

Your Name = [Ankila Kumari

GIS 5555 Basic Spatial Analysis

] internet_id = [kuma0389]

Time_Spent = [40 mins] (after-class)

[Windows+Shift+S for screenshot of your analysis]

[Fill the above-listed info and then submit the completed document in Canvas (try to include all analysis results that can help reflect your workflow and thoughts, i.e., images, information about data, your statements, etc.)]

Assignment for Lab 4a

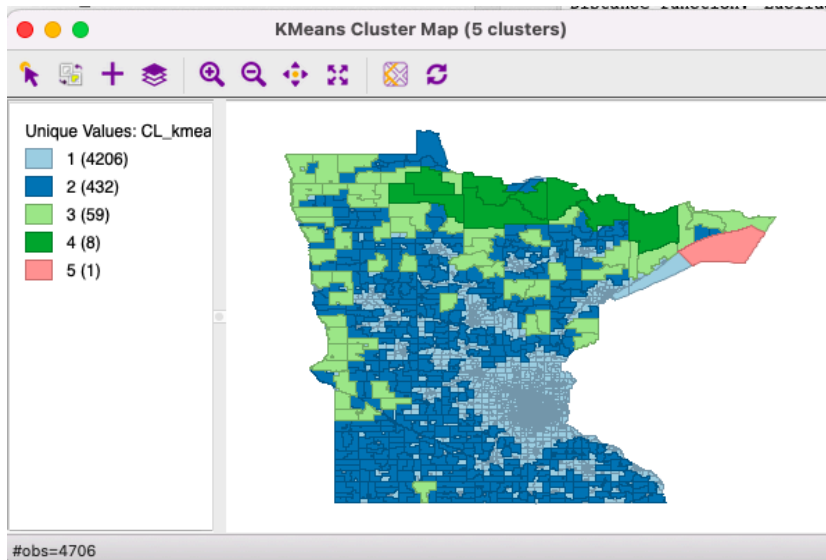
“Spatial Clustering”

Please choose carefully on your selected set of variables. We expect the selected variables to represent multiple characteristics of your geographical units. There should be some grouped variations to highlight the purpose of this lab exercise.

➤ Task 1 K-means clustering

- Select at least three variables that are of different attributional meanings for following analysis. Explain the chosen variables.
- Fix k as five, conduct K-means using K-means++ algorithm and 1000 initialization re-runs
 - Read the summary table, especially the ratio of between to total sum of squares (BSS/TSS)
- Change the data transformation from “Standardization” (z-score) to “Range standardization”, which may be more robust to the effect of outliers, how is the overall BSS/TSS ratio changing for your data?

Answers - I clustered ALAND, AWATER, and COORD_X using K-means++ ($k=5$) with 1000 iterations. The initial BSS/TSS ratio was 0.733 (z-score), improving to 0.808 with range standardization, showing better separation after reducing outlier effects. Cluster sizes were uneven (e.g., one cluster had only 1 observation).



KMeans Clustering Settings

Input:

Select Variables

Variable
ALAND
AWATER
NUM_N_ueen
COORD_X
COORD_Y
Z
Z_LAG
Z_LAG_arch
LISA_I
LISA_CL
LISA_P
I
Water_lag

☐ Use geometric centroids ☐ Auto Weighting

Weighting: 0 1 1

Select Spatial Weights: tl_2023_27_bg

Parameters:

Number of Clusters: 5

Minimum Bound: ☐ ALAND 1 10%

Transformation: Range Standardize

Initialization Method: KMeans++

Initialization Re-runs: 1000

Use Specified Seed: ☒ Change Seed

Maximum Iterations: 1000

Distance Function: Euclidean

Output:

Save Cluster in Field: CL_kmeans

Run Close

Summary

Method: KMeans
Number of clusters: 5
Initialization method: KMeans++
Initialization re-runs: 1000
Maximum iterations: 1000
Transformation: Range Standardize
Distance function: Euclidean

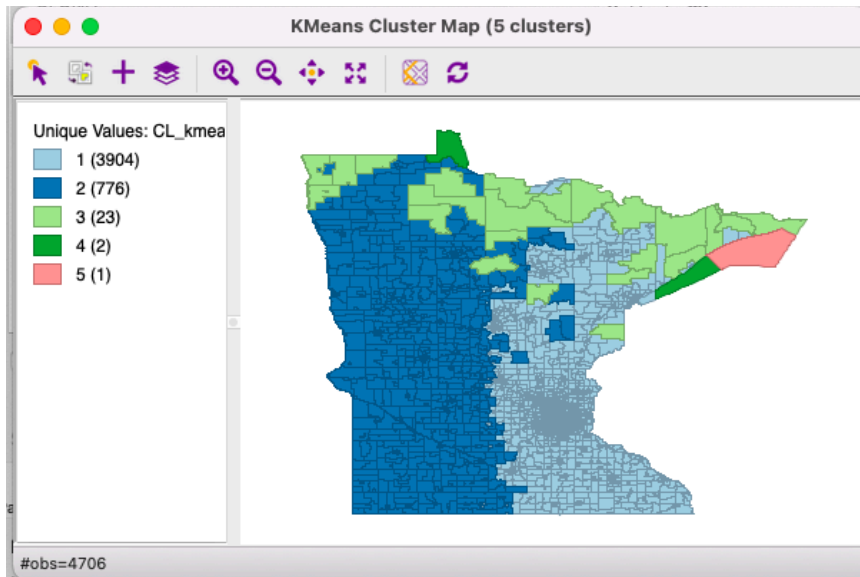
Cluster centers:

	ALAND	AWATER	COORD_X
C1	1.10081e+07	1.49016e+06	-93.4866
C2	2.29935e+08	1.11574e+07	-94.485
C3	7.13637e+08	3.84853e+07	-94.8809
C4	2.31348e+09	1.7425e+08	-93.4594
C5	0	4.18524e+09	-90.2894

The total sum of squares: 8.50052
Within-cluster sum of squares:

	Within cluster S.S.
C1	0.682052
C2	0.388217
C3	0.252262
C4	0.30924
C5	0

The total within-cluster sum of squares: 1.63177
The between-cluster sum of squares: 6.86875
The ratio of between to total sum of squares: 0.808039



KMeans Clustering Settings

Input:

Select Variables

ALAND
AWATER
NUM_N_ueen
COORD_X
COORD_Y
Z
Z_LAG
Z_LAG_arch
LISA_I
LISA_CL
LISA_P
I
Water_lag

☐ Use geometric centroids

Weighting: 0 1

Select Spatial Weights: tl_2023_27_bg

Parameters:

Number of Clusters: 5

Minimum Bound: ☐ ALAND

Transformation: Standardize (Z)

Initialization Method: KMeans++

Initialization Re-runs: 1000

Use Specified Seed: ☒ Change Seed

Maximum Iterations: 1000

Distance Function: Euclidean

Output:

Save Cluster in Field: CL_kmeans

Summary

Number of clusters: 5
Initialization method: KMeans++
Initialization re-runs: 1000
Maximum iterations: 1000
Transformation: Standardize (Z)
Distance function: Euclidean

Cluster centers:

	ALAND	AWATER	COORD_X
C1	1.72642e+07	1.02137e+06	-93.2284
C2	1.34391e+08	6.20947e+06	-95.4467
C3	1.48684e+09	1.28923e+08	-93.5307
C4	1.79949e+08	1.49052e+09	-93.1263
C5	0	4.18524e+09	-90.2894

The total sum of squares: 14115
Within-cluster sum of squares:

	Within cluster S.S.
C1	1484.76
C2	1415.86
C3	829.912
C4	36.0374
C5	0

The total within-cluster sum of squares: 3766.57
The between-cluster sum of squares: 10348.4
The ratio of between to total sum of squares: 0.733152

Method: KMeans
Number of clusters: 5
Initialization method: KMeans++
Initialization re-runs: 1000
Maximum iterations: 1000
Transformation: Range Standardize
Distance function: Euclidean

Cluster centers:

	ALAND	AWATER	COORD_X
C1	1.10081e+07	1.49016e+06	-93.4866
C2	2.29935e+08	1.11574e+07	-94.485
C3	7.13637e+08	3.84853e+07	-94.8809
C4	2.31348e+09	1.7425e+08	-93.4594
C5	0	4.18524e+09	-90.2894

The total sum of squares: 8.50052
Within-cluster sum of squares:

	Within cluster S.S.
C1	0.682052
C2	0.388217
C3	0.252262
C4	0.30924
C5	0

The total within-cluster sum of squares: 1.63177
The between-cluster sum of squares: 6.86875
The ratio of between to total sum of squares: 0.808039

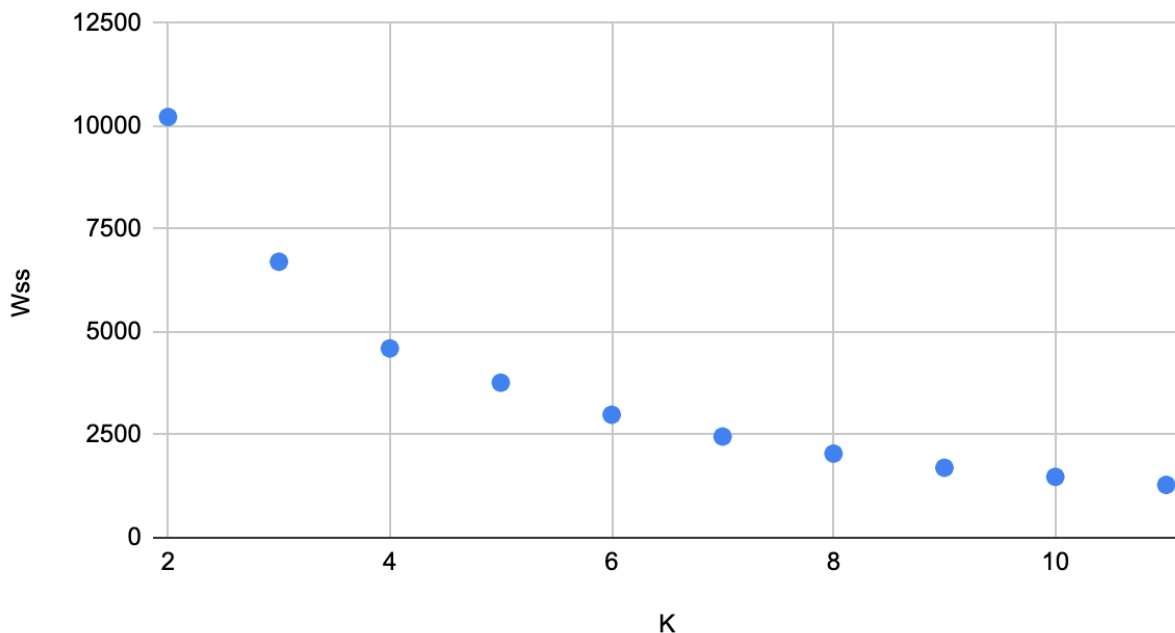
➤ **Task 2 Elbow plot to determine the cluster numbers in K means**

- Conduct a K-means clustering using your fixed customized settings, but change the number of clusters (k) from 1 to 10, record all tuples of $\langle k, BSS/TSS \rangle$ in a .csv table
- Visualize the elbow plot and discuss the optimal number of clusters in this scenario.

Ans - Testing $k=2$ to 12 , I recorded BSS/TSS ratios and created an elbow plot. The optimal $k=5$ was chosen where the curve flattened, balancing cluster separation and simplicity. Higher k values showed diminishing returns.

K	WSS	BSS/TSS	Change
2	10222.9	0.275745	10222.62426
3	6705.36	0.524948	6704.835052
4	4598.7	0.674197	4598.025803
5	3766.57	0.733152	3765.836848
6	2987.15	0.788371	2986.361629
7	2457.25	0.825912	2456.424088
8	2043.01	0.855259	2042.154741
9	1698.87	0.879641	1697.990359
10	1478.24	0.895271	1477.344729
11	1282.86	0.909114	1281.950886

Wss vs. K



➤ Task 3 Agglomerative clustering

- Perform the default agglomerative clustering using Ward-linkage and $k=5$
 - What's the reported BSS/TSS?

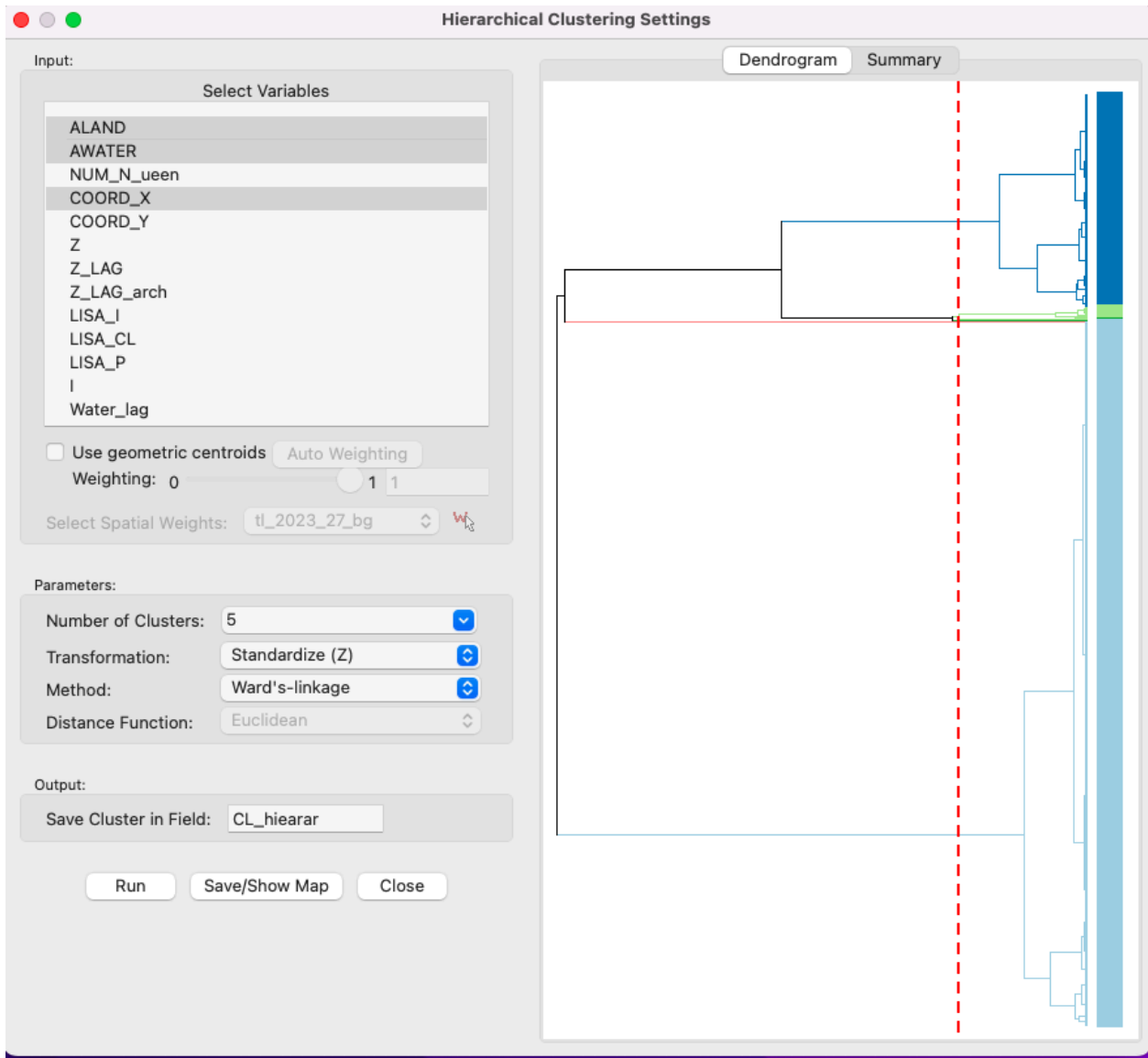
Ans- The total within-cluster sum of squares: 4081.21

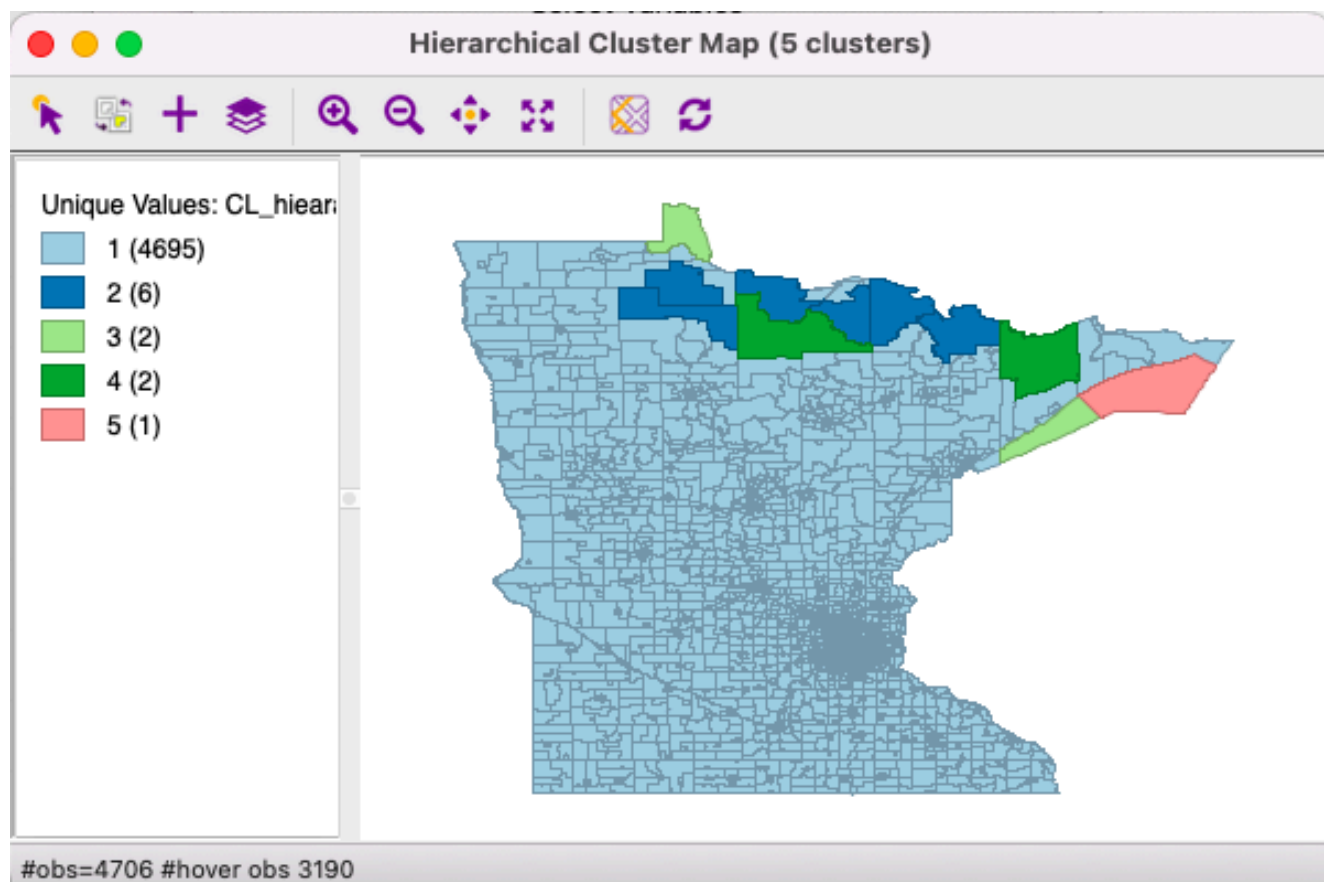
The between-cluster sum of squares: 10033.8

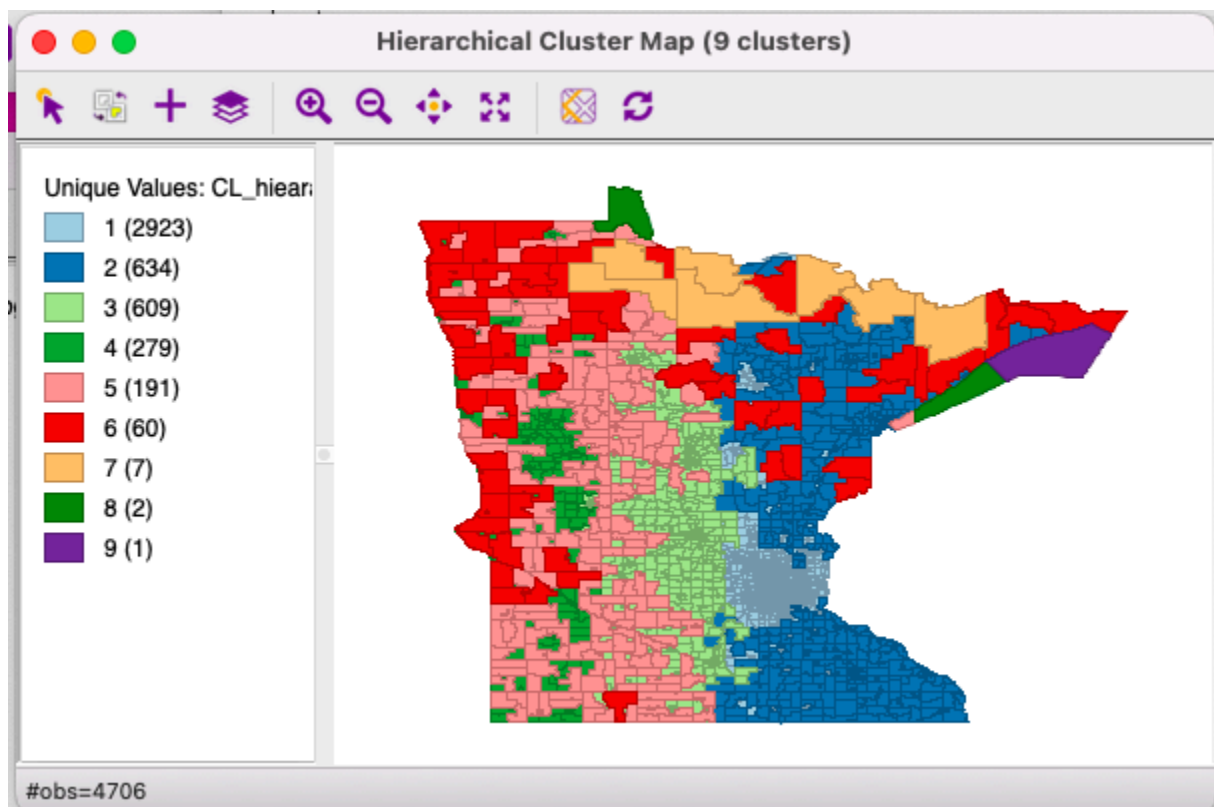
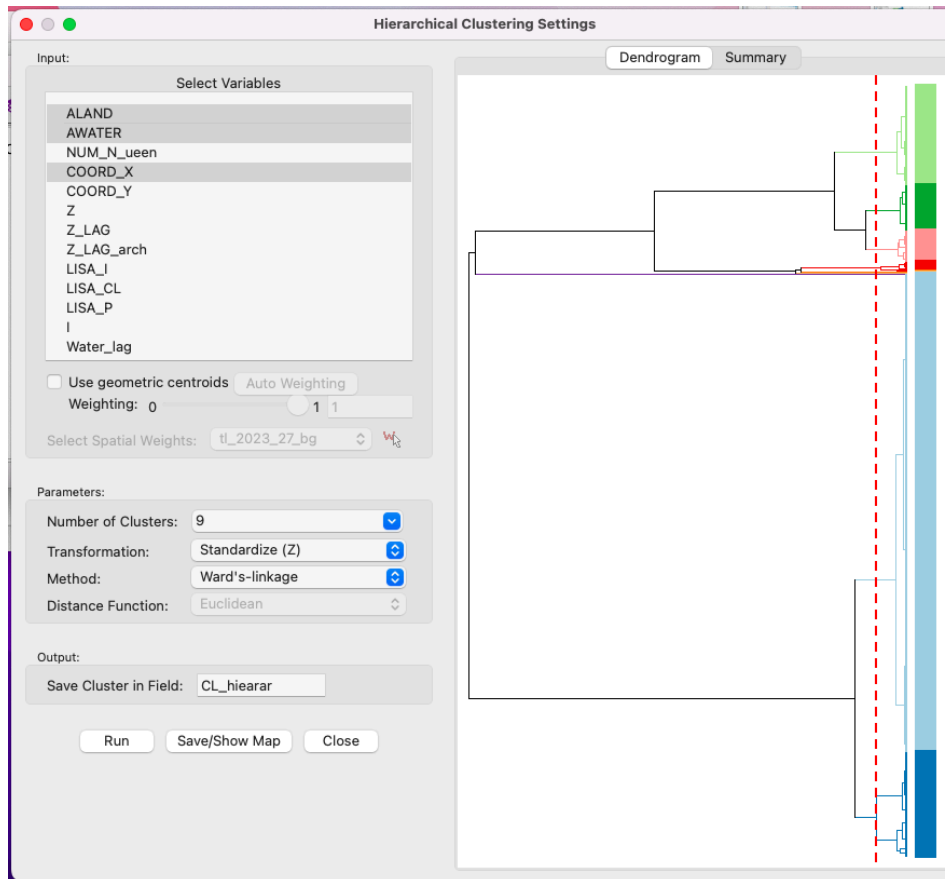
The ratio of between to total sum of squares: 0.71086

- Show the cluster map & the dendrogram
 - Change k interactively via the dendrogram, show the new cluster map and the new dendrogram
- Fix all your clustering settings besides the linkage option; Which one out of the four linkages yield the best clustering results for you? Why?

Answers- Using **Ward's linkage ($k=5$)**, the **BSS/TSS ratio was 0.711**, with clusters more balanced than K-means. Comparing linkages, **Ward's performed best**; single linkage produced singletons, while complete/average linkages had poorer ratios (0.25–0.42).







ical Clustering Settings

Dendrogram

Summary

```

-----
Number of clusters: 9
Transformation: Standardize (Z)
Method: Ward's-linkage
Distance function: Euclidean

Cluster centers:

```

	ALAND	AWATER	COORD_X
C1	3.68231e+06	335586	-93.2737
C2	7.44581e+07	2.66751e+06	-92.5215
C3	3.78805e+07	3.77279e+06	-94.4492
C4	4.54025e+07	3.44931e+06	-96.0602
C5	2.71272e+08	1.23675e+07	-95.3437
C6	7.23608e+08	3.5264e+07	-94.9746
C7	2.42127e+09	1.94438e+08	-93.483
C8	1.79949e+08	1.49052e+09	-93.1263
C9	0	4.18524e+09	-90.2894

```

The total sum of squares: 14115
Within-cluster sum of squares:

```

	Within cluster S.S.
C1	150.681
C2	499.661
C3	182.444
C4	106.444
C5	208.461
C6	474.545
C7	232.385
C8	36.0374
C9	0

```

The total within-cluster sum of squares: 1890.66
The between-cluster sum of squares: 12224.3
The ratio of between to total sum of squares: 0.866053

```

```

-----
Number of clusters: 2
Transformation: Standardize (Z)
Method: Ward's-linkage
Distance function: Euclidean

```

```

Cluster centers:

```

	ALAND	AWATER	COORD_X
C1	1.62974e+07	751228	-93.1396
C2	1.29047e+08	1.41561e+07	-95.0047

```

The total sum of squares: 14115
Within-cluster sum of squares:

```

	Within cluster S.S.
C1	1068.19
C2	9476.9

```

The total within-cluster sum of squares: 10545.1
The between-cluster sum of squares: 3569.91
The ratio of between to total sum of squares: 0.252916

```

Hierarchical Clustering Settings

Input:

Select Variables

ALAND

AWATER

NUM_N_ueen

COORD_X

COORD_Y

Z

Z_LAG

Z_LAG_arch

LISA_I

LISA_CL

LISA_P

I

Water_lag

☐ Use geometric centroids

Auto Weighting

Weighting:

0

1

1

Select Spatial Weights:

tl_2023_27_bg

W_i

Parameters:

Number of Clusters:

5

Transformation:

Standardize (Z)

Method:

Complete-linkage

Distance Function:

Euclidean

Output:

Save Cluster in Field:

CL_hiearar

Run

Save/Show Map

Close

Dendrogram

Summary

ical Clustering Settings

Dendrogram

Summary

 Number of clusters: 9
 Transformation: Standardize (Z)
 Method: Ward's-linkage
 Distance function: Euclidean

Cluster centers:

	ALAND	AWATER	COORD_X
---	-----	-----	-----
C1	3.68231e+06	335586	-93.2737
C2	7.44581e+07	2.66751e+06	-92.5215
C3	3.78805e+07	3.77279e+06	-94.4492
C4	4.54025e+07	3.44931e+06	-96.0602
C5	2.71272e+08	1.23675e+07	-95.3437
C6	7.23608e+08	3.5264e+07	-94.9746
C7	2.42127e+09	1.94438e+08	-93.483
C8	1.79949e+08	1.49052e+09	-93.1263
C9	0	4.18524e+09	-90.2894

The total sum of squares: 14115
 Within-cluster sum of squares:

	Within cluster S.S.
---	-----
C1	150.681
C2	499.661
C3	182.444
C4	106.444
C5	208.461
C6	474.545
C7	232.385
C8	36.0374
C9	0

The total within-cluster sum of squares: 1890.66
 The between-cluster sum of squares: 12224.3
 The ratio of between to total sum of squares: 0.866053

 Number of clusters: 2
 Transformation: Standardize (Z)
 Method: Ward's-linkage
 Distance function: Euclidean

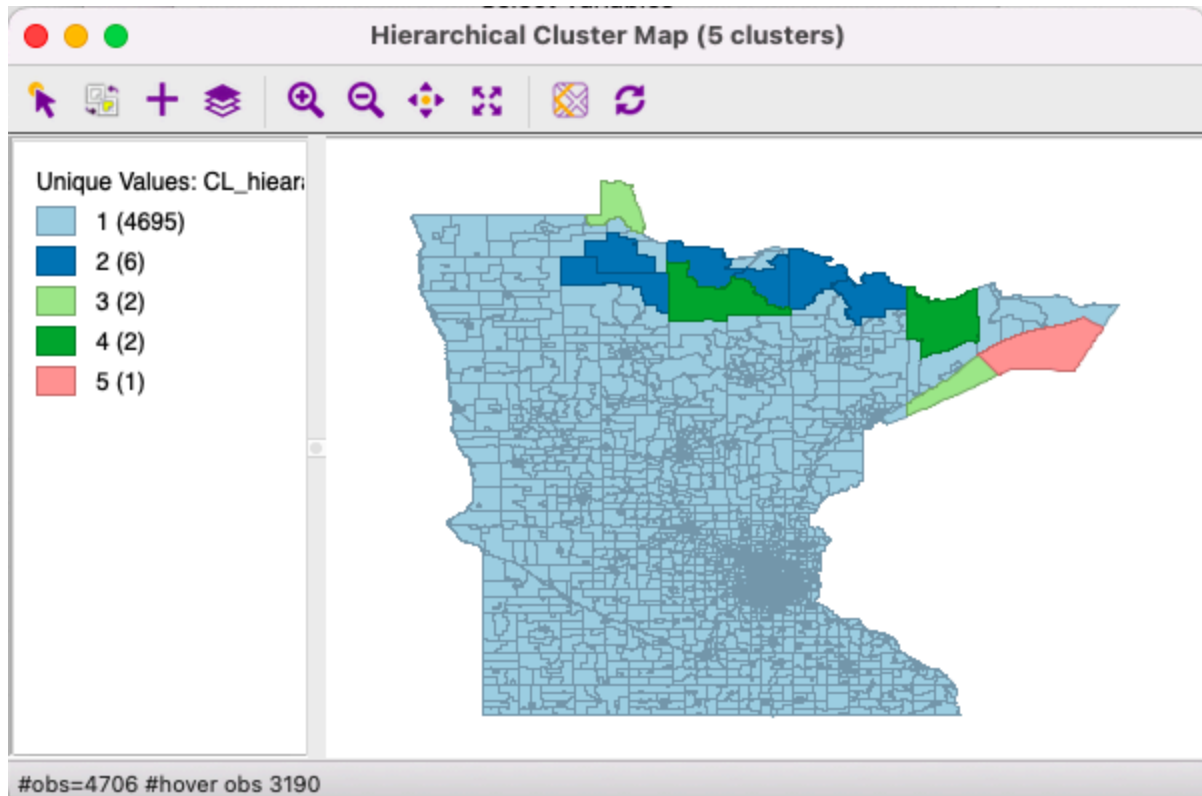
Cluster centers:

	ALAND	AWATER	COORD_X
---	-----	-----	-----
C1	1.62974e+07	751228	-93.1396
C2	1.29047e+08	1.41561e+07	-95.0047

The total sum of squares: 14115
 Within-cluster sum of squares:

	Within cluster S.S.
---	-----
C1	1068.19
C2	9476.9

The total within-cluster sum of squares: 10545.1
 The between-cluster sum of squares: 3569.91
 The ratio of between to total sum of squares: 0.252916



➤ **Task 4 Spatial constraint hierarchical clustering**

- Create a spatial weights matrix for your data
- Conducting SCHC for your data with z-standardized variables and the Ward's linkage.
 - Select your number of clusters via dendrogram
 - Read the BSS/TSS and show the cluster map

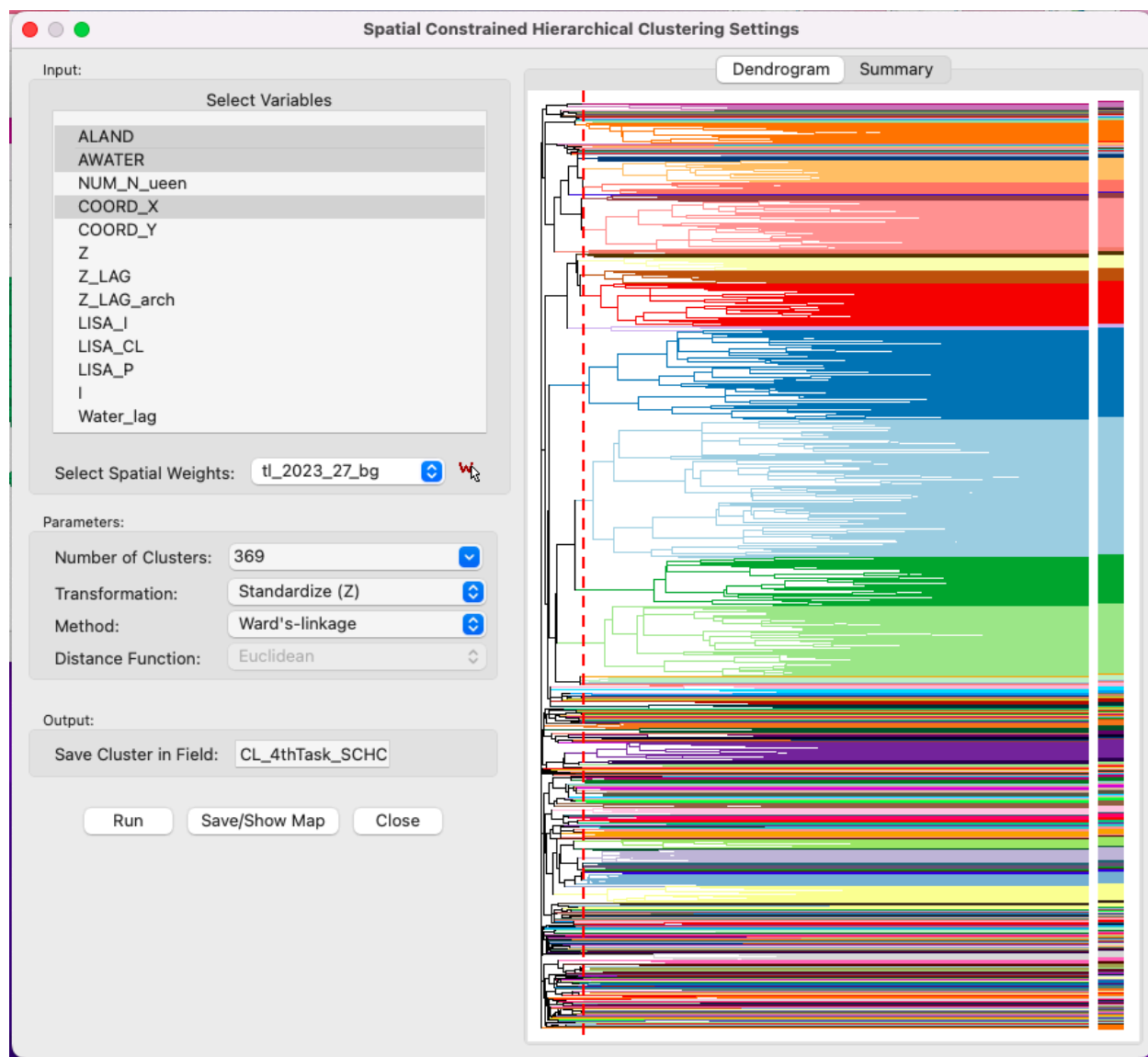
Ans- The total within-cluster sum of squares: 84.1869

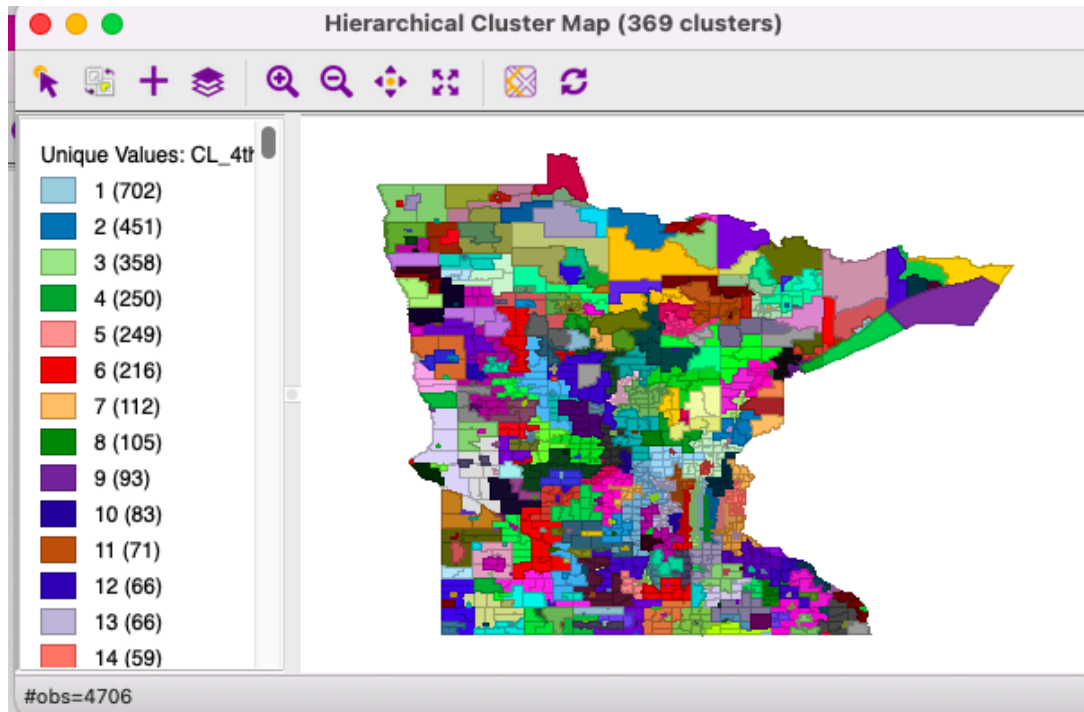
The between-cluster sum of squares: 14030.8

The ratio of between to total sum of squares: 0.994036

- What will be the results from the same clustering setting but without the spatial contiguity constraint? Explain from both the cluster map and the summary statistics.

Answer- With **Ward's linkage and spatial weights**, SCHC achieved a **BSS/TSS of 0.994**, creating contiguous clusters. Without constraints, the ratio dropped to **0.711**, showing spatial contiguity significantly improved cluster compactness.





Spatial Constrained KMeans Clustering Settings

Input:

Select Variables:

- ALAND
- AWATER
- NUM_N_ueen
- COORD_X
- COORD_Y
- Z
- Z_LAG
- Z_LAG_arch
- LISA_I
- LISA_CL
- LISA_P
- I
- Water_lag

Select Spatial Weights:

Parameters:

Number of Clusters:

Minimum Bound: ☐ ALAND 10%

Transformation:

Initialization Method:

Initialization Re-runs:

Use Specified Seed: ☒ Change Seed

Maximum Iterations:

Distance Function:

Output:

Save Cluster in Field:

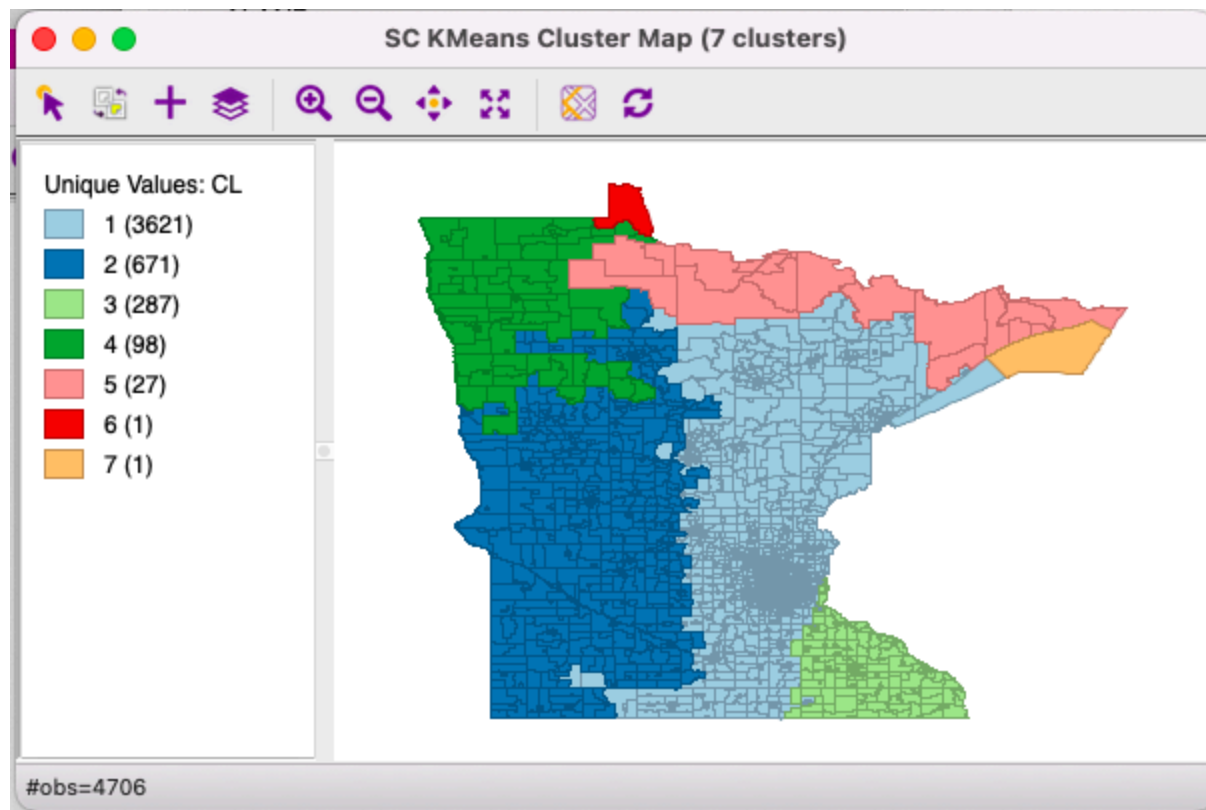
Summary

```
-----
Method: SC KMeans
Number of clusters: 7
Weights: tl_2023_27_bg
Initialization method: KMeans++
Initialization re-runs: 150
Maximum iterations: 1000
Transformation: Standardize (Z)
Distance function: Euclidean

Cluster centers:
-----
ALAND      AWATER      COORD_X
C1 1.91797e+07 1.76954e+06 -93.3026
C2 1.02758e+08 5.71277e+06 -95.3649
C3 4.88817e+07 836572 -92.3384
C4 2.90985e+08 8.28881e+06 -96.2137
C5 9.23667e+08 8.23765e+07 -92.7421
C6 3.59897e+08 1.2349e+09 -94.9684
C7 0 4.18524e+09 -90.2894

The total sum of squares: 14115
Within-cluster sum of squares:
-----
C1 2182.07
C2 840.812
C3 119.057
C4 491.49
C5 1583.98
C6 0
C7 0

The total within-cluster sum of squares: 5217.41
The between-cluster sum of squares: 8897.59
The ratio of between to total sum of squares: 0.630364
```



➤ **Task 5 SKATER and OTHERS**

- Implement SKATER for your data
- Save the minimum spanning tree after pruning and load it as a connectivity graph for visualization
- Explore the other settings provided in SKATER that haven't been discussed in the Tutorial (Or explore the REDCAP method), record any results or insights that come to your mind.

*Answers- **SKATER** (k=5) yielded a **BSS/TSS of 0.63**, pruning a minimum spanning tree to form regions. Clusters were less compact than SCHC but offered unique spatial groupings. The saved tree visualized connectivity between clusters.*

Skater Settings

Input:

Select Variables

ALAND

AWATER

NUM_N_veen

COORD_X

COORD_Y

Z

Z_LAG

Z_LAG_arch

LISA_I

LISA_CL

LISA_P

I

Water_lag

Select Spatial Weights:

tl_2023_27_bg

Parameters:

Number of Regions:

5

Minimum Bound:

☐ ALAND

1

10%

Min Region Size:

Distance Function:

Euclidean

Transformation:

Standardize (Z)

Output:

Save Cluster in Field:

CL_skater

(Optional)

☐ Save Complete Spanning Tree

Run

Save Spanning Tree

Close

Summary

Number of clusters: 5

Weights: tl_2023_27_bg

Minimum region size: 1

Distance function: Euclidean

Transformation: Standardize (Z)

Cluster centers:

	ALAND	AWATER	COORD_X
C1	2.38136e+07	1.59093e+06	-93.2655
C2	1.64233e+08	1.33378e+07	-95.6362
C3	1.16816e+08	1.52724e+06	-95.235
C4	2.69521e+09	6.58771e+07	-94.6249
C5	0	2.96569e+09	-90.7869

The total sum of squares: 14115

Within-cluster sum of squares:

	Within cluster S.S.
C1	2535.83
C2	2606.53
C3	303.96
C4	107.396
C5	597.129

The total within-cluster sum of squares: 61

The between-cluster sum of squares: 7964.1

The ratio of between to total sum of squares:

Skater Settings

Input:

Select Variables

ALAND

AWATER

NUM_N_ueen

COORD_X

COORD_Y

Z

Z_LAG

Z_LAG_arch

LISA_I

LISA_CL

LISA_P

I

Water_lag

Select Spatial Weights:

tl_2023_27_bg

Parameters:

Number of Regions:

5

Minimum Bound:

☐

ALAND

1

10%

Min Region Size:

Distance Function:

Euclidean

Transformation:

Standardize (Z)

Output:

Save Cluster in Field:

CL_skater

(Optional)

☐

Save Complete Spanning Tree

Run

Save Spanning Tree

Close

Summary

Number of clusters: 5

Weights: tl_2023_27_bg

Minimum region size:

Distance function: Euclidean

Transformation: Standardize (Z)

Cluster centers:

	ALAND	AWATER	COORD_X
C1	2.38136e+07	1.59093e+06	-93.2655
C2	1.64233e+08	1.33378e+07	-95.6362
C3	1.16816e+08	1.52724e+06	-95.235
C4	2.69521e+09	6.58771e+07	-94.6249
C5	0	2.96569e+09	-90.7869

The total sum of squares: 14115

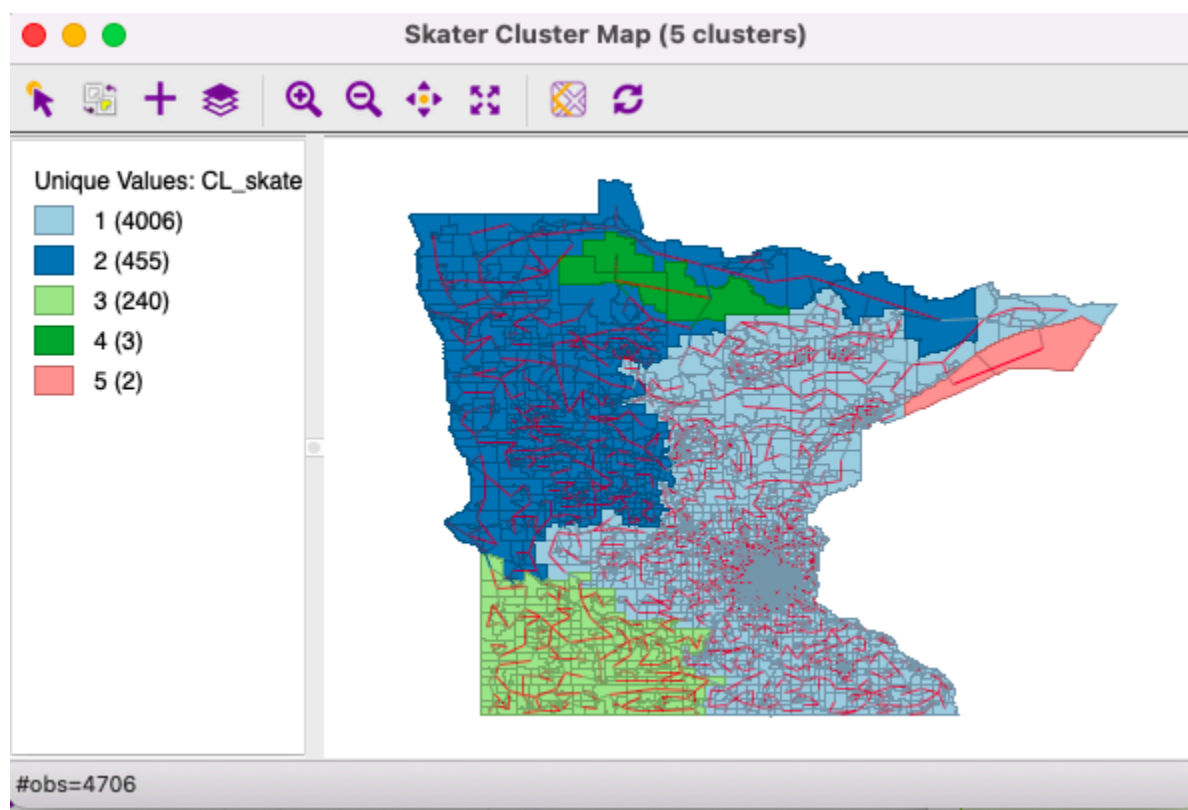
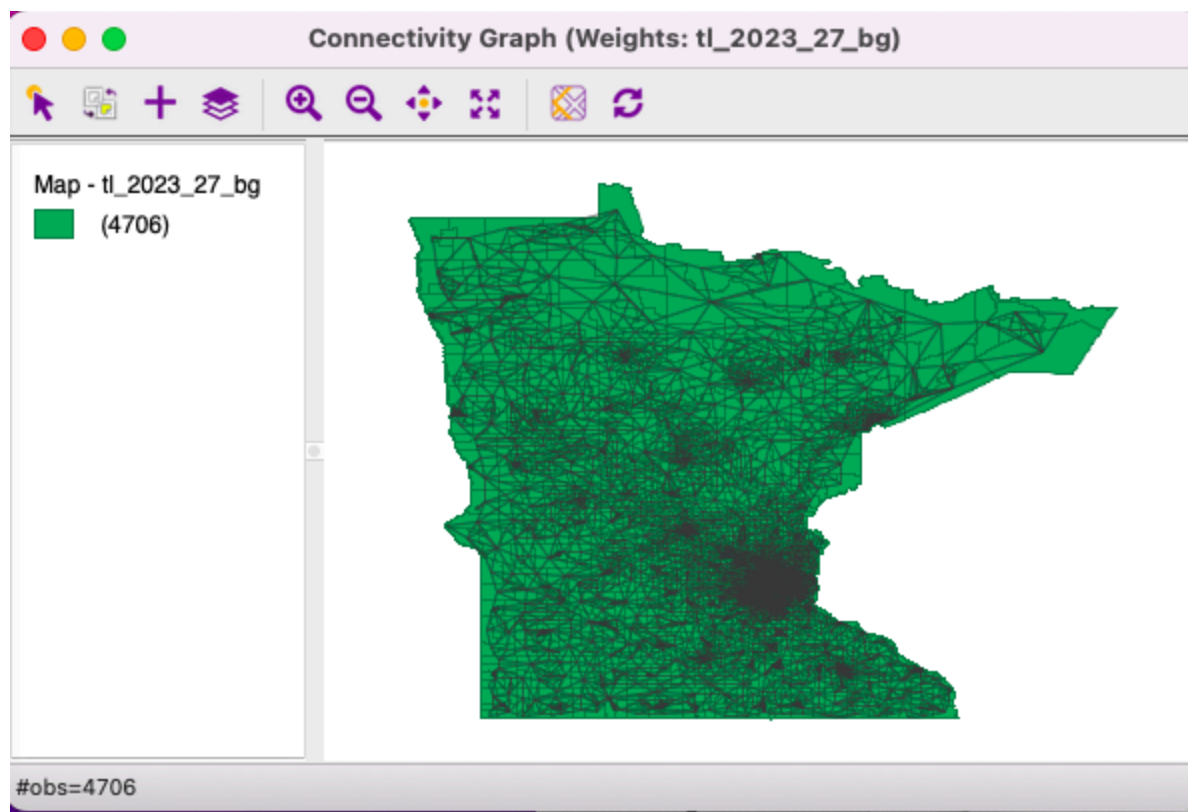
Within-cluster sum of squares:

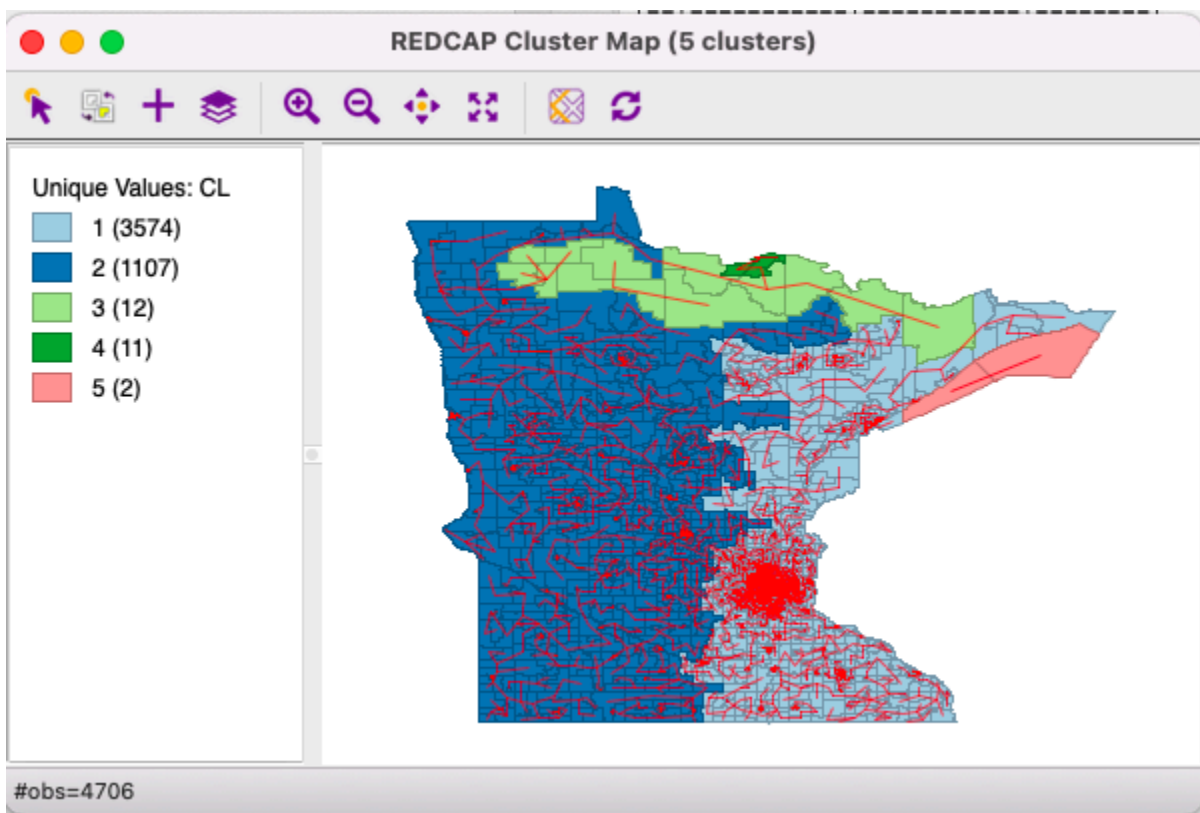
	Within cluster S.S.
C1	2535.83
C2	2606.53
C3	303.96
C4	107.396
C5	597.129

The total within-cluster sum of squares: 61.

The between-cluster sum of squares: 7964.1

The ratio of between to total sum of squares:





REDCAP Settings

Input:

Select Variables

Skater

ALAND

AWATER

NUM_N_ueen

COORD_X

COORD_Y

Z

Z_LAG

Z_LAG_arch

LISA_I

LISA_CL

LISA_P

I

Select Spatial Weights:

tl_2023_27_bg

Parameters:

Number of Regions:

5

Method:

FullOrder-WardLinkage

Minimum Bound:

☐ Skater

1

10%

Min Region Size:

Distance Function:

Euclidean

Transformation:

Standardize (Z)

Output:

Save Cluster in Field:

CL

(Optional)

☐ Save Complete Spanning Tree

Run

Save Spanning Tree

Close

Summary

Weights: tl_2023_27_bg

Method: FullOrder-WardLinkage

Minimum region size:

Save cluster in field: CL

Transformation: Standardize (Z)

Distance function: Euclidean

Cluster centers:

	ALAND	AWATER	COORD_X
C1	1.77629e+07	927451	-93.1426
C2	1.0916e+08	7.34235e+06	-95.0557
C3	1.77139e+09	1.23572e+08	-94.231
C4	6.02967e+07	7.32181e+06	-93.404
C5	0	2.96569e+09	-90.7869

The total sum of squares: 14115

Within-cluster sum of squares:

	Within cluster S.S.
C1	1539.09
C2	2648.38
C3	628.52
C4	7.38595
C5	597.129

The total within-cluster sum of squares: 5420.5

The between-cluster sum of squares: 8694.5

The ratio of between to total sum of squares: 0.61