Dwelling on Wikipedia

Investigating time spent by global encyclopedia readers

Nate TeBlunthuis (nathante@uw.edu) [1,2,3]
Tilman Bayer [2,3]
Olga Vasileva [2]

August 23rd, 2019

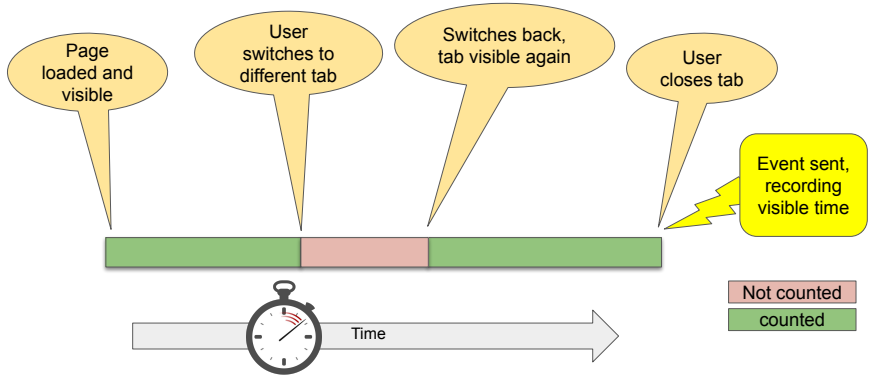
University of Washington[1]
Department of Communication
Wikimedia Foundation[2]
Community Data Science Collective[3]

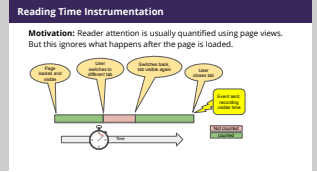
[Thanks to the NSF GRFP #2016220885, the WMF Analytics Engineering team, Zareen Farooqui, and more]

I'm Nate TeBlunthuis and I'm here with Tilman Bayer to share our work on this project to understand the Wikipedia readership through an analysis of a novel metric. I'm a PhD student at the University of Washington and did the central part of this analysis as a contractor for Wikimedia where I worked with Tilman and also with Olga Vasileva.

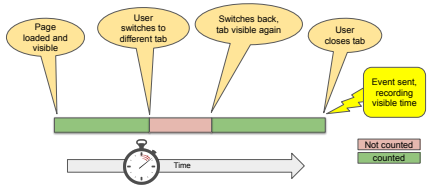
Reading Time Instrumentation

Motivation: Reader attention is usually quantified using page views. But this ignores what happens after the page is loaded.





Reading Time Instrumentation



Dwell time yields a new metric, capturing reader engagement.

Useful e.g. in development of new software features (first example: A/B test of design change for page issues templates on mobile site).

Reading Time Instrumentation
Dwell time yields a new metric, capturing reader engagement.
Useful e.g. in development of new software features (first example: A/B test of design change for page issues templates on mobile site).

Page views may often be the wrong success metric. For page issues change (making warnings about e.g. NPOV problems more prominent, we found a small increase in dwell time, corresponding to increased attention for those warnings.

In this talk we'll focus on a different result about reader behavior in general.

Data for Understanding Wikipedia Readership

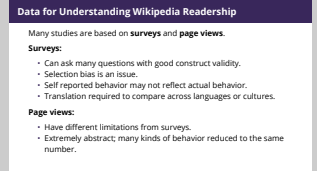
Many studies are based on **surveys** and **page views**.

Surveys:

- Can ask many questions with good construct validity.
- Selection bias is an issue.
- Self reported behavior may not reflect actual behavior.
- Translation required to compare across languages or cultures.

Page views:

- Have different limitations from surveys.
- Extremely abstract; many kinds of behavior reduced to the same number.



Surveys are a great tool for doing social science, but like all methods, they have limitations.

It's well-known that people don't always reliably report their own behavior. Especially when it comes to behaviors that are socially desirable.

Also, you have to worry about selection bias, especially when you invite a large number of people to take the survey compared to the number of people who actually take the survey.

Complementary methods

Surveys and behavioral data can collectively increase our confidence in research findings.

Lemmerich et al. (2019) conducted an international survey of Wikipedia readers: ”Why the world reads Wikipedia.” One standout finding was that readers in the “Global South” said they are more likely to engage in in-depth reading.

Can we observe behavioral evidence of this?

Complementary methods
Surveys and behavioral data can collectively increase our confidence in research findings.
Lemmerich et al. (2019) conducted an international survey of Wikipedia readers: "Why the world reads Wikipedia." One standout finding was that readers in the "Global South" said they are more likely to engage in in-depth reading.
Can we observe behavioral evidence of this?

I’ll just point out that when it comes to doing cross-national surveys translation becomes and issue. Surveys can be sensitive to nuances of meaning in how questions are worded. So when it comes to this question in particular I think it’s useful to have some behavioral data to back it up.

We’re going to come back to this question! But first we’re going on a detour to talk about the data we’re using and some descriptive analysis.

How good is this data?

It has some limitations:

1. Missing older browsers (Android browser,chrome < 39, Safari, iOS < 11.3.
2. Respects “Do Not Track”
3. Anomalous large amount of missing data on mobile
4. Doesn’t perfectly capture “reading.” Only measures that the page is visible.

We collected sampled 0.1% of page views from 2017-11-20 through 2018-10-25.

How good is this data?
It has some limitations:
1. Missing older browsers (Android browser,chrome < 39, Safari, iOS < 11.3.
2. Respects “Do Not Track”
3. Anomalous large amount of missing data on mobile
4. Doesn’t perfectly capture “reading.” Only measures that the page is visible.
We collected sampled 0.1% of page views from 2017-11-20 through 2018-10-25.

How long do people read?

How long do people read?

Total time spent reading Wikipedia by all of humanity:

670,000 years per year

32 years during this talk

How long do people read?

Total time spent reading Wikipedia by all of humanity:
670,000 years per year
32 years during this talk

└How long do people read?

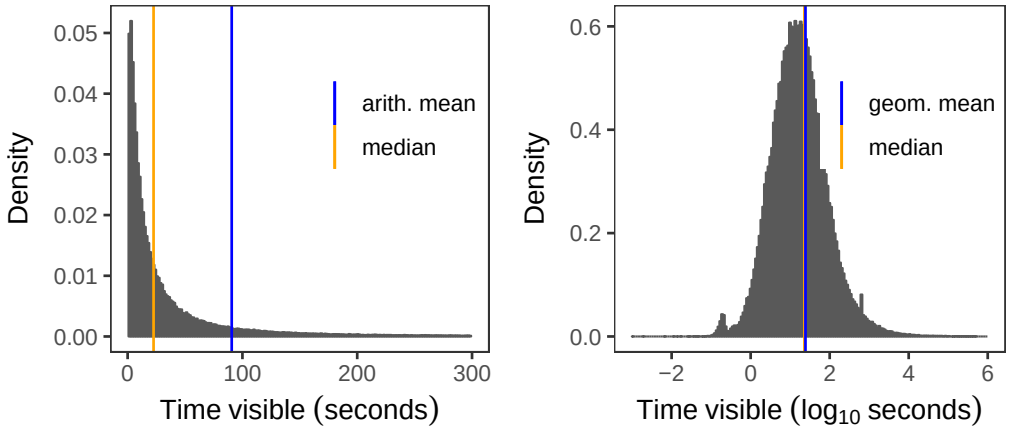
November 2017 through October 2018 (not including apps)

Based on mean time per page

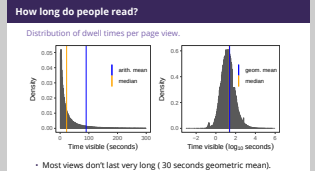
This talk: 25 minutes (not accounting for daily variations)

How long do people read?

Distribution of dwell times per page view.



- Most views don't last very long (30 seconds geometric mean).



The skewness means that the geometric mean is a better metric than the arithmetic mean (average).

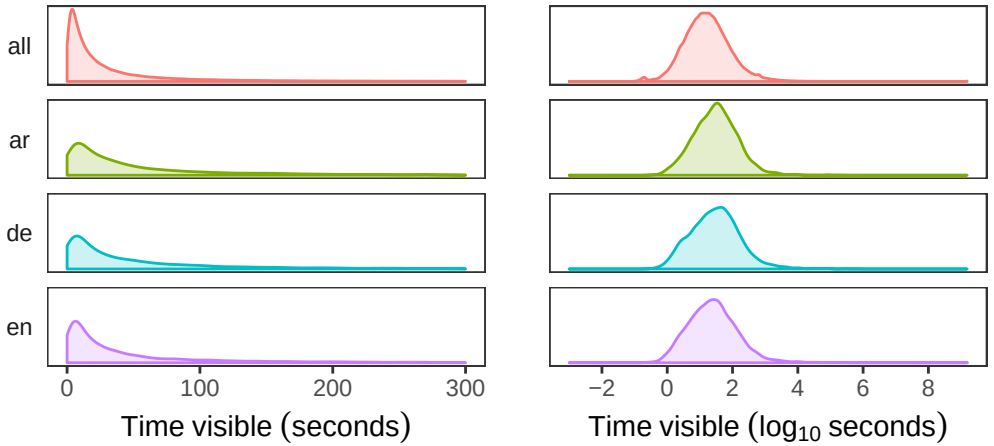
Reading time is skewed

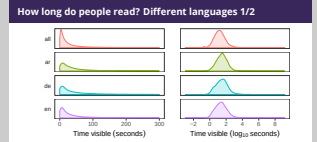
For a Wikipedia article that is 20,000 bytes long, these typical 30 seconds would only suffice to read through less than 5% of the text.

Reading time is skewed
For a Wikipedia article that is 20,000 bytes long, these typical 30 seconds would only suffice to read through less than 5% of the text.

Assuming a reading speed of around 250 words per minute and an average word length of 5 characters in English, not including spaces and punctuation.

How long do people read? Different languages 1/2

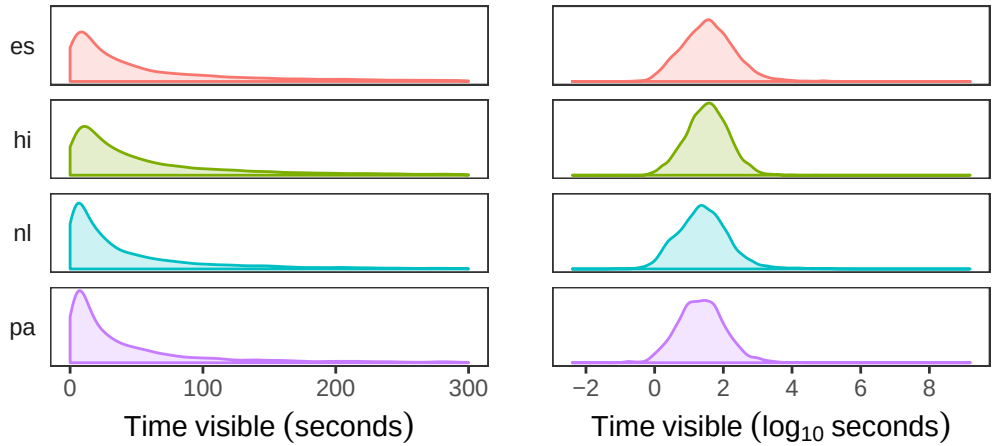


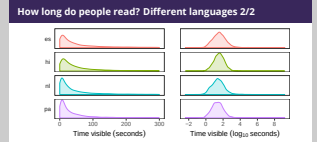


└How long do people read? Different

Kernel density plots of the distribution of dwell times on a selection of wikis. Spanish, Hindi, and Arabic appear to have longer reading times while English and Punjabi appear to have somewhat shorter reading times. In general, the distribution is very skewed, as these example wikis demonstrate.

How long do people read? Different languages 2/2





Our online supplement has more wikis.

How can reading time data help us understand global readership?

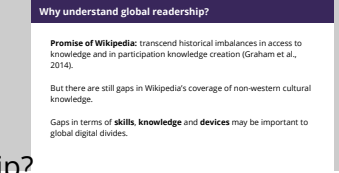
Now we're returning to the earlier question from Lemmerich et al.'s "Why the world reads Wikipedia". Readers from the Global South say that they are more likely to engage in deeper information seeking tasks compared to readers from the Global North. Does this mean they are likely to read for longer?

Why understand global readership?

Promise of Wikipedia: transcend historical imbalances in access to knowledge and in participation knowledge creation (Graham et al., 2014).

But there are still gaps in Wikipedia's coverage of non-western cultural knowledge.

Gaps in terms of **skills**, **knowledge** and **devices** may be important to global digital divides.



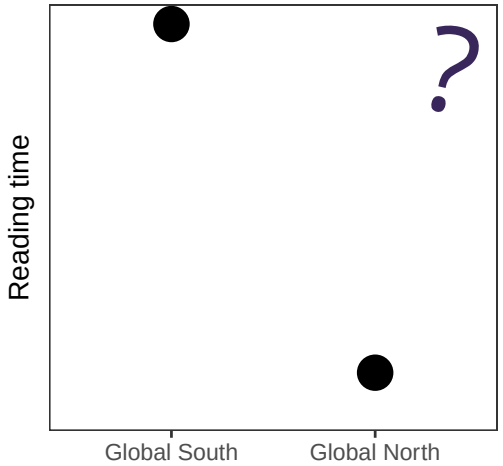
└Why understand global readership?

Wikipedia promises to advance over traditional modes of knowledge production in which dominant western attitudes shape what people and places will be included and how they will be represented in authoritative sources like encyclopedias Graham et al., 2014. In theory, peer production can empower people around to the world to add their local knowledge of their places to Wikipedia. Yet even as global access to Wikipedia grows, it is slow to fulfill these promises. Gaps in coverage of cultural knowledge reflect and reinforce structural digital divides at many levels that "disadvantage many of the world's informational

Hypotheses

Development and Reading Time

H1: Readers in the Global South are more likely to spend more time reading each page they visit compared to readers in the Global North.



Development and Reading Time
H1: Readers in the Global South are more likely to spend more time reading each page they visit compared to readers in the Global North.
?
Reading time
Global South
Global North

This is based on the survey finding, but also supports intuitions that knowledge gaps between Global South and Global North information contexts exist that can be filled by Wikipedia.

Mobile vs desktop devices

Desktop devices have advantages for in-depth understanding. GS

readers may experience relatively limited access to desktop devices. When they do have access, will they read for longer?

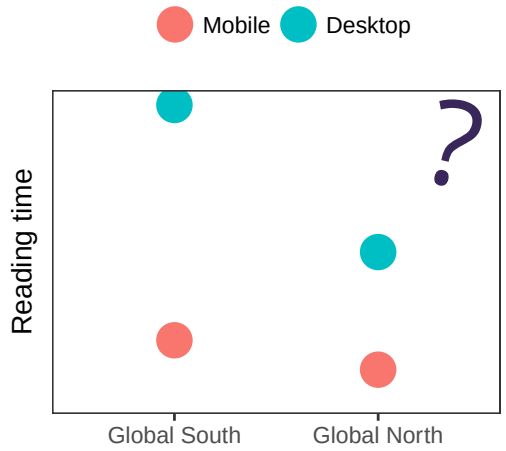
Mobile vs desktop devices

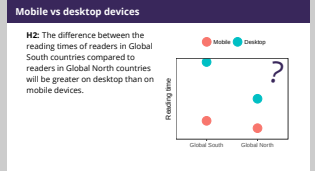
Desktop devices have advantages for in-depth understanding. GS

readers may experience relatively limited access to desktop devices. When they do have access, will they read for longer?

Mobile vs desktop devices

H2: The difference between the reading times of readers in Global South countries compared to readers in Global North countries will be greater on desktop than on mobile devices.





If desktop devices have advantages for reading to gain in-depth understanding then users may be more likely to choose these devices for such tasks (when they have the choice).

Global South readers may also experience gaps limiting their access to desktop devices, and when they do have access may be likely to take advantage of such opportunities by reading longer.

Therefore, we expect users in countries within the Global South designation (or with lower HDIs) to read even longer

Last-in-session page views

Reading times at the end of a session are longer. Is that because of

“screen-and-glean” behavior? If so and if Global South readers do more

in-depth reading, then will we find longer reading times in the last-in-session page view?

Last-in-session page views

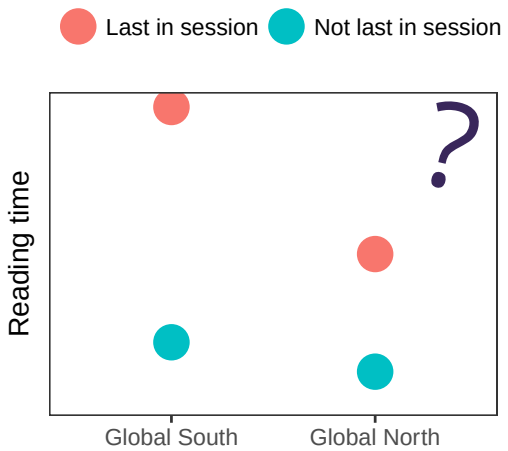
Reading times at the end of a session are longer. Is that because of

“screen-and-glean” behavior? If so and if Global South readers do more

in-depth reading, then will we find longer reading times in the last-in-session page view?

Last-in-session page views

H3: The difference between reading times in countries with lower HDI and countries with higher HDI will be greater on the last page view in a session than on other page views.



Analytic Plan

Country from MaxMind GeoIP database. **Global North / South**

defined by the international telecommunications union. **Human**

Development Index from the UN. **Page length** in wikitext bytes. **Last**

in session Is the page view the last before the browser tab closes?
Mobile vs Desktop devices: based on endpoint.

We tested the following hypotheses using log-linear regression and a simple non-parametric analysis on a stratified sample over 285 Wikimedia projects.

Analytic Plan

Country from MaxMind GeoIP database. **Global North / South** defined by the international telecommunications union. **Human Development Index** from the UN. **Page length** in wikitext bytes. **Last in session** Is the page view the last before the browser tab closes? **Mobile vs Desktop devices:** based on endpoint.

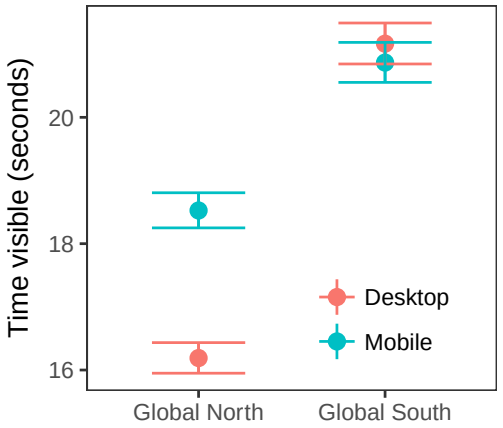
We tested the following hypotheses using log-linear regression and a simple non-parametric analysis on a stratified sample over 285 Wikimedia projects.

We use the Human Development Index (HDI) and the Global South/Global North regional classification to comparing countries separated by varying levels of development. We recognize that both are insufficient for defining economic development. These concepts and our measures of them only provide an incomplete understanding of the unique cultures within an information-seeking context.

We hope that this work provides a basis of study that may be continued with work that takes into account individual cultural context, internet accessibility, and internal inequality.

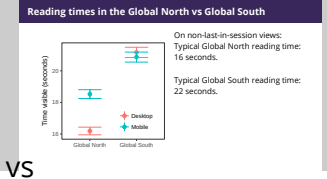
Results

Reading times in the Global North vs Global South

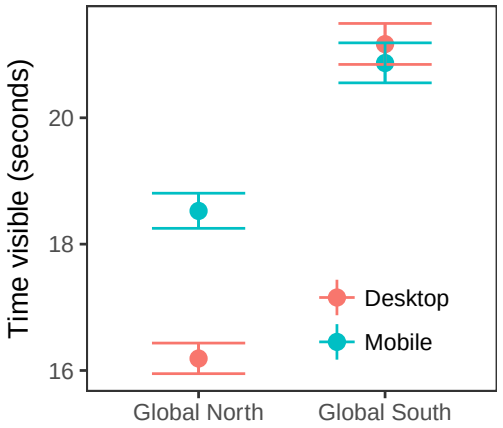


On non-last-in-session views:
Typical Global North reading time: 16 seconds.

Typical Global South reading time: 22 seconds.



For non-last-in-session page views, a prototypical reader on a desktop device in a country with an HDI one standard deviation below the mean is predicted to spend about 25 seconds on a given non-last-in-session page view compared to the predicted 18 seconds spent by an average reader in a country with an HDI one standard deviation above the mean.

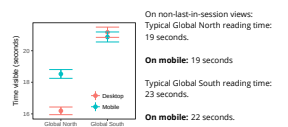


On non-last-in-session views:
Typical Global North reading time:
19 seconds.

On mobile: 19 seconds

Typical Global South reading time:
23 seconds.

On mobile: 22 seconds.



For non-last-in-session page views, a prototypical reader on a desktop device in a country with an HDI one standard deviation below the mean is predicted to spend about 25 seconds on a given non-last-in-session page view compared to the predicted 18 seconds spent by an average reader in a country with an HDI one standard deviation above the mean.

Last-in-session

People read longer in last-in-session views.

But we didn’t observe the hypothesized amplification between global-south and global-north readers.

Last-in-session

People read longer in last-in-session views.

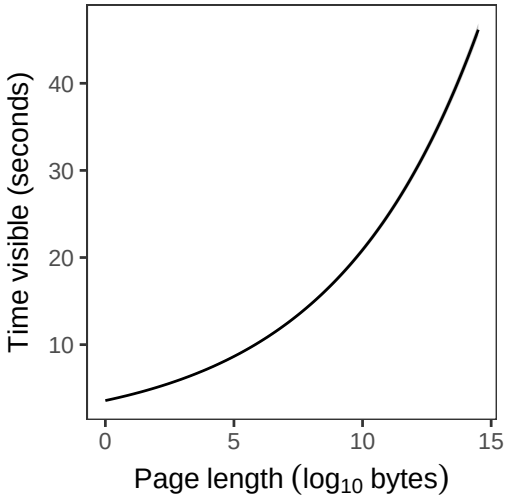
But we didn't observe the hypothesized amplification between global-south and global-north readers.

In another analysis we did in the paper, we found another piece of evidence against the “screen-and-glean” model. This helps explain why we might not observe evidence of the hypothesis.

We included page length in our models. How much longer do you think people read on longer pages?

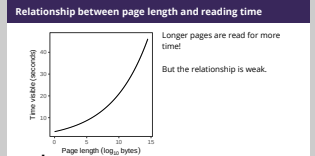
We included page length in our models. How much longer do you think people read on longer pages?

Relationship between page length and reading time



Longer pages are read for more time!

But the relationship is weak.



If a page were to double its length, our model would predict a marginal increase in reading times of a factor of 1.2. For example, a page with 10000 bytes has a predicted reading time of 25 seconds, which for a page with twice that length (20000 bytes) increases to 30 seconds.

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Discussion
Readers in the Global South dwell on pages longer than readers in the Global North.

Our findings provide evidence behavioral data to corroborate findings from Lemmerich et al's "How the world reads Wikipedia" that say that Global South readers are more likely to read for deeper information seeking tasks. Global South readers read for more time on average compared to Global North readers. And they do this on the kinds of devices that we expect will be associated with deeper information seeking tasks.

We thought that deeper information seeking would be associated with last-in-session views and so the GS/GN gap would be greater there. This was based on a

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

Discussion
Readers in the Global South dwell on pages longer than readers in the Global North.
Especially on Desktop devices

Our findings provide evidence behavioral data to corroborate findings from Lemmerich et al's "How the world reads Wikipedia" that say that Global South readers are more likely to read for deeper information seeking tasks. Global South readers read for more time on average compared to Global North readers. And they do this on the kinds of devices that we expect will be associated with deeper information seeking tasks.

We thought that deeper information seeking would be associated with last-in-session views and so the GS/GN gap would be greater there. This was based on a

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

These findings support the notion that readers in the Global South read for deeper information seeking.

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

These findings support the notion that readers in the Global South read for deeper information seeking.

Our findings provide evidence behavioral data to corroborate findings from Lemmerich et al's "How the world reads Wikipedia" that say that Global South readers are more likely to read for deeper information seeking tasks. Global South readers read for more time on average compared to Global North readers. And they do this on the kinds of devices that we expect will be associated with deeper information seeking tasks.

We thought that deeper information seeking would be associated with last-in-session views and so the GS/GN gap would be greater there. This was based on a

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

These findings support the notion that readers in the Global South read for deeper information seeking.

We didn't try to account for whether people read in a first language.

Discussion
Readers in the Global South dwell on pages longer than readers in the Global North.
Especially on Desktop devices
These findings support the notion that readers in the Global South read for deeper information seeking.
We didn't try to account for whether people read in a first language.

Our findings provide evidence behavioral data to corroborate findings from Lemmerich et al's "How the world reads Wikipedia" that say that Global South readers are more likely to read for deeper information seeking tasks. Global South readers read for more time on average compared to Global North readers. And they do this on the kinds of devices that we expect will be associated with deeper information seeking tasks.

We thought that deeper information seeking would be associated with last-in-session views and so the GS/GN gap would be greater there. This was based on a

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

These findings support the notion that readers in the Global South read for deeper information seeking.

We didn't try to account for whether people read in a first language.

Reading time, and similar measures of reader behavior can be useful. They should be used with care.

Discussion

Readers in the Global South dwell on pages longer than readers in the Global North.

Especially on Desktop devices

These findings support the notion that readers in the Global South read for deeper information seeking.

We didn't try to account for whether people read in a first language.

Reading time, and similar measures of reader behavior can be useful. They should be used with care.

Our findings provide evidence behavioral data to corroborate findings from Lemmerich et al's "How the world reads Wikipedia" that say that Global South readers are more likely to read for deeper information seeking tasks. Global South readers read for more time on average compared to Global North readers. And they do this on the kinds of devices that we expect will be associated with deeper information seeking tasks.

We thought that deeper information seeking would be associated with last-in-session views and so the GS/GN gap would be greater there. This was based on a

Thank you!

Questions?

Further information:

Paper: OpenSym 2019 proceedings

We plan to release a public dataset consisting of periodically aggregated dwell times per page - please visit [[phab:T230642]] and comment to help us do it!

Thank you!

Questions?

Further information:

Paper: OpenSym 2019 proceedings

We plan to release a public dataset consisting of periodically aggregated dwell times per page - please visit [[phab:T230642]] and comment to help us do it!

Backup Slides

Percentiles

wiki	5%	25%	50%	75%	95%
all wikis	1.8	8.0	25.0	75.1	439.1
ar	5.2	5.2	21.5	69.9	371.7
de	14.1	14.1	14.1	56.6	482.7
en	37.2	37.2	37.2	37.2	262.4
es	23.3	23.3	23.3	65.5	616.4
hi	2.5	11.4	31.4	82.6	360.5
nl	6.1	6.1	15.9	60.1	441.8
pa	2.0	7.2	19.5	55.4	303.1

Percentiles

wiki 5% 25% 50% 75% 95%
all wikis 1.8 8.0 25.0 75.1 439.1
ar 5.2 5.2 21.5 69.9 371.7
de 14.1 14.1 14.1 56.6 482.7
en 37.2 37.2 37.2 37.2 262.4
es 23.3 23.3 23.3 65.5 616.4
hi 2.5 11.4 31.4 82.6 360.5
nl 6.1 6.1 15.9 60.1 441.8
pa 2.0 7.2 19.5 55.4 303.1

Percentiles for reading times (in seconds) on selected Wikipedia editions

Non-Parametric Results

Economic-region	Desktop	Last-in-session	Time-visible
North	False	False	20.1
South	False	False	21.5
North	True	False	16.1
South	True	False	21.8
North	False	True	28.1
South	False	True	28.7
North	True	True	39.8
South	True	True	43.6

Table 1: Table of median reading times by last-in-session, economic region, and device type. Reading times in the Global South are greater than in the Global North in all categories, and are markedly greater on desktop compared

Non-Parametric Results

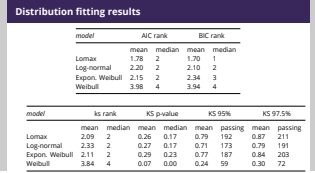
Economic-region Desktop Last-in-session Time-visible
North False False 20.1
South False False 21.5
North True False 16.1
South True False 21.8
North False True 28.1
South False True 28.7
North True True 39.8
South True True 43.6

Table 1: Table of median reading times by last-in-session, economic region, and device type. Reading times in the Global South are greater than in the Global North in all categories, and are markedly greater on desktop compared to mobile devices.

Distribution fitting results

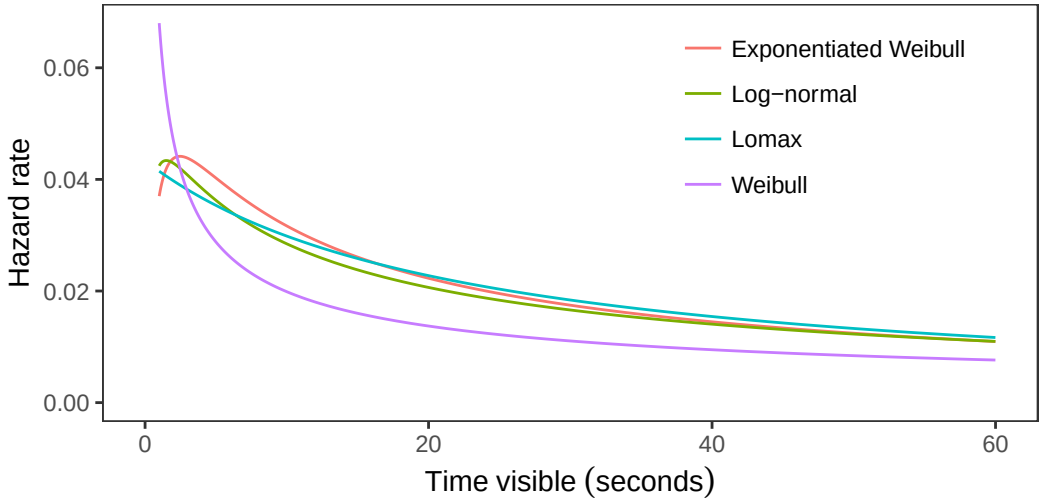
model	AIC rank		BIC rank	
	mean	median	mean	median
Lomax	1.78	2	1.70	1
Log-normal	2.20	2	2.10	2
Expon. Weibull	2.15	2	2.34	3
Weibull	3.98	4	3.94	4

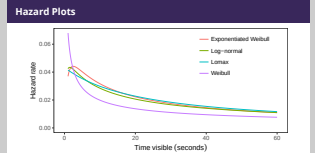
model	ks rank		KS p-value		KS 95%		KS 97.5%	
	mean	median	mean	median	mean	passing	mean	passing
Lomax	2.09	2	0.26	0.17	0.79	192	0.87	211
Log-normal	2.33	2	0.27	0.17	0.71	173	0.79	191
Expon. Weibull	2.11	2	0.29	0.23	0.77	187	0.84	203
Weibull	3.84	4	0.07	0.00	0.24	59	0.30	72



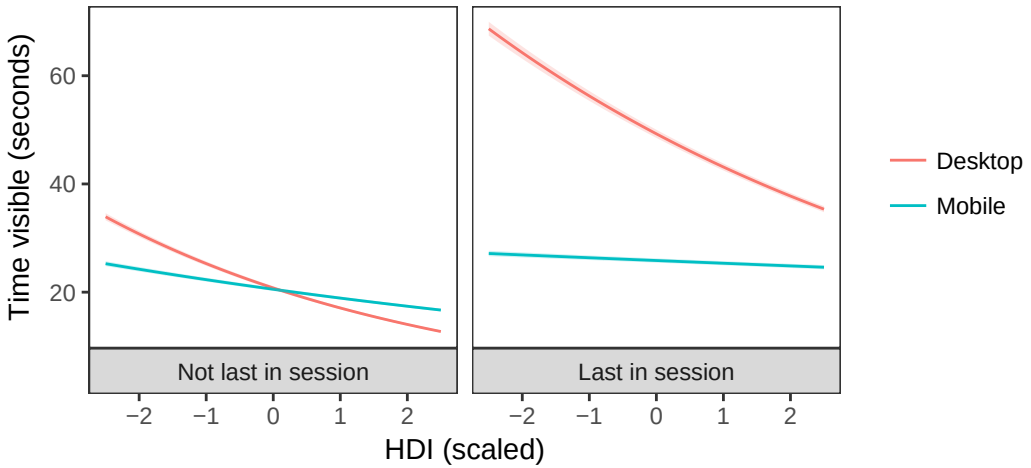
Goodness of fit statistics resulting from the model selection process on 242 wikis. The Lomax, log-normal, and exponentiated Weibull distributions fit the data reasonably well, but the Lomax most often fits the best. The ”mean” columns under KS 95%, and KS 97.5% refer to the proportion of wikis passing KS-tests at the 95% and 97.5% significance levels, and the ”passing” columns states the absolute number.

Hazard Plots

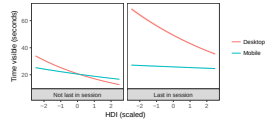




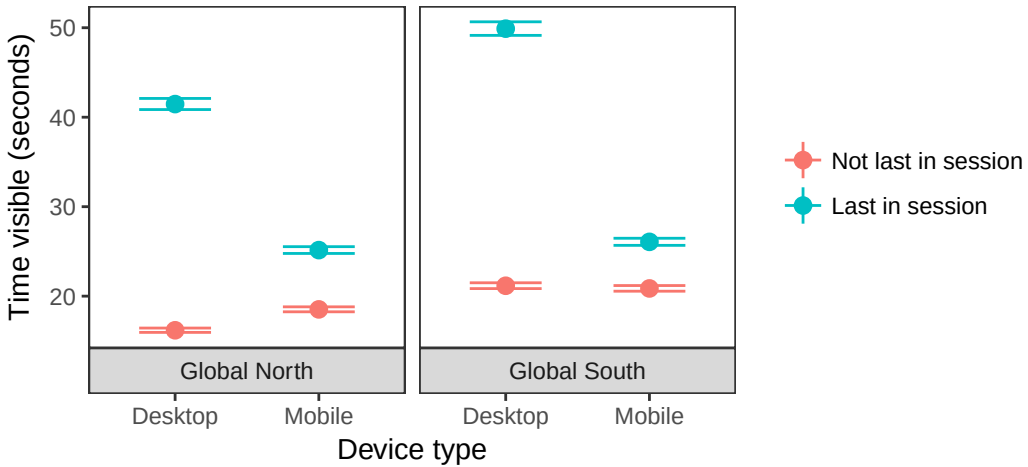
Hazard functions for the parametric models estimated on English Wikipedia. The exponentiated Weibull model (the best fit to the data) indicates that the hazard rate increases in the first seconds of a page view, after which we observe negative aging.



2019-08-17 Dwelling on Wikipedia

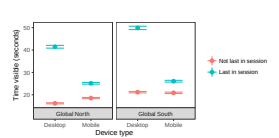


Marginal effects plot showing the relationship between HDI and reading time predicted by *model 1a*. The negative slope of the lines shows that lower-HDI readers have longer reading times, and the difference in slopes between devices shows that the relationship between HDI and reading time is more pronounced on desktop devices. The ribbons reflect 95% confidence intervals of the model coefficients. The x-axis units represent standard deviations from the mean HDI.



2019-08-17

Dwelling on Wikipedia



Marginal effects plot showing dwell times on Wikipedia pages predicted by our regression model. Compared to readers in the Global North, readers in the Global South spend substantially more time reading when on desktop devices.