

# GeoDa Workshop

## Part I

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 IR01HS021752-01AI



# introduction and overview

## getting started

## geovisualization

## multivariate EDA

## spatial weights



# Introduction and Overview



**motivation and history**

**functionality**

**what is new in 1.8**

**sample data sets**



# Motivation and History



- **vision**

to provide a software tool that facilitates the exploration and analysis of geospatial data as a transition from simple description and visualization to structured exploration and formal modeling

no GIS needed, but compatible with GIS data structures





# From SpaceStat to CyberGIS: Twenty Years of Spatial Data Analysis Software

Luc Anselin<sup>1</sup>

## Abstract

This essay assesses the evolution of the way in which spatial data analytical methods have been incorporated into software tools over the past two decades. It is part retrospective and prospective, going beyond a historical review to outline some ideas about important factors that drove the software development, such as methodological advances, the open source movement and the advent of the Internet and cyberinfrastructure. The review highlights activities carried out by the author and his collaborators and uses SpaceStat, GeoDa, PySAL, and recent spatial analytical web services developed at the ASU GeoDa Center as illustrative examples. It outlines a vision for a spatial econometrics workbench as an example of the incorporation of spatial analytical functionality in a cyberGIS.

## Keywords

spatial analysis, methods, spatial statistics and spatial econometrics, geographic information science, modeling in GIS

Anselin (2012)



- brief history

antecedents

SpaceStat (1992), DynESDA (1998)

legacy GeoDa (2003)

Windows XP only, built on ESRI MapObjects



- open GeoDa

- open source and cross-platform

- hosted on Github

- native look and feel on each OS

- exploits many open source libraries

- WxWidgets, Boost, GDAL, etc.

- first release Version 1.0 in 2011, 1.8 in 2016

- close to 200,000 users worldwide



# GeoDa on Github

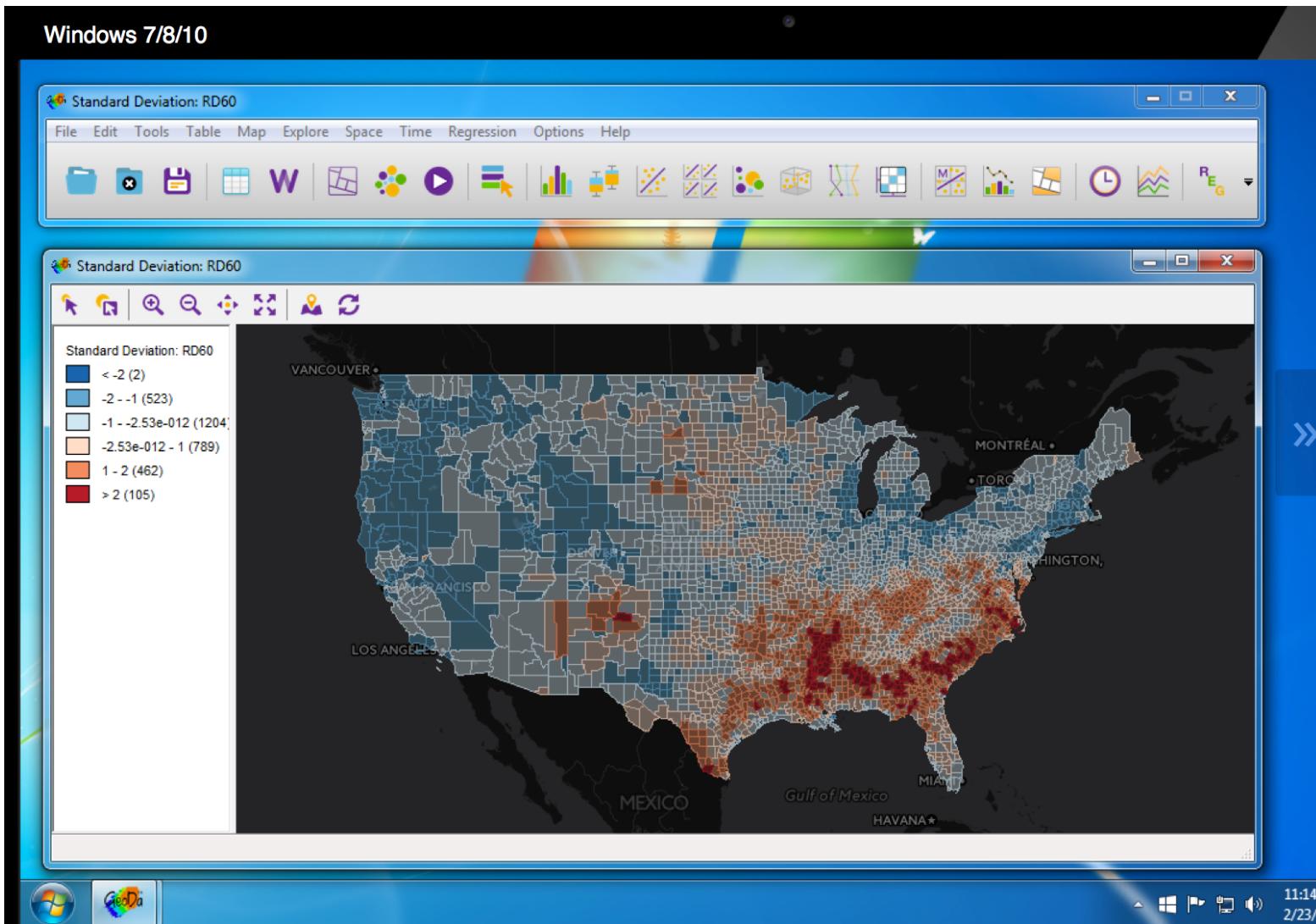
GeoDa (TM): Software providing an introduction to spatial data analysis.

[View on GitHub](#)[Download .zip](#)[Download .tar.gz](#)

[geodacenter.github.io](http://geodacenter.github.io)



Windows 7/8/10

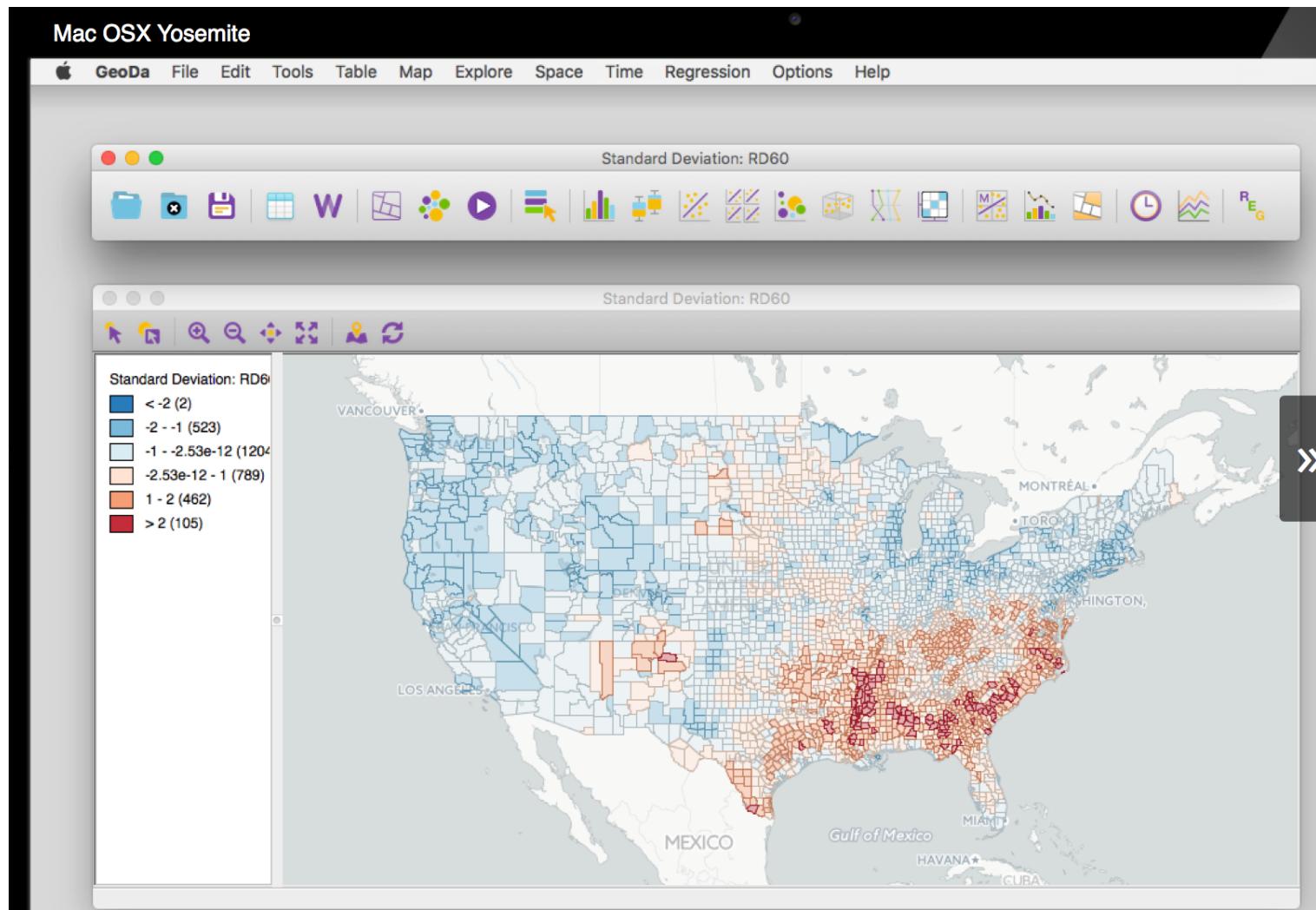


# GeoDa for Windows



Copyright © 2016 by Luc Anselin, All Rights Reserved



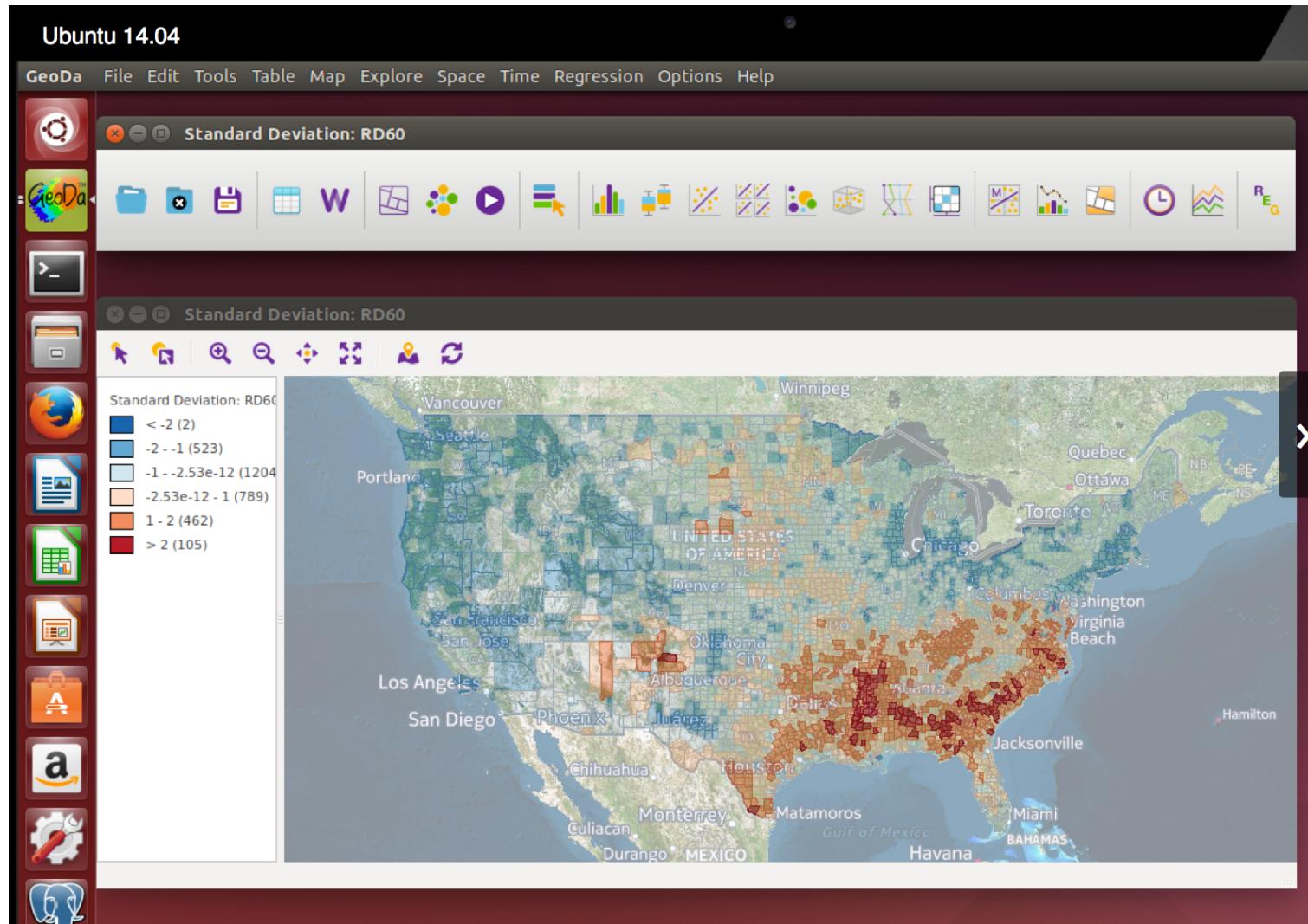


# GeoDa for Mac OSX



Copyright © 2016 by Luc Anselin, All Rights Reserved





## GeoDa for Linux (Ubuntu)



# Functionality



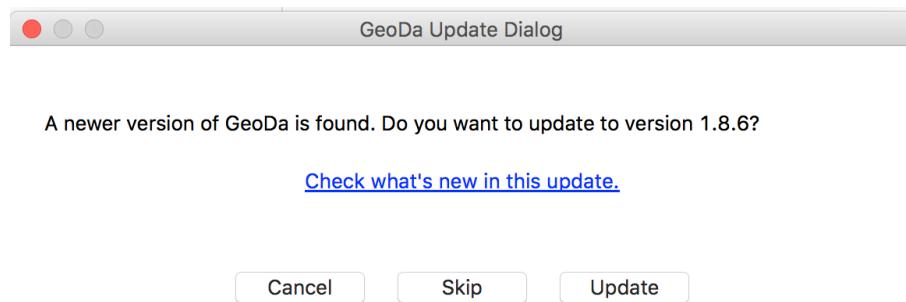
**GeoDa** File Edit Tools Table Map Explore Space Time Regression Options Help

## the GeoDa Menu



## the GeoDa toolbar





automatic updates



Copyright © 2016 by Luc Anselin, All Rights Reserved



- distinctive features

- multiple views on the data

- dynamic linking and brushing

- observations selected in one view are immediately selected in all views

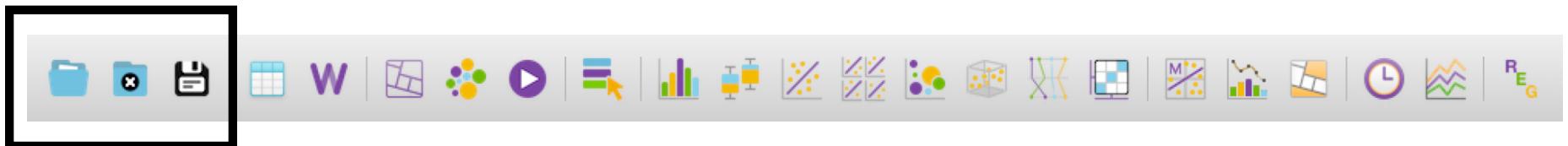
- assessing spatial heterogeneity

- structural breaks across space

- space-time exploration



- **data entry**



supports multiple file formats

GDAL/OGR

data base connections

PostGIS, Oracle

web feature service (WFS)

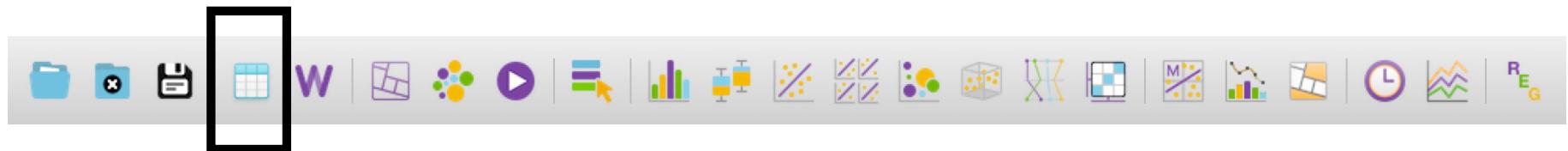
CartoDB data sets

“Export” serves as file format converter

project file



# • data manipulation



# table functionality

## join different tables (dbf, csv)

## new variable creation and transformation

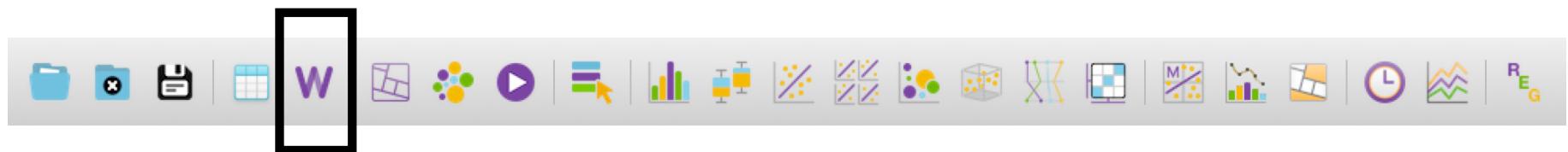
# queries/selection of observations

## export selected observations

# edit variable properties



- **weights manager**



create spatial weights

contiguity based, distance based

load spatial weights from external files

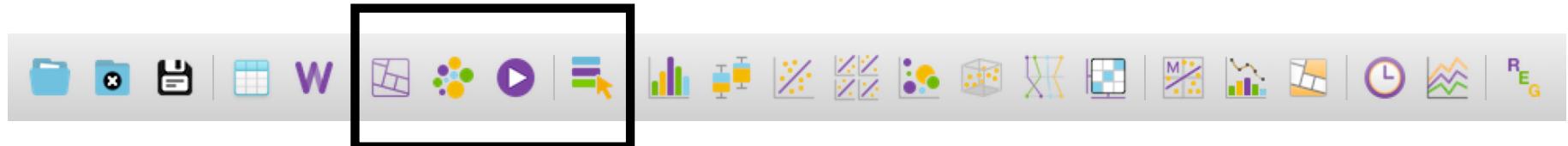
weights summary

connectivity histogram

interactive connectivity map



- mapping and geovisualization



choropleth maps

rate maps and rate smoothing

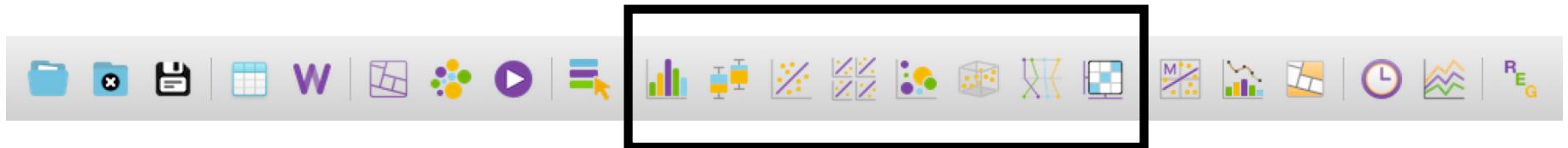
cartogram

map movie

interactive custom break editor



# • exploratory data analysis (EDA)



# histogram

# box plot

# scatter plot (with lowess smoothing)

# scatter plot matrix

# bubble chart

# 3-D scatter plot

# parallel coordinate plot (PCP)

# conditional plot (map, histogram, scatter plot)

- spatial autocorrelation analysis



## global spatial autocorrelation

Moran scatterplot, univariate and differential (space-time)

## spatial correlogram

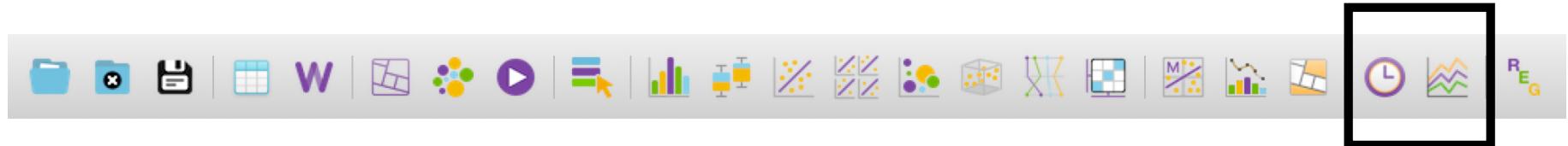
## local spatial autocorrelation

univariate and differential local Moran cluster map

## local G cluster map



# ● space-time analysis



time manager

time editor

time player

grouping variables

creating space-time variables

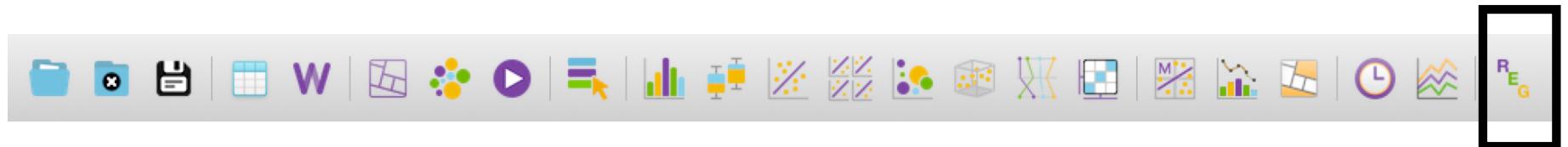
animations

averages tool

structural breaks, treatment effects, DID



- spatial regression



OLS regression with spatial diagnostics

tests for spatial effects

ML estimation of spatial lag model

ML estimation of spatial error model

residuals and predicted value maps



# What is new in 1.8



- new features
  - automatic updates
  - CartoDB data connection
  - project files
  - background layers for maps
  - lowess scatter plot smoother
  - scatter plot matrix
  - spatial weights connectivity map
  - non-parametric spatial autocorrelation
  - differential (space-time) spatial autocorrelation
  - averages tool
  - pooled space-time regression/weights



- improved features

- time manager

- space-time analysis

- table editor

- variable properties editor

- category editor

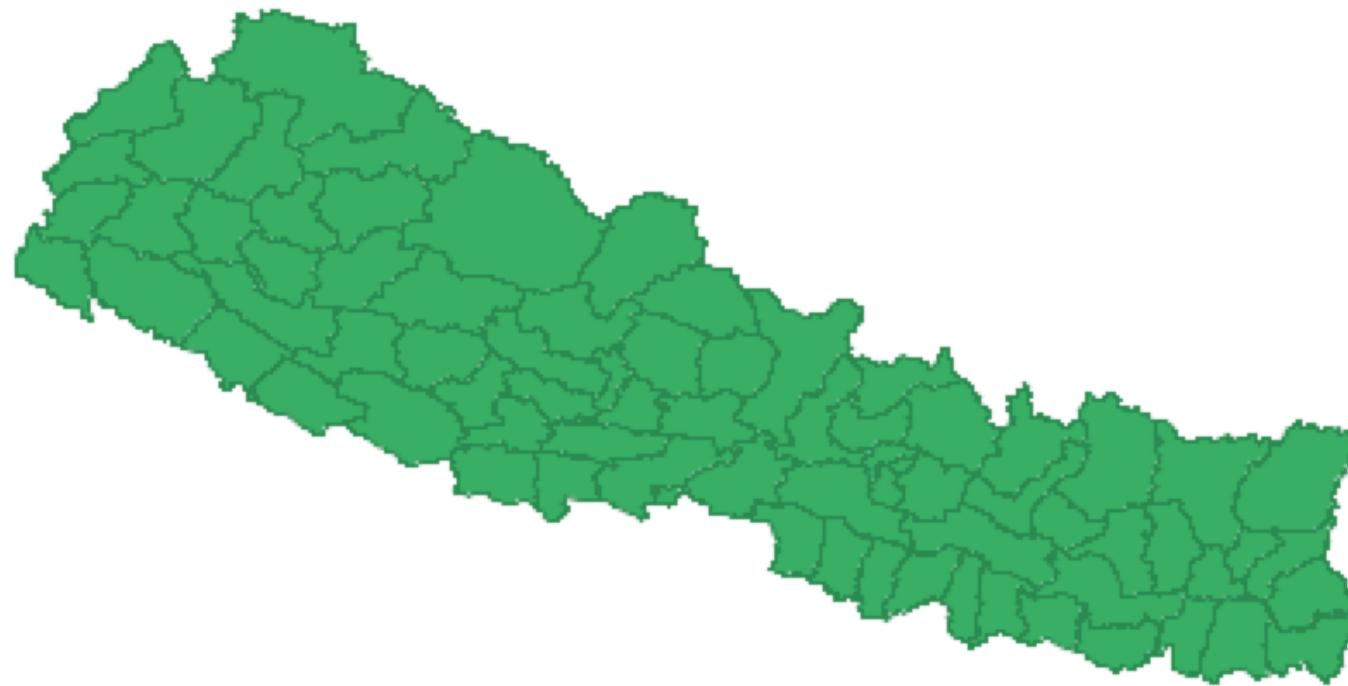
- animation

- many under-the-hood improvements



# Sample Data Sets





75 Nepal districts



- Nepal districts

n = 75

most data for 2013

cross-sectional only

socio-economic characteristics

open Nepal data: [data.opennepal.net](http://data.opennepal.net)

variables

deprivation index, poverty index, per capital income,  
malnourished kids, life expectancy, percentage  
without safe water





55 NYC sub-boroughs



- NYC Sub-Boroughs

n = 55

three time periods: 2002, 2005, 2008

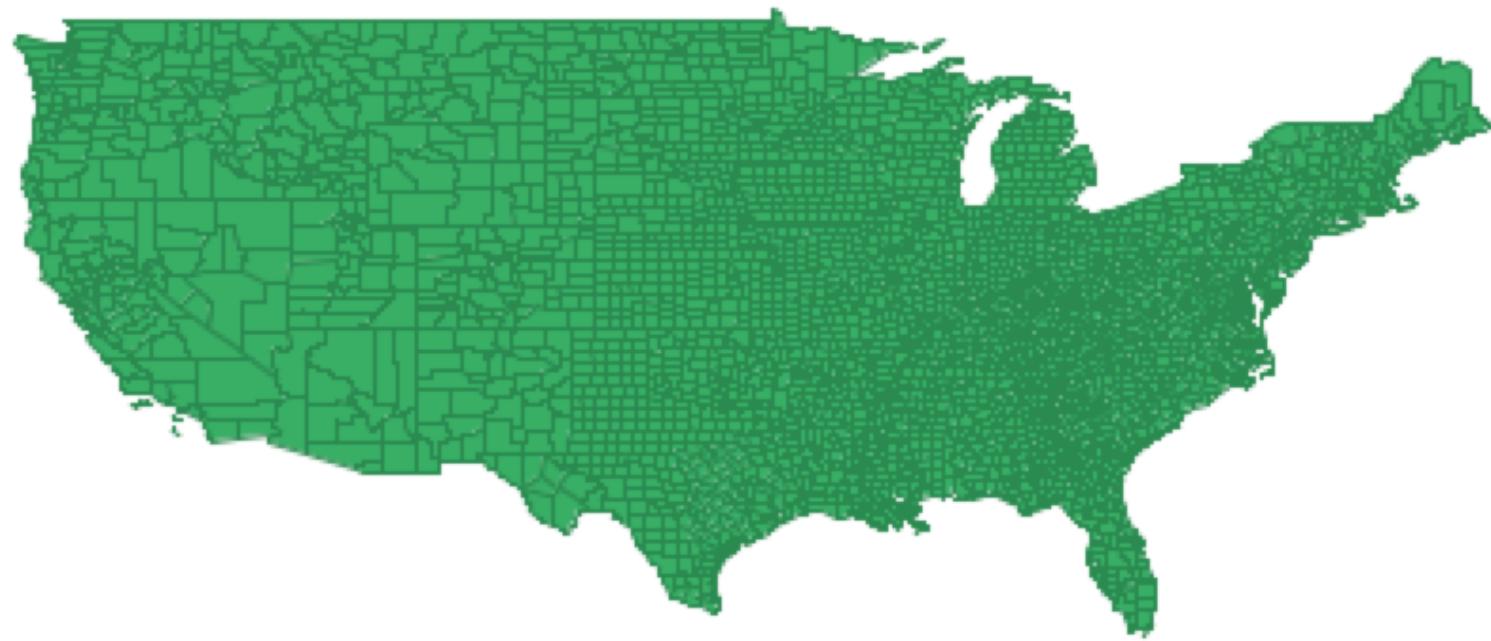
socio-economic characteristics

Furman Center: [furmancenter.org](http://furmancenter.org)

variables

household size, percent kids, median rent, rent burden, percent renter, percent owner, percent no high school, unemployed, ethnicity, years in neighborhood, percent foreign





3085 US counties



- US Counties

n = 3085

four time periods: 1960, 1970, 1980, 1990

homicides and socio-economic characteristics

geodacenter.asu.edu sample data sets

variables

homicide rate, resource deprivation index,  
unemployment rate, population component index,  
divorce rate, median age, families below poverty,  
Gini index, female headed households



# Getting Started



Copyright © 2016 by Luc Anselin, All Rights Reserved



**data input**

**creating spatial data sets**

**data cleanup**

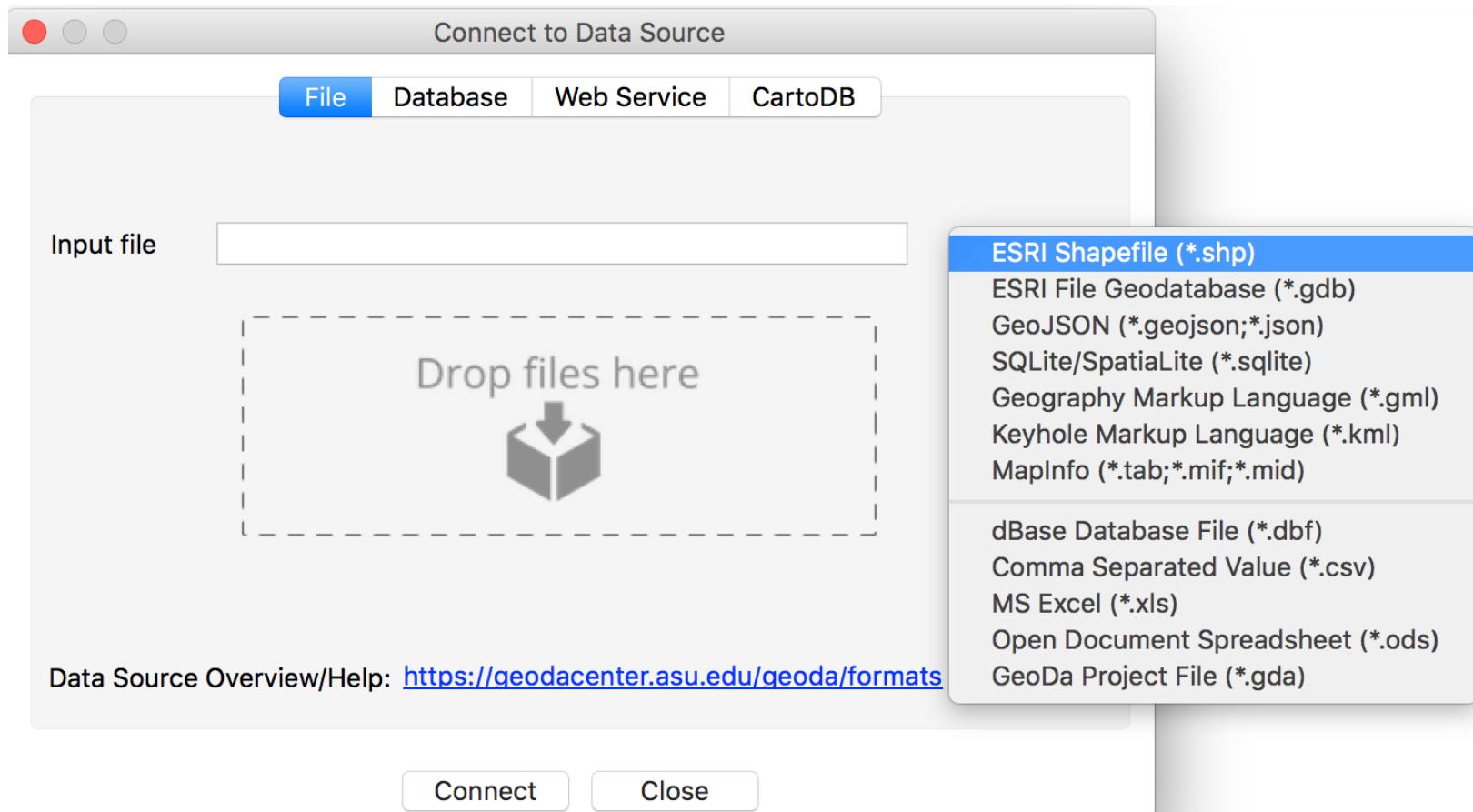
**creating/transforming variables**

**selection**



# Data Input





load different file types



Connect to Data Source

File Database Web Service CartoDB

Database Type: PostgreSQL/PostGIS Database

Database Host: Oracle Spatial Database MySQL Spatial Database

Database Port: 5432

Database/Instance Name: [REDACTED]

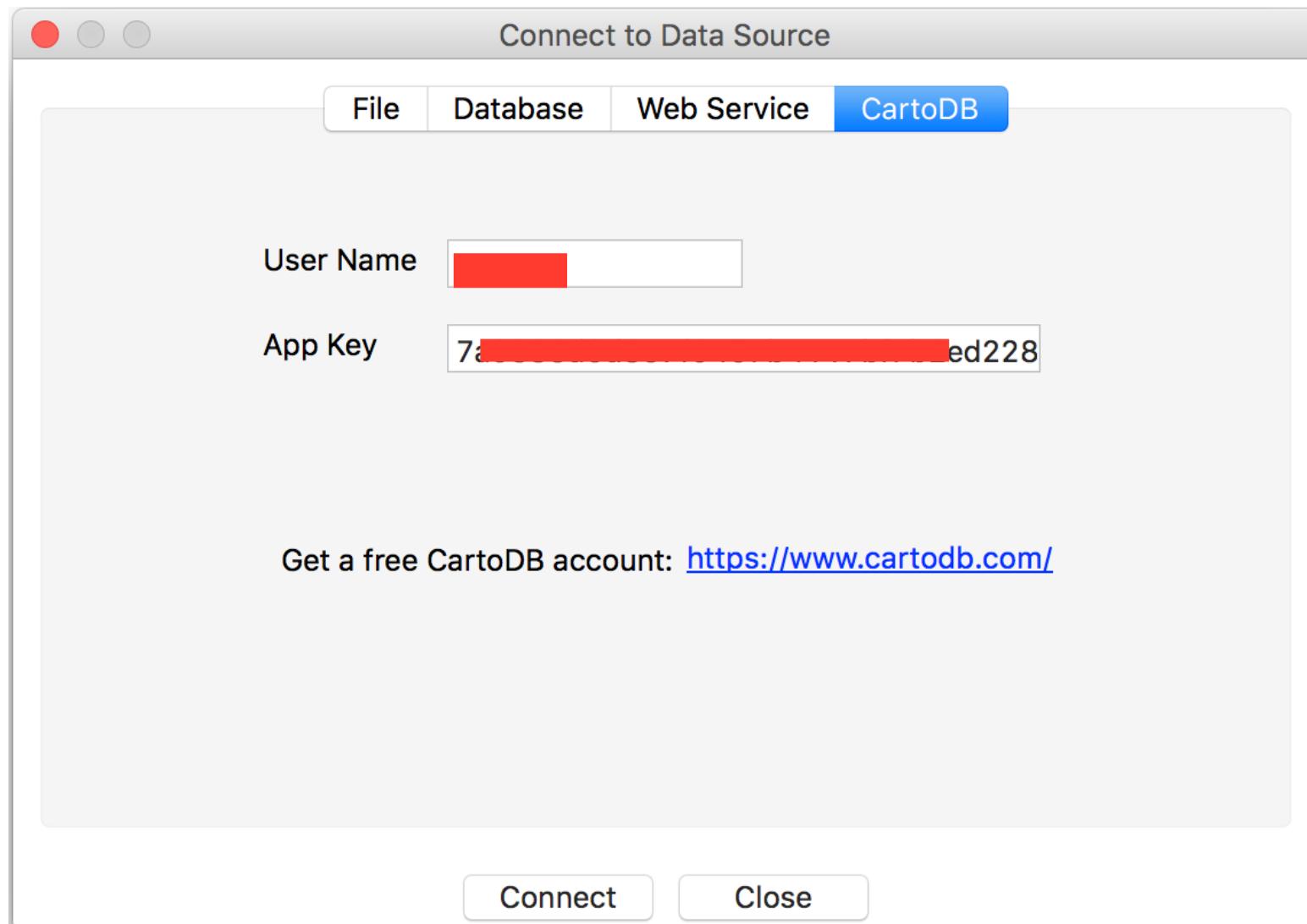
User name: luc

Password: [REDACTED]

Connect Close

connect to a spatial data base



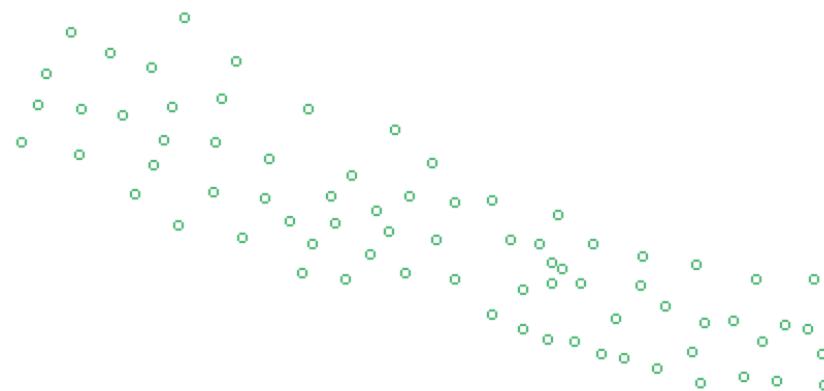
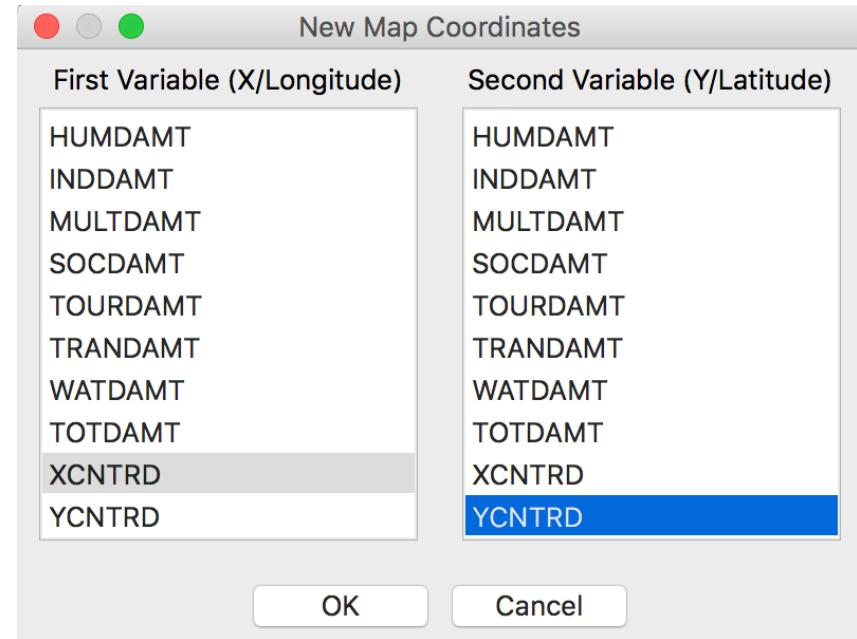
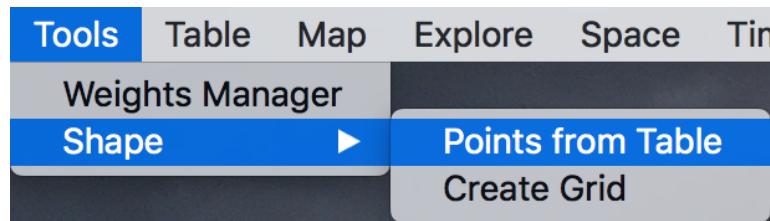


connect to CartoDB data bases



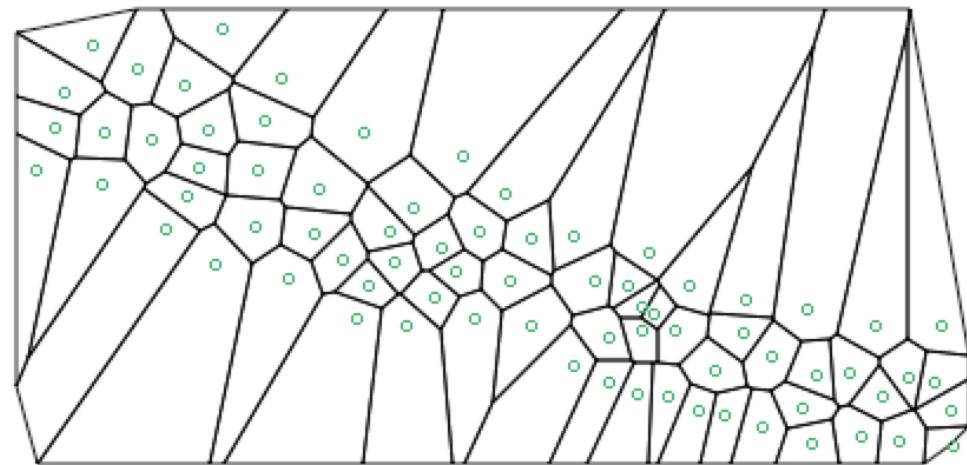
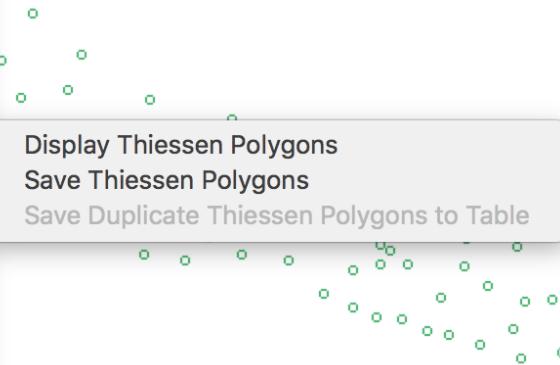
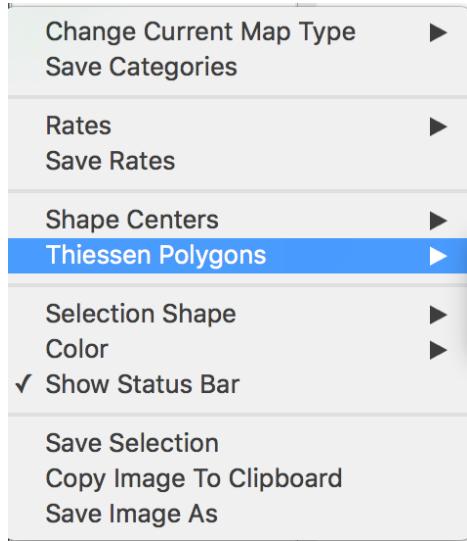
# Creating Spatial Data Sets





creating a point layer from x,y coordinates





## Thiessen polygons from points



# Data Cleanup



HLTHSOCWK
193
182
19,63
340
451

14753	182	132
15590	19,63	4619
23894	340	378

14753	182	132
15590	1963	4619
23894	340	378

edit values in table



Variable Calculation  
Add Variable  
Delete Variable(s)  
Rename Variable "GNI"  
**Edit Variable Properties**

GNI	integer
HLTHSOCWK	<b>string</b>
HOTELREST	integer

HLTHSOCWK	strin ▾
HOTELREST	real
MANUF	inte...
MINQUAR	date
OTHSVCE	<b>string</b>

HLTHSOCWK
193
182
1963
340
451

GNI	integer
HLTHSOCWK	integer
HOTELREST	integer

changing the variable type



# Creating/Transforming Variables



	BOYRATIO
1	0.116546
2	0.067674
3	0.082438
4	0.076907
5	0.055722

Variable Calculation

- Add Variable
- Delete Variable(s)
- Rename Variable "POPULATION"
- Edit Variable Properties

Variable Calculation

Special    Univariate    **Bivariate**    Spatial Lag    Rates

Result    Add Variable    =    Variable / Constant    Operator    Variable / Constant

BOYRATIO    BOYG1\_5    DIVIDE    POPULATION

BOYRATIO = BOYG1\_5 / POPULATION

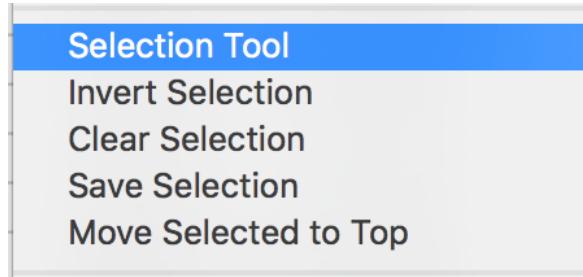
Apply    Close

## variable calculation



# Selection





Selection Tool

Selection

New Selection    Select From Current Selection    Append To Current Selection

Selection Variable: POVINDEX   Time:

Select All In Range:  <= POVINDEX <=

Select All Undefined: POVINDEX  
Invert Selection  
Add Neighbors To Selection   Weights:   
Clear Selection

Assign Values to Currently Selected / Unselected

Add Variable  
Target Variable:  Time:   
 Selected =     Unselected =

Apply

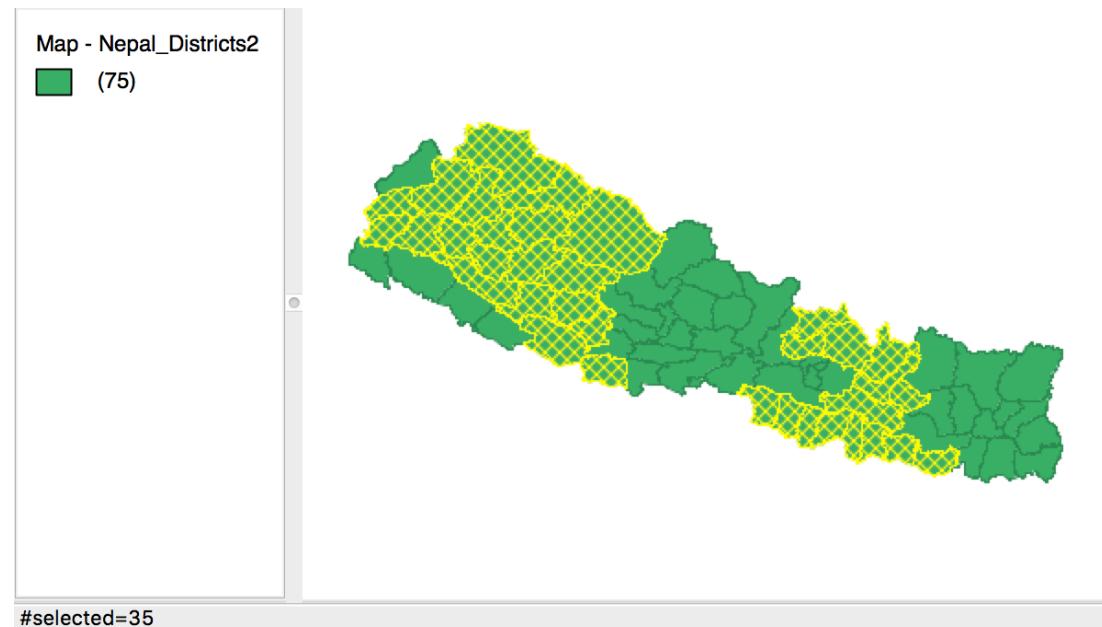


# selection tool



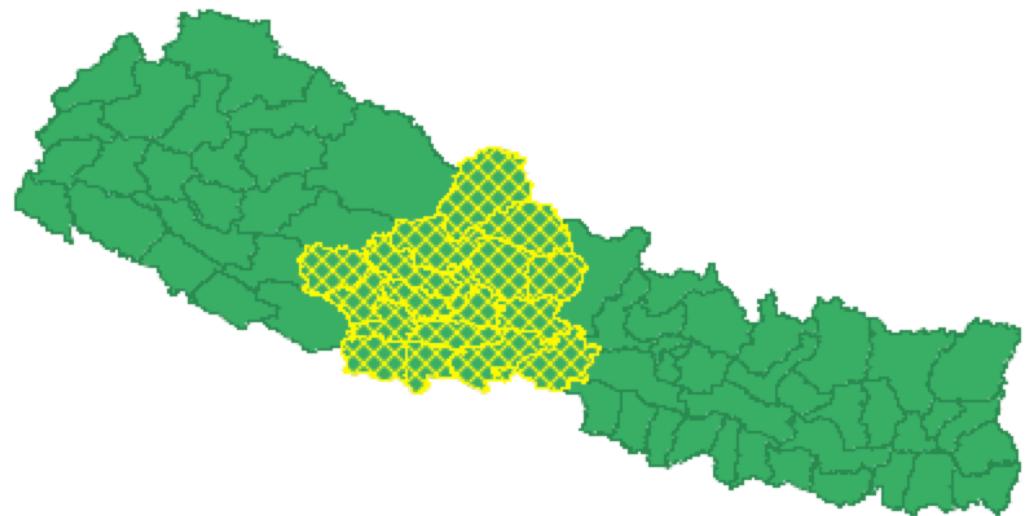
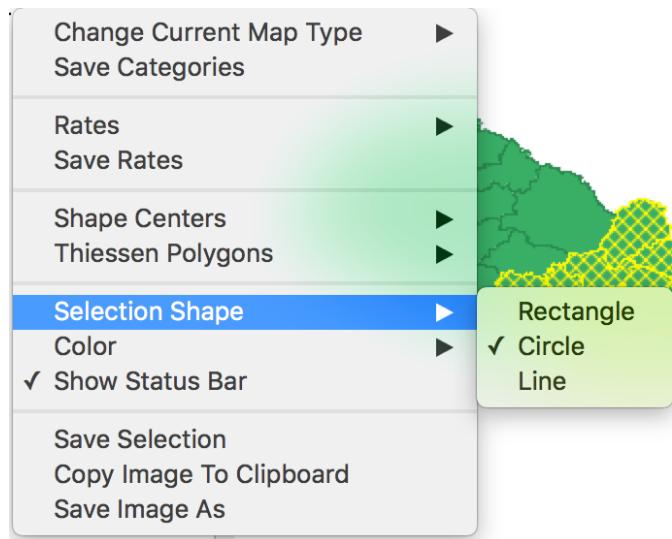
Table - Nepal\_Districts2

	DEPECProv	POVINDEX	PCINC	PCINCPPP	PCINCMP	MALKIDS
9	23.340000	19.180000	1172	1894	84715	16.2000
10	24.710000	35.660000	672	1086	48586	37.1000
11	28.570000	42.240000	941	1520	67973	45.0000
12	32.930000	38.030000	687	1110	49657	46.6000
13	27.740000	41.720000	580	938	41951	43.6000
14	33.070000	35.700000	571	922	41256	44.0000
15	26.590000	44.750000	421	681	30436	43.6000
16	32.400000	36.350000	588	951	42524	44.4000
17	24.360000	43.860000	501	809	36177	37.7000
18	35.910000	37.950000	509	822	36767	36.3000
19	21.780000	40.090000	916	1480	66175	39.6000
20	23.320000	28.440000	873	1410	63079	25.1000
21	20.910000	36.370000	757	1223	54706	37.2000
22	22.290000	46.430000	468	757	33840	39.7000
23	28.880000	24.800000	951	1537	68748	42.0000



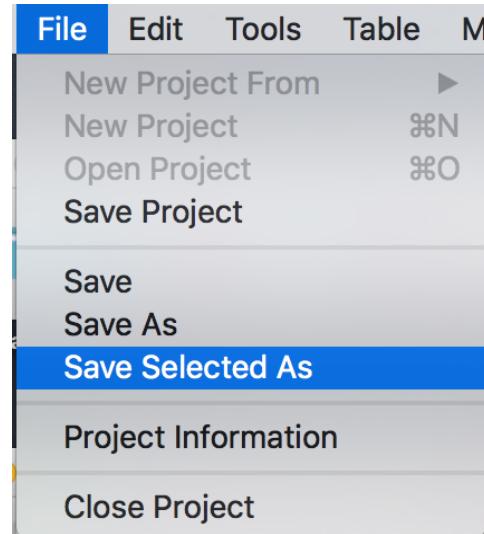
selection in table and map - linking





## selection by shape





save selected observations as new layer



# Geovisualization



Copyright © 2016 by Luc Anselin, All Rights Reserved



maps

linking and brushing

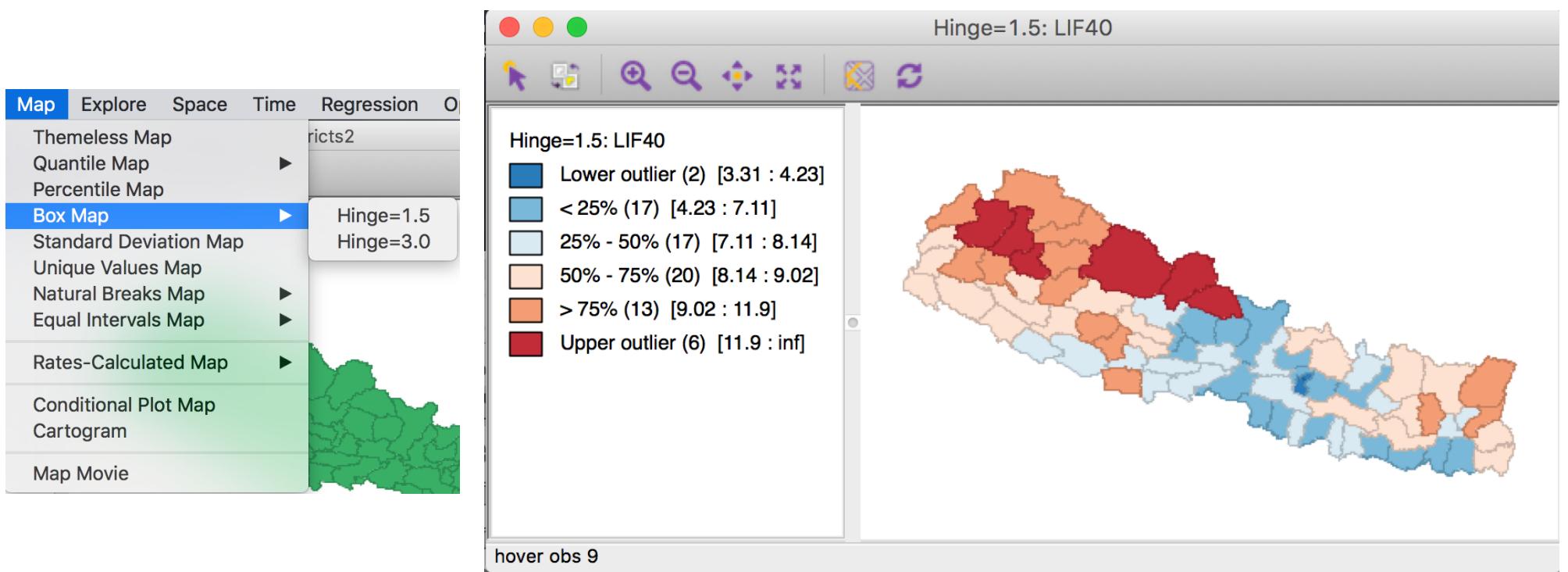
cartogram and map movie

category editor



# Maps

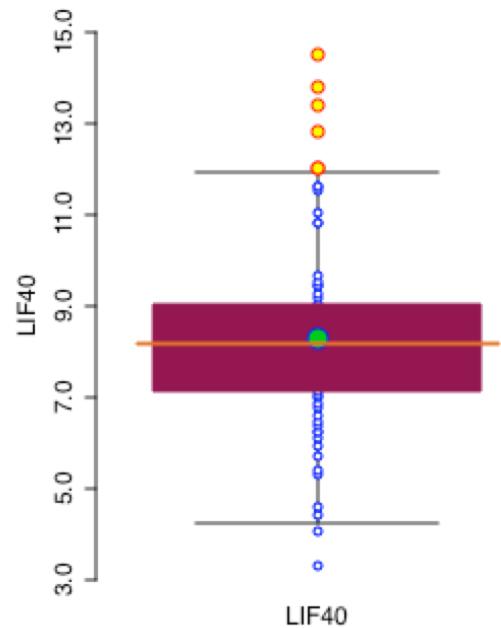




## map menu - box map

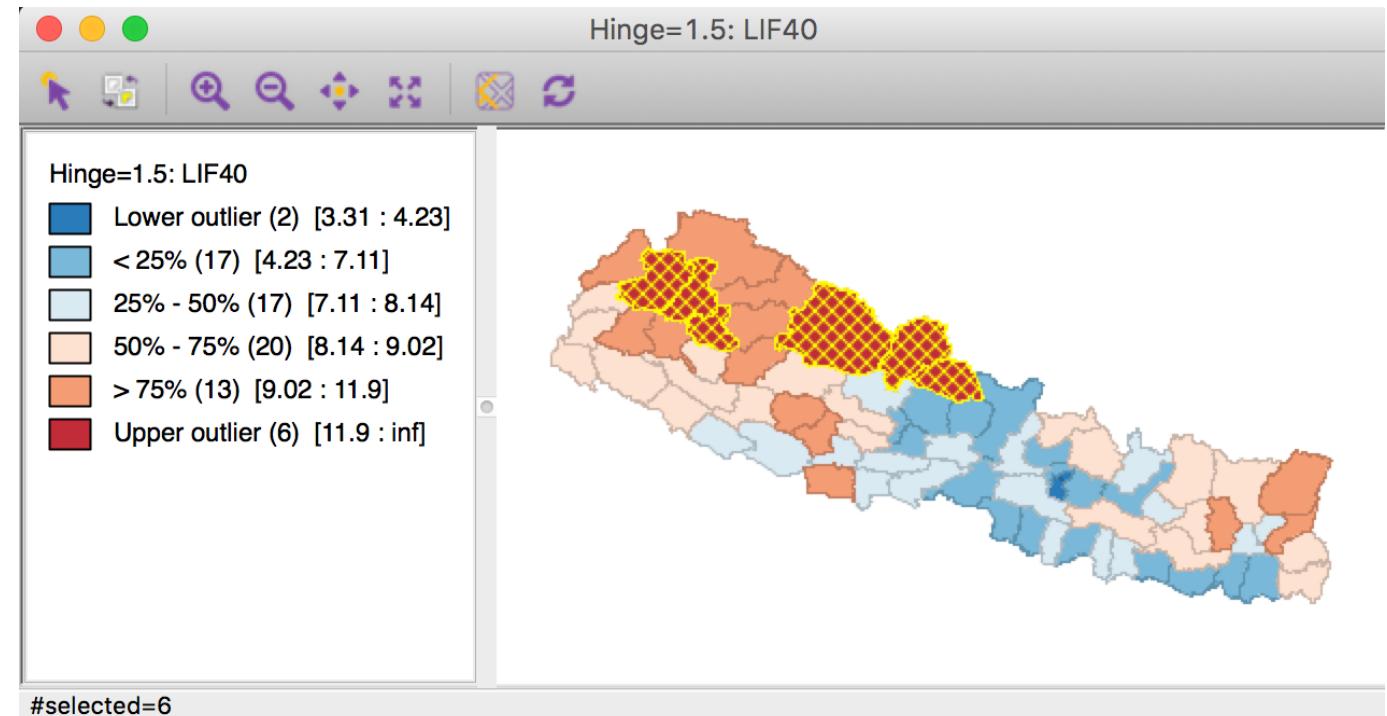


● ○ ● Box Plot (Hinge=1.5): LIF40



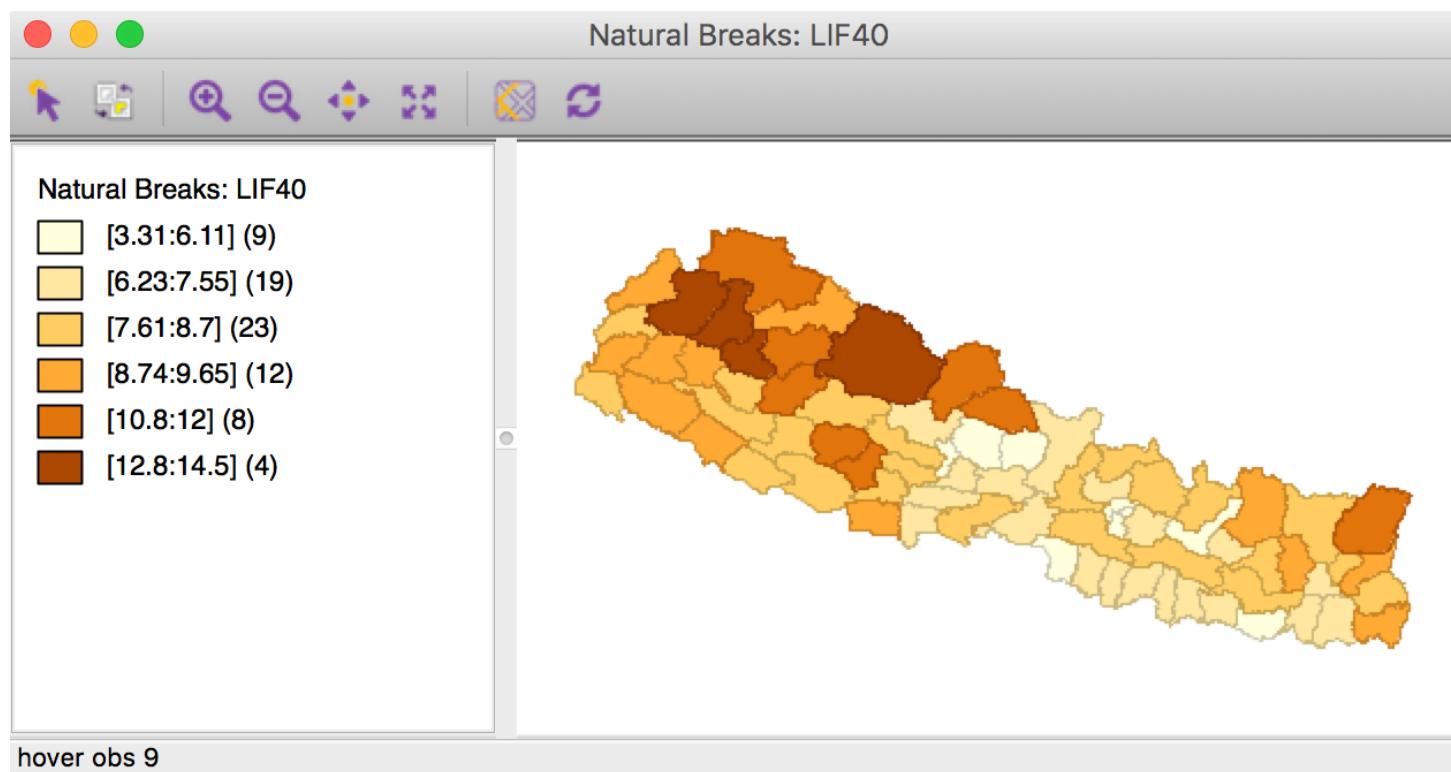
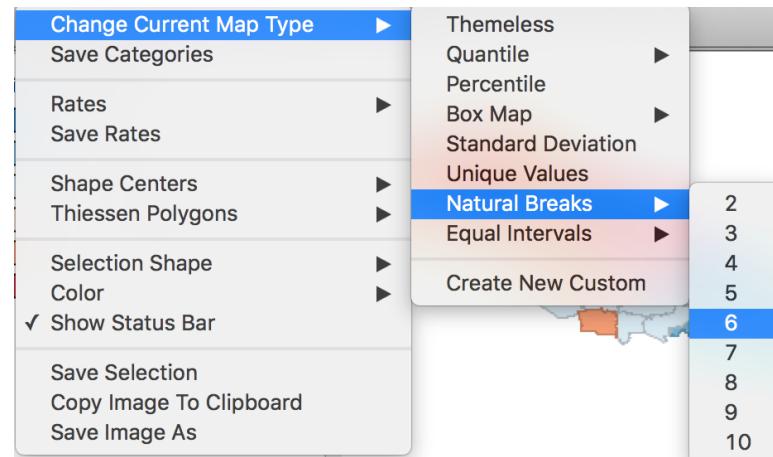
LIF40  
min 3.31  
max 14.48  
Q1 7.105  
median 8.14  
Q3 9.02  
IQR 1.915  
mean 8.277  
s.d. 2.175

#selected=6



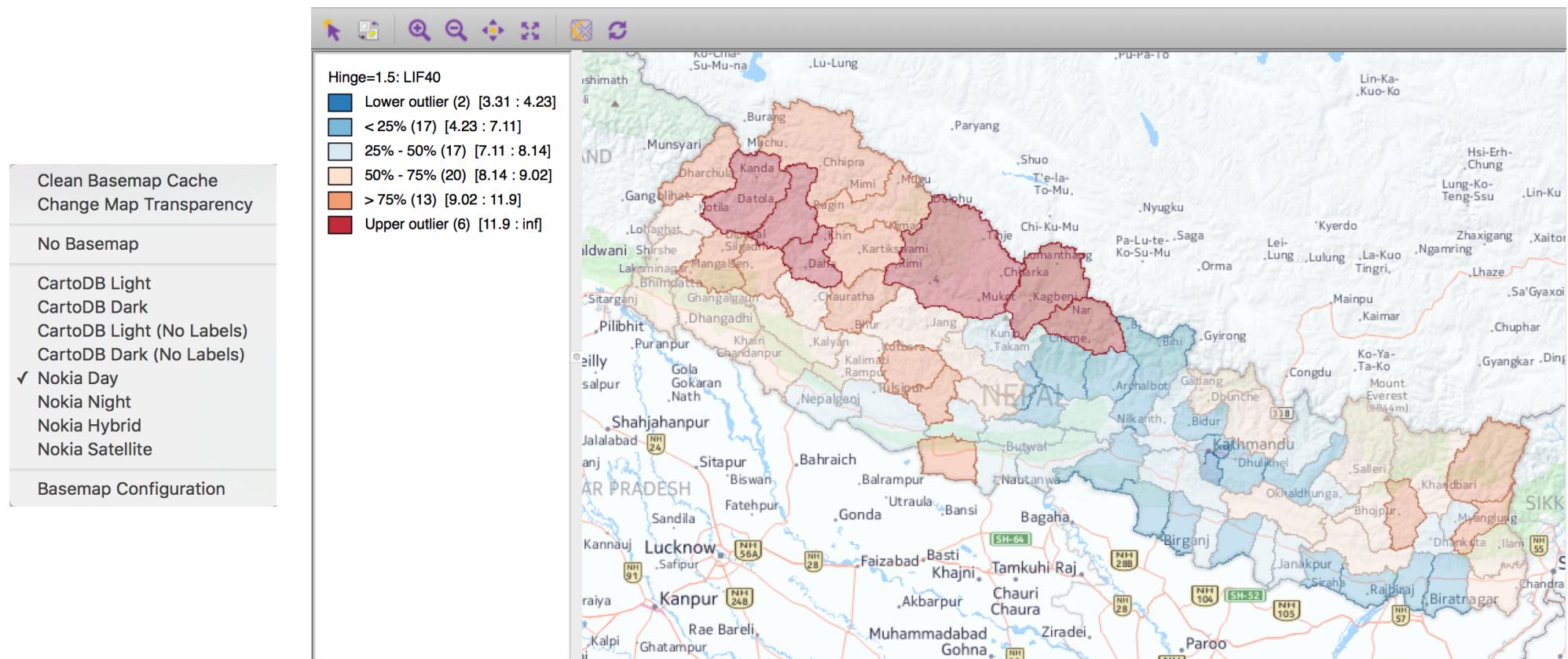
box map principle





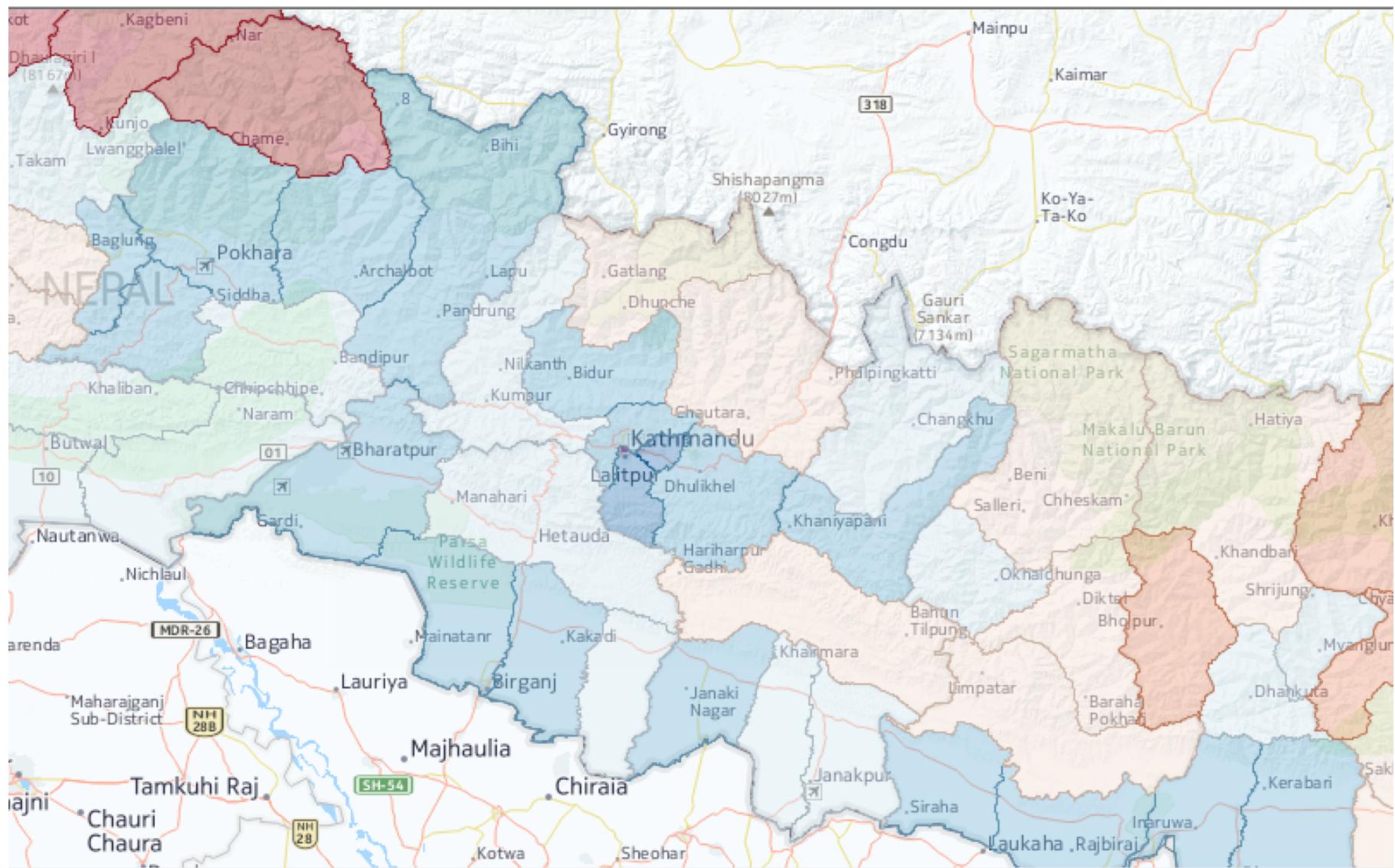
## natural breaks map





# basemap





zoom in

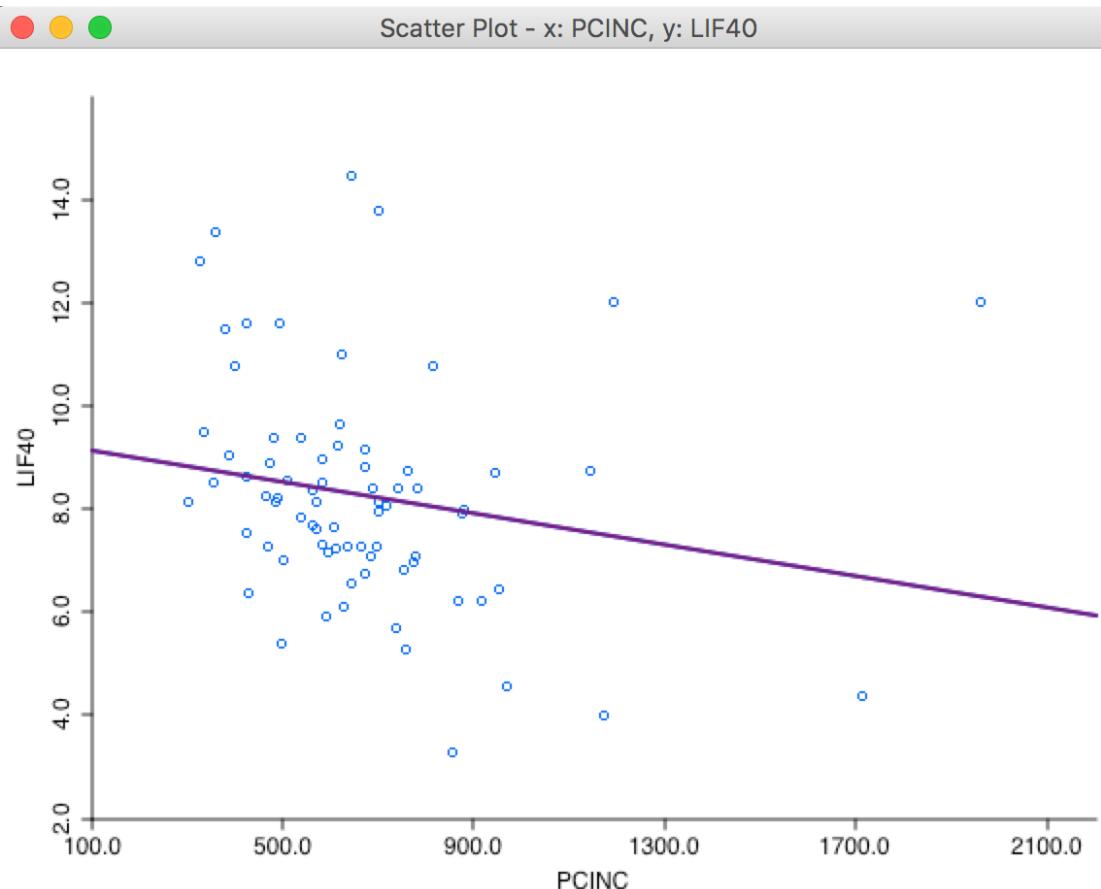
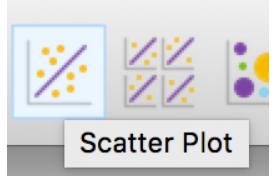
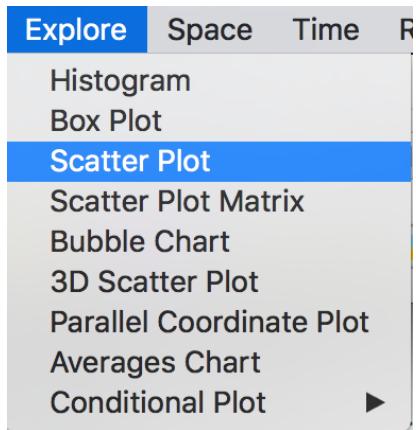


# Linking and Brushing



Copyright © 2016 by Luc Anselin, All Rights Reserved

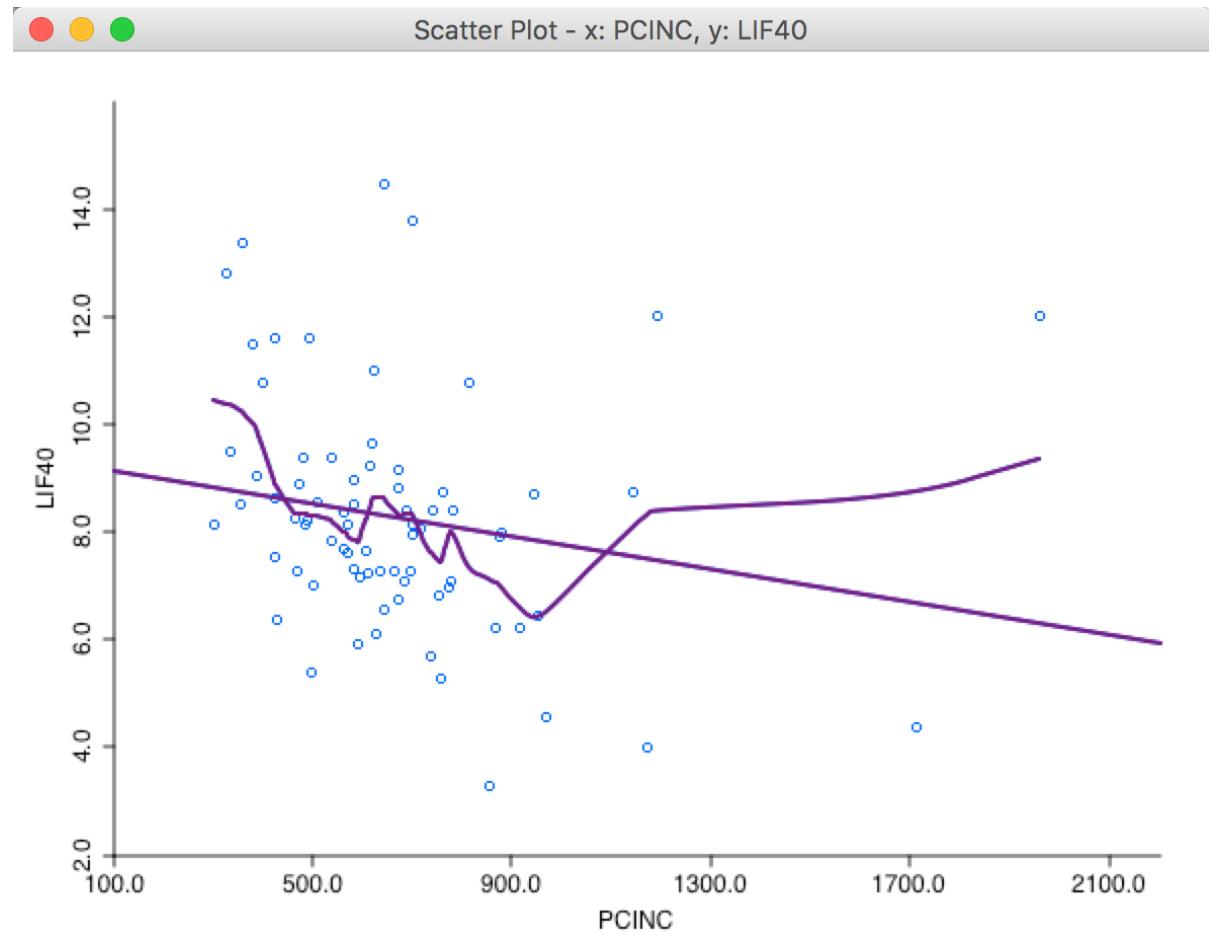
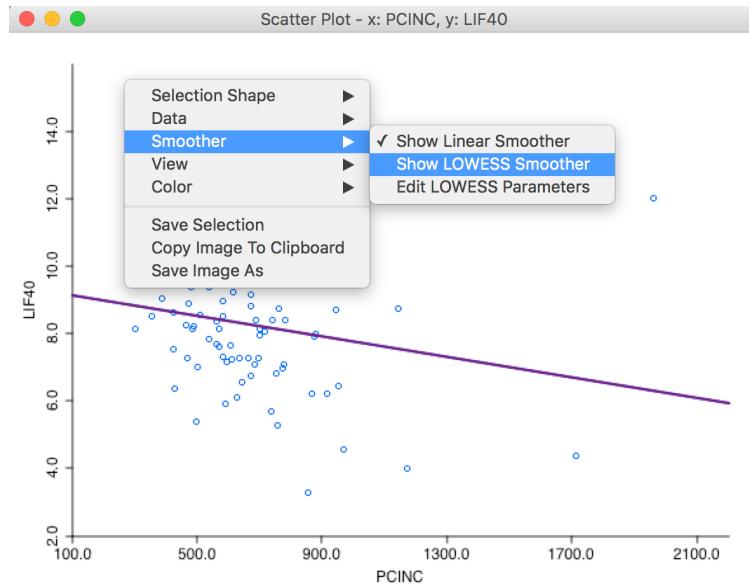




#obs	R <sup>2</sup>	const a	std-err a	t-stat a	p-value a	slope b	std-err b	t-stat b	p-value b
75	0.0366	9.29	0.658	14.1	0	-0.00152	0.000911	-1.67	0.0999
0	0	0	0	0	0	0	0	0	0
75	0.0366	9.29	0.658	14.1	0	-0.00152	0.000911	-1.67	0.0999

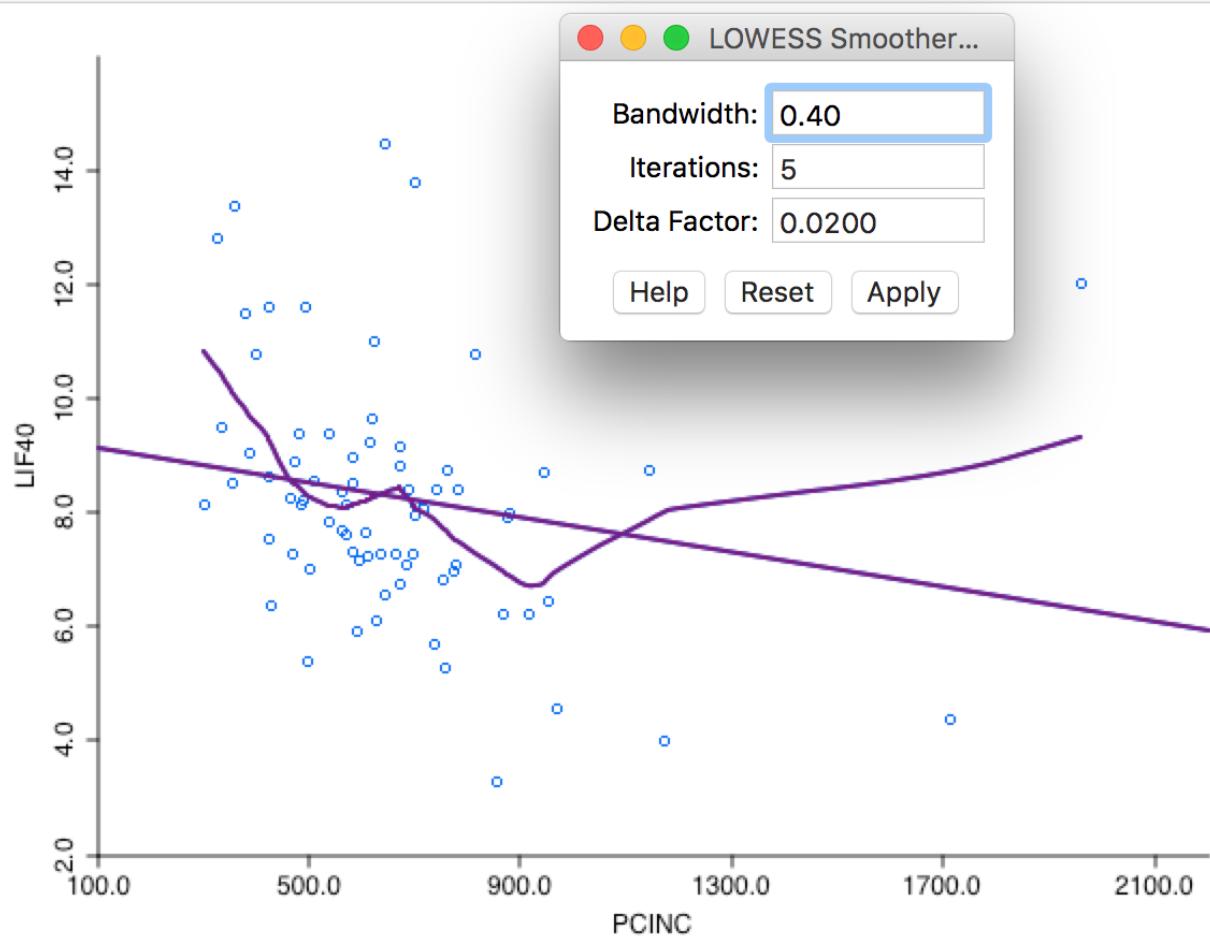
# scatter plot





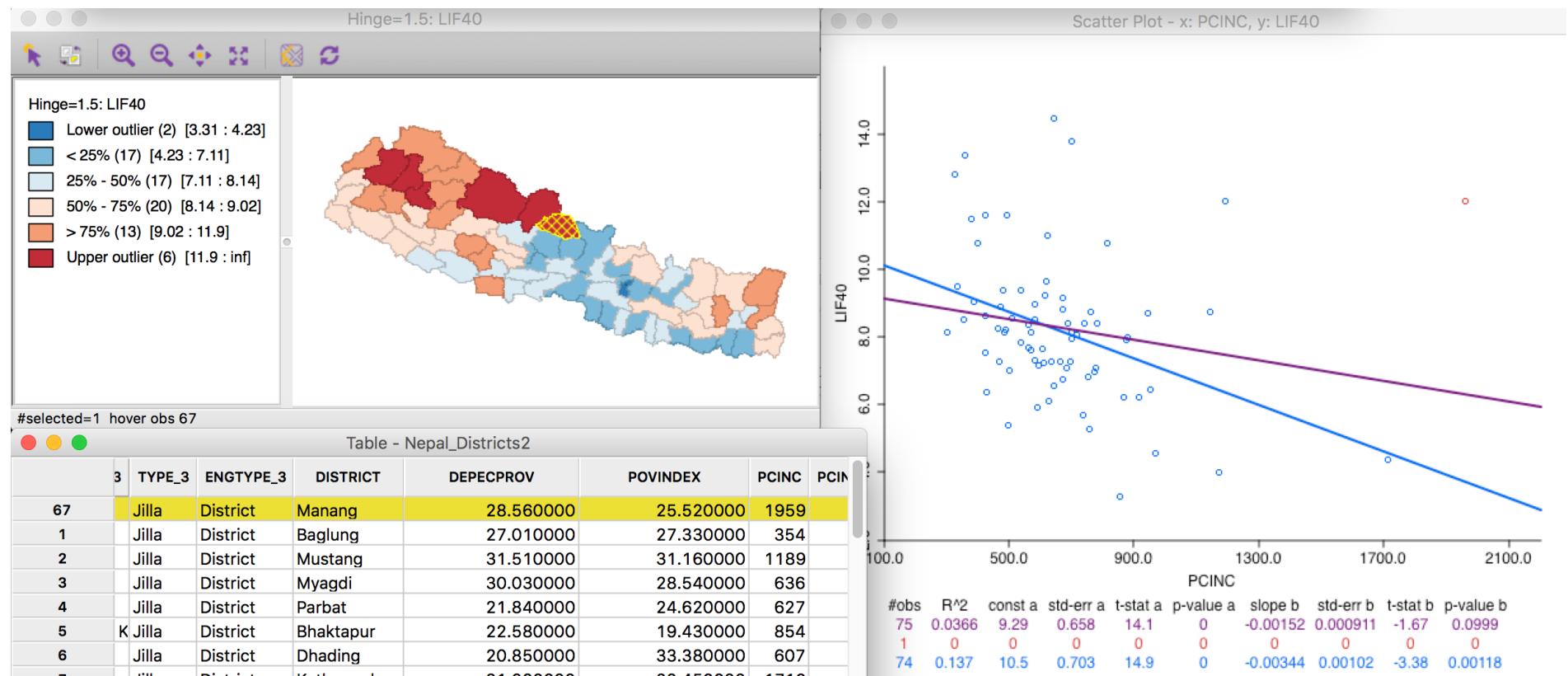
## lowess smoother





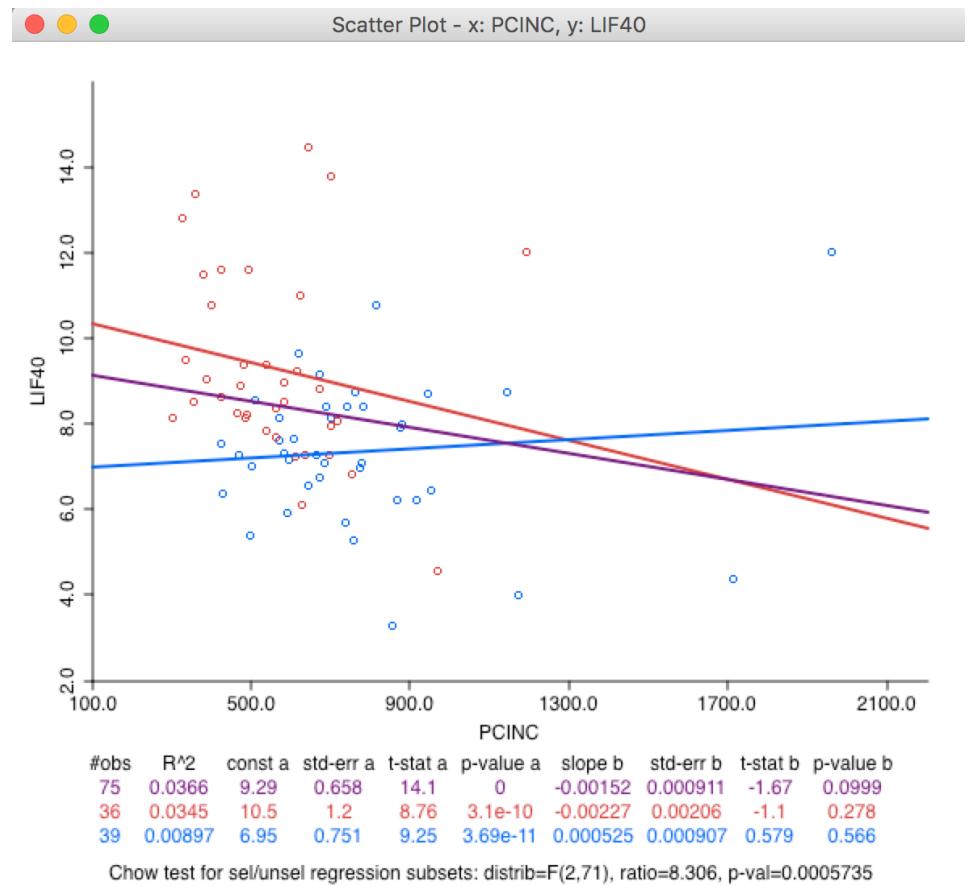
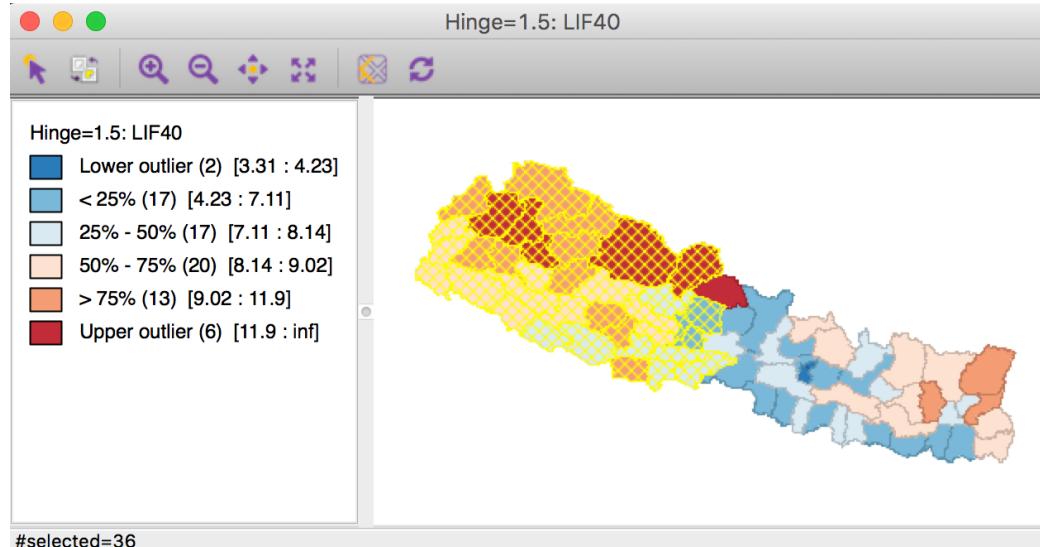
customize lowess bandwidth





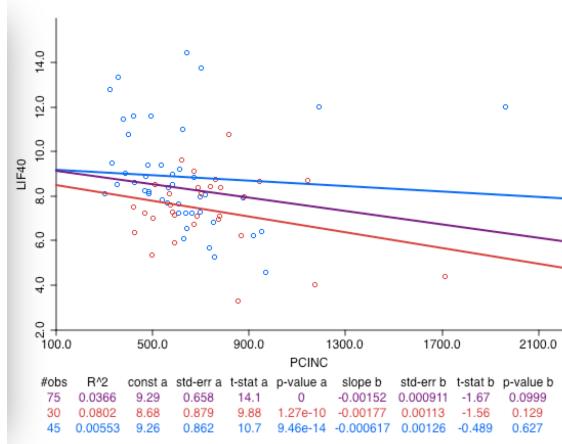
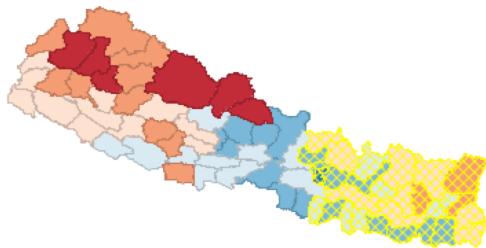
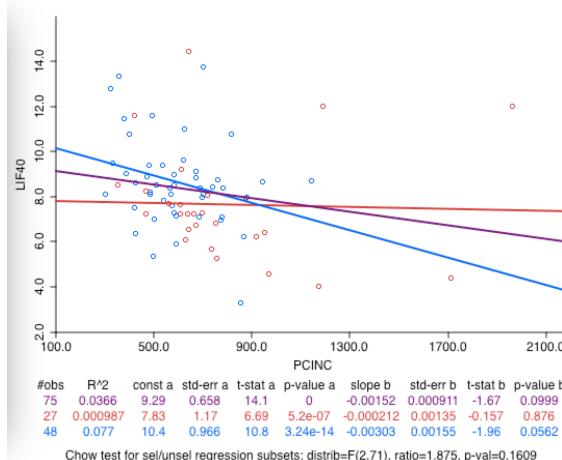
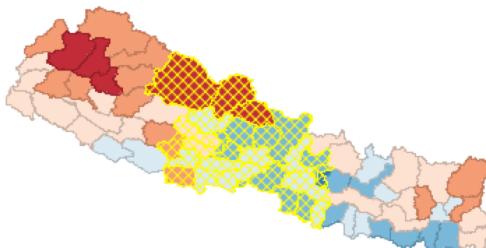
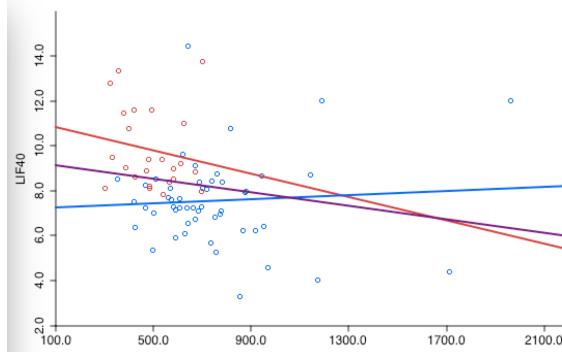
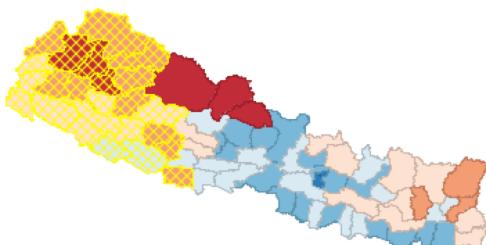
## effect of outlier - linking





## test for structural stability



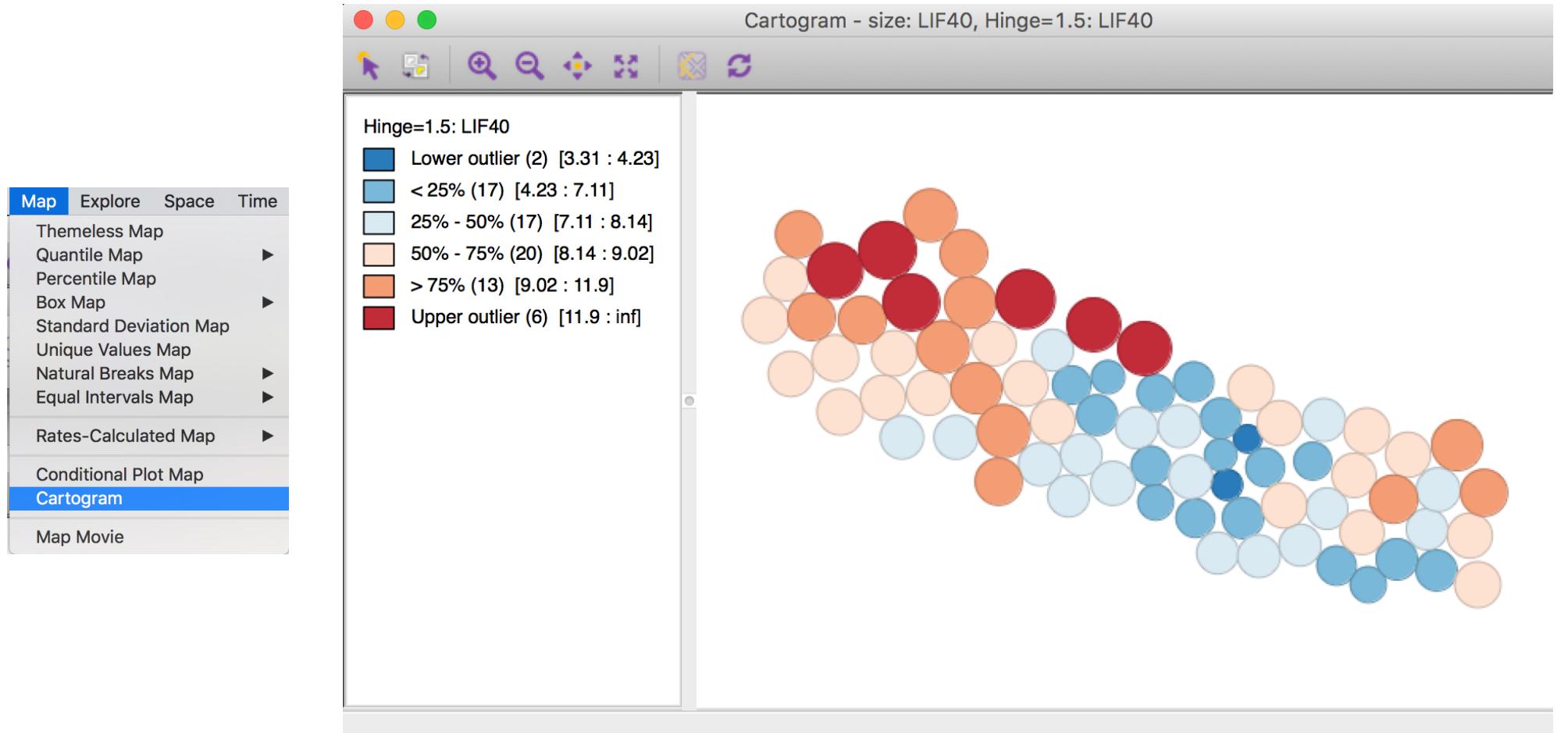


# Cartogram and Map Movie



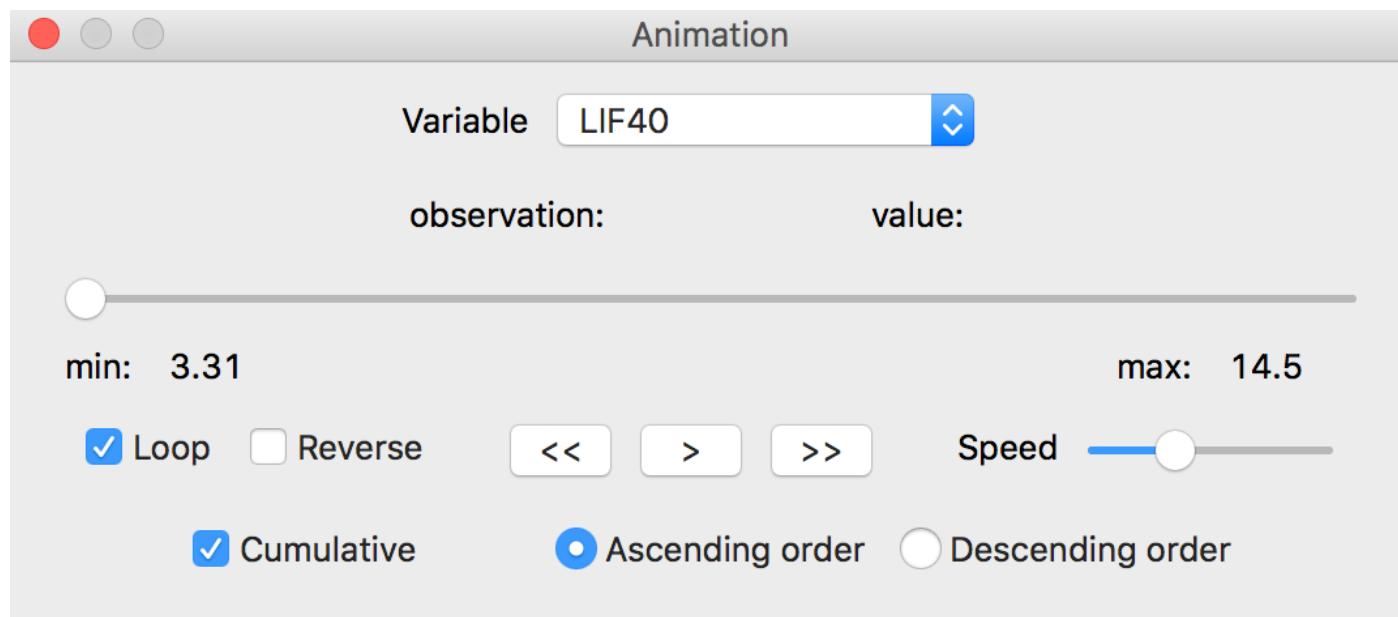
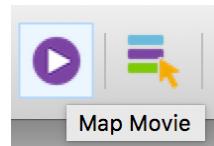
Copyright © 2016 by Luc Anselin, All Rights Reserved





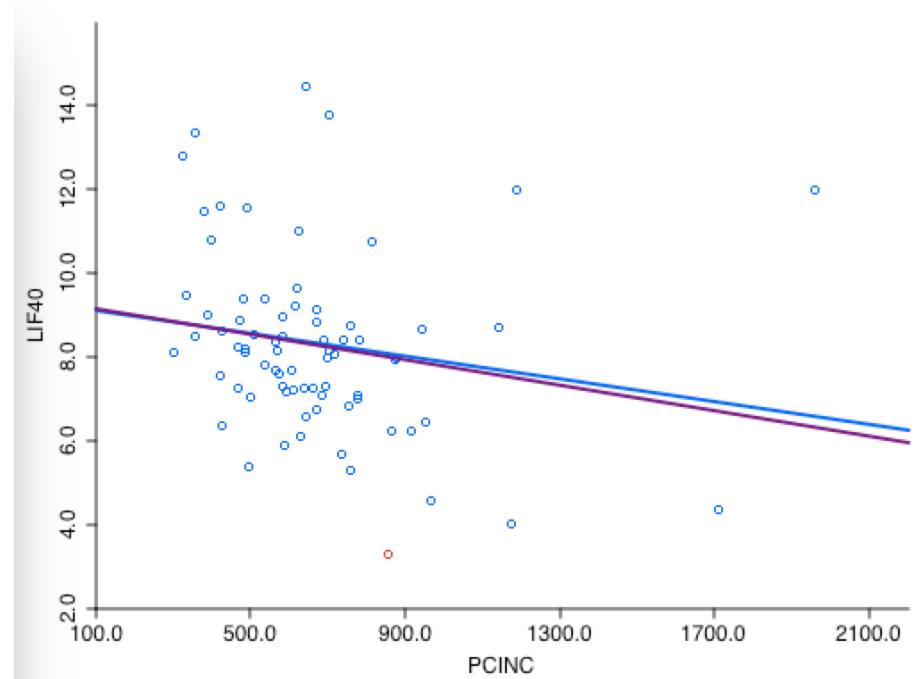
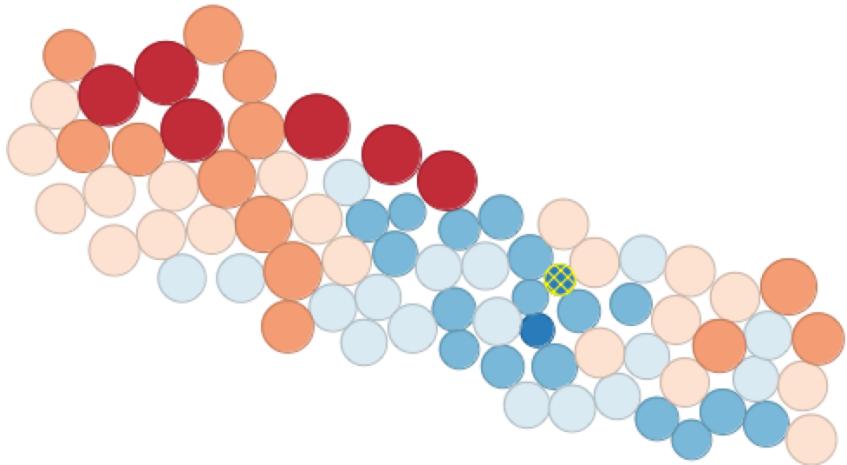
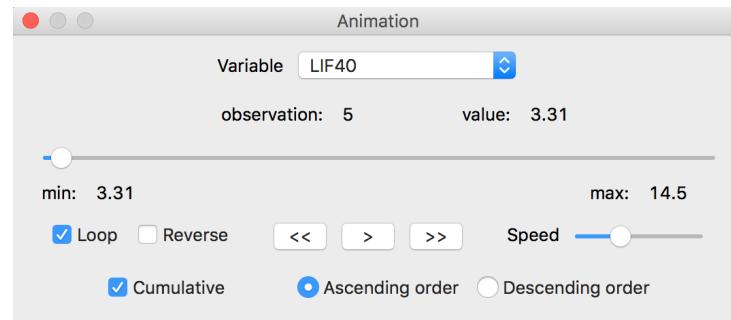
cartogram





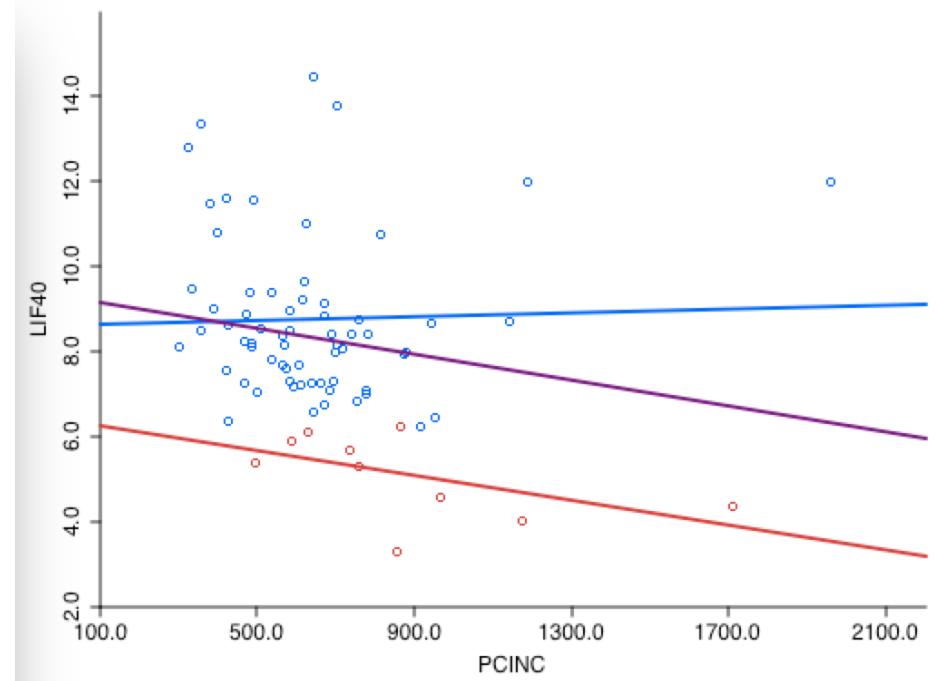
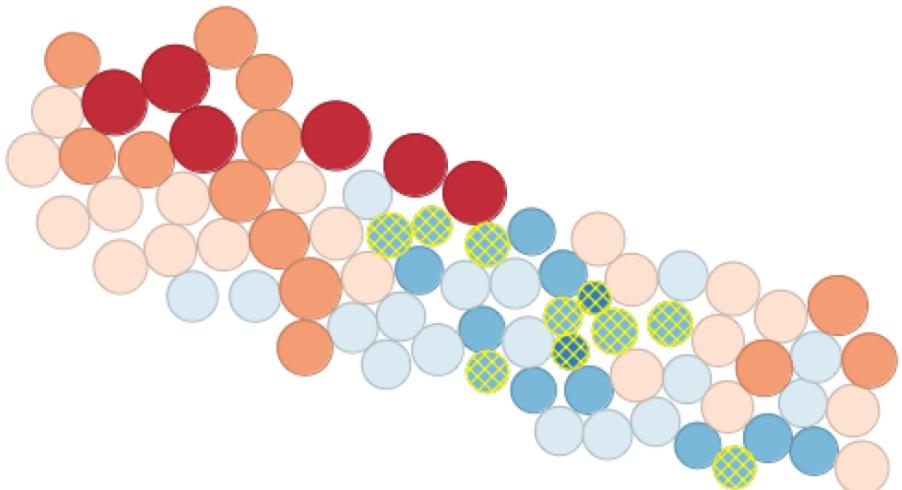
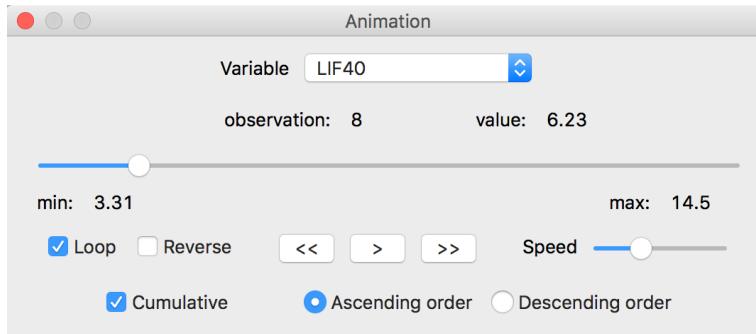
## animation - map movie





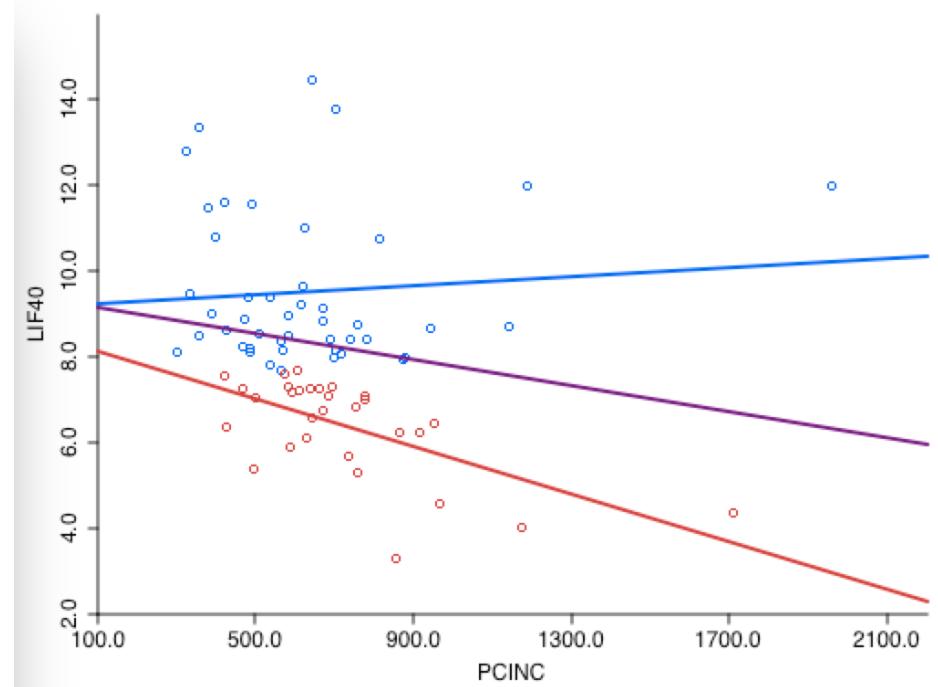
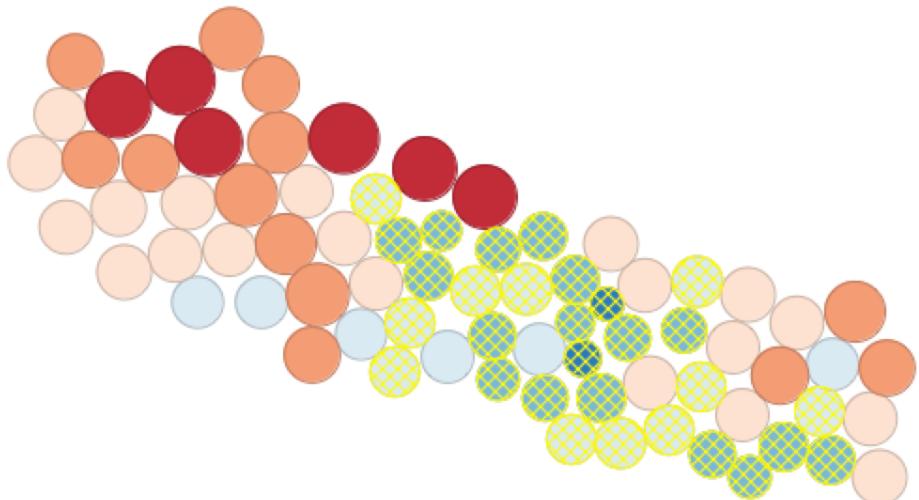
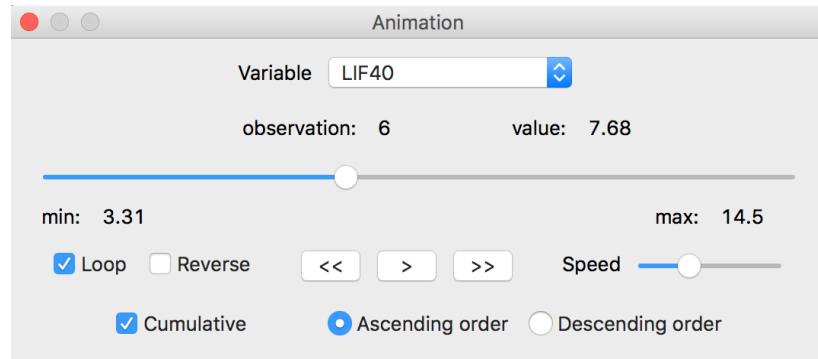
## animation - one observation





## animation - ten observations





## animation - thirty observations

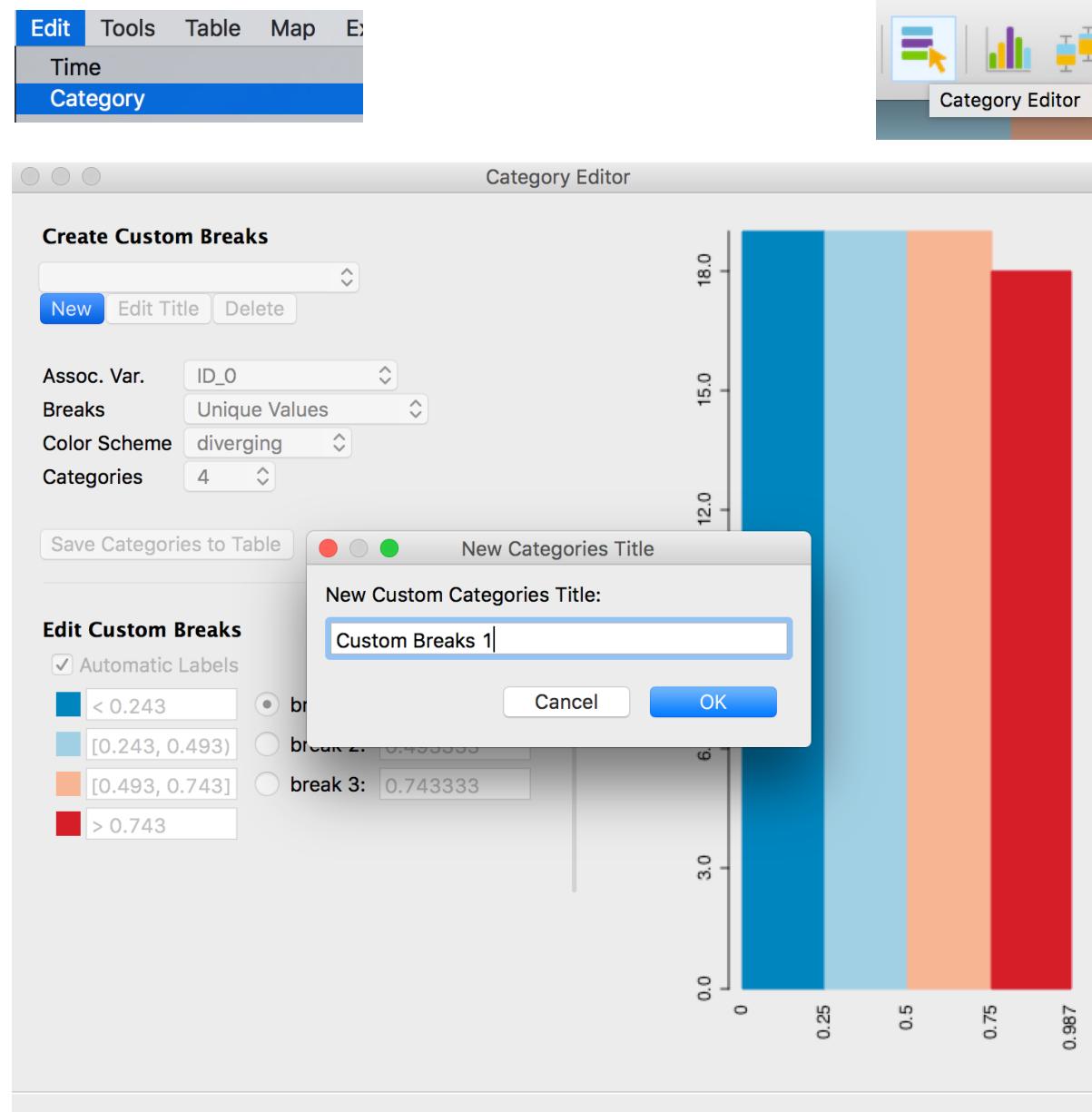


# Category Editor



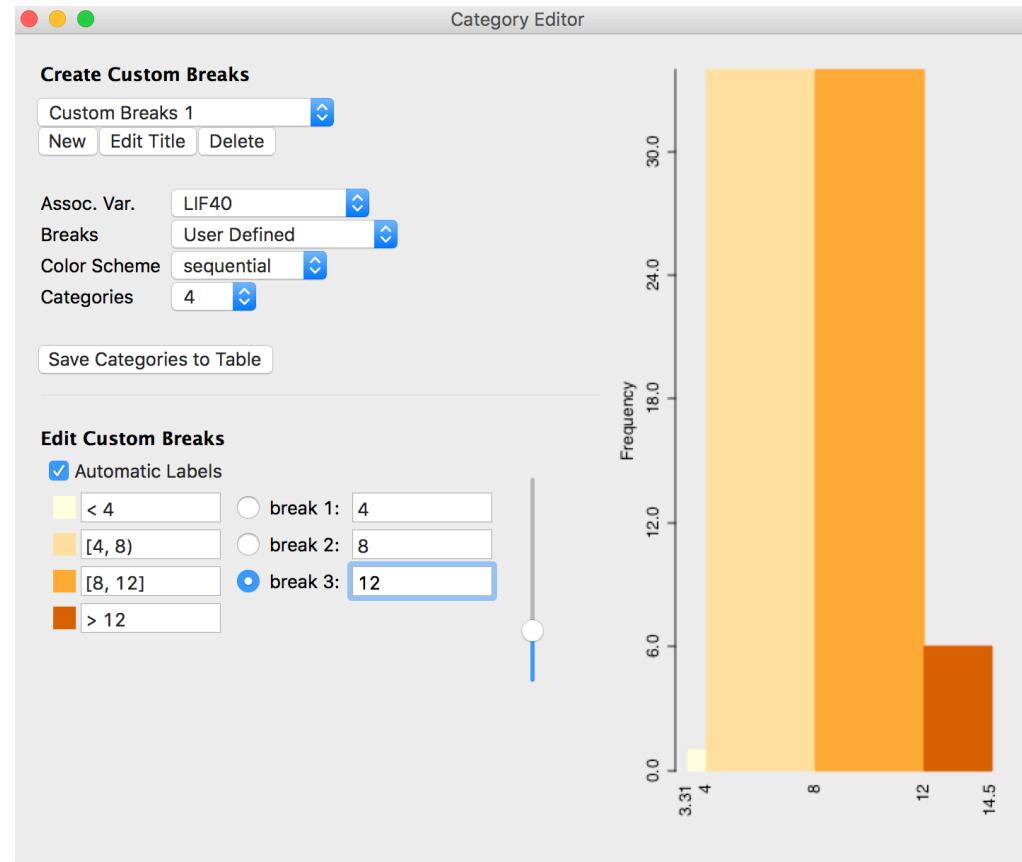
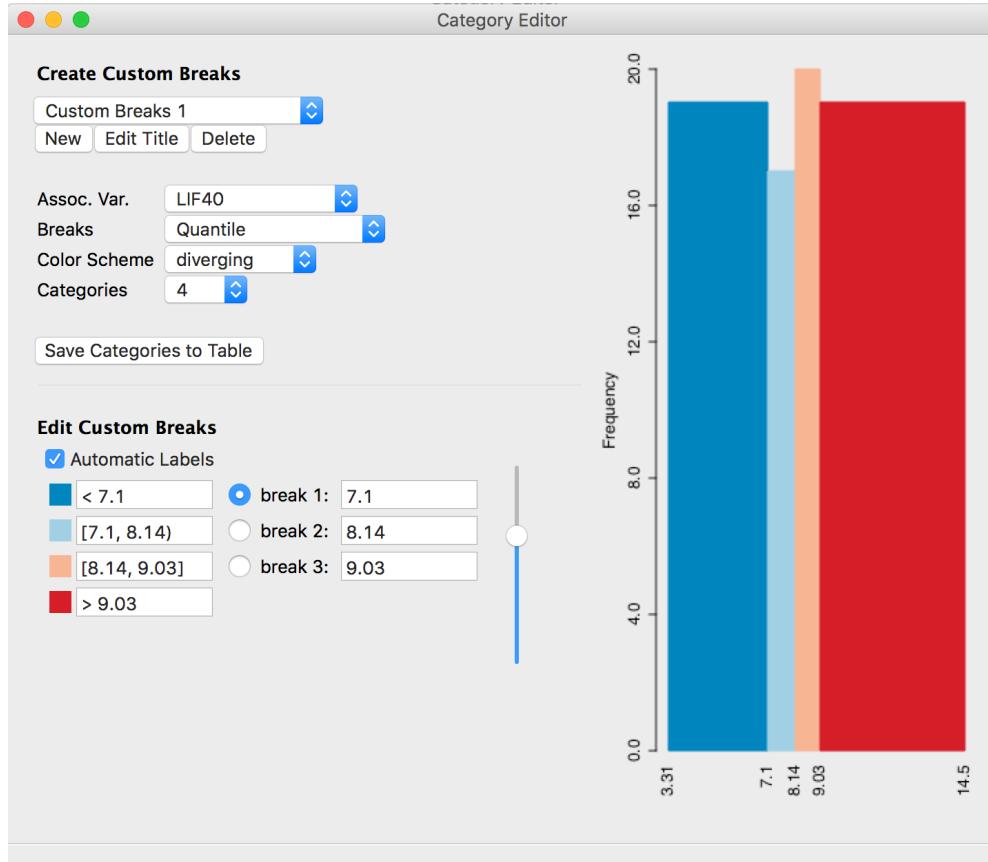
Copyright © 2016 by Luc Anselin, All Rights Reserved





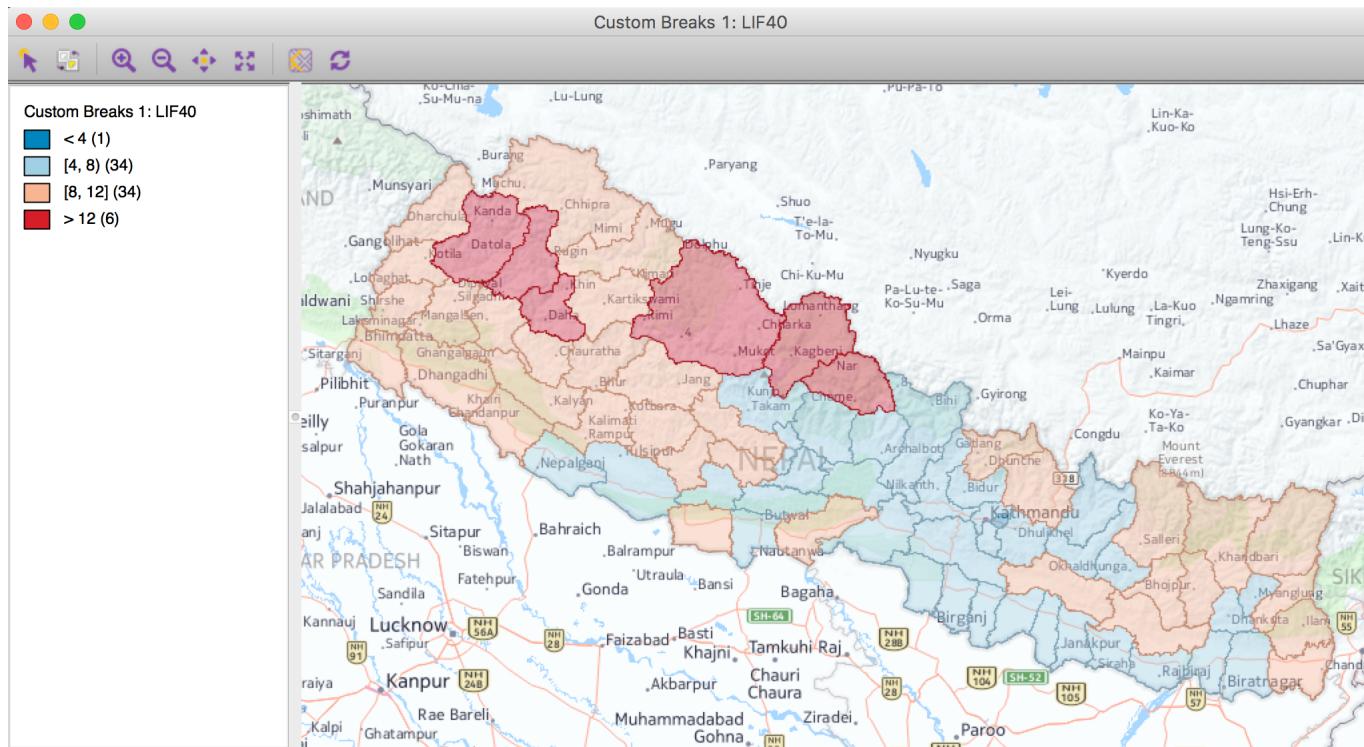
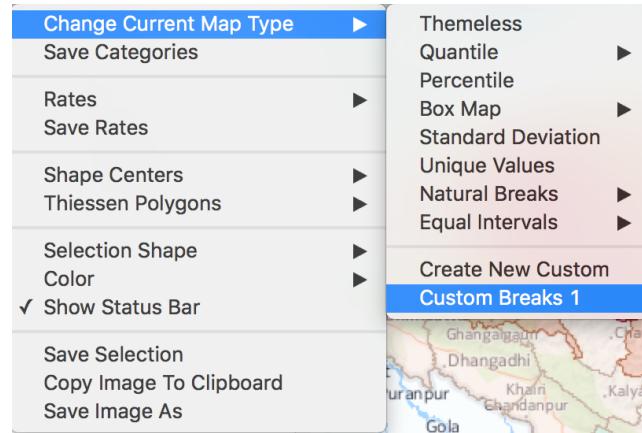
## category editor





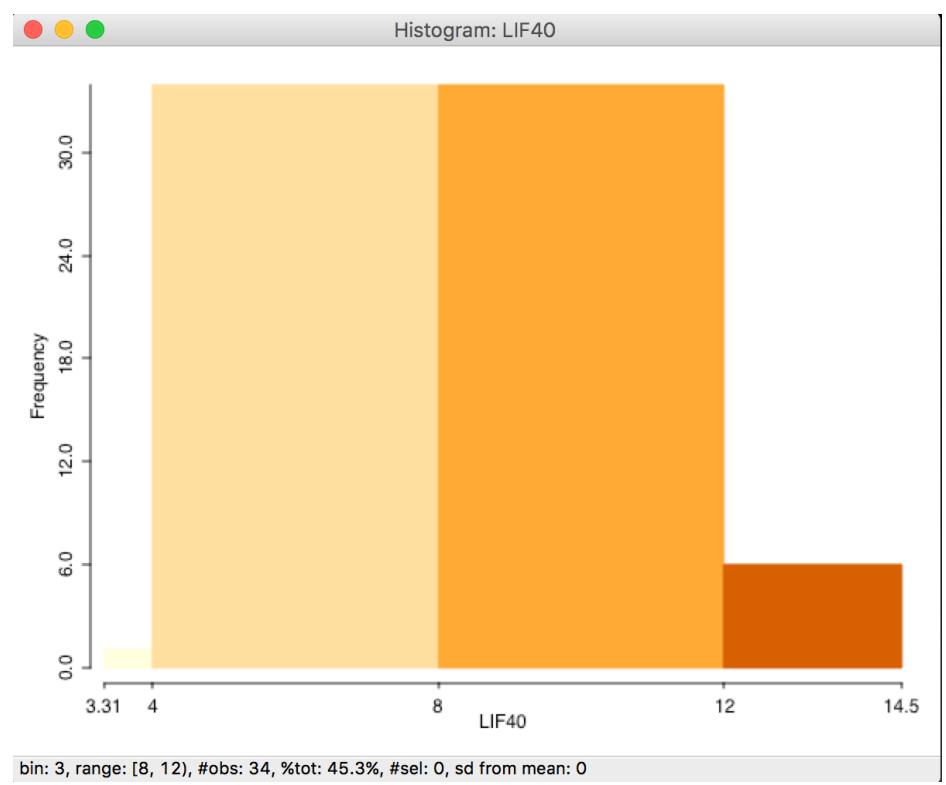
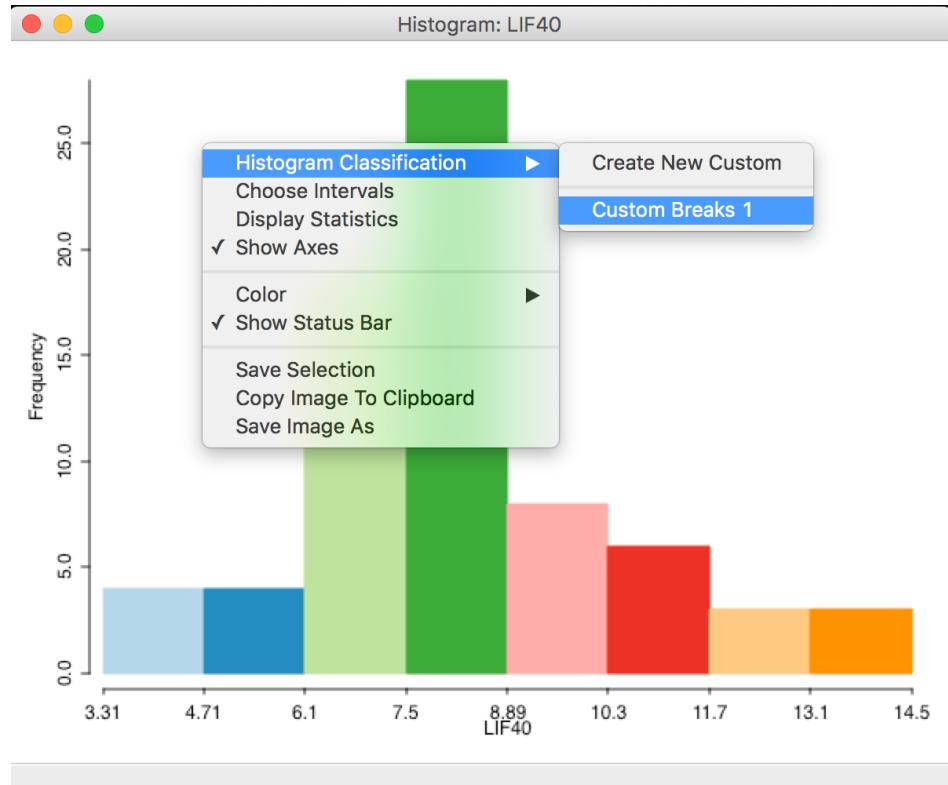
# category editor - setting break points





## custom categories in map





## histogram with custom breaks



# Multivariate EDA



Copyright © 2016 by Luc Anselin, All Rights Reserved



scatter plot matrix

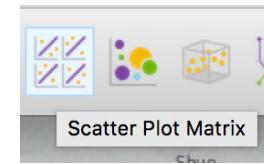
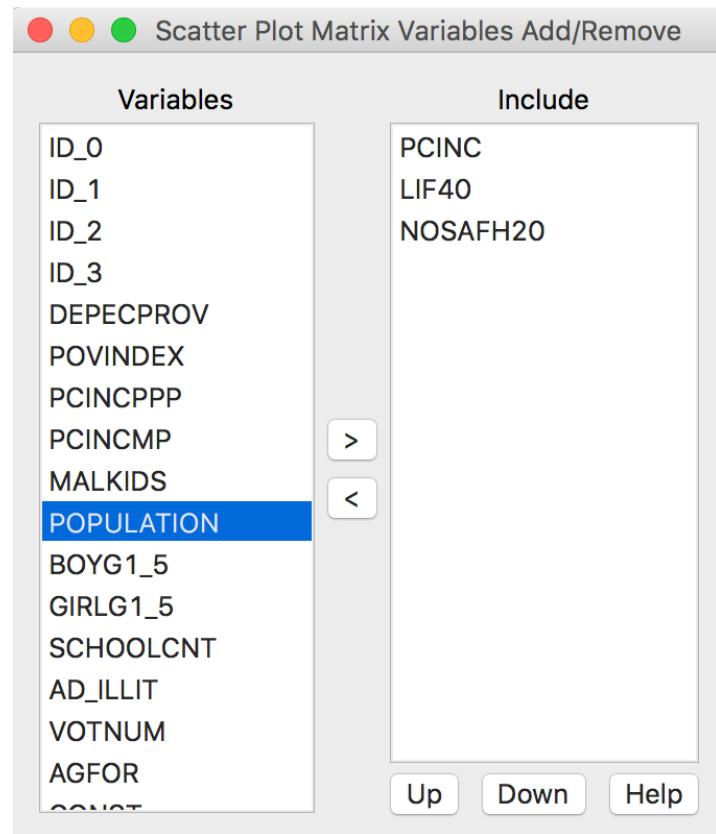
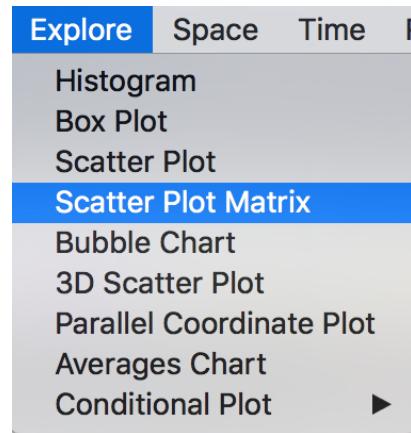
parallel coordinate plot

conditional plots



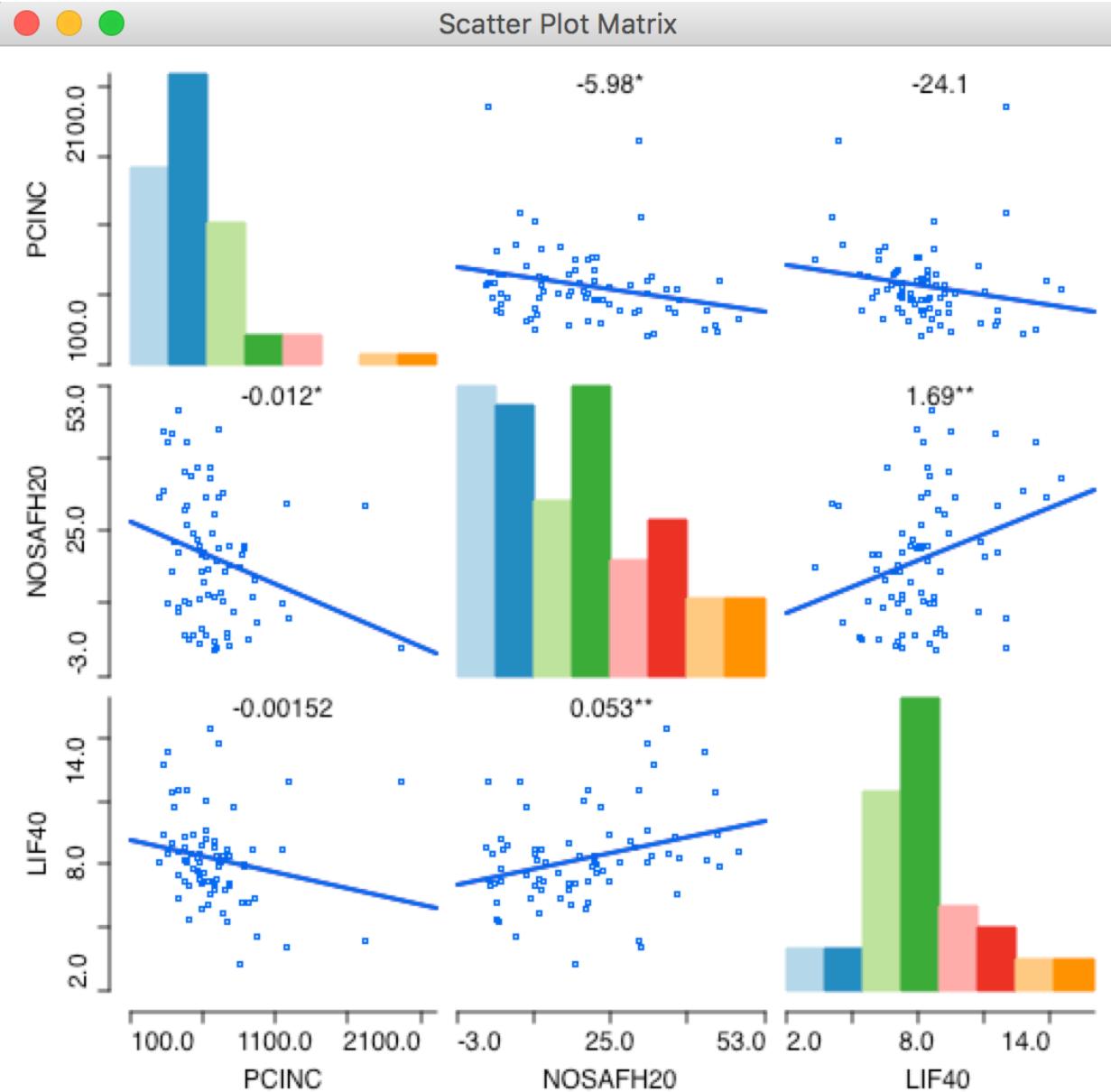
# Scatter Plot Matrix





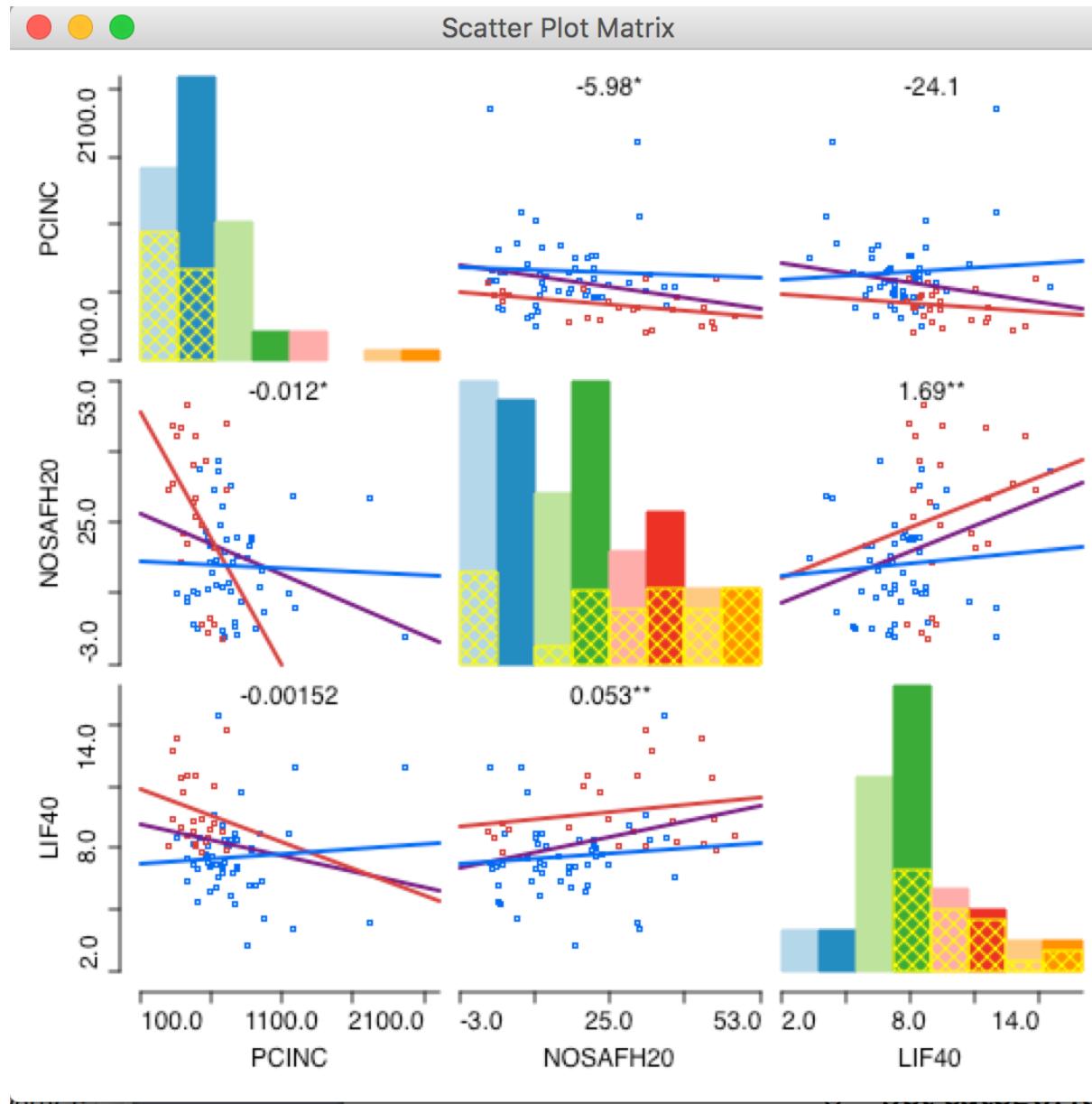
## scatter plot matrix setup





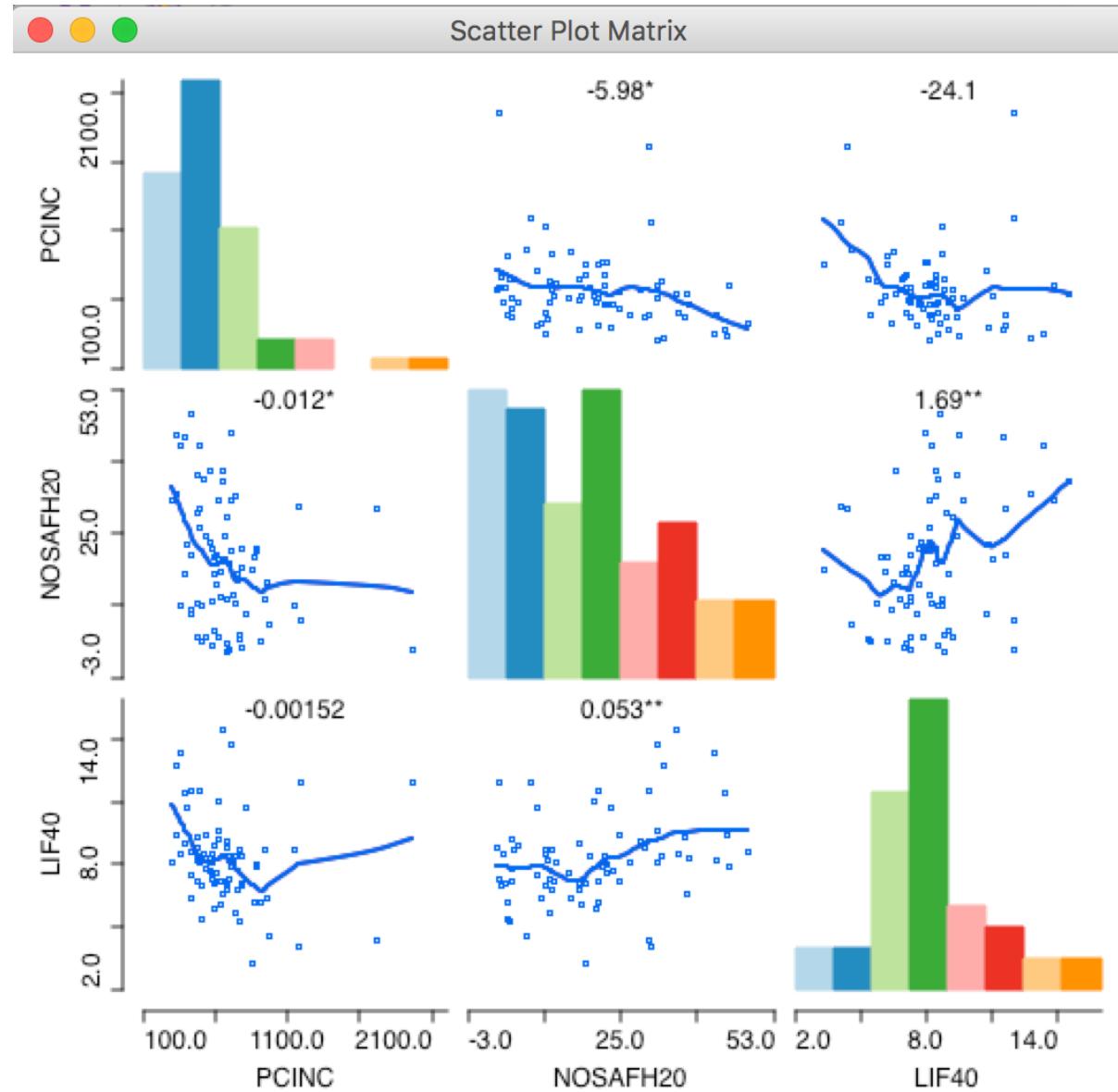
scatter plot matrix - linear fit





scatter plot matrix with selection





scatter plot matrix with lowess fit



# Parallel Coordinate Plot (PCP)



Explore Space Time R

- Histogram
- Box Plot
- Scatter Plot
- Scatter Plot Matrix
- Bubble Chart
- 3D Scatter Plot
- Parallel Coordinate Plot**
- Averages Chart
- Conditional Plot

Parallel Coordinate Plot

Exclude      Include

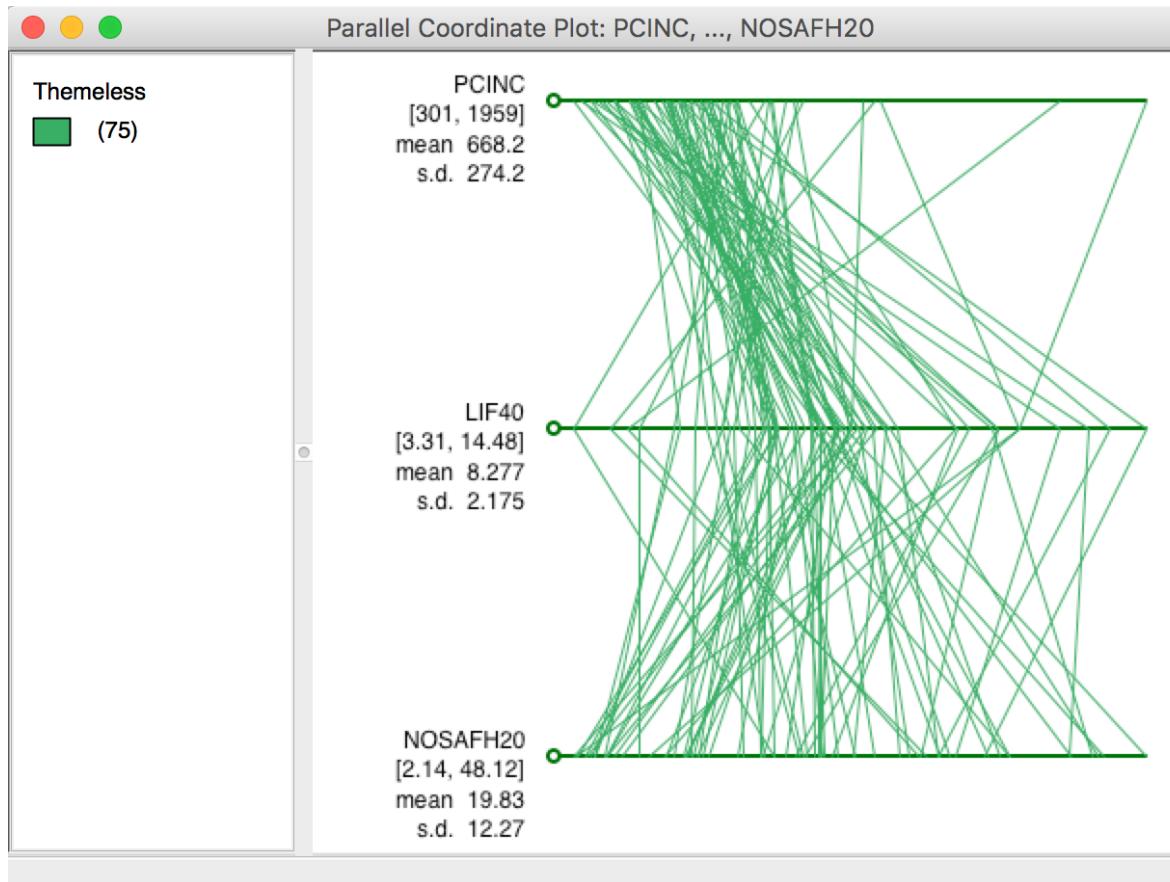
Exclude		Include
ID_0	>>	PCINC
ID_1	>	LIF40
ID_2	<	NOSAFH20
ID_3	<<	
DEPECPROV		
POVINDEX		
PCINCPPP		
PCINCMPP		
MALKIDS		
<b>POPULATION</b>		
BOYG1_5		
GIRLG1_5		
SCHOOLCNT		

OK      Close



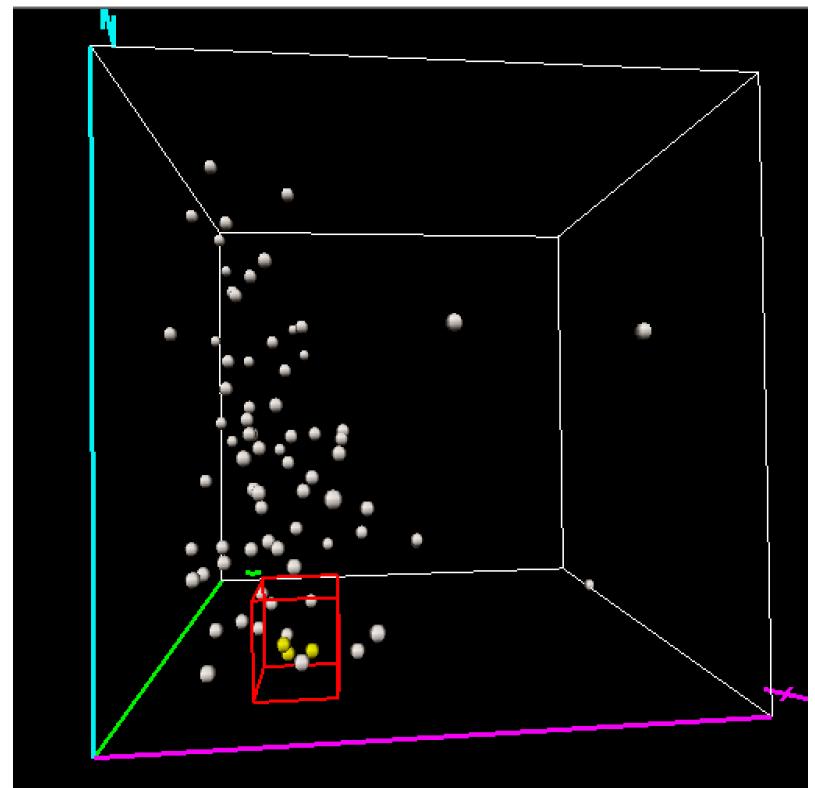
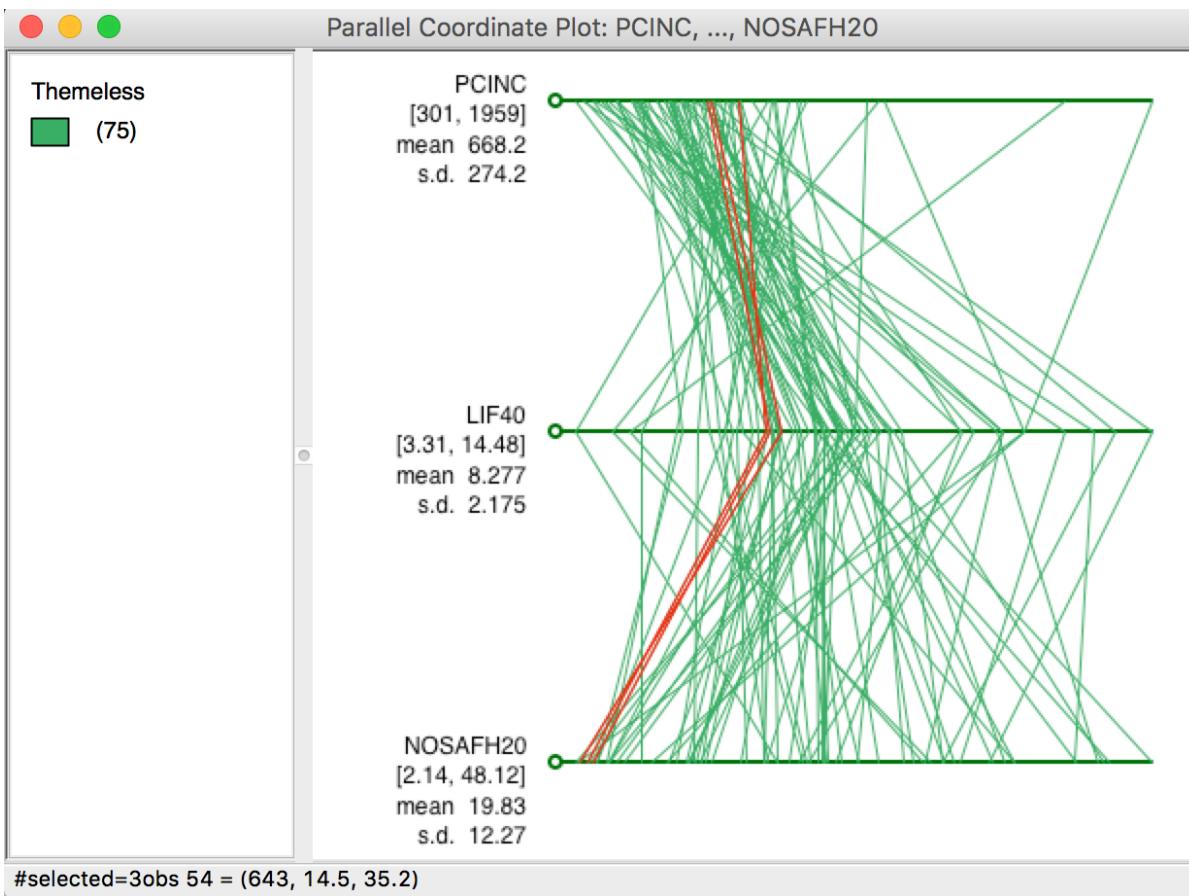
## pcp setup





## parallel coordinate plot





## selected observations in pcp



# Conditional Plots



Copyright © 2016 by Luc Anselin, All Rights Reserved



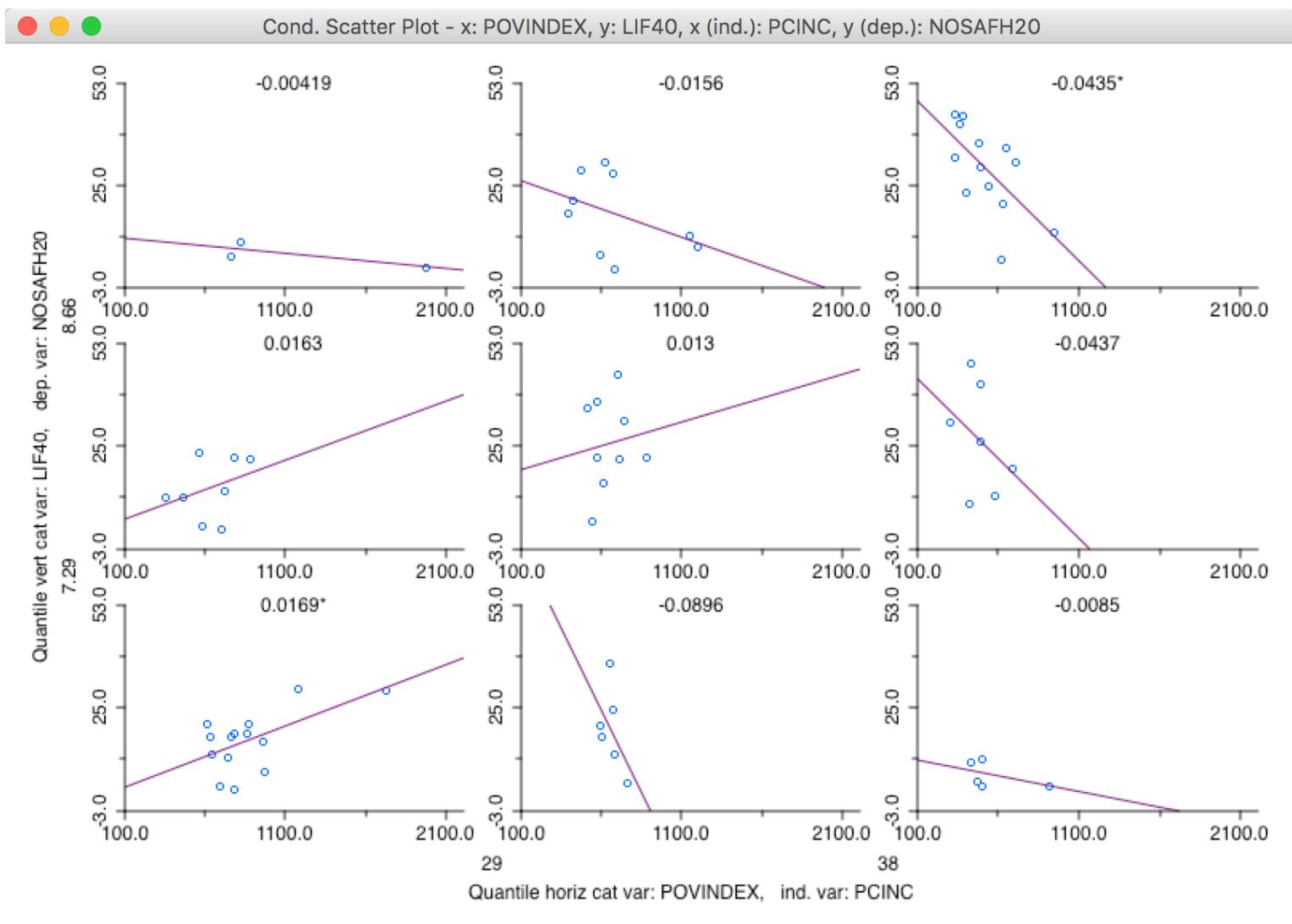
Conditional Scatter Plot Variables

Horizontal Cells	Vertical Cells	Independent Var (x-axis)	Dependent Var (y-axis)
ID_1	POVINDEX	ID_3	PCINCMPP
ID_2	PCINC	DEPECPROV	MALKIDS
ID_3	PCINCPPP	POVINDEX	LIF40
DEPECPROV	PCINCMPP	PCINC	NOSAFH20
POVINDEX	MALKIDS	PCINCMPP	POPULATION
PCINC	LIF40	PCINCMPP	BOYG1_5
PCINCPPP	NOSAFH20	MALKIDS	GIRLG1_5
PCINCMPP	POPULATION	LIF40	SCHOOLCNT
MALKIDS	BOYG1_5	NOSAFH20	AD_ILLIT
LIF40	GIRLG1_5	POPULATION	VOTNUM

OK Cancel

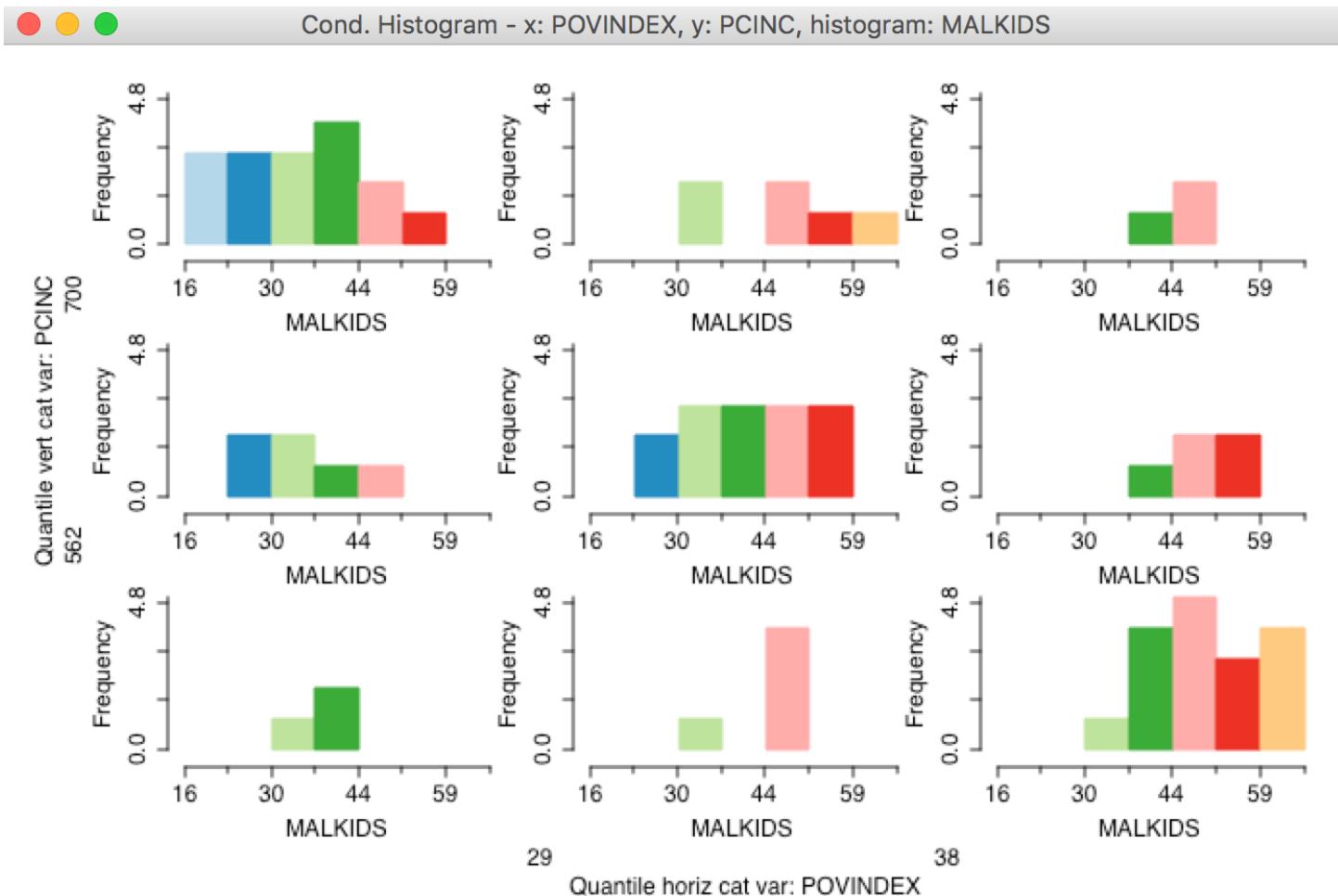
## conditional scatter plot setup





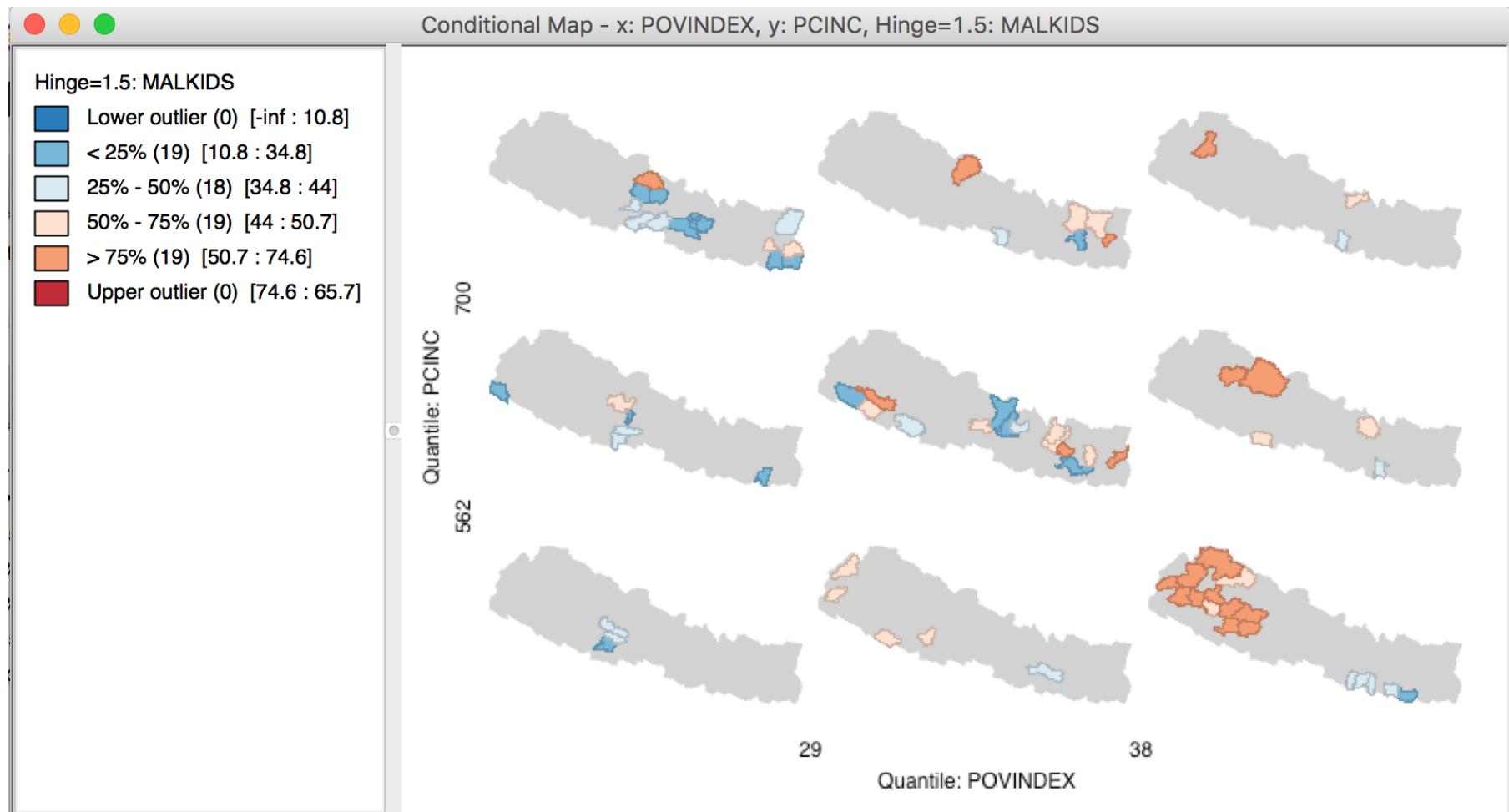
## conditional scatter plot





# conditional histogram





conditional map



