Spatial Economics – Assignment 2

Gustav Pirich (h11910449) Peter Prlleshi () Filip Lukijanovic ()

April 2, 2024

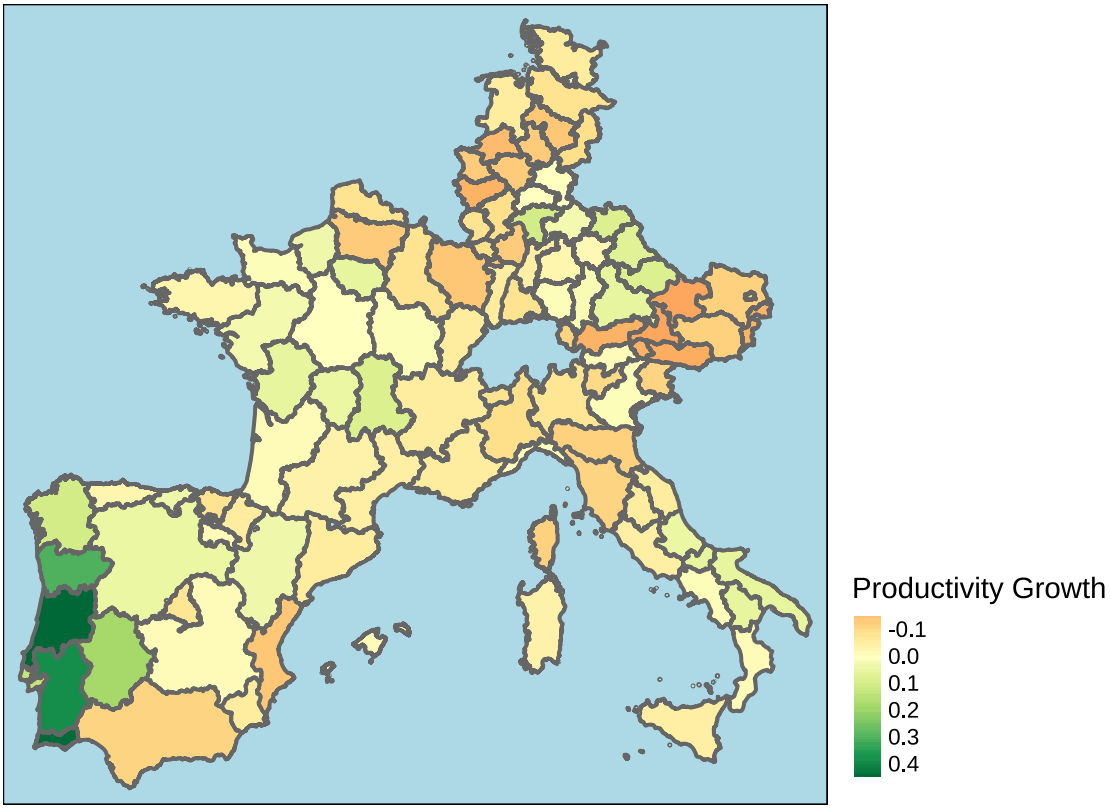
Contents

The code that was used in compiling the assignment is available on GitHub at https://github.com/gustavpirich/spatial_econ/blob/main/02_assignment/02_assignmnet.Rmd.

Exercise A

Calculate the growth rate of productivity from 1980 to 2013 and create a map that shows the productivity growth for each region.

The map shows the productivity growth rates in the NUTS-2 regions for the selected countries. We can see that many regions especially in Germany, Austria, and France exhibited negative productivity growth over the selected time period. Notably, Portugal's productivity has been growing the fastest. We suspect that the negative growth rates can be explained by the fact that high-income countries had a high baseline productivity to being with, while Portugal had a low baseline productivity. This could be evidence of convergence among productivity differences across Europe.



Generate three different spatial weights matrixes using (i) a distance threshold, (ii) smooth distance-decay, and iii) a contiguity-based measure.

(i) Distance Threshold

First, we create a binary distance threshol spatial weights matrix based. Any region is being assigned a '1', if the center of any other region is less than 3 km away. Note that we have chosen this value so that every region has a neighbor. We use the nb2mat function from the 'spdep' package. We row-normalize the matrix.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.4377  0.8376  1.0191  1.1436  1.4705  2.6527
```

(ii) Smooth-Distance Decay

Next, we create a spatial weights matrix based on a smooth distance-decay. We use the following simple distance decay function $w_{i,j}(d) = 1/d$. We calculate the weights for each neighboring region based on the k=1 nearest neighbors. We do not row-normalize the matrix.

(iii) Contiguity-based measure

Finally, we calculate a contiguity based measure, which we row normalize as well.

Compare the matrices; use your knowledge of graph theory and linear algebra

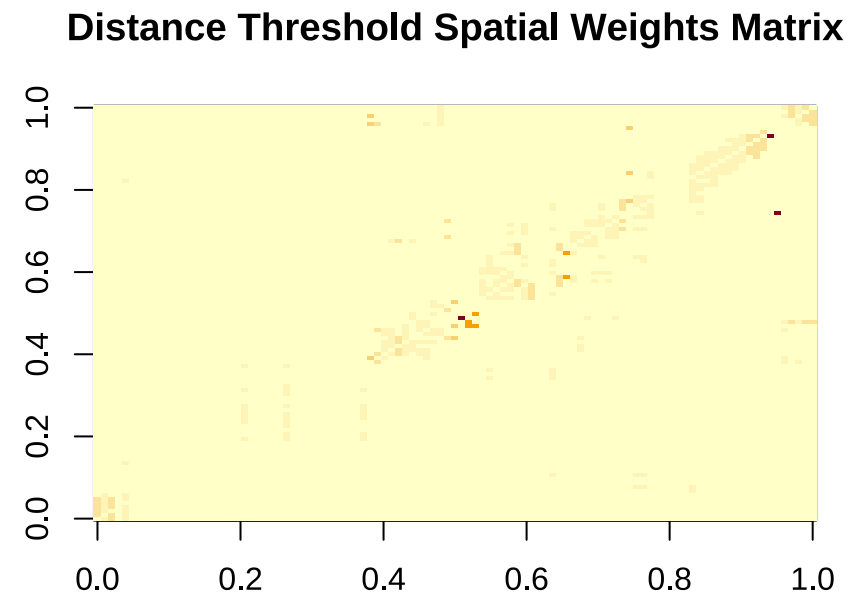
Distance Threshold Graph: Number of vertices: 103 Number of edges: 514 Average path length: 0.6750662 Graph density: 0.09784885 Average degree: 9.980583 Max Eigenvector Centrality: 1 Min Eigenvector Centrality: 0.008694479 Average Eigenvector Centrality: 0.1497447 Most Central Unit (Vertex ID): 49 Most Central Unit (Vertex Name):

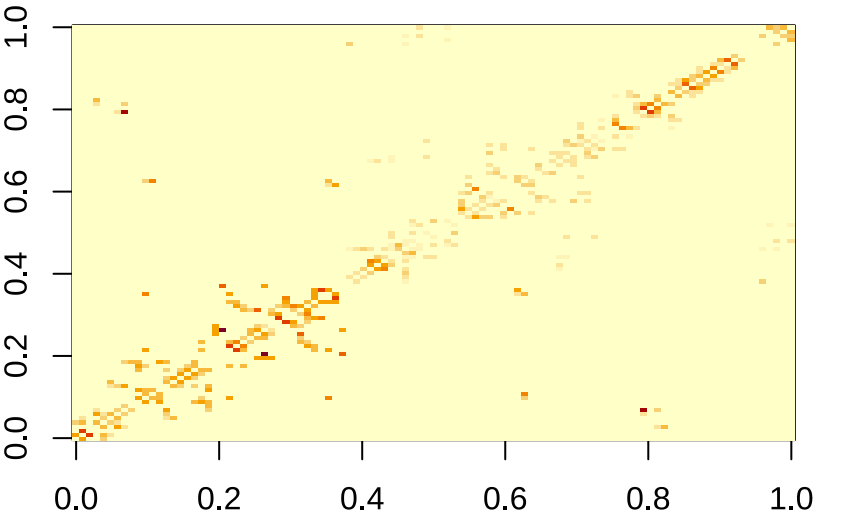
Smooth Distance-Decay Graph: Number of vertices: 103 Number of edges: 40 Average path length: 4.531818 Graph density: 0.007614696 Average degree: 0.776699 Max Eigenvector Centrality: 1 Min Eigenvector Centrality: 0 Average Eigenvector Centrality: 0.05949516 Most Central Unit (Vertex ID): 35 Most Central Unit (Vertex Name):

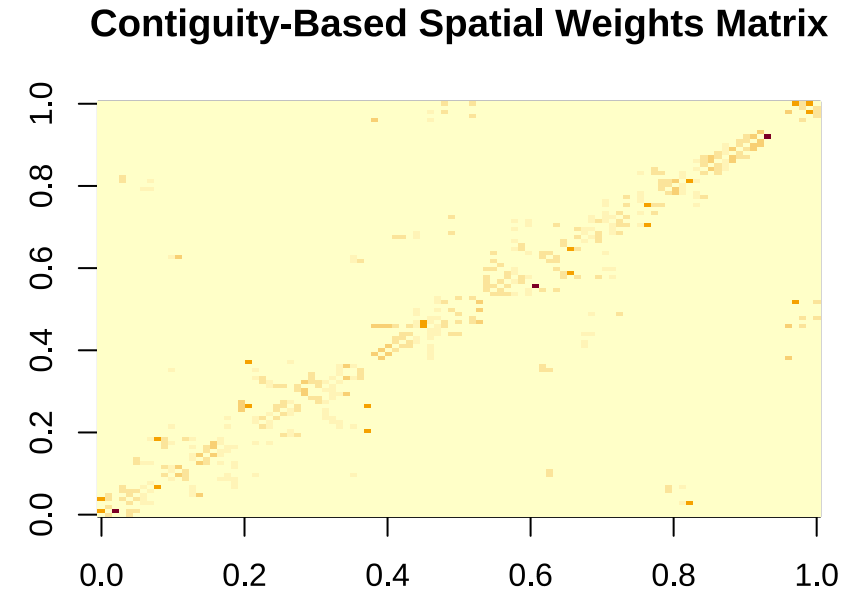
Contiguity-Based Graph: Number of vertices: 103 Number of edges: 222 Average path length: 1.288804 Graph density: 0.04226156 Average degree: 4.31068 Max Eigenvector Centrality: 1 Min Eigenvector Centrality: 1.697653e-16 Average Eigenvector Centrality: 0.08975192 Most Central Unit (Vertex ID): 48 Most Central Unit (Vertex Name): We compare the matrices based on a set of characteristics.The Distance Threshold Graph has 514 edges, indicating a higher level of connections between vertices compared to the other two. This suggests that many vertices are within the distance threshold to be considered neighbors. Smooth Distance-Decay Graph has only 40 edges, significantly fewer than the other graphs. This implies a much stricter criterion for considering two vertices to be connected, likely reflecting a much tighter control over distance decay effects. Contiguity-Based Graph strikes a balance with 222 edges, suggesting its connectivity criteria (likely based on shared boundaries or proximities) result in a moderate level of connections. Average Path Length Distance Threshold Graph shows a very low average path length (0.6750662), meaning that, on average, vertices are closely connected, allowing for short paths between them. Smooth Distance-Decay Graph has a much higher average path length (4.531818), indicating that paths between vertices are generally longer, reflecting the sparse connections. Contiguity-Based Graph has an average path length (1.288804), which is between the other two, suggesting moderate ease of movement across the graph compared to the other types. Graph Density Distance Threshold Graph has a density of 0.09784885, which is higher than the others, reflecting its higher number of edges and suggesting a relatively dense network. Smooth Distance-Decay Graph is the sparsest, with a density of 0.007614696, aligning with its fewer edges and longer path lengths. Contiguity-Based Graph has a density of 0.04226156, which, while lower than the distance threshold graph, suggests a moderately dense network, possibly due to its criteria for connections. Average Degree Distance Threshold Graph has a high average degree (9.980583), suggesting that, on average, each vertex is directly connected to about 10 other vertices, indicating a high level of connectivity. Smooth Distance-Decay Graph has a low average degree (0.776699), showing that each vertex is connected to less than one vertex on average, emphasizing its sparse nature. Contiguity-Based Graph has an average degree of 4.31068, indicating moderate connectivity among its vertices. Comparison Summary Connectivity: The Distance Threshold Graph is the most connected, followed by the Contiguity-Based Graph, with the Smooth Distance-Decay Graph being the least connected. Cohesion: Measured by average path length and graph density, the Distance Threshold Graph shows the highest cohesion, indicating that it's easier to traverse between any two vertices. The Smooth Distance-Decay Graph has the lowest cohesion, with the Contiguity-Based Graph in the middle. Application Suitability: The Distance Threshold Graph may be best suited for scenarios where the emphasis is on the close and dense connections between locations. The Smooth Distance-Decay Graph might be preferred for studies focusing on the impact of distance, favoring sparse connections that significantly weaken with distance. The Contiguity-Based Graph could be suitable for applications where geographical contiguity or proximity is a primary concern, such as regional planning. Each graph type, therefore, offers distinct advantages and drawbacks, depending on the specific requirements of your spatial analysis or modeling efforts.

Plot the matrix

We now plot the three spatial weight matrices. We see that the distance decay weight matrix is symmetric. The distance decay matrix is







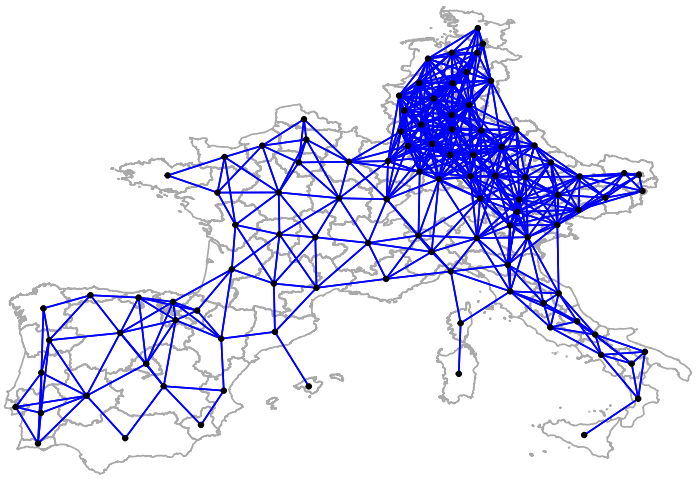
Try to visualize the network they represent

Let us first visualize the distance based spatial matrix. The first plot shows the map of Europe and the blue lines indicate the connections. The map shows the connectivity in Europe based on the distance threshold.

The second map visualizes the network based on the distance decay matrix. However, the edges do not display the intensity of connections, but just the connectivity to neighboring regions.

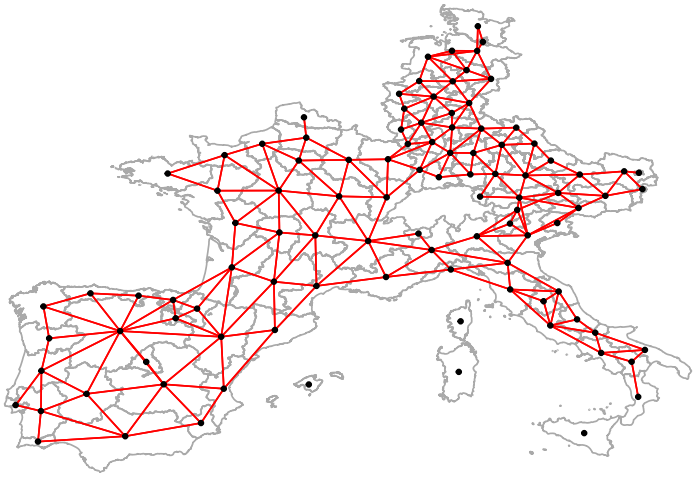
The last map displays the queen contiguity based measure. Interestingly, the islands in the middle sea are not being counted as neighbors. This should caution the use of this network,m as it seems implausible that Silicy for example is mainland italian provinces.

Distance Threshold



Distance Decay

Queen Contiguity



Test	Moran_I	Expectation	Variance	p_value
Distance Threshold	0.5938	-0.0098	0.0026	0
Smooth Distance-Decay	0.4768	-0.0102	0.0048	0
Contiguity-Based	0.5380	-0.0102	0.0045	0

Compute a suitable measure of spatial autocorrelation for productivity growth using these matrices. Point out differences, if there are any.

We calculate Global Moran's I as a measure of spatial autocorrelation for all three spatial weight matrices. All three matrices display the strong positive spatial autocorrelation between 0.62 - 0.54, which are all highly statistically significant with p-values < 0.01. Thus there is strong evidence for the presence of sizeable levels of spatial autocorrelation. This result is robust to the choice of the spatial weights matrix.

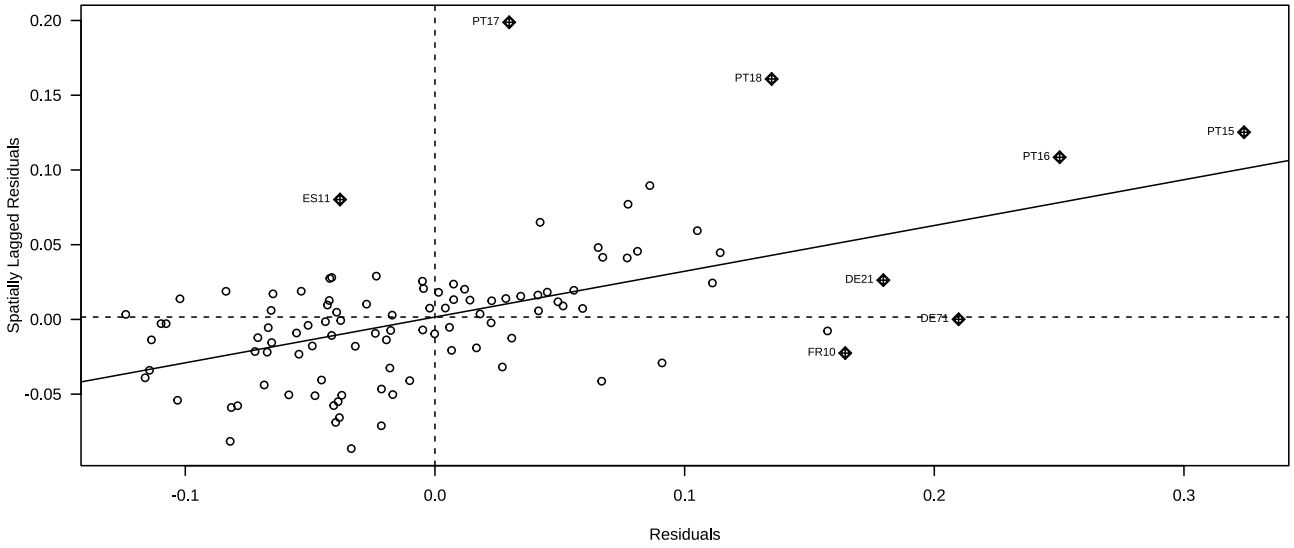
Estimate a linear regression model using OLS.

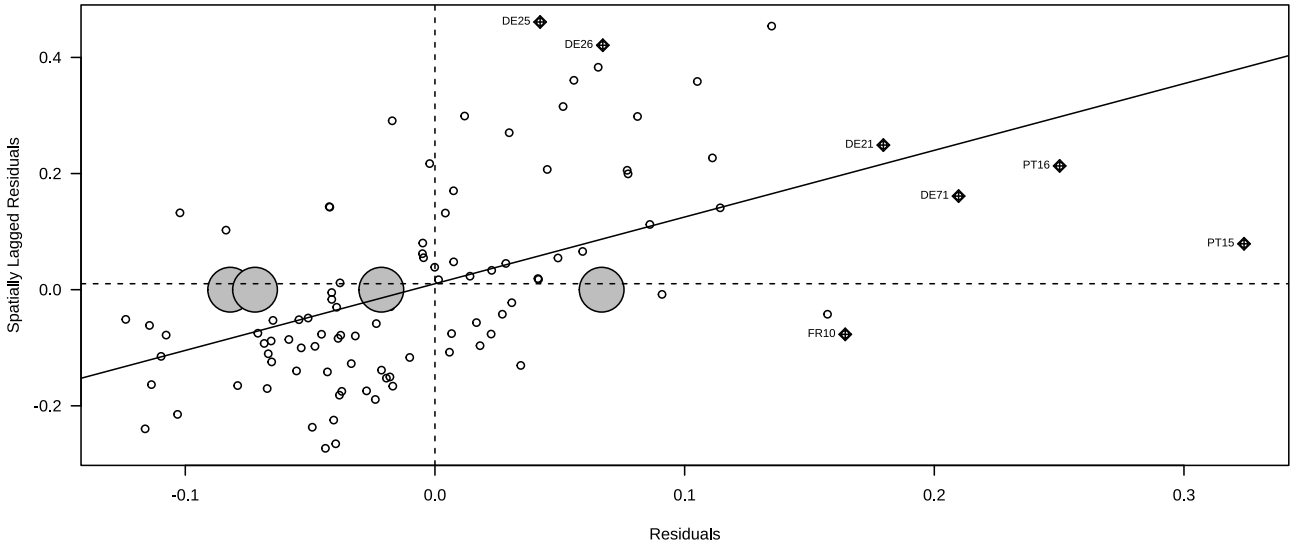
We estimate the specified model and obtain the following output.

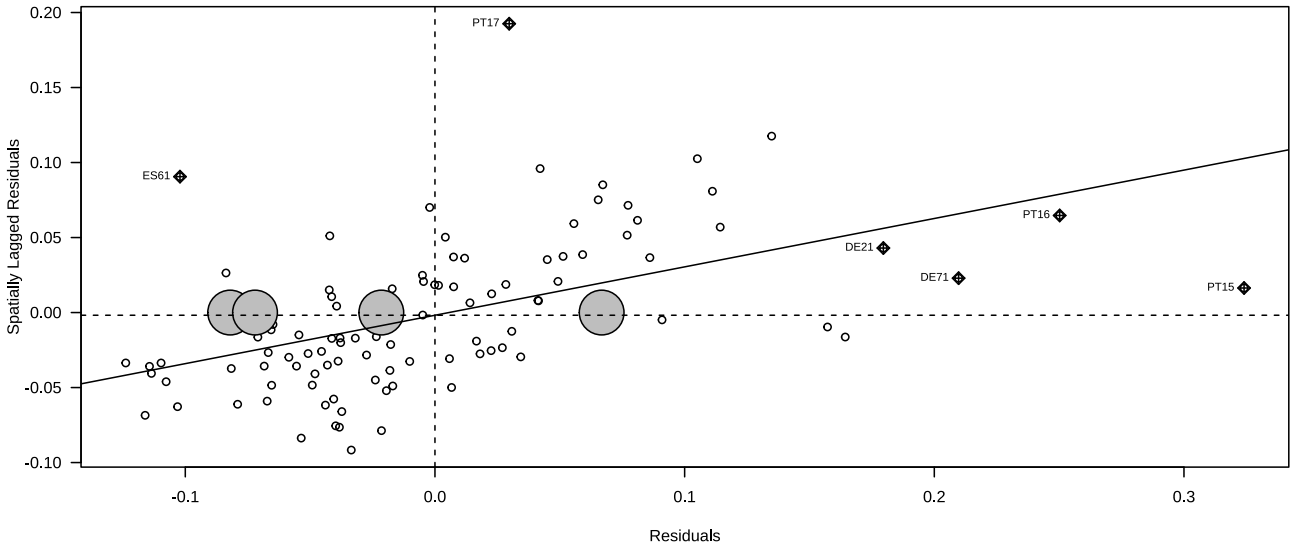
Table 1:

	Dependent variable:
	prod_growth
pr80b	-0.253^{***}
	(0.025)
lninv1b	0.032^{***}
	(0.008)
lndens.empb	0.007
	(0.009)
Constant	0.314^{***}
	(0.061)
Observations	103
R^2	0.528
Adjusted R^2	0.514
Residual Std. Error	0.080 (df = 99)
F Statistic	36.983^{***} (df = 3; 99)
Note:	*p<0.1; **p<0.05; ***p<0.01

We observe strong evidence for spatial autocorrelation. This holds for all spatial weights matrices used. The different weighting schemes highlight differnt countries. Thus the neglect fo this spatial dimension might give rise to bias in the OLS estimated coefficients.

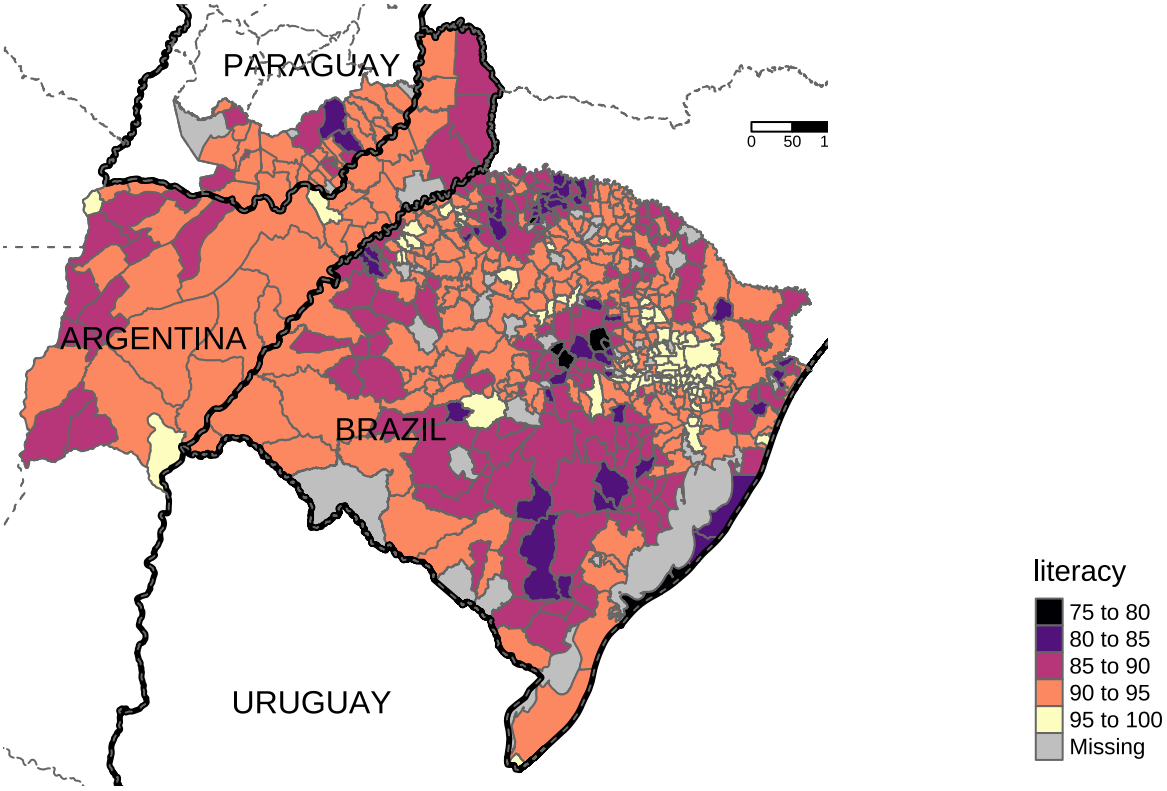

Spatially Lagged Residuals
Residuals
PT17
PT18
PT15
PT16
ES11
DE21
DE71
FR10


Spatially Lagged Residuals
Residuals
DE25
DE26
DE21
PT16
DE71
PT15
FR10


Spatially Lagged Residuals
Residuals
PT17
ES61
PT16
DE21
DE71
PT15

Exercise B

Creating maps



We now replicate table 2

```
summary(lm(illiteracy ~ distmiss + as.character(arg) + as.character(par) +
as.character(corr) + as.character(ita), data = literacy_Arg_Bra_Par))
```

```
##
## Call:
## lm(formula = illiteracy ~ distmiss + as.character(arg) + as.character(par) +
##     as.character(corr) + as.character(ita), data = literacy_Arg_Bra_Par)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.4934 -2.9187 -0.7474  2.5482 15.1243
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           9.830559   0.388595  25.298   <2e-16 ***
## distmiss             -0.003051   0.001565  -1.950   0.0517 .
## as.character(arg)1   -1.857182   1.019179  -1.822   0.0690 .
## as.character(par)1   -1.025812   1.317133  -0.779   0.4364
## as.character(corr)1   1.793435   1.265105   1.418   0.1569
## as.character(ita)1   -0.152454   1.461438  -0.104   0.9170
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.002 on 543 degrees of freedom
##   (29 observations deleted due to missingness)
## Multiple R-squared:  0.01168,    Adjusted R-squared:  0.002579
## F-statistic: 1.283 on 5 and 543 DF,  p-value: 0.2695
```