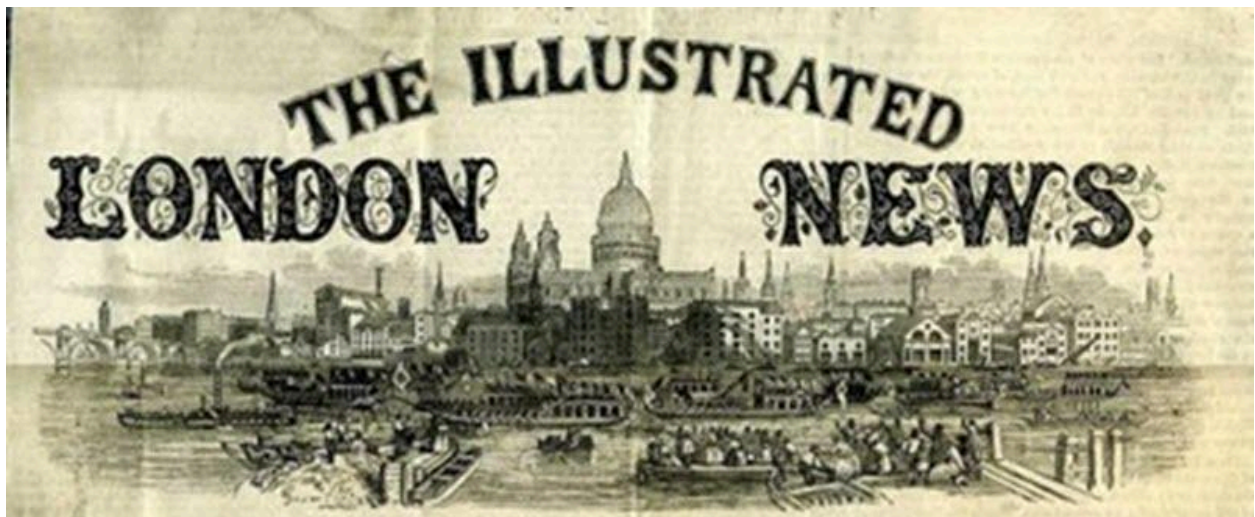


Tracking Archaeology in *The Illustrated London News*

Sarah Ketchley



This digital project was prompted by my interest in the reporting of archaeology in *The Illustrated London News*, a publication notable for its fine illustrations and contributions by some of the preeminent figures of the day. [Gale Primary Sources](#) offers access to the entire run of the newspaper, covering the period 1842–2003. This project essay describes my workflow for the preliminary investigation of the data: initial content set creation, cleaning, analysis, export, and visualization. At the outset, the research questions were necessarily broad:

- Which words were most prevalent in articles reporting about archaeological digs?
- What themes or topics are most prevalent in the dataset?
- What was the overall feeling about this type of reporting? Was it reported favorably?
- Is it possible to identify which archaeologists were directly contributing to the publication and how many contributions they made?

Engaging in the practical process of curation and analysis offers opportunities to refine these questions, and almost inevitably suggests new avenues for future exploration.

Building a Dataset

I built my dataset by searching the [Gale Primary Sources](#) archive using Advanced Search in [Gale Digital Scholar Lab](#) for keywords, including “archaeology,” which appears for the first time in 1881; “excavations”; and “ruins.” I also searched by site, excavator, and civilization; for example, “Layard,” “Assyria,” “Nineveh,” “Sumerian,” “Egypt,” “tomb,” etc. While the results weren’t comprehensive, I ended up with an initial content set of 2,513 primary source documents—mostly newspaper articles with some advertisements. This corpus comprised a collection of optical character recognition (OCR) texts generated from original scans of the ILN Historical Archive, accessed via the [Gale Digital Scholar Lab](#) and [Gale Primary Sources](#).

I wanted to examine the content of the documents to see if I could identify recurrent themes or topics, along with the most common words and expressions of positive or negative sentiment in the dataset. I opted to work with Tableau to generate multiple visualizations for display on an interactive dashboard. To create the statistical data underlying these visualizations, I used the Lab to run a series of text-mining analyses with the tools available in the platform using the following workflow:

1. In order to perform statistical analysis on my content set, I needed to **clean** the texts to remove recurrent OCR errors, and also to remove stop words, which are the most common words in the English language and not of interest for my research. I ran a preliminary nGrams analysis to identify prevalent OCR errors: I configured the tool to return only unigrams (single words), and exported the results as a CSV. I was able to identify and weed out some regularly occurring OCR errors reflected in the CSV, which I pasted into the stop word list in the Clean tool. I continued to iterate on this process, removing and correcting the base OCR texts.

Once I created a clean configuration that removed most of the most common OCR text errors, I ran the collected dataset through my analysis tools.

2. Topic Modeling: The Lab uses a tool called MALLET to perform LDA topic modeling on the corpus of texts. The algorithm iterates through the “bag of words,” or collected textual data, and identifies terms that are topically similar, then groups them together. The configuration I chose to run was 30 topics, each containing 20 topic terms, while also applying the cleaning configuration I’d created. I chose this number of topics because I wanted to move beyond what the algorithm would find most obvious, and instead discern connections that were less apparent in the dataset. I exported the results of this analysis run as a CSV, as well as a second “topic proportion by document” analysis spreadsheet, which I didn’t end up using for this run of visualizations but which nonetheless provides a wealth of granular detail about my documents.

	A	B	C	D	E	F	G
1	topic_name	topic_id	topic_allocation_count	topic_allocation_ratio	topic_coherence	topic_corpus_dist	topic_document_entropy
2	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
3	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
4	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
5	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
6	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
7	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
8	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
9	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
10	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
11	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
12	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
13	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
14	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
15	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
16	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
17	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
18	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
19	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
20	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
21	Rome and Roman	0	0.0784	0.006	-399.0384	3.4582	5.5557
22	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
23	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
24	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
25	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
26	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
27	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
28	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
29	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
30	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
31	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
32	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
33	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
34	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
35	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
36	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472
37	London Institutions	1	0.1574	0.2341	-435.9915	2.8971	5.8472

Gale Digital Scholar Lab Topic Modeling CSV Output.

I examined individual articles returned in the results list using the Document Terms output alongside the Documents by Topic pop-up.

Topic 19		TERMS	COUNT	PROBABILITY	DOCS
IDENTIFIED IN		tomb	185	0.0331	35
57 Documents		sarcophagus	56	0.01	15
TOPIC MEASURES		royal	53	0.0095	17
Tokens	5583	Egypt	45	0.0081	26
Document Entropy	3.4009	Carter	44	0.0079	8
Average Word Length	6.5	funerary	41	0.0073	15
Coherence	-246.6864	scenes	40	0.0072	10
Uniform Distance	3.4393	Egyptian	40	0.0072	23
Corpus Distance	2.5642	Horemheb	39	0.007	5
Exclusivity	0.6908	Akhenaten	37	0.0066	13
		Pharaoh	36	0.0065	20
		chamber	36	0.0065	16
		burial	36	0.0065	15
		Amarna	32	0.0057	12
		room	31	0.0056	7
		coffin	31	0.0056	8
		Dynasty	27	0.0048	11
		corridor	26	0.0047	6
		rooms	25	0.0045	6
		King	25	0.0045	15

Documents associated with...

Topic 19

Includes term: tomb

Showing documents 10/35

- A Refuge for Christians
- Books of the Day
- Finds from Biblical Lachish
- Ras Shamra Yields New Treasure to the Spade
- Men who perform the "spade work" of history: British names famous in the field of Archaeology
- Digging up the past
- Archæology of the Month
- The Making of an Archæologist
- The Fate of Jericho Revealed by the Spade
- A New Page Opened in Ancient History

3. **Sentiment Analysis:** Gale Digital Scholar Lab visualizes sentiment across time using the AFINN sentiment lexicon, which ranks documents as positive, neutral, or negative on a scale of +5 to -5, based on the words included in the text. I ran this tool on my cleaned dataset and exported the results as a CSV.

	A	B	C	D	E
1	docid	document title	publication year	sentiment score	relative mean score
2	HN3100297680	Archv¶ological Discovery for Years	1933	2.4	3.0687444
3	HN3100523109	Awards	1984	2.3888888	3.0508146
4	HN3100522845	Hotel Guide	1984	2.3333333	2.9611664
5	HN3100520327	Archaeology Awards	1978	2.2162163	2.772179
6	HN3100105681	Archv¶ology	1879	2.1666667	2.6922224
7	HN3100559116	in the Field of Mayan Archv¶ology: One of	1947	2.142857	2.6538014
8	HN3100304840	Animal Sculpture in	1935	2.1363637	2.6433234
9	HN3100283757	and Other	1930	2.1176472	2.613121
10	HN3100183131	The King's Tour	1903	2.1052632	2.5931375
11	HN3100256276	Fashions and Fancies	1923	2.032258	2.4753315
12	HN3100359177	Nimrud Looked like	1951	2	2.4232779
13	HN3100305298	Bronze Age Pottery	1935	2	2.4232779
14	HN3100279345	News	1928	2	2.4232779
15	HN3100538155	Original Monument in Bronze Has Lately	1887	2	2.4232779
16	HN3100294863	Fountain-Head of	1932	2	2.4232779
17	HN3100360321	Civilisation in	1951	1.9180328	2.2910101
18	HN3100553758	of Ancient Sculpture	1926	1.9166666	2.2888057
19	HN3100446399	Afghanistan	1879	1.882353	2.233435
20	HN3100413042	Discoveries at Igbo	1970	1.88	2.2296379
21	HN3100408282	Greek Coinage...	1966	1.875	2.2215695
22	HN3100294412	Rome: Relics of	1932	1.8695652	2.2127998
23	HN3100274467	News	1927	1.8571428	2.192754

I anticipated that visualizing this data would give me a sense of how archaeology and archaeological reporting was presented in a popular publication; ideally, I'd like to compare this with other contemporary newspaper reporting while also taking into account authorship—whether the material was written by an archaeologist or by a staff reporter.

4. **Clustering:** This analysis is carried out using the k-means clustering algorithm. The documents were grouped in 20 clusters, according to the algorithm's ranking of proximity or similarity in document content or other factors, and I again downloaded the CSV output.

	A	B	C
1	cluster	docid	document title
2	Cluster-1	HN3100249693	Meiringen
3	Cluster-1	HN3100502954	Dealing with the Oriental Cylinder Seals, Appearing in Our Last Issue
4	Cluster-1	HN3100429318	Lectures
5	Cluster-1	HN3100120034	to a Fellowship at Lincoln College; and Mr. W. M. Ramsay, M. A., Fellow of
6	Cluster-1	HN3100042114	Nineveh Antiquities
7	Cluster-1	HN3100505169	The New Year Honours
8	Cluster-1	HN3100555365	(illustrated in Our Issue of Nov. 26) Found at Beisan during Excavations by the
9	Cluster-1	HN3100499709	A Window on the World
10	Cluster-1	HN3100135502	CausVC, upon Mr. Augustus Wollaston Franks, M. A., of Trinity College, C. B., F. R.
11	Cluster-1	HN3100073518	S., Keeper of British and MediVlval Antiquities. &c., at the British Museum; and
12	Cluster-1	HN3100186253	A Bronze Money-chest Has Lately Been Discovered in the Excavations at Pompeii
13	Cluster-1	HN3100544733	Roman Military Antiquities in Scotland: Fortifications and Military Pits
14	Cluster-1	HN3100000160	The Ears of a God: Wonderful Prayer-Tablets and Other Discoveries at Memphis
15	Cluster-1	HN3100000160	Saturday Last
16	Cluster-1	HN3100194222	Latest Excavations in the Sinai Peninsula: Monuments from 4600 B. C.
17	Cluster-1	HN3100096092	ArchVlology
18	Cluster-1	HN3100522845	Shetland's Own Good Hotel Guide
19	Cluster-1	HN3100259126	Grindel-Wald
20	Cluster-1	HN3100116524	A Buried City of the Exodus
21	Cluster-1	HN3100418510	Swans Art Treasures Tours
22	Cluster-1	HN3100058333	Eleven Inches High, Having upon It the Name of Xerxes in Two Languages?one
23	Cluster-1	HN3100428727	Explore
24	Cluster-1	HN3100396843	TV Schweppsicolor
25	Cluster-1	HN3100087381	Calendar for the Week Ending June 8
26	Cluster-1	HN3100065694	Flags of Which Were Scattered about More Than 200 Skeletons

I found this output more challenging to visualize in a meaningful way, and ultimately opted to drop clustering in favor of topic modeling, which provided a thematic breakdown sufficiently detailed to be of interest.

5. Finally, I downloaded the **metadata** for all the documents, which included author, title, date and place of publication, and document ID. I wanted to see if I could identify whether archaeologists wrote regularly for the newspaper, and the content and context they provided compared with more formal academic publications. This is where the real challenge began!

Working with Dates in Excel

I looked at the date formats and noted that post-1900 dates were written numerically, while pre-1900 were text. This prompted the methodological question of how to clean up these variances, an issue that was compounded by Excel not recognizing dates before January 1, 1900.

I was able to find a couple useful resources online that helped me solve the problem, but I was surprised by how complex it turned out to be.

This [article](#) provided good background context, while this [post](#) ultimately solved my issue.

I began by manually splitting out the text dates into a separate column.

D	E	F	G	
Title		Publication Date	Publisher	Place
ted London News.	April 8, 1882	April 8, 1882		Londc
ted London News.		7/22/1961		Londc
ted London News.		8/17/1935		Londc
ted London News.		2/6/1937		Londc
ted London News.		1/29/1938		Londc
ted London News.	February 3, 1877	February 3, 1877		Londc
ted London News.		10/27/1973		Londc
ted London News.		11/27/1982		Londc
ted London News.		2/13/1960		Londc
ted London News.		4/26/1986		Londc
ted London News.		3/27/1937		Londc
ted London News.		9/12/1964		Londc
ted London News.		10/3/1970		Londc
ted London News.		8/2/1958		Londc
ted London News.	August 7, 1847	August 7, 1847		Londc
ted London News.		12/3/1960		Londc
ted London News.	March 10, 1888	March 10, 1888		Londc
ted London News.	June 24, 1876	June 24, 1876		Londc
ted London News.		7/28/1984		Londc
ted London News.		10/29/1983		Londc
ted London News.		9/1/1962		Londc
ted London News.		3/17/1923		Londc
ted London News.		4/19/1930		Londc
ted London News.		5/25/1968		Londc
ted London News.	November 5, 1870	November 5, 1870		Londc
ted London News.		10/8/1966		Londc
ted London News.		8/15/1925		Londc
ted London News.		2/3/1951		Londc
ted London News.		5/16/1970		Londc
ted London News.		7/12/1969		Londc
ted London News.	November 4, 1871	November 4, 1871		Londc
ted London News.		10/27/1928		Londc
ted London News.		2/28/1925		Londc
ted London News.		7/2/1955		Londc
ted London News.		5/21/1938		Londc
ted London News.	August 28, 1852	August 28, 1852		Londc
ted London News.		10/27/1984		Londc
ted London News.		1/22/1949		Londc

I then created an additional three columns in order to perform the text-to-number conversion and to get Excel to appropriately render dates before January 1, 1900.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
23	British	London	00 Jan 1900	00 Jan 1900	TRUE			2/12/1918		London, Unib. William Wat	Illustrated	Lo The Illustrate	GALE HN31000312		
24	British	London	00 Jan 1900	00 Jan 1900	TRUE			8/17/1957		London, Unib. William Wat	Illustrated	Lo The Illustrate	GALE HN31000312		
25	British	London	00 Jan 1900	00 Jan 1900	TRUE			1864-03-05	March 5, 1864	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100063870		
26	British	London	00 Jan 1900	00 Jan 1900	TRUE			2/24/1923		London, Unib. Percy S. New	Illustrated	Lo The Illustrate	GALE HN3100025242		
27	British	London	00 Jan 1900	00 Jan 1900	TRUE			1850-06-22	June 22, 1850	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100025088		
28	British	London	00 Jan 1900	00 Jan 1900	TRUE			9/23/1905		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100020384		
29	British	London	00 Jan 1900	00 Jan 1900	TRUE			6/7/1991		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100023979		
30	British	London	00 Jan 1900	00 Jan 1900	TRUE			6/18/1955		London, Unib. Theresa Gee	Illustrated	Lo The Illustrate	GALE HN3100027496		
31	British	London	00 Jan 1900	00 Jan 1900	TRUE			9/16/1967		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100024040		
32	British	London	00 Jan 1900	00 Jan 1900	TRUE			11/21/1936		London, Unib. From the Wa	Illustrated	Lo The Illustrate	GALE HN3100012093		
33	British	London	00 Jan 1900	00 Jan 1900	TRUE			3/12/1938		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100014964		
34	British	London	00 Jan 1900	00 Jan 1900	TRUE			1877-09-01	September 1, 1877	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100010490		
35	British	London	00 Jan 1900	00 Jan 1900	TRUE			9/24/1966		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100054483		
36	British	London	00 Jan 1900	00 Jan 1900	TRUE			1868-02-29	February 29, 1868	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100071364		
37	British	London	00 Jan 1900	00 Jan 1900	TRUE			4/19/1958		London, Unib. Tigueil Yadin	Illustrated	Lo The Illustrate	GALE HN3100085846		
38	British	London	00 Jan 1900	00 Jan 1900	TRUE			12/1/1962		London, Unib. David Oates	Illustrated	Lo The Illustrate	GALE HN3100099021		
39	British	London	00 Jan 1900	00 Jan 1900	TRUE			1877-05-05	May 5, 1877	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100020647		
40	British	London	00 Jan 1900	00 Jan 1900	TRUE			11/27/1965		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100040692		
41	British	London	00 Jan 1900	00 Jan 1900	TRUE			1900-01-01		London, Unib. Margaret Lai	Illustrated	Lo The Illustrate	GALE HN3100020571		
42	British	London	00 Jan 1900	00 Jan 1900	TRUE			3/1/1964		London, Unib. James Melis	Illustrated	Lo The Illustrate	GALE HN3100040913		
43	British	London	00 Jan 1900	00 Jan 1900	TRUE			1/18/1969		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100050916		
44	British	London	00 Jan 1900	00 Jan 1900	TRUE			12/7/1957		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100028445		
45	British	London	00 Jan 1900	00 Jan 1900	TRUE			10/14/1961		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100039466		
46	British	London	00 Jan 1900	00 Jan 1900	TRUE			7/17/1926		London, Unib. Harold J. She	Illustrated	Lo The Illustrate	GALE HN3100027025		
47	British	London	00 Jan 1900	00 Jan 1900	TRUE			5/2/1931		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100056200		
48	British	London	00 Jan 1900	00 Jan 1900	TRUE			8/31/1974		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100041826		
49	British	London	00 Jan 1900	00 Jan 1900	TRUE			6/28/1905		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100062461		
50	British	London	00 Jan 1900	00 Jan 1900	TRUE			11/13/1954		London, Unib. George E. M	Illustrated	Lo The Illustrate	GALE HN3100072279		
51	British	London	00 Jan 1900	00 Jan 1900	TRUE			3/29/1975		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100014099		
52	British	London	00 Jan 1900	00 Jan 1900	TRUE			1850-05-01		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100079514		
53	British	London	00 Jan 1900	00 Jan 1900	TRUE			2/6/1909		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100040474		
54	British	London	00 Jan 1900	00 Jan 1900	TRUE			2/21/1959		London, Unib. Margaret Gu	Illustrated	Lo The Illustrate	GALE HN3100049648		
55	British	London	00 Jan 1900	00 Jan 1900	TRUE			7/5/1930		London, Unib. Edith Hughes	Illustrated	Lo The Illustrate	GALE HN3100036181		
56	British	London	00 Jan 1900	00 Jan 1900	TRUE			7/13/1957		London, Unib. Dr. Ghoshma	Illustrated	Lo The Illustrate	GALE HN3100088078		
57	British	London	00 Jan 1900	00 Jan 1900	TRUE			10/24/1936		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100011801		
58	British	London	00 Jan 1900	00 Jan 1900	TRUE			5/7/1966		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100048603		
59	British	London	00 Jan 1900	00 Jan 1900	TRUE			11/1/1941		London, Unib. Charles E. By	Illustrated	Lo The Illustrate	GALE HN3100025459		
60	British	London	00 Jan 1900	00 Jan 1900	TRUE			1850-04-06	April 6, 1850	London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100024434		
61	British	London	00 Jan 1900	00 Jan 1900	TRUE			12/3/1966		London, United Kingdom	Illustrated	Lo The Illustrate	GALE HN3100040662		

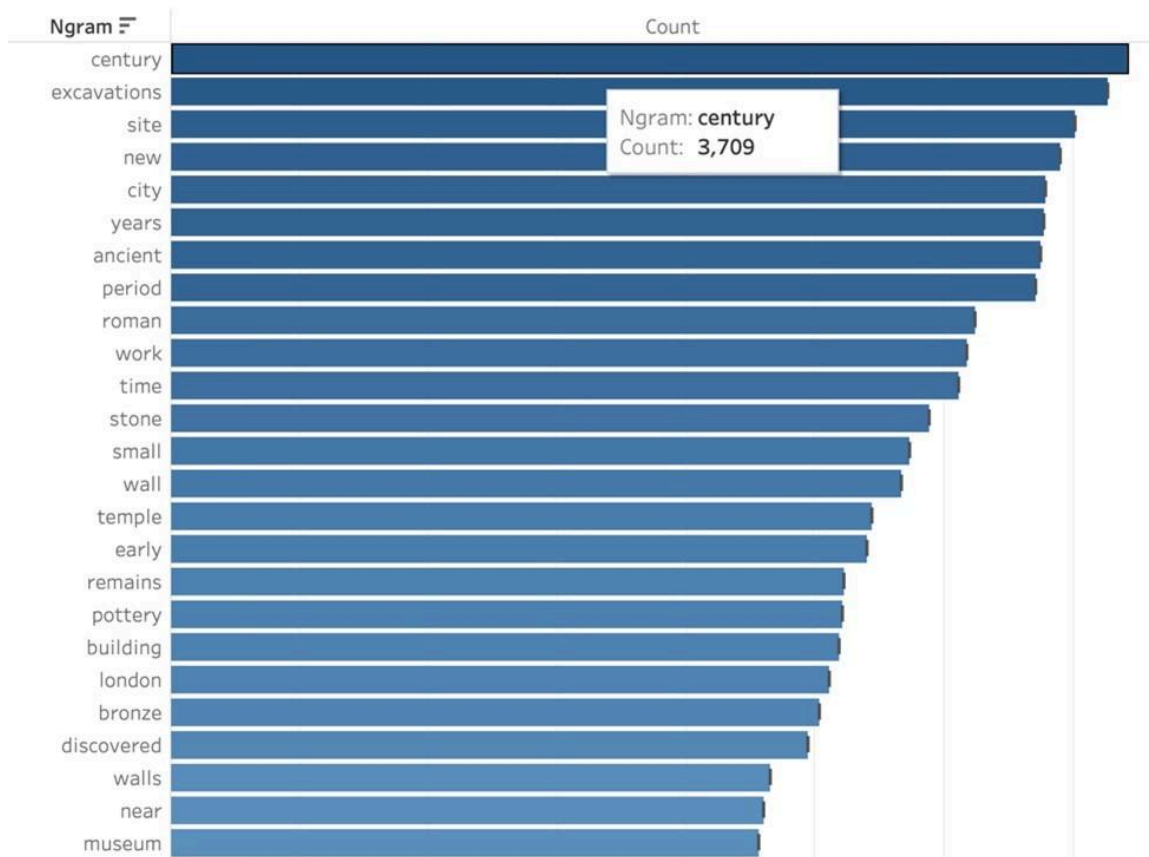
G247 Columns E, F, and G are the added columns, and the final formula is shown in the function bar. One issue that arose was that blank fields were replaced with "1900-01-01," and trying to clean these up with find and replace didn't work. Again, I had to do this manually, but the outcome was a table with all dates standardized in the format YYYY-MM-DD.

Using Exported CSV Data in Tableau

Ngrams

To answer the question "Which words were most prevalent in articles reporting about archaeological digs?" I visualized the nGrams CSV output in Tableau. A bar chart proved to be the most effective visualization.

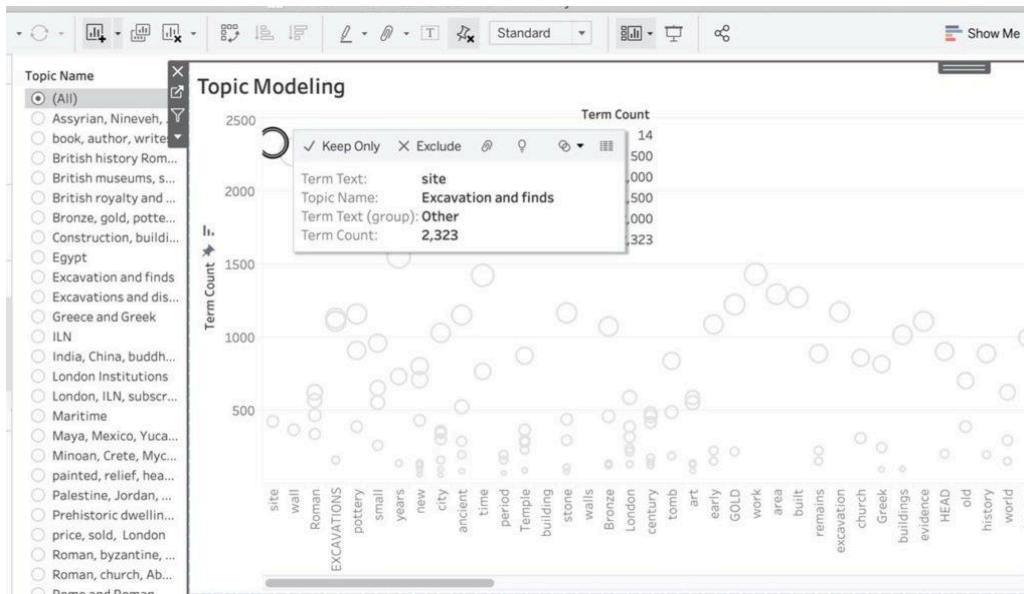
Ngrams



Following data cleanup, the most common terms included “century,” “excavations,” and “site,” which one would probably expect in documents related to archaeology. The first civilization mentioned is “Roman,” and the city is “London,” which, given that the newspaper was published in the UK and Roman archaeological artifacts are regularly found, wasn’t surprising. Further refinements in the terms may yield interesting results about the civilizations most commonly reported on and whether the focus was on the material culture or the process of archaeology itself.

Topic Modeling

I used a circle view in Tableau to visualize my topic modeling analysis to answer the question “What themes or topics are most prevalent in the dataset?” I found the option to display all or to zoom in on a single theme to be most helpful.



Topic modeling is a qualitative analysis, so it's incumbent on the researcher to decide what the connections are between the terms the algorithm comes up with and to then name the topic appropriately. In this case, the most common words are grouped around the theme that I named "excavations and finds." The term displayed above is "site," which occurs 2,323 times in the dataset.

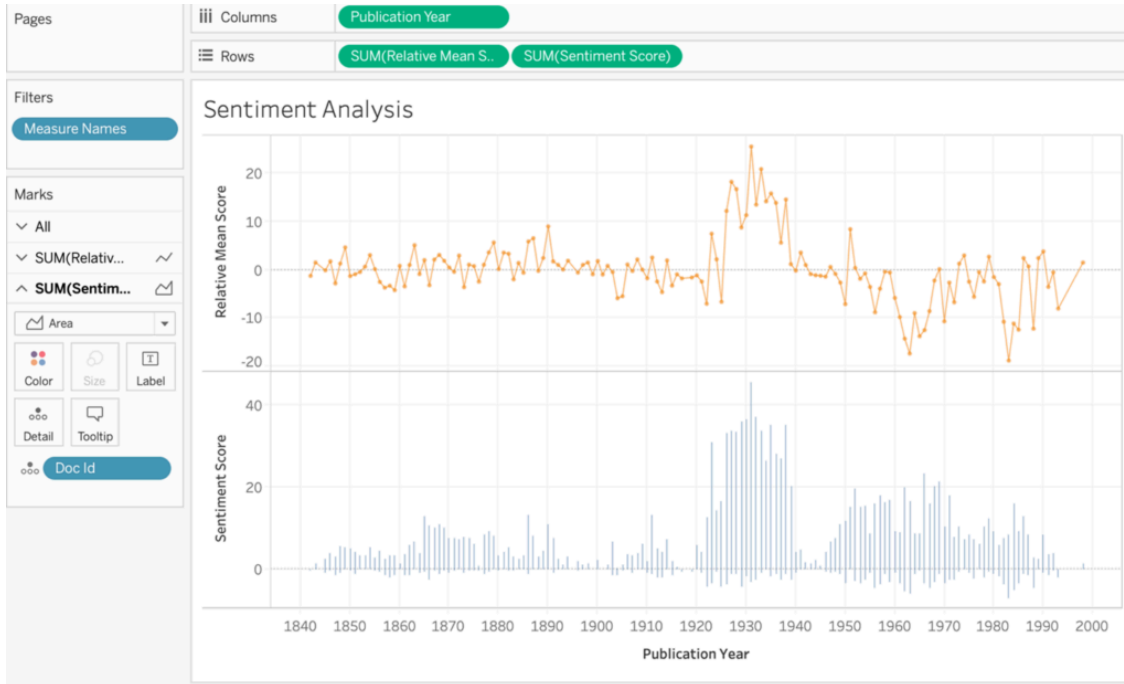
I plan to do more work on this visualization, returning to the [Digital Scholar Lab](#) in conjunction with Tableau to experiment with various topic measures available in the platform. The granularity of this data could allow for more detailed analysis results as I continue to refine the OCR content.

Tables

Abc	Topic Name
#	Topic Id
Abc	Term Text
Abc	term docIds
	Topic Name (group)
	Term Text (group)
Abc	<i>Measure Names</i>
#	Topic Allocation Count
#	Topic Allocation Ratio
#	Topic Coherence
#	Topic Corpus Dist
#	Topic Document Entropy
#	Topic Eff Num Words
#	Topic Exclusivity
#	Topic Rank 1 Docs
#	Topic Token Doc Diff
#	Topic Tokens
#	Topic Uniform Dist
#	Topic Word Length
#	Term Coherence
#	Term Corpus Dist

Sentiment Analysis

I used the sentiment analysis output to answer the questions “What was the overall feeling about this type of reporting?” and “Was it reported favorably?” I imported the CSV and created a sentiment analysis visualization in Tableau using two line charts to show the relative mean sentiment score and the sentiment score. The data was visualized over time, between 1842 and 2003, the full run of the newspaper. On the whole, the sentiment is overwhelmingly positive.



Selecting the most positive report brings up the detail of the point, along with the relevant Gale Doc ID.

Sentiment Analysis



I was then able to go into the Lab to find the document and figure out what made it so positive.

Revelations of the Byzantine Genius: Sculpture; Unique Marble Inlay

Source Citation
"Revelations of the Byzantine Genius: Sculpture; Unique Marble Inlay." Illustrated London News, 11 Apr. 1931, p. 611. The Illustrated London News Historical Archive, 1842-2003.
link.gale.com/apps/doc/HN3100001717/DSLAB?u=nomi&sid=DSLAB&id=084782be. Accessed 17 Dec. 2020.
Gale Document Number:GALE|HN3100001717

ACTIVE CONTENT SET
Select Content Set
MANAGE | ANALYZE

DOC EXPLORER
33 - 33/ 33 Go
Original Document Scan
OCR Text
SEARCH WITHIN
Search within document

REVELATIONS OF THE BYZANTINE GENIUS :
SCULPTURE ; UNIQUE MARBLE INLAY

Document Text
OCR Confidence: 95% Learn how this text was created

REVELATIONS OF THE BYZANTINE GENIUS. SCULPTURE; UNIQUE MARBLE INLAY By COURTESY OF MR. STANLEY CASSON AND MACRIDY BEY. (SEE ARTICLE, PAGE 612.) THESE photographs (numbered in sequence from those opposite) also illustrate Mr. Stanley Casson's article on page 612 describing the very important additions to the world's knowledge of Byzantine art resulting from recent excavations in the double Church of St. Mary Panachrantos at Constantinople. Figs. 5 and 6 above show examples of sculpture and decorative work in marble finer than anything previously known of the periods they represent. But the greatest and most surprising discovery was that shown in Fig. 4. Two small chapels were found hidden in the roof, and near one of them, lying face-downwards on the floor, was a superb icon with a full-length figure of St. Eudokia worked in inlaid marbles of various colours, on a background [Continued below, Cominsea.] of plain white marble, and bearing her name inscribed in Greek. The face and hands are wrought in rose-coloured marble, with the features and detail lightly engraved; while the halo is yellow and the robe purple and green, studded with stones to resemble mosaic. As a specimen of this hitherto unknown method it is absolutely unique, and constitutes one of the most remarkable works of Byzantine art that has ever come to light. We may add that a reproduction of it, H r a O 14, - o FIG. 4. "THE FIRST OF ITS KIND," ILLUSTRATING "A WHOLLY NEW TECHNIQUE IN BYZANTINE ART": AN ICON OF ST. EUKOKIA (WITH HER NAME IN GREEK LETTERS) WROUGHT IN COLOURED MARBLE INLAY "OF EXTRAORDINARY RICHNESS AND BEAUTY," (TENTH CENTURY), I. L. FIG. 5. EXQUISITE BYZANTINE SCULPTURAL DECORATION OF THE TENTH CENTURY - MARBLE MULLIONS OF THE CENTRAL APSE IN THE NORTH CHURCH OF ST. MARY PANACHRANTOS, CONSTANTINOPLE. (SEE FIG. 8, PAGE 618, 14.) FIG. 6. "SUPERB ARCHITECTURAL DECORATION IN MARBLE" REVEALED BY THE REMOVAL OF PLASTER AND WHITEWASH: AN UPWARD VIEW OF THE INTERIOR OF A CUPOLA IN THE NORTH CHURCH. (SEE DETAIL DRAWING, FIG. 7, PAGE 612.) This icon, in the full beauty of its actual colouring, will be given in a future issue of "The Illustrated London News."

Words like "genius," "unique," "important," and "finer" contributed to the positivity of the document. Looking at the period between the world wars, there was a lot of positive reporting on archaeology, and it was a time of great activity and discovery. This visualization did a nice job of capturing the spirit of discovery, while the two line charts gave a good sense of the pervading sentiment in the newspaper.

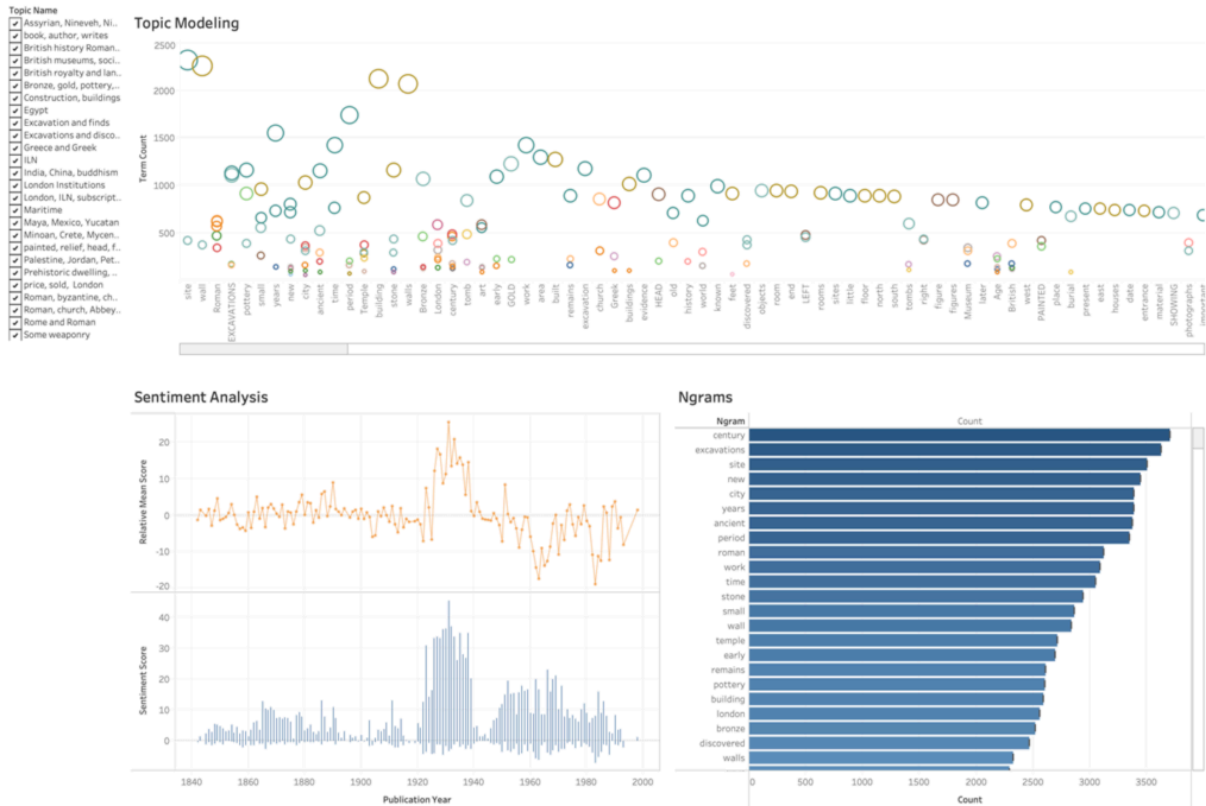
5. My final question pertained to the archaeologists themselves: "Is it possible to identify which archaeologists were directly contributing to the publication and how many contributions they made?"

The answer is yes, but interestingly, fewer of the articles prior to 1900 have an author's name associated with them. I didn't realize that the articles started having bylines after this date until I looked at this visualization. There are a number of well-known archaeologists authoring articles, including Max Mallowan, who excavated in Iraq and was the second husband of Agatha Christie; Howard Carter/Percy Newberry/Harry Burton/A. Mace (Tutankhamun); Henry Frankfort (early Egypt); J. D. S. Pendlebury (Egypt and Crete); John

Garstang (Egypt); and many more. It was fascinating to see how many archaeologists wrote articles for the publication. I plan to compare authorship of similar articles from other contemporary newspapers—for instance, *The Times* of London—to see if the ILN is an anomaly or the norm.

Tableau Dashboard

Tracking archaeology and archaeological reporting in *The Illustrated London News* 1842-2003



My final Tableau dashboard included the Ngrams, sentiment analysis, and topic modeling analyses generated by the raw exported data. Ngrams showed that while the general terms “century,” “excavation,” and “site” were the most common in my dataset, the most-reported excavations were those with Roman finds. This pattern was reinforced when interacting with the topic model: again, the biggest topics included those describing general archaeological terms, while the algorithm also did a good job of grouping by civilization or type of find, including Egyptian, Assyrian, Roman, Greek, Etruscan, and pottery, bronze/metal, tomb, etc., and giving a snapshot of some of the more significant archaeological activities of the day. Finally, sentiment analysis indicated that the period

between the two World Wars was most positive in nature: there were a large number of excavations underway during this period with well-reported finds. While I didn't include my analysis of authors in the final dashboard, I was able to identify the main contributors to the newspaper, who also happened to be some of the more famous practicing archaeologists of the day. This opened a new research angle to me, as I thought about comparing published excavation reports by these individuals and their more popular writing.