MBA USP ESALQ

Spatial Analysis II

Rafael de Freitas Souza

Luiz Rodriguez Fantini 005.374.619-81

Spatial Data

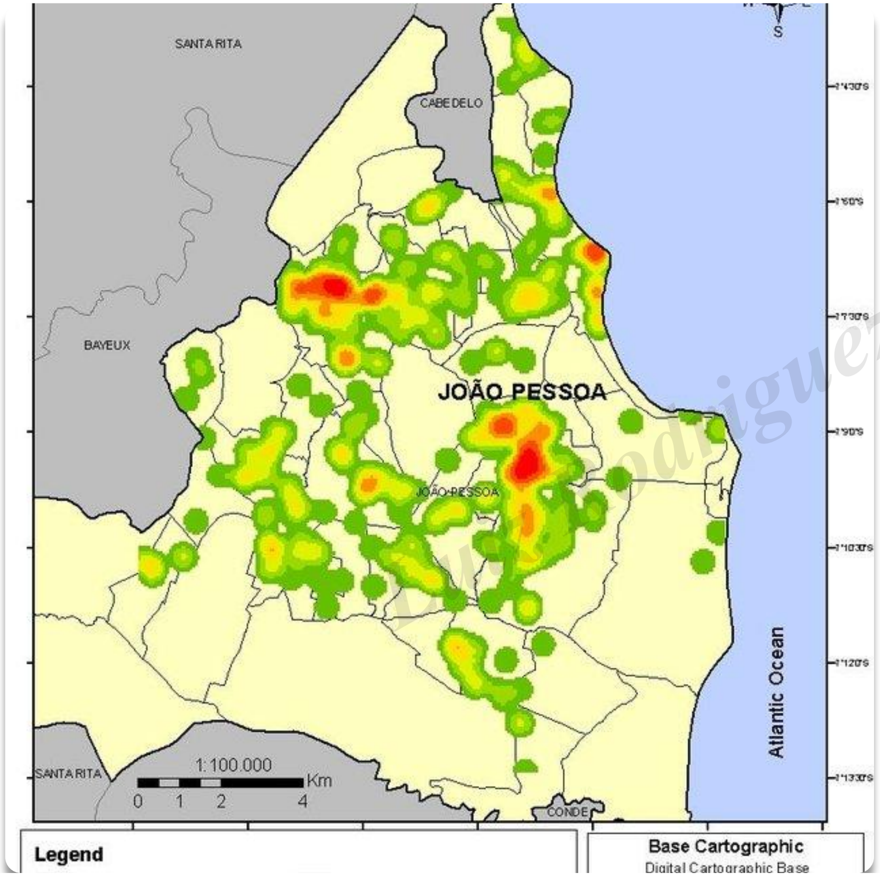
- Our subject will approach **the spatial data**, differently from the logic studied until now. However, what is the difference between data and spatial data?

*According to Fotheringham, Brunsdon and Charlton (2000) the **spatial data** indicates how much a certain phenomenon varies, concerning where this variation occurs; and **the non-spatial data** indicates how much a certain phenomenon varies, without concerning where it occurs.*

- **First Law of Geography:** "everything is related to everything else, but near things are more related than distant things." (Tobler, 1970).

Luiz Rodrigues Fantini 005.374.619-81

Example of the Functioning of Data Spatiality

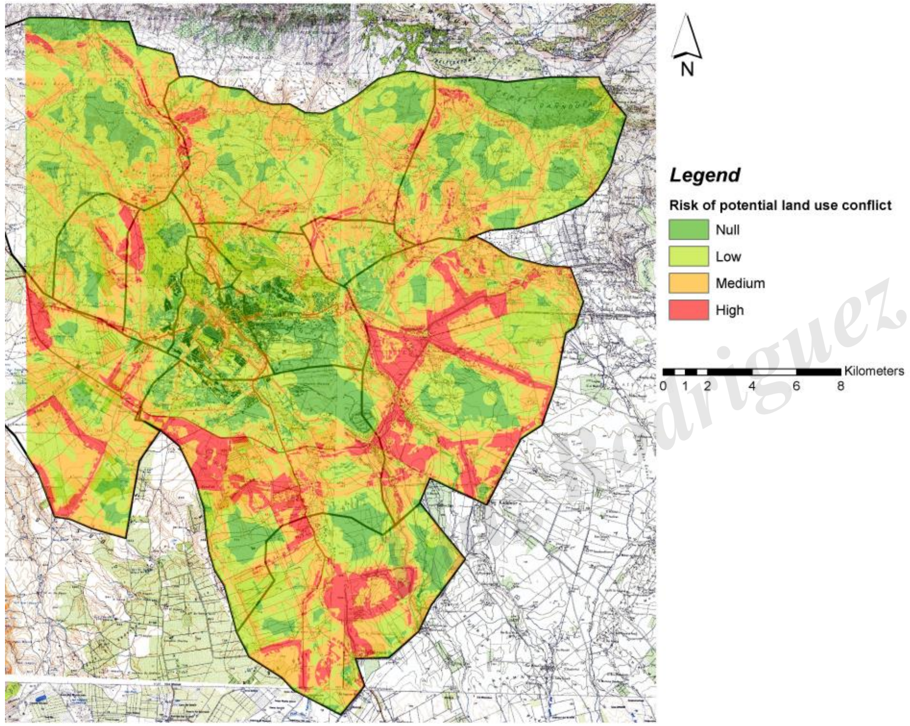


Santana and Sá (2012)

- The example on the left regards the study of crimes in João Pessoa/PB in 2012.

Fantini 005.374.619-81

Example of the Functioning of Data Spatiality

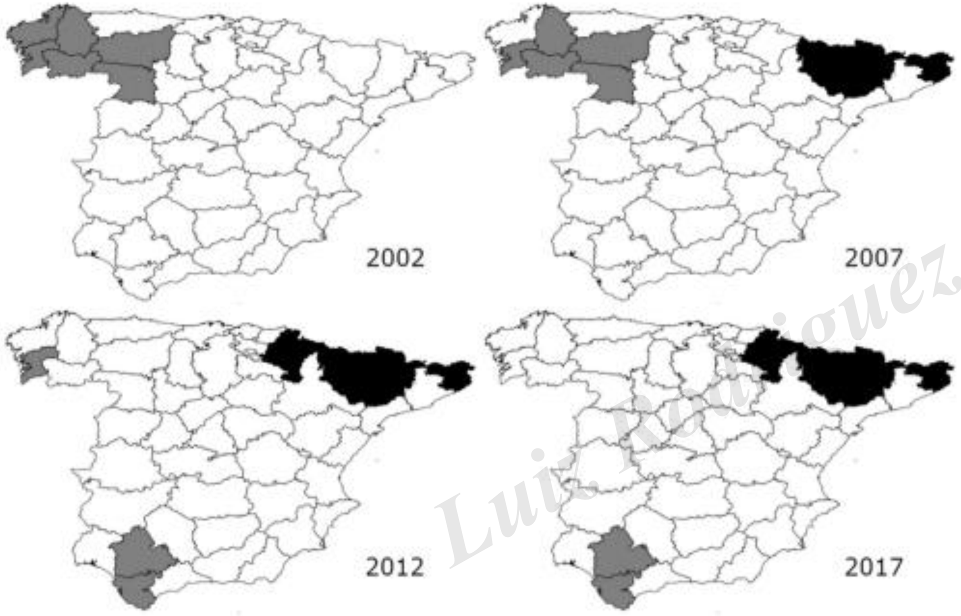


Debolini, Valette, François and Chery (2015)

- The example on the left refers to the study of the potential of conflicts generated due to the land use in Mequinez, Moroccos.

Rodriguez Fantini 005.374.619-81

Example of the Functioning of Data Spatiality



Almeida, Golpe and Álvarez (2020)

- The example on the left refers to spatial *clusters* regarding the pricing and purchase of tobacco at the frontier of Spain, from 2002 to 2017.

Luiz Rodriguez Fantini 005.374.619-81

The Structuring of the Spatial Database

id	Zip code	expenses	children	latitude	longitude
Customer 1	04537-120	4,356.90	0	-23.5928180	-46.6787473
Customer 2	04533-020	3,999.01	1	-23.5872953	-46.6878729
...					
Customer 3	05507-000	1,653.88	2	-23.5663135	-46.7125806
Customer 4	05503-120	1,522.65	2	-23.5701952	-46.7133214
...					
Customer 5	03612-160	871.32	2	-23.5167665	-46.5381829
Customer 6	03614-030	888.09	3	-23.5163167	-46.5331962

Luiz Rodriguez Fantini 005.374.619-81

MBA USP ESALQ

The Use of R in an similar way to a *Software* of the *Geographic Information Systems* (GIS) Type.

So, by what was demonstrated, is there a need to use a pre-established map?

The answer is NO. The map is a graphic resource that, depending on the purposes of your study, it can or cannot exist. In contrast, the consideration of geographical positions is really important.

- In the course, we will learn to deal with objects of *shapefiles*, *simple features*, *spatial points* and *rasters* types.

Luiz Rodrigues Fantini 005.374.619-81

Attention with Object Classes

Remember that it's necessary to use the `class()` function to observe in which class a certain object of R language belongs.




In the examples above, the classes of objects would be the forms; the objects would be the cookies. Each object class of R has their own specificities.

Attention with the Choice of a *Coordinate Reference System (CRS)*

- In a simple way, a CRS indicates coordinates on the Earth surface (e.g. longitude and latitude);

- It is necessary to pay attention in your choice since there are coordinates that consider geodesic measures, others consider euclidean measures; others consider the Earth's center of mass; others assume arbitrary points.

Luiz Rodrigues Fantini 005.374.619-81

Objects of the *Shapefile* Class

Shapefiles: presentation

- The *shapefiles* are **files** that contain information of a certain geography, and include its location and format, commonly used in GIS *softwares*, as mentioned by Lansley and Cheshire (2018).

- To summarize, the *shapefiles* correspond to sets of files that allow the existence of a map that are linked to a database. These files have extensions *.shp, *.shx, *.dbf e *.prj, at least.
  - *.shp: file that contains the geometry, that is, the polygons that will make the map;
  - *.dbf: file that contains the database;
  - *.shx: file that relates the files *.shp and *.dbf;
  - *.prj: file that describes which geographical projection system the map uses.

Luiz Rodriguez Fantini 005.374.619-81

Opening a *shapefile* in R

- Main library adopted by the course: `rgdal`

- Basic process for opening *shapefiles* in R:

```
readOGR (dsn = 2"shapefile_sp"2, layer = 3"estado_sp"3)
```

Name of the folder where the files that form the *shapefile* are.

Main nomenclature of files that form the *shapefile*

Luiz Rodriguez Fantini 005.374.619-81

Objects of the *Simple Feature* Class

Generating a *simple feature* object in R

- According to Pebesma (2018) the objects of the *simple feature* type refer, in general, to the *data frames* that have vectors with information of geographic reference.
- Main library adopted by the course: `sf`
- Therefore, the first step is to have a data frame that contains columns regarding the geographical position of the observations.
- Then, it's possible to convert the data frame into a simple feature object using the `st_as_sf()` function:

```
st_as_sf(x = seu data frame aqui,
coords = variables of longitude and latitude here,
crs = coordinate reference system here)
```

Luiz Rodriguez Fantini 005.374.619-81

Objects of the *Spatial Points* Class

Generating a *spatial points* object in R

- They are similar objects to the sf class, that can or cannot have a linked database.
- The main library adopted by the course is the **sp**;
- The sp objects are usually obtained using the `SpatialPoints()` or `SpatialPointsDataFrame()` functions:

```
SpatialPoints(coords = variables of longitude and latitude here,
proj4string = projection reference system here)

SpatialPointsDataFrame(data = your data frame here,
coords = variables of longitude and latitude here,
proj4string = projection reference system here)
```

Luiz Rodriguez Fantini 005.374.619-81

Objects of the *Raster* Class

Luiz Rodriguez Fantini 005.374.619-81

Rasters: presentation

- *Raster* objects are those that contain images with descriptions of each of the pixels that make them. This type of resource presents a georeferenced image (format *.tiff, *.jpeg, *.bmp, etc.) instead of polygons or points.
- The main library adopted by the course will be the **`raster`**;
- Basic process for opening *rasters* in R:

```
raster(raster = "raster_sp/relevo_sp.tif")
```

Luiz Rodriguez Fantini 005.374.619-81

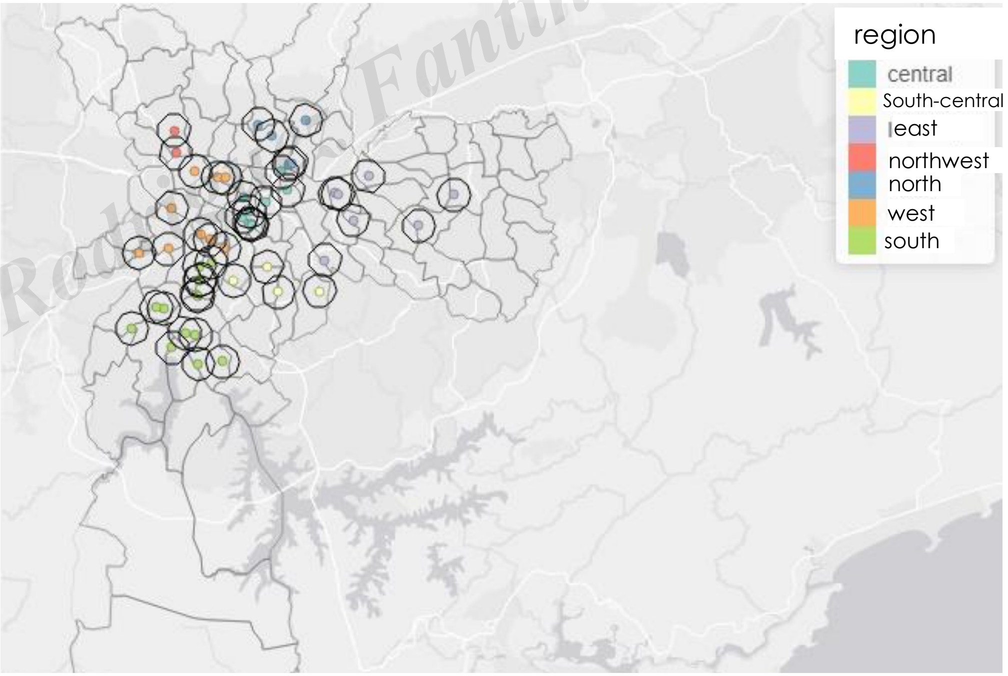
Some techniques of Spatial Analysis

Rodriguez Fantini 005.374.619-81

MBA USP ESALQ

Buffer Analysis

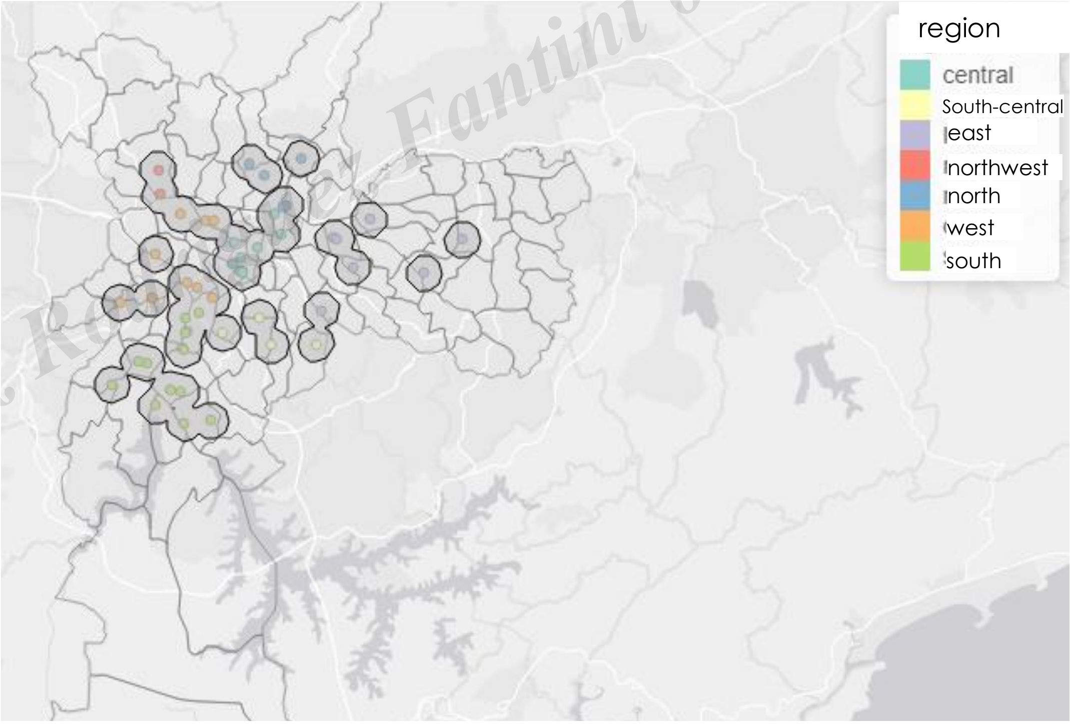
- *Buffering* is a technique to measure distances outward from the center of a certain geographical point.



Luiz Rod... Fantini 005.374.619-81

Buffer Union

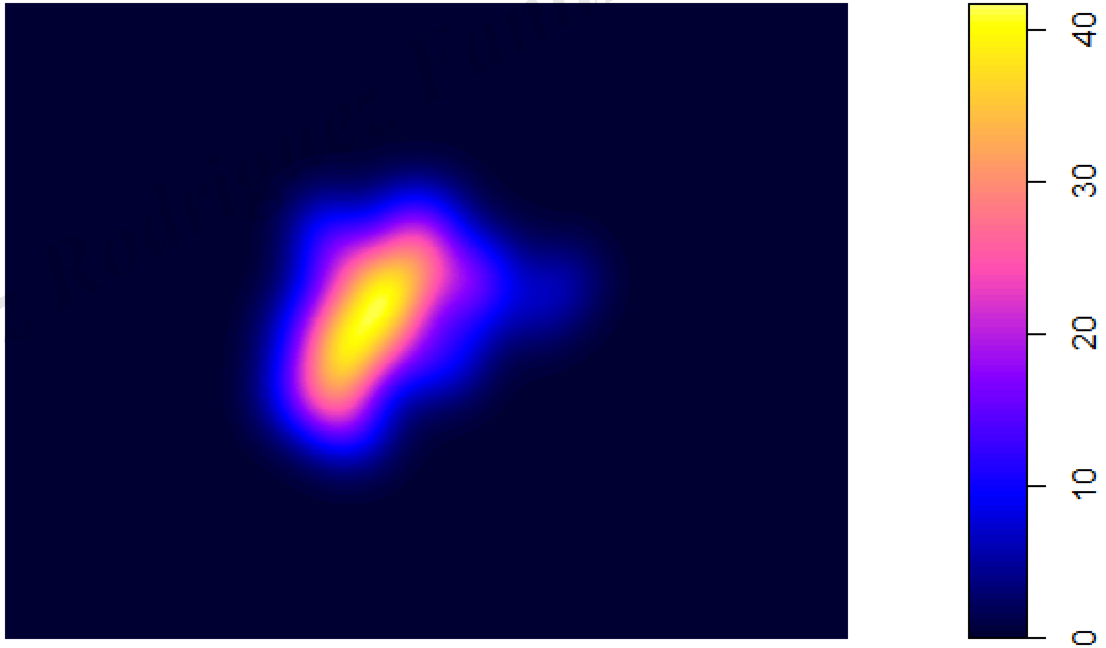
- *Buffer Union* is a technique that combines the points of *outputs* of the *buffering* technique.



Luiz Roberto ... Fantini 005 374.619-81

Kernel Densities

- The *kernel densities* technique measures the density of the presence of points of interest in a certain geographical area.



Kernel Densities

- According to Silverman (1986), the calculation of densities for a location $(x, y)$is done in the following way:

$$Density = \frac{1}{(radius)^2} \sum_{i=1}^{n} \left[\frac{3}{\pi} \cdot pop_i \left(1 - \left(\frac{dist_i}{radius} \right)^2 \right)^2 \right]$$

in which:

$i = 1,2, \ldots, n$and it refers to locations in addition to (x, y); Include only points in the sum if they are inside the distance of the place(x, y);

pop_iit indicates the population of observations in the field i;

$dist_i$it points the distance between the point $i$and the location(x, y).

The calculated density is multiplied by the number of points or by the sum of the population field, if there is any.

Luiz Rodriguez Fantini 005.374.619-81

Standardizing the concept of *datum*

Rodriguez Fantini 005.374.619-81

Datum – Reference points, Model of the Shape of the Earth, Geographic Coordinate Systems

- In the field of Spatial Analysis, we can understand the concept of *datum* as a set of information that includes a system of reference points on the earth's surface that it's associated (or it should be associated) to the model of the shape of the Earth (planar, ellipsoidal, etc.) to be able to define a system of geographic coordinates.
- The most commonly used *datum* refers to the World Geodetic System 1984 (WGS 84), also known as WGS 1984, EPSG:4326.
- In R language, all the holistic base of *datum* is, as a rule, summarized to the CRS component.

First of all: what is the WGS 84 or the WGS 84 EPSG:4326 actually?

- Therefore, it can be stated that the WGS 84 is configured in a type of standard used by Cartography that has a system of terrestrial coordinates (latitude and longitude), assuming the Earth as an ellipsoid – due to its surface and altitudes - and a *equipotential gravitational surface* (geoid) for its sea level.

- Is the WGS 84 the only standard possible?
  - No!
    - There are other standards that can be used, depending on how the format of the Earth is assumed and how its coordinates are measured, and also from which point of reference these coordinates are guide (e.g. EPSG:3857, EPSG:7789, Córrego Alegre, etc.)

driguez Fantini 005.374.619-81

Examples of Distortions that can be caused by a Carelessness in *Datum* (CRS, for R)

Area of South America: 17,840,000 km²

Antarctica area: 14,200,000 km²

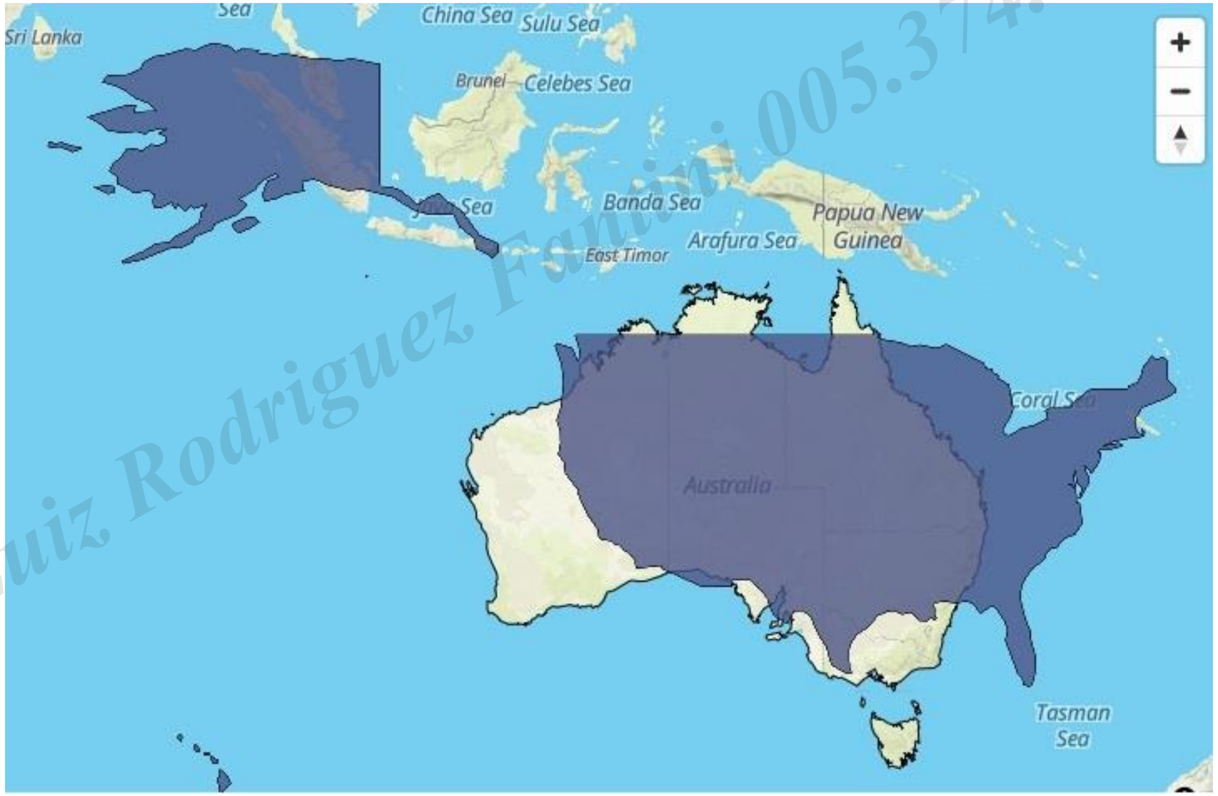
Luiz Rodrigues Faritini 005.374.619-81

Other Examples of Distortions: Comparisons between the areas of Brazil and Australia

Source: Website https://www.mylifeelsewhere.com

Luiz Rodriguez Fantini 005.374.619-81

Other Examples of Distortions: Comparisons between the areas of Australia and the USA



Source: Website https://www.mylifeelsewhere.com

Luiz Rodriguez Fanini 005.374.619-81

Other Examples of Distortions: Comparisons between the areas of the USA and Chile

Luiz Rodriguez Fantini 005.374.619-81

Source: Website https://www.mylifeelsewhere.com

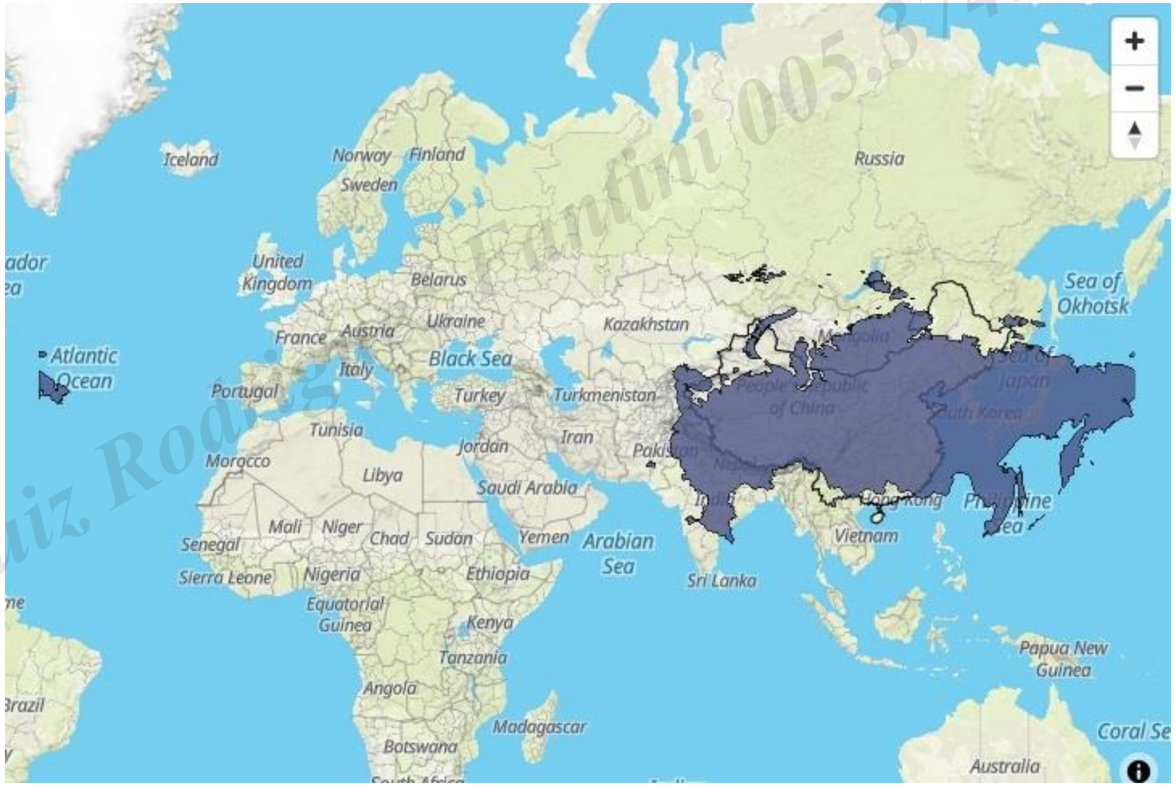
Other Examples of Distortions: Comparisons between the areas of Europe and Chile

Source: Website https://www.mylifeelsewhere.com

Luiz Rodriguez Fantini 005.374.619-81

Other Examples of Distortions: Comparisons between the areas of China and Russia

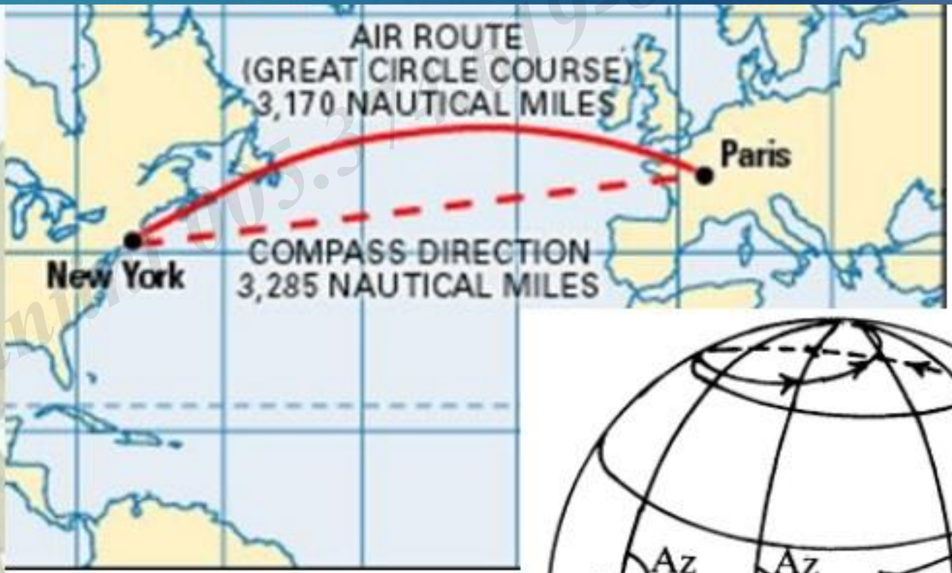
Luiz Rodrigues Fantini 005.374.619-81



Source: Website https://www.mylifeelsewhere.com

MBA USP ESALQ

Other Examples of Distortions: Comparisons of distances assuming a straight line or an arc



Source: Universidade de Lisboa

References

Almeida, A.; Golpe, A. A.; Álvarez, J. M. M. (2020).

A spatial analysis of the Spanish tobacco consumption distribution: Are there any consumption clusters? *Public Health*, 186, 28-30.

Debolini, M.; Valette, E.; François, M.; Chery, J.-P. (2015). Mapping land use competition in the rural–urban fringe and future perspectives on land policies: A case study of Meknès (Morocco). *Land Use Policy*, 47, 371-381.

Fotheringham, A. S.; Brunsdon, C.; Charlton, M. (2000). Quantitative Geography: Perspectives on spatial data analysis. Longres: Sage Publications.

Lansley, G.; Cheshire, J. (2018). Challenges to representing the population from new forms of consumer data. Geography Compass, 12(7), 1-13.

Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal, 10(1), 439-446.

Santana, A. M.; Sá, L. A. C. M. (2012). *Spatial Analysis of the Crime Distribution: A Case Study in João Pessoa-Paraíba-Brazi*l Apresentado no 8º Congresso da Fédération Internacionale des Géomètres. Montevidéu, novembro de 2012.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall.

Luiz Rodriguez Fantini 005.374.619-81