Eurographics Conference on Visualization (EuroVis) 2016
E. Bertini, N. Elmqvist, and T. Wischgoll
(Guest Editors)

Short Paper

Spiral Theme Plot

Shenghui Jiang[1], Shiaofen Fang[1] and Shaun Grannis[2]

[1]Department of Computer & Information Science, Indiana University Purdue University Indianapolis
[2]Regenstrief Institute

Abstract

We introduce a new visualization method for temporal data, Spiral Theme Plot, by combining ThemeRiver method, spiral patterns, and scatter plot technique. Similar to ThemeRiver, data in different categories (themes) are visualized in different bands, but also in a spiral pattern. Themes are stacked along a spiral curve, which represent the time axis. Individual data points are plotted within the regions of the themes, with various visual features. In addition to showing the overall theme patterns over time, this approach also shows plotting patterns within the themes. Compared to ThemeRiver, Spiral Theme Plot can accommodate longer time axis, and more importantly, can provide periodic patterns that are typically not available in ThemeRiver.

Keywords: Temporal data visualization, Theme River, Scatter plot.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Picture/Image Generation—Line and curve generation

1. Introduction

The visualization of time-varying data [TA15, AMST11, Sch11] has been an important way to detect patterns of events over time. Traditional time-varying data visualization focuses on showing either the attribute changes over time of individual data items (e.g. documents) or the overall trend over time of a collection of data items (theme). There are however other features in time-varying data that have not been sufficiently addressed. For example:

1) Given the changes from patterns of theme, it is often important to show the distribution of these changes within the themes that are related to individual data attributes.
2) Time dimension can be very long. Traditional straight time-axis does not provide the optimal space usage for long historical data.
3) Time-varying data often exhibit periodic patterns, which is not very well presented in most current approaches.

Healthcare data is a good example of applications of time-varying data visualization [GM11]. Although patterns of changes of the overall occurrences of different diseases are of great interest, the demographic distributions within these themes are often more important. For healthcare applications, it is very common to have very long period of historical data in order to build or validate a meaningful hypothesis. The ability to present the data span over a long period of time is therefore very desirable. Finally, periodic pattern is of critical importance for healthcare data as both seasonal patterns and deviation from seasonal patterns are important features which healthcare data analysts will look for.

Using healthcare data as a pilot study, we demonstrate a new time-varying data visualization technique using a spiral time-line, which can both stretch to a very long period of time interval, and represent periodic variations in the data.

In the rest of this paper, related work will be discussed in Section 2. In Section 3, we will describe our Spiral Theme Plot technique in detail. Some visualization results for healthcare data will be shown in Section 4. Conclusions and future work will be discussed in Section 5.

2. Related Work

The visualization of time-varying data is challenging because it is difficult to encode the time axis while still maintaining the data context. The simplest set of techniques are line chart [Pla05], which display observations on the y-axis against equally spaced time intervals on the x-axis, and their variations such as Horizon graph [HKA09], Time-series plot [Tuf83], and SparkClouds [LRKC10]. A comparison of these variations can be found in [JME10].

ThemeRiver [HHN00], or stacked graphs [BW08, DGWC10, Wat05], is another class of time-series visualization techniques. In a stacked graph, multiple series (themes) are represented as layers stacked one on another. The variation in the width of each layer represents the value changes of each series. It can provide an overview for trend of each individual series, as well as the comparisons of all the series. Enhancements of the stacked graph styled techniques have been applied in different time-series visualization applications, particularly in text visualization [SWL*10, CLT*11],

DOI: 10.2312/eurovisshort.20161170

by filling the empty space inside each layer with word clouds, or by adding splitting/merging branches between layers to show the inter-layer relations during their evolutions.

Spiral curves have been applied in visualizing time-series data, but were primarily used for generating periodic patterns of single data items [WAM01]. Other earlier work with simple use of spiral curves to represent time includes Berlin's Spiral Graph [Ber83], Spiral patterns in calendar visualization [MRD94], Pixel-based spiral patterns for database visualization [KK94], and spatiotemporal spiral event visualization [HH00].

3. Spiral Theme Plot

We developed a new time-series data visualization method called Spiral Theme Plot by integrating stacked graphs (ThemeRiver) [HHN00] and spiral patterns [WAM01] to plot time-series data as points in stacked spiral graphs. Time is represented as a base spiral curve. Multiple time series themes (for example, diseases or topics) are represented as stacked spiral graphs along the base spiral curve. Individual data points are plotted within the regions of the themes as points with various visual attributes, including radius, color and shape.

Spiral Theme Plot can represent time as a structured periodic variable, which presents intuitive visual clarity on time movement. Taking healthcare data as an example, we can represent diseases as themes, and plot patients as data points. This allows patient data during multiple years to be plotted periodically such that seasonal patterns or abnormal patterns for seasonal diseases can be easily detected. When plotting patient records, a key attribute, for example "age", will be represented as radius. Other attributes such as race and gender, can be represented as color and shape of the dots, respectively.

The base spiral curve is:

$$\begin{cases} x = r(\theta)\sin\theta \\ y = r(\theta)\cos\theta \end{cases} \quad (1)$$

Where $r(\theta)$ is a monotonic continuous radius function of angle θ. When $r(\theta)$ is a linear function $r(\theta) = a + b\theta$, the gap between spirals is a constant $2\pi b$, which can be estimated based on the maximum cumulative width of the themes (see Figure 1).

The width of the theme at a particular angle is determined by the total occurrences of the disease at that particular time. The boundary curve of each theme can then be interpolated by spline curves. This interpolation is done by splitting the time axis into a fixed number of segments. The maximum width of each segment is used as an interpolation point (See Figure 2). This leads to a discrete set of interpolation points from which the spline curve can be generated as the boundary curve of the theme. When plotting a point for each patient, the width of the theme needs to be computed first in order to determine the proper radius of the point. Although this width information can theoretically be computed from the spline representations, it is more efficient to simply check the color values along the normal direction of the spiral curve to estimate the width of a theme at each angle.

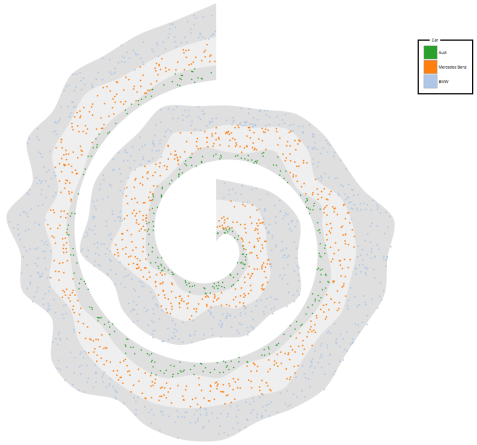

Figure 1: *A spiral theme plot of car sales data with 3 stacked themes.*

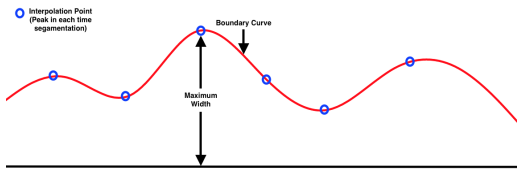

Figure 2: *Interpolation point defined by maximum width within each time segmentation.*

There are two potential issues that may need to be addressed in some cases. First, a uniform base spiral curve sometime may not be ideal when the width variation of themes is large. In these cases, a non-constant gap between consecutive spiral cycles may provide better spatial efficiency. These can be implemented by making the radius function $r(\theta)$ a non-linear function, which can be estimated by interpolating the maximum cumulative widths at different angles.

A second issue comes from the nature of spiral curves. For each theme, the area available for plotting the data points for a given angle interval within the theme becomes larger as the angle (time) increases, even if the width of the theme does not increase. This may lead to a wrong visual impression regarding the density of the data points at different times. Although this can be easily fixed by reducing the speed of the angle increase with time (i.e. angle does not move linearly with time), it is only applicable when the periodic property of the data is not important since angle and time are now not synchronized.

With these two potential issues in mind (and proper adjustments as needed), this technique can be very effective in visualizing large time-series data sets that may have periodic properties, and that may need closer explorations of the data points over a long period of time.

4. Healthcare Application

We have implemented and tested the Spiral Theme Plot technique using a large public health dataset: Notifiable Condition Detector (NCD) System. The Regenstrief Institute implemented and maintains an unparalleled HIE-based, automated electronic lab reporting (ELR) and case-notification system for over ten years in the State of Indiana. The NCD receives real-time HL7 version 2 clinical transactions daily, including diagnoses, laboratory studies, and transcriptions from hospitals, national labs and local ancillary service organizations. The NCD dataset contains 833,710 public health notifiable cases spanning more than 10 years from among 439,547 unique patients. In order to comply with the patient privacy policies and protocols, certain attributes of the data have been altered or perturbed.

Figure 3 shows the result of three diseases, Influenza, Malaria and Coronary artery disease (CAD), over a three year period (December 2009 to December 2012), along with the spiral time curve. From the visualization, we can see that these three diseases have clear seasonal patterns. For example, Influenza peaks in late winter and early spring. Malaria occurs mostly during late summer months, and CAD peaks during late fall or early winter. Mouse-over operations have also been implemented to highlight individual themes and details of individual patients. Figure 4 shows a highlight of CAD and a popup window shows details of one patient record. Figure 4 also shows that CAD patients are mostly older (concentration of dots near the outer boundary of the theme). Figure 5 shows a Spiral Theme Plot for HFMD (hand-foot-and-mouse disease), Mumps, and CAD over a 3-month (June to August) period in 2014. It shows a clear age pattern for all three conditions —younger patients for HFMD and Mumps, and older patients for CAD.

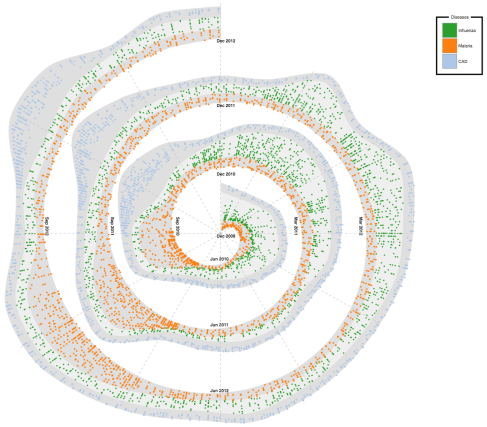

Figure 3: *The Spiral Theme Plot of Influenza, Malaria and CAD over a 3-year period shows seasonal patterns.*

One special feature of healthcare data is that the same patient may be included (e.g. hospital visits) multiple times for multiple diseases or for the same disease at different times. In these cases, it is desirable to show connections of these multiple occurrences from the same patient. Lines or curves are drawn to connect these plotted dots, as shown in Figure 6.

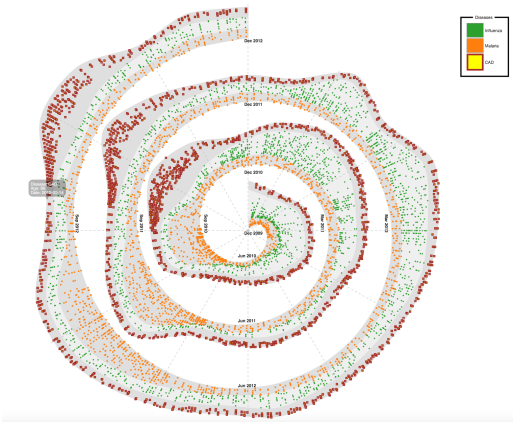

Figure 4: *Highlights of CAD patients show both seasonal and age patterns.*

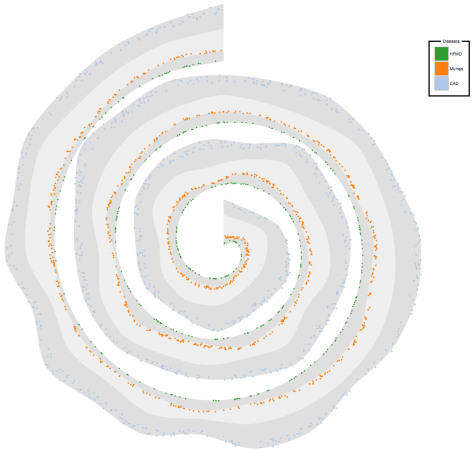

Figure 5: *The Spiral Theme Plot of HFMD, Mumps and CAD over a 3-month period shows age patterns.*

5. Conclusions

The Spiral Theme Plot technique is a combination of several information visualization methods including ThemeRiver, Spiral Plot and Scatter Plot. It provides a more comprehensive visualization for time-series data that shows not only the global trends of themes, but also patterns of individual attributes and the potential periodic patterns. For public health data with large patient databases, this particular combination satisfies several key requirements for visualizing time-series patient records. This technique has now been implemented as a tool in a web based healthcare data visualization system.

In the future, we would like to develop additional interactive operations associated with the Spiral Theme Plot for more sophisticated data exploration. We would also like to apply evaluation techniques to assess the effectiveness of this technique, particularly for healthcare applications.

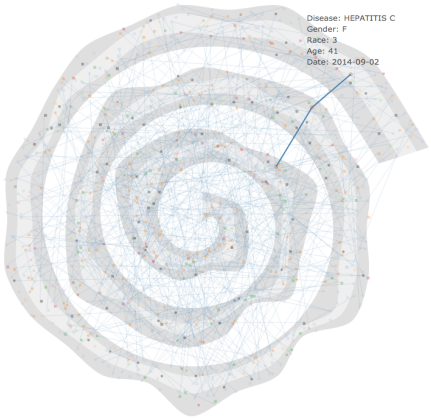

Figure 6: *A Spiral Theme Plot with connecting lines for multiple occurrences of the same patients.*

Acknowledgements

The project is supported by US Department of the Army, award W81CWH-13-1-0020. We gratefully acknowledge the contribution of Jennifer Williams; and the Regenstrief Institute's data engineers.

References

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of Time-Oriented Data (Human-Computer Interaction Series)*. Springer, 2011. URL: http://www.timeviz.net/. 1

[Ber83] BERTIN J.: *Semiology of Graphics*. University of Wisconsin Press, 1983. 2

[BW08] BYRON L., WATTENBERG M.: Stacked graphs – geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics 14*, 6 (Nov. 2008), 1245–1252. URL: http://dx.doi.org/10.1109/TVCG.2008.166, doi:10.1109/TVCG.2008.166. 1

[CLT*11] CUI W., LIU S., TAN L., SHI C., SONG Y., GAO Z., QU H., TONG X.: Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics 17*, 12 (Dec 2011), 2412–2421. doi:10.1109/TVCG.2011.239. 1

[DGWC10] DORK M., GRUEN D., WILLIAMSON C., CARPENDALE S.: A visual backchannel for large-scale events. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov 2010), 1129–1138. doi:10.1109/TVCG.2010.129. 1

[GM11] GROSSMAN C., MCGINNIS J. M.: *Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary*. The National Academies Press, Washington, DC, 2011. doi:10.17226/12912. 1

[HH00] HEWAGAMAGE K. P., HIRAKAWA M.: Pattern browser: Spiral-based interactive visualization using timelines on a geographical map. 1679–1686. 2

[HHN00] HAVRE S., HETZLER B., NOWELL L.: Themeriver: visualizing theme changes over time. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on* (2000), pp. 115–123. doi:10.1109/INFVIS.2000.885098. 1, 2

[HKA09] HEER J., KONG N., AGRAWALA M.: Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2009), CHI '09, ACM, pp. 1303–1312. URL: http://doi.acm.org/10.1145/1518701.1518897, doi:10.1145/1518701.1518897. 1

[JME10] JAVED W., MCDONNEL B., ELMQVIST N.: Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov 2010), 927–934. doi:10.1109/TVCG.2010.162. 1

[KK94] KEIM D. A., KRIEGEL H. P.: Visdb: database exploration using multidimensional visualization. *IEEE Computer Graphics and Applications 14*, 5 (Sept 1994), 40–49. doi:10.1109/38.310723. 2

[LRKC10] LEE B., RICHE N. H., KARLSON A. K., CARPENDALE S.: Sparkclouds: Visualizing trends in tag clouds. *IEEE Transactions on Visualization and Computer Graphics 16*, 6 (Nov 2010), 1182–1189. doi:10.1109/TVCG.2010.194. 1

[MRD94] MACKINLAY J. D., ROBERTSON G. G., DELINE R.: Developing calendar visualizers for the information visualizer. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology* (New York, NY, USA, 1994), UIST '94, ACM, pp. 109–118. URL: http://doi.acm.org/10.1145/192426.192470, doi:10.1145/192426.192470. 2

[Pla05] PLAYFAIR W.: *The Commercial and Political Atlas and Statistical Breviary*. New York: Cambridge University Press, 2005. 1

[Sch11] SCHULZ H. J.: Treevis.net: A tree visualization reference. *IEEE Computer Graphics and Applications 31*, 6 (Nov 2011), 11–15. URL: treevis.net, doi:10.1109/MCG.2011.103. 1

[SWL*10] SHI L., WEI F., LIU S., TAN L., LIAN X., ZHOU M. X.: Understanding text corpora with multiple facets. In *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on* (Oct 2010), pp. 99–106. doi:10.1109/VAST.2010.5652931. 1

[TA15] TOMINSKI C., AIGNER W.: The timeviz browser, 2015. URL: http://survey.timeviz.net/. 1

[Tuf83] TUFTE E. R.: *The Visual Display of Quantitative Information*. Graphics Press, 1983. 1

[WAM01] WEBER M., ALEXA M., MULLER W.: Visualizing time-series on spirals. In *Information Visualization, 2001. INFOVIS 2001. IEEE Symposium on* (2001), pp. 7–13. doi:10.1109/INFVIS.2001.963273. 2

[Wat05] WATTENBERG M.: Baby names, visualization, and social data analysis. In *Proceedings of the Proceedings of the 2005 IEEE Symposium on Information Visualization* (Washington, DC, USA, 2005), INFOVIS '05, IEEE Computer Society, pp. 1–. URL: http://dx.doi.org/10.1109/INFOVIS.2005.7, doi:10.1109/INFOVIS.2005.7. 1