

# What is digital humanities?


---

AND WHAT IS IT FOR?

INTRODUCTION FOR THE WORKSHOP IN DIGITAL HUMANITIES

PEETER TINITIS

05.09.2017

A solid orange horizontal bar at the bottom of the slide.

# What is digital humanities?

---

Common definitions:

**Digital humanities (DH)** is an area of scholarly activity at the intersection of [computing](#) or [digital technologies](#) and the disciplines of the [humanities](#), it includes the systematic use of digital resources in the [humanities](#), as well as the reflection on their application.<sup>[1][2]</sup> DH can be defined as new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing. (Wikipedia, EN)

Das Fach **Digital Humanities** (deutsch: „digitale Geisteswissenschaften“) umfasst die Anwendung von computergestützten Verfahren und die systematische Verwendung von digitalen Ressourcen in den [Geistes-](#) und [Kulturwissenschaften](#) sowie die Reflexion über deren Anwendung. Es handelt sich um ein interdisziplinär ausgerichtetes Fach, dessen Vertreterinnen und Vertreter sowohl durch eine traditionelle Ausbildung in den Geistes- und Kulturwissenschaften ausgewiesen sind als auch durch ihre Vertrautheit mit einer Reihe von einschlägigen Konzepten, Verfahren und Standards der [Informatik](#). (Wikipedia, DE)

# What is digital humanities?

---

Definitions abound.

<http://www.whatisdigitalhumanities.com/>

## What Is Digital Humanities?

DH is a tool, a process, and a perspective.

*Monica Boyd*

---

NB: Refresh the page to get a new definition. Quotes were pulled from participants from the [Day of DH](#) between 2009-2014. As of January 2015, the database contains 817 rows and randomly selects a quote each time the page is loaded. If you want to do something cool with the data, I am providing a download for the CSV I compiled [here](#).

Made by [Jason Heppler](#). Problems? Questions? I'm [@jaheppler](#) on Twitter.

# Interest in digital humanities

Relatively new keyword.

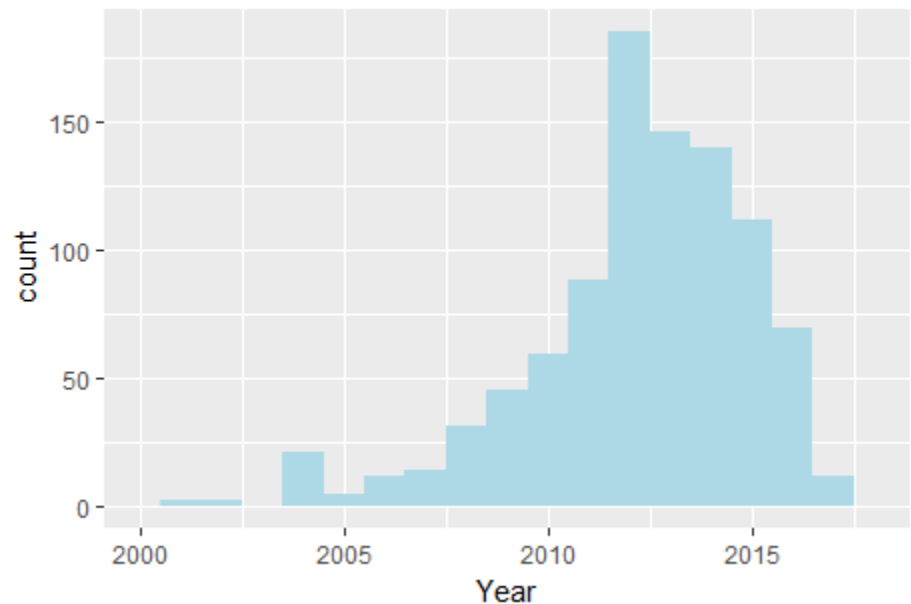
Considerable interest since mid 2000s.

Shown most cited papers with DH in title

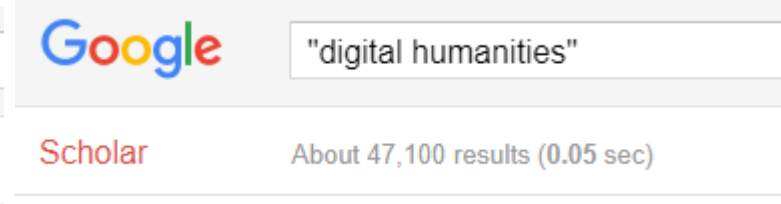
Data only top 1,000 most cited of the 2,640 total

Google Scholar finds 47,100 papers containing it.

Papers with the phrase /digital humanites/ in title



Data retrieved via ***Publish or Perish***, Harzing (2007).



# Google ngram findings

Has important  
predecessors though.

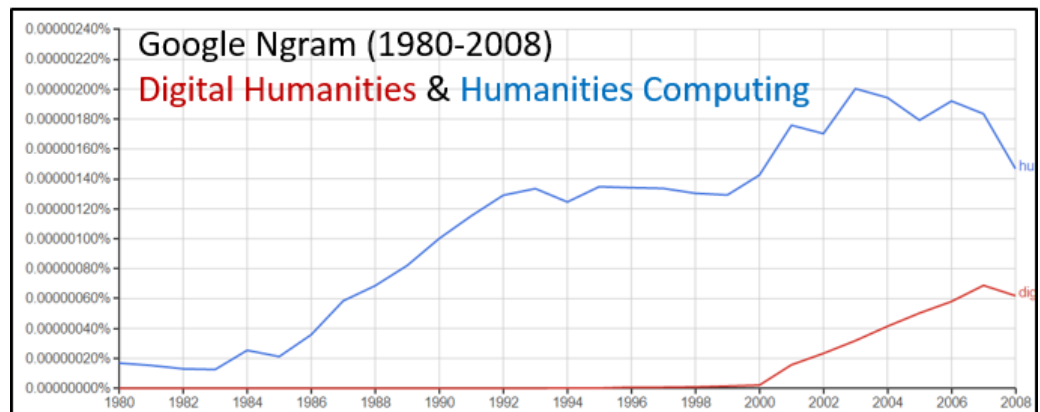
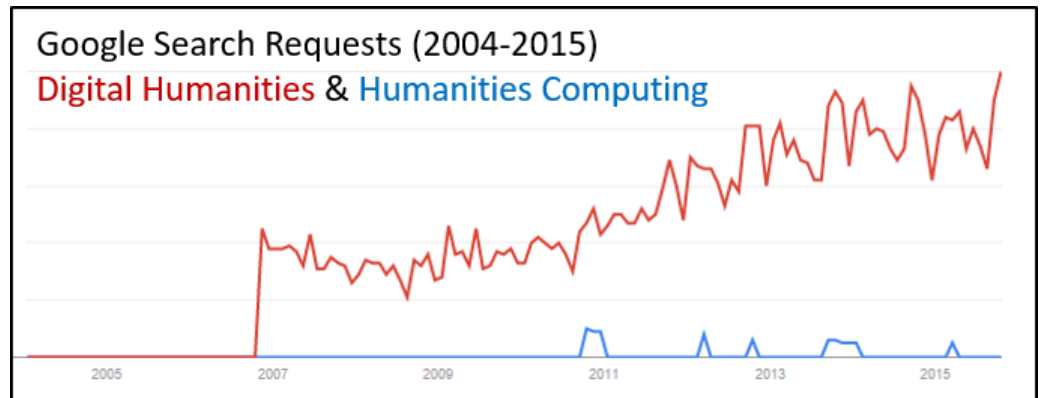


Image by Dave King 2015

# The traditon goes far back in fact

It has had a continuous tradition since 1960s

“Computers and the Humanities” (CHum),

“Digital Humanities Quarterly” (DHQ),

“Literary and Linguistic Computing” (LLC)  
(now “Digital Scholarship in the Humanities”) (DSH).

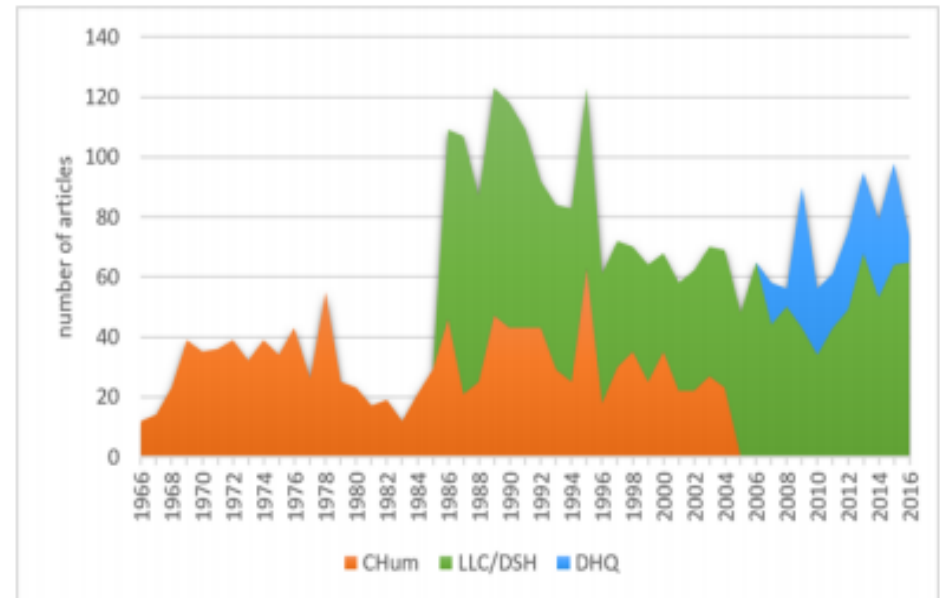
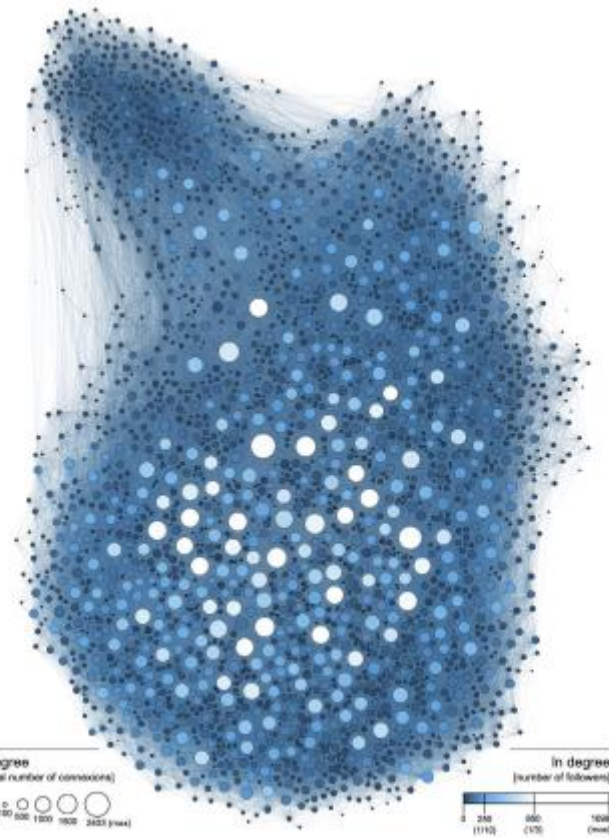


Figure 1. In total, 3,068 journal articles: *CHum* (1,195 articles with 26,033 citations), *LLC/DSH* (1,633 articles with 28,501 citations), and *DHQ* (240 articles with 4,289 citations)

# Finding the digital humanists

A visualization of 2,500 twitter users that identify themselves as digital humanities

Figure 1. Digital humanities network on Twitter: 2,500 users following each other.



**melissa terras** ✓

@melissaterras

Professor of Digital Humanities @UCL |  
Director of @UCLDH | computer science  
v!



**Sarah Bond**

@SarahEBond

Assistant Professor of Classics at the  
University of Iowa and contributor to  
@Forbes. Digital Humanities, Late...



**Andrew Prescott**

@Ajprescott

Digital humanities enthusiast at  
University of Glasgow. Writes on archives,  
lib



**Scott B. Weingart** 🌱

@scott\_bot

Program Director, Digital Humanities @  
Carnegie Mellon University Libraries.  
Historian of science; librarian; circus...

Grandjean, 2016

# Perhaps a useful definition

---

Using computer assistance to do humanities in any step of the research cycle.

This includes

- 1) planning,
- 2) data storage and access,
- 3) analysis,
- 4) writing up and publication,
- 5) sharing the research
- 6) ... (etc).





# Practically

---

This includes:

- A variety of tools just to facilitate regular humanities research.  
(searchable, shareable, transparent)
- New kinds of research questions, and new information to be found
- Procedures of analysis and interpretation
- Data storage and accessibility
- Presentation of research results (interactivity, multimodality etc)

# Close vs distant reading

---

## Close reading

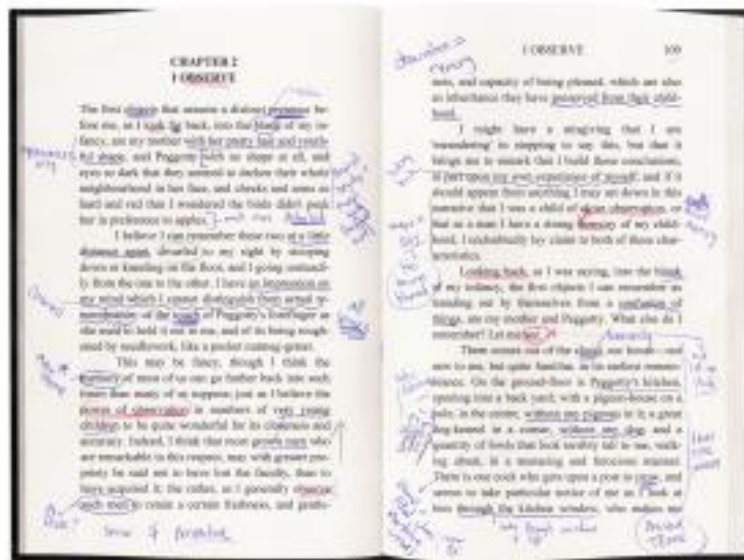
Close reading involves concentrated, thoughtful and critical analysis that focuses on significant details garnered from one text or a small sample of texts (e.g. a novel or collection of novels) with an eye toward understanding key ideas and events, as well as the underlying form (words and structure).

## „Distant reading“ ?

The use of various statistical, data analysis, and visualization techniques to understand the aggregate patterns of large(r) collections of text at a “distance.” [This could] “place the literary field literally in front of our eyes — and show us how little we know about it” (Moretti 2007).

Definitions from King (2015)

# Close reading

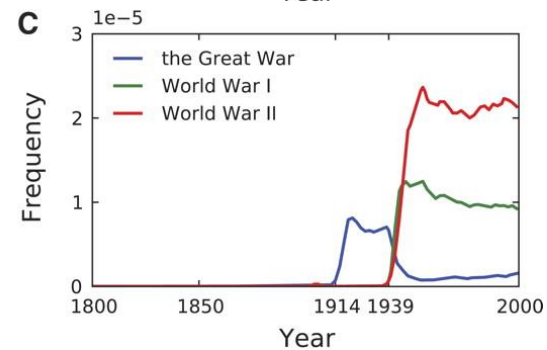
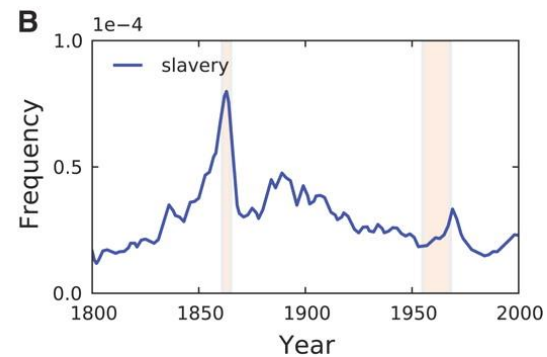
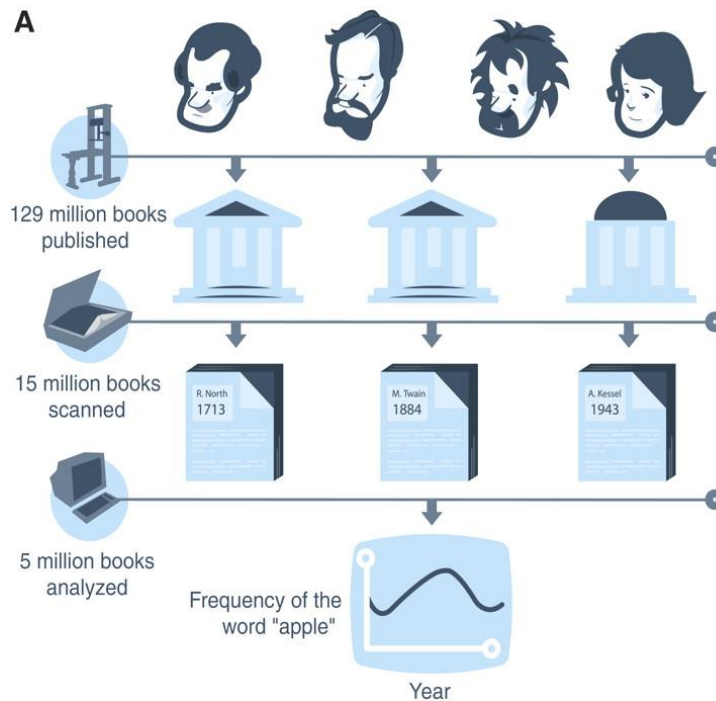


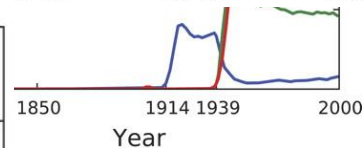
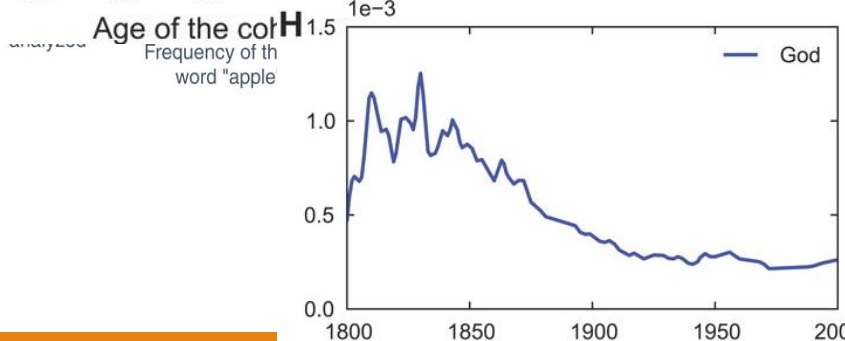
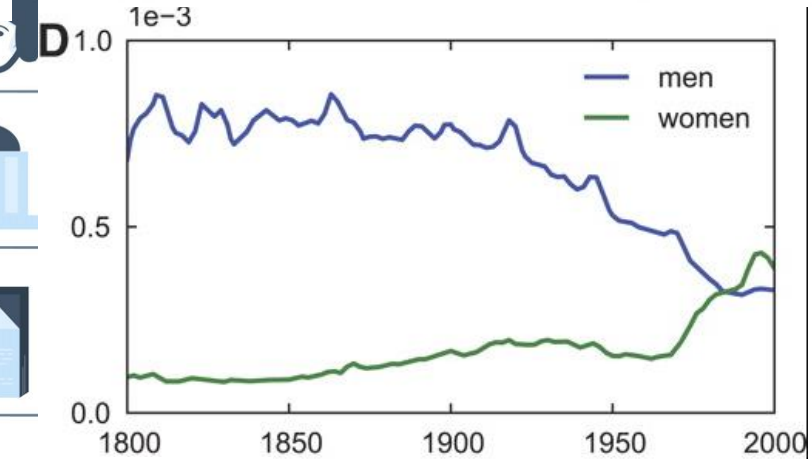
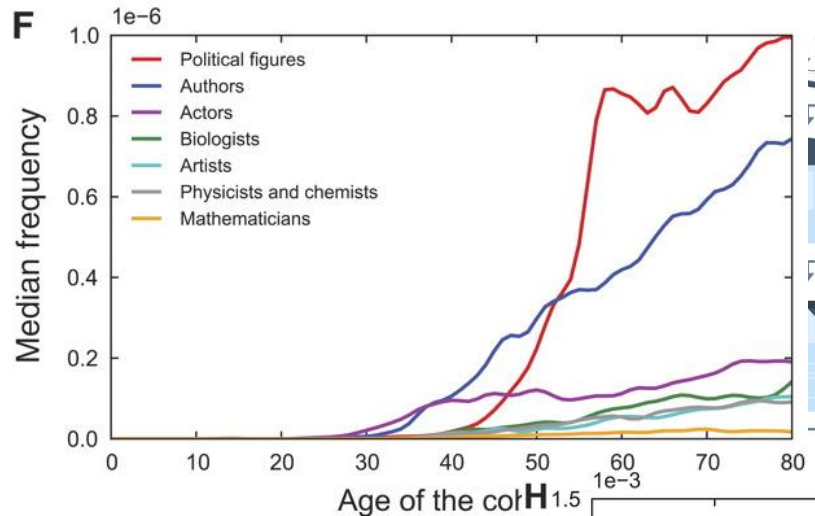
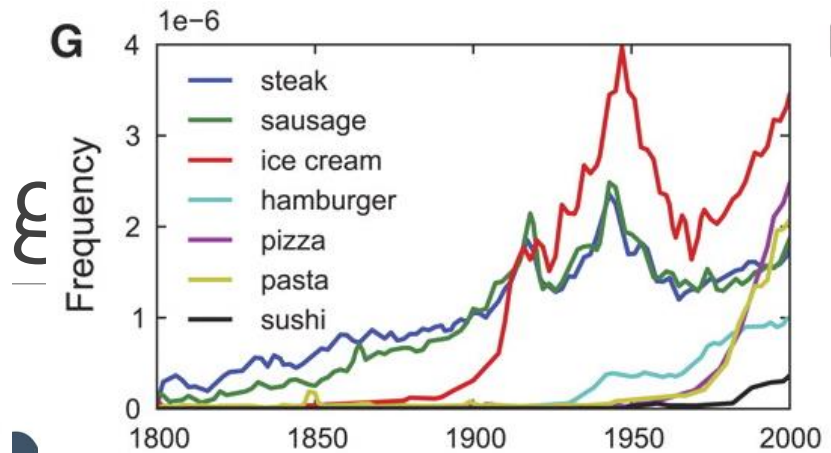
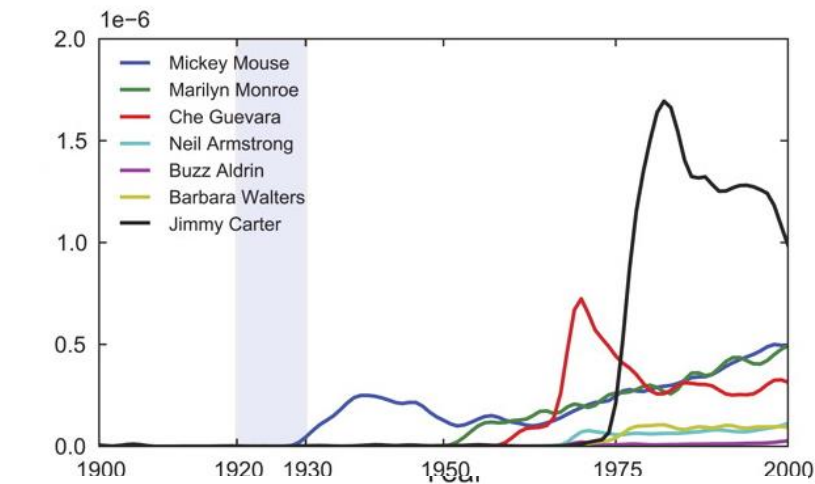
(a) Traditional close reading.



(b) Digital close reading with eMargin.

# Distant reading





<https://books.google.com/ngrams/>

# What's in a name

---

Digital Humanities is a name that has been chosen

- Not computational humanities
- Not digitized humanities.

But is that important?

And it also has a self-reflective streak (+ tools!)

**~~*“What isn’t DH?”*~~**

Visconti (2016)

# Current issues

---

## Keeping contact with the traditional humanities

- Skills for computation and humanities are often in different groups
- Is it a different kind of humanities or the same kind of humanities?

## Too many tools?

- People keep building and building new tools, and it's possible to spend entire time just finding the right ones.

## Education

- Computers, interdisciplinary collaboration and teamwork are often not core of humanities education.
- Courses are abound but content will vary a lot.

Otherwise, many simple things have not yet been done.



# Learning Digital Humanities

---

It has a learning curve, but with some effort and proper teaching, both programming and math are learnable.

- But not everyone needs to know everything – teamwork is more important, and figuring out the meaningful places to go.
- For this, also traditional humanists need to be aware of the possibilities.

But once you get started on this, a whole new world opens up.

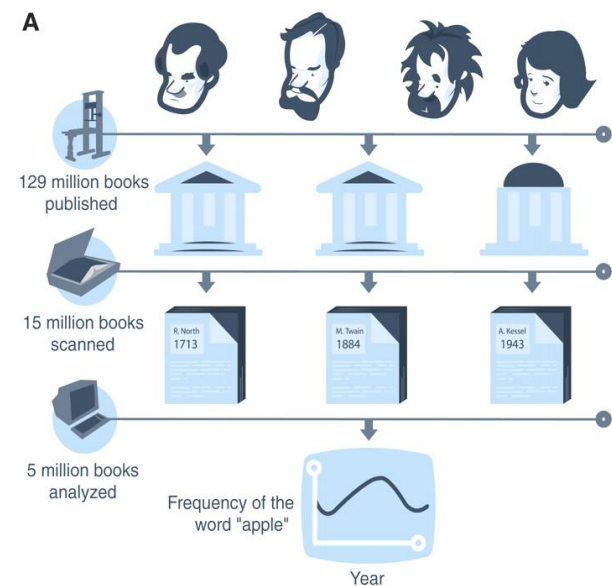
- Datasets every day
- Repeatable steps of analysis
- Can start collaborating in all steps of analysis



# How does it work?

Can be understood in workflows:

- 1) Think about what you want to do
  - Need to propose good questions
  - And also questions that can be addressed with data
- 2) Work with data
  - For this, someone needs to prepare the data
  - Maybe go back to step 1.
- 3) Figure out things that can be done and do them
  - Set up a workflow, collaborate
- 4) See what came out and publish
  - If you had reasonable questions, it is also ok if your results are unsurprising or messy
- 5) And we already know more – repeat.



# Why all this data?

1) We're very good at telling stories to each other and ourselves, would be a good idea to check.

- !!! The relationship between your data and the world can vary. !!!

2) The tools open up a new world.

3) It saves time (for all)!



## HYPOTHESIS TESTING

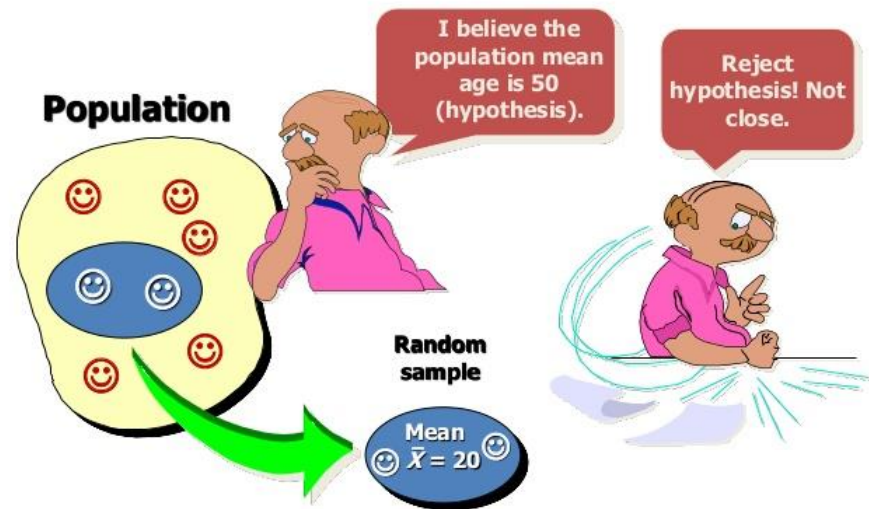


Image by Hakeem-Ur- Rehman (2015)

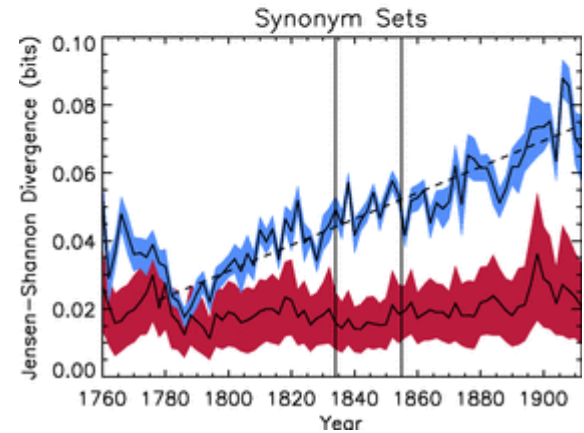
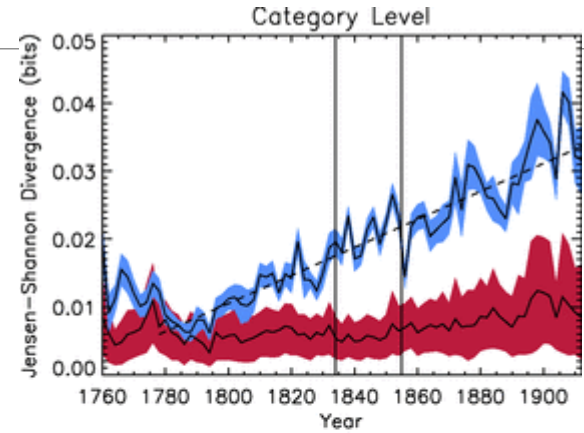
# Some example studies

---

CASE STUDY AFTER CASE STUDY

A solid orange horizontal bar at the bottom of the slide.

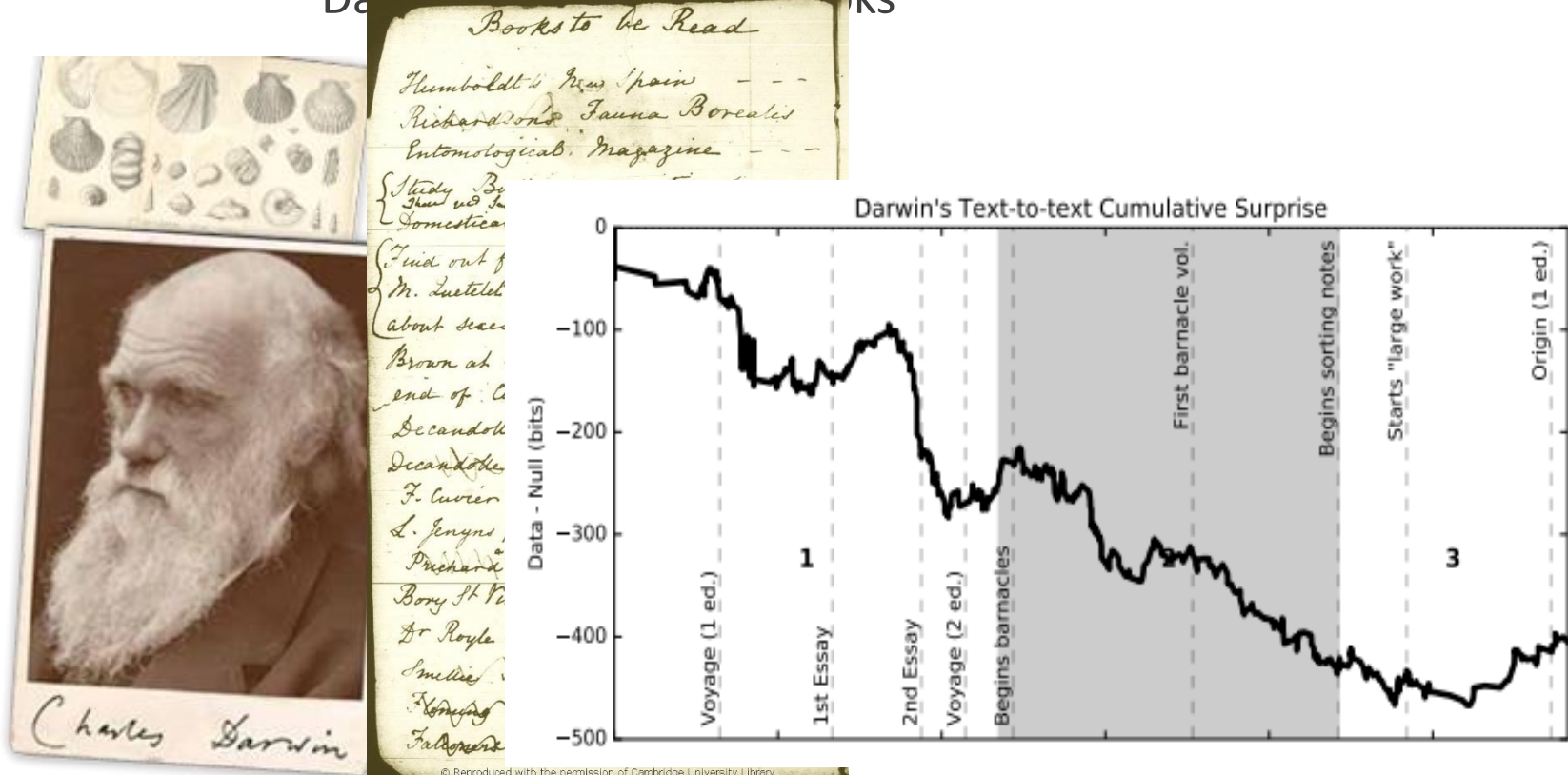
# Civilizing process in Old Bailev's



Klingenstein et al 2014 [The civilizing process in London's Old Bailey](#)

# Darwin's book-to-book surprise

## Darwin's reading notebooks



Murdock et al. (2017), [Exploration and exploitation of Victorian science in Darwin's reading notebooks.](#)

# Authorship attribution

**Shakespeare** \* **Ben Jonson**  
(1564 – 1616) (1572 - 1637)



Shakespeare was romantic Ben Jonson was Classicist

*The Spanish Tragedie:  
OR,  
Hieronimo is mad againe.*  
Containing the lamentable end of Don Horatio, and  
Belimperia with the painful death of Hieronimo.  
Newly corrected, amended, and enlarged with new  
Additions of the Painter part, and others, as  
a lark of life been divers times acted.



LONDON,  
Printed by W. White, for I. White and T. Langley,  
and are to be sold at their Shop over against the  
Swanhead without New-gate. 1615.

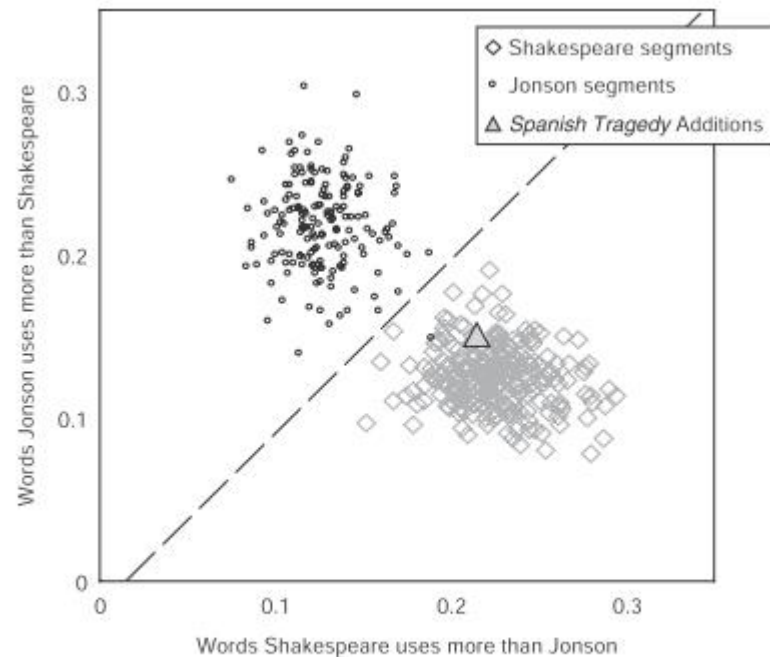
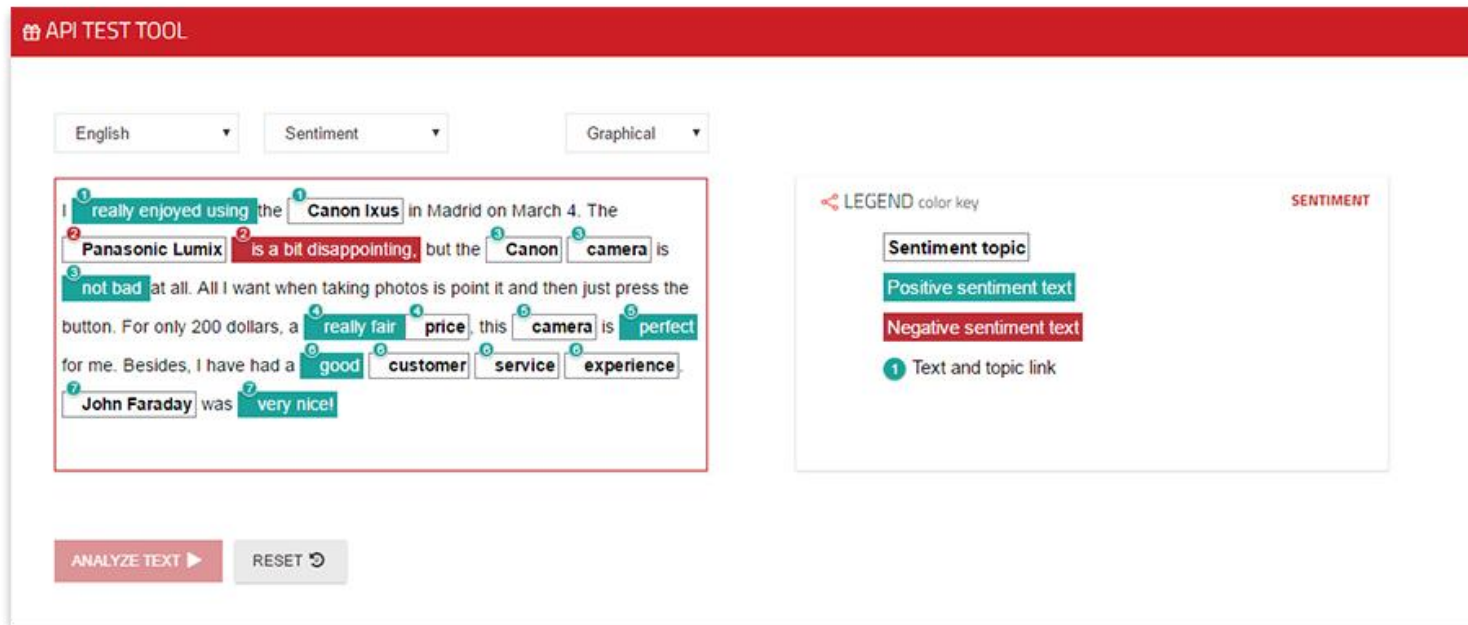


Figure 8.1 Lexical-words test: 2500-word Shakespeare segments versus 2500-word Jonson segments, with the 1602 Additions to *The Spanish Tragedy*.



# Sentiment analysis of storylines

Words and phrases with known sentiment can be picked out in texts



Bitext tool

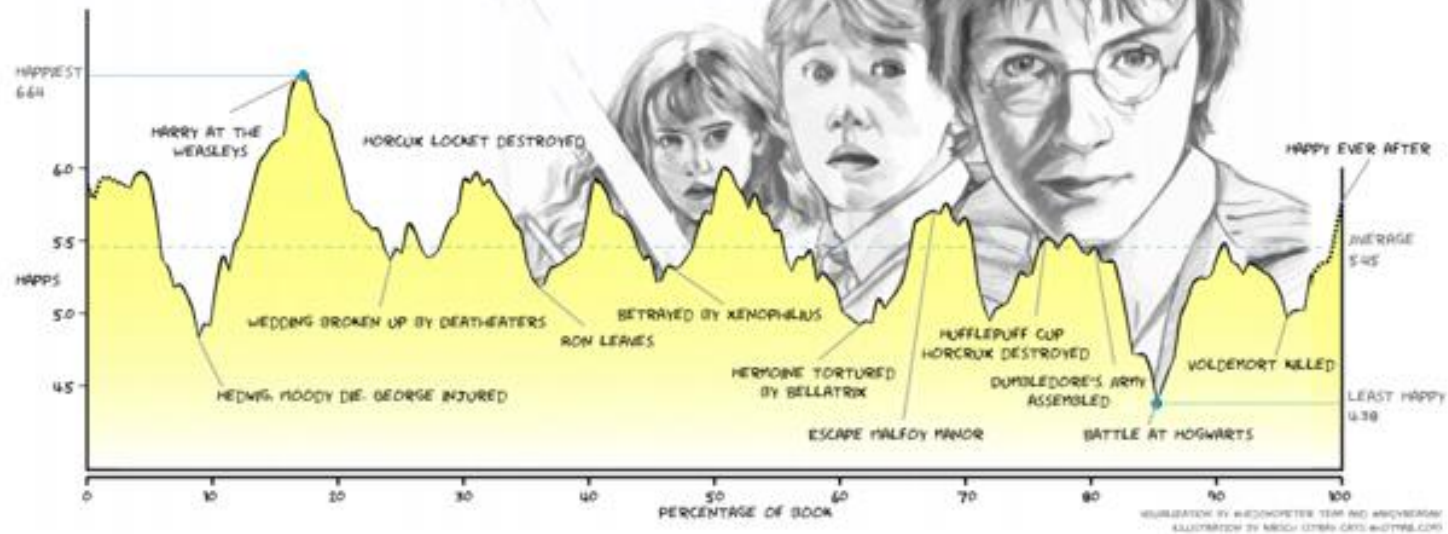
Reagan et al. 2016 The emotional arcs of stories are dominated by six basic shapes  
David Robinson 2017, Examining the arc of 100,000 stories: a tidy analysis

# Sentiment analysis of storylines

Words and phrases with known sentiment can be picked out in texts

API TEST TOOL

## Harry Potter and the Deathly Hallows by J.K. Rowling



Reagan et al. 2016 The emotional arcs of stories are dominated by six basic shapes  
David Robinson 2017, Examining the arc of 100,000 stories: a tidy analysis

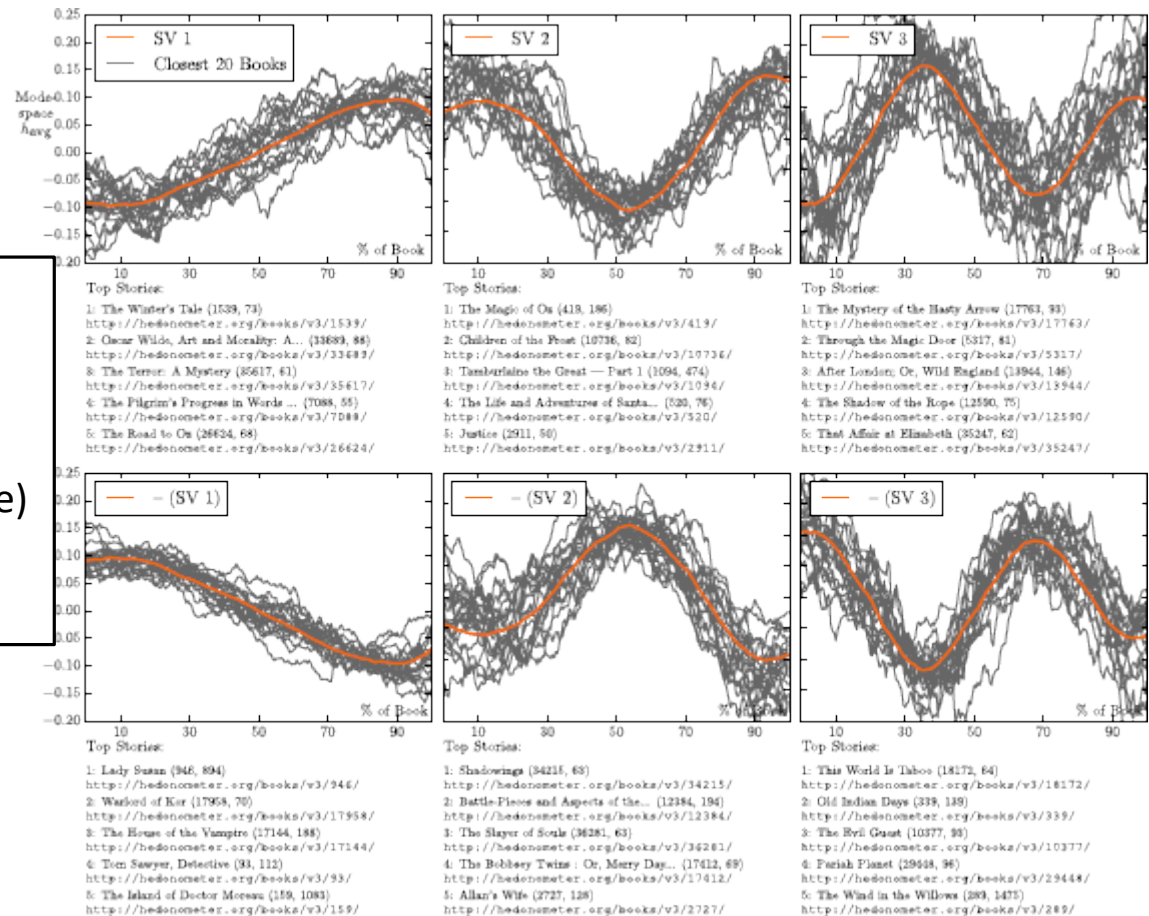
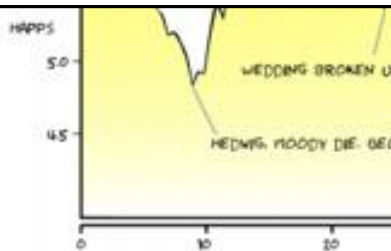


# Sentiment analysis of storyline

Words and phrases

API TEST TOOL

1. Rags to Riches (rise)
2. Riches to Rags (fall)
3. Man in a Hole (fall then rise)
4. Icarus (rise then fall)
5. Cinderella (rise then fall then rise)
6. Oedipus (fall then rise then fall)



Reagan et al. 2016 The emotional arcs of stories are dominated by six basic shapes

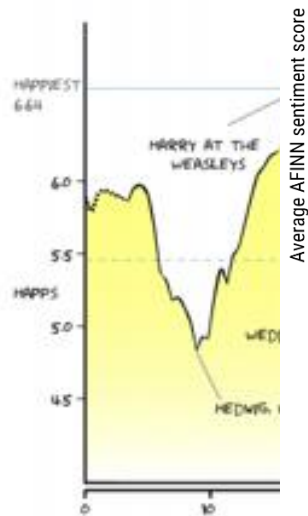
David Robinson 2017, Examining the arc of 100,000 stories: a tidy analysis

# Sentiment analysis of storylines

Words and ph

API TEST TOOL

Harry Potter  
by J.K. Rowling



0.0

-0.2

-0.4

25%

50%

75%

100%

Position within a story

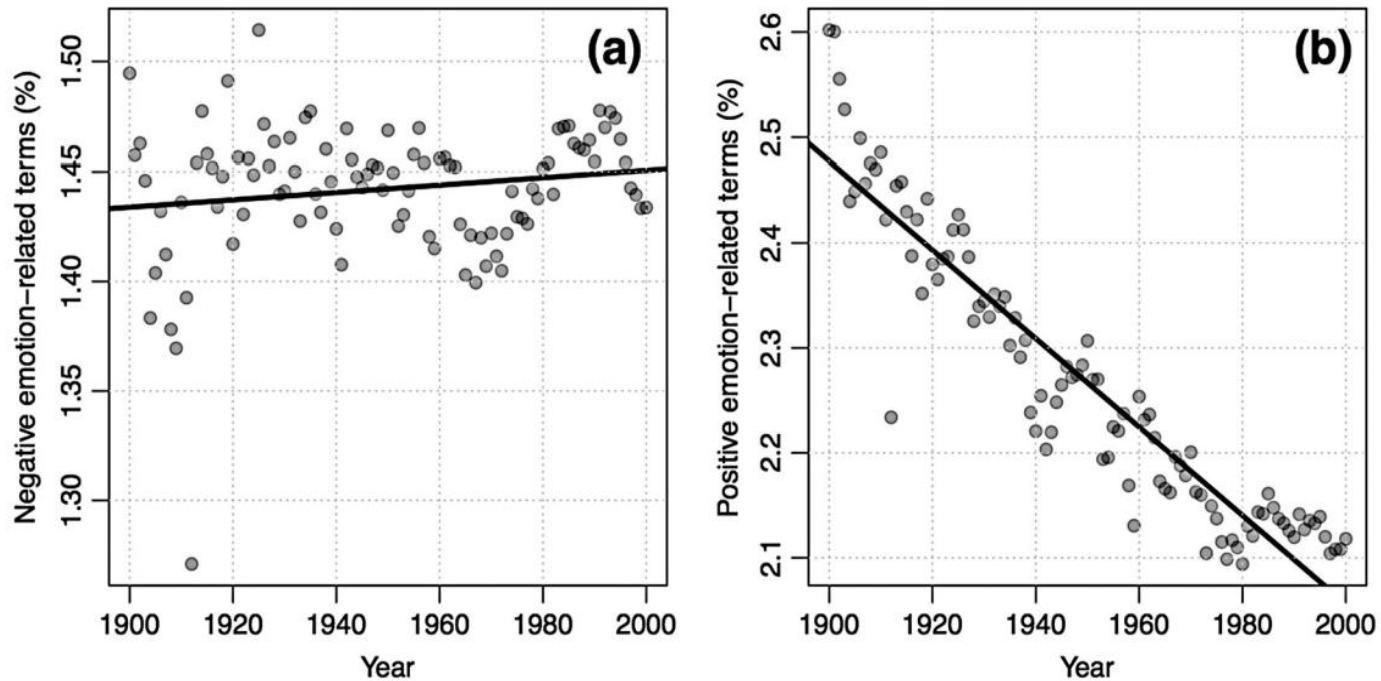
out in texts



BY MICHAEL PETER TEAM AND MICHAEL PETER TEAM  
FROM THE MICHAEL PETER TEAM AND MICHAEL PETER TEAM

Reagan et al. 2016 The emotional arcs of stories are dominated by six basic shapes  
David Robinson 2017, Examining the arc of 100,000 stories: a tidy analysis

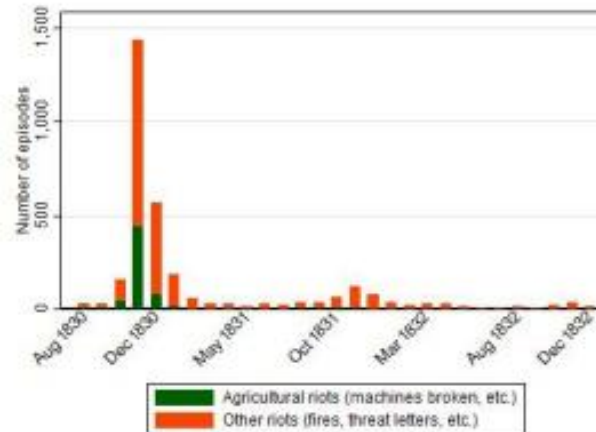
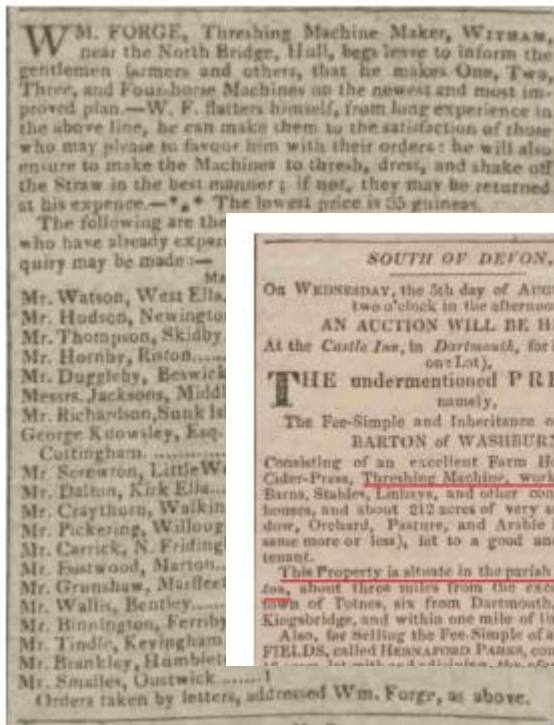
# Birth of the cool



**Figure 3.** Emotionality changes in Anglophone literature, for the Google Books corpus. (a): negative emotions-related terms. (b): positive emotions-related terms. Solid lines represent linear regressions of the data.

# Riots against the machines 1830-1832

Diffusion of technology from newspaper texts -> more precise history.

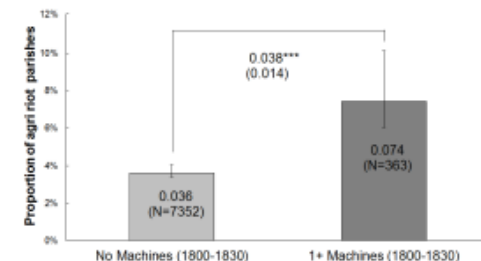
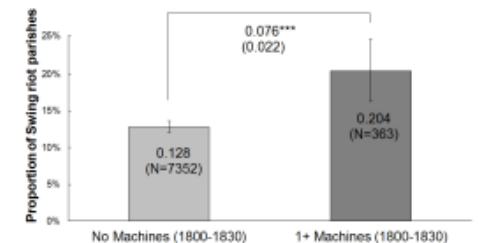


## RAGE AGAINST THE MACHINES: LABOR-SAVING TECHNOLOGY AND UNREST IN ENGLAND, 1830-32\*

Bruno Capretini  
University of Zurich

Hans-Joachim Voth  
University of Zurich and CEPR

First version: December 2016  
This version: January 2017



# Economic progress from interest in encyclopedias

Ordering of encyclopedia and economic growth of cities

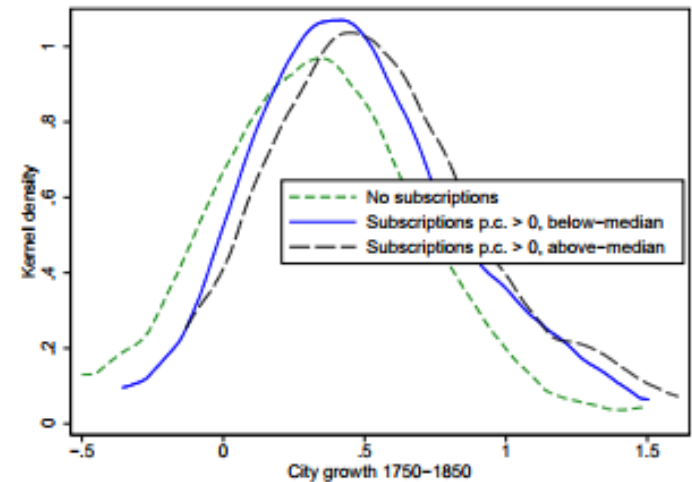
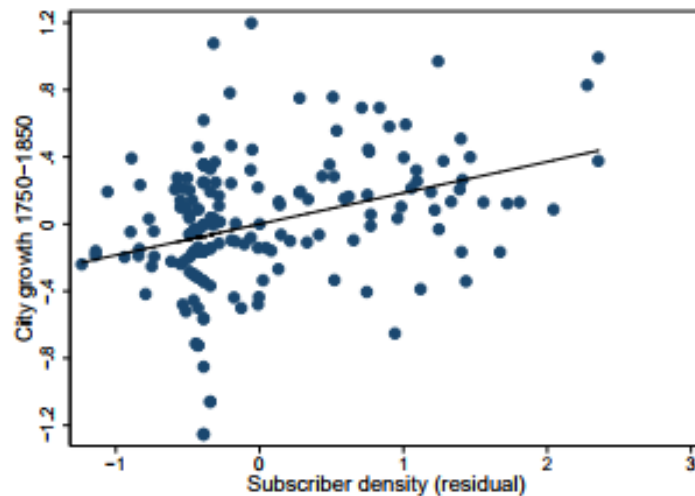
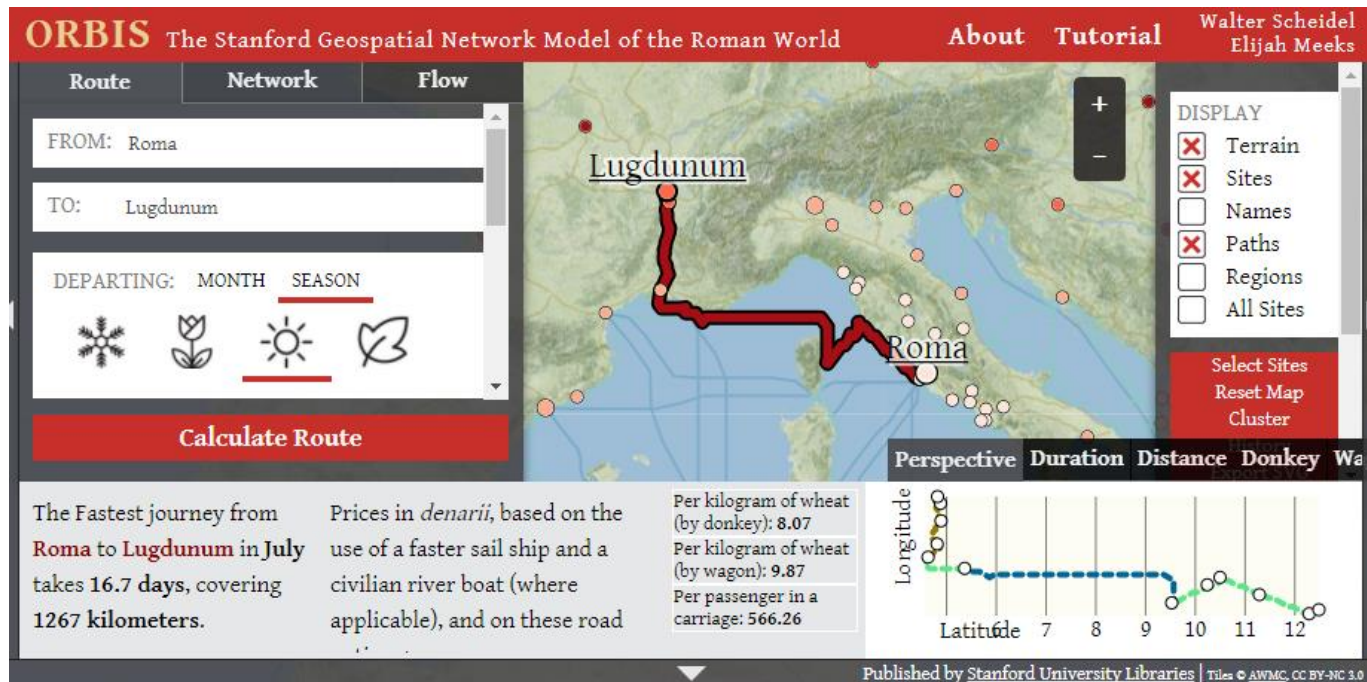


Figure 3: Subscriptions and city growth, 1750-1850

Figure 2: *Encyclopédie* subscriptions and city growth, 1750-1850

# Interactive apps

Explore the travelling roads in Ancient Europe



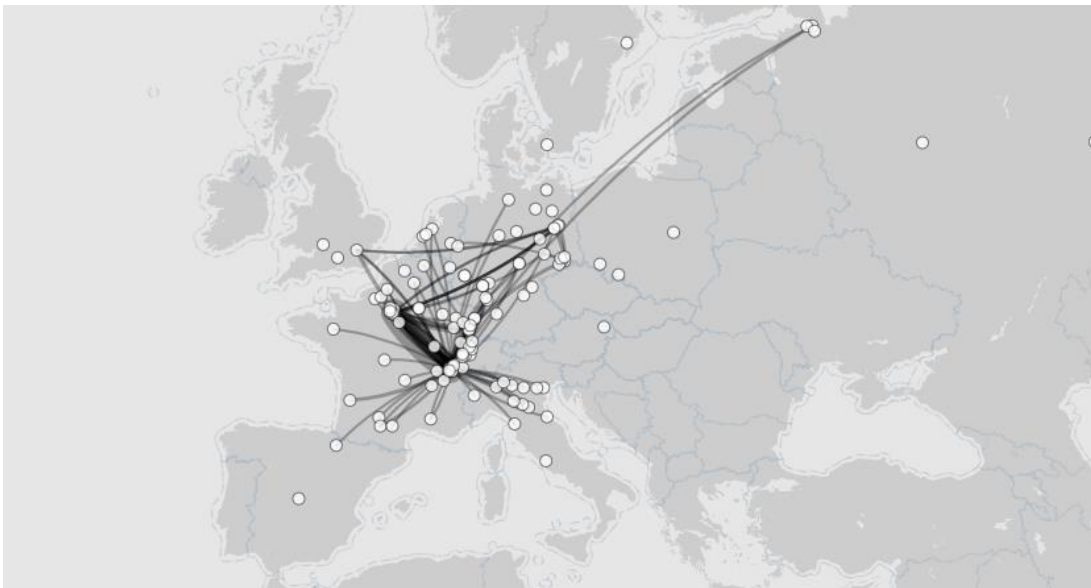
<http://orbis.stanford.edu/>



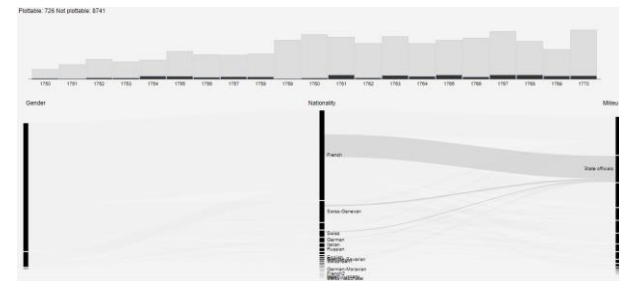
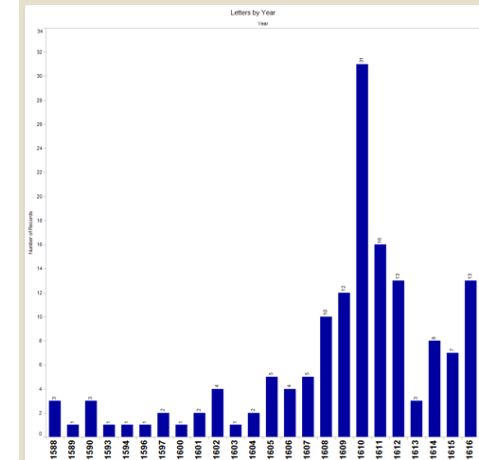


# Historical social networks

The connections within the Republic of Letters.



Number of letters sent by Galileo per year



Voltaire's correspondence network visualized by Mapping the Republic of Letters' RPLVIZ tool.  
<http://ink.designhumanities.org/voltaire/>

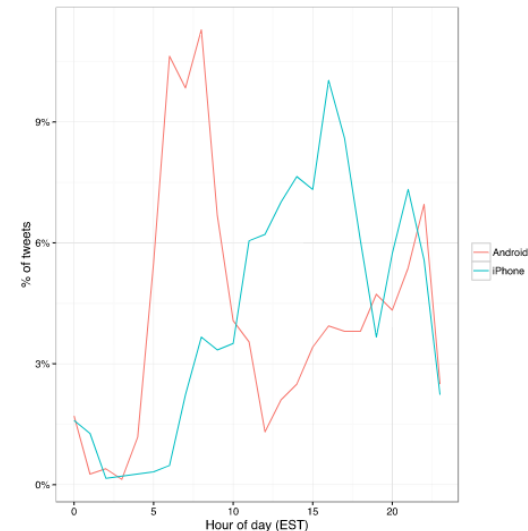
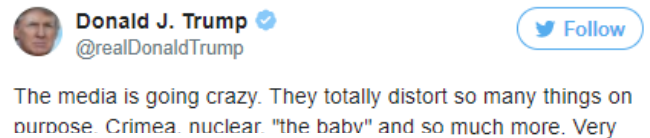
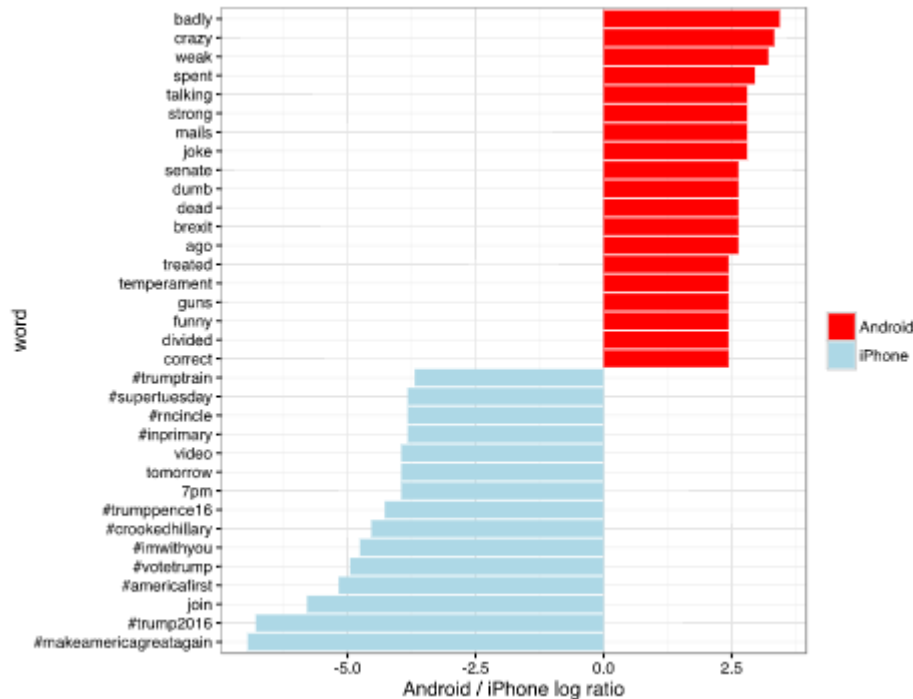
Some of the Republic of letters tools were used in studying the panama papers





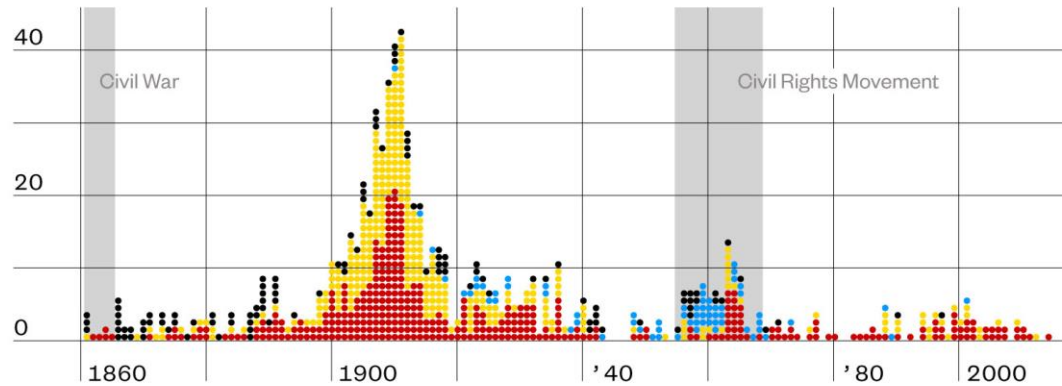
# Text analysis of Trump's tweets confirms he writes only the (angrier) Android half

I don't normally post about politics (I'm not particularly savvy about polling, which is where data science has had the largest impact on politics). But this weekend I saw a hypothesis about Donald Trump's twitter account that simply begged to be investigated with data:



## Confederate iconography came long after the Civil War

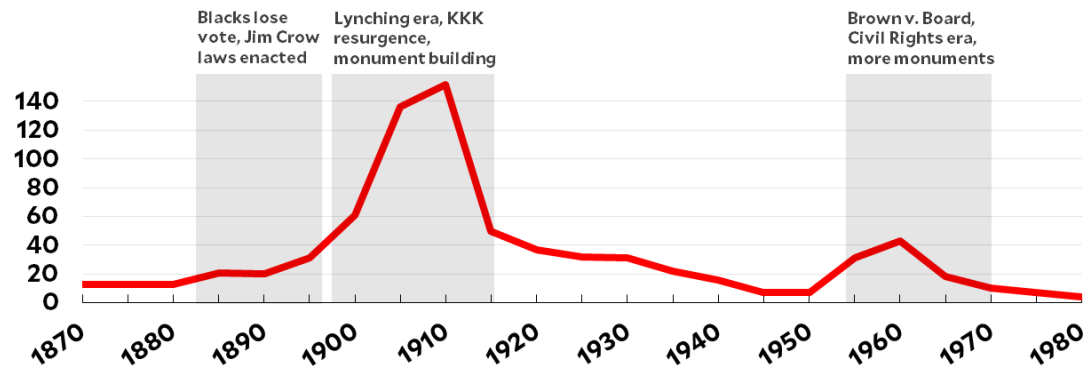
● Monument ● Monument on courthouse grounds ● School ● Other



Source: As of April 2016, Southern Poverty Law Center

## Confederate Monument and Statue Building

By Five-Year Periods, 1870-1980

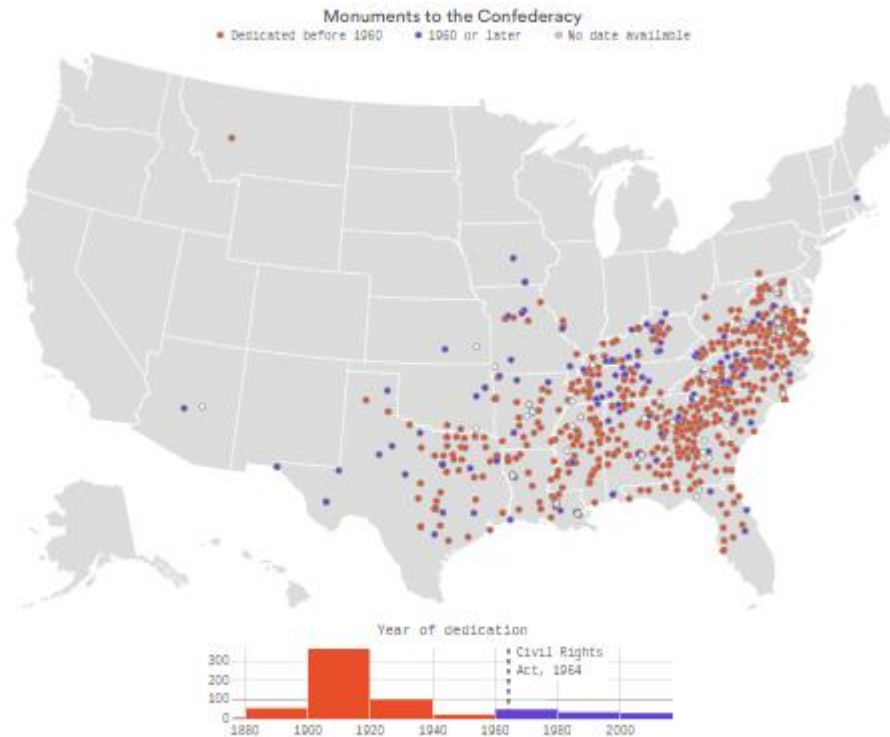


Source: Southern Poverty Law Center

Mother Jones

# Modern use

---



Note: Map does not include monuments for which location data was not available; Data: Southern Poverty Law Center; Graphic: Chris Canipe / Axios

# What is it

---

A bit of a mess, but with good momentum.

Has grand goals, and a has a warm hearth.

Hard work but with many possible benefits.



# What else

---

## The openness attitude

- Keep your work open, usually can only help

## Using tools for timesavers

- Takes time, and looking for one is a risk, but finding the right one is worth it.

## It's all online

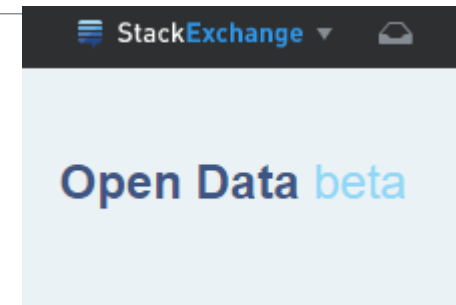
- Easy to connect and ask for help
- Easy to look for others working on similar topics

# How to connect



# adminannouncements  
# annotation  
# announcements  
# beginners  
# codeofconduct  
# coding  
# conferences

**Digital Humanities Slack**  
[TinyUrl.com/DHSlack](https://tinyurl.com/DHSlack)



StackExchange ▾

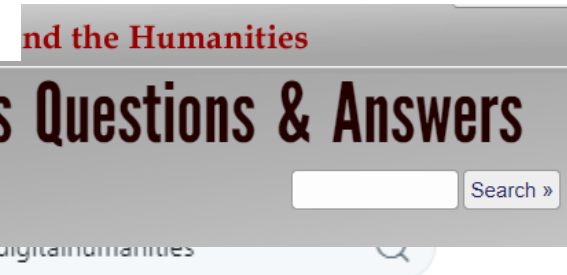
Open Data beta

## The Programming Historian



goodreads

**Popular Digital Humanities Books**  
(showing 1-50 of 171)



nd the Humanities

**Digital Humanities Questions & Answers**

Ask a Question

Search »

Home Notifications

Messages

#digitalhumanities

**#digitalhumanities**

# That's all

---

Thanks for listening!