



**RIPE** Réseaux IP  
Européens

RIPE 89 Meeting | 29 October 2024



# Unboxing the APNIC per AS User Population Dataset

Loqman Salamatian  
Calvin Ardi  
Matt Calder  
Vasilis Giotsas  
Ethan Katz-Bassett  
Todd Arnold

# Shoutout to amazing collaborators



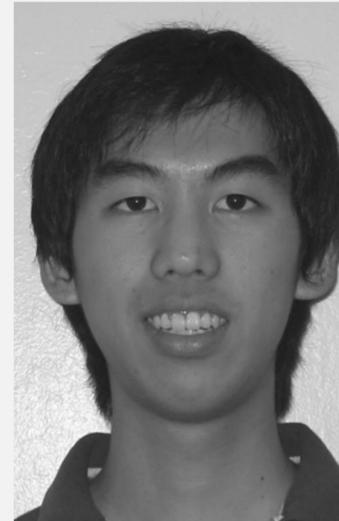
Loqman Salamatian



Matt Calder



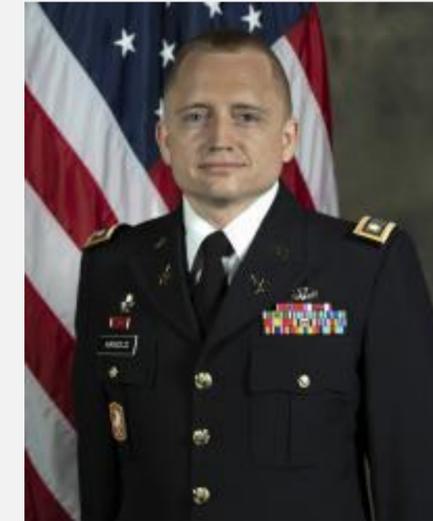
Calvin Ardi



Ethan  
Katz-Bassett



Todd Arnold



It's important to understand the impact of incidents and actions in terms of people affected



**FNN Fast News Netw** @fastnews · 2m  
#BREAKING 1 million + people are without power in Florida from Hurricane #Helene #flwx x.com/fastnewsnet/st...

**Florida**  
Customers Tracked: 11,399,159 Customers Out: 1,036,553 Last Updated: 2024-09-26 11:31:42 PM

# Lack of access to drinking water affects 33 million people in Brazil

World Water Day is celebrated this Friday, March 22

EDRO PEDUZZI

Published on 22/03/2024 - 11:02



## Thousands more evacuated in Greece as high winds and heat fuel wildfires

19,000 people, mostly tourists, moved by buses and boats out of path of fire in southern Rhodes, as evacuations begin on other islands

News story

## £1,000 yearly tax cut for households from today

27 million people across the UK will benefit from a yearly tax cut worth hundreds of pounds from today, meaning a household with two average earners will save nearly £1,000 per year.

## Tens of thousands of MBTA riders now qualify for half-price fares

September 03, 2024

By [Andrea Perdomo-Hernandez](#)



The MBTA is set to launch a new reduced fare program this week, dramatically widening the pool of people who qualify and potentially cutting transit costs in half for an additional 60,000 riders.



# We need to assess Internet phenomena in the context of number of users affected

- How many users were affected during a disruption?
- How many users can benefit from a routing policy change?
- What is the market share of an Autonomous System (AS) in terms of users served?
- How many users can resolve a DNS name?
- How many users are affected by a BGP hijack?



# We need to assess Internet phenomena in the context of number of users affected

- How many users were affected during a disruption?
- How many users can benefit from a routing policy change?
- What is the market share of an Autonomous System (AS) in terms of users served?
- How many users can resolve a DNS name?
- How many users are affected by a BGP hijack?

**How can we find how many users are served by a network?**

# The APNIC AS Customers Population Dataset



- The only *publicly* available dataset on user populations per ASN

<https://stats.labs.apnic.net/aspop>

**Visible ASNs: Customer Populations (Est.)**

Date: 25/10/2024

Rank	ASN	AS Name	CC	Users (est.)	% of country	% of Internet	Samples
1	AS55836	RELIANCEJIO-IN Reliance Jio Infocomm Limited	<a href="#">IN</a>	290,192,682	48.51	6.874	134,050,096
2	AS4134	CHINANET-BACKBONE No.31,Jin-rong Street	<a href="#">CN</a>	250,488,743	30.9	5.934	32,540,107
3	AS45609	BHARTI-MOBILITY-AS-AP Bharti Airtel Ltd. AS for GPRS Service	<a href="#">IN</a>	152,766,815	25.54	3.619	70,568,307
4	AS4837	CHINA169-BACKBONE CHINA UNICOM China169 Backbone	<a href="#">CN</a>	136,733,542	16.87	3.239	17,762,571
5	AS9808	CHINAMOBILE-CN China Mobile Communications Group Co., Ltd.	<a href="#">CN</a>	121,330,521	14.97	2.874	15,761,619
6	AS197207	MCCI-AS	<a href="#">IR</a>	50,598,427	78.35	1.199	3,216,566
7	AS29465	VCG-AS	<a href="#">NG</a>	44,721,981	63.83	1.059	9,735,273
8	AS7922	COMCAST-7922	<a href="#">US</a>	43,581,470	16.86	1.032	26,024,492
9	AS4812	CHINANET-SH-AP China Telecom Group	<a href="#">CN</a>	42,259,090	5.21	1.001	5,489,729
10	AS8452	TE-AS TE-AS	<a href="#">EG</a>	34,980,119	60.08	0.829	11,998,216
11	AS7713	TELKOMNET-AS-AP PT Telekomunikasi Indonesia	<a href="#">ID</a>	34.126.735	29.67	0.808	16.776.974



# The APNIC AS Customers Population Dataset

- The only *publicly* available dataset on user populations per ASN

<https://stats.labs.apnic.net/aspop>

**Visible ASNs: Customer Population**

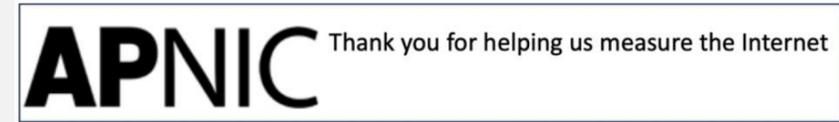
Date: 25/10/2024

Rank	ASN	Users (est.)	% of country	% of Internet	Samples
1	AS55836 RELIANCEJIO-IN Reliance Jio Info	10,192,682	48.51	6.874	134,050,096
2	AS4134 CHINANET-BACKBONE No.31,Jin	10,488,743	30.9	5.934	32,540,107
3	AS45609 BHARTI-MOBILITY-AS-AP Bharti A	8,276,815	25.54	3.619	70,568,307
4	AS4837 CHINA169-BACKBONE CHINA UN	6,733,542	16.87	3.239	17,762,571
5	AS9808 CHINAMOBILE-CN China Mobile C	4,130,521	14.97	2.874	15,761,619
6	AS197207 MCCI-AS	3,059,427	78.35	1.199	3,216,566
7	AS29465 VCG-AS	4,721,981	63.83	1.059	9,735,273
8	AS7922 COMCAST-7922	3,581,470	16.86	1.032	26,024,492
9	AS4812 CHINANET-SH-AP China Telecom C	2,259,090	5.21	1.001	5,489,729
10	AS8452 TE-AS TE-AS	34,980,119	60.08	0.829	11,998,216
11	AS7713 TELKOMNET-AS-AP PT Telekomunikasi Ind	34,126,735	29.67	0.808	16,776,974

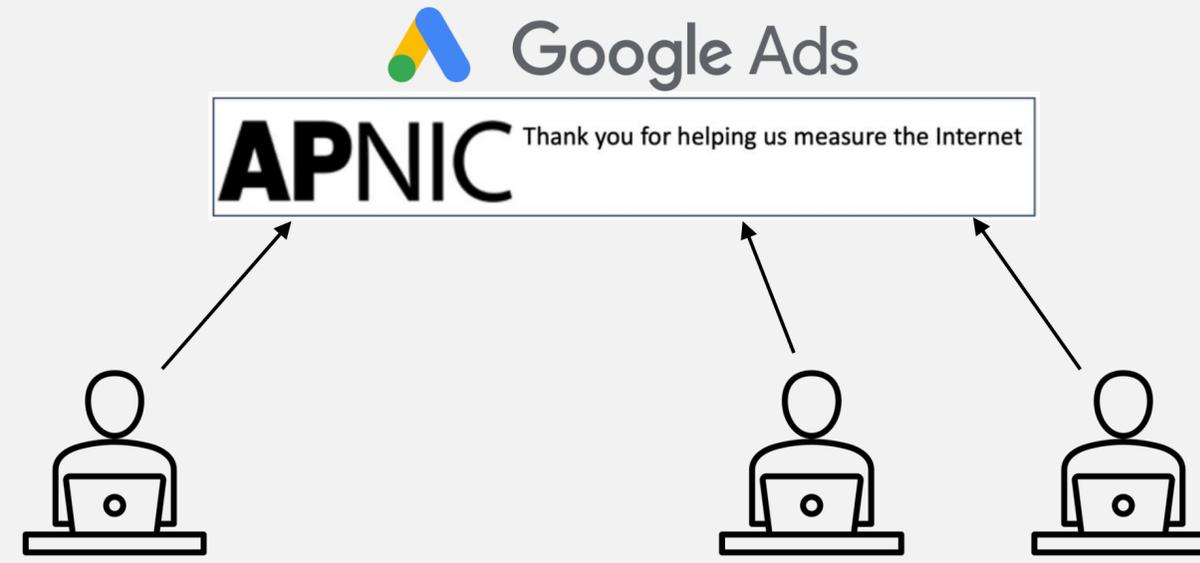


**How accurate are the AS population estimates?**

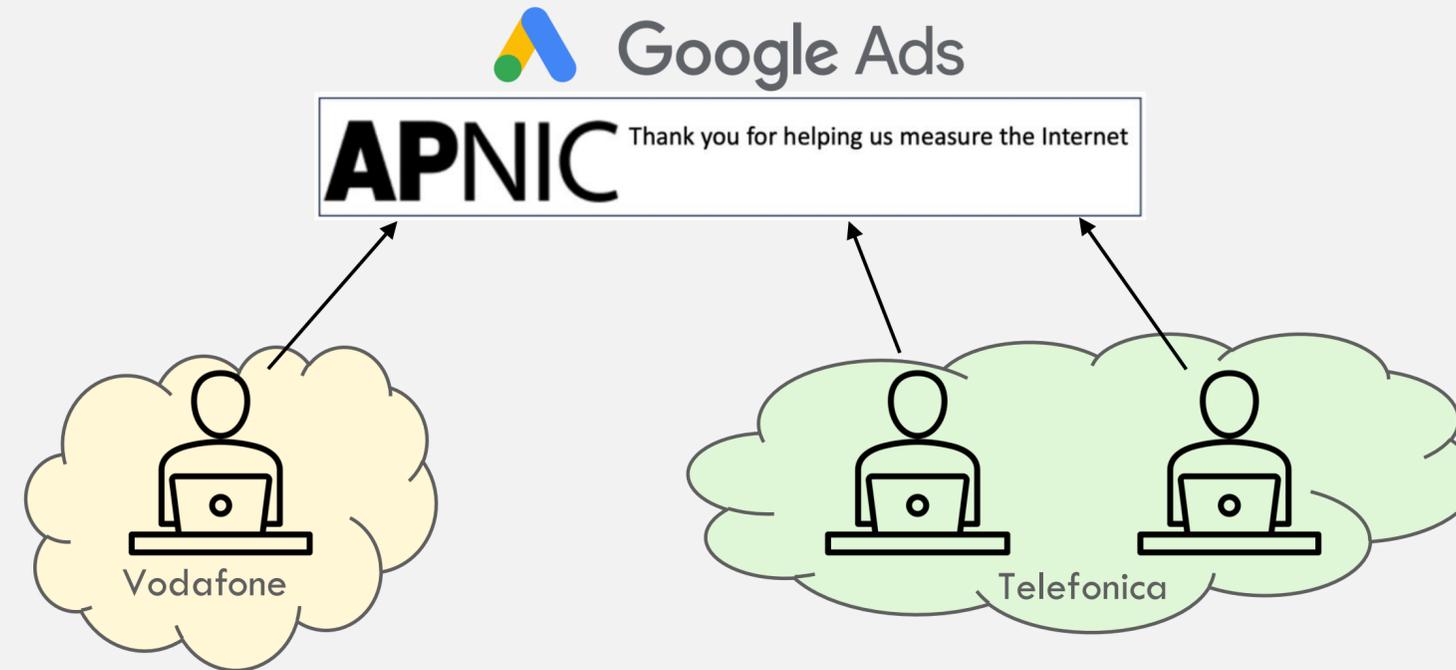
# The APNIC AS Customers Population Dataset



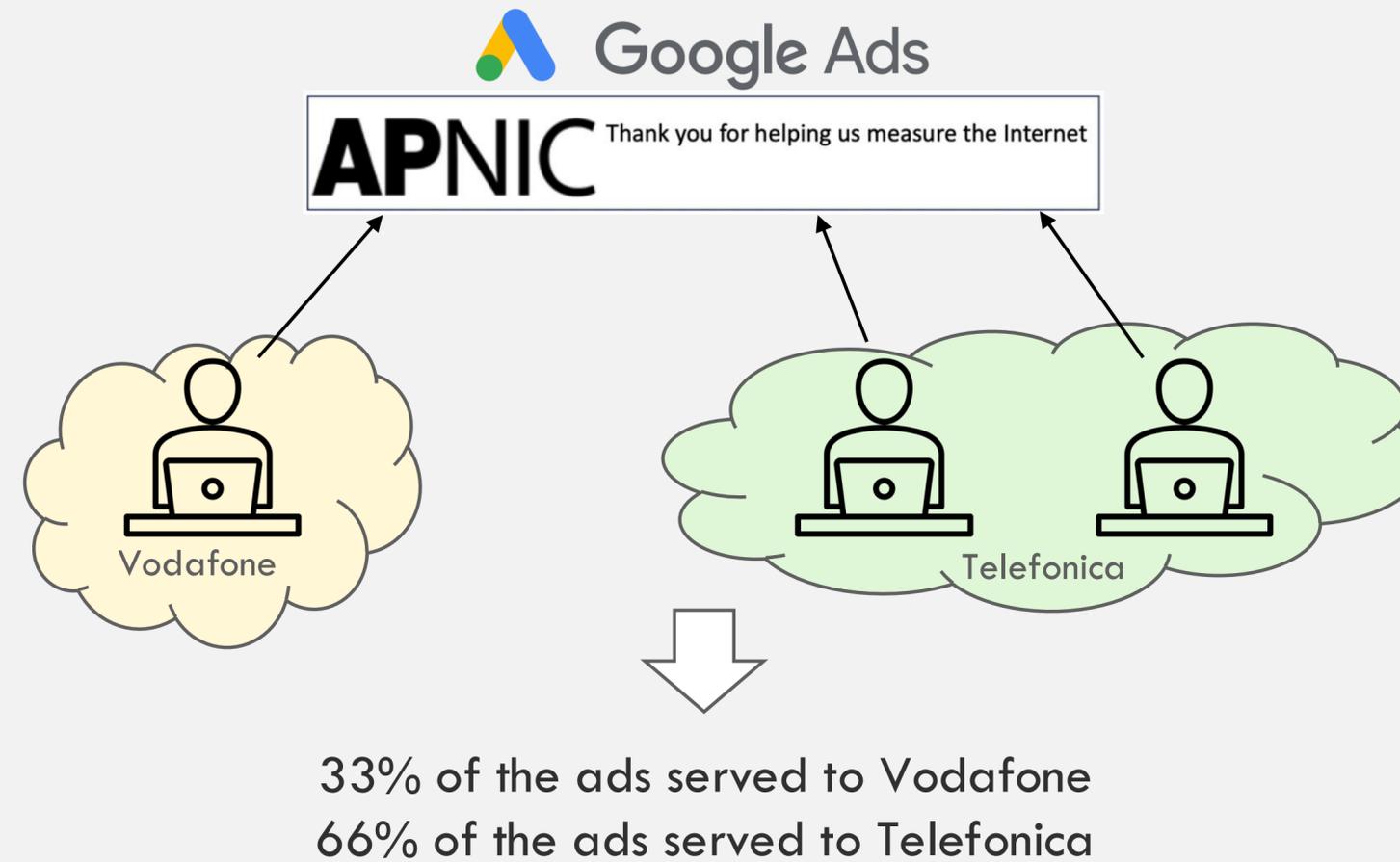
# The APNIC AS Customers Population Dataset



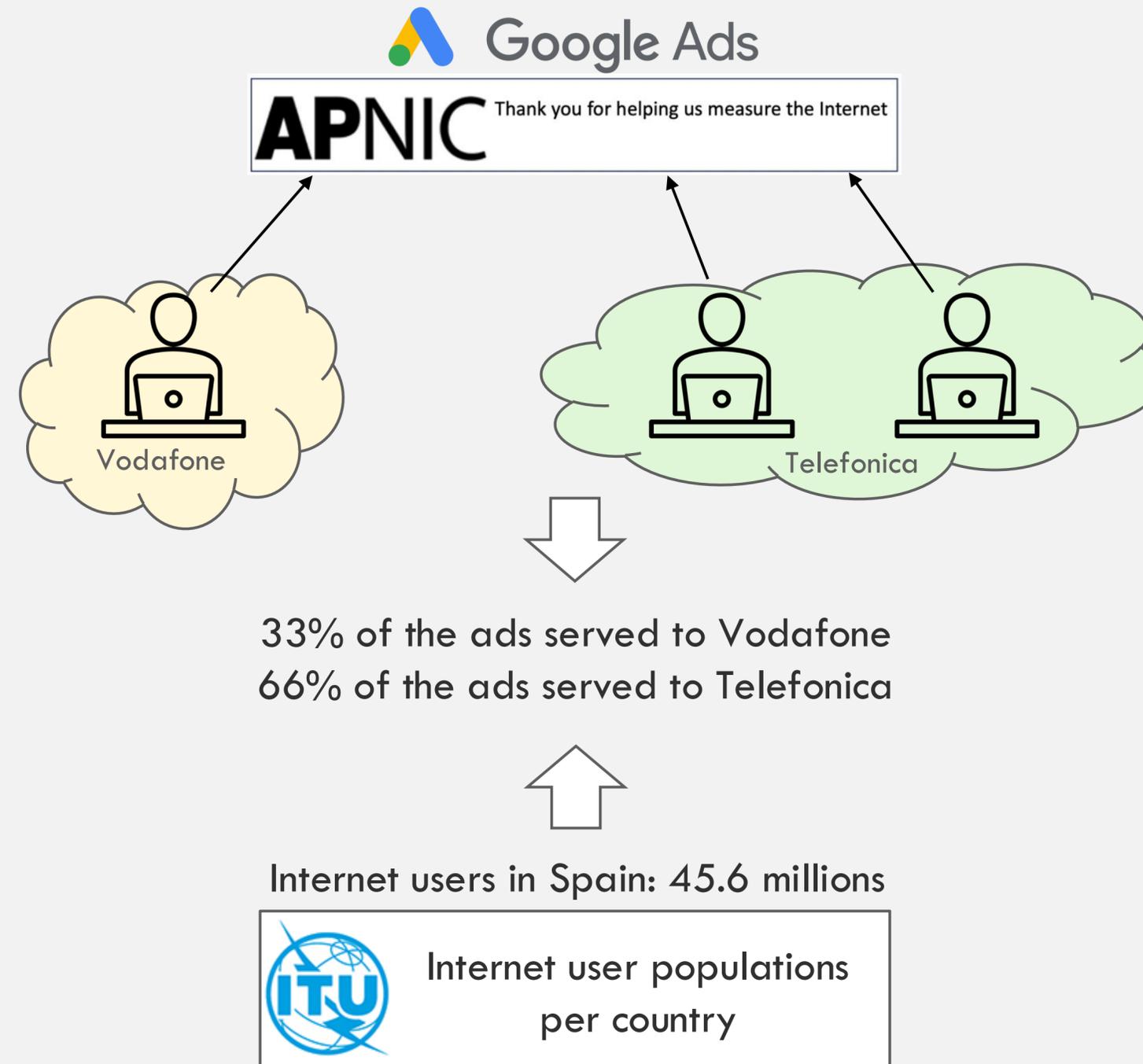
# The APNIC AS Customers Population Dataset



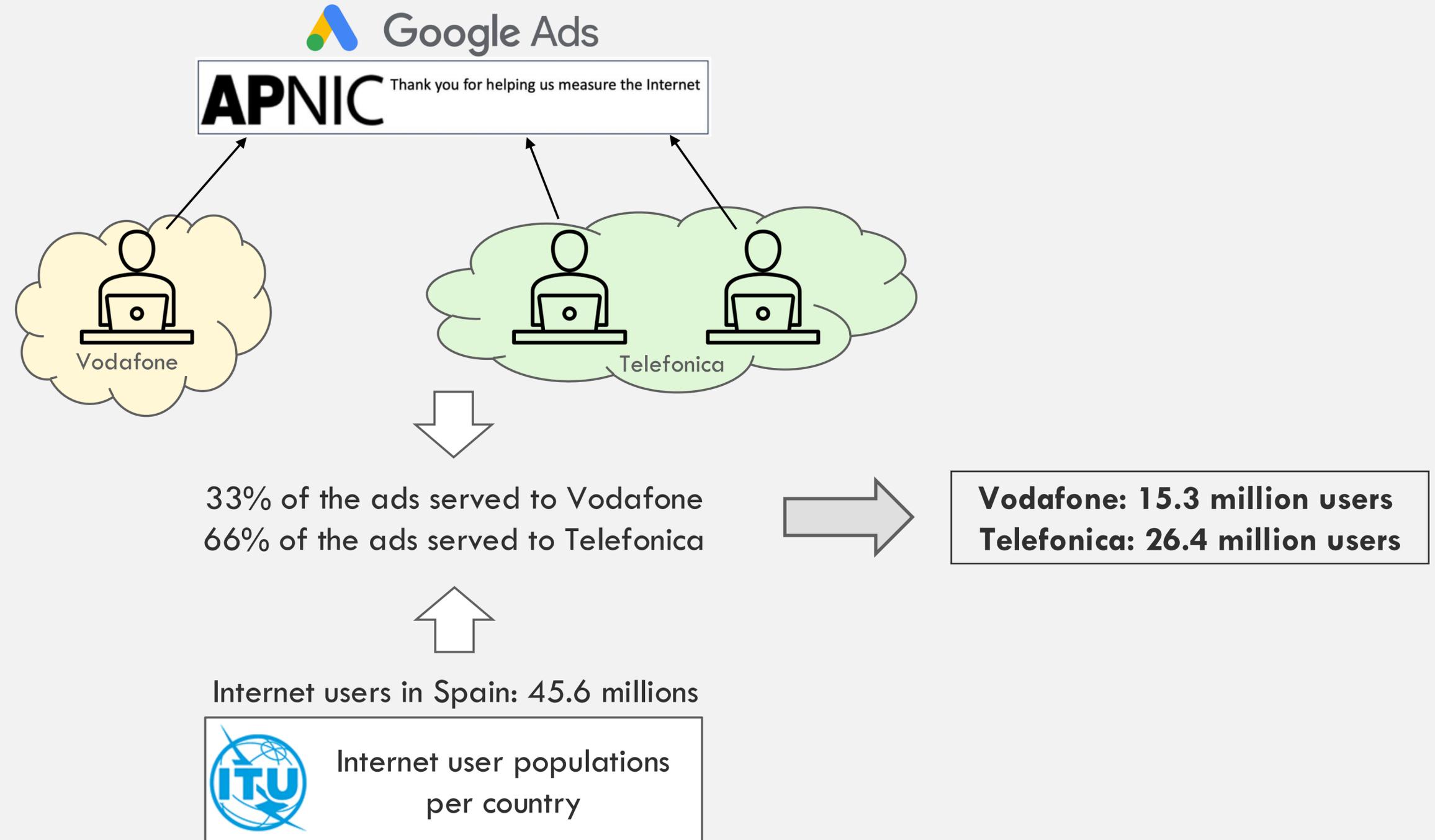
# The APNIC AS Customers Population Dataset



# The APNIC AS Customers Population Dataset



# The APNIC AS Customers Population Dataset



# Potential Biases



- **Non-Uniform Ad Placement:** Google Ads' reach varies across countries, potentially leading to inaccurate user estimates where Google services are less prevalent.
- **Accuracy of ITU-T Estimates:** Fluctuations in ITU-T's Internet user estimates can impact the APNIC dataset's accuracy.
- **Incorrect IP geolocation.**
- **Incorrect IP-to-AS mapping.**

# To understand the APNIC dataset accuracy, we combine four data sources



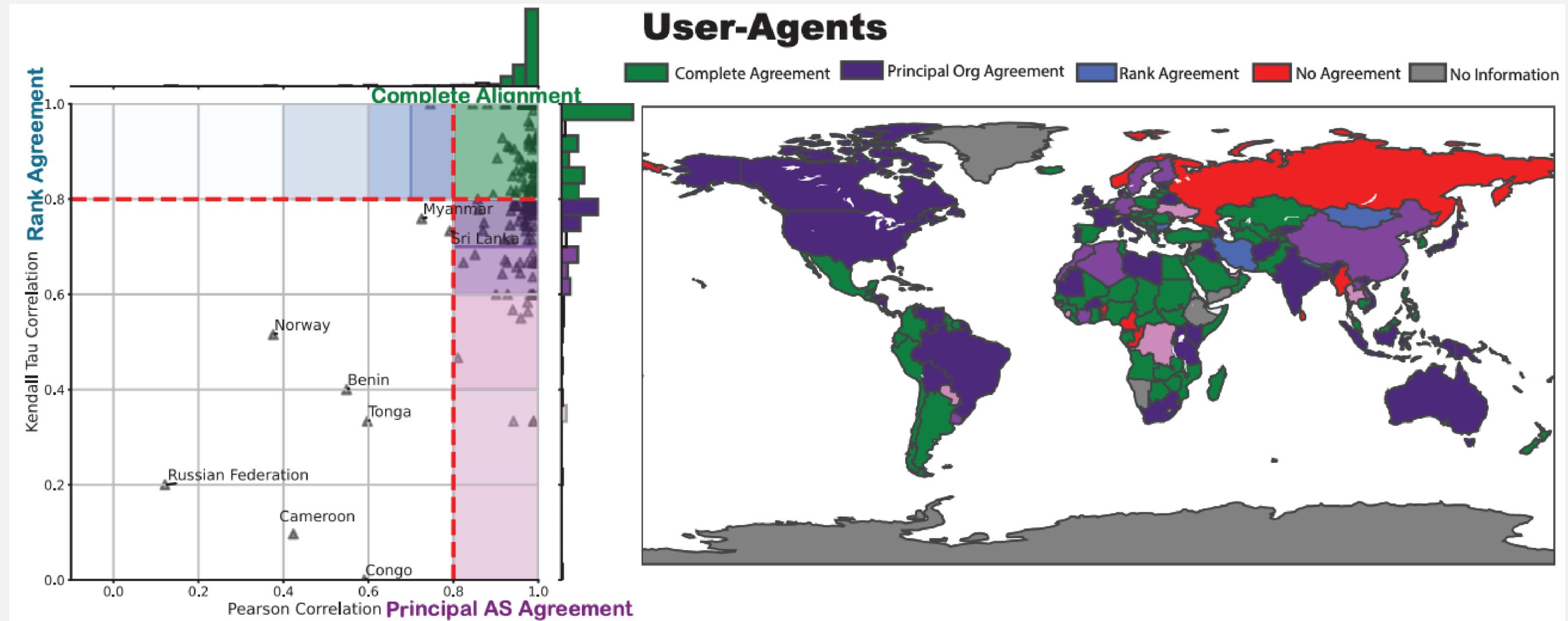
Provider	Data	Availability
Global CDN	Distinct User Agents per ASN	Proprietary
	HTTP traffic per ASN	Proprietary
Manual survey	Broadband subscribers	Public
PeeringDB	Cumulative IXP peering Capacity	Public
M-Lab	Speed tests per ASN	Public

# Comparing APNIC user estimates with the Broadband Subscribers survey



- There is a strong linear correlation for 14/20 (70%) countries.
- Poor correlation for Russia, Brazil, Japan, Poland, South Korea, China
- APNIC tends to overestimate the number of users for mobile broadband providers
  - Ad blocking is less popular in mobile devices

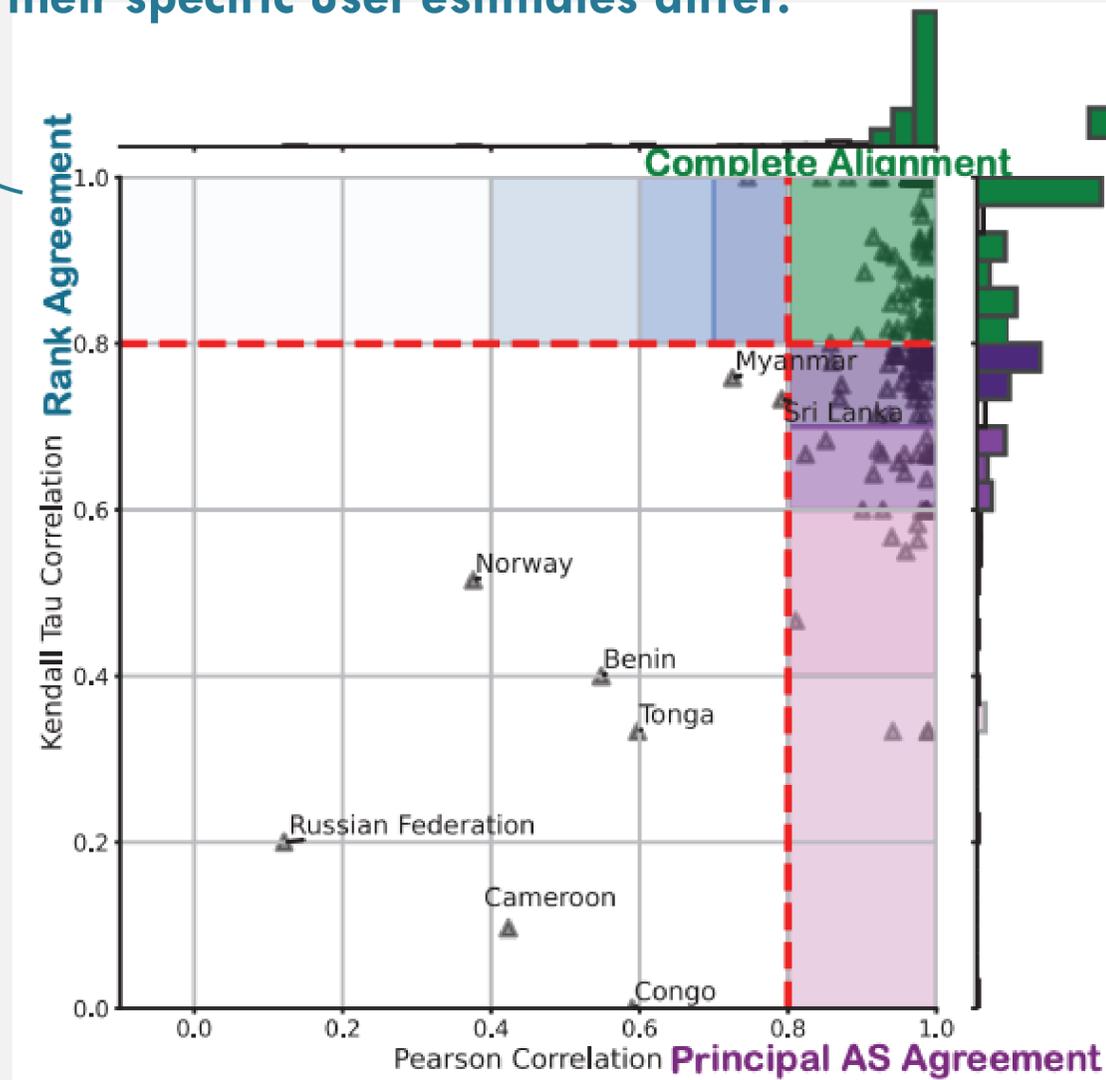
# Evaluating the agreement between the APNIC and the CDN datasets at the country level



# Evaluating the agreement between the APNIC and the CDN datasets at the country level

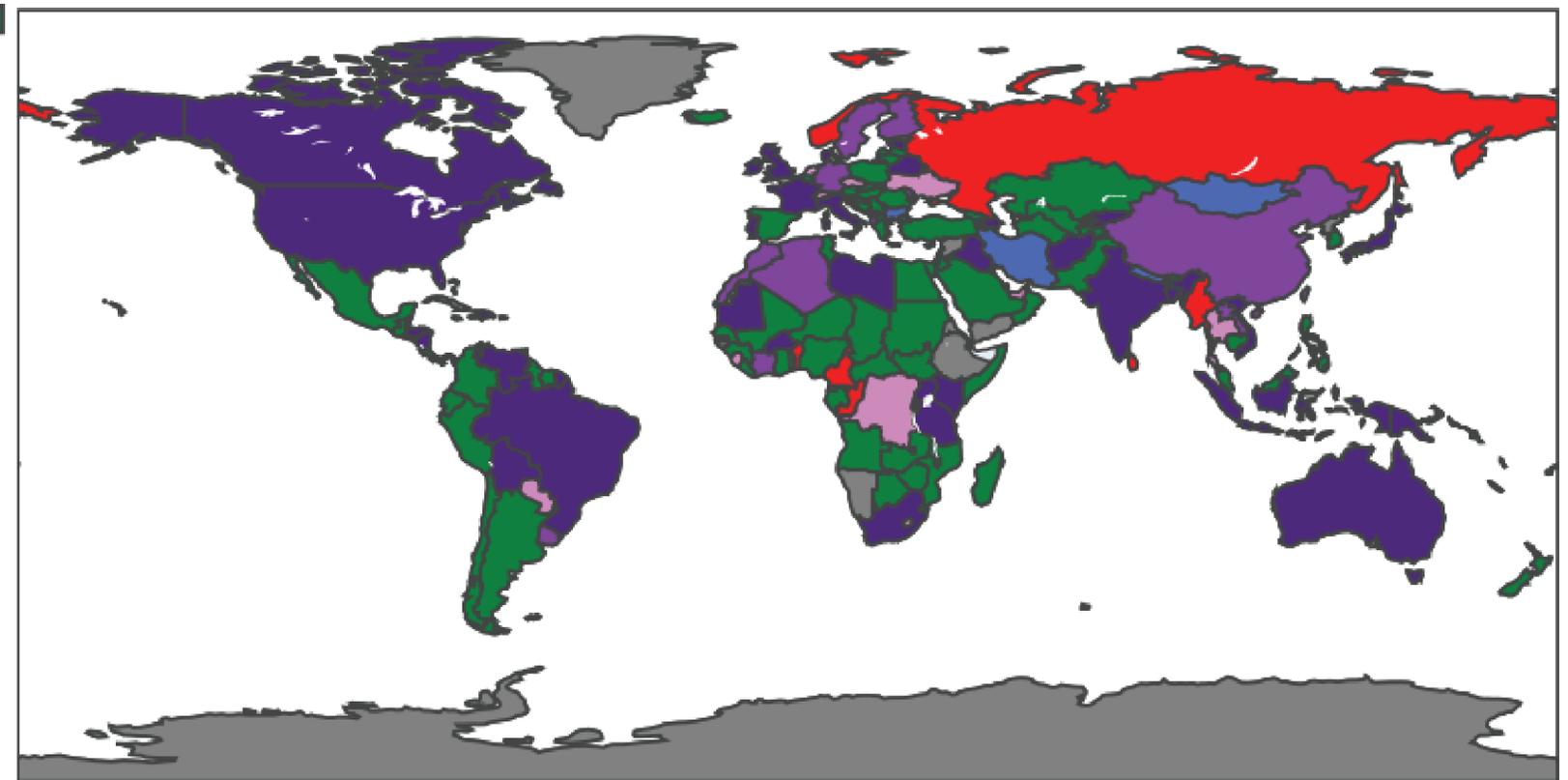


Both datasets identify similar organization order, even if their specific user estimates differ.



## User-Agents

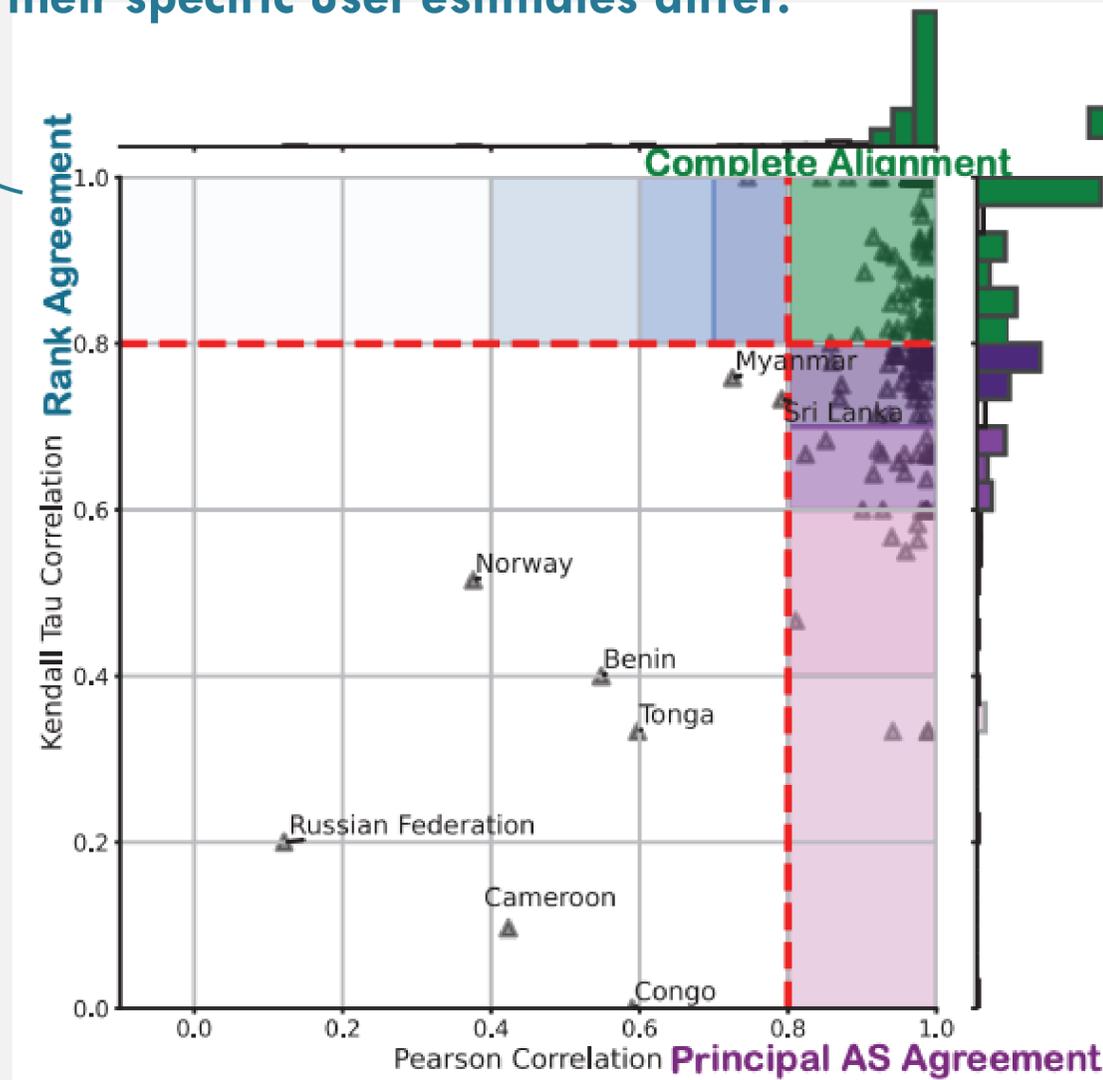
Complete Agreement Principal Org Agreement Rank Agreement No Agreement No Information



# Evaluating the agreement between the APNIC and the CDN datasets at the country level

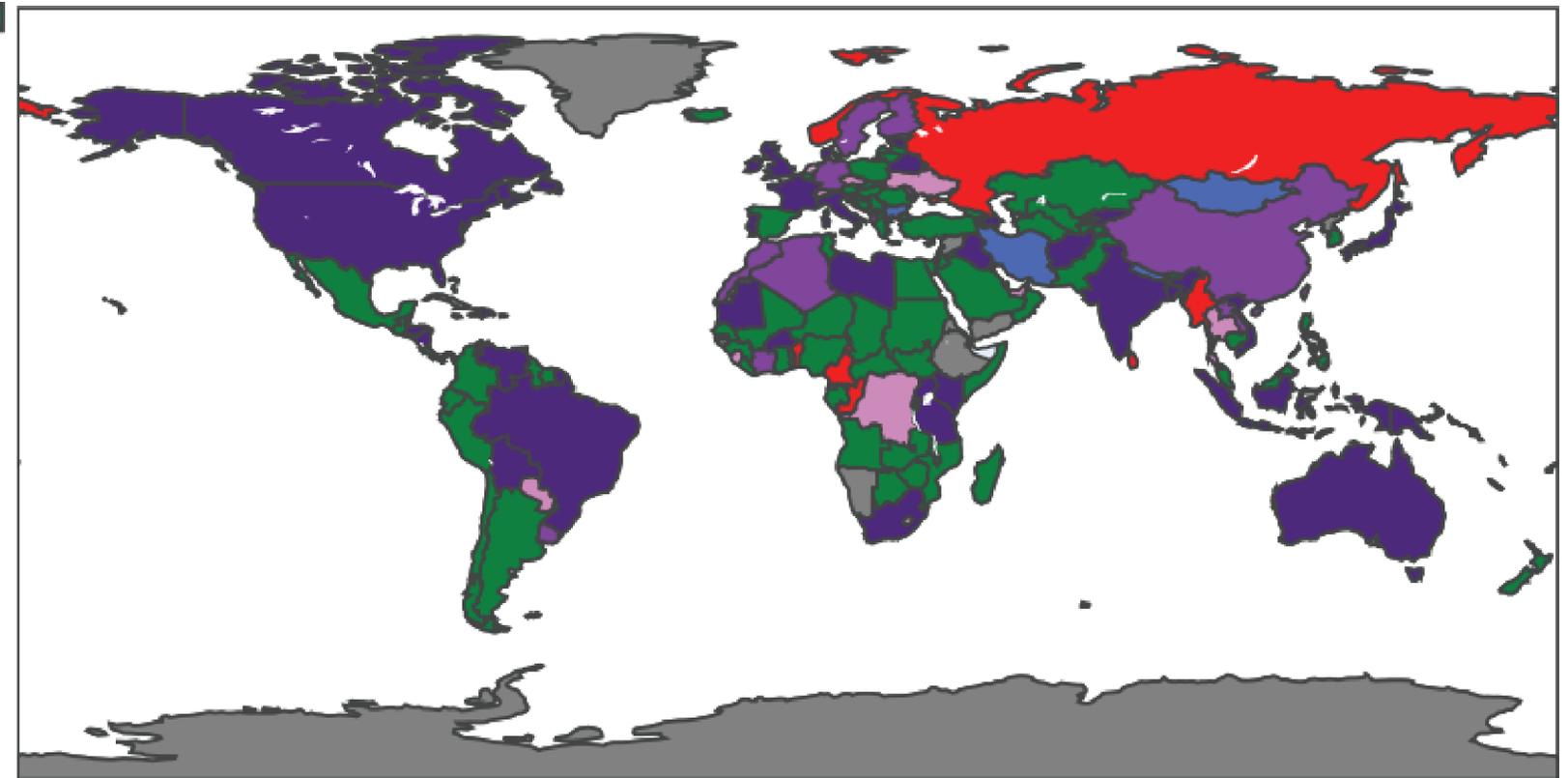


Both datasets identify similar organization order, even if their specific user estimates differ.



## User-Agents

Complete Agreement Principal Org Agreement Rank Agreement No Agreement No Information

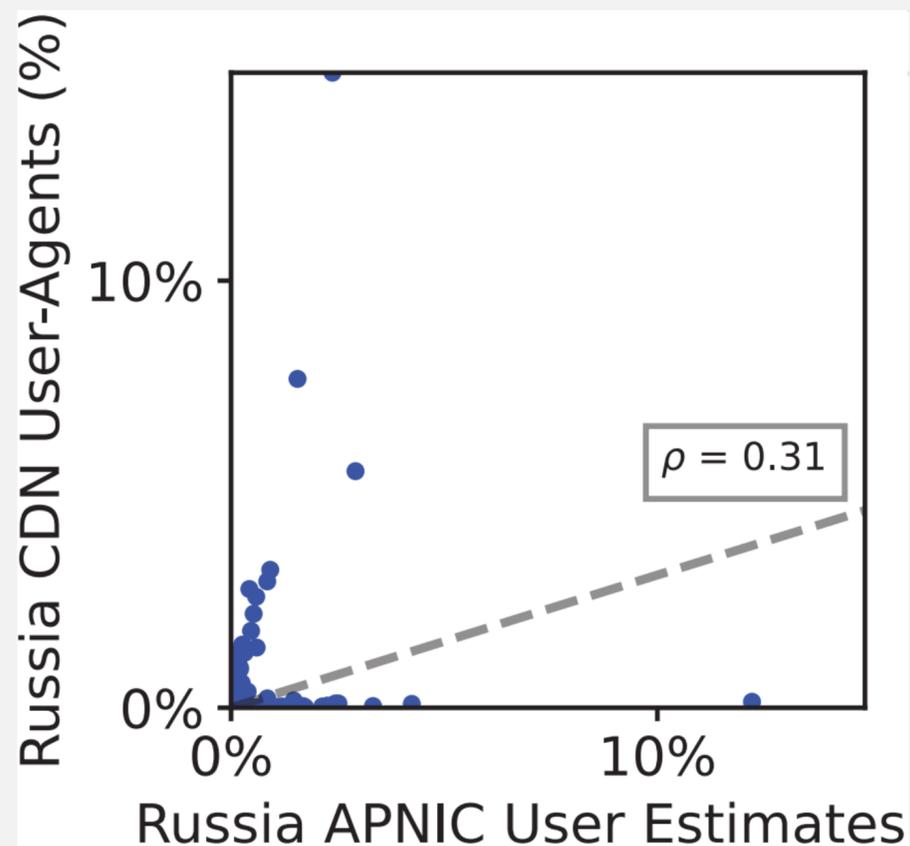


Agreement between the two datasets regarding the most significant networks within each country

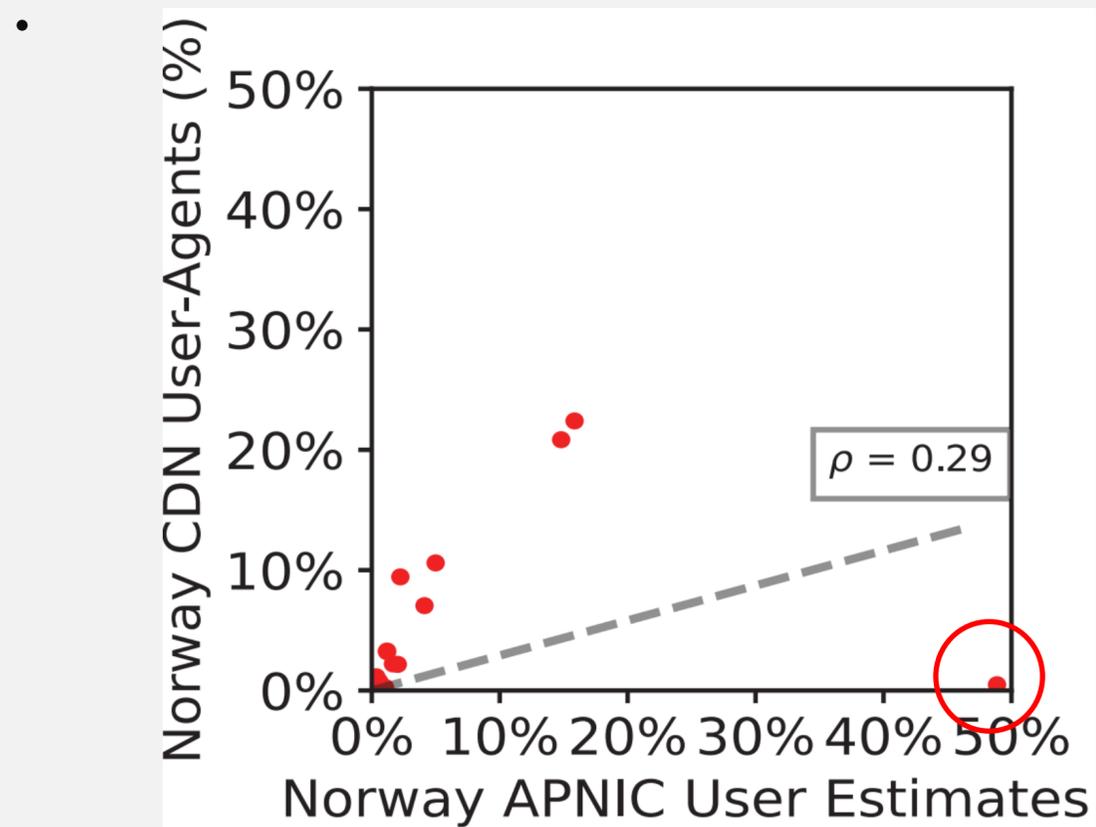


# Understanding the outliers

**Russia:** Discrepancies are due to Yandex's market dominance, Russia's isolated Internet efforts, and Google's reduced presence following the Ukraine conflict.



**Norway:** Overrepresentation is caused by VPN traffic routing through a few IP addresses in Norway, leading to misinterpretation in the APNIC data.





# How to make the best use of APNIC's estimates?

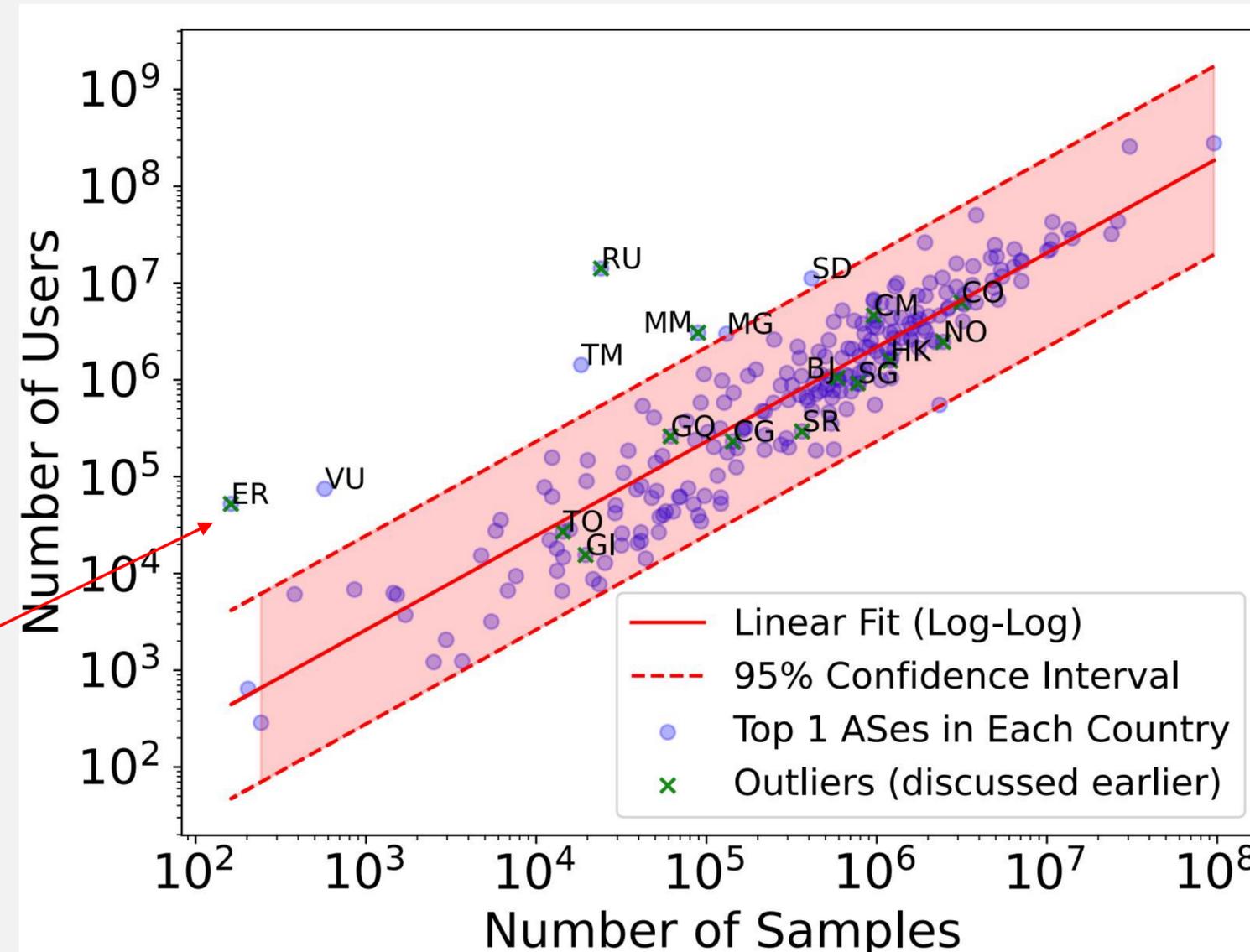
## Take 3 points into consideration

- The number of ad samples per estimated user can impact the accuracy
- The APNIC estimates are computed daily, but the published numbers are averaged over a period of 60 days
- Geolocation seems to be responsible for misplacing many users to the wrong countries

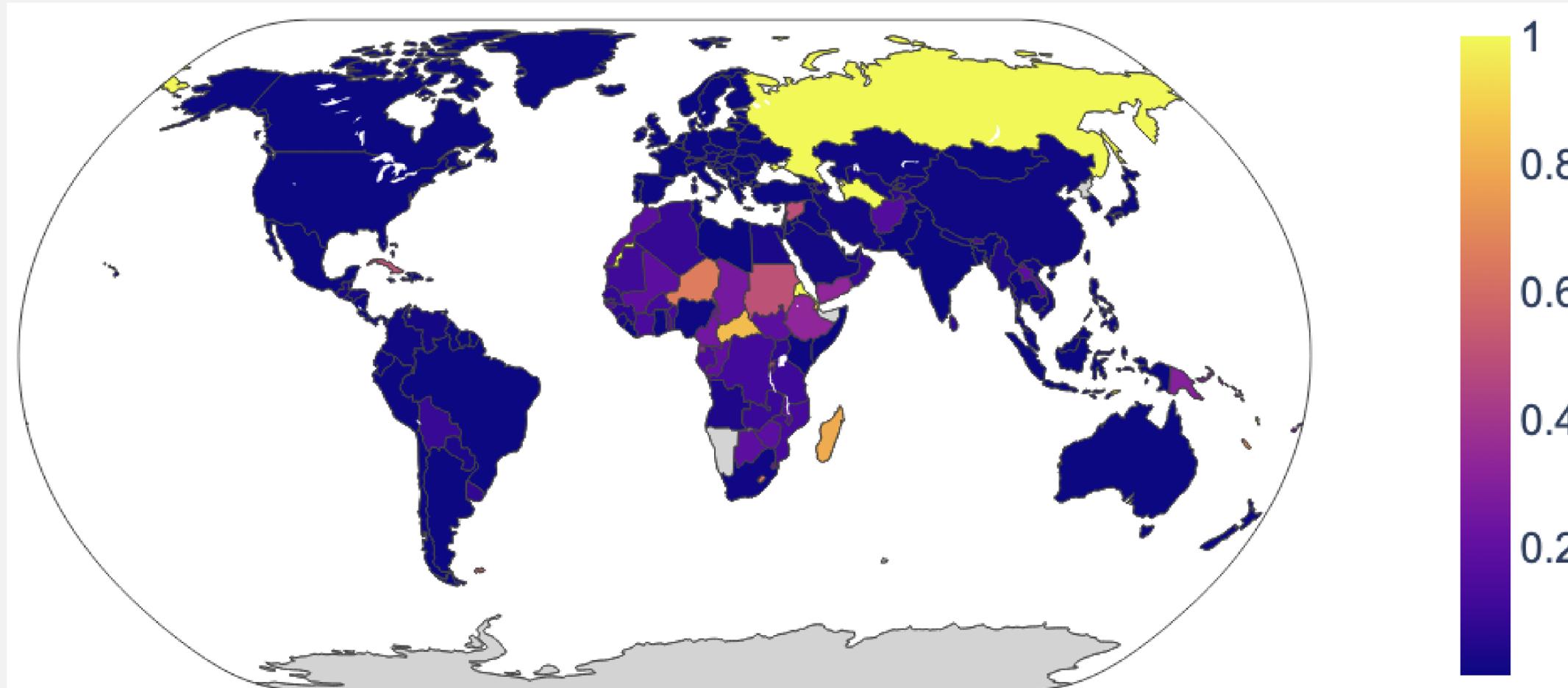
# Users-to-samples ratio per country



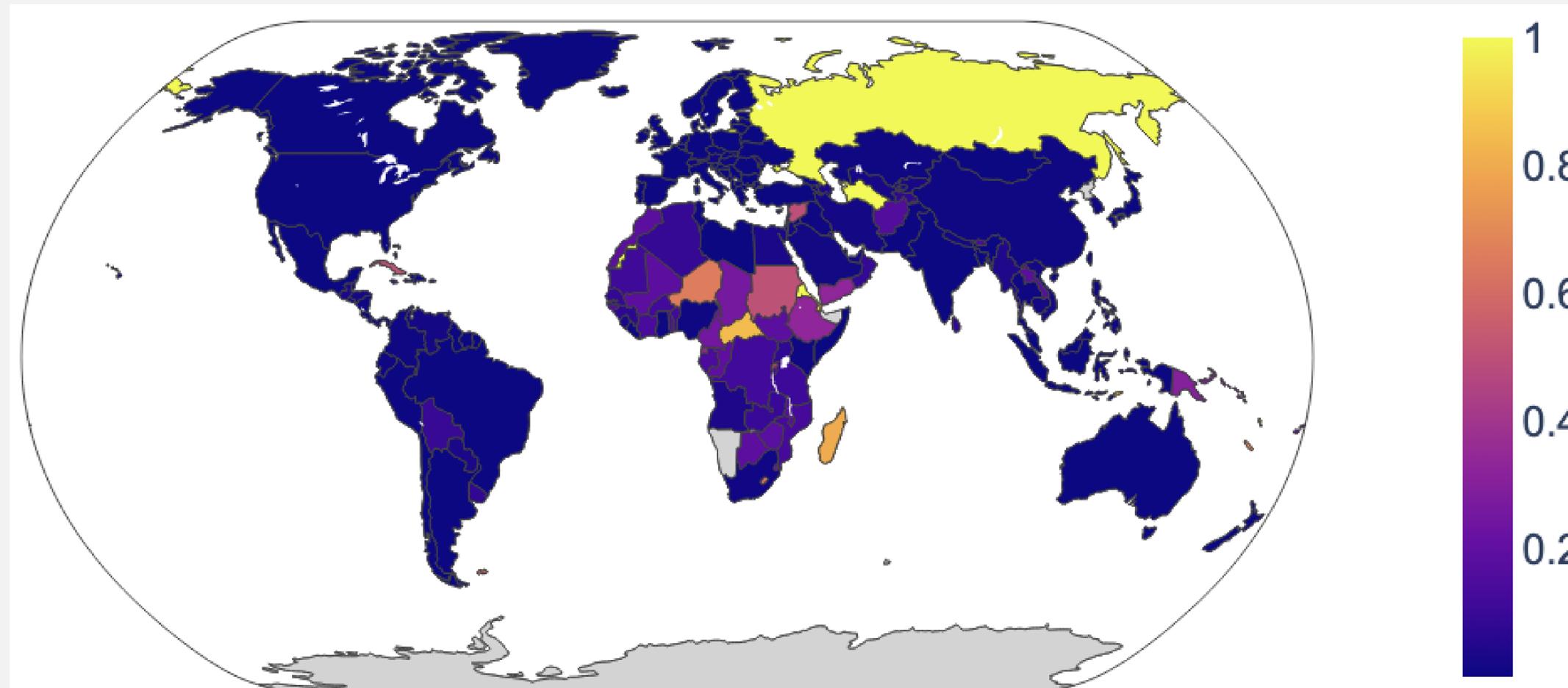
Each sample point in Eritrea maps to  $\approx 1000$  users



# Fraction of days across 2024 where the User-to-Sample ratio did not lie in the estimated confidence interval



# Fraction of days across 2024 where the User-to-Sample ratio did not lie in the estimated confidence interval

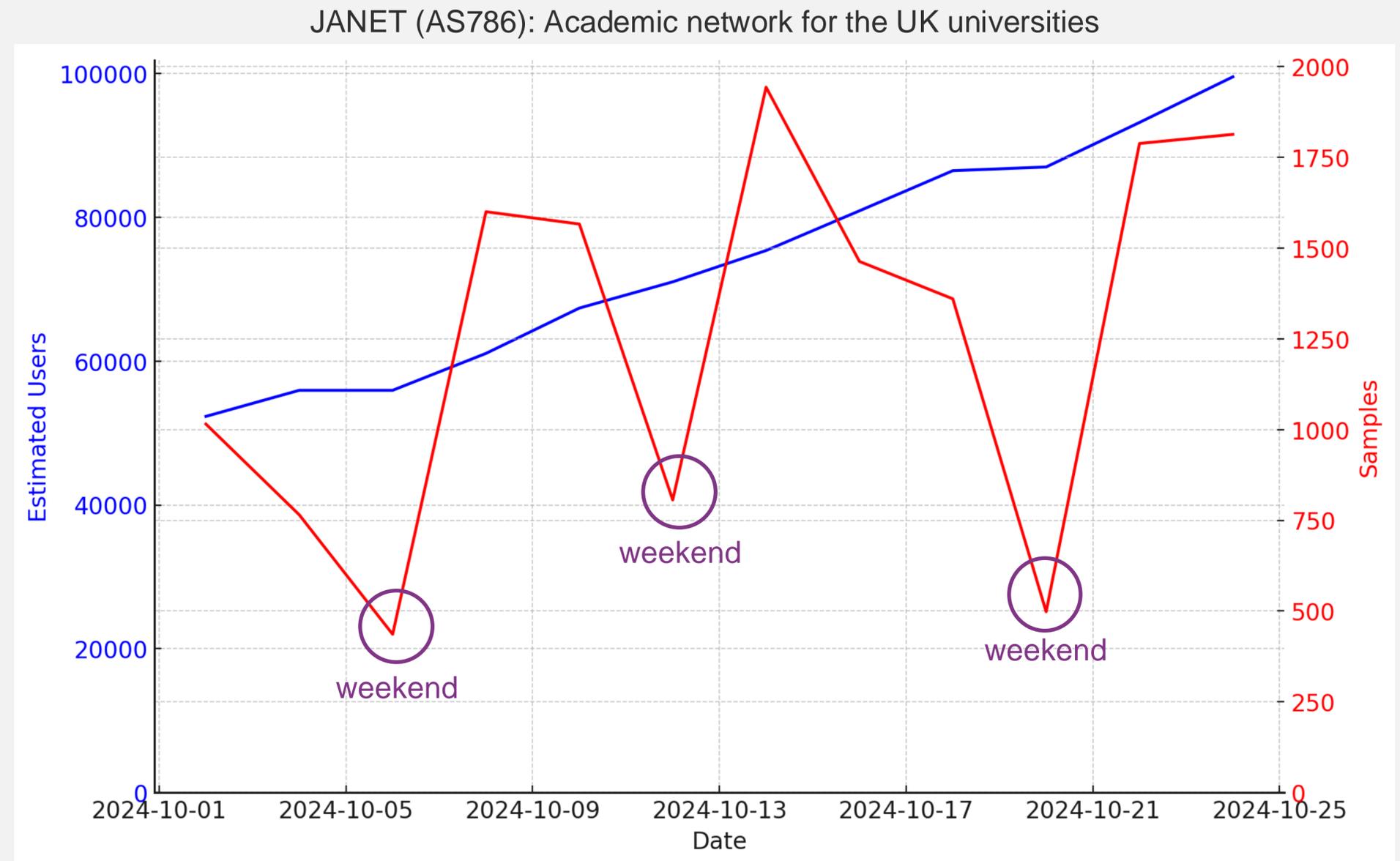


**Recommendation:** For countries with large fluctuations, select the day with the lowest users-to-samples ratio in a 60-day period

# Consider the 60-day smoothing if you need user estimates at finer granularity



- Consider JANET (UK academic network)
- The **samples** reflect clearly the weekends when universities have fewer people
- Why does the number of estimated users is monotonically increasing?



# Conclusion



- APNIC dataset works well in countries with sufficient Google Ads data.
- Accuracy improves by verifying the user-to-sample ratio.
- Understating nuances is important to avoid misuse of data.
- Ensure APNIC data aligns with external datasets discussed in the paper (IP Geo, PeeringDB, M-Lab).



[https://github.com/Burdantes/unboxing\\_apnic](https://github.com/Burdantes/unboxing_apnic)