

DiffPin: Interactive Specification of References for Comparative Small-Multiple Displays

F. Spechtenhauser and H. Piringer

VRVis Zentrum für Virtual Reality und Visualisierung, Austria

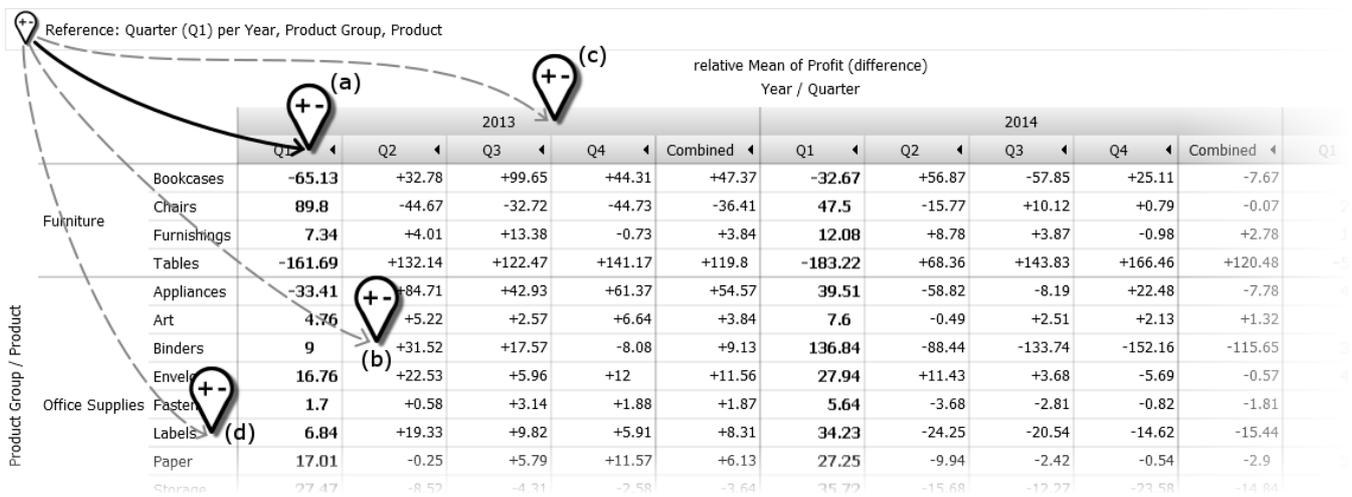


Figure 1: Average profit in an example of sales data, partitioned by Year/Quarter (X axis) and Product Group/Product (Y axis). The quarter "Q1" is used as reference for computing differences. The pins indicate possible reference types which can be specified by DiffPin: columns (a), single cells (b), columns with sub-structure (c), rows (d).

Abstract

Small-multiple displays are frequently used for visual comparison by juxtaposition. Explicitly encoding differences between cells enables to convey also minor differences in the data, but requires to specify references for the comparison which is non-trivial in views with a complex layout. This paper proposes DiffPin - an interaction technique to flexibly specify references for cells in small-multiple views. DiffPin enables an easy and fast specification of references on different levels of the visualization (e.g. cell, row, column), even in complex small-multiple views containing nested axes or subtotals. The paper provides considerations for visual encodings in comparative small-multiple displays using DiffPin, and proposes strategies for visually discriminating reference data from compared data for specific visualization types. Initial qualitative user feedback indicates that users are intuitively able to use the provided interactions to quickly and efficiently perform comparative tasks.

CCS Concepts

•Human-centered computing → Interaction techniques; Visual analytics;

1. Introduction

When dealing with multi-variate data, pivotization is commonly used for analysis by splitting the data using categorical attributes (dimensions, [Tho02]) and aggregating values of numerical attributes (measures). The resulting pivotized data is commonly

used for creating small-multiple displays, such as bar charts or heat maps. Juxtaposition of multiple displays supports a comparison between categories. However, when dealing with a larger number of categories, comparison by juxtaposition becomes increasingly difficult, and small differences in values may not be spotted [GAW*11]. To overcome these drawbacks, many ap-

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

proaches [LKH10, JDKW15, KSB*16, BBW16] rely on explicitly encoding differences in the visualization.

This paper builds upon previous work by Kehrer et al. [KPBG13], who define a model for specifying references for comparative small-multiple displays. Similar to Polaris [STH02] and Tableau [MHS07, Tab], their model is based on displays organized in a matrix, hierarchically subdivided into rows and columns by dimensions assigned to its axes (see Fig. 1). The resulting graphics (*cells*) of the matrix display the data for the corresponding category combinations. The model of Kehrer et al. provides a flexible definition of semantically meaningful reference graphs between these cells. This enables to explicitly encode the differences of each cell to its corresponding reference in a displayed table.

Our motivation for this work is to provide an easy way to interactively specify such references directly in the small-multiple display, so that users can intuitively explore differences between categories. The contributions can be summarized as follows:

- We propose a new interaction technique called DiffPin - a draggable pin to interactively set references for computed differences directly in the visualization.
- We characterize types of references, which can be defined using DiffPin for visualizations of pivoted data-tables.
- We propose considerations and strategies for visually encoding the resulting comparative small-multiple displays.
- We summarize initial qualitative user feedback from domain experts in industrial manufacturing, which suggests the usefulness of DiffPin for their daily tasks.

2. Related Work

The literature provides a variety of examples for the usefulness of small-multiple displays for comparison tasks [MHS07, RFF*08, APP11, vdEvW13]. Gleicher et al. [GAW*11] provide a taxonomy of visual designs for comparison, grouping them into three (combinable) categories: juxtaposition, superposition, and explicit encodings. Results of their survey suggest that the use of small-multiple displays for comparison by juxtaposition may place too much comparative burden on the user's memory. Explicitly encoding differences reduces this comparative burden, but requires the definition of relationships between the displayed objects.

The commercial system Tableau [MHS07, Tab] provides a flexible definition of small-multiple displays, based on the table algebra defined by Stolte et al. [STH02]. The current version 10.5 of Tableau supports an explicit encoding of absolute or percentage-based differences between categories. The references for the computed differences can be set via a dialog window, which allows to specify fixed categories of dimensions, or to use the first, previous, next or last category of dimensions as references. However, this dialog is decoupled from the small-multiple display.

Tominski et al. [TFJ12] propose interaction techniques inspired by real-world behaviour of people. Their approach allows to flexibly arrange multiple displays on the screen. This enables to compare views similar to naturally used comparisons such as side-by-side, shine-through and folding, covering all categories of Gleicher

et al.'s taxonomy. The approach, however, is limited to a small number of selected views to be used for a detailed comparison.

The contributions of this paper are based on previous work by Kehrer et al. [KPBG13], who proposed a formal model for structure-based comparisons in small-multiple displays. The following section describes the model in more detail. While their work focuses on a formal definition and application-oriented discussion of their model, this paper addresses some of their suggestions for future work: to provide an interactive specification of references for the underlying model and, in addition, to select appropriate visual encodings for resulting visualizations.

3. A Model for Structure-based Comparison

The model defined by Kehrer et al. is built upon the table algebra by Stolte et al. [STH02]. This algebra defines a matrix configuration T by a *cross operation* (\times). Using this operation, the matrix is partitioned into rows and columns by performing the Cartesian product of the categories of n dimensions X_1, \dots, X_n assigned to the x and y axes of the matrix (see Fig. 1):

$$T = X_1 \times \dots \times X_n = \{(t_1, \dots, t_n) | t_i \in X_i\}, \quad (1)$$

where $X_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,l_i}\}$ denotes a dimension with l_i categories. The composition of multiple dimensions, such as "Year" and "Quarter" in the x axis of Fig. 1, can be used to define a hierarchy of categories. Each cell in the resulting matrix is formally described by an n -tuple $t = (t_1, \dots, t_n)$, which represents a unique combination of the categories of the assigned dimensions.

The key idea of the model by Kehrer et al. is to define a reference pattern which uniquely specifies a referred cell for each cell of the small-multiple view. The pattern consists of a separate reference specification for each independent dimension of the view. For each dimension, the reference category can be inferred from the referring cell in three ways: using an absolute category for all cells (e.g., all cells refer to "Q1"), using the same category as in the referring cell (i.e. "self"-reference, e.g., cells where Quarter = "Q1" refer to "Q1", cells where Quarter = "Q2" refer to "Q2", etc.), or using the previous/next category for ordered dimensions (e.g., "Q2" cells refer to "Q1", "Q3" cells refer to "Q2", etc.). A reference pattern of the view in Fig. 1 could, for example, be (Product = self, Year = self, Quarter = "Q1"). A cell ("Furniture/Tables", "2014", "Q3") would thus refer to the cell ("Furniture/Tables", "2014", "Q1").

In Fig. 1, "Product Group" is a dimension used to hierarchically structure the categories of the specific products. Such nesting dimensions provide no independent information and are not part of the Cartesian product. They are consequently not part of the reference specification because their referred category is defined by their nested category (e.g., Product = "Tables" implies that Product Group = "Furniture").

DiffPin supports to interactively specify reference patterns containing absolute and self-references, which matches the most common use cases. It does not support to define order-dependent references. Extending the interaction to such references (e.g., based on gestures) will be an aspect of future work.

4. Interactive Reference Specification Using DiffPin

The idea for the interaction technique DiffPin was inspired by natural behaviour. Diffpin supports the metaphor of pointing at the screen, when suggesting to use a certain part of a small-multiple display visualization as reference for computed differences. Via drag and drop, a pin-shaped icon (see Fig. 1) can be placed directly onto different visual elements of a small-multiple display. That visual element defines for which dimensions which absolute categories are referred to. For example, dropping DiffPin onto the row label "Tables" specifies Product = "Tables" as absolute reference while all other dimensions are defined as self references. The following list characterizes different reference types which can interactively be specified by DiffPin:

- **Single cell.** Dropping DiffPin onto a single cell makes all cells refer to the selected cell for comparison (Fig. 1b).
- **Single row / column.** When dropping DiffPin onto a caption of a row / column, the corresponding category is used as absolute reference for all cells of the small-multiple display. In Fig. 1d, dropping DiffPin onto the row "Office Supplies / Labels" defines the reference (Product = "Labels", Year = self, Quarter = self), and allows to compare the average profit of each product to "Office Supplies / Labels" per quarter.
- **Row / column in context.** In Fig. 1a, the pin is dropped onto a category ("Q1") which is defined in the context of its parent dimension ("Year"). In such a case, the full context of that axis is used as reference by default ("2013 / Q1"). This is consistent with the behaviour for single rows / columns. Alternatively, users may define references without context, e.g., to compare the quarters of each year to the corresponding first quarter in the example (Product = self, Year = self, Quarter = "Q1"). In our implementation, pressing the modifier key *CTRL* while using DiffPin specifies references without context. The displayed table in Fig. 1 shows the result of the stated example.
- **Row / column with sub-structure.** Dropping DiffPin onto a row or column with a sub-structure (see Fig. 1c) enables to compare the entire sub-structure. For example, dropping DiffPin onto the caption "2013" (Fig. 1c) compares each quarter and each product to its corresponding counterpart in 2013 (Product = self, Year = "2013", Quarter = self).
- **Comparison between hierarchy levels.** It is sometimes necessary to compare categories to the aggregate of their parent category. In the model by Kehrer et al., such comparative scenarios require an intermediate node of the hierarchy, called *intermediate aggregates*. For example, comparing the average profit per quarter to the average profit of the whole year requires to represent "entire year" as an intermediate aggregate on the same level as Quarter. In our implementation, such aggregates can optionally be enabled per dimension, and serve as drop-target for DiffPin like any other category (see "Combined" category in Fig. 1 and Fig. 2). As a shortcut, dropping DiffPin onto the label of the Y axis defines a reference of each cell to the intermediate aggregate of the entire row (e.g. this would be the product in Fig. 1). Likewise, dropping DiffPin onto the label of the X axis defines a reference to the intermediate aggregate of the entire column.

The interaction technique DiffPin for defining references is independent of the chosen type of comparison. For example, the com-

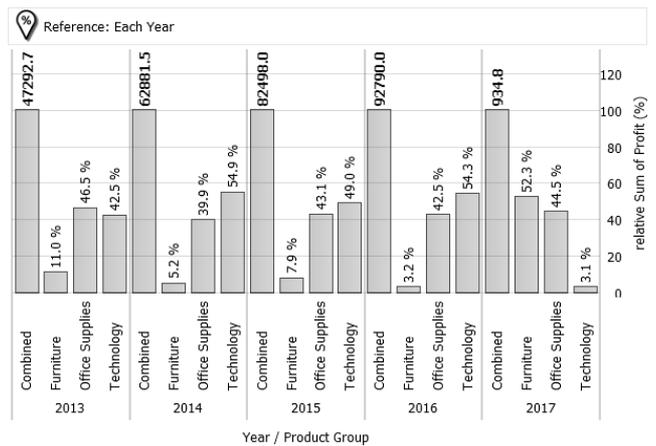


Figure 2: Bar chart comparing the sum of profit of each product group to the overall ("Combined") sum of profit per year in percent.

parison may be expressed as differences (Fig. 1) or percentages (Fig. 2). In Fig. 2, the intermediate aggregate of Product Group per Year (titled "Combined") is used as reference, thus representing 100 %. All other cells show their sum of profits in percent with respect to the intermediate aggregate of their reference.

5. Visual Encoding

Once a reference pattern has been defined using DiffPin, a subsequent challenge for comparative small-multiple views is to visually encode the structure of the comparison (i.e., "what is compared to what?") and the compared data as such. The following list summarizes our considerations about goals of the visual encoding:

- G1 Maintain the layout of the small-multiple view for all reference patterns.** Specifying references should not affect the layout. This is inherent considering that the definition of the layout is based on the table algebra which is independent of the reference specification in the model by Kehrer et al., but also fosters consistency for the user when rapidly switching between different comparison scenarios.
- G2 Provide a concise summary description of the reference pattern.** Describing the reference pattern in a human-readable way is helpful for explaining the effect of DiffPin during exploration and is typically necessary as caption text of diagrams in reports.
- G3 Facilitate an effective visual discrimination of reference cells from cells displaying relative information** such as differences or percentages. We emphasize that a discrimination is not possible with order-dependent references (e.g., each quarter referring to the quarter before), as in such cases each cell is both a referring and referenced cell. This was one reason why we excluded order-dependent references in this work.
- G4 Display original data for reference cells.** In addition to the cells displaying relative information (e.g., the difference), it is typically necessary to include the absolute data of the reference cells as important context information. Due to the heterogeneous characteristics between absolute and relative infor-

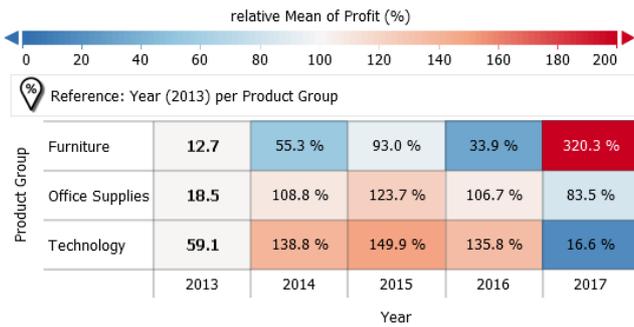


Figure 3: A heat map for a percentage-based comparison of average profits of product groups over multiple years, using the year "2013" as reference.

mation such as different scales, integrating both in one small-multiple view can be challenging in general.

To address these goals, we discuss strategies for visual encodings with a focus on text tables, bar charts and heat maps as examples. A comprehensive exploration of the overall design space is out of scope of this paper.

Encoding original and relative information with a common scaling is in general not effective. Therefore, we consistently use the relative scale for all cells, including the reference cells, and augment them with the corresponding original values by a text-label using a bold font-type. In Fig. 2, for example, all reference cells represent 100 % (encoded by the length of the bar), and their original values are displayed on top of each bar. This allows to visually discriminate reference values from differences (G3), as well as to keep context information in form of original values of reference cells (G4). Fig. 3 shows the same strategy for a heat map, where a divergent color map [War12] is used to encode percentages. This maps the central value "100 %" of reference cells to the neutral color white.

For summarizing the reference pattern, we display a textual description at the top of the visualization (G2, see examples in Fig. 1, 2 and 3). The description is composed as follows:

Reference: $d_1(c_1), d_2(c_2), \dots, d_n(c_n)$ per D_1, D_2, \dots, D_m ,

where d_i denotes a dimension where an absolute category (c_i) is used as reference, and D_j denotes a dimension where each cell refers to its own category of the dimension. In some cases, when using intermediate aggregates as references, we shorten this description. In Fig. 2, for example, the description "Reference: Each Year" is used instead of the description "Reference: Product Group (Combined) per Year".

6. User Feedback

For evaluation, DiffPin was deployed as new feature of our visual analysis system in industrial quality management. A team of seven experts had been using this system for a year and were familiar to the definition and visualization of pivot tables in the software. Frequent tasks of these users include the investigation of key perfor-

mance indicators (KPIs) of products in order to improve the product's quality, to reduce production rejects and, ultimately, to reduce production costs. An important use case in this scenario is comparing KPIs for plants, products, or materials, and to identify sources of found deviations.

Initial feedback from the users confirms the usefulness of DiffPin. After a 15-minute introduction to the feature during a supervised workshop, the domain experts were able to make use of the interaction technique to investigate their own data. By comparing KPIs of products, broken down by properties such as product type or production order, one expert instantly found an interesting pattern which he promptly forwarded to colleagues for further investigation. DiffPin allowed them to explore deviations in KPIs in much shorter time than using their previous tools such as Microsoft's Excel. For the first time, they were able to interactively explore differences in investigated KPIs, and to quickly generate a number of difference-visualizations which would otherwise have taken a whole day to produce. One user was especially enthusiastic about the new feature, stating "With a few mouse clicks to the desired results - thank you for this amazing feature!" in an e-mail two days after the workshop. Since then, DiffPin has stayed to be a frequently used feature until today, which is a period of four months.

The users still report that the flexibility of DiffPin covers their most important comparative scenarios. The lack of support for order-based references was not considered a shortcoming for their tasks.

7. Conclusion and Future Work

This paper proposes a new interaction technique called DiffPin, a draggable pin which can be used to specify references in small-multiple displays. We explain how DiffPin can be used to set different types of references directly in the visualization and reflect on goals for visualizing comparative views. Feedback from domain experts in industrial quality management confirms the usefulness of DiffPin for their daily tasks.

DiffPin can generally be applied to small-multiple displays based on the algebra described in Section 3. For example, integrating DiffPin into existing software such as Tableau [Tab] would conceptually be possible and facilitate current capabilities for specifying relative measures.

Future work includes the definition of additional interactions for DiffPin. We plan to extend the concept in order to cover all types of references as defined by the model of Kehrer et al. [KPBG13] (e.g., order-based references). In addition, future work includes a further exploration of the design space for comparative small-multiple views, as well as conducting a controlled user study.

Acknowledgements

This research is funded by the COMET K1 program - Competence Centers for Excellent Technologies (854174) by BMVIT, BMWFW, Styria, Styrian Business Promotion Agency - SFG and Vienna Business Agency. This work has also been supported by the Austrian Funding Agency (FFG) within the scope of the K-project DEXHELPP (843550). The COMET Programme is managed by FFG.

References

- [APP11] ARCHAMBAULT D., PURCHASE H., PINAUD B.: Animation, small multiples, and the effect of mental map preservation in dynamic graphs. *IEEE Transactions on Visualization and Computer Graphics* 17, 4 (2011), 539–552. 2
- [BBW16] BECK F., BURCH M., WEISKOPF D.: A matrix-based visual comparison of time series sports data. In *VMV* (2016). 2
- [GAW*11] GLEICHER M., ALBERS D., WALKER R., JUSUFI I., HANSEN C. D., ROBERTS J. C.: Visual comparison for information visualization. *Information Visualization* 10, 4 (2011), 289–309. 1, 2
- [JDKW15] JAREMA M., DEMIR I., KEHRER J., WESTERMANN R.: Comparative visual analysis of vector field ensembles. In *IEEE Conference on Visual Analytics Science and Technology (VAST)* (2015), IEEE, pp. 81–88. 2
- [KPBG13] KEHRER J., PIRINGER H., BERGER W., GRÖLLER M. E.: A model for structure-based comparison of many categories in small-multiple displays. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2287–2296. 2, 4
- [KSB*16] KRUEGER R., SUN G., BECK F., LIANG R., ERTL T.: Traveldiff: Visual comparison analytics for massive movement patterns derived from twitter. In *IEEE Pacific Visualization Symposium (PacificVis)* (2016), IEEE, pp. 176–183. 2
- [LKH10] LAMPE O. D., KEHRER J., HAUSER H.: Visual analysis of multivariate movement data using interactive difference views. In *VMV* (2010), vol. 10, pp. 315–322. 2
- [MHS07] MACKINLAY J., HANRAHAN P., STOLTE C.: Show me: Automatic presentation for visual analysis. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007). 2
- [RFF*08] ROBERTSON G., FERNANDEZ R., FISHER D., LEE B., STASKO J.: Effectiveness of animation in trend visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008). 2
- [STH02] STOLTE C., TANG D., HANRAHAN P.: Polaris: A system for query, analysis, and visualization of multidimensional relational databases. *IEEE Transactions on Visualization and Computer Graphics* 8, 1 (2002), 52–65. 2
- [Tab] Tableau software. <https://www.tableau.com/>. Accessed: 2018-02-26. 2, 4
- [TFJ12] TOMINSKI C., FORSELL C., JOHANSSON J.: Interaction support for visual comparison inspired by natural behavior. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2719–2728. 2
- [Tho02] THOMSEN E.: *OLAP solutions: building multidimensional information systems*. John Wiley & Sons, 2002. 1
- [vdEvW13] VAN DEN ELZEN S., VAN WIJK J. J.: Small multiples, large singles: A new approach for visual data exploration. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 191–200. 2
- [War12] WARE C.: *Information visualization: perception for design*. Elsevier, 2012. 4