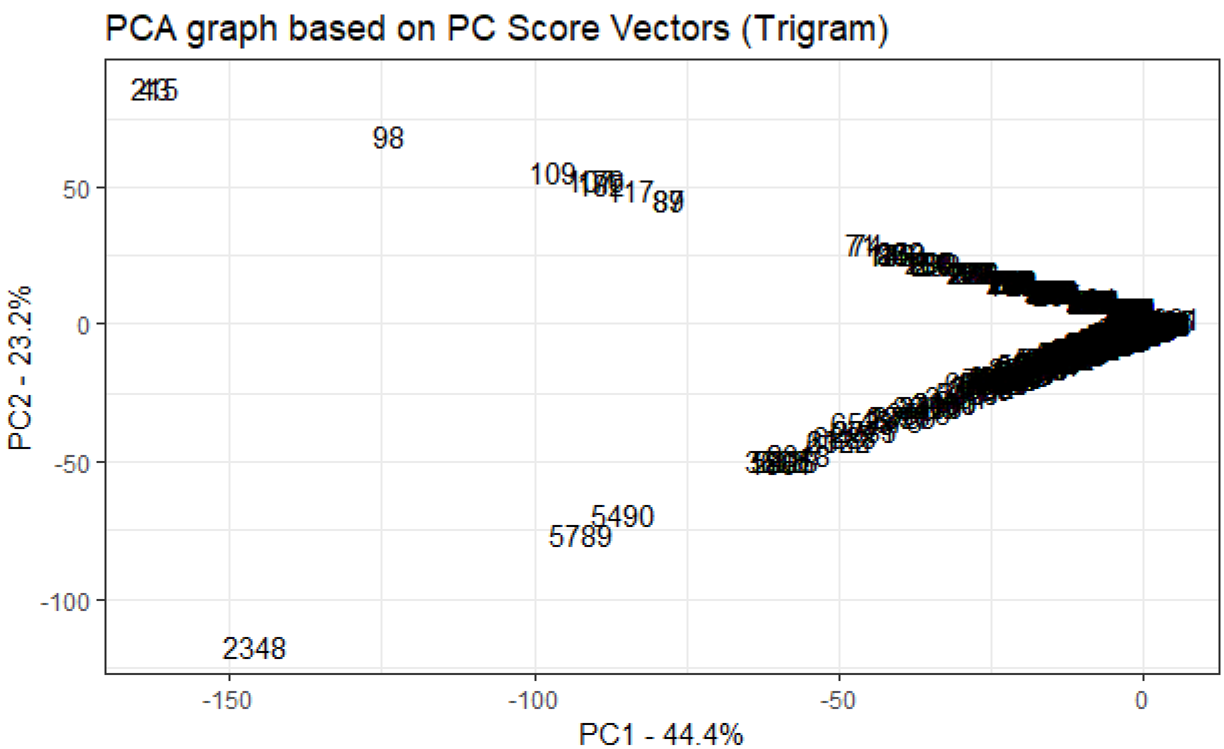

Figure 5.1: Sorted PC Score Table

	Sample <fctr>	X <dbl>	Y <dbl>
215	215	-162.10082	86.39957
43	43	-162.10047	86.40429
2348	2348	-145.72648	-117.11329
98	98	-123.78613	69.30068
109	109	-96.84132	55.96061
5789	5789	-92.40581	-76.29229

6 rows

Figure 5.2: Sorted PC Score Table

	Sample <fctr>	X <dbl>	Y <dbl>
2348	2348	-145.72648	-117.11329
5789	5789	-92.40581	-76.29229
5490	5490	-85.53847	-69.31127
1861	1861	-59.65608	-49.62395
3606	3606	-60.20695	-49.41768
5806	5806	-59.24784	-49.32371
2338	2338	-58.65339	-49.20767
2318	2318	-56.48891	-47.42540
2332	2332	-49.99918	-42.08106
6123	6123	-50.03721	-42.08013

1-10 of 10 rows

Figure 6: Most Influential Words Based on Loading Scores

PC1

[1] "u_s_department" "state_case_no" "house_select_benghazi"
[4] "state_dept_produced" "dept_produced_to" "subject_to_agreement"
[7] "agreement_on_sensitive" "select_benghazi_comm" "produced_to_house"
[10] "to_agreement_on"

PC2

[1] "state_case_no" "u_s_department" "house_select_benghazi"
[4] "state_dept_produced" "dept_produced_to" "subject_to_agreement"
[7] "agreement_on_sensitive" "select_benghazi_comm" "produced_to_house"
[10] "to_agreement_on"

PC3

```
 [1] "hdr22_clintonemail_com" "abedinh_state_gov"      "state_gov_to"
 [4] "gov_to_h"               "abedin_huma_abedinh"    "huma_abedinh_state"
 [7] "h_hdr22_clintonemail"   "clintonemail_com_to"    "to_h_sent"
[10] "turkey_armenia_text"
```

PC4

```
 [1] "mills_cheryl_d"    "millscd_state_gov" "cheryl_d_millscd"  "d_millscd_state"
 [5] "to_mills_cheryl"   "state_gov_to"      "gov_to_h"          "cheryl_d_sent"
 [9] "cheryl_d_subject"  "state_gov_sent"
```

PC5

```
 [1] "jiloty_lauren_c"      "turkey_armenia_text"  "subject_re_mashabane" "re_mashabane_call"
 [5] "re_turkey_armenia"    "subject_re_turkey"    "jilotylc_state_gov"   "sent_sun_aug"
 [9] "sun_aug_30"           "lauren_c_jilotylc"
```

PC6

```
 [1] "the_united_states"    "the_u_s"              "the_state_department" "around_the_world"
 [5] "united_states_and"    "the_middle_east"      "jiloty_lauren_c"      "as_well_as"
 [9] "u_s_department"       "state_case_no"
```

PC7

```
 [1] "state_gov_sent"         "to_h_sent"              "turkey_armenia_text"
 [4] "h_hrod17_clintonemail"  "hrod17_clintonemail_com" "clintonemail_com_sent"
 [7] "jiloty_lauren_c"        "re_turkey_armenia"      "subject_re_turkey"
[10] "state_gov_subject"
```

PC8

```
 [1] "sullivan_jacob_j"       "abedinh_state_gov"      "hdr22_clintonemail_com"
 [4] "huma_abedinh_state"     "abedin_huma_abedinh"    "sullivanjj_state_gov"
 [7] "h_hdr22_clintonemail"   "clintonemail_com_to"    "jacob_j_sullivanjj"
[10] "j_sullivanjj_state"
"
```

PC9

```
 [1] "the_u_s"                "state_gov_sent"         "clintonemail_com_sent"
 [4] "h_hrod17_clintonemail"  "hrod17_clintonemail_com" "turkey_armenia_text"
 [7] "huma_abedinh_state"     "abedin_huma_abedinh"    "re_turkey_armenia"
[10] "subject_re_turkey"
```

PC10

```
 [1] "the_u_s"                "the_united_states"      "u_s_embassy"
```

```
 [4] "state_gov_sent"          "huma_abedinh_state"      "abedin_huma_abedinh"
 [7] "abedinh_state_gov"       "clintonemail_com_sent"   "h_hrod17_clintonemail"
[10] "hrod17_clintonemail_com"
```

Figure 7: Topic Outcomes from STM Process

Topic 1 Top Words:
- Highest Prob: clinton, secretary, hillary, state, president, question, department, case, date, video
- FREX: clinton, video, branch, wjc, gore, hillary, question, school, mrs, book
- Lift: quarter, gore, branch, wjc, clinton, teachers, angola, mother, video, mrs
- Score: clinton, quarter, gore, wjc, branch, mrs, school, secretary, hillary, video

Topic 2 Top Words:
- Highest Prob: senate, house, republican, bill, vote, republicans, democrats, health, care, party
- FREX: senate, republicans, boehner, democrats, reid, republican, voters, vote, senators, bill
- Lift: reid, boehner, ohio, baucus, corker, unfavorable, republicans, beck, democrats, senate
- Score: reid, boehner, republican, democrats, republicans, senate, voters, vote, unfavorable, ohio

Topic 3 Top Words:
- Highest Prob: case, de, david, craig, kelly, zelaya, m, unclassified, thomas, sent
- FREX: craig, kelly, zelaya, de, goldman, shannon, brazilian, arturo, sean, thomas
- Lift: janice, los, rio, jacobs, lautenberg, brazilian, shannon, goldman, oas, valenzuela
- Score: los, zelaya, craig, goldman, brazilian, honduras, janice, shannon, rio, kelly

Topic 4 Top Words:
- Highest Prob: pm, secretary, office, meeting, state, w, department, room, en, arrive
- FREX: arrive, route, depart, room, residence, en, office, pm, meeting, briefing
- Lift: outer, arrive, route, residence, depart, mini, camera, dimartino, kitty, room
- Score: outer, depart, arrive, route, residence, pm, en, room, jiloty, office

Topic 5 Top Words:
- Highest Prob: call, sent, huma, subject, abedin, text, re, turkey, sun, sheet
- FREX: armenia, sheet, davutoglu, text, turkey, call, ashton, amy, tina, philip
- Lift: scanlon, kaidanow, armenia, amy, tina, davutoglu, sheet, esther, ashton, sponsor
- Score: kaidanow, davutoglu, armenia, turkey, huma, scanlon, abedin, ashton, sheet, call

Topic 6 Top Words:
- Highest Prob: american, case, state, doc, department, date, unclassified, one, right, party
- FREX: saudi, muslim, american, arabia, movement, religious, islamic, district, anti, america
- Lift: district, muslims, arabia, activists, kingdom, christian, muslim, religion, saudi, movement
- Score: district, jewish, orthodox, muslim, saudi, tea, jews, muslims, arabia, religious

Topic 7 Top Words:
- Highest Prob: state, department, date, unclassified, case, doc, release, part, full, pm
- FREX: unclassified, date, doc, department, state, case, release, full, part, b6
- Lift: delivered, doc, unclassified, date, case, release, department, full, state, part
- Score: delivered, unclassified, state, doc, department, case, date, release, b5, pm

Topic 8 Top Words:

Highest Prob: women, rights, state, human, people, africa, department, case, melanne, date
FREX: women, rights, gay, africa, human, ecumenical, melanne, turkish, patriarchate, verveer
Lift: patriarchate, patriarch, ecumenical, festival, congo, girls, uganda, marriage, gay, gender
Score: ecumenical, women, patriarchate, melanne, verveer, gay, patriarch, verveerms, festival, orthodox

Topic 9 Top Words:

Highest Prob: report, embassy, media, state, p, press, ambassador, update, department, sbu
FREX: sbu, embassy, report, guard, kabul, pa, n, contract, ca, eap
Lift: hussein, eap, guards, sbu, guard, ca, contract, consular, clips, ds
Score: hussein, embassy, sbu, kabul, nea, eap, guards, guard, pa, update

Topic 10 Top Words:

Highest Prob: sent, subject, message, original, re, com, clintonemail, gov, call, state
FREX: thu, lauren, jiloty, jilotylc, clintonemail, original, dec, hrod17, message, list
Lift: mashabane, thru, vermarr, thu, lauren, jilotylc, jiloty, followup, pis, dec
Score: mashabane, jiloty, lauren, clintonemail, jilotylc, original, message, gov, com, dec

Topic 11 Top Words:

Highest Prob: obama, president, mr, white, administration, state, house, department, policy, case
FREX: mr, obama, koch, white, company, campaign, romney, adviser, administration, bush
Lift: heyman, koch, panetta, ft, romney, emanuel, corporate, mr, advisers, charles
Score: obama, heyman, koch, romney, mr, bush, white, administration, emanuel, campaign

Topic 12 Top Words:

Highest Prob: haiti, january, haitian, children, people, un, government, relief, assistance, au
FREX: haitian, haiti, au, port, earthquake, relief, preval, children, donors, reconstruction
Lift: haitians, bellerive, le, meghann, preval, haitian, haiti, au, earthquake, jean
Score: haiti, le, haitian, bellerive, preval, meghann, haitians, earthquake, au, port

Topic 13 Top Words:

Highest Prob: com, www, us, email, message, mail, please, blair, sent, may
FREX: kris, shanghai, mail, cherie, intended, recipient, balderston, www, email, expo
Lift: windrush, cherie, notify, sender, kris, balderston, pavilion, shanghai, expo, recipient
Score: windrush, cherie, www, kris, blair, shanghai, balderston, expo, sender, pavilion

Topic 14 Top Words:

Highest Prob: huma, abedin, sent, subject, gov, state, abedinh, message, original, re
FREX: abedin, huma, valmoro, valmorou, lona, abedinh, schedule, oscar, talk, tomorrow
Lift: tonite, oprah, pdb, valmorou, hanleymr, valmoro, hanley, oscar, flores, lona
Score: huma, abedin, abedinh, oprah, valmorou, lona, valmoro, clintonemail, gov, pdb

Topic 15 Top Words:

Highest Prob: president, government, secretary, state, department, case, date, united, unclassified, doc
FREX: elections, sudan, colombia, delegation, cooperation, interim, council, colombian, coup, summit
Lift: pan, colombia, colombian, uribe, sudan, delegation, coup, parliamentary, hemisphere, latin
Score: pan, colombia, honduras, colombian, uribe, elections, sudan, honduran, election, government

Topic 16 Top Words:

Highest Prob: know, can, think, get, just, good, re, like, see, m
FREX: think, good, hope, ve, ll, happy, going, know, get, great
Lift: mikulski, evergreen, talbott, pir, strobe, preines, excited, wonderful, happy, terrific
Score: mikulski, strobe, pir, preines, talbott, ve, think, evergreen, know, thank

Topic 17 Top Words:

Highest Prob: development, policy, foreign, department, state, global, diplomacy, affairs, international
FREX: programs, development, usaid, pdpa, diplomacy, resources, agency, global, priorities, affairs
Lift: pdpa, audiences, programs, hunger, evaluation, collaboration, strategies, expertise, usaid, agriculture

Score: pdpa, development, diplomacy, usaid, programs, global, agency, food, strategic, resources

Topic 18 Top Words:

Highest Prob: state, benghazi, sensitive, information, dept, agreement, produced, select, waiver, comm
FREX: benghazi, waiver, comm, redactions, libyan, dept, select, produced, sensitive, libya
Lift: aqim, juwali, jcp, gnc, ntc, qaddafi, magariaf, libyan, comm, benghazi
Score: redactions, benghazi, foia, comm, waiver, select, magariaf, libyan, keib, libya

Topic 19 Top Words:

Highest Prob: china, world, countries, economic, can, global, us, new, freedom, chinese
FREX: china, freedom, chinese, internet, economic, climate, berlin, economy, countries, asia
Lift: diseases, emissions, chronic, berlin, burden, disease, beijing, china, pacific, google
Score: diseases, china, chronic, economic, freedom, economy, chinese, global, countries, berlin

Topic 20 Top Words:

Highest Prob: sullivan, jacob, sent, subject, gov, state, message, original, sullivanjj, re
FREX: sullivanjj, sullivan, jacob, feltman, jake, muscatine, lissa, slaughter, anne, harold
Lift: reinesp, hongju, jacobi, feltman, slaughtera, sullivanjj, edits, declassify, sullivann, gis
Score: reinesp, sullivan, jacob, sullivanjj, muscatine, lissa, b5, jake, gov, slaughter

Topic 21 Top Words:

Highest Prob: blair, uk, british, britain, eu, week, former, memo, us, palau
FREX: britain, palau, british, inquiry, blair, chilcot, saddam, sir, memo, bush
Lift: breakthrough, whitehall, chilcot, palau, inquiry, saddam, britain, sarkozy, sir, miliband
Score: breakthrough, chilcot, blair, palau, britain, whitehall, saddam, uk, eu, sir

Topic 22 Top Words:

Highest Prob: afghanistan, pakistan, military, iran, us, security, afghan, nuclear, united, states
FREX: afghanistan, afghan, pakistan, nuclear, taliban, iran, military, mcchrystal, pakistani, diplomats
Lift: shore, cables, afghanistan, mcchrystal, afghan, taliban, pakistani, weapon, nuclear, qaeda
Score: shore, mcchrystal, taliban, pakistan, afghan, afghanistan, nuclear, military, iran, iraq

Topic 23 Top Words:

Highest Prob: news, state, sent, subject, ses, department, ap, reuters, fw, date
FREX: reuters, ap, ses, news, mahogany, o_shift, korea, north, ii, korean
Lift: external, reuters, mahogany, ap, o_shift, ses, korean, tehran, kim, korea
Score: external, ses, reuters, mahogany, o_shift, ap, abedinh, abedin, news, huma

Topic 24 Top Words:

Highest Prob: cheryl, mills, subject, sent, millscd, fw, gov, cc, pm, fyi
FREX: mills, cheryl, millscd, cdm, nora, toiv, crowley, patrick, kennedy, ellen
Lift: laszczych, psa, joanne, katie, mills, ellen, cheryl, millscd, nora, toiv
Score: cheryl, mills, millscd, psa, cdm, toiv, nora, gov, joanne, mchale

Topic 25 Top Words:

Highest Prob: minister, prime, european, europe, germany, berlusconi, bank, bangladesh, melanne
FREX: berlusconi, bangladesh, germany, greece, sri, european, lanka, euro, europe, debt
Lift: berlusconi, grameen, euro, greece, lanka, hasina, sri, bangladesh, italy, debt
Score: grameen, hasina, berlusconi, bangladesh, euro, sri, lanka, melanne, european, greek

Topic 26 Top Words:

Highest Prob: deal, uk, sid, brown, party, cameron, gordon, sbwhoeop, ireland, dup
FREX: cameron, dup, shaun, ireland, brown, sinn, labour, tories, northern, fein
Lift: cameron, unionist, parades, bravo, devolved, belfast, uup, sinn, shaun, dup
Score: dup, devolved, tories, sinn, labour, shaun, cameron, clegg, guardian, fein

Figure 8

Top ic #	Topic Label	Topic defining words	Positive Word %
1	Clinton's family/US Politics	Wjc, clinton, hillary, gore	0.7777778
2	US Politics/Senate	reid, boehner, republicans, democrats	**0.7500000**
3	South America	brazilian, zelaya, rio, los, de, arturo	0.6666667
4	Logistics	arrive, depart, route, meeting	0.6666667
5	Armenia/Turkey	armenia, turkey, davutoglu	**0.6000000**
6	Muslims/Middle East	muslim, islamic, saudi, movement, anti	0.5454545
7	Unclassified emails	unclassified, date, doc, release	1.0000000
8	Women's rights	women, gender, melanne, verveer	0.8000000
9	Kabul Embassy Reports	kabul, embassy, hussein, report	**0.3333333**
10	Hillary Clinton's emails	clintonemail, hrod17, original, message	0.7777778
11	Obama-Romney race	obama, romney, campaign, adviser	0.6250000
12	Haiti Earthquake	haiti, haitian, earthquake, relief	0.5454545
13	Email Errors	notify, sender, recipient, email	0.7500000
14	Huma Abedin/HRC's emails	huma, abedin, lona, valmoro, oscar	**0.7142857**
15	Foreign elections	sudan, colombia, coup, elections	0.4000000
16	Thank you messages	happy, wonderful, good, think	0.8666667
17	Diplomacy/Aid	diplomacy, development, affairs, usaid	0.9285714
18	Benghazi	benghazi, libya, sensitive, redactions	**0.4444444**
19	China/economy	china, chinese, global, economy	0.5333333
20	Emails from Jacob Sullivan	jacob, sullivan, jake, j	0.3333333
21	UK	uk, britain, british, eu	0.7777778
22	Iran, Pakistan, Afghanistan	iran, pakistan, afghanistan, nuclear	**0.6250000**

23	North Korea	north, korea, news, reuters	**0.1666667**
24	Internal White House emails	cheryl, mills, nora, crowley	0.4285714
25	South Asia/Economy	bangladesh, sri, lanka, euro, debt	0.5000000
26	UK elections	uk, labour, tories, cameron	0.7000000
27	Palestine/Israel	israel, palestine, gaza, netanyahu	0.6153846

Figure 9: Biplot for Topic 5

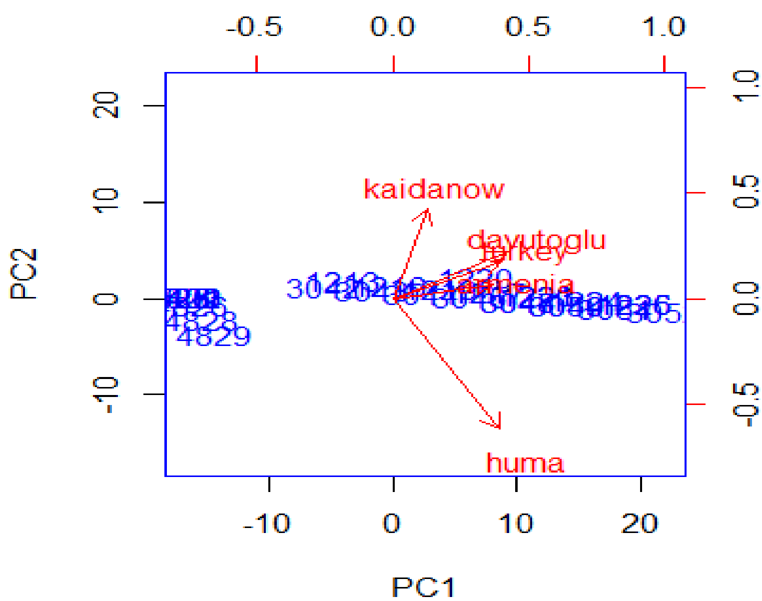

Figure 10: Scree Plot for Topic 5

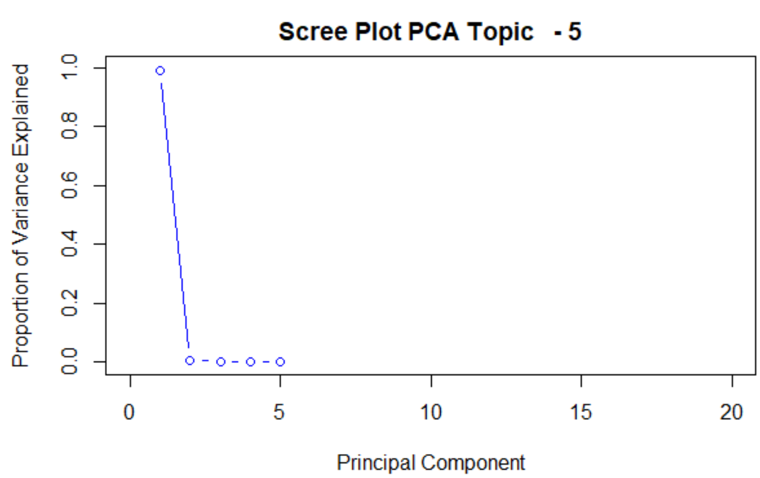

Figure 11: Biplot for Topic 14

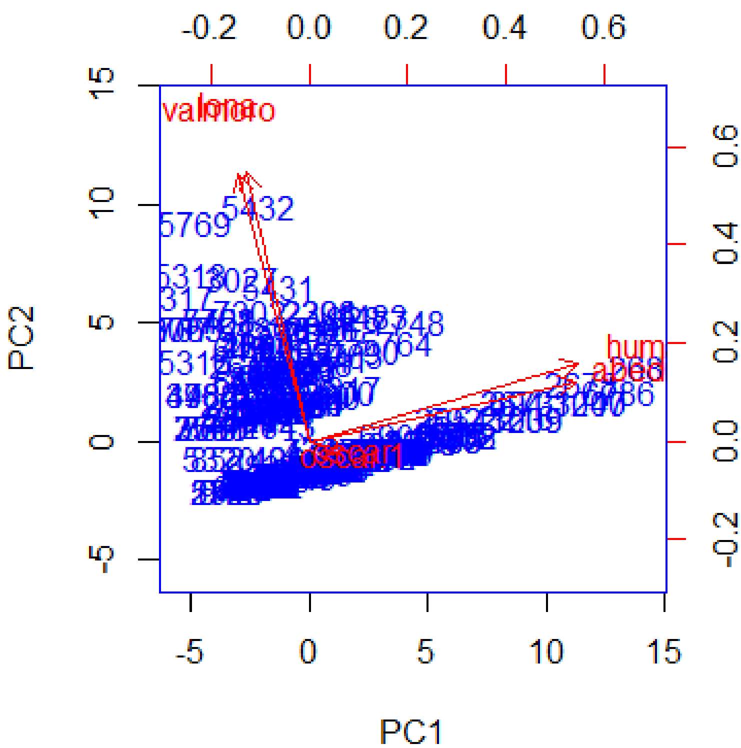

Figure 12: Scree Plot for Topic 14

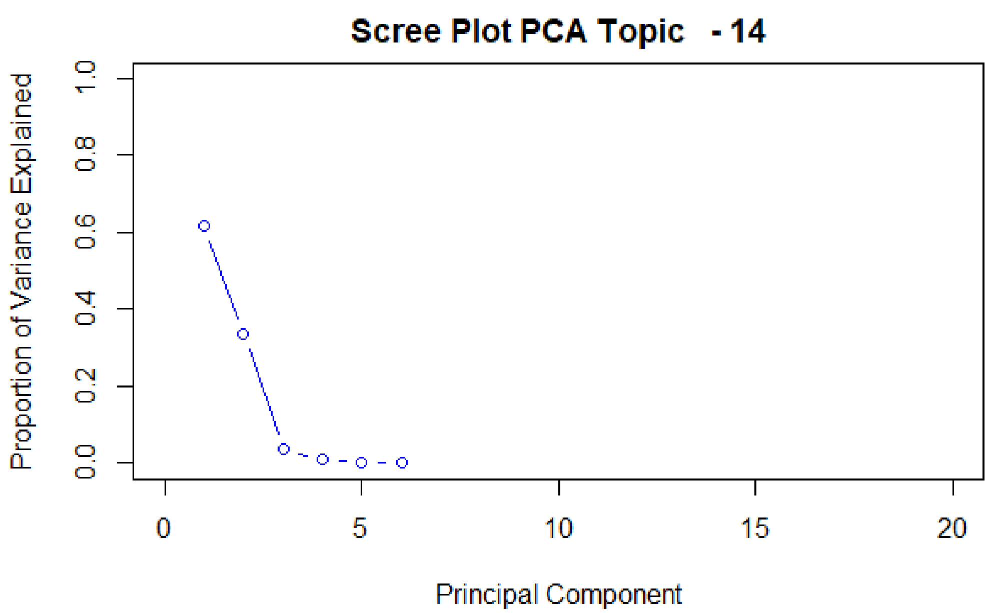

Figure 13: Biplot for Topic 18

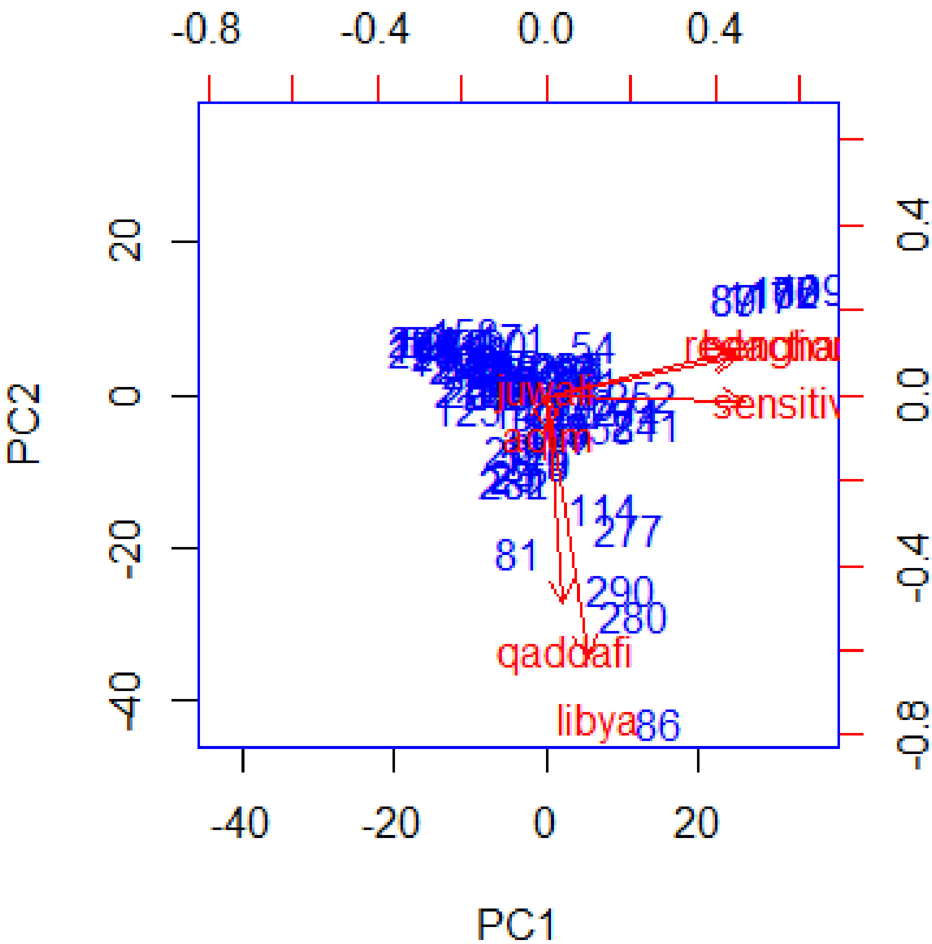

Figure 14: Scree plot for Topic 18

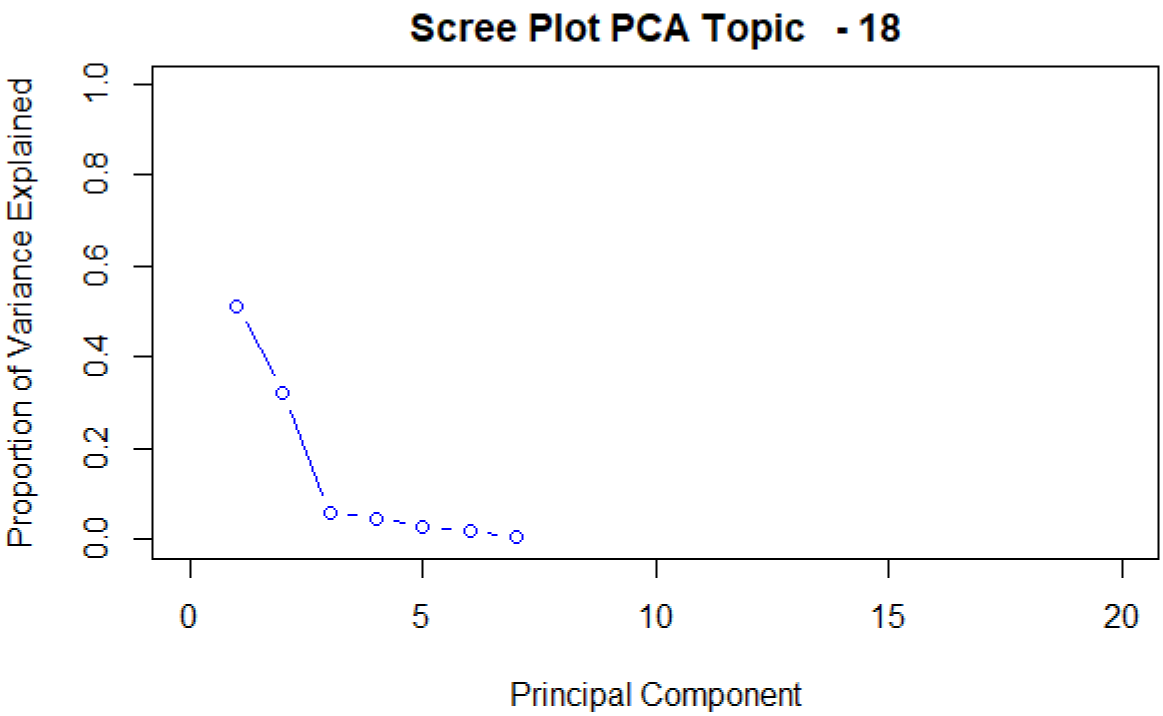

Figure 15: Sorted PC Score Table for Topic 18

	Sample <fctr>	X <dbl>	Y <dbl>
109	109	35.27835	14.08762
107	107	31.84661	13.63663
176	176	31.84661	13.63663
182	182	31.84661	13.63663
117	117	28.41487	13.18565
87	87	24.98313	12.73466

6 rows

Bibliography

"18 Revelations from Wikileaks' Hacked Clinton Emails." *BBC News*, BBC, 27 Oct. 2016, www.bbc.com/news/world-us-canada-37639370.

Zapotosky, Matt. "FBI Interviewed Top Clinton Aide Huma Abedin after Saying Investigation of Her Boss Had Concluded." *The Washington Post*, WP Company, 2 Mar. 2018, www.washingtonpost.com/world/national-security/fbi-interviewed-top-clinton-aide-huma-abedin-after-saying-investigation-of-her-boss-had-concluded/2018/03/02/1be1cd04-1be0-11e8-b2d9-08e748f892c0_story.html.

George Washington University. semesterinwashington.gwu.edu/lona-valmoro.