GeoDa Workshop
Part 2

Luc Anselin

GeoDa Center
School of Geographical Sciences and Urban Planning
Arizona State University

after July 1, 2016
Center for Spatial Data Science
University of Chicago

Acknowledgments

NSF OCI-1047916

AHRQ 1R01HS021752-01A1

spatial weights

spatial autocorrelation

space-time exploration

averages tool (treatment effect analysis)

spatial regression

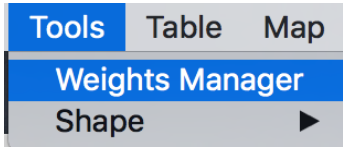
Spatial Weights

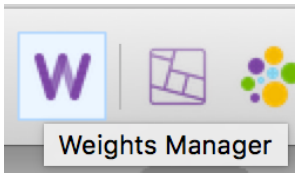
creating weights

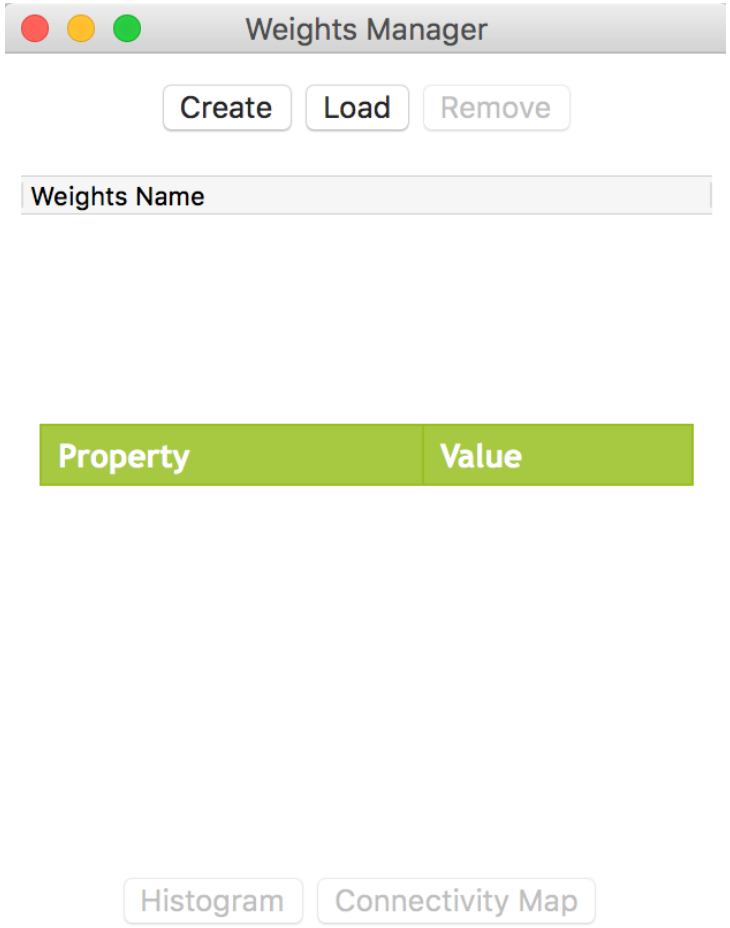
weights properties

project file

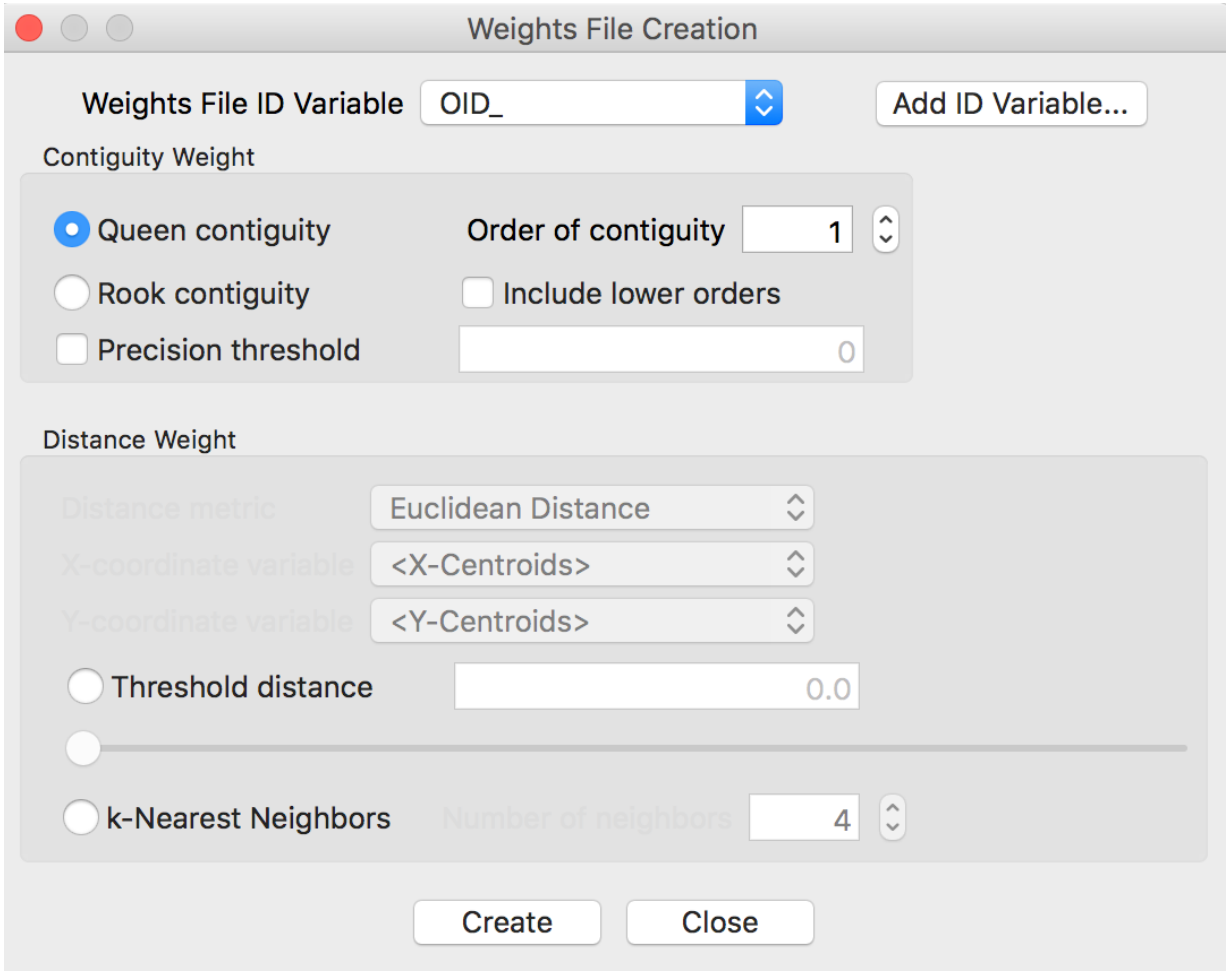
Creating Weights



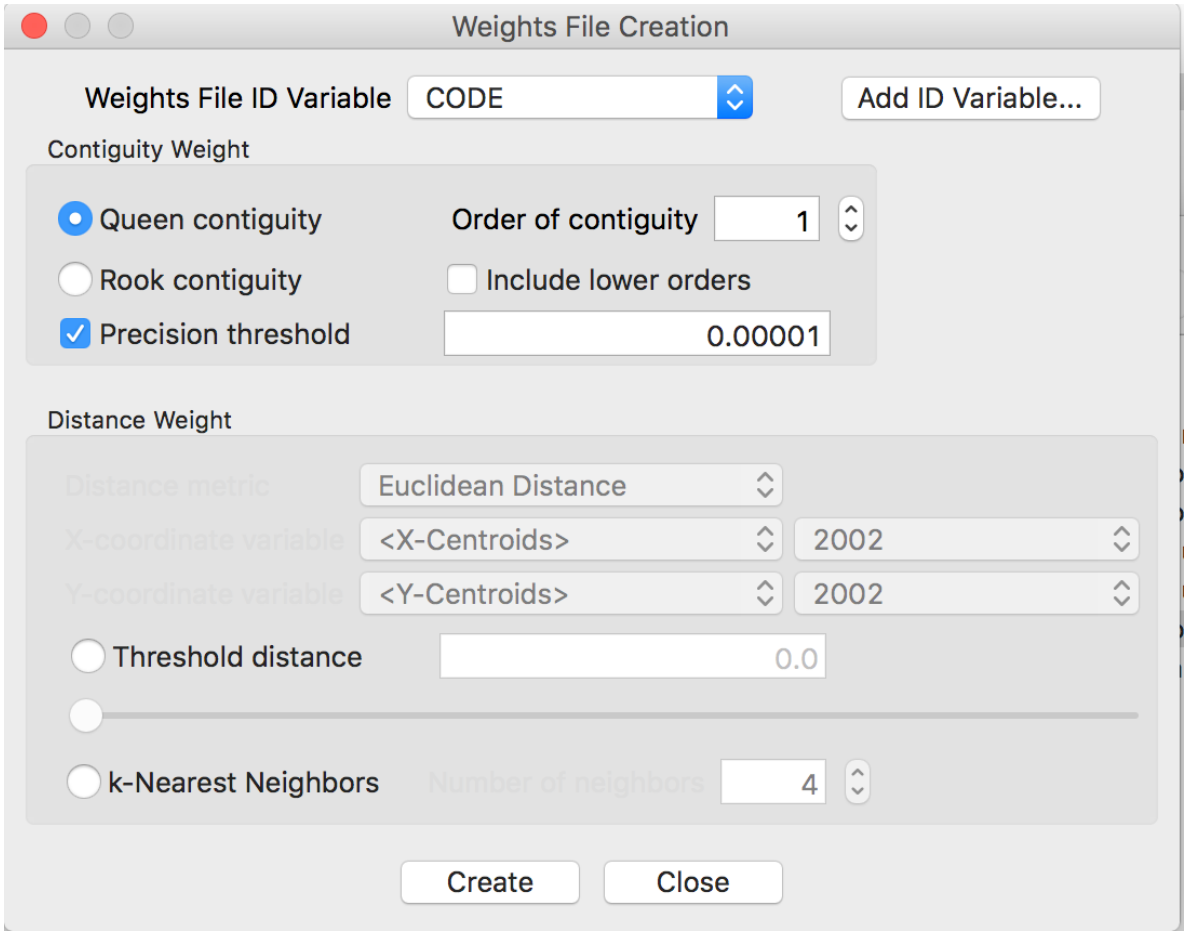




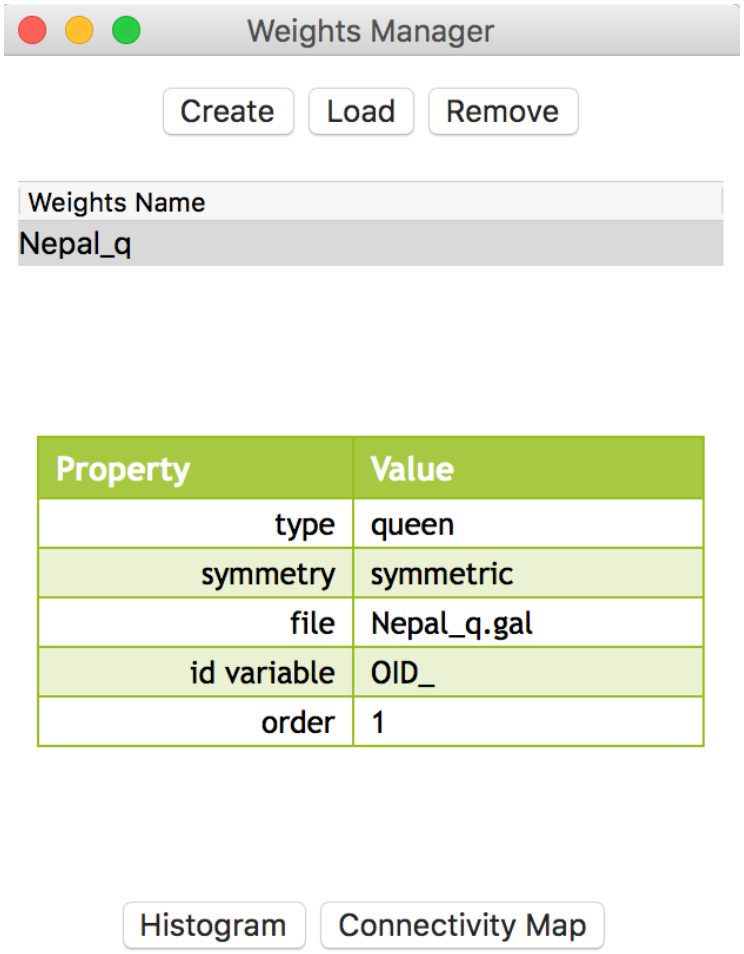
weights manager



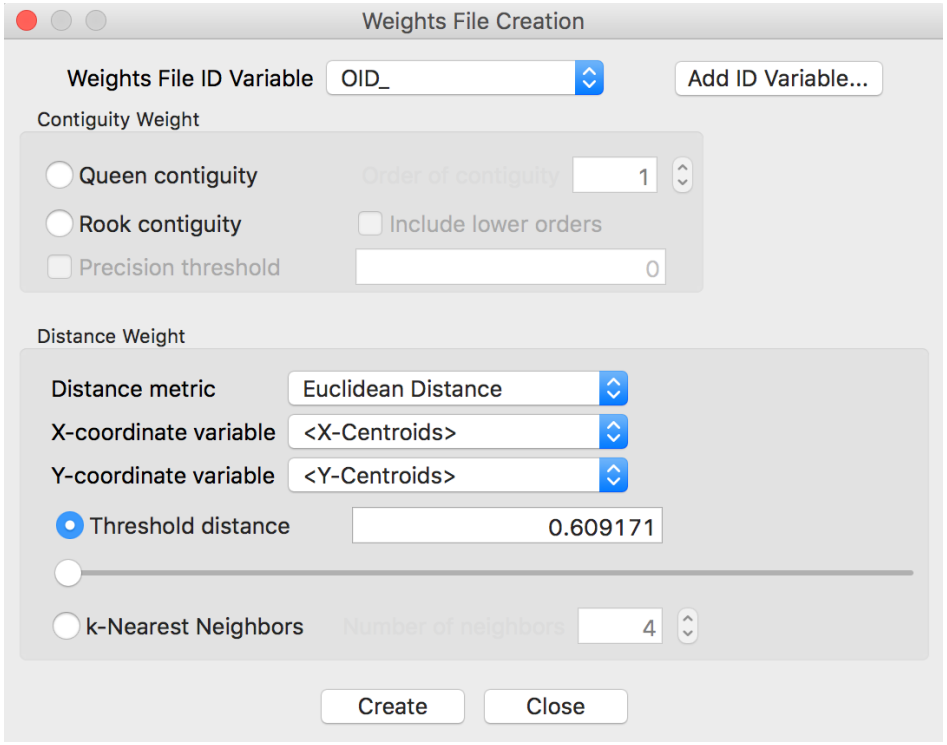
contiguity weights

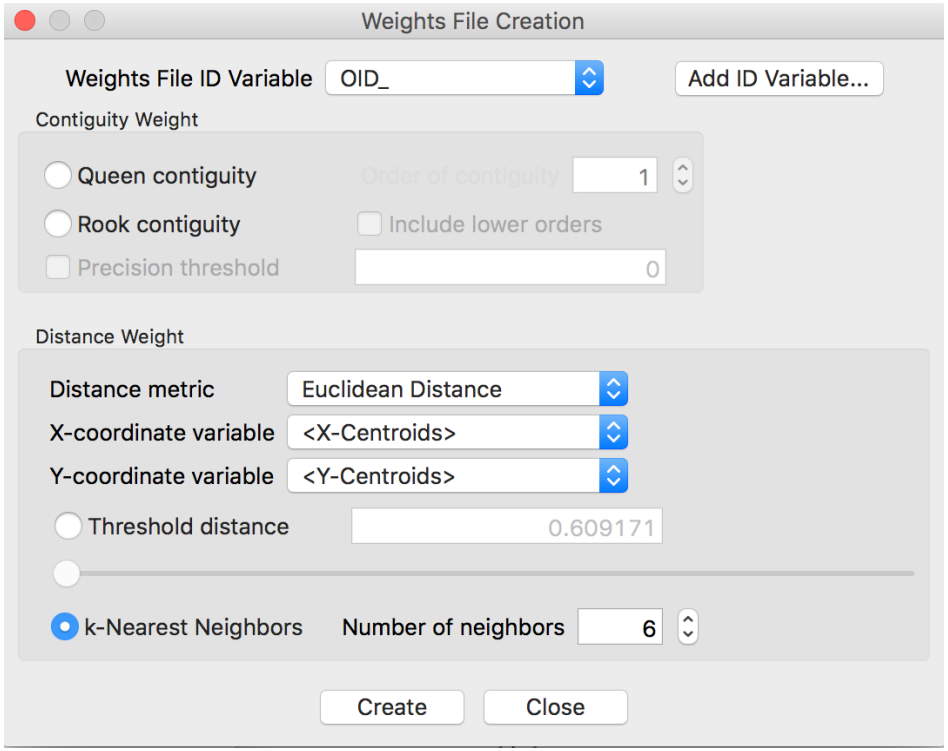


precision threshold (NYC example)

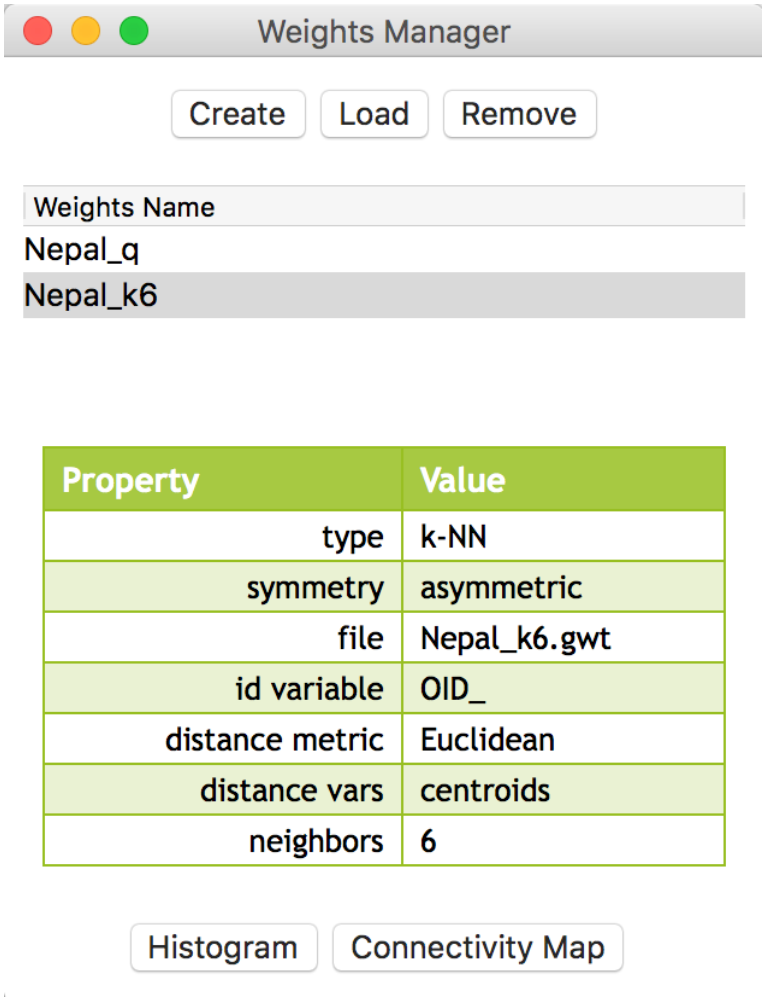


weights in weights manager



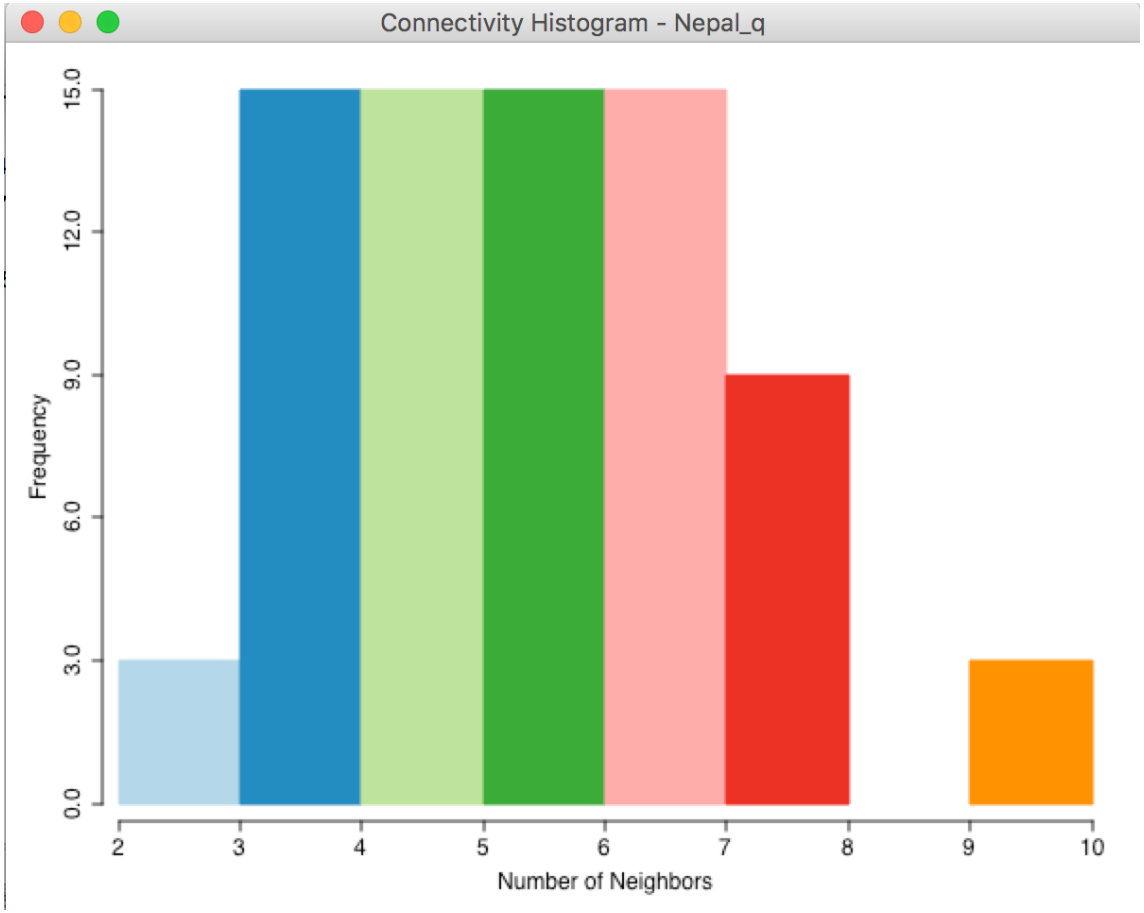


distance-based weights

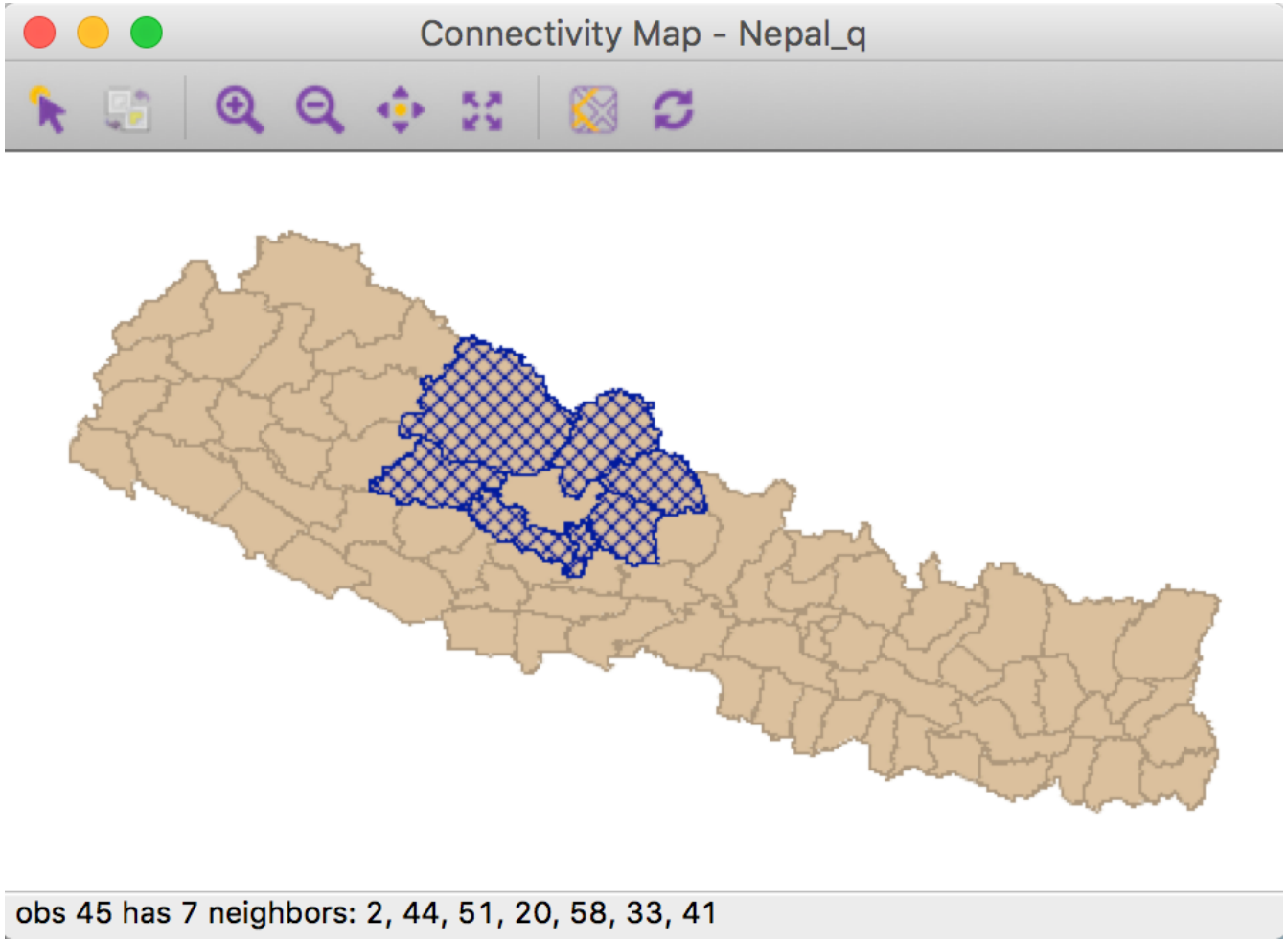


multiple weights in weights manager

Weights Properties

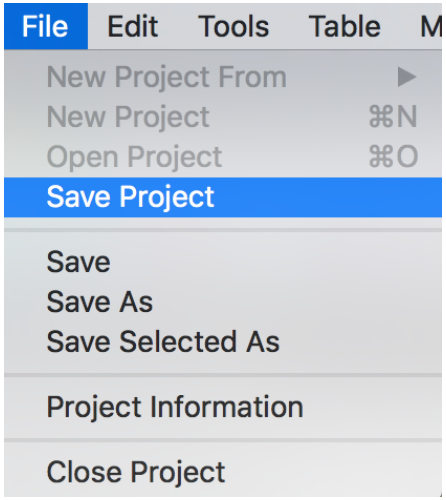


connectivity histogram

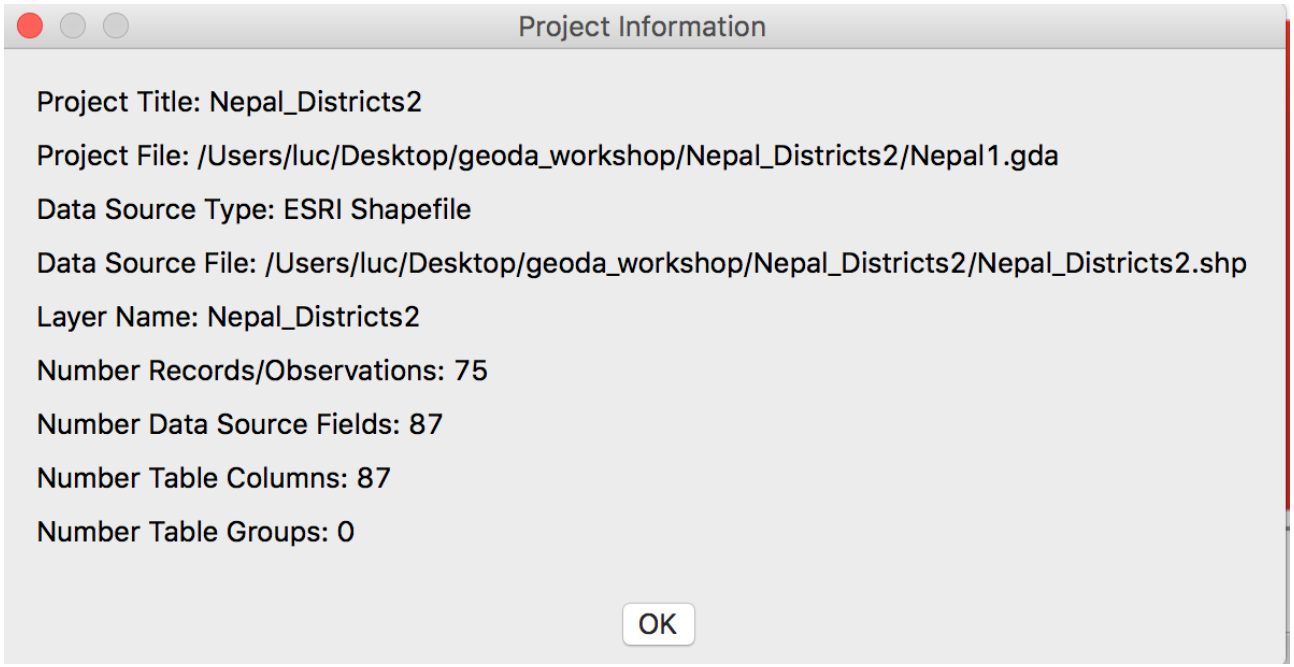


connectivity map

Project File



creating a project file



project information

```
<weights_entries>
    <weights>
        <title>Nepal_q</title>
        <default/>
        <meta_info>
            <weights_type>queen</weights_type>
            <order>1</order>
            <inc_lower_orders>true</inc_lower_orders>
            <path>Nepal_q.gal</path>
            <id_variable>OID_</id_variable>
            <symmetry>symmetric</symmetry>
        </meta_info>
    </weights>
    <weights>
        <title>Nepal_k6</title>
        <meta_info>
            <weights_type>knn</weights_type>
            <dist_metric>euclidean</dist_metric>
            <dist_units>mile</dist_units>
            <dist_values>centroids</dist_values>
            <num_neighbors>6</num_neighbors>
            <path>Nepal_k6.gwt</path>
            <id_variable>OID_</id_variable>
            <symmetry>asymmetric</symmetry>
        </meta_info>
    </weights>
</weights_entries>
```

weights entries in project file

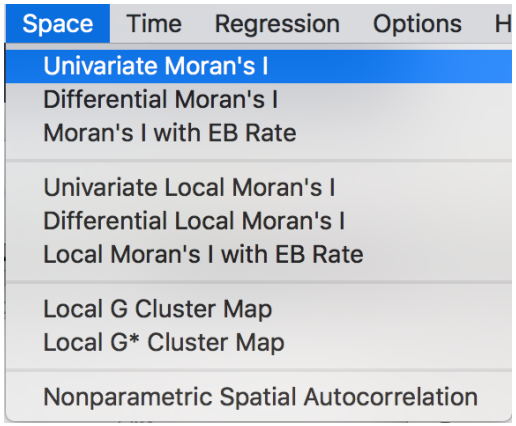
Spatial Autocorrelation

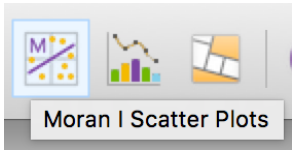
Moran scatter plot

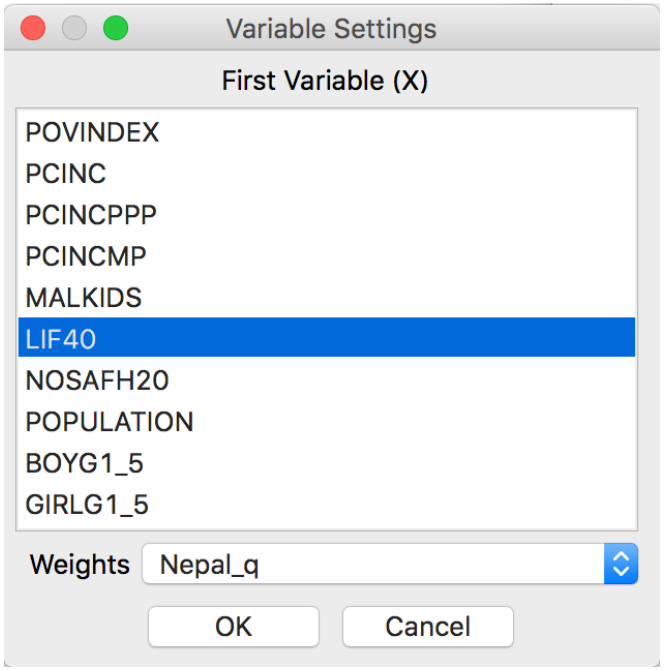
spatial correlogram

local spatial autocorrelation

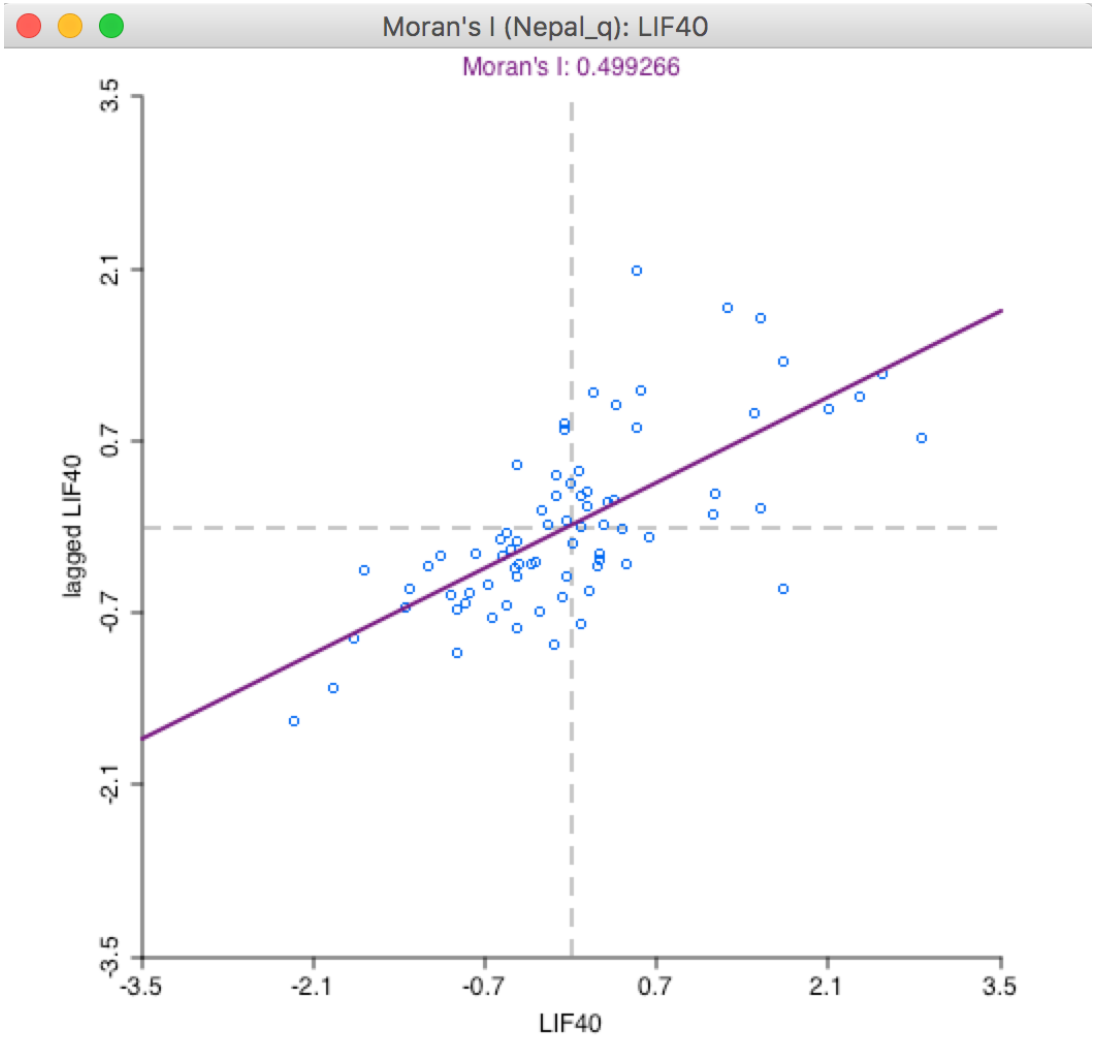
Moran Scatter Plot



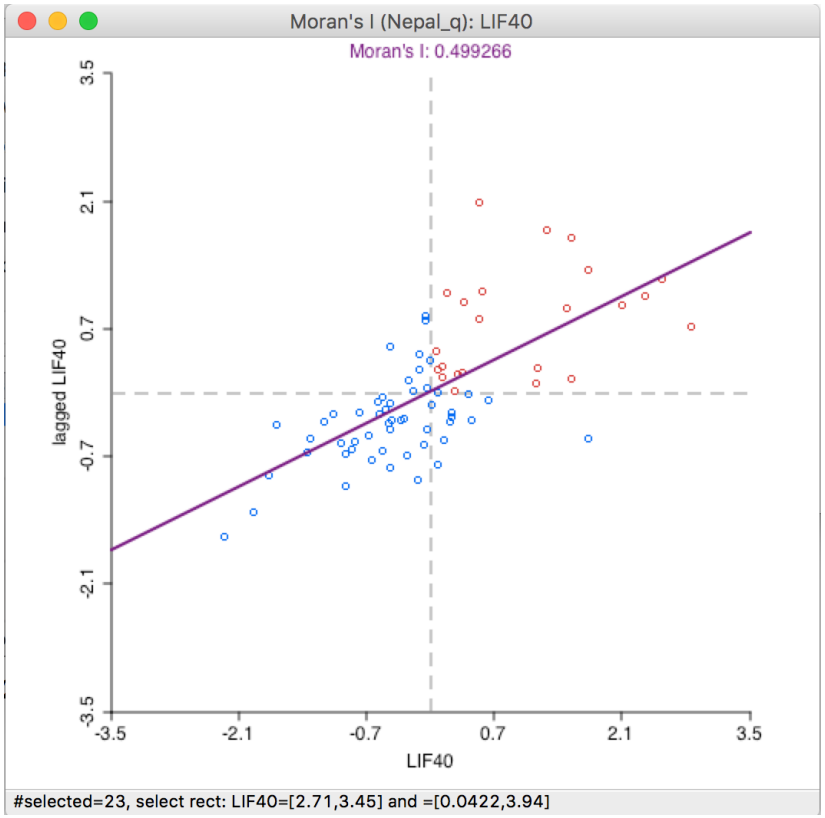


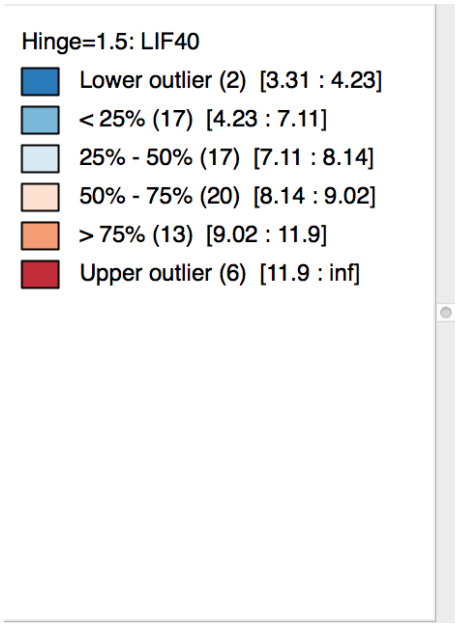


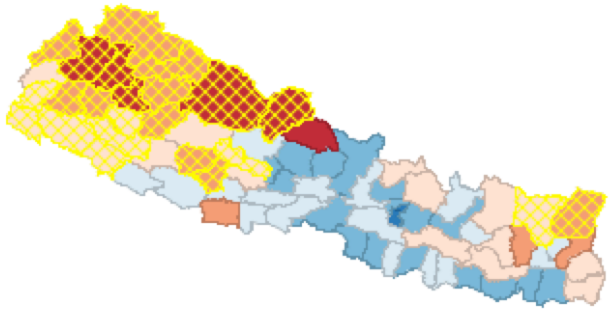
moran scatter plot setup



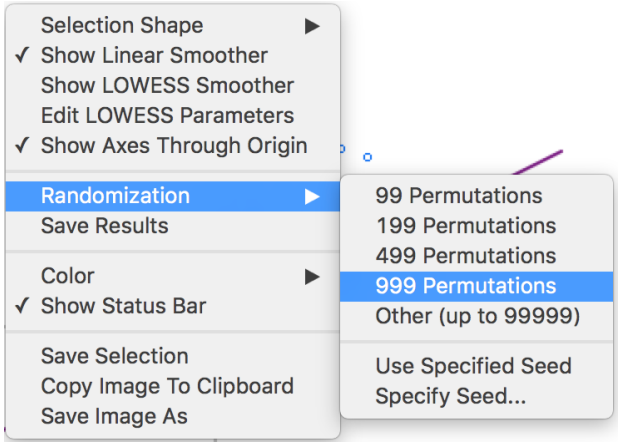
Moran scatter plot

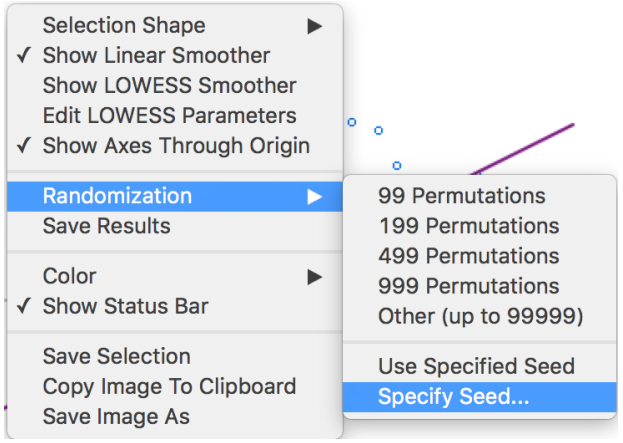


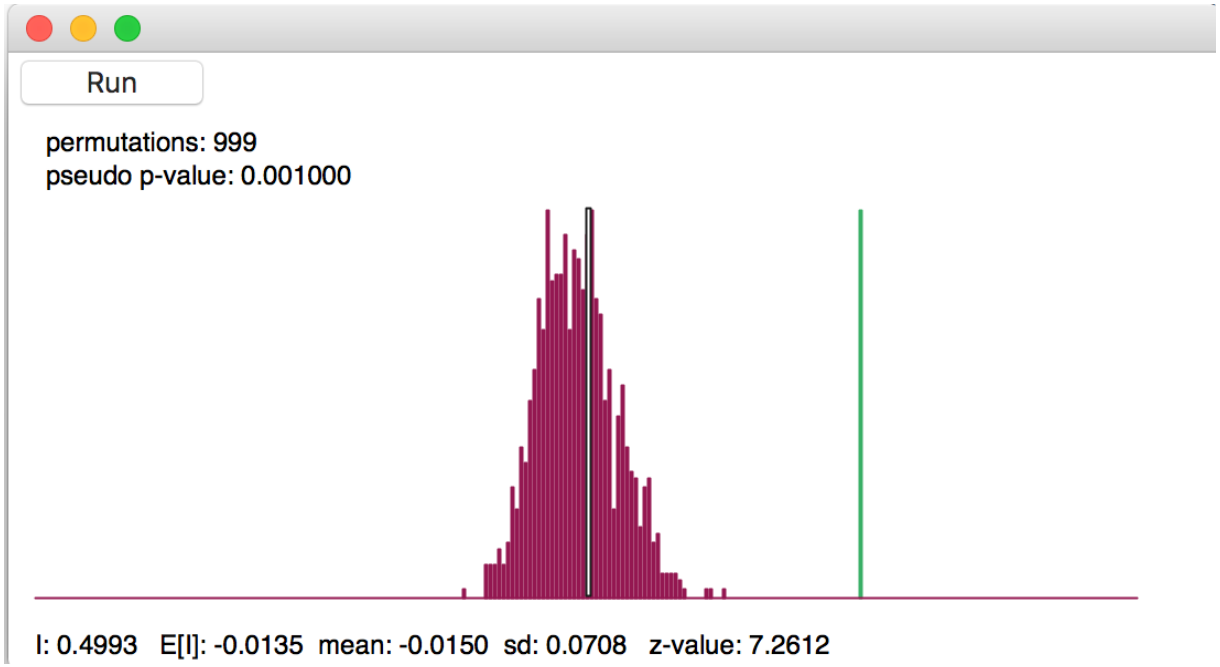




Moran scatter plot, high-high locations

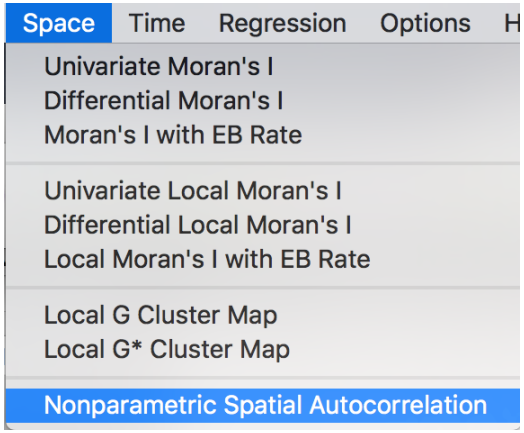


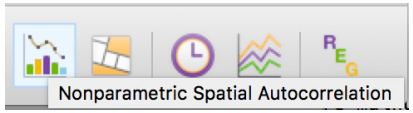


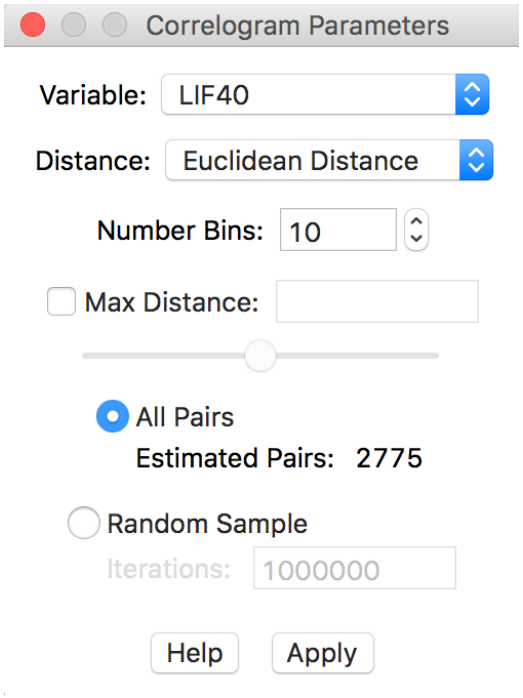


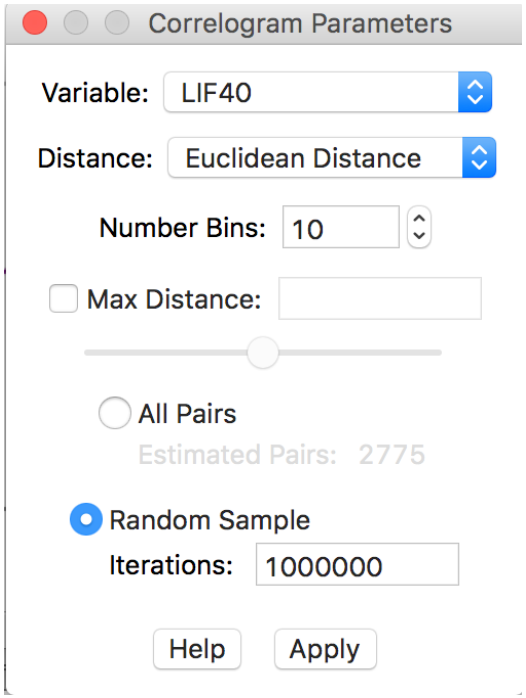
permutation inference

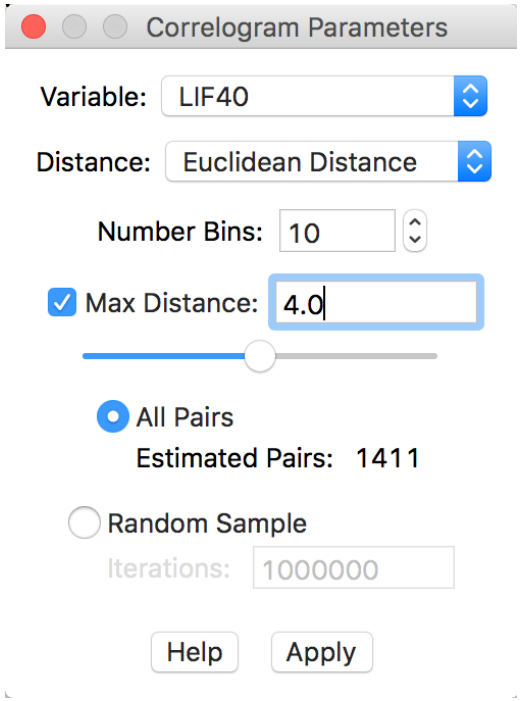
Spatial Correlogram



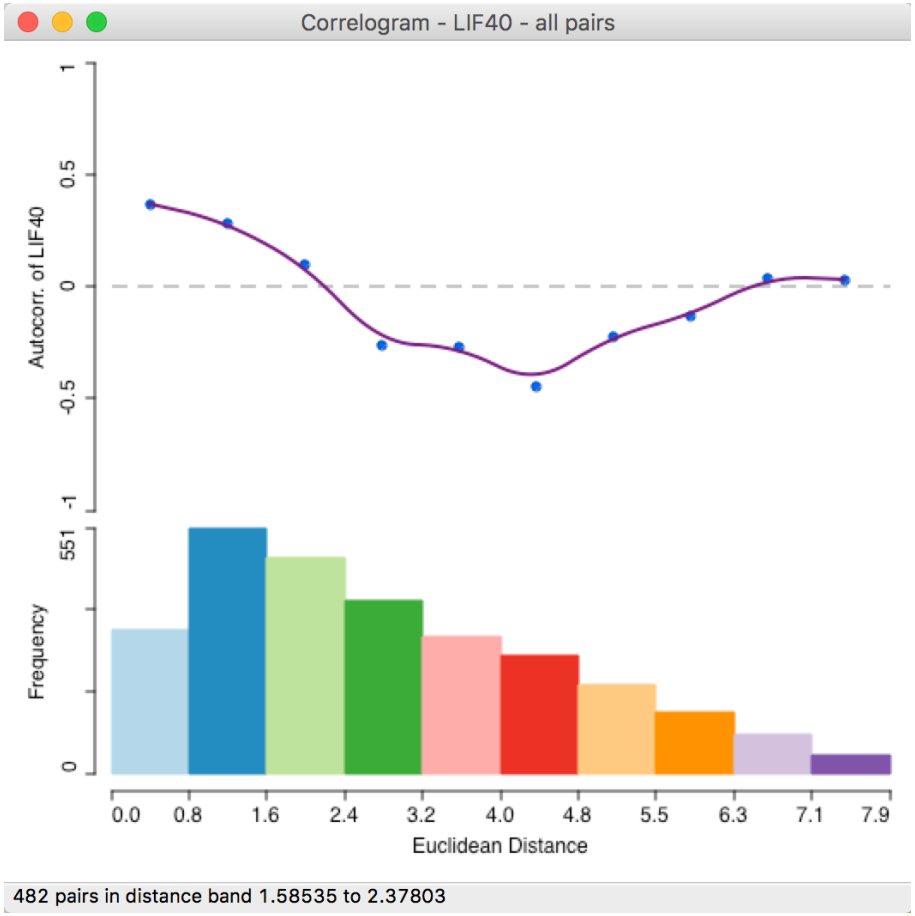


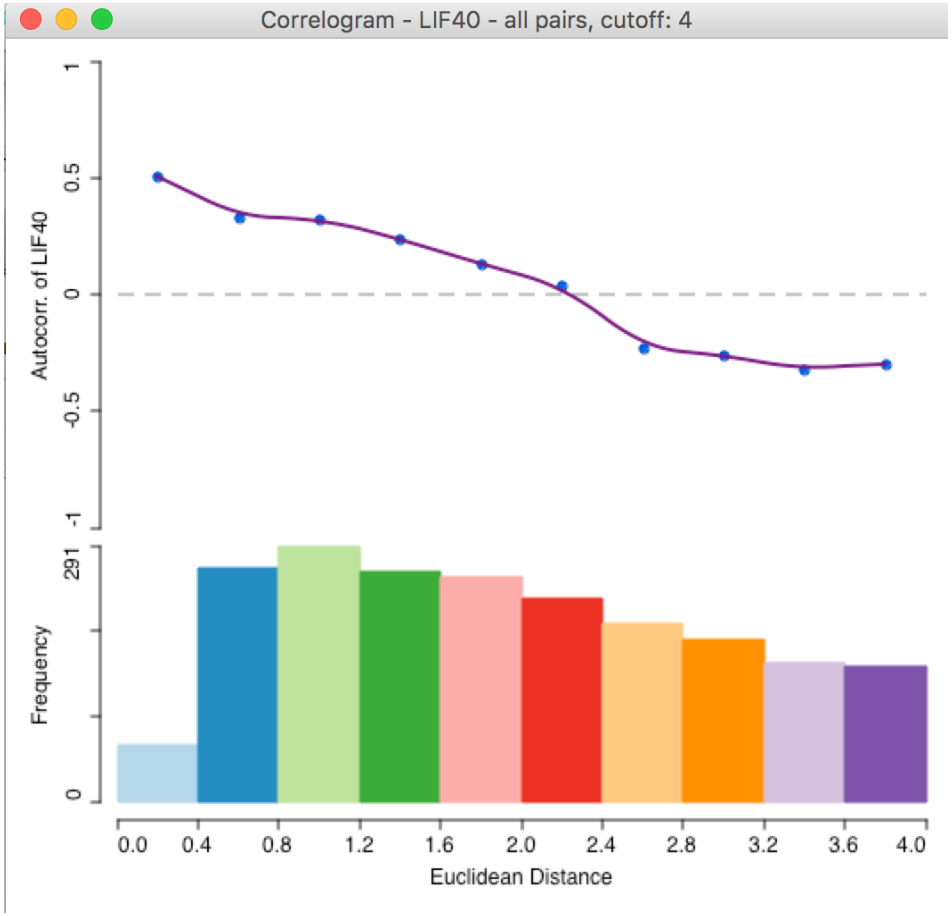






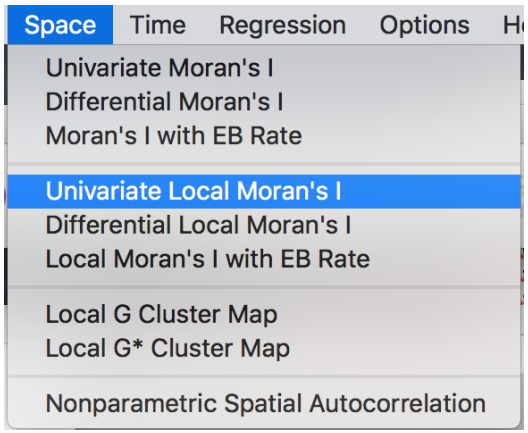
spatial correlogram setup

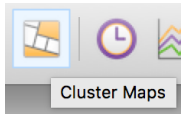


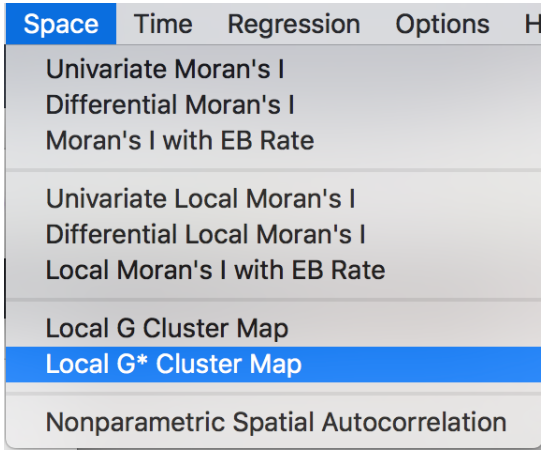


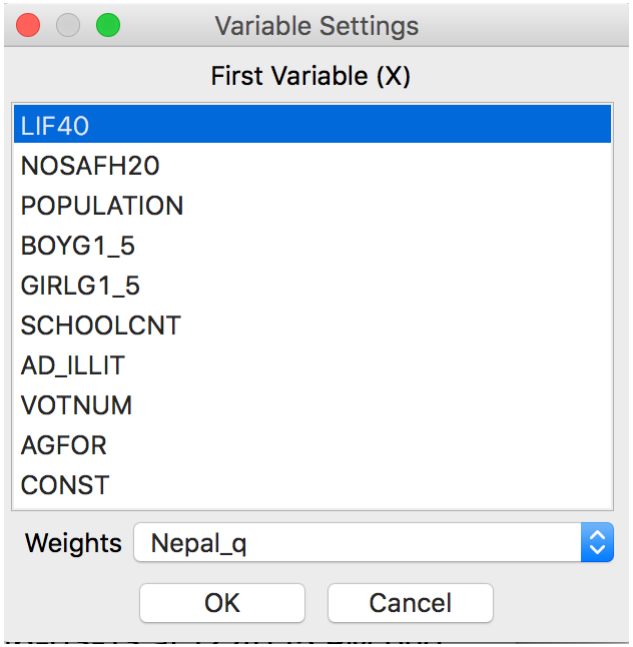
spatial correlogram

Local Spatial Autocorrelation

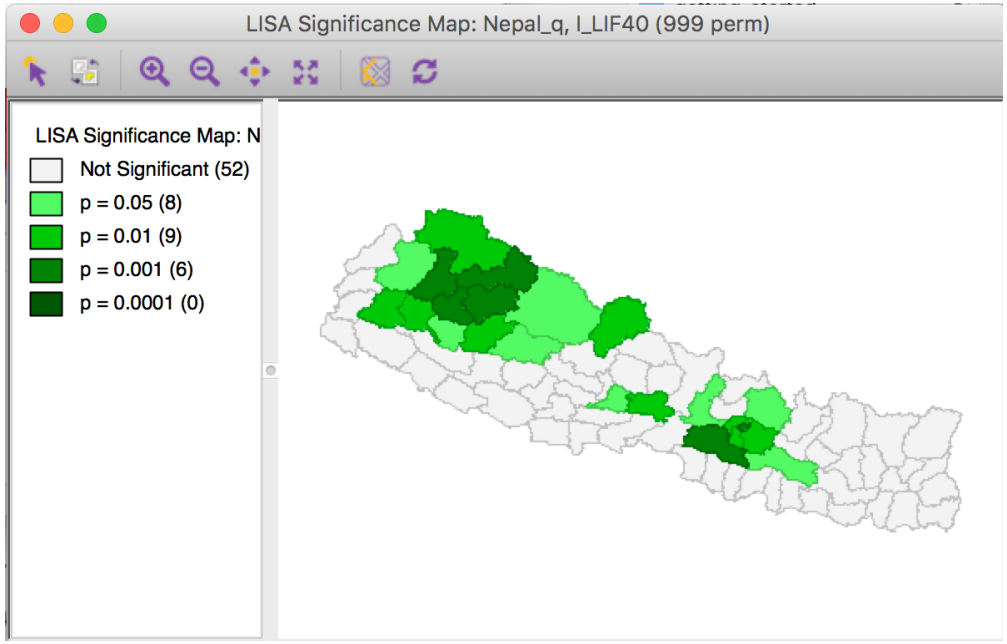


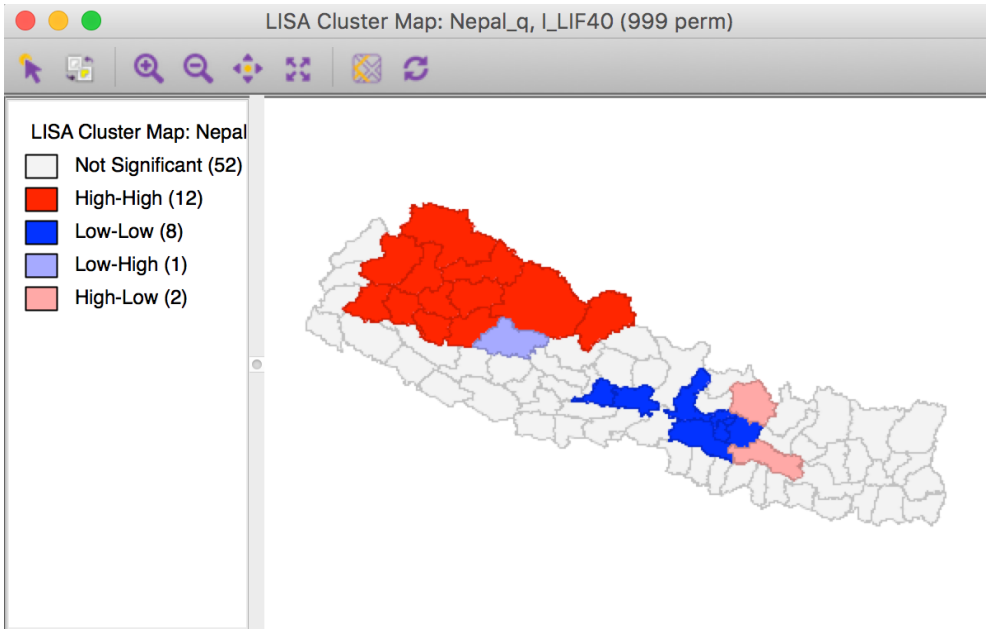




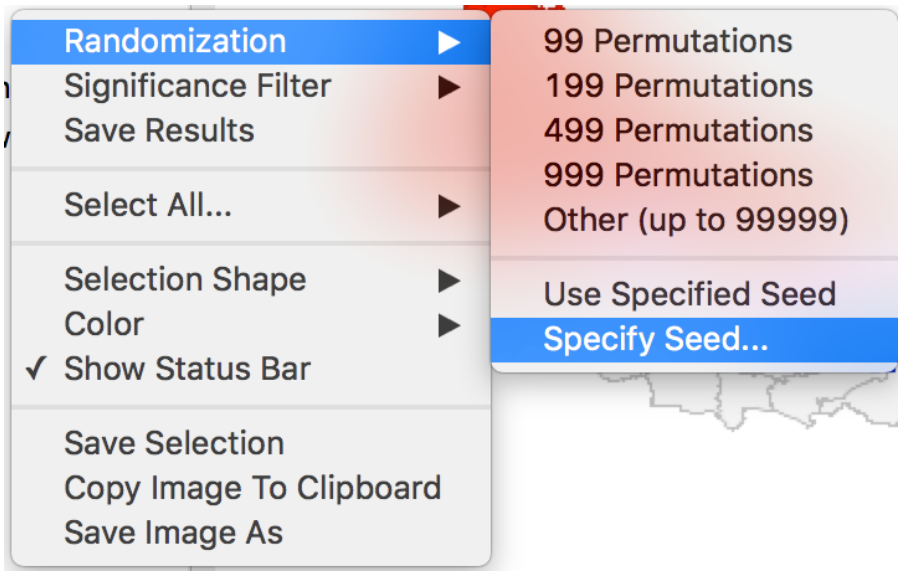


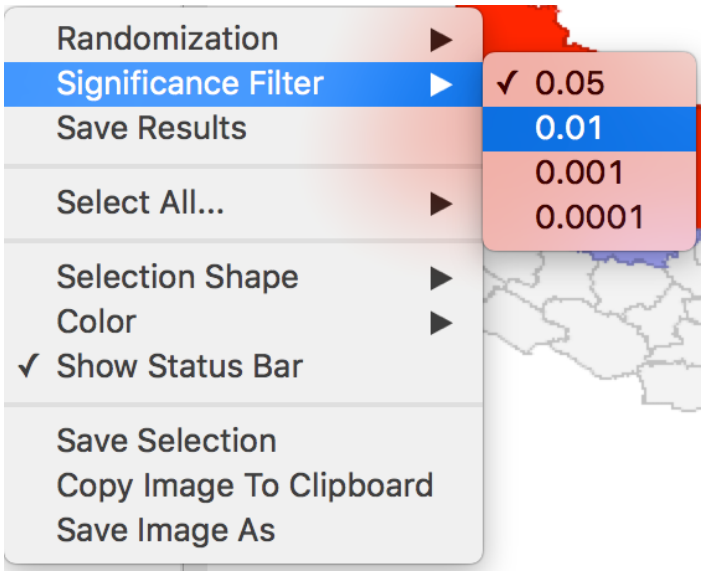
local spatial autocorrelation setup



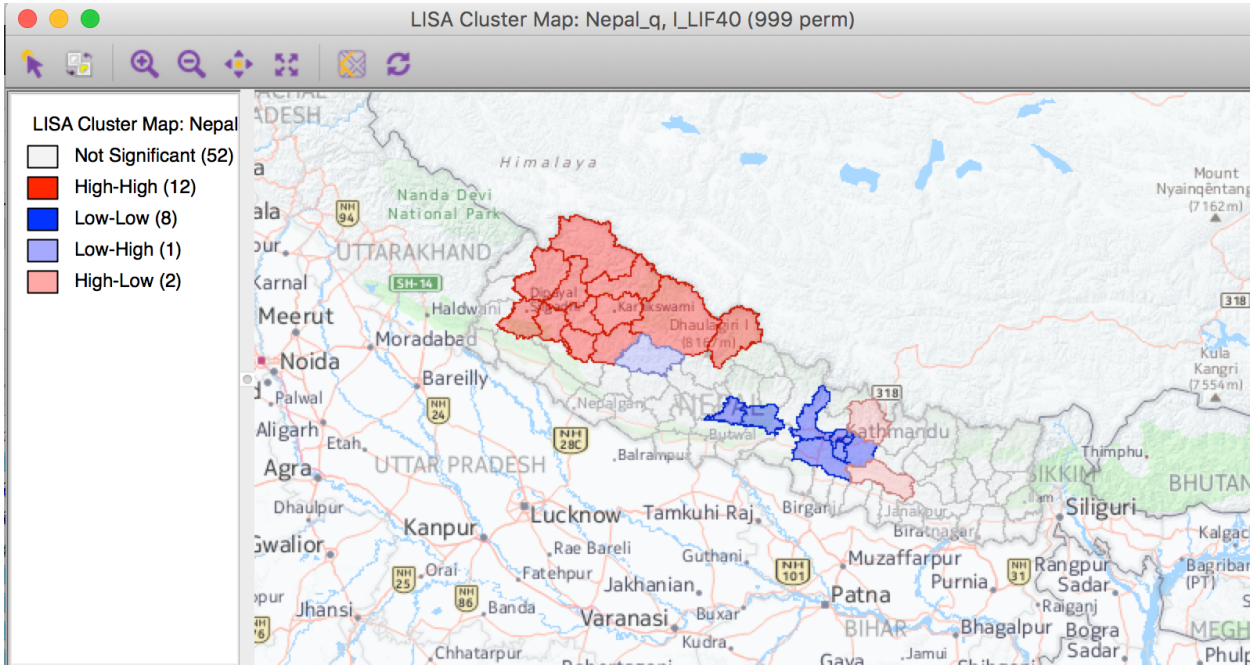


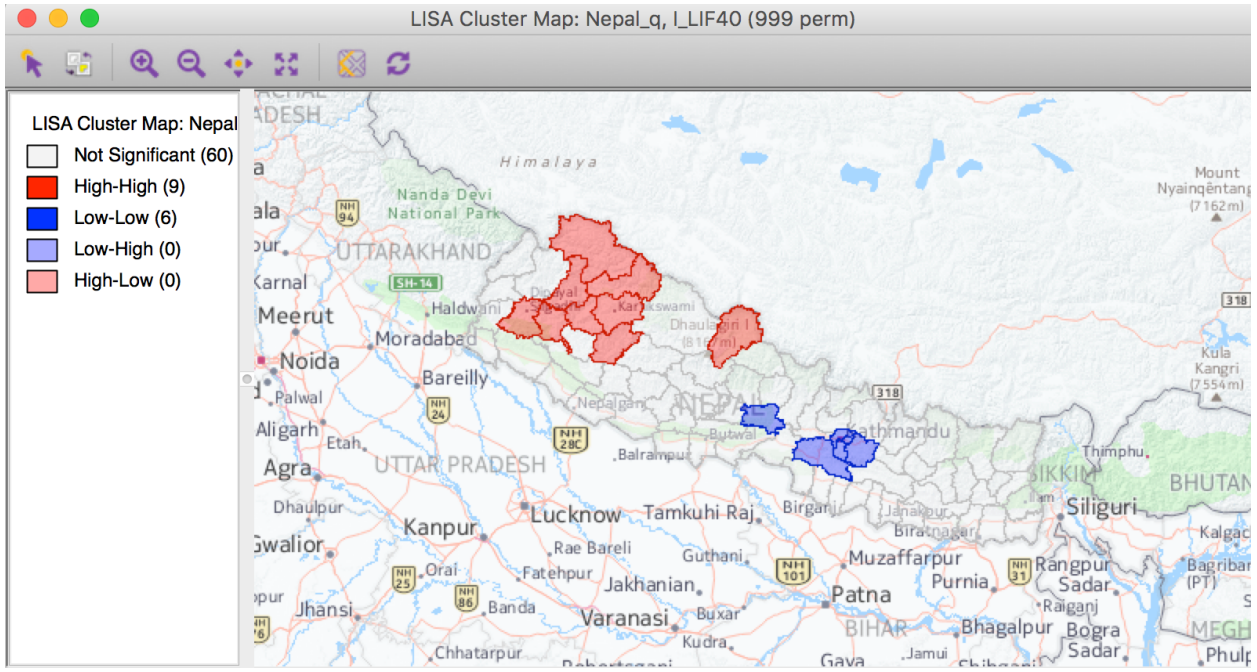
local moran significance and cluster map



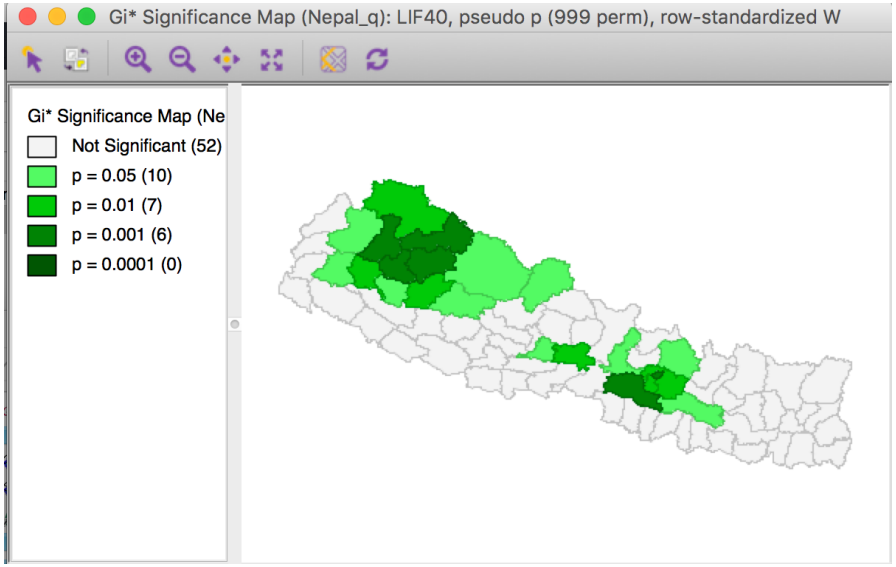


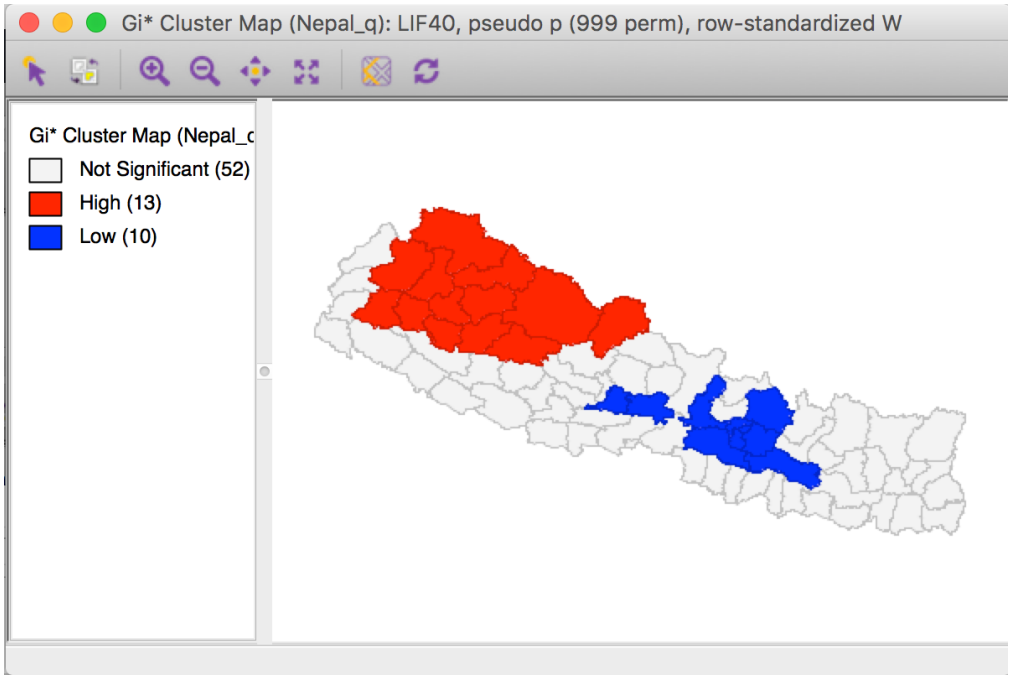
local spatial autocorrelation options



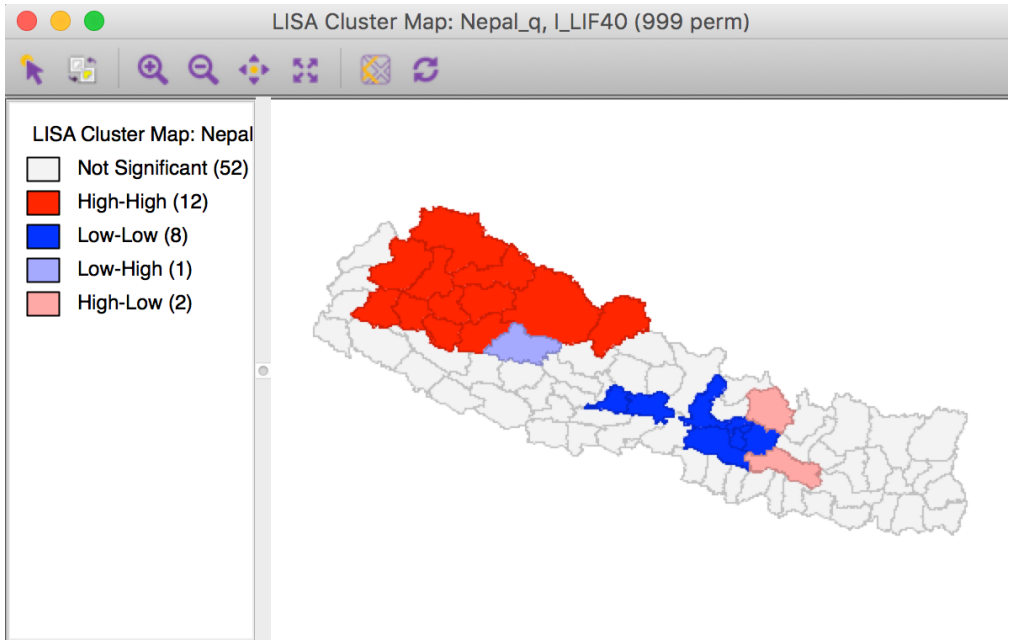


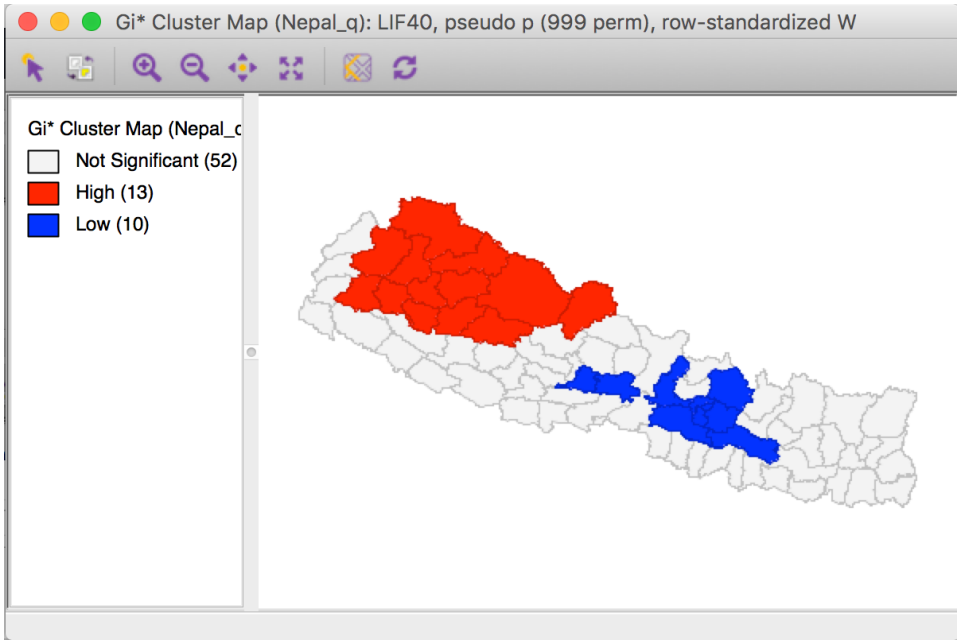
base map included, p=0.05 and 0.01





local G* significance and cluster map





local moran vs local G*

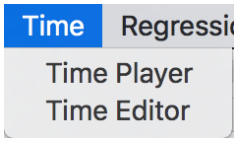
Space-Time Exploration

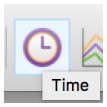
time editor

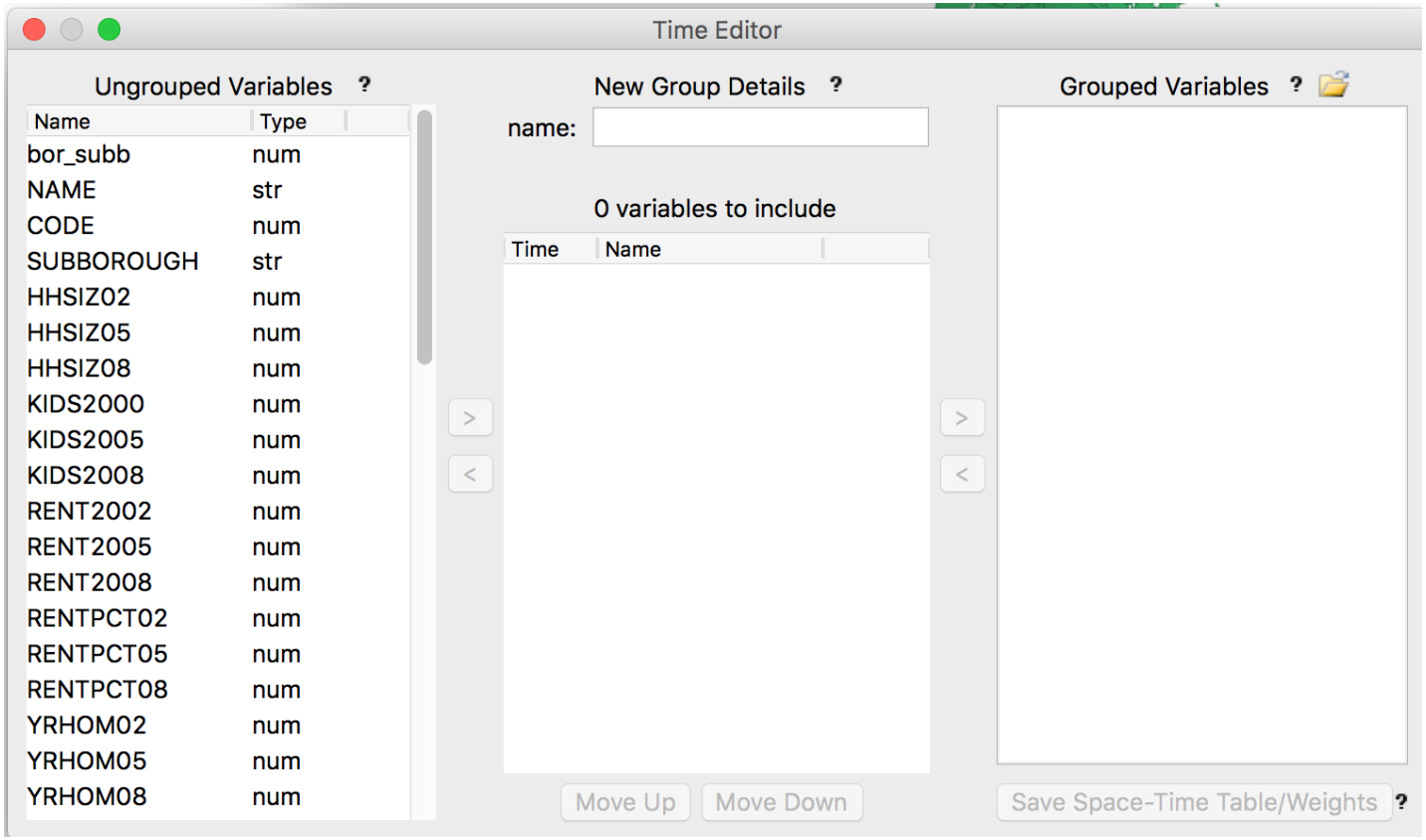
time player

differential spatial autocorrelation

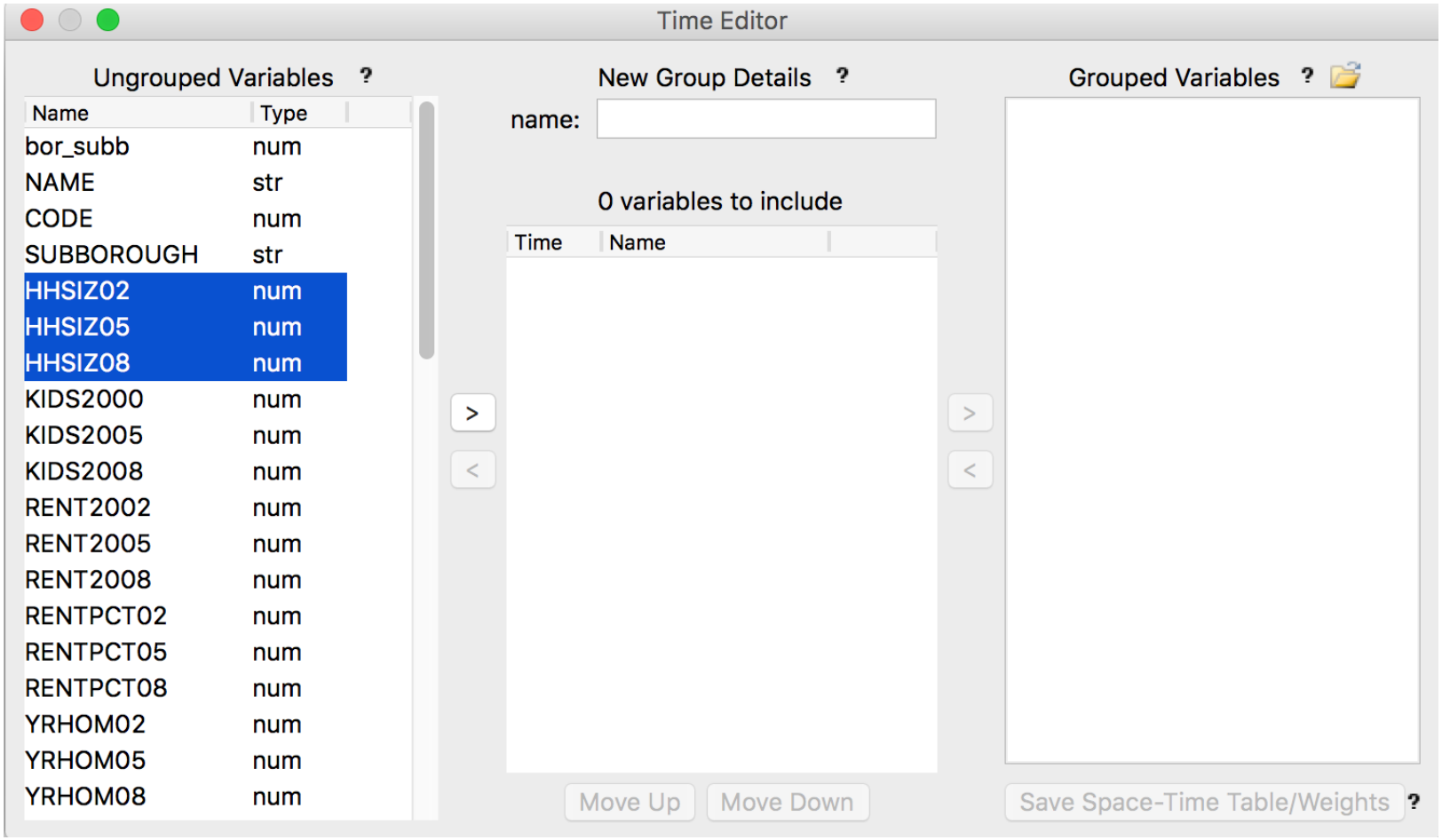
Time Editor



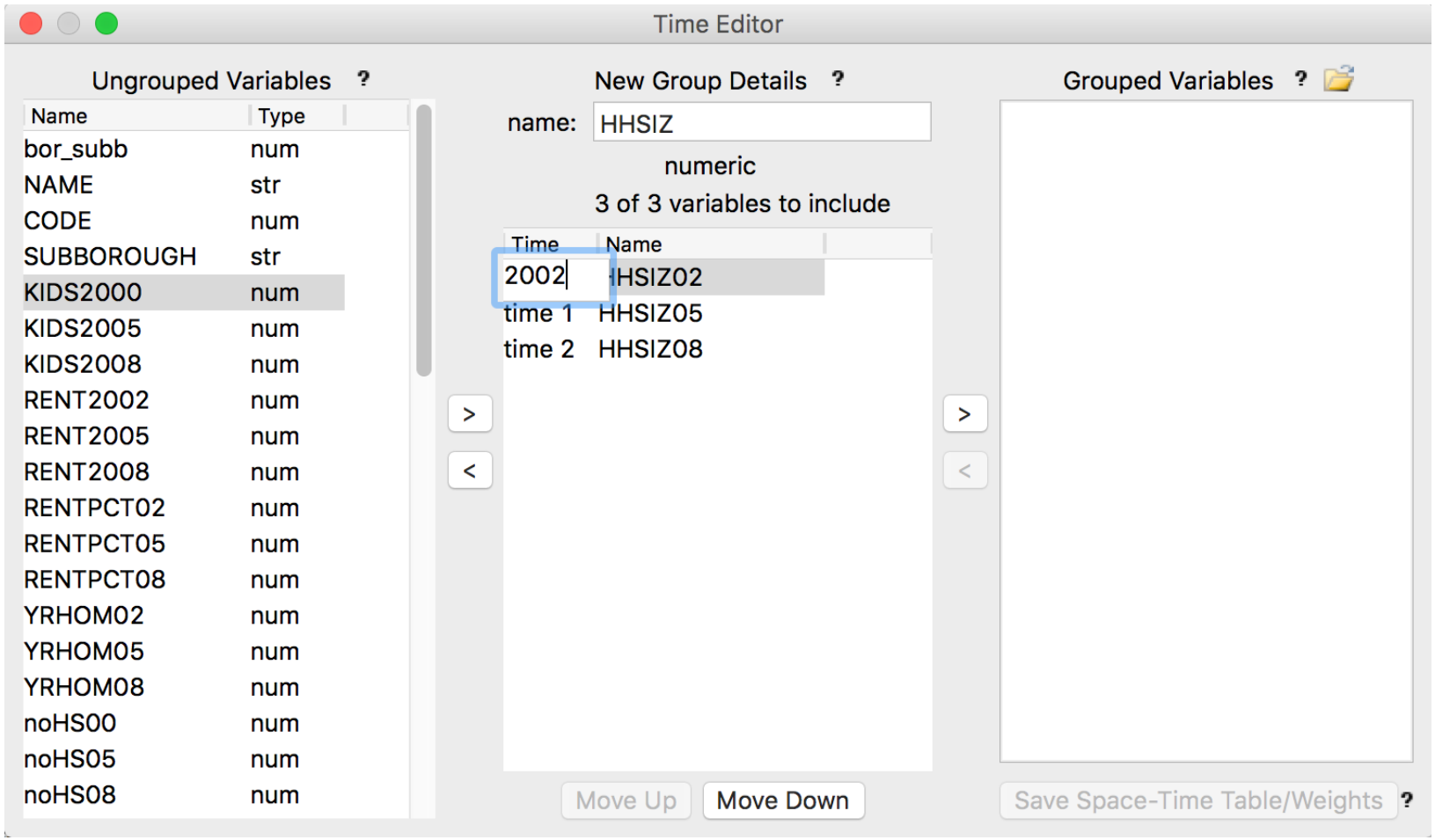




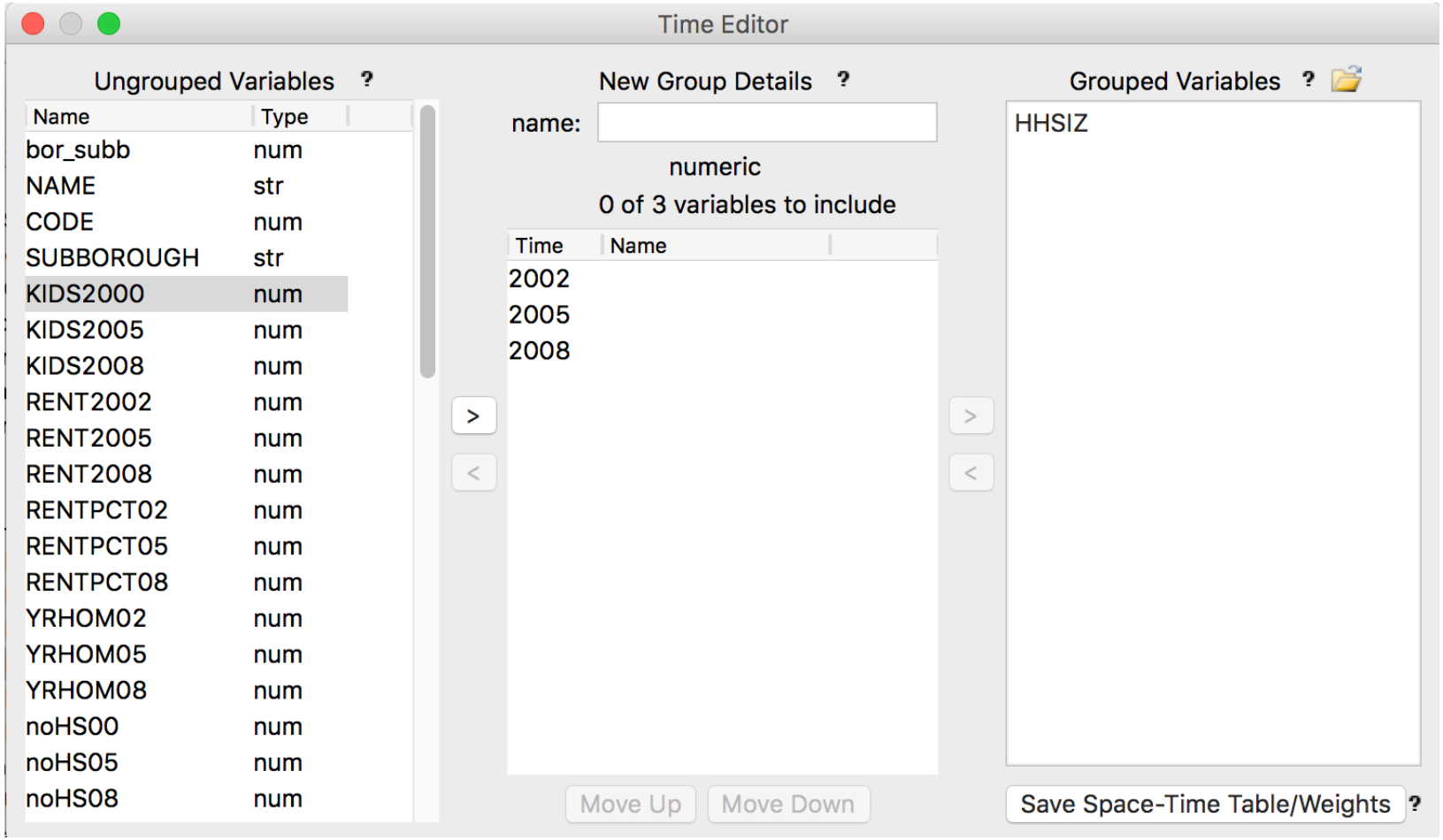
time editor setup



select variables to be grouped



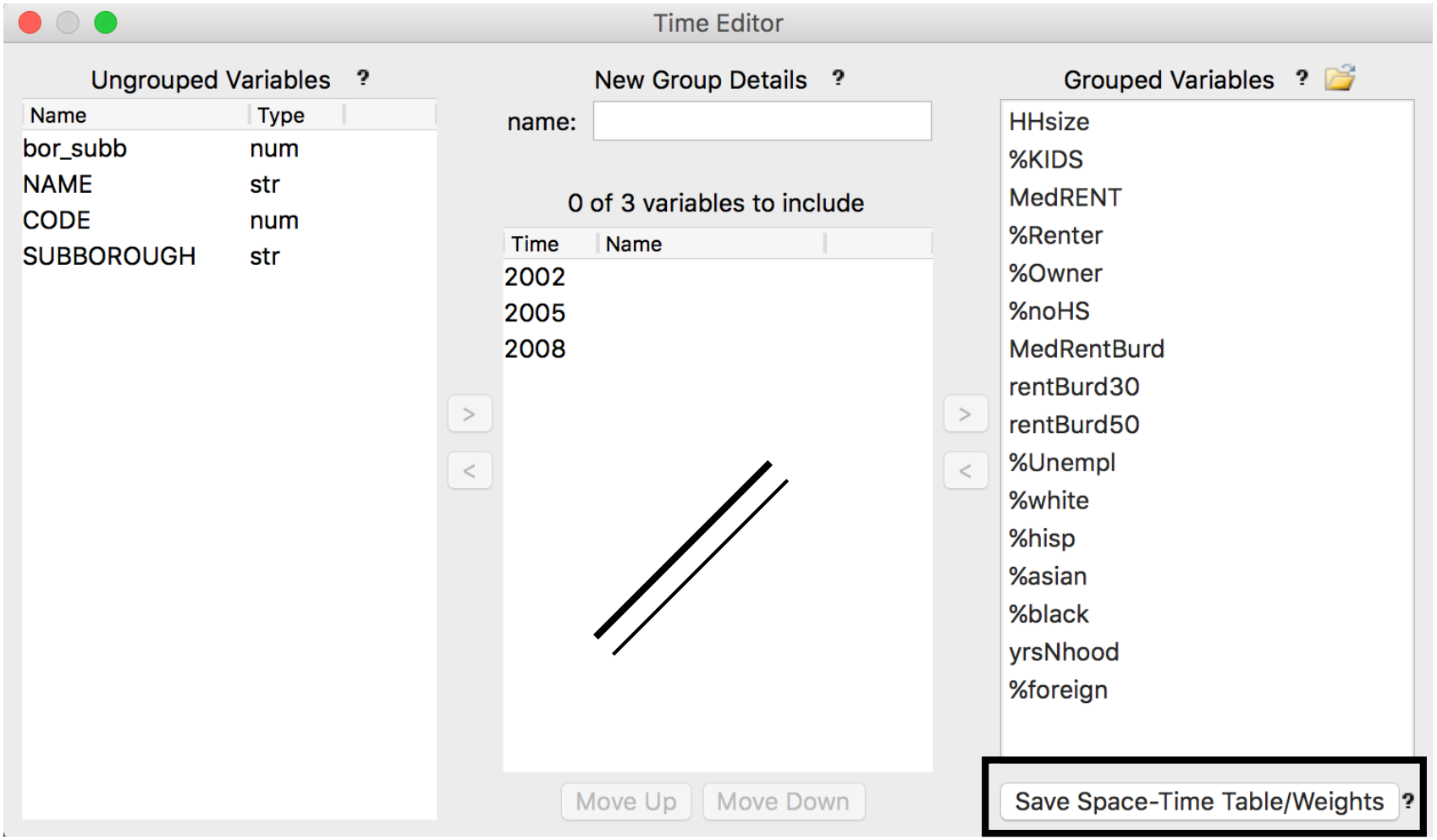
edit variable name and time labels



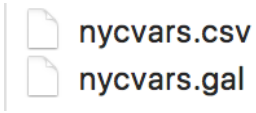
grouped variable

```
        <variable_order>
            <time_ids>
                <id>2002</id>
                <id>2005</id>
                <id>2008</id>
            </time_ids>
            <var>bor_subb</var>
            <var>NAME</var>
            <var>CODE</var>
            <var>SUBBOROUGH</var>
            <group>
                <name>HHsize</name>
                <var>HHSIZ02</var>
                <var>HHSIZ05</var>
                <var>HHSIZ08</var>
            </group>
```

grouped time variables in project file



grouped variables from project file




```
STID,CODE,TIME,HHsize,%KIDS,MedRENT
1,401,2002, 2.303200000000000,28.45
2,210,2002, 2.307200000000000,26.30
3,411,2002, 2.670800000000000,30.72
4,203,2002, 2.479900000000000,45.03
5,413,2002, 2.888700000000000,42.53
6,211,2002, 2.595700000000000,31.93
7,212,2002, 3.086600000000000,41.13
8,202,2002, 2.023200000000000,24.74
9,216,2002, 2.646400000000000,51.68
```

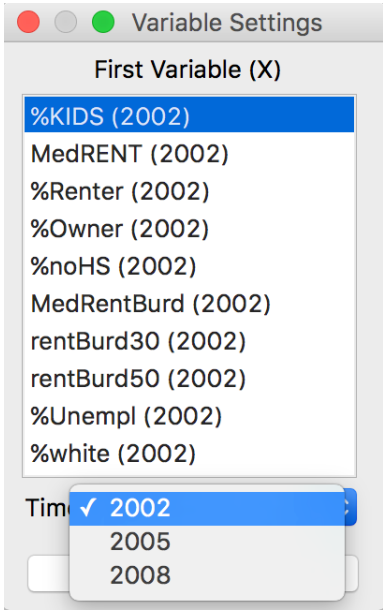
```
0 165 nycvars STID
1 1
2
2 2
3 1
3 1
2
4 2
6 5
```

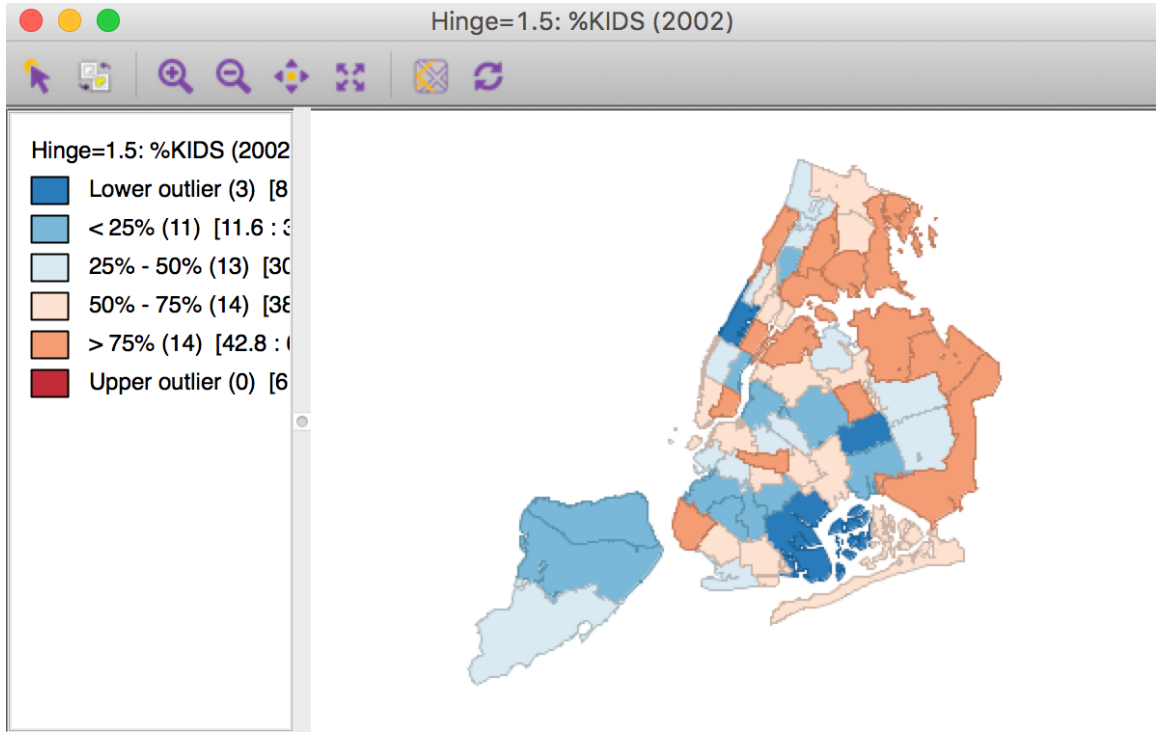
space-time ID, space ID, time ID

n = 55 x 3, space-time ID

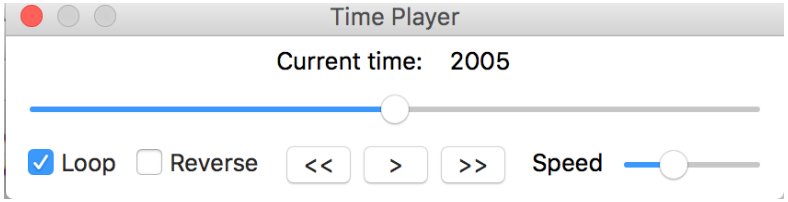
saved space-time table/weights

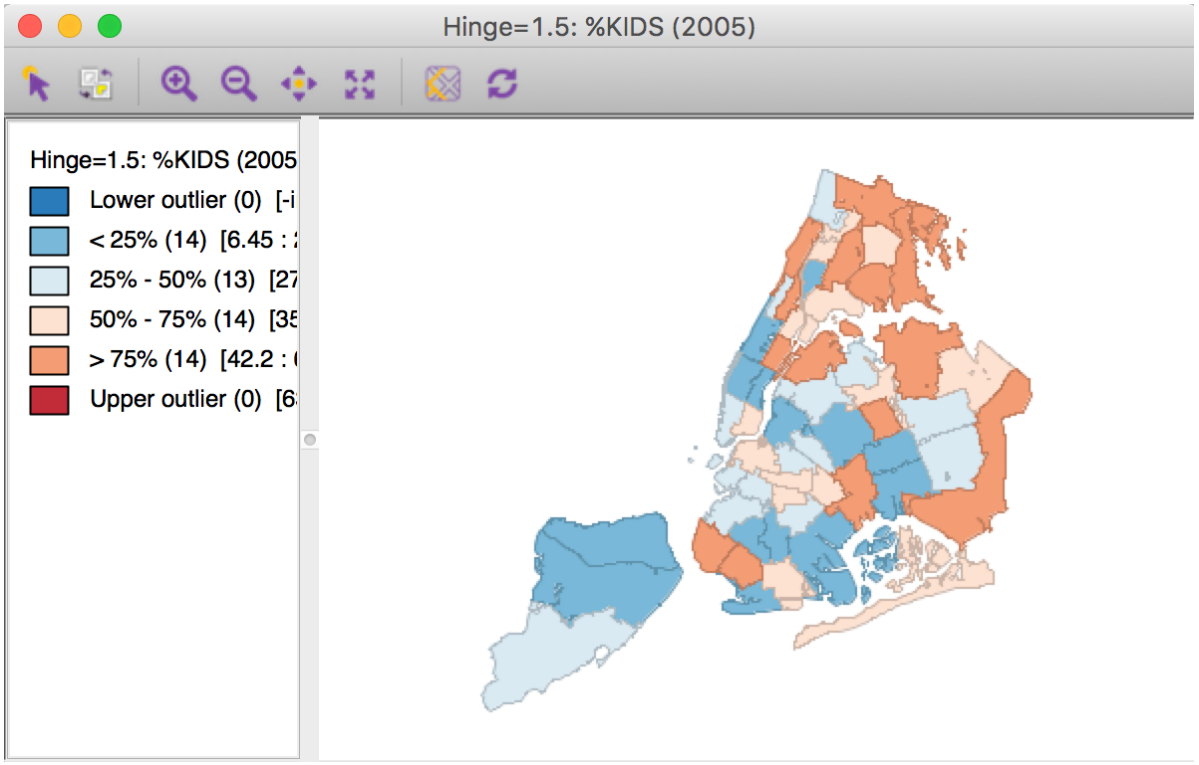
Time Player



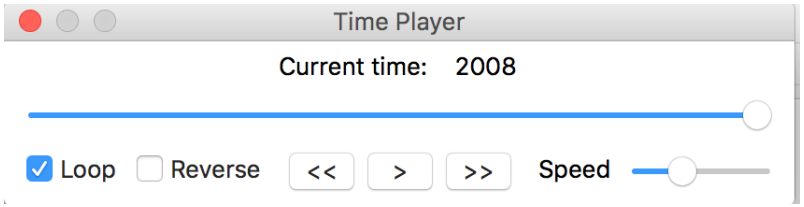


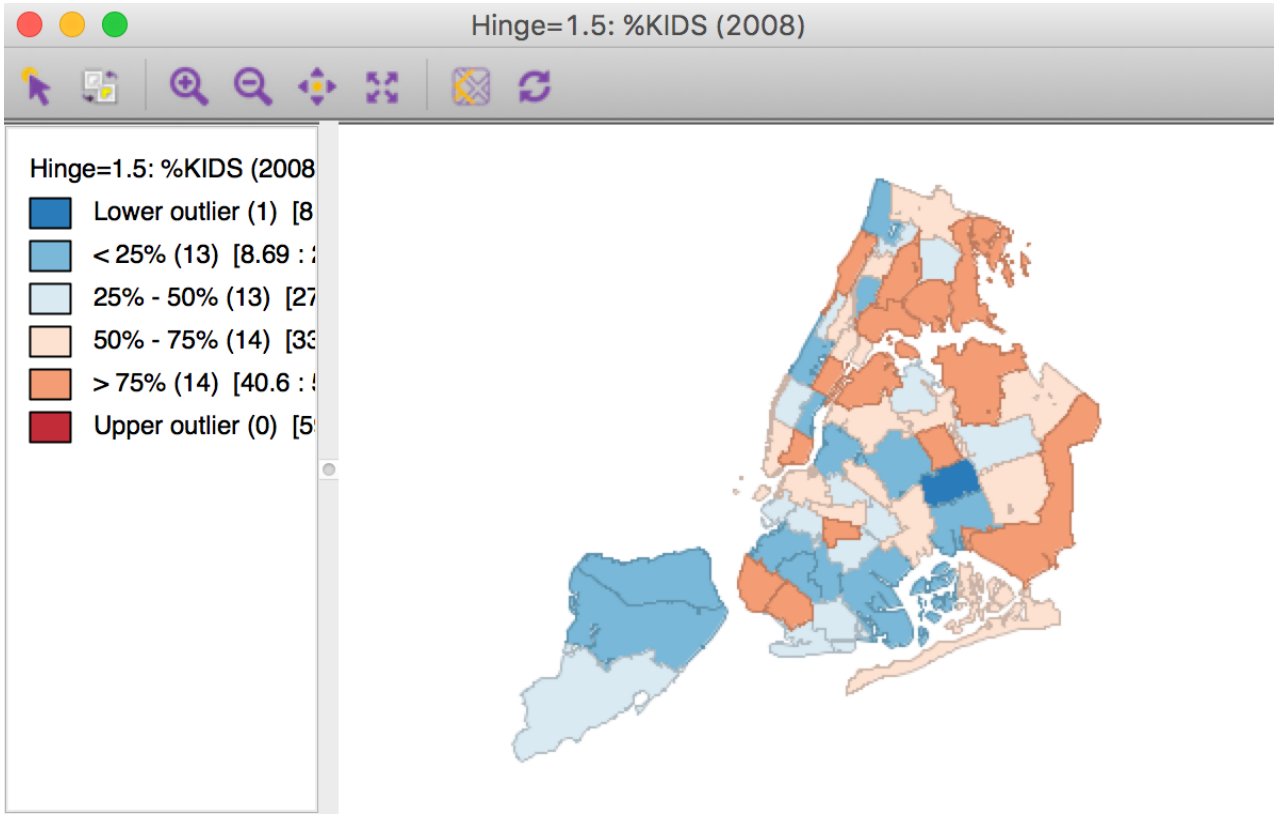
box map with time variable



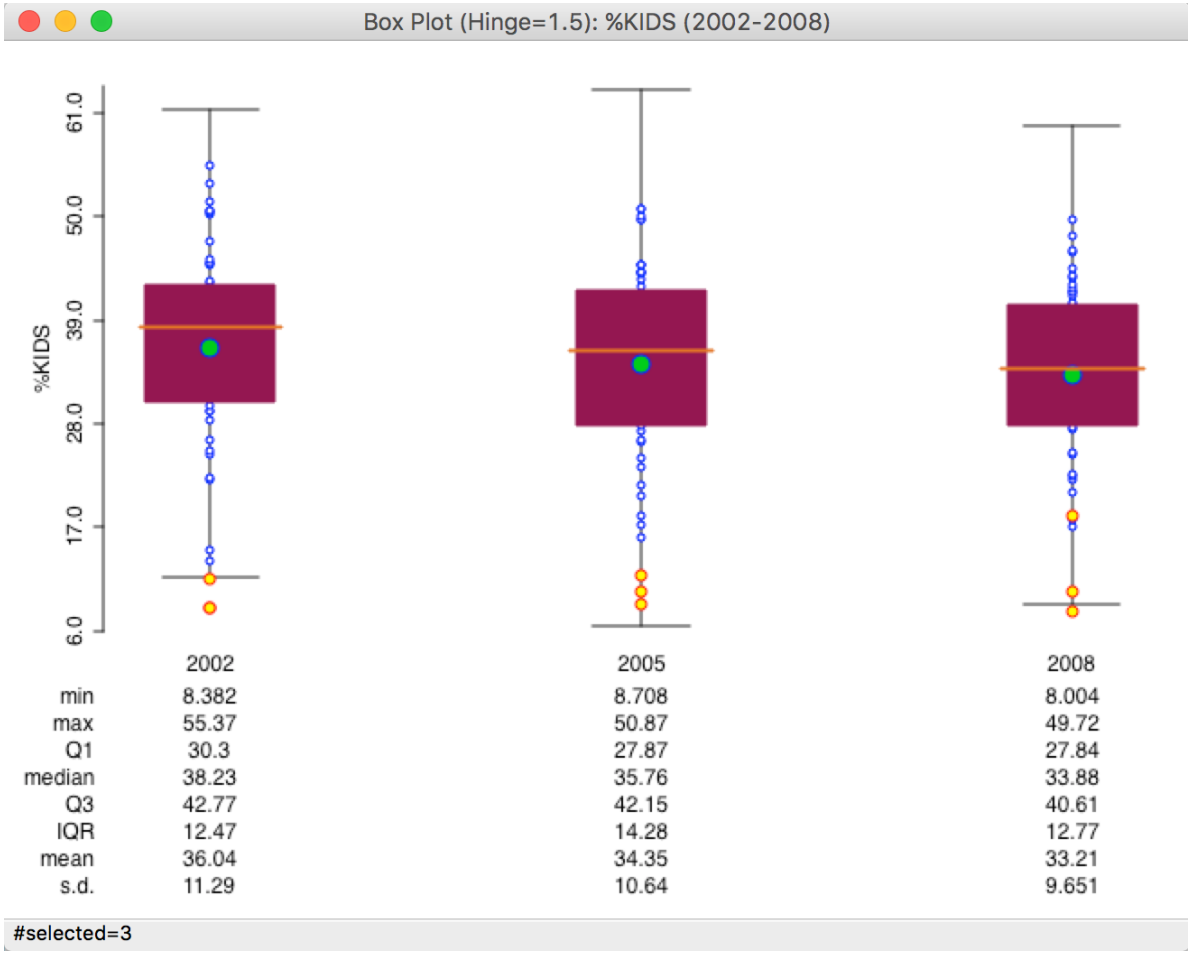


box map 2005

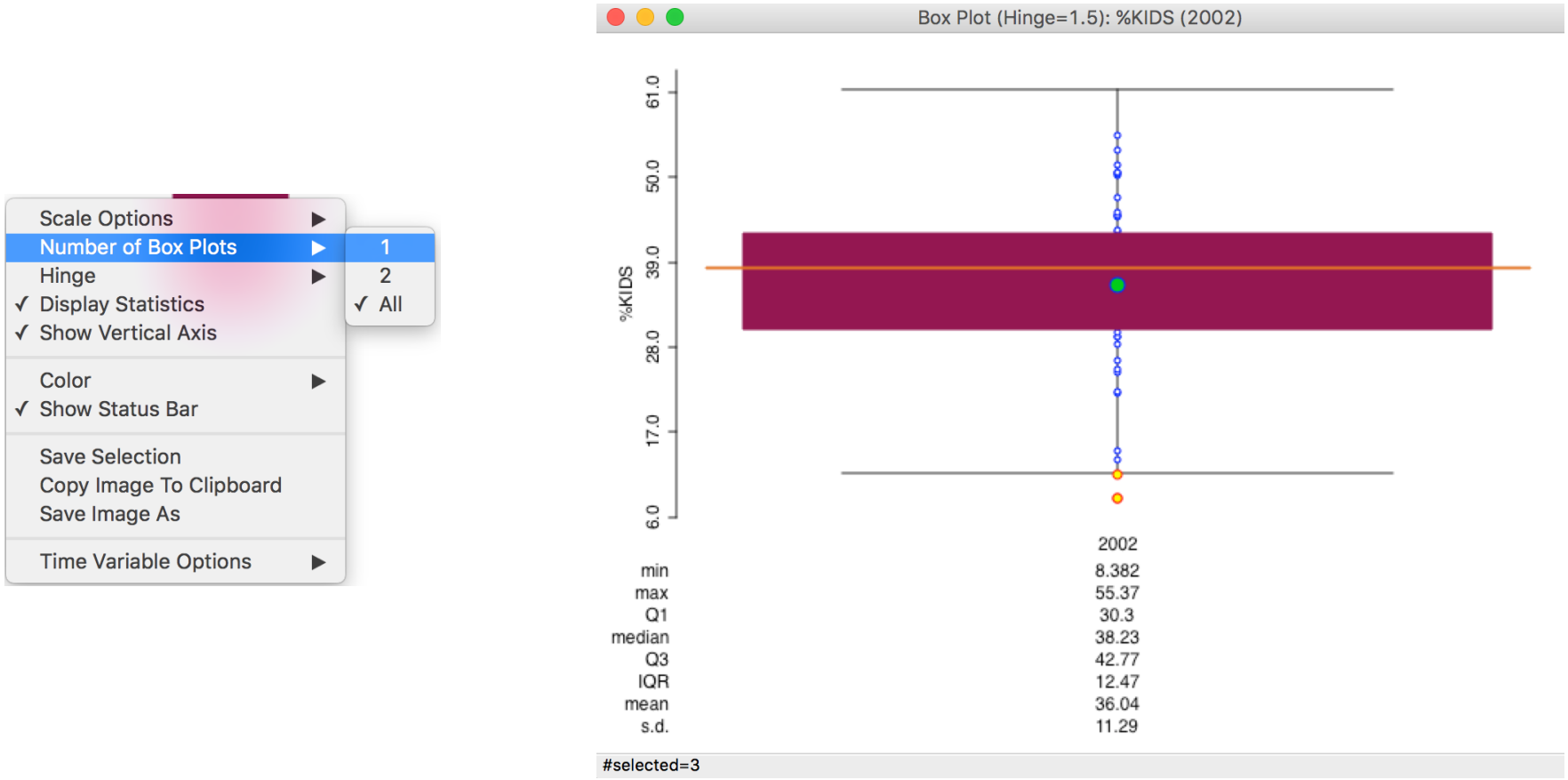




box map 2008



three period box plot



number of box plots option

Differential Spatial Autocorrelation

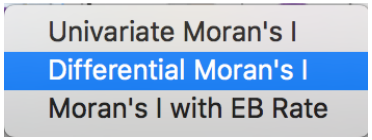
- Principle

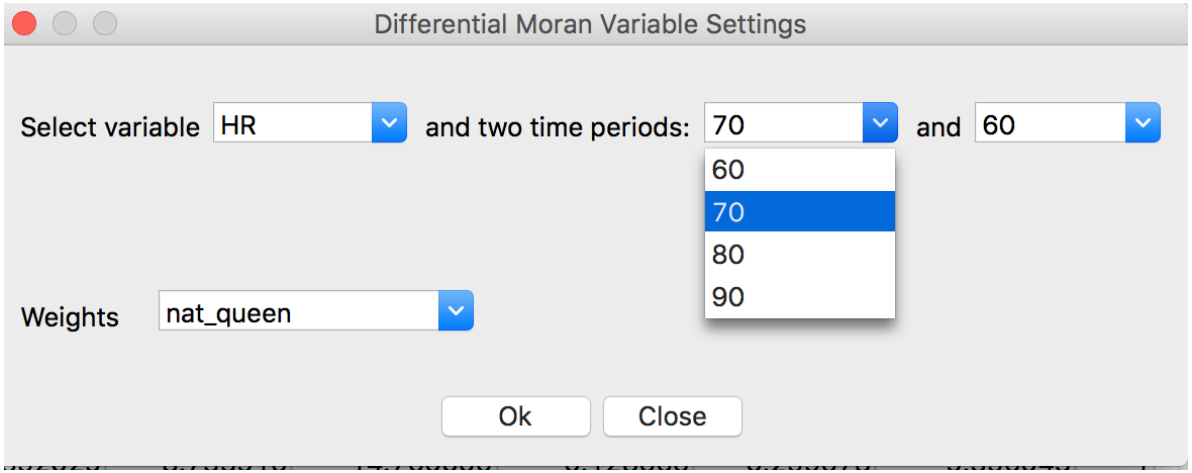
  spatial autocorrelation on change over time, i.e., on $y_t - y_{t-1}$

  Moran scatterplot for $y_t - y_{t-1}$

  Local Moran maps for $y_t - y_{t-1}$

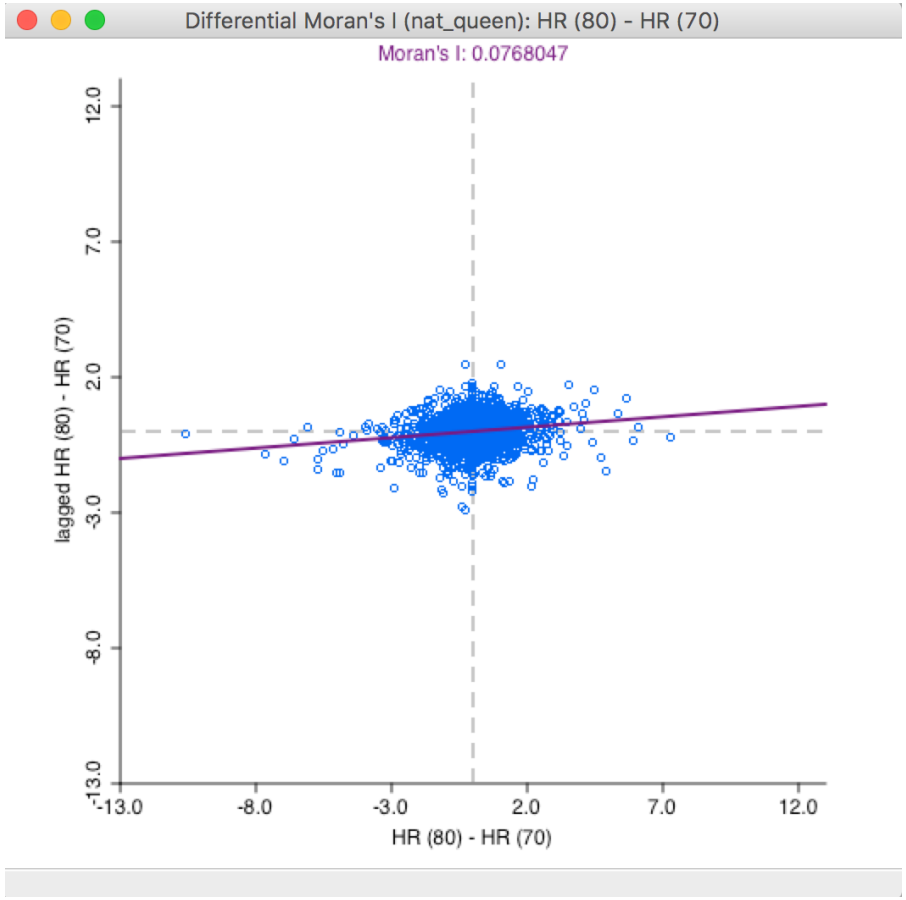
  replacement of bivariate Moran

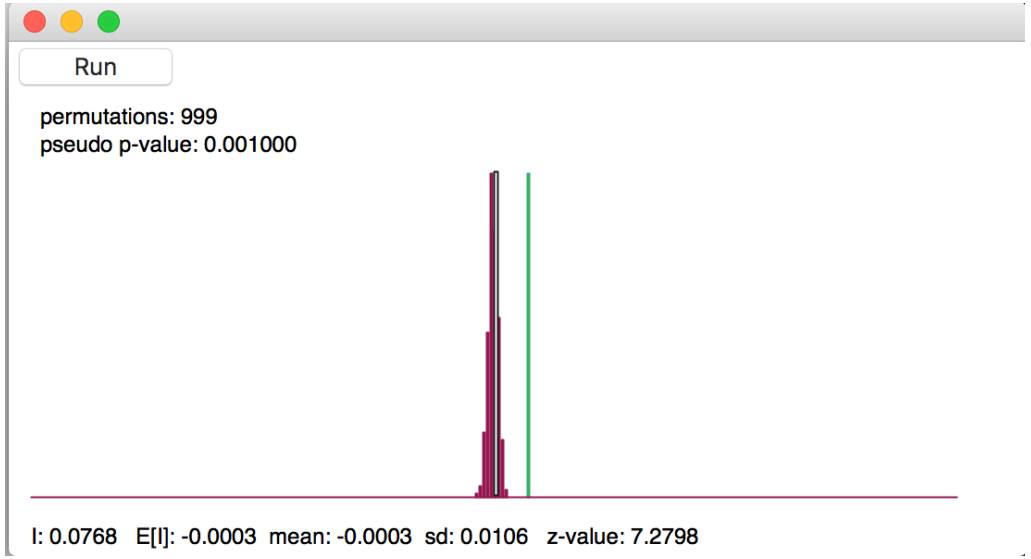




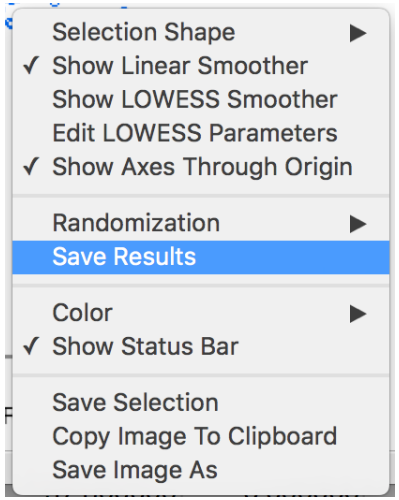
differential Moran's I (scatter plot)

variable selection using two time periods





differential Moran scatter plot



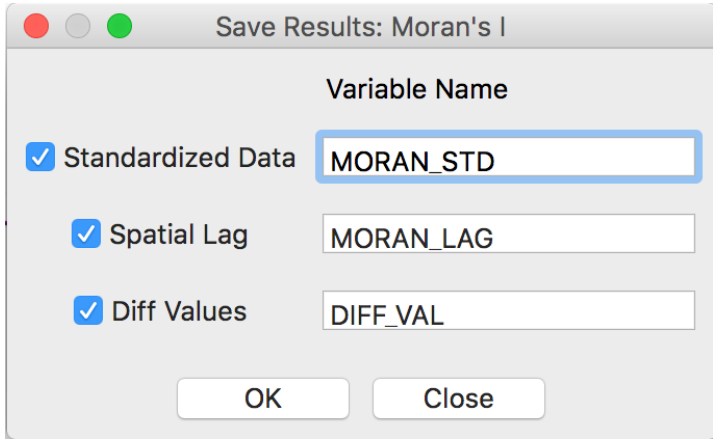
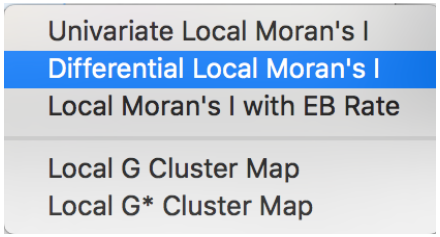
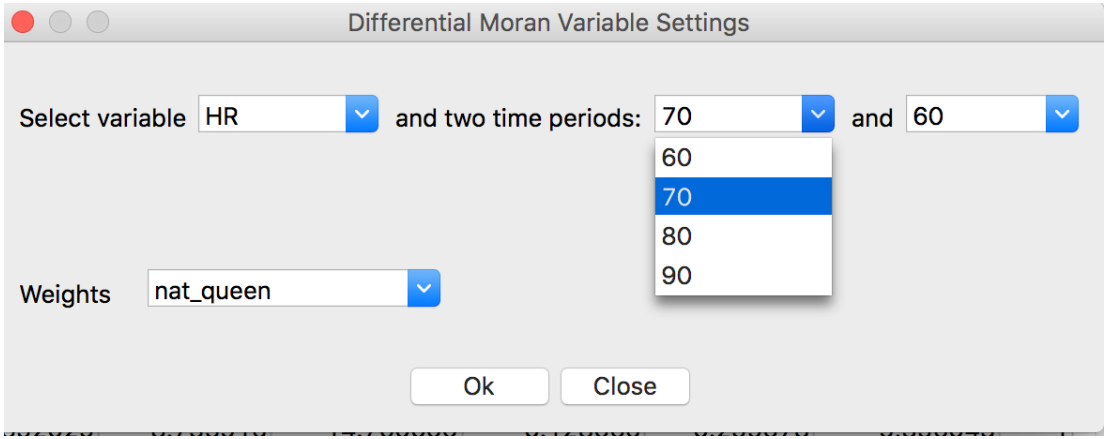


Table - natregimes

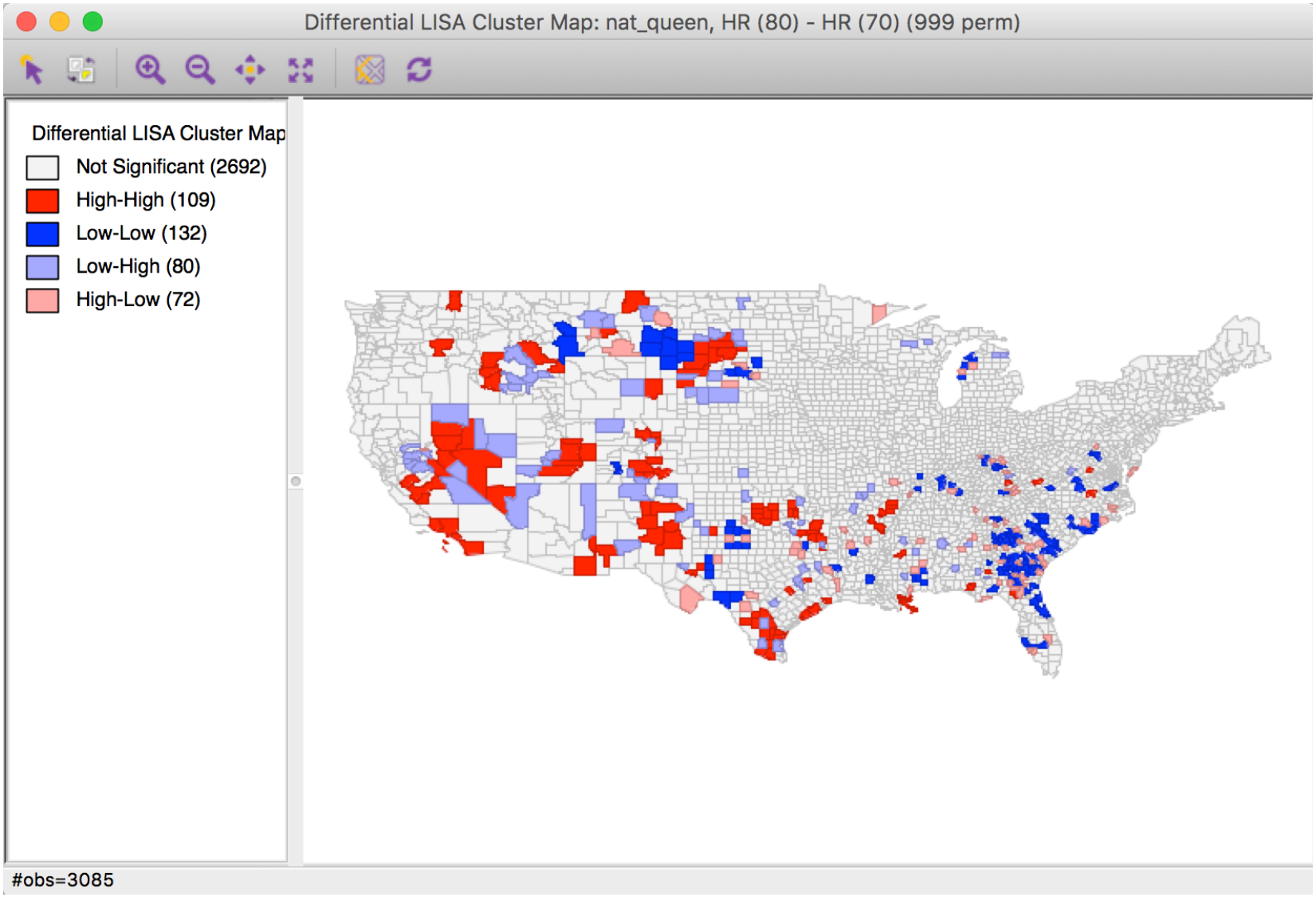
	FH (60)	West	MORAN_STD	MORAN_LAG	DIFF_VAL
35	11.279621	0	1.2198860	0.0255470	8.8558270
58	10.053476	1	2.4456820	0.0914580	17.2087420
99	9.258437	1	0.1456410	0.8732250	1.5356160
40	9.039900	1	0.2101460	0.4930810	1.9751710
63	8.243930	1	1.0605420	0.3052060	7.7700080

save lag and standardized and unstandardized first difference





differential local Moran



differential local Moran cluster map

local clustering of the change

Averages Tool

- Principle

  simple treatment effects analysis

  compare mean of a variable in selected vs unselected set of observations

    cross-section: selected vs unselected

    all observations: one time period vs another time period

    selected: one time period vs another time period

  simple F test on difference in means

- Difference-in-Differences Test

  dummy variable regression

    cross-sectional case: selected observations = 1

    all observations: second time period = 1

    selected at different points in time

      selected = 1, second time = 1, interaction = 1

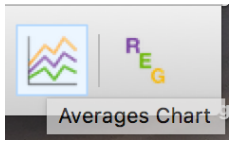
- Save Dummy

  creates new data set with dummy variables

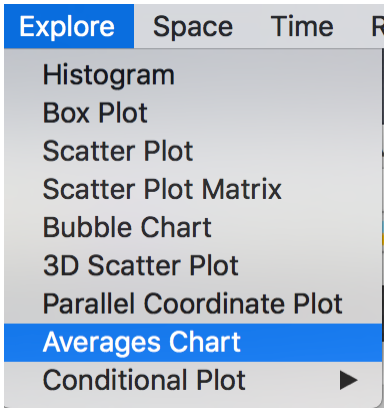
  in space-time case a space-time data set is created with a matching space-time spatial weights file (if a weights file is specified)

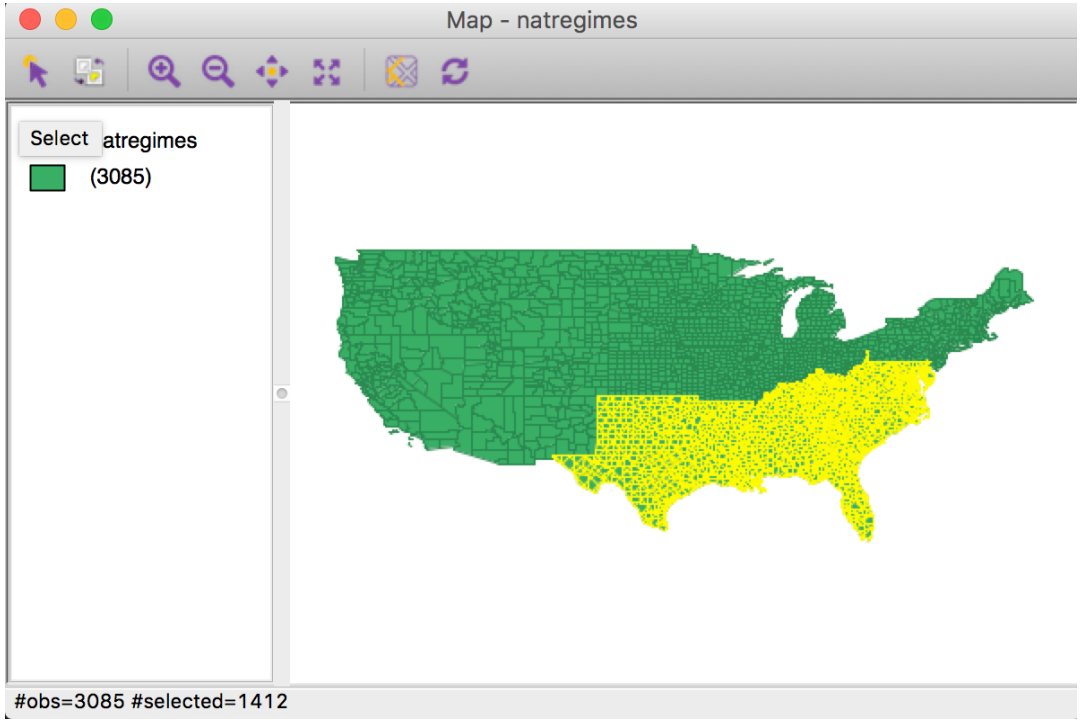
    this allows for space-time regression in the regression modules using the new data file and matching spatial weights file

    weights file is block-diagonal between time periods



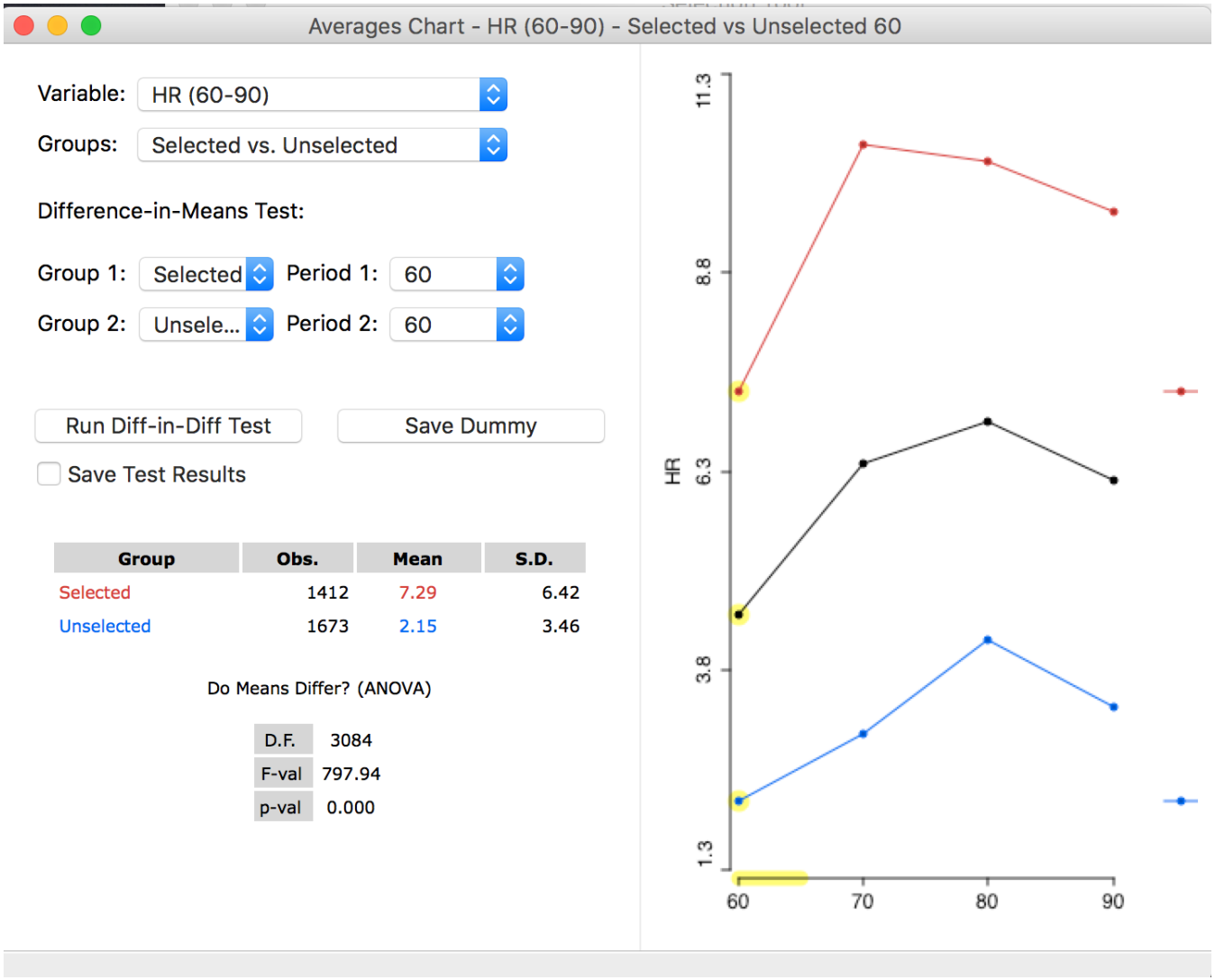
averages chart icon and menu selection



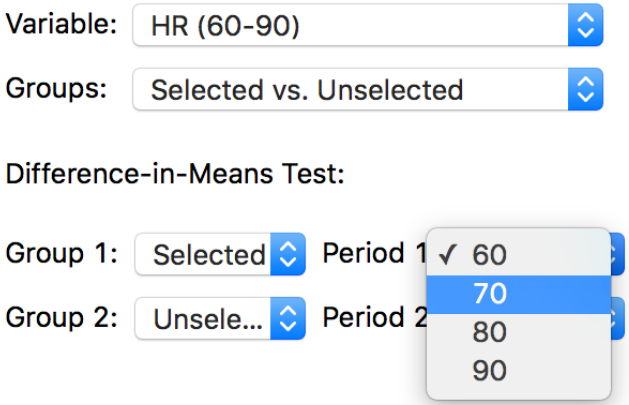


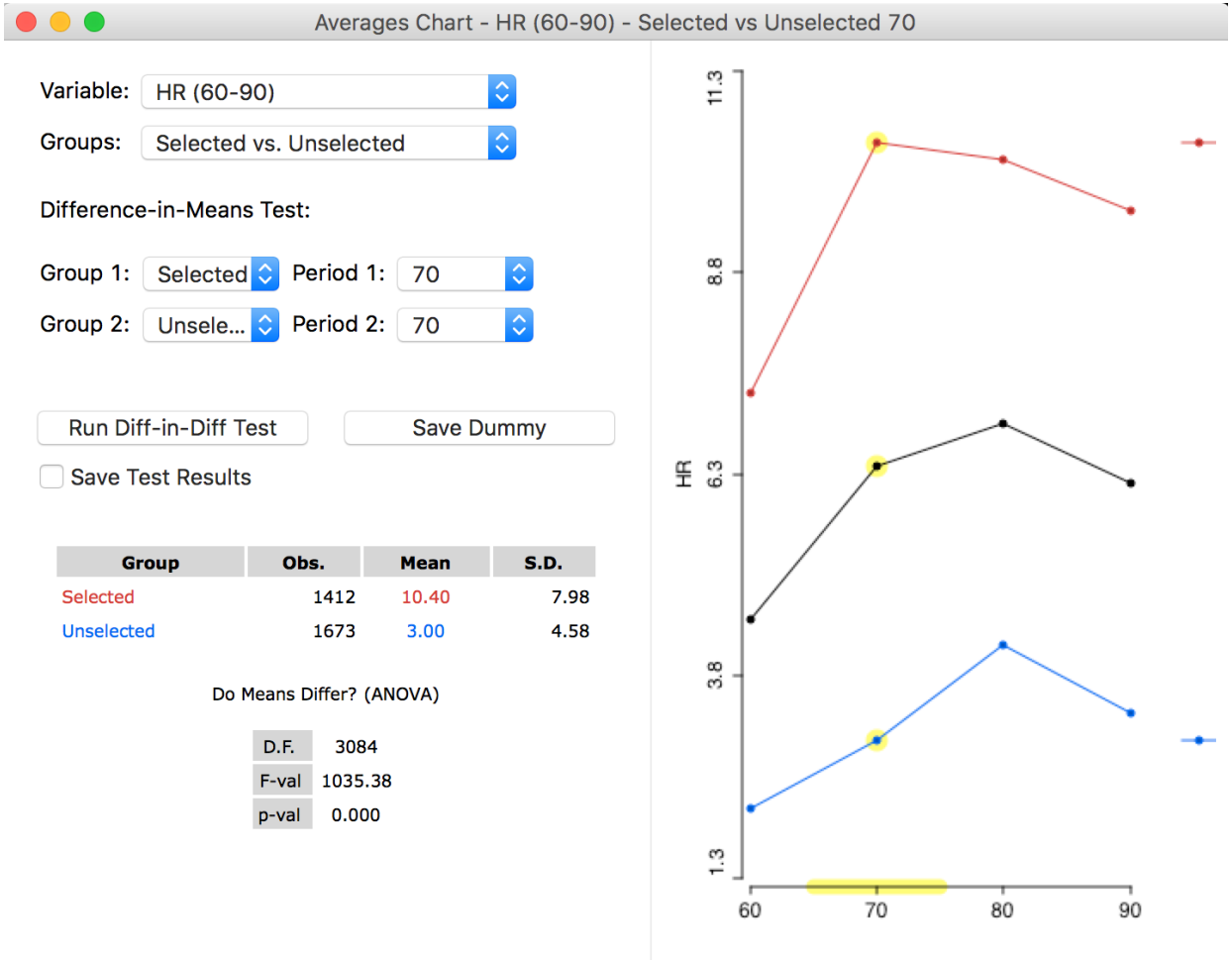
selection: south = 1

Cross-Section

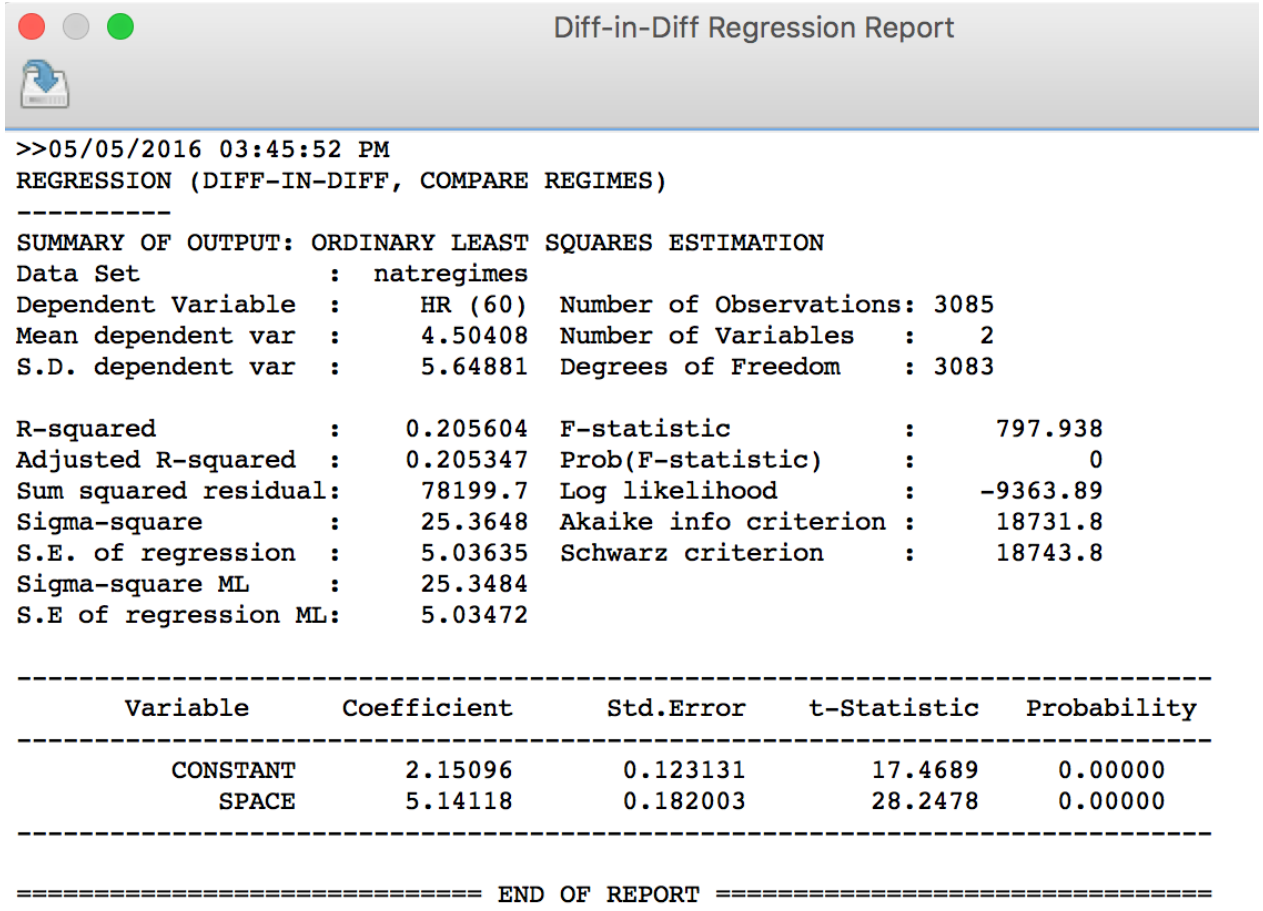


selected vs unselected for HR in 60

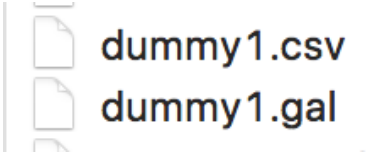




select different time period



run diff-in-diff test = dummy variable regression

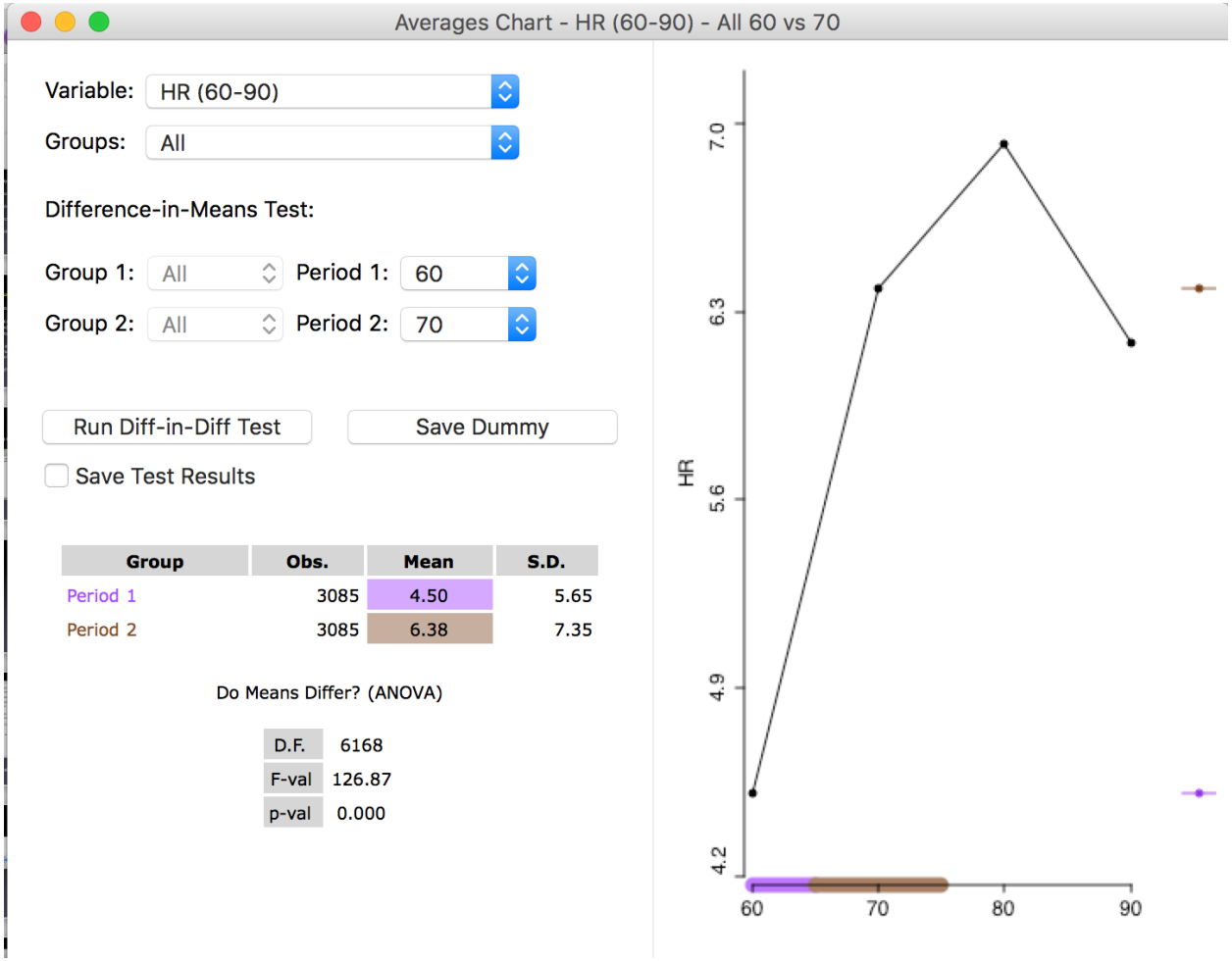



```
STID,FIPSNO,PERIOD,HR,SPACE
1,27077,60,        0.000000000,0
2,53019,60,        0.000000000,0
3,53065,60,        1.863863416,0
4,53047,60,        2.612330199,0
5,53051,60,        0.000000000,0
6,16021,60,        0.000000000,0
7,30053,60,        7.976389886,0
8,30029,60,        1.011173467,0
9,30035,60,       11.529038766,0
10,30101,60,        0.000000000,0
```

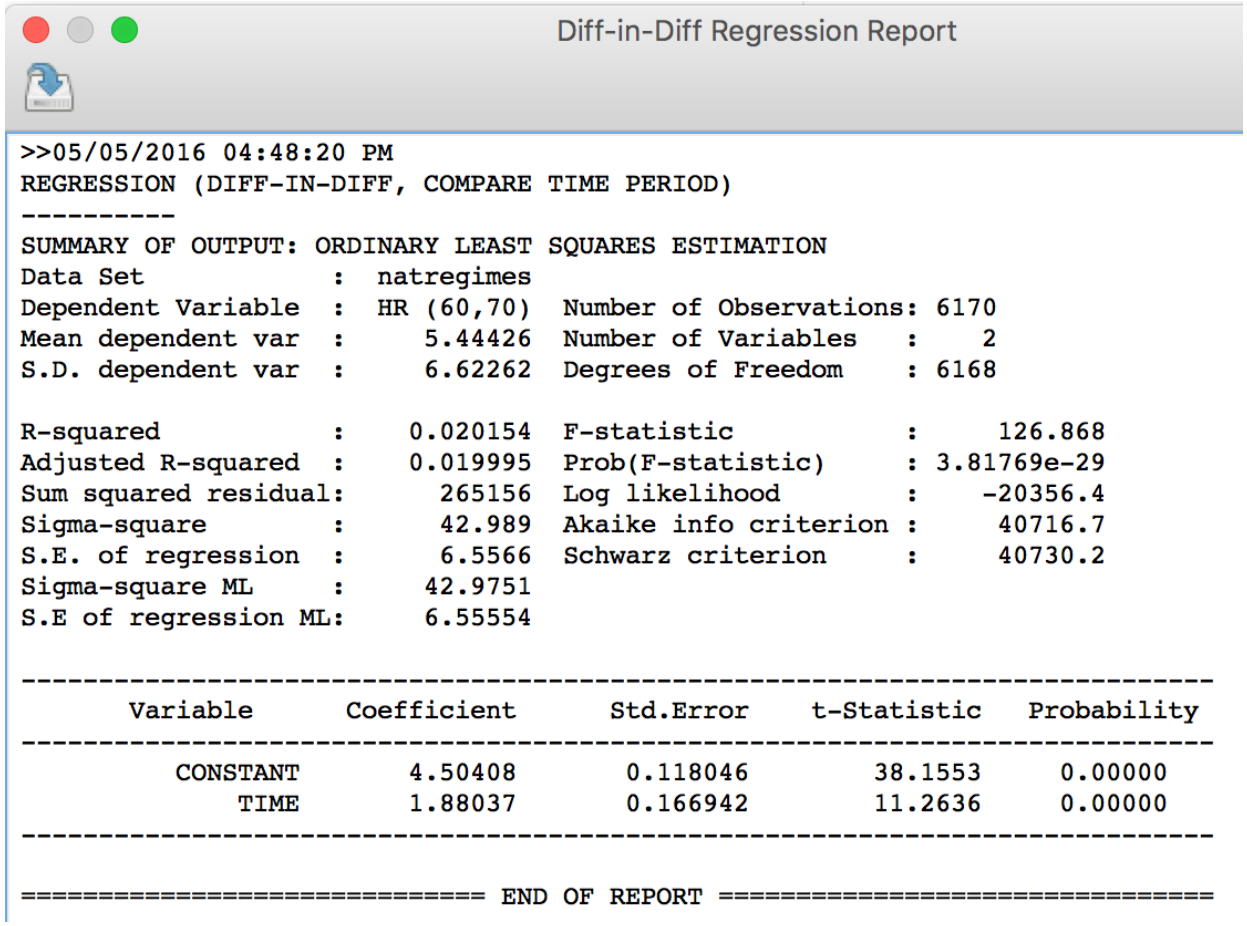
```
0 3085 dummy1 STID
1 3
30 22 40
2 3
69 2 3
3 4
1 62 4 69
4 7
69 1 27 31 42 55 68
5 4
2 5 28 62
```

files created by save dummy

All Observations - Two Periods



comparison of all observations in two time periods



run diff-in-diff test = dummy variable regression

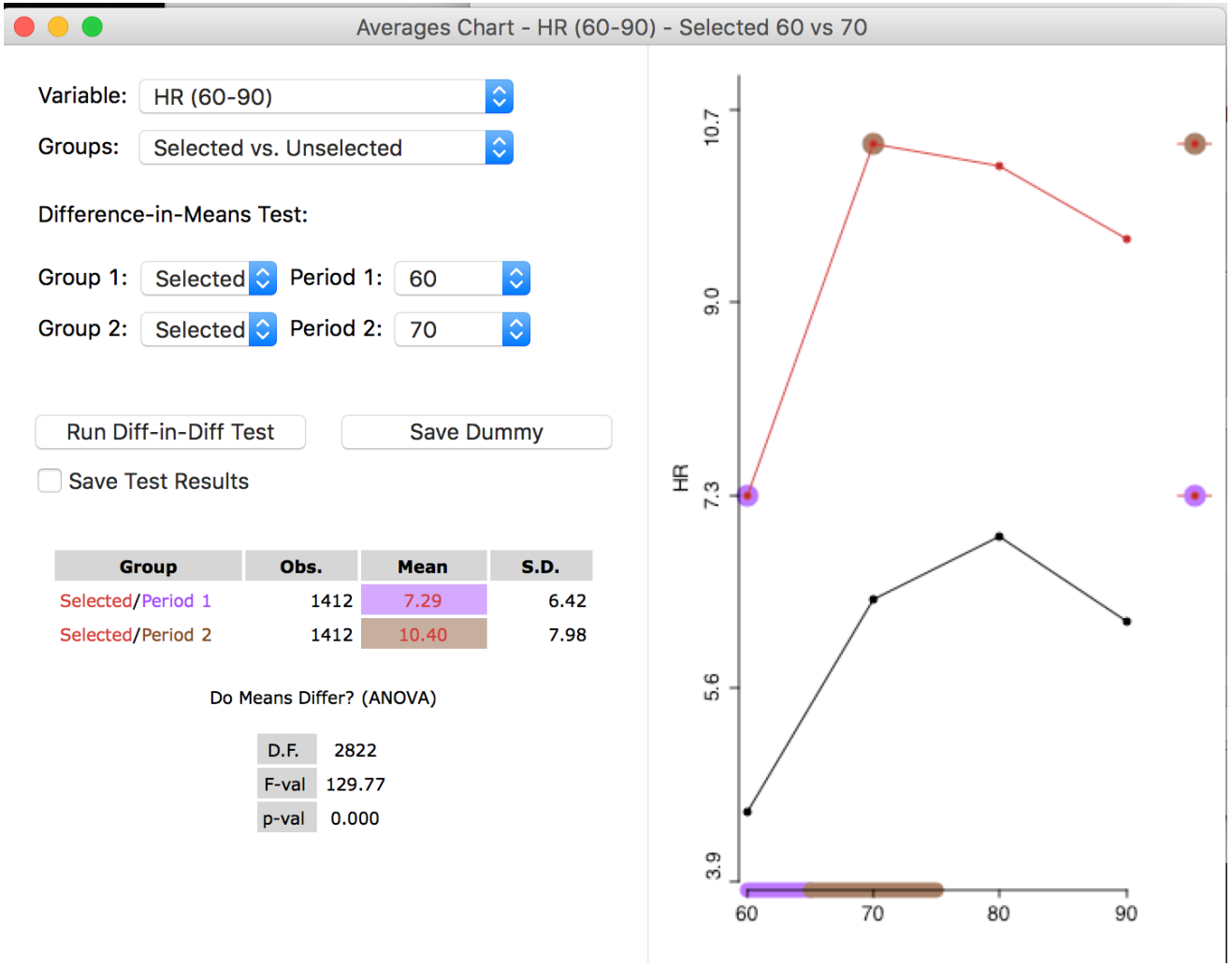
```
STID,FIPSNO,PERIOD,HR,TIME
1,27077,60,       0.000000000,0
2,53019,60,       0.000000000,0
3,53065,60,       1.863863416,0
4,53047,60,       2.612330199,0
5,53051,60,       0.000000000,0
6,16021,60,       0.000000000,0
7,30053,60,       7.976389886,0
8,30029,60,       1.011173467,0
9,30035,60,      11.529038766,0
```

```
0 6170 dummy2 STID
1 3
30 22 40
2 3
69 2 3
3 4
1 62 4 69
4 7
69 1 27 31 42 55 68
5 4
2 5 28 62
```

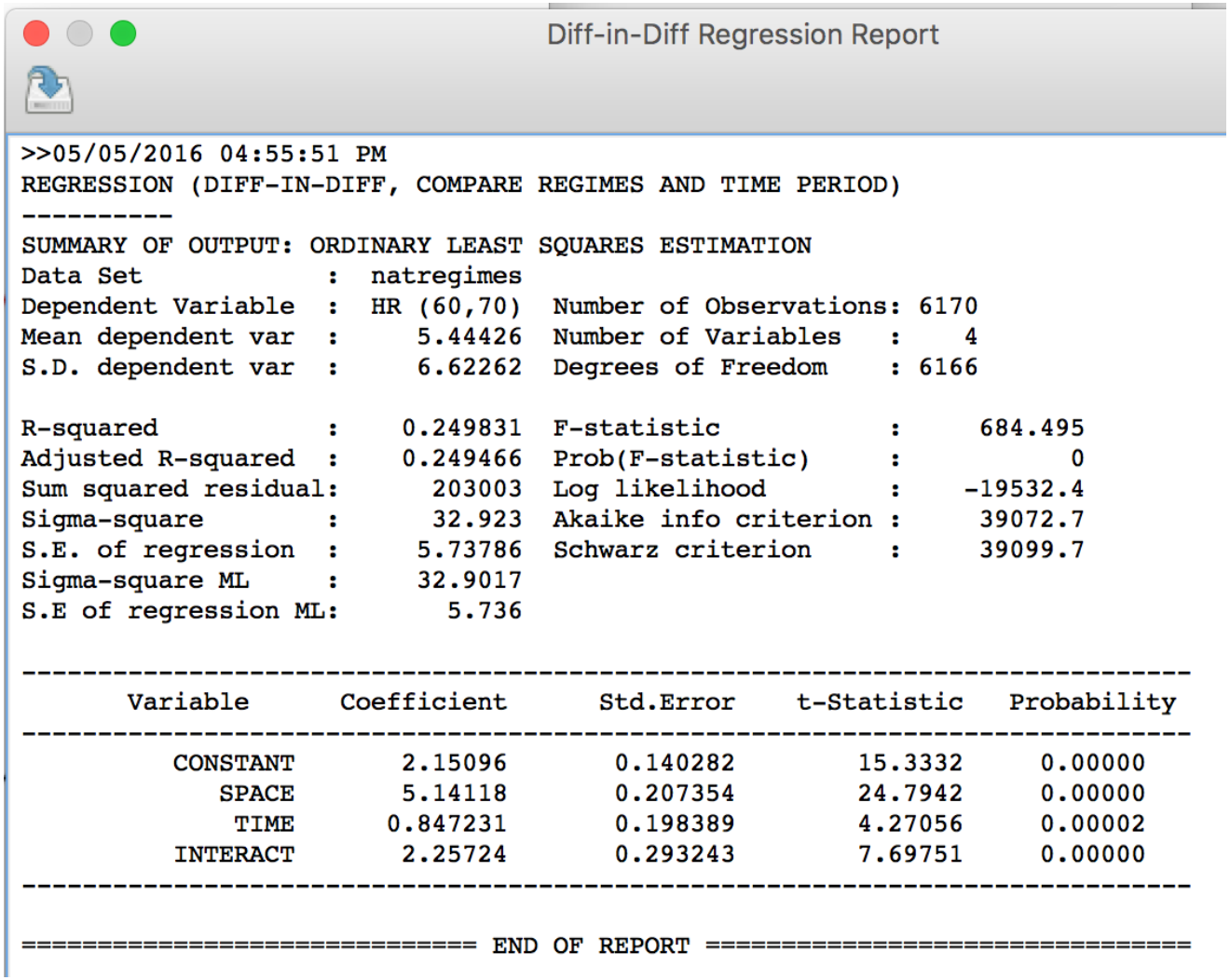
save dummy

space-time data set and gal weights file

Selected - Two Periods



comparison of selected observations in two time periods



run diff-in-diff test = space and time dummies and interaction

Spatial Regression

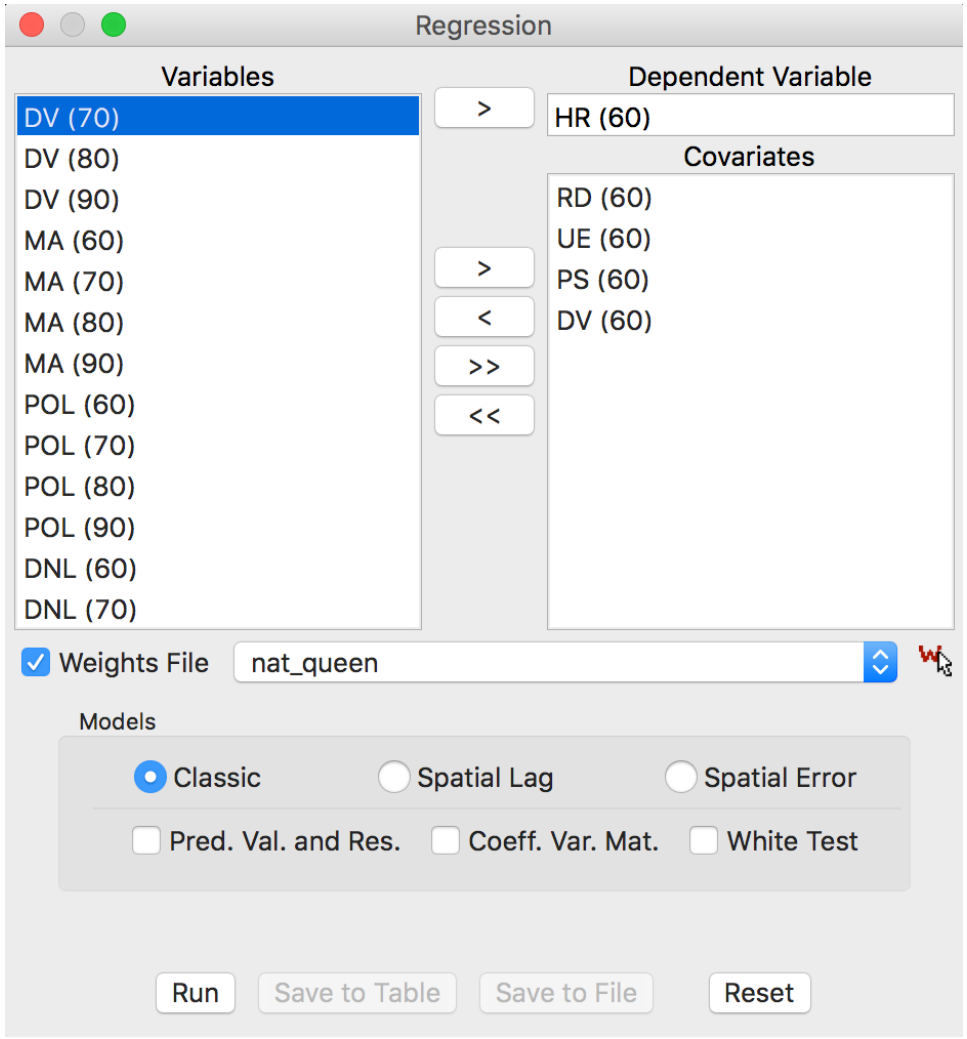
- Functionality

  OLS regression with diagnostics for spatial effects

  ML regression of spatial lag and spatial error models

    save residuals and predicted values

    limitation: only for intrinsically symmetric weights



regression toolbar icon and interface

Regression Report

```
>>05/05/2016 05:57:42 PM
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  natregimes
Dependent Variable  :     HR (60)  Number of Observations: 3085
Mean dependent var  :     4.50408  Number of Variables   :    5
S.D. dependent var  :     5.64881  Degrees of Freedom    : 3080

R-squared           :    0.220243  F-statistic           :     217.488
Adjusted R-squared  :    0.219231  Prob(F-statistic)     :           0
Sum squared residual:     76758.7  Log likelihood        :     -9335.2
Sigma-square        :     24.9217  Akaike info criterion :     18680.4
S.E. of regression  :     4.99216  Schwarz criterion     :     18710.6
Sigma-square ML     :     24.8813
S.E of regression ML:     4.98811

------------------------------------------------------------------------------
       Variable      Coefficient      Std.Error    t-Statistic   Probability
------------------------------------------------------------------------------
          CONSTANT       2.65084       0.252745        10.4882      0.00000
           RD (60)       2.87011      0.0987288        29.0706      0.00000
           UE (60)    -0.0367479      0.0364174       -1.00908      0.31300
           PS (60)      0.702509      0.0940004        7.47318      0.00000
           DV (60)       1.07469      0.0976658        11.0037      0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER   5.711199
TEST ON NORMALITY OF ERRORS
TEST                   DF           VALUE             PROB
Jarque-Bera             2       296191.3107          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                   DF           VALUE             PROB
Breusch-Pagan test      4           460.2628          0.00000
Koenker-Bassett test    4            18.6704          0.00091

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : nat_queen
   (row-standardized weights)
TEST                          MI/DF        VALUE          PROB
Moran's I (error)            0.1906       17.8535       0.00000
Lagrange Multiplier (lag)         1       382.1303       0.00000
Robust LM (lag)                   1        70.8931       0.00000
Lagrange Multiplier (error)       1       313.2932       0.00000
Robust LM (error)                 1         2.0560       0.15161
Lagrange Multiplier (SARMA)       2       384.1863       0.00000
================================ END OF REPORT ================================
```

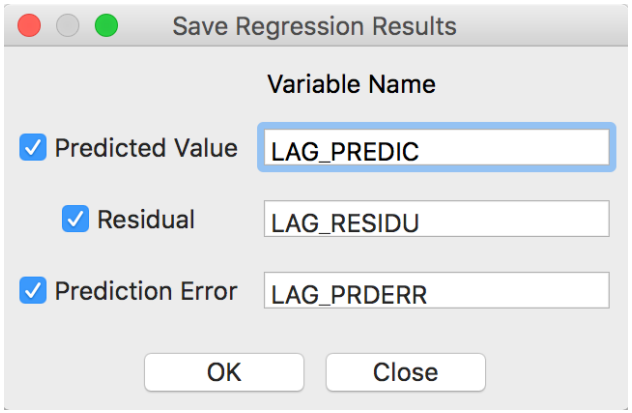
ols regression results with diagnostics

```
>>05/05/2016 05:58:11 PM
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : natregimes
Spatial Weight      : nat_queen
Dependent Variable  :     HR (60)  Number of Observations: 3085
Mean dependent var  :     4.50408  Number of Variables   :    6
S.D. dependent var  :     5.64881  Degrees of Freedom    : 3079
Lag coeff.   (Rho)  :    0.387045

R-squared           :    0.309955  Log likelihood        :    -9191.23
Sq. Correlation     : -            Akaike info criterion :     18394.5
Sigma-square        :     22.0186  Schwarz criterion     :     18430.7
S.E of regression   :      4.6924

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
         W_HR (60)      0.387045      0.0231936        16.6876      0.00000
         CONSTANT       1.32898       0.251275         5.28893      0.00000
          RD (60)       1.88275       0.106873        17.6167       0.00000
          UE (60)    0.00261764       0.0342308      0.0764703      0.93904
          PS (60)       0.49365       0.0893103        5.52736      0.00000
          DV (60)      0.735294       0.0924862        7.95031      0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF       VALUE         PROB
Breusch-Pagan test                      4        649.4843      0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : nat_queen
TEST                                    DF       VALUE         PROB
Likelihood Ratio Test                   1        287.9311      0.00000
================================ END OF REPORT =================================
```

ML estimation spatial lag

```
>>05/05/2016 05:58:30 PM
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL ERROR MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : natregimes
Spatial Weight      : nat_queen
Dependent Variable  :     HR (60)  Number of Observations: 3085
Mean dependent var  :    4.504075  Number of Variables   :    5
S.D. dependent var  :    5.648806  Degrees of Freedom    : 3080
Lag coeff. (Lambda) :    0.391016

R-squared           :    0.300728  R-squared (BUSE)      : -
Sq. Correlation     : -            Log likelihood        :-9212.716168
Sigma-square        :     22.3131  Akaike info criterion :     18435.4
S.E of regression   :     4.72367  Schwarz criterion     :     18465.6

------------------------------------------------------------------------------
       Variable       Coefficient     Std.Error       z-value    Probability
------------------------------------------------------------------------------
          CONSTANT       2.96545      0.309134        9.59279      0.00000
           RD (60)       2.44082      0.129618        18.8308      0.00000
           UE (60)     0.0359107     0.0413815       0.867796      0.38551
           PS (60)      0.650578      0.111279        5.84636      0.00000
           DV (60)      0.693085      0.109803        6.31207      0.00000
            LAMBDA      0.391016     0.0243657        16.0478      0.00000
------------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                    DF      VALUE        PROB
Breusch-Pagan test                      4       613.6859     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL ERROR DEPENDENCE FOR WEIGHT MATRIX : nat_queen
TEST                                    DF      VALUE        PROB
Likelihood Ratio Test                   1       244.9628     0.00000
============================== END OF REPORT ================================
```

ML spatial error model



save to table for spatial lag model