**Analysis author:** *Kayla Aburida*
**Title: Lab IV: Laboratory 3: The Bivariate and Empirical Bayesian rates local Moran's I, k means partitioning**
**Dataset source:** *Brenden Hurley, GW Geography*
**Dataset location:** *Nepal*
**Date(s) of analysis:** *10/18/2023- 10/25-2023*
**Dataset time span: 2012- 2013**
**Dataset Scale:** *Districts in Nepal*

**Analysis author:** *Kayla Aburida*
**Title: Lab IV: Laboratory 3: The Bivariate and Empirical Bayesian rates local Moran's I, k means partitioning**
**Dataset source:** *Brenden Hurley, GW Geography*
**Dataset location:** *France*
**Date(s) of analysis:** *10/18/2023- 10/25-2023*
**Dataset time span: 1830s**
**Dataset Scale:** *Counties in France*

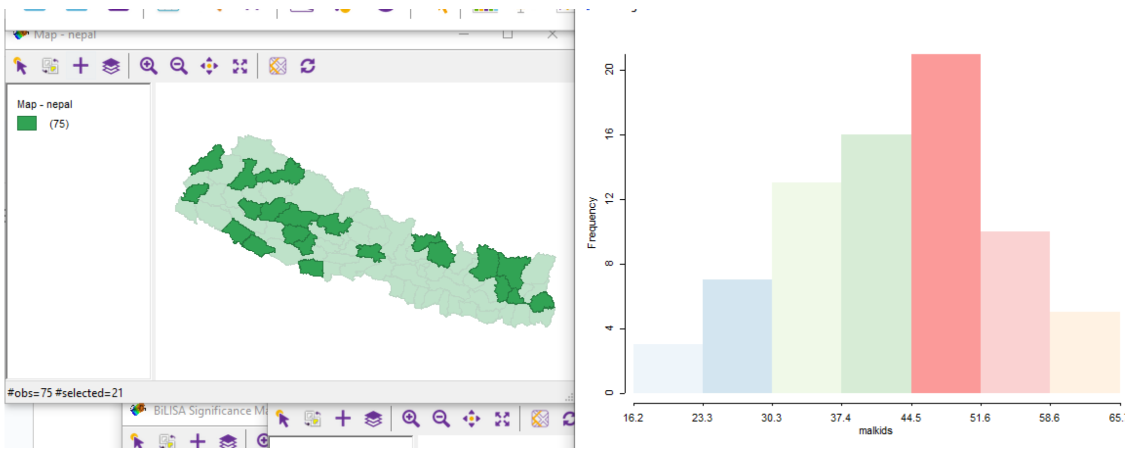
**Analysis author:** *Kayla Aburida*
**Title: Lab IV: Laboratory 3: The Bivariate and Empirical Bayesian rates local Moran's I, k means partitioning**
**Dataset source:** *Brenden Hurley, GW Geography*
**Dataset location:** *Texas*

**Date(s) of analysis:** ***10/18/2023- 10/25-2023***
**Dataset time span: 2020**
**Dataset Scale:** ***Counties in Texas***

**Introduction**

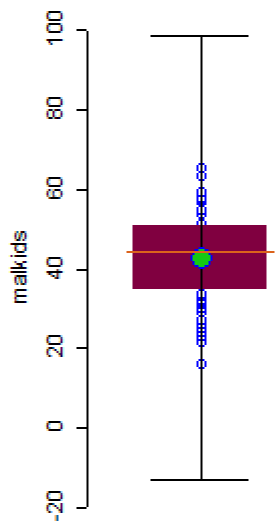
*Histogram*

A histogram is a diagram consisting of rectangles whose area is proportional to the frequency of a variable and whose width is equal to the class interval. This histogram represents the percentage of kids within Nepal who are malnourished. This histogram represents a roughly normal distribution

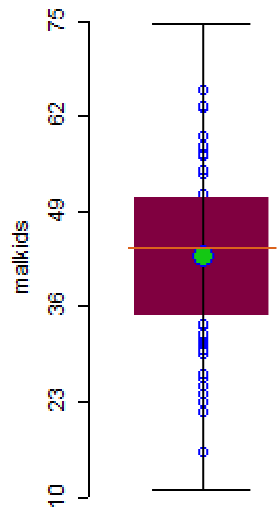



Explanation of what a boxplot is including what a "hinge" is, and two screenshots.

A box plot is a method for graphically demonstrating numerical data's locality, spread and skewness groups through their quartiles. The hinge is the box that stretches the lower hinge (the 25th percentile) to the upper hinge (the 75th quartile), therefore containing the middle portion of data in the distribution. Boxplots in Geoda also fetch data for min, max, Q1, median, Q3, IQR, mean, and s.d.These boxplots represent the number of adults who are illiterate in Nepalese districts.



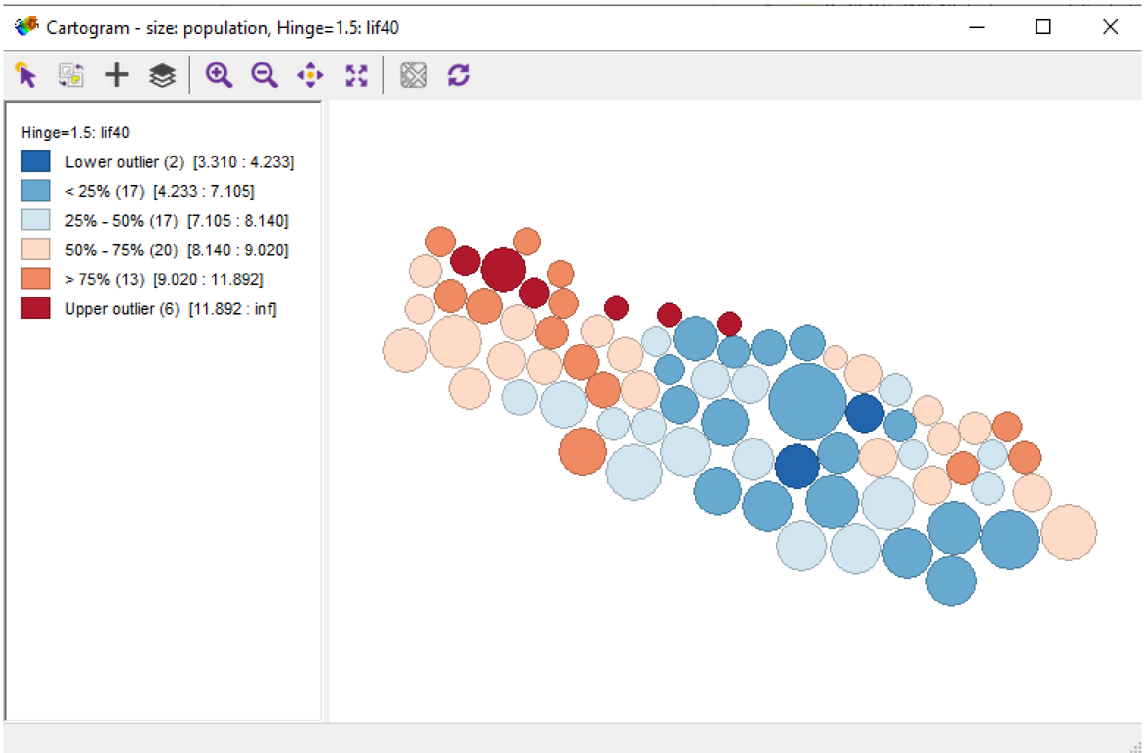

| | malkids |
|---|---|
| min | 16.2000 |
| max | 65.7000 |
| Q1 | 34.7500 |
| median | 44 |
| Q3 | 50.7000 |
| IQR | 15.9500 |
| mean | 42.8053 |
| s.d. | 10.9697 |




| | malkids |
|---|---|
| min | 16.2000 |
| max | 65.7000 |
| Q1 | 34.7500 |
| median | 44 |
| Q3 | 50.7000 |
| IQR | 15.9500 |
| mean | 42.8053 |
| s.d. | 10.9697 |

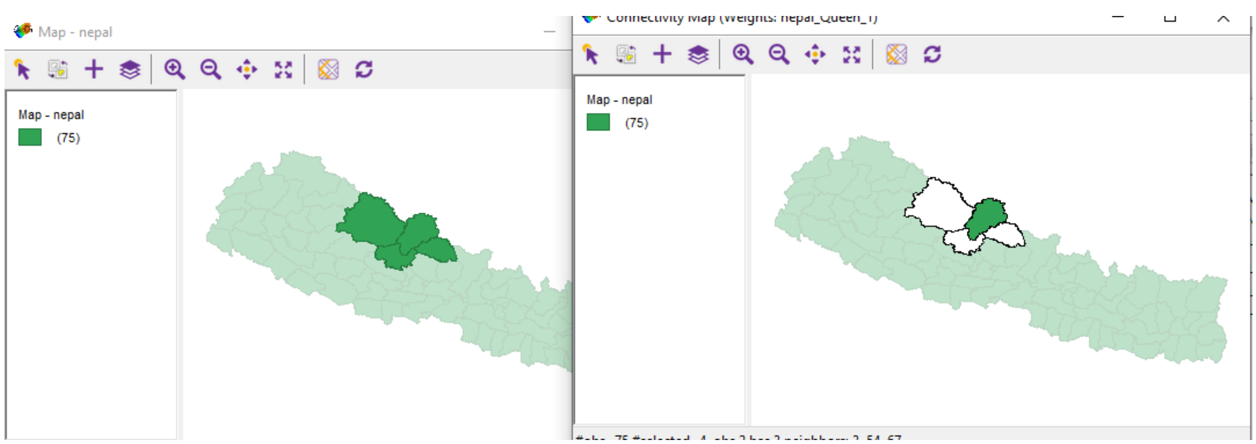
*Cartogram*

A cartogram is a map type where the original layout of the areal unit is replaced by a geometric form (usually a circle, rectangle, or hexagon) that is proportional to the value of the variable for the location. This cartogram represents population sizes, and people not expected to live past 40 years old in Nepal.




**Bivariate Moran's I**

Screenshot of your connectivity map for a Queen's Case first order.

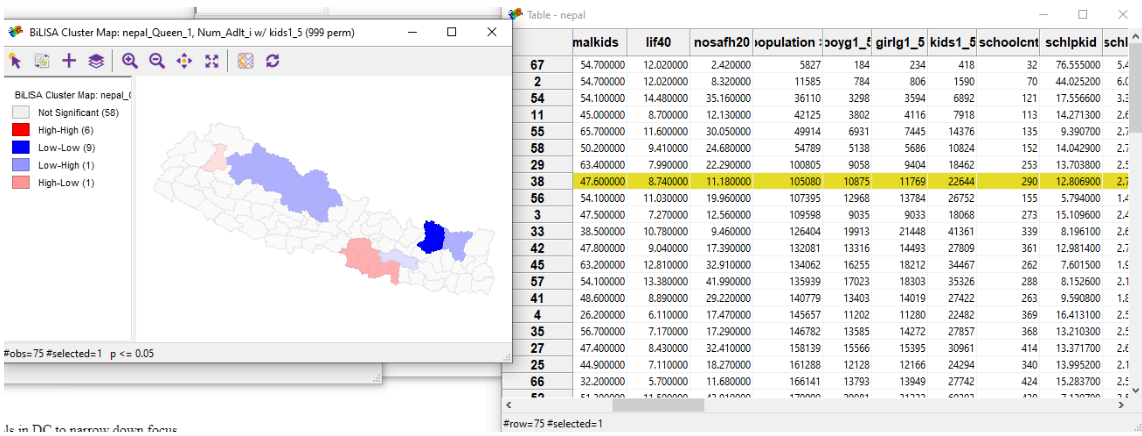



*Bivariate Moran's I.*

Bivariate local Moran's I describes the statistical relationship between one variable at a location and a spatially lagged second variable at neighboring locations, and geographically weighted regression (GWR) allows regression coefficients to vary at each observation location. We will examine bivariate Moran’s I to see if there is a spatial correlation between the number of kids enrolled in grades 1-5 and the number of adults who are illiterate in Nepal.

*A highlighted county that helps explain spatial lag (with screenshot), and an explanation from you.*

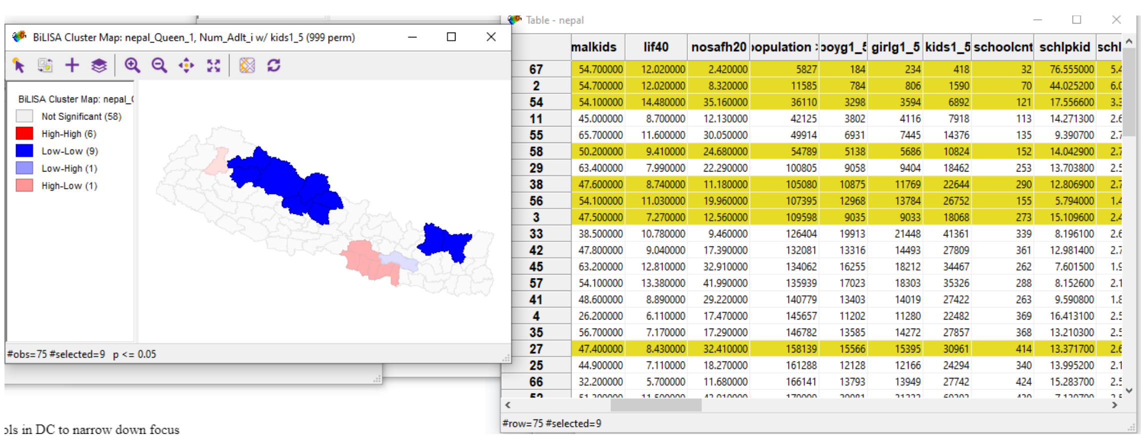
In the below we see the Bivariate Moran's I showing spatial lag as events (the number of kids enrolled in grades 1-5) predict an increased likelihood of other events (the number of adults who are illiterate in Nepal. Here we see a variety of districts that are significant, identifying High-high (high number of kids enrolled in grades 1-5 surrounded by high number of adults who are illiterate), low-low (a low number of kids enrolled in grades 1-5 surrounded by low number of adults who are illiterate), high-low (a high number of kids enrolled in grades 1-5 surrounded by low number of adults who are illiterate), and low-high (a low number of kids enrolled in grades 1-5 surrounded by high number of adults who are illiterate)




| | malkids | lif40 | nosafh20 | population | boyg1_5 | girlg1_5 | kids1_5 | schoolcnt | schlpkid | schl |
|---|---|---|---|---|---|---|---|---|---|---|
| 67 | 54.700000 | 12.020000 | 2.420000 | 5827 | 184 | 234 | 418 | 32 | 76.555000 | 5.4 |
| 2 | 54.700000 | 12.020000 | 8.320000 | 11585 | 784 | 806 | 1590 | 70 | 44.025200 | 6.0 |
| 54 | 54.100000 | 14.480000 | 35.160000 | 36110 | 3298 | 3594 | 6892 | 121 | 17.556600 | 3.3 |
| 11 | 45.000000 | 8.700000 | 12.130000 | 42125 | 3802 | 4116 | 7918 | 113 | 14.271300 | 2.6 |
| 55 | 65.700000 | 11.600000 | 30.050000 | 49914 | 6931 | 7445 | 14376 | 135 | 9.390700 | 2.7 |
| 58 | 50.200000 | 9.410000 | 24.680000 | 54789 | 5138 | 5686 | 10824 | 152 | 14.042900 | 2.7 |
| 29 | 63.400000 | 7.990000 | 22.290000 | 100805 | 9058 | 9404 | 18462 | 253 | 13.703800 | 2.5 |
| 38 | 47.600000 | 8.740000 | 11.180000 | 105080 | 10875 | 11769 | 22644 | 290 | 12.806900 | 2.7 |
| 56 | 54.100000 | 11.030000 | 19.960000 | 107395 | 12968 | 13784 | 26752 | 155 | 5.794000 | 1.4 |
| 3 | 47.500000 | 7.270000 | 12.560000 | 109598 | 9035 | 9033 | 18068 | 273 | 15.109600 | 2.4 |
| 33 | 38.500000 | 10.780000 | 9.460000 | 126404 | 19913 | 21448 | 41361 | 339 | 8.196100 | 2.6 |
| 42 | 47.800000 | 9.040000 | 17.390000 | 132081 | 13316 | 14493 | 27809 | 361 | 12.981400 | 2.7 |
| 45 | 63.200000 | 12.810000 | 32.910000 | 134062 | 16255 | 18212 | 34467 | 262 | 7.601500 | 1.9 |
| 57 | 54.100000 | 13.380000 | 41.990000 | 135939 | 17023 | 18303 | 35326 | 288 | 8.152600 | 2.1 |
| 41 | 48.600000 | 8.890000 | 29.220000 | 140779 | 13403 | 14019 | 27422 | 263 | 9.590800 | 1.8 |
| 4 | 26.200000 | 6.110000 | 17.470000 | 145657 | 11202 | 11280 | 22482 | 369 | 16.413100 | 2.5 |
| 35 | 56.700000 | 7.170000 | 17.290000 | 146782 | 13585 | 14272 | 27857 | 368 | 13.210300 | 2.5 |
| 27 | 47.400000 | 8.430000 | 32.410000 | 158139 | 15566 | 15395 | 30961 | 414 | 13.371700 | 2.6 |
| 25 | 44.900000 | 7.110000 | 18.270000 | 161288 | 12128 | 12166 | 24294 | 340 | 13.995200 | 2.1 |
| 66 | 32.200000 | 5.700000 | 11.680000 | 166141 | 13793 | 13949 | 27742 | 424 | 15.283700 | 2.5 |

*Screenshot of your cluster map, along with your highlighted table, and an interpretation of your results.*

These districts were selected because they represent low-low (a low number of kids enrolled in grades 1-5 surrounded by a low number of adults who are illiterate). The reason why there were many low-low districts, specifically in the northwest, was because they tended to have lower populations as well. When running this function, without empirical bayes the computer does not account for this fact.




| | malkids | lif40 | nosafh20 | population | boyg1_5 | girlg1_5 | kids1_5 | schoolcnt | schlpkid | schl |
|---|---|---|---|---|---|---|---|---|---|---|
| 67 | 54.700000 | 12.020000 | 2.420000 | 5827 | 184 | 234 | 418 | 32 | 76.555000 | 5.4 |
| 2 | 54.700000 | 12.020000 | 8.320000 | 11585 | 784 | 806 | 1590 | 70 | 44.025200 | 6.0 |
| 54 | 54.100000 | 14.480000 | 35.160000 | 36110 | 3298 | 3594 | 6892 | 121 | 17.556600 | 3.3 |
| 11 | 45.000000 | 8.700000 | 12.130000 | 42125 | 3802 | 4116 | 7918 | 113 | 14.271300 | 2.6 |
| 55 | 65.700000 | 11.600000 | 30.050000 | 49914 | 6931 | 7445 | 14376 | 135 | 9.390700 | 2.7 |
| 58 | 50.200000 | 9.410000 | 24.680000 | 54789 | 5138 | 5686 | 10824 | 152 | 14.042900 | 2.7 |
| 29 | 63.400000 | 7.990000 | 22.290000 | 100805 | 9058 | 9404 | 18462 | 253 | 13.703800 | 2.5 |
| 38 | 47.600000 | 8.740000 | 11.180000 | 105080 | 10875 | 11769 | 22644 | 290 | 12.806900 | 2.7 |
| 56 | 54.100000 | 11.030000 | 19.960000 | 107395 | 12968 | 13784 | 26752 | 155 | 5.794000 | 1.4 |
| 3 | 47.500000 | 7.270000 | 12.560000 | 109598 | 9035 | 9033 | 18068 | 273 | 15.109600 | 2.4 |
| 33 | 38.500000 | 10.780000 | 9.460000 | 126404 | 19913 | 21448 | 41361 | 339 | 8.196100 | 2.6 |
| 42 | 47.800000 | 9.040000 | 17.390000 | 132081 | 13316 | 14493 | 27809 | 361 | 12.981400 | 2.7 |
| 45 | 63.200000 | 12.810000 | 32.910000 | 134062 | 16255 | 18212 | 34467 | 262 | 7.601500 | 1.9 |
| 57 | 54.100000 | 13.380000 | 41.990000 | 135939 | 17023 | 18303 | 35326 | 288 | 8.152600 | 2.1 |
| 41 | 48.600000 | 8.890000 | 29.220000 | 140779 | 13403 | 14019 | 27422 | 263 | 9.590800 | 1.8 |
| 4 | 26.200000 | 6.110000 | 17.470000 | 145657 | 11202 | 11280 | 22482 | 369 | 16.413100 | 2.5 |
| 35 | 56.700000 | 7.170000 | 17.290000 | 146782 | 13585 | 14272 | 27857 | 368 | 13.210300 | 2.5 |
| 27 | 47.400000 | 8.430000 | 32.410000 | 158139 | 15566 | 15395 | 30961 | 414 | 13.371700 | 2.6 |
| 25 | 44.900000 | 7.110000 | 18.270000 | 161288 | 12128 | 12166 | 24294 | 340 | 13.995200 | 2.1 |
| 66 | 32.200000 | 5.700000 | 11.680000 | 166141 | 13793 | 13949 | 27742 | 424 | 15.283700 | 2.5 |

*Answer to Question: if we have somewhat misleading results with this analysis of absolute population numbers, what would be a better way to do this?*

A better way to run an analysis with population is to normalize it by using Empirical Bayes. In essence, the EB technique consists of computing a weighted average between the raw rate for each county and the state average, with weights proportional to the underlying population at risk.

**Local Moran's with EB Rates**

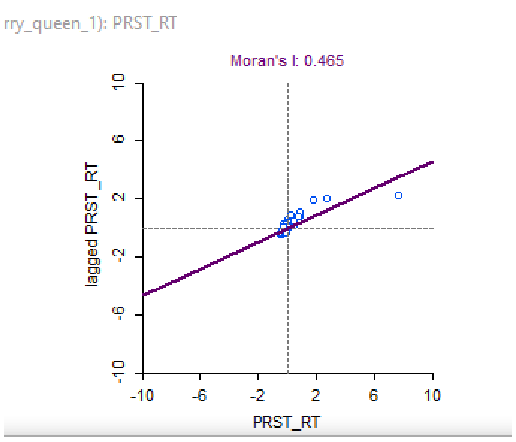
We will be examining a comparison of pre-calculated rates of donations to the poor/population in 1831 France (*DNTNS_RT*) *vs.* allowing the Moran's I with Empirical Bayesian Rates to do the calculation; and a comparison of pre-calculated rates of prostitutes/population (*PRST_RT*) around France (1816 – 1834) *vs.* again allowing the Moran's with Empirical Bayesian rates to do the calculation. Two things to note:

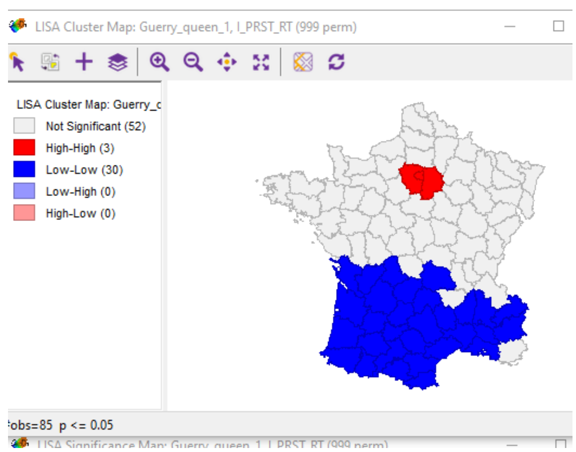
*Explanation of what the Moran's with EB rates is doing.*

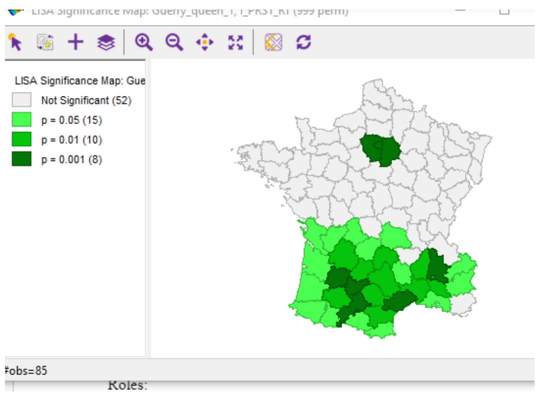
The Local Moran's with EB rates is determining if there is a spatial relationship with donations to the poor/ population in 1891 France and a comparison of pre-calculated rates of prostitutes/population (*PRST_RT*) around France (1816 – 1834) *vs.* again allowing the Moran's with Empirical Bayesian rates to do the calculation. The EB rates will help smooth spurious outliers, in essence, the EB technique consists of computing a weighted average between the raw rate for each county and the state average, with weights proportional to the underlying population at risk. Areas like these counties that have smaller populations compared to the other areas on the data set will typically have their rates adjusted considerably, where larger counties with larger populations will hardly change.

A screenshot of the regular univariate local Moran's I (like on page 18).

<u>Prostitutes in Paris: 1830's</u>

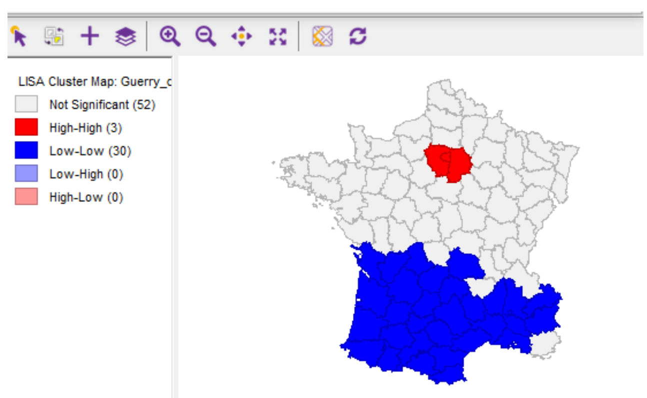


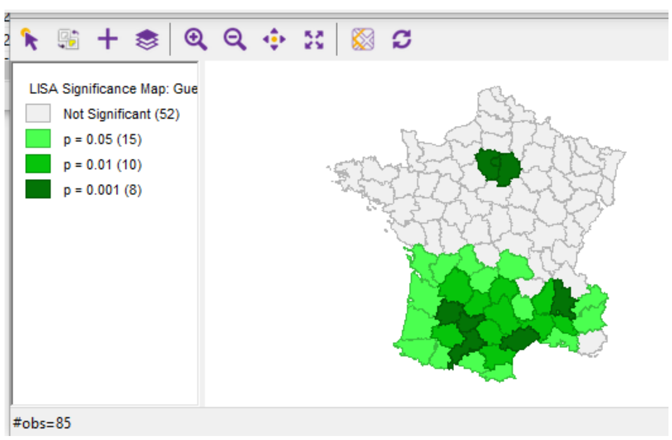






A screenshot of the EB rate-corrected local Moran's I (like on page 19).

<u>Prostitutes in Paris: 1830's</u>


<u>Prostitutes in Paris: 1830's</u>

*Explanation of your results.*

There are high-high rates of spatial autocorrelation of prostitutes in France surrounding the capital city, paris. Whereas the Low-low rates were not in metropolitan areas in 1830's, so there won't be high rates of prostitutes.

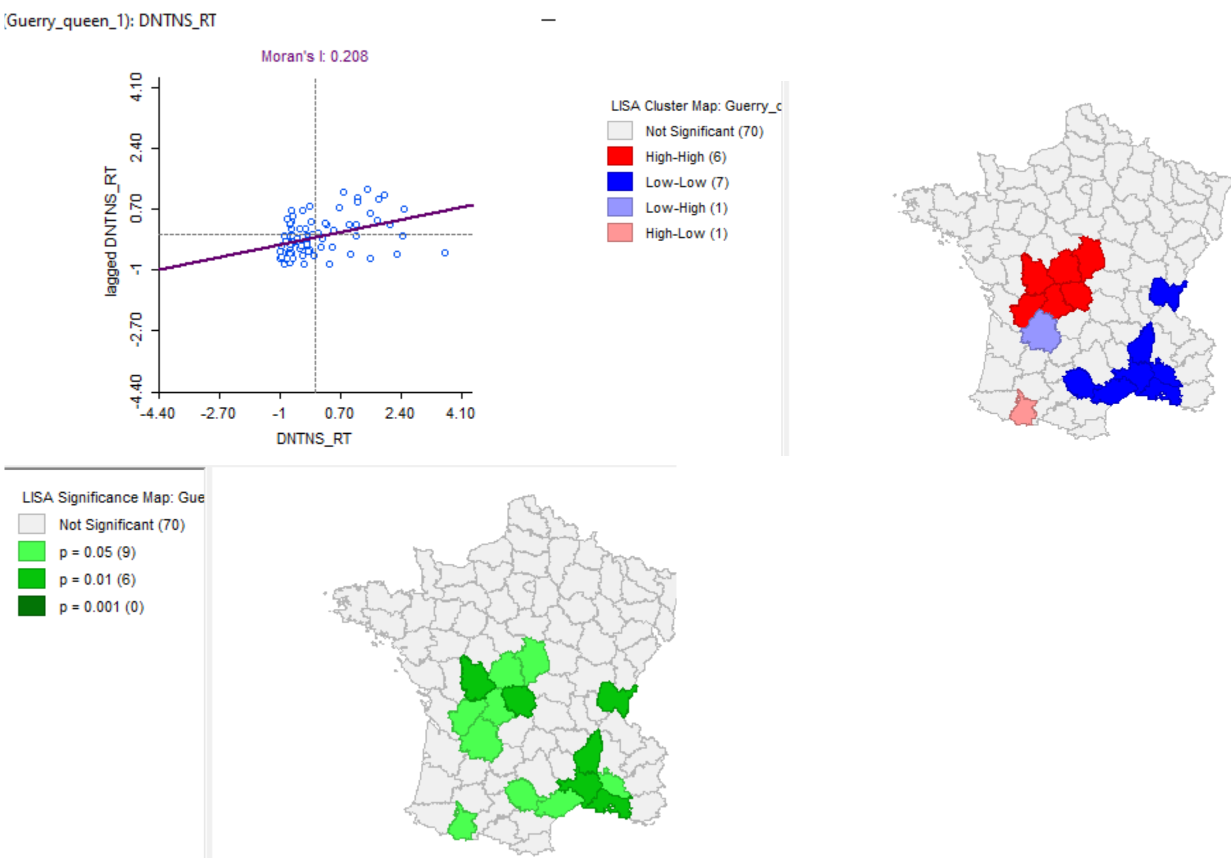
Prostitutes in Paris: 1830's

*Answer to the question (asked on page 20): Why do you think the results are almost identical?*

Due to the fact that these counties look relatively uniform in size, the EB function won't change rates considerably.

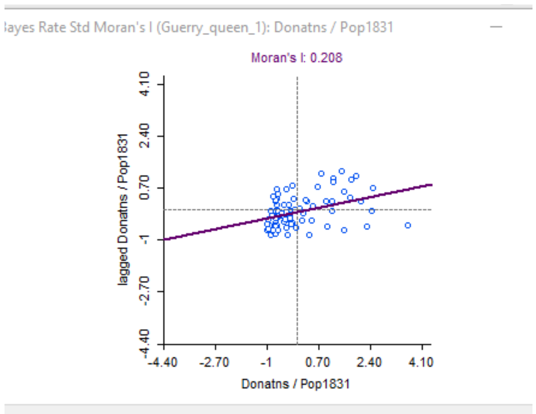
Donations to the Poor

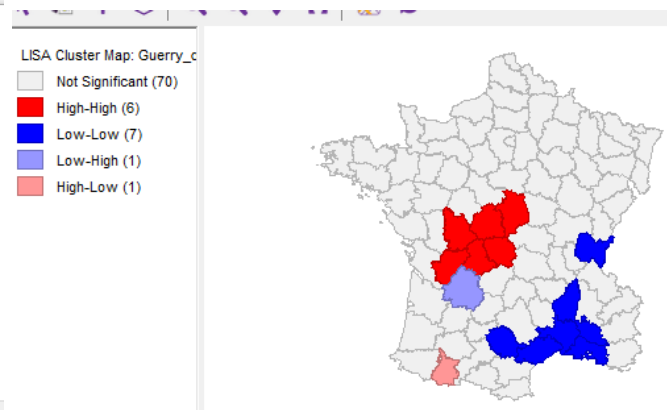
A screenshot of the regular univariate local Moran's I (like on page 20).

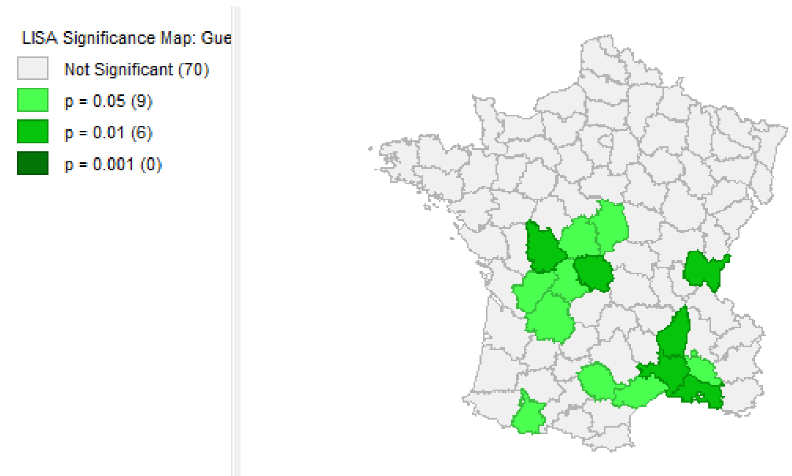



Donations to the Poor

A screenshot of the EB rate-corrected local Moran's I (like on page 21).




Donations to the Poor

*Explanation of your results.*

High rates of donations to the poor are highly related to the county's high rates near them in western central France. Where there is a low amount of donations with other county's also giving low donations is centered towards the southeast of France. These might have to do with wealth distribution in the country.

Donations to the Poor

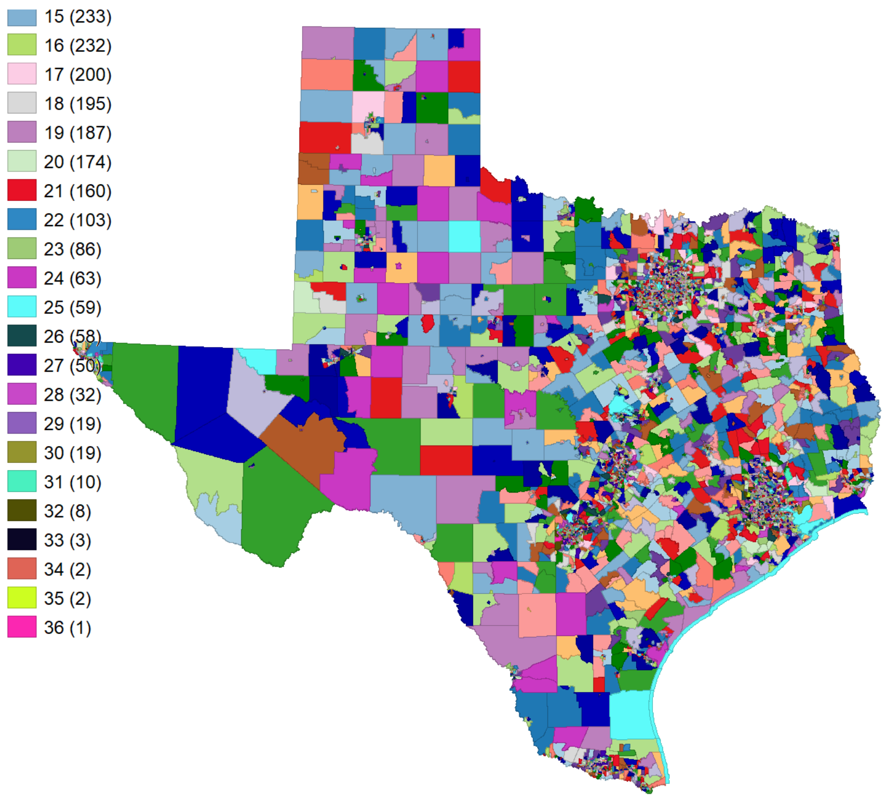
*Answer to the question (asked on page 21): Why do you think the results here actually are identical?*

Again there must be an evenness between the population size in the counties of France so the spurious outliers are so insignificant that EB doesn't have to correct for the issue of small counties having small populations in relation to the counties in the data set

**K means clustering**

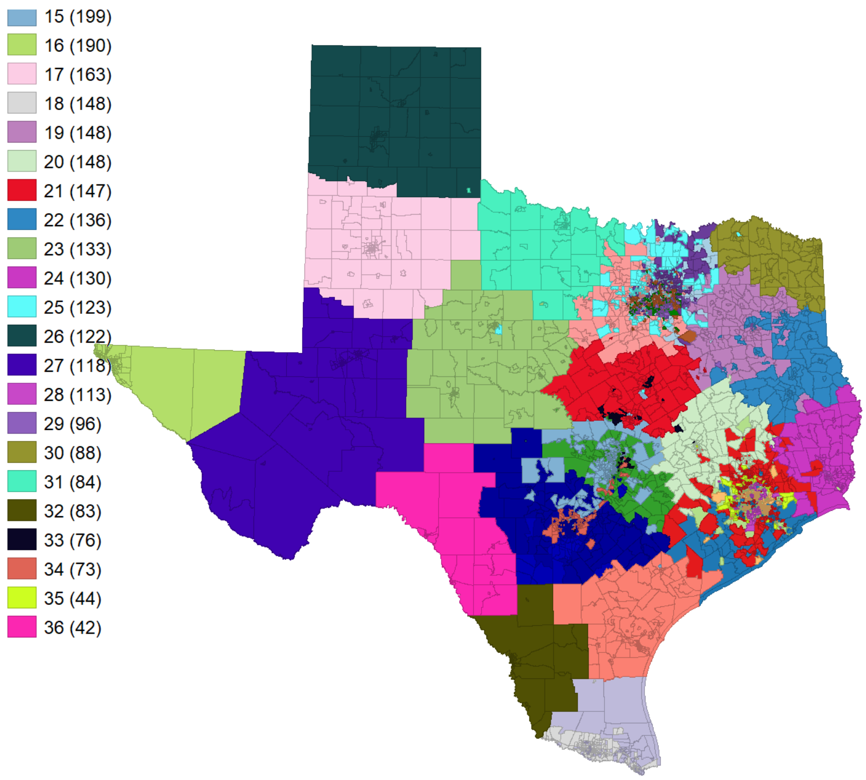
*Your first map is based on the total population without geometric centroids, and your printout. Explain* *your printout as on pages 18 and 19 here.*

This identifies districts based on certain criteria, typically k-means groups similar data points together and discover underlying patterns. Here there isn't much grouping going on because we haven't used weights. Essentially this map is not grouped based on its neighbors.



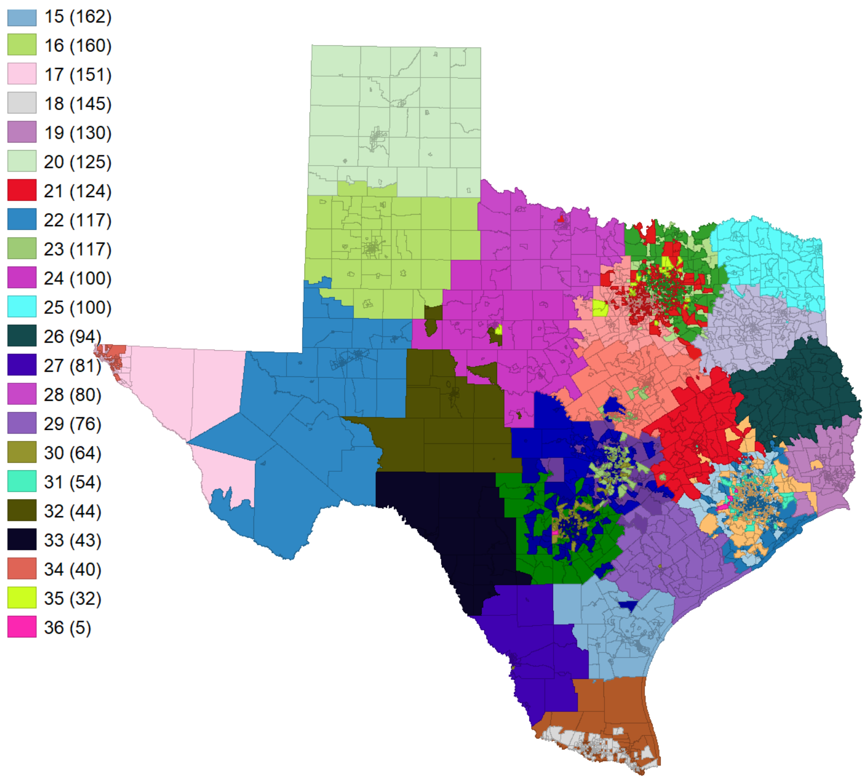

*Your second map with centroids weighted at 0.65, and print out. What did your ratio of between to the total sum of squares do in comparison to the first map?*

The ratio between the total sum of squares of my centroids weighted at 0.65 compared to my first map got smaller.



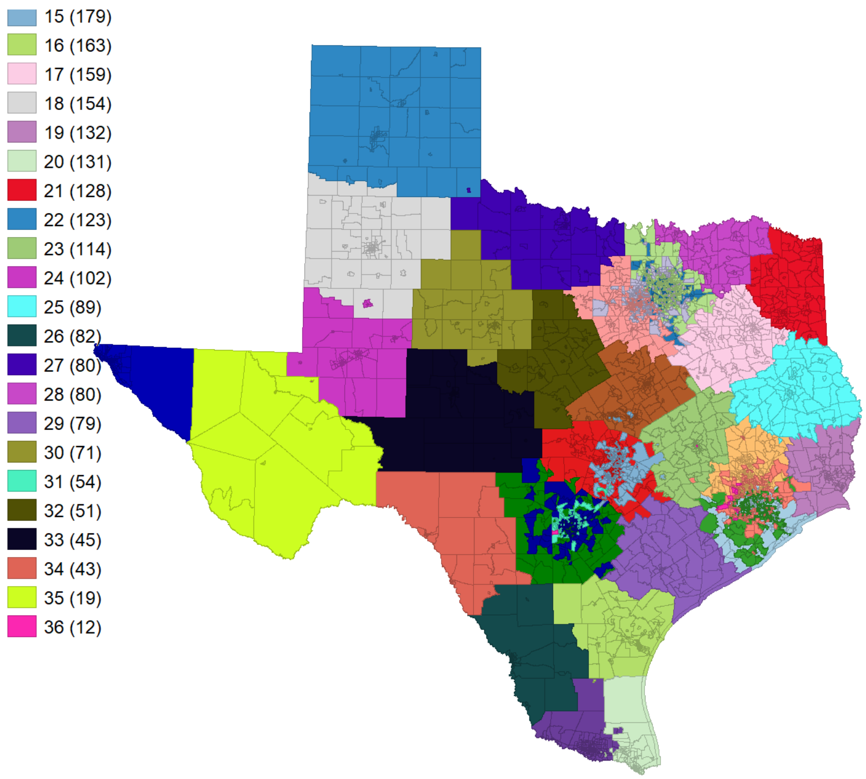

*Your second map with centroids weighted at 0.85, and print out. What did your ratio of between to the total sum of squares do in comparison to the first map?*

The ratio between the total sum of squares from my centroids weighted at .85 compared to my first map got smaller.




*Your final map of clusters by ethnicity is weighted at 0.65, and print out. Did you find this map was closer to the districts as seen in the reference image?*

Yes, I think the final ethnicity weighted at 0.65 map looked a lot closer to the districts in the reference image.


15 (179)
16 (163)
17 (159)
18 (154)
19 (132)
20 (131)
21 (128)
22 (123)
23 (114)
24 (102)
25 (89)
26 (82)
27 (80)
28 (80)
29 (79)
30 (71)
31 (54)
32 (51)
33 (45)
34 (43)
35 (19)
36 (12)