

Linked Micromap Plots for South America – General Design Considerations and Specific Adjustments

Gráficas de micromapas enlazados para Suramérica – Consideraciones
generales de diseño y ajustes específicos

JÜRGEN SYMANZIK^{1,a}, XIAOTIAN DAI^{1,b}, MARC H. WEBER^{2,c},
QUINN PAYTON^{3,d}, MICHAEL G. MCMANUS^{4,e}

¹DEPARTMENT OF MATHEMATICS AND STATISTICS, UTAH STATE UNIVERSITY, LOGAN, UTAH,
USA

²OFFICE OF RESEARCH AND DEVELOPMENT, NATIONAL HEALTH AND ENVIRONMENTAL
EFFECTS RESEARCH LABORATORY, U.S. ENVIRONMENTAL PROTECTION AGENCY, CORVALLIS,
OREGON, USA

³DEPARTMENT OF STATISTICS, OREGON STATE UNIVERSITY, CORVALLIS, OREGON, USA

⁴OFFICE OF RESEARCH AND DEVELOPMENT, NATIONAL CENTER FOR ENVIRONMENTAL
ASSESSMENT, U.S. ENVIRONMENTAL PROTECTION AGENCY, CINCINNATI, OHIO, USA

Abstract

Linked micromap (LM) plots have been in use in the United States of America (USA) since their introduction in 1996 as an effective way to display statistical summaries associated with regional spatial units. However, LM plots were always hard to create by non-experts. The introduction of the *micromap* R package has simplified the construction of LM plots for arbitrary geographic regions by facilitating the use of external Geographic Information System (GIS) features (such as shapefiles) as the basis for the maps. In this article, we will introduce LM plots for countries from South America. However, spatial representations of features are often not immediately suitable for LM plots, even after some automated simplification of the boundaries of the map regions. A common problem is that relatively small geographic regions are often not visible when plotted in LM plots. Thus, it is necessary to enlarge small regions and display them on the outside of the

^aProfessor. E-mail: symanzik@math.usu.edu

^bPhD Student. E-mail: xiaotian.dai@aggiemail.usu.edu

^cGeographer. E-mail: Weber.Marc@epa.gov

^dPhD Student. E-mail: paytonq@science.oregonstate.edu

^eEcologist. E-mail: McManus.Michael@epa.gov

main map. We introduce some algorithmic guidelines on how small regions can be addressed in LM plots for South America. Moreover, we will provide recommendations how to include areas into LM plots that are far away from the main geographic region.

Key words: Geographic Data, LM Plots, Map Visualization, R, Statistical Maps, Visualization Software.

Resumen

Las gráficas de micromapas enlazados (LM por sus siglas en inglés) han sido usados en Estados Unidos desde su introducción en 1996 como una forma efectiva de presentar resúmenes estadísticos asociados con unidades espaciales regionales. Sin embargo, las gráficas LM son difíciles de crear por no expertos. La introducción al paquete R micromap ha simplificado la construcción de gráficos LM para regiones geográficas arbitrarias al facilitar el uso de Sistemas de Información Geográficos (GIS por sus siglas en inglés) como la base para los mapas. En este artículo, se presentan gráficos LM para los países de Suramérica. Sin embargo, las representaciones espaciales están a menudo no disponibles para los gráficos LM, incluso después de simplificaciones automatizadas de los límites de las regiones. Un problema común es que regiones geográficamente pequeñas a menudo no son visibles en los gráficos LM. Entonces, se hace necesario ampliar estas regiones pequeñas y mostrarlas por fuera del mapa principal. Se introducen algunas guías algorítmicas de cómo considerar regiones pequeñas en los gráficos LM de Suramérica. Adicionalmente, se dan recomendaciones de cómo incluir áreas que se encuentran lejanas de la principal región geográfica en los gráficos LM

Palabras clave: datos geográficos, gráficos LM, mapas estadísticos, R, Software de visualización, visualización de mapas.

1. Introduction

Linked micromap (LM) plots were first introduced in 1996 (Olsen, Carr, Courbois & Pierson 1996, Carr & Pierson 1996) to overcome some of the limitations of choropleth maps. Rather than focusing on a single detailed geographic map, there are multiple small maps (micromaps) in an LM plot. Areas in these small maps are linked via color to the names of these areas and to one or more statistical panels. The statistical panels may contain any type of statistical plot, such as dot plots with or without confidence intervals or error bars, boxplots, bar charts, line charts or time series plots, scatterplots, and others. With the inclusion of informative statistical plots, information loss that frequently occurs in choropleth maps is no longer a problem in LM plots. For a detailed discussion of LM plots, the reader is referred to Symanzik & Carr (2008) and Carr & Pickle (2010). A detailed motivational example with a hypothetical LM plot can be found in Gebreab, Gillies, Munger & Symanzik (2008).

In LM plots, the boundaries of the areas are often simplified. Moreover, small geographic areas are enlarged (such Washington, D.C., in LM plots for the USA), and regions far away from the main geographic region are shifted towards the main

area (such as Alaska and Hawaii for the USA). Therefore, another limitation of choropleth maps (where small areas are usually hard to spot) has been resolved in LM plots. Choropleth maps make it difficult to compare more than just one statistical variable. As LM plots allow one to draw multiple statistical panels side-by-side and link to the map, this problem does not exist in LM plots.

Numerous government agencies in the USA have used LM plots for the display of their data. Initially, the U.S. Environmental Protection Agency (EPA) planned to use LM plots for the display of their hazardous air pollutant (HAP) data on the cumulative exposure project (CEP) web page in 1999 (Symanzik, Wong, Wang, Carr, Woodruff & Axelrad 2000, Symanzik, Axelrad, Carr, Wang, Wong & Woodruff 1999, Symanzik, Carr, Axelrad, Wang, Wong & Woodruff 1999). Unfortunately, this project was never finalized. Therefore, credit must be given to the U.S. Department of Agriculture — National Agricultural Statistics Service (USDA–NASS) for the first release of interactive, web-based LM plots for their 1997 Census of Agriculture data. LM plots at this web site still are accessible today (<http://www.nass.usda.gov/research/sumpant.htm>). Most extensively, the National Cancer Institute (NCI) has used LM plots since April 2003 (Wang, Chen, Carr, Bell & Pickle 2002, Carr, Chen, Bell, Pickle & Zhang 2002, Carr, Bell, Pickle, Zhang & Li 2003) to provide online access to their collection of cancer data (<http://www.statecancerprofiles.cancer.gov/micromaps>). More recently, researchers at the U.S. Department of Housing and Urban Development (HUD) have used LM plots for the mapping of household data from the American Community Survey (Mast 2013).

Computer code to construct LM plots (originally for S-Plus, nVizN, and Java) has been made available since their introduction in 1996, as summarized in Symanzik & Carr (2008), Section 1.5. Numerous developments of code for the production of LM plots in the statistical computing environment R (R Core Team 2014) followed, as summarized in Symanzik & Carr (2013). However, Payton, Weber, McManus & Olsen (2012) observed: “*Producing LMplots [...] has typically been somewhat difficult, and therefore LMplots have seen limited use.*” Only since the introduction of the two R packages *micromap* (Payton & Olsen 2014, Payton, McManus, Weber, Olsen & Kincaid 2014) and *micromapST* (Carr & Pearson 2014, Pickle, Pearson & Carr 2014) have LM plots become relatively easy to produce in R by non-experts. While *micromapST* is focused on LM plots for the United States, *micromap* can be used for any geographic regions, as long as the necessary geographic boundary information for the areas shown in the maps is available. Therefore, the LM plots presented in this article are based on the *micromap* R package.

Carr et al. (2002) stated that ‘a “*micromap*” can be any spatial representation from a human body caricature to a communication network.’ The focus of this article is on regional LM plots that will be introduced in Section 2 for South America derived from the Global Administrative Areas (GADM) database. A key task for regional LM plots in many cases is the adjustment of existing shapefiles. We will first discuss an ad-hoc approach and then present a new algorithm that allows a user to enlarge small areas in maps, followed by a discussion how areas far away from the main geographic area and enlarged small areas can be shifted to a new location in the map. Examples of newly created regional LM plots for South

America, in particular for Argentina, Brazil, and Uruguay, follow in Section 3. We finish with a discussion and outlook on possible future work in Section 4. R code segments can be found in the Appendix.

2. Regional LM Plots

LM plots can be used for all kinds of geographic areas, such as groups of countries, states within the USA, counties within a state of the USA, ecoregions, and many more. Under *regional LM plots*, we understand LM plots that are related to subregions within a particular country, such as states or counties within the USA. In the past, such regional LM plots were extensively used for data from the USA. Only a few other examples of regional LM plots have been created in the past, such as for France (Bonnal, Favard, Laurent & Ruiz-Gazen 2011) and for Korea (Ahn 2013, Han, Park, Mun, Choi, Symanzik, Gebreab & Ahn 2014). A possible explanation for this limited use of regional LM plots for other countries could be the initial effort to prepare boundaries of the subregions that are meaningful and clearly visible in a small map. Once created, these boundaries can be used again whenever a new LM plot for the same region has to be created. However, the initial preparation of the boundary files used to be very time-consuming.

Theoretically, any available shapefiles that outline the boundaries of the subregions could be used as a basis for the maps in LM plots. However, there are several reasons why this results in map displays that are not well suited for LM plots – unless further modifications are made to these shapefiles. First, boundaries often contain too much detailed information that is of little use in a small map and rather appears as a thick black line or black area in a small map. Next, regions that are already small and hardly visible in a big map become practically invisible in a small map. Finally, areas that are far away from the geographic main area have to be shifted closer to this area, or otherwise the central area of the map would contain a geographic region (such as an ocean or a neighboring country) that is of little relevance for the regional LM plot.

In the past, boundaries suitable for LM plots typically were created manually with the help of tools available in Geographic Information Systems (GIS). Once created, these boundaries were rarely modified again. In this article, we will demonstrate how this processing of shapefiles can be done (almost automatically) in R. In the next subsection, we will present LM plots based on ad-hoc simplified maps. In Subsection 2.2, we present an algorithm that allows a user to enlarge small areas in maps. In Subsection 2.3, we discuss how areas far away from the main geographic area and enlarged small areas can be shifted. Figure 1 shows the adjustments of boundaries for Brazil in three steps.

2.1. Ad-Hoc Versions

For the basic process of simplifying regions for use in LM plots for geographic features in R, several methods are available. These include R functions such as `thinnedSpatialPoly` in the *maptools* R package (Bivand & Lewin-Koh 2014), `dp` in the *shapefiles* R package (Stabler 2013), `generalize.polys` in the *GISTools*

R package (Brunsdon & Chen 2014), and the `gSimplify` function in the *rgeos* R package (Bivand & Rundel 2014). These simplification functions all use the Douglas–Peucker algorithm (Douglas & Peucker 1973) for point simplification and require a weeding tolerance for simplification of polygons. The point simplification algorithm is fast and simplifies lines by keeping critical points to depict shapes and removing other points. It is, however, a fairly rudimentary simplification approach with limitations, and handling of polygon boundaries when simplifying in R is fairly poor (for example, slivers and gaps are introduced as the level of simplification increases and there are no readily available means in R to control this, and alternatively when using topology preserving techniques in R, feature shapes can quickly become distorted beyond recognition). An alternative approach to simplifying regions in R is to use the topology-preserving simplification approach in the online tool *MapShaper* (Harrower & Bloch 2006), accessible at <http://www.mapshaper.org/> and <https://github.com/mbloch/mapshaper>, an approach that is suggested in both the *micromap* and *surveillance* (Höhle, Meyer & Paul 2013) R packages. We show an example of ad-hoc simplification for Brazil in the upper right map in Figure 1.

2.2. Enlargement of Small Areas

In most instances, shapefiles brought into R to create LM plots adhere to realistic boundaries of subregions, and many subregions of an area being depicted may well be too small to be discernible in an LM plot (for instance Washington D.C. in the United States as mentioned earlier). We use a simple function within R to apply a minimum size threshold for all subregions in a region which is provided in the Appendix. Any subregions falling below this threshold are then enlarged accordingly through a rescale function which is provided in the Appendix and makes use of the `gbuffer` function in the *rgeos* R package.

Applying this rescaling to any given subregion in a *SpatialPolygonsDataFrame* involves dissolving the subregion into the most appropriate neighboring subregion, and then replacing that subregion with its rescaled replacement which has further been “clipped” to any neighboring boundaries where necessary. We make use of functions in the *rgeos* R package to do the dissolving, clipping, and enlarging of subregions in a region. Further, we have found it necessary to simplify each particular subregion that needs rescaling prior to simplifying an entire region. This is due to line artifacts that may be created when simplifying, which will show up in the map after rescaled features are added back to the map. The steps we have found to work for rescaling are to:

- Dissolve feature to be rescaled into neighboring feature using the *rgeos* R package function `unionSpatialPolygons`.
- Align slots of data and polygons after dissolving since the output of dissolving is just a polygon object in R without the associated data (spatial objects in R are composed of “slots” – so that a spatial object has separate slots for attribute data, coordinates, coordinate reference system, etc.).

- Rescale the feature of interest.
- Clip the rescaled feature with dissolved background subregion using the *rgeos* R package `gIntersection` function and then cut out the area of the rescaled feature from the dissolved background region using the *rgeos* R package `gDifference` function.
- Align slots of data and polygons again.

The result of our rescaling process, applied to the smallest regions in Brazil, is shown in the lower left map in Figure 1. Overall, seven regions in Brazil have been enlarged: Alagoas, Distrito Federal (the federal district), Espirito Santo, Paraiba, Rio de Janeiro, Rio Grande do Norte, and Sergipe. Four of these enlarged regions are located in the northeast, two are located in the central east, and the federal district is located almost in the center of Brazil.



FIGURE 1: Boundaries for the regions of Brazil: (upper left) Original boundaries, based on the shapefile from GADM; (upper right) boundaries after some simplification using the Douglas–Peucker algorithm; (lower left) boundaries after simplification and rescaling; and finally (lower right) boundaries after simplification and rescaling with example shifting of the federal district. Note that maps in the bottom row appear larger as some of the small islands to the east of Brazil have been removed in the thinning process for these maps.

2.3. Shifting of Far Away and Enlarged Small Areas

In some instances, these enlarged small areas will require shifting in order to maintain the integrity of the map. Furthermore, in order to construct LM plots, it may sometimes be necessary to bring some subregions within closer proximity to the main area being depicted for proper display in an LM plot (for example Alaska and Hawaii in the United States as mentioned previously). If areas simply need to be shifted, we apply a shift function which takes an offset value for x and y coordinates and applies it to each of the “slots” in a spatial data frame in R .

An additional complication is determining in a somewhat automated way where to place the shifted subregion. We do this by comparing the bounding box of the region and the convex hull of the main area being depicted, finding the largest empty area to place a shifted subregion. Lastly, shifted subregions may need to be rescaled as well, either enlarging or shrinking to be visible in proportion with the main map area.

Our shifting process applied to the federal district of Brazil is shown in the lower right map in Figure 1. However, this map has been created for demonstration purposes only as we believe that just the enlargement of this region is sufficient and no further shifting is needed. This is not always the case.

3. Examples for South America

For this article, we define South America according to *The World Factbook* of the CIA, accessible at https://www.cia.gov/library/publications/the-world-factbook/wfbExt/region_soa.html. However, we left out the Falkland Islands (Islas Malvinas), French Guiana, and the South Georgia and South Sandwich Islands. A different source may use a different definition of the countries that make up South America.

3.1. Outline of Boundaries for South American Countries

Figure 2 shows the boundaries of twelve countries of South America, after the simplification of boundaries, as discussed in Subsection 2.1. These modified boundaries eventually can be used for LM plots by interested readers. For some countries, such as Bolivia, Paraguay, and Suriname, this basic step will result in meaningful boundaries for LM plots. For a country such as Argentina, it seems to be necessary to further enlarge the capital area as discussed in Subsection 2.2. However, for a country such as Uruguay, this is optional, depending on the size of the actual figure that will show an LM plot for Uruguay, i.e., on a full-page figure, the capital region would be well discernible while in a smaller figure, this wouldn't be the case. In Subsection 3.2, we do not enlarge the capital region of either country and just work with the ad-hoc versions of the simplified boundaries. For a country such as Ecuador, it may be advantageous to shift one of the regions (the Galapagos province) closer to the main area as discussed in Subsection 2.3. For

some of the remaining countries, such as Chile and Colombia, multiple adjustment steps may be required to produce boundaries that are well suited for LM plots.

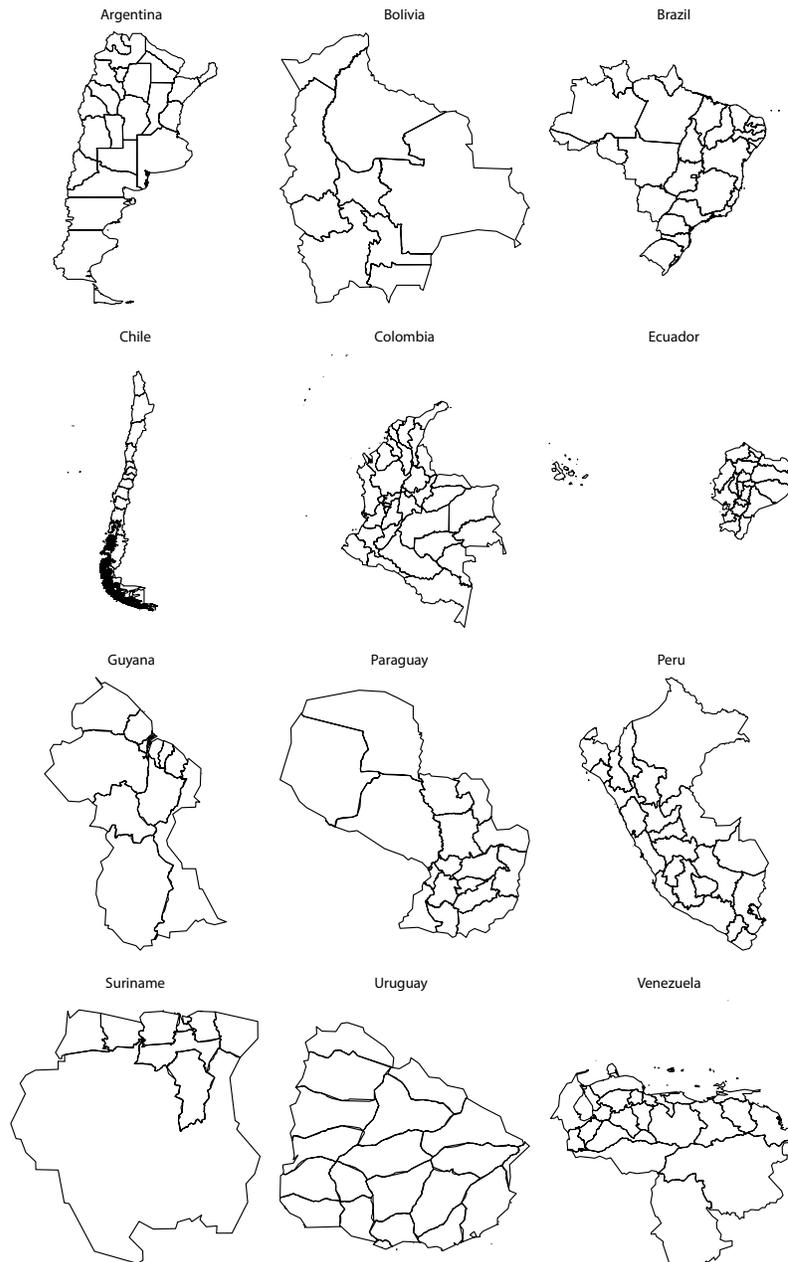


FIGURE 2: Boundaries for twelve countries from South America, after some ad-hoc boundary simplifications.

In LM plots for the United States, the 50 states are typically partitioned into 10 perceptual groups, showing 5 states in each map. The countries of South America have between 9 and 33 administrative divisions and therefore require different partitionings of these regions into the maps. According to the recommendations from Symanzik & Carr (2008), Table 1.2, our Table 1 shows possible partitionings into small perceptual groups for the regions of these twelve countries.

TABLE 1: Full symmetry partitionings with targeting groups of size 5. Column “Partitioning 1” puts the smallest counts in the middle. Full-symmetry alternatives that avoid small counts appear in column “Partitioning 2”.

| Country | # Regions | Partitioning 1 | Partitioning 2 |
|-----------|-----------|----------------|----------------|
| Argentina | 24 | 5 5 4 5 5 | |
| Bolivia | 9 | 4 1 4 | 3 3 3 |
| Brazil | 27 | 5 5 3 1 3 5 5 | 4 4 4 3 4 4 4 |
| Chile | 15 | 5 5 5 | |
| Colombia | 33 | 5 5 5 3 5 5 5 | 4 5 5 5 5 5 4 |
| Ecuador | 24 | 5 5 4 5 5 | |
| Guyana | 10 | 5 5 | |
| Paraguay | 18 | 5 4 4 5 | 4 5 5 4 |
| Peru | 26 | 5 5 3 3 5 5 | 5 4 4 4 4 5 |
| Suriname | 10 | 5 5 | |
| Uruguay | 19 | 5 4 1 4 5 | 4 4 3 4 4 |
| Venezuela | 25 | 5 5 5 5 5 | |

3.2. Ad-Hoc LM Plot Examples for Argentina and Uruguay

Argentina is the eighth largest country in the world, the second largest country in Latin America, and the largest among the Spanish-speaking nations. Argentina is subdivided into 23 provinces and one autonomous city. Hence, there are 24 administrative regions overall in Argentina that will form the basis for the LM plots for this country.

Figure 3 shows the population of each province in 2010 and the population change from 2001 to 2010, based on the report from the National Institute of Statistics and Census of Argentina in the first two statistical panels (from left to right). The data set has been obtained from the Wikipedia web page at http://en.wikipedia.org/wiki/List_of_Argentine_provinces_by_population. The rightmost statistical panel shows the ratio of the population in the year 2010 to that of the year 2001 for each province. The 24 provinces of Argentina are sorted in the LM plot in the order of their population in the year 2010, starting with the province of Buenos Aires that had the largest population, and ending with the province of Tierra del Fuego that had the smallest population. The central and northern provinces are the most populated regions in Argentina. There is a huge gap in population between the province with the largest population, Buenos Aires, and the province with the second largest population, Cordoba. As mentioned in Subsection 3.1 the federal region (Ciudad de Buenos Aires) is practically invisible in the maps based on the ad-hoc boundaries and needs some enlargement, as discussed in Subsection 2.2.

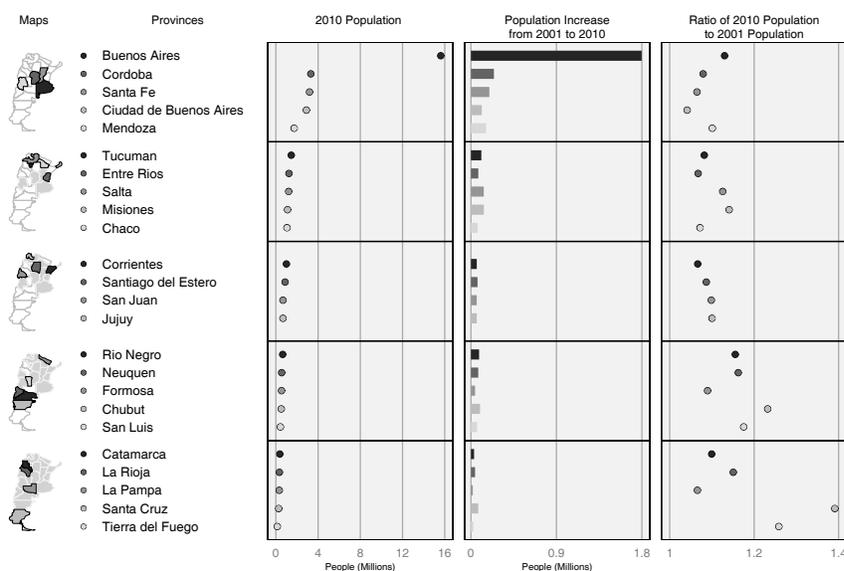


FIGURE 3: LM plot for Argentina, showing the 2010 population and the ratio of the 2010 to 2001 population as dot plots in the left and right statistical panels and a bar chart for the total population increase in the central statistical panel. The provinces are sorted according to the 2010 population in the first statistical panel. In the maps, we use some incremental shading so that each new map also indicates by the light gray color which provinces appeared in the previous perceptual groups, i.e., in maps above the current map.

All provinces had a population increase from 2001 to 2010. For the province of Buenos Aires, this was an increase of more than 1.5 million people, while for all other provinces, the increase was less than 500,000 people each. The population in the year 2010 and the population change from 2001 to 2010 may not be positively correlated because the provinces with the largest population in 2010 also had the largest population increase. In fact, the patterns in the first and third statistical panel of this figure suggest that the population in 2010 and the ratio of 2010 population to 2001 population are negatively correlated. Provinces with small populations in 2010 had encountered the largest (relative) population increases from 2001 to 2010.

Uruguay is a country located in the southeastern part of South America. It is bordered by Argentina to the west, Brazil to the north and east, and the Atlantic Ocean to the south and southeast. Uruguay is subdivided into 19 departments. Figure 4 shows the area and the population in 2011 of each department. The data set has been obtained from the Wikipedia web page at <http://en.wikipedia.org/wiki/Uruguay>. The departments in the LM plot are split into four groups and one median department, based on their area rankings. For the first three perceptual groups (i.e., about 14 departments), the population in each department is about the same, independently from the size of the department. However,

interestingly, the last group of departments in the south facing the Rio de la Plata show a strong negative correlation between their area and population. Montevideo, the capital department, has by far the smallest area, but had a population of about 1.3 million which was more than twice the population of Canelones, that had the second largest population. As mentioned in Subsection 3.1, it may be beneficial to enlarge the capital department in the maps, depending on the size of the figure that shows the LM plot. As a side note, the shapefiles used for this LM plot associated the area covered by the Lagoon Mirim (Laguna Merín) to Rivera and not to any of its three bordering departments, Cerro Largo, Rocha, and Treinta y Tres.

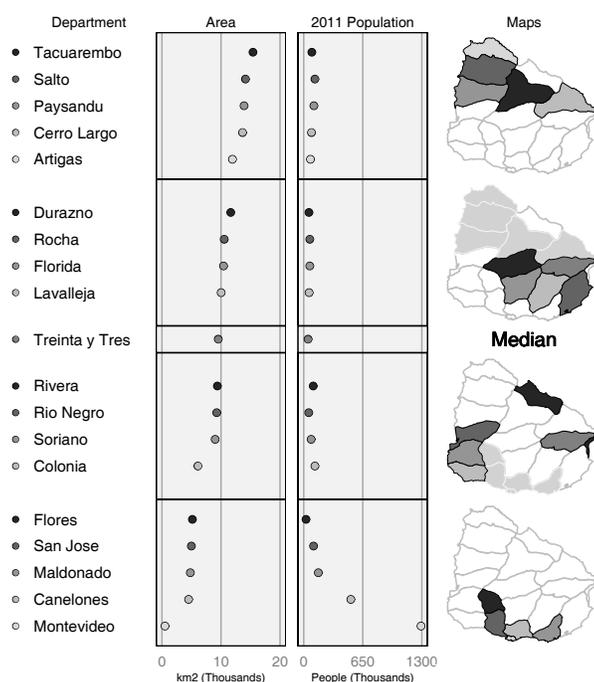


FIGURE 4: LM plot for Uruguay, showing the area and the 2011 population as dot plots in two statistical panels. The departments are sorted according to the area. In the maps, we use some two-sided shading to emphasize departments that are larger than the median size in the two perceptual groups at the top and departments that are smaller than the median size in the two perceptual groups at the bottom.

3.3. LM Plot Examples for Brazil after the Enlargement of Small Areas

Brazil, officially the Federative Republic of Brazil, is the largest country in both South America and the Latin American region. It is the world’s fifth largest

country, both by geographical area and by population. Overall, Brazil is one of the most important countries in South America that was ranked seventh in the world in 2012 with respect to the nominal gross domestic product (GDP), according to data from *The World Bank* (<http://databank.worldbank.org/data/download/GDP.pdf>). Brazil is a union of 27 federative units, split into 26 states and one federal district. The LM plots are based on these 27 administrative units. The maps in the following two LM plots make use of the boundaries after simplification and rescaling, as shown in Figure 1 (lower left).

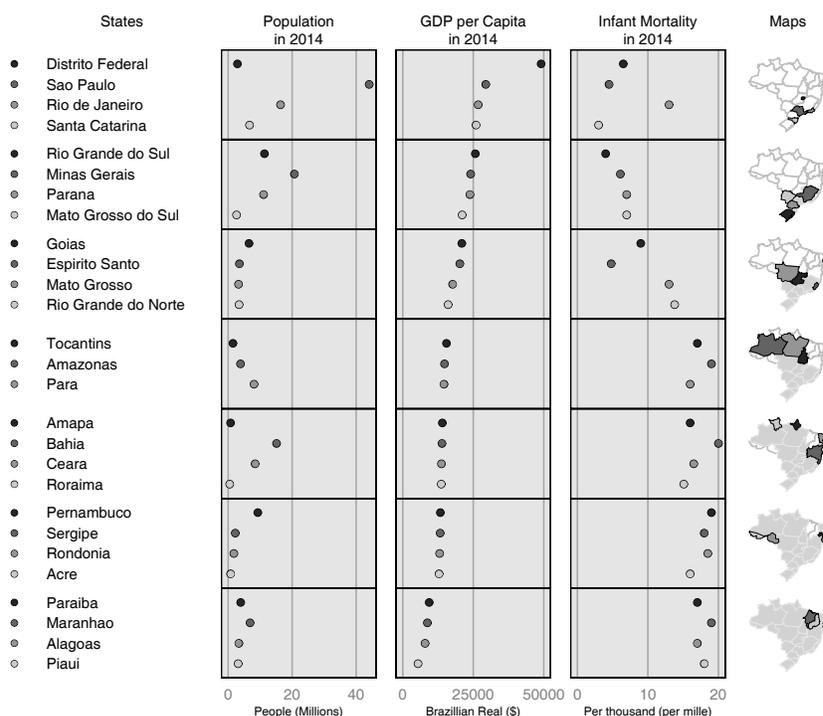


FIGURE 5: LM plot for Brazil, showing the 2014 population, the GDP in 2014, and the infant mortality in 2014 as dot plots in three statistical panels. The states are sorted according to the 2014 GDP in the middle statistical panel. In the maps, we use some incremental shading so that each new map also indicates by the light gray color which states appeared in the previous perceptual groups, i.e., in maps above the current map.

The first LM plot example for Brazil shown in Figure 5 is based on economic and health data from 2014 for each state, in particular population, the GDP per capita, and infant mortality (per thousand). The data have been obtained from the Wikipedia web page at http://en.wikipedia.org/wiki/States_of_Brazil. The states are sorted in order of their GDP per capita in 2014, with the highest value of about 50,000 Brazilian Real found in the federal district (Distrito Federal), which has some of the lowest total population. Overall, the highest GDP per capita

can be found in the southeastern and southern states, while the lowest GDP can be found in the western, northern, and northeastern states. When comparing GDP per capita with the infant mortality, two patterns emerge. In the first three perceptual groups (i.e., about 12 states), there is a strong negative association between GDP per capita and infant mortality. But, there are also two outliers: Rio de Janeiro has an unusual high infant mortality and Espirito Santo has a somewhat lower infant mortality than what might be expected. For the remaining 15 states in the last four perceptual groups, there is hardly any noticeable relationship between GDP per capita and infant mortality. Rather, for these states, the infant mortality somewhat arbitrarily ranges from about 15 to 20 per thousand. However, the GDP per capita is about three times as high in Tocantins, the highest-ranked of these 15 states, compared to Piaui, the lowest-ranked state.

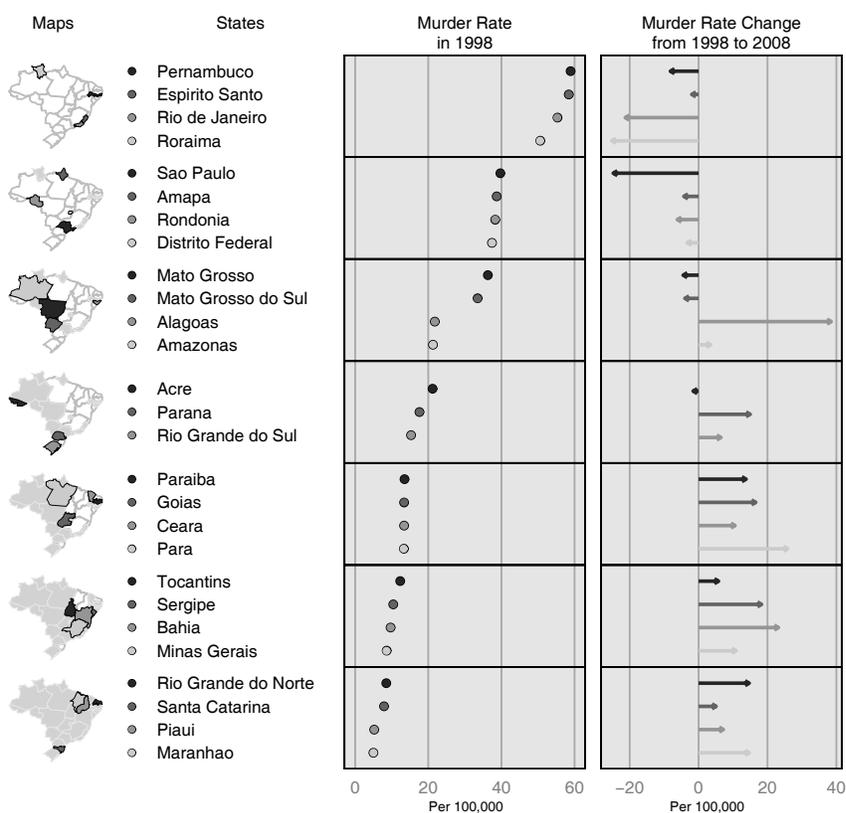


FIGURE 6: LM plot for Brazil, showing the 1998 murder rate and the murder rate change from 1998 to 2008 in two statistical panels. The first one shows a dot plot while the second one shows arrows for the changes. In the maps, we use incremental shading again.

The second LM plot example for Brazil shown in Figure 6 is based on the murder rates for each state. The data have been obtained from the Wikipedia

web page at http://en.wikipedia.org/wiki/List_of_Brazilian_states_by_murder_rate. Data are available for the time period from 1998 to 2008. The first statistical panel shows the murder rates in 1998 and has been used for sorting of the states. Murder rates have been similar in some of the neighboring states, but no major national pattern can be discerned. The second statistical panel shows the murder rate changes per 100,000 from 1998 to 2008. Here, arrows are used to show the direction of the change. Interestingly, the murder rates of the states with the highest murder rates in 1998 dropped in 2008, while the states with low murder rates in 1998 experienced higher murder rates in 2008.

Different users may have different preferences how to arrange the panels in an LM plot, e.g., map panel — name panel — statistical panel(s), name panel — statistical panel(s) — map panel, or some other meaningful arrangement. On the NCI LM plot web site, a user can change the placement of the map panel via the options menu. In the *micromap* R package, the arrangement of the different panels can be done relatively easily via two arguments in the plotting function. In Figures 3 and 6, we show the maps on the left and in Figures 4 and 5, we show the maps on the right. Any reader who carefully compares the designs of these four LM plots will notice that we also have experimented with some of the other LM plot design features that are provided by the *micromap* R package. For more details, the reader is referred to the help functions of this R package.

4. Conclusion and Outlook

We have demonstrated in this article how a set of existing geographic representations of the administrative regions for countries in South America can be modified within R for use in LM plots. Specifically, we have shown how the boundaries of regions can be simplified, small areas rescaled to be more visible, and far away areas shifted to appear next to the overall region of interest. Further, we have introduced automated techniques to implement these steps in R and construct a set of regional LM plots for South America. Some details are left to the reader.

One limitation of this approach in R involves the introduction of slivers and gaps in the simplification techniques available which complicate rescaling and replacement of subregions. The incorporation of these functions and methods for simplifying, rescaling, and shifting in R, however, adds greater utility to the production of LM plots in combination with the ability to read in and consume any shapefile or *SpatialPolygonsDataFrames* in R and produce an LM plot using the *micromap* R package.

For now, we have posted a zip file that contains all R code, shapefiles, and data files at <http://www.math.usu.edu/~symanzik/>. After unzipping, this file will allow the reader to fully reproduce the simplified shapefiles and all figures from this article. We are currently testing our approach for different geographic regions and shapefiles from different sources. Further work likely will include the addition of these newly developed functions to the *micromap* R package and the development of specialized boundaries, similar to Brazil. Also, it might be worthwhile to create boundaries for regional micromaps outside of R in specialized GIS software and make these boundaries available to interested R users.

Acknowledgment

The information in this article has been funded wholly (or in part) by the US Environmental Protection Agency. It has been subjected to review by the National Health and Environmental Effects Research Laboratory's Western Ecology Division and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

[Recibido: mayo de 2014 — Aceptado: octubre de 2014]

References

- Ahn, J. Y. (2013), Visualizing Statistical Information using Korean Linked Micromap Plots, in S.-H. Cho, ed., 'Proceedings of IASC-Satellite Conference for the 59th ISI WSC & The 8th Conference of IASC-ARS', Asian Regional Section of the IASC, pp. 219-221.
- Bivand, R. S. & Lewin-Koh, N. (2014), *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8-29.
*<http://CRAN.R-project.org/package=maptools>
- Bivand, R. S. & Rundel, C. (2014), *rgeos: Interface to Geometry Engine — Open Source (GEOS)*. R package version 0.3-4.
*<http://CRAN.R-project.org/package=rgeos>
- Bonnal, L., Favard, P., Laurent, T. & Ruiz-Gazen, A. (2011), 'Pourquoi le coût de l'éducation est-il plus élevé en zone rurale? Le cas de la région Midi-Pyrénées', *Revue d'Économie Régionale & Urbaine* **2011/5**, 887-910.
- Brunsdon, C. & Chen, H. (2014), *GISTools: Some further GIS Capabilities for R*. R package version 0.7-2.
*<http://CRAN.R-project.org/package=GISTools>
- Carr, D. B., Bell, B. S., Pickle, L. W., Zhang, Y. & Li, Y. (2003), The State Cancer Profiles Web Site and Extensions of Linked Micromap Plots and Conditioned Choropleth Map Plots, in 'Proceedings of the Third National Conference on Digital Government Research', Digital Government Research Center (DGRC), pp. 269-273.
*<http://dl.acm.org/citation.cfm?id=1123196>
- Carr, D. B., Chen, J., Bell, B. S., Pickle, L. W. & Zhang, Y. (2002), Interactive Linked Micromap Plots and Dynamically Conditioned Choropleth Maps, in 'Proceedings of the Second National Conference on Digital Government Research', Digital Government Research Center (DGRC), pp. 61-67.
*<http://dl.acm.org/citation.cfm?id=1123098>

- Carr, D. B. & Pearson, J. B. (2014), *micromapST: Linked Micromap Plots for U.S. States*. R package version 1.0.3.
*<http://CRAN.R-project.org/package=micromapST>
- Carr, D. B. & Pickle, L. W. (2010), *Visualizing Data Patterns with Micromaps*, Chapman & Hall/CRC, Boca Raton, FL.
- Carr, D. B. & Pierson, S. M. (1996), 'Emphasizing statistical summaries and showing spatial context with micromaps', *Statistical Computing and Statistical Graphics Newsletter* **7**(3), 16–23.
- Douglas, D. H. & Peucker, T. K. (1973), 'Algorithms for the reduction of the number of points required to represent a digitized line or its caricature', *The Canadian Cartographer* **10**(2), 112–122.
- Gebreab, S. Y., Gillies, R. R., Munger, R. G. & Symanzik, J. (2008), 'Visualization and interpretation of birth defects data using linked micromap plots', *Birth Defects Research (Part A): Clinical and Molecular Teratology* **82**, 110–119.
- Han, K. S., Park, S. J., Mun, G. S., Choi, S. H., Symanzik, J., Gebreab, S. & Ahn, J. Y. (2014), 'Linked micromaps for the visualization of geographically referenced data', *ICIC Express Letters* **8**(2), 443–448.
- Harrower, M. & Bloch, M. (2006), 'MapShaper.org: A map generalization web service', *IEEE Computer Graphics and Applications* **26**(4), 22–27.
- Höhle, M., Meyer, S. & Paul, M. (2013), *surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena*. R package version 1.7–0.
*<http://CRAN.R-project.org/package=surveillance>
- Mast, B. D. (2013), 'Visualizing same-sex couple household data with linked micromaps', *Cityscape: A Journal of Policy Development and Research* **15**(2), 267–271.
- Olsen, A. R., Carr, D. B., Courbois, J. P. & Pierson, S. M. (1996), Presentation of Data in Linked Attribute and Geographic Space, in '1996 Abstracts, Joint Statistical Meetings, Chicago, Illinois', American Statistical Association, Alexandria, VA, p. 271.
- Payton, Q. C., McManus, M. G., Weber, M. H., Olsen, A. R. & Kincaid, T. M. (2014), 'micromap: A Package for Linked Micromaps', *Journal of Statistical Software, Under Review* **15**(11).
*<http://www.jstatsoft.org/v07/i11/>
- Payton, Q. C. & Olsen, A. R. (2014), *micromap: Linked Micromap Plots*. R package version 1.8.
*<http://CRAN.R-project.org/package=micromap>

- Payton, Q. C., Weber, M. H., McManus, M. G. & Olsen, A. R. (2012), Linked Micromaps: Statistical Summaries in a Spatial Context, *in* 'Water: One Resource — Shared Effort — Common Future, 8th National Monitoring Conference, April 30–May 4, 2012, Portland, Oregon', National Water Quality Monitoring Council.
- Pickle, L. W., Pearson, J. & Carr, D. B. (2014), 'micromapST: Exploring and communicating geospatial patterns in U.S. State data', *Journal of Statistical Software, Under Review* **15**(11).
*<http://www.jstatsoft.org/v07/i11/>
- R Core Team (2014), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<http://www.R-project.org/>
- Stabler, B. (2013), *shapefiles: Read and Write ESRI Shapefiles*. R package version 0.7.
*<http://CRAN.R-project.org/package=shapefiles>
- Symanzik, J., Axelrad, D. A., Carr, D. B., Wang, J., Wong, D. & Woodruff, T. J. (1999), HAPs, Micromaps and GPL — Visualization of Geographically Referenced Statistical Summaries on the World Wide Web, *in* 'Annual Proceedings (ACSM–WFPS–PLSO–LSAW 1999 Conference CD)', American Congress on Surveying and Mapping.
- Symanzik, J. & Carr, D. B. (2008), Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data, *in* C. Chen, W. Härdle & A. Unwin, eds, 'Handbook of Data Visualization', Springer, Berlin, Heidelberg, pp. 267–294 & 2 Color Plates.
- Symanzik, J. & Carr, D. B. (2013), Linked Micromap Plots in R, *in* S.-H. Cho, ed., 'Proceedings of IASC–Satellite Conference for the 59th ISI WSC & The 8th Conference of IASC–ARS', Asian Regional Section of the IASC, pp. 213–218.
- Symanzik, J., Carr, D. B., Axelrad, D. A., Wang, J., Wong, D. & Woodruff, T. J. (1999), Interactive Tables and Maps — A Glance at EPA's Cumulative Exposure Project Web Page, *in* '1999 Proceedings of the Section on Statistical Graphics', American Statistical Association, Alexandria, VA, pp. 94–99.
- Symanzik, J., Wong, D., Wang, J., Carr, D. B., Woodruff, T. J. & Axelrad, D. A. (2000), Web-based Access and Visualization of Hazardous Air Pollutants, *in* 'Geographic Information Systems in Public Health: Proceedings of the Third National Conference August', Agency for Toxic Substances and Disease Registry, San Diego, California, pp. 18–20.
- Wang, X., Chen, J. X., Carr, D. B., Bell, B. S. & Pickle, L. W. (2002), 'Geographic statistics visualization: Web-based linked micromap plots', *Computing in Science & Engineering* **4**(3), 90–94.

Appendix. R Code Segments

We have read in our spatial boundaries for Brazil from the Global Administrative Areas (GADM) database at <http://gadm.org/>. Alternatively, the boundaries for the countries from South America shown in Figure 2 were obtained from the DIVA-GIS database at <http://www.diva-gis.org/> that is partially based on GADM version 1.0, but this database also provides access to spatial data from other sources. For Brazil, we simply used the convenience of reading in R *SpatialPolygonsDataFrames* directly using code such as:

```
R> con <- url("http://gadm.org/data/rda/BRA_adm1.RData")
R> print(load(con))
R> close(con)
```

One could download boundaries as shapefiles just as easily and then read these into R using a function such as `readOGR` in the *rgdal* R package or `readShapePoly` in the *maptools* R package.

For our R functions mentioned in this article, we provide our newly written R code below. After some more testing with other regions and countries, this code and functionality will be added to the *micromap* R package. Ideas and R code for the shifting of a subregion are based on a discussion started on June 22, 2013, at <http://gis.stackexchange.com/questions/64187/how-to-shift-polygons-by-updating-their-details-in-r>.

Function to determine the percent of area of each subregion in a region:

```
R> AreaPercent <- function(x) {
  tot_area <- sum(sapply(slot(x, "polygons"), slot, "area"))
  sapply(slot(x, "polygons"), slot, "area") / tot_area * 100
}
```

Function to rescale a subregion:

```
R> RescaleArea <- function(x, unit, rescale_factor) {
  b <- gBuffer(x[unit, ], width = rescale_factor, byid = TRUE)
  return(b)
}
```

Function to shift a subregion:

```
R> ShiftArea <- function(region, subregion, latlon, offset) {
  for (i in 1:length(region@polygons[[subregion]]@Polygons)) {
    cds <- slot(slot(slot(region, "polygons")[[subregion]],
                      "Polygons")[[i]], "coords")
    # get coords for vertices and label points
    l1 <- slot(slot(slot(region, "polygons")[[subregion]],
                    "Polygons")[[i]], "labpt")
    l2 <- slot(slot(region, "polygons")[[subregion]],
```

```
      "labpt")
# shift vertices coords
cds[, latlon] <- cds[, latlon] + offset
# shift lapt1 (centerpoint)
newl1 <- l1[latlon] + offset
# shift lapt2 (centerpoint)
newl2 <- l2[latlon] + offset
# put these shifted points back into the polygon object
slot(slot(slot(region, "polygons")[[subregion]],
           "Polygons")[[i]], "coords") <- cds
slot(slot(slot(region, "polygons")[[subregion]],
           "Polygons")[[i]], "labpt") <- newl1
slot(slot(region, "polygons")[[subregion]],
      "labpt") <- newl2
}
return(region)
}
```