POVERTY MAPPING IN LATVIA

RESULTS FOR CONSULTATION AND VALIDATION

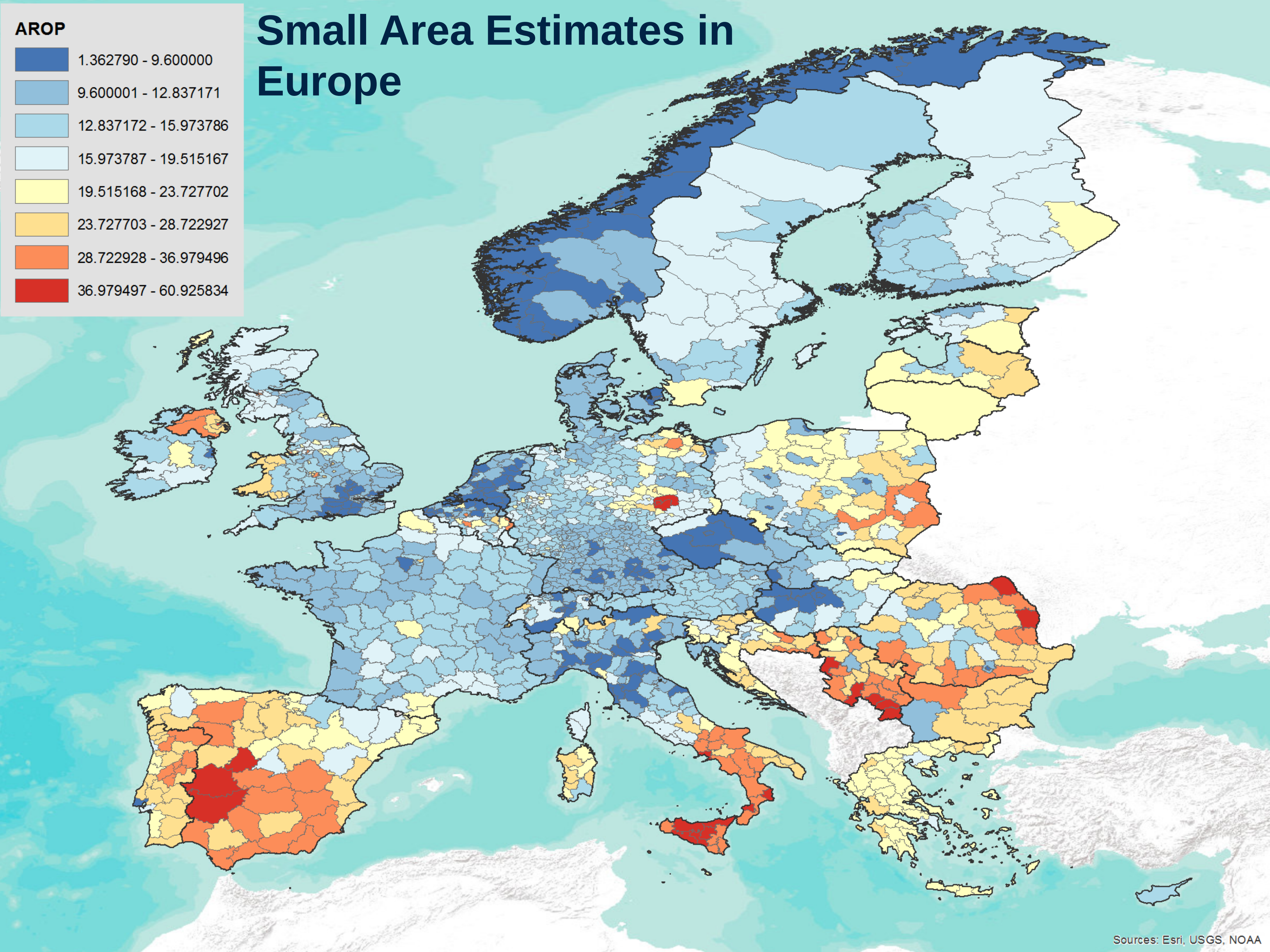
William Seitz
Economist
The World Bank

WORLD BANK GROUP

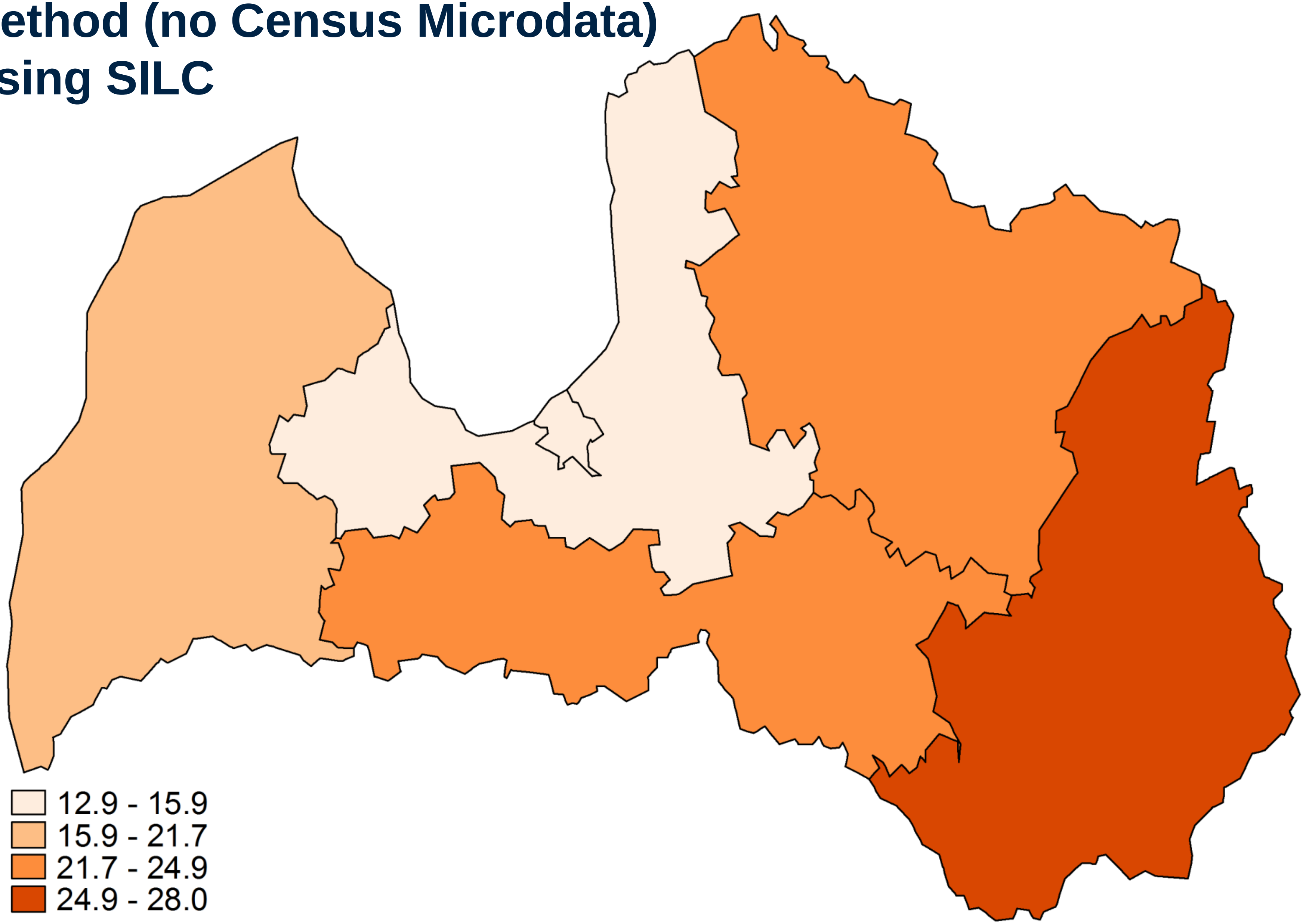
Outline

- What are Poverty Maps? Why Make Them?
- Analytical Approach
- Validation
- AROP Results
- Results for Subpopulations
- Absolute Per Capita Results
- Prices and Cost of Living

Small Area Estimates in Europe



Rates Available with Area-Based Method (no Census Microdata) Using SILC



Objective: Getting more local estimates

- Poverty and household welfare information often not available for lower administrative units
- Sample sizes are too small, At-Risk-of-Poverty (AROP) rates only representative at national and region level

- **But sometimes, the precise location of the poor is important**

WORLD BANK GROUP

The Basic Approach to Small Area Estimation (SAE)

Survey:

- Welfare measure
- Not representative at lower level

+

Census:

- Full coverage
- No monetary welfare measure

→

Poverty Mapping:

- Simulating a measure of welfare from household survey in the census, using statistical methods
- Geographically disaggregated estimates of: poverty, number of poor, average income/consumption, inequality
- Considerable data requirements

SAE Maps are Useful in Many Applications

- Build awareness and help identify leading and lagging areas of a country
- Strengthen accountability
- Decentralization of governance in many countries requires information on smaller administrative units
- Achieve better geographic targeting of resources
- Impact maps: updating poverty maps over time for monitoring
- Inform design of policy and program more broadly

There are Significant Data Requirements

- Survey and census have variables in common
- Common variables are sufficiently correlated with the welfare measure
- Survey and census can be linked at the target area level
- Census includes variables that capture location specific effects
- Census has large coverage

- **Latvia data meet all of these requirements**

WORLD BANK GROUP

Analytical Approach

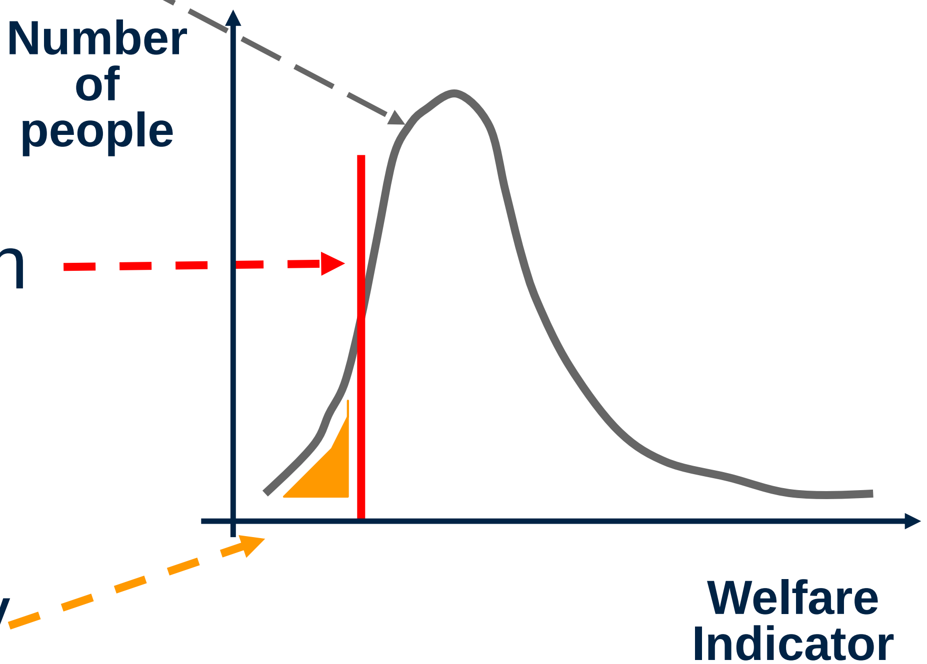
WORLD BANK GROUP

Measuring the AROP Rate

1. The income measure: used to derive a distribution of living standards

2. The **poverty line**: a threshold below which individuals are classified as poor

3. The **poverty measure**: a summary statistics of poverty in population



The ELL Approach

- Methodological papers
  - Elbers, Lanjouw and Lanjouw (2003, Econometrica)
  - Hentschel et al. (2000) and ELL (2000, 2002, 2003, 2008)
  - Elbers and van der Weide (2014, World Bank)

The ELL Approach in a Nutshell (1)

- We want to estimate W, the welfare measure of interest (AROP), for a lau-2 area for which it is not observed.
- $W(y, m)$ is a function of AE incomes y and adult equivalent size of the household m.
- We can estimate W by evaluating the expected value of y_{ch} given household characteristics and income model
- We model (the log of) adult equivalent income as:

$$y_{ch} = X'_{ch}\beta + u_{ch},$$

where $u_{ch} = \mu_c + \epsilon_{ch}$ and $E[u_c^2] = \sigma_\mu^2 + \sigma_\epsilon^2$

- household *h* residing in area *c*
- There is an area component (μ_c) and a household component (ϵ_{ch}) of the error term
- A consistent estimator of μ from the income regression.

WORLD BANK GROUP

The ELL Approach in a Nutshell (2)

- ELL permit the variance of idiosyncratic errors $\varepsilon_{c,h}$ to vary between households (to account for heteroscedasticity), by considering the variance as a function of household and area characteristics
- We do not observe the variance, so we cannot use it as a dependent variable. But we do observe the squared residual, which satisfies: $\sigma^2_{\varepsilon,ch} = E\left[\varepsilon^2_{ch}|z_{ch}\right]$, where z_{ch} denotes the vector with household and area characteristics
- ELL propose work with a logistic transformation of ε^2_{ch} which yields a more symmetric distribution because the “raw” squared error tends to have a heavily skewed distribution:

$$ln\left(\frac{\varepsilon^2_{ch}}{A-\varepsilon^2_{ch}}\right) = z^T_{ch}\alpha + \epsilon_{ch}$$

WORLD BANK GROUP

The ELL Approach in a Nutshell (3)

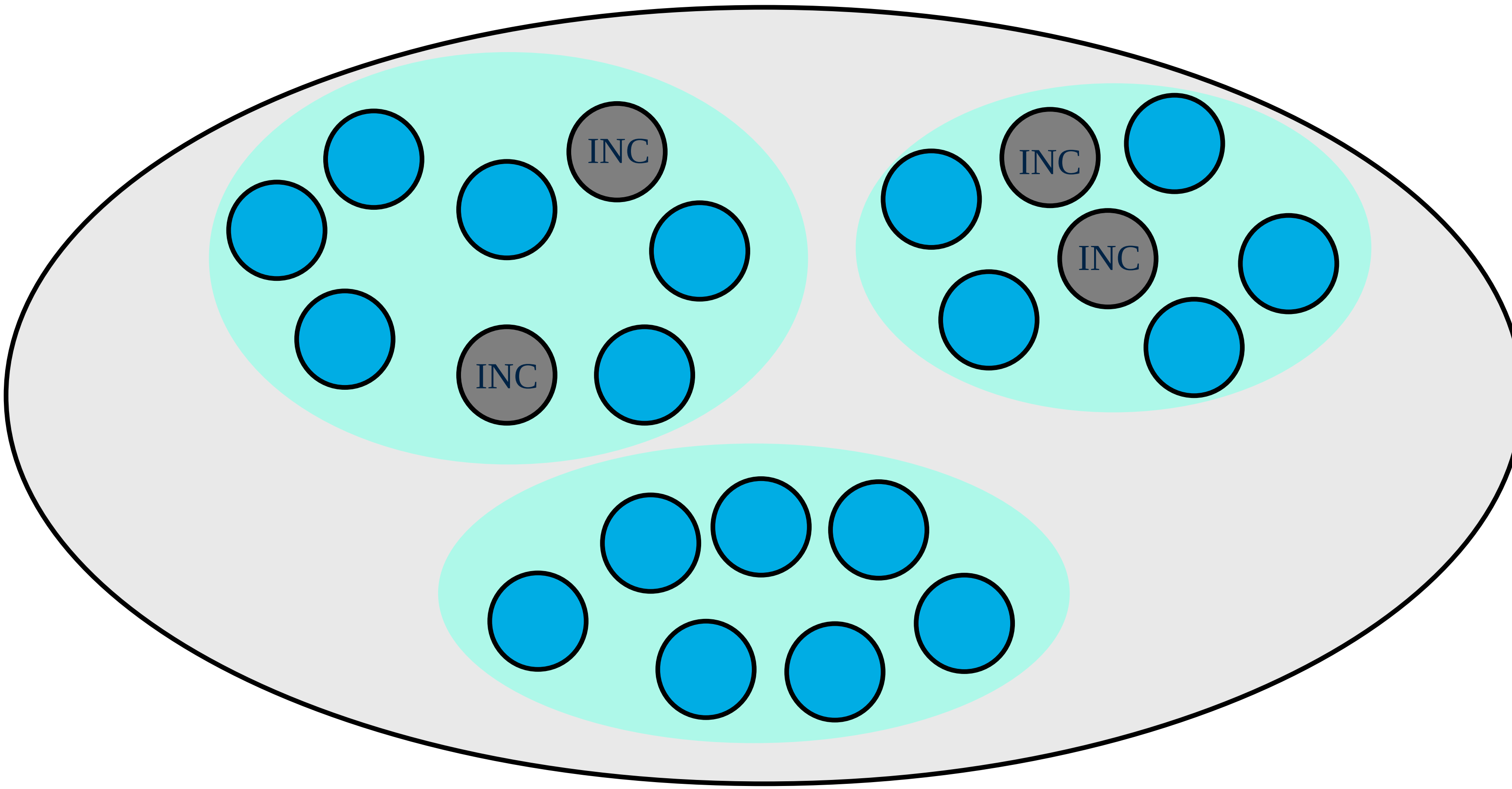
- Use simulation to obtain point estimates and standard errors
- For each replication r simulate:

$$\tilde{y}^r = X^{'}\widetilde{\beta}^r + \tilde{\mu}_c^r + \tilde{\varepsilon}_{ch}^r$$

- where model parameters and errors are drawn from their corresponding distributions
- Evaluate welfare $\widehat{W}^{(r)}$ for each replication
- Averaging over *R* simulated values given the point estimate of *W*
- Standard deviation over simulated values provides estimates of standard errors

WORLD BANK GROUP

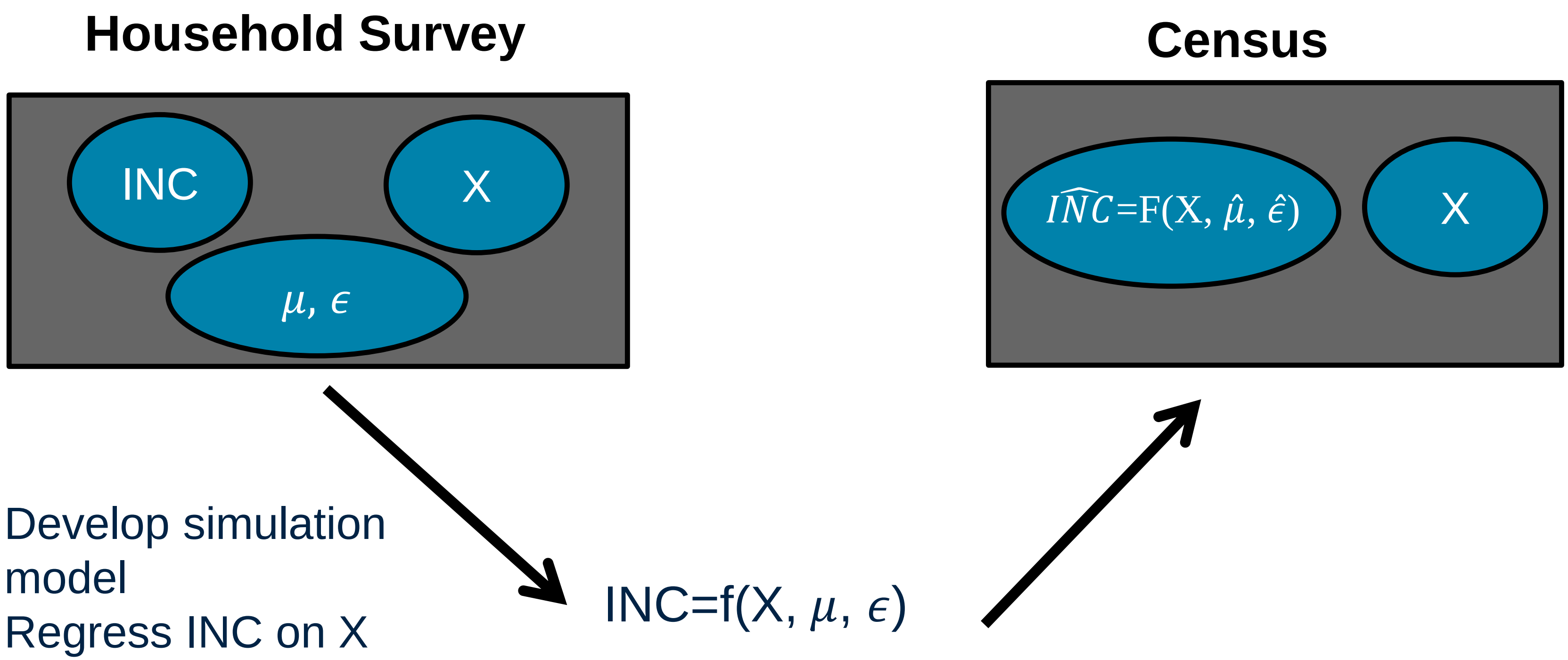
Visualizing The Problem



Interviewed in SILC

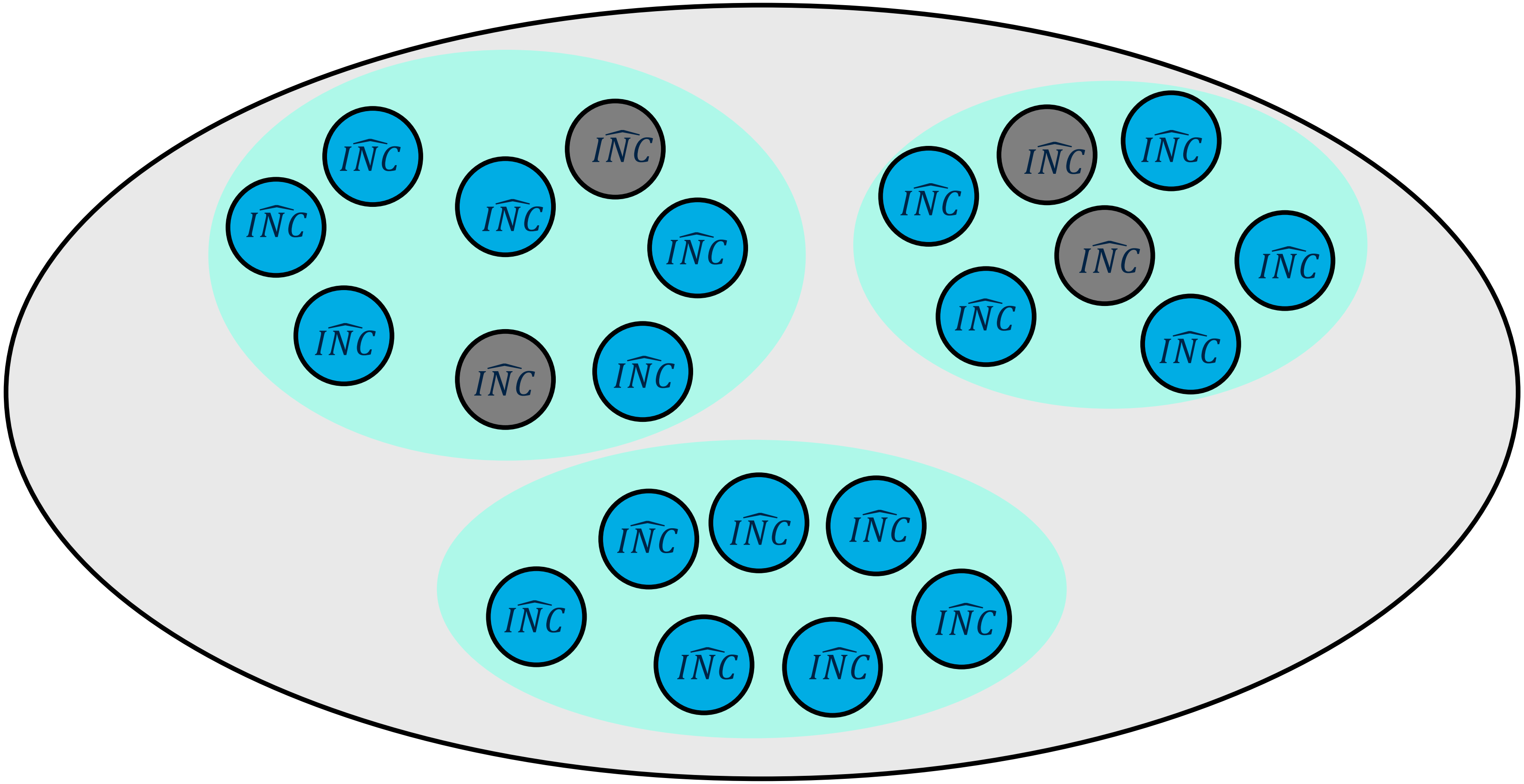
Not interviewed in SILC

The Basics of ELL (Elbers, Lanjouw, and Lanjouw 2003)



$\widehat{INC}$: **Simulated** income (replicated many times)
X: Poverty correlates like employment, education
μ, ϵ: Terms relating to the area and model error (replicated many times)

Visualizing The Problem



Interviewed in SILC

Not interviewed in SILC

To Recap

- The model generates simulated income for each household in the census, for Latvia, simulated 100 times
- Estimate share that are poor, or other indicators of interest
- The method also estimates how precise the results are

The ELL Approach has Some Limitations

Idiosyncratic error $(W - \mu)$: increases for smaller populations → limit to the degree of geographic disaggregation

Model error $(\mu - \hat{\mu})$: depends on the precision of model estimates, and the distance between household characteristics X in the Census and the survey sample

Thus, key ingredients for the method to work:

- Large enough sample size
- Good match between Survey and Census characteristics
- Model that can explain well variation in income
- Survey and Census should not be far apart in time

WORLD BANK GROUP

Variables in the SILC usually match the Census – Individual Level

	Survey			Census		
	Mean of Sum	Mean of Mean	Mean of Max	Mean of Sum	Mean of Mean	Mean of Max
Male	1.19	0.44	0.83	1.08	0.44	0.73
Female	1.36	0.56	0.95	1.30	0.56	0.85
Age 0 to 6	0.27	0.06	0.22	0.15	0.04	0.13
Age 1 to 14	0.59	0.14	0.38	0.34	0.09	0.24
Age 15 to 24	0.52	0.13	0.37	0.32	0.11	0.25
Age 25 to 64	1.77	0.56	0.87	1.29	0.55	0.78
Age 65 and Above	0.45	0.21	0.35	0.44	0.25	0.35
Born In Latvia	2.25	0.87	0.94	2.03	0.82	0.89
Born in Another Country	0.30	0.13	0.23	0.35	0.18	0.27
Citizen	2.17	0.84	0.91	2.04	0.84	0.89
Not Latvian Citizen	0.38	0.16	0.25	0.34	0.16	0.24
Not Married	1.14	0.38	0.66	0.88	0.42	0.61
Married	1.18	0.43	0.56	0.90	0.39	0.49
Employed	1.30	0.41	0.75	0.96	0.40	0.63
Unemployed	0.30	0.09	0.24	0.22	0.09	0.19

WORLD BANK GROUP

Variables in the SILC Usually Match the Census – Household Level

	Survey	Census
Household Size	2.41	2.38
Household Size^2	7.77	7.74
Log of Household Size	0.72	0.70
Individual House	0.26	0.24
Semi-Detached House	0.04	0.02
Other Building	0.68	0.72
Central Heating	0.56	0.70
Other Heating	0.21	0.30
No Heat	0.23	0.00
Water Connection	0.84	0.89
No Water Connection	0.16	0.11
Bath Exists	0.81	0.78
Without Bath	0.19	0.22
Indoor Toilet Exists	0.82	0.98
Floor Space	62.84	66.06

The Main (Beta) Model for the AROP Simulation

ln_y	Coef.	Std. Err.	p-value
Latgale	-.026581	.0302564	0.380
Pier_ga	-.0302659	.0275517	0.272
Vidzeme	-.060611	.033539	0.071
age_65pl_sum	.1248789	.0149927	0.000
bath_exists	.1897565	.0266665	0.000
employed_mean	.6323528	.0313044	0.000
female_sum	.0505583	.0174216	0.004
floor_space	.0021672	.0003055	0.000
indiv_house	-.0604117	.0279342	0.031
lau_pct_cens	.0198969	.0045149	0.000
lau_toilet_flush	.2290859	.1040184	0.028
lva_ctzn_mean	.1222214	.0257429	0.000
male_mean	.3032487	.0351027	0.000
post_secondary_max	.2267947	.0376416	0.000
post_secondary_mean	.2487576	.054885	0.000
unemployed_mean	-.6504381	.0475351	0.000
Constant	6.23166	.185533	0.000

WORLD BANK GROUP

The Residual (Alpha) Model for the AROP Simulation

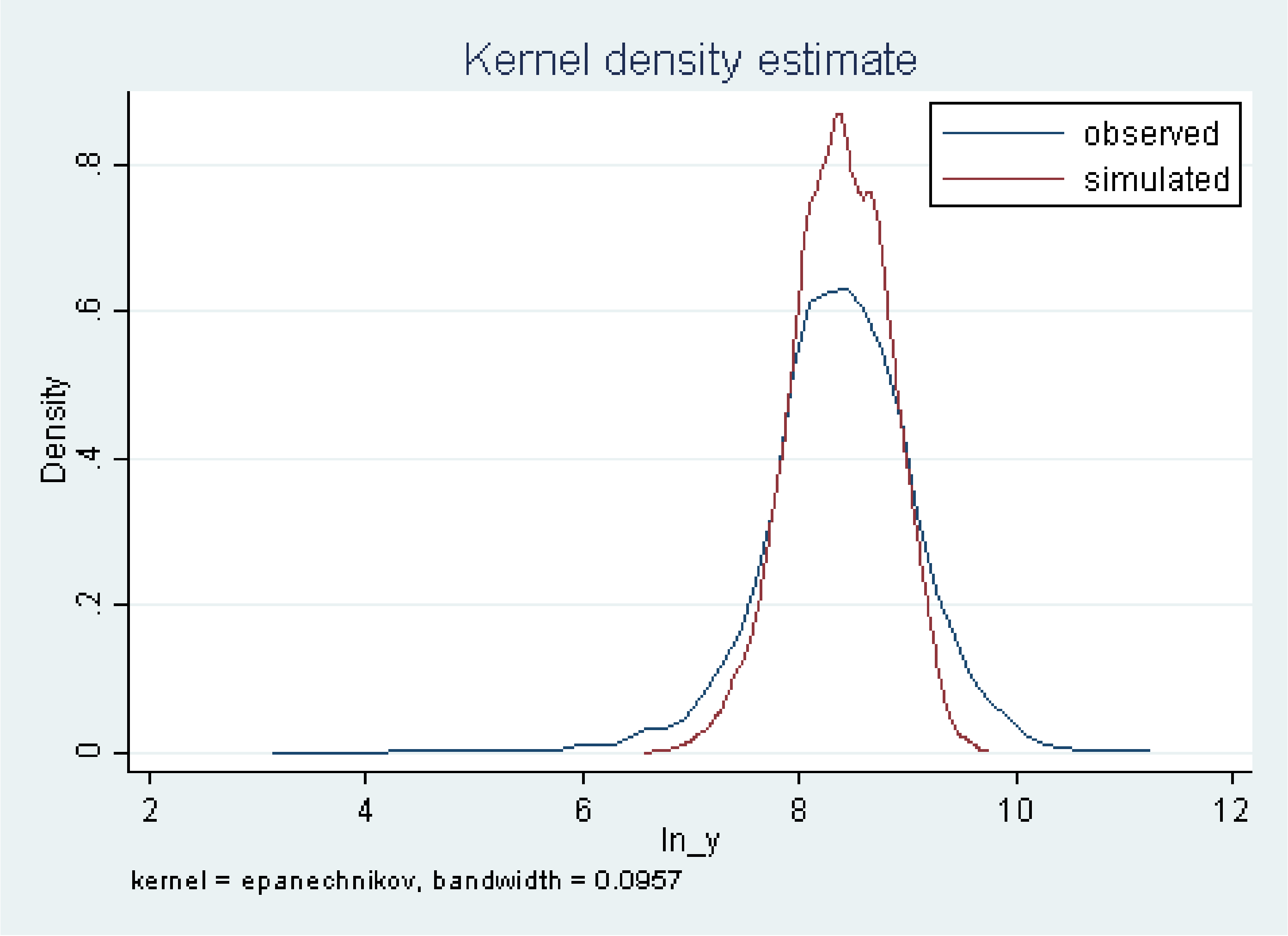
Residual	Coef.	Std. Err.	p-value
age_65pl_mean	-1.375691	.1211237	0.000
employed_mean	-.5761464	.1458661	0.000
female_sum	-.1306706	.0659954	0.048
floor_space	.0025451	.0010641	0.017
post_secondary_max	-.3967177	.1661145	0.017
post_secondary_mean	1.268396	.2430074	0.000
unemployed_mean	.5646785	.2148894	0.009
Constant	-5.87342	.1455299	0.000

Validation

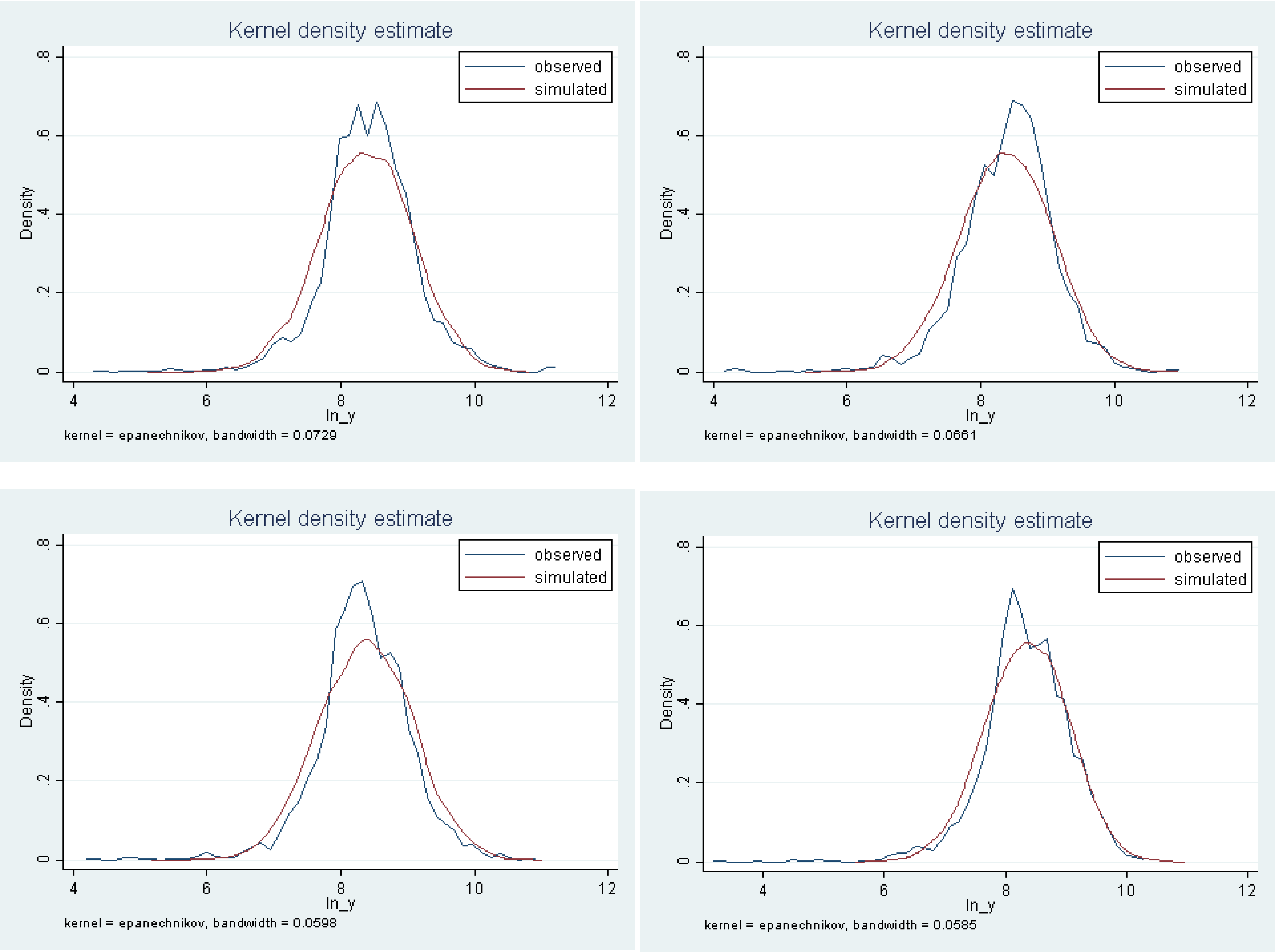
Outline

- Two Standard Methods
  - Simulating within the SILC
  - Aggregating the Census Results and Comparing to the Census
- Third option in Latvia
  - Compare Census estimates in the Census with results in the SILC

Simple Conditional Prediction Peaks at the Mean, Does Less Well at the Tails



WORLD BANK GROUP



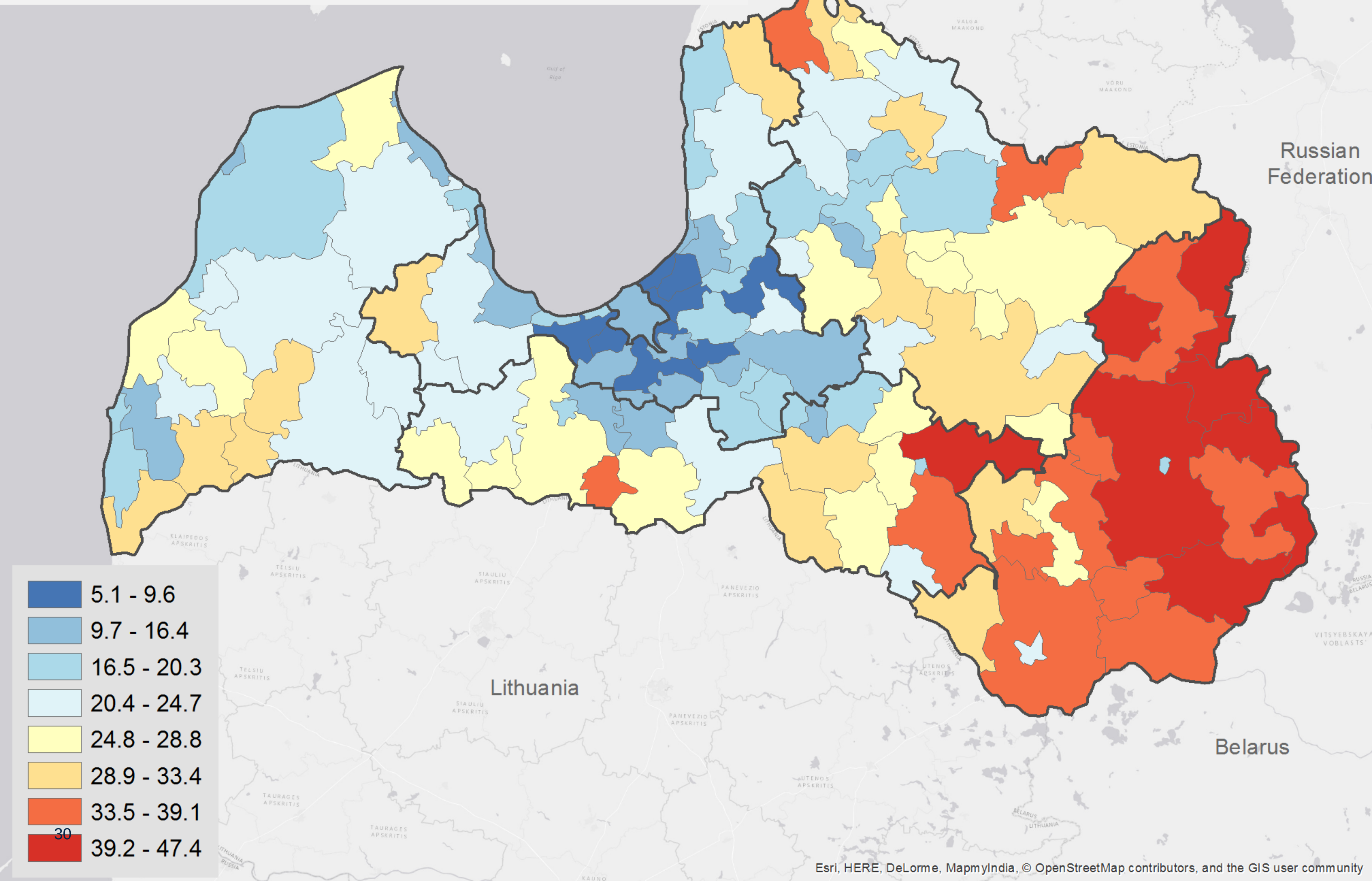
Kernel density estimate
observed
simulated
Density
ln_y
kernel = epanechnikov, bandwidth = 0.0729
Kernel density estimate
observed
simulated
Density
ln_y
kernel = epanechnikov, bandwidth = 0.0661
Kernel density estimate
observed
simulated
Density
ln_y
kernel = epanechnikov, bandwidth = 0.0598
Kernel density estimate
observed
simulated
Density
ln_y
kernel = epanechnikov, bandwidth = 0.0585

Pretty Close Match Between Direct vs. Simulated Estimates

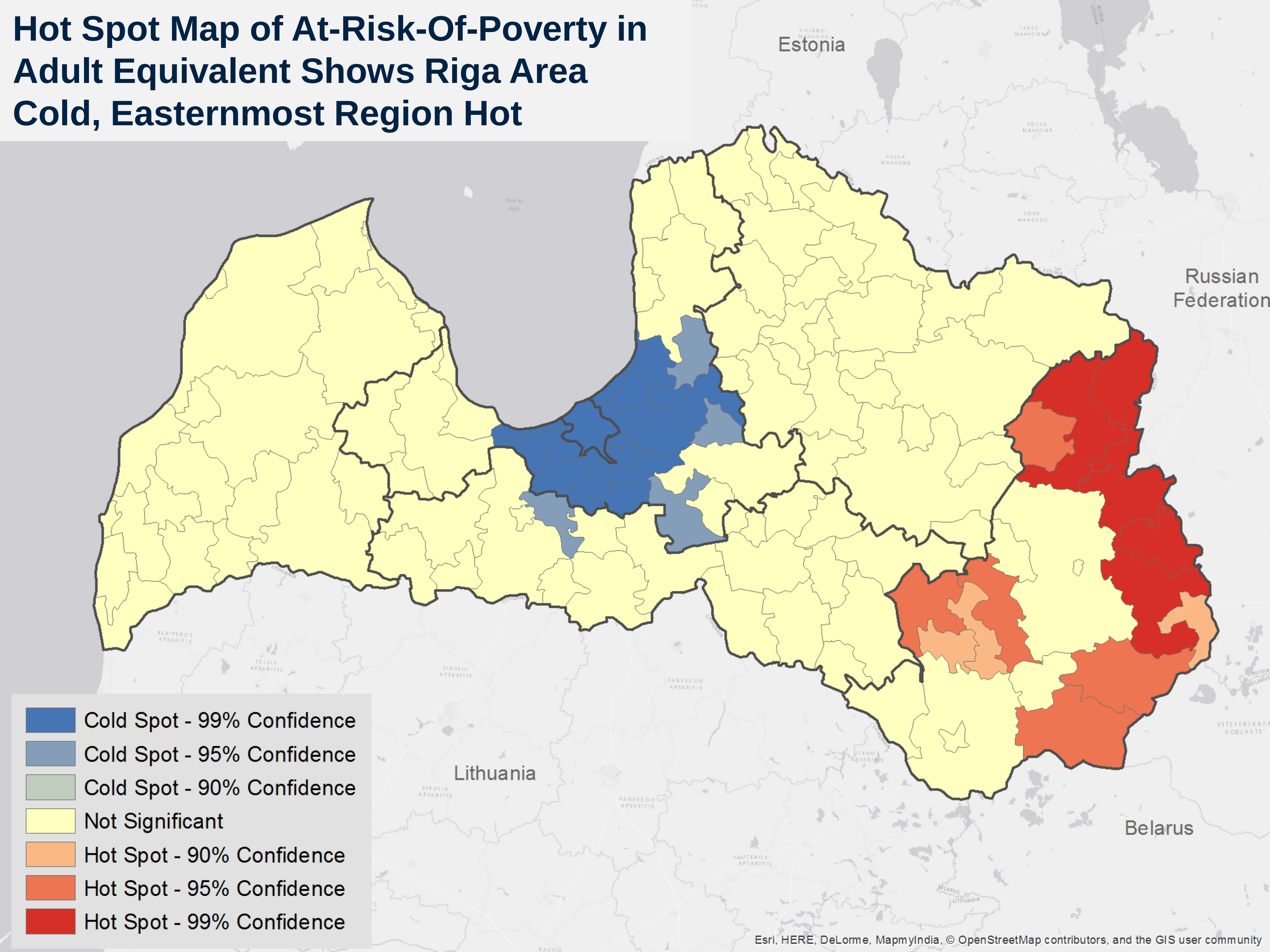
	AROP					
	Direct		Povmap V1		Povmap V2	
	AROP	Std. Err.	AROP	Std. Err.	AROP	Std. Err.
Riga	12.9	0.7	12.7	0.6	12.7	0.6
Pieriga	15.5	1.2	14.9	2.8	15.4	3.1
Kurzeme	19.8	1.4	20.4	3.1	20.5	3.2
Zemgale	22.2	1.6	23.0	3.8	23.1	4.0
Vidzeme	27.8	2	22.2	4.0	23.6	4.4
Latgale	28.6	1.7	30.0	3.0	30.4	4.1
National	19.2	0.5	18.9	2.6	19.2	2.7

AROP Results

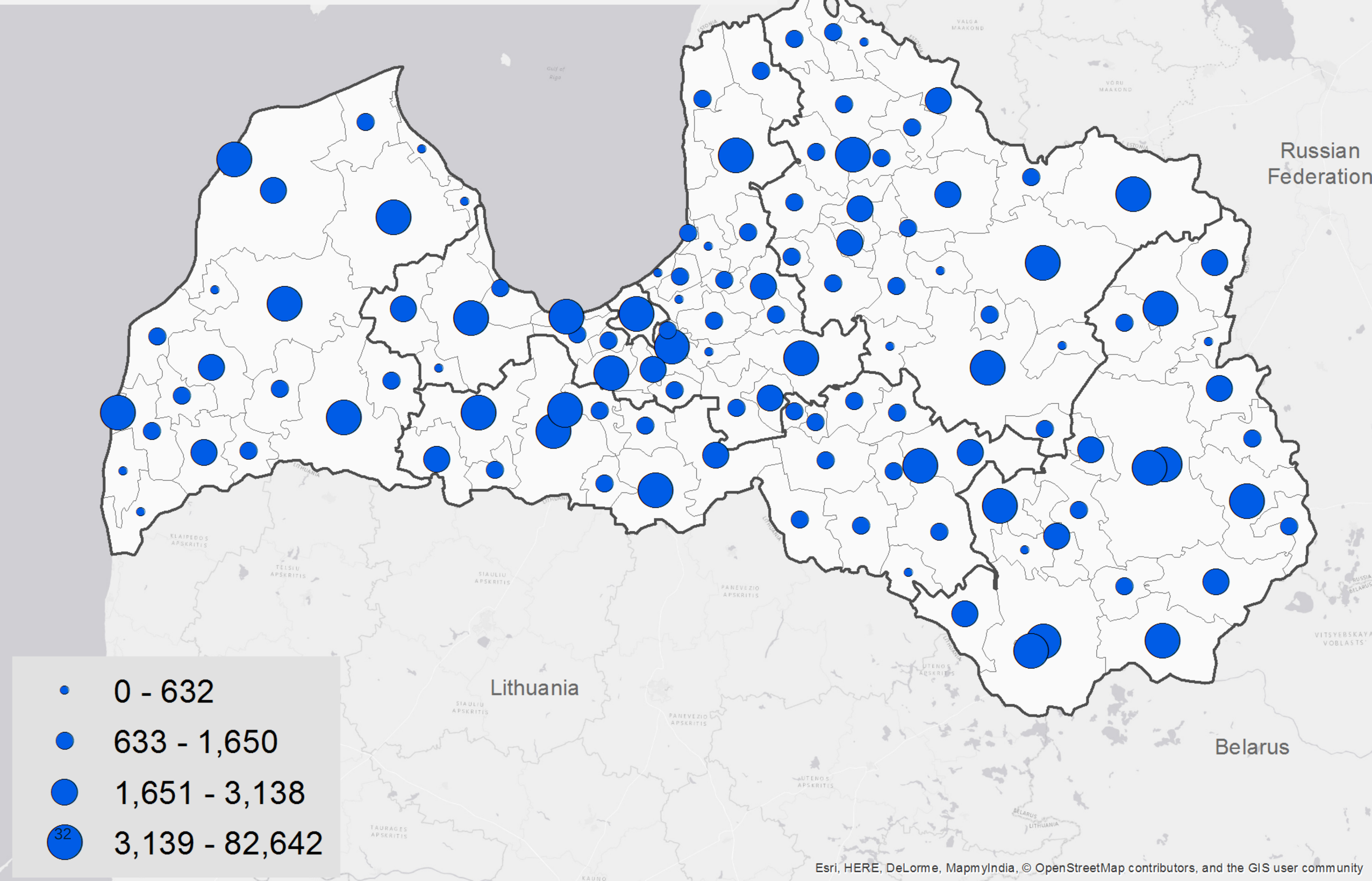
At-Risk-Of-Poverty in Adult Equivalent Shows Riga with Lowest Rates, Easternmost Region with Highest.



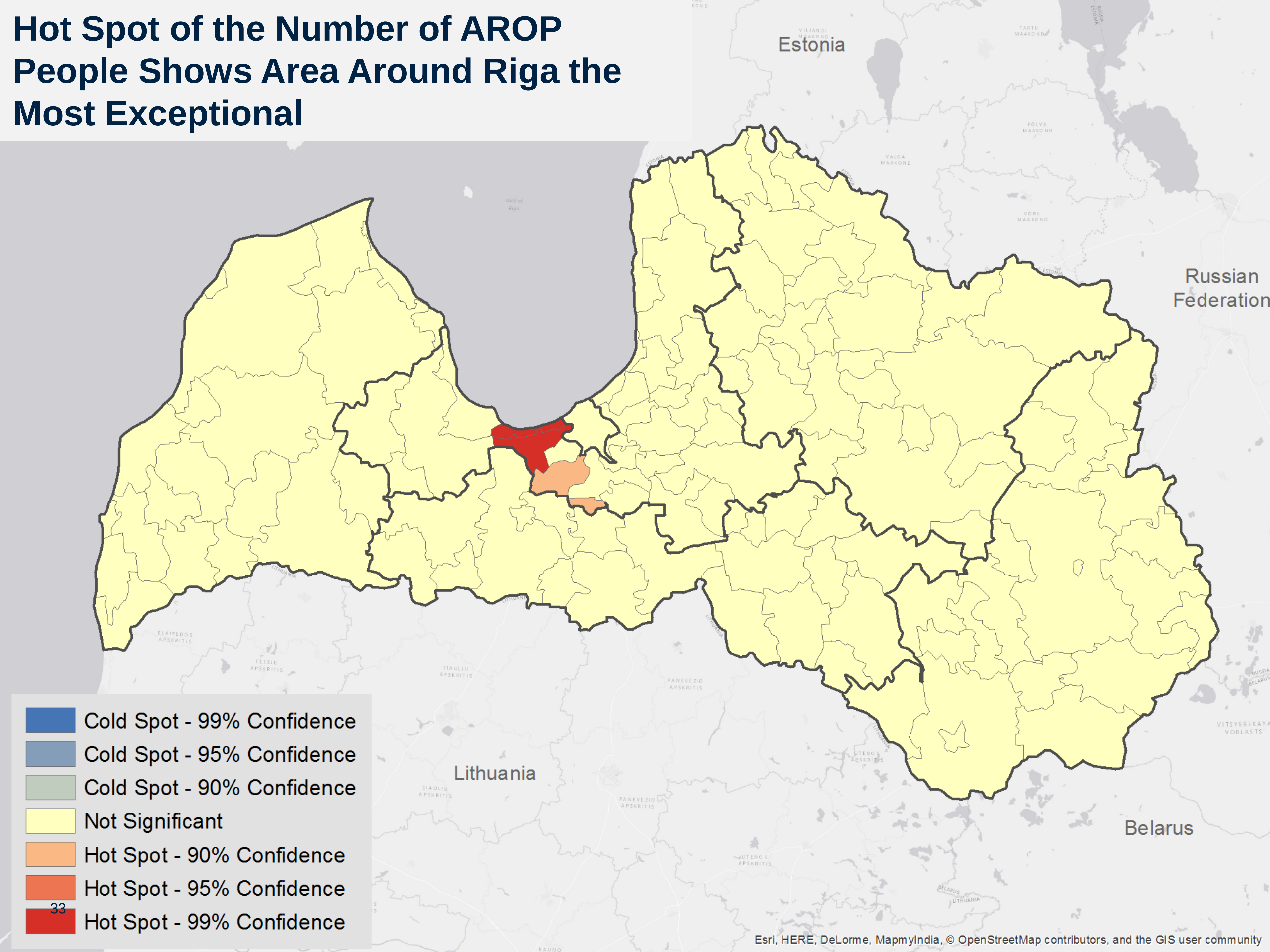
Hot Spot Map of At-Risk-Of-Poverty in Adult Equivalent Shows Riga Area Cold, Easternmost Region Hot



The Number of People At-Risk-Of-Poverty Shows Remarkable Geographic spread, Concentrated in Dense Areas.

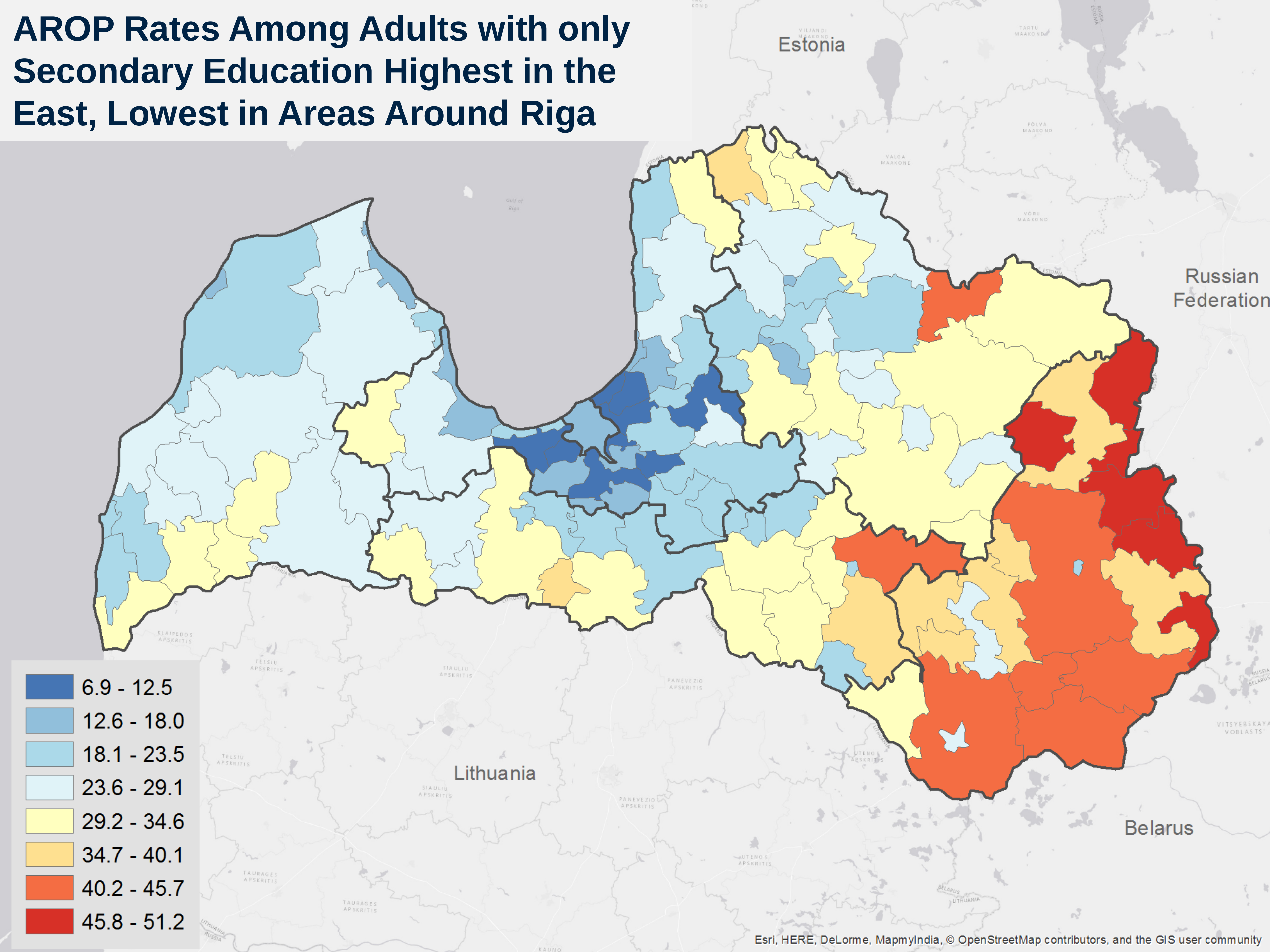


Hot Spot of the Number of AROP People Shows Area Around Riga the Most Exceptional

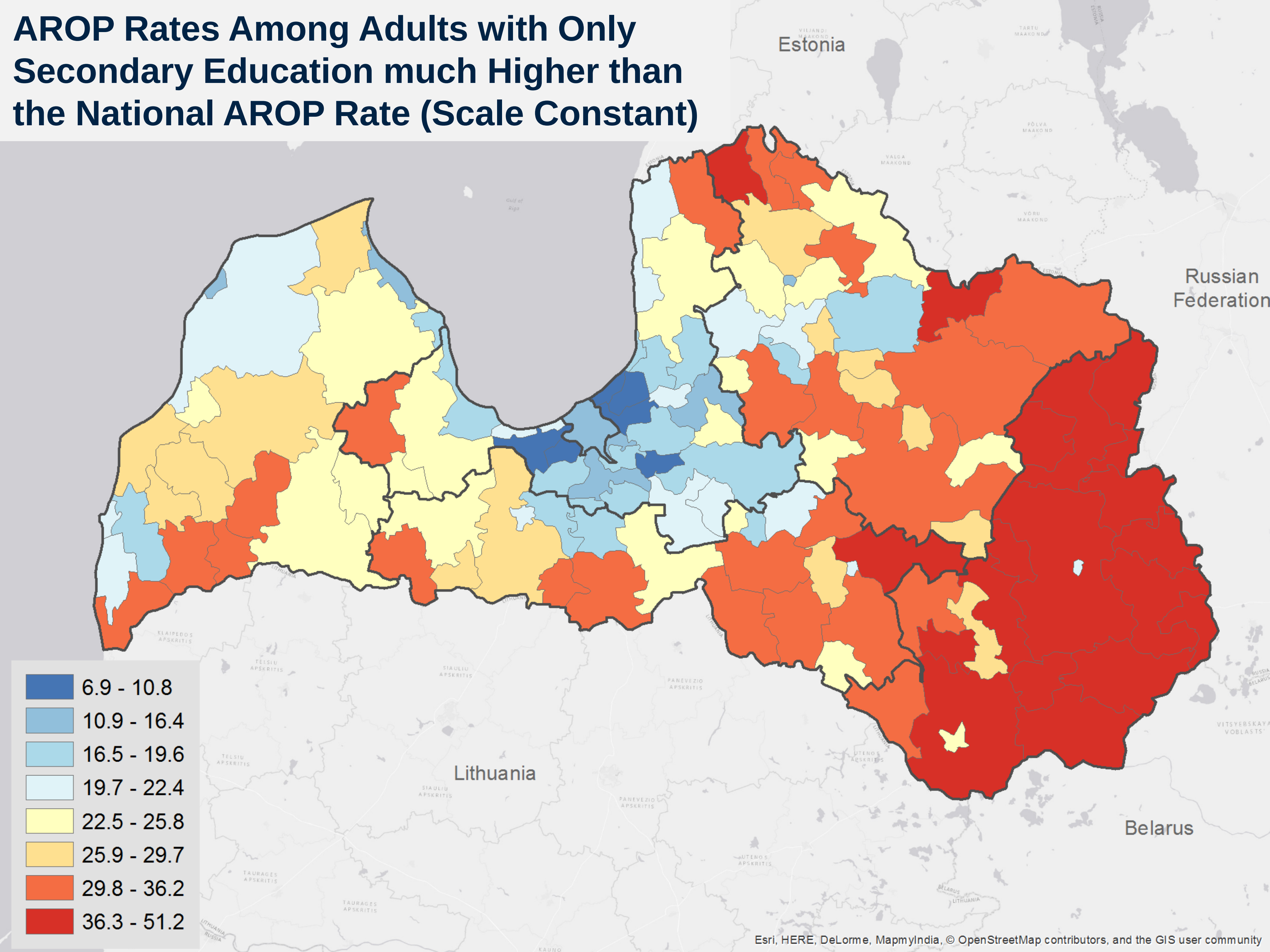


Subpopulations

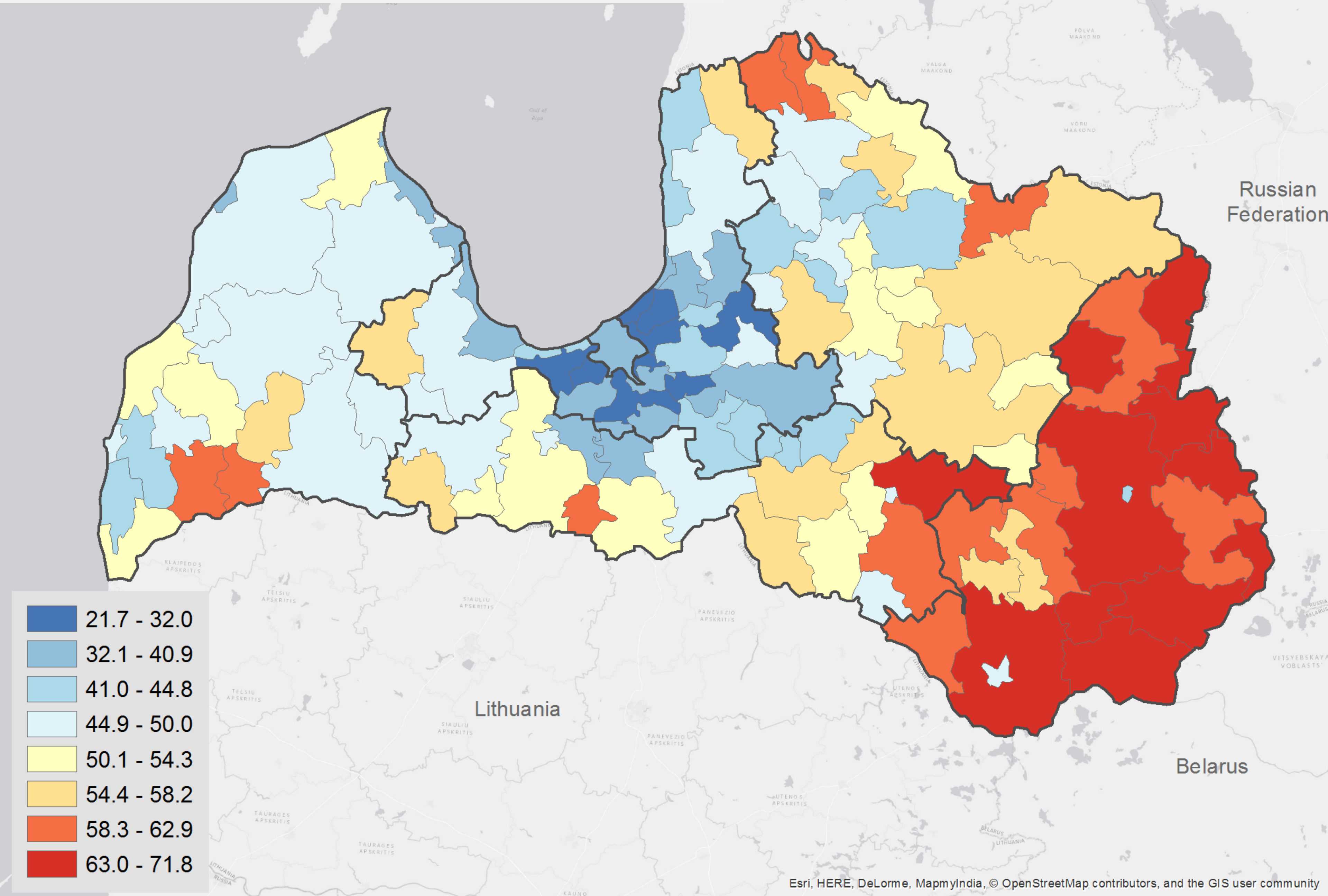
AROP Rates Among Adults with only Secondary Education Highest in the East, Lowest in Areas Around Riga



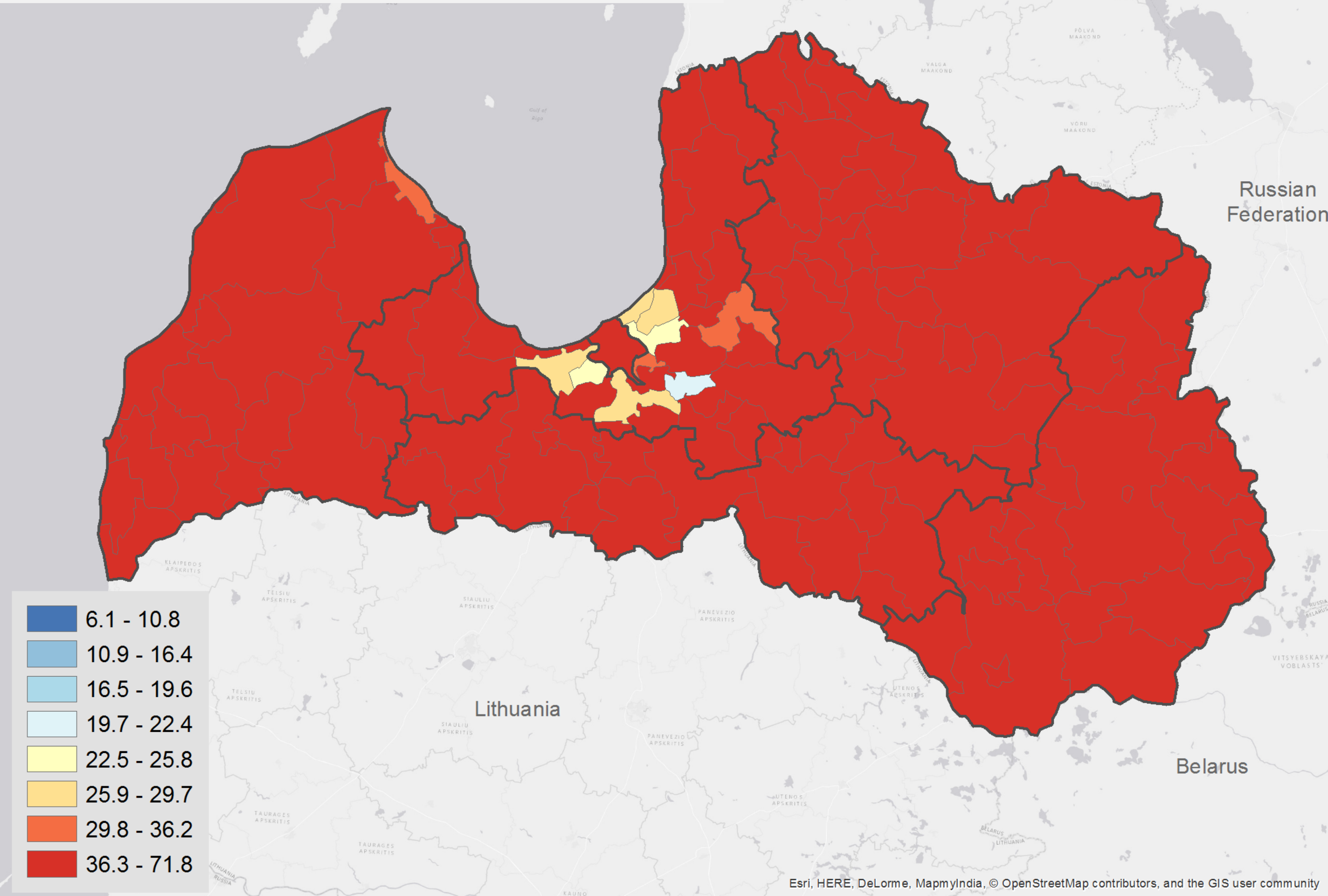
AROP Rates Among Adults with Only Secondary Education much Higher than the National AROP Rate (Scale Constant)



AROP Rates Among Unemployed Also Concentrated in the East



AROP Rates Among Unemployed Much higher than National Average



Per Capita PPP Results

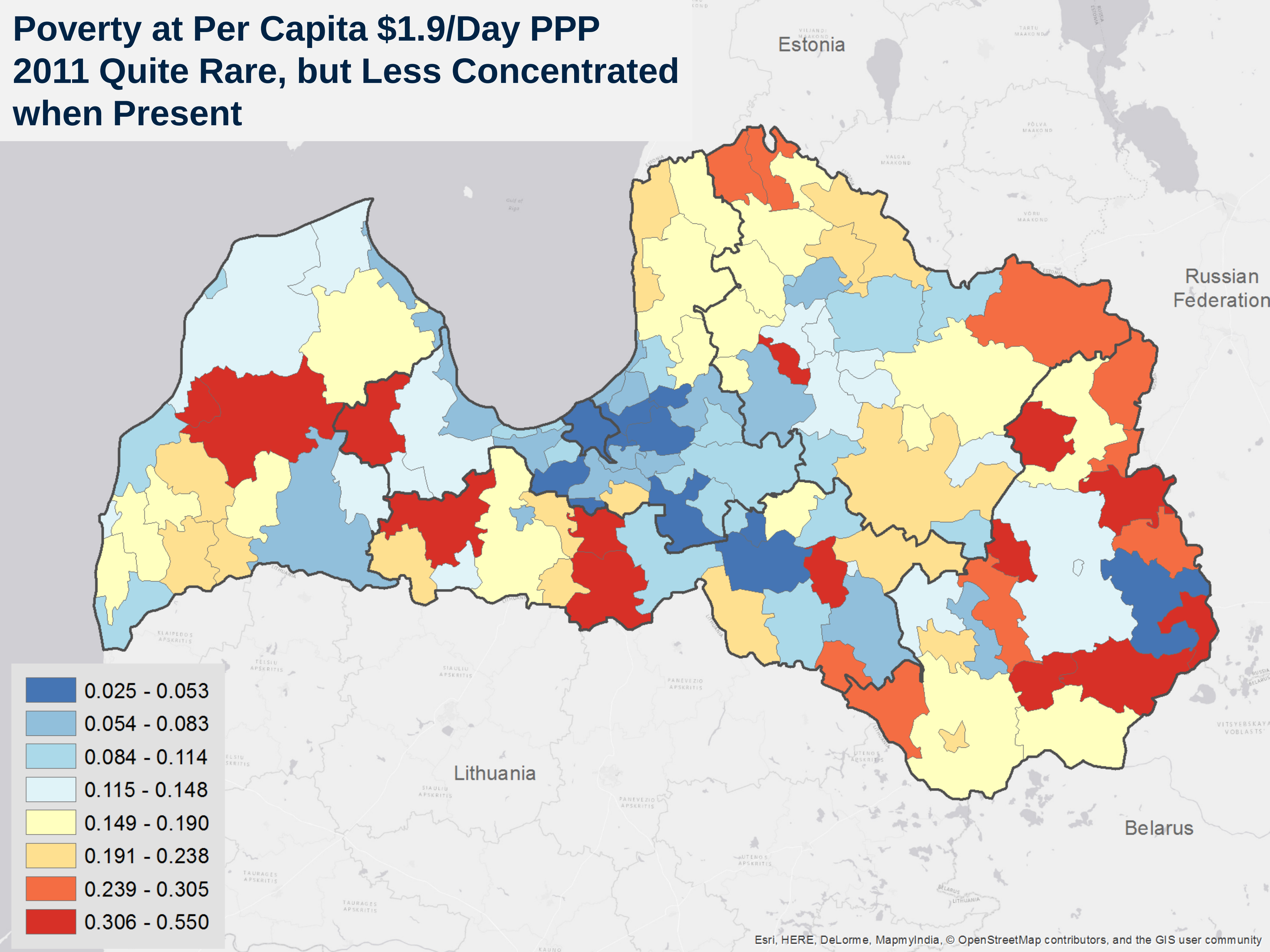
Poverty Using PPP and Per Capita Lines

Per Capita $1.9/Day PPP 2011

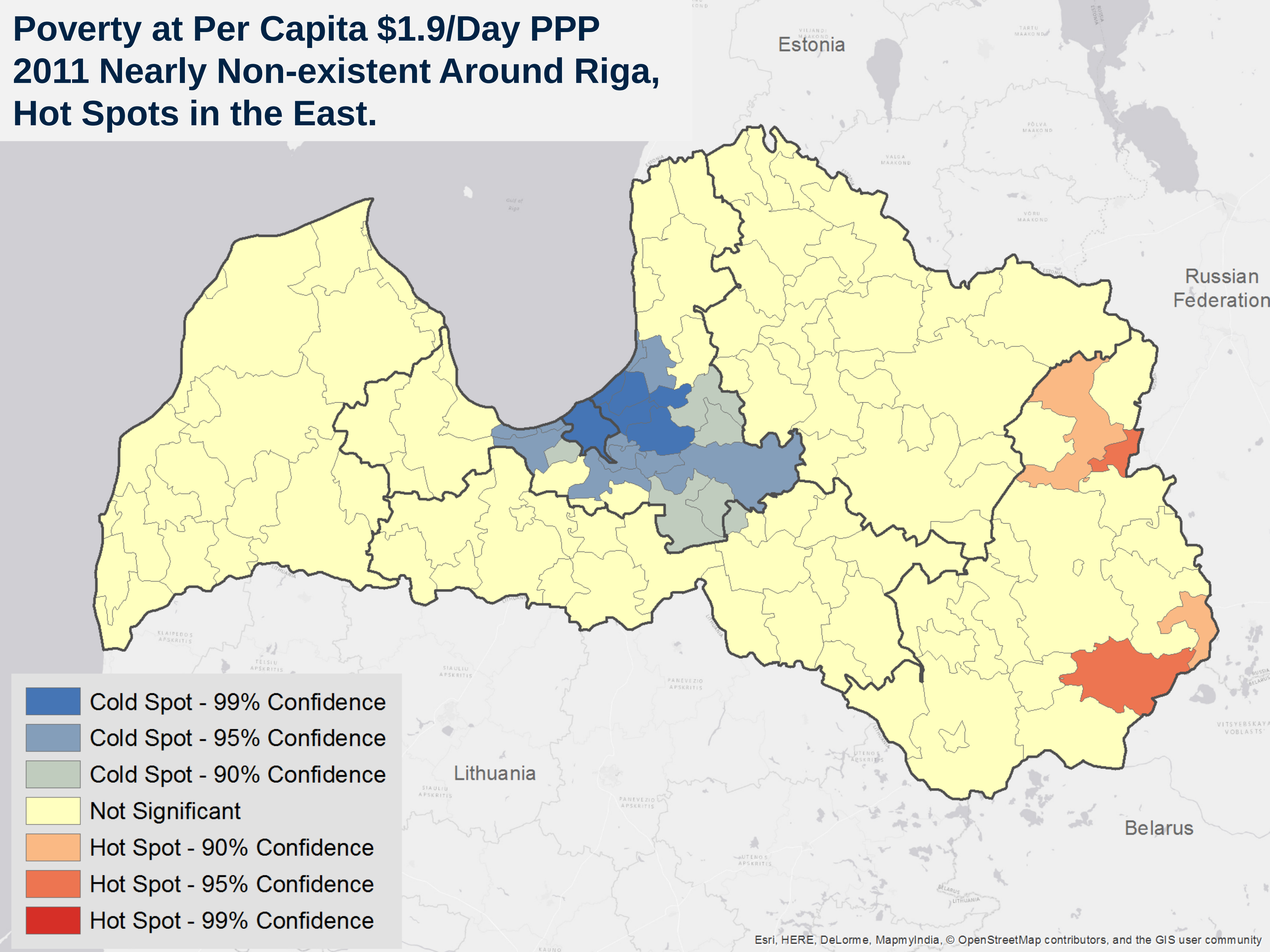
	SILC		Poverty Map	
	Rate	Std. Err.	Rate	Std. Err.
Kurzeme	1.09%	0.35%	0.12%	0.06%
Latgale	1.60%	0.39%	0.23%	0.10%
Pieriga	0.96%	0.32%	0.09%	0.06%
Riga	0.52%	0.16%	0.05%	0.02%
Vidzeme	1.23%	0.41%	0.14%	0.09%
Zemgale	1.07%	0.36%	0.16%	0.09%
National	0.97%	0.12%	0.12%	0.06%

WORLD BANK GROUP

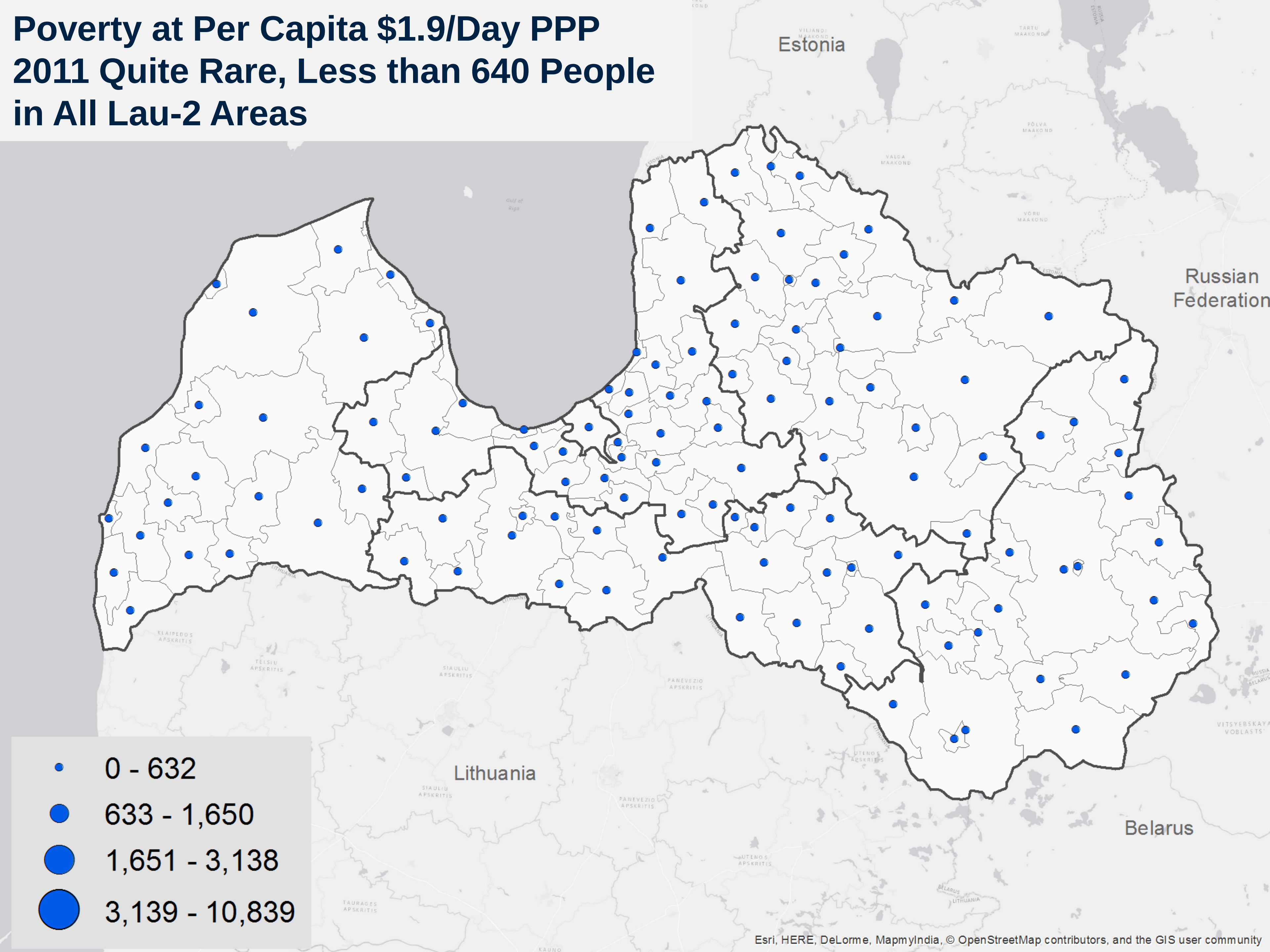
Poverty at Per Capita $1.9/Day PPP 2011 Quite Rare, but Less Concentrated when Present



Poverty at Per Capita $1.9/Day PPP 2011 Nearly Non-existent Around Riga, Hot Spots in the East.



Poverty at Per Capita $1.9/Day PPP 2011 Quite Rare, Less than 640 People in All Lau-2 Areas



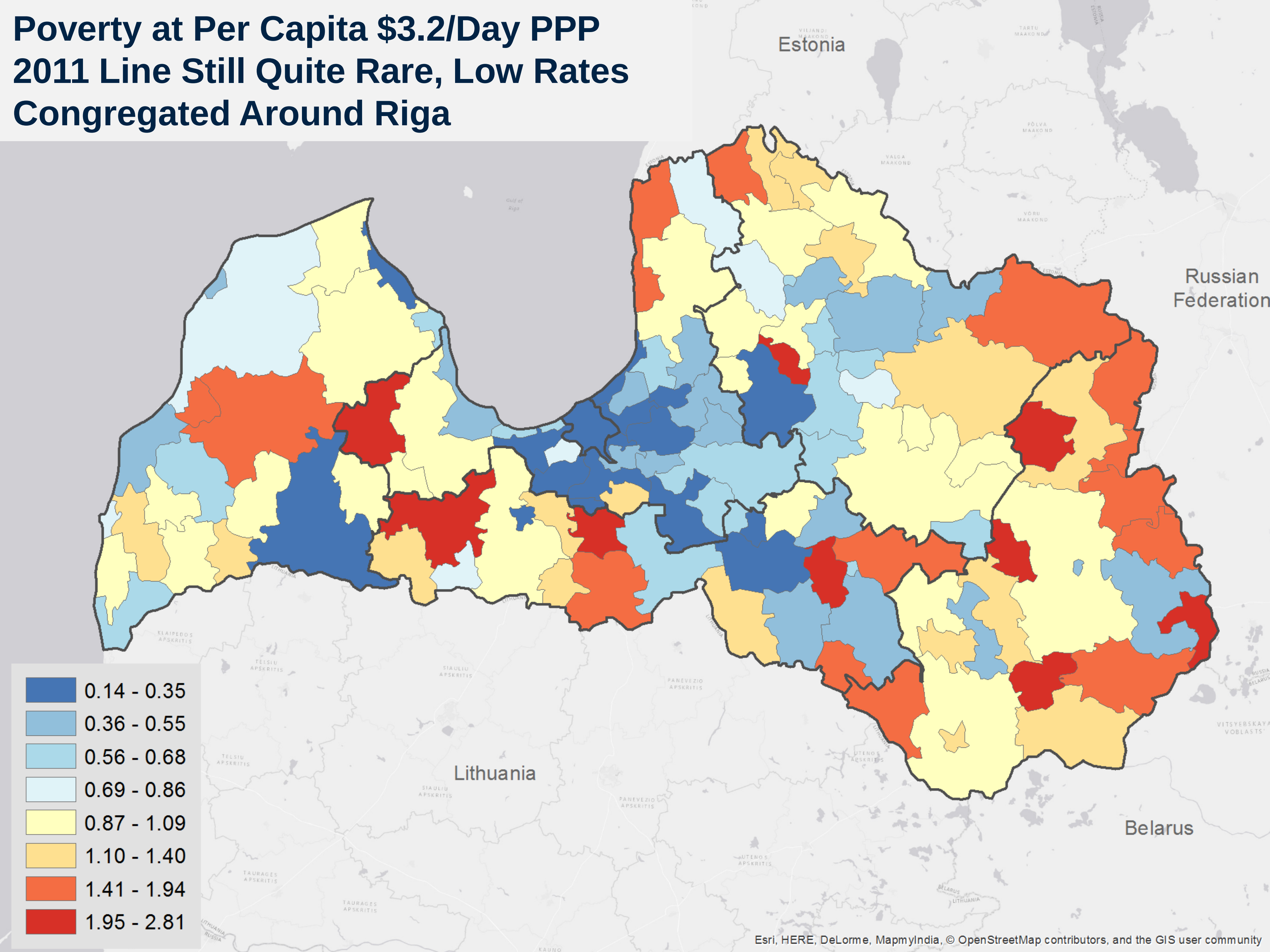
Using PPP and Per Capita Lines

Per Capita $3.2/Day PPP 2011

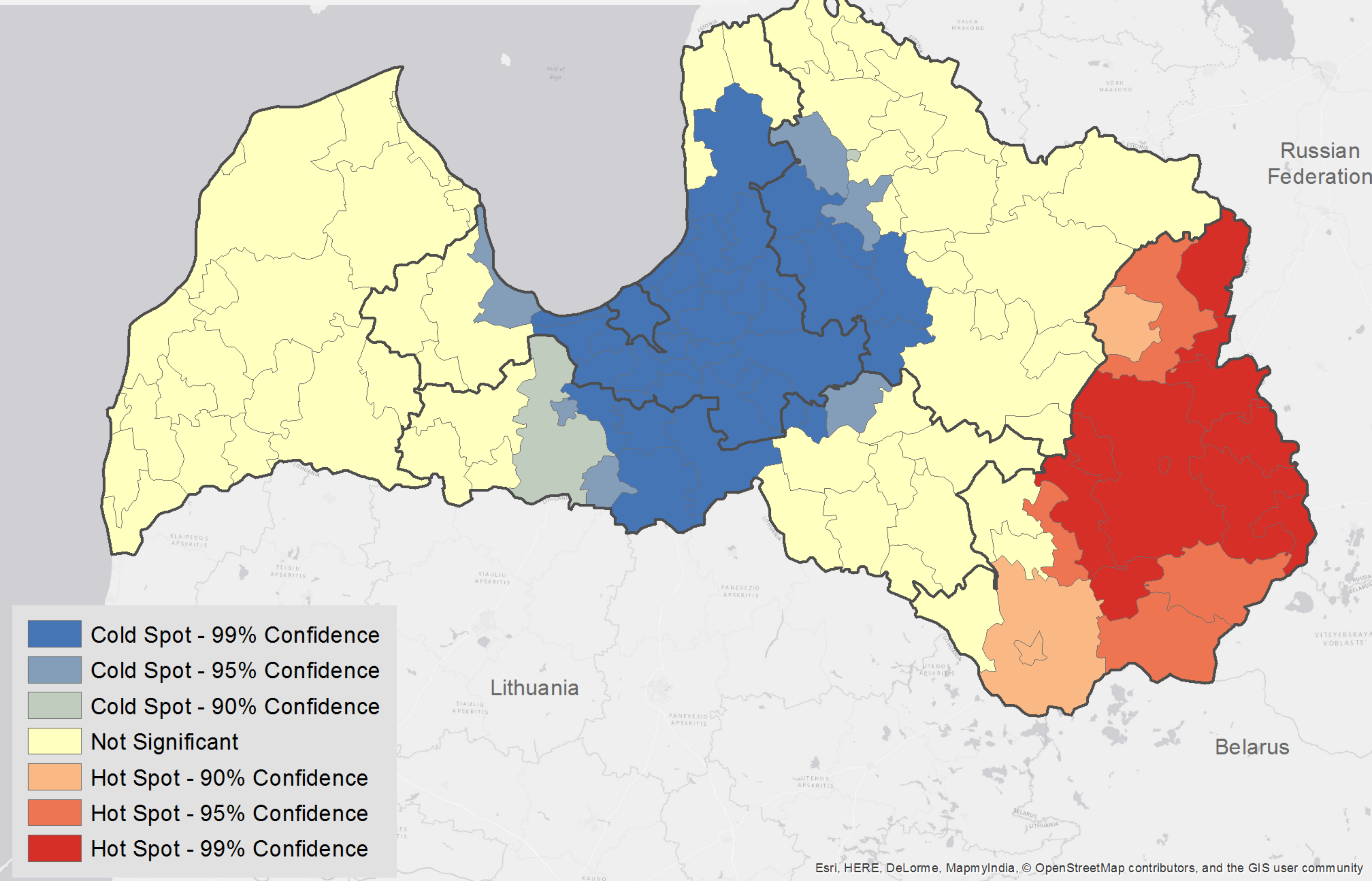
	SILC		Poverty Map	
	Rate	Std. Err.	Rate	Std. Err.
Kurzeme	1.22%	0.36%	0.74%	0.23%
Latgale	3.13%	0.54%	1.30%	0.35%
Pieriga	1.51%	0.41%	0.51%	0.21%
Riga	0.95%	0.21%	0.31%	0.06%
Vidzeme	2.33%	0.57%	0.83%	0.31%
Zemgale	2.65%	0.56%	0.96%	0.30%
Total	1.76%	0.16%	0.68%	0.20%

WORLD BANK GROUP

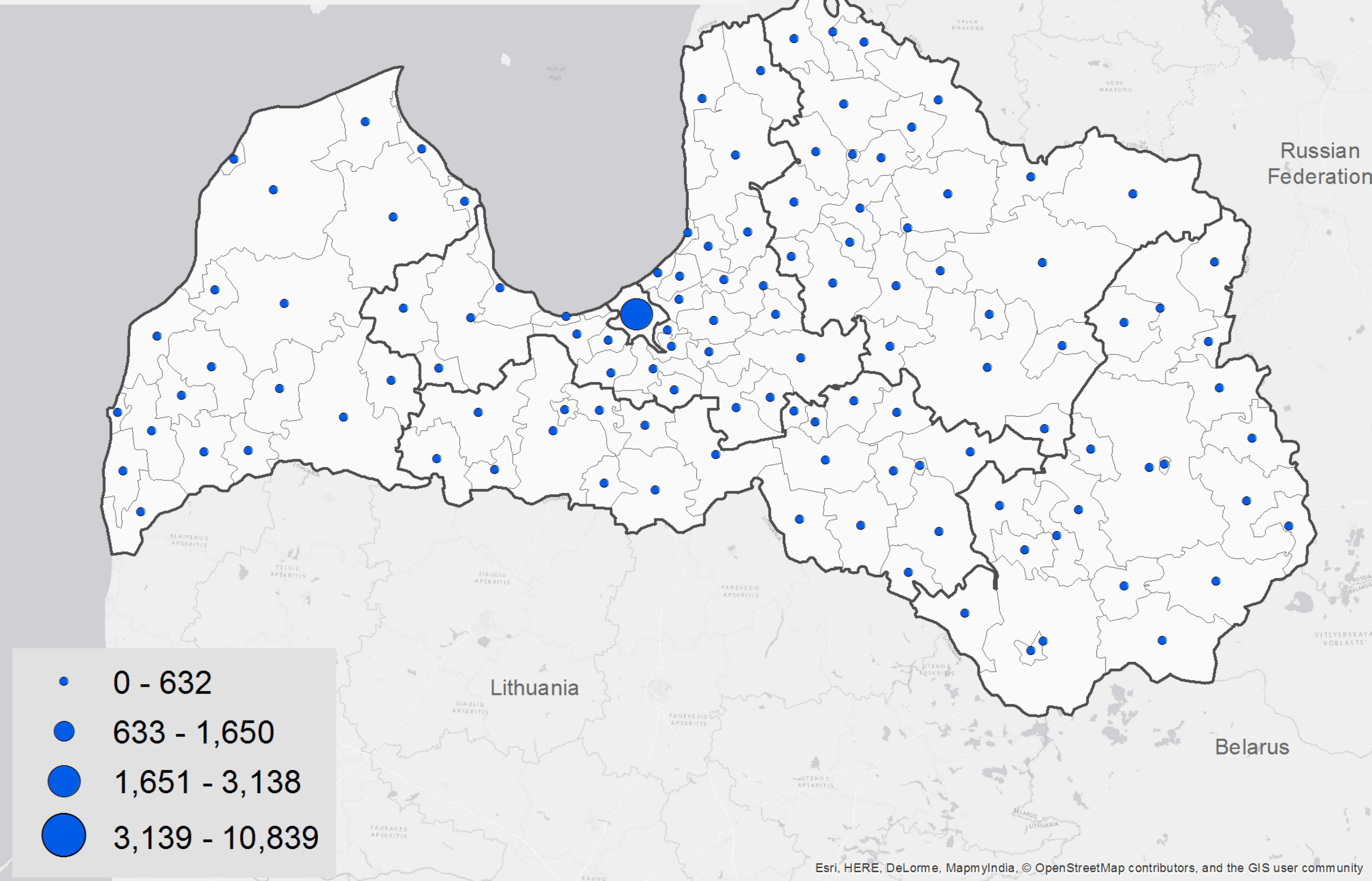
Poverty at Per Capita $3.2/Day PPP 2011 Line Still Quite Rare, Low Rates Congregated Around Riga



Poverty at Per Capita $3.2/Day PPP 2011 More Clearly Concentrated in East When Using Hot Spot Analysis



Number of Poor at Per Capita $3.2/Day PPP 2011 Still Rare, Only Breaks 633 People Total

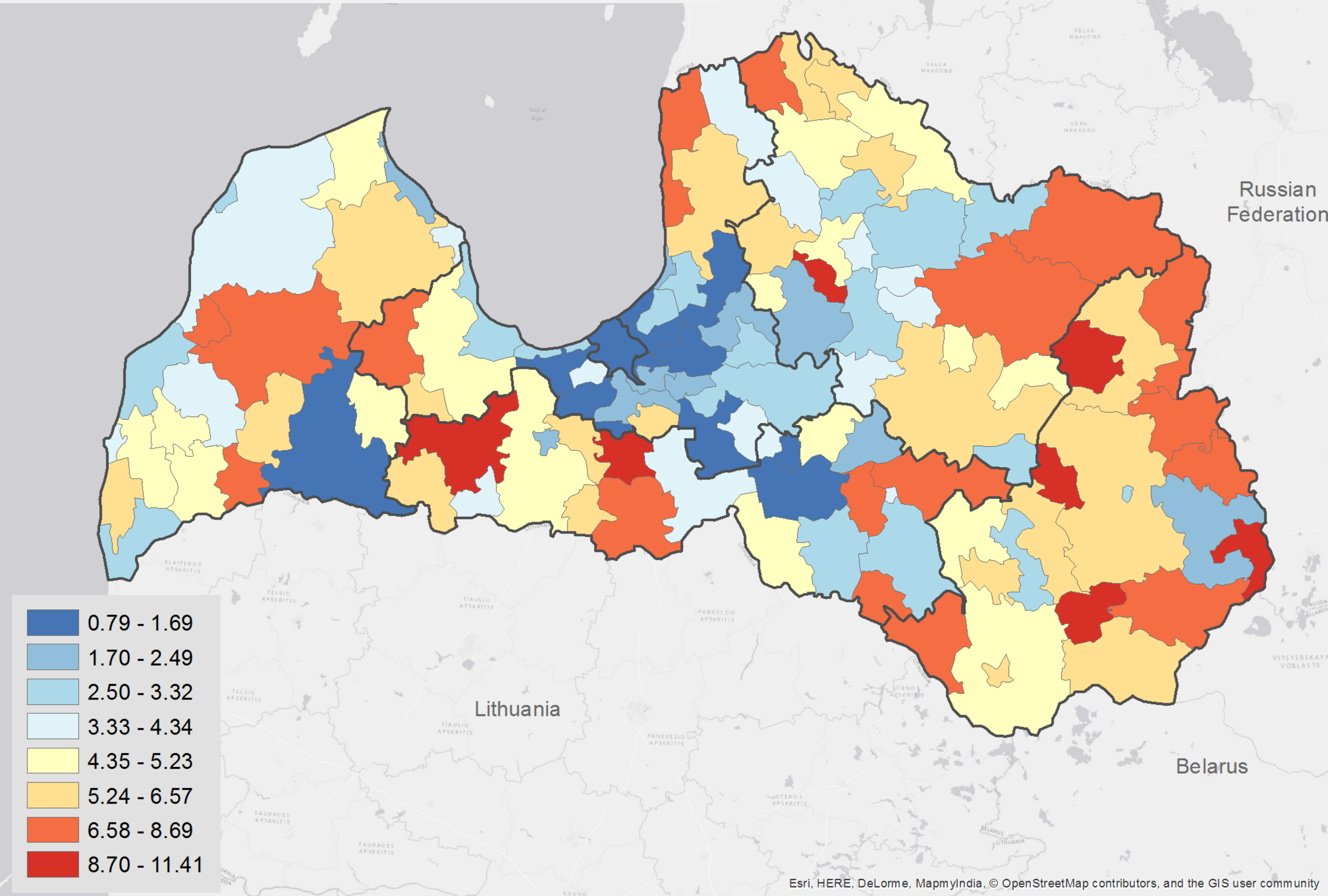


Using PPP and Per Capita Lines

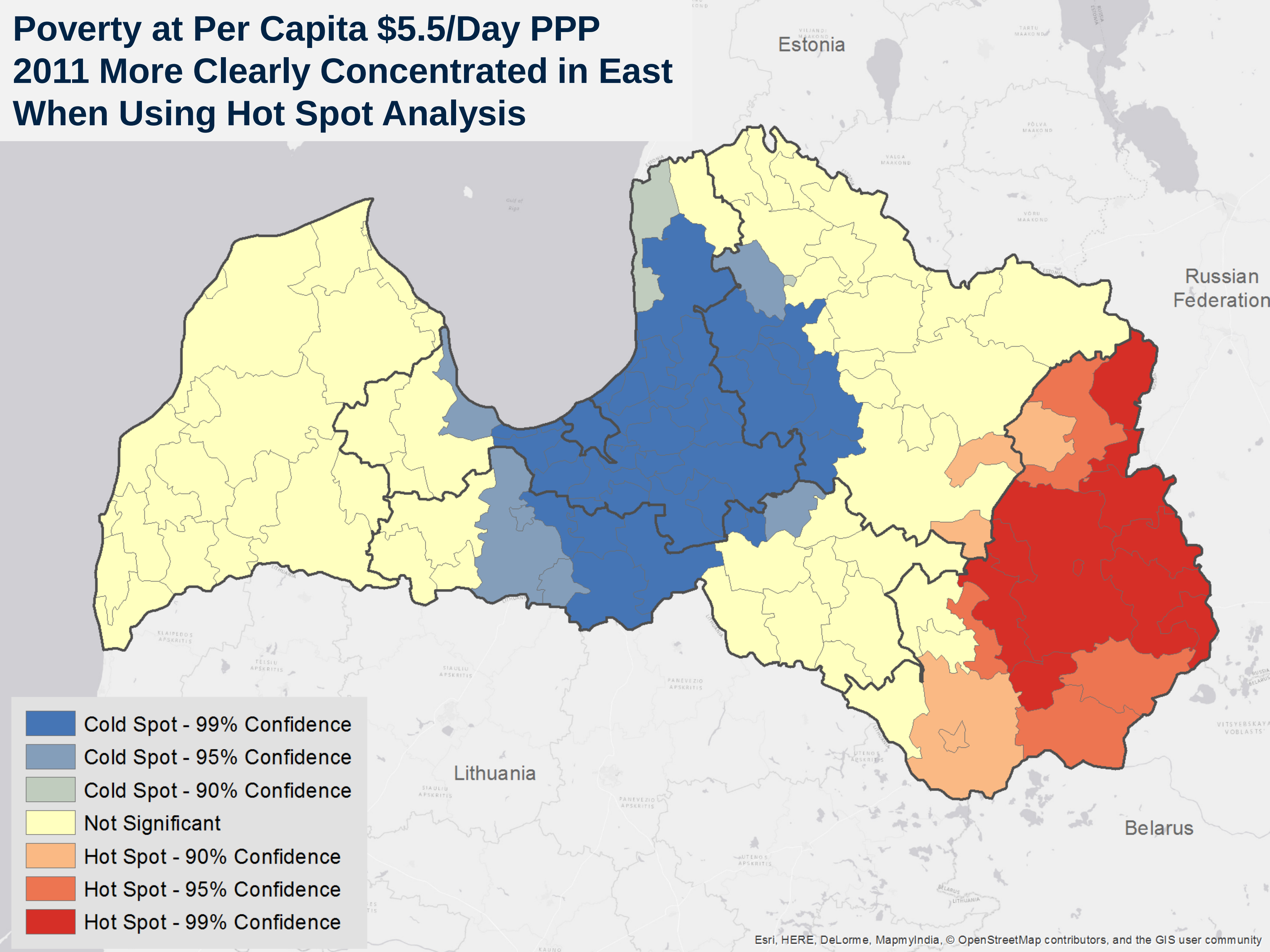
Per Capita $5.5/Day PPP 2011

	SILC		Poverty Map	
	Rate	Std. Err.	Rate	Std. Err.
Kurzeme	4.14%	0.66%	3.84%	0.81%
Latgale	7.37%	0.81%	5.90%	1.14%
Pieriga	3.34%	0.60%	2.68%	0.73%
Riga	2.58%	0.35%	1.68%	0.19%
Vidzeme	7.47%	0.99%	4.14%	1.03%
Zemgale	9.07%	1.00%	4.74%	1.04%
Total	4.92%	0.27%	3.38%	0.70%

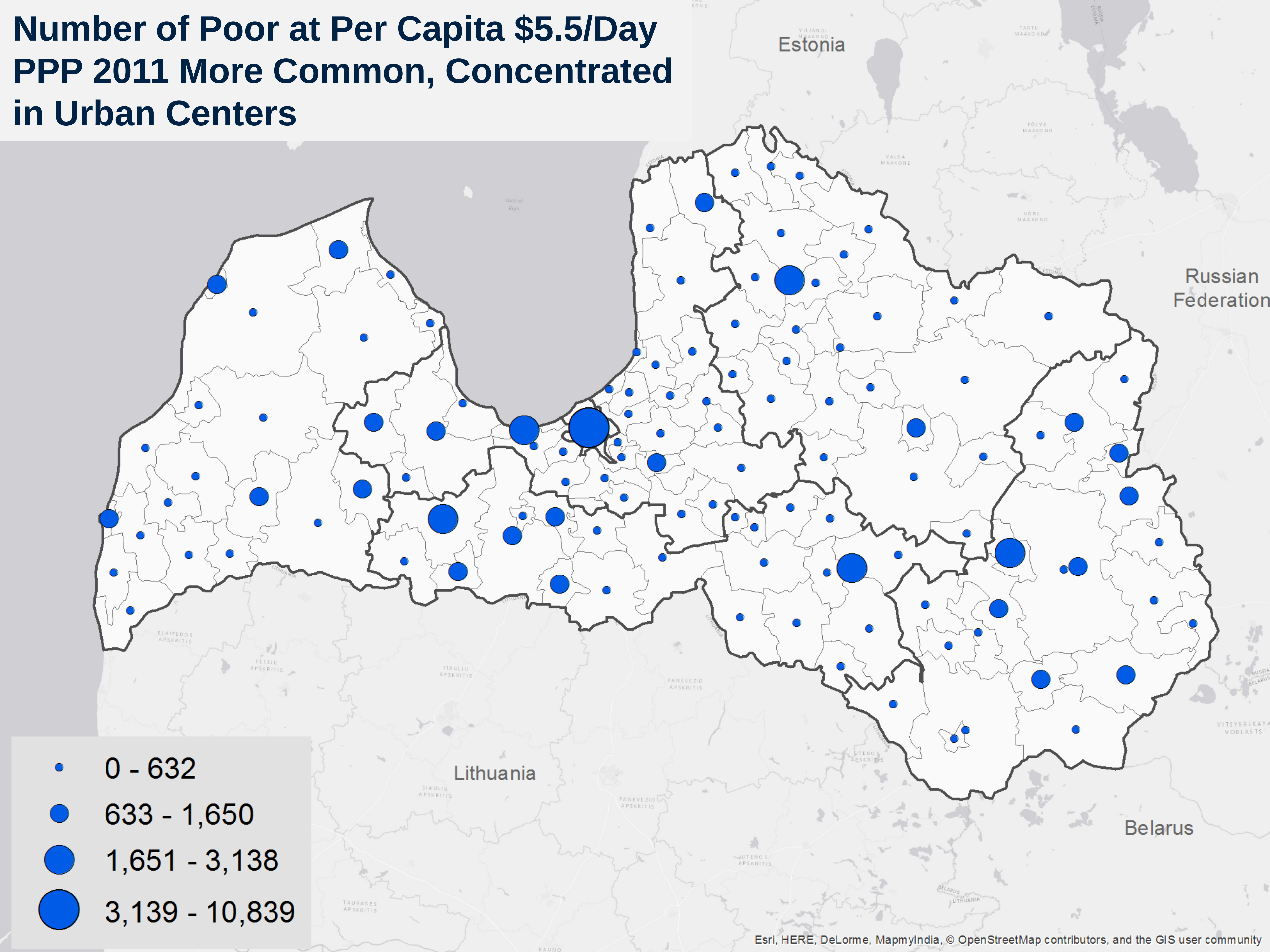
Poverty at Per Capita $5.5/Day PPP 2011 Follows Similar Spatial Pattern.



Poverty at Per Capita $5.5/Day PPP 2011 More Clearly Concentrated in East When Using Hot Spot Analysis



Number of Poor at Per Capita $5.5/Day PPP 2011 More Common, Concentrated in Urban Centers

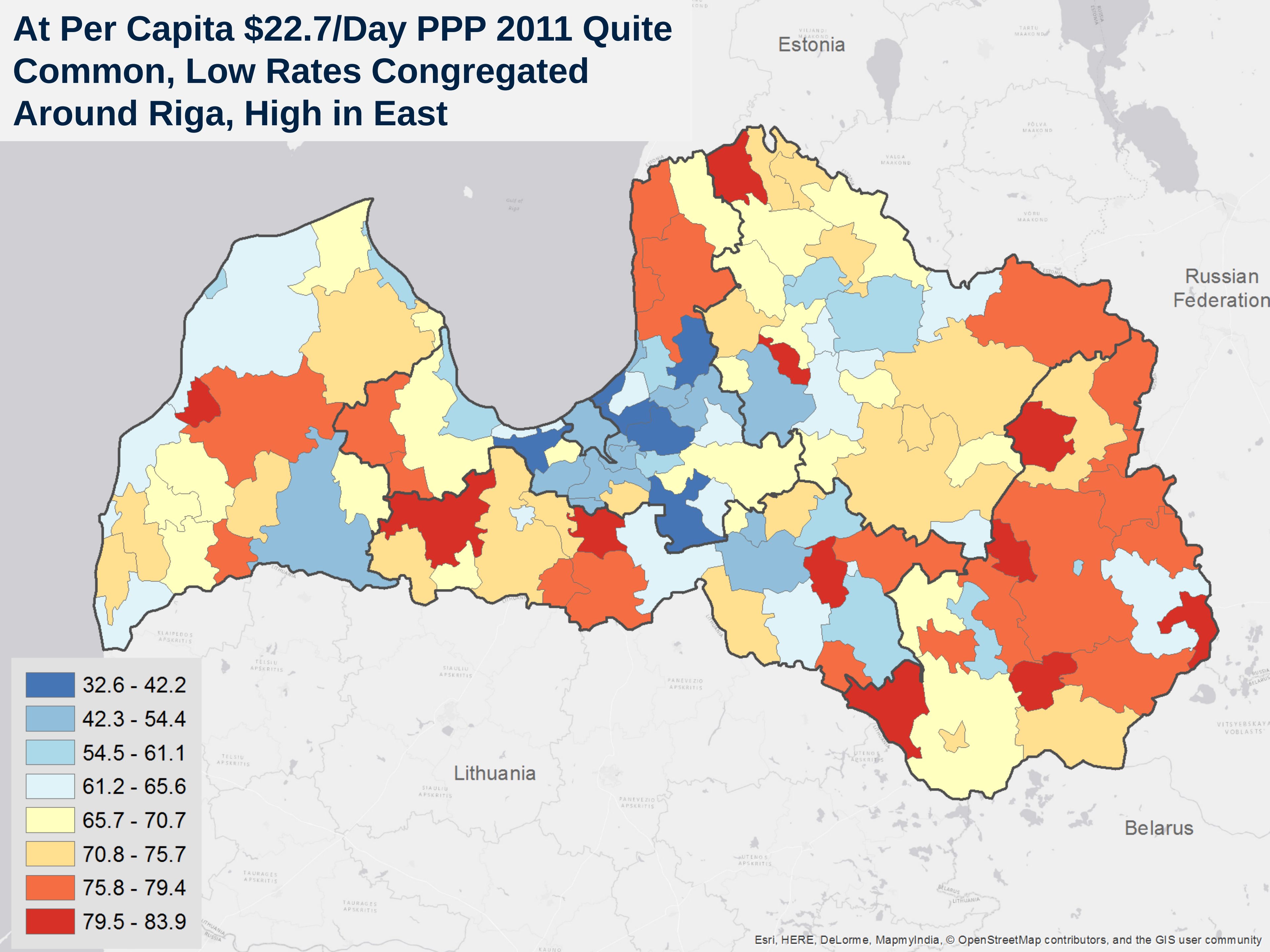


Using PPP and Per Capita Lines

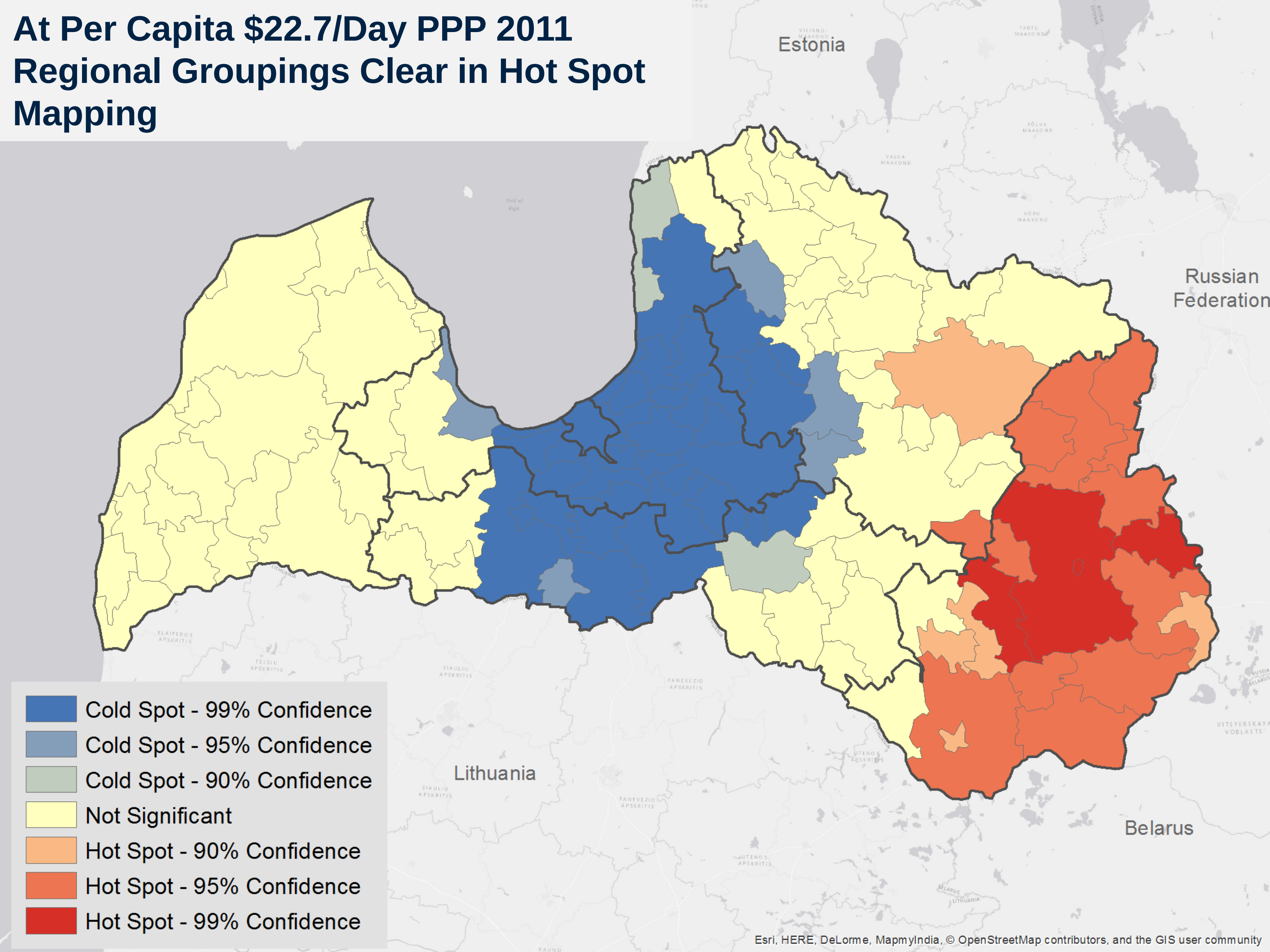
Per Capita $22.7/Day PPP 2011

	SILC		Poverty Map	
	Rate	Std. Err.	Rate	Std. Err.
Kurzeme	60.79%	1.62%	63.80%	4.23%
Latgale	74.64%	1.34%	72.38%	3.84%
Pierīga	59.64%	1.63%	54.92%	5.40%
Rīga	48.89%	1.09%	47.30%	1.22%
Vidzeme	67.91%	1.75%	66.12%	5.20%
Zemgale	66.22%	1.65%	68.20%	4.59%
Total	60.22%	0.61%	58.96%	3.56%

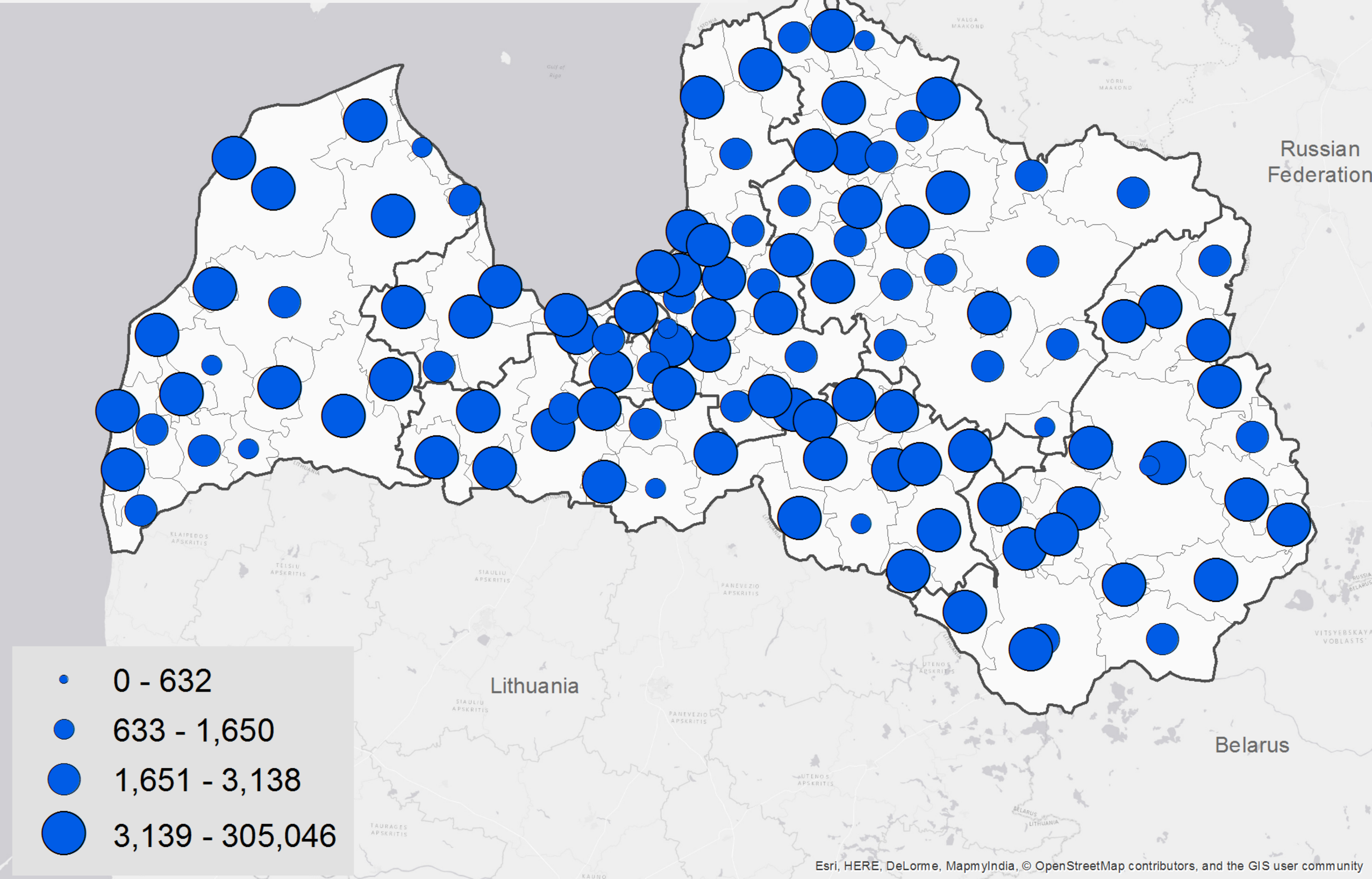
At Per Capita $22.7/Day PPP 2011 Quite Common, Low Rates Congregated Around Riga, High in East



At Per Capita $22.7/Day PPP 2011 Regional Groupings Clear in Hot Spot Mapping



At Per Capita $22.7/Day PPP 2011 All Areas have Signifiant Populations Living Below Lines



Prices and the Cost of Living

WORLD BANK GROUP

Spatial differences in Prices Can be Very Important for Poverty

What is a price index?

- A price index is useful in separating real income from nominal income
- Cost of living indexes allow for interpersonal welfare comparisons when the costs of living vary over time and space

Why should we consider cost of living across space?

- Ignoring price levels may lead to mistaken information as to where a country’s poor reside
  - The AROP threshold within a member state is the same everywhere, whether a family lives in the capital or in a rural area
  - It is not uncommon for welfare programs to be based on a level of income
  - It is implicitly assumed that the same income will yield the same level of utility across space. However, if two households face different price levels, welfare cannot be compared across time or space.

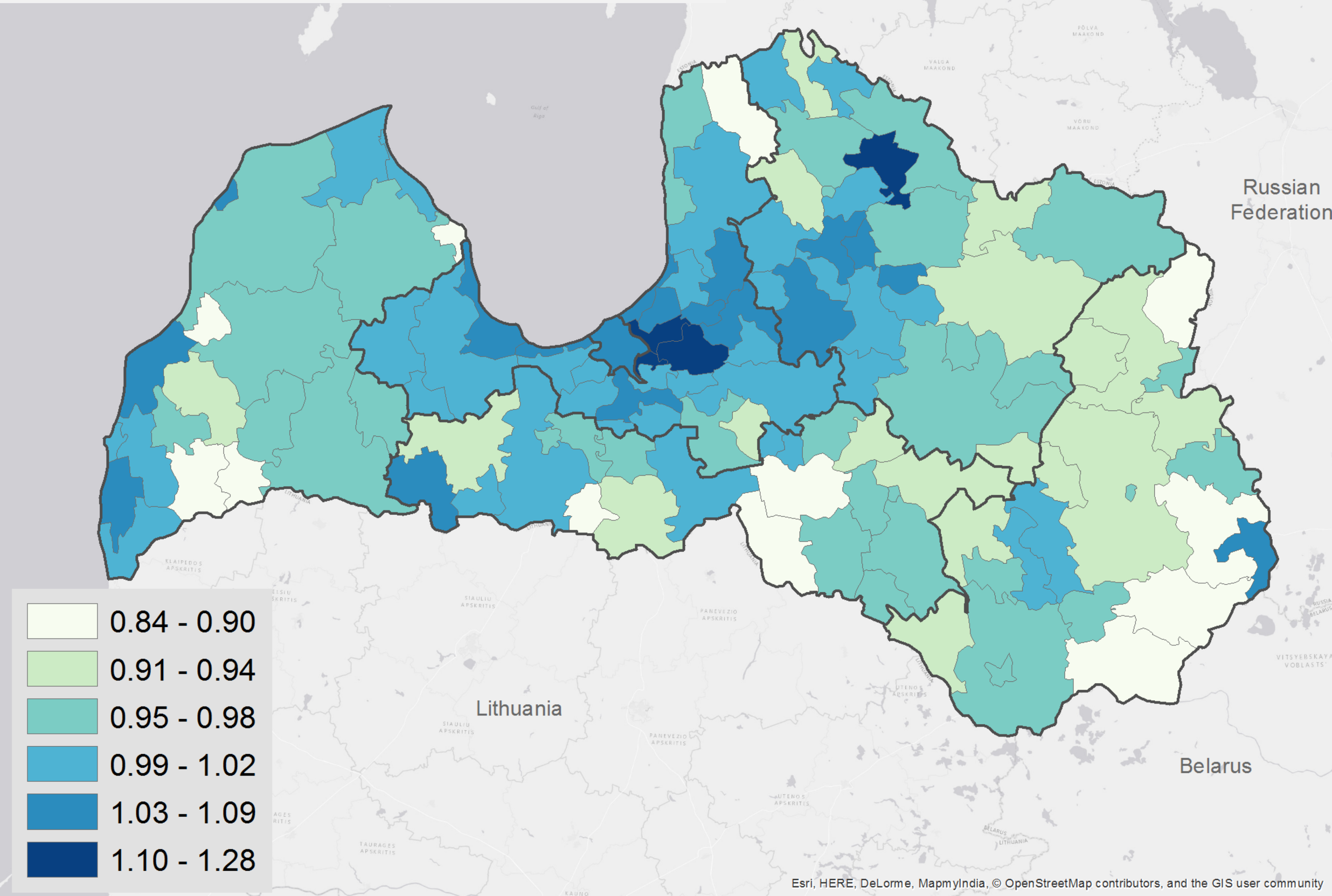
The HBS and SILC Allow for Estimating the Importance of Spatial Differences in Prices in Latvia

How to incorporate it into our work?

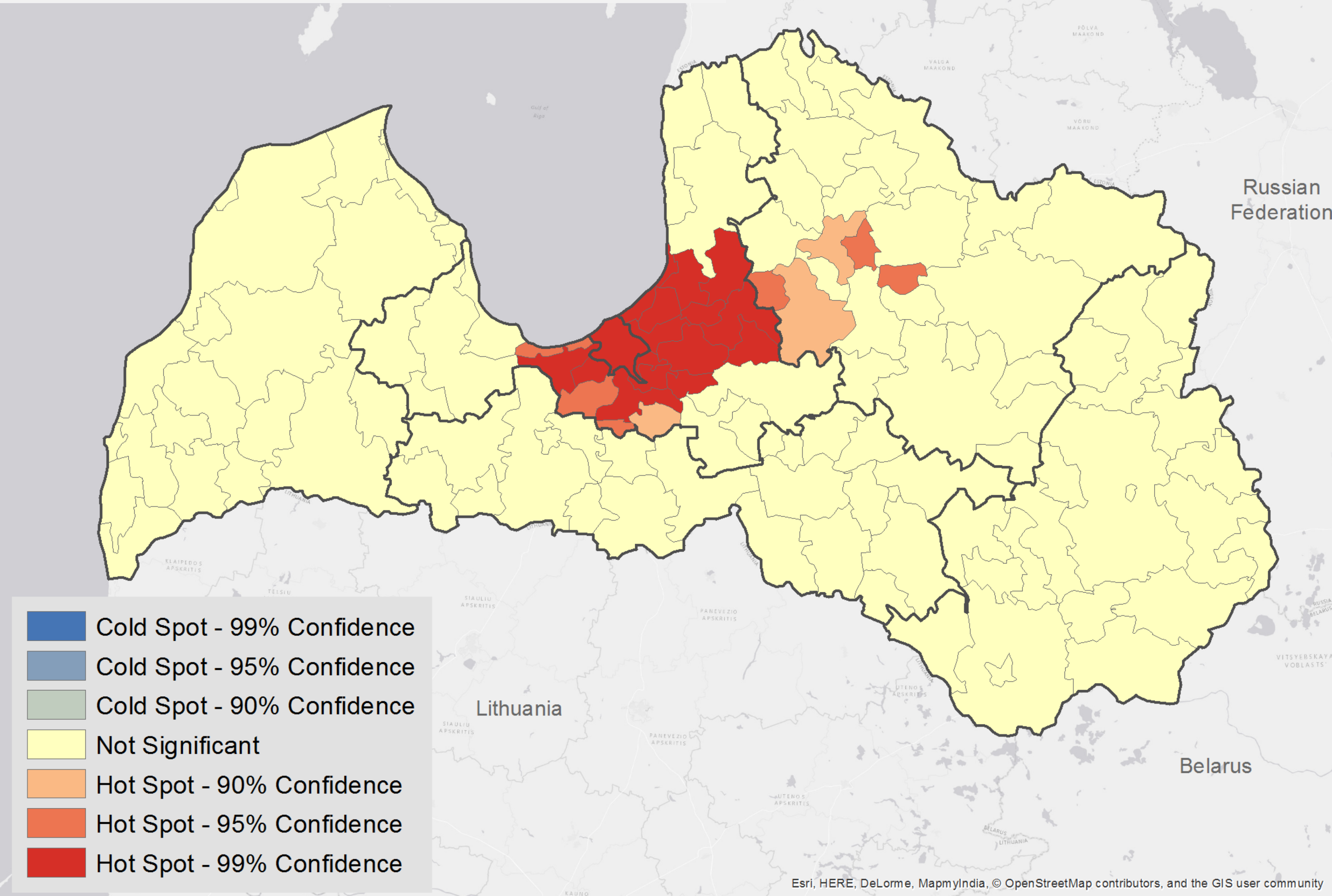
1. Use a Paasche index based on food expenditures in the HBS
   - This reveals how much better or worse off would be an individual who moved to the base region
   - For locations not in the HBS, a distance weighted average of the observed municipalities is used
2. Differences in food alone may not paint the whole picture
   - Use imputed rents relative to the national average (both SILC and HBS)

WORLD BANK GROUP

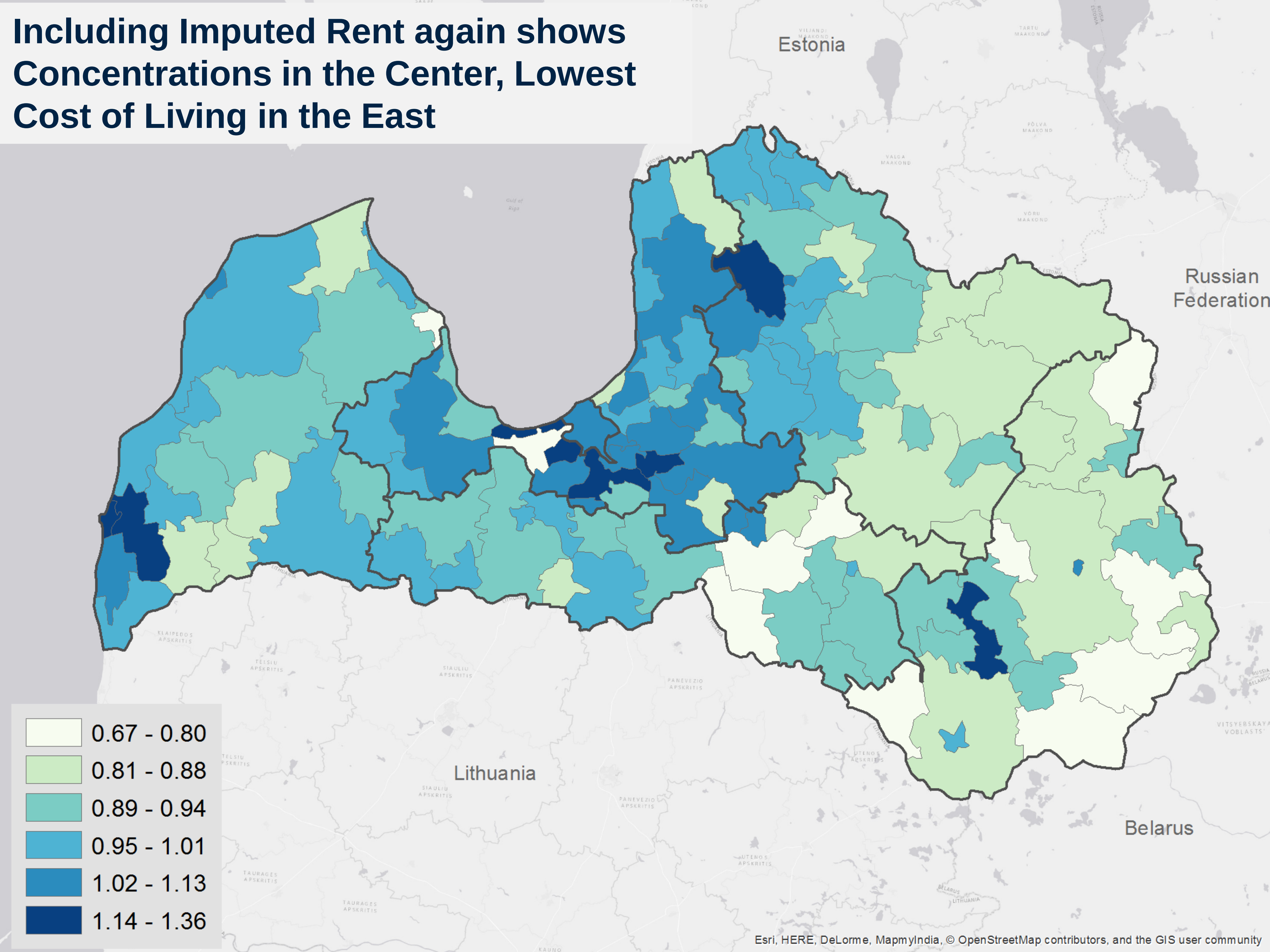
Food Prices are Much Higher in the Center



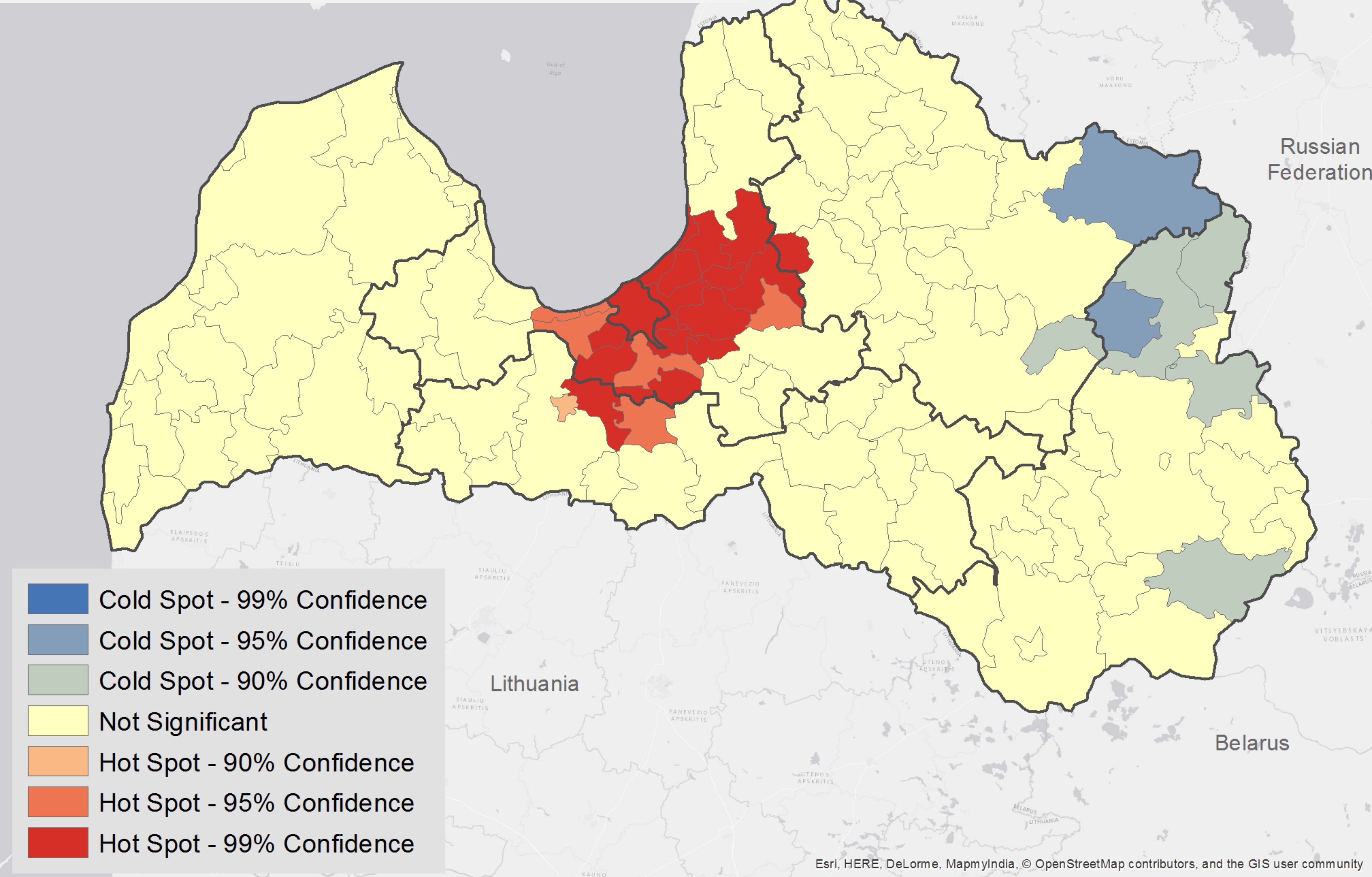
Hot Spot Maps Show that Food Prices are Highest Around Riga



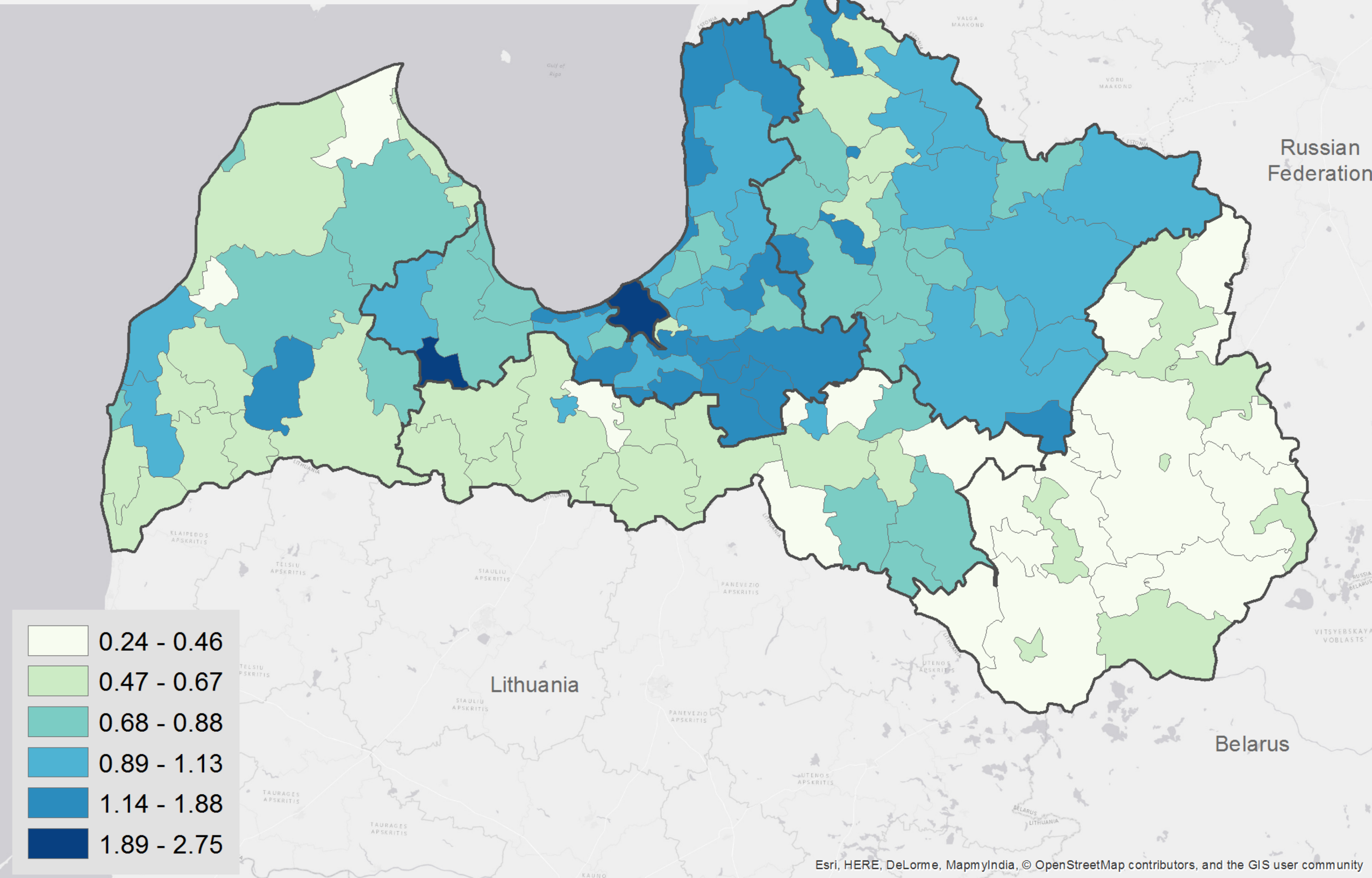
Including Imputed Rent again shows Concentrations in the Center, Lowest Cost of Living in the East



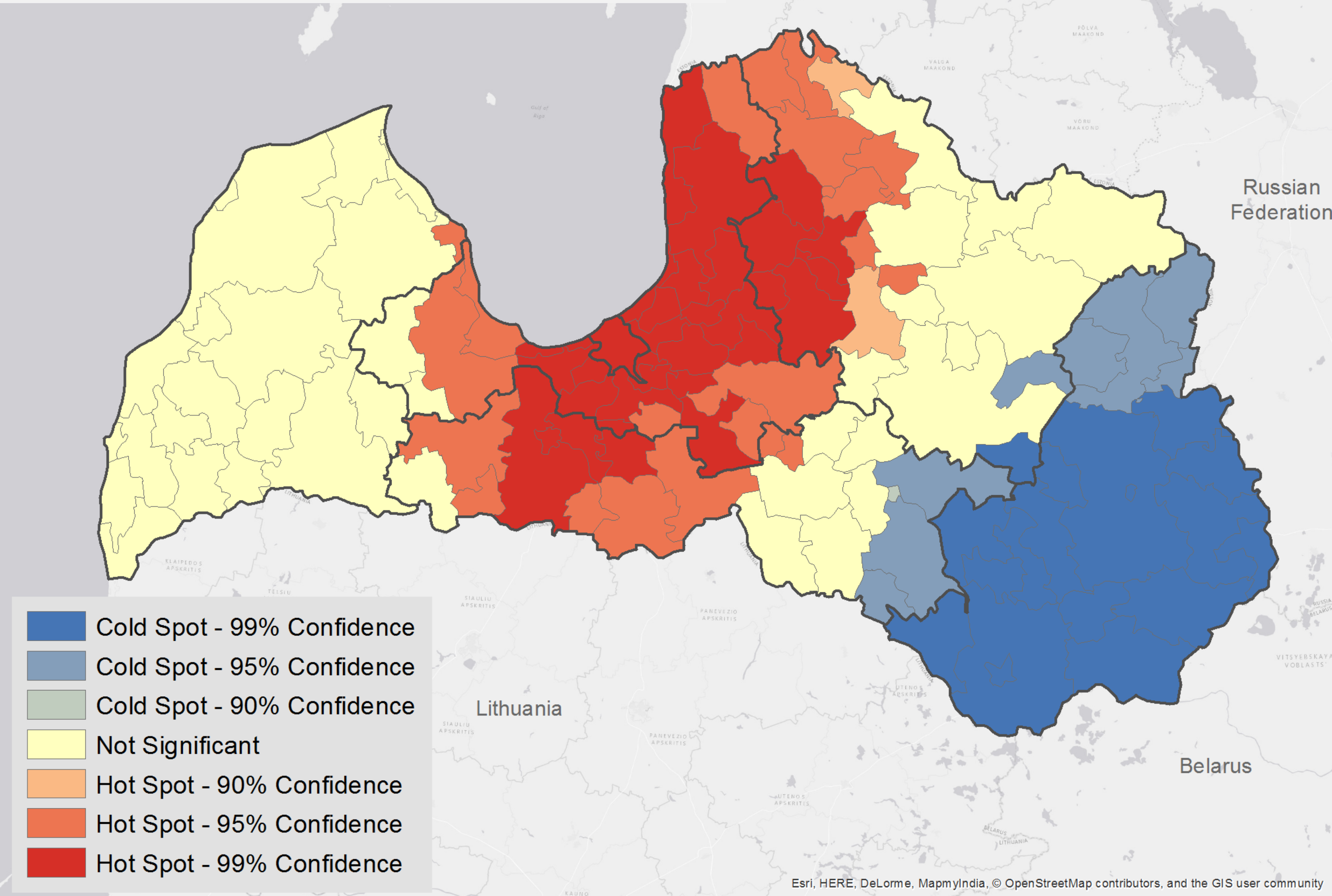
Hot Spot Map Again Shows High Concentrations in the Center, Lowest Cost of Living in the North East



Rent in the SILC Shows Similar High Concentrations in the Center, Lowest in the North East



Hot Spots Show Even Larger Regional Divide in Imputed Rent in the SILC

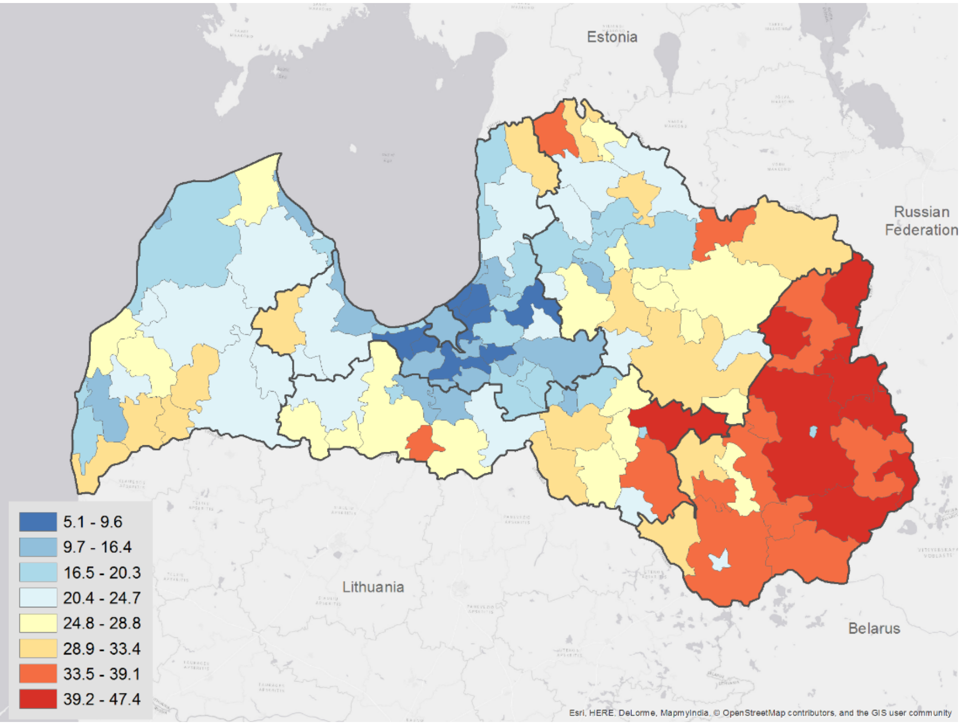


Results with (Food-based) Spatial Deflation

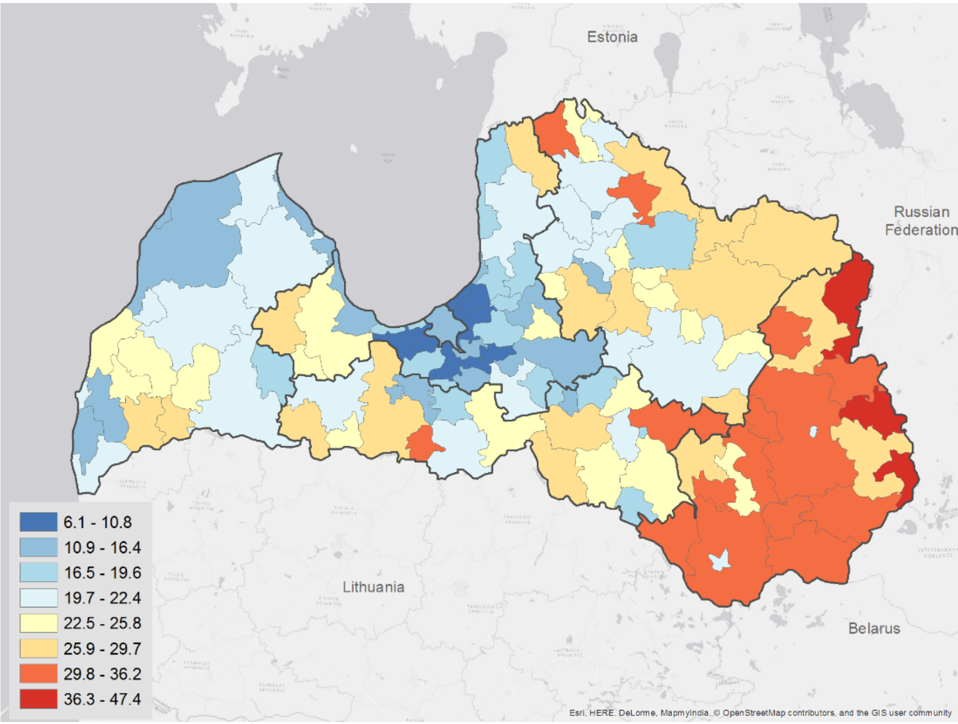
	AROP			
	Rdef		Povmap V2	
	AROP	Std. Err.	AROP	Std. Err.
Riga	12.8	2.5	12.7	0.6
Pieriga	16.1	3.9	15.4	3.1
Kurzeme	19.7	4.6	20.5	3.2
Zemgale	20.8	4.7	23.1	4.0
Vidzeme	22.9	5.3	23.6	4.4
Latgale	27.2	5.4	30.4	4.1
National	18.4	4.0	19.2	2.7

Poverty Map with Spatial Deflation Narrows the Divergence Between East and Center

No Spatial Deflation



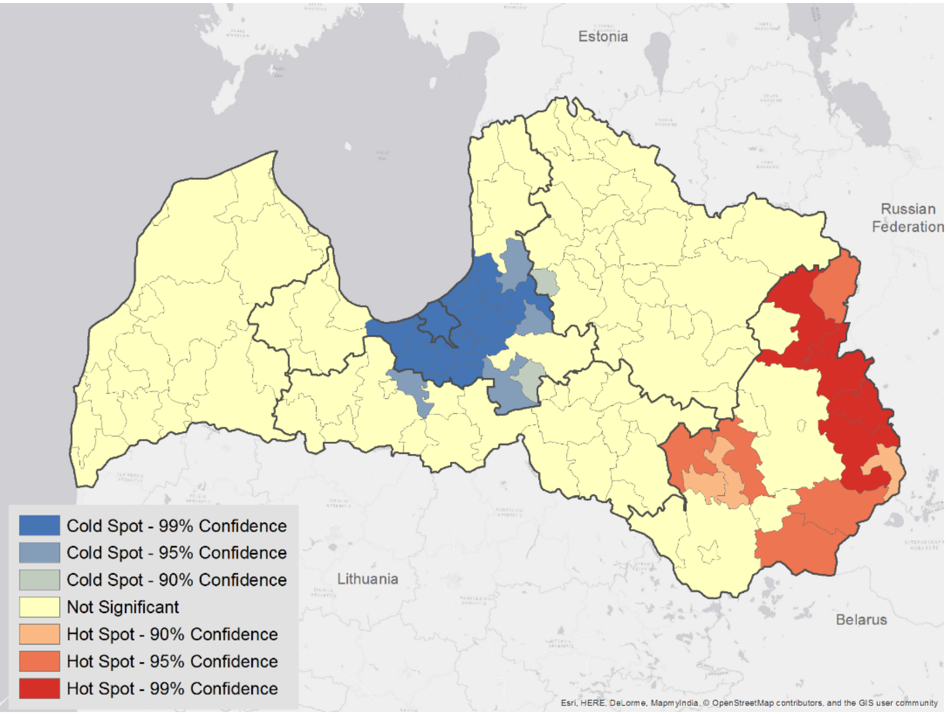
With Spatial Deflation



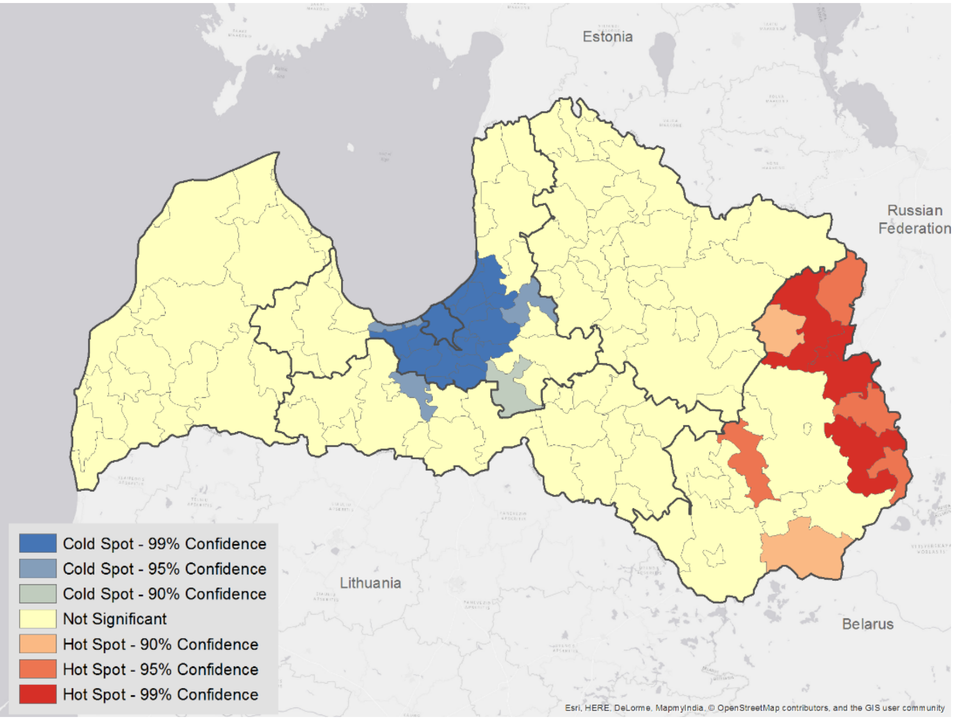
WORLD BANK GROUP

Poverty Map with Spatial Deflation Narrows the Size of Hot/Cold Spots

No Spatial Deflation



With Spatial Deflation



WORLD BANK GROUP

THANK YOU

APPENDIX: STEP BY STEP

- Bootstrap the survey (unless "parametric drawing" of the model parameters is used).
- Estimate β by means of OLS, and extract the residuals.
- Regress residuals from (2) on the area dummies (i.e., estimate Fixed-Effects (FE) model), and extract the residuals.
- Estimate the unconditional variance parameters of the nested error model (σ_η^2 and σ_ε^2) by applying Henderson-method-III (see Henderson, 1953), which uses the residuals the previous two steps.

APPENDIX: STEP BY STEP

- If heteroskedastic household errors are assumed, then: (a) derive estimates of the household errors by subtracting the area averages from the residuals (i.e., deviations from the area mean residual), (b) apply a logistic transformation to the errors derived under (a) to obtain the left-hand-side (LHS) of the regression (also referred to as the "alpha-model") that will be used to predict the conditional variance of ε_{ch}, denoted by $\sigma^2_{\varepsilon,ch}$, and (c) ensure that the unconditional variance is still equal to σ^2_{ε}, i.e., $E[\sigma^2_{\varepsilon,ch}] = \sigma^2_{\varepsilon}$.
- Given estimates of the unconditional variance σ^2_{η} and conditional variance $\sigma^2_{\varepsilon,ch}$, construct the covariance matrix $\Omega = E[\eta\eta^T + \varepsilon\varepsilon^T|x] = \sigma^2_{\eta} I_n + diag(\sigma^2_{\varepsilon,ch})$, which is used to obtain the GLS estimator for β.

APPENDIX: STEP BY STEP

- At this stage estimates have been obtained for all the model parameters: $\tilde{\beta}^{(r)}, \tilde{\sigma}_{\eta}^{2,(r)}, and\ \tilde{\sigma}_{\varepsilon,ch}^{2,(r)}$. Next, draw the area errors and the household idiosyncratic errors: $\tilde{\eta}_c^{(r)}$ and $\tilde{\varepsilon}_{ch}^{(r)}$ from their respective normal distributions with variances $\tilde{\sigma}_{\eta}^{2,(r)}\ and\ \tilde{\sigma}_{\varepsilon,ch}^{2,(r)}$.
- All necessary components to compute the round *r* simulated (log) income values for all households in the population census are available: $\tilde{y}_{ch}^{(r)} = x_{ch}^T\tilde{\beta}^{(r)} + \tilde{\eta}_c^{(r)} + \tilde{\varepsilon}_{ch}^{(r)}$.
- With the simulated household income data, compute the poverty and inequality measures as if the population census came with household income data from the start.
- This yields simulated poverty and inequality measures for each of the *R* simulation rounds. The averages of the simulated poverty and inequality measures provide the point estimates and the standard deviations provide the estimates of the corresponding standard errors.