

# Community Data Science meets Digital Heritage

## Controversies about World Heritage Sites in the Online Community of Wikipedia

### Controversy [\[edit\]](#) [\[hide\]](#)

Controversy resolved and moved to the footnote. — [Malaiya](#) (talk • contrib • actions • log • sig) 02:39, 27 November 2008 (UTC)

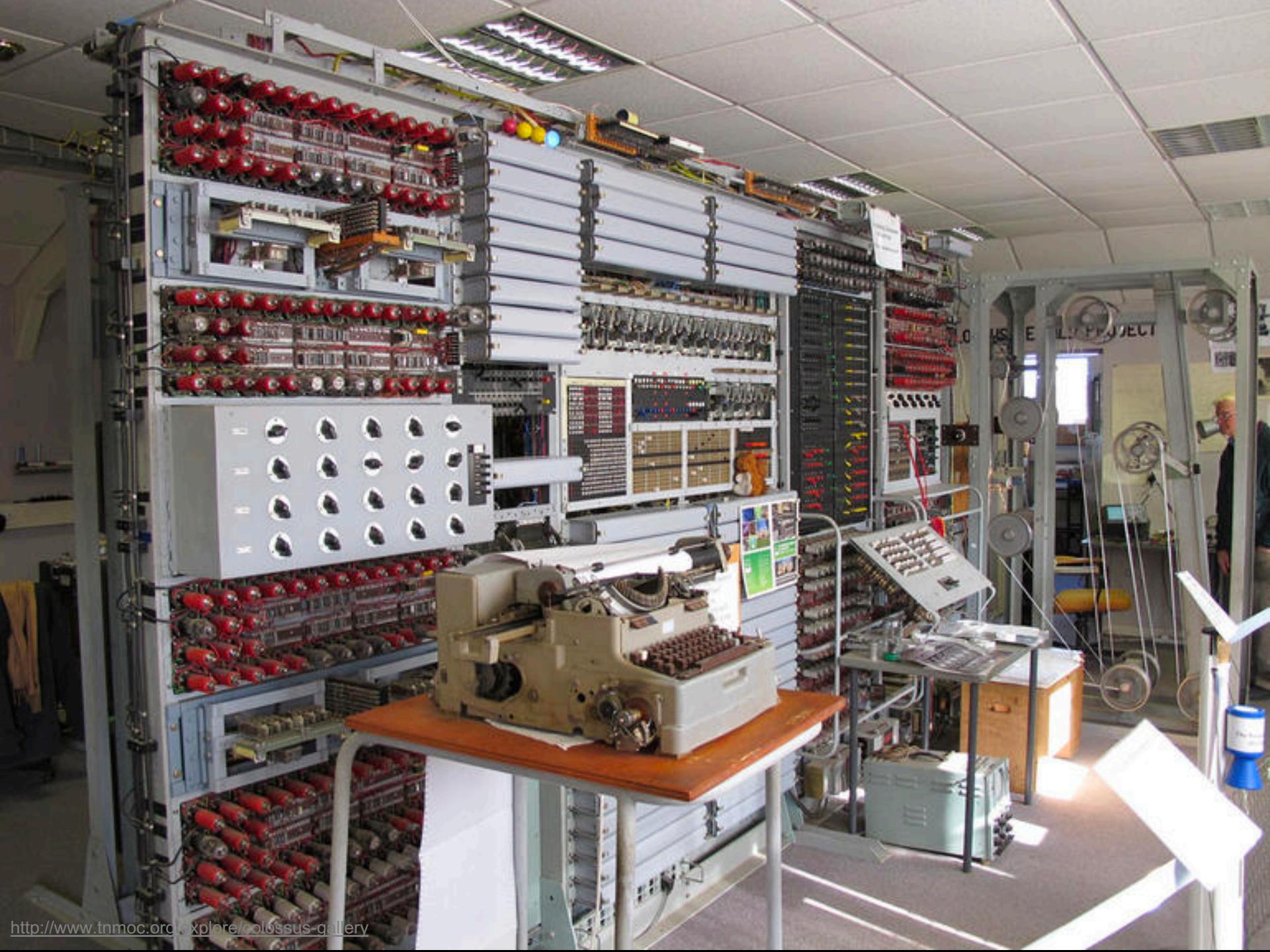
Ben Marwick  
Prema Smith  
University of Washington  
@benmarwick

Digital Heritage in a World of Big Data  
13 and 14 May 2019, at University of Stirling

#HeritageBigData



# Digital Heritage



- Heritage as the processes and outcomes of engaging with elements of the past – material and immaterial – and attributing social and cultural meanings to them in the present (Smith 2006; Harrison 2013)
- Digital heritage as the interactions of this kind that are enabled by the Internet and the traces of such processes (Bonacchi and Krzyzanska 2019)
- Represents a major turn from post-modernism, critical theory, & deeply qualitative, towards digital heritage ontologies, data-intensive ethnographies, role of heritage scholars

**UNESCO**  
**World**  
**Heritage**



- UNESCO established in 1945 “to end global conflict and help the world rebuild materially and morally.” Since 1972 World Heritage Convention has had a remit to protect world heritage and archaeology
- Currently 845 cultural sites on the World Heritage List
- WHL is much less about preserving heritage than about political and economic transactions, sites are “a proxy for territory, sovereignty, and security as well as international corporate agreements” (Meskell 2016)
- What sites have captured scholarly attention, and why?



- The 1992 inscription of Angkor (Cambodia) was supported by exiled sympathizers of the genocidal Khmer Rouge regime, hoping to bolster territorial claims
- Mapungubwe (2003, South Africa), state-supported mining with BRICS led to the prioritisation of development goals over cultural assets
- Timbuktu (Mali), extensive destruction of shrines by rebels in 2012 left the WHC gridlocked and frustrated by their inability to help, similarly in Kandy (Sri Lanka) , Palmyra (Syria), etc.
- Preah Vihear (2008) on a disputed part of Thai-Cambodia border, WikiLeaks documents the hyperconnectivity of international actors



<https://news.stanford.edu/2018/11/19/stanford-scholar-examines-unesco-world-heritage-program/>

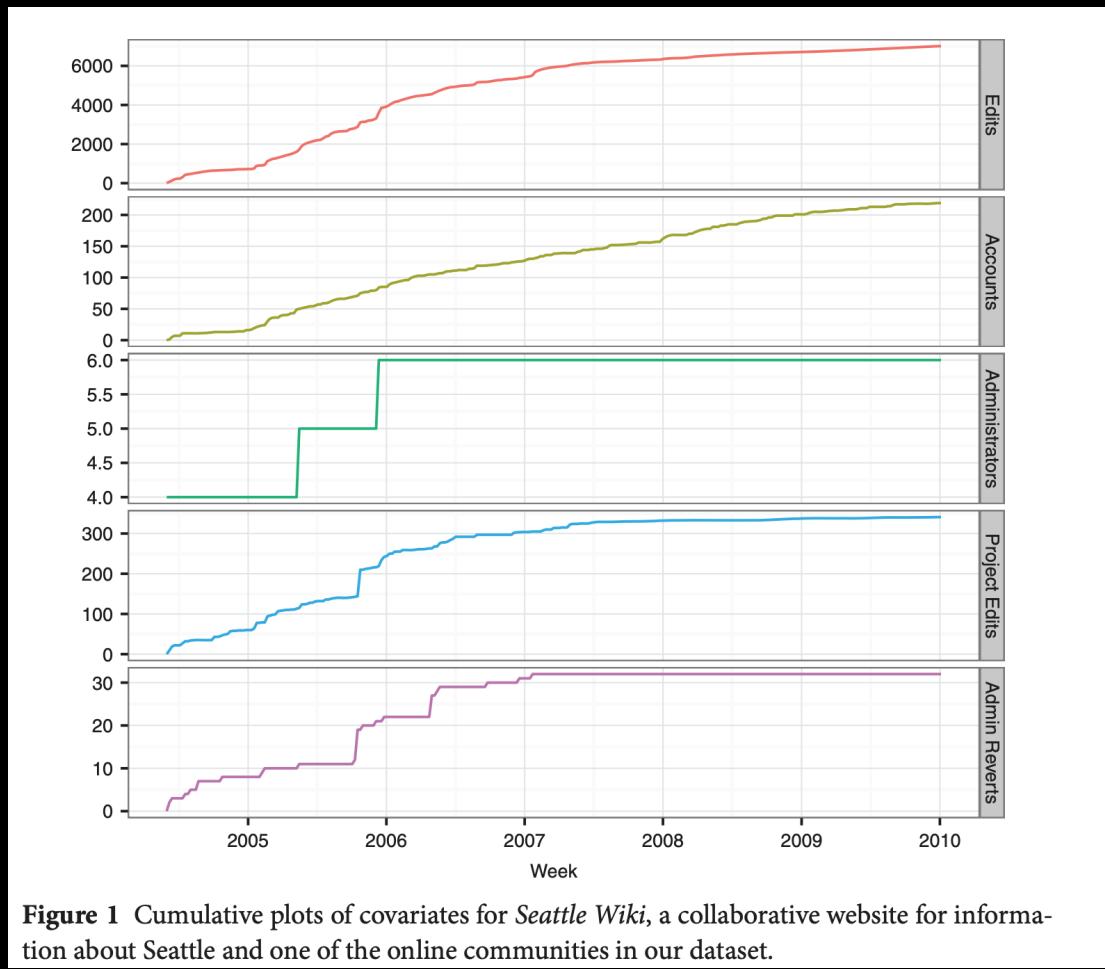
# **Community Data Science**



<https://wiki.communitydata.science>



- Quantitative and qualitative methods **to the study of online communities**
- Focused on how the design of communication and information technologies shape fundamental social outcomes with broad theoretical and practical implications
- “big data” quantitative analyses at the intersection of communication, sociology, and human-computer interaction.



**Figure 1** Cumulative plots of covariates for *Seattle Wiki*, a collaborative website for information about Seattle and one of the online communities in our dataset.

JOURNAL OF  
COMMUNICATION

Journal of Communication ISSN 0021-9916

ORIGINAL ARTICLE

## Laboratories of Oligarchy? How the Iron Law Extends to Peer Production

Aaron Shaw<sup>1</sup> & Benjamin M. Hill<sup>2</sup>

1 Department of Communication Studies, Northwestern University, Evanston, IL 60208, USA  
 2 Department of Communication, University of Washington, Seattle, WA 98195, USA

Peer production projects like Wikipedia have inspired voluntary associations, collectives, social movements, and scholars to embrace open online collaboration as a model of democratic organization. However, many peer production projects exhibit entrenched leadership and deep inequalities, suggesting that they may not fulfill democratic ideals. Instead, peer production projects may conform to Robert Michels' "iron law of oligarchy," which proposes that democratic membership organizations become increasingly oligarchic as they grow. Using exhaustive data of internal processes from a sample of 683 wikis, we construct empirical measures of participation and test for increases in oligarchy associated with growth in wikis' contributor bases. In contrast to previous studies, we find support for Michels' iron law and conclude that peer production entails oligarchic organizational forms.



## Welcome to [Wikipedia](#),

the free encyclopedia that anyone can edit.

5,853,913 articles in English

### From today's featured article



Hypothetical depiction

**Paranthodon** was a [stegosaurian](#) dinosaur that lived in present-day [South Africa](#) between 139 and 131 [million years ago](#), during the [Early Cretaceous](#).

Discovered in 1845, it was one of the first stegosaurs found. The only remains of the [genus](#), a partial skull, isolated teeth, and fragments of vertebrae, were found in the [Kirkwood Formation](#). British paleontologist [Richard Owen](#) initially identified the fragments as those of the [pareiasaur Anthodon](#). After remaining untouched for years in the [British Museum of Natural History](#), the partial skull was identified by South African

paleontologist [Robert Broom](#) as belonging to a different genus; he named the specimen [Palaeoscincus africanus](#). Several years later, Hungarian paleontologist [Franz Nopcsa](#), unaware of Broom's new name, similarly concluded that it represented a new [taxon](#), and named it [Paranthodon owenii](#). The genus name combines the [Ancient Greek](#) *para* (near) with the genus name *Anthodon*, to represent the initial referral of the remains. ([Full article...](#))

Recently featured: "[Faces](#)" (*Star Trek: Voyager*) · [Maine Centennial half dollar](#) · [Myst III: Exile Archive](#) · [By email](#) · [More featured articles](#)

### Did you know...

- ... that [Charles Darwin](#) contributed funds to the building of [New Zealand's oldest surviving church](#) (pictured)?
- ... that despite spending the majority of his career with the [Russian Northern Fleet](#), [Aleksandr Moiseyev](#) was appointed commander of the [Black Sea Fleet](#) in 2018?
- ... that [Martin Luther](#) wrote "[Nun komm, der Heiden Heiland](#)" as a close paraphrase of the Latin hymn "[Veni redemptor gentium](#)", with a melody derived from the [plainchant](#)?
- ... that, inspired at a young age by Polish [footballer Roman Kosecki](#) of [Galatasaray S.K.](#), in Turkey, [Nurcan Çelik](#) became a player and later established her own football club?
- ... that the [Sword of State of South Carolina](#) is listed in the [FBI National Stolen Art File](#)?
- ... that after being influenced by [Wahhabism](#), [Mubarez-ud-Daulah](#) wanted to overthrow the [Nizam of Hyderabad](#) and the British in India?
- ... that the 1897 [supernatural novel](#) [Gulabsinh](#) by [Gujarati-language](#) writer [Manilal Dwivedi](#) is an adaptation of the mystical novel [Zanoni](#) by English writer [Edward Bulwer-Lytton](#)?
- ... that [Tim Morgan](#)'s *Veins* are made of steel and glass?



Christ Church in Russell, New Zealand

[Archive](#) · [Start a new article](#) · [Nominate an article](#)

- [Arts](#)

- [Biography](#)

- [Geography](#)

- [History](#)

- [Mathematics](#)

- [Science](#)

- [Society](#)
- [Technology](#)
- [All portals](#)

### In the news

• [Judd Trump \(pictured\)](#) wins the [World Snooker Championship](#), defeating [John Higgins](#) in the final.



Judd Trump

• The [United Nations](#) reports that, due to human activity, the Earth's [biodiversity](#) has suffered a catastrophic decline unprecedented in [human history](#).

• At least 29 people are killed in [clashes at the Gaza-Israel border](#).

• [Aeroflot Flight 1492](#) crash-lands and bursts into flames at [Sheremetyevo International Airport](#), Moscow, killing 41 of the 78 people on board.

**Recent deaths:** [Yuan Baohua](#) · [Sprent Dabwidio](#) · [David Montgomery](#) · [Georg Katzer](#) · [Jean Vanier](#) · [Seymour Nurse](#)

[Other recent events](#) · [Nominate an article](#)

### On this day

#### May 11

• [868](#) – A copy of the [Diamond Sutra](#) (frontispiece pictured) was printed in China, making it the world's oldest dated complete printed book.



Frontispiece of the *Diamond Sutra*

• [1745](#) – [War of the Austrian Succession](#): French forces defeated those of the [Pragmatic Allies](#) at the [Battle of Fontenoy](#), in the [Austrian Netherlands](#) in present-day Belgium.

• [1889](#) – An [attack upon a U.S. Army paymaster and escort](#) resulted in the theft of over \$28,000 and the award of two [Medals of Honor](#).

• [1997](#) – [Deep Blue](#) became the first [computer](#) to win a match against a [world chess champion](#), when it defeated [Garry Kasparov](#) in [six games](#).

• [2013](#) – Two car bombs by unknown perpetrators [exploded](#) in [Reyhanlı](#), Turkey, resulting in 52 killed and 140 injured.

[Johann Friedrich Blumenbach](#) (b. 1752) · [Juliette Récamier](#) (d. 1849) · [Zenna Henderson](#) (d. 1983)

More anniversaries: [May 10](#) · [May 11](#) · [May 12](#)

[Archive](#) · [By email](#) · [List of historical anniversaries](#)

### Today's featured picture

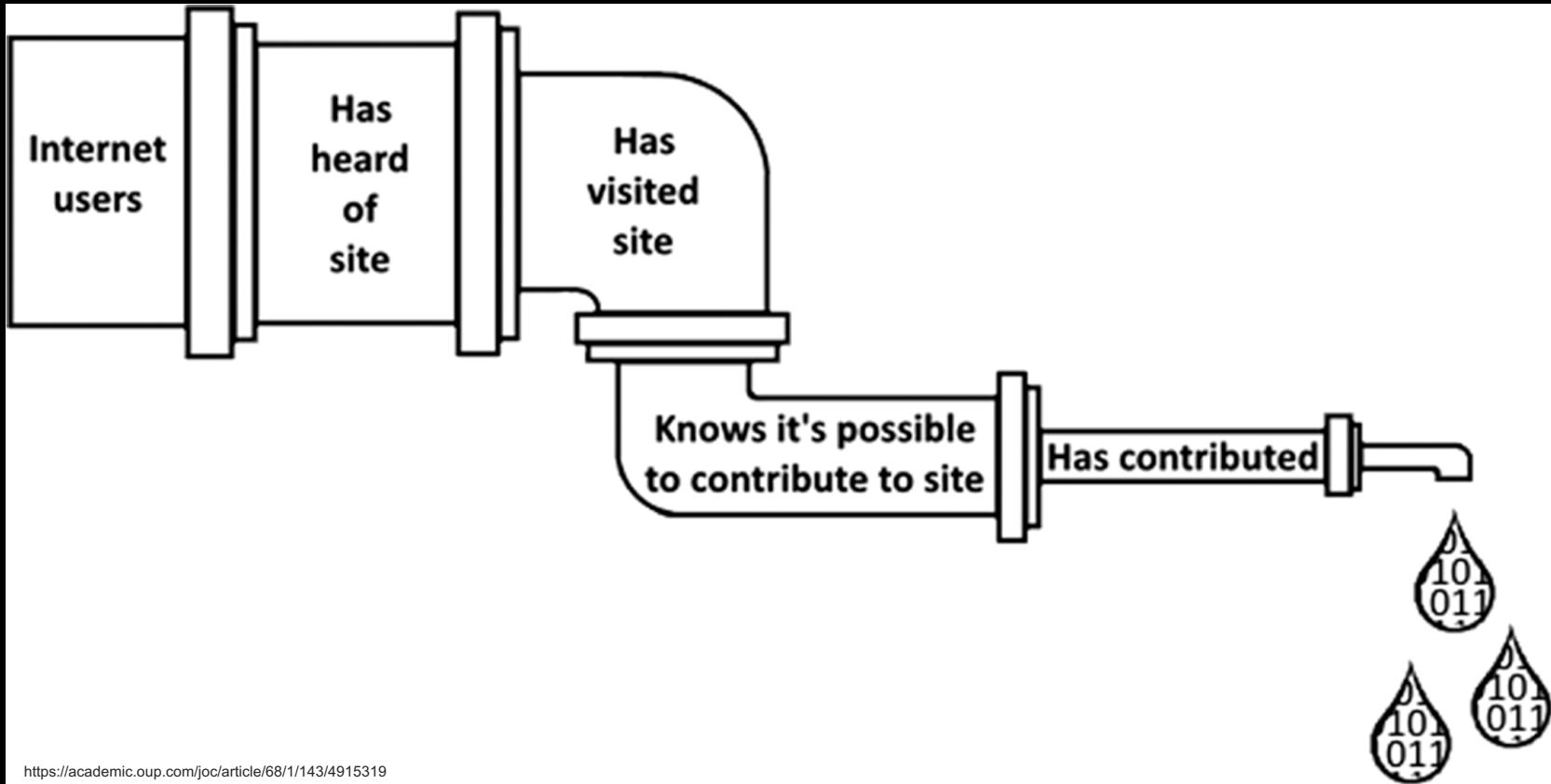
- An online ‘encyclopedia’ anyone can edit - if they can learn and follow the norms of the community.
- Reliability of knowledge a topic of ongoing interest
- Page views indicate public interest in movies, stock prices, animals, the flu
- Digital inequalities mean that online environments often exacerbate existing patterns of social exclusion
- A ‘context of heritage production’, a trace of human activity resulting from engagement with the past (Bonacchi et al 2018)

Permission level	Wikipedia users
Most permissions	Developer/System administrator
	Steward
	Check user
	Oversight
	Bureaucrat
	Administrator/Sysop
	Bot
	Registered user
	Newly registered user
	Anonymous user
No permissions	Blocked user

<https://journals.sagepub.com/doi/pdf/10.1177/1461444810365297>

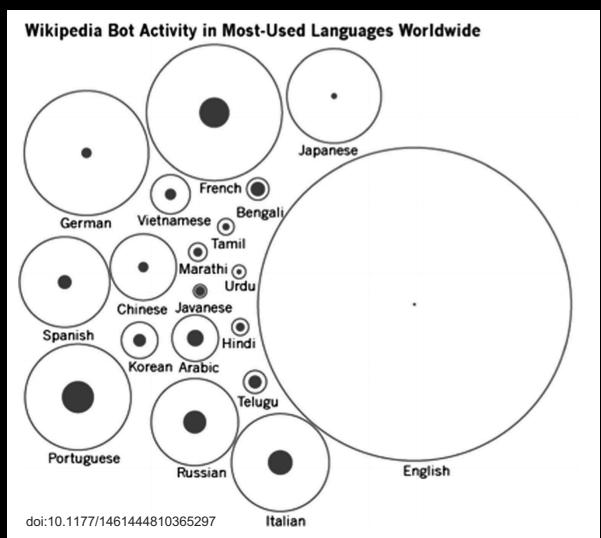
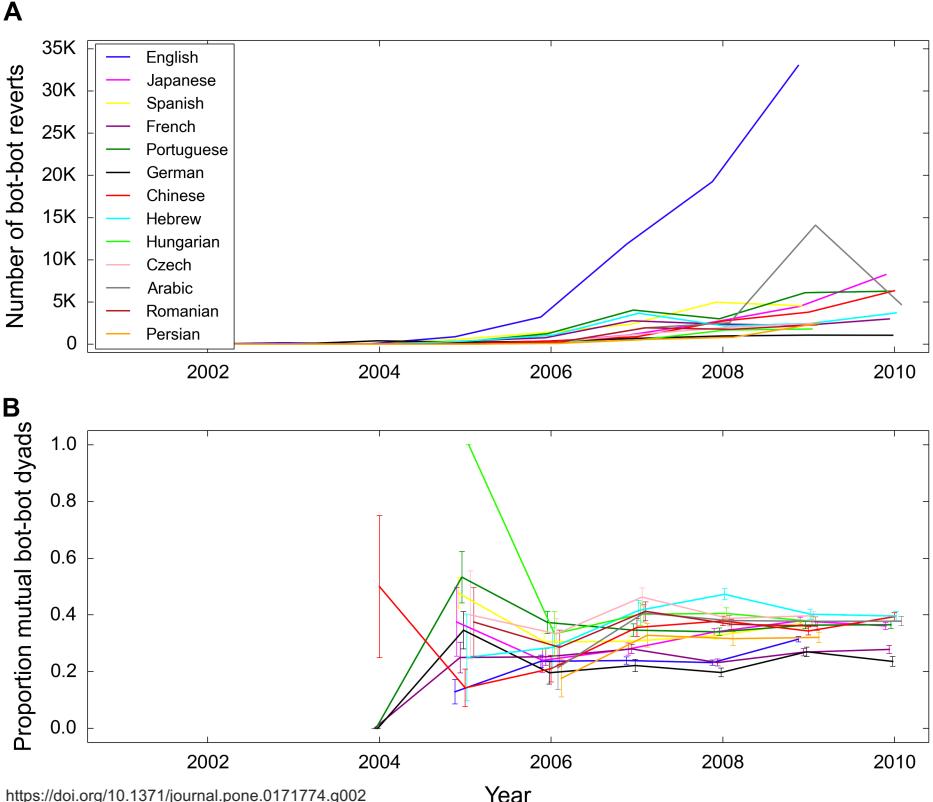


[https://en.wikipedia.org/wiki/Reliability\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Reliability_of_Wikipedia)



<https://academic.oup.com/joc/article/68/1/143/4915319>

Individuals do not automatically engage in content production on even the most participatory parts of Wikipedia, but must have the necessary knowledge to do so



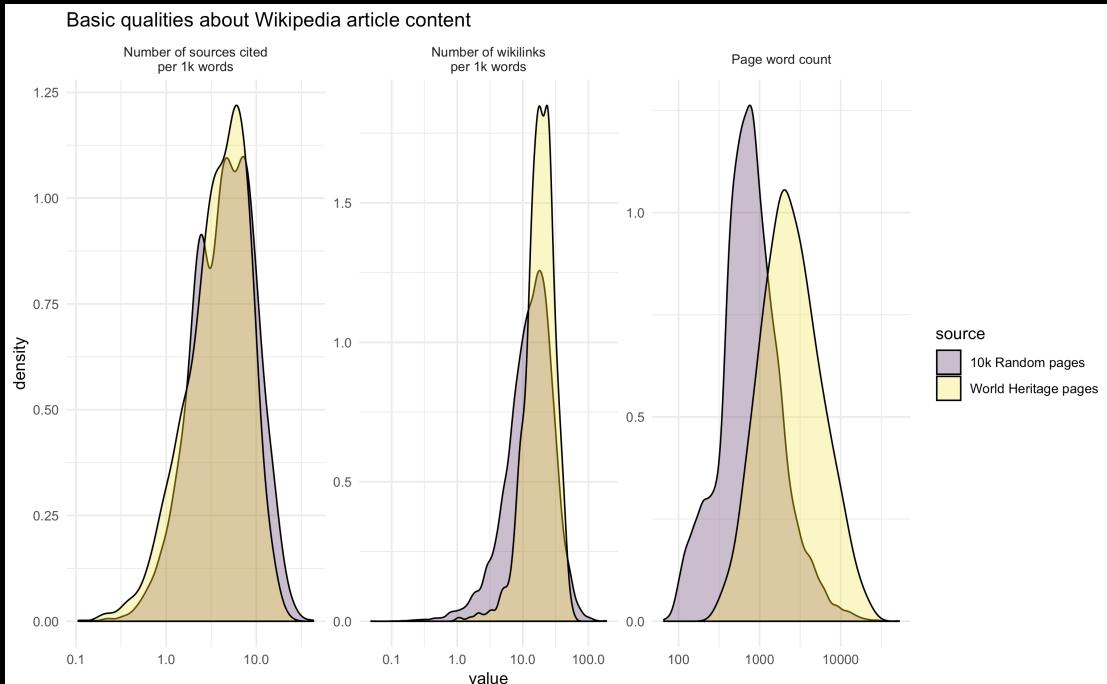
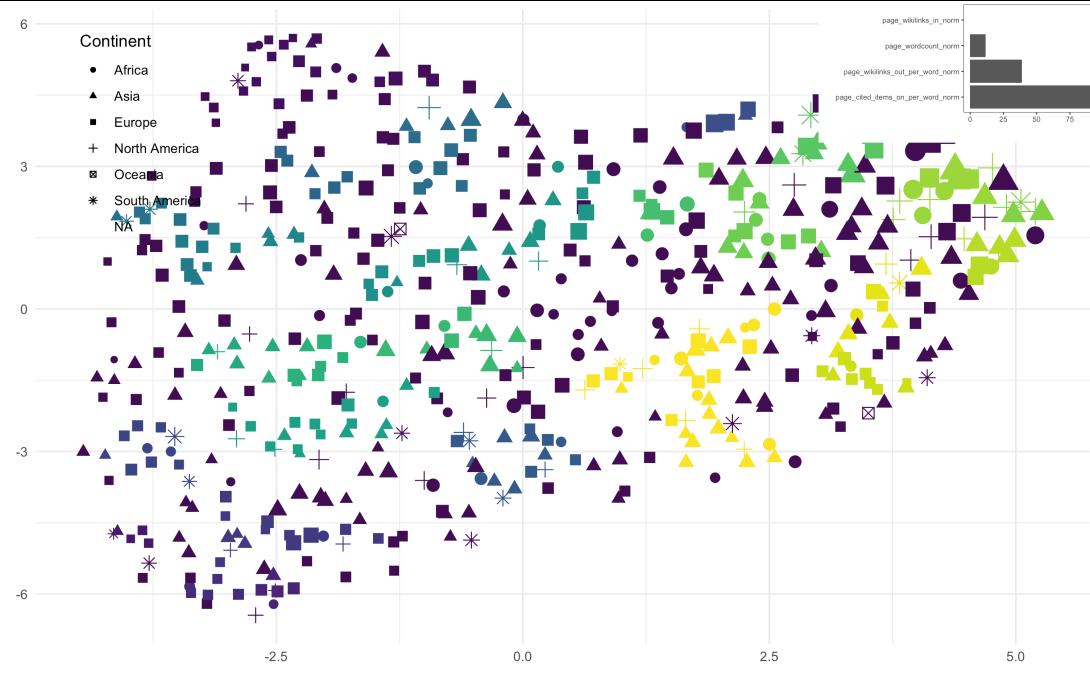
- A sociotechnical system of content production and maintenance with human and non-human actors: people, bots, and “software-assisted” human editors
- Bots! Software that automatically modify articles, perhaps battling each other?
- Editing mechanics allow for fine-grained revision control, identity management, freezing and reverting
- Talk pages document the process of content generation
- Conflict detection among Wikipedia editors is active field of study

# **UNESCO World Heritage Sites on Wikipedia**

A preliminary sketch of:

- Content: size, links-out, sources cited
- Consumption: page views, links-in
- Production: edits, editors, vandalism, reverts, bots, talk pages
- Spatial patterns: places represented, origin of edits
- Temporal patterns: age of pages, spikes in editing

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction, with Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to identify clusters, Random Forests to identify variable loadings

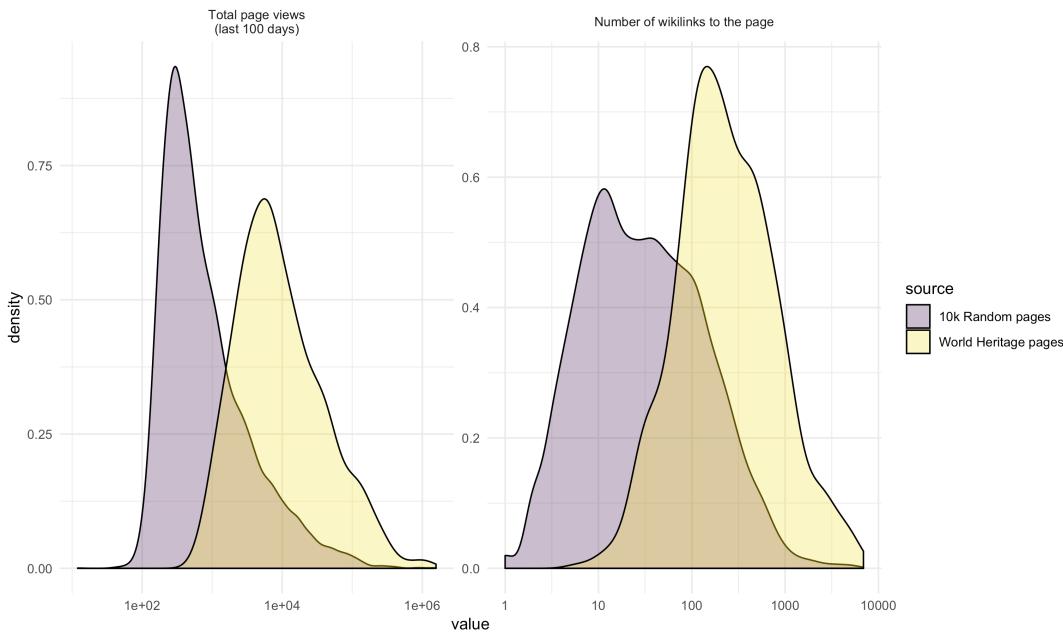


Overall, contents of WHS pages are similar to typical WP pages with

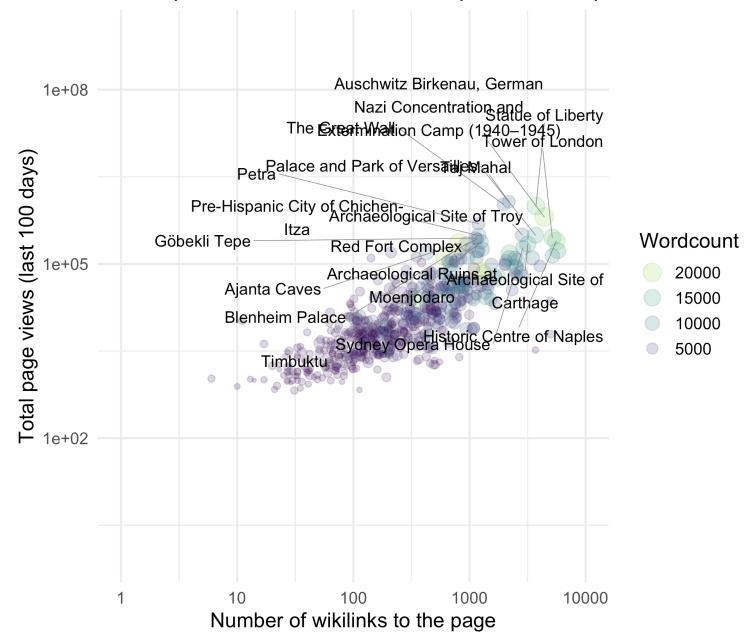
- Number of sources cited on the page
- Number of links to other WP pages

But tend to be longer than typical pages

### Basic qualities about consumption of Wikipedia articles



### Basic qualities about the consumption of Wikipedia articles



WHS pages get far more page views and links from other WP pages: they are central to WP

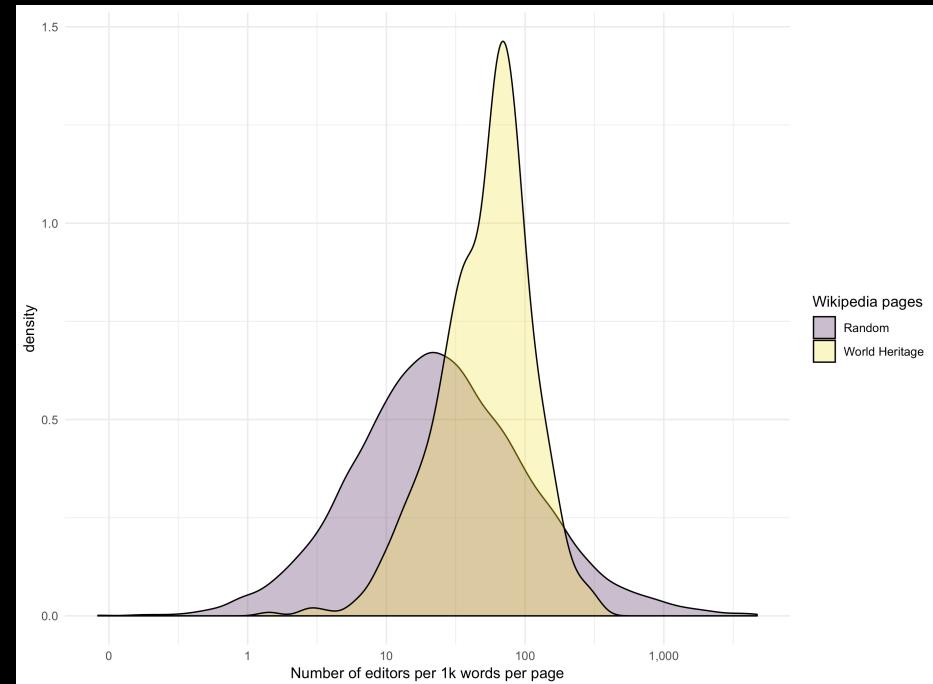
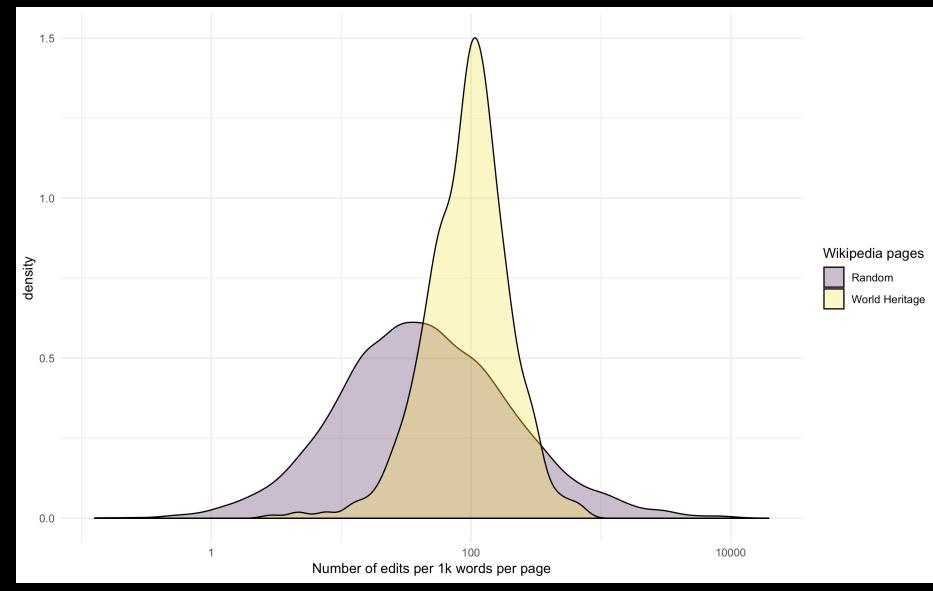
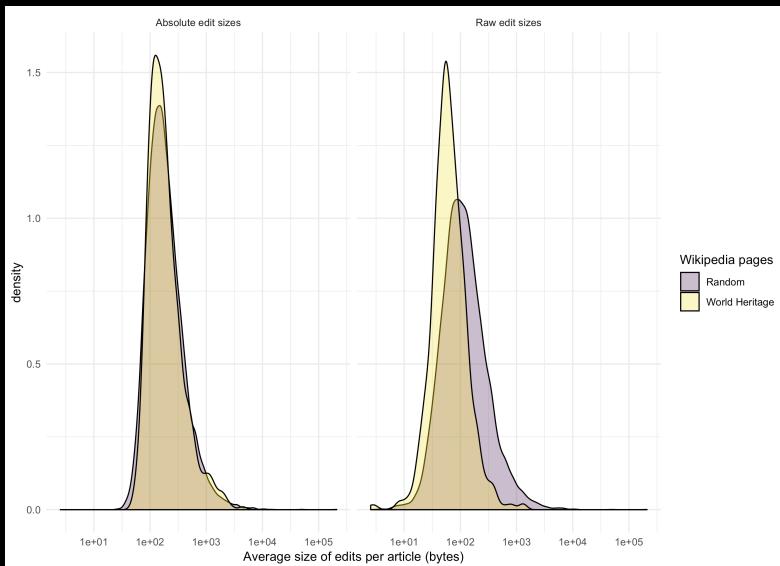
The most intensively consumed WHS pages are largely absent from the scholarly discussion of UNESCO World Heritage, with the exception of Timbuktu

The production of WHS pages differs from typical WP pages with substantially more

- Number of edits per 1k words per page
- Number of editors per 1k words per page

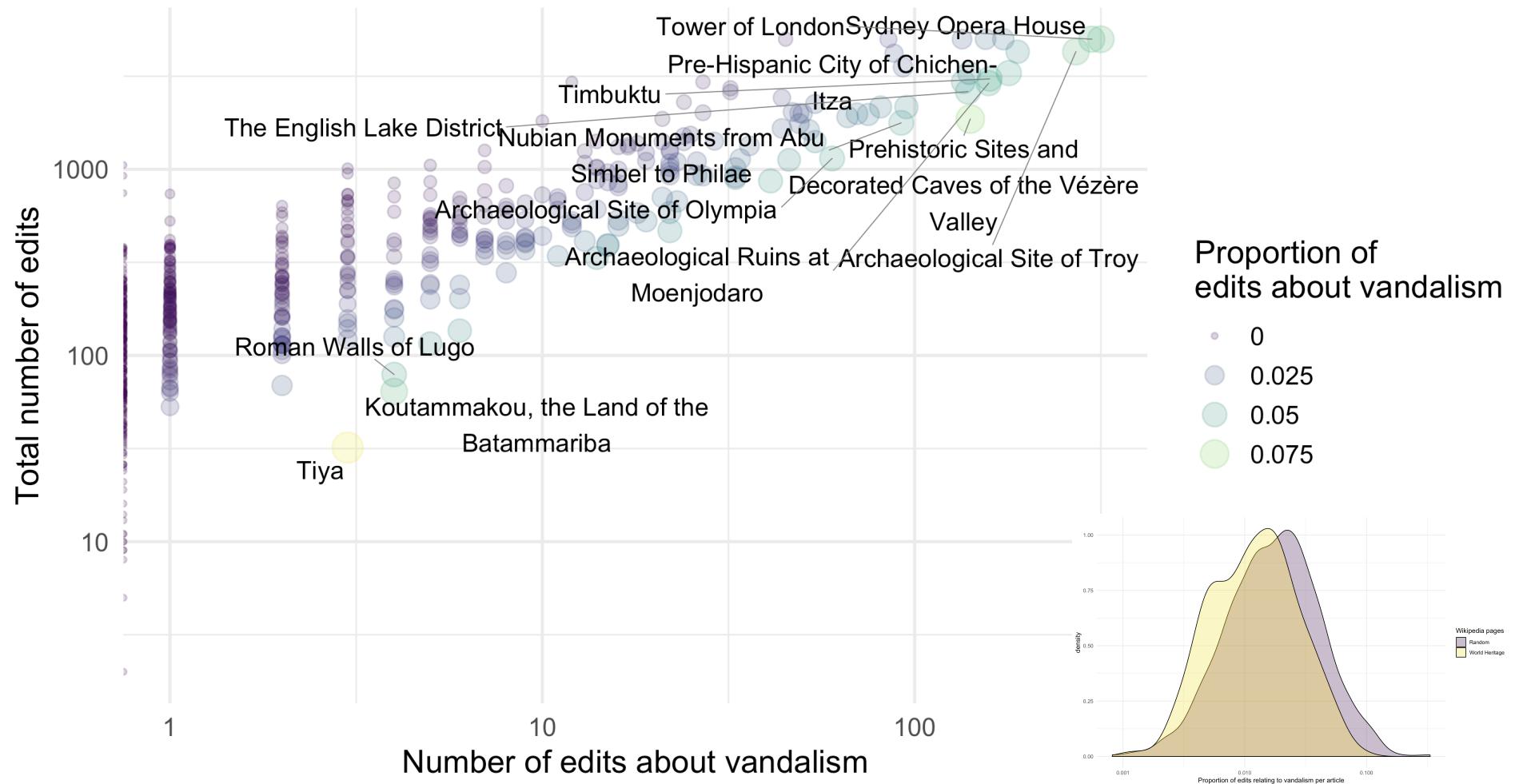
They are intensively word-smithed by a diverse community

The size of edits are about the same

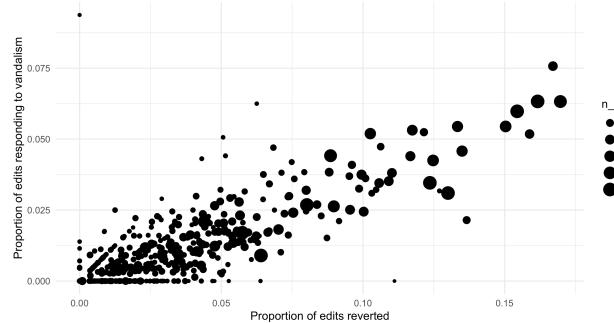
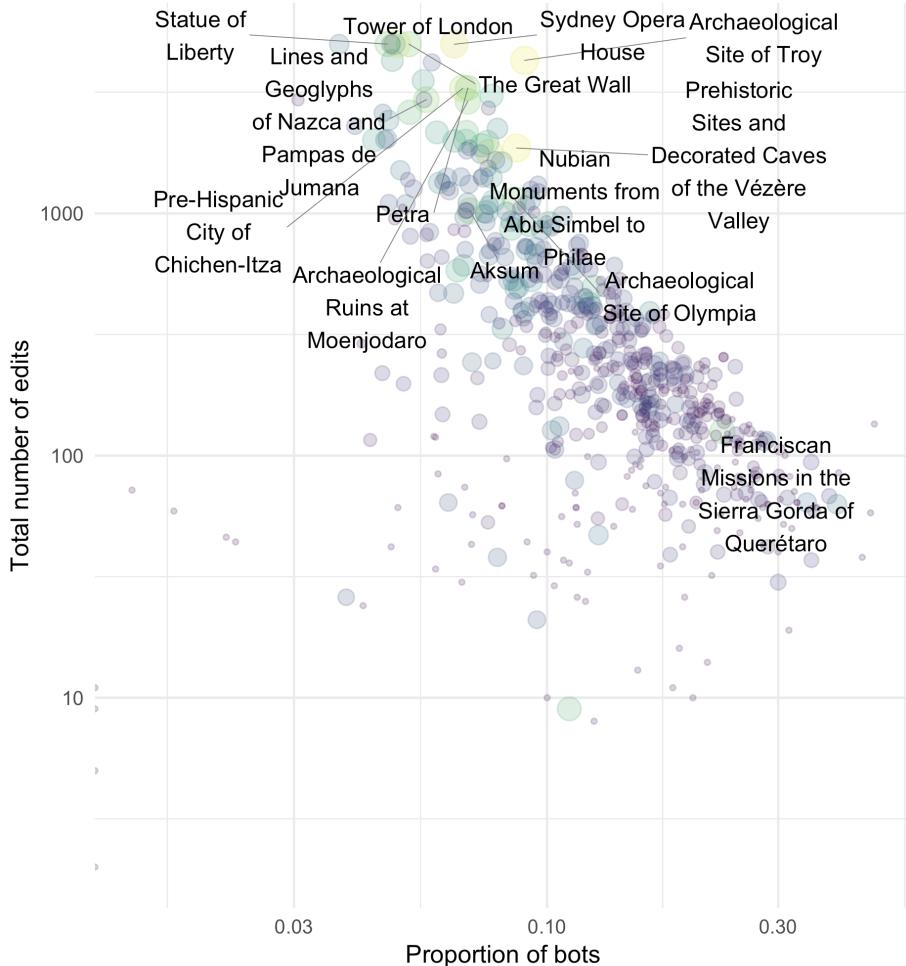


WHS pages receive less vandalism than other WP pages. The most vandalized pages are the high profile WH sites.

### Editing about vandalism on Wikipedia articles

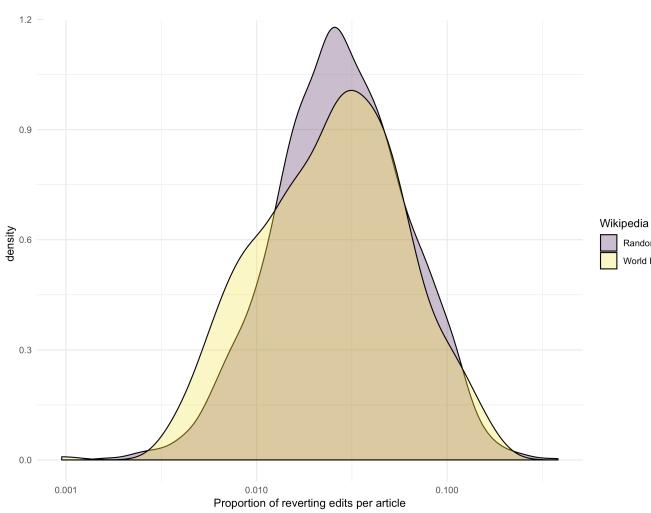


## Reverted edits on Wikipedia articles



Proportion of edits reverted

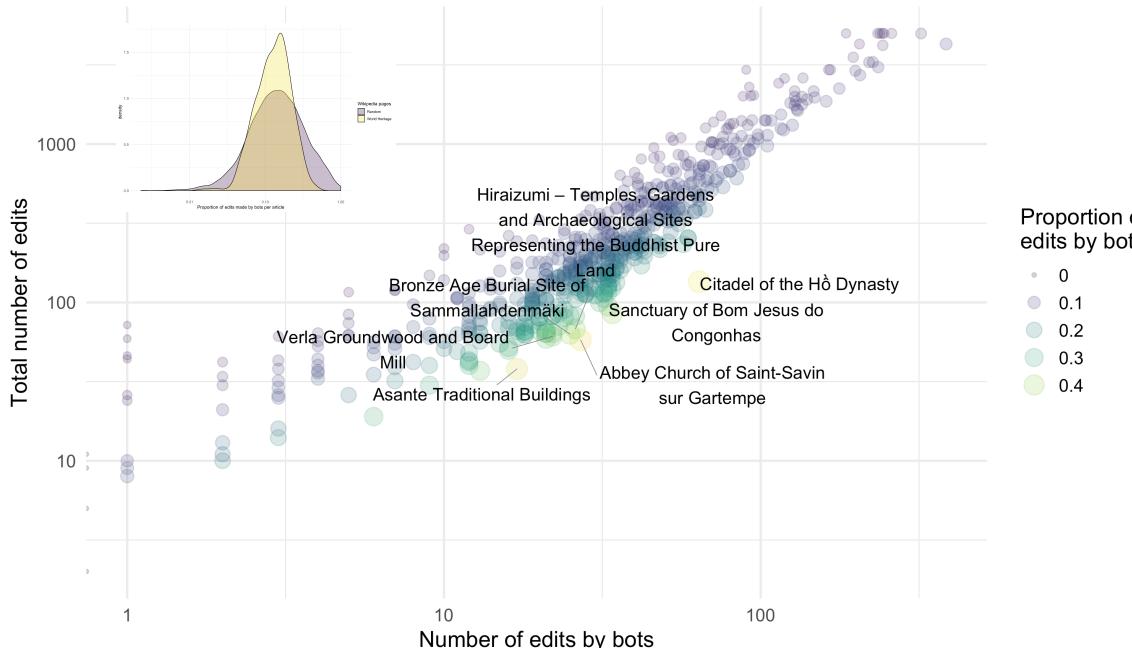
- 0
- 0.04
- 0.08
- 0.12
- 0.16



Wikipedia pages

WHS pages have about the same number of edits reverted compared to other WP pages. The most reverted pages are the high profile WH sites. Reverts correlate with vandalism.

## Bot activity on Wikipedia articles



Frequency of bot activity is no different from typical WP pages

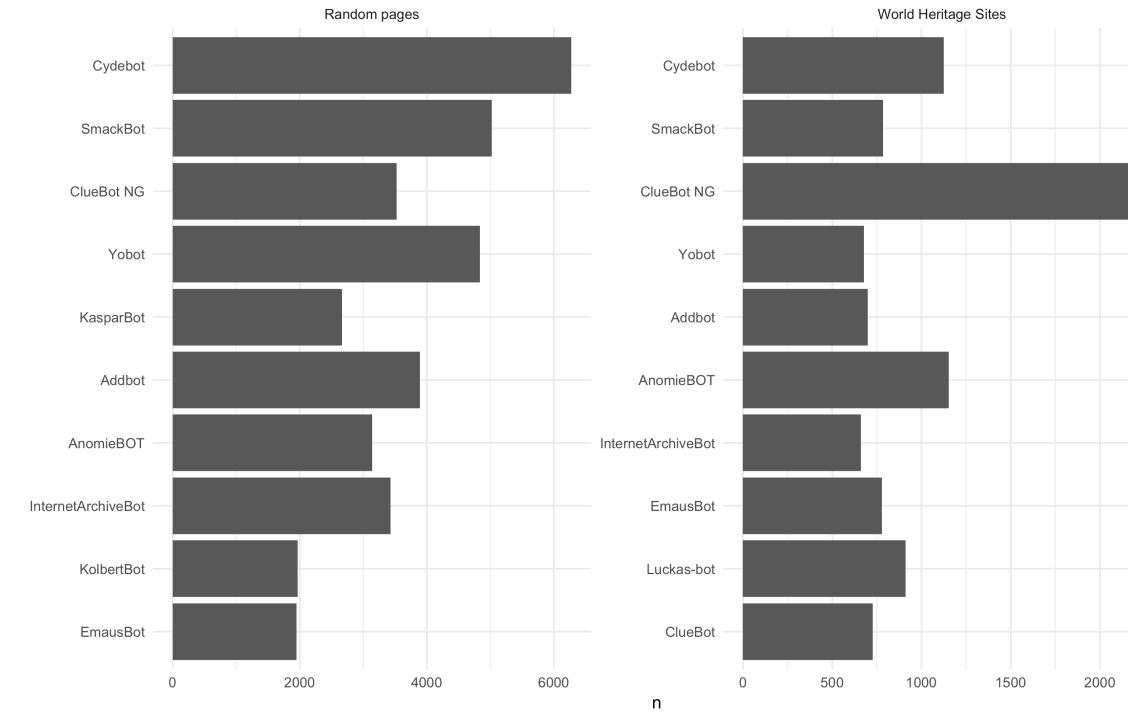
Bots are most active on smaller, low profile WHS pages

ClueBot NG: Vandalism detection and reverting

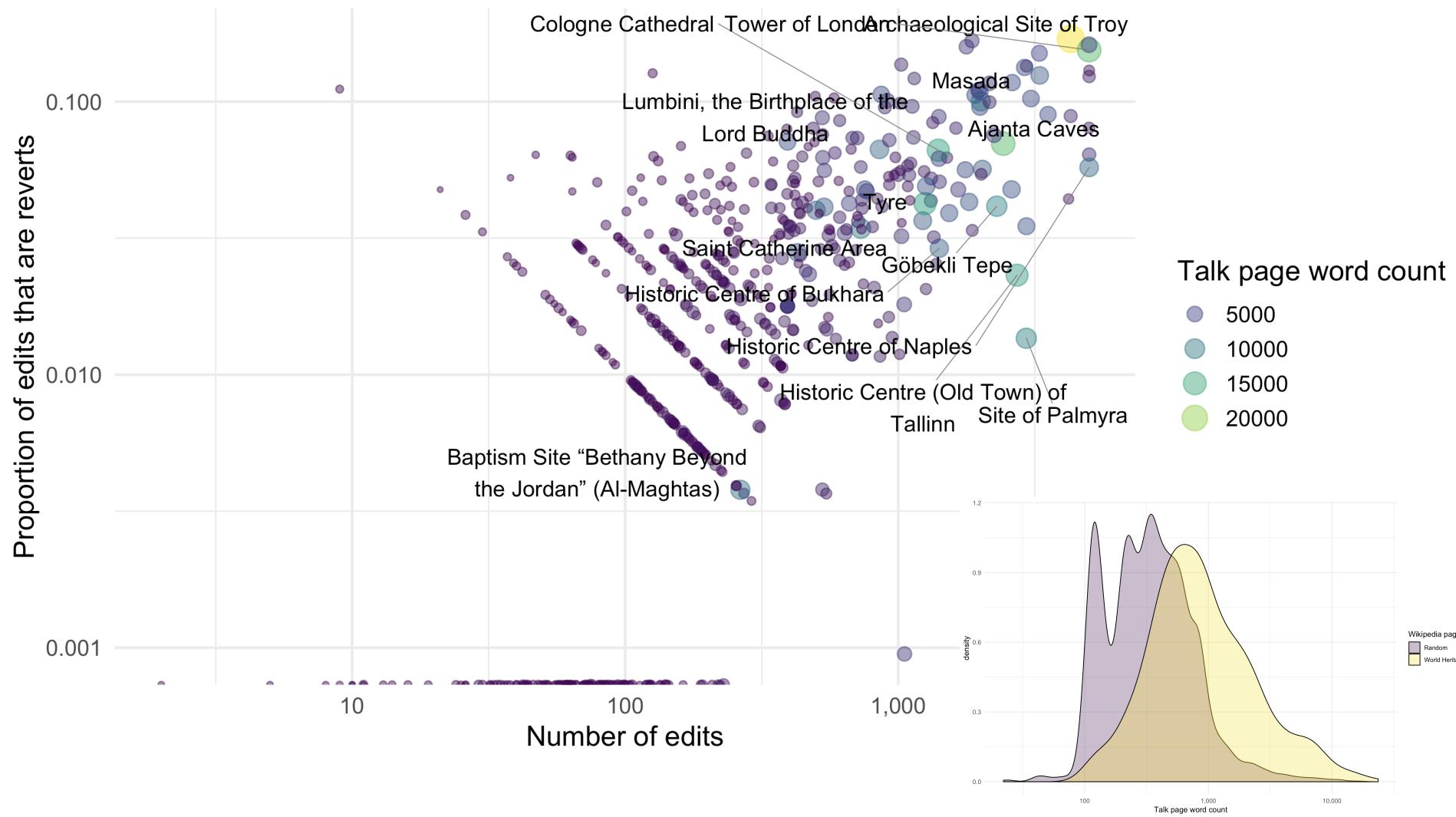
Cydebot: category edits

Smackbot: dates in templates

AnomieBOT: reference syntax



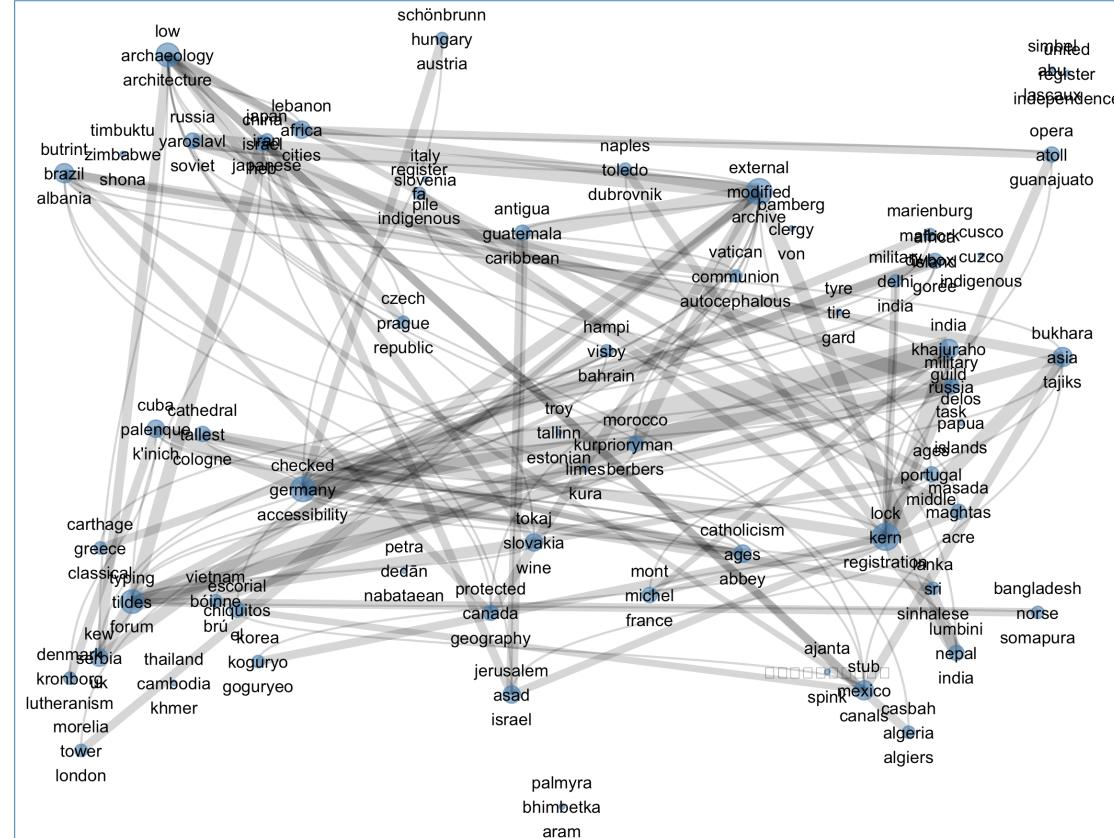
Talk page length is correlated with controversy. WHS talk pages are longer than typical WP pages. The longest talk pages are high profile sites.



## Top Topics



Expected Topic Proportions



Topics on talk pages are highly local. Prominent topics on many pages are about article administration, editing processes, and quality control

## Bias [\[edit\]](#) [\[hide\]](#)

---

- Just across from [Thailand](#) - This is **not** across but Thailand and Khmher both agree to separate the building into 2 sides. So, the people from Thailand can enter without giving any passport from one side. Not that it is comepletely belong to Khmer

And there are more places that is not neutral. Please change — 125.24.54.76 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 04:28, 2 March 2008 (UTC)

- Dear 125.24.54.76. Thank you very much for your comments. Preah Vihear has been a controversial place for such a long time! But I must respectfully disagree with your analysis. The World Court ruled in 1962 that Preah Vihear is in Cambodia. Thailand accepted that decision. Today the temple is still in Cambodia (the border line passes only a few meters from the base of the temple's lower entrance steps). As you point out, Thailand and Cambodia cooperate today to allow people to enter the temple from Thailand without visas. But you are not right in suggesting that Cambodia has surrendered sovereignty over the temple. Cambodia is like a person who owns a forest but allows neighbors to come into the forest whenever they want to enjoy the cool air. But the forest still belongs to that person! So please check your information. I invite you to contact the Thai Ministry of Foreign Affairs. People there will confirm that under international law, the temple is the possession of Cambodia, not Thailand. May I ask, were you the person who changed the article to say that Preah Vihear is in Thailand? If so, I hope that you will restore the correct wording. Thank you very much for your interest. —

Preceding [unsigned](#) comment added by — 151.200.254.107 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 02:49, 4 March 2008 (UTC)

- That is so wrong. Thailand never accepted the decision by the World Court, but Thailand could not do much even though it disagreed. Furthermore, the staircase and a few more meters actually belong to Thailand even by the World Court judgement. Also, the world heritage proposal by the Cambodian government include 4.6 sq km territory that do belong to Thailand. This area isn't part of the temple itself, but around it. Even if the whole temple may belong to Cambodia, the proposal which UNESCO has accepted includes Thai territories. —Preceding [unsigned](#) comment added by — 125.25.31.24 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 11:00, 8 July 2008 (UTC)
- Hello 125.25.31.24. If you mean that Thailand never emotionally accepted the World Court's decision, you're right--many Thais continue to feel that the temple was stolen from them. But if you mean that the Thai government never legally accepted the decision, you're wrong. Check the statements of Foreign Minister Thanat Khoman in 1962. You will see that Thanat, speaking for Thailand, accepted the court's decision, however reluctantly. Thai troops withdrew and the site was turned over to the Cambodians. In the last 46 years, no Thai government has made any serious attempt to assert sovereignty over the temple. In the recent dispute over World Heritage proposal, Thailand's claim is not that it owns the temple but that it owns adjacent land that Cambodia was including in its application. — 151.200.17.117 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 03:57, 9 July 2008 (UTC)
- Dear 125.25.31.24. The Cambodian government didn't include the 4.6 sq km territory in the application because they wanted it to be approved quickly, as they didn't want it to drag on any further. The Cambodian government never ever ceded the disputed territory to Thailand, so get your facts straight. —Preceding [unsigned](#) comment added by — 125.238.40.173 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 10:49, 17 July 2008 (UTC)
- Simply, Thailand has accepted that 'the temple' is in Cambodian territory according to the court. But the present day dispute is about the area down to the hill and around the hill. Right, Cambodia has claimed its sovereignty over it, same as Thailand, she has never recognized this. Nothing about this in the court decision.

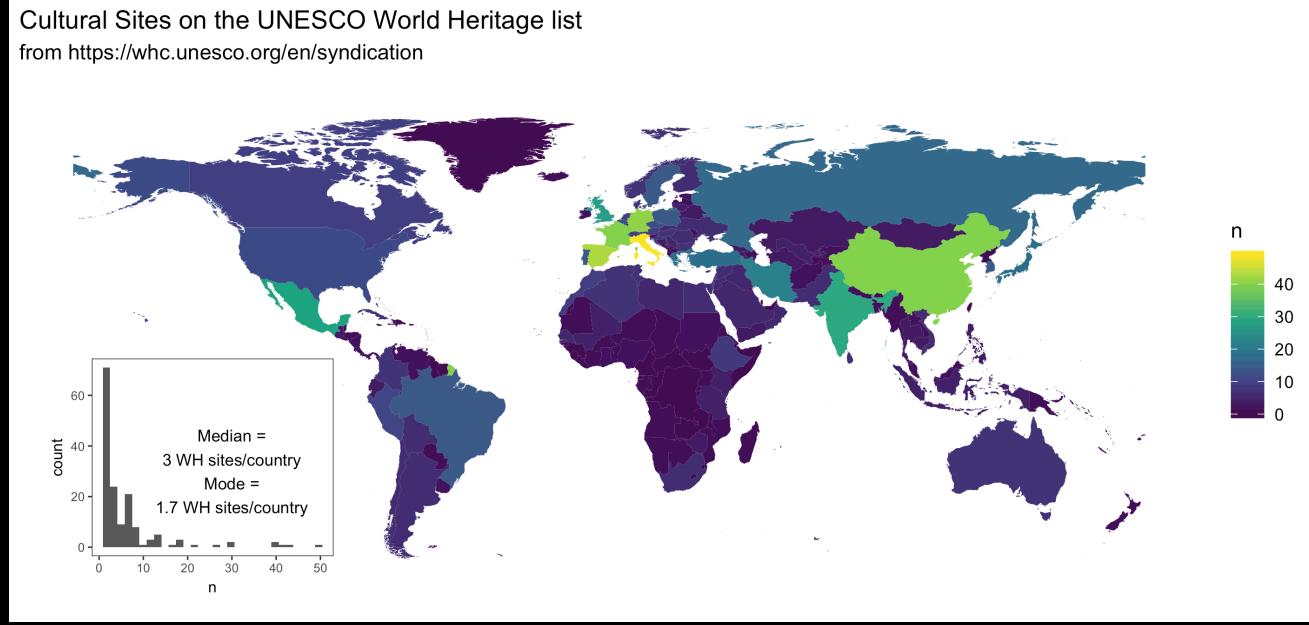
And now Thailand has step up to show her right over this area. And yes, this could be seen in several perspectives or features. —Preceding [unsigned](#) comment added by — [Xtramanx](#) ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 07:24, 19 July 2008 (UTC)

Wrong. Thailand CLAIMS the border has never been demarcated. This is, in a very limited way, correct, it is also immaterial as the border has most definitely been MAPPED. In 1907. That 101-year old map was judged by the ICJ to show where the border is. Thailand continues to occupy land the 1907 map and 1962 judgement show to be Cambodian. Thailand's deploying soldiers into this area is a clear border violation, an illegal act under international law and a demonstrable insult to Cambodia, Cambodian national soveriegnty and Cambodian dignity. —Preceding [unsigned](#) comment added by — 203.144.65.25 ([talk](#) • [contrib](#) • [actions](#) • [log](#) • [sig](#)) 06:49, 20 July 2008 (UTC)

# Spatial patterns on WP are similar to WHL

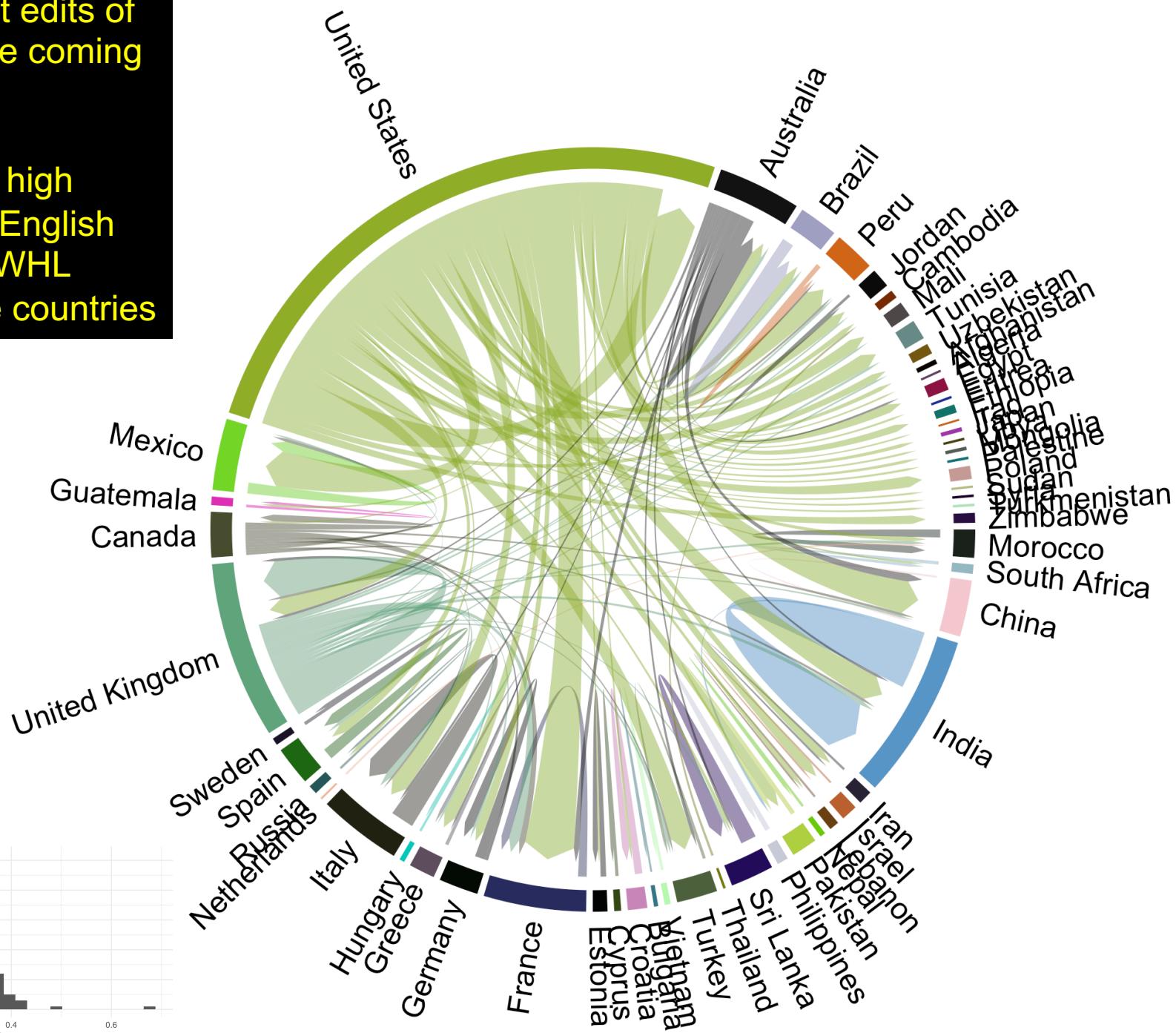
Europe and East/South Asia most are represented

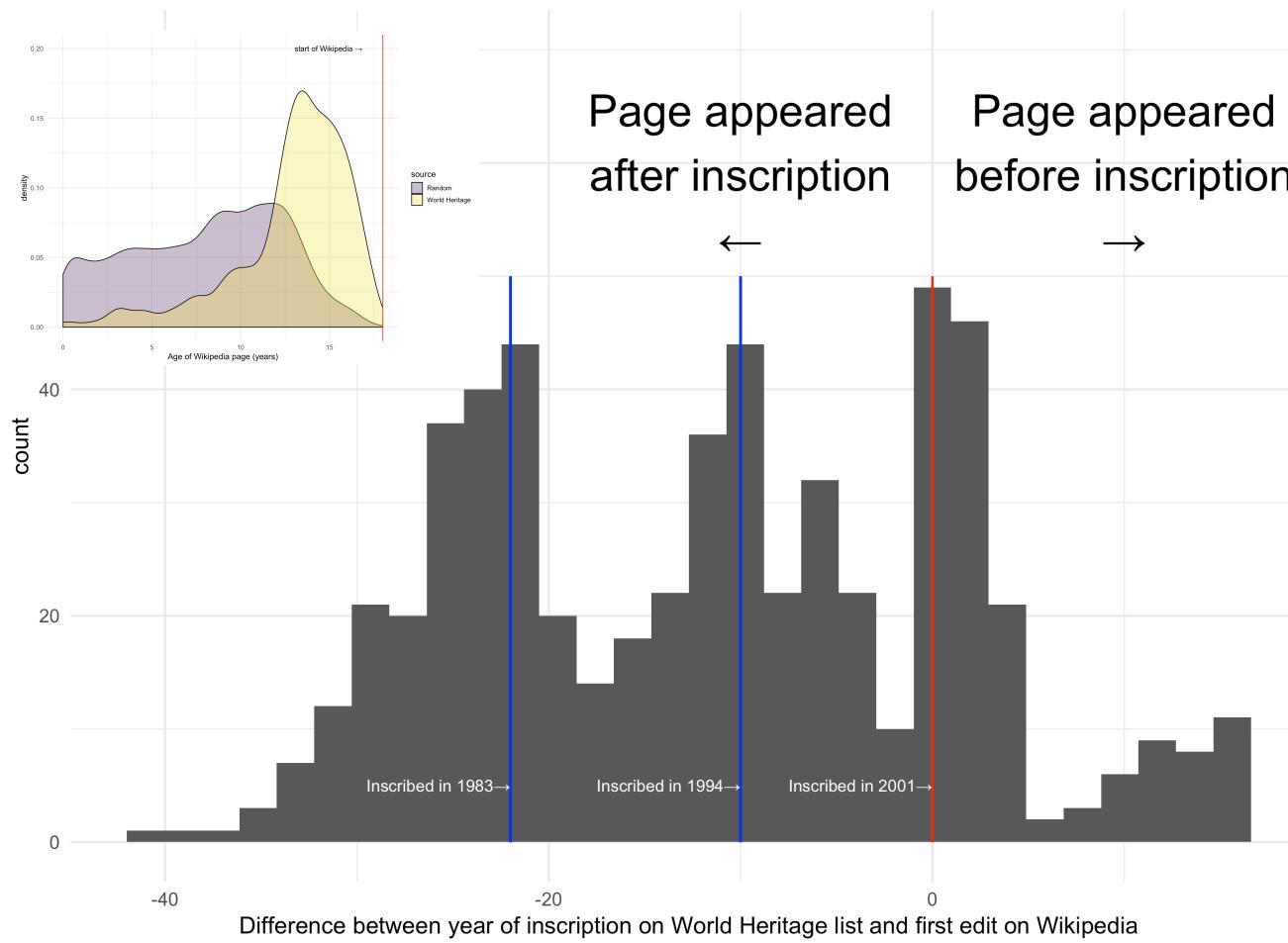
But where are the edits to these pages coming from?



By far the most edits of WHL pages are coming from the US.

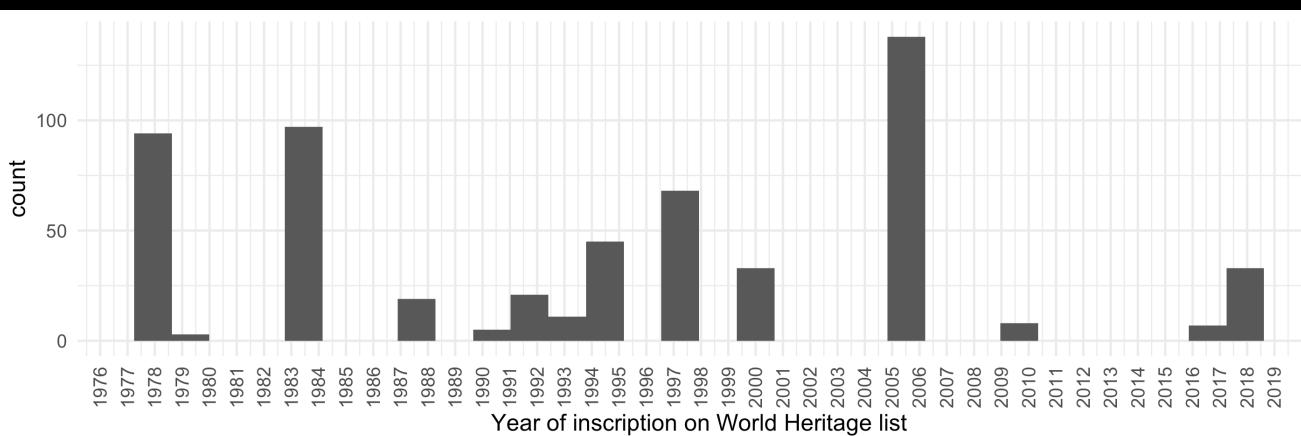
Countries with high proportions of English speakers edit WHL pages in those countries

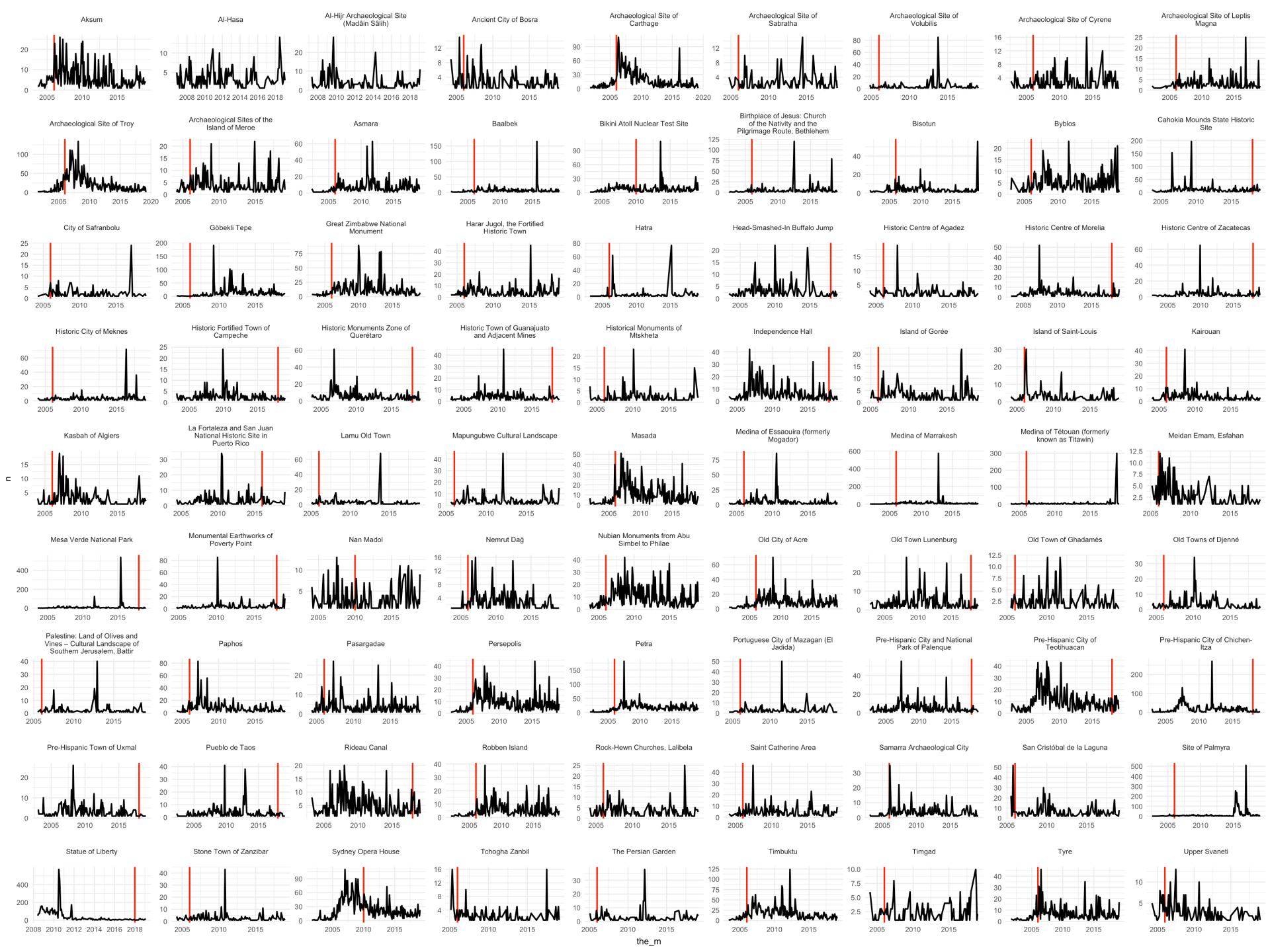




Most WHL pages appeared early in the history of WP

Most WHL pages appeared well after the site's inscription





- Content: WHL pages longer than typical
- Consumption: heavily viewed & linked to
- Production: intensively word-smithed, less vandalised & more controversial than typical, but most disputes are highly local or about WP technicalities. Scholarly concerns are not well-represented
- Spatial patterns: locations are representative, but most editors are based in the US
- Temporal patterns: pages are generally older than typical, editorial activity is very spiky, but not correlated with inscription events.

Broader implications:

Agency of participants is highly constrained by the technical characteristics of editing, bots, and consensus-valuing culture

WP offers few opportunities for heritage activism occurs, but we see behind-the-scenes activity on the talk pages.

Talk pages have great potential for close reading of heritage activism and disputes.

Scholarly concerns about state-like actors, violence & destruction, deal-making, etc. are poorly represented. Concerns on WP are hyper-local, process-centred.

The tension between the exquisite structure of WP and the mangle of community practice of WP makes sampling and reproducibility challenging