City Research Online

City, University of London Institutional Repository

Citation: Andrienko, N. & Andrienko, G. (2023). It's about Time: Analytical Time Periodization. Computer Graphics Forum, 42(6), e14845. doi: 10.1111/cgf.14845

This is the published version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: https://openaccess.city.ac.uk/id/eprint/30603/

Link to published version: https://doi.org/10.1111/cgf.14845

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

City Research Online: http://openaccess.city.ac.uk/ publications@city.ac.uk

DOI: 10.1111/cgf.14845

COMPUTER GRAPHICS *forum*
Volume 0 (2023), number 0 pp. 1–20

It's about Time: Analytical Time Periodization

Natalia Andrienko[1,2] and Gennady Andrienko[1,2]

[1]Fraunhofer Institute IAIS, Sankt Augustin, Germany
{natalia.andrienko, gennady.andrienko}@iais.fraunhofer.de
[2]City University of London, London, UK

Abstract

This paper presents a novel approach to the problem of time periodization, which involves dividing the time span of a complex dynamic phenomenon into periods that enclose different relatively stable states or development trends. The challenge lies in finding such a division of the time that takes into account diverse behaviours of multiple components of the phenomenon while being simple and easy to interpret. Despite the importance of this problem, it has not received sufficient attention in the fields of visual analytics and data science. We use a real-world example from aviation and an additional usage scenario on analysing mobility trends during the COVID-19 pandemic to develop and test an analytical workflow that combines computational and interactive visual techniques. We highlight the differences between the two cases and show how they affect the use of different techniques. Through our investigation of possible variations in the time periodization problem, we discuss the potential of our approach to be used in various applications. Our contributions include defining and investigating an earlier neglected problem type, developing a practical and reproducible approach to solving problems of this type, and uncovering potential for formalization and development of computational methods.

Keywords: multivariate time series, time, visual analytics, visualization

CCS Concepts: [Human-centred computing → Visual analytics]: Visualization application domains—Visual analytics

1. Introduction

Periodisation is defined in dictionaries as the act or process of dividing history into periods (e.g. [Col22]). "Periodisation is a form of historical understanding, designed as a historiological tool for making the past understandable, intelligible, and meaningful by dividing it into compartments" [Sat01]. The concept of time periodization reflects the view of time as both a continuum and a process of perpetual change, which implies that any description of time needs to emphasize continuity at some points and difference at others [Blo54]. Attempts to represent continuity and differences result in defining time periods and boundaries between them.

Time periodization is pertinent not just to history. Generally, any process can be divided into meaningful phases. Dividing the time during which something exists, functions, or happens into periods that can be characterized synoptically and compared to each other [Hay98] is an instrument for analytical abstraction and synthesis of general knowledge. In this function, time periodization is used in various domains, for example:

- in marketing – to identify patterns and trends in consumer behaviour, such as seasonal variation in demands and changes in preferences and interests, in order to develop effective marketing strategies and make sensible decisions about promotions, launching new products, or pricing adjustments;
- in crime analysis – to grasp and characterize variations of criminal activities over time, which helps police offices to deploy their resources more effectively and law enforcement agencies to develop strategies for crime prevention and policing;
- in finance – to understand the developments in stock prices, interest rates, and other financial indicators for developing reasoned investment strategies and making timely decisions about buying and selling;
- in sport analysis – to define different strategies applied by players or teams in sport games, which can help their opponents to prepare to forthcoming games.

Data-driven time periodization, which is practised in these and other application domains, is based on analysis of time series and/or

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

event sequence data with the help of computational methods, such as clustering, time series decomposition, and breakpoints detection (e.g. [BPJ*20]). However, these methods alone may not produce results that are meaningful to human analysts and convenient for using in the further analysis or for presenting and explaining to others. Analysts need to review the results and adjust them in accord with domain knowledge and requirements of the further use. Results of computations can be improved by varying method parameters [BPJ*20] but still may not be fully satisfactory. This is where visual analytics combining computational techniques with interactive visual interfaces can be of great help.

Using visual analytics approaches, a human analyst can intelligently and flexibly combine information about varying characteristics of a studied phenomenon with domain knowledge of the nature of the changes and their relationships to the structure and properties of time [AAD*10, AAF*20]. For example, an analyst dealing with a social phenomenon expects differences between weekdays and weekends, whereas a climate researcher is aware of possible effects of the solar cycle. An analyst may decide to ignore casual abrupt variations of characteristics that break continuity of time periods and unite the differing time steps with their neighbours. However, the analyst may also deem the outstanding time steps to require a special consideration and therefore decide to organize them in a separate non-contiguous time period. To make such decisions and implement them, analysts need appropriate visual representation of data reflecting the temporal variation and tools enabling interactive division of the time and/or modification of results of algorithmic division.

It is not unusual that dynamic phenomena under analysis consist of heterogeneous parts whose characteristics vary in somewhat different ways. For example, students and working people may have different rhythms of social activities, groups of customers may differ in their shopping behaviour, and development trends may vary among sectors of economy. Like in periodization of historical time, where scholars bring together individual developments of different countries, analysts of time-related data may need to create an overarching division of time that respects essential features of temporal variation pertinent to different parts but is uncomplicated and easily interpretable. As a starting point, an analyst may consider several periodizations made individually for each of the heterogeneous parts. The analyst can note similarities, assess the importance of differences, take suitable time periods from different divisions, and make adjustments to capture important patterns of temporal variation of each part while ignoring irrelevant individual features. There is a need in interactive visual interfaces that can support integration of several time divisions into a single overall division.

In this paper, we propose a workflow and a combination of tools for interactive division of time into meaningful and manageable periods, which may include creation and subsequent integration of multiple different divisions. To make the concepts clearer to the readers, we begin with introducing a motivating example and then present a general formulation of the time periodization problem (Section 2). After an overview of the related work in Section 3, we use the example to present the proposed workflow for solving the problem and define a combination of tools can support the fulfilment of this workflow (Section 4). We deem important to note that we strive to define the tools in a general way, that is in terms of their functions, acknowledging the possibility of different realisations. In Section 5, we test the workflow by applying it to a different example, which allows us to demonstrate how the approach may vary depending on properties of the data. This is followed by a discussion of our contribution in Section 6 and final conclusion in Section 7.

2. Problem Statement

2.1. Motivating example

The SIMBAD project ("Combining Simulation Models and Big Data Analytics for ATM Performance Analysis", https://www.simbad-h2020.eu/) aims to develop methods and models for analysing and understanding the performance of the Air Traffic Management (ATM) system in Europe. This project responds to the need for representing the annual variation of air traffic with simulation models that can be used for prediction and planning. However, the ATM system is highly complex, with over 30,000 daily flights affected by the decisions and interactions of multiple stakeholders and external variables such as seasonality, weather and variation of demands. Consequently, the traffic flows on different days can vary substantially, and a single simulation model cannot accurately represent the system's performance throughout the year. This highlights the need to develop a system of multiple models that can represent different periods of the year accurately. To be practically usable, the models need to be relatively few in number, and it must be clear when to use each model and why. To achieve this, it is necessary to divide the 1-year time span into well-defined periods so that each period's traffic can be accurately represented by a single simulation model. This task has been in the focus of our research.

The data under analysis describe air traffic situations that existed in numerous different time steps, such as days of a year. A traffic situation in each time step is a spatial distribution of the air traffic. It is characterized by a number of variables with values referring to different spatial objects or regions, such as airports or airspace sectors. These variables capture various aspects of air traffic, including traffic volumes and performance indicators such as flight delays.

The temporal variation of traffic characteristics can vary significantly across different parts of the traffic network or subspaces of the airspace, which complicates the task of time division. For instance, seasonal or weekly variation patterns may be more pertinent to some elements of the network (airports, regions, connections, etc.) than to others. Furthermore, in some parts of the network, weekly variation patterns may be more pronounced during certain seasons. The diversity of the temporal patterns is illustrated by the map fragments shown in Figure 1, where time series of daily flight counts by airports are represented by mosaic glyphs with colours representing normalized deviations of the flight counts from the airport-specific means. It is crucial to take into account the heterogeneity of the temporal variation patterns across the network while performing time division. Otherwise, the result may unify time steps with situations that appear similar on the overall scale but differ considerably in some part(s) of the network. Statistical measures alone may be insufficient to reveal such differences. It requires human expert knowledge and reasoning to explore the heterogeneity of temporal variation patterns across the network and define time periods accordingly.

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

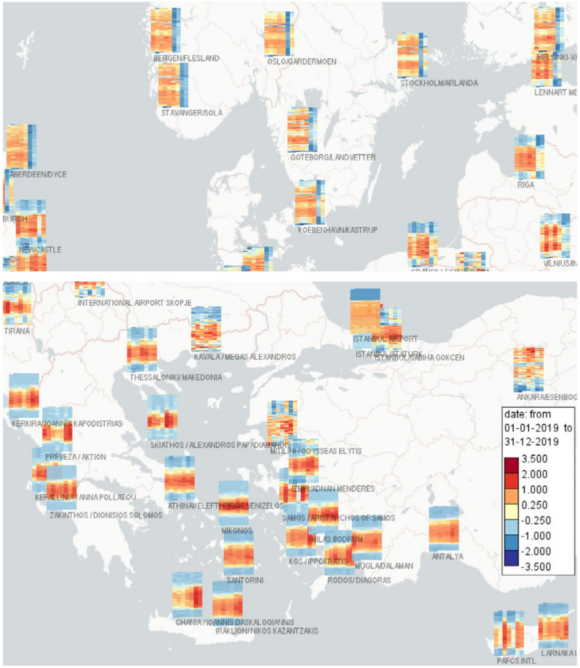

Figure 1: *Fragments of a map showing time series of daily counts of flights (normalized by transforming to z-scores) served in airports. The time series are represented by mosaic glyphs with columns corresponding to 7 days of a week and rows to 53 weeks of a year. The values are encoded using a diverging colour scale with shades of blue used for negative values (i.e. below the means) and red for positive values (above the means). The map fragments demonstrate heterogeneity of the temporal variation patterns.*

2.2. General formulation

Context: A dynamic phenomenon involves multiple components or objects with time-varying characteristics. The changes of the characteristics of each component are represented by values of one or more attributes (variables) specified for consecutive time steps, which may be time moments or intervals. Hence, there is a dataset consisting of time series of attribute values referring to different objects or components. The data cover a long extent of time, so that the time series include a large number of time steps.

Problem: To understand and describe how the whole dynamic phenomenon evolves over time, taking into account that its components may have different temporal patterns of changes, the long sequence of time steps needs to be reduced to a small number of time periods encapsulating relatively stable states of the phenomenon or continuing tendencies of its development. This means that characteristics of the phenomenon or consecutive changes of those need to be similar within each period and differ from other periods. The division of the time into periods must take into account essential differences between development patterns of diverse components or objects in the phenomenon.

General idea of the approach: First, the set of objects or components is divided into subsets whose members have common patterns of changes. Next, time division is performed separately for each subset based on its common pattern of change. Finally, multiple divisions are merged into a singe division by a human expert seeking an appropriate balance between simplicity and accuracy of capturing different patterns of change. This approach needs to be supported by an appropriate combination of computational, visual and interactive techniques arranged in a workflow.

3. Related Work

Problem decomposition based on partitioning data sets into subsets is one of main principles of data analysis and model building [AAF*20] (sect.13.8). Partitioning usually divides a data set into a manageable number of subsets, such that items within each subset are similar and differ from items in other subsets. Geographic partitioning was a key theme in geospatial data analysis research for decades [AAD*10] resulting in methods that take into account specific features of the geographic space and distributions of spatial phenomena.

For temporal data, partitioning of time has been considered in the context of the problem of data abstraction. Abstraction of temporal data results in a sequence of time intervals with discrete labels assigned to them [HÖ2]. Aigner et al. state that the labels need to convey key ideas [AMM*08, AMST11]. This aligns with the AI view of temporal abstraction as a type of a generic interpretation task that aims at deriving higher-level concepts [Sha97].

The problem of time partitioning and abstraction can be considered either as a supervised or as an unsupervised task [HÖ2]. Supervised abstraction means finding time intervals where data have certain properties defined a priori and assigning corresponding labels from a given set. Unsupervised time abstraction means that no concepts are defined in advance; they have to be learned from data. Our work relates to the task of unsupervised time abstraction.

For supervised abstraction, there are methods detecting predefined shapes in univariate time series, such as constancy, increase, and decrease, where increase or decrease can be linear, convex, or concave [HÖ2]. Aigner et al. [AMM*08, AMST11] describe examples of transformation of numeric values to categorical labels using domain-specific rules. Computational techniques are used to find sufficiently long time spans of stable values. The theory of knowledge-based temporal abstraction (KBTA) developed by Shahar [Sha97] suggests that the task is done by matching data from time intervals with patterns defined in domain ontology. Temporal abstractions can also be defined using event calculus [CC99]. To support this, Chittaro et al. [CC01] developed a visual interface where users can specify relationships between time intervals and logical relationships between conditions. The visual language was later extended to allow specification of durations, granularities, and other time-related features [CO12]. Sacchi et al. [SLCB07] propose a data mining framework for identifying recurring temporal relationships between occurrences of user-specified patterns of interest.

For unsupervised abstraction, a possible general approach is to identify similar parts in the time series by means of clustering [HÖ2]. A well-known example is the work by van Wijk and van

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Selow [vWvS99], who applied clustering to data characterising days of a year and showed the resulting time clusters in a calendar view, which revealed weekly and seasonal patterns of data variation. The idea of clustering time steps based on their features was further developed for spatially referenced time series of attribute values [GCML06, AAB*10], dynamic mobility graphs [vBR*16], and time series of spatial distributions of movement data [AAFW17]. These approaches implement a common workflow: define an appropriate similarity measure (distance function) for the features of the time steps, apply some clustering method, and visualize cluster membership of the time steps and summary features for the clusters.

Grouping of time steps can be based on user-specified queries, which provides the advantages of full user control, understanding of the grouping principles and ease of interpretation of the resulting subsets of time steps, which, however, may not form continuous periods. Computational methods for performing time queries on large datasets were developed in temporal databases [Gad88, JCG*92]. Following the ideas of interactive and dynamic queries [Shn94, HS04, BAP*05], a Time Mask query interface allows interactive selection of time intervals with user-defined properties [AAC*17]. A user observes an immediate feedback in visual displays of the data after modifying query conditions. Later extensions of time query operations [AAA*21] enable ignoring intermittent brief intervals of query satisfaction or absence of satisfaction and adding temporal buffers before and/or after intervals selected by a query. These operations may help users to build continuous time periods from disjoint pieces.

Partitioning of numeric time series into time periods can be achieved using algorithms for time series segmentation [GYD*19]. It can be combined with interactive visual techniques [BDB*16, BBB*18], which may involve embedding of time steps into a low-dimensional (typically 2D) space based on similarity of their characteristics. The simplest embedding is a scatter plot where positions correspond to values of two attributes and points representing consecutive time steps are connected by lines [Phi58]. Projection of more complex temporal data, such as values of multiple attributes, graphs, statistical or spatial distributions, or traffic situations, is done by means of dimensionality reduction (DR) methods, such as MDS[Kru64] or t-SNE [vdMH08]. DR can be applied to individual time steps (e.g. [vdEHBvW16]) or to previously clustered and aggregated data [HWX*10, BWK*13]. Connecting consecutive points by lines in such projection forms a so-called Time Curve [BSH*16]. Different motifs in Time Curves can be interpreted semantically as gradual or rapid changes, stable states, oscillation, stagnation, etc. [BSH*16, vdEHBvW16]. Bernard et al. [BWS*12] assign colours to the positions in the projection space and represent the temporal variation of the value combinations by variation of colours along the time axis. Recent extensions of Time Curve, which include depiction of the point density and representation of distribution statistics [BSP*22], allow assessing a likely number of clusters in the data. This work connects two general approaches to analysis of complex time-related data: time clustering and time embedding [AAF*20].

Our work builds on these approaches by combining their strengths. The research on the application of time step embedding [BSH*16, vdEHBvW16] proved to be especially important for developing our approach.

4. Workflow and Methods

We present an analytical workflow that involves several steps. First, we clean the data to remove any occasional variations that may obscure the general patterns. Next, we divide the data into subsets that exhibit coherent patterns of temporal variation. Once we have identified these subsets, we define specific time periods for each of them. After this, we integrate all the subsets into a unified set of time periods. Finally, we characterize the specific features of each period to gain a deeper understanding of the data. We have developed this workflow using the example of the air traffic data, specifically, the daily counts of the flight departures from airports of Europe and a few neighbouring countries. The dataset covers the entire year of 2019, which was the last year of normal air traffic before the Covid-19 pandemic resulted in travel restrictions.

To examine the temporal variation in flight departures independent of airport sizes and capacities, we normalized the time series by converting the flight counts into z-scores. This involved calculating the differences between the flight counts and the means of the time series, and then dividing these differences by the respective standard deviations. To provide a visual representation of the time series, we created glyphs in the form of a matrix. Each glyph consists of seven columns for the days of the week and 53 rows for the weeks of the year. We used a diverging colour scale [HB03] with shades of blue for negative values (below the means) and red for positive values (above the means) to depict the values of each cell in the matrix.

4.1. Step 1. Clean data from occasional variations

Any real-world data set requires careful investigation of data properties and assessment of data quality [LAW*18], which includes detection of incorrect, incomplete, duplicate or otherwise erroneous data and fixing the problems thus revealed. For time series data, it is necessary to investigate rapid changes and other kinds of surprising patterns that may be indicative of errors or anomalies [GAM*14]. Although these unusual patterns may not be inherently wrong, they may need to be ignored, particularly when the analysis task requires generalization. Our goal is to capture the general features of the development of air traffic in Europe over the year, and thus, we need to detect and discard cases of extraordinary temporal variation that might have occurred due to specific local conditions or occasional circumstances. To achieve this, we use a 2D embedding technique to project the time series into a low-dimensional space based on similarities between them expressed by an appropriate distance metric. Exceptional data samples are spatially separated from the bulk in such an embedding. In our example, we use Euclidean distance as the distance metric, Sammon's mapping [Sam69] as the projection method, and create a projection plot representing the time series by mosaic glyphs, as shown in Figure 2.

On the left side of the projection plot in Figure 2, we see two oddly painted mosaics, marked by black frames. The upper glyph corresponds to Milano Linate, which has a blue stripe indicating very low values in the period from August to October. The lower mosaic corresponds to Istanbul Ataturk airport, indicating a sharp drop in the number of flights after the beginning of the year. In both cases, the manner of temporal variation is surprising. To investigate

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

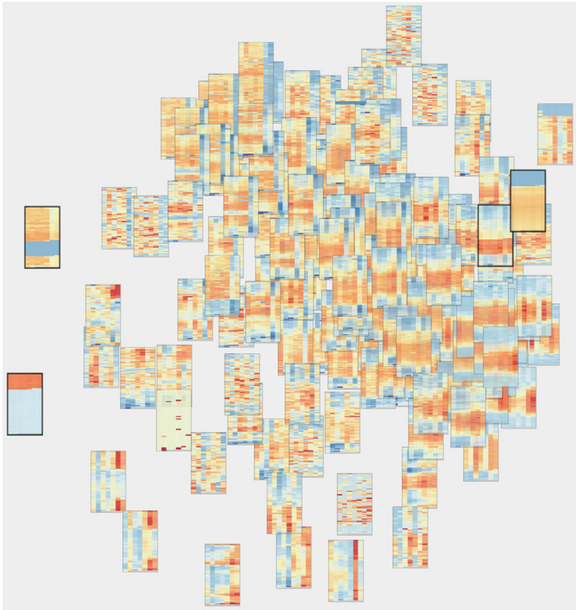

Figure 2: *2D space embedding (projection) of the time series reveals two extraordinary cases of temporal variation (marked by black frames on the left). The encoding of the time series is the same as in Figure 1*

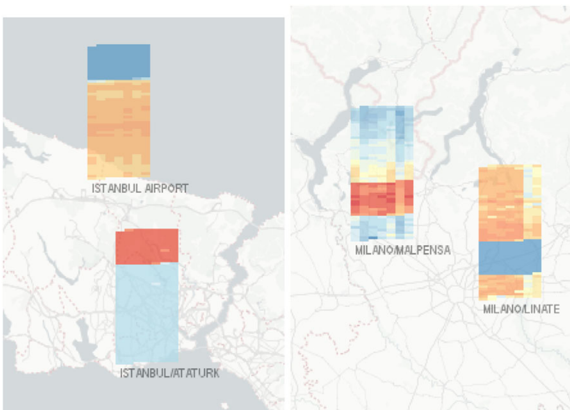

Figure 3: *Special cases: airports with peculiar temporal patterns of changes and their geographical neighbours.*

further, we locate these airports on the map and examine their neighbourhoods, as shown in Figure 3. We notice that the glyphs of the Istanbul Airport and Milano Malpensa demonstrate opposite colour patterns with respect to the glyphs of Istanbul Ataturk and Milano Linate, respectively. The most likely explanation is that Istanbul Ataturk and Milano Linate were not used for long time periods, and their neighbouring airports were used instead. The glyphs of these neighbouring airports are located on the right of the projection plot in Figure 2; they are also marked by black frames. These two glyphs not separated from others because their respective temporal patterns are not extraordinary. Many airports experience an increase in the number of flights during the summer and early autumn months, such as Milano Malpensa, and low traffic in the winter, as is the case with Istanbul Airport.

To ensure that our analysis captures the general features of the European air traffic over the year, we exclude the airports that exhibit unusual temporal patterns as being irrelevant to the overall manner of air traffic changes throughout the year.

4.2. Step 2. Divide data into subsets with coherent temporal patterns

The air traffic data exhibits diverse patterns of temporal variation with varying interplay between seasonal and weekly time cycles, as revealed by both the map (Figure 1) and projection plot (Figure 2). While some airports experience an increase in traffic during summer, others experience a decrease, and some have no seasonal variation. Similarly, the weekend traffic decreases in some airports and increases in others, while there are some airports with no weekly variation. Furthermore, some airports have differing amounts of flights on Saturdays and Sundays. To obtain an overall time periodization that adequately captures the differences between seasons and days of the week across all or almost all airports, we group similar time series together using partition-based clustering, such as the popular k-means algorithm [HW79].

Partition-based clustering algorithms aim to divide data into a specified number of groups, called clusters. In our case, we are interested in grouping similar time series with distinct patterns of temporal variation. Since neither the map (Figure 1) nor the projection (Figure 2) reveal clear-cut groups of similar time series suggesting the "right" number of clusters k, we apply the strategy of multiple trials. This involves increasing k until the clustering result does not reveal substantially different patterns [AAF*20]. To some extent, this strategy is similar to what is suggested in the elbow method [Tho53]: increase k until only small decreases of a numeric measure of the approximation error can be gained in the following steps. The reliability of the elbow method is questioned (e.g. in [Sch22]) as the error decreases with increasing k in nearly the same way irrespective of the dataset properties. Instead of relying on a statistical measure, we emphasize the importance of clustering results to be meaningful for analysts. Therefore, the criterion is novelty and importance of patterns being uncovered when k increases.

To apply this criterion, we need an expressive visualization of aggregated temporal patterns of members of different groups. We use the technique demonstrated in Figure 4. For each group, the display includes a matrix with columns corresponding to the 7 days of the week (from 1 for Monday to 7 for Sunday) and rows to the 53 weeks of the year. The squares in the cells represent aggregated values, such as the group averages for the corresponding days. Each group has been given a distinct colour, which is used for painting the squares representing positive numbers, whereas white squares stand for negative numbers. The sizes of the squares are proportional to the absolute values. Since the individual time series consist of z-scores, large coloured squares correspond to increased amounts of traffic compared to the airport-specific means, large white squares to reduced amounts of traffic, and small squares to amounts around the means.

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

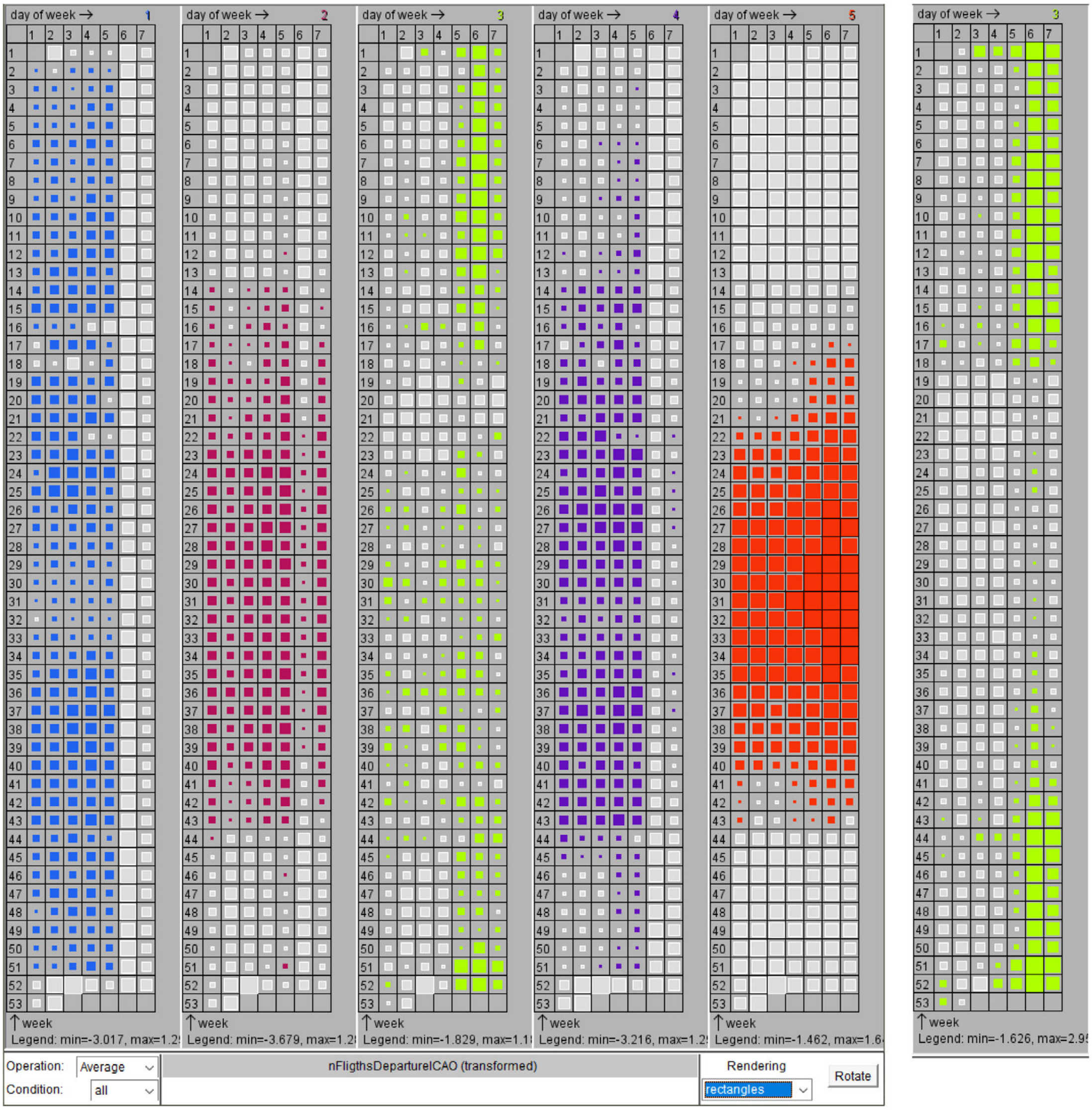

Figure 4: *Time aggregate matrices with rows corresponding to weeks and columns to days of the week show temporal variation of the aggregated flight counts in five clusters of airports. On the right is the matrix of cluster 3 after cleaning from time series with irregular patterns. Coloured and white rectangles represent, respectively, high or low traffic compared to the mean.*

After conducting multiple trials, we found that the result with five clusters (shown in Figure 4) was the most suitable. All temporal patterns are distinct, and making more clusters does not reveal substantially differing patterns. We used the colours of the clusters to paint the dots representing the airports on the map in Figure 5. Although the clusters are not well separated geographically, there are regions where members of particular clusters are more prevalent. For example, cluster 1 (blue) is more prevalent in the north of Europe, cluster 5 (red) in the south, and clusters 2 (burgundy) and 4 (purple) in the centre and on the British Isles. Cluster 3 (light green) is more dispersed.

Figure 6 shows a 2D projection of the time series involved in clustering. The colours of the glyph boundaries indicate the cluster affiliations. The glyphs for cluster 3 are highly dispersed in the projection space and even appear on opposite sides of it, indicating that the time series are quite dissimilar. This explains why the aggregate temporal pattern of cluster 3 in Figure 4 is not as pronounced as the others. By visual comparison, we find out that only three glyphs located at the right edge of the plot have similar patterns of temporal variation, exhibiting high traffic amounts during the weekends of the cold seasons. These patterns correspond to Marrakech, Tenerife and Gran Canaria airports. The remaining time series are dissimilar to each other and members of other clusters. It is reasonable to exclude these exceptional time series from the process of defining general time periods. After removing these time series, the time aggregate matrix of cluster 3 changes, as shown on the right side of Figure 4.

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

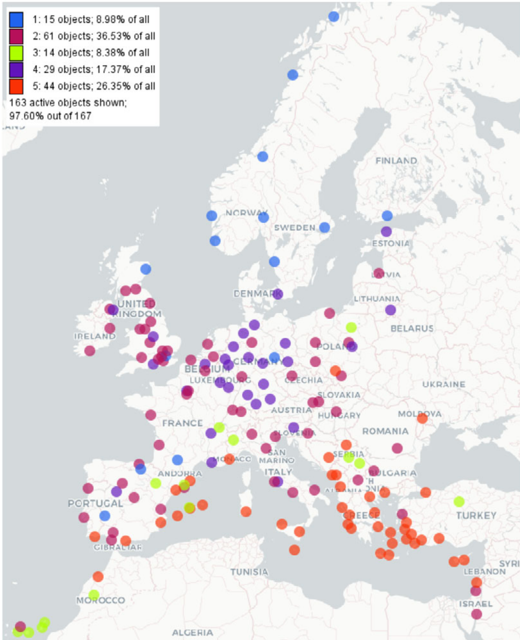

Figure 5: *The cluster membership of the airports is represented by circle colours.*

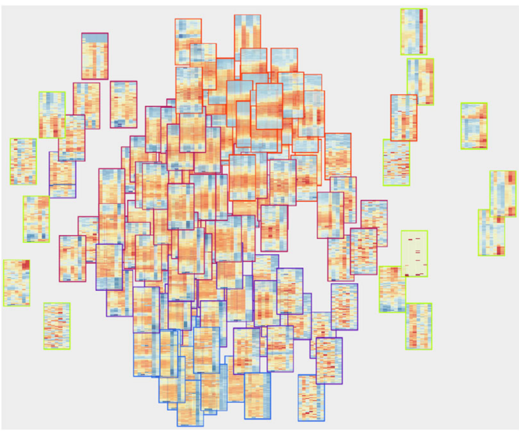

Figure 6: *In a 2D projection, cluster affiliations of the airports are represented by colours of the glyph boundaries. The members of cluster 3 (light green) are highly dispersed in the projection space due to their low similarity.*

4.3. Step 3. Define time periods for the data subsets

Figure 4 suggests that each group of airports may require its specific division of the time into periods of low, average, and high traffic intensity. A general tool to fulfil partitioning tasks is partition-based clustering, for example using k-means. For time division, clustering needs to be applied to data characterising time steps, that is to the distributions of the flights over the airports in our case. We describe each time step by feature vectors consisting of the airport-associated values and apply clustering to these feature vectors separately for each group of airports. To choose a suitable number of clusters and check whether clustering results adequately capture the general patterns of temporal variation, we use 2D projections of the time steps obtained by applying dimensionality reduction (Sammon's mapping) to the same feature vectors. To see the patterns, we encode the positions in the projection space by colours using a continuous two-dimensional colour scale as shown in Figure 7. These colours are then used in a display called time arranger, which is presented in Figure 8. In a time arranger, time steps are represented by coloured blocks arranged in a matrix layout, where columns correspond to days of the week and rows to weeks of the year. In the upper part of Figure 8, the colours of the blocks are taken from the 2D projections of the time steps shown in Figure 7. It should be noted that the colour assignments for the time steps based on projection are not consistent across airport groups. This is because the time steps for each group are described by a unique combination of variables (corresponding to the airports of this group). There is no way to project them onto a common space, making it impossible to achieve colour consistency.

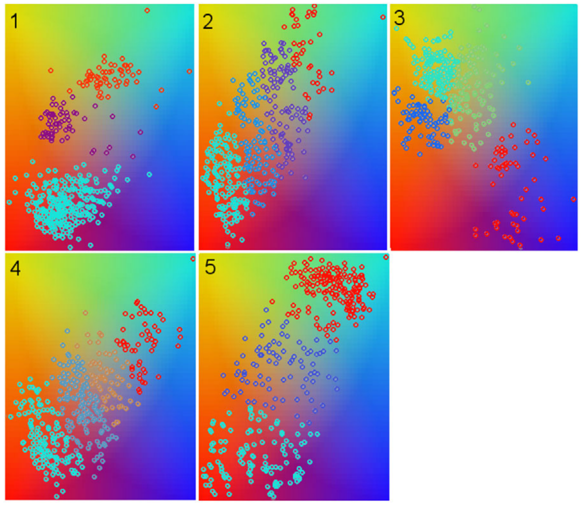

Figure 7: *2D projections of the time steps based on the similarities of the flights distributions over the airports of each group. The dots representing the time steps are coloured according to their membership in k-means time clusters.*

The patterns of continuous colour variation observed in the time arrangers reveal the interplay of different patterns of seasonal and weekly variation of air traffic in the groups of airports. The presence of seasonal changes is particularly prominent in groups 2 and 5 and is not observed in group 1. However, group 1 shows high dissimilarity of Saturdays and Sundays from the weekdays and from each other. Saturdays are substantially different from the other days in groups 2 and 4, whereas Tuesdays and Fridays are dissimilar to the other weekdays in groups 3 and 5. As the observed patterns are complex, a pragmatic strategy is to define time periods that capture seasonal changes and characterize each period in terms of the specific weekly variation of that period. This may imply the need to model the traffic of different days of the week in each time period.

In the lower row of time arrangers in Figure 8, the blocks are coloured according to the results of partition-based clustering, with

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

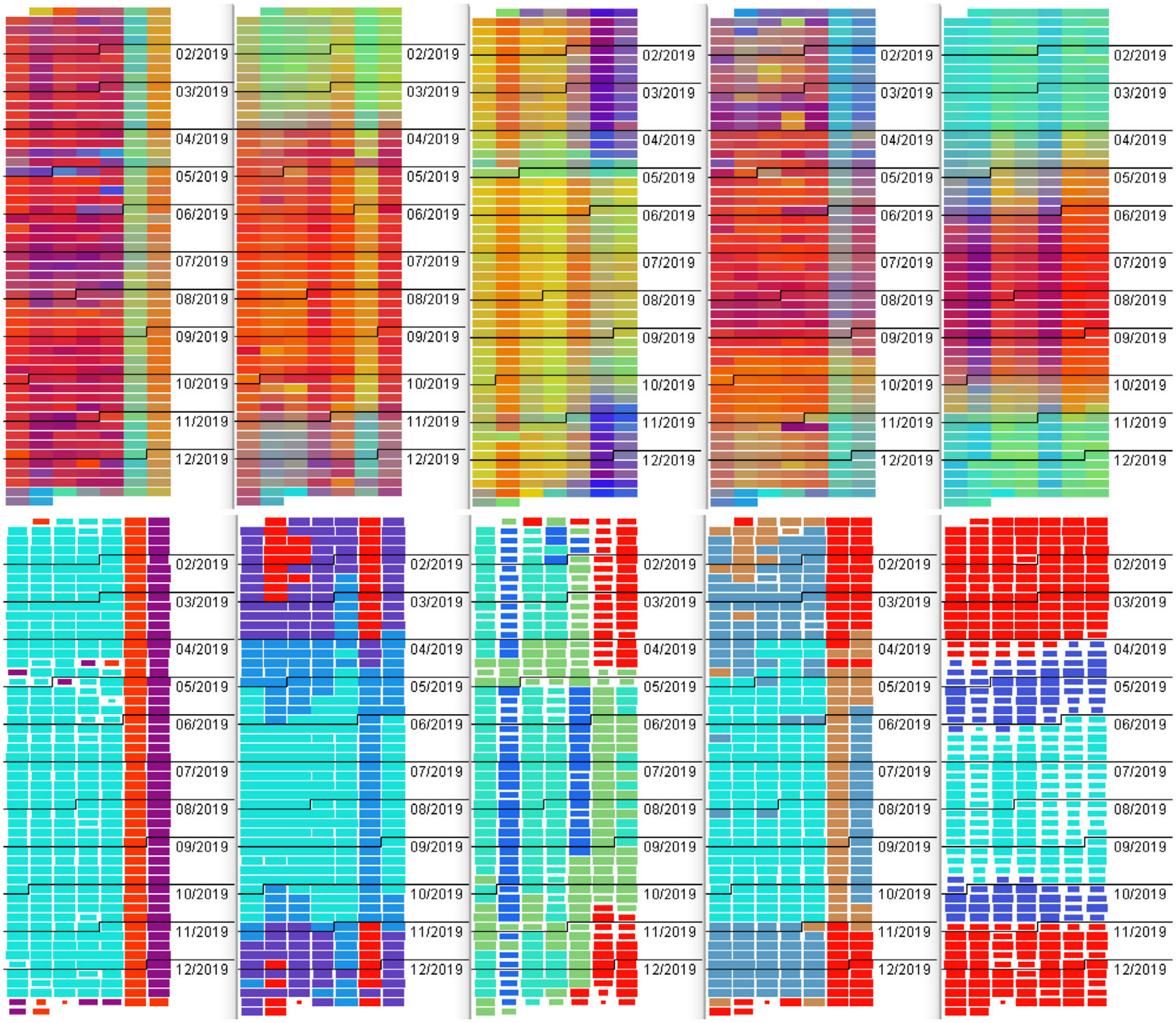

Figure 8: *The time steps are represented by coloured bricks arranged in rows corresponding to the weeks and columns to days of the week. In the upper row of time arrangers, the bricks are coloured according to the positions of the time steps in the projections (Figure 7). In the lower row, the brick colours represent the cluster affiliations of the time steps.*

each colour representing one of the clusters. The same colour-coding is used for the dots representing the time steps in the projection plots in Figure 7. Additionally, the size of the blocks in the lower row of time arrangers indicates the proximity of the time steps to the centres of their respective clusters. Larger blocks correspond to core cluster members, while smaller blocks indicate borderline members.

The purpose of partition-based clustering is to facilitate time periodization for analysts by generating draft partitions rather than requiring them to define time periods from scratch. However, these partitions should not be considered strict and precise, as peripheral members of a cluster may be more similar to members of other clusters than to their own core members. Therefore, analysts may need to modify cluster boundaries to adapt to their analysis goals. In suggesting draft divisions, it is desirable for the clustering results to be simple and easily interpretable. The clustering results shown in the lower part of Figure 8 adhere to this criterion, with 3 or 4 clusters for each group of airports that capture the most prominent features of the observed temporal variation in the corresponding time arranger in the upper row.

To achieve the goal of defining a universal time periodization, one might consider using algorithmic clustering to obtain a common time division for all airports. However, it is often difficult to guarantee that all significant patterns have been accurately captured in the clustering result. Figure 9 illustrates this point. We applied the k-means clustering algorithm to the daily value distributions of all five airport groups combined (after the exclusion of the outliers), with the number of clusters increasing from 6 (left) to 10 (right). To ensure colour consistency across multiple clustering runs, we used similarity-aware colour assignment to the clusters [AAFW17]. The visible pattern mainly reflects the variations of the largest groups 2

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

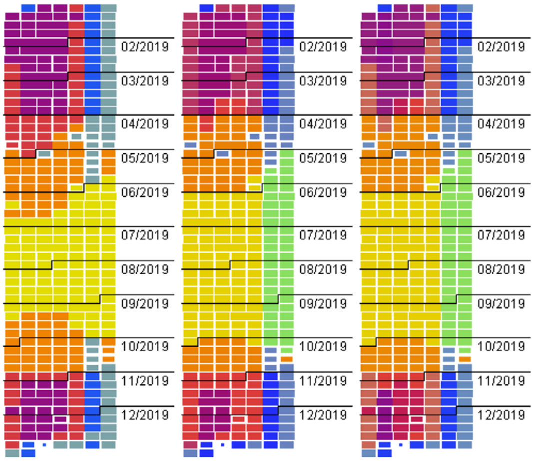

Figure 9: *Results of automatic clustering of the days by similarity of the daily value distributions over the airports from all groups. From left to right: 6, 8, and 10 clusters.*

and 5, while the smaller groups remain under-represented. This experiment demonstrates that automated clustering cannot replace a human analyst's thoughtful definition of unified time periods. Each subset of data must receive appropriate attention to ensure reliable time periodization.

4.4. Step 4. Define integrated time periodization

This step of the time periodization workflow is performed using the timeline display as demonstrated in Figure 10. In the display, the horizontal dimension represents time, and the horizontal bars represent sequences of time steps (in this case, days). The colours of the bar segments correspond to the spatial projections of the steps (Figure 7) in the upper part of Figure 10 and to the colours of the automatically generated time clusters (Figure 8, bottom) in the lower part of the figure.

The linear arrangement of the time step colours in the timeline display creates stripy patterns that highlight weekly variation but make it difficult to perceive longer time periods. To alleviate this issue, a time mask filter [AAC*17] can be applied to hide the weekends. Figure 12 demonstrates the effect of this filtering. The continuous bar in the original timeline display is transformed into a sequence of rectangles. Although continuity is broken, the spaces between the rectangles do not attract as much attention as the colourful stripes in Figure 10, allowing for better perceptual unification of neighbouring rectangles with similar colours, as prompted by the Gestalt laws [Met06].

Figure 11 displays the interactive tools used to define time periods. The main operations include importing time partitions (referred to as "time classes") from an existing division, creating a new time class, and defining an interval by dragging along the time axis. The interval gets the label of the currently active time class, indicated by a selected radio button. Any time steps previously labelled by other time classes are re-labelled.

To define time periods, we use the interactive tools as follows. First, we import the partitions from the time division for airport group 5 (Figure 10, lowest bar). The imported divisions appear in a dedicated row of the timeline display, highlighted in the middle of Figure 12. We then select the radio button of class 2 (cyan) and redefine the period using brushing to simplify the definition of the summer time period, which now stretches from June till the end of September. Next, we consider the time clusters and corresponding projection-based colour variations of the other airport groups. The

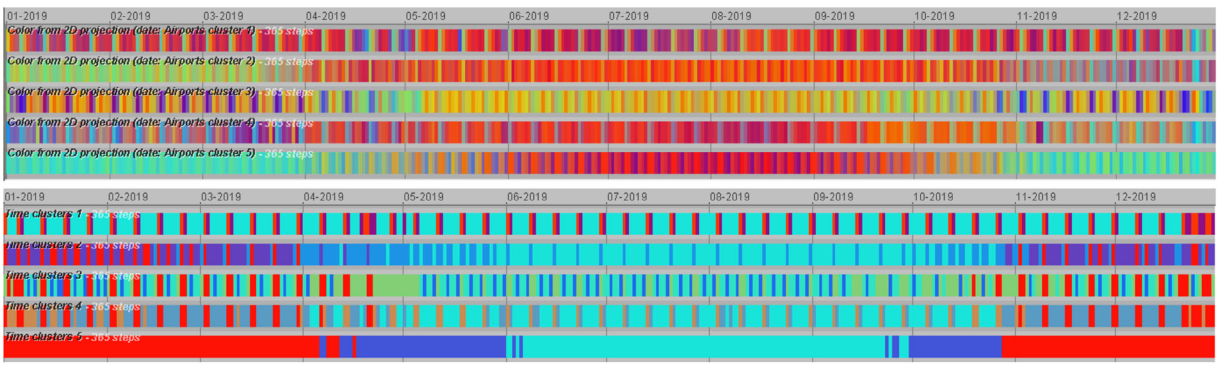

Figure 10: *In a timeline display, each horizontal bar represents a sequence of time steps represented by coloured segments. The colours in the upper part correspond to the positions of the time steps in the 2D projections shown in Figure 7. In the lower image, the colours correspond to the time clusters presented in Figure 8.*

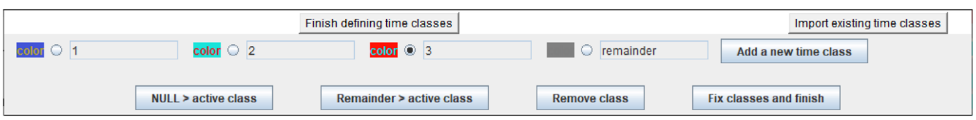

Figure 11: *Controls for interactive time periodization.*

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

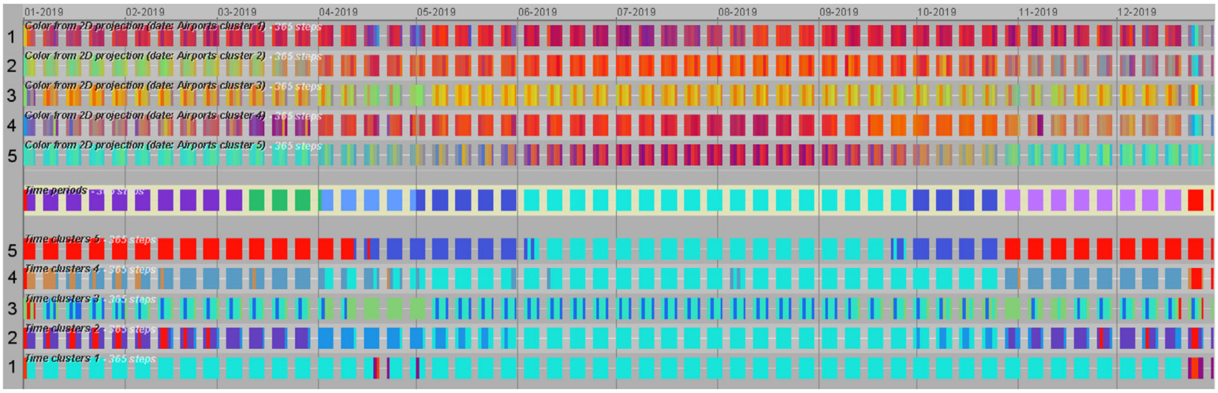

Figure 12: *In the highlighted row of the time line display, time periods have been interactively defined by considering the continuous colour variation patterns in the upper part of the display and the divisions of the time steps into clusters shown in the lower part. Saturdays and Sundays have been filtered out to facilitate focusing on long-term variation.*

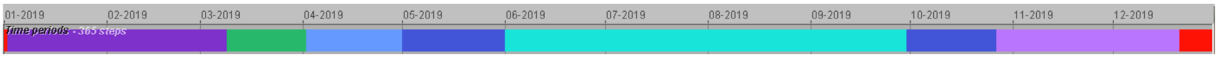

Figure 13: *The time periods that have been defined.*

divisions into time clusters are shown in the lower part of the timeline display and the colour variations in the upper part. The rows are arranged symmetrically with respect to the central row where the periods are defined.

Using the operations of adding a new time class and brushing along the time axis, we define time periods as shown in the middle of Figure 12. The red colour is assigned to New Year and Christmas, while purple and violet periods reflect variations in groups 2 and 4. The time interval from the second week of March to the end of April is divided into two periods (green and light blue) to reflect variations in groups 3 and 4. May and October are considered sufficiently similar to have the same time class (blue). After defining the periods, we cancel the time filter and make the periods continuous by brushing over them.

The final periods are shown in Figure 13 along the timeline and arranged by weeks and days of the week in Figure 14, left. Two of the seven time periods (coloured in red and blue) consist of two disjoint parts appearing in the first and second halves of the year. However, as mentioned earlier, there is weekly variation of the air traffic in each period, requiring separate modelling of traffic for days with substantial differences in flight distribution. To understand how each period should be modelled, we need to compare the flight distributions of different days of the week. A possible approach is described in the next section.

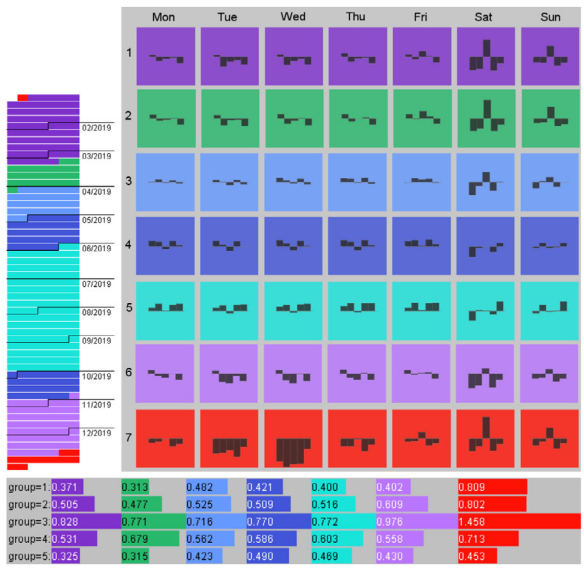

Figure 14: *On the left, the time periods are arranged weekly, similar to Figure 8. On the right, the time periods are shown in rows of a matrix. Each row corresponds to a statistical summary of flight distributions for different days of the week (matrix columns). The bar charts show the median relative flight amounts (i.e. normalized differences from the mean) for five groups of airports. At the bottom, standard deviations from the period means for the airport groups are represented by bar lengths.*

4.5. Step 5. Characterize the time periods

The purpose of dividing the time span of a dynamic phenomenon into periods is to represent its development as a series of states or trends that occurred at specific times and in a particular sequence. These states or trends should be easy to comprehend and describe. During the process of time periodization, we did not consider the characteristics of the phenomenon at different times, but instead focused on indications that suggested when one state or trend was replaced by another. Now it is time to describe the time periods in terms of the data. However, since the data are complex, they need

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

to be aggregated in order to facilitate understanding. To capture the differences between days of the week within each time period, data aggregates must be computed for all combinations of time periods and days of the week. Flight distributions for each combination can be aggregated by groups of airports. The resulting aggregates are displayed in a matrix in Figure 14. The rows correspond to the time periods and columns to days of the week. Each cell contains a bar chart with five bars corresponding to the groups of airports, displaying the deviations of flight amounts from the average. The bar charts at the bottom of Figure 14 represent averaged standard deviations from the period's means.

The matrix in Figure 14 serves a dual purpose. Firstly, it provides insight into the general characteristics of flight distributions across the time periods, allowing for comparisons to be made. For example, periods 4 and 5 exhibit increased flight amounts in all airport groups except for group 3, while period 3 is near average, and periods 1 and 6 show relatively low air traffic. Additionally, there are visible differences between weekends and weekdays for each period. Notably, group 3 experiences high amounts of flights in periods 1-3 and 6-7, while traffic in the other groups significantly decreases.

Secondly, the matrix shows how to break down the modelling task for each period. For all periods, separate models are required for Saturdays and Sundays. However, for periods 1, 2, and 7, two models can be used to represent Saturdays and Sundays together. The Christmas and New Year period (7) requires separate models for each weekday. For periods 3–5, one model can represent all weekdays. Periods 1 and 2 require specific models for Monday + Thursday, Tuesday + Wednesday, and Friday, while period 6 requires weekdays to be divided between Monday + Friday and Tuesday to Thursday.

The specific aim of this analysis was to assist domain experts and modelling specialists in representing air traffic throughout the year using a sufficient number of simulation models. The result was highly valued by our project partners, who are domain experts in the field. They found the time periods and corresponding distributions to be very clear and well-aligned with their domain knowledge, indicating that our approach was effective in achieving its goal.

Taking a more general perspective, we have developed a systematic approach for analysing and describing the development of dynamic phenomena. The key idea is to decompose the development process into manageable components by identifying relatively stable states or trends. To achieve this, we divided time into periods where the state or trend of the phenomenon remained relatively constant. We devised a combination of techniques to facilitate time periodization and defined a sensible sequence of steps to arrive at our desired outcome. This analytical workflow enabled us to better understand the phenomenon and effectively model its behaviour over time. With the aim of exploring the potential generalization of our approach, we shall now apply it to a different dataset.

5. Testing and Elaborating the Approach

In order to validate and refine our approach, we will apply it to a new dataset that differs from the air traffic example in several ways. First, this dataset is multivariate, containing six time-dependent attributes. Second, although the weekly cycle strongly affects the temporal variation of the data, we will disregard the cyclic patterns in our analysis because we focus on the long-term development of the phenomenon. Third, the phenomenon under investigation is expected to involve prolonged trends, in which the attributes change continuously in a steady manner, in addition to the relatively stable states as found in the air traffic example. Unlike the air traffic case, we do not have specific users or applications in mind for the analysis results of this study. Our primary aim from the research perspective is to assess the generalization potential of our approach for other applications. The goal of the presented analysis scenario will be to understand the evolution of the society's reaction to a pandemic. We shall describe the analysis in less detail than in the previous example, giving primary attention to the adaptations of the workflow and methods to the differing properties of the second usage scenario.

5.1. COVID-19: Google Mobility Trends

We shall analyse a subset of the publicly available data set of Google Mobility Trends [Goo22]. The data is gathered from anonymized information provided by apps such as Google Maps, which records how people's movements have changed during the pandemic. The data consist of daily visitor numbers to specific categories of places (e.g. grocery stores, parks, train stations, etc.) relative to baseline days before the pandemic outbreak. Baseline days represent a normal value for each day of the week and are given as the median value over the five-week period from 3 January to 6 February 2020. The data thus consist of the deviations from the normal values expressed in percent of the normal values. Positive values mean increased numbers of visits to a certain category of places and negative values have the opposite meaning.

The subset we use for testing our approach contains data for 60 countries of Europe, Asia, and North America. The data cover the time period of the length of 501 days from 15 February 2020 (Saturday) to 30 June 2021 (Wednesday). The time graphs in Figure 15 show the time series of the values of the six attributes included in the Google dataset. All attribute names in the original data end with a constant expression 'percent change from baseline'. We shall thus use only the first parts of the attribute names, which are 'retail and recreation', 'grocery and pharmacy', 'parks', 'transit stations', 'workplaces', and 'residential'. Each time graph in Figure 15 corresponds to one of the attributes. The horizontal axis represents time. The temporal variation of the values for each country is represented by a line in grey. The thick black line shows the variation of the daily median from the whole set of 60 countries.

The prominent "saw teeth" pattern visible in all time graphs emerges due to the usual weekly cycle of human activities. However, these obvious weekly fluctuations are not of interest for the analysis. The time needs to be divided into periods capturing the major patterns and trends in the development of the phenomenon (i.e. human mobility behaviour) irrespective of the weekly variation.

5.2. Disregarding weekly fluctuations

As in the first usage scenario, it is valid to expect that the character of temporal variation is not the same in all components of the phenomenon, that is countries. Hence, we shall try to divide the set

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

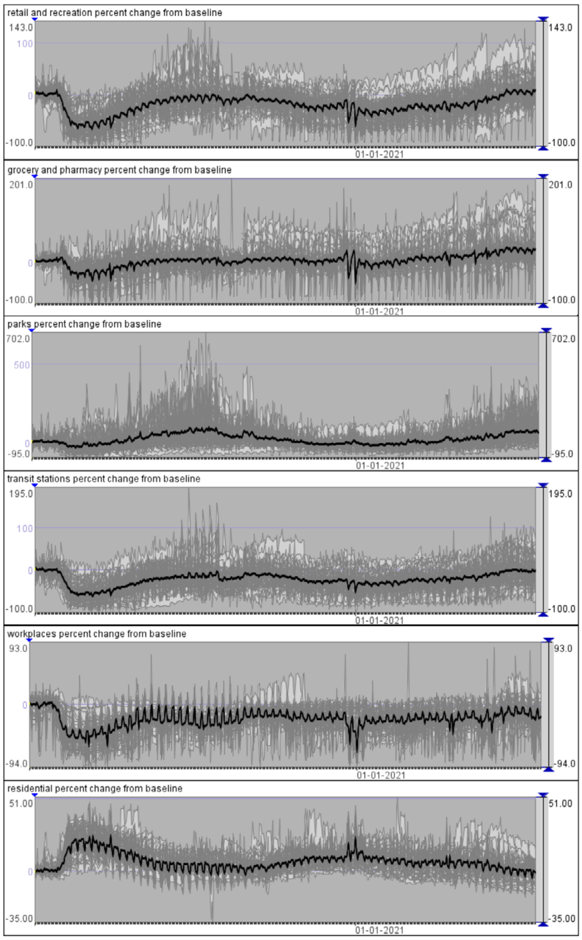

Figure 15: *Mobility trends: daily time series.*

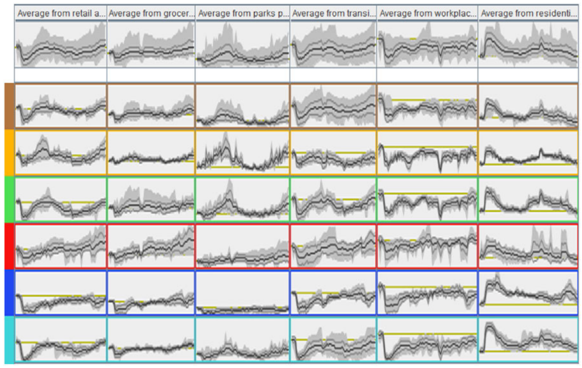

Figure 16: *Time series of the weekly averages for the whole set of countries (top) and for 6 groups of countries.*

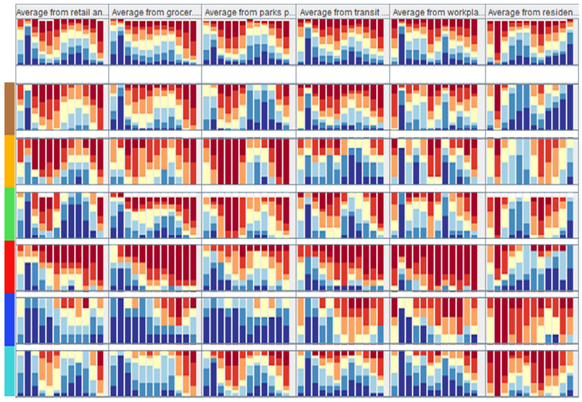

Figure 17: *The time series of the weekly averages represented in an aggregated form by segmented bars.*

Figure 18: *The colours of the circles represent the affiliations of the countries to the groups. The circles are positioned in the locations of the country capitals.*

of countries into groups with similar mobility behaviours over time. As before, we shall use partition-based clustering. However, to disregard the weekly fluctuations and, simultaneously, reduce the dimensionality of the data, we apply clustering to the time series of weekly averages computed from the original time series of daily values. The clustering is applied to the time series of all attributes taken together, that is to 60 feature vectors consisting of $72 \times 6 = 432$ values (72 weeks and 6 attributes). After multiple trials, we obtain a reasonable division of the countries into 6 groups.

The weekly time series for the groups are shown in the time graphs in Figure 16. The time graphs are arranged in rows by the groups (the topmost row corresponds to the whole set of countries) and in columns by the attributes. The multi-colour bar on the left shows the colours assigned to the groups of countries. The same weekly time series are represented in an aggregated form by segmented graphs in Figure 17. Each bar covers a time interval of 6 weeks. The coloured segments represent the proportions of the countries with attribute values belonging to different ranges. The segments are painted using a diverging colour scale from ColorBrewer [HB03], with shades of blue corresponding to lower value ranges and shades of red to higher value ranges. The differences between the groups of countries are clearly seen in both figures.

The affiliations of the countries to the groups are represented by the colours of the circles on the map in Figure 18. The colours are the same as in Figure 16 and Figure 17.

For defining the groups of countries, it was sufficient to use the low resolution data, that is the weekly averages. However, for defining time periods, it is preferable to use the original daily resolution, because important changes in the development trends may be lost due to the aggregation by the weeks. Nevertheless, we need to disregard the obvious differences between the human mobility behaviours on the weekdays and on the weekends. For this purpose, we apply time filtering, which excludes the data for Saturdays and

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

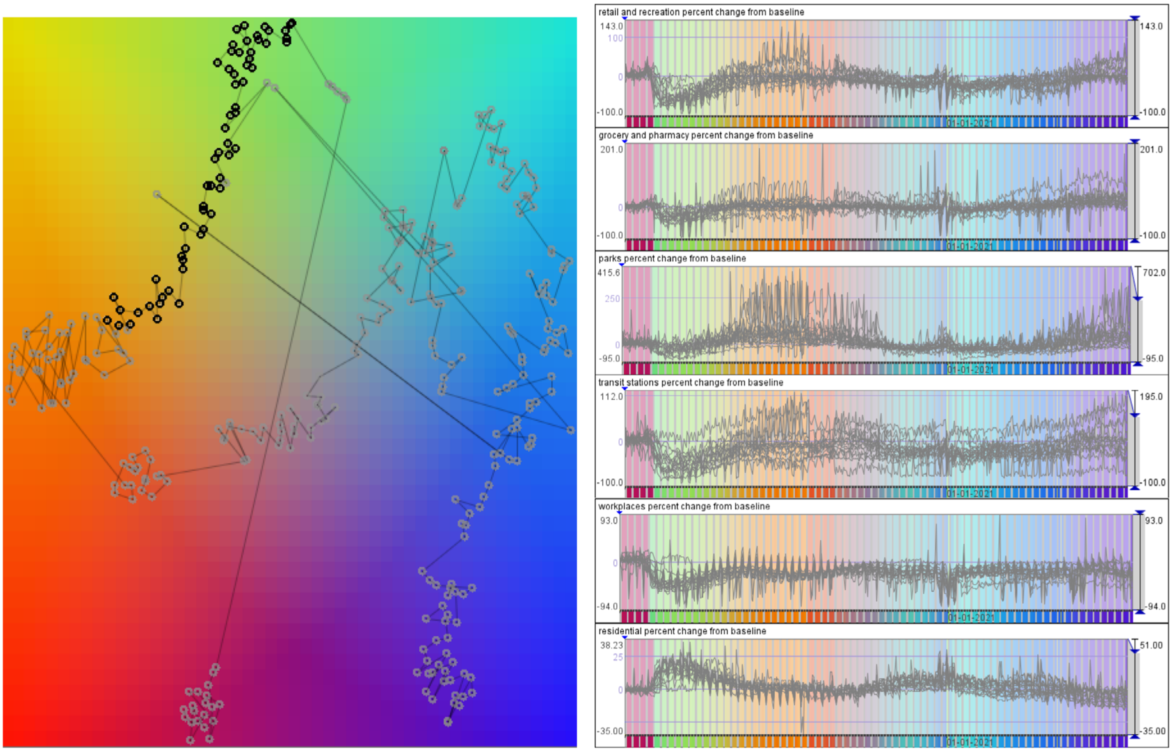

Figure 19: *Left: A 2D projection of the combinations of attribute values corresponding to the country group 1. Points representing consecutive time steps (excluding weekends) are connected by lines. Right: Colour variation from the 2D projection is reflected in the backgrounds of the time graphs showing the variation of the attribute values for the countries of the group 1.*

Sundays from the analysis. The following steps of the analysis will be applied to the filtered data.

5.3. Dealing with multiple attributes

In contrast to our first scenario, our analysis here involves studying patterns that encompass changes in several attributes. However, it is not feasible to consider the variation of every attribute in the same way as we did previously. We need a representation that integrates values of all six attributes. Therefore, we rely on low-dimensional (2D) projections of the data, similarly to the first scenario (Figure 7). In this case, however, we apply the dimensionality reduction algorithm to the combinations of values of all six attributes. We do this for each group of countries and thus obtain six projections, which are coloured as in the first scenario. For example, a projection for group 1 is shown on the left of Figure 19. To interpret and validate the projection results, we transfer the colours from the projections to time graphs showing the attribute values, as demonstrated on the right of Figure 19 for the country group 1. As can be seen, the colour variations effectively capture both the sudden shift in mobility behaviours that occurred in mid-March 2020 and the gradual changes that took place over time.

Seeing that the character of the colour variation is representative of the joint development of the multiple attributes, we can rely on it in performing the time division. For this purpose, colours from the projections are transferred to a timeline display, which is shown at the top of Figure 20. As noted before, the data have been filtered to exclude the values for Saturdays and Sundays, which have not been included in the projections. Therefore, the colouring of the horizontal bars in the timeline display is interrupted by grey segments that correspond to the weekends. However, the regularly positioned breaks in the colouring can be easily ignored without hindering the perception of the colour variation along a bar induced by Gestalt laws, such as figure-ground, similarity, proximity, and continuation [Met06].

5.4. Dealing with development trends

In the previous example, the phenomenon being studied could be broken down into distinct states that corresponded to different seasons and days of the week, with no gradual transitions between them. In the current example, when weekends are ignored, there are also periods of relative stability that can be treated as stable states. These states appear in a 2D projection as compact groups of close points [BSH*16, vdEHBvW16]. However, transitions between states are generally smooth and prolonged, with small changes occurring over time. Such sequences of small changes, called *trends*, appear in a projection as elongated arrangements of points resembling paths. One of such "paths" is highlighted in Figure 19, left. It is important to identify not only stable states but also periods of different trends in the joint development of the attributes. Partition-based clustering may not be effective in

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

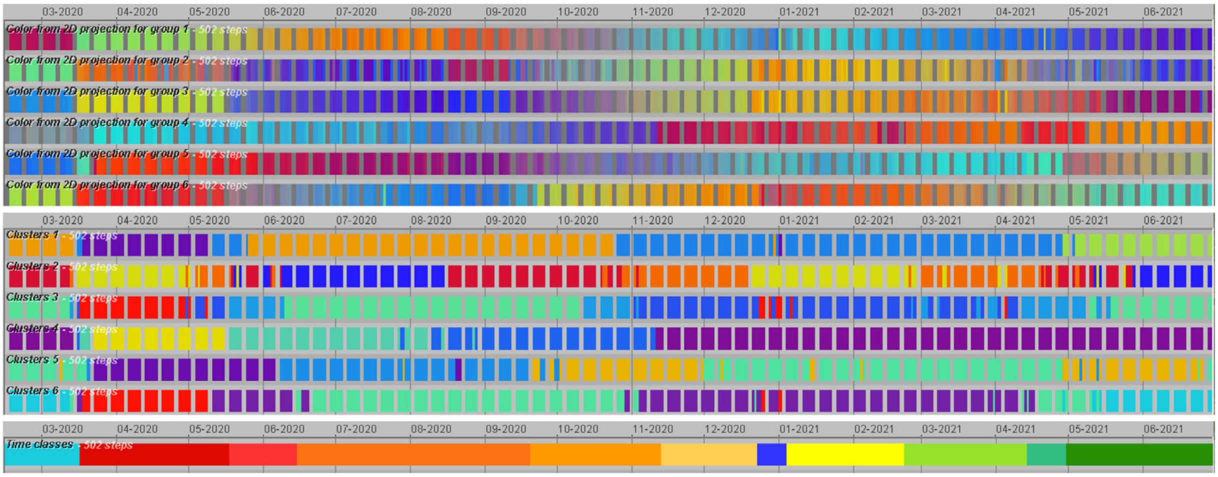

Figure 20: *Top: Colour variation from 2D projections of the combinations of attribute values corresponding to 6 groups of countries. Middle: Results of time steps clustering for the groups of countries. Bottom: Unified time periods defined interactively.*

identifying these periods, as it only considers pairwise similarities and cannot account for the sequential arrangement of data items. This can result in arbitrary breaks in the sequence, such as dividing a "path" into two or more "round" "clusters".

To analyse trends in the development of multiple groups, we need to combine information from different projections. This is done by transferring the colours of the positions in the projections to a timeline display, as shown in the upper part of Figure 20. In this display, a trend in a projection is represented as a bar segment with coherent colour changes along the time axis. For example, the "path" highlighted in Figure 19, left, is translated to a segment in the upper bar of the display in Figure 20, top, where shades of green gradually change to more yellowish tones and then to orange. By observing multiple coloured bars in the timeline display, we can detect abrupt changes in the colour shades, which indicate changes in the state or trend of a group. The time step when the change occurred is a candidate for becoming a boundary between two time periods, but we also look at the other bars to see if they contain abrupt colour changes around the same time step. If they do, we set a common break that takes the average or earliest time step of change from several groups. By repeating this procedure, we can define the time periods that reflect the stable states and trends in the development of multiple groups.

While clustering may not always successfully distinguish between stable states and prolonged trends, it can still be a helpful tool for time periodization. It can generate time clusters that serve as a starting point for time division, instead of creating time periods from scratch in an entirely interactive manner. In the middle of Figure 20, the timeline display shows the results of partition-based clustering of the time steps. By comparing these cluster-based bars with the projection-based bars, we can identify which clusters adequately encompass some of the stable states or trends visible in the projection-based bars. These clusters can then be imported and used as an initial time division. In our example, we began by using the time clusters created for group 5 and added a few clusters generated for group 6. We then continued to modify the initial division interactively until we arrived at the final time periods shown at the bottom of Figure 20.

In the first scenario, we characterized the time periods by averaging the flight distributions (Figure 14). This approach was reasonable because the development of air traffic mainly consists of stable states that can be adequately represented by average distributions. However, averaging is not suitable for representing development trends. To understand and describe what was happening in the different time periods, we use time graphs with the periods represented by background colours, as shown in Figure 21. To simplify the task of characterizing trends, we represent the variation of the mobility behaviours in each group of countries by time series of the medians of the value distributions in consecutive time steps. To smooth out weekly fluctuations, we applied temporal smoothing to the original time series before aggregating them by the groups. The spaces between the lines representing the group time series are painted in gray. This allows us to see clearly when groups behaved similarly and when their behaviours significantly diverged. To investigate developments in any particular group, we interactively select the group, and the corresponding lines in the time graphs become marked in black. For example, in Figure 21, the lines for group 1 are marked. This approach enables us to describe time periods in terms of common and differing states or trends of groups of countries.

5.5. Conclusion from both scenarios

We have considered two distinct scenarios in order to investigate the time periodization problem comprehensively, reveal possible variants related to features of the data and focus of the analysis, and determine how these variations affect the analysis workflow and the use of analytical techniques. The results of our studies are summarized in Table 1.

6. Discussion

In this paper, we have introduced time periodization as a general type of analysis problem and presented two examples of solving it

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

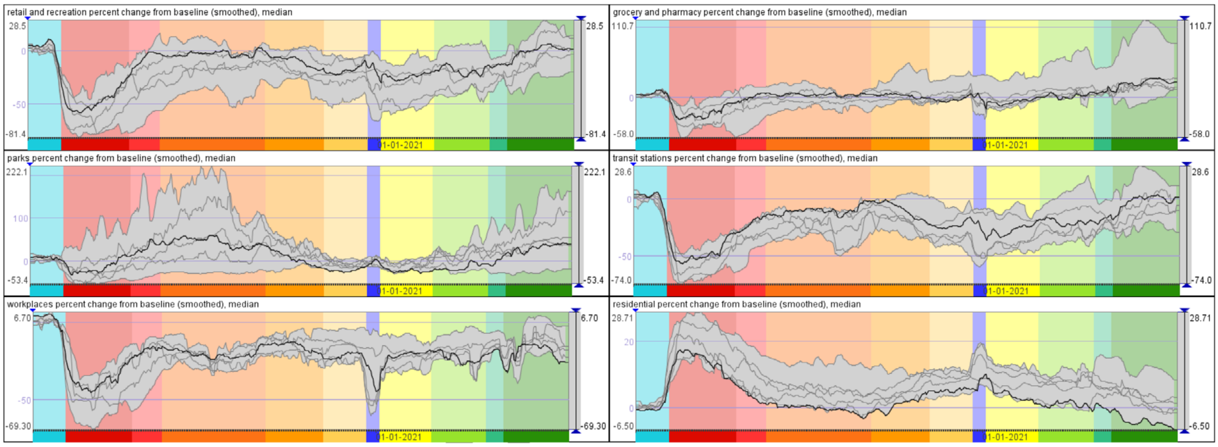

Figure 21: *Time periods are represented by background colours in the time graphs showing temporal patterns of the mobility development in the six groups of countries. Each group is represented in each time graph by a time series of the medians of the value distributions in consecutive time steps. Smoothing by a sliding window of the length of 7 days have been applied to the original time series before the aggregation to hide the weekly fluctuations.*

Table 1: *Variations of data features, analysis focuses, and the corresponding analysis operations.*

Data or task features	Challenges & opportunities	Technique, way of application
Data: single time-variant attribute, periodic variation	Opportunity: investigate and compare periodic patterns of value variation using compact representations	time matrix, applied to groups of components (Figure 4)
Data: multiple time-variant attributes	Challenge: patterns of joint variation of attribute values are hard to visualise, perceive, and compare	Multiple time graphs and/or segmented bar charts arranged by the attributes and groups of components (Figs. 16,17)
Data: multiple time-variant attributes, periodic variation	Challenge: patterns of periodic variation of all attributes are hard to investigate visually	Time filtering: select parts of the time cycles for visualisation and investigation (Figs. 19,20)
Task feature: periodic variation is irrelevant	Challenge: difficult to ignore periodic fluctuations and focus on overall trends Opportunity: data can be simplified by filtering or smoothing	Time filtering: exclude irrelevant parts of the time cycles from the analysis (Figs. 19, 20) Temporal smoothing: by a sliding window of the length equal to the time cycle length (Figure 21)
Data development feature: relatively stable states and short transitions; subtask: define periods	Opportunity: different states are easy to distinguish visually and computationally (by clustering)	Projection and clustering applied to the value distributions in the time steps Projection: shows the approximate number of states (Figure 7) Partition-based clustering: separates the states (Figure 8)
Data development feature: relatively stable states and short transitions; subtask: characterize periods	Opportunity: stable states can be summarised and visualised in an aggregated form	Summarisation of value distributions for time periods Visualisation of the summarised value distributions (Figure 14)
Data development feature: includes trends and prolonged smooth transitions between states	Challenge: different trends involving multiple attributes are difficult to detect visually Challenge: partition-based clustering cannot adequately separate trends	Projection applied to the value distributions in the time steps Projection plot: shows states as point clusters and trends as paths (Figure 19) Projection colours transferred to timeline display: states and trends are recognised from colour similarities and variation along the time axis (Figure 20)

for data with different characteristics of temporal variation. With the aim to provide a flexible and adaptable framework for time periodization, which can accommodate the diverse properties of data and analysis objectives, we have devised a generic workflow that comprises abstract steps or operations rather than relying on specific methods and software tools.

6.1. Workflow and methods

The steps of the workflow are: (1) detect diverse patterns of temporal development among components of the studied phenomenon and define groups of components with similar patterns; (2) represent the temporal development of each group by colour variation; (3) jointly

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

consider the colour variations of the groups, identify stable states and trends in the group developments, and interactively define time periods enclosing states or trends of multiple groups.

The first step can be performed with the help of partition-based clustering applied to the time series of attribute values associated with different components of the phenomenon. Clustering is a commonly used technique for analysing multidimensional data and multiple time series. Visualizations of cluster summaries are utilized to assess the quality and interpretability of the clustering results and find appropriate parameter settings. Different visualizations can be suggested depending on the number of attributes and the presence or relevance of cyclic temporal variations; see Table 1.

The second step relies on projecting time steps onto a 2D space based on the similarities of the corresponding value distributions. Here we build on the research performed by Bach et al. [BSH*16] and van den Elzen et al. [vdEHBvW16], who used projections for analysis of complex dynamic phenomena, classified possible spatial patterns that can emerge in such projections, and explained their meanings. However, our analysis problem requires joint consideration of patterns from multiple projections, which cannot be achieved by using several projection plots alone. To address this issue, we propose a visualization that combines these patterns in a single view by encoding the positions in the plots with colours, which are used to represent the positions in a timeline display. This approach is similar to the use of scarf plots, where colours represent eye-tracking positions over a computer screen [AACF22].

As is well known, representing high-dimensional data in low-dimensional embeddings is prone to errors, which can be of two types: missed similarity, where similar items are positioned too far apart, and false similarity, where dissimilar items are positioned too close together [KP11]. Achieving an appropriate trade-off between these two types of errors depends on the analysis goals and data properties. Temporally-dependent data, in particular, often exhibit similarity between neighbouring time steps, so preserving local neighbourhoods can be important for capturing their temporal structure. Neighbour embedding (NE) algorithms [Yan05, YPK13], which aim to place nearest neighbours close together in the embedding space at the expense of higher distortions of mid- and long-range distances, can be well-suited for this purpose. Examples of NE methods include t-SNE [vdMH08], which we found helpful in the second usage scenario that involved prolonged trends (after excluding irrelevant weekends). However, when such trends are absent, other projection methods, such as MDS [Kru64] or Sammon's mapping [Sam69], may be more appropriate. It is worth noting that the impact of different embedding methods and their parameters on the results requires further investigation.

To represent positions in an embedding, we use a 2D colour space called Cube Diagonal Cut B-C-Y-R (Blue-Cyan-Yellow-Red), which was rated highly in a task-based evaluation study of 22 colour maps [BSM*15]. It provides about 585 noticeably different colours, which is a relatively high number, and is therefore well suited for localization and identification tasks, which are critical for matching colours in the timeline display to positions in the embedding space. However, this colour map is not optimal for accurately representing distances between points in the 2D space. Colour maps like CIELUV or CIELAB [Sch07] are designed for perceptual uniformity, but have a more limited range of distinguishable colours, ranging from 193 to 423 colours in different versions [BSM*15]. In the proposed approach to time periodization, precise estimation of projection distances is not necessary. A broader range of distinguishable colours can help detect periods of stability, gradual changes, and abrupt changes more effectively.

Partition-based clustering of time steps based on similarity of data distributions can be used to obtain an initial variant of time division subject to interactive revision. This is more convenient and efficient than to define time periods from scratch. Clustering may be especially suitable when the development of the phenomenon under study consists of several relatively stable states and short transitions between them, as we had in the first usage scenario. According to [BSH*16] and [vdEHBvW16], stable states appear in a projection as dense concentrations of points (see Figure 7), which can suggest suitable parameter settings for the partition-based clustering. In the presence of development trends or long smooth transitions between states (Figure 19, left), partition-based clustering may be of limited utility.

There are numerous options available for designing and implementing interactive techniques that facilitate the creation and editing of time periods. While exploring the entire design space is beyond the scope of our research, we would like to emphasize the importance of enabling time filtering. Time filtering is valuable for handling periodic changes and can help analysts ignore specific times (e.g. holidays) or, conversely, concentrate on specific times.

6.2. Generalizability and application scope

Time periodization can serve two main purposes: description and prediction [ALA*18]. Description involves creating a simplified and understandable representation of a behaviour that can be used for comparison or communication to various stakeholders. On the other hand, predictive modelling is useful when a behaviour is expected to repeat, especially periodically. For example, in our first usage scenario involving air traffic, the behaviour is known to occur repeatedly each year; so, time periodization was used to decompose the task of predictive modelling into subtasks focusing on periods of relative stability of the air traffic distribution properties. Our second usage scenario analysing changes in mobility behaviour in response to a pandemic was primarily focused on description rather than prediction.

Here are a few additional examples of tasks where time periodization is relevant and our approach could be useful. It could be used to identify and describe patterns in car parking occupancy over time, such as peak hours or days with lower demand. By dividing the data into meaningful time periods, analysts can identify trends and fluctuations in occupancy that can be used to optimize the use of available parking spaces through the development of a pricing policy. For example, by identifying the times when demand is highest, parking facilities can increase prices to maximize revenue.

Our approach can help companies to identify patterns in the shopping behaviours of different customer groups over time, such as

© 2023 The Authors. Computer Graphics Forum published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

changes in preferences or spending habits. By defining relevant time periods, analysts can gain insights into the factors that drive customer behaviour and develop more targeted marketing strategies for different customer segments. This can help companies to increase engagement and sales by better understanding and meeting their customers' needs.

Time periodization can also be useful in studying and describing the playing behaviour of a football team over the course of a game. By dividing the game into meaningful time periods, journalists and analysts working for media can characterize and compare team behaviours in a way that is interesting and understandable to the public. Coaches and team analysts can use these insights to identify which aspects of the team's playing behaviour are contributing to their success or failure and develop more effective strategies for future games.

6.3. Constraints and limitations

While our approach can be applied to a wide range of dynamic phenomena represented by multivariate time series, there are certain constraints and limitations to consider. One key assumption is the existence of meaningful time periods (i.e. differentiable stages in the development or functioning), which may not always be the case. Another constraint of the approach is the number of components that can be considered. The approach includes separate time division for each component and interactive integration of multiple divisions into a unified sequence of periods. This requires a human analyst to consider all variations together. However, as the number of components increases, the complexity of the analysis may become unmanageable. This may call for prioritizing the most relevant components for the analysis and exploring the trade-offs between complexity and interpretability. Additionally, the use of dimensionality reduction to handle multiple attributes can lead to specific challenges, such as sparsity issues when dealing with high-dimensional data.

One more constraint is related to the number of time steps in the time series, which is technically limited by the display resolution, as each time step requires at least one pixel in the timeline view. Additionally, the nature and rate of colour variation along the timeline can further limit the length of time series that can be effectively analysed, as a high number of abrupt changes can be difficult for human perception.

Also, the granularity of the time steps [AMST11] needs to match the intended scale of the analysis. For example, if the behaviour being studied occurs on an annual cycle, using minute-by-minute or even hourly data may not be appropriate. In such cases, it may be necessary to aggregate or downsample the data to a coarser temporal resolution. However, the use of data reduction techniques may lead to loss of information and potential bias, which need to be carefully evaluated. The choice of appropriate granularity depends on the specific properties of the phenomenon being studied, such as the duration of the cycles or the rate of change of the variables.

The existing limitations highlight the importance of careful consideration of the specific problem domain and the potential trade-offs between the accuracy and interpretability of the resulting periodization.

6.4. Directions for extension and future research

To address the constraints regarding time series length and granularity, our approach can be extended to involve a combination of computational and interactive visual techniques that allow for exploration and refinement of the time periodization. By iteratively adjusting the granularity of the time steps and, as a result, the length of the time series, analysts can refine their understanding of the underlying patterns in the data. For example, COVID-19 data could be initially aggregated by weekly time intervals to identify main stages in the pandemic's development. Analysts could then examine each stage in more detail and possibly further subdivide it using the daily data.

As a potential direction for future research, we see a possibility to formalise the time division problem as a problem of multi-criteria optimization and to develop computational methods for solving this problem and suggesting candidate solutions to analysts. Possible criteria include

- number of time periods – to be minimized;
- homogeneity of periods in terms of chosen similarity measures – to be maximized;
- homogeneity of changes between consecutive time steps within the periods – to be maximized;
- number and frequency of distinct transitions between periods – to be minimized;
- regularity of repeated transitions between time periods – to be maximized.

7. Conclusion

In this paper, we have presented our approach to time periodization for analysing and modelling behaviours of dynamic phenomena represented by multivariate time series. Our investigation was motivated by the observation that the problem of time division has not been fully considered in visual analytics and data science research. Through examples, we have defined the problem and developed a practical approach to solve it.

We have also explored various properties of data and goals of analysis that can impact the choice of analysis techniques and ways of using them. While our approach relies heavily on visual and interactive techniques, our investigation has revealed the possibility of formalizing this problem. This could open up opportunities to develop computational methods that can assist analysts in their work.

Overall, our research contributions include the definition and investigation of a previously neglected problem type, a practical and reproducible approach to solving problems of this type, and the potential for formalization and development of computational methods. We believe that our approach can be useful in various fields, including transportation, marketing, and sports analysis, and we hope that it can inspire future research in the area of time periodization.

Acknowledgements

This work was partially funded by Federal Ministry of Education and Research of Germany and the state of North-Rhine Westphalia

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as part of the Lamarr Institute for Machine Learning and Artificial Intelligence, Lamarr22B, by EU in project CrexData (grant agreement 101092749) and SoBigData++, and by German Science Foundation (DFG) in priority research program SPP 1894 on VGI.

Open access funding enabled and organized by Projekt DEAL.

References

[AAA*21] Andrienko G., Andrienko N., Anzer G., Bauer P., Budziak G., Fuchs G., Hecker D., Weber H., Wrobel S.: Constructing spaces and times for tactical analysis in football. *IEEE Transactions on Visualization and Computer Graphics 27*, 4 (2021), 2280–2297. https://doi.org/10.1109/TVCG.2019.2952129

[AAB*10] Andrienko G., Andrienko N., Bremm S., Schreck T., Von Landesberger T., Bak P., Keim D.: Space-in-time and time-in-space self-organizing maps for exploring spatiotemporal patterns. *Computer Graphics Forum 29*, 3 (2010), 913–922. https://doi.org/10.1111/j.1467-8659.2009.01664.x

[AAC*17] Andrienko N., Andrienko G., Camossi E., Claramunt C., Cordero Garcia J. M., Fuchs G., Hadzagic M., Jousselme A.-L., Ray C., Scarlatti D., Vouros G.: Visual exploration of movement and event data with interactive time masks. *Visual Informatics 1*, 1 (2017), 25–39. https://doi.org/10.1016/j.visinf.2017.01.004

[AACF22] Andrienko N., Andrienko G., Chen S., Fisher B.: Seeking patterns of visual pattern discovery for knowledge building. *Computer Graphics Forum 41*, 6 (2022), 124–148. https://doi.org/10.1111/cgf.14515

[AAD*10] Andrienko G., Andrienko N., Demsar U., Dransch D., Dykes J., Fabrikant S. I., Jern M., Kraak M.-J., Schumann H., Tominski C.: Space, time and visual analytics. *International Journal of Geographical Information Science 24*, 10 (2010), 1577–1600. https://doi.org/10.1080/13658816.2010.508043

[AAF*20] Andrienko N., Andrienko G., Fuchs G., Slingsby A., Turkay C., Wrobel S.: *Visual Analytics for Data Scientists*. Springer, 2020. https://doi.org/10.1007/978-3-030-56146-8

[AAFW17] Andrienko G., Andrienko N., Fuchs G., Wood J.: Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data. *IEEE Transactions on Visualization and Computer Graphics 23*, 9 (Sep. 2017), 2120–2136. https://doi.org/10.1109/TVCG.2016.2616404

[ALA*18] Andrienko N., Lammarsch T., Andrienko G., Fuchs G., Keim D., Miksch S., Rind A.: Viewing visual analytics as model building. *Computer Graphics Forum 37*, 6 (2018), 275–299. https://doi.org/10.1111/cgf.13324

[AMM*08] Aigner W., Miksch S., Müller W., Schumann H., Tominski C.: Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics 14*, 1 (2008), 47–60. https://doi.org/10.1109/TVCG.2007.70415

[AMST11] Aigner W., Miksch S., Schumann H., Tominski C.: *Visualization of Time-oriented Data*. Springer, 2011. https://doi.org/10.1007/978-0-85729-079-3

[BAP*05] Buono P., Aris A., Plaisant C., Khella A., Shneiderman B.: Interactive pattern search in time series. In *Visualization and Data Analysis* (2005), vol. *5669* of SPIE Proceedings, SPIE, pp. 175–186. https://doi.org/10.1117/12.587537

[BBB*18] Bernard J., Bors C., Bögl M., Eichner C., Gschwandtner T., Miksch S., Schumann H., Kohlhammer J.: Combining the automated segmentation and visual analysis of multivariate time series. In *EuroVA@ EuroVis* (2018), pp. 49–53. https://doi.org/10.2312/eurova.20181112

[BDB*16] Bernard J., Dobermann E., Bögl M., Röhlig M., Vögele A., Kohlhammer J.: Visual-interactive segmentation of multivariate time series. In *Proceedings of the EuroVis Workshop on Visual Analytics* (Goslar, DEU, 2016), Eurographics Association, p. 31–35. https://doi.org/10.2312/eurova.20161121

[Blo54] Bloch M.: *The Historian's Craft*. Manchester University Press, 1954.

[BPJ*20] Bastidon C., Parent A., Jensen P., Abry P., Borgnat P.: Graph-based era segmentation of international financial integration. *Physica A: Statistical Mechanics and its Applications 539*, (2020), 122877. https://doi.org/10.1016/j.physa.2019.122877

[BSH*16] Bach B., Shi C., Heulot N., Madhyastha T., Grabowski T., Dragicevic P.: Time curves: Folding time to visualize patterns of temporal evolution in data. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan 2016), 559–568. https://doi.org/10.1109/TVCG.2015.2467851

[BSM*15] Bernard J., Steiger M., Mittelstädt S., Thum S., Keim D., Kohlhammer J.: A survey and task-based quality assessment of static 2D colormaps. In *Visualization and Data Analysis 2015* (2015), Kao D. L., Hao M. C., Livingston M. A., Wischgoll T., (Eds.). vol. *9397*, International Society for Optics and Photonics, h, p. 93970M. https://doi.org/10.1117/12.2079841

[BSP*22] Brich N., Schulz C., Peter J., Klingert W., Schenk M., Weiskopf D., Krone M.: Visual analytics of multivariate intensive care time series data. *Computer Graphics Forum 41*, 6 (2022), 273–286. https://doi.org/10.1111/cgf.14498

[BWK*13] Bernard J., Wilhelm N., Krüger B., May T., Schreck T., Kohlhammer J.: Motionexplorer: Exploratory search in human motion capture data based on hierarchical aggregation. *IEEE Transactions on Visualization and Computer Graphics 19*, 12 (2013), 2257–2266. https://doi.org/10.1109/TVCG.2013.178

[BWS*12] Bernard J., Wilhelm N., Scherer M., May T., Schreck T.: TimeSeriesPaths: Projection-based explorative

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

analysis of multivariate time series data. *Journal of WSCG* (2012), 97–106.

[CC99] Combi C., Chittaro L.: Abstraction on clinical data sequences: an object-oriented data model and a query language based on the event calculus. *Artificial Intelligence in Medicine 17*, 3 (1999), 271–301. https://doi.org/10.1016/S0933-3657(99)00022-6

[CC01] Chittaro L., Combi C.: Visual definition of temporal clinical abstractions: A user interface based on novel metaphors. In *Artificial Intelligence in Medicine* (Berlin, Heidelberg, 2001), Quaglini S., Barahona P., Andreassen S., (Eds.), Springer Berlin Heidelberg, pp. 227–230. https://doi.org/10.1007/3-540-48229-6_32

[CO12] Combi C., Oliboni B.: Visually defining and querying consistent multi-granular clinical temporal abstractions. *Artificial Intelligence in Medicine 54*, 2 (2012), 75–101. https://doi.org/10.1016/j.artmed.2011.10.004

[Col22] Collins: Periodization. In *Collins Online English Dictionary*. Collins, 2022. [Online; accessed 8-August-2022]. URL: https://www.collinsdictionary.com/dictionary/english/periodization

[Gad88] Gadia S. K.: A homogeneous relational model and query languages for temporal databases. *ACM Transaction on Database Systems 13*, 4 (October 1988), 418–448. https://doi.org/10.1145/49346.50065

[GAM*14] Gschwandtner T., Aigner W., Miksch S., Gärtner J., Kriglstein S., Pohl M., Suchy N.: Timecleanser: A visual analytics approach for data cleansing of time-oriented data. In Proceedings of the 14th International Conference on Knowledge Technologies and Data-driven Business (2014), pp. 1–8.

[GCML06] Guo D., Chen J., MacEachren A., Liao K.: A visualization system for space-time and multivariate patterns (VIS-STAMP). *IEEE Transactions on Visualization and Computer Graphics 12*, 6 (2006), 1461–1474. https://doi.org/10.1109/TVCG.2006.84

[Goo22] Google: Google mobility trends, 2022. [Online; accessed 7-September-2022]. URL: https://ourworldindata.org/covid-google-mobility-trends

[GYD*19] Gharghabi S., Yeh C.-C. M., Ding Y., Ding W., Hibbing P., LaMunion S., Kaplan A., Crouter S. E., Keogh E.: Domain agnostic online semantic segmentation for multi-dimensional time series. *Data Mining and Knowledge Discovery 33*, 1 (2019), 96–130. https://doi.org/10.1007/s10618-018-0589-3

[HÖ2] Höppner F.: Time series abstraction methods - a survey. In *Informatik Bewegt: Informatik 2002-32. Jahrestagung Der Gesellschaft Für Informatik e.v. (GI)* (2002), GI, p. 777–786.

[Hay98] Haydu J.: Making use of the past: Time periods as cases to compare and as sequences of problem solving. *American Journal of Sociology 104*, 2 (1998), 339–371. https://doi.org/10.1086/210041

[HB03] Harrower M., Brewer C. A.: Colorbrewer.org: An online tool for selecting colour schemes for maps. *The Cartographic Journal 40*, 1 (2003), 27–37. https://doi.org/10.1179/000870403235002042

[HS04] Hochheiser H., Shneiderman B.: Dynamic query tools for time series data sets: Timebox widgets for interactive exploration. *Information Visualization 3*, 1 (2004), 1–18. https://doi.org/10.1057/palgrave.ivs.9500061

[HW79] Hartigan J. A., Wong M. A.: Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics) 28*, 1 (1979), 100–108. https://doi.org/10.2307/2346830

[HWX*10] Hu Y., Wu S., Xia S., Fu J., Chen W.: Motion track: Visualizing variations of human motion data. In *IEEE Pacific Visualization Symposium (PacificVis)* (2010), pp. 153–160. https://doi.org/10.1109/PACIFICVIS.2010.5429596

[JCG*92] Jensen C. S., Clifford J., Gadia S. K., Segev A., Snodgrass R. T.: A glossary of temporal database concepts. *SIGMOD Rec. 21*, 3 (September 1992), 35–43. https://doi.org/10.1145/140979.140996

[KP11] Kaski S., Peltonen J.: Dimensionality reduction for data visualization [applications corner]. *IEEE Signal Processing Magazine 28*, 2 (2011), 100–104. https://doi.org/10.1109/MSP.2010.940003

[Kru64] Kruskal J. B.: Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika 29*, 1 (Mar. 1964), 1–27. https://doi.org/10.1007/BF02289565

[LAW*18] Liu S., Andrienko G., Wu Y., Cao N., Jiang L., Shi C., Wang Y.-S., Hong S.: Steering data quality with visual analytics: The complexity challenge. *Visual Informatics 2*, 4 (2018), 191–197. https://doi.org/10.1016/j.visinf.2018.12.001

[Met06] Metzger W.: *Laws of Seeing*. MIT Press, Cambridge, MA, USA, 2006.

[Phi58] Phillips A. W.: The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861–1957. *Economica 25*, 100 (1958), 283–299. https://doi.org/10.1111/j.1468-0335.1958.tb00003.x

[Sam69] Sammon J. W.: A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers 18*, 5 (May 1969), 401–409. https://doi.org/10.1109/T-C.1969.222678

[Sat01] Sato M.: Time, chronology, and periodization in history. In *International Encyclopedia of the Social & Behavioral Sciences*, Smelser N. J., Baltes P. B., (Eds.). Pergamon, Oxford, 2001, pp. 15686–15692. https://doi.org/10.1016/B0-08-043076-7/02609-7

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

[Sch07] SCHANDA J.: *Colorimetry: understanding the CIE system*. John Wiley & Sons, 2007.

[Sch22] SCHUBERT E.: Stop using the elbow criterion for k-means and how to choose the number of clusters instead, 2022. https://doi.org/10.48550/ARXIV.2212.12189

[Sha97] SHAHAR Y.: A framework for knowledge-based temporal abstraction. *Artificial Intelligence 90*, 1 (1997), 79–133. https://doi.org/10.1016/S0004-3702(96)00025-2

[Shn94] SHNEIDERMAN B.: Dynamic queries for visual information seeking. *IEEE Software 11*, 6 (1994), 70–77. https://doi.org/10.1109/52.329404

[SLCB07] SACCHI L., LARIZZA C., COMBI C., BELLAZZI R.: Data mining with temporal abstractions: Learning rules from time series. *Data Mining and Knowledge Discovery 15*, 2 (October 2007), 217–247. https://doi.org/10.1007/s10618-007-0077-7

[Tho53] THORNDIKE R.: Who belongs in the family? *Psychometrika 18*, 4 (December 1953), 267–276. https://doi.org/10.1007/BF02289263

[vBR*16] VON LANDESBERGER T., BRODKORB F., ROSKOSCH P., ANDRIENKO N., ANDRIENKO G., KERREN A.: Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan 2016), 11–20. https://doi.org/10.1109/TVCG.2015.2468111

[vdEHBvW16] VAN DEN ELZEN S., HOLTEN D., BLAAS J., VAN WIJK J. J.: Reducing snapshots to points: A visual analytics approach to dynamic network exploration. *IEEE Transactions on Visualization and Computer Graphics 22*, 1 (Jan 2016), 1–10. https://doi.org/10.1109/TVCG.2015.2468078

[vdMH08] VAN DER MAATEN L., HINTON G.: Visualizing data using t-SNE. *Journal of Machine Learning Research 9*, 86 (2008), 2579–2605. URL: http://jmlr.org/papers/v9/vandermaaten08a.html

[vWvS99] VAN WIJK J. J., VAN SELOW E. R.: Cluster and calendar based visualization of time series data. In *Proc. IEEE Symposium on Information Visualization (InfoVis)* (1999), pp. 4–9. https://doi.org/10.1109/INFVIS.1999.801851

[Yan05] YANG L.: Data embedding techniques and applications. In *Proceedings of the 2nd International Workshop on Computer Vision Meets Databases* (New York, NY, USA, 2005), CVDB '05, Association for Computing Machinery, pp. 29–33. https://doi.org/10.1145/1160939.1160948

[YPK13] YANG Z., PELTONEN J., KASKI S.: Scalable optimization of neighbor embedding for visualization. In *ICML (2)* (2013), vol. *28* of JMLR Workshop and Conference Proceedings, JMLR.org, pp. 127–135.

© 2023 The Authors. *Computer Graphics Forum* published by Eurographics - The European Association for Computer Graphics and John Wiley & Sons Ltd.

14678659, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/cgf.14845 by City University Of London Library, Wiley Online Library on [01/06/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License