



Hillary Clinton's Emails

What is it all about?

Project 4 | Natural Language Processing | Vanessa Hu

TIMELINE

January 2009

clintonemail.com is registered, as
Hillary Clinton became the
Secretary of State

May 2015

The State Department began to
release Hillary Clinton's emails in
batches, as the controcy grew

March 2015

NY Times breaks the story of
Clinton's use of a private email server

Summer 2016

FBI conducts investigation on
Clinton's emails during the
presidential election season

Emails in the dataset are from 2009 to 2012

WORKFLOW

Data Acquisition

Focused on only the emails sent by Hillary

SQLite

01

Topic Modeling

CorEx, LSA, Word2Vec
Three Topics

LDA

03

Visualization

Topics, Sentiment,
Circle of Trust, Dates

Tableau

05

spaCy

02

Text processing

REGEX, Lemmatization
TF-IDF

VADER

04

Sentiment Analysis

NLTK VADER
TEXTBLOB

CIRCLE OF TRUST



19%

Huma Abedin

Long-time Aid and Policy Advisor



17%

Cheryl Mills

Chief of Staff and Counselor



16%

Jake Sullivan

Director of Policy Planning

12%

Lauren Jiloty

Personal Assistant

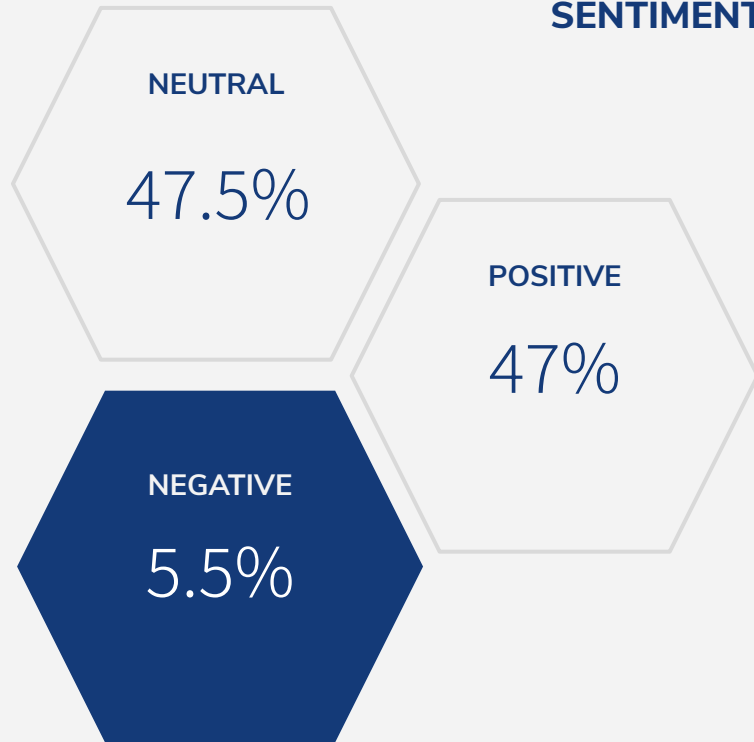
7%

Lona Valmoro

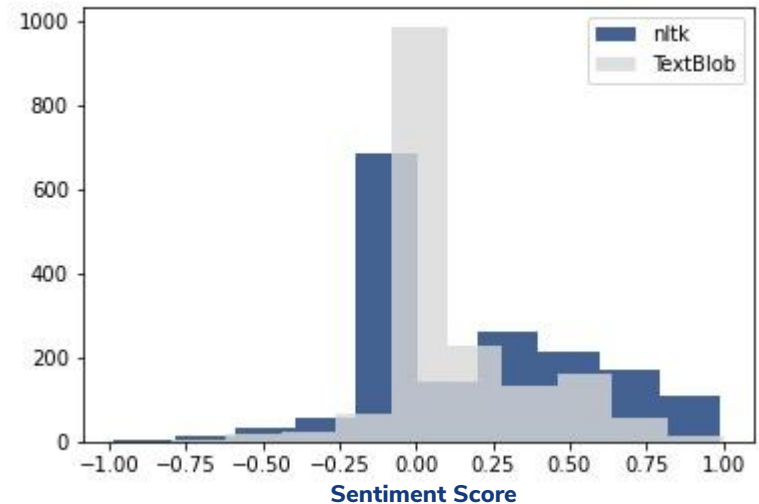
Personal Assistant

70%+

SENTIMENT ANALYSIS



NLTK Vader vs TextBlob



Policy Related

Advisors

24.5%

Email ID 608:

Is the 50k figure State and AID? Kurdistan election results? Can you find out for me how many Nazi prison camps we had in US during WW2 and what we did w the detainees--did we ever try any? Did we deport them after war ended? Re Sudan: how do you pronounce Abyei? What is % and dollar increase in 2010 budget for ops and aid? I don't understand the answer to the prostitution policy question. What's the simplest way to say it?

Action Items

Staff

43%

Email ID 892:

Melanne--I sent you the emails from Burns forwarding both Kay Warren's and Lynn Lusi's emails expressing their concerns about the partnership w Norway that I announced. I don't want to do anything that undermines Heal Africa. Will we be sending \$ to them for training too? Let me know how we can answer their anxiety since I admire all they're doing. H

Schmooze

→ Allies

32.5%

Email ID 1096:

Do you recall Nooni Ali(?), the ten year old Yemeni girl who got herself divorced? I met her at the Glamour awards last year. There was a CNN story last few days about how unhappy she is, still living at home, not attending school and quite angry that her life is not better. Is there any way we can help her? Could we get her to the US for counselling and education?

Tableau Demo

github.com/vanessa920/nlp_hillary_email

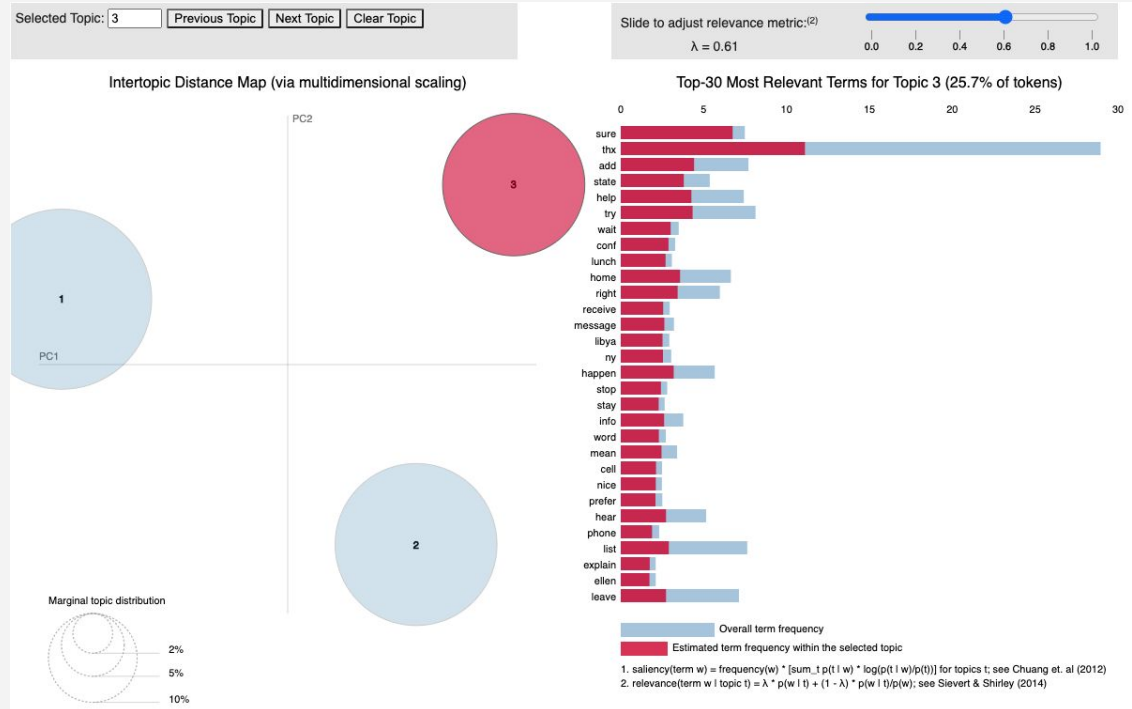


TAKE-AWAYS

- Negative sentiment score could be really valuable when trying to filter information quickly
- When doing topic modeling, if it is hard to summarize the topics through just words, try including the corpus docs to understand the context
- SQL works better than Pandas dataframe when querying from relational database directly
- Tableau visualization works really well with multi-dimensional dataset

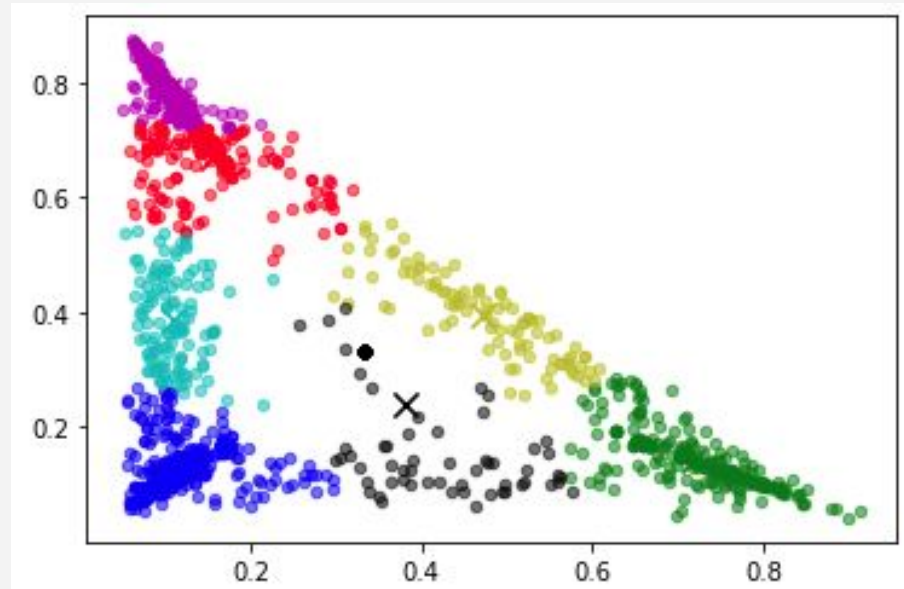
APPENDIX

pyLDA Topic Modeling Chart



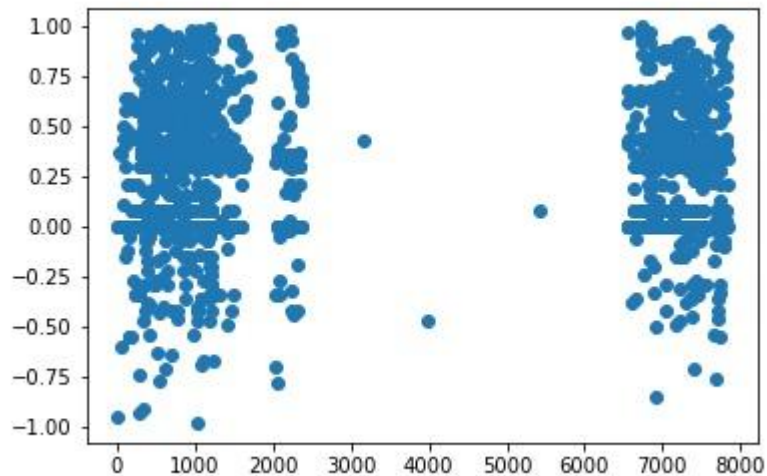
APPENDIX

K-Means Cluster Chart

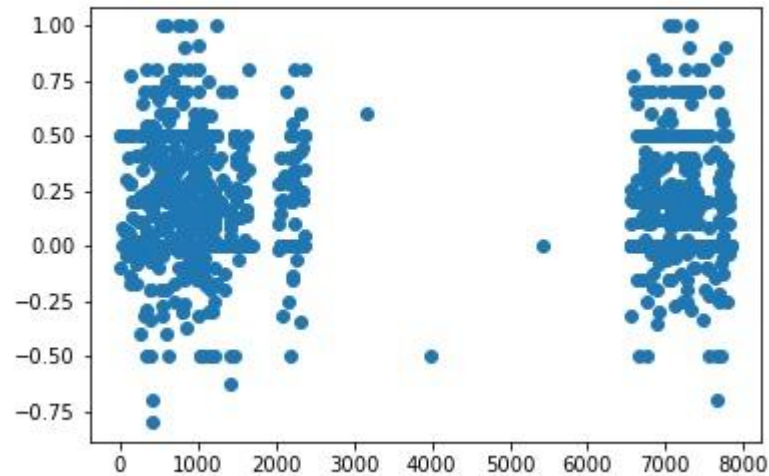


APPENDIX

NLTK Vader and TextBlob Scatter Chart



NLTK Vader Sentiment



TextBlob Sentiment

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

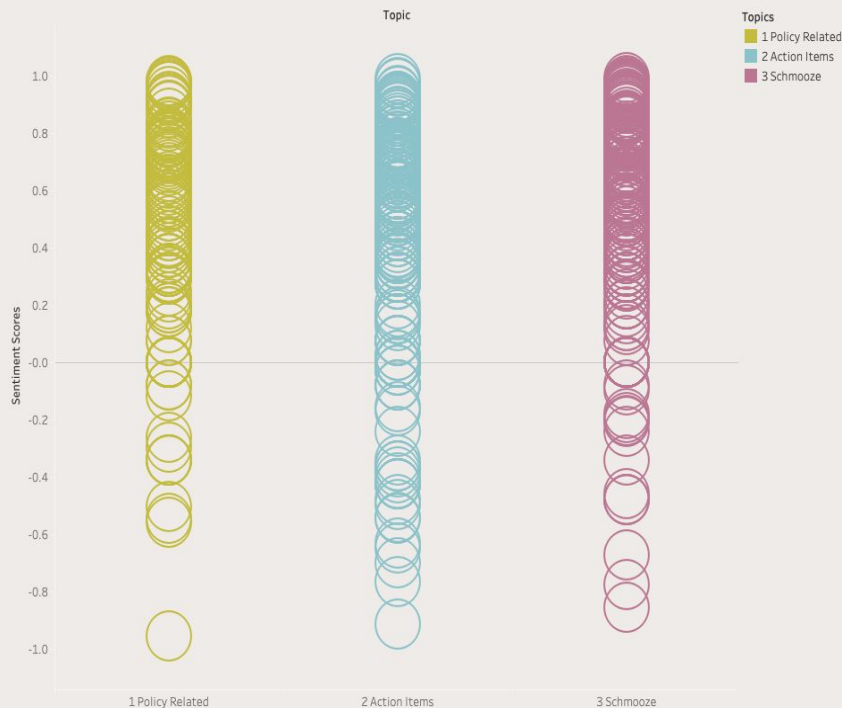
by Vanessa Hu

The emails in this analysis were sent by Hillary Rodman Clinton between 2009 and 2012. This scatter plot includes an overview of sentiment scores and three topics. The

Let's take a look at all the emails sent by Hillary under 3 topics as an overview. In general, she conducts policy related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

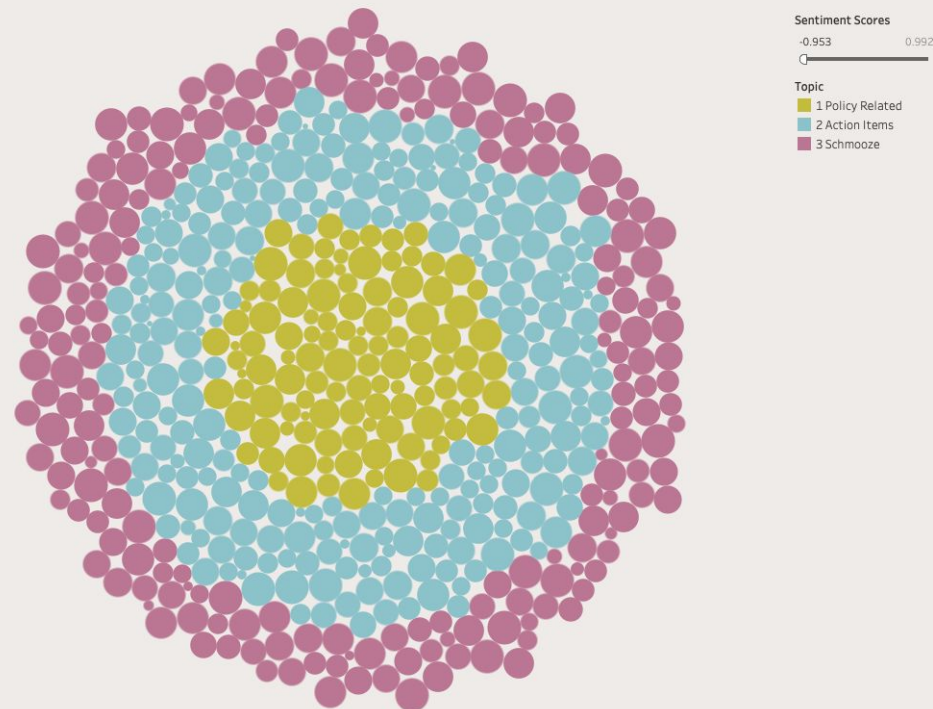
by Vanessa Hu

The emails in this analysis were sent by Hillary Rodman Clinton between 2009 and 2012. This scatter plot includes an overview of sentiment scores and three topics. The

Let's take a look at all the emails sent by Hillary under 3 topics as an overview. In general, she conducts policy related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.



APPENDIX

Tableau Story Line

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

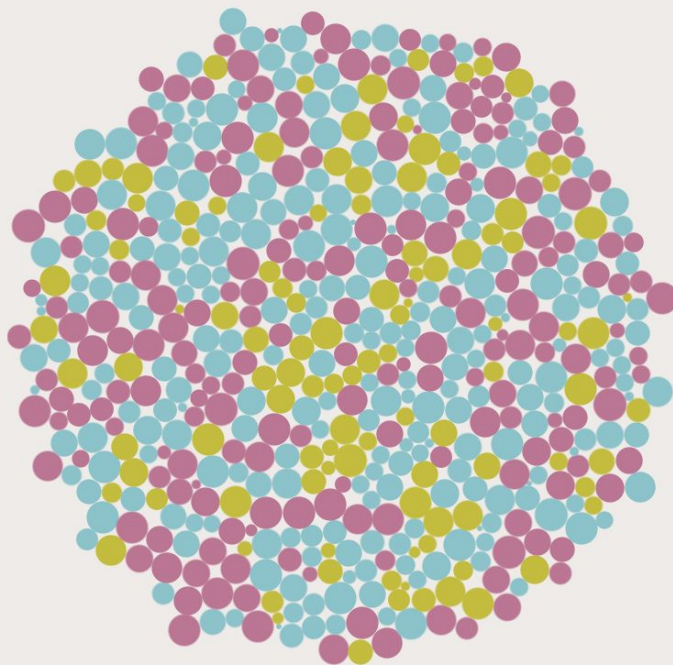
by Vanessa Hu

The emails in this analysis were sent by Hillary Rodman Clinton between 2009 and 2012. This scatter plot includes an overview of sentiment scores and three topics. The

Let's take a look at all the emails sent by Hillary under 3 topics as an overview. In general, she conducts policy related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

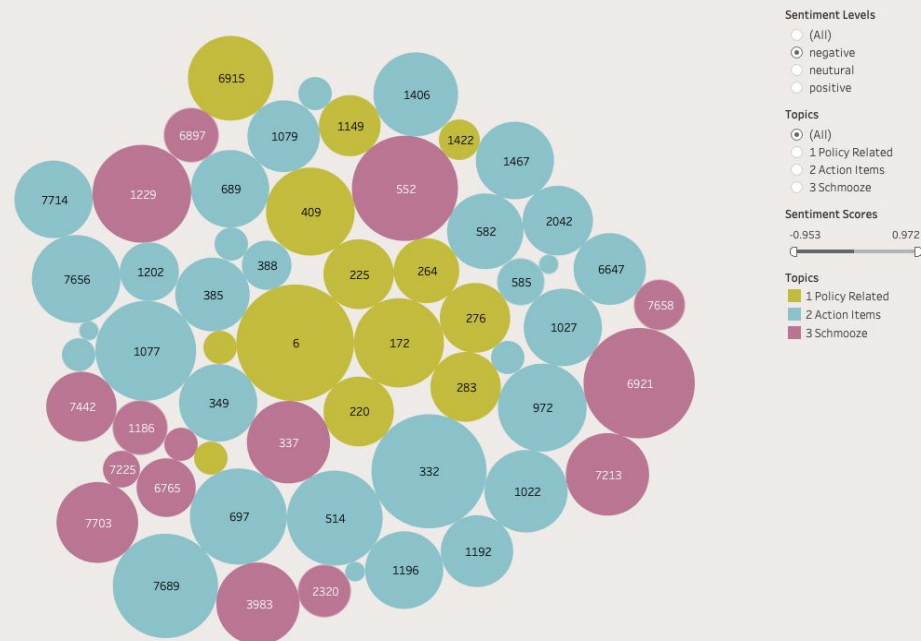
by Vanessa Hu

The emails in this analysis were sent by Hillary Rodman Clinton between 2009 and 2012. This scatter plot includes an overview of sentiment scores and three topics. The

Let's take a look at all the emails sent by Hillary under 3 topics as an overview. In general, she conducts policy related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.



APPENDIX

Tableau Story Line

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

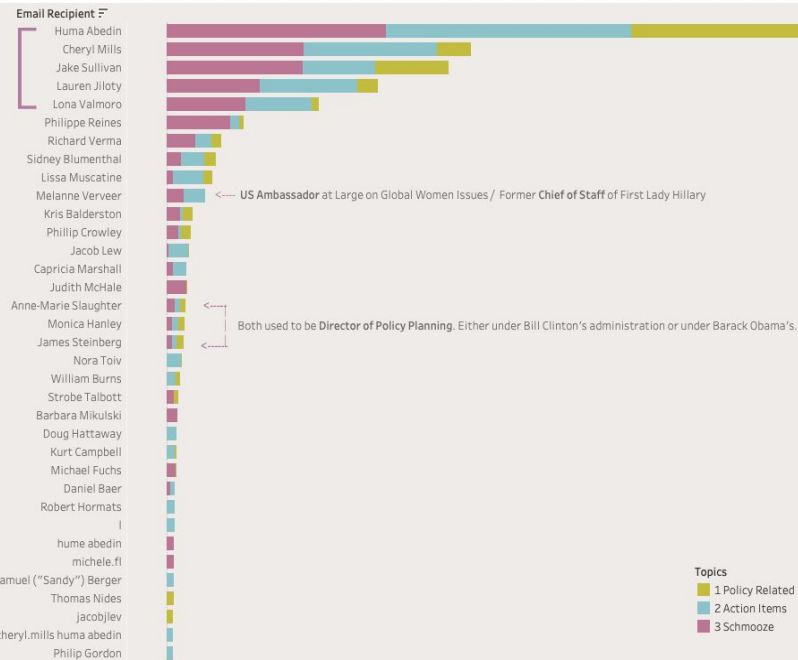
by Vanessa Hu

The emails in this analysis were sent by Hillary Rodham Clinton between 2009 and 2012. This scatter plot includes an overview of sentiment scores and three topics. The

Let's take a look at all the emails sent by Hillary under 3 topics as an overview. In general, she conducts policy related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

by Vanessa Hu

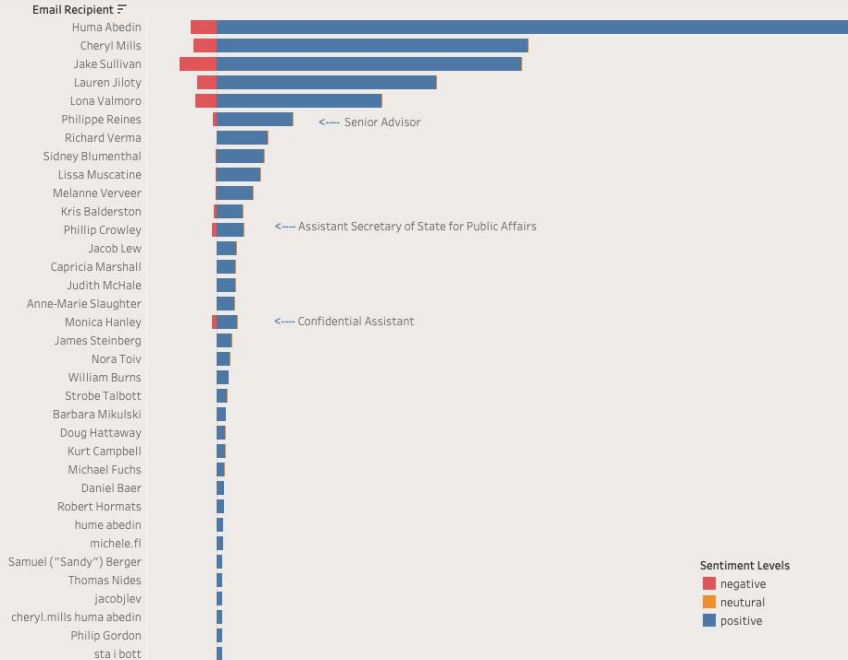
take a look at all the emails by Hillary under 3 topics as an overview. In general, she conducts related conversation with advisors, coordinates action items

This breakdown chart with customized parameters allows users to view and compare between each topic and sentiment levels.

Let's explore Hillary's circle of trust. She sent out more than 70% of her emails to just 5 people, her long-time aids and policy advisors.

By using sentiment analysis, we can more clearly see who her close confidants are. The emails with negative sentiment scores indicate closer relationships with less

Another sentiment network each s



APPENDIX

Tableau Story Line

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

by Vanessa Hu

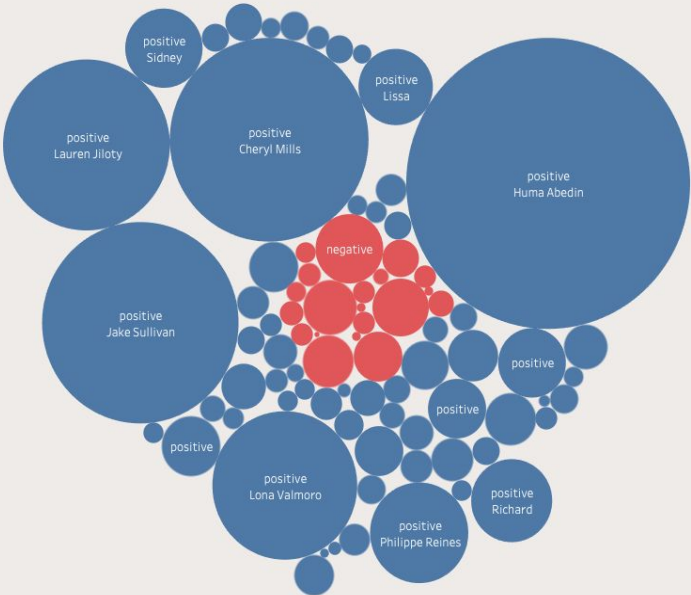
breakdown chart with
mixed parameters allows
to view and compare between
topic and sentiment levels.

Let's explore Hillary's circle of trust.
She sent out more than 70% of her
emails to just 5 people, her long-
time aids and policy advisors.

By using sentiment analysis, we can
more clearly see who her close
confidants are. The emails with
negative sentiment scores indicate
closer relationships with less

Another bubble chart to visualize
sentiments throughout her
network. Choose the filter to check
each sentiment level.

Three
sent t
repre
conve
match



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

by Vanessa Hu

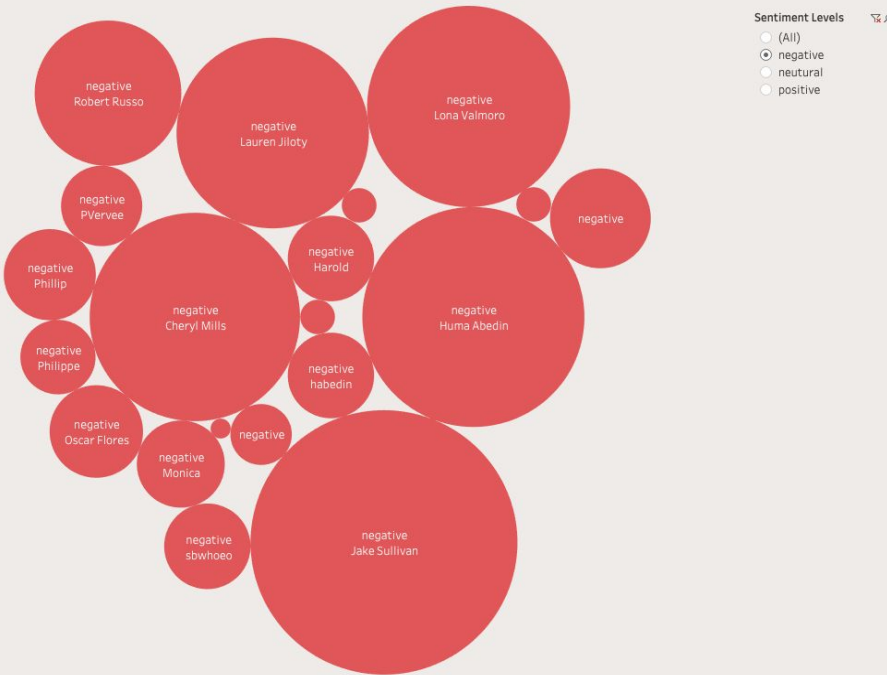
breakdown chart with
mixed parameters allows
to view and compare between
topic and sentiment levels.

Let's explore Hillary's circle of trust.
She sent out more than 70% of her
emails to just 5 people, her long-
time aids and policy advisors.

By using sentiment analysis, we can
more clearly see who her close
confidants are. The emails with
negative sentiment scores indicate
closer relationships with less

Another bubble chart to visualize
sentiments throughout her
network. Choose the filter to check
each sentiment level.

Three
sent t
repre
conve
match



APPENDIX

Tableau Story Line

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

by Vanessa Hu

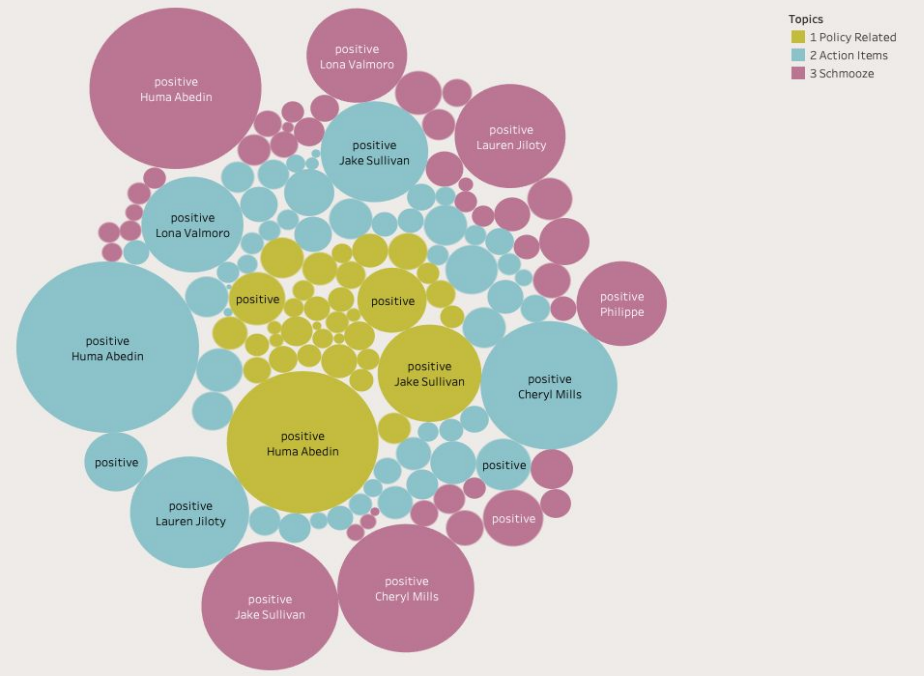
explore Hillary's circle of trust. ant out more than 70% of her s to just 5 people, her long-ids and policy advisors.

By using sentiment analysis, we can more clearly see who her close confidants are. The emails with negative sentiment scores indicate closer relationships with less

Another bubble chart to visualize sentiments throughout her network. Choose the filter to check each sentiment level.

Three topics from the emails Hillary sent to her network. The bubble size represents how closely the conversion with that person matches the topic.

Timeli The pr are rel beginn Hillary



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

by Vanessa Hu

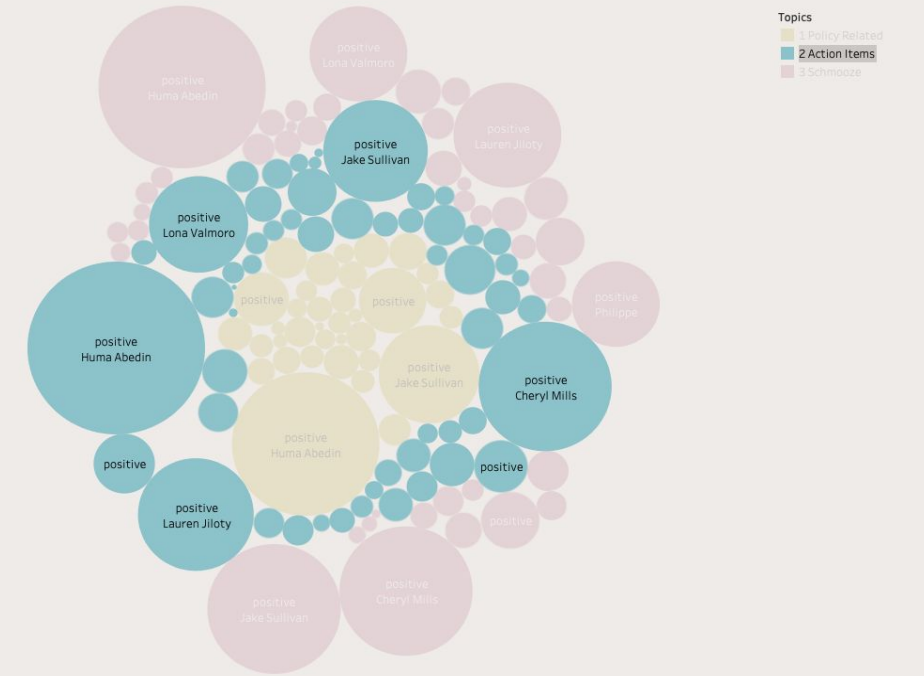
explore Hillary's circle of trust. ant out more than 70% of her s to just 5 people, her long-ids and policy advisors.

By using sentiment analysis, we can more clearly see who her close confidants are. The emails with negative sentiment scores indicate closer relationships with less

Another bubble chart to visualize sentiments throughout her network. Choose the filter to check each sentiment level.

Three topics from the emails Hillary sent to her network. The bubble size represents how closely the conversion with that person matches the topic.

Timeli The pr are rel beginn Hillary



APPENDIX

Tableau Story Line

NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

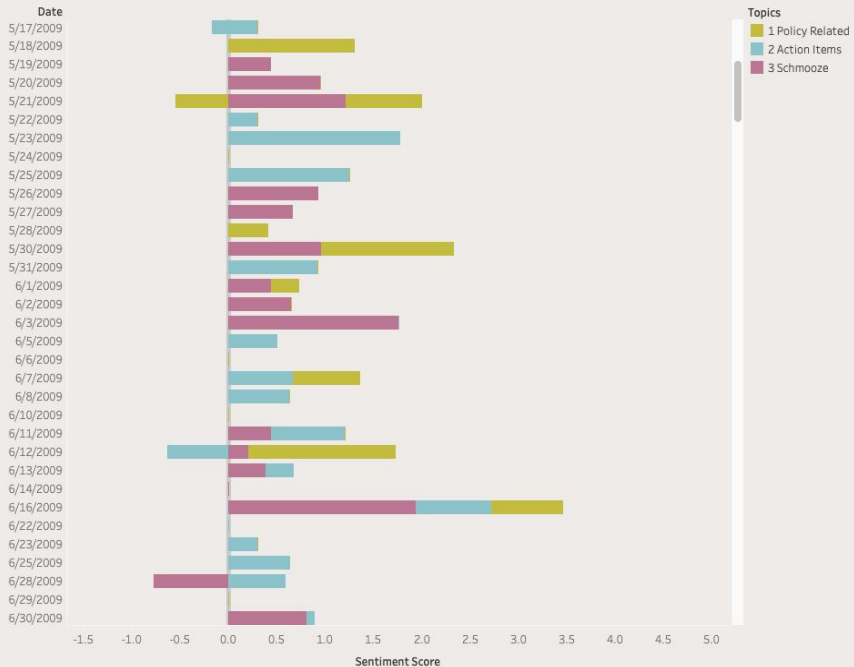
by Vanessa Hu

By using sentiment analysis, we can more clearly see who her close confidants are. The emails with negative sentiment scores indicate closer relationships with less

Another bubble chart to visualize sentiments throughout her network. Choose the filter to check each sentiment level.

Three topics from the emails Hillary sent to her network. The bubble size represents how closely the conversion with that person matches the topic.

Timeline with Topics and Sentiment. The percent composition of topics are relatively stable in 2009 and the beginning of 2010. Emails sent from Hillary are missing from the second



NLP Topic Modeling and Sentiment Analysis on Hillary Clinton's Emails

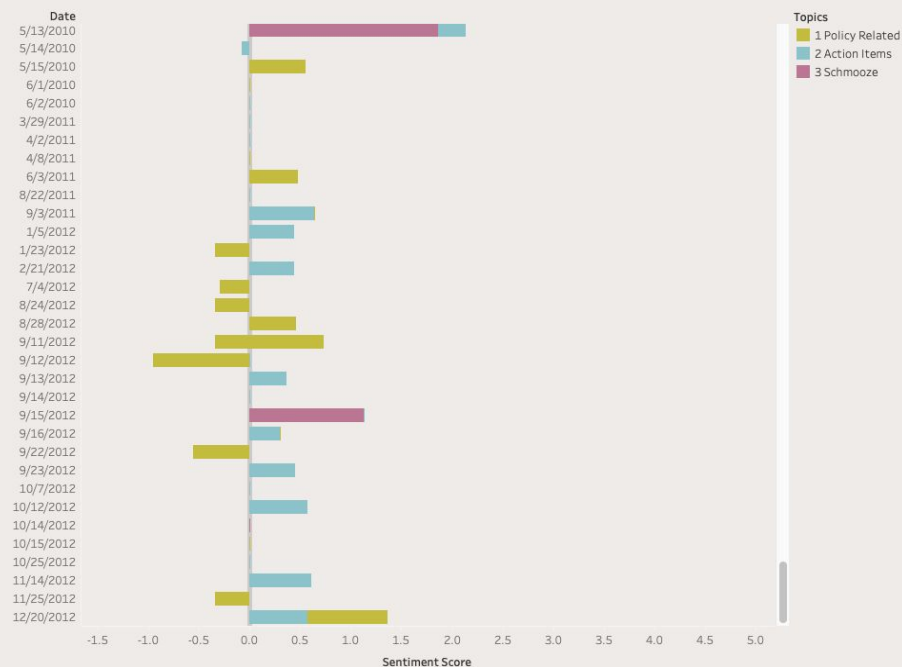
by Vanessa Hu

By using sentiment analysis, we can more clearly see who her close confidants are. The emails with negative sentiment scores indicate closer relationships with less

Another bubble chart to visualize sentiments throughout her network. Choose the filter to check each sentiment level.

Three topics from the emails Hillary sent to her network. The bubble size represents how closely the conversion with that person matches the topic.

Timeline with Topics and Sentiment. The percent composition of topics are relatively stable in 2009 and the beginning of 2010. Emails sent from Hillary are missing from the second



APPENDIX

Tableau Story Line

Fonts & colors used

This presentation has been made using the following fonts:

Nunito Sans

(<https://fonts.google.com/specimen/Nunito+Sans>)

Assistant

(<https://fonts.google.com/specimen/Assistant>)

#143a78

#d9d9d9

#f3f3f3

#ee5c00