[CV-04-012] Multivariate Mapping

Abstract

Bivariate and multivariate maps encode two or more data variables concurrently into a single symbolization mechanism. Their purpose is to reveal and communicate relationships between the variables that might not otherwise be apparent via a standard single-variable technique. These maps are inherently more complex, though offer a novel means of visualizing the nuances that may exist between the mapped variables. As information-dense visual products, they can require considerable effort on behalf of the map reader, though a thoughtfully-designed map and legend can be an interesting opportunity to effectively convey a comparative dimension.

This chapter describes some of the key types of bivariate and multivariate maps, walks through some of the rationale for various techniques, and encourages the reader to take an informed, balanced approach to map design weighing information density and visual complexity. Some alternatives to bivariate and multivariate mapping are provided, and their relative merits are discussed.

Keywords: cartography, cognition, contiguity, Gestalt psychology, map design techniques, symbolization

Author & citation

Nelson, J. (2020). Multivariate Mapping. The Geographic Information Science & Technology Body of Knowledge (1st Quarter 2020 Edition), John P. Wilson (ed.). DOI: 10.22224/gistbok/2020.1.5.

This Topic is also available in the following editions:

DiBiase, D., DeMers, M., Johnson, A., Kemp, K., Luck, A. T., Plewe, B., and Wentz, E. (2006). Multivariate Displays. The Geographic Information Science & Technology Body of Knowledge. Washington, DC: Association of American Geographers.

Acknowledgements: The author would like to thank Dr. Robert Roth for the invitation and patient guidance, the reviewers for their thoughtful input, and colleagues at Esri for their support.

Explanation

1. Definitions
2. Introducing Multivariate Mapping
3. Multivariate Structures
4. Examples
5. Practical Considerations
6. Alternatives to Multivariate Mapping
7. Summary

1. Definitions

archetypal legend: A key consisting of the map's multivariate symbols in various rather extreme conditions, annotated with the interesting properties of each.

bivariate mapping: a form of multivariate mapping specific to encoding two data variables into a single product, for the purposes of investigating a relationship.

Chernoff Faces: The technique of encoding multiple data dimensions as varying symbolic features of a human (or humanoid) face, developed by statistician Herman Chernoff.

cognitive load: the effort associated with storing information in working memory while it is processed.

intra-symbol encoding: graphically representing multiple visual variables within a single discrete symbol. Gestalt psychology refers to this more generally as integral encodings.

inter-symbol encoding: graphically representing multiple visual variables concurrently via different geometric dimensions (point centroid, polygon fill, polygon stroke, etc). Gestalt psychology refers to this more generally as separable encodings.

multivariate mapping: a technique of encoding cartographic visual variables from multiple data variables into a single product, for the purposes of investigating a relationship.

natural mapping: when the representative dimension echoes a literal quality of the represented data, like vehicle seat adjustment buttons positioned in the shape of the seat itself.

qualitative data: measures of types, represented as names or categories.

quantitative data: measures of values, expressed as numbers.

redundant cues: encoding a single phenomenon with multiple concurrent visual variables.

small multiples: A series of small data visualization instances varying in a key attribute allowing for easy comparison.

spatial contiguity principle: learning is improved when corresponding words and graphics are presented near rather than far from each other.

working memory: a cognitive system responsible for temporarily holding information available for processing.

2. Introducing Multivariate Mapping

Bivariate and multivariate mapping is a method of thematic mapping that simultaneously encodes two variables (**bivariate**), or more (**multivariate**), into the symbolization of a map. The advantages of multivariate mapping include the increased amount of information conveyed in one representation and the opportunity for the map reader to directly infer

relationships between two or more geographic phenomena. Increasing the amount of information concurrently conveyed in a map, however, generally increases the visual complexity of the map and asks more of the reader. The map maker must balance the benefits and insights afforded by bivariate and multivariate mapping with their expectations of the audience's ability and willingness to engage effectively with the map.

3. Multivariate Structures

3.1 Inter-symbol encoding

Perhaps the most straightforward bivariate technique is simply to show two data variables concurrently, each with a distinct geometric channel via a thematic map combination. Figure 1, for example, uses a graduated point symbol to represent healthcare costs per county while the county's fill color represents the proportion of the population that is uninsured. These two representations could well exist independently of each other, but their appearance in a single map invites comparison. Inter-symbol encoding can also be understood as "thematic map combinations."

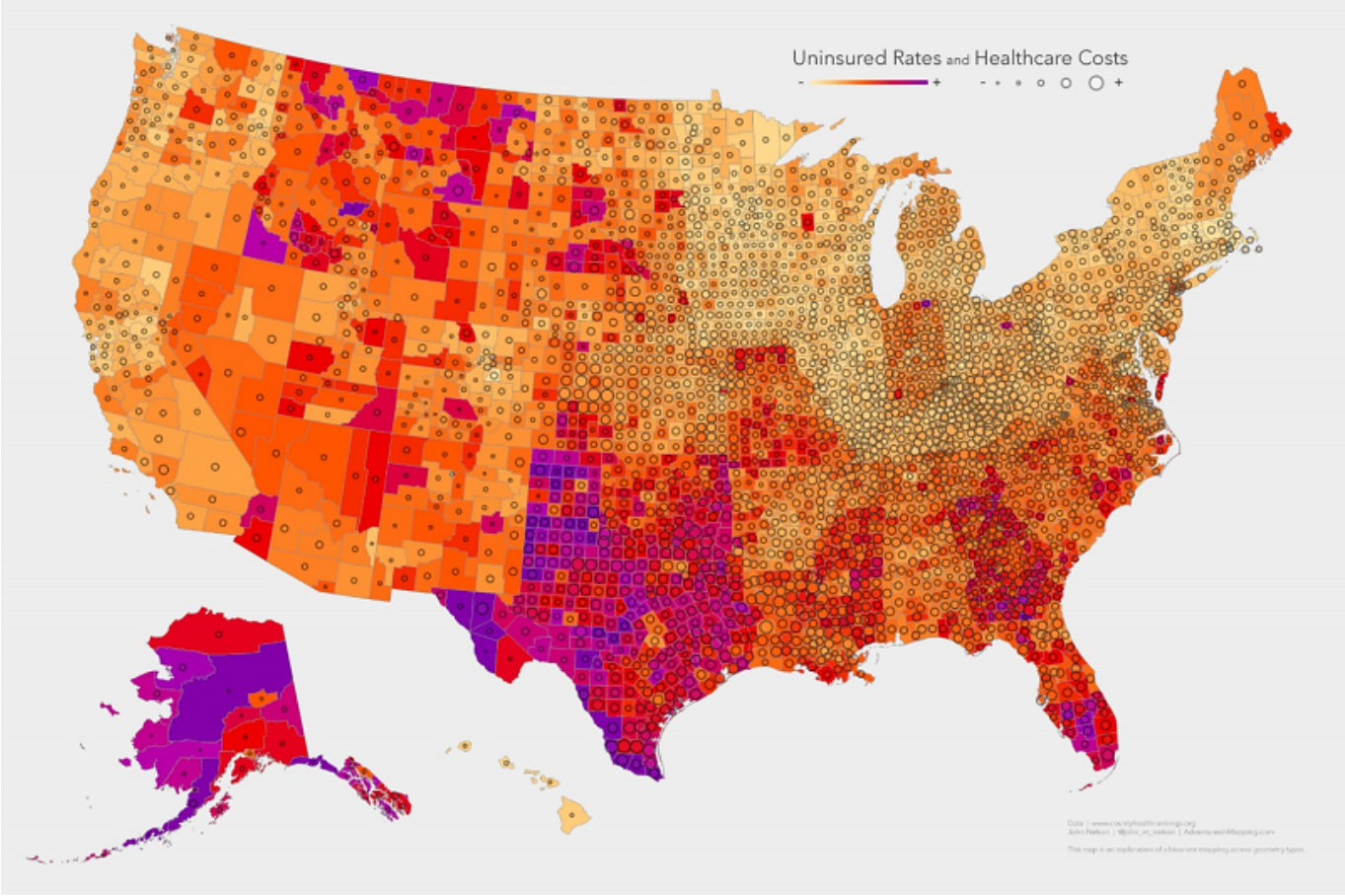


Figure 1: An inter-symbol presentation of two visual dimensions, each tied to a unique geometric channel (polygon fill and point centroid, in this case). Source: author.

3.2 Intra-Symbol Encoding

The map maker can manipulate multiple visual variables (see **Symbolization & the**

Visual Variables) of a single feature type to achieve **intra-symbol encoding** (Nelson, 2000). The work of cartographers of the past is a rich resource for innovative multivariate devices. Figure 2, an environmental map from 1852, encodes wind direction and strength, precipitation, and season via multivariate point symbols. Currents are shown as line features, encoding strength, temperature, season, and month. This, in the context of coastal and bathymetric features. One might note that this map is actually two multivariate symbol layers, wind and currents, shown concurrently—an inter-symbol coupling of intra-symbol encodings. The result is a data-dense product that requires a thorough description of symbolization and some determination on the part of the map reader.

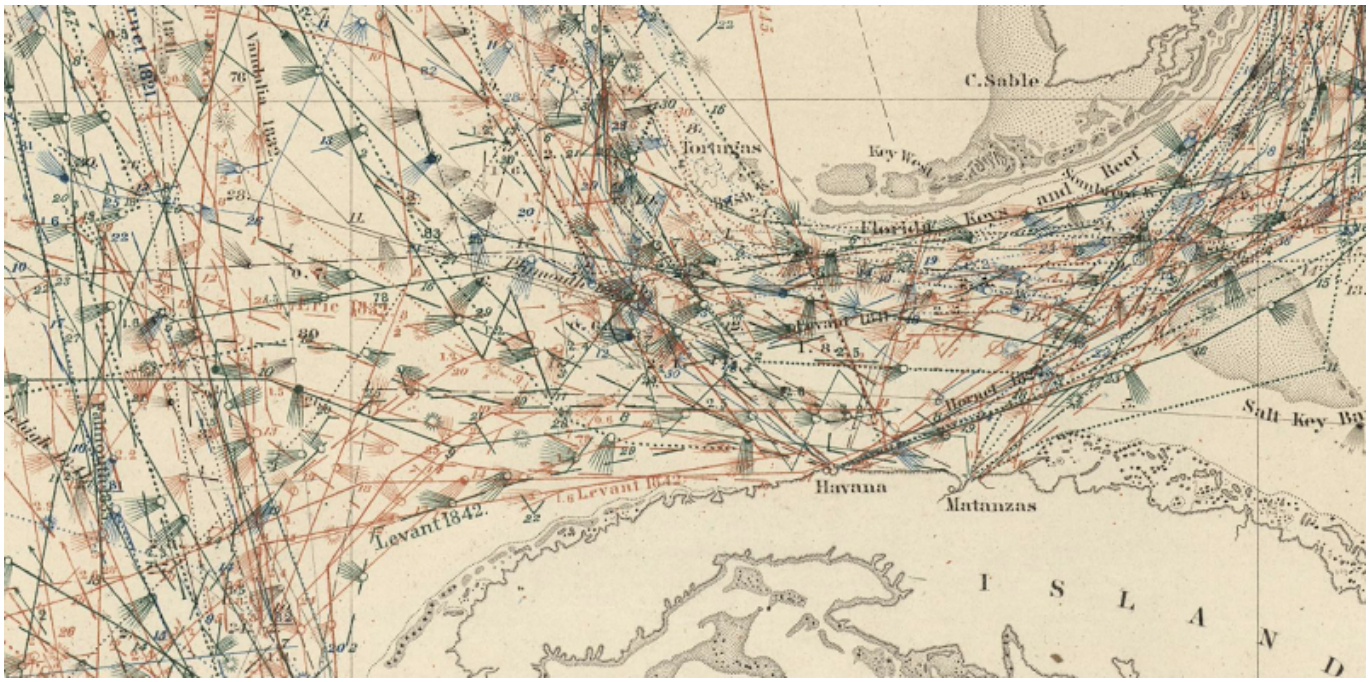


Figure 2: An excerpt from Matthew Fontaine Maury's Wind and Current Chart of the North Atlantic. Source: the David Rumsey Map Collection.

4. Examples

4.1 Bivariate Choropleths

Perhaps the most prevalent example of bivariate mapping is the bivariate choropleth, a method that blends two color ramps into a single choropleth map that varies color hue and color value. Consider the following contiguous United States county-level choropleth maps showing rates of low birth weight (Figure 3) and rates of obesity (Figure 4).

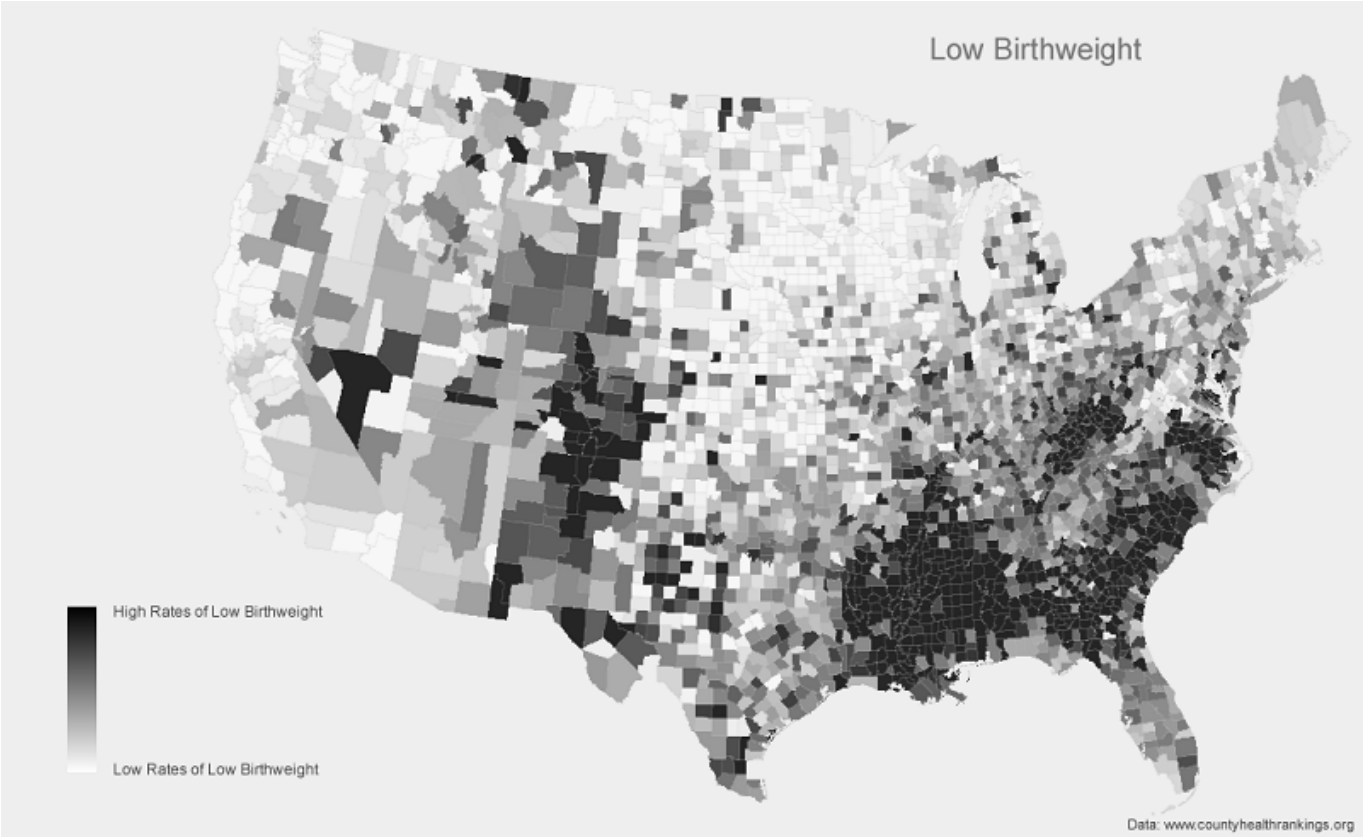


Figure 3: Low birthrate represented as a choropleth. Source: author.

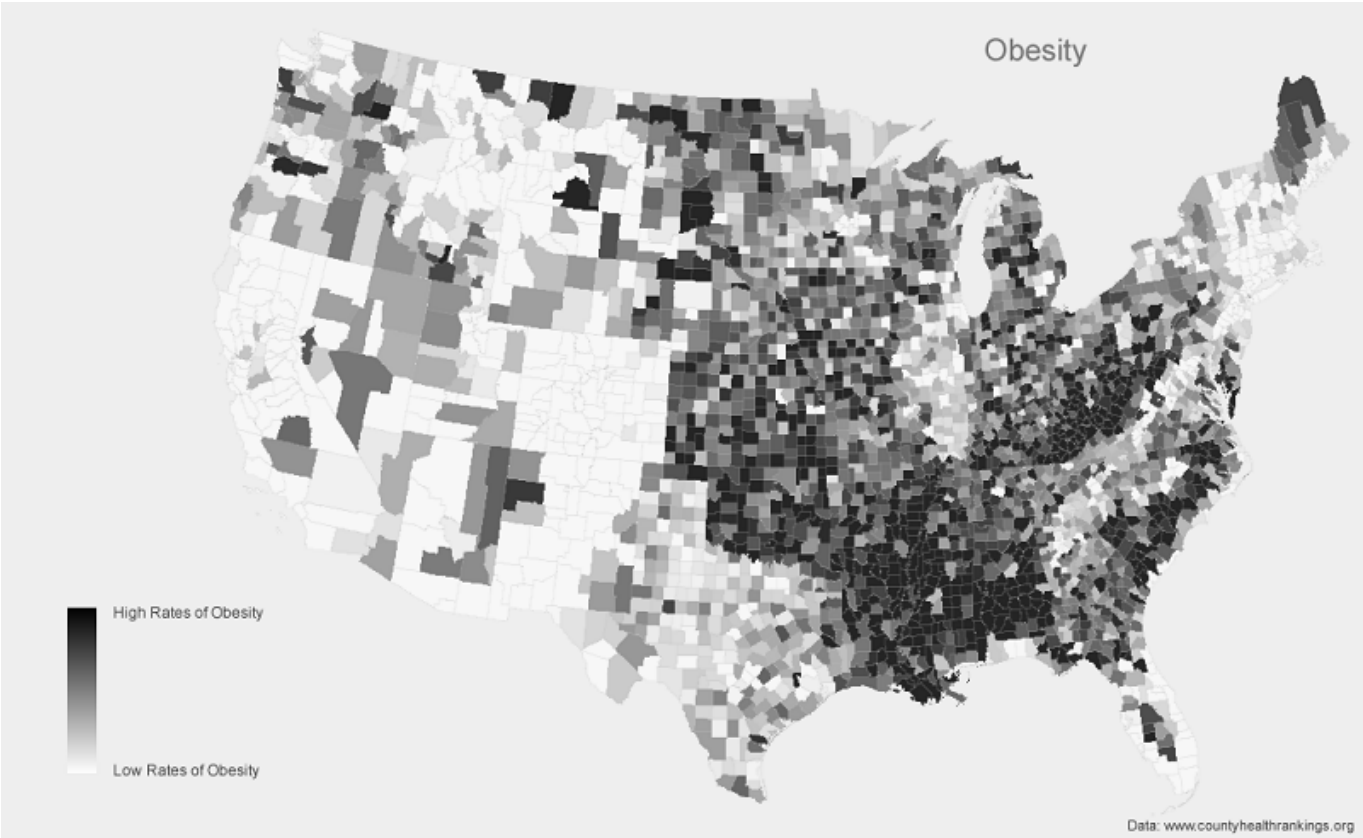


Figure 4: Obesity represented as a choropleth. Source: author.

It would be asking much of a reader to visually compare the geographic nature of these two variables via glancing between these two distinct maps presented adjacently. Visualizing the relationship of two variables across two separate visualizations invites errors of registration between visual anchors in each, requires the working memory to retain an impression of one map while viewing the other, and in general fails to tease out the finer relationships between the two (MacEachren et al. 1998).

4.1.1 Bivariate Relationships

When interpreting a bivariate map of variables A and B, the reader can conceptualize four distinct groups, with optional intermediate ranges between these groups (Elmer, 2012). They are:

- High A and High B
- High A and Low B
- Low A and High B
- Low A and Low B

Take, for example, the bivariate choropleth shown in Figure 5, which uses a matrix of colors to represent two distinct variables: Low Birth weight and Obesity. Low birth weight is represented as light (low) to dark (high) values of a blue hue, while Obesity is represented as light (low) to dark (high) values of an orange hue. The resulting matrix of mixed colors creates a visual product whereby a map reader can identify areas with high rates of both variables, low rates of both variables, and areas with an asymmetric relationship between the two mapped variables.

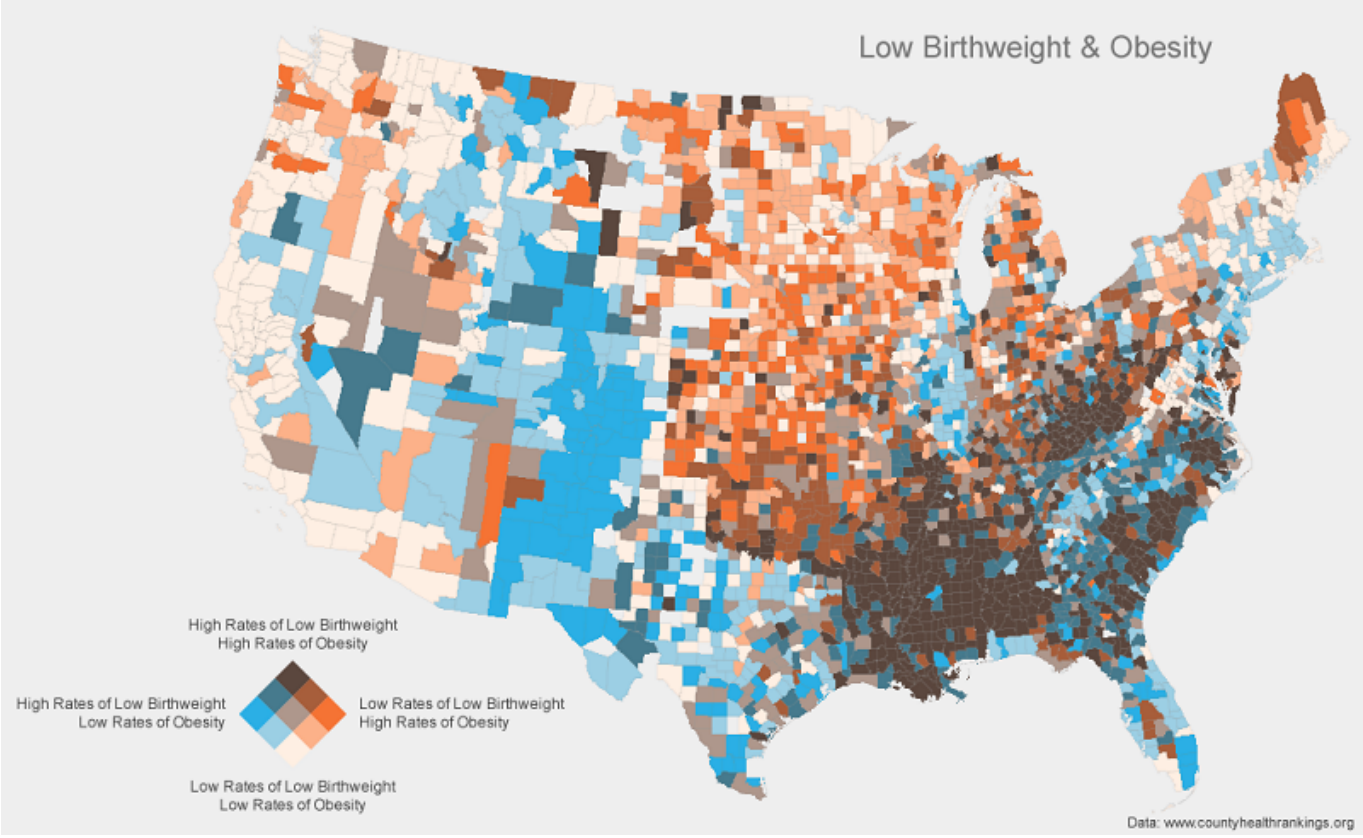


Figure 5: Low birthweight and obesity represented together as a single bivariate choropleth map. Source: author.

A bivariate map like this allows map readers to immediately see where the variables of low birth weight and obesity are consistent with each other (Figure 6), and where they differ (Figure 7). While this map is a more complex product than either of the maps individually, it is a tool that encourages direct and categorized visual comparison, availing the map reader of new dimensions of discovery and questioning.

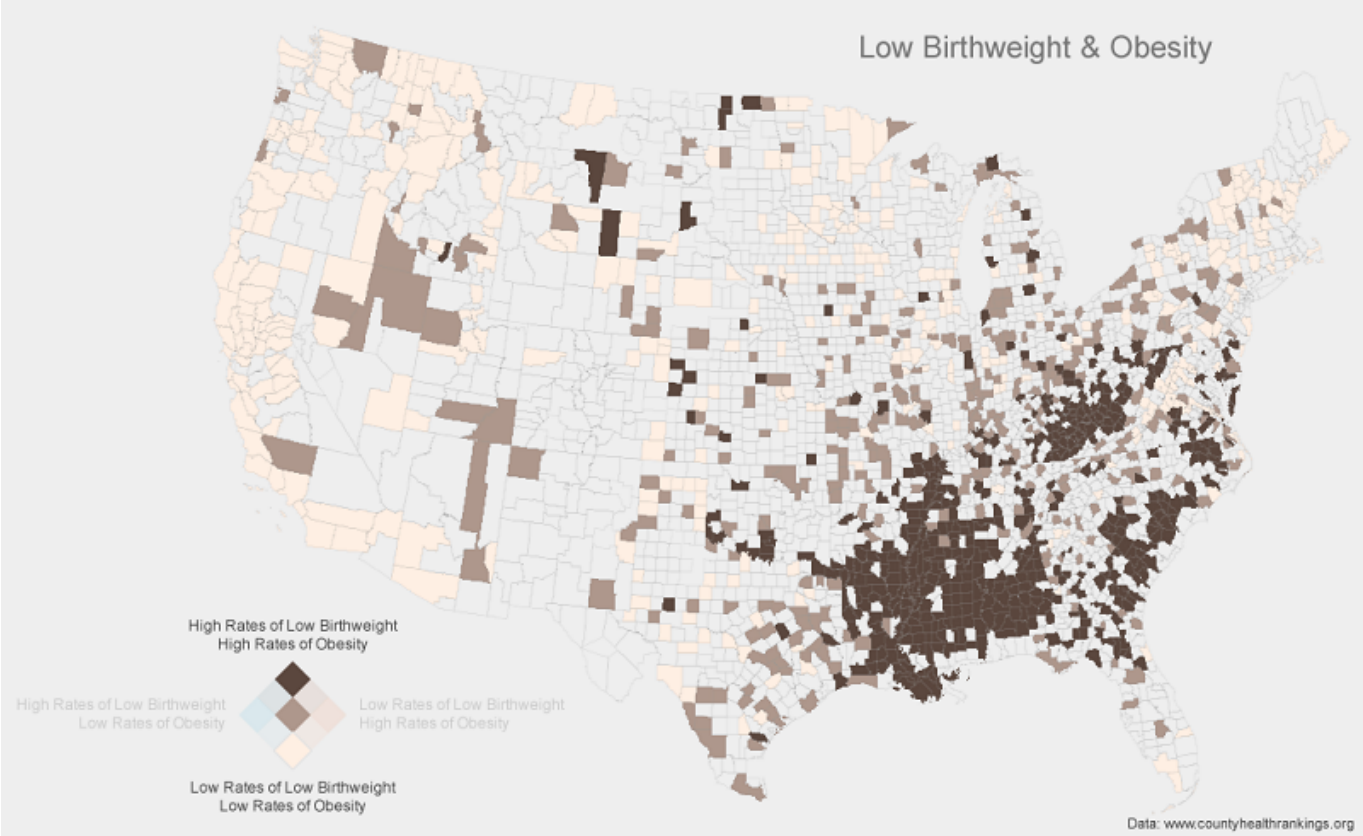


Figure 6: The bivariate choropleth map, highlighted to illustrate the overall strength of the variable. Source: author.

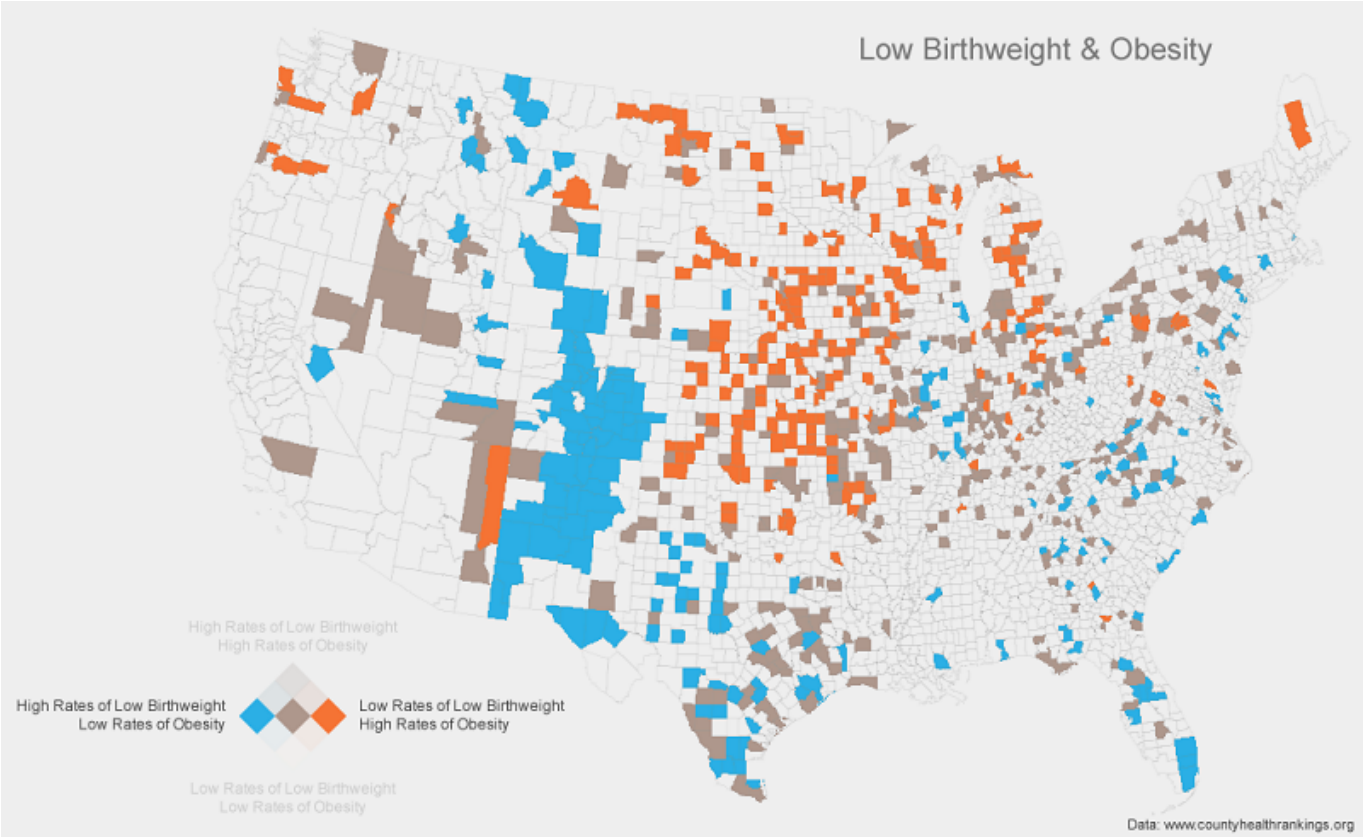


Figure 7: The bivariate choropleth map, highlighted to illustrate the relative strength of the variable. Source: author.

4.1.2 Bivariate Normalization

A popular method of communicating the political landscape of election maps has been a “value-by-alpha” (Roth et al., 2010) technique whereby hue denotes a comparative proportion, the proportion of Republican versus Democrat votes for example, while opacity is keyed to the overall number of votes, effectively normalizing the map by population (see Statistical Mapping). Figure 8 illustrates another sort of vote. In this example, hue is keyed to the difference in per capita spending on beer (amber) versus wine (purple) while opacity is keyed to overall spending. This map shows preference as well as overall engagement.

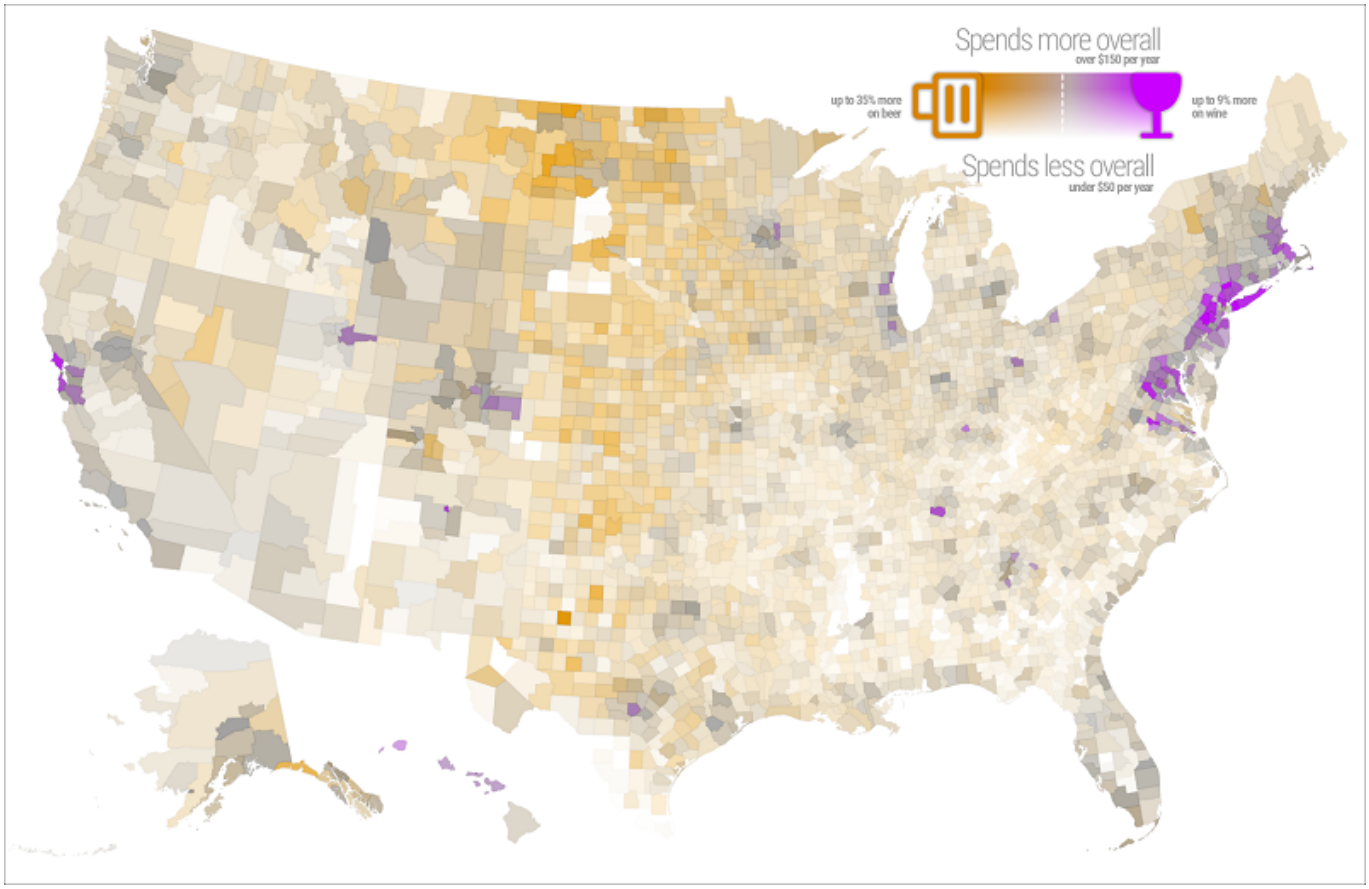


Figure 8: A value-by-alpha bivariate map showing beer or wine preference and overall spending. Source: author.

4.2 Trivariate Choropleth Maps

Taking this another step, the map maker can employ a third color to represent a third variable. This is the basis of process printing's use of halftones of cyan, magenta, yellow, and a key (usually black) colored ink to blend into a full colored visual (Sharma 2003). The result is a visually complex product that may require a more practiced effort on behalf of the reader. While the variate nomenclature could be extrapolated to call this technique "trivariate" mapping, any "variate" beyond two is generally known as "multivariate" mapping.

For example, rates of smoking, obesity, and excessive drinking are represented via graduated opacities of cyan, magenta, and yellow in Figure 9. The overall strength of each constituent hue is indicative of the presence of the phenomena and their relative mixing into derivative colors indicates a three-part nature of their relationship. With some effort, the map reader can potentially identify areas of strong or weak individual concentration and the blending of these variables.

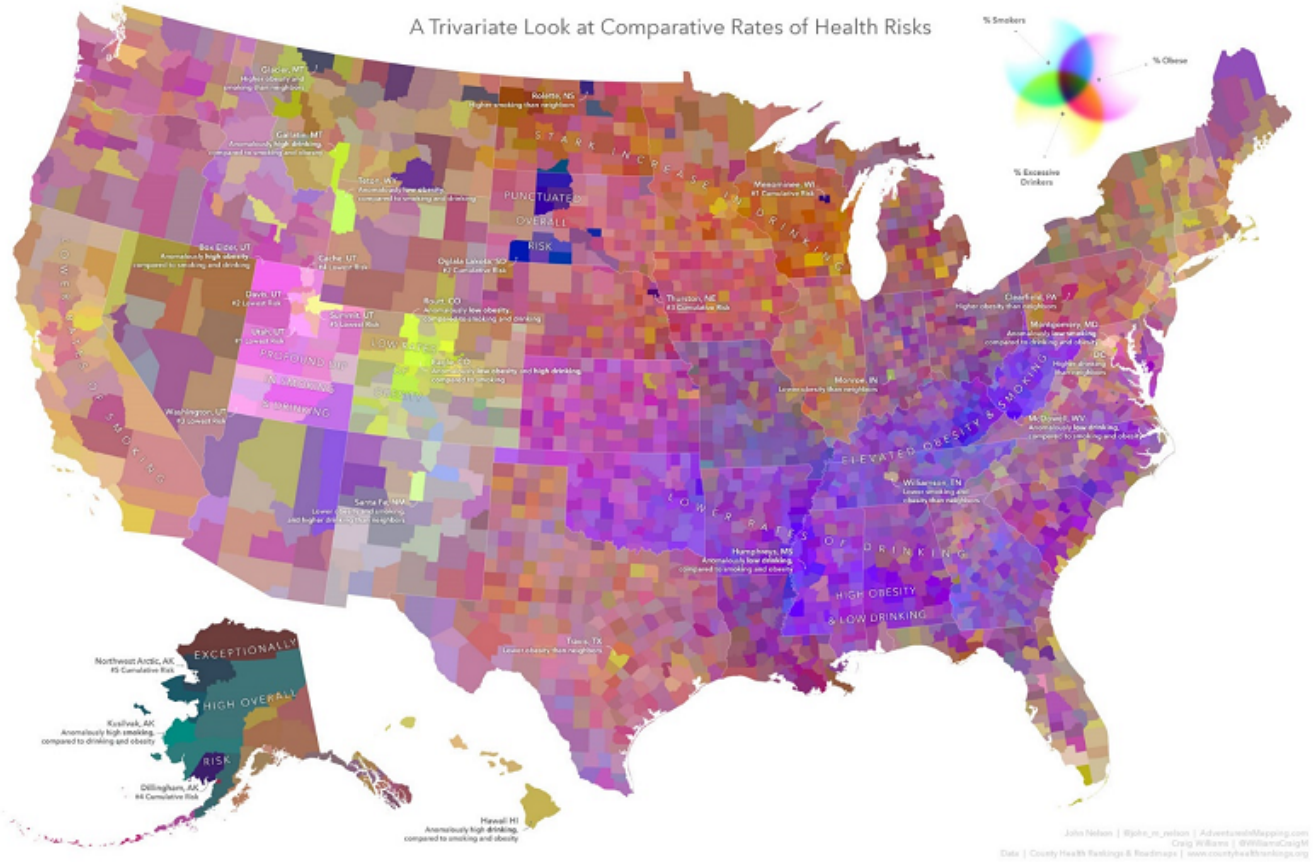


Figure 9: A trivariate choropleth using variable rates of three constituent hues, to concurrently represent the relative prevalence of three phenomena via blending, in the same way cyan, magenta, and yellow inks contribute to the impression of a colored image in process printing. Source: author.

A general risk of these approaches is they produce a potentially intimidating visual if the concept of three-part color mixing new to the map reader. Each additional variable adds to the complexity of the map and the cognitive load required of the map reader to confidently process and interpret the growing array of results. The three-color multivariate map is also inherently inaccessible to those with color deficiencies (see Color Theory). For these reasons, generous descriptive labelling within the map is particularly helpful.

4.3 Coincident Multivariate Symbolization

The map maker can consider visual variables in addition to color to utilize additional resources of the human visual system. In the example of Figure 10, the cumulative intensity of tornado activity is represented as graduated circle size while the ring is colored to denote month-of-year. Scale is an effective means of showing quantity, a natural mapping of a quantity to size.

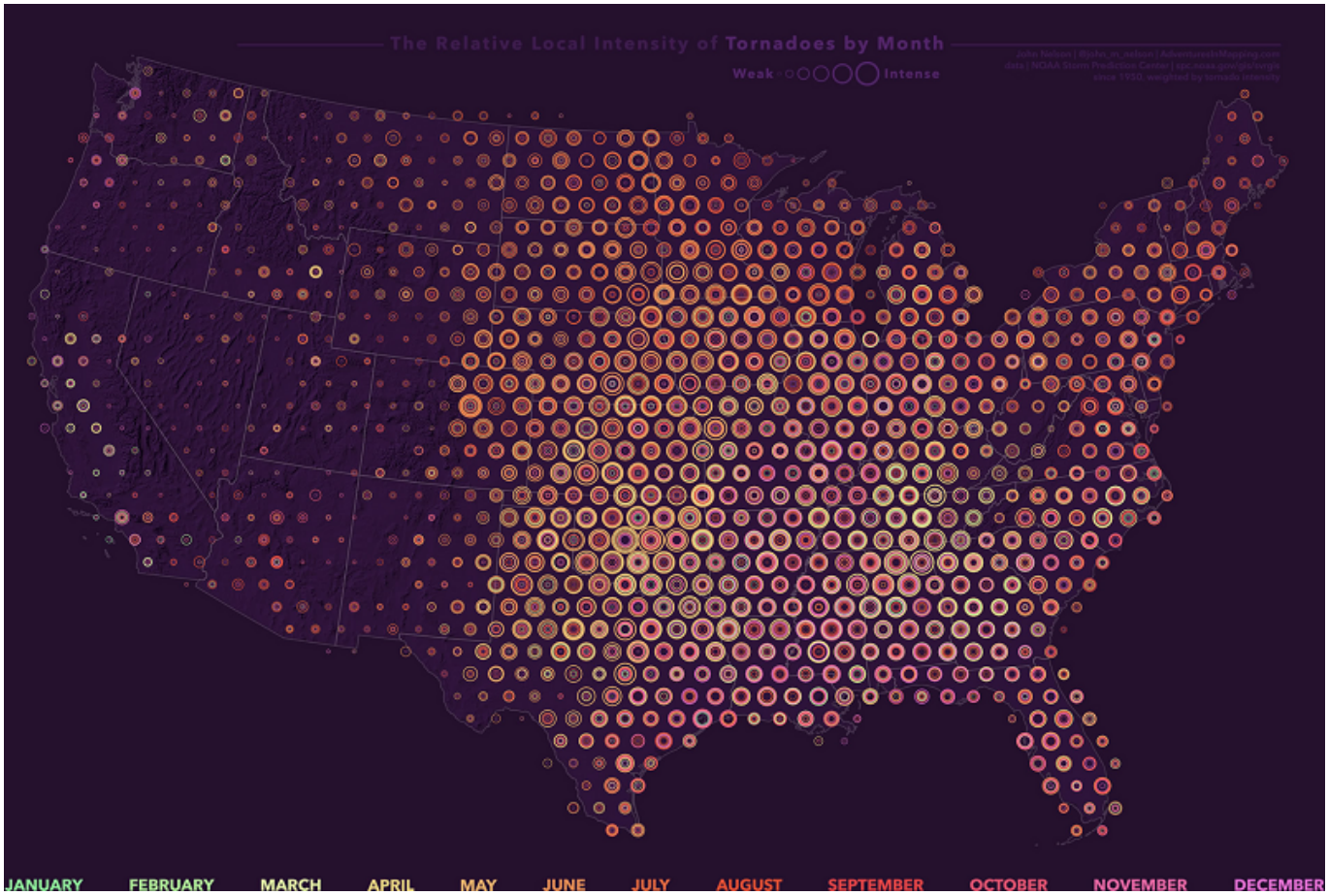


Figure 10: A graduated symbol map of seasonal tornado intensity. Source: author.

When twelve instances of these bivariate symbols are placed concurrently, for each month of the year, their cumulative footprint also serves as a derivative visual dimension of density. The map reader may, with some familiarization, benefit from a local, as well as regional, impression of seasonal activity.

4.4 Chernoff Faces

Herman Chernoff, a statistician seeking to harness the human mind's remarkable ability to intuit minute facial expression components into a holistic impression of nonverbal communication, theorized that statistical information could be represented via cartoonish facial components to communicate large amounts of information and the subtle variations within (Bruckner, 1978). This concept can be applied to maps, with debatable communicative efficacy (Elmer, 2013). The Chernoff faces in Figure 11 encode the proportion of six statistical categorizations of happiness factors at the national level, as enumerated in the United Nations' World Happiness Report.

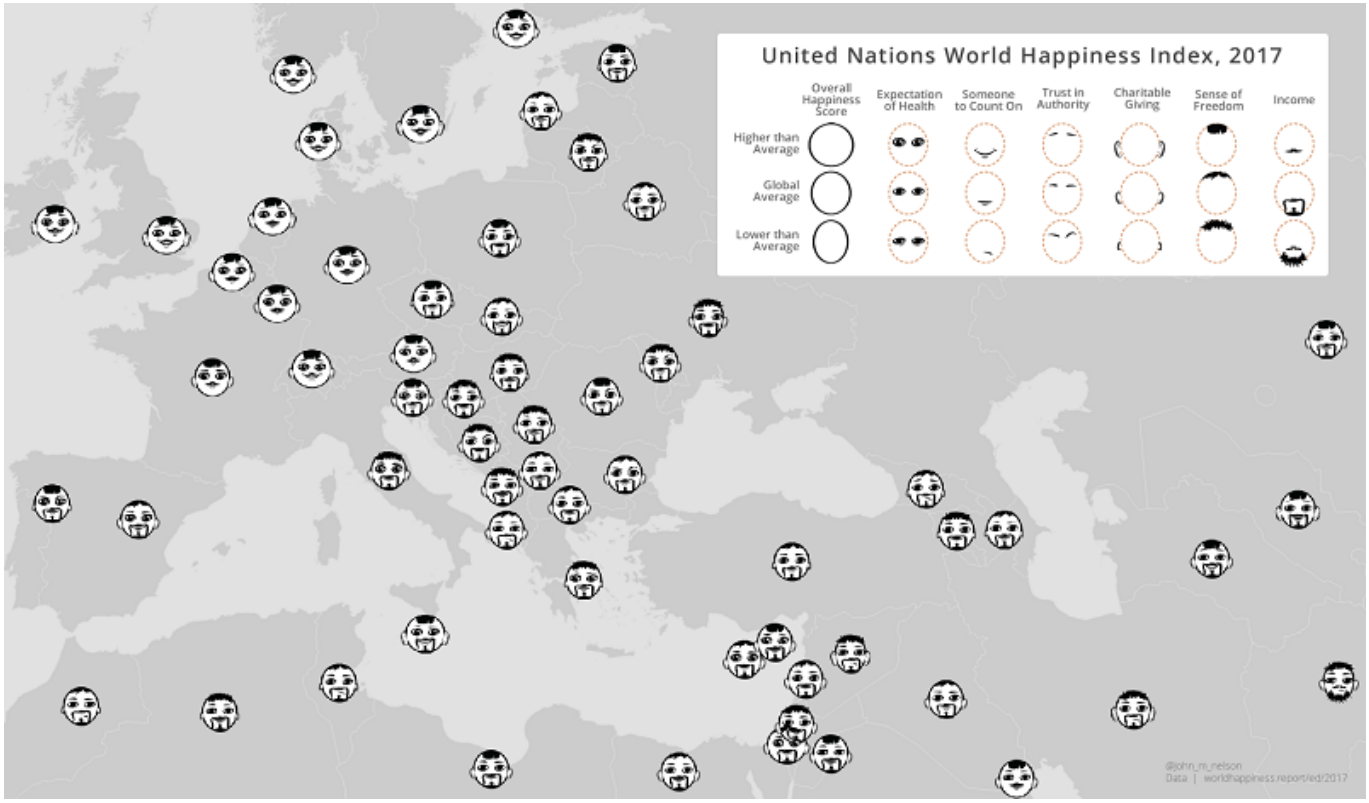


Figure 11: A Chernoff map showing overall happiness and the relative strength of six happiness influences. Source: author.

Because some amount of **natural mapping**, when the representative dimension echoes a literal quality of the represented data (Norman, 1988), can be applied to the Chernoff components, there is a presumed ease of interpretation for some happiness factors. For example, the eyes appear along a range of healthiness commensurate with the health scores of the underlying data. While literal tie-ins to the data can be convenient conveyance, they often stray into realms of cuteness or metaphor (larger ears to represent an awareness of the needy, for example) and any intrinsic value of the symbol is lost. Furthermore, their presumed benefit—our ability to detect minute variations in the human likeness and ascribe meaning therefrom—can invite interpretations of incidental expressions that distract or mislead.

5. Practical Considerations

5.1. Legends

Providing a key by which the map reader can interpret a bivariate or multivariate map is as important as creating an effective symbology scheme itself. Multivariate maps are more complex devices, elevating the importance of an effective legend.

A common, and technically straightforward, approach segregates the visual variables and provides a discrete key for each (Figure 12). The map reader must couple these dimensions in their mind and apply their mental image of the symbolization as they interpret the map (or they focus on any one individual data dimensions of the map at a time).



Figure 12: A segregated legend depicting a bivariate scheme of graduated, colored, circles. Source: author.

A popular approach for bivariate choropleth maps that rely solely on color is an array of color values, where the axes are labeled for their representative variable (Figure 13). This approach benefits from the concise presentation of variables and values. Alternatively, the map maker might instead label the extreme corners of the bivariate relationship, leapfrogging the abstraction of variable names and ranges in favor of a more literal description of the relationship, confident the map reader can intuit the nature of ranges and intermediate values. Note also how the color array is rotated so that higher values orient to “up”.

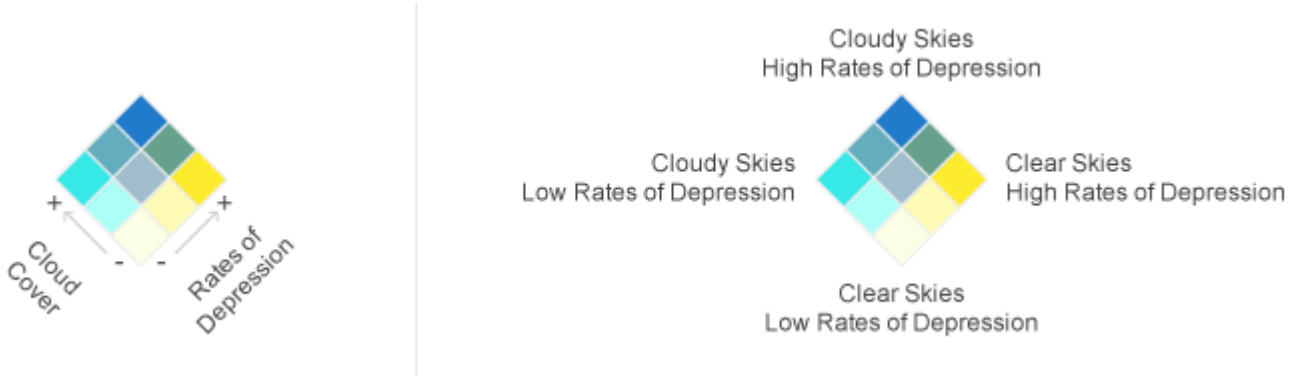


Figure 13: A popular color array legend approach for bivariate choropleth maps. Variable and range labeling at left and comparative relationship labeling at right. Source: author.

The map maker must consider what approach best balances efficiency and communication. In some cases, a hybrid approach that labels variables and their corners may suit the map best. The map legend in Figure 14 takes a rather verbose approach, naming data variables and ranges, visual variables, and labeled relationships.

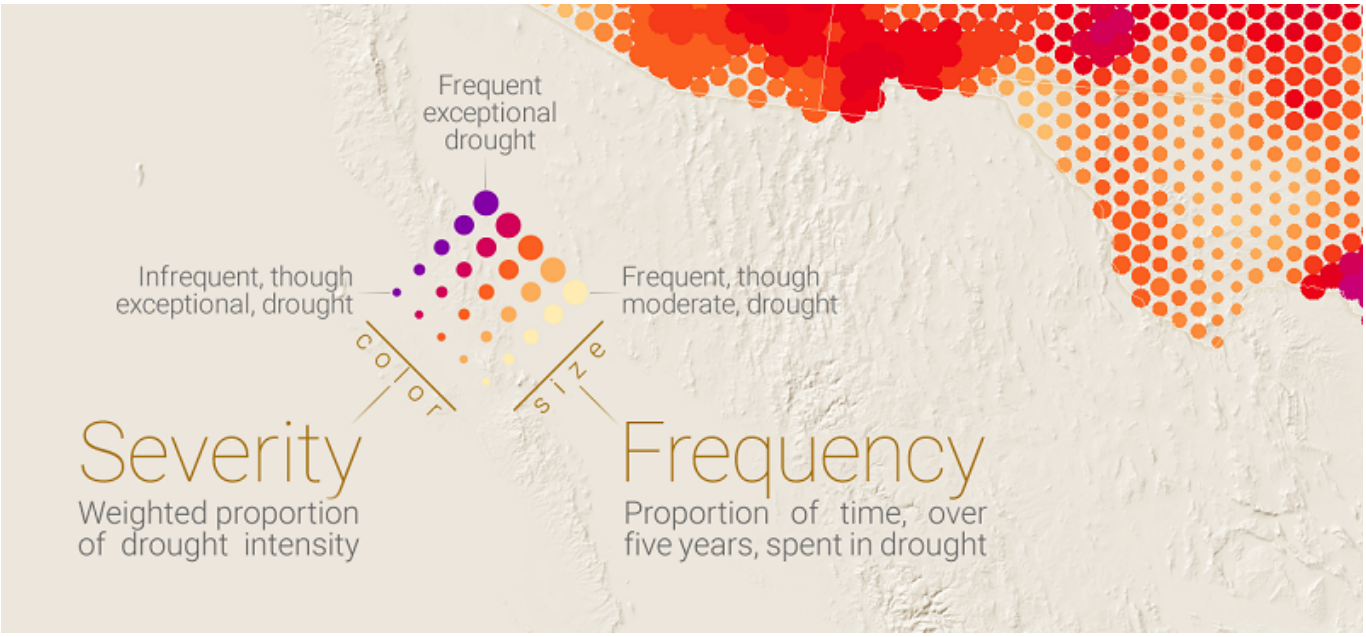


Figure 14: A bivariate legend keying data variables and ranges, visual variables, and relationship extremes. Source: author.

Multivariate maps are laden with information and potential insights. An extension of the comparative labeling legend strategy described in Figure 14 is to focus on those insights as for instance examples. Annotating exemplary symbols directly within the map can offer an effective means of illuminating dimensionally rich symbolization, while also benefitting from the principle of **spatial contiguity**, the idea that we learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen (Moreno & Mayer, 1999). An extension of this conversational approach can be to include an **archetypal legend**, a small set of notable symbol conditions with a brief description of the interesting character of each combination, like those in Figure 15, which correspond to the map in Figure 16.

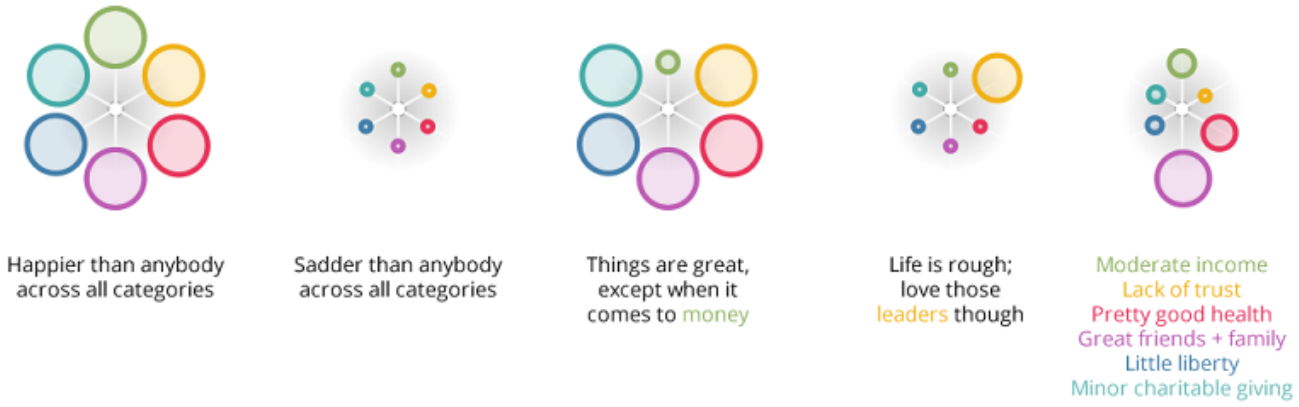


Figure 15: An annotated set of archetypal symbols, whether in a dedicated legend, or called out within the map itself, can serve as a conversational conveyance of the phenomenon. Source: author.

5.2. Redundant Cues

Because multivariate maps are inherently more information-dense products then their single-variable counterparts, **redundant cues** (when more than one visual dimension is used to encode the same data variable) can be used to offer the reader more opportunities for understanding. For example, Figure 16 shows six variables (the same six happiness factors described in Figure 11's Chernoff Faces). Each symbol comprises six circles oriented radially around a center point. Color and offset orientation are redundant cues representing categorical affiliation—two visual variables denoting the same variable. Additionally, the strength of each component is represented by the symbol size and offset distance. Both variables (category and strength) are given two visual dimensions. Redundant cues can help reinforce the visual interpretation of a thematic variable and can be particularly accommodating to those with color deficiencies (Ware, 2004).



Figure 16: A multivariate map showing the proportional rate of six happiness influences. Redundant cues offer the reader more visual dimensions for interpretation of category (color and orientation) as well as variable strength (size and offset distance). Source: author.

6. Alternatives to Multivariate Mapping

Multivariate thematic maps are inherently more complex than their single-variable cousins. As such the map maker ought to consider carefully if a multivariate map is altogether the best method in light of other options. If the presentation method allows, the author may consider a narrative presentation of standard single-variable maps, wrapped with additional

interpretation and context provided by the author (see Narrative and Storytelling, forthcoming).

Small multiple maps, a series of small maps of identical area arranged to show a single changing variable for the purposes of comparison, can be an effective visualization method for comparative phenomena (Tufte 1990). The small multiple series in Figure 17 is an alternative to the multivariate map of seasonal tornado magnitude shown in Figure 10.

Animation is another alternative to multivariate mapping, wherein a changing sequence progresses through the univariate elements of an otherwise multivariate map.



Figure 17: A small multiple as an alternative to the multivariate approach. Source: author.

7. Summary

Multivariate maps encode more than one variable into a single map. They are potent tools for presenting information compactly within a single product so the reader can directly explore the relationship between mapped variables. Multivariate maps afford a reduced dependence on working memory, and introduction of potential errors of registration, to the alternative of comparing multiple separate maps.

The inherent complexity of multivariate maps, however, can overwhelm a reader's ability, or inclination, to engage in a meaningful way. The map maker is encouraged to consider carefully if this is a beneficial mechanism. If the map author is mindful of their audience and the nature of their task, is thoughtful in legend design, and strategic in visual variable choices, multivariate maps are uniquely positioned to be an insight-rich communication

tool.

References

Bruckner, L.A. (1978). On Chernoff faces. In P. C. C. Wang (Ed.), Graphical presentation of multivariate data (pp. 93-121). New York: Academic Press.

Elmer, M. (2013). The Trouble With Chernoff. Blog posting. Last accessed December 4, 2023.

MacEachren, A. M., Brewer, C. A., and Pickle, L. W. (1998). Visualizing Georeferenced Data: Representing Reliability of Health Statistics. Environment and Planning A: Economy and Space. 30(9), 1547-1561.

Moreno, R., & Mayer, R.E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. Journal of Educational Psychology, 91, 358-368.

Nelson, E. (2000). The Impact of Bivariate Symbol Design on Task Performance in a Map Setting. Cartographica, 37(4), 61–78.

Norman, D. A. (1988). The Design of Everyday Things. 1st Edition. New York: Basic Books.

Roth, R. E., Woodruff, A. W., & Johnson, Z. F. (2010) Value-by-alpha Maps: An Alternative Technique to the Cartogram. The Cartographic Journal. 47(2), 130-140.

Sharma, G. and Bala, R. (Eds.) (2003). Digital Color Imaging Handbook, 1st Edition. CRC Press, Florida.

Tufte, E. R. (1990). Envisioning Information. Cheshire, CT: Graphics Press.

Ware, C. (2004), Information Visualization: Perception for Design. 2nd ed., Morgan Kaufmann, California.