



Journal of Statistical Software

January 2015, Volume 63, Issue 2.

<http://www.jstatsoft.org/>

micromap: A Package for Linked Micromaps

Quinn C. Payton
Oregon State
University

Michael G. McManus
US Environmental
Protection Agency

Marc H. Weber
US Environmental
Protection Agency

Anthony R. Olsen
US Environmental
Protection Agency

Thomas M. Kincaid
US Environmental
Protection Agency

Abstract

The R package **micromap** is used to create linked micromaps, which display statistical summaries associated with areal units, or polygons. Linked micromaps provide a means to simultaneously summarize and display both statistical and geographic distributions by linking statistical summaries to a series of small maps. The package contains functions dependent on the **ggplot2** package to produce a row-oriented graph composed of different panels, or columns, of information. These panels at a minimum typically contain maps, a legend, and statistical summaries, with the color-coded legend linking the maps and statistical summaries. We first describe the layout of linked micromaps and then the structure required for both the spatial and statistical datasets. The function `create_map_table` in the **micromap** package converts the input of an `sp SpatialPolygonsDataFrame` into a data frame that can be linked with the statistical dataset. Highly detailed polygons are not appropriate for display in linked micromaps so we describe how polygon boundaries can be simplified, decreasing the time required to draw the graphs, while retaining adequate detail for detection of spatial patterns. Our worked examples of linked micromaps use public health data as well as environmental data collected from spatially balanced probabilistic surveys.

Keywords: linked micromaps, GIS, spatial.

1. Introduction

A series of recent articles discuss the merits, limits, and opportunities for collaboration in the fields of statistical graphics and information visualization, or InfoVis ([Few January 20,](#)

2011; Gelman and Unwin 2013a,b; Kosara 2013; Murrell 2013; Wickham 2013). InfoVis is part of computer science that explores, analyzes, and presents large amounts of data (Kosara 2013; Wickham 2013). A type of graphic that uses strengths from both fields is a linked micromap. In “Visualizing Data Patterns with Micromaps”, Carr and Pickle (2010) first present principles of data visualization design, such as the use of small multiples (Tufte 1983), and then apply them to georeferenced statistical summaries. Such summaries occur when environmental data are collected and aggregated over areal units, or polygons. Examples include crop statistics per state, public health data such as number of cancer cases per county, or monitoring data collected among reporting regions from a spatially balanced probabilistic survey (Olsen, Kincaid, and Payton 2012). For these data, displaying both the geographic and statistical distributions is of interest, and linked micromap plots, often just called micromaps, are a way to visualize such geographically referenced data. (Carr and Pierson 1996; Carr, Olsen, Courbois, Pierson, and Carr 1998; Carr, Wallin, and Carr 2000; Symanzik and Carr 2008; Carr and Pickle 2010).

Micromaps are row-oriented graphs with different panels, or columns, of information, which typically contain at a minimum maps, a color-coded legend, and statistical summaries (Figure 1) and are a more informative alternative to choropleth maps (Carr and Pierson 1996; Symanzik and Carr 2008). Unlike a choropleth map in which color is associated with the magnitude of a variable, the color-coded legend in a micromap provides the link between the maps and statistical summaries. Two features of a micromap are the sorting and grouping in the display of the data. The sorting is done on the values of a statistical variable, and the data are grouped so only few, typically four to six, areal units and their statistical summaries are presented across a row as a perceptual group (Carr and Pickle 2010). This use of perceptual groups allows the viewer to discern cumulative geographical patterns. For example, in the poverty and education micromap (Figure 1) the accumulation of gray polygons in the map panel reveals that, except for Virginia, all of the southeastern states have poverty levels greater than the median poverty level. Another significant advantage of this type of visualization over choropleth maps is that multiple variables can be simultaneously displayed.

The means with which users may readily produce micromaps has been varied. Some United States agencies have web sites or downloadable software to make micromaps, but analysts cannot access the software itself (<http://statecancerprofiles.cancer.gov/micromaps/>, <http://gis.cancer.gov/tools/micromaps/>, http://nass.usda.gov/Charts_and_Maps/). Analysts can use geographic information system (GIS) software to make the map component of micromaps, and statistical software to produce the statistical summaries, and then bring those two components together to produce a micromap. Symanzik and Carr (2008) noted that most major statistical packages could produce micromaps, the requirements being: availability of polygons, or boundary files, ease of production, quality of appearance, and control of output format. In R (R Core Team 2014), the **grid** package (Murrell 2005) has been used to produce micromaps of large financial datasets (Blunt 2006). A new package, **micromapST**, allows users to create linked micromaps specifically for US states and the District of Columbia (Carr, Pearson, and Pickle 2013).

Here we describe the **micromap** package (Payton and Olsen 2014), created for R and available on CRAN, which has been created to enable users to create linked micromaps for any areal spatial dataset. Similar to other packages such as **GGally** (Schloerke, Crowley, Cook, Hofmann, Wickham, Briatte, and Marbach 2013), the **micromap** package was designed to take advantage of the graphical functionality of the **ggplot2** package. The **ggplot2** package allows

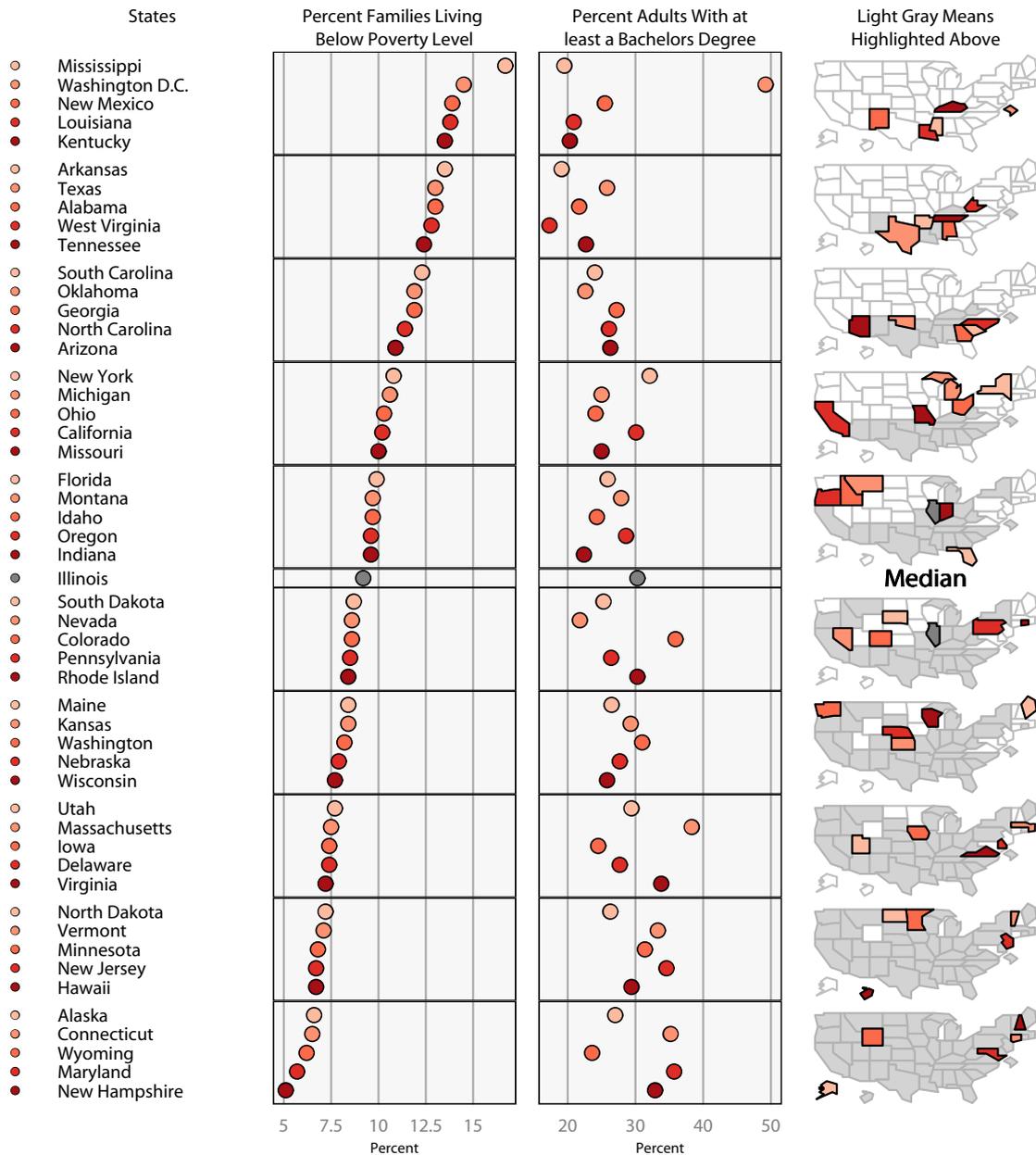


Figure 1: A micromap of poverty and education level in the United States illustrating the components of a micromap, which include panels, or columns, a color legend, labels for the polygons, statistical summaries of the polygons, and maps that share a color code across the panels giving a spatial context to the summary statistics. The data are sorted by poverty level and displayed by perceptual groups of five states per group. The data are from the 2006–2010 American Community Survey and were obtained from: <http://statecancerprofiles.cancer.gov/micromaps/>.

for a predictable alignment of panels and perceptual groups, leaving the user free to concern themselves purely with the finer aesthetic details of the plot. A challenge to working with statistical and spatial data noted by Wickham (2009) is the need to match the identifiers in the statistical data to the corresponding identifiers in the spatial data. The general idea behind the **micromap** package is to make this as intuitive as possible while allowing the user to make aesthetic changes to the entire plot (perceptual group number and size, color coding, header sizes, etc.) rather than panel by panel. This should create a seamless feel among the various panels being produced and assemble them effortlessly into a single display.

Our objective with the **micromap** package is to simplify the making of linked micromaps so this type of visualization can readily be used by analysts to represent georeferenced statistics (Carr and Pickle 2010). We meet that objective by describing four steps to create micromaps. First, we identify what geoprocessing needs to be done to the spatial data prior to making a micromap. Second, we describe how in the **micromap** package the spatial polygon data frame is restructured so the spatial data can be linked to the statistical data. Third, we use the `mmpplot` function to call the two data frames, spatial and statistical, to create the different panels and rows of perceptual groups for a draft micromap. Producing this draft micromap allows the user to check that the data frames are structured correctly and to decide if coding for a more refined, publication quality micromap is warranted. Finally, we provide examples of three types of micromaps, and also describe how users can create their own types. Section 2 of this paper describes the preparation of the spatial data. Section 3 covers the main functions in the **micromap** package. Section 4 gives examples of the three micromap plot types, including the new type called a group categorized micromap. Finally, Section 5 describes challenges and future developments to making micromaps.

2. Preparation of spatial data

2.1. Map simplification

The polygons displayed in the maps of micromaps only need to be detailed enough to convey shape and relative position of the areas and provide a spatial framework by identifying neighboring polygons (Carr and Pickle 2010). In fact, overly detailed polygons will create difficulty in display and interpretation of linked micromap plots. The early applications of micromaps linked statistical estimates to administrative polygons such as countries and states (Carr and Pierson 1996; Carr, Olsen, Courbois, Pierson, and Carr 1998). An example of simplified administrative polygons suitable for use in micromaps is the United States State Visibility Base Map developed by Monmonier (1993) available in the **maps** package (Becker, Wilks, Brownrigg, and Minka 2013). These generalized polygons, or map caricatures, are faster to draw for micromaps (Carr and Pierson 1996; Carr, Wallin, and Carr 2000).

For users working with other areal units such as watershed polygons or ecoregions, those data are often overly detailed and will render the map difficult to read (Gebreab, Gillies, Munger, and Symanzik 2008). Such data will also slow the `mmpplot` function in drawing the map panel. Several approaches can be used to generalize, or smooth, the polygons so they contain fewer vertices. One approach relies on working with GIS software, such as ArcMap (ESRI 2013), that has simplification functions. The resulting simplified shapefile can be read into R using the `readOGR` function from the **rgdal** (Bivand, Keitt, and Row-

ingson 2013) package that is loaded with the **micromap** package or one of the functions in the other R spatial libraries that read in shapefile data. Another approach is to use some of the topology-preserving simplification tools available in Javascript such as **mapshaper** (Bloch 2014). Also see descriptions at <http://www.jasondavies.com/maps/simplify/> and <http://bost.ocks.org/mike/simplify/>. Finally, one may simplify polygons within R using polygon simplification functions such as the `thinnedSpatialPoly` function in **maptools** (Bivand and Lewin-Koh 2013), `dp` in the **shapefiles** package (Stabler 2013), `generalize.polys` in **GISTools** (Brundson and Chen 2012) or the `gsimplify` function in **rgeos** (Bivand and Rundel 2013). The simplification functions in both ArcGIS and R use the Douglas-Peucker algorithm (Douglas and 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. Additionally in ArcGIS a bend simplify algorithm is available (Wang and Muller 1998) that applies shape recognition techniques to simplify lines and tends to produce more cartographically pleasing results than the point remove algorithm. Simplifying polygons without preserving topology can result in visual artifacts in the maps such as overlapping polygons, slivers and gaps. Due to the diversity of simplification functions in ArcGIS, we have found that to be the most efficient way to simplify polygons.

2.2. Simplification approaches

Both the GIS software and R approaches are illustrated in the **micromap** package user guide. In the GIS software approach and using the `thinnedSpatialPoly` function in R a user can specify a minimum area of polygons to retain in simplifying, and care is needed to avoid mistakenly deleting any polygons that should appear on the micromap.

There are two methods for simplifying polygons in R. The **maptools** method:

```
R> library("micromap")
R> load("Micromap_Example_Code_From_Paper_Data.RData")
R> library("maptools")
R> Polys_v1 <- thinnedSpatialPoly(txeco, tolerance = 7000, minarea = 0.001,
+   avoidGEOS = TRUE)
```

and the **rgeos** method:

```
R> library("rgeos")
R> Polys_v2 <- gSimplify(txeco, 7000, topologyPreserve = TRUE)
R> Polys_v2 <- SpatialPolygonsDataFrame(Polys_v2, txeco@data)
```

In the user guide we also provide details on steps for simplifying very large spatial datasets. Another geoprocessing step that may be necessary is to manipulate the layout and size of the polygons that are to be brought into the **micromap** package. For the map caricature, small polygons may need to be enlarged while retaining sufficient features so they are recognizable (Carr and Pierson 1996). For example, in Figure 1 notice how Alaska and Hawaii are offset to be below the southwest border of the United States, and that Washington D.C. is enlarged and set just off the east coast. These steps may require some degree of editing within a GIS application or R, and functions for offsetting and enlargement of polygons may be included in future version of the **micromap** package.

3. Functions in the micromap package

The user accessible functions contained in the **micromap** package can be classified into two categories: two main plot creation functions and a small number of utility functions for more meticulous control over that process. The plot creation functions, `mmpplot` and `mmgroupedplot`, allow the user to create the most commonly created linked micromaps with all aesthetic preferences being specified using a single list of arguments provided to a single function call. The utility functions within the **micromap** package allow for ad hoc adjustments and fine tuning. The **micromap** package is aimed at facilitating the consolidation of statistical data, in the form of a data frame, with GIS spatial objects. To simplify this process, we have included the `create_map_table` function which transforms a spatial object into a flat data frame object. The `create_map_table` function is similar to the **ggplot2** package's `fortify` function. The `create_map_table` function creates a data frame object which is specifically designed for use with the `mmpplot` and `mmgroupedplot` functions. This entails organizing collections of polygons which include “holes”, “islands”, and what we refer to as “plugs” — that is, islands that overlay the “hole” from another collection of polygons. Plotting each of these polygons can require a significant amount of memory; therefore we have attempted to eliminate the redundancy inherent in plotting multiple polygons where only one will be visible. These features of `create_map_table` offer significant advantages in comparison to the `fortify` function. The output from the `create_map_table` is used with the `mmpplot` and `mmgroupedplot` functions to draw the polygons for the map panel.

There is a considerable amount of variation in the level of complexity a user may prefer in presenting their statistical data graphically. With this in mind, the **micromap** package has been designed to facilitate as many needs as possible. For those wishing to make simple use of the most common graphs associated with spatial data (box plots, dot plots, and bar plots), these graph types have been included along with various easily modifiable customizations. However, in consideration of users who may desire more unique or specialized graphs for their data, a series of specialized functions are included to allow a user to create new and unique graph types. The user guide that accompanies the **micromap** package walks through the process in great detail. Users are encouraged to submit any unique graph additions to the creators for inclusion in further updates of the **micromap** package.

The `print` method associated with `mmpplot` objects is designed to allow users to make post-hoc changes and reformat output files of the plots they've created. The **ggplot2** package is designed to allow a user to create objects which may then be adjusted (altering color schemes, axis labels, etc.) without having to rerun the code used to create the plot in its entirety. The `mmpplot` and `mmgroupedplot` functions return these **ggplot2** objects as a list within a custom `mm` class object. Each **ggplot2** object in this list may then be adjusted which allows the user to then reassemble the linked micromap easily. This function also allows a user to take a previously designed micromap and print directly to a file without needing to recreate the plot in R.

4. Examples

The following examples encompass the three basic types of plots available in the **micromap** package, which include a box plot, dot plot, and bar plot. The dot and bar plots can also be displayed with standard errors or confidence limits. Displaying such measures of variation

is an advantage micromaps have over choropleth maps (Symanzik and Carr 2008). The first example is from the US EPA National Lakes Assessment (NLA) in which during the summer of 2007 the US Environmental Protection Agency, states, tribes, and other partners conducted a nationwide survey of the condition of lakes in the conterminous US (US Environmental Protection Agency 2009). The NLA used a spatially balanced probabilistic survey of 1028 lakes in the lower 48 states and as part of that design lakes were stratified by nine National Aquatic Resources Survey (NARS) reporting regions. A variable measured at each lake was pH, with lower values indicating greater acidity and a pH of 7 being neutral. We used the **micromap** package to produce a micromap showing two panels of statistical summaries that are linked to the nine reporting regions (Figure 2).

The polygons of the reporting regions were simplified as in Section 2, and a map table created from the WSA9 (the nine NARS reporting regions) SpatialPolygonsDataFrame using the `create_map_table` function. That map table now has an ID variable that is used to link to the statistical data.

```
R> wsa.polys <- create_map_table(WSA9, "WSA_9_NM")
R> head(wsa.polys)
```

	ID	region	poly	coordsx	coordsy	hole	plug
1	Coastal Plains	1	623	1465123	-830258.1	0	0
2	Coastal Plains	1	623	1480642	-868249.9	0	0
3	Coastal Plains	1	623	1498419	-864833.4	0	0
4	Coastal Plains	1	623	1463862	-823721.8	0	0
5	Coastal Plains	1	623	1505524	-871703.7	0	0
6	Coastal Plains	1	623	1512576	-885503.2	0	0

In the statistical data `NLA_pH.ds5`, `ecoregions` is the variable that corresponds to the ID variable (Table 1).

The lines of code below produce Figure 2. The first few lines of code (through the specification of the `map.link`) can be used to create a draft micromap which a user may use to explore data and evaluate the usefulness of a more detailed micromap display. In the micromap code, the sequential order of the panel types corresponds directly to the sequential order of the list of panel data. That is, `"dot_legend"` panel type does not need input data and so is associated with an NA in the data list, `"labels"` is associated with `"Ecoregions"`, `"box_summary"` is associated with the five number summary of pH from the statistics data frame, etc. The map also has an NA designation in the panel data list, since the source of that data is specified in the `map.data` argument. The `map.link` connects the reporting regions polygon labels in `stat.data` to their corresponding ID polygons in `map.data`.

```
R> mmpplot(stat.data = NLA_pH.ds4, map.data = wsa.polys,
+ panel.types = c("dot_legend", "labels", "box_summary", "dot_cl", "map"),
+ panel.data = list(NA, "Ecoregions",
+ list("Min", "Q1", "Median", "Q3", "Max"),
+ list("Estimate", "LCB95Pct", "UCB95Pct"), NA),
+ ord.by = "Estimate", grouping = 3, median.row = FALSE,
+ map.link = c("Ecoregions", "ID"), plot.height = 7, plot.width = 7,
+ colors = brewer.pal(6, "BuPu")[4:6],
```

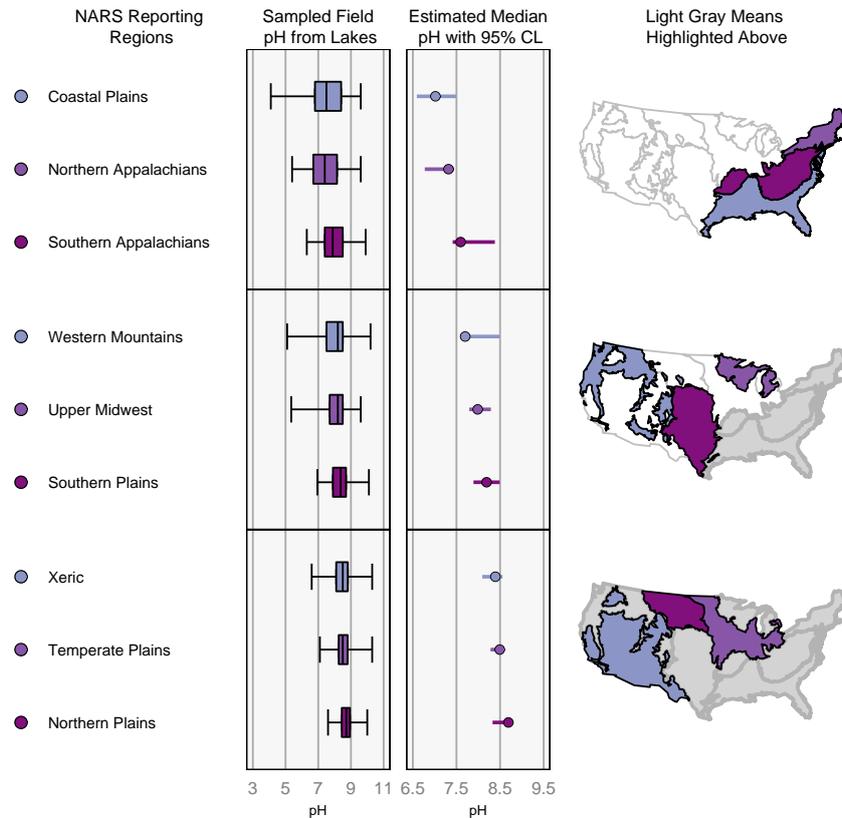


Figure 2: The sampled and population estimates of lake pH among the nine NARS Reporting Regions from the NLA. This micromap used both the "box_summary" and the "dot_cl" options to produce the two panels of statistical summaries. The population estimates were obtained using the **spsurvey** package.

```

+   map.color2 = "lightgray",
+   panel.att = list(
+     list(1, point.type = 20, point.size = 2),
+     list(2, header = "NARS Reporting\n Regions", panel.width = 1.25,
+       align = "left", text.size = 0.8),
+     list(3, header = "Sampled Field pH\n from Lakes",
+       graph.bgcolor = brewer.pal(5,"Greys")[1],
+       xaxis.ticks = list(3.0, 5.0, 7.0, 9.0, 11.0),
+       xaxis.labels = list(3.0, 5.0, 7.0, 9.0, 11.0),
+       xaxis.title='pH', graph.bar.size = 0.4),
+     list(4, header = "Estimated Median pH\n with 95% CL",
+       graph.bgcolor = brewer.pal(5,"Greys")[1],
+       xaxis.ticks = list(6.5, 7.5, 8.5, 9.5),
+       xaxis.labels = list(6.5, 7.5, 8.5, 9.5), xaxis.title = "pH"),
+     list(5, header = "Light Gray Means\n Highlighted Above",
+       inactive.border.color = gray(.7), inactive.border.size = 2,
+       panel.width = 2)))

```

	Estimate	LCB95Pct	UCB95Pct	Ecoregions	Min	Q1	Median	Q3	Max	N
1	7.02	6.59	7.49	Coastal Plains	4.10	6.80	7.50	8.40	9.60	100
2	7.32	6.77	7.45	Northern Appalachians	5.40	6.70	7.40	8.15	9.60	93
3	8.69	8.33	8.80	Northern Plains	7.60	8.44	8.72	8.94	10.00	65
4	7.60	7.41	8.38	Southern Appalachians	6.30	7.40	7.88	8.50	9.90	116
5	8.19	7.89	8.49	Southern Plains	6.95	7.90	8.37	8.70	10.10	126
6	8.49	8.28	8.60	Temperate Plains	7.10	8.25	8.50	8.80	10.30	136
7	7.99	7.79	8.29	Upper Midwest	5.35	7.70	8.20	8.50	9.60	146
8	7.70	7.59	8.49	Western Mountains	5.10	7.50	8.20	8.50	10.20	155
9	8.39	8.09	8.56	Xeric	6.60	8.10	8.50	8.81	10.30	84

Table 1: National lake assessment pH median estimates and five number summaries of pH for the nine regions.

	Subpopulation	Category	PTL.NResp	PTL.Estimate.P	PTL.LCB95Pct.P	PTL.UCB95Pct.P
1	National	LEAST DISTURBED	586	58.08	52.68	63.48
2	National	INTERMEDIATE DISTURBANCE	219	23.74	19.28	28.20
3	National	MOST DISTURBED	223	18.18	13.93	22.43
4	CPL	LEAST DISTURBED	42	47.70	32.23	63.18
5	CPL	INTERMEDIATE DISTURBANCE	38	37.04	22.24	51.85
6	CPL	MOST DISTURBED	22	15.25	5.10	25.40
7	NAP	LEAST DISTURBED	71	79.72	70.70	88.73
8	NAP	INTERMEDIATE DISTURBANCE	15	14.60	6.02	23.18
9	NAP	MOST DISTURBED	7	5.68	1.71	9.65

Table 2: An excerpt of rows from the table of National Lake Assessment water chemistry estimates by disturbance categories. Only phosphorus level estimates are displayed.

The first statistical panel displays the five-number summary of pH values from the sampled lakes as a box plot for each of the nine ecoregions. The second statistical panel displays the population estimates of the median pH and confidence limits, with these estimates and interval bounds calculated using the **spsurvey** package (Stevens and Olsen 2004; Kincaid and Olsen 2012). The micromap is sorted from the lowest to highest estimated median pH. The nine NARS reporting regions are split into perceptual groups of three regions each. That grouping, along with the linking through color of labels, statistical summaries, and ecoregion polygons assists the viewer in noting that estimated median pH values of ~ 7.5 or less occur in the eastern portion of the United States. We can also see from the box plot that increases in medians of the sampled pH values are associated with smaller interquartile ranges. Although this micromap displays just a small number of polygons, it allows viewers to discern both statistical and spatial patterns.

Statistical and spatial data that are comprised of a large number of polygons, such as states that have many counties, can be a challenge to display in a typical micromap layout of maps, region labels, and statistical panels (Carr 2001). For our second example, we examine public health data for the 255 counties of Texas, specifically incidence of leukemia and emissions data from roads and airports at the county level (Senkayi, Sattler, Rowe, and Chen 2014). Displaying all the counties for the state of Texas in a standard micromap layout would extend far beyond the vertical margins of a page layout. Creating such a micromap requires subsetting the data (done through row indexing the ordered data) in order to plot the data in groups over a series of micromap panels, as well as modifying micromap parameters for text and element sizes to work with the multi-panel layout (Figure 3). The ability to produce multi-panel micromaps extends the utility of micromaps for display of data composed of many regions.

Whereas the previous examples had a one-to-one relationship between a polygon and a statistical estimate, this last example, which uses the NLA data, has a one polygon to many statistical estimates relationship in which each reporting region is crossed with a categorical group that identifies a disturbance condition. The structure of the dataset is shown in Table 2.

To visualize this data requires a new style of micromap. This “group categorized” micromap we believe was first used in the Wadeable Streams Assessment (US Environmental Protection Agency 2006) and provides an alternative to pie charts or stacked bar charts. The `mmgroupedplot` function has been included in the **micromap** package to allow a user to create this type of map. Ten maps are displayed in the micromap in Figure 4, with the map at the top presenting the national summary statistics for the percent estimates, with confidence limits, of lakes that occurred in the categories of least disturbed, intermediate disturbance, and most disturbed based on their values in total phosphorus, total nitrogen, and turbidity.

The remaining maps show a specific reporting region and the estimates among the three categories. This comparison of a national summary to a region summary lets the viewer note that most of the lakes in the Northern Plains are in the most disturbed category based on their total phosphorus and total nitrogen values.

Besides the micromaps described here, users can create their own micromap by creating a new panel type. The **micromap** package user guide provides an example detailing the creation of an arrow linked micromap panel type which may be used to illustrate a temporal change in statistical estimates.

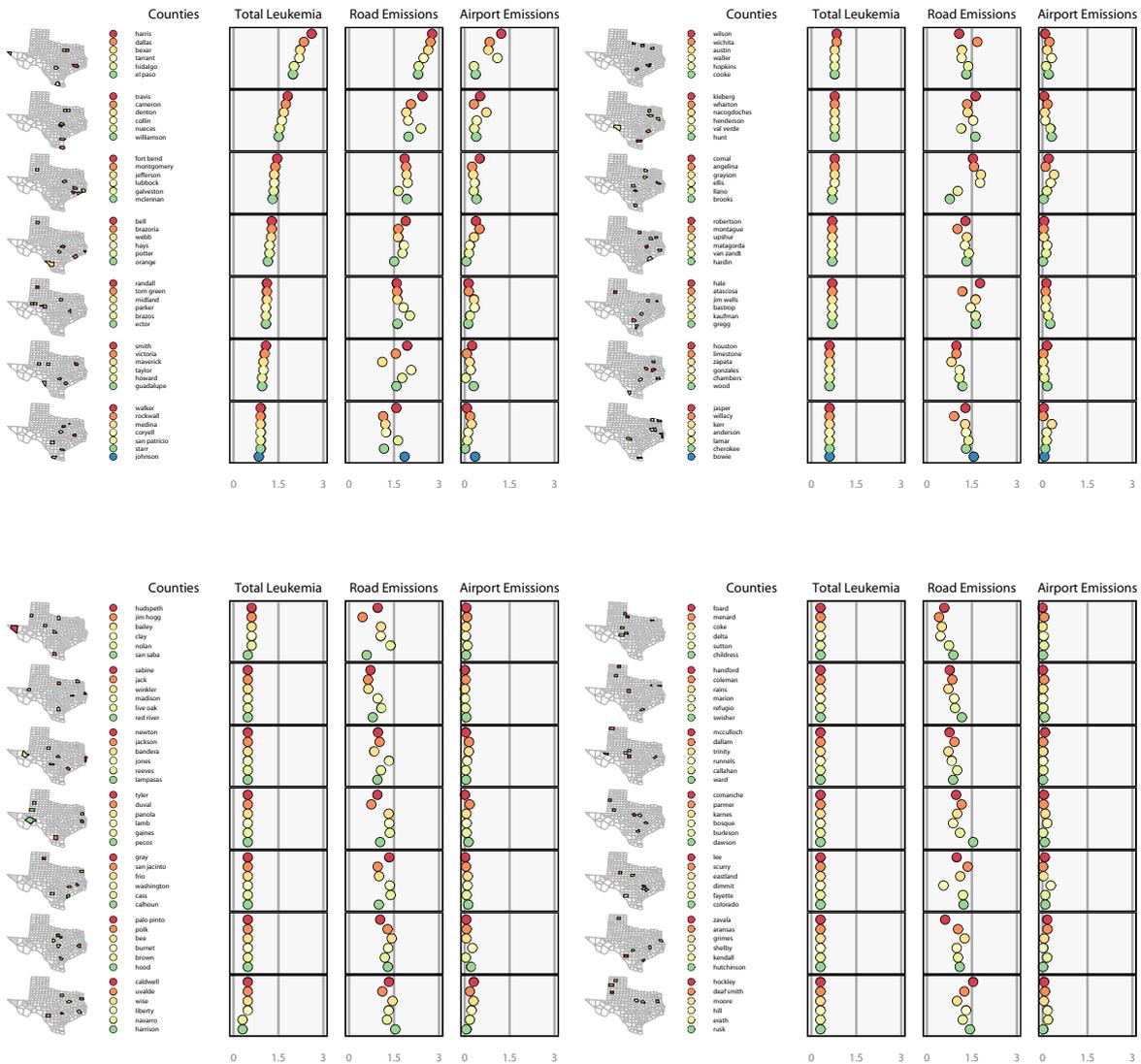


Figure 3: A linked micromap of county public health data. With so many areal units, linked micromaps of subsets of polygons are made and arranged on the page layout. A $\log(x + 1)$ transformation was used on the variables.

5. Challenges and future developments

A future challenge with the **micromap** package is to expand its capabilities to include plot types such as histograms, time series data, and possibly symbol plots. Shortly after micromaps were introduced, Carr (2001) noted the value of having micromaps in an interactive mapping environment. While that intent goes beyond the scope of the **micromap** package, some efforts have been made to create dynamic micromaps (Wang, Chen, Carr, Bell, and Pickle 2002; Hurst, Symanzik, and Gunter 2003). Interactive tables and micromaps are available for cancer statistics at the Pennsylvania Cancer Atlas (<http://www.geovista.psu.edu/grants/CDC/>) and at the United States Atlas of Cancer prototype (<http://www.geovista.psu.edu/grants/CDC/national.html>). Recently, the authors of the **GeoXP** package for interactive graphics

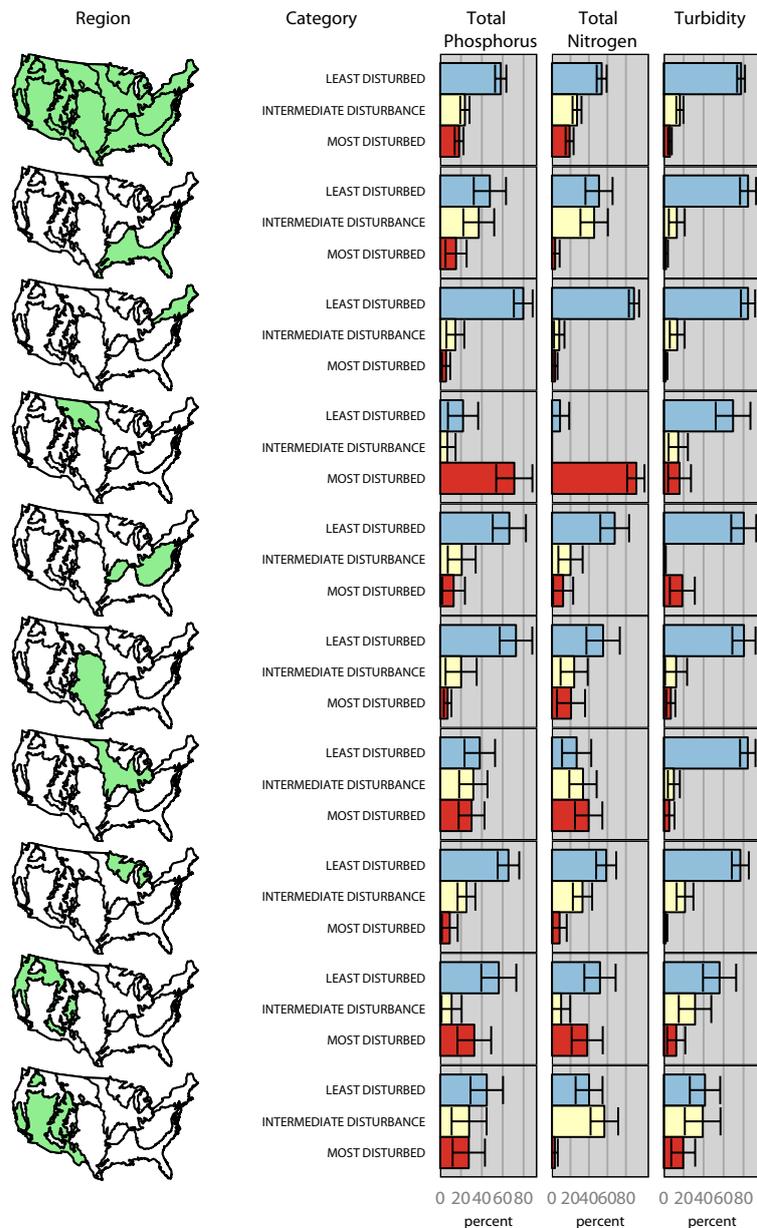


Figure 4: A grouped categorized micromap of water chemistry from the NLA. National and NARS Reporting Regional patterns are categorized by site condition of least, intermediate, and most disturbed based on the values of total phosphorus, total nitrogen, and turbidity.

for exploratory spatial data analysis mentioned that a new tool for that package might include a micromap display (Laurent, Ruiz-Gazen, and Thomas-Agnan 2012).

We believe publication quality micromaps can be made using the **micromap** package based on the four steps we have outlined. First, working in GIS software or R, users need to simplify polygons for display in a micromap. Fortunately, once such map caricatures are made they can be reused (Carr, Wallin, and Carr 2000). Second, in the **micromap** package, the spatial polygon data frame is converted to a map table that includes an identifying variable to link

the map table to the statistical data frame. Third, only approximately ten lines of code need to be used with the `mmp` function to render a draft micromap. Finally, users have detailed control over the layout, quality of appearance, and output format of the final micromap, and they can create their own types of micromaps. The four steps make for convenient production of micromaps, and we hope that visualizing georeferenced statistics in this manner will become more frequently used (Peterson 2011; Robbins 2012). The versatility of the micromap package is that analysts can use it on a range of areal units such as watersheds and ecoregions, as well as administrative units from other countries. This flexibility may require some modification of both the geospatial data and the micromap plot aesthetics in order to ultimately produce a publication quality figure.

Acknowledgments

Thanks to Sala Senkayi for use of Texas Leukemia data.

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.

References

- Becker RA, Wilks AR, Brownrigg R, Minka TP (2013). *maps: Draw Geographical Maps*. R package version 2.3-2, URL <http://CRAN.R-project.org/package=maps>.
- Bivand R, Keitt T, Rowlingson B (2013). *rgdal: Bindings for the Geospatial Data Abstraction Library*. R package version 0.8-6, URL <http://CRAN.R-project.org/package=rgdal>.
- Bivand R, Lewin-Koh N (2013). *maptools: Tools for Reading and Handling Spatial Objects*. R package version 0.8-23, URL <http://CRAN.R-project.org/package=maptools>.
- Bivand R, Rundel C (2013). *rgeos: Interface to Geometry Engine – Open Source (GEOS)*. R package version 0.2-16, URL <http://CRAN.R-project.org/package=rgeos>.
- Bloch M (2014). *mapshaper: A Tool for Topologically Aware Shape Simplification*. URL <http://www.mapshaper.org/>.
- Blunt G (2006). *Using grid Graphics to Produce Linked Micromap Plots of Large Financial Datasets*. PowerPoint presentation at the User! Conference, Vienna, Italy.
- Brundson C, Chen H (2012). *GISTools: Some Further GIS Capabilities for R*. R package version 0.7-1., URL <http://CRAN.R-project.org/package=GISTools>.
- Carr D, Pearson J, Pickle L (2013). *micromapST: Linked Micromap Plots for US States*. R package version 3.0-2, URL <http://CRAN.R-project.org/package=micromapST>.
- Carr DB (2001). "Designing Linked Micromap Plots for States with Many Counties." *Statistics in Medicine*, **20**(9-10), 1331– 1339.

- Carr DB, Olsen AR, Courbois JP, Pierson S, Carr DA (1998). “Linked Micromap Plots: Named and Described.” *Statistical Computing & Graphics Newsletter*, **9**, 24–32.
- Carr DB, Pickle LW (2010). *Visualizing Data Patterns with Micromaps*. Chapman & Hall/CRC, Boca Raton.
- Carr DB, Pierson S (1996). “Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps.” *Statistical Computing & Graphics Newsletter*, **7**, 16–23.
- Carr DB, Wallin JF, Carr DA (2000). “Two New Templates for Epidemiology Applications: Linked Micromap Plots and Conditioned Choropleth Maps.” *Statistics in Medicine*, **19**(17–18), 2521–38.
- Douglas D, Peucker T (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.
- ESRI (2013). *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute, Redlands. URL <http://www.arcgis.com/>.
- Few S (January 20, 2011). “Are InfoVis and Statistical Graphics Really that Different?” URL http://www.perceptualedge.com/files/are_infovis_and_statistical_graphics_really_all_that_different.pdf.
- Gebreab SY, Gillies RR, Munger RG, Symanzik J (2008). “Visualization and Interpretation of Birth Defects Data Using Linked Micromap Plots.” *Birth Defects Research Part A – Clinical and Molecular Teratology*, **82**(2), 110–119.
- Gelman A, Unwin A (2013a). “Infovis and Statistical Graphics: Different Goals, Different Looks.” *Journal of Computational and Graphical Statistics*, **22**(1), 2–28.
- Gelman A, Unwin A (2013b). “Tradeoffs in Information Graphics.” *Journal of Computational and Graphical Statistics*, **22**(1), 45–49.
- Hurst J, Symanzik J, Gunter L (2003). “Interactive Federal Statistical Data on the Web Using “nVIZn”.” *Computing Science and Statistics*, **35**.
- Kincaid TM, Olsen AR (2012). *spsurvey: Spatial Survey Design and Analysis*. R package version 2.5, URL <http://CRAN.R-project.org/package=spsurvey>.
- Kosara R (2013). “InfoVis Is so Much More: A Comment on Gelman and Unwin and an Invitation to Consider the Opportunities.” *Journal of Computational and Graphical Statistics*, **22**(1), 29–32.
- Laurent T, Ruiz-Gazen A, Thomas-Agnan C (2012). “GeoXp: An R Package for Exploratory Spatial Data Analysis.” *Journal of Statistical Software*, **47**(2), 1–23. URL <http://www.jstatsoft.org/v47/i02/>.
- Monmonier M (1993). *Mapping It Out: Expository Cartography for the Humanities and Social Sciences*. The University of Chicago Press, Chicago.
- Murrell P (2005). *R Graphics*. Chapman & Hall/CRC, Boca Raton.

- Murrell P (2013). “InfoVis and Statistical Graphics: Comment.” *Journal of Computational and Graphical Statistics*, **22**(1), 33–37.
- Olsen AR, Kincaid TM, Payton Q (2012). “Spatially Balanced Survey Designs for Natural Resources.” In RA Gitzen, JJ Millspaugh, AB Cooper, DS Licht (eds.), *Design and Analysis of Long-Term Ecological Monitoring Studies*, pp. 126–150. Cambridge University Press, Cambridge.
- Payton Q, Olsen AR (2014). *micromap: Linked Micromap Plots*. R package version 1.8, URL <http://CRAN.R-project.org/package=micromap>.
- Peterson G (2011). “Micromap Software.” URL <http://www.gretchenpeterson.com/blog/micromap-software/>.
- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Robbins N (2012). “Linked Micromaps for Geographically Referenced Data.” URL <http://www.forbes.com/sites/naomirobbs/2012/05/23/linked-micromaps-for-geographically-referenced-data>.
- Schloerke B, Crowley J, Cook D, Hofmann H, Wickham H, Briatte F, Marbach M (2013). *GGally: Extension to ggplot2*. R package version 0.4.4, URL <http://CRAN.R-project.org/package=GGally>.
- Senkayi S, Sattler ML, Rowe N, Chen VC (2014). “Investigation of an Association Between Childhood Leukemia Incidences and Airports in Texas.” *Atmospheric Pollution Research*, **5**, 189–195.
- Stabler B (2013). *shapefiles: Read and Write ESRI Shapefiles*. R package version 0.7, URL <http://CRAN.R-project.org/package=shapefiles>.
- Stevens DL, Olsen AR (2004). “Spatially Balanced Sampling of Natural Resources.” *Journal of the American Statistical Association*, **99**(465), 262–278.
- Symanzik J, Carr DB (2008). “Interactive Linked Micromap Plots for the Display of Geographically Referenced Statistical Data.” In C Chen, W Hardle, A Unwin (eds.), *Handbook of Data Visualization*, pp. 267–294. Springer-Verlag, Berlin.
- Tufte ER (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire.
- US Environmental Protection Agency (2006). “Wadeable Streams Assessment: A Collaborative Survey of the Nation’s Streams.” *Technical report*, Office of Water, US Environmental Protection Agency.
- US Environmental Protection Agency (2009). “National Lakes Assessment: A Collaborative Survey of the Nation’s Lakes. EPA 841 - R -09 - 001.”
- Wang XS, Chen JX, Carr DB, Bell BS, Pickle LW (2002). “Geographic Statistics Visualization: Web-Based Linked Micromap Plots.” *Computing in Science & Engineering*, **4**(3), 90–94.
- Wang Z, Muller JC (1998). “Line Generalization Based on Analysis of Shape Characteristics.” *Cartography and Geographic Information Science*, **25**(1), 3–15.

Wickham H (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.

Wickham H (2013). “Graphical Criticism: Some Historical Notes.” *Journal of Computational and Graphical Statistics*, **22**(1), 38–44.

Affiliation:

Quinn Payton
Department of Statistics
Oregon State University
44 Kidder Hall
Corvallis, OR 97331, United States of America
E-mail: paytonq@science.oregonstate.edu