A Novel Method for the Depiction of Multivariate Data through Flow Maps

Alberto Debiasi* Bruno Simões† Raffaele De Amicis‡

Fondazione Graphitech

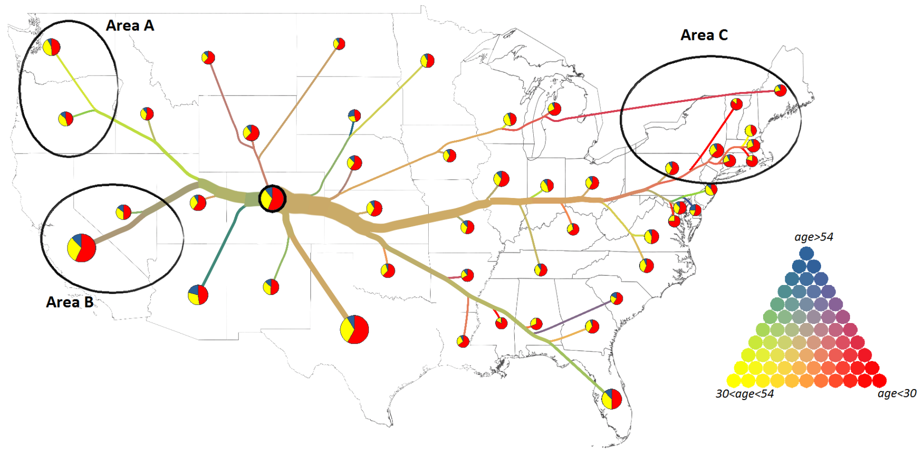


Figure 1: Migratory flows from Colorado, by age groups: people younger than 30 (red), between 30 and 54 (yellow), and over 54 (blue).

ABSTRACT

Quantitative flow maps in cartography are graphical representations that depict the movement of a phenomenon from one location to another. These thematic maps were originally created to depict univariate origin-destination datasets. In this paper, we describe a novel technique for the visualization of multivariate geographical data in flow maps. Our solution crystallizes on color blending techniques and pie charts. To evaluate the relevance and impact of this study, a use case is presented.

Index Terms: I.3.3 [COMPUTER GRAPHICS]: Picture/Image Generation —Line and curve generation; I.3.6 [COMPUTER GRAPHICS]: Methodology and Techniques —Interaction techniques

1 INTRODUCTION

Flow maps depict the movement of phenomena between geographic locations [6]. The movement of something from one location to another is represented as a straight line or curve – flow line.

Flow lines sharing the same destination are aggregated together and their thickness is defined as proportional to the sum of the flows' magnitude they represent – Sankey flow maps. In this paper, we propose a solution that combines the value of color-based techniques, as in choropleth maps, with the aggregation aspect of Sankey flow maps. This method does not require any form of user interaction or content animation as it depicts multivariate information in a single static image. Additionally, this method requires less screen space than small multiples techniques because only one map is needed. The main limitation of our approach is that number of attributes that can be visualized in simultaneous. Nevertheless, this is a limitation that can be tacked by combining colors and shape patterns.

*e-mail: alberto.debiasi@graphitech.it

†e-mail: bruno.simoes@graphitech.it

‡e-mail: raffaele.de.amicis@graphitech.it

2 OUR APPROACH

In this section, we explain how to extend the algorithm for the automatic generation of Sankey Flow map, called *Force Directed Flow Map Layout (FDFML)* [2], with support for the visualization of multiple attributes (without modifying its layout). Our idea capitalizes on enriching the nodes' representation with color schemes.

The original *FDFML* algorithm requires as an input the geographic position of the origin (the root node r of the tree t) and n geographical destinations (corresponding to the leaves l_1, l_2, ..., l_n), respectively with flow magnitudes $magn(l_1)$, $magn(l_2)$, ..., $magn(l_n)$. The output is simply a tree t with additional "intermediate" nodes as well as their flow magnitudes – a list of control points used to generate Cubic Splines.

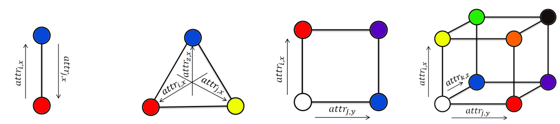
In our extended version of the algorithm, it supports as an input, for each leaf node, a list of complementary attributes, i.e. their union is equal to the whole magnitude. Let $a_{i,j}(l) \in [0,1]$ be the value of the phenomena j of the attribute with index i assigned for each destination l, such that $\sum_{i=1}^{h} a_{i,j}(l) = 1$ where h is at most 3.

The multivariate cases supported are: [C1] Two attributes that compose the whole amount. Each leaf l has two values: $a_{1,x}(l)$ and $a_{2,x}(l)$ which sum is 1. [C2] Three attributes that sum up to the whole amount. Each leaf l has three values: $a_{1,x}(l)$, $a_{2,x}(l)$ and $a_{3,x}(l)$ which sum is 1. [C3] Two independent attributes. Each leaf l has two values: $a_{i,x}(l)$ and $a_{j,y}(l)$ where x and y are different phenomena. [C4] Three independent attributes. Each leaf l has three values: $a_{i,x}(l)$, $a_{j,y}(l)$ and $a_{k,z}(l)$ where x, y and z are different phenomena.

2.1 Color models

This work has its inspiration in choropleth maps and color-based techniques. In particular, colors are used to convey information about the attributes of the geographical data (depicted in the flow map). We use the color technique defined by Gossett et al. [5], based on a subtractive color space, with Red (R), Yellow (Y) and Blue (B) as primary colors.

In C1, we use a R and B color pattern defined by the mapping function $C_1(a_{1,x}, a_{2,x}) -> ryb(255 * a_{1,x}, 0, 255 * a_{2,x})$, see Figure 2a. In C2, we use a triaxial graph composed by three colors: R, Y and B. The mapping function is $C_2(a_{1,x}, a_{2,x}, a_{3,x}) -> ryb(255 * a_{1,x}, 255 * a_{2,x}, 255 * a_{3,x})$, see Figure 2b. The color domain includes only colors that sum up to 255. In C3, we use the square pattern composed by the R and B colors. The color mapping function is $C_3(a_{i,x}, a_{j,y}) -> ryb(255 * a_{i,x}, 0, 255 * a_{j,y})$, see Figure 2c. Finally, in C4 we use the color cube as a color domain and the mapping function $C_4(a_{i,x}, a_{j,y}, a_{k,z}) -> ryb(255 * a_{i,x}, 255 * a_{k,z}, 255 * a_{j,y})$, see Figure 2d.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Figure 2: RYB interpolation patterns.

In some cases users might find difficult to distinguish between range of colors and consequently to perceive which attribute is greater than the others. In our work, as shown in the section 3, each primary color is normalized by taking as interval its minimum and maximum value. In C1, the two attributes have a common range of values. Therefore the formula is:

$norm\_attr_{var,x}(l) = attr_{var,x}(l) - min(attr_{var,x})/(1 - (min(attr_{1,x}) + min(attr_{2,x})))$

In C2 the formula to normalize each attribute is:

$norm\_attr_{var,x}(l) = attr_{var,x}(l) - min(attr_{var,x})/(1 - (min(attr_{1,x}) + min(attr_{2,x}) + min(attr_{3,x})))$

In C3 and C4, the attributes are independent, hence each of them is normalized accordingly with its minimum and maximum value.

2.2 Flow representation

In cartography, the visual primitives used to describe the movement of entities in a flow map are the flow lines. Instead of represent the attributes as parallel lines each of those depicted with a primary color, we decided to keep the single flow representation for each destination and working on its color. In particular, the color of the flow line is the result of the colors that identify its attributes. Hence, the color of each flow line derives from the attributes of the intermediate nodes. To do so we first calculate iteratively their values starting from the leaf nodes in the following way:

For a given attribute, $a_{i,x}(n) = magn(n) / \sum_{m \in child} magn(m) * a_{i,x}(m)$ where *child* is the set of children of the node n. Then the color of each curve segment is calculated using the appropriate mapping function, as described in the subsection 2.1, taking as input the attributes of the end point of the segment, that can be an intermediate or a leaf node.

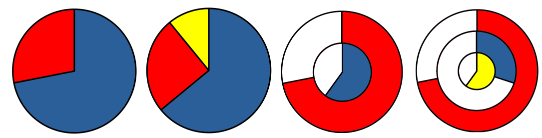
2.3 Nodes representation

Flow map end nodes – used to depict destinations – can be optionally extended with additional visual information about the node. For example, its size can be proportional to the magnitude of the related destination.

The pie chart is a circular graph that breaks down a total amount into more detailed slices of information, illustrating numerical proportion. Although this representation is not always suitable, see [3, 4], in this case it can enrich our node representation:

(1) Nodes are usually depicted as circle, thus a pie chart can be easier to visualize without changing the layout of the flow map. (2) The number of attributes is at most three, thus this representation is not misleading. (3) Percentage ratios are easy to understand. (4) Rethinks the space area used to represent node representation.

Each leaf node is depicted as a pie chart where each slice identifies the amount of entities moving to that destination. Depending on the number of attributes to depict, different variants of pie chart are used, as shown in Figure 3.



(a) Case 1 (b) Case 2 (c) Case 3 (d) Case 4

Figure 3: Different node representations using pie charts.

3 Evaluation

We demonstrate our method depicting some data from County-to-County Migration Flows dataset [1]. In order to emphasize the differences between the attributes, their values are normalized following the procedure described in section 2.1, thus the color does not identifies exactly the percentages of each attributes.

In Figure 1 it is possible to know the age of the people moving from the Colorado. By looking at the pie charts, it is possible to extract the percentage of people for each group. Meanwhile looking to the flows it is possible to understand the same information but also for aggregated destinations. For example looking at the main split in the west of the Colorado, the flow directed to north (area A) has more migrants with age between 30 and 54 respect to the migrants directed in California and Nevada (area B) that have more young people. The migrations directed in the north east (area C) are characterized with high percentage of young people.

4 Conclusion

A color scheme blending is used in conjunction with the aggregation aspects of flow maps and pie charts, to visualize multi-attribute information. Each primary color corresponds to an attribute. The color compositing technique works only if the viewer is able to discern the individual components that combine the final color image [5]. Further work requires an exhaustive evaluation study to validate our work.

References

[1] U. C. Bureau. County-to-county migration flows, 2012.

[2] A. Debiasi, B. Simões, and R. De Amicis. Supervised force directed algorithm for the generation of flow maps. In *Proceedings of the WSCG 2014 - 22nd International Conference on Computer Graphics*, 2014.

[3] S. Few. Show me the numbers. *Analytics Pres*, 2004.

[4] S. Few. Save the pies for dessert. *Perceptual Edge Visual Business Intelligence Newsletter*, pages 1–14, 2007.

[5] N. Gossett and B. Chen. Paint inspired color mixing and compositing for visualization. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 113–118. IEEE, 2004.

[6] T. A. Slocum, R. B. McMaster, F. C. Kessler, and H. H. Howard. *Thematic cartography and geovisualization*. Pearson Prentice Hall Upper Saddle River, NJ, 2009.