

Blue Noise Plots

¹Christian van Onzenoort

²Gurprit Singh

¹Timo Ropinski

³Tobias Ritschel

¹Ulm University

²Max-Planck Institute for Informatics, Saarbrücken

³University College London

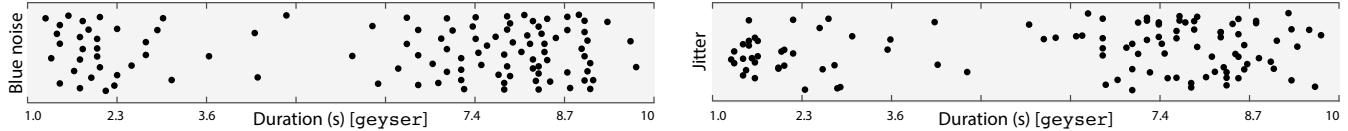


Figure 1: Blue Noise Plots (**left**) prevent clutter and provide visually more appealing results than frequently used jitter plots (**right**). Importantly, Blue Noise Plots are unbiased, in the sense that no data point is ever changed and strictly all points of the sample are presented.

Abstract

We propose Blue Noise Plots, two-dimensional dot plots that depict data points of univariate data sets. While often one-dimensional strip plots are used to depict such data, one of their main problems is visual clutter which results from overlap. To reduce this overlap, jitter plots were introduced, whereby an additional, non-encoding plot dimension is introduced, along which the data point representing dots are randomly perturbed. Unfortunately, this randomness can suggest non-existent clusters, and often leads to visually unappealing plots, in which overlap might still occur. To overcome these shortcomings, we introduce Blue Noise Plots where random jitter along the non-encoding plot dimension is replaced by optimizing all dots to keep a minimum distance in 2D i. e., Blue Noise. We evaluate the effectiveness as well as the aesthetics of Blue Noise Plots through both, a quantitative and a qualitative user study. The Python implementation of Blue Noise Plots is available [here](#).

1. Introduction

Consider depicting a univariate data set, e. g., observed ages in a cohort, on paper. While we could simply report first order statistics, such as for instance the mean, this would be an oversimplification in many important cases [Tal07]. Instead, we would like to show all data points, and thus ask for the optimal way to represent these with dots in a two-dimensional plot.

A *strip plot* simply plots data points on a single horizontal axis, while an observer is free to apply domain knowledge to judge what is the density, what might be modes or what might be outliers. Recent efforts, to communicate data to a wider public and not just to experts have resulted in such plots to be used increasingly in print media, television and on the web. Strip plots are most effective when the number of data points is still low enough to be displayed, but high enough to represent the information. The first row of Fig. 2 shows examples of strip plots on a univariate toy data set. Unfortunately, the main disadvantage of using strip plots to depict such data sets, is *clutter* which often leads to overdraw. As such, if two dots x_i and x_j are closer than what the printer, display or the human visual system can reliably discern, the advantage is lost, since not all data points are effectively communicated. For instance, the data set $\{1, 2, 4\}$ results in the same visual representation, as the data set $\{1, 2, 2, 4, 4\}$,

because the individual dots representing the values 2 and 4 would not be discernible when plotting the second data set, due to overdraw.

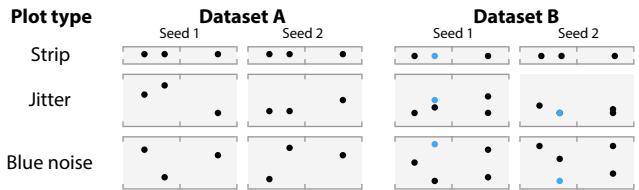


Figure 2: Three different plotting approaches (**rows**) for two different univariate datasets at two different random seeds (**columns**). Strip plots always fail to convey the datasets, as they look the same. Jitter plots, may sometimes fail: dataset B looks like dataset A for seed 2 while they are different. Our Blue Noise Plot never fails.

As a remedy, *jitter plots* have been proposed [Cha18, Tuk77], which introduce an additional, non-encoding dimension, along which dots are randomly perturbed. For our example, of univariate data points plotted in two dimensions, this means to simply move the respective dots vertically by a random amount. The second row of Fig. 2 shows examples of such jitter plots on the before data sets. While jitter plots usually reduce the amount of clutter and are

easy to implement, they have three main drawbacks. First, the introduced randomness leads to gaps and clusters, which might be falsely perceived as features present in the data set. Second, no minimum distance between dots is enforced, which in the worst case might lead to overlap, something we have observed in many real world examples. Lastly, jitter plots do often look visually unappealing.

By introducing Blue Noise Plots, we provide a solution to the three main drawbacks of jitter plots. A conceptual comparison is made in the last row of Fig. 2. Computer graphics has explored *blue noise* [Uli88], that is, dot patterns, that are still random, but without dots getting to close to each other. By proposing a modified Lloyd relaxation algorithm [Llo82], we can extend jitter plots to become Blue Noise Plots. Importantly, Blue Noise Plots are *unbiased*, in the sense that we make no assumption on how data points are represented, no data point is ever changed and strictly all data points of a data set are presented, setting it apart from other methods [BS04, BW08, KHD*10, MG13, MPOW17], that either re-sample or change the way data points are presented.

To investigate the impact of Blue Noise Plots on task performance and visual appeal, we have conducted a crowd sourced user study, whereby the obtained results indicate, that Blue Noise Plots are beneficial over conventional jitter plots. Furthermore, we have performed a qualitative study assessing the visual appeal of Blue Noise Plots, and discuss objective quality measures.

2. Previous Work

Our work addresses the visualization problem of plotting univariate data sets using the computer graphics methodology of blue noise, both of which we will review now.

Visualizing univariate data sets. To visualize univariate data sets, several different plot types exist. We distinguish between direct and aggregated visualizations. While direct methods allow for the depiction of the individual data points forming the data set, aggregated methods only communicate the data set per se, often in an approximate manner. Several well known plots fall into the later category, such as for instance histograms [Pea94], *box plots* [Spe69], and *violin plots* [HN98]. Nevertheless, this category is not in the focus of our work, and we instead aim for direct techniques, which explicitly convey the existing data points. One of these techniques, that is frequently used are strip plots [Cle85], where data points are represented as symbols, usually dots or circles. As stated above, these symbols are simply plotted all along the same dimension, independent of the occurring densities. While strip plots are an intuitive way to convey univariate data sets in a direct manner, they come with severe limitations, as they suffer from overdraw in dense regions. As a consequence, they might introduce distortion, as the maximum density to be communicated is limited by this overdraw. There are two main approaches to deal with the problem of overdraw. The size and or alpha blending value of the representing symbols can be altered, or data points can be transformed such that overdraw is reduced. While naturally, these approaches can be combined, we solely focus on transformation approaches in this paper.

Among the techniques applying transformations, the *stacked dot plot* is the most basic, as it simply stacks the shown dots, which effectively resembles a bar chart for non-continuous data [Cle93, Wil99].

For continuous data sets, more elaborate packing schemes need to be employed, in order to obtain an acceptable layout. So-called *beeswarm plots* exploit such packing which enables them to obtain a stacked and dense representation without the need for binning, but their simple construction leads to false features and clustering.

In contrast to the stacked representations described above, jitter plots randomly distribute all points in a given range along an additional, non-encoding dimension [Tuk77]. Thus, it also becomes possible to reduce the amount of overdraw, to a degree dependent on the plot size and the occurring densities. Jittering originally goes back to Chamber et al. [Cha18], while Tukey and Tukey additionally exploit constraints [TT90].

Recently, several visualization approaches have been proposed, which deal with the shortcomings of dot plots in general. Bachthaler and Weiskopf have introduced continuous scatter plots, which sacrifice the discrete nature of scatter plots, in order obtain a dense visualization [BW08]. Mayorga and Gleicher go further by automatically grouping some dots, while keeping others [MG13], which is combined with interactive exploration. Along a similar line, Bertini and Santucci introduce non-uniform samplings in order to communicate density in 2D scatter plots [BS04]. Keim et al. introduce generalized scatter plots, where the dot locations are modified in order to reduce overlap [KHD*10], which in contrast to our approach, this adds bias to the actual data values. Micallef present an optimization approach for the perceptual optimization of scatter plots [MPOW17]. By exploiting task-dependent cost functions, they are able to obtain satisfying scatter plot designs, which compete with those crafted by humans. Yuan et al. [YXX*20] and Rapp et al. [RPD20] also address the problem of overdraw in scatter plots where they select a subset of data points from a large data set such that the resulting patterns follows the density, yet has blue noise. Our work does not select subsets, but shows all data points, and introduce an additional, non-encoding dimension so that dots can become blue noise in the first place.

Some forms of box plot also communicate individual data points, e.g., when they are outliers, and thus can be considered a combination of aggregated and direct visualization. Many of the contributions made in this paper, can also be applied in this context.

Blue noise. Dot patterns are often described in terms of their expected power spectrum profile. Patterns exhibiting mostly high-frequency content in their power spectrum are characterized as *blue noise*. The resulting spatial distribution of dots respect some minimum distance giving perceptually pleasing patterns [Yel83]. Consequently, blue noise has been widely adopted in many computer graphics applications including halftoning [Uli88], stippling [Sec02], artistic packing [RRS13], anti-aliasing [DW85] and variance reduction for Monte Carlo rendering [SOA*19]. Typically, such methods are free to place dots in whatever arrangement, only their correlation and in some cases an importance function is relevant. Also the dot count is typically not fixed strictly. An exception is the work by Reinert et al. [RRS13], who initialize Lloyd relaxation with coordinates that are correlated with a feature vector of a specific set of elements. In comparison to our approach, this is biased, i. e., dots are free to change position or order, defying the purpose of a jitter plot altogether, which is to link individual dots to definitive data points. Our work shares similarity with Reinert et al.

Algorithm 1 Lloyd relaxation	Algorithm 2 Jitter Plots	Algorithm 3 Blue Noise Plots
<pre> 1: $P \leftarrow \text{uniform}()$ 2: repeat 3: $V = \text{voronoi}(P)$ 4: for $p_i \in P$ do 5: $p_i \leftarrow 0$ 6: for $q_j \in V_i$ do 7: $p_i \leftarrow p_i + q_j / V_i$ 8: end for 9: end for 10: 11: until converged 12: plot(P) </pre>	$+$	<pre> 1: $P \leftarrow \text{stack}(X, \text{uniform}())$ 2: 3: 4: 5: 6: 7: 8: 9: 10: 11: plot(P) </pre>

Figure 3: Our approach (**right**) is a combination of Lloyd relaxation (**left**) and jitter plots (**middle**) with a data constraint extension. The function $\text{stack}(A, B)$ stacks vector A on top of vector B . The function $\text{unstack}(A, B)$ returns the vector A with the dimensions from B removed.

[RRSG15] that treat different dimensions in the dot optimization differently, but without complying to data points, only optimizing for different spectra. In this work, we show that dots distributed while obeying to blue noise provide better data visualizations. Therefore, we optimize the dot layout using Lloyd relaxation [Llo82] to obtain a blue noise distribution. However, unlike traditional approaches, our optimization works by keeping the encoding dimension fixed, while optimizing the dot positions along the other, non-encoding dimension. Our optimization runs in higher dimension than the data and uses a novel distance metric to emphasize the non-encoding dimension to guide the optimization. We already have mentioned Hu et al. [HSVK*19] as a rare example of a visualization paper that relates to blue noise. Their task might be easier than ours, as they assume sampling from multiple importance functions to produce multi-class blue noise, without adhering to ordinality or coordinates.

3. Blue Noise Plots

We will here give a formal definition of our approach, starting from in- and output (Sec. 3.1), to a variational formulation with constraints (Sec. 3.2) and our implementation to minimize it (Sec. 3.3). The section concludes with several extensions, such as centrality, resembling bee swarm plots, automatically choosing the plot height (Sec. 3.4), as well as introducing a multi-class version (Sec. 3.5).

3.1. Input and Output

Input to our method is a set $X = \{x_i \in \mathbb{R}\}$ of univariate data points, where values are assumed to be associated with the horizontal axis. Output is a set $Y = \{y_i \in \mathbb{R}\}$ of scalar jitter values associated with the vertical axis. We write $P = \{\mathbf{p}_i = (x_i, y_i) \in \mathbb{R}^2\}$ to denote the combination of data values along the horizontal (encoding) dimension, and jitter value along the vertical (non-encoding) dimension into a set of two-dimensional dots.

3.2. Cost Function

Optimization is performed for the set of vertical jitter values Y of an output Blue Noise Plot, given the univariate input data set X .

We minimize the cost

$$\arg \min_Y \sum_{y_i \in Y} \mathbb{E}_{\mathbf{q} \sim V(\mathbf{p}_i, P)} \kappa(\mathbf{p}_i, \mathbf{q}), \quad (1)$$

the sum of the expected value \mathbb{E} of the κ -distance from the i -th output dot \mathbf{p}_i to all sites \mathbf{q} in its Voronoi cell $V(\mathbf{p}_i, P)$ in respect to all other dots P .

Uniform metric. For classic Blue Noise Plots, we use

$$\kappa(\mathbf{p}_1, \mathbf{p}_2) = \|(\mathbf{p}_1 - \mathbf{p}_2)^T \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\|_1. \quad (2)$$

Here, the constant diagonal matrix, emphasises the non-encoding dimension along the vertical direction. We will in Sec. 3.4 introduce further metrics to realize other variants.

3.3. Optimization

While two-dimensional Lloyd relaxation [Llo82] would minimize the distance cost, it unfortunately does not adhere to the hard constraints. In Lloyd relaxation, after a random initialization (Alg. 1, Line 1) every dot is iteratively replaced by its Voronoi cell center (Alg. 1, Line 2 to 10), followed by a re-computation of the Voronoi cells (Alg. 1, Line 3) in a expectation-maximization procedure [DLR77]. Running it directly, will loose the information present in the data sets, such as trivially done by jitter plots (Alg. 2).

Our main contribution is a solver that extends Lloyd relaxation to produce dot patterns with even distribution that adhere to the data, as explained in (Alg. 3). Including the hard constraint is intuitive: use both dimensions for the relaxation (Alg. 3, Line 1) but never update the encoding dimension (Alg. 3, Line 10). During optimization, dots move vertically, with a single degree of freedom per dot, but their cost computation, including the Voronoi construction, involves 2D.

Note, how this is different from optimizing only an 2D dot pattern and not involving the 1D information. The information is not updated, but it is crucial to include it in the distance computation, such as to satisfy the Lloyd objective in what is perceived: 2D space.

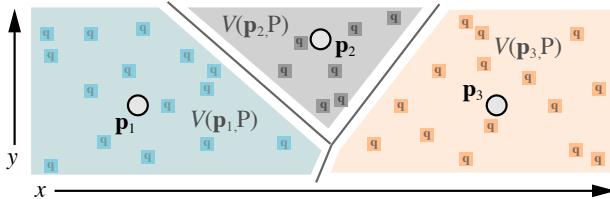


Figure 4: Lloyd relaxation involves Voronoi cells V (blue, grey and orange areas) computed on the output dot pattern P . These V are sampled using sites \mathbf{q} shown here as rectangles. The optimization relates every dot \mathbf{p}_i to all the \mathbf{q}_j in its cell $V(\mathbf{p}_i, P)$.

In practice, different options to construct Voronoi regions exist. We follow the approach from Balzer et al [BSD09] but without the capacity constraint. We use 8,192 random 2D points \mathbf{q} to discretize the domain. Note that faster GPU methods exist, that make use of regular grids [HIKL*99, LNW*10].

3.4. Automatic Height

The output dots P are from a domain that typically is wider than high when using the horizontal axis as the encoding axis, and the vertical axis as the non-encoding dimension. This is, as most datasets have only a couple of different data points per interval in the data dimension, compared to the total number of data points. Unfortunately, it might not be obvious how to choose an appropriate height, both for jitter plots, as well as for Blue Noise Plots for a given data set. The desired distance between dots is part of the problem definition. Making recommendations how to choose the distance of two dots such that they become visually discernible is clearly an often-encountered visualization challenge, but out-of-scope for this work. We will assume it to be known and use a distance of two-times the dot radius in all results we show. We will now show, how to choose the plot height automatically and optimally.

We assume access to the *density* function $d(x) \in \mathbb{R} \rightarrow \mathbb{R}^+$, defined on the domain of the data distribution (black in Fig. 5). Typically, we are only given a sample of the distribution not the true density. Hence the density function needs to be estimated from the sample, for which we use

Kernel Density Estimation in practice. Please note, that we do not in general rely on density estimation, unless the plot height is chosen automatically or we optimize for centrality (Sec. 3.4.1).

The optimal height depends on the maximal density d_{\max} , the total number of data points n , and the desired distance r between dots. The desired distance is chosen by the user (pink in Fig. 5). It depends on the output medium, whereby a typical choice is to make it twice the size of a dot, so they become discernable. We note that at the data coordinate with d_{\max} , we need to “stack” $r \cdot d_{\max} \cdot n$ dots (orange in Fig. 5). Ignoring optimal packings with efficiency around

0.9, and assuming a conservative efficiency of 1 instead, stacking such dots with radius r needs to be $r^2 \cdot d_{\max} \cdot n = h$ (green, Fig. 5).

3.4.1. Centrality

The point density in a Blue Noise Plot might vary. Alternatively, we can restrict the points to not use all the space available due to the non-encoding dimension. To this end, for a fixed width (manual or automatic), we limit dots to move less along the non-encoding dimension, resembling the appearance of beeswarm plots. We refer to such a plot as having *centrality*.

Choosing a varying height is based on the generalization of the aforementioned automatic height (Sec. 3.4). Instead of choosing a single height value h , we choose height as a function $h(x)$ of the data coordinate x itself.

This idea is conveniently realized by changing the metric itself to be non-uniform. How distant two dots are, is depending on the density at the coordinate $x_{12} = (x_1 - x_2)/2$ between them:

$$\kappa(\mathbf{p}_1, \mathbf{p}_2) = \|(\mathbf{p}_1 - \mathbf{p}_2)^T \begin{pmatrix} d(x_{12}) & 0 \\ 0 & 1 \end{pmatrix} \|_1 \quad (3)$$

Note, that \mathbf{p}_1 and \mathbf{p}_2 are typically close, so even while the function is not a metric for all pairs, it locally is as the density function is smooth. In particular it can be chosen arbitrarily smooth by using a smooth kernel in the density estimation, that, in the limit, corresponds to the constant height.

3.5. Multi-Class

For data sets comprised of different classes, Blue Noise Plots can be extended to multi-class blue noise [Wei10]. Here, all data points in one class maximize Eq. 1, as well as all possible unions of all data points in all classes do. The solver implements this by alternating between the individual classes and their unions.

4. Results

We present both qualitative and quantitative results of our work.

Single-class. We show results of our, as well as existing approaches on typical data sets in Fig. 6. We see, that our approach does minimize the amount of visual clutter in the form of overlap for all examples. While this is particularly the case in denser regions, even in sparser regions (for example in Fig. 6, c) on the right), dots are more evenly distributed over the available domain, supporting the perception of individual dots.

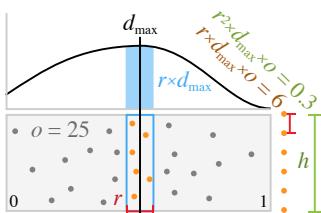


Figure 5: Auto-height (see text).

Automatic height. As described in Sec. 3.4, we dynamically adapt the height of the plot, depending on the number of dots and their distribution. While Fig. 7 shows an example of differently sized subsets of the geyser data set using a constant height for the plot, Fig. 8, shows the dynamic adaption. We do so to compromise between a compact plot and room for the dots to relax and therefore reduce overlap. If the number of data points gets large, retaining a fixed distance is only possible at the expense of a high plot.

Centrality. Fig. 9 shows results where the height is chosen automatically, but varying with the data dimension. Depending on the

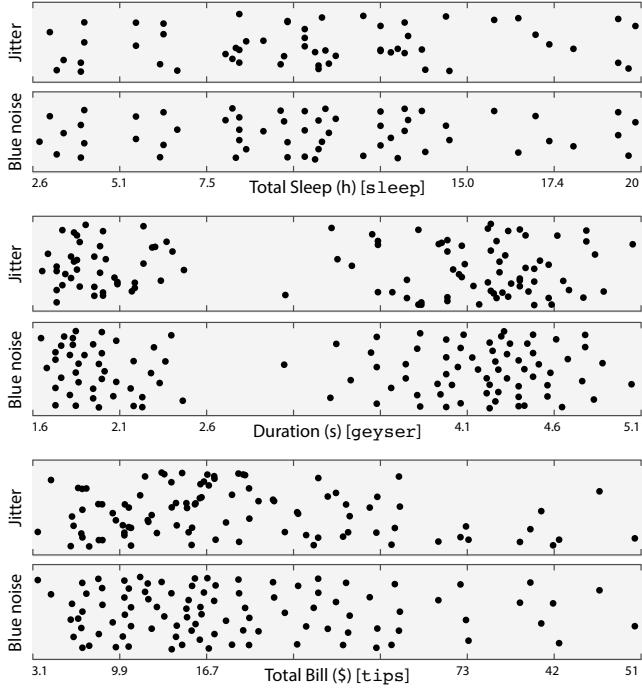


Figure 6: Comparison of three different data sets, each of them visualized using a traditional jitter plot and our Blue Noise Plot.

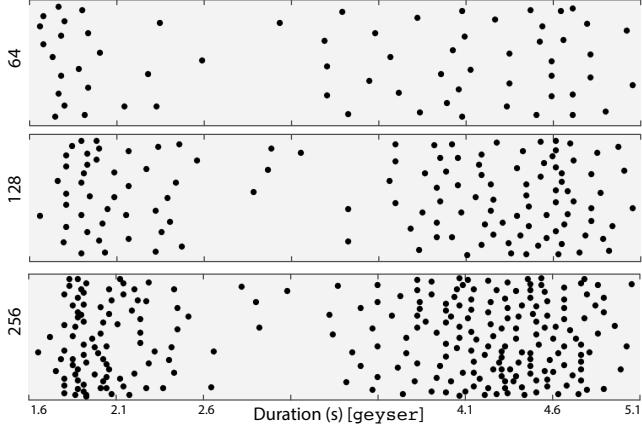


Figure 7: Comparison of plots with different numbers of dots. All plots are drawn using the same height, but with different numbers of dots. These plots show a random subset of the geyser data set, visualized using our Blue Noise Plot. Here a) shows 64 dots, b) shows 128 dots and finally c) shows 256 dots.

reliability of the underlying density estimation, this can be an effective additional cue. At any rate, adding blue noise improves upon jitter in readability and aesthetics.

Multi-class. Finally, we show an extension to multiple classes of data points. Here, every input point additionally has a class label. We use our method to produce a plot that is blue noise for all classes jointly, as well as for every class on its own. Fig. 10 shows

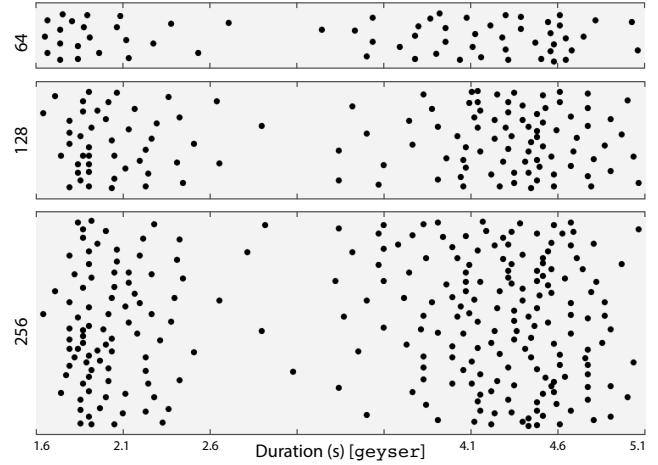


Figure 8: Comparison of adaptive plots with different numbers of dots. These plots show a random subset of the geyser data set, visualized using our Blue Noise Plot. Here a) shows 64 dots, b) shows 128 dots and finally c) shows 256 dots.

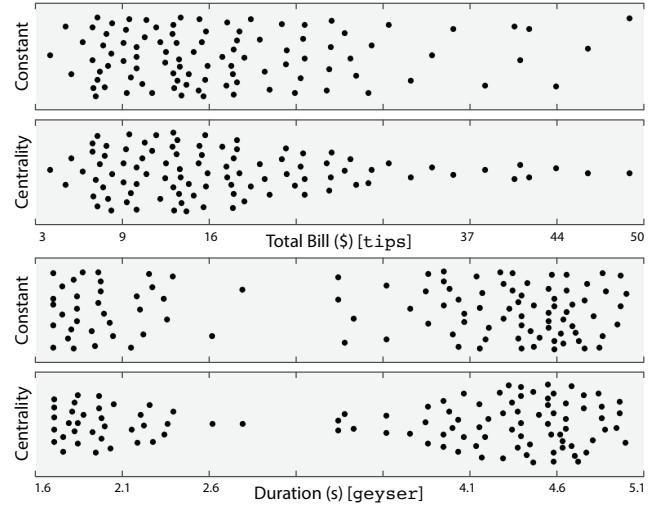


Figure 9: Optimal constant plot height (first and third) and a varying height (second and fourth), both for Blue Noise Plots.

an example of this where the first two rows show the blue noise distributed dots of the individual classes. The third row, shows the final plot, where the first two rows are combined. Fig. 11 shows more examples of our Blue Noise Plots, encoding multiple classes. Here, we can also see that our approach nicely distributes all the dots, as well as the dots for the individual classes over the entire domain. Further, Fig. 11 and Fig. 12 show examples of quantized data. While the blue noise pattern is less prominent in this case, this shows another strength of our approach. In contrast to jitter plots, where the overlap between dots is amplified by this type of data distribution, our approach spreads out the dots vertically.

Icons. Inspired by approaches which represent data points with more complex primitives instead of dots [HHD03, RRS13], we have

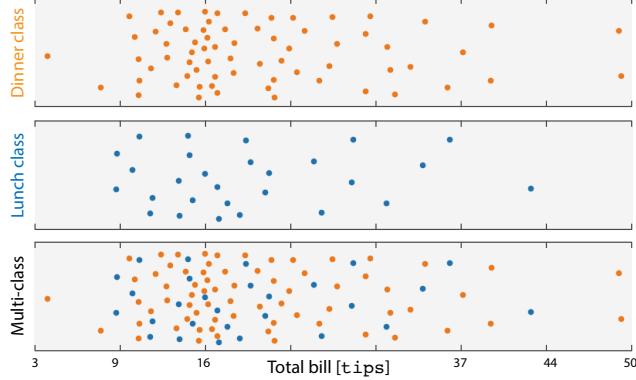


Figure 10: Multi-class Blue Noise Plot for the *tips* data set with two classes: dinner and lunch encoded into color. The first two rows show the individual, the third the combined plot. Please note, how this is three visualizations of one set, fulfilling all intra- and inter-class, as well as the data constraints simultaneously.

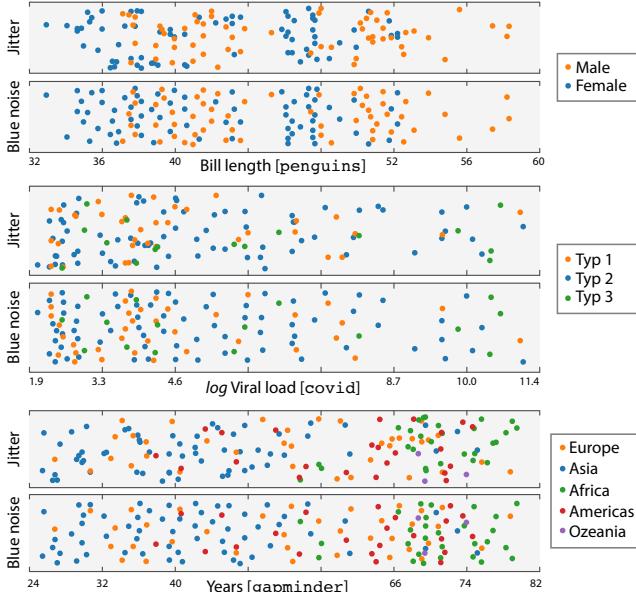


Figure 11: Different examples of multi-class data sets, visualized using jitter plots as well as Blue Noise Plots.

used our method to place icons as seen in Fig. 13. Here, every data point has a unique icon, making relations visible without bias or clutter.

Parameter choice. A typical result of a data set containing 256 data points, as shown in this paper, requires 40 iterations of Lloyd relaxation, with 8,192 Voronoi samples, resulting in a total time of six seconds for a naive, non-parallel implementation.

Analysis. We analyze of our results both from the graphics perspective using the spectral quality of Blue Noise Plot as dot patterns, as well as with overlap measures used to analyze plots.

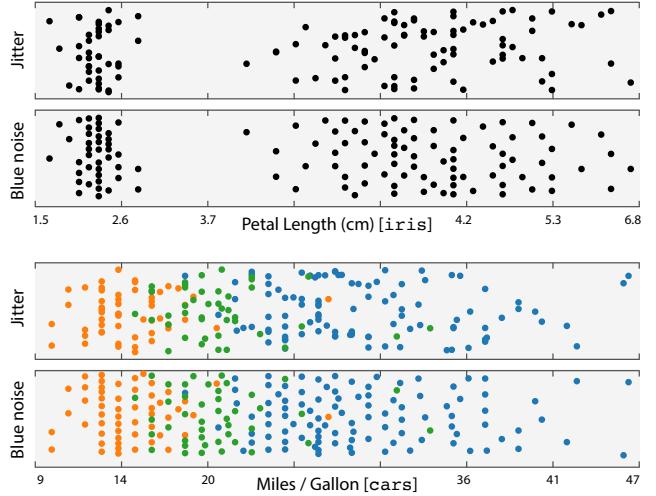


Figure 12: Quantized data sets, such as shown here, where not many x values are shared by a data point, are difficult to optimize for, but worth addressing: besides being less visually appealing in many cases, according to our study, jitter plots, due to clutter, are more difficult to read.

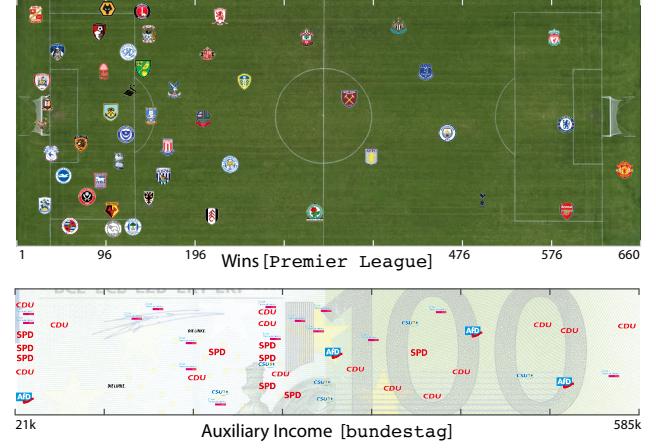


Figure 13: Blue Noise Plots can further be used to position extended primitives, instead of dots, here, little icons depicting soccer clubs or political parties.

To perform spectral analysis, we compute the expected power spectrum of dots obtained from the *geyser* dataset. We generate 100 different realizations of the dot patterns, compute their power spectrum and average these power spectra to get the expected spectrum. Fig. 14 shows these expected power spectra for OUR (right) and JITTER (left). We compare this against vanilla Lloyd relaxation (middle), which is not a plotting method, as it does not produce an unbiased result, but can serve as some upper bound of what we could achieve when using it as a backbone. For JITTER, the spectrum is flat like *white noise*. OUR approach gives a dark region in the middle of the spectrum confirming the *blue noise* behavior. The bright line in the middle is due to the non-uniform density of dots along the horizontal axis, where they obey to the data values. If we

run Lloyd relaxation without constraining the data along the horizontal axis, the dark region in the middle gets larger (middle) and we get uniform density points. That is why, there is no bright line in the middle spectrum. The anisotropic structure of the dark region is evident due to the non-square domain of the plot. A domain of length L has a valid spectrum at only $1/L$ -th frequencies [SSC*20]. In Fig. 14, since the plot along x -axis $\in [0, 1)$ and y -axis $\in [0, 0.2)$, the spectrum is valid only at integer frequencies along the x -axis and every $(1/0.2 =) 5$ -th frequency along y -axis. Lastly, the central dark line in the Lloyd relaxation spectrum (middle) implies denser stratification of the x -axis w.r.t. the y -axis.

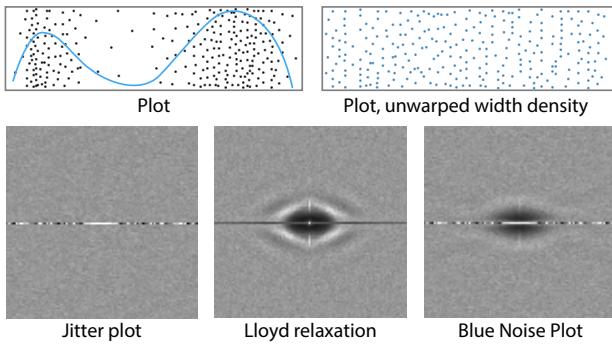


Figure 14: The top part shows the plot, its density function as a blue line and the unwarped plot points. When performing this on jitter, blue noise based on Voronoi and Blue Noise Plot, we find the three spectra seen. A well-distributed dot set should exhibit a low energy (black) in the low-frequency regions (center). While being inferior to Lloyd relaxation, we fair substantially better than jitter.

We also analyzed our plots using a point overlap metric presented for scatter plots [vOHR20]. In Fig. 15 we see this overlap metric (less is better) applied to the result of OUR and JITTER at different dot counts and for different datasets. This quantifies what was hinted at before: with JITTER, one can get almost-acceptable results as well as very bad ones (as seen by the high variance; now in a while JITTER might discover an accidental Blue Noise Plot) while OUR is consistently providing a low variance with less overlap. When dot count increases, variance of JITTER becomes less, but the gap to OUR becomes even wider.

5. User Study

To evaluate Blue Noise Plots, we conducted two user studies. The first is a preference study (Sec. 5.1), indicating that Blue Noise Plots are considered more appealing over jitter plots. The second is a threshold experiment (Sec. 5.2),

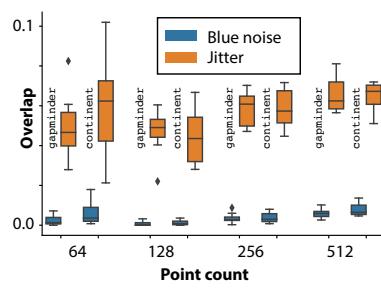


Figure 15: Overlap analysis (see text).

confirming that users are performing better to perceive the underlying distribution when using our method. In both experiments, we compare OUR approach (Alg. 3) to a JITTER baseline (Alg. 2).

5.1. Preference Study

Methods. To evaluate the visual preference, we conducted a user study with a total of $N = 12$ participants (3 female, 9 male, $M_{age} = 27.92, SD = 3.26$). These participants were recruited from a university setting, but without a particular expertise in visualization. They were presented with nine different data sets (*tips*, *titanic*, *iris*, *penguin*, *geyser*, *car*, *gapminder*, *tooth*, and *diamond*) visualized using both, OUR, as well as JITTER treatment, presented in a randomized side-by-side layout. They were asked two questions: First, to rate which one is “more visually appealing” on a choice-enforcing four-point Likert scale, ranging from “Strongly agree” to “Strongly disagree”. Second, to indicate which treatment, if any, they prefer.

Analysis. Analyzing responses to the first question using a Mann–Whitney U test, we find OUR ($Mdn = 2.0, IQR = 1.0$) to be significantly more visually appealing compared to JITTER plots ($Mdn = 1.5, IQR = 2.0, U = 4091.5$, significant $p < .01$). Looking at the individual responses, we find clear preferences (significant $p < .05$) for *tips* (difference of JITTER and OUR of 1.25) *penguin* (0.75) *iris* (0.50) *tooth* (0.58) *gapminder* (0.58) and lower responses (no significance) for *car* (0.08) *titanic* (0.16) *geyser* (0.25) and *diamonds* (0.25).

For the second question we found a preference of OUR technique in 62.04% of all responses, in 9.26% of the cases a preference towards the JITTER plot and 28.7% without a preference. When further analysing responses to the second question, for the individual data sets, we find preferences of (*diamond*: 83.33%, *gapminder*: 75.0%, *geyser*: 75.0%, *penguin*: 83.33%, *tips*: 83.33%). These data sets show evenly-distributed points (examples are seen in the Fig. 6, Fig. 8, and Fig. 11, indicating that our approach does support these situations the most. For sparse data sets, participants responded that they do not prefer one of the techniques (*car*: 58%, *tooth*: 58%), possibly due to the fact, that these data sets do not suffer from overdraw.

Free-text responses. Afterwards, we gave participants the option to respond to the following question using a free text field: “Do you prefer one of the options? If yes, why?”.

While analyzing the free text responses, we found that our participants appreciated the Blue Noise Plot not only being “prettier”, but also for being easier to understand. They for example stated, that the Blue Noise Plot is “definitely prettier”, it looks “cleaner and less noisy”, and “more organized”. Besides this aesthetic aspects, they also stated that the a Blue Noise Plot is “easier to understand”, that dots are “more easily distinguishable”, and “easier to count”.

This indicates that our approach might not only be more visually pleasing but also improves the understanding of the data, informing the next study to confirm these subjective judgements. Further studies would be required to understand preference for variants of our approach, such as centrality or multi-class patterns.

5.2. Performance Study

Methods. A total of $N = 232$ participants from the Amazon MTurk Masters population were simultaneously shown a dot plot on the left and two variants of a distribution to the right (Fig. 16). They were tasked to indicate in a two-alternative forced choice, which variant of the distribution corresponds to the dot plot. Dot plots were, randomly, either using OUR or JITTER. Distributions comprised of B-spline curves with five uniformly-placed control points drawn as line plots. Their variants result from choosing a random control point in every trial and perturbing it vertically by an offset O . In every trial, a staircase procedure (QUEST, [WP83]) was conducted to estimate the threshold of O i.e., at which level of difference of the reference distribution, different dot plots allow humans to answer correctly in 75 % of the cases. A successful treatment, would have a lower such threshold, which is the dependent variable we record in units of just-noticeable differences (JND) [OJEF18, HYFC14, KH15].

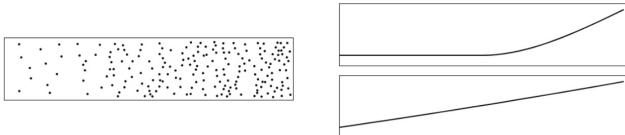


Figure 16: Experimental stimulus, showing a dot plot of a given data set on the left, and two reference distributions to the right.

Data preparation. For 72 participants the threshold experiment did not converge after 100 trials. In a staircase procedure without bounds this indicates they clicked randomly as any deterministic response will ultimately converge to a value, be it high or low. Filtering resulted in 160 valid responses. Based on timings from piloting, participants were paid \$2, for a target rate of \$8/hour.

Analysis. A Mann-Whitney U test finds a significantly smaller threshold for Blue Noise Plots ($Mdn = .34$, $IQR = .16$), compared to the jitter plot ($Mdn = .38$, $IQR = .22$, $U = 2698.0$, $p = .044$), rejecting the hypothesis that they have identical perceptual thresholds to convey a distribution as a dot plot.

Discussion. At first, this study design can appear contrived, and it can be asked why not perform a direct comparison. However, there is no single reliable offset O that is valid over all subjects, their viewing conditions, stimuli, training effects, etc. Hence no O could also be found in a pilot study or using any other process. Consequently, the difference to study needs to adapt to the conditions, and this is exactly what a staircase procedure does.

Next, one could wonder why JND is a measure of success. JND is the smallest change a channel (from algorithm over display to the human visual system) encodes. An efficient visual channel –such as we want our technique to be– aims to reproduce as many different values as possible, to maximize the entropy, realizing communication with a high bandwidth. Hence, our smaller JNDs indicate the task was made easier as detailed by van Onzenooldt et al. [OJEF18].

6. Discussion

Lloyd relaxation backbone. We use Lloyd relaxation as an admittedly simple means to achieve a blue noise spectrum. Many other refined techniques have been proposed

[DGBOD12, BSD09, Fat11, SGBW10, ÖG12, ZHWW12, LSM*19] to produce better blue noise patterns, in particular those, that support non-uniform importance. Strikingly, adding data constraints has not found consideration in the literature we are aware of. Lloyd, however, as an expectation maximization process is well suited to enforce constraints iteratively. We further could also use other discretizations of the domain, but this one is particularly amendable to the non-uniform metric used. We hope adding data constraints for visualization purposes could become a new and important sub-task to consider in computer graphics point design methods to come.

Point quality. We have shown many examples that clearly outperform jitter plots as a baseline found in countless papers printed every day. We further analyzed the dot quality according to state-of-the-art dot correlation metrics. Still the result quality of even the most naïve blue noise method could be considered superior to ours, but this is not a plausible comparison to make. General but biased graphics techniques can remove or add dots, move them freely, etc. making the task much easier than our unbiased setting. But even if there is a gap, it is not clear if the patterns we produce are actually any close to the best patterns we can hope for even with those additional constraints. A reader is encouraged to apply the blue noise Turing test: is it obvious how to move the points to make the pattern actually better for a human? We think, yes, maybe, but in many cases only by diminishing returns compared to what is the improvement over jitter. Future work might find optimization approaches to get point sets that are unbiased in our sense, yet at even higher spectral quality.

Visualization impact. Hu et al. [HSVK*19] and Reinert et al. [RRS13] have made links between placement of primitives according to data and distribution quality. Our work is ignorant of the way data points are ultimately presented, so it would be important to have a loop back and ask what size, color, icons or animation would allow for efficient visualization of a dot set, given the spectrum is now high-quality. In particular, our approach can cover higher dimensions, leading to further visualization questions. We think both our and their work will open up new problems and solutions in visualization optimizing for aesthetics and clutter avoidance.

7. Conclusions

We improve the visual appeal and functionality of jitter plots, by re-casting their randomization into an optimization procedure to put dots “nicely”, resulting in improved visual appeal and depiction of univariate data sets. During our user studies, we found that our Blue Noise Plot were not only considered to be visually more appealing compared to frequently used jitter plots, but easier to interpret. Our user study also supports our hypothesis that our plots enable a more accurate estimation of univariate data sets, compared to jitter plots.

While we use one encoding data dimension and one additional, non-encoding dimension to target the important case of 2D visualization, other combinations are possible. For 3D [SLC*18] or tangible [LIRC12] visualization, an optimization could be extended to fix two data dimensions and optimize a third one. In other future work, instances of randomization in visualization, e.g., in user interfaces, and Human-Computer interaction, even including the physical world, could be moved forward into problems where information is neither placed regular, nor random, but inspired by blue noise.

References

- [BS04] BERTINI E., SANTUCCI G.: By chance is not enough: preserving relative density through nonuniform sampling. In *Proc. InfoVis* (2004), pp. 622–629.
- [BSD09] BALZER M., SCHLÖMER T., DEUSSEN O.: Capacity-constrained point distributions: a variant of Lloyd’s method. *ACM Trans Graph* 28, 3 (2009), 1–8.
- [BW08] BACHTHALER S., WEISKOPF D.: Continuous scatterplots. *IEEE Trans Vis and Comp Graph* 14, 6 (2008), 1428–1435.
- [Cha18] CHAMBERS J. M.: *Graphical methods for data analysis*. CRC Press, 2018.
- [Cle85] CLEVELAND W. S.: *The elements of graphing data*. Wadsworth Publ. Co., 1985.
- [Cle93] CLEVELAND W. S.: *Visualizing data*. Hobart Press, 1993.
- [DGBOD12] DE GOES F., BREEDEN K., OSTROMOUKHOV V., DESBRUN M.: Blue noise through optimal transport. *ACM Trans Graph* 31, 6 (2012), 1–11.
- [DLR77] DEMPSTER A. P., LAIRD N. M., RUBIN D. B.: Maximum likelihood from incomplete data via the EM algorithm. *J Royal Statistical Society: Series B (Methodological)* 39, 1 (1977), 1–22.
- [DW85] DIPPÉ M. A. Z., WOLD E. H.: Antialiasing through stochastic sampling. In *Proc. SIGGRAPH* (1985), p. 69–78.
- [Fat11] FATTAL R.: Blue-noise point sampling using kernel density model. *ACM Trans Graph* 30, 4 (2011), 1–12.
- [HHD03] HILLER S., HELLWIG H., DEUSSEN O.: Beyond stippling—methods for distributing objects on the plane. In *Comp Graph Forum* (2003), vol. 22, pp. 515–22.
- [HIKL*99] HOFF III K. E., KEYSER J., LIN M., MANOCHA D., CULVER T.: Fast computation of generalized voronoi diagrams using graphics hardware. In *Proc. SIGGRAPH* (1999), pp. 277–86.
- [HN98] HINTZE J. L., NELSON R. D.: Violin plots: a box plot-density trace synergism. *The American Statistician* 52, 2 (1998), 181–184.
- [HSVK*19] HU R., SHA T., VAN KAICK O., DEUSSEN O., HUANG H.: Data sampling in multi-view and multi-class scatterplots via set cover optimization. *IEEE Trans Vis and Comp Graph* 26, 1 (2019), 739–748.
- [HYFC14] HARRISON L., YANG F., FRANCONERI S., CHANG R.: Ranking visualizations of correlation using weber’s law. *IEEE Trans Vis and Comp Graph* 20, 12 (2014), 1943–52.
- [KH15] KAY M., HEER J.: Beyond weber’s law: A second look at ranking visualizations of correlation. *IEEE Trans Vis and Comp Graph* 22, 1 (2015), 469–78.
- [KHD*10] KEIM D. A., HAO M. C., DAYAL U., JANETZKO H., BAK P.: Generalized scatter plots. *Information Visualization* 9, 4 (2010), 301–11.
- [LIRC12] LEE B., ISENBERG P., RICHE N. H., CARPENDALE S.: Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Trans Vis and Comp Graph* 18, 12 (2012), 2689–2698.
- [Llo82] LLOYD S.: Least squares quantization in PCM. *IEEE Trans Information Theory* 28, 2 (1982), 129–137.
- [LNW*10] LI H., NEHAB D., WEI L.-Y., SANDER P. V., FU C.-W.: Fast capacity constrained voronoi tessellation. In *Proc. ACM i3D* (2010).
- [LSM*19] LEIMKÜHLER T., SINGH G., MYSZKOWSKI K., SEIDEL H.-P., RITSCHEL T.: Deep point correlation design. *ACM Trans Graph (Proc. SIGGRAPH Asia)* 38, 6 (2019), 1–17.
- [MG13] MAYORGA A., GLEICHER M.: Splatterplots: Overcoming overdraw in scatter plots. *IEEE Trans Vis and Comp Graph* 19, 9 (2013), 1526–38.
- [MPW17] MICALLEF L., PALMAS G., OULASVIRTA A., WEINKAUF T.: Towards perceptual optimization of the visual design of scatterplots. *IEEE Trans Vis and Comp Graph* 23, 6 (2017), 1588–99.
- [ÖG12] ÖZTIRELI A. C., GROSS M.: Analysis and synthesis of point distributions based on pair correlation. *ACM Trans Graph* 31, 6 (2012), 1–10.
- [OJEF18] ONDOV B., JARDINE N., ELMQVIST N., FRANCONERI S.: Face to face: Evaluating visual comparison. *IEEE Trans Vis and Comp Graph* 25, 1 (2018), 861–71.
- [Pea94] PEARSON K.: Contributions to the mathematical theory of evolution. *Phil. Trans of the Royal Society of London. A* 185 (1894), 71–110.
- [RPD20] RAPP T., PETERS C., DACHSBACHER C.: Void-and-cluster sampling of large scattered data and trajectories. *IEEE Trans Vis Comp Graph* 26, 1 (2020), 780–9.
- [RRS13] REINERT B., RITSCHEL T., SEIDEL H.-P.: Interactive by-example design of artistic packing layouts. *ACM Trans Graph* 32, 6 (2013).
- [RRSG15] REINERT B., RITSCHEL T., SEIDEL H.-P., GEORGIEV I.: Projective blue-noise sampling. *Comp Graph Forum* (2015).
- [Sec02] SECORD A.: Weighted voronoi stippling. In *Proc. NPAR* (2002), p. 37–43.
- [SGBW10] SCHMALTZ C., GWOSDEK P., BRUHN A., WEICKERT J.: Electrostatic halftoning. 2313–27.
- [SLC*18] SICAT R., LI J., CHOI J., CORDEIL M., JEONG W.-K., BACH B., PFISTER H.: DXR: A toolkit for building immersive data visualizations. *IEEE Trans Vis and Comp Graph* 25, 1 (2018), 715–25.
- [SOA*19] SINGH G., OZTIRELI C., AHMED A. G., COEURJOLLY D., SUBR K., DEUSSEN O., OSTROMOUKHOV V., RAMAMOORTHI R., JAROSZ W.: Analysis of sample correlations for Monte Carlo rendering. *Comp Graph Forum (Proc. EGSR)* 38, 2 (2019).
- [Spe69] SPEAR M. E.: *Practical charting techniques*. McGraw-Hill, 1969.
- [SSC*20] SINGH G., SUBR K., COEURJOLLY D., OSTROMOUKHOV V., JAROSZ W.: Fourier analysis of correlated Monte Carlo importance sampling. *Computer Graphics Forum* 39, 1 (2020), 7–19.
- [Tal07] TALEB N. N.: *The black swan: The impact of the highly improbable*, vol. 2. Random House, 2007.
- [TT90] TUKEY J., TUKEY P.: *Strips displaying empirical distributions: I. textured dot strips*. Tech. rep., Bellcore Technical Memorandum, 1990.
- [Tuk77] TUKEY J. W.: *Exploratory data analysis*, vol. 2. Reading, MA, 1977.
- [Uli88] ULICHNEY R. A.: Dithering with blue noise. *Proc. IEEE* 76, 1 (1988).
- [vOHR20] VAN ONZENOORT C., HUCKAUF A., ROPINSKI T.: On the perceptual influence of shape overlap on data-comparison using scatterplots. *Computers & Graphics* (2020).
- [Wei10] WEI L.-Y.: Multi-class blue noise sampling. *ACM Trans Graph* 29, 4 (2010), 1–8.
- [Wil99] WILKINSON L.: Dot plots. *The American Statistician* 53, 3 (1999), 276–281.
- [WP83] WATSON A. B., PELLI D. G.: QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics* 33, 2 (1983), 113–20.
- [Yel83] YELLOTT J. I.: Spectral consequences of photoreceptor sampling in the rhesus retina. *Science* 221, 4608 (1983).
- [YXX*20] YUAN J., XIANG S., XIA J., YU L., LIU S.: Evaluation of sampling methods for scatterplots. *IEEE Trans Vis and Comp Graph* (2020).
- [ZHWB12] ZHOU Y., HUANG H., WEI L.-Y., WANG R.: Point sampling with general noise spectrum. *ACM Trans Graph* 31, 4 (2012), 1–11.