

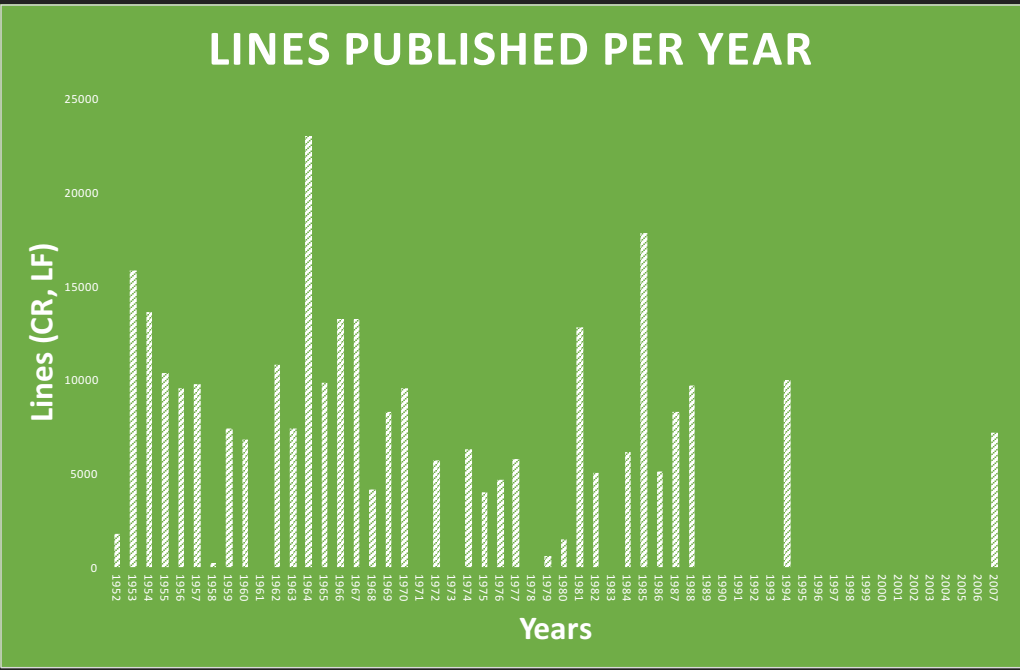
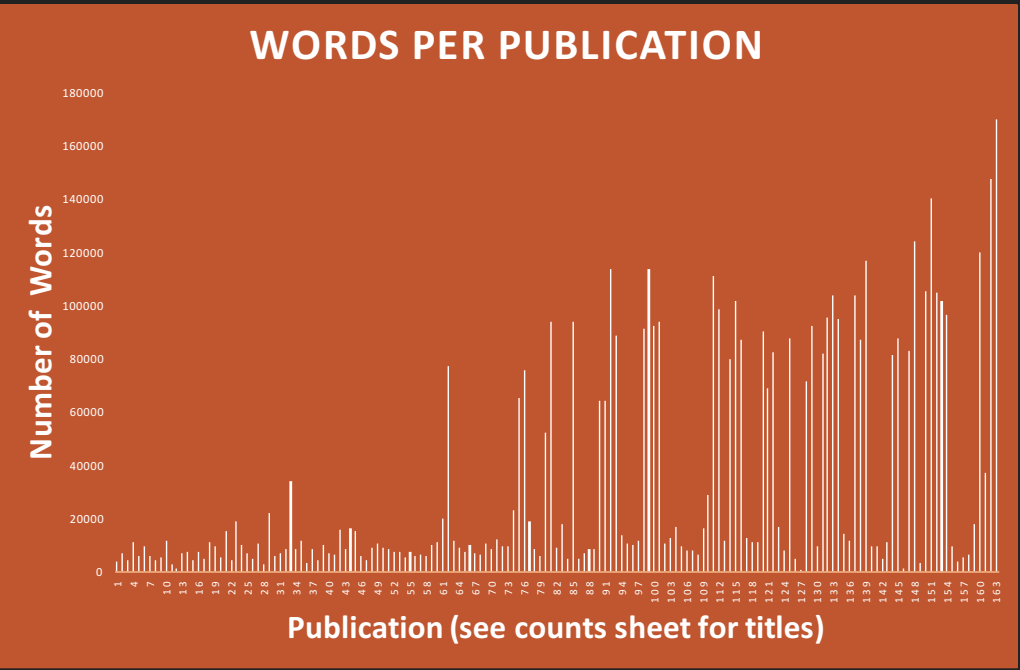
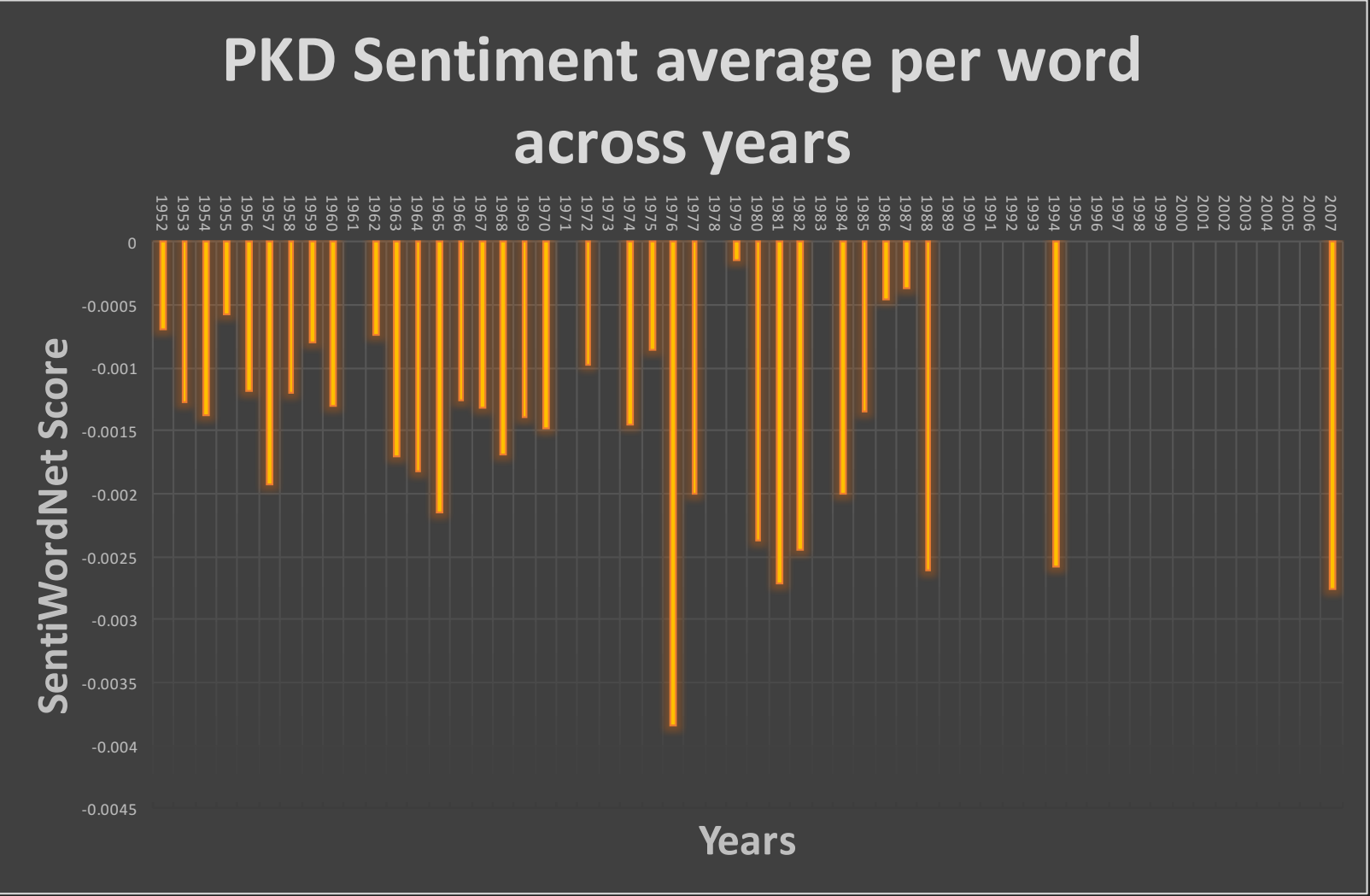
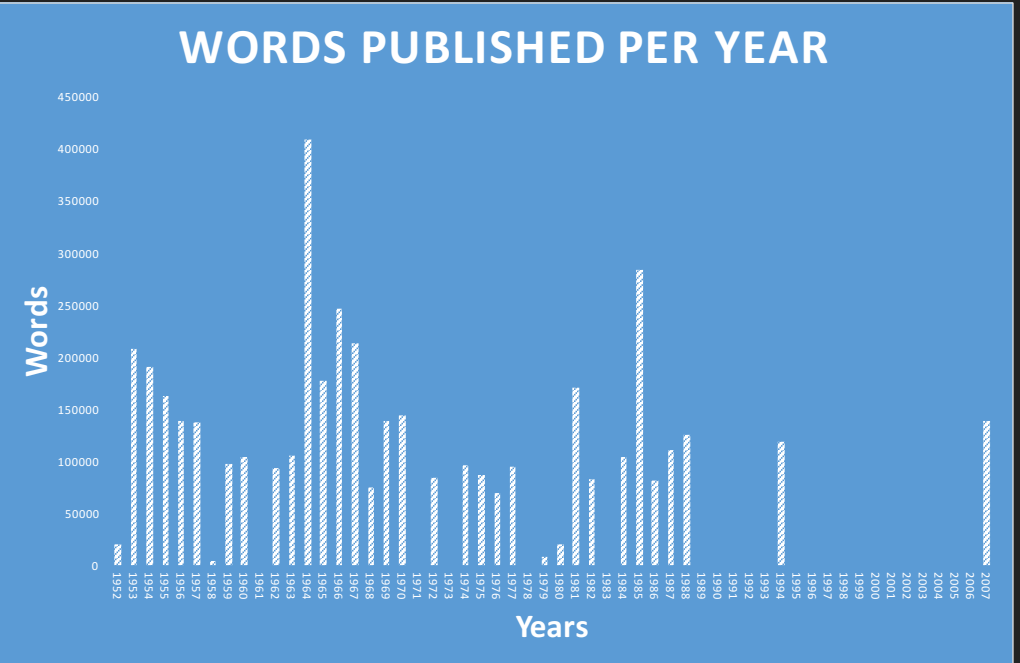
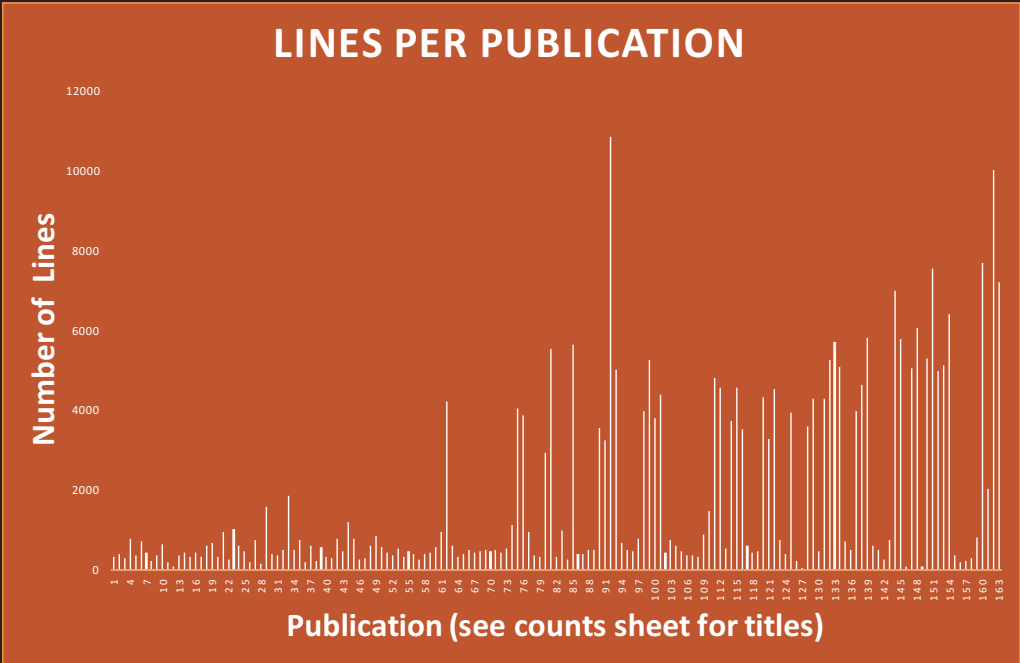
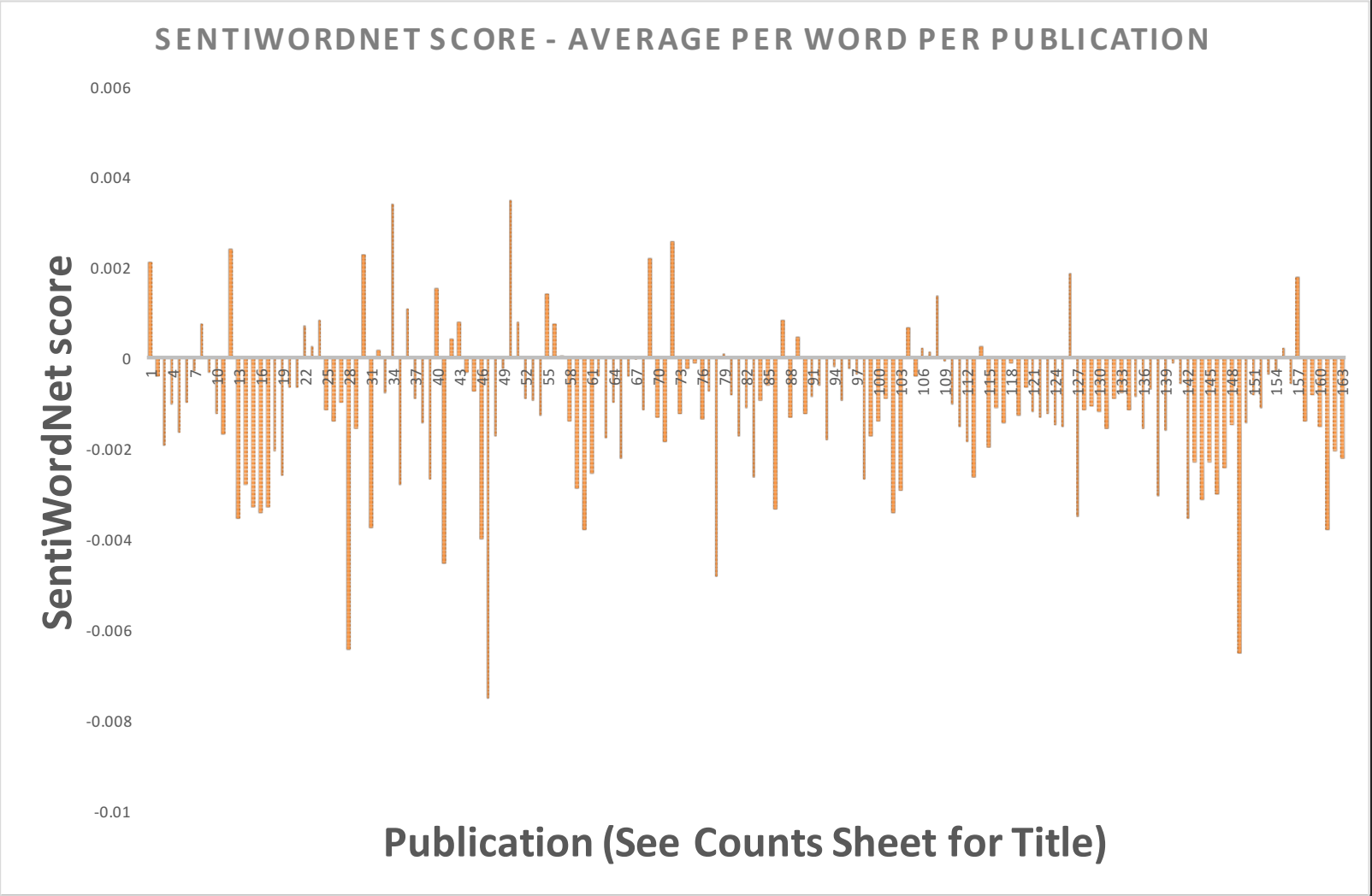
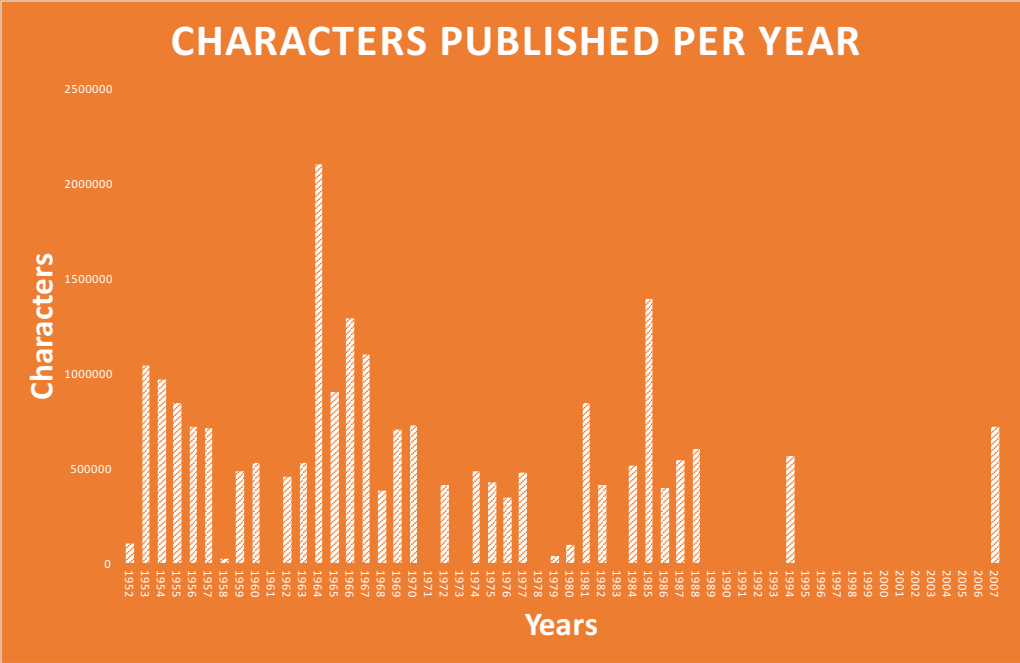
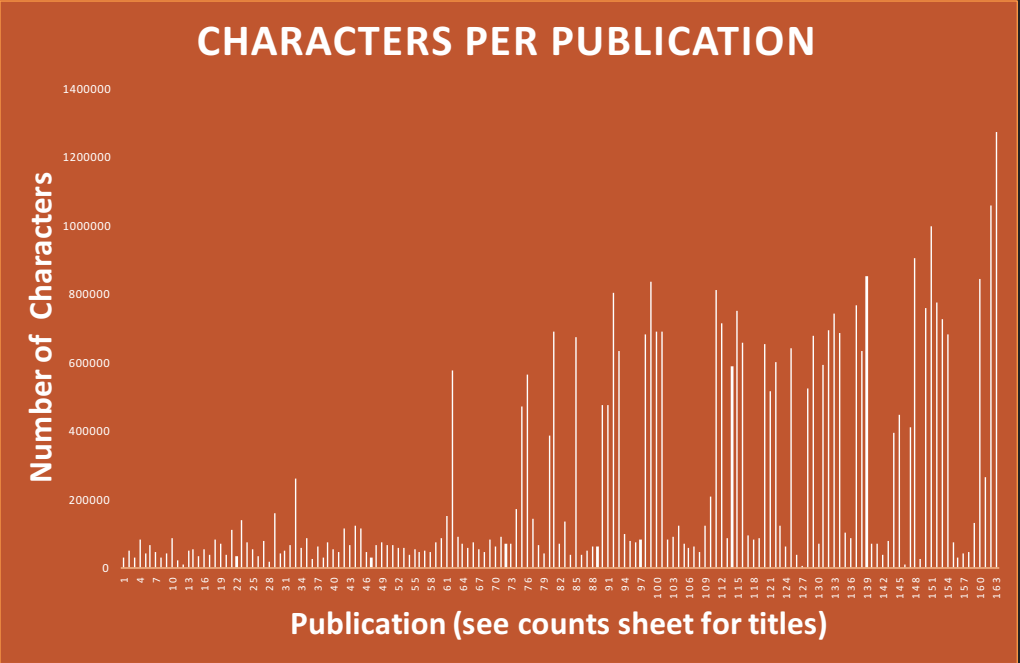
WORKSHOP 1: PROCESSING / CLEANING DATA

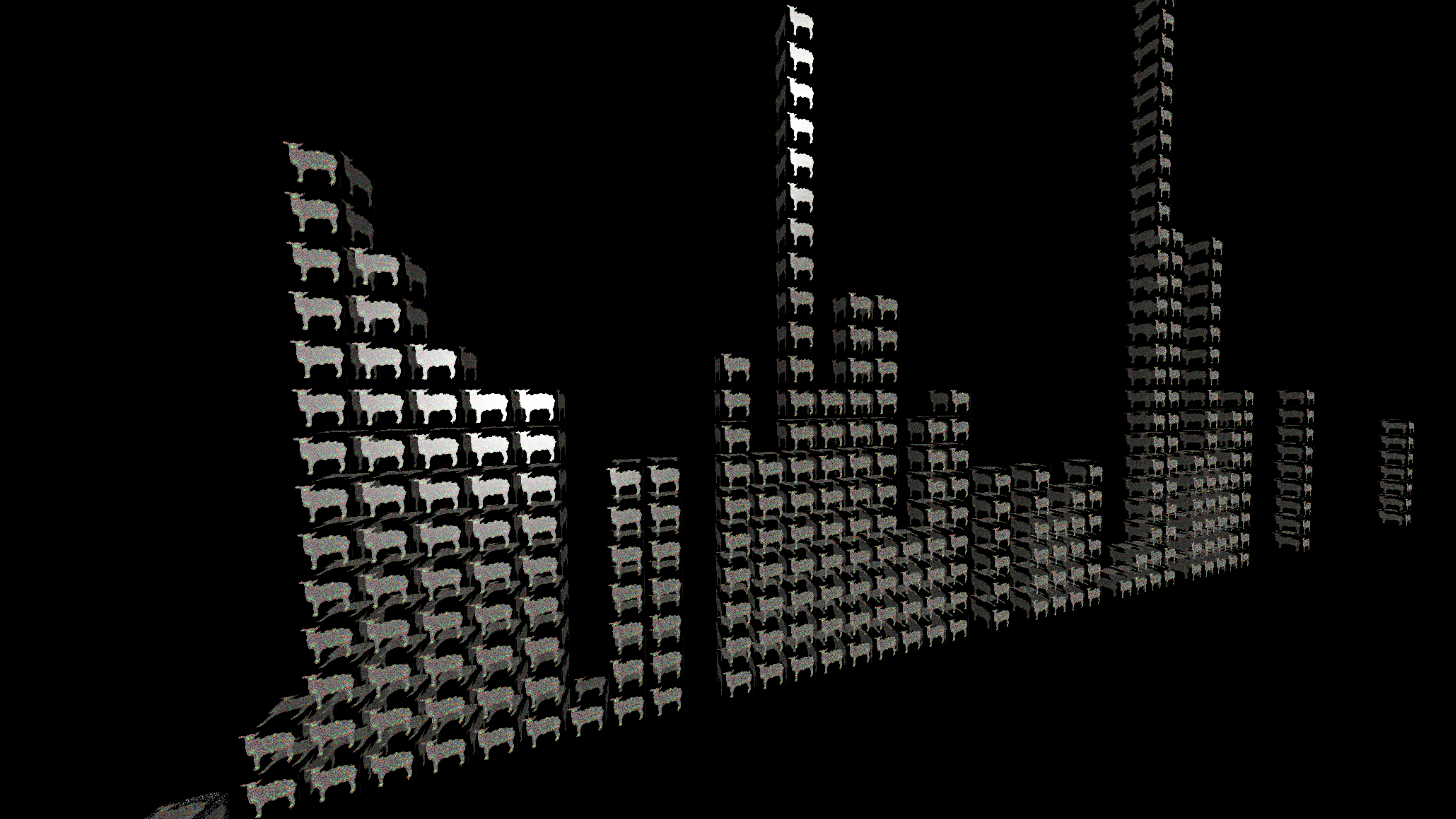
SHL OPEN WORKSHOP SERIES

PLAN FOR THE SESSION

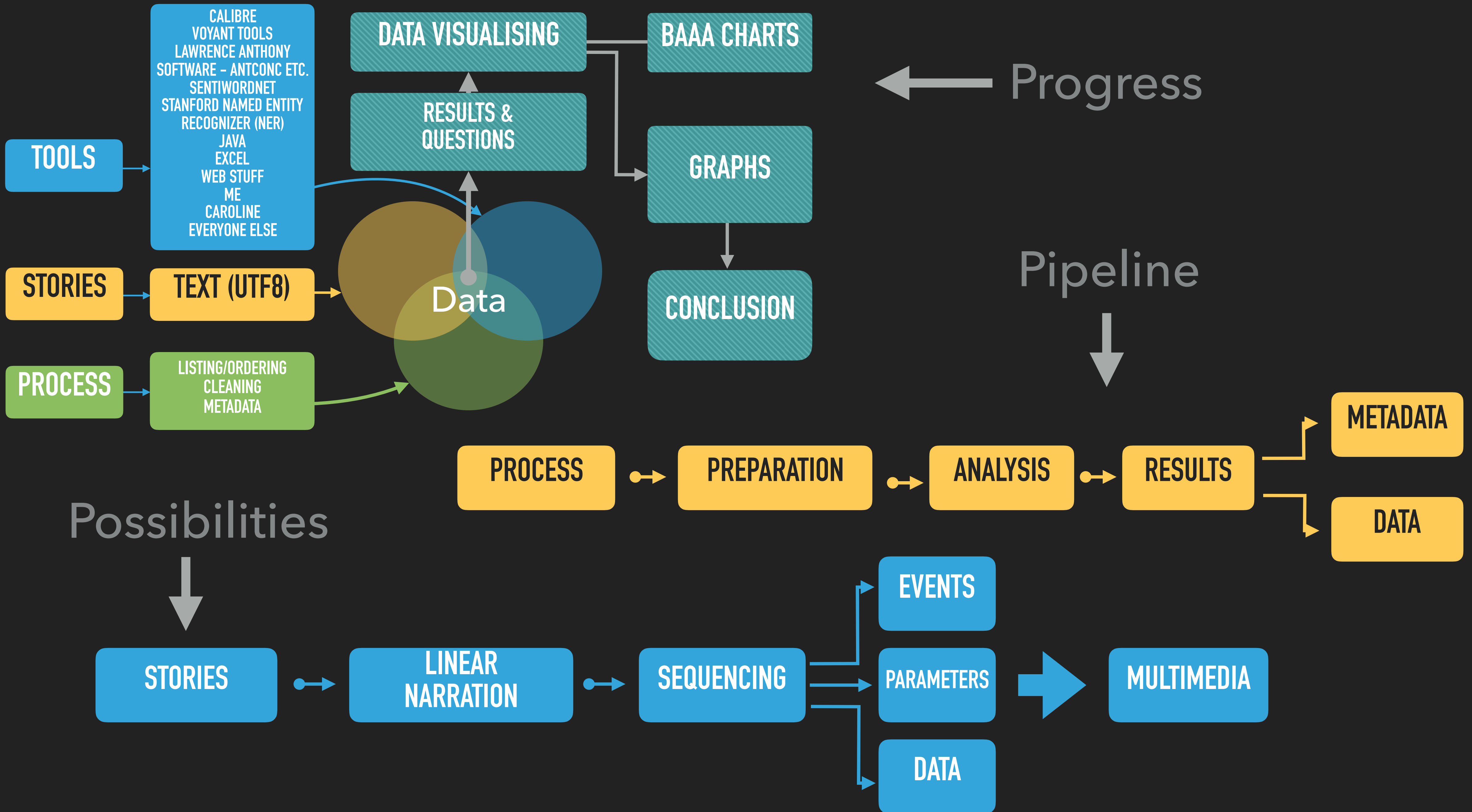
- ▶ Introduction - Short talk on what, how, why and some things to consider
- ▶ Practical 1. Calibre - eBook management tool
- ▶ Practical 2. Laurence Anthony's Tools
- ▶ Practical 3. Voyant Tools
- ▶ Practical 4. Fun with Text Editors
- ▶ Conclusion

SHOULD WE COUNT ELECTRIC SHEEP? : WHAT WE CAN SAY WITH NUMBERS : MAY 2017









MO 12TH OF MAY

phpMyAdmin

Recent

Favorites

New

androidhive

cotest

information_schema

mo12may

New

mo_scores

Columns

Indexes

mo_sentiment

Columns

New

mo_id

sentiment_count

sentiment_sum

sentiment_words

Indexes

mo_stanford_ner

Columns

New

mo_id

ner_tags

ner_words

Indexes

mo_words

Columns

New

char_count

line_count

mo_id

mo_text

word_count

Indexes

motest

mysql

obv

performance_schema

phpmyadmin

sierraleonedb

slapidb

sys

Browse

Structure

SQL

Search

Insert

Export

Import

Privileges

Operations

Tracking

Triggers

Showing rows 0 - 24 (2386 total, Query took 0.0006 seconds.)

SELECT * FROM `mo_scores`

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

1 > >> | Restore column order | Number of rows: 25 | Filter rows: Search this table

Sort by key: None

+ Options

		char_count	mo_id	word_count	line_count	sentiment_count	sentiment_sum	sentiment_av	sentiment_av2	sentiment_percent
<input type="checkbox"/>		64914	12th May 2015_twitter	12706	3720	1680	21.116688020231418	0.0016619461687573916	0.012569457154899653	13.2220997
<input type="checkbox"/>		6579	Mass Observation Diary May 12 2014	1380	95	322	10.871736312927476	0.007878069791976432	0.033763156251327565	23.3333333
<input type="checkbox"/>		3897	MT_2010_1	819	60	175	-3.668145009481978	-0.00447880953538703	-0.020960828625611303	21.3675213
<input type="checkbox"/>		4334	MT_2010_10	908	86	225	1.7957070358697405	0.001977650920561388	0.007980920159421069	24.7797356
<input type="checkbox"/>		1691	MT_2010_100	373	41	74	-0.5007525668923754	-0.0013425001793361268	-0.006766926579626694	19.839142
<input type="checkbox"/>		4396	MT_2010_101	831	63	227	-0.29686069263957016	-0.00035723308380213015	-0.0013077563552403972	27.3164861
<input type="checkbox"/>		4253	MT_2010_102	885	56	223	4.815581717805553	0.005441335274356557	0.021594536851145978	25.1977401
<input type="checkbox"/>		4297	MT_2010_103	879	70	227	2.5627518909212124	0.0029155311614575794	0.011289655907141904	25.8248009
<input type="checkbox"/>		4287	MT_2010_104	906	82	218	1.3710800635437341	0.0015133334034699052	0.006289358089650156	24.0618101
<input type="checkbox"/>		4569	MT_2010_105	937	65	228	4.41768636792945	0.004714713306221399	0.019375817403199342	24.3329775
<input type="checkbox"/>		4546	MT_2010_106	900	86	199	0.9261868754422593	0.001029096528269177	0.004654205404232459	22.1111111
<input type="checkbox"/>		4808	MT_2010_107	970	88	262	5.273208911155809	0.00543629784655238	0.02012675156929698	27.0103092
<input type="checkbox"/>		4376	MT_2010_108	911	73	208	-1.3691102203589751	-0.0015028652254214874	-0.006582260674802765	22.8320526
<input type="checkbox"/>		4563	MT_2010_109	833	55	201	-10.31003800136212	-0.012376996400194622	-0.05129372139981154	24.1296518
<input type="checkbox"/>		4330	MT_2010_11	891	69	240	0.14821275770342673	0.00016634428474009735	0.0006175531570976114	26.9360269
<input type="checkbox"/>		4239	MT_2010_110	889	66	222	-1.9846309879292663	-0.0022324308075694784	-0.008939779224906606	24.9718785
<input type="checkbox"/>		4182	MT_2010_111	852	63	179	0.4523356194974461	0.0005309103515228241	0.0025270146340639447	21.0093896
<input type="checkbox"/>		2591	MT_2010_112	603	110	131	-4.759144299107386	-0.007892444940476594	-0.03632934579471287	21.7247097
<input type="checkbox"/>		3027	MT_2010_113	624	53	155	2.780148192869423	0.004455365693700999	0.01793643995399628	24.8397435
<input type="checkbox"/>		4495	MT_2010_114	881	74	188	-1.5276647769094138	-0.0017340122325873028	-0.008125876472922414	21.339387
<input type="checkbox"/>		3233	MT_2010_115	700	61	162	2.889997186195793	0.004128567408851133	0.017839488803677736	23.1428571
<input type="checkbox"/>		4501	MT_2010_116	896	96	203	-2.7467521745875443	-0.0030655716234235983	-0.013530798889593814	22.65625
<input type="checkbox"/>		4897	MT_2010_117	907	73	221	-4.6985247169830675	-0.005180291859959281	-0.021260292837027454	24.3660418
<input type="checkbox"/>		7675	MT_2010_118	1356	137	399	-1.3971833233111848	-0.0010303711823828796	-0.0035017125897523426	29.4247787
<input type="checkbox"/>		5749	MT_2010_119	1146	147	246	-0.7940971547437667	-0.0006929294544011926	-0.003228037214405556	21.4659685

☐ Check all With selected:

1 > >> | Restore column order | Number of rows: 25 | Filter rows: Search this table

Query results operations

Database Explorer

- New
- androidhive
- cotest
- information_schema
- mo12may
 - New
 - mo_scores
 - Columns
 - Indexes
 - mo_sentiment
 - Columns
 - New
 - mo_id
 - sentiment_count
 - sentiment_sum
 - sentiment_words
 - Indexes
 - mo_stanford_ner
 - Columns
 - New
 - mo_id
 - ner_tags
 - ner_words
 - Indexes
 - mo_words
 - Columns

Showing rows 0 - 24 (2386 total, Query took 0.0006 seconds.)

```
SELECT * FROM `mo_scores`
```

1 > >> | Restore column order | Number of rows: 25 | Filter rows: Search this table

Sort by key: None

+ Options

		char_count	mo_id	word_count	line_count	sentiment_count	sentiment_sum
<input type="checkbox"/>	Edit Copy Delete	64914	12th May 2015_twitter	12706	3720	1680	21.1166880202314
<input type="checkbox"/>	Edit Copy Delete	6579	Mass Observation Diary May 12 2014	1380	95	322	10.8717363129274
<input type="checkbox"/>	Edit Copy Delete	3897	MT_2010_1	819	60	175	-3.6681450094819
<input type="checkbox"/>	Edit Copy Delete	4334	MT_2010_10	908	86	225	1.79570703586974
<input type="checkbox"/>	Edit Copy Delete	1691	MT_2010_100	373	41	74	-0.50075256689237
<input type="checkbox"/>	Edit Copy Delete	4396	MT_2010_101	831	63	227	-0.296860692639570
<input type="checkbox"/>	Edit Copy Delete	4253	MT_2010_102	885	56	223	4.8155817178055
<input type="checkbox"/>	Edit Copy Delete	4297	MT_2010_103	879	70	227	2.56275189092121
<input type="checkbox"/>	Edit Copy Delete	4287	MT_2010_104	906	82	218	1.37108006354373
<input type="checkbox"/>	Edit Copy Delete	4569	MT_2010_105	937	65	228	4.417686367929
<input type="checkbox"/>	Edit Copy Delete	4546	MT_2010_106	900	86	199	0.92618687544225
<input type="checkbox"/>	Edit Copy Delete	4808	MT_2010_107	970	88	262	5.2732089111558

Search this table

☐ Profiling [[Edit inline](#)] [[Edit](#)] [[Explain SQL](#)]

count	sentiment_count	sentiment_sum	sentiment_av	sentiment_av2	sentiment_percent
3720	1680	21.116688020231418	0.0016619461687573916	0.012569457154899653	13.2220997
95	322	10.871736312927476	0.007878069791976432	0.033763156251327565	23.3333333
60	175	-3.668145009481978	-0.00447880953538703	-0.020960828625611303	21.3675213
86	225	1.7957070358697405	0.001977650920561388	0.007980920159421069	24.7797356
41	74	-0.5007525668923754	-0.0013425001793361268	-0.006766926579626694	19.839142
63	227	-0.29686069263957016	-0.00035723308380213015	-0.0013077563552403972	27.3164861
56	223	4.815581717805553	0.005441335274356557	0.021594536851145978	25.1977401
70	227	2.5627518909212124	0.0029155311614575794	0.011289655907141904	25.8248009
82	218	1.3710800635437341	0.0015133334034699052	0.006289358089650156	24.0618101
65	228	4.41768636792945	0.004714713306221399	0.019375817403199342	24.3329775
86	199	0.9261868754422593	0.001029096528269177	0.004654205404232459	22.1111111
88	262	5.273208911155809	0.00543629784655238	0.02012675156929698	27.0103092

mo_sentiment

- Columns
 - New
 - mo_id
 - sentiment_count
 - sentiment_sum
 - sentiment_words
- Indexes

mo_stanford_ner

- Columns
 - New
 - mo_id
 - ner_tags
 - ner_words
- Indexes

mo_words

- Columns
 - New
 - char_count
 - line_count
 - mo_id
 - mo_title
 - word_count
- Indexes

motest

mysql

obv

performance_schema

phpmyadmin

sierraleonedb

slapidb

sys

<input type="checkbox"/>				MT_2010_110	{"LOCATION":["West End of Glasgow","Australia"],"PERSON":	My name is xxx, I'm 29 and I live with my boyfrien...
<input type="checkbox"/>				MT_2010_111	{ "LOCATION": ["Brighton", "Cameron", "Grey Gardens"], "ORGANIZATION":	
<input type="checkbox"/>				MT_2010_112	["Asda", "Nestle", "Renault", "Labour"], "PERSON":	
<input type="checkbox"/>				MT_2010_113	["Laurie", "Jessie", "Vincent", "Laurie", "Laurie", "Vic", "Bob", "Vincent", "Jessie", "Vincent	
<input type="checkbox"/>				MT_2010_114	, "Jessie", "Vincent", "Emma", "Arthur", "Cameron", "Vincent", "Emma", "Arthur", "Vince", "Bob	
<input type="checkbox"/>				MT_2010_115	the", "Vincent", "Caroline", "Clegg", "Jessie", "Vincent", "Jessie", "Finn", "Vincent", "Joseph	
<input type="checkbox"/>				MT_2010_116	, "Jessie", "Joseph", "Pickle", "Jessie", "Finn", "Vincent", "Caroline", "Jessie", "Joseph", "L	
<input type="checkbox"/>				MT_2010_117	aurie", "Vince", "Jessie", "Laurie", "Helen", "Laurie", "Wolf Hall"] }	
<input type="checkbox"/>				MT_2010_118	{ "LOCATION": ["Lewes", "London", "Dunkarry", "Lewes", "Brighton", "Brighton", "Britain", "Gardner	Gender: Female
<input type="checkbox"/>				MT_2010_119	Street", "China", "Madagascar", "Madagascar"], "ORGANIZATION": ["Infinity Cafe", "Infinity	City: <...
<input type="checkbox"/>				MT_2010_120	Health", "Tories"], "PERSON": ["Cameron", "Clegg", "Brown", "Obama", "Tory", "Lapsang", "Sara	May 12th 2010
<input type="checkbox"/>				MT_2010_121	Maitland", "Gormley", "Dad", "Carluccio", "Maggie", "Gordon Brown"] }	ian
<input type="checkbox"/>				MT_2010_122		45
<input type="checkbox"/>				MT_2010_123		g in <LOCATIO...
<input type="checkbox"/>				MT_2010_124		first day of the new Conservative/Lib-Dem gov...
<input type="checkbox"/>				MT_2010_125		BIOGRAPHY

SELECT * FROM articles
WHERE MATCH (title,body)
AGAINST ('database' IN
NATURAL LANGUAGE MODE);

1 > >> | Number of rows: 25 | Filter rows: Search this table

Query results operations

MO 12th May Diaries API

Examples

[get_diary_list/.....](#)

Call returns a list of mo diary ids - first test example for mo API.

[get_diary_id_list/.....](#)

Call returns a list of diary ids (content of the mo_id field from the mo_scores table).

[get_diary_entry](#)

Call returns a diary score entry. Diary specified by the "diary" segment of the url, which is the mo_id value.

[get_diary_entry/diary/...](#)

Call returns a diary entry. Diary specified by the "diary" segment of the url, which is the mo_id value.

[get_diary_sentiwords/diary/...](#)

Call returns a diary entry's sentiment bearing word list - a csv description of the sentiment bearing words and their sentiwordnet values and POS assignments. Diary specified by the "diary" segment of the url, which is the mo_id value.

Mass Observation 12th of May Diaries Macroscope

An SHL development project

This tool has been created to explore approaches to the presentation of information derived from the Mass Observation 12th of May Diaries. This demonstration site is designed to illustrate how a range of tools for textual and data analysis, might be combined to create a working 'macroscope' – an environment where 'big data' can be explored both at scale, and at the level of the single datum. Its purpose is to allow a new 'open eyed' way of working with data of all sorts – to allow macro-patterns and clusters to be identified; while single words and phrases can be fully contextualised. This experiment uses tools created for a previous investigation, this implementation is part of an assessment of the reuse opportunities for the technology.

[Graphs](#)[Diary](#)[Extracted Data](#)

MO id: MT_2014_220



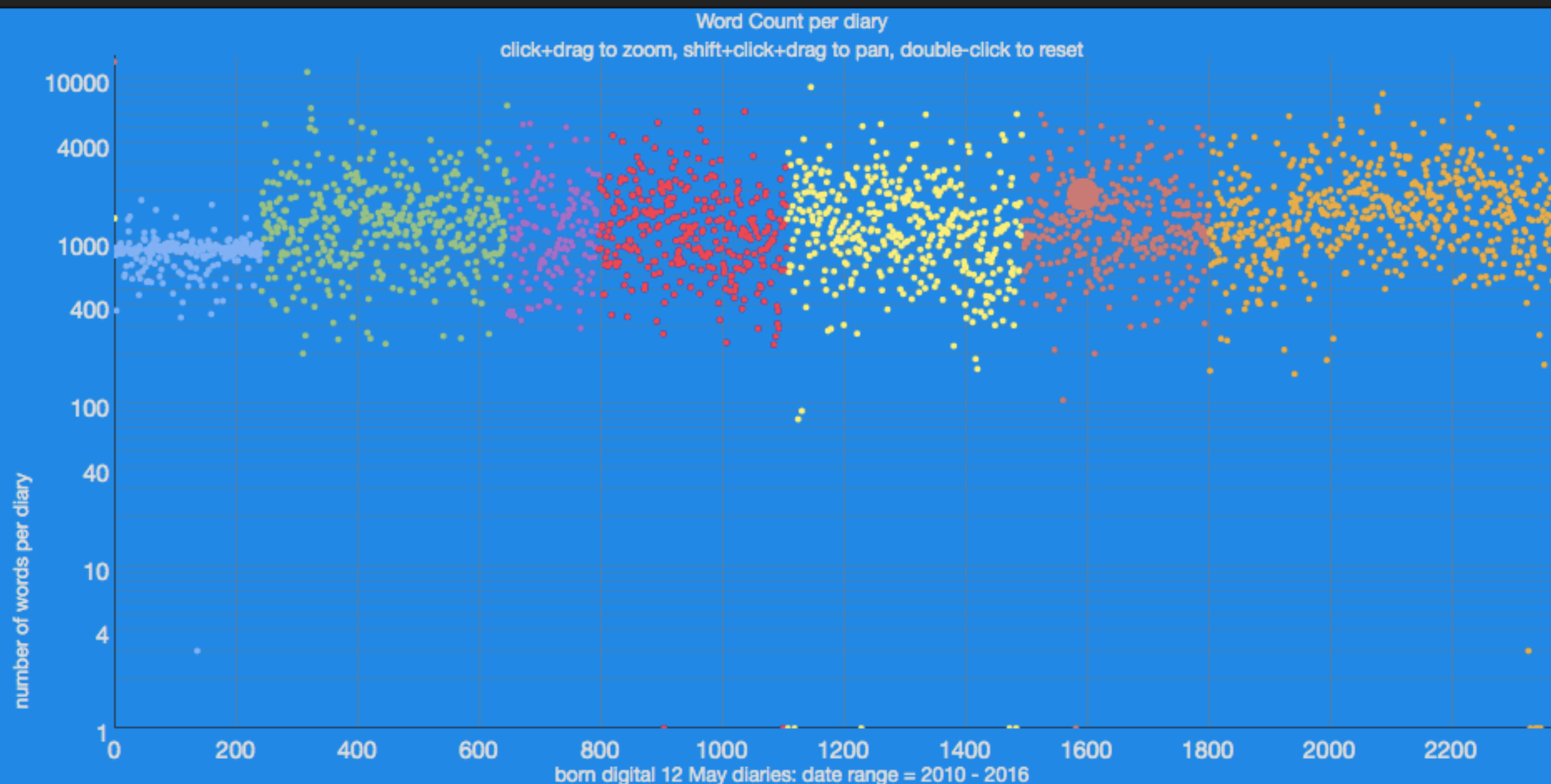
Graphs

filters ▾

data: 2386 records

Filter 1

- ☒ 2010
- ☒ 2011
- ☒ 2012
- ☒ 2013
- ☒ 2014
- ☒ 2015
- ☒ 2016
- ☒ word count
- ☐ sentiment count
- ☐ sentiment %



mo_id: MT_2015_193

word count: 1942

char #: 9360

line #: 156

senti #: 493

senti sum:

4.909952398704223

senti av:

0.0025282968067477977

senti av2:

0.009959335494329053

senti %: 25.3861997

Mass Observation 12th of May Diaries Macroscope

An SHL development project

This tool has been created to explore approaches to the presentation of information derived from the Mass Observation 12th of May Diaries. This demonstration site is designed to illustrate how a range of tools for textual and data analysis, might be combined to create a working 'macroscope' – an environment where 'big data' can be explored both at scale, and at the level of the single datum. Its purpose is to allow a new 'open eyed' way of working with data of all sorts – to allow macro-patterns and clusters to be identified; while single words and phrases can be fully contextualised. This experiment uses tools created for a previous investigation, this implementation is part of an assessment of the reuse opportunities for the technology.

[Graphs](#)[Diary](#)[Extracted Data](#)

MO Id: MT_2014_220



Diary: MT_2014_220

Mass Observation Diary – 12th May 2014

Richard Lovett

Age 68 yrs

Living in Brighton

Retired Police Officer

Married for 46 years with two Children and three Grand Children .

Christian

Brighton and Hove Albion Fan and Sussex Cricket supporter

Volunteer HealthWalks leader

Keen amateur photographer

Cyclist

Traveller

Avid reader.

0700 Up as usual to feed cat, empty dishwasher and make cup of tea. Back to sit in bed with my wife, watched BBC News on TV. Read today's reading from scripture together using the Anglican lectionary app on our iPads.

0800 sunny start to day, so my wife is hanging out washing, while I make my breakfast porridge.

Breakfast with my wife, then caught up with "Words with friends" games on iPad before shower

Word count: 885

Char count: 4397

Line count: 82

Sentiment count: 213

Sentiment sum: 7.309430885602352

Sentiment av: 0.008259243938533731

Sentiment av2: 0.034316576927710576

Sentiment %: 24.0677966

Mass Observation 12th of May Diaries Macroscope

An SHL development project

This tool has been created to explore approaches to the presentation of information derived from the Mass Observation 12th of May Diaries. This demonstration site is designed to illustrate how a range of tools for textual and data analysis, might be combined to create a working 'macroscope' – an environment where 'big data' can be explored both at scale, and at the level of the single datum. Its purpose is to allow a new 'open eyed' way of working with data of all sorts – to allow macro-patterns and clusters to be identified; while single words and phrases can be fully contextualised. This experiment uses tools created for a previous investigation, this implementation is part of an assessment of the reuse opportunities for the technology.

Graphs

Diary

Extracted Data

MO id: MT_2014_220

Named Entity Recognition Tags

Person:

Richard Lovett

Hove

Hove

Olly

Olly

Minecraft

I. Mushroom

Oscar Garcia

Victoria Coren-Mitchell

John SIMM

Organisation:

Hove Albion

Sussex Cricket

BBC News on TV

Church House

St Patricks

Albion

BB

Named Entity Recognition Words

Mass Observation Diary – 12th May 2014

Richard Lovett

Age 68 yrs

Living in Brighton

Retired Police Officer

Married for 46 years with two Children and three Grandchildren

Brighton and Hove Albion Fan and Sussex Cricket supporter

Volunteer HealthWalks leader

Keen amateur photographer

Cyclist

Traveller

Avid reader.

0700 Up as usual to feed cat, empty dishwasher and read newspaper with wife, watched BBC News on TV. Read today's reading 'The Word' and lectionary app on our iPads.

0800 sunny start to day, so my wife is hanging out with friends

Breakfast with my wife, then caught up with "Words with Friends" and getting dressed

Sentiment bearing words and results

12th, Adjective, 0.0

years, Noun, 0.0

supporter, Noun, 0.0

leader, Noun, 0.0

amateur, Adjective, 0.08333333333333333

photographer, Noun, 0.0

reader, Noun, 0.0

as, Adverb, -0.125

usual, Adjective, 0.125

cat, Noun, -0.015330705212439772

empty, Adjective, 0.23

dishwasher, Noun, 0.0

cup, Noun, 0.02299605781865966

tea, Noun, 0.0

bed, Noun, 0.009198423127463865

wife, Noun, 0.0

today, Noun, 0.08333333333333333

reading, Noun, 0.02299605781865966

scripture, Noun, 0.0

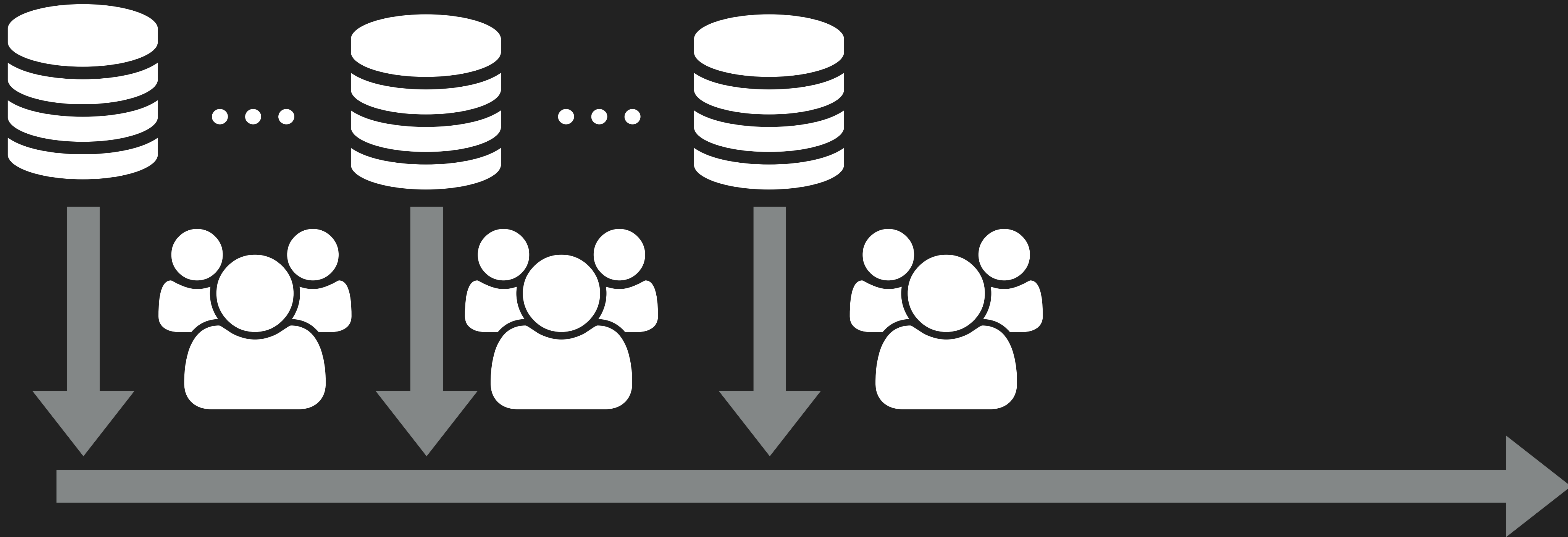
together, Adverb, 0.0

sunny, Adjective, 0.5

start, Noun, 0.0

OLD BAILEY





(SHL) Old Bailey Voices API

OBV API

This API, and the database it connects to are works in progress. The API began as a way to provide web access to data collected for the Old Bailey Voices of Authority project for a VR puppet show, this is still the primary purpose. The desire for more data to increase the story telling possibilities means the database and API continue to grow. The database now includes sources from the Old Bailey Online and the Old Bailey Corpus (2). The work that created these data sets represents considerable effort that has taken place over more than two decades. This project is entirely dependent on the source data and the work that created it.

The work is an output the Sussex Humanities Lab at the University of Sussex. If you find the API useful it was written by Ben Jackson with assistance from Tim Hitchcock. If you wish to create applications based on it please contact me (find me on the Sussex Uni web pages) - this access point is my development site and is likely to change without notice, this site is where I try out ideas so changes may occur while you're reading. Please contact me if you want more reliable use of the work.

Catagories

- obo = information taken from Old Bailey Online ~ (1674 - 1913)
- obc = information taken from Old Bailey Corpus (2) ~ (1720 - 1913)
- obv = information taken from Old Bailey Voices of Authority ~ (1720 - 1880)

[get_obc_producer/.....](#)

Returns an obc producer entry

[get_obv_trialid/defendant/t17800112-1-defend47/format/json](#)

Returns a trial from a defendantid -

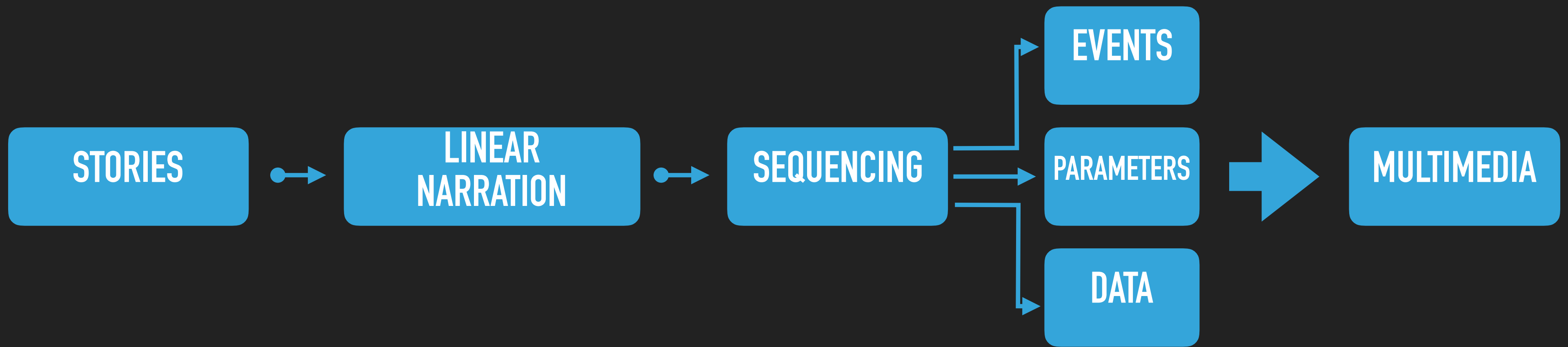
[get_obv_session_list/.....](#)

Returns a list (227) of obv session dates -

[get_obv_session_trial_list/.....](#)

Returns a list of trial IDs, uses an obv session date to specify the session paper

[get_obv_trials_wordcounts/](#)





Defendant: ANN THORNTON

Indictment: stealing a pair of leather pumps value 3 s two pair of men s leather shoes value 12 s two leather shoes value 6 s and one stuff shoe value 2 s the property of James Belcher Ball privately and secretly in the shop of the said James

Offence: theft - shoplifting

Outcome: guilty imprison whipping

Age: unknown

Occupation: spinster

Gender: female

Session date: 23rd February 1780

Trial ID: t17800223-1

Crime Location: Kensington

⚙️

▶️

🔊

🖼️

🏰

👤

📄

🔄

🚩

🗨️

❓

✕

THE WHOLE PROCEEDINGS ON THE KING's Commission of the Peace, Oyer and Terminer, and Gaol Delivery for the CITY of LONDON; AND ALSO, The Gaol Delivery for the County of MIDDLESEX; HELD AT JUSTICE HALL in the OLD BAILEY, On Wednesday the 23d of February, 1780, and the following Days; Being the THIRD SESSION in the Mayoralty of The Right Honble BRACKLEY KENNET, LORD MAYOR OF THE CITY OF LONDON.

TAKEN IN SHORT-HAND BY JOSEPH GURNEY , And Published by Authority.

NUMBER III. PART I.

LONDON:

Printed for JOSEPH GURNEY (the PROPRIETOR) And Sold by M. GURNEY, No. 34, Bell-yard, near Temple-Bar, MDCCLXXX.

THE WHOLE PROCEEDINGS UPON THE

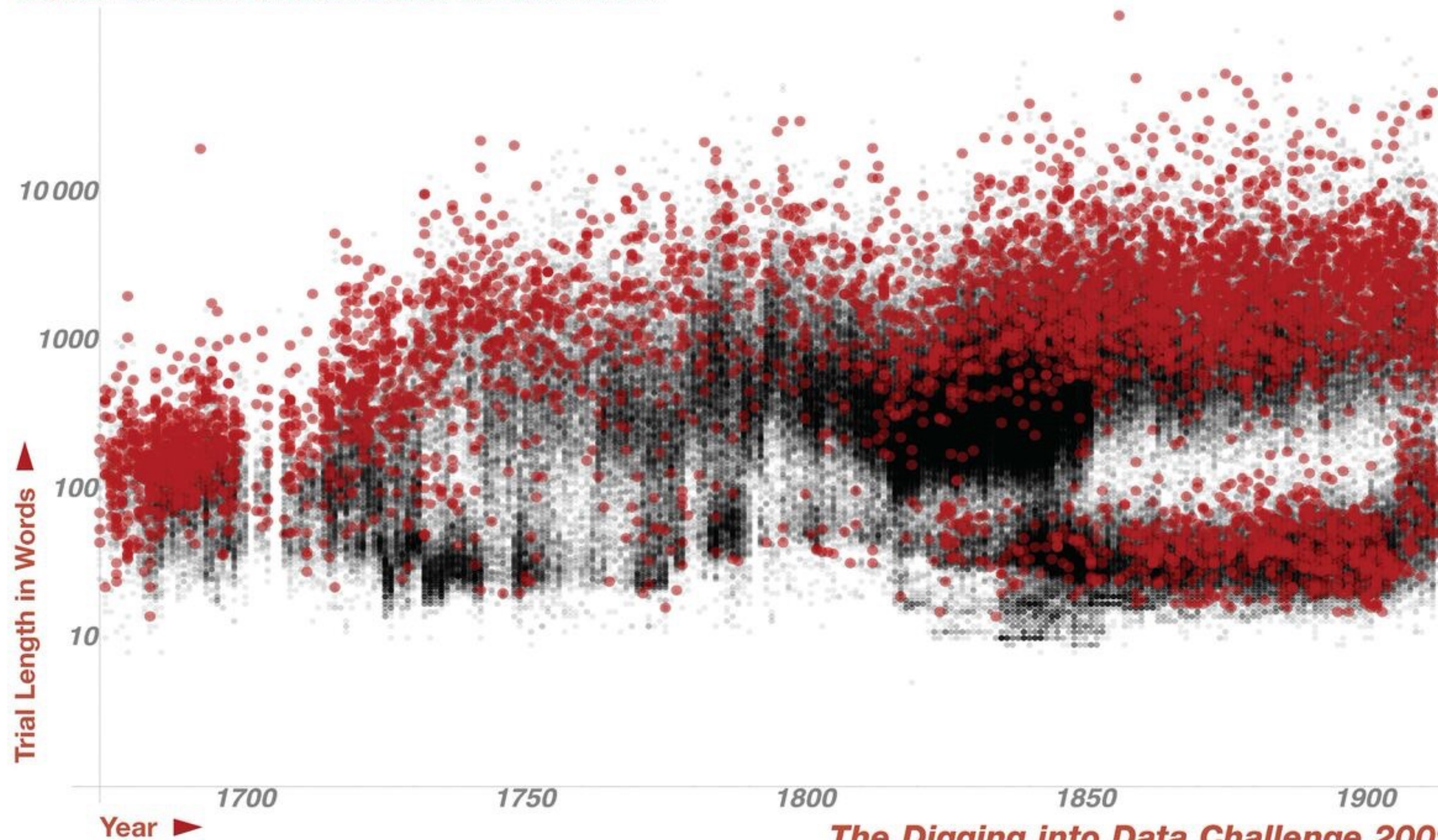
KING's Commission of the Peace, Oyer and Terminer, and Gaol Delivery for the CITY of LONDON, &c.

BEFORE the Right Honourable BRACKLEY KENNET, LORD MAYOR of the City of London; The Honourable Sir RICHARD PERRYIN , Knt. one of the Barons of his Majesty's Court of Exchequer; The Honourable FRANCIS BULLER , one of the Justices of his Majesty's Court of King's Bench; Mr. Serjeant ADAIR, Recorder; and others his Majesty's Justices of Oyer and Terminer of the City of London and Justices of the Gaol Delivery of Newgate, holden for the said City and County of Middlesex.

London. Jury.

Data Mining With Criminal Intent

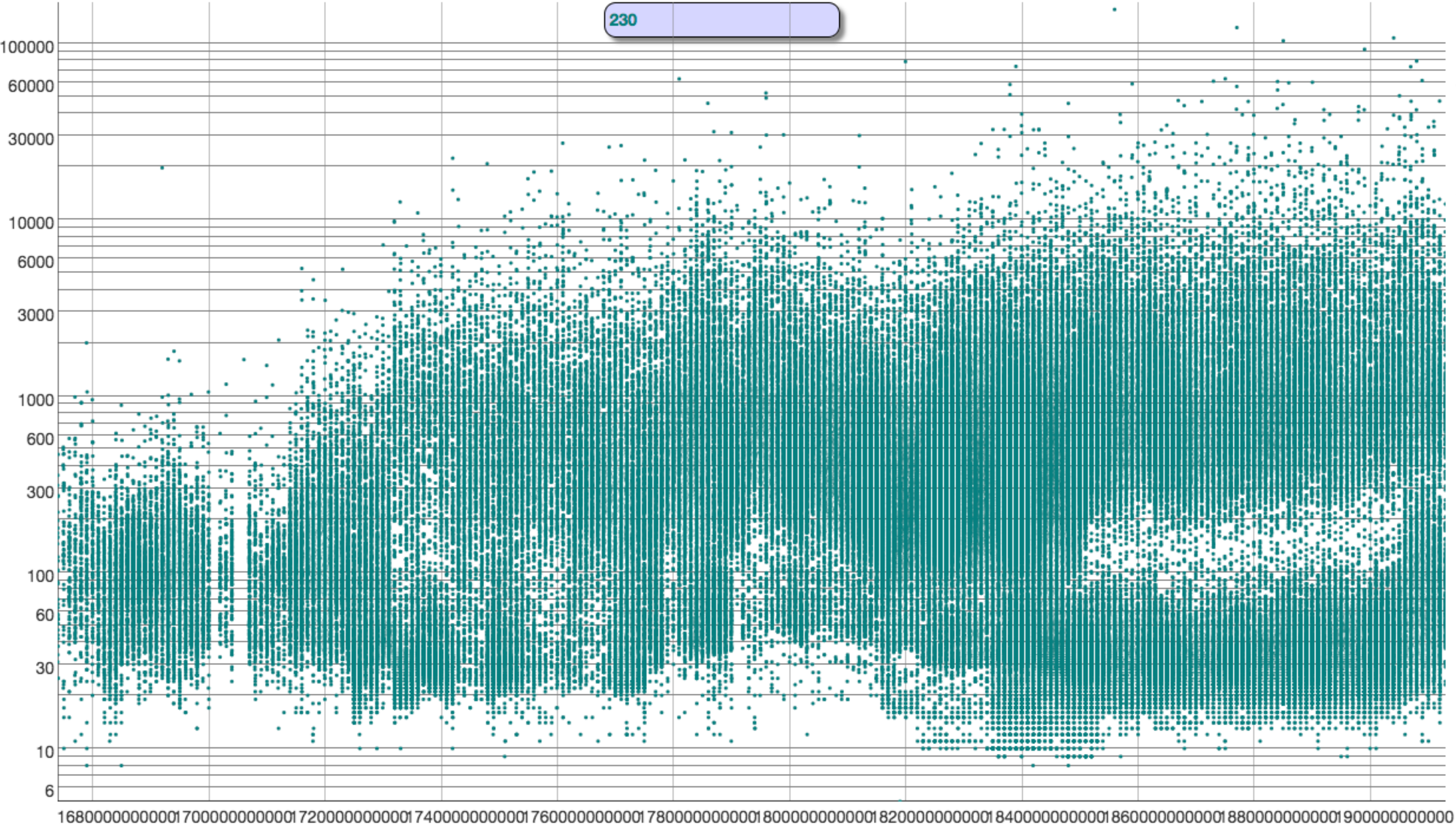
Cyril Briquet • Dan Cohen • Frederick Gibbs • Tim Hitchcock • Jamie McLaughlin • Geoffrey Rockwell
Joerg Sander • Robert Shoemaker • John Simpson • Stéfan Sinclair • Sean Takats • William J. Turkel



Length in words of trials involving 'killing' (red) versus all other trials (black) from the Old Bailey Proceedings, plotted with a logarithmic scale for the x axis. 197,745 Trials.

The Digging into Data Challenge 2009

NEH • JISC • SSHRC



WILLIAM AXFORD
Offence: theft - grandLarceny
Outcome: guilty imprison houseOfCorrection
Age: 17
Occupation: apprentice to a waterman
Gender: male
Session date: 2nd June 1802
Trial ID: t18020602-1
<http://tim.benskitchen.com/ob1.php trial Id = t18020602-1>
Crime Location: Scotland-yard
Indictment: feloniously stealing on the 28th of April a jacket value 30s a pair of breeches value 18s a waistcoat value 5s a hat value 10s and two handkerchiefs value 1s

THE WHOLE PROCEEDINGS ON THE KING'S Commission of the Peace, Oyer and Terminer, and Goal Delivery FOR THE CITY OF LONDON; AND ALSO, The Goal Delivery FOR THE COUNTY OF MIDDLESEX, HELD AT JUSTICE-HALL, IN THE OLD-BAILEY, ON WEDNESDAY, the 2d of JUNE, 1802, and following Days, BEING THE FIFTH SESSION IN THE MAYORALTY OF The Right Honourable SIR JOHN EAMER , KNIGHT, LORD MAYOR OF THE CITY OF LONDON.
TAKEN IN SHORT-HAND BY RAMSEY & BLANCHARD,

LONDON: PRINTED AND PUBLISHED, By Authority of the CORPORATION of the CITY of LONDON, By W. WILSON, St. Peter's-Hill, Little Knight-Rider-Street, Doctors' Commons.

1802.

THE WHOLE PROCEEDINGS ON THE KING'S Commission of the Peace, Oyer and Terminer, and Goal Delivery FOR THE CITY OF LONDON, &c.

BEFORE the Right Honourable Sir JOHN EAMER , KNIGHT, LORD-MAYOR of the City of LONDON; Sir NASH GROSE , Knight, one of the Justices of his Majesty's Court of King's Bench; Sir ROBERT GRAHAM , Knight, one of the Barons of his Majesty's Court of Exchequer; Sir WILLIAM STAINES , Knight, JOHN BOYDELL , Esq. BROOK WATSON , Esq. and WILLIAM NEWMAN , Esq. Aldermen of the said City; Sir JOHN WILLIAM ROSE, Knight, Serjeant at Law, Recorder of said City; JOHN PERRING , Esq. JAMES SHAW , Esq. and GEORGE CLARKE , Esq. Aldermen of the said City; and JOHN SILVESTER , Esq. Common-Serjeant of the said City: His Majesty's Justices of

OldBaileyVoices.org

Telling stories with Old Bailey Data: An SHL development

This content has been created to explore approaches to the presentation of information derived from the proceedings of Old Bailey Trials (Sessions Papers). The datasets that this work relies on are output from 'The Old Bailey Online' and related projects.

This demonstration site is designed to illustrate how a range of tools for textual and data analysis, might be combined to create a working ‘macroscopic’ – an environment where ‘big data’ can be explored both at scale, and at the level of the single datum. Its purpose is to allow a new ‘open eyed’ way of working with data of all sorts – to allow macro-patterns and clusters to be identified; while single words and phrases can be fully contextualised.

The site builds on the Old Bailey Online dataset, which encompasses accounts of some 197,745 trials held at the Old Bailey in London, between 1674 and 1913. The Proceedings contain some 127,000,000 words of accurately transcribed texts, which has in turn, been substantially marked-up in XML to encode the administrative process (crime, verdict, punishment, etc) reflected in the accounts. Additionally, text that purports to reflect direct speech has also been marked up to allow for analysis of verbal-linguistic change.

The Old Bailey dataset has been chosen because it combines a consistent textual record at scale; because it incorporates structured data; because it includes varieties of text types (speech vs administration); and because it represents a uniquely consistent dataset reflecting historical change over 240 years. But the purpose of this demonstrator is suggest a more generic approach to ‘big data’ as a whole – an approach in which large scale patterns can be rapidly explored, and their constituent elements identified; or where single words or phrases (or just datum) can be radically contextualised within the full body of

USING MULTIMEDIA TO EXPLORE OLD BAILEY DATA

SHL MACROSCOPE ALPHA 1

Presented by Tim Hitchcock

▶

0:00 / 1:38

◀

⏮

⏭

⏸

🔊

⛶

⋮



filters

records: 21023

Series

offence

gender

verdict

breaking peace

damage

deception

kill

miscellaneous

royal offences

sexual

theft

violent theft

male

female

gender other

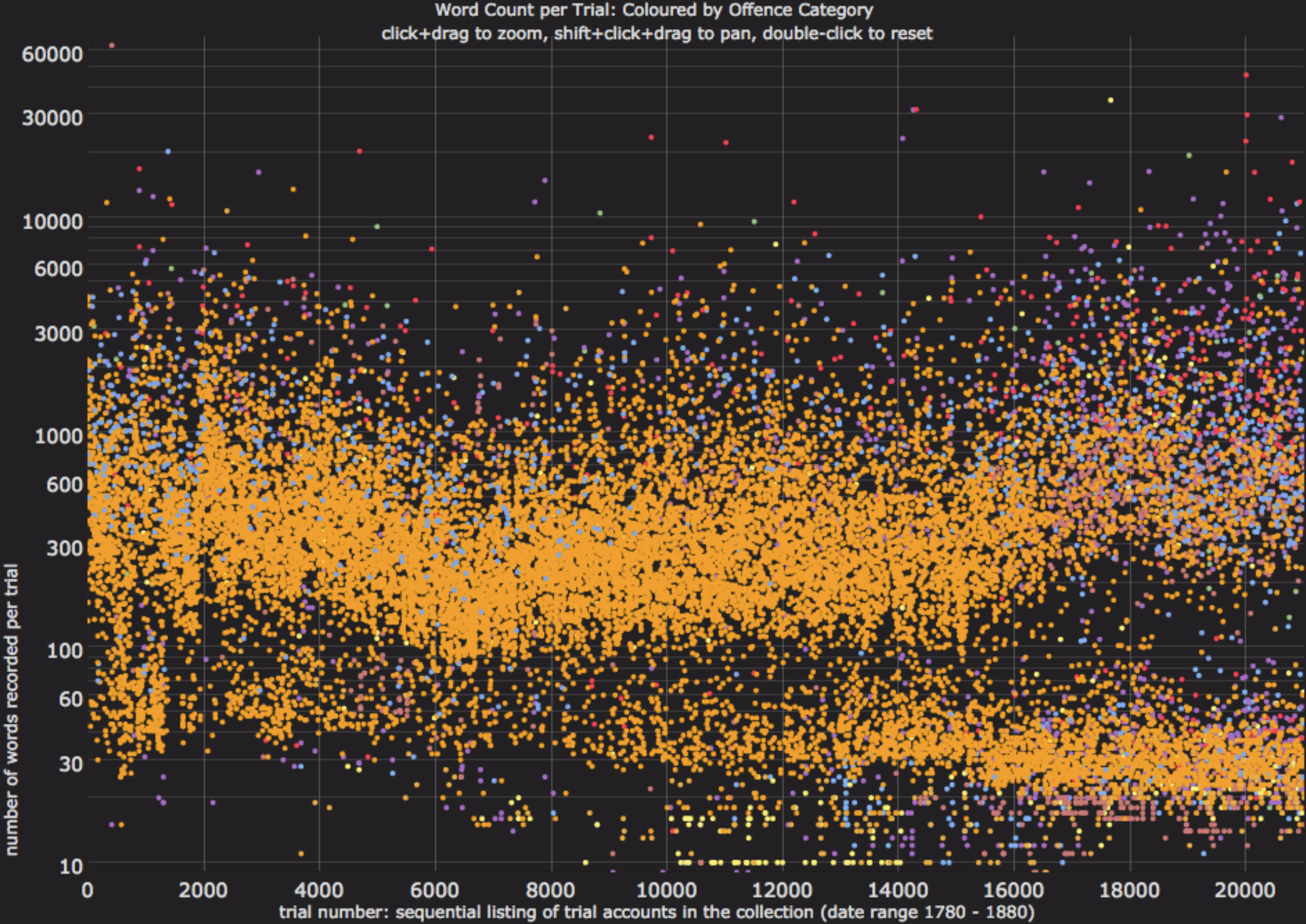
guilty

not guilty

verdict other

all

none



TrialId: t17800112-1

Wordcount: 809

Defendant: MARY DYER

Gender: female

Offence: theft

subcategory: grandLarceny

Verdict: guilty

subcategory:

Punishment: imprison

subcategory: newgate

Crime Location: No. 7, in Great Suffolk-street, Charing-cross

Indictment: stealing five linen sheets value 15 s and three tablecloths value 21 s

filters

records: 21023

Series

offence

gender

verdict

☒breaking peace

☒damage

☒deception

☒kill

☒miscellaneous

☒royal offences

☒sexual

☒theft

☒violent theft

☒male

☒female

☒gender other

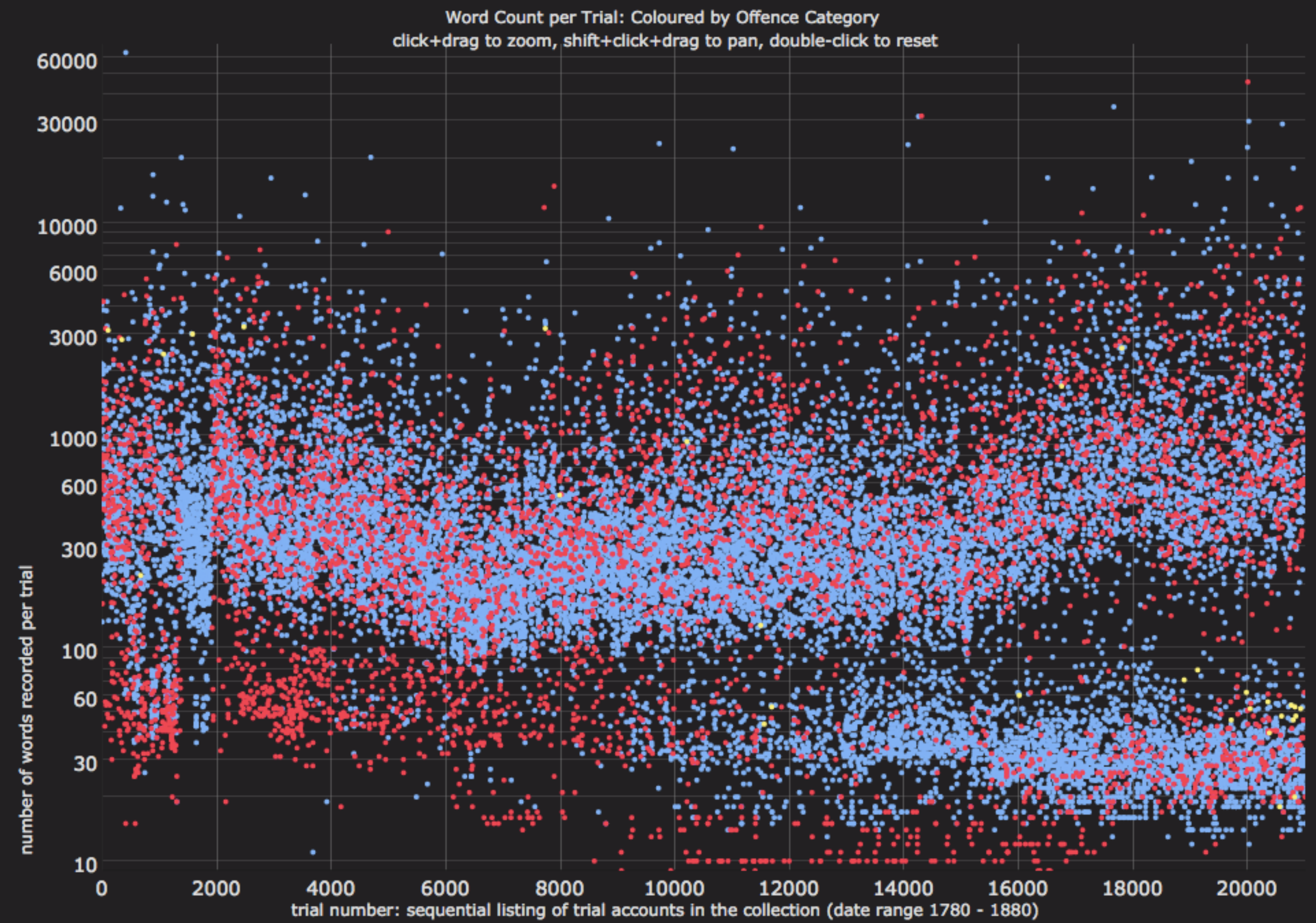
☒guilty

☒not guilty

☒verdict other

all

none





filters

records: 4272

Series

offence

gender

verdict

☒breaking peace

☒damage

☒deception

☒kill

☒miscellaneous

☒royal offences

☒sexual

theft

violent theft

☒male

☒female

☒gender other

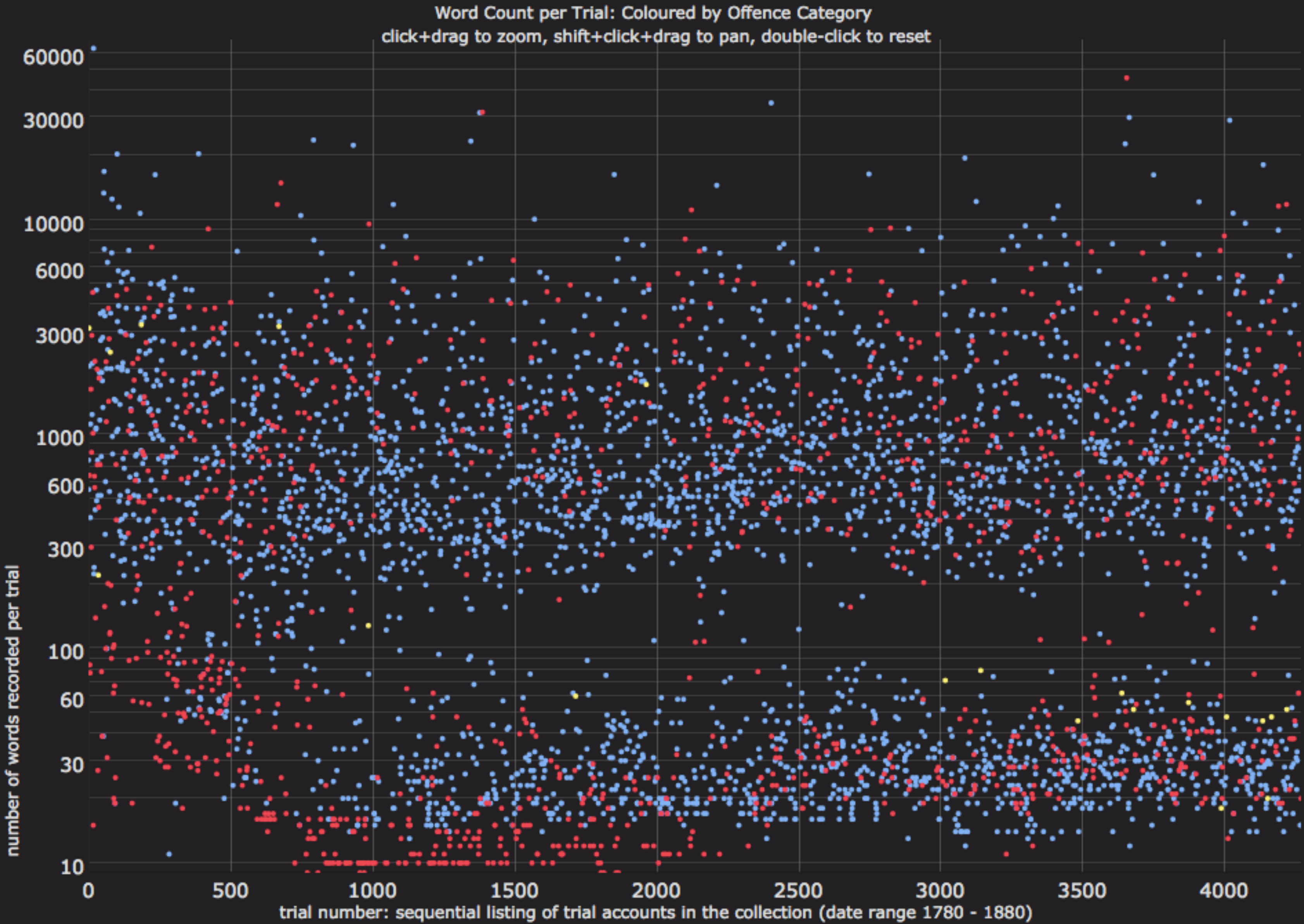
☒guilty

☒not guilty

☒verdict other

all

none





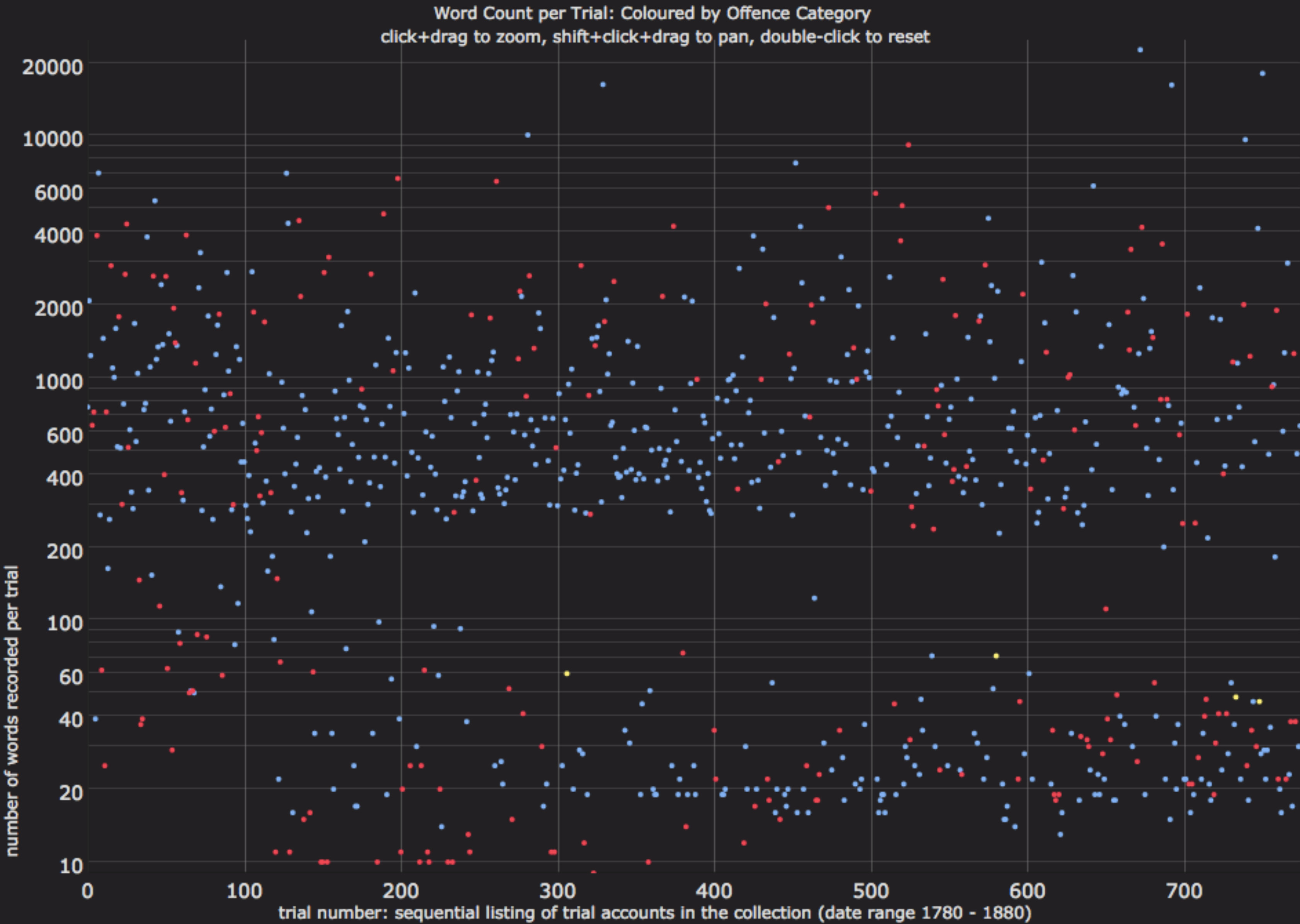
filters

records: 775

Series

- ☐ offence
 - ☐ gender
 - ☒ verdict
-
- ☒ breaking peace
 - ☒ damage
 - ☒ deception
 - ☒ kill
 - ☒ miscellaneous
 - ☒ royal offences
 - ☒ sexual
 - ☐ theft
 - ☐ violent theft
-
- ☐ male
 - ☒ female
 - ☒ gender other
-
- ☒ guilty
 - ☒ not guilty
 - ☒ verdict other

all none





filters

records: 94 [speech lines](#)

Series

☐ offence

☐ gender

☒ verdict

☐ breaking peace

☐ damage

☐ deception

☒ kill

☐ miscellaneous

☐ royal offences

☐ sexual

☐ theft

☐ violent theft

☐ male

☒ female

☒ gender other

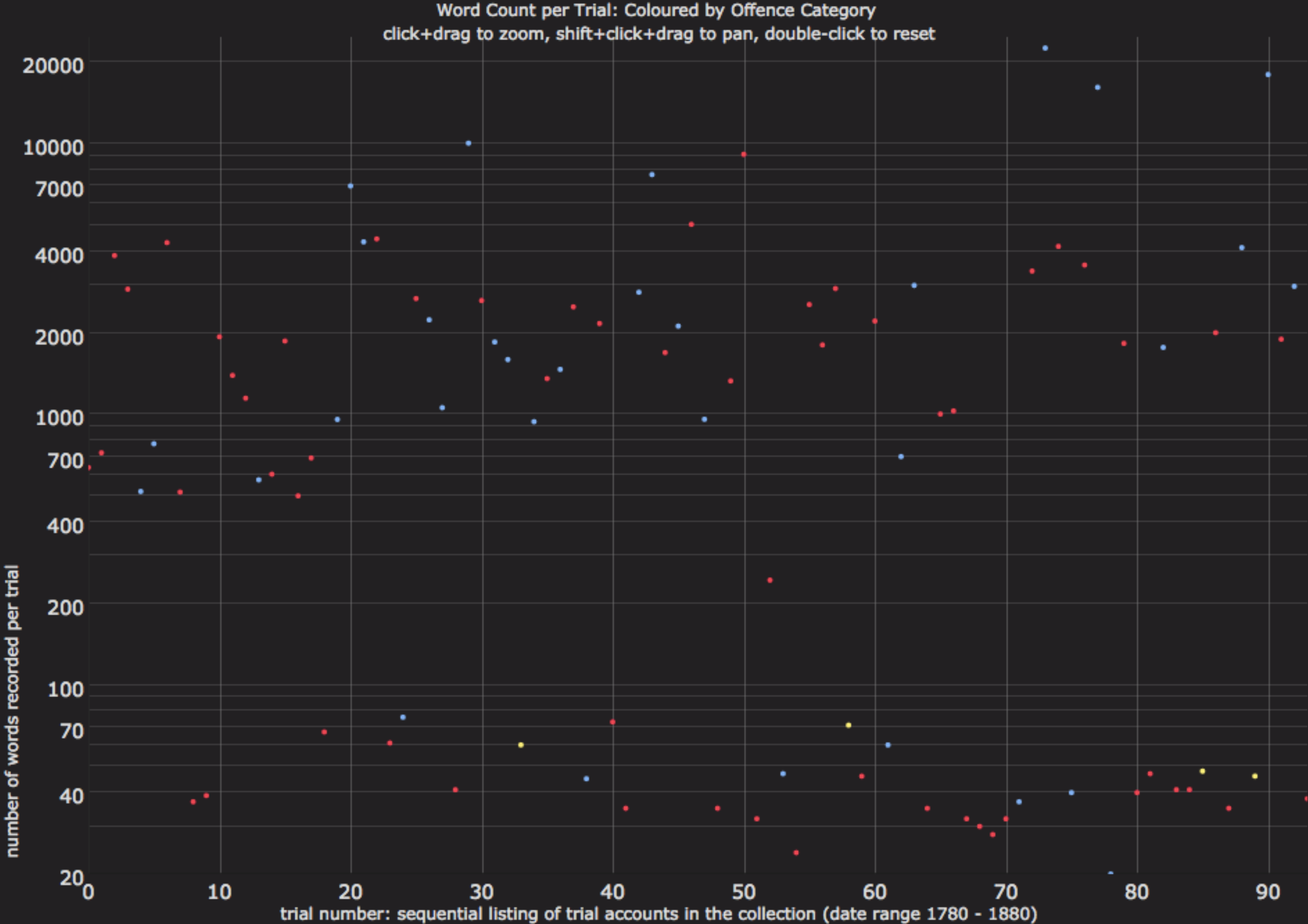
☒ guilty

☒ not guilty

☒ verdict other

all

none





filters

records: 32 [speech lines](#)

Series

☐ offence

☐ gender

☒ verdict

☐ breaking peace

☐ damage

☐ deception

☒ kill

☐ miscellaneous

☐ royal offences

☐ sexual

☐ theft

☐ violent theft

☐ male

☒ female

☒ gender other

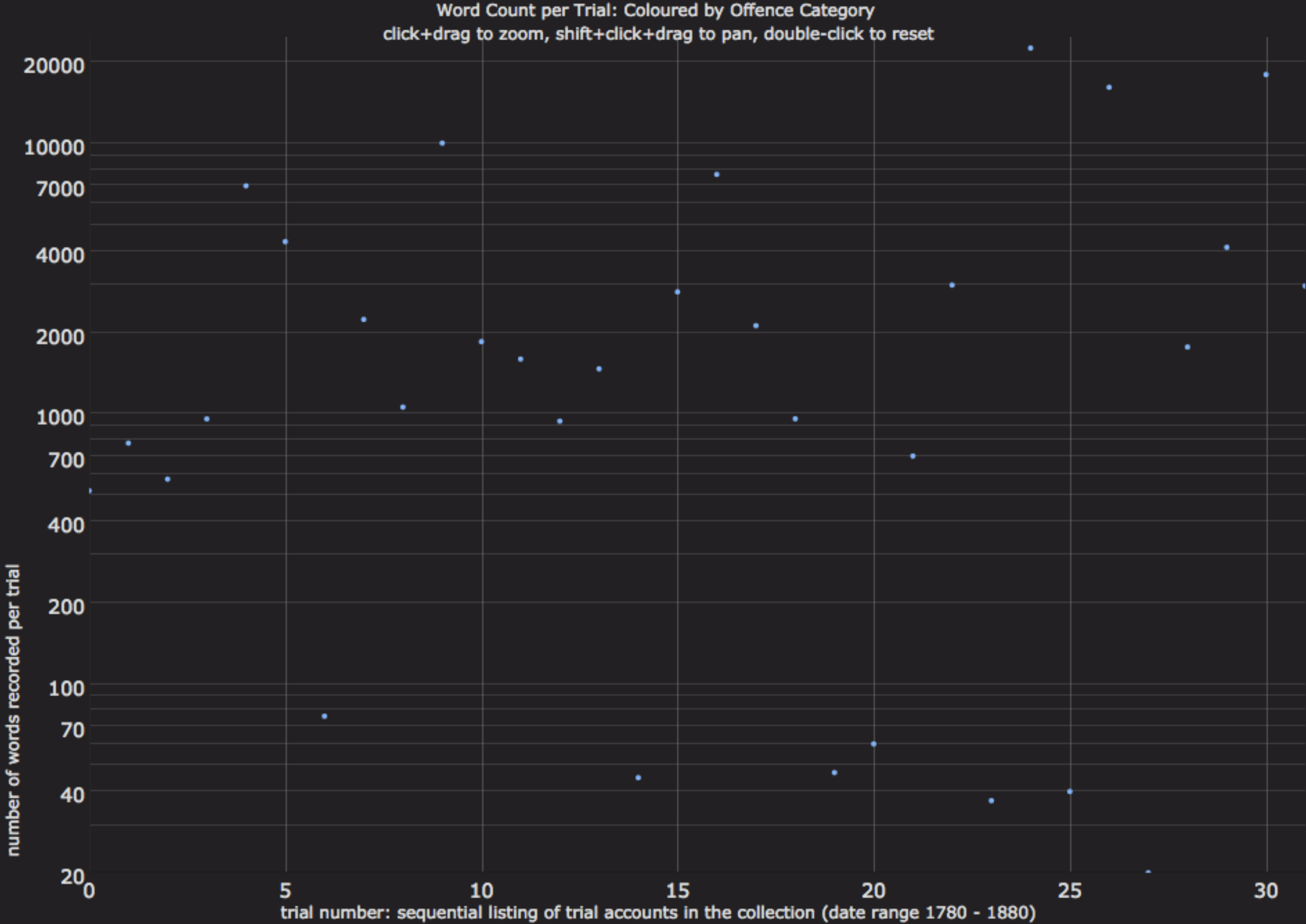
☒ guilty

☐ not guilty

☐ verdict other

all

none





filters

records: 32 [speech lines](#)

Series

offence

gender

verdict

breaking peace

damage

deception

kill

miscellaneous

royal offences

sexual

theft

violent theft

male

female

gender other

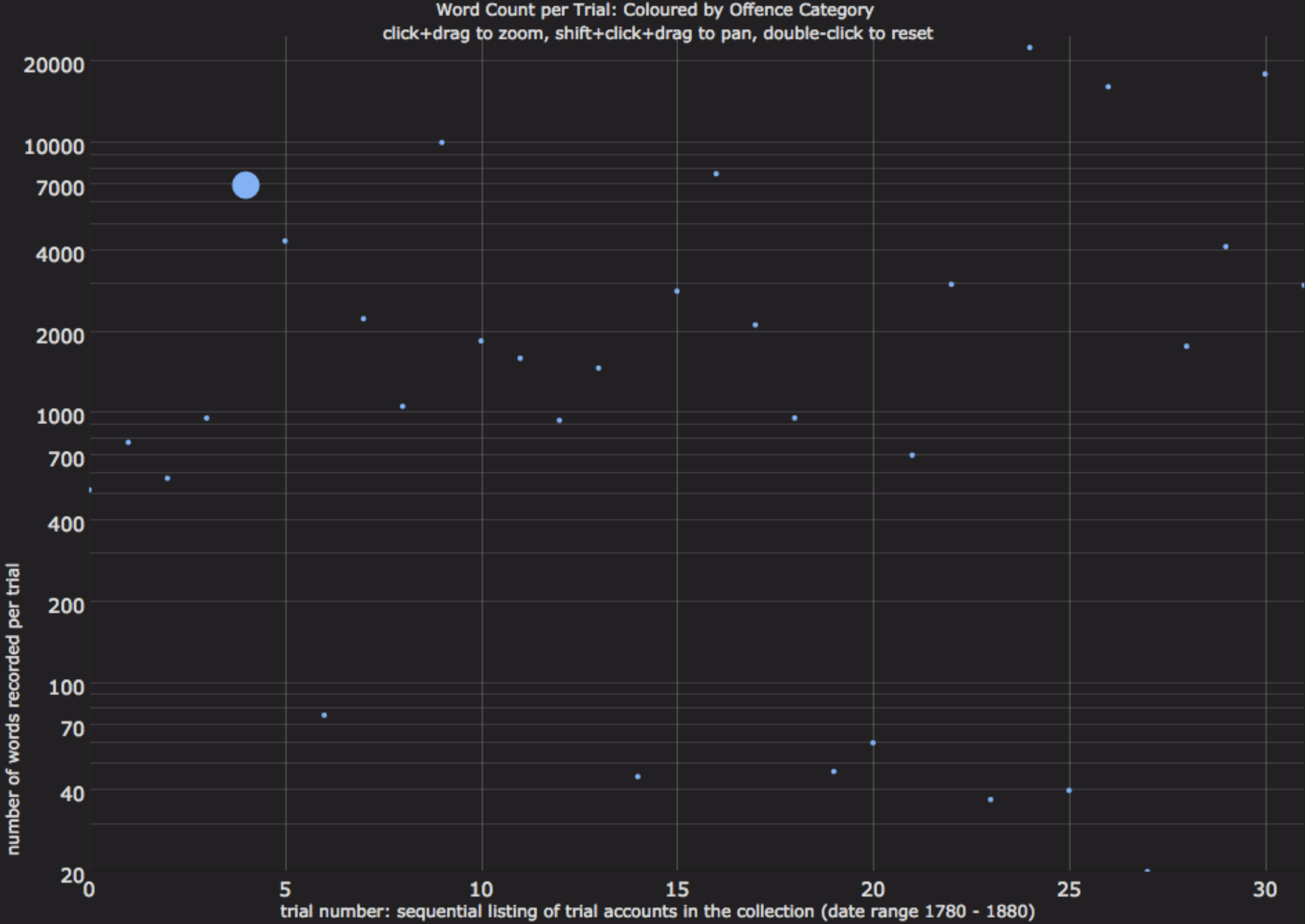
guilty

not guilty

verdict other

all

none



TrialId: t18350202-472

Wordcount: 6985

Def: ISABELLA HOPES

Sex: female

Off: kill

OSub: murder

Ver: guilty

VSub: withRecommendation

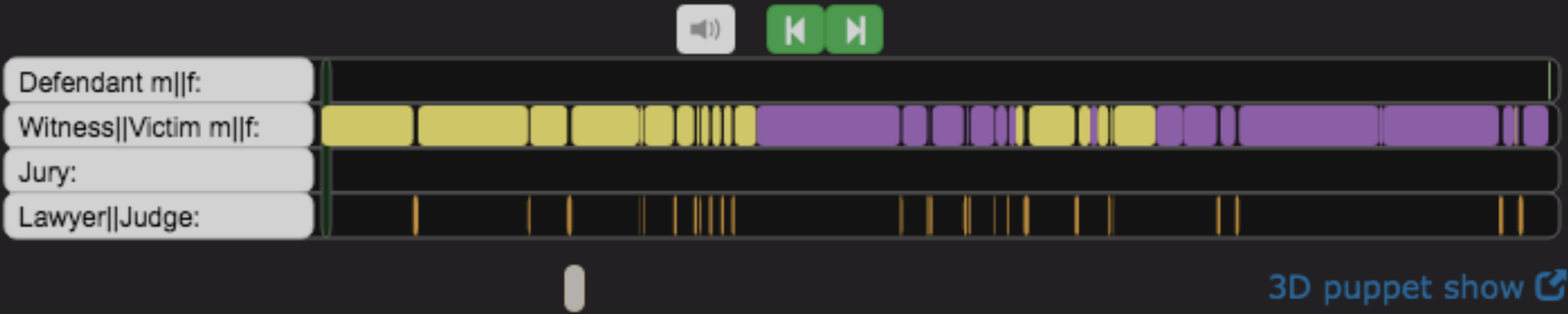
Pun: death

PSub:

Indictment: for that she on the 6th of November at Edmonton Middlesex unlawfully maliciously and feloniously did administer to and cause to be taken by Elizabeth Cambridge certain poison to wit one drachm of white arsenic the same being a deadly poison with intent feloniously wilfully and of her malice aforethought to kill and murder her against the Statute

Trial

t18350202-472



Account

[old bailey online](#) - [voyant tools](#)

472. ISABELLA HOPES was indicted for that she on the 6th of November, at Edmonton, Middlesex, unlawfully, maliciously, and feloniously did administer to, and cause to be taken by Elizabeth Cambridge, certain poison, to wit, one drachm of white arsenic, the same being a deadly poison, with intent feloniously, wilfully, and of her malice aforethought, to kill and murder her, against the Statute.

2nd COUNT, calling that which was administered "a certain destructive thing."

MR. CLARKSON conducted the Prosecution.

ELIZABETH CAMBRIDGE . I carry on the business of a market-gardener at Tottenham, Middlesex. The prisoner was in my service in November last. On the 5th of November I had lost 3l. 10s.; I had lost money before, but did not distrust the prisoner—I cannot exactly say how much I lost altogether—in consequence of suspicion, I examined her clothes after she was gone to bed, between ten and eleven o'clock on the 5th of November—(we had no regular agreement for any wages)—I found £7 or £8 in her pocket—the money was in sovereigns and half-sovereigns, and there was some silver—she was asleep—I awoke her, and told her I felt very uncomfortable, that I missed the

Sessions Paper

2nd February 1835 - [voyant tools](#)

CENTRAL CRIMINAL COURT.
SESSIONS PAPER.
WINCHESTER, MAYOR.
FOURTH SESSION, HELD FEBRUARY 2, 1835.
MINUTES OF EVIDENCE,
Taken in Short-hand,
BY HENRY BUCKLER.
LONDON:
GEORGE HEBERT, CHEAPSIDE.
W. TYLER, PRINTER, IVY-LANE.
1835.
THE
WHOLE PROCEEDINGS
On the King's Commission of the Peace,
OYER AND TERMINER, AND GAOL DELIVERY
FOR

Summary

Defendant: ISABELLA HOPES - [digital panopticon](#)

Session date: 2nd February 1835
Offence: kill - murder
Outcome: guilty death
Age: 20
Occupation: unknown
Gender: female
Word count: 6985
Spoken word count: 6485

Indictment: for that she on the 6th of November at Edmonton Middlesex unlawfully maliciously and feloniously did administer to and cause to be taken by Elizabeth Cambridge certain poison to wit one drachm of white arsenic the same being a deadly poison with intent feloniously wilfully and of her malice aforethought to kill and murder her against the Statute

Text Tools

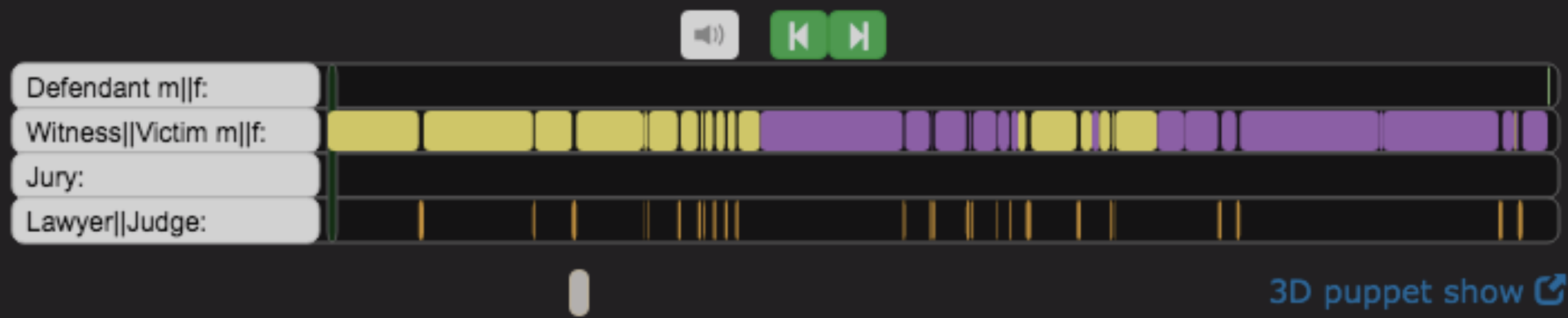
🔍 Selection

Hint! Select some text in the page

-

Trial

t18350202-472



Account

[old bailey online](#) - [voyant tools](#)

472. ISABELLA HOPES was indicted for that she on the 6th of November, at Edmonton, Middlesex, unlawfully, maliciously, and feloniously did administer to, and cause to be taken by Elizabeth Cambridge, certain poison, to wit, one drachm of white arsenic, the same being a deadly **poison**, with intent feloniously, wilfully, and of her malice aforethought, to kill and murder her, against the Statute.

2nd COUNT, calling that which was administered "a certain destructive thing."

MR. CLARKSON conducted the Prosecution.

ELIZABETH CAMBRIDGE . I carry on the business of a market-gardener at Tottenham, Middlesex. The prisoner was in my service in November last. On the 5th of November I had lost 3l. 10s.; I had lost money before, but did not distrust the prisoner—I cannot exactly say how much I lost altogether—in consequence of suspicion, I examined her clothes after she was gone to bed, between ten and eleven o'clock on the 5th of November—(we had no regular agreement for any wages)—I found £7 or £8 in her pocket—the money was in sovereigns and half-sovereigns, and there was some silver—she was asleep—I awoke her, and told her I felt very uncomfortable, that I missed the

Sessions Paper

2nd February 1835 - [voyant tools](#)

CENTRAL CRIMINAL COURT.

SESSIONS PAPER.

WINCHESTER, MAYOR.

FOURTH SESSION, HELD FEBRUARY 2, 1835.

MINUTES OF EVIDENCE,

Taken in Short-hand,

BY HENRY BUCKLER.

LONDON:

GEORGE HEBERT, CHEAPSIDE.

W. TYLER, PRINTER, IVY-LANE.

1835.

THE

WHOLE PROCEEDINGS

On the King's Commission of the Peace,

OYER AND TERMINER, AND GAOL DELIVERY

FOR

Summary

Defendant: ISABELLA HOPES - [digital panopticon](#)

Session date: 2nd February 1835
Offence: kill - murder
Outcome: guilty death
Age: 20
Occupation: unknown
Gender: female
Word count: 6985
Spoken word count: 6485

Indictment: for that she on the 6th of November at Edmonton Middlesex unlawfully maliciously and feloniously did administer to and cause to be taken by Elizabeth Cambridge certain poison to wit one drachm of white arsenic the same being a deadly poison with intent feloniously wilfully and of her malice aforethought to kill and murder her against the Statute

Text Tools

🔍 Selection

poison

[google ngrams](#) - [historical thesaurus](#)

🔍 Old Bailey Voices

🔍 Sessions Papers

🔍 Ordinary's Accounts

Found 89 Speech utterances in Old Bailey Voices containing the search term: poison

[speech lines](#)

Largest number of mentions of "poison" in a single trial is: 8

• t17810425-12 (1)

2nd COUNT, calling that which was administered "a certain destructive thing."

MR. CLARKSON conducted the Prosecution.

ELIZABETH CAMBRIDGE . I carry on the business of a market-gardener at Tottenham, Middlesex. The prisoner was in my service in November last. On the 5th of November I had lost 3l. 10s.; I had lost money before, but did not distrust the prisoner—I cannot exactly say how much I lost altogether—in consequence of suspicion, I examined her clothes after she was gone to bed, between ten and eleven o'clock on the 5th of November—(we had no regular agreement for any wages)—I found £7 or £8 in her pocket—the money was in sovereigns and half-sovereigns, and there was some silver—she was asleep—I awoke her, and told her I felt very uncomfortable, that I missed the money, and I said, "There is only you and your brother in my house"—(his name is Walker, he is a half-brother)—she said, "I hope you do not distrust me, Mrs. Cambridge"—my answer was there was only her and her brother in the house—I said, "Do you think your brother's key opens my box?"—she said, "I do not know, his key is in the box, and you can try it"—I did so, and it would not fit it—I told her I was deficient in my payments; I was going to take the money to London with me next morning, and did not know what I should do for want of it, and asked if she had any she could lend me—she said, "How much do you want?"—I said, "Three or four sovereigns"—I had not at that time told her I had discovered the money in her pocket—I asked her if I should hand her pocket to her—she said yes, if I pleased—I handed her her pocket—she took out a purse, and wished to give me what I had asked for, but she wished to keep the remainder, as I thought, out of my sight; it appeared so—she gave me four sovereigns—I save she had more, and said to her, "It is well for you, Isabella, that you have got more money than you knew of; I am afraid this is all my money which I have lost at different times."—I answered, "This money is not yours, it is the money of the Court."

Taken in Short-hand,
BY HENRY BUCKLER.

LONDON:

GEORGE HEBERT, CHEAPSIDE.

W. TYLER, PRINTER, IVY-LANE.

1835.

THE

WHOLE PROCEEDINGS

On the King's Commission of the Peace,

OYER AND TERMINER, AND GAOL DELIVERY

FOR

The City of London,

AND GAOL DELIVERY FOR THE

COUNTY OF MIDDLESEX, AND THE PARTS OF THE
COUNTIES OF ESSEX, KENT, AND SURREY, WITHIN THE
JURISDICTION

OF THE

CENTRAL CRIMINAL COURT.

Before the Right Honourable HENRY WINCHESTER, LORD MAYOR of the City of London; Sir John Vaughan, one of His Majesty's Justices of the Court of Common Pleas; Sir John Patteson, Knt., one of the Justices of His Majesty's Court of King's Bench; John Ansley, Esq.; George Scholey, Esq.; John Atkins, Esq.; Sir William Heygate, Bart; William Venables, Esq., Aldermen of the said City of London; the Honourable Charles Ewan Law, Recorder of the said City; Thomas Kelly, Esq; Samuel Wilson, Esq.; Sir Chapman Marshall, Knt; and John Pirie, Esq., Aldermen of the said City; John Mirehouse, Esq., Common Sergeant of the said City; and William St. Julien Arabin, Sergeant at Law; His

🔍 Selection

poison

google ngrams ↗

-

historical thesaurus ↗

🔍

Old Bailey Voices

🔍

Sessions Papers

🔍

Ordinary's Accounts

Found 89 Speech utterances in Old Bailey Voices containing the search term: poison

speech lines ↗

Largest number of mentions of "poison" in a single trial is: 8

•

t17810425-12

(1)

•

t17820515-27

(1)

•

t17850406-1

(1)

•

t17911026-63

(1)

•

t18000528-23

(1)

•

t18021027-34

(2)

•

t18060702-45

(1)

•

t18150405-18

(4)

•

t18261026-215

(1)

•

t18300218-6

(3)

•

t18350202-472

(5)

•

t18360404-906

(6)

•

t18451027-1974

(2)

•

t18451027-2153

(3)

•

t18451027-2154

(2)

•

t18451215-201

(4)

•

t18460706-1445

(3)

•

t18481218-261

(1)

•

t18500506-1032

(8)

•

t18520510-575

(1)

•

t18530103-223

(1)

•

t18570706-817

(2)

•

t18601126-38

(4)

filters

records: 21023 [speech lines](#)

Series

offence

gender

verdict

breaking peace

damage

deception

kill

miscellaneous

royal offences

sexual

theft

violent theft

male

female

gender other

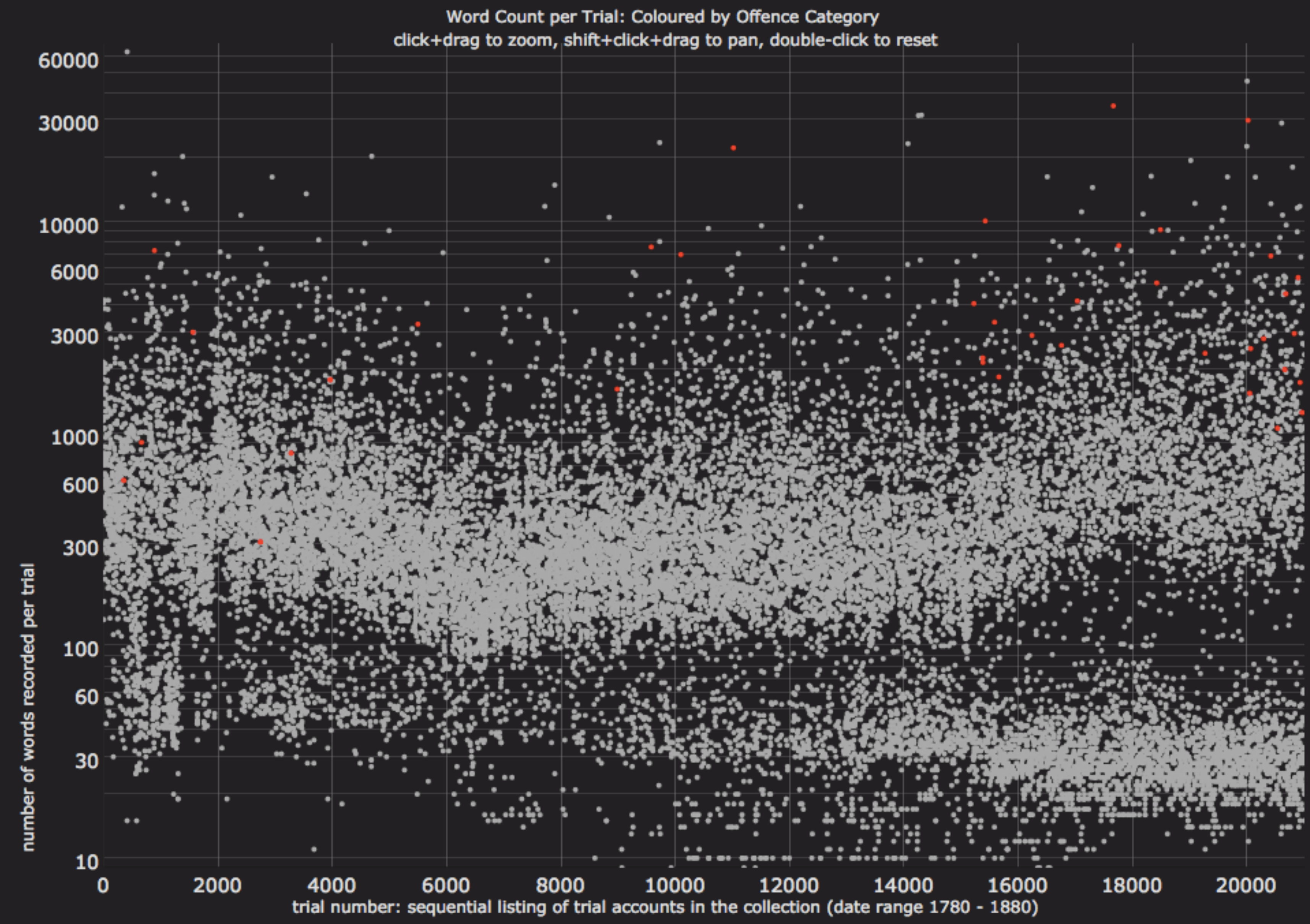
guilty

not guilty

verdict other

all

none



Graph these comma-separated phrases: ☐ case-insensitive

between and from the corpus with smoothing of [Search lots of books](#)

Share

Tweet

[Embed Chart](#)



Search in Google Books:

[1680 - 1809](#) [1810 - 1873](#) [1874 - 1885](#) [1886 - 1913](#) [1914 - 1930](#) [poison](#) English

[Home](#) > [Search](#) >

SELECT CATEGORY

You searched for:

Word: **poison**

[Refine your search](#)

Timeline visualization **Mini-timelines:** ☒ On ☐ Off

Sort results:

There are a total of **30** results.

WORD RESULTS:

01.02.04.04.03 | 07 vt.

Come by death :: Kill by specific method :: by poisoning **poison** (a1400–)

01.03.01.04.22 vt.

Cause to be ill :: Poison **poison** (1605–)

01.03.03.05.02.07 n.

Healing/cure :: Medicinal potion/draught **poison** (1377–1579)

01.07.01.22.04 | 02 vt.

Be food for :: Salt manufacture :: perform other salt-manufacturing tasks **poison** (1885–)

01.07.02.02 n.

Timeline visualization

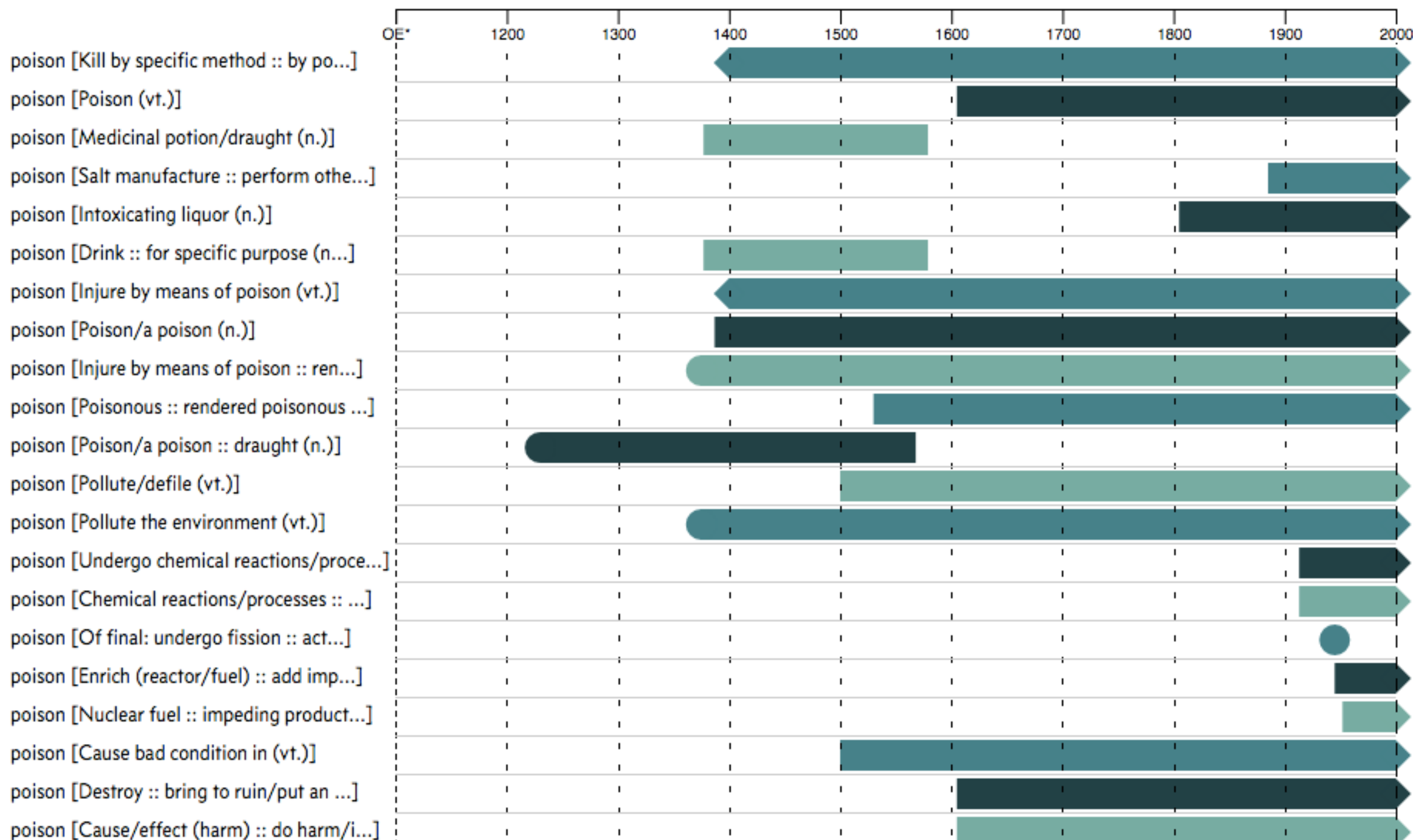


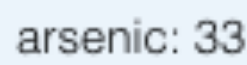
The timeline takes your search results and uses bars to represent stretches of time and circles for single dates. Should you wish to narrow your search down by date or by part of speech, please use our [Advanced Search](#). Select a bar or circle by clicking on it to see the underlying Historical Thesaurus dates used to create the visualization and also the full category title should it be truncated. Click on a word label to view the category.

▼ More

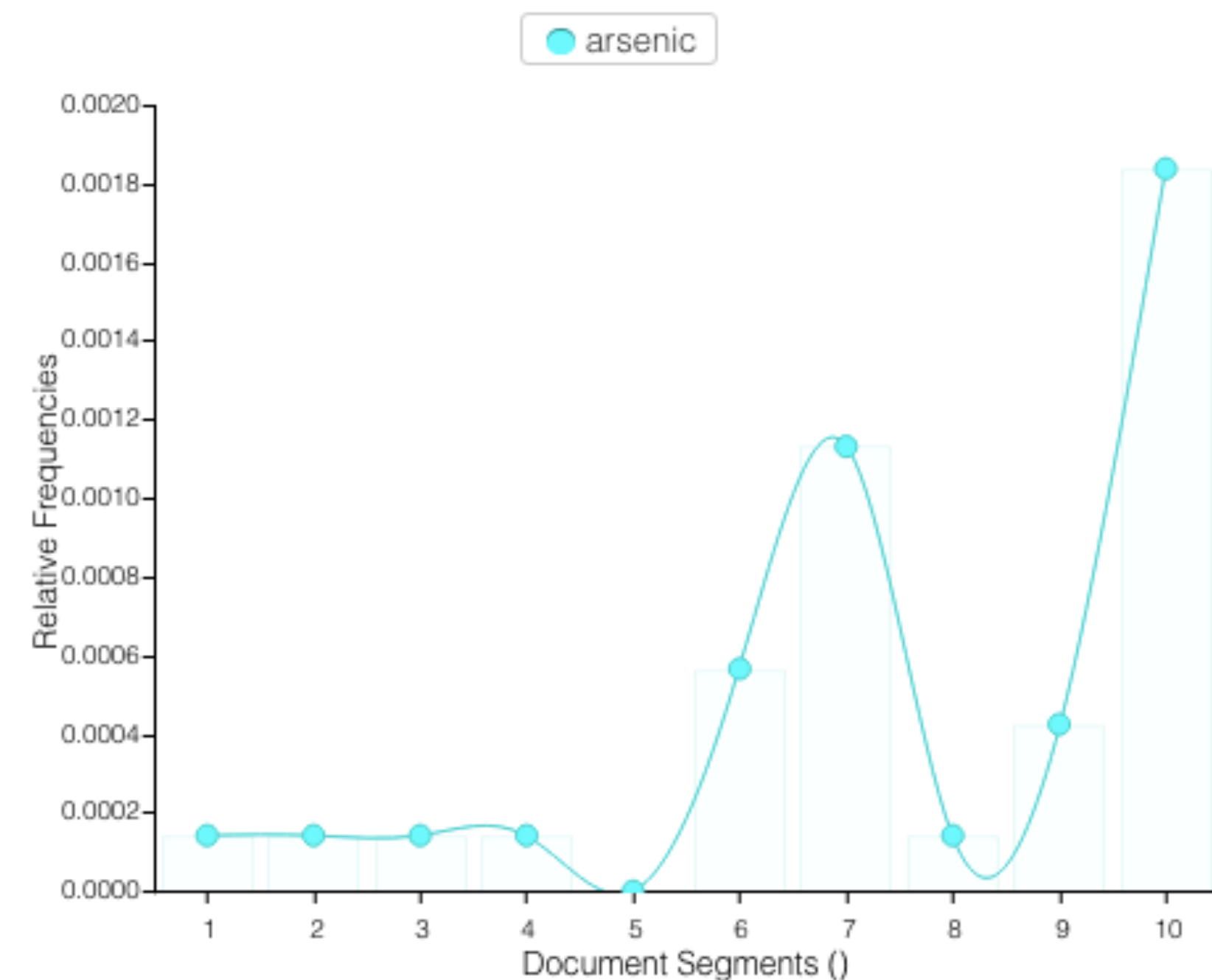
Sort by: **↓¹ Thesaurus Category** **⌚ First attested date** **↓^A₂ Alphabetically** **↔ Length of attestation**

Save Image (SVG)



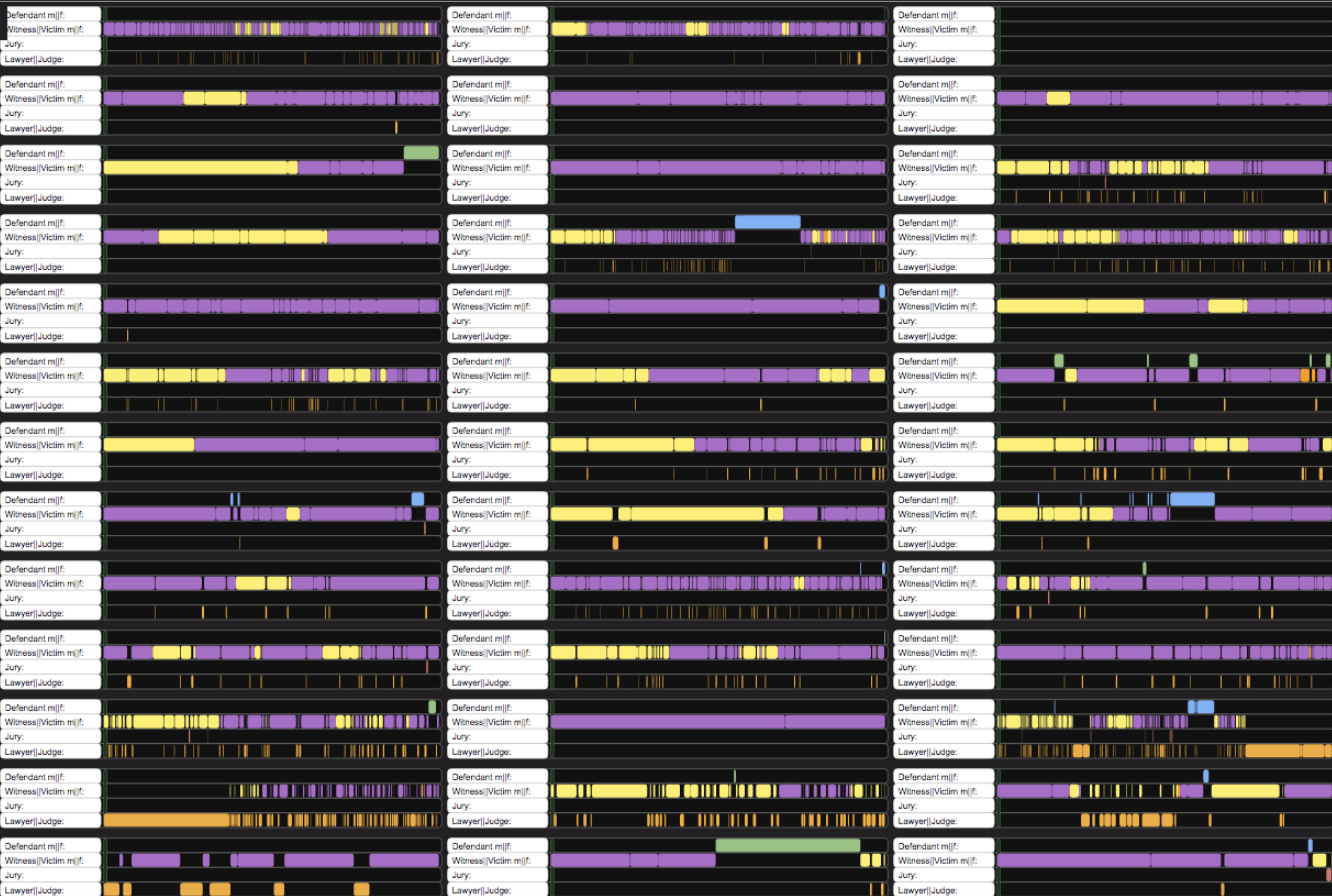


ELIZABETH CAMBRIDGE . I carry on the business of a market-gardener at Tottenham, Middlesex. The prisoner was in my service in November last. On the 5th of November I had lost 3l. 10s.; I had lost money before, but did not distrust the prisoner—I cannot exactly say how much I lost altogether—in consequence of suspicion, I examined her clothes after she was gone to bed, between ten and eleven o'clock on the 5th of November—(we had no regular



Navigation icons: back, forward, search, and other controls.

9



This money they swear to, is my own, I have saved up at a shilling a time. When I met this gentleman first, he was with a black woman with a white gown and white coat on. What he had, was entirely unbuttoned. I was at a distance, against the rails. I went down towards Pall-Mall; I stood upon the stone of a door in Gloucester-court. He asked if there was any house he could go into; I said there was a house there. I knocked at No. 3, and went in. He said, My dear, I have no money; I have been with a black woman; my money is all gone. He pulled out his pocket, and said, I have got a snuff box, and a watch, and a pin valued at so much, and a pocket-book at so much, which he could not part with. I said, if he had no money, I would not go with him. I said, As you have no money, I do not chuse to give my carcase up to you for nothing; and I hope you will give me liberty to get some water, for I am dry. He said yes; but he would keep my cloak till I came back. What he offered to me, was what is not fit: he is a man neither fit for God nor the devil; he is neither fit for a black woman, nor a white woman. What he expressed to me, put a shock upon my spirits, and frightened me. I went to open the door. The maid said, Take this man with you. I said, No; I do not chuse to have any more to do with him. I ran out at the door: I heard no more of the gentleman till I went out; and then the constable took me; and said, I had robbed him of so much. It must be meer wickedness, because, when he took me into the house, he said, he had not any, and because I would not condescend to his will, he charges me with this. I did not come into close quarters; I was at arm's length from him.



CALIBRE EBOOK MANAGER

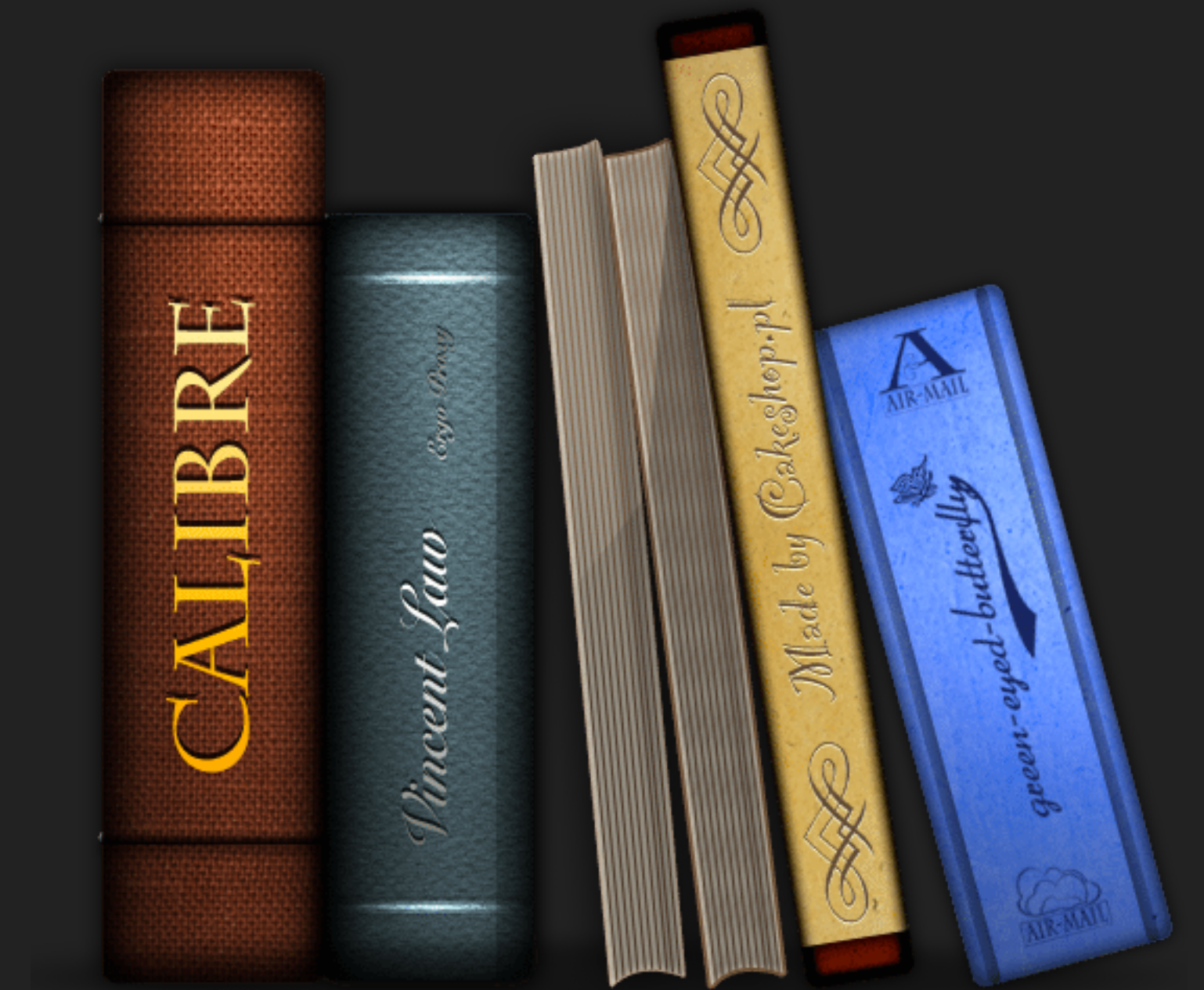
- ▶ Input

AZW, AZW3, AZW4, CBZ, CBR, CBC, CHM, DJVU, DOCX, EPUB, FB2, HTML, HTMLZ, LIT, LRF, MOBI, ODT, PDF, PRC, PDB, PML, RB, RTF, SNB, TCR, TXT, TXTZ

- ▶ Output

AZW3, EPUB, DOCX, FB2, HTMLZ, OEB, LIT, LRF, MOBI, PDB, PMLZ, RB, PDF, RTF, SNB, TCR, **TXT**, TXTZ, ZIP

- ▶ UTF-8





SUPPORT CALIBRE

ABOUT CALIBRE DEMO DOWNLOAD HELP WHAT'S NEW GET INVOLVED BUGS LANGUAGE

Click to watch calibre demo

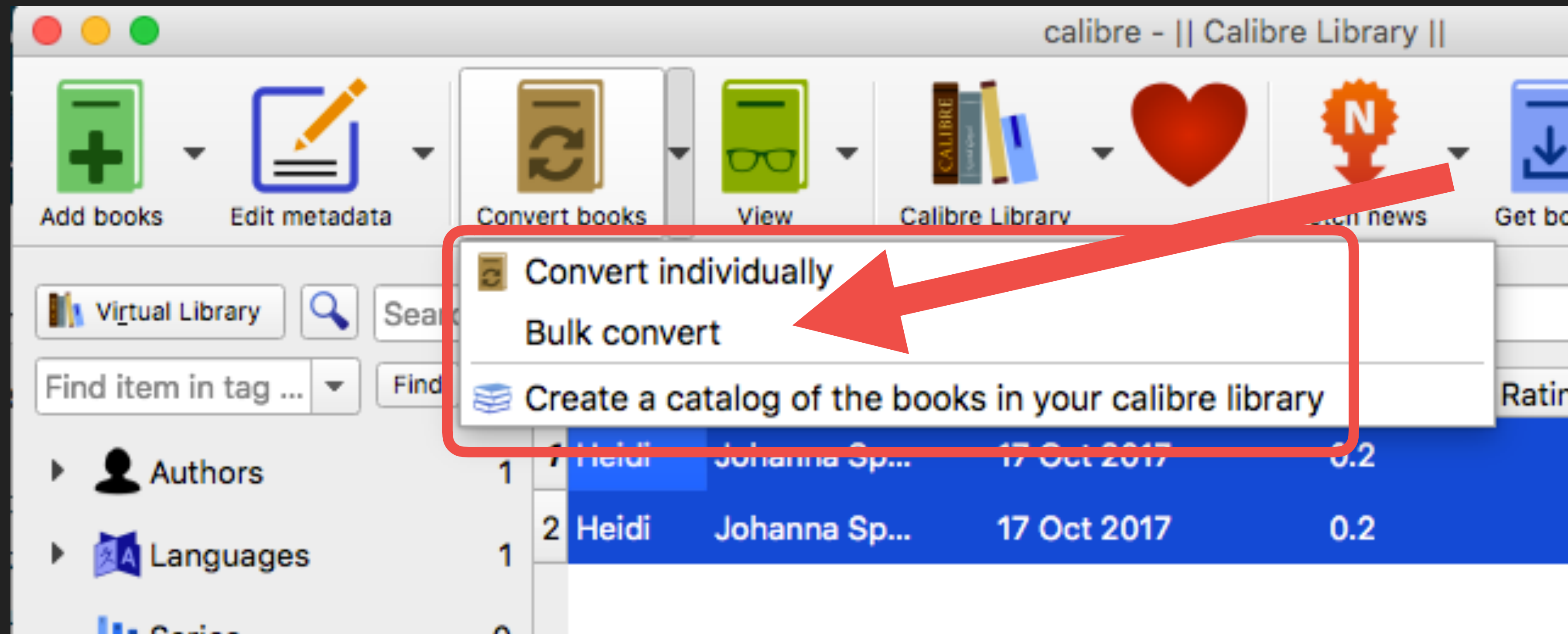


CALIBRE IS THE
ONE STOP SOLUTION
TO ALL YOUR
E-BOOK NEEDS.

DOWNLOAD CALIBRE ►

GET INVOLVED ►

Get DRM free ebooks



Where did it go? Do you trust it? How will you check?

BATCH JOBS

WORKSHOP MATERIALS

<http://benskitchen.com/txt/>

Heidi - eBook examples

Mass Observation 12th May Diaries (2015)

Old Bailey Sessions Papers

Please be respectful with your use of Mass Observation content.

TASK 1 – CALIBRE

- ▶ Add and remove content from the Calibre Library (Try the eBooks)
- ▶ Use Bulk Convert to process multiple files
- ▶ Export your text as UTF-8
- ▶ Move the results to a folder that you can use later today (Stick them on your desktop in a folder called output1) (Hint! retain a clear distinction between source files and output files)

Laurence Anthony's Website

[Home](#) [Resume](#) [Publications](#) [Software](#) [Classes](#) [Photo Albums](#) [Links](#) [Contact](#)

Software



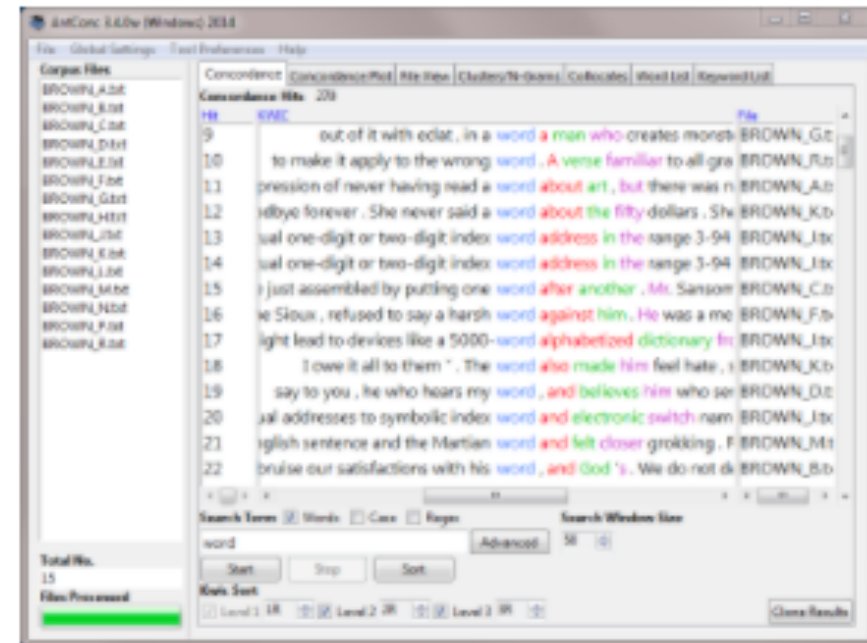
AntConc

A freeware corpus analysis toolkit for concordancing and text analysis.

[\[AntConc Homepage\]](#) [\[Screenshots\]](#) [\[Help\]](#)

Downloads:

- [Windows \(3.4.4\)](#)
- [Macintosh OS X 10.7-10.12 \(3.4.4\)](#)
- [Macintosh OS X 10.6 \(3.4.1\)](#)
- [Linux \(3.4.3\)](#)
- [Older versions](#)



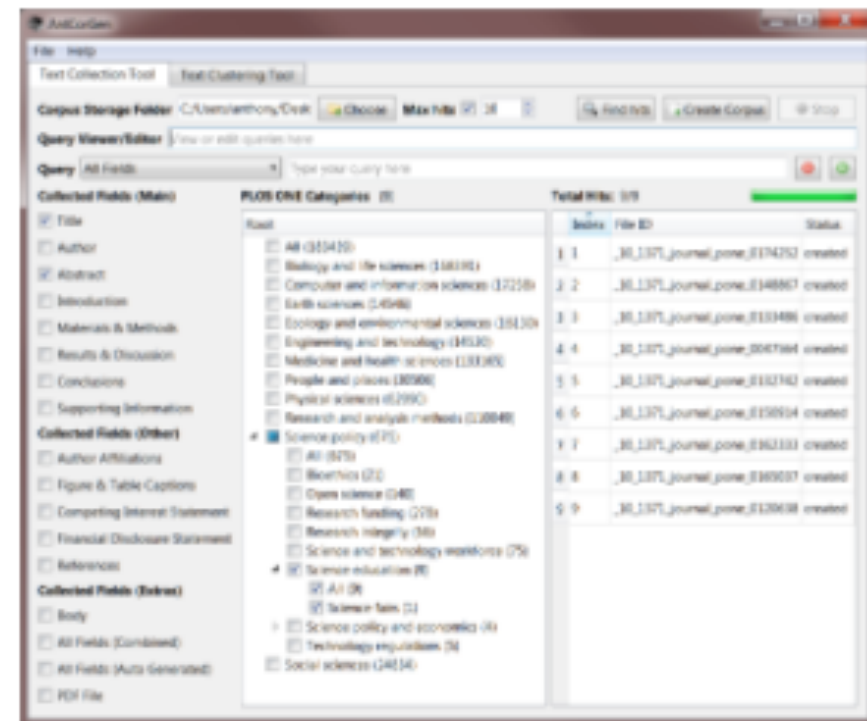
AntCorGen

A freeware discipline-specific corpus creation tool.

[\[AntCorGen Homepage\]](#) [\[Screenshots\]](#) [\[Help\]](#)

Downloads:

- [Windows \(1.1.0\)](#)
- [Windows 64-bit \(1.1.0\)](#)
- [Macintosh OS X \(1.1.0\)](#)
- [Linux \(1.1.0\)](#)
- [Linux 64-bit \(1.1.0\)](#)
- [Older versions](#)



AntPConc

A freeware ****parallel**** corpus analysis toolkit for concordancing and text analysis using UTF-8 encoded text files.

[\[AntPConc Homepage\]](#) [\[Screenshots\]](#) [\[Help\]](#)



Laurence Anthony Tools

AntFileConverter AntConc TagAnt

An alternative file conversion tool (because “PDF is a terrible format to convert from”)

TASK 2: ANT TOOLS

- ▶ Task 2a
- ▶ Run AntFileConverter
- ▶ Convert the 12th May 2015 Folder
- ▶ Check your results (You should have errors, I don't expect everyone to see them all.)



- ▶ Task 2b
- ▶ Run AntConc
- ▶ Try a single document, then try the whole folder



- ▶ Task 2c
- ▶ Run TagAnt
- ▶ Tag at least 1 document
- ▶ View your results



TASK 3 – VOYANT TOOLS

NB! Voyant can process pdf and word docs without cleaning

- ▶ Launch a 'Stand Alone' version of the Voyant Server
(Please don't upload MO docs to the public Voyant server!)
- ▶ Load a Session Paper, or MO document or folder of documents (Start Small)
- ▶ Explore the visualisation tools and layout options

TASK 4 – USE CONTENT IN THE SESSIONS PAPERS FOLDER FOR THIS ACTIVITY

Use your text editor to remove all xml tags - except paragraph tags

My Suggested process for this is:

1. Replace the tags you want to keep (<p></p>) with your own identifiers
2. Remove remaining XML markup
3. Replace your own identifiers (put the <p>s back in)
4. Deal with 'whitespace'
5. Consider doing this to a folder of documents

<[^>]+>

^(?:[\t]*(?:\r?\n|\r))+

WORKSHEET

<https://benskitchen.com/txt/>

FUN WITH TEXT EDITORS: SOLUTION

All of these operations use the Find & Replace feature, either in normal mode or in regular expression mode.

Steps:

1. Replace the opening paragraph tag (<p>) with a custom identifier (e.g ****ben****)
2. Replace the closing paragraph tag (</p>) with a second custom identifier (e.g ****/ben****)
3. This time use regular expression mode with the following pattern (without quotes) "**<[^>]+>**" to remove all remaining XML tags . NB the regular expression mode icon is normally dot star e.g. **.***
4. Re-insert the paragraph tags so your data has paragraph. (Otherwise you'll be left with one massive paragraph)
5. Use Regular expression mode to Remove unwanted white space and blank lines. (without quotes) "**^(?:[\t]*(?:\r?\n|\r))+**"