Visualizing citizens' security level from open data

I. Extended abstract

Over the past decades, the intensification of digital transformation has expanded the capacity to collect, integrate, and analyze massive volumes of data from heterogeneous sources. This technological evolution has had implications for social security systems, which can be defined as the ensemble of policies and programs designed to ensure citizens' welfare and protection. Among the various challenges in this domain, in this work, we focus on the translation of complex datasets into insights to support timely and informed decision-making. The convergence of Business Intelligence (BI) tools and data visualization techniques can support the governance of urban safety, particularly in environmental and road security.

This study focuses on the application of BI platforms to publicly available datasets concerning air quality and traffic accidents, aiming to promote safer, more sustainable environments. The work builds on the methodological foundations of visual analytics [1] and user experience for visual decision support [2], and proposes interactive dashboards and visual exploration to analyze multifaceted data spaces. The amount of data might involve progressive data abstraction and multi-level aggregation, enabling both expert and non-expert users to detect hidden patterns, anomalies, and temporal-spatial correlations that may otherwise remain obscured in raw tabular data.

One of the goals of this contribution is to allow people access to data-driven insights through visualization interfaces that are characterized by clarity, interpretability, and responsiveness. The results obtained from the analysis of urban air pollution and traffic accident data, extracted from open government repositories, reveal how visual narratives can indicate relationships between environmental parameters and mobility risks. For example, temporal synchronizations between peaks in NO_2 levels and increased road accident rates highlight the potential of cross-domain data integration for preventive intervention design. By combining visualization strategies with BI practices, and more specifically, by using dashboards [3], [4], urban safety monitoring can lead to more accessible information, also to the citizens.

The study also promotes the adoption of visual analytics as a component of smart city infrastructures, because the interpretation of complex, real-time data can provide timely information that can improve the quality of governance. On the other hand, data visualization in civic contexts can benefit from such results, particularly when employed not just as a means of presentation but as an instrument for exploration [5], [6], sensemaking, and participatory decision-making.

Data visualization supports cognitive processes and facilitates interaction with large and complex datasets by extending human working memory. Indeed, visual encoding reduces cognitive effort by transforming abstract or symbolic data into spatial representations that can be more manageable and intuitive to interpret.

Through visualizations, a large volume of information can be compressed into a limited space, which is typical when dealing with multidimensional data or time-dependent phenomena. Through such transformations, it becomes possible to see trends, anomalies, and relationships that are not immediately apparent in raw, textual formats.

A further aspect is the capacity of visualization to show patterns. When information is organized visually, especially in ways that reflect temporal, spatial, or categorical associations, underlying structures often emerge with greater clarity. These patterns support both exploratory and confirmatory data analysis by offering visual cues that can guide interpretation.

Unlike static diagrams, interactive visualization systems also support manipulation and exploration. They enable users to engage dynamically with data by adjusting parameters, filtering subsets, or switching perspectives. This interactivity makes visualization not only a medium of communication but also a query instrument.

According to Tominski and Schumann [7], an effective visualization must accurately represent the data, allow users to meet their analytical goals, and offer a favorable cost-benefit balance in terms of time and cognitive effort.

Temporal data visualization poses distinct challenges within data analysis. Unlike static or purely categorical information, time-oriented data requires an understanding of temporal structures and relationships and also the careful design of visual representations that must faithfully reflect their dynamics and variability.

Aigner et al. propose a structured approach to analyzing and visualizing data with temporal components [8]. They emphasize the importance of aligning temporal semantics with visual encodings that respect both the granularity and the context of the data.

This paper considers two datasets to demonstrate how temporal visualization can be applied to real-world data. These two datasets are only a part of the available open data that can be downloaded and analyzed, but they already contain interesting information and reveal patterns that can be useful for future decisions.

The first dataset involves air quality monitoring data collected by ARPA Puglia [9]. This case illustrates the importance of mapping continuous environmental phenomena over time, using graphical techniques that highlight periodic trends and critical thresholds. The data, collected through a network of

distributed monitoring stations, contains temporally segmented measurements of atmospheric pollutants associated with specific locations and time intervals. These observations offer a view of the spatial and temporal variability in air quality across various sites in the Apulia region.

Among the monitored pollutants, three were selected for in-depth analysis due to their relevance for public health and urban environmental policy. The first is benzene (C_6H_6), a volatile organic compound (VOC) commonly found in gasoline, cigarette smoke, and industrial emissions. Benzene is classified as a carcinogen and poses risks to both the nervous and reproductive systems. Its presence in the atmosphere is closely linked to vehicular traffic and industrial activities.

The second pollutant is particulate matter with a diameter of less than 10 micrometers (PM_{10}). These airborne particles originate from a combination of combustion processes (such as vehicle emissions, residential heating, and industrial output), mechanical disruptions (like construction and road dust), and natural sources (including wind-blown soil and agricultural activity). PM_{10} particles are particularly concerning due to their ability to penetrate the respiratory system and contribute to chronic health issues.

The third compound analyzed is ground-level ozone (O_3). Although ozone exists naturally in the atmosphere, its concentration at lower altitudes can rise significantly due to anthropogenic pollutants, especially in urban environments. High levels of ozone can irritate the respiratory tract and exacerbate conditions such as asthma, posing heightened risks for children, the elderly, and individuals with preexisting respiratory conditions.

By mapping fluctuations in pollutant concentration over time and across locations, it is possible to identify recurring patterns, episodic anomalies, and seasonal cycles.

The second dataset concerns traffic accident data from the Bari metropolitan area, with the goal of revealing temporal distributions and patterns related to the occurrence of accidents. This dataset documents road traffic accidents that occurred within the city during the calendar year 2023. The temporal coverage spans from January 1 to December 31, offering an annual view of traffic-related accidents across the urban territory. The organizational structure proposed by Aigner et al. was used as a guide in selecting appropriate visualization methods, such as time-series plots, calendar views, or stratified temporal aggregations, depending on the nature of the events and the required level of temporal granularity.

This dataset was selected for its relevance in assessing urban safety and understanding the dynamics of vehicular accidents in a densely populated city. It enables explorations of spatial and temporal patterns in accident occurrence, providing support for evaluating the effectiveness of municipal traffic policies and identifying areas requiring targeted interventions.

By integrating this data with environmental and infrastructural variables, the analysis contributes to a broader framework for urban safety optimization and policy evaluation.

Fig. 1 shows a heatmap of accidents grouped by week, each week divided by hour of the day so that it is possible to

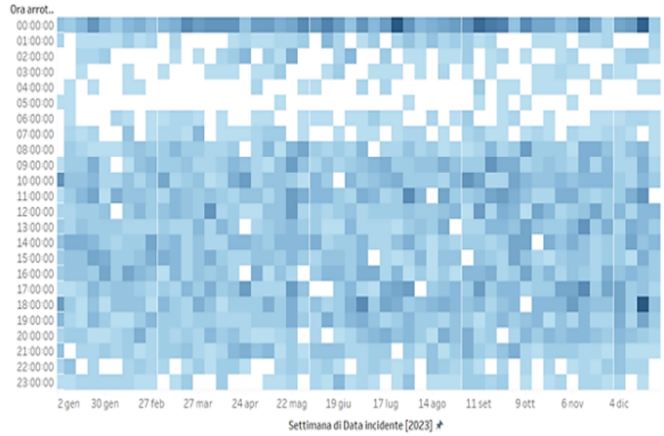

Fig. 1. Heatmap of accidents by week and hour of the day with hourly granularity

compare the different moments in the day, weekly. The color gradient indicates the number of accidents, which varies from zero (white) to 16 (dark blue). The interval from 1:00 to 7:00 a.m. contains many zero values, while the higher presence of accidents is from 8:00 a.m. to 7:00 p.m. There are a few outliers: one is at 6:00 p.m. right before Christmas, and another visible pattern is that many accidents occur from 0:00 to 1:00 a.m. While one would think that the first hour of the day is the most dangerous, a second investigation revealed that sometimes the accidents are not recorded with the time when they occurred but only with the date, so it is assumed by the system that they occurred at 00:00 a.m. Ignoring for a while this aspect, it is useful to see the overview of one year of accidents and see the patterns that appear

References

[1] Daniel Keim, Jörn Kohlhammer, Glenn Ellis, and Florian Mansmann, editors. *Mastering the Information Age: Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[2] Robert Amar and John Stasko. Knowledge precepts for design and evaluation of information visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 11(4):432–442, 2005.

[3] Stephen Few. *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly Media, 2006.

[4] Rob Kitchin, Tracey Lauriault, and Gavin McArdle. Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science*, 2(1):6–28, 2015.

[5] Jason Dykes, Alan M MacEachren, and Menno-Jan Kraak, editors. *Exploring Geovisualization*. Elsevier, 2010.

[6] Alan M MacEachren, Anthony Robinson, Steve Hopper, Susan Gardner, Ryan Murray, Mark Gahegan, and Elizabeth Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.

[7] Christian Tominski and Heidrun Schumann. *Interactive Visual Data Analysis*. CRC Press, 2020.

[8] Wolfgang Aigner, Silvia Miksch, Heidrun Schumann, and Christian Tominski. *Visualization of Time-Oriented Data*. Springer, second edition edition, 2023.

[9] ARPA Puglia. Air quality monitoring data, 2023. Accessed 2024-12-01.