Report on Wikipedia Portal Visitors

A Look At Session Lengths

Mikhail Popov (Analysis & Report)

2 June 2016

Executive Summary

The way our event logging (EL) schema currently works on Wikipedia Portal (wikipedia.org) is: randomly decide if the user gets enrolled in EL (1 in 200); if they are selected, set a timer for 15 minutes; if they come back to the page BEFORE their session's expiration, renew the timer for another 15 minutes, but if they come back AFTER the expiration, stop tracking their actions and clear the local storage.

The 15 minute timer was an initial guess when the EL schema was initially put together. The purpose of this report is to provide a thorough reference point if/when we decide to adjust the EL timer on the Wikipedia Portal.

Specifically, we found that the most common session length is approximately 10 seconds, and that majority of the sessions are shorter than 1 minute. English is the most most-preferred language in our users, many of whom are from United States. Around 80% of the English-using visitors' sessions are shorter than 1 minute, and same for United States visitors, while only 45% of the Russian-using visitors' sessions are shorter than that.

Data

The event logging (EL) data was extracted from the database using the script data.R and refined into sessions using refine.R.

Throughout the report "session length" is calculated as the arithmetic difference between the last recorded event and the first recorded event from a user, uniquely identified by a combination of: session ID, user agent, and language preferences. These are used in tandem to correct for possible session ID clashes.

Results

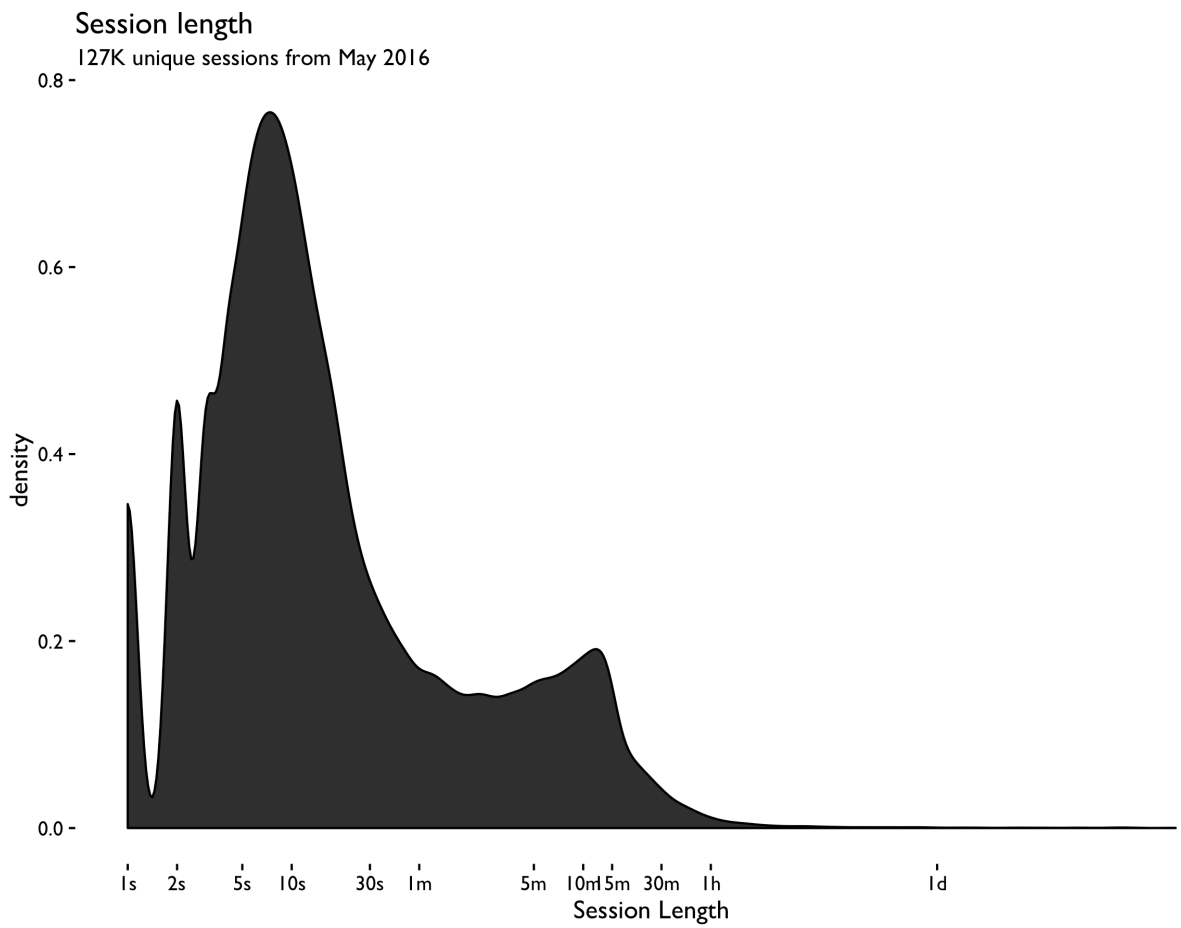

Figure 1: The peak is at 10 seconds, which is a session length that we will see in other figures. Very few sessions last longer than 30 minutes.

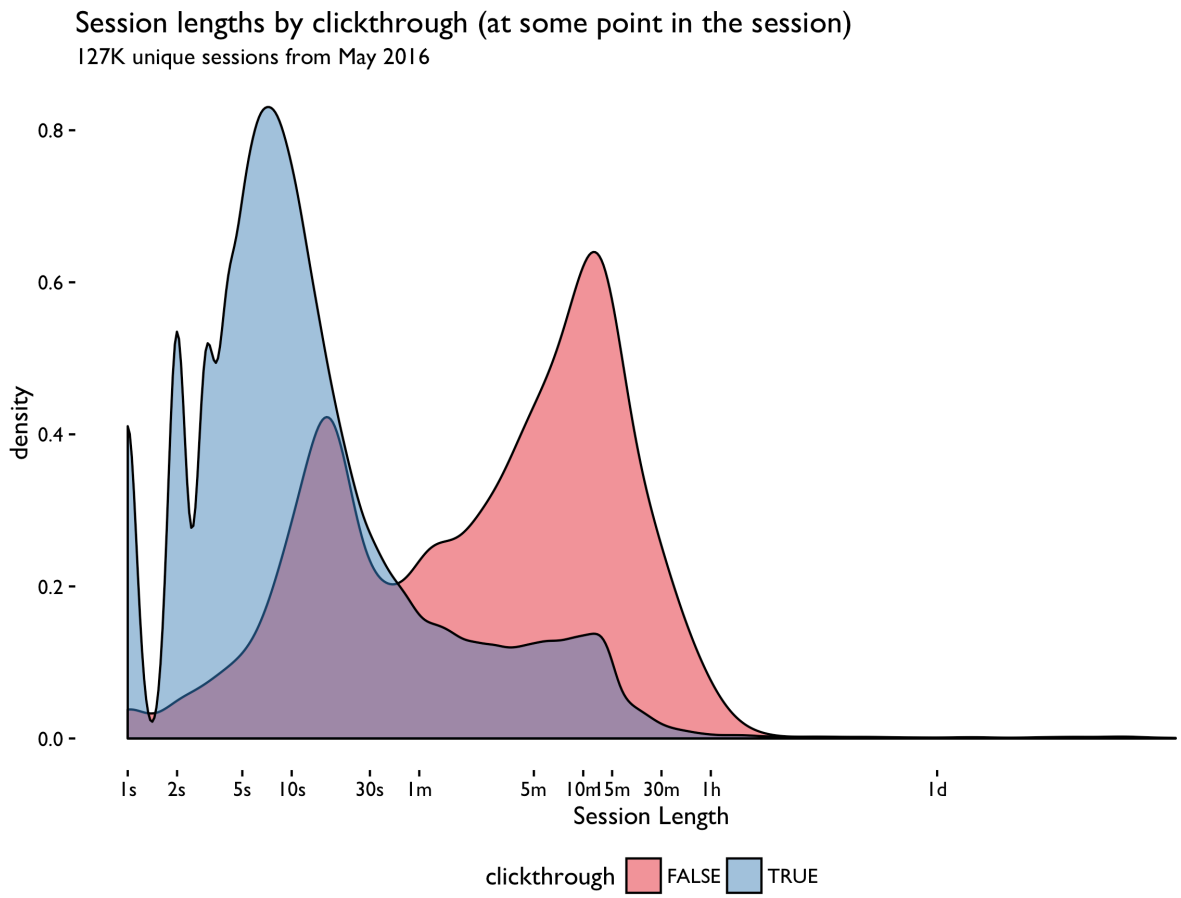

Figure 2: The second peak around the 15 minute point in the previous figure is mostly explained by the sessions that visited multiple times but never clicked through.

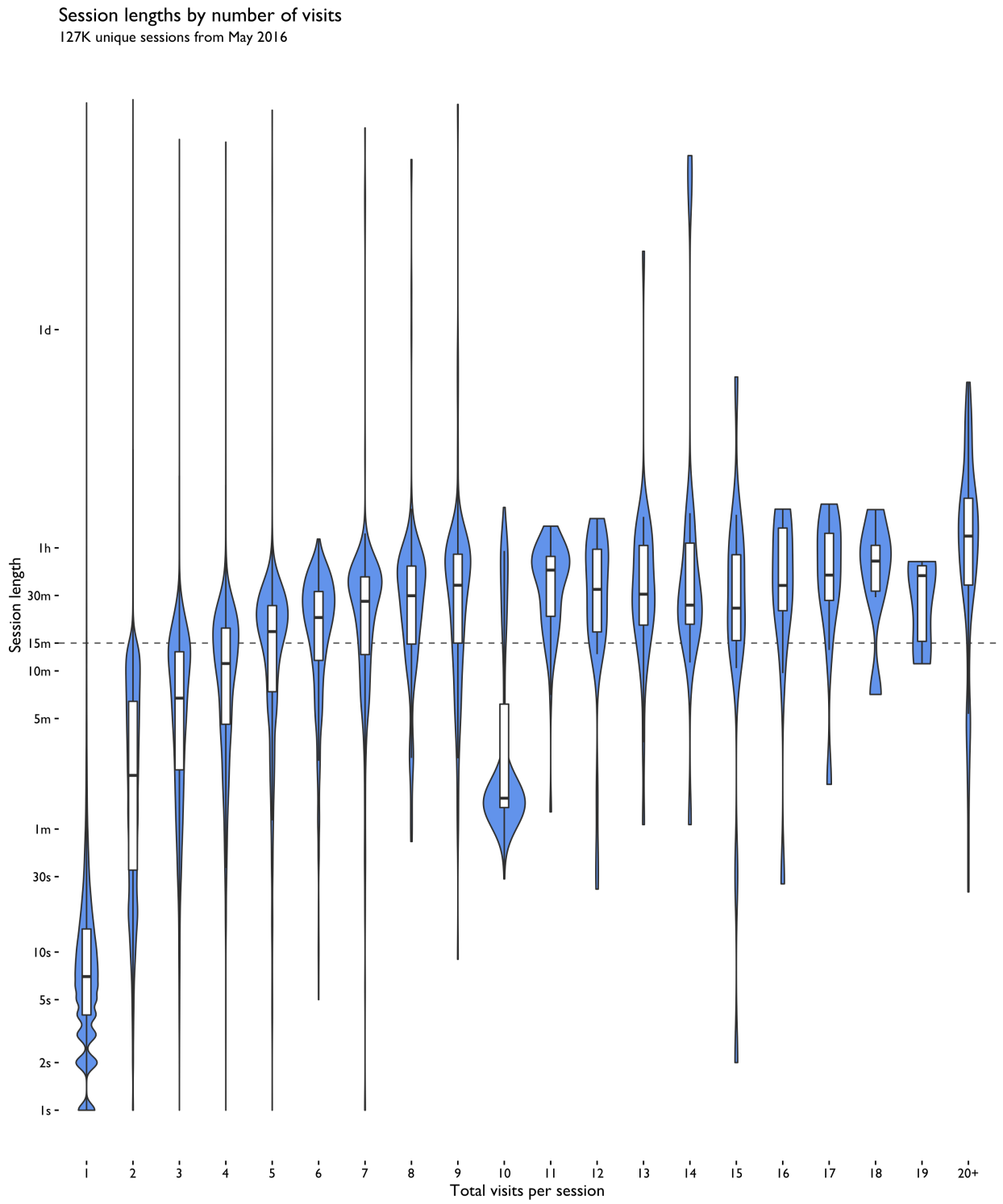

Figure 3: A large portion of the 1-visit sessions lasted between 5 and 10 seconds, with most of those sessions lasting between 5 and 30 seconds. Most of the 2-visits sessions lasted between 30 seconds and 10 minutes (or 15 minutes, if we want to be more inclusive). Then, as number of visits per session increases, the session length also increases at a slow rate.

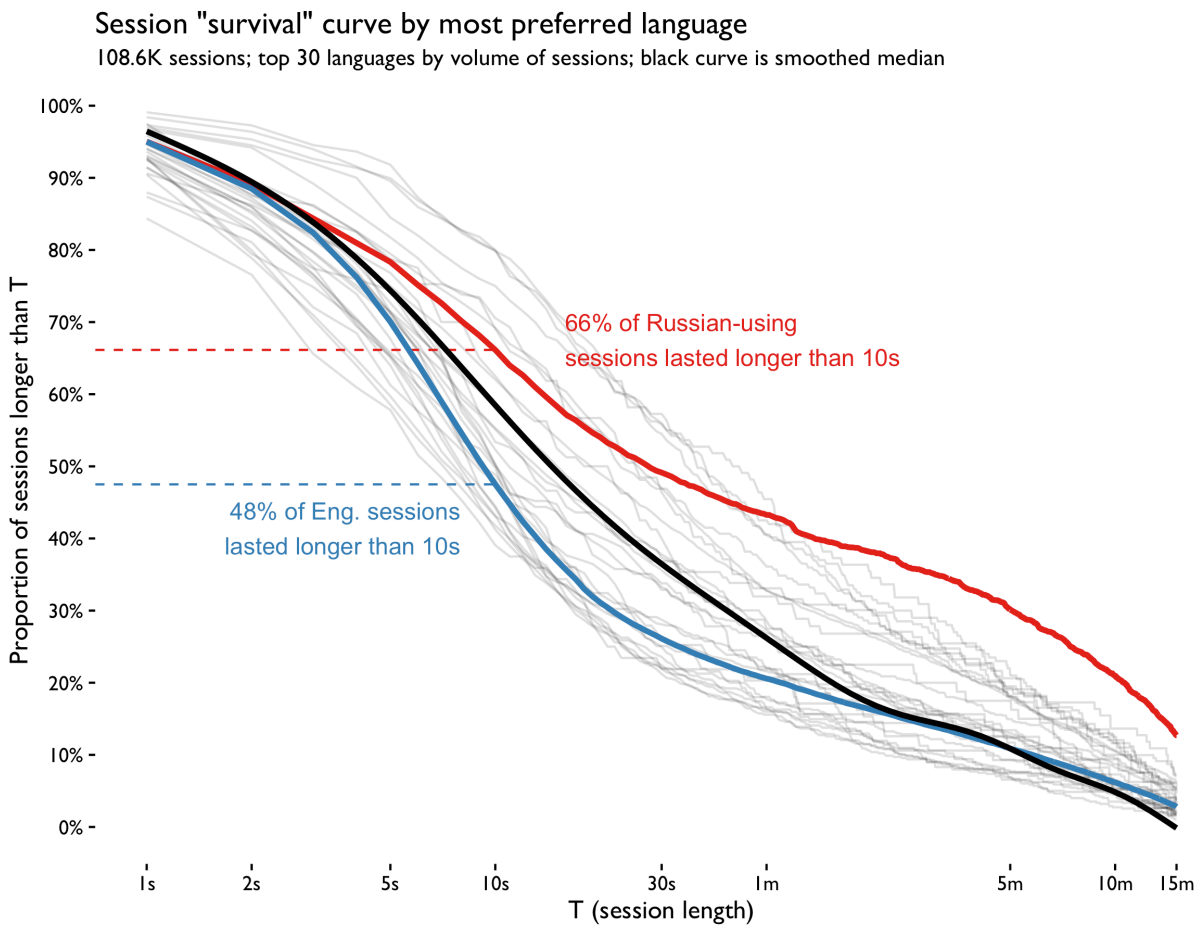

Figure 4: This is a survival curve (broken down by most preferred language) that we can use to get an idea of how users of different languages (such as English and Russian) have sessions of very different lengths. Specifically, English-using visitors have shorter sessions than the average (median) of all languages, while Russian-using visitors have longer sessions.

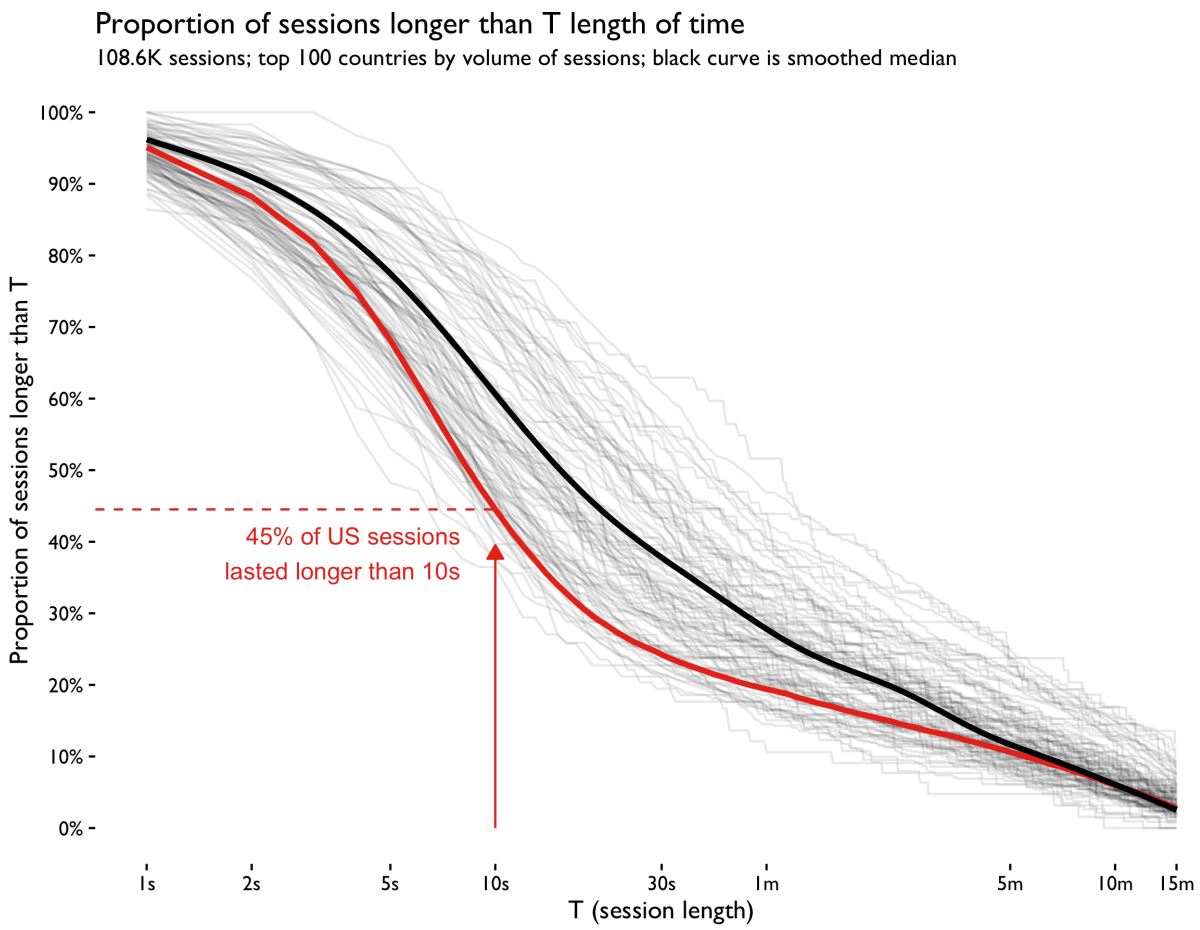

Figure 5: This is a survival curve (broken down by most preferred language) that we can use to get an idea of how users of different languages (such as English and Russian) have sessions of very different lengths. Specifically, English-using visitors have shorter sessions than the average (median) of all languages, while Russian-using visitors have longer sessions.

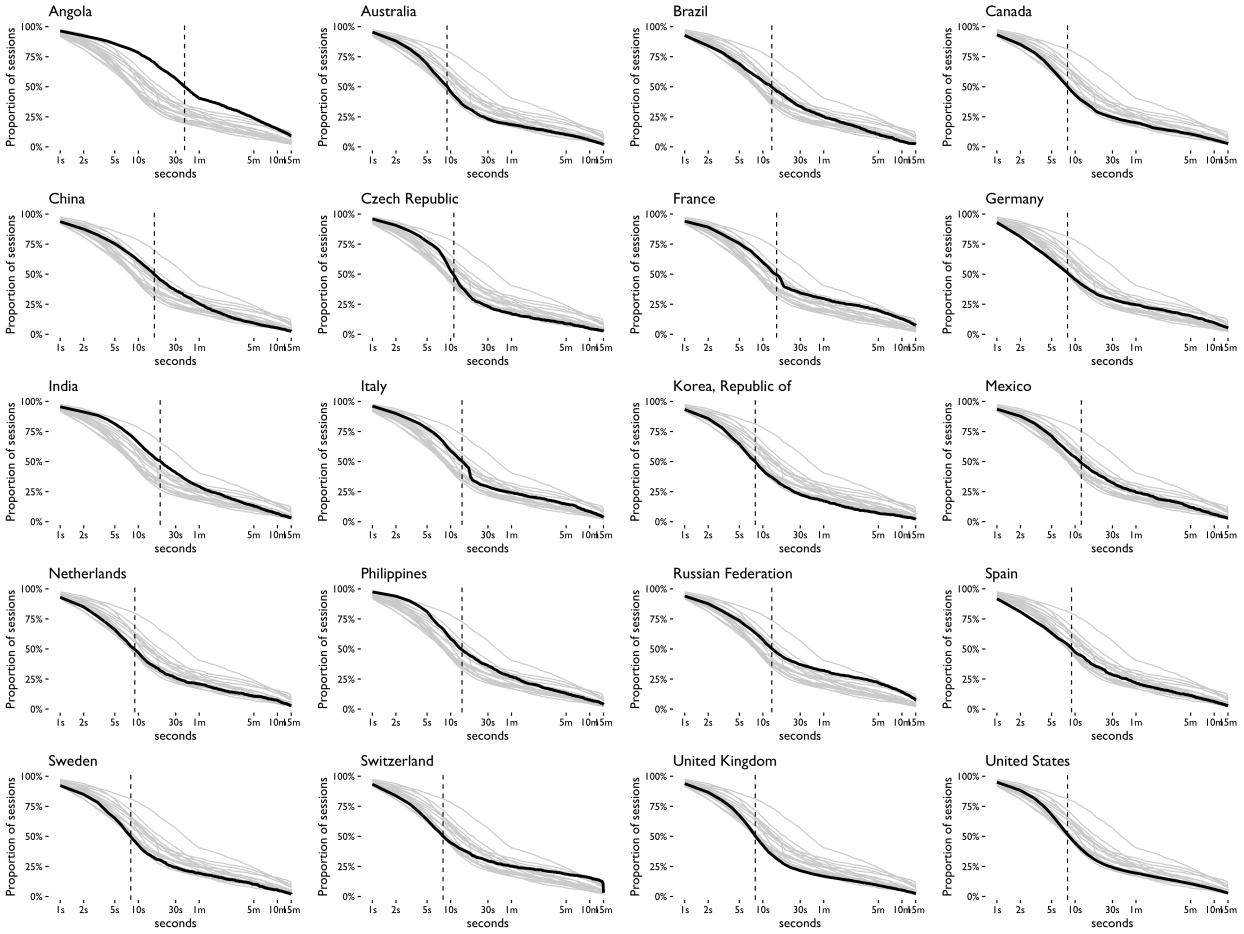

Figure 6: These are the top 20 countries (by volume of sessions) with the dashed vertical line indicating the time point at which approximately half (the median) of the sessions have ceased, a different time point for each country. For example, this point is 7 seconds for United States & United Kingdom, but 11 seconds for Russia.

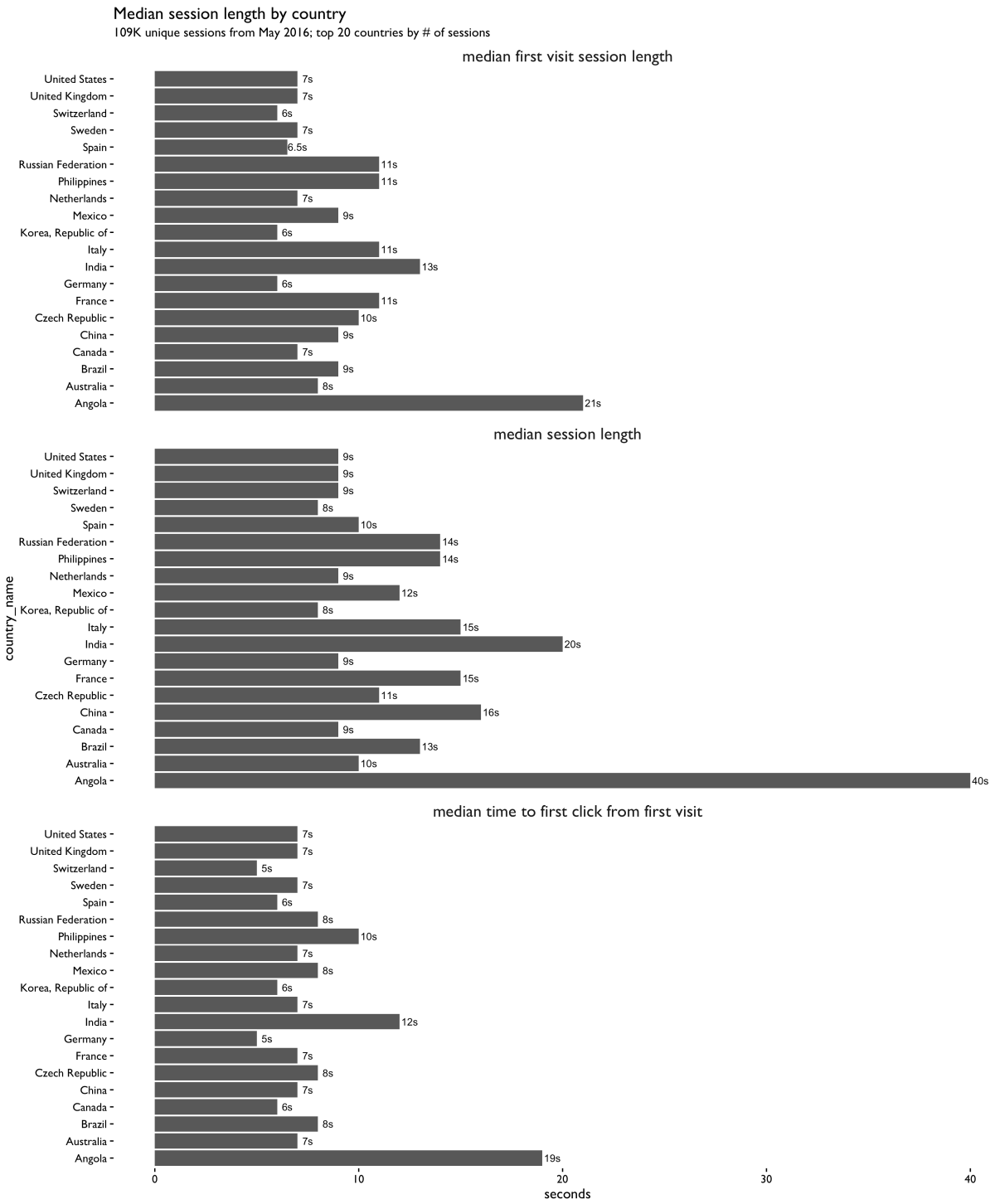

Figure 7: Median first visit session length, median session length, and median time to first click from initial landing by country.

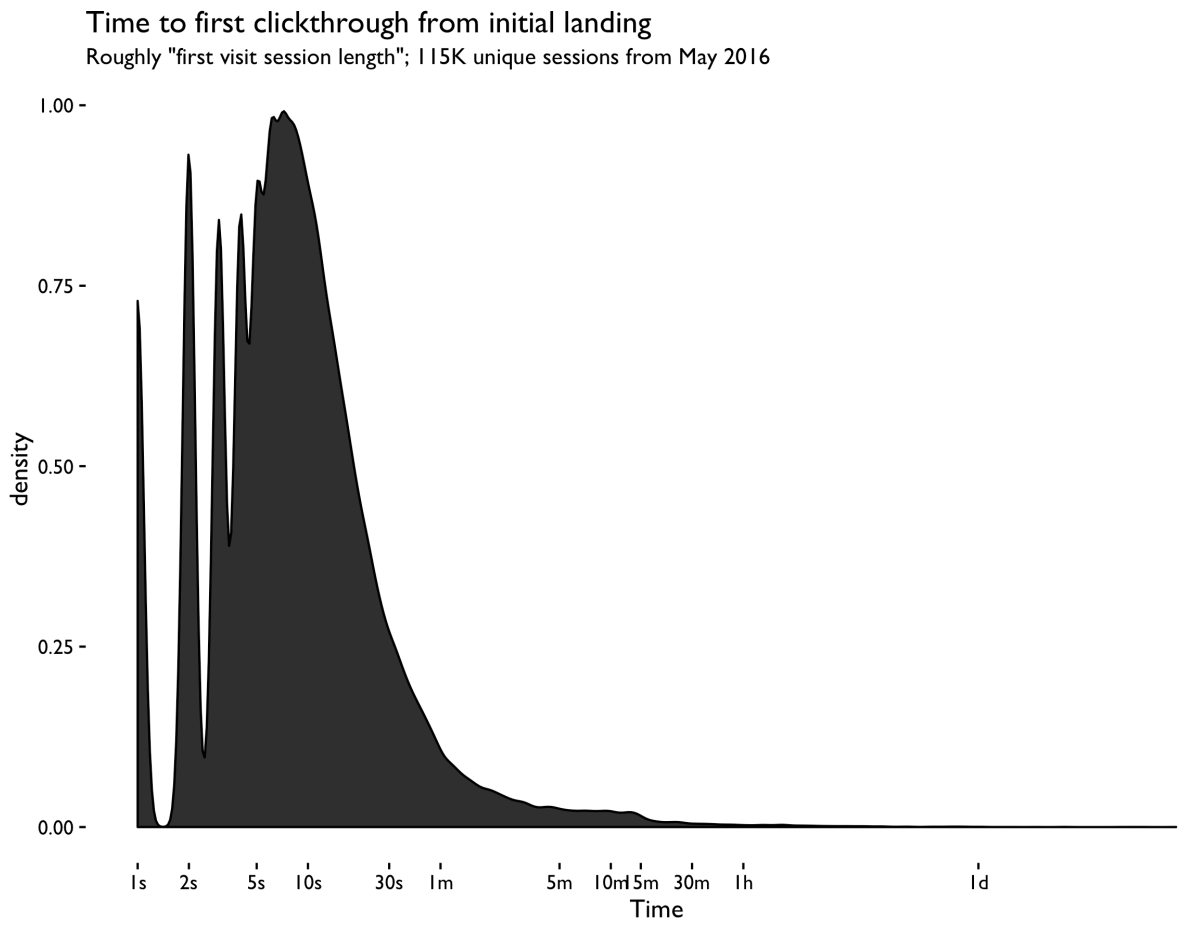

Figure 8: A vast majority of sessions take between 1 and 60 seconds to click from their initial landing on the Portal page, with a peak aroun d the 10 second point.

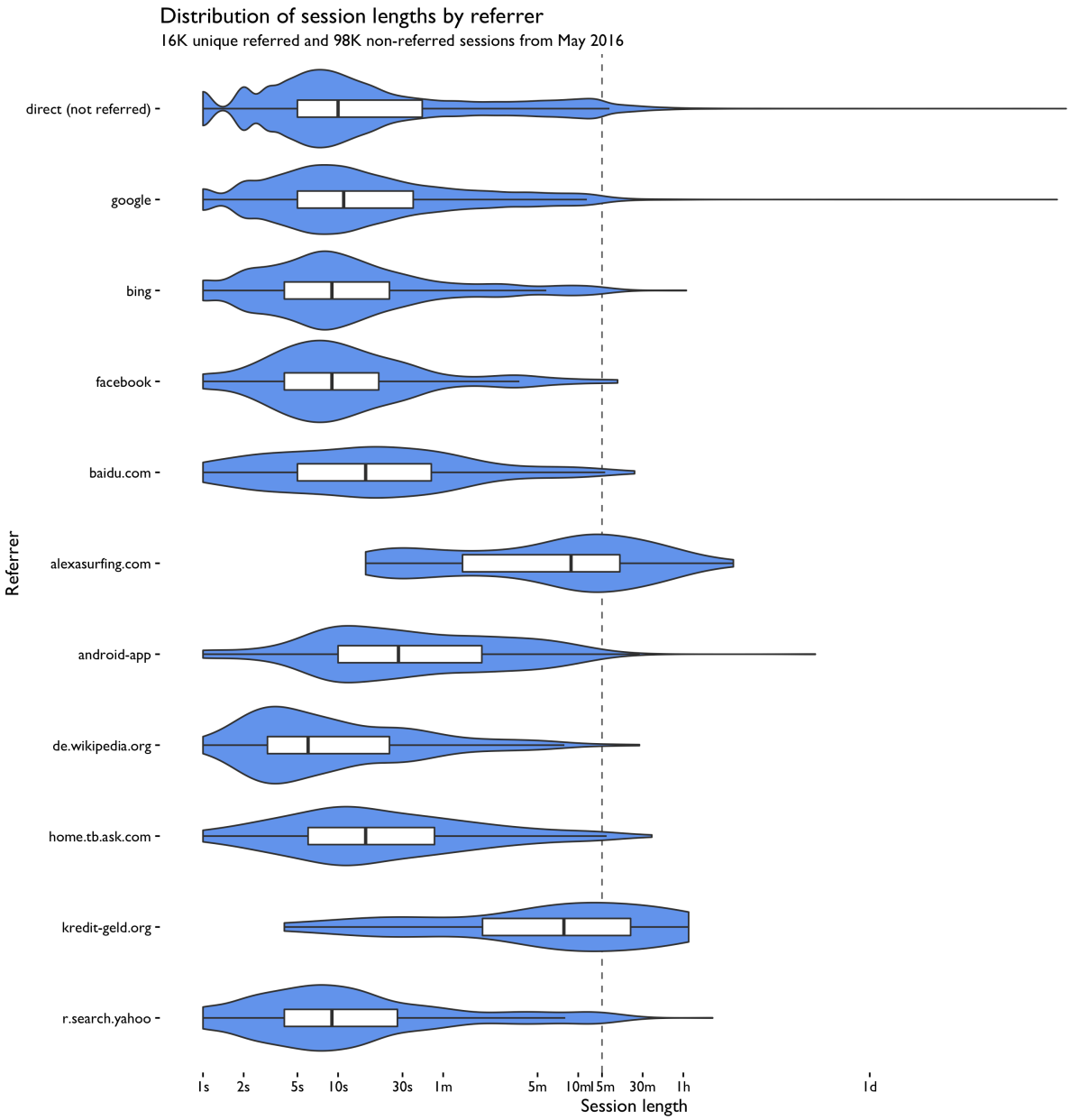

Figure 9: Two of the top ten referrers (alexasurfing.com and kredit-geld.org) did not have any clickthgroughs, and the destributions of session lengths referred by those two particular referrers are drastically (visibly) different from the rest of the session length distributions, including direct traffic. We grouped the various Google, Bing, and Yahoo websites together by omitting the TLD extensions.