Guided Visual Exploration of Cyclical Patterns in Time-series

Davide Ceneda*
TU Wien, Austria

Theresia Gschwandtner
TU Wien, Austria

Silvia Miksch
TU Wien, Austria

Christian Tominski
University of Rostock, Germany

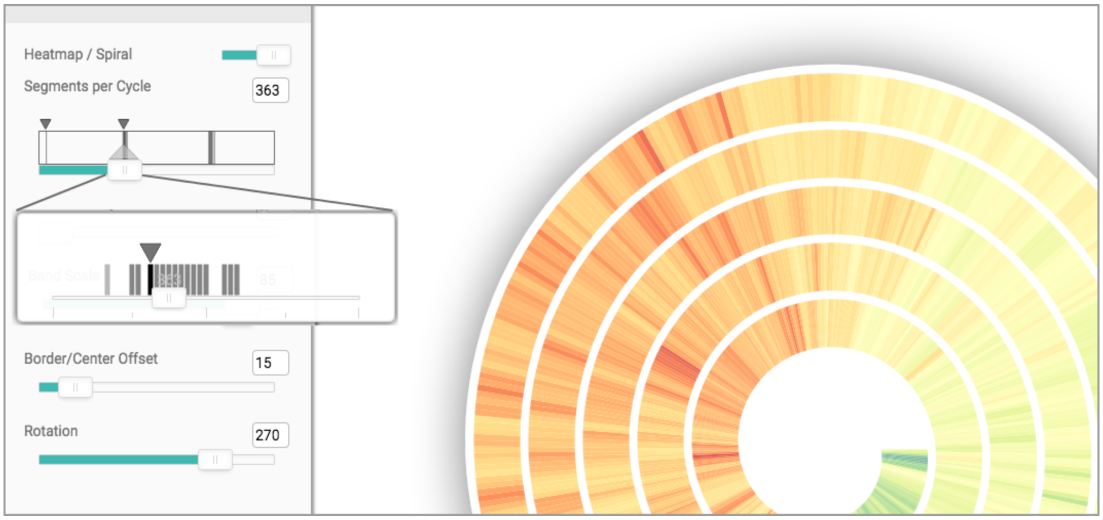


Figure 1: Our guided solution for the exploration of cyclical patterns in time-series. A classical spiral plot is enhanced with data-driven guidance mechanisms to support the identification of patterns. This figure is cropped to see the slider which is used to modify the cycle length displayed in the spiral plot. We statistically determine cycle lengths that reveal strong patterns and visually indicate these interesting cycle lengths while the user interacts with the slider.

Abstract

The analysis of cyclical patterns in time-series data can be challenging. Interactive visualization techniques allow the analysts to identify recurring behavior by exploring the data. However, this kind of analysis usually requires time and considerable mental effort, because it relies completely on the user. Alternatively, automatic statistical methods can be employed to identify periodicities in the data. However, different methods could yield inconsistent results, and analysts still need to investigate the data visually.

Guidance techniques have the potential to partially alleviate the burden on the user and leverage a more profitable analysis. In this paper, we describe the design and the implementation of a data-driven guidance technique to support the visual exploration of cyclical patterns in time-series. We use statistical results to support the visual analysis by providing guidance directly at the point of interacting with the visualization. In particular, we enrich a spiral plot with visual cues that suggest how the spiral must be configured to bring to light cyclical patterns. We evaluated our solution in a qualitative user-study, which showed that guidance can enhance the data exploration. The participants developed a deeper understanding of the data and had an increased confidence in the analysis outcome.

Index Terms: Human-centered computing—Visual analytics; Human-centered computing—User interface design; Human-centered computing—User studies; Mathematics of computing—Time series analysis; Information systems—Decision support systems;

*e-mail: davide.ceneda@tuwien.ac.at

1 Introduction

Visual data sciences enable the user to wisely take decisions by visually communicating data-derived information. This is usually achieved by a combination of statistical methods, computational models (e.g., machine learning and neural networks), and interactive visual means. In such situations, the correct choice of different visual and statistical parameters is central to a successful or a fruitless analysis.

A common example in which the configuration of the tool and of the visual representation is crucial to the success of the analysis is the exploration of time-series. The correct choice of parameters, e.g., the visual aggregation or alignment of data points, is fundamental to spot patterns and derive trends. Setting these parameters is not necessarily trivial and users may find it difficult to determine suitable values that allow them to progress with the task at hand. One way to help users in such situations is to provide guidance.

Guidance has been defined as a *"computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual analytics (VA) session"* [7, p.2]. Guidance does not solve problems for the users, but it helps them to overcome difficult situations on their own. This very definition makes it challenging to formally test or evaluate the effect of guidance methods. In fact, several guidance methods can be found in the literature, but only very little evidence is available for the benefits of guidance, particularly in the context of time-series analysis.

This work sets out to address this gap in the literature. The problem we focus on is the visual analysis and exploration of cyclical patterns in univariate time-series. Given this problem, our goal is to improve the analysis by providing guidance. We evaluate the effect of guidance by means of a qualitative user-study, which we used to test not only the improvements in terms of performance, but also the effects of the guidance on the user's mental state.

There are two alternatives for finding cyclical patterns in time-series. First, we can employ an algorithm and let it detect cycles for us. However, there are various algorithms and the results they deliver may not be consistent. Second, we can employ an interactive visual solution where cyclical patterns unfold before us if we configure the visual arrangement appropriately. However, the appropriate configuration is typically not known upfront and must be found manually, which may require lengthy phases of trial and error.

Our idea is to combine the two alternatives to a kind of *guided visual exploration of cyclical patterns*. We will use a spiral plot as the basic visual representation, because it naturally supports the detection of cycles. To assist in the detection, algorithms will compute candidates of cyclical patterns. This information will then be utilized to guide the user towards display configurations that will potentially make cycles visible in the spiral. To this end, we pick up the idea of scented widgets and embed visual cues into the user interface and the visualization. Users are free to follow the provided suggestions or their own intuition.

To achieve our original goal of finding evidence for the benefits of guidance, a qualitative study has been conducted. We hypothesized that guidance will be beneficial for actually finding cyclical patterns in the data and for the users' trust and confidence in the obtained analysis results. The feedback acquired during the study confirms that guidance has indeed a positive effect.

Taken together, this paper makes two main contributions. First, we enhance a spiral visualization with visual cues for guiding the visual exploration of cyclical patterns in time-series data. Second, we conduct a study that provides first evidence that guidance can positively influence visual data analysis activities. As far as we know, this is the first work that deals with designing and evaluating guidance for the analysis of time series.

2 Related Work

The detection and exploration of cyclical patterns in univariate time-series can be supported by visual representations on the one hand and by algorithmic methods on the other hand.

Visualization of Cyclical Patterns. Visualizing data has always been a natural way for spotting cyclical behaviors and patterns [1]. Visualizing these data in spiral form is a common solution for representing periodic time-series [10]. Cycles in the data are directly mapped to cycles of the spiral which allows an easy visual detection of patterns. Furthermore, spiral displays maintain the linear temporal progression of the data.

Carlis and Konstan [5] presented a first technique to represent different kind of time-series data on a spiral. Time points are presented as circular elements, whereas time intervals are shown as filled bars that indicate the start and the end of an interval. The technique also allows for the visualization of multivariate time-series. To this end, spirals are stacked in 2.5 dimensions and each attribute is displayed by colored bars. For large datasets, a spiral visualization may not fit the screen. Weber et al. [21] enriched complement spirals with a 3D helix visualization allowing an easy interaction and exploration (zooming and navigation) through time. Tominski et al. use 3D helix icons to represent cyclical data on a map [18]. While the x- and y-axes are used to encode the spatial dimension, the z-axis encodes time. Additional interactive means allow the inspection of selected spirals and the detection of patterns. A good overview of previous work on spiral representations is given by Tominski and Schumann [19]. They also suggest enhancing spiral displays with two-tone coloring to allow users to quickly and accurately read data values.

The spiral encoding is not the only means to visualize recurrent patterns. Van Wijk and van Selow [20] proposed a visualization technique to visualize temporal data on a calendar. The calendar view gives the user an overview, while detailed information and possible cyclical patterns are represented as stacked line charts. In a similar manner, Lammarsch et al. proposed GROOVE [13], which uses a pixel-based visualization in calendric structure to facilitate the detection of temporal patterns. Bach et al. [2] designed a foldable timeline to show similar data points close to each other, without losing the sequential temporal order of the data. Other authors recently explored the design of univariate time-series for different purposes: Bremer et al. [4] focus on storytelling, while Bögl focus on model building activities for event prediction. Finally, Zhao et al. [25] support the visual exploration of outliers and anomalies in time-series by using customizable lenses that transform, on the fly, the area under analysis. All these solutions have in common the design and implementation of different visualizations specifically tailored to a selected task. Besides the visual considerations, which we describe in the next sections, we focus on the design of guidance enhanced widgets to support the exploration of cycles in time-series.

Algorithmic Detection of Cyclical Patterns. Besides visual methods, cyclical phenomena in temporal data can also be computed by algorithmic means. In this context, the discrete Fourier transform (DFT) is commonly used [23]. It approximates the signal with a linear combination of basic functions. Similarly, the wavelet transform deconstructs efficiently a time signal into its different scale components [8]. The wavelet transform deals better with peaks and discontinuous functions, while the DFT works properly on stationary signals. A further statistical algorithm for detecting cycles is described by Sokolove et al. [16]. They construct a chi square periodogram (CSP) and showed its efficiency in spotting periodical and circadian rhythms.

Using algorithmic detection methods requires users being trained in choosing an appropriate method and setting its parameters properly based on the needs of the domain and the task. Furthermore, the algorithms' output often consists of long lists of numbers and probabilities, which are not easy to read and interpret.

Guidance. Providing guidance is one of the main aims of human-computer interaction, information visualization, and visual analytics [7, 9, 15]. A prominent example of a guidance mechanism is the concept of scented widgets [22], which influenced our work considerably. The basic idea is to improve the exploration process by showing information scents (e.g., previews of the data distribution or other users' interaction choices) directly in the visual interface.

The guidance technique presented by Baudisch [3] visualizes the presence and proximity of off-screen navigation targets by means of glyphs at the display border. A similar approach is also proposed by Zellweger [24]. The aim is to steer the analysis, giving the user the information about interesting data that is not yet visualized. A similar solution is presented by May et al. [14], where glyphs are used to indicate the shortest path towards unexplored meaningful data regions. Gladisch et al. [12] utilize flexible degree-of-interest (DoI) functions to support the navigation in hierarchical graphs. They use visual cues that not only indicate the navigation targets, but also why the system considers them relevant.

Considering existing techniques that support the discovery of cyclical patterns, two key issues remain to be addressed. Visual representations still need to be configured manually, while there is little to no indication on how to do it so that cyclical patterns can be revealed. When using algorithmic means there are different alternatives to choose from and it is unclear which algorithms to use and how to set their parameters. As a consequence, users often need several iterations of trial-and-error interactions until cyclical patterns are visually revealed.

To overcome these drawbacks, we do not propose new visual representations of cyclical data. Instead, we focus on enhancing the spiral visualization with guidance to get a better idea of the most prominent cyclical patterns and push the exploration forward. Our guidance solution constitutes a combination of the DFT and the CSP, which provides users with both, a comprehensive overview and

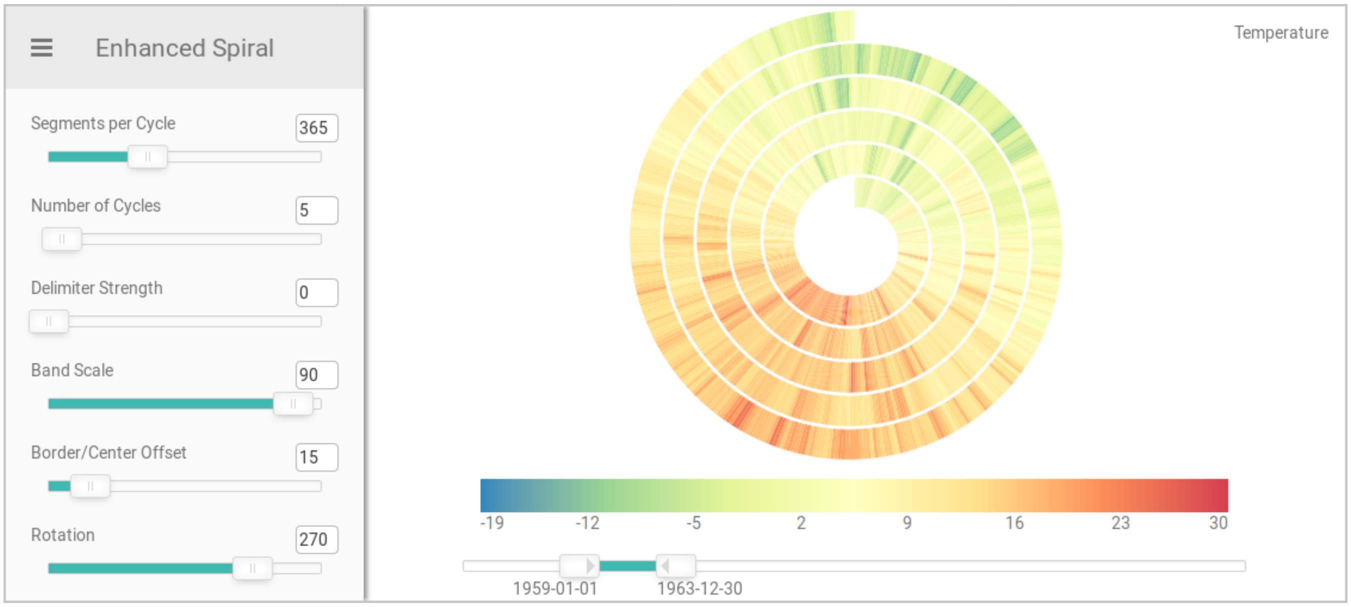


Figure 2: Spiral visualization (center) and user interface with sliders (left). The spiral currently visualizes a subset of 1.825 temperature values from a dataset with more than 25.000 days worth of data. There are 365 segments per spiral cycle to emphasize the yearly temperature fluctuation.

precise hints of the cycles directly at the point of interacting with the sliders that control the parameters of the visualization.

3 Guiding the Detection of Cyclical Patterns

We describe (i) the computational detection of cyclic patterns and (ii) the integration of visual cues to guide the user.

3.1 Visual Encoding and Task

The starting point of our approach is a visual representation that is capable of showing cyclical patterns in univariate time-series. In our case, we use a spiral [19]. The choice of the spiral is mostly due to the fact that there is a natural correspondence between cyclical patterns and the quasi-circular shape of the spiral. Furthermore, it was a minor design choice, as our main focus is on the design and provision of guidance. Alternatively, a rectangular calendar representation can be used as well (see Figure 3).

A spiral consists of a number of cycles and a number of segments per cycle. Each segment visualizes a time-dependent data value by means of color-coding. Figure 2 shows an example with five cycles and 365 segments per cycle, which corresponds to 1.825 data points. If the data consist of more data points than can be mapped to the spiral, the range selector at the bottom can be employed to specify which portion of the data should be shown.

The spiral's layout and its appearance can be adjusted via sliders as illustrated on the left-hand side of Figure 2. The available parameters include the number of cycles, the center offset, the width of the spiral band, and the rotation. Among the parameters, the number of segments per cycle is of primary importance in this work, as it determines the visual appearance of cyclical patterns.

In order to make a cyclical pattern visible, the number of segments per cycle must be adapted to the length of this pattern. Figure 4 show differently configured spirals, according to different cycle length values. As can be seen, visual patterns emerge only if the number of segments is set to 27 or 28, which matches closely the four-weekly cyclical pattern in the data.

So, when searching for cyclical patterns with the help of a spiral (or any other cyclical visualization), the user has to deal with the question of how many segments per cycle should be displayed so that the patterns become clearly visible. This question characterizes the knowledge gap that we seek to tackle by means of our guidance mechanism.

In summary, the task we want to support is the appropriate configuration of the spiral parameters, in particular the number of segments displayed in each cycle. Typically, a user would follow an interactive trial-and-error procedure: (1) adjust the the cycle length, that is, number of segments per cycle, (2) check the spiral for the emergence of cyclical patterns, (3) go back to step (1) unless all possible cycle lengths have been tested. However, exhaustively testing for all possible cycle lengths would be impractical as it would require a substantial amount of time and a considerable mental effort, especially when the data are composed of multiple overlapping cyclical patterns of different lengths. Furthermore, users may also overlook some of the patterns due to the increased cognitive load.

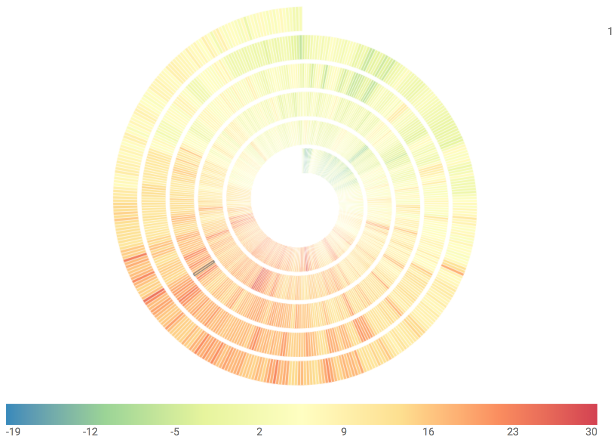
In the next section, we will describe in detail how the guidance mechanism is designed and implemented.

3.2 Guidance Design

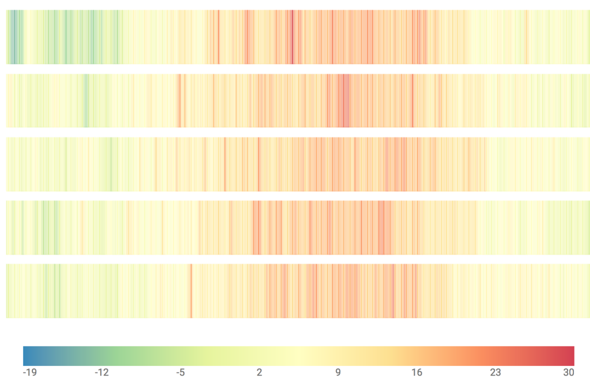
In line with the definition of guidance as given by Ceneda et al. [6, 7], we characterize our solution as an *orienting* support via visual cues that aims at solving an *unknown target* problem for which the knowledge gap is a *data* problem. The unknown target refers to the number of segments the users should visualize in each spiral cycle to easily spot cyclical patterns. We use the results of algorithmic computations as the *input* to this guidance process.

Algorithms The two algorithms we chose (DFT and CSP) are well suited to data that contain stationary cycles, i.e., the same cycles characterize the whole dataset. To better explain what results they deliver, we consider a temperature dataset with ca. 25.000 daily measurements, as used in Figure 2.

The DFT algorithm computes a discrete set of cycle lengths. For our data, it will likely compute a cycle length of 365, which corresponds to a yearly cycle. Other values may be computed as well, such as 30, which represents a monthly cycle. This is reasonable considering that the temperatures of a geographic area have a fixed alternation.

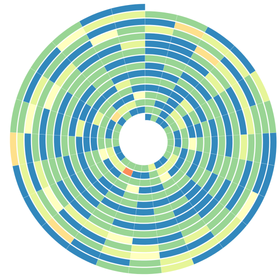


(a) Spiral visualization

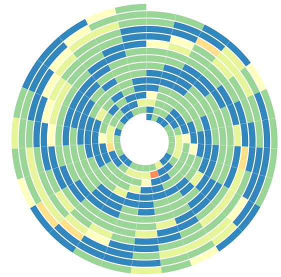


(b) Calendar-based visualization

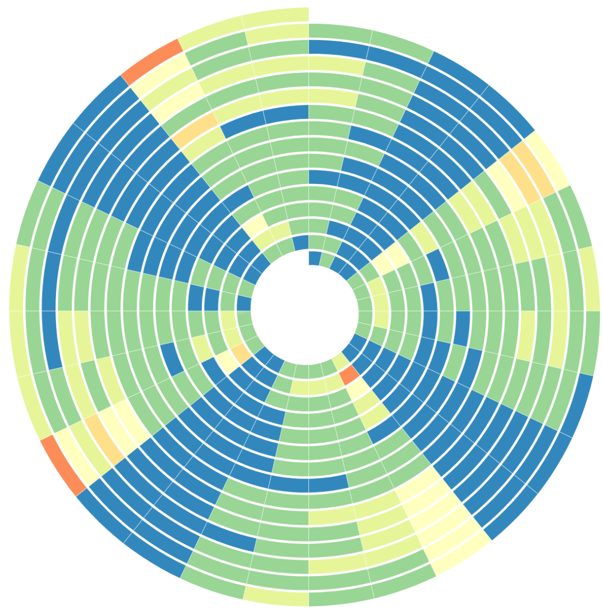
Figure 3: Spiral and calendar-based visualization of the same data and cycles. Our tool allow an easy on-the-fly switch between different visual encodings of the time-series. The cycle length value is set to 365 segments (days) and cycles clearly appear in both visualizations.



(a)



(b)



(c)

Figure 4: Visual exploration of cyclical patterns. Adjusting the number of segments per spiral cycle will make cyclical patterns visible. a) 25 segments: No patterns are visible. b) 27 segments: A cyclical pattern starts to emerge as segments begin to align in the spiral. c) 28 segments: A cyclical pattern is completely visible as the segments are fully aligned to form four-weekly cycles.

The CSP, on the other hand, produces candidates of cycle lengths and associates them with a probability. As the algorithm lists all possible cycle lengths, including those with a probability next to zero, we truncate this list to the most likely candidates ($p < 0.01$). For the temperature example, the CSP algorithm would assign high probabilities to cycle lengths close to 365 (e.g., [360..370]), or close to 30. Those values might represent oscillations of the cycle length, but might also reflect particular cases such as leap years.

Using DFT and CSP makes for a nice combination of algorithmic results. The DFT provides concrete indications, whereas the CSP complements them with probable suggestions. We chose CSP not only because it complements the results of the DFT algorithm, but also because it additionally reports recurring patterns that might also be worth investigating. For our example dataset, cycle lengths around 730 would have a high probability, which corresponds to a biannual cycle.

The value of our solution originates from the combination of the output of such algorithms. Note that domain-specific methods could be included as well. However, care must be taken not to generate too many suggestions, which could have a detrimental effect on guidance. Next we describe how the results of DFT and CSP are integrated and visually communicated to the user.

Communicating Guidance Suggestions. There are two options for communicating the guidance suggestions: Incorporate them in the main visualization (i.e., in the spiral) or in the user interface (i.e., in the sliders).

Sliders are a very common means of interaction and for this reason most users are familiar with them. In our case, sliders are the first and only means with which users can interact. Sliders provide direct feedback as the results of a parameter change are immediately reflected in the visualization. Sliders already incorporate a scale, provide an easy navigation, and inherently communicate concepts like direction and distance so that users know intuitively where to direct the interaction and how much to move the slider's handle. Finally, sliders can be easily customized to include additional information [11,22]. This makes them excellent user interface components to incorporate guidance.

Alternatively, guidance suggestions can be encoded directly in the main visualization. However, one must be careful not to interfere too much with the actual visual encoding of the data. In other words, our options of communicating guidance suggestions within the spiral are limited due to possible misinterpretation problems. Furthermore, keeping separated the visual encoding of the data and the encoding

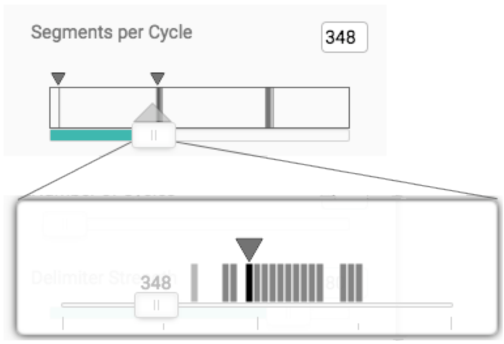


Figure 5: Guidance suggestions represented in a slider. The statistical output of CSP algorithm is used to delineate areas of interest while the different shades of gray/black encode probabilities. The output of the DFT algorithm is encoded with small triangles pointing to precise points within these areas. A further triangular shape handle, is used to show precisely the currently selected value. In the bottom part of the image, a zoomed view of the selection allows for a fine-grained value selection.

of the guidance allows an easier evaluation of our solution. For these reasons, we focused our effort towards including the guidance suggestions directly at the sliders.

Enhancing Sliders. We use the area located immediately above the sliders to encode the guidance suggestions. The visual encoding can be seen in Figure 5. We indicate the results of DFT and CSP with complementary visual cues, as the algorithms also complement each other. The output of the DFT algorithm is represented with a set of downward pointing triangles. If the DFT algorithm computes, for instance, yearly and monthly cycles, triangles at the top of the slider would point to the position corresponding to the cycle lengths of 365 and 30 days.

Additionally, we encode the interesting cycle lengths and their associated probabilities as reported by the CSP algorithm. The different cycle lengths are represented by bars (see Figure 5). The probabilities are grouped into high, medium, and low probability, and are encoded with a gray-scale from dark to light, where a darker gray represents a higher probability. A gray-scale encoding has been chosen to distinguish it from the color-coding of values in the spiral itself. As the participants to our evaluation will suggest, these visual cues efficiently communicate an intuitive overview of possibly interesting cycle lengths to look into.

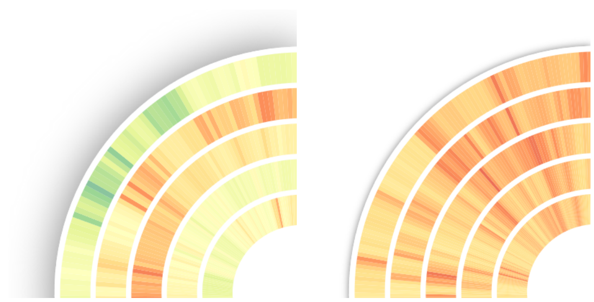
However, sliders usually do not have much display space at their disposal. This can make it difficult to discern the visual cues and select concrete cycle lengths. Therefore, we implemented a focus+context solution with which the user can access a larger and more detailed slider on demand. When operating the handle of the regular slider, a downward movement of the mouse cursor will trigger the appearance of a zoomed slider below the regular slider. As can be seen in Figure 5, the zoomed slider (bottom) shows only a subrange of the regular slider (top). Thus, it is easier to discern the different probabilities and the precise position of the suggested cycle lengths. During the interaction, control is handed over to the zoomed slider so that the selection of a cycle length can be done with increased precision. Both sliders are synchronized and the changes are immediately reflected in both of them.

In sum, our slider communicates where suitable cycle lengths are located in the parameter space and how confident the algorithms are about them, and it facilitates the precise interactive selection of the the indicated cycle lengths.

Enhancing the Spiral While the enhanced slider already communicates the suggested cycle lengths quite well, it is still necessary to look back and forth between the spiral and the slider to visually confirm if cyclical patterns do indeed emerge for a selected cycle length. To facilitate this task, we added a simple visual cue directly into the spiral.

In particular, we add a glow to the spiral border to indicate how far the current configuration is off from the closest suggested cycle length. If the currently set number of displayed cycles is far away from the one suggested by the algorithms, the glow will have a larger radius to make it more blurry. When approaching a suggested cycle length, the radius will decrease to the point where the glow is very thin and sharp to indicate that a suitable configuration has been reached. As illustrated in Figure 6, the glow provides additional guidance, but only minimally interferes with the actual visualization of the data.

We think that users could benefit from this additional visual cue by getting a glimpse of the closeness of interesting cycles while simultaneously interacting with the slider. Moreover, the glow serves as a visual indicator even before cyclical patterns start to form in the spiral.



(a) Wide glow: Far off. (b) Narrow glow: good match.

Figure 6: The distance to a suggested cycle length candidate is encoded as a glowing ring surrounding the spiral. a) A wide glow indicates that the current configuration is far off from a suggested cycle length. b) A narrow glow suggests that number of segments per cycle matches the cycle length in the data.

3.3 Implementation

The visualization and the guidance are coded in plain javascript, using html canvas elements to draw the output generated by the two selected algorithms. To support the computation of interesting cycles, the system communicates with a back-end server that manages the datasets and runs the algorithms. The back-end is written using the R language for statistical computing [17] and different extensions (Rook, and Rjson) to manage the net communication and the encoding of the messages. The DFT algorithm is also implemented in R, and is part of a package for the analysis of time-series. On the other hand, we found a standalone implementation of the CSP which we adapted to fit the characteristics of our datasets.

The evaluation and its results are described in the next section.

4 Evaluation

To evaluate the effectiveness of our guidance solution, we conducted a semi-structured qualitative study. We asked the participants to solve one task with two different datasets, at first without assistance and then supported by the guidance suggestions as introduced before. We collected qualitative feedback, in written form, after each phase to understand how the users performed under the two different conditions. We conducted the evaluation having in mind the following hypotheses:

H1 The guidance mechanism brings benefits to the analysis in terms of an increased trust and confidence of the user.

H2 The implementation of guidance relieves the users from mental workload, allowing them to focus their attention on reasoning and on confirming work hypotheses.

In literature, there are many ways to evaluate a given approach. However, this is not true for guidance, as there are no methods to state if the provided guidance is appropriate. The novelty of our evaluation comes from the integration of classic metrics (execution time, correctness) with the evaluation of how much the visual integration of guidance in the spiral affected the users mental state (i.e., users trust and confidence). In summary, we wanted to evaluate not only if the designed guidance had positive effects on the performances but also if the guidance was in line with the users way of reasoning, if the guidance distracted the user, if it facilitated insight discovery, and the users trust and confidence towards the guidance suggestions.

Study Participants. We recruited six visualization experts. The study participants we recruited are all pursuing their doctorate in the field of visualization/visual analytics and are familiar with the spiral plot as well as the interaction means we used in the experiment (i.e., the sliders). However, besides their participation in the user study, they were not involved in any phase of the design nor in the development of our solution. We chose six experienced users because they are representative users we think should use the tool. Furthermore, they already knew the visualization metaphor (i.e., the spiral) we used as base for the guidance. In this way the users were not distracted by the main visualization, but focused their attention (and provided useful feedback) on the guidance suggestions.

Datasets. We used two datasets for the study, one with real data and one with artificial data. The first dataset contains information about the weather in a German city. It comprises about 25.000 entries and different measurements, like humidity, air pressure, or quantity of precipitation, from which we chose the temperature measurement for our study. The dataset spreads over multiple years and contains multiple cycles (monthly, yearly, biannual). The interface and the dataset are shown in Figure 2 where a yearly cycle is displayed. We chose this dataset, because it represents a typical scenario in which data science is usually applied.

We created the second dataset artificially by modeling a sinus wave with a cycle length of 13 days. We use this artificial dataset to simulate the case of cycles not following exactly the common calendar subdivision. We further introduced a certain amount of white noise into the data to avoid that the cycles are too easily recognizable at first glance. The noise is introduced to simulate real measurements and to bring the visual discovery of cyclical patterns in line with real world datasets.

The implemented algorithms worked well for both datasets used in the study. Moreover, since we knew the cycles and the respective lengths contained in both datasets we could use them as a ground truth for evaluating the findings of the study.

Procedure. We subdivided the study participants randomly into two groups and asked all of them them to solve the same task with the two different datasets and under two different conditions: with and without guidance. To mitigate learning effects, the order of the dataset was switched in the two groups. So, one group performed the task first without guidance using the *weather dataset* and then with guidance using the *artificial dataset*. For the second group, the datasets were switched. A graphical representation of the study procedure is portrayed in Figure 7.

At the beginning of the study, all participant received a short introduction, and they had the possibility to interact with the system to get familiar with the operation of the sliders. After that, we presented them the following task:

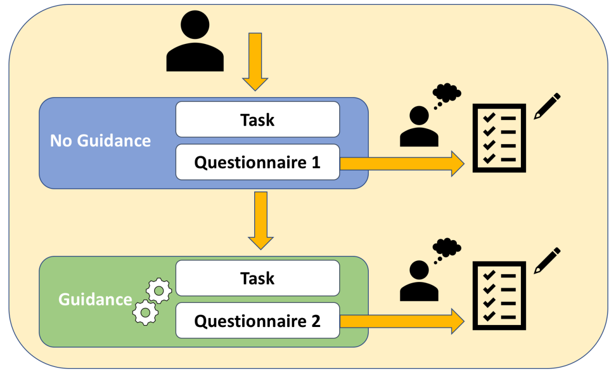


Figure 7: The procedure of the user study. We asked the participants to perform the same task under different conditions (guidance/no guidance). After each execution we asked questions regarding their trust and confidence, to understand if guidance improved the analysis.

> *Find all the cyclical patterns present in the dataset and report, for each of them, the cycle length. You are also encouraged to report your thoughts about the results you find (e.g., is it a relevant cycle?) and to think aloud.*

They had to find all the cyclical patterns in the dataset and possibly reason about their relevance (e.g., the relevance of a cycle that is a multiple of another cycle may be considered less relevant). We further asked them to think aloud and explain their reasoning while pursuing the task. An external person sat next to the participants taking notes, but without interfering with the execution.

The task was at first carried out without guidance support. Hence, the study participants could interact with the sliders but did not receive any suggestions of possibly interesting cycle lengths. Then we asked them three questions to get a first idea about how they performed the analysis and get their impressions. We asked them 1) if whey followed a specific strategy 2) if they felt confident about the results they found and 3) if they believed they could detect all relevant cycles.

In the second phase of the study, the participants performed the same task but on the other dataset. This time the guidance suggestions were enabled to assist the data analysis. Before starting the task, we also gave them a short introduction on the encoding of the guidance suggestions. After this second phase, we asked them six questions targeted at discovering if the guidance affected the way they performed the task. In particular, we asked them 1) if they followed a different analysis strategy (compared to the first phase) and if the guidance influenced their choice. We wondered 2) whether they felt more confident about the results they found and 3) if they felt that they may have missed some important results. We further investigated 4) the trust the users had in the analytical algorithms and asked them 5) if they felt that the suggestions were leading them into unwanted analysis paths, towards wrong results, or 6) if they found the suggestions to be useful to solve the task and complying their way of reasoning. Finally, we collected comments about the guidance solution and the design choices.

4.1 Results

From the questionnaires and from the observation of the participants, we can derive how they performed under the two conditions. In particular, the questionnaires allowed us to evaluate the effects of the guidance and how it changed the analysis strategy. Summarizing, the study shows that our combination of visual and algorithmic means is effective and has a positive impact on the data analysis. In the following, we report the main findings.

Detection of Cyclical Patterns. Although we did not focus our evaluation on correctness of task nor completion times, we noticed that when supported with guidance the users found more cycles and, more importantly, reasoned more about the results. We noticed they were more inclined to formulate hypothesis e.g., what phenomenon was reported in the data, and to reason aloud (H2). With guidance, they were able to rate the relevance of the cycles they found, order them, and reason about recurrences and multiple cycles. Without guidance they had more approximated answers and in most of the cases did not find all the cycles.

Analysis Strategies. Without guidance, the participants followed quite closely the trial-and-error workflow outlined in Section 3.2. When guidance was provided, they changed their strategy.

Without any support, the main way of solving the task was characterized by first getting an overview of the dataset followed by an exploration phase during which different values for the cycle length were tested one by one. After this exploration, the users proceeded with a deeper inspection of a selected group of cycle length values. Some of them explored the values from the highest to the lowest, while others performed it the other way around. Without guidance the analysis was mainly characterized by a thorough exploration of all the possible cycle lengths, followed by a confirmation phase in which the most promising cycle lengths were inspected in detail.

The introduction of guidance had substantial effect on the analysis strategy. Users spent most of the time evaluating the cycles suggested by the algorithms instead of exploring all the different cycle length values. We further noticed that the participants developed a deeper understanding of the data. They reasoned more about the possible meaning of the phenomena described in the data and developed more research hypotheses than without guidance (H2). This change of strategy was reported by the participants in the questionnaires.

Confidence and Trust. Although all participants were visual analytics experts, the majority of them reported an increased confidence in the results they found when guidance was enabled (H1). They also reported to have the impression to complete tasks faster and in an easier way, although we did not record timings precisely.

The majority of them reported to have trust in the guidance suggestions (H1). However, we noticed that when guidance was provided, the confirmatory analysis of the suggested cycle length was usually followed by a further exploration of the data. Aiming to investigate in detail this behavior, we asked follow up questions regarding the participants' trust in the guidance suggestions (questions 4 and 5). For most of them, the main problem in trusting the provided suggestions was the lack of knowledge about the algorithms that led to these suggestions. Some stated they felt the guidance may have missed some findings or did not display all the important results. This explains why they continued the analysis and did not just rely completely on the provided suggestions.

However, the participants also reported that their confidence in the guidance mechanism increased after a certain time spent in the analysis. This feedback confirms the positive effects of guidance (H1). It also confirms the general advantage of combining algorithms and interactive visual means, which is the core idea behind visual data science. On the other hand, it shows that the guidance suggestions should be effectively encoded in order to be trusted.

Visual Encoding. The visual encoding was appreciated by all but one participant. All understood the slider encoding and used it effectively. Also the gesture to activate the detailed view of the slider was well accepted. One user reported that he would have preferred the suggested cycle length candidates visualized directly at the spiral, while all the others preferred the sliders and did not mind to jump back and forth between the slider and the spiral when interacting. A couple of users stated that the guidance suggestions were too small and all agreed that bigger sliders would have improved the interaction even more. However, they also stated that the focus+context visualization (the zoomed slider) effectively improved the interaction.

Distance Encoding. The most controversial part of the encoding was the glow around the spiral that we used to encode the distance to a suggested cycle length. While the majority of the users evaluated positively the provision of visual cues in the sliders, just one third of them actually used the glow for solving the task. Most of the users noticed this cue only after interacting with the visualization for a while. We observed that when using the glow, the exploration of the dataset proceeded faster. Participants were also more precise in their interactions, when moving towards suggested values, and gained awareness fast when the distance was growing.

Other observations. Five out of six participants reported that they did not have the impression that the suggestions were misleading, nor that they contained errors. One of the participants reported that this kind of guidance perfectly fits the task of finding the length of the cycles, while it may be inadequate for more complex tasks. Another one reported that, when guidance was provided, he lost some time verifying the correctness of a suggestion, while he was already almost certain that it was not worth following that hint. Finally, a user stated that he felt constrained by the guidance, although the guidance itself did not really impose any constraints to the interaction.

5 Discussion and Future Developments

In this section we discuss the findings of the user-study and describe possible extensions of our guidance solution.

Guidance Encoding. From the feeback we received, we learned that the design choice of providing visual guidance directly in the slider was very well understood and appreciated by the study participants. However, they also highlighted the importance of having visual hints directly on the spiral. In particular the border, may be well suited for incorporating further cues to steer the exploration process, thus increasing the degree of guidance. For example, it could be used to encode the direction of the movement, or to visually combine the encoding of the distance with the encoding of the guidance. However, we reserve this idea to future extensions.

Cyclical Data Encoding. We chose a spiral display to visualize recurring patterns. Spirals are commonly used for this task and the study participants were already familiar with this visual encoding. Alternative choices, would have included heat-maps and calendar views. In our study, the users had the possibility to change visual metaphor, among the cited ones, but nobody did it. However, we did not test whether the study participants preferred one encoding over the others, as our main focus was on evaluating the usefulness of our guidance design. We think that the results of this design might be valid for different visualization types, as they influence mainly the interaction and where to look for cycles. However, further research is needed to verify this hypothesis and test whether our solution may be generalized. Moreover, we designed and evaluated our solution using univariate time-series data only. Although we think that this is not a limitating factor, another research direction would be to investigate solutions for multivariate time-series which might demand for different visualizations.

Distance Encoding. Visualizing the distance to the next suggested cycle length by a glow around the spiral was controversial. Some of the study participants evaluated this encoding positively as a hint to fine-tune their selection with the help of the detailed slider. However, the majority of study participants found it either less useful, did not use it all, or used it only after a while. Since advantages of using the main visualization for guidance suggestions were frequently mentioned, we still believe that there is considerable potential here. However, the design space of how to visualize guiding cues is huge and constitutes an interesting direction for further research.

Calendars and Granularity Encoding. A couple of participants reported difficulties understanding the granularity of the datasets. Both datasets used for the evaluation contained daily measurements. However, the operation of continuously modifying the cycle length, in search for patterns, was interpreted as a temporal aggregation by two study participants. By now, we do not support the modification of the displayed granularity nor the use of calendars. However, these two additions could significantly improve the understanding of the dataset and of the guidance suggestions. On the other hand, this would add up to the number of visualization parameters that need to be set by the user, and thus, would add complexity to the task. Nevertheless, we believe that these might be valuable additions and we carefully consider different designs to integrate these features, such as an overlay visualization of a calendar view onto the main spiral visualization.

Highlighting Related Patterns. One of the suggestions coming from the study participants was to further sort and filter the results of the analytical algorithms. In particular, for large datasets it may happen that the probabilistic algorithm finds a set of cycles repeated over the dataset. In this case, the visual guidance in the sliders can easily get cluttered with these repetitions of multiples, although the number of relevant cycles is low. One solution to the aforementioned problem could be an automatic sorting and pre-selection of the most important cycles. However, this way we might miss some important yet repeated patterns. A better solution would be to visually highlight multiple related patterns subsequently to the selection of one of them, to clearly communicate their relation.

6 Conclusion

We described the design and implementation of data-driven guidance to detect and support the exploration of cyclical patterns in time-series data. A qualitative user-study confirmed our hypotheses and provided first evidence of the added value of this guidance support. Another valuable outcome of this study is that it sheds light on some sensitive issues in the context of guidance that require special consideration (such as not pointing users to irrelevant exploration paths). Still, our findings suggest that guidance has the potential to improve analysis outcomes by unburdening users from repetitive browsing tasks so they can focus on a deeper exploration of interesting phenomena in the data. Moreover, our study participants reported to be more satisfied with and confident in the patterns they found when guidance was provided. We conclude that our work presents initial insights into the value of guidance in visual data sciences and point to a number of further research directions in this field.

Acknowledgements

This work was supported by the Centre for Visual Analytics Science and Technology CVAST, and funded by the Austrian Austrian Science Fund (FWF), grant P31419-N31.

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer, 2011. doi: 10.1007/978-0-85729-079-3

[2] B. Bach, C. Shi, N. Heulot, T. Madhyastha, T. Grabowski, and P. Dragicevic. Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE transactions on visualization and computer graphics*, 22(1), 2016.

[3] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 481–488. ACM, 2003.

[4] M. Brehmer, B. Lee, B. Bach, N. H. Riche, and T. Munzner. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE transactions on visualization and computer graphics*, 23(9):2151–2164, 2017.

[5] J. V. Carlis and J. A. Konstan. Interactive visualization of serial periodic data. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, pp. 29–38. ACM, 1998.

[6] D. Ceneda, T. Gschwandtner, T. May, S. Miksch, M. Streit, and C. Tominski. Guidance or No Guidance? A Decision Tree Can Help. In C. Tominski and T. von Landesberger, eds., *EuroVis Workshop on Visual Analytics (EuroVA)*. The Eurographics Association, 2018. doi: 10.2312/eurova.20181107

[7] D. Ceneda, S. Miksch, T. Gschwandtner, H.-J. Schulz, M. Streit, T. May, and C. Tominski. Characterizing guidance in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):111–120, 2017.

[8] K.-P. Chan and A. W.-C. Fu. Efficient time series matching by wavelets. In *Data Engineering, 1999. Proceedings., 15th International Conference on*, pp. 126–133. IEEE, 1999.

[9] A. Dix, J. Finlay, G. Abowd, and R. Beale. *Human-Computer Interaction*. Pearson Education, 3rd ed., 2004.

[10] G. M. Draper, Y. Livnat, and R. F. Riesenfeld. A survey of radial methods for information visualization. *IEEE transactions on visualization and computer graphics*, 15(5):759–776, 2009.

[11] S. G. Eick. Data visualization sliders. In *Proceedings of the 7th Annual ACM Symposium on User Interface Software and Technology*, UIST '94, pp. 119–120. ACM, New York, NY, USA, 1994. doi: 10.1145/192426.192472

[12] S. Gladisch, H. Schumann, and C. Tominski. Navigation Recommendations for Exploring Hierarchical Graphs. In G. Bebis, R. Boyle, B. Parvin, B. Koracin, Darkoand Li, F. Porikli, V. Zordan, J. Klosowski, S. Coquillart, X. Luo, M. Chen, and D. Gotz, eds., *Advances in Visual Computing: Proceedings of the International Symposium on Visual Computing (ISVC)*, vol. 8034 of *Lecture Notes in Computer Science*, pp. 36–47. Springer, 2013. doi: 10.1007/978-3-642-41939-3_4

[13] T. Lammarsch, W. Aigner, A. Bertone, J. Gärtner, E. Mayr, S. Miksch, and M. Smuc. Hierarchical temporal patterns and interactive aggregated views for pixel-based visualizations. In *Information Visualisation, 2009 13th International Conference*, pp. 44–50. IEEE, 2009.

[14] T. May, M. Steiger, J. Davey, and J. Kohlhammer. Using signposts for navigation in large graphs. *Computer Graphics Forum*, 31(3pt2):985–994, 2012. doi: 10.1111/j.1467-8659.2012.03091.x

[15] S. Smith and J. Mosier. Guidelines for designing user interface software. Technical Report ESD-TR-86-278, Mitre Corporation, Bedford MA, 1986.

[16] P. G. Sokolove and W. N. Bushell. The chi square periodogram: its utility for analysis of circadian rhythms. *Journal of theoretical biology*, 72(1):131–160, 1978.

[17] R. C. Team. R: A language and environment for statistical computing. r foundation for statistical computing, vienna, austria. 2013, 2014.

[18] C. Tominski, P. Schulze-Wollgast, and H. Schumann. 3d information visualization for time dependent data on maps. In *Information Visualisation, 2005. Proceedings. Ninth International Conference on*, pp. 175–181. IEEE, 2005.

[19] C. Tominski and H. Schumann. Enhanced interactive spiral display. In *SIGRAD 2008. The Annual SIGRAD Conference Special Theme: Interaction; November 27-28; 2008 Stockholm; Sweden*, number 034, pp. 53–56. Linköping University Electronic Press, 2008.

[20] J. J. Van Wijk and E. R. Van Selow. Cluster and calendar based visualization of time series data. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, pp. 4–9. IEEE, 1999.

[21] M. Weber, M. Alexa, and W. Müller. Visualizing time-series on spirals. In *Infovis*, vol. 1, pp. 7–14, 2001.

[22] W. Willett, J. Heer, and M. Agrawala. Scented widgets: Improving navigation cues with embedded visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1129–1136, 2007. doi: 10.1109/TVCG.2007.70589

[23] J. Wu and S. Wei. *Time series analysis*. Hunan Science and Technology Press, ChangSha, 1989.

[24] P. T. Zellweger, J. D. Mackinlay, L. Good, M. Stefik, and P. Baudisch. City lights: contextual views in minimal space. In *CHI'03 extended abstracts on Human factors in computing systems*, pp. 838–839. ACM, 2003.

[25] J. Zhao, F. Chevalier, E. Pietriga, and R. Balakrishnan. Exploratory analysis of time-series with chronolenses. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2422–2431, 2011.