

ILLUMINATING THE ILLUMINATED

Dr. Joss Wright^{1,2}

¹Oxford Internet Institute
<https://www.oii.ox.ac.uk>

²Weird Data Science
<https://www.weirddatascience.net>

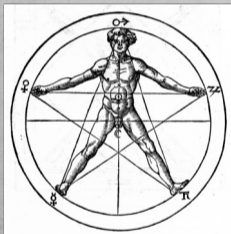
ARCANE HISTORY

- The history of the written word is a history of magic, secrets, symbolism and the occult.
- From earliest recorded divination, through symbols and hieroglyphs, encoded and enciphered information has been preserved and protected.
- Late medieval Europe, and the transition to the Renaissance, was rife with new philosophies, discoveries, and experimentation.
- Individuals and societies explored new ways of thinking, and inscribed their ideas, practices, rituals.



Hermes Trismegistus - 17th Century Print

OCCULT WRITING



Heinrich Cornelius Agrippa -
Pentagram and the Human Body.
from "De Occulta Philosophia
libri III"

- Many notable individuals in Europe both explored the world in new ways, and left their writings behind.
- Heinrich Cornelius Agrippa - "De Occulta Philosophia libri III" (1533) was foundational in hermetic philosophy and the dawn of science.
 - ▶ Alchemy. Astronomy. Magic and religion.
 - ▶ Student, alongside alchemist and physician Paracelsus, of Johannes Trithemius.
- John Dee - alchemist and court astronomer to Elizabeth I.
 - ▶ Recorded his conversations with angels in an unknown "angelic" script, later termed Enochian.
 - ▶ Owned a copy of the enciphered "Book of Soyga" (Bodley MS. 90B), a treatise on magic, demonology, astrology, the names of angels, and others.

CRYPTIC FOUNDATIONS

- Hiding the meaning of occult works was well-established.
 - ▶ Caesar's cipher of antiquity.
 - ▶ Alchemical and other magical works guarded with extreme secrecy.
 - ▶ Secret societies - Copiale cipher of the Oculists. (1730).
Deciphered in 2011.
- Johannes Trithemius (1462-1516), German Benedictine abbot, wrote his *Steganographia* and *Polygraphia* in, and concerning, cryptography.
 - ▶ Third volume of *Steganographia* not fully decoded until 1998.
(Reeds)
- al-Kindi (801CE, Baghdad) in his "Manuscript on Decyphering Cryptographic Messages" (841CE) details means to decrypt ciphers via frequency analysis of texts.
 - ▶ Included methods to analyse ciphers widely considered unbreakable until the 20th century in Europe.



Trithemius - Chart from "Steganographia".
Feniarth MS 423p. in the hand of John Dee.

LIBER EX MACHINA

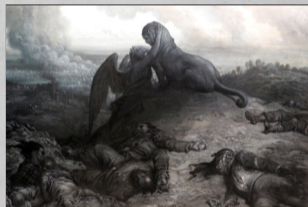


Wilfrid Voynich. (1865-1930)

- Wilfrid Voynich (1865-1930).
 - ▶ Born in modern-day Lithuania, then in the Russian Empire.
 - ▶ Escaped penal servitude in Siberia following revolutionary activities, eventually becoming an antiquarian, and highly successful rare book dealer in London and later New York.
- Discovered the Voynich Manuscript during the sale of holdings by the Jesuit Monastery at the Villa Mondragone, Frascati, in 1912.
- Kept the manuscript until his death. It now resides in Yale's Beinecke Rare Book and Manuscript Library. (MS 408)

INK AND SKIN

- Small volume illustrated with depictions of plants, herbs, human figures, astronomical, and astrological symbols.
- Almost entirely in an unknown script, barring a small number of words apparently in Latin and High German.
 - ▶ Approximately twenty-five glyphs in the script.
 - ▶ Paragraph construction indicates a left-to-right orthography.
- Has resisted all attempts at interpretation by cryptographers, historians, and linguists.



Dore - The Enigma (1871)



Handwritten text in a circular border around the diagram on the right page. The text is written in a cursive script and appears to be a commentary or explanation of the diagram's elements.



Handwritten text in a cursive script, likely a commentary or explanation of the diagrams on the adjacent pages. The text is arranged in several columns, following the layout of the illustrations.

A HIDEIOUS AUTHENTICITY



Domenico Fetti - Portrait of a
Scholar

- Scholars have debated the Voynich Manuscript's provenance, authenticity, meaning, and linguistic properties for over a century.
 - ▶ Likely mention by Athanasius Kircher, 1639, replying to Georgius Barschius of Prague, via Theodor Moretus. Refers to a "libellum steganographici mysterisi".
 - ▶ Later scholars have claimed various languages, and have delineated the manuscript into chapters based on diagrams, but no robust decryptions have been shown.
- Radio carbon dating has shown the manuscript to be constructed from contemporary 15th Century materials.

TREMULOUS QUERIES

- Is the manuscript a genuine text, or an artificial or meaningless text?
- Do its linguistic features support the apparent chapters suggested by the diagrams?
- Does the text reveal any patterns in underlying linguistic shifts?



Van Bentam - Explosion in the Alchemist's Laboratory. (c.1620)

IPSA SCIENTIA POTESTAS EST

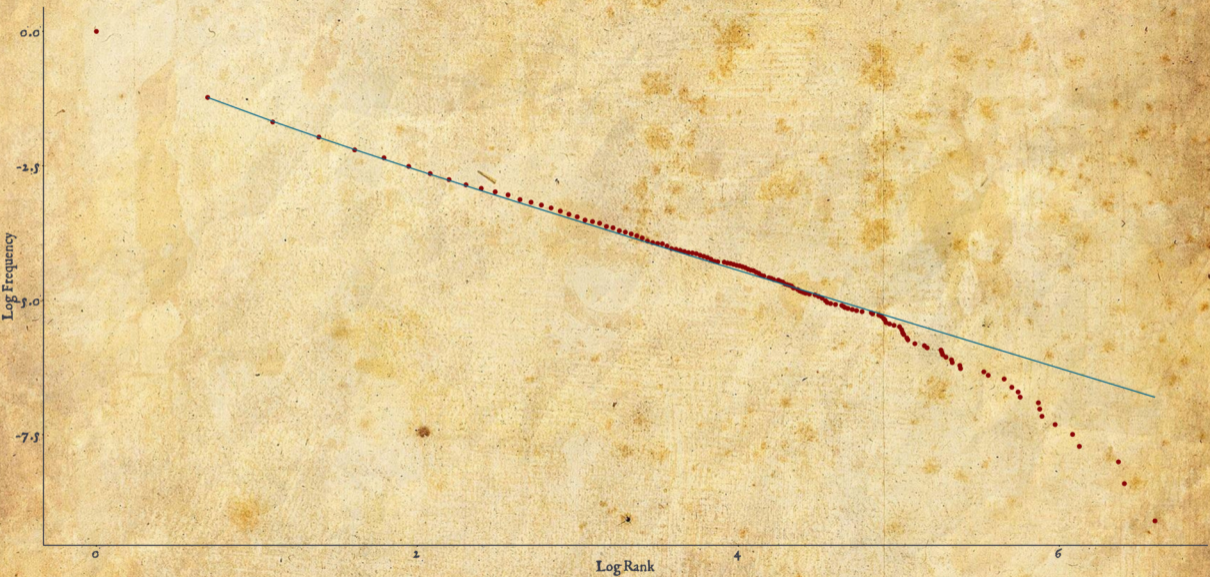


Atlas bearing the heavens in the form of an armillary sphere. from William Cunningham. The Cosmographical Glasse. London 1559.

- Human languages are not random, they follow surprisingly consistent statistical patterns.
- Zipf's Law (1932) identified power law distributions word frequencies of human languages.
 - ▶ Frequency $\propto \frac{1}{\text{Rank}}$
 - ▶ The most frequent word is twice as frequent as the second most frequent word, and so on.
- Often evidenced by linearity on a log-log plot, although this demonstrates a lack of evidence against, rather than evidence for, a power law distribution.

Fitted Power Law of Voynich Corpus

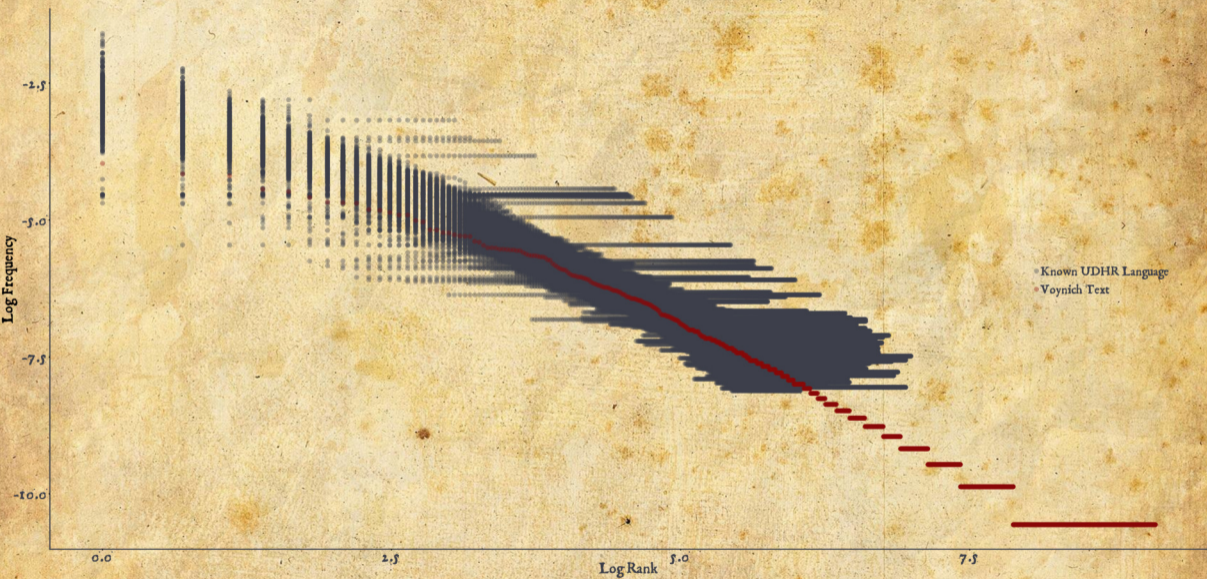
<http://www.weirddatascience.net> | @WeirdDataSci



Data: <http://www.voynich.nu>

Voynich Manuscript Rank Frequency Distribution against UNDHR Translations

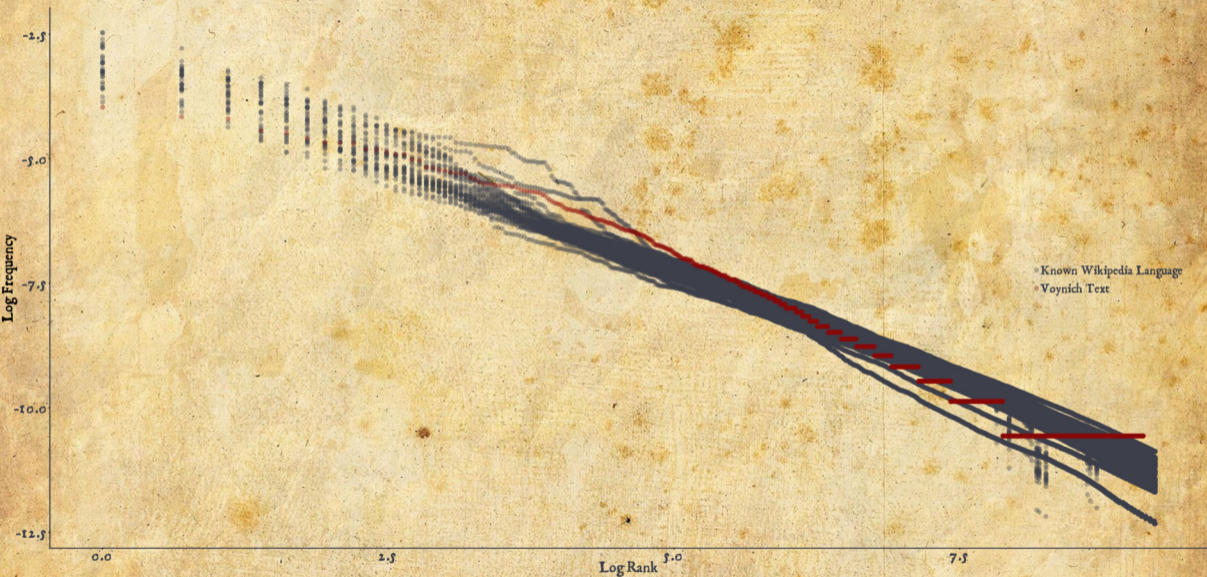
<http://www.weirddatascience.net> | @WeirdDataSci



Data from: <http://www.voynich.nu> | <http://unicode.org/udhr/>

Voynich Manuscript Rank Frequency Distribution against Wikipedia Corpora

<http://www.weirddatascience.net> | @WeirdDataSci



Data from: <http://www.voynich.nu> | <http://wikipedia.org>

SPURIOUS ACCUSATIONS

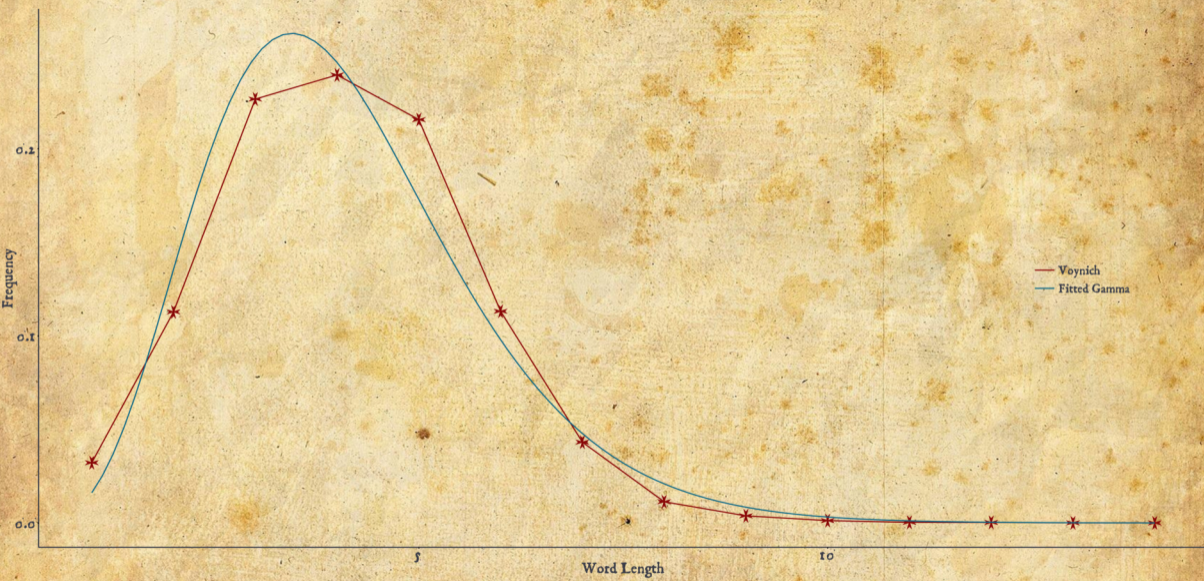


Dor - Untitled illustration from
Cinderella from the book
Les Contes de Perrault (1662)

- A random distribution of letters, plus a space, also results in a Zipf-like distribution.
- Human languages, however, often display certain deviations from power laws.
- Sigurd, Eeg-Olofsson, and van Weijer, (2004), demonstrate that shortest words are not the most common.
 - ▶ For English, Swedish, and German, words between 3 and 5 letters in length dominate.
 - ▶ Approximated by an appropriately parameterised Gamma distribution.

Word Length Frequency Plot of Voynich Corpus

<http://www.weirddatascience.net> | @WeirdDataSci



Data: <http://www.voynich.nu>

RECKLESS ASSUMPTIONS



Dor -- The Inferno. Canto XXI.
"Be none of you outrageous."

- Without uncovering the meaning of the Voynich text, we can never be sure that the manuscript is not a hoax.
- Statistical properties of the text could result from chance, but these properties were not known at the time of the manuscript's apparent creation.
- Physical and statistical evidence points strongly to its authenticity.

LINEAR DISCRIMINATION

- Pseudo-chapters have been identified by scholars due to the subject of the illustrations.
 - ▶ Astronomical. Astrological. Herbal. Pharmaceutical.
- Do the properties of the text support a topical separation of the manuscript?
- Apply `_topic modelling_` to classify folios in the manuscript.



Louis Le Breton - Stolas, from
de Flanzy's "Dictionnaire
infernal". (1263)

LINEAR DISCRIMINATION



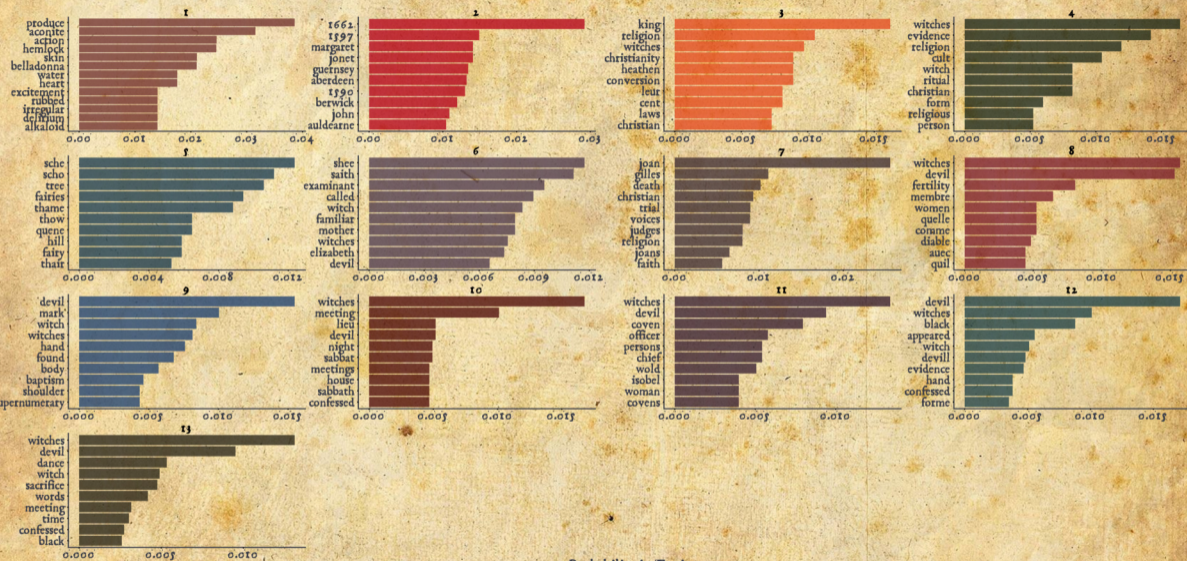
Buer, the "demon who teaches
Moral and Natural Philosophy".

- An unsupervised machine learning method.
- Models each unit, such as page or chapter, as a bag of words, drawn at random.
- Allows us to classify each page by sorting into the 'bag' that matches it best.
- Used to categorise and cluster large volumes of text.
 - ▶ Without having to read the words.

"The Witch Cult in Western Europe" (Murray, 1921) Topic Model

<http://www.weirddatascience.net> | @WeirdDataSci

Term



Probability in Topic

Voynich Manuscript Topic Model (12 Topics)

<http://www.weirddatasci.net> | @WeirdDataSci

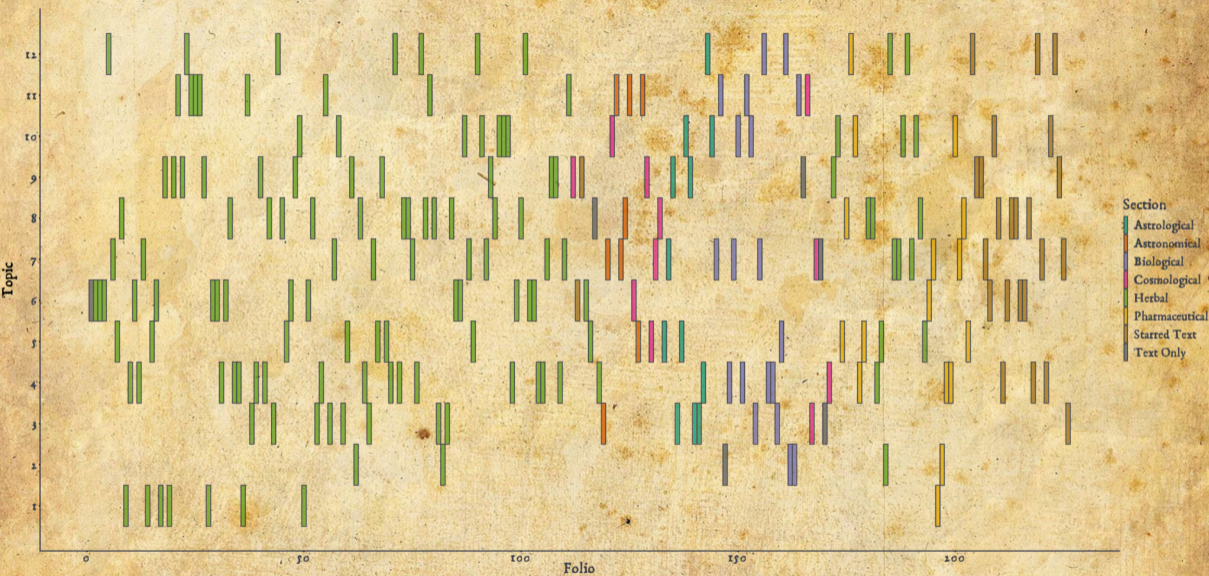
Term



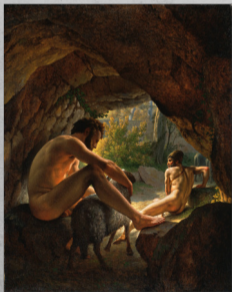
Probability in Topic

Voynich Manuscript Topic Heatmap

<http://www.weirddatascience.net> | @WeirdDataSci



TOPICAL REMEDIATION



Christoffe Wilhelm Eckersberg -
Ulysses Fleeing the Cave of
Polyphemus (1812)

- As expected, pages fall into different categories.
- Patterns of topics do suggest shifting subject focus throughout the manuscript.
- Does not strongly mirror visually identified chapters, but some minor clustering apparent.
- Dig deeper.

TEMPORA MUTANTUR

- To rend the veil further, draw on tools from `_stylometry_`, combined with a Bayesian statistical approach to `_changepoint analysis_`.
- Scrutinize word length frequency to identify significant shifts in textual content of the manuscript.
- Stylometrics assesses raw textual features to reveal subtle variations in use of language.
 - ▶ Word frequency, particularly of short words.
 - ▶ Usage of the letter "i". (Greenstadt et al.)



Gustave Doré - The Adventure with
the Windmills, from
Illustrations to Don Quixote.
(1863)

POST EQUITEM SEDET



¶ Los cinco libros de effocabo e inuencible cauallero
 Tirante el blanco de roca (alaba: Cauallero dela Ba-
 rroterra. El qual por su otra caualleria alcágo a fer por
 cipe e celar del imperio de greca.

Frontispiece of the first
 Castilian-language
 translation of Tirant lo
 Blanc, printed by Diego de
 Gumiel. (1511)

- Exemplar: Tirant lo Blanc - a chivalric romance written by the Valencian knight Joanot Martorell. (1490)
- Posthumously completed and published by Marti Joan de Galba.
- Scholarly debate concerning the extent of Galbas contributions.
- A 2006 analysis (Riba & Ginebra) examined frequency of word lengths to indicate a shift in authorship.
 - ▶ Identified a stylistic shift between chapters 371 and 382.

DEFINED CONSTRAINTS

- Consider the text as a multinomial distribution.
- Word length as the category, number of words per page is the number of trials.
- Model an entire text as two multinomial distributions, with two sets of parameters.
 - ▶ Where in the text do two fitted multinomial distributions meet such that they best describe the overall?



Giovanni Battista Piranesi -
Carceri. Folder 7 (1745)

TEMPORAL FRACTURES



Gustave Doré - The Deluge

Formally:

$$\Omega \sim \text{Multinomial}(t < c? \theta_e, \theta_1)$$

$$\theta_e \sim \text{Dirichlet}(\alpha)$$

$$\theta_1 \sim \text{Dirichlet}(\alpha)$$

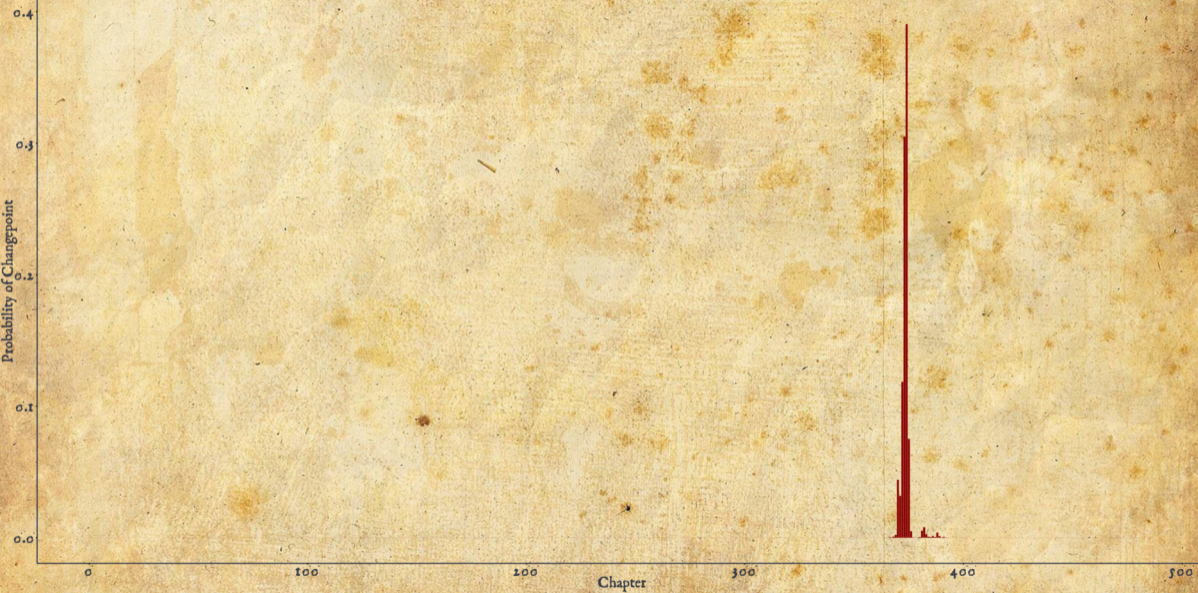
$$c \sim \text{Uniform}(1, t)$$

$$\alpha = [0.6, \dots]_{[T]}$$

Implemented in the Stan probabilistic programming language for Bayesian model fitting and assessment.

Tirant lo Blanc Chapter Change Probability

<http://www.weirddatascience.net> | @WeirdDataSci



Panta Rhei

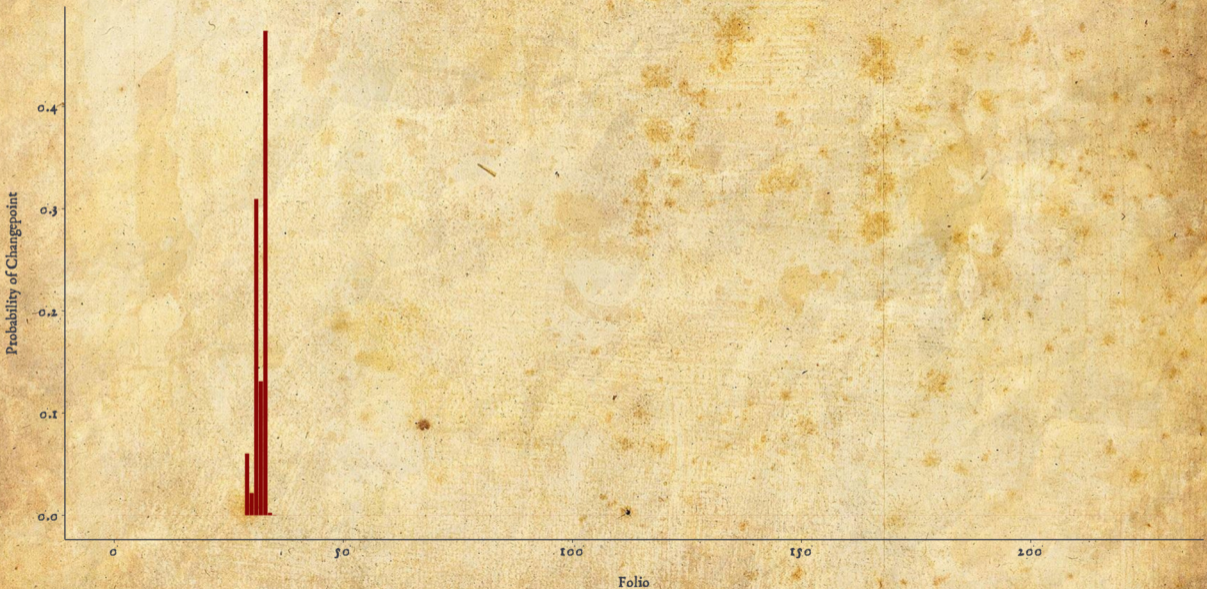


Buer, the "demon who teaches
Moral and Natural Philosophy".

- The model reproduces the findings of a known authorship shift in Tirant lo Blanc.
- Does it find significant shifts in the Voynich?

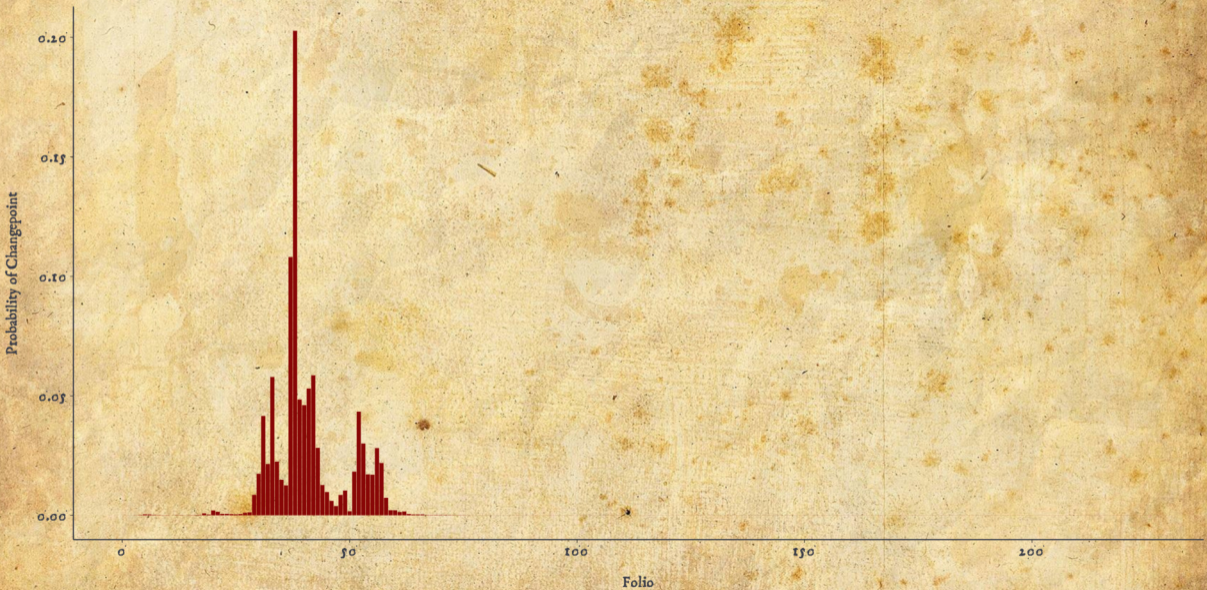
Voynich Folio Word-Length Changepoint

<http://www.weirddatascience.net> | @WeirdDataSci



Voynich Folio Changepoint Probability - Topic Model (12 topics)

<http://www.weirddatascience.net> | @WeirdDataSci



Handwritten text in a medieval script, likely a form of Gothic or similar, arranged in several lines at the top of the page.



Handwritten text in a medieval script, arranged in several lines at the top of the page, interspersed with diagrams.



HIDEOUS CONCLUSIONS

- Perhaps the statistics were the real horror all along.
- Statistical and physical evidence lends itself to credence in the Voynich Manuscript's authenticity.
- New evidence for a significant textual shift early in the Voynich Manuscript.
 - ▶ Extension to multiple changepoints.
- Unknown terrors detailed in cryptic glyphs within arcane manuscripts uncovered in ancient monasteries are not free from our scrutiny.



Dore - The Inferno, Canto XXXIV. Satan,
trapped in the frozen central zone in
the Ninth Circle of Hell.