

Geoprocessing and Spatial Analysis

Training Workshop provided by the IUSSP Early Career Taskforce



Organizer and instructor:

Igor Cavallini Johansen

University of Campinas, Brazil
igorcav@unicamp.br

Tuesday 10 August 2021

12:00-13:30 UTC



IC
2021
count
down
event

Does space matter?

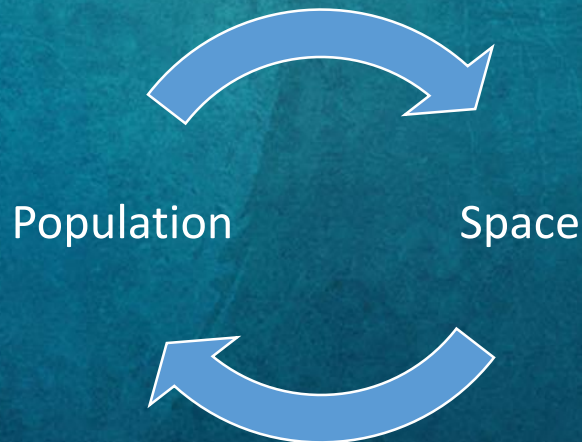
- 1- What do we mean by space?
- 2- How can space change population dynamics?
 - Morbidity and mortality
 - Fertility
 - Migration
- 3- What is geoprocessing?
- 4- What is spatial analysis?
- 5- Hands-on training



1- What do we mean by space?

- Location where we find population groups with certain characteristics
- Place of residence, study area, working place, commuting path, etc
- Why are these people in that location? Why do they have that characteristics? Bidimensional relationship:

Population can change the space where it is located



Space also can play a role on population dynamics (morbidity/mortality, fertility and migration)



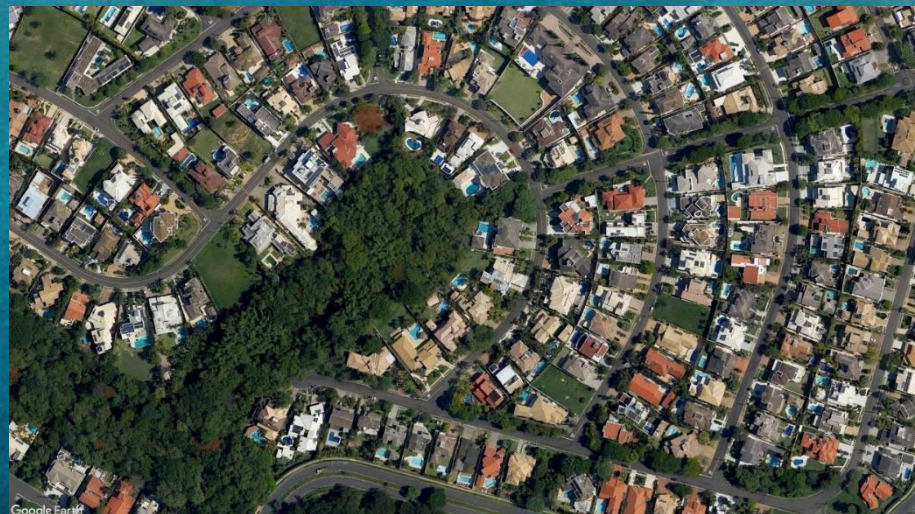
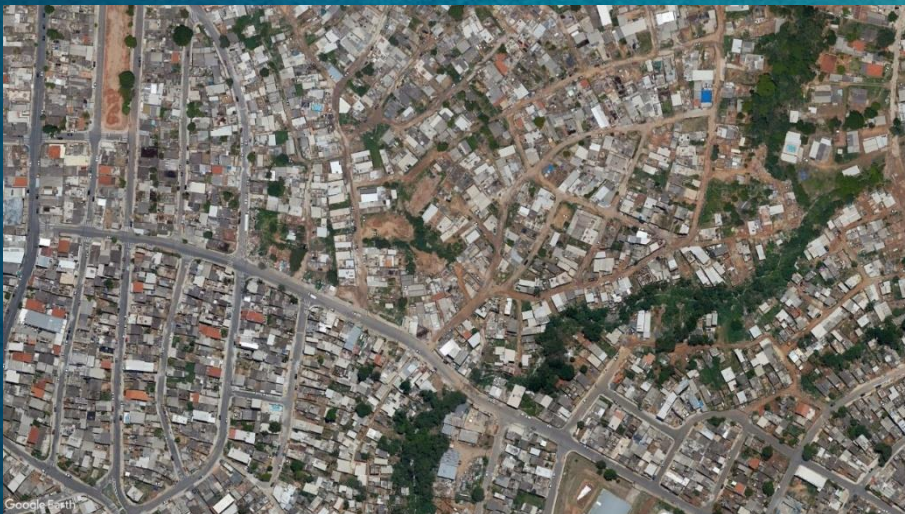
A



- So... Does space matter?

- Why?
- How?
- When?

Two neighborhoods in Campinas municipality, São Paulo state, Brazil (same scale)



2021
count
down
event

2- How can space change population dynamics?

- Morbidity and mortality
 - Depending on where you live or where you spent an amount of time of your day (work, school, etc.) your chances to get sick and/or to die due to specific diseases may differ



2021
count
down
event

Artículos/Investigación • Rev. salud pública 20 (3) May-Jun 2018 • <https://doi.org/10.15446/rsap.V20n3.54315> [COPY](#)

Environmental and demographic determinants of dengue incidence in Brazil [link](#)

Determinantes ambientales y demográficos asociados con la incidencia del dengue en Brasil

Igor Cavallini Johansen Roberto Luiz do Carmo Luciana Correia Alves Maria do Carmo Dias Bueno

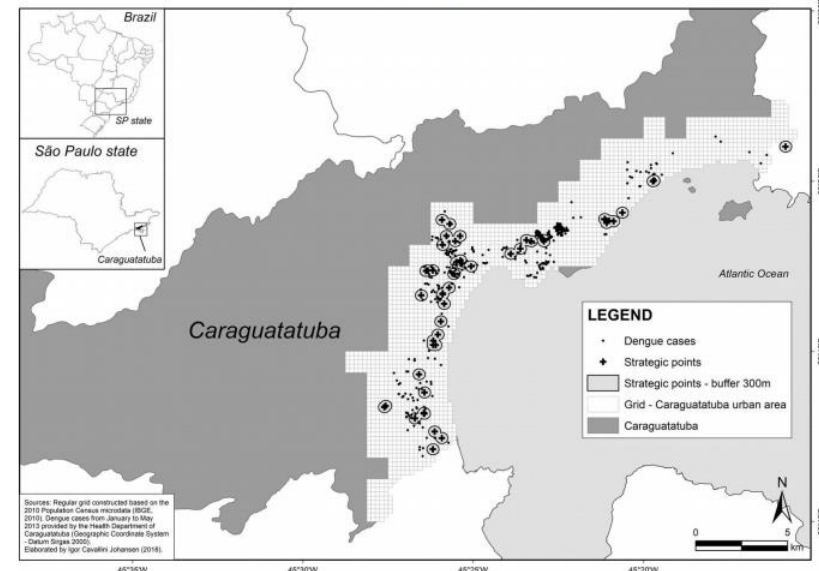
[ABOUT THE AUTHORS](#)

Table 1. Ratio of the incidence rate of the zero-inflated negative binomial regression models for dengue. Campinas, São Paulo, Brazil, January to May 2013

Variables	Ratio of the incidence rate	p-value
waterinad	0.13	0.057
sewinad	1.28	0.658
wasteinad	0.00	0.059
proximity	1.67	0.047
up3mw	71.71	0.000
nwhite	4.36	0.041
notowned	0.08	0.000

Source: Own elaboration based on software R outputs

Figure 1. Caraguatatuba municipality, in the State of São Paulo, Brazil. Regular grid of the urban area and dengue cases from January to May 2013 and 300m buffer from strategic points



Notes: a The cells correspond only to the urban area. Because dengue is primarily an urban disease, only this area was examined. b The considerable portion of the territorial municipality not classified as urban comprises mainly a vegetation-covered area of the Serra do Mar State Park.

Association between proximity to industrial chemical installations and cancer mortality in Spain

Ana Ayuso-Álvarez ^{a, b}, Javier García-Pérez ^{c, d}, José-Matías Triviño-Juárez ^e, Unai Larrinaga-Torrentegui ^f, Mario González-Sánchez ^c, Rebeca Ramis ^{c, d}, Elena Boldo ^{c, d}, Gonzalo López-Abente ^{c, d}, Iñaki Galán ^{g, h}, Pablo Fernández-Navarro ^{c, d, i}

Show more

+ Add to Mendeley



Share



Cite

<https://doi.org/10.1016/j.envpol.2019.113869>

Get rights and content

Under a Creative Commons license

open access

<https://doi.org/10.1016/j.envpol.2019.113869>



2021
count
down
event

Excess risk of cancer mortality was found in the vicinity of chemical facilities.

Industrial chemical installations and cancer mortality in Spain

Location of Chemical Industries



Data

-Industry: IPCC & E-PRTR, (2007-2010)
-Municipal cancer mortality (32 types of cancer) (Spanish Statistical Office (Instituto Nacional de Estadística – INE), 1999-2008)

Statistical analysis

-Bayesian Hierarchical Poisson Regression Models.
-Methods of Bayesian Inference: INLA & MCMC
-Distance between facilities and municipal centroids as a proxy of exposure to pollution



Relative risks of cancer mortality in towns situated at ≤5 km from different categories of chemical industries

	MEN						WOMEN					
	INLA ^a			MCMC ^b			INLA ^a			MCMC ^b		
	RR ^c	95%CrI	RR ^c	95%CrI	RR ^c	95%CrI	RR ^c	95%CrI	RR ^c	95%CrI	RR ^c	95%CrI
Chemical facilities for organic chemical transformation	Tumors											
	Colorectal	1.11	1.05-1.16	1.07	1.02-1.13	1.08	1.02-1.14	1.05	1.00-1.11			
	Gallbladder	1.20	1.03-1.40	1.16	1.00-1.35	1.13	1.00-1.27	1.11	1.00-1.25			
	Ovary						1.30	1.02-1.20	1.10	1.02-1.13		
Chemical facilities of explosives and pyrotechnic products	Ovary						1.11	1.04-1.22	1.10	1.04-1.21		
	Breast						1.06	1.00-1.13	1.06	1.00-1.11		
Chemical facilities of Fertilizers	Pleural	2.45	1.52-3.94	2.18	1.45-3.60	1.95	1.04-3.53	1.92	1.04-3.59			
Chemical facilities for inorganic chemical production	Breast						1.10	1.03-1.18	1.09	1.03-1.17		

^a: Integrated Nested Laplace Approximations

^b: Markov Chain Monte Carlo Method

^c: Relative risks adjusted for municipal population size (rural, urban and semi-urban), illiteracy rates, farmer rates, unemployment rates, average numbers of person per household and average income for each municipality. 95%CrI=95% credible interval.

• Fertility

- Different distribution of population characteristics accross space and over time (and in some cases also due to environment – ex. pollution or resources availability) can play a role on fertility and reproductive health.

THE LANCET

Global Health

Volume 9, Issue 6, June 2021, Pages e802–e812



Articles

Geographical distribution of fertility rates in 70 low-income, lower-middle-income, and upper-middle-income countries, 2010–16: a subnational analysis of cross-sectional surveys

Carla Pezzulo PhD ^a, Kristine Nilsen PhD ^a, Alessandra Carioli MSc ^a, Natalia Tejedor-Garavito PhD ^a, Sophie E Hanspal MSc ^a, Theodor Hilber BSc ^d, William H M James PhD ^{a, e}, Corrine W Ruktanonchai PhD ^{a, f}, Victor Alegana PhD ^{g, h}, Alessandro Sorichetta PhD ^a, Adelle S Wigley MSc ^a, Graeme M Hornby MSc ^{a, b}, Prof Zoe Matthews PhD ^c, Prof Andrew J Tatem PhD ^a

Show more

+ Add to Mendeley

Share

Cite

[https://doi.org/10.1016/S2214-109X\(21\)00082-6](https://doi.org/10.1016/S2214-109X(21)00082-6)

Get rights and content

open access

[https://doi.org/10.1016/S2214-109X\(21\)00082-6](https://doi.org/10.1016/S2214-109X(21)00082-6)

Methods

We did a subnational analysis of ASFRs and TFRs from the most recent publicly available and nationally representative cross-sectional Demographic and Health Surveys and Multiple Indicator Cluster Surveys collected between 2010 and 2016 for 70 low-income, lower-middle-income, and upper-middle-income countries, across 932 administrative units. We assessed the degree of global spatial autocorrelation by using Moran's *I* statistic and did a spatial cluster analysis using the Getis-Ord *G*_i^{*} local statistic to examine the geographical clustering of fertility and key selected fertility determinants. Descriptive analysis was used to investigate the distribution of ASFRs and of selected determinants in each cluster.

Findings


TFR varied from below replacement (2·1 children per women) in 36 of the 932 subnational regions (mainly located in India, Myanmar, Colombia, and Armenia), to rates of 8 and higher in 14 subnational regions, located in sub-Saharan Africa and Afghanistan. Areas with high-fertility clusters were mostly associated with areas of low prevalence of women with secondary or higher education, low use of contraception, and high unmet needs for family planning, although exceptions existed.




 2021
 count
 down
 event


• Migration

- Space is intrinsic to migration processes. Expelling and receiving areas have specific features that are relevant for peoples' choice to move out/in.



Social Networks

Volume 64, January 2021, Pages 181-193



Exploratory social-spatial network analysis of global migration structure

Hossein Akbari ✉

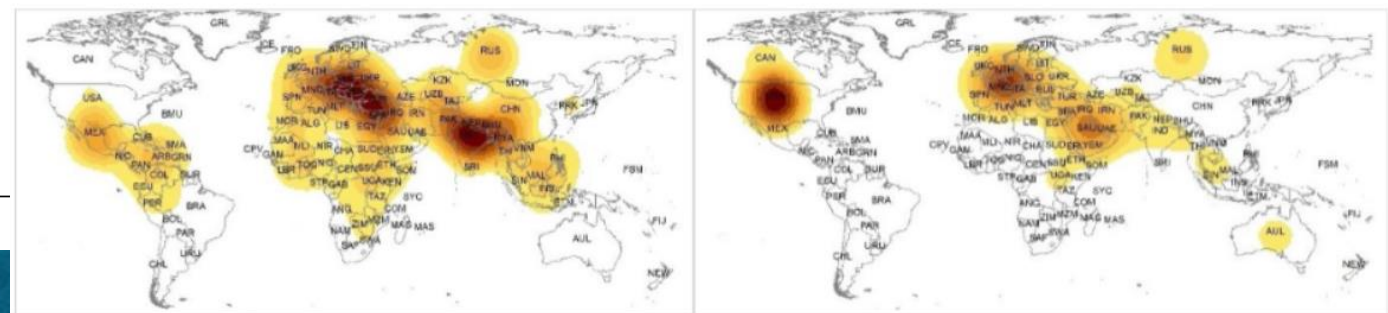
[Show more](#) ▾

[+ Add to Mendeley](#)
[Share](#)
[Cite](#)

<https://doi.org/10.1016/j.socnet.2020.09.007>



IC
2021
count
down
event



Map 1. Spatial distribution of the most important immigrants and emigrants regions based on the "in" (right map) and "out" (left map) centrality degrees using Kernel Density analysis in ARC/GIS.

3- What is geoprocessing?

- Part of a Geographic Information System (GIS)
 - “A GIS can be defined as a computer system with the capacity to **capture, store, analyze, and display geographically-referenced information**. In other words, it is an informatics for storing and managing data that has been identified according to location.” (MUSA et al., 2013)
- Utilized in a variety of fields including natural, social, engineering, and especially health sciences by incorporating physical, biological, cultural, demographic and economic data (and combining them => geoprocessing).



- In any GIS project: **preparing the data** and **performing analysis** are key. Geoprocessing operations;

- Still part of geoprocessing operations: **documentation!** (recording information about data and methodology);

- You may also use some geoprocessing tools in the last step of your project (specifically, cartography tools).



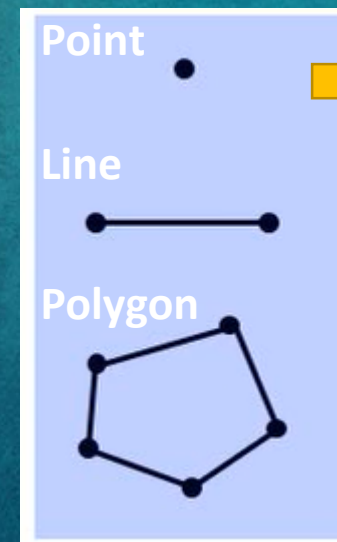
- Spatial data

- Raster images

- pixels
 - satellite images
 - aerial photos

- Vector data

- Features: points, lines and polygons
 - Attributes: size, type, length, etc.
- (Castro, 2017)



- * Events – crimes, accidents, flu cases
- * Sample from a surface – air quality monitors, house sales
- * Objects – county centroids



2021
count
down
event

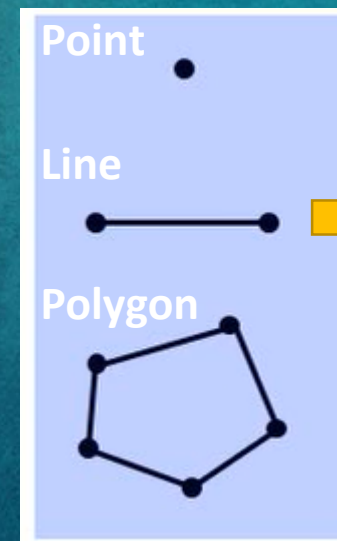
- Spatial data

- Raster images

- pixels
- satellite images
- aerial photos

- Vector data

- Features: points, lines and polygons
- attributes: size, type, length, etc.
(Castro, 2017)



- * Collection of several sequenced locations
- * e.g. Roads, rivers, fault line, drains, railroad, walls along borders, oil pipelines



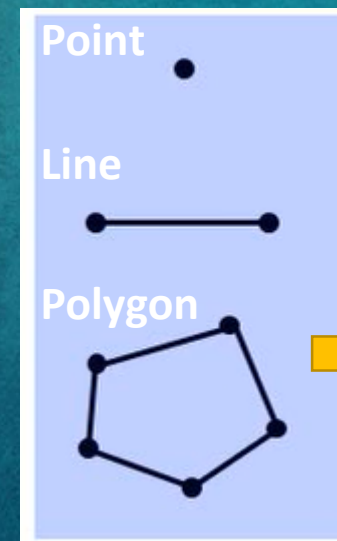
- Spatial data

- Raster images

- pixels
- satellite images
- aerial photos

- Vector data

- Features: points, lines and polygons
- attributes: size, type, length, etc.
(Castro, 2017)

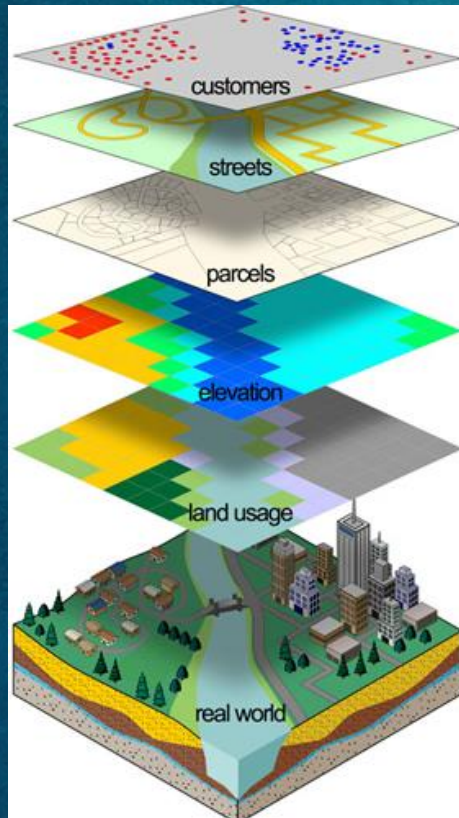


* Areal aggregations

* e.g. country, state, county, census blocks/tracks, school district, patch of vegetation, lakes, agricultural fields, refugee camp, flooding zones



- How does a GIS work? Layers' overlay:



Then you can combine data to answer your research questions.

Overlay analysis

Ex. If you have 2 layers with the same spatial unit (polygons):

Layer 1 – Population density per km²

Layer 2 – Tuberculosis incidence

Logical (boolean) operations: relate spatial data in different layers

* What areas consist of both **high population density** and **high tuberculosis** incidence? This would select only polygons which meet both conditions.

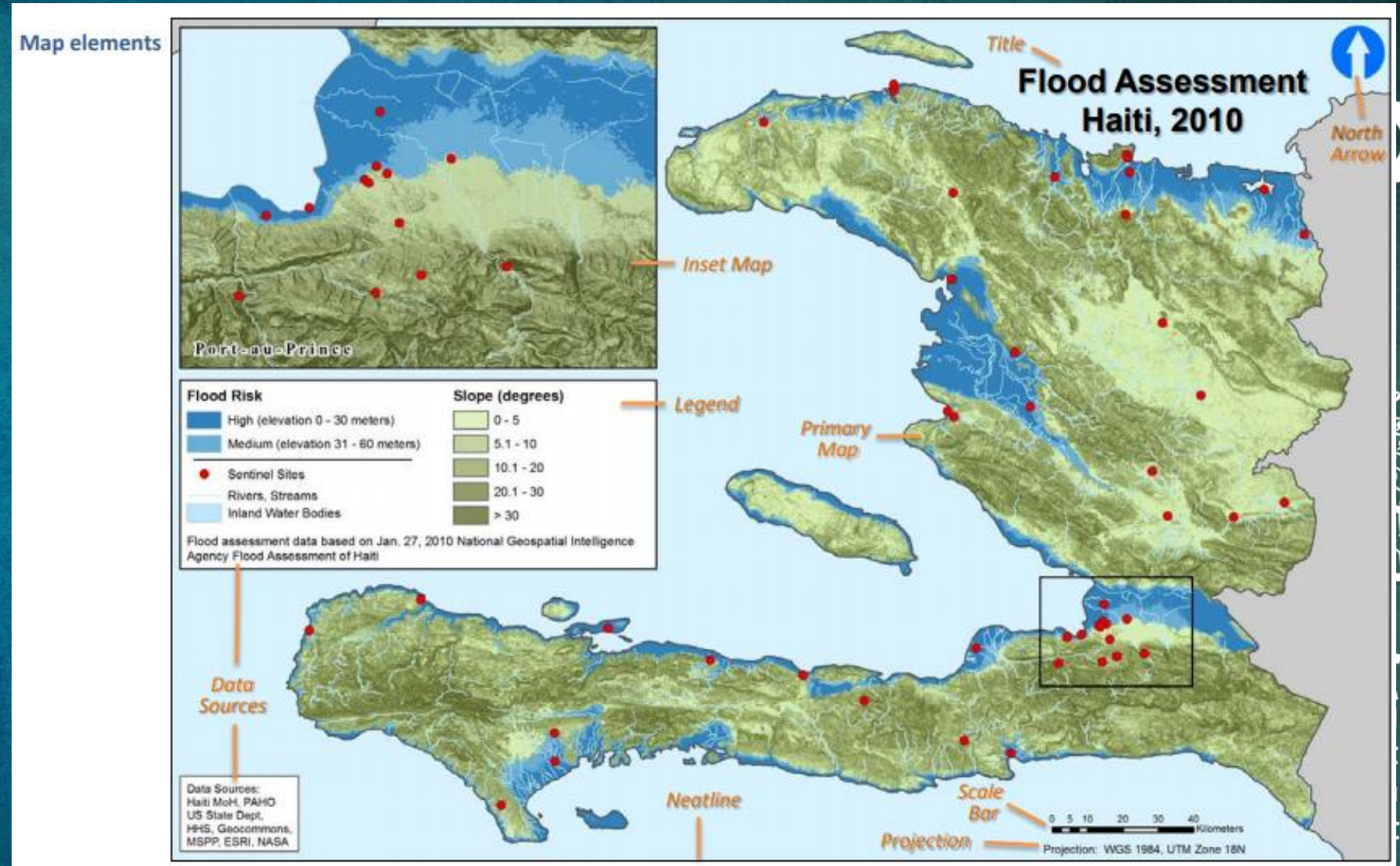


- GIS makes possible, but it is necessary to be careful combining data:
 - From different sources
 - From different years
 - In different scales
 - With different quality
 - Collected for different purposes than yours!



ic
2021
count
down
event

- Cartography

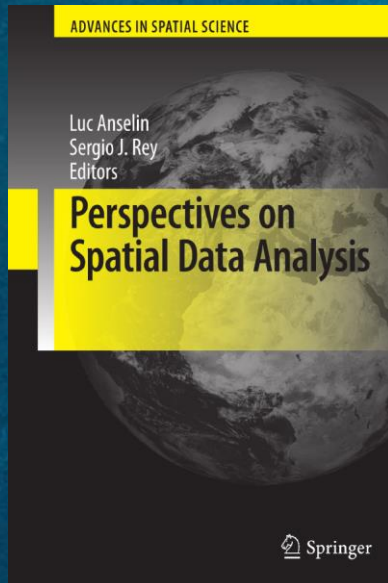


Reference: CDC, 2012

(https://www.cdc.gov/dhds/maps/gisx/resources/cartographic_guidelines.pdf)

4- What is spatial analysis?

- Tobler (1970) referred to “the first law of geography: everything is related to everything else, but near things are more related than distant things.”



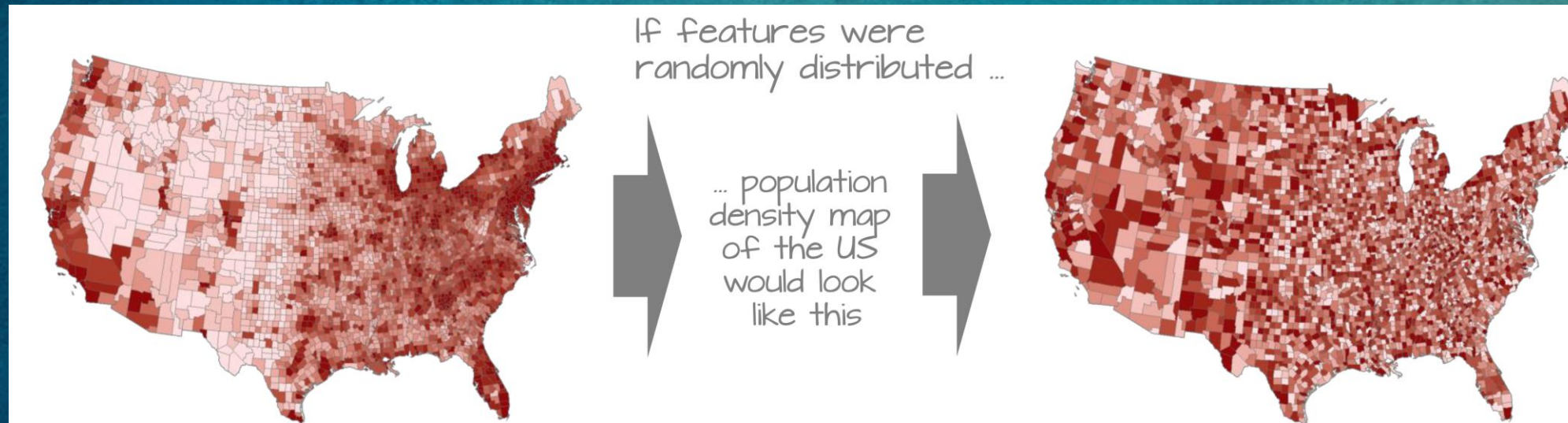
<https://www.springer.com/gp/book/9783642019753>

- Be careful:

Spatial association does not necessarily imply causality. Two things that are associated may be involved in a causal relationship, or there may be other hidden variables that cause the association. Although correlation is not causality, it provides evidence of causality that can (and should) be assessed in light of theory and/or other evidence. (MILLER, 2003)



- Spatial autocorrelation:
 - Propensity for attribute values in neighboring areas to be similar.



Gimond (2021)

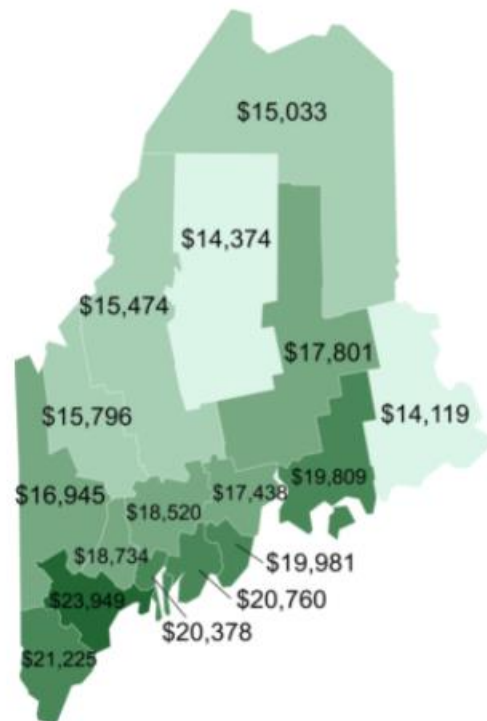


2021
count
down
event

- Discern clustered regions from non-clustered regions may not always be so obvious.
- Quantitative and objective approach to quantifying the degree to which similar features cluster and where such clustering occurs.
- One popular test of spatial autocorrelation is the Moran's I test.



- Global Moran's I: measures spatial autocorrelation based on both feature locations and feature values simultaneously



P-value **0.01**, rejecting the null hypothesis of income values as randomly distributed across each county

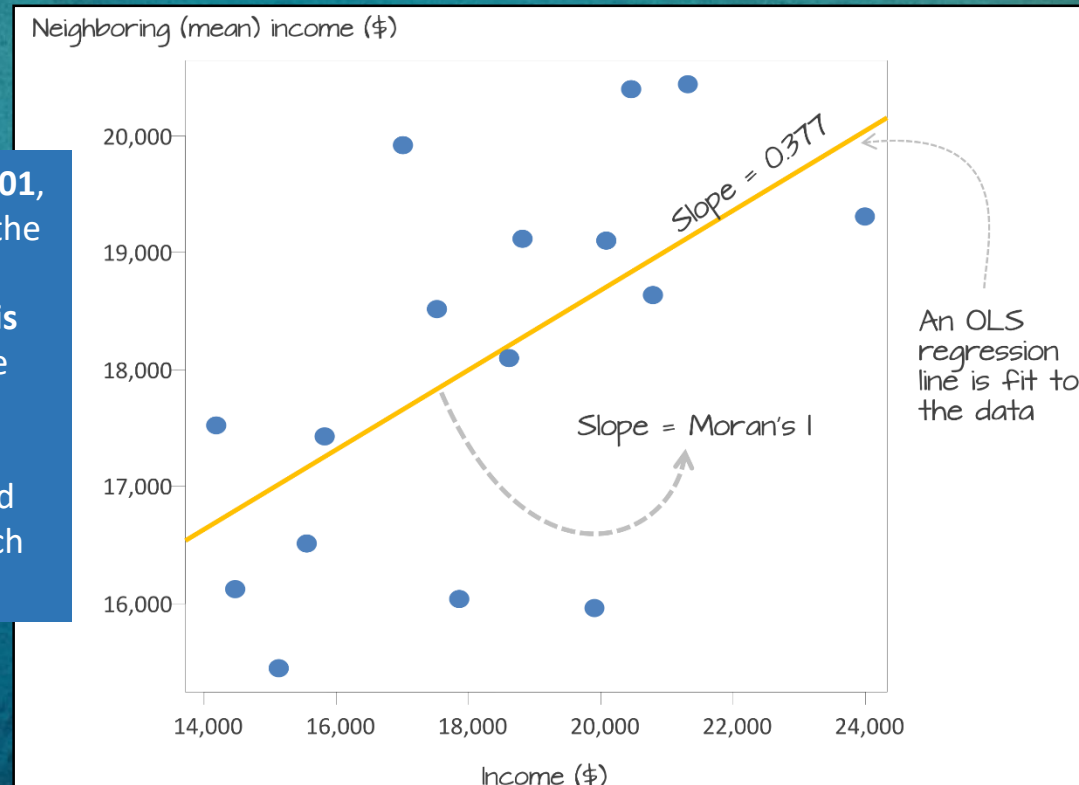


Figure 13.1: 2010 median per capita income aggregated at the county level.

- Local Moran's I: *localized* measure of autocorrelation—i.e. a map of “hot spots” and “cold spots”.

We use Monte Carlo techniques to assess whether or not the High-High values and Low-Low values are significant at a confidence level of 0.05 (i.e. for P-value= 0.05).



2021
count
down
event

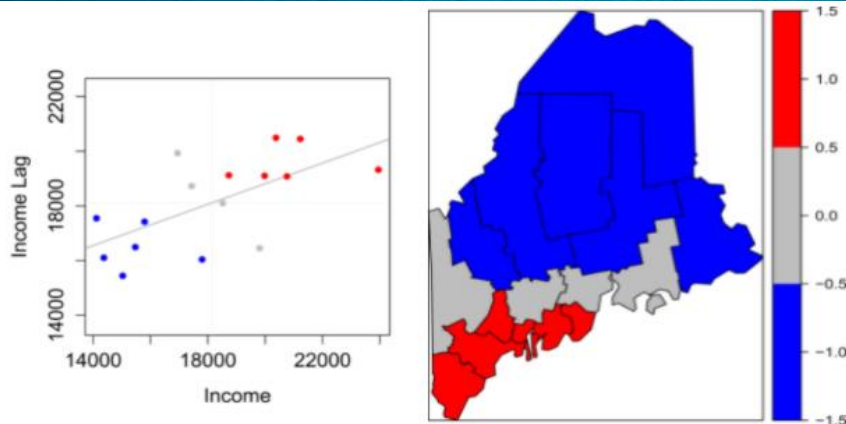


Figure 13.6: Red points and polygons highlight counties with high income values surrounded by high income counties. Blue points and polygons highlight counties with low income values surrounded by low income counties.

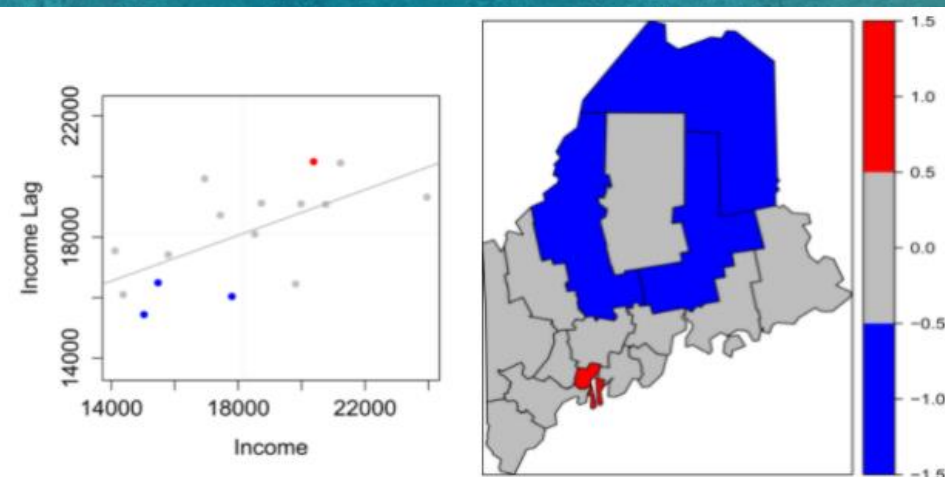


Figure 13.7: Significantly High-High and Low-Low clusters with P-values less than or equal to 0.5.

- Are there other measures of spatial autocorrelation? See Getis-Ord indexes (Getis-Ord, 1992; Ord-Getis, 1995).



- Are there other spatial analysis techniques? See Variogram and Krigging, Geographically Weighted Regression, Space-time analysis (Fischer, Getis, 2010)



- Kernel density estimation
Data smoothing – transforms points
into a raster (surface)



- Softwares for geoprocessing and spatial analysis?



- ArcGIS: <https://www.arcgis.com/index.html>



- QGIS: <https://qgis.org/en/site/>



- Geoda: <https://geodacenter.github.io/>



- TerraView: <http://www.dpi.inpe.br/terralib5/wiki/doku.php>



- R: <https://www.r-project.org/>

- [...]



5- Hands-on training


- Be sure you have downloaded the latest version of QGIS to your computer
 - Webiste: <https://qgis.org/en/site/forusers/download.html>
 - Select the option highlighted in the image below
 - Click to open the installer [5 min.]
 - Click “Run”, then “Next”
 - Select “I accept the terms in the Licence Agreement”
 - Click “Next”, then “Next” again
 - Click “Install”, then “Yes” [5 min.]
 - Click “Finish”
 - Open the folder in your Desktop
 - Double click “QGIS 3.16.9”



2021
count
down
event

Download for Windows

QGIS in OSGeo4W (recommended for regular users):

 [OSGeo4W Network Installer](#)

In the installer choose **Express Install** and select **QGIS** to install the *latest release* or **QGIS LTR** to install the *long term release*. The express installations have several optional packages including non-free software. To avoid those you have to use the **Advanced Install** and choose **qgis** and/or **qgis-ltr** in the desktop section.


NOTE FOR EXISTING USERS: OSGeo4W v2 (previously known as testing) is now the regular repository. The latest QGIS release is only available here, as it already requires dependencies not available in the old repository. The long term release is additionally also available in the old repository using the same dependencies as before (see below). This also includes a 32-bit version, which OSGeo4W v2 does not support.

CAUTION: Upgrades of old setups using the new repository are not supported. You need to do a fresh install or use a different directory.

CAUTION: Windows 7 no longer works as we are now using Python 3.9, which dropped support for it.


Standalone installers (MSI) from OSGeo4W packages (recommended for new users)

Latest release (richest on features):

 [QGIS Standalone Installer Version 3.20](#)

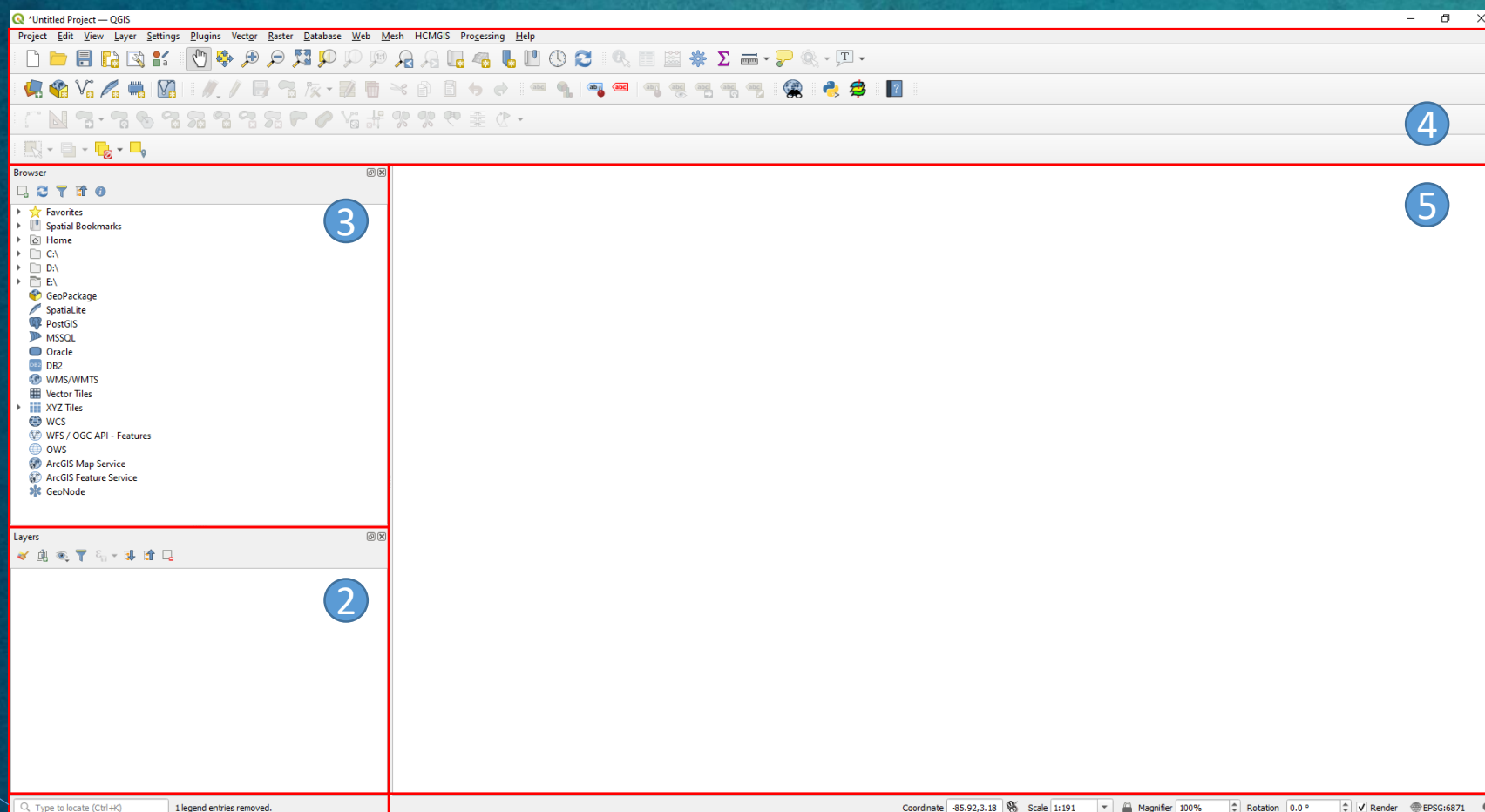
sha256

Long term release (most stable):

 [QGIS Standalone Installer Version 3.16](#)

sha256

- Click “Project” -> “New project”



- 1 Locator bar
- 2 Layers bar
- 3 Browser bar
- 4 Toolbars
- 5 Map canvas
- 6 Status bar



2021
count
down
event

- Save the folder with workshop files into your computer and unzip it. Link to download:

https://drive.google.com/file/d/1MybwL0Gy_chXi5QgHdgBOqpHVZQP15gd/view?usp=sharing



ic
2021
count
down
event

Acknowledgments

- To Tom Legrand, Nico van Nimwegen, Paul Monet, Mary Ellen Zuppan from the International Union for the Scientific Study of Population (IUSSP)
- To my colleagues of the the Early Career Taskforce at IUSSP
- To the São Paulo Research Foundation (FAPESP)
- To the audience, during and after the workshop



References

Anselin, L., & Rey, S. J. (2010). Perspectives on spatial data analysis. In Perspectives on Spatial Data Analysis (pp. 1-20). Springer, Berlin, Heidelberg.

Castro, M. C. (2017). Introduction to Spatial Methods for Public Health (Spring course). Harvard School of Public Health. Boston, USA.

Centers for Disease Control and Prevention. (2012). Cartographic guidelines for public health. Centers for Disease Control and Prevention.

Fischer, M. M., & Getis, A. (Eds.). (2010). Handbook of applied spatial analysis: software tools, methods and applications. Berlin: springer. Available in: <https://link.springer.com/book/10.1007/978-3-642-03647-7>. Accessed in 1 Aug. 2021.

Getis, A. and J.K. Ord. 1992. "The Analysis of Spatial Association by Use of Distance Statistics" in Geographical Analysis 24(3).

Gimond, M. (2021). Intro to GIS and Spatial Analysis. Available in: <https://mgimond.github.io/Spatial/index.html>. Accessed in 01 Aug. 2021.

Miller, H. J. (2004). Tobler's first law and spatial analysis. Annals of the Association of American Geographers, 94(2), 284-289.

Musa, G. J., Chiang, P. H., Sylk, T., Bavley, R., Keating, W., Lakew, B., ... & Hoven, C. W. (2013). Use of GIS mapping as a public health tool—from cholera to cancer. *Health services insights*, 6, HSI-S10471.

Ord, J.K. and A. Getis. 1995. "Local Spatial Autocorrelation Statistics: Distributional Issues and an Application" in Geographical Analysis 27(4).

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240.



Additional learning references

QGIS Training Manual: https://docs.qgis.org/3.16/en/docs/training_manual/

A Gentle Introduction to GIS: https://docs.qgis.org/3.16/en/docs/gentle_gis_introduction/

Cartographic concepts and definitions (UCLA): <https://guides.library.ucla.edu/c.php?g=180224&p=5416287>

Design principles for cartography (ESRI): <https://www.esri.com/arcgis-blog/products/product/mapping/design-principles-for-cartography/>

Cartography guide: <https://www.axismaps.com/guide>

Coursera: <https://www.coursera.org/specializations/gis-mapping-spatial-analysis>

Spatial Data, Spatial Analysis, Spatial Data Science (lecture by Luc Anselin at the University of Chicago): https://www.youtube.com/watch?v=MmCYeJ27DsA&ab_channel=GeoDaSoftware

Geospatial Analysis 6th Edition, 2020 update: <https://www.spatialanalysisonline.com/HTML/index.html>

Journal to follow:

Spatial Demography

<https://www.springer.com/journal/40980>



2021
count
down
event

Free spatial data sources

- Shapefiles by country: <https://gadm.org/maps.html>
- Different layers with a variety of scales: <https://www.openstreetmap.org/>
- See countries's statistical office for shapefiles of each country
- Gridded Population of the World: <https://sedac.ciesin.columbia.edu/data/collection/gpw-v4>
- Satellite imagery: <https://earthexplorer.usgs.gov/>



Thank you!



Igor Cavallini Johansen (UNICAMP)
igorcav@unicamp.br



ic
2021
count
down
event



IUSSP International Population Conference

5-10 December 2021

A virtual conference with 25 general
themes and a special focus on COVID-19

Registration opens on 15 September