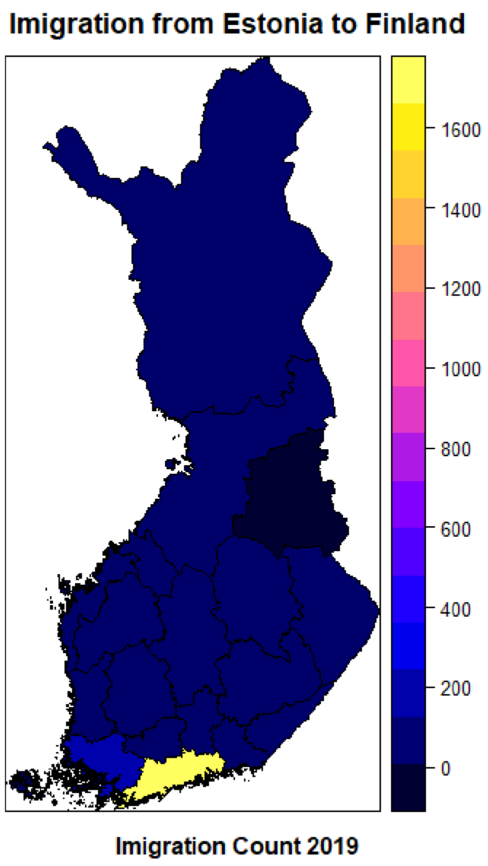
GWR FINLAND IMMIGRATION FROM ESTONIA

1. Opening the Shapefile and showing the Dependent Variable



2. Apply OLS, to understand how much the variables explain the variability of Immigration Rate in Finland.

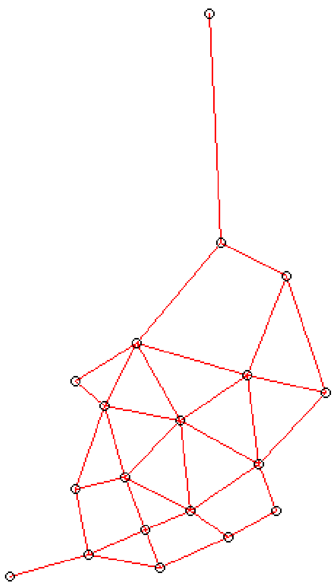
```
Call:
lm(formula = im_2019 ~ unem_.19 + gdp_eur + proxkm + pop.65 +
    X.male, data = finland)

Residuals:
   Min     1Q Median     3Q    Max
-302.9 -146.3  -25.0  154.5  434.4

Coefficients:
               Estimate  Std. Error t value Pr(>|t|)
(Intercept) -1317.13031   979.30184  -1.345  0.20163
unem_.19       45.65502    34.22833   1.334  0.20516
gdp_eur         0.04602     0.01384   3.326  0.00547 **
proxkm         -0.21568     0.13706  -1.574  0.13961
pop.65        -26.49806    23.74297  -1.116  0.28461
X.male          3.03866     4.85307   0.626  0.54207
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 229.2 on 13 degrees of freedom
Multiple R-squared:  0.7296,    Adjusted R-squared:  0.6256
F-statistic: 7.014 on 5 and 13 DF,  p-value: 0.00221
```

RSquared explains 27.26% of the variation in Immigration Rate.

Now with counts 2019, RSquared explains 72.96% of the variation in Immigrants count 2019

The P-value shows that it is not statistically significant so we keep finding for spatial dependence.

3. Create the Weights File, in GeoDa



Gabriel Graph

4. Apply Lagrange Model for testing spatial dependence

```
        Lagrange multiplier diagnostics for spatial dependence

data:
model: lm(formula = im_2019 ~ unem_.19 + gdp_eur + proxkm + pop.65 + X.male, data = finland)
weights: weights

LMerr = 0.83741, df = 1, p-value = 0.3601
```

It is insignificant, but we keep working. LMError explains 83% of variation.

5. Apply Spatial Error Model

It gives Nagel Kerke pseudo-P-value 77.98% that explains the variability in Immigration. It increased due to a new variable Lag Error which gives more explanatory power.

```
Residuals:
     Min       1Q   Median       3Q      Max
-331.627  -79.756   10.546   84.239  290.700

Type: error
Coefficients: (asymptotic standard errors)
                Estimate   Std. Error z value   Pr(>|z|)
(Intercept) -149.9307651  677.9443605 -0.2212  0.8249718
unem_.19      85.7981609   21.6055431  3.9711 0.00007154 ***
gdp_eur        0.0292771    0.0097058  3.0164  0.0025577 **
proxkm        -0.3656471    0.0882345 -4.1440 0.00003412 ***
pop.65       -67.3806041   19.6544206 -3.4283  0.0006074 ***
X.male         7.8838485    3.7458946  2.1047  0.0353206 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Lambda: -0.8824, LR test value: 3.9119, p-value: 0.047947
Approximate (numerical Hessian) standard error: 0.26186
    z-value: -3.3697, p-value: 0.00075249
Wald statistic: 11.355, p-value: 0.00075249

Log likelihood: -124.6574 for error model
ML residual variance (sigma squared): 22663, (sigma: 150.54)
Nagelkerke pseudo-R-squared: 0.77989
Number of observations: 19
Number of parameters estimated: 8
AIC: 265.31, (AIC for lm: 267.23)
```

6. Apply Durbin Model, to find average effects, interaction effects + Common Factor Hypothesis

```
           Likelihood ratio for spatial linear models

data:
Likelihood ratio = 17.622, df = 5, p-value = 0.00346
sample estimates:
Log likelihood of x Log likelihood of y
          -115.8465           -124.6574
```

If it significant it means our variables are good for explanation. P-value 0.003

The Durbin model, adds Lag variables which help to give more explanatory power.

```
Type: mixed
Coefficients: (numerical Hessian approximate standard errors)
                Estimate   Std. Error z value      Pr(>|z|)
(Intercept)  891.6938665 1217.1861540  0.7326      0.463811
unem_.19      64.1287488   23.0336997  2.7841      0.005367 **
gdp_eur        0.0422322    0.0104769  4.0310 0.00005554738 ***
proxkm        -1.1702625    0.2172535 -5.3866 0.00000007179 ***
pop.65       -52.5552575   22.3633333 -2.3501      0.018770 *
X.male         7.5252708    3.4991365  2.1506      0.031507 *
lag.unem_.19 166.0468367   40.8870003  4.0611 0.00004883877 ***
lag.gdp_eur   -0.0046024    0.0215140 -0.2139      0.830606
lag.proxkm     0.5572931    0.3508625  1.5884      0.112207
lag.pop.65  -145.0562617   30.8944912 -4.6952 0.00000266327 ***
lag.X.male    18.9315480    7.0973630  2.6674      0.007644 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Rho: -0.60097, LR test value: 2.9864, p-value: 0.083964
Approximate (numerical Hessian) standard error: 0.30711
    z-value: -1.9568, p-value: 0.050367
Wald statistic: 3.8292, p-value: 0.050367

Log likelihood: -115.8465 for mixed model
ML residual variance (sigma squared): 10404, (sigma: 102)
Nagelkerke pseudo-R-squared: 0.91293
```

The variables that not helps to the model are Lag Unemployment Rate, and Lag GDP. Anyways, the model explanatory power increased to 91.29%.

7. Now, we apply effects. Direct, Indirect, and total impacts

Impact measures (mixed, exact):

	Direct	Indirect	Total
unem_.19	39.87015132	103.90243136	143.77258268
gdp_eur	0.04793653	-0.02443211	0.02350442
proxkm	-1.41009205	1.02721827	-0.38287378
pop.65	-30.96690214	-92.46546978	-123.43237193
X.male	4.78390629	11.74158732	16.52549361

Simulated p-values:

	Direct	Indirect	Total
unem_.19	0.186088	0.00851646	0.00000046134
gdp_eur	0.000101	0.07922062	0.0503092
proxkm	0.000032269	0.01252687	0.0083312
pop.65	0.271133	0.00024238	0.00005690437
X.male	0.247685	0.02324785	0.0107115

- *Unemployment is Significant, Indirect and Total*
  Increases of Unemployment cause increase of Immigration to Helsinki. So, regions that changes cause Effects in the surrounding regions.

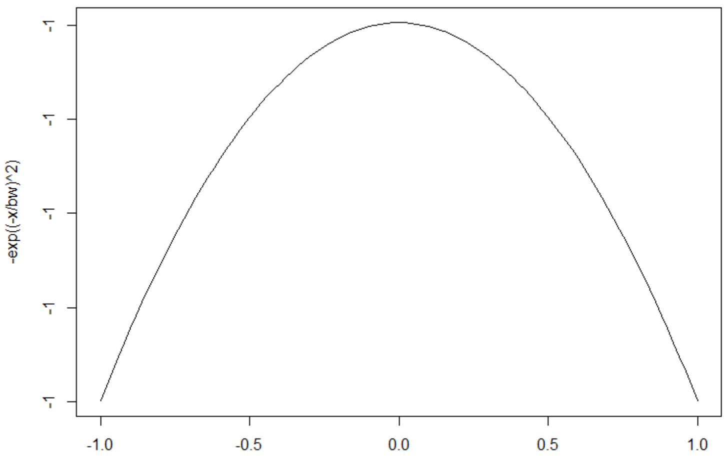
- *GDP is Significant, in Direct.*
  Small changes in GDP causes changes of Immigration, but not relevant.

- *Proximity all Significant. More in Direct.*
  The more proximity (density) less Immigration.

- *Pop over 65 Significant, Indirect and Total.*
  Increases in elderly decreases the Immigration to Finland

- *Male pop Significant in Total.*
  Increases in Males population increases in Immigration.

8. Get the Bandwidth



9. GWR

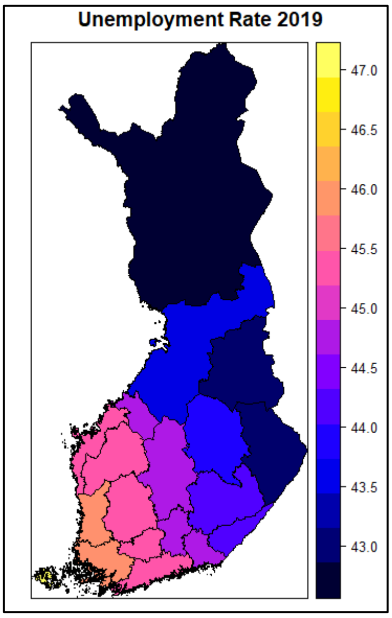
What happen with Migration if distance to Helsinki changes?

Changes in migration in one location will change in another location because of the way the effect travels back and forth. Spatial connectivity causes heterogeneity.

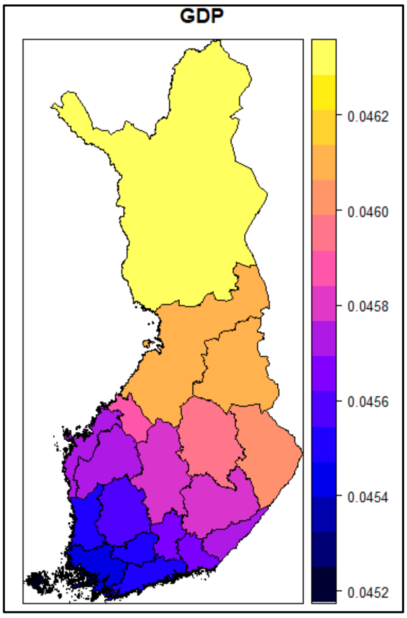
OLS

```
gwr(formula = im_2019 ~ unem_.19 + gdp_eur + proxkm + pop.65 +
    X.male, data = finland, bandwidth = bw, hatmatrix = TRUE)
Kernel function: gwr.Gauss
Fixed bandwidth: 2231982
Summary of GWR coefficient estimates at data points:
                    Min.      1st Qu.       Median      3rd Qu.         Max.     Global
X.Intercept. -1388.045475 -1308.946056 -1287.694022 -1270.534555 -1249.029773 -1317.1303
unem_.19        42.849475    43.919821    44.707462    45.392117    46.934043    45.6550
gdp_eur          0.045247     0.045572     0.045718     0.045894     0.046285     0.0460
proxkm          -0.237887    -0.228633    -0.222577    -0.217043    -0.197288    -0.2157
pop.65         -28.237652   -27.528727   -26.932848   -26.173595   -23.092576   -26.4981
X.male           2.830048     3.017137     3.078543     3.123382     3.150819     3.0387
Number of data points: 19
Effective number of parameters (residual: 2traceS - traceS'S): 6.418765
Effective degrees of freedom (residual: 2traceS - traceS'S): 12.58124
Sigma (residual: 2traceS - traceS'S): 229.1208
Effective number of parameters (model: traceS): 6.218124
Effective degrees of freedom (model: traceS): 12.78188
Sigma (model: traceS): 227.3154
Sigma (ML): 186.4444
AICc (GWR p. 61, eq 2.33; p. 96, eq. 4.21): 278.0285
AIC (GWR p. 96, eq. 4.22): 258.8069
Residual sum of squares: 660469.1
Quasi-global R2: 0.738499
```

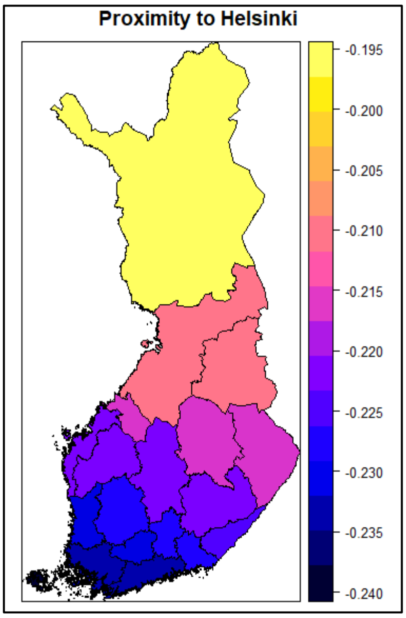
Global is the regression coefficient. But we study the local variation/heterogeneity of the changes by regions.



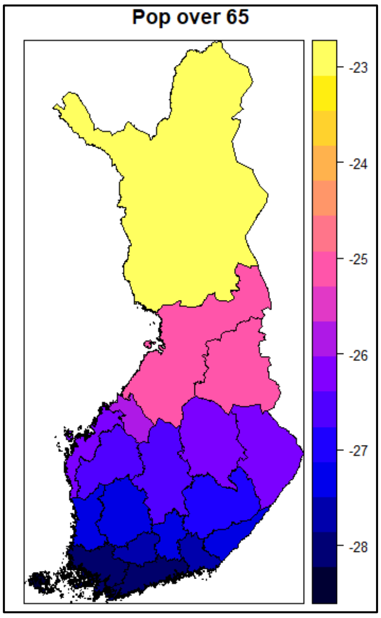
The changes in Unemployment affects more the immigration in south-west of Finland. Maybe it is worse in Estonia.



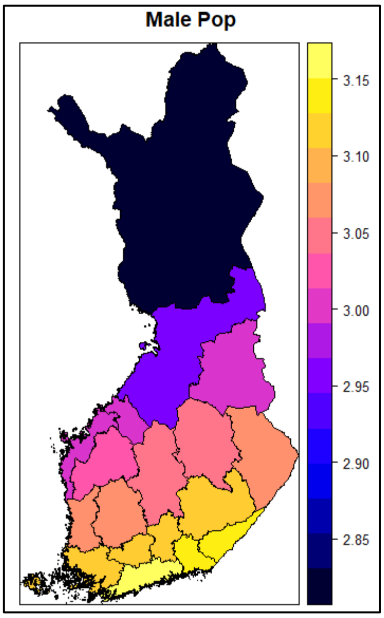
The changes in GDP are not relevant in changes of Immigration. But most likely to be affected in the north of Finland.



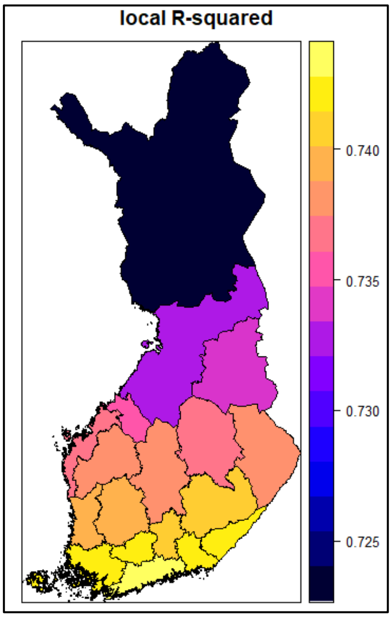
More density/proximity increases the Immigration more likely at the south Finland.



Increase in pop over 65 causes less Immigration at the South of Finland. People normally Migrate to cities more livable with more youngers.

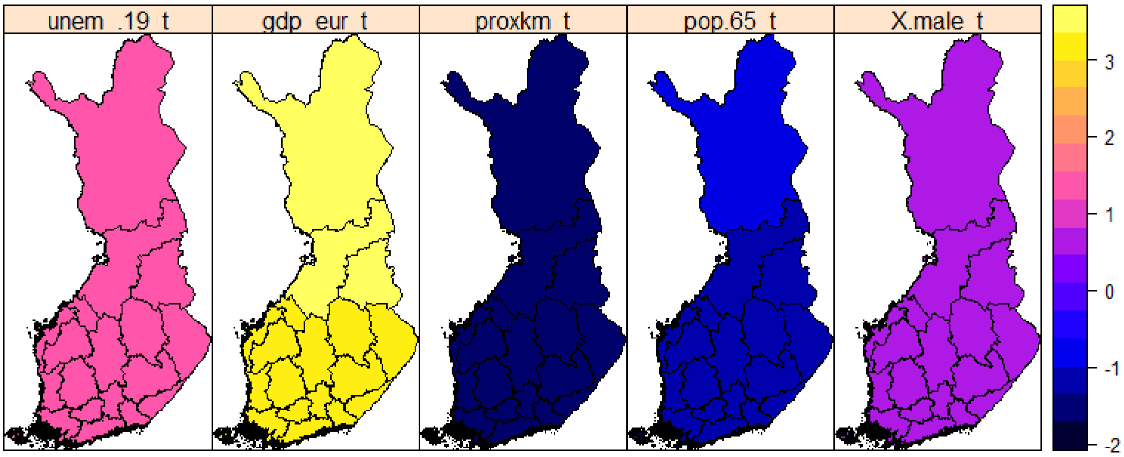


Increases in Male population increases Immigration at South Finland.



Quality of the regression. The set of explanatory variables doesn't explain Immigration the best in Northern Finland. The Southern Finland has the best explanatory power based on our variables to explain Immigration. Over 74%.

10. T-Values. Higher t values mean higher statistical significance.



Values closer to 0 means no explanatory power.

Better variables to explain Immigration in Finland are GDP, then Unemployment, then Proximity.

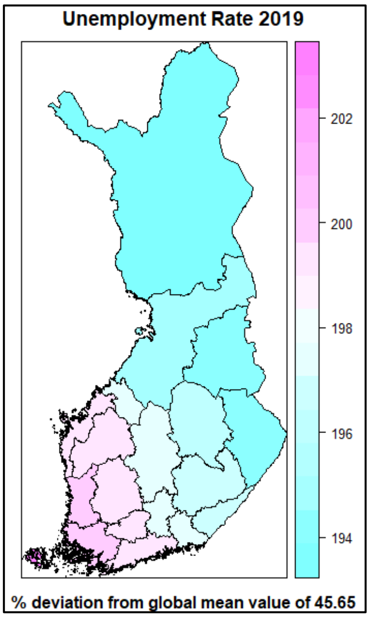
On the other hand, Population over 65 and Male population are bad predictors.

P-values.

- T-value 3 P-value 0.0105 Significant 99%
- T-value 2 P-value 0.0675 Significant 95%
- T-value 1 P-value 0.336
- T-value 0 P-value 1
- T-value -1 P-value 0.336
- T-value -2 P-value 0.0675 Significant 95%

11. Non-stationary effect.

Anomaly of effect of the Variable in Immigration, at particular location.



It reveals Niches of changes. Likes South-West. And North-East.

2 regions where the effect of Unemployment affect the Immigration sufficiently different. These 2 regions work by different mechanisms of Unemployment which explains Immigration.

With other variables, it will give the same regionalization.