

AU Datathon 2020

Stonewall 50 Collection

or

“We had some problems.”

Or

“I want to be the first to throw a brick at this dataset.”

Research Questions:

- What stories about the Stonewall uprising circulated around its 50th anniversary in 1969?
- How was the anniversary commemorated?
- Who is identified as important to the Stonewall story?
 - named entity recognition
- Can we identify competing narratives?
 - topic modeling

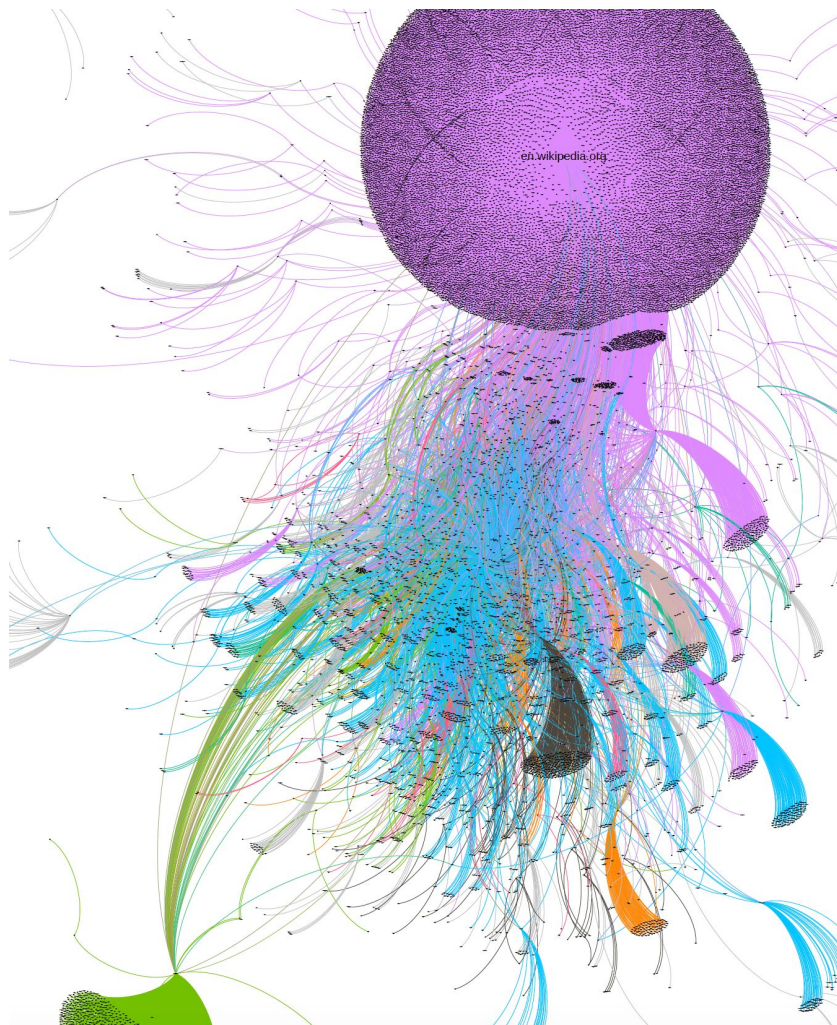
What went well?

- Team members leveraged tools, techniques and resources from their existing repertoires in combination with AUL tools
- Ian and Nick helped a lot with AU-tool tips and outputs



Collected by: Columbia University Libraries

Collector: Columbia University Libraries



[illegible]

Subsets of Data: content of all *Making Gay History* episodes/minisodes on Stonewall



Subsets of Data: content of all *Making Gay History* episodes/minisodes in dataset



Subsets of Data: corpus of *Making Gay History* episodes/minisodes on Stonewall versus all episodes/minisodes as reference corpus

Keyword Types: 21			Keyword Tokens: 6440		Search Hits: 0
Rank	Freq	Keyness	Effect	Keyword	
1	331	+ 173.99	0.0124	em	
2	347	+ 104.79	0.0129	stonewall	
3	119	+ 60.23	0.0045	narration	
4	98	+ 49.68	0.0037	craig	
5	98	+ 43.95	0.0037	christopher	
6	81	+ 43.61	0.0031	cr	
7	183	+ 35.98	0.0069	street	
8	2828	+ 35.21	0.09	the	
9	90	+ 29.52	0.0034	village	
10	54	+ 29.06	0.002	kl	
11	54	+ 26.77	0.002	rodwell	
12	53	+ 22.75	0.002	greenwich	
13	87	+ 22.68	0.0033	march	
14	49	+ 21.96	0.0019	crowd	
15	1368	+ 21.46	0.0475	of	
16	341	+ 21.2	0.0127	people	
17	42	+ 21.06	0.0016	avenue	
18	39	+ 20.99	0.0015	mm	
19	39	+ 20.99	0.0015	sr	
20	57	+ 20.55	0.0022	riots	
21	82	+ 19.93	0.0031	liberation	

Subsets of Data: corpus of *Making Gay History* episodes/minisodes on Stonewall versus all episodes/minisodes as reference corpus

Keyword Types: 21		Keyword Tokens: 6440		Search Hits: 0
Rank	Freq	Keyness	Effect	Keyword
1	331	+ 173.99	0.0124	em
2	347	+ 104.79	0.0129	stonewall
3	119	+ 60.23	0.0045	narration
4	98	+ 49.68	0.0037	craig
5	98	+ 43.95	0.0037	christopher
6	81	+ 43.61	0.0031	cr
7	183	+ 35.98	0.0069	street
8	2828	+ 35.21	0.09	the
9	90	+ 29.52	0.0034	village
10	54	+ 29.06	0.002	kl
11	54	+ 26.77	0.002	rodwell
12	53	+ 22.75	0.002	greenwich
13	87	+ 22.68	0.0033	march
14	49	+ 21.96	0.0019	crowd
15	1368	+ 21.46	0.0475	of
16	341	+ 21.2	0.0127	people
17	42	+ 21.06	0.0016	avenue
18	39	+ 20.99	0.0015	mm
19	39	+ 20.99	0.0015	sr
20	57	+ 20.55	0.0022	riots
21	82	+ 19.93	0.0031	liberation



hosted by Eric Marcus
(‘EM’ in the transcript for
Stonewall episodes,
‘Eric’ in the transcripts
for other episodes)

Subsets of Data: content of all *Making Gay History* episodes/minisodes on Stonewall

Concordance Hits:

- “marsha” (i.e. Marsha P. Johnson) = 55
- “sylvia” (i.e. Sylvia Rivera) = 47
- “craig” (i.e. Craig Rodwell) = 98

Subsets of Data: content of all *Making Gay History* episodes/minisodes in dataset

Concordance Hits:

- “marsha” (i.e. Marsha P. Johnson) = 259
- “sylvia” (i.e. Sylvia Rivera) = 164
- “craig” (i.e. Craig Rodwell) = 102

Hit	KWIC	File	
133	interview at the beginning of Season 2 of Making	Gay History, but we know thousands of you have	makinggayhist
134	Brody. Thanks Hal! Stay in touch with Making	Gay History by signing up for our newsletter at	makinggayhist
135	. Thank you, Michael! Stay in touch with Making	Gay History by signing up for our newsletter at	makinggayhist
135	Brody. Thanks Hal! Stay in touch with Making	Gay History by signing up for our newsletter at	makinggayhist
136	your generous donation. Stay in touch with Making	Gay History by signing up for our newsletter at	makinggayhist
138	Matilda! EM Narration: Stay in touch with Making	Gay History by signing up for our newsletter at	makinggayhist
139	Stonewall, here is a handy factsheet that Making	Gay History co-produced. The final page has more	makinggayhist
139	Stonewall, here is a handy factsheet that Making	Gay History co-produced. The final page has more	makinggayhist
140	which they lived, check out the following Making	Gay History episodes and the accompanying episode not	makinggayhist
140	which they lived, check out the following Making	Gay History episodes and the accompanying episode not	makinggayhist
141	City. I'm Eric Marcus. This is Making	Gay History. Exactly 50 years ago, cops raided the St	makinggayhist
141	. It's also an audio book, and Making	Gay History fans may recognize some of the voices	makinggayhist
142	you will. So this is about how Making	Gay History got its name. I wrote a book 30	makinggayhist
154	bit of Making Gay History backstory, or Making	Gay History history, if you will. So this is	makinggayhist
154	from us for this landmark season of Making	Gay History. I'd like to say "Hi and	makinggayhist
153	lot this season. Pick a significant moment in	gay history in the 1960s or '70s, Craig was	makinggayhist
154	lot this season. Pick a significant moment in	gay history in the 1960s or '70s, Craig was	makinggayhist
155	: I'm Eric Marcus and this is Making	Gay History. In the early morning hours of June 28, 1	makinggayhist
155	who among the agitators we featured on Making	Gay History inspires them, and I'm going to	makinggayhist
154	theme music was composed by Fritz Meyers. Making	Gay History is a co-production of Pineapple Street	makinggayhist
155	theme music was composed by Fritz Meyers. Making	Gay History is a co-production of Pineapple Street	makinggayhist
156	theme music was composed by Fritz Meyers. Making	Gay History is a co-production of Pineapple Street	makinggayhist
157	theme music was composed by Fritz Myers. Making	Gay History is a co-production of Pineapple Street	makinggayhist
158	reating those hateful newspaper headlines. Making	Gay History is a co-production of Pineapple Street	makinggayhist
159	stories about his life in the movement. Making	Gay History is made possible by support from the	makinggayhist
160	AIDS on May 14, 1992. He was 41 years old. Making	Gay History is made possible by support from the	makinggayhist
160	a monthly gay dinner table. Hi Kay! Making	Gay History is made possible by support from the	makinggayhist
160	one of the organizers of that march. Making	Gay History is made possible by support from the	makinggayhist
160			makinggayhist

podcast boilerplate

us honor their memory. #Stonewall50 pic.twitter.com/Zx1o2oXLZy View photo · Entertainment Weekly @EW Jun 5 Replying to @EW On the anniversary of the 1969 Stonewall riots, Ruby Rose and 5 other LGBTQ trailblazers share their stories and celebrate the heroes who fought for equality: share.ew.com/eUDA66H #Stonewall50 pic.twitter.com/aXgrOA9Doz View conversation · NYPD NEWS @NYPDnews Jun 6 Watch as @NYPDONeill makes a statement today about the events at the Stonewall Inn, 50 years ago. #Stonewall50 #WorldPride2019 #Pride2019 pic.twitter.com/LBAMPoXCLi View photo · Madonna @Madonna Jun 3 For once the rumours are true..... Madame X is coming to Pride Island..... June 30th 🏳️🌈🏳️🌈 @NYCPride #WorldPrideNYC #Stonewall50 #WorldPride2019 #prideisland pic.twitter.com/VD5BhPkQYJ View photo · Stacey Abrams @staceyabrams Jun 7 🏳️🌈 The spirit of #Stonewall50 lives in all of us, and I will continue to be your ally every step of the way, using my voice to lift up others. Happy #Pride, and see you in the parade! 🏳️🌈 pic.twitter.com/INb0apRWZ6 View photo · Human Rights Campaign @HRC Jun 4 "I want everyone -- regardless of race, sexuality, gender, ability, or gender identity -- to have the same rights." Longtime @HRC supporter @BillieJeanKing talks to @thedailybeast about #Stonewall50. thedailybeast.com/billie-jean-ki... View summary · The Daily Beast @thedailybeast Jun 8 On #Stonewall50, "Star Trek: Discovery" star @wcruz73 says: "I am inspired by all the LGBTQ people of color who ignited the revolution" trib.al/gv5awBy View summary · i-D @i-D Jun 4 50 years on from the Stonewall riots, circumstances for NYC's LGBT+ culture have changed exponentially. 🙌🏳️🌈 i-D speaks to eyewitnesses and members of the elderly LGBT+ community at New York's Sage Centre. ♥ #Stonewall50 Watch the full film here:youtu.be/zHK2WwINSic pic.twitter.com/o3gtCh2Yie View photo · The Daily Beast @thedailybeast Jun 10 Actor @Alancumming on #Stonewall50: "I don't see why I should thank someone for treating me equally" trib.al/zlYlj3r View summary · SteffyAmondiMS @SteffyAmondi_MS 11m Happy Pride Month! This year, we celebrate the 50th anniversary of the Stonewall uprising - a significant moment that forever changed the LGBTQ Community. #Stonewall50 pic.twitter.com/GY9divzR3P View photo · The Daily Beast @thedailybeast Jun 3 EXCLUSIVE: @realamberheard on #Stonewall50: "I would love to see a world not needing the protection o

We now know that there are nearly 29K (!!!) rainbow emojis within the Stonewall 50 Collection. Thanks, Nick!

What didn't go so well?


Challenges

- Internet connectivity in general
- Ambitions outstripped time and technical resources
- Lots of noise in the dataset (but Alex Thurman kept us from going down rabbit holes to nowhere)
- The size of the dataset prevented many tools and methods from working properly without further filtering and cleaning... requiring more time and technical resources.

```
[ ] tokenizer = Tokenizer(inputCol="content", outputCol="tokenized_content")
countTokens = udf(lambda tokenized_content: len(tokenized_content), IntegerType())

[ ] makinggayhistory_webpages_content_tokenized = tokenizer.transform(makinggayhistory_webpages_content)\
    .select("url", "content", "tokenized_content")\
    .withColumn("tokens", countTokens(col("tokenized_content")))

makinggayhistory_webpages_content_tokenized.show()
```

 -----
AttributeError Traceback (most recent call last)
<ipython-input-55-50c9a70d42ec> in <module>()
----> 1 makinggayhistory_webpages_content_tokenized = tokenizer.transform(makinggayhistory_webpages_content)\n 2\n 3 makinggayhistory_webpages_content_tokenized.show()

----- 2 frames -----
/usr/local/lib/python3.6/dist-packages/pandas/core/generic.py in __getattr__(self, name)
5177 if self._info_axis._can_hold_identifiers_and_holds_name(name):
5178 return self[name]
-> 5179 return object.__getattribute__(self, name)
5180
5181 def __setattr__(self, name, value):

AttributeError: 'Series' object has no attribute '_jdf'

[SEARCH STACK OVERFLOW](#)

What would we do with more time?

What would we do with more time?

- Clean up the data *a lot*
- Topic model for competing narratives
- Compare to older sets of data (e.g. Sarah's queer Geocities data)
- Named entity recognition for key individuals, places, and organizations involved in commemoration events and historical essays and testimony
- Really, ideally getting the Notebooks working a bit more smoothly with some better-filtered and better-formatted data.

laad 39m On #Stonewall50 we reflect on how far we've come and the fight ahead. Together
th 50 years ago DATE today DATE , a raid on the Stonewall Inn sparked the # Sto
ut the LGBT+ GPE community owes them a huge debt of gratitude. 🏳️🌈 Thank you 🏳️🌈
started 50 years ago DATE today DATE , are a lesson about protest in the fight for fr
wall50 pic.twitter.com/nLwMw6ttjw MONEY View photo · Dustin Lance Black @DLanceBl
orked to organize our LGBTQ+ movement. We must also be thankful for the movements tha
KjL1wxu View photo · Fflur Arwel PERSON @FflurGwen 4h CARDINAL We stand on
the liberation and rights of lgbt people. 🙌🏳️🌈 #Stonewall50 pic.twitter.com/8eVe0dgagg Vie
ility and diversity We can only protect those rights with eternal vigilance The struggle for LG
a bunch of LGBTQ ORG people fought back against discrimination, prejudice and hate