



Does the Post-Soviet Space Still Exist?

An Unsupervised Machine Learning Approach to Classifying Countries

Minsk 2021

The Question

MOTIVATION BEHIND RESEARCH: The upcoming 30th anniversary of the 1991 Belavezha Accords is again attracting the attention of researchers and policymakers to the topic of the post-Soviet transformation. The question about the consistency of this region's characteristics compared to other countries and regions or their conceptual differences, based on the theory of institutional change and persistence, is still not fully answered.

Benefiting from available open-source data and machine learning (ML) algorithms, we propose to take a new look at this question and **test two main hypotheses.**

TWO HYPOTHESES TO TEST:

H1: The Post-Soviet Space Still Exists

ACCEPT

If the ML algorithm classifies all or most of the post-Soviet states into a separate group(s), we will have solid empirical grounds to continue applying the concept of the post-Soviet states in social science research and in applied analysis.

REJECT

If, however, the findings are different, we will get a factual confirmation of the increasingly popular argument about the end of the post-Soviet transit.

H2: The Post-Soviet Space is Homogeneous

ACCEPT

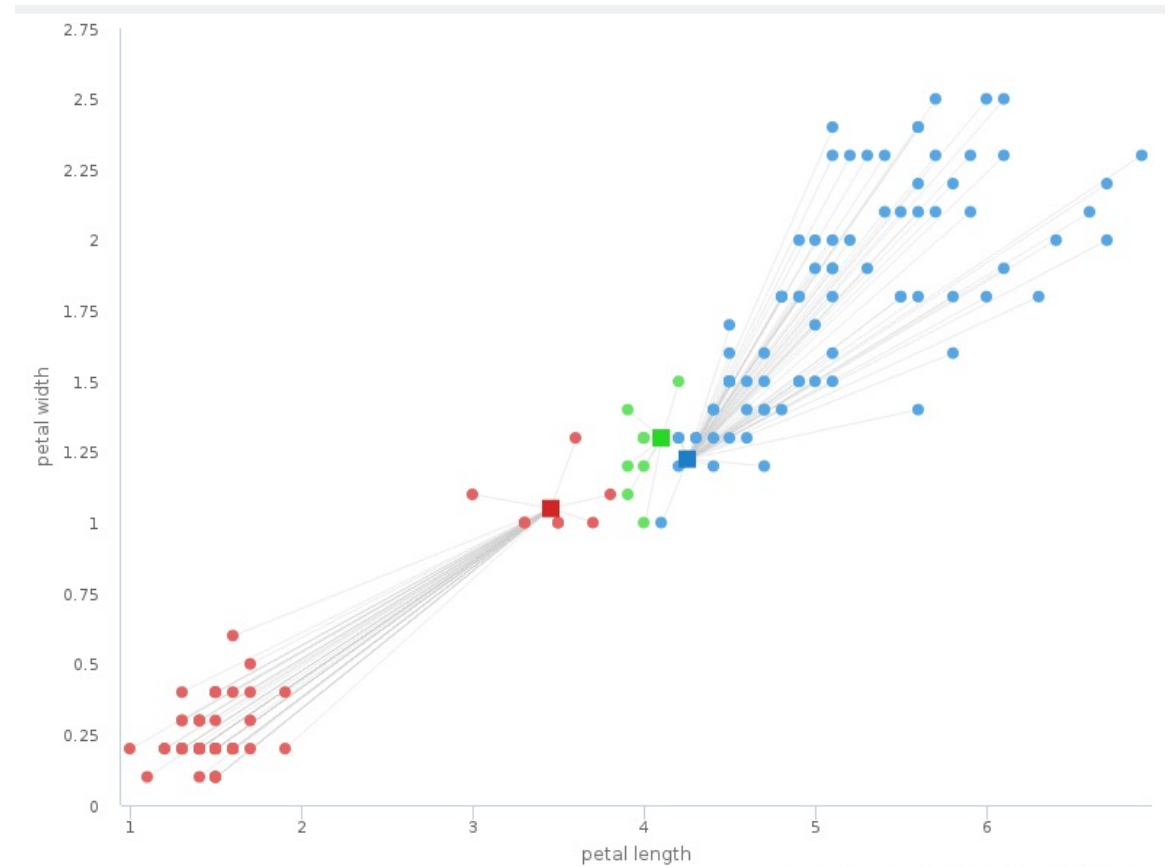
The Silhouette coefficients do not allow us to choose different clusters when we apply the ML algorithm to 15 post-Soviet states.

REJECT

There are different clusters among 15 post-Soviet states, defined by the ML algorithm.

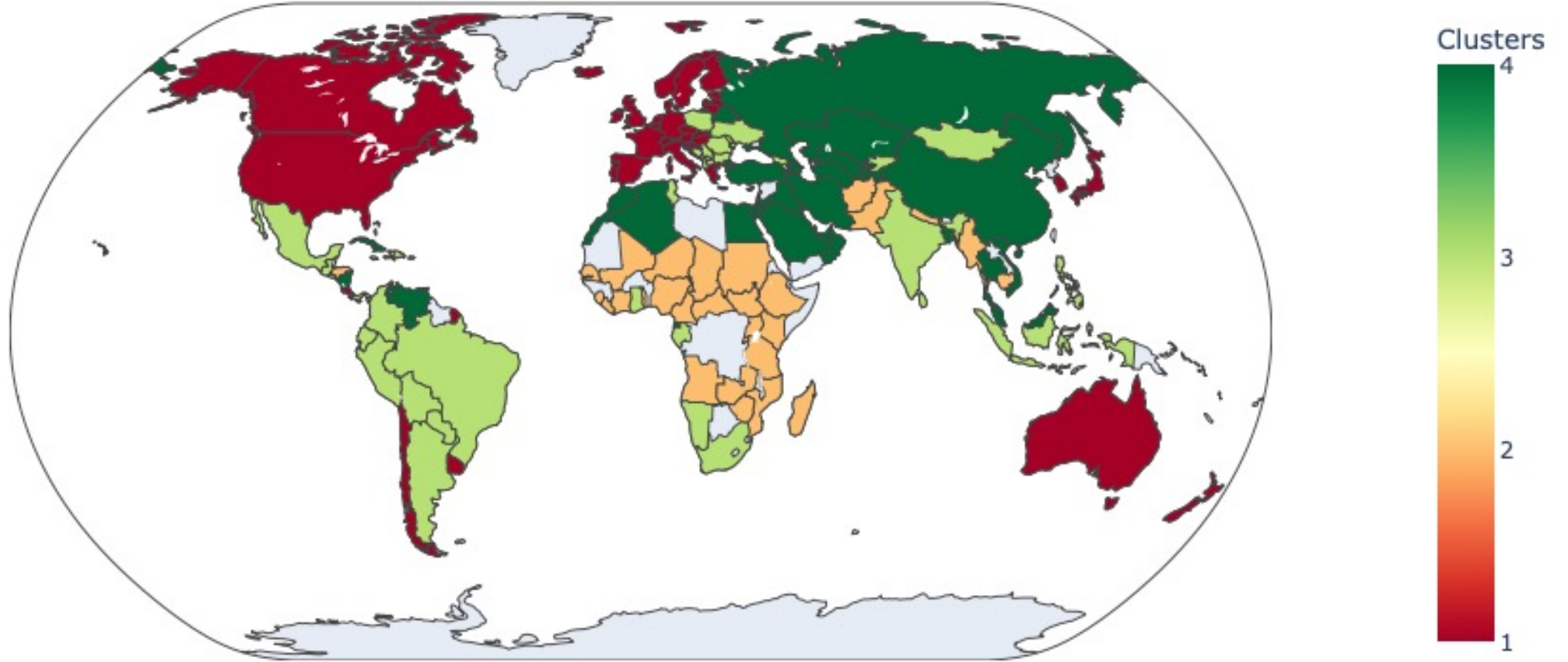
Methodology

Unsupervised ML algorithms (k-Mean++) is used to define data-driven clusters of countries. The algorithm is informed by a vector of country-specific characteristics. To define a list of characteristics, the method of an expert elicitation has been applied.



H1: The Post-Soviet Space Still Exists

H1: Geographical Visualization of Clusters



H1: Homogeneity of Post-Soviet Region

The machine learning algorithm (K-Mean++) assigned all 15 post-Soviet countries into three separate clusters (out of four). The assignment is based on parameters closely related to the main characteristics of the post-Soviet area. As a result, we have empirical ground to **reject Hypothesis 1 (H1)** and **accept the counterhypothesis about the end of the post-Soviet transit.**

CLUSTER 1 (40 COUNTRIES)

Estonia	<ul style="list-style-type: none">- Countries with the highest GDP per capita, but the lowest GDP growth rate;- Service economies where an agro sector plays a minor role;- High government expenditures and government debt;- Low level of shadow economy and income inequality;- Developed democratic institutions;- Majority of people have higher education;- Mostly Christian religion but with the highest share of atheists;- Low level of military expenditures.
Latvia	
Lithuania	

CLUSTER 2 (31 COUNTRIES)

	<ul style="list-style-type: none">- Least developed countries with low GDP per capita;- 65%, in average, of population employed in the agro sector;- Extremely high level of poverty;- Decreasing population;- 60%, in average, of rural population;- Lowest share of people with high education;- Highly religious societies;- Societies with a male dominance and low women participation;- High level of military expenditures.

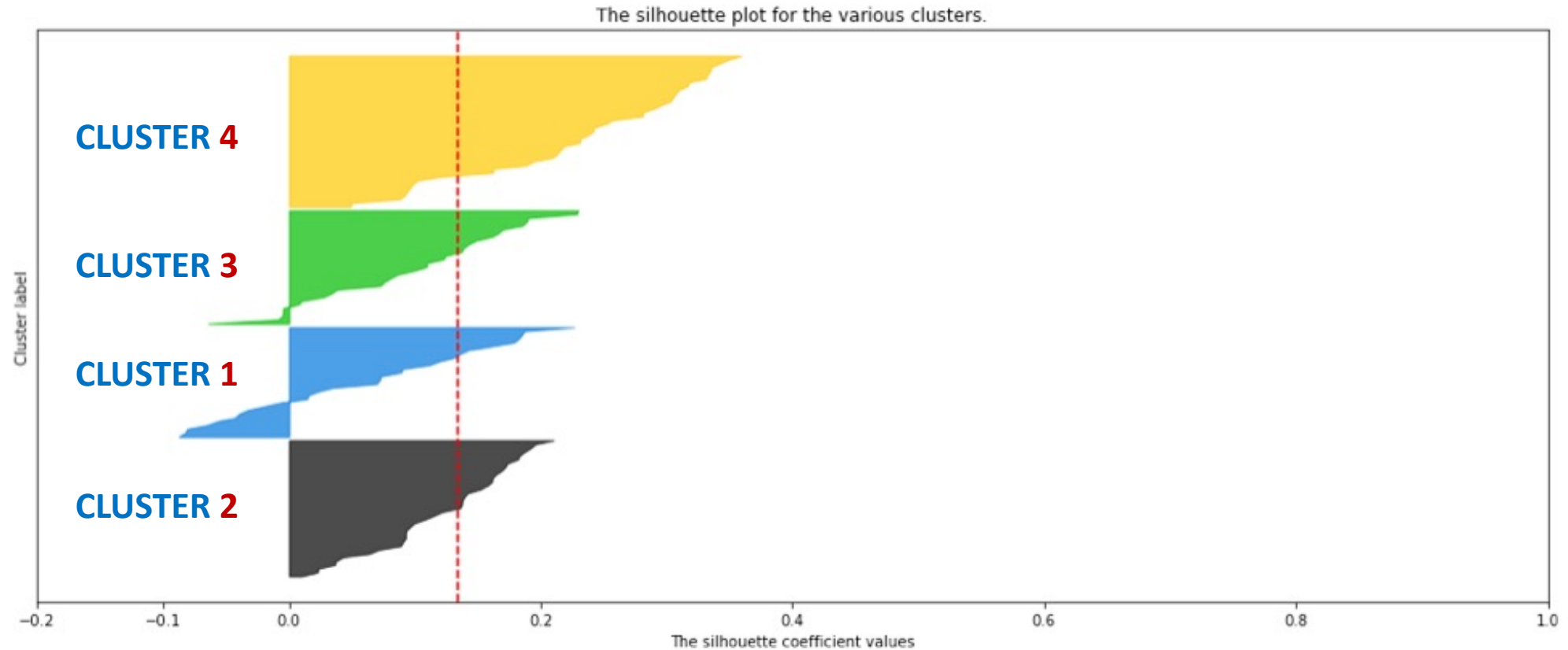
CLUSTER 3 (39 COUNTRIES)

Armenia	<ul style="list-style-type: none">- Countries with an average GDP growth rate, but with less than average GDP per capita;- Low level of participation in economy and thus, high level of shadow economy;- High level of income inequality;- Higher than average level of development of democratic institutions;- Mostly decentralized states;- Average level of military expenditures.
Georgia	
Kyrgyz Republic	
Moldova	
Ukraine	

CLUSTER 4 (29 COUNTRIES)

Azerbaijan	<ul style="list-style-type: none">- Countries with a high share of industrial production, median level of GDP per capita, and average GDP growth;- High participation of government in economy;- The lowest level of development of democratic institutions;- Low level of freedom of religion and expression;- High level of migration;- Mostly patriarchal society with low level of women participation;- The highest level of military expenditures.
Belarus	
Kazakhstan	
Russia	
Tajikistan	
Turkmenistan	
Uzbekistan	

H1: Stability of Clusters



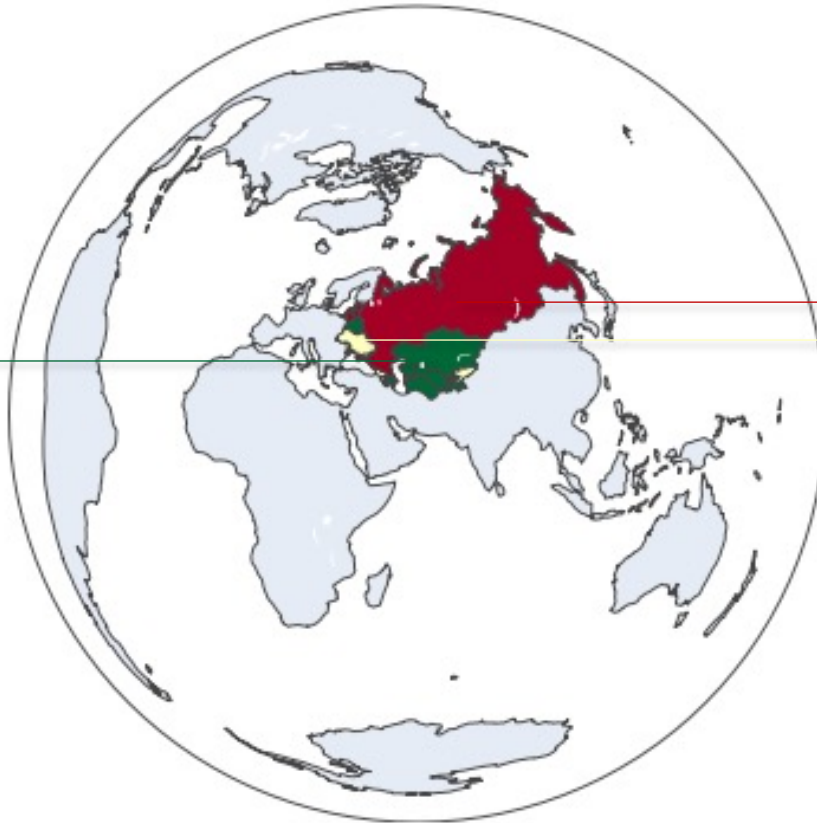
Silhouette coefficients (as these values are referred to as) near 1.0 indicate that the sample is far away from the neighboring clusters. A value of 0.0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster.

H2: The Post-Soviet Space is Homogeneous

H2: Three Groups of Post-Soviet Countries

Geographical Visualization of Clusters (Post-Soviet Area)

The **post-Soviet** region is **not** homogeneous.



WESTERNIZED

Estonia
Latvia
Lithuania
Russian Federation

INTERMEDIATE

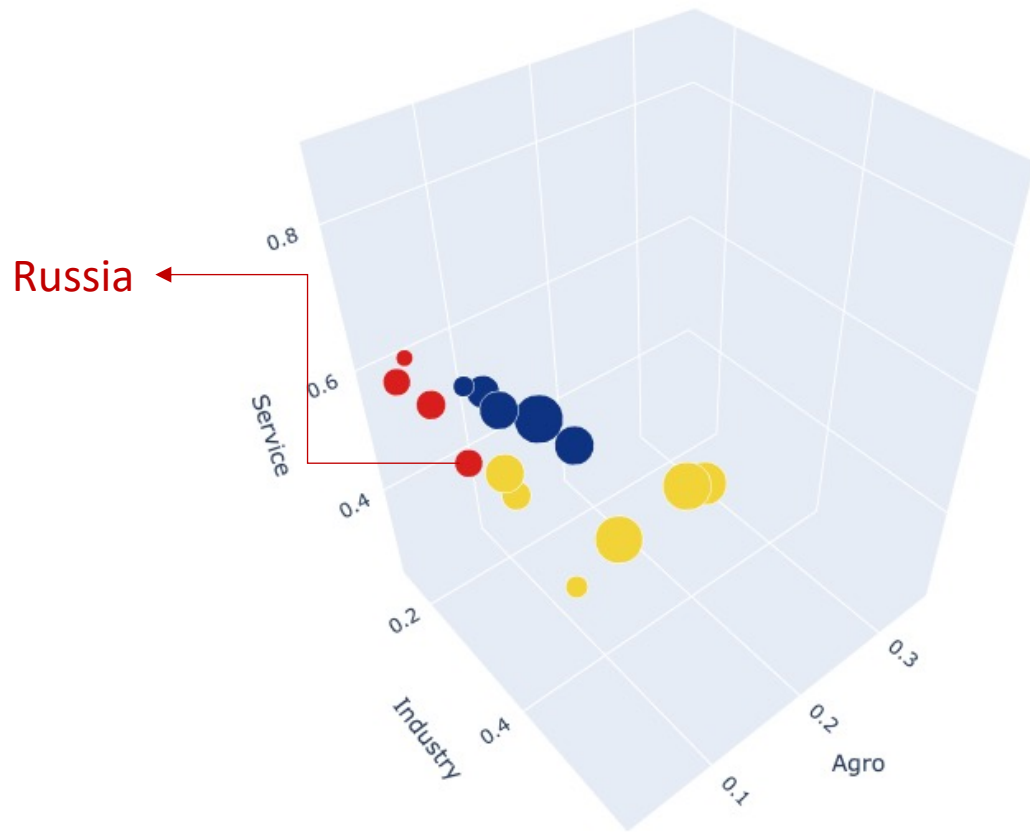
Armenia
Georgia
Kyrgyz Republic
Moldova
Ukraine

CONSERVATIVE

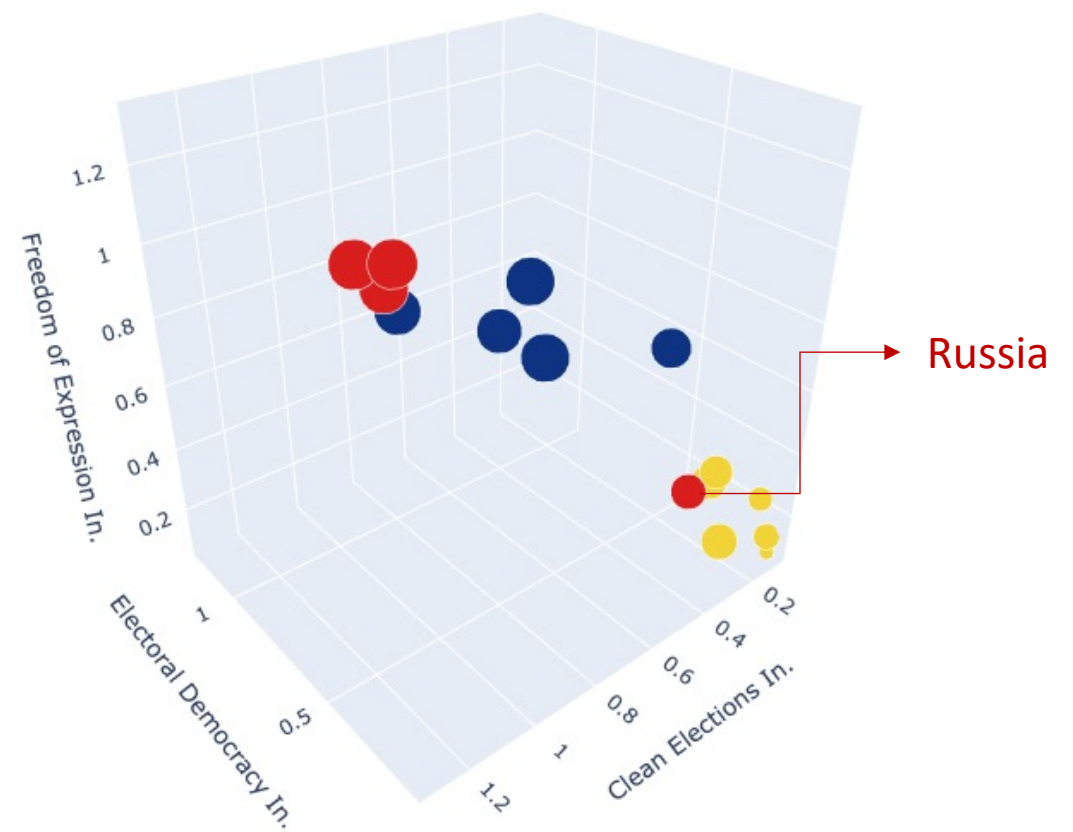
Azerbaijan
Belarus
Kazakhstan
Tajikistan
Turkmenistan
Uzbekistan

H2: The Case of Russia

ECONOMIC STRUCTURE

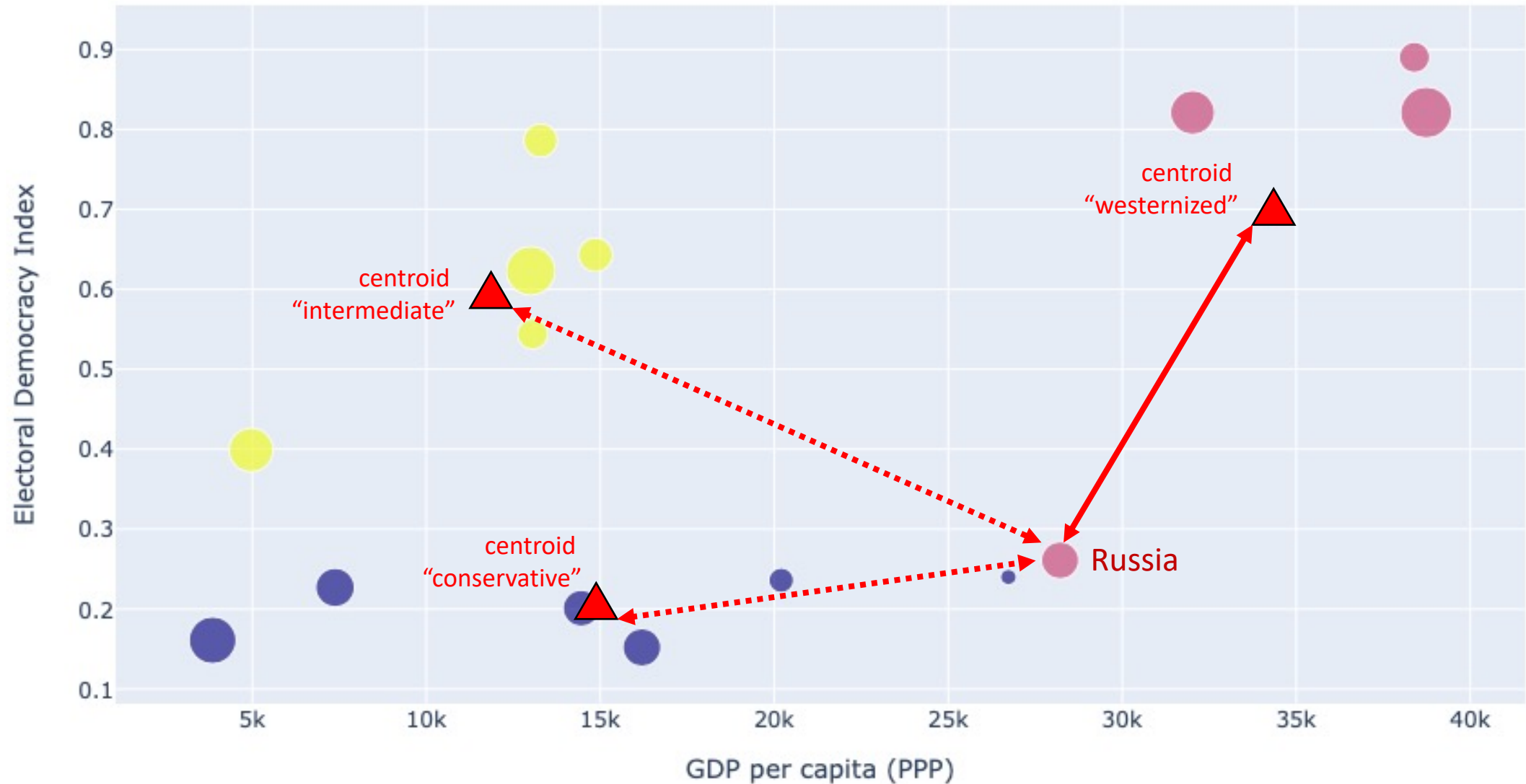


DEMOCRATIC INSTITUTIONS



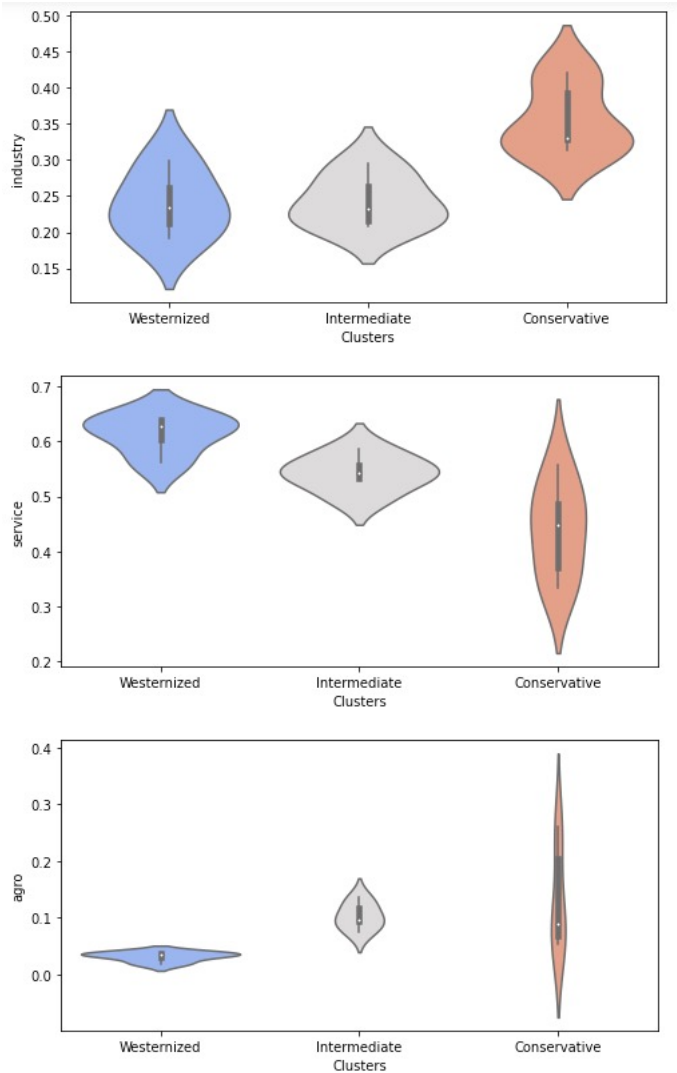
H2: The Case of Russia - 2

GDP per capita (PPP) vs Electoral Democracy Index

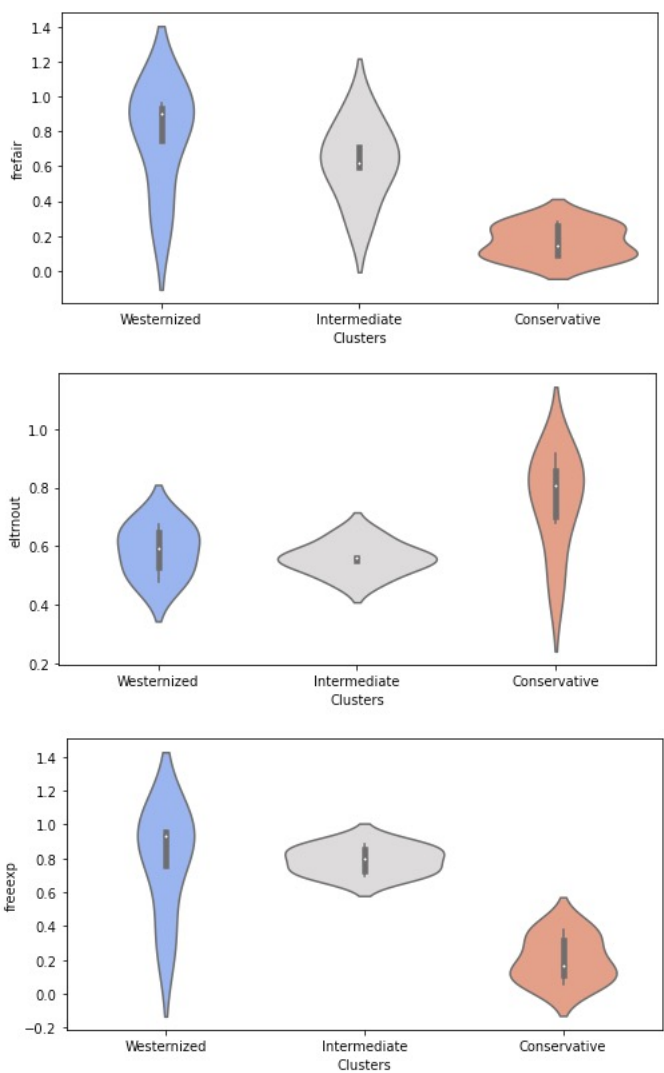


H2: Dissection of Clusters

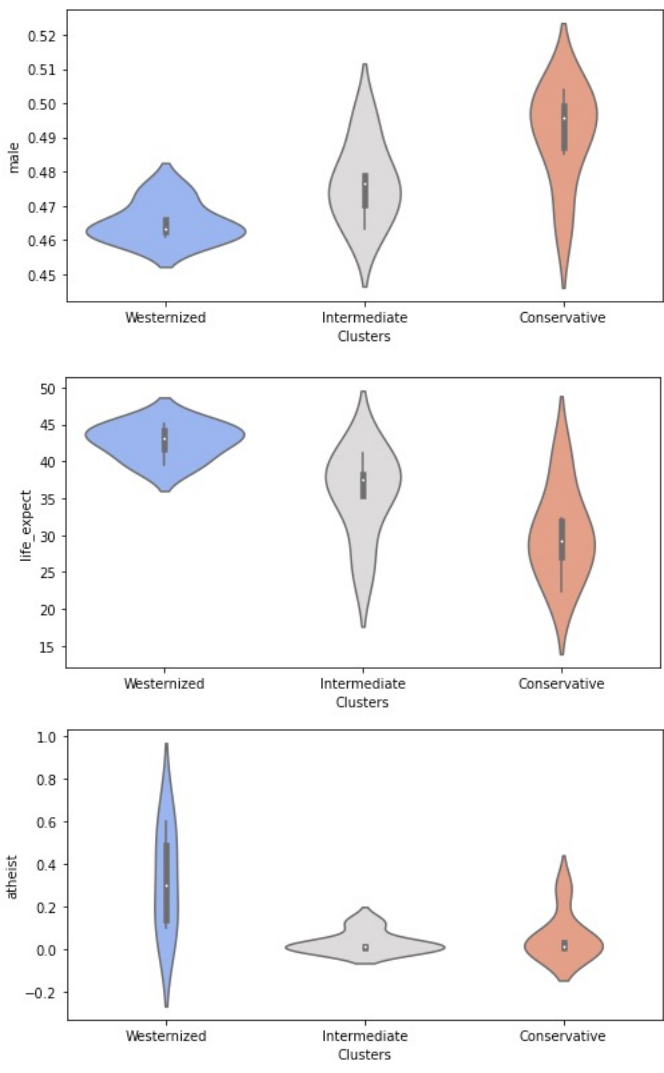
ECONOMIC INDICATORS



POLITICAL INDICATORS



DEMOGRAPHIC INDICATORS



How can the research be improved?

- A more sophisticated method of dimensionality reduction can be used;
- Instead of the centroid-based clustering algorithm, which can handle only clusters with spherical or ellipsoidal symmetry, the density-based algorithm can be implemented;
- Add more indicators to analysis;
- Use not only descriptive attributions, but also interaction characteristics of countries;
- Check consistency of clusters over time.



Does the Post-Soviet Space Still Exist?

An Unsupervised Machine Learning Approach to Classifying Countries

Minsk 2021