



# DATA, DIGITAL HUMANITIES, AND ENGLISH STUDIES

Dr. Elizabeth Davis  
Department of English  
[eadavis@uga.edu](mailto:eadavis@uga.edu)

# English Studies: What Do We Do?

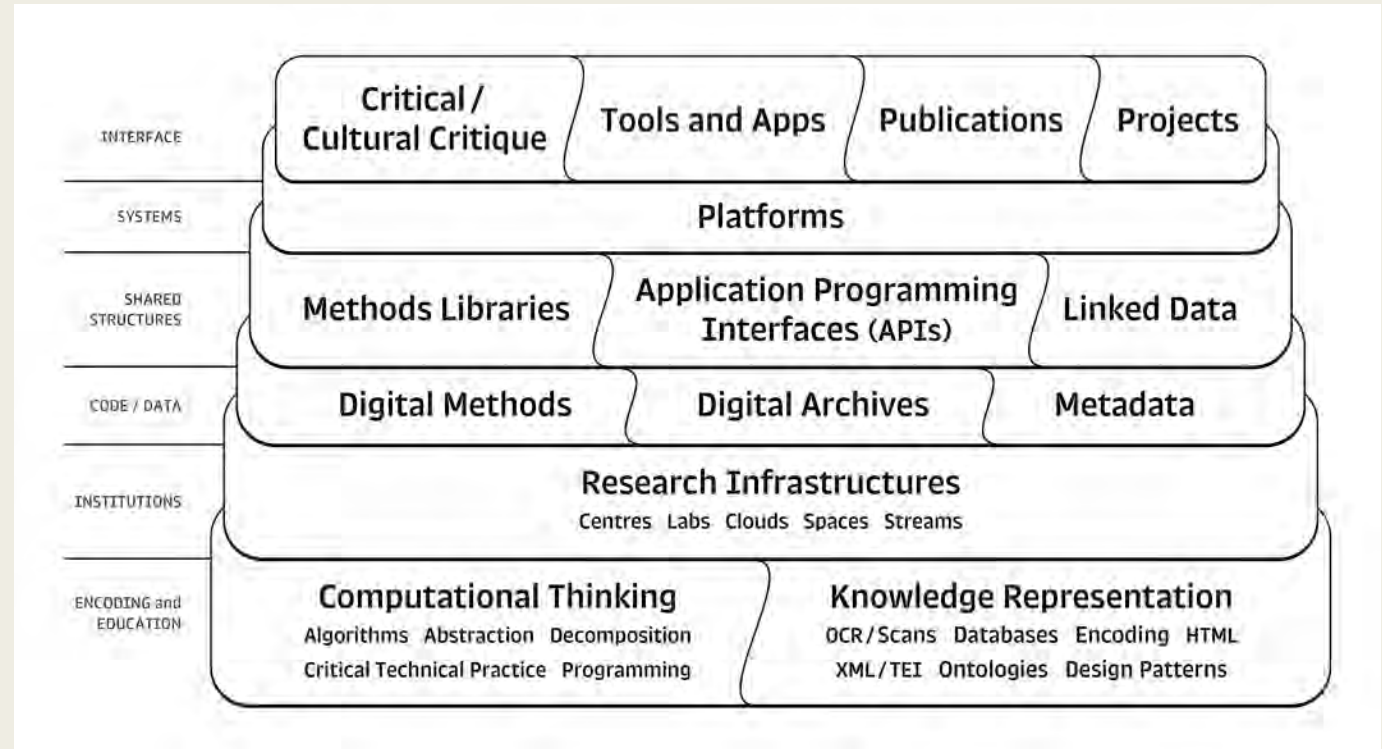
- Literary Studies
- Rhetoric & Composition (aka, Writing Studies)
- Creative Writing
- Linguistics

["A woman standing, facing the left in profile, wearing a hat and a lace collar, reading from a prayer book"](#) by Abraham Bosse is licensed under [CC0 1.0](#)



# So What is Digital Humanities?

- Application of digital tools and methods to humanities study
  - *Digital archives*
  - *Network analysis*
  - *Quantitative analysis*
  - *Tool-building*








# Archives

- [William Blake Archive](#)
- [Project Gutenberg](#)

An aerial photograph of a river valley. The river is a dark, winding line that meanders through a patchwork of green agricultural fields. The fields are separated by thin, light-colored lines, likely fences or roads. The overall scene is a high-angle, wide-area view, illustrating the concept of 'birds-eye view' or 'distant reading' mentioned in the text.

# Digital Texts, Distant Reading

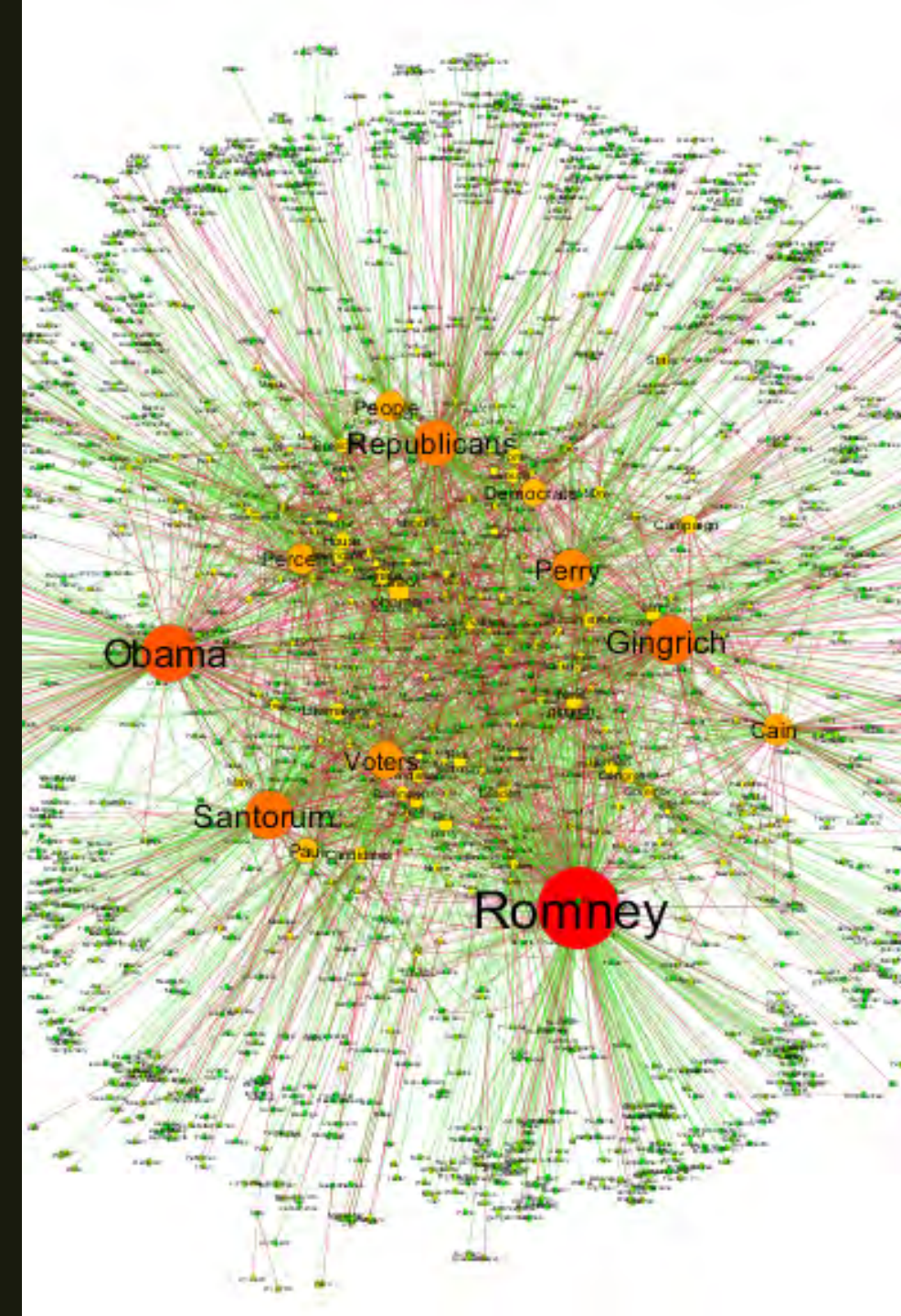
- Too many texts, too little time...
  - *Close reading: human*
  - *Distant reading: computer*

*The birds-eye view: patterns, landscapes, the big picture we can't see up close*



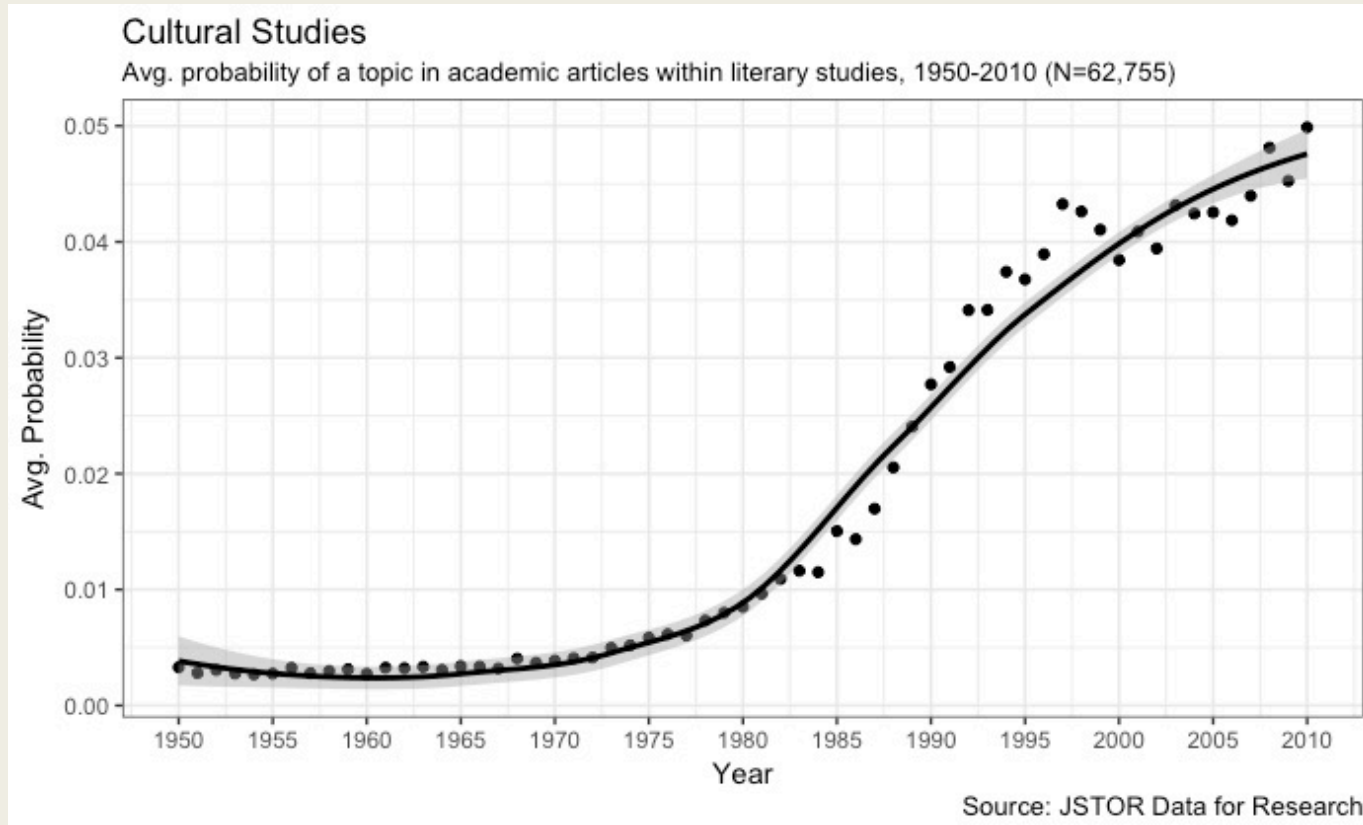
# Network Analysis

- Mapping relationships
- Visualizing connections
- Interactive mapping



Narrative Network of US Election 2012 - Nodes indicate noun phrases, links go from subject to object, color expresses relation of support or opposition. Appeared in: "Automated analysis of the US presidential elections using Big Data and network analysis; S Sudhahar, GA Veltri, N Cristianini; Big Data & Society 2 (1), 1-28, 2015"

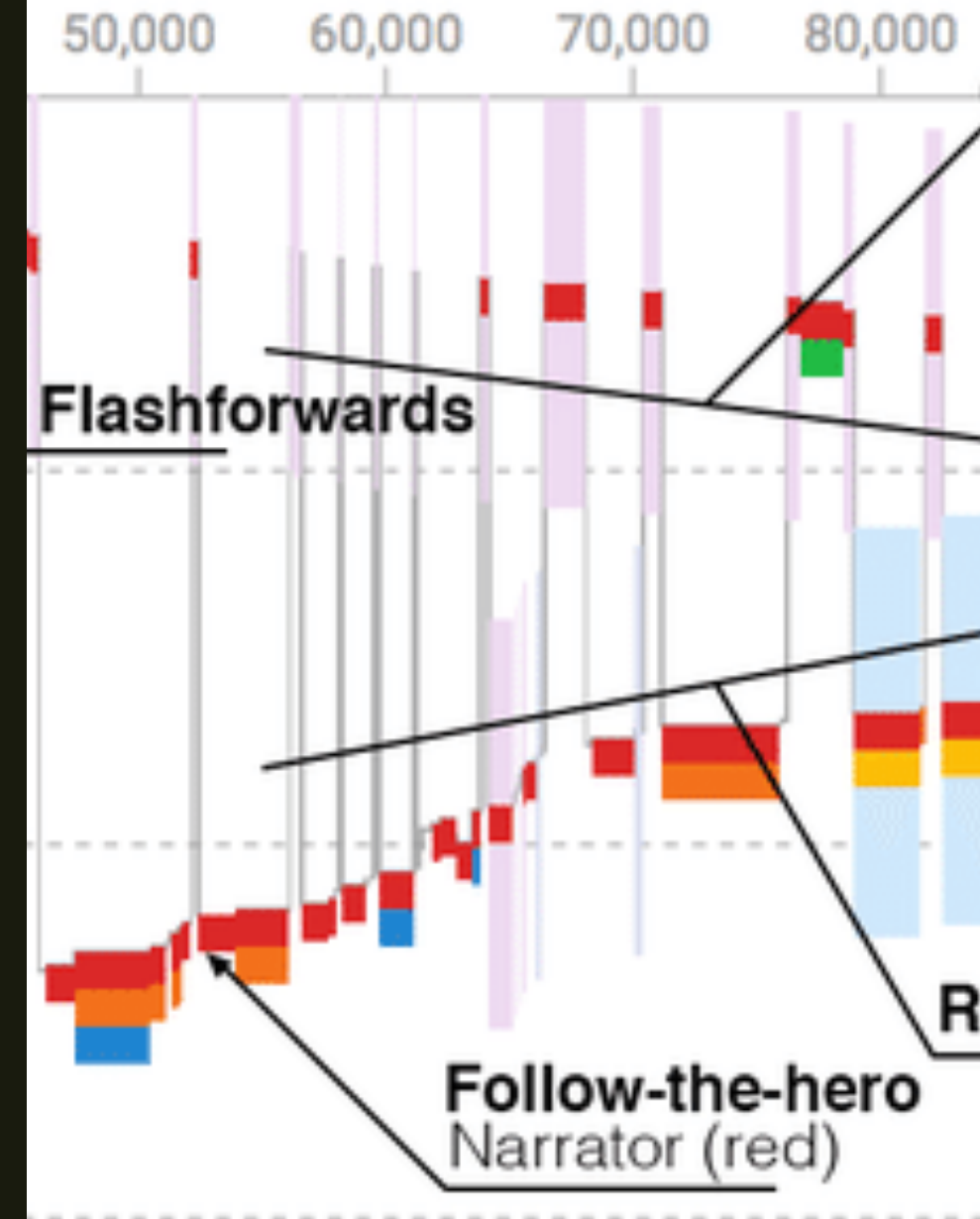
# Quantitative Analysis



- Cultural phenomena – as reflected in language/word usage
  - *Topic modeling*
- Natural language processing
  - *Gender bias*
  - *Specific events*
  - *Narrative analysis*

# Potential for Data-Driven Analysis

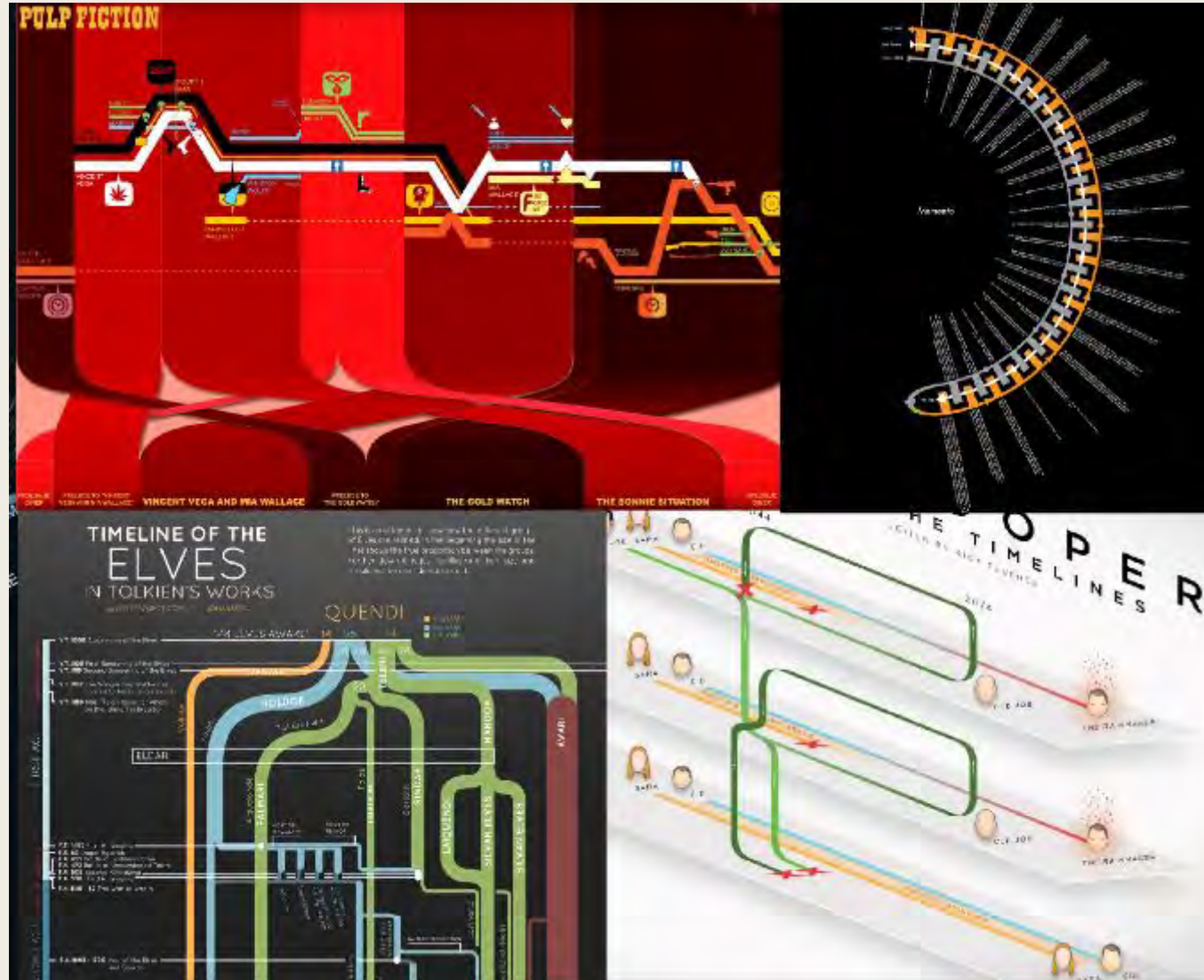
- scale up - the numbers game; abstraction; pattern recognition
- genre analysis through statistics (e.g., defining "sci-fi")
- prediction of narrative plot through natural language processing
- determine levels of "innovation"
- machine learning to identify gender (of author, of characters) to identify trends
- identify character stereotypes
- distinguishing author's voice from characters' voices
- stylistic distinctions to attribute authorship
- identifying adaptations/appropriations
- finding the small places of order - which can help us see the noise and consider noise as "an object of knowledge" (Moretti)





# Narrative Theory + Data

*“We crawled movie scripts from [The Internet Movie Script Database \(IMSDB\)](#). We parsed the scripts [to] extract story data including scenes and character dialogues. Once scene and character information is parsed, our system further extracts semantic metadata from the script...”*



*[Kim, et al., Storycurves](#)*



# Data Analysis as Supplement

“Close reading and historical research delimit a set of literary and social patterns from which the text is easily excluded. Machine learning, on the other hand, suggests that there is some relation to haiku at the level of statistical pattern—a subtle yet consistently present pattern of words and collocations of words. This is influence as a kind of statistical likelihood, where words and other stylistic features are seen to be uniquely distributed across different types of texts. These latent, nonexplicit traces of influence are precisely what the machine is good at detecting and are impossible for the individual reader to identify on a large scale.”

([Long & So, Literary Pattern Recognition](#))



BUT...  
GIGO





# Critique: Oversimplification

*"For instance, what is a novel, what does 'published' mean in this context (date of appearance, editing, composition, acquisition, review, distribution), and how was the 'year' determined. Statistical methods come into play after these decisions have been made, counting objects whose identity was established by interpretative decisions. Many aspects of constructed-ness are in play. But the graphical presentation of supposedly self-evident information (again, formulated in this example as 'the number of novels published in a year') conceals these complexities, and the interpretative factors that bring the numerics into being, under a guise of graphical legibility."*

Johanna Drucker, "Humanities Approaches to Graphical Display"



# Critique: Instrumentality

*“...it is remarkable how little the field [DH] draws on contemporary science-technology studies to enrich its discussion of tools, building, and instrumentality through new understandings of the way researchers, technicians, processes, communication media, and literal instruments come together in what Andrew Pickering calls the “mangle of practice” that is inextricably linked to society and culture”*

*Alan Liu, “Where Is Cultural Criticism in the Digital Humanities?”*



# Critique: Reductive

*“Computational literary criticism is prone to fallacious overclaims or misinterpretations of statistical results because it often places itself in a position of making claims based purely on word frequencies without regard to position, syntax, context, and semantics. Word frequencies and the measurement of their differences over time or between works are asked to do an enormous amount of work, standing in for vastly different things.”*

Nan Z. Da, “The Computational Case Against Computational Literary Studies”



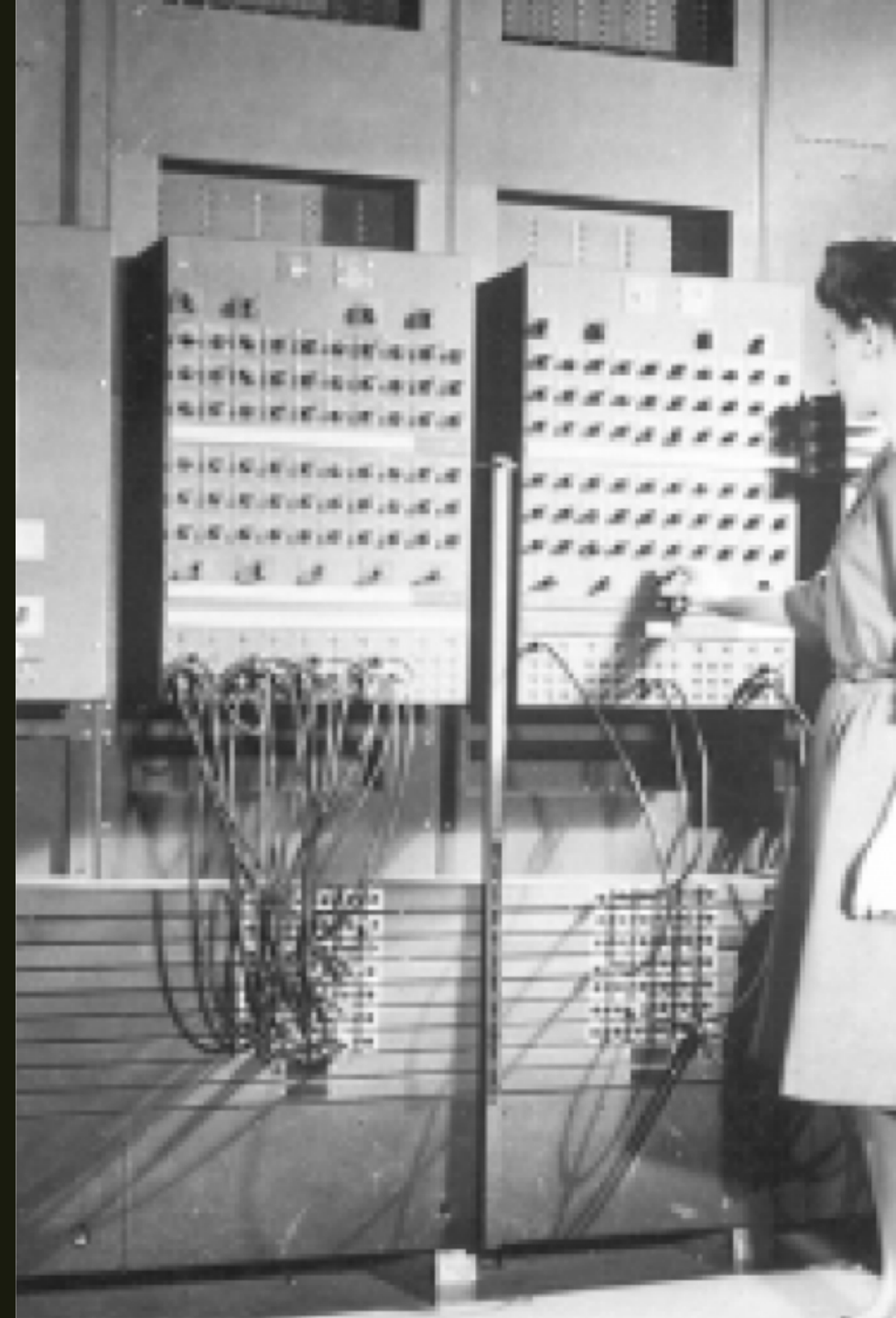
# Domain Expertise and Big Data

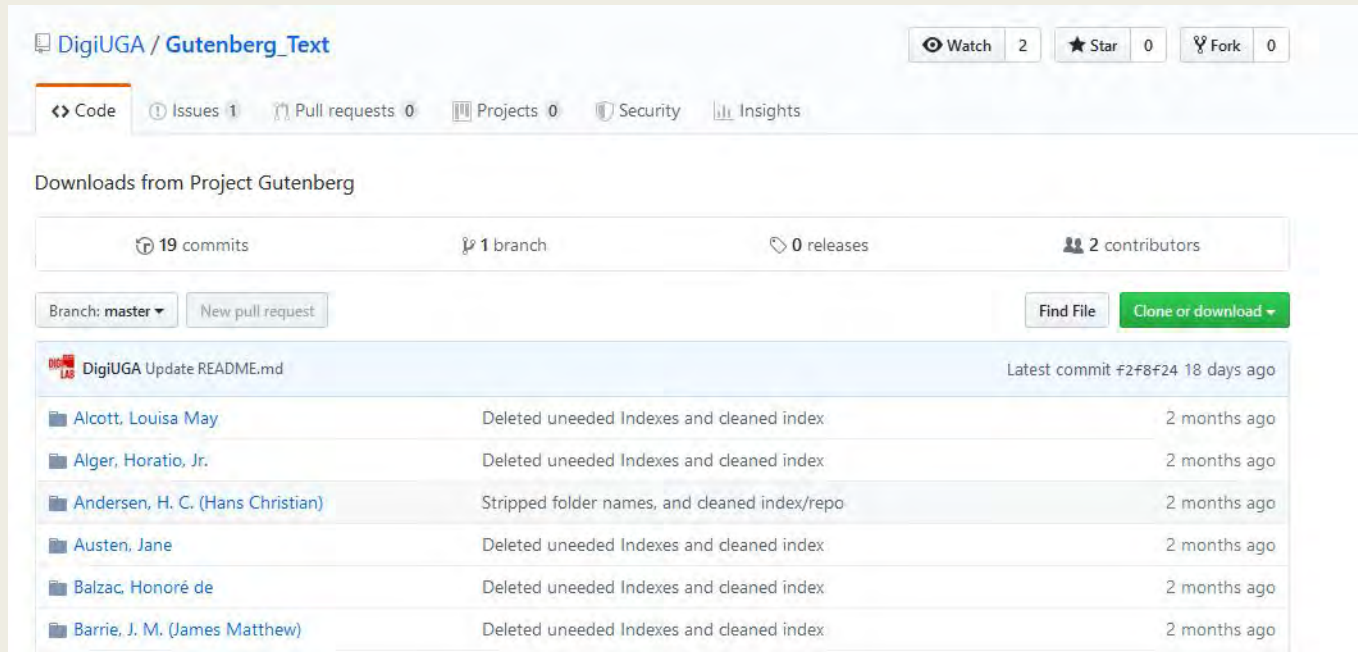
*“Simply put, the computer needs a human to tell it what language consists of – that is, which objects to count.”*

([Roland, UC-Berkeley](#))

*“Literary critics explain and interpret how meaning is made. They analyze texts in ways that illuminate the ideologies texts contain and thereby enable critique of them. Literary critics don’t take data at its word.”*

([Pressman & Swanstrom, “The Literary And/As Digital Humanities”](#))





# What Can We Ask About Textual Data?

- Genre?
- Form?
- Historical events?
- Translation?
- Language use?
- Allusions?
- Appropriation?
- Influence

*“Cultural analytics is not a virtual humanities that replaces the interpretive skills developed by scholars over centuries with mathematical rigor. It is an augmented humanities that, at its best, presents new kinds of evidence, often invisible to even the closest reader, alongside carefully considered theoretical arguments, both working in tandem to produce new critical work.”*

[Mark Algee-Hewitt, Critical Inquiry Online Forum](#)