Towards building capacity on Spatial Analysis

A data-wrangling focused workshop

Manuel Lopez Bugueno

Graduate School of International Development, Nagoya University

May 8, 2024

Outline

Outline

Workshop objectives

1. To develop capacity on geographic and economic data wrangling for spatial analysis
2. To get familiar with data analysis tools: Stata, Python, GeoDa
3. To manage basic spatial econometrics' concepts:
   - Exploratory spatial data analysis (ESDA)
     - ★ Visualizing simple spatial distributions
     - ★ Weight matrixes
     - ★ Global spatial autocorrelation
     - ★ Local spatial autocorrelation
     - ★ HH, LL, HL, LH clusters
4. To perform ESDA

Today's outline

1. Applied spatial econometric concepts
2. Review of basic geographic, satellite and economic data sources
3. Basic data wrangling in Stata

Outline

MSMEs are key for growth but they face historic and structural challenges

- 98.6% of registered companies with sales
- 55.3% of employment
- 14.55% of total registered sales
- Structural challenges: necessity, innovation and added value capacities, rigidity, access to markets, finance, others

Source: Author's work using official data, 2021

Literature review

A wide range of literature has been developed over the pas two decades noting the relevance of MSMEs to growth:

- They are key agents for diffusing knowledge and innovations, boosting competition, and contributing to employment (Beck et al., 2005) (Acs et al., 2008)

Literature review

A wide range of literature has been developed over the pas two decades noting the relevance of MSMEs to growth:

- They are key agents for diffusing knowledge and innovations, boosting competition, and contributing to employment (Beck et al., 2005) (Acs et al., 2008)
- They can help bridge the gap between knowledge and total factor productivity (Solow, 2007)
- Influenced by the stage of development of a country (Porter,):
  - Factor-driven.
  - Efficiency-driven.
  - Innovation-driven.
- Their contribution is positively related to the level of human capital. As human capital level increases, more innovative high-growth potential firms can be expected (Qian Acs, 2013)
- Evidence also shows that birth rate is also relevant for development at the subnational level (Bruce et al., 2009)

Literature review

Developing countries:

- Labor-intensive, low-tech activities, less-qualified human capital $\rightarrow$ lower growth (Acs et al., 2008).
- Institutional challenges. Empirical evidence for ASEAN+1 and Latin America (Ha et al., 2008) (Acs and Amoros, 2008).
- Key paper by Cravo (2015) applies spatial methods to the municipal in Brazil (Cravo, 2015)

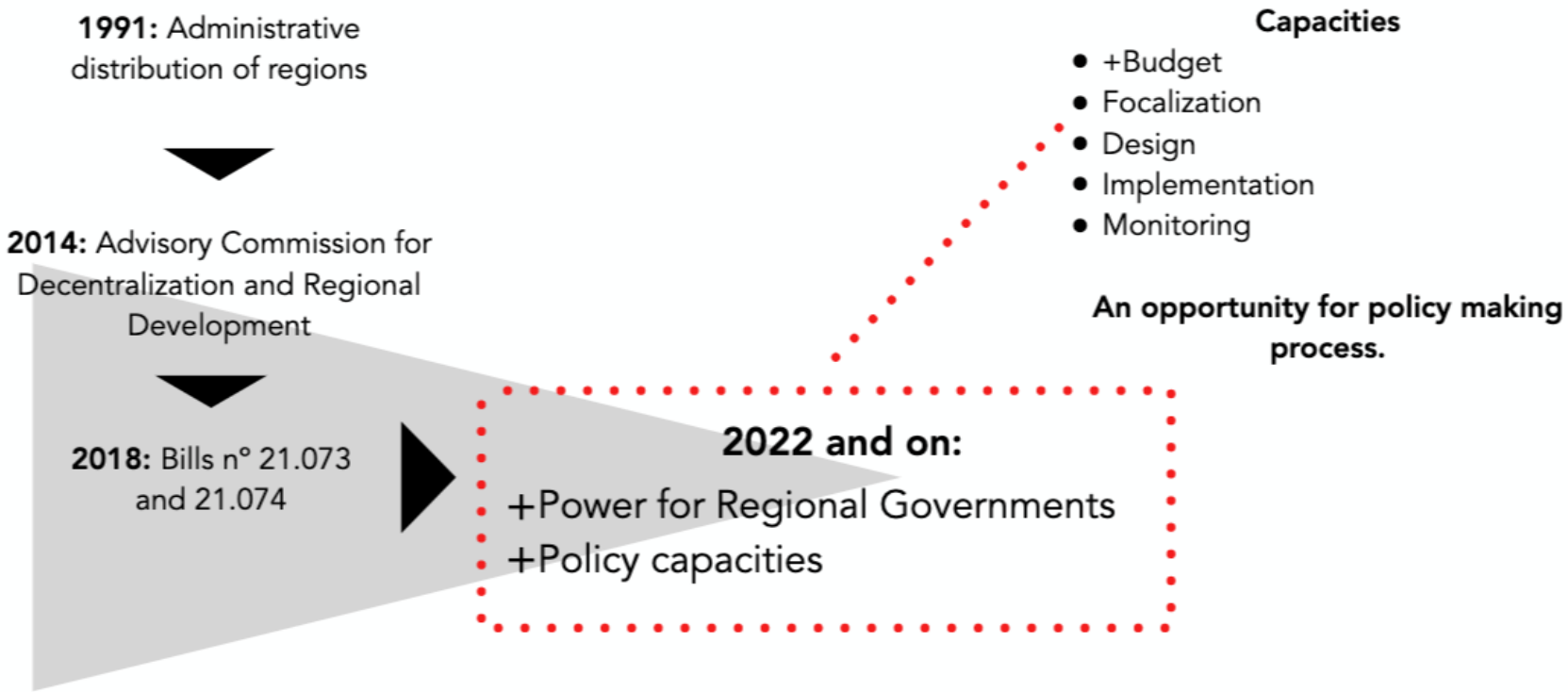
Chile: strong institutional infrastructure

- Support in key stages: starting-up a company, scaling business models, digital transformation, productivity, others.
- Nº 1 in SME Development Services in Latin America in 2019 $\rightarrow$ OECD.

Chile: strong institutional infrastructure facing new challenges

- Lack of a formal policy/strategy that defines strategic intent, short-term, long-term goals.
- Lack of a formal and binding governance.
- Hundreds of instruments, mostly designed centrally.

Decentralization process: An opportunity for regional development



Source: Author's work based on literature.

Outline

This case's research design

1. Objective
   - To provide an exploratory standpoint on spatial interactions related to MSMEs economic activity.
     - ★ Question 1 → How can Chilean policy makers target policy actions?
     - ★ Question 2 → Does space play a relevant role when evaluating MSMSE economic activity?
     - ★ Question 3 → Where does MSME activity concentrate over space?
2. Data
   - Sales - IRS, 2021
   - Population projections - NIS, 2021
   - Offical map adm3 - UREDE, 2021
3. Methods
   - Exploratory Spatial Data Analysis

This case's research outputs

- Crossectional analysis of the MSME economic activity in Chile, 2020 (**IMPORTANT:** only one variable analyzed).
- Global spatial autocorrelation analysis $\rightarrow$ Global cluster patterns $\rightarrow$ Moran's I.
- Local spatial autocorrelation analysis $\rightarrow$ Local Cluster Analysis.

Outline

Global Spatial Auto-correlation Analysis

- Up to what extent geographical units influence their neighbours → Overall clustering patterns.
- Measured by Moran'I Statistic: -1 to 1.
- 2020 data analysis.
- Weight matrix k=4.

IMPORTANT! Before continuing: What's Moran's I?

- Moran's I is the indicator that measures global spatial autocorrelation.
- Meaning that global spatial autocorrelation refers to the existence of regional relations at an overall level.
- In simple words, it measures how strong is the relationship of the regions at study in terms of a particular variable.
- From an overall standpoint, it helps answers the question "Is the role of space relevant when analyzing a specific variable?".

NOTE: More formal and detailed definitions are contained in the materials that will be provided later on.

Global Spatial Auto-correlation Analysis - Sales

Moran's I k=4, 10% significance - Sales

- 2018 $\rightarrow$ 0.105

Global Spatial Auto-correlation Analysis - Sales

Moran's I k=4 - Sales

- 2005 $\rightarrow$ 0.110
- 2020 $\rightarrow$ 0.105

What could be derived from such Moran's I value?

Local Indicator of Spatial Auto-correlation Analysis (LISA)

- Moran's I provides overall vision. LISA $\rightarrow$ what cities are clustering at significant levels.
- Weight matrix: Knn=4.
- Significance $\rightarrow$ 0.10 $\rightarrow$ many 1-city clusters.

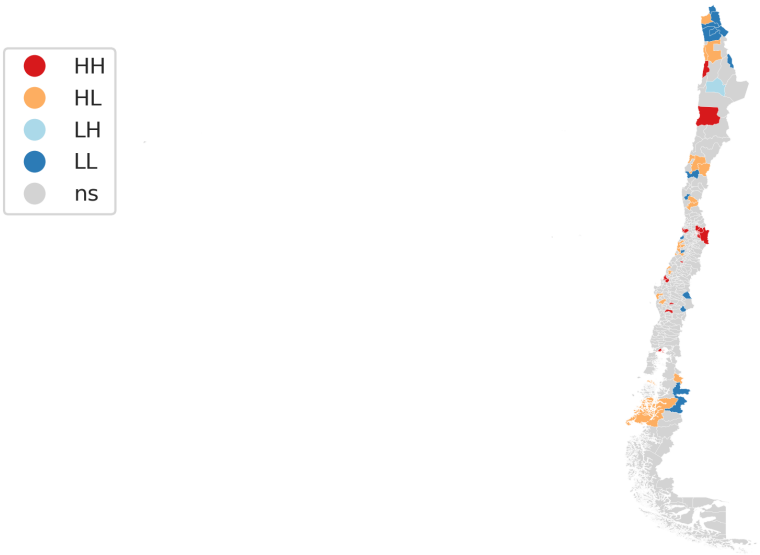
IMPORTANT! Before continuing: What's LISA?

- LISA is a statistical method used to identify spatial clusters of high or low values within a dataset.
- It helps reveal where similar values are geographically concentrated.
- It helps pinpoint specific locations where certain attributes exhibit significant spatial relationships, enabling insights into local patterns, hotspots, and potential spatial dependencies within the data..
- From an overall standpoint, it helps answers the questions "Where are spatial regimes located?" or "Where should I target a policy?".

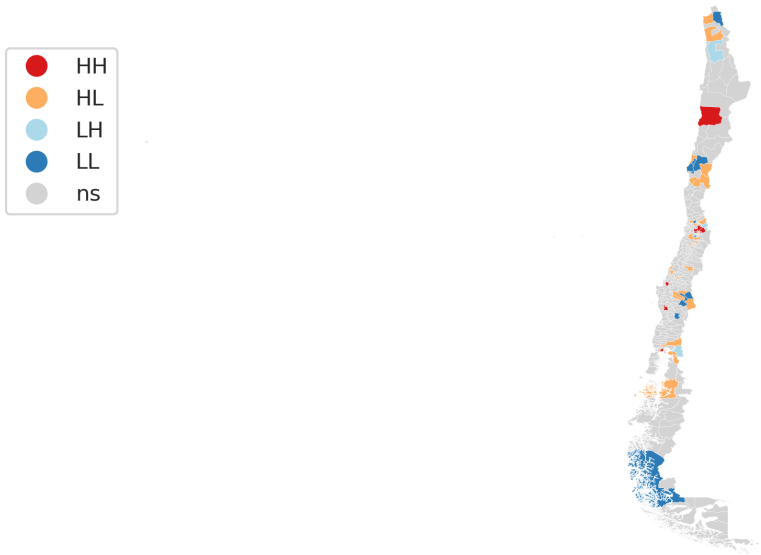
NOTE: More formal and detailed definitions are contained in the materials that will be provided later on.

Local Spatial Auto-correlation Analysis (LISA) - Sales

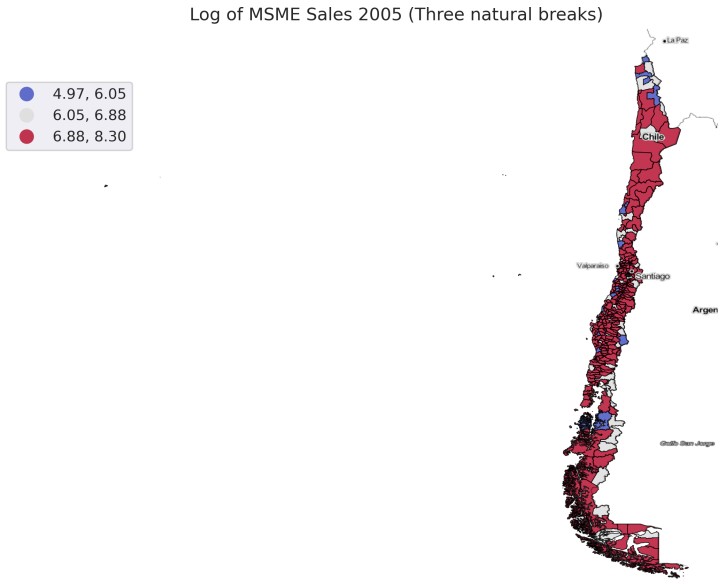
Spatial clusters and outliers knn 4 - Sales 2005 (p<0.10)

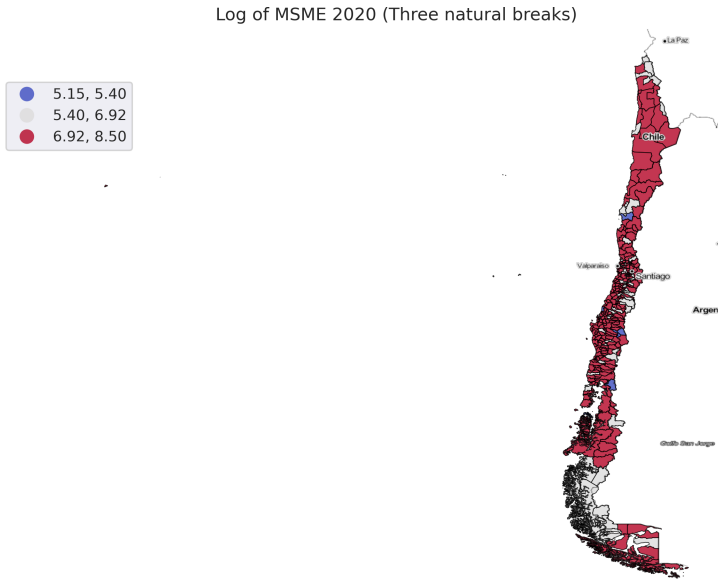


Spatial clusters and outliers knn4 - Sales 2020 (p<0.10)



Overall Spatial Patterns - Sales





Local Spatial Auto-correlation Analysis (LISA) - Sales

Number of cities by cluster and year - Sales

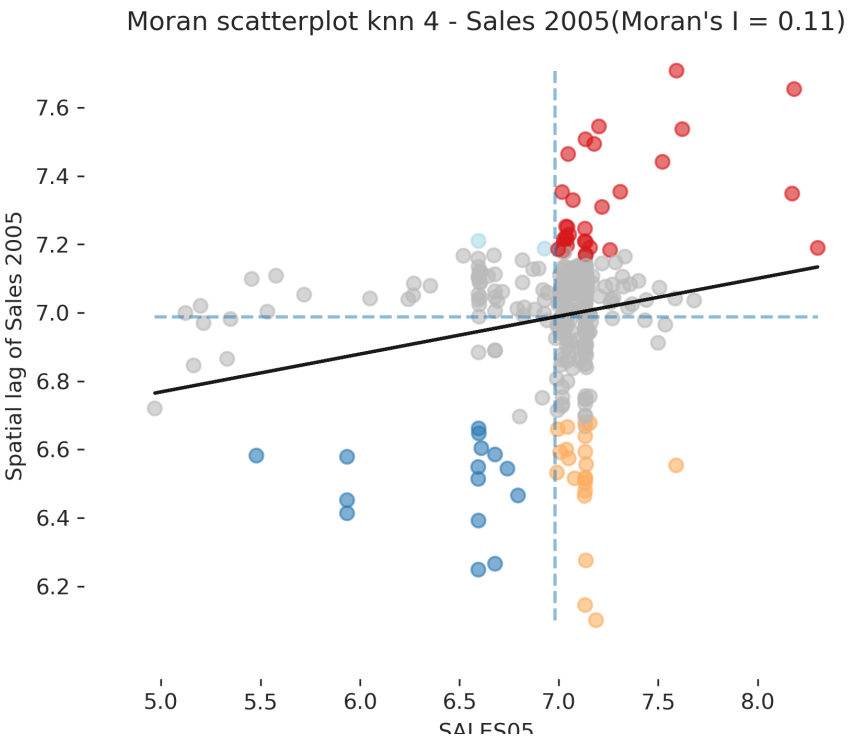
Quadrant	2005	2020
HH	32	24
HL	20	18
LL	15	10
LH	3	9
Total significant	70	61
Non-significant	275	284

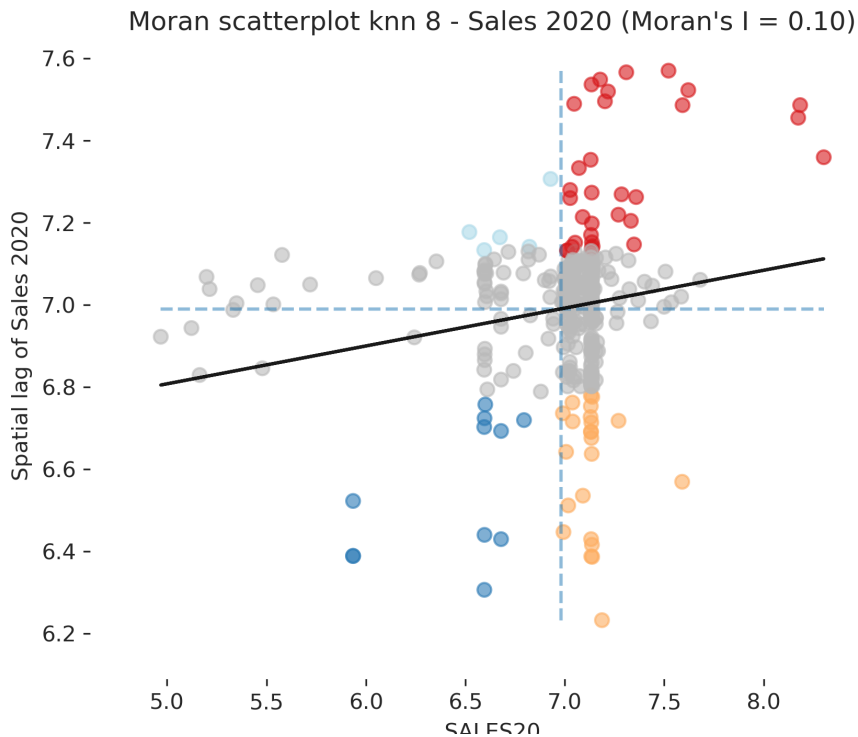
Source: Author's calculation

The most notorious patterns:

- HH cluster $\rightarrow$ slight decrease Metropolitan, central and northern macrozone.
- LL cluster $\rightarrow$ slight decrease, highly concentrated in Austral macrozone by 2020.

Local Spatial Auto-correlation Analysis





The graph in the right says knn 8 and it must say knn 4.

Outline

Basic data wrangling?

Geographic data:

- Where can I get a map for the country/region I want to analyze?
- Where can I get satellite, economic, and demographic data **(fast)**?

Basic data wrangling?

Geographic data:

- Where can I get a map for the country/region I want to analyze?
- Where can I get satellite, economic, and demographic data **(fast)**?

GeoQuery.

Basic data wrangling:

Managing data:

- How can I get my data prepared for ESDA?
- How can I check basic summary Statistics?

Let's go to Stata.

Outline

Basic data wrangling:

- Get the map, annual ntl, and population for the country of your interest. Use 2 years, the same for each variable.
- Perform basic transformations.
- Make summary statistics for your variables of interest.
- Reshape dataset

References

Acs, Z. and Amoros, J. (2008).
Entrepreneurship and competitiveness dynamics in latin america.
Small business Economics, pages 305–322.

Acs, Z., Desai, S., and Hessels, J. (2008).
Entrepreneurship, economic development and institutions.
Small business Economics, pages 219–234.

Beck, T., Demirguc-Kunt, A., and Levine, R. (2005).
Smes, growth, and poverty: Cross-country evidence.
Journal of Economic Growth, pages 199–229.

Bruce, D., Deskins, J., Hill, B., and Rork, J. (2009).
(small) business activity and state economic growth: Does size matter?
Regional studies.