Engagement with Wikipedia Portal

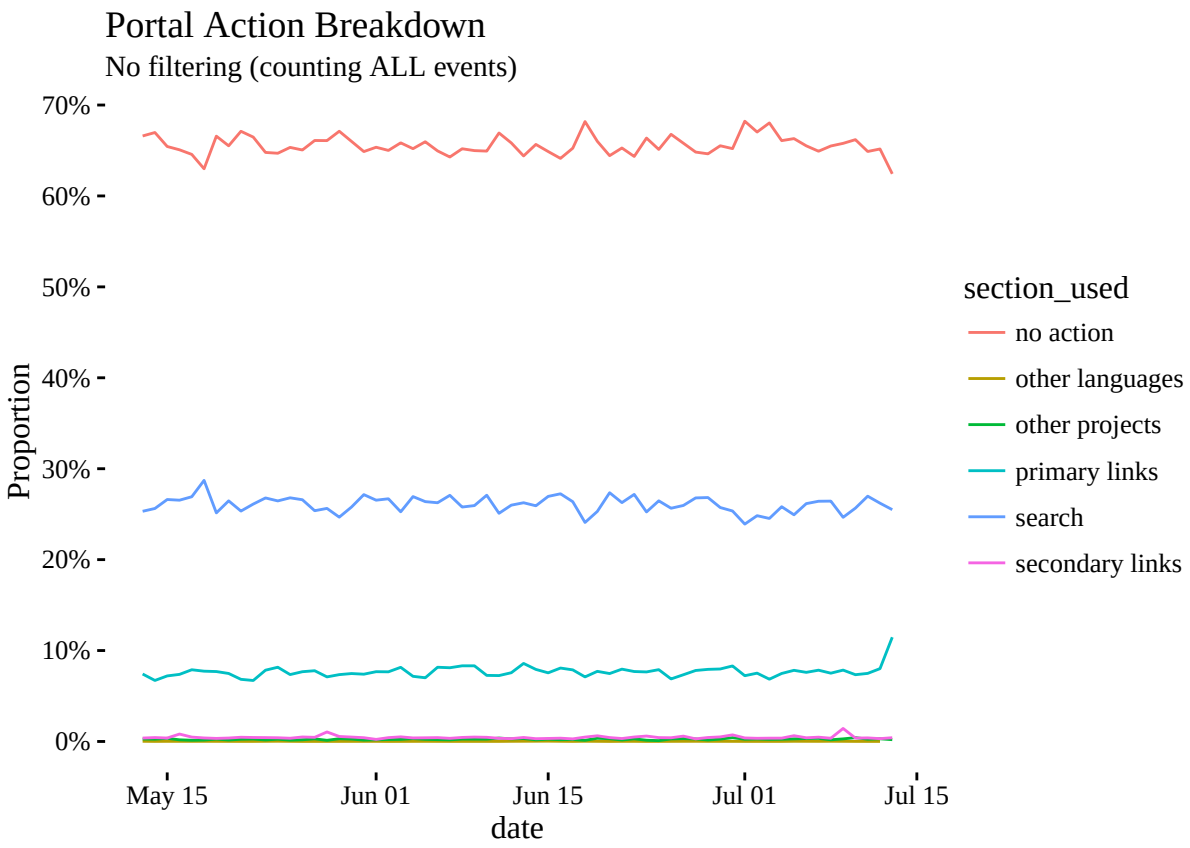
Mikhail Popov

20 July 2016

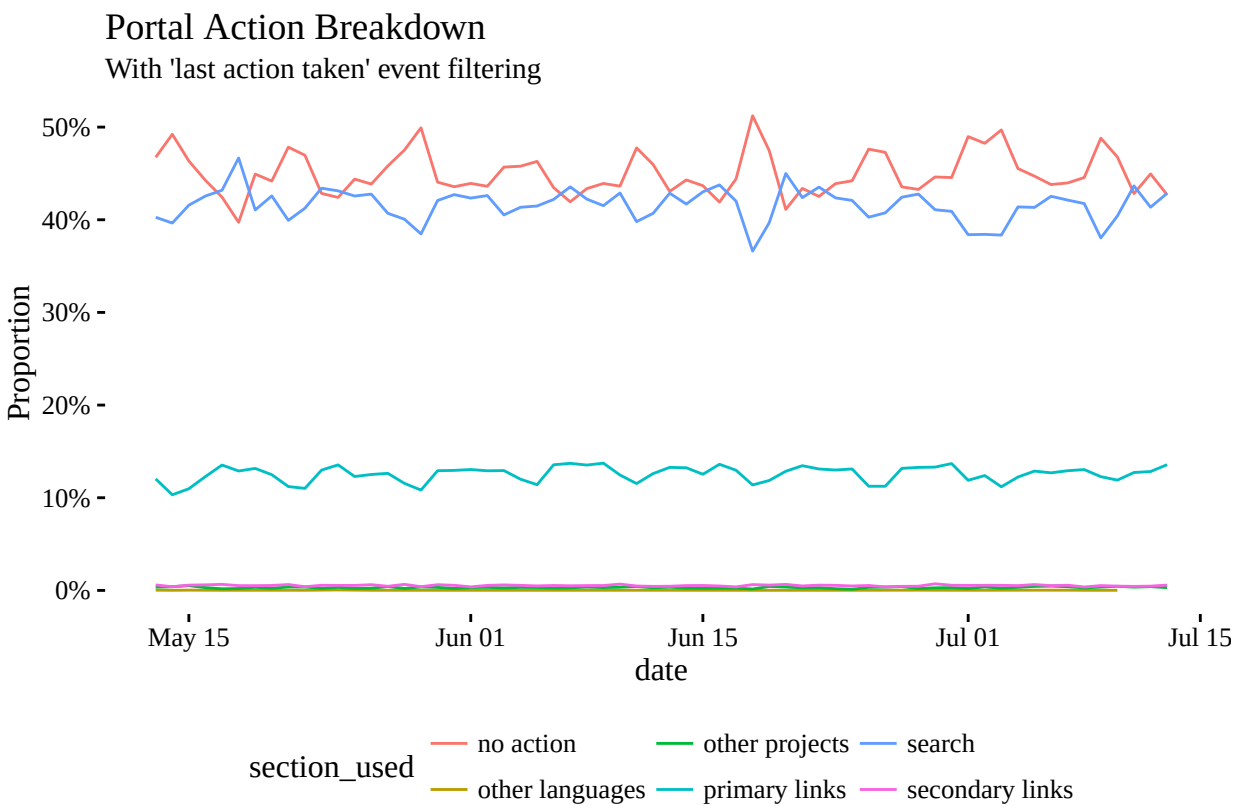
Summary

In this brief report we respond to the issue reported in T139109 wherein there was a huge spike in no-actions. The issue was caused by an accidental change to the data collection/aggregation scrips. Namely, there was a "last action performed in session" filtering applied to events that was removed in a patch because it was snaply determined as being unnecessary. The change has since then been reverted (as a result of the analysis performed in this report). We conclude with a proposal of an alternative metric, either the "most commonly clicked section per visit" or "most common section clicked on across all visits per session".

Exploratory Data Analysis



The difference is actually caused by a line deletion from the script that processes the data fetched from the EL database. Here's the same data but with a filtering that the script previously had.



The reasoning behind that change was that this panel is supposed to be counts/proportions of ALL events, even if one user had a single session with hundreds or thousands of events. Previously, the breakdown referred to the last action in each session. A separate dashboard panel was created for "actions on first visit". This change has been reverted, the data removed, and backfilled on 14 July 2016.

Table 1: Top 10 sessions by number of clicks per session from 13 May 2016 to 13 July 2016. Of note is the user who clicked 526 times on the primary links on their first visit to the page.

country	device	os	browser	visit	clickthroughs	sections used	most used section	clicks on that section
Malaysia	Asus T00I	Android	Chrome Mobile	1st	526	1	primary links	526
United States	Generic Smartphone	Android	Chrome Mobile	1st	95	2	primary links	61
Thailand	Other	Windows 7	Chrome	1st	80	1	search	80
Germany	Generic Smartphone	Android	Chrome Mobile	4th	74	2	secondary links	58
India	Lenovo P70-A	Android	Chrome Mobile	1st	61	2	secondary links	42
NA	Kindle	Android	Amazon Silk	1st	44	1	search	44
NA	Other	Windows Vista	Chrome	2nd	37	1	search	37
Germany	Generic Smartphone	Android	Chrome Mobile	6th	36	3	secondary links	28
Viet Nam	Other	Windows 7	Chrome	1st	34	1	search	34
Spain	Generic Smartphone	Android	Chrome Mobile	4th	32	2	primary links	16

Table 2: Top 10 sessions by number of visits from 13 May 2016 to 13 July 2016. Of note is the user whose session lasted 23 minutes but they visited the Wikipedia Portal 279 times. Of another note is the number of sessions with many visits but not any clickthroughs, hence the impact on the metric when the filter was removed.

visits	session length	country	device	os	browser	maximum sections used	most used sections	most used section	total clicks
279	22m 29s	Russian Federation	Lenovo A316i	Android	Android	1	secondary links	secondary links	2
274	1h 23m 37s	Russian Federation	Other	Windows 7	Chrome	1	search	search	270
249	22h 33m 2s	Germany	Other	Windows 7	IE	0			0
199	11h 9m 5s	NA	Other	Linux	PhantomJS	0			0
152	1h 38m 37s	United States	Other	Windows 10	Edge	0			0
152	1h 38m 56s	United States	Other	Windows 10	Edge	0			0
151	1h 52m 5s	United States	Other	Windows 10	Edge	0			0
150	1h 31m 2s	United States	Other	Windows 10	Edge	0			0
149	1h 49m 19s	United States	Other	Windows 10	Edge	0			0
144	39s	India	Other	Windows 10	Edge	0			0

Why unfiltered "no action" numbers are so big

How engagement varies by country

Potential metrics as alternatives or in addition to current ones

A major problem with the current metric is that there is not a very strong interpretation or story being told. All it tells us is the proportion of sessions where a particular section was the last thing clicked. A user who searches 100 times but then clicks on a primary link as the last action before their session expires will show up as a primary link in the daily clickthrough breakdown. Below are some potential alternatives that have better stories to tell about how Wikipedia Portal visitors engage with the page.

As we saw above, some visits can lead to many clickthroughs. In the "most commonly clicked section per visit" metric, we find the most clicked section (if the user clicked on multiple sections after landing on the page once) per visit. All visits contribute equally and independently to the metric, with no accounting for multiple visits in one session. In the "most common section clicked on across all visits per session" metirc, we find the most used section per visit as before, but then look at each session individually and find which section they used the most across all the visits that session has. For example, if the user visits the Wikipedia Portal 10 times, 7 of which resulted in them using search while 3 resulted in them going to a Wikipedia in a particular language, then their whole session would contribute to "search".

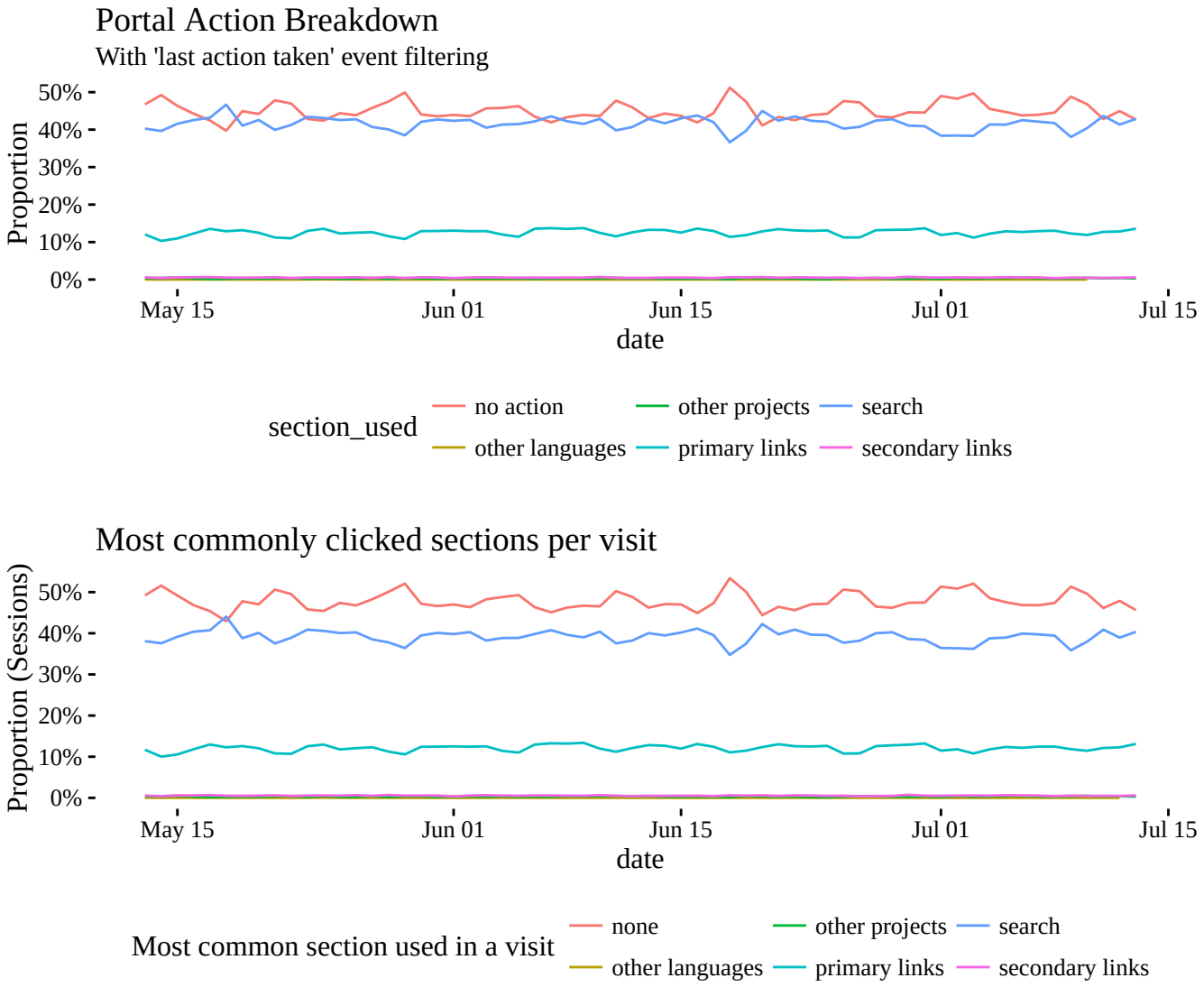


Figure 1: Two proposed metrics.

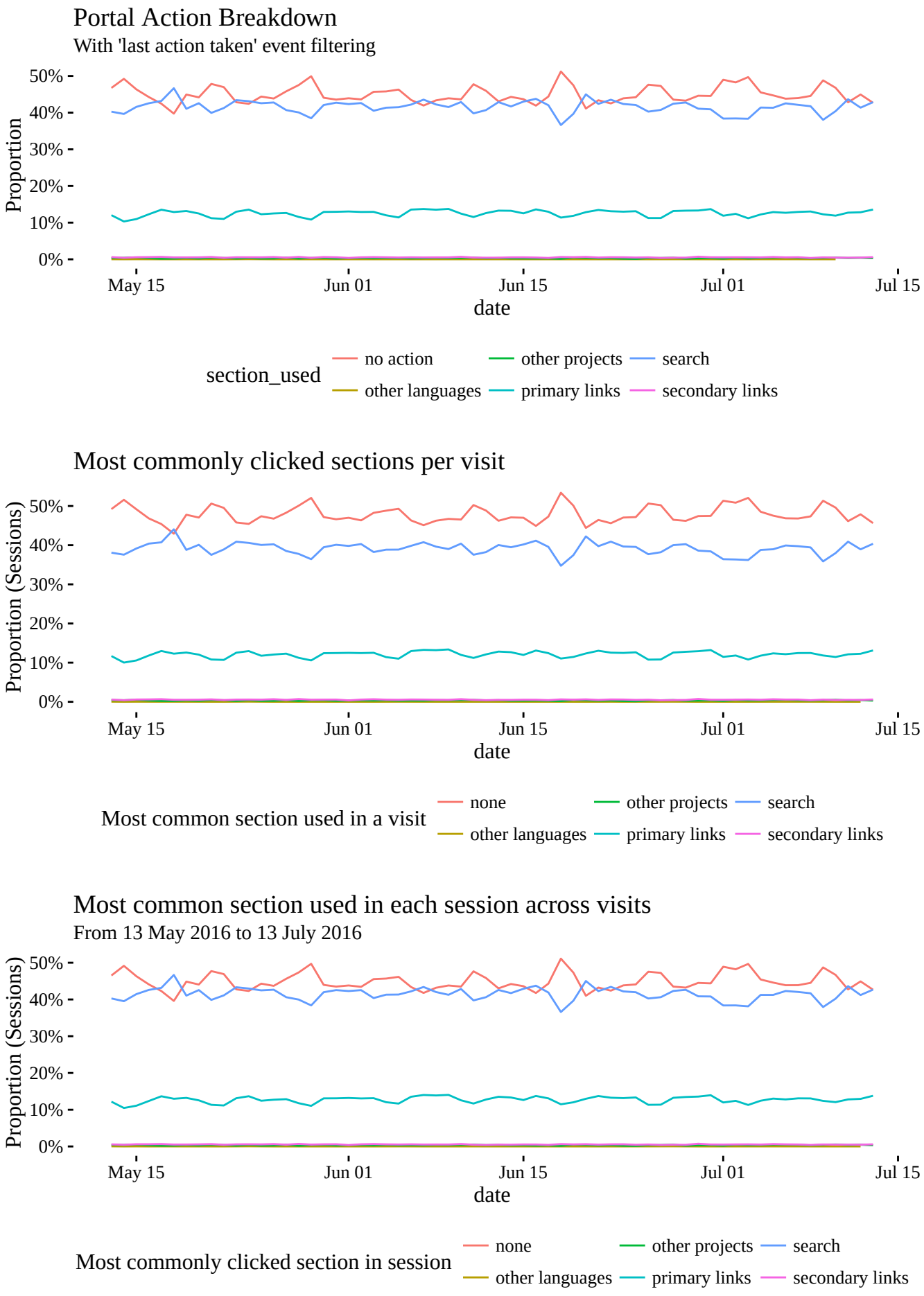


Figure 2: Comparison of current metric with two proposed.