Cartogram Visualization for Bivariate Geo-Statistical Data

Sabrina Nusrat, Md. Jawaherul Alam, Carlos Scheidegger, Stephen Kobourov

Abstract—We describe *bivariate cartograms*, a technique specifically designed to allow for the simultaneous comparison of two geo-statistical variables. Traditional cartograms are designed to show only a single statistical variable, but in practice, it is often useful to show two variables (e.g., the total sales for two competing companies) simultaneously. We illustrate bivariate cartograms using Dorling-style cartograms, yet the technique is simple and generalizable to other cartogram types, such as contiguous cartograms, rectangular cartograms, and non-contiguous cartograms. An interactive feature makes it possible to switch between bivariate cartograms, and the traditional (monovariate) cartograms. Bivariate cartograms make it easy to find more geographic patterns and outliers in a pre-attentive way than previous approaches, as shown in Figure 2. They are most effective for showing two variables from the same domain (e.g., population in two different years, sales for two different companies), although they can also be used for variables from different domains (e.g., population and income). We also describe a small-scale evaluation of the proposed techniques that indicates bivariate cartograms are especially effective for finding geo-statistical patterns, trends and outliers.

Index Terms—Geo-visualization, Cartograms, Bivariate maps.

1 Introduction

A cartogram, or a value-by-area map, is a representation of a map in which geographic regions are modified to reflect some geo-referenced statistic, such as population or income. Specifically, geographic regions, such as countries and states, are scaled by area in order to visualize the given statistical data, while attempting to keep the overall result readable and recognizable. Cartograms have a fairly long history with early variants dating back over a century [49]. The main appeal of cartograms is that they combine *statistical and geographical* information in the same visualization. Unlike standard visualizations for statistical data, such as bar charts and pie charts (which are great for displaying quantitative data), cartograms also show geographical data. Thus, by the very design of cartograms, they make it possible to provide a simultaneous overview of both statistics and geography: statistical patterns, trends and outliers can be seen in the sizes of the regions, while geographical patterns, trends and outliers are embedded in the map itself.

The overwhelming majority of cartograms show one variable at a time and there is little work on cartograms that display multiple variables. The term "bivariate cartogram" has been applied before to augmented cartograms, where region areas represent one variable of interest and a second variable is realized by color [19], [62]. Thus one attribute is used to proportionally re-scale the area of each state, and a second attribute is shown as a choropleth thematic map, with colors and color-shades; see Fig. 4. Glyphs and texture patterns on the map have also been used to represent the second variable [64], [65]. Similar thematic maps showing two variables with a combination of different visual encodings such as size, shape, and hue have also been proposed [62]. Such bivariate maps make it difficult to effectively convey the individual distributions and the correlations between them [24]. In all of the approaches above, the viewer has to compare different methods for representing the underlying data – size and color, size and texture – in order to make a comparison across variables. However, magnitude comparison of attributes with different encodings is particularly difficult [54]. With this in mind, we are interested in designing bivariate cartograms that effectively represent two variables and encode the attributes in the same fashion.

In this paper, we propose a simple yet novel approach for designing bivariate cartograms in which both variables are encoded as areas. We use two complementary colors to show the relation between two variables (whether one is smaller or larger than the other). Our main contributions are: (i) a new simple visualization technique to generate bivariate cartograms; (ii) a technique that can be applied to most standard cartogram types; (iii) a visualization with visual properties that can be detected rapidly, making it easy to find outliers in a pre-attentive way; (iv) implementations of the new visualization technique using several standard cartogram types; (v) a small-scale evaluation of the effectiveness of the proposed technique. The evaluation is based on two types of visualization tasks, and compares the proposed bivariate cartogram visualization against two cartogram-based visualization alternatives: side-by-side monovariate cartograms and shaded cartograms. Even though our evaluation is limited (e.g., by the number of different tasks, by the number of alternative techniques), the results are encouraging. Bivariate cartograms were more effective than the cartogram-based visualization alternatives in meta-data extraction tasks, such as finding outliers and summarization.

- *S. Nusrat, C. Scheidegger and S. Kobourov are at the University of Arizona, Tucson, USA.*
E-mail: {sabrinanusrat, cscheid, kobourov}@cs.arizona.edu
- *Md. Jawaherul Alam is working as a Software Development Engineer at Amazon, USA. E-mail: jawaher@amazon.com*

2 Background

2.1 Cartograms

According to Tobler [60] the term "cartogram" dates back to at least 1868 and was used to mean statistical maps, or choropleth maps [31], [52]. In 1934 Raisz gave a formal definition of value-by-area cartograms, although only rectangular cartograms were considered [53]. Cartograms are studied in the visualization literature [30], [35], [39] and in several cartography textbooks [22], [57]; see a recent survey [49]. There is a wide variety of methods to generate cartograms, which can be broadly categorized by type: contiguous, non-contiguous, Dorling, and rectangular. In contiguous cartograms the original geographic map is modified by deforming the boundaries to change areas. Among these cartograms, the most popular method is the diffusion-based method proposed by Gastner and Newman [32]. Others of this type include CartoDraw by Keim et al. [40] and constraint-based continuous cartograms by House and Kocmoud [35]. In circular-arc cartograms by Kämper et al. [38], the straight-line segments of a map are replaced by circular arcs. The curvature of the circular arcs is used to "inflate" regions with less area than required and "deflate" those with more area than required. This provides pre-attentive visual cues about regions that have grown/shrunk. Non-contiguous cartograms are generated by starting with the regions of the given map and scaling down each region independently until the desired areas are obtained [50]. Dorling cartograms [25], [26] schematize regions using circles which have areas proportional to the given statistical data. In order to avoid overlaps, circles might need to be moved away from their original geographic locations. Demers cartograms [9] are a related variant, where squares are used in place of circles. Rectangular cartograms schematize regions using rectangles and date back to 1934 [53]. Unlike the others types above which modify the given geographic map, rectangular cartograms create a contact-of-rectangles representation of the dual graph of the input map [15], [63]. Mosaic cartograms [17] redraw the input geographical map as a tiling of the plane, using simple tiles (e.g., squares or hexagons). A detailed description of the cartogram types for which we have implemented our bivariate method can be found in Section 5.2.

2.2 Cartograms and Perception

The impact of parameters such as area, color, and texture on map visualization and understanding has been studied in visualization and cartography. This is relevant to cartograms as different algorithms generate different types of shapes (circles, rectangles, irregular polygons). Bertin [8] was one of the first to provide systematic guidelines to test visual encodings. Cleveland and McGill [21] extended Bertin's work with human-subjects experiments showing significant accuracy advantage for position judgments over both length and angle judgments, which in turn proved to be better than area judgments. Stevens [58] showed that subjects perceive length with minimal bias, but underestimate differences in area. This finding is further supported by Cleveland et al. [20], and Heer and Bostock [34]. These results were consistent with the findings of "judgment of size" by Teghtsoonian [59].

Dent [23] surveyed work on magnitude estimation, highlighting the tendency of humans to estimate lengths correctly, but underestimate areas and volumes. Perceptual tests led Flannery [29] to use apparent scaling of circles (rather than absolute scaling) to compensate for underestimation. However, others argue for absolute scaling. Tufte demands to tell truth about the data: "The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented" [61]. Krygier [42] suggests that "good legend design could eliminate the perceptual problem." These studies indicate that although there are non-trivial area perception issues, it is possible to deal with them with good design, proper legends, and clear labels.

Pre-attentive processing refers to the intuitive notion that certain visual properties are detected rapidly and accurately by the low-level visual system. For maps and cartograms, pre-attentive tasks include boundary detection and target detection, where the main feature is color. Color pre-attentiveness depends on the *saturation*, and *size of color patch*, as well as the *degree of difference from surrounding colors* [16], [33]. In the now common red-blue US election cartograms (with states colored red or blue, depending on whether republicans or democrats win), one can quickly see patterns such as the overwhelmingly democratic coastal states, and outliers such as inland democratic states. These studies provide background to our work since the concept of pre-attentiveness and area perception are relevant to our study.

3 Related Work

3.1 Bivariate Mapping in Charts and Maps

In cartography, thematic mapping is used to show the variation of statistical attributes across space. The best practices regarding thematic map design address the choice of color schemes [11], [13], [51], the means of assigning data into classes [14], [37], and algorithms for perceptual scaling of proportional symbols [12], [29]. While most thematic maps show a single variable, there is also work on multivariate maps [41] and bivariate maps [27]. Identifying patterns and recognizing spatial relationships among the variables is an important feature of bivariate mapping. However, showing multiple variables on a map often makes the visualization cluttered and hard-to-read, especially when there are multiple symbols, glyphs, and colors [62]. In this section, we consider several approaches for bivariate mapping, along with their strengths and limitations.

Scatterplots: In a traditional scatterplot, the values of the two given variables determine the (x, y)-coordinate of every data point. By examining the plot, one can often spot correlations, clusters, and outliers. A scatterplot, however, cannot show geographic patterns, trends, and outliers and it is not clear how to combine value-by-map visualizations with scatterplots [4]. For example, Fig. 1 (left) shows a scatterplot of the number of McDonald's stores and Starbucks stores in the US. Each point represents a state, and the x and y coordinate values denote the number of McDonald's and Starbucks shops, respectively. Similarly Fig. 1 (right) shows a scatterplot of the number of McDonald's stores and Starbucks stores per 100,000 residents in each state. From

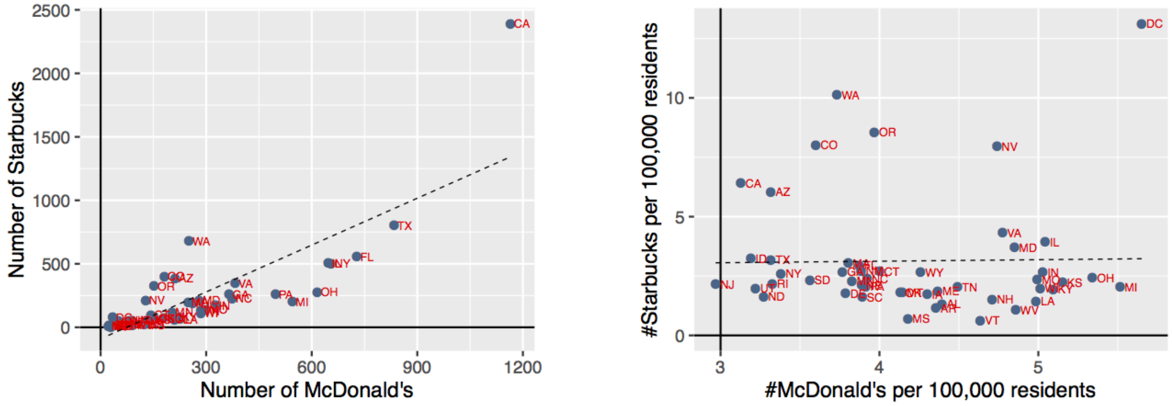
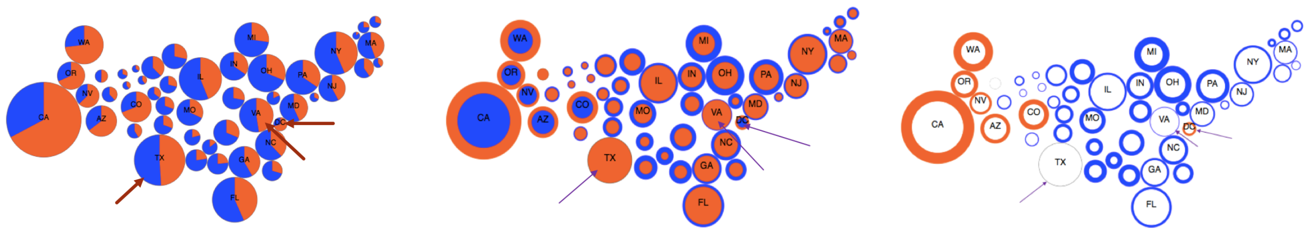


Fig. 1: Scatterplots showing the number of McDonald's and Starbucks shops (left), and number of McDonald's and Starbucks shops per 100,000 residents (right). In both scatterplots, the dotted regression line shows the general pattern. A careful observation shows that the states above the line are mostly from the Pacific/West coast. However, it's difficult to identify geographic patterns and outliers from this chart.



(a) (b) (c)

Fig. 2: Visualizing bivariate geographic data (blue represents number of McDonald's and orange represents number of Starbucks) with (a) pie-charts, (b) filled concentric circles, and (c) our bivariate cartogram. Consider three states: TX has almost the same number of McDonald's and Starbucks, VA has slightly higher number of McDonald's, and DC is a geographic outlier (having more Starbucks than McDonald's, whereas its neighbors have more McDonald's than Starbucks). All of these cases are clearly visualized in (c), in which TX is a gray circle, VA has a light blue ring, and DC is an orange circle surrounded by blue circles. Our design (c) also shows several geographic patterns: Starbucks has a higher density in the West (more orange circles, McDonald's in the East and Midwest (more blue circles), and the difference is the greatest in California (more Starbucks) and in Michigan (more McDonald's), as indicated by the thick rings around CA and MI.

these visualizations we can see that CA has highest numbers of McDonald's and Starbucks, and DC has the highest numbers per resident. While the scatterplots provide some distribution and clustering information (e.g., the dotted regression line shows the distribution pattern), they do not represent the underlying geographic information. Thematic maps and cartograms allow us to also see geographic trends and patterns.

Bivariate Maps: Bivariate maps are often described as a combination of two univariate map symbols. Several combinations of visual variable pairs have been used for bivariate mapping, such as size, shape, and hue [36], [62]. Nelson enumerates several bivariate map types, according to the combinations of the visual variables, and provides a typology of bivariate symbols [46]. For example, a choropleth map with superimposed symbols (e.g., graduated circles) one variable is encoded by color hue, and the other by the size of the [27]. In bivariate choropleth maps both variables are shown by colors. Other bivariate maps use symbols, such as bar charts or pie charts, overlaid on top of each region of a given map [3], [12].

In such visualizations, however, it is inherently difficult to make comparisons, find trends, and spot outliers. For example, we can use pie charts to visualize the bivariate maps of McDonald's and Starbucks (see Fig. 2(a)). Here, although DC is a geographic outlier, it is hard to spot. In other words, since all charts use both colors, such visualizations are not pre-attentive. Specifically, by the nature of the design, in Fig. 2(a) it is difficult to spot DC as an outlier, since every state is represented by both colors orange and blue. In contrast, in Fig. 2(c) each state is associated with just one color, which makes it easy to spot the odd orange circle (DC) in a sea of blue. In this design we follow Healey, et al.s observation that single-hue variation can be pre-attentively processed [33]. Attempting to combine pie charts with value-by-area visualizations will likely result in visualizations that are difficult to interpret, as color comparisons are fundamentally different than area comparisons.

There are other techniques, such as the use of glyphs to display multivariate data in the shape of a human face [18]. The individual parts, such as eyes, ears, mouth and nose in these "Chernoff faces" represent values of the variables by their shape, size, placement and orientation. In cartography, these face glyphs represent data on a map following the traditional methods of thematic representation [25]. A serious criticism of the use of face glyphs is that they can overload the viewer with information. There are other criticisms, such as, the dangers of conveying unintentional emotional messages, and racial stereotypes [45].

Note that a bivariate map is, by its nature, more visually complex than a univariate map. The visual complexity of a map, as defined by MacEachren [43], describes the degree of intricacy created by the map elements. The complexity increases the cognitive workload for the map reader, and if the map is too difficult to process mentally then this affects the utility of the map. As pointed out by Fisher, a bivariate map is effective only as long as the difficulty of comprehending two or more variables does not exceed the value of being able to relate them [28].

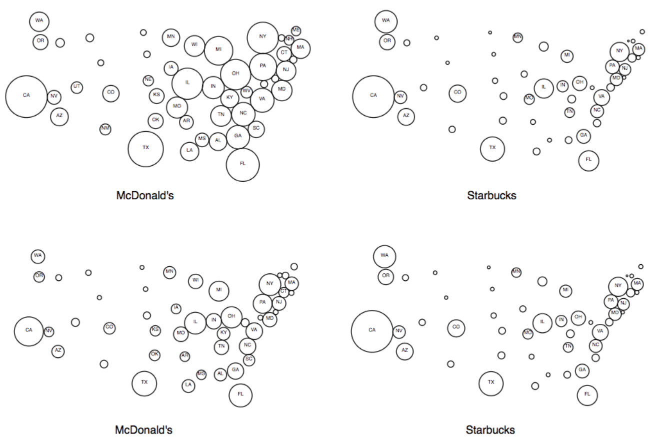


Fig. 3: Side-by-side monovariate cartograms for McDonald's and Starbucks shops, using individual normalization (top), and the same normalization (bottom). In the bottom row, two circles of the same area correspond to same number of shops in both cartograms.

3.2 Bivariate Data Visualization Using Cartograms

Side-by-side cartograms: A simple way to show bivariate data is to create a cartogram for each statistical variable and place the two of them side-by-side; see Fig. 3. While in general "small multiples" visualizations can be useful, it has been shown that side-by-side maps are not very effective [54]. Consider the side-by-side maps, showing the exact number of McDonald's and Starbucks shops, in the top row of Fig 3: it is difficult to see that the number of Starbucks is greater in the West, because circles of the same size correspond to different number of shops in the two cartograms.

We could scale both the cartograms using the same normalization unit (e.g., the maximum of both dataset); see the bottom row of Fig 3. This makes it plausible that patterns can be seen, because now circles of the same size correspond to the same number of shops in the two cartograms. However, the visualization is still difficult to analyze as we need to compare pairs of states in order to see patterns in the bivariate dataset.

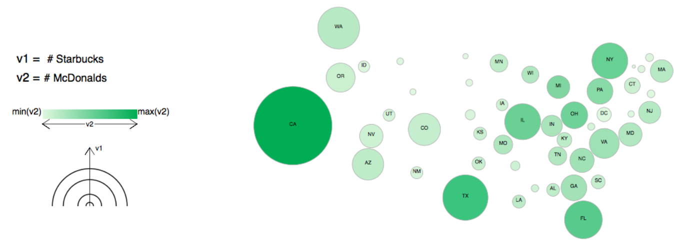


Fig. 4: Shaded cartogram for McDonald's and Starbucks.

Shaded cartograms: Another possible way to show the two variables is with a cartogram in which one variable is represented by size and the other by color gradation [62]. The major difficulty here is comparing values encoded by area to values encoded by color. For example, by examining Fig. 4 we can see that the Midwest has many McDonald's shops (darker shade of green), but it is difficult to spot that Starbucks outnumber McDonald's in all Pacific coast states.

To summarize, while several different map-based bivariate and multivariate visualizations have been proposed, there are no earlier methods encoding bivariate data using the standard value-by-area interpretation of cartograms. All the above approaches to simultaneously show two variables seem to have inherent limitations. Our goal is to design a simple value-by-area visualization for *bivariate data*, i.e., a visualization that shows two geo-statistical datasets on top of a geographic map.

4 Our Approach

We are interested in a visualization that simultaneously shows both statistical datasets as well as the underlying geography, so that we can find patterns, trends and outliers in the statistical variables and also in the geography. In statistics, an *outlier* is an observation point that is distant from other observations. We would like to be able to see both statistical outliers, as well as *geographic outliers*, defined as geographic regions with different statistical properties from their neighbors. With this in mind, we propose bivariate cartograms. Our simple visual encoding uses *size* to represent both variables and *color* to depict the binary relation between them (greater or smaller).

We implemented our technique for the major types of cartograms: contiguous, non-contiguous, Dorling, Demers and rectangular. Additional examples can be found online [5]; next we we describe the details for Dorling cartograms. In Section 5.2, we discuss how this approach can be generalized to any cartogram type.

4.1 Design Considerations

Consider a visual encoding, such as a Dorling cartogram, where the variables are represented by circle size (the larger the circle, the bigger the value); see Fig 5. The two variables are matched to two complementary pair of colors, blue and orange, as recommended for quantitative data in maps [11]. Variable 2 is larger and it is represented by a larger circle. In the "winner-takes-all" approach, each state is given the color of the larger variable. In this case, we are not encoding both

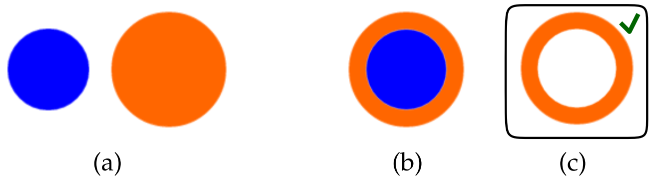


Fig. 5: (a) Two variables encoded by circle size (here Variable 2 $>$ Variable 1). (b) Combining the encodings for the two variables as colored concentric circles with the smaller one on top of the larger one. (c) Combining the encodings as concentric circles with the larger circle colored and the smaller one blank. Now the ring shows only the color of the larger circle, and the thickness of the ring encodes how much larger Variable 2 is, compared to Variable 1.

the variables simultaneously, and we cannot compare the two variables (there would be no difference in the encoding when one variable is 5% larger or 90% larger than the other).

We want to combine the variables in a way that clearly shows their binary relation (which is larger/smaller), and the difference between them. To do this, we could place the smaller circle on top of the larger one, using the same center. Then the inner blue circle shows the smaller data value, but covers a large part of the image. This type of overlay symbol has been used in geography, although as Brewer et al. [12] point out, it has serious limitations: "Overlay symbol construction is awkward where amounts are near equal because symbols are almost the same size but the slightly smaller one will take visual precedence." It also goes against Tufte's principle that the representation of numbers, as physically measured on the surface of the graphic itself, should be proportional to the numerical quantities represented [61]. Fig. 2(b) illustrates this approach for showing two geographic datasets: number of McDonald's and number of Starbucks in each state in the US.

Another way to combine the two circles is to fill the larger circle with its associated color (orange, in this example) and leave the inner circle uncolored. Now the ring between the larger circle and the inner circle is filled with the color of the larger circle. In this way, the thickness of the ring gives an estimate of the magnitude of one variable compared to the other, while the color of the ring is determined by the larger variable. We use this encoding in our design. The rules for color encoding of the ring are shown in Table 1.

Color	Criteria
Blue	If variable 1 is sufficiently larger than variable 2
Orange	If variable 2 is sufficiently larger than variable 1
Gray	Otherwise

TABLE 1: Color encoding for bivariate cartograms.

Figure 6(a) illustrates how the size and color of a circle change depending on the data values. Here, variable 1 is plotted on the x-axis, and variable 2 on the y-axis. If we move away from the origin along the x-axis, the value of v_1 increases; similarly, if we move away from the origin along the y-axis, the value of v_2 increases. Consider the case when v_2 has a fixed value of 4 and v_1 varies from 2 to 6. We begin with a thick orange ring (v_2 is larger than v_1). When the two values are nearly equal (in this case, 4), the ring turns gray. Eventually, the ring turns blue and its thickness increases.

The usual interpretation of circle size remains valid – the bigger the circle, the larger the data value. But now we can read the "bivariate" encoding as follows. Whether variable 1 is larger than variable 2 is encoded with a binary choice of colors – blue or orange. The difference between these two values is represented by the thickness of the "ring." In summary, there are two important features:

1) **Color of ring:** A blue ring means that variable 1 is (at least 3%) larger than variable 2; an orange ring means that variable 2 is (at least 3%) larger than variable 1; a gray ring means that they are roughly equal (within 3% of each other).
2) **Thickness of ring:** The thicker the ring, the bigger the difference between the two variables. In particular, the areas of the two circles for each state are proportional to the values of the two variables in that state, and the area of the ring is proportional to the difference.

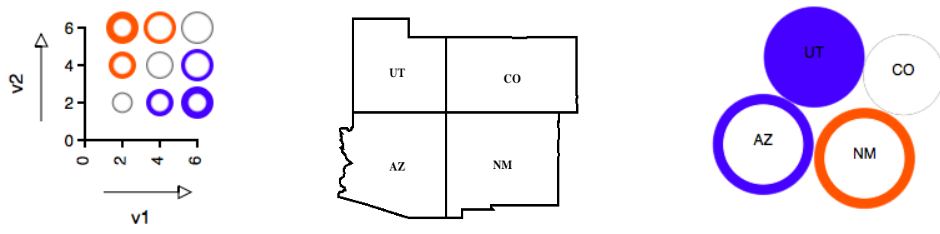
A variable is sufficiently larger if it is at least 3% larger than the other variable. This threshold value works well for most of the US datasets we considered. For different datasets and different maps, different thresholds values might be more appropriate, and the constant can easily be changed. Automatically determining the best threshold value would be an interesting problem for future work.

The bivariate cartogram using our approach is shown in Fig. 2(c). Details of implementation of this approach in generating bivariate Dorling cartograms is described in the following section.

4.2 Generating Bivariate Dorling Cartograms

In the original method of Dorling [25], the layout of the circles is based on predefined geographical constraints: circles try to stay in close contact with their original geographic neighbors and circles do not overlap. Unlike this approach, in many web implementations of Dorling cartograms, locality (keeping circles as close to their original positions as possible) is preferred over topology (keeping circles in close contact with their original geographic neighbors). We decided to offer an explicit balance between locality and topology using a force-directed layout, which considers three types of forces: (i) a repulsive force between each pair of overlapping circles, (ii) an attractive force (locality), keeping each circle center close to the original geographic center for the corresponding region, and (iii) an attractive force (topology) that keeps neighboring pairs of circles close to each other. We provide a slider with which the viewer can control the ratio between the two attractive forces; see Fig. 7 (more examples here [5]).

In addition to changing the force-directed layout for Dorling cartograms, we also modified the collision detection technique for bivariate data so that a collision is detected whenever the larger of the pair of circles for a state collides with a circle for another state. We also normalize the data values to meet the following goals: (i) different datasets with different ranges should be comparable, and (ii) the circles for different datasets should fit in the visualization window. Specifically, we normalize both datasets using the



(a) (b) (c)

Fig. 6: (a) Legend for the bivariate encoding: Variable 1 ($v1$) is represented by blue circles and Variable 2 ($v2$) is represented by orange circles. Larger circles correspond to larger values. If we move away from the origin along the x-axis, the value of $v1$ increases; therefore the thickness of the blue rings also increases. Analogously, the thickness of the orange rings increases along the y axis. When the two values are nearly equal, the ring is gray. (b) Geographic map for the four corners region of the USA: Utah (UT), Arizona (AZ), New Mexico (NM) and Colorado (CO). (c) Bivariate cartogram for the four corners region with $(v1, v2)$ values: $(80, 0)$, $(80, 50)$, $(50, 80)$, and $(50, 51)$. Note that the ring fills the entire circle when the value of one variable is 0, as in the case of $v2$ here.

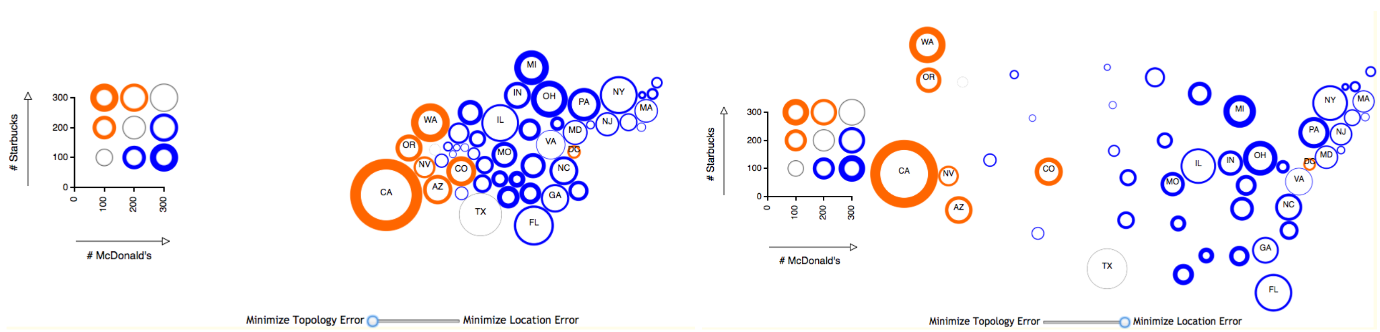


Fig. 7: A slider provides the viewer an explicit balance between locality and topology. On the left side, topology error is minimized, while on the right side location error is minimized.

maximum overall value (in either dataset), so that a state with this maximum value has a pre-specified radius (e.g., one unit). This implies that two circles of the same area correspond to same value in both cartograms. Thus the resulting representation can show the change in the data value for each state.

Note that Fig. 2 represents absolute numbers: circles are directly proportional to the number of Starbucks or McDonald's. In *per-capita* cartograms, the data is scaled by the population, in order to explore the effects of population density. For example, Fig. 8 shows the per-capita bivariate cartogram of Starbucks and McDonald's. Each circle represents the number of McDonald's and Starbucks shops, per 100,000 residents. Note that the colors of the circles match those in the absolute numbers cartogram, as the binary relation (more Starbucks or more McDonald's) is not affected by the per-capita normalization. However, with the per-capita bivariate cartogram we can see additional information. For example, although California (CA) is a large state with large population, in the per capita bivariate cartogram, CA is average-sized. Other states have higher per-capita numbers of Starbucks and McDonald's, most notably Washington (WA) and the District of Columbia (DC).

Fig. 9 illustrates another example of a bivariate cartogram, showing the population in the US in 1930 and 1950 (before and after the Second World War). Once again, the sizes of the circles indicate that the Northeast (e.g., New York and Pennsylvania) and Midwest (Illinois and Michigan) had the largest population. The thickness of the rings show that the fastest growth in population was in California and Florida. Finally, the colors indicate that in most states, the population increased in this 20 year period. The exceptions are several Great Plains states, such as Kansas and Nebraska, where there is little change, and North Dakota, South Dakota and Oklahoma, where population decreased.

Interaction: We augmented our design for bivariate cartograms with simple interactive features, such as showing data values on mouse-over events. Another interactive feature makes it possible to switch between bivariate cartograms, and the traditional (monovariate) cartograms for each dataset; see Fig. 10. Here, the viewer can choose to see either one of the datasets or, both of them simultaneously. This additional feature allows the viewer to see geographic patterns and distribution individually in either dataset, as well as simultaneously in the bivariate dataset. Note that we have implemented these interactions for all the three visualization techniques. However, for the main (timed) part of the study, the interactivity was disabled in order to fairly compare the effectiveness of the three static techniques. With these interactions enabled, both time and accuracy will be very high, as exact numerical values (shown on mouseover events) are easy to compare.

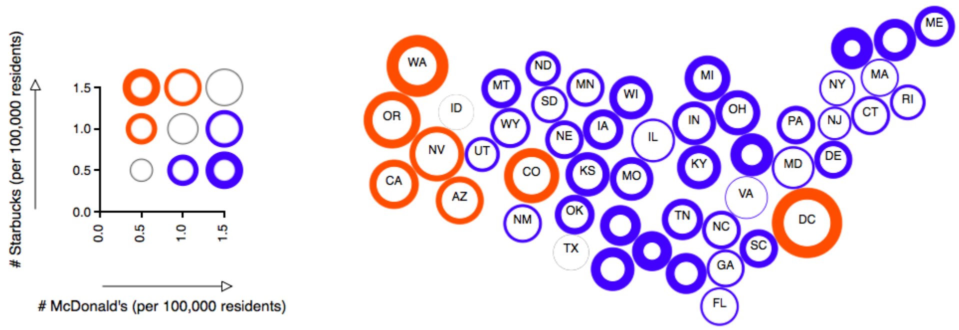


Fig. 8: A per-capita bivariate cartogram showing the distribution of Starbucks and McDonald's shops per 100,000 residents in the US. From this cartogram we see that it is not California, but Washington (WA) and the District of Columbia (DC) that have the most Starbucks per capita.

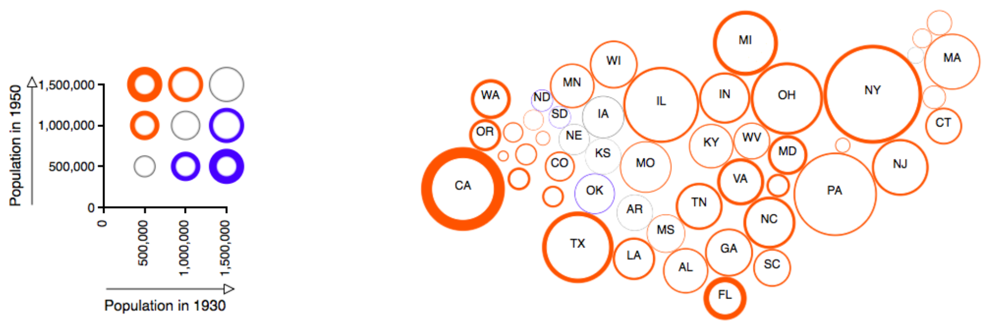


Fig. 9: A bivariate cartogram showing US population in 1930 and 1950. The population increased in most states, except for several Great Plains states, such as Kansas and Nebraska (where there is no change), and North Dakota, South Dakota and Oklahoma (where population decreased).

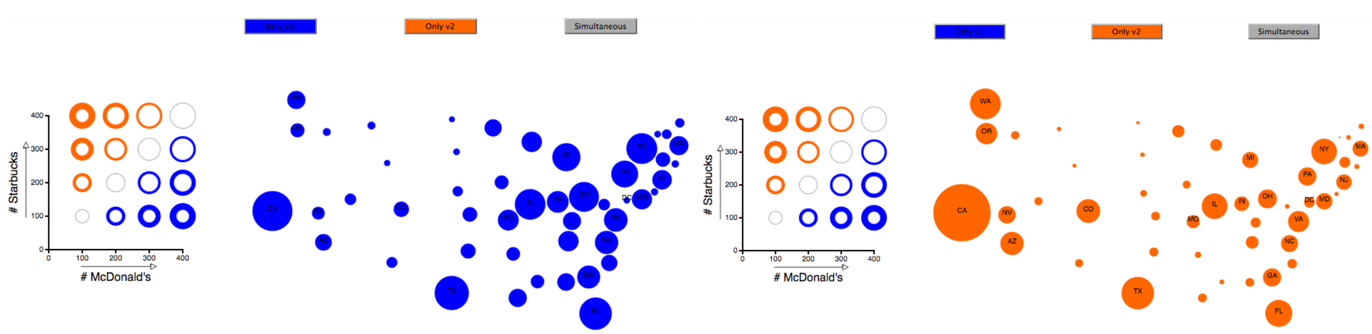


Fig. 10: Interactive features make it easy to switch between the bivariate view and the monovariate view for variable 1 (left), or variable 2 (right). The simultaneous view of the bivariate cartogram is shown in Fig. 2(c).

5 Extensions and Generalizations

One of the advantages of simple techniques is that they can often be easily extended and generalized. We briefly discuss how to extend the proposed techniques to data from different domains and how to generalize to different cartogram types.

5.1 Extensions to Different Domains

In the examples shown so far, both variables use the same scale (e.g., the number of Starbucks and McDonald's shops), or are from the same domain (population in two different years). The proposed techniques can be extended to datasets from different domains. Consider, for example, population and GDP data, which have very different ranges of values. In this case a different normalization must be used for each dataset, in order to make them comparable to each other in the visualization. We compute the average value for each dataset and map this average value to a circle with pre-defined radius (e.g., one unit). Since the average values of both datasets are mapped to the same radius, the resulting visualization shows for each state the contribution of that state to the total value for each variable.

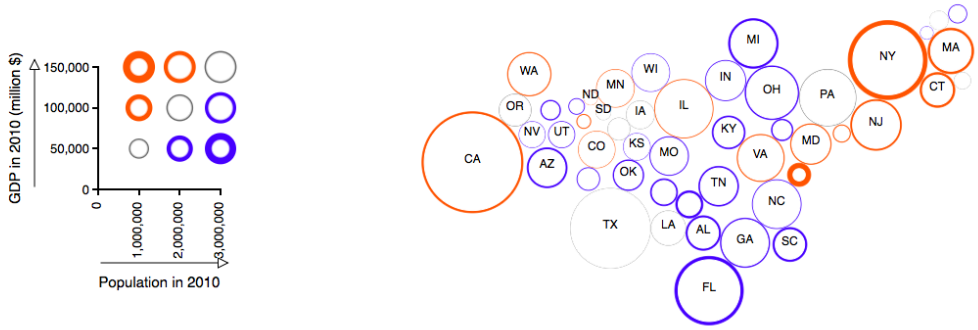


Fig. 11: A bivariate cartogram showing the relative distribution of population and GDP of the US in 2010. In general, coastal states contribute more GDP, with notable exceptions of a few inland states with major cities: Illinois (Chicago), Colorado (Denver), Minnesota (Minneapolis).

We illustrate this in Fig. 11, where two datasets are normalized, so that a state with population equal to the average population and a state with GDP equal to the average GDP, both have equal circle radius (25 units). Large blue rings, such as CA and NY indicate that these two states have large population and large GDP, with GDP dominating the comparison. The large orange ring for FL indicates large population and GDP, however here, population dominates the comparison. In this way, we can compare the contribution of each state to the total population and the contribution of each state to the total GDP.

Specifically, for the Dorling bivariate cartogram which encodes two datasets X and Y from different domains, for each state S the sizes of the two circles and the colors of the rings can be described as follows:

(i) The areas of the two circles for S are proportional to the values of $\frac{X(S)}{\sum X}$ and $\frac{Y(S)}{\sum Y}$, where $X(S)$ and $Y(S)$ denote the scalar X and Y values for the state S respectively, and $\sum X$ and $\sum Y$ denote the total values over all the states in the map. In other words, the two circle areas are proportional to the fraction of the *contribution* of S to the total values of X and Y, respectively. The area of the ring is proportional to the difference in the contribution $|\frac{X(S)}{\sum X} - \frac{Y(S)}{\sum Y}|$ in S.

(ii) The color of a ring represents which variable contributes more. A blue ring indicates that the circle for $X(S)$ is at least 3% larger than the circle for $Y(S)$; a red ring indicates that the circle for $Y(S)$ is at least 3% larger; a gray ring indicates that the state contributes roughly equally (within 3% of each other) to both variables.

5.2 Cartogram Types

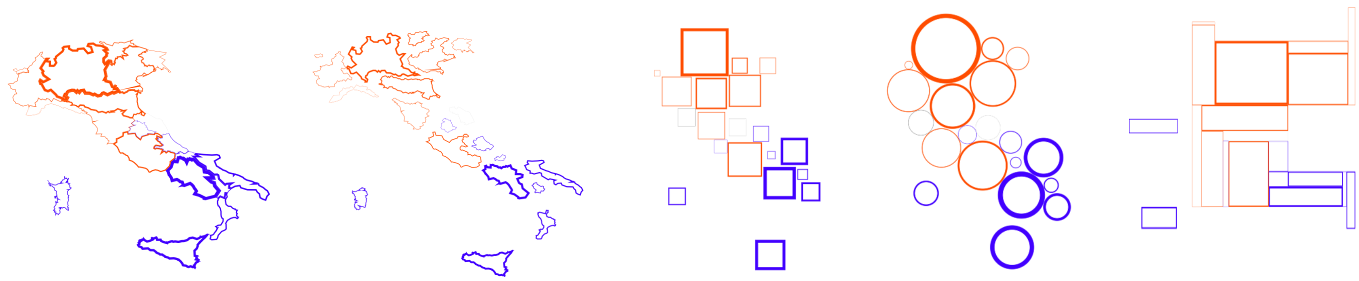
We described the proposed techniques using Dorling cartograms, yet they do generalize to all the major cartogram types. In particular, we designed and implemented the bivariate cartogram encoding for four other types of cartograms: contiguous, non-contiguous, Demers and rectangular cartograms; see Fig. 12. We begin with a brief description of the four cartogram types for which we implemented our bivariate cartogram encoding.

Contiguous Cartograms: These cartograms deform the regions of a map (by pulling and pushing the boundaries), so that the desired areas are obtained, while adjacencies are maintained. The original map is often recognizable, but the shapes of some countries might be distorted. We use the diffusion-based algorithm of Gastner and Newman [32]. The input map is projected onto a distorted grid, computed in such a way that the areas of the countries match the pre-defined values. This distorted grid is obtained by an iterative diffusion process, where quantities flow from one grid cell to another until a balanced distribution is reached.

Non-Contiguous Cartograms: These cartograms are created by starting with the regions of a map, and scaling down each region independently, so that the desired size/area is obtained. They satisfy area and shape constraints, but do not preserve the topology of the original map. The non-contiguous cartograms method of Olson [50] scales down each region in place (centered around the original geographical centroid), while preserving the original shapes. For each region, the density (statistical data value divided by geographic area) is computed and the highest-density region is chosen as the anchor: its area remains unchanged while all other regions become smaller.

Demers Cartograms: A Demers cartogram [9] is a variant of a Dorling cartogram, where squares are used in place of circles. Demers cartograms have no cartographic errors, but do not preserve shapes. Cartographic error measures the relative distortion of the area of each modified region from the desired statistic [49]. Since squares can be packed more compactly than circles, Demers cartograms can capture the underlying map topology better than Dorling cartograms.

Rectangular Cartograms: Rectangular cartograms represent regions with rectangles. These are "topological cartograms" where the adjacency relation between the regions of the map is represented by the dual graph of the map, and that graph is used to obtain a schematized representation with rectangles. In rectangular cartograms there is often a trade-off between achieving zero (or small) cartographic error and preserving the map properties (relative position of the regions, adjacencies between them). In our design, we use a state-of-the-art rectangular cartograms algorithm [15]. There are several options for this type of algorithm and we choose the variant where the generated cartogram preserves



(a) Contiguous (b) Non-contiguous (c) Demers (d) Dorling (e) Rectangular

Fig. 12: Bivariate cartograms of Italy showing population (blue) and GDP (orange) of Italy in 2011 using, respectively, contiguous, non-contiguous, Demers, Dorling and rectangular cartograms. Northern Italy contributes more to GDP than population (hence more orange) and Southern Italy contributes more to population than GDP (more blue).

topology (adjacencies) at the possible expense of some cartographic error.

Fig. 12 illustrates the bivariate cartogram encoding on contiguous, non-contiguous, Demers, Dorling, and rectangular cartograms, where each cartogram shows the population and GDP of the regions of Italy in 2011.

5.3 Generalization to Other Cartogram Types

We briefly summarize a general algorithm for bivariate cartograms from any standard monovariate cartogram type.

1) First we normalize the two datasets. If the datasets are from the same domain, we normalize both using the overall maximum value. Otherwise, we normalize each dataset using the respective averages (as described in Section 5.1. Let v_1 and v_2 denote the functions for the two normalized datasets.
2) Define the two functions v_{max} and v_{min}, where for each region S, $v_{max}(S) = \max\{v_1(S), v_2(S)\}$, and $v_{min}(S) = \min\{v_1(S), v_2(S)\}$.
3) Design a cartogram of the specified type using v_{max} as the statistical weight, i.e., one where each region S is a polygon $P(S)$ with area proportional to $v_{max}(S)$.
4) Inside each polygon $P(S)$, draw an inner polygon $Q(S)$ with area proportional to $v_{min}(S)$, by *inward-offsetting* the polygon $P(S)$. *Inward Offsetting* $P(S)$ by t units, shifts each edge of $P(S)$ by t units towards the inside of the polygon. For the cartogram types, where the polygons or shapes are regular (such as circles in Dorling cartograms, and rectangles in rectangular cartograms), one can find the value of t such that inward-offsetting $P(S)$ by t units yields a polygon with area proportional to $v_{min}(S)$ in constant time. For other cartogram types (such as contiguous and non-contiguous cartograms), we use a numerical approach to find a value of t so that the inner polygon has area approximately proportional to $v_{min}(S)$. Computing the inward offset for a polygon requires geometric algorithms and data structures, provided in the CGAL library [1].
5) Color the area between $P(S)$ and $Q(S)$ for each region S (a ring, in the case of Dorling cartograms), using the rules from Table 1.

Note that there is a natural limitation to this generalization, when going from "nice" shapes such as circles, squares and rectangles, to arbitrary shapes. For non-convex shapes, in particular, inwards offsetting may result in disconnected interior representations.

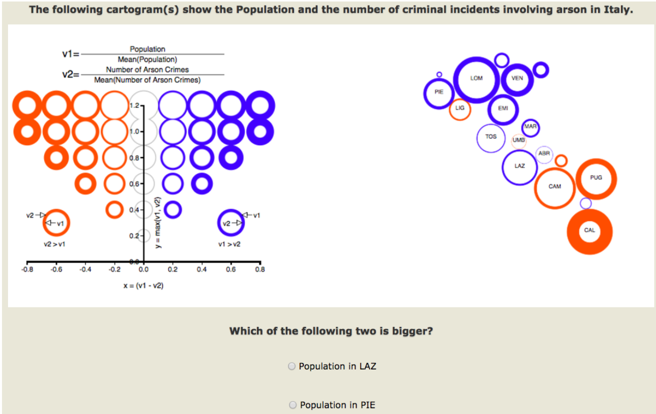
6 Experiment and Evaluation

To validate our cartogram technique, we conducted two small studies: a pilot study on 8 participants, and a full-scale controlled experiment on 23 participants. The participants were asked to perform tasks using our bivariate cartograms, and two of the other visualizations for showing bivariate data (side-by-side cartograms and shaded-cartograms). The questions involved "comparison" tasks, as well as meta-data tasks such as "summarize" and "find outlier." Given our design decisions, we expected that participants would perform better with bivariate cartograms for summarizing patterns and finding outliers. Since the bivariate cartogram design inherits area-perception issues of standard cartograms, we did not expect that participants would perform area-comparisons better, but we anticipated that performance would not be worse than with the other two visualizations. We evaluated the effectiveness of the different visualizations by measuring accuracy and completion time for the visualization tasks, as well as via subjective preferences and participant feedback.

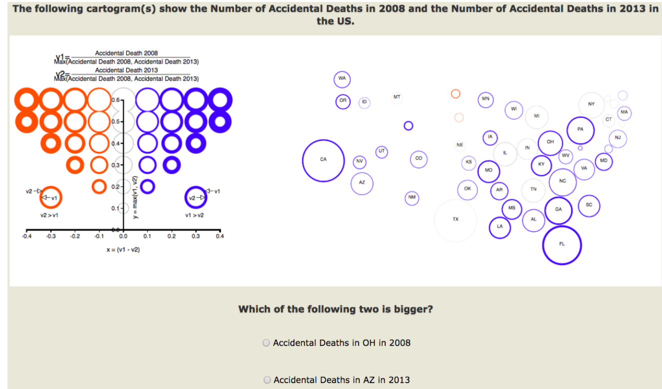
6.1 Visualization Tasks

There is a large number of task taxonomies in cartography, information visualization and human-computer interaction [10], [55], [56]. Visualization tasks are defined and classified, often depending on the context and scope of the tasks [7], [66]. Visualization tasks have also been categorized across different dimensions [10], [56]. These taxonomies provide general guidelines for visualizations, but are not specifically designed for cartograms.

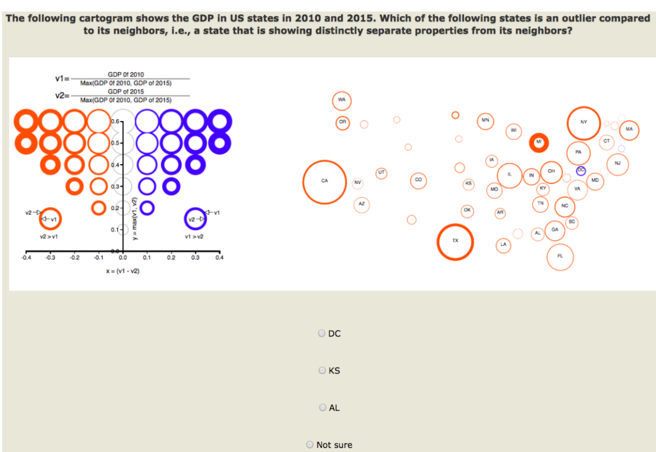
A recent task taxonomy for cartograms adapts tasks from cartography and information visualization and adds new cartogram-specific tasks [48]. This taxonomy categorizes cartogram tasks into four groups based on the design dimensions of cartograms: group 1 is related to the shape preservation of regions in cartograms, group 2 focuses on *comparison* tasks, group 3 checks for topology preservation in cartograms, and group 4 is associated with meta-data extraction, such as *finding outliers* and *summarization*.



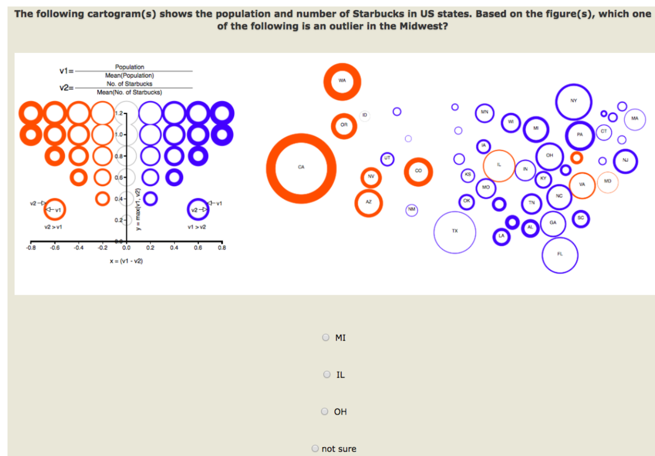
(a) *Compare-within* task



(b) *Compare-across* task



(c) *Summarize* task



(d) *Summarize* task

Fig. 13: Example tasks on bivariate cartograms.

As our design does not impact the shape recognition (group 1) or topology preservation (group 3), we used comparison tasks (group 2) and meta-data tasks (group 4). From these groups we choose the most commonly used tasks in cartogram evaluations: *compare* and *summarize*. In bivariate cartograms, we encode two variables at the same time, making it possible to compare data points within the same set, or between datasets. Therefore we split the generic task *compare* into two subcategories: *compare within* the same dataset and *compare across* datasets. Similarly, for meta-data extraction tasks, we include "summarization" as well as "find outlier" questions; see Fig. 13 for example tasks.

6.2 Visualization Techniques

We compare across techniques, using Dorling cartograms for consistency. **Bivariate Cartograms:** these are generated using our method, described in Section 4. **Side-by-side (Monovariate) Cartograms:** for this visualization we used two standard Dorling cartograms, placed side-by side, to show the two datasets; see Fig. 3. The cartograms are normalized; see Section 3.1 for more details. **Shaded-Cartogram Visualization:** one dataset is realized by circle areas, and the other by color gradation; see Fig. 4.

6.3 Pilot Study

We first conducted a pilot study with 8 participants. They were all university students with background in computer science and electrical engineering. We described the problem and the visualization and asked the participants to perform several tasks and answer multiple-choice questions. After the meeting, we asked the participants to comment on the proposed bivariate cartograms. Most of the feedback was positive, especially noting that it is easier to see overall patterns and find outliers. Some specific comments included: "Nice! The bivariate is a kind of 'highlight' on the dataset", "This makes it easier to make comparisons", "This method of visualization definitely made the tasks easier."

These comments also provided some useful suggestions, such as issues with label size. These recommendations generated both formative suggestions for improvement and summative feedback in terms of visual encodings.

6.4 Controlled Experiment

We conducted an experiment [6] to evaluate the three visualization techniques, via quantitative measurements of task accuracy and completion time. Together, the qualitative and quantitative evaluation required about 30 minutes.

6.4.1 Hypotheses Formulation

Our hypotheses are informed by prior cartogram evaluations, perception studies, and popular critiques of cartograms. We formulate the following hypotheses for bivariate data:

H1: For questions that involve detecting outliers, summarizing the results and understanding patterns in data, participants will likely perform better (in terms of completion time and accuracy) with bivariate cartograms.This hypothesis is based on the observation that the bivariate cartograms present the dataset in a pre-attentive way, with a goal to making geographic trends and outliers stand out.

H2a: For comparison within the same domain, there will be no discernible impact of the visualization technique on performance (completion time and accuracy).

H2b: For comparing between different domains, both bivariate cartograms and side-by-side cartograms are likely to outperform the shaded-cartograms. There should be no discernible difference between bivariate and side-by-side cartograms. For the *compare-across* task, participants have to compare between two datasets. In the shaded cartogram one dataset is encoded in area and the other in color. For all the other comparison tasks in all the visualization techniques, participants compared data encoded in the same variable (either area or color). Magnitude comparison of attributes with different encodings has proved to be very difficult [54]. It has also been noted that for representing numerical data, size is preferable than color [44].

6.4.2 Participants and Datasets

We recruited 23 participants: 16 male and 7 female; 14 between the ages of 18–25 and 9 between 25–40. The highest completed education levels were: 2 high school, 9 undergraduate, 10 Masters and 2 PhD. Since some of our tasks require the subjects to identify regions highlighted with orange and blue colors, all participants were first tested for color blindness. None of the participants had any issue with the colors we used.

To reduce possible bias, we used three country maps (USA, Germany and Italy) and few different statistics: population, GDP, number of Starbucks, crime rates in different years and number of accidental deaths in different years in the USA; population and GDP of Germany; and population and number of arson-related crimes in Italy. We used a within-subject experimental design. For each subject, questions were selected from all the visualization types and all the tasks. For each of the three tasks (*compare-within*, *compare-across*, and *summarize*), the questions were drawn from a pool of questions involving all cartograms used for the task. Also, to guard against possible bias, questions within each set of tasks were randomized for each participant. For each type of task and each type of cartogram, two questions were asked. Therefore, each participant answered 18 task-related questions = 3 tasks $\times$ 3 cartograms $\times$ 2.

For the *comparison* tasks, pairs of circles were selected to avoid cases that are too easy or too hard. Specifically, the larger circle was at least 1.3, and at most 2.5 times larger than the smaller circle. This range was determined based on prior cartogram evaluations [47] and the pilot study.

In addition to the the visualization tasks, we assessed subjective preferences and logged verbal and written participant feedback. We asked the participants to subjectively rate each visualization type, at the beginning of the experiment, before the task-related questions, and then again at the end of the task-related questions. Towards the end of the study, we also asked the participants to select one of the three visualizations with which they would like to perform additional tasks. Finally we also collected feedback and comments from them at the end.

7 Results and Data Analysis

We use ANOVA F-tests at the significant level $\alpha = 0.05$ to carry out the statistical analysis. The within-subject independent variables are the three visualization methods and the two dependent measures are the participants' average completion times and error percentages, shown in the last two columns of Table 2. The null hypothesis is that the visualization methods do not affect completion times and error rates. When the probability of the null hypothesis (p-value) is less than 0.05 (or, equivalently the F-value is greater than the critical F-value, $F_{cr} = F_{0.05}(2, 66) = 3.14$), the null hypothesis is rejected. In case the null hypothesis is rejected, paired t-tests are utilized for the post-hoc analysis, with Bonferroni correction on the significance level $\alpha = 0.05$. For pairwise comparison between 3 visualizations (i.e., 3 different pairs), we conclude that there is a significant difference in the mean completion time (resp. mean error rate), if the computed t-value is greater than the critical t-value, $t_{cr} = t_{0.05/3}(22) = 2.59$.

There is strong evidence in support of Hypothesis 1, based on the results of the *summarize* task. In particular, there is statistically significant improvement in error rates for the bivariate cartograms over both the traditional visualizations, although there is no significant impact on completion time; see Table 2. This implies that for finding trends, outliers and summarizing results, the participants found the bivariate cartograms more helpful than the others.

Hypothesis H2a is partially supported by the experimental evidence, since there are no statistically significant differences between the visualization methods in terms of error rate for the *compare-within* task; see Table 2). Bivariate cartograms required significantly more time for this task, but we suspect that this might be attributed to the learning curve involved with the new visualization (in particular since for the later two tasks in the experiment, there is no significant difference in terms of completion time and the variances in both error rate and completion time decreased for bivariate cartogram in the later tasks.

Hypothesis H2b is not supported by the results in the experiment, since there are no statistically significant differences between the visualization methods in terms of error rate or completion time for the *compare-across* task; see Table 2. This implies that there is neither significant improvement nor deterioration when using bivariate cartograms for this type of task.

In general, the experimental evidence suggests that for visualization of bivariate data, our proposed method is capable of showing patterns, trends and summary more effectively than the traditional approaches while it does not negatively impact other aspects of the visualization.

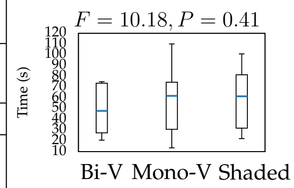
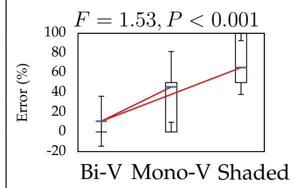
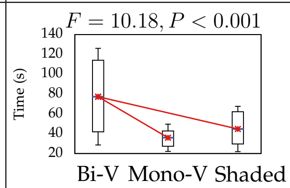
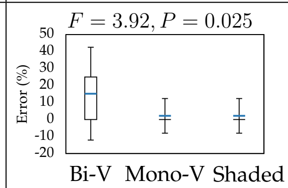
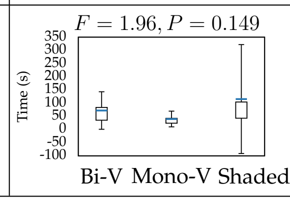
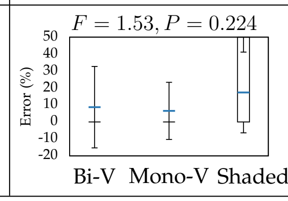
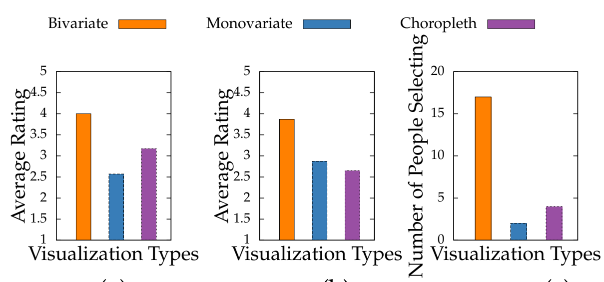
		Question	Time (s)	Error %
H1	*Summarize: given a vis...*	...for the number of accidental deaths in the US in 2008 and 2013, which state has the biggest decrease?		
		...for crimes in the US in 2000 and 2012, which state is an outlier?		
		...for the US GDP in 2010 and 2015, which state is an outlier?		
		...for the population and GDP, which statements are true?		
		...for the population and number of starbucks in the US, which state is an outlier?		
		...for the population and number of starbucks in the US, which region has largest starbucks density?		
		...for the population and number of arson-related crimes in Italy, which statements are true?		
H2a	*Compare-within (one domain)*	Which of the two specified data is bigger? (The two data are selected from the same dataset; see Fig. 13(a))		
H2b	*Compare-across (two domains)*	Which of the two specified data is bigger? (The two data are selected from different datasets; see Fig. 13(b))		

TABLE 2: The last two columns show average completion time in seconds and error percentage for the different techniques, along with the F and p values from ANOVA F-tests. The critical value of F is 3.14. The bottom and top of the boxes and the blue band represent first quartile, third quartile and mean. The distance between whiskers and the band shows standard deviation. The red line segments indicate statistically significant relationships obtained using paired t-tests with Bonferroni correction. The critical value of t is 2.59.

7.1 Feedback and Subjective Ratings

In addition to quantitative measurements of time and error, we also gathered feedback and subjective preferences from participants. All participants mostly gave positive feedback on the effectiveness of bivariate cartograms.



(a) (b) (c)

Fig. 14: Subjective ratings for the visualizations before the study (a), after the study (b), and number of participants selecting a visualization for remaining tasks (c).

At the beginning of the experiment, after the introduction of the three visualization methods used, the participants rated each visualization type on Likert scales (excellent = 5, good = 4, average = 3, poor = 2, very poor =1); see Fig. 14(a). Again after performing all the visualization tasks, we collected their rating on the same Likert scale; see Fig. 14(b). The purpose was to see whether their preferences change after performing the tasks. Fig. 14(a)–(b) shows that the participants prefer the bivariate cartograms over the other two techniques both before and after performing the tasks, although their preference between the two traditional visualizations changes.

At the end of the study, we also asked the participants to select one of the three visualizations which would be used for an additional group of three questions. The purpose of these questions was to see which cartogram the subject preferred the most, and choose to work on among the three variants. 74% participants chose bivariate cartograms.

Finally we also collected feedback from the participants about the effectiveness and design decisions for bivariate cartograms. All participants found bivariate cartograms to be clearer, and easier to interprete than other methods of visualization. One participant wrote “Bivariate cartograms are definitely superior to the other visualization methods for most types of questions.” Another participant wrote “I like them. They make it easier to see the differences in the variables than the other types at a glance.” Few participants mentioned the difficulty in comparing across dataset “ It was difficult, however, to answer questions in which a comparison is required between the inner ring of one data point and the outer ring of another.”

Although bivariate cartograms mosty received positive response, some participants mentioned the learning process in bivariate cartograms “It’s much better than others to show two datasets in the same time. But the ‘learning curve’ for reading bivariate cartograms seems steeper than the others.”, and some other participant wrote “Difficult to learn, but can answer question with more certainty than the other two techniques."

8 Limitations

Standard (monovariate) cartograms have several known limitations. For example, the statistical data might force regions to become too small. Another example is that some cartogram types and some extreme statistical distributions might make it difficult to recognize the underlying geography. All of these directly affect bivariate cartograms. Legends, labels, and basic interaction techniques can be used to alleviate some of these known issues.

Area perception problems affect cartograms in general and bivariate cartograms in particular. Further, for non-convex shapes in contiguous and non-contiguous settings, our technique might result in disconnected regions.

Cartograms, like other statistical maps are prone to the modifiable areal unit problem (MAUP) [2]: the scale at which one chooses to analyze and visualize information can produce different results. This problem is also inherent in bivariate cartograms.

Our approach encodes the difference between the two statistics in the boundary of the regions of a map. This makes it difficult to generalize the approach to more than two variables and also limits the scalability of the approach (e.g., when there are many regions, most of the area will be white space).

When comparing two datasets from different domains, our visualization does not show the variance of the datasets directly. Exploring different encodings and ways to show more information about the two datasets seems like a challenging task that is worth exploring.

We described the proposed approach focusing on Dorling cartograms, although the approach generalizes to most cartogram types. We designed and implemented bivariate cartograms for several major types of cartograms (contiguous, non-contiguous, Dorling, Demers and rectangular), but only evaluated the Dorling variant.

Finally, the evaluation was limited to a particular subset of tasks, particular choices for the task settings, and involved a small number of participants. There are numerous limitations when considering the types of possible geographic maps (e.g., more countries, continents, or even synthetic maps) and the types of statistical data shown on these maps (e.g., more extreme area changes, more control over such changes).

9 Conclusions

We described simple cartogram-based techniques for encoding two geo-referenced datasets, allowing the viewer to simultaneously explore statistical and geographical correlations, patterns, trends, and outliers. We also performed a small-scale evaluation of the proposed techniques based on *summarize* and *compare* tasks. Despite several limitations, our results indicate that these representations provide a useful way to communicate geo-statistical information. The bivariate cartogram, in particular, provides significantly better performance (in terms of time and error) for meta-data extraction tasks, such as finding outliers and summarization. Comparison tasks were more time-consuming and error-prone across all techniques, although such issues might be alleviated with interactions. The proposed techniques can be extended to data from different domains and generalized to different cartogram types, although it is not clear how to generalize our approach to more than 2 variables.

We believe that the simplicity of the approach, combined with the relatively straight-forward interpretation of the resulting cartograms, makes a compelling case for the use of value-by-area representation of bivariate geo-referenced data. We make the implementations of the proposed visualizations available online [5]; the experimental study is also available [6].

References

[1] The computational geometry algorithms library. http://www.cgal.org.

[2] An introduction to the modifiable areal unit problem. http://gispopsci.org/maup/. [Online; accessed 12/28/2014].

[3] London squared map. http://aftertheflood.co/projects/london-squared-map.

[4] Scatterplot: Outlier. http://www.itl.nist.gov/div898/handbook/eda/section3/scattera.htm.

[5] Bivariate cartograms. http://bivariate.weebly.com, 2016.

[6] User Evaluation Study for Cartograms. http://cartogram.cs.arizona.edu:5775/, 2017.

[7] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. *IEEE Symposium on Information Visualization*, pages 111–117, 2005.

[8] J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. The University of Wisconsin Press, Madison, 1983.

[9] I. Bortins, S. Demers, and K. Clarke. Cartogram types. http://www.ncgia.ucsb.edu/projects/Cartogram_Central/types.html, 2002.

[10] M. Brehmer and T. Munzner. A multi-level typology of abstract visualization tasks. *IEEE Transaction on Visualization and Computer Graphics*, 19(12):2376–2385, 2013.

[11] C. Brewer. Color chart use in map design. *Cartographic Perspectives*, (04):3–10, 1989.

[12] C. Brewer and A. J. Campbell. Beyond graduated circles: varied point symbols for representing quantitative data on maps. *Cartographic Perspectives*, (29):6–25, 1998.

[13] C. Brewer, A. M. MacEachren, L. W. Pickle, and D. Herrmann. Mapping mortality: Evaluating color schemes for choropleth maps. *Annals of the Association of American Geographers*, 87(3):411–438, 1997.

[14] C. Brewer and L. Pickle. Evaluation of methods for classifying epidemiological data on choropleth maps in series. *Annals of the Association of American Geographers*, 92(4):662–681, 2002.

[15] K. Buchin, B. Speckmann, and S. Verdonschot. Evolution strategies for optimizing rectangular cartograms. In *Geographic Information Science*, volume 7478, pages 29–42, 2012.

[16] T. C. Callaghan. Interference and dominance in texture segregation: Hue, geometric form, and line orientation. *Perception & psychophysics*, 46(4):299–311, 1989.

[17] R. G. Cano, K. Buchin, T. Castermans, A. Pieterse, W. Sonke, and B. Speckmann. Mosaic drawings and cartograms. *Computer Graphics Forum*, 34(3):361–370, 2015.

[18] H. Chernoff. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68(342):361–368, 1973.

[19] J. H. Claessen and J. J. Van Wijk. Flexible linked axes for multivariate data visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2310–2316, 2011.

[20] W. S. Cleveland, C. S. Harris, and R. McGill. Judgments of circle sizes on statistical maps. *Journal of the American Statistical Association*, 77(379):541–547, 1982.

[21] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *J. of the American stat. assoc.*, 79(387):531–554, 1984.

[22] B. Dent. *Cartography: Thematic map design*. New York: McGraw Hill, 1999.

[23] B. D. Dent. Communication aspects of value-by-area cartograms. *The American Cartographer*, 2(2):154–168, 1975.

[24] T. Donnellan. Thematic cartography and geographic visualization. *Journal of Spatial Science*, 49(2):135–136, 2004.

[25] D. Dorling. *The visualization of spatial social structure*. PhD thesis, Department of Geography, University of Newcastle-upon-Tyne, U.K., 1991.

[26] D. Dorling. *Area Cartograms: Their Use and Creation*, volume 59 of *Concepts and Techniques in Modern Geography*. University of East Anglia, 1996.

[27] M. E. Elmer. *Symbol Considerations for Bivariate Thematic Mapping*. PhD thesis, University of Wisconsin–Madison, 2012.

[28] H. T. Fisher. Mapping information: The graphic display of quantitative information. *ABT Books, Cambridge, MA*, 1983.

[29] J. J. Flannery. The relative effectiveness of some common graduated point symbols in the presentation of quantitative data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 8(2):96–109, 1971.

[30] G. Fuchs and H. Schumann. Visualizing abstract data on maps. In *Information Visualisation*, pages 139–144, 2004.

[31] H. G. Funkhouser. Historical development of the graphical representation of statistical data. *Osiris*, 3:269–404, 1937.

[32] M. Gastner and M. Newman. Diffusion-based method for producing density-equalizing maps. In *Proceedings of The National Academy of Sciences of the USA*, volume 101, pages 7499–7504, 2004.

[33] C. G. Healey, K. S. Booth, and J. T. Enns. Harnessing preattentive processes for multivariate data visualization. In *Graphics Interface*, pages 107–107. CANADIAN INFORMATION PROCESSING SOCIETY, 1993.

[34] J. Heer and M. Bostock. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *CHI*, pages 203–212. ACM, 2010.

[35] D. H. House and C. J. Kocmoud. Continuous cartogram construction. In *IEEE Symp. Information Visualization*, pages 197–204, 1998.

[36] D. Howard and A. M. MacEachren. Interface design for geographic visualization: Tools for representing reliability. *Cartography and Geographic Information Systems*, 23(2):59–77, 1996.

[37] G. F. Jenks and F. C. Caspall. Error on choroplethic maps: definition, measurement, reduction. *Annals of the Association of American Geographers*, 61(2):217–244, 1971.

[38] J.-H. Kämper, S. G. Kobourov, and M. Nöllenburg. Circular-arc cartograms. In *IEEE Pacific Visualization Symposium (PacificVis'13)*, pages 1–8, 2013.

[39] D. Keim, S. North, and C. Panse. Cartodraw: A fast algorithm for generating contiguous cartograms. *IEEE Transactions on Visualization and Computer Graphics*, 10(1):95–110, 2004.

[40] D. Keim, S. North, C. Panse, and J. Schneidewind. Visualizing geographic information: VisualPoints vs. CartoDraw. *Information Visualization*, 2(1):58–67, 2003.

[41] S. Kim, R. Maciejewski, A. Malik, Y. Jang, D. S. Ebert, and T. Isenberg. Bristle maps: A multivariate abstraction technique for geovisualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(9):1438–1454, 2013.

[42] J. Krygier. Perceptual scaling of map symbols — making maps: DIY cartography. https://makingmaps.net/2007/08/28/perceptual-scaling-of-map-symbols/, 2007.

[43] A. M. MacEachren. Map complexity: Comparison and measurement. *The American Cartographer*, 9(1):31–46, 1982.

[44] A. M. MacEachren. *How maps work: representation, visualization, and design*. Guilford Press, 1995.

[45] I. Muehlenhaus. *Web cartography: map design for interactive and mobile devices*. CRC Press, 2013.

[46] E. Nelson. The impact of bivariate symbol design on task performance in a map setting. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 37(4):61–78, 2000.

[47] S. Nusrat, M. J. Alam, and S. Kobourov. Evaluating cartogram effectiveness. *IEEE Transactions on Visualization and Computer Graphics*, 2016.

[48] S. Nusrat and S. Kobourov. Task taxonomy for cartograms. *EuroVis (short paper)*, 2015.

[49] S. Nusrat and S. Kobourov. The state of the art in cartograms. In *Eurographics Conference on Visualization (EuroVis)-STARs*, 2016.

[50] J. Olson. Noncontiguous area cartogram. *The Professional Geographer*, 28(4):371 – 380, 1976.

[51] J. M. Olson and C. A. Brewer. An evaluation of color selections to accommodate map users with color-vision impairments. *Annals of the Association of American Geographers*, 87(1):103–134, 1997.

[52] G. Palsky. Des chiffres et des cartes: Naissance et developpement de la cartographie quantitative franqaise au XIXe siécle. *Paris: Comite des travaux historiques et scientifique*, 1996.

[53] E. Raisz. The rectangular statistical cartogram. *Geographical Review*, 24(3):292–296, 1934.

[54] K. Ramachandran. Visualizing and comparing multivariate scalar data over a geographic map. Master's thesis, North Carolina State University, 2009.

[55] R. E. Roth. An empirically-derived taxonomy of interaction primitives for interactive cartography and geovisualization. *IEEE Transaction on Visualization and Computer Graphics*, 19(12):2356–2365, 2013.

[56] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. *IEEE Transaction on Visualization and Computer Graphics*, 19(12):2366–2375, 2013.

[57] T. Slocum. *Thematic cartography and visualization*. Prentice Hall, 1999.

[58] S. S. Steven. On the psychophysical law. *Psychological Review*, 64(3), 1957.

[59] M. Teghtsoonian. The judgment of size. *The American journal of psychology*, pages 392–402, 1965.

[60] W. Tobler. Thirty five years of computer cartograms. *Annals of Association of American Geographers*, 94:58–73, 2004.

[61] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, U.S.A., 1983.

[62] J. A. Tyner. *Principles of Map Design*. Guilford Press, 2010.

[63] M. van Kreveld and B. Speckmann. On rectangular cartograms. *Computational Geometry*, 37(3):175–187, 2007.

[64] C. Ware. *Information Visualization: Perception for Design*. Interactive Technologies. Elsevier Science, 2004.

[65] C. Ware. Quantitative texton sequences for legible bivariate maps. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1523–1530, 2009.

[66] S. Wehrend. Visual cues: Practical data visualization, 1993.

Sabrina Nusrat is a PhD student at the Department of Computer Science at the University of Arizona. In 2012 she obtained a Masters degree in Computer Science from the University of Saskatchewan. Her current research interest is in visualization and visual analytics, with a focus on geo-referenced visualization.

Md. Jawaherul Alam received his PhD in Computer Science from the University of Arizona in 2015. He is now working as a software development engineer at Amazon, USA. His research interests include algorithms for graphs and maps, graph drawing, and information visualization.

Carlos Scheidegger is an Assistant Professor at the Department of Computer Science at the University of Arizona. He holds a PhD from the University of Utah, where he worked on software infrastructure for scientific collaboration. His research interests are in large-scale data analysis, information visualization and, more broadly, what happens "when people meet data".

Stephen Kobourov is a Professor at the Department of Computer Science at the University of Arizona. He received a BS degree in Mathematics and Computer Science from Dartmouth College and MS and PhD degrees from Johns Hopkins University. His research interests include information visualization, graph theory, and geometric algorithms.