



CONVOCO ! EDITION 2011

GEOGRAPHIES OF THE WORLD'S KNOWLEDGE

Dr. Corinne M. Flick and
the Convoco Foundation in cooperation with
Oxford Internet Institute / University of Oxford



Oxford Internet Institute
University of Oxford

Geographies of the World's Knowledge

Copyright © Oxford Internet Institute in cooperation with Dr. Corinne M. Flick and the Convoco Foundation 2011

This publication is released under the Creative Commons Attribution-NonCommercial-NoDerivs [CC BY-NC-ND] license. A summary of the license terms and the full legal text is available from <http://creativecommons.org/licenses/by-nc-nd/3.0/>



This publication should be cited as: Graham, M., Hale, S. A. and Stephens, M. (2011) Geographies of the World's Knowledge. Ed. Flick, C. M., London, Convoco! Edition.

The information in this publication has been published in good faith and every effort has been made to ensure its accuracy. The authors can accept no responsibility for any error or misinterpretation. All liability for loss, disappointment, negligence or damage caused by reliance on the information contained in this publications is hereby excluded to the fullest extent permitted by law.

Editor: Dr. Corinne M. Flick

Foreword: Dr. Corinne M. Flick

Authorship, visualisations and analysis by Mark Graham, Scott A. Hale, Monica Stephens and Viktor Mayer-Schönberger

Cover design: Camao AG, München

Booklet design: Kunika Kono

Printing and binding: Technique Print Group

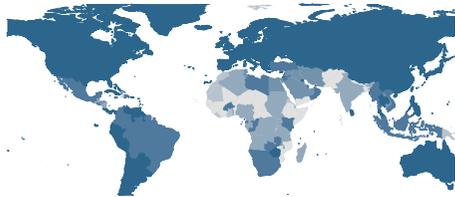
Printed in Great Britain

Data for "Time-series of the Distribution of Biographies on Wikipedia over the Last Five Centuries" provided by Adrian Popescu. Data for "User-generated Content in Google" provided by Matthew Zook.

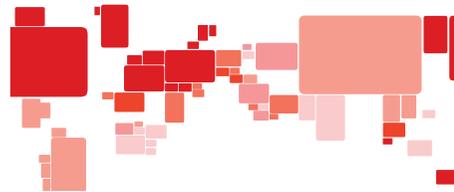
Contents

5 Foreword

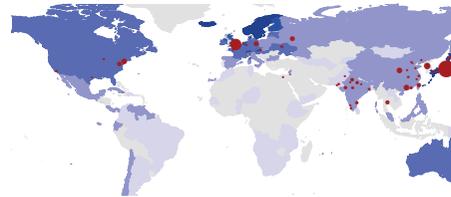
7 Introduction



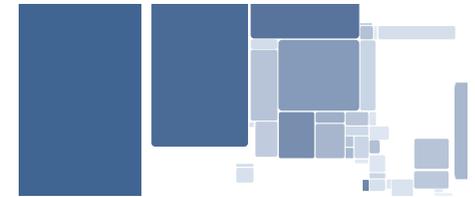
8 Literacy and Gender



10 Internet Penetration



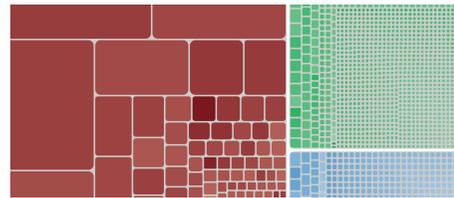
12 The World's Newspapers



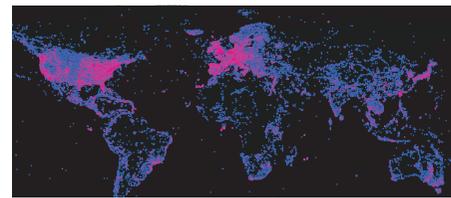
14 The Location of Academic Knowledge



16 Academic Knowledge and Language



18 Academic Knowledge and Publishers



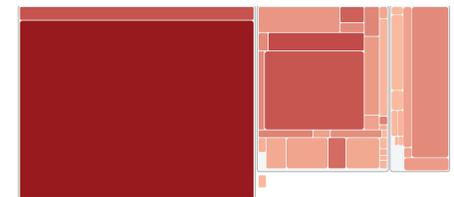
20 Mapping Flickr



22 The Distribution of all Wikipedia Articles



24 Time-series of the Distribution of Biographies on Wikipedia over the Last Five Centuries



26 User-generated Content in Google

28 Appendix

Foreword

As the early 20th-century German philosopher, Walter Benjamin, once remarked, “When media changes, society changes, too.”

What does this actually mean? Is the mere technical invention of new media sufficient to transform society - or do new media change the ways in which human beings communicate, irrespective of the content they transmit?

These are the questions Convoco will ponder in 2011 under the title of “Who owns the knowledge of the world?” — because communication is first and foremost about knowledge. However, one must be aware that knowledge is only relevant to the extent that it is conveyed and accessible to others. This has always been true, even before new media and modern communication tools became involved. Communication can take effect only by sharing knowledge, and knowledge is created by sharing it. Knowledge kept to oneself is irrelevant and without effect.

Convoco means “I bring together.” Its mission is an invitation to share ideas and thoughts through communication.

Access to knowledge is as important to the individual’s development as it is for the development of a society and its economic success as a whole. Knowledge has always meant power. At the same time it creates tolerance, encourages an understanding of the other and builds bridges. Knowledge creates identities.

Reflections about the ownership of the knowledge in the world are linked to the question “Who is responsible for knowledge?” This has to be taken into consideration, because knowledge can also be neglected or get lost, and it may be despised or banned. This is an extremely important point, since knowledge not only determines the future of humankind, but is also simultaneously a part of its cultural heritage.

In this respect one might for instance ask, “Who controls Wikipedia entries?” The good old times when anyone could intervene and share knowledge on this medium are long gone. Nowadays Wikipedia is based on a pyramidal structure of volunteers who manage the system. There is a clear hierarchy within this structure and control is exercised concerning what is accepted and what not. For the external author it has become almost impossible to add or modify an article with a lasting effect. The critical point here is that Wikipedia loses a considerable part of its legitimacy, which is, indeed, based on the concept of participation. This should give food for thought.

“Geographies of the World’s Knowledge”, which you hold in your hands today, is a joint venture between Convoco and the Oxford Internet Institute. It reflects upon the distribution of the world’s knowledge, and discusses ten key words, from literacy to user-generated content.

Data, evaluated in an unprecedented way, shows the current distribution of knowledge in the different parts of the globe. Some of the implications of this are surprising, others are worrying. The maps visualise where the foci of knowledge – and, thus, the forces of innovation and economic growth – are located. Thanks to this scientific visualisation the most important factors involved can be grasped at a glance. It also provides evidence of a paradigm shift towards a new visual language in the domain of sciences.

Proper evaluation must follow the analysis and presentation of these results, which imply questions like “Who should own the knowledge of the world?” and “For what purpose can one permit it to be used?” The answers to these questions will point the way ahead into the future.

Convoco aims to provide orientation. Orientation brings security. Einstein said that it does not matter, whether you understand the world; you must find your bearings in it. I hope that my initiative, Convoco, will make a contribution towards finding this orientation.

Dr. Corinne Michaela Flick, June 2011

Introduction

The sum of the world's knowledge is always expanding, doubling every few years. While remarkable, this prompts important questions: Where is this knowledge being created? Who has access to it? How is it being distributed?

And most importantly: How does our ability to access and produce codified knowledge change due to advances in information and communication technologies? American historian Elizabeth Eisenstein wrote about the advent of the printing press that a fifty-year-old person in the early 16th century would have seen more books being printed in her lifetime than in all of human history. The print revolution is not only a success story of the dramatic rise in the production of books. It happened in tandem with a revolution of access to knowledge in Europe. Prior to the invention of the printing press, the Catholic Church acted as an intermediary to knowledge as it controlled schools, scribes, and libraries. The instigators of the reformation, cognisant of the power of print, advocated for a more direct access of knowledge to people. Taken out of the hands of religious intermediaries, knowledge became more accessible – at least in Europe.

Despite these fundamental shifts, the world's codified knowledge remained concentrated among educated and affluent people in industrialized nations. This concentration spanned all phases in the cycle of information — from the production of knowledge to its distribution and access. Changing these patterns is inevitably hard, as access to knowledge is often a prerequisite to create new knowledge.

Unsurprisingly, the geographies of knowledge remain largely characterized by strong core-periphery patterns. For example approximately one-third of commercial innovation in the US is concentrated in two small regions: Silicon Valley in California, and the Greater Boston area on the East Coast. These two regions house the world's leading universities – established institutions to access, convey, and create knowledge.

The Internet raised hopes that knowledge might become more accessible, reaching less affluent groups, as well as those farther away from the producers and sources of information. Many commentators speculated that people outside of industrialised nations would be able to gain access to all networked and codified knowledge, thus mitigating the world's concentration of knowledge.

These early expectations remain largely unrealised. Knowledge institutions, including producers and distributors, around the world remain concentrated. The reason may partially lie in the economic power of large knowledge institutions, and the existing structure

of intellectual property rights. Together they prevent a more equitable access to (and equitable creation of) knowledge. As a result, many have once again pinned their hopes for change on technological innovation. A new generation of Internet tools have enabled not only global access to knowledge, but also the opportunity to establish new institutions for the sharing, peer-production and remixing of knowledge. The promise, in other words, is for a much more equitable production and distribution of the world's knowledge.

We approach this exciting and complex topic through ten visualisations – each one a snapshot highlighting a distinct dimension of the global distribution of knowledge. Visualisations empower us to quickly grasp complex dynamics and gain unexpected insights. We start with the preconditions to access knowledge in the digital age through literacy and Internet penetration. We continue by examining the established institutions of knowledge – from mass media to academic publishers, and discover concentrations of knowledge in different forms than might be commonly expected. Finally, we visualise peer-produced and user-generated knowledge – through new institutions of knowledge like Wikipedia that are founded on a need for broad participation. But even here we can observe significant concentrations of knowledge, leading to explosive questions.

We offer these visualisations as a contribution to the important debate on the production of, and access to, the world's knowledge. The information in this booklet has not yet been presented using the visualisations that we employ, and many of the datasets we use were collected specifically for this project. The result we hope is substantially more than the sum of its parts.

We ultimately hope to show that although information can now be produced almost anywhere, we should never assume that information is produced everywhere. Much of the world remains, both literally and figuratively, absent from the global map of knowledge.

Viktor Mayer-Schönberger and Mark Graham
Oxford Internet Institute
University of Oxford

Literacy and Gender

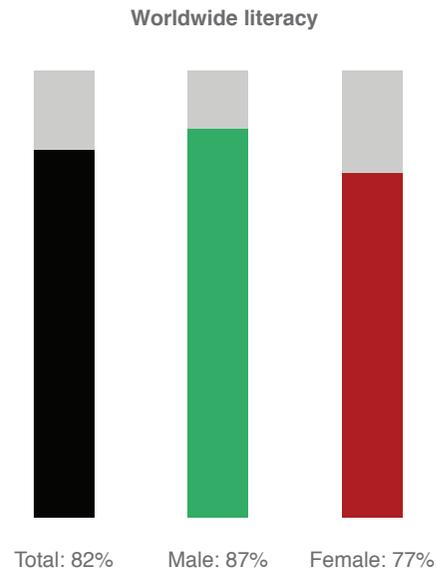
Literacy is an often overlooked factor in examining the flows, production and consumption of knowledge. This map visualises overall literacy rates and rates of literacy by gender around the world.

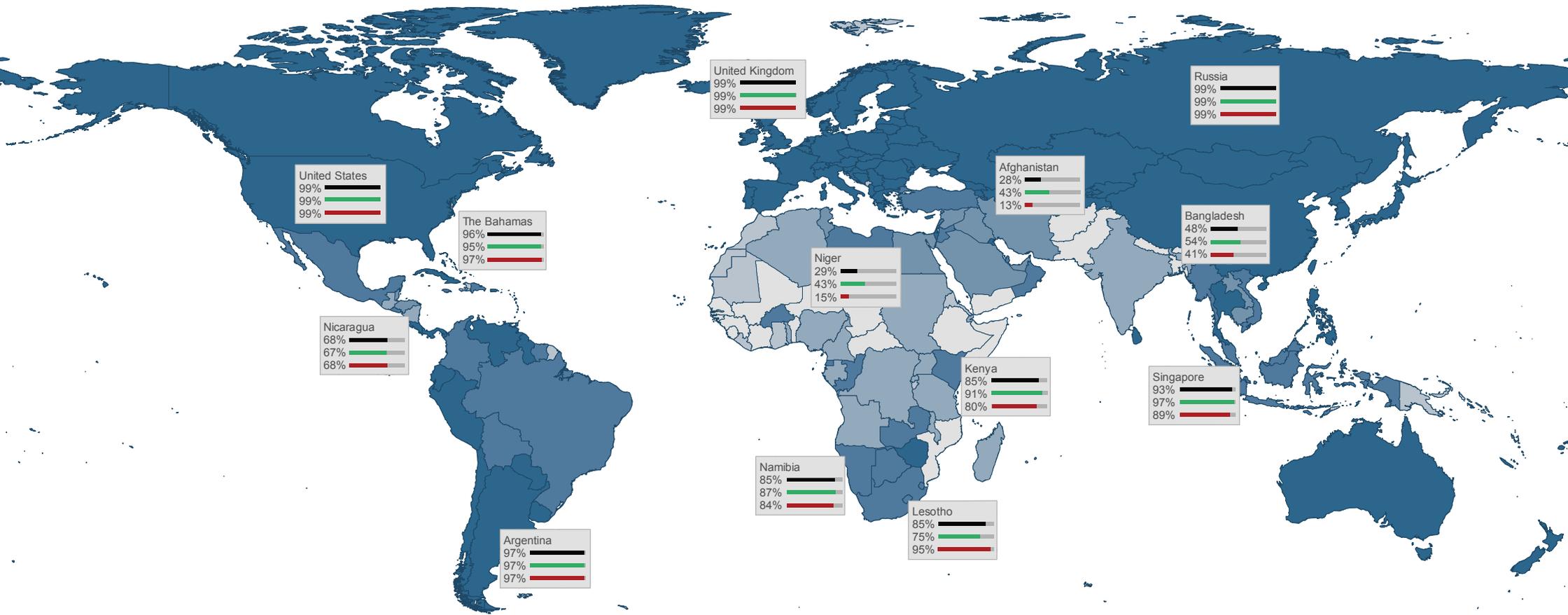
Data

It is a challenge to obtain statistics that uniformly define “literacy.” This map employs the CIA Factbook’s index of literacy rates, which merges data from several national studies. The most common definition of literacy is the number of people over 15 years of age who can read and write. However, a small number of countries report data using different criteria.

Findings

Many countries in the world have attained high rates of literacy. Large parts of the Americas, Europe and Asia are characterized by literacy rates that exceed 90%. However, there are still significant parts of the world in which more than a third of the adult population are unable to read and write. In particular, there are only a few countries in Africa, the Middle East and South Asia with full or high levels of literacy. Surprisingly, there are still twenty-one countries in which literate people are a minority of the population. It is also important to point out that there are significant differences in literacy rates between genders with females having lower rate of literacy than males. Globally, 82% of people are literate; but this statistic masks the fact that 87% of the world’s men can read and write whilst only 77% of women are literate. This contrast is especially pronounced in Afghanistan and Niger. In both countries, about three times more men than women are literate.





Adult literacy by gender



Adult literacy by country



Internet Penetration

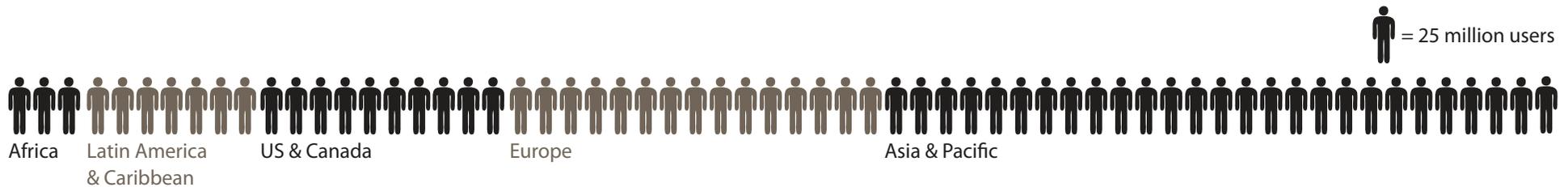
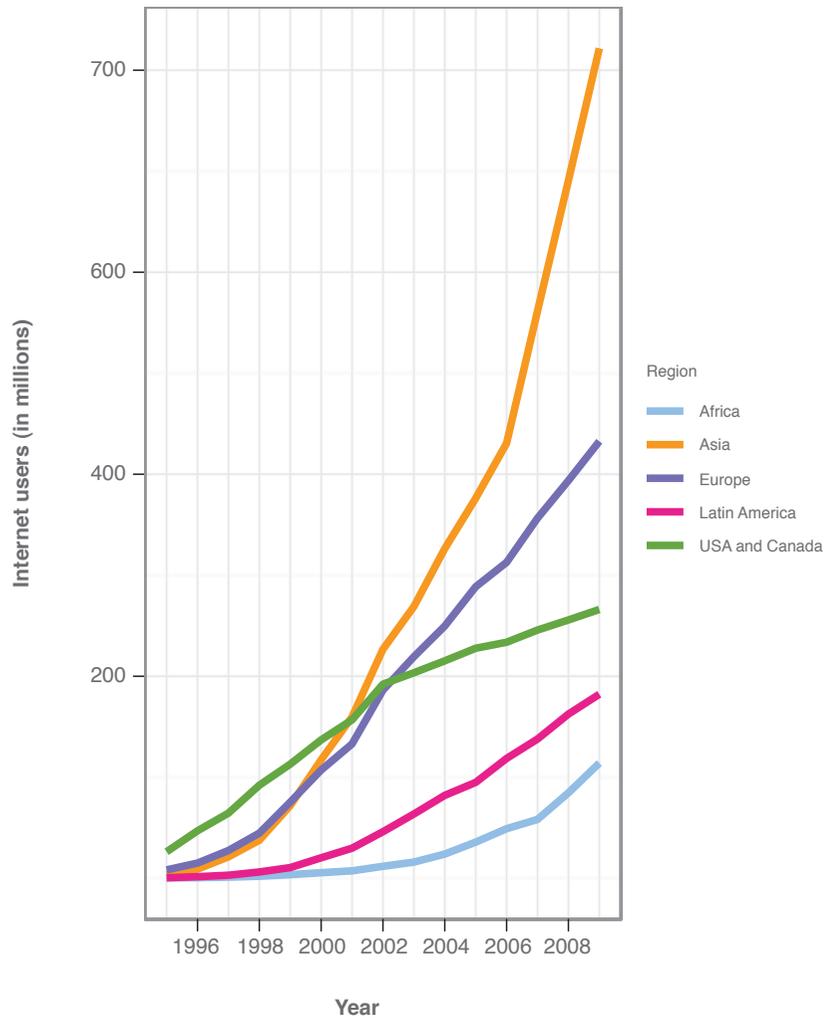
This visualisation illustrates the raw number of Internet users in each country as well as the percentage of the population with Internet access.

Data

This map uses 2008 statistics from the World Bank, which has tracked the number of Internet users per country and the number of Internet connections per 100 people since the 1990's as part of its Worldwide Governance Indicators project. The data are visualised with a cartogram in which the size of each country is drawn based on its proportion of global Internet users. The shading of each country reflects its Internet penetration rate: darker shades indicate higher levels of Internet usage amongst the population. Countries with online populations of less than approximately 2 million have been removed from the map.

Findings

We see that the map of the world's online population presents an interesting picture of the locations of Internet users. China has the world's largest total number of Internet users (there are currently over 400 million users in China) despite its relatively low penetration rate. The map also starkly illustrates the relatively small number of users in South America and Africa. The visualisation causes South America to shrink to a size that is smaller than the United States, and Africa to skew unrecognizably on the map. We also see that there are very few countries in the Global South with high Internet penetration rates. This indicates future growth in the total number of Internet users will most likely come from areas that are currently underrepresented.



The World's Newspapers

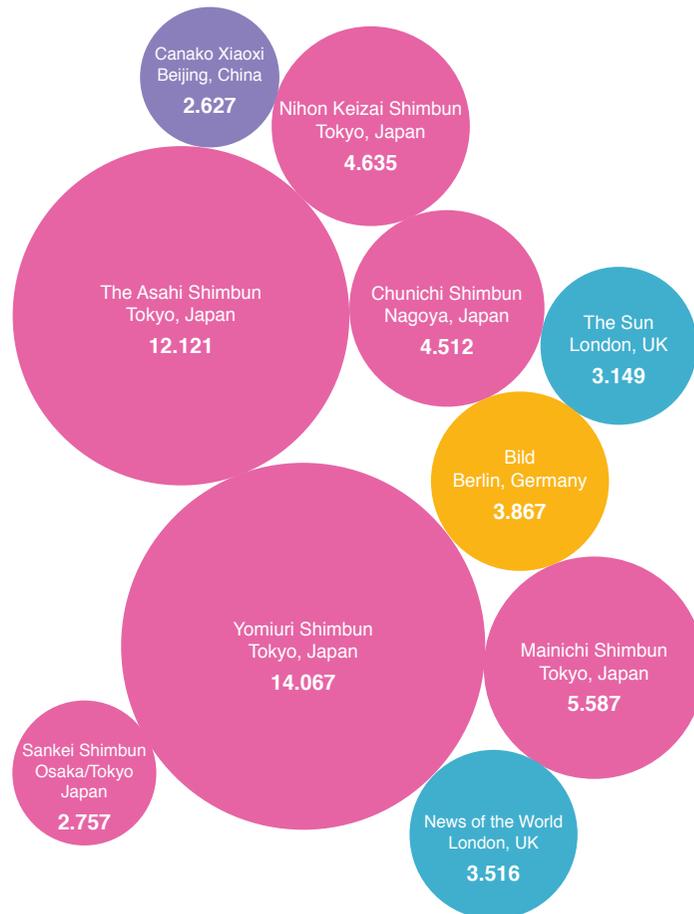
The importance and visibility of traditional media is often overlooked. This map visualises the world's 100 largest newspapers as well as the number of physical papers printed daily in each country.

Data

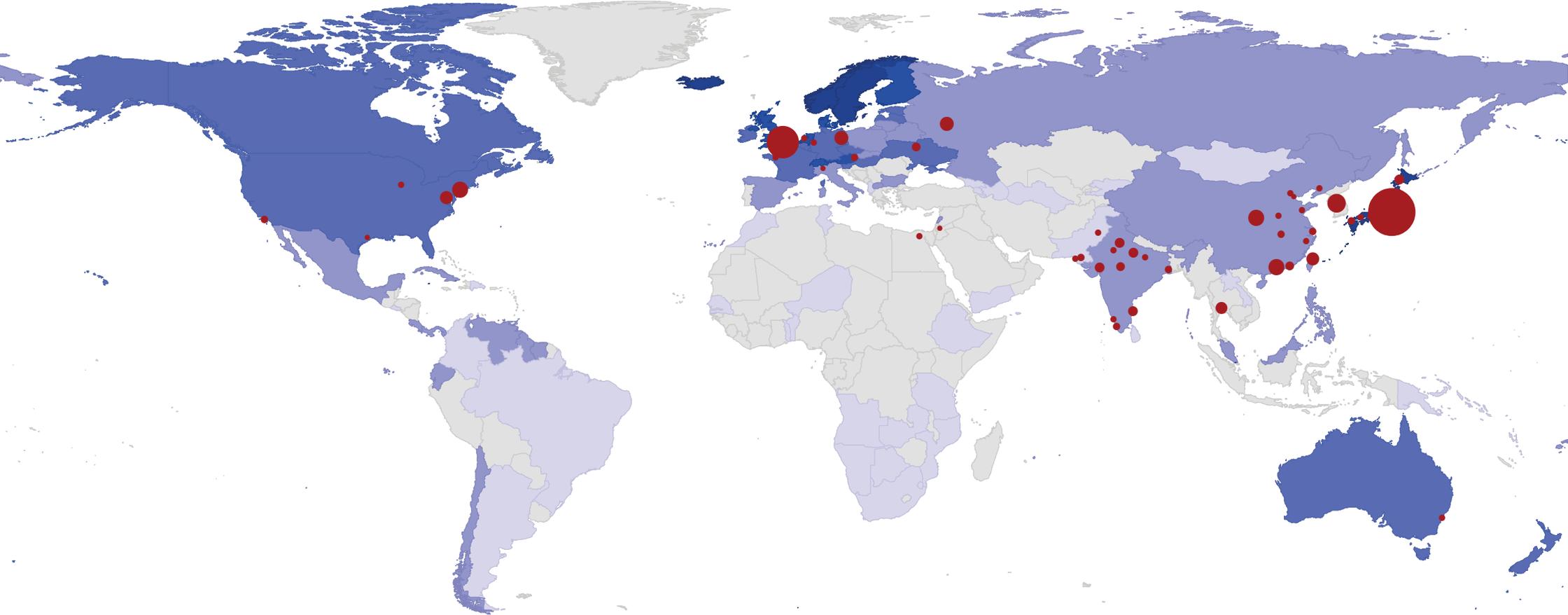
This map combines two data sources. First, it draws on 2005 data from the World Bank detailing the number of daily newspapers printed per 1000 people in each country (i.e. total printed copies rather than the diversity of newspapers). The World Bank defines a daily newspaper as a paper that is published at least four times a week. Second, the map uses data from Newspapers24.com in order to visualise the main offices of the world's 100 largest print newspapers. Statistics were aggregated in cases where cities shared multiple entries.

Findings

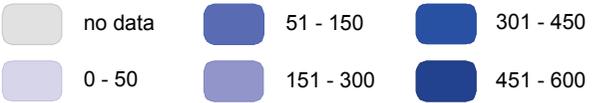
Perhaps unsurprisingly, this visualisation reveals a geographic pattern that is not dissimilar from the map of literacy on page 9. The world's wealthiest countries tend to have more per capita printed newspapers than the rest of the world. Scandinavia and Japan in particular stand out for having more daily newspaper prints (per person) than anywhere else on Earth. The locations of the world's 100 largest papers, in contrast, reveal some surprising patterns. The five largest papers by circulation are all in Japan. Indeed, a majority of newspapers on the list are in Asia: a fact that could be partially explained by the large number of mega-cities on that continent.



Circulation of world's largest newspapers (in millions)



Newspapers published per 1,000 people



Circulation of largest newspapers



The Location of Academic Knowledge

The production and publication of academic knowledge has distinct geographies. This graphic visualises the locations of academic journals listed in Thompson Reuters' Web of Knowledge: the most important and influential collection of academic content.

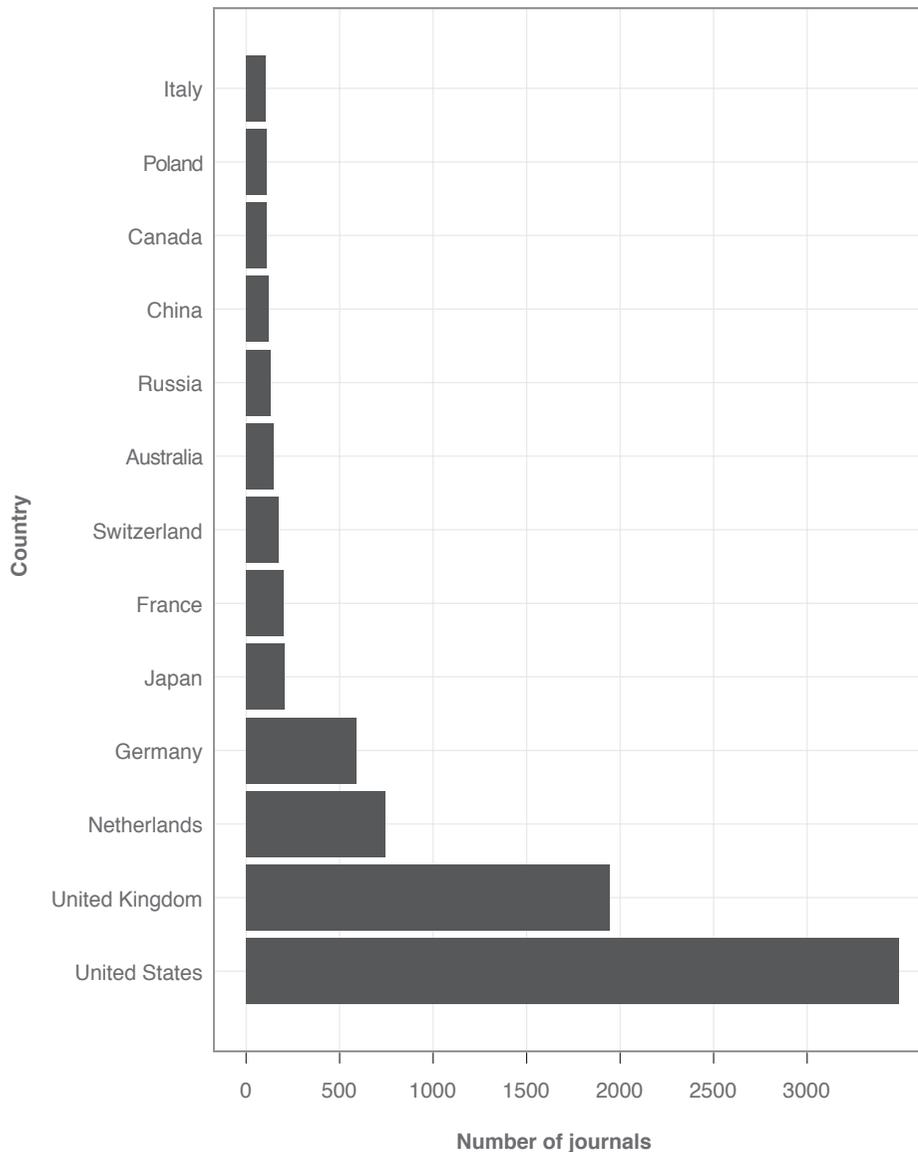
Data

This map uses data from the Web of Knowledge Journal Citation Reports (JCR) from 2009, allowing us to measure the locations and impact factors of journals. The JCR Science Edition contains references from over 7,300 journals in science and technology. The JCR Social Sciences edition contains references from over 2,200 journals in the social sciences. A reference for each of the 9,500 journals in the sciences and social sciences was downloaded to extract the journal's location. A cartogram is used in which each country is represented by a box that is sized according to the number of journals published from within it. The shading of each country indicates the average impact factor (a measure of how often articles within a journal are cited) of journals within that country.

Findings

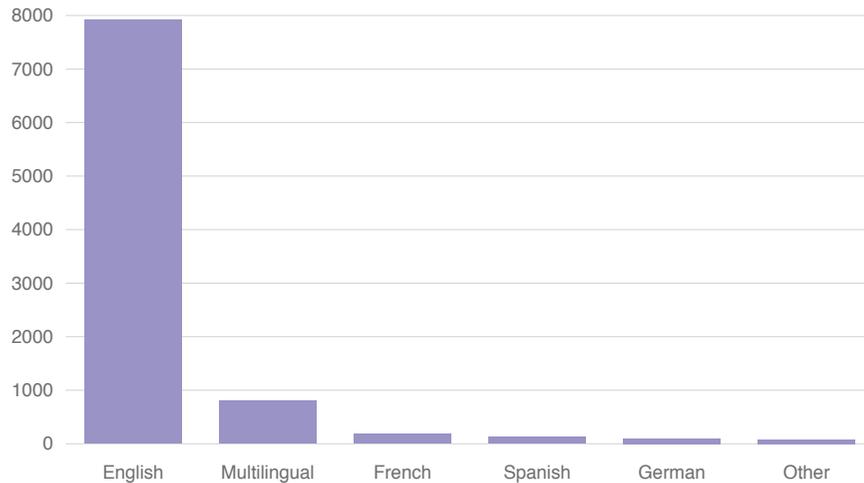
This map reveals a staggering amount of inequality in the geography of the production of academic knowledge. The United States and the United Kingdom publish more indexed journals than the rest of the world combined. Western Europe, in particular Germany and the Netherlands, also scores relatively well. Most of the rest of the world then scarcely shows up in these rankings. One of the starkest contrasts is that Switzerland is represented at more than three times the size of the entire continent of Africa. The non-Western world is not only under-represented in these rankings, but also ranks poorly on average citation score measures. Despite the large number and diversity of journals in the United States and United Kingdom, those countries manage to maintain higher average impact scores than almost all other countries. It is important to note that the 9,500 journals included in this map do not represent the entirety of all published journals. However, given the influence of the JCR, and its claims to provide a "systematic, objective means to critically evaluate the world's leading journals,"¹ it remains crucial to understand the geography of academic knowledge.

¹ http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/

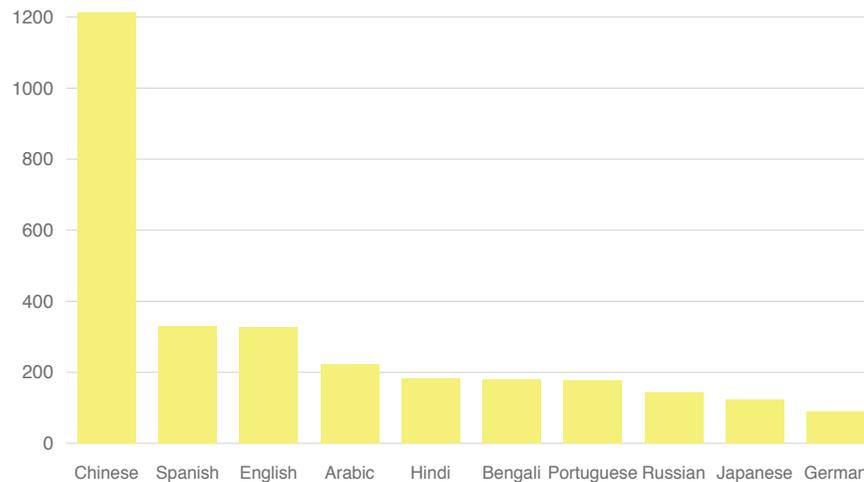


Academic Knowledge and Language

Number of journals by language



Number of native speakers (in millions)



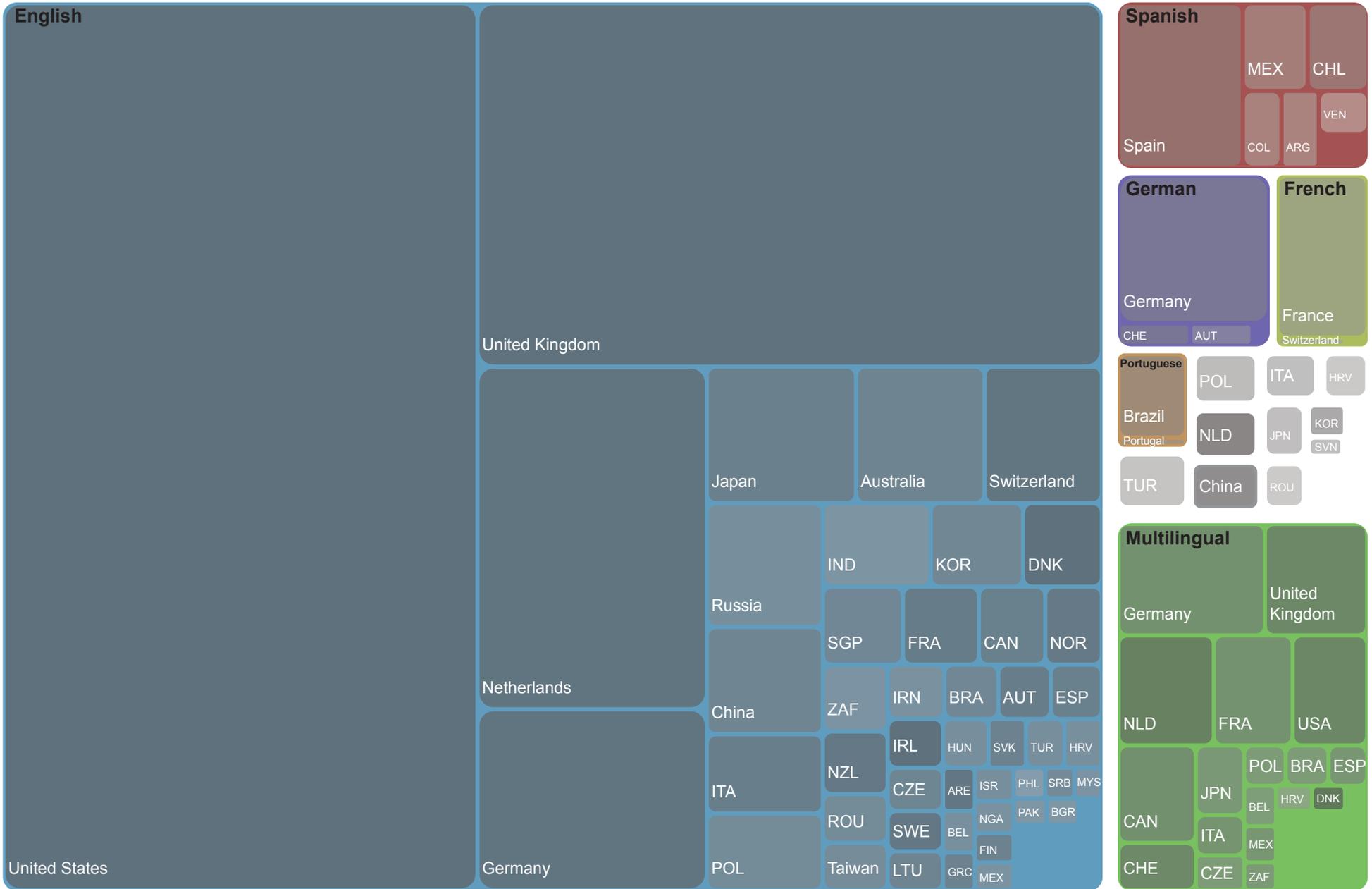
This graphic visualises the role that language plays within the reproduction of academic knowledge in scientific journals. This visualisation segments academic journals by language and country and shades each country by the average impact factor of the journals published within it.

Data

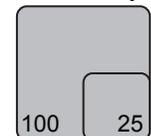
This graphic also draws from the Thompson Reuters Web of Knowledge Journal Citation Reports described further on page 14. The data from Thompson Reuters contains journals in twenty-two distinct languages. The impact of a journal is calculated as the number of citations that articles in the journal received in 2008 and 2007 divided by the total number of articles published in the journal over the same time period. To maintain clarity, most countries that contained only a small number of journals were omitted from the visualisation.

Findings

Building on the map of journals on page 15, this graphic highlights the dominance of the Anglophone sphere in academic publishing. English is a dominant language in academic publishing not only because of large number of journals published in the United States and United Kingdom. This visualisation demonstrates that English-language journals are also published from much of the rest of the world (English accounts for 86% of academic journals in the dataset). While this dataset is not a comprehensive sample of all journals published in the world, the fact that the Web of Knowledge dataset is so important to institutional and individual rankings in academia only serves to reinforce the dominance of the English language within the sciences and social sciences.



Number of journals

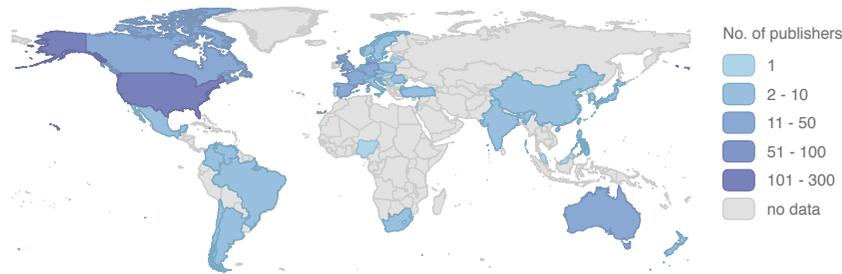


Impact factor

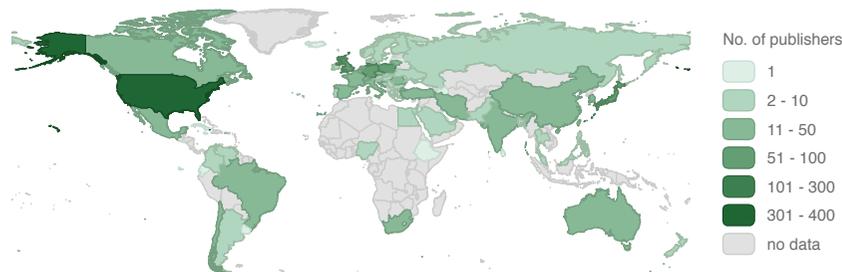


Academic Knowledge and Publishers

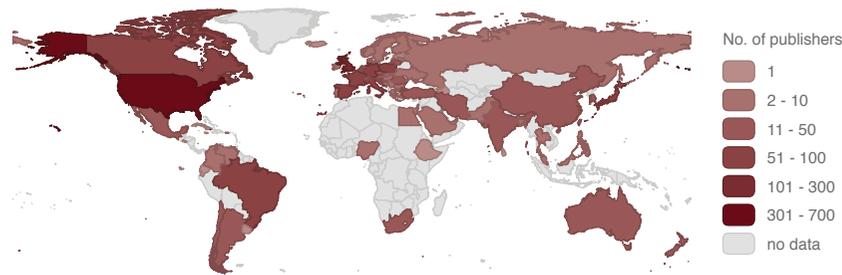
Publishers of social science journals



Publishers of science journals



Publishers of social science and science journals



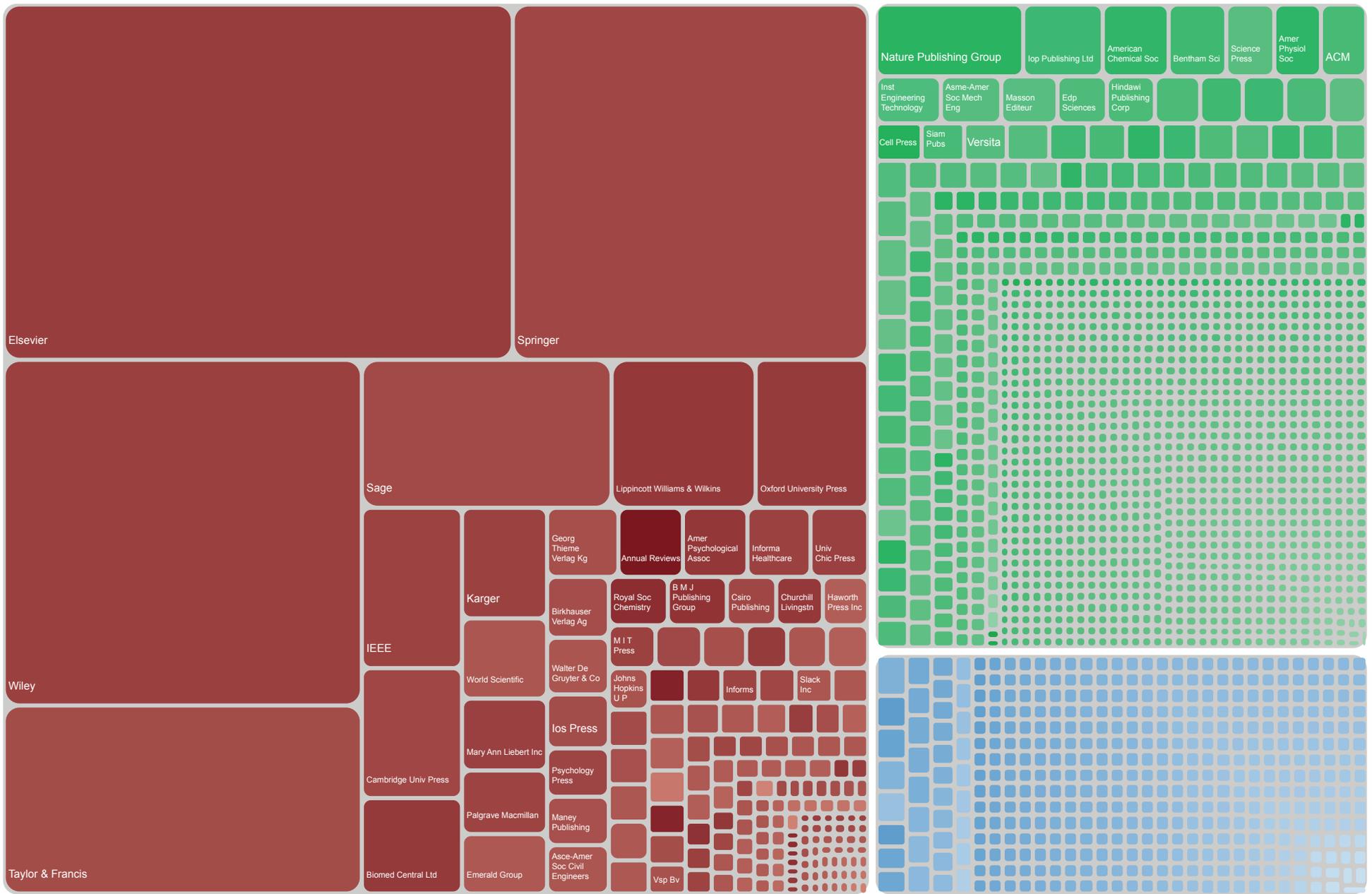
This series of graphics depicts the control of academic journals in the Web of Knowledge index by publishers. Mapping academic publishers allows us to understand the geography of who controls the printing and dissemination of academic knowledge.

Data

This visualisation uses data from the Web of Knowledge Journal Citation Reports (JCR) from 2009. Two types of visualisations are paired together on this graphic. The choropleth maps illustrate the number of publishers in each country with darker colours indicating more publishers. The treemaps indicate the number of journals published by each publisher. This graphic is segmented into three categories: publishers of science journals, publishers of social science journals and publishers of both.

Findings

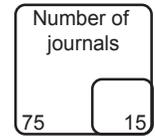
Despite the absence of linguistic and geographic diversity in academic publishing, there remains a surprising lack of concentration amongst journal publishers. Within the groups of publishers that focus only on journals in the sciences or social sciences, the publication of journals is distributed through many organisations and companies. The larger group of publishers that control both science and social science journals, on the other hand, are characterised by a greater degree of clustering (i.e. fewer organisations controlling relatively large numbers of journals). Springer, Wiley-Blackwell, Elsevier and Taylor & Francis control a large amount of the academic publishing market and all have relatively high average citations scores.

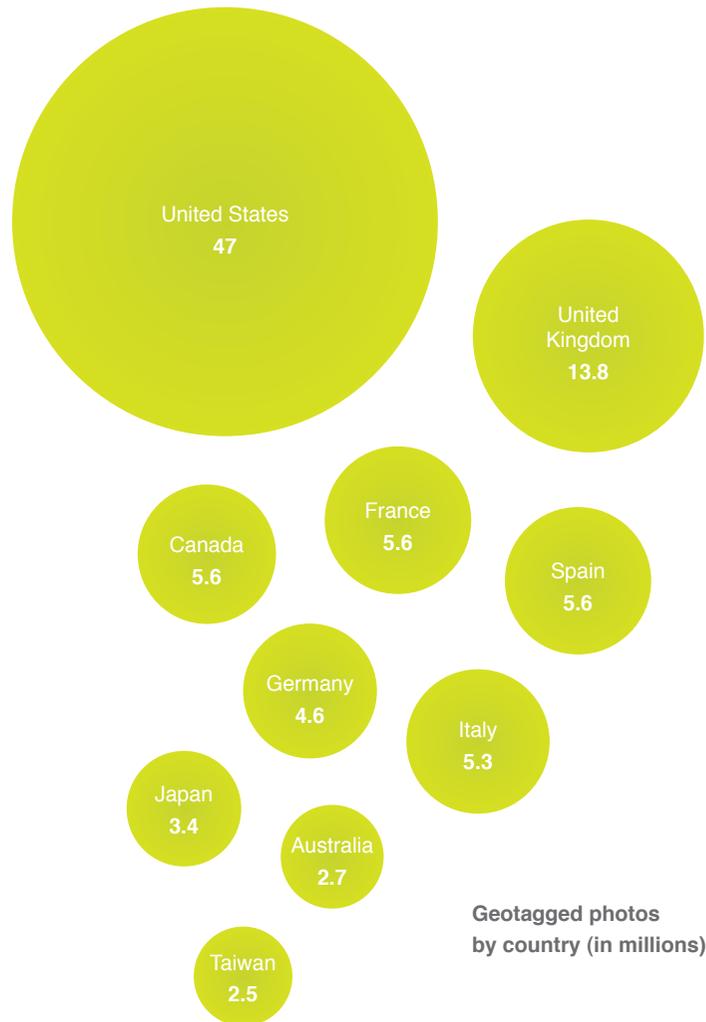


● Publisher of science and social science journals

● Publisher of only science journals

● Publisher of only social science journals





Mapping Flickr

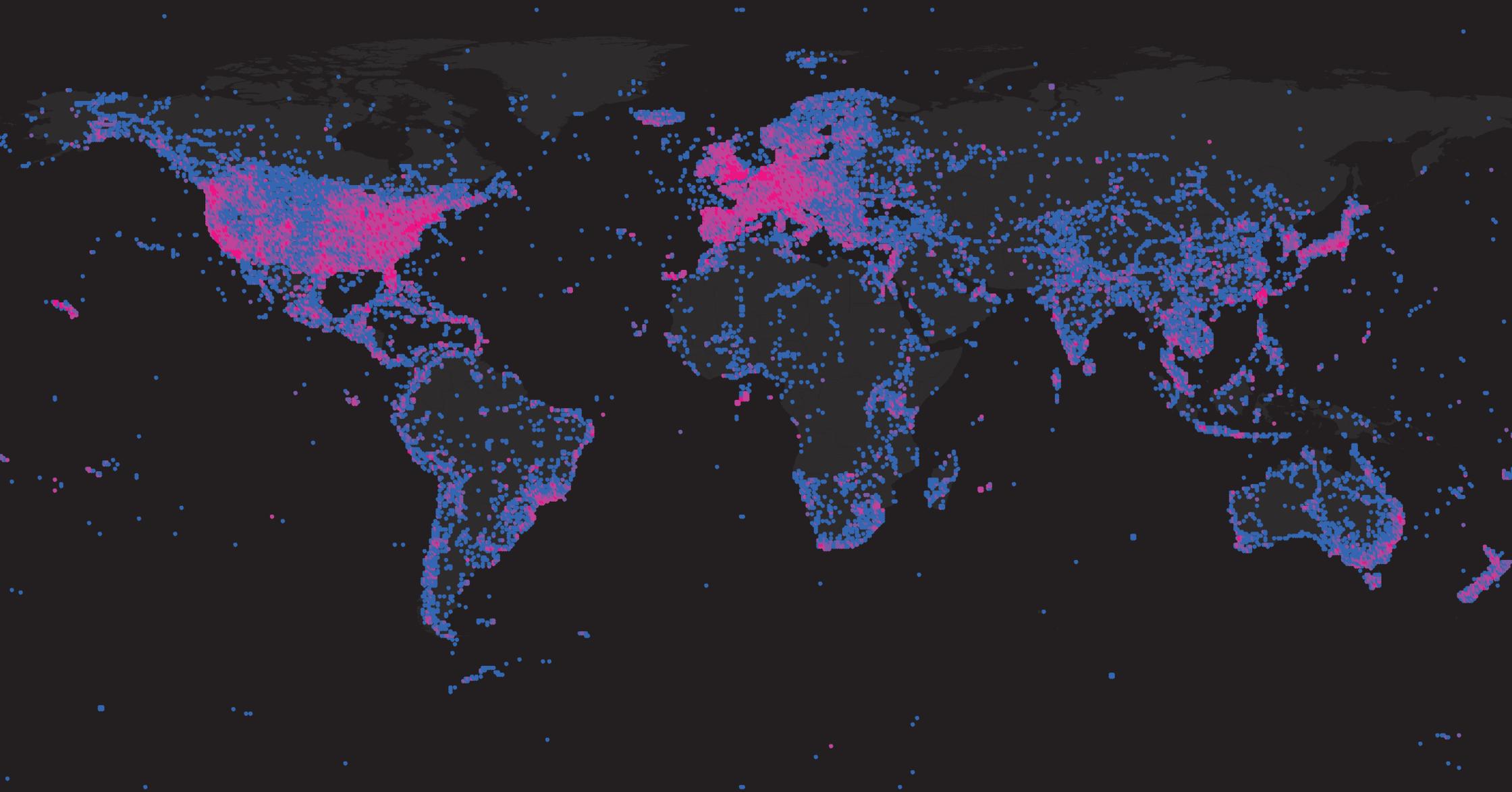
Images are an important form of knowledge that allow us to develop understandings about our world. Flickr is the world's most used and most popular public repository of photographs and currently hosts over five billion images. This map reveals the global geographic distribution of geotagged images on the platform, and thus reveals the density of visual representations and locally depicted knowledge of all places on our planet.

Data

Each point on the map indicates the total number of geotagged photographs uploaded to Flickr at that location. All geotagged photos were collected through Flickr's application programming interface (API) using a custom designed program on April 1, 2011. Our program downloaded the count of geotagged photographs in every 0.5 x 0.5 degree latitude-longitude square on the Earth's surface. Due to the convergence of lines of longitude at the poles, our sampling boxes become larger as they approach the equator.

Findings

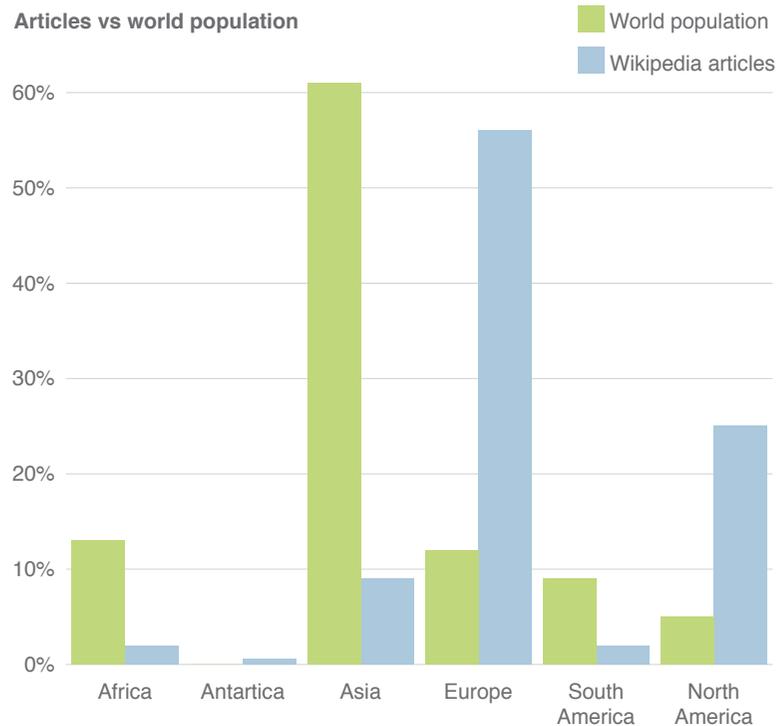
As might be expected, the largest concentrations of photographs can be found in some of the world's most populated places. Much of Western Europe, North America and East Asia are covered by dense layers of virtual representation. However, the density of photographs is not simply a factor of population. Potentials for participation are limited by both censorship of the platform (e.g. in Iran) and the presence of more popular services (e.g. in China). In the rest of the world, the patterns on this map point to images being mostly created by people in the world's wealthiest and most highly connected regions. The stark contrast in the number of photographs covering Mexico and the United States and the lack of dense clusters of images in most populated parts of Africa (e.g. coastal West Africa) reinforces this point.



Number of photographs • 5 - 100 • 101 - 500 • 501 - 10,000 • 10,001 - 1,000,000

The Distribution of all Wikipedia Articles

Articles vs world population



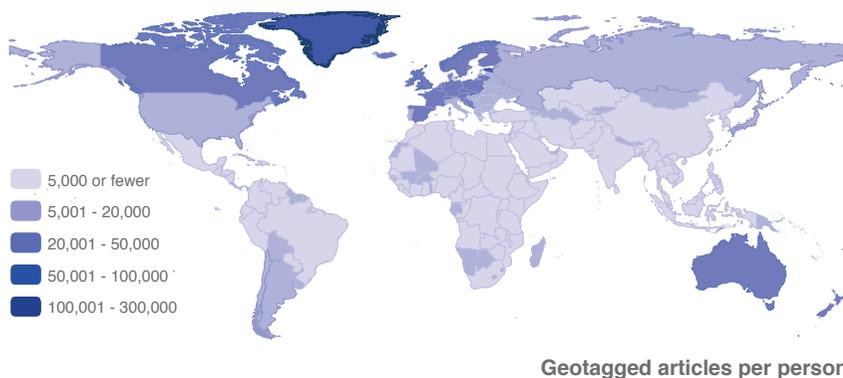
Wikipedia is one of the world's largest and most important repositories of crowdsourced knowledge. This map uncovers the distinct geographies of that information.

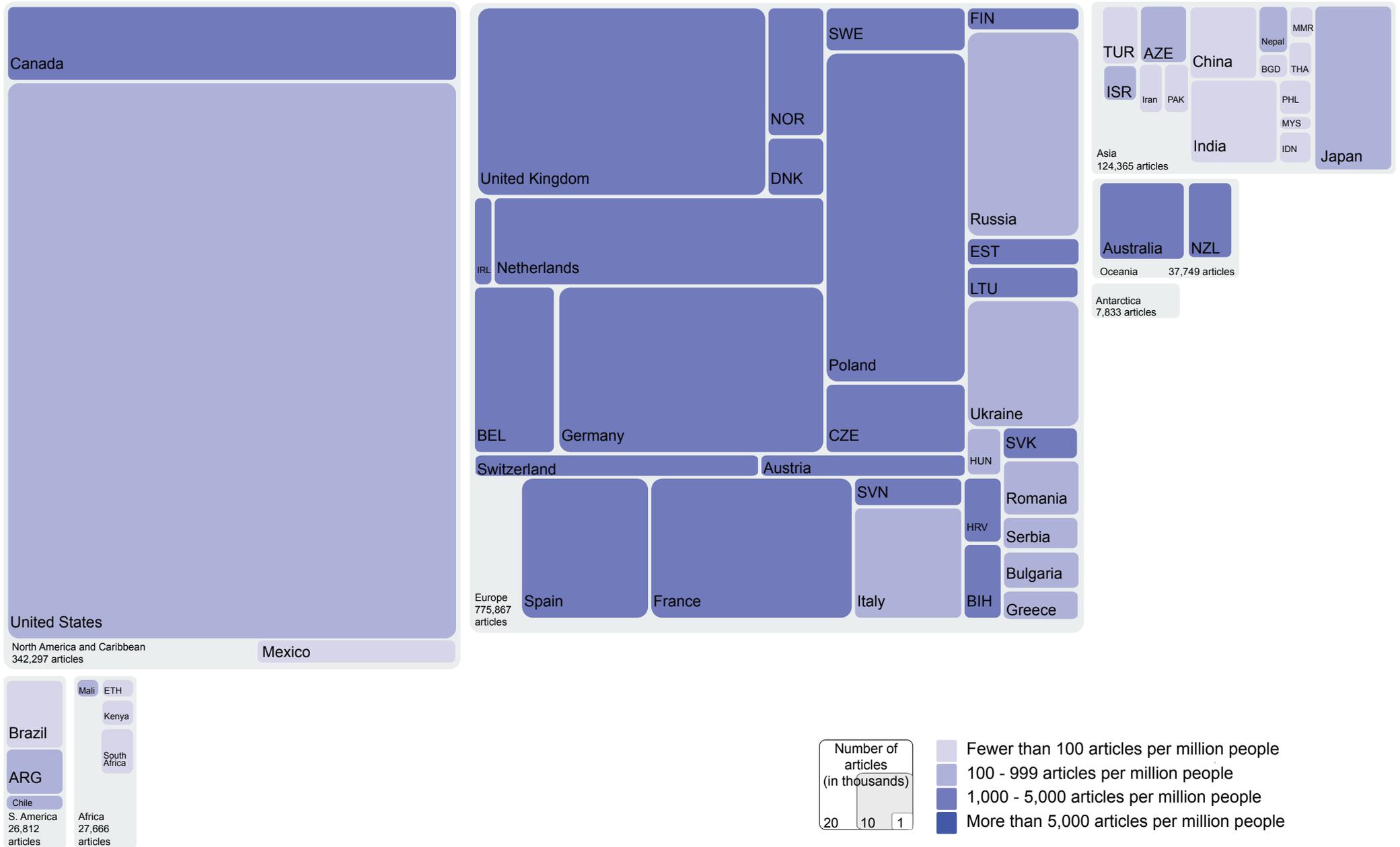
Data

Most Wikipedia articles about places, events or any other locatable articles are geotagged with a pair of latitude and longitude coordinates. We downloaded the list of approximately 1.5 million articles from a 2010 database of Wikipedia and joined them to a file containing the boundaries of every country in order to determine the total number of geotagged articles in every country. Although not every article can be geotagged (e.g. there would be no reason to tag articles on 'blueberries' or 'zombies') there is no reason to assume that there is any systematic or geographic bias in the tagging of articles. The dataset that we are using contains geotagged articles written in any language version of Wikipedia.

Findings

There is a clear and highly uneven geography of information in Wikipedia. Europe and North America are home to 84% of all articles. Anguilla has the fewest number of geotagged articles (four), and indeed most small island nations and city states have less than 100 articles. However, it is not just microstates that are characterised by extremely low levels of wiki representation. Almost all of Africa is poorly represented in the encyclopaedia. There are remarkably more Wikipedia articles (7,800) written about Antarctica than any country in Africa or South America. Even China, which is home to the world's biggest population of Internet users and is the fourth largest country on Earth contains fewer than 1% of all geotagged articles. Because of the high visibility of Wikipedia in online information ecosystems, countless decisions are made and countless opinions are formed based on information available in the encyclopaedia. It is thus important to point out the digital terra incognita that covers much of the world and reproduces existing representational asymmetries.





Time-series of the Distribution of Biographies on Wikipedia over the Last Five Centuries

In order to further explore geographic differences of knowledge about the world in Wikipedia, this time-series of maps visualises the distribution of articles about people in the encyclopaedia from the sixteenth century until the present-day.

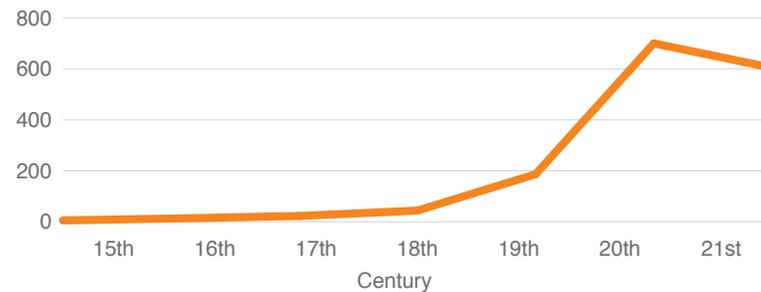
Data

These graphics visualise references to places within 1,575,865 biography articles in the English version of Wikipedia. Every biography was georeferenced by counting the number of references to place names in each person's biography and then mapping only the most mentioned place in each article. Ranking of placenames was conducted not only using the English version of the article, but also using the equivalent in up to seven languages (English, German, French, Dutch, Spanish, Italian and Portuguese). The method therefore aims not to provide a comprehensive overview of the geography of articles about people in all Wikipedia language versions, but rather offers a look at the geographic focus of the subjects of key Western European languages.

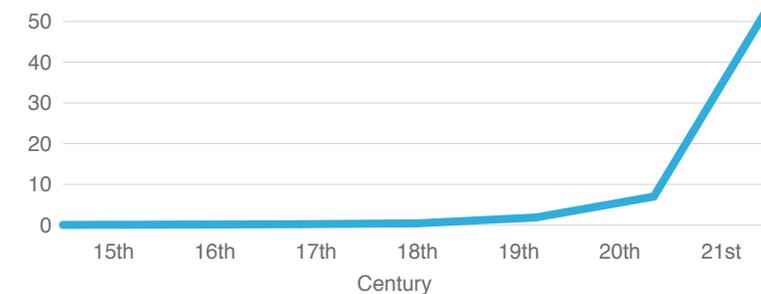
Findings

Articles about people in Wikipedia are highly likely to reference particular parts of the world (the US and Western Europe), reinforcing many of the patterns seen in our other graphics. Nonetheless, this is a geography of people that is in no way reflective of the actual distribution of population on our planet. Wikipedia guidelines specify that biographies should only be about notable people and this map suggests that Wikipedia editors believe there to be vastly more notable people in Europe and North America than anywhere else in the world. It is therefore clear that much more work needs to be done before Wikipedia can attain its stated goal of storing the “the sum of all human knowledge.”

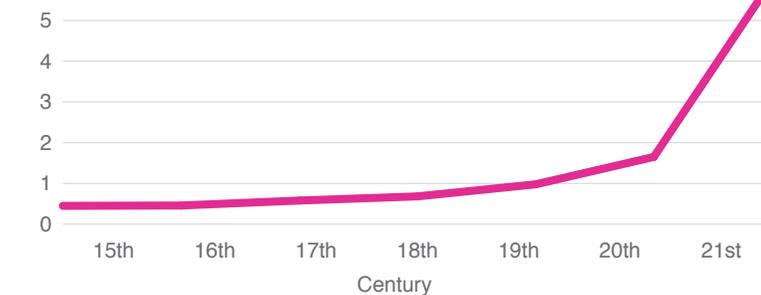
Number of Wikipedia biographies (in thousands)



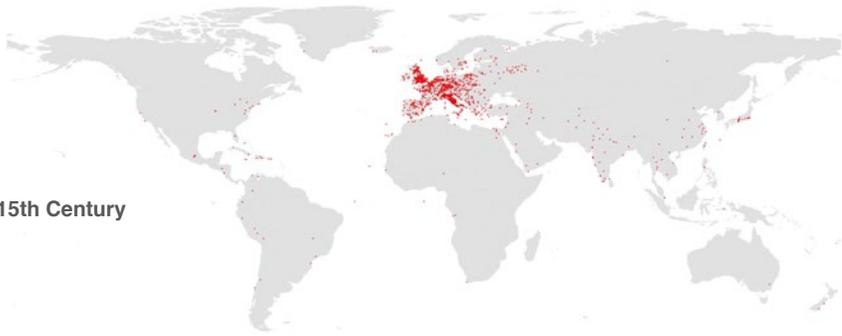
Biographies per year (in thousands)



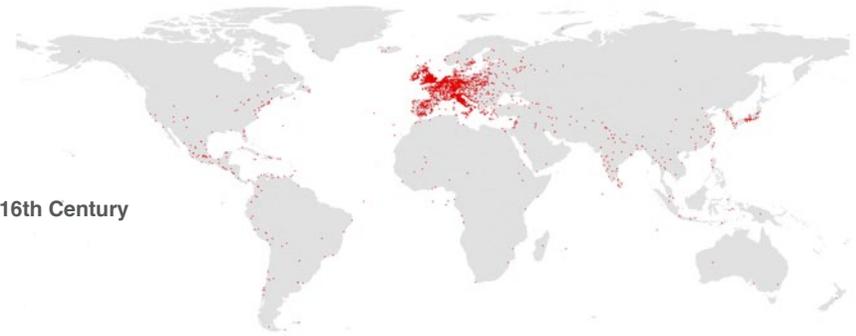
World population (in billions)



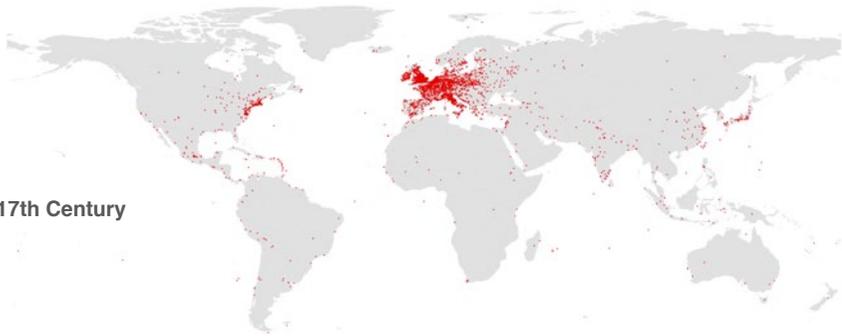
15th Century



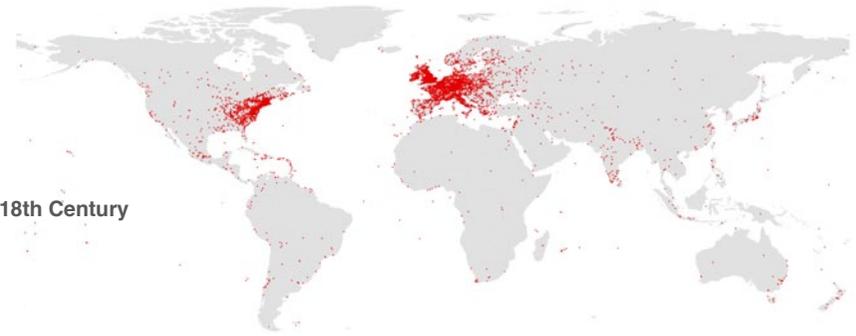
16th Century



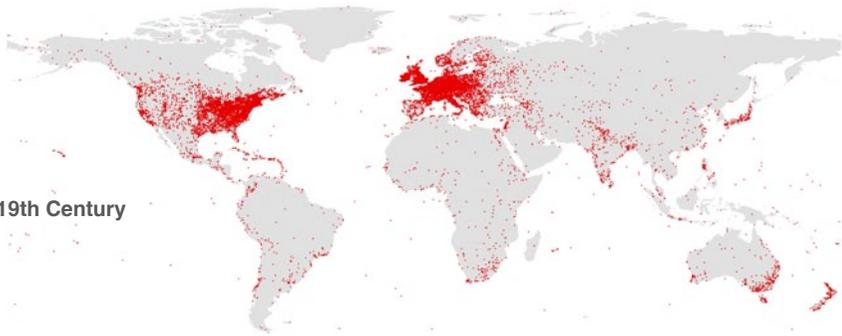
17th Century



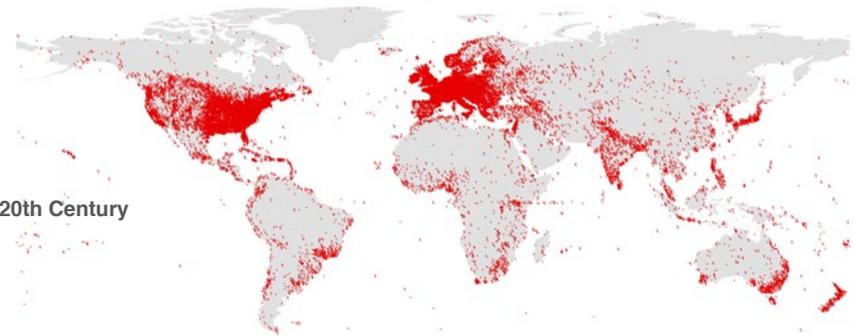
18th Century



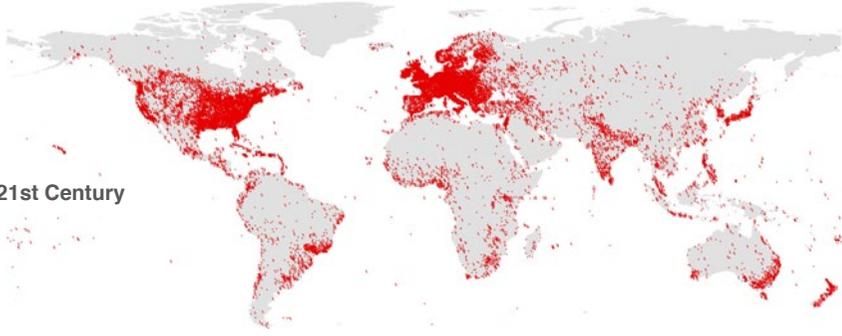
19th Century



20th Century



21st Century



User-generated Content in Google

Knowledge indexed by Google plays a key role in how we perceive our material, offline environment. This map visualises the amount of user-generated content indexed by Google in 2009 in a sample of 250,000 points around the world.

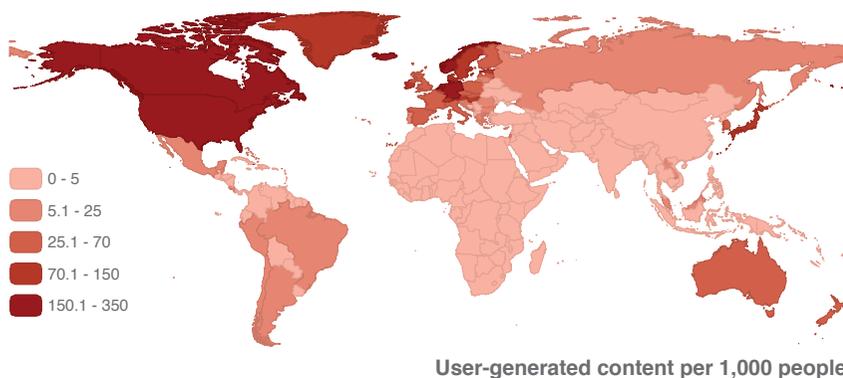
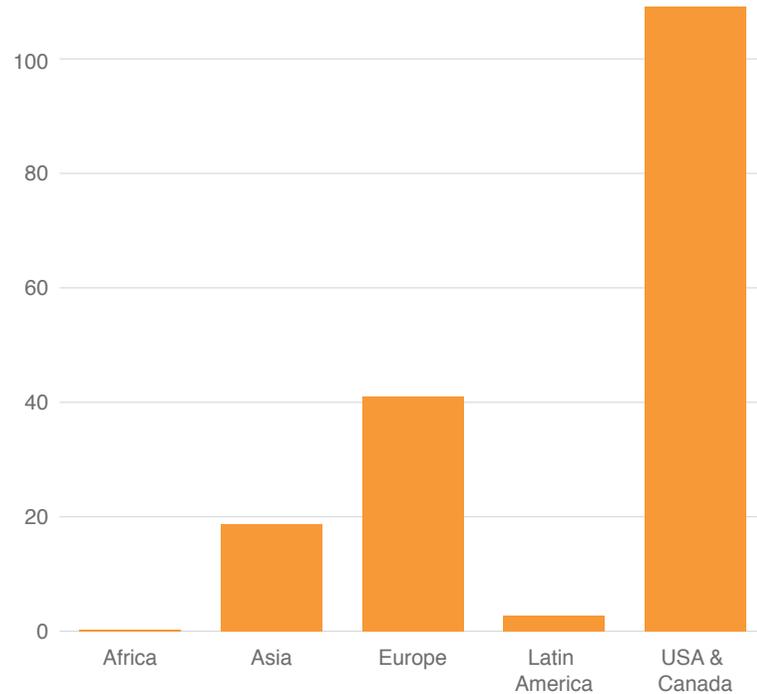
Data

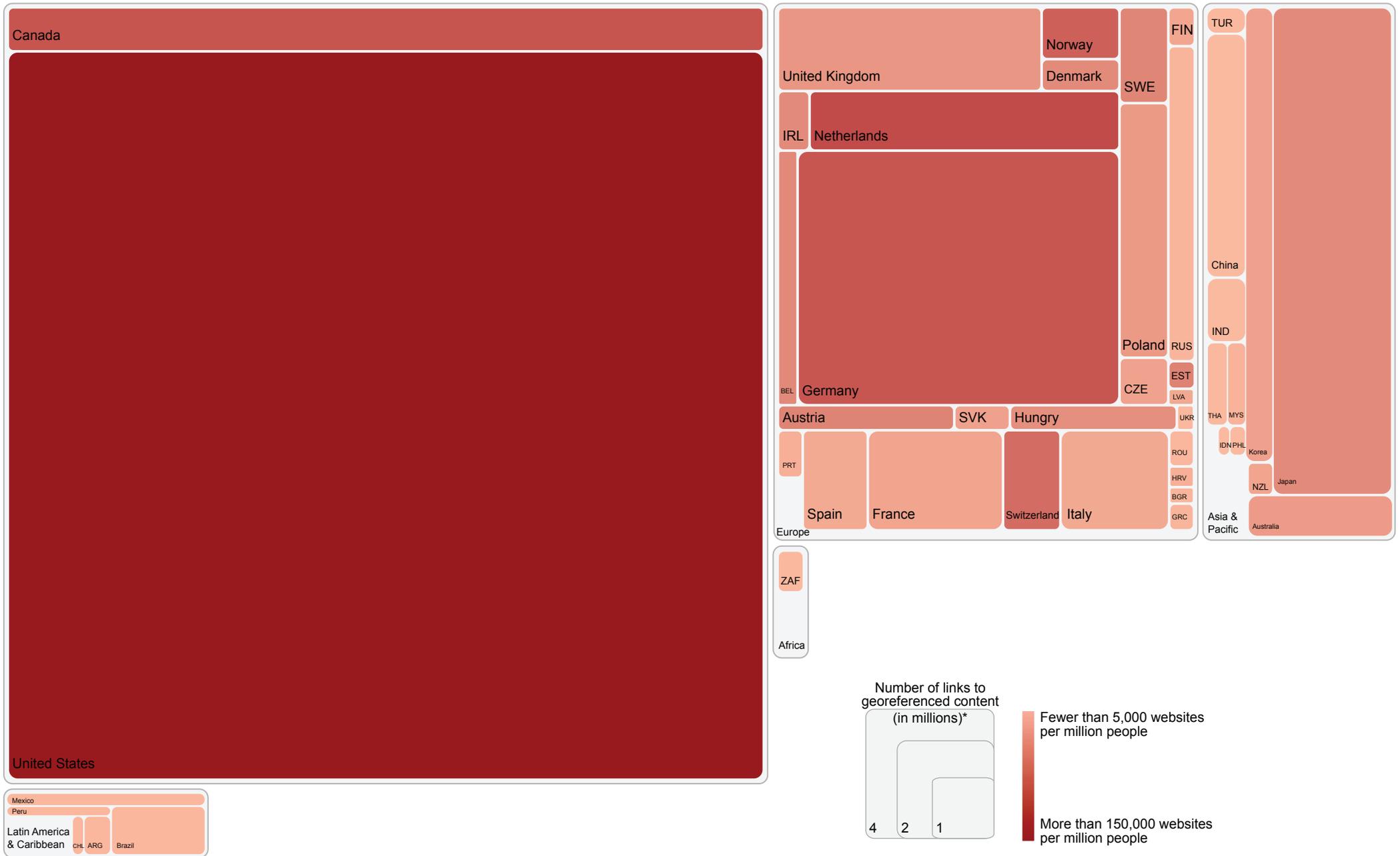
In order to measure the amount of user-generated content indexed by Google in each country, a dataset was created based on a 0.25 x 0.25 degree grid of all the land mass in the world (roughly 250,000 points). A buffer was then constructed for each point using a sliding variable size based on the great circle distance to neighbouring points in the grid pattern. It was important to adjust this value in order to compensate for decreasing distance between longitudes as the software moves from the equator to the poles. For each point and buffer combination a search was run in Google Maps to measure the total number of hits for user-generated content at each location (as defined by Google).

Findings

This graphic again highlights the staggering amount of unevenness in the production and dissemination of information. The United States, and to a lesser extent Europe and Japan, are home to the bulk of the world's content (90%). The poor rankings of some countries are undoubtedly due to censorship or competing platforms, but even in the rest of the world we see that large swathes of territory unambiguously have very little content created about them. We can therefore see that user-generated content is far from being a simple mirror of either population density or human activity.

User-generated content by region (in millions)





Appendix

Country codes used in this booklet

ARE	United Arab Emirates	HRV	Croatia	ROU	Romania
ARG	Argentina	HUN	Hungary	RUS	Russia
AUT	Austria				
AZE	Azerbaijan	IDN	Indonesia	SAU	Saudi Arabia
		IND	India	SDN	Sudan
BEL	Belgium	IRL	Ireland	SGP	Singapore
BGD	Bangladesh	IRN	Iran, Islamic Republic of	SRB	Serbia
BGR	Bulgaria	ISR	Israel	SVK	Slovakia
BIH	Bosnia and Herzegovina	ITA	Italy	SVN	Slovenia
BLR	Belarus			SWE	Sweden
BRA	Brazil	JPN	Japan	SYR	Syrian Arab Republic
CAN	Canada	KOR	Korea, Republic of	THA	Thailand
CHE	Switzerland			TUN	Tunisia
CHL	Chile	LTU	Lithuania	TUR	Turkey
CHN	China	LVA	Latvia		
COL	Colombia			UKR	Ukraine
CZE	Czech Republic	MEX	Mexico	URY	Uruguay
		MMR	Myanmar	USA	United States
		MYS	Malaysia		
DNK	Denmark			VEN	Venezuela, Bolivarian Republic of
DZA	Algeria	NGA	Nigeria	VNM	Vietnam
		NLD	Netherlands		
ECU	Ecuador	NOR	Norway	ZAF	South Africa
ESP	Spain	NZL	New Zealand		
EST	Estonia				
ETH	Ethiopia	PAK	Pakistan		
		PER	Peru		
FIN	Finland	PHL	Philippines		
FRA	France	POL	Poland		
		PRT	Portugal		
GRC	Greece				



WWW.CONVOCO.CO.UK