Geovisualization of land use and land cover using bivariate maps and Sankey flow diagrams

Georgianna Strode,[a] Victor Mesev,[a] Benjamin Thornton,[a] Marjorie Jerez,[a] Thomas Tricarico,[a] and Tyler McAlear[a]

[a] *Florida State University, Tallahassee, FL, USA; gstrode@fsu.edu, vmesev@fsu.edu*

Abstract: The terms 'land use' and 'land cover' typically describe categories that convey information about the landscape. Despite the major difference of land use implying some degree of anthropogenic disturbance, the two terms are commonly used interchangeably, especially when anthropogenic disturbance is ambiguous, say managed for-estland or abandoned agricultural fields. Cartographically, land use and land cover are also sometimes represented interchangeably within common legends, giving with the impression that the landscape is a seamless continuum of land use parcels spatially adjacent to land cover tracts. We believe this is misleading, and feel we need to reiterate the well-established symbiosis of land uses as amalgams of land covers; in other words land covers are subsets of land use. Our paper addresses this spatially complex, and frequently ambiguous relationship, and posits that bivariate cartographic techniques are an ideal vehicle for representing both land use and land cover simultaneously. In more specific terms, we explore the use of nested symbology as ways to represent graphically land use and land cover, where land cover are circles nested with land use squares. We also investigate bivariate legends for representing statistical covariance as a means for visualizing the combinations of land use and cover. Lastly, we apply San-key flow diagrams to further illustrate the complex, multifaceted relationships between land use and land cover. Our work is demonstrated on data representing land use and cover data for the US state of Florida.

Keywords: bivariate maps, land use/land cover, statistical legends

1. Introduction

The land use/ land cover dichotomy is firmly related to anthropogenic disturbance; yet there are very few parts of the Earth where human presence is not evident at some level. Instead, the distinction typically has taken on a more ontological approach, where an emphasis is on the epistemological understanding of the types of data that are collected and organized (Comber et al, 2005; Fisher & Comber 2005). To this end, land cover is defined less as a term to identify 'natural' landscapes and more on the products related to the geology, ecology and hydrology of the landscape. Similarly, land use is not overly simplified as 'artificial' and instead is complex ramifications resulting from the interplay of economic, socio-political and cultural activities. Nevertheless, and semantics aside, the land use/ land cover debate is commonly operationalized within GIScience as both an anthropogenic distinction, and by the unilateral assumption that land cover categories can be spatial subsets of land use categories; for instance, the land cover categories of grass, trees, and water can be spatial subsets of a larger land use category of park (see Anderson et al 1976 for full description of land use/ land cover categories at various spatial scales).

Our work is a reaction to the frequent blurring of semantic lines that distinguish land cover categories from land use when operationalized by cartographic software. Instead of maps that represent both adjacently, we seek to construct bivariate maps that represent both simultaneously. The technique employs nested symbology where circles representing land cover fit within larger squares representing land use. These circle/square combinations are de-signed to illustrate the duality of their mixed occurrence, emphasizing the complex interplay of land use and land cover at the same geographic locations. We find this type of bivariate symbology to be most efficiently plotted on gridded raster frames. In addition, we use statistical legends to further represent this close interplay of land use and cover by plotting their percentage co-occurrences. This facilitates an immediate visual and quantitative identification of the exceptionally high and the exceptionally low co-occurrences of land use/ land cover co-occurrences. These levels of land use/ land cover interactions are further illustrated with Sankey flow diagrams. Using D3 open source software and modifying line widths, color changes, and adding intermediary (or secondary) types we demonstrate the many permutations that link land use categories with land cover. In all, our goal is to draw attention to the many complex and symbiotic relationships of land use and land cover categories using innovative cartographic techniques. We argue that co-occurrence nested symbology demonstrates this complex interplay.

2. Bivariate Mapping

The chief advantage of using bivariate maps is to display two spatial phenomena within one map, and thereby eliminate the inefficient and at times annoying imposition

of continually glancing at univariate maps side-by-side when judging spatial contrasts, comparisons, changes, and relationships (Carstensen 1986). Instead, bivariate maps allow more convenient visual and statistical comparisons between data pairs, especially when searching for causality in socioeconomic applications– elderly populations and locations of healthcare facilities, ethnic minorities and housing stock, or levels of youth populations and crime spots. The key is to construct bivariate maps that not only emphasize visual contrasts between variables but also preserve the data gradients of each individual dataset, and where variation in one does not overly impair the ability to read the other (Ware 2009). Bivariate mapping is also a logical choice for time series maps, especially during non-linear temporal bivariate variations (Schroeder 2010).
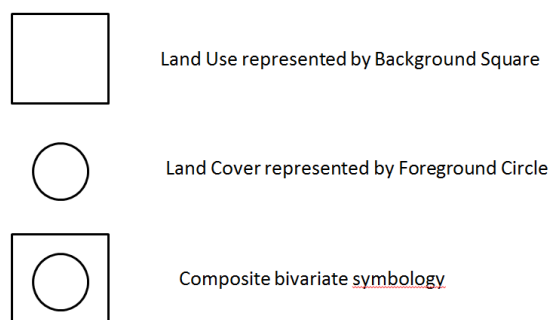

Fig. 1. Point symbology representing land use categories as squares and land cover categories as circles.

Bivariate maps obviously use similar visualization techniques to univariate maps, such as sequential color schemes to represent increases or decreases in data values (Brewer 1994). However, bivariate maps are more conducive to experimental symbology design; manipulating size, color, texture, and arrangement of graphic points, lines and areas (Strode et al 2016; Slocum et al 2005). Our research explores possibilities in designing symbology that represents bivariate data at the same location; in other words overlapping graphic symbology. The process can be related to Gestalt; a psychological term signifying unified whole, and where visualization principles describe how to organize spatial elements into groups (Arnheim 1954; Bradley 2014). We apply the Gestalt principle 'figure/ground,' where we use one dataset as the foreground symbol and one as the background symbol (with the option of toggling between the two). This is a nested process where one symbol is completely within another. Fig. 1 illustrates our chose of circle symbols to represent land cover categories, completely surrounded by square symbols representing land use categories (this order satisfies the assumption of land cover being subsets of land use).

Operationalizing nested symbology requires software that can manipulate large data sets from multiple sources quickly and efficiently. Visual analytics are a set of software tools programmed to transform data into graphic representations for the purposes of highlighting trends, groups, uniformity and outliers in datasets. They are also used to design to customize statistical map legends that quantitatively organize and summarize data interactions (Bostock 2015).

3. Data & Methodology

Our goal is to construct bivariate maps that represent combinations of land use and land cover categories in a manner that seems more seamless and intuitive. We chose the Gestalt principle that allows us to implement nested symbology using the foreground/background contrast. For plotting the circles/squares we chose a gridded frame, simply because regular grids are systematic, can be normalized, preserve data and visually simpler. We also selected colors to represent land use and land cover categories that are consistent with institutionalized mapping agencies so that comparisons are possible with conventional, non-nested symbology maps.
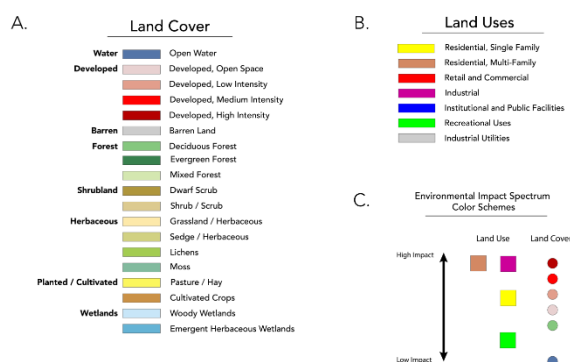

Fig. 2. Commonly used color schemes: A. Land cover scheme. B. Land use scheme. C. Land use and land cover categories along high impact (high disturbance) and low impact (low disturbance) color continuum.

The data we used to demonstrate bivariate mapping of land use and land cover are derived from national and state mapping collections. The land cover data are from the 2010 US National Land Cover Database, using the categories open water, wetlands, forest, vegetation, barren, and development, in four intensities of open, low, medium, and high. The only amendment we made was to combine the categories of shrub, planted/cultivated, and herbaceous into one category named vegetation in order to simplify visualization. The land use data are from the 2014 State of Florida Department of Revenue Cadastral Database, and the categories we selected are the same as the ones used by the property appraiser: residential, commercial, industrial, agricultural, institutional, government, and miscellaneous Fig. 2 (A) and Fig. 2 (B) shows both land cover and land use categories. These types of land use data were preferred over data from more traditional zoning maps because cadastral data have a higher spatial resolution which allows land use codes to be assigned directly by the local property appraiser. Moreover, cadastral data include sufficient information to support land information across jurisdictional boundaries (von Meyer et al, 2002). In contrast, zoning maps are a more generalized and do not ensure that all properties within a planning zone are of the same land use type. For instance, it is possible for a low resolution zoning map to contain properties not within the planning category. Once

extracted, both land use and land cover data sets are normalized to a 1-km grid using the predominant value. From a data performance perspective, gridded maps preserve data density and are less susceptible to the modifiable areal unit problem (Mennis 2003).

In terms of color coordination we employed shades that are commonly used by the US and State of Florida mapping collections, and indeed are also consistent with the US Geological Survey (Jeer 1997; USGS 2006). Fig. 2 (C) shows both land use (squares) and land cover (circles) colors in a sequence relative to a continuum that de-scribes the level of impact land use and land cover have on the landscape. We label this continuum low to high impact, and is akin to levels of anthropogenic disturbance in the case of land use (low disturbance is recreational shown in green, most disturbance is industrial shown in pink), and development when measuring land cover (least developed is water shown in blue and most developed is shown in pink). The complete legend is illustrated by Fig. 4 where colored land use squares and colored land cover circles overlap in 63 combinations. The goal is to create a legend palate that produces a robust array of permutations that mirror the complex interplay of land use and land cover in reality.
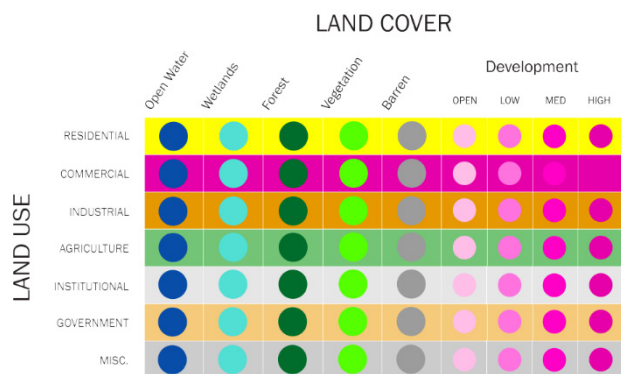

Fig. 4. Legend showing 63 combinations of land use and land cover categories.
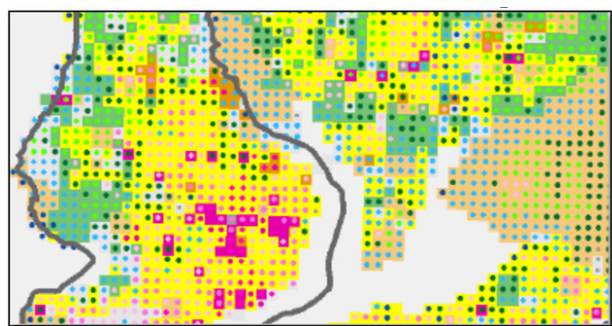

Fig. 5. Bivariate map using color and symbology.

Using the legend from Fig. 4, and applying a gridded frame we can generate a bivariate map of land use and land cover data from the two mapping agencies shown as Fig. 5. Residential areas are identifiable by a yellow background and can have many types of land cover; for instance, yellow backgrounds with green circles are forested residential. Areas with fuchsia background indicate high impact or development. Medium pink

circles represent medium development. Pale pink circles show residential areas with more open space. A green background indicates the land is used for agriculture; and within this there are several types of land cover, forest, vegetation, and wetland all shown as circles of dark green, medium green, and pale blue respectively. Governmental lands are represented by a tan background with land cover of open water (blue), vegetation (medium green), and wetland (pale blue).
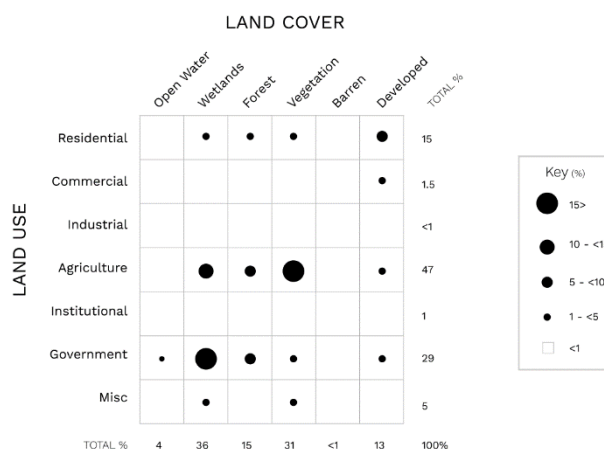

Fig. 6. Statistical legend quantifies the data distributions.

We can further develop our legend from Fig. 4 by incorporating quantitative information. Fig. 6 illustrates a leg-end incorporating the statistical covariance of the data from land use and land cover. Black circles are graduated in proportion to the percentage of land use/ land cover co-occurrence. For example, a very high percentage of land (>15%) is used as agriculture and covered in vegetation, a slightly lower percentage (5% to less than 10%) is agricultural wetlands, and governmental use with wetlands, while no institutional land use has barren land cover.
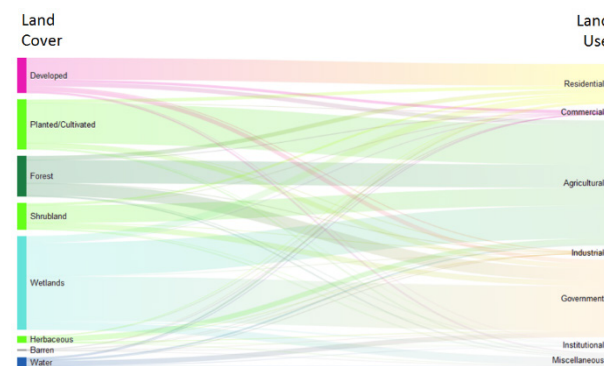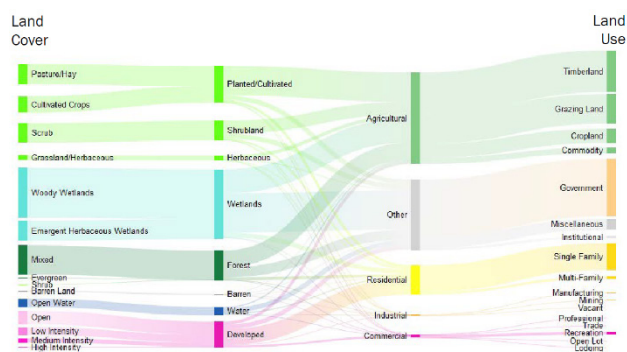

Fig. 7. Primary Sankey diagram.

Fig. 8. Secondary Sankey diagram.

Our final cartographic technique for illustrating the complex interaction of land use and land cover categories is with the use of Sankey diagrams. These are flow diagrams where the width of the lines is proportional to the quantity of flow. They have been used in many applications, including hydrology, trade, and even the energy efficiency of steam engines. Here we used open software called D3 to generate Sankey diagrams that demonstrate the complex interaction of land use and land cover; specifically how land cover types can be part of more than one land use category. We edited the D3 software code to allow the colors of the lines to be grade smoothly from land cover to use. Like all Sankey diagrams, wider lines represent more data combinations; thinner lines represent fewer inter-actions. For consistency we used the same colors as the ones we used to create the bivariate map. Figs 7 and 8 illustrate primary and secondary Sankey diagrams of the complex many-to-many relationships between land use and land cover for the data representing the State of Florida. Primary diagrams illustrate the relationships between broad categories of land use and land cover; for instance, forest land cover forms predominantly agricultural land use, but is also a component of government land use (e.g. national parks), and plays a minor role in residential and commercial. Adding two more levels of land use and land cover categories produces a secondary Sankey diagram (Fig. 8). Forest is now drawn from mixed land cover and is again predominantly a component of agricultural land use, and as a component of residential is also part of single family and multi-family residential.

4. Conclusions

We have demonstrated how cartographic techniques employing bivariate maps, statistical legends and Sankey diagrams can illustrate the complex interplay and overlap of land use and land cover. The two are common terms describing the landscape and unfortunately have been used interchangeably to mean the same landscape. The distinction of the presence of anthropomorphic disturbance should be implicitly upheld in land use, so too is the assumption that land covers are subsets of land use. Our use of nested symbology graphically illustrates how the two can co-occur at the same geographic location; this is a valuable contribution to understanding that land cover is land use. For the first time we can describe land use and land cover as parts of each other; For instance, agricultural vegetation implies anthropogenic disturbance of the land for agrarian purposes but that vegetation is a major land cover component. Similarly, the residential land use is primarily developed land cover but also includes small components of wetland, forest and vegetation land cover. Statistical legends add quantifiable information that identifies the co-occurrence percentages of land use and land cover. This type of information summarizes landscapes in terms of the type and level of mixing; uniform co-occurrences would indicate more 'natural' landscapes (i.e. less anthropogenic disturbance). The level of land use/ land cover mixing is also illustrated by Sankey diagrams that visually demonstrate how land cover types are components of land use and in what proportion.

Looking forward we intend on continuing this line of research by exploring visual analytics to further enhance the graphic appeal of bivariate maps. One improvement would be to automate the nested symbology where users could toggle between land use and land cover in search of spatial patterns in the data sets.

5. References

Anderson, J.R., Hardy, E.E., Roach, J.T., Witmer, R.E. (1976). A Land Use and Land Cover Classification system for use with Remote Sensor Data (Geological Survey Circular 671). US Department of the Interior and US Geological Survey.

Arnheim, R. (1954). Art and Visual Perception: A Psychology of the Creative Eye. Cengage Learning.

Bostock, M. (2015). D3.js - Data-Driven Documents. D3.js - Data-Driven Documents. N.p., n.d. Web.

Bradley, S. (2014). Design principles: visual perception and the principles of Gestalt. Retrieved February 24, 2017, from https://www.smashingmagazine.com/2014/03/design-principles-visual-perception-and-the-principles-of-gestalt/

Brewer, C.A. (1994). Color use guidelines for mapping and visualization. In A. MacEachren, Taylor, D.R. (Ed.) Visualization in Mod-ern Cartography. Pergamon.

Carstensen, L. W. (1986). Bivariate choropleth mapping: The effects of axis scaling. The American Cartographer 13(1): 27-42.

Comber, A., Fisher, P., Wadsworth, R. (2005). You know what land cover is but does anyone else?....an investigation into semantic and ontological confusion. International Journal of Remote Sensing, 26(1), 223–228.

Fisher, P.F., Comber, A.J., Wadsworth, R. (2005). Land use and land cover: contradiction or complement. In P. Fisher (Ed.), Re-Presenting GIS (pp. 85–98). Wiley: Hoboken, New Jersey.

Jeer, S., Bain, B. (1997). Traditional Color Coding for Land uses. Report–American Planning Association.

Mennis, J. (2003). Generating surface models of population using dasymetric mapping. The Professional Geographer, 55(1), 31–42.

Schroeder, J.P. (2010). Bicomponent trend maps: a multivariate approach to visualizing geographic time series. Cartography and Ge-ographic Information Science, 37(3), 169–187.

Slocum, T., McMaster, R., Kessler, F., Howard, H. (2005). Thematic Cartography and Geographic Visualization (2nd Ed.). Upper Saddle River, New Jersey: Pearson.

Strode, G., Thornton, B., Mesev, V. Johnson, N. (2016). Bivariate mapping: Examples on Florida covariance data using scatterplots from open source visual analytic tools. The Florida Geographer 47, 1–17.

United States Geological Survey (2006). Product Legend. Multi-Resolution Land Characteristics Consortium.

Von Meyer, N., Ader, B., Nagy, Z., Stage, D., Ferguson, B., Benson, K., Johnson, B., Kirkpatrick, S., Stevens, R., Mates, D. (2002). Parcel Identifiers for Cadastral Core Data: Concepts and Issues. Federal Geospatial Data Committee Cadastral Committee.

Ware, C. (2009). Quantitative texton sequences for legible bivariate maps. ICCC Transactions on Visualization and Computer

Graphics, 15(6), 1523–1529.