

Making Sense of the Emergence of Manslaughter in British Criminal Justice

Tim Hitchcock <hitchcock_dot_t_at_gmail_dot_com>, Professor Emeritus of Digital History, University of Sussex 
<https://orcid.org/0000-0002-4429-5870>

William J. Turkel <william_dot_j_dot_turkel_at_gmail_dot_com>, Professor of History, The University of Western Ontario 
<https://orcid.org/0000-0003-4734-9894>

DOI: <https://doi.org/10.63744/4t6z4ckh5xxf>

Abstract

Manslaughter emerged as a new and distinct category of crime amongst those tried at the Old Bailey in London in the first half of the nineteenth century. From being a rare charge in 1800, manslaughter came to represent over 60% of all trials for ‘killing’ by the 1850s. This article describes the methodologies used by the authors to explore this phenomenon via trials included in the *Old Bailey Online*. It details the use of unsupervised clustering, embeddings and relevance measures to map the changing language associated with the charge of manslaughter; and more importantly, describes the application of top-down ‘sense making’ methodologies to the resulting analysis. Along the way it argues for the importance of including qualitative judgements by subject specialists in the process of developing quantitative analyses. *Inter Alia* it suggests that the rise of manslaughter was the result of a complex set of forces including changing statute law, the rise of a professional police, changes in the administration of coroners’ courts, and a growing public intolerance of violence.

Introduction

The first half of the nineteenth century witnessed a fundamental transformation in the character of the British criminal justice. From a system dominated by judicial discretion, in which large numbers of people were sentenced to hang for minor property offences, the ‘bloody code’, Britain all-but abolished capital punishment for crimes apart from murder and treason within the British Isles. The processes that drove this transformation were complex. In part, these changes were the result of a series of Acts of Parliament, frequently referred to as “Peel’s Acts”. Passed between 1826 and 1832, these Acts effectively consolidated the statutory basis of the criminal justice system, established the Metropolitan Police, and renewed Britain’s commitment to hard labour, reformatory prisons, and criminal transportation as an alternative to hanging. The significance of this legislative revolution has long been acknowledged by historians [Devereaux 2023, chps. 8 & 9]. But, as part of the same transition, court behaviour – who was tried for what crimes – changed in ways that cannot be reliably attributed to the development of statute law. A prime example of this second sort of change is provided by the charge of “manslaughter”. This article uses the *Old Bailey Proceedings* and a methodology based on clustering and ‘sense making’ to interrogate the origins and significance of this new pattern of court behaviour, and to illustrate the methodologies used by the authors.

It is designed to describe the methodologies the authors used to explore this historical problem. It will not provide a clear explanation of precisely how or why manslaughter became a common charge at the Old Bailey in the first half of the nineteenth century. A comprehensive answer to these questions is beyond the scope of this article. Instead, it details the author’s collaborative research journey towards a better understanding of the issues raised, arguing along the way that by acknowledging and describing the specific roles of both domain and methodological expertise in that particular journey, we can help digital humanists in general to consider more precisely the roles of each.

“Manslaughter” had existed in common and statute law from at least the thirteenth century, but before 1800 was rarely tried at the Old Bailey [Coldiron 1950]; [Wiener 1999]; [Handler 2007]; [Ferner and McDowell 2006]. As Greg Smith comments, “Until the latter half of the eighteenth-century homicides by accident or manslaughter were not clearly distinguished within the law” [Smith 1999, 11]. And while there is strong evidence that early modern coroners handed down verdicts of manslaughter relatively frequently, these did not result in felony trials [Sharpe and Dickinson 2016]; [Wiener 2004, 25, fn63].

3

In the eighteenth century a conviction for manslaughter was normally punished with branding or burning on the hand or a small fine. This was increased to up to a year’s imprisonment by the Penitentiary Act of 1779, and forty-three years later, in 1822, increased to a maximum of three years imprisonment (19 Geo. III, c.74; 3 Geo. IV c.38). This was in turn increased to a maximum sentence of transportation for life following the passage of the Offences Against the Persons Act in 1828 (9 Geo. IV c. 31). Cases of manslaughter were also reserved to the superior courts following the Quarter Sessions Act of 1842 (5 & 6 Vic. c.38). But this apparently inexorable rise in the severity of punishment associated with manslaughter, and associated legal seriousness was not reflected in the behaviour of the courts. The vast majority of guilty verdicts for manslaughter continued to be punished by small fines, and periods of imprisonment ranging from a few days to a few weeks. At the Old Bailey, of the 140 guilty verdicts handed down between 1829 and 1850 — following the passage of the Offences Against the Persons Act — only four resulted in the maximum punishment of transportation for life.

4

By the end of the first half of the nineteenth century manslaughter had become the most likely charge to be brought in relation to any variety of unlawful killing. In 1850, in England and Wales as whole, 52 people were tried for murder, against 192 tried for manslaughter [UK Parliamentary Papers 1850, 53].

5

As an essentially new category of crime — at least at the Old Bailey — an exploration of “manslaughter” allows us to interrogate change at both the level of the trial (charge, verdict, punishment, etc.) and at the level of trials as genre or types of texts. Were these ‘new’ crimes, or just the same old crimes, labelled differently? Or are we looking at a genuinely new category of offence, reflecting a new intolerance of types of violence that would previously have gone unpunished? Or does the emergence of manslaughter simply reflect the changing administrative and legal framework within which these trials were conducted: changing statute law, the rise of the police as prosecutors, and the changing role of coroner’s inquests? In this article we describe how our ongoing study of the emergence of manslaughter, using the 453 trials for manslaughter heard at the Old Bailey between 1800 and 1850, allowed us to experiment with proof-of-principle tools that incorporate semantic interaction, model steering, and interactive visualization. More generally, our current work on manslaughter has provided an opportunity to reflect on the role of computation in the processes by which we make sense of evidence of the past. And although we have sought to explore the importance of different factors in driving change in the use of manslaughter by the court, this article is intended primarily to emphasize the microhistorical potential of the authors’ methodological approach, rather than a straightforward quantitative analysis of legal change.

6

Methods

The *Old Bailey Online* (Old Bailey Online) provides a consistent record of trials held at London’s Central Criminal Court between 1674 and 1913, making accounts of 197,752 individual trials available for computational analysis. These trials form a substantial record of the changing character of the British criminal justice system as it responded to the development of statute law, legal practice, and wider social norms. Created as a digital resource between 1998 and 2008, the important aspect of the *Old Bailey Online* for this article is the extensive XML mark-up of the 197,000 trials — each an unique collection of words. As a result, the site facilitates both statistical analysis on the basis of the mark-up (in this instance 453 trials for ‘manslaughter’), and linguistic analysis using all the tools of the Digital Humanities. We have been collaborating on research using the *Old Bailey Online* archive since 2007, self consciously combining statistical and linguistic analysis in pursuit of new questions, and new answers. Our typical practice involves iteratively writing bespoke code that draws equally on Hitchcock’s subject expertise in British social, legal, and criminal history, and Turkel’s methodological skills with programming, data science, and linguistics. Along the way we have created almost 2 million derivative files, including indexes, frequencies, n-grams, stable random projections, visualizations, and

7

other representations. Much of this work was undertaken as part of the *Data Mining with Criminal Intent* project (Data Mining with Criminal Intent) which effectively re-engineered the *Old Bailey Online* around an API that ensured the dataset on which this article relies is publicly available in a consistent form. For much of that work we have treated the *Proceedings* as a “massive text object”, something that can be understood in its entirety [Hitchcock and Turkel 2021a]. Since there are more than 127 million words of text, no one has been or will be able to read and make sense of them all.

One of the techniques that we have used several times in different ways is unsupervised clustering, since it does not require the laborious preparation of training sets of data. This is a very common way to study large document collections in the digital humanities so we will only mention a handful of examples. [Schmidt 2018] demonstrated Stable Random Projections, a kind of document sketch, by using the technique to perform unsupervised clustering on approximately 13.6 million volumes from the Hathi Trust. Schmidt showed that the volumes clustered at different scales, zooming in successively from language to subject, to genre, to date or style, to author, and finally to individual work. Unsupervised clustering is also very useful within more homogenous document collections. [Green, Feinerer, and Burman 2013] used the technique to discover underlying genres in the articles published in the journal *Psychological Review* between 1894 and 1903. The story of American Psychology in this period has usually been told in terms of competing schools representing behaviorist, structuralist, and functionalist positions. Green and his colleagues described a much more finely detailed collection of genres changing over time. As would be expected, some of these could be related to the better-known schools. The authors showed, for example, that representatives of the “notoriously diverse” functionalist approach shared a common vocabulary. Other genres (such as a Scripture/statistical cluster that popped up in 1894-95) reflected the fact that “Fashionable topics and approaches, then as now, experienced a ‘flash’ and then quickly faded away” (p 187). We ourselves used unsupervised clustering on stable random projections of the Old Bailey manslaughter trials to first get a sense of what kinds of situations might lead to an indictment for manslaughter [Hitchcock and Turkel 2021b].

Unsupervised techniques are wonderful for discovery. As the analyst or subject expert becomes increasingly familiar with a set of documents, however, they soon become aware of misfits between the clustering and their sense of the ground truth that the documents depict. Typical practice at this point is to adjust the parameters of unsupervised clustering and run the model again, in hopes that a better fit will be achieved. In *The Dangerous Art of Text Mining*, Jo Guldi argues that “a new *sensibility*” is required rather than “just another algorithm”. We need to embrace the “perspectival nature of algorithms” rather than “ignoring occlusions, silences, and uncertainties of interpretation in an eagerness to validate a new method in the field”. She refers to this sensibility with the phrase “critical search” [Guldi 2023, 124–125].

The bottom-up process from sources to information via searching, filtering, finding, reading, and extracting is nicely covered by the “foraging loop” portion of an influential model of the research process developed by [Pirulli and Card 2005]. The authors used cognitive task analysis and a think aloud protocol to study the research practices of intelligence analysts. They refer to foraging as a loop because the outputs of the process can be iteratively fed back into research to drive further information extraction. The result is an increasingly bespoke set of collections of trials whose shared character is obvious to the domain expert, but which do not conform to a formal, pre-defined, taxonomy. We note, however, that the bottom-up processes of information foraging have been extensively automated in the digital humanities, whereas top-down ones (i.e., critical search) have not [Tukel 2025]. Pirulli and Card describe a second loop, the “sense making” loop, which builds on the structured outputs of the foraging loop to create even higher-level structures like schemata and support for reasoning with explicit hypotheses. These, in turn, can be fed back in a top-down manner to direct the search for evidence, support, and relations. While automation of these sensemaking processes may appear in some systems designed for military and intelligence use, they are not at all common in the digital humanities. Instead, the sensemaking loop remains firmly in the domain of traditional, non-digital scholarship.

In what follows it is important to understand that our goal is not to design a system that is good for all digital humanists, or to argue that one technique is better than another for qualitative exploration of large document collections. Instead, we describe some experiments that we did to add top-down sensemaking techniques to our own bespoke code and research practice in pursuit of an answer to a specific historical problem. We are always happy to share our code as

open source,^[1] but our aim is to broaden practice in the digital humanities to include a consideration of these kind of techniques. Digital humanists already make use of topic modeling and unsupervised clustering; we are suggesting that they may wish to consider incorporating top-down practices like those sketched below which make use of semantic interaction, model steering, and interactive visualization driven by expert judgments. Note that the semantics in these techniques are not provided by a computational technique, but by the human in the loop.

Case Study – Manslaughter at the Old Bailey

In order that we should have some machine-readable information that reflected Hitchcock's growing understanding of nineteenth-century manslaughter, he wrote a two- to three-line précis for 71 of the 150 manslaughter trials of the 1830s (see Appendix). These were selected for both their length (over around 250 words), and their chronological distribution through the decade. Some obvious categories emerged from this process of annotation, including fights (especially pub and alcohol-related fights), medical malpractice, traffic accidents, domestic violence, and cases where a child was killed. We used this additional information to label trials and to understand how well various sensemaking techniques might work for us in practice. We did not, however, provide the information to the system in question. In a figure below, we show trials as nodes in a network color-coded with green for fights and orange for accidents. This information allows us to see immediately how the trials were clustered, but the labels were not used as inputs for the clustering algorithm.

12

In our explorations, we built on two basic assumptions. The first, which comes from information retrieval and text mining, is that similarity between keywords or keyphrases in documents is one way of establishing semantic similarity between the documents themselves. We worked with two different ways of automatically extracting keywords. The first was RAKE (Rapid Automatic Keyword Extraction) which is unsupervised and operates on an individual document in a single pass [Rose et. al 2010]. It splits the text at stop words and phrase delimiters, returning multiword keywords. RAKE works very well for abstracts and news articles. For our purposes, however, RAKE keywords are so specific that they do not generalize across trials. This is due, in part, to the fact that the trials contain large amounts of reported testimony. For example, high scoring RAKE terms from the manslaughter trials included a few phrases like "lord chamberlain's pay office" which might conceivably play a role in another trial, but also many phrases like "good tempered humane sort", "old lady joked afterwards", and "poor man groaning dreadfully", which were unique to a particular trial. The other way that we chose keywords was by calculating the well-known TF-IDF (Term Frequency Inverse Document Frequency) measure [Manning et al. 2008]. Since criminal trials contain names and location data which can be relatively distinctive — and thus would have a high TF-IDF — but which are irrelevant to our task, we limited potential keywords to those that appear in the dictionary. [Turkel 2011] provides an online demonstration of the use of the TF-IDF measure on a sample Old Bailey trial. Our second basic assumption, which comes from machine learning and information visualization, is that spatial metaphors make it easier to visualize and reason about document similarity and difference. Items that are like one another are positioned nearer in space; those that are less similar are more distant. This visualization usually makes use of dimension reduction.

13

Having adopted these two assumptions, we were able to experiment with semantic interaction and model steering, using a review article by [Endert, Bradel, and North 2013] as our jumping off point. Adopting the spatial metaphor, the authors describe 'direct manipulation' in terms of three principles. First, that the object under study must be continuously represented. In our practice to date, this has not been the case. After studying the output of an unsupervised clustering, we might return to the code that generated it and alter parameters before rerunning it. In the direct manipulation model, however, the researcher should be able to see how the changes that they make affect the visualization in real time. Second, direct manipulation requires that the user press buttons or make physical actions (like adjusting sliders or dragging interface elements) rather than inputting commands with complex syntax. Again, our prior practice violated this principle. Finally, there should be "rapid incremental reversible operations whose impact on the object of interest is immediately visible" [Endert, Bradel, and North 2013, 6]. This allows the researcher to test interpretations in real time and backtrack as necessary.

14

From the programmer's perspective, one obvious way to implement direct manipulation of unsupervised clustering or dimension reduction would be to give the user a manipulable interface with a graphic display of the clusters and sliders or buttons to parameterize the algorithm. Endert and colleagues give an example of an interactive interface for PCA

15

(Principal Component Analysis). There are several problems with this approach, however, as they note.

First, many analysts aren't experts in complex mathematical models and thus don't understand the meaning of the parameters for the interactive controls. Second, analysts think about and understand the documents at the semantic level, yet the interactive controls for the models operate at the lower syntactic level of the model parameters. This creates a mismatch. Third, when analysts haven't yet gained a good understanding of the documents and their insights are still informal, they don't yet have a basis for expressing their inputs into the formal model parameters (p8).

16

They suggest that direct manipulation can be applied instead at three levels of interactivity within a spatialization: direct manipulation of constraints, of parameter weighting, or of a machine learner.

17

As an example of direct manipulation of spatialization constraints, [Endert, Bradel, and North 2013] discuss the Dust & Magnet tool of [Yi et al. 2005]. In this visualization, the particles of dust represent documents or other entities of interest from a dataset. A 'shake dust' mechanism separates particles so that they do not occlude one another. The researcher can then place 'magnets' into the space which attract (or optionally repel) dust particles to the extent that they contain particular keywords or have other relevant data dimensions. Yi and colleagues showed examples of the system in use with a relatively structured dataset of 77 cereals, where each item was labeled with twelve attributes including manufacturer, type, calories, protein, fat, and so on. Placing magnets for sugar and fat in the space, for example, makes it easy for the analyst to visually survey the dataset along these dimensions and identify cereals that are low in sugar and fat (since they are not attracted by either magnet) but, say, high in protein and vitamins. One aspect of their tool is that it is animated so that dust particles move as the user moves the magnets, and the more a particle is attracted to a magnet, the more quickly it moves. We experimented with a simplified implementation of the dust & magnet idea by separating trials with a pair of magnets for fight-like and accident-like keywords. We did not implement the animation, however. While appreciating the directness of the metaphor and its ease of use, our sense was that the technique was better suited to a structured dataset and to research questions that are more easily approached through multivariate information visualization.

18

We next turned to direct manipulation of parameter weighting for spatialization. The ForceSPIRE system for text analytics ([Endert, Fiaux, and North 2012]) involves many parameters that can be directly manipulated, including searching, highlighting terms, annotating, and moving documents in the space. Each of these is directly connected to the underlying dimension reduction model. We chose to experiment with the idea of using keyword highlighting to increase the weight of the keywords in the distance metric. For spatialization, ForceSPIRE uses a force-directed graph model with documents as vertices. Edges represent keywords shared between documents. In our implementation we started by fully interconnecting all trials that shared high TF-IDF keywords, then used term highlighting as a way of increasing the strength of connections so related documents would move closer to one another in the space. For reasons that will be discussed below, we next limited trial document interconnections to highlighted terms only, starting with fully connected graphs, and we eventually settled on a scheme whereby each trial document that contained a highlighted keyword was linked to the trial document which had the highest TF-IDF value for that keyword. Part of the interface for our proof-of-principle system is shown in Figure 1.

19

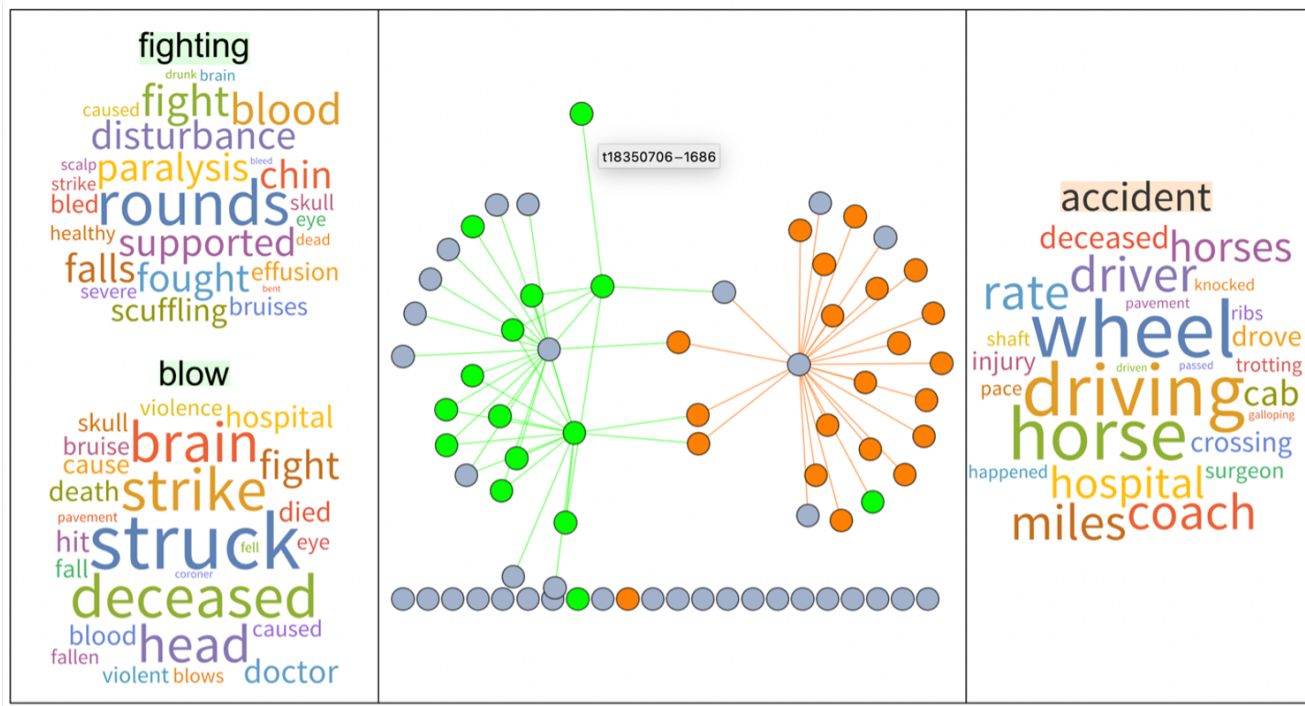


Figure 1. Term Highlighting Introduces Edges in Force-Directed Layout. Orange nodes represent trials that the analyst has determined are traffic accidents, green nodes represent fights. Gray nodes have not been categorized yet. These categories are not provided to the spatialization algorithm. They provide the human analyst with a visual sense of the effectiveness of particular semantic interactions. The left and right panels show the three highlighted keywords, and other high TF-IDF terms that are highly predictive of each highlighted term (and vice versa).

Each trial document is represented as a vertex in the graph shown in the center panel. If you hover the mouse cursor over a node, a tooltip displays the Old Bailey trial number. In this screenshot the mouse was over the uppermost green colored vertex which links to trial *t18350706-1686*. In our system these trial IDs link to local copies of trial texts and other derived representations; they also can be used to directly retrieve page images, text or XML from the Old Bailey Online site. On the right-hand panel, the keyword ‘accident’ has been highlighted by the researcher in orange. As a result, all the orange-colored trials except one have been joined to a central node (the trial where the keyword ‘accident’ has the highest TF-IDF measure). On the left-hand panel the two keywords ‘fighting’ and ‘blow’ have been highlighted in green. Multiple highlights in the same color allow the researcher to group related terms and deal with synonymy. All the green-colored trials except one have been joined to one or both of the central nodes for ‘fighting’ and ‘blow’. Since some trials contain both accident and fight keywords, the two subgraphs are joined by edges. The orange and green coloring of nodes reflects the annotations made by Hitchcock manually (so we can see how well the method works) but this information is *not* provided to the spatialization algorithm. Note that there are several trials along the bottom of the center panel which do not contain any highlighted keywords and are thus not connected to the graph.

The lefthand and righthand panels contain a further refinement that we adopted from TextTonic, another interactive visualization system that uses semantic interactions for exploring very large unstructured document collections [Paul et al. 2018]. In TextTonic, the spatialization is applied to RAKE keywords rather than documents, and the user directly manipulates clustering and display of keywords through searching, dragging them, ‘pinning’ a keyword in place, and resizing a term to increase its importance. Double clicking a keyword provides a snippet of sample text and double clicking the sample gives the full source. From TextTonic we took the idea of using an *F*-score to assess the degree to which one keyword predicts another. We used the balanced *F1* score which weights precision and recall as equally important and takes their harmonic mean. It is calculated by determining the number of true positives (*tp*: documents which contains keyword *a* and also contain keyword *b*), the number of false positives (*fp*: documents which contain keyword *a* but do not contain keyword *b*), and the number of false negatives (*fn*: documents which do not contain keyword *a* but do contain keyword *b*). The *F1* score is then $2tp/(2tp+fp+fn)$. It ranges between 0 and 1 and is commutative, that is $F1(a,b)=F1(b,a)$.

In the annotated manslaughter trials of the 1830s, the high-TF-IDF term ‘accident’ is highly predictive of the high-TF-IDF term ‘wheel’ and vice versa, $F1(\text{accident}, \text{wheel})=0.732$. The term ‘fighting’ predicts the term ‘drunk’ with the lower $F1$ value of 0.308. Both ‘accident’ and ‘fighting’ predict ‘pavement’ with $F1$ values 0.343 and 0.133 respectively, and ‘deceased’ with $F1$ values of 0.444 and 0.231 respectively. In our proof-of-principle system, when a term is highlighted, the researcher is given a list of other terms that are highly predicted by the selected term. Here the information is displayed in the form of word clouds, but it could be presented as a sorted list or in some other fashion. This facilitates searching and highlighting but it serves another purpose. After highlighting the single term ‘accident’ the researcher can see immediately that the language of accidents in the manslaughter trials is largely a language of traffic accidents (since we see ‘wheel’, ‘driving’, ‘driver’, ‘horse/s’, ‘coach’, ‘cab’, ‘pace’, ‘trotting’, and ‘galloping’). The language of fights, on the other hand, emphasizes both activity (‘blow/s’, ‘strike’, ‘struck’, ‘falls’, ‘hit’) and damage (‘paralysis’, ‘blood’, ‘bled’, ‘bleed’, ‘bruise/s’, ‘chin’, ‘skull’, ‘eye’, ‘head’, ‘brain’).

22

In unsupervised clustering and in visualizations that use semantic interaction, typical practice is to use all the available information in spatialization. For this reason, we originally experimented with fully interconnected force-directed graph layouts. That is to say that the length of an edge between two vertices was determined by considering all shared keywords, and there was an edge between every pair of documents that shared at least one keyword. The effects of direct manipulation were then applied by changing relative weights in the complete set of interactions. As mentioned earlier, we gradually abandoned this approach in favor of a much more sparsely connected graph, where connections are made through the process of sensemaking. It was consideration of the $F1$ scores that led us to abandon complete interconnection. In our application, a pair of criminal trials may share incidental details such as surnames. To us it does not matter if one Mr. Smith is the victim of a traffic accident and a different Mr. Smith kills someone in a fight. So even though a pair of trials may share the keyword ‘smith’ we do not want that to affect the spatialization. Similarly, since someone in a manslaughter trial is certain to be ‘deceased’ (although that word may not always be used), we do not want that keyword to affect spatialization either. Starting with fully interconnected graphs allows unsupervised clustering to work, but it also obscures the connections that will become important for interpretation in a mass of connections that are incidental. By allowing the sensemaking work of the researcher to guide spatialization from the top-down, a clearer picture of the researcher’s interpretation can form more quickly.

23

Let’s return to Figure 1 and explore some of the other ways that it reflects and guides an emerging understanding of the manslaughter trials of the 1830s. First note that there is a single green ‘fight’ node that is linked to the ‘accident’ subgraph but not to the other fights. This is trial *t18300218-65* (t18300218-65). Hitchcock’s annotation read “watchhouse cell fight, involving two men, with hospital surgeon giving evidence. Defendant does not give evidence. N[ot] G[uilt].” When Turkel checked the trial text he found that the keyword ‘accident’ was used three times in the text and had a high TF-IDF. The surgeon giving evidence in the case stated “the wound in the [victim’s] head was entirely the cause of his death; I should think the most probable cause of the wound was falling on the stone cill described by the witness - the blow was a severe one, but I cannot form an opinion whether he himself might not have caused it, by stumbling against the cill, or by being pushed.” In this instance, the intention was to identify an edge case, an obviously anomalous trial, that could be used by the domain expert to interrogate how the system was working. Systems like ForceSPIRE and TexTonic are often tested by asking analysts to “think aloud” while using them. Since there are only two of us, this happens naturally in conversation. But to give a sense of what these conversations entail, Hitchcock provided a written description of his thinking in annotating this case as a fight:

24

On the Wilkinson trial, I can see why the TF-IDF came up as an accident, and I suspect I was second guessing the text myself. There are two things that made me label it a fight. First, the quantity of blood; and second, the general rush to declare it an accident. In my experience it is pretty hard to kill yourself by accidentally falling over and hitting your head on a ledge halfway up a wall (I have tried); and if you do manage it, getting blood all over the place seems odd. And if it was an accident, why did Wilkinson not call out, and instead waited until the watchhouse keeper came down 45 minutes after putting Wilkinson in the cell? And if the dead guy could chat with three different people in the three weeks before his death, including about the seriousness of his wound, why he didn’t say something about how it happened, is just a bit weird. Just looking at the account,

it feels like the sort of thing where there were lots of opposing narratives presented at the coroner's inquest (hence the trial), that gradually got edited out on the way to court. On the face of it, there is no obvious reason it has come to trial at all. But again, I suspect that this is just me second guessing!

Since our proof-of-principle system allows us to use one form of direct manipulation for semantic interaction (term highlighting) while keeping another form separate from the spatialization process (trial color-coding and annotation), it allows us to see and discuss mismatches that are directly relevant to expert interpretation.

As further examples, consider the two trials color coded as fight and accident respectively that are not connected by the three terms that have been highlighted so far. Hitchcock's annotation for the fight (*t18350817-1883*) reads "Domestic violence, wife dies following a fight, with alcohol abuse involved. A fair amount of medical evidence. The defendant does not give evidence. Guilty - 1 year." As terms for domestic violence are highlighted, this node may become attached to a new subgraph. Hitchcock's annotation for the accident (*t18351123-131*) reads "Shipping accident on the Thames at Deptford. Male victim, male defendant. Long detailed statement from the second mate - but no other witnesses. Medium length. No medical testimony, or Defendant's statement. N[ot] G[uilt]." Unlike most of the accidents in the accident subgraph this case was not related to carriage traffic. Instead, a man was killed when a steamer paddle boat struck a launch. Future term highlighting might draw this case into the accident subgraph or might lead to a new focus on shipboard killings.

Finally, consider the trial at the center of the accident subgraph, which has not been color-coded as an accident. In this case (*t18300916-154*) the term 'accident' is used thirteen times, giving it the highest TF-IDF value for that term among all the annotated manslaughter cases. But this is an accidental death of a different kind. Hitchcock's annotation reads "Medium to long trial, male defendant and victim. Death occurs a while after the def. shoves the victim, who falls over and hurts his hip - later dying of dysentery. Lots of medical evidence, no defendant statement. N[ot] G[uilt]." As with the other cases discussed here, the proof-of-principle system directed our attention to trials that were highly relevant for our emerging interpretation, precisely because the spatialization was driven solely by that interpretation and not obscured by incidental connections.

One other aspect of our system which has not been shown is that it provides keyword in context (KWIC) listings for high TF-IDF terms. For example, the analyst might confirm that 'rounds' is being used in a fight-related sense in the manslaughter trials (Figure 2). This feature can be used to determine which terms to highlight, because they used unambiguously, and which might confuse the boundaries of interpretive categories.

would	beat	hardaway	in	two	more	rounds	at	the	last	round	edwards	fell
them	and	saw	six	or	seven	rounds	it	up	peared	a	fair	fight
clenched	his	fist	for	about	two	rounds	they	had	a	regular	fight	gaywood
worst	of	it	after	about	two	rounds	they	made	a	sort	of	stand

Figure 2. KWIC listing for 'rounds'

The terms that we highlighted for our example reflect the categories that Hitchcock was developing as he read and annotated some of the manslaughter trials. Other terms, like 'surgeon', 'inquest', 'spirits', or 'drunk' also occur frequently in the manslaughter trials, but they crosscut these categories. In future versions of our system, we plan to add a tabbed interface that allows the researcher to work with different sets of highlighted terms, creating new associations and proximities to investigate in parallel.

Historical Significance

Our work with manslaughter continues. Thus far, this work has allowed us to identify significant clusters of trials – suggesting the charge was being used as a portmanteau category bringing together dispirit forms of killing – domestic violence, drink related fights, traffic and industrial 'accidents' and medical malpractice. The circumstances that brought these varieties of killing into the larger category of 'manslaughter' remain complex and elusive. In part this transition

resulted from developments in statute law. The 1822 Felonies Act raised the maximum sentence for manslaughter from one to three years in prison, and five years later, the 'Offences Against the Person Act' of 1828 substantially increased this to transportation for life (3 George IV, c.38; 9 George IV, c.31). The punishment of transportation for life was re-enacted as part of the Offences Against the Person Act of 1827 (7 William IV & 1 Victoria, c.85). The 1820s and 1830s also saw the abolition of the 'Bloody Code', with its reliance on exemplary execution as the principal form of punishment, and its overwhelming focus on property crime. By 1841 only treason and murder remained subject to the death penalty. And in the place of execution, imprisonment and transportation emerged as forms of punishment that could be readily calibrated to the severity of the crime. In combination, these developments made manslaughter an attractive charge and verdict, as the punishment could range from a day in jail or small fine, to imprisonment or transportation for life [Wiener 1999]. But, nowhere in the legislation was domestic violence, medical malpractice or industrial and traffic accidents identified as the object of statute law.

Other factors were also involved. The creation of the Metropolitan Police in 1829, and their increasing role in the prosecution of crime also played a part [Emsley 2010]. Approximately a third of all prosecutions for manslaughter after 1829 either mention the presence of a police officer or feature an officer as a witness. The presence of a 'bobby on the beat', with extensive local knowledge, could help explain increasing numbers of domestic violence prosecutions.

31

Similarly, the changing role of coroners, whose inquests made the preliminary judgement about the cause of death might help account for growing numbers of accidental death and medical malpractice cases. In Westminster, for example, the 1830s witnessed a substantial increase in the overall number of inquests held, and the number per 100,000 of the population. An increasing percentage of these inquests were returning verdicts of accidental death — including industrial and traffic accidents — that were recorded as homicides resulting in a trial for manslaughter at the Old Bailey [White Greenwald and Greenwald 1983, 59 Table 2], [UK Parliamentary Papers 1838]. The passage of two pieces of legislation in 1836 simply re-enforced this trend. The first statute mandated fees for the payment of medical witnesses, making inquests into complex cases easier to conduct. And the second, the 'Birth and Death Registration Act', required reporting all deaths within five days and prohibited burial in the absence of either a coroner's inquest, or registrar's certificate [Greenwald and White Greenwald 1981]. These pieces of legislation made it more likely that an inquest would be held, with a commensurate upward pressure on coroner's verdicts resulting in a charge of manslaughter.

32

Finally, and more difficult to tie down, there was, in these decades, a rising intolerance of violence of all kinds, including domestic violence and pub brawls of the sort frequently prosecuted as 'manslaughter' [Wiener 2004]. It is important to note, however, that none of these explanations fully account for the dramatic rise in manslaughter cases at the Old Bailey, and their role in the subcategories of 'manslaughter' identified here is at best tangential.

33

Conclusion

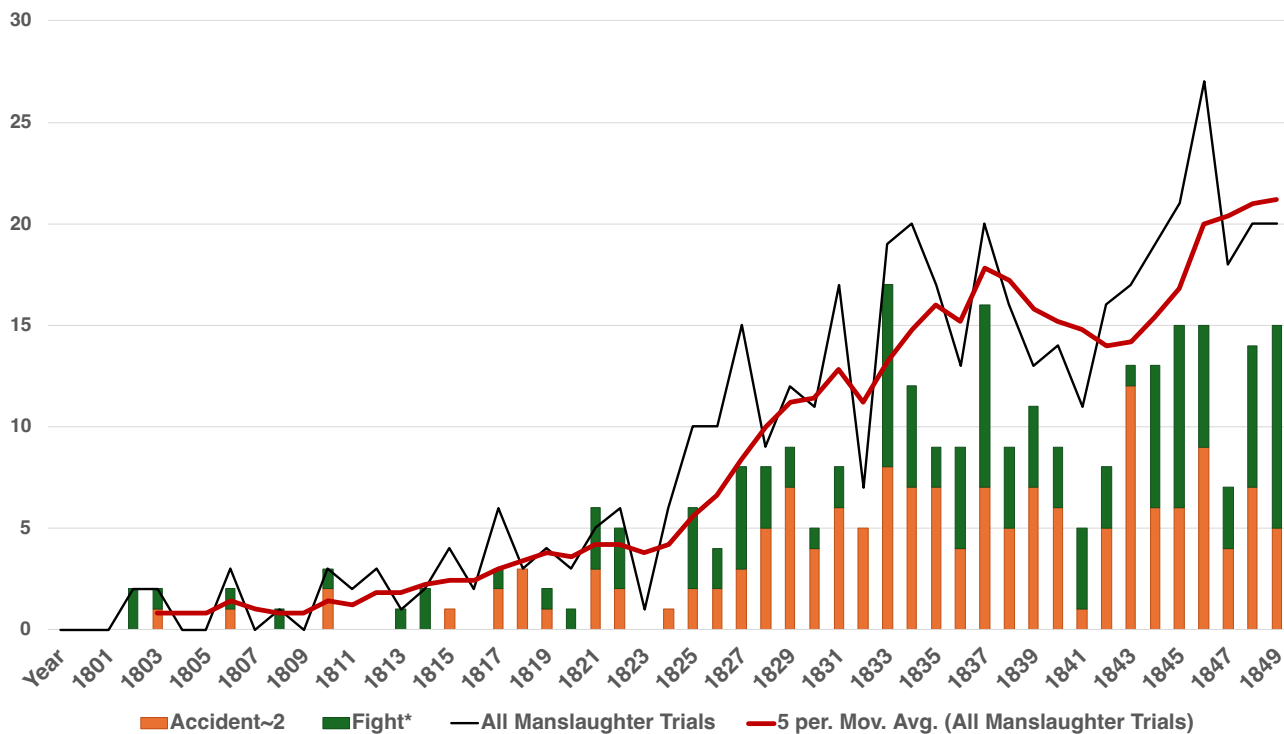


Figure 3. Manslaughter Prosecutions, 1800-1850. Keywords: 'Accident' and 'Fight' (Counting by 'Trial').

The unexplained rise in manslaughter prosecutions remains a significant historical problem. The methodologies outlined above have allowed us to identify the nature of the changes effecting the charge of manslaughter more clearly; to separate out the different sorts of trials that made up the whole. In part, what we have described reflects the unique working practices of two historians and digital humanists, who recognise the importance of *both* applying new methodologies to historical data, and detailed, domain specific expert engagement with primary material. The process described here does not provide a simple model that we expect others to follow; but instead argues for the need to more fully incorporate top-down, human engagement – sense making – into the research journeys at the core of the Digital Humanities. It is an attempt to make more transparent and explicit the dialogue between the historical context, in this instance the legal and administrative process that led to each trial, and the methodologies used to explore the data, and to create from that dialogue a new historical analysis. And if we do not yet have a comprehensive understanding of the forces that contributed to the evolution of the system of British criminal justice and the rise of manslaughter; we do have a clear set of working practises that provides a roadmap in the right direction.

Appendix

To retrieve all the Manslaughter trials from the 1830s: https://www.oldbaileyonline.org/search/crime?month_gte=1&month_lte=12&offence=manslaughter&year_gte=1830&year_lte=1839#results

Table of Annotations

These trials were selected for annotation from the 150 manslaughter trials heard at the Old Bailey in the 1830s on the basis of length (approximately over 250 words) and their chronological distribution over the decade.

"t18300218-65", watchhouse cell fight, involving two men, with hospital surgeon giving evidence. Defendant does not give evidence. NG.

"t18300708-65", male lodger kills housekeeper's 4-year-old daughter by giving her a large glass of rum. Apothecary gives evidence. Defendant gives evidence briefly. NG.

"t18300708-71", two men, skittle ground, drunken argument, - a surgeon gives evidence. Defendant does not give evidence. NG.

"t18310908-226", Coronation Day fireworks, 10-year-old boy knocked down; accidental discharge of a gun held by a man, leads to his death. Surgeon gives evidence. Defendant does not give evidence. NG.

"t18311020-150", Man throws mattress & bed clothes out window – along with his infant female child – domestic violence, some drink. A surgeon gives evidence. Defendant does not give evidence. Guilty – Transportation for life.

"t18311201-125", Fight among ostler and coachman, in a stable yard. Surgeon gives evidence. Defendant does not give evidence. Guilty. Fined 1s.

"t18320216-114", Female customer kills a pub landlady, by pulling her down. Surgeon gives evidence. Defendant gives evidence. Guilty – 1 year confinement.

"t18330214-131", Death by drinking brandy at a public house – testing whether he was coaxed to do so. The defendant appears to be the landlord but does not give evidence. Medical student gives evidence. NG.

"t18330411-209", An apparent mix up in prescriptions leads to a woman being given a large dose of prussic acid, which kills her. A medical malpractice case made up primarily of medical witnesses? The defendant does not give evidence. NG.

"t18330905-126", Mother accused of killing her infant child through neglect/starvation. Primarily medical evidence. The defendant does not give evidence. NG.

"t18340220-80", Man forced to drink large amounts of alcohol, on a boat, later dies. Defendant does not give evidence. No medical evidence. NG.

"t18340410-187", Drunken pub altercation. Two men. Defendant does not give evidence. No medical evidence. Guilty – 3 months

"t18340410-189", Woman kills a female child – issues of neglect, bruising, and an underlying medical condition. A lot of medical evidence. The defendant does not give evidence. NG.

"t18340904-160", Mother accused of scalding her child, who later dies. A lot of medical evidence. The defendant does not give evidence. NG.

"t18350706-1686", A rather formal fight with 'seconds'. Two male defendants, who do not give evidence. Some medical evidence. 'Acquitted' tagged as NG.

"t18350817-1883", Domestic violence, wife dies following a fight, with alcohol abuse involved. A fair amount of medical evidence. The defendant does not give evidence. Guilty – 1 year.

"t18350921-1938", Pub fight involving two men. Starts with medical evidence. Defendant doesn't give evidence. NG.

"t18351123-150", Workhouse death where a female nurse (probably ward nurse), assaults a 77-year-old man. He dies several days later. A lot of medical evidence. The defendant does not give evidence. NG.

"t18360229-723", Widowed mother of two boys, kills the younger one with a poker blow to the head. Older child the main witness (v. affecting). A lot of medical evidence. Guilty. Fined 1s.

"t18360919-2160", Husband accused, following his wife falling into a fit and dying.

Some medical evidence. Defendant does not give evidence. NG.
"t18361024-2373", Two very drunk men, in a pub fight. Strong medical evidence. Defendant does not give evidence. Guilty, Fined 1s.
"t18361128-208", A fight on a boat in harbour between two men; also involves striking a dog and the body ending in the river. No medical evidence, but a lot on how the main blow was struck (i.e. which hand etc). The defendant does not give evidence. Guilty – Confined one week.
"t18300415-168", Traffic accident involving two very posh coaches and some drunkenness. Defendant and victim both male. No medical evidence. No Defendant statement. NG.
"t18300527-112", Traffic accident – with lots of witnesses; quite long. Defendant and victim both male. 7 lines of medical testimony. No defendant statement. NG.
"t18300916-252", Traffic accident, involving a lot of drink and some malice. Numerous witness statements, and some medical testimony. Defendant and victim both male. No defendant statement. Widow asked that the body not be opened 'if it could be avoided'. NG.
"t18301028-176", A traffic accident, with a doctor on the scene almost immediately. This is a long account with numerous witness statements, including several from doctors, etc. Defendant and victim both male. The defendant does not give evidence. NG.
"t18310217-104", A VERY long medical malpractice case involving a post treatment infection leading to death. All the evidence is essentially medical in character. The victim in female, defendant male. There is a long defendant statement that claims there is a professional rivalry, and a good dozen character witnesses. See https://en.wikipedia.org/wiki/John_St._John_Long NG.
"t18310908-219", A long, detailed traffic accident case – good on the rules of the road. Defendant and victim both male. Minimal medical testimony; no defendant's testimony. NG.
"t18311201-161", Traffic accident. A short account, female victim, male cab driver. No medical testimony but includes brief defendant's statement. NG.
"t18320906-104", Traffic accident involving a painter up a ladder (knocked over by a dray). Long, with numerous witness statements, and a short bit of medical evidence. Male defendant and victim. The defendant makes a short statement. NG.
"t18321129-53", A traffic accident involving an elite female victim. Long, with numerous witness statements, and some medical testimony. Male defendant does not make a statement. NG.
"t18330411-184", Long traffic accident case, involving two omnibuses and a road worker. Both defendant and victim were male. Includes a defendant statement but no medical evidence. Does mention deodands. Guilty 3 months.
"t18330704-129", Traffic accident, curds and whey seller, female, killed by male cart driver. Numerous witnesses, but not long. Some medical evidence. NG.
"t18330905-198", Traffic accident involving a 6-year-old girl, and a male omnibus driver. Long, with numerous witness statements; including an extended defendant statement. Short medical testimony. NG.
"t18340220-48", Traffic accident involving a cab – both victim and defendant are

male. A long case with numerous witnesses, and an extended defendant's statement. Guilty 1 year.
"t18340904-157", Traffic accident involving a male cab driver, and a 6-year-old female victim. Long, with numerous witness statements, and a short bit of medical testimony. Defendant gives brief statement. Guilty 9 months.
"t18340904-158", Traffic accident, medium length, male victim and defendant. No defendant statement, some medical testimony. Guilty 3 months in Newgate.
"t18341124-164", Traffic accident involving an overloaded cart. Both victim and defendant are male. The case is a long one with numerous witnesses, including a brief bit of medical testimony, and a very short defendant statement. Guilty – 1 month.
"t18341124-191", Traffic accident, male defendant, male victim. Quite long with 5 witnesses. No medical evidence or defendant statement. Guilty
"t18350615-1520", - a fight between three men, in a domestic context involving a poker and a sword – mistaken initially as a 'domestic'. A very long trial, with a lot of medical evidence, including the claim that an amputation shortened the victim's life – and is mentioned by the jury at the end. Guilty x2, Judgement Respited.
"t18350817-1910", Traffic accident – a couple crossing the road, hit by a cab, killing the husband. Reasonably long, but relatively few witnesses. No Defendant's statement, and one medical witness. Guilty, 2 months. 6th Jury, Common Sergeant.
"t18350921-2013", Traffic accident involving a cab – male defendant, elderly male victim. Long, including numerous witnesses, the defendant's statement, and some medical evidence. Guilty, 6 months. 3rd jury, Recorder.
"t18351026-2288", Traffic accident, involving a cart. Male defendant and victim. Long, with numerous witnesses, some medical evidence and a defendant's statement. Not Guilty. 2nd Jury, Recorder.
"t18351123-131", Shipping accident on the Thames at Deptford. Male victim, male defendant. Long detailed statement from the second mate – but no other witnesses. Medium length. No medical testimony, or Defendant's statement. NG.
"t18300527-161", Very short, Male defendant accused of killing his wife (I assume). Dismissed after limited medical evidence. NG.
"t18300708-134", A short trial, revolving around a blow made by a potboy to a porter in a pub. Female pub landlady only witness, with evidence of remorse etc. Victim died 8 weeks after incident. No medical evidence. NG.
"t18310217-208", Very short, Male defendant accused of killing his wife (I assume). Dismissed after limited medical evidence. NG.
"t18311020-105", A stub – 'the offence having been committed in 'Kent, the witnesses were not examined'. NG.
"t18320906-103", Very short. A husband and wife appear to be charged with killing their child. A 'bad' inquisition leads to acquittal. NG.
"t18320906-262", Very short, Male defendant accused of killing his wife (I assume). Dismissed after limited medical evidence. NG.
"t18321129-156", Very short, Male defendant accused of killing a woman with a different last name. Dismissed after limited medical evidence. NG.

"t18341205-338", Thirteen words – male defendant, male victim, no evidence. NG

"t18300916-152", Medium length trial, of a soldier who kicks a young girl who had been playing on some 'chains'. No medical evidence, or statement from the defendant. Issues of identifying the girl and soldier. NG.

"t18300916-154", Medium to long trial, male defendant and victim. Death occurs a while after the def. shoves the victim, who falls over and hurts his hip – later dying of dysentery. Lots of medical evidence, no defendant statement. NG.

"t18301028-112", see above t18310217-104. Another VERY long medical malpractice case involving the same defendant, and a post treatment infection leading to death. All the evidence is essentially medical in character. The victim in female, defendant male. There is a long defendant statement that claims there is a professional rivalry, and a good dozen character witnesses.

https://en.wikipedia.org/wiki/John_St._John_Long

"t18310106-88", Medium length trial, involving an elderly woman being frightened, bricks through windows, and white sheets. She dies three months later. Defendant male, victim female, some medical testimony. NG.

"t18310630-65", Medium length, involving a street fight between two men. There is some medical evidence. Guilty. 6 months.

"t18320105-213", A fight between a building labourer and an elderly man, in a small court. Involves a large crowd, cry of murder and about four weeks between assault and the victim's death. Guilty. 12 months.

"t18320517-119", Two women assault a third woman, who dies the next morning. There is some question of whether a true bill had been found. After initial witness statements most of this long trial is given over to medical testimony from half a dozen doctors. NG.

"t18320705-112", A female neighbour kills a 6-month-old boy, by dropping a box, that knocks the child from its grandmothers arms. The child dies immediately. A lot of witnesses, medium length. A short defendant statement. Guilty. 1 month.

"t18320906-102", Two sisters in law – one assaulting the other. Lots of witnesses. Some medical evidence determining 'natural causes'. NG.

"t18320906-161", A very long trial in which one of twelve cannons ordered to celebrate the Reform Act, blows up on firing, killing someone. Defendant is the cannon maker. Lots of technical evidence on making cannons, from a lot of witnesses. No medical evidence or defendant's statement. Guilty. 10 days.

"t18320906-257", Two youngish boys charged with the death of a third. Takes place in a paper manufacturing, and involves a steam engine (the victim is exposed to the vented steam). A long, case with lots of witnesses. Some medical evidence suggesting steam could cause 'excitement' of the brain. The victim dies a week after the main event. Not guilty.

"t18321129-154", Medium length. A late-night drunken brawl involving around five men. Both defendant and victim are men. A fair number of witnesses, including two doctors. Guilty 2 months.

"t18321129-155", Medical malpractice, with a female medical practitioner contributing to the death of a 3-year-old girl, with a scorbutic disease. Revolves around the use of a plaster. Lots of medical evidence, and long statements from the

parents. Written statement from the defendant. NG.
"t18330103-177", Following a christening, the parents are on a pub crawl, are thrown out by the bartender, and the child is hit by the door and dies. Quite short, with strong medical evidence. NG.
"t18340703-102", Medium to long trial. Two couples get into a fight on the street; one man is killed, and his wife badly beaten. Both the men and women involved, but the defendant and victim are both male. Lots of witnesses, some medical testimony. Guilty. 3 months.
"t18340703-103", A street fight, male defendant and victim. Defendant claims his mother was insulted, and he struck out. Involves drink, honour and random violence. Some medical testimony. Longish trial. NG.
"t18340703-145", A drunken fight outside a pub on payday. Reasonably long, with both male and female witnesses. No medical evidence. NG.
"t18350615-1464", Male defendant and victim. A fight on shipboard, as the ship is being brought into a dock. A trip, an open-handed blow and a knife. Longish, with quite a few witnesses, and some medical testimony. Guilty 1 Year.
"t18350706-1682", A man and woman, an unmarried couple, get in a fight with a soldier outside a pub, leading to the soldier's death. Longish, with long statements. Some medical evidence. NG.

Table 1.

Notes

[1] All our code has been implemented in *Mathematica / Wolfram Language*. Implementations of some of the techniques described below (like RAKE, TF-IDF, and KWIC) are provided in [Turler 2020] with open source, open content, and open access licensing. The semantic interaction techniques will be covered in a companion volume to be released in late 2025 or early 2026.

Works Cited

- Coldiron 1950** Coldiron, W.H. (1950) "Historical development of manslaughter," *Kentucky Law Journal* 38, no. 4 (1950): 527–50. Available at: <https://uknowledge.uky.edu/klj/vol38/iss4/2/>.
- Devereaux 2023** Devereaux, S. (2023) *Execution, state and society in England, 1660-1900*. Cambridge, England: CUP.
- Emsley 2010** Emsley, C. (2010) *Crime and society in England, 1750-1900*, 4th ed. Harlow, England: Longman.
- Endert, Bradel, and North 2013** Endert, A., Bradel, L., and North, C. (2013) "Beyond control panels: Direct manipulation for visual analytics," *IEEE Computer Graphics and Applications*, 6-13.
- Endert, Fiaux, and North 2012** Endert, A., Fiaux, P., and North, C. (2012) "Semantic interaction for visual text analytics," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 18(12), pp. 2879-2888.
- Ferner and McDowell 2006** Ferner, R.E., and McDowell, S.E. (2006) "Doctors charged with manslaughter in the course of medical practice, 1795-2005: A literature review," *Journal of the Royal Society of Medicine* 99, no. 6, pp. 309–14.
- Foyster 2005** Foyster, E. (2005) *Marital violence: An English family history, 1660-1857*. Cambridge.
- Green, Feinerer, and Burman 2013** Green, C.D., Feinerer, I., and Burman, J.T. (2013) "Beyond the schools of psychology 1: A digital analysis of psychological review, 1894-1903," *Journal of the History of the Behavioral Sciences*, Vol. 49(2), pp. 167-189.
- Greenwald and White Greenwald 1981** Greenwald, G.I. and White Greenwald, M. (1981) "Medicolegal progress in inquests of felonious deaths: Westminster, 1761-1866," *The Journal of Legal Medicine*, 2, no. 2, pp. 193-264. Available at: <https://www.tandfonline.com/doi/abs/10.1080/01947648109513328>.
- Guldi 2023** Guldi, J. (2023) *The dangerous art of text mining: A methodology for digital history*. Cambridge: Cambridge University Press.

- Handler 2007** Handler, P. (2007) "The law of felonious assault in England, 1803-61," *Journal of Legal History* 28, no. 2, pp. 183–206.
- Hitchcock and Turkel 2021a** Hitchcock, T. and Turkel, W.J. (2021a) "The Old Bailey proceedings, 1674-1913: Text mining for evidence of court behavior." Annotated article, *Models of Argument-Driven Digital History*. Available at: <https://doi.org/10.31835/ma.2021.09>.
- Hitchcock and Turkel 2021b** Hitchcock, T. and Turkel, W.J. (2021b) "Studying the historical emergence of manslaughter in English law using stable random projections and tag parameter spaces," *Association for Computing in the Humanities*.
- Manning et al. 2008** Manning, C.D., Raghavan, P., and Schütze, H. (2008) *Introduction to information retrieval*. Cambridge: Cambridge University Press.
- Paul et al. 2018** Paul, C.L., Chang, J., Endert, A., Cramer, N., Gillen, D., Hampton, S., Burtner, R., Perko, R., and Cook, K.A. (2018) "TexTonic: Interactive visualization for exploration and discovery of very large text collections", *Information Visualization*. Available at: <https://doi.org/10.1177/1473871618785390>.
- Pirolli and Card 2005** Pirolli, P. and Card, S. (2005) "The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis", *Proceedings of International Conference on Intelligence Analysis*,(5).
- Rose et. al 2010** Rose, S., Engel, D., Cramer, N., and Cowley, W. (2010) "Automatic keyword extraction from individual documents", in Michael W. Berry and Jacob Kogan (eds.) *Text mining: Applications and theory*, pp. 3-20. New York: John Wiley and Sons.
- Schmidt 2018** Schmidt, B. (2018) "Stable Random Projection: Lightweight, General-Purpose Dimensionality Reduction for Digitized Libraries", *Journal of Cultural Analytics*. Available at: <https://culturalanalytics.org/article/11033-stable-random-projection-lightweight-general-purpose-dimensionality-reduction-for-digitized-libraries>.
- Sharpe and Dickinson 2016** Sharpe, J.A. and Dickinson, J.R. (2016) "Revisiting the 'violence we have lost': Homicide in seventeenth-century Cheshire", *The English Historical Review* 131, no. 549, pp. 293–323. <https://doi.org/10.1093/ehr/cew122>
- Smith 1999** Smith, G.T. (1999) "The state and the culture of violence in London, 1760-1840" (University of Toronto, PhD thesis.) Available at: <https://tspace.library.utoronto.ca/handle/1807/13127>.
- Turkel 2011** Turkel, W.J. (2011) "Term weighting with TF-IDF". Wolfram Demonstrations Project. Available at: <https://demonstrations.wolfram.com/TermWeightingWithTFIDF/>.
- Turkel 2020** Turkel, W.J. (2020) *Digital research methods with mathematica*, 2nd revised edition. <https://williamjturkel.net/digital-research-methods-with-mathematica/>
- Turkel 2025** Turkel, W.J. (2025) "Incorporating the sensemaking loop into bespoke tools for digital history", *Historia y Grafía*, 64, "Historia digital: en la frontera del sur y norte global". Available at: <https://www.revistahistoriaygrafia.com.mx/index.php/HyG/article/view/543>.
- UK Parliamentary Papers 1838** UK Parliamentary Papers (1838) "Returns of the coroners inquests and of the magistrates inquests or inquiries for murders and homicides committed in England and Ireland from 1st January 1827 to 1st December 1837".
- UK Parliamentary Papers 1850** UK Parliamentary Papers (1850) Tables of the Number of Criminal Offenders in England and Wales, 1850'.
- White Greenwald and Greenwald 1983** White Greenwald, M. and Greenwald, G.I. (1983) "Coroners' inquests: A source of vital statistics: Westminster, 1761-1866," *Journal of Legal Medicine*, 4, no. 1, pp. 51-86. Available at: <https://www.tandfonline.com/doi/abs/10.1080/01947648309513373>.
- Wiener 1999** Wiener, M.J. (1999) "Judges v. jurors: Courtroom tensions in murder trials and the law of criminal responsibility in nineteenth-century England", *Law and History Review* 17, no. 3, pp. 467–506. Available at: <https://doi.org/10.2307/744379>.
- Wiener 2004** Wiener, M.J. (2004) *Men of blood: Violence, manliness, and criminal justice in victorian England*. Cambridge: Cambridge University Press.
- Yi et al. 2005** Yi, J.S., Melton, R., Stasko, J., and Jacko, J.A. (2005) "Dust & magnet: Multivariate information visualization using a magnet metaphor", *Information Visualization*, pp. 1-18.

Recommendations

DHQ is testing out three new article recommendation methods! Please explore the links below to find articles that are related in different ways to the one you just read. We are interested in how these methods work for readers—if you would like to share feedback with us, please complete our short evaluation survey. You can also visit our documentation for these recommendation methods to learn more.

SPECTER Recommendations

Below are article recommendations generated by the SPECTER model:

1. Interpreting Measures of Meaning: Introducing Salience Differentiated Stability, 2022, Hugo Dirk Hogenbirk, University of Groningen, Departement of the History of Philosophy; Wim Mol
2. Recognition and Analysis of the Proceedings of the Greek Parliament after WWII, 2024, Epameinondas-Konstantinos Barmounis, Athens University of Economics and Business; John Pavlopoulos, Athens University of Economics and Business; Panos Louridas, Athens University of Economics and Business; Dritsa Konstantina, Athens University of Economics and Business
3. Criminocorpus. Un projet numérique pour l'histoire de la justice, 2018, Marc Renneville, CLAMOR. Center for Digital Humanities and History of Justice (CNRS); Jean-Lucien Sanchez, CLAMOR. Center for Digital Humanities and History of Justice (CNRS); Sophie Victorien, CLAMOR. Center for Digital Humanities and History of Justice (CNRS)
4. From Semi-structured Text to Tangible Categories: Analysing and annotating death lists in 18th century newspaper issues, 2023, Claudia Resch, Austrian Academy of Sciences; Nina C. Rastinger, Austrian Academy of Sciences; Thomas Kirchmair, Austrian Academy of Sciences
5. Category Development at the Interface of Interpretive Pragmalinguistic Annotation and Machine Learning: Annotation, detection and classification of linguistic routines of discourse referencing in political debates, 2023, Michael Bender, Technical University of Darmstadt; Maria Becker, University of Heidelberg; Carina Kienes, Technical University of Darmstadt; Marcus Müller, Technical University of Darmstadt

DHQ Keyword Recommendations

Below are article recommendations generated by DHQ Keywords:

1. Matching Computational Analysis and Human Experience: Performative Arts and the Digital Humanities, 2020, Jan-Hendrik Bakels, Freie Universität Berlin; Matthias Grotkopp, Freie Universität Berlin; Thomas Scherer, Freie Universität Berlin; Jasper Stratil, Freie Universität Berlin
2. How Technology Means: Texts, History, and Their Associated Technologies, 2018, Joshua L. Mann, Expositus
3. Reflections on the Ethics of Research with the Registers of Liberated Africans in the Indian Ocean, 2025, Matthew S. Hopper, California Polytechnic State University
4. Predicting the Past, 2018, Tobias Blanke, King's College London, Department of Digital Humanities
5. The Radical Historicity of Everything: Exploring Shakespearean Identity with Web 2.0, 2009, Katheryn Giglio, University of Central Florida (English Department); John Venecek, University of Central Florida Libraries

TF-IDF Recommendations

Below are article recommendations generated by the TF-IDF Model:

1. Mining for the Meanings of a Murder: The Impact of OCR Quality on the Use of Digitized Historical Newspapers, 2014, Carolyn Strange, Australian National University; Daniel McNamara; Josh Wodak, Australian National University; Ian Wood, Research School of Computer Science, Australian National University
2. Prison Writer as Witness: Can DH Read for Social Justice?, 2021, Doran Larson, Hamilton College
3. Sight and Sound: Counter-mapping the U.S.-Mexico Border Crisis, 2022, Lacey Schauwecker, University of Southern California
4. Naming Slavery in a Digital Public History Project in Mali in the Context of Increased Violence Against Those Who Refuse to Be Called "Slaves", 2025, Marie Rodet, SOAS, University of London; Mamadou Sène Cissé, Association Donkosira, Mali
5. Missed Connections: The Collective Novel and the Metropolis, 2011, J.J. Butts, Assistant Professor of English, Wartburg College



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.