



HAL
open science

Navigating Multi-Attribute Spatial Data Through Layer Toggling and Visibility-Preserving Lenses

Karelia Salinas, Jean-Daniel Fekete, Luis Gustavo Nonato

► To cite this version:

Karelia Salinas, Jean-Daniel Fekete, Luis Gustavo Nonato. Navigating Multi-Attribute Spatial Data Through Layer Toggling and Visibility-Preserving Lenses. Posters of VIS 2024 - IEEE Visualization, Oct 2024, Virtual, United States. hal-04744681

HAL Id: hal-04744681

<https://inria.hal.science/hal-04744681v1>

Submitted on 19 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Navigating Multi-Attribute Spatial Data Through Layer Toggling and Visibility-Preserving Lenses

Karelia Salinas*
ICMC, University of São Paulo

Jean-Daniel Fekete†
Université Paris-Saclay, CNRS, Inria, LISN

Luis Gustavo Nonato‡
ICMC, University of São Paulo

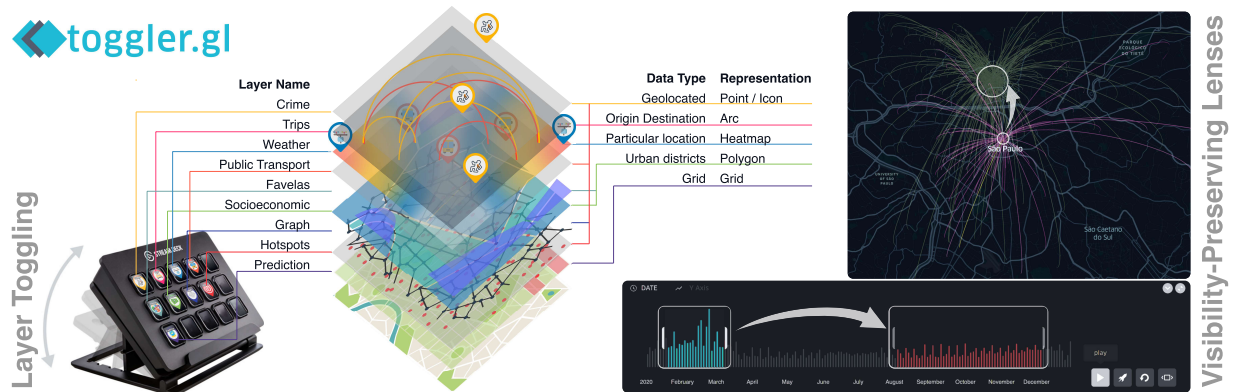


Figure 1: City databases are rendered into spatially aligned layers overlaid to compose a view to evaluate data used in model-based predictions (layer #9) of crimes. Each layer visualizes a specific type of spatial data using a dense representation. Each of the nine layers is assigned to a button in the button box to interactively toggle its visibility on the screen to explore the spatial correlation between layers, relying on our retinal persistence. On the top-right, the Visibility-Preserving Lens dynamically adjusts the brush radius based on the graph link density at the mouse position to control the graph legibility. For 1D attribute filtering, such as temporal value range sliders, the brush size dynamically adapts to the local distribution also to control the legibility. Animation can be used to sweep over the whole range with controlled legibility. The results of the prediction can thus be compared to all the data layers and assessed locally and globally.

ABSTRACT

This work introduces two interaction techniques for examining multi-attribute spatial data for visual analytic applications: *Layer Toggling* and *Visibility-Preserving Lenses*. With the rise of open city data and services, gathering information at the city scale, like transportation, modeling, and predicting city events such as crime and traffic, has become feasible. Visually assessing the quality of these models requires correlating features, predictions, and ground truth both globally and locally, which is challenging with multi-attribute data due to occlusion. To address this, we introduce *Layer Toggling* for instantaneously changing layer visibility using a physical button box, allowing comparison of spatially coherent views using retinal persistence. *Visibility-preserving lenses* dynamically adjust to the density of revealed features, facilitating the exploration of spatial (2D) and quantitative data (1D), such as temporal attributes. We validate our approach on a use case visualizing urban data from São Paulo city across multiple data layers. Our methods support user exploration and expert analysis tasks, especially in validating and interpreting prediction algorithm outcomes at both global and local scales.

Index Terms: Multi-attribute spatial data, Spatial data, Toggling interaction technique, Spatial data exploration, Model validation, Visual analytics

1 MOTIVATION

Urban data is a clear example of multi-attribute data, utilized more and more for predicting various phenomena such as accidents, crime, and mobs. To anticipate these phenomena, various predictive models have been put forth [3, 4, 1]. These models can be assessed along three dimensions: the success rate (How Much?), the reasons behind their success (Why?), and the specific locations where they

perform well (Where?). Machine learning accuracy metrics address the “How Much” aspect, while explainability techniques tackle the “Why” question. To address the “Where” question, we leverage 2D spatial data visualization of many attributes and datasets. Additionally, predictive models rely not only on historical data but also integrate urban variables such as socioeconomic data, points of interest, and climate. In the context of machine learning, such integration is handled by the assembly of large feature vectors. From a visualization standpoint, the integration can be structured as a set of overlapping layers, each associated with a specific data type (e.g., origin-destination graphs for traffic, heatmaps for crime and real-estate value, vector data for road maps, etc.)

Traditionally, visual analytic applications opt for space multiplexing to deal with the complexity of the problem, dividing the screen into multiple sections [2, 4, 6] to show multiple maps. This practice limits the resolution of each attribute/dataset, making it challenging to observe spatial correlations between features. In contrast, employing the entire screen space to showcase detailed information allows for the perception of both local and global correlations among features. However, this approach poses additional challenges, particularly concerning the accumulation of data within the same spatial framework, leading to visual clutter and occlusion that impede effective analysis.

2 INTERACTIVE TIME-MULTIPLEXING

In this work, we propose a visualization methodology where each layer occupies the entire screen space, avoiding spatial multiplexing to maintain a good spatial resolution. The multi-attributes are handled by efficiently selecting the visibility and order of layers, using dynamic lenses for filtering items while making the layers easily readable.

Our use case comprises several steps. Initially, we collect data using the CityHub library [6]. Subsequently, we applied a classification algorithm in Python, specifically XGBoost with Random Undersampling, achieving a Gmean performance of 0.89. To visually present databases and results, we employ the `kepler.gl`

*e-mail: kareliavs17@gmail.com

†e-mail: Jean-Daniel.Fekete@inria.fr

‡e-mail: gnonato@icmc.usp.br

library [5], but the approach is applicable to other platforms that can enhance scalability. Our visualization setup incorporates customized filtering and navigation functions:

- **Layer Toggling:** A set of databases, each containing a unique data type and a representation, is handled as layers (Figure 2). The layers fill the screen and can overlap, displaying simultaneously. Clutter is managed by adjusting map overlay and layer blending options (normal, additive, subtractive). To switch seamlessly between layers, we developed *Layer Toggling*, which enables an easy comparison of features. Three versions of manipulation exist: mouse, keyboard, and button box. The button box Elgato Stream Deck Classic was chosen for its superior navigation. It features 15 customizable LCD keys and an adjustable stand, offering efficient layer control (see left of Figure 1). Each key corresponds to a layer, with an icon for easy identification.

- **Visibility-Preserving Lenses** We’ve implemented two dynamic filtering lenses that are adaptable based on an initial parameter, either spatial or temporal.

Spatial: A brush lens that highlights groups of points.

The lens size is adjusted based on point density. Inspired by [7], we insert the points into a QuadTree, in geolocated layers. The KNN algorithm uses the cursor position and a fixed number to identify the nearest points. We then measure the distance to the farthest point from the cursor and set this as the brush radius.

Temporal: Our system incorporates a folding window with a histogram for analyzing temporal data. We’ve integrated dynamic adjustment of animation intervals based on density. This ensures consistent data representation on the screen in each frame of the animation. Additionally, the animation reverses direction at the end of the histogram to maintain continuity.

In summary, our contributions include:

- Proposing a novel analysis method involving data organization into layers. The tendency is to aggregate data, but subdividing it into layers yields a deeper understanding of datasets.
- Introducing new techniques such as Layer Toggling navigation and dynamic filtering of spatial and temporal data.
- Visually evaluate the performance of a predictive model, detecting where and why it fails or succeeds.

3 RESULTS

We conducted a preliminary usability study on a use case to assess the effectiveness of our tool for crime analysis in São Paulo, involving ten graduate students. Key highlights included its utility, layered design (90% agreement for its ability to overlay and deactivate data), and interactivity. We evaluated three “Layer Toggling” devices, with the Stream Deck emerging as the preferred choice across all categories. The technique was appreciated notably in terms of ease of use, novelty, ergonomics, avoiding eye fixation, and reliance on muscular memory. The mouse ranked second, while the keyboard was least preferred. They rated our methodology as satisfactory, with six students reaching the maximum score.

Additionally, we compared our dynamic implementations of visibility-preserving lenses with the original static version. The dynamic spatial brush outperformed the static one due to its flexibility and ease of use when comparing densely populated areas with sparser ones. Although the histogram with animation provided a smoother transition and novelty, it was rated equally in ease of use compared to the static version. This was attributed to the additional instructional requirement to understand the window’s adaptability.

Regarding the layer evaluating prediction algorithms, it was visually detected that the majority of errors were located in the city’s peripheries. This can be attributed to the unique characteristics of isolated areas and the proportion of these cases.

The final comments primarily focused on making the auxiliary device more accessible and enhancing its applicability to other urban problems. Adaptability to various spatial issues is feasible, including dimensionality reduction, clustering, and cartography.

Layer name	Visualization	Icon	Representation	Data type	S/D Task	User
Crime Point			Point	Geolocated	Dynamic	Normal User
Trips Arc			Arc	Origin Destination		
Weather Heatmap			Heatmap	Particular Location		
Public Transport Icon			Icon	Geolocated	Static	Exploration Normal User
Favelas Geoson			Polygon	Urban Districts		
Socioeconomic Geoson						
Graph dead end Point			Point	Geolocated	Analysis Analyst	Analyst
Hotspots Point						
Prediction Grid			Grid	Grid		

Figure 2: This table lists all layers under consideration, each from a distinct database. The first column shows the interface buttons, the second displays the map, and the third column showcases the icon representing each layer in the button box. The fourth column illustrates the data representation based on available types in Kepler, such as point, arc, etc. The fifth column denotes the type of data contained in each database, such as geolocated, Origin Destination, etc. Moving on, the sixth column classifies the data as dynamic or static, depending on update frequency and temporal variability. The last two columns outline the tasks associated with each layer and the user’s interaction with them.

ACKNOWLEDGMENTS

This work was supported by FAPESP (#2020/07012-8, #2022/09091-8) and CNPq (#307184/2021-8). The opinions, hypotheses, conclusions and recommendations expressed in this material are the responsibility of the authors and do not necessarily reflect the views of FAPESP and CNPq.

REFERENCES

- [1] N. Ferreira, J. Poco, H. T. Vo, J. Freire, and C. T. Silva. Visual exploration of big spatio-temporal urban data: A study of new york city taxi trips. *IEEE transactions on visualization and computer graphics*, 19(12):2149–2158, 2013. 1
- [2] R. Gao, H. Tao, H. Chen, W. Wang, and J. Zhang. Multi-view display coordinated visualization design for crime solving analysis: Vast challenge 2014: Honorable mention for effective use of coordinated visualizations. In *2014 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 321–322. IEEE, 2014. 1
- [3] G. Garcia, J. Silveira, J. Poco, A. Paiva, M. B. Nery, C. T. Silva, S. Adorno, and L. G. Nonato. Crimanalyzer: Understanding crime patterns in são paulo. *IEEE Transactions on Visualization and Computer Graphics*, 27(4):2313–2328, 2019. 1
- [4] G. Garcia-Zanabria, M. M. Raimundo, J. Poco, M. B. Nery, C. T. Silva, S. F. A. de Abreu, and L. G. Nonato. Cripav: Street-level crime patterns analysis and visualization. *IEEE Trans. Vis. Comp. Graph.*, 2021. 1
- [5] S. He. From beautiful maps to actionable insights: introducing kepler.gl, uber’s open-source geospatial toolbox. *Uber Eng*, 29, 2018. 2
- [6] K. Salinas, T. Gonçalves, V. Barella, T. Vieira, and L. G. Nonato. City-hub: A library for urban data integration. In *2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, vol. 1, pp. 43–48. IEEE, 2022. 1
- [7] B. Vad. k-nearest-neighbor search using d3 quadtrees. <https://observablehq.com/@11b411/k-nearest-neighbor-search-using-d3-quadtrees>, 2024. Accessed: February 23, 2024. 2