Phone subscriber data
✓ As of January 2019
✓ 29 million persons
✓ 380 thousands surnames
VARIATION OF SURNAMES

Variation = \frac{\text{Number of the unique surnames}}{\text{Number of subscribers in the region}}

It highly depends on population.

In a large region, we expect a large variety of surnames, but the above formula negates it because of the high denominator.
There are 30 pupils in the class. If they have non-repeating surnames, then there are 30 different surnames. $30 \text{ surnames} / 30 \text{ pupils} = 1$.

If there are 16 Ivanovs, then we get 15 different surnames ($14 \text{ are not Ivanovs} + \text{Ivanov}$). $15 / 30 = 0.5$. It means that the larger the index, the more variation of surnames.

When the number of pupils is the same, there are no problems with comparing the indices. But if the classes are different by number of pupils, the index does not work.

How to calculate the variation that could be compared between different classes?
VARIATION OF SURNAMES

The smallest region has 7000 subscribers.

We are randomly picking up 7000 subscribers from each region.

We repeat the procedure one thousand times.

We calculate the average and median number of distinctive surnames in each region.

We divide the median to the subsample value (seven thousand).
The top-5 surnames in some regions of Russia
In the top 1000 in the regions:
Omsk
Kurgan
Kemerovo
Chelyabinsk
Sverdlovsk
Zabaikalsky Krai

2970th place in Bashkiria
CONCLUSIONS

The regions with a high variety of surnames are located within the Russian main belt of settlement, which has fostered a more active exchange and admixture of the population.

Our analysis has shown a diverse palette of surnames in Russia, which is determined by

- various geographical landscapes,
- historical pathways of settlement,
- ethnic, cultural and religious plurality.