European Journal of Operational Research 000 (2017) 1–13

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

On Mathematical Optimization for the visualization of frequencies and adjacencies as rectangular maps

Emilio Carrizosa[a], Vanesa Guerrero[a,*], Dolores Romero Morales[b]

[a] Instituto de Matemáticas de la Universidad de Sevilla (IMUS), Seville, Spain
[b] Copenhagen Business School, Frederiksberg, Denmark

ARTICLE INFO

Article history:
Received 11 March 2016
Accepted 6 July 2017
Available online xxx

Keywords:
Mixed Integer Linear Programming
Visualization
Multidimensional Scaling
Rectangular maps
Frequencies and adjacencies

ABSTRACT

In this paper, we address the problem of visualizing a frequency distribution and an adjacency relation attached to a set of individuals. We represent this information using a rectangular map, i.e., a subdivision of a rectangle into rectangular portions so that each portion is associated with one individual, their areas reflect the frequencies, and the adjacencies between portions represent the adjacencies between the individuals. Due to the impossibility of satisfying both area and adjacency requirements, our aim is to fit as well as possible the areas, while representing as many adjacent individuals as adjacent rectangular portions as possible and adding as few false adjacencies, i.e., adjacencies between rectangular portions corresponding to non-adjacent individuals, as possible. We formulate this visualization problem as a Mixed Integer Linear Programming (MILP) model. We propose a matheuristic that has this MILP model at its heart. Our experimental results demonstrate that our matheuristic provides rectangular maps with a good fit in both the frequency distribution and the adjacency relation.

1. Introduction

It is critical to enable analysts to observe and interact with data, using appropriate visualization tools (Keim et al., 2008; Liu, Cui, Wu, & Liu, 2014). Operations Research arises as a powerful area of knowledge to give answers to new challenges in Visualization (Mortenson, Doherty, & Robinson, 2015; Olafsson, Li, & Wu, 2008).

A natural and frequent task is to depict a set of individuals $V = \{v_1, \ldots, v_N\}$, to which there is attached a frequency distribution, $\omega = (\omega_1, \ldots, \omega_N)$, with $\sum_{r=1}^{N} \omega_r = 1$, see, e.g., Spence and Lewandowsky (1991). Market share, vote intention or population rates, just to name a few, are usual examples. In order to visualize frequencies, a common approach is to consider a bounded region of the plane and to subdivide it into portions $\mathbf{P} = (P_1, \ldots, P_N)$ of common shape whose areas represent the frequencies. Well-known visualization tools for this kind of data are the classic pie or fan charts, Fig. 1(a) and (b) respectively, and rectangular maps (Baudel & Broeksema, 2012; Heilmann, Keim, Panse, & Sips, 2004), see Fig. 1(c) and (d). In this kind of representations, holes are not allowed, thus, receiving the name of planar space-filling visualization maps.

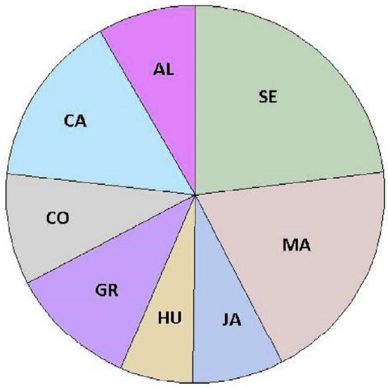
A planar space-filling map to visualize the frequencies attached to individuals in a bounded set Ω of the plane can be found by constructing the portions of the desired area and putting them together to fill Ω. This is straightforward in the case of the pie or fan charts: for a permutation $\sigma(1), \sigma(2), \ldots, \sigma(N)$ of the indices $1, 2, \ldots, N$, portions of areas proportional to $\omega_{\sigma(1)}, \omega_{\sigma(2)}, \ldots, \omega_{\sigma(N)}$ are placed sequentially in Ω. The only freedom in such planar space-filling visualization maps is thus the choice of the permutation, which can be made according to different *seriation* criteria as exposed in Hahsler (2017). For the case of rectangular maps, where Ω is the unit square and portions are rectangles, the same approach, illustrated in Fig. 1(c), can be used, where the rectangular portions go all the way from North to South (or, by rotation, from West to East, for instance). While pie and fan charts only admit different sequential arrangements, rectangular maps allow more freedom than the choice of a permutation, as illustrated in Fig. 1(d). The flexible layout offered by rectangular maps is also desirable when, in addition to frequencies, we are interested in visualizing proximity, measured by adjacencies, which is the subject of this paper.

The nature of the proximity can be diverse, a classical example being geographical proximity. A well-known problem in Cartography is the representation of geographical regions with relatively simple shapes, such as rectangles, whose areas represent a magnitude such as population rates or vote intention, as well as the relative position between regions is maintained,

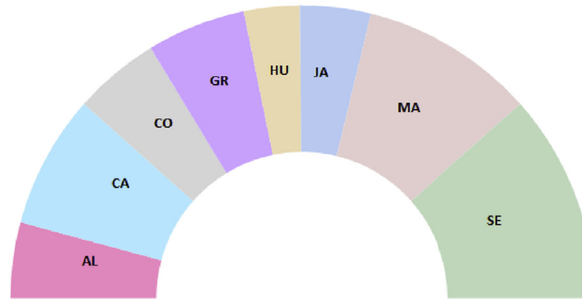
* Corresponding author.
E-mail addresses: ecarrizosa@us.es (E. Carrizosa), vguerrero@us.es (V. Guerrero), drm.eco@cbs.dk (D. Romero Morales).

http://dx.doi.org/10.1016/j.ejor.2017.07.023

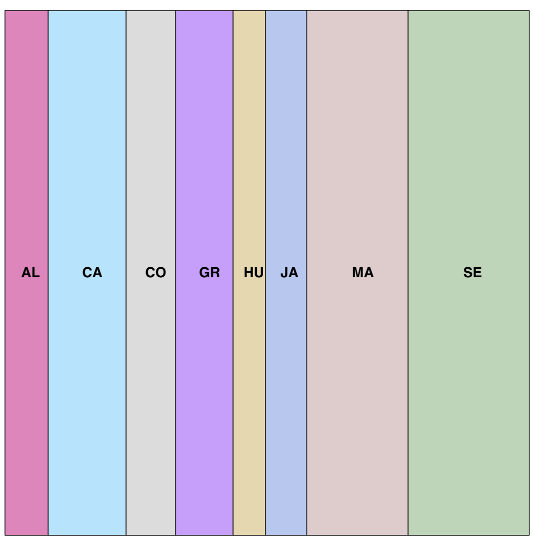
Please cite this article as: E. Carrizosa et al., On Mathematical Optimization for the visualization of frequencies and adjacencies as rectangular maps, European Journal of Operational Research (2017), http://dx.doi.org/10.1016/j.ejor.2017.07.023



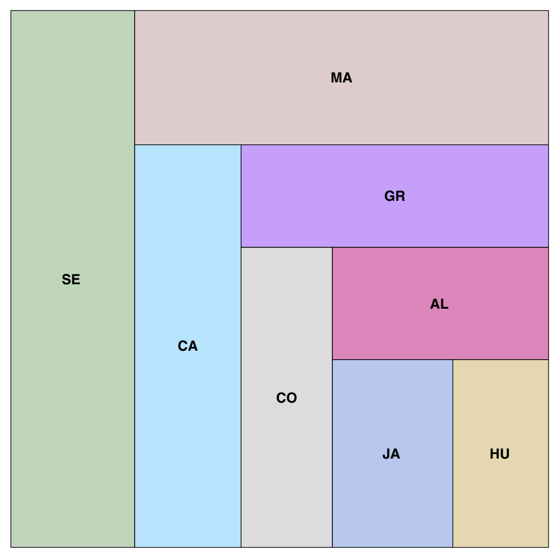
(a) Pie Chart



(b) Fan Chart



(c) Rectangular map



(d) Rectangular map

Fig. 1. Examples of planar space-filling visualization maps.

(Nusrat & Kobourov, 2016; Tobler, 2004). One of the most popular visualization tools for this is rectangular cartograms, which were first introduced in Raisz (1934) and have been further investigated in, e.g., Buchin, Speckmann, and Verdonschot (2012); Eppstein, Mumford, Speckmann, and Verbeek (2009); Heilmann et al. (2004); Kreveld and Speckmann (2007). The approaches developed in the literature to obtain rectangular cartograms take advantage of the geographical relative positions of the individuals (countries, cities, etc.) to obtain a cartogram, and thus their approaches cannot be directly extended to more general data structures. Fig. 2 depicts a rectangular cartogram for the geographical area of the states in the U.S. built using the Recmap package in R (Panse, 2016), see Fig. 2(b).

When dealing with proximity, a common approach in the literature has been to represent *close* individuals as *close* portions in the visualization map, see Abbiw-Jackson, Golden, Raghavan, and Wasil (2006); Carrizosa, Guerrero, and Romero Morales (2017a); 2017b); Duarte, Sikansi, Fatore, Fadel, and Paulovich (2014); Gómez-Nieto et al. (2014); Hahsler (2017) and references therein. A grid map (Eppstein, van Kreveld, Speckmann, & Staals, 2015) is a visualization tool that represents as accurately as possible the adjacencies present in a geographical dataset by assigning exactly one cell of the grid to each individual, although frequencies are not taken into account. Fig. 2(c) depicts the grid map built for the 48 contiguous states in the U.S., see Fig. 6-L_2^2 in Eppstein et al. (2015), representing 56 adjacencies of the 105 present in the actual map, see Fig. 2(a). With the methodology described in Section 4, we are able to represent 63 adjacencies of the 105 present in Fig. 2(a), see Fig. 2(d). In this paper, our goal is to propose a mathematical optimization formulation and a suitable solution approach to build rectangular maps to visualize the frequency distribution $\omega = (\omega_1, \ldots, \omega_N)$ and the proximity between the individuals, measured by an adjacency matrix $E = (e_{rs})$. As far as the authors are aware, this is a novel problem in the literature.

Throughout this paper, the weighted graph $G = (V, E, \omega)$ will model the set V of individuals, attached with the binary relation (adjacency) E and the frequency distribution ω. Similarly, we denote by $G^{\mathbf{P}} = (V, E^{\mathbf{P}}, \omega^{\mathbf{P}})$, the weighted graph associated with the