Extending Scatterplots to Scalar Fields

Shenghui Cheng[1], Pengcheng Cui[2] and Klaus Mueller[1]

[1]Visual Analytics and Imaging (VAI) Lab, Computer Science Department, Stony Brook University

[2]Mathematics Department, Shandong University, China

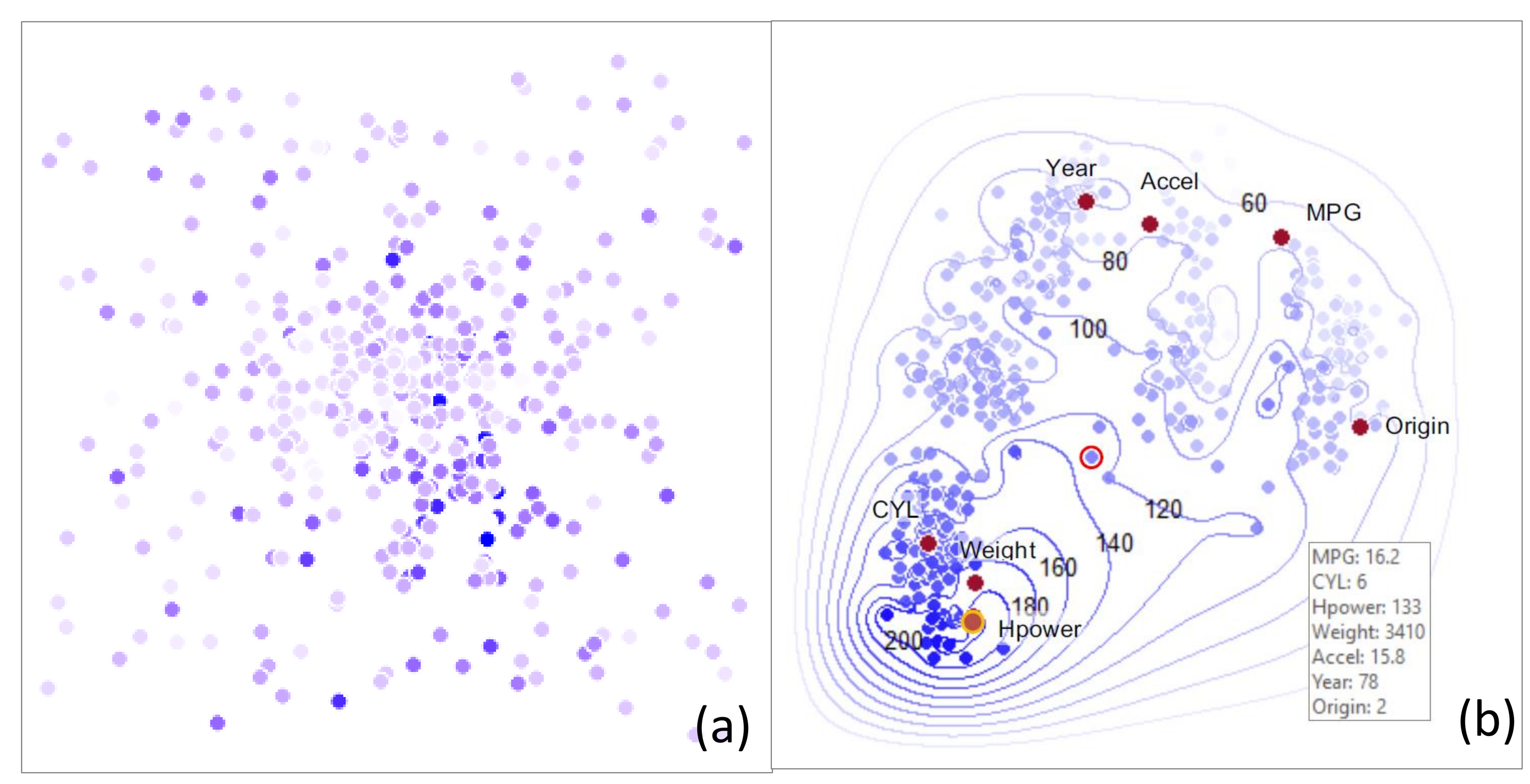
Introduction

Embedding high-dimensional data into a 2D canvas is a popular strategy for their visualization. It allows users to appreciate similarity relationships among the data points by their spatial organization on the 2D display. In this work we consider the case where the collection of data points also serves as a scalar field for one of the dataset's attributes, essentially forming samples of the attribute's continuous function in this embedded space. We study methods by which this continuous function can be estimated from the discrete samples, making certain assumptions on the function's smoothness in high-dimensional space. Our method allows users to create distance fields, iso-contours, topographic maps, and even extrapolations to embed the possibly odd-shaped point assembly into a filled rectangular region.

Approach

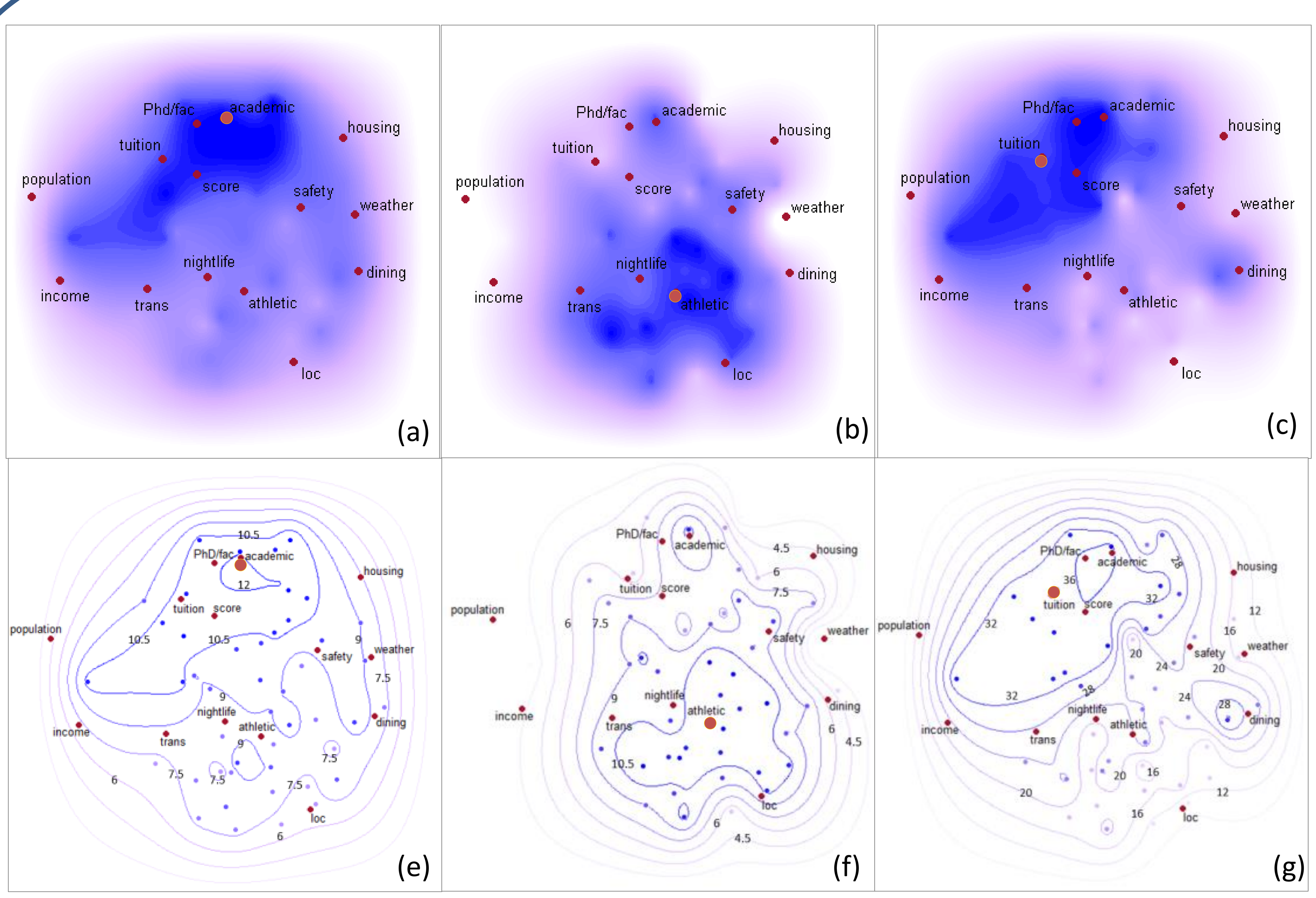
Coloring a bivariate scatterplot with the values of a third attribute will give rise to a random confetti-like arrangement of colored points. The same is true when attempting to color a 2D layout generated by multi-dimensional scaling (MDS). Panel (a) shows such a display where we mapped the third variable to brightness. We can easily observe that bright and less bright points are distributed across the display without clear structure. To derive a better and more organized display, we utilized the joint data-attribute embedding we recently published in [1] (see panel (b)).

To fill the empty space and obtain a continuous scalar field, we devised a new estimation method called OIE-AKDE (an original (O), interpolating (I), extrapolating (E) variant of adaptive kernel density estimation (AKDE).) A helpful illustration often used in scalar fields is the iso-contour – a closed curve indicating a certain level of the scalar attribute. A set of iso-contours, one for each level at some level spacing, can then act as a topographic map. For example, if a user desires a car with "HPower" between 120~140, (s)he can simply find the car of interest based on the iso-contour, e.g., the red circled one - "Peugeot 604s" with its values shown in the box (right).



[1] S. Cheng, K. Mueller, "The Data Context Map: Fusing Data and Attributes into a Unified Display," IEEE TVCG, 22(1): 121-130, 2016.

Case Study



Selecting a good university is significant for students in their final years in high school. Take Peter, for example, who prioritizes the following factors in his selection: low tuition, good academics, and good athletics. He wishes to attend university with good academics (>9), good athletics (>9), and low tuition (<$18,000).

Peter uses our interface to visualize data with college-related factors. In the frames on the left, the large red nodes represent the attributes while the small blue points represent the colleges. We apply OIE-AKDE to estimate the full field according to the attributes "academics", "athletics", and "tuition", respectively (see panels (a), (b), (c)). We observe that the scalar field is smooth for all attributes. In order to force extrapolations, we consider the data values to get smaller as we approach the border, and we assume the values of all border points to be zero in each dimension. By means of this setting, the estimation function will bend to zero and fade to border. Panels (e) (f), (g), show the associated contour maps.

Contact: {shecheng, mueller}@cs.stonybrook.edu

This research was partially supported by NSF grant IIS 1527200, the MSIP Korea, under the "IT Consilience Creative Program (ITCCP)" supervised by NIPA, and BNL LDRD grant 16-041.