GeomLab 2020

July 21, 2020

Contents

1 Motivation

Motivated by the COVID-19 disease, several outlets printed two-dimensional depictions of data which usually contained multidimensional datasets. These datasets usually compromise {*infections, recoverd, death*} rates with regard to the *country* and sometimes also the evolution over *time*.

In this lab we focus on depicting data in a meaningful way. Besides its quantitative
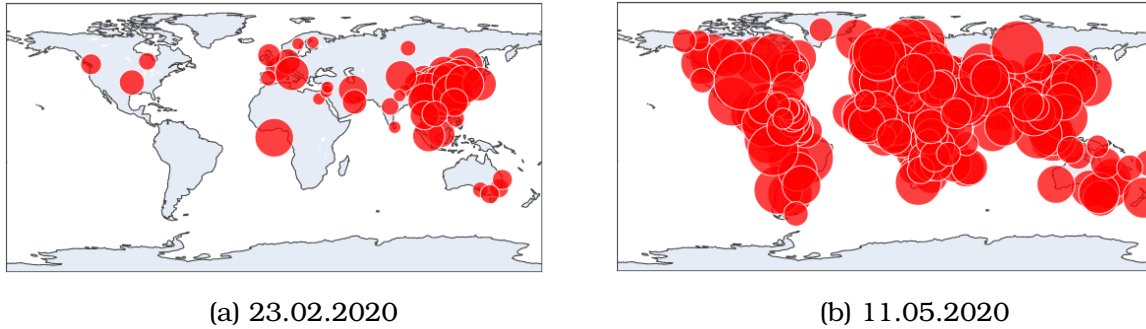

(a) 23.02.2020   (b) 11.05.2020

Figure 1: Comparison of a typical scatterplot depicting the *log* of confirmed cases as the radius.

nature this task has some qualitative properties. Similar to many solutions of problems in computational geometry, like deciding if a point is contained in a polygon or the construction of a convex hull for a set of points, the quality of the scatterplot is obvious to the human observer. Not so for a machine. While Fig. 1a is highly informative for a human or potential policy maker, Fig. 1b bears little information content. Despite these facts both figures result from the very same algorithm depicting data from the disease spread of COVID-19 during 2020 [Algo im Appendix].

To investigate the quantification of visual quality mainly Tufte in 1983 in *The visual display of quantitative information graphics press* and Miller et al in *The Need For Metrics In Visual Information Analysis* provided fundamental work. Based on this, follow-up work and further considerations from visibility problems and sorting we develop a novel approach to visualize multidimensional data in a scatterplot which transports as much information content as possible.

Our main resource will be the paper by Sergio Cabello at al *Algorithmic Aspects of Proportional Symbol Maps*. In this paper the hardness of maximizing utility measures with regards to physically realizable drawings was discussed and another type of drawing given by a stacking order was defined. For those drawings some algorithms which we will briefly discuss were derived. Those algorithms where derived for proportional Symbol Maps which only allow us to depict a single quantity. We will try to find more general glyphs and use slightly adapted algorithms to get optimal result with regards to those glyphs and moreover multidimensional data. Hereby, our main focus will be nested circles, centered nested circles and pie charts.

2 Problem Statement

Sergio Cabello at al discussed proportional Symbol Maps. For those we have n circles/disks in R^2 and the radius or size of the circles corresponds to some quantity which is associated to the center point of the circle. The problem is to find a drawing which retains as much of the information about the quantity and the location of the center as possible.

For the definition of a drawing they describe two approaches. One is given by *physically realizable drawings* and the other is *stacking drawings*. As *stacking drawings* are the natural perception and therefore natural restriction of a physically realizable drawing, we define the physically realizable drawing first. A drawing is physically realizable if the order in which the circles or disks – bounded by a perimeter – are stacked upon each other is uniquely determined at every face and no two of such orders conflict. In particular, any two or more disks with a common intersection have a unique ordering. Intutively, this can be imagined by very flexible disks, which somehow maintain their perimeter in the direction of beeing looked at. A stacking drawing is inutively the same thing with stiff disks.

A stacking drawings is realized by a stacking order which is given by a bijection $\phi : [n] \to [n]$. This bijection describes the order in which the circles are drawn. Circles which are drawn later cover parts over the circles which are drawn earlier. Therefore our task is to find a stacking order such that as little as possible of every circle is covered.

Now we need to discuss what we mean by covering as little as possible. We want that the viewer of the map should be able to identify the size of the circles and the center point. To achieve this goal, we could try to maximize the area of the circles which is not covered, but there are cases in which that would not allow the viewer to get all of the information. The approach we will use is to maximize the visible boundary of our glyph. We can define the visible and occluded boundary for circles formally with

$$U_i^{vis} = \partial C_i \setminus \bigcup_{\phi(j) > \phi(i)} C_j$$

$$U_i^{occ} = \partial C_i \cap \bigcup_{\phi(j) > \phi(i)} C_j.$$

Some results which are given in *Algorithmic Aspects of Proportional Symbol Maps* are the following

Theorem 1. *It is NP-hard to decide if a given collection of congruent disks has a physically realizable drawing where at least some given length of the perimeter of each disk is visible*

This implies that trying to maximize the minimal U_i^{vis} with respect to physically realizable drawings is NP-hard.

Theorem 2. *It is NP-hard to decide if a given collection of disks has a physically realizable drawing whose total visible perimeter is at least a given value.*

This implies that trying to maximize the sum over all the U_i^{vis} with respect to physically realizable drawings is NP-hard.

For stacking orders they give a positive result with regards to the minimum:

Theorem 3. *Given n disks in the plane, a stacking order maximizing the boundary length of the disk that is least visible can be computed in $O(n^2 \log n)$ time.*

If a polynomial algorithm for the sum of all the U_i^{vis} exists remains open to debate.

We won't try to construct an algorithm for the sum problem. Instead we will be looking at glyphs which can depict more information. The first glyph we will discuss are nested disks. Formally every glyph G^i is given by a set of circles $G_1^i, ..., G_k^i$ such that $r_{j+1} > r_j$ and $G_{j+1}^i \subset G_j^i$ for all $j \in \{1, ..., k-1\}$. Often times figures using nested disks also center all the discs in the same point. The second type of glyphs we will use are pie charts. A pie chart is formally given by a circle C_i and a set $\{\alpha_1^i, ...\alpha_k^i\}$ of dividing lines which we can encode as angle in $[0, 2\pi]$.

In the next section we will be discussing an easy general approach for maximizing special utility functions for glyph stacking order drawings and then apply it to our glyphs and some utility functions.

3 Stacking order algorithms for more general Glyphs

We will be focusing on utility functions which follow the following formula: Maximize the local utility of the worst glyph. In the example of the disk glyphs the local utility would just be the visible perimeter.

We can use a simple greedy algorithm to solve the problem. At each time step just calculate for all discs their visible perimeter as if they were the bottommost disk and then choose the disc C_i with the largest visible perimeter and then draw it. In the next timestep we look at the set of disks without C_i. This is done until we have processed all disks.

We can now formalize this approach for more general objects and prove that the greedy algorithm is optimal with respect to the utility.

Let D be a finite set and for $d \in D$ and $S \subseteq D$ the function $\Gamma(d, S) \mapsto \Re_{\geq 0}$ with the property, that for every $d \in D$ we have $S' \subseteq S => \Gamma(d, S') \geq \Gamma(d, S)$.

The set D corresponds to our set of glyphs and Γ corresponds to a utility function, which depends on the insertion order, i.e, the set S describes all the glyphs which lie

```
1  ALG(finite set D)
2      x := argmax_{d∈D} Γ(d, D \ d)
3      return [x, ALG(D \ x)]
```

Figure 2: The greedy algorithm

above d. Let $\phi : \{1,...,n\} \mapsto D$ be a stacking order. Then the for every glyph $d_i := \phi(i)$ the set of glyphs above d_i is given by $\{d_j \in D | j > i\}$.

Theorem 4. *For a given set D and a local utility function Γ the greedy Algorithm as described in Figure 2 returns a stacking order which maximizes*

$$\min_{d_i \in D} \Gamma(d_i, \{d_j \in D | j > i\}).$$

Proof. existiert muss noch getext werden ☐

We can use this theorem for object of any kind. In particular for the glyphs we discussed before. We now only need to define some reasonable utility functions for our different glyphs and calculate them locally with reasonable run-time.

3.1 Centered nested disks

For centered nested disks we can easily define local utility measures. Every glyph G corresponds to k nested circles therefore we could just calculate the visible perimeter for the k circles and then just combine them by choosing the minimum of the k circles or the sum of the visible perimeter of the k circles. For some set S of objects above G_i we then get formally:

$$\Gamma_{min}(G^i, S) = \min_j {}^{vis}G^i_j$$

$$\Gamma_{sum}(G^i, S) = \sum_{j=1}^{k} {}^{vis}G^i_j$$

We may also be interested in the relative utility which is given by normalizing the size of the circles. This relative approach stops the small circles from dominating big circles. In our applications there is close to no different between the relative utility and the normal one.

For Γ_{min} it may happen that for all glyphs the smallest circle may be completely covered. Then the utility would be zero for all of the circles. If that happens the algorithm has an optimal solution of value zero. If that happens, we perform the greedy step again and ignore the smallest circle in all of the glyphs until we get a value bigger than zero. This heuristic gives better results than just choosing a random circle.

Remark. *It is NP-hard to calculate physically realizable drawing where $\min \Gamma_{min}$ or $\min \Gamma_{sum}$ is maximized since disc glyphs are a special case of nested discs and both utility measures coincide and are equal to the visible circumference.*

3.2 Nested disks

For dense regions centered nested discs loose a lot of information since the small circles can easily be completely covered up. Therefore we will now derive an algorithm which guarantees, that every nested circle is at least visible. To achieve that we will allow the circles to be moved inside each other. We still want to have $r_{j+1} > r_j$ and $G_{j+1}^i \subset G_j^i$ for all j but the centers can be different.

Our approach is quite simple. We first just look at the biggest circle of each glyph. For those we perform our greedy algorithm with the local utility given by the size of the maximal **continuous** visible perimeter. After we have done that we can take the point p on the circle which is in the middle of the continuous visible perimeter piece and connect it to the center of the big circle by a line. Now every center point for a nested inner circle is uniquely defined by positioning it on that line such that the circle touches the point p. The final glyphs look a little bit like the *Hawaiian earring* topological space.
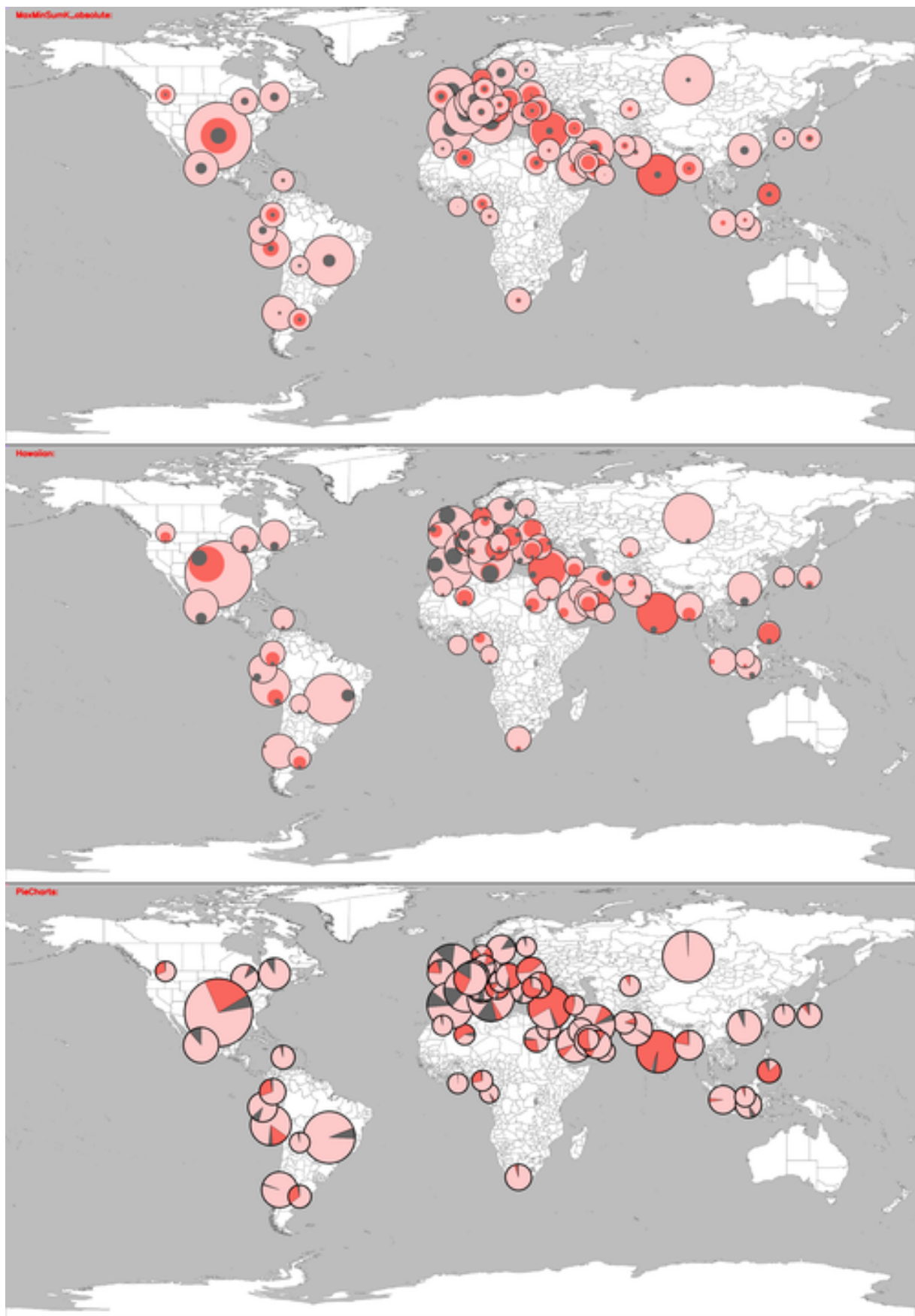
The resulting drawing guarantees that the perimeter of every circle is at least a little bit visible. The downside of this construction is that the drawing look a little bit strange to the viewer and its more difficult to focus the center of a glyph.
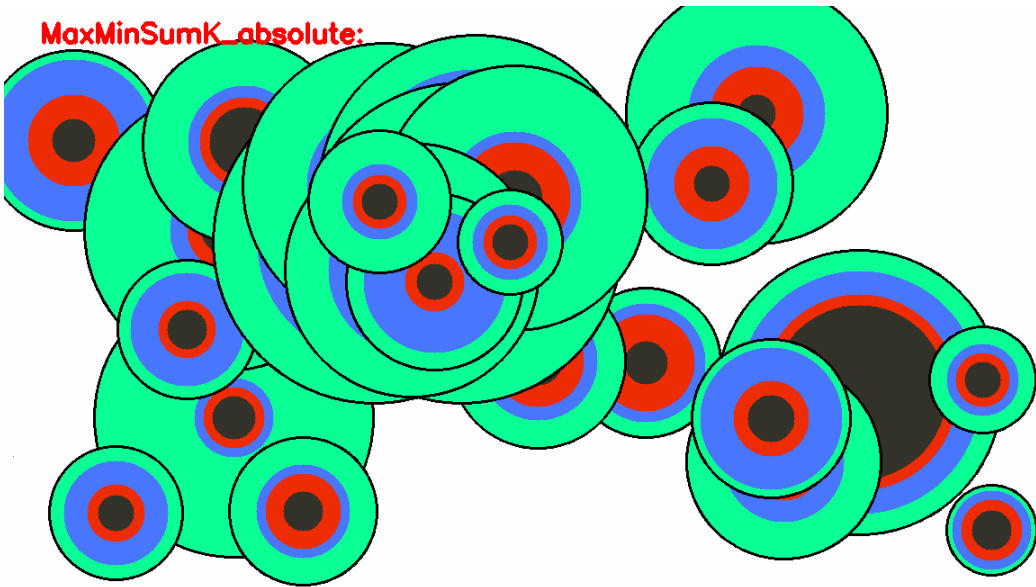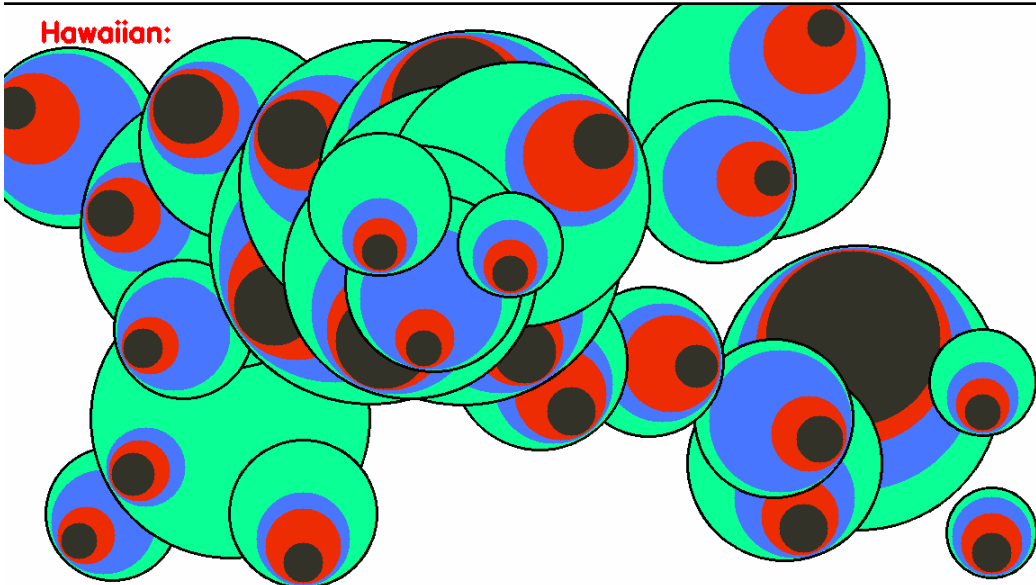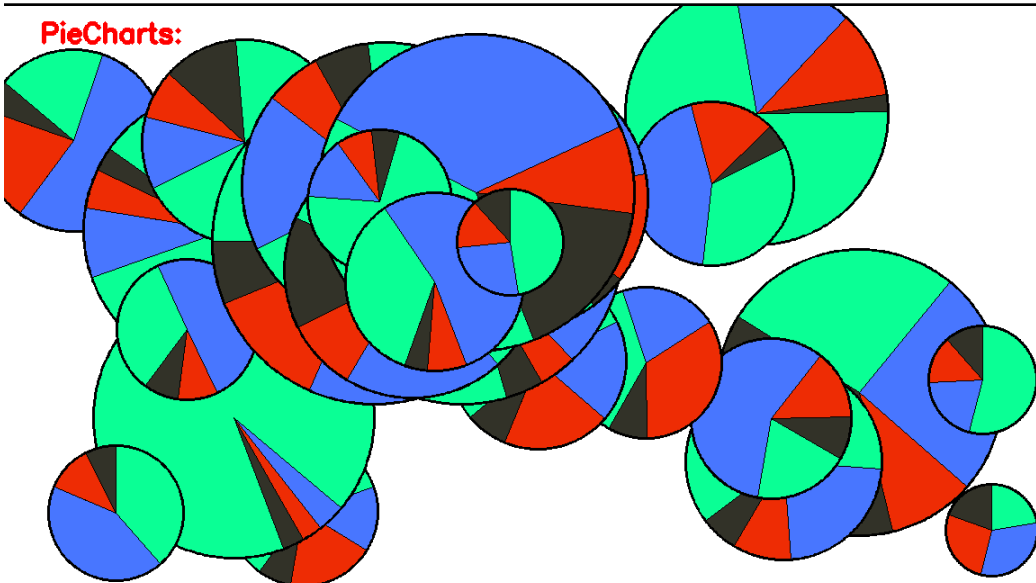
3.3 Pie charts

TBD

4 Experimental validation
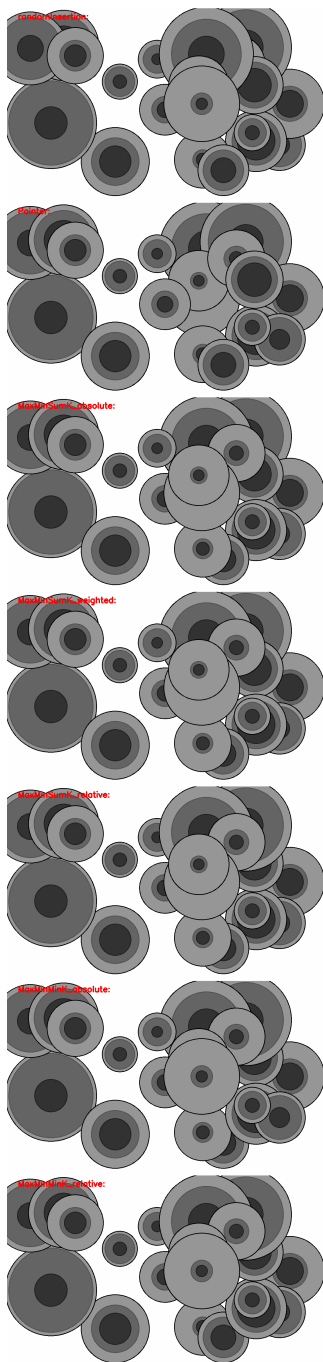
MaxMinSumK_absolute:

Hawaiian:

PieCharts:

5 Sources

Tufte, Edward R. "The visual display of quantitative information graphics press." Cheshire, Connecticut (1983).

Miller, Nancy, et al. "The need for metrics in visual information analysis." Proceedings of the 1997 workshop on New paradigms in information visualization and manipulation. 1997.

Brath, Richard. "Metrics for effective information visualization." Proceedings of VIZ'97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium. IEEE, 1997.

Tatu, Andrada, et al. "Visual quality metrics and human perception: an initial study on 2D projections of large multidimensional data." Proceedings of the International Conference on Advanced Visual Interfaces. 2010.

Urribarri, D. K., & Castro, S. M. (2016). Prediction of data visibility in two-dimensional scatterplots. Information Visualization, 16(2), 113–125. doi:10.1177/1473871616638892

Coeurjolly, D., Miguet, S., & Tougne, L. (2004). 2D and 3D visibility in discrete geometry: an application to discrete geodesic paths. Pattern Recognition Letters, 25(5), 561–570. doi:10.1016/j.patrec.2003.12.002

Yang-Pelaez J and Flowers WC. Information content measures of visual displays. In: Proceedings of the IEEE symposium on information vizualization 2000 (INFOVIS'00), 2000, pp. 99–103. Washington, DC: IEEE Computer Society.