Spatial Economics – Assignment 2

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

April 2, 2024

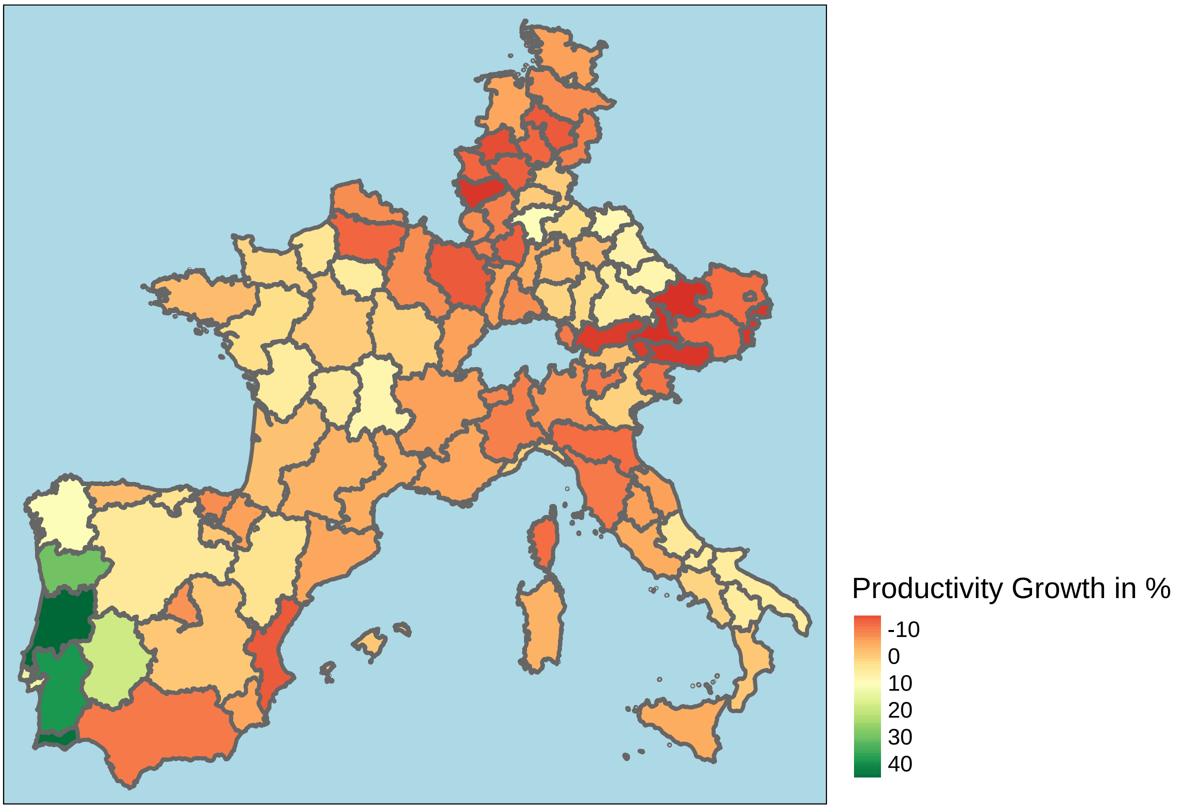
Contents

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 NUTS-2 regions for the selected countries. We can see that many regions especially in West Germany, Austria, and France exhibited negative productivity growth over the selected 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 begin with, while Portugal started from a rather low baseline productivity. Thus we can interpret this as productivity convergence 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 threshold spatial weights matrix based. Any region is being assigned a '1' with respect to another region, if it is less than 3 km away. We have chosen this threshold so that every region has a neighbor. We use the nb2mat function from the 'spdep' package. We row-normalize the matrix.

```
coords <- st_coordinates(st_centroid(EU27))

#checking the maximum distance as to include all observations which have a matrix
nb1 <- knn2nb(knearneigh(coords, k = 1))

dist1 <- nbdists(nb1, coords)

distw <- dnearneigh(coords, 0, 3)

#creating matrix based on distance threshold up to 3 kilometers
dist_w_matrix <- nb2mat(distw, style="W", zero.policy=TRUE)
```

Table 1: Summary of Distance Threshold Graph

Property	Value
Number of vertices	103
Number of edges	514
Average path length	0.6750662
Graph density	0.09784885
Average degree	9.980583
Max Eigenvector Centrality	1.00
Min Eigenvector Centrality	0.008694479
Average Eigenvector Centrality	0.1497447
Most Central Unit (Vertex ID)	49

Table 2: Summary of Smooth Distance-Decay Matrix

Property	Value
Number of vertices	103
Number of edges	1266
Average path length	0.49328
Graph density	0.2410051
Average degree	24.58252
Max Eigenvector Centrality	1
Min Eigenvector Centrality	0.001002582
Average Eigenvector Centrality	0.2836761
Most Central Unit (Vertex ID)	23

(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 $w_{i,j} = 1/d_{i,j}^{\lambda}$, where d denotes the distance between observation i and j, and λ is the distance decay parameter. By ease of convention we set $\lambda = 0$. We calculate the weights for each neighboring region based on the k=20 nearest neighbors. We do *not* row-normalize the matrix.

```
k1 <- knearneigh(coords, k=20)
k2 <- knn2nb(k1)

dists <- nbdists(k2, coords)

ids <- lapply(dists, function(d){1/d})

decay_weights_matrix_list <- nb2listw(k2, glist = ids, style = "B", zero.policy = TRUE)

decay_weights_matrix <- listw2mat(decay_weights_matrix_list)
```

(iii) Contiguity-based measure

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

```
# Create a contiguity-based spatial weights matrix
queen_weights <- poly2nb(EU27, queen = TRUE)

contig_w_matrix <- nb2mat(queen_weights, style="W", zero.policy=TRUE)
```

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

We can gain deeper insights into these spatial weights matrices as well as the networks they represent by comparing key measures of the graphs that are derived from them.

We compare the matrices based on a set of characteristics. We compare the row-normalized matrices for the

Table 3: Summary of Contiguity-Based Graph

Property	Value
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	0
Average Eigenvector Centrality	0.08975192
Most Central Unit (Vertex ID)	48

queen contiguity and distance threshold matrix. Note that normalization procedure does not preserve the structure of the network.

Number of edges

The Smooth Distance-Decay Graph has the most edges (1266). The Distance Threshold Graph has fewer edges (514) than the Smooth Distance-Decay Graph, implying stricter criteria for edge creation based on a fixed distance threshold. The Contiguity-Based Graph has the fewest edges (222), since only directly contiguous or neighboring entities are connected, leading to a more sparse graph structure.

Average path length

The Contiguity-Based Graph has the highest average path length (1.2888), reflecting the sparse connectivity where nodes are less directly connected. The Distance Threshold Graph has a medium average path length (0.6751). The Smooth Distance-Decay Graph has the lowest average path length (0.4933), indicative of a denser network where nodes are more directly accessible to one another.

Graph density

Consistent with the number of edges, the Smooth Distance-Decay Graph is the densest (0.2410), followed by the Distance Threshold Graph (0.0978), with the Contiguity-Based Graph being the least dense (0.0423).

Average degree

Again, the Smooth Distance-Decay Graph shows the highest average degree (24.5825), the Distance Threshold Graph shows a medium degree (9.9806), and the Contiguity-Based Graph has the lowest (4.3107).

Minimum Eigenvector Centrality

The Contiguity-Based Graph shows the most significant variation in centrality (minimum near zero), reflecting a few very poorly connected nodes, or nodes that only connect to other low-influence nodes. The Smooth Distance-Decay Graph and the Distance Threshold Graph have higher minimum values, indicating a more uniform distribution of node influence.

Average Eigenvector Centrality

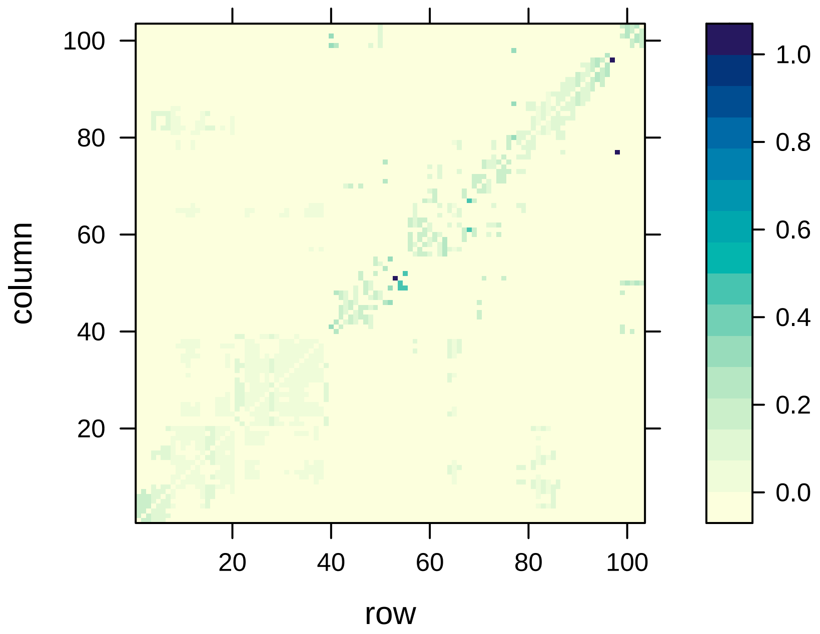
Higher on average in the Smooth Distance-Decay Graph (0.2837), suggesting that, on average, nodes are better positioned or more influential within the network. It's lowest in the Contiguity-Based Graph (0.0898), consistent with its sparse and uneven connectivity.

Looking at the most central unit through eigenvector centrality shows that different nodes are identified as most central in each graph, reflecting the impact of the underlying connection logic on the perceived importance or centrality of nodes.

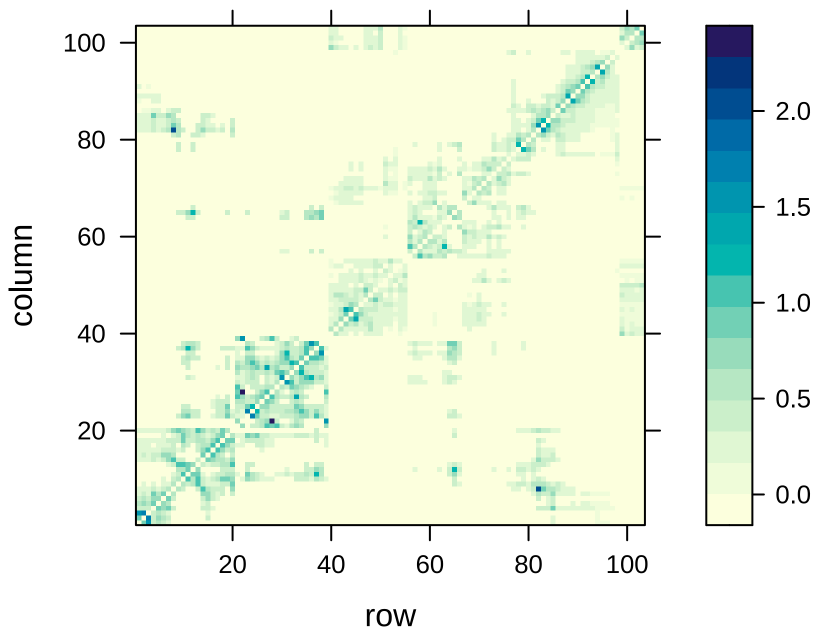
Plot the matrix

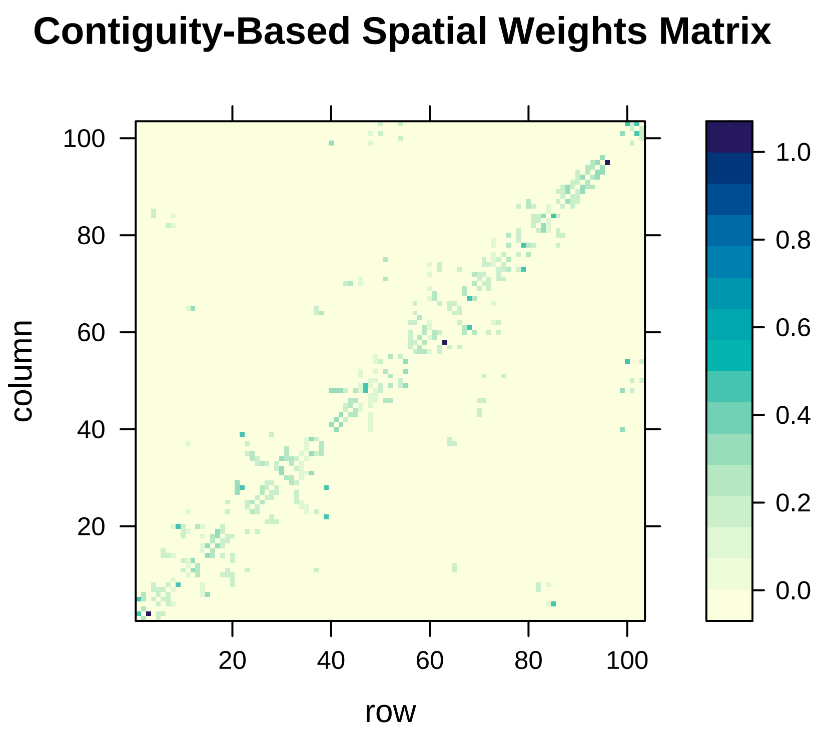
We now plot the three spatial weight matrices. We see that the distance decay weight matrix is symmetric. The distance decay matrix is

Distance Threshold Spatial Weights Matrix



Smooth Distance-Decay Spatial Weights Matrix





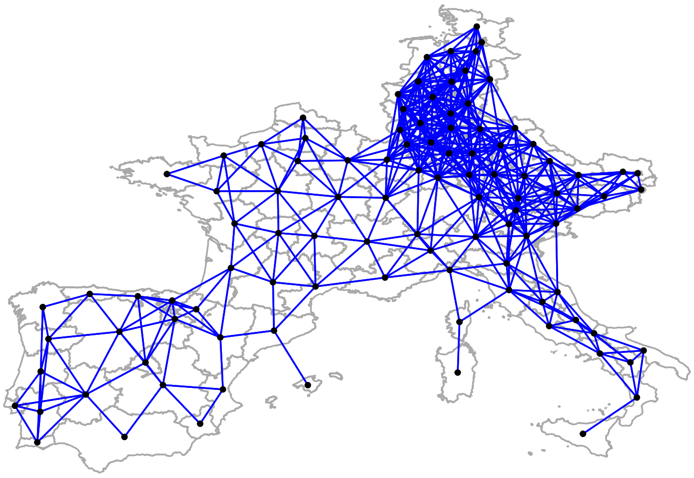
Try to visualize the network they represent

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.

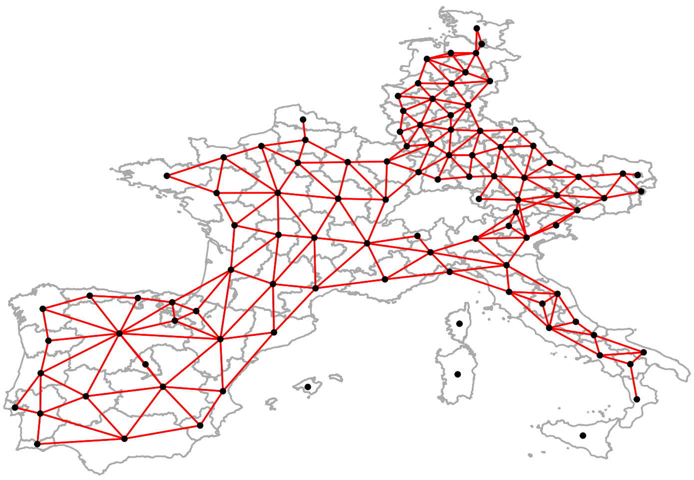
The last map displays the queen contiguity based measure. The islands in the middle sea are not being counted as neighbors. This should caution the use of this network, as it seems implausible that Siciliy for example is not connected to the mainland Italian provinces.

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	0.2925	-0.0098	0.0010	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.

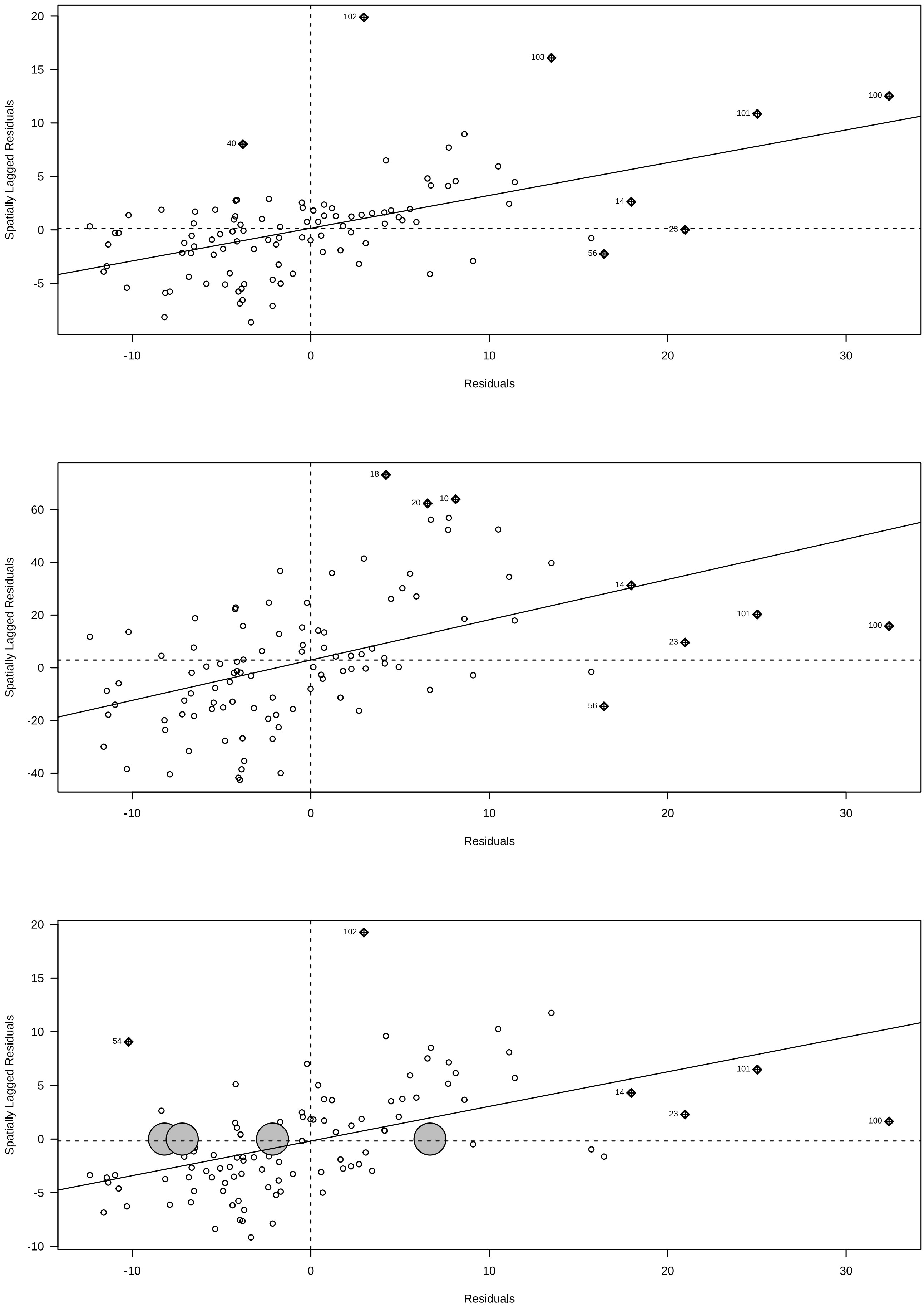
Estimate a linear regression model using OLS.

We estimate the specified model and obtain the following output.

Table 4:

	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:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

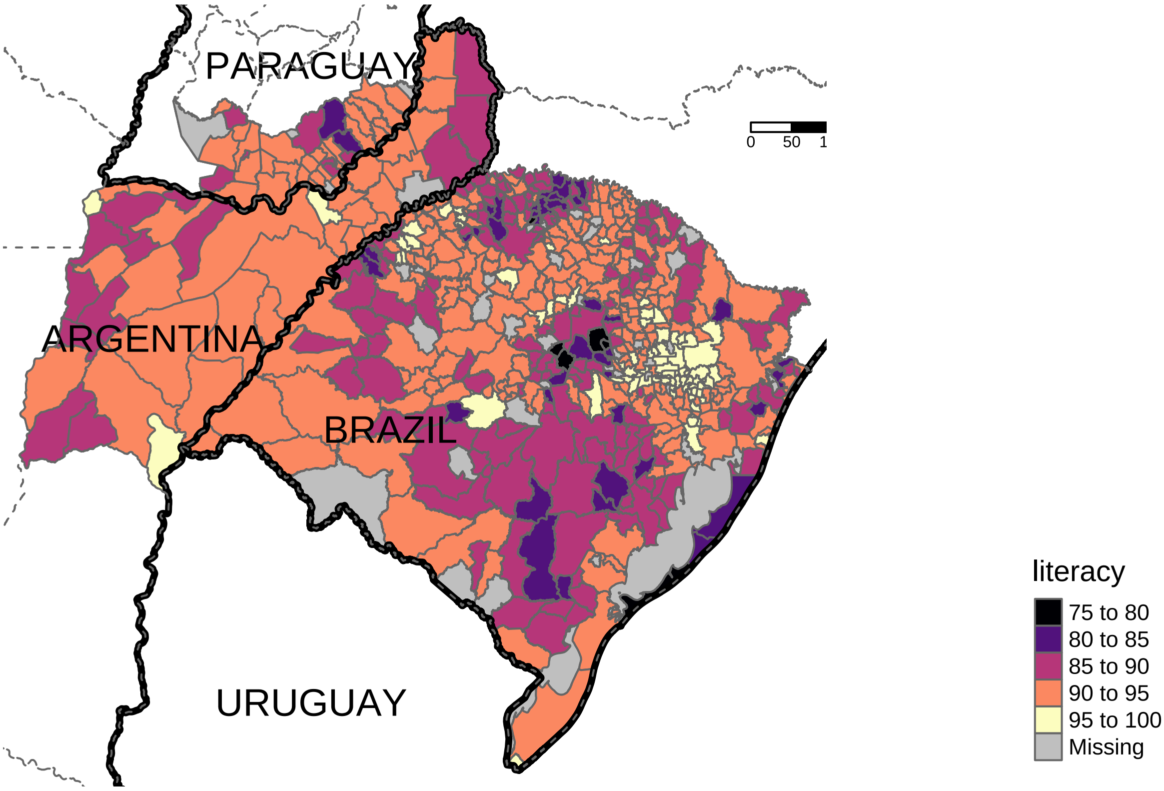
We observe strong evidence for spatial autocorrelation. This holds for all spatial weights matrices used. The different weighting schemes highlight differnt countries. Thus the neglect fo this spatial dimension might give rise to bias in the OLS estimated coefficients.



Spatially Lagged Residuals
Residuals
102
103
100
101
40
14
23
56
Spatially Lagged Residuals
Residuals
18
20
10
14
101
100
23
56
Spatially Lagged Residuals
Residuals
102
54
101
14
23
100

Exercise B

Creating maps



We now replicate table 2

```
list_2 <- dat2 %>%
  select(COUNTRY, geometry, NAME_1, NAME_2) %>%
  mutate(country = ifelse(COUNTRY == "Brazil", "BRA", COUNTRY)) %>%
  rename("muni" = "NAME_2") %>%
  mutate(state = ifelse(NAME_1 == "Rio Grande do Sul", "RS", NAME_1))

literacy_Arg_Bra_Par_2 <- literacy_Arg_Bra_Par %>%
  left_join(list_2, by = c("muni",  "state", "country"))

mod1 <- lm(illiteracy ~ (lati) + (longi) + distmiss + state, data = literacy_Arg_Bra_Par_2)

mod2 <- lm(illiteracy ~ (lati) + (longi) + distmiss + state + coast + river + slope + rugg +
alti + tempe + preci + area, data = literacy_Arg_Bra_Par_2)


bra <- literacy_Arg_Bra_Par_2 %>%
             filter(country == "BRA")

mod3 <- lm(illiteracy ~ (lati) + (longi) + distmiss + mesorregi, data = bra)

mod4 <- lm(illiteracy ~ (lati) + (longi) + distmiss + mesorregi + coast + river + slope +
rugg + alti + tempe + preci + area, data = bra)


arg <- literacy_Arg_Bra_Par_2 %>%
             filter(country == "Argentina")

mod5 <- lm(illiteracy ~ (lati) + (longi) + distmiss, data = arg)

mod6 <- lm(illiteracy ~ (lati) + (longi) + distmiss + coast + river + slope + rugg + alti +
tempe + preci + area, data = arg)
```

```r
par <- literacy_Arg_Bra_Par_2 %>%
            filter(country == "Paraguay")

mod7 <- lm(illiteracy ~ (lati) + (longi) + distmiss + state, data = par)

mod8 <- lm(illiteracy ~ (lati) + (longi) + distmiss + state + coast + river + slope + rugg +
alti + tempe + preci + area, data = par)

stargazer(mod1, mod2, mod3, mod4, mod5, mod6, mod7, mod8)
```

```
##
## % Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac
## % Date and time: Sat, Apr 13, 2024 - 14:05:31
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lcccccccc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  & \multicolumn{8}{c}{\textit{Dependent variable:}} \\
## \cline{2-9}
## \\[-1.8ex] & \multicolumn{8}{c}{illiteracy} \\
## \\[-1.8ex] & (1) & (2) & (3) & (4) & (5) & (6) & (7) & (8)\\
## \hline \\[-1.8ex]
##  lati & 0.556$^{**}$ & 0.072 & 0.359 & 3.206$^{**}$ & 0.043 & $-$6.724$^{**}$ & 3.540$^{**}$ & $-$7.
##   & (0.251) & (0.782) & (0.403) & (1.473) & (0.643) & (2.599) & (1.622) & (4.381) \\
##   & & & & & & & & \\
##  longi & $-$1.108$^{***}$ & $-$1.007 & $-$1.716$^{***}$ & $-$5.054$^{***}$ & 0.054 & 6.704$^{***}$ &
##   & (0.269) & (0.687) & (0.369) & (1.521) & (0.496) & (1.934) & (1.017) & (6.364) \\
##   & & & & & & & & \\
##  distmiss & 0.011$^{***}$ & 0.011$^{**}$ & 0.020$^{***}$ & 0.031$^{***}$ & 0.013 & 0.055$^{***}$ & $
##   & (0.004) & (0.005) & (0.006) & (0.008) & (0.008) & (0.019) & (0.023) & (0.029) \\
##   & & & & & & & & \\
##  stateItapúa & 2.154 & 3.619$^{**}$ &  &  &  &  &  &  \\
##   & (1.350) & (1.693) &  &  &  &  &  &  \\
##   & & & & & & & & \\
##  stateMisiones & 1.017 & 1.297 &  &  &  &  & 0.200 & $-$2.053 \\
##   & (1.572) & (1.865) &  &  &  &  & (1.481) & (1.661) \\
##   & & & & & & & & \\
##  stateMisiones1 & 2.061 & 3.733$^{**}$ &  &  &  &  &  &  \\
##   & (1.565) & (1.881) &  &  &  &  &  &  \\
##   & & & & & & & & \\
##  stateRS & 5.341$^{***}$ & 6.027$^{***}$ &  &  &  &  &  &  \\
##   & (1.521) & (1.798) &  &  &  &  &  &  \\
##   & & & & & & & & \\
##  coast &  & 0.209 &  & $-$3.864$^{**}$ &  & $-$0.764 &  & 21.631$^{***}$ \\
##   &  & (0.989) &  & (1.848) &  & (2.857) &  & (7.366) \\
##   & & & & & & & & \\
##  river &  & 1.465$^{**}$ &  & 1.645$^{**}$ &  & 10.574$^{***}$ &  & $-$4.913 \\
##   &  & (0.741) &  & (0.802) &  & (2.672) &  & (4.569) \\
##   & & & & & & & & \\
##  slope &  & $-$0.00001 &  & 0.00002 &  & $-$0.053$^{*}$ &  & $-$0.071$^{***}$ \\
##   &  & (0.0002) &  & (0.0002) &  & (0.028) &  & (0.021) \\
##   & & & & & & & & \\
##  rugg &  & $-$0.00000 &  & $-$0.00000 &  & 0.001$^{*}$ &  & 0.002$^{***}$ \\
##   &  & (0.00000) &  & (0.00000) &  & (0.001) &  & (0.001) \\
##   & & & & & & & & \\
##  alti &  & 0.006 &  & 0.005 &  & 0.062$^{***}$ &  & 0.038$^{**}$ \\
```

```
##   &  & (0.004) &  & (0.005) &  & (0.011) &  & (0.014) \\
##   & & & & & & & & \\
##  tempe &  & 0.058 &  & 0.057 &  & 0.915$^{***}$ &  & 0.842$^{***}$ \\
##   &  & (0.079) &  & (0.099) &  & (0.208) &  & (0.204) \\
##   & & & & & & & & \\
##  preci &  & $-$0.003 &  & $-$0.002 &  & $-$0.019$^{**}$ &  & $-$0.012$^{**}$ \\
##   &  & (0.002) &  & (0.002) &  & (0.007) &  & (0.005) \\
##   & & & & & & & & \\
##  area &  & 0.0001 &  & $-$0.0004 &  & $-$0.0003 &  & 0.002$^{***}$ \\
##   &  & (0.0002) &  & (0.0003) &  & (0.0002) &  & (0.001) \\
##   & & & & & & & & \\
##  mesorregi &  &  & $-$0.418$^{*}$ & $-$0.216 &  &  &  &  \\
##   &  &  & (0.216) & (0.261) &  &  &  &  \\
##   & & & & & & & & \\
##  Constant & $-$40.669$^{***}$ & $-$59.567 & 1,724.668$^{*}$ & 755.296 & 11.554 & 31.195 & 121.370$^{
##   & (13.191) & (36.199) & (919.888) & (1,111.741) & (20.002) & (72.708) & (60.387) & (208.813) \\
##   & & & & & & & & \\
## \hline \\[-1.8ex]
## Observations & 549 & 548 & 467 & 467 & 42 & 42 & 40 & 39 \\
## R$^{2}$ & 0.042 & 0.073 & 0.056 & 0.095 & 0.099 & 0.695 & 0.145 & 0.652 \\
## Adjusted R$^{2}$ & 0.029 & 0.047 & 0.048 & 0.071 & 0.028 & 0.583 & 0.047 & 0.491 \\
## Residual Std. Error & 3.948 (df = 541) & 3.916 (df = 532) & 4.101 (df = 462) & 4.050 (df = 454) & 2.
## F Statistic & 3.370$^{***}$ (df = 7; 541) & 2.793$^{***}$ (df = 15; 532) & 6.880$^{***}$ (df = 4; 46
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{8}{r}{$^{*}$p$<$0.1; $^{**}$p$<$0.05; $^{***}$p$<$0.01} \\
## \end{tabular}
## \end{table}
```

Exercise C

Table 5:

	Dependent var			
	illiteracy			
	(1)	(2)	(3)	(4)
lati	0.556**	0.072	0.359	3.206**
	(0.251)	(0.782)	(0.403)	(1.473)
longi	−1.108***	−1.007	−1.716***	−5.054***
	(0.269)	(0.687)	(0.369)	(1.521)
distmiss	0.011***	0.011**	0.020***	0.031***
	(0.004)	(0.005)	(0.006)	(0.008)
stateItapúa	2.154	3.619**		
	(1.350)	(1.693)		
stateMisiones	1.017	1.297		
	(1.572)	(1.865)		
stateMisiones1	2.061	3.733**		
	(1.565)	(1.881)		
stateRS	5.341***	6.027***		
	(1.521)	(1.798)		
coast		0.209		−3.864**
		(0.989)		(1.848)
river		1.465**		1.645**
		(0.741)		(0.802)
slope		−0.00001		0.00002
		(0.0002)		(0.0002)
rugg		−0.00000		−0.00000
		(0.00000)		(0.00000)
alti		0.006		0.005
		(0.004)		(0.005)
tempe		0.058		0.057
		(0.079)		(0.099)
preci		−0.003		−0.002
		(0.002)		(0.002)
area		0.0001		−0.0004
		(0.0002)		(0.0003)
mesorregi			−0.418*	−0.216
			(0.216)	(0.261)
Constant	−40.669***	−59.567	1,724.668*	755.296
	(13.191)	(36.199)	(919.888)	(1,111.741)
Observations	549	548	467	467
R^2	0.042	0.073	0.056	0.095
Adjusted R^2	0.029	0.047	0.048	0.071
Residual Std. Error	3.948 (df = 541)	3.916 (df = 532)	4.101 (df = 462)	4.050 (df = 454)
F Statistic	3.370*** (df = 7; 541)	2.793*** (df = 15; 532)	6.880*** (df = 4; 462)	3.976*** (df = 12; 454)

Note: