# DiffMatrix: Matrix-based Interactive Visualization for Comparing Temporal Trends

Hyunjoo Song[1], Bongshin Lee[2], Bohyoung Kim[1] and Jinwook Seo[1]

[1]Seoul National University, Seoul, Korea
[2]Microsoft Research, Redmond, WA, USA

## Abstract

*Line graphs have been commonly used for visualizing temporal trends in time series data. Since comparing trends is one of the main tasks for analyzing multiple temporal trends, many efforts have been made to enhance visual representations of line graphs to help people efficiently compare multiple temporal trends. However, as the number of line graphs increases, the overlap makes it difficult to perform comparison and other analyses. In this paper, we introduce DiffMatrix, a matrix-based interactive visualization designed to support effective analyses of a large number of time series data. It employs four visual representations for each cell in the matrix to show the difference between two time series—dual lines, diff line, diff area, diff heatmap—and a detail view to support more indepth analyses on individual line graphs. DiffMatrix allows users to seamlessly switch between these representations that best support their tasks. We also report possible future work we identified through case studies with three real-world time series datasets with a large number of series.*

Categories and Subject Descriptors (according to ACM CCS): H.5.0 [Information Interfaces and Presentation (e.g., HCI)]: General—

## 1. Introduction and Related Work

In line graphs, comparing quantitative values on the curves can be done accurately because it is a position perception along a common vertical scale that is known to be the most accurate perceptual task [CM84]. Due to such a perceptual advantage, line graphs have been commonly used to visualize temporal data from multitudinous fields such as science, engineering, finance, management, and medicine.

However, a line graph is not the most appropriate choice for comparing two time series. One has to perceive the vertical distance between two curves to compare values in each series, which makes the task a much less accurate length perception than a position perception. Moreover, as shown in the curve-difference chart example [CM84], our brain is not wired to judge the vertical distance between the two curves, but to judge the shortest distance between them.

Since perceiving the difference between the series is one of the main tasks for time series data analysis [WJE10], it is important to visualize the difference in a more efficient way. It is possible to show the difference itself as a line graph, transforming the comparison task into the most ac-

curate position perception task. In this case, however, it is necessary to provide users with a way to access the corresponding source series. Even more challenging becomes the simultaneous comparative analysis task with many time series data. A single line graph with many series shown as multiple curves often suffers from severe overlap of the curves.

Efforts have been made to deal with this overlapping problem. One popular approach to address the issue is to make a line graph for each series (instead of showing all series in one line graph) and align them by time to spur comparison, often by using small multiples [Tuf83] [Tuf06]. While this approach resolves the overlapping problem, it makes the comparison task much less accurate because the curves do not share the same vertical scale any more. Also small multiples do not scale well unless a miniature representation (e.g., sparkline) is used because they take more space as the number of series increases. Stacked graphs [BW08] are also free from the overlapping problem. However, estimating difference values in a stacked graph is quite challenging because we have difficulty in even perceiving the individual values (i.e., vertical lengths) accurately for the same reason as in the curve-difference chart example. An-
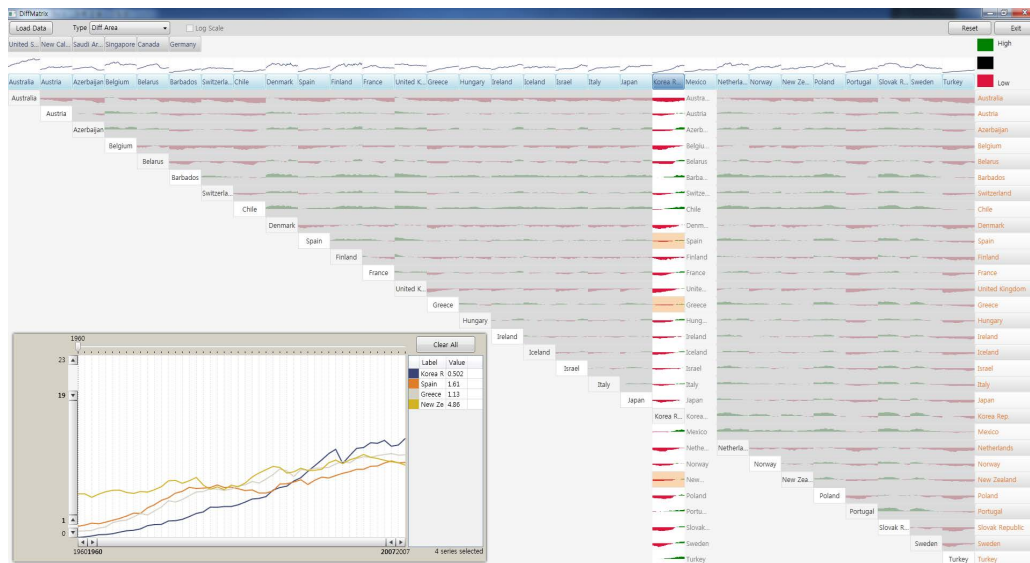
**Figure 1:** *DiffMatrix showing pairwise differences of $CO_2$ emission among 36 countries using "diff area" graph. The column for Korea Rep. is unfolded, and the selected temporal series (highlighted in orange background) are shown in the line graph detail view.*

other more scalable approach in term of the number of series to show simultaneously is to use color to encode the values [HLK07] [TSK05]. However, since the color perception is much less accurate than the length perception [CM84], this color-coding based approach is a less efficient choice for performing quantitative comparisons. There were some efforts to improve the scalability of line-based representations by using a color mapped angular histogram [ZGR11] or by using novel aggregated representations [EGS09] [KMH07]. There also have been several other studies involving about a teen of time series [JHA09] [WJE10]. However, the number of time series was very small and they often used a synthetic dataset to control variations for the controlled experiment.

In this paper, we present DiffMatrix (Fig. 1), a more scalable visualization based on a matrix representation that can deal with up to about 50 time series at once. It employs four visual representations in each cell using the small multiples technique—the standard line graph representation with two curves (dual lines) and three representations for showing differences directly (diff line, diff area, and diff heatmap), and a detail view to support more indepth analysis on individual line graphs. DiffMatrix allows users to seamlessly switch between these representations to best support their tasks.

## 2. DiffMatrix

DiffMatrix design was inspired by the human perception of the pattern of differences between curves in a line graph [CM84]. People tend to perceive minimum distance as difference between two curves rather than vertical distance.

Thus, showing difference directly using position encoding is a better way to help people correctly perceive the change of differences between two series. Therefore, DiffMatrix provides three visual representations—diff line, diff area, and diff heatmap—to display the difference of two series directly, as well as the one visualizing two individual curves in a cell.

### 2.1. Matrix-based Overview and Line Graph Detail View

We designed DiffMatrix based on a matrix representation with each column and row representing one series (Fig. 1). Each cell in the matrix view shows the difference between two corresponding series, one for the row and the other for the column. DiffMatrix provides a compact overview of all possible pairs with small multiples based on sparkline and heatmap, using the upper right triangular part of the matrix view. At the top of each column, it also shows the corresponding series as a sparkline to provide a visual reference to the original series. Since sparkline or heatmap visualization lacks the capability of showing data values in detail, DiffMatrix provides a detail view as a standard line graph to enable users to perform more targeted analyses on individual series, at the bottom left part. It can currently support up to about 50 series on a common wide screen monitor of a resolution of 1980x1080.

#### 2.1.1. Visual Representations for the Overview

DiffMatrix supports four visual representations (dual lines, diff line, diff area, and diff heatmap; Fig. 2) for each cell
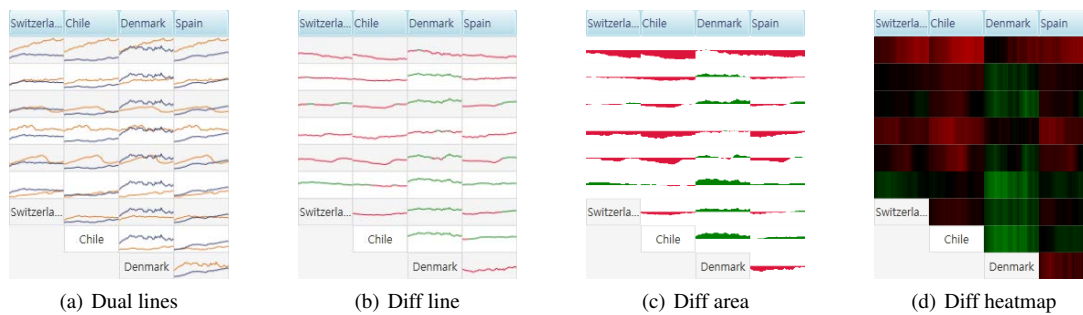
(a) Dual lines      (b) Diff line      (c) Diff area      (d) Diff heatmap

**Figure 2:** *Four visual representations for each cell in the matrix to help users see the difference between two time series.*

of the matrix view, allowing users to choose the most appropriate one depending on the tasks. They are embedded as sparklines or heatmaps for a compact display of the pairwise comparison results. While the standard line graph shows the two corresponding series (Fig. 2(a)), the other three representations directly show the difference derived from them. The difference is calculated by subtracting the column series from the row series. We designed these three difference representations to take the perceptual advantage of showing difference directly in supporting comparison tasks.

The three visual representations encode the difference values using a couple of visual variables. Diff line graph utilizes the most accurate position perception to show the actual difference values (upper half for positive and lower half for negative values) and the color perception to denote the sign of values (green for positive and red for negative) (Fig. 2(b)). Diff area graph also uses positional encoding in the same manner and uses the same color coding, but it fills the area between the curve and the baseline with corresponding color (Fig. 2(c)). Diff heatmap uses only colors to encode the actual difference values using a linear color mapping from red to green for minimum to maximum (Fig. 2(d)).

#### 2.1.2. Line Graph Detail View

DiffMatrix provides the detail view as a line graph to enable users to accurately recognize data values. Users can click on a cell to view the corresponding series in the detail view. They can click again on the selected item to cancel the selection. For multiple selections of series, DiffMatrix supports the standard multiple item selection mechanism (control/shift + click).

The detail view includes several controls to support tasks for retrieving data values or performing more targeted analyses on individual series. With the range sliders along the x- and y-axes, users can adjust the time duration of interest and the value range of interest, respectively. Since the detail view is tightly coupled with the matrix view, when users change the time duration of interest or the value range of interest, all the cells in the matrix are also updated accordingly.

The legend table, which is at the right side of the detail view, consists of three columns: assigned color for each series, series name, and the value at the time point of interest, which can be selected through the slider at the top of the detail view. Light grey vertical lines are also drawn in the background of the line graph to give a visual cue of time points, and the current time point is indicated using a black dashed line moving with the slider. Users can click on the column header of the legend table to sort it. They can also move the mouse over the curve of a series to retrieve the data value directly from the series at a specific time point through the tooltip.

Note that DiffMatrix is a superset of a standard line graph. For example, users can select all the sparklines for individual series from the top of each column, or select an entire row or column from the main matrix. In other words, they can see all the series in a line graph visualization but with additional functionalities described above.

### 2.2. Interaction and Animation

#### 2.2.1. Unfolding and Folding

As mentioned before, DiffMatrix uses only the upper right triangular part of the matrix view to show the overview of all possible pairs. Since the difference value is always computed by subtracting the column series from the row series, users have to follow an 'L' shape path to compare one series with all the other series. Furthermore, they have to mentally flip differences for the cells on the horizon part of the path. To address this problem, DiffMatrix allows users to unfold the 'L' shape path by clicking on the corresponding column label (Fig. 1). Users can fold it back to the 'L' shape by clicking again on the selected column label.

Upon selection, the label is highlighted with dark blue background color. In the meantime, all the cells on the horizontal part of the 'L' shape path for the selected series are moved to the lower left triangular part of the matrix view, being vertically aligned on the selected column. The difference visualization is properly flipped for the unfolded cells.

All unfolded cells are highlighted by graying out remaining cells. Similarly, users can select a series by clicking on the corresponding row label. In this case, all the cells on the vertical part of the path for the selected series are unfolded to the lower left triangular part of the matrix view to be horizontally aligned on the selected row.

### 2.2.2. Filtering and Reordering

When a large number of series are shown, users may want to filter out some series to focus on the series of their interest. DiffMatrix allows users to filter out series by dragging the corresponding column label out of the column header to the topmost row reserved for the filtered-out series (Fig. 1). Similarly, users can drag the column label from the topmost row to the column header row to restore the filtered-out series. While we only implemented the filtering capability for the labels in the column header, the result of filtering interaction keeps synchronized with the labels in the row header.

Since the ordering of the matrix can significantly influence the human perception of patterns in the matrix [Ber81], DiffMatrix allows users to reorder series in the matrix view. Users can reorder series with similar drag-and-drop interaction. They can simply drag the column label of a series to a different position. Since the order of series should be the same for row and column, changed order of series is synchronized between row and column. All the interactions mentioned above cause drastic changes in the visualization. To help users maintain their locus of attention and follow changes, DiffMatrix provides smooth animated multi-step transitions when significant changes happen.

## 3. Discussion and Future Work

We conducted case studies, analysing three real world time series data from World Bank and Eurostat websites—CO2 emission, government debt of EU countries, and inflation data—with DiffMatrix (Fig. 1). Each data consists of up to 50 series and 49 time points for each series at most and has different characteristics that help us emphasize various aspects of DiffMatrix. We can summarize the relative advantages of the three visual representations for difference perception: (1) the diff heatmap shows a least cluttered overview revealing the interesting spots and similar series effectively. (2) the diff area works well in identifying similar series while emphasizing the detailed trend better. (3) the diff line is good for revealing any subtle differnces.

We also identified the followings important issues that we think deserve further discussions and future work.

### 3.1. Scalability

To support realistic analysis, it is important to design and develop an interactive visualization system that can handle a broad range of real-world datasets that often have a large number of time series. There are two levels of scalability to consider in DiffMatrix: the number of time series and the number of time points in each series. While they are intertwined with each other, we believe that the former is more challenging than the latter. The sparkline curves and the heatmap representation we employed for each cell are compact and tolerant to a large number of time points in terms of showing an overview of each series. We could make the diff area representation even more scalable by adopting the idea of mirrored representation in the horizon graph [TSK05].

A large number of time series causes problem in system performance. For example, it will take more computing resources to animate a much larger number of cells. We could improve performance by using acceleration techniques such as parallelization with GPGPU. However, as the number of series increases, the number of cells in the matrix view quadratically increases, which puts more cognitive load on users. A possible remedy might be to aggregate similar columns or rows by clustering them.

### 3.2. L Shape Path for a Column

DiffMatrix computes the difference between two time series by subtracting the column series from the row series. Therefore, the matrix view is symmetrical about the diagonal, but two cells with reversed elements in the pair have same absolute difference values with a different sign, i.e., sign of [A (column), B (row)] = -sign of [B (column), A (row)].

We decided to use only half (i.e., upper right triangular part) of the matrix to spare space for the detail view. However, this requires users to follow an 'L' shape path when they want to compare one series with all the other series. To alleviate this issue, we provide users with the capability to unfold the 'L' shape path as described in section 3.2.1. However, this hinders users from identifying the interesting overall pattern pre-attentively. A possible remedy might be to enable users to switch between the full square matrix view and the current upper triangular matrix view.

## 4. Conclusion

Perceiving difference between curves in a standard line graph is not as easy as it seems. When there are many curves to compare, the standard line graph suffers from severe overlap of curves. In this paper, we propose showing the difference directly in a pairwise fashion using a matrix-based visualization with details shown in a line graph. We introduce DiffMatrix, an interactive visualization system that we designed to support the effective comparative analysis of large temporal datasets focusing on accurate and efficient perception of difference. DiffMatrix employs four visual representations for each cell in the matrix to show the difference between two series—dual lines, diff line, diff area, and diff heatmap—and a detail view to support more indepth analyses on individual line graphs.

**References**

[Ber81] BERTIN J.: *Graphics and graphic information-processing*. Walter de Gruyter, 1981. 4

[BW08] BYRON L., WATTENBERG M.: Stacked graphs - geometry and aesthetics. *IEEE TVCG (InfoVis 2008) 14*, 6 (2008), 1245–1252. 1

[CM84] CLEVELAND W. S., MCGILL R.: Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association 79* (1984), 531–554. 1, 2

[EGS09] E. GRUNDY M. W. JONES R. S. L. R. P. W., SHEPARD. E. F.: Visualization of sensor data from animal movement. *Computer Graphics Forum (EuroVis 2009) 28*, 2 (2009), 815–822. 2

[HLK07] H. LAM T. M., KINCAID R.: Overview use in multiple visual information resolution interfaces. *IEEE TVCG (InfoVis 2007) 13*, 6 (2007), 1278–1285. 2

[JHA09] J. HEER N. K., AGRAWALA M.: Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proc. CHI 2009* (2009), pp. 1303–1312. 2

[KMH07] K. MATKOVIC D. GRACANIN Z. K., HAUSER H.: Color lines view:an approach to visualization of families of function graphs. In *Proc. IV2007* (2007), pp. 59–64. 2

[TSK05] T. SAITO H. N. MIYAMURA M. Y. H. S. Y. H., KASEDA T.: Two-tone pseudo-coloring: Compact visualization for one-dimensional data. In *Proc. InfoVis 2005* (2005), pp. 173–180. 2, 4

[Tuf83] TUFTE E. R.: *The visual display of quantitative information*. Graphics Press, 1983. 1

[Tuf06] TUFTE E. R.: *Beautiful evidence*. Graphics Press, 2006. 1

[WJE10] W. JAVED B. M., ELMQVIST N.: Graphical perception of multiple time series. *IEEE TVCG (InfoVis 2010) 16*, 6 (2010), 927–934. 1, 2

[ZGR11] Z. GENG Z. PENG R. S. L. R. W., ROBERTS J. C.: Angular histograms: Frequency-based visualizations for large, high dimensional data. *IEEE TVCG 17*, 12 (2011), 2572–2580. 2