This is a preprint of the following chapter: C. Tominski et al., "Understanding Temporal Aspects of Digital Trace Data with Visual Analytics", published in "Digital Trace Data Research in Information Systems: Foundations, Methods, and Applications", edited by B. Wurm and J. Mendling, 2026, Springer reproduced with permission of Springer. The final authenticated version is available online at: https://doi.org/10.1007/978-3-032-05497-5_6

Understanding Temporal Aspects of Digital Trace Data with Visual Analytics

Christian Tominski[0000−0001−7704−355X]
Wolfgang Aigner[0000−0001−5762−1869]
Silvia Miksch[0000−0003−4427−5703]
Heidrun Schumann

Abstract Digital trace data are almost inevitably tied to the dimension of time simply because tracing only makes sense in relation to time. The dimension of time is therefore an essential aspect that deserves special attention in any attempt to understand digital trace data. This chapter focuses on understanding the temporal aspects of digital trace data using visual analytics. The premise of visual analytics is to combine visualization, interactivity, and data analysis methods into coherent tools that help people make sense of data. As a basis for developing of such tools, this chapter first specifies the characteristics of time and time-oriented data. Then, specific visualization, interaction, and data analysis techniques are introduced and illustrated with practical examples. The chapter concludes with an outline of challenges for future research in visual analytics in the context of time-oriented digital trace data.

1 Introduction

Analyzing the traces that we leave in the digital world can tell us a lot about human individuals, human behavior, and human society in general. The actual trace data can be diverse. For example, the browsing history that is collected when we explore the Internet contains a lot of information about our interests. Posts on social media platforms are most likely related to our daily concerns. Transaction data collected

Christian Tominski
University of Rostock, Germany, e-mail: christian.tominski@uni-rostock.de

Wolfgang Aigner
St. Pölten University of Applied Sciences, Austria, e-mail: wolfgang.aigner@fhstp.ac.at

Silvia Miksch
TU Wien, Austria, e-mail: silvia.miksch@tuwien.ac.at

Heidrun Schumann
University of Rostock, Germany, e-mail: heidrun.schumann@uni-rostock.de

by our favorite payment service contain valuable information about our daily needs. GPS tracks that we may collect with our phones add a spatial component that tells us where we have been at different times of the day and, on a larger scale, at different stages of our lives.

As we study the various digital trace data, we quickly realize that they have one aspect in common: time. Digital traces are generally associated with some notion of time. This makes the dimension of time an essential data facet to consider when making sense of digital trace data.

Making sense of any reasonably large and complex data can be a challenging endeavor. In this chapter, we take a *visual analytics* (VA) approach to making sense of data. The core idea of VA is to combine the strengths of the human cognitive system with the enormous power of modern computing machines to solve data analysis tasks (see Thomas and Cook, 2005; Keim et al., 2008). The machines do what they can do best: fast, accurate, and repeatable computations on large amounts of data. Humans contribute their ability to visually identify patterns and trends, their problem-solving skills, and their creativity. Consequently, VA implements a *human-in-the-loop* strategy to facilitate data analysis activities. To this end, VA solutions seamlessly integrate visualization, interactivity, and data analysis to help humans make sense of data. Thomas and Cook (2005) define VA as follows:

> Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces. People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible, and understandable assessments; and communicate assessment effectively for action.
>
> Thomas and Cook (2005, p. 4)

Defined in this way, VA is an inherently interdisciplinary field that combines diverse research areas such as human-computer interaction (HCI), usability engineering, cognitive and perceptual science, information visualization, scientific visualization, databases, data mining, statistics, knowledge discovery, data management, and knowledge representation (see Keim et al., 2010).

In this chapter, we introduce the three key components of VA, visualization, interactivity, and data analysis, and shed some light on how they can support the understanding of digital trace data, where our particular interest will be in the temporal aspects of the data. In Section 2, we first characterize the dimension of time and time-oriented data, which is essentially what digital trace data are. In Section 3, we will discuss how time-oriented data can be mapped to expressive and effective visual representations. Section 4 explores interactivity for human-in-the-loop data analysis in more detail. How analytical computation can support the data analysis, especially for larger datasets, is the topic of Section 5. We conclude with a discussion of research challenges for VA of digital trace data in Section 6.

As we can only convey a limited amount of information in this chapter, the interested reader is also referred to the authors' recent open-access book[1] for more details on how visualization, interactivity, and data analysis can support the understanding of time and time-oriented data (see Aigner et al., 2023).

[1] https://www.timeviz.net

2 Time and Time-Oriented Data

Time is a fundamental phenomenon that is deeply embedded in human existence (see Lenz, 2005). It has been a central theme across various scientific disciplines for centuries. Observable through natural cycles, like day-night succession and seasonal changes, it influences virtually all aspects of life on earth. Early human artifacts, such as bone engravings used as calendars, demonstrate our long-standing fascination with time (see Wills, 2012; Aigner et al., 2023). The breadth of time-related research is vast but despite the diversity of theories, common themes emerge: time is unidirectional (arrow of time) and provides order to events.

Particularly in the context of digital trace data, the notion of time is of central importance. The way time itself and the information related to time are modeled in the recorded traces determines the conclusions we can draw from them. Reasoning about the order and sequence of data elements, temporal patterns and recurrences, rates of change, or synchronization events are of interest when making sense of the records of activity left by our interactions with digital information systems.

In order to support such sensemaking processes with appropriate VA solutions, we must first understand how the physical dimension of time and associated data can be modeled in a way that facilitates interactive visualization and analysis using computer systems.

2.1 Characteristics of the Dimension of Time

The conceptualization and representation of temporal aspects in information systems do not aim to achieve an exact replication of time as a physical dimension. Instead, the primary objective is to develop temporal models that effectively capture the relevant phenomena and facilitate the specific analytical tasks. That said, it is crucial to recognize the inherent diversity in temporal modeling approaches. There exists no single, universally applicable model or taxonomy of time that can be considered definitively correct. The domain of temporal representation in information systems is characterized by a variety of approaches (see Hajnicz, 1996). The selection and implementation of a particular temporal model are contingent upon the specific application domain and the unique requirements of the problem space under consideration.

In the context of this chapter, we therefore present the overall design aspects of modeling time, and not a particular model. Figure 1 provides an overview of the fundamental design aspects for time and related abstractions.

2.1.1 Fundamental Design Aspects

In order to identify the key design elements related to time, we draw upon the research of Frank (1998) and Goralwalla et al. (1998), who proposed fundamental

orthogonal dimensions for categorizing various types of time. The following paragraphs will provide an exploration of four crucial aspects: scale, scope, arrangement, and viewpoint.

Scale – ordinal vs. discrete vs. continuous. First, let us examine the scale along which elements of time are given. In an *ordinal* time domain, only the relative order (before, after) between elements is captured, but no quantifiable distances. An example statement is: "Valentina posted a comment to the tweet *before* it was commented by Vincent." In a *discrete* time domain, we can determine quantifiable temporal distances. Common in information systems, it uses a smallest unit (or *chronon*), such as, seconds or milliseconds. An example for discrete time would be: "Vincent was born in 2006 and Emilia was born in 2012." With that, we do not only know that Emilia was born *after* Vincent, but also that Vincent is *6 years older* than Emilia. In contrast to discrete time domains with a smallest possible unit, *continuous*

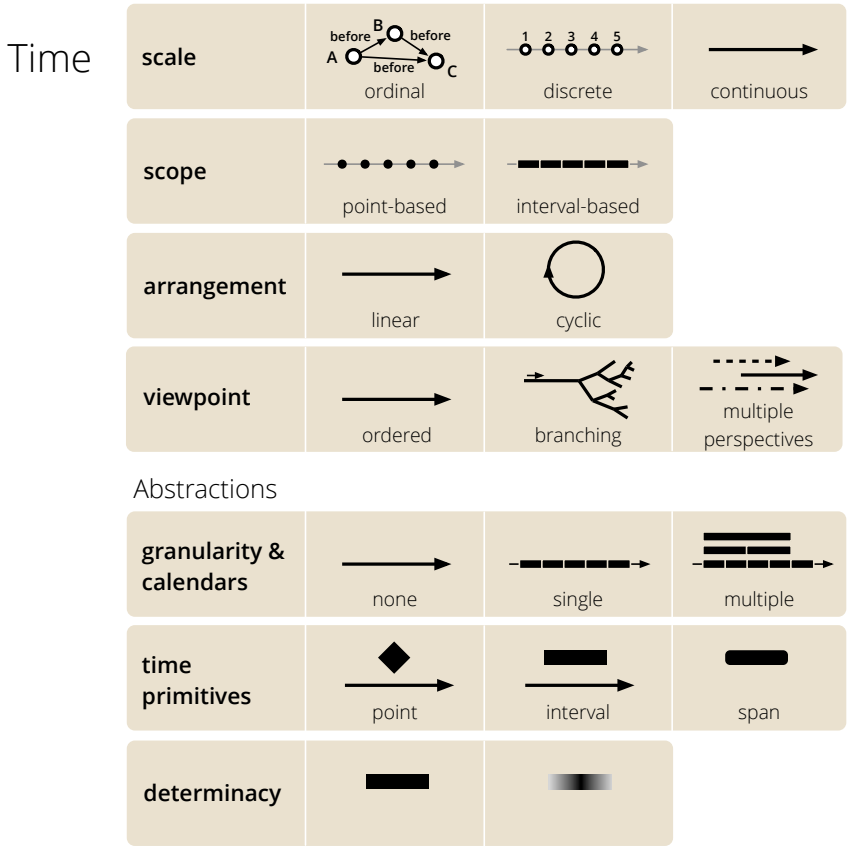


Fig. 1: Overview of design aspects for modeling time.

time domains allow for arbitrary precision by mapping time to real numbers, with infinite divisibility between any two points in time (also referred to as *dense time*).

Scope – point-based vs. interval-based. The scope of time domains can be categorized as point-based or interval-based. *Point-based* time domains are analogous to discrete points in space with zero temporal extent. In contrast, *interval-based* time domains represent subsections of time with a duration greater than zero. This concept is related to granularity. For example, "October 23, 2012" in a point-based domain with granularity of days refers to exactly that date, while in an interval-based hourly domain, it represents the entire 24-hour interval.

Arrangement – linear vs. cyclic. The third design aspect is the arrangement of the time domain. Time is typically perceived as linear, progressing from past to future with unique predecessors and successors. However, periodicity is common in data, such as seasonal variations or monthly averages. For cyclic time, the domain consists of recurring time values, where any value both precedes and succeeds another (e.g., winter comes before as well as after summer). Strictly cyclic data that ignore linear progression are rare (e.g., data records for the day of the week not considering month or year). More common are serial periodic data, combining periodic and linear progression, such as daily website traffic averages over weeks.

Viewpoint – ordered vs. branching vs. multiple perspectives. The fourth subdivision concerns views of the modeled time. In most cases, *ordered* time domains are used, which can capture sequences of events based on a total or partial ordering. *Branching* time makes it possible to split up and unfold several strands of time. This is useful for comparing alternative scenarios in decision-making processes, applicable to both future and past events. In addition, *multiple perspectives* enable us to model simultaneous, even contradictory views of time, which can be necessary for example to unpack contradictory traces recorded about the same real-world event. Temporal databases often use multiple perspectives, distinguishing between valid time (when a fact is true in reality) and transaction time (when it is stored in the information system).

2.1.2 Abstractions: Granularities, Time Primitives, and Determinacy

The previous section outlined design aspects for modeling the time domain's scale, scope, arrangement, and viewpoints. We will now focus on three additional aspects: the hierarchical organization of time, the definition of specific time elements used to link data to time, and temporal determinacy.

Granularity and calendars – none vs. single vs. multiple. To manage time's complexity, different levels of granularity are used. Granularities are (often human-made) time *abstractions* such as minutes, hours, or days that map time values to larger or smaller conceptual units. A system of multiple interrelated granularities mapping fine-grained abstractions to coarser abstractions is called a calendar, often including cyclic elements for concise expression of time values. Mappings between

granularities can be regular (e.g., 60 seconds to one minute) or irregular (e.g., days to months, which vary depending on the specific month and year). Calendars, like the Gregorian calendar, typically use cyclic elements (day, month, year) that repeat as time progresses. Overall, a time model can be characterized as using no, a single, or multiple granularities for defining the elements of time.

Time primitives – point vs. interval vs. span. Time primitives are the basic elements used to relate data to time. They serve as an intermediate layer between data elements and the time domain. A time *point* is a single moment in time (e.g., May 23, 1977). Note, however, that time points may have a duration depending on the time model used and, they can be defined at various granularity levels (e.g., day or minute). An *interval* is a portion of the time domain, typically specified by two time points, a beginning and an end. Intervals can also be specified by a beginning plus a duration or by a duration plus an end. A time *span* is an unanchored, directed primitive, representing an amount of time in terms of a number of elements in a given granularity (e.g., 10 days, 150 minutes). When the granularity is irregular (e.g., months), the exact length of a span can only be determined when it is anchored to the time domain.

Determinacy – determinate vs. indeterminate. Uncertainty is a crucial aspect of time-oriented data (see Jensen et al., 1998), particularly when dealing with incomplete information or granularity conversions. Time specifications can be either *determinate*, in which case we have complete knowledge of all temporal aspects, or *indeterminate*, meaning we have don't-know-(exactly)-when information. Indeterminate information includes inexact knowledge, future planning data, or imprecise time specification. Indeterminacy can be explicitly specified or implicitly present in multiple granularities. For example, a minute-based time specification of a social media post becomes indeterminate when viewed at a granularity of seconds, as the exact time within the minute is unknown.

2.2 Characteristics of Time-Oriented Data

After discussing the aspect of modeling time itself, we now move on to untangle the question of characterizing time-oriented data. Time-oriented data refers to data values associated with time primitives. The pyramid framework by Mennis et al. (2000) offers a useful cognitive model for such data. It is based on three perspectives: location (where), time (when), and theme (what). These data aspects form objects at a higher cognitive level, along with their taxonomy (classification; super-/subordinate relationships) and partonomy (interrelationships; part-whole relationships).

The choice of perspective depends on the phenomenon and analysis purpose. Following the aspects of measurement described by Hand (2016), for example, one might consider time-related conceptual entities (objects) for representational measurement, or practical phenomena like user engagement indirectly by tracking interaction intensity over time for pragmatic measurement. While no single model

suits all applications, fundamental design alternatives can be identified to characterize time-oriented data. Figure 2 provides an overview of the design aspects we are considering here.

Scale – quantitative vs. qualitative. The type of scale along which data variables are measured can be categorized into quantitative and qualitative. Quantitative variables use a metric range (discrete or continuous) that enables numeric comparisons and distance calculations. Qualitative variables, on the other hand, comprise either an unordered (nominal) or ordered (ordinal) set of data values, but no notion of distance between values. Understanding the characteristics of the data scale is crucial for designing appropriate visual representations, as we will see in Section 3.

Frame of reference – abstract vs. spatial. It is useful to differentiate between abstract and spatial data. *Abstract* data lacks the *where* aspect of the pyramid framework, meaning it is not inherently linked to a spatial location. *Spatial* data, conversely, contains an intrinsic spatial layout, incorporating the *where* aspect in its data model. This distinction significantly impacts the visualization approach for time-oriented data. For spatial data, the mapping of data to the screen is largely predetermined by the spatial frame of reference, which makes it challenging to effectively incorporate the temporal (*when*) aspect. For abstract data, without a predetermined spatial mapping, the primary task is to create an expressive spatial layout that prominently features the time domain.

Kind of data – events vs. states. This criterion addresses whether the data represents events or states. *Events* are markers of change, such as a passenger's boarding of a plane, while *states* are phases of continuity between events, like a plane being in flight. In the context of digital trace data, transactions and interactions primarily follow the notion of events, while sensor data or regular snapshot of an information

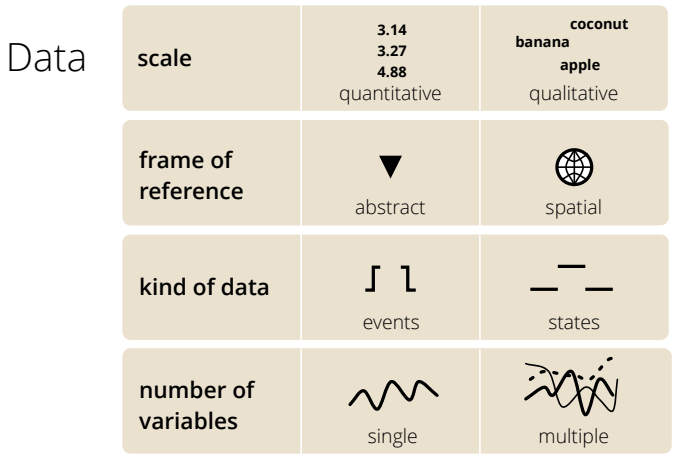


Fig. 2: Overview of design aspects for modeling data.

system's status fit the notion of states. Although events and states are intrinsically linked, it's crucial to clearly communicate which aspect–events, states, or a combination of both–are being visualized.

Number of variables – single vs. multiple. This criterion focuses on the number of time-dependent variables present in the data. There is a significant distinction between univariate data, where each time primitive is associated with a *single* data value, and multivariate data, where *multiple* data values exist per time point. This difference is crucial in data visualization, as the range of applicable methods for univariate data is considerably broader than for multivariate data. The complexity of representing multiple variables simultaneously often limits the available visualization techniques for multivariate time-oriented data.

2.3 Data in Relation to Time

Last but not least, let us look at how a trace dataset can be related to time. A dataset's time dependency can be described along the two dimensions of *internal time* T_i and *external time* T_e as defined by Steiner (1998). *Internal time* describes how time is represented inside the dataset. *External time* is extrinsic to the data model and describes how a dataset evolves over time. The following classification depends on the number of time primitives in both internal and external time as shown in Figure 3.

Static non-temporal data. When both internal and external time domains consist of only a single time primitive each, the data are considered time-independent or

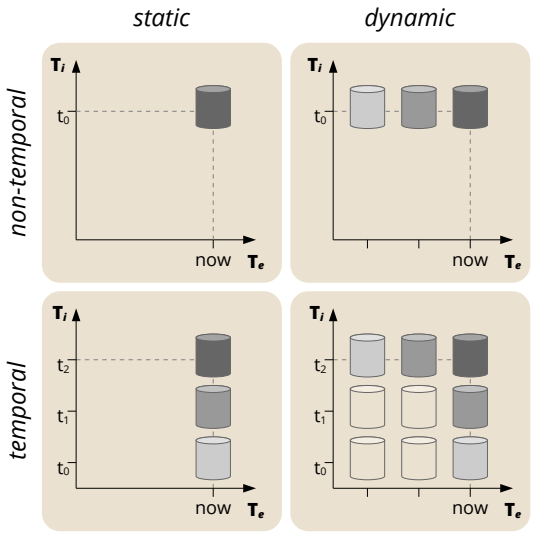


Fig. 3 Data in relation to internal and external time, T_i and T_e. Adapted from Steiner (1998). CC BY

static. This type of data, being unaffected by temporal changes, falls outside the scope of this chapter's focus on temporal aspects of digital trace data.

Static temporal data. When internal time contains multiple time primitives while external time has only one, the data are considered time-dependent. Static temporal data are the typical case of digital trace data representing a historical view (or trace) of how an observed environment and our interactions with it have evolved over time.

Dynamic non-temporal data. When internal time contains only one primitive, but external time comprises multiple primitives, the data are dependent on external time. This scenario represents dynamic data, where the dataset itself changes over time without keeping a history, that is, outdated values are overwritten.

Dynamic temporal data. When both, internal and external time domains consist of multiple time primitives, the data are considered bi-temporally dependent. This means the data contains variables that depend on internal time, while the dataset itself evolves over external time. Typically, internal and external time are closely coupled and can be mapped to each other. An example of bi-temporally dependent data are tweets and likes about stock market trading (internal time) that are updated daily with new records (external time).

3 Visualization

Undeniably it would be quite difficult for humans to understand the data outlined in the previous section by simply looking at a binary or alphanumeric data file. Therefore, the premise of VA is to represent the data visually. The human perceptual system is highly sophisticated and specifically suited to detect visual patterns (see Ware, 2020). Visualization seeks to exploit this ability and help people interpret and understand complex data. For a visualization to be successful, three main criteria must be met (see Ward et al., 2015; Tominski and Schumann, 2020):

- expressiveness,
- effectiveness, and
- efficiency.

The *expressiveness* criterion demands that exactly the relevant information contained in the data be displayed; no more and no less. To obtain *effective* visual representations that actually help in understanding the data, visualizations must take human cognition into account. Finally, the *efficiency* criterion requires that the benefits of a visualization outweigh its costs. Expressiveness, effectiveness, and efficiency are criteria that any visualization should strive to meet.

3.1 Visual Mapping

Conceptually, a visualization transforms raw data into image data. This process is commonly called the *visualization pipeline* (see Hauser and Schumann, 2009; Moreland, 2013). It typically consists of three steps, as shown in Figure 4:

1. filtering,
2. mapping, and
3. rendering.

The filtering step prepares the raw input data for processing through the remaining steps of the pipeline. This includes data cleansing, the selection of relevant data, data enrichment, reduction, interpolation, and grouping. At the core of the visualization pipeline is the mapping step. It literally maps the prepared data to a suitable visual representation. This is the most crucial step as it largely influences the expressiveness and effectiveness of the resulting visual representation. Finally, the rendering step generates actual image data using computer graphics techniques.



Fig. 4: The visualization pipeline.

Because the mapping step is so crucial, let us look at it in more detail for the particular case of time-oriented data. The input of the mapping step are the data with all their different characteristics as described in Section 2. The essence of the mapping step is to map time and data variables to *visual variables*. Just as it needs a data table to hold the data, it needs *graphical marks* to represent the visual variables. For example, for a simple point plot, time (a data variable) is mapped to position (a visual variable) of dots (the graphical marks) along the horizontal display axis, and a time-dependent variable is mapped to the dots' vertical position. In general, the mapping step can employ different graphical marks, including 0D points, 1D lines, 2D areas, and 3D volumes, which can hold a variety of visual variables, including position, length, angle and slope, width, lightness, and size (see Munzner, 2014).

Figure 5 shows some examples of how visual variables can be used to represent time. We have already mentioned *position* as a prominent visual variable. *Length* is particularly useful for encoding the duration of time intervals. *Slope* and *angle* can be used for cyclic representations of time. The *size* of point marks can also encode time. Varying the *width* or *lightness* of line marks can be useful for visualizing time along traces, and simple *connections* works as well. *Containment*, *symbol*, and *texture* can be applied when there are only a few time primitives provided that appropriate text labels are included. While *position*, *length*, *slope*, and *angle* allow for quantitative assessment, the remaining visual variables are mostly suited for qualitative inferences about time.

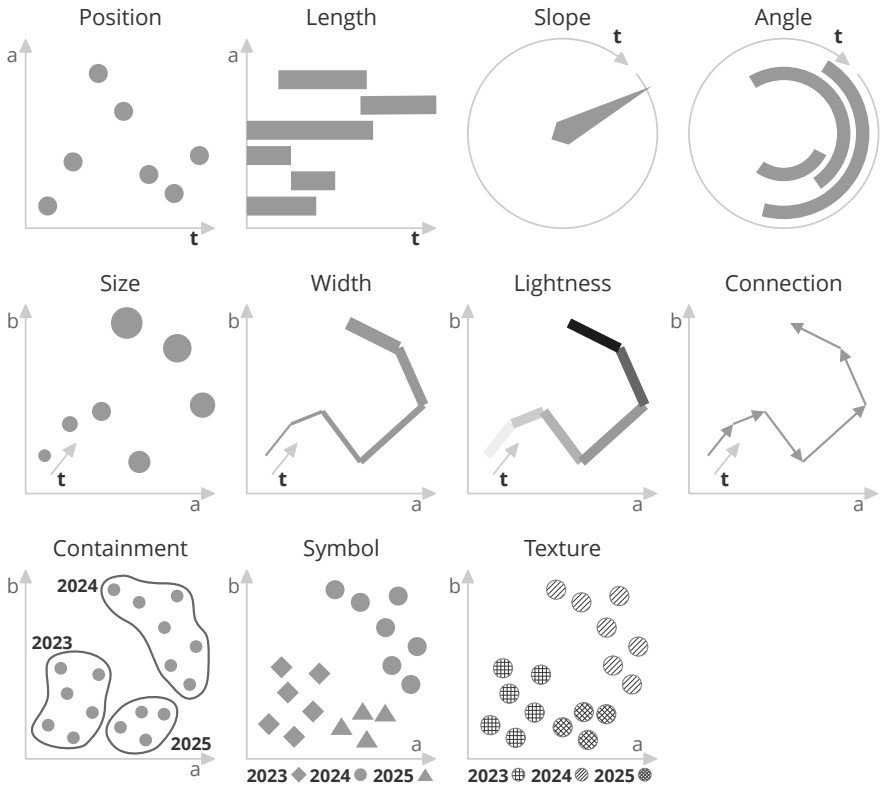


Fig. 5: Examples of visual variables for mapping time (**t**) and time-dependent variables (a and b).

There is another way to map time, but we cannot show it in a static figure. It is the mapping of time to time, which results in a dynamic visual representation that shows the data's evolution over time by dynamically changing the visualization over time. Such dynamic (or animated) visualizations, if properly designed, can communicate overall temporal dynamics and general trends quite well. Yet, comparing the data for multiple time points can be cumbersome and getting an overview might be difficult. Therefore, the decision whether to use static (mapping time to marks and visual variables) or dynamic (mapping time to time) visualizations must be made based on the specific analysis tasks at hand.

When designing a visualization, the choice of an appropriate visual mapping of time and time-dependent variables is crucial. Perceptual studies by Cleveland and McGill (1984), which were later replicated by Heer and Bostock (2010), suggest that position is a quite effective and universally applicable visual variable. Other visual variables vary in their effectiveness depending on the data they encode (see data scale in Section 2.2). Length and angle are better suited for quantitative data.

Color lightness and saturation allow us to interpret ordinal data well. For nominal data, symbol and color hue are the most suited visual variables. Yet, it should also be noted that our understanding of human perception is far from complete and that the effectiveness of a visualization is influenced by many different factors beyond visual variables (see McColeman et al., 2022).

Nevertheless, the perceptual ranking of visual variables can inform the visual mapping. Given the importance of time for digital trace data, it seems useful to map time to the most effective visual variables position, length, or angle. Time-dependent data variables can then be mapped to the next effective visual variables (e.g., size, color lightness, or color hue) depending on the data characteristics. If the trace data include spatial information, the visual variable position usually must reflect the spatial frame of reference and cannot be used otherwise. In such cases, alternative solutions need to be applied, for example, mapping time to time or using a 3D visualization called *space-time cube* where the first two dimensions encode space, while the third display dimension offers an additional axis for mapping time to the 3D z-coordinate (see Kraak, 2003; Tominski et al., 2005; Bach et al., 2017).

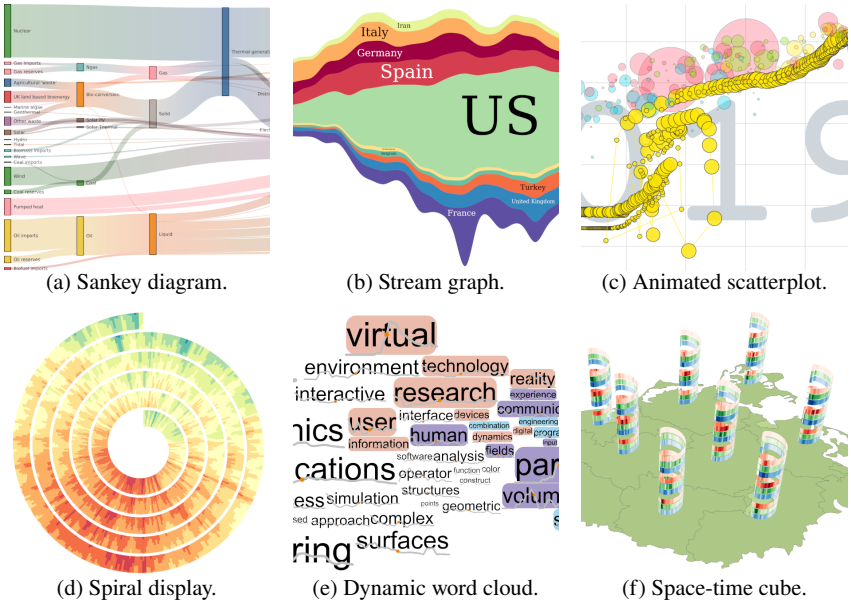
3.2 Visualization Techniques

The different options for mapping time and data to marks and visual variables span a huge design space. So, it is not surprising that there is also a wealth of visualization techniques for time and time-oriented data for different data characteristics, analysis goals, and application domains. Figure 6 illustrates a few examples that may be useful in the context of trace data, but which we cannot describe in detail here. A comprehensive survey of these and more than 150 other visualization techniques is provided in Aigner et al. (2023)'s book and the TimeViz Browser (see Tominski and Aigner, 2023).

Finding a suitable visualization technique for a particular real-world problem in the large zoo of possible solutions can be challenging. For a successful search, one would need to address three key questions:

- What is visualized? (data level)
- Why is it visualized? (task level)
- How is it visualized? (presentation level)

These questions relate to the different characteristics of time and time-oriented data (the what), the different analytical tasks one seeks to solve (the why), and the different options for the visual representation (the how). However, non-experts may find it difficult to deal with the theoretical intricacies of these questions. Therefore, it makes sense to simplify and focus on the what and the how, which are easier to decide. Because analytical tasks are rarely predefined and usually vary throughout the data analysis, the why aspect is not further considered here, but will surface again when we talk about interactivity in Section 4.



(a) Sankey diagram. (b) Stream graph. (c) Animated scatterplot.

(d) Spiral display. (e) Dynamic word cloud. (f) Space-time cube.

Fig. 6: Examples of visualization techniques. Aigner et al. (2023)

For further simplification, the what and the how are broken down into three practical key criteria: (1) time and (2) data, the what, and (3) visual representation, the how. Each key criterion has two sub-criteria with two corresponding categories each, which gives us a simple and accessible overall structure:

- **Time**
  - *Primitives* – points vs. intervals
  - *Arrangement* – linear vs. cyclic
- **Data**
  - *Number of variables* – single vs. multiple
  - *Frame of reference* – abstract vs. spatial
- **Visual representation**
  - *Mapping of time* – static vs. dynamic
  - *Dimensionality* – 2D vs. 3D

The time and data criteria pick up four selected aspects (i.e., type of time primitives, arrangement of time, number of variables, and frame of reference) from Section 2, while the visual representation criterion refers to two key aspects of the visual mapping (i.e., mapping of time and dimensionality).

Using this simple scheme, the TimeViz Browser can offer a structured filtering mechanism that allows us to partition the space of solutions and thus to guide

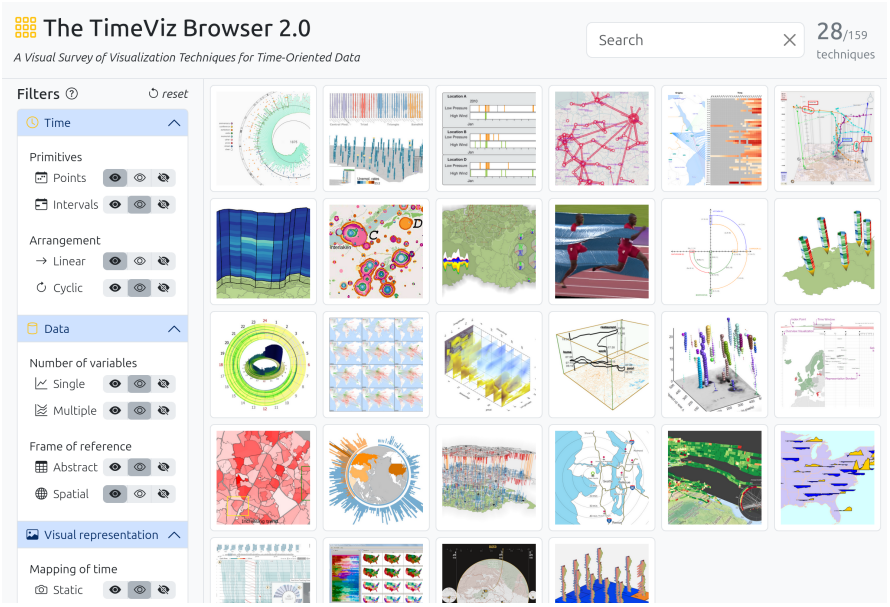


Fig. 7: The TimeViz Browser offers filters (left) to guide search of suitable visualization techniques (center) in a corpus of more than 150 possible solutions.

the search for suitable visualization techniques. Figure 7 shows the corresponding interface of the TimeViz Browser. A data analyst can now simply take the data specification of the problem at hand and set the filter criteria accordingly, and the TimeViz Browser will only show techniques that match the analysts needs.

3.3 Visualization Examples

While the TimeViz Browser covers visualization techniques for time and time-oriented data in general, the next two sections will describe in more detail two examples that may be of particular interest in the context of digital trace data.

3.3.1 Visualizing Financial Traces in Time

One form of digital trace data is financial flows (see Howison et al., 2011). A financial flow documents the transfer of an amount of money by a user's interaction with an information system. As an example, federal transparency laws in Austria mandate the disclosure of financial flows from public institutions to media entities. Although the resulting media transparency dataset, comprising quarterly transfers, is publicly

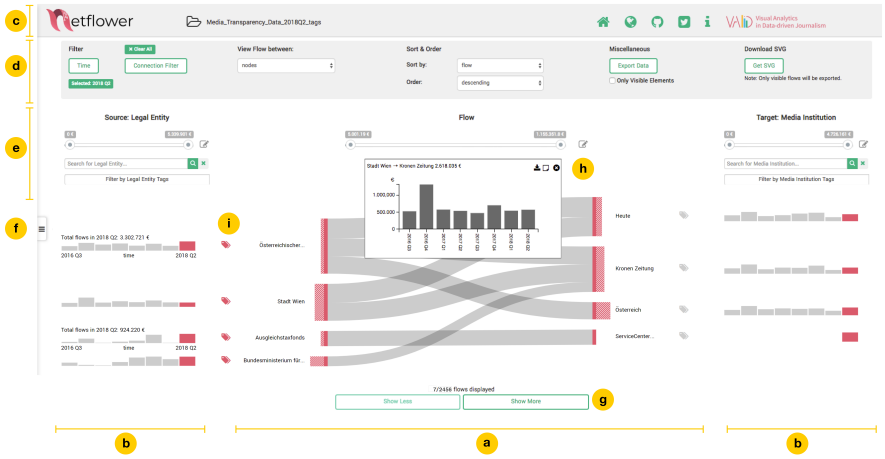


Fig. 8: Netflower visualization of quarterly flows of money between government institutions (left) and media outlets (right) by Stoiber et al. (2019). CC BY

available, it is too extensive for line-by-line examination. Simple analysis of the largest flows is insufficient, as many questions of interest involve temporal changes and complex many-to-many relationships between entities and media.

In general, financial flow traces can be conceptualized as time-dependent attributes on the edges of a bipartite network spanned by accounts and the transactions between them. The dataset is characterized as dynamic temporal data, containing internal time information while evolving externally. Its temporal aspect consists of time points on a discrete, interval-based, linear time domain with quarter and year granularities. The data associated with time elements is both quantitative and qualitative, abstract (lacking spatial information), records states (quarterly sums rather than individual transactions), and involves multiple variables.

Traditional data analysis tools like spreadsheets are inadequate for this complex data structure. Netflower, introduced by Stoiber et al. (2019), is a visualization tool designed for exploring such financial flow traces. Figure 8 shows an screenshot of the tool. The central visualization (a) is a Sankey diagram, with originating nodes on the left and destination nodes on the right of a bipartite network. Line widths are proportional to flow quantities summed over the selected time interval. Netflower emphasizes scalability for large graphs and supports exploration of temporal dynamics through details-on-demand, filters and tags, auxiliary views, and a provenance notebook (d-f, i). A time interval filter (d) allows selection of the period shown in the Sankey view. Upon flow selection, a bar graph (h) displays flow values for each time point. Bar graph sparklines (b) on both sides of the nodes provide a temporal overview of outgoing (source nodes) and incoming flows (target nodes). Time itself is represented as linear, interval-based time-axis using the granularity month, both in the sparklines to the left and right of the Sankey diagram (i), as well as in the details-on-demand panel for selected flows (h).

Using a visual exploration tool as this effectively addresses the challenges of investigating complex, time-dependent data, enabling sensemaking and deeper insights.

3.3.2 Visualizing Movement Traces in Space-Time

In the introduction, we have already mentioned GPS traces as a specific example of digital trace data. Such data are also known as movement data. Here we consider movement that is constrained to certain natural or anthropogenic features of the environment, such as rivers or roads. In addition to the temporal component (discrete, point-based, linear time), movement data contain spatial information. We also already mentioned the challenge of visualizing data with temporal *and* spatial aspects, which is mainly due to the need to visualize the spatial frame of reference along with the actual spatiotemporal data (see Andrienko et al., 2013).

A typical visualization for movement data is to display them on a map as 2D traces, where time is mapped along the traces and color encodes a quantitative data variable (e.g., speed), as shown on the left in Figure 9. From such a visual representation, we can infer where the movement took place and how fast things were moving (red for slow, green for fast). However, since the movement is constrained to roads, there is a lot of overplotting, and we cannot really be sure how many traces are shown, which places where visited before others (i.e., movement direction), and whether there are any hidden patterns.



Fig. 9: Visualization of GPS traces. Left: Plain color-coded traces on a 2D map. Right: Stacking-based 3D visualization by Tominski et al. (2012). CC BY

An alternative representation addressing the aforementioned challenges is shown on the right of Figure 9. The visualization now shows the map in the base plane of a 3D space. On top of the map, individual traces are stacked as colored 3D bands along

the third dimension. Again, speed is encoded by color, and additionally small arrows indicate temporal order, which allows us to see the movement direction. Individual trajectories are now much easier to trace visually. We can also see potentially interesting patterns where red and green colors alternate frequently, which may indicate stop-and-go traffic or hint at sensor problems.

While the overplotting could be reduced in the 3D visualization, we now have to deal with 3D occlusion. We may also be dissatisfied with the temporal information being visible only qualitatively as temporal order indicated by the small arrows. These issues can be addressed through interactivity. For example, 3D navigation can be used to resolve 3D occlusion and view the data from different perspectives. More details about interactivity in general, and in particular how quantitative temporal properties of movement data can be made accessible via interaction, will be discussed in the next section.

4 Interactivity

The previous examples already indicated that interactivity plays a crucial role when analyzing temporal aspects of digital trace data. In general, VA follows a human-in-the-loop approach, meaning that human analysts incorporate their expertise into the analysis process to dynamically steer it according to varying analysis tasks and desired outcomes. To develop a comprehensive understanding, analysts create and explore multiple visualizations, each showing a specific part or perspective of the data. Even before the widespread use of computers, cartographer and visualization pioneer Bertin (1981) formulated the following:

> A graphic is not "drawn" once and for all; it is "constructed" and reconstructed until it reveals all the relationships constituted by the interplay of the data. The best graphic operations are those carried out by the decision-maker himself.
>
> Bertin (1981, p. 16)

To give structure to the process outlined by Bertin, one can resort to the *information seeking mantra* proposed by Shneiderman (1996):

> "Overview first,
> zoom and filter,
> then details-on-demand."
>
> Shneiderman (1996, p. 2)

According to this mantra, the data analysis process typically begins with an overview of the entire dataset. To gain deeper insights, analysts utilize interaction techniques to drill down into specific areas of the data. By moving back and forth between the overview and the details of different parts of the data, analysts can gradually acquire more and more knowledge about the data.

Going from overviews to details and back is but one reason for interactivity. In fact, analysts may require or desire interaction for a number of reasons. Yi et al. (2007) provide a detailed description of the corresponding interaction intents:

Select – Mark something as interesting When we see something interesting in a visualization, we often want to mark and highlight it, be it low-level trace items or higher-level behavioral abstractions.

Explore – Show me something else As trace data are often only visualizable in parts, it is necessary to navigate the data in order to explore different portions of time and space.

Reconfigure – Show me a different arrangement The different characteristics of digital trace data (e.g., linear vs. cyclic time or abstract vs. spatial frame of reference) make it necessary to create different arrangements or layouts.

Encode – Show me a different representation Similarly, the visual encoding of data values can be subject to interactive adjustments (e.g., use a different color mapping or visualization technique) to adapt to changing analysis tasks.

Abstract/Elaborate – Show me more or less detail This interaction intent reflects the previously mentioned visual information seeking mantra: Switching between overview and detailed views of traces.

Filter – Show me something conditionally Filtering enables analysts to narrow the visualization to display only the trace data that align with their specific search conditions, thereby eliminating irrelevant information.

Connect – Show me related items When a potentially interesting finding is made, it is usually helpful to interactively collect, compare, and evaluate related information in the trace data.

A study by Van Berkel et al. (2024) confirms that the availability of suitable interaction techniques for these intents significantly impacts data analysis activities. On top of the aforementioned visual-analysis-specific intents, Yi et al. (2007) identify two more general interaction intents of high relevance:

Undo/Redo – Let me go to where I have already been As users interact with a visualization and navigate the data they usually want to be able to return to previous states of the analysis process.

Change configuration – Let me adjust the interface In addition to interacting with the visualization, it is also often necessary to configure the overall interface of the system.

These high-level user intents characterize interactivity from the perspective of *why* the interaction takes place. Next, we look at *how* interactions between humans and VA systems are actually carried out.

4.1 Basic Concepts

The interaction between an analyst and a VA system can be illustrated with the *human action cycle* model by Norman (2013), which is depicted in Figure 10. Based on a goal, the execution phase of the action cycle involves formulating an intent, devising a plan, and executing it to cause a change in the world. In our case, the world

corresponds to the VA system, which reacts with a response that is typically visual in nature. During the evaluation phase, the human will perceive the change, interpret it, and evaluate the outcome with the intended change. If the outcome matches the intended change, the interaction was successful. If not, the cycle begins anew, usually with an adjusted execution phase to achieve the desired outcome.

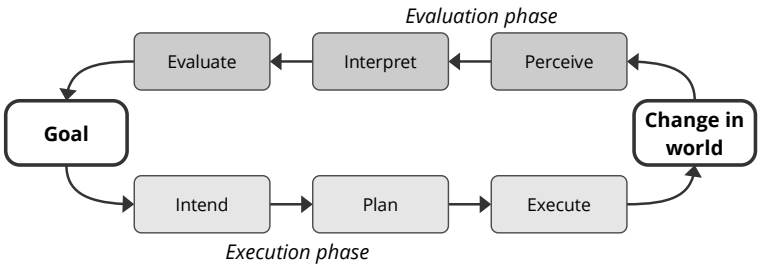


Fig. 10: The human action cycle according to Norman (2013).

The human action cycle in VA systems is traditionally designed for mouse and keyboard interaction and regular desktop computers. Yet, digital trace data can also be explored in other settings, such as mobile devices, large touch displays, or high-resolution display walls. In such settings, alternative interaction modalities become relevant, including multi-touch interaction, pen-based interaction, spoken commands, gaze-based interaction, or full-body gestures. More information about interaction for mobile and large-scale visualizations can be found in the works by Brehmer et al. (2021) and Belkacem et al. (2024), respectively.

Regardless of the environment, for interactivity to be a useful asset in the VA toolbox, the action cycle must be as smooth as possible. Therefore, it is essential to keep the costs of interaction low. Lam (2008) describes in detail the physical (e.g., moving the eyes frequently between different parts of the display) and mental (e.g., interpreting a complex change in the visualization) costs associated with each step of the action cycle. The smoothness of the action cycle also hinges on the VA system's ability to generate a response quickly. Liu and Heer (2014) found that a latency of 0.5s can already adversely affect the analysis process. Therefore, it makes sense to consider *progressive* VA methods (see Fekete et al., 2024), which aim to maintain interactive response rates when larger trace data need to be explored.

What is conceptually captured by the human action cycle model is practically reflected in the user interface of VA systems. The user interface is the channel through which the analyst and the VA system exchange information (i.e., interactive input and visual feedback). The key components of a VA user interface are the visualization views and interaction controls.

Visualization views are responsible for showing the visual data representations. Understanding complex data, such as trace data, typically requires multiple views that present different aspects and different levels of detail of the data. Relevant concepts in this regard are overview+detail, focus+context, and multiple coordinated views.

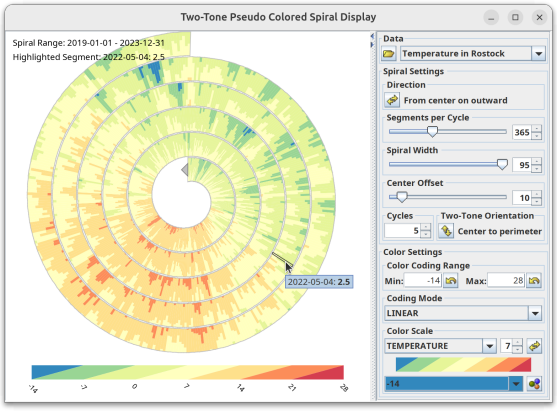


Fig. 11 User interface with a spiral view (left) and a control panel (right). CC BY

Overview+detail methods present overview and details separately, whereas *focus+context* methods smoothly integrate detail and overview in a single visual representation. Zoom interaction is commonly applied to switch between overview and detail. Cockburn et al. (2009) provide a comprehensive review of these different methods and discuss their advantages and disadvantages.

The concept of *multiple coordinated views* allows us to show different aspects of the data, for example, different temporal granularities, different arrangements for linear and cyclic behaviors, or dedicated views for spatial traces. To help analysts develop a consistent overall picture from multiple views, view coordination is indispensable. View coordination means that an interaction being initiated in one view is automatically propagated to all other views, which update themselves accordingly. For example, selecting data items in one view will highlight the selected data in all views. Another practical example is browsing in time. When navigating to a particular point in time in one view, all coordinated views follow the navigation automatically. Baldonado et al. (2000) provide further valuable guidelines for the use multiple coordinated views.

Interacting with visualization views requires appropriate interaction controls. They can be integrated directly into the visualization to facilitate *direct manipulation* (see Shneiderman, 1983; Saket et al., 2020), or they can be assembled into dedicated control panels that accompany the visualization.

Figure 11 illustrates a simple example with a spiral view (see Tominski and Schumann, 2008) that is complemented with a control panel. The spiral view itself affords the selection of time points and the navigation in time through *direct manipulation*. The control panel's sliders, buttons, drop boxes, and custom controls can be used for adjusting the parameters that control the visual mapping.

When interacting with a VA system, there are certainly operations that are carried out more frequently than others. Ideally, frequent operations are designed to have the least interaction costs (e.g., by allowing direct manipulation or by placing controls prominently in the user interface).

4.2 Interaction Techniques

Interaction techniques implement concrete solutions allowing analysts to satisfy their interaction intents. Next, we look at selected examples that are particularly relevant when exploring the temporal aspects of trace data.

Navigating time A primary concern for any time-oriented data analysis is to navigate the dimension of time. Because it is so fundamental, navigating time should be preferably be possible directly in the visualization. An interesting direct navigation method is DimpVis (see Kondo and Collins, 2014). Figure 12 illustrates how DimpVis works on a point plot. The user grabs a dot and a path shows up indicating the selected data item's trajectory through time. In order to navigate in time, the user simply drags the dot along the path until the desired moment in time is reached. As such, DimpVis is similar to a slider control, but the sliding operates on a curved path, rather than a straight line.

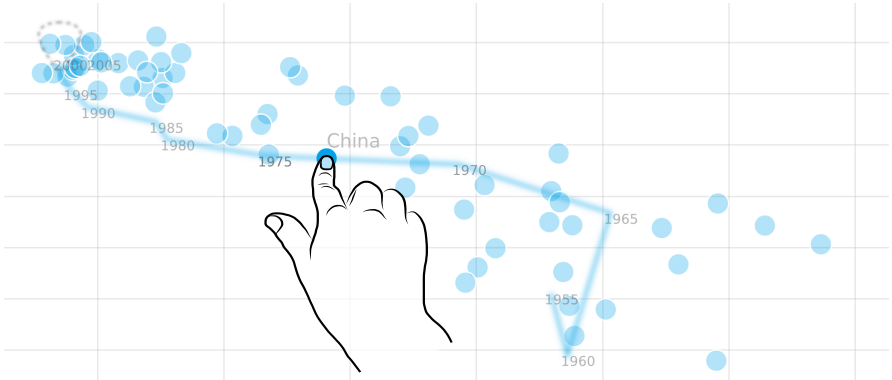


Fig. 12: Navigation through time by dragging a data item along a path. Generated with DimpVis by Kondo and Collins (2014). CC BY Aigner et al. (2023)

Dynamic queries Similar to navigating to specific points in time, it is often necessary to query the visualization with respect to specific data conditions. In other words, the analysts explores not only the $1D$ time dimension, but the entire nD data space. Dynamic queries support such an exploration by restricting the visualization to the data that match the query conditions (see Shneiderman, 1994). All other data are dimmed or even filtered so that the relevant data stand out.

The query conditions are usually specified using dedicated mechanisms. Sliders are effective for filtering time and numerical data, text boxes are useful for querying categorical data (e.g., data tags). Conditions can also be combined via logic operators (e.g., AND, OR) for multidimensional data querying. Ideally, the visual interface allows users to specify filter conditions directly in the visualization, as we will see in the next example.

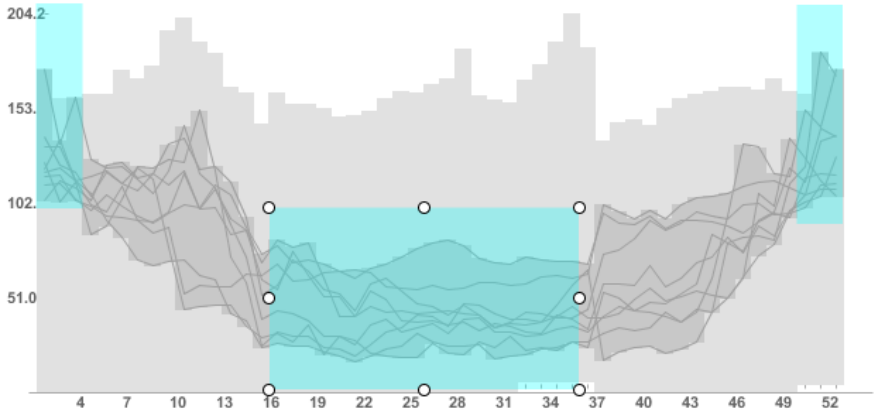


Fig. 13: Three timeboxes (in cyan) are used to dynamically query stock data. The corresponding conditions demand that stock values be high at the beginning of the year, low in the middle, and again high at the end of the year. Generated with TimeSearcher by Hochheiser and Shneiderman (2004). Aigner et al. (2023)

Timeboxes (see Hochheiser and Shneiderman, 2004) filter temporal variables from multivariate line plots. Query conditions are defined by creating one or more time boxes that specify value ranges and time intervals. Figure 13 shows three such time boxes in cyan. The visualization displays only the plots that overlap with the time boxes, essentially performing AND-combined range queries on the data. The rest of the data are highly abstracted as gray bars in the background.

Interactive lenses Navigating in time or performing dynamic queries change a visualization globally. More local visualization adjustments can be achieved with the help of interactive lenses. Much like a magnifying glass, an interactive lens can be moved across a visualization to adjust it in various ways. Tominski et al. (2017) define a conceptual model for lenses and review more than 50 different techniques for visualization purposes. In general, a lens first performs a selection on the visualization to determine what is to be affected by the lens. Then a lens effect is computed on the selection to create a locally adjusted visual representation. Finally, a join operation merges the lens effect with the main visualization, which usually also includes visual feedback about the shape, size, and position of the lens. The advantage of interacting with lenses is that they are lightweight tools that can be easily applied to satisfy local information needs without disturbing the overall visualization too much.

For a concrete example, we look at the time lens (see Tominski et al., 2012), which makes temporal details in digital GPS traces visible. Consider the map visualization with color-coded GPS traces in Figure 14. The lens defines a query region and the visual feedback shows how the GPS data are distributed over time. The interior shows a detailed view of all queried GPS points, and links connect them to a time

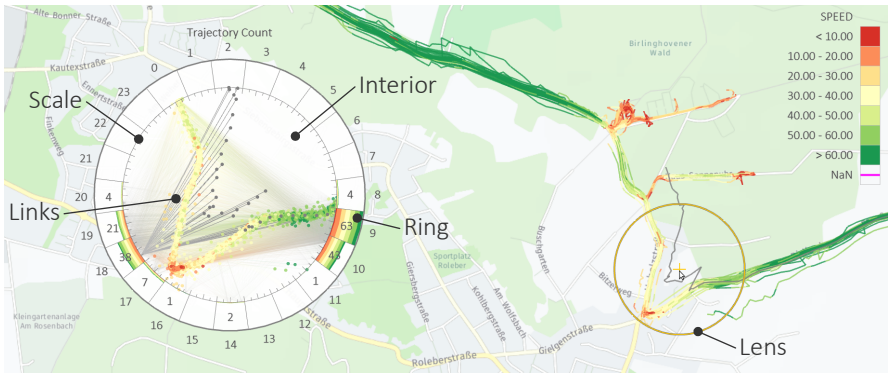


Fig. 14: The time lens provides access to temporal information about GPS traces displayed on a map. Tominski and Schumann (2020)

scale showing their exact time. The ring visualizes the temporal distribution of the GPS points. One can nicely see that most of them were taken between 9 and 10 as well as 18 and 19 o'clock, which indicates commuting behavior.

User guidance The more complex the data, the more complex the analysis. This can lead to situations where analysts are unsure of how to proceed with the data exploration. By providing guidance, analysts can be assisted and steady analytical progress can be made (see Schulz et al., 2013; Ceneda et al., 2017).

To give again an example, a common challenge when exploring temporal data is determining the appropriate level of granularity to focus on when identifying specific temporal phenomena. Zooming often allows analysts to look at different granularities, but where can interesting findings be made? To assist in answering this question, user guidance can be provided through so-called heterogeneity bands (see Luboschik et al., 2012). The hypothesis is that local differences (or heterogeneity) between adjacent abstraction levels may indicate areas where a more detailed visual inspection would be beneficial.

Figure 15 shows a line plot of a time series with 3.6 million time points, which are accessible at 21 pre-computed levels of granularity. Below the line plot are two stacks of heterogeneity bands for two different difference metrics. The color coding shows where larger differences (according to two the metrics) are located. Aligned with the peaks in the line plot, the heterogeneity bands show interesting notches suggesting to the analyst that the tips of the peaks may exhibit interesting behavior yet to be explored. In this example, the guidance makes it easier for the analyst to make an informed decision to continue the data exploration at the tips of the peaks.

This section gave a brief overview of the role of interactivity for VA in the context of data with temporal dependencies. Next, we will describe how analytical computations can support the understanding of temporal data.

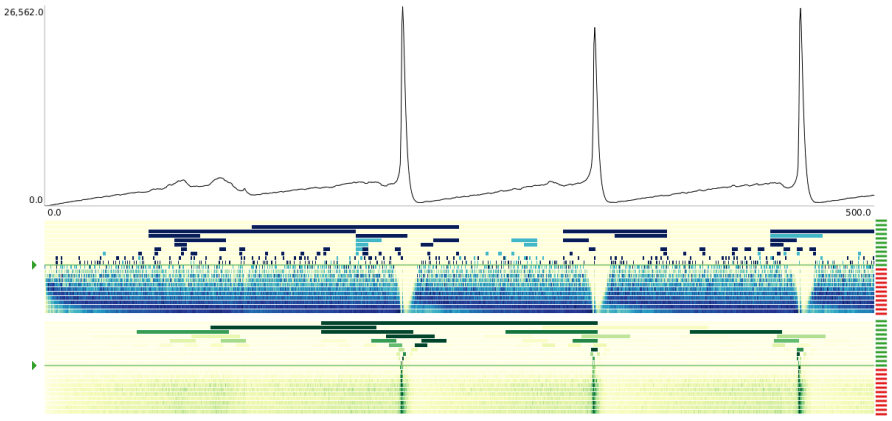


Fig. 15: Heterogeneity bands below the line plot may guide the data exploration towards interesting findings. Generated with software by Luboschik et al. (2012). Tominski and Schumann (2020)

5 Analytical Computations

Purely interactive and visual data analysis works well unless the problem to be solved exceeds a certain size and complexity. With massive, heterogeneous, dynamic, and ambiguous data, it becomes increasingly difficult to create overview visualizations without losing interesting patterns, and human observers may have a hard time interpreting and understanding the data.

In order to emphasize that it is not sufficient to just retrieve and display data using a purely visual and interactive approach, Shneiderman's information seeking mantra has been revised and expanded by Keim et al. (2006) as follows:

> Analyze First -
> Show the Important -
> Zoom, Filter and Analyse Further -
> Details on Demand.
>
> Keim et al. (2006, p. 6)

This VA mantra makes it clear that for supporting comprehensive understanding, it is necessary to computationally analyze the data according to aspects of interest, to show the most relevant features of the data, and to provide interaction methods that allow users to access details on demand.

This section will briefly illustrate how automatic analytical computations can be utilized to facilitate the exploration and analysis of larger and more complex time-oriented data. In general, such calculations are concerned with extracting useful information from the data. More specifically, Laxman and Sastry (2006) characterize the following categories of temporal data analysis approaches:

Classification Given a predefined set of classes, the goal of classification is to determine to which class a data element, sequence, or subsequence belongs. A specific instance of classification is *segmentation and labeling*. It transforms point-based time into interval-based time and assigns class labels to the derived time segments. For example, in the analysis of spatiotemporal movement traces (see Section 3.3.2) segmentation and labeling can classify certain movement behaviors, such as acceleration phases or slower movement in bends.

Clustering Clustering is concerned with grouping data into clusters based on similarity. This requires an appropriate similarity measure, which generally must be chosen depending on the data to be clustered. In the context of time-oriented data, it usually makes sense to cluster similar time series or subsequences of them. But also spatial aspects of the data may be used for clustering. For example, the 3D visualization of movement shown earlier in Figure 9 (right) is based on clustering (and stacking) traces with similar spatial shapes to reduce visual clutter.

Search & retrieval These approaches encompass searching for a-priori specified queries in possibly large volumes of data. Search & retrieval can be applied to locate exact matches for an example query or approximate matches. In the latter case, similarity measures are again needed to define the degree of exactness or fuzziness of the search, for example, to find customers whose spending patterns over time are similar but not necessarily equal to a given spending profile.

Pattern discovery While search & retrieval requires a predefined query, pattern discovery is concerned with automatically discovering interesting patterns in the data without any a-priori assumptions. The term *pattern* usually covers a variety of meanings (see Andrienko et al., 2021), including sequential pattern and periodic pattern, but also temporal association rules. Often, frequently occurring patterns are of interest, for example, when analyzing the impact of promotional activities on increased customer behavior. But rare patterns can also be interesting because they may indicate failures or malicious behavior.

Prediction An important task in analyzing time-oriented data is the prediction of likely future behavior. The goal is to infer from trace data collected in the past how the data, and thus the phenomenon to be understood, will evolve in the future. To achieve this goal, one has to build a predictive model for the data. Examples of such models are autoregressive models, non-stationary and stationary models, or rule-based models.

There are various ways to implement these approaches (e.g., rule-based systems, machine learning, artificial intelligence), but we do not have the space to discuss them here. In the following, we will therefore focus on three practical examples of how VA can be supported through temporal data abstraction, clustering time series, and time series segmentation and labeling. For more information and a broader view of analytical computations on time-oriented data, the interested reader is referred to Mitsa (2010) and Ali et al. (2019).

5.1 Temporal Data Abstraction

The basic idea of *temporal data abstraction* is to use qualitative values, classes, or concepts instead of raw data for visualization and further analysis (see Lin et al., 2007; Combi et al., 2010). Such abstractions make it easier to deal with data size and data complexity. To arrive at appropriate data abstractions, several tasks must be performed, including selecting relevant information, filtering out unneeded information, calculating aggregations, and sorting.

For a brief example of temporal data abstraction, let us look at medical data. Figure 16 shows data from monitoring newborns requiring artificial ventilation. Three time-dependent variables are plotted as points against a horizontal time axis: S_aO_2 (arterial oxygen saturation), $P_{tc}CO_2$ (transcutaneous partial pressure of carbon dioxide), and P_aCO_2 (arterial partial pressure of carbon dioxide). S_aO_2 and $P_{tc}CO_2$ are measured continuously at a regular rate but with different frequencies. New values for P_aCO_2 arrive irregularly and some values for $P_{tc}CO_2$ are missing.

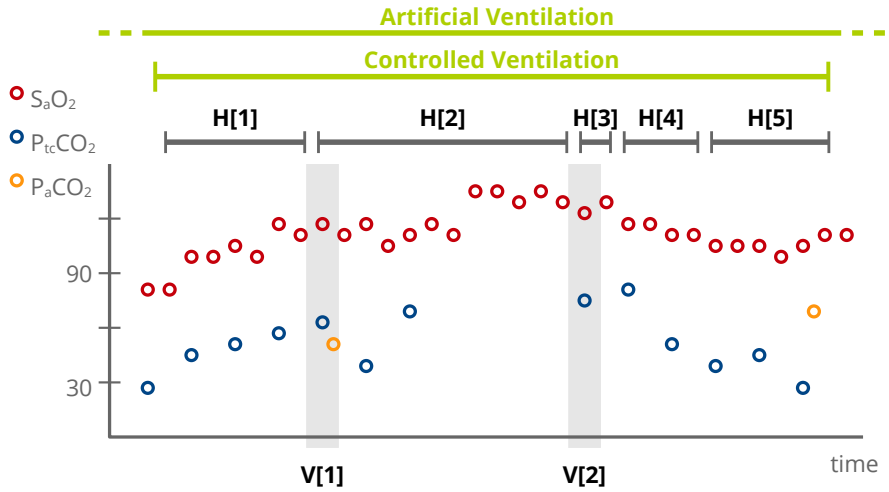


Fig. 16: Temporal data abstraction with vertical temporal abstractions V[1] and V[2] and horizontal temporal abstractions H[1]–H[5] in the context of "artificial ventilation" and its sub-context "controlled ventilation". Aigner et al. (2023)

The aim of temporal data abstraction is to arrive at qualitative values or patterns over time intervals. *Vertical* temporal abstraction (illustrated as V[1] and V[2] in Figure 16) considers multiple variables over a particular time point and combines them into a qualitative value or pattern. For example, V[1] could be a particular state of the ventilated infant according to the three values of S_aO_2, $P_{tc}CO_2$, and P_aCO_2. In this case, three quantitative values from the data were abstracted to a single qualitative state. *Horizontal* temporal abstraction (illustrated as H[1]–H[5])

infers a qualitative value or pattern from one or more variables and a corresponding time interval. For example, the fact that $P_{tc}CO_2$ has decreased from 82 to 55 in a few minutes could result in H[4]: "$P_{tc}CO_2$ is decreasing too fast." Usually, such abstractions are context-dependent, marked in Figure 16 as the context of artificial ventilation and sub-context of controlled ventilation.

5.2 Clustering Time Series

Applying clustering methods is another way to develop understanding through data abstractions (i.e., the clusters) rather than examining individual data values. This makes it possible to analyze much larger datasets. Appropriate *distance* or *similarity measures* lay the groundwork for clustering. These measures are profoundly application-dependent and may involve average geometric distance, longest common subsequences, and also probabilistic models. Based on computed distances, clustering methods create groups of data, where the number of available techniques is large, including hierarchical clustering, partitional clustering, and sequential clustering (see Liao, 2005).

The diversity of available methods makes the selection and appropriate parametrization of algorithms a challenging task. Therefore, it is essential to validate clustering results carefully on a regular basis and ensure the reliability of the clustering process. More details on clustering methods and distance measures can be found in the work by Jain et al. (1999), Gan et al. (2007), and Xu and Wunsch II (2009).

To illustrate the concept of clustering, we use a prominent classic but still relevant work on understanding labor and energy consumption data by van Wijk and van Selow (1999). Their goal was to identify common and uncommon behaviors in datasets with very many time series and to understand their distribution over time. Simply drawing line plots for all time series is not a satisfactory solution in this case because of the resulting severe overplotting. Clustering the time series and using a calendar-based visualization can help overcome this challenge. Let's see how.

Given k daily time series describing some observed variable over a day, the clustering process starts with these k daily series as the initial clusters. Figure 17 shows them at the bottom for an example with $k = 7$. The next step is to compute the differences between all pairs of clusters and to merge the two most similar clusters into a new cluster (i.e., an aggregated representative of the two clusters). This step runs repeatedly and results in a *clustering hierarchy* with $2k - 1$ clusters, where the hierarchy root represents the entire dataset as an aggregated abstraction.

By visualizing the clustered time series in two different views as shown in Figure 18, we may now engage in two analysis tasks: (1) assess similarities among daily behaviors and (2) locate common and uncommon days in time. The first task is facilitated by the basic line plots on the right. They show a selected number of clusters, where each plot uses a unique color. To accomplish the second task, a calendar display (left) is used where individual days are color-coded according to cluster affiliation. With the help of these visualizations, analysts can see the daily

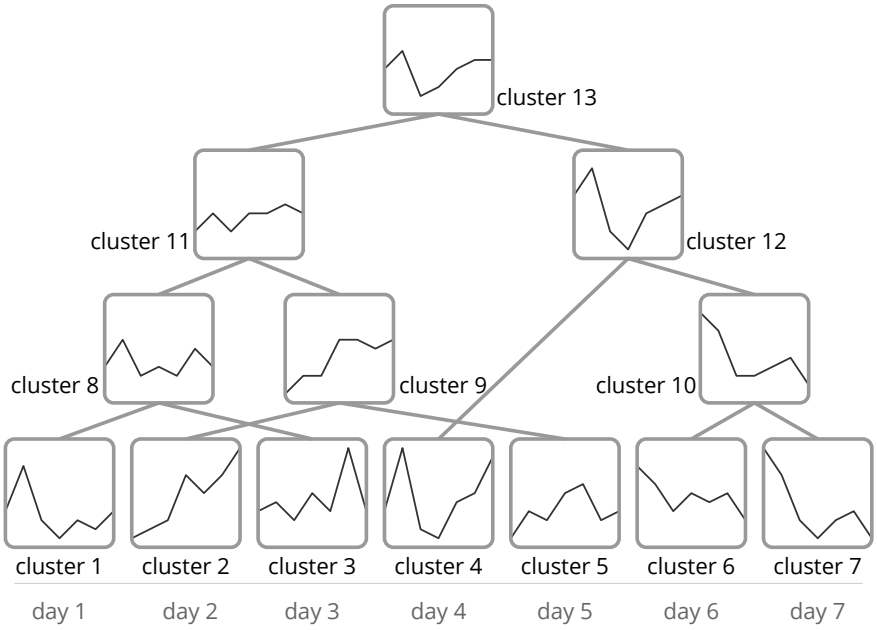


Fig. 17: A clustering hierarchy is generated by repeatedly merging (bottom to top) the two most similar time series into new clusters. The root cluster is an aggregated representative time series of the entire dataset. Aigner et al. (2023)

behaviors represented by the clusters (line plots) and also understand when during a year these behaviors occur (calendar). In combination with interaction methods, the visualization can be adjusted and the data be explored further. For example, to assess similarities, one can select a day from the calendar and, based on the clustering hierarchy, similar days (and clusters) are automatically retrieved and visualized.

For the example of the labor data in Figure 18, van Wijk and van Selow (1999) could draw several conclusions from the cluster and calendar-based analysis:

- Employees follow office hours quite strictly and work between 8:30 am and 5:00 pm in most cases.
- Fewer people work on Fridays during summer (cluster 718).
- During weekends and holidays only very few people are at work (cluster 710).
- It is common practice to take a day off after a holiday (cluster 721).

In summary, the process of clustering can be a critical step in understanding complex trace data. Next, we come to our third and final example of automatic analytical computation and outline the power of time series segmentation and labeling.

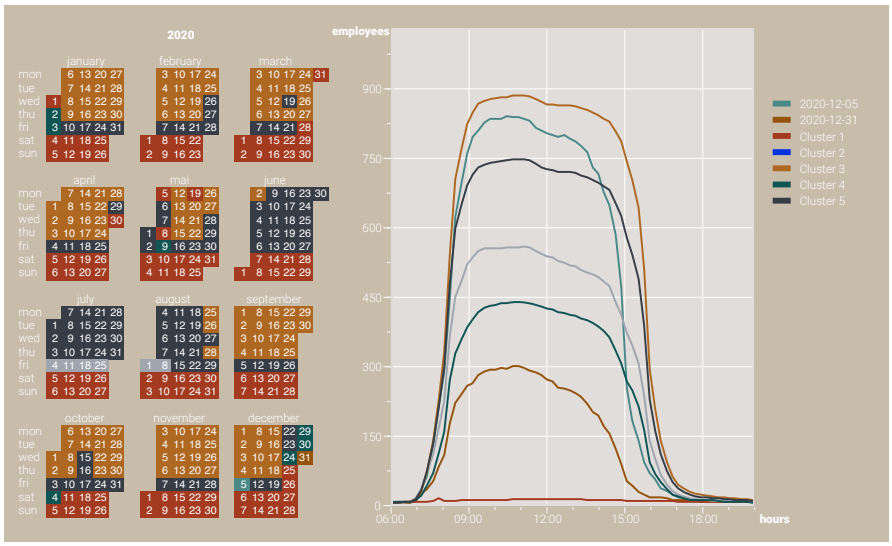


Fig. 18: Visual analysis of the number of employees at work. The calendar on the left shows cluster affiliation by different colors. Plots for selected days and clusters are visualized on the right. Adapted from van Wijk and van Selow (1999).

5.3 Segmentation and Labeling of Multivariate Time Series

The process of segmentation and labeling consists of dividing a long time series into segments (the segmentation) and assigning a class label to each segment (the labeling). The resulting abstractions are simpler than the raw data, can be visualized more compactly, and are therefore easier to understand.

Segmentation algorithms can be distinguished by the type of data they process. There are algorithms for discrete domains (e.g., Cohen et al., 2002) and algorithms for continuous domains (e.g., Lin et al., 2002). Once the segments have been computed, they can be labeled using *supervised* methods, which use already labeled training data, or *unsupervised* methods, which seek hidden structures in unlabeled data autonomously. Supervised methods apply a model to partially or completely label segments (see Xing et al., 2010). Unsupervised methods calculate a grouping that can be used for further aggregation and analysis (see Liao, 2005).

To demonstrate how segmentation and labeling can be integrated into a VA interface, we refer to the VisSect project (see Bernard et al., 2018). This project is concerned with multivariate time series of radiation observations. Figure 19 shows an overview of the interface. On the left side (A), we can see the main steps of the general segmentation workflow. It starts with the selection of the data source and moves on to the parameter visualization and uncertainty analysis, all the way to the user feedback module. In the central part (B), the interface shows three different visualizations highlighting different data characteristics. On the right side (C), we

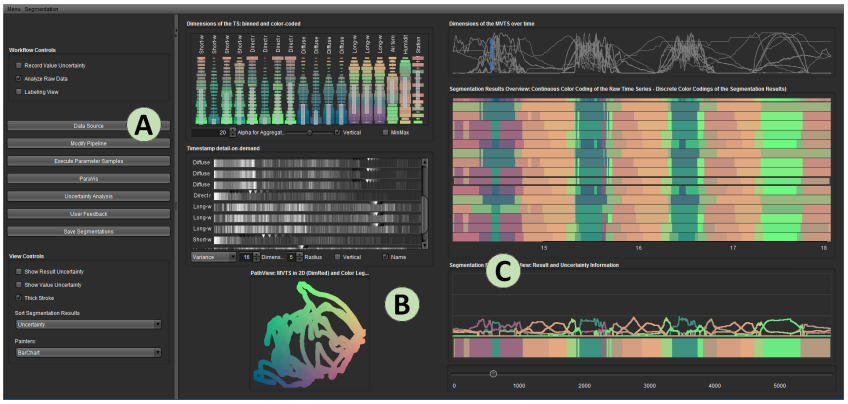


Fig. 19: Segmentation and labeling of radiation observations. (A) General segmentation workflow. (B) Data perspective including raw data, characteristics of individual variables, and dimensionality-reduced plot for patterns. (C) Overview and details of the multivariate time series and the segmentation results. CC BY Aigner et al. (2023)

see the raw data as line plots at the top, and in the middle, colored bars visualize the segmentation results for different parametrizations of the chosen segmentation and labeling algorithms. A detailed inspection of an interactively selected segmentation outcome is possible with the bottom view in part (C).

Taken together, the VisSect interface implements exactly the combination of visual, interactive, and automatic means that is advocated by the VA approach. While the analytical computations (here segmentation and labeling) do their automatic work, the human stays in the loop and can inspect different visualizations to evaluate the produced results and, if necessary, re-parameterize the whole workflow to iteratively optimize the analysis outcome.

At this point, we have completed the VA circle with visual, interactive, and analytical methods. We delineated and analyzed the temporal aspects of trace data through the lens of the VA approach, which encompassed a range of viewpoints and perspectives. Most of the solutions we have demonstrated come from a vibrant VA research community. The next section will shed some light on what future research challenges might be promising to address in the context of VA for digital trace data.

6 Challenges

In the following, we take a brief look at topics for future research with a focus on visualization, interactivity, and analytical computations for time-oriented data.

Cover the specifics of time more broadly While there are already many powerful visualization techniques for trace data with temporal dependencies (the TimeViz

Browser has a corpus of more than 150 techniques), some areas of the time and data characterization spectrum from Section 2 are underrepresented. Future research could develop more techniques for interval-based time, branching time, and multiple perspectives, for displaying raw data and data abstractions simultaneously, and for visualizing data in their spatial frame of reference. Future work could also focus on the multi-scale nature of time and devise new ways to explore, combine, and switch between different levels of abstraction of data and time.

Another question is how to combine multiple visualizations that emphasize specific data aspects. Classic multi-view approaches work with side-by-side arrangement of views. A promising alternative is to consider smooth transitions between views (see Tominski et al., 2021). Exploring the design space for transitioning between non-trivial representations of trace data seems an interesting research direction.

Consider more data aspects On top of the temporal and spatial dependencies in trace data, a third data aspect can be quite relevant: semantic relationships between data items. These are typically modeled as edges between nodes in a graph. When relationships change over time, we are dealing with dynamic graphs or temporal networks (see Holme and Saramäki, 2012). Beck et al. (2017) provide a comprehensive overview of visualization techniques for dynamic graphs, which, however, primarily represent changes in the graph topology (i.e., creation and removal of edges). How to visualize the changes of time-dependent variables being associated with nodes and edges remains a challenging research topic even today.

Furthermore, data quality and uncertainty can be quite important for understanding trace data, and integrating these aspects into VA is the goal of ongoing research. For example, Gschwandtner et al. (2016) study different visual representations for temporal uncertainty. But we still lack generalizable insight on how data quality and uncertainty can be represented and affect understanding. Future work in this direction could strengthen the user's confidence in VA and the obtained findings.

Communicate on more channels In addition to visualizing trace data, other forms of communication are possible. For example, the data could be communicated via sound or haptic sensations (e.g., Braille interfaces). Although these mappings are conceivable in principle, their feasibility and usefulness need to be investigated.

There are similarities between visualization and sonification (i.e., the mapping of data to sound). For example, theoretical studies by Enge et al. (2021) conceptualize space as substrate for visualization marks and time as a substrate for auditory marks. Later, Enge et al. (2022) instantiated their theoretical approach in an exploratory data analysis tool for multivariate data. However, more research is needed to explore how auditory and visual representations can be fruitfully combined for data analysis.

Support a broader range of interactions Similar to using more perceptual channels, there is also room to explore different ways of interacting with data (see Lee et al., 2021). Modalities such as touch and speech are just two examples of modern interaction approaches. Future research could explore these new interaction modalities for their usefulness in the context of trace data analysis.

In addition, interaction for collaborative data analysis is a hot research topic. Large high-resolution displays combined with small mobile displays seem to be particularly suitable for collaboration (see Horak et al., 2018). Large displays naturally lend themselves to large-scale interactive data exploration, while small mobile displays can be used for detailed inspection of data subsets. However, interaction at large scales (e.g., via gaze or physical navigation) and at small scales (e.g., via touch or wrist gestures) requires tailored interaction designs (see Brehmer et al., 2021; Belkacem et al., 2024). Achieving a seamless interaction experience in such cross-device scenarios is a challenge for future work.

Better support for computational analysis We have seen that analytical methods are crucial for understanding large time-oriented data. However, these methods tend to behave like black boxes. They operate on some input data and produce some analytical results, but what happens inside them is often unclear. Therefore, Mühlbacher et al. (2014) propose to open analytical black boxes and make them more transparent and steerable for users. In this regard, Piccolotto et al. (2023) concluded that studying the influence of spatial and temporal parameters on analytical computations is a particular challenge for the future.

Because trace data are typically large, applying analytical methods to them often entails long-running computations. New progressive and incremental approaches as described by Fekete et al. (2024) can help make these computations manageable and steerable. However, it remains a task for future work to either develop new temporal analysis methods that are inherently progressive or to revise and adapt existing methods to give them progressive capabilities where possible.

Focus more on user needs Despite the advances in interactive visual data analysis, appropriate solutions can only be obtained if the data analysis workflows of the users are sufficiently understood. It is generally accepted that the first step in developing VA solutions should be a sound requirements analysis of the given problem domain (see Munzner, 2014; Miksch and Aigner, 2014). This includes specifying (1) the characteristics of the data, (2) the characteristics of the users, and (3) the intentions and tasks of the users. Properly describing, storing, and using this knowledge (see Federico et al., 2017) can help automate the selection or development of VA solutions that match the users, data, and tasks.

In this chapter, we described the TimeViz Browser as a tool for selecting visualization techniques based on data characteristics. However, selection based on analysis tasks is not yet possible, let alone selection based on user characteristics. This would require an easy-to-use way to specify tasks and users, as well as studies to determine which techniques are appropriate for which specifications. Both of these aspects have great potential for future research.

Guide users better From studies by Börner et al. (2019) we know that limited visualization literacy can impede access to valuable information and complicate problem solving and decision making . Therefore, in addition to user-centered design, it is also important to guide users, especially in the context of VA where multiple visualizations, interactions, and computations work together.

To mitigate this, onboarding and guidance play increasingly important roles (see Ceneda et al., 2017; Stoiber et al., 2022). In the initial stage of using VA solutions, onboarding helps people get familiar with new visual representations, interaction techniques, and computational steps. During the use of VA solutions, guidance helps users in situations where they have difficulty making analytical progress on their own. A few onboarding and guidance methods already exist in the literature. In Section 4.2, we briefly illustrated how guidance can look like for multi-scale data exploration. Yet, more research is needed to identify effective onboarding and guidance methods for trace data more broadly. The ultimate goal would be to arrive at true mixed-initiative solutions where both the system and the user can contribute to the progress of the data analysis.

Understand what works Finally, in order to guide users or suggest appropriate techniques to users based on data and tasks, we need more knowledge about what works (see Franconeri et al., 2021). This requires evaluation studies–many of them. Evaluations can test whether the data characteristics are sufficiently communicated, and whether the visual representation, interaction techniques, and analysis methods match the tasks, expectations, and cognitive abilities of the users.

The literature provides a wealth of methodologies for conducting evaluation studies in general (see Lazar et al., 2017). However, Plaisant (2004) and Isenberg et al. (2013) point out that thorough evaluation requires the combined consideration of multiple criteria. Therefore, sophisticated VA solutions with uncertainty visualizations, cross-device interaction, or user guidance can make comprehensive evaluation studies extremely difficult and expensive. Addressing this issue specifically for the visual analysis of trace data remains a challenge for future work. New evaluation strategies such as ICE-T (see Wall et al., 2019) could be helpful in this regard.

7 Conclusion

This chapter shed some light on how VA can support the understanding of temporal aspects of digital trace data. We introduced the basic characteristics of time and time-oriented data and the fundamental methods to visualize them. We saw that interactivity is indispensable in human-in-the-loop knowledge generation processes and how automatic analytical computations can boost data analysis. Finally, we outlined relevant research challenges for VA in the context of digital trace data.

The key take-home messages of this chapter can be summarized as follows:

1. There is more to time than just a linear series of time points, as indicated by the various characteristics of time and time-oriented data.
2. VA combines visualization, interactivity, and automatic analytical computations to support human-in-the-loop sense-making.
3. Visualization design is an intricate process, in particular when trace data with multiple variables need to be understood in time and space.

4. There are many different visualization techniques for time-oriented data. Many are cataloged in the TimeViz Browser.[2]
5. Interactivity is crucial, but also comes at a cost. It allows us to scrutinize and validate our hypotheses. Yet, repetitive manual work should be avoided.
6. Analytical computations can generate abstractions that allow us to handle large and complex data. Interpretable abstractions support human understanding.
7. There is much more to do at the intersection of VA and digital trace data research. Interdisciplinarity seems to be crucial to the success of future research endeavors.

Acknowledgements Parts of this work were funded by the German Research Foundation (DFG) through grant [514630063], the Vienna Science and Technology Fund (WWTF) through grant [10.47379/ICT19047], and the Austrian Science Fund (FWF) through the grants [10.55776/P35767], [10.55776/P33531], and [10.55776/I5622].

References

Aigner, W., S. Miksch, H. Schumann, and C. Tominski (2023). *Visualization of Time-Oriented Data*. Second Edition. Springer. DOI: 10.1007/978-1-4471-7527-8. URL: https://timeviz.net.

Ali, M., A. Alqahtani, M. W. Jones, and X. Xie (2019). "Clustering and Classification for Time Series Data in Visual Analytics: A Survey". In: *IEEE Access* 7, pp. 181314–181338. DOI: 10.1109/ACCESS.2019.2958551.

Andrienko, G., N. Andrienko, P. Bak, D. Keim, and S. Wrobel (2013). *Visual Analytics of Movement*. Springer. DOI: 10.1007/978-3-642-37583-5.

Andrienko, N., G. Andrienko, S. Miksch, H. Schumann, and S. Wrobel (2021). "A Theoretical Model for Pattern Discovery in Visual Analytics". In: *Visual Informatics* 5.1, pp. 23–42. DOI: 10.1016/j.visinf.2020.12.002.

Bach, B., P. Dragicevic, D. W. Archambault, C. Hurter, and S. Carpendale (2017). "A Descriptive Framework for Temporal Data Visualizations Based on Generalized Space-Time Cubes". In: *Computer Graphics Forum* 36.6, pp. 36–61. DOI: 10.1111/cgf.12804.

Baldonado, M. Q. W., A. Woodruff, and A. Kuchinsky (2000). "Guidelines for Using Multiple Views in Information Visualization". In: *Proceedings of the Conference on Advanced Visual Interfaces (AVI)*. ACM Press, pp. 110–119. DOI: 10.1145/345513.345271.

Beck, F., M. Burch, S. Diehl, and D. Weikopf (2017). "A Taxonomy and Survey of Dynamic Graph Visualization". In: *Computer Graphics Forum* 36.1, pp. 133–159. DOI: 10.1111/cgf.12791.

Belkacem, I., C. Tominski, N. Médoc, S. Knudsen, R. Dachselt, and M. Ghoniem (2024). "Interactive Visualization on Large High-Resolution Displays: A Survey". In: *Computer Graphics Forum* 43.6, e15001. DOI: 10.1111/cgf.15001. URL: https://doi.org/10.1111/cgf.15001.

[2] https://browser.timeviz.net

Bernard, J., C. Bors, M. Bögl, C. Eichner, T. Gschwandtner, S. Miksch, H. Schumann, and J. Kohlhammer (2018). "Combining the Automated Segmentation and Visual Analysis of Multivariate Time Series". In: *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)*. Eurographics Association, pp. 49–53. DOI: 10.2312/eurova.20181112.

Bertin, J. (1981). *Graphics and Graphic Information-Processing*. Translated by William J. Berg and Paul Scott. de Gruyter.

Börner, K., A. Bueckle, and M. Ginda (2019). "Data Visualization Literacy: Definitions, Conceptual Frameworks, Exercises, and Assessments". In: *Proceedings of the National Academy of Sciences* 116.6, pp. 1857–1864. DOI: 10.1073/pnas.1807180116.

Brehmer, M., B. Lee, J. Stasko, and C. Tominski (2021). "Interacting with Visualization on Mobile Devices". In: *Mobile Data Visualization*. Ed. by B. Lee, R. Dachselt, P. Isenberg, and E. K. Choe. CRC Press, pp. 67–110. DOI: 10.1201/9781003090823-3.

Ceneda, D., T. Gschwandtner, T. May, S. Miksch, H.-J. Schulz, M. Streit, and C. Tominski (2017). "Characterizing Guidance in Visual Analytics". In: *IEEE Transactions on Visualization and Computer Graphics* 23.1, pp. 111–120. DOI: 10.1109/TVCG.2016.2598468.

Cleveland, W. S. and R. McGill (1984). "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods". In: *Journal of the American Statistical Association* 79.387, pp. 531–554. DOI: 10.1080/01621459.1984.10478080.

Cockburn, A., A. Karlson, and B. B. Bederson (2009). "A Review of Overview+Detail, Zooming, and Focus+Context Interfaces". In: *ACM Computing Surveys* 41.1, 2:1–2:31. DOI: 10.1145/1456650.1456652.

Cohen, P., B. Heeringa, and N. Adams (2002). "Unsupervised Segmentation of Categorical Time Series into Episodes". In: *Proceedings of the International Conference on Data Mining (ICDM)*. IEEE Computer Society, pp. 99–106. DOI: 10.1109/ICDM.2002.1183891.

Combi, C., E. Keravnou-Papailiou, and Y. Shahar (2010). *Temporal Information Systems in Medicine*. Springer. DOI: 10.1007/978-1-4419-6543-1.

Enge, K., A. Rind, M. Iber, R. Höldrich, and W. Aigner (2021). "It's about Time: Adopting Theoretical Constructs from Visualization for Sonification". In: *Proceedings of the International Audio Mostly Conference (AMI)*. ACM Press, pp. 64–71. DOI: 10.1145/3478384.3478415.

Enge, K., A. Rind, M. Iber, R. Höldrich, and W. Aigner (2022). "Towards Multimodal Exploratory Data Analysis: SoniScope as a Prototypical Implementation". In: *Proceedings of the Eurographics / IEEE Conference on Visualization (EuroVis) - Short Papers*. Eurographics Association, pp. 67–71. DOI: 10.2312/evs.20221095.

Federico, P., M. Wagner, A. Rind, A. Amor-Amoros, S. Miksch, and W. Aigner (2017). "The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics". In: *Proceedings of the IEEE Conference on Visual*

Analytics Science and Technology (VAST). IEEE Computer Society, pp. 92–103. DOI: 10.1109/VAST.2017.8585498.

Fekete, J.-D., D. Fisher, and M. Sedlmair, eds. (2024). *Progressive Data Analysis*. Eurographics Association. DOI: 10.2312/pda.20242707.

Franconeri, S. L., L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman (2021). "The Science of Visual Data Communication: What Works". In: *Psychological Science in the Public Interest* 22.3, pp. 110–161. DOI: 10.1177/15291006211051956.

Frank, A. U. (1998). "Different Types of "Times" in GIS". In: *Spatial and Temporal Reasoning in Geographic Information Systems*. Ed. by M. J. Egenhofer and R. G. Golledge. Oxford University Press, pp. 40–62.

Gan, G., C. Ma, and J. Wu (2007). *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability. Society for Industrial and Applied Mathematics. DOI: 10.1137/1.9780898718348.

Goralwalla, I. A., M. T. Özsu, and D. Szafron (1998). "An Object-Oriented Framework for Temporal Data Models". In: *Temporal Databases: Research and Practice*. Ed. by O. Etzion, S. Jajodia, and S. Sripada. Springer, pp. 1–35. DOI: 10.1007/bfb0053696.

Gschwandtner, T., M. Bögl, P. Federico, and S. Miksch (2016). "Visual Encodings of Temporal Uncertainty: A Comparative User Study". In: *IEEE Transactions on Visualization and Computer Graphics* 22.1, pp. 539–548. DOI: 10.1109/TVCG.2015.2467752.

Hajnicz, E. (1996). *Time Structures: Formal Description and Algorithmic Representation*. Vol. 1047. Lecture Notes in Computer Science. Springer. DOI: 10.1007/3-540-60941-5.

Hand, D. J. (2016). *Measurement: A Very Short Introduction*. Oxford University Press. DOI: 10.1093/actrade/9780198779568.001.0001.

Hauser, H. and H. Schumann (2009). "Visualization Pipeline". In: *Encyclopedia of Database Systems*. Ed. by L. Liu and M. T. Özsu. Boston, MA: Springer US, pp. 3414–3416. DOI: 10.1007/978-0-387-39940-9_1133.

Heer, J. and M. Bostock (2010). "Crowdsourcing Graphical Perception: Using Mechanical Turk to Assess Visualization Design". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM Press, pp. 203–212. DOI: 10.1145/1753326.1753357.

Hochheiser, H. and B. Shneiderman (2004). "Dynamic Query Tools for Time Series Data Sets: Timebox Widgets for Interactive Exploration". In: *Information Visualization* 3.1, pp. 1–18. DOI: 10.1057/palgrave.ivs.9500061.

Holme, P. and J. Saramäki (2012). "Temporal Networks". In: *Physics Reports* 519.3, pp. 97–125. DOI: 10.1016/j.physrep.2012.03.001.

Horak, T., S. K. Badam, N. Elmqvist, and R. Dachselt (2018). "When David Meets Goliath: Combining Smartwatches with a Large Vertical Display for Visual Data Exploration". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM Press. DOI: 10.1145/3173574.3173593.

Howison, J., A. Wiggins, and K. Crowston (2011). "Validity Issues in the Use of Social Network Analysis with Digital Trace Data". In: *Journal of the Association for Information Systems* 12.12, p. 2. DOI: 10.17705/1jais.00282.

Isenberg, T., P. Isenberg, J. Chen, M. Sedlmair, and T. Möller (2013). "A Systematic Review on the Practice of Evaluating Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12, pp. 2818–2827. DOI: 10.1109/TVCG.2013.126.

Jain, A. K., M. N. Murty, and P. J. Flynn (1999). "Data Clustering: A Review". In: *ACM Computing Surveys* 31.3, pp. 264–323. DOI: 10.1145/331499.331504.

Jensen, C. S., C. E. Dyreson, M. H. Böhlen, J. Clifford, R. Elmasri, S. K. Gadia, F. Grandi, P. J. Hayes, S. Jajodia, W. Käfer, N. Kline, N. A. Lorentzos, Y. G. Mitsopoulos, A. Montanari, D. A. Nonen, E. Peressi, B. Pernici, J. F. Roddick, N. L. Sarda, M. R. Scalas, A. Segev, R. T. Snodgrass, M. D. Soo, A. U. Tansel, P. Tiberio, and G. Wiederhold (1998). "The Consensus Glossary of Temporal Database Concepts – February 1998 Version". In: *Temporal Databases: Research and Practice*. Ed. by O. Etzion, S. Jajodia, and S. Sripada. Springer, pp. 367–405. DOI: 10.1007/bfb0053710.

Keim, D., G. Andrienko, J.-D. Fekete, C. örg, J. Kohlhammer, and G. Melançon (2008). "Visual Analytics: Definition, Process, and Challenges". In: *Information Visualization: Human-Centered Issues and Perspectives*. Ed. by A. Kerren, J. T. Stasko, J.-D. Fekete, and C. North. Springer, pp. 154–175. DOI: 10.1007/978-3-540-70956-5_7.

Keim, D., J. Kohlhammer, G. Ellis, and F. Mansmann, eds. (2010). *Mastering the Information Age – Solving Problems with Visual Analytics*. Eurographics Association. URL: https://diglib.eg.org/handle/10.2312/14803.

Keim, D. A., F. Mansmann, J. Schneidewind, and H. Ziegler (2006). "Challenges in Visual Data Analysis". In: *Proceedings of the International Conference Information Visualisation (IV)*. IEEE Computer Society, pp. 9–16. DOI: 10.1109/IV.2006.31.

Kondo, B. and C. Collins (2014). "DimpVis: Exploring Time-varying Information Visualizations by Direct Manipulation". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 2003–2012. DOI: 10.1109/TVCG.2014.2346250.

Kraak, M.-J. (2003). "The Space-Time Cube Revisited from a Geovisualization Perspective". In: *Proceedings of the 21st International Cartographic Conference (ICC)*. The International Cartographic Association (ICA), pp. 1988–1996. URL: https://icaci.org/files/documents/ICC_proceedings/ICC2003/Papers/255.pdf.

Lam, H. (2008). "A Framework of Interaction Costs in Information Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 14.6, pp. 1149–1156. DOI: 10.1109/TVCG.2008.109.

Laxman, S. and P. S. Sastry (2006). "A Survey of Temporal Data Mining". In: *Sādhanā* 31, pp. 173–198. DOI: 10.1007/bf02719780.

Lazar, J., J. H. Feng, and H. Hochheiser (2017). *Research Methods in Human-Computer Interaction*. 2nd ed. John Wiley & Sons, Ltd.

Lee, B., A. Srinivasan, P. Isenberg, and J. T. Stasko (2021). "Post-WIMP Interaction for Information Visualization". In: *Foundations and Trends in Human-Computer Interaction* 14.1, pp. 1–95. DOI: 10.1561/1100000081.

Lenz, H. (2005). *Universalgeschichte der Zeit*. Wiesbaden, Germany: Marixverlag, p. 575.

Liao, T. W. (2005). "Clustering of Time Series Data – A Survey". In: *Pattern Recognition* 38.11, pp. 1857–1874. DOI: 10.1016/j.patcog.2005.01.025.

Lin, J., E. Keogh, S. Lonardi, and P. Patel (2002). "Finding Motifs in Time Series". In: *Proceedings of the SIGKDD Workshop on Temporal Data Mining*, pp. 53–68. URL: https://cs.gmu.edu/~jessica/Lin_motif.pdf.

Lin, J., E. J. Keogh, L. Wei, and S. Lonardi (2007). "Experiencing SAX: A Novel Symbolic Representation of Time Series". In: *Data Mining and Knowledge Discovery* 15.2, pp. 107–144. DOI: 10.1007/s10618-007-0064-z.

Liu, Z. and J. Heer (2014). "The Effects of Interactive Latency on Exploratory Visual Analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 2122–2131. DOI: 10.1109/TVCG.2014.2346452.

Luboschik, M., C. Maus, H.-J. Schulz, H. Schumann, and A. Uhrmacher (2012). "Heterogeneity-Based Guidance for Exploring Multiscale Data in Systems Biology". In: *Proceedings of the IEEE Symposium on Biological Data Visualization (BioVis)*. IEEE Computer Society, pp. 33–40. DOI: 10.1109/BioVis.2012.6378590.

McColeman, C. M., F. Yang, T. F. Brady, and S. Franconeri (2022). "Rethinking the Ranks of Visual Channels". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1, pp. 707–717. DOI: 10.1109/TVCG.2021.3114684.

Mennis, J. L., D. J. Peuquet, and L. Qian (2000). "A Conceptual Framework for Incorporating Cognitive Principles into Geographical Database Representation". In: *International Journal of Geographical Information Science* 14.6, pp. 501–520. DOI: 10.1080/136588100415710.

Miksch, S. and W. Aigner (2014). "A Matter of Time: Applying a Data-Users-Tasks Design Triangle to Visual Analytics of Time-Oriented Data". In: *Computers & Graphics* 38, pp. 286–290. DOI: 10.1016/j.cag.2013.11.002.

Mitsa, T. (2010). *Temporal Data Mining*. Chapman & Hall/CRC. DOI: 10.1201/9781420089776.

Moreland, K. (2013). "A Survey of Visualization Pipelines". In: *IEEE Transactions on Visualization and Computer Graphics* 19.3, pp. 367–378. DOI: 10.1109/TVCG.2012.133.

Mühlbacher, T., H. Piringer, S. Gratzl, M. Sedlmair, and M. Streit (2014). "Opening the Black Box: Strategies for Increased User Involvement in Existing Algorithm Implementations". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12, pp. 1643–1652. DOI: 10.1109/TVCG.2014.2346578.

Munzner, T. (2014). *Visualization Analysis and Design*. A K Peters/CRC Press. DOI: 10.1201/b17511.

Norman, D. A. (2013). *The Design of Everyday Things*. Revised and expanded edition. Basic Books.

Piccolotto, N., M. Bögl, and S. Miksch (2023). "Visual Parameter Space Exploration in Time and Space". In: *Computer Graphics Forum* 42.6. DOI: 10.1111/cgf.14785.

Plaisant, C. (2004). "The Challenge of Information Visualization Evaluation". In: *Proceedings of the Conference on Advanced Visual Interfaces (AVI)*. ACM Press, pp. 106–119. DOI: 10.1145/989863.989880.

Saket, B., S. Huron, C. Perin, and A. Endert (2020). "Investigating Direct Manipulation of Graphical Encodings as a Method for User Interaction". In: *IEEE Transactions on Visualization and Computer Graphics* 26.1, pp. 482–491. DOI: 10.1109/TVCG.2019.2934534.

Schulz, H.-J., M. Streit, T. May, and C. Tominski (2013). *Towards a Characterization of Guidance in Visualization*. Poster at IEEE Conference on Information Visualization (InfoVis). Atlanta, USA.

Shneiderman, B. (1983). "Direct Manipulation: A Step Beyond Programming Languages". In: *IEEE Computer* 16.8, pp. 57–69. DOI: 10.1109/mc.1983.1654471.

Shneiderman, B. (1994). "Dynamic Queries for Visual Information Seeking". In: *IEEE Software* 11.6, pp. 70–77. DOI: 10.1109/52.329404.

Shneiderman, B. (1996). "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In: *Proceedings of the IEEE Symposium on Visual Languages*. IEEE Computer Society, pp. 336–343. DOI: 10.1109/VL.1996.545307.

Steiner, A. (1998). "A Generalisation Approach to Temporal Data Models and their Implementations". PhD thesis. Swiss Federal Institute of Technology.

Stoiber, C., D. Ceneda, M. Wagner, V. Schetinger, T. Gschwandtner, M. Streit, S. Miksch, and W. Aigner (2022). "Perspectives of Visualization Onboarding and Guidance in VA". In: *Visual Informatics* 6 (1), pp. 68–83. DOI: 10.1016/j.visinf.2022.02.005.

Stoiber, C., A. Rind, F. Grassinger, R. Gutounig, E. Goldgruber, M. Sedlmair, Š. Emrich, and W. Aigner (2019). "netflower: Dynamic Network Visualization for Data Journalists". In: *Computer Graphics Forum* 38.3, pp. 699–711. DOI: 10.1111/cgf.13721.

Thomas, J. J. and K. A. Cook (2005). *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society.

Tominski, C. and W. Aigner (2023). *The TimeViz Browser – A Visual Survey of Visualization Techniques for Time-Oriented Data*. Version 2.0. URL: https://browser.timeviz.net.

Tominski, C., G. Andrienko, N. Andrienko, S. Bleisch, S. I. Fabrikant, E. Mayr, S. Miksch, M. Pohl, and A. Skupin (2021). "Toward Flexible Visual Analytics Augmented through Smooth Display Transitions". In: *Visual Informatics* 5.3, pp. 28–38. DOI: 10.1016/j.visinf.2021.06.004.

Tominski, C., S. Gladisch, U. Kister, R. Dachselt, and H. Schumann (2017). "Interactive Lenses for Visualization: An Extended Survey". In: *Computer Graphics Forum* 36.6, pp. 173–200. DOI: 10.1111/cgf.12871.

Tominski, C., P. Schulze-Wollgast, and H. Schumann (2005). "3D Information Visualization for Time Dependent Data on Maps". In: *Proceedings of the International Conference Information Visualisation (IV)*. IEEE Computer Society, pp. 175–181. DOI: 10.1109/IV.2005.3.

Tominski, C. and H. Schumann (2008). "Enhanced Interactive Spiral Display". In: *Proceedings of the Annual Conference of the Swedish Computer Graphics Association (SIGRAD)*. Linköping University Electronic Press, pp. 53–56. URL: https://ep.liu.se/ecp/034/013/ecp083413.pdf.

Tominski, C. and H. Schumann (2020). *Interactive Visual Data Analysis*. AK Peters Visualization Series. CRC Press. DOI: 10.1201/9781315152707.

Tominski, C., H. Schumann, G. Andrienko, and N. Andrienko (2012). "Stacking-Based Visualization of Trajectory Attribute Data". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12, pp. 2565–2574. DOI: 10.1109/TVCG.2012.265.

Van Berkel, N., B. Tag, R. M. Jacobsen, D. Russo, H. C. Purchase, and D. Buschek (2024). "Impact of Interaction Technique in Interactive Data Visualisations: A Study on Lookup, Comparison, and Relation-seeking Tasks". In: *International Journal of Human-Computer Studies* 192, p. 103359. DOI: 10.1016/j.ijhcs.2024.103359.

Van Wijk, J. J. and E. R. van Selow (1999). "Cluster and Calendar Based Visualization of Time Series Data". In: *Proceedings of the IEEE Symposium Information Visualization (InfoVis)*. IEEE Computer Society, pp. 4–9. DOI: 10.1109/INFVIS.1999.801851.

Wall, E., M. Agnihotri, L. Matzen, K. Divis, M. Haass, A. Endert, and J. Stasko (2019). "A Heuristic Approach to Value-Driven Evaluation of Visualizations". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1, pp. 491–500. DOI: 10.1109/TVCG.2018.2865146.

Ward, M. O., G. Grinstein, and D. Keim (2015). *Interactive Data Visualization: Foundations, Techniques, and Applications*. 2nd ed. A K Peters/CRC Press. DOI: 10.1201/b18379.

Ware, C. (2020). *Information Visualization: Perception for Design*. 4th ed. Morgan Kaufmann. DOI: 10.1016/C2016-0-02395-1.

Wills, G. (2012). *Visualizing Time - Designing Graphical Representations for Statistical Data*. Springer. DOI: 10.1007/978-0-387-77907-2.

Xing, Z., J. Pei, and E. Keogh (2010). "A Brief Survey on Sequence Classification". In: *SIGKDD Explorations Newsletter* 12.1, pp. 40–48. DOI: 10.1145/1882471.1882478.

Xu, R. and D. C. Wunsch II (2009). *Clustering*. John Wiley & Sons. DOI: 10.1002/9780470382776.

Yi, J. S., Y. ah Kang, J. T. Stasko, and J. A. Jacko (2007). "Toward a Deeper Understanding of the Role of Interaction in Information Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6, pp. 1224–1231. DOI: 10.1109/TVCG.2007.70515.