

# EARLY MODERN LITERARY STUDIES



## **Shakestats: Writing About Shakespeare Between the Humanities and the Social Sciences**

Jeffrey R. Wilson

Harvard University

[jeffreywilson@fas.harvard.edu](mailto:jeffreywilson@fas.harvard.edu)

This essay assesses the role of quantitative data in Shakespeare studies. I make the deflationary case that statistics do indeed have a part to play in both our classrooms and our scholarship, but it is not a particularly revolutionary role. Statistics are best used in Shakespeare studies to ask questions but not to answer them. In other words, quantitative data can help us understand where we need to direct qualitative analysis, but statistics won't interpret things for us. This is not, I suspect, a controversial statement. It might even be painfully obvious, but it needs to be said because it tempers the claims of both the giddy grad student who believes computer-aided analysis can revolutionize literary studies (it can't) and the befuddled professor emeritus who fears computers will de-humanize the humanities (they won't).

### **I. Statistics in the Digital Humanities: Humanities Computing, Distant Reading, and Culturomics**

Behind this tension between qualitative and quantitative thinking is what C.P. Snow identified in 1956 as the problem of the 'two cultures'. His well-known critique of the humanists who do not understand the science they reject actually climaxes in an allusion to Shakespeare:

A good many times I have been present at gatherings of people who, by the standards of the traditional culture, are thought highly educated and who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold: it was also negative. Yet I was asking something which is the

scientific equivalent of: *Have you read a work of Shakespeare's?*... So the great edifice of modern physics goes up, and the majority of the cleverest people in the western world have about as much insight into it as their neolithic ancestors would have had.<sup>1</sup>

To this day, the humanist's suspicion and rejection of scientific ideas and methodologies he or she may not understand remains a central challenge to the digital humanities, which has involved several sub-fields over the years, including humanities computing, distant reading, and culturomics.

With precedents in Victorian philology and Russian formalism, the digital humanities which embrace the quantitative analysis that traditional humanities scholarship shuns is not particularly new. The founding father of digital humanities is often identified as Roberto Busa, an Italian Jesuit who, in the 1940s (when computers were the size of walls), started developing software that could search his massive, 56-volume concordance of the complete works of St. Thomas Aquinas.<sup>2</sup> Busa's *Index Thomisticus* was completed in the 1970s and, between the 60s and 70s, the digital humanities started gaining institutional credibility with the creation of international conferences (such as the *Literary Data Processing Conference* held in Yorktown Heights, NY in 1964), academic centers (like the Centre for Literary and Linguistic Computing established at Cambridge in 1963), dedicated journals (*Computers and the Humanities*, founded in 1966), and scholarly monographs (including Dolores M. Burton's *Shakespeare's Grammatical Style: A Computer-Assisted Analysis of 'Richard II' and 'Antony and Cleopatra'* [1973]).<sup>3</sup>

---

The author would like to thank Hugh Craig, Gabriel Egan, Laura Estill, Kyle Vitale, and the audience at the 2017 MLA Convention panel "Teaching Shakespeare: New Digital Challenges and Solutions" for comments and conversations about the ideas presented in this article.

<sup>1</sup> C.P. Snow, *The Two Cultures* (1956; Cambridge: Cambridge University Press, 1959), pp. 14-15.

<sup>2</sup> See Roberto Busa, 'Complete Index Verborum of St. Thomas', *Speculum: A Journal of Medieval Studies* 25.1 (1950), 424-5; Busa, 'The Annals of Humanities Computing: The Index Thomisticus', *Computers and the Humanities* 14 (1980), 83-90; Busa, 'Foreword: Perspectives on the Digital Humanities', in *A Companion to Digital Humanities*, ed. by Susan Schreibman, Ray Siemens, and John Unsworth (Oxford: Blackwell, 2004), pp. xvi-xxi; and Stephen E. Jones, *Roberto Busa, S.J., and The Emergence of Humanities Computing: The Priest and the Punched Cards* (New York: Routledge, 2016).

<sup>3</sup> See Susan Hockey, 'The History of Humanities Computing', in *A Companion to Digital Humanities*, pp. 3-19. These developments occurred on the backdrop of a renewed search for a unified theory of knowledge, what Edward Wilson called *Consilience: The Unity of Knowledge* (New York: Alfred A. Knopf, 1998). In the introduction to *Creating Consilience: Integrating the Sciences and the Humanities* (Oxford: Oxford University Press, 2012), editors Edward Slingerland and Mark Collard identified a

In 2000, Franco Moretti rattled the literary studies establishment with his disparagement of ‘close reading’ as ‘a theological exercise – very solemn treatment of very few texts taken very seriously’; the antidote, Moretti argued, is ‘distant reading’ done by computers, which not only allows you to analyze massive amounts of non-canonical texts, but also ‘allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems’.<sup>4</sup> It is unclear why Moretti felt the promotion of ‘distant reading’ entailed the rejection of ‘close reading’. Academic gamesmanship seems to have played a big part, as opposed to any logically compelling reason; more charitably, perhaps it was a shrewd calculation that, in order to get the funding needed to establish the Literary Lab at Stanford University, he would need to distance himself as much as possible from the unfundable humanities.<sup>5</sup> In any event, Moretti missed an opportunity to consider the relationship between the ‘distant reading’ enabled by computers, which is directed to questions of fact, and the ‘close reading’ that only humans can do, directed to questions of meaning.

In 2011, a group of young social scientists from Harvard University led by Jean-Baptiste Michel partnered with Google to create a digital tool for investigating cultural trends quantitatively, a field they dubbed ‘culturomics’. Drawing from Google Books, Michel and company created ‘a corpus of 5,195,769 digitized books containing ~4% of all books ever published’.<sup>6</sup> With this database, they (and anyone) could generate quantitative data about trends in language use over time through a publicly accessible and easy-to-use program called the ‘Ngram Viewer’. As they explained their lingo, ‘A 1-gram is a string of characters uninterrupted by a space; this includes words (“banana”,

---

‘second wave consilience’, ‘one in which humanists and scientists work together as equal partners in constructing a shared framework for inquiry’ (p. 4).

<sup>4</sup> Franco Moretti, ‘Conjectures on World Literature’, *New Left Review* 1 (2000), 54-68 (p. 57-8). See also Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005); Stephen Ramsay, *Reading Machines: Toward an Algorithmic Criticism* (Champaign: University of Illinois Press, 2011); Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History* (Champaign: University of Illinois Press, 2013); and Moretti, *Distant Reading* (London: Verso, 2013). For representative critiques, see Kathryn Schulz, ‘The Mechanic Muse: What is Distant Reading?’ *The New York Times Sunday Book Review* (June 24, 2011), <http://www.nytimes.com/2011/06/26/books/review/the-mechanic-muse-what-is-distant-reading.html>; and Stanley Fish, ‘Mind Your P’s and B’s: The Digital Humanities and Interpretation’, *The New York Times* (Jan. 23, 2012), <https://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation>.

<sup>5</sup> The more charitable reading was voiced by Cary Wolfe, ‘Scale and Literary Studies: A Conversation with Franco Moretti’ (Floriana, Malta: Annual Conference of the Society for Literature, Science, and the Arts, June 2015).

<sup>6</sup> Jean-Baptiste Michel, et al., ‘Quantitative Analysis of Culture Using Millions of Digitized Books’, *Science* 331.5014 (Jan. 14, 2011), 176-82 (p. 176).

“SCUBA”) but also numbers (“3.14159”) and typos (“excesss”). An n-gram is a sequence of 1-grams, such as the phrases “stock market” (a 2-gram) and “the United States of America” (a 5-gram).<sup>7</sup> As an example, they pointed out that the phrase ‘the Great War’ fell out of usage in the early 1940s as the phrases ‘World War I’ and ‘World War II’ came into usage. The Ngram viewer can search different languages (English, French, Chinese, etc.), and in some cases even different dialects (e.g., American English, British English). But there are limitations in the program: the Ngram Viewer is not very reliable before the year 1700, and mistranscriptions in Google Books can lead to inaccurate results.<sup>8</sup> In contrast to Moretti’s disparagement of traditional humanistic thought, however, Michel and his team concluded with a call for qualitative analysis to be applied to the quantitative analysis done in culturomics: ‘Culturomic results are a new type of evidence in the humanities. As with fossils of ancient creatures, the challenge of culturomics lies in the interpretation of this evidence’.<sup>9</sup>

## II. Digital Shakespeare

The rise of the digital humanities has impacted Shakespeare studies in several ways.<sup>10</sup> First, the digital age has increased access to Shakespearean texts and contexts, whether through free publicly available yet reliably edited online versions of Shakespeare’s works (such as the *Internet Shakespeare Editions*)<sup>11</sup> or through electronic versions of early printed books easily accessible for academics through institutional subscriptions (such as *Literature Online* and *Early English Books Online*).<sup>12</sup> Second, the digital age has fostered communication in the Shakespearean community through online platforms

---

<sup>7</sup> Ibid, 176.

<sup>8</sup> For example, the ‘long s’ (ſ) popular in English books in the seventeenth century is often mistranscribed in Google Books as an ‘f’ rather than an ‘s’. In the original version of the Ngram Viewer, this led to misleading results when searching for a term such as ‘Shakespeare’, but the problem was corrected in a 2009 re-release of the program. The older and newer versions are both still available.

<sup>9</sup> Michel et al., 181.

<sup>10</sup> For general overviews, see the companion essays, Ian Lancashire, ‘The State of Computing in Shakespeare’, *The Shakespearean International Yearbook* 2 (2002), 89-110; and Brett D. Hirsch and Hugh Craig, “‘Mingled yarn’: The State of Computing in Shakespeare 2.0”, *The Shakespearean International Yearbook* 14 (2014), 3-35.

<sup>11</sup> See Michael Best, ‘The Internet Shakespeare Editions: Scholarly Shakespeare on the Web’, *Shakespeare* 14 (2008), 221-33.

<sup>12</sup> See Steven Hall, ‘Literature Online: Building a Home for English and American Literature on the World Wide Web’, *Computers and the Humanities* 32.4 (1998), 285-301; Diana Kichuk, ‘Metamorphosis: Remediation in Early English Books Online (EEBO)’, *Literary and Linguistic Computing* 22.3 (2007), 291-303.

such as the listserv SHAKSPER (founded in 1990).<sup>13</sup> Third, and most importantly, the digital age has brought about new tools and methods for quantifying our studies of Shakespeare in statistics.

On this third point, the digital is a distraction. While the digital humanities is a controversial and polarizing subject, it is only relevant here because it makes statistical analysis faster and easier. Statistics (about Shakespeare or otherwise) predated and can operate independently of computers, which is why Martin Mulleer preferred the term 'literary infomatics' to 'digital' or 'computer' criticism.<sup>14</sup> The rise of digital tools has made statistics more accessible for more people, including our students, who are my focus in the last third of this article. Students can now do on their laptops modes of statistical interpretation that previously only professional scholars with institutional support and finely curated databases could do. As I see it, the impact of the digital humanities will not be the creation of new forms of interpretation; instead, it will be the mainstreaming of a previously niche mode of analysis. That is why it is most appropriate to discuss the statistical analysis of Shakespeare as a pedagogical concern.

Broadly speaking, statistics have been brought to bear in Shakespeare studies in two ways. The first approach does a statistical analysis of Shakespeare's text itself, while the second involves a statistical consideration of Shakespeare's reception. Taking up each field in turn, the two sections that follow gather together some data sets about Shakespeare and his afterlife. My goal is not to do some sort of meta-analysis: the parameters and approaches of the studies cited are too various. Instead, my goal is to bring together in one place a series of quantitative studies that call out for further explanation. I hope to nudge our use of these data from historical observation to literary interpretation, bringing statistical studies of Shakespeare back to the first-order questions about meaning and value that brought us all to Shakespeare in the first place.

### **III. Text-Based Shakestats**

The most common concern in computer-aided textual studies is authorship, including questions of collaboration. Thanks to methodologies pioneered by MacDonald P. Jackson, Brian Vickers, and Hugh Craig, among others, stylometric analysis aided by large scale data processing of words and phrases is allowing us to speak with increasing

---

<sup>13</sup> See Hardy M. Cook, 'Behind the Scenes with SHAKSPER: The Global Electronic Shakespeare Conference', *College Literature* 36.1 (2009), 105-20.

<sup>14</sup> Martin Mueller, 'Digital Shakespeare, or Towards a Literary Informatics', *Shakespeare* 4 (2008), 284-301.

certainty about which texts Shakespeare wrote, which ones he didn't, and which ones show signs of collaboration.<sup>15</sup> Computer-based attribution in Shakespeare studies will always be haunted, however, by its first major instance: a failure. In 1996, Donald Foster used computer-based analysis to claim Shakespeare wrote *A Funeral Elegy* (published in 1612 with its author listed as 'W.S.'), but Foster recanted after Gilles Monsarrat demonstrated in 2002, using old-fashioned language analysis, that the author was actually John Ford.<sup>16</sup> (As Matthew Steggle has discussed, much of the debate about *A Funeral Elegy*, including Foster's recantation, took place on the online listserv SHAKSPER.<sup>17</sup>) In any event, it was recently international news that *The New Oxford Shakespeare* would contain 40 plays, 12 involving co-authors, and two additional sets of scenes that Shakespeare added to existing plays.<sup>18</sup> Computer-aided statistical analysis allowed the Oxford editors to conclude that Shakespeare had a hand in *The Tragedy of M. Arden of Faversham*; that *Titus Andronicus* was a collaboration between Shakespeare, Peele, and possibly Middleton; that Nashe contributed to *1 Henry VI*, while Marlowe and Shakespeare collaborated on all three of the *Henry VI* plays; that Shakespeare had a hand in the anonymous *Edward III*; that there were a number of collaborators, including Shakespeare, on *The Passionate Pilgrim*; that Shakespeare probably added some scenes to Kyd's *The Spanish Tragedy*; that he also added some scenes to Munday's *Sir Thomas More*; that Shakespeare may have contributed to a lost version of Jonson's *Sejanus*; that *Measure for Measure*, *All's Well that Ends Well*, *Timon*, and *Macbeth* all had contributions from Middleton; that *Pericles* had

---

<sup>15</sup> I will not list the extensive scholarship of these researchers, but for representative examples see *Shakespeare, Computers, and the Mystery of Authorship*, ed. by Hugh Craig and Arthur Kinney (Cambridge: Cambridge University Press, 2009); Brian Vickers, 'Shakespeare and Authorship Studies in the Twenty-First Century', *Shakespeare Quarterly* 62.1 (2011), 106-42; MacDonald P. Jackson, 'Authorship and the Evidence of Stylometrics', in *Shakespeare beyond Doubt: Evidence, Argument, Controversy*, ed. by Paul Edmondson (Cambridge: Cambridge University Press, 2013), pp. 100-10; and Santiago Segarra, Mark Eisen, Gabriel Egan, and Alejandro Ribeiro, 'Attributing the Authorship of the *Henry VI* Plays by Word Adjacency', *Shakespeare Quarterly* 67.2 (2016), 232-56.

<sup>16</sup> See Donald W. Foster, 'A *Funeral Elegy*: W[illiam] S[hakespeare]'s "Best-speaking witnesses"', *PMLA* 111 (1996), 1080-1105; G.D. Monsarrat, 'A *Funeral Elegy*: Ford, W.S., and Shakespeare', *The Review of English Studies* 53.210 (2002), 186-203; William S. Niederkorn, 'A Scholar Recants on His "Shakespeare" Discovery', *New York Times* (June 20, 2002): B1 and B5.

<sup>17</sup> See Matthew Steggle, "'Knowledge will be multiplied": Digital Literary Studies and Early Modern Literature', in *A Companion to Digital Literary Studies*, ed. by Susan Scheibman and Ray Siemens (Oxford: Blackwell, 2008), pp. 98-101.

<sup>18</sup> See *The New Oxford Shakespeare: Authorship Companion*, ed. by Gary Taylor and Gabriel Egan (Oxford: Oxford University Press, 2017); Dalya Alberge, 'Christopher Marlowe Credited as One of Shakespeare's Co-Writers', *The Guardian* (Oct. 23, 2016),

<https://www.theguardian.com/culture/2016/oct/23/christopher-marlowe-credited-as-one-of-shakespeares-co-writers>.

contributions from Wilkins; and that Shakespeare and Fletcher collaborated on *Henry VIII*, *The Two Noble Kinsmen*, and the lost play *Cardenio* (again, debate on SHAKSPER ensued).

While affirming the importance of computer-aided attribution studies, I want to point out that the questions such studies ask and answer are historical questions, not literary questions. Up to this point, statistical analysis has largely left untouched questions about how a text works, how it conveys meaning, and how it creates significance. These are literary questions, and they are the kinds of questions students tend to care about – and, I would add, they are the kinds of questions I care about, identifying more as a Shakespearean critic than a Shakespearean scholar.

Beyond questions of authorship and attribution, however, critics such as Hugh Craig have mobilized statistics to ask compelling literary questions. In 2000, Craig gleefully used data processing to argue against the post-structural theorists proclaiming ‘the death of the author’: ‘In an analysis of affinities between 100 plays by various authors from the Shakespearean period, based on frequencies of very common words, authorship emerged as distinctly more important than genre or date in grouping plays’.<sup>19</sup> Also in 2000, Craig suggested that increased modal verbs like *can*, *may*, and *should* in later Renaissance drama indicate an increased interest in human subjectivity and thus ‘are a crude but measurable indicator of the shift from early to modern’.<sup>20</sup> Then, in a 2011 essay, Craig countered the persistent myth that Shakespeare had a preternaturally large vocabulary by using statistics to show that ‘Shakespeare introduce[d] “fresh” words – that is, words he has not used before – at about the same rate as his contemporaries’.<sup>21</sup> We only believe Shakespeare’s vocabulary was massive, Craig suggested, because he produced many more works than his contemporaries. And Craig’s most recent book, co-authored with Brett Greatley-Hirsch, *Style, Computers, and Early Modern Drama: Beyond Authorship* (2017), is explicitly concerned with what I have called ‘literary’ rather than ‘historical’ questions.<sup>22</sup>

---

<sup>19</sup> Hugh Craig, ‘Is the Author Really Dead? An Empirical Study of Authorship in English Renaissance Drama’, *Empirical Studies of the Arts* 18.2 (2000), 119-34 (p. 119). Craig used a comparison between Shakespeare and Fletcher to further support this claim in ‘Style, Statistics and New Models of Authorship’, *Early Modern Literary Studies* 15.1 (2010), <http://purl.oclc.org/emls/15-1/craistyl.htm>.

<sup>20</sup> Hugh Craig, ‘Grammatical Modality in English Plays from the 1580s to the 1640s’, *English Literary Renaissance* 30.1 (2000), 32-54 (p. 33).

<sup>21</sup> Hugh Craig, ‘Shakespeare’s Vocabulary: Myth and Reality’, *Shakespeare Quarterly* 62.1 (2011), 53-74 (p. 58).

<sup>22</sup> See Hugh Craig and Brett Greatley-Hirsch, *Style, Computers, and Early Modern Drama: Beyond Authorship* (Cambridge: Cambridge University Press, 2017).

To be sure, authorship studies can serve interpretation, as John Jowett illustrated in 2003 when using a statistical analysis of the scenes in *Timon* credited to Shakespeare and Middleton to reflect on the relationship between economic themes in the play and Middleton's literary indebtedness to Shakespeare.<sup>23</sup> Jonathan Hope and Michael Witmore also used computer reading to ask specifically literary questions in one 2004 study.<sup>24</sup> Using a word analysis program called DocuScope, developed at Carnegie Mellon University, Hope and Witmore discovered that Shakespearean genres are identifiable, not just according to content and tone (the usual, qualitative basis of generic classification, but not things computers can detect), but also according to language. When Shakespeare's plays were processed, Docuscope 'divided the plays almost exactly according to the folio genres of History (Group 1) and Comedy (Group 2)',<sup>25</sup> as Hope and Witmore represented in Figure 1.<sup>26</sup> The authors noted and discussed the anomalies (*Comedy of Errors*, *A Midsummer Night's Dream*, and *The Tempest* were grouped with the histories, *Henry VIII* with the comedies), and observed that Docuscope was not able to place the tragedies, leading Hope and Witmore to remark on 'the "inbetween" status of the Tragedies'.<sup>27</sup> How did Docuscope make its classifications? Among other factors, it counted kinds of speech, as determined by the linguistic forms common in certain kinds of discourse (e.g., a lot of imperative verbs ['Syracusan, say in brief what cause'] indicate social interaction, while a density of first-person pronouns ['I'll utter what my sorrows give me leave'] indicate inner thinking, which can, of course, still occur in moments of social interaction). It was specifically because linguistic word forms were the data used by Docuscope that Hope and Witmore were surprised by the results:

Comedies have significantly more first person forms than either Histories or Tragedies. Perhaps the difference between Comedies and Histories on this feature is not surprising and could have been predicted, but it seems to us that the difference between the Comedies and Tragedies is counter-intuitive: Comedy is supposed to be the genre of society; Tragedy of the individual. Soliloquy might have been expected to boost the frequency of first person in the

---

<sup>23</sup> See John Jowett, 'Middleton and Debt in *Timon of Athens*', in *Money and the Age of Shakespeare: Essays in New Economic Criticism*, ed. Linda Woodbridge (New York: Palgrave Macmillan, 2003), pp. 219-36.

<sup>24</sup> Jonathan Hope and Michael Witmore, 'The Very Large Textual Object: A Prosthetic Reading of Shakespeare', *Early Modern Literary Studies* 9.3 (2004), 6.1-36.

<sup>25</sup> Ibid, 22.

<sup>26</sup> In a nice moment of replication, the principle opposition of comedy and history (as opposed to comedy and tragedy) was confirmed using a different methodology by Hugh Craig, "'Speak, that I may see thee": Shakespeare Characters and Common Words', *Shakespeare Survey* 61 (2008), 281-8.

<sup>27</sup> Hope and Witmore, 24.



Tragedies. If we see Early Modern plays as one of the places where the modern ‘self’ is constructed, perhaps we have been looking in the wrong place with our concentration on *Hamlet*: should we turn instead to *Comedy of Errors* and *The Two Gentlemen of Verona*?<sup>28</sup>

As Hope and Witmore’s study illustrated, ‘computer reading can surprise us, and perhaps prompt us to ask different questions about Shakespeare’s texts’, which they continued to illustrate with additional DocuScope readings of Shakespeare.<sup>29</sup>

Group 1	Group 2
King Richard II	Taming of the Shrew
3 Henry VI	All’s Well That Ends Well
King John	Two Gentlemen of Verona
Henry V	Merchant of Venice
1 Henry IV	Twelfth Night
2 Henry IV	As You Like It
1 Henry VI	Winter’s Tale
2 Henry VI	Love’s Labour’s Lost
Richard III	King Henry VIII
Comedy of Errors	Much Ado About Nothing
Midsummer’s Night Dream	Measure for Measure
Tempest	

Figure 1: The DocuScope sorting of Shakespeare’s Plays into Group 1 (Histories) and Group 2 (Comedies), in Hope and Witmore, ‘The Very Large Textual Object’. Image from *Early Modern Literary Studies*.

Perhaps the most well-known computer-aided essay about Shakespeare, Moretti’s ‘Network Theory, Plot Analysis’ (2011), also took aim at a question of real literary meaning. Moretti began by citing ‘culturonomics’ as evidence of the rise of quantitative

<sup>28</sup> Ibid, 28.

<sup>29</sup> Ibid, 9. See also Jonathan Hope and Michael Witmore, ‘Shakespeare by the Numbers: On the Linguistic Texture of the Late Plays’, in *Early Modern Tragicomedy*, ed. by Subha Mukherji and Raphael Lyne (Woodbridge: D.S. Brewer, 2007), pp. 133-53; Jonathan Hope and Michael Witmore, ‘Quantification and the Language of Later Shakespeare’, *Société Française Shakespeare Actes du Congrès* 31 (2014), 123-49; Michael Witmore, Hope Jonathan, and Gleicher Mike, ‘Digital Approaches to the Language of Shakespearean Tragedy’, in *The Oxford Handbook of Shakespearean Tragedy*, ed. by Michael Neill and David Schalkwyk (Oxford: Oxford University Press, 2016), pp. 316-335; and Jonathan Hope and Michael Witmore, ‘Books in Space: Adjacency, EEBO-TCP, and Early Modern Dramatists’, in *Early Modern Studies after the Digital Turn*, ed. by Laura Estill, Diane K. Jakacki, and Michael Ulliot (Tempe: Arizona Center for Medieval & Renaissance Studies, 2016), pp. 9-34.

studies in the humanities, but he then went in a different direction. Instead of using culturomics to track words over time, Moretti sought to apply network theory (the use of graphs to represent complex systems) to the relationships among characters in Shakespeare's *Hamlet* (see Figure 2).<sup>30</sup> Curiously, the computer-generated visualizations of the character networks in *Hamlet* were, Moretti wrote, too 'clumsy', so 'the networks in this study were all made by hand'.<sup>31</sup> Are we not back in the realm of old-fashioned qualitative analysis? Moretti's essay is digital humanities at its worst: a façade of electronic gimmickry pushing aside both rigorous qualitative analysis and replicable quantitative analysis.

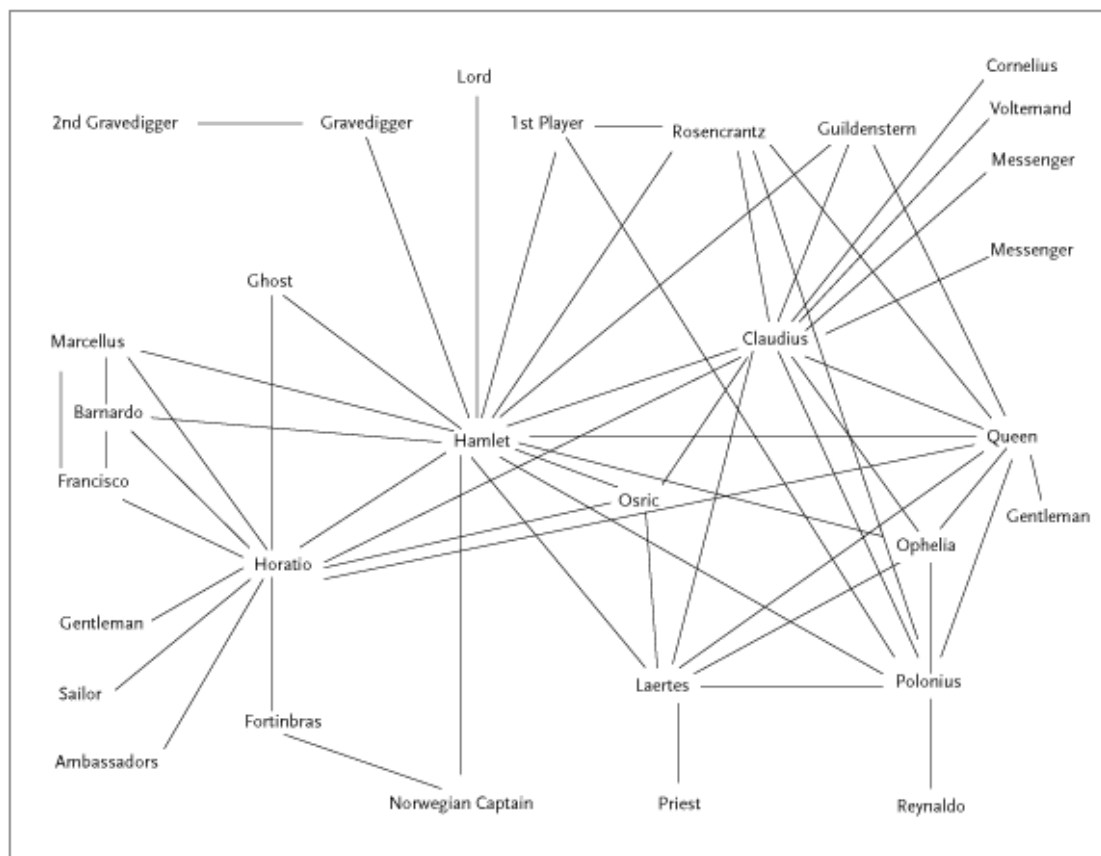


Figure 2: Franco Moretti, 'The *Hamlet* Network', in 'Network Theory, Plot Analysis', 81. Image from *New Left Review*.

Based on characters who speak to each other, Moretti identified two clusters of characters in *Hamlet* – one cluster of the Danish court insiders grouped around Claudius, Gertrude, and Ophelia, another of the outsiders involving Horatio, Fortinbras, and the Ghost – all tied together by Hamlet. Take away Hamlet, and the two groups of characters split with almost no connections between the court insiders and the outsiders. Go one step further and take Claudius away from the insiders and that group remains

<sup>30</sup> Franco Moretti, 'Network Theory, Plot Analysis', *New Left Review* 68 (2011), 80-102 (p. 80).

<sup>31</sup> *Ibid*, 82.

intact: the members of the Danish court all have relationships with all the other members, so that network survives. But take away Horatio from the outsiders and that group falls apart: Horatio is the only thing tying together a disparate group of autonomous individuals. Thus, 'Horatio's space – ambassadors, messengers, sentinels, talk of foreign wars, and of course the transfer of sovereignty at the end – all this announces what will soon be called, not Court, but State'.<sup>32</sup>

Moretti's study does not reveal anything new to someone with a basic understanding of Shakespeare's play.<sup>33</sup> It may have revealed something new to Moretti, and for this reason his energetic form of pseudo-scientific literary analysis can be valuable as an interpretive tool for students in the classroom. But consider Moretti's claim that 'we would never think of discussing *Hamlet* – without Hamlet'.<sup>34</sup> That statement magnificently ignores one of the best Shakespeare books published in the last few decades, Margreta de Grazia's *'Hamlet' without Hamlet* (2007).<sup>35</sup> What Moretti presented as his digitized epiphany – that *Hamlet* is really about international relations – de Grazia presented three years earlier in both textually grounded and historically specific terms. I suspect Moretti's reading of *Hamlet* was so cursory because he was trying to use Shakespeare to illustrate his approach to literature, when what we really need is a demonstration of the way that his approach to literature can illuminate Shakespeare. For Shakespeareans, the kinds of questions that matter are the kind Moretti shrugged off in a footnote after observing that the legitimate rulers (Old King Hamlet and Fortinbras) are in the outsider group with little connection to the insider group and its illegitimate ruler (Claudius): 'Why the balance is not there – why choose a ghost and a Norwegian as figures of legitimacy – is a different question, on which network theory has probably nothing to say'.<sup>36</sup> That is a literary question pertaining to meaning and significance, and it is not a question Moretti's computer can answer. From where I stand, it was a missed opportunity for the human behind the computer to become a humanist and think through this issue in humanistic as opposed to scientific terms.

---

<sup>32</sup> Ibid, 92.

<sup>33</sup> The obviousness of Moretti's observations is seen as a virtue, rather than a shortcoming, by Lisa Rhody, 'A Method to the Model: Responding to Franco Moretti's Network Theory, Plot Analysis', *Magazine Modernisms* (Aug. 22, 2011), <https://magmods.wordpress.com/2011/08/22/a-method-to-the-model-responding-to-franco-moretti-s-network-theory-plot-analysis>. Rhody claims, 'The paper doesn't argue that we should be surprised; instead, this information acts as a type of control'.

<sup>34</sup> Moretti, 86.

<sup>35</sup> See Margreta de Grazia, *Hamlet without Hamlet* (Cambridge: Cambridge University Press, 2007).

<sup>36</sup> Moretti, 88 n. 6.

That is what I sought to do in an essay titled ‘Is *Hamlet* a Sexist Text? Overt Misogyny vs. Unconscious Bias’.<sup>37</sup> I began with the observation that ‘although roughly half of the human population is made up of women, they make up roughly 7 percent of the characters in *Hamlet* and speak roughly 8 percent of the lines in the play’. These data points were confirmed with additional evidence both quantitative and qualitative:

Claudius has 552 lines in the play, Gertrude (the analogous female character) only 157. Horatio has 294 lines, Ophelia (the analogous female character) only 173. Looking at substantive speeches is even more telling: Claudius has 47 of his 102 speeches that run for three or more lines (46 percent), Gertrude only 16 of her 69 (23 percent). Indeed, when Gertrude does speak, it is often for one-line affirmations of things male characters have already said: ‘Ay, amen’, ‘It may be, very like’, ‘So he does indeed’. As for Ophelia, her one-liners are so peppered with ‘my lord’ (which she says more than half the times she speaks: 30 out of 58 speeches) that they are only half-liners.

But the delivery of statistics is not – cannot be – the end of a literary study. Statistics are merely the occasion for reflection and eventually argument:

*Hamlet* is not misogynistic in the sense that it promotes the superiority of men and the inferiority of women. In fact, *Hamlet* critiques misogyny and patriarchy by configuring them with tragedy, yet the Shakespeare who wrote *Hamlet* still held an unconscious bias against women. In other words, *Hamlet* exhibits a structural sexism that is different from and more difficult to discern than the overt sexism of misogyny and patriarchy. *Hamlet* is therefore a powerful literary example of the way that, even when someone is trying to be ethically progressive, unconscious bias can remain.

I am buoyed, not by the whiz-bang technobabble of Moretti and the digital revolutionaries, who want computers to do our interpretation for us, but by more grounded opportunities for computationally literate close readings that use quantitative analysis as a springboard for qualitative investigation. I have in mind projects like ‘The New Variorum Shakespeare Digital Challenge’, which provides XML files and schema for Shakespeareans to search, study, and manipulate.<sup>38</sup> Or Cyrus Mulready’s assignment

---

<sup>37</sup> Jeffrey R. Wilson, ‘Is *Hamlet* a Sexist Text? Overt Misogyny vs. Unconscious Bias’, unpublished essay.

<sup>38</sup> See the MLA Committee on the New Variorum Edition of Shakespeare, ‘The New Variorum Shakespeare Digital Challenge’, *MLA.org* (2012-16), <https://www.mla.org/About-Shakespeare-Digital-Challenge>,

for his students, ‘Close and Distant Reading’, which asks for an interpretation of one of Shakespeare’s sonnets that exploits both traditional literary exegesis and innovative data-mining.<sup>39</sup> Or Marcus Nordlund’s recent book, *The Shakespearean Inside* (2017), a study of Shakespeare’s soliloquies and asides which asked, as I have, ‘What if we could produce new data of real literary interest about Shakespeare’s complete plays?’<sup>40</sup>

#### IV. Reception-Based Shakestats

Shifting attention from Shakespeare’s text to its reception, statistical studies have sought to quantify Shakespearean publication, performance, education, and scholarship. Looking at early print culture, for example, Lukas Erne’s *Shakespeare and the Book Trade* (2013) tallied editions of the most popular early-modern dramatists to reveal that ‘Shakespeare, with seventy-four editions, out-publishe[d] all his contemporaries by more than 50 percent’, as illustrated in Figure 3.<sup>41</sup> Erne used this and other empirical evidence in the service of his overarching argument that Shakespeare ‘witnessed and was not indifferent towards his rise to popularity in the book trade, which clearly exceeded that of other dramatists, as he must have noticed’.<sup>42</sup>

---

Us/Governance/Committees/Committee-Listings/Publications/Committee-on-the-New-Variorum-Edition-of-Shakespeare/The-New-Variorum-Shakespeare-Digital-Challenge.

<sup>39</sup> Cyrus Mulready, ‘Close and Distant Reading Assignment’,

<https://cyrusmulready.wordpress.com/pedagogy-resources/close-and-distant-reading-assignment/>.

<sup>40</sup> Marcus Nordlund, *The Shakespearean Inside: A Study of the Complete Soliloquies and Solo Asides* (Edinburgh: Edinburgh University Press, 2017), p. 1.

<sup>41</sup> Lukas Erne, *Shakespeare and the Book Trade* (Cambridge: Cambridge University Press, 2013), p. 37. Erne’s book was a more Shakes-centric version of the wider concerns about the popularity of Renaissance playbooks more generally. See Peter W.M. Blayney, ‘The Publication of Playbooks’, in *A New History of Early English Drama*, ed. by John D. Cox and David Scott Kastan (New York, NY: Columbia University Press, 1997), pp. 383-422; Alan B. Farmer and Zachary Lesser, ‘The Popularity of Playbooks Revisited’, *Shakespeare Quarterly* 56.1 (2005), 1-32; and Peter W.M. Blayney, ‘The Alleged Popularity of Playbooks’, *Shakespeare Quarterly* 56.1 (2005), 33-50.

<sup>42</sup> Erne, p. 10.

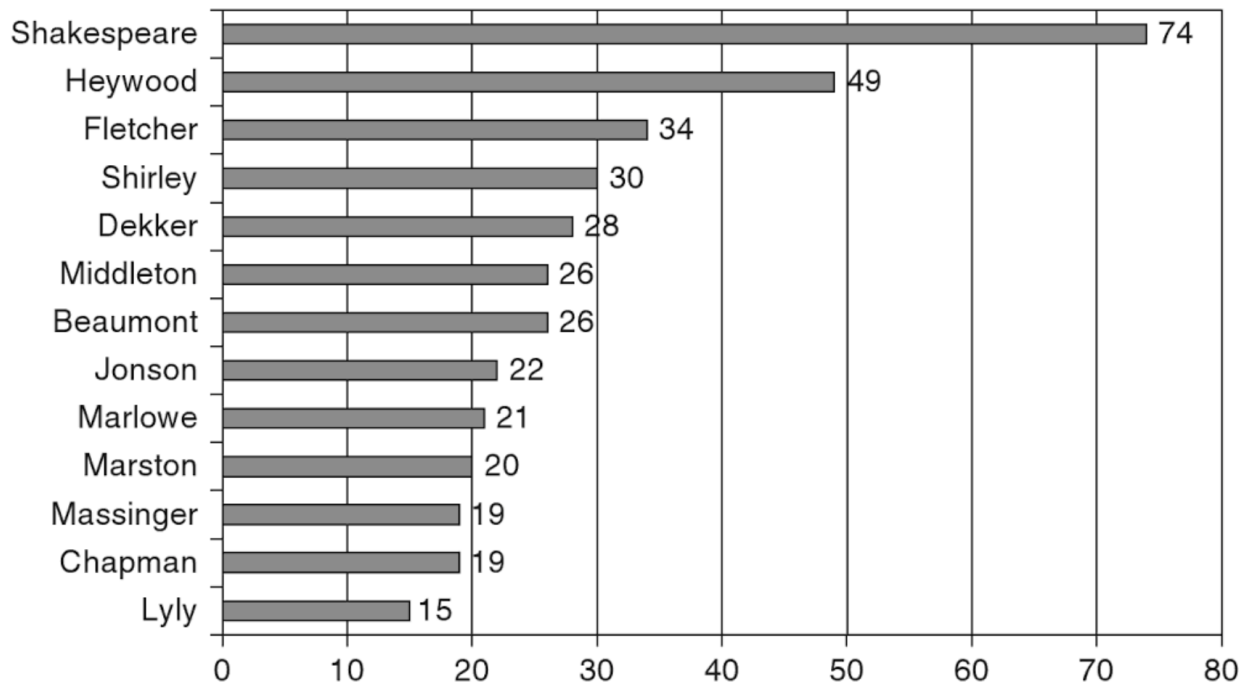
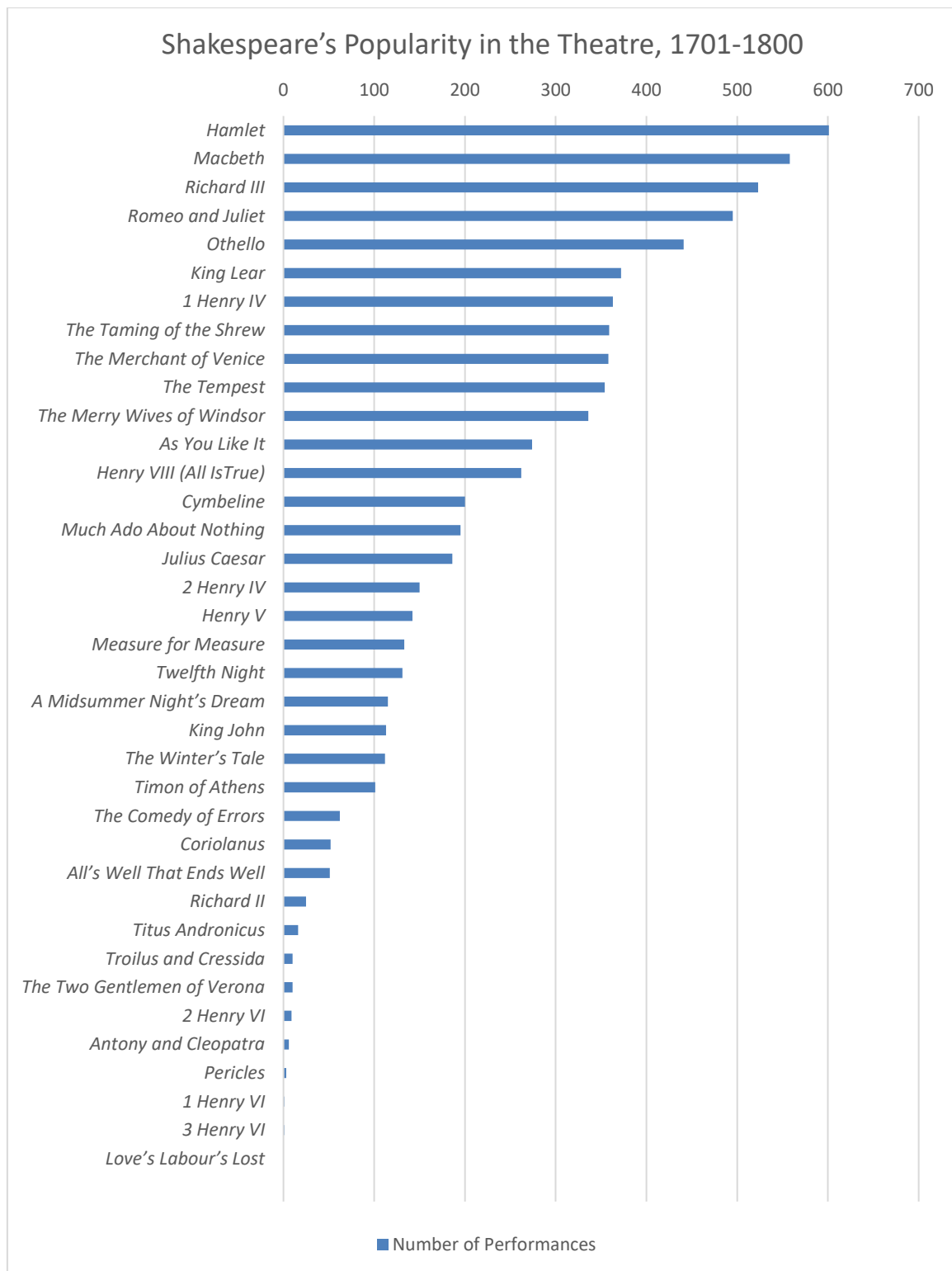


Figure 3: Lukas Erne, 'Editions of Playbooks to 1642', in *Shakespeare and the Book Trade*, 38. Image from Google Books.

Shifting from the seventeenth to the eighteenth century, and from publication to performance, Charles Hogan's pre-digital study, *Shakespeare in the Theatre: 1701-1800* (1952), tabulated the number of times each of Shakespeare's plays was acted in order to determine the relative popularity of each.<sup>43</sup> As Figure 4 shows, it was all about the tragedies in the eighteenth century: *Hamlet*, *Macbeth*, *Richard III*, *Romeo and Juliet*, *Othello*, and *King Lear* were the six most popular plays. But the histories also made a decent showing: *Richard III*, *1 Henry IV* and *Henry VIII*, for example, were ranked third, seventh, and thirteenth.

<sup>43</sup> Charles Beecher Hogan, *Shakespeare in the Theatre, 1701-1800*, vol. 2, *A Record of Performance in London, 1751-1800* (Oxford: Clarendon Press, 1952), pp. 715-19.



*Figure 4: Shakespeare's Popularity in the Theatre, 1701-1800, based on data from Charles Beecher Hogan, Shakespeare in the Theatre, 1701-1800, 715-19.*

Contrast that eighteenth-century data set with what we see when we turn to the recent past. In 2016, the journalist Dan Kopf, writing for the content tracking website *Priceonomics*, wanted to know ‘What Is Shakespeare’s Most Popular Play?’<sup>44</sup> Kopf crunched the numbers from the amateur Shakespeare enthusiast Eric Minton’s catalogue of nearly 2,000 Shakespearean performances from 2011-16, 75 percent of which were in the United States (‘according to our research, his list of productions is the most comprehensive available’).<sup>45</sup> Figure 5 shows that *A Midsummer Night’s Dream* was the ‘revealed preference’ of these twenty-first century audiences, followed in the top five by *Romeo and Juliet*, *Twelfth Night*, *Hamlet*, and *Taming of the Shrew*. The rankings of *Richard III*, *1 Henry IV*, and *Henry VIII* dropped way down to fourteenth, twenty-seventh, and thirty-fifth. Minton’s data also included the location of each production, which allowed Kopf to compare the most popular Shakespeare plays in the U.S. to those abroad. He found that *Romeo and Juliet* and *Twelfth Night* were disproportionately popular in the U.S., while *King Lear* and *The Comedy of Errors* were disproportionately unpopular (see Figure 6).

---

<sup>44</sup> Dan Kopf, ‘What Is Shakespeare’s Most Popular Play?’ *Priceonomics* (Sept. 22, 2016), <https://priceonomics.com/what-is-shakespeares-most-popular-play/>.

<sup>45</sup> It is worth noting that the World Shakespeare Bibliography (which has better global coverage than Kopf) lists 3,390 productions.



## Which Shakespeare Play Is Performed Most?

Based on Notable Professional Performances Documented on Shakespeareances Since 2011

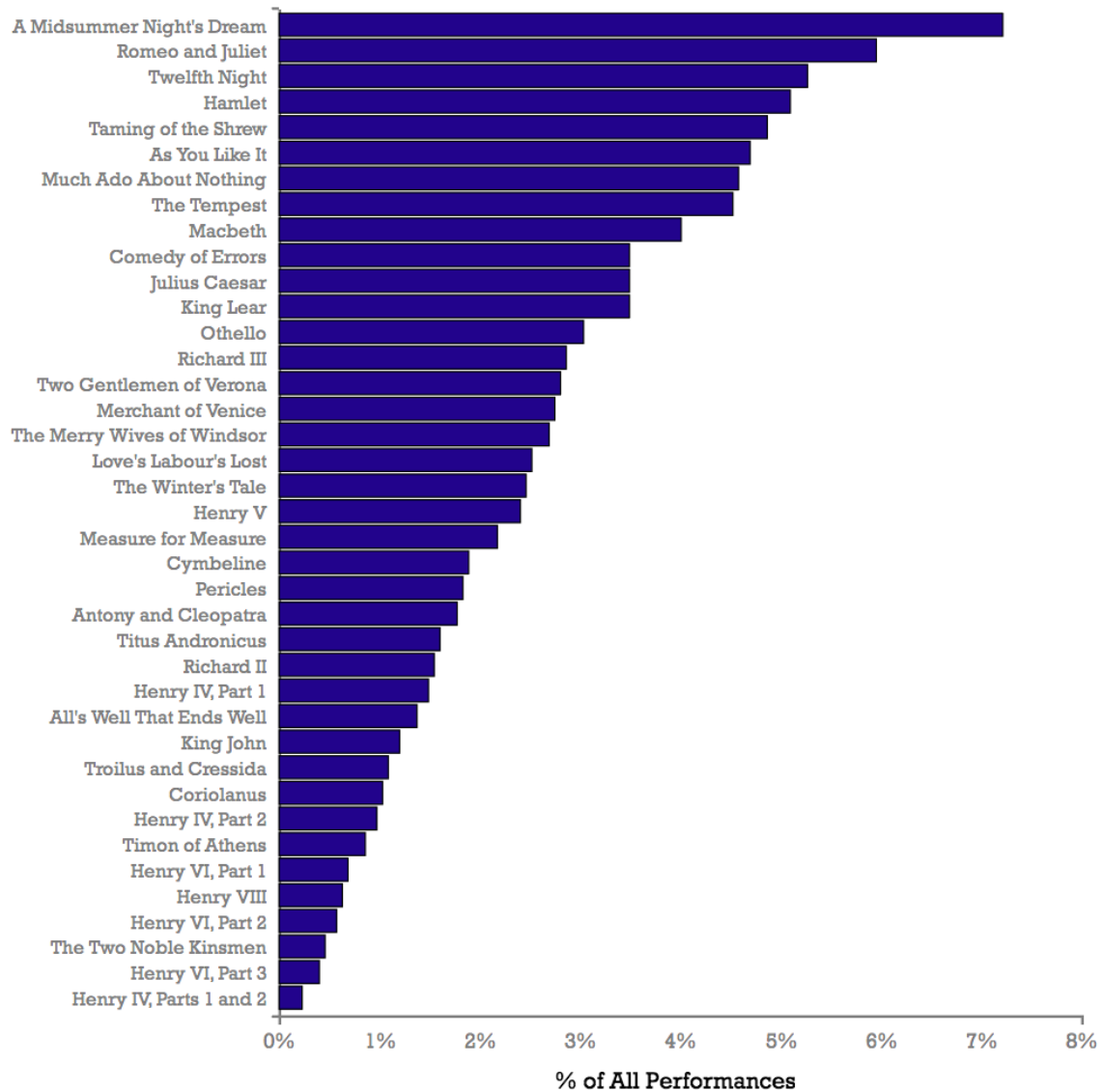


Figure 5: Dan Kopf, 'Which Shakespeare Play Is Performed Most?' in 'What Is Shakespeare's Most Popular Play?' (priceconomics.com).

## Most Performed Shakespeare Plays: U.S. Versus the World

Based on Notable Professional Performances Documented on Shakespeareances Since 2006 Shakespeare Plays

U.S.			Non U.S.		
Rank	Play	% of All Performances	Rank	Play	% of All Performances
1	A Midsummer Night's Dream	7.1%	1	A Midsummer Night's Dream	7.7%
2	Romeo and Juliet	6.7%	2	Hamlet	5.9%
3	Twelfth Night	5.8%	3	King Lear	5.9%
4	Hamlet	4.9%	4	As You Like It	4.8%
5	Much Ado About Nothing	4.9%	5	Taming of the Shrew	4.8%
6	Taming of the Shrew	4.9%	6	Comedy of Errors	4.5%
7	As You Like It	4.7%	7	The Tempest	4.0%
8	The Tempest	4.7%	8	Othello	3.7%
9	Macbeth	4.2%	9	Macbeth	3.5%
10	Julius Caesar	3.6%	10	Much Ado About Nothing	3.5%

Figure 6: Dan Kopf, 'Most performed Shakespeare Plays: U.S. Versus the World' in 'What Is Shakespeare's Most Popular Play?' (priceconomics.com).

Turning from the stage to the classroom, we have seen a number of empirical reports about Shakespeare in the curriculum. In 2006, Neill Thew of the Higher Education Academy's English Subject Centre produced a report, *Teaching Shakespeare: A Survey of the Undergraduate Level in Higher Education*, which surveyed 51 British institutions to determine how Shakespeare was being taught.<sup>46</sup> Although '89% of respondents consider[ed] their students at best adequately and often poorly prepared for their studies', the report found that Shakespeare was ubiquitously taught: 'Fully 73% of respondents run one or more compulsory courses including significant study of Shakespeare at Level 1. 80% of respondents offer optional courses devoted to Shakespeare at Levels 2+' (3). According to Figure 7, the most common works taught were *Hamlet*, *The Tempest*, *The Sonnets*, *Twelfth Night*, and *A Midsummer Night's Dream*.

<sup>46</sup> See Neill Thew, *Teaching Shakespeare: A Survey of the Undergraduate Level in Higher Education* (York: Higher Education Academy, 2006).

	Whole text: Level 1	Extract: Level 1	Whole text: Levels 2+	Extract: Levels 2+	Respondent total
Hamlet	8	0	15	4	23
The Tempest	5	0	17	1	22
Poetry: the Sonnets	5	5	9	2	20
Twelfth Night	7	1	13	2	19
Midsummer Night's Dream	4	0	11	4	18
Othello	4	1	12	2	16
King Lear	4	0	12	2	16
Henry V	3	0	12	1	16
Measure for Measure	3	0	11	3	16
Winter's Tale	5	0	11	0	15
Macbeth	4	0	9	3	14
Henry IV, Part 1	5	0	7	2	14
As You Like It	5	0	7	1	13
Merchant of Venice	5	0	8	0	12
Richard II	4	1	8	1	12
Richard III	2	0	10	2	12
Taming of the Shrew	4	0	8	1	11
Antony and Cleopatra	2	0	7	2	10
Julius Caesar	2	0	8	0	9
Titus Andronicus	2	0	8	0	9
Coriolanus	2	0	7	1	9
Romeo and Juliet	1	0	7	1	9
Much Ado About Nothing	4	0	5	0	8
Henry IV, Part 2	1	0	5	2	8
Cymbeline	1	0	6	0	7
Poetry: other poems	3	0	4	0	7
Love's Labours Lost	1	0	5	0	6
All's Well that Ends Well	1	0	4	1	6
Troilus and Cressida	1	0	4	1	6
Comedy of Errors	2	0	4	0	5
Henry VI, Part 1	0	0	4	0	4
Two Gentlemen of Verona	0	0	4	0	4
Merry Wives of Windsor	0	0	3	1	4
Pericles, Prince of Tyre	0	0	3	1	4
Henry VI, Part 2	0	0	3	0	3
Henry VI, Part 3	0	0	3	0	3
Henry VIII	0	0	3	0	3
Timon of Athens	0	0	3	0	3
King John	0	0	2	0	2

Figure 7: Neill Thew, 'The Texts Ranked in Order of Popularity', in *Teaching Shakespeare*, 29.

In the U.S., the American Council of Trustees and Alumni (ACTA) painted a different picture. In a series of empirical reports – *The Shakespeare File: What English Majors Are Really Studying* (1996), *The Vanishing Shakespeare* (2007), and *The Unkindest Cut: Shakespeare in Exile* (2015) – the ACTA found that the number of schools

requiring a course in Shakespeare had decreased from 23 of 70 schools surveyed in 1996 (33 percent), to 15 of 70 in 2007 (21 percent), to 4 of 52 in 2015 (8 percent).<sup>47</sup> On the one hand, the resentful identity politics of the ACTA are obvious: ‘While Shakespeare and other canonical authors are no longer required, many institutions such as Rice, Vassar, and Vanderbilt go further and require students to study “non-canonical traditions”, “race, gender, sexuality, or ethnicity”, and “ethnic or non-Western literature”’.<sup>48</sup> On the other hand, the shoddy science of the ACTA is also clear: they shout doom-and-gloom for Shakespeare because the required sole-author Shakespeare course is being phased out while turning a blind eye to: (1) the inevitable place of Shakespeare in required surveys of British medieval and Renaissance literature; (2) optional single-author courses on Shakespeare; (3) the frequent inclusion of Shakespeare in the topics courses that the ACTA despises; and (4) the infusion of Shakespeare in courses across the humanities. I suspect the decline of the required Shakespeare course stems from Shakespeare’s already established ubiquity as much as the desire for cultural and intellectual diversity which the ACTA (bizarrely) spurns. In other words, Shakespeare’s massive popularity – his total infusion throughout the curriculum – is what is leading to the decline in the required single-author Shakespeare course.

Indeed, research conducted in December 2011 by the British Council and the Royal Shakespeare Company reported that ‘over half the world’s school children study Shakespeare’.<sup>49</sup> Many children in the world do not go to school, of course, and this number relies heavily on the fact that 21 million school children in China study a passage from *The Merchant of Venice* each year.<sup>50</sup> The British Council and the Royal Shakespeare Company are packaging facts to fit their preconceived motives just like the ACTA. A more responsible, less political approach was taken in a 2013 essay by Jonathan Burton, who reported surveying 400 American high school English teachers to

---

<sup>47</sup> See American Council of Trustees and Alumni, *The Shakespeare File: What English Majors Are Really Studying* (1996), [https://www.goacta.org/images/download/shakespeare\\_file.pdf](https://www.goacta.org/images/download/shakespeare_file.pdf); *The Vanishing Shakespeare* (2007), [https://www.goacta.org/images/download/vanishing\\_shakespeare.pdf](https://www.goacta.org/images/download/vanishing_shakespeare.pdf); and *The Unkindest Cut: Shakespeare in Exile* (2015), [https://www.goacta.org/images/download/The\\_Unkindest\\_Cut.pdf](https://www.goacta.org/images/download/The_Unkindest_Cut.pdf).

<sup>48</sup> ACTA, *Unkindest Cut*, 5.

<sup>49</sup> See the British Council, *Culture is Great Britain* (Dec. 2012), 17. Available at <https://www.britishcouncil.org/sites/default/files/c485-uk-2012-programme.pdf>.

<sup>50</sup> On the dubiousness of this statistic, see Paul Prescott, ‘Shakespeare and the Dream of Olympism’, in *Shakespeare on the Global Stage: Performance and Festivity in the Olympic Year*, ed. by Paul Prescott, Erin Sullivan (London: Bloomsbury, 2015), pp. 1-38 (pp. 10-11).

find a preponderance of Shakespearean tragedy.<sup>51</sup> *Romeo and Juliet* appeared in 93 percent of all ninth-grade classes; taken together, five of Shakespeare's tragedies (*Macbeth*, *Hamlet*, *Romeo and Juliet*, *Julius Caesar*, and *Othello*) accounted for 90 percent of all Shakespearean plays assigned in high schools.

One particularly important slice of the education industry, academic scholarship on Shakespeare, has recently been studied empirically in Laura Estill, Dominic Klyve, and Kate Bridal's 'A Statistical Analysis of Writing about Shakespeare, 1960–2010' (2015). Proclaiming itself to be 'the first to present quantitative evidence about directions in late twentieth-century Shakespeare studies', this essay argued that 'the relative popularity of Shakespeare plays reveals our critical preoccupations and concerns'.<sup>52</sup> Like dutiful social scientists, the authors start by discussing their methodology, emphasizing the reliability and limitations of the data about Shakespeare scholarship they collected from the *World Shakespeare Bibliography*. Analyzing those data, they illustrated that *Hamlet* is by far the most popular play among academics, accounting for about 15 percent of all Shakespeare scholarship published between 1960 and 2010 (see Figure 8). Indeed, the tragedies are the most popular Shakespearean genre in scholarship (see Figure 9) and, on a related note, the problem plays are more popular in scholarship than they are in performance: *Measure for Measure* and *Troilus and Cressida* are more likely to be written about than performed.<sup>53</sup> Earlier, statistics from Hogan and Koph illustrated the popularity of lighter comedies in performance (such as *A Midsummer Night's Dream* and *Twelfth Night*), but *The Merchant of Venice* is the most popular Shakespearean comedy in scholarship (see Figure 10). Estill, Klyve, and Bridal also did regression analyses to discuss trends in Shakespeare scholarship over time. Since the 1960s, the popularity of *Julius Caesar* has significantly declined, while the popularity of *The Tempest* has significantly risen (see Figure 11).

---

<sup>51</sup> Jonathan Burton, 'Shakespeare in Liberal Arts Education', *The Rock: Whittier College Magazine* (Fall, 2013), 42-6.

<sup>52</sup> Laura Estill, Dominic Klyve, and Kate Bridal, "'Spare your arithmetic, never count the turns': A Statistical Analysis of Writing about Shakespeare, 1960–2010", *Shakespeare Quarterly* 66.1 (2015), 1-28 (pp. 1-2).

<sup>53</sup> *Ibid*, 19.

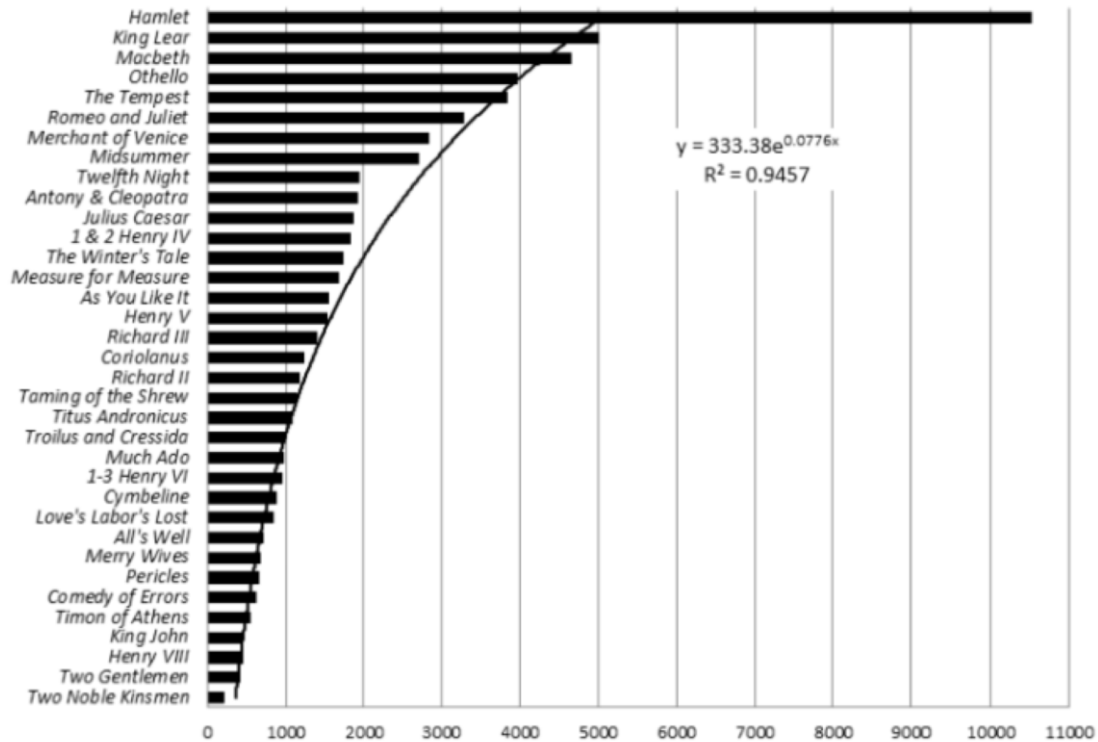


Figure 8: Laura Estill, Dominic Klyve, and Kate Bridal, 'Total Number of Publications About Each of Shakespeare's Plays (1960–2010), Sorted by the Number of Times They Have Been Written About', in 'A Statistical Analysis of Writing about Shakespeare, 1960–2010', 10.

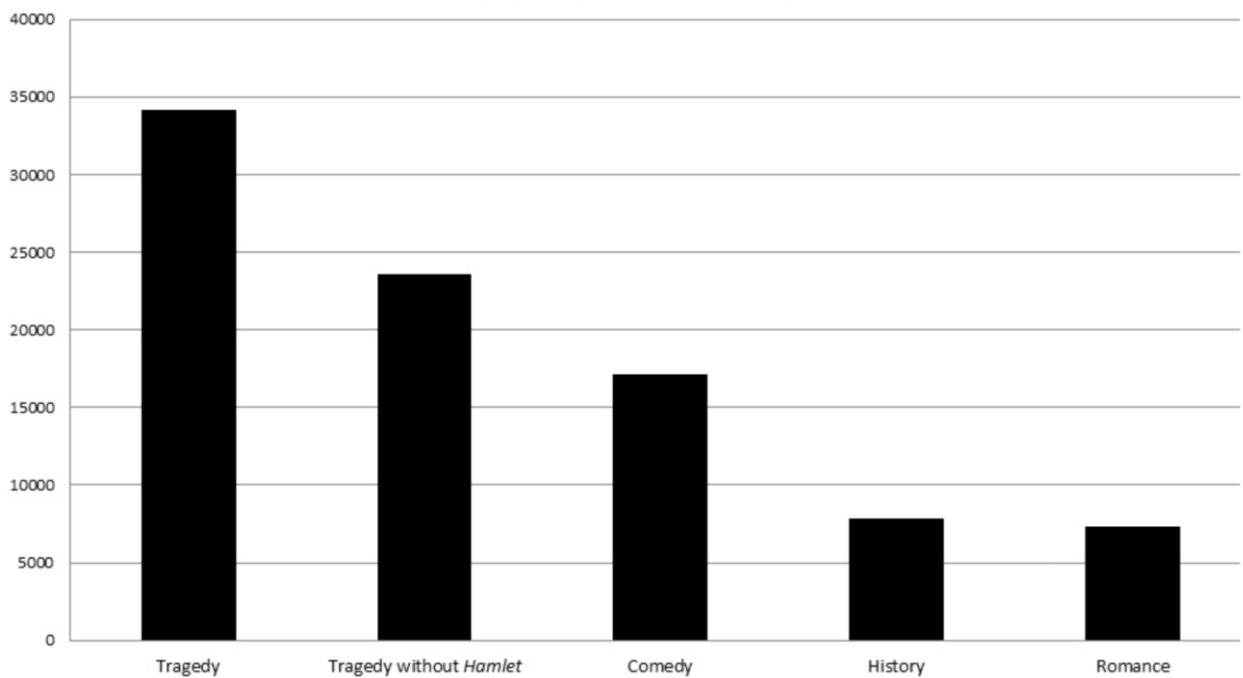


Figure 9: Laura Estill, Dominic Klyve, and Kate Bridal, 'Total Publications Sorted by Genre, Including Romance (Following *The Riverside Shakespeare*)', in 'A Statistical Analysis of Writing about Shakespeare, 1960–2010', 12.

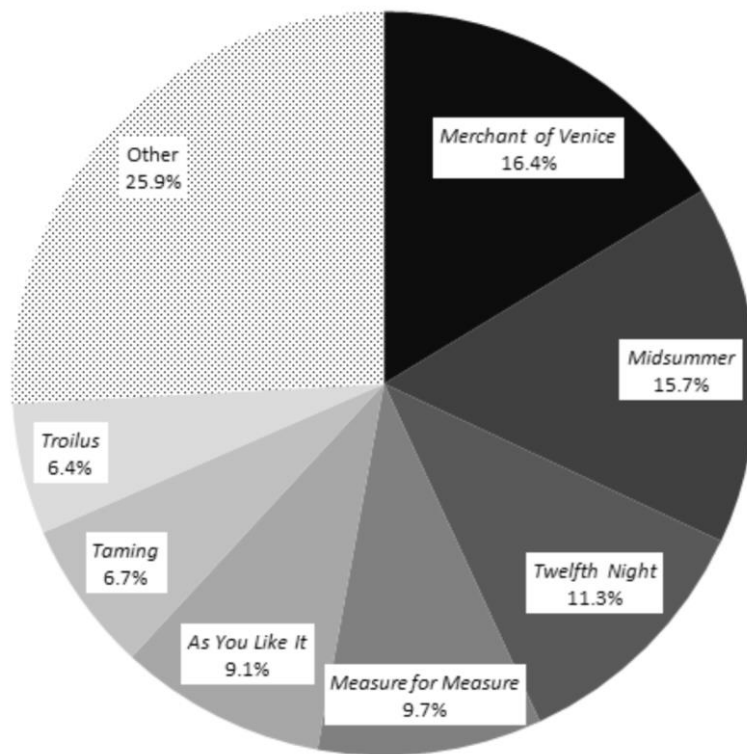


Figure 10: Laura Estill, Dominic Klyve, and Kate Bridal, 'Popularity of Comedies (As a Percentage of All Comedies), 1960-2010', in 'A Statistical Analysis of Writing about Shakespeare, 1960-2010', 17.

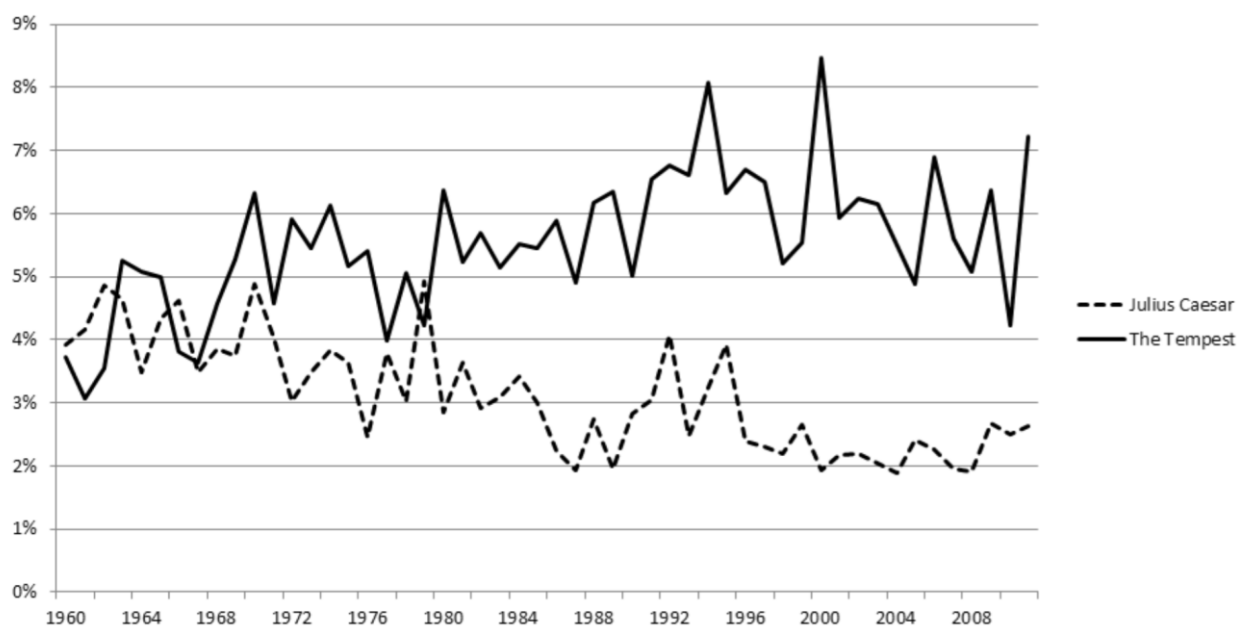


Figure 11: Laura Estill, Dominic Klyve, and Kate Bridal, 'A Comparison of the Popularity of *Julius Caesar* and *The Tempest*', in 'A Statistical Analysis of Writing about Shakespeare, 1960–2010', 25.

What I like about Estill, Klyve, and Bridal's article – that it gives to Shakespeareans a series of empirically demonstrated facts that need to be explained – is also what I dislike. I dislike the presentation of empirical data as if that were a satisfying conclusion to a literary study. This goal of disseminating facts also informs Eric Johnson's BardMetrics initiative at the Folger Shakespeare Library, 'devised to collect, store, and make available data pertaining to the publication and performance of Shakespeare in modern times'.<sup>54</sup> I am myself guilty of this *just the fact, ma'am* approach in this article. Rather than separate the collection and interpretation of data, however, I like thinking about the possibility that Shakespearean scholars can observe methods of statistical analysis and incorporate those methods into their own studies. Statistical methods can provide us with raw data about where, when, and sometimes even how Shakespeare is popular – but not why. The data need to be interpreted, as in all social sciences. Thus, empirical research is the beginning, not the end, of the new Shakespeare reception history that is underway.

## V. Questions of Fact and Questions of Meaning

Clearly, I share with David Hoover the concern that 'all too often quantitative studies fail to address problems of real literary significance'.<sup>55</sup> As Thomas Rommel wrote in his account of the digital humanities in literary studies, 'Data are collected that allow for and require further analysis and interpretation by the researcher'.<sup>56</sup> In his account of culturomics for *The Chronicle of Higher Education*, Geoffrey Nunberg agreed: 'The data don't wear their cultural significance on their sleeves; they need cultural historians to speak for them.... These new results are very often just intriguing quantitative nuggets that call out for narrative explication'.<sup>57</sup> And I think Matthew Kirschenbaum has identified a sensible way forward: 'Data mining and machine learning are best

---

<sup>54</sup> See 'BardMetrics', <http://folgerpedia.folger.edu/BardMetrics>.

<sup>55</sup> David L. Hoover, 'Quantitative Analysis and Literary Studies', in *A Companion to Digital Literary Studies*, ed. Susan Schreibman and Ray Siemens (Oxford: Blackwell, 2008),

<sup>56</sup> Thomas Rommel, 'Literary Studies', in *A Companion to Digital Humanities*, ed. by Susan Schreibman, Ray Siemens, Ray, and John Unsworth (Oxford: Blackwell, 2004), 94.

<sup>57</sup> Geoffrey Nunberg, 'Counting on Google Books', *Chronicle of Higher Education* (Dec. 16, 2010), <http://www.chronicle.com/article/Counting-on-Google-Books/125735/>.



understood in terms of “provocation” – the potential for outlier results to surprise a reader into attending to some aspect of a text not previously deemed significant’.<sup>58</sup> Thus, whether looking at Shakespeare’s texts or their receptions, the best use of statistics in Shakespeare studies, both in the classroom and in scholarly criticism, is to employ them at the beginning of the research process as a way to identify information in need of interpretation: quantitative data can help Shakespeareans ask questions they didn’t know needed to be asked. The answers to those questions, though, will probably come from traditional, qualitative, humanities-style interpretation.

## VI. Shakestats in the Classroom

This attitude toward quantitative data informs an assignment I’ve designed for students in my *Why Shakespeare?* course at Harvard University. In this course, I invite Shakespeare lovers and haters alike to consider the question of Shakespeare’s massive popularity by looking into the relationship between his methods of artistic creation and the values of the modern world. Why is Shakespeare the most celebrated writer in England, the West, and indeed the whole world? For their final paper, students are asked to identify a modern culture, group, or setting in which Shakespeare is (or isn’t) popular. The driving question is always *why*. Why is Shakespeare prominent (or not) in the specified modern instance? What does (or doesn’t) Shakespeare offer to the culture in question? What are the culture’s assumptions, desires, and commitments that bring it to Shakespeare (or not)? To answer these questions, students consider three fields of evidence: they do a literary study of one or more of Shakespeare’s original plays, a cultural study of the modern culture in question, and a historical study of the modern reception of Shakespeare in that culture (in criticism, productions, allusions, adaptations, appropriations, etc.). Thus, we combine literary studies with cultural studies to do reception history.

But this all begins with an empirical demonstration of a Shakespearean phenomenon in the modern age. Students mine one of several databases – such as Early English Books Online (EEBO), the Short Title Catalog, Eighteenth Century Collections Online (ECCO), Google Books, the MLA International Bibliography, the World Shakespeare Bibliography, a certain disciplinary database, or a certain newspaper – in an effort to generate some quantitative nugget of truth that calls out for qualitative explanation.

---

<sup>58</sup> Matthew G. Kirschenbaum, ‘The Remaking of Reading: Data Mining and the Digital Humanities’, in *The National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation* (Baltimore, MD: Oct. 11, 2007), <http://www.csee.umbc.edu/~hillol/NGDM07/abstracts/talks/MKirschenbaum.pdf>.

Using Estill, Klyve, and Bridal's 'A Statistical Analysis of Writing about Shakespeare' as a model, I introduce the assignment in class by reviewing the differences between argumentative research papers written in the humanities and empirical research papers published in the sciences. Argumentative essays in the humanities are often about new interpretations of old texts that are already widely known, I point out, while empirical research in the sciences is often about the presentation of new data that isn't already known (usually with some interpretation added on top). This is an overly simplistic distinction, but argumentative essays are mostly about persuasion (thus they have a thesis that is argued), while empirical papers are largely about discovery (thus they have a hypothesis that is tested). These different foci result in not only different citational styles (often MLA or Chicago style in the humanities and APA style in the social and natural sciences), but also different organizational structures (argumentative essays are usually organized according to *introduction*, *body*, and *conclusion*, while empirical papers usually follow the structure of *introduction*, *method*, *results*, and *discussion*).

I then have some fun quizzing the class on Shakespeare trivia, which is really an opportunity to review some databases we can use to generate empirical data about Shakespeare's reception, and to discuss some methodologies for doing so. I start with an easy one: *Which writer in the Western canon receives the most scholarly attention?* The answer, of course, is Shakespeare, but it is astounding just how much more attention Shakespeare receives than other canonical Western writers. As Figure 12 shows, he is almost four times more popular in scholarship than the next most popular western writer, Dante, who is followed in order by Joyce, Chaucer, Milton, Dickens, Faulkner, Beckett, Woolf, Proust, Hemingway, Spenser, and Dostoevsky. These data were retrieved using the *MLA International Bibliography*. I make a point of showing that searches for 'Dante' also returned results related to the nineteenth-century English artist Dante Gabriel Rossetti, which needed to be subtracted to get an accurate tally for Dante Alighieri. It is an opportunity to talk about data integrity and the need to be aware of liabilities and limitations in data and methodology when doing empirical research.

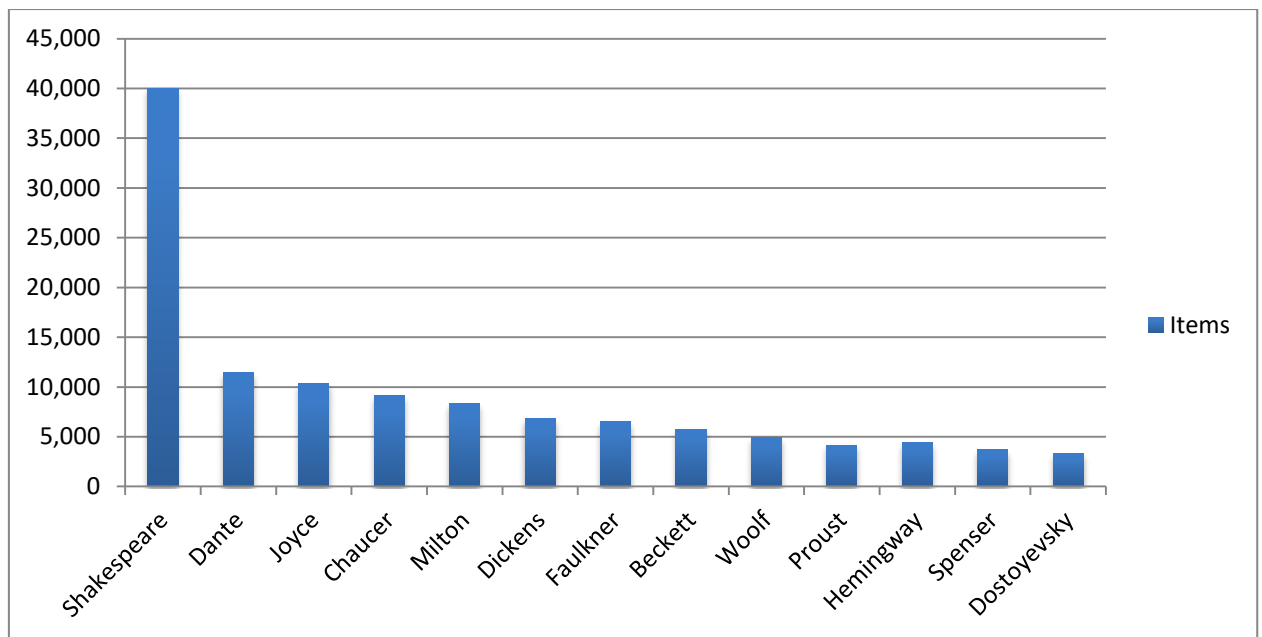


Figure 12: A Chart of Academic Pieces on Authors in the Western Canon, According to the MLA International Bibliography (as of November 10, 2015).

Who was more popular in the seventeenth century, Shakespeare or Jonson? According to the Shakespeare scholar Hugh Grady, the Restoration era displayed ‘a consensus ranking Shakespeare beneath Jonson and Fletcher’.<sup>59</sup> According to a tally of editions in EEBO, however, Shakespeare was considerably more popular than Jonson in the seventeenth century (see Figure 13). In all but one decade (the 1640s), Shakespeare had more editions published than Jonson. Again, the collection of these data is dicey, most especially because of misattributions in EEBO, which had to be accounted for in my research and which provides an opportunity to illustrate that data is only as good as the methodology used to gather it. But this empirical report could prompt questions to be pursued in a qualitative study. If Shakespeare was, in fact, more popular than Jonson during the seventeenth century, where does the claim that Jonson was more highly regarded come from? It is entirely possible that Shakespeare was *more popular* yet Jonson was *more highly regarded*. Was there a disconnection between public, book-buying audiences and the elite, book-writing critics in the seventeenth century? Was Shakespeare more popular with general audiences even though Jonson was more popular with literary critics? A further study on this question would have to consider the fact that *popularity* and *reputation* are not synonymous.

<sup>59</sup> Grady, p. 269.

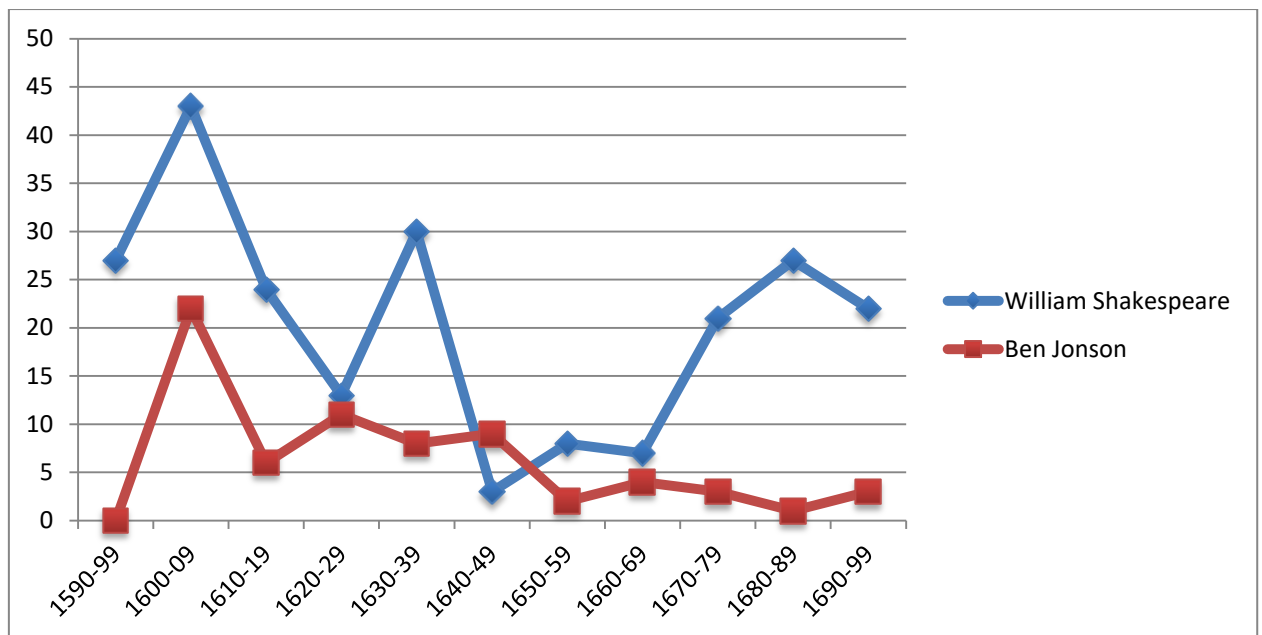
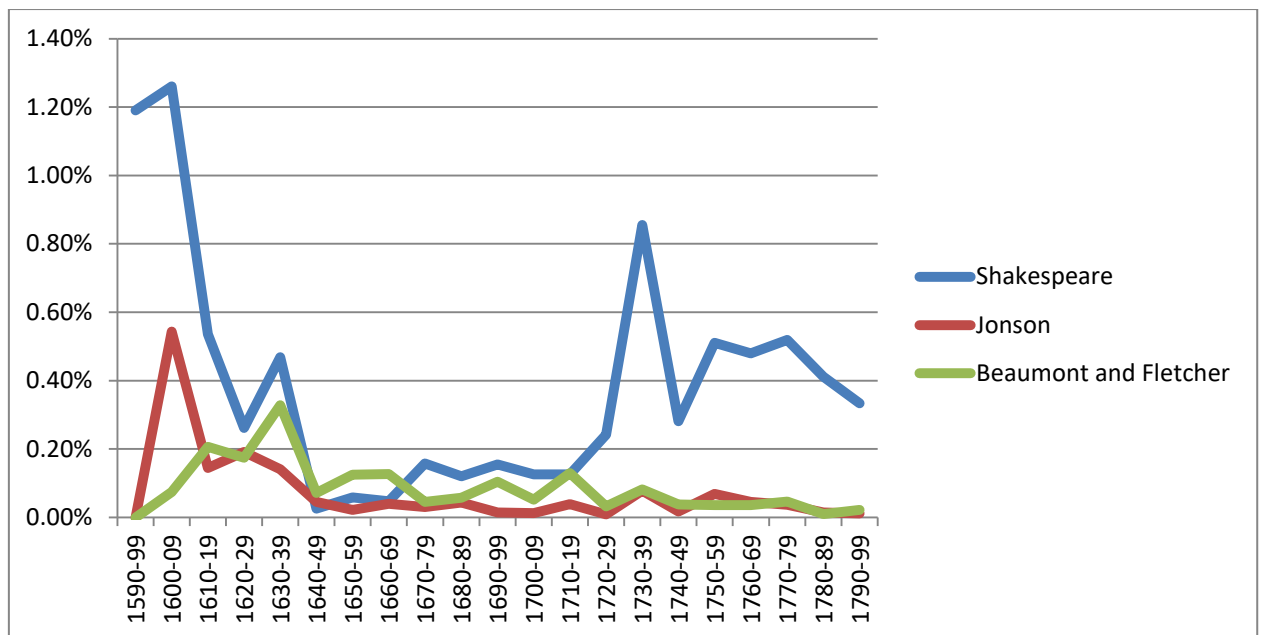


Figure 13: Number of Editions for Shakespeare and Jonson by Decade in the Seventeenth Century, According to EEBO.

*What about Beaumont and Fletcher?* An answer was sought by using the English Short Title Catalog to calculate the relative prominence of editions from Shakespeare, Jonson, Beaumont, and Fletcher (i.e., the percentage of total books published each year by each author). Represented in Figure 14, this research revealed that Shakespeare was massively popular in his own day, almost tripling the popularity of Jonson (and confirming the look at the raw numbers in Figure 13). All the dramatists took a hit after the theatres were closed in 1642, but Fletcher and Beaumont were markedly more popular during that period than Shakespeare and Jonson. But why did the less popular Shakespeare and the more popular Beaumont and Fletcher swap places between the 1660s and 1670s? Shakespeare remained slightly more popular than Beaumont and Fletcher, and markedly more popular than Jonson, until the 1720s, at which time Shakespeare experienced a remarkable surge in popularity (surely related to the publication of Nicholas Rowe's edition of Shakespeare's works in 1709, Alexander Pope's edition in 1725, and Lewis Theobald's edition in 1733).



*Figure 14: Relative Frequency of Publications (Percent of All Published Works) by Shakespeare, Jonson, and Beaumont and Fletcher According to the English Short Title Catalog.*

*In what year did Shakespeare's rise to fame begin?* If we simply type 'William Shakespeare' into Google's Ngram Viewer, we get a misleading result (see Figure 15). Shakespeare's name was not spelled consistently until around the start of the twentieth century. In an effort to account for this limitation, I included a search using the most popular spellings (*Shakespeare, Shakspeare, Shakespear, Shakspeare, Shaxper, and Shakspere*). As illustrated in Figure 16, Shakespeare was relatively unpopular in the eighteenth century, grew in popularity during the nineteenth century, and remained very popular during the twentieth century. These quantitative tidbits provide the basis for qualitative questions. Why wasn't Shakespeare very popular until at least the third quarter of the eighteenth century? Are the peaks and valleys evident in Shakespeare's general rise in popularity during the nineteenth century significant? Did things happen that led Shakespeare to rise in the 1850s, fall in the 1860s, rise back up even higher in the 1870s-80s, and then fall again in the 1890s? Given Shakespeare's decrease in popularity during WWII and increase in the post-war period, how does Shakespeare relate to war? Did all literature exhibit a similar pattern, or just Shakespeare? Why did Shakespeare's popularity steadily decline from 1950-70? Why did Shakespeare's popularity steadily increase from 1980-2000? Can we recognize patterns in the increase and decline of Shakespeare's popularity that would allow us to predict Shakespeare's popularity in the future?



Figure 15: A Google Ngram search for ‘William Shakespeare’ in the years 1700-2008 (the chart is misleading because Shakespeare’s name was not spelled consistently until around the start of the twentieth century). Image from Google.

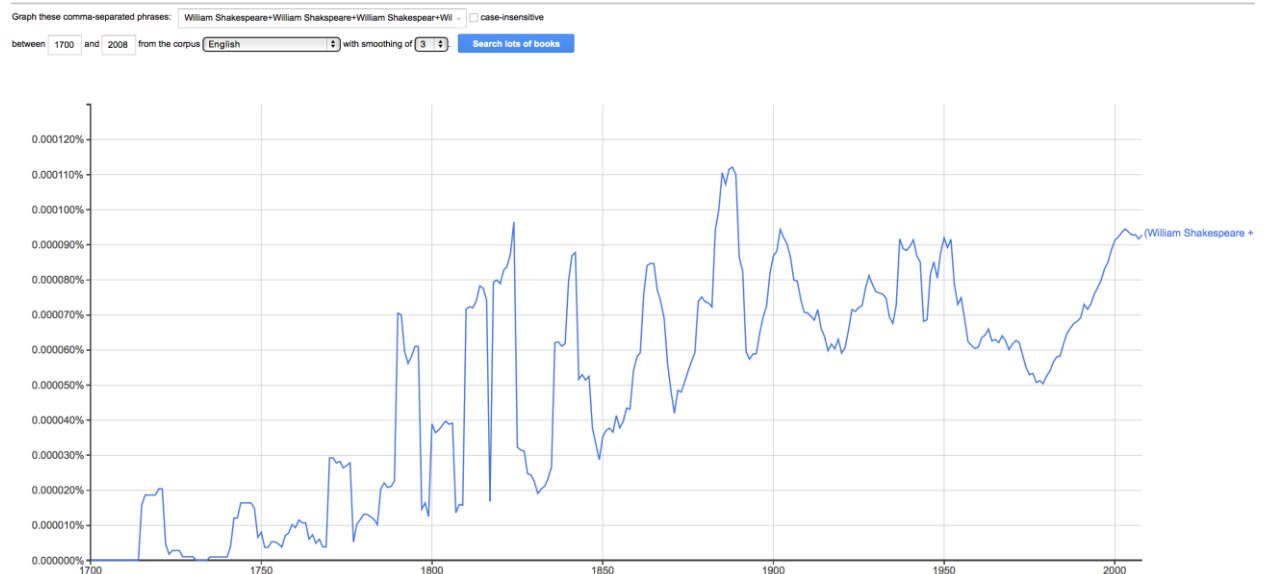


Figure 16: A Google Ngram search for ‘William Shakespeare+William Shakspeare+William Shakespear+William Shakspeare+William Shaxper+William Shakspere’ in the years 1700-2008 (controlling for the various spellings of Shakespeare’s name). Image from Google.

*What has been the most commercially successful Shakespearean film?* We can look at box office totals thanks to sites like the Internet Movie Database (IMDB), but this question is complicated by two issues. First, what constitutes a ‘Shakespearean film’?

Should we include *productions* (using Shakespeare's text as the basis for its screenplay, even if the setting is radically changed), *adaptations* (taking inspiration from one of Shakespeare's plays to create an original story), and films *about Shakespeare* (presenting him as a character or his art as subject matter)? Second, any comparison of box office totals across time is complicated by inflation. The \$42 million earned by Zeffirelli's *Romeo and Juliet* in 1968 is more remarkable than the \$46 million earned by Lurhmann's *Romeo + Juliet* in 1996. Adjusting for inflation, Figure 17 reveals that the most commercially successful Shakespearean film to date is *The Lion King*, a very loose adaptation of *Hamlet*. Indeed, most commercially successful Shakespearean films are adaptations (7 of the top 10) rather than straightforward productions or films about Shakespeare. In terms of straightforward productions, the most successful, *Romeo and Juliet* (1968), was more than four times as profitable as the second most successful, *Romeo + Juliet* (1996), when adjusting for inflation. Two questions emerged from this research. First, why are Shakespearean adaptations which change a lot about Shakespeare's original plays more commercially viable than Shakespearean productions which retain the original script? Second, why is *Romeo and Juliet* Shakespeare's most commercially viable text?

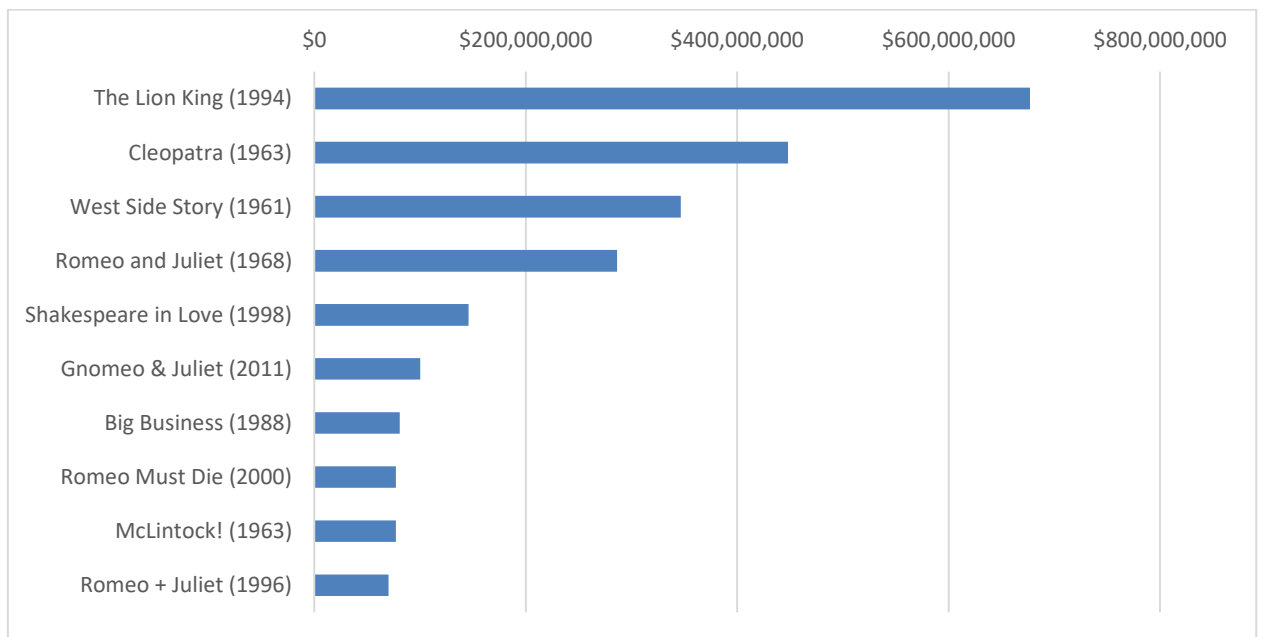
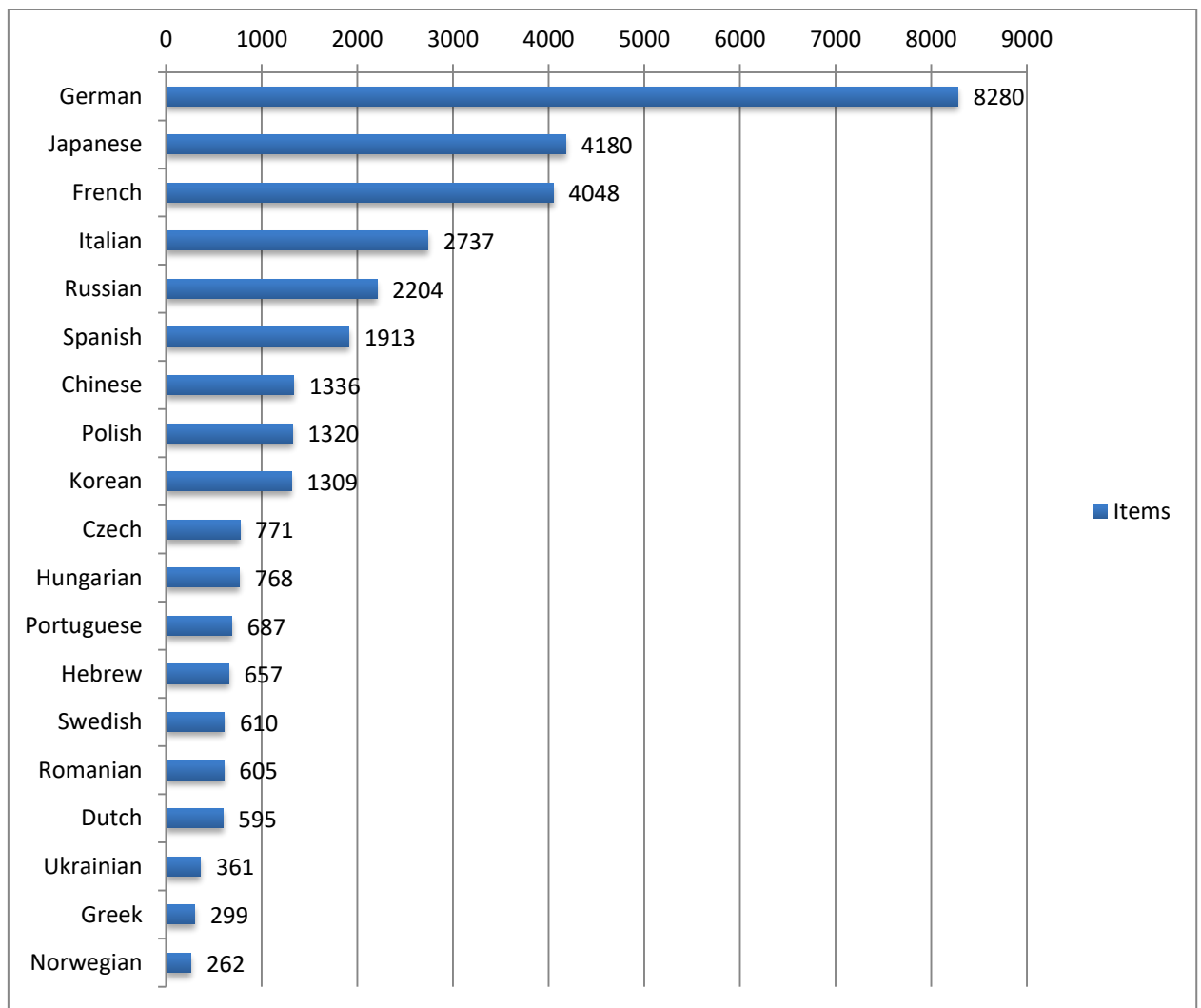


Figure 17: The Most Commercially Successful Shakespearean Films, According to IMDB.

*Other than English, what is the most common language for Shakespeare scholarship?* Before searching in the World Shakespeare Bibliography, I hypothesized that languages with similar cultural contexts to English would exhibit the greatest affinity for Shakespeare, and that generally there would be more results for a language the more common that language is. Therefore, I expected Western European languages like

German and French to have the most results, potentially followed by languages with many speakers like Chinese and various Indian languages. Testing revealed that some aspects of this pattern do exist, as Figure 18 shows. French, German, Italian, and Spanish were all in the top ten, and German had easily the most results, with more than twice the results of the second most common language. But there were some significant divergences from my expectations. Most obviously, there was a high number of results for certain East Asian languages, specifically Japanese and Korean. It makes sense that Chinese would rank high, since it is the most widely spoken language in the world, but Japanese is the third most-represented language in the World Shakespeare Bibliography while being the ninth most-popular language in the world; Korean is the tenth most-represented language in the Bibliography while being the seventeenth most-popular language in the world. How did international relations after World War II and the Korean War affect Shakespeare's popularity in Japan and Korea?



*Figure 18: The Top 20 Languages (Other than English) Represented in the World Shakespeare Bibliography.*



After providing students with sample empirical reports such as these, I allow them to generate their empirical data themselves, or to work off data from others (me, other students, or previous scholars). If they use previously generated data, however, they must be able to replicate the results by either following the original researcher's methodology or conducting corroborating empirical research. This step in the assignment allows us to discuss the role of replication in the creation of scientific knowledge – and the so-called replication crisis in the sciences<sup>60</sup> – even as we remain focused on the need for qualitative analysis in the pursuit of a literarily and culturally significant argument. For example, one student used data presented in class about Shakespeare's relative unpopularity during the Interregnum and Restoration to argue that 'the calamitous – indeed unprecedented – upheaval during the English Civil War engendered a powerful desire for order and stability that not only led to the Restoration but also influenced culture during the Restoration, particularly a popular dislike for the perceived lack of order in Shakespeare's plays as well as the critical acceptance of the rigid, but stable, Neoclassical approach to literature'.<sup>61</sup>

Another student used data presented in class about the disproportionate prominence of Shakespeare in Japan to give a Shakespeare-inflected reading of the cultural shift from the feudalistic Tokugawa Japan (1603-1867) to the imperial Meiji Japan (1868-1912): he argued that the theme of honor in a text like *Hamlet*, first translated into Japanese in 1874, spoke directly to the Bushido code of the samurai warrior class who led the Meiji restoration.<sup>62</sup> A third student asked the all-important question – *Why?* – about Estill, Klyve, and Bridal's data showing that the popularity of *The Tempest* has significantly increased since the 1970s (see Figure 11). We can't just chalk it up to the invention of the post-colonial critique of the play, she reasoned, since that critique was advanced by Latin American authors in the early- to mid-twentieth-century. The popularity of *The Tempest* only took off once the post-colonial reading was advanced by white academics who looked like the colonizers, not the colonized, and who presented the post-colonial reading in objective, academic articles rather than subjective, personal essays,

---

<sup>60</sup> See Tom Bartlett, 'Replication Crisis in Psychology Research Turns Ugly and Odd', *Chronicle of Higher Education* (June 23, 2014), <http://www.chronicle.com/article/Replication-Crisis-in/147301>.

<sup>61</sup> Xavier Gonzalez, 'Shakespeare Unpopular? Historical Forces Behind Shakespeare's 'Unrestored' Critical and Popular Reputation during the Restoration of the English Monarchy', unpublished essay for Jeffrey Wilson's *Why Shakespeare?* class (Cambridge, MA: Harvard University: Spring 2015), 1-19. All student essays cited below refer to this class and will be noted according to date.

<sup>62</sup> Trevor Jones, 'Hamlet the Samurai: A Look at Hamlet's Popularity in Japan', unpublished essay for Wilson, *Why Shakespeare?* (Fall 2015), 1-14.

something she called ‘the colonialism of the postcolonial criticism of *The Tempest*’.<sup>63</sup> A fourth student used the statistic from the British Council stating that 21 million Chinese students study *The Merchant of Venice* in secondary school as the entryway for a qualitative study of the late-twentieth-century Chinese adaptations: influenced by Marx and Mao, these Chinese adaptations minimized the dominant theme of religious conflict in *Merchant* (Christians vs. Jews) in order to emphasize the economic theme of class conflict and to present the play as a tragedy of capitalism.<sup>64</sup>

In most cases, however, students generate their data themselves. In fact, one student showed, contrary to the British Council, that *Hamlet*, not *The Merchant of Venice*, is the most popular play in China, at least in academic scholarship; moreover, *Hamlet* is disproportionately popular in China compared to Western countries (see Figure 19). Yet scholarship on Shakespeare did not emerge in China until after the death of Mao Zedong in 1976 (see Figure 20). This student used these empirical nuggets to launch his argument that *Hamlet* became popular in China because of the play’s affinity with the Confucianism that China sought to reconnect with in the wake of Mao’s failed Cultural Revolution.<sup>65</sup>

---

<sup>63</sup> Sarah Horne, ‘The Colonialism of Postcolonial Criticism of *The Tempest*’, unpublished essay for Wilson, *Why Shakespeare?* (Fall 2015), 1-12.

<sup>64</sup> John Ball, ‘*The Merchant of Venice*: A Western Answer to China’s Cultural Revolution’, unpublished essay for Wilson, *Why Shakespeare?* (Fall 2016), 1-13, citing the Royal Shakespeare Company, ‘Students From Around the World: World Shakespeare Festival’ (2011), <https://www.youtube.com/watch?v=X0wzYI0hu40>.

<sup>65</sup> Luca Schroeder, ‘The Tragic Prince as Confucian Hero: *Hamlet*, Bardolatry, and the Great Cultural Revolution’, unpublished essay for Wilson, *Why Shakespeare?* (Fall 2014), 1-15.

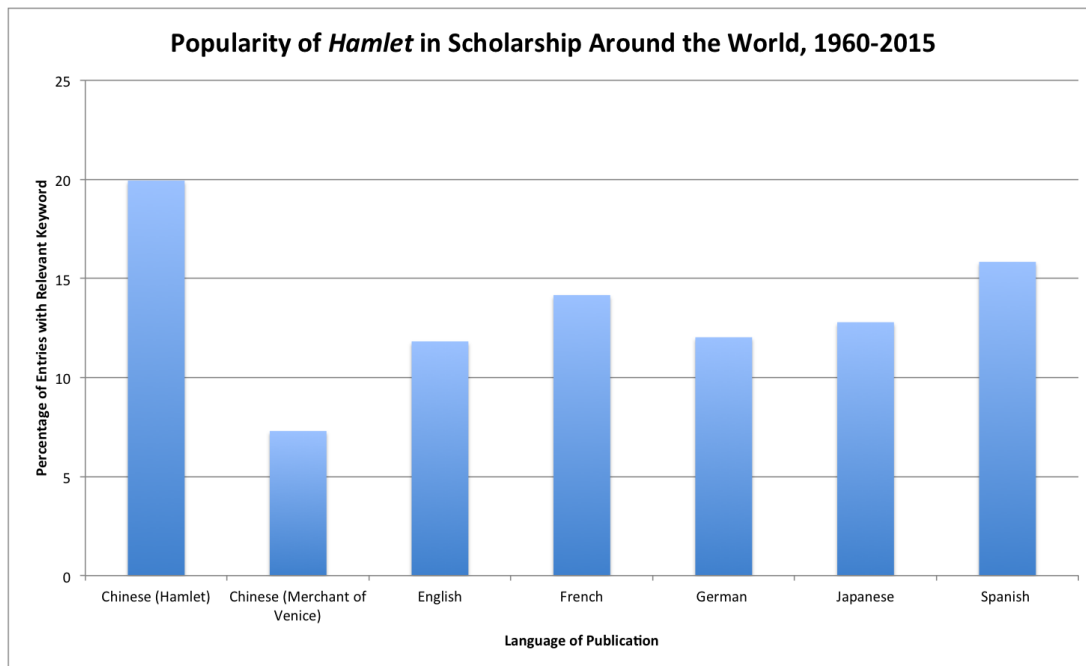


Figure 19: Comparison of the proportion of scholarship and theatrical productions listed in the World Shakespeare Bibliography Online corresponding to the keyword ‘Hamlet’ across the literature in select major Western languages, Japanese, and Chinese.

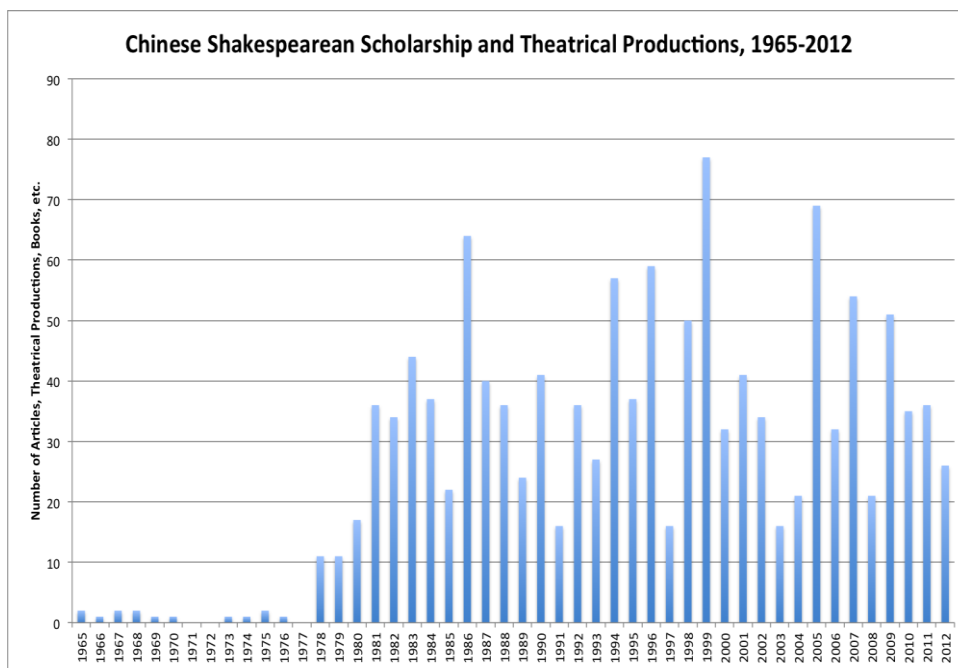


Figure 20: Number of Chinese language entries by year listed in the World Shakespeare Bibliography Online.

A different student, looking at an Ngram locating the start of Shakespeare’s popularity in France in the 1840s-60s, complicated the artistic explanation (France finally rejected

neo-classicism, allowing Shakespeare's non-neo-classical plays to thrive) with a political explanation:

Changes in the European organization of diplomatic relations and power structures, which were developed during the Congress of Vienna in 1815, allowed Shakespeare's rise in popularity in France in the mid 19th century. On the one hand, the novel balance of powers shifted the Anglo-French relationship from one of intense competition to one of collaboration and mutual respect, allowing the French to overlook Shakespeare's anti-French tendencies. On the other hand, the introduction of soft power combined with England's cultural influence lead the French elite to use Shakespeare as means to prove their intellectual prowess, which was necessary as France felt the need to reestablish its worth as a European superpower after Napoleon's defeat in Waterloo.<sup>66</sup>

Another student produced empirical evidence showing that U.S. courts on the federal and state level have cited Shakespeare more than 800 times. Why? What does a 400-year-old English playwright who lived under a system of common law have to do with modern American justice? This student worked up from his own close readings of a number of those citations to argue that modern U.S. legal writers find Shakespeare appealing because 'Shakespeare's works – especially comedies like *The Merchant of Venice* and *Measure for Measure* – represent the shortcomings of legal objectivism and formalism within judicial proceedings in the spirit of critical legal studies'.<sup>67</sup> Someone else scoured the music database *Genius* to discover that, in 2016, there were more allusions to Shakespeare's name in hip hop music than ever before, and also more allusions to Shakespeare than to recent pop-culture celebrities such as Beyoncé, Justin Bieber, Kim Kardashian, Monica Lewinsky, and Donald Trump (see Figure 21). What was Shakespeare doing in hip hop? She argued that the hip hop community is drawn to Shakespeare because his journey from provincial playwright to immortal bard is Western culture's most prominent symbol of upward social mobility and fame achieved through art, an obsession in much hip hop music.<sup>68</sup>

---

<sup>66</sup> Jeanne Olivier, 'Shakespeare's Popularity in France: The Implications of Changing Power Structures', unpublished essay for Wilson, *Why Shakespeare?* (Spring 2016), 4.

<sup>67</sup> Nikolas Paladino, 'Shakespeare in Law: The Bard as a Critical Legal Scholar', unpublished essay for Wilson, *Why Shakespeare?* (Fall 2016), 2.

<sup>68</sup> Sally O'Keeffe, 'Societal Outliers of Intelligence: Shakespeare's Popularity in Hip-Hop Culture', unpublished essay for Wilson, *Why Shakespeare?* (Fall 2016), 1-12.

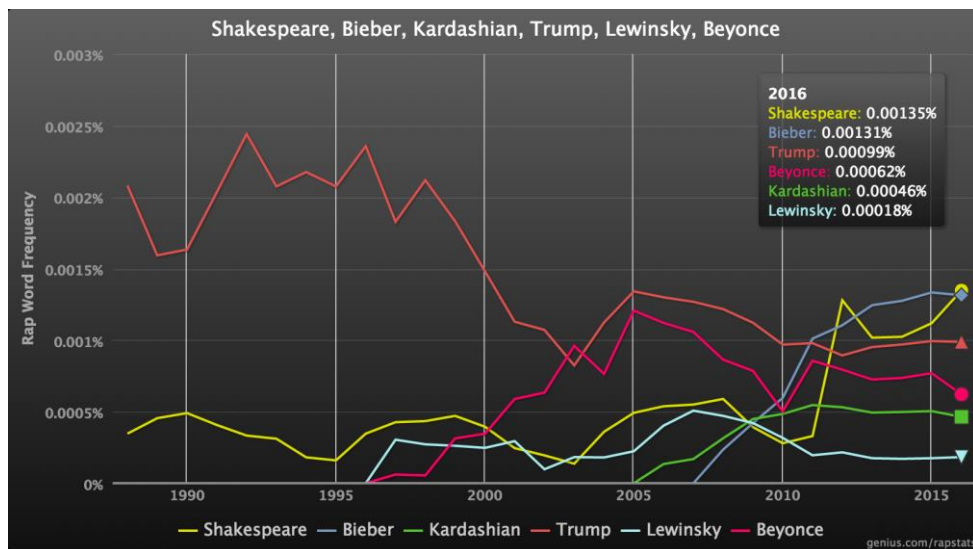


Figure 21: Frequency of Name Mentions in Rapurce: ‘Rap Stats’, According to Genius.

Sometimes, students’ arguments are imaginative and provocative but not fully demonstrated, as when one student suggested, intriguingly, that the high number of Jewish Shakespeare scholars – more than 9 percent of the 152 academics listed on the Wikipedia Shakespearean scholars page are Jewish, compared to the roughly 2.2 percent of the U.S. population that is Jewish – may point to an affinity between the Talmudic hermeneutic tradition of Biblical interpretation and academic Shakespearean criticism.<sup>69</sup> It is also important to note that, in this assignment, as in all scientific research, quantification can go awry. Poorly conceived data-gathering projects can produce misleading results which can be stubbornly defended by deeply invested researchers. Discussions of methodology and data quality are crucial. Or sometimes there is no data available. I have received plenty of *Why Shakespeare?* essays that, after failed or unhelpful empirical reports, still generated compelling arguments about Shakespeare’s modern manifestations. Perhaps the most memorable was the student fascinated by pornographic appropriations of Shakespeare. He could not generate any reliable data, but he did contact the producer of the most popular Shakespeare porn, *A Midsummer Night’s Cream*, to interview him.<sup>70</sup> This interview led to the discovery that Shakespeare was a way for this art-school drop-out turned pornographer to reconnect with the artistic origins of his interest in film. This student was not doing quantitative analysis, but his fieldwork did resemble a recognizable form of social science research.

<sup>69</sup> Eric Wasserman, ‘The Jewish Tradition of Interpreting Texts Lends Itself Well to Shakespeare’, unpublished essay for Wilson, *Why Shakespeare?* (Fall 2015), 1-15.

<sup>70</sup> Matthew Mandel, ‘This is Shakespeare XXX: The Artistic Impulse in Porn Parody’, unpublished essay for Wilson, *Why Shakespeare?* (Spring 2016), 1-11.

## VII. Conclusions

For students, quantitative analysis of Shakespeare's text and reception can serve as an important illustration that literary studies and the humanities in general are not discourses devoid of fact and truth filled with whims of subjective opinion, as some detractors say when levelling charges of airy relativism. Literary studies are often written to provide new perspectives on information everyone already knows about, and new facts about texts, such as those unearthed through empirical investigation, can contribute to new understandings of culturally significant documents. Additionally, quantitative analysis of Shakespeare or any other literature can serve as a gateway for more advanced forms of empirical inquiry in the sciences and social sciences.

If institutionalized and funded, this sort of research could, I suppose, produce a repository of facts about Shakespeare, his works, and their reception. I can see potential value in the creation and funding of something like a 'Shakespeare Fact-ory' – akin to the Literary Lab at Stanford – designed to produce reliable empirical information that can be disseminated to Shakespearean scholars and critics to interpret and incorporate into their research. At the same time, I think a more promising and more likely route forward would be for Shakespearean scholars to acknowledge, embrace, study, and master methods of quantitative thinking about literature and to fold them into our other modes of analysis. These methods would therefore not represent a radical break. Rather, the application of quantitative methods will enhance but not replace the kind of qualitative analysis traditionally done in the humanities.

Thus, I expect quantitative analysis to be simply incorporated into Shakespeare studies without much trouble. My guess is that distant reading will not be a methodological revolution in the way that, for example, new historicism in the 1980s was a movement which involved a rethinking of the way we do literary studies. As digital literacy becomes more and more infused in our daily lives, distant reading will simply become another tool in the toolkit of literary studies.