





The Interactive Filter and Display of Hillary Clinton's Emails

About “her emails”

During the 2016 US presidential election campaign, Hillary Clinton's emails were the source of much debate and controversy – and may have played a decisive role in her defeat by Donald Trump. **Christopher Salahub** and **Wayne Oldford** have built a tool to analyse the contents of her private server. What more can we learn?

The 2016 US presidential election was one of the most contentious in history, and a major political flashpoint was the existence and possible content of Hillary Clinton's private email server. This dogged her bid for the White House, and was doubtless a contributing factor in her defeat by Donald Trump.

Clinton's use of a private server, while serving as Secretary of State from 2009 to 2013, was revealed during Congressional investigation of an attack on US government facilities in Benghazi, Libya. The attack, which took place on 11–12 September 2012, left four Americans dead, prompting criticism of the State Department's – and Clinton's – handling of the situation. An FBI investigation into the email server

ABOVE Hillary Clinton at an event in Pennsylvania in October 2016, during her campaign for the US presidency.

followed and subsequently ended only two days before election day. In that time, Clinton's political opponents took full advantage of the email controversy; Trump in particular used it as a basis to label his opponent “crooked”.

There are of course genuine and valid concerns about a senior official using a private email server. Questions we might ask include: Were all government-related emails preserved for the historical record? Who corresponded with the Secretary of State, and on what topics? Were classified documents shared with anyone having insufficient security clearance to see them?

As statisticians, we might be able to help answer such questions. All emails turned over by Clinton's team from

Presented by: Christopher Salahub

Joint work with R. Wayne Oldford

Based on work published in

Significance June 2018



The Emails

- **Roughly half of all emails from clintonemail.com (Registered January 13, 2009)**
 - **Other half irrevocably deleted**
- **A corpus of 32,795 documents as of May 2018**
 - **Emails and attachments**

Getting the Emails

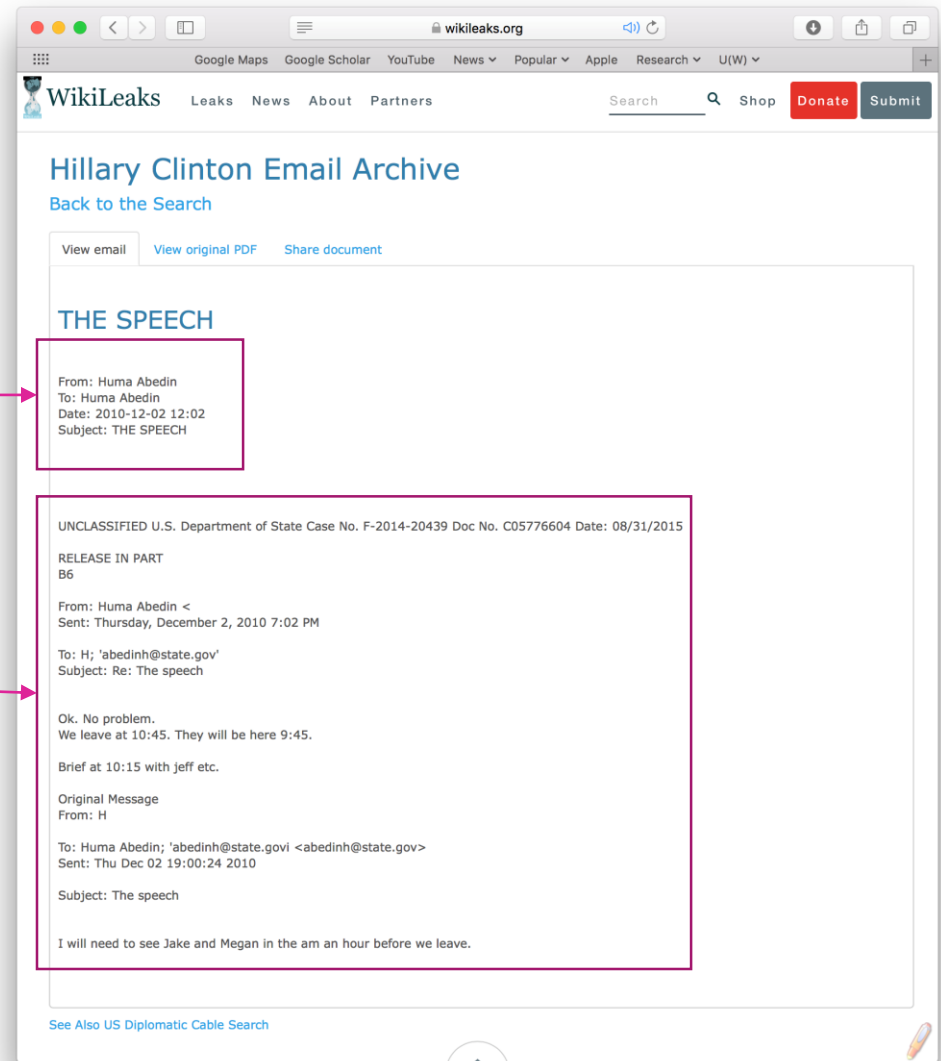
- Released by the State Department (FOIA request)
- Processed from PDF to HTML by Wikileaks
- Wikileaks HTML extracted and processed



The screenshot shows the WikiLeaks website interface. At the top, there is a navigation bar with the WikiLeaks logo, links for 'Leaks', 'News', 'About', and 'Partners', a search bar, and buttons for 'Shop', 'Donate', and 'Su'. The main heading is 'Hillary Clinton Email Archive'. Below the heading is a paragraph of text: 'On March 16, 2016 WikiLeaks launched a searchable archive for over 30 thousand emails & email attachments sent to and from Hillary Clinton's private email server while she was Secretary of State. The 50,547 pages of documents span from 30 June 2010 to 12 August 2014. 7,570 of the documents were sent by Hillary Clinton. The emails were made available in the form of thousands of PDFs by the US State Department as a result of a Freedom of Information Act request. More PDFs were made available on February 29, 2016, and a set of additional 995 emails was imported up to February 2, 2018.' To the right of the text is a cartoon illustration of a woman sitting at a laptop, with a globe and dollar signs around her, and a hand holding a document above her. Below the text is a search bar with two tabs: 'Search by Terms in Email' and 'Search by Email-ID'. The search bar contains the text 'Search terms throughout whole of email'. To the right of the search bar is a note: 'You can use boolean operators to search emails. (Example: *syria libya* will show results containing both words. *syria | libya* will show results with either words, while *syria !libya* will show results containing

Final Data

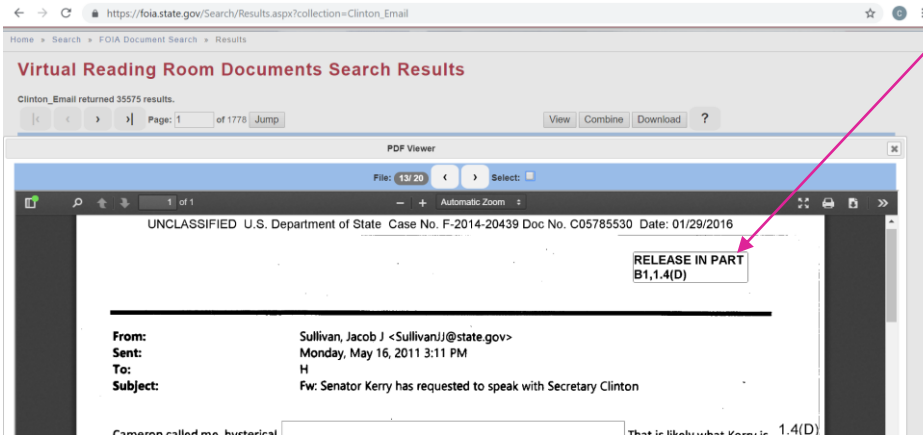
- **Header**
 - Subject, email addresses, date and time
- **Body**
 - Email content



Auxiliary and Synthetic Data

- FOIA redaction codes
- Term frequency
- TF-IDF (term “importance”)
- Official travel schedule

RELEASE IN PART
B1,1.4(D)



Putting it All Together

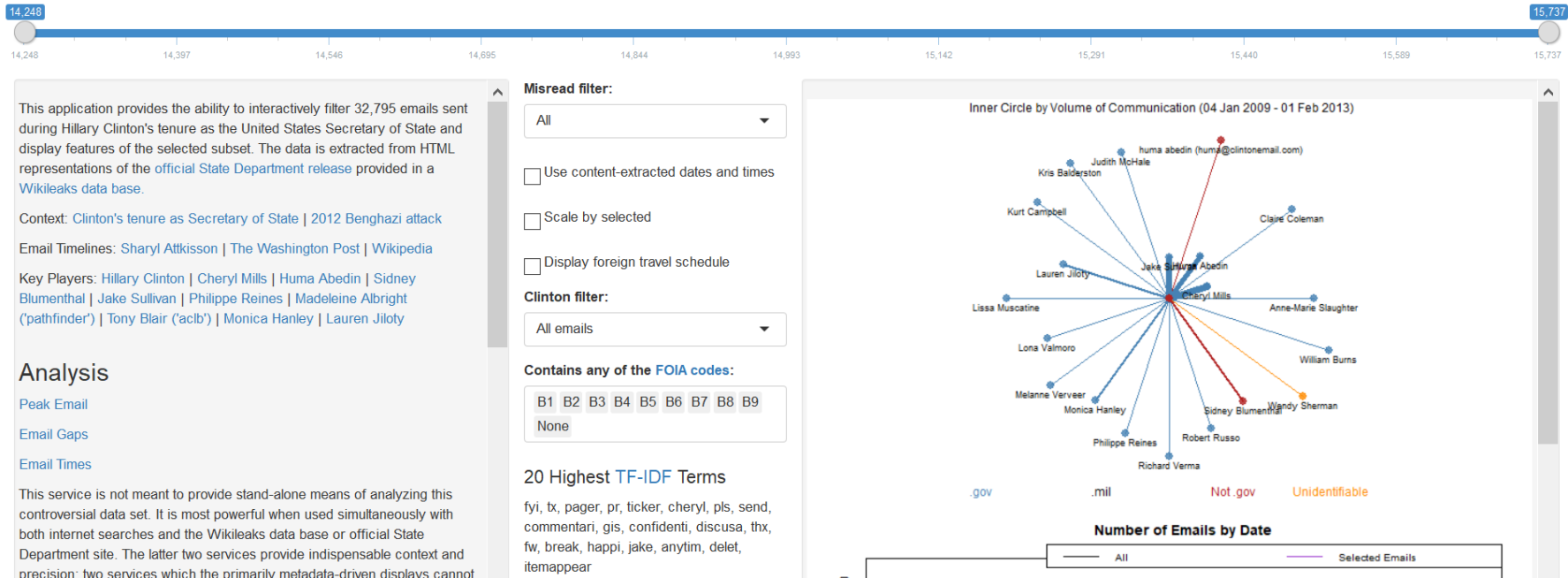
Secretary Clinton's Email (Source: Wikileaks)

Christopher D. Salahub and R. Wayne Oldford: [Interactive Filter and Display of Hillary Clinton's Emails: A Cautionary Tale of Metadata](#)

Application Version 1.6 | Data Version 2.1 | May 3, 2018

Date Range:

04 Jan 2009 - 01 Feb 2013



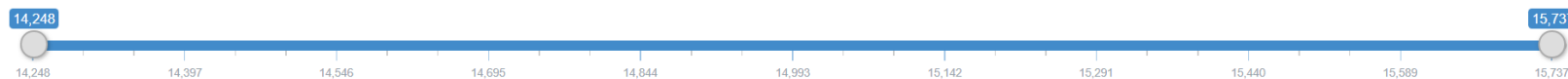
<https://shiny.math.uwaterloo.ca/sas/clinton/>



Features: Filter the Data

Date Range:

04 Jan 2009 - 01 Feb 2013



Misread filter:

All ▼

Use content-extracted dates and times

Scale by selected

Display foreign travel schedule

Clinton filter:

All emails ▼

Contains any of the **FOIA codes:**

B1 B2 B3 B4 B5 B6 B7 B8 B9
None

Date

Wikileaks data errors

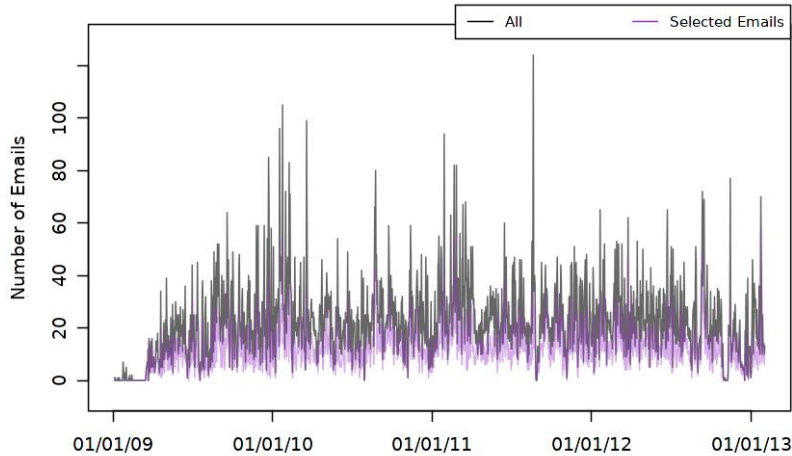
Display settings

Clinton to/from

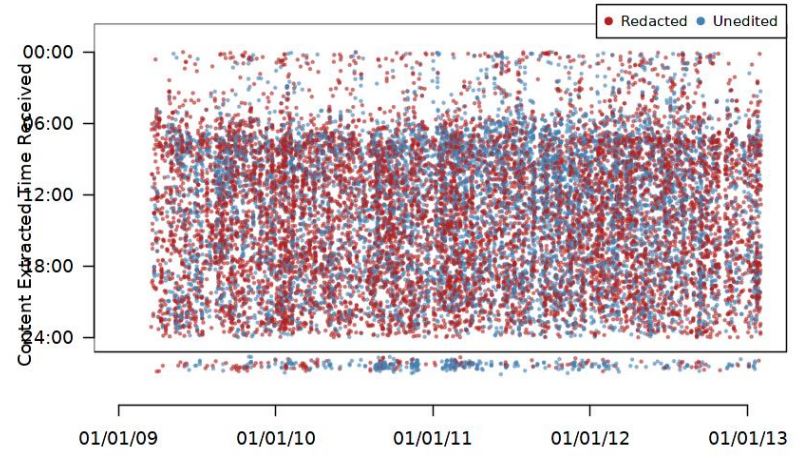
FOIA codes

Features: Interactively Visualize

Number of Emails by Date

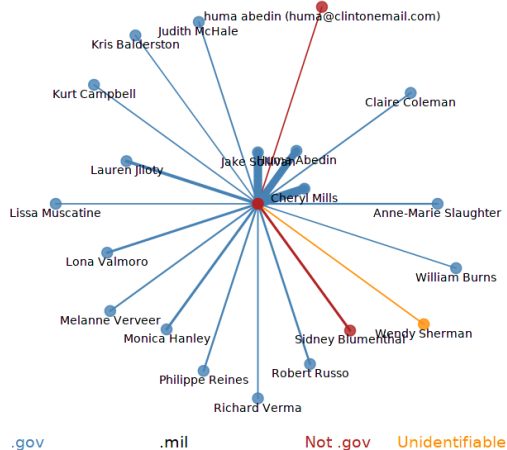


Email Received Times Extracted from Content



Date (dd/mm/yy)
04 Jan 2009 - 01 Feb 2013

Inner Circle by Volume of Communication (04 Jan 2009 - 01 Feb 2013)



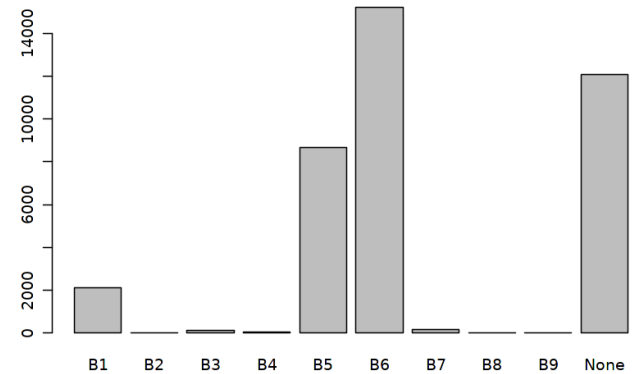
20 Highest TF-IDF Terms

fyi, tx, pager, pr, ticker, cheryl, pls, send, commentari, gis, confidenti, discusa, thx, fw, break, happi, jake, anytim, delet, itemappear

20 Highest Frequency Terms

secur, peopl, year, nation, countri, obama, support, offic, clinton, make, report, polit, american, unit, foreign, hous, talk, minist, public, day

FOIA Redaction Codes Appearing in Selected Emails



04 Jan 2009 - 01 Feb 2013

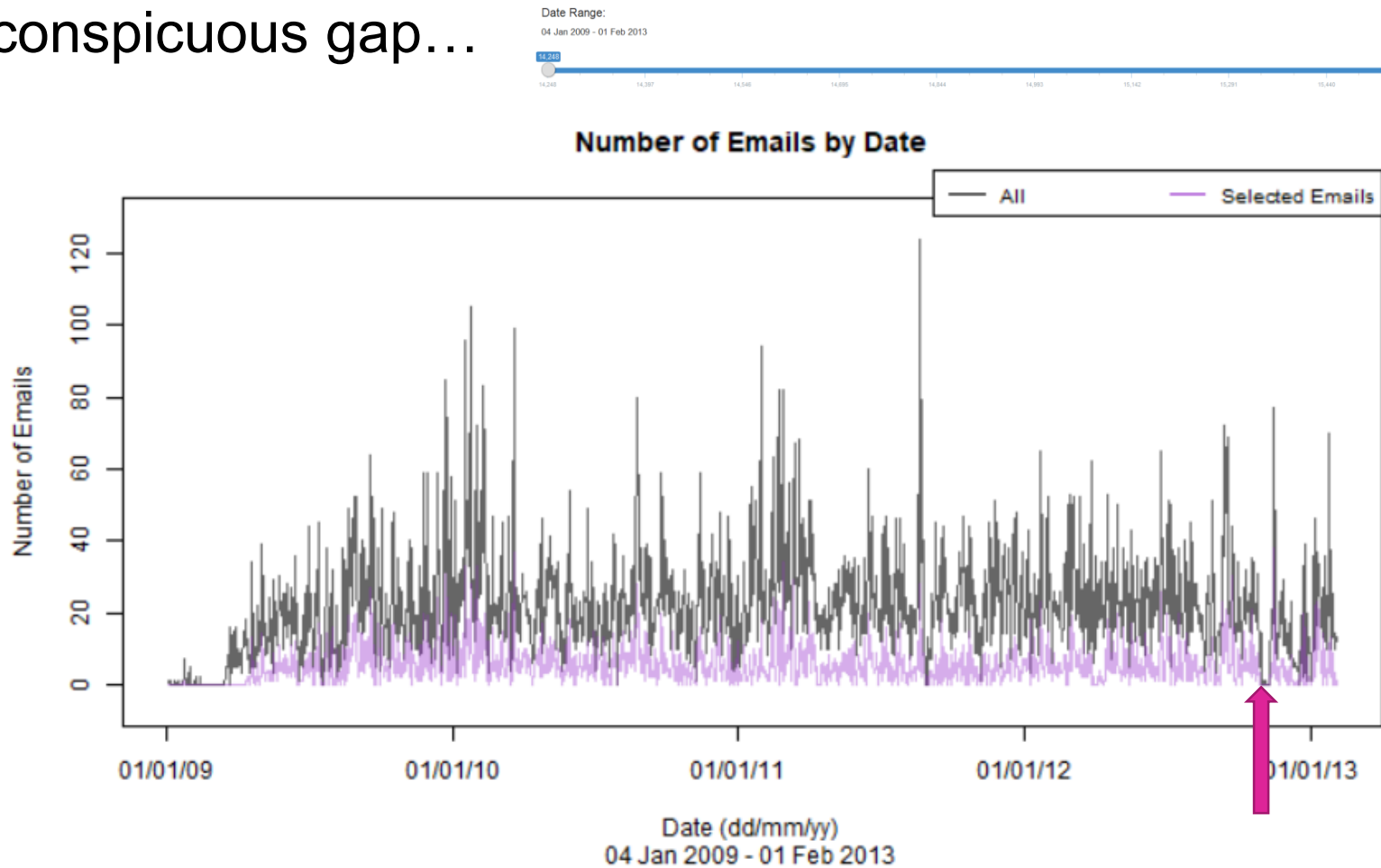


What should we do with this?

Explore!

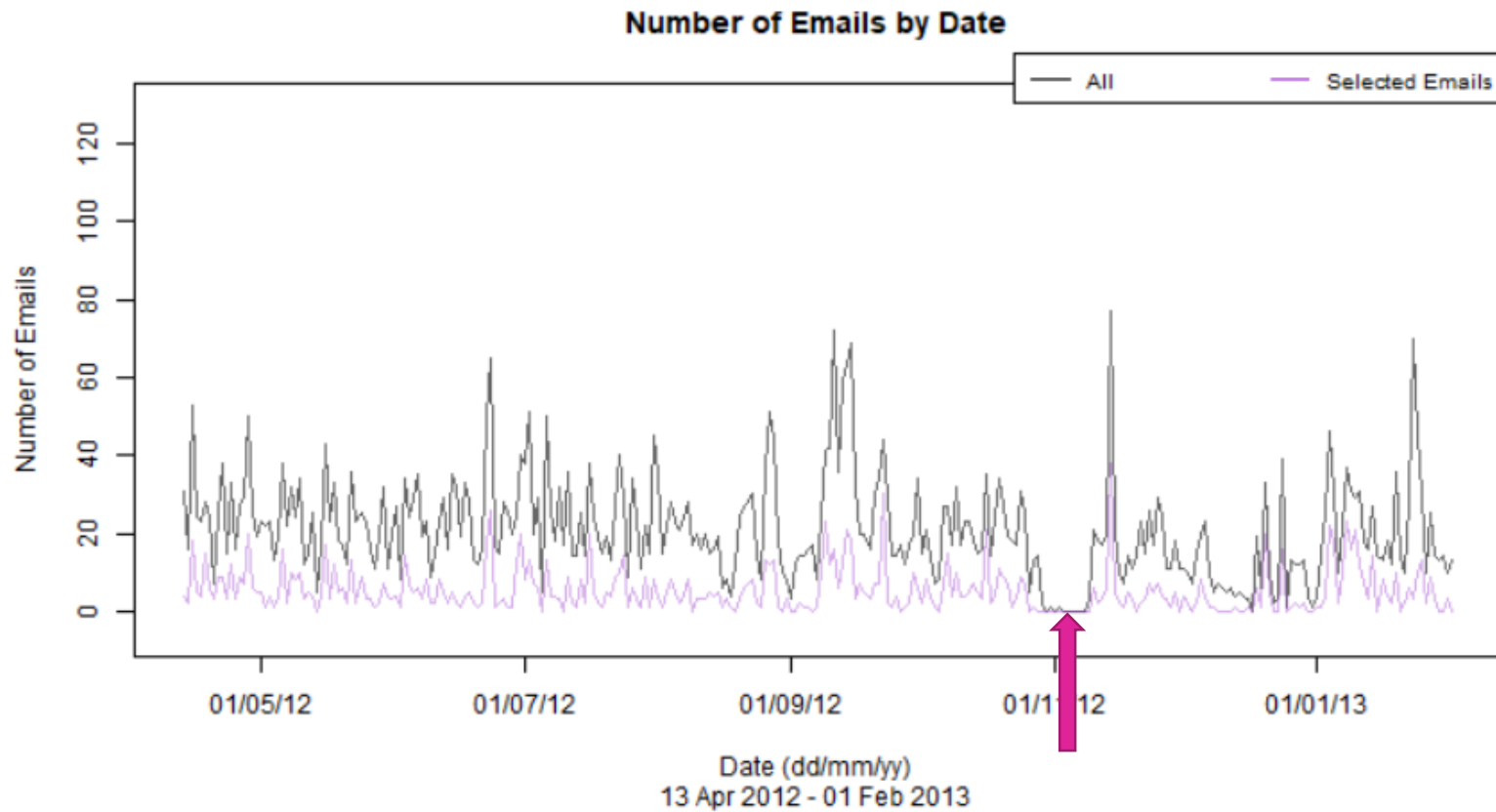
Missing Emails?

- A conspicuous gap...



Missing Emails?

- Filter by time...



Missing Emails?

- But is this gap of six days (Nov. 3-9 2012) “significant?”
 - In other words: would we expect such a gap to be present as the result of random sampling?

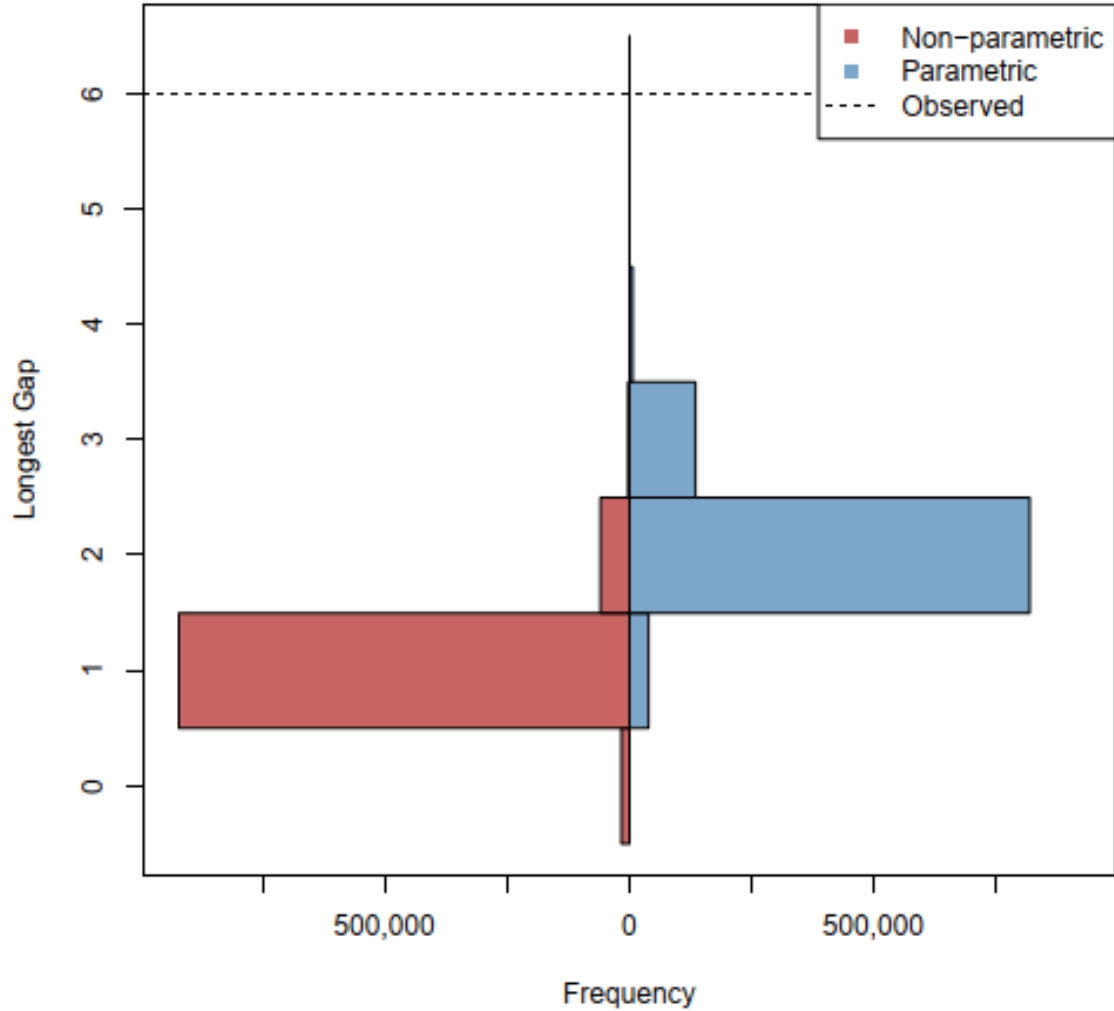


A Tale of Two Tests

- Parametric
 - Estimate the probability of a day with no email
 - Simulate independent random samples and note the largest gap
- Non-parametric
 - Duplicate each email
 - Remove gaps by shifting dates
 - Sample half of these email dates and note the largest gap
 - Meant to mimic the selection process



A Tale of Two Tests



Why the Gap?

By SHARYL ATTKISSON / CBS NEWS / October 18, 2012, 12:40 PM

POLITICS

EXCLUSIVE: CIA operators were denied request for help during Benghazi attack, sources say



By Jennifer Griffin

Published October 26, 2012

Fox News



Ad closed by Google

Stop seeing this ad

AdChoices

Advertisement: Faith or fellatio? British religionists demand to submit to secular indoctrination

More from Fox News

Libya consulate: Was security added or taken away?

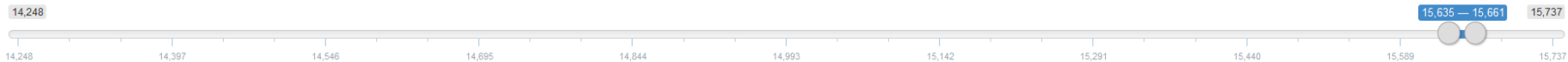
Lt. Col. Andy Wood led a 16-member Special Forces site security team responsible for

Why the Gap?

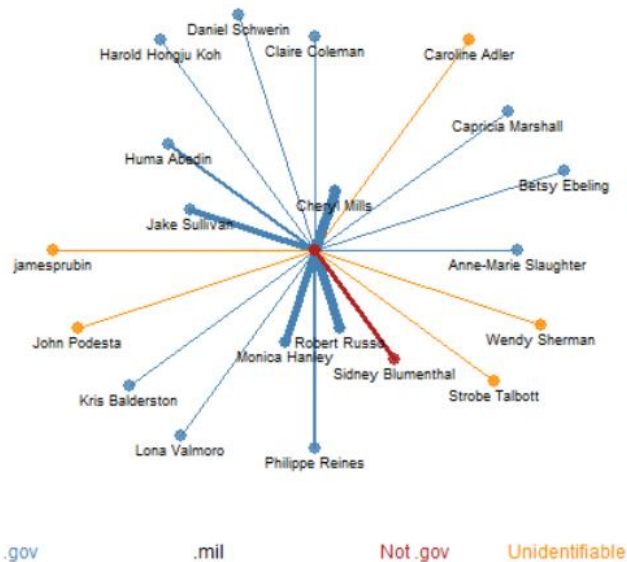
- Event seems unlikely by chance...
 - What is happening around the event?

Date Range:

22 Oct 2012 - 17 Nov 2012



Inner Circle by Volume of Communication (22 Oct 2012 - 17 Nov 2012)



20 Highest TF-IDF Terms

birthday, fbi, cantor, pager, tx, petraeus, happi, oct, scandal, hama, casualti, resign, agent, **sharia**, nla, awar, **ansar**, investig, fyi, mangoush

20 Highest Frequency Terms

forc, fbi, petraeus, israel, investig, secur, nation, cantor, polit, day, obama, report, sourc, al, mr, attack, inform, isra, year, peopl



Confirmation Bias

- We should be extra careful with such a polarized topic
- Perform a coarse Google search, maybe something controversial is always going on?
 - 10 month year combinations
 - First page of results examined
 - Verdict: many months had controversial headlines, none related to the emails except for December 2012



"March 2009" news "State Department"



Conclusions

- Data has inherent limitations
 - Context and expertise are critical
- Statisticians as facilitators
 - Assisting experts and interested individuals
- Look at the data for yourself:
 - <https://github.com/Salahub/Clinton-Emails>



Thanks for listening! Any questions?

Who's Who: The Spiral Network Plot

Inner Circle by Volume of Communication (04 Jan 2009 - 01 Feb 2013)



Huma Abedin
Photo: Maggie Hallahan



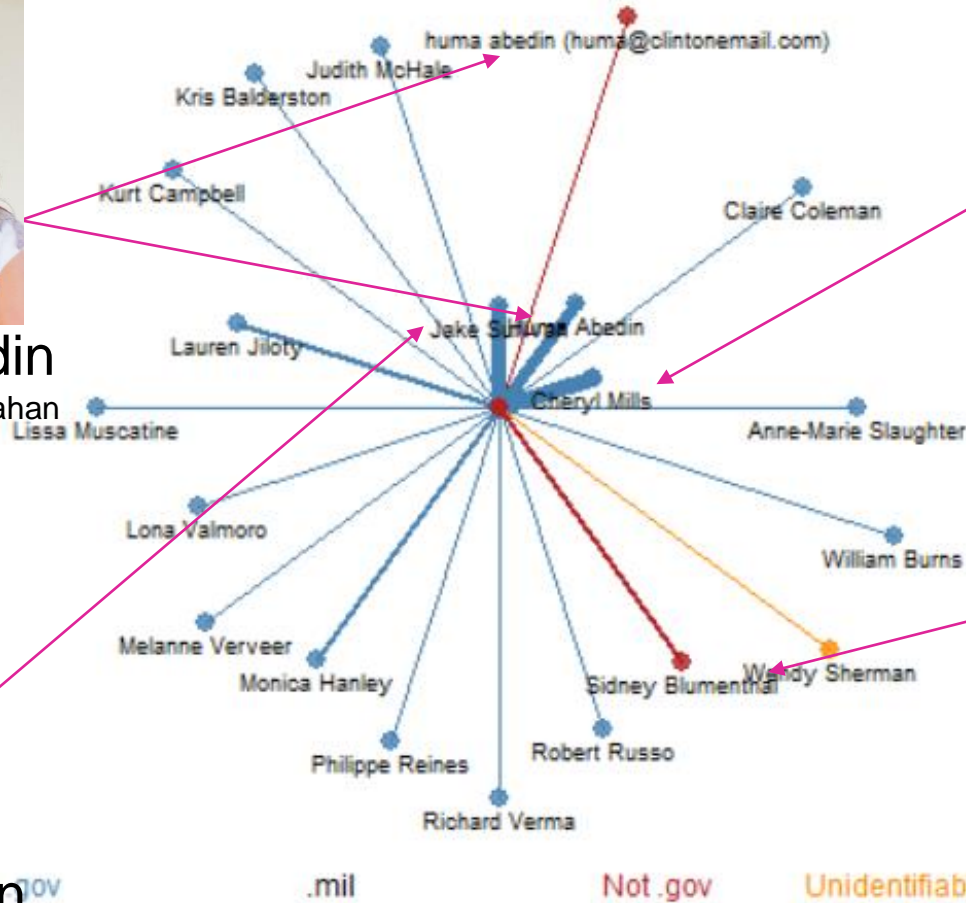
Cheryl Mills
Photo: US State Dept.



Jake Sullivan
Photo: The White House

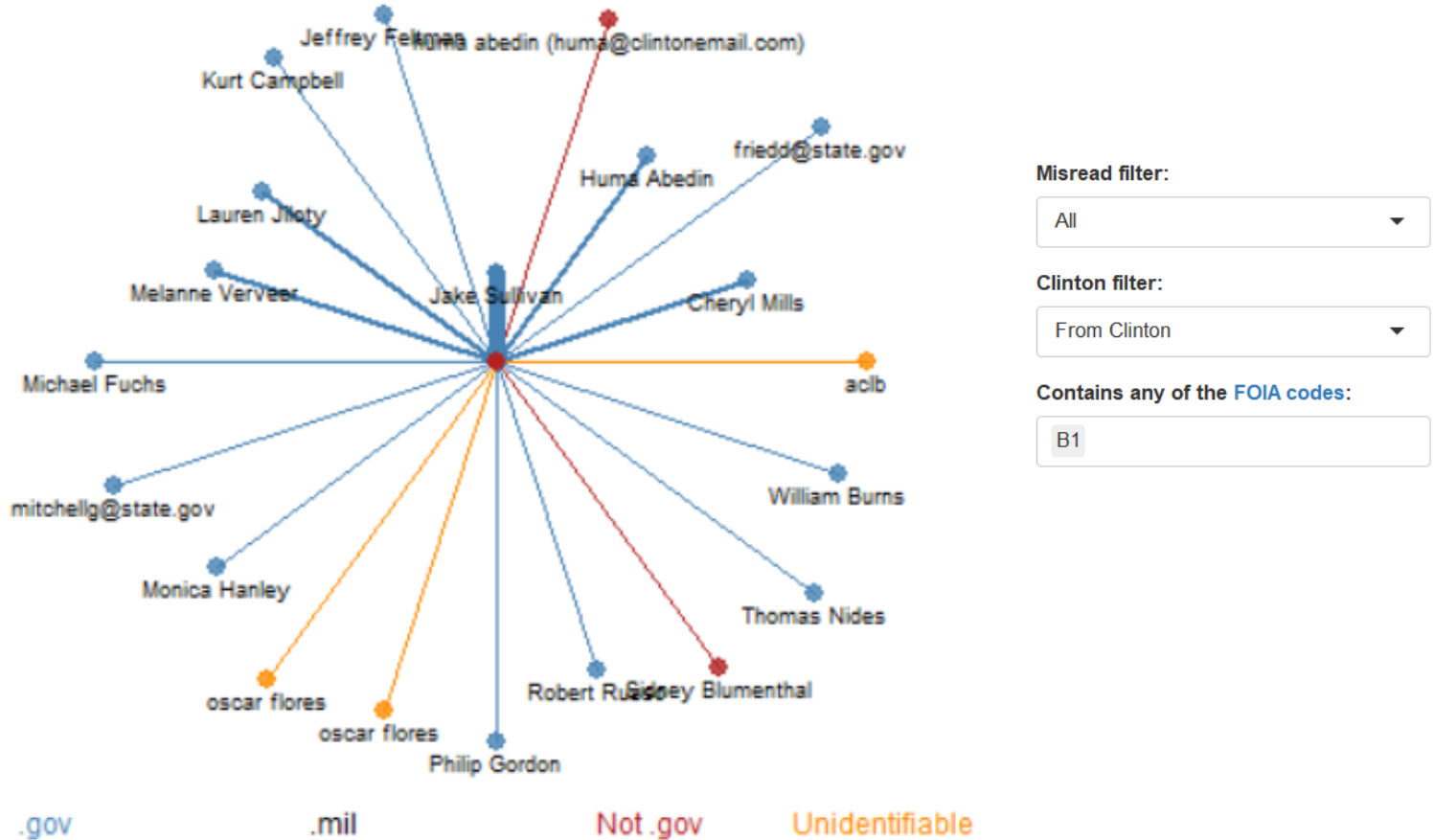


Sidney Blumenthal
Photo: Larry D. Moore



A More Interesting Who's Who

Inner Circle by Volume of Communication (04 Jan 2009 - 01 Feb 2013)



A More Interesting Who's Who

(b)(1) (A) specifically authorized under criteria established by an Executive order to be kept secret in the interest of national defense or foreign policy and (B) are in fact properly classified to such Executive order

Source: <https://vault.fbi.gov/explanation-of-exemptions>

Misread filter:

All

Clinton filter:

From Clinton

Contains any of the FOIA codes:

B1

Inner Circle by Volume of Communication (04 Jan 2009 - 01 Feb 2013)

