Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Rare and famous: Understanding vascular plant awareness globally through Wikipedia page views

Arka Pal

palarka@fld.czu.cz

Czech University of Life Sciences Prague

Matilda J M Brown

Royal Botanic Gardens, Kew

Steven P Bachman

Royal Botanic Gardens, Kew

Article

Keywords:

Posted Date: July 15th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10279865/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Increased public awareness is crucial to mitigate the ongoing loss of biodiversity worldwide, and conservation culturomics offers a way to quantify this awareness and its implications for conservation. Large online data repositories provide an opportunity to measure the awareness, even for groups such as plants that receive less attention in global conservation analyses. We used Wikipedia page views of vascular plants across ten language editions (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian, and Spanish) as a proxy for public awareness and quantified the differences across extinction risk, documented human use, and geographic range size. We found that overall, threatened species tend to have fewer page views, contrary to expectations. However, when we accounted for the covariates of range size and human use, this effect is reversed, with threatened species having more page views than non-threatened species. Unsurprisingly, useful species with larger range sizes (mostly non-threatened and Data Deficient) tend to receive more page views than those without documented use, suggesting that people are inclined to search online for plants that provide provisioning services. Our results strongly suggest that there is heightened public awareness of threatened plants in their local-language on Wikipedia. We present the first global study of vascular plants using Wikipedia page views as a proxy for public awareness, through conservation culturomics to understand human-plant relationships.

Introduction

The social dimensions of biodiversity conservation, such as public awareness of biodiversity, are increasingly being recognised as crucial for addressing biodiversity loss and tackling sustainability challenges[1,2]. Conservation culturomics has emerged as an approach to support biodiversity conservation strategies by gaining insights into human-nature interactions, usually derived by analysing some of the world's major digital corpora, such as Facebook[3], X (formerly Twitter)[4–6], YouTube[7,8], Flickr[9], Google Trends[10–15] and Wikipedia[16–22]. With approximately 67% of the world's population connected to the internet[23], the way in which human populations access biodiversity information through these digital corpora can be used to generate metrics such as biodiversity 'awareness'[21]. However, these approaches have yet to be fully harnessed for certain components of biodiversity, such as plants, fungi and invertebrates.

We focus here on Wikipedia, an open-access online encyclopaedia that has 63 million articles in 310 languages and approximately two billion unique monthly visits[18,24]. The scale and geographic scope of Wikipedia make it an appropriate proxy for exploring public awareness of biodiversity. While the analysis of Wikipedia page views [total number of visits to an article page in the given time period[25]] has been widely used to investigate awareness in animals, like reptiles[16], birds[17,18], and reef fishes[22], limited attempts have been made to apply this approach to understand awareness of plants[26], despite 45.1% of plant species predicted to be threatened with extinction[27], and 27,139 plants already assessed as threatened worldwide[28]. This lack of research attention could be attributed to Plant Awareness Disparity

(PAD)[26,29], which refers to the human tendency to overlook and undervalue plants in the surroundings, leading to an inability to recognise the importance of plants for biodiversity conservation[30]. Nonetheless, it is important to address this gap and to explore different dimensions of plant awareness, such as geography/language, extinction risk and plant utility.

We expected threatened[31–33] and useful plant species[34,35,26] to garner more page views, which we equate to higher public awareness. We also expected that geographic range size would influence page view patterns, with wider-ranged species assumed to be more common or 'popular' species[22,35]. It has also been shown that different Wikipedia language editions may influence patterns[16], with the expectation that page views are highest with dominant 'local' Wikipedia language editions that overlap a species range. We therefore harvested page views on Wikipedia's top ten language editions (Arabic, Chinese, English, French, German, Italian, Japanese, Portuguese, Russian and Spanish)[25], range sizes for all [59,435] vascular plants containing non-zero Wikipedia page views in the form of the number of botanical countries from the World Checklist of Vascular Plants[36], and plant use from the World Checklist of Useful Plants[20]. Threat status was obtained from the IUCN Red List of Threatened Species[28], which, although not comprehensive for plants, includes a large sample of 59,435 species, approximately 20% of all described plants. We utilised a standardised culturomics workflow (Supplementary Figure **S1**) to explain differences in plant awareness based on extinction risk status, documented use, geographic range distribution and local-language Wikipedia.

We first examined the difference in page views between threatened (Vulnerable, Endangered, Critically Endangered), Data Deficient, and non-threatened (Least Concern and Near Threatened) species, without considering covariates, i.e., the general pattern. We then modelled this relationship using fixed-effect generalised linear models to assess the influence of covariates on species page views, comparing threatened and non-threatened plants while controlling for covariates (see **Methods**). In the main text, we use the Spanish language edition to visualise our findings; this is the language edition with the second-highest page views overall (after English), but a stronger link to the geographic distribution of Spanish[25]; results and corresponding figures for all language editions are presented in the Supplementary Information.

Results

Non-threatened species had the highest median page views (Spanish – 3037; 95% CI [260, 401714.50]), followed by Data Deficient species (Spanish – 1409; 95% CI [16, 793033.55]). Threatened plants had the lowest median page views (Spanish – 1091; 95% CI [53, 66377]). Most language editions (8) exhibited a pattern in which threatened plants had the lowest median page views, followed by Data Deficient species, and non-threatened species had the highest median page views (Fig. 1; Spanish; all selected languages, Supplementary Figure **S3**). Exceptions were English and Portuguese, where Data Deficient page views were lowest.

Model results (after accounting for covariates) revealed that threatened species consistently had higher page views than non-threatened species (Fig. 2), and that this difference increased with range size (except for the Japanese edition (Supplementary Figure **S2[vii]**, Supplementary Table **S1**). Generally, we observed that views were highest when the species had a documented use and when the local-language Wikipedia overlapped with its range distribution (e.g., for Spanish: smaller range, 0.7 ± 0.26, $p < 0.01$; larger range, 0.84 ± 0.33, $p < 0.01$). Contrastingly, species without documented human use and no range overlap had the lowest page views on the respective language Wikipedia (Spanish: smaller range, 0.59 ± 0.3, $p > 0.05$; larger range, 0.66 ± 0.34, $p > 0.05$), and the difference between threatened and non-threatened plants became indiscernible (Fig. 2). Best-fit lines through the scatterplots reveal the pattern across language editions. In all combinations, the page views increased with increasing range size (Supplementary Figure **S2[i-ix]**).

To better understand differences in extinction risk across language editions, we plotted the estimated marginal means of the coefficients for the pairwise contrast between threatened and non-threatened species, treating them as fixed effects and including the covariates, as shown in Fig. 3. Results clearly showed that, across languages, model coefficients for threatened species page views were significantly higher than those for non-threatened species, except in Italian, Japanese, and Russian Wikipedia. Although not significant, trends for these languages were generally higher for threatened plants. Most language editions exhibited a positive correlation between the coefficients and the range size; however, the French, German, Italian and Russian editions showed a negative correlation. The model summary for all languages is provided in Supplementary **Table S1**.

The contrast in Italian, Japanese and Russian editions was not significant across any covariate combination. In Portuguese, when no overlap of species range and local-language occurs, non-threatened species have higher page views than threatened species at smaller ranges (no documented use, -0.54 ± 0.45, $p > 0.05$; documented use, -0.31 ± 0.3, $p > 0.05$), but the contrast reverses at larger ranges (no documented use, 1.16 ± 0.59, $p > 0.05$; documented use, 1.4 ± 0.38, $p < 0.001$). This trend is not observed when the range and local-language overlap. The scenario reverses for Russian edition, where we observe higher page views of threatened plants at smaller ranges [no overlap - (no documented use, 0.5 ± 0.43, $p > 0.05$; documented use, 0.47 ± 0.42, $p > 0.05$) | overlap - (no documented use, 0.52 ± 0.48, $p > 0.05$; documented use, 0.49 ± 0.47, $p > 0.05$)], but is higher for non-threatened plants at larger ranges [no overlap - (no documented use, -0.18 ± 0.56, $p > 0.05$; documented use, -0.21 ± 0.49, $p > 0.05$) | overlap - (no documented use, -0.16 ± 0.61, $p > 0.05$; documented use, -0.19 ± 0.54, $p > 0.05$)]. It is also similar in the German edition at smaller ranges [no overlap - (no documented use, 0.61 ± 0.42, $p > 0.05$; documented use, 0.75 ± 0.35, $p < 0.05$) | overlap - (no documented use, 0.37 ± 0.87, $p > 0.05$; documented use, 0.51 ± 0.85, $p > 0.05$)] and larger ranges [no overlap - (no documented use, -0.04 ± 0.6, $p > 0.05$; documented use, 0.1 ± 0.38, $p > 0.05$) | overlap - (no documented use, -0.28 ± 0.9, $p > 0.05$; documented use, -0.14 ± 0.78, $p > 0.05$)]. However, the Japanese edition showed the highest anomaly: as range size increased, page views for non-threatened species decreased, while they increased for threatened species. Here, we see that when species range does not overlap with local-language, non-threatened plants page views are higher for smaller ranges (no documented use, -0.73 ± 0.57, $p > 0.05$;

documented use, -0.73 ± 0.51, $p > 0.05$), whereas threatened page views are higher at larger ranges (no documented use, 0.8 ± 0.58, $p > 0.05$; documented use, 0.8 ± 0.47, $p > 0.05$). It is very similar when the species range and local-language overlap - the non-threatened species page views are higher than threatened ones at smaller ranges (no documented use, -1.15 ± 0.9, $p > 0.05$; documented use, -1.14 ± 0.9, $p > 0.05$) and higher for threatened species at larger ranges (no documented use, 0.39 ± 1.13, $p > 0.05$; documented use, 0.39 ± 1.11, $p > 0.05$).

Discussion

Our analysis suggests public awareness of plant species, as measured by Wikipedia page views, is highest for threatened species in local-language editions of Wikipedia. People were interested in searching for local plants with high extinction risk in their native language, irrespective of the species' documented uses or the geographic extent of their range. Our study corroborates previous findings that, once covariates are accounted for, threatened plant species receive higher public attention[31–33]. Our study provides a novel insight into the public perception of plant awareness and its implications for plant conservation.

Extinction risk, geographic range and plant use

We expected that threatened plant species would have higher public awareness and, consequently, higher page views. We found that threatened species had the lowest median page views among other extinction risk categories when plotted against page views (Supplementary Figure **S3**). Additionally, threatened species, with or without documented use, consistently had the lowest median page views, while non-threatened and Data Deficient species recorded as useful had the highest Wikipedia page views (Supplementary Figure **S4**), which clearly illustrates people's interest in searching for beneficial plants, consistent with a previous study showing that plants useful to people stimulate more searches[35]. Mouquet *et al*[22] also found that species with wider distributions have higher page views. We find that for non-threatened and Data Deficient plants, range size was positively associated with page views because these species are mostly used by humans for food. Contrastingly, threatened plants have a more restricted range size compared to non-threatened and Data Deficient ones, as most threatened plants are assessed based on small range size[37]. However, the smaller-ranged threatened species (< 10 botanical countries) had high page views, which was anomalous compared to the non-threatened or Data Deficient species, where page views increased with increasing range size (Supplementary Figure **S5**). The pattern reveals a juxtaposition of people's interest in plants with their extinction risk and the useful, widespread plants.

Modelled page view patterns

We tested our hypothesis by analysing contrasts in page views between threatened and non-threatened species, using the other covariates as fixed effects in Poisson GLMs to explore the impact globally, with regional changes evident through variations in language editions. We find that although threatened plant

species tend to have lower page views than non-threatened species, this can be attributed to correlated differences in range size, documented use and the dominant, locally-used Wikipedia language. After accounting for these factors, we found that overall, threatened species tend to have similar or more (but never fewer) page views than non-threatened species, with exceptions in a few languages (Fig. 3). Overall, the models suggest that people tend to search for plants spanning a larger geographic area, as indicated by the increase in page views with increasing range size (Fig. 2). They also mostly search for plants that are beneficial and can be observed around them, and look them up in the local-language edition of Wikipedia.

The anomalous trends in Japanese Wikipedia (Supplementary Figure **S2[vii]**) can be argued to result from the interest of specific groups of people rather than a wider general audience[38]. Additionally, no impact of range size on non-threatened species page views may be attributed to the country's geographic location – Japan is an island country, with most searches occurring from within[25]. The general public's attention would most likely be focused on local non-threatened plants, leading to a weaker association with those that are more widely distributed. Conversely, threatened endemic plants have comparatively lower page views, which increase with range size, indicating people's interest in globally threatened plants. The anomaly can also be attributed to the insufficient species data in the Japanese Wikipedia for adequate analysis (Supplementary Table **S2**). Similarly, the Portuguese edition of Wikipedia (Supplementary Figure **S2[viii]**), which receives the highest traffic from Brazil[25]—one of the most biodiverse countries on Earth—shows higher page views for non-threatened plants with smaller ranges, likely reflecting greater public interest in endemic species that are not necessarily at risk of extinction[39]. This trend is less pronounced than in the Japanese edition and varies with increasing range size: page views for threatened species increase significantly, mostly for plants with documented use (Fig. 3). The German, Italian and Russian editions, where we observe a trend of decreasing contrast between the threatened and non-threatened species page views with increasing range size, can be attributed to the restricted geographic expanse of the local-language used. The use of these editions is confined to a small number of countries in a particular geographic region, leading people to search less for local plants with extinction risk. In contrast, they tend to search for widespread, non-threatened, and useful plants, thereby reducing the contrast between threatened and non-threatened plants. Contrastingly, language editions used across large world regions, such as English, French and Spanish, show a distinct pattern - threatened plants consistently have higher page views than non-threatened plants. Hence, we clearly observe a phytogeographical effect of the language edition influencing the page views of vascular plants – local-languages used across different phytogeographic regions show more differences between threatened and non-threatened plants than those present within one phytogeographic region[40].

People's perspective on plant awareness

Only a few previous studies focused on the significance of plants as indicators to measure public awareness of biodiversity[34,35,26]. Our study uncovers marked human interest biases in plants, based on

their extinction risk, documented use, geographic range size, and the use of local-language Wikipedia, which is evident from the fact that the mean page views difference between threatened and non-threatened plants decreased as the covariate combinations turned from positive to negative (Fig. 3).

Though the results present novel findings, they have limitations. Wikipedia data does not represent all demographics, as only 67% of the global population has internet access[23]—which excludes indigenous tribes, forest communities, and the poorest groups in low- and middle-income countries—and gender disparities exacerbate the data bias in these regions, since women often possess critical knowledge about plants. Even in well-connected nations, a divide between urban and rural areas reduces the impact of cultural interests on page views, which often underrepresent complex and multifaceted aspects, such as the cultural significance of plants[41,42,22,38,43]. Biases can also occur in countries where Wikipedia is not the dominant platform. For example, a Chinese language encyclopaedia (Baidu Baike) is more commonly used in China compared to Wikipedia. However, user location and engagement data are not easily accessible, hindering understanding of search motivations[21]. Language editions serve as proxies for geographical variation, but can overlook key stakeholders or misinterpret search motives, weakening signals of public awareness[44]. Consequently, only a broad overview of public attention based on Wikipedia page views is reliably achievable, given these unavoidable biases.

Our data captures the absolute level of public awareness of plants, collected as one-time information over the study period. While we primarily analysed absolute page views, there are likely temporal variations that could offer deeper insights into changes in viewership over time, especially during or after significant events such as flagship events or nature documentaries[45,35,19,17]. Fluctuations in page views for specific species can indicate increased awareness or a heightened need for conservation efforts, helping us identify neglected or priority species and their causal factors[31,21,16]. Beyond page views, metadata such as edits and references from crowd-sourced platforms like Wikipedia can serve as valuable indicators for large-scale digital monitoring of biodiversity awareness, contributing to research in conservation culturomics[44,43,46]. When combined with other digital resources such as the Global Biodiversity Information Facility (GBIF)[47] and Plants of the World Online (POWO)[48], which provide habitat and observation data, our findings can shed more light on human interest in threatened species.

Conclusion

We present the first comprehensive study of people-plant awareness, using Wikipedia page views from the top ten language editions as a measure of interest in plants. Our novel approach examines the effects of extinction risk, documented human use, range size (the number of botanical countries), and local-languages on Wikipedia on public awareness of threatened plants. Globally, we find that people search for plants that are useful to them and are spread over large geographic regions. On a finer scale, our empirical data suggests that people are more interested in plants with extinction risk, which are local to their region. Plant awareness has steadily risen among younger demographics over the past decade, as a consequence of enhanced dissemination of science communication to the public through social

media platforms, which grew further after the COVID-19 pandemic[35]. Further enhancement of botanical interest in humans, complemented by Red List-related media outreach, can persuade governments and non-governmental organisations (NGOs) to augment funding allocation to plants. This is because popular attitude is a key factor in species prioritisation and financial investments[33]. Our data can provide a benchmark for assessing plant popularity, serving as a key tool for conservationists to advocate for increased funding for the conservation of species and their habitats. A shift in public perception and awareness towards plants can be instrumental in their conservation efforts, persuading governments and non-governmental entities to increase resource allocation[17,33,49]. This is a vital component in mitigating global biodiversity loss.

Conservation culturomics is increasingly used to gain insights into human-nature interactions, with Wikipedia serving as a powerful tool despite its limitations. It can play a vital role in achieving the targets set by the Global Biodiversity Framework (GBF) and Sustainable Development Goals (SDG) by raising public awareness and supporting efforts to halt biodiversity loss. Overall, our study provides valuable baseline information on global human-plant interactions, laying the groundwork for future research on plant conservation.

Methods

Extinction risk, geographic range and plant use

Extinction risk status was obtained from The IUCN Red List of Threatened Species (hereafter the Red List) *version 2025 – 2*[28]. The Red List was queried for all vascular plant species (Phylum: *Tracheophyta*) on a global geographical scope, excluding species categorised as Extinct (EX) and Extinct in the Wild (EW), and matched species names from this vascular plant Red List to the World Checklist of Vascular Plants (WCVP) *version 14 (2025-1)*[50,51] using exact and fuzzy matching functions from the rWCVP package[51], leaving us with 62,458 matched and accepted species. We grouped extinction risk into three categories: threatened (Critically Endangered, Endangered and Vulnerable), non-threatened (Least Concern and Near Threatened), and Data Deficient[37]. Geographic ranges were obtained from the WCVP dataset using Level 3 of the World Geographical Scheme for Recording Plant Distribution (WGSRPD)[52], often referred to as 'botanical countries'[36]. We excluded species with ranges coded as extinct or doubtful. Plant use data was obtained from the World Checklist of Useful Plants (WCUP)[53] and appended to the WCVP dataset to provide a use score, where 1 indicates a species documented as useful according to WCUP, and 0 indicates a species was not listed and therefore coded as non-useful.

Wikipedia page views

To obtain page views per species, we first made calls to the Wikidata API using the binomial names in our dataset, which returned Wiki Q identifiers (QID)[54]. We applied this approach to the top ten language editions on Wikipedia by search traffic volume, which were English, Spanish, German, Russian, French,

Italian, Chinese, Portuguese, Arabic and Japanese (in descending order)[25]. A total of 571 species had no QID and were therefore excluded. The QIDs generated for 62,458 species were used to retrieve their article titles from each Wikipedia language. Article titles were then used to download human-generated (excluding bot-generated: automated views using computer-controlled accounts) monthly page views from 1st July 2015 (earliest available archived page views) to 30th June 2025, and these were summed to obtain the total page views. A total of 3,023 species that had Wikipedia pages without any page views in any one language edition were excluded from analysis to avoid the zero-inflation factor, resulting in 59,435 species.

The 'local-language' Wikipedia version, that is, the dominant Wikipedia language version used in a region, was determined using Wikidata analytics[25] that contains page views by Wikipedia language edition across countries. For each 'botanical country', we coded each Wikipedia language edition as 1 when there were at least 10% of all searches in one of the languages used for the study and 0 when there were less than 10% of all searches. We used R packages 'httr', 'jsonlite', 'sf' and 'rnaturalearth' to query and visualise data from Wikidata and Wikipedia[55–58]

Modelled page view patterns

We fitted Poisson generalised linear models (GLMs) to test the response of species page views to the predictors for every language, with and without interaction among the predictors (extinction risk, documented use, range size and local-language Wikipedia). Subsequently, we performed tests for overdispersion of the models using the R package 'DHARMa'[59] and re-fitted the models using quasi-Poisson GLM. We used the Quasi-Akaike Information Criterion (QAIC) for model selection and selected the best-fitting models to explain pageview variations using covariate predictors[60,61]. The quasi-Poisson model was tested for seven terms: the covariates individually, and the interactions between extinction risk and use, extinction risk and range size, and extinction risk and local-language Wikipedia, using the 'MuMIn' package[62] to test the effect of the threatened and non-threatened contrast across the covariates. We selected the best models with delta QAIC values < 6[60,61] to eliminate nested models and clarify variation across language editions, and we averaged these models with model uncertainty. Thereafter, we calculated the least-squares means of the models utilising the 'emmeans' package[63] after model weighting and Tukey adjustment, to compare the pairwise comparisons between threatened and non-threatened species page views across all covariates and language editions. A complete workflow visualisation is provided in Supplementary **Figure S6.**

We carried out the study using R programming software *version 4.5.0* [64]. **Table S3** provides a comprehensive list of all the packages used for data handling and statistical analyses. All scripts are made available at https://github.com/Arkapal7/Plant-awareness.

Declarations

Funding

AP was funded by the Internal Grant Agency (IGA) of the Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague (Grant number - 43970/1312/3146). MJMB and SPB were funded by the Kew Research Fellowship.

Competing interests

The authors declare no competing interests.

Author Contribution

S.P.B. and M.J.M.B. conceived the study, with contributions from A.P. A.P. carried out data extraction, processing and analyses, prepared visualisations and illustrations and led the initial writing of the manuscript; M.J.M.B. contributed to data processing and analyses and helped to draft the manuscript and critically revised it; S.P.B. helped to draft the manuscript and critically revised it. All authors gave final approval for the publication of the manuscript and agree to be held accountable for the work performed therein.

Acknowledgements

The authors would like to thank the Royal Botanic Gardens, Kew, for providing the WCVP and WCUP datasets, and Wikipedia and Wikimedia for granting access to Wiki QIDs, article titles, and monthly page views. A.P. would like to thank the Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague (43970/1312/3146) and the HIVE Project (43/2312/25) for funding and logistical assistance. A.P. would also like to thank Saptarshi Chakrabarti for his contributions to the initial support and code review, Melina Dominguez Olivares for her ideas and support during the study, Agnish Kumar Das for reviewing the draft of the manuscript, and Dr Andrew Liebhold for his constant help and encouragement.

Data Availability

The complete dataset used for analysis has been uploaded to figshare and freely available at the following URL: https://figshare.com/s/1d4ea9b5b54952f22863?file=64179835

Code availability

The complete R codes for Wikipedia data downloading, reconciliation with other datasets, processing, visualisation and analysis are available at: https://github.com/Arkapal7/Plant-awareness

References

1. Soriano-Redondo, A. et al. Harnessing online digital data in biodiversity monitoring. *PLoS Biol.* **22**, e3002497 (2024).
2. Adams, V. M. Social media data for biodiversity conservation. *Trends Ecology Evolution*. **39**, 16–18 (2024).
3. Chowdhury, S. et al. Insights from citizen science reveal priority areas for conserving biodiversity in Bangladesh. *One Earth*. **6**, 1315–1325 (2023).
4. Acerbi, A. et al. Sentiment analysis of the Twitter response to Netflix's Our Planet documentary. *Conserv. Biol.* **37**, e14060 (2023).
5. Evans, L. C., Greenwell, M. P., Boult, V. L. & Johnson, T. F. Characterizing the trophy hunting debate on Twitter. *Conserv. Biol.* **37**, e14070 (2023).
6. Ishida, S. et al. Twitter Mining for Detecting Interest Trends on Biodiversity: Messages from Seven Language Communities. *Sustainability* **15**, 12893 (2023).
7. Huertas Herrera, A. et al. Social media reveal visitors' interest in flora and fauna species of a forest region. *Ecosyst. People*. **19**, 2155248 (2023).
8. Magalhães, A. L. B., de Azevedo, C. S., Maceda-Veiga, A. & Patoka, J. The scientist eyes: monitoring YouTube™ to quantify aquatic pet release in Brazil. *Aquat. Ecol.* **58**, 73–88 (2024).
9. Havinga, I. et al. Social media and deep learning reveal specific cultural preferences for biodiversity. *People Nat.* **5**, 981–998 (2023).
10. Jarić, I. et al. Invasion Culturomics and iEcology. *Conserv. Biol.* **35**, 447–451 (2021).
11. Correia, R. A. & Di Minin, E. Tracking worldwide interest in sustainable development goals using culturomics. *PLOS Sustain. Transform*. **2**, e0000070 (2023).
12. De Oliveira Caetano, G. H. et al. Evaluating global interest in biodiversity and conservation. *Conserv. Biol.* **37**, e14100 (2023).
13. Felgentreff, E. S., Buchholz, S. & Straka, T. M. From science to society to practice? Public reactions to the insect crisis in Germany. *People Nat.* **5**, 660–667 (2023).
14. Moore, M. J. & Hyman, A. A. What can conservation culturomics tell us about factors driving public interest in aquatic endangered species. *Biol. Conserv.* **289**, 110397 (2024).
15. Yaman, B. & Van Huynh, A. Public attention towards declining global amphibian species. *Biol. Conserv.* **290**, 110472 (2024).
16. Roll, U. et al. Using Wikipedia page views to explore the cultural importance of global reptiles. *Biol. Conserv.* **204**, 42–50 (2016).
17. Mittermeier, J. C., Roll, U., Matthews, T. J. & Grenyer, R. A season for all things: Phenological imprints in Wikipedia usage and their relevance to conservation. *PLoS Biol.* **17**, e3000146 (2019).
18. Mittermeier, J. C., Roll, U., Matthews, T. J., Correia, R. & Grenyer, R. Birds that are more commonly encountered in the wild attract higher public interest online. *Conservat Sci. Prac.* **3**, e340 (2021).

19. Fernández-Bellon, D. & Kane, A. Natural history films raise species awareness—A big data approach. *Conserv. Lett.* **13**, e12678 (2020).
20. Pironon, S. et al. Toward Unifying Global Hotspots of Wild and Domesticated Biodiversity. *Plants* **9**, 1128 (2020).
21. Millard, J. W., Gregory, R. D., Jones, K. E. & Freeman, R. The species awareness index as a conservation culturomics metric for public biodiversity awareness. *Conserv. Biol.* **35**, 472–482 (2021).
22. Mouquet, N. et al. Low human interest for the most at-risk reef fishes worldwide. *Sci. Adv.* **10**, eadj9510 (2024).
23. International Telecommunication Union. Statistics. *ITU* (2024). https://www.itu.int:443/en/ITU-D/Statistics/Pages/stat/default.aspx
24. Wikipedia *Wikipedia: The Free Encyclopedia* (2024).
25. Wikipedia:Pageview statistics. Wikipedia. (2025).
26. Vardi, R., Mittermeier, J. C. & Roll, U. Combining culturomic sources to uncover trends in popularity and seasonal interest in plants. *Conserv. Biol.* **35**, 460–471 (2021).
27. Bachman, S. P., Brown, M. J. M., Leão, T. C. C., Lughadha, N., Walker, B. E. & E. & Extinction risk predictions for the world's flowering plants to support their conservation. *New Phytol.* **242**, 797–808 (2024).
28. IUCN. The IUCN Red List of Threatened Species. *IUCN Red List. Threatened Species* (2025). https://www.iucnredlist.org/en
29. Parsley, K. M. Plant awareness disparity: A case for renaming plant blindness. *PLANTS PEOPLE PLANET.* **2**, 598–601 (2020).
30. Schussler, E. E. & Olzak, L. A. It's not easy being green: student recall of plant and animal images. *J. Biol. Educ.* **42**, 112–119 (2008).
31. Van Huynh, A. Effect of IUCN Red List category on public attention to mammals. *Conserv. Biol.* **37**, e14050 (2023).
32. Betts, J. et al. A framework for evaluating the impact of the IUCN Red List of threatened species. *Conserv. Biol.* **34**, 632–643 (2020).
33. Davies, T. et al. Popular interest in vertebrates does not reflect extinction risk and is associated with bias in conservation investment. *PLOS ONE*. **13**, e0203694 (2018).
34. Kacprzyk, J., Clune, S., Clark, C. & Kane, A. Making a greener planet: nature documentaries promote plant awareness. *Ann. Botany*. **131**, 255–260 (2023).
35. Burke, R. et al. Botanical boom: A new opportunity to promote the public appreciation of botany. *Plants People Planet.* **4**, 326–334 (2022).
36. Brummitt, R. K., Pando, F., Hollis, S. & Brummitt, N. A. *World Geographical Scheme for Recording Plant Distributions* (Hunt Inst. for Botanical Documentation, 2001).

37. IUCN Standards and Petitions Committee. *Guidelines for Using the IUCN Red List Categories and Criteria*. 122 (2024). https://www.iucnredlist.org/documents/RedListGuidelines.pdf
38. Akasaka, M., Kubo, T. & Soga, M. How well does online information-seeking behavior indicate public conservation orientation? Taxonomy and personal characteristics matter. *J. Nat. Conserv.* **66**, 126123 (2022).
39. Brito-Júnior, V. M. & Albuquerque, U. P. Beyond Traditions: The Potential of Big Data in Assessing Interest in Medicinal Plants on the Internet. *Econ. Bot.* **79**, 65–78 (2025).
40. Carta, A., Peruzzi, L. & Ramírez-Barahona, S. A global phylogenetic regionalization of vascular plants reveals a deep split between Gondwanan and Laurasian biotas. *New Phytol.* **233**, 1494–1504 (2022).
41. Wang, J. & Lin, P. C. Spatial Spillover of the Global Internet Penetration Rate and the Digital Gender Divide. *Sage Open.* **14**, 21582440241255448 (2024).
42. George, D. A. S. Universal Internet Access: A Modern Human Right or a Path to Digital Colonialism. (2024). https://doi.org/10.5281/ZENODO.10970024 doi:10.5281/ZENODO.10970024.
43. Correia, R. A. et al. Digital data sources and methods for conservation culturomics. *Conserv. Biol.* **35**, 398–411 (2021).
44. Mittermeier, J. C., Correia, R., Grenyer, R., Toivonen, T. & Roll, U. Using Wikipedia to measure public interest in biodiversity and conservation. *Conserv. Biol.* **35**, 412–423 (2021).
45. Jarić, I., Crowley, S. L., Veríssimo, D. & Jeschke, J. M. Flagship events and biodiversity conservation. *Trends Ecology Evolution*. **39**, 106–108 (2024).
46. Jarić, I. et al. iEcology: Harnessing Large Online Resources to Generate Ecological Insights. *Trends Ecology Evolution*. **35**, 630–639 (2020).
47. GBIF. The Global Biodiversity Information Facility. What is GBIF? (2025). https://www.gbif.org/
48. POWO. Plants of the World Online. *Plants of the World Online* (2025). https://powo.science.kew.org/
49. Martín-Forés, I., Martín-López, B. & Montes, C. Anthropomorphic Factors Influencing Spanish Conservation Policies of Vertebrates. *International Journal of Biodiversity* 142670 (2013). (2013).
50. Govaerts, R., Nic Lughadha, E., Black, N., Turner, R. & Paton, A. The World Checklist of Vascular Plants, a continuously updated resource for exploring global plant diversity. *Sci. Data*. **8**, 215 (2021).
51. Brown, M. J. M. et al. r WCVP: a companion R package for the World Checklist of Vascular Plants. *New Phytol.* **240**, 1355–1365 (2023).
52. TDWG. World Geographical Scheme for Recording Plant Distributions (WGSRPD). Biodiversity Information Standards (TDWG). (2025).
53. Diazgranados, M. et al. World Checklist of Useful Plant Species. (2020).
54. Wikidata, W. Identifiers. (2025). https://www.wikidata.org/wiki/Wikidata:Identifiers
55. Ooms, J. The jsonlite Package: A Practical and Consistent Mapping Between JSON Data and R Objects. Preprint at (2014). https://doi.org/10.48550/arXiv.1403.2805
56. Wickham, H. httr: Tools for Working with URLs and HTTP version 1.4.7 from CRAN. (2023). https://rdrr.io/cran/httr/

57. Massicotte, P., South, A. & Hufkens, K. rnaturalearth: World Map Data from Natural Earth. (2023).
58. Pebesma, E. & Bivand, R. *Spatial Data Science: With Applications in R* (Chapman and Hall/CRC, 2023).
59. Hartig, F. DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. (2024).
60. Richards, S. A. Dealing with overdispersed count data in applied ecology - Richards – 2008 - Journal of Applied Ecology - Wiley Online Library. *J. Appl. Ecol.* 2664.2007.01377.x (2008). https://doi.org/doi/full/10.1111/j.1365-
61. Richards, S. A. Testing Ecological Theory Using the Information-Theoretic Approach: Examples and Cautionary Results. *Ecology* **86**, 2805–2814 (2005).
62. Bartoń, K. & MuMIn Multi-Model Inference. 1.47.5 (2010). https://doi.org/10.32614/CRAN.package.MuMIn
63. Lenth, R. V. & emmeans Estimated Marginal Means, aka Least-Squares Means. (2025).
64. R Core Team. R: The R Project for Statistical Computing. *R Foundation for Statistical Computing* (2023). https://www.r-project.org/

Figures

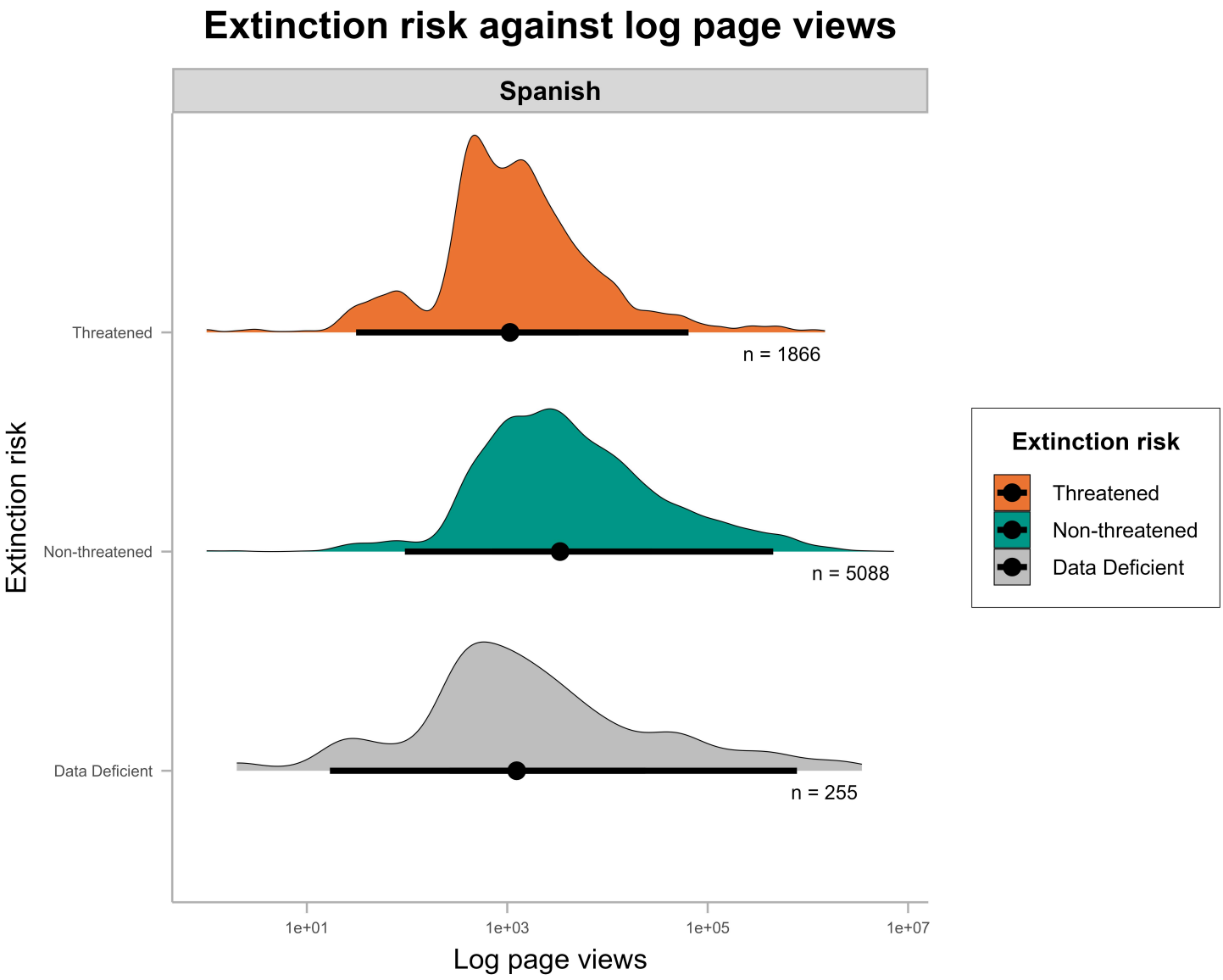


Figure 1

Wikipedia page views (Spanish language edition; $\log_{10}$ transformed) across extinction risk categories. The solid black dots denote the median page views in each category, and the black lines denote the 95% confidence interval. The distribution of page views density is shown in the coloured areas. The orange, green and grey plots show the page views distribution for threatened, non-threatened and Data Deficient species, respectively. 'n' denotes the total species count in each category.

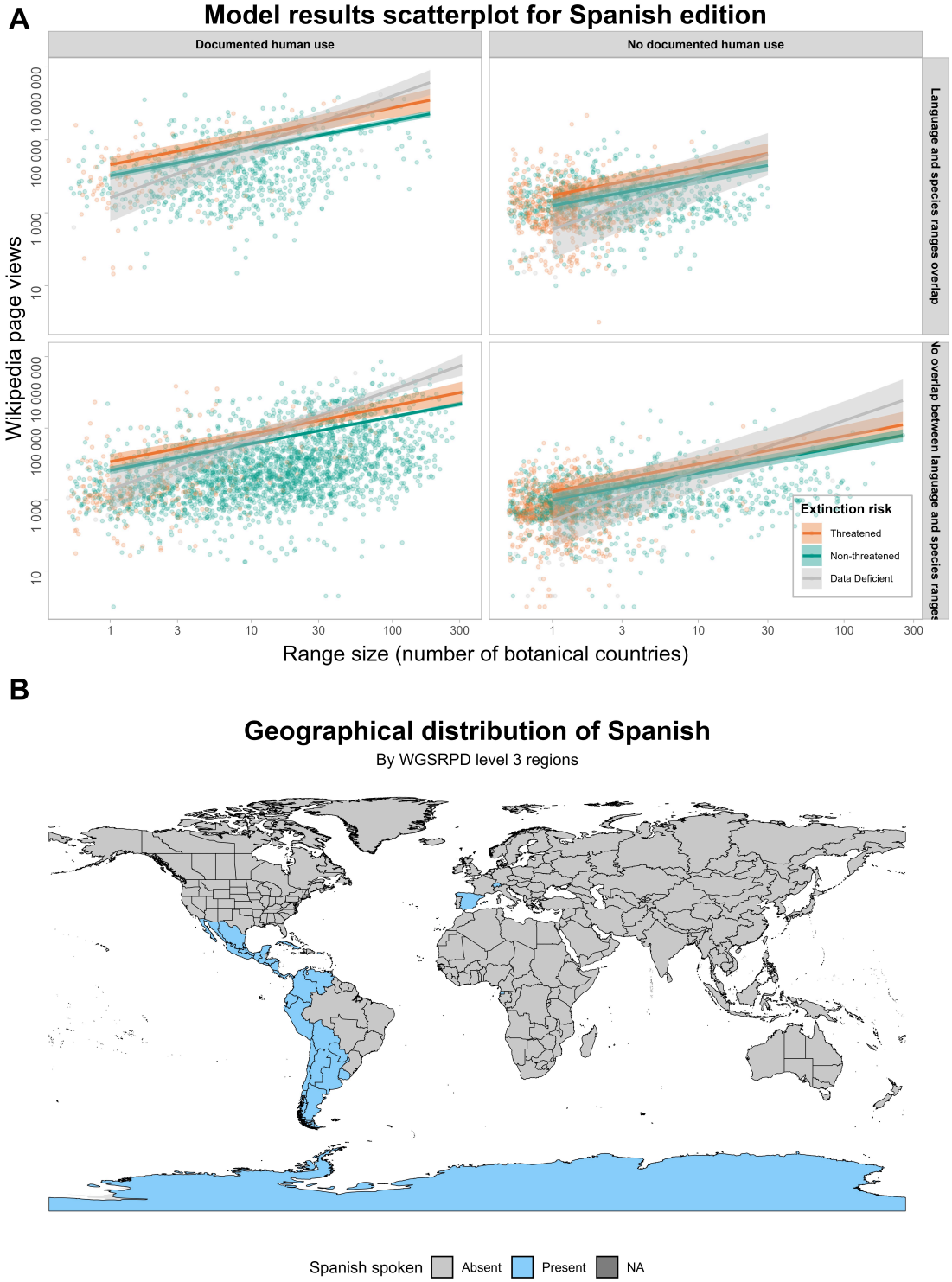


Figure 2

Effect of covariates on page views in the local-language Wikipedia. *(A) Page views across extinction risk, range size, documented use and geographic overlap; shown here for the Spanish language edition. The threatened species are shown in red-orange, Data Deficient species in grey, and non-threatened species in green. Shaded ribbons show the standard error. (B) The geographical distribution (botanical countries) of the Spanish language Wikipedia, where the Spanish edition of Wikipedia comprises >10% of total Wikipedia*

views. Blue represents the botanical countries where >10% of the Spanish Wikipedia is used. Grey indicates countries where <10% of Wikipedia views are from the Spanish language edition.

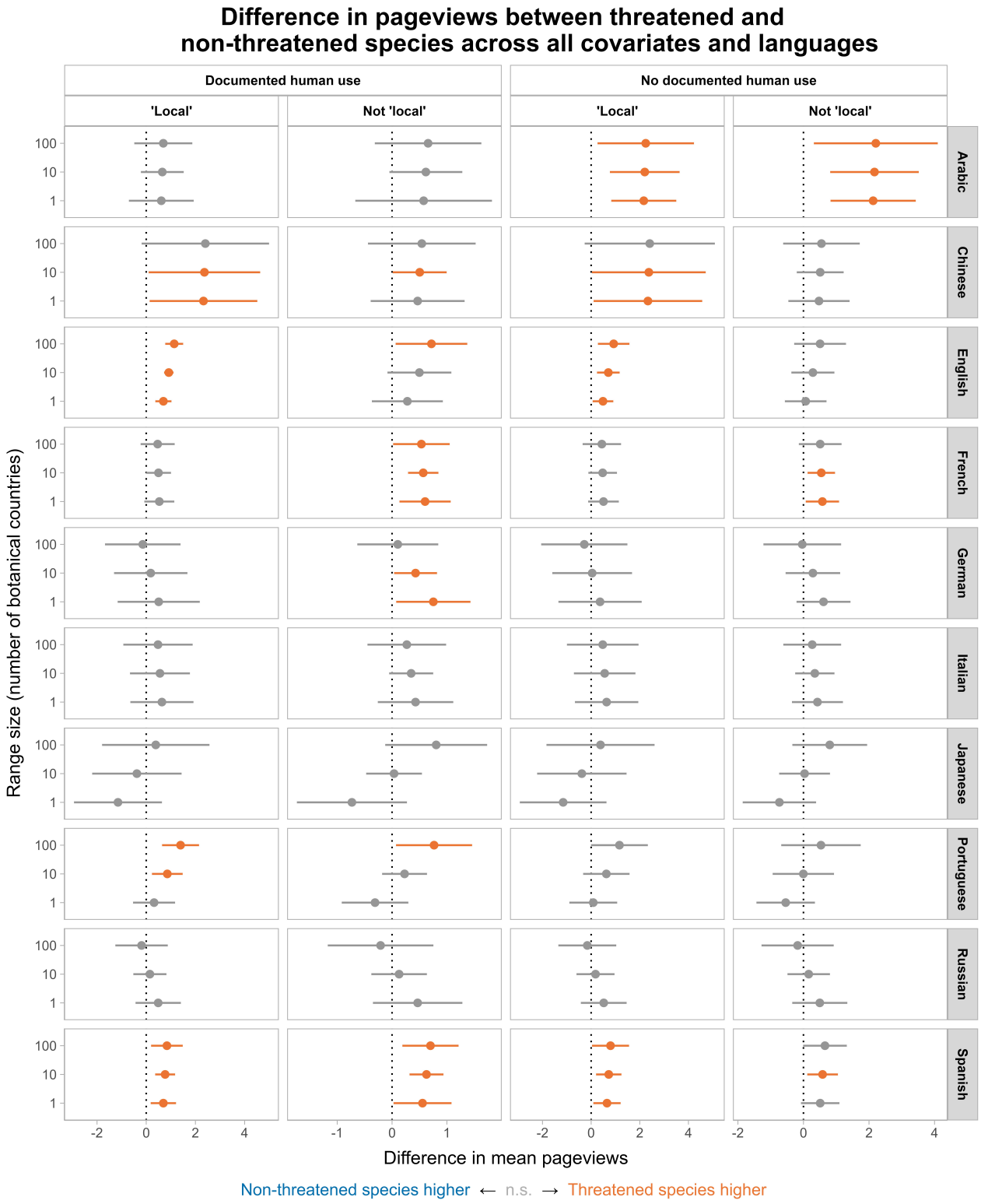


Figure 3

Pairwise contrasts of threatened and non-threatened plants across all languages. Coefficient plot showing the least-squares mean coefficient for all Wikipedia language editions across different range

sizes (botanical countries) in $\log_{10}$ scale. The orange-red dots represent the significantly positive coefficient values for threatened plants, and the orange-red lines define the 95% confidence intervals of the coefficients. The grey-coloured dots and lines indicate the coefficient values and standard error, respectively, for threatened plants that are not significantly different from those of non-threatened species. The vertical black-dashed lines indicate the zero values of the coefficients. ‘Local’ refers to the species whose ranges overlap with the language Wikipedia used. ‘Not local’ indicates no overlap between the range and geographical distribution of the language on Wikipedia.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PlantAwarenessSupplementaryTables.docx
- PlantAwarenessSupplementaryFigures.docx