

Influence of user location on revert

By [Mykola Trokhymovych](#) (Research Intern)

Introduction

There is an assumption that users from different locations have different patterns of behavior for different Wikipedia languages. It can result in various revisions quality they create to update Wikipedia pages. The reasons for that can be, for example, cultural (knowledge of the language), political (relations between countries), using VPN, etc. At the same time, [policies](#) should be followed to create quality content on Wikipedia. In case of geographical features influence revision quality, they should be considered in the models that aim to automatically evaluate the quality of revisions.

In this brief report, I observe how the location of anonymous users correlates with the quality of their revisions. As a proxy to the quality, I use the revert rate. Revert usually means that the suggested revisions have some issues, so it is canceled. I use anonymous users, as their IP is open to the public when they change the Wikipedia page so that the approximate location can be inferred. Also, anonymous users usually have a significantly higher revert rate, and their revision more frequently includes vandalism. I observe the spaces for languages where I am a native or fluent speaker (enwiki, ukwiki, ruwiki).

Data preparation details

I use *the mediawiki_history* data set and collect revision history with such rules:

- event_entity = 'revision'
- wiki_db in (ukwiki, ruwiki, enwiki)
- snapshot dated 2022-07
- the observation period is 2021-08-01 – 2022-08-01 (one year)
- consider only revisions with page_namespace = 0 (content)
- Filter bots
- leave only revisions by anonymous users

Important note: for enwiki, I am taking 0.5 random samples of a dataset to speed up processing and analysis.

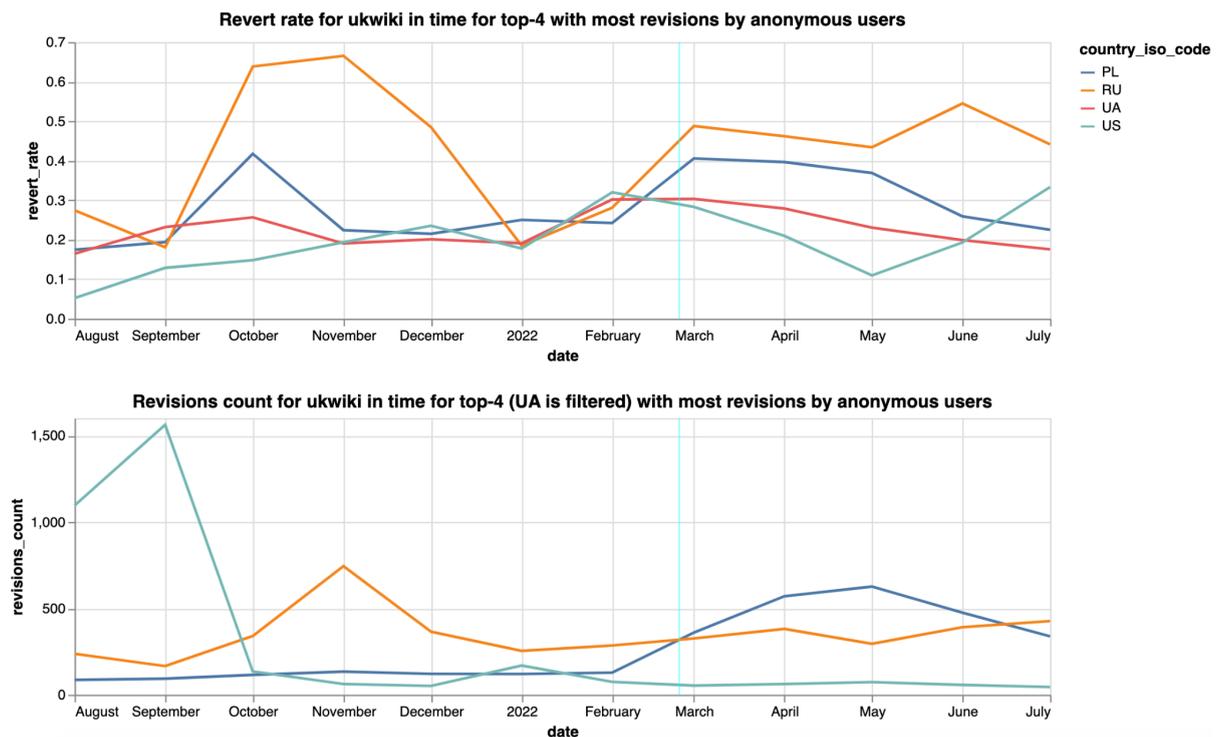
After that, I used the [GeoLite2-City](#) database to extract location features from the IP address of anonymous users, which is actually their user_name.

Results analysis

General stats and insights

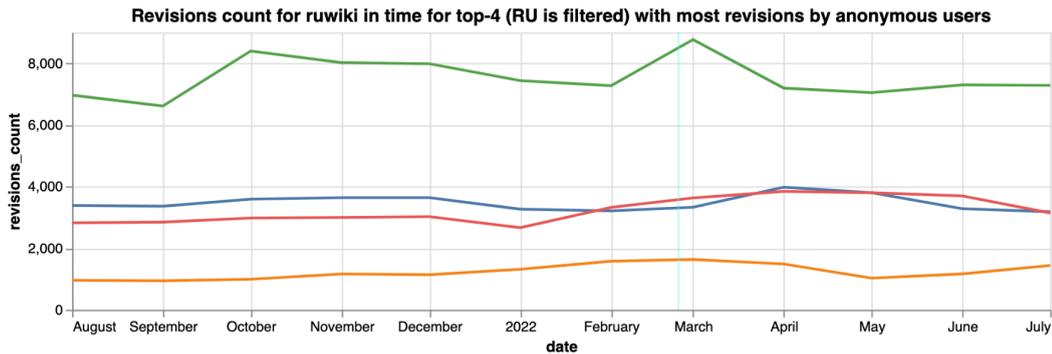
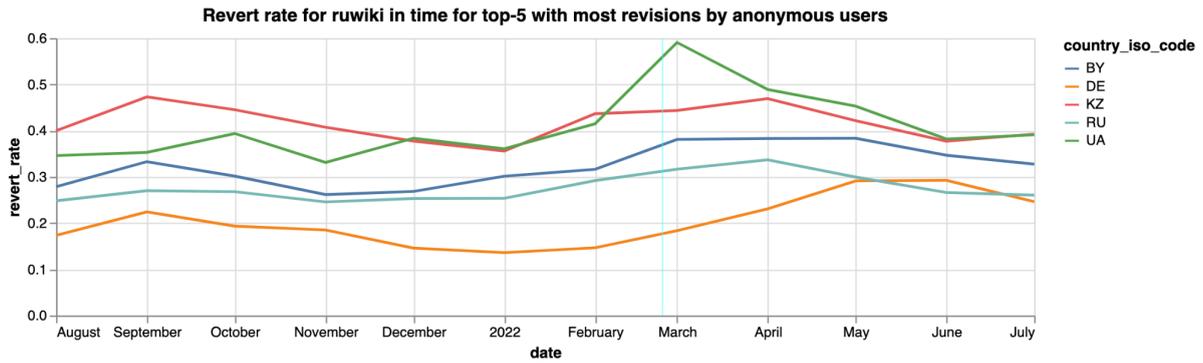
wikidb	Number of anonymous revisions	Revert rate
enwiki	3.943.709	0.288
ruwiki	996.728	0.303
ukwiki	153.573	0.224

How revert rate changes in time



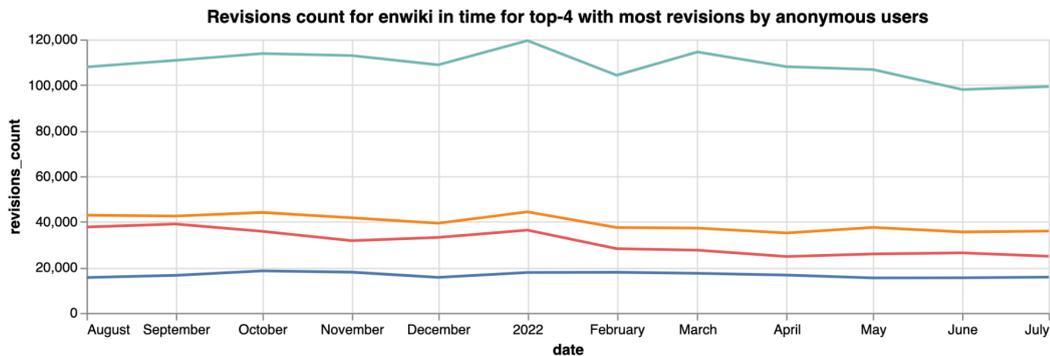
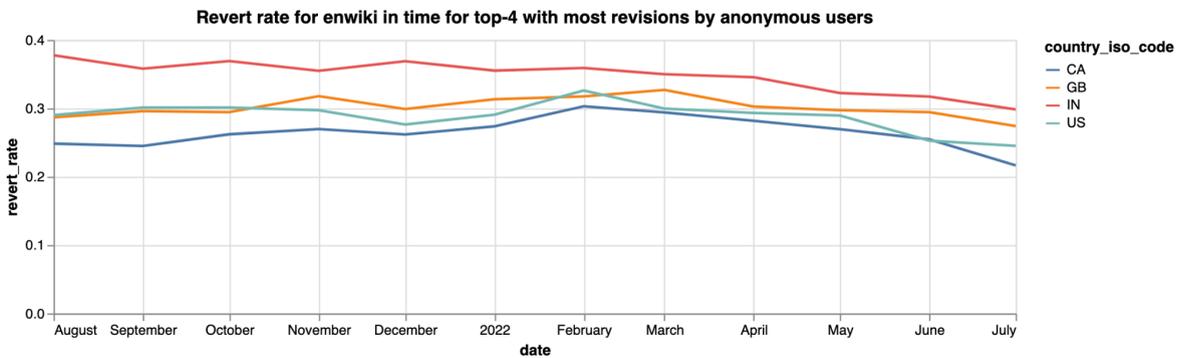
Summary for ukwiki:

- Top frequent countries are UA, RU, US, PL
- We see that with the start of [Russian invasion of Ukraine](#) the revert rate increased almost twice
- Another observation is that number of revisions from PL increases up to 5 times. (Poland was the main destination for Ukrainian refugees)



Summary for ruwiki:

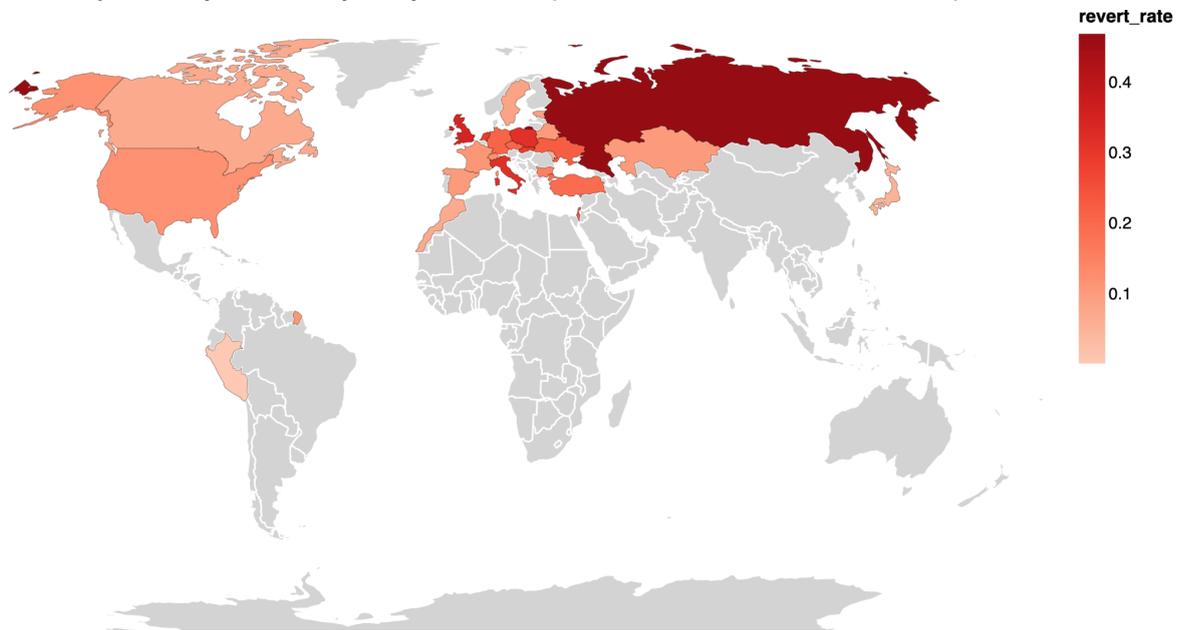
- Top frequent countries are RU, UA, KZ, BY, and DE
- We see that with the start of [Russian invasion of Ukraine](#), the revert rate from Ukrainian users has spiked, increasing up to 60% of revert rate in March.
- It should be mentioned that the revert rate in different countries has a similar dynamic.



Revert rate per country map

Observing ukwiki:

Revert rate per country for ukwiki by anonymous users (countries with more than 200 revisions)

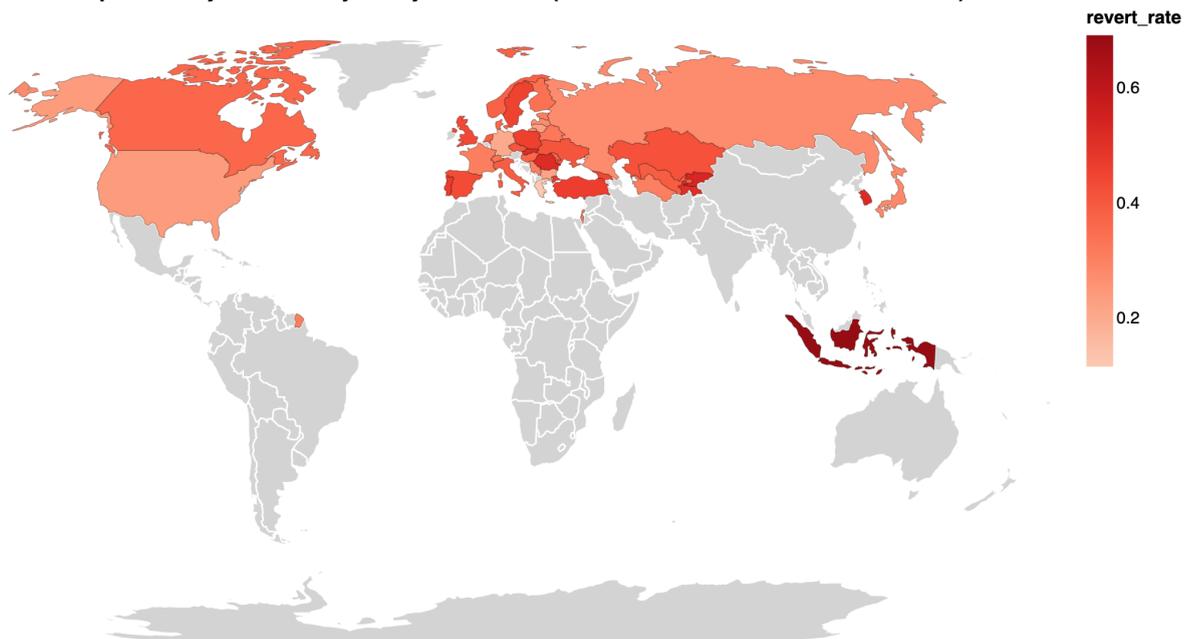


Top-10 countries and their statistics for ukwiki:

country_iso_code	revert_rate	revision_count
UA	0.224952	123382
RU	0.469121	4210
US	0.121547	3439
PL	0.315856	3166
JP	0.051257	3024
DE	0.212527	2315
GB	0.346851	1159
ES	0.104555	966
CZ	0.209003	933
CA	0.071823	905

Observing ruwiki:

Revert rate per country for ruwiki by anonymous users (countries with more than 200 revisions)

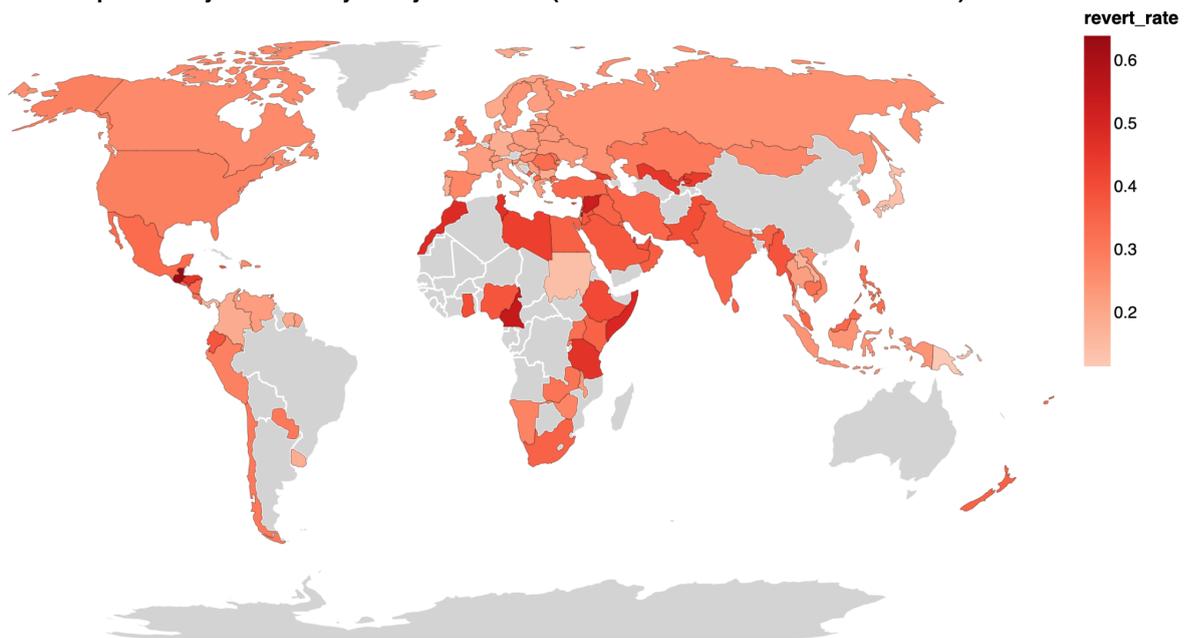


Top-10 countries and their statistics for ruwiki:

country_iso_code	revert_rate	revision_count
RU	0.275676	723878
UA	0.409718	90267
BY	0.324082	41653
KZ	0.417961	38762
DE	0.202036	14834
US	0.238090	8732
AZ	0.487976	5988
PL	0.460948	5467
MD	0.373358	4797
UZ	0.392514	4782

Observing enwiki:

Revert rate per country for enwiki by anonymous users (countries with more than 200 revisions)



Top-10 countries and their statistics for enwiki:

country_iso_code	revert_rate	revision_count
US	0.289173	1304636
GB	0.301681	473650
IN	0.350727	371178
CA	0.265748	200103
AU	0.299613	153154
PH	0.327915	103167
ID	0.244319	75029
DE	0.183062	67081
IT	0.210057	61488
BR	0.218194	59493