Dwelling on Wikipedia: Investigating time spent by global encyclopedia readers

Nathan TeBlunthuis
Wikimedia Foundation
San Francisco, United States
University of Washington
Seattle, United States
nathante@uw.edu

Tilman Bayer
Wikimedia Foundation
San Francisco, United States
haebwiki@gmail.com

Olga Vasileva
Wikimedia Foundation
San Francisco, United States
ovasileva@wikimedia.org

ABSTRACT

Much existing knowledge about global consumption of peer-produced information goods is supported by data on Wikipedia page view counts and surveys. In 2017, the Wikimedia Foundation began measuring the time readers spend on a given page view (dwell time), enabling a more detailed understanding of such reading patterns. In this paper, we validate and model this new data source and, building on existing findings, use regression analysis to test hypotheses about how patterns in reading time vary between global contexts. Consistent with prior findings from self-report data, our complementary analysis of behavioral data provides evidence that Global South readers are more likely to use Wikipedia to gain in-depth understanding of a topic. We find that Global South readers spend more time per page view and that this difference is amplified on desktop devices, which are thought to be better suited for in-depth information seeking tasks.

ACM Classification Keywords

H.5.3. Information Interfaces and Presentation (e.g. HCI): Group and Organization Interfaces – Computer-supported cooperative work

Author Keywords

Peer production, Wikipedia, Readership, Digital divides, Quantitative methods, Web analytics, Dwell time

INTRODUCTION

How do Wikipedia readers vary across different geographic and developmental contexts? A recent study of readers of different Wikipedia language editions found that readers in countries with a lower human development index (HDI) were more likely to read for in-depth understanding compared to readers in high-HDI countries [18]. However, this study is limited by the use of self-reported data, which can be biased by effects of social desirability and self-selection due to the volunteer nature of web-based surveys [1, 14, 16, 27]. This study provides additional support for this finding from large-scale observation of reading behavior across contexts with varying levels of development.

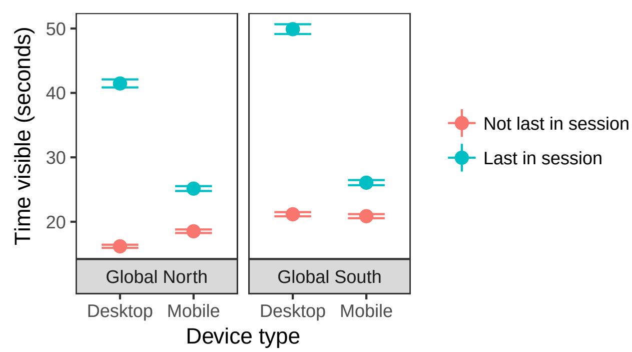

Figure 1. Marginal effects plot showing dwell times on Wikipedia pages predicted by our regression model. Compared to readers in the Global North, readers in the Global South spend substantially more time reading when on desktop devices.

Wikipedia contributors generally start as Wikipedia readers Therefore understanding and better supporting readership is important for the continued growth of the Wikimedia movement [28]. In 2017, the Wikimedia Foundation's web team introduced new instrumentation to measure the amount of time Wikipedia readers spend on the pages they view. We utilize this newly available data source, which provides additional information over the widely used page view data. With reading times in our field of view, it becomes clear that not all views are created equal. Some page views seem to involve in-depth reading, yet most are quite short.

We begin our analysis by evaluating the quality of the adopted approach for measuring reading times. We find limitations including a high rate of missing data on mobile devices and a low rate of invalid (missing or negative) measurements. However, we believe that the data can be generally informative as long as these limitations are considered. We then present a summary of the data and estimate the total time spent reading Wikipedia.

OpenSym '19, The 15th International Symposium on Open Collaboration
August 20–22, 2019, Skövde, Sweden

ACM ISBN 978-1-4503-6319-8/19/08.
DOI: https://dx.doi.org/10.1145/3306446.3340829

Next we evaluate probability models for reading time data. In addition to validating assumptions that underlie the use of parametric statistics and regression models used in answering our research questions, model selection can also help evaluate theorized data generating processes that predict when a given model will be a good fit for the data [20, 33]. For instance, Liu et al. (2010) analyze dwell times using Weibull models, finding evidence for "screen-and-glean" patterns in which people first spend a short amount of time to assess a web page, and then decide whether to read it in-depth [19]. We evaluate several probability distributions on the data from Wikipedia readers, and find that the Weibull model is not a good fit, but that the log-normal distribution fits the data well enough to justify using the geometric mean as a metric.

Finally, we return to our study of global reading behavior. Consistent with the results of Lemmerich et al., we find that readers in countries with lower HDI or in the so-called Global South spend more time reading per page view compared to readers in the Global North or in countries with higher HDI [18]. Moreover, this difference is amplified where we would expect users to consume information in depth: on the desktop (non-mobile) site. While we also hypothesized that the difference would likewise be greater in the last page-view in a session, this idea was not supported by our data analysis. We demonstrate these patterns using both multivariate regressions and a simple non-parametric analysis.

BACKGROUND

Wikipedia readership

Reading behavior on Wikipedia has been studied extensively, with a 2014 literature review listing 99 publications by 2011 [22]. Page view count data is central to this body of work when it comes to quantifying the attention readers give to particular topics or entire Wikipedia language editions. According to Priedhorsky et al., "the most common application [of page view data] is detection and measurement of popular news topics or events," with other uses including forecasting attempts (of e.g. box office revenues) and the study of Wikipedia's own processes [29]. As an example of research using it to examine information imbalances, building on earlier work by Gorbatâi and others, Warncke-Wang et al. compared page view data with article quality ratings, and found "misalignment between supply and demand", as the Wikipedia articles with the most views were often not the highest quality [9, 35]. Other, less frequently used research strategies include using click streams and session lengths [12, 24].

Surveys are another important source of information about Wikipedia readership [22]. As mentioned in the introduction, such voluntarily self-reported data are subject to participation and social desirability biases. Participation biases from self-selection may have had significant effects in the case of a previous Wikipedia reader and editor survey [14].

Some previous research on Wikipedia readership has already used an approximation of reading time that assumes that the end of a page view is always marked by a new web-request originating from the same IP and user agent [32]. Apart from the limitations arising from using the IP/user agent combination as a substitute for a user ID, this approach also does not allow measuring the dwell time for the last page view in a session.

Dwell times and information seeking

It has long been observed that page view numbers can paint a misleading picture of the amount of attention spent by web readers, or the information value a web site provides to them. An early study of search engine users found in 2003 that typical reading times were "substantially less than has been previously reported using survey data" [15]. In more recent years, metrics based on page dwell time (or total time spent on a site) have been adopted more widely. A prominent example is the online publishing platform Medium.com, which in 2013 declared "Total Time Reading" (TTR) as their "Only Metric That Matters." Distancing themselves from widely adopted web analytics metrics such as page views or active users, they argue that the act of reading should be seen as the most relevant form of user engagement for content websites [6].

Much prior work on web page dwell times focuses on applications in information retrieval and content recommendation (e.g. [17, 36, 37]). Long dwell times can signal successful information retrieval in search applications because they suggest that the user has found sought information [17]. Liu et al. analyzed dwell time data collected through a web browser plugin to characterize types of web content [19]. However, factors beyond content may influence dwell times including psychological processes of decision making and individualized styles of content consumption [37]. As we compare Wikipedia readers using mobile and desktop devices it is worth noting that dwell times are likely to be longer on desktop computers compared to mobile devices [36].

Global device and knowledge gaps

We seek to understand differences in Wikipedia's audience between the areas roughly known as the Global North and the Global South. Lemmerich et al. show empirical differences between self-reported information seeking behavior between such contexts [18]. These differences are likely related to digital divides or gaps between the knowledge, information and technology resources commonly available in different contexts, which can lead to systematic differences in reading behavior.

For people to use the Internet (or Wikipedia), they have to be able to connect to it, but not all forms of access are equally suited for a given task [34]. Deursen et al. suggest that personal computers will be better for in-depth information seeking, while mobile devices, which are often close at hand, have advantages for social interaction [34]. As Internet access becomes more ubiquitous, gaps in skills and knowledge about how to use the Internet are increasingly salient digital dividers and can be reinforced by device gaps [7, 13]. For instance, in many parts of the non-western world, mobile phones diffused before PCs, and skills for PC usage may be less widespread [21, 25]. We contribute new information about the interaction between device use around the world and how people read Wikipedia.

Gaps in skills and knowledge may also help explain gaps in who contributes to Wikipedia [31]. 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 [10]. 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 peripheries" [10]. These gaps in Wikipedia's coverage help motivate a better understanding of global readership.

In this paper, we use the Human Development Index (HDI) and the Global South/Global North regional classification as means of comparing countries separated by varying levels of development. We recognize that both are insufficient for defining economic development. Furthermore, these concepts and our measures of them only provide an incomplete understanding of the unique identities and motivations of cultures within an information-seeking context. What's more, they do not take into consideration inequality within a geographic region due to minority populations, which may affect the utility of averages such as GDP, income, and life expectancy. 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.

METHODS

Collecting reading time data

Our data collection instrument, the reading depth plugin uses the page visibility API to measure *time visible*, the total amount of time that the page was in a visible browser tab.[1] The instrument also records a second candidate measure of reading time: *total time*. This is simply the entire time the page was loaded in the browser. We used this variable for data validation and in robustness checks. We chose to focus on *time visible* because it excludes time when the user could not possibly have been reading the page. This is similar to the client-side approach described in Yi et al. (2014) [36].

Beginning November 20th 2017, we logged events from a 0.1% sample of visitor sessions.[2] The sampling rate was increased to 10% on September 25, 2018 to support future studies at a higher level of granularity.

[1]See https://meta.wikimedia.org/wiki/Schema:ReadingDepth *archived at* https://perma.cc/JK75-Y6DH and https://developer.mozilla.org/en-US/docs/Web/API/Page_Visibility_API *archived at* https://perma.cc/79PB-389J

[2]Sessions are based on a random identifier recorded in the browser's *sessionStorage*, which expires at the end of each browser session. This is more privacy-friendly than the common approach (as used in e.g. Google Analytics) of tracking users via a cookie, in that the session identifier is not sent with every request to Wikimedia servers. It also differs from session cookies in that a new identifier will be used for links opened in a new browser tab or window.

Since we care about the reading behavior of humans, we identify bots using user agent strings and exclude them from all of our analyses.[3]

Missing data

We are only able to collect data from web browsers that support the APIs on which the instrument depends. Also, we excluded certain user agents that were found to send data unreliably in our testing, namely the default Android browser, versions of Chrome earlier than 39, Safari, and all browsers running on versions of iOS older than 11.3. We also do not collect data from browsers that have not enabled JavaScript or that have enabled Do Not Track.[4]

Even when the above conditions are met, in some cases we are still not able to collect data. Sometimes we observe a page loaded event, indicating that a user in our sample opened a page, but we do not observe a corresponding event indicating that the user has left the page (a page unloaded event). This issue affects 57% percent of records on the mobile site and about 5% of records on the desktop site. The likely explanation for why many mobile views are affected is that many mobile browsers will fail to send a page-unloaded event in certain situations, such as when the user closes the browser app using the app switcher.[5] We only include page views for which we observe exactly 1 page loaded event and 1 page unloaded event and remove 0.016% of page unloaded events where, for unknown reasons, the instrument recorded a page visible time that was less than 0 or undefined.

Taking a sample

Because Wikipedia is so widely read, even a 0.1% sample results in an amount of data exceeding the statistical requirements of this analysis. We therefore conduct our analysis on random sub-samples of the collected data.

To ensure that all Wikipedia language projects are fairly and adequately represented in our sample, we use stratified sampling by assigning a *weight* to each group that adjusts the probability that members of the group are chosen in the sample. This introduces a *known bias* in the resulting sample, which is corrected using the *weights* in ways analogous to weighted averaging. For estimating total reading time, and for distribution selection, we stratify by wiki, taking up to 20,000 data points for each wiki and excluding wikis that have fewer than 300 data points. This leaves us with 242 wikis in our sample. In the multivariate analysis below, we stratify by wiki, by the country of the reader's approximate location, and by whether or not we think that the user is on a mobile device. We sample up to 200 data points for each stratum and analyze a sample of 285 wikis.

[3]See https://meta.wikimedia.org/wiki/Research:Page_view/Tags#Spider *archived at* https://perma.cc/3NSL-X6L2

[4]See https://en.wikipedia.org/wiki/Do_Not_Track *archived at* https://perma.cc/J368-ZYBD

[5]We are planning to remedy this issue in future versions of the instrumentation, by making use of alternatives to the page unloaded event available in modern browsers, e.g. the Page Lifecycle API introduced in Google Chrome in 2018.

Ethical considerations

Our approach in this paper relies on large-scale observational data collected by monitoring the behavior of Wikipedia visitors. We neither see nor speak to the humans on the other side of the screen. In addition to the empirical limitations discussed below, this approach is subject to epistemic limitations. It makes those behaviors that we can observe through browser APIs visible, while obscuring those we cannot. It cannot speak to how people in different countries understand their experience of Wikipedia [11]. Furthermore, "big data" approaches carry critical and novel ethical risks that are not easily understood in conventional informed-consent and human subjects research frameworks [4].

Wikimedia's privacy policy endeavors to clearly communicate that the information we use here will be collected, but we do not consider this an ethical license to use this data however we see fit.[6] We chose an analysis that we believe poses minimal risk to Wikipedia visitors' expectations, trust, and autonomy [8].[7] Each observation of individuals in our study was aggregated with many others at a high level of granularity. We chose to study the country level partly because our geolocation measure is most accurate at that level, but also because it is very coarse. We do not track people from one session to another, and do not look at the content of the pages they visit other than the page length. We exclude people from our analysis who indicate a wish for privacy by enabling Do Not Track in their browsers, and will discard any session identifiers remaining in the data collected for this analysis after it is complete.

DISTRIBUTION OF READING TIMES

Here we present summary statistics and a high level description of reading behavior on Wikipedia in terms of dwell times. When someone opens a given page on Wikipedia, how long do they typically stay on the page? Are reading times highly skewed? How much does reading behavior vary across different language editions of Wikipedia? How much time does all of humanity spend reading Wikipedia?

Wikipedia as a whole

In general, the distribution of reading times is very skewed (see Figure 2). The median reading time is 25 seconds and the 75th percentile is 75.1 seconds. This skewness pushes the arithmetic mean far from most of the mass of the distribution. Therefore, the geometric means, medians, and other percentiles have more utility within our discussion of reading times.

Total time spent

Based on our data, we estimate that humanity spent about *672,349 years* reading Wikipedia from November 2017 through October 2018. We calculated this estimate as the product of the mean reading time on each Wikipedia wiki by the number of page views on that wiki, excluding readers using the mobile apps and identified bots. It is possible that some people leave Wikipedia pages visible in their browsers for extended periods of time without reading. To make our estimates of total reading time in this section somewhat conservative, we rounded all page views down to 1 hour.

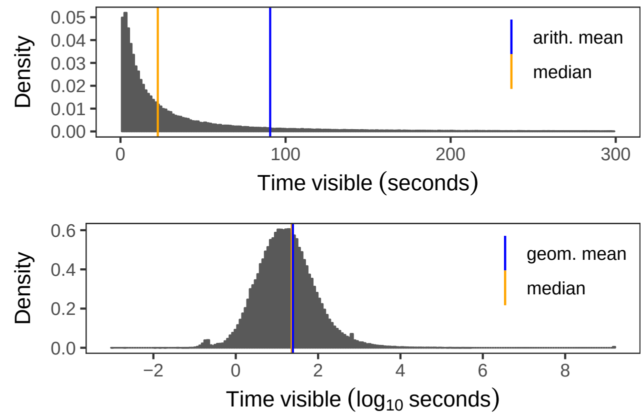

Figure 2. The distribution of dwell times across 242 language editions of Wikipedia. The top chart shows a histogram of dwell times less than one hour long (the x-axis is truncated to 300 seconds for clarity). In this chart we can see that the median dwell time is about 25 seconds long and that the distribution of dwell times is very skewed, with the arithmetic mean far from the median. The y-axis represents the probability that a given page view is in a given box. In the lower figure, the dwell times are log-transformed and the data appear bell-shaped, with some skew to the right.

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

Table 1. Percentiles for reading times (in seconds) on selected Wikipedia editions

Variation between different language editions

Figure 3 shows kernel density estimates of the distribution of page visible times on several Wikipedia language editions selected to highlight projects of different sizes and of different cultures. These are Arabic (ar), German (de), English (en), Spanish (es), Hindi (hi), Dutch (nl) and Punjabi (pa). As above, we place unscaled data side-by-side with log-transformed data. Only the log-transformed plots show the full range of the data. Similar kernel density plots for other languages as well as box-and-whisker plots are available in our online supplement.[8]

[6]See https://foundation.wikimedia.org/wiki/Privacy_policy *archived at* https://perma.cc/C4VQ-HWRT

[7]We followed the Wikimedia Foundation's (WMF) guidelines and processes for conducting research. As it is not a federally funded institution, research at the WMF is not supervised by an institutional review board (IRB).

[8]Available at https://w.wiki/5Jo.

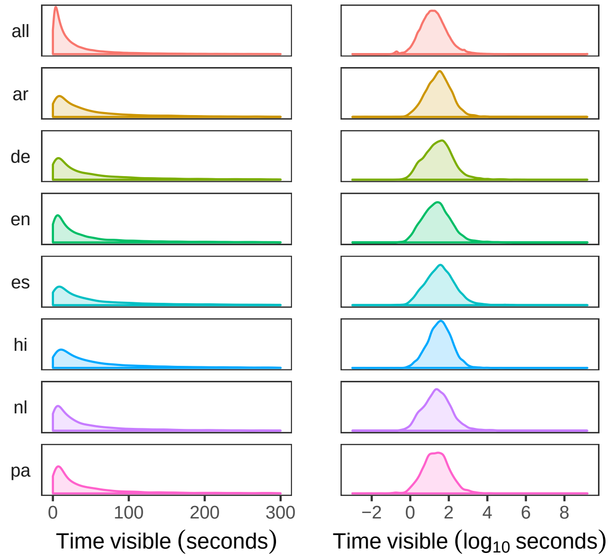

Figure 3. Kernel density plots of the distribution of dwell times on a selection of wikis. Spanish, Hindi, and Arabic appear to have longer reading times while English and Punjabi appear to have somewhat shorter reading times. In general, the distribution is very skewed, as these example wikis demonstrate.

UNIVARIATE MODEL SELECTION

Motivation

Analysts of reading times on Wikipedia will wish to make parametric assumptions to justify the use of statistical models for evaluating experiments, drawing comparisons between different samples of reading times, and performing multivariate analyses as we do below. This requires assuming a probability distribution with interpretable parameters such as mean, variance, and shape parameters. Fitting parametric distributions to data allows us to estimate these parameters and to statistically test changes in the parameters. However, parametric models can mislead if they don't fit the data well. Below, we evaluate several models.

Candidate models

We consider the following distributions in our model-selection process.

Log-normal distribution: This is a normal distribution, but on a logarithmic scale. Differences in means between log-normal samples can be tested using t-tests. Such advantages make the log-normal distribution a common choice in analyzing skewed data, even when it is not a perfect fit.

Lomax (Pareto Type II) Distribution: Datasets on human behavior often exhibit power-law distributions, meaning that the probability of extreme events, while still low, is much greater than would be predicted by a normal (or log-normal) distribution [5]. We fit the Lomax Distribution, a commonly used long-tailed distribution with two parameters that assumes that power law dynamics occur over the whole range of the data.

Weibull Distribution: Liu et al. model reading times on web pages using a Weibull Distribution [19]. This model has two parameters: λ, a scale parameter, and k, a shape parameter. The Weibull distribution can be a useful model because of the intuitive interpretation of k. If $k > 1$, then reading behavior exhibits positive aging, which means that the longer someone stays on a page, the more likely they are to leave the page at any moment. Conversely $k < 1$ is interpreted as negative aging, which means that as someone remains on a page, they become less likely to leave the page at any given moment. The Weibull distribution is often used in the context of reliability engineering for modeling the chances that a given part will fail at a given moment.

Exponentiated Weibull Distribution: The Weibull model assumes that the rate of readers leaving a page changes monotonically over time. This implies there must be either negative aging, positive aging, or no aging. It excludes more complicated dynamic processes where positive aging gives way to negative aging after a point in time. The exponentiated Weibull distribution is a three-parameter generalization of the Weibull distribution that relaxes this constraint [23]. The extra degree of freedom will allow this model to fit a greater range of empirical distributions compared to the two-parameter Weibull model.

Methods

Our method for model selection is inspired in part by Liu et al., who compared the log-normal distribution to the Weibull distribution of dwell times on a large sample of web pages [19]. They fit both models to data for each web page, and then compare two measures of model fit: the log-likelihood, which measures the probability of the data given the model (higher is better), and the Kolmogorov-Smirnov distance (KS-distance), which is the maximum difference between the model CDF and the empirical CDF (lower is better). For the sample of web pages they consider, the Weibull model outperformed the log-normal model in a large majority of cases according to both goodness-of-fit measures.

Similar to the approach of Liu et al., we fit each of the models we consider on reading time data, separately for each Wikipedia project [19]. In addition to the KS-distance, we also use KS-tests of the null hypothesis that the model is a good fit for the data to evaluate goodness-of-fit [5]. For the samples sizes we use, passing the KS-test is a high bar.

Adding parameters can increase model fit without improving out-of-sample predictive performance or explanatory power. To make fair comparison between models with different numbers of parameters, we use the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) instead of the log-likelihood. Both criteria attempt to quantify the amount of information lost by the model (lower is better), by evaluating the log likelihood, and adding a penalty for the model parameters. The difference between AIC and BIC is that BIC maintains the penalty for larger sample sizes.[9]

Following Liu et al., we build these goodness-of-fit measures for each wiki and rank them from best to worst [19]. For each distribution, we report the mean and median of these ranks.

[9]We provide a more detailed example of this procedure in our online supplement at https://w.wiki/5Jo.

In addition, we report the mean and median p-values of the KS-tests as well as the number and proportion of wikis that pass the KS-test for each model.

We fit the models using SciPy. The exponentiated Weibull, Weibull, and Lomax models were fit using maximum likelihood estimation and the log-normal distributions were fit using the method of moments.

RESULTS

Table 2 below shows the results of this procedure. The Lomax, exponentiated Weibull, and Log-normal all fit the data reasonably well. All pass the KS-test for many wikis, and are in a three-way tie for best median rank according to AIC. Despite this, none of our candidate models pass the KS test for all wikis: There are 28 wikis where all 4 models fail to pass at the 95% level, and 13 wikis where they all fail at the 97.5% level.

The Lomax distribution is the best fit across all wikis according to all metrics. With only 2 parameters, it has a lower AIC and BIC than the three-parameter exponentiated Weibull distribution and passes the KS-test 79% of the time at the 95% confidence level. The exponentiated Weibull model fits the data better than the log-normal model in terms of passing KS-tests and with respect to AIC. However, the log-normal is better in terms of BIC, which imposes a greater penalty on the additional parameter of the exponentiated Weibull model.

The Weibull model fits substantially worse than the Lomax, log-normal, and exponentiated Weibull in terms of all of our goodness-of-fit metrics. In this respect, our results differ from those of Liu et al., who observed the Weibull model fitting dwell time data better than the Log-normal model [19]. We observe that for dwell times on Wikipedia, the Log-normal model is the better fit. While substantially worse than the Lomax model, the Log-normal model still passes the KS-test at the 95% level for about 71% of wikis in the sample.

Discussion

We found that the Lomax, exponentiated Weibull, and log-normal models all fit the data within reason. We now discuss how each of these models can be applied to understanding Wikipedia reading behavior.

Lomax (Pareto Type II) Distribution: That the Lomax model fits well suggests that Wikipedia reading times may follow a power law. Mitzenmacher (2004) describes several possible data generating processes for power law (Pareto) and log-normal distributions [20]. Rich-get-richer dynamics such as preferential attachment are commonly associated with power law distributions, and a mixture of Log-normal distributions can also generate a power law [20]. Deeper exploration of potential power-law dynamics in reading behavior is a potential avenue for future research.

Log-Normal Distribution: The log-normal model does not fit the data perfectly, but it fits well enough to be useful. It frequently passes KS-tests, and is preferred to the exponentiated Weibull by the BIC. Even though the Lomax model typically fits the data better, assuming a log-normal model justifies using t-tests to compare differences in geometric means when evaluating experiments. Furthermore, assuming log-normality can help justify using ordinary least squares to estimate regression models in multivariate analysis (as we do below) instead of models that require maximum likelihood estimation.

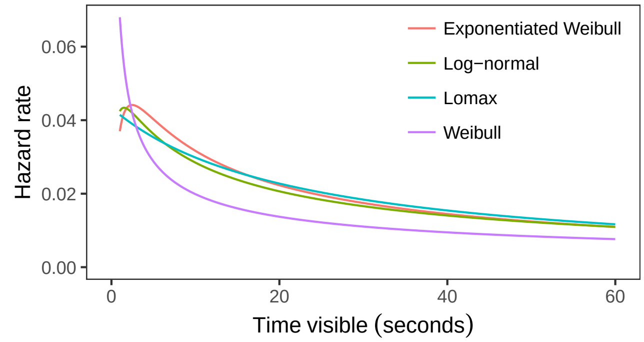

Figure 4. Hazard functions for the parametric models estimated on English Wikipedia. The exponentiated Weibull model (the best fit to the data) indicates that the hazard rate increases in the first seconds of a page view, after which we observe negative aging.

Weibull Distribution: The Weibull model did not fit the data well. While Liu et al. observed that the Weibull model outperformed the log-normal model on their datasets, we (along with [37]) observe the opposite. However, the exponentiated Weibull model generalizes the Weibull, is a good fit for the data, and can help us explain why the Weibull does not fit the data well.

Exponentiated Weibull Distribution: The exponentiated Weibull has 3 parameters [23]. Two are shape parameters ($\alpha > 0$ and $\gamma > 0$) and one is a scale parameter ($\lambda > 0$). The major qualitative distinctions in interpreting the model depend on the shape parameters. In many cases the parameters can be interpreted in terms of a transition from negative to positive aging (or visa-versa) after some threshold. However, if either $\gamma > 1$, $\alpha < 1$ or $\gamma < 1$, $\alpha > 1$ then qualitative interpretation may require closer inspection of estimated hazard functions.

Inconveniently, we estimated $\alpha > 1$ and $\gamma < 1$ for all but one of the 285 Wikipedia projects we analyzed. This limits the usefulness of exponentiated Weibull models for large-scale analysis on many wikis because the parameters are outside the area where the model leads directly to intuitive qualitative interpretations. However, by plotting the estimated hazard function we can see over what range of the data the hazard function is decreasing or increasing, accelerating or decelerating.

In figure 4 we observe that, on English Wikipedia, the log-normal and exponentiated Weibull models both indicate a brief period of positive aging, during which the instantaneous rate of page-leaving increases, followed by negative aging. This helps explain why the Weibull model is not a good fit for the data compared to the log-normal and exponentiated Weibull models: the Weibull distribution cannot model a non-monotonic hazard function. While Liu et al. found it to

model	AIC rank		BIC rank		ks rank		KS p-value		KS 95%		KS 97.5%	
	mean	median	mean	median	mean	median	mean	median	mean	passing	mean	passing
Lomax	1.78	2	1.70	1	2.09	2	0.26	0.17	0.79	192	0.87	211
Log-normal	2.20	2	2.10	2	2.33	2	0.27	0.17	0.71	173	0.79	191
Expon. Weibull	2.15	2	2.34	3	2.11	2	0.29	0.23	0.77	187	0.84	203
Weibull	3.98	4	3.94	4	3.84	4	0.07	0.00	0.24	59	0.30	72

Table 2. 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.

be a good model for the distribution of dwell times in data collected through a web browser plugin, our analysis suggests that the behavior of Wikipedia readers may be somewhat more complex. Perhaps whereas Liu et al. operationalized "screen-and-glean" as a monotonically decreasing hazard function, Wikipedia readers require more than 1 or 2 seconds to "screen" the page and during these first few moments, their hazard of leaving it increases.

READING TIME AND GLOBAL CONTEXTS

Now we return to our analysis of Wikipedia readers in a global context. Our analysis is most closely inspired by Lemmerich et al.'s large-scale global survey of Wikipedia readers. They found that readers in lower-HDI countries are more likely to use Wikipedia in educational contexts and for intrinsic learning, but not for fact-checking [18]. Such motivations and contexts are likely to involve longer sessions and dwell times compared to fact-checking [18, 32]. Therefore, we predict that readers in lower-HDI countries and in the Global South are more likely to have longer dwell times on Wikipedia articles.

H1: Readers in countries with lower HDI (or the Global South) are more likely to spend more time reading each page they visit compared to readers in countries with higher HDI (or the Global North).

We also test a second prediction of the theory that Global South readers are more likely to use Wikipedia for in-depth understanding. 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). Furthermore, 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 on desktop devices.

H2: The difference between the reading times of readers in countries with lower HDI compared to readers in higher-HDI countries will be greater on desktop than on mobile devices.

Based on the "screen-and-glean" model of information seeking behavior that Liu et al. observed on the web [19], we propose that reading of articles for in-depth understanding is most likely to take place in the last page view in a session. Differences in reading time in other page views might be attributable to less efficient "screening"—gaps in the skills required to efficiently sift through Wikipedia pages to find the page with the information sought. However, the final page view in a session may reflect "gleaning"—information consumption. If so, then the last page view in a session provides an opportunity to isolate information consumption from information seeking.

Therefore if the gap between low and high development context readers is attributable to types of information seeking tasks, and in-depth reading tasks require more time spent "gleaning," then we predict that the gap between reading time in low versus high HDI countries will also be amplified on the last page view in a session.

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.

On the other hand, a "skills gap" with respect to information screening may drive an opposite result. The gap between reading times in the Global South and the Global North may shrink on the last page view in a session if Global South readers are less efficient at filtering information.

Methods and measures

The EventLogging system records the date and time the page was viewed. We include *Day-Of-Week* and *Month* as statistical controls for seasonal and weekly reading patterns. Including *NthInSession* statistically adjusts for the number of pages a reader has viewed so far in the session. *Revision Length*, the size of the wiki-page, measured in bytes, roughly accounts for the amount of content on the page. We use two other measures from the instrument to statistically adjust for page load time: *time till first paint*, the time from the request until the browser starts to render any part of the page; and *dom interactive time*, the time from the request until the user can interact with the page.[10]

We obtain the *page length*, measured in bytes at the time the page was viewed, by merging the EventLogging data with the edit history. To understand how reading behavior on *mobile* devices differs from behavior on non-mobile (i.e. desktop) devices, we assume that visitors to mobile web-hosts (e.g. en.m.wikipedia.org) are using mobile devices and that visitors to non-mobile web-hosts (e.g. en.wikipedia.org) are on non-mobile (desktop) devices.

[10]See `https://developer.mozilla.org/en-US/docs/Web/API/PerformanceNavigationTiming/domInteractive` *archived at* `https://perma.cc/RRA8-8SQG`, DOM refers the page's "document object model" structure

We determine the approximate country in which a reader is located from the MaxMind GeoIP database which is integrated with the Wikimedia analytics pipeline.[11] We use the United Nations' human development index (*HDI*) to measure the development level of the country.[12] We lack geolocation data before March 3rd 2018, which limits our analysis of reading times in the global context to the period from then until September 28th 2018. We standardize the HDI by centering to 0 and scaling it by the standard deviation (taken at the country level) because the partial residual plots of interaction term between (unscaled) HDI and mobile were very skewed. This also allows us to interpret results in terms of standard deviations.

We also use the established regional classifications of Global North and Global South[13] as a second, dichotomous, measure of development. Finally, the EventLogging instrumentation retains a session token with which we measure whether or not a given page view is the *last-in-session*. We also statistically adjust for the number of pages viewed in the session so far (*Nth in session*).

Models

We test the three hypotheses using two regression models that differ only in how they represent economic development. *Model 1a* uses the human development index (HDI) and *model 1b* uses the Global North / Global South regional classification. Here is the specification of *model 1a*:

$$
\begin{gathered}
Y = B_0 + B_1 HDI + B_2 Mobile + B_3 Mobile\ x\ HDI \\
+ B_4 RevisionLength + B_5 DayOfWeek + B_6 Month \\
+ B_7 NthInSession + B_8 LastInSession \\
+ B_9 HDI\ x\ LastInSession + B_{10} Mobile\ x\ LastInSession \\
+ B_{11} FirstPaint + B_{12} DomInteractiveTime
\end{gathered}
$$

The formula for *model 1b* is the same except for using *GlobalNorth* terms instead of *HDI*.

We consider **H1** supported if $B_1 < 0$ in both models; **H2** if $B_3 > 0$; and **H3** if $B_9 < 0$. Because interaction terms can be difficult to interpret qualitatively, we will present marginal effect (ME) plots to assist in qualitative interpretation of the observed relationships [26].

We explored alternative model specifications that include higher order terms and additional interaction terms. We choose to present *model 1a* and *model 1b* because more complex models neither substantively improve the explained variance and the predictive performance nor lead to qualitatively different conclusions. We fit both models using weighted ordinary least squares estimation in R on a stratified sample of size 9,873,641.

[11] See https://wikitech.wikimedia.org/wiki/Analytics/Systems/Cluster/Geolocation *archived at* https://perma.cc/C36T-2E4E

[12] From http://hdr.undp.org/en/data *archived at* https://perma.cc/SLQ3-HS8S. The HDI is a number between 0 and 1.

[13] See https://meta.wikimedia.org/wiki/List_of_countries_by_regional_classification *archived at* https://perma.cc/WHN7-GB9D

Non-parametric Analysis

Our multivariate regression analysis assumes a parametric model and as we saw in the univariate analysis above, the assumption of log-normality may not be valid for every Wiki. Therefore, we also provide a simple non-parametric analysis based on median reading times. Unlike the regression analysis, the non-parametric analysis does not include statistical controls or afford statistical hypothesis tests, but it avoids having to depend on assumptions about the distribution. We construct a 3x3 table of users depending on whether they are in the Global North or Global South, on a mobile or desktop device, or on the last page view in their session. The medians of each cell of the table validate that our findings are not driven by the normality assumption alone.

RESULTS

We use marginal effects (ME) plots to interpret our regression models.[14] A marginal effects plot shows how the model's predicted outcome varies with respect to one or more of the predictors when other terms of the model are held constant at some typical value [26]. Since we are interested in comparing reading times between last-in-session page views and other page views, we create two marginal effects plots for each model: one for last-in-session page views and one for non-last-in-session page views. Similarly, we also break down predicted reading times by device type.

For each marginal effects plot, the y-axis shows the model predicted values and the x-axis shows the values of the predictor variables. In the marginal effects plots shown here, uncertainty intervals represent confidence intervals of the parameter estimates, not uncertainty about the model predictions. Uncertainty about model predictions in this case is generally very high, as our models explain only a small fraction (about 7%) of the variance in reading times.

Hypothesis 1: Global context and reading times

We find support for **H1**: that readers in higher-HDI countries ($B = -0.20$, $SE = 0.002$) or in the Global North ($B = -0.27$, $SE = 0.002$) are likely to spend less time on each page than readers in lower HDI countries or in the Global South. For illustration, our ME plot for *model 1a* (figure 5) shows that, 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. Similarly, per our ME plot for *model 1b* (figure 1), for last-in-session page views on desktop devices, a prototypical Global North reader is predicted to spend around 42 seconds per page view compared to the 50 seconds spent by a prototypical Global South reader.

Hypothesis 2: Global context and mobile devices

We also find support for **H2**: that readers in the Global North ($B = 15$, $SE = 0.002$) or higher-HDI ($B = 0.11$, $SE = 0.002$) countries are likely to spend

[14] Full regression tables are available in the appendix.

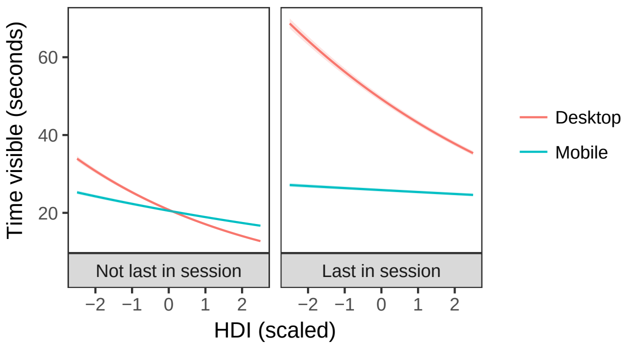

Figure 5. Marginal effects plot showing the relationship between HDI and reading time predicted by *model 1a*. The negative slope of the lines shows that lower-HDI readers have longer reading times, and the difference in slopes between devices shows that the relationship between HDI and reading time 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.

even less time reading compared to Global South or lower-HDI readers when they are on a desktop device compared to a mobile device. This is clearly visible as a differences in slopes in figure 5. Indeed, for pages views other than the last-in-session, the predicted reading times for prototypical readers in countries 1 standard deviation below the mean decreases from 25 seconds on desktop devices to 22 seconds on mobile devices, but the reverse is true for readers in higher-HDI countries. In a country 1 standard deviation above the mean, an otherwise comparable reader is predicted to read for about 19 seconds on mobile and about 17 seconds on desktop. The ME plot for *model 1b* (figure 1) shows that for the prototypical reader, the gap between Global South and Global North is greater on desktop devices (about 5 seconds) than on mobile devices (about 3 seconds).

Hypothesis 3: Global context and last-in-session

Based on the "screen-and-glean" results by Liu et al, we expected in-depth reading to be most likely in the last page view in a session, and thus predicted **H3:** the difference in reading times between lower-HDI countries and higher-HDI countries will be amplified in the last page view in a session. However, we do not find support for this hypothesis, which would have been indicated by a negative regression coefficient for the interaction term between development and last-in-session. Instead we find a positive coefficients for *HDI:Last in session* (B $= 0.63$, $SE = 0.002$) in *model 1a* and for *Global North:Last in session* (B $= 0.08$, $SE = 0.002$) in *model 1b*.

Non-parametric Analysis

Table 3 shows the median time pages are visible by the user's economic region, device and whether a page is the last viewed in the user's session. Consistent with **H1**, median users in the Global South spend more time on pages compared to median users in the Global North regardless of device or session stage. Consistent with **H2**, the difference between Global South and Global North users is clearly more pronounced on desktop compared to mobile. In contrast to the prediction of **H3**, but in line with the findings from our parametric analysis, we do not observe an accentuation of the difference between Global South and Global North users in the last page view in a session.

Page length

In addition to the above results on reading times and global contexts, we also examined how reading times relate to page length. The association between page length and reading times is small and positive (B $= 0.17, SE = 0.0004$). Pages on Wikipedia vary greatly in length: from just a few bytes up to 2,000,000 bytes. 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.[15]

LIMITATIONS

Two important technical limitations of our dwell time data affect our ability to compare reader behavior between mobile phone and PC devices. The first is missing data on mobile devices, discussed above. This missing data likely introduces a negative bias to our measures of reading time on mobile devices because we believe observations are more likely to be lost when users switch tasks from the browser, and subsequently return to reading. This bias may be quite significant as the issue affects a large proportion of our sample.

The second limitation occurs when readers leave a page visible in the browser at times when they are not directly reading it. For example, a user may have multiple windows visible while only looking at one of them, or may leave a browser window visible and move away from the computer for a long period of time. In general, the best we can hope to observe

[15] See our online supplement at `https://w.wiki/5Jo` for a marginal effects plot. Page length refers to the size of the wikitext source of the page measured in bytes. Not every byte corresponds to a character of readable text. Wikitext source also includes code for formatting, using templates, or embedding images. Additionally, some characters, especially in non-Latin alphabets, may take up multiple bytes. Still our results confirm that for longer Wikipedia articles, only a fraction of the text is read in a typical page view. 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), these 30 seconds would only suffice to read through less than 1000 of these 20000 bytes [2, 3].

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 3. 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.

is that a page is visible in a browser. We cannot, through this instrument alone, know with confidence that an individual is reading. This limitation leads to positive bias in our measures of reading time. To partially address this limitation, we fit regression models on data with dwell times greater than 1 hour removed (assuming that it contains a higher ratio of those "visible but not reading" cases), and found that our results were not substantively affected by the change.

It is possible that this positive bias may correlate with our analytic variables. Perhaps last-in-session views may be particularly subject to this source of bias and may contribute to the gap we observe between reading times in last-in-session page views compared to others. We designed our analysis of H1 and H2 to account for differences between last-in-session and other page views, and found that the sign of the observed differences remained the same whether the view was the last in a session or not. We did not find support for H3, which considered differences within last-in-session page views.

Additional steps could be taken to construct new measures of reading that would not suffer this limitation through browser instrumentation to track mouse movements or scroll positions. However, such steps should be taken with care as additional data collection may negatively affect users in terms of privacy, browser responsiveness, page load times, and power consumption.

Finally, readers should keep in mind that we analyzed observational, not experimental, data with the intention to describe correlations between our variables, not to demonstrate causal relationships. We used ordinary least squares analysis, but future analysis might better account for the hierarchical structure of our data using multilevel modeling.

Alternative explanations

Furthermore, there are several plausible alternative explanations that we cannot rule out in the presented analysis. The observed reading time gap between more and less developed countries may be due to factors other than the types of information seeking tasks in which readers are engaged. For instance, if readers experience knowledge gaps in less developed countries, they may be likely to read in languages that are not their primary language, and thus spend more time reading regardless of task [10]. A future iteration of this project may partially address this limitation by accounting for whether a Wikipedia edition is a common primary language in the reader's country.

Another alternative explanation may be that the gap between readers in more and less developed countries is partly due to time spent on exploration ("screening") rather than on content consumption ("gleaning"). Our finding rejecting **H3**, suggests this, as Global South readers have longer dwell times on non-last-in-session page views compared to Global North readers. We also observe shorter non-last-in-session page views on desktop devices compared to mobile for Global North readers, but for Global South readers such page views are about the same length no matter what device is used. This unexpected result would be consistent with a skills gap experienced by Global South readers who may have greater difficulty finding sought information, especially when using desktop devices [34]. The present analysis offers only tentative support for this claim, but we suggest it as an avenue for future research.

Global South readers may also be more sensitive to the price of downloading data and thus they may avoid opening pages that they are unlikely to read in-depth. Future work might use data from the Wikipedia Zero project to study the relationship between price sensitivity and Wikipedia audiences. More generally, drawing conclusions about information seeking from our analysis rests on strong assumptions about relationships between task type and reading times. Future work on information seeking behavior on Wikipedia testing these assumptions would help validate such conclusions.

DISCUSSION AND CONCLUSION

In an analysis of novel data from Wikipedia, measuring the time that web pages are visible in the browser window as an approximation of reading time, we investigated patterns of reader behavior across global contexts and found systematic differences consistent with greater use for in-depth understanding in lower-HDI countries compared to higher-HDI countries. We believe this analysis should strengthen confidence in similar findings from surveys of reader behavior because our data have complementary strengths and limitations compared to self-report data.

We conclude that Global South readers are more likely to engage in in-depth information seeking when reading Wikipedia compared to Global North readers. Consistent with Lemmerich et al.'s survey results [18], we find that readers in lower-HDI countries have longer reading times than readers in higher-HDI countries, and that this difference is greater for users of non-mobile (desktop) devices.

The observed relationships are quite similar whether measured using the human development index (HDI) or dichotomized economic region (Global South / Global North). These relationships are supported not only by the regression models, but also by non-parametric analysis. While Wikipedia readers increasingly use mobile devices to visit Wikipedia, they are likely to spend the most time reading when they are in the last page view of a desktop session. This is exactly when we expect them to gain in-depth understandings of topics.

We lack evidence to fully explain our findings in terms of structural and socioeconomic differences between the Global North and Global South. One possibility is that the gap in reading times reflects differences in information seeking and content understanding skills [34, 31]. That we did not observe the gap between global contexts widen in last-in-session page views tentatively suggests that Global South readers are more likely to struggle to find and filter information on Wikipedia compared to Global North readers.

However, given the evidence that Wikipedia readers in the Global South are more likely to engage in deeper information seeking tasks [18], we conjecture that the gap in reading times may be explained by the quality and accessibility of the information on Wikipedia relative to alternatives available in the reader's contexts. Wikipedia may not be perfect, but given historical inequalities in education, and knowledge production between the Global South and Global North [10], it still might be competitive compared to other sources, especially when it comes to encyclopedic content about the Global South, content in local languages, and information not otherwise available for free to Internet users. This would explain why Global South readers would be more likely to choose Wikipedia when seeking in-depth information. Future research might test this hypotheses in audience surveys or by adapting approaches previously applied to gender comparisons on English Wikipedia [30].

Another contribution of this study is to vet the reading time data to understand its limitations and to conduct model selection to justify parametric assumptions for future analysts. We found a high rate of missing data on mobile, among other less significant irregularities. Future analysts should keep this in mind and work to improve the coverage. We found that the log-normal distribution often fits the data well, and therefore adopted the use of geometric means as a metric for comparing samples reading times. This also helped support our decision to adopt ordinary least squares regression analysis for multivariate comparison. However, we also found that exponentiated Weibull and Lomax probability models were often an even better fit. Future researchers might explore how reader behavior may generate data in processes consistent with these models.

The reading time data we used in this study is a promising tool for future researchers to improve upon studies of page views for understanding Wikipedia's audiences. For example, recent research has shown widespread misalignment between how often articles are visited and the quality of those articles [35]. However, we have observed that not all views are created equal. Future studies on the relationship between content production and content consumption on Wikipedia might use reading time data to learn about how content consumption might change depending on article quality.

ACKNOWLEDGEMENTS

We are grateful to the anonymous reviewers, whose observations helped improve the paper. Specials thanks to the web team at the Wikimedia Foundation that built the instrumentation, to Zareen Farooqui who conducted initial data quality vetting as part of her data analyst Outreachy internship at WMF, and to the Foundation's Analytics Engineering team for supporting the data analysis infrastructure used in this work. Thanks to those who provided comments on various stages of this research, including Kaylea Champion, other members of the Community Data Science Collective, Johnathan Morgan, Aaron Halfaker, Isaac Johnson, Miriam Redi, Abbey Ripstra, and other members of the Wikimedia research team. Special thanks to Benjamin Mako Hill for his comments and advice. This work was completed while Nathan TeBlunthuis was a PhD student at the University of Washington and in his capacity as a Wikimedia Foundation contractor and affiliate. It was also supported by the National Science Foundation (GRFP-2016220885).

REFERENCES

1. Judd Antin and Aaron Shaw. 2012. Social Desirability Bias and Self-Reports of Motivation: A Study of Amazon Mechanical Turk in the US and India. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2925–2934. DOI: http://dx.doi.org/10.1145/2207676.2208699
2. Timothy I. Bell. 2001/00/00. Extensive Reading: Speed and Comprehension. *Reading Matrix: An International Online Journal* 1, 1 (2001/00/00).
3. Vladimir V. Bochkarev, Anna V. Shevlyakova, and Valery D. Solovyev. 2012. Average Word Length Dynamics as Indicator of Cultural Changes in Society. *arXiv:1208.6109 [cs]* (Aug. 2012). http://arxiv.org/abs/1208.6109
4. danah boyd and Kate Crawford. 2012. Critical Questions For Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon. *Information, Communication & Society* 15, 5 (June 2012), 662–679. DOI:http://dx.doi.org/10.1080/1369118X.2012.678878
5. A. Clauset, C. Shalizi, and M. Newman. 2009. Power-Law Distributions in Empirical Data. *SIAM Rev.* 51, 4 (Nov. 2009), 661–703. DOI: http://dx.doi.org/10.1137/070710111
6. Pete Davies. 2013. Medium's Metric That Matters: Total Time Reading. (Nov. 2013). https://medium.com/data-lab/mediums-metric-that-matters-total-time-reading-86c4970837d5
7. Alexander J. A. M. Van Deursen, Ellen Helsper, Rebecca Eynon, and Jan A. G. M. van Dijk. 2017. The Compoundness and Sequentiality of Digital Inequality. *International Journal of Communication* 11 (Jan. 2017), 452–473.
8. Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society* 4, 1 (Jan. 2018), 2056305118763366. DOI: http://dx.doi.org/10.1177/2056305118763366
9. Andreea D. Gorbatâi. 2011. Exploring Underproduction in Wikipedia. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration (WikiSym '11)*. ACM, New York, NY, USA, 205–206. DOI: http://dx.doi.org/10.1145/2038558.2038595
10. Mark Graham, Bernie Hogan, Ralph K. Straumann, and Ahmed Medhat. 2014. Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers* 104, 4 (July 2014), 746–764. DOI:http://dx.doi.org/10.1080/00045608.2014.910087
11. Mark Graham and Taylor Shelton. 2013. Geography and the Future of Big Data, Big Data and the Future of Geography. *Dialogues in Human Geography* 3, 3 (Nov. 2013), 255–261. DOI: http://dx.doi.org/10.1177/2043820613513121

12. Aaron Halfaker, Os Keyes, Daniel Kluver, Jacob Thebault-Spieker, Tien Nguyen, Kenneth Shores, Anuradha Uduwage, and Morten Warncke-Wang. 2015. User Session Identification Based on Strong Regularities in Inter-Activity Time. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 410–418. DOI: http://dx.doi.org/10.1145/2736277.2741117

13. Eszter Hargittai. 2002. Second-Level Digital Divide: Differences in People's Online Skills. *First Monday* 7, 4 (April 2002). DOI: http://dx.doi.org/10.5210/fm.v7i4.942

14. Benjamin Mako Hill and Aaron Shaw. 2013. The Wikipedia Gender Gap Revisited: Characterizing Survey Response Bias with Propensity Score Estimation. *PLoS ONE* 8, 6 (June 2013). DOI: http://dx.doi.org/10.1371/journal.pone.0065782

15. Bernard J Jansen and Amanda Spink. 2003. An Analysis of Web Documents Retrieved and Viewed. In *International Conference on Internet Computing*. CSREA Press, Las Vegas, Nevada, 65–69.

16. Sara Kiesler and Lee S. Sproull. 1986. Response Effects in the Electronic Survey. *Public Opinion Quarterly* 50, 3 (Jan. 1986), 402–413. DOI: http://dx.doi.org/10.1086/268992

17. Youngho Kim, Ahmed Hassan, Ryen W. White, and Imed Zitouni. 2014. Modeling Dwell Time to Predict Click-Level Satisfaction. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining (WSDM '14)*. ACM, New York, NY, USA, 193–202. DOI: http://dx.doi.org/10.1145/2556195.2556220

18. Florian Lemmerich, Diego Sáez-Trumper, Robert West, and Leila Zia. 2019. Why the World Reads Wikipedia: Beyond English Speakers. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)*. ACM, New York, NY, USA, 618–626. DOI: http://dx.doi.org/10.1145/3289600.3291021

19. Chao Liu, Ryen W. White, and Susan Dumais. 2010. Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '10)*. ACM, New York, NY, USA, 379–386. DOI: http://dx.doi.org/10.1145/1835449.1835513

20. Michael Mitzenmacher. 2004. A Brief History of Generative Models for Power Law and Lognormal Distributions. *Internet Mathematics* 1, 2 (Jan. 2004), 226–251. DOI: http://dx.doi.org/10.1080/15427951.2004.10129088

21. Philip M. Napoli and Jonathan A. Obar. 2014. The Emerging Mobile Internet Underclass: A Critique of Mobile Internet Access. *The Information Society* 30, 5 (Oct. 2014), 323–334. DOI: http://dx.doi.org/10.1080/01972243.2014.944726

22. Chitu Okoli, Mohamad Mehdi, Mostafa Mesgari, Finn Årup Nielsen, and Arto Lanamäki. 2014. Wikipedia in the Eyes of Its Beholders: A Systematic Review of Scholarly Research on Wikipedia Readers and Readership. *Journal of the Association for Information Science and Technology* 65, 12 (2014), 2381–2403. DOI:http://dx.doi.org/10.1002/asi.23162

23. Manisha Pal, M. Masoom Ali, and Jungsoo Woo. 2006. Exponentiated Weibull Distribution. *Statistica* 66, 2 (2006), 139–147. DOI: http://dx.doi.org/10.6092/issn.1973-2201/493

24. Ashwin Paranjape, Robert West, Leila Zia, and Jure Leskovec. 2016. Improving Website Hyperlink Structure Using Server Logs. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16)*. ACM, New York, NY, USA, 615–624. DOI: http://dx.doi.org/10.1145/2835776.2835832

25. Katy E. Pearce and Ronald E. Rice. 2013. Digital Divides From Access to Activities: Comparing Mobile and Personal Computer Internet Users. *Journal of Communication* 63, 4 (Aug. 2013), 721–744. DOI: http://dx.doi.org/10.1111/jcom.12045

26. Thomas B. Pepinsky. 2018. Visual Heuristics for Marginal Effects Plots. *Research & Politics* 5, 1 (Jan. 2018), 2053168018756668. DOI: http://dx.doi.org/10.1177/2053168018756668

27. Derek L. Phillips and Kevin J. Clancy. 1972. Some Effects of "Social Desirability" in Survey Studies. *Amer. J. Sociology* 77, 5 (March 1972), 921–940. DOI: http://dx.doi.org/10.1086/225231

28. Jennifer Preece and Ben Shneiderman. 2009. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Transactions on Human-Computer Interaction* 1, 1 (2009), 13–32.

29. Reid Priedhorsky, Dave Osthus, Ashlynn R. Daughton, Kelly R. Moran, Nicholas Generous, Geoffrey Fairchild, Alina Deshpande, and Sara Y. Del Valle. 2017. Measuring Global Disease with Wikipedia: Success, Failure, and a Research Agenda. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 1812–1834. DOI: http://dx.doi.org/10.1145/2998181.2998183

30. Joseph Reagle and Lauren Rhue. 2011. Gender Bias in Wikipedia and Britannica. *International Journal of Communication* 5, 0 (Aug. 2011), 21. https://ijoc.org/index.php/ijoc/article/view/777

31. Aaron Shaw and Eszter Hargittai. 2018. The Pipeline of Online Participation Inequalities: The Case of Wikipedia

Editing. *Journal of Communication* 68, 1 (Feb. 2018), 143–168. DOI:http://dx.doi.org/10.1093/joc/jqx003

32. Philipp Singer, Florian Lemmerich, Robert West, Leila Zia, Ellery Wulczyn, Markus Strohmaier, and Jure Leskovec. 2017. Why We Read Wikipedia. In *Proceedings of the 26th International Conference on World Wide Web - WWW '17*. 1591–1600. DOI: http://dx.doi.org/10.1145/3038912.3052716

33. Michael P. H. Stumpf and Mason A. Porter. 2012. Critical Truths About Power Laws. *Science* 335, 6069 (Feb. 2012), 665–666. DOI: http://dx.doi.org/10.1126/science.1216142

34. Alexander J. A. M. van Deursen and Jan A. G. M. van Dijk. 2015. Toward a Multifaceted Model of Internet Access for Understanding Digital Divides: An Empirical Investigation. *The Information Society* 31, 5 (Oct. 2015), 379–391. DOI: http://dx.doi.org/10.1080/01972243.2015.1069770

35. Morten Warncke-Wang, Vivek Ranjan, Loren Terveen, and Brent Hecht. 2015. Misalignment Between Supply and Demand of Quality Content in Peer Production Communities. In *Ninth International AAAI Conference on Web and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10591

36. Xing Yi, Liangjie Hong, Erheng Zhong, Nanthan Nan Liu, and Suju Rajan. 2014. Beyond Clicks: Dwell Time for Personalization. In *Proceedings of the 8th ACM Conference on Recommender Systems (RecSys '14)*. ACM, New York, NY, USA, 113–120. DOI: http://dx.doi.org/10.1145/2645710.2645724

37. Peifeng Yin, Ping Luo, Wang-Chien Lee, and Min Wang. 2013. Silence Is Also Evidence: Interpreting Dwell Time for Recommendation from Psychological Perspective. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*. ACM, New York, NY, USA, 989–997. DOI: http://dx.doi.org/10.1145/2487575.2487663

Table 4. Regression tables for models 1a and 1b.

	Model 1a	Model 1b
Intercept	1.3660 (0.0085)***	1.3791 (0.0085)***
Global North		−0.2680 (0.0022)***
mobile : Global North		0.1490 (0.0024)***
mobile : Last in Session	−0.6332 (0.0021)***	−0.6349 (0.0021)***
Global North : Last in Session		0.0830 (0.0024)***
Human development index	−0.1961 (0.0018)***	
mobile : HDI	0.1133 (0.0019)***	
HDI : Last in Session	0.0632 (0.0019)***	
Revision length (bytes)	0.1752 (0.0004)***	0.1758 (0.0004)***
time to first paint	−0.0164 (0.0006)***	−0.0171 (0.0006)***
time to dom interactive	0.0025 (0.0009)**	0.0024 (0.0009)**
mobilemobile	−0.0118 (0.0023)***	−0.0142 (0.0023)***
sessionlength	−0.0001 (0.0000)***	−0.0001 (0.0000)***
Last in session	0.8632 (0.0023)***	0.8575 (0.0023)***
nthinsession	0.0002 (0.0000)***	0.0002 (0.0000)***
dayofweekMon	0.0939 (0.0020)***	0.0926 (0.0020)***
dayofweekSat	0.0169 (0.0020)***	0.0175 (0.0020)***
dayofweekSun	0.0322 (0.0020)***	0.0332 (0.0020)***
dayofweekThu	0.0561 (0.0019)***	0.0548 (0.0019)***
dayofweekTue	0.0349 (0.0020)***	0.0326 (0.0020)***
dayofweekWed	0.0757 (0.0019)***	0.0743 (0.0019)***
usermonth4	0.0095 (0.0096)	0.0083 (0.0096)
usermonth5	0.0108 (0.0095)	0.0104 (0.0095)
usermonth6	−0.0102 (0.0097)	−0.0103 (0.0097)
usermonth7	−0.0494 (0.0097)***	−0.0491 (0.0097)***
usermonth8	−0.0119 (0.0097)	−0.0121 (0.0097)
usermonth9	0.0382 (0.0076)***	0.0370 (0.0076)***
usermonth10	−0.0004 (0.0075)	0.0010 (0.0075)
R^2	0.0721	0.0725
Adj. R^2	0.0720	0.0725
Num. obs.	9873641	9873641
RMSE	14.2330	14.2297

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$