Visualizing Streaming of Ordinal Big Data

João Moreira*, Henrique Ferreira, Daniel Gonçalves
INESC-ID, Instituto Superior Técnico, Universidade de Lisboa
* Email: joaopedro.moreira@inesc-id.pt

2022 International Conference on Graphics and Interaction (ICGI) | 979-8-3503-4573-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICGI57174.2022.9990170

***Abstract*— Horizontal transitions are used in the Streaming of Big Data when there is the need to change the aggregation level of the data being presented. For example, data in a heat map may be aggregated into a line chart. Although these transitions have already been studied for quantitative streamed big data, ordinal data remains unchecked. In this study, we conducted an empirical study to explore horizontal transitions for ordinal data using Graceful Degradation, a concept that allows an overview of the received data at different periods via different levels of aggregation. We chose four visual idioms (Histogram, Ordinal Scatter Plot, Heat Map, and Line Chart), created several transitions between them, and tested how effectively could people perceive data in each idiom before, during, and after each corresponding transition. Participants had to watch numerous videos showcasing the idioms and transitions, and then they had to answer a questionnaire for us to measure how effective was their perception. All the four idioms tested were effective, and we were able to define numerous design guidelines for the creation of horizontal transitions in Streaming of Ordinal Big Data.**

***Index Terms*—Information Visualization, Big Data Streaming, User Study, Horizontal Transitions**

I. Introduction

Information analysis is essential for data recognition and decision-making in multiple areas of scientific study. The advancement of information technologies, such as smartphones or IoT devices, has originated a growth in the amount of accessible and valuable data created and stored by increasingly more entities. When the dimensions of these datasets defy computationally their representation, it is safe to say that the dataset belongs in Big Data. Then, when data arrives in real-time, it belongs to Data Streaming.

Let us imagine that a company needs to keep track of how many people access their website and how they navigated it. They have stored data since the company started working, and have now millions of records to be visualized. However, visualizing Big Data is an ongoing challenge because it is hard to represent everything simultaneously without producing visual clutter, compromising the overall analysis. This situation means that Big Data visualizations are obliged to find adequate aggregation techniques to effectively represent the whole dataset by reducing the overall complexity of the data. For example, if not enough aggregation is applied, the visualization may be forced to depict too many elements, which most likely affects the system's hardware performance. Besides, the overdrawing of elements will make the visualization impossible to understand. In both cases, the analysis is compromised. However, the solution should not always apply high aggregation levels because the visualization might lose the ability to present valuable information. It should therefore be dependent on the intended detail designers want to convey.

Now let us imagine that, at a certain moment, the company started monitoring in real-time the same data records that they had been storing since the beginning, and they were receiving thousands of records per minute. This time, when information is being received in real-time, regardless of being Big Data, designers are faced with new challenges. Streaming is characterized by the data's moment of creation during the observation of the visualization. The primary concern is the representation of the data as soon as it is received, and without it constantly changing due to the new data. One visualization must be able to process information and show it as it is received simultaneously. Unlike Big Data, visualizations must use adequate visual idioms and internal structures to manage information in real-time.

The company was now facing two different challenges, Big Data and Data Streaming (Big Data Streaming) and new issues need now to be solved. Since the data is streamed, it cannot be processed before the start of the visualization. Then, because it is Big Data, it needs to be aggregated to be perceptible. Therefore, besides processing the data, the data must be grouped in real-time. However, this grouping needs to be handled carefully. If grouped too soon, the system could, again, lose necessary information. If too late, the visualization will hold too many visual elements, thus producing visual clutter. Finally, the data timestamps should be explicit in the visualization. Newer data is usually the focus of Streaming visualizations. However, since Big Data information is aggregated, older data may also contain important information. Therefore, the visualization should allow an overview of how the data evolves by presenting different data periods. The company then decided that it needed a way to visualize the data gathered since the beginning, and how data was evolving in different periods.

Currently, there is yet no final solution to this issue. However, one system in development called VisMillion (fig. 1) has been taking the first steps in Big Data Streaming Visualizations, in hopes of allowing people to have an overview of data at different periods. However, all the work conducted in VisMillion has assumed that data is quantitative. We question how well would people perceive data depicted using this prototype if ordinal data was applied instead. Therefore, our goal was to **choose visual idioms for streaming of ordinal big data**, we designed **horizontal transitions** between those

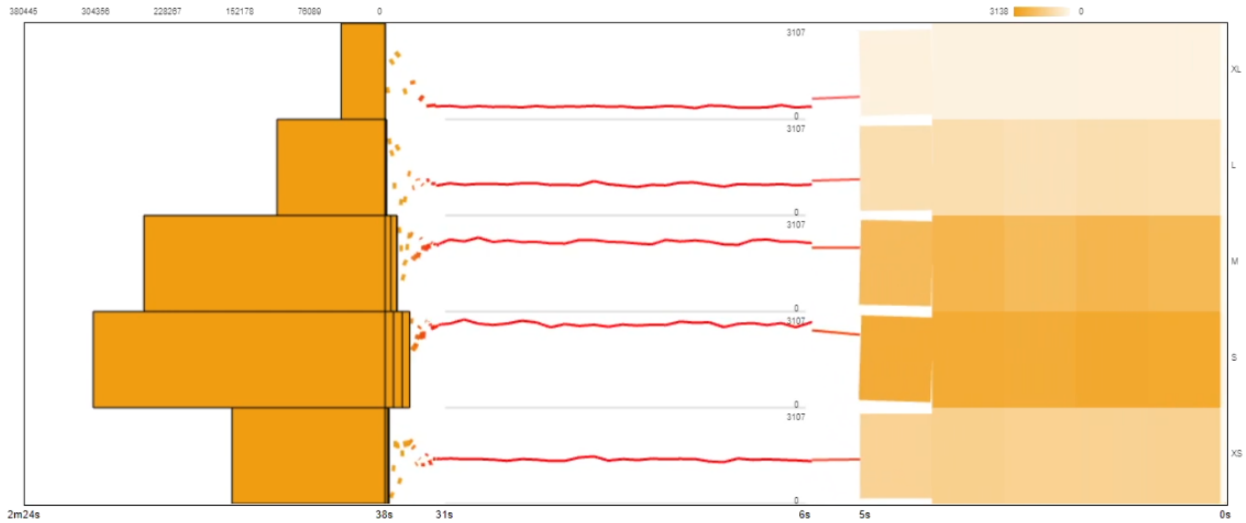

Fig. 1. Our prototype of VisMillion depicting the concept of Graceful Degradation from right to left. In this example, the streamed data arrives at the Heat Map, then it goes to the Line Chart via a horizontal transition, and then it goes to the Histogram with another horizontal transition. Each visual idiom corresponds to a different aggregation level, which represents different periods.

idioms, and then we conducted a **user study** to measure how accurately would people perceive both the information depicted in the idioms and transitions between them. Our major contribution is **design guidelines for choosing visual idioms and horizontal transitions for streaming of ordinal big data**.

II. Related Work

Big Data Streaming imposes several challenges in different domains [15], [16], [21] such as the social networks [23], industry [13], educational sector, healthcare system [1], [20], financial transaction, national security [18], oil and gas industries, and transportation [27], [32]. The datasets are often dynamic and characterized by high variety and volatility [5], [7], [21], [28], and their processing usually involves five steps: cleaning, aggregation, encoding, storage, and access [5], [7], [13], [15], [21], [28]. However, additional V's such as Visualization and Value have also been used to improve the definition of big data [24]. For visualization, systems must be designed to support real-time interaction, quick data processing, visual scalability, user assistance, and personalization [5], [13], [17], [29]. Machine learning, for example, is used to process vasts amounts of data to feed systems who output structures that help to predict and analysis trends and patters via visualization techniques [13], [21], [29], [31]. These insights can then be used for effective decision-making and reporting [13]. However, people must be able to have an effective overview of the data depicted, and must be allowed to filter any detail if needed [13], [29].

A. Challenges

There are a lot of popular visualization tools [21], [29], but the majority of visualizations systems cannot handle the size of Big Data datasets [5], [13], [15], [29] because of limited computational and memory resources [5], which often lead to overloading issues [5], [17], [21], [28]. Therefore, data reduction techniques are usually applied. However, these must be carefully applied. On one hand, interesting data patterns may be lost due to careless aggregations. On the other, if not enough aggregations are applied, the visualization may become too dense or cluttered [7]. Regarding streaming, systems need to allow real-time data exploration [5], [13], [17], which may not be easy to accomplish due to the variability of streamed datasets structures [5].

B. Solutions

Depending on the data structure, different visual idioms may be used [8], [9], [12]. Dimension reductions techniques are often designed with hierarchical visual idioms like the Tree Map or Circle packing [5], [10]. Still, other solutions also deal with the high data volume, variety, and dynamics, like the Parallel Coordinates, or any Stacked Graph. The last type in particular works better for temporal data [7] because they allow to see data over time [14]. Then, designers may also employ specialized software to deal with data storage, managing, and analyzes [15], such as the MapReduce, Haddop, or specialized techniques just as the PASS (Preserving Anomaly and Semantics Sampling) [2].

Particularly in streaming, designers focus on having data seen in separate timespans, each for a specific analysis [15], [19]. If possible, people should be able to see data as it arrives [4], [17], [33], for example, to detect anomalies in real-time [10], [30]. Furthermore, the data must be processed without burdening the system's memory too much [7], [13]. Moreover, if different timespan windows exist, designers should employ techniques to preserve a viewer's mental map of incremental results [11]. To ensure this preservation, designers may choose to use animation techniques to help users track how the data evolves. [34]. For example, grouping objects by predefined trending directions and by clustering their moving trends.

C. Graceful Degradation

Recently, the concept of Graceful Degradation [3] has been applied to Big Data Streaming in an innovative system called VisMillion [26]. It was designed to depict time-series big data streamed datasets in different modules, each using a different visual idiom. With this concept, as data gets older, it gets aggregated into different visual idioms. The core idea is that recent data might need more detail than older data. In VisMillion, each module presents data with a different aggregation level. The data flows from right (most recent data) to left (older data), where the left-most module depicts information since the visualization started working.

However, the initial proposal had abrupt cuts between modules, meaning that data got aggregation between visual idioms, but there were no visual cues to preserve the viewer's mental map of what was happening. Therefore, to enhance VisMillion, horizontal animated transitions were implemented [25]. Horizontal transitions are used between two modules to depict how information gets aggregated. For example, between a heat map that shows data distribution and a histogram that aggregates data into categories. Then, vertical transitions were also added to VisMillion [6], [22] to allow one module to change its own visual idiom. Vertical transitions are used in one module to change how the information is displayed. For example, using a heat map, a trend shift might not be detected. Therefore, a line chart would better emphasize this anomaly.

D. Discussion

As we have seen, Big Data Stream is an ongoing challenge due to its datasets properties. There are ongoing attempts to deal with variety and volatility [10], [15], [17], [21], and to deal with Visualization, and Value [5], [13]. Then, in some cases, viewers might be able to see information in different periods [15]. From our literature review, VisMillion was the system that has been actively trying to provide a complete solution to big data stream regarding the visualization phase. It receives Streams of Big Data that can be visualized with different aggregations levels [26] using graceful degradation [3], and it supports animation when data shifts between modules [25] or when viewers need to see information differently [6], [22]. However, the studies conducted until now assumed only quantitative data, and it is known that datasets in Big Data Streams can vary significantly, and there is the need to choose appropriate visualizations [12]. Therefore, we decided to improve VisMillion by adding support to ordinal big data.

III. VisMillion

VisMillion was conceived to support the Streaming of Big Data. Pereira et al. [26] created a functional web-based prototype. Our solution uses the three.js library to create, manipulate, and display 3D or 2D elements. It is built on top of WebGL, a JavaScript API that accelerates graphic elements rendering in a web browser by shifting the drawing of the elements to the GPU.

A. Dataset generator

Finding appropriate Big Data datasets for Streaming scenarios is not easy. Since our focus was not yet to test our prototype with actual data, we used our dataset generator to create, send, and manipulate ordinal data using a python server script. For example, it allows the creation of a dataset whose values have a positive trend.

B. Elements

The visual representations of data displayed in the visualization are individual elements created by idioms and transitions that manipulate these representations to convey information as intended. All elements are instances of the three.js library, and the required elements were the following:

- **Dots:** Simple, small rectangular planes that represent individual data points.
- **Lines:** Rectangular plane, where its height and angle of rotation are given by two x and y coordinates. The line thickness can also be changed.
- **Rectangles:** Rectangular plane, with a single position. In this case, the rectangle's size or color can be changed. It is also possible to draw borders surrounding the rectangle, accomplished by creating lines for each border. This approach means the rectangles are comprised of five geometries instead of one.
- **Polygon:** This element is needed to create non-rectangle polygons, and it is possible through a buffer geometry by providing a list of vertices positions. This element allows more flexibility since it does not restrict the shape of the visual representation.

C. Visual Idioms

All idioms that we implemented shared an x-axis that encodes the data's timestamp, constantly updated in real-time (fig. 2). Therefore, the visual elements move towards the end of the visualization from the right (newer data) to the left (older data) of the visualization. The y-axis also encodes the ordinal value for the complete visualization.

1) Ordinal Scatter Plot: Shows every ordinal value received in real-time. The dots are restrained into invisible lines representing each current ordinal value. This idiom is intended to depict the arrival of data points and their distribution over time, as each data point is represented in one dot. However, since it applies no aggregation techniques, it designed only as the starting idiom using Graceful Degradation.

2) Heat Map: The length of each cell encodes a period, and the number of points inside the cell determines its color saturation. This idiom is suitable for various periods, but most suitable as a replacement for the Ordinal Scatter Plot.

3) Line Chart: We present, stacked vertically, one line chart for each possible ordinal value. Each data point in those line charts corresponds to the number of occurrences of that particular value over a certain time interval. This allows us to see the evolution of the distribution of values through time. The vertical scale of the line charts is adapted in real-time so that it can accommodate the values it has to depict at every

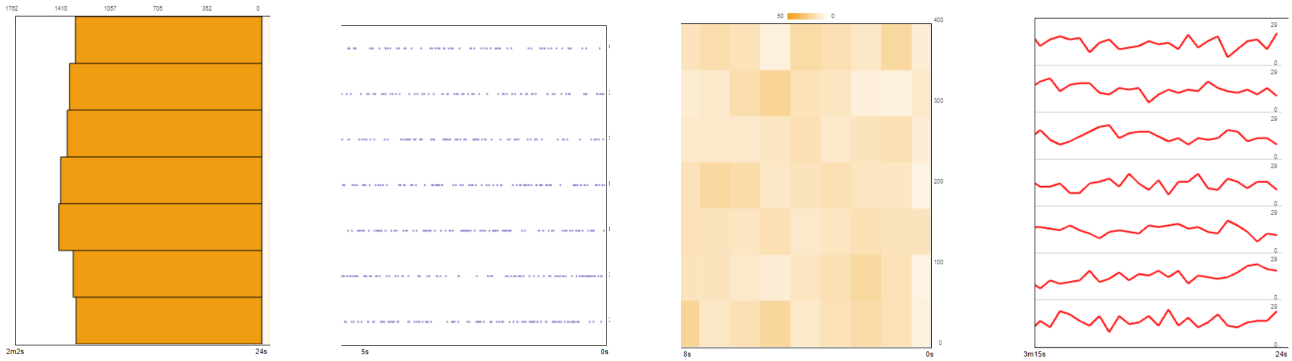

Fig. 2. Visual idioms used in our user study. From left to right, Histogram, Ordinal Scatter Plot, Heat Map, and Line Chart.

moment, taking care to use the same scale for all lines to facilitate comparisons. The idiom was designed to be used as the middle module.

4) Histogram: It worked as the final visual idiom of the data path because it contained information since the visualization initialization. Each bar represented the number of data points for one ordinal value, and the maximum length of the bars was updated to fit the available space in the prototype.

D. Transitions

The transitions referred to in this work are called **horizontal transitions**, which represent the continuity between two idioms, by transforming the elements that represent the data from one idiom to the following idiom's element properties [11], [34]. The goal is to perceptibly show that the information is the same in different aggregation levels. We designed several horizontal transitions, always considering the elements present in the origin and destination visual idioms.

1) Growing Bars: It was applied between the Ordinal Scatter Plot and Ordinal Line Chart (fig. 3). The dots converge on top of a bar in the left boundary of the transition. As a result, the bar will grow to its value in the ordinal line chart. This bar is attached to a line that connects to the following idiom's latest line, creating a seamless connection where there is always a line being produced and growing with each entering dot, giving the idea that they are being grouped by pilling up in a bar whose height is its number of dots.

2) Grouped Dots: It was applied between the Ordinal Scatter Plot and Ordinal Line Chart (fig. 4). The dots converge at the beginning of the respective line segment, giving the idea that the dots group together on the group's future on the ordinal line chart. In addition, the lines' opacity grows the closer they get to the following idiom.

3) Pilot Lines: It is applied between the Ordinal Scatter Plot and Histogram (fig. 5). By scaling the height to zero, the dots convey being merged into a single point. The ordinal values follow a line, and the scaling will almost be unperceivable. After the first third, a single rectangle is created with the dots' size in the following idiom. Once these rectangles hit the idiom boundary, they enlarge to a horizontal bar "pushed" to the following idiom.

4) Morphing: It was applied when the destination visual idiom was either a Heat Map or an Ordinal Line Chart (fig. 6). The elements that leave in the first idiom gradually transform into the elements present in the second. This idea could be done by changing color, opacity, size, and rotation. If more than one element was needed to represent the next element, the first would only morph into a second portion. The element morphing eases the visual transition, and the proportion clarifies how much data is being grouped.

5) Line Squeeze: It was applied between the Ordinal Line Chart and the Heat Map (fig. 7). In this transition, the lines will move towards the boundary while horizontally straightening. Then, after they reach their position in the Heat Map, they increase in height almost as if they were squeezed against the boundary. Multiple lines will likely represent one cell so that each line will depict just a portion.

6) Squares: It was applied when the starting visual idiom was the Heat Map (fig. 8). To understand the number of points inside the expelled element, it divides itself into smaller squares proportional to the total received points in that interval. If the ending idiom is the Ordinal Line Chart, the created squares will transform into small segments by rotating and gradually resizing themselves. If instead, the ending transition is a Histogram, then the squares will enter and increase the size of a bar, like being pushed to the following idiom.

7) Dissolving Lines: It begins in an ordinal line chart and ends in a Histogram, which has the same logic as the Squares transition for the same ending idiom (fig. 9). The only difference is that instead of splitting the element into squares, it "dissolves" the lines into segments during the transition.

8) Stacking Bars: If the ending idiom represents data through rectangles, then this transition is applied (fig. 10). Here, the starting elements transform into rectangles. After transforming, the resulting rectangles make their way to the end of the transition, where they stack on top of other rectangles, giving the idea that they are being aggregated into a single bar.

IV. User Study

To evaluate the effectiveness of the proposed visual idioms and transitions, we conducted a user study with 24 participants via questionnaires created with Google Forms. There were in total six combinations of visual idioms that we could test. Between the Ordinal Scatter Plot and the Histogram, or Ordinal Line Chart. Between the Heat Map and the Histogram, and Ordinal Line Chart. Finally, between the Ordinal Line Chart and the Histogram, and Heat Map. For each combination, we created a questionnaire, and each participant filled a questionnaire in an order generated using the Latin Square distribution.

Each questionnaire started by explaining both visual idioms used in the corresponding combination. After participants learned how each idiom worked, they had to answer several questions regarding the data presented in each one. Since each idiom is used for different tasks, the corresponding questions will be presented in the following sections.

Then, the transitions created for that combination were presented, and, again, participants had to answer several questions. All idioms and transitions were presented via recorded videos, thus ensuring that the data presented was always the same between different participants (please see this link [1] to watch the videos used, plus a demonstration of how VisMillion worked). Furthermore, it allowed them to rewatch the videos if they needed. Since it was not our goal to compare combinations between themselves, each questionnaire had its set of questions created according to the corresponding idioms and transitions presented. Finally, the data presented in the videos were randomly generated to fit the needs of the study, at 100 points per second. Again, since each transition conveys different information, the corresponding questions will also be presented in the following sections.

[1]https://bit.ly/3PtVsiR

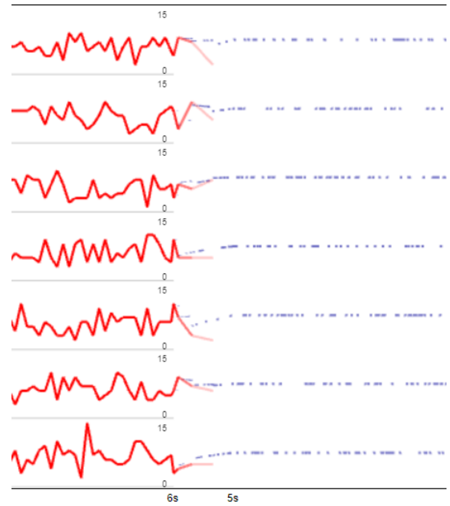

Fig. 4. Grouped Dots transition applied between the ordinal scatter plot and the line chart.

A. *Results – Visual Idioms*

For the Ordinal Scatter Plot, the video showed an almost binomial distribution on one of the ordinal values, and then suddenly, an agglomeration of points emerged on a distinct ordinal value. The first question tested if participants were able to identify which ordinal value had more data points. The second question tested if they could identify any ordinal value where the number of data points changed significantly. If they did, they indicated which ordinal value via a third question. Participants answered with 100% accuracy the first and second questions, and the third with 83% accuracy. Therefore, **the Ordinal Scatter Plot was an effective visual idiom**.

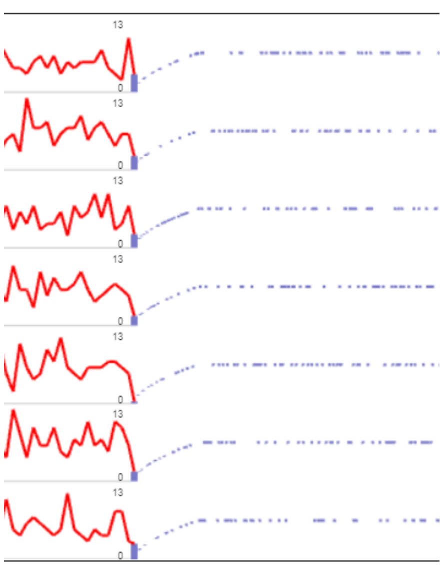

Fig. 3. Growing Bars transition applied between the ordinal scatter plot and the line chart.

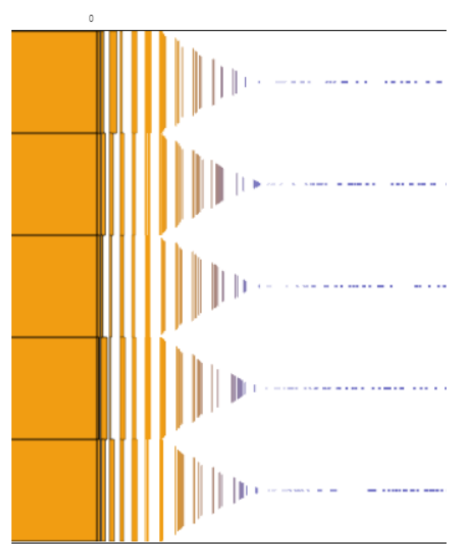

Fig. 5. Pilot Lines transition applied between the ordinal scatter plot and the histogram.

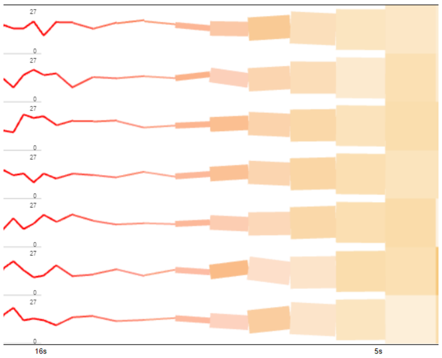

Fig. 6. Morphing transition applied between the heat map and the line chart.

The Heat Map's video followed the same logic as the Ordinal Scatter Plot, and the results were similar. 100% for the first question, 95.7% for the second, and 87.0% for the third. Therefore, **the Heat Map was an effective visual idiom**.

For the Ordinal Line Chart, only the first question was equal to the Ordinal Scatter Plot and Heat Map. The second question asked if they could detect if there were any ordinal values with a positive or negative trend. Then, the third question asked if participants could read the value encoded with the line. The first two questions had high accuracy values, 95.7%, and 78.3%. However, the third only had 56.5%. Therefore, **the Ordinal Line Chart was effective to convey information, except to decode the exact average presented**.

Finally, the Histogram's video demonstrated a simple binomial distribution of points around one ordinal value. The idiom was then tested with just two questions. In the first question, we asked participants which ordinal value had more data points, and in the second we asked the exact number encoded in one particular ordinal value. Accuracy was high for both questions, 100%, and 88.9% respectively. Therefore, **the Histogram was an effective idiom to convey information**.

B. Results - Transitions

The questions regarding the transitions always targeted each transition's aggregation, fluidity, and logic. Aggregation represented how well participants understood how the data shifted from one idiom to another. Fluidity represented how smooth participants thought the transition was. Finally, the logic represented what exactly was happening to the data between visual idioms. In the first question, participants had to answer with a Likert scale from 1 (Totally Disagree) to 5 (Completely Agree) to "The data points are being aggregated." In the second question, again with a Likert scale from 1 (Little Smooth) to 5 (Very Smooth), to "How smooth is the transition?". Finally, participants had to select one true sentence from a set of three. Each one explained how the transition worked, but only one was correct.

In some combinations, we proposed more than one transition (for example, between the Ordinal Scatter Plot and the Ordinal Line Chart). To understand if there were statistically significant differences (when two transitions were proposed) for aggregation and fluidity, each combination of transitions underwent a *Wilcoxon Signed-Rank Test* except for one transition that found a non-normal difference median distribution, in which we tested with the *Sign Test*. Then, for the logic, the question's result had to be tested with the *McNemar's Test*. However, no significant differences were found in any combination for the logic test.

1) From the Ordinal Scatter Plot: To the Ordinal Line Chart, we proposed two transitions. The Growing Bars had statistically significantly better results with $p < 0.0005$ in aggregation and fluidity. Growing Bars had 5 (1) on aggregation and fluidity. Grouped Dots had 4 (1) on aggregation and 3 (2) on fluidity. Therefore, **the Growing Bars transition was better than the Grouping Dots between the Ordinal Scatter Plot and the Ordinal Line Chart**.

2) From the Heat Map: To the Ordinal Line Chart, we proposed two transitions. Morphing had statistically significant better results with $p < 0.008$ in fluidity, but not on aggregation. Morphing had 4 (2) on fluidity, and Squares had 3 (1) on

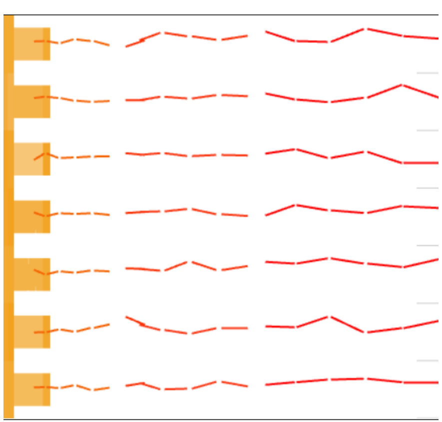

Fig. 7. Line Squeeze transition applied between the line chart and the heat map.

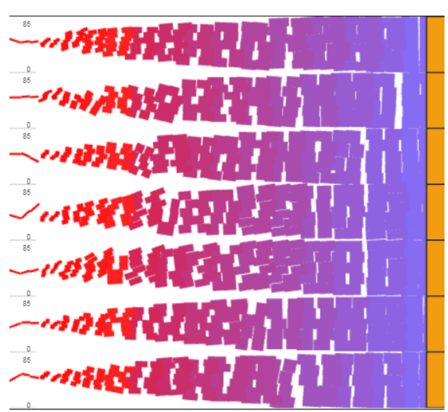

Fig. 8. Squares transition applied between the heat map and the line chart.

TABLE I
THE MOST SUITABLE TRANSITIONS. HEAT MAP (HM), LINE CHART (LC), HISTOGRAM (H), AND SCATTER PLOT (SP) COMBINATIONS.

	HM	LC	H
SP	-	Growing Bars	Pilot Lines
HM	-	Morphing	Stacking Bars
LC	Morphing	-	Stacking Bars

fluidity. Therefore, **Morphing was a more suitable transition between the Heat Map and the Ordinal Line Chart**. To the Histogram, The Stacking Bars had statistically significant better results with $p < 0.038$ in aggregation and $p < 0.005$ in fluidity. Stacking Bars had 5 (1) on aggregation and 4 (1) on fluidity. Squares had 4 (2) on aggregation and 4 (1) on fluidity. Therefore, **the Stacking Bars transition is a more suitable transition between the Heat Map and Histogram**.

3) From the Ordinal Line Chart: To the Heat Map, we proposed two transitions. The Morphing had statistically significant better results with $p < 0.002$ in aggregation and $p < 0.001$ in fluidity. Morphing had 4 (2) in aggregation and fluidity. Line Squeeze had 3 (2) in aggregation and 3 (1) in fluidity. Therefore, **the Morphing is a more suitable transition between the Ordinal Line Chart and Heat Map**. In the Histogram, we proposed two transitions, and there were no significant differences in aggregation and fluidity. The preference values decided that **the Stacking Bars is a more suitable transition between the Ordinal Line Chart and Histogram**.

C. Summary

The Ordinal Scatter Plot, Heat Map, Line Chart, and Histogram all showed positive results in the questionnaires, with most questions answered correctly by most of the users, proving to be good representations for ordinal data presented in VisMillion. The Line Chart had good results in understanding trends and for value comparison, yet the identification of the exact values returned poor results. The most suitable transitions of the suggested idiom combinations can be found in table I. From our work, we were able to define several design guidelines for Ordinal Big Data Streaming:

- The Ordinal Scatter Plot, Heat Map, Line Chart, and Histogram are effective idioms.
- Line Chart is not good to identify specific means;
- The Growing Bars transition should be used between the Ordinal Scatter Plot and the Line Chart;
- The Pilot Lines transition should be used between the Ordinal Scatter Plot and the Histogram;
- The Morphing transition should be used between the Heat Map and Line Chart and between the Line Chart and the Heat Map.
- The Stacking Bars transition should be used between the Heat Map and Histogram and between the Line Chart and the Histogram;

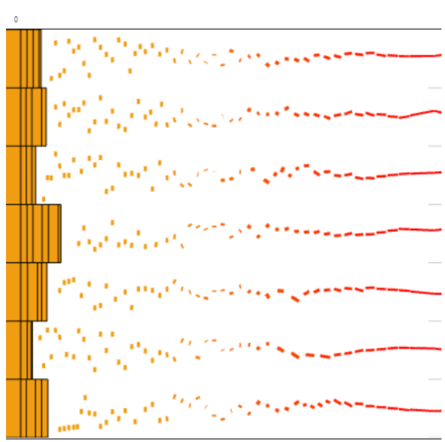

Fig. 9. Dissolving Lines transition applied between the line chart and the histogram.

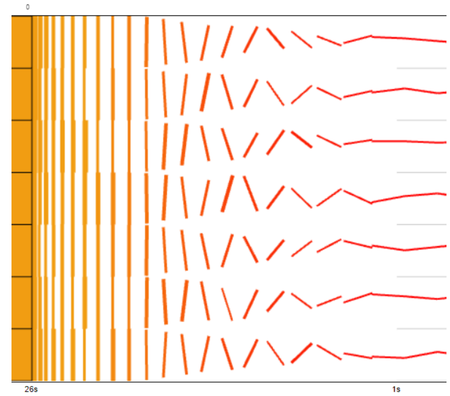

Fig. 10. Stacking Bars transition applied between the line chart and the histogram.

V. DISCUSSION AND FUTURE WORK

Our prototype was able to fulfill some major goals of Big Data Stream: real-time interaction, visual scalability, and personalization [5], [13], [17], [29]. It now supports horizontal transitions for ordinal and quantitative data, and vertical transitions for quantitative data, and allows customizing modules with different visual idioms. However, VisMillion does not yet support data processing and user assistance. Our data was generated by use to test certain scenarios in our user study.

Furthermore, participants were able to see effectively see data via the proposed visualizations, both in individual idioms and corresponding transitions, which is also important in the big data stream pipeline [13], [21], [29], [31]. Finally, VisMillion effectively conveyed information, giving the wanted overview these systems need to have [13], [29]. We believe the future for VisMillion is to integrate the missing components needed for Big Data Stream. First, real streamed datasets are usually dynamic [5], and their size leads to unprepared systems overflowing [5], [13], [15], [29]. Then, due to this dynamic restriction, VisMillion must be able to adapt according to the data it received to help users detect anomalies [10], [30].

VI. Conclusion

Big Data Streaming visualizations are currently a challenge in research. There is the need to develop systems that handle vast datasets in real-time, represent them without producing visual clutter, and allow people to overview data across time. VisMillion works as a solution by presenting information in different periods, each with a specific visual idiom, via Graceful Degradation, resulting in different levels of aggregation between them. We proposed visual idioms and transitions for the streaming of ordinal big data. We explored via a user study which visual idioms and transitions were the most effective, and defined guidelines for choosing combinations.

Acknowledgements

This work was partially supported by FCT through projects PTDC/CCI-CIF/28939/2017, SFRH/BD/143496/2019, and UIDB/50021/2020.

References

[1] N. E. aboudi and L. Benhlima. Big data management for healthcare systems: Architecture, requirements, and implementation. *Advances in Bioinformatics*, 2018:1–10, jun 2018.

[2] S. Ahmed, M. J. Islam, and H. Rajan. Semantics and anomaly preserving sampling strategy for large-scale time series data. *ACM/IMS Transactions on Data Science*, 2(4):1–25, nov 2021.

[3] M. A. Alves. Graceful degradation : Journey of a concept, from fault-tolerance to information loss. In *2021 16th Iberian Conference on Information Systems and Technologies (CISTI)*. IEEE, jun 2021.

[4] J. Angskun, S. Tipprasert, and T. Angskun. Big data analytics on social networks for real-time depression detection. *Journal of Big Data*, 9(1), may 2022.

[5] N. Bikakis. Big data visualization tools, 2018.

[6] F. Castanheira, J. Moreira, D. Mendes, and D. Goncalves. Evaluating transitions for streaming big data. In *2021 International Conference on Graphics and Interaction (ICGI)*. IEEE, nov 2021.

[7] G. Chawla, S. Bamal, and R. Khatana. Big data analytics for data visualization: Review of techniques. *International Journal of Computer Applications*, 182(21):37–40, oct 2018.

[8] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, sep 1984.

[9] A. Corallo, A. M. Crespino, M. Lazoi, and M. Lezzi. Model-based big data analytics-as-a-service framework in smart manufacturing: A case study. *Robotics and Computer-Integrated Manufacturing*, 76:102331, aug 2022.

[10] Y. P. Faniband, I. Ishak, and S. M. Sait. A review of open source software tools for time series analysis, 2022.

[11] T. Fujiwara, J.-K. Chou, Shilpika, P. Xu, L. Ren, and K.-L. Ma. An incremental dimensionality reduction method for visualizing streaming multidimensional data. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):418–428, jan 2020.

[12] M. Golfarelli and S. Rizzi. A model-driven approach to automate data visualization in big data analytics. *Information Visualization*, 19(1):24–47, jul 2019.

[13] F. Gurcan and M. Berigel. Real-time processing of big data streams: Lifecycle, tools, tasks, and challenges. In *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, oct 2018.

[14] Y. He and H. Li. Optimal layout of stacked graph for visualizing multidimensional financial time series data. *Information Visualization*, 21(1):63–73, sep 2021.

[15] A. C. Ikegwu, H. F. Nweke, C. V. Anikwe, U. R. Alo, and O. R. Okonkwo. Big data analytics for data-driven industry: a review of data sources, tools, challenges, solutions, and research directions. *Cluster Computing*, mar 2022.

[16] W. Jiang and J. Luo. Big data for traffic estimation and prediction: A survey of data and tools. *Applied System Innovation*, 5(1):23, feb 2022.

[17] T. Kolajo, O. Daramola, and A. Adebiyi. Big data stream analysis: a systematic literature review. *Journal of Big Data*, 6(1), jun 2019.

[18] A. Lavalle, M. A. Teruel, A. Maté, and J. Trujillo. Improving sustainability of smart cities through visualization techniques for big data from IoT devices. *Sustainability*, 12(14):5595, jul 2020.

[19] C. K. Leung, Y. Wen, C. Zhao, H. Zheng, F. Jiang, and A. Cuzzocrea. A visual data science solution for visualization and visual analytics of big sequential data. In *2021 25th International Conference Information Visualisation (IV)*. IEEE, jul 2021.

[20] C. K. Leung, Y. Zhang, C. S. Hoi, J. Souza, and B. H. Wodi. Big data analysis and services: Visualization on smart data to support healthcare analytics. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, jul 2019.

[21] A. Mohamed, M. K. Najafabadi, Y. B. Wah, E. A. K. Zaman, and R. Maskat. The state of the art and taxonomy of big data analytics: view from new big data framework. *Artificial Intelligence Review*, 53(2):989–1037, feb 2019.

[22] J. Moreira, F. Castanheira, D. Mendes, and D. Gonçalves. Designing animated transitions for dynamic streaming big data. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS - Science and Technology Publications, 2022.

[23] D. Paschalides, D. Stephanidis, A. Andreou, K. Orphanou, G. Pallis, M. D. Dikaiakos, and E. Markatos. MANDOLA. *ACM Transactions on Internet Technology*, 20(2):1–21, may 2020.

[24] R. Patgiri and A. Ahmed. Big data: The v's of the game changer paradigm. In *2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, dec 2016.

[25] T. Pereira, J. Moreira, D. Mendes, and D. Goncalves. Evaluating animated transitions between contiguous visualizations for streaming big data. In *2020 IEEE Visualization Conference (VIS)*. IEEE, oct 2020.

[26] G. Pires, D. Mendes, and D. Goncalves. VisMillion: A novel interactive visualization technique for real-time big data. In *2019 International Conference on Graphics and Interaction (ICGI)*. IEEE, nov 2019.

[27] J. Rafael, J. Moreira, D. Mendes, M. Alves, and D. Gonçalves. Graceful degradation for real-time visualization of streaming geospatial data. *EuroVis 2021 - Short Papers*, 2021.

[28] A. M. Rahmani, E. Azhir, S. Ali, M. Mohammadi, O. H. Ahmed, M. Y. Ghafour, S. H. Ahmed, and M. Hosseinzadeh. Artificial intelligence approaches and mechanisms for big data analytics: a systematic study. *PeerJ Computer Science*, 7:e488, apr 2021.

[29] T. R. Rao, P. Mitra, R. Bhatt, and A. Goswami. The big data system, components, tools, and technologies: a survey. *Knowledge and Information Systems*, 60(3):1165–1245, sep 2018.

[30] Z. Saeed, R. A. Abbasi, I. Razzak, O. Maqbool, A. Sadaf, and G. Xu. Enhanced heartbeat graph for emerging event detection on twitter using time series networks. *Expert Systems with Applications*, 136:115–132, dec 2019.

[31] J. Wang, Y. Yang, T. Wang, R. S. Sherratt, and J. Zhang. Big data service architecture: A survey. *Journal of Internet Technology*, 21(2):393–405, 2020.

[32] Q. Xu, L. Xiang, H. Wang, X. Guan, and H. Wu. GeoMapViz: a framework for distributed management and geospatial data visualization based on massive spatiotemporal data streams. *IOP Conference Series: Earth and Environmental Science*, 1004(1):012017, mar 2022.

[33] I. M. Yazici and M. S. Aktas. A novel visualization approach for data provenance. *Concurrency and Computation: Practice and Experience*, 34(9), jul 2021.

[34] Y. Zheng, W. Wu, N. Cao, H. Qu, and L. M. Ni. Focus+context grouping for animated transitions. *Journal of Visual Languages & Computing*, 48:61–69, oct 2018.