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

March 2015

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

May 2015

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

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

Text processing

REGEX, Lammentization
TF-IDF

spaCy
02

Topic Modeling

CorEx, LSA, Word2Vec
Three Topics

LDA
03

Sentiment Analysis

NLTK VADER
TEXTBLOB

VADER
04

Visualization

Topics, Sentiment,
Circle of Trust, Dates

Tableau
05

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

NEUTRAL

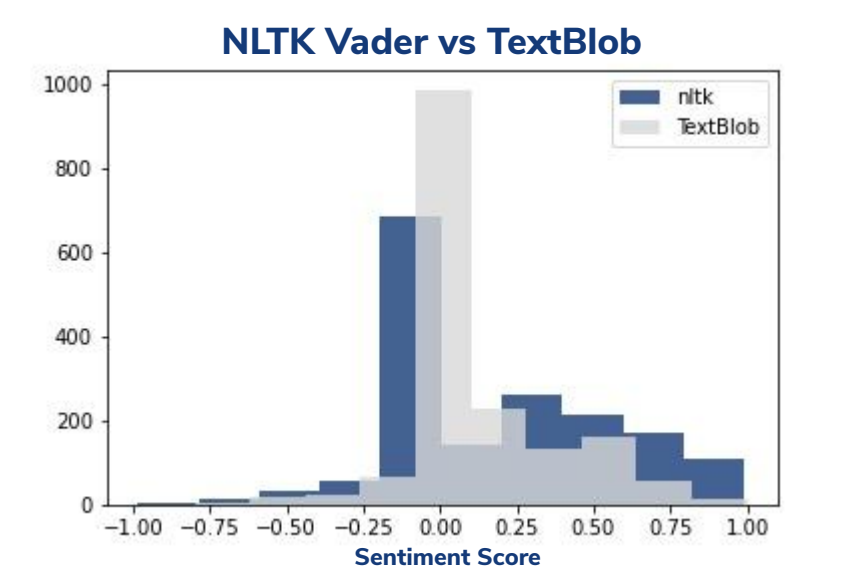
47.5%

POSITIVE

47%

NEGATIVE

5.5%



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?

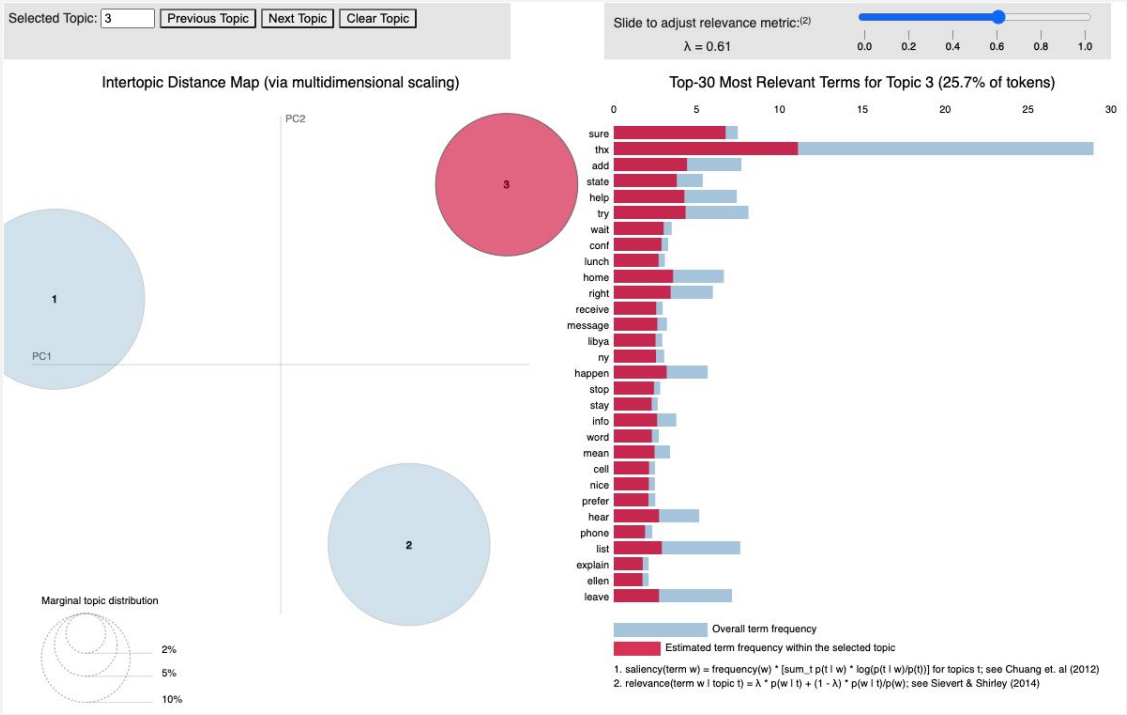
Tableau Demo

github.com/vanessa920/nlp_hillary_email

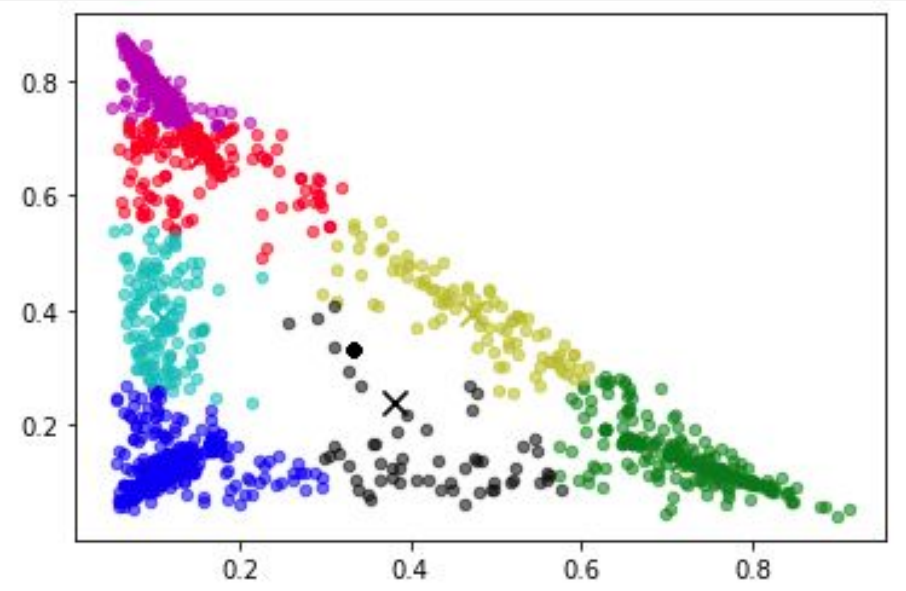
TAKE-AWAYS

- 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

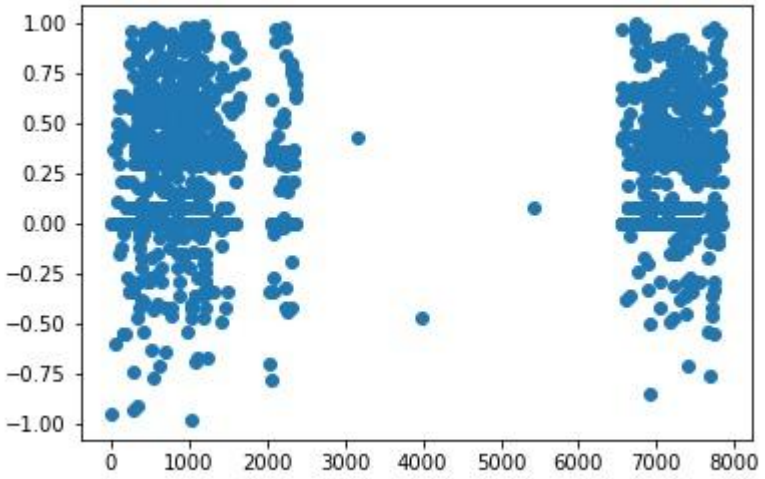


APPENDIX
K-Means Cluster Chart

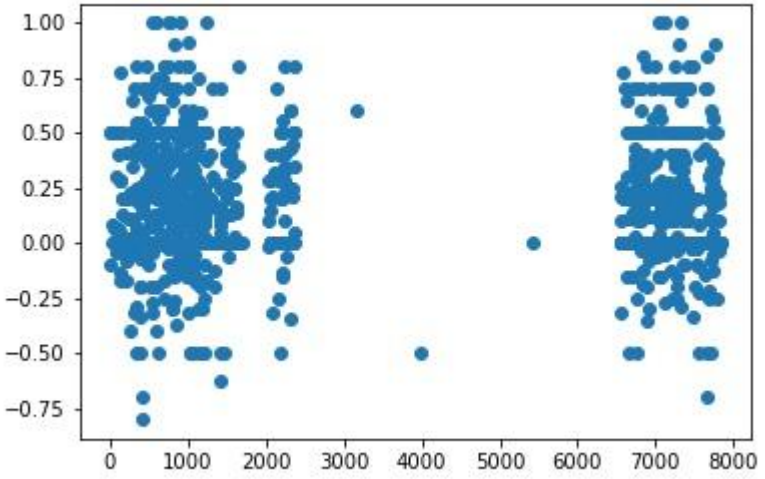


APPENDIX

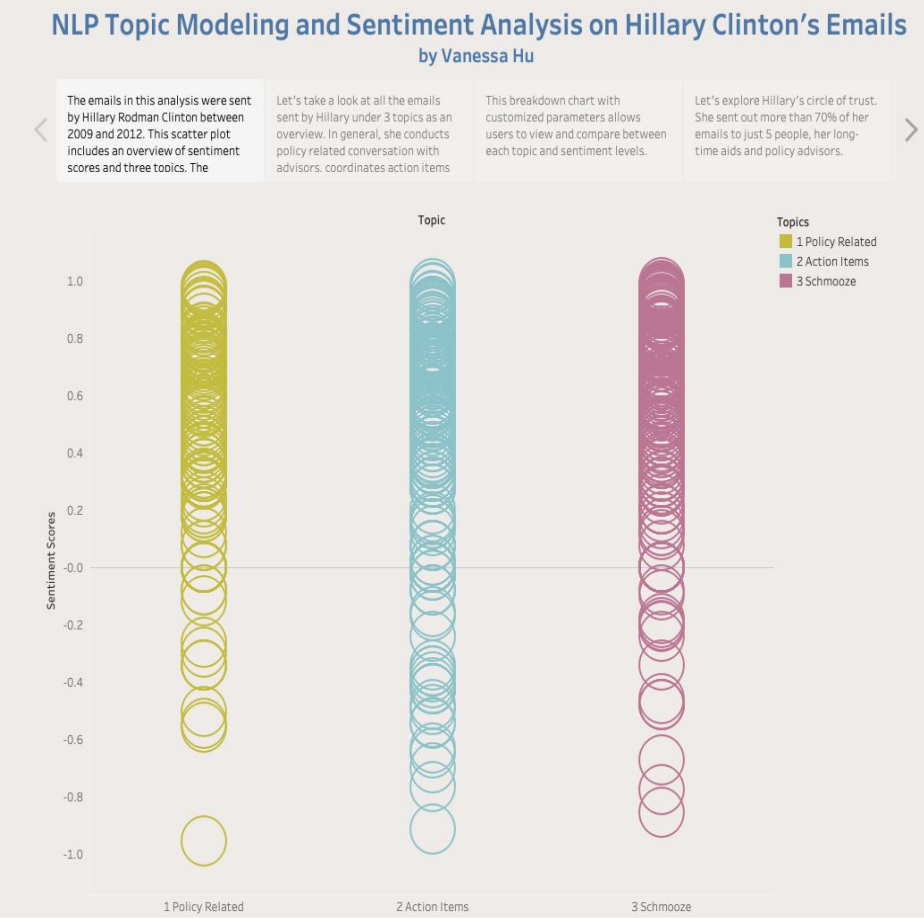
NLTK Vader and TextBlob Scatter Chart

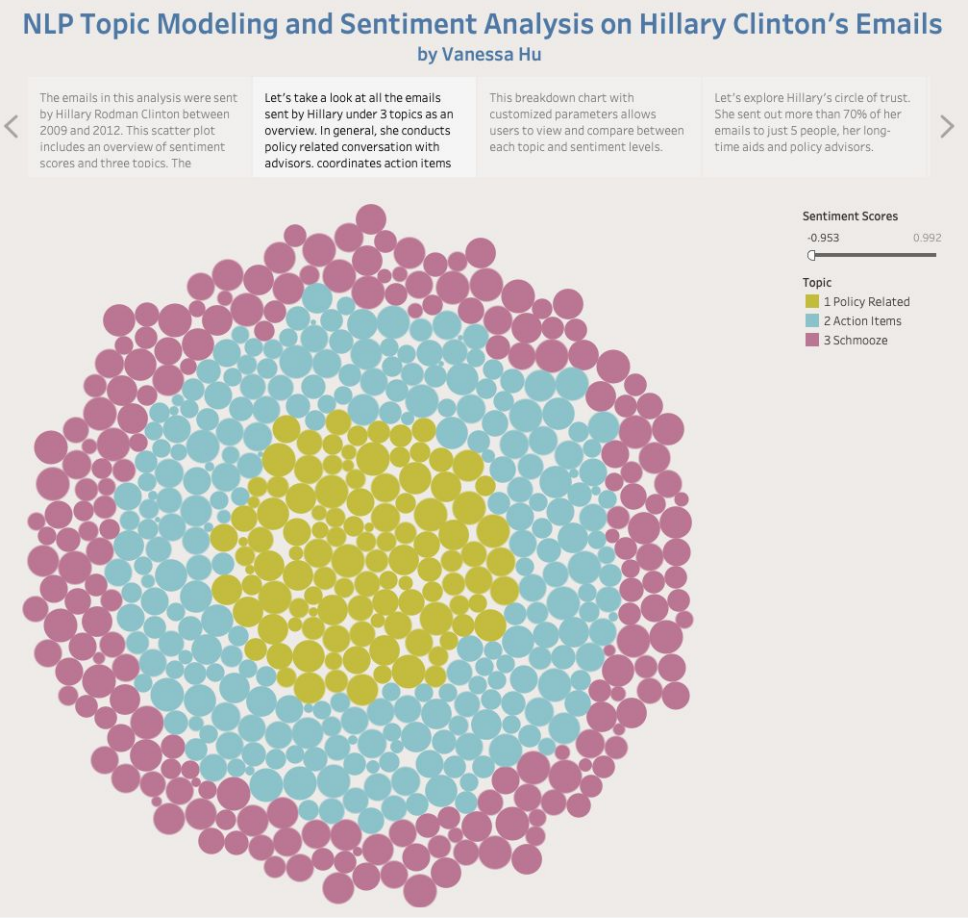


NLTK Vader Sentiment



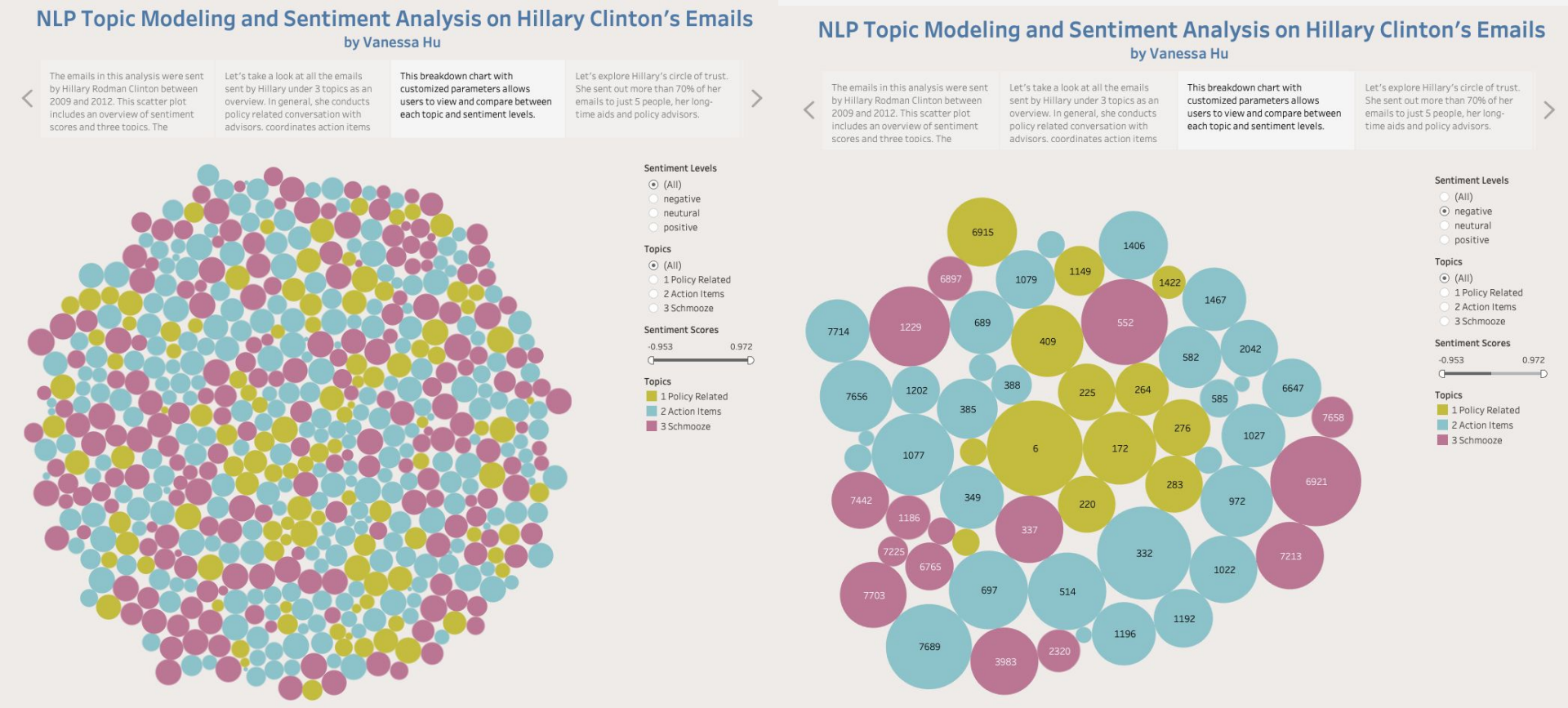
TextBlob Sentiment





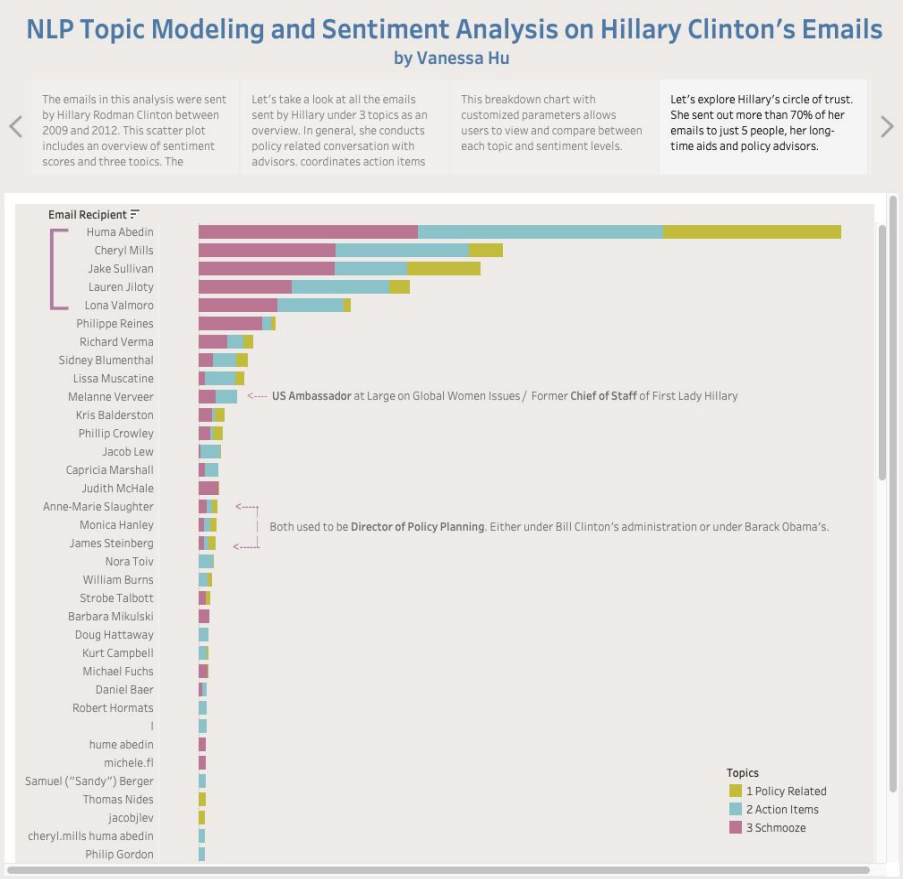
APPENDIX

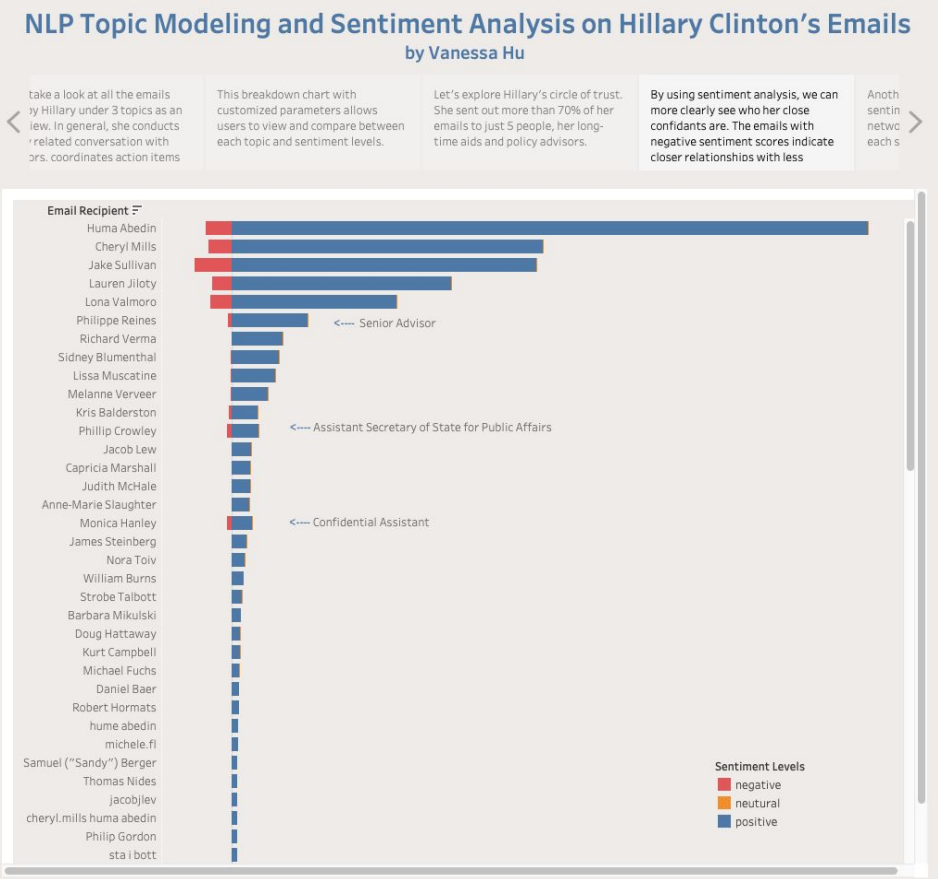
Tableau Story Line



APPENDIX

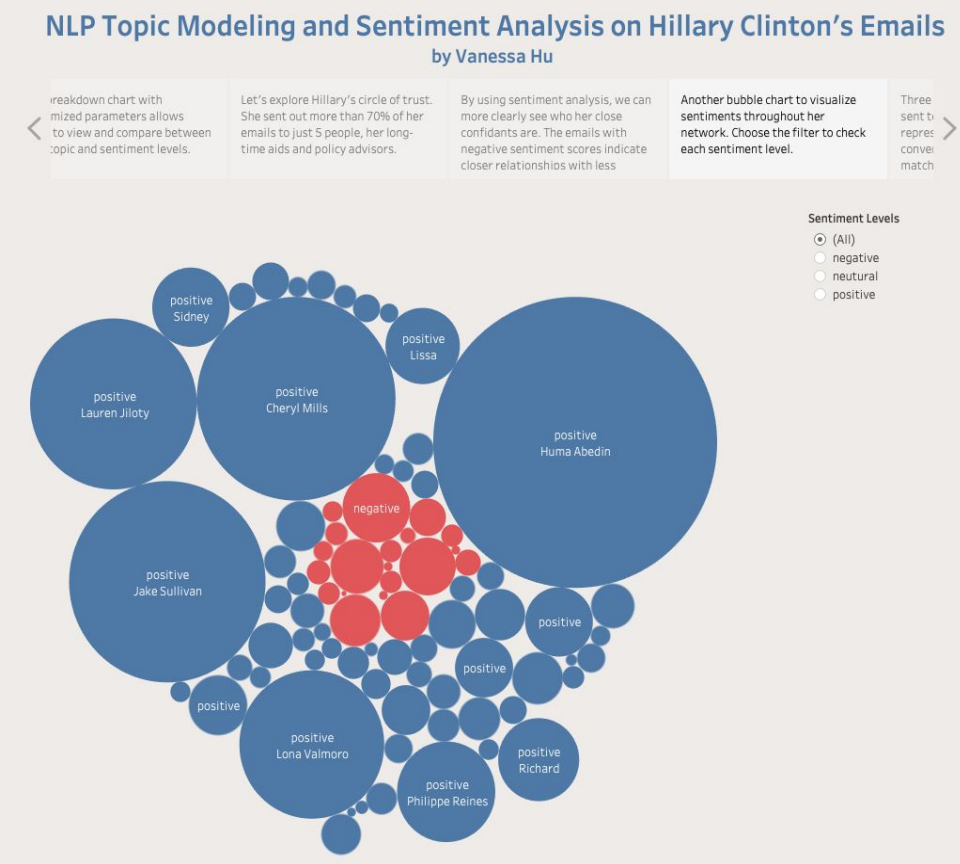
Tableau Story Line

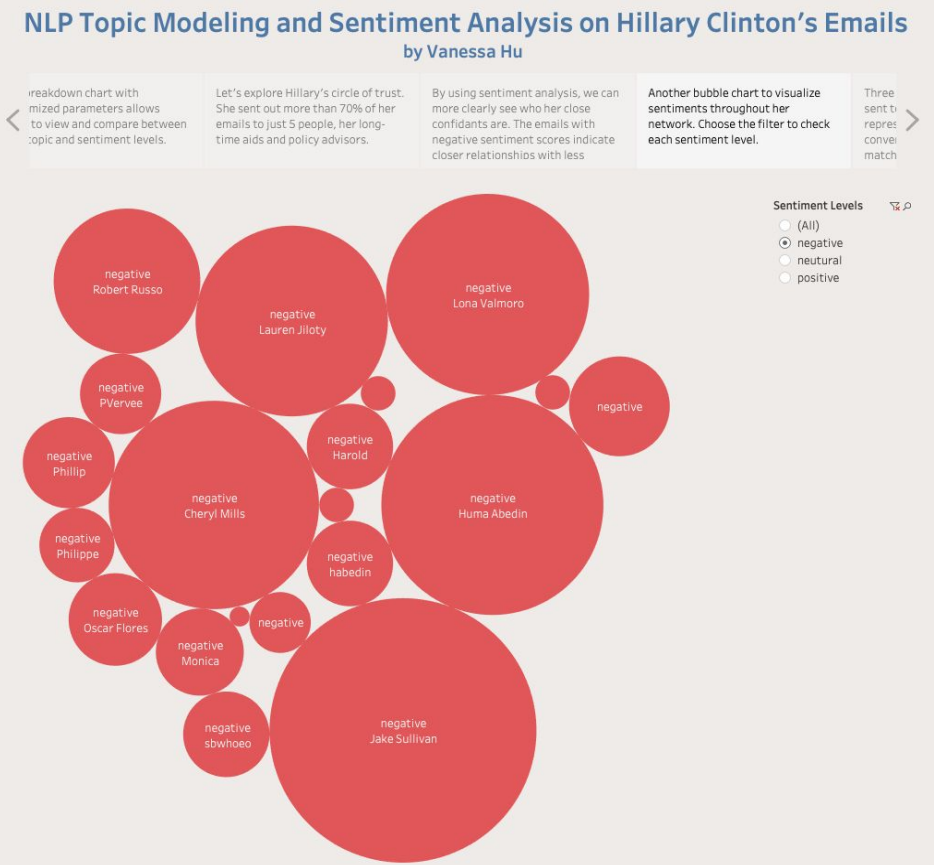




APPENDIX

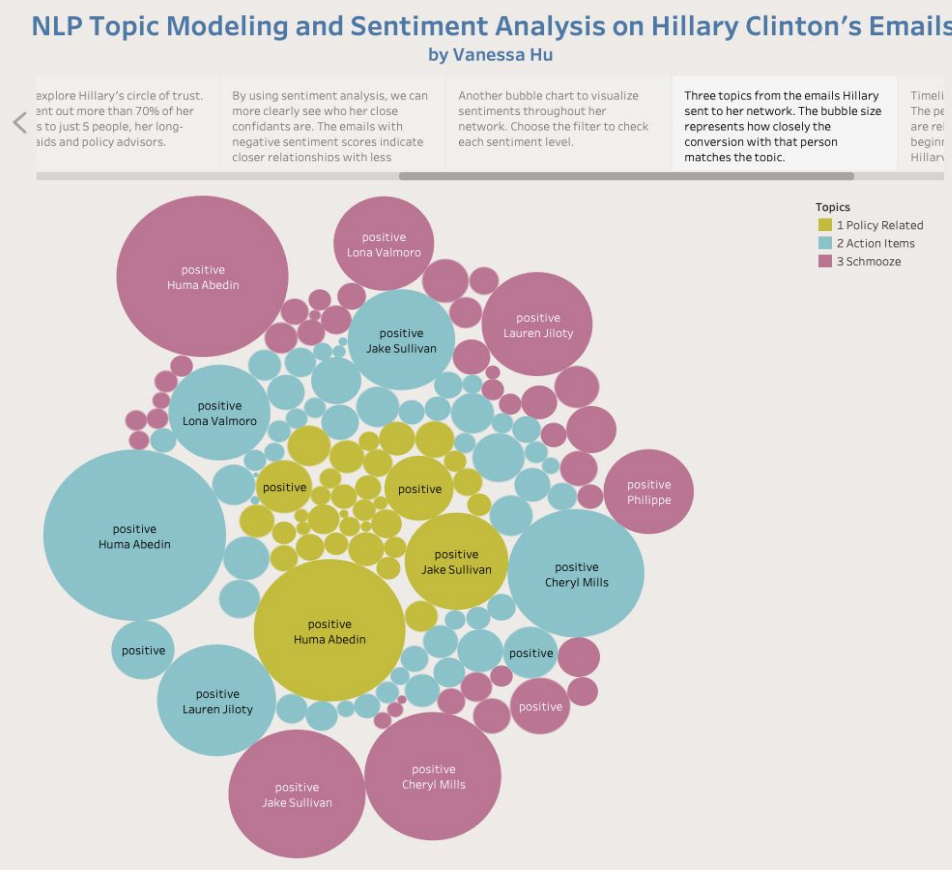
Tableau Story Line

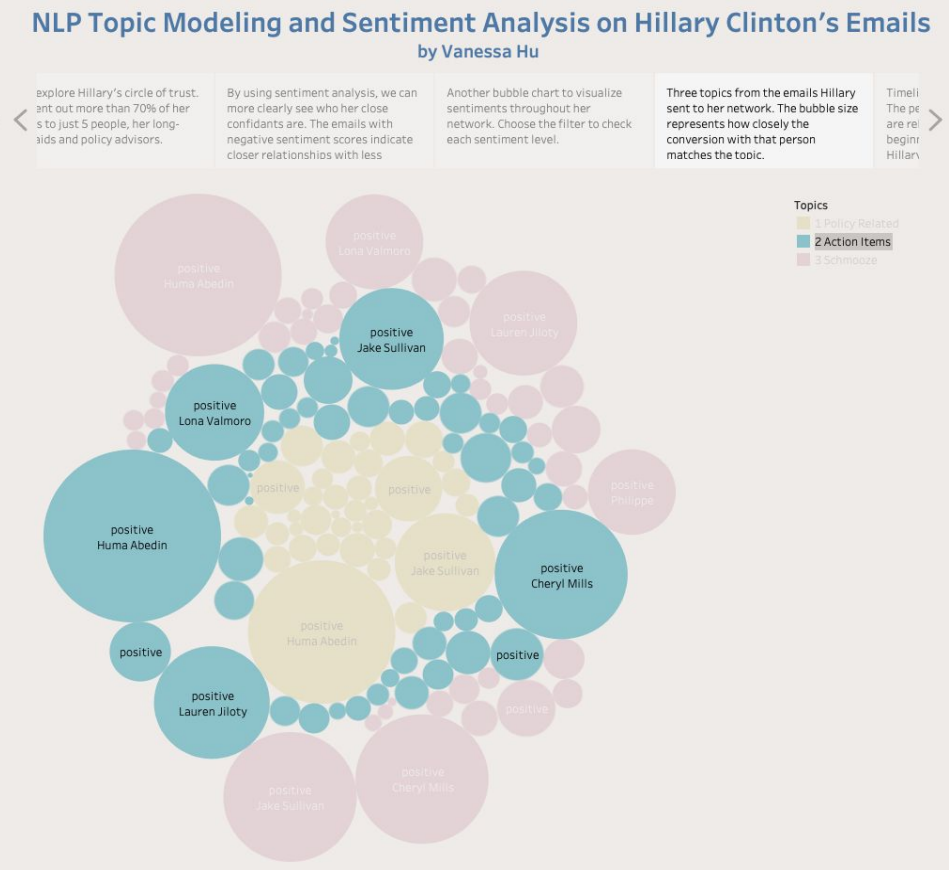




APPENDIX

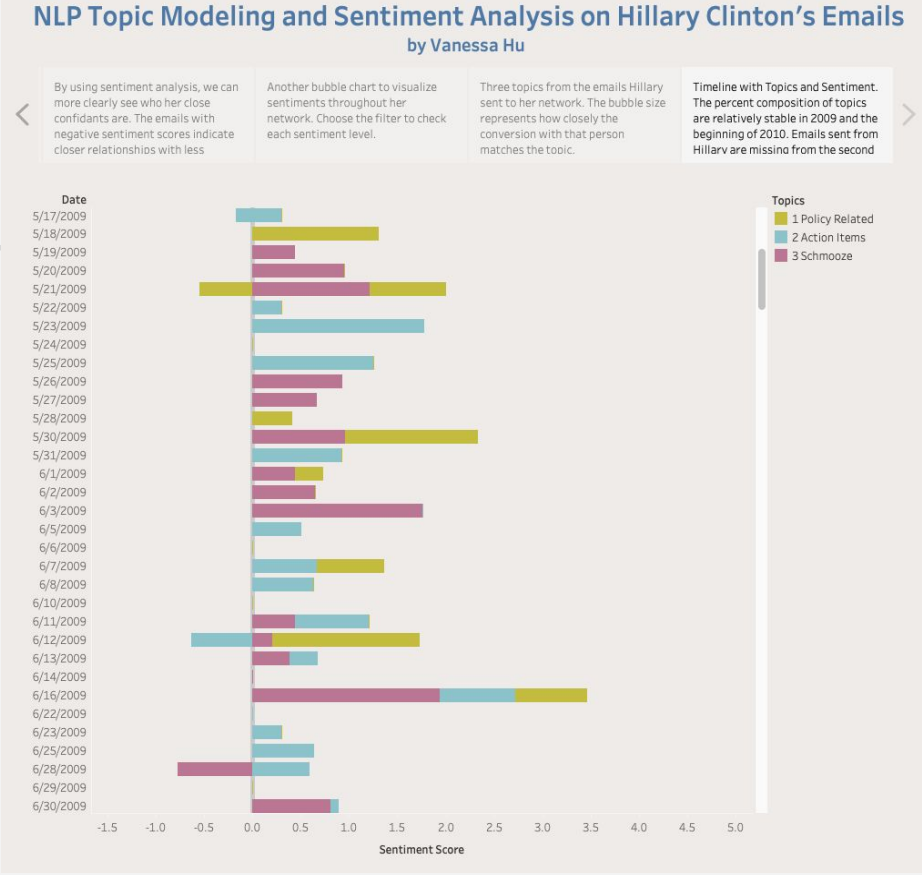
Tableau Story Line

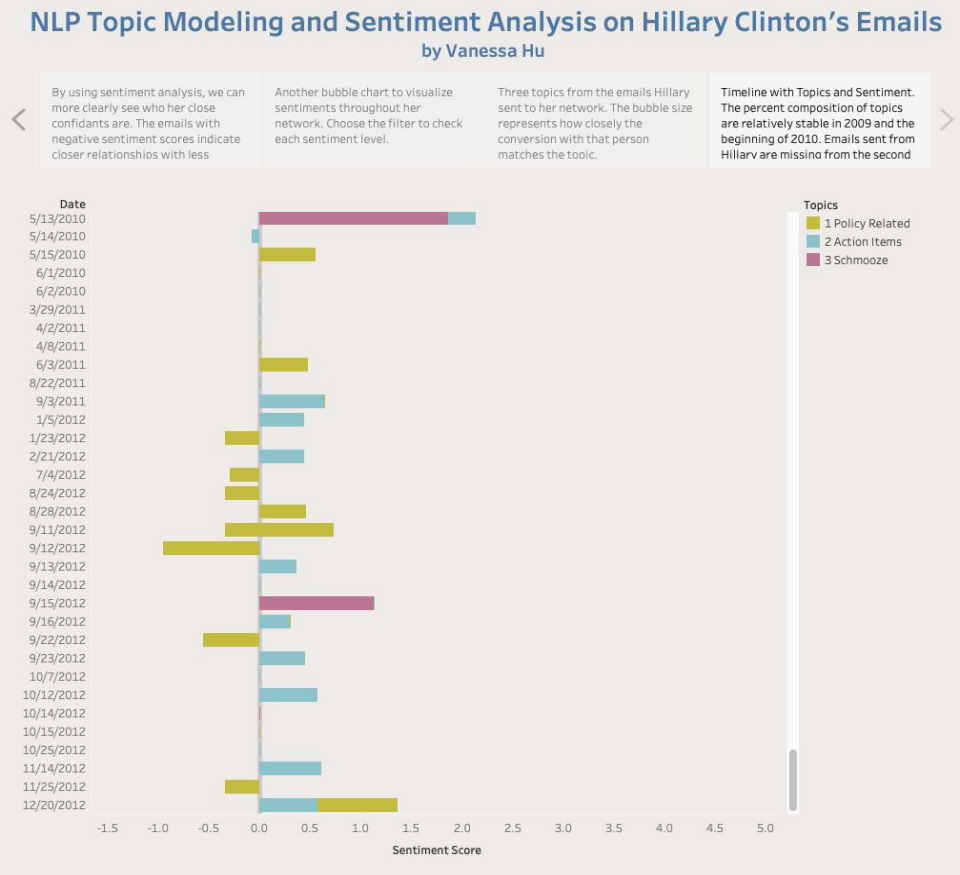




APPENDIX

Tableau Story Line





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