Stat

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

Visualizing uncertainty in areal data with bivariate choropleth maps, map pixelation and glyph rotation†

Lydia R. Lucchesi and Christopher K. Wikle*

Received 15 May 2017; Accepted 24 May 2017

In statistics, we quantify uncertainty to help determine the accuracy of estimates, yet this crucial piece of information is rarely included on maps visualizing areal data estimates. We develop and present three approaches to include uncertainty on maps: (1) the bivariate choropleth map repurposed to visualize uncertainty; (2) the pixelation of counties to include values within an estimate's margin of error; and (3) the rotation of a glyph, located at a county's centroid, to represent an estimate's uncertainty. The second method is presented as both a static map and visuanimation. We use American Community Survey estimates and their corresponding margins of error to demonstrate the methods and highlight the importance of visualizing uncertainty in areal data. An extensive online supplement provides the R code necessary to produce the maps presented in this article as well as alternative versions of them. Copyright © 2017 John Wiley & Sons, Ltd.

Keywords: sample surveys; spatial statistics; statistical graphics; visualization

1 Introduction

In statistics, we quantify uncertainty to help determine the accuracy of estimates. Yet this crucial piece of information is rarely included on a map of areal data, and the best methods for displaying it are still not clearly defined (MacEachren et al., 2005). On a choropleth map, latitude and longitude require two dimensions, and colours symbolizing the estimates fill the defined spaces. Although it is difficult to effectively add more information to such a visual presentation, there is increasing interest in creating maps that include uncertainty and thus provide a more accurate depiction of the statistics (Bonneau et al., 2014). In this article, we present three approaches to visualizing uncertainty: (1) the bivariate choropleth map repurposed to visualize uncertainty; (2) the pixelation of counties to include values within an estimate's margin of error; and (3) the rotation of a glyph, located at a county's centroid, to represent an estimate's uncertainty. In addition to static maps, we present a visuanimation, a concept developed by Genton et al. (2015). It is produced by animating the second method and embedding the dynamic map within this paper. We have also produced an extensive vignette to demonstrate the R commands necessary to produce each of the visualizations. This is included in the online supplement.

Department of Statistics, University of Missouri, Columbia, MO 65211, USA

***Email: wiklec@missouri.edu**

†Requires Adobe Reader

The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

1.1 The power of visuals

Few (2009) explains the limitations of our visual working memory to highlight the necessity and power of statistical graphics. He explains that, unless we commit information to our long-term memory, we can only consider about three pieces of information at a time (e.g. three numbers). However, if the individual pieces of information are condensed into a meaningful visual, we can comprehend complex ideas and information within our visual working memory. For example, on a line graph, the trend line representing possibly hundred of values constitutes as only one of the three pieces of information (Few, 2009). Similarly, Tufte (1986) discusses the power of US county choropleth maps. They display thousands of numbers simultaneously, making it possible to detect what is happening locally, regionally and nationally. From Few (2009) and Tufte (1986), we see that, instead of working to remember more than three numbers in our visual working memory, we can use maps to comprehend the main trends among thousands of numbers.

In uncertainty visualization, we look to compare the trends among both estimates and errors. For example, consider visualizing survey statistics from the US Census Bureau American Community Survey (ACS). Every year, a 5-year multi-year estimate of poverty rates is released for all counties, and each estimate includes a margin of error that can be used to calculate its 90% confidence interval. The simple comparison of two poverty estimates and their margins of error highlights the importance of considering uncertainty when displaying estimates on a map. In 2015, Kenedy County, TX, USA, had an estimated poverty rate among families of 20.0% with a confidence interval of (7.5%, 32.5%). Los Angeles County, CA, USA, had an estimated poverty rate among families of 14.3% with an interval of (14.2%, 14.4%). These two estimates have vastly different uncertainties because of their different population sizes. Thus, in a more populous area like Los Angeles County, a larger survey sample size is obtained, reducing the margin of error. We can be fairly confident that around 14% of Los Angeles families really do live below the poverty line. Conversely, we cannot be as sure about the proportion of families living in poverty in Kenedy County, which has a population of less than 500 people. The county's estimate has a large margin of error, the result of a smaller survey sample size. Although a table of values could provide us access to these estimates and errors, with over 3000 counties, it could not provide us with an idea of the overall trends that exist across the country. This requires a map that displays the estimates and errors simultaneously.

1.2 Uncertainty visualization in spatial statistics

There has been a consistent concern throughout the history of uncertainty visualization that adding the uncertainty measure to a map will clutter the visualization or make the variable of interest difficult to interpret (MacEachren et al., 2005). Recently, uncertainty visualization in spatial analysis has been gaining attention from statisticians, and researchers are summarizing past and present visualization methods to help us gauge what has worked, what has not worked and what we should focus on in the future (e.g. Genton et al., 2014). One summary by MacEachren et al. (2005) stands out as especially comprehensive for areal spatial data, and although the article is 12 years old, most of the methods they mention are still the common approaches used today.

MacEachren et al. (2005) discuss a wide variety of methods. Several of them involve texture and fog encoded to represent uncertainty. For example, closely spaced parallel lines or dense fog layered on top of a choropleth or heat map might represent areas of high uncertainty. Other options involve changing the appearance of colours or resolution; areas of high uncertainty could be less saturated in colour or blurry. They mention displaying several side-by-side maps that depict different predictions, known as multiple realizations, where the differences between the maps suggest the uncertainty. They also describe uses of animation. For example, in Fisher (1993), the longer a value appeared on the map, the more certain it was (MacEachren et al., 2005).

Bivariate schemes, map pixelation and glyphs are also discussed in this overview by MacEachren et al. (2005), and the methods we present in this article are motivated by the ones they mention but have novel components as discussed in Section 2.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

The ISI's Journal for the Rapid
Dissemination of Statistics Research

2 Methods

This section outlines the three approaches to visualizing uncertainty on maps developed here. Before explaining these approaches, we describe historical uses of bivariate colour schemes, pixelation and glyphs to highlight how we have repurposed, altered or extended these techniques for uncertainty visualization. The maps throughout this section display estimated poverty rates among families and their margins of error and serve as examples for the three methods. We use ACS 5-year estimate data; however, these uncertainty visualization methods are applicable to many different types of areal spatial data as well as output from spatial statistical models with predictive error estimates.

2.1 The bivariate choropleth map for uncertainty visualization

In the 1970s, the US Census Bureau (USCB) produced some of the first known bivariate choropleth maps, which were used to visualize two variables such as death rate and population density. Examples of USCB bivariate choropleth maps can be found in Fienberg (1979, see Figure K in that article) and Olson (1981, see Figures 1, 4 and 5 in that article). Although the innovation seemed to be appreciated by some, the maps were not successful. Fienberg (1979) argued that interpreting them was difficult because of the 4×4 colour key grid that covered a wide range of colours: red, orange, yellow, green, blue and purple. User studies conducted by Wainer & Francolini (1980) proved Fienberg's claim. They asked participants to identify what was happening at certain locations on either a USCB bivariate map or two univariate maps and concluded that the error rates for the bivariate maps were higher, especially when the colour keys were removed from the maps after several trials. Remembering the arrangement of the 4×4 grid required memorization because a glance was not enough to internalize it (Wainer & Francolini, 1980). This finding was significant because colour keys should not require extensive time and effort to understand.

Shortly after the introduction of this method and the initial criticisms and experiments that followed, Trumbo (1981) emphasized that exploring colour as a way to visualize two variables on a map should not be abandoned completely. Furthermore, MacEachren (1992) commented that not all applications of a bivariate choropleth map have been explored and explicitly mentioned uncertainty visualization as one of these unexplored applications. Additional literature suggested potential improvements to these maps, including limiting the number of colour bins to nine (Eyton, 1984), adding a detailed description to aid viewers (Olson, 1981) and choosing more interpretable colours (Brewer, 1994).

Importantly, although the bivariate choropleth map is generally used to display the relationship between two variables, we use it to visualize one variable and its uncertainty. Researchers have experimented with bivariate colour schemes for uncertainty visualization; however, the approaches are different than the one developed here (Figure 1). For example, the bivariate technique described by MacEachren et al. (2005), which was developed by Howard & MacEachren (1996), is based on using a diverging colour scheme to represent uncertainty and the darkness/lightness of a colour to represent the variable of interest. Retchless & Brewer (2016), using a 4×5 grid, represented the variable of interest with several colours and the uncertainty by changing the saturation or value of these colours or by adding different amounts of white or gray to them. It is important to note that these two bivariate colour schemes were not used for choropleth maps. The bivariate choropleth map developed here is novel in that it visualizes uncertainty and combines the three refinements noted in the previous paragraph, as well as a grid rotation, to provide the information in a simple, clear format to improve interpretability.

Instead of using a large colour key grid with many colours, we limit the colour key grid to 3×3 and two single hue colour palettes, which we naturally blend by averaging the three components of the RGB colour codes. We also rotate the grid 45 degrees. This rotation has been applied to a bivariate colour scheme in a news article (Keating & Karklis, 2016), and we apply it to the grid so that the colour bin representing the highest values in both categories is at the

The ISI's Journal for the Rapid Dissemination of Statistics Research
(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

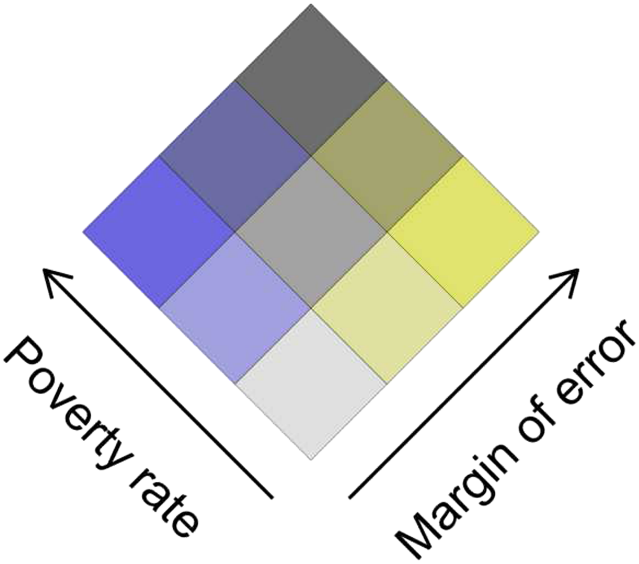


Figure 1. The 3×3 colour grid encodes estimated poverty rates and margins of error. Each square is an average of blue, representing poverty rate, and yellow, representing margin of error. See Figure 2 for an example of a colour grid that includes the numerical bounds for each colour bin.

top of the grid. Ultimately, our colour grid is intended to be more manageable and intuitive than the grids from the 1970s. In the US county maps that demonstrate our approach, we focus on estimated county poverty rates and their margins of error, and we average the values of blue, which represent poverty rate, and yellow, which represent margin of error, to produce the 3×3 colour key grid in Figure 1.

Each county is assigned one of the nine colours depending on its estimated rate and error. For example, counties with low poverty rates and small margins of error are represented by the light gray at the bottom of the grid (Figure 1), which is an average of the lightest blue and the lightest yellow. Conversely, the dark gray at the top of the grid is an average of the darkest blue and the darkest yellow, and it represents counties with high poverty rates and large margins of error. Figure 2 shows many clusters of colour on the US county map, highlighting regional trends in poverty and uncertainty.

A vignette that provides and describes the R code needed to produce several versions of the bivariate choropleth map is located in Supporting Information Section A. In particular, Supporting Information Section A.1 of the vignette includes the code necessary to make a tercile map (Figure 2), and Supporting Information Section A.2 demonstrates how to plot each of the nine colours from the tercile map individually. Supporting Information Section A.3 demonstrates how to make a tercile map that highlights significant changes in the estimates over time, and Supporting Information Section A.4 describes how to exchange terciles for equal intervals. Generally, though, we use terciles because with equal intervals an outlier might force the majority of the data into one or two of the colour bins, leading to a lack of different colours on the map. Throughout the vignette, we focus on poverty rates and margins of error, but maps with different variables could be easily produced.

2.2 Map pixelation for uncertainty visualization

MacEachren et al. (2005) discuss map pixelation for uncertainty visualization. They describe an approach by Evans (1997), in which maps that only included the most certain pixels, that used colour saturation to encode a pixel's

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

The ISI's Journal for the Rapid
Dissemination of Statistics Research

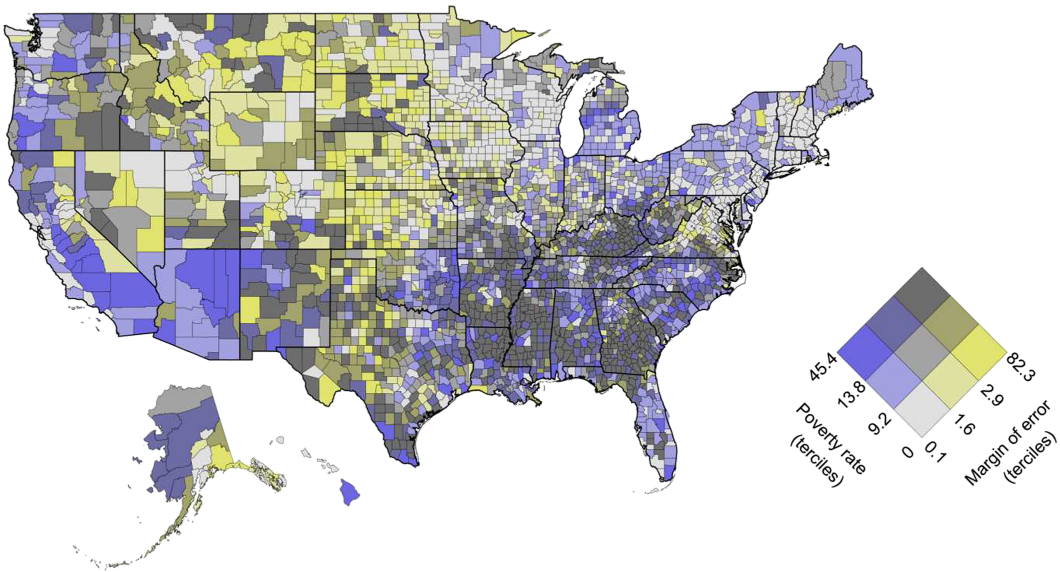


Figure 2. This US poverty map displays the percentage of families whose income was below the poverty level in 2015, and it includes uncertainty in its representation. Terciles divide the estimates and their margins of error into three categories. Each square in the 3×3 colour grid is an average of blue, representing poverty rate, and yellow, representing margin of error, and in this case, all four corners of the colour grid have a strong presence on the map. The South is dark gray, meaning that it has high poverty rates and large margins of error, while the Southwest is bright blue, meaning that it has high poverty rates and small margins of error. The Midwest is bright yellow because of its low poverty rates and large margins of error, while the upper East Coast is light gray because of its low poverty rates and small margins of error.

uncertainty, or that flickered between a visual with all pixels and one with an overlay of the most certain pixels encoded by full colour saturation were created. Our approach is similar in that we use pixelation and flickering. We also focus on the use of texture, one of the visual variables developed in Bertin (1983), and on the concept that maps can be designed so that it is difficult to draw conclusions about the variable of interest in areas of high uncertainty (Retchless & Brewer, 2016). In the method we present, texture represents areas of uncertainty, and including values within an estimate's margin of error prevents map users from viewing the estimates as all equally certain and as the only values the variable of interest could be. Fisher (1993) developed this idea of including other values for the variable of interest in his animated soil maps; he animated the maps to display other soils that could be present in areas classified as only one soil type. Our approach to map pixelation for uncertainty visualization includes both static and dynamic maps.

In this method, we pixelate the counties and randomly assign each pixel in a county a value within the estimate's margin of error. Counties with high uncertainty appear pixelated because the margin of error spans a wide range of values and therefore colours. Conversely, counties with low uncertainty appear smooth because the margin of error spans a small range of values, so the colours filling the pixels do not differ greatly. The lightest colour in a county represents the county's estimate minus the margin of error, and the darkest colour represents the county's estimate plus the margin of error. The colours in-between represent values within the 90% confidence interval. The caption for Figure 3 describes how the colour range is defined for Cedar County, Missouri, to demonstrate the method.

It is important for the individual pixels to be small enough so that colour differences within a county are not attributed to different locations in the county. For example, if a large pixel covering one-fifth of the county is a darker green than the pixel below it, map users would most likely interpret that location as having a higher rate of poverty among families; however, the estimate is for the entire county. If the pixels are small enough, this misinterpretation should be avoided.

The ISI's Journal for the Rapid Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

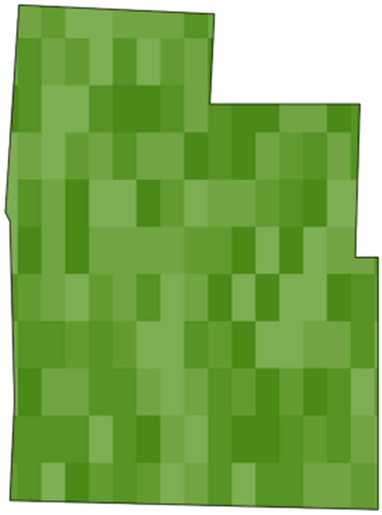


Figure 3. This polygon corresponds to Cedar County, MO. In 2015, it had an estimated poverty rate among families of 21.4% and a margin of error of 4.1%. The lightest green represents the value 17.3% ($21.4 - 4.1$), and the darkest green represents the value 25.5% ($21.4 + 4.1$). The green colours in-between represent values within the interval (17.3%, 25.5%).

In Figure 4, which is the full Missouri map, the counties containing major Missouri cities appear less pixelated than their surrounding counties because they have some of the smallest margins of error in the state, which are tied closely to a county's population. We also explore the concept of visuanimation developed in Genton et al. (2015), which discusses the effectiveness of animated visuals for multidimensional data and encourages researchers to embed visuals within an article to provide readers with immediate access to dynamic data exploration. We have embedded a visuanimation to demonstrate how the pixelated map can be animated to visualize uncertainty (Movie 1). The pixels flicker between randomly assigned values within an estimate's margin of error. Areas of high uncertainty have visible movement among the pixels, and areas of low uncertainty do not.

Supporting Information Section B of the online supplement is a vignette that includes and describes the R code necessary to produce both static and dynamic maps. Supporting Information Section B.1 of the vignette produces a static map, and Supporting Information Section B.2 focuses on animating the pixels to flicker.

2.3 Glyph rotation for uncertainty visualization

One long-standing critique of county choropleth maps is the misleading emphasis they place on a county's size relative to other counties. In Section 1.1, we mentioned Tufte's appreciation of county choropleth maps because of their ability to visualize such a large quantity of numerical information. However, he has also been a critic of the choropleth map because of its misleading emphasis on size; for example, one issue is that the highly populated counties are generally the smallest geographically and are therefore easily overlooked (Tufte, 1986). We created a visual that includes spatial information for the estimates but eliminates the element of geographic area. In the following method, we use rotated glyphs, which are located at the counties' centroids, to allow map users to focus on the estimates, their uncertainties and spatial trends. Ultimately, it removes the element of geographic size to prevent the perception that larger counties have a greater significance. All counties are represented by a glyph of the same size, and the rotation of each glyph represents the degree of uncertainty.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

The ISI's Journal for the Rapid
Dissemination of Statistics Research

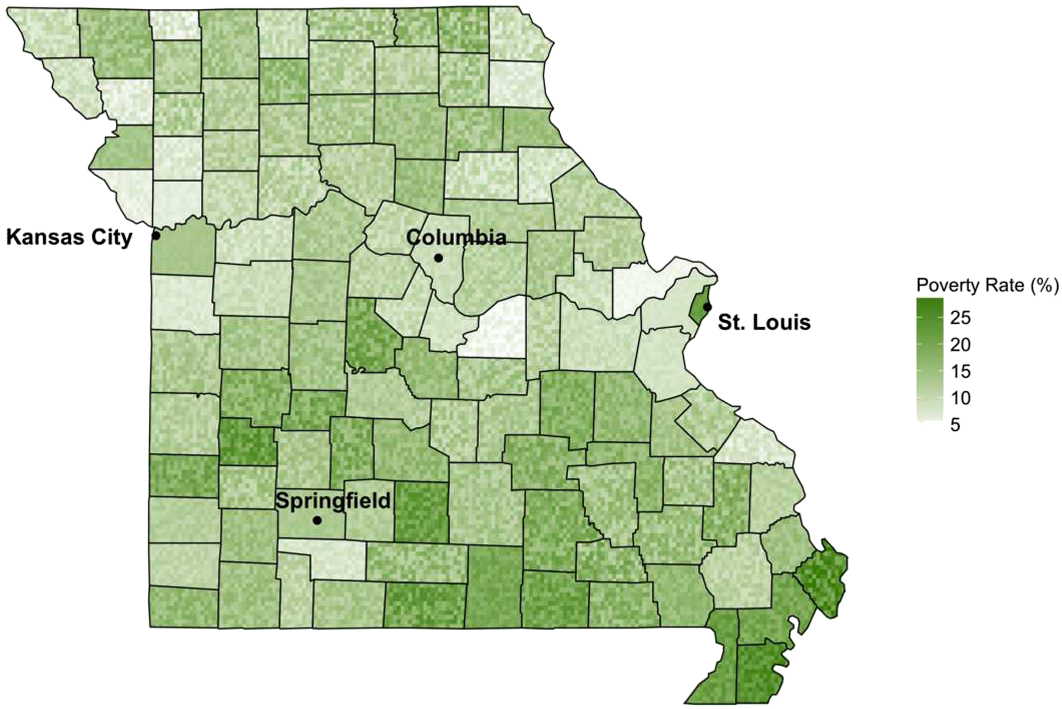


Figure 4. This Missouri poverty map displays the percentage of families whose income was below the poverty level in 2015, and it includes uncertainty in its representation. Each pixel in a county is assigned a colour within the estimate's margin of error. Areas of high uncertainty appear pixelated because the margin of error spans a wide range of colours. Areas of low uncertainty appear smoother because the differences in colour are minimal. For example, counties containing major Missouri cities, which have some of the smallest margins of error in the state, appear less pixelated than their surrounding counties.

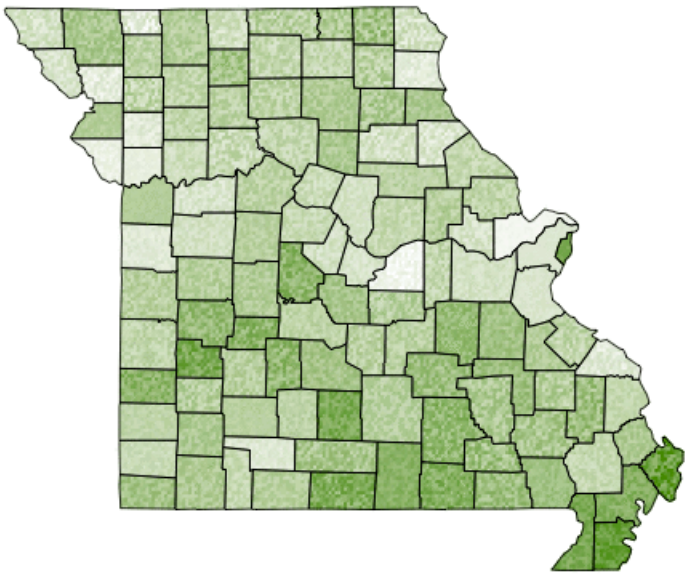
Glyphs have been used to represent uncertainty before. For example, MacEachren et al. (2005) showed that colour saturation, edge definition and transparency of a circular glyph can represent uncertainty, and Bonneau et al. (2014) describe altering the shapes of glyphs or changing the radius of a conical glyph based on its uncertainty value. Another use of glyphs by Wittenbrink et al. (1996) explored visualizing uncertainty in environmental data with arrows. For example, the direction of an arrow indicates the predicted direction of an ocean current, and the width of the arrowhead represents the uncertainty of the prediction. However, because the ACS estimates do not involve a directional aspect, we can use the rotation of the glyph to represent uncertainty and not the variable of interest.

The glyph in Figure 5 was chosen for several reasons. It allows map users to see differences in rotation, unlike a circular glyph. It has substantial surface area, making the colours that fill the glyphs easily identifiable and comparable. Also, when upright, the glyph naturally has a positive implication, such as low uncertainty, and when tipped upside down, it has a negative implication, such as high uncertainty (e.g. imagine an ice cream cone that is turned over). On the map in Figure 6, areas of extreme concordance and discordance in glyph rotation are noticeable. For example, the glyphs tipped horizontally in the middle of the map stand out as a cluster of similar uncertainties. As for discordance, in the southeast corner of the state, sharp contrasts in rotation highlight an area that has important differences in uncertainty among neighbouring counties.

The R code for this map is provided in Supporting Information Section C. It demonstrates how to create the glyph, place it at a county centroid, rotate it accordingly using the rotation matrix and fill it with the proper colour.

Stat
The ISI's Journal for the Rapid
Dissemination of Statistics Research

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150



Movie 1. In this visuanimation, the pixels flicker between randomly assigned values within an estimate's margin of error. Areas of high uncertainty have visible movement, while areas of low uncertainty do not. For example, there is not visible movement around St. Louis (Figure 4) because the counties have small margins of error. (To ensure that the visuanimation will play properly, use Adobe Acrobat Reader to view the pdf file.)

Figure 5. The colour and rotation of this glyph are used to represent an estimate and its uncertainty.

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

The ISI's Journal for the Rapid Dissemination of Statistics Research

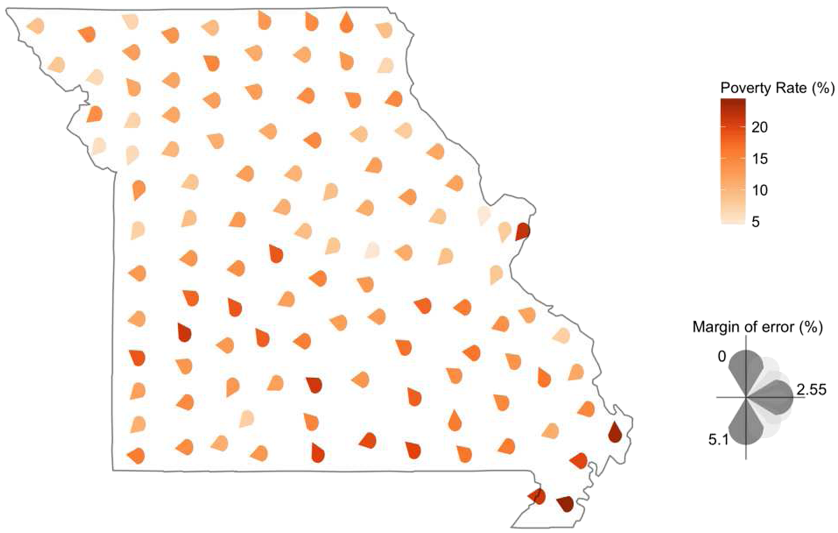


Figure 6. This Missouri map displays the percentage of families whose income was below the poverty level in 2015, and it includes uncertainty in its representation. The colour of each glyph represents the estimated poverty rate among families, and the rotation of it represents the estimate's margin of error.

3 Discussion

Although the bivariate choropleth map has been a controversial method since the 1970s, we utilized refinements suggested by previous research and repurposed it to include uncertainty. Our method mathematically combines two single hue colour palettes to obtain one new palette; limits the colour scheme to a 3×3 grid, which is rotated 45 degrees; and offers the option to add a temporal element to the map by bolding county borders. It allows map users to detect trends in the variable of interest and its uncertainty, and users can track numerical information for both of these measures.

In the past, map pixelation has also been used to visualize uncertainty. We present a novel approach to this by pixelating counties and randomly assigning these pixels values within an estimate's margin of error. It leads to counties with either visible or invisible pixelation, depending on the uncertainty. Although users can see which counties have high uncertainties, they cannot determine the exact quantities of the margins of error by looking at the pixelated map. In regard to this aspect, map pixelation falls short compared with the bivariate choropleth map and the glyph rotation method. Conversely, the pixelated map is easily animated, which can emphasize differences in uncertainty. Visuanimation is one way to overcome the difficulty of adding additional elements to a map, and it is worth exploring further in uncertainty visualization (Genton et al., 2015).

In the exploration of glyph rotation for uncertainty visualization, we remove the misleading elements of size from a county choropleth map to provide an unbiased visualization of the ACS estimates, their uncertainties and the spatial relationships among them. An interesting extension of this technique would be to display the glyphs on an equal area cartogram (tile grid map), which would further emphasize the unimportance of geographic area in this type of visual.

Conducting user studies to determine whether these three methods effectively communicate uncertainty is the next step in this data visualization research. For example, Wainer & Francolini (1980), Olson (1981) and Retchless & Brewer (2016) conducted studies in which participants actively used the maps to draw conclusions and answer questions. These experiments led to important insights on how map users interpreted the maps or perceived uncertainty,

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

and before implementing regular use of the methods we present, quantifying their effectiveness in a user study is a critical next step.

In conclusion, maps play an important role in spatial statistics; however, when they fail to include uncertainty, they lack a crucial piece of information that should greatly influence the conclusions we draw about the variable of interest. In the approaches we present, the bivariate choropleth map, map pixelation and glyph rotation, we not only visualize uncertainty but also demonstrate why it is so important to include uncertainty on maps displaying areal data estimates.

Acknowledgements

This work was supported by the US National Science Foundation (NSF) and the US Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program.

References

Bertin, J (1983), *Semiology of Graphics*, *University of Wisconsin Press*, Madison, WI.

Bonneau, G-P, Hege, H-C, Johnson, CR, Oliveira, MM, Potter, K, Rheingans, P & Schultz, T (2014), *Overview and State-of-the-art of Uncertainty Visualization*, Scientific Visualization, *Springer*.

Brewer, CA (1994), 'Color use guidelines for mapping and visualization', *Visualization in Modern Cartography* in MacEachren, AM & Taylor, DRF (eds), Elsevier Science, Tarrytown, NY, pp. 123–147.

Evans, BJ (1997), 'Dynamic display of spatial data-reliability: does it benefit the map user?' *Computers & Geosciences*, **23**(4), 409–422.

Eyton, JR (1984), 'Complementary-color, two-variable maps', *Annals of the Association of American Geographers*, **74**(3), 477–490.

Few, S (2009), *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, *Analytics Press*, Oakland, CA.

Fienberg, SE (1979), 'Graphical methods in statistics', *The American Statistician*, **33**(4), 165–178.

Fisher, PF (1993), 'Visualizing uncertainty in soil maps by animation', *Cartographica: The International Journal for Geographic Information and Geovisualization*, **30**(2-3), 20–27.

Genton, MG, Castruccio, S, Crippa, P, Dutta, S, Huser, R, Sun, Y & Vettori, S (2015), 'Visuanimation in statistics', *Stat*, **4**(1), 81–96.

Genton, MG, Johnson, C, Potter, K, Stenchikov, G & Sun, Y (2014), 'Surface boxplots', *Stat*, **3**(1), 1–11.

Howard, D & MacEachren, AM (1996), 'Interface design for geographic visualization: tools for representing reliability', *Cartography and Geographic Information Systems*, **23**(2), 59–77.

Keating, D & Karklis, L (2016), 'The increasingly diverse United States of America', *Washington Post*.

MacEachren, AM (1992), 'Visualizing uncertain information', *Cartographic Perspectives*, **13**, 10–19.

MacEachren, AM, Robinson, A, Hopper, S, Gardner, S, Murray, R, Gahegan, M & Hetzler, E (2005), 'Visualizing geospatial information uncertainty: what we know and what we need to know', *Cartography and Geographic Information Science*, **32**(3), 139–160.

Stat

(wileyonlinelibrary.com) DOI: 10.1002/sta4.150

The ISI's Journal for the Rapid
Dissemination of Statistics Research

Olson, JM (1981), 'Spectrally encoded two-variable maps', *Annals of the Association of American Geographers*, **71**(2), 259–276.

Retchless, DP & Brewer, CA (2016), 'Guidance for representing uncertainty on global temperature change maps', *International Journal of Climatology*, **36**(3), 1143–1159.

Trumbo, BE (1981), 'A theory for coloring bivariate statistical maps', *The American Statistician*, **35**(4), 220–226.

Tufte, ER (1986), *The visual display of quantitative information*, *Graphics Press*, Cheshire, CT, USA.

Wainer, H & Francolini, CM (1980), 'An empirical inquiry concerning human understanding of two-variable color maps', *The American Statistician*, **34**(2), 81–93.

Wittenbrink, CM, Pang, AT & Lodha, SK (1996), 'Glyphs for visualizing uncertainty in vector fields', *IEEE transactions on Visualization and Computer Graphics*, **2**(3), 266–279.

Supporting Information

Additional supporting information may be found online in the supporting information tab for this article.