

平行坐标综述

赖楚凡 袁晓如

摘 要:

关键词:

Parallel Coordinates

Chufan Lai Xiaoru Yuan

Abstract:

Keywords:

1 引言

在现实世界中，同一事物往往有多个不同的属性，要完整地描述该事物，需要记录其各个属性的信息。而这类具有多属性（维度）信息的数据，就称为多变量数据（Multivariate Data），或称高维数据（High-dimensional Data）。在传统的数据分析中，高维数据一般利用统计方法加以处理，并通过表格的方式呈现。但传统方法过于依赖自动算法，既无法结合用户的知识与判断，也不利于对抽象的高维信息的认知和理解。而信息可视化技术（Information Visualization）通过对高维数据的视觉呈现，辅以交互式的数据处理手段，能使用户充分参与到数据分析过程中，既加深了用户对数据的理解，也提高了分析的效率和可靠性。

高维数据一直是信息可视化领域的研究重点之一，其独有的两个特性：即维度数高、抽象性强，为可视化设计带来了巨大挑战。一方面，维度数的增长使得数据的信息量急剧增大，有限的显示空间只能表达一部分数据特征，而自动算法也会因为所谓的“维数灾难”（过高的复杂度、过低的采样率等 [4]）而难以应用。另一方面，由于人无法感知高于三维的拓扑空间，高维的数据特征（如聚类、流型等）往往抽象而难以理解。

针对这些挑战，领域内提出了各类高维可视化方法[12]，包括降维投影[11]、散点图矩阵[7]、雷达图[15]、星型坐标[20]等等。而由Inselberg等人提出的平行坐标 [17]（如图1）也因其简单直观、可拓展性强等优点而被广泛应用于高维数据的可视分析中。

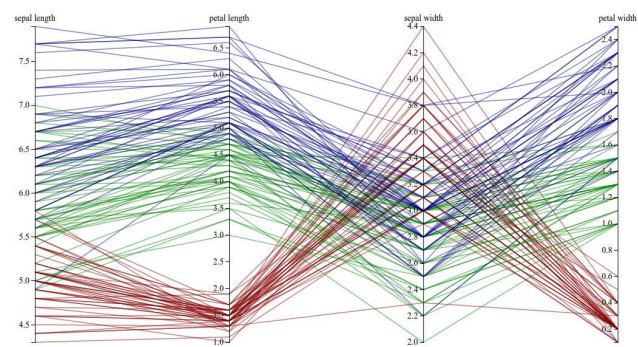


图 1: 平行坐标示例

然而，相比于其他高维可视化形式，平行坐标的设计也存在种种缺点，譬如展现的信息量较少、容易出现视觉混杂（Visual Clutter）、数据特征不直观等等（详见第3节）。针对这些问题，领域内的学者们做了大量的研究工作，并提出了各种方法对平行坐标进行改进和延展。

本文基于一些已有的综述论文 [12] [14] [26] [29]，并结合近年来的最新进展，将对平行坐标方面的研究工作进行比较、总结和展望。在文章

的余下部分里：第2节介绍了平行坐标的基本原理，并阐述其设计上存在的主要问题；第3节就平行坐标的不同部分，回顾了领域学者们为解决前述问题所作出的研究努力；第4节则介绍了平行坐标在不同领域内的应用现状；第5节基于已有的研究和应用，探讨了仍然存在的挑战，以展望未来的研究动向和应用前景；最后，第6节对全文内容进行总结。

2 基本原理及主要问题

在这一节里，我们将从四个方面简要介绍平行坐标的设计。在此过程中，我们会逐步阐述平行坐标的几个主要问题。作进一步的探讨前，这里先简要介绍Jarke提出的的可视分析模型 [27]（如图2）。在一个可视分析流程中，数据（Data）驱动了可视化展示（Visualization），用户对视图进行感知（Perception）并从中获取知识（Knowledge）。基于已有知识，用户可以通过交互探索（Exploration）来改变视图的参数设置（Specification），以产生新的可视化来发掘更多知识。由此可见，可视分析流程主要由四部分组成，分别是数据、视图、视觉感知和交互。本节余下内容以及第3节的内容都将按这四个部分展开。

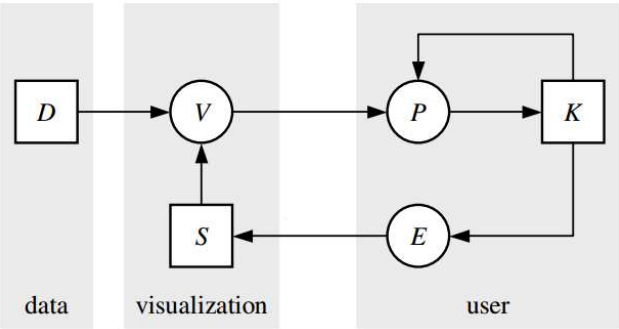


图 2: 可视分析模型。本图修改自Jarke 等人的论文 [27]。

2.1 数据

假设一个数据集有 m 个样本，每个样本有 n 个维度，则它可以记成一个 $m \times n$ 的矩阵，其中每一行是一个样本，每一列代表一个维度。这样的矩阵又称为数据矩阵（data matrix），是平行坐

标所处理的典型高维数据。数据的维度大致分为两种，即数值型维度（Continuous Dimension，具有无数种取值，如身高、体重）和类别型维度（Categorical Dimension，只有有限的取值，如国籍）。而一般的坐标轴只适用于数值型维度，类别型数据需要转换成数值型来表示，但往往会造成一定的误导（详见 2.2 节），这使得平行坐标不适合表现类别型数据。其次，高维数据的产生往往伴随着其他类型的信息，譬如高维时空数据、高维网络数据、高维流场数据等等。如何拓展平行坐标以表现这些混合高维数据，也是研究的重点之一。除了不同的数据类型，数据的隐私性、不确定性，对大数据的处理等等，都是在数据方面的研究会涉及的课题。

2.2 视图

正如第1节所提到，人无法感知高于三维的拓扑关系。而传统的笛卡尔坐标系限定了各坐标轴必须相互垂直，这使得高维的坐标空间极为抽象和难以理解。平行坐标解决这一难题的方法，顾名思义，就是将各个维度的坐标轴平行放置，如图3所示。一个高维数据在各个坐标轴上均有取值，连接各取值点所形成的一条折线，即表达了这个数据。事实上，基于线和点的转化关系，平行坐标和笛卡尔坐标之间有很好的几何对称性，有兴趣的读者可参阅相关文献 [17] [18]。

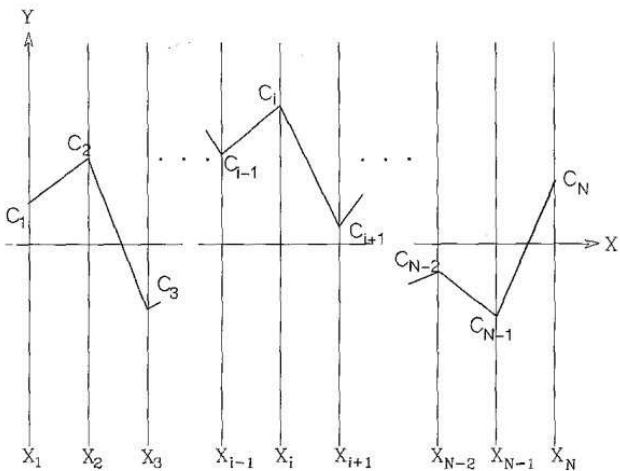


图 3: 平行坐标的构造原理。本图来源于Inselberg 等人的论文 [17]。

和其他高维可视化方法相比，平行坐标有几个突出的优点。首先，由于坐标轴的数量只受限

于屏幕的横向分辨率，它相比于散点图矩阵等可以容易容纳更多维度，具有很强的可扩展性。其次，它完整地表达了数据在每个维度上的取值，相比于降维投影、雷达图等具有较低的不确定性和信息损失。最后，平行坐标构造简单、无需自动算法辅助，既方便使用也容易理解。

但相对地，平行坐标也存在一些明显的局限和缺点：

其一，当一个维度上有多个相同的取值时，多条折线汇合到一个点，直接观察将无法区分不同的数据走向。而在类别型的数据中，类似问题出现于每一个轴，导致大量线段完全重叠，用户甚至无法辨别其中的数据量（见图4）。这正是平行坐标不适合表现类别型数据的原因。

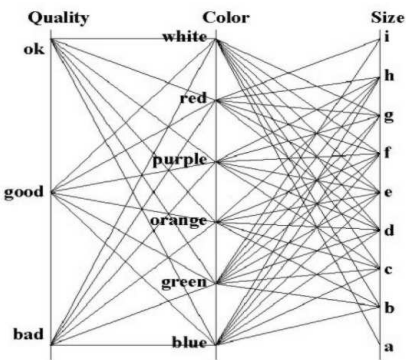


图 4: 类别型数据中的线段重合问题。本图来源于Rosario 等人的论文 [25]。

其二，每个数据用一条折线来表示，这会比散点形式占用更多的显示空间。当数据量增大时，视图很容易发生图元堆叠、导致视觉混杂（如图 5），用户将无法从中得到有用的信息。由此可见，平行坐标对数据规模的扩展性较差。

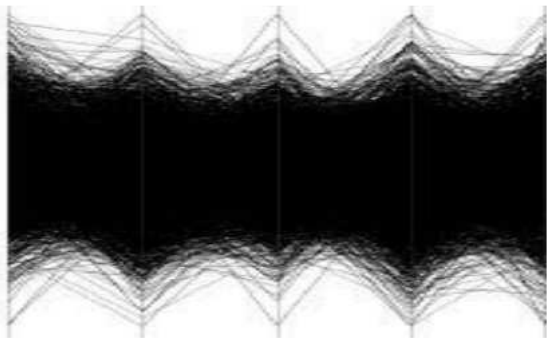


图 5: 视觉混杂问题。本图来源于Artero 等人的论文 [1]。

其三，平行坐标能够表现相邻维度间的关系，但信息量偏低。假设数据有 n 个维度，它们

两两组合可得到 $\frac{n(n-1)}{2}$ 个二维子空间，散点图矩阵能容纳其中的所有信息。但在平行坐标里，用户无法观察两个不相邻维度间的关系,因此只能看到 $(n-1)$ 个二维子空间。换言之，平行坐标只表现了 $\frac{2}{n}$ 的维度关系，不恰当的坐标轴排序很容易掩埋重要的相关性信息（如图6）。

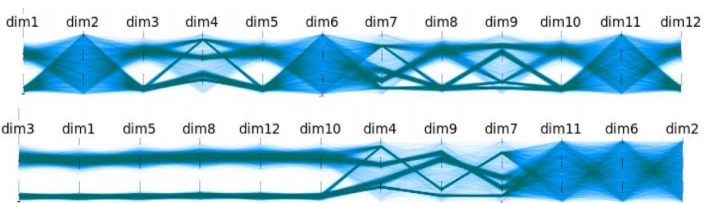


图 6: 不同坐标轴顺序对平行坐标的影响。其中上、下图分别为重排序前、后的视图，可见排序后各个维度的相关性和差异性更为明显。本图修改自Ferdosi 等人的论文 [10]。

最后，除了排序上的变化，坐标轴的放置还存在许多可能性。平行坐标的出现启发了大量对非垂直坐标系、甚至是非直线坐标轴的研究，这些都将在第 3 节中加以讨论。

2.3 感知

针对平行坐标的感知研究主要有两种，一是视觉感知的有效性评估，二是数据特征的刻画和提取。

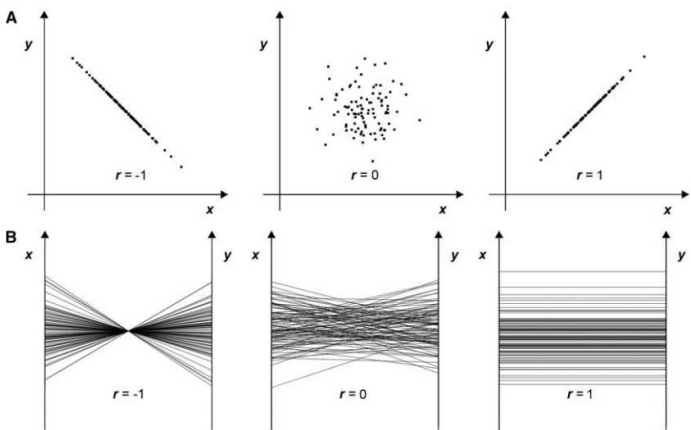


图 7: 维度相关性的解读。A、B两行分别显示了不同的相关关系下数据在散点图和平行坐标中的分布。本图来源于Li 等人的论文 [23]。

在平行坐标里，能观察到的数据特征主要有两种，即聚类信息和相关关系。一方面，相似的折线形成线簇，代表着不同的聚类，反映了数据

趋势、异常点等行为。但要观察到聚类，必须先处理好前面提到的视觉混杂问题，使隐含的聚类能被凸显出来。另一方面，相邻坐标轴之间的线段走向能反映这两个维度的相关关系（如图7）。但用户对相关性的感知是否准确，和其他方法相比是否更有效等等，都需要通过实验来评估。除此之外，还有很多有意义的数据特征无法通过简单的聚类或相关来描述。而已有的特征也需要相应的自动提取算法，来遍寻和评估不同的参数设置（如不同的维度排序），以找到特征明显、信息量大的视图。

2.4 交互

基于已有的观察，用户通过交互来改变视图，以发掘更多信息。平行坐标提供两种基本的交互功能，即数据刷选和坐标轴排序。对于刷选，用户需要给每个坐标轴限定取值范围[24]。这相当于在高维空间中建立一个超立方体（Hypercube）包围盒，同时满足各维度条件的数据才会被选中。这样的方式虽然逻辑简单，但操作繁琐、缺乏灵活性，用户既无法选中任意形状的数据区域，也无法根据其他数据特征来选择数据。除了刷选，用户还可以手动调整坐标轴的顺序。但这在数据维度较高时并不现实，用户很难通过交互来发现合适的排序。

3 平行坐标的相关研究

针对前一节中提出的各种问题，领域学者们做了许多研究工作来对平行坐标进行改进和延展。这一节将详细地回顾这些工作，并对它们进行比较和总结。为了保持行文结构、方便读者对照理解，这一节依然按数据、视图、感知和交互四个部分来展开。

3.1 数据

对数据的研究主要关注三个方面：不同的数据类型、数据隐私与不确定性、以及大数据的绘制效率。

3.1.1 不同的数据类型

正如 2.2 节所述，由于折线重叠问题，平行坐标并不适合表现类别型数据。在这种数据里，大量样本具有相同的取值，它们相互之间无法被区分。但事实上，这些数据的分析重点并不在于单个样本，而是同一类样本的集合，因此需要在平行坐标里加入表达集合关系的元素。Kosara 等人 [21] [5] 利用线段宽度来映射取值相同的样本数量，提出了平行集合（Parallel Set）的形式（如图8(a)）。而Hofmann等人 [16] 指出，折线的宽度并不对应它所代表的数据量（比较图8(b)中的 a 、 b 两段）。他们提出用统一的角度来绘制折线，即统一角度图（Common Angle Plot）的形式，使用户能通过直线部分的宽度来准确感知集合的大小。

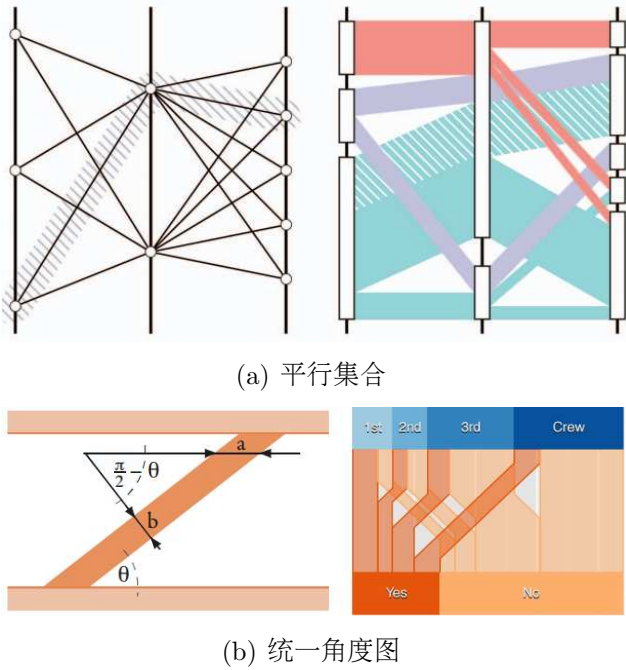


图 8: 平行集合及其改进版本，用以在平行坐标框架下表现类别型数据。图(a)、(b)分别来自论文 [21] 和 [16]。

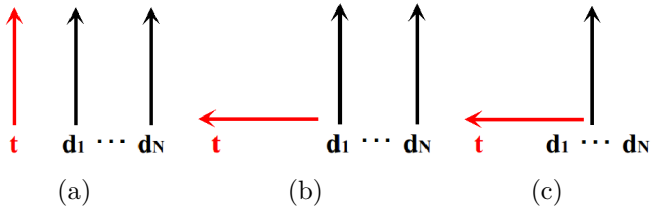


图 9: 时变平行坐标的三种静态展示方式：(a).时间轴平行于属性轴；(b).时间轴在坐标平面内垂直于属性轴；(c).时间轴垂直于坐标平面。

除了类别型数据，时变数据也是一类重要的拓展，具体方法可分为静态和动态两种。其中静态的展示需要引入时间轴、并同时绘制所有时刻的数据，只是对时间轴的处理各有不同（见图9）。

最简单的方式莫过于将时间轴看作一个平行轴，但各个时刻的数据被同等对待，不利于凸显数据的变化信息。Johansson等人 [19]提出利用不透明度和色调等来区分数据变动的时间范围（如图10）。用户可以直观地看出变化的频率和趋势，但不同时刻的数据依然会相互遮挡。

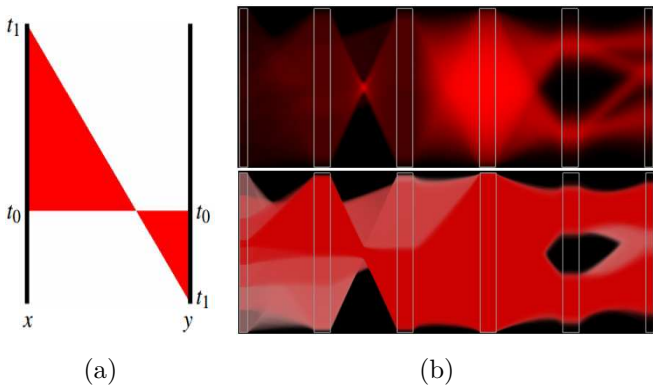


图 10: 用不透明度和色调刻画数据变动。其中折线变化用多边形表示（如图(a)），不透明度刻画了变化的频率，不同的色调则代表不同的时刻（如图(b)）。本图来源于Johansson 等人的论文 [19]。

为防止遮挡，也可以结合折线图（Line Chart），用多个轴表示一个维度的时变信息 [9] [8]（如图11），但同一维度的重复出现会降低视图的可扩展性。

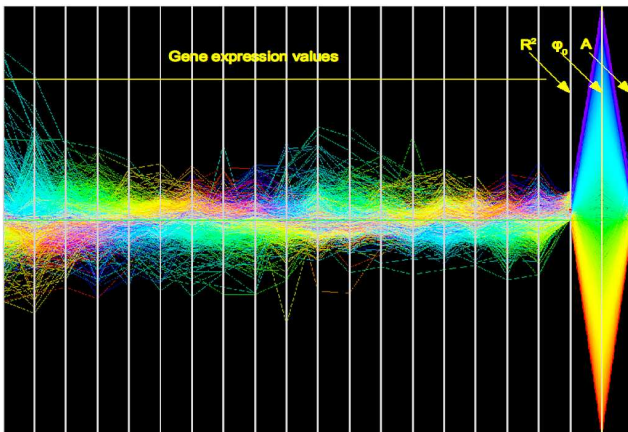


图 11: 在平行坐标中嵌入折线图。本图来源于Dietzsch 等人的论文 [8]。

进一步地，Wegenkittl等人 [28]令时间轴垂直于坐标平面，提出了三维时变平行坐标的形式

（如图12）。其中每条折线沿时间轴演化会形成一个曲面，该曲面即表达了时变高维的所有信息。但多个不同的曲面很容易交错重叠而难以观察，使其不适用于数据量大的情形。

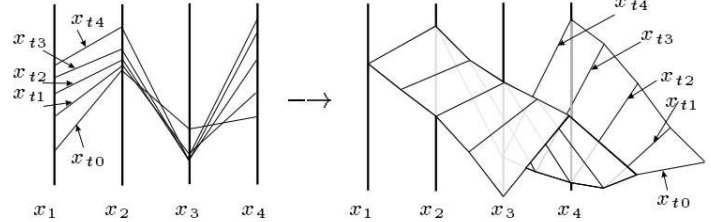


图 12: 三维时变平行坐标。图中 x_1 至 x_4 表示不同维度， t_1 至 t_4 为不同的时刻。本图来源于Wegenkittl 等人的论文 [28]。

除了静态展示，还可以用动画来表现数据变化 [3]。但视觉的瞬变使用户难以同时跟踪多个维度的变动，在维数较高、变动剧烈的情况下尤为如此。为每个维度提供额外的时变直方图 [6]可以缓解这一问题，但这又割裂了时变信息与维度关系。总的来说，以上方法都各有优劣，但还没有一个较为全面的方式来展现时变平行坐标。

还有一类拓展在于连续型数据。在科学数据（Scientific Data）里，所有样本可能都来自同一个数据域（Data Domain，如流场、物体表面等）。平行坐标能够逐个描绘这些样本，但无法展现通过插值重构出来的连续数据域，即所谓的“连续型数据”（Continuous Data）。绘制连续型数据的概念最早由Bachthaler等人 [2]提出并应用在散点图中。他们基于离散的样本，建模估计、并以颜色表现了视图各处的密度。基于这一研究，Heinrich等人利用散点图和平行坐标的几何对称性，发展出连续型平行坐标 [13]的形式（如图13）。后来的研究 [22]则进一步揭示了这两种连续视图之间的密切联系。

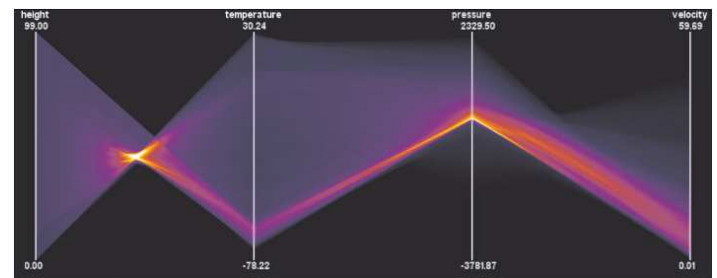


图 13: 连续型平行坐标。本图来源于Heinrich 等人的论文 [13]。

相比于传统形式，连续型平行坐标可以在样本较少的情况下重构出精度较高的原数据域，而且能避免视图混杂问题（见 3.4 节），也具有较好的可扩展性和渲染效率。但它无法给出单个样本的信息，也不支持传统的交互和分析方法（如刷选、聚类等），在可用性上有一定的局限。

3.1.2 数据隐私与不确定性

3.2 视图

3.3 感知

3.4 交互

4 平行坐标的应用

5 探讨与展望

6 结论

参考文献

[1] A. O. Artero, M. C. F. de Oliveira, and H. Levkowitz. Uncovering clusters in crowded parallel coordinates visualizations. In *Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on*, pages 81–88. IEEE, 2004.

[2] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1428–1435, 2008.

[3] N. Barlow and L. J. Stuart. Animator: A tool for the animation of parallel coordinates. In *Information Visualisation, 2004. IV 2004. Proceedings. Eighth International Conference on*, pages 725–730. IEEE, 2004.

[4] R. E. Bellman and S. E. Dreyfus. *Applied dynamic programming*. 1962.

[5] F. Bendix, R. Kosara, and H. Hauser. Parallel sets: Visual analysis of categorical data. In *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*, pages 133–140. IEEE, 2005.

[6] J. Blaas, C. P. Botha, and F. H. Post. Extensions of parallel coordinates for interactive exploration of large multi-timepoint data sets. *Visualization and Computer Graphics, IEEE Transactions on*, 14(6):1436–1451, 2008.

[7] W. C. Cleveland and M. E. McGill. *Dynamic graphics for statistics*. CRC Press, Inc., 1988.

[8] J. Dietzsch, J. Heinrich, K. Nieselt, and D. Bartz. Spray: A visual analytics approach for gene expression data. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on*, pages 179–186. IEEE, 2009.

[9] R. M. Edsall. The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics & Data Analysis*, 43(4):605–619, 2003.

[10] B. J. Ferdosi and J. B. Roerdink. Visualizing high-dimensional structures by dimension ordering and filtering using subspace analysis. In *Computer Graphics Forum*, volume 30, pages 1121–1130. Wiley Online Library, 2011.

[11] I. K. Fodor. A survey of dimension reduction techniques. 2002.

[12] G. Grinstein, M. Trutschl, and U. Cvek. High-dimensional visualizations. In *Data mining conference KDD workshop 2001*, pages 7–19, 2001.

[13] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 15(6):1531–1538, 2009.

[14] J. Heinrich and D. Weiskopf. State of the art of parallel coordinates. *STAR Proceedings of Eurographics*, 2013:95–116, 2013.

- [15] P. Hoffman, G. Grinstein, and D. Pinkney. Dimensional anchors: a graphic primitive for multidimensional multivariate information visualizations. In *Proceedings of the 1999 workshop on new paradigms in information visualization and manipulation in conjunction with the eighth ACM international conference on Information and knowledge management*, pages 9–16. ACM, 1999.
- [16] H. Hofmann and M. Vendettuoli. Common angle plots as perception-true visualizations of categorical associations. *Visualization and Computer Graphics, IEEE Transactions on*, 19(12):2297–2305, 2013.
- [17] A. Inselberg. The plane with parallel coordinates. *The Visual Computer*, 1(2):69–91, 1985.
- [18] A. Inselberg. *Parallel coordinates*. Springer, 2009.
- [19] J. Johansson, P. Ljung, and M. D. Cooper. Depth cues and density in temporal parallel coordinates. In *EuroVis*, volume 7, pages 35–42, 2007.
- [20] E. Kandogan. Star coordinates: A multidimensional visualization technique with uniform treatment of dimensions. In *Proceedings of the IEEE Information Visualization Symposium*, volume 650, page 22, 2000.
- [21] R. Kosara, F. Bendix, and H. Hauser. Parallel sets: Interactive exploration and visual analysis of categorical data. *Visualization and Computer Graphics, IEEE Transactions on*, 12(4):558–568, 2006.
- [22] D. J. Lehmann and H. Theisel. Features in continuous parallel coordinates. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):1912–1921, 2011.
- [23] J. Li, J.-B. Martens, and J. J. Van Wijk. Judging correlation from scatterplots and parallel coordinate plots. *Information Visualization*, 9(1):13–30, 2010.
- [24] A. R. Martin and M. O. Ward. High dimensional brushing for interactive exploration of multivariate data. In *Proceedings of the 6th Conference on Visualization’95*, page 271. IEEE Computer Society, 1995.
- [25] G. E. Rosario, E. A. Rundensteiner, D. C. Brown, M. O. Ward, and S. Huang. Mapping nominal values to numbers for effective visualization. *Information Visualization*, 3(2):80–95, 2004.
- [26] H. Siirtola and K.-J. Räihä. Interacting with parallel coordinates. *Interacting with Computers*, 18(6):1278–1309, 2006.
- [27] J. J. Van Wijk. The value of visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 79–86. IEEE, 2005.
- [28] R. Wegenkittl, H. Löffelmann, and E. Groller. Visualizing the behaviour of higher dimensional dynamical systems. In *Visualization’97., Proceedings*, pages 119–125. IEEE, 1997.
- [29] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization.