

The Empire of the Old Bailey Online:  
Why Zero Matters

by

Matthew Alleyne Dodd

A thesis submitted to the Faculty of Graduate and Postdoctoral  
Affairs in partial fulfillment of the requirements for the degree of

Master of Arts

in

History with Specialization in Digital Humanities

Carleton University  
Ottawa, Ontario

© 2018, Matthew Alleyne Dodd

## **Abstract**

Can the methods of digital, quantitative analysis today be made to communicate with earlier eras of quantitative history? This thesis isolates one database – the Old Bailey Online, a massive collection of published proceedings from one of London, England’s busiest court houses – and tests ways in which I can, and cannot, analyze its data to make a meaningful comparison with the quantitative analysis the legal historian John Beattie performed in the 1980s on records pertaining to Surrey and Sussex. In this thesis I am concerned with what we learn from this process of interrogating two different data sets and quantitative methodologies. With certain caveats, I find that a quantitative approach to the Old Bailey records does not generate findings for London that are significantly different than Beattie’s for Surrey and Sussex. Even if my current results are to accept the null hypothesis, the importance of “zero” in this case becomes that we now know where *not* to focus our research – not on looking for statistical difference in crime between these two areas in this period and, instead, perhaps focusing on qualitative data regarding the people who experienced and had ideas about crime in this historical context.

## **Acknowledgements**

I would like to express my deepest gratitude to Dr. Danielle Kinsey and Dr. Shawn Graham for their extraordinary support and supervision throughout this thesis process. Without your continued guidance I would never have accomplished so much and this thesis would never have explored such a varied corpus of material. I would like to also thank Dr. Kinsey for recommending that I apply for this Master's program; your belief in me and my abilities is truly astounding.

I would also like to acknowledge the incredible knowledge and willingness to share that knowledge of Dr. Kinsey for explaining how cultural histories exist in concert with the local and the individual rather than just the global. I would also like to acknowledge Dr. Graham's willingness to teach and demonstrate so many skills, including statistics and digital techniques, without which this thesis would never have been completed.

I would like to thank my friends and family, but especially my mum, Lisa Dodd, who spent countless hours listening to me, reading material which you had no training in, and supporting such a difficult undertaking without complaint. Your support of my work and willingness to read every paper I have ever submitted for any class is incredible.

Finally, I would like to thank Dr. Ian Milligan and Dr. Tracey Lauriault for engaging with my work and helping to clean up and improve my final document.

## Table of Contents

<b>Abstract.....</b>	<b>ii</b>
<b>Acknowledgements .....</b>	<b>iii</b>
<b>Table of Contents .....</b>	<b>iv</b>
<b>List of Tables .....</b>	<b>vi</b>
<b>List of Illustrations.....</b>	<b>vii</b>
<b>List of Appendices.....</b>	<b>viii</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1    The Old Bailey Online.....	2
1.2    Beattie’s Crime and Courts of England.....	5
1.3    Apples and Oranges and Statistics .....	7
<b>Chapter 2: Methodology.....</b>	<b>10</b>
2.1    Nature of the Data Sources.....	10
2.2    Collecting the Data for the Current Project.....	12
2.3    Caveats and Limitations .....	17
2.4    Statistical Tests.....	21
<b>Chapter 3: Results.....</b>	<b>26</b>
3.1    Categorizations .....	26
3.2    A Quick Word on Data Speaking for Itself.....	27
3.3    What Questions Should We Ask .....	28
<b>Chapter 4: Discussion .....</b>	<b>34</b>
4.1    Tentative Thoughts on Historical Implications .....	34
4.2    Thoughts on Methodological Implications.....	40
<b>Chapter 5: Conclusion.....</b>	<b>45</b>

<b>Appendices.....</b>	<b>48</b>
Appendix A Glossary .....	48
<b>Bibliography .....</b>	<b>50</b>

## List of Tables

Table 1 Summary data table of Beattie versus the Old Bailey Online (for a full breakdown include gender and other demographics see GitHub repo document raw_data_for_analysis.xlsx sheet(s) oldBailey_data and beattie_data.....	17
Table 2 An example of working through the chi-square test of association.....	24
Table 3 Table example of working out Kendall's Tau. ....	25
Table 4 Hypothesis table and summary of data to one significant digit– Highlighted hypothesis denote those with strong associations across all tests.....	33

## **List of Illustrations**

Illustration 1 Count of men charged with violent crimes at the Old Bailey Online ..... 39

## List of Appendices

Appendix A Glossary.....	48
--------------------------	----

## Chapter 1: Introduction

Can the methods of digital, quantitative analysis today be made to communicate with earlier eras of quantitative history? This thesis isolates one database – the Old Bailey Online, a massive collection of published proceedings from one of London, England’s busiest court houses – and tests ways in which I can, and cannot, analyze its data to make a meaningful comparison with the quantitative analysis the legal historian John Beattie performed in the 1980s on records pertaining to Surrey and Sussex. With certain caveats, I find that a quantitative approach to the Old Bailey records does not generate findings for London that are significantly different than Beattie’s for Surrey and Sussex. In terms of quantitative analysis, crime in London in the long eighteenth century looks very similar to how Beattie characterized crime in Surrey and Sussex in the same period. Instead of dwelling on the implications this has for the historiography of crime in Britain, in this thesis I am concerned with what we learn from this process of interrogating two different data sets and quantitative methodologies. In the era in which Beattie wrote, it was unthinkable to allow others open access to one’s raw data and the conventions of historical argumentation did not foreground quantitative methodology so much as the conclusions of that analysis. In contrast, the digital world and my interrogation of the Old Bailey Online allows me to leave a visible trail of breadcrumbs, as it were, or paths into the work that future researchers can follow, reproduce, and further. Even if my current results are to accept the null hypothesis, the importance of “zero” in this case becomes that we now know where *not* to focus our research – not on looking for statistical difference in crime between these two areas in this period and, instead, perhaps focusing on qualitative data regarding the people who experienced and had ideas about crime in

this historical context. Thus, this thesis is also a gesture towards developing norms of open access and reproducible research (which while increasingly common in other fields) continue to be relatively rare for History as an academic discipline.

## 1.1 The Old Bailey Online

The Proceedings of the Old Bailey Online 1674-1913, first available in 2003, is a collection of the published court proceeding documents from cases held before the Central Criminal Court of England and Wales. These mainly referred to cases involving “The City” (an administrative unit thought to comprise the confines of the medieval expanse of London) as well as Greater London (the area outside of “The City” before 1834).<sup>1</sup> The Old Bailey Online digitized these records through double entry transcription: passages where the transcriptions differed were automatically tagged for correction. Passages were then semantically tagged and marked to enable search and study.

The Old Bailey Online has become the go-to digital hub for historical crime research about Britain, it seems. Many studies have used this resource for both quantitative and qualitative work.<sup>2</sup> Searching for citations of this resource gives an indication of the impact of having an easily digitized corpus - a Google Scholar search for the Old Bailey Online returns 574 citations or roughly 38 per year of its existence.<sup>3</sup>

---

<sup>1</sup> Allyson May, *The Bar and the Old Bailey, 1750-1850* (Chapel Hill: University of North Carolina Press, 2003), 7.

<sup>2</sup> see for instance Howard, S. (2015). Bloody Code: Reflecting on a decade of the Old Bailey online and the digital futures of our criminal past. *Proceedings of the National Academies of Science* <http://www.pnas.org/content/pnas/early/2014/06/11/1405984111.full.pdf>; Klingenstein, S., Hitchcock, T., & DeDeo, S. (2014). The civilizing process in London’s Old Bailey. *Proceedings of the National Academy of Sciences*, 201405984.; Hitchcock, T., & Turkel, W. J. (2016). The Old Bailey Proceedings, 1674–1913: text mining for evidence of court behavior. *Law and History Review*, 34(4), 929-955; Gray, D. (2014). Putting undergraduates on trial: using the Old Bailey Online as a teaching and assessment tool. *SOLON Law, Crime and History*, 4(1), pp.104-113

<sup>3</sup> Citation metrics are notoriously difficult to assess. Google Scholar is the least restricted in what it indexes and searches, which suits our purpose here because it is the *penetration* of this resource into broader consciousness that I am concerned with here.

What are the consequences of this centrality? On the one hand, the centrality of the Old Bailey to historical work about crime is understandable – it was a central court. And at first glance, it makes sense to assume that London’s history and, therefore, the history of the Old Bailey is so distinct that it can stand alone, without qualification and contextualization. The question as to whether or not crime as found in the records of the Old Bailey is representative of crime in general in the eighteenth century seems to be almost a non-starter because, at first glance, it seems reasonable to hypothesize that London’s crime statistics would be different than other locations in England, especially the more rural Surrey and Sussex. As the administrative center of a massive transnational empire, it stands to reason that London was the singularly most critical metropole in the United Kingdom. In terms of thinking about the rise of modern governmentality and the surveillance state à la Michel Foucault or Max Weber, one might assume that the effects of government oversight and modern bureaucracy would be particularly obvious in the city that houses Parliament, witnessed unprecedented population growth from the seventeenth century until the twentieth century, and was renowned for its seedy underbelly. That is to say, the records of the Old Bailey Online are particular to the extraordinary conditions of London and might reasonably be thought to be too exceptional to stand in for the broader English context. The early nineteenth-century travel-writer, James Grant, for example, referred to Londoners as a separate “race” from the rest of Britain – if people at the time thought the society and culture of London was so distinct, perhaps then the proceedings of the Old Bailey, like Londoners themselves, might be thought of as in its own standalone category.<sup>4</sup>

---

<sup>4</sup> James Grant, *The Great Metropolis*, Volume 1, (New York: Saunders and Otley, 1837), 19.

On the other hand, are these assumptions sound, even if they echo the assumptions of the time? Was London really that special when it came to crime and to what degree does data from the Old Bailey reflect trends elsewhere in England? Is there a danger that the analysis of Old Bailey Proceedings might become synonymous with crime and justice in British history, just because its data is so readily accessible, and is that analytically justifiable? As Ian Milligan demonstrates in the ‘Illusory order of OCR’,<sup>5</sup> digitized resources can become overused in historical research by virtue of their accessibility and the norms of the discipline which often conceal when a scholar is consulting a digital version. Milligan finds in his studies of Canadian history that citations of *The Globe and Mail* increased ten times post-digitization while the citations of Canadian newspapers *not* digitized declined. This over-representation of *The Globe and Mail* in historical research can distort our understanding of print culture in the past, during which *The Globe and Mail* was only one among several newspapers available. In terms of the Old Bailey, it was one of eleven different courts. It dealt with only those crimes perpetrated within its ambit. It was a place in a particular space, at a particular time. And yet, by virtue of being online, it assumes a weight in the world of scholarship that is perhaps unwarranted. It draws to itself attention and energy and potentially occludes other kinds of research, making the Old Bailey Online its own kind of ‘Empire.’ In order to evaluate the centrality of this resource, and perhaps the centrality of London’s experience of crime within British history, what is required is a *foil*, a point of comparison with another body of scholarship. Unfortunately, at this time there is no

---

<sup>5</sup> Ian Milligan, “Illusory Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010,” *Canadian Historical Review* 94, no. 4 (2013): 540–569, quote from 542.

online resource that works to decentralize London and the Old Bailey in historicizing crime British crime in the eighteenth century. Thus, I had to turn to older historiographies to contextualize the history of the Old Bailey as well as the Old Bailey Online. This is where John Beattie's work on crime in Sussex and Surrey enters.

## 1.2 Beattie's Crime and Courts of England

John M. Beattie's *Crime and the Courts of England 1660-1800*, is not a database but rather a monograph about the counties of Surrey and Sussex in the southeast of England, which abut the greater London area. There is some fuzziness and overlap in the jurisdictions covered. There is an issue of sample size and case representation. Lack of a discussion on sample sizes and representation in Beattie's work makes specificity difficult and precision within statistical analysis less valid. Jurisdictional overlap occurred between Surrey and Sussex and Surrey and Greater London. It was often difficult to know where a case should be tried and what local legal considerations must be made. For the purposes of my data, jurisdictional overlap occurred often in Beattie's work where he makes claims of counts that might be within Greater London, specifically Southwark, which may also be represented in the Old Bailey Online. There is also some fuzziness in the time periods covered, but if I keep in mind the differences between a monograph and a database, and I use the monograph as a guide to *minimal assumptions* (that is, I let the ways Beattie constructed his categories guide how I interrogate the *Old Bailey Online*), I can assess the question 'is London, in terms of criminal court information, truly statistically significantly distinct from the rest of England?' If I know the answer to *that* question, we will have a better sense of whether or not this 'Empire of the Old Bailey' is warranted.

Beattie's work was published in 1986 and used data that Beattie had gathered himself from written records in the archives of Surrey and Sussex. Beattie's work argues that the reason why crime was committed began with resource availability. Specifically, Beattie argues that "above all, poverty pressed with particular urgency [...] especially when the large population of men and women who lived at the margins of security suffered from shortages of work."<sup>6</sup> Poverty, that is a lack of both money and food resources, is the single most significant reason for why crime occurred according to Beattie. Beattie's argument is first and foremost an argument about property rather than justice or security. His secondary argument is that the administration of criminal law was to regulate and demonstrate authority from the wealthy to the poor at all levels of society. Beattie's claim that the administration of law was for the purposes of authority is his most cited argument.

Beattie's *Crime and the Courts in England 1660-1800* is cited over 1200 times according to Google Scholar.<sup>7</sup> It is an interesting parallel to note that the annual rate of citation for *Crime and the Courts in England 1660-1800* is roughly the same as the Old Bailey Online and while only a conjecture one wonders what a study of citation patterns would reveal about historians use of *others'* quantitative work, a kind of washing-of-the-hands-well-glad-someone-has-done-it! The associated authority of citations is one that has been granted to Beattie for decades and questioned only rarely. In a 2009 study, the British Medical Journal found that "Citation is both an impartial scholarly method and a

---

<sup>6</sup> John M. Beattie, *Crime and the Courts in England 1660-1800* (Princeton: Princeton University Press, 1986), 14.

<sup>7</sup> "Citations of Crime and the Courts in England, 1600-1800," Google Scholar, accessed May 15, 2018, [https://scholar.google.ca/scholar?cites=12854185621251261658&as\\_sdt=2005&sciodt=0.5&hl=en](https://scholar.google.ca/scholar?cites=12854185621251261658&as_sdt=2005&sciodt=0.5&hl=en).

powerful form of social communication. Through distortions in its social use that include bias, amplification, and invention, citation can be used to generate information cascades resulting in unfounded authority of claims.”<sup>8</sup>

### 1.3 Apples and Oranges and Statistics

A monograph is the result of sometimes decades-long engagement with a corpus of material, a long wrestling with the archives. It is a solitary work. The work discusses the sources and uses passages from these to illustrate the broader discussion. As a work of so-called ‘cliometrics’<sup>9</sup>, Beattie’s monograph is not concerned to illustrate *process* so much as the conclusions. Indeed, consistent with the genre of academic monographs, publishers did not and still do not want to publish reams of “raw data” but rather argument and takeaway message. This process is the mark of the monograph, which is understood to be the author’s own work and intellectual property. It defines an author’s expertise but does not necessary show the process of analysis in action. In digital history as it is coming to be practiced today, the *process* is a valuable output in its own right: we call it ‘paradata’.

Digital history has roots in cliometrics, stylometry, GIS (geographic information system), archaeology, and others.<sup>10</sup> Cliometrics is the application of economic and mathematical methods and formula to the study of history, particularly social history. In this form, all conclusions followed an equation and presented a number, or series of numbers, as the final answer to the question. Paradata is the third of the “data triplets”:

---

<sup>8</sup> Steven Greenberg, “How citation distortions create unfounded authority: analysis of a citation network,” *British Medical Journal* 339 (March 2009): 1. <https://doi:10.1136/bmj.b2680>.

<sup>9</sup> A Greif, “Cliometrics after 40 years,” *The American Economic Review* 87 (2) (1997): 400-403.

<sup>10</sup> Graham, S., Milligan, I., & Weingart, S. (2015). *Exploring big historical data: The historian's microscope*. World Scientific Publishing Company.

data (the raw material), metadata, (the description of the raw data), and paradata (the description of the collection process of the data). Had Beattie provided the paradata document, the option of reproducing his work would have been feasible. Thus, for me to use Beattie as a foil in the comparisons that follow (see chapter two, Methods) and put my work in conversation with the broader digital history community, I need to make my work open and transparent, providing sufficient breadcrumbs so that the work is *replicable*. Chapter two is my *paradata*, my methods. Comparing the *Old Bailey Online* to *Crime and the Courts of England* is an apples-to-oranges type of comparison. In Chapter two I discuss how certain statistical approaches borrowed from archaeology (which often deals with incomplete or seemingly incommensurate data) can enable this comparison.

I formed hypotheses which I tested using different statistical methods in an effort to make two disparate sources communicate. I chose four statistics tests for this process, the first of which was Chi-Squared. This simple test of association between two variables is simple and was powerful enough to begin my exploration of the data. It was also simple enough that any scholar attempting to reproduce or replicate my work would have no issue doing so. In cases where I was using numbers below five I opted to use Guttman's Lambda as my test of association. I did so because, again, it is simple enough to replicate but also powerful enough to determine if a next step approach might yield a significant result. Beyond these two simple tests of association I used both Phi or Cramer's V and Kendall's Tau. In this instance I needed a method which used similar variables/numbers to Chi-Squared to determine the strength of the association between

the variables. Like using Guttman's Lambda for small values in Chi-Squared, I employed Kendall's Tau to test for the strength of a relationship where small values were possible.

All of the data consumed, manipulated, and analyzed in my study is freely available in all of its forms on the GitHub repository set up for my research. A subsequent scholar can take a copy of all of my data and see what choices I made with it. It is archived with the Zenodo data archiving service in the state I completed as of August 2018.

In Chapter Three I present the results of this analysis, in Chapter Four I discuss what we should take away from this experience, having found that London is not all that different from Surrey and Sussex, given the confines of this experiment.

## **Chapter 2: Methodology**

This thesis attempts to grapple with material developed in two different eras in history. The data comes from two sources, Beattie's *Crime and the Courts of England, 1660-1800* and the Proceedings of the Old Bailey Online. The goal of this thesis is to use the work of Beattie as a foil for the Proceedings of the Old Bailey Online. In this chapter I will discuss the nature of the data, the collection of the data, sample sizes and representiveness, rationale for the testing methods, and provide the steps which I undertook as part of this research. This chapter explains my decision-making process and my choices in tools and methods for attempting to compare and work with two different data sets. This research is focused on the quantitative aspect of criminal court cases rather than the qualitative meanings within the texts.

### **2.1 Nature of the Data Sources**

Beattie counted and tabulated crimes tried at the Surrey and Sussex Assizes and Quarter Sessions between 1660 and 1800; his data are represented in the form of summary tables. Surrey and Sussex are two counties located south of Greater London (the geographic region, not the City of London). Surrey is directly south of Greater London and is a landlocked rural area, with very low population density and small towns. Sussex, however, is directly south of Surrey, and is on the southern coast of England and is more densely populated with a higher population. The geographic borders of both counties remained largely unchanged throughout Beattie's date range, although some

overlap between the two (and Surrey and Greater London) did occur. Beattie's data, therefore, does occasionally contain material from London.

The Old Bailey Online's data takes the form of criminal court text records from London's Central Court (the Old Bailey) which have been digitized through double-entry transcription. The texts were then marked up following the XML (eXtensible Mark Up) conventions of the Text Encoding Initiative.<sup>11</sup> This means that rather than parsing tables of keys and values in a database, the texts *themselves* form the database, making them able to be searched according to the *semantic* information they contain: "show me crimes committed by women within the City of London itself over the following time-period". It is key-word search where the machine understands (thanks to the tagging) the *context* of the words on the page. Thus, the semantic information is both *machine readable* and *human readable* enabling both close and distant reading simultaneously.<sup>12</sup> According to the *Old Bailey Online* itself, there are more than 197,000 cases included in their corpus.

The data encompasses the written record of the Old Bailey (London's Central Court) which was published between 1674 and 1913. This is a larger date range than Beattie's material by about one hundred years more. I use Beattie's range to guide my

---

<sup>11</sup> <http://www.tei-c.org/> "The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics."

<sup>12</sup> M. H. Beals, a historian of migration at Loughborough University in the UK, shares a variety of resources with her students on getting started with the basics of TEI for historical research at <http://mhbeals.com/tei-for-close-reading-can-it-work-for-history/>

collection from the *Old Bailey Online*. Material after 1800 was removed from the analysis.

In terms of geographic borders, the Old Bailey Court generally concerned itself with the City of London and less so (but including) Greater London. There are (and were) eleven courts in London, thus the Old Bailey does not encompass, nor define, London, although it is the oldest of the eleven. The *Proceedings of the Old Bailey Online* were chosen for exploration due to their accessibility and the comparative possibilities of material which seemed to closely match Beattie's work in terms of content, topic, and general categories.

## **2.2 Collecting the Data for the Current Project**

I prepared a Github repository as a kind of work space for this project. 'Git' is a piece of software one uses on one's own computer for managing the evolution of a software project. 'Github' is an online centralized storage for the folders or 'repositories' kept under version control using Git. By making these folders available online, multiple individuals can contribute to the development of a project and can observe the evolution of my project. What's more, another researcher could take a 'fork' or copy of my research and continue the development in new directions. Github keeps track of these 'forks' and 'branches' thus providing a new kind of citation trail enabling the history of development to be visualized and tracked.<sup>13</sup>

The entirety of my research process is contained within my Github repository, "The Empire of the Old Bailey v.0.1.1," at

---

<sup>13</sup> see van Strien, Daniel. 2016 An Introduction to Version Control Using Github Desktop, *The Programming Historian* <https://programminghistorian.org/en/lessons/getting-started-with-github-desktop>.

<https://github.com/matthewdodd/matthewdodd.github.io>. That repository contains not only the experiments and results that are discussed more fully below, but also the dead ends and experiments that failed. This is done in deliberate contrast to the dictates of the ‘monograph’ form followed by Beattie, wherein only the final compiled data are shared. It is my contribution to what the historian W. Caleb McDaniel calls the promise of ‘Open Notebook History’.<sup>14</sup> As an aside, this openness or willingness to share process and blind alleys marks the difference between digital history and digitizing history. In so many forms we now see the claim of being “digital history” is actually merely digitizing history: websites, apps, games, mapping, and scanning. This was not the digital history that I wanted to do. Mine is a different digital history and it is in dialogue with the approaches employed by William Turkel, a professor of digital and computational history, in asking non-quantitative questions and asking questions that are not of counts, but of context and text such as sentiment analysis.<sup>15</sup> It also speaks to the work of Tim Hitchcock, one of the principal developers of the *Old Bailey Online* and a professor of digital and cultural history, in asking quantitative questions about counts of data and the meanings of those counts.

In what follows, when I refer to various file names, these may be found in my repository.<sup>16</sup> Beattie’s data were manually transcribed from his book and the tables he presents. Each table, which contains material which can be compared has been copied

---

<sup>14</sup> McDaniel, W. Caleb. 2013 Open Notebook History <http://wcm1.web.rice.edu/open-notebook-history.html>

<sup>15</sup> William Turkel, “Data Mining with Criminal Intent,” *Criminal Intent*, last accessed September 16, 2018, <http://criminalintent.org/wp-content/uploads/2011/09/Data-Mining-with-Criminal-Intent-Final1.pdf>.

<sup>16</sup> Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

into the document raw\_data\_for\_analysis.xlsx. Each table in Beattie was transcribed directly into an Excel spreadsheet document. The only occurrence in which this was not a direct copy was when certain tables followed the same, or a similar structure, such that I could modify them into two tables and I did so for the following:

- 9.1 (Capital Punishments in Surrey),
- 9.3 (Noncapital Punishment in Surrey),
- 9.5 (Capital Punishments for Property in Surrey),
- 9.6 (Noncapital Punishments for Property in Surrey),
- 9.7 (Pardons and Capital Punishment in Surrey),
- 10.1 (Capital Punishment in Surrey),
- 10.3 (Noncapital Punishments for Property Offenses in Surrey),
- 10.7 (Noncapital Punishments for Property Offences in Surrey),
- 10.9 (Capital Punishment in Surrey),
- 10.10 (Noncapital Punishments for Property Offenses in Surrey), and
- 10.14 (Noncapital Punishments for Property Offenses in Surrey).

In these cases, I organized by date range and offence (grouped by male, female, or both) since it was not possible to group everything under a single date range banner.

Furthermore, despite Beattie's title using the upper bound of 1800, his data often exceeds that to 1802. While this is not huge, it is nonetheless important to ensure date consistency across data sources.

The *Old Bailey Online*'s data was gathered through a series of downloads meant to capture first the unique case identifiers and then download the XML marked up version of the case. I did this through wget, which is a Command Line tool which can

take an input file containing web URLs and then output the web pages as results. The first input file was called `obapi_wget.txt` and is contained in the folder “wget” in my GitHub repository.<sup>17</sup> The first output files contained the lists of the unique case identifiers. I modified this output into three other files (called `obapi_wget_2.txt`, `obapi_wget_3.txt`, and `obapi_wget_4.txt`) with the Old Bailey API URL (<https://www.oldbaileyonline.org/obapi/text?div=>) prefixed to the unique case identifiers. I reran wget using these three new files and forced the output to be contained under the folder “xml” under “wget”. Thus, I had now downloaded all of the Old Bailey Online’s 197,745 cases.

I determined that for *counting* the cases in the way that Beattie had I would need to structure these files into a single database. While there are many possible solutions, one that is commonly used amongst historians working with data is MySQL.<sup>18</sup> I wrote a ‘structured query language’ (SQL) script to insert every `<interp>` XML tagged line into a table in the MySQL database. The `<interp>` tag contains the information entered by the historians from the Old Bailey Online project from which I can generate comparable categories to Beattie’s work.

The resulting table had three columns, each capable of storing 200 characters. The columns were called `id`, `type`, and `value`. The reason each was made so large was that, while inefficient in terms of storage size, this would enable the data to be inserted without overflow or `data_too_long` errors. Following the data insertion, which took approximately

---

<sup>17</sup> Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

<sup>18</sup> see for instance Blackadar, Jeff. 2018 Getting Started with MySQL and R. *The Programming Historian* <https://programminghistorian.org/en/lessons/getting-started-with-mysql-using-r>

forty minutes, the resulting table was over eleven million rows long. This data collection was then queried through a long SQL select statement to output the table found in raw\_data\_for\_analysis.xlsx under the “oldBailey\_data” sheet.<sup>19</sup> The statement used was the following:

```
select p2.primary_id,
       to_date(substr(p2.primary_id,2,8),'yyyymmdd') as datestamp,
       p2.unique_id,
       p2.first_name,
       p2.last_name,
       p2.age,
       p2.gender,
       p2.occupation,
       cc.breaking_peace,
       cc.damage,
       cc.deception,
       cc.kill,
       cc.misc,
       cc.royal_offences,
       cc.sexual,
       cc.theft,
       cc.violent_theft,
       vc.guilty,
       vc.not_guilty,
       vc.misc_verdict,
       vc.special_verdict,
       pc.corporal,
       pc.death,
       pc.imprison,
       pc.misc_punish,
       pc.no_punish,
       pc.transport
from dodd.person_2 p2
join dodd.crimecount cc on cc.primary_id = p2.primary_id
join dodd.verdictcount vc on vc.primary_id = p2.primary_id
join dodd.punishcount pc on pc.primary_id = p2.primary_id
order by 1;
```

---

<sup>19</sup> Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

While complex, what this statement says in plain English is to list the case identifier, the person involved, the crime that was committed, the finding (guilty/not guilty/miscellaneous verdict/special verdict), and the sentence that was passed down using a zero to represent false and a one to represent true for each of the columns after occupation. This result has 967,340 records for individuals contained within the Old Bailey Online. From here, I summed the records which aligned with Beattie’s tables and added the numbers to my hypothesis tables (an example of which can be read below).

<b>Data Source</b>	<b>Date Range Type</b>	<b>Date Range</b>	<b>Number of Cases</b>
Beattie	Full Range	1660-1802	~8,300
The Old Bailey Online	Full Range	1674-1913	197,745
Beattie	Overlap Range	1674-1802	~7,400
The Old Bailey Online	Overlap Range	1674-1802	34,298

**Table 1 Summary data table of Beattie versus the Old Bailey Online (for a full breakdown include gender and other demographics see GitHub repo document raw\_data\_for\_analysis.xlsx sheet(s) oldBailey\_data and beattie\_data.**

While a table carries a certain rhetorical authority, these numbers are only valid within certain caveats and limitations.

### **2.3 Caveats and Limitations**

The conversion of data from XML files into a database necessitates making transformations or perhaps better, ‘interpretations’, when the data is not a direct match. In converting the data as I have done, I have ignored all of the other textual markers which are contained in the court cases and left behind only those items which were tagged. In effect, I have converted a qualitative written work into a quantitative statistical database. I have attempted make the Old Bailey Online resemble, as far as possible, Beattie’s work.

Beattie's work is my baseline for minimizing my assumptions about the data. This does ignore any other tags within the XML documents, but the <interp> tags seem to encompass the same data as the other tags, including tags such as <place\_name id="t16740429-1-crimeloc2">Hide-Park</place\_name> being identical to <interp id="t16740429-1-crimeloc2" type="placeName" value="Hide-Park">. <sup>20</sup> This different tagging system could be useful in a mapping context, but for the purposes of quantitative research is not necessary.

In terms of sample sizes, it seems like reverse engineering Beattie's work might yield the extent of his sample sizes, but this is not possible because we do not have enough of the context to do so. Beattie himself does not explain his sample size or how many cases or individuals fall under which category, he merely provides a table with final data; no methodology or model is provided. <sup>21</sup> It is almost impossible, therefore, to properly ascertain whether comparative samples can be found, and if they can be found, if they are truly valid. Despite this, I have taken Beattie's material at face value and assumed that what he has found is in fact valid for the area he claims to encompass. I do so for two reasons: the first is that I have no justification for not believing his work the second is that for this kind of exploration to work I must have two sources of relatively comparable data. The obvious limitations in these assumptions eloquently makes the

---

<sup>20</sup> This example is found in the document "t16740429-1.xml" under the folders "wget/xml". Matthew Dodd, "The Empire of the Old Bailey v.0.1.1," last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

<sup>21</sup> John M. Beattie, *Crime and the Courts in England 1660-1800* (Princeton: Princeton University Press, 1986), 14.

argument for more transparency and openness in historians' work, especially when dealing with quantitative data.

The Old Bailey Online, however, is straightforward when it comes to sample size and years: following the sheet I noted above (`raw_data_for_analysis.xlsx` under the "oldBailey\_data") counts are simple enough to find within the confines of dates or other categories.<sup>22</sup> There were 197,745 cases digitised as part of the Old Bailey Online system (although I have 197,751 documents; I have seven documents which are not marked as being a "trialAccount", these are instead marked as being a "punishmentSummary"). Contained within these cases the counts of men, women, and those not identified can be determined, their specific crimes, punishments, occupations, ages, even locations of the crime and their homes are possibly contained. The Old Bailey Online, as it is translated into my documents, contains information primarily from the City of London, but also from greater London and even beyond due to the punishment of Transportation sending individuals outside of England and across the British Empire.

The Old Bailey Online is the oldest of the eleven courts in London, but this does not necessarily make it representative of London, nor even of the cases that occurred in London. What it does mean, however, is that there is access to the largest digitized repository of London criminal court data and that this data can be reasonably assumed to be close in representativeness to the rest of London, especially given the geographic markers in the Old Bailey Online's data. As far as representativeness goes for Beattie: as noted earlier, Beattie's work must be accepted as representative of Surrey and Sussex's

---

<sup>22</sup> Matthew Dodd, "The Empire of the Old Bailey v.0.1.1," last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

Quarter Sessions and Assizes (their court system) for the years that Beattie uses (1660 CE to 1800/1802 CE). Both sources share a similar representativeness in terms of geographic boundaries: neither sources boundaries have changed much since 1660, despite huge population growth, especially in the nineteenth and twentieth centuries. As for categories such as gender, age, and class: these are generally well represented. Beattie makes it a point to discuss the history of the crimes, primarily property offenses, and to discuss how, where, and why the crime would have been punished and who was involved. The Old Bailey Online does not make a point of this discussion, but it does include every record from 1674 CE to 1913 CE which contains those persons of different genders, ages, and classes who had found themselves before the court, either as defendants or as part of the prosecution. Therefore, while there is no explicit justification of sample size or representation, both sources have endeavored to include large amounts of material within the confines of the categories that they have determined.

As noted above, there are issues with Beattie's work in that we do not have the methodology for how he came to his conclusions and all we are able to use are his final tables. This does raise certain issues, including that a proper comparison between Beattie and the Old Bailey Online necessitates certain assumptions being made for the Old Bailey Online, including that Beattie has been honest with his numbers and that what he states as a title and explanation for his tables is valid and within the bounds he sets out. Other issues, however, include the conversion of XML files to a spreadsheet where the counts may not be quite as they are from Beattie, given different assumptions or different

methods for determining values and categories. These, however, deal with issues within the data, there are other issues which need to be addressed.

The categories used by Beattie and the Old Bailey Online are similar, but not the same. Beattie does not employ a category/subcategory system which the Old Bailey Online does and this makes it difficult to ascertain where a delineation should exist; should I use only categories, only subcategories, or some combination of both? In the end, I made the decision to use only the categories from Beattie and the Old Bailey, as I felt that with the subcategories Beattie would not fit as cleanly. Given that this thesis is an attempt to use Beattie as a foil for the Old Bailey Online, using any other method is unfair. Finally, Beattie's categories are often grouped under different dates or different locations, so it is difficult to compare Surrey, Sussex, or Surrey and Sussex to London. To this end, I have made the assumption that when left unspoken the data is for both counties. Very often, however, the data is explicitly only for Surrey, while Sussex is left out of the picture. To minimize issues, I have noted what assumptions I have made throughout my statistical analysis where one must be made.

## **2.4 Statistical Tests**

I employed four basic statistical tests in my research:

- Chi-Squared,
- Phi/Cramér's V,
- Guttman's Lambda, and
- Kendall's Tau.<sup>23</sup>

---

<sup>23</sup> Locke, Gary, *Digging Numbers* - full reference here. Also, Shennan, S. (1997). *Quantifying archaeology*. University of Iowa Press.

These are test for association between variables, both the strength of the association and the direction of that association. I chose these because tests of association are appropriate for categorical data distributions and are used often in other fields (e.g. archaeology) where comparisons between difficult-to-compare bodies of evidence are desired.<sup>24</sup> Chi-Squared is a test of association between two variables. While it is possible in Excel or other data software packages to press a button and obtain a result, I instead constructed tables in Excel that lay out each step in the calculation to arrive at the final Chi-Square value. These tables can be copy and pasted into new spreadsheets to work with new data, or readers may explore my tables and alter the values to see for themselves the effect of having more or less data for given categories.

One situation where the Chi-Square test fails is in cases where a given categorical count is less than five. In such cases I opted to use Guttman's Lambda as my test of association. I did so because, again, it is simple enough to replicate but also powerful enough to determine if a next step approach might yield a significant result. Beyond these two simple tests of association I used both Phi/Cramér's V and Kendall's Tau. While Chi-Square will tell the researcher whether or not an association exists, these two tests enable me to determine whether or not an association is *strong* and in what *direction* the association lies. Like using Guttman's Lambda for small values in Chi-Squared, I employed Kendall's Tau to test for the strength of a relationship where overly small values were present. I organized my work using multiple hypotheses and then forming the particular statistical tests into tables, such as the example below which comes from the document `new_data_worksheet_2.xlsx`.<sup>25</sup>

In this example below, I arranged my data to reflect a hypothesis concerning ‘being male’ and being convicted of a property offence. I put my counts of cases where the individual is male and convicted in the top left cell, and being convicted while being ‘not male’ in the cell below, and so on. This is my observed distribution. The ‘expected’ table calculates from these observations what a normal distribution under these circumstances could be expected to look like. The chi-square statistic then becomes the sum of the squared differences between these two distributions, divided by the expectations. To determine whether or not this statistic is significant at various strengths, one compares the value to a table of pre-computed values. In the example table below, the chi-square statistic is so large, we could say with 99% certainty that there definitely is an association between being male and being convicted of property offences. The other statistics I mentioned above could tell us how *strong* that association is, and which way the *directionality* lies: is it that being male means that one will be convicted? Or is it that if there is a property crime, a male will be sought out for conviction? The document `new_data_worksheet_2.xlsx` in my GitHub repo contains the tables for working out the tests of association for my fourteen hypotheses concerning the data that I lay out in Table 3 in Chapter Three.<sup>26</sup>

---

<sup>24</sup> Locke, Gary, Digging Numbers - full reference here. Also, Shennan, S. (1997). Quantifying archaeology. University of Iowa Press. 115.

<sup>25</sup> Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

<sup>26</sup> `new_data_worksheet_2.xlsx`. Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

OBSERVED		property offence	not property offence		EXPECTED		property offence	not property offence	
male		742655	32423	775078	male		740771	34307	775078
not male		169226	9808	179034	not male		171110	7924	179034
		911881	42231	954112			911881	42231	954112
							x <sup>2</sup> =	576.7	
O	E	o-e	(o-e) <sup>2</sup>	(o-e) <sup>2</sup> /e					
	742655	740771	1884	354787					
	169226	171110	-1884	354787					ACCEPT THE HYPOTHESIS WITH CONFIDENCE GREATER THAN 99.9%
	32423	34307	-1884	354787					
	9808	7924	1884	354787					

**Table 2 An example of working through the chi-square test of association**

For working out the *direction* of an association (once evidence of there actually *being* an association is determined) I employed Kendall's Tau. Below is a table that works out how this statistic is calculated. Such tables as used in this thesis are collected in `new_data_worksheet_2.xlsx` in my GitHub repo.

OBSERVED	property offence	not property offence	
male	742655	32423	775078
not male	169226	9808	179034
	911881	42231	954112
$P = 742655 * 9808 =$	7283960240	$P - Q =$	1797145642
$Q = 32423 * 169226 =$	5486814598	$P + Q + TR =$	38509646511
$TR = 742655 * 32423 +$ $169226 * 9808 =$	25738871673	$P + Q + TC =$	1.38765E+11
$TC = 742655 * 169226 +$ $32423 * 9808 =$	1.25995E+11	$SQRT (P + Q + TC)(P + Q + TR)$	73101321570
	$tau-b = (P - Q) /$	$SQRT (P + Q + TC)(P + Q + TR)$	0.0
Very Weak			

**Table 3 Table example of working out Kendall's Tau.**

In total, working from Beattie's data as a baseline for finding comparable data in the Old Bailey Online, I generated fourteen different hypotheses concerning patterns of crime and punishment. It is important to note that these are not hypotheses that emerge from a pre-existing historical question that I had, but rather cases where I could make these two bodies of scholarship *speak* to one another. We turn now to how these hypotheses were generated, and the results uncovered. All data and statistical tests (including dead ends and experiments) are contained within the following files in the repository; the tables germane to chapter three are contained in the file `new_data_worksheet_2.xlsx` within the GitHub repository used for this research.<sup>27</sup>

<sup>27</sup> `new_data_worksheet_2.xlsx`. Matthew Dodd, "The Empire of the Old Bailey v.0.1.1," last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

## Chapter 3: Results

In this chapter I discuss the results of using Beattie as a foil for exploring the Old Bailey Online. It is worth noting that historians generally come to data with their hypotheses *already formed*; what is novel here is that I am asking: what questions can historians reasonably ask *across different kinds of quantitative studies*. I open with a discussion of categorization and the commensurability of the two data sources. As a result of this discussion, I am able to find fourteen potential historical hypotheses that can be asked of *both* data sources. I present the results of my chosen statistical tests on these fourteen hypothesis. In summary, I find that while I can make these two sources intelligible to one another, the patterns observed in Beattie's work seem to be present in the Old Bailey Online. The urban and the rural seem broadly similar, at least as far as *this* experiment goes.

### 3.1 Categorizations

For Beattie and the Old Bailey Online, the researchers used a series of categorizations for organizing the kinds of crimes or the kinds of people accused of/committing the crimes. The Old Bailey Online offers an explicit list of its categories and their subcategories, but Beattie does not provide *his* list; we deduce it from his tables. For instance, Beattie finds a distinction between 'rape' and 'attempted rape,' while the Old Bailey Online classifies all sexual crimes as just 'sexual crimes.' The Old Bailey Online does, however, sometimes use subcategories for crimes - eg a sexual crime may have been 'rape' or 'sodomy.' Furthermore, Beattie acknowledges ignoring categories,

including assault and riot to instead focus on homicide, rape, and infanticide.<sup>28</sup> The ‘why’ of these categorizations is not always apparent, although in the instance above regarding ignoring categories, Beattie claims this is because these are relatively clear categories and that they are interesting.<sup>29</sup> Beattie claims that there are broad categories under which cases may be grouped including offenses which were particularly serious (treason, homicide, larceny, and arson) and those that were lesser charges (assault, riot, fraud, and embezzlement).<sup>30</sup> For instance, it is generally accepted that a court case has a standard outcome: the defendant(s) are pronounced guilty, not guilty, or some special verdict may be rendered. Issues, however, are already obvious: if one falls under the special verdict category, how can it then be subcategorized? One point of comparison that *is* constant between the two sources is *gender*, and that became the wedge into the problem of making these two representations of data speak to one another.

### 3.2 A Quick Word on Data Speaking for Itself

Use of technology to perform a task is not the same thing as critically examining the use of the technology to perform that task. John Russell and Merinda Hensley have used the term buttonology to describe this phenomenon. “Knowing how to upload texts into a tool like Voyant does not help researchers think about what texts should be uploaded, how selecting data relates to a research question or even what constitutes an effective research question.”<sup>31</sup> By the same token, merely putting data into a database and

---

<sup>28</sup> John M. Beattie, *Crime and the Courts in England 1660-1800* (Princeton: Princeton University Press, 1986), 76.

<sup>29</sup> John M. Beattie, *Crime and the Courts in England 1660-1800* (Princeton: Princeton University Press, 1986), 76.

<sup>30</sup> John M. Beattie, *Crime and the Courts in England 1660-1800* (Princeton: Princeton University Press, 1986), 5.

<sup>31</sup> Voyant Tools is an online text mining and visualization tool developed by Stefan Sinclair of McGill University and Jeffrey Rockwell of the University of Alberta. The tool is located at <https://voyant->

then running visualization tools against that data is not sufficient. ‘Data’ is Latin for *things given*; Johanna Drucker reminds us that a better word that *ought* to be used is *capta*, ‘*that which has been captured or gathered*’.<sup>32</sup>

Thinking of our data as *capta*, we see more clearly that we have to bring some sort of interpretative framework to know *which* questions can be asked. In our context, questions that *do not* percolate out of a database include those questions of context, of individuals, and of interaction and conversation. In particular, questions relating to poverty and homelessness of supposed thieves that were under the age of twelve do not readily percolate out, instead they must be investigated and found. The context of the theft is often lost or was never recorded in the first place, especially when the only material that is readily available is that which is tagged. It is possible to find a court case unique identification code and link that information to individuals and their specific information, including age, gender, occupation, relation to the case, charge, conviction, and punishment.

### **3.3 What Questions Should We Ask**

Because the two works did not employ similar levels of granularity, I employed a tactic of minimizing assumptions. In each stage of an analysis, the (metaphorical) multiplication of assumptions decreases the levels of (metaphorical) confidence we can have in our results. By minimizing assumptions consciously, from the outset, I increase the chances that my results are meaningful. As Beattie was in general (despite my

---

[tools.org/](https://tools.org/). John Russell and Merinda Hensley, “Beyond Buttonology: Digital humanities, digital pedagogy, and the ACRL Framework,” *College and Research Libraries News* 78, no. 11 (2017), <https://crln.acrl.org/index.php/crlnews/article/view/16833/18427>.

<sup>32</sup> Drucker, J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1), 1-21.

comments above) not as granular as the Old Bailey Online, I used Beattie as the baseline for the questions I might ask. Granularity in this case meant that I was unable to separate a term into constituent parts such as ‘sexual crimes’ not being broken down into ‘rape,’ ‘sodomy,’ ‘sexual assault,’ etc. As indicated above, ‘gender’ is recorded in sufficient numbers that a statistical investigation becomes possible; other instances where the *capta* can be made to speak to each other are detailed below.

Hypothesis	Data Source	Counts	Chi-Squared	Phi / Cramer's V	Kendall's Tau	Guttman's Lambda	Result	new_data_worksheet_2.xlsx file in GitHub repo, range
there is an association between being male and property offences	Old Bailey	male & prop. = 742655 male & not prop. = 32423 total = 775078	576.7	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A2:J18 phi_cramersv R = A7:E10 kendalls_tau R = A7:D18
	Beattie	male & prop. = 3619 male & not prop. = 4415 total = 8034	0.1	0	0		Reject; no relationship	chisquared R = L1:BM18 phi_cramersv R = G6:AI10 kendalls_tau R = F6:AC18
there is an association between being guilty and property offences	Old Bailey	guilty & prop. = 763945 guilty & not prop. = 34216 total = 798170	391.5	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A20:J36 phi_cramersv R = A12:E15 kendalls_tau R = A20:D31
	Beattie	guilty & prop. = 2571 guilty & not prop. = 526 total = 3097	35.1	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = L20:U36 phi_cramersv R = G12:K15 kendalls_tau R = F20:I31
there is an association between being male and being sentenced to death for property offences	Old Bailey	male & prop. = 240131 male & not prop. = 19407 total = 259538	250.1	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A39:J55 phi_cramersv R = A17:E20 kendalls_tau R = A33:D44
	Beattie	male & prop. = 233 male & not prop. = 52 total = 285	1.7	0	0		Reject; low relationship; both increase	chisquared R = L39:U55 phi_cramersv R = G17:K20 kendalls_tau R = F33:I44
there is an association between being male and found guilty in Homicide	Old Bailey	male & guilty = 18322 male & not guilty = 5950	33.3	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A58:J74 phi_cramersv R = A22:E25 kendalls_tau R = A47:D58

Cases (or for Manslaughter)		total = 24272						
	Beattie	male & guilty = 136 male & not guilty = 171 total = 307	0.6	0	0		Reject; low relationship; both increase	chisquared R = L58:U74 phi_cramersv R = G22:K25 kendalls_tau R = F47:I58
there is an association between being male and found guilty in Property Cases vs Murder Cases	Old Bailey	male & prop. = 618259 male & murder = 18322 total = 636581	22.8	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A76:J92 phi_cramersv R = A27:E30 kendalls_tau R = A60:D71
	Beattie	male & prop. = 2079 male & murder = 52 total = 2131	3.8	0	0	0	Accept w/ 90% confidence; low relationship; inverse increase	chisquared R = L76:U92 phi_cramersv R = G27:K30 kendalls_tau R = F60:I71 chisquared R = W76:AC90
there is an association between being rape/attempted rape and pre/post 1740	Old Bailey	x	x	x	x	x	Cannot be tested	
	Beattie	pre-1740 & rape = 20 pre-1740 & attempt. rape = 19 total = 39	5.2	x	0.2		Accept w/ 95% confidence; both increase	chisquared R = L94:U110 kendalls_tau R = F73:I84
there is an association between being sentenced to death for different crimes as a male	Old Bailey	male & prop. = 240131 male & not prop. = 19407 total = 259538	250.1	0	0		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A112:J128 phi_cramersv R = A32:E35 kendalls_tau R = A86:D97
	Beattie	male & prop. = 409 male & not prop. = 72 total = 481	3.8	0.1	0.1		Accept w/ 90% confidence; low relationship; inverse increase	chisquared R = L112:U128 phi_cramersv R = G32:K35 kendalls_tau R = F86:I97
there is an association between being guilty and being the accused	Old Bailey	x	x	x	x	x	Cannot be tested	
	Beattie	primary & guilty = 145	7.2	x	0.2		Accept w/ 99% confidence;	chisquared R = L130:U146 kendalls_tau R = F99:I110

		primary & not guilty = 126 total = 271					both increase	
there is an association between male homicide guilty verdicts pre 1740 and post 1740	Old Bailey	pre-1740 & guilty = 3187 pre-1740 & not guilty = 610 total = 3797	168.3	0.1	0.1		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A148:J164 phi_cramersv R = A37:E40 kendalls_tau R = A112:D123
	Beattie	pre-1740 & guilty = 75 pre-1740 & not guilty = 91 total = 166	19.9	0.3	-0.3		Accept w/ 99.9% confidence; moderately strong relationship; inverse increase	chisquared R = L148:U164 phi_cramersv R = G37:K40 kendalls_tau R = F112:I123
there is an association between being male and being sentenced to death	Old Bailey	male & death = 143300 male & not death = 226969 total = 370269	1300.3	0.1	0.1		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A166:J182 phi_cramersv R = A42:E45 kendalls_tau R = A125:D136
	Beattie	male & death = 942 male & not death = 803 total = 1745	0.2	0	0		Reject; low relationship; both increase	chisquared R = L166:U182 phi_cramersv R = G42:K45 kendalls_tau R = F125:I136
there is an association between being male and being sentenced to transportation	Old Bailey	male & trans. = 142528 male & not trans. = 227741 total = 370269	575.4	0	0		Accept w/ 99.9% confidence; low relationship; inverse increase	chisquared R = A184:J200 phi_cramersv R = A47:E50 kendalls_tau R = A138:D149
	Beattie	male & trans. = 383 male & not trans. = 1362 total = 1745	0.3	0	0		Reject; low relationship; inverse increase	chisquared R = L184:U200 phi_cramersv R = G47:K50 kendalls_tau R = F138:I149
there is an association between being male and	Old Bailey	male & prison = 46315	10.4	0	0		Accept w/ 99% confidence; low	chisquared R = A202:J218 phi_cramersv R = A52:E55 kendalls_tau R = A151:D162

being sentenced to imprisonment		male & not prison = 380790 total = 427105					relationship; inverse increase	
	Beattie	male & prison = 24 male & not prison = 1721 total = 1745	49.3	0.2	-0.2		Accept w/ 99% confidence; low relationship; inverse increase	chisquared R = L202:U218 phi_cramersv R = G52:K55 kendalls_tau R = F151:I162
there is an association between being female and being sentenced to death for infanticide	Old Bailey	female & infant. = 62 female & not infant = 2472 total = 2534	258.1	0.1	0.1		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A220:J236 phi_cramersv R = A57:E60 kendalls_tau R = A164:D175
	Beattie	female & infant. = 4 female & not infant = 33 total = 37	52.4	0.3	0.3		Accept w/ 99.9% confidence; strong relationship; both increase	chisquared R = L220:U236 phi_cramersv R = G57:K60 kendalls_tau R = F164:I175
there is an association between being charged with Homicide and found guilty (1660-1740 & 1740-1802)	Old Bailey	1660-1740 & guilty = 3185 1660-1740 & not guilty = 608 total = 3793	256.8	0.1	0.1		Accept w/ 99.9% confidence; low relationship; both increase	chisquared R = A238:J254 phi_cramersv R = A62:E65 kendalls_tau R = A177:D188
	Beattie	1660-1740 & guilty = 27 1660-1740 & not guilty = 91 total = 119	17.5	0.3	-0.3		Accept w/ 99.9% confidence; strong relationship; inverse increase	chisquared R = L238:U254 phi_cramersv R = G62:K65 kendalls_tau R = F177:I188

**Table 4 Hypothesis table and summary of data to one significant digit– Highlighted hypothesis denote those with strong associations across all tests**

## Chapter 4: Discussion

While this thesis is not a historiographic treatise on crime in London in the long eighteenth century, it does appear on this reading that crime in London, Surrey, and Sussex appear very similar. I am not concerned with the historiographic implications of this finding, but rather am concerned with what the results imply for digital history methods.

That said, I will gather some *initial* observations on the historiographic results as a *jumping off* point for a future engagement with this data. Given that my transcriptions of Beattie's data, and my interpretations/translations and analysis of this data are openly shared on the web, the following observations could become a useful starting point for an undergraduate or graduate student's first engagement with digital history. These initial observations are more in the vein of thinking-out-loud of potential ways of taking this research further and are offered in that spirit.

### 4.1 Tentative Thoughts on Historical Implications

In most tests the results from both Beattie and the Old Bailey Online were the same; this is a null result, but this is not a bad result. A null result for a tested hypothesis is not a failure; rather it supports the idea of both sources sharing similarities with reference to the hypothesis being tested.

In such cases where a tested hypothesis resulted in a different conclusion between Beattie and the Old Bailey Online, it seems often different due to gender between defendants. Gender is a critical factor in crime and is often highlighted in Beattie's work. The difference was seen after employing the Chi-Squared statistical analysis test to determine if there is an association between two variables which intersect. In cases where

there was a value close to zero (five or less) I also employed Guttman's Lambda which accounts for small values. In the first instance of a difference, the hypothesis being tested stated that "there is an association between being male and found guilty in Homicide Cases (or for Manslaughter)." For the data in the Old Bailey Online, this was to be accepted with a confidence greater than 99.9%, while for Beattie, this was to be outright rejected. The other two cases used similar hypotheses: "there is an association between being male and being sentenced to death" and "there is an association between being male and being sentenced to transportation." Like the above, for the data in the Old Bailey Online, this was to be accepted with a confidence greater than 99.9%, while for Beattie, this was to be outright rejected. Is this something in the data or something in the methodology?

To test this, I employed Phi/Cramér's V, in an effort to determine if there was any strength to the relationships. To remind the reader, Cramér's V is a similar comparative tool to Chi-squared in that it is a measure of association between two variables. It differs, however, in that it provides the strength of the relationship (the assessment of the effect size) between zero and one.<sup>33</sup> While Chi-squared gave a reason to believe that there were differences between Beattie and the Old Bailey Online, Phi/Cramér's V was crucial in determining whether that relationship had any strength to it. In each of the three cases the association between the two variables in the Old Bailey Online data was *very weak*. In Beattie's material the strength of the relationship was either *moderately strong* or *strong*. I have taken this to mean that it is very strongly opposed to being associated. I further

---

<sup>33</sup> Rebecca Warner, *Applied Statistics: From Bivariate Through Multivariate Techniques* (Washington DC: Sage, 2013), 334.

pushed this line of questioning by exploring Kendall's Tau. Kendall's Tau is like Chi-squared in that it tests for association, but it can be used for very low counts, of which there was one case being explored. Further, Kendall's Tau tests for the direction of an association.<sup>34</sup> This means that it can be discovered if the gender is dependent on the crime or vice versa. Kendall's Tau testing revealed that males sentenced to transportation or to imprisonment had a negative correlation, which "indicates that as the rank of one variable increases the other one decreases."<sup>35</sup> This means that, in this instance, as either the number of males increased, or the numbers of those sentenced to a certain punishment increased, the other should decrease. In the case of being male and found guilty for homicide, the positive correlation which "indicat[es] that the ranks of both variables increase together."<sup>36</sup> Thus, as the number of homicides increases the number of males which are found guilty increases.

There are a wide variety of reasons for why this may be the case. Simplest, is that men are most often associated with violent crime. Put another way:

With greater physical strength combined with greater aggressiveness, men are and have always been far more seriously violent than women. Perpetrators of homicide, excepting the special case of infanticide, have in almost all times and places been largely male, often overwhelmingly so. It is in fact a cliché of criminology that violent criminals are far more likely to be male than female. [...] Thus, it is safe to say that homicide, whether the victims are female or male, is and as far as we can ascertain always has been a highly gendered behaviour.<sup>37</sup>

---

<sup>34</sup> Rebecca Warner, *Applied Statistics: From Bivariate Through Multivariate Techniques* (Washington DC: Sage, 2013), 322.

<sup>35</sup> Nicola Crichton, "Dialysis diet and fluid non-adherence questionnaire," *Journal of Clinical Nursing* 10 (2001): 715.

<sup>36</sup> Nicola Crichton, "Dialysis diet and fluid non-adherence questionnaire," *Journal of Clinical Nursing* 10 (2001): 715.

<sup>37</sup> Martin J. Wiener, *Men of Blood: Violence, Manliness and Criminal Justice in Victorian England* (Cambridge: Cambridge University Press, 2004), 1-2.

The Old Bailey Online and Beattie data also showed a correlation between women and infanticide, as Wiener briefly remarks above. Specifically, in cases in which a criminal was sentenced to death for infanticide, it was always the case, with no exception, that the criminal was female. While males were never found guilty for infanticide, females were found guilty for homicide, just in very low proportions. For the Old Bailey Online it was the case that for every nine males found guilty of homicide there would be two females. For Beattie, however, for every seven males found guilty there was one female found guilty. This does not mean that London cannot be used to create a model for the rest of England; merely that London appears more normalized around the gender-gap in violent crime.

The gender gap bears exploration due to the way it appears in both data sets. In terms of understanding the gender-gap in crime during the Victorian-era, it must first be remembered that the gap primarily consists of those found guilty of a violent crime. Violent crime is not merely assault or homicide; rather it can be non-physical and still be violent. Conventionally, we might assume that there was a belief that women were moral, spiritual, and religious while men were stronger, energetic, and rational.<sup>38</sup> This also led to women being considered weak and fragile and requiring the protections of men. Thus, “much of the toughening of the criminal law [...] was already stimulated [...] by a new concern about violence against women. [...] Nineteenth-century courts – in spite of their all-male composition – focused more and more on men.”<sup>39</sup> Violent crime was not

---

<sup>38</sup> Martin J. Wiener, *Men of Blood: Violence, Manliness and Criminal Justice in Victorian England* (Cambridge: Cambridge University Press, 2004), 29.

<sup>39</sup> Martin J. Wiener, *Men of Blood: Violence, Manliness and Criminal Justice in Victorian England* (Cambridge: Cambridge University Press, 2004), 29.

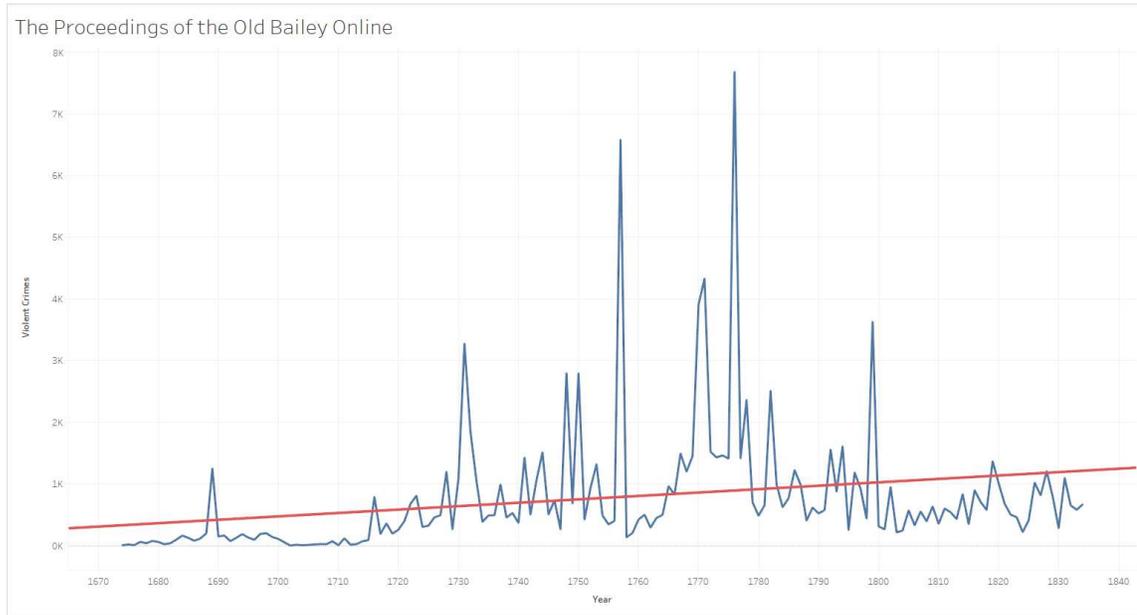
necessarily on the rise. In fact, it was in serious decline in proportion to the population since the medieval ages.<sup>40</sup> Despite this, records of violent crime, especially of men committing violent crime, appear to be on the rise and become more prominent. These records come from the Old Bailey Online data set due to the larger sum of court cases and the inability to data model Beattie's work. This makes it a difficult task to produce comparable examples of crime changing over time; it is necessary to demonstrate that crime was a fluctuating issue.

Illustration One (below) shows this trend with peaks at 1757, 1771, 1776, 1812, and 1820. Each of these years represented a major event in Great Britain. 1757 saw William Pitt the Elder resign from the government, be recalled to the government, and the passage of significant legislation relating to mutiny, militia, supply, and taxation. 1771 and 1776 fall in the middle of the American Revolutionary War and include Prince Henry marry a commoner, legislation relating to copper counterfeiting, and the publication of Adam Smith's *The Wealth of Nations*. 1812 saw conflict on two continents with British forces assaulting the United States of America and fighting against Napoleon in Europe. Finally, 1820 saw a new monarch after the Regency Era, the Radical War in Scotland began, the Cato Street conspiracy was exposed, and the leaders executed or transported, and the Union Suspension Bridge was officially opened between England and Scotland. Each of these years represented political and societal upheaval and turmoil, but this does

---

<sup>40</sup> Manuel Eisner, "Long-Term Historical Trends in Violent Crime," *Violence Research Centre – University of Cambridge*, 85. <https://www.vrc.crim.cam.ac.uk/vrcresearch/paperdownload/manuel-eisner-historical-trends-in-violence.pdf>.

not adequately explain the huge spikes. ‘Violent Crimes’ is a combination of ‘kill,’ ‘sexual offenses,’ and ‘violent theft’ without consideration of guilt.



**Illustration 1 Count of men charged with violent crimes at the Old Bailey Online**

When there is a consideration of ‘guilty’ and ‘not guilty’ there is an obvious dichotomy. When the years with spikes of guilty are visible, not guilty is almost nonexistent. Following those spikes, however, are usually years where not guilty is returned most often. Why is this the case and why do the numbers not sit closely to the trend line? One might draw a connection to who the Recorder of London was at the time. The Recorder of London was the senior judge present at the Central Criminal Court. One could imagine that this made a difference in ascribing guilt. During the period under consideration (1673-1834) there were twenty-three Recorders of London, of which fourteen served less than ten years and only nine served more than ten years.<sup>41</sup>

---

<sup>41</sup> John Noorthouck, "Addenda: The Recorders of London," in *A New History of London Including Westminster and Southwark*, (London: R Baldwin, 1773), 893-894. British History Online, accessed July 10, 2018, <http://www.british-history.ac.uk/no-series/new-history-london/pp893-894>.

Furthermore, the longer length of service was heavily weighted towards the later years rather than the early years. 1753 to 1778, which saw one of the two major periods in which spikes of violent crime charges occurred were under the purview of only two Recorders, Sir William Moreton and Sir James Eyre.<sup>42</sup> Both Moreton and Eyre were qualified in the practice of law and in treasury, which may have seen them act to preserve the rule of law, especially as it stands in relation to monetary claims.

#### **4.2 Thoughts on Methodological Implications**

The conjectures above are merely that, conjectures. Finding no real difference overall, statistically, in the patterns of crime recorded is more meaningful in what it causes us to reflect on in terms of methodology. Is a ‘null result’ a failure?

There are two ways to conceive of ‘failing’: that a failure to produce a suspected result means a complete loss for the process and conclusion or that a failure produces something that can be recovered and shared for others to learn from. In this first sense a failing is something which has produced an unexpected or unusable result which does not further the hypothesis or which renders a previous belief void. Even in this sense the failure may still be usable. Furthermore, a failure is usually one that, in the digital humanities, halts progress or stymies useful discussion by placing an arbitrary limit on something. The other failing is one that, despite not resulting in the expected outcome, still provides a way for others to understand their own methodology and to understand

---

<sup>42</sup> Douglas Hay, "Eyre, Sir James (bap. 1734, d. 1799)." *Oxford Dictionary of National Biography*. Eds. Leslie Stephen, Sidney Lee, and Christine Nicholls. Oxford, UK: Oxford University Press, 2004. [http://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=3395&context=scholarly\\_works](http://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=3395&context=scholarly_works).

where a process becomes derailed. Below, I explore how negative results can still lead to progress.

Productive failings are never a negative because they still allow for needed discussion, learning, and growth. There are pedagogical issues which can cause researchers and students to fear failing and Shawn Graham posits that “This is perhaps a result of high-stakes testing and the ways we as educators have inculcated all-or-nothing grading in our courses.”<sup>43</sup> In this sense, it must be possible that there is a way in which one can work and not find the result that one initially set out to find and yet remain successful.

A simple option for this is to maintain an open notebook or to maintain access to a source control system for the work under consideration, such as the GitHub repository used for my research.<sup>44</sup> According to Graham there is a taxonomy of fails which are compiled as technological failure, human failure, failure as artifact, failure as epistemology, and failing to share.<sup>45</sup> It is often the case that failing to share ones raw work or ones methodology can lead others to fail in the same ways and without having already discussed it, this can lead to still others failing again. To fail openly, therefore, is to share a way forward that may not include that particular method, but this remains a

---

<sup>43</sup> Shawn Graham, “Failing Productively in Digital Archaeology,” Electric Archaeology, last modified March 14, 2017, <https://electricarchaeology.ca/2017/03/14/failing-productively-in-digital-archaeology/>.

<sup>44</sup> Matthew Dodd, “The Empire of the Old Bailey v.0.1.1,” last modified July 28, 2018, <https://doi.org/10.5281/zenodo.1322132>.

<sup>45</sup> Shawn Graham, “Failing Productively in Digital Archaeology,” Electric Archaeology, last modified March 14, 2017, <https://electricarchaeology.ca/2017/03/14/failing-productively-in-digital-archaeology/>.

productive failure. Thus, the simplest way to fail productively is to document a process, either openly or closed for later reveal, regardless of ultimate results.<sup>46</sup>

For Beattie and the Old Bailey Online the failures do not lie solely in human or technological failures, but rather in that fifth failure: failure to share. Had Beattie merely compiled his rough work and his raw data into a journal or notebook and later published this, or published this as a companion monograph, there would be fewer issues with reproducibility and replicability.

Likewise, the Old Bailey Online provides substantial corpora on definitions and history as it is relevant to the material, but it provides no paradata document (recall that ‘paradata’ are the data that describe the process and decision making in representing the data). Had Beattie provided something akin to a paradata document, the option of reproducing his work would have been feasible. Had the Old Bailey Online provided the paradata document, understanding the decisions made in the construction of the database by the developers would have required less interpretation. Thus, despite the failure to share, there is still the productive failure that is learning from this experience: scholars must endeavour to share all forms of data (data, metadata, and paradata) to ensure that future failures or misinterpretations are not repeated.

Peter Suber argues that open access to materials in the digital age is a public good and that open access actually grants the original author more control over the integrity of their work rather than the mere license of access to their work.<sup>47</sup> Had Beattie made his raw material available then, according to Suber, his ability to preserve the integrity of the

---

<sup>46</sup> Shawn Graham, “Open Notebook Research in Digital Archaeology,” *Electric Archaeology*, last modified March 15, 2017, <https://electricarchaeology.ca/2017/03/15/open-notebook-research-in-digital-archaeology/>.

<sup>47</sup> Peter Suber, *Knowledge Unbound* (Cambridge: MIT Press, 2016), 3,26.

work would have been much stronger than his chosen method of integrity preservation: hiding all of the raw work. This, however, only addresses potential issues with replicability and not with reproducibility.

Reproducibility in science is “the calculation of quantitative scientific results by independent researchers using the original data and methods.”<sup>48</sup> Scholarship is strengthened by providing conventions and technical infrastructure to facilitate openness. To do this, according to Marwick, Boettiger, and Mullen, a research compendium should be provided. A research compendium follows three principles: organized following the prevailing standards of the scholarly community, a clear separation of data, method, and output and an unambiguous expression of the relationship between the three, and the specifics of the computational environment used for the original analysis. This, however, only works when the scholar has the data to work with, which is not the case with Beattie. Beattie could have released a companion document with all of his raw material, which could have been digitized.

Replicability is the ability to perform the same tests and methods of transformation on another dataset that was performed on a previous dataset. Finding significance in one data source is a fairly straightforward endeavour. Finding it in two data sources is just as straightforward. Finding significance in *across* two disparate data sources without knowing how the data was originally catalogued, having access to the original data, knowing the original methodologies, or knowing the reasons for the final

---

<sup>48</sup> Ben Marwick, Carl Boettiger, and Lincoln Mullen, “Packaging data analytical work reproducibly using R (and friends),” *The American Statistician* 72 (2018): 80.

output combine to make significance a more difficult prospect. Revisiting work, is not a duplication of effort, but rather a crucial step for verification and validation of results.

## Chapter 5: Conclusion

To see anything in the Old Bailey Online through the data requires a deep dive into the murky waters in which that data resides; it requires a foil against which its strengths and weaknesses can be tested. From there the specifics of that which is under investigation may be found by random search, luck, or some combination of experience and training in the Old Bailey Online. In this latter case, it is often that the work which results is authored, or co-authored, by Tim Hitchcock who is one of the primary individuals responsible for the Old Bailey Online in its current state.

To this end, Hitchcock performs these deep dives to formulate a quantitative history *via patterns in word use* of the material, in particular through a text analysis initiative. In this initiative Hitchcock finds that the shortest trial is only eight words, that the longest is over 155,000 words, and that the entire Old Bailey Online corpus has over 127,000,000 words.<sup>49</sup> Hitchcock, and the other authors, claims that this research is to show:

A detailed description of the Proceedings as a single massive text object, [...] to compare these measures of a changing text to statistics reflecting the behavior of the court (patterns of prosecution and convictions), [...] and] to use these two measures in combination to both test the reliability of the Proceedings as evidence of court room practice at the Old Bailey in the eighteenth century, and of changing court behavior in the nineteenth.<sup>50</sup>

Text analysis serves a different purpose than what I have sought to achieve here.

Hitchcock also argues that “academic history writing lost the plot” when it comes to

---

<sup>49</sup> Tim Hitchcock and William J. Turkel, "The Old Bailey Proceedings, 1674-1913: Text Mining for Evidence of Court Behavior," *Law and History Review* 34, no. 4 (11, 2016): 929-955, <http://proxy.library.carleton.ca/login?url=https://search-proquest-com.proxy.library.carleton.ca/docview/1824680975?accountid=9894> (accessed June 19, 2018).

<sup>50</sup> Tim Hitchcock and William J. Turkel, "The Old Bailey Proceedings, 1674-1913: Text Mining for Evidence of Court Behavior," *Law and History Review* 34, no. 4 (11, 2016): 929-955,

confronting digital resources. Hitchcock argues that there are inherent issues with algorithm-driven research, poor OCR, and selection bias. Furthermore, “academic historians have largely failed to respond effectively to these challenges and suggests that while they have preserved the form of scholarly good practice, they have ignored important underlying principles.”<sup>51</sup> Hitchcock argues that in many ways the rigour of academic writing and research through evidence has fallen to the wayside in the quest for easier ways to research and faster results from a search engine. The secondary critique, however, is that the technology itself does not allow for a researcher to accurately access the original work and thus cannot verify anything that they are provided by an electronic resource. Sources need to be put into dialogue with one another.

In this research I have found that it was almost always the case that the Old Bailey Online and Beattie’s datasets agreed about the relationship and strength of variables. The Empire of the Old Bailey marches on. In those few cases in which they did not, it was often the case that those failings were not the fault of the data, but rather the fault of the ways in which material was organized. I suggest that open research practices could mitigate these kinds of issues in the future, and in choosing to conduct my research, warts-and-all, in the open and making my process and data available, the example of this present thesis moves the discipline forward.

This experiment suggests that London may not be as epiphenomenal a site of social and cultural expression as many in the eighteenth and nineteenth centuries (as well

---

<http://proxy.library.carleton.ca/login?url=https://search-proquest-com.proxy.library.carleton.ca/docview/1824680975?accountid=9894> (accessed June 19, 2018).

<sup>51</sup> Tim Hitchcock, “Confronting the Digital: or How Academic History Writing Lost the Plot,” *The Journal of the Social History Society* 10, no. 1 (May 2015), 9.

as historians today), assumed. While I cannot overstate my findings and conclude that crime in London was the same as crime elsewhere in England, I would encourage future scholars to question their assumptions about London as categorically different from the rest of Britain and to explore this problem further. I would also encourage others to critically approach, as best they can, the analytical processes and transparency of historiographical “chestnuts” such as John Beattie’s *Crime and the Courts of England* to see how we can make these earlier works speak to the databases and digital methodologies of today. In this way, we can begin to evaluate the representativeness of online resources such as the Old Bailey Online and start to understand how digital accessibility might shift our understandings of “what happened” in the past and how.

## Appendices

### Appendix A Glossary

1. XML: eXtensible Markup Language is a machine and human readable data format for facilitating the accurate tagging, or identifying, of particular information in a larger text.
2. API: an Application Programming Interface is a software connection between a website and a program which facilitates the transfer of data over the internet.
3. TEI: Text Encoding Initiative is an organization dedicated to forming usable standards for text encoding, in particular XML, which can be implemented to facilitate ease of integration in data.
4. SPARQL: SPARQL Protocol and RDF Query Language is a programming language designed to facilitate the retrieval of data and information stored in an RDF format.
5. RDF: Resource Descriptive Framework is a specification for metadata data models. Specifically, it is a standard for the way in which metadata (data which describes other data) should be formatted.
6. RDBMS: Relational Database Management System based on the relationship model. This means different pieces of data modelled in relation to each other.
7. MySQL: an open-source relational database management system designed for storing complex datasets.
8. Oracle RDBMS: Oracle developed relational database management system designed for storing complex datasets. This is an enterprise database solution.

While a single, personal-use only version does exist, this is usually a very expensive RDBMS.

9. SQL: Structured Query Language is a programming language designed to facilitate the retrieval of data from many modern relational databases.
10. Schema: a database schema is the primary structure of the database including all objects, constraints, functions, and more, but excluding tablespaces (storage) and roles.
11. OCR: Optical Character Recognition is a software process designed to scan in a printed document and analyze the written content on the page and convert it into digital content which is then searchable and modifiable.
12. JSON: JavaScript Object Notation is an open-standard file format used for language independence and simplicity of reading either by a computer or by a human.
13. Metadata: this is the definition and constraints which go into understanding what a piece of data is meant to be understood as.
14. ICPSR: Inter-university Consortium for Political and Social Research is a support organization for funding requests, research, and instruction in data science principles. It also maintains a data archive and other specialized collections.

## Bibliography

- “About this Project.” London Lives 1690 to 1800. Last modified March 2018.  
<https://www.londonlives.org/static/Project.jsp>.
- Anderson, Deborah. “Standards.” Digital Humanities Data Curation. Accessed July 28, 2018. <https://guide.dhcurator.org/contents/standards/>.
- Beattie, John M. *Crime and the Courts in England 1660-1800*. Princeton: Princeton University Press, 1986.
- Blackadar, Jeff. “Introduction to MySQL with R.” The Programming Historian. Last modified May 19, 2018. <https://programminghistorian.org/en/lessons/getting-started-with-mysql-using-r>.
- “Citations of Crime and the Courts in England, 1600-1800.” Google Scholar. Accessed May 15, 2018.  
[https://scholar.google.ca/scholar?cites=12854185621251261658&as\\_sdt=2005&sciodt=0,5&hl=en](https://scholar.google.ca/scholar?cites=12854185621251261658&as_sdt=2005&sciodt=0,5&hl=en).
- Crichton, Nicola. “Dialysis diet and fluid non-adherence questionnaire.” *Journal of Clinical Nursing* 10 (2001): 707-715.
- Crymble, Adam. “Does Your Historical Collection need a Database-Driven Website?” *Digital Humanities Quarterly* 9, no. 1 (2015).  
<http://www.digitalhumanities.org/dhq/vol/9/1/000206/000206.html>.
- Dodd, Matthew. “The Empire of the Old Bailey v.0.1.1.” Last modified July 28, 2018.  
<https://doi.org/10.5281/zenodo.1322132>.
- Drucker, J. “Humanities approaches to graphical display.” *Digital Humanities Quarterly* 5(1) 2011, 1-21.
- Eisner, Manuel. “Long-Term Historical Trends in Violent Crime.” *Violence Research Centre – University of Cambridge*.  
<https://www.vrc.crim.cam.ac.uk/vrcresearch/paperdownload/manuel-eisner-historical-trends-in-violence.pdf>.
- Feilden, Tom. “Most scientists 'can't replicate studies by their peers'.” BBC News, February 22, 2017. <http://www.bbc.com/news/science-environment-39054778>.
- Fletcher, Mike and Gary R. Lock. *Digging Numbers: Elementary Statistics for Archaeologists*. Exeter: Short Run Press, 1994.
- Fotopoulou, Aikaterini, Martin A. Conway, and Mark Solms. “Confabulation: Motivated reality monitoring.” *Neuropsychologia* 45, no. 10 (2007): 2180-2190.  
<https://doi.org/10.1016/j.neuropsychologia.2007.03.003>.
- Graham, Shawn. “Failing Productively in Digital Archaeology.” *Electric Archaeology*. Last modified March 14, 2017. <https://electricarchaeology.ca/2017/03/14/failing-productively-in-digital-archaeology/>.
- Graham, Shawn. “Open Notebook Research in Digital Archaeology.” *Electric Archaeology*. Last modified March 15, 2017.  
<https://electricarchaeology.ca/2017/03/15/open-notebook-research-in-digital-archaeology/>.
- Graham, Shawn, Ian Milligan, and Scott Weingart. “Final Draft.” Exploring Big Historical Data: The Historian’s Macroscope. Accessed June 19, 2018.  
[http://www.themacroscope.org/?page\\_id=584](http://www.themacroscope.org/?page_id=584).
- Grant, James. *The Great Metropolis*. Volume 1. New York: Saunders and Otley, 1837.

- Gray, Drew. *Crime, Policing and Punishment in England, 1660-1914*. New York: Bloomsbury Publishing, 2016.
- Gray, Drew. "Putting undergraduates on trial: using the Old Bailey Online as a teaching and assessment tool." *SOLON Law, Crime and History*, 4(1) 2014. 104-113.
- Greif, A. "Cliometrics after 40 years." *The American Economic Review* 87 (2) (1997): 400-403.
- Greenberg, Steven. "How citation distortions create unfounded authority: analysis of a citation network." *British Medical Journal* 339 (March 2009): 1-14. <https://doi:10.1136/bmj.b2680>.
- Hay, Douglas. "Eyre, Sir James (bap. 1734, d. 1799)." *Oxford Dictionary of National Biography*. Eds. Leslie Stephen, Sidney Lee, and Christine Nicholls. Oxford, UK: Oxford University Press, 2004. [http://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=3395&context=scholarly\\_works](http://digitalcommons.osgoode.yorku.ca/cgi/viewcontent.cgi?article=3395&context=scholarly_works).
- Hirstein, William. *Brain Fiction: Self-Deception and the Riddle of Confabulation*. Cambridge: MIT Press, 2005.
- Hitchcock, Tim. "Confronting the Digital: or How Academic History Writing Lost the Plot." *The Journal of the Social History Society* 10, no. 1 (May 2015), 9-23.
- Hitchcock, Tim and William J. Turkel. "The Old Bailey Proceedings, 1674-1913: Text Mining for Evidence of Court Behavior." *Law and History Review* 34, no. 4 (11, 2016): 929-955, <http://proxy.library.carleton.ca/login?url=https://search-proquest-com.proxy.library.carleton.ca/docview/1824680975?accountid=9894> (accessed June 19, 2018).
- Howard, S. "Bloody Code: Reflecting on a decade of the Old Bailey online and the digital futures of our criminal past." *Proceedings of the National Academies of Science*. <http://www.pnas.org/content/pnas/early/2014/06/11/1405984111.full.pdf>
- "ICPSR." Inter-university Consortium for Political and Social Research. Accessed June 19, 2018. <https://www.icpsr.umich.edu/icpsrweb/>.
- Lincoln, Matthew. "Confabulation in the Humanities." Last modified March 21, 2015. <https://matthewlincoln.net/2015/03/21/confabulation-in-the-humanities.html>.
- Martini, Beatrice. "Decolonizing technology: A reading list." Accessed August 5, 2018. <https://beatricemartini.it/blog/decolonizing-technology-reading-list/>.
- Marwick, Ben, Carl Boettiger, and Lincoln Mullen. "Packaging data analytical work reproducibly using R (and friends)." *The American Statistician* 72 (2018): 80-88. <https://doi:10.1080/00031305.2017.1375986>.
- May, Allyson. *The Bar and the Old Bailey, 1750-1850*. Chapel Hill: University of North Carolina Press, 2003.
- McDaniel, Caleb. *Open Notebook History*. 2013. <http://wcm1.web.rice.edu/open-notebook-history.html>
- Milligan, Ian. "Illusionary Order: Cautionary Notes for Online Newspapers." Ian Milligan. Last modified March 26, 2012. <https://ianmilligan.ca/2012/03/26/illusionary-order-cautionary-notes-for-online-newspapers/>.
- Milligan, Ian. "New Article: "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010." Ian Milligan. Last modified November 19, 2013. <https://ianmilligan.ca/2013/11/19/new-article-illusionary->

- order-online-databases-optical-character-recognition-and-canadian-history-1997-2010/.
- Morse, Christopher. "Research Databases and the Future of Digital Humanities Applications." Digital Arts and Humanities Harvard. Accessed May 15, 2018. <http://darthcrimson.org/research-databases-future-digital-humanities-applications/>.
- Moscovitch, Morris. "Confabulation." In *Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past*, edited by Joseph T. Coyle, 226-251. Cambridge: Harvard University Press, 1979.
- Noorthouck, John. "Addenda: The Recorders of London," in *A New History of London Including Westminster and Southwark*, (London: R Baldwin, 1773), 893-894. British History Online, accessed July 10, 2018, <http://www.british-history.ac.uk/no-series/new-history-london/pp893-894>.
- The Old Bailey Online. "Access to Old Bailey Online Data." Accessed June 6, 2018. <https://www.oldbaileyonline.org/static/Data.jsp>.
- The Old Bailey Online. "Bibliography: Judicial Procedures." Last modified March 2018. <https://www.oldbaileyonline.org/static/JudicialBibliography.jsp>.
- The Old Bailey Online. "Crimes tried at the Old Bailey." Last modified March 2018. <https://www.oldbaileyonline.org/static/Crimes.jsp>.
- The Old Bailey Online. "Publications that Cite the Old Bailey Proceedings Online." Last modified March 2018. <https://www.oldbaileyonline.org/static/Publications.jsp>.
- Russell, John and Merinda Hensley. "Beyond Buttonology: Digital humanities, digital pedagogy, and the ACRL Framework." *College and Research Libraries News* 78, no. 11 (2017). <https://crln.acrl.org/index.php/crlnews/article/view/16833/18427>.
- Suber, Peter. *Knowledge Unbound*. Cambridge: MIT Press, 2016.
- van Strien, Daniel. An Introduction to Version Control Using Github Desktop, *The Programming Historian* 2016. <https://programminghistorian.org/en/lessons/getting-started-with-github-desktop>.
- Warner, Rebecca. *Applied Statistics: From Bivariate Through Multivariate Techniques*. Washington DC: Sage, 2013.
- Weber, Max. "Politics as Vocation." Speech given January 28, 1919 in Munich, Bavaria. <http://anthropos-lab.net/wp/wp-content/uploads/2011/12/Weber-Politics-as-a-Vocation.pdf>.
- Weingart, Scott. "Ghosts in the Machine." Last modified November 1, 2015. <http://scottbot.net/ghosts-in-the-machine/>.
- Wiener, Martin J. *Men of Blood: Violence, Manliness and Criminal Justice in Victorian England*. Cambridge: Cambridge University Press, 2004.
- Turkel, William. "Data Mining with Criminal Intent." *Criminal Intent*. last accessed September 16, 2018. <http://criminalintent.org/wp-content/uploads/2011/09/Data-Mining-with-Criminal-Intent-Final1.pdf>.
- Underwood, Ted. "Distant Reading and Recent Intellectual History." In *Debates in the Digital Humanities*, edited by Matthew K. Gold. Minneapolis: University of Minnesota Press, 2016. <http://dhdebates.gc.cuny.edu/debates/text/95>.