Summary Document for the Bulletin Video Entitled:

Creating Storytelling Visualizations for the COVID-19 Pandemic Using Feature-Action Design Patterns

Panagiotis D. Ritsos* — Bangor University, UK
Saiful Khan — STFC, UK
Scott Jones — Imperial College London, UK
Benjamin Bach — University of Edinburgh, UK
Julie Meikle — University of Oxford, UK
Jonathan C. Roberts — Bangor University, UK
Jo Wood — City University, UK
Min Chen — University of Oxford, UK

In this bulletin video, we summarize a novel technique for authoring storytelling visualization [3–5]. The technique was developed by one of the teams in the RAMPVIS project [1, 2], which provided visualization support to epidemiological modeling during the COVID-19 pandemic. The team explored the prevailing approaches, in the UK and internationally, for creating public-facing visualizations related to the pandemic. This ranged from those produced by a number of governments (e.g., the four home nations in the UK), organizations (e.g., WHO, UK ONS), universities (e.g., Johns Hopkins dashboards), media outlets (e.g., FT Coronavirus tracker), and non-commercial web services (e.g., Worldometers). The team concluded that we should complement, but not duplicate, the existing effort, and defined our goal as to inform the public through advanced storytelling visualization [1].

The commonly-adopted authoring approach for creating storytelling visualization requires the author of a storyboard to gain a good understanding of the data to be visualized in the story. We soon realized that this approach would not be scalable as there were many datasets for individual regions and these datasets were changing on a daily basis. The author of a storyboard might study and understand the data of one region, but could not do so for every region. The author might understand the data for a specific period, but it would be burdensome to study the data and update the story every day.

This led to the development of a new technique for authoring storytelling visualization, which are briefly described in the following sections (Figure 1(b)).

1 Meta-Authoring

The process of meta-authoring requires a different approach to traditional storytelling; authors need to think abstractly. With traditional visual storytelling, a person would explore the data in a visualization tool, save states as storyboards, which would be played. However, with meta-authoring the author needs to (1) explore the data, (2) turn specific story items into generalized story *features*, and (3) encode related *actions*. With time-series data, such as those encountered during the pandemic, such features can be peaks, dips, rising and falling segments, local max/min, etc. The corresponding actions can be defined by the creator; in our use cases we highlighted or labeled features (e.g., circle peaks, color segments, etc.), provided text messages via annotations or adjacency, or initiated animations.

2 Algorithmic Pipeline

We consider two categories of time series data: a) *Numerical time series* (NTS), which includes many commonly-encountered time

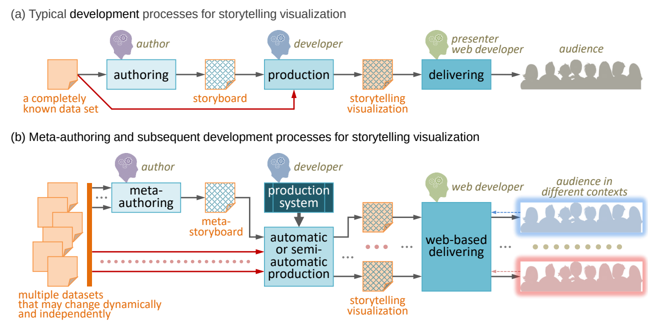

Figure 1: Comparison of two storytelling paths. At the top (a), a typical path where the author creates a storyboard based on a known dataset, which is then developed as a web-based visualization, usually for a specific target audience. At the bottom (b), our approach where the author creates a meta-storyboard that works with multiple, dynamic and often partially unknown datasets. The meta-storyboards are then converted by a developer following a set of rules which facilitate the automatic or semi-automatic depiction of user-driven, web-based stories for different target audiences.

series during the COVID-19 pandemic, such as daily, accumulated, normalized, and k-day moving average data, with semantics such as number of cases, hospitalizations, fatalities, vaccinations, etc., and, b) *Categorical time series* (CTS), which extends the notion of NTS by considering each data point at time t can have a categorical or nominal value. An example of a categorical time series, may feature major healthcare policy changes, such as 'lockdowns' or vaccination programme starts etc. Combinations of time series data of these two categories are processed as follows:

a) Features such as peaks, falls, rising and falling segments of the time series etc. are detected, based on a story author's definition.
b) Event features in the categorical time series are typically predefined and ranked (e.g., vaccination programme start ranked higher than an announcement for a major healthcare policy change) by the story author.
c) Each detected data feature is given a rank and each ranking value is converted to a Gaussian curve. Likewise, ranked semantic events are converted individually to Gaussian curves.
d) Gaussian curves are combined using Gaussian mixture models (e.g., a max-model is used within a time series and a mean-model is used between time series).
e) The story is divided in segments according to the combined importance curve. The number of segments is defined by the story author.

Subsequently, a technical developer translates the specifications to feature-action data patterns in a lookup table, where *features* are categorical labels of all features that might be detected in the NTS and CTS that users can select. Corresponding *actions* are the

*E-mail of the corresponding author: p.ritsos@bangor.ac.uk

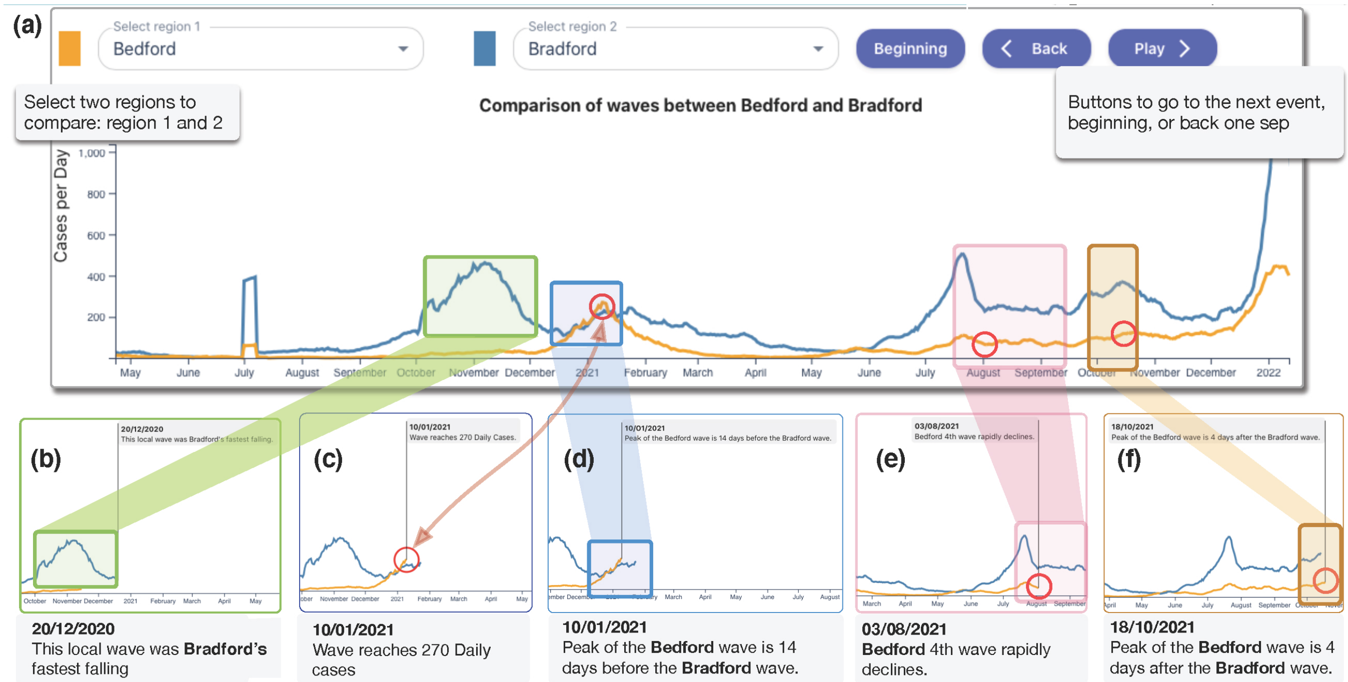

Figure 2: Comparison story demonstration, where (a) depicts the final frame. The story is shown in stages, moving key 'features', and alternating 'actions' between region 1 and 2. The insets (b-f) depict several key event features, which are incrementally shown as the story progresses; (b) a single feature and action about Bradford (region 2); (c) story action focusing on peaks, with data specific to the local site; (d and f) comparison feature showing differences in terms of days; (e) feature comparison based on calculated data.

programming references of the software components or events for displaying various visual artifacts, such as data, axes, highlights, texts, etc. The story's length is controlled by how many segments the Gaussian curve is split into and how many features there are in the lookup table.

3 Story Example

We have created a number of storyboards using our approach, to create stories for the COVID-19 pandemic. For example, a comparison story (Figure 2) allows viewers to compare epidemic waves in Bedford (100,000 inhabitants, 50 miles from London) and Bradford (350,000 inhabitants, 150 miles North of Bedford), both in England, under the same COVID-19 restrictions and reasonably close; yet on different rail routes to London and distant enough to have separate ecosystems. Investigating them could indicate if the pandemic was moving towards or away from London.

In this example each line-graph corresponds to a location. Because there are two time series, and the comparison between both is the key narrative element, the author chose to alternate between features in each timeline. Features to be detected need to be relevant to the characteristic of a wave; e.g., constant case increase over a 14 day period. Messaging actions for this story have two main goals: a) to highlight features on each time series, and b) to highlight comparisons between the time series. In this context, we created two story variants, revolving around a comparison between the regional vs regional and regional vs national profile of the wave, such as peak dates, and rise and fall rates.

4 Conclusion

With this bulletin video, we present a new method for creating storytelling visualizations, addressing challenges of creating generic storyboards that can be applied to data streams of different regions, and can respond to different (data) features automatically using different actions. We have investigated the feasibility of this new method through storyboards developed in the COVID-19 context. We are in the process of submitting a paper based on this work to a journal, and we will make a pre-print version available on arXiv.

Acknowledgments

We thank all other VIS volunteers for their contributions to various RAMPVIS activities since May 2020. This work was supported in part by the UKRI/EPSRC grant EP/V054236/1.

References

[1] M. Chen, A. Abdul-Rahman, D. Archambault, J. Dykes, A. Slingsby, P. D. Ritsos, T. Torsney-Weir, C. Turkay, B. Bach, A. Brett, H. Fang, R. Jianu, S. Khan, R. S. Laramee, L. Matthews, P. H. Nguyen, R. Reeve, J. C. Roberts, F. Vidal, Q. Wang, J. Wood, and K. Xu. RAMPVIS: Answering the challenges of building visualisation capabilities for large-scale emergency responses. *Epidemics*, 39:100569, 2022. doi: 10.1016/j.epidem.2022.100569

[2] J. Dykes, A. Abdul-Rahman, D. Archambault, B. Bach, R. Borgo, M. Chen, J. Enright, H. Fang, E. E. Firat, E. Freeman, T. Gönen, C. Harris, R. Jianu, N. W. John, S. Khan, A. Lahiff, R. S. Laramee, L. Matthews, S. Mohr, P. H. Nguyen, A. A. M. Rahat, R. Reeve, P. D. Ritsos, J. C. Roberts, A. Slingsby, B. Swallow, T. Torsney-Weir, C. Turkay, R. Turner, F. P. Vidal, Q. Wang, J. Wood, and K. Xu. Visualization for epidemiological modelling: Challenges, solutions, reflections & recommendations. *Philosophical Transactions of the Royal Society A*, 380:2233, 2022. doi: 10.1098/rsta.2021.0299

[3] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013. doi: 10.1109/MC.2013.36

[4] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics & Applications*, 35(5):84–90, 2015. doi: 10.1109/MCG.2015.99

[5] C. Tong, R. C. Roberts, R. Borgo, S. P. Walton, R. S. Laramee, K. Wegba, A. Lu, Y. Wang, H. Qu, Q. Luo, and X. Ma. Storytelling and Visualization: An Extended Survey. *Information*, 9:65, 2018. doi: 10.3390/info9030065