Technical Section

Bi-Scale density-plot enhancement based on variance-aware filter

Huaiwei Bao [a], Xin Chen [a], Kecheng Lu [b], Chi-Wing Fu [c], Jean-Daniel Fekete [d], Yunhai Wang [b,*]

[a] *Shandong University, China*
[b] *Renmin University of China, China*
[c] *Chinese University of Hong Kong, China*
[d] *University Paris-Saclay, CNRS, Inria, LISN, France*

ARTICLE INFO

ABSTRACT

We present Bi-Scale density Plot (BSP), a new technique to enhance density plots by efficiently optimizing the local density variance in high- and mid-density regions while providing more details in low-density regions. When visualizing large and dense discrete point samples, scatterplots and thematic maps are often employed and we need density plots to further provide aggregated views. However, in the density plots, local patterns such as outliers can be filtered out and meaningful structures such as local density variations can be broken down. The key innovations in BSP include (i) the unified bin–summarize–decompose–combine framework for interactively bi-scale enhancing density plots through combining large- and small-scale density variations; and (ii) the variance-aware filter, which is reformulated based on the edge-preserving image filter, for maintaining the relative data density while reducing the excessive variability in the density plot. Further, BSP can be adopted with a 2D colormap, allowing simultaneous exploration of the enhanced structures and recovering the absolute aggregated densities to improve comparison and lookup tasks. We empirically evaluate our techniques in a controlled study and present two case studies to demonstrate their effectiveness in exploring large data.

1. Introduction

Scatterplots and thematic maps are among the most effective techniques for visualizing discrete data points in 2D; it is one of the most fundamental and important visualization techniques. Yet, for large and dense data, overplotting leads to visual clutter and obscures the data distribution (Fig. 1(a)). So, another representation can be used: the *2D density plot* (also called *density map* and *heatmap*, we use the term *density plot*). Instead of visualizing each point separately like a scatterplot or a thematic map would, a density plot counts the number of points falling inside a pixel or a small area (called a *bin*) and visually encodes the count using intensity.

Density plots fully resolve the overplotting issue by converting a scatterplot into a density field and presenting the density field via a colormap. We refer to such density plots as continuous density plots (CDP). Such density plots can effectively show global patterns (e.g., trends and clusters), especially the ones in high-density areas. However, they often miss local patterns such as outliers in nearly empty areas (see the red box in Fig. 1(b)) and can hide important visual structures (called "structures" in the remaining), i.e., meaningful local density variations (see the blue box in Fig. 1(b)). This is because of the limited intensity resolution of existing displays and of our vision system; the

highest density is mapped to the highest (or lowest) visible intensity, saturating the colormap and compressing the intensity variation of mid- and low-densities, making them less distinguishable.

A few techniques have attempted to address some of these issues. Splatterplots (SPP) [1] and Sunspot plots (SSP) [2] show outliers by overlaying discrete points in sparse regions onto the density plot, but the high-density regions still saturate the color map. For example, the road structures enclosed by the blue box on the top right of Figs. 1(c), 1(d) are not clearly visible.

Another common approach to reveal hidden details in density plots is to use interaction. Continuous zooming [1] allows users to explore regions of interest in detail, but at the expense of losing the global context. This issue can be mitigated by computing two density fields: one with a large kernel to provide an overview and another with a small kernel to provide a detailed view [3]. Yet, using two separate views not only consumes more screen space but also increases the cognitive load, requiring the user to correlate the two views when examining local structures. Color lens [4] optimizes the colormap by consuming a small screen area (the lens) and applying the color map globally (clamped to the display gamut) to augment the density plot, mitigating the two views problem. However, this technique still requires a long
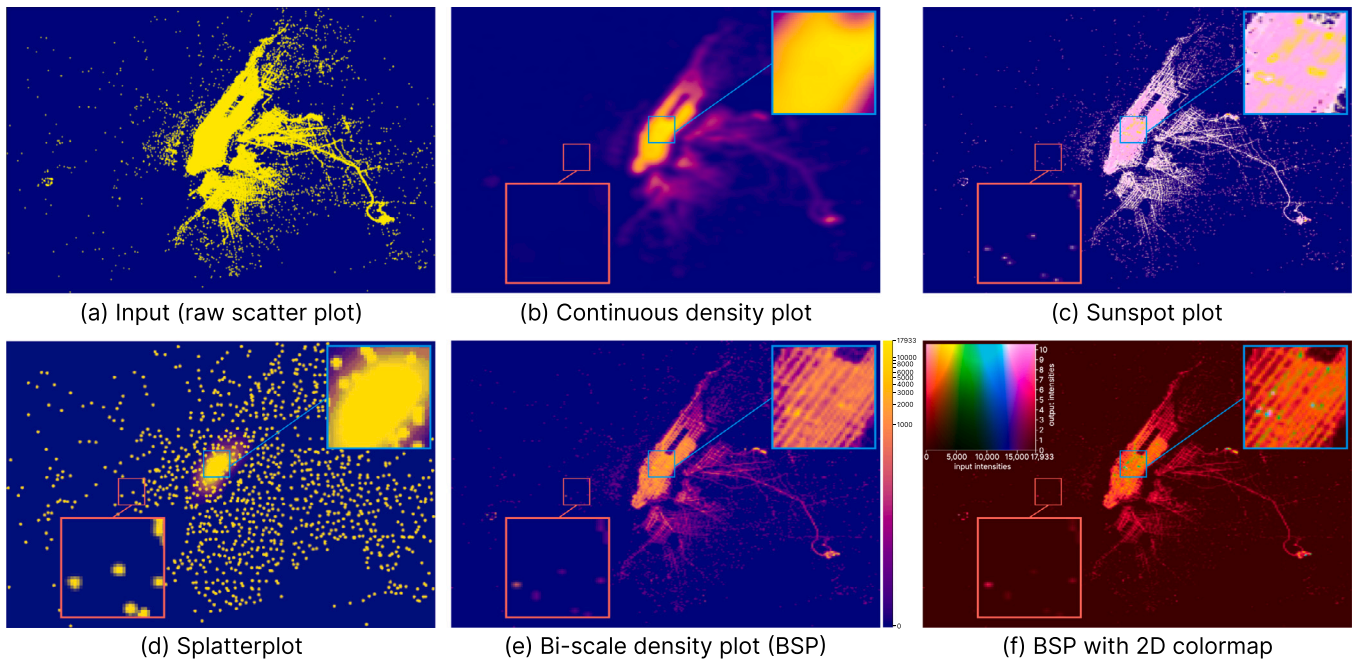

Fig. 1. Different methods for visualizing the "New York TLC Trip" data with two million points. (a) The scatterplot suffers from overdrawing; overlapping visual marks obscure the major structures. (b) A density plot using a Gaussian kernel reveals the global patterns, but the high density in the center (i.e., Manhattan) hides peripheral local structures in the blue box, while outliers in the red box are missed. (c) The Sunspot plot better preserves the outliers but cannot reveal local density variations; see, e.g., the blue box. (d) Splatterplots abstract the density variations in high-density regions, hiding local structures, but show point samples in low-density regions; see the red box. (e) Our Bi-Scale density Plot (BSP) provides more details in low-density regions and reveals more underlying structures; and (f) The BSP enriched by our default 2D colormap, additionally allows looking up and comparing absolute density values and local density variations.

and tedious interactive exploration of the full visualization in order to locate interesting patterns.

In this article, we present *Bi-Scale* density Plots (BSP), a new bi-scale density plot visualization technique. According to the scatterplot task taxonomy [5], an effective visualization design must facilitate the comparison and estimation of numerosity while enabling the identification of outliers and local structures. This requires preserving relative data density and the ability to discern absolute density values while highlighting outliers and structures, as demonstrated in prior studies [2,6–8]. Consequently, we design our density plots in terms of the following four requirements for exploring scatterplots at the overview level:

DR1: providing an overview that *maintains the relative data density* in mid- and high-density regions as much as possible;

DR2: *preserving and revealing outliers* in low-density regions;

DR3: supporting *accurate lookup and comparison* of absolute density values; and

DR4: improving the *readability of structures* in regions of interest.

DR2 is met in SPP [1] and SSP [2]. DR1 and DR2 are met by the scatterplot sampling techniques proposed by Chen et al. [6], but not DR3 and DR4. To the best of our knowledge, these four requirements have not been met by any density plot technique.

We first propose the *bin–summarize–decompose–combine* framework, which outlines the density plot enhancement pipeline and facilitates the efficient computation of bi-scale density plots for large datasets. This framework not only illustrates the common components of existing techniques, such as SPP and SSP, but also elucidates why they suffer from the *ambiguity* issue identified by Kindlmann & Scheidegger [9]. This inspires our approach to decomposing the density field into coarse- and fine-scale layers, thereby better aligning with the design requirements. Second, we revisit the guided filter [10], an edge-preserving smoothing filter, and reformulate it as a variance-aware filter. This adaptation allows for the efficient preservation of structures in high-variance density regions while effectively reducing noise in low-variance density regions. Compared to a conventional Gaussian filter,

the variance-aware filter enables a more effective decomposition of the density field into a base layer that captures the major structures and dynamics of high-density regions, alongside a detail layer that highlights minor structures and outliers in low-density regions. This design permits users to interactively combine the two layers with a specified weight, enhancing the readability of structures of interest without compromising the relative data density. As shown in Fig. 1(e), most outliers are effectively enhanced while the major structures remain clearly visible.

As the variance-aware filter inherently changes the absolute intensities of distant regions, it prevents their accurate comparison. We propose a new 2D colormap for visualizing our density plots: the original density and the enhanced density are mapped to hue and lightness, respectively. The structures can then be perceived efficiently through the lightness channel, while the absolute input densities can be read and compared across the visualization using the hue channel (see Fig. 1(f)).

We evaluated our approach in a controlled study in terms of density-driven analysis tasks and in two case studies. First, we compared our density plots with the state-of-the-art density visualizations (e.g., SPP [1] and SSP [2]). Our method achieves a comparable result in preserving relative densities and outliers while better revealing the local patterns. Second, we show that interactive improvement by adjusting weights on the detail layer and the control points of 2D colormaps enables viewers to flexibly explore structures of interest.

We provide a web-based implementation[1] of BSP to demonstrate its effectiveness. To summarize, our main contributions are as follows:

- We introduce the *bin–summarize–decompose–combine* framework for describing density plot visualization pipelines.
- We propose a bi-scale technique to enhance density plots with the variance-aware filter.

---

[1] https://anonymous-372464.github.io/Filter-basedDensityMapEnhancement/.

- We present a new 2D colormap for visualizing density plots that simultaneously display the enhanced and absolute densities.
- We empirically validate our technique in a controlled study and present two case studies to demonstrate its effectiveness.

2. Related work

Our method relates to the topics of density-based visualizations and edge-preserving image enhancement.

2.1. Density visualization

Density visualizations represent a pivotal technique frequently employed to tackle the issue of overplotting problems in scatterplots. Below, we discuss the generation, enhancement, interaction, and evaluation of density visualizations.

Generation of Density Plots. The generation of density plots includes two principal strategies: adjusting point opacities and employing density estimation techniques. Adjusting point opacity creates semi-transparent visualizations, which can reveal high-density areas. Matejka et al. [11] proposed to set the opacity value based on a user-driven model. Besides, Micallef et al. [12] advocated for an optimization approach, considering opacity, mark size, and additional visual attributes together to align with specific data characteristics and analytical objectives. However, it is difficult to distinguish varying densities in the outputted density plots. Instead, density estimation techniques compute a continuous density field through Kernel Density Estimation [13] (KDE), and the result is often shown as a color-coded plot. For simplicity, we term the resulting visualizations Continuous Density Plots (CDP).

As KDE-based techniques are not scalable to large data [14], Wickham [15] proposed a bin–summarize–smooth framework for handling large data while maintaining the data fidelity. Rather than directly working on the original data, this approach first condenses the original data into a set of discrete densities and then applies a Gaussian-like smoothing to obtain a continuous density field. By doing so, the resulting density field can clearly show similar global data patterns as the one resulting from KDE. Recently, Jo et al. [16] extended this approach to multi-class density maps and provide a declarative rendering model. However, the smoothing step might reduce the local variability, so some important structures could be missed (see Fig. 1(b)). In this work, we revisit this framework and extend it to better account for the major structures and density variations and also to allow for interactively choosing detailed vs. smoothed results.

Enhancement of Density Plots. Traditional continuous density plots have some problems, for example, they often fail to depict local variations and outliers. SPP [1] introduces an approach to preserve outliers. In sparse regions, it displays data points through sub-sampling, while in dense regions, it uses closed smooth contours. However, as SPP is developed for multi-class scatterplots, it mistakenly assigns the same color to some regions, leading to ambiguity in understanding densities (see the red and blue boxes in Fig. 1(d)).

SSP [2] generates two density fields by applying KDE with distinct kernel types to the input data. Following the computation, it renders these density fields using varied colormaps and then combines them adaptively. This process leverages shading and shape cues to preserve the relative differences. The honeycomb plots [8] further enhance the visualization by improving the shading to show outliers more clearly. However, Both methods still face challenges in effectively conveying local density variations (see the blue box in Fig. 1(c)) and have the ambiguity problem as SPP (see the pink areas in Fig. 1(c) in the two boxes).

Recent studies have explored density plot techniques tailored to specific fields. Feng et al. [17] introduced the topology density map,

Table 1
Comparing our bi-scale density plots with prior techniques on the four design requirements (see Section 1), where ✓ and × indicate whether the technique accounts for a given requirement or not.

	DR1	DR2	DR3	DR4
CDP	✓	×	×	×
SPP [1]	×	✓	×	×
SSP [2]	✓	✓	×	×
BSP	✓	✓	✓	✓

enhancing density visualization by incorporating road network topology and traffic conditions. This approach improves clarity and precision in dense urban environments, providing more intuitive insights for urban analysis. Similarly, Xue et al. [18] developed line-based density plots that employ color schemes to represent data density and highlight similar regions, facilitating the identification of trends and patterns. However, the specific focus of these methods limits their versatility, restricting their use as general solutions across diverse applications.

Our bi-scale density plots (BSP) address this problem in three aspects: (i) introducing a new pipeline for computing density fields in a bi-scale way; (ii) reformulating the classic edge-preserving image filter, the guided filter, as variance-aware filter to satisfy DR1 and DR2; and (iii) the 2D colormap design for showing absolute (input) densities and filter densities simultaneously. With this new pipeline, users can interactively combine the coarse- and fine-scale density fields to explore details of interest. Table 1 summarizes the difference among these variants of density plots along the four design requirements (DR1-DR4).

Interaction of Density Plots. A large kernel typically leads to smooth visualizations but might lose some local variations, and vice versa for a small kernel. So, interaction mechanisms are often used to mitigate this issue. Willems et al. [3] provided overview and detail by simultaneously visualizing two density fields with kernels of different sizes. Mayorga and Gleicher [1] introduced continuous zooming for smoothly revealing abstracted details, where details emerge as zooming occurs. Staib et al. [19] devised a focus+context technique that explicitly encodes out-of-focus distance based on the depth of field. Our approach provides a new perspective for enhancing details in density fields, supporting magnifying regions of interest; see Section 4.2.

Evaluation of Density Plots. Trautner et al. [2] compared five distinct visual designs with two tasks: density estimation and density comparison. Their findings revealed that SPP is comparably effective to the KDE-based continuous density plots. Besides, Sarikaya and Gleicher [5] outlined a collection of tasks unique to scatterplots and assessed the efficiency of scatterplots, contour plots, and SPP in supporting aggregation-level tasks. In this work, we further expand the evaluation method to support our research.

2.2. Edge-preserving image enhancement

Image smoothing is a fundamental operation in image processing and computer vision. It aims to remove noise from the image while preserving the major structures. The widely-used Gaussian filter computes a weighted sum of pixel values in a local neighborhood. However, since it has non-zero weights for all pixels, even when they are irrelevant and far from the neighborhood center, over-smoothing and edge detail loss often occur. Hence, various edge-preserving image filters [20] have been proposed to preserve the high-contrast edges while removing the small-contrast ones. In this work, we adopt the guided filter [10], a simple and popular filter, which can effectively achieve good edge-preserving smoothing like some other filters, such as the bilateral filter, yet almost avoiding the "gradient reversal" artifacts near the edges. By considering a discrete density field as an image, we reformulate the guided filter to be density variance-aware, aiming at reducing the noise in the density field and maintaining the relative data density.
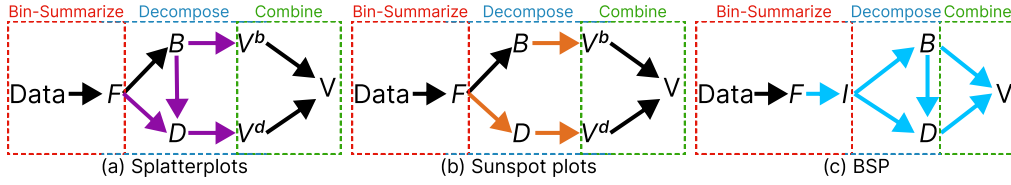
Fig. 2. Illustrating our unified bin–summarize–transform–decompose–combine framework, where the bin–summarize stage first produces the density field F. Then, we can choose to map F to a new density field I by a global transformation or not. After that, we decompose the result into base layer B and detail layer D, together with the associated rendering results V^b and V^d, respectively. Finally, V^b and V^d are combined to yield the final plot V. The black arrows indicate the common paths for all methods, whereas the colored arrows indicate technique-specific paths.

To enhance image details, Durand and Dorsey [21] proposed to decompose an image into two fields: a base layer (edge-preserving smoothing result) and a detail layer (difference between the input image and smoothing result). Combining the detail layer with a large weight on the base layer can emphasize the details while preserving the large-scale features in the result. We show that this bi-scale decomposition can help facilitate the exploration of details in the density field.

3. Bin–summarize–transform–decompose–combine framework

In this section, we introduce our bin–summarize–decompose–combine framework for generating enhanced density plots from large and dense bivariate data. We then formulate SPP [1], SSP [2], and our proposed Bi-Scale Density-Plot (BSP) under this framework. Also, we present our variance-aware filter and explain how it better characterizes structures.

Building upon the bin–summarize–smooth approach of Wickham [15], our framework consists of the following three stages:

Bin–summarize We bin and summarize the data into a 2D density field F, in which each value denotes the density in an associated discrete bin. To improve the contrast in low-density regions, F can undergo optional preprocessing, e.g., logarithm transforms, or intensity histogram equalization.

Decompose From the discrete density field F, we generate the base (B) and detail (D) fields, called layers, that capture the coarse- and fine-scale density variations, respectively. Wickham's approach only creates one base layer at this stage, that undergoes a *smooth* filter. Other methods use image processing techniques to create and separate the fields, sometimes improving or filtering each field separately.

Combine We further combine the processed detail layer D and the base layer B to produce the final density plot. The combination is not limited to alpha blending the two layers; our BSP method provides a more sophisticated operation.

To visualize a bivariate dataset, the first step produces a density field F: $R^2 \rightarrow R$ in screen space, where F_i is the number of data samples in bin i on the screen. To explore how the generated density plot supports DR1 and DR2, we define the *relative data density* [22] between two regions having the same area Ω_i and Ω_j as 1 if $\sum_{k \in \Omega_i} F_k > \sum_{k \in \Omega_j} F_k$, as -1 if $\sum_{k \in \Omega_i} F_k < \sum_{k \in \Omega_j} F_k$, and as 0 otherwise. As there is no clear definition of outliers, we follow SPP to treat data points in low-density regions as outliers, but we do not set a specific density threshold. To reduce the colormap influence on the perception of density plots, we follow Trautner et al. [2] that use the perceptually uniform colormaps [23]: Plasma and gray for illustration and evaluation, respectively.

3.1. Revisit existing density plots

Below, we show that we can re-write the existing enhanced density plots techniques under our bin–summarize–decompose–combine framework. They all rely on computing a KDE, that we perform in image space. Wickham [15] explains that calculating the KDE of a density

field in data space produces results similar to calculating it in image space with a complexity $O(n)$ (n the number of data points) for the former and $O(1)$ (a constant that depends on the image resolution) for the latter.

Splatterplots. Like traditional density plots, SPP first computes a continuous density field B by smoothing F using a kernel function K:

$$B_i = \frac{1}{\sum_{j \in \Omega} K(x_i - x_j)} \sum_{j \in \Omega} K(x_i - x_j) F_i, \tag{1}$$

where x_i is the coordinate of bin i and Ω is the whole density field. By default, K is a Gaussian function:

$$K(d) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{d^2}{2\sigma^2}), \tag{2}$$

where $K(d)$ depends on distance $d = |x_i - x_j|$ and σ is the kernel bandwidth. The larger σ, the smoother the density field B.

To reduce the visual complexity of the continuous density plot, SPP decomposes the density field B to obtain a detail field D:

$$D_i = \begin{cases} B_i & \text{if } B_i \leq \lambda \\ 0 & \text{otherwise,} \end{cases}$$

where λ is a threshold. Next, we generate the base layer by assigning a constant color to the bins with density larger than λ:

$$V_i^b = \begin{cases} c & \text{if } \lambda \leq B_i \\ M(B_i) & \text{otherwise,} \end{cases}$$

where M is a 1D color mapping function from density to color; and c is the foreground color. Then, the contours that enclose all pixels of color c are computed and overlaid onto the density visualization V^b. The detail visualization V^d is constructed with a set of data points subsampled from the detail field D; see, e.g., Fig. 1(d).

Finally, the *Combine* stage performs alpha blending:

$$V = \alpha V^b + (1 - \alpha) V^d \tag{3}$$

where $\alpha \in [0, 1]$ is a constant value. In doing so, the resulting plot V contains three kinds of visual primitives: color-coded densities, discrete isolated points, and polygons filled with a constant color.

Sunspot Plots. SSP can also be formulated under the same framework. We can generate the base field B by applying a Gaussian kernel to F, and then the detail field D by applying a discrete kernel to F. Then, they are transformed into images V^b and V^d:

$$V^b = M^b(B), V^d = M^d(D), \tag{4}$$

where M^b and M^d are color mapping functions, which can be the same, although the value ranges of B and D are usually different. Finally, V^b and V^d are combined through alpha blending, but the opacity of each pixel i is proportional to the smoothed density value B_i.

Fig. 2 illustrates the process of constructing these two plots under the bin–summarize–decompose–combine framework. We can see that they both might suffer from ambiguity [9] since similar pixel colors in the two layers might actually correspond to different density values, see Figs. 1(c), 1(d). Although this issue can be alleviated by using non-intersecting colormaps, their combination might lead to misleading colors in the overlapping regions due to alpha blending.
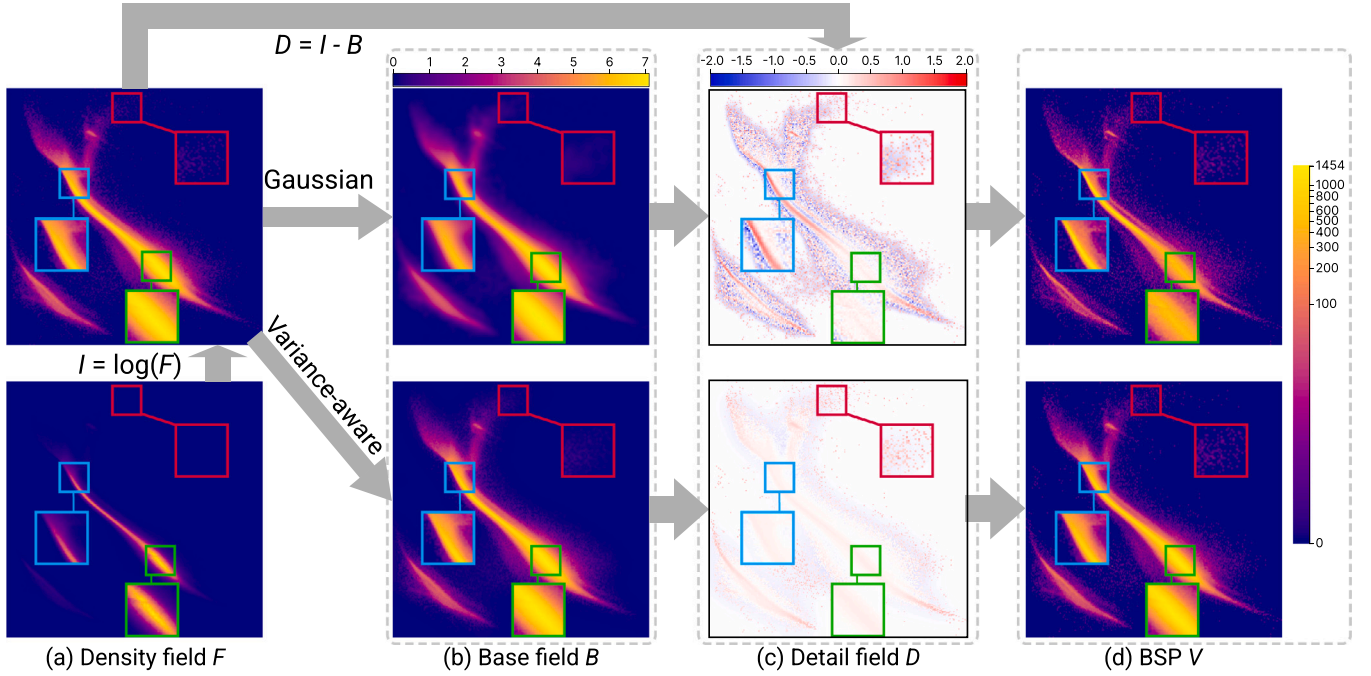
Fig. 3. The bi-scale density-plot enhancement approach, applied to two pipelines, a Gaussian filter on top and our variance-aware filter on the bottom. First, we bin–summarize the input data and generate the density field I via a logarithmic transform. Then, we decompose I into the base and detail layers, using our variance-aware filter (bottom path). Finally, we magnify the detail layer and incorporate it into the base layer to produce the bi-scale density plot (BSP) to better reveal the outliers (see the ones in the red box).

3.2. Bi-scale density plot enhancement

Fig. 3 illustrates the pipeline of our method, corresponding to the three steps in Fig. 2(c). First, we transform the bin–summarized density field F to the logarithm domain to expand the low-density and compress high-density range. Next, we decompose it into the base and detail fields ($B\&D$). Finally, we weight the detail field (ωD) to enhance local details, combine it with the base field, and colorize the combined result (I') using a colormap. These steps are formulated as follows:

$$I = \log(F + 1),$$
$$B = K(I), \quad D = I - B,$$
$$I' = B + \omega D, \quad V = M(I'), \tag{5}$$

where K is the filter function, ω is a weight, and M is a 1D colormap. Note that B and D in Fig. 3 are encoded using different colormaps in the illustration. Using a proper default ω, the fine-scale structures captured by D can help enhance the details in the final density plot V, which is referred to as a bi-scale density plot, or BSP. By interactively adjusting ω, users can control the relative importance of the local details.

Using the Gaussian kernel (see Eq. (2)) as the filter, however, might over-enhance some major structures, leading to breaking DR1. For example, the Gaussian smoothed B loses the salient edges in I, and these edges are extracted to the detail layer D (see the blue box at the top of Fig. 3(c)). Thereby, enhancing the detail with Eq. (5) overly magnifies the red edges and violates the relative data densities. Obviously, the maximal data density in the green box is larger than that in the blue box in F (see Fig. 3(a)), but it becomes smaller at the top of Fig. 3(d). In the following, we show how our density variance-aware filter could help address this problem.

3.2.1. Reformulated guided filter: Variance-aware filter

The original guided filter [10] was designed to smooth out noises in natural images while preserving edges, without suffering from gradient reversal artifacts. Aiming at preserving the relative densities (DR1) and enhancing outliers (DR2) in the final density plots, we propose to reformulate the guided filter in terms of density variance. With the

reformulation, we can show that our variance-aware filter can better maintain the relative data densities in high-density variance areas, while reducing excessive variability in low-variance areas.

After obtaining the log-transformed density field I, we construct overlapping tiles of size $h \times h$ over the field, so each bin (pixel) in I is covered by at most h^2 tiles. Assuming that there is a local linear model between I and the smoothed field B, we can express B_i as a linear transform of the pixels I_i contained in the tile Ω_k centered at bin k:

$$B_i = a_k I_i + b_k, \forall i \in \Omega_k, \tag{6}$$

where i is a bin index and (a_k, b_k) are the linear coefficients of tile Ω_k. By finding a best-fit linear relation between the I and the D, this model has proven to be useful in removing excessive variability [24]. However, improperly choosing (a_k, b_k) removes major structures, in a way similar to a Gaussian filter. To avoid this issue, one can constrain the filtered result to be as close as possible to the input density field.

To that end, the linear coefficients (a_k, b_k) can be determined by minimizing the difference between the smoothed B and input I:

$$E(a_k, b_k) = \sum_{i \in \Omega_k} \left((a_k I_i + b_k - I_i)^2 + \tau a_k^2 \right), \tag{7}$$

where τ is a regularization parameter penalizing large a_k. Setting the derivative of $E(a_k, b_k)$ with respect to a_k and b_k to zero yields the following solution:

$$a_k = \frac{\sigma_k^2}{\sigma_k^2 + \tau}, \quad b_k = (1 - a_k)\mu_k, \tag{8}$$

where μ_k and σ_k^2 are the mean and variance of I in Ω_k. Since τ is a positive number, a_k and b_k have ranges $[0, 1)$ and $(0, \mu_k]$, respectively. Substituting Eq. (8) into Eq. (6), we obtain

$$B_i = \mu_k + (I_i - \mu_k)a_k = \mu_k + (I_i - \mu_k)\frac{\sigma_k^2}{\sigma_k^2 + \tau}. \tag{9}$$

For a tile Ω_k with high-density variance ($\sigma_k^2 \gg \tau$), we have $a_k \approx 1$, $b_k \approx 0$, and $B_i \approx I_i$; and for Ω_k with low-density variance ($\sigma_k^2 \ll \tau$), $a_k \approx 0$, $b_k \approx \mu_k$, and $B_i \approx \mu_k$. In other words, for areas with a high-density variance, the density values are preserved, whereas for areas
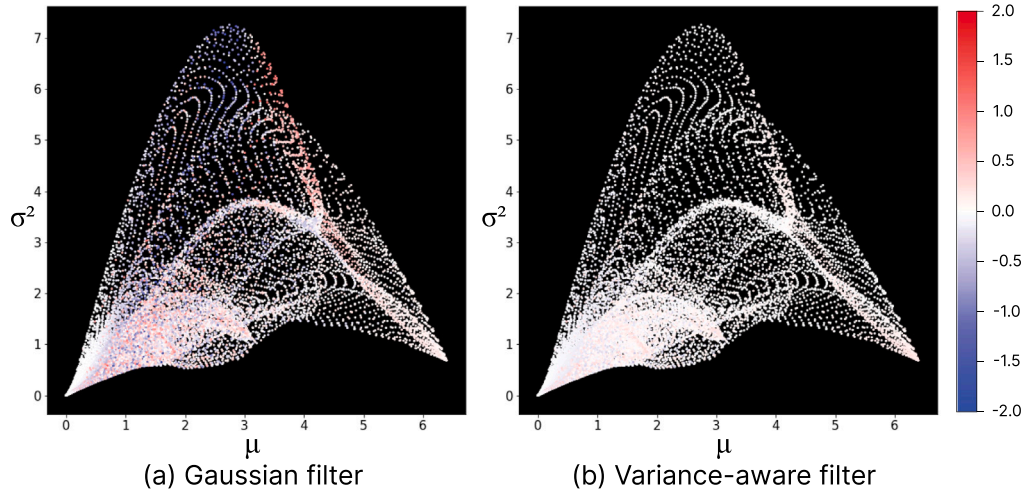
Fig. 4. The scatterplots visualize the mean density (μ) and density variance (σ^2) of points representing the bins k of a lower-resolution grid over I in Fig. 3. Their color encodes the density difference $D = I - B$ in Fig. 3(c) for the Gaussian filter (a) and the variance-aware filter (b). Most points in (a) have large differences (more red/blue), while our method limits this difference; points with large differences in (b) occur mainly in the areas with low mean values and low variances.

with a small-density variance, the density values are moved mostly around the mean. Though the low-variance areas are smoothed, their density change is also small due to the small variance. So, the relative data densities can be largely preserved.

As bin i is contained in multiple overlapping tiles, B_i in Eq. (6) may have different values when computed in different tiles. This problem can be circumvented by averaging all possible $a_k I_i + b_k$ for all tiles Ω_k that cover i. By symmetry of the tile, all centers k of Ω_k covering i are covered by Ω_i, so Eq. (6) can be re-written as

$$\bar{a}_i = \frac{1}{|h^2|} \sum_{k \in \Omega_i} a_k, \quad \bar{b}_i = \frac{1}{|h^2|} \sum_{k \in \Omega_i} b_k, \quad B_i = \bar{a}_i I_i + \bar{b}_i.$$

To learn how the variance-aware filter preserves the density variance, we compute the detail field D, and then study how D is related to μ_k and σ_k^2 of I. As shown in Fig. 4(b), our variance-aware filter mainly alters low-density, low-variance regions, while the Gaussian filter alters the density field regardless of the density variance, smoothing high-variance regions largely (see red and blue points in Fig. 4(a)). Such smoothing causes edge loss, as highlighted by the blue box at the top of Fig. 3(b). Due to such differences, our filter model can better preserve the density variance. For example, the density variances in the blue and green boxes at the top of Fig. 3(a) are changed from (5.58, 1.84) to (4.55, 1.69) and (5.35, 1.73) by the Gaussian filter and the variance-aware filter, respectively.

Parameters h and τ. Tile height h determines the region size, whereas τ determines if a tile region is low- or high-variance. Fig. 5 shows their influence on the filtered result: a larger τ smooths out more details and h does not significantly impact the filtering results. This is reasonable, since a large τ means more regions are smoothed, while our filter model maintains the density variations regardless of the tile size. Empirically, we set $h = 20$ and $\tau = 0.16$, which work well for most scatterplots because the density of bins containing outliers is usually 1 and the corresponding tile is nearly empty ($\mu_k \approx 0$). According to Eq. (5), density variances of outlier bins are $(\log(2) - \log(1))^2 \approx 0.09 \ll 0.16$, while the ones of other bins are not less than $(\log(3) - \log(1))^2 \approx 0.23 > 0.16$. To treat points in bins with higher density as outliers, users can increase the value of τ accordingly.

3.2.2. Detail enhancement

Following Eq. (5), we can obtain the bi-scale density plots to meet DR2 by combining the base layer B and the weighted detail layer D. A large weight ω helps to more clearly show the outliers in low-density regions but may dim the major structures captured by the base layer.
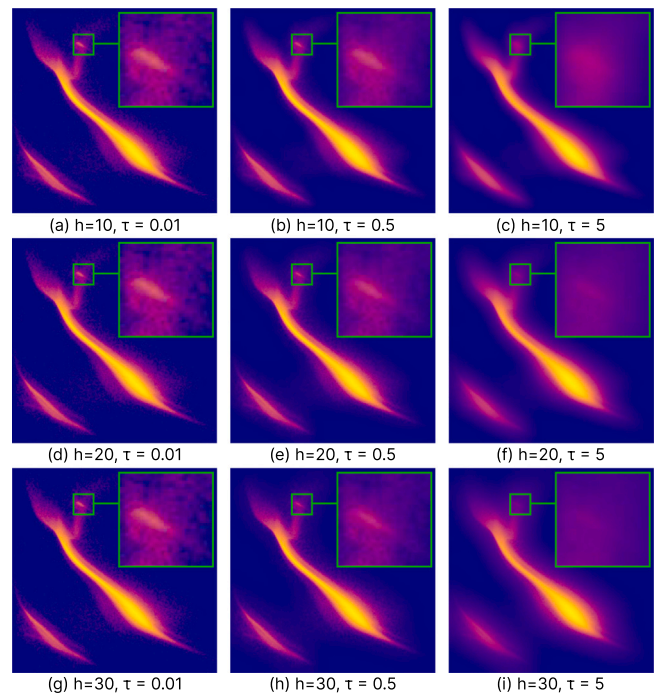

Fig. 5. Effect of parameters h and τ on the filtered result B. A large τ results in over-smoothing and a small τ cannot effectively reduce variability, whereas h has little impact on the results.

By default, we use $\omega = 3$ for all tested data in our experiments. Since some D_i are negative, the resulting I' might be negative, which is meaningless for a density plot. To remedy this issue, we set I' to zero when $I' < 0$.

Fig. 3(d) compares the results after detail enhancement between the Gaussian filter and the variance-aware filter. Both filters perform similarly for low-density regions (see the red boxes) where the variance is small. Yet, they have very different behaviors for the other regions, where the Gaussian filter overly smooths out the details but our filter model almost preserves the intensity. Accordingly, their resulting detail layers shown in Fig. 3(c) have different value ranges: [−2.01, 1.56] (Gaussian) vs. [−0.28, 1.05] (variance-aware). Boosting the detail layer with weight $\omega = 3$, both density plots highlight the details in Fig. 3(d)
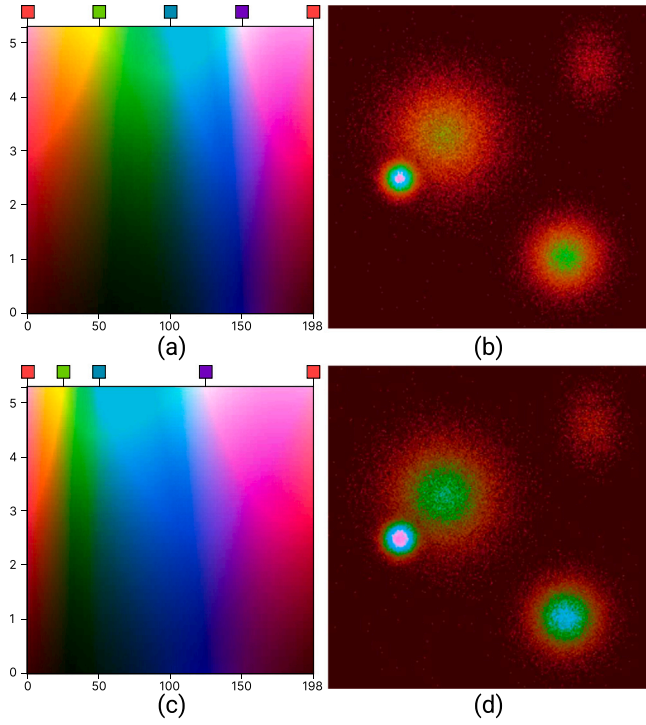
Fig. 6. (a) Our default 2D colormap with the absolute values of the input density field and the ones of the enhanced field encoded by hue on the *x*-axis and luminance on the *y*-axis and its resulting colorized density plot of a synthetic dataset displayed shown in (b). (c) The colormap changed by moving more hues to the low-density ranges and (d) its resulting density plot better reveals two inner green clusters.

the color space [26], where hues cannot be distinguished, we follow the Munsell sample book [27] to consider the luminance range only above the value of 24, and linearly map the density range in I' to this range in the *y*-axis. For the density range in F, we map it to the whole hue range on the *x*-axis and set the chroma channel in the CIELCh space as a constant value of 100 as the default. Fig. 6(a) shows the 2D colormap, where different hues represent different density value ranges (e.g., green for values between 50 and 100). Using this colormap, the density plot of the synthetic dataset shown in Fig. 6(b) facilitates users to better estimate the absolute density values and compare them to distant regions. For instance, the cluster in green on the bottom right is significantly denser than the large cluster on its upper left, while the top-right cluster has low-density values but its structure is clearly revealed.

By specifying different hue values to different control points on the *x*-axis, users are allowed to interactively add, delete, or move key points to construct a new colormap to show different densities of interest. An example is shown in Figs. 6(c), 6(d), where more control points are moved to the range [0,50] for exploring low-density clusters. From Fig. 6(d), we can see an inner green cluster within each of two large red clusters.

4. Evaluation

We implement a web-based interactive prototype using JavaScript and OpenCV.js [28]. Owing to the efficiency of our bin–summarize–decompose–combine framework, this prototype can generate BSPs for data with millions of point samples in less than 200 ms on a PC with an Intel Core i5-4590 3.3 GHz CPU and 24 GB memory, whereas SPP and SSP both require GPUs to achieve similar speeds because their implementations provided by the original authors directly deal with point samples in the data space rather than a bin–summarized density field. Besides supporting interactive parameter adjustment (combination weight ω, tile height h, and regularization parameter τ), it provides two interactions: (i) brush a region and assign a different weight ω to enhance/lower the importance of the structures, and (ii) specify different hue to control points in the 2D colormap (see Fig. 6(c)).

To demonstrate the effectiveness of our approach, we conducted a controlled study to compare BSP with state-of-the-art density plot visualizations and performed two case studies for interactively exploring two real-world datasets. For all tested data, we compute the density plots with the resolution 256×256. The full evaluation results, including screenshots, the website, and the analysis code, can be found in the supplemental material.

4.1. Controlled study

Density plots inherently support two categories of abstract analysis tasks [5] related to scatterplots: open-ended browsing and aggregation judgments. We evaluate the effectiveness of BSP by choosing representative abstract tasks from each category to design our experimental tasks. Based on them, we created five concrete experimental tasks: *density comparison*, *density estimation*, *find maximum*, *count outliers*, and *search for pattern*; they correspond to two aggregate-level tasks (numerosity comparison/estimation and identify anomalies) and two browsing tasks (explore data and search for a known motif) as listed in Fig. 8(a). These tasks are also well aligned with the four design requirements (see Section 1), so the experimental results also indicate how BSPs meet these requirements.

Density Plot Techniques. We compared our two BSP techniques with three existing density-plot techniques as baselines: CDP, SPP, and SSP. We include two versions of SSP in our experiment. SSP can use the same colormap for both layers, referred to as *SSP-s*, or two different colormaps, referred to as *SSP-d*.

Based on our framework, each version of BSP is determined by two factors: filters (Gaussian filter or variance-aware filter) and colormaps
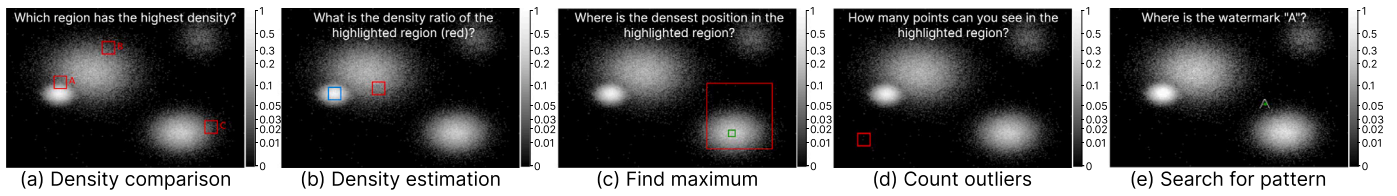
but the one with the Gaussian filter undesirably changes the relative data density. For example, the highest density of the green box at the top of Fig. 3(a) is (7.28), which is greater than that of the blue box (6.88). Yet, their order is reversed by the Gaussian filter at the top of Fig. 3(d) (7.71 < 8.65), whereas the relative data density between them is preserved by our filter model (7.74 > 7.14) at the bottom of Fig. 3(d). The reason is that the differences within these highest-density bins yielded by our model are close to zero, avoiding falsely changing the relative data density when boosting D. Also, since a_k is always less than one, the gradient of the detail layer has the same direction as the input density field:

$$\partial_x B = a_k \partial_x I,$$
$$\partial_x D = \partial_x I - \partial_x B = (1 - a_k)\partial_x I.$$

This ensures our filter model does not suffer from the gradient reversal artifacts, thereby better preserving the local relative densities. When a_k and b_k are replaced by the averaged coefficients $\bar{a}_i$ and $\bar{b}_i$, we obtain $\partial_x D \approx (1 - \bar{a}_i)\partial_x I$ and thus the above observation still holds.

3.3. Bivariate colormaps

With the detail enhancement, BSP meets DR1, DR2, and DR4. Yet, this process is not a perfect linear map from densities to screen intensities, thus preventing direct estimation of absolute densities from the plots. Hence, instead of using a 1D colormap, we propose a new bivariate color mapping strategy employing the perceptually uniform CIELCh color space [25].

In our method, we use the bivariate colormap to simultaneously encode the intensities of the input and output density plots. As Figs. 6(a), 6(c) show, the density values of F is mapped to the *x* axis encoded with the hue channel, and I' is mapped to the *y* axis encoded with the luminance channel. Also, to avoid the dark and white regions of

Fig. 7. Illustration of the five tasks and the corresponding questions we asked the participants. Images are generated by using our BSP-VF-1D method on the synthesized dataset, where the red boxes and the blue box are contained in the stimuli, and the green box in (c) and green point in (e) are the participants' input.

(1D or 2D colormaps). Hence, we include four versions: BSP with Gaussian filter and 1D colormaps (BSP-GS-1D) and BSP with Gaussian filter and 2D colormaps (BSP-GS-2D), BSP with variance-aware filter and 1D colormaps (BSP-VF-1D), and BSP with variance-aware filter and 2D colormaps (BSP-VF-2D). To reduce the color influence on the perception, we render the density fields with a gray colormap for all density plots except SSP-d, BSP-GS-2D, and BSP-VF-2D.

For SPP and SSP, we use the default parameters provided by the implementations of the original authors. For a fair comparison, we set the same bandwidth $\sigma = 2$ for methods involved in the Gaussian kernel (CDP, BSP-GS-1D, BSP-GS-2D), $h = 20, \tau = 0.16$ for methods involved in the variance-aware filter (BSP-VF-1D, BSP-VF-2D), and $\omega = 3$ for all BSP methods.

To summarize, we compare 8 techniques: CDP, SPP-s, SPP-d, SSP, BSP-GS-1D, BSP-GS-2D, BSP-VF-1D, and BSP-VF-2D.

Tasks & Measures. Our five analysis tasks are:

1. *Density comparison*: Following the methodology of Trautner et al. [2], we highlighted three regions (A, B, C) using a 50×50 px red boxes (see Fig. 7(a)) and asked participants to choose one region with the highest density. Each question offers three possible answers (A, B, or C), and only one answer is right. We defined the error measure as 0 if the participant's response was correct and 1 otherwise.

2. *Density estimation*: Following the methodology of Trautner et al. [2], we highlighted one region using a 50×50 px red box (see Fig. 7(b)) and asked participants to estimate the density ratio to the densest region on a range between 0 and 1, with the value of 0 corresponding to the region with the lowest-density and 1 to the highest. Note that we refined this task by showing the highest-density region in blue after the pilot study. The error was computed as the absolute difference between the participant's answer and the actual density ratio of the highlighted region to the highest-density region.

3. *Find maximum*: We highlighted a region using a 250×250 px red box, and asked participants to find a small region within it with the highest density, marked by a 25×25 px green box (see Fig. 7(c)). The error was computed as the difference between the density value in the selected position and the actual maximum in the highlighted region, normalized to $[0, 1]$ by dividing the density difference by the maximum density.

4. *Count outliers*: We highlighted a sparse region using a 50×50 px red box (see Fig. 7(d)) and asked participants to count the points within it. The error was computed as the absolute difference between the participant's answer and the actual number of points.

5. *Search for pattern*: We embedded a letter "A" into the dataset by adopting a spatial-domain watermarking technique [29]. Specifically, we compute the mean density within the bounding box of the given letter, then add one-quarter of the mean density to the densities of the letter related foreground pixels, and finally encode the modified density field with the gray colormap. Given such density plots, we asked the participants to find the position of the watermark in the density plot (see Fig. 7(e)). The error was evaluated as binary: 0, if the selected position fell within the bounding box of the watermark, and 1, otherwise.

As shown in Fig. 8(a), the first and third tasks are related to numerosity comparison and the second is about numerosity estimation. Unlike the task of *density comparison* that only gives a judgment of relative density among three regions, *density estimation* requires the estimation of the absolute density values, and *find maximum* further involves estimating the density of all small areas within the given large area. The other two tasks correspond to the abstract analysis tasks of *identify anomalies* and *search for known motif*, respectively. Fig. 7 shows a sample screenshot for each task with the size of 900×600 pixels, which ensures that each individual point can be seen clearly.

Datasets. We collected four datasets with varying density distributions, including two synthetic datasets and two real-world datasets. The first synthetic dataset with 255,147 points is generated by mixing four Gaussian clusters with the maximal densities varying from [14, 198], while the other consists of only one large, sparse cluster with 2001 records collected from Abbas et al. [30]. The two real-world datasets are the *2014 Boston Marathon* dataset [31] with 31,945 records and the *Person Activity* dataset [32] with 98,568 records. For each dataset and each task, we created two stimuli with different randomly-selected regions highlighted. As a result, we have a total of

5 tasks × 8 techniques × 4 datasets × 2 stimuli = 320 trials.

Hypotheses. We expect our approaches to outperform the state-of-the-art methods in revealing underlying structures while maintaining relative data density and outliers. Trautner et al. [2] show that SSP performs similarly to CDP on density estimation tasks, and we expect that our BSP-VF-2D performs even better than them due to the bivariate colormaps. Hence, we had the following hypotheses:

H1: BSP-VF-* are comparable to CDP and SSP-d in the *density comparison* task, and perform better than SSP-s and SPP.

H2: BSP-*-2D outperform the other methods in the *density estimation* task.

H3: BSP-VF-* outperform the other methods in the *find maximum* task.

H4: BSP-* are comparable to SSP and perform better than SPP and CDP in the *count outliers* task.

H5: BSP-VF-* perform the best, followed by BSP-GS-*, and the others have the worst performance in the *search for pattern* task.

Apparatus. The study was conducted on a quad-core PC with a 27" LCD widescreen with a mouse and a keyboard as the input and a 3840×2160 pixel monitor with 163 PPI as the output, which was calibrated for faithful color reproduction. All participants were seated at ~60 cm from the display in a nearly constantly-illuminated room.

Pilot Study. We conducted a pilot study with eight participants from our laboratory to try out our experimental design. Before the study, we explained the tasks to the participants and instructed participants to use our 1D and 2D colormaps to estimate density values. Then, we only used one synthetic dataset, so each participant had to complete $5 \times 8 \times 2 = 80$ trials, taking around 15 min. We performed a follow-up interview with each participant and asked them if there was any design factor that limited their efficiency and accuracy in completing the task and if they had any suggestions for improving the study.

The answers hinted at one major factor that influenced the results: a strong learning effect. The learning effect was caused by two different
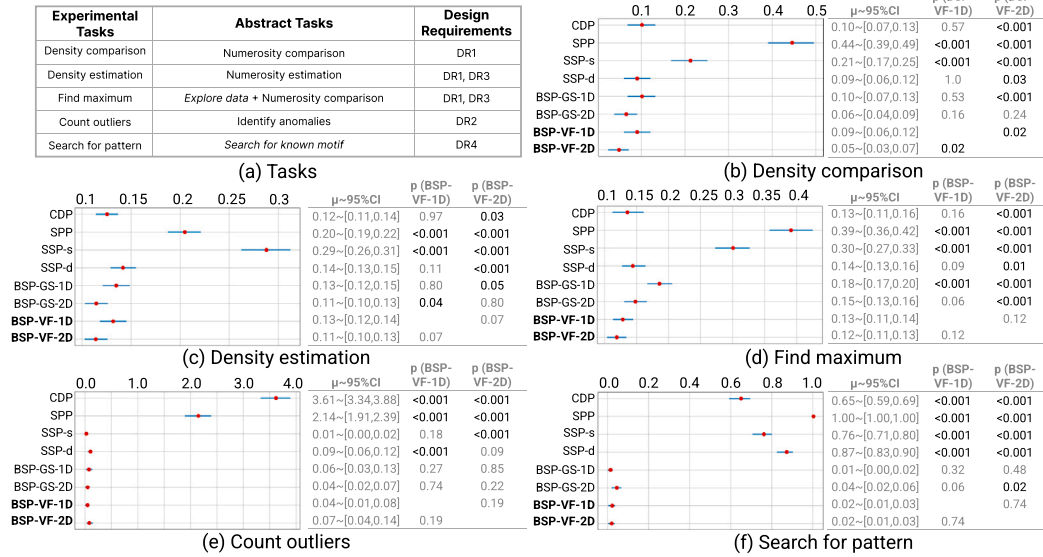
Fig. 8. The experimental tasks and results of each task in our controlled study. (a) The table shows how five experimental tasks correspond to the abstract tasks and design requirements. (b-f) The results of the controlled study, where we show the confidence interval plots and statistical tables for the tasks. Error bars represent 95% confidence intervals. Each table shows the statistical test results of our experimental methods, showing the mean with 95% confidence interval and *p*-value from the Mann–Whitney test with two target methods (BSP-VF-1D and BSP-VF-2D).

highlighted regions on the same dataset used by the eight methods so the participants could easily remember the highlighted regions and corresponding answers. Also, almost all participants indicated that the *density estimation* task was challenging because of two factors. One is that it was hard for them to find the reference region with the highest density. The other was that the tick value shown on the colormap legend was the absolute density value, which cannot be directly used for estimating a normalized density value in the range [0,1] for task 2.

Such findings led us to create more variations of density plots. Yet, adding more stimuli would lead to too many trials. To address this issue, we flipped the density plots either horizontally or vertically to create variations while maintaining the tasks equivalent; the local structures were not changed. For the *density estimation* task, we explicitly show the highest density region marked by a blue box (see Fig. 7(b)) and normalize the tick value of the colormap legend to [0,1] for all tasks.

Participants. We recruited 45 participants, yielding a power of 1 at the effect size Cohen's f = 0.25 and 70% order effect coverage: 35 males and 10 females, aged 18 to 39, including 32 undergraduates, 11 masters, and 2 Ph.D. in different majors. None have color vision deficiencies.

Procedure. We used the Touchstone2 [33] tool to design a *within-subject* experiment, in which the order of tasks was fixed, and the methods, datasets, and stimuli were counterbalanced with a Latin square. This ensures that each participant has a different trial order of all stimuli for each task, which helps avoid systematic errors and minimize random errors [34]. Each participant went through the following steps: (i) review a consent page and the task instructions, followed by completing four training trials, which include the density plots generated by BSP with 2D colormaps; (ii) complete each trial as accurately as possible, where each participant was asked to take a 2–3 min break after finishing each task; and (iii) provide demographic information. The four training trials were identical to the subsequent real test. On average, the participants took 35 min to finish all the trials (min: 25 and max: 45).

We implemented different response mechanisms for the five tasks. For the *density comparison* and *count outliers* tasks, participants clicked a letter or number they thought was the correct answer from multiple options displayed below the visualization, while participants entered their estimated values in the *density estimation* task. For the *find maximum* and *search for pattern* tasks, participants clicked the possible position in the density plot.

Analysis. Following the previous study [2], we did not assume that the underlying data fits the normality assumption and hence analyzed the results using 95% confidence intervals. We performed the *Mann–Whitney* test for pairwise comparisons between methods to learn if they have significant differences.

Results. Figs. 8(b)–8(f) show the experimental results for the five tasks, where two target methods used by the *Mann–Whitney* test and the p-values with significant differences are highlighted in bold. The response time for all methods is similar, and thus, we do not show it.

Fig. 8(b) shows the results of the *density comparison* experiment. There is no significant difference between BSP-VF-1D and CDP, SSP-d, and BSP-GS-*, while the differences between BSP-VF-1D and SSP-s and SPP are significant. More specifically, we found that BSP-VF-2D even outperforms CDP and SSP-d. The results overall confirm **H1** that BSP-VF-* is comparable to CDP and SSP-d and performs better than SSP-s and SPP. Also, BSP-VF-2D performs similarly to BSP-GS-2D and outperforms BSP-*-1D, indicating that our bivariate colormaps can help reduce the error in relative density comparison.

Fig. 8(c) shows the results of the *density estimation* experiment. BSP-VF-2D and BSP-GS-2D perform similarly and slightly better than CDP, BSP-*-1D, and SSP-d, followed by SPP, and SPP-s is the worst. The statistical differences between BSP-VF-2D and SSP and SPP are highly significant, and BSP-VF-1D is significantly worse than BSP-GS-2D. The results largely confirm **H2**, indicating that all BSP-*-2D methods outperform the other methods in density estimation.

Fig. 8(d) shows the results of the *find maximum* experiment. BSP-VF-* perform the best and slightly better than BSP-GS-2D, SSP-d, and CDP, followed by BSP-GS-1D and SSP-s, and SPP is the worst. The statistical differences between BSP-VF-2D and all the other methods except BSP-VF-1D are significant, and the ones between BSP-VF-1D and BSP-GS-1D, SSP-s, and SSP are significant. The results largely accept **H3**, indicating that the variance-aware filter helps for locating the correct maxima.
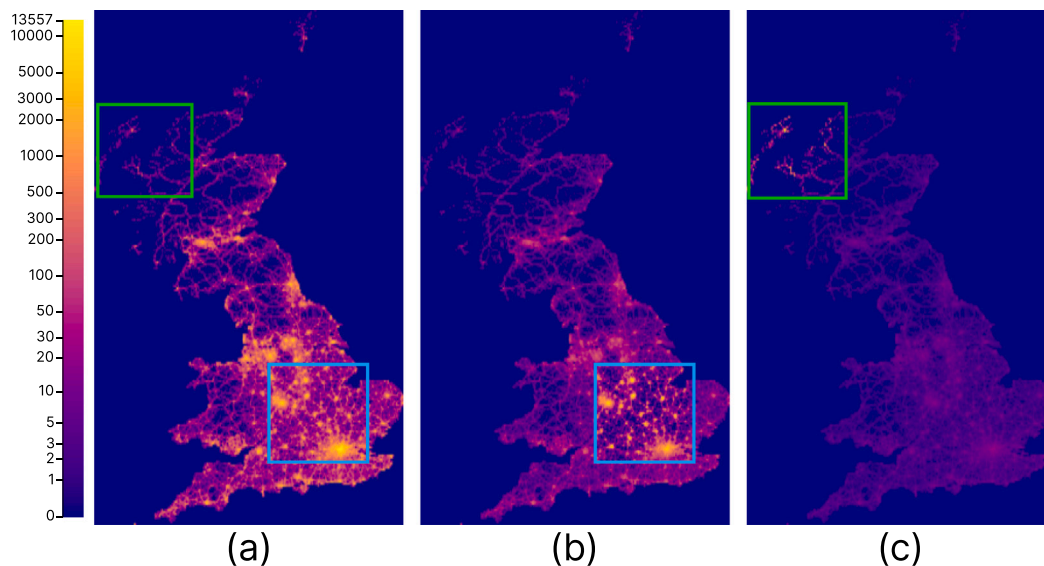
Fig. 9. Density plots of the "UK Road Safety" dataset. (a) The BSP generated with default parameters and (b,c) the results of highlighting the area above London (blue box in (b)) and the northwest area of Scotland (green box in (c)) with a large local ω (20) for the selected regions.

As for the *count outliers* task, we observed that BSP-* performs similarly to SSP-* and is significantly better than CDP and SPP (see Fig. 8(e)), which confirms **H4**. For the *search for pattern* task, all BSP methods perform similarly and significantly better than the others in revealing local structures (see Fig. 8(f)). Yet, there is no significant difference between BSP-VF-* and BSP-GS-* methods, and thus, the results partially support **H5**.

In summary, we have the following controlled-study results:

1. BSP performs significantly better than the other methods in preserving local patterns and is comparable to SSP in revealing outliers;
2. BSP with the variance-aware filter can better preserve the local maximum, and BSP with bivariate colormaps make density comparison and estimation more accurate;
3. Combining BSP with the variance-aware filter and bivariate colormaps performs the best for all tasks.

Discussion. From the above results, we can see that all BSP methods are comparable or better than CDP, SPP, and SSP in preserving relative data densities, outliers, and local patterns. The only reason is that our BSP enhancement strategy is more effective than the previous strategies (see Fig. 2). Although SSP is comparable to BSP in preserving outliers, it cannot capture the watermark in the detail layer and enhance it. One reason for the poor performance of SPP is that it is designed for abstracting the visual complexity of multi-class scatterplots rather than single-class scatterplots.

The results show that BSP-VF-* only performs significantly better than the other methods on the *find maximum* task and performs similarly as BSP-GS-*, SSP-d, and CDP on the *density comparison/estimation* tasks. This is unexpected, since we initially thought that the variance-aware filter could better preserve relative data densities. After carefully investigating the tasks of *density comparison/estimation* and *find maximum*, we found that the former is mainly about the estimation of mean value within a region (50×50 px), where humans have a strong ability in making efficient judgments in scatterplots [35]. For the latter one, the user needs to identify the maximum within a smaller region (25×25 px), and our variance-aware filter can explicitly preserve them, especially for the structures with high variances. The results in Figs. 8(b), 8(c) show that bivariate colormaps facilitate density comparison and estimation more accurately.

4.2. Case studies

Using our interactive system, we conducted two case studies on real-world datasets using our BSPs with the variance-aware filter model and bivariate colormaps.

UK-Road-Safety. We used the annual road safety data published by the UK government [36] from 2005 to 2017 and obtained a density field containing 2,047,256 records, with the geographical locations (longitude and latitude) mapped to x and y axes, and the accident frequency at each location mapped to the density. As Fig. 9(a) shows, we first generated a density plot with the variance-aware filter and default parameters, in which most of the relative data densities are preserved. However, the road structures above London (the blue box) and low-density regions like the northwest area of Scotland (the green box) are not shown clearly. To further explore details in these regions, we set a large weight $\omega = 20$ for the selected regions. As Figs. 9(b), 9(c) shows, the road structures in the selected regions are further enhanced and become clearer, while the contrast of other regions is lowered.

arXiv. arXiv [37] contains abundant scholarly articles in the fields of physics, mathematics, computer science, etc. We took a snapshot of the repository and used the UMAP technique [38] to project it into a 2D scatterplot based on the similarity between document contents [39]; it contains 1,625,064 articles. We first visualized the density field using BSP with the default parameters. As shown in Fig. 10(a), the structures of major clusters (e.g., the green box) were not clear, and almost all pixels were red or green, indicating that the default hue mapping scheme shown in Fig. 10(b) does not show the absolute density distributions well. Hence, we adjusted the control points on the x-axis of the colormap, spreading more hues ranging in $[90, 270]$ to the absolute density range $[100, 400]$ (see Fig. 10(d)). As a result, some interesting structures, like the cross in the green box, became visible in Fig. 10(c). After further investigating the articles around this cross, we found that the path from ① to ④ consisted of articles from *Astrophysics* to *General Relativity* & *Quantum Cosmology*, while the path from ③ to ② consisted of articles that also started from *Astrophysics* but connected to *High Energy Physics*. We can therefore speculate about the relationship between these sub-fields in Physics.
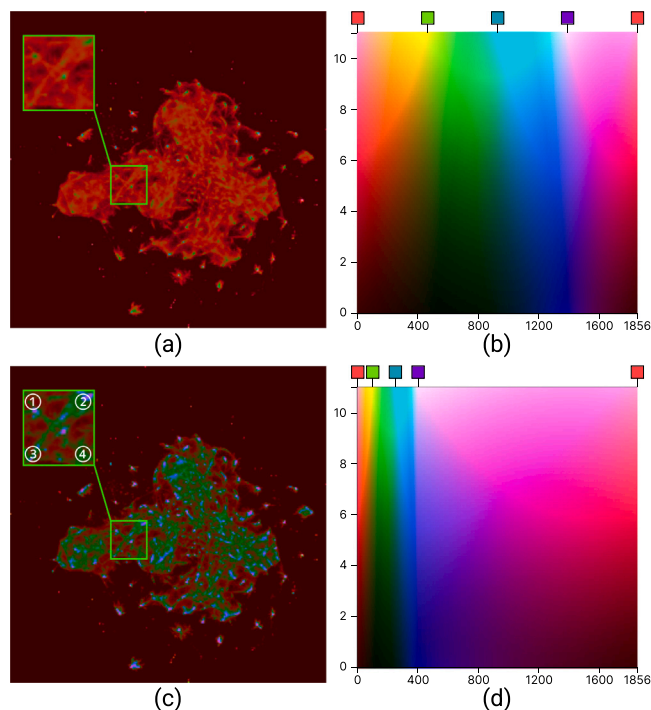
Fig. 10. Interactive exploration of the *arXiv* dataset with their corresponding bivariate colormaps. (a) Our BSP plot generated by the original bivariate colormap (b); and (c) improved BSP that shows more structures generated by adjusting the bivariate colormap with more hue in the value range $[100, 400]$ (d).

5. Conclusion & future work

We presented bi-scale density plots (BSP), a new technique to enhance density plots at the overview level of scatterplots. We first developed the unified bin–summarize–decompose–combine framework for enhancing density plots. Then, we took the density variance into the guided filter model and formulated the variance-aware filter, which explicitly preserves both the data maxima and relative densities. To improve the efficiency of density comparison and lookup, we further adopted a 2D colormap strategy using the hue and lightness channels to encode the absolute densities and improved densities. We evaluated the approach through a controlled user study, which demonstrates that our BSPs help better preserve the relative data densities, maxima, outliers, and local patterns. Also, we developed a web-based tool that implements our approach and demonstrated its effectiveness for the interactive exploration with two case studies.

There are still some limitations in our technique. First, our default parameters and 2D colormaps cannot reveal all the important structures, as shown in Section 4.2. We plan to infer better default parameters based on density distributions automatically. Second, our controlled study tested two state-of-the-art density enhancement methods: Splatterplots (SPP) and Sunspot plots (SSP). This choice was made to mitigate fatigue in participants. Whether other density plot improvement techniques (e.g., shaded SSP [2]) would lead to similar results remains to be seen. Third, the bivariate colormap limits us to single-class scatterplots; however, our framework with the variance-aware filter can be easily extended for multi-class scatterplots. Also, our framework is not limited to the Gaussian filter and guided filter. We will explore the possibility of adopting state-of-the-art learning-based filters such as [40].

CRediT authorship contribution statement

Huaiwei Bao: Writing – review & editing, Investigation, Formal analysis, Data curation, Conceptualization. **Xin Chen:** Writing – review & editing, Writing – original draft, Software, Methodology, Conceptualization. **Kecheng Lu:** Validation, Methodology, Formal analysis, Conceptualization. **Chi-Wing Fu:** Writing – original draft, Visualization, Validation, Methodology, Conceptualization. **Jean-Daniel Fekete:** Writing – original draft, Methodology, Data curation, Conceptualization. **Yunhai Wang:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Conceptualization.

Declaration of competing interest

I have nothing to declare.

Acknowledgments

Appendix A. Supplementary material

The full evaluation results, user study website, and the analysis code are available at https://osf.io/7qbhe/?view_only=ffe2ab92df354c43b9 84953e42ce3af0.

Data availability

I have shared the link to my data/code at

Bi-Scale Density-Plot Enhancement based on Variance-Aware Filter (O riginal data) (GitHub)
Bi-Scale Density-Plot Enhancement based on Variance-Aware Filter (O riginal data) (OSF)

References

[1] Mayorga A, Gleicher M. Splatterplots: Overcoming overdraw in scatter plots. IEEE Trans Vis Comput Graph 2013;19(9):1526–38. http://dx.doi.org/10.1109/TVCG.2013.65.

[2] Trautner T, Bolte F, Stoppel S, Bruckner S. Sunspot plots: Model-based structure enhancement for dense scatter plots. Comput Graph Forum 2020;39(3):551–63. http://dx.doi.org/10.1111/cgf.14001.

[3] Willems N, Van De Wetering H, Van Wijk JJ. Visualization of vessel movements. Comput Graph Forum 2009;28(3):959–66. http://dx.doi.org/10.1111/j.1467-8659.2009.01440.x.

[4] Elmqvist N, Dragicevic P, Fekete J-D. Color lens: Adaptive color scale optimization for visual exploration. IEEE Trans Vis Comput Graph 2011;17(6):795–807. http://dx.doi.org/10.1109/TVCG.2010.94.

[5] Sarikaya A, Gleicher M. Scatterplots: Tasks, data, and designs. IEEE Trans Vis Comput Graph 2018;24(1):402–12. http://dx.doi.org/10.1109/tvcg.2017.2744184.

[6] Chen X, Zhang J, Fu C-W, Fekete J-D, Wang Y. Pyramid-based scatterplots sampling for progressive and streaming data visualization. IEEE Trans Vis Comput Graph 2022;28(1):593–603. http://dx.doi.org/10.1109/TVCG.2021.3114880.

[7] Padilla L, Quinan PS, Meyer M, Creem-Regehr SH. Evaluating the impact of binning 2d scalar fields. IEEE Trans Vis Comput Graphics 2016;23(1):431–40. http://dx.doi.org/10.1109/TVCG.2016.2599106.

[8] Trautner T, Sbardellati M, Stoppel S, Bruckner S. Honeycomb plots: Visual enhancements for hexagonal maps. In: Bender J, Botsch M, Keim DA, editors. Vision, modeling, and visualization. The Eurographics Association; 2022, p. 65–73. http://dx.doi.org/10.2312/vmv.20221205.

[9] Kindlmann G, Scheidegger C. An algebraic process for visualization design. IEEE Trans Vis Comput Graph 2014;20(12):2181–90. http://dx.doi.org/10.1109/TVCG.2014.2346325.

[10] He K, Sun J, Tang X. Guided image filtering. IEEE Trans Pat Anal Mach Int 2013;35(6):1397–409. http://dx.doi.org/10.1109/TPAMI.2012.213.

[11] Matejka J, Anderson F, Fitzmaurice G. Dynamic opacity optimization for scatter plots. In: Proceedings of the SIGCHI conference on human factors in computing systems. ACM; 2015, p. 2707–10. http://dx.doi.org/10.1145/2702123.2702585.

[12] Micallef L, Palmas G, Oulasvirta A, Weinkauf T. Towards perceptual optimization of the visual design of scatterplots. IEEE Trans Vis Comput Graph 2017;23(6):1588–99. http://dx.doi.org/10.1109/TVCG.2017.2674978.

[13] Silverman BW. Density estimation for statistics and data analysis. London: Chapman & Hall; 1986.

[14] Heer J. Fast & accurate Gaussian kernel density estimation. In: 2021 IEEE visualization conference. VIS, IEEE; 2021, p. 11–5. http://dx.doi.org/10.1109/VIS49827.2021.9623323.

[15] Wickham H. Bin-summarise-smooth: a framework for visualising large data. Tech. Rep, 2013, had.co.nz.

[16] Jo J, Vernier F, Dragicevic P, Fekete J-D. A declarative rendering model for multiclass density maps. IEEE Trans Vis Comput Graph 2019;25(1):470–80. http://dx.doi.org/10.1109/TVCG.2018.2865141.

[17] Feng Z, Li H, Zeng W, Yang S-H, Qu H. Topology density map for urban data visualization and analysis. IEEE Trans Vis Comput Graphics 2021;27(2):828–38. http://dx.doi.org/10.1109/TVCG.2020.3030469.

[18] Xue Y, Paetzold P, Kehlbeck R, Chen B, Kwan KC, Wang Y, Deussen O. Reducing ambiguities in line-based density plots by image-space colorization. IEEE Trans Vis Comput Graphics 2024;30(1):825–35. http://dx.doi.org/10.1109/TVCG.2023.3327149.

[19] Staib J, Grottel S, Gumhold S. Enhancing scatterplots with multi-dimensional focal blur. Comput Graph Forum 2016;35(3):11–20. http://dx.doi.org/10.1111/cgf.12877.

[20] Jain P, Tyagi V. A survey of edge-preserving image denoising methods. Inf Syst Front 2016;18:159–70. http://dx.doi.org/10.1007/s10796-014-9527-0.

[21] Durand F, Dorsey J. Fast bilateral filtering for the display of high-dynamic-range images. In: ACM trans. graph.. SIGGRAPH, 2002, p. 257–66. http://dx.doi.org/10.1145/566654.566574.

[22] Bertini E, Santucci G. Give chance a chance: Modeling density to enhance scatter plot quality through random data sampling. Inf Vis 2006;5(2):95–110. http://dx.doi.org/10.1057/palgrave.ivs.9500122.

[23] Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng 2007;9(03):90–5.

[24] Kambhatla N. Local models and Gaussian mixture models for statistical data processing. Oregon Graduate Institute of Science and Technology; 1996.

[25] Schanda J. CIE colorimetry. Color: Underst CIE Syst 2007;25–78. http://dx.doi.org/10.1002/9780470175637.

[26] Sharma G, Rodríguez-Pardo CE. The dark side of CIELAB. In: Color imaging XVII: displaying, processing, hardcopy, and applications. Vol. 8292, SPIE; 2012, p. 94–103. http://dx.doi.org/10.1117/12.909960.

[27] Munsell AH. A color notation. Munsell color company; 1919.

[28] Taheri S, Vedienbaum A, Nicolau A, Hu N, Haghighat MR. OpenCV.js: Computer vision processing for the open web platform. In: Proceedings of the 9th ACM multimedia systems conference. 2018, p. 478–83. http://dx.doi.org/10.1145/3204949.3208126.

[29] Nikolaidis N, Pitas I. Robust image watermarking in the spatial domain. Signal Process 1998;66(3):385–403. http://dx.doi.org/10.1016/S0165-1684(98)00017-6.

[30] Abbas MM, Aupetit M, Sedlmair M, Bensmail H. *Clustme*: A visual quality measure for ranking monochrome scatterplots based on cluster patterns. Comput Graph Forum 2019;38(3):225–36. http://dx.doi.org/10.1111/cgf.13684.

[31] Mill B. Boston Marathon raw data. 2014, https://github.com/llimllib/bostonmarathon.

[32] Kaluža B, Mirchevska V, Dovgan E, Luštrek M, Gams M. An agent-based approach to care in independent living. In: International joint conference on ambient intelligence. Springer; 2010, p. 177–86. http://dx.doi.org/10.1007/978-3-642-16917-5_18.

[33] Eiselmayer A, Wacharamanotham C, Beaudouin-Lafon M, Mackay WE. *Touchstone2*: An interactive environment for exploring trade-offs in HCI experiment design. In: Proceedings of the SIGCHI conference on human factors in computing systems. 2019, p. 1–11. http://dx.doi.org/10.1145/3290605.3300447.

[34] Cox DR, Reid N. The theory of the design of experiments. CRC Press; 2000, http://dx.doi.org/10.1201/9781420035834.

[35] Gleicher M, Correll M, Nothelfer C, Franconeri S. Perception of average value in multiclass scatterplots. IEEE Trans Vis Comput Graph 2013;19(12):2316–25. http://dx.doi.org/10.1109/TVCG.2013.183.

[36] Department for Transport (DfT), UK. Road safety. 2023, https://www.data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data.

[37] McKiernan G. Arxiv.org: the Los Alamos National Laboratory e-print server. Int J Grey Lit 2000;1(3):127–38. http://dx.doi.org/10.1108/14666180010345564.

[38] McInnes L, Healy J, Melville J. UMAP: Uniform manifold approximation and projection for dimension reduction. 2018, http://dx.doi.org/10.48550/arXiv.1802.03426, arXiv preprint arXiv:1802.03426.

[39] Caillou P, Renault J, Fekete J-D, Letournel A-C, Sebag M. CARTOLABE: A web-based scalable visualization of large document collections. IEEE Comput Graph Appl 2021. http://dx.doi.org/10.1109/MCG.2020.3033401, URL https://hal.inria.fr/hal-02499006.

[40] Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior. In: Proc. IEEE conf. on comp. vis. and pat. rec.. 2018, p. 9446–54. http://dx.doi.org/10.1109/CVPR.2018.00984.