Aggregated Parallel Coordinates: Integrating Hierarchical Dimensions into Parallel Coordinates Visualisations

Keith Andrews
Institute for Information Systems and Computer Media,
Graz University of Technology,
Inffeldgasse 16c,
8010 Graz, Austria
kandrews@iicm.edu

Majda Osmić
AVL List GmbH,
Hans-List-Platz 1,
8020 Graz, Austria
majda.osmic@avl.com

Gerhard Schagerl
AVL List GmbH,
Hans-List-Platz 1,
8020 Graz, Austria
gerhard.schagerl@avl.com

ABSTRACT

Aggregated Parallel Coordinates (APC) are an extension of standard parallel coordinates, which supports the visualisation and exploration of hierarchies within numerical dimensions. Such datasets can occur when data is available at several granularities and these can be grouped or aggregated in some way (mean, sum, max) to form higher levels of abstraction. While existing parallel coordinates techniques can be used to visualise individual dimensions of such data, they have no provision for interactively expanding and collapsing such hierarchically aggregated dimensions.

In race car engineering, specialised simulation software is used to derive a car setup for optimal performance during a race. Such simulations generate complex, high-dimensional datasets with a large number of records, many of which are inherently hierarchically grouped. The APC software was implemented in C$^\sharp$ as a Microsoft WPF user control library. It has been integrated into AVL's data visualisation tool called SimBook and is being used to explore the output of race car simulations.

CCS Concepts

•**Human-centered computing → Visualization techniques; Information visualization;**

Keywords

information visualisation, parallel coordinates, race car simulation, hierarchical relationships, drill-down, roll-up, aggregate dimensions

1. INTRODUCTION

Multi-dimensional datasets are very common in information visualisation. Sometimes, multi-dimensional datasets have inherent hierarchies, particularly within numerical dimensions. Such datasets can occur, for example, when data is available at several granularities and these can be grouped or aggregated in some way (mean, sum, max) to form higher levels of abstraction. Think of sales data for three products at local, regional, and national level. This kind of data is frequently encountered in data mining and is called a *data cube* [4]. Operations inside a data cube to navigate an inherent hierarchy and access either coarser (more aggregate) or finer (less aggregate) data are known as *roll-up* and *drill-down* respectively.

In visual analytics, where techniques from information visualisation often provide a front-end interface to back-end analytics processing, roll-up and drill-down interactions are usually available generically, outside of any specific visualisation, and the resulting aggregate or individual dimensions are then visualised. However, when an analyst is already working inside a particular visualisation, it can make sense to provide access to such operations directly inside the visualisation.

I-Know 2015 21–23 October 2015, Graz, Austria
© 2015 ACM. ISBN 978-1-4503-2138-9. . . $15.00
DOI: 10.1145/1235

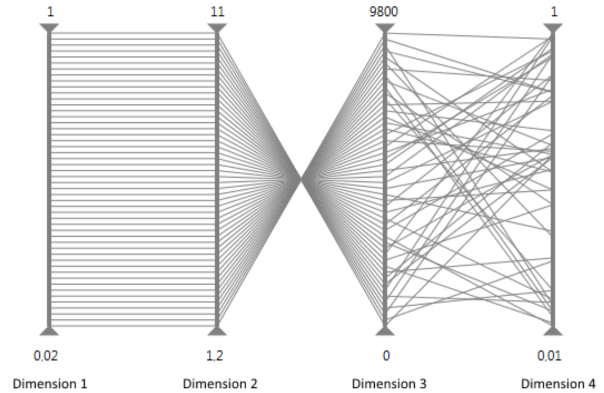

Figure 1: Parallel coordinates. Dimensions are represented by parallel vertical lines, records are represented by horizontal polylines. Similar records have similar polylines.

2. PARALLEL COORDINATES

Parallel coordinates [5] are used to visualise multi-dimensional data comprising dimensions and records. The dimensions of the dataset are visualised as axes placed in parallel to one other, whereas records are visualised as polylines strung between the axes, as shown in Figure 1. Since the axes are placed in parallel to each other, an arbitrary number of dimensions can be visualised. Records with similar values have similar polylines. Neighbouring dimensions displaying relatively few edge crossings is an indication of a possible (positive) correlation between the two dimensions. Many polyline segments crossing in the middle, on the other hand, is an indication

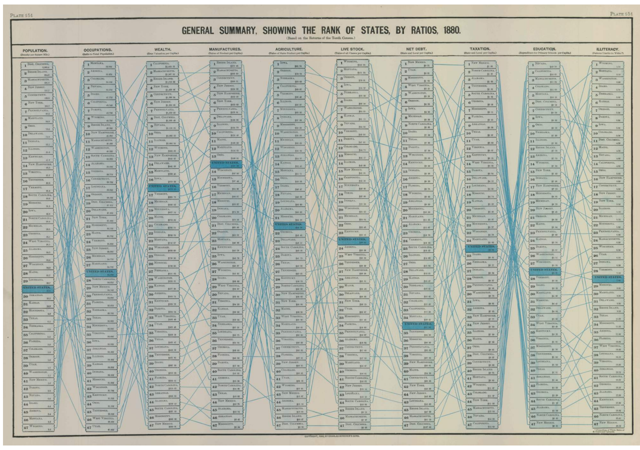

Figure 2: One of Henry Gannett's parallel coordinates illustrations for the 1883 edition of Scribner's Statistical Atlas of the United States [6, Plate 151]. The ten vertical dimensions show different statistics for each of the 47 states. Each state appears once in each ranking and its instances are connected by a line.

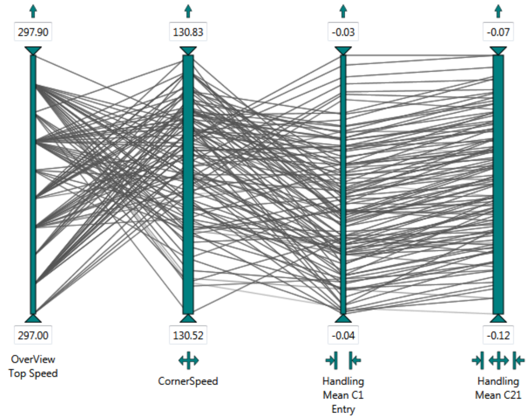

Figure 3: The four different types of axis in Aggregated Parallel Coordinates (APC). The leftmost axis is atomic and does not belong to a hierarchy. The second axis is a root axis which can be expanded. The third axis is a leaf axis, which does not itself contain children, but has a button for collapsing back to its parent. The rightmost axis is an inner axis having both its own children and a parent.

of a possible negative correlation. This can be seen in Figure 1.

Parallel coordinates were first used by Henry Gannett in 1883 for several illustrations in Scribner's Statistical Atlas of the United States [6], which drew on data from the 1880 census. The most striking example is Plate 151, reproduced in Figure 2, which shows 10 different statistics (such as Population, Occupations, and Illiteracy) for the (then) 47 states. Each vertical dimension represents a statistic, each state appears once in each dimension, and its instances are connected by line segments. Two years later, in 1885, Maurice d'Ocagne described the theoretical and mathematical properties of parallel coordinates [2]. Alfred Inselberg independently re-invented the technique in 1959 and went on to popularise it [8], culminating in his definitive book in 2009 [7].

As with most information visualisation techniques, the visual representation itself is only half the story. Interaction with the visual representation is the essential complement. The following interactive features are commonly provided with parallel coordinates: record highlighting, selection, and brushing, record filtering, axis inversion, axis re-ordering, adding and removing axes, displaying histograms on axes, displaying a mean line, zooming, and an undo/redo mechanism. Common extensions to parallel coordinates include semi-transparency for records and clustering of records.

3. MULTIDIMENSIONAL HIERARCHICAL DATA

In race car simulation, a huge number of input and output parameters are considered. A typical resulting dataset might contain a dozen or more input parameters (dimensions), a few hundred output parameters (also dimensions) and several hundred records. Input parameters often include tyre camber angle, spring stiffness, position of the front and rear wings, etc. Output parameters often include the overall lap time, but also a large variety of others such as aerodynamic efficiency, handling, and tyre grip. Each record corresponds to a run of the simulation with a set of unique input values. Table 1 shows part of a typical race car simulation dataset.

Some output parameters exhibit a hierarchical relationship, which sometimes analysts would like to see as an aggregate (roll-up) rather than seeing each parameter individually. For example, a track can be divided into straights and corners. The corners of a track are usually divided into three segments: entry, mid, and exit. Handling is usually calculated for every segment of every corner of a track (see Table 1b), but an analyst might prefer to look at say the mean value over an entire corner. Or the mean value over all entry segments of all corners. Or the mean value for all straights over the whole track.

4. AGGREGATED PARALLEL COORDINATES

The Aggregated Parallel Coordinates (APC) software developed in this project [9] implements many of the standard features found in other parallel coordinates packages. The distinctive new contribution in APC is the support for an analyst to explore (expand and collapse) aggregate dimensions inline inside the parallel coordinates plot. Aggregate axes are visually distinguished by a slightly wider axis to indicate that they have children. The four types of axis found in APC are:

- *atomic*: having neither parent nor child (standard).
- *root*: having children, but no parent.
- *leaf*: having a parent, but no children.
- *inner*: having both a parent and children.

These can be seen in Figure 3. Root, leaf, and inner axes are *aggregate* axes and have a special button located directly beneath them to expand and/or collapse the corresponding dimension. An aggregate axis displays record values which are the aggregate (i.e. mean, minimum, or maximum) of each corresponding child record. The range values used for the top and bottom of an aggregate axis are always derived from the extreme values of all direct children, so as to avoid problems with excluded records upon expansion.

Since axes can be moved freely across the plot, expanded axes which have the same parent do not necessarily have to remain located next to each other. Identifying which axes have the same parent is possible by hovering over a Collapse button, which highlights all sibling axes, as shown in Figure 4. In case a sibling axis has itself been expanded (i.e. replaced by its children), all its children

No.	...	CamberFL	CamberFR	CamberRL	CamberRR	ToeFL	ToeFR	ToeRL	ToeRR	Lap Time	...
1		1.6	3.1	1.8	3.2	2.7	1.5	5.3	-0.61	98.1238	
2		2.8	2.5	2.6	2.4	3.6	2.6	4.1	-0.32	99.3412	
...		...								...	
1000		1.3	4.2	3.1	3.8	5.1	3.8	2.6	-0.49	98.9615	

(a) Input parameters Camber and Toe are typically defined separately for all 4 vehicle wheels (FL, FR, RL, RR) and analysts do not typically aggregate them. Lap Time is often an important output parameter.

No.	...	Handling C1Entry	Handling C1Mid	Handling C1Exit	...	Handling C21Mid	Handling C21Exit	...
1		-0.2	-0.3	-2.0		-0.3	-0.8	
2		-2.8	-1.1	-1.0		-0.1	-0.3	
...		...				...		
1000		-0.5	-0.7	-0.6		-0.9	-0.9	

(b) Handling is an output parameter usually calculated for all three segments (entry, mid, and exit) of every corner of the track. An analyst might want to aggregate either by corner or by segment.

Table 1: A typical race car simulation dataset has a dozen or more input parameters (dimensions), a few hundred output parameters (also dimensions) and several hundred records. Here, some example dimensions (columns) are shown for a dataset with 1000 records (rows).

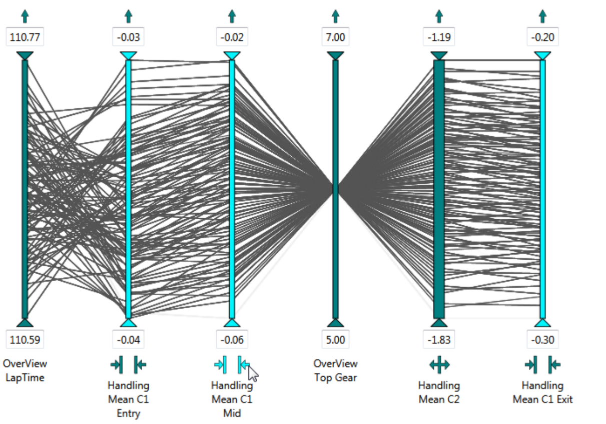

Figure 4: Hovering over a Collapse button highlights all of its descendants, i.e. all the axes which will be collapsed by clicking on it, in a lighter shade of blue. This interaction allows for easy identification of hierarchically related dimensions.

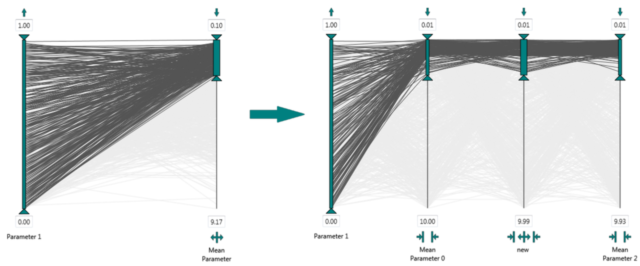

Figure 5: Expanding an aggregate axis. Any scaling, orientation, and slider values are inherited by its child axes.

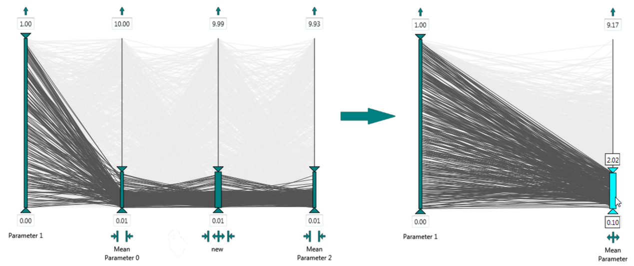

Figure 6: Collapsing an aggregate axis. The minimum and maximum slider values of all children are calculated. The sliders of the parent axis are then set to these values.

are also highlighted, since they would be removed from the view as well, if the Collapse button were to be pressed.

When an aggregate axis is expanded, its child axes are inserted into the plot at the same position in the plot, replacing the aggregate axis. If the parent axis was inverted at the moment of expanding, all its children will also be inverted. The children also inherit any scaling and slider positions from the parent. This behaviour is shown in Figure 5.

When a parent axis is collapsed, its new state is derived from its children. If the majority of children are inverted at the moment of collapsing, the aggregate axis will also be inverted. In case the number of inverted children is the same as the number of non-inverted children, the parent axis will have the same state as it had before collapsing. The minimum and maximum values of all sliders on child axes are calculated, and the sliders of the collapsed parent are placed in the corresponding position, as shown in Figure 6. The same applies when deriving scaling from child axes.

The APC software was implemented in C$^\sharp$ as a Microsoft WPF user control library, so that it can be easily integrated into any WPF application. The APC widget was developed in an agile development process, with frequent updates, and race car engineers providing direct feedback.

5. APPLICATION OF APC IN SIMBOOK

SimBook is a tool developed by AVL Racing, which is especially designed for visual exploration of race car simulation data. The APC user control (widget) was integrated into SimBook so that it could be used and tested by race car engineers on real datasets. The SimBook user interface is shown in Figure 7.

The Parameter Overview panel provides a hierarchical tree-view all parameters within the dataset. At the top level, parameters are

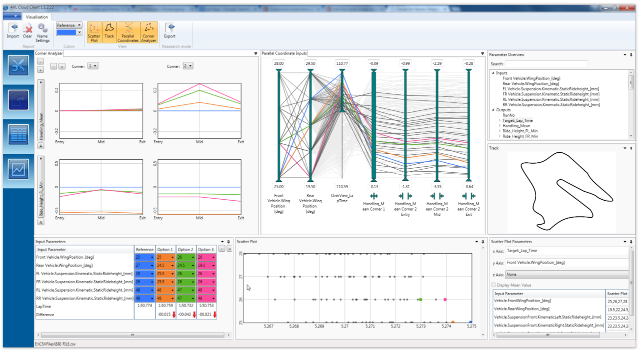

Figure 7: The SimBook tool for visual exploration of race car simulation data. Aggregated Parallel Coordinates are in the central panel. The parameter Handling Mean for Corner 2 has been expanded to reveal its Entry, Mid, and Exit components.

grouped into two categories: input and output. Input parameters are part of the car setup. Output parameters are calculated by the simulation software based on the car setup and sometimes have hierarchical structure within the dimensions.

In the setup table (the Input Parameters panel), input parameters are displayed as rows, while columns represent different combinations of the input parameters (setups). Each setup corresponds to one configuration of the input parameters, or in other words, one record in the dataset. The setup table allows a reference setup to be compared against a small number of alternative setups.

The Parallel Coordinates panel is initially populated with the currently selected set of input parameters and lap time (as axes) and the entire set of records. This allows easy comparison of lap times for all setups, and subsequent elimination of those setups with too high lap times. Additional parameters (dimensions) can be added to the parallel coordinates plot simply by dragging the parameter from the tree view and dropping it onto the Parallel Coordinates panel. If a parameter containing sub-dimensions is dropped onto the panel, the parameter is displayed as an aggregate axis. Depending on the needs of the analyst, parameters at any level of hierarchy can be individually added to the view.

6. RELATED WORK

The authors are not aware of any other software providing for drill-down and roll-up of aggregated parameters (dimensions) directly inside a parallel coordinates visualisation. These kind of interactions are available more generically (outside of any specific visualisation) in visual analytics tools such as Tableau [11] or Spotfire [10].

In their state-of-the-art survey, Heinrich and Weiskopf [5] refer to aggregation within parallel coordinates in the sense of aggregation of data *records* into (possibly hierarchical) clusters. Fua, Ward and Rundensteiner [3] describe a technique called "Hierarchical Parallel Coordinates", which generates and displays a cluster hierarchy of records in a parallel coordinates visualisation. However, hierarchically related *dimensions* in our sense are completely distinct from a cluster hierarchy of *records*.

7. FUTURE WORK

The current implementation of APC utilises internal data structures which explicitly encode any hierarchical relationships between dimensions. This is fine for systems which maintain such relationships internally. To facilitate the storage and exchange of datasets with hierarchical aggregate dimensions, a non-standard version of CSV files can be used today, but further in the future, it might be possible to read input directly as a data cube, using say the RDF Data Cube Vocabulary [1].

8. CONCLUDING REMARKS

Aggregated Parallel Coordinates (APC) emerged from a real need in a real project for race car engineers to be able to drill-down and roll-up aggregate dimensions from within a parallel coordinates plot. There are potentially many other scenarios in visual analytics where this might be the case and where APC could prove a useful technique.

References

[1] Richard Cyganiak and Dave Reynolds. *The RDF Data Cube Vocabulary*. W3C Recommendation. 16th Jan 2014. http://w3.org/TR/vocab-data-cube/.

[2] Maurice d'Ocagne. *Coordonnées Parallèles et Axiales: Méthode de transformation géométrique et procédé nouveau de calcul graphique déduits de la considération des coordonnées parallèlles*. Gauthier-Villars, 1885. 93 pages. https://archive.org/details/coordonnesparal00ocaggoog.

[3] Ying-Huey Fua, Matthew O. Ward and Elke A. Rundensteiner. "Hierarchical Parallel Coordinates for Exploration of Large Datasets". In: *Proc. 10th IEEE Visualization Conference (Vis'99)*. (San Francisco, CA, USA). IEEE Computer Society, 24th Oct 1999, pages 43–50. ISBN 078035897X. doi:10.1109/VISUAL.1999.809866. http://davis.wpi.edu/xmdv/docs/vis99_HPC.pdf.

[4] Jim Gray et al. "Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals". In: *Data Mining and Knowledge Discovery* 1.1 (1st Mar 1997), pages 29–53. ISSN 1384-5810. doi:10.1023/A:1009726021843. http://arxiv.org/pdf/cs/0701155v1.

[5] Julian Heinrich and Daniel Weiskopf. "State of the Art of Parallel Coordinates". In: *Proc. Eurographics 2013*. (Girona, Spain). Eurographics. 6th May 2013, pages 95–116. doi:10.2312/conf/EG2013/stars/095-116. http://joules.de/files/heinrich_state_2013.pdf.

[6] Fletcher Willis Hewes and Henry Gannett. *Scribner's Statistical Atlas of the United States*. 743–745 Broadway, New York, NY, USA: Charles Scribner's Sons, 1883. http://www.loc.gov/item/a40001834.

[7] Alfred Inselberg. *Parallel Coordinates: Visual Multidimensional Geometry and Its Applications*. Springer, 8th Oct 2009. 554 pages. ISBN 0470856181.

[8] Alfred Inselberg. "The Plane with Parallel Coordinates". In: *The Visual Computer* 1.4 (Dec 1985), pages 69–91. doi:10.1007/BF01898350.

[9] Majda Osmić. "Aggregated Parallel Coordinates – Multi-Dimensional Information Visualisation in Race Car Engineering". Master's Thesis. Institute for Information Systems and Computer Media (IICM): Graz University of Technology, 21st May 2015. http://ftp.iicm.tugraz.at/pub/theses/mosmic-2015-msc.pdf.

[10] Spotfire. *Spotfire*. 21st Jun 2015. http://spotfire.com/.

[11] Tableau. *Tableau*. Tableau Software. 21st Jun 2015. http://tableau.com/.