Wikipedia Portal Traffic and JavaScript Support

Mikhail Popov (Analysis & Report), Julien Girault (UX Engineering)

February 11, 2016

Executive Summary

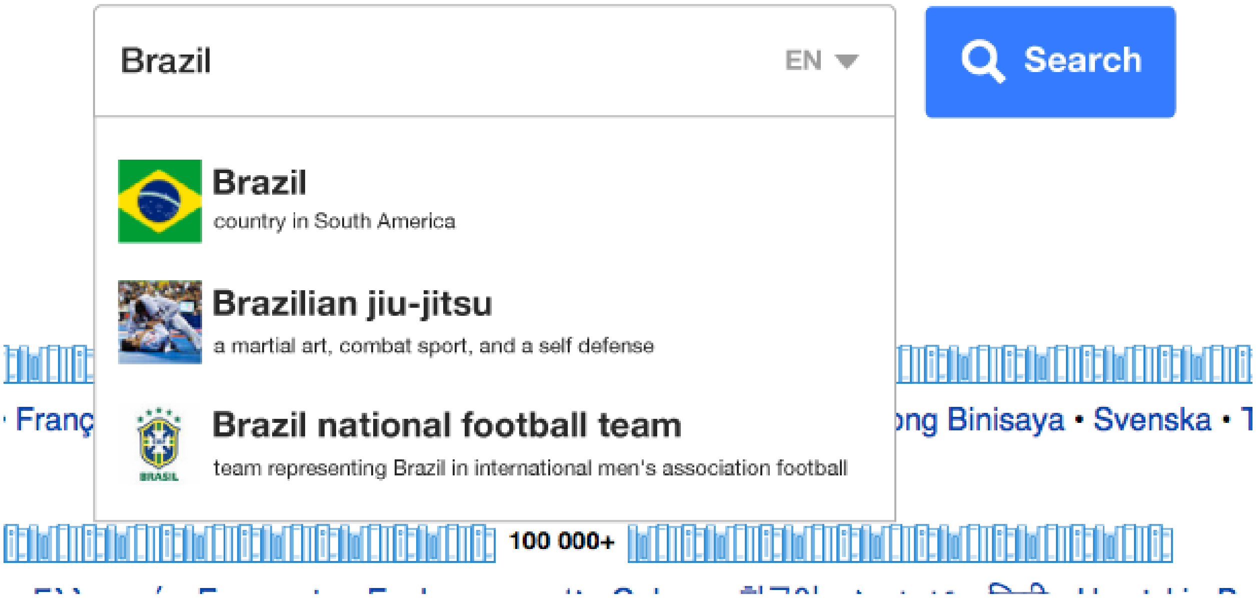
On 5 February 2016 we deployed a patch to the Wikipedia Portal moving the inline JavaScript into a separate file, which enabled us to finally measure the proportion of traffic with JS support separate from the overall traffic to the Portal. This report covers logs of HTTP request from 5 Feb to 10 Feb, 2016.

Overall, 93% of the requests made to the Wikipedia Portal have JS support. However, a large component (45%) of this overall percentage is accounted by traffic from United States, which has an overall proportion of 96%. The remaining 55% of the traffic from 234 other countries show a lot of variation in JS support, with 86.5% on average.

We also performed an analysis of browser usage and learned that approx. 75% of the traffic comes from users with relatively modern browsers, with a few exceptions such as Internet Explorer 8 (3.2% of total traffic). Of those 17 browsers, 14 had populations with more than 93% JS support. That is, less than 7% of those browsers' users had turned off JavaScript for privacy/bandwidth/other reasons. Interestingly, only 80% of Opera Mini 7 traffic and 60% of Android 4 / Chrome Mobile 30 traffic had JS support.

Introduction

The Discovery team relies on JavaScript (JS) for event logging and search autocompletion, but until now we did not know what proportion of our users support JS and have it turned on. It is especially important for the Wikipedia Portal team to know this proportion, as it dictates which conceptual avenues to pursue and how much support to provide those improvements.



Brazil

Brazil
country in South America

Brazilian jiu-jitsu
a martial art, combat sport, and a self defense

Brazil national football team
team representing Brazil in international men's association football

Figure 1: Mock-up showing thumbnail images and Wikidata descriptions in typeahead suggestion box using JavaScript.

Methods

To identify what proportion of Wikipedia Portal requests lack JavaScript support (T124655), we decided to compare the number of sampled requests to the recently deployed JS file (see T124713) relative to the number of sampled requests to the Wikipedia Wordmark. The rationale for this being that if a user visits wikipedia.org, their browser will only download the JS file if the browser supports JS **ANND** the user has left it enabled, but their browser will definitely make a request for the Wordmark file. Note: some countries made more requests for the JS file than the wordmark image, which suggests that certain numbers and percentages should be taken with a grain of salt, and should be regarded as rough estimates.

Also, while we filtered out as much bot and spider traffic as we could prior to aggregation and analysis, but we would like the reader to note that our databases of bots' user agents (UAs) are not complete.

Results

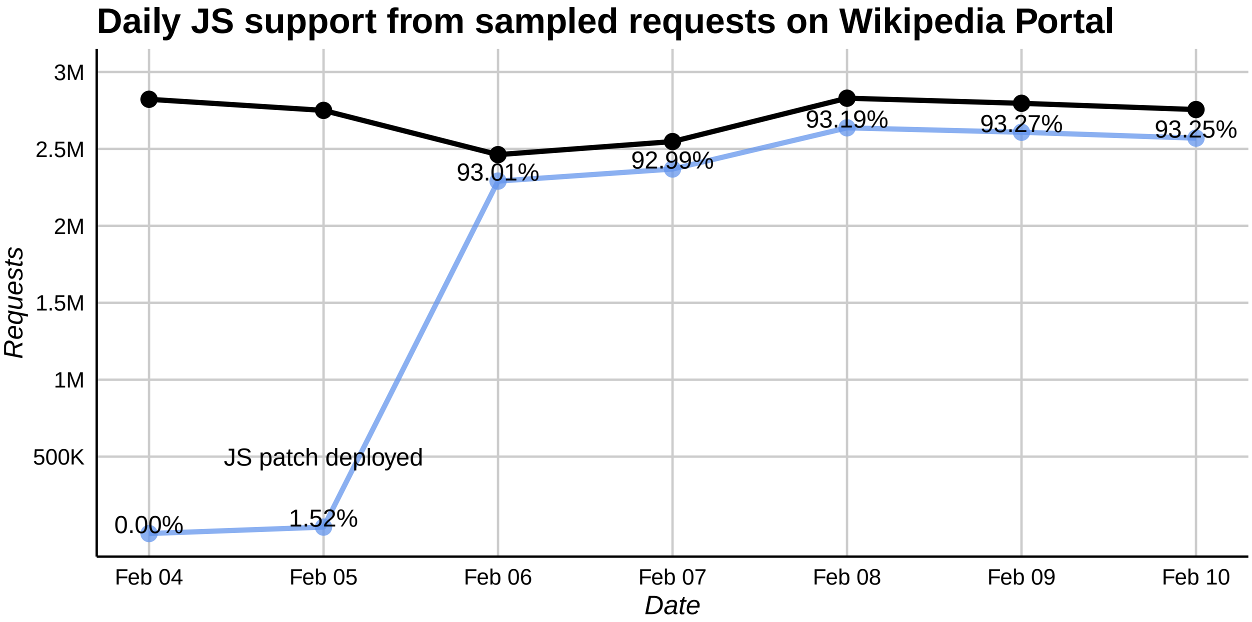


Figure 2: The 5 days of recorded data suggests that the overall proportion of Wikipedia Portal requests with JS support is approximately 93%.

Table 1: Summary of JS support % across 235 countries.

Min %	25th Percentile	Average %	Median %	75th Percentile	99th Percentile
50.00%	81.09%	86.53%	88.35%	93.77%	100.00%

The median in Table 1 (88.4%) is different than Fig. 1's 93% because in the figure we used the total of all requests, while the numbers in Table 1 reflect a distribution of JS support in requests broken up by countries.

Figure 3 is a *choropleth* – a map shaded according to a statistic – for which we excluded 199 countries with less than 100 sampled Wordmark requests to make the proportion estimates more reliable. The rationale for this is that, for example, if we recorded 11 Wordmark requests and 10 or 11 JS requests, that would yield 91% and 100% JS support, respectively, but those estimates are unreliable due to a small sample that is most likely not representative of the entire population. A more inclusive choropleth, containing 235 countries, is also provided.

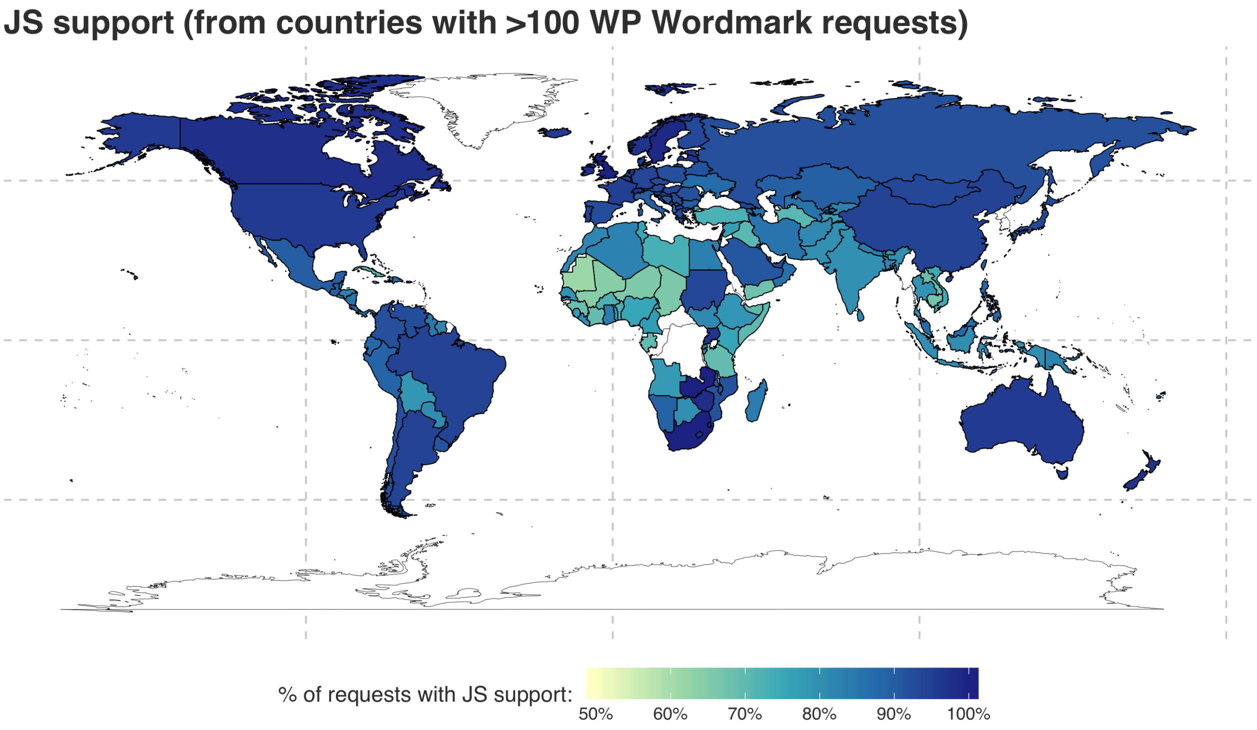

JS support (from countries with >100 WP Wordmark requests)
% of requests with JS support:
50%
60%
70%
80%
90%
100%

Figure 3: Approximate JS support across 199 countries. The estimate is the median of 5 days of data.

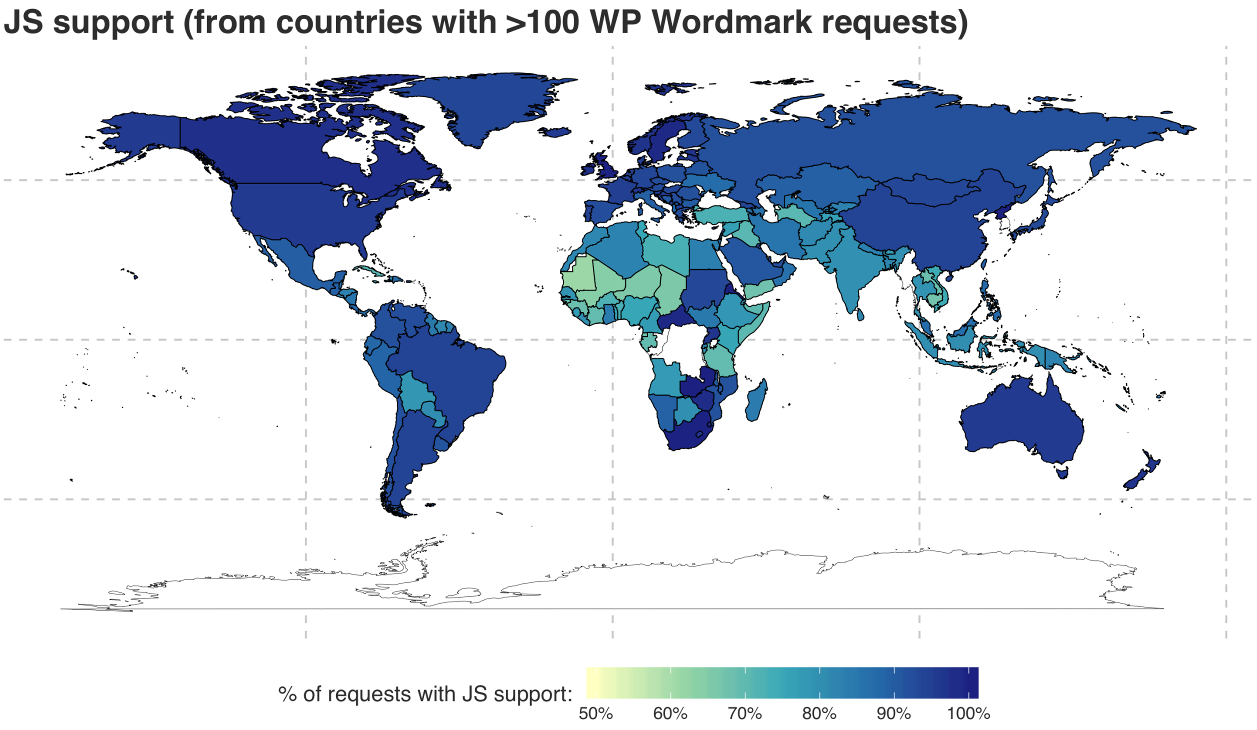
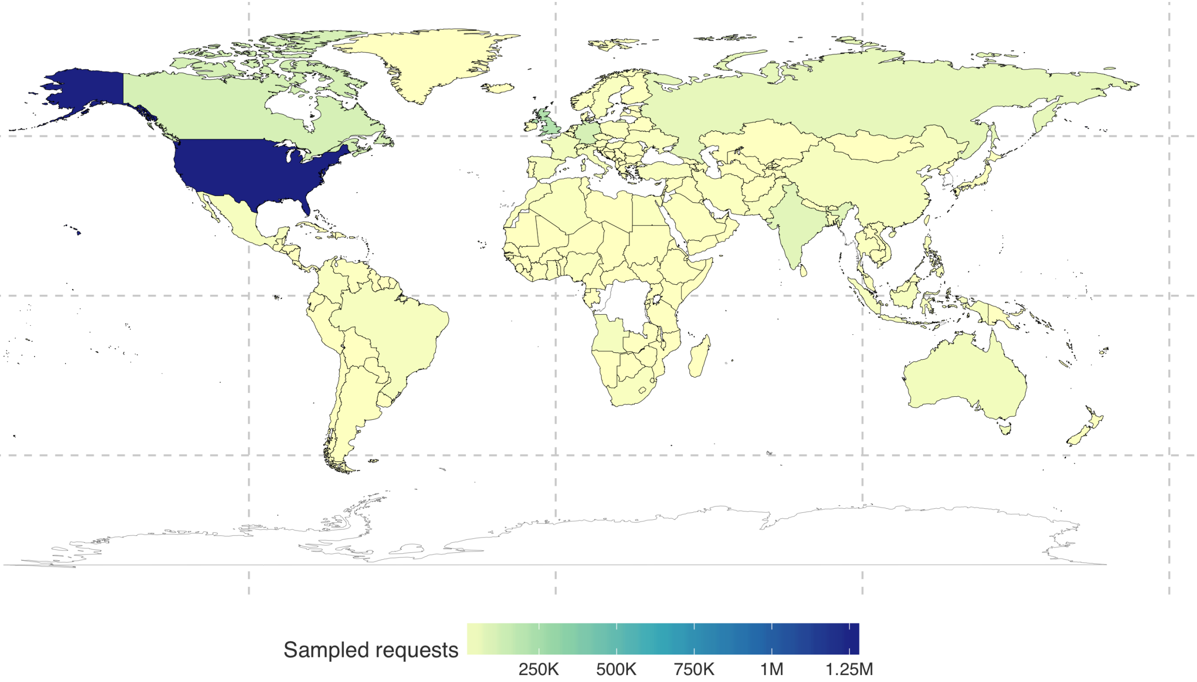

JS support (from countries with >100 WP Wordmark requests)
% of requests with JS support:
50%
60%
70%
80%
90%
100%

Figure 4: Approximate JS support across all 235 countries. The estimate is the median of 5 days of data.

Sources of sampled requests (median of 5 days)



U.S. accounted for 45% of the traffic to Wikipedia Portal in this dataset, with U.K. coming in second place with 8%, so it makes sense to look at the map excluding U.S. traffic:

Sources of sampled requests (median of 5 days, excluding U.S.)

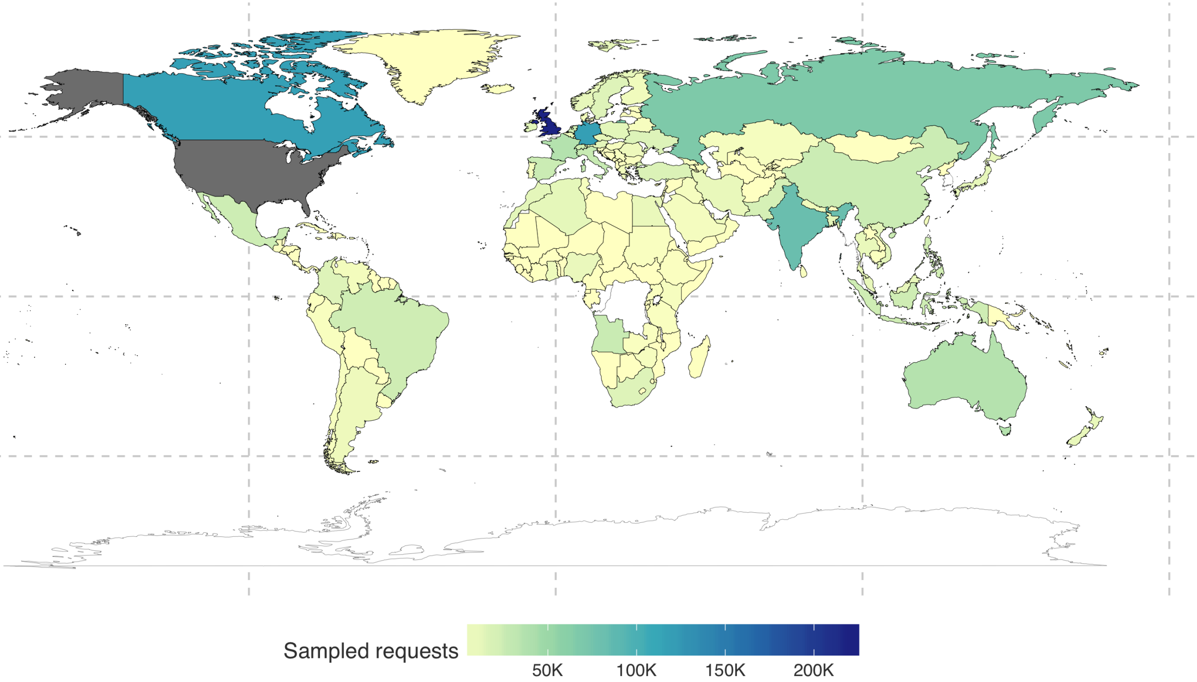


Table 2: Top 20 countries by % of traffic and the corresponding approximate percentage of requests ('users') with JS support.

country	Wordmark Rqsts	JS Rqsts	% of Total	Approx. % with JS support (JS/Wordmark)
United States	1.26M	1.22M	45.41%	96.15%
United Kingdom	218.67K	220.41K	7.86%	100.00%
Germany	118.16K	110.5K	4.25%	93.52%
Canada	116.68K	114.08K	4.19%	97.77%
India	81.98K	64.59K	2.95%	78.79%
Russia	69.02K	63.16K	2.48%	91.51%
France	37.47K	35.5K	1.35%	94.72%
Australia	37.05K	35.51K	1.33%	95.84%
Italy	27.51K	24.56K	0.99%	89.26%
Angola	26.84K	20.78K	0.96%	77.42%
Spain	25.89K	23.75K	0.93%	91.75%
China	25.7K	23.67K	0.92%	92.08%
Mexico	24.58K	22.22K	0.88%	90.38%
Indonesia	24.07K	18.92K	0.87%	78.58%
Brazil	24.03K	22.37K	0.86%	93.11%
Netherlands	23.96K	22.57K	0.86%	94.17%
Switzerland	23.12K	22.23K	0.83%	96.13%
Philippines	21.89K	19.41K	0.79%	88.68%
Pakistan	20.32K	16.1K	0.73%	79.19%
Turkey	18.07K	13.04K	0.65%	72.13%

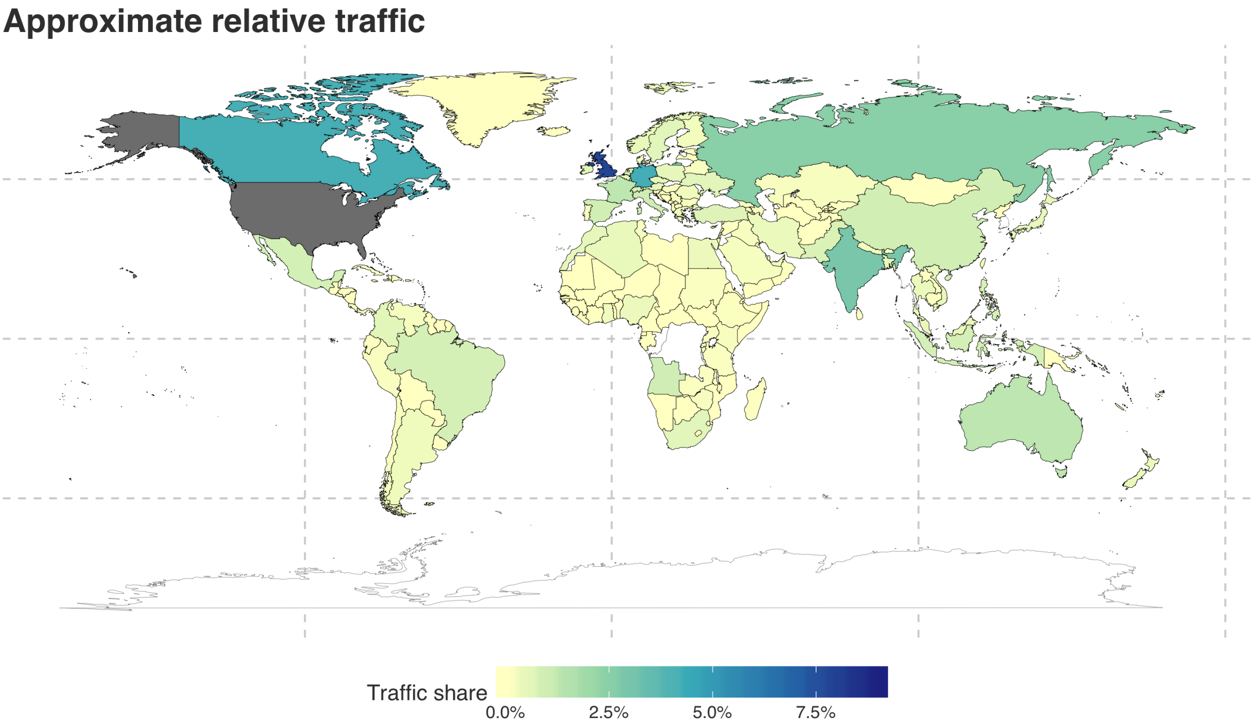


Figure 5: Median relative traffic across the world (sans U.S.) from 5 days of data.

Browsers and JS

As of 9 Feb 2016, the browsers in the table below account for at least 1% of daily traffic and together make up around 75% of the traffic. Furthermore, we can see in Fig. 6 that the proportion of JS support in some of the browsers is around 60% and 80%. We suspect this may be due to users turning off JavaScript in their browser settings for privacy and bandwidth reduction.

Browser	Median Wordmark Requests Per Day	Median Daily Share
Chrome 48	484.63K	17.32%
Mobile Safari 9	192.17K	6.87%
IE 11	191.31K	6.84%
Safari 9	183.58K	6.56%
Firefox 43	133.99K	4.79%
Android 4	127.12K	4.54%
Safari 8	118.32K	4.23%
Firefox 44	111.81K	4.00%
Chrome Mobile 30	98.32K	3.51%
IE 8	88.55K	3.16%
Firefox 24	87.41K	3.12%
Safari 5	76.77K	2.74%
Chrome Mobile 48	54.14K	1.93%
Opera Mini 7	39.19K	1.40%
Chrome 47	36.74K	1.31%
Mobile Safari 8	34.54K	1.23%
Edge 13	28.33K	1.01%

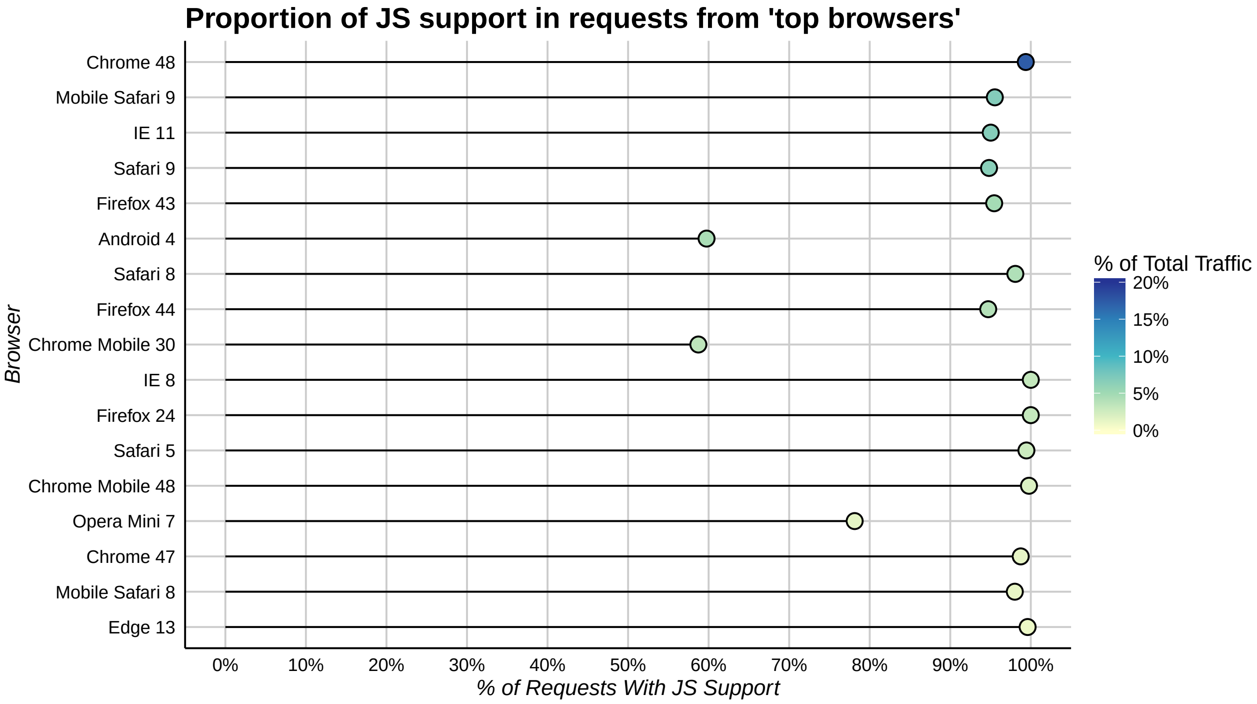
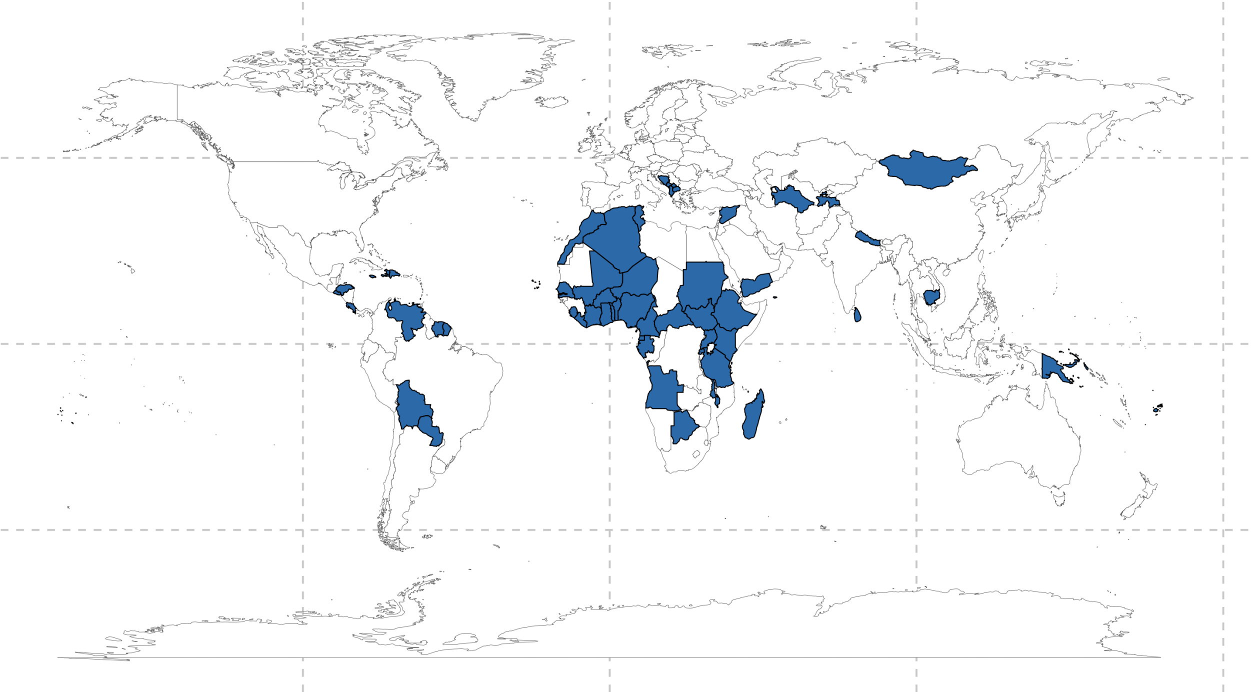


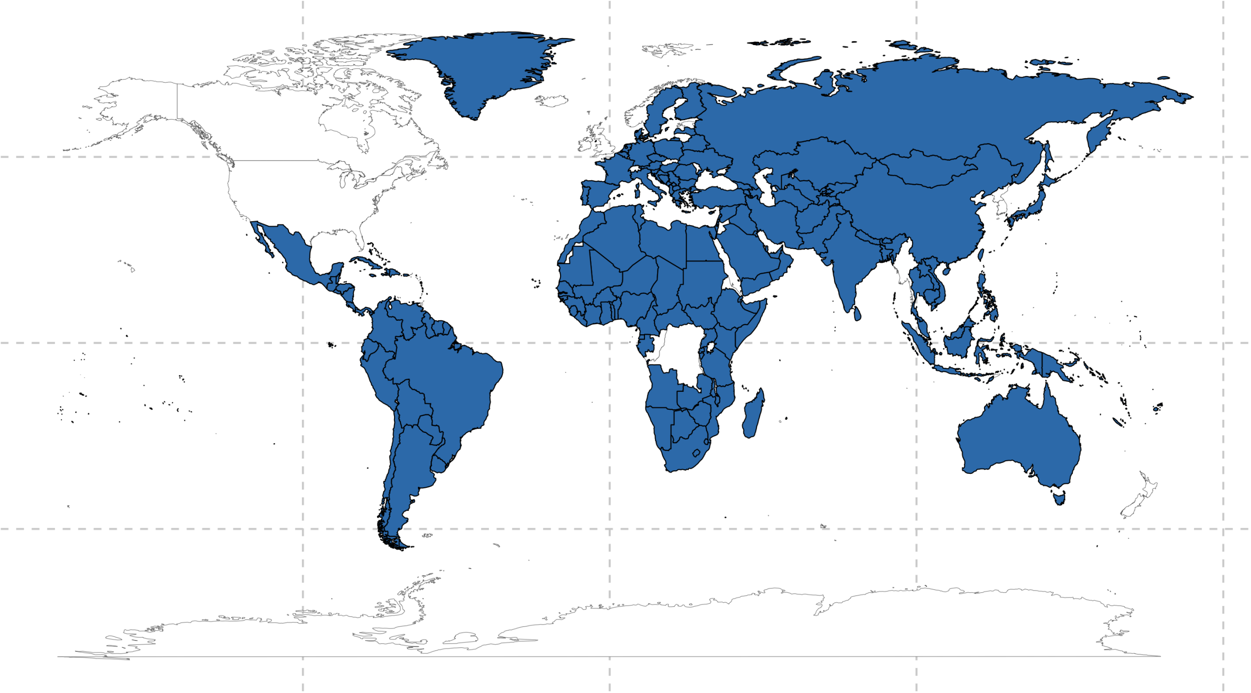
Figure 6: Proportion of JS requests from 'top browsers'.

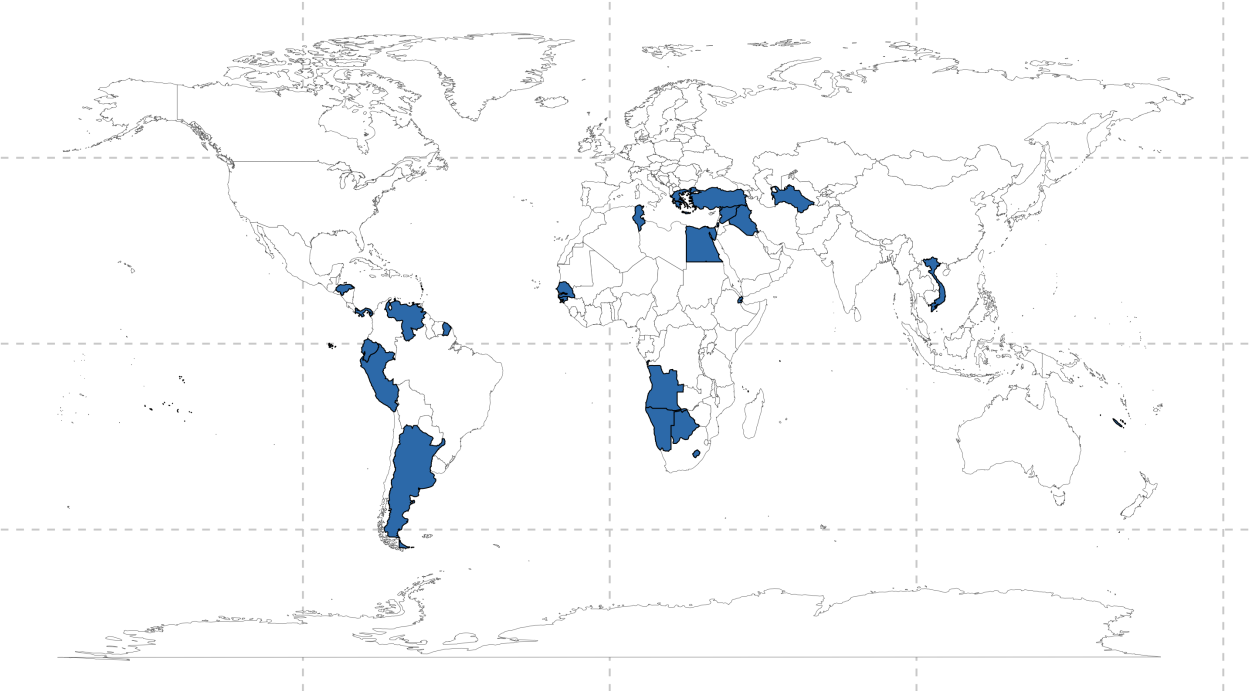
The remainder of the pages show the popularity of each of the "meta top N" browsers across the world. To clarify, the "meta top N" list was compiled by getting the top 10 browsers in each country and then getting the top N browsers by number of countries where it was in the top 10. We decided to show the popularity of $N = 16$ browsers.

67 countries where Android 2 is one of the top 10 browsers

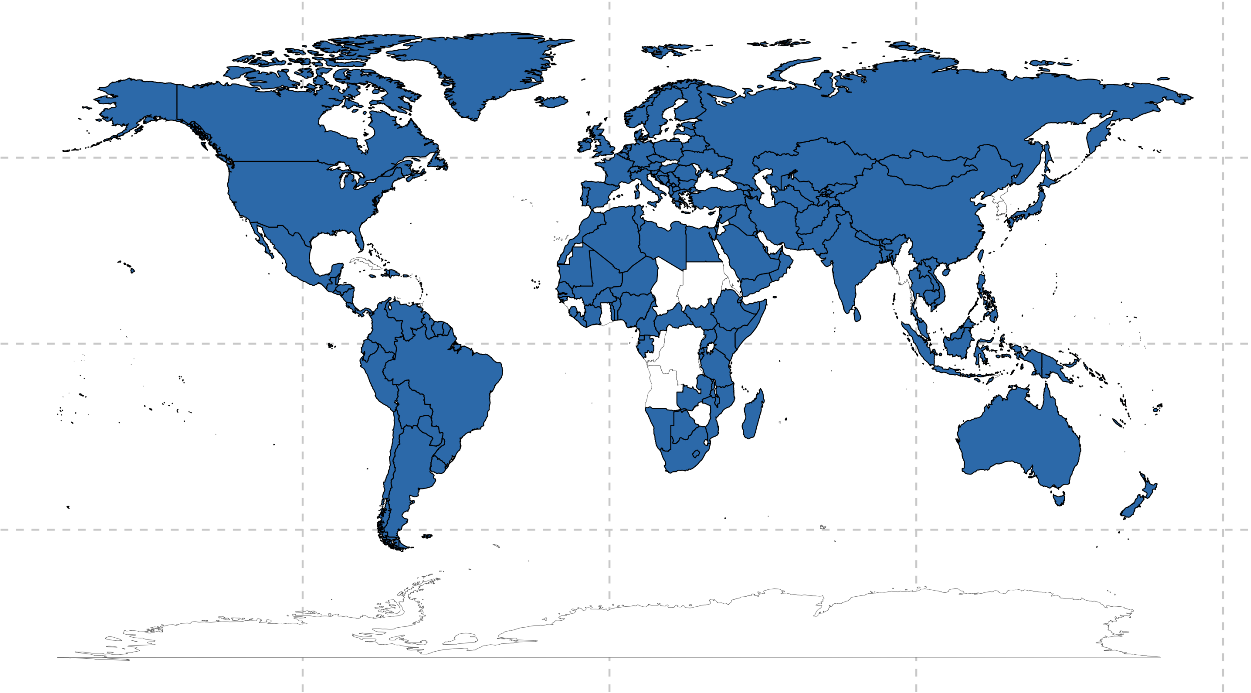


206 countries where Android 4 is one of the top 10 browsers



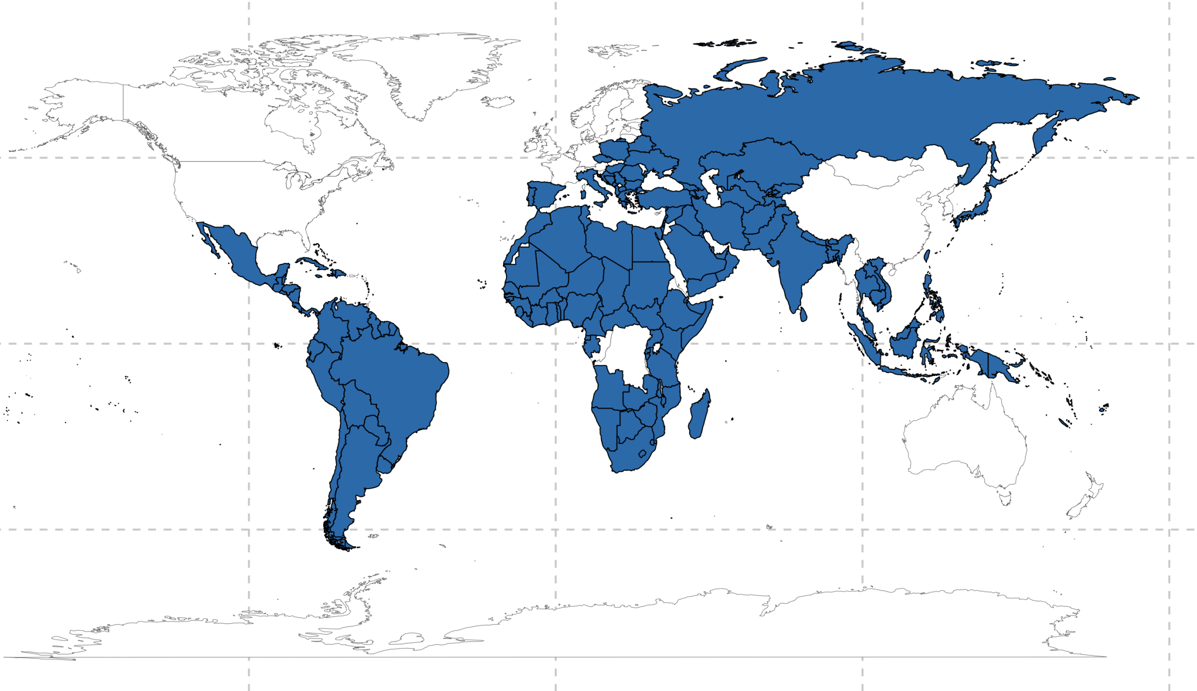


35 countries where Chrome 30 is one of the top 10 browsers

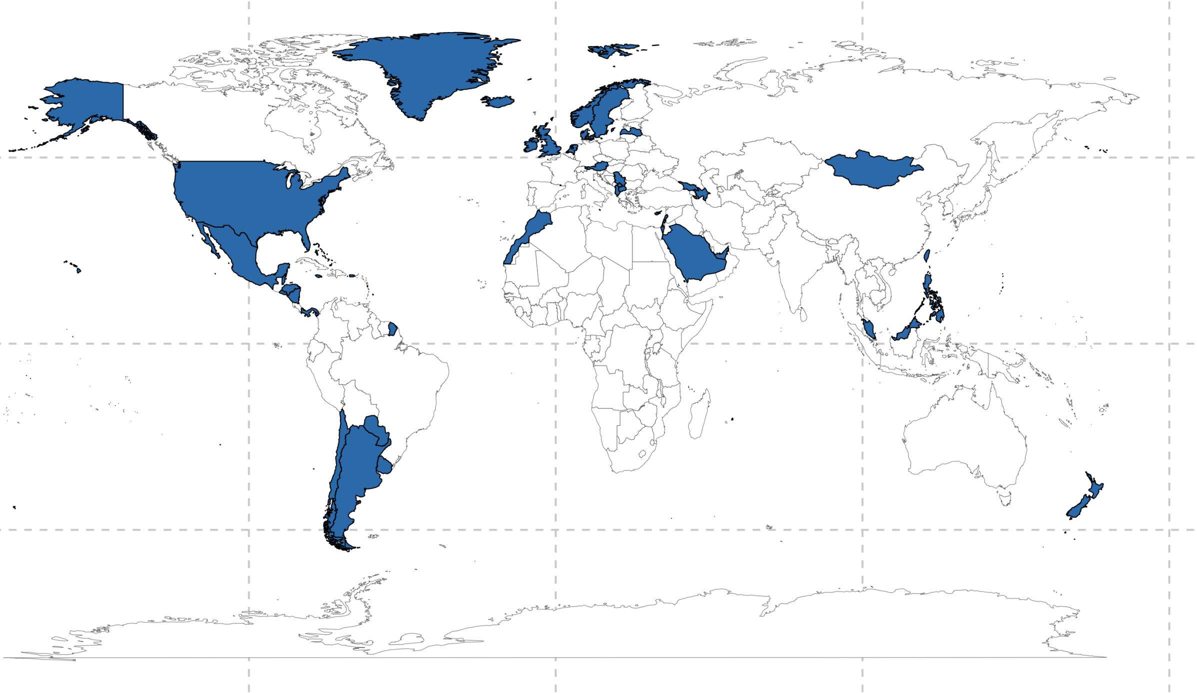


220 countries where Chrome 48 is one of the top 10 browsers

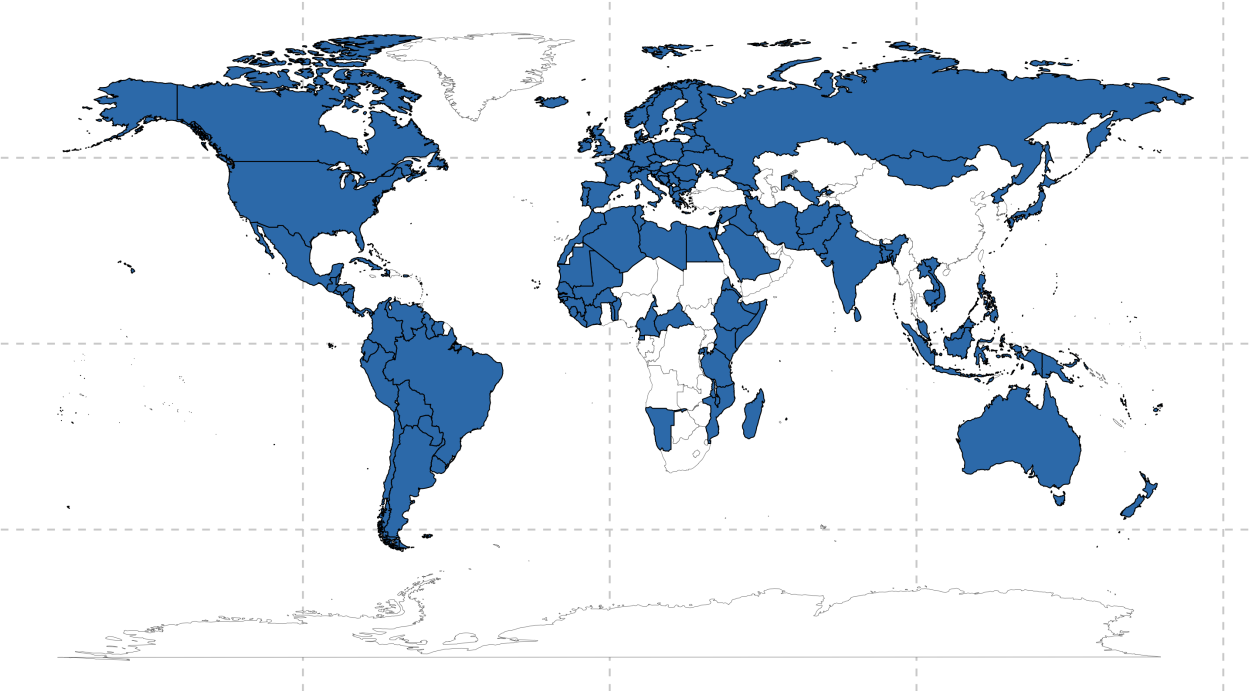
186 countries where Chrome Mobile 30 is one of the top 10 browsers



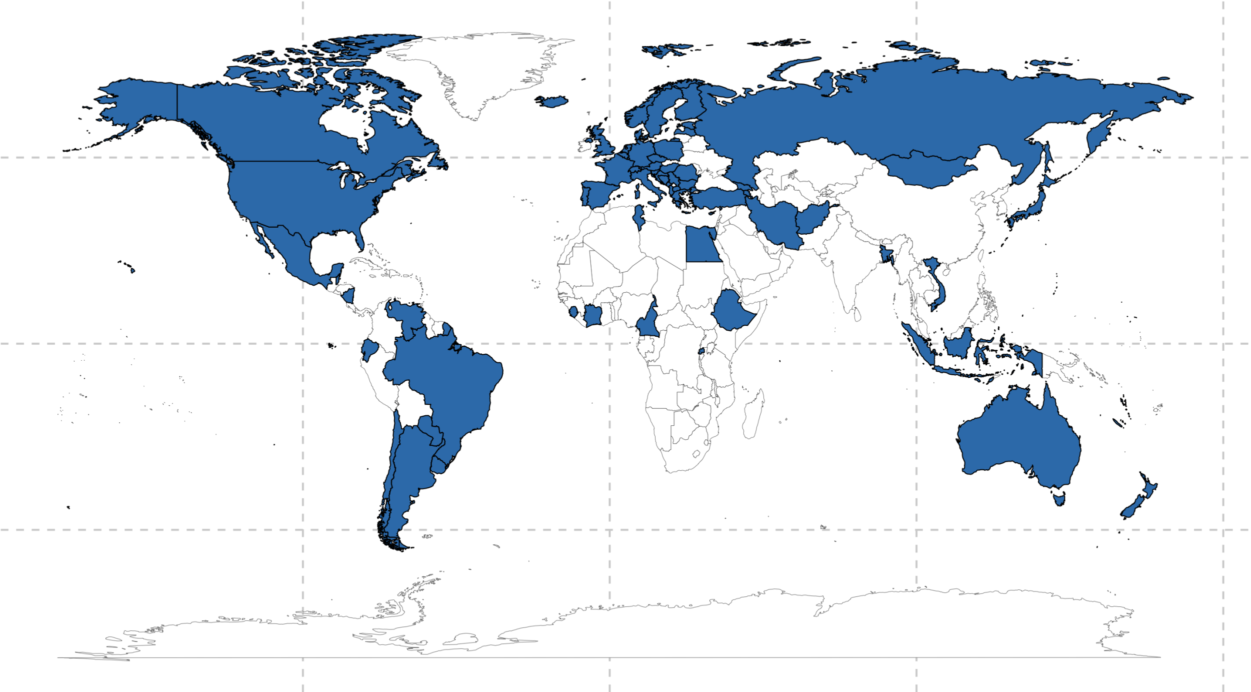
59 countries where Chrome Mobile 48 is one of the top 10 browsers



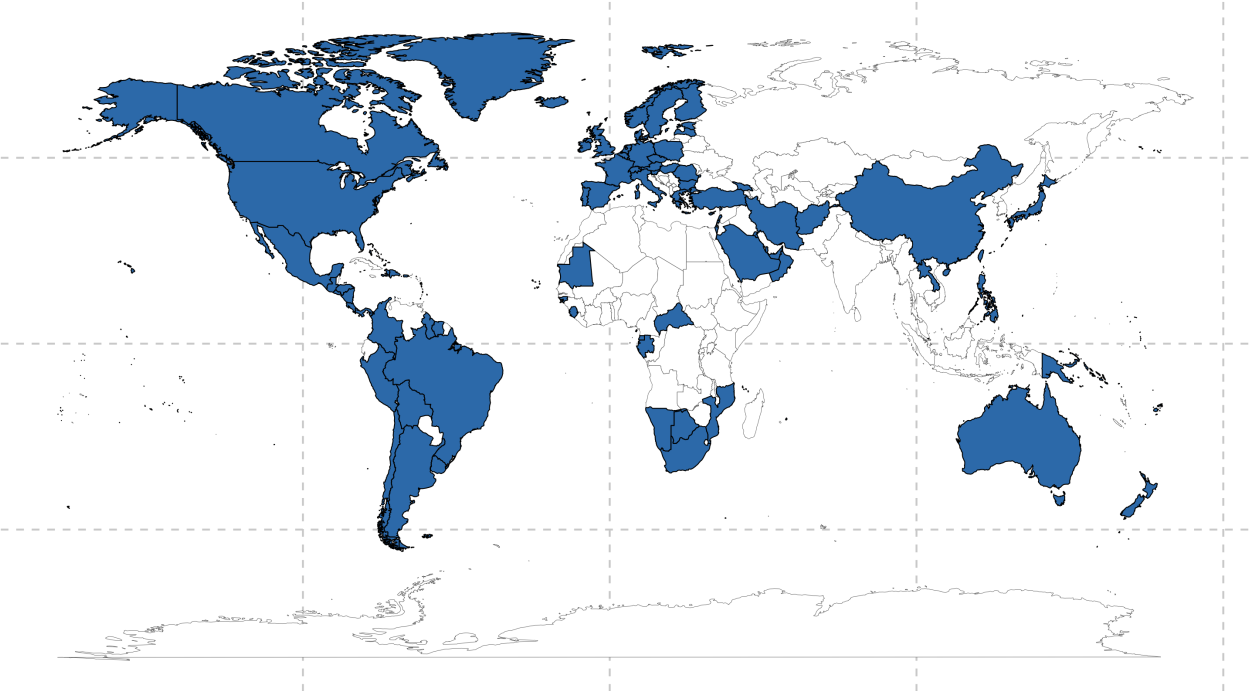
169 countries where Firefox 43 is one of the top 10 browsers



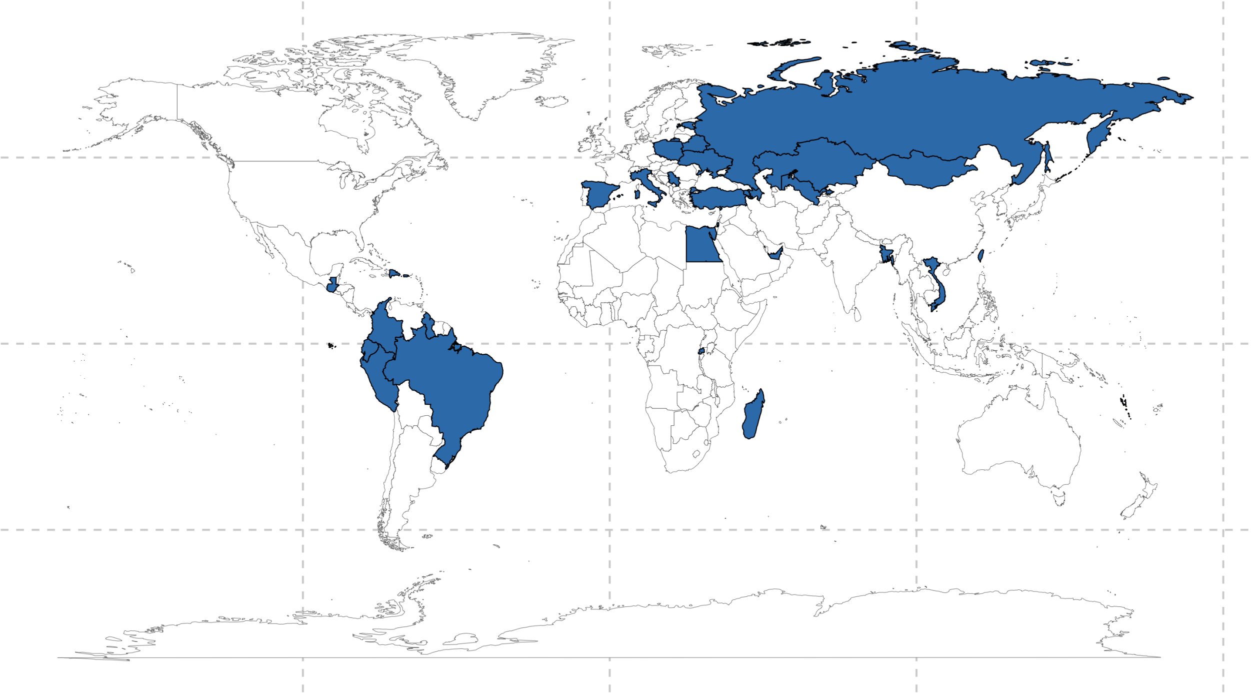
93 countries where Firefox 44 is one of the top 10 browsers



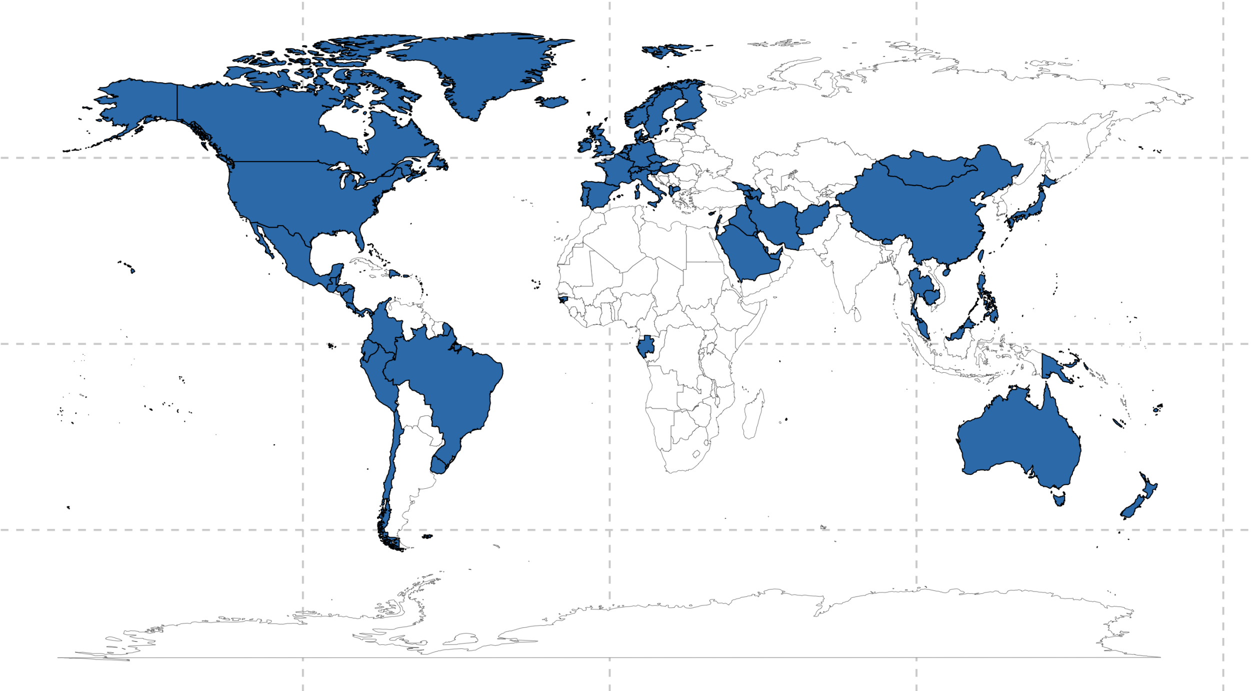
133 countries where IE 11 is one of the top 10 browsers



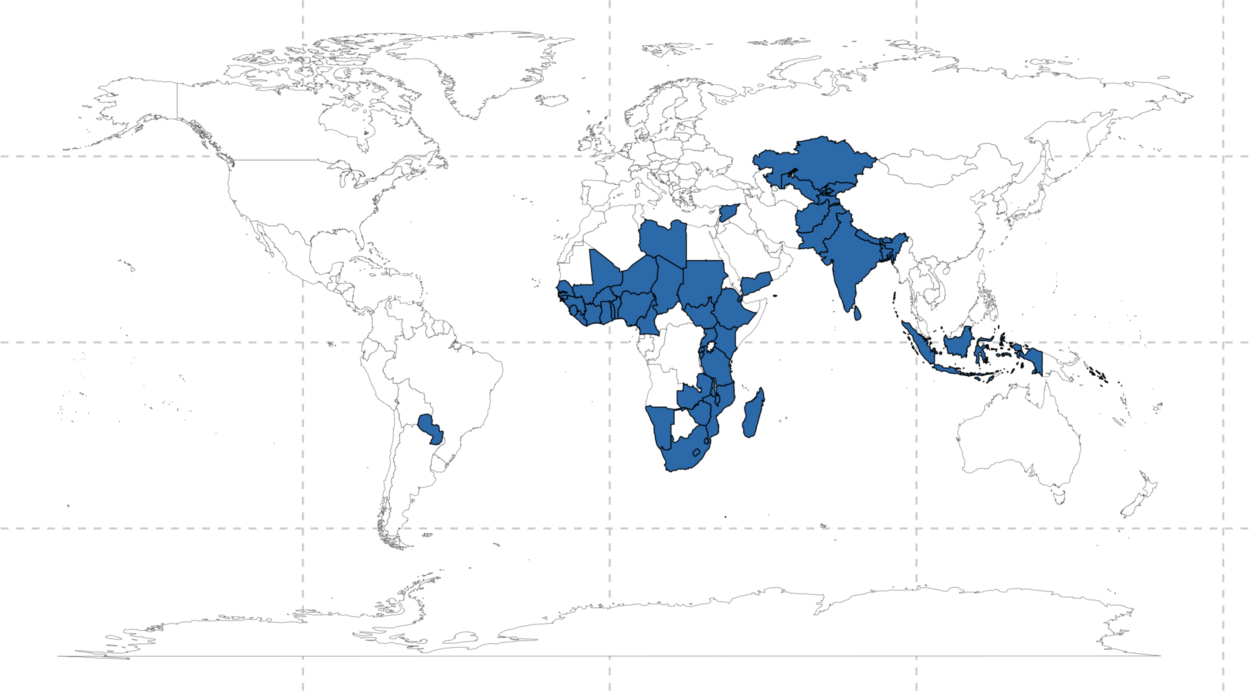
38 countries where IE 7 is one of the top 10 browsers



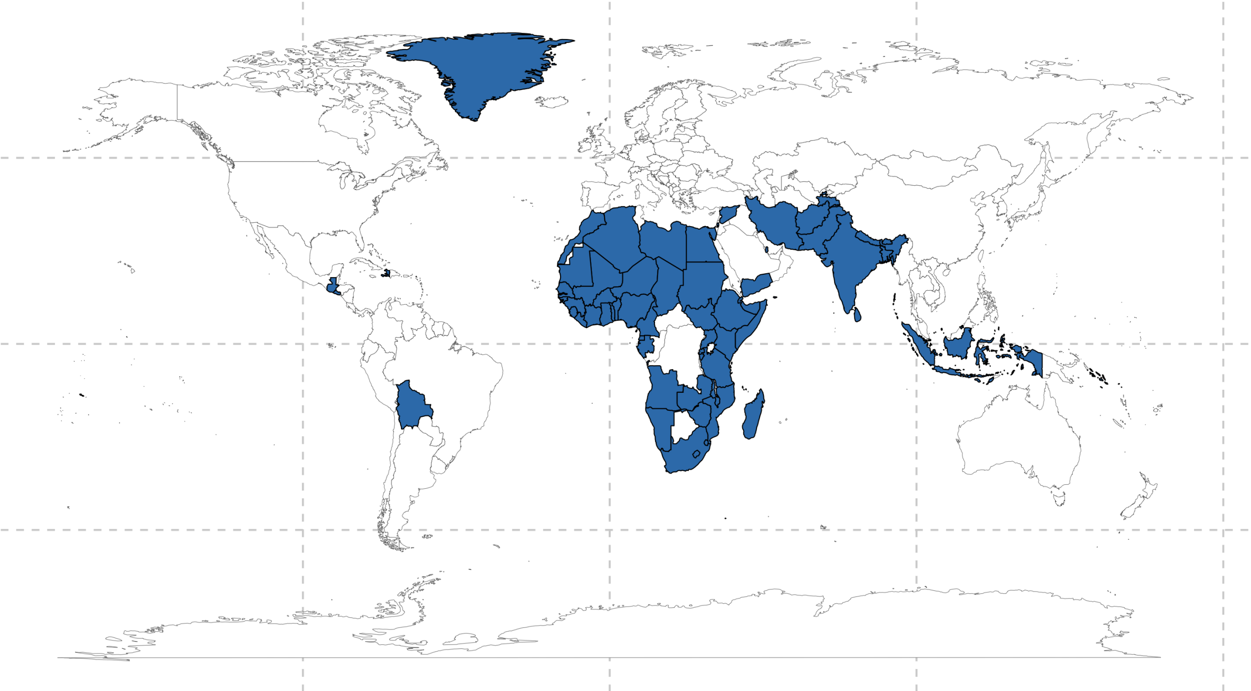
126 countries where Mobile Safari 9 is one of the top 10 browsers



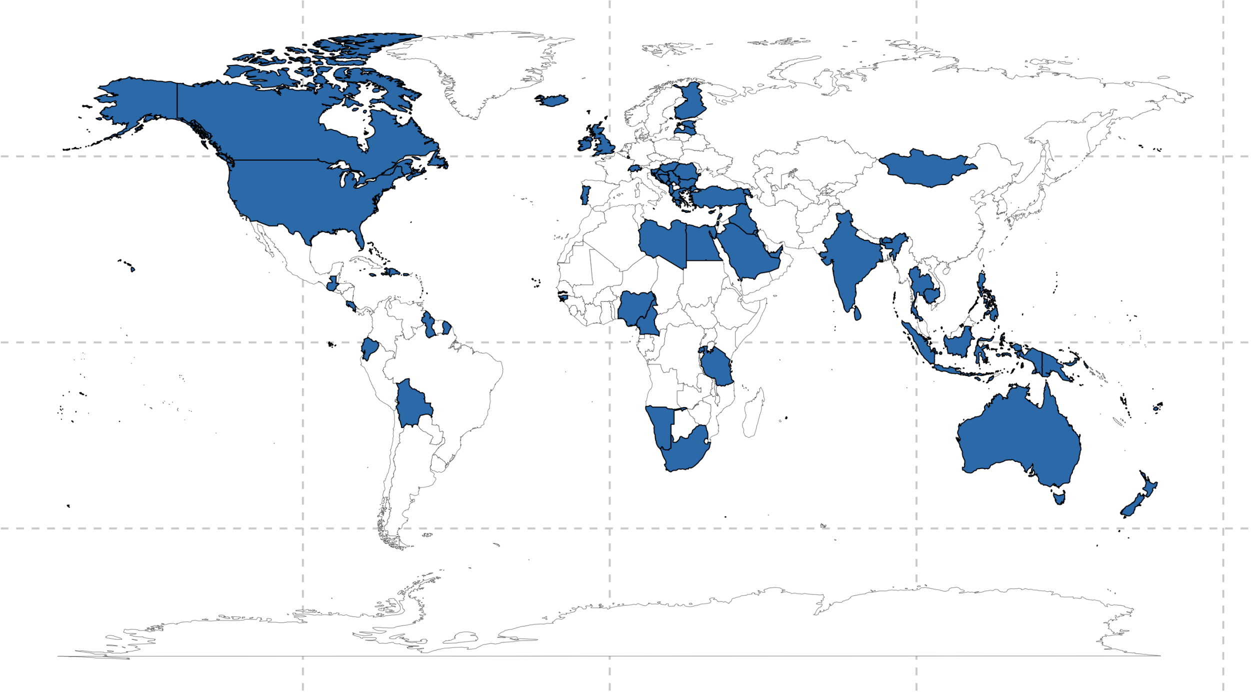
54 countries where Opera Mini 4 is one of the top 10 browsers



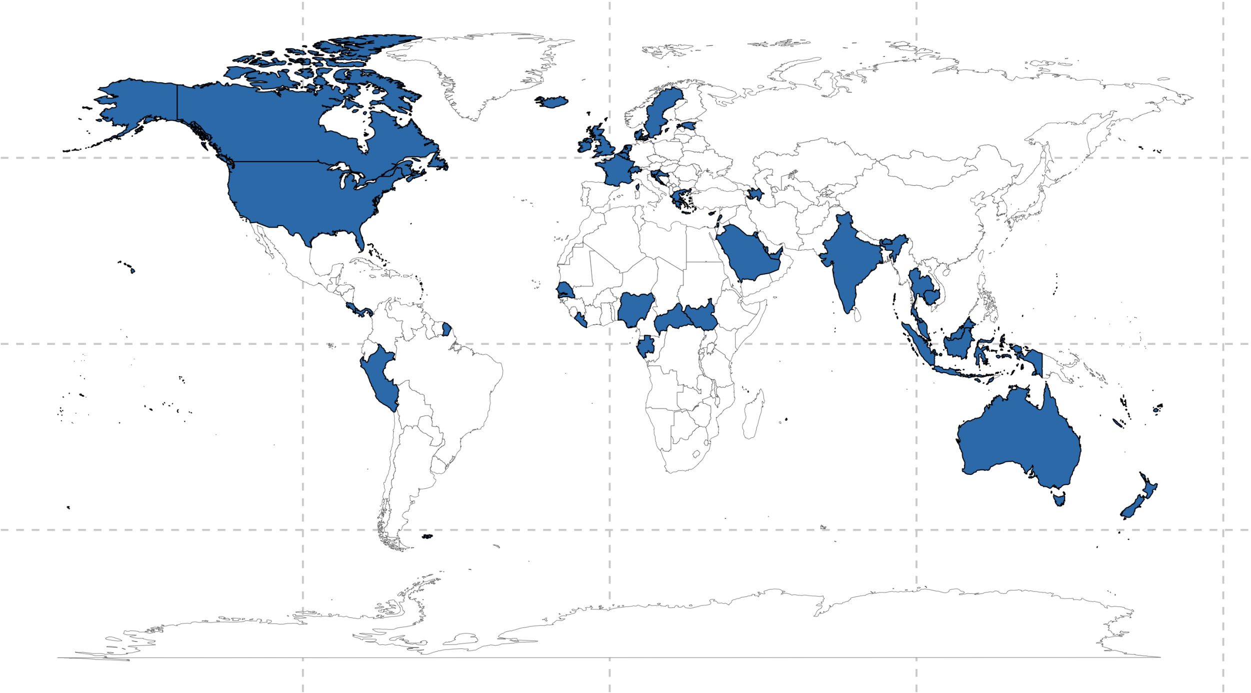
71 countries where Opera Mini 7 is one of the top 10 browsers



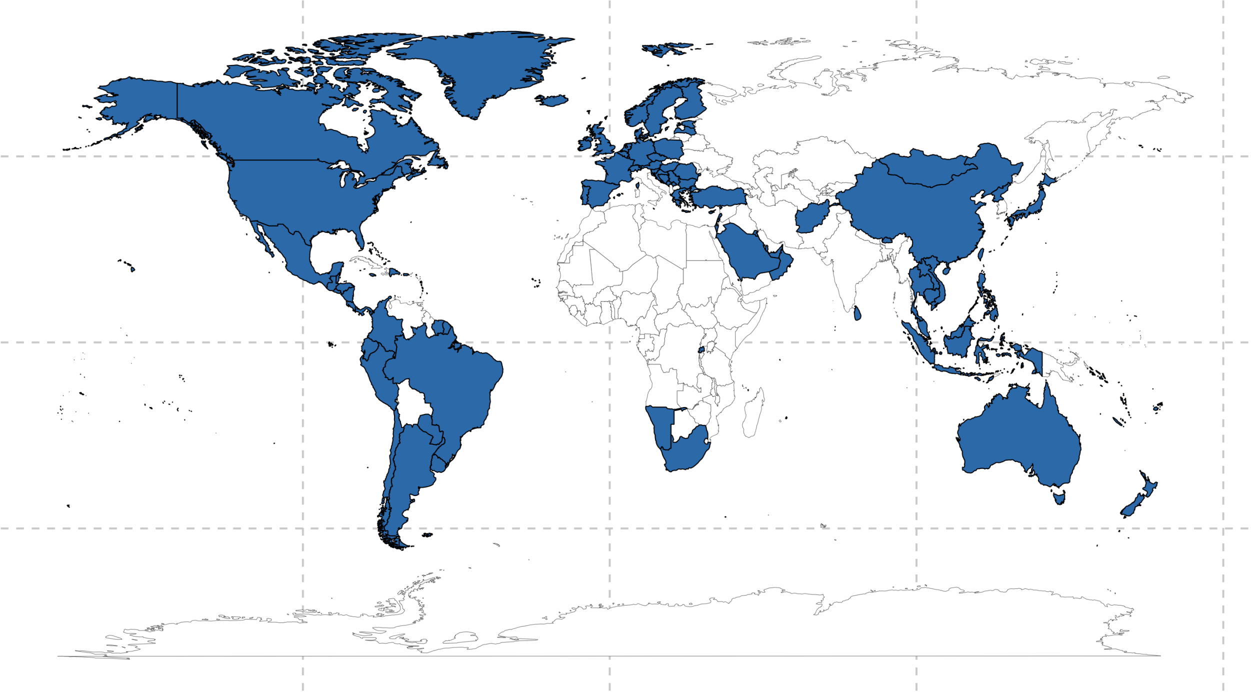
104 countries where Safari 5 is one of the top 10 browsers



87 countries where Safari 8 is one of the top 10 browsers



146 countries where Safari 9 is one of the top 10 browsers



Appendix

The following Hive query was used to aggregate the request logs:

```
ADD JAR /home/ebernhardson/refinery-hive-0.0.21-SNAPSHOT.jar;
CREATE TEMPORARY FUNCTION is_spider AS 'org.wikimedia.analytics.refinery.hive.IsSpiderUDF';
CREATE TEMPORARY FUNCTION is_wikimedia AS 'org.wikimedia.analytics.refinery.hive.IsWikimediaBotUDF';
CREATE TEMPORARY FUNCTION ua_parser AS 'org.wikimedia.analytics.refinery.hive.UAParserUDF';
USE wmf;
SELECT date, country_code, country, browser, file, COUNT(1) AS requests FROM (
  SELECT
    TO_DATE(ts) AS date,
    geocoded_data['country_code'] AS country_code,
    geocoded_data['country'] AS country,
    CONCAT(ua_parser(user_agent)['browser_family'], ' ',
           ua_parser(user_agent)['browser_major']) AS browser,
    CASE
      WHEN INSTR(uri_path, '/portal/wikipedia.org/assets/js/index-') > 0 THEN 'index_js'
      ELSE 'wikipedia_wordmark' END
    AS file
  FROM webrequest
  WHERE webrequest_source IN('text', 'mobile')
   AND year = 2016 AND month = 2 AND day > 4
   AND uri_host IN('wikipedia.org', 'www.wikipedia.org')
   AND (
     INSTR(uri_path, '/portal/wikipedia.org/assets/js/index-') > 0
     OR
     INSTR(uri_path, '/portal/wikipedia.org/assets/img/Wikipedia_wordmark') > 0
   )
   AND NOT is_spider(user_agent) AND NOT is_wikimedia(user_agent)
   AND NOT (ua_parser(user_agent)['device_family'] = 'Spider')
   AND http_status IN('200', '202', '301', '302', '304')
   AND NOT(geocoded_data['country'] = 'Unknown')
 ) AS requests
 GROUP BY date, country_code, country, browser, file;
```