

Wikipedia Portal Traffic and JavaScript Support

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

February 10, 2016

Daily Downloads of Wordmark and JS File

To identify what proportion of Wikipedia Portal requests lack JavaScript (JS) 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 may or may not make a request for the JS file depending on their settings and JS support, but their browser will definitely make a request for the JS 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.

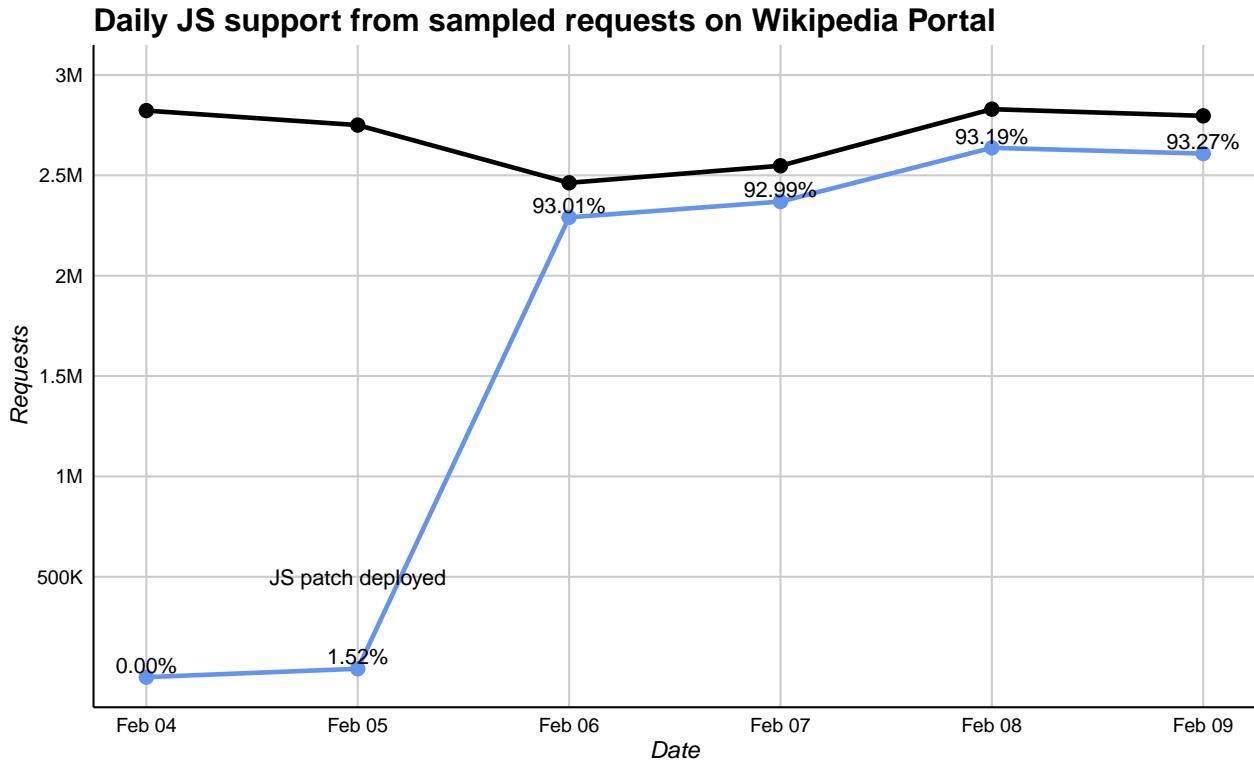
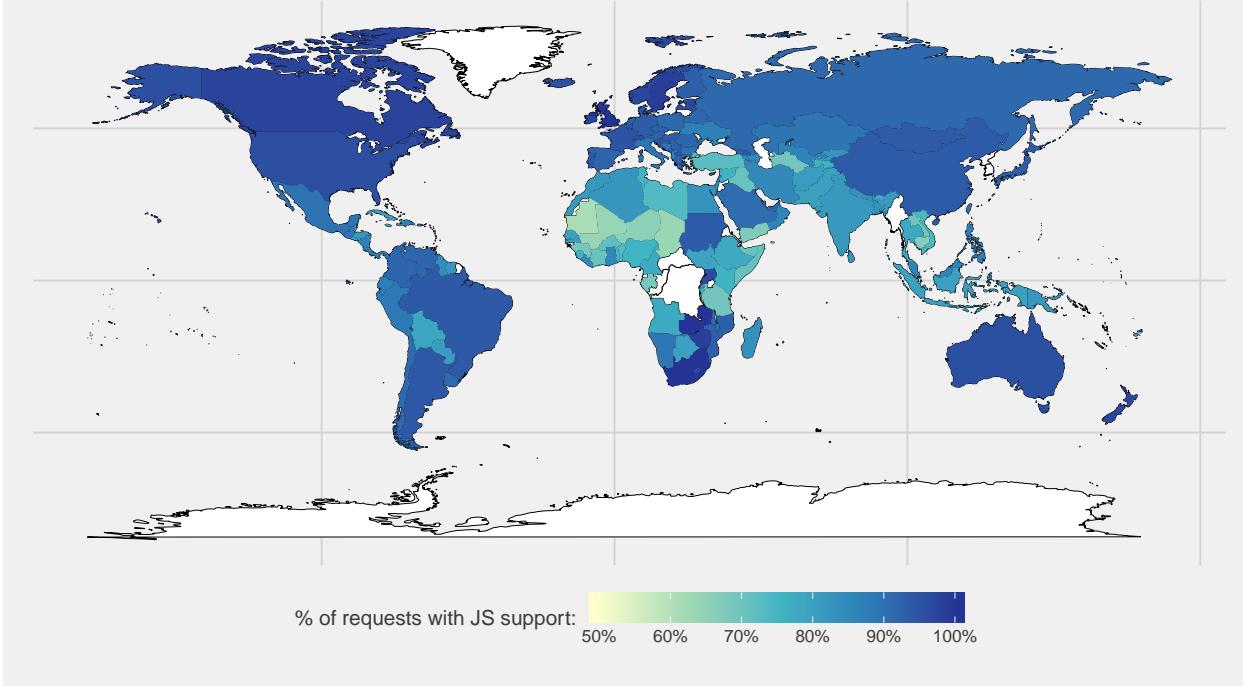


Figure 1 suggests that, overall, 93% of Wikipedia Portal requests support JS. In this report, we break that 93% down by countries and browsers. The following **Table 1** summarizes JS support across 235 countries:

Min %	25th Percentile	Average %	Median %	75th Percentile	Max %
50.00%	80.75%	86.36%	88.31%	93.66%	100.00%

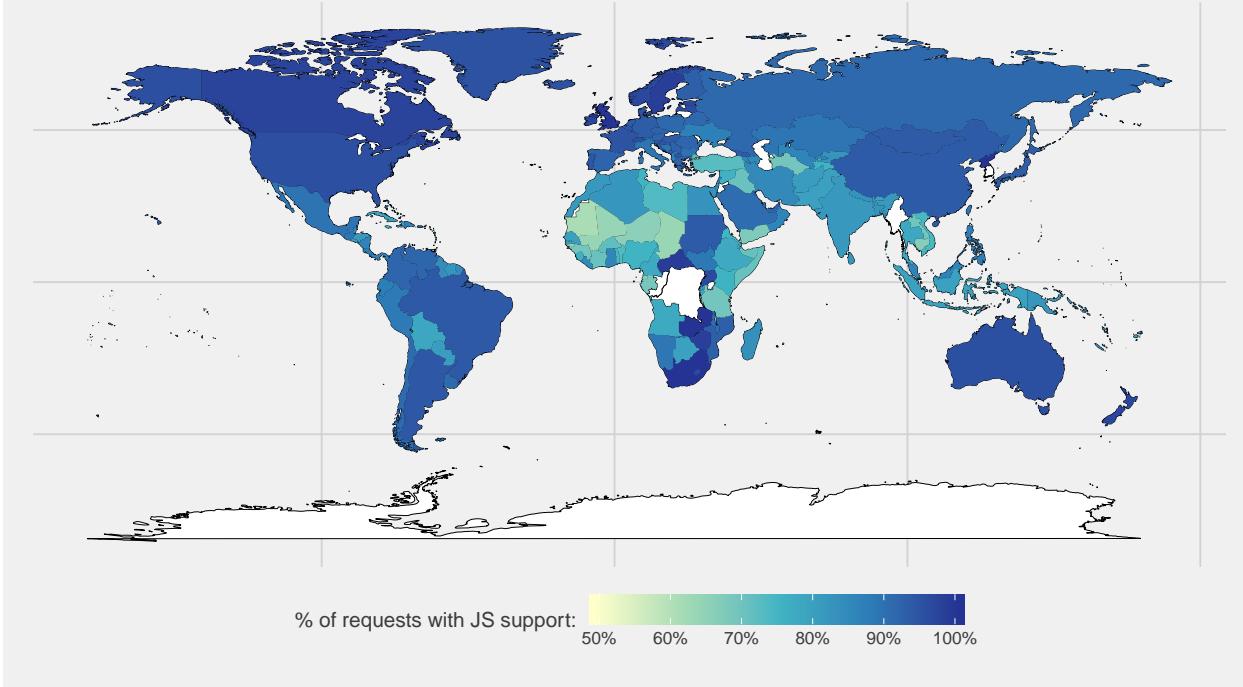
The median (88.31%) different than Fig. 1's 93% because in the figure we used the total of all requests, while the numbers in Tbl. 1 reflect a distribution of JS support in requests broken up by countries.

JS support (from countries with >100 WP Wordmark requests)

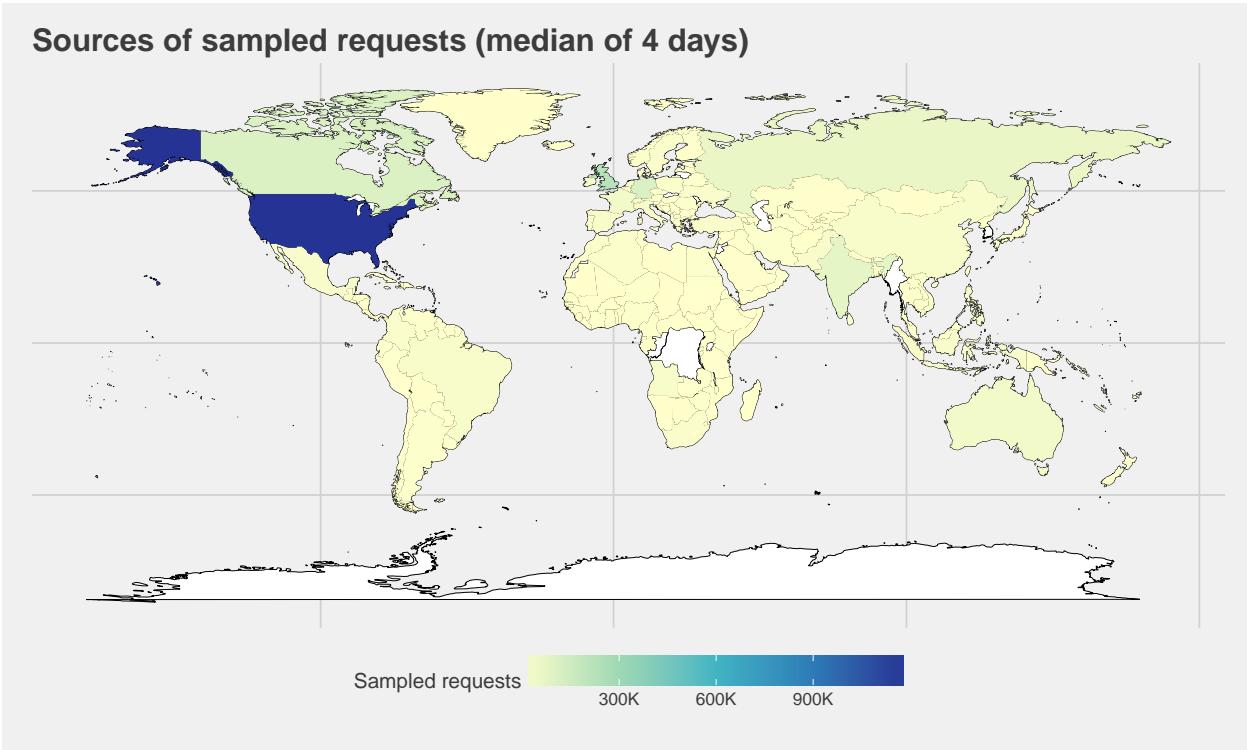


For the *choropleth* – a map shaded according to a statistic – above, we excluded 199 countries with less than 100 sampled Wordmark requests to make the proportion estimates more reliable. For example, if we recorded 11 Wordmark requests and 10 or 11 JS requests, that yields 91% and 100% JS support, 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 provided below:

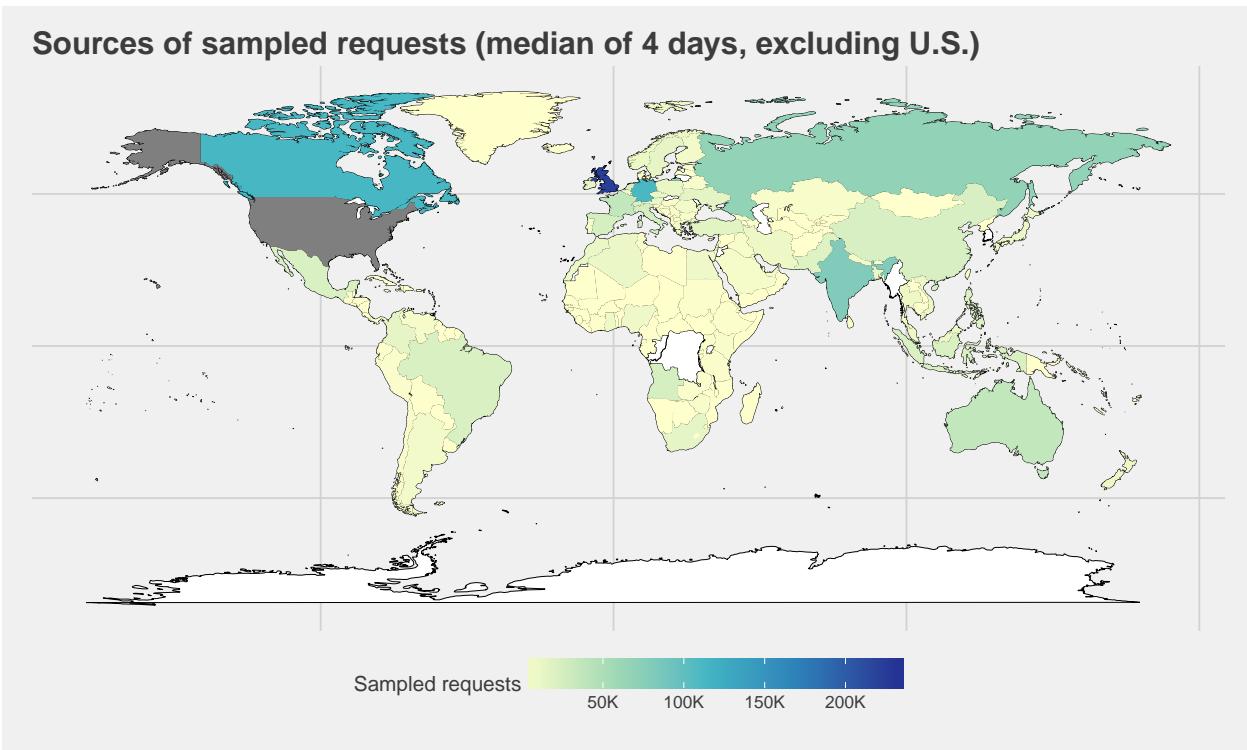
JS support (from countries with >100 WP Wordmark requests)



Worldwide Wikipedia Portal Traffic

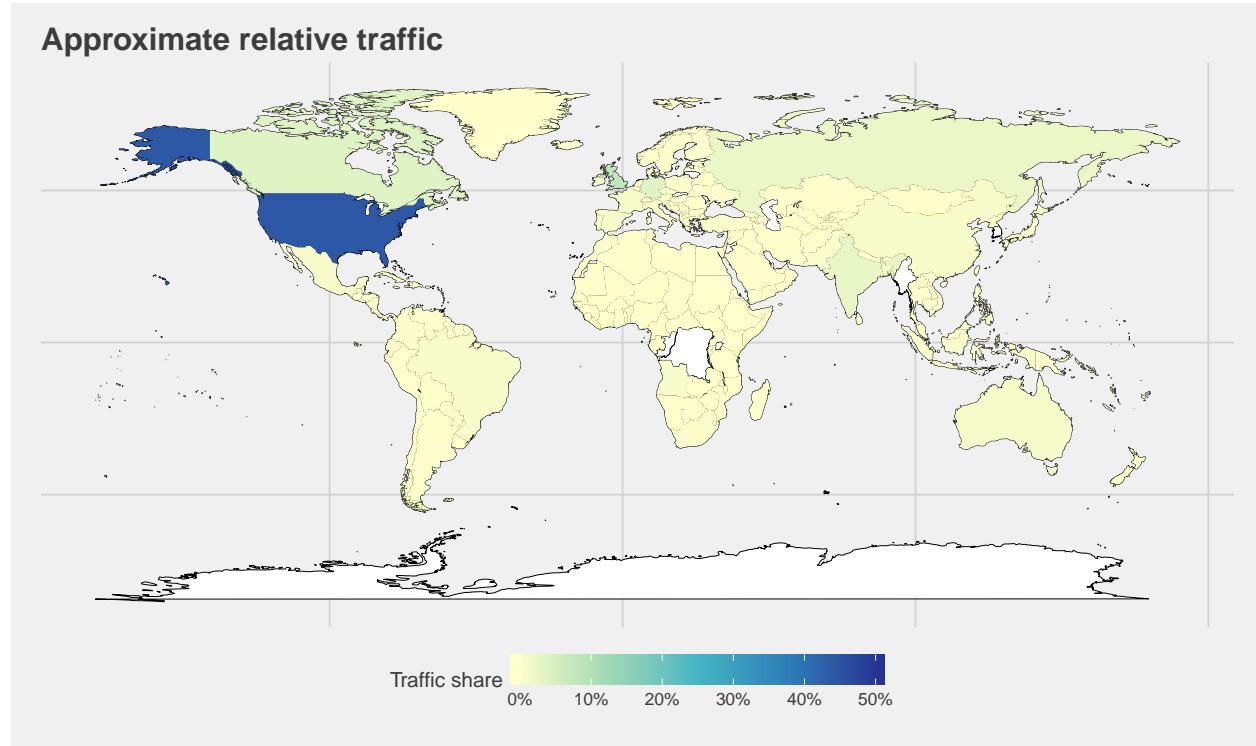


U.S. accounted for approximately 44% of the total traffic to the portal, with U.K. coming in second at 8%, so it also makes sense to look at a version of this map excluding U.S. traffic.



The following is a table showing the top 20 countries (by % of traffic) and the corresponding approximate percentage of users with JS support.

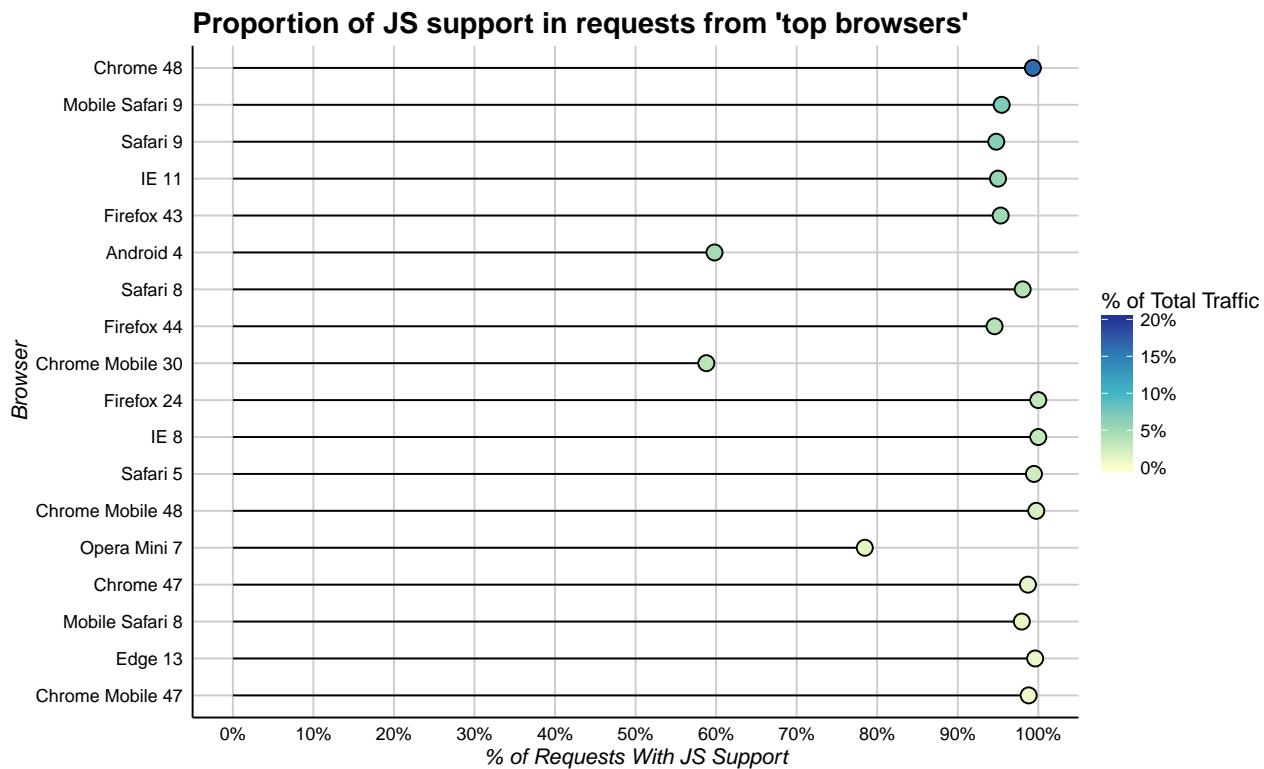
country	Wordmark DLs	JS DLs	% of Total	Approx. % with JS support (JS/Wordmark)
United States	1.18M	1.13M	44.06%	95.58%
United Kingdom	222.41K	222.97K	8.30%	100.00%
Canada	112.53K	109.57K	4.20%	97.37%
Germany	111.98K	104.29K	4.18%	93.13%
India	80.65K	64.33K	3.01%	79.77%
Russia	69.21K	63.55K	2.58%	91.82%
France	37.27K	35.35K	1.39%	94.85%
Australia	37.14K	35.54K	1.39%	95.69%
Italy	27.1K	24K	1.01%	88.57%
Angola	25.62K	19.75K	0.96%	77.12%
Spain	25.38K	23.32K	0.95%	91.89%
China	24.89K	23.04K	0.93%	92.56%
Brazil	23.89K	22.35K	0.89%	93.57%
Indonesia	23.68K	18.8K	0.88%	79.39%
Mexico	23.58K	21.09K	0.88%	89.42%
Switzerland	22.49K	21.52K	0.84%	95.67%
Netherlands	22.42K	20.91K	0.84%	93.28%
Philippines	21.57K	18.97K	0.81%	87.94%
Pakistan	20.31K	16.01K	0.76%	78.83%
Turkey	18.1K	13.07K	0.68%	72.20%



Browsers and JS

As of 9 Feb 2016, the following browsers account for at least 1% of daily traffic.

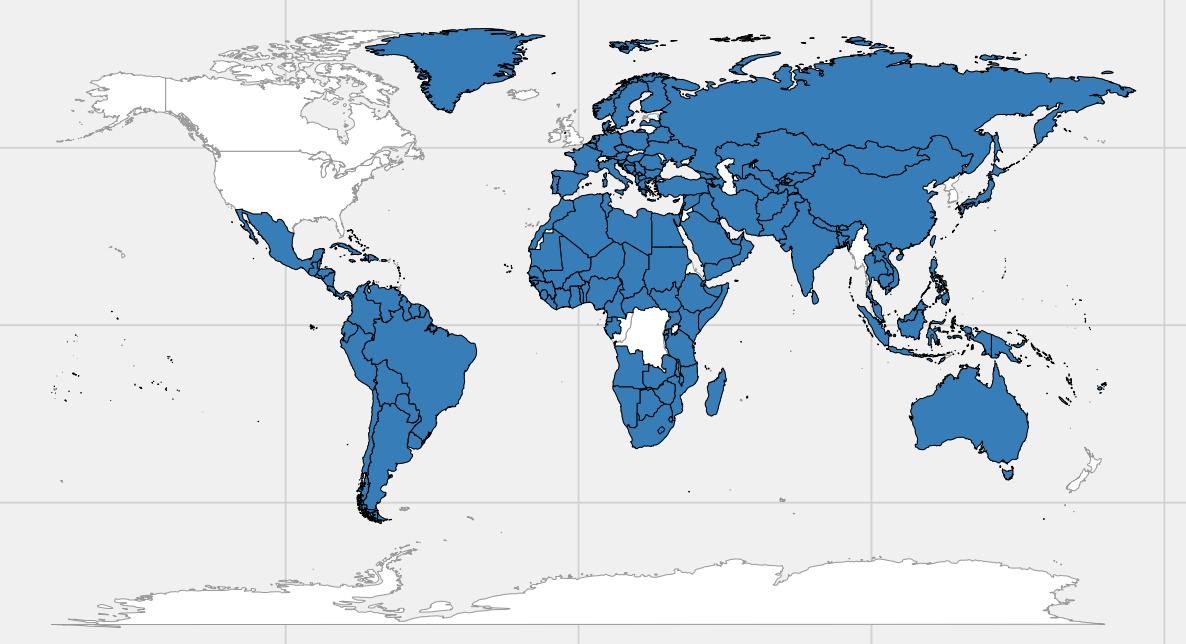
Browser	Median Wordmark Requests Per Day	Median Daily Share
Chrome 48	432.09K	16.26%
Mobile Safari 9	192.23K	7.23%
Safari 9	175.76K	6.61%
IE 11	151.95K	5.72%
Firefox 43	134.37K	5.06%
Android 4	125.52K	4.72%
Safari 8	108.86K	4.10%
Firefox 44	107.7K	4.05%
Chrome Mobile 30	97.76K	3.68%
Firefox 24	88.57K	3.33%
IE 8	87.54K	3.29%
Safari 5	71.2K	2.68%
Chrome Mobile 48	56.01K	2.11%
Opera Mini 7	40.29K	1.52%
Chrome 47	36.98K	1.39%
Mobile Safari 8	34.77K	1.31%
Edge 13	28.56K	1.07%
Chrome Mobile 47	26.71K	1.00%



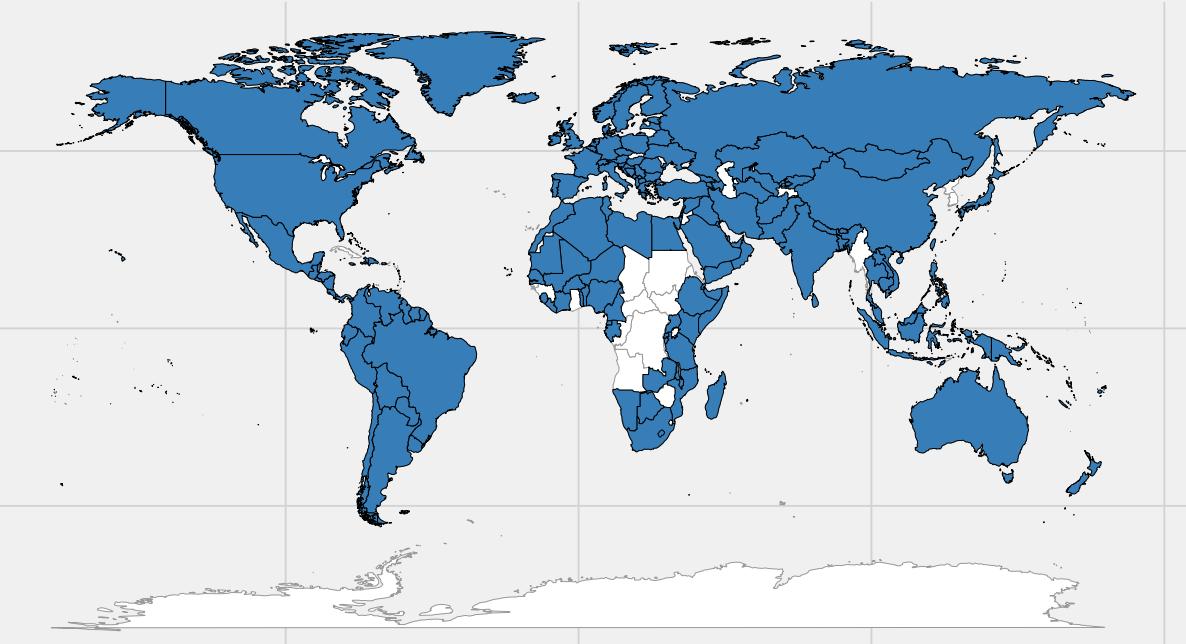
The reader might notice that the proportion of JS support in some of the browsers is around 60% and 80%. We suspect this may be due to users turning off JavaScript in their browser settings for privacy and bandwidth reduction.

The next few pages show the popularity of each of the “meta top 10” browsers across the world. To clarify, the “meta top 10” list was compiled by getting the top 10 browsers in each country and then getting the top 10 browsers by number of countries where it was in the top 10.

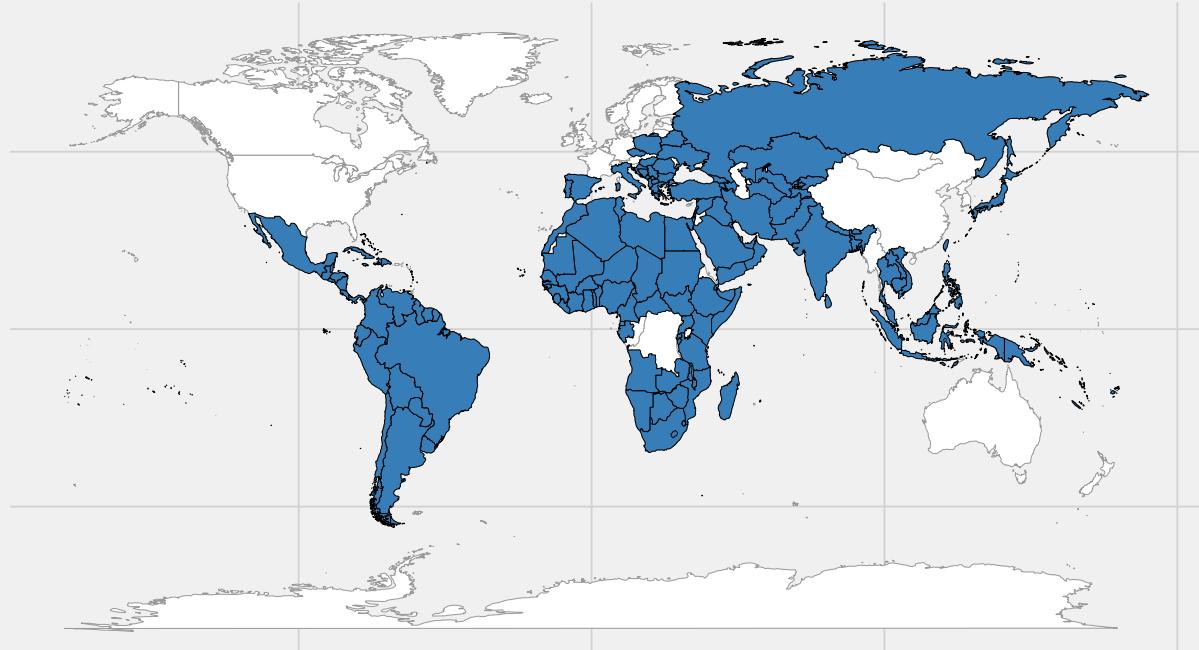
Countries where Android 4 is one of top 10 browsers



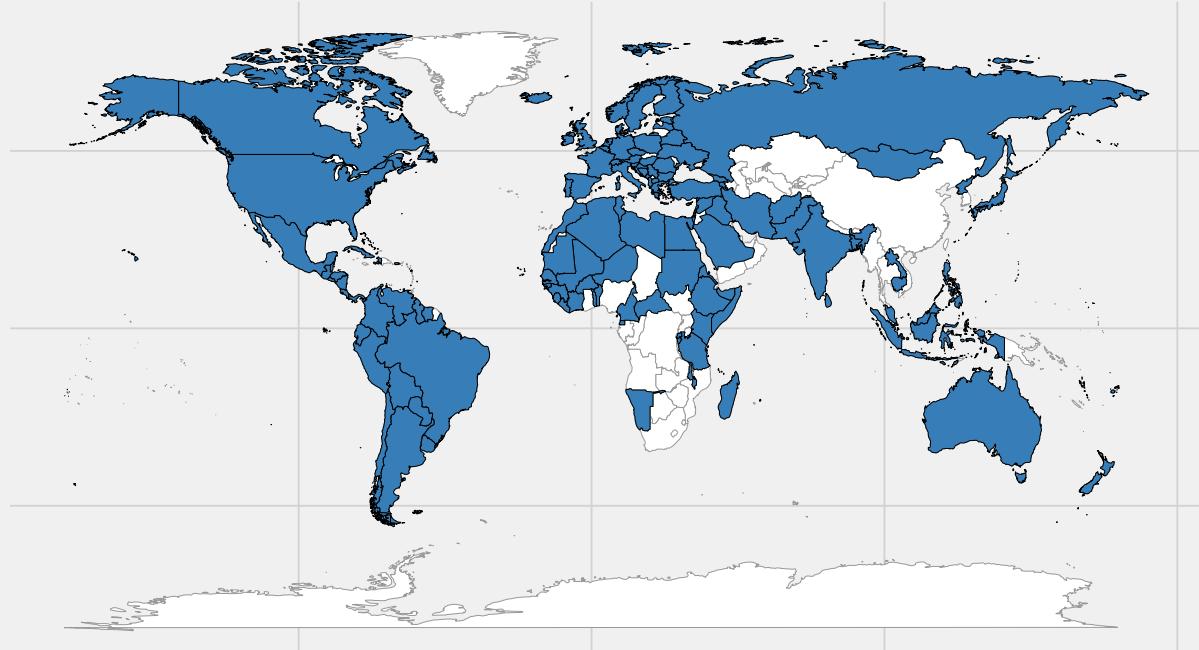
Countries where Chrome 48 is one of top 10 browsers



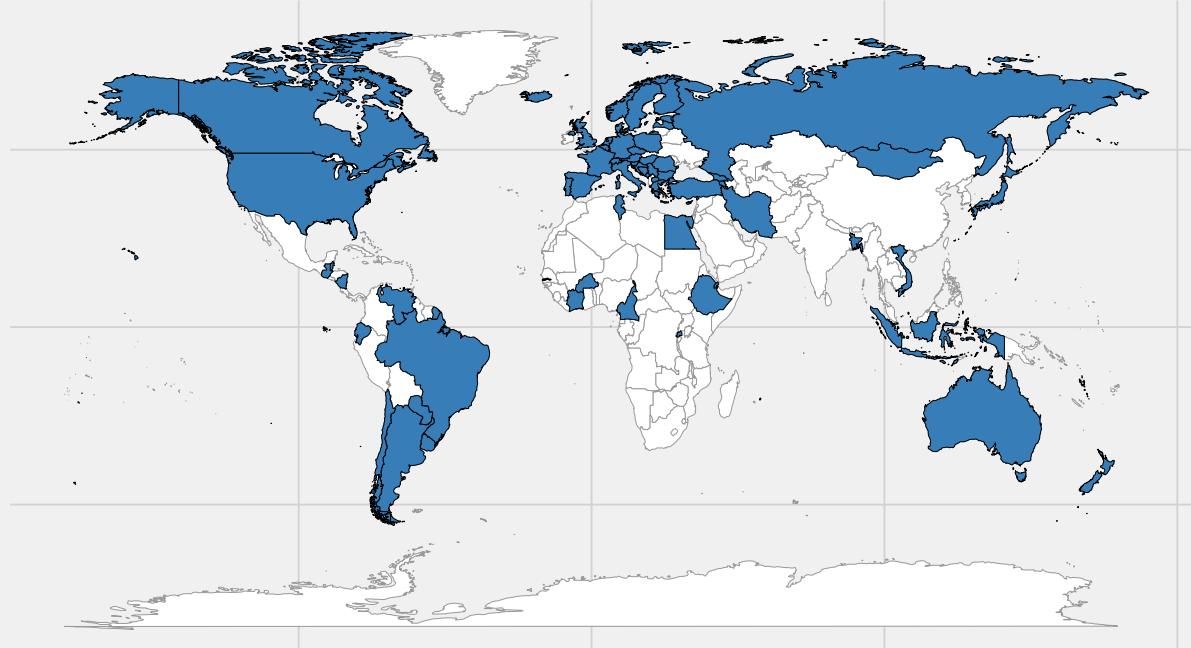
Countries where Chrome Mobile 30 is one of top 10 browsers



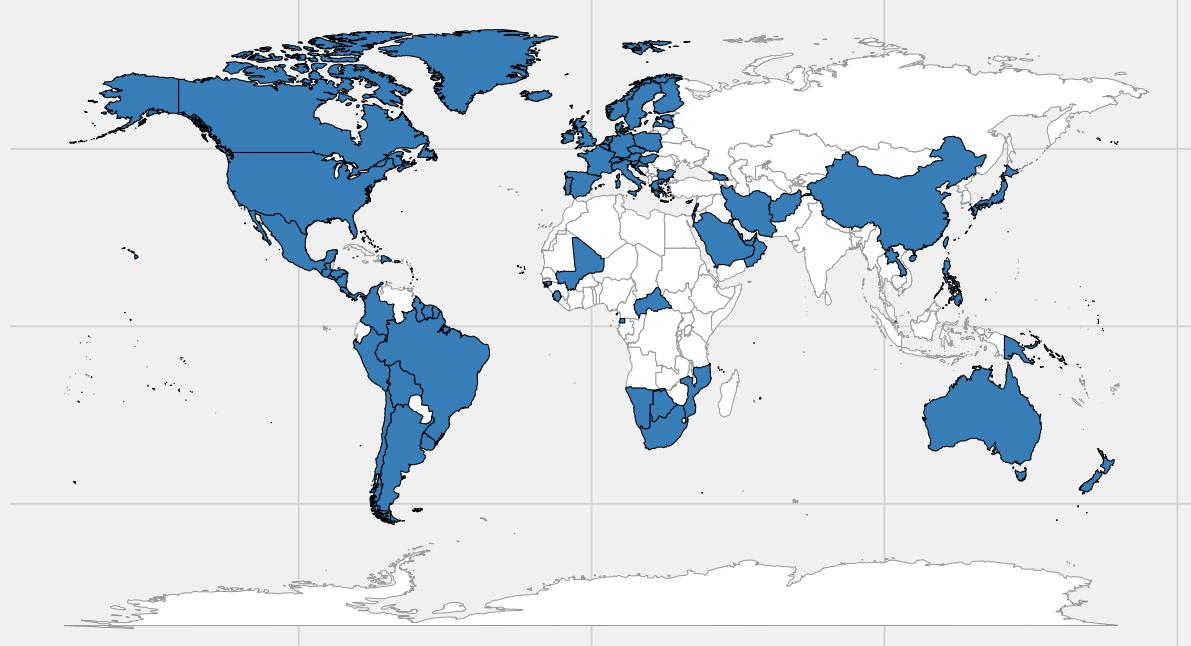
Countries where Firefox 43 is one of top 10 browsers



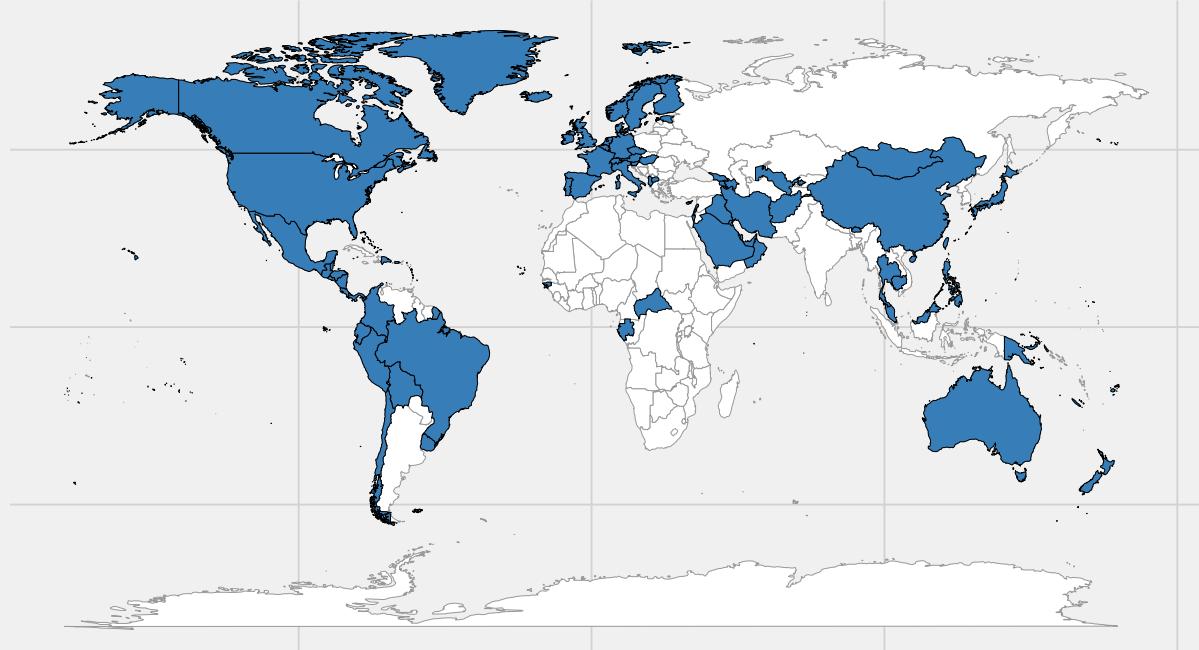
Countries where Firefox 44 is one of top 10 browsers



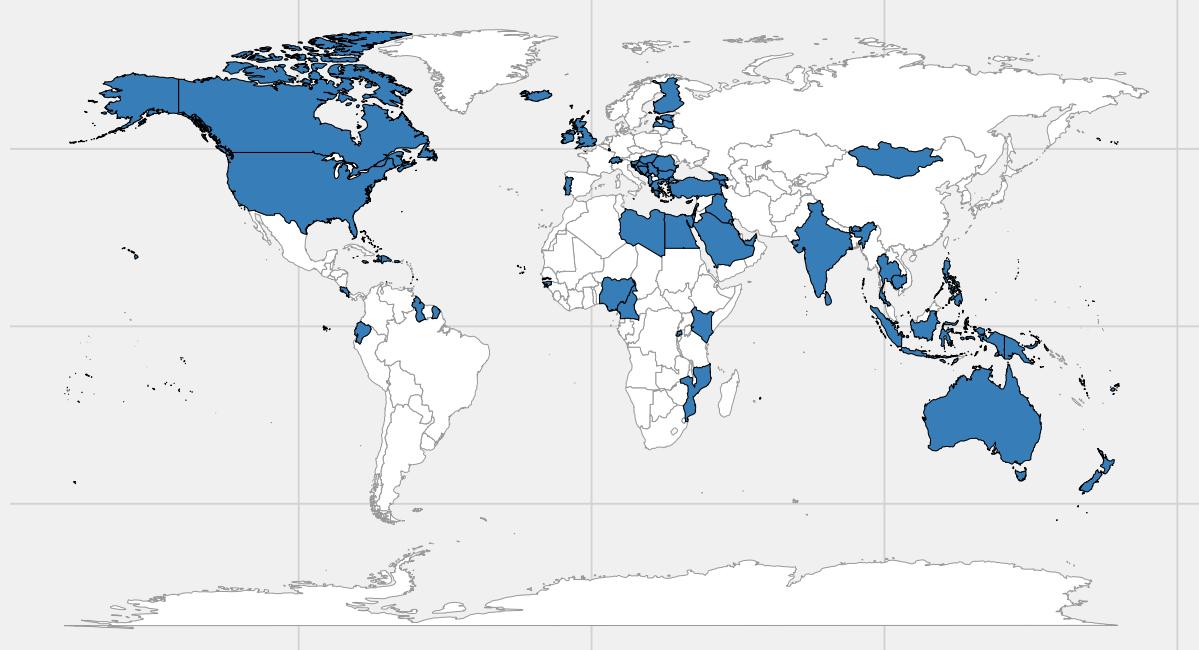
Countries where IE 11 is one of top 10 browsers



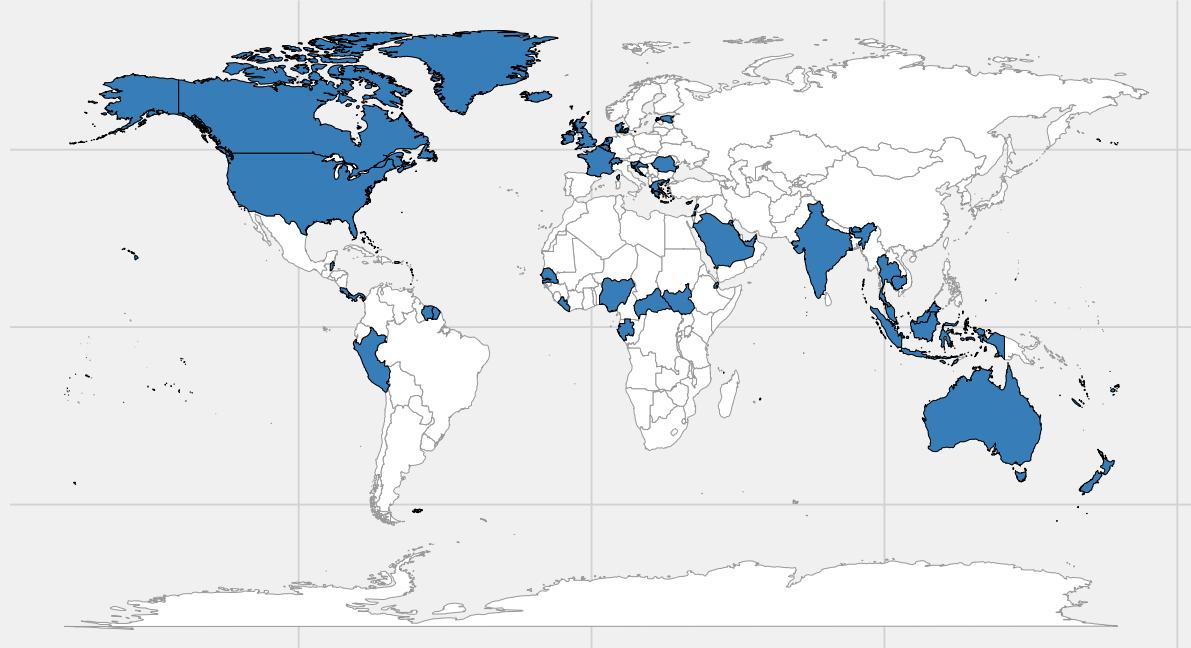
Countries where Mobile Safari 9 is one of top 10 browsers



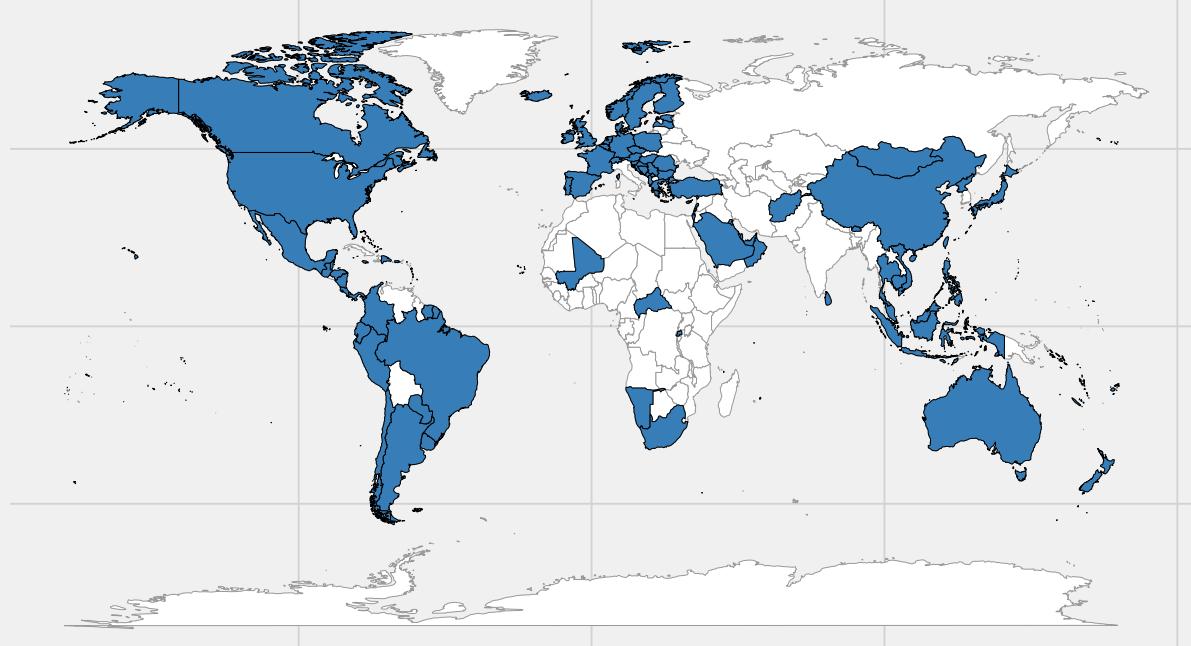
Countries where Safari 5 is one of top 10 browsers



Countries where Safari 8 is one of top 10 browsers



Countries where Safari 9 is one of 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;
```