Data storytelling for univariate time series in automotive sensor data

Abstract
With growing data volumes, visually analysing univariate time series data from sensor measurements becomes increasingly challenging for vehicle manufacturers. This paper explores data storytelling as a method to increase the efficiency of insight comprehension for complex data. Through expert interviews, we identify the trends most common in univariate time series created with a durability tracing method developed for automotive durability testing. The data storytelling process is tested on two domain-specific scenarios, to investigate whether data storytelling can be integrated into the analysis process to identify key insights and reduce the cognitive load for mechanical engineers. Furthermore, we implement dynamic trend recognition methods specifically tailored for univariate time series analysis to evaluate the outcome of data storytelling in the use case of automotive development data. This research contributes to the field of visual data analysis by demonstrating the value of data storytelling for domain-specific data stories and providing practical guidance on implementing trend detection algorithms in TypeScript for web-based visualization tools. The findings hold significant value for professionals in the automotive and durability industries, as well as researchers in data visualization and analysis.

Keywords
data storytelling, univariate time series, data visualizations, trend detction, sensor data

1. Introduction

The ever-growing size of databases in the automotive and mobility sectors necessitates effective methods for data evaluation, algorithm development and insight visualization. In the domain of durability testing, engineers rely heavily on visualizations to extract valuable insights from complex measurement data. However, as data volumes increase, visually analysing key information becomes increasingly challenging [1]. Data stories can increase the efficiency of insight comprehension when compared to traditional visualization methods, regardless of a reader's data literacy [2]. By reducing visual clutter, data storytelling highlights the key insights that contribute to data understanding [3]. However, data storytelling goes beyond mere visualization techniques. When data representations in the form of charts and diagrams are combined with storytelling elements, they can reduce the cognitive load of synthesizing large amounts of new information [4]. The cognitive load can be defined as the extent of mental resources used, for example, when identifying patterns in a dataset [5]. Furthermore, data storytelling can influence the decision-making process and make insight communication more persuasive [6, 7]. The visual storytelling process includes the exploration of the domain data to identify key insights, compiling visualizations and storytelling elements to write a compelling narrative and communicating the story to the target audience [8]. This process can be divided into a

LWDA'24: Lernen, Wissen, Daten, Analysen. September 23–25, 2024, Würzburg, Germany

CEUR Workshop Proceedings (CEUR-WS.org)

sequence of five steps, defined as understanding the domain context, choosing an appropriate visualization, removing visual clutter, focusing the viewer attention and data storytelling [9].

This research investigates the potential of data storytelling in enhancing the visual analysis of automotive sensor data. Driving styles significantly influence the forces exerted on vehicle components. Drivers tend to adapt their driving behaviour over time, leading to inconsistencies when repeatedly driving the same route. During durability testing, this driver variability can influence test results, which compromises the early detection of potential component failures. In this research, data storytelling is used to analyse the level of damage to a vehicle during a test drive. Furthermore, a second data story highlights the key insights in the univariate time series of each sensor measurement, to investigate the causes for the vehicle damage. Our methodology involved expert interviews with engineers, implementation of automated trend detection algorithms to highlight key data points, and the creation of an interactive website. To evaluate the effectiveness of data storytelling for univariate tine series automotive sensor data, we assessed the final implementation of the interactive website through a final expert interview.

The following Section 2 introduces the data model of the use case and presents the trend detection methods based on the results of the expert interviews. Section 3 outlines the data storytelling process and presents the findings of the data visualization methods. A brief discussion of the evaluation results and directions for further research are provided in Section 4.

2. Data Model and Analytics

During automotive durability testing, vehicles navigate a diverse set of terrains, including forest paths, cobblestones and smooth asphalt. For each test drive, acceleration sensors and strain gauges mounted on the vehicle measure the external forces on the vehicle [10]. The durability-transfer (DT) method utilizes these measurements to calculate the loads and stresses on individual vehicle components [11]. The results can be used to assess component endurance strength by comparing fatigue behaviour to target values or recording stress events as a univariate time series. A univariate time series can be defined as a vector $x \in \mathbb{R}^n$ representing n measurements at constant time intervals. These measurements occur at discrete points in time $t = (t_1, t_2, \ldots, t_n)$. Each element x_i is a single measurement at time t_i.

To identify the key insights relevant to mechanical engineers in the univariate time series during durability testing, a guided interview with several domain experts was conducted. The transcripts were analysed with the qualitative content analysis method [12, 13]. The most frequently types of insights mentioned during the interview are local peaks, offsets, oscillations, drifts, outliers and linear trends. These trends can be detected using the analytical methods described below.[1]

[1]The implementation of the trend detection methods will be made publicly available upon acceptance.

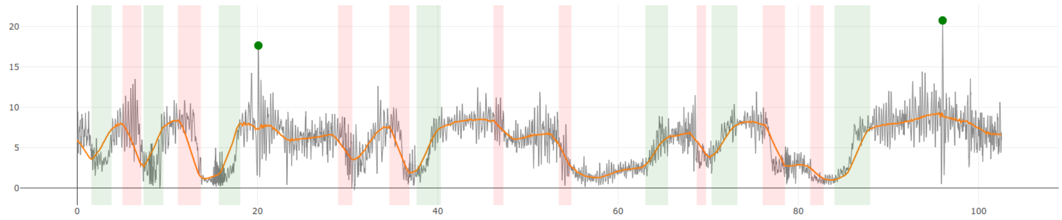


Figure 1: Trend detection in a univariate time series: moving average, linear trends and outliers

Most trend detection methods use the moving average $a \in \mathbb{R}^n$, which calculates the arithmetic mean of $m \in \mathbb{N}$ data points on either side of a central value x_i in a univariate time series $x \in \mathbb{R}^n$. Each element a_i of the moving average time series can be calculated as follows:

$$a_i = \frac{1}{2m+1} \sum_{k=-m}^{m} x_{i+k}, \quad i = m+1, \ldots, n-m \tag{1}$$

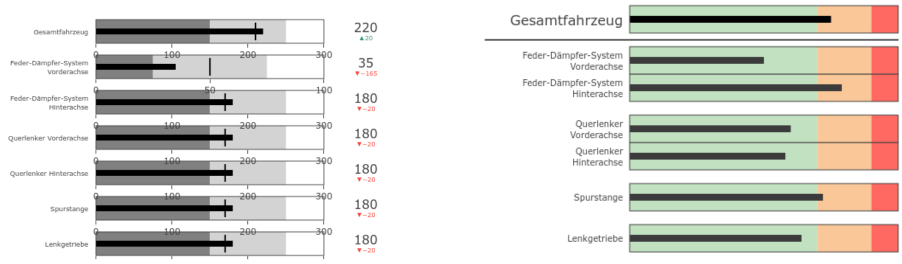
The elements outside the given range are calculated by adjusting the data range window [14]. The moving average, which is visualized in figure 1, can be used to identify local peaks in a univariate time series. Experts named local peaks during the interview as the most common trend mechanical engineers look for in time series data. These maxima and minima are commonly caused by variations in the driving style, for example a sudden increase in acceleration force due to alternating road surfaces. Local peaks can be detected through comparison with neighbouring data points [15]. A data point t_i is a maximum if its value x_i is greater than both the preceding and succeeding point x_{i-1} and x_{x+i}. To identify unique local peaks, the set of maxima is filtered with a domain specific minimum height. Next, given a pre-defined distance threshold and starting with the highest valued maxima, any local peaks deemed non-essential are removed. Finally, local peaks exceeding a minimum distance from the moving average are selected:

$$E = \{t_i \mid t_i \in M \wedge x_i > a_i + b\} \tag{2}$$

where E is the set of local peaks, M is the set of all filtered maxima, $a_i \in \mathbb{R}$ is the moving average at time t_i and $b \in \mathbb{R}$ is the buffer value. Minimal local peaks can be identified similarly.

Local peaks can also be outliers, which are visualized in Figure 1 with round markers. Experts defined outliers as data points outside the expected value range. Outliers can be caused by malfunctioning sensors or sudden vehicle component failure. To identify outliers, the set of local peaks is filtered by upper and lower limits, which are calculated based on percentage deviations specified by a mechanical engineer.

Based on the moving average, increasing and decreasing linear trends in a univariate time series x can be detected. The starting point of an increasing linear trend is identified when the moving average value a_i of a data point t_i is smaller than the values a_{i-1} and a_{i+1} of the two neighbouring points. Subsequent points are considered part of the same trend, if their moving average value is higher than their left neighbour and lower than the right neighbour. Each linear trend must include a predefined minimum number of data points and exceed the height difference $h = (\max(x) - \min(x)) \cdot p$, where p is a pre-defined percentage. Examples of linear trends detected in a univariate time series can be seen in Figure 1.



(a) Traditional visualization (b) Data story with narrative

Figure 2: Bullet charts visualizing the damage load on a test vehicle

During the interview, experts emphasized the importance of offset and drift detection, which can occur when vehicle components are plastically deformed. Drifts are identified if the moving average of the first data point t_1 significantly differs from the moving average of the last point t_n. Offsets, characterized by more erratic changes over time, are detected if the average of the second half of the univariate time series exceeds the sum of a buffer and the average of the first half. Further analysis can detect oscillations between subsequent minimum and maximum values in a univariate time series. The trend detection method considers both the minimum amplitude and maximum distance between local peaks. Only if both the distance and height criteria are satisfied, the combined minimum-maximum pair are classified as an oscillations and returned.

3. Data Storytelling

This section explores the application of data storytelling principles to automotive sensor data, to demonstrate how data stories can improve the communication of complex information compared to traditional visualizations. We examine two domain-specific scenarios, each including a corresponding visualization with an insight generating narrative. The first use case analyses vehicle damage, based on insights gained from value counting methods. The second data story utilizes the trend detection methods introduced in Section 2, to highlight key insights in a univariate time series representing individual measurement sequences.

3.1. Visual analysis of vehicle damage

In the first use case, the data story should enable mechanical engineers to quickly assess the current level of damage to a test vehicle. Following the five steps of data storytelling, analysing the domain context led to choosing an appropriate visualization. In this case, each test vehicle has an upper damage limit based on previous test results, above which the damage to the vehicle becomes critical. The visualization should therefore represent the previously calculated damage load as a percentage of the total limit value. Bullet charts are particularly suitable for visualizing measurements in relation to a target range and are ideal for displaying multiple damage loads simultaneously due to their space efficiency [16]. Figure 2a displays an example of a traditional

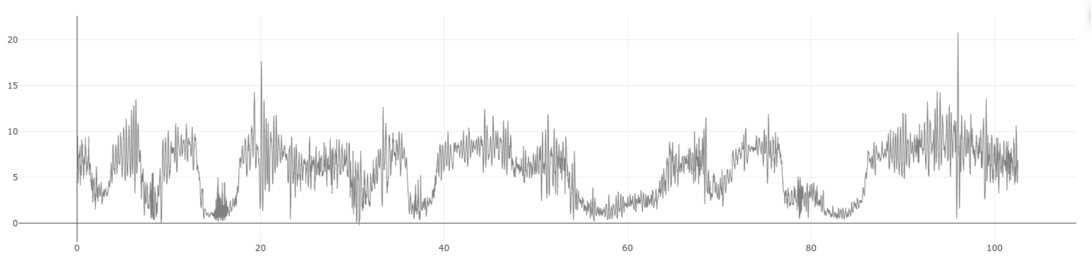


Figure 3: Traditional visualization of a univariate time series without data storytelling

bullet chart visualizing the damage load data.

During the third step of the data storytelling process, the visual clutter in Figure 2a is removed. Based on the design principles of data storytelling, the delta values, redundant value labels and tick marks are eliminated as they have limited relevance for the damage assessment [17]. To focus the attention of viewers on critical components, related vehicle sensors are grouped together. Next, the whitespace between graphs is scaled up to increase clarity and readability in the visualization. Finally, the total damage to the vehicle is emphasized with a larger, dark gray font and visually separated from the graphs of the individual components.

In the final step of the storytelling process, the data story narrative is integrated into the bullet chart visualization. As can be seen in Figure 2b, the current level of damage to the test vehicle is now presented as a percentage of the upper damage limit. To highlight the urgency for vehicle maintenance, the warning zone starting at 70% total damage and the critical zone starting at 90% total damage are coloured in yellow and red respectively. To increase insight comprehension, the colours of the traffic light system are chosen for quick recognition. Finally, critical damage levels of the test vehicle trigger a warning message on the interactive website. The warning highlights the affected vehicle components and recommends vehicle maintenance as a call to action.

3.2. Visual analysis of univariate time series data

Following the first use case, the second data story continues the narrative by analysing how the current damage level to the test vehicle occurred. To expedite the analysis of the sensor data measured during each test drive, each univariate time series is examined individually. To start the data storytelling process, the domain context of the data story was analysed. The forces on the test vehicle are recorded with acceleration and gauge sensors at a frequency of 500 Hz. The loads and stresses on each vehicle component can then be calculated by the DT method, creating a univariate time series as defined in Section 2. Time series data are an integral part of operational durability research and are commonly displayed as line graphs to visualize the change over time [18]. Therefore, this well established chart type was chosen, to reduce the cognitive load for mechanical engineers during the visual analysis. Figure 3 visualizes a traditional time series chart with a univariate time series.

During the third step of the data storytelling process, the visual clutter in the graph is eliminated. As each sensor measurement shares a similar value range, the x-axis was centred in the graph to better visualize oscillations and the lines of both axes were removed. The number of ticks on the x-axis was limited to four ticks per graph, and the commonly used grid lines in the background were decolorized. To focus the attention of viewers during the fourth step of data storytelling, the trend detection methods from Section 2 were implemented dynamically, to highlight key insights in the univariate time series. As can be seen in Figure 4, interesting features such as local peaks and the moving average are emphasized with green markers. To integrate the data story narrative during the final step of the process, a hover function was implemented to enable users to easily read the values of each data point. Finally, the dynamic trend detection methods were implemented as an interactive side menu, to allow users to visually explore the univariate time series. These interactive tools communicate of the key insights in the sensor data measurements and explain the cause of the vehicle damage presented in the previous use case.

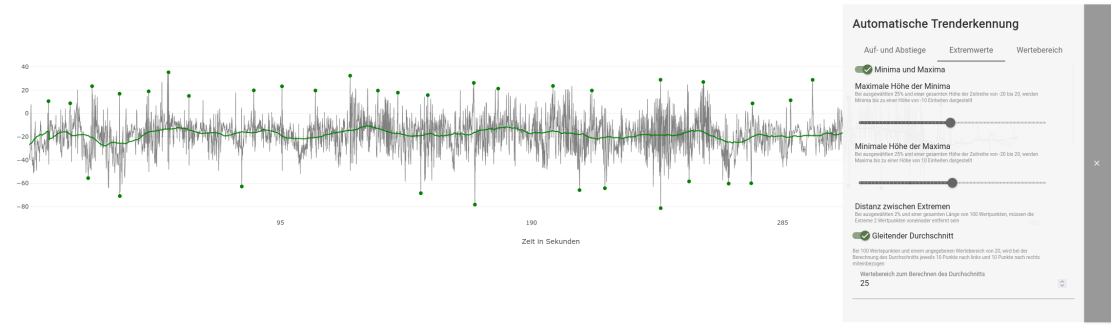


Figure 4: Data story with sensor measurements and interactive trend detection module

4. Conclusion and future research

This study investigated the effectiveness of data storytelling for detecting and communicating insights from univariate automotive sensor time series data. The results of the data storytelling process were evaluated in a final expert interview, revealing positive feedback on their user-friendliness, comprehensibility, and goal orientation. Data storytelling was considered an effective tool for generating analytical insights and communicating significant details during the decision-making process. Furthermore, experts agreed that data storytelling facilitated efficient analysis and reduced cognitive load for mechanical engineers compared to traditional data exploration methods.

Significantly, this research highlights the suitability of data storytelling for automotive sensor data. However, the core principles of the data storytelling process can be readily adapted to other domains, offering a versatile tool for knowledge extraction and communication. Future research will focus on integrating Large Language Models (LLMs) into the data storytelling process. By incorporating LLMs' ability to generate narratives, the communication of data insights within the visual analysis process can be improved.

References

[1] R. Ding, S. Han, Y. Xu, H. Zhang, D. Zhang, Quickinsights, in: P. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, T. Kraska (Eds.), Proceedings of the 2019 International Conference on Management of Data, ACM, New York, NY, USA, 2019, pp. 317–332. doi:10.1145/3299869.3314037.

[2] H. Shao, R. Martinez-Maldonado, V. Echeverria, L. Yan, D. Gasevic, Data storytelling in data visualisation: Does it enhance the efficiency and effectiveness of information retrieval and insights comprehension?, in: F. F. Mueller, P. Kyburz, J. R. Williamson, C. Sas, M. L. Wilson, P. T. Dugas, I. Shklovski (Eds.), Proceedings of the CHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2024, pp. 1–21. doi:10.1145/3613904.3643022.

[3] R. Rosenholtz, Y. Li, J. Mansfield, Z. Jin, Feature congestion, in: W. Kellogg, S. Zhai, G. van der Veer, C. Gale (Eds.), Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, 2005, pp. 761–770. doi:10.1145/1054972.1055078.

[4] B. Dykes, Effective data storytelling: How to drive change with data, narrative, and visuals, Wiley, Hoboken New Jersey, 2020.

[5] J. L. Plass, R. Moreno, R. Brünken, Cognitive load theory, Cambridge University Press, Cambridge and New York, 2010.

[6] A. V. Pandey, A. Manivannan, O. Nov, M. Satterthwaite, E. Bertini, The persuasive power of data visualization, IEEE transactions on visualization and computer graphics 20 (2014) 2211–2220. doi:10.1109/TVCG.2014.2346419.

[7] M. K. Daradkeh, An empirical examination of the relationship between data storytelling competency and business performance, Journal of Organizational and End User Computing 33 (2021) 42–73. doi:10.4018/JOEUC.20210901.oa3.

[8] B. Lee, N. H. Riche, P. Isenberg, S. Carpendale, More than telling a story: Transforming data into visually shared stories, IEEE computer graphics and applications 35 (2015) 84–90. doi:10.1109/MCG.2015.99.

[9] C. N. Knaflic, Storytelling with data: A data visualization guide for business professionals, Wiley, Hoboken New Jersey, 2015.

[10] S. Keil (Ed.), Dehnungsmessstreifen, Springer Fachmedien Wiesbaden, Wiesbaden, 2017. doi:10.1007/978-3-658-13612-3.

[11] monalysis, Durability transfer method, 2024. URL: https://www.monalysis.de/uk/.

[12] P. Mayring, Qualitative Inhaltsanalyse: Grundlagen und Techniken, neuausgabe ed., Beltz Verlagsgruppe, Weinheim, 2022.

[13] U. Kuckartz, S. Rädiker, Qualitative Inhaltsanalyse Methoden, Praxis, Computerunterstützung, Grundlagentexte Methoden, 5. [überarbeitete] auflage ed., Beltz Juventa, Weinheim, 2022.

[14] R. J. Hyndman, Moving averages, in: M. Lovric (Ed.), International Encyclopedia of Statistical Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011, pp. 866–869. doi:10.1007/978-3-642-04898-2{\textunderscore}380.

[15] Samuel Pröll, Finding peaks in noisy signals, 2022. URL: https://www.samproell.io/posts/signal/peak-finding-python-js.

[16] J. Schwabisch, S. Ribecca, Financial times visual vocabulary: Designing with data, 2024.
[17] J. A. Schwabish, Better data visualizations: A guide for scholars, researchers, and wonks, Columbia University Press, New York, 2021.
[18] S. Götz, K.-G. Eulitz, Betriebsfestigkeit, Springer Fachmedien Wiesbaden, Wiesbaden, 2020. doi:10.1007/978-3-658-31169-8.