



Regional Inequalities in Education in China: Could Virtual Learning Changes the Situation?

Yuming Liu **Adviser: Richard Evans**
MACSS Program, University of Chicago, June 2020



Abstract

Because of the unbalanced developments between coastal and inland provinces and between urban and rural areas, regional imbalance has been a major issue in China for decades since the reform and opening-up. Spatial inequalities are in areas like education and health care. Our paper would focus on the regional inequalities in education. We built several models using data collected from the National Yearbooks and applying Machine Learning algorithms to explain the causes of education inequalities. From the results, we determined whether the popularization of virtual learning could help reduce inequalities and reshape the educational resource allocation in the next decade.

Education Gini Index

The **Education Gini Index** was introduced by Thomas and etc. (2000), and we generate the Education Gini Index based on their function:

$$E_L = \frac{1}{\mu} \sum_{i=2}^n \sum_{j=1}^{i-1} p_i |y_i - y_j| p_j$$

where

E_L is the Education Gini based on education attainment distribution;

μ is average years of schooling for the sample population;

p_i and p_j stand for the proportions of population with certain levels of schooling;

y_i and y_j are years of schooling of specific levels;

n is the number of levels.

Average Year of Schooling

We calculate the **Average Year of Schooling (AYS)** from the function modified from the function to calculate Education Gini Index:

$$AYS = \sum_{i=1}^n y_i p_i$$

where

AYS is the Average Year of Schooling based on education attainment distribution;

p_i stands for the proportion of population with a certain level of schooling;

y_i is years of schooling of a specific level;

n is the number of levels. (Fleisher and etc. 2010)

Virtual Learning Trends

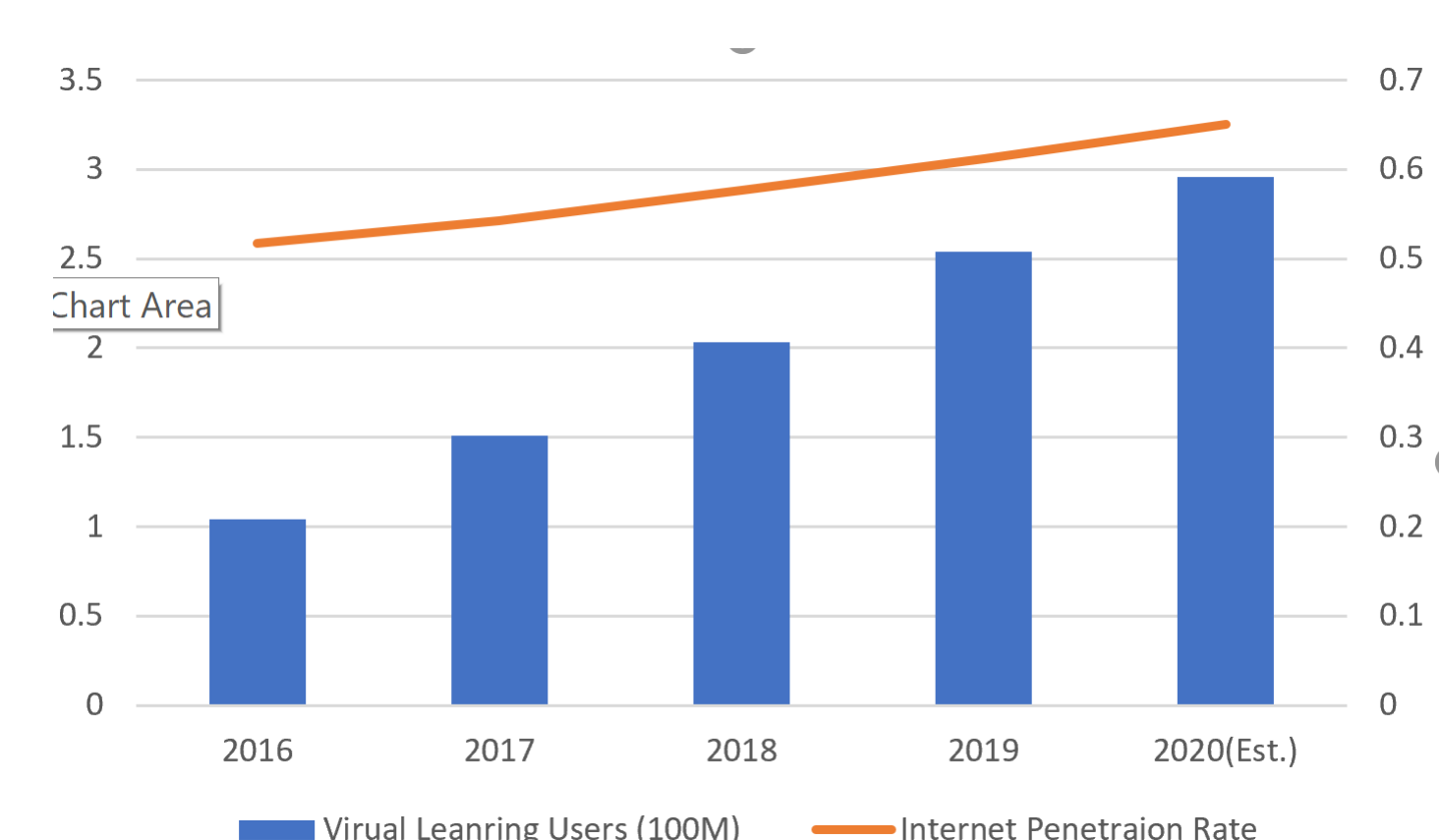


Figure 0.1. The development of Internet Penetration Rate and Virtual Learning Users in China (2016-2020)

Spatial Visualization of Education Gini

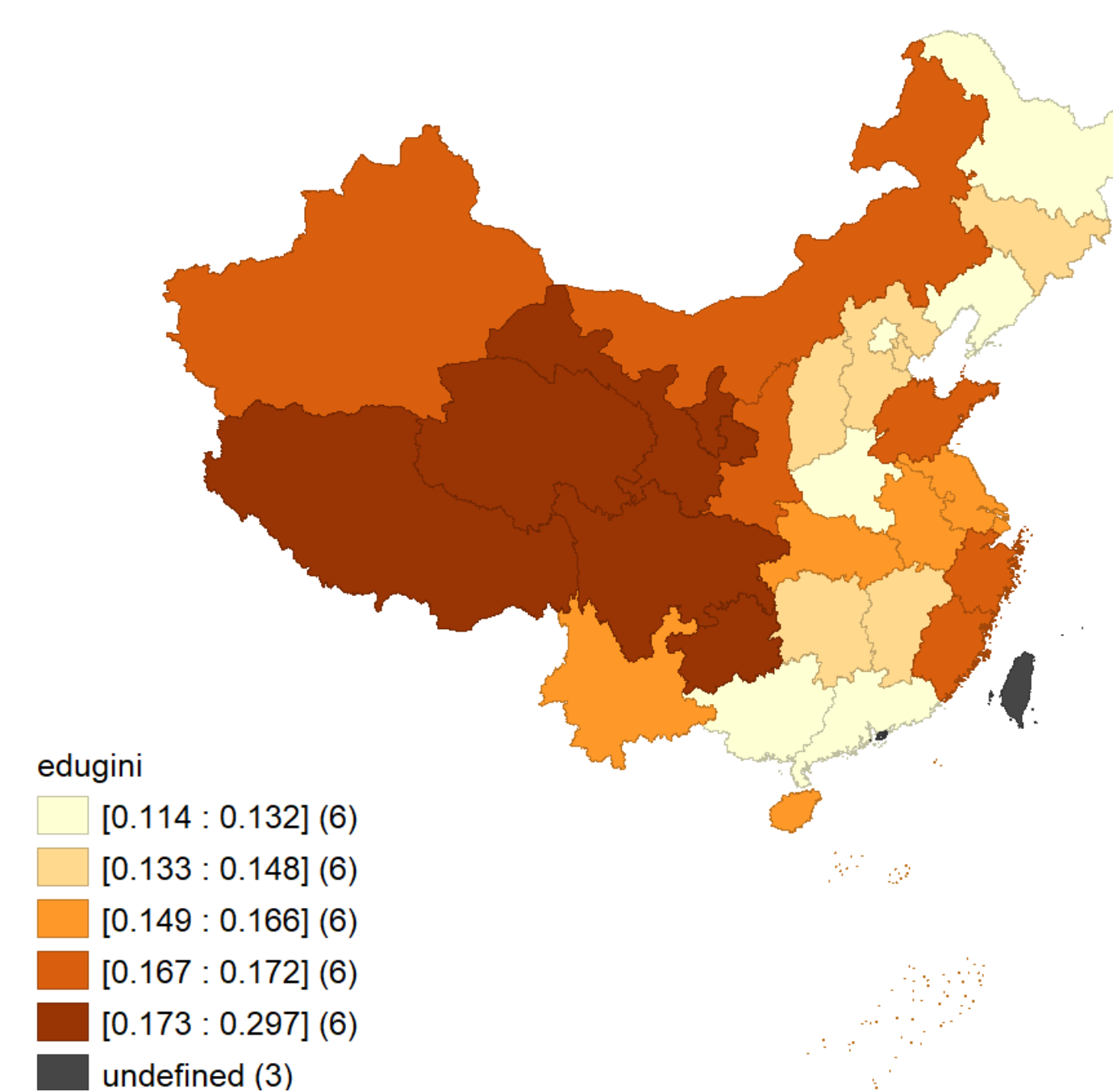


Figure 0.2. Education Gini Index by Region in China, 2018

Spatial Visualization of AYS

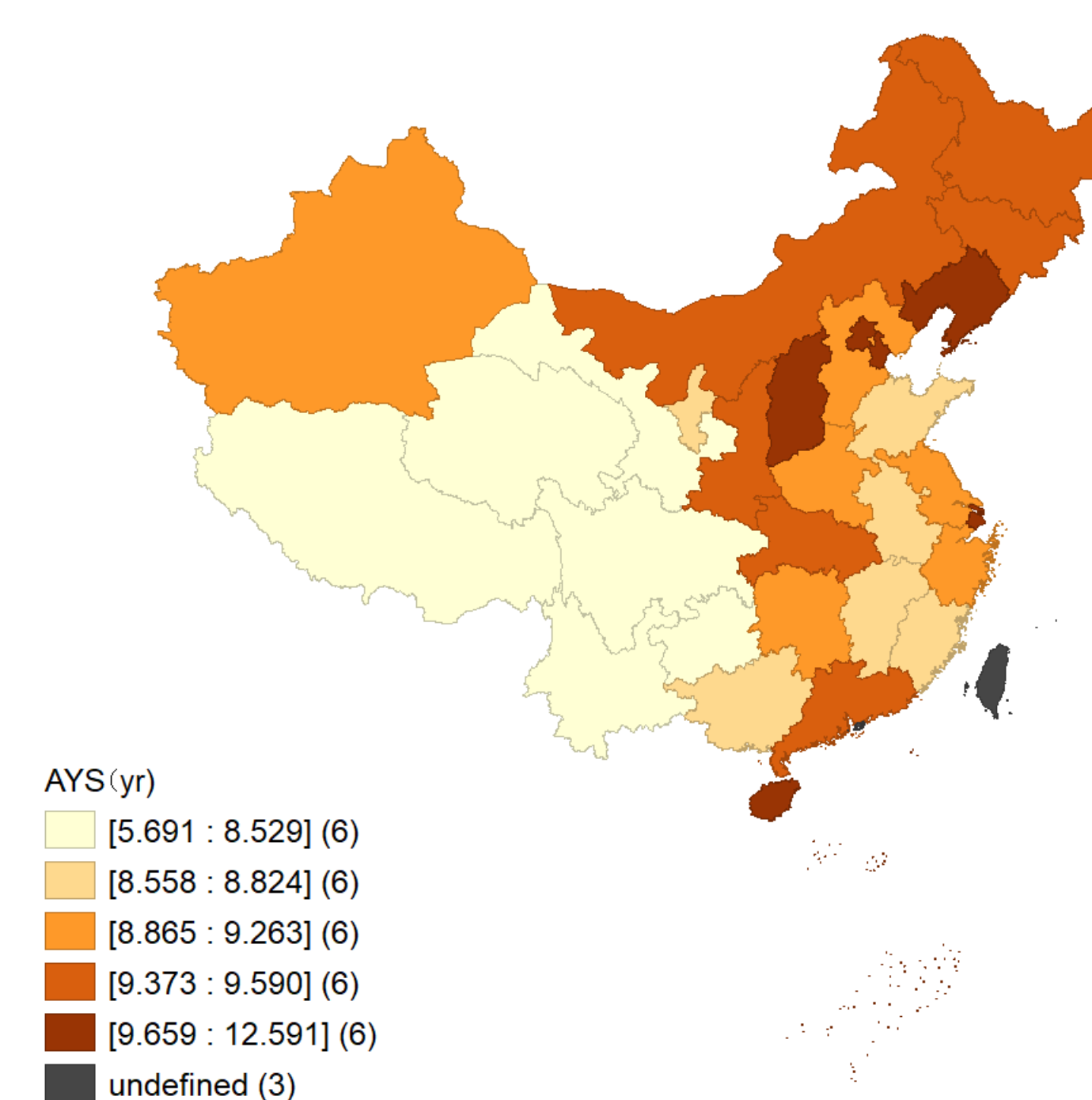


Figure 0.3. Average Year of Schooling (AYS) by Region in China, 2018

OLS Model Results

Table 6: Combined OLS Model Results

	Value	SE	t	p
virt_pct (constant)	80.5174	2.227	36.153	0.000
rural_pct	-0.0854	0.008	-10.234	0.000
Multiple R-squared	0.783			
Model p-value	0.00			
Intercept	0.4446	0.066	6.729	0.000
AYS(Model)	-0.0331	0.010	-3.424	0.002
DPI	4.655e-06	1.13e-06	4.132	0.000
Education&Entertainment Consumption	-3.769e-05	1.48e-05	-2.545	0.017
Number of Primary Schools	-3.487e-06	1.06e-06	-3.303	0.003
Media Attention of Virtual Learning	-0.2975	0.153	-1.939	0.064
Multiple R-squared	0.698			
Model p-value	0.00			

Figure 0.4. Combined OLS Model Results

2020 Education Gini Index Predictions (Random Forest)

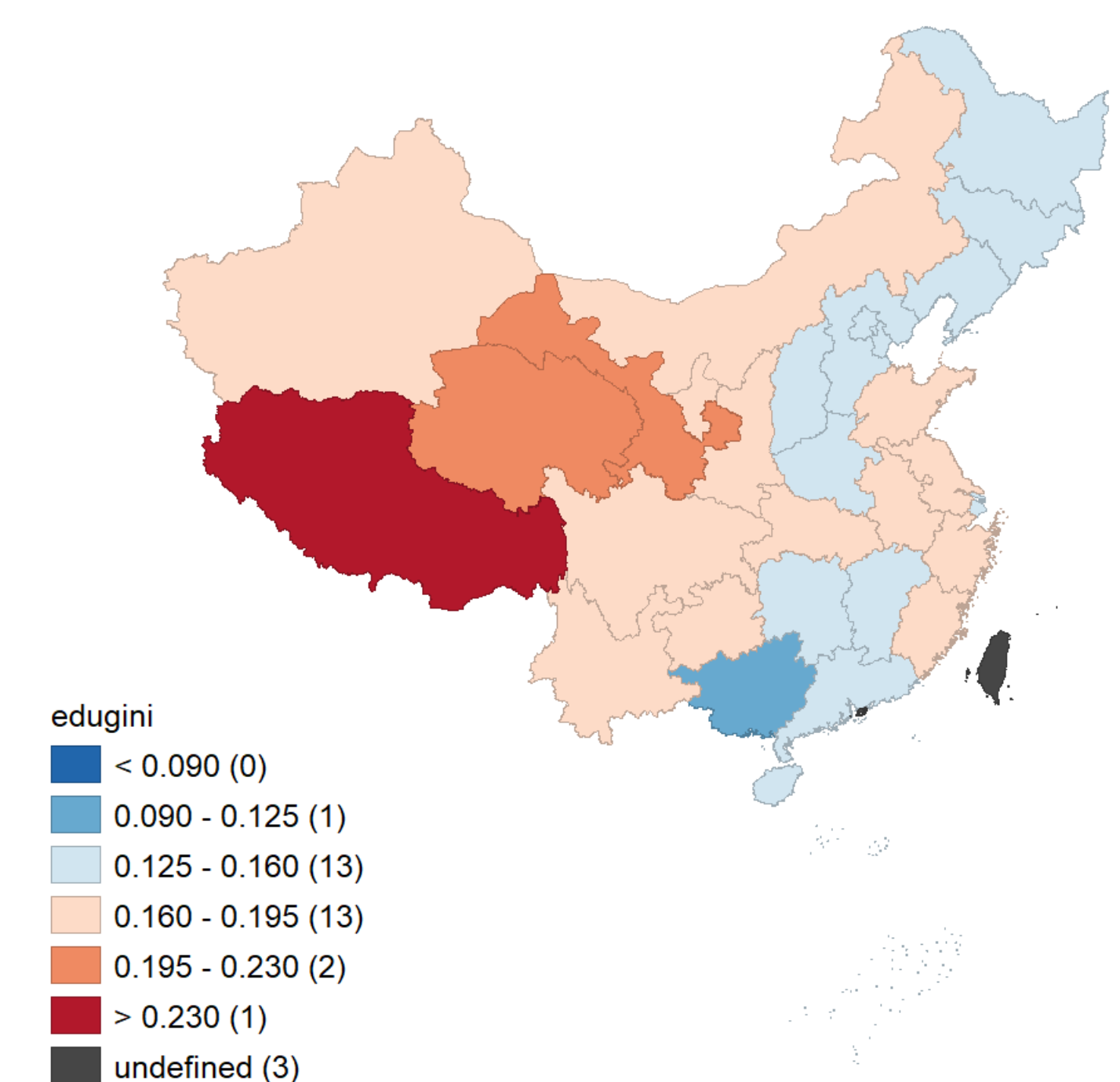


Figure 0.5. Standard Deviation of Education Gini Index by Region in China, 2018

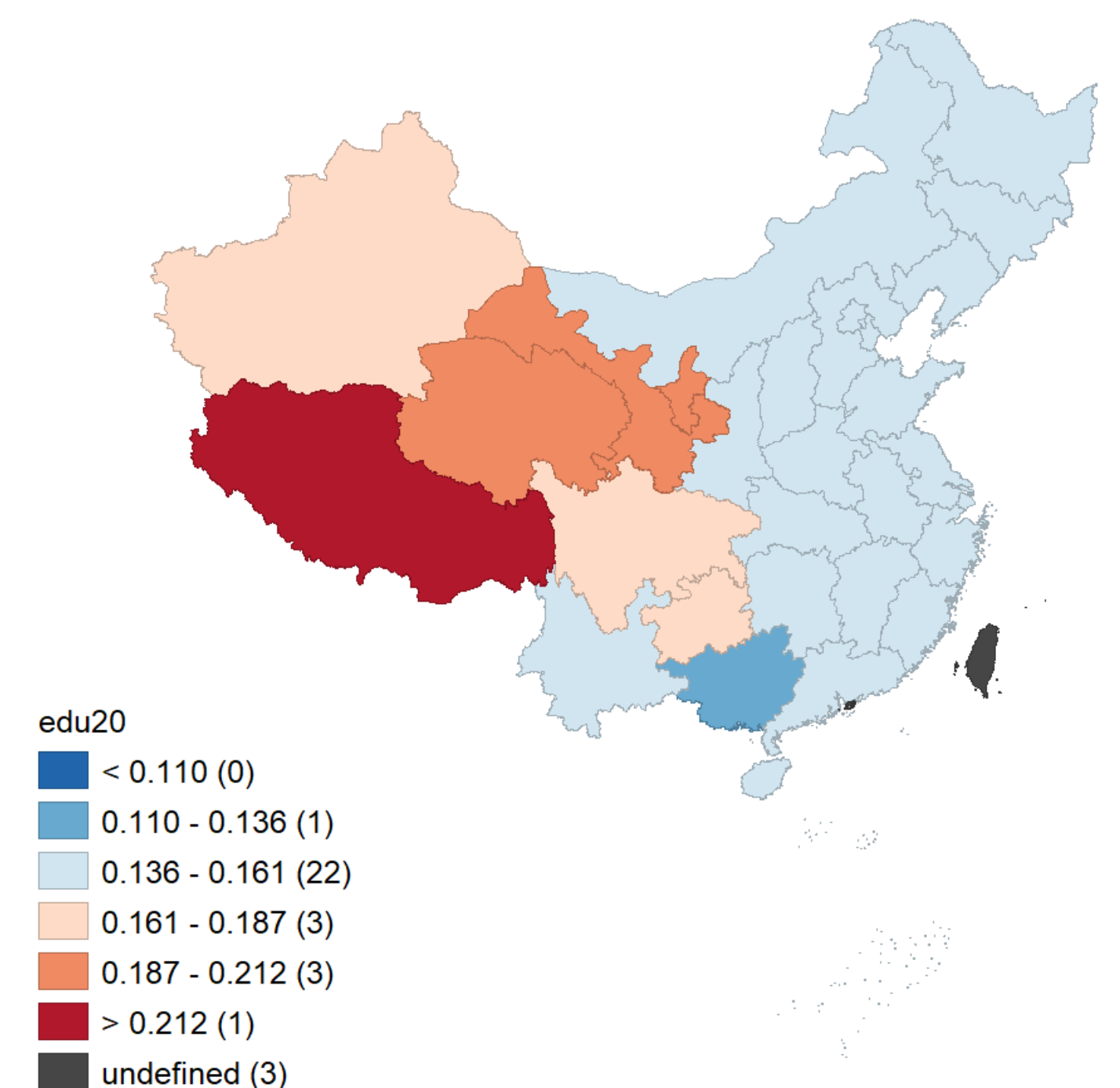


Figure 0.6. Predicted Standard Deviation of Education Gini Index by Region in China, 2020

Summary

The prediction results from the Machine Learning Algorithms with the highest model scores show that the increase of virtual learning resources and virtual learning users would help to eliminate the regional inequalities in education. Although the western provinces would remain their Education Gini Indexes higher than the others, the gaps between them could be narrowed by the popularization of virtual learning.

Acknowledgements

Thank my family, my friends, and Dr. Richard Evans for offering me remarkable supports during this period!