GeomLab 2020

July 20, 2020

Contents

1 Motivation **2**

2 Problem Statement and State of the Art **3**

2.1 Problem statement and constraints 3

2.2 Previous work 3

3 Methodology **4**

3.1 Cost functions 4

3.2 Selection of algorithms and complexity 4

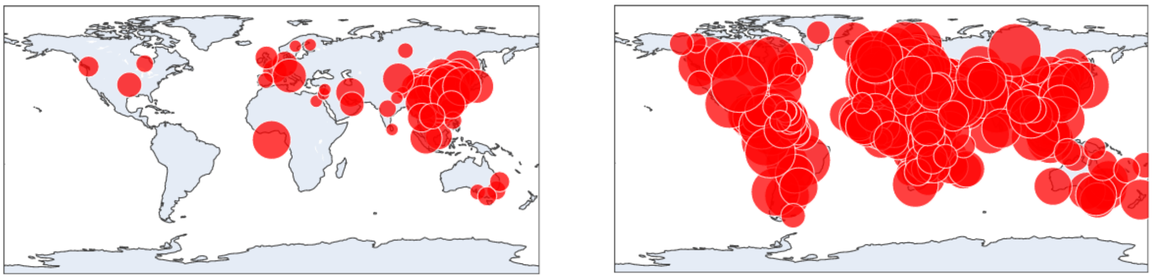
3.3 Optimalitaet fuer Pie Chart 4

4 Experimental validation **7**

5 Sources **10**

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*.



(a) 23.02.2020 (b) 11.05.2020

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

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

2 Problem Statement and State of the Art

2.1 Problem statement and constraints

1. Represention and visibility of scatter plots with glyphs
2. Represention and visibility of multi-dimensional data in nested discs
3. Represention and visibility of pie charts glyphs
4. Represention and visibility of (one?) other type of glyphs

2.2 Previous work

1. See papers in Slack Channel

3 Methodology

3.1 Cost functions

1. Information content covered by the data,
2. information content of the data in the visualization,
3. information capacity of a visualization,
4. topological information content.

3.2 Selection of algorithms and complexity

1. Random inseration
2. Painter
3. MaxMinSumK/absolute - Sum of visible circumference/hull
4. MaxMinSumK/weighted - weighted such that big circles/glyphs not dominate
5. MaxMinSumK/relative - Sum of relative circumference/hull
6. MaxMinMinK/absolute - MaxMin of minimal visible circumference/hull of each circle/glyph
7. MaxMinMinK/relative

3.3 Optimalitaet fuer Pie Chart

Let D be a finite set and for $d \in D$ and $S \subseteq D$ the function $\phi(d, S) \mapsto \Re_{\geq 0}$ with the property, that for every $d \in D$ $S' \subseteq S => \phi(d, S') \geq \phi(d, S)$.
Let the stacking-order $\pi : \{1, ..., n\} \mapsto D$ be bijective.

Theorem 1. *The algorithm*

```
ALG(finite set D) {
  x := argmax_{d \in D} \phi(d, D \setminus {d})
  return [x, ALG(D \setminus {x})]
}
```

returns a stacking-order, which maximizes $min_{d \in D} \phi(d, \{d \ \in D | d' > d\})$.

Proof. Given an optimal stacking-order s and output of ALG s^* .

1. case: $s = s^*$. Done.

2. case: $s \neq s^*$. $s \neq s^*$ implies, that there is a lowest $k \in N$, such that $s(k) \neq s^*(k)$. Let $s(k) = d_k$ and $s^* = d_k^*$. Hence, $k^* = s^{-1}(s^*(k)) = s^{-1}(d_k^*)$ is the index of d_k^* with respect to s. It follows $k < k^*$.

   We modify s such, that at point in time, we insert disc d_k^* and shift all others towards the end. Let this new stacking-order be s'.

   Comparing s and s', we find that s' selects all discs later. That means, that all ϕ-values are at least bigger.

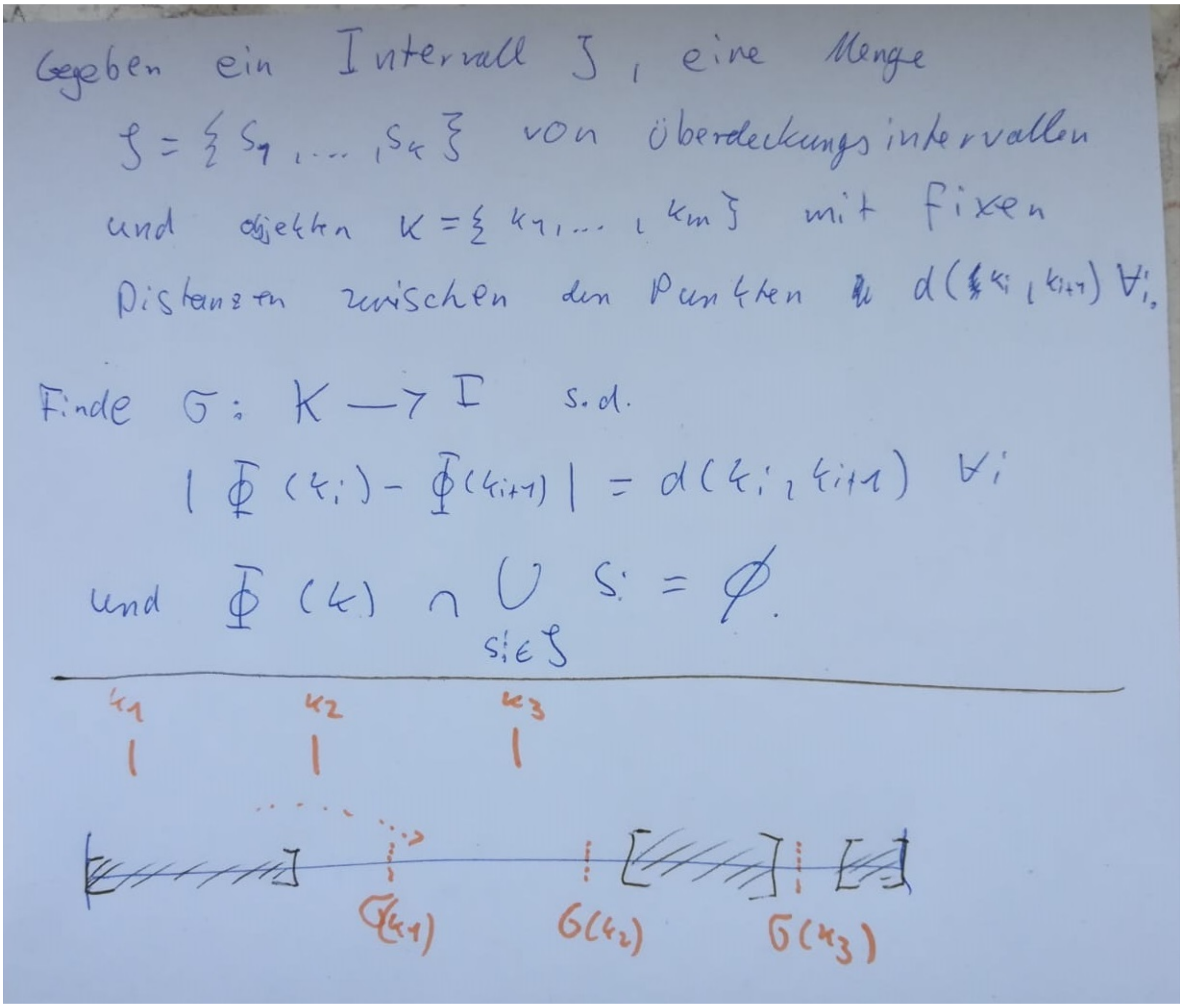
   Now d_k^* is chosen in s', as well as in s^* at position k. As d_k^* was chosen by s^*, d_k^* has to maxmize ϕ. Therefore, it is bigger or equal to d_k, as s and s^* are identical except at position k.

□

Remark. *ϕ may be the visible glyph boundary, the visible glyph boundary, the visible surface or the distance of...*

1. Rotation

2. Physics/Heuristics based approaches



Gegeben ein Intervall $\mathcal{J}$, eine Menge
$\mathcal{S} = \{s_1, \dots, s_k\}$ von überdeckungsintervallen
und objekten $K = \{k_1, \dots, k_m\}$ mit fixen
Distanzen zwischen den Punkten $d(k_i, k_{i+1})$ $\forall i$.

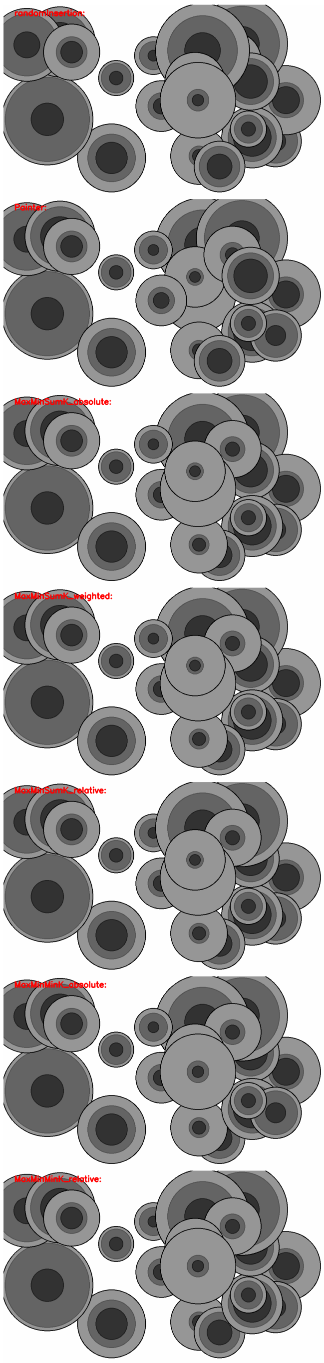
Finde $\sigma: K \to I$ s.d.

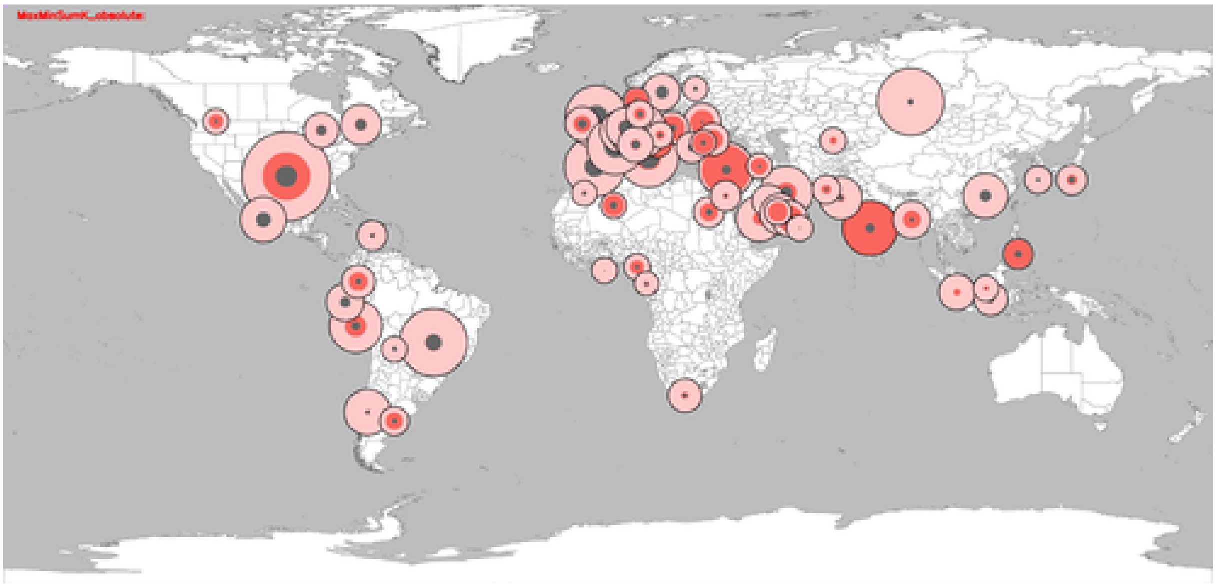
$$|\Phi(k_i) - \Phi(k_{i+1})| = d(k_i, k_{i+1}) \quad \forall i$$

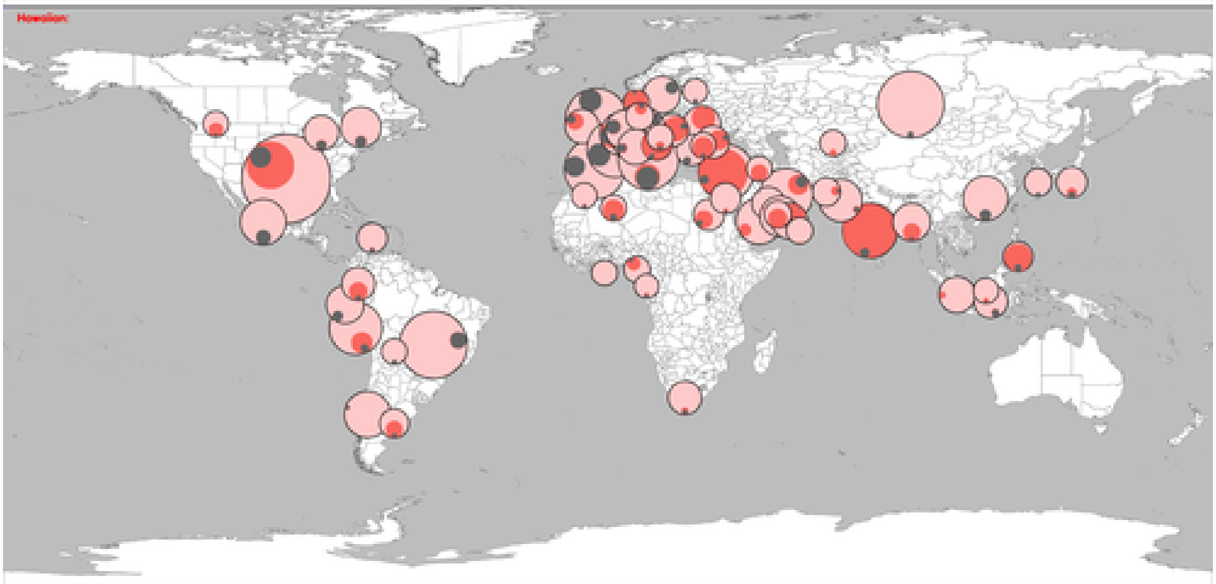
und $\Phi(K) \cap \bigcup_{s_i \in \mathcal{S}} s_i = \emptyset$.

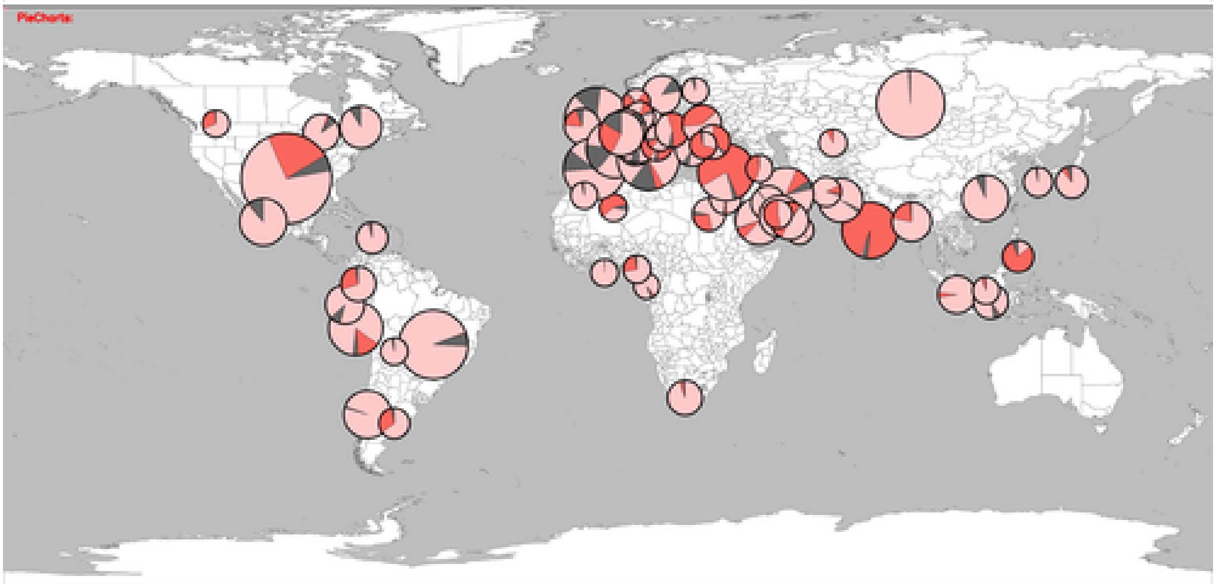
Figure 2: Trennlinienproblem

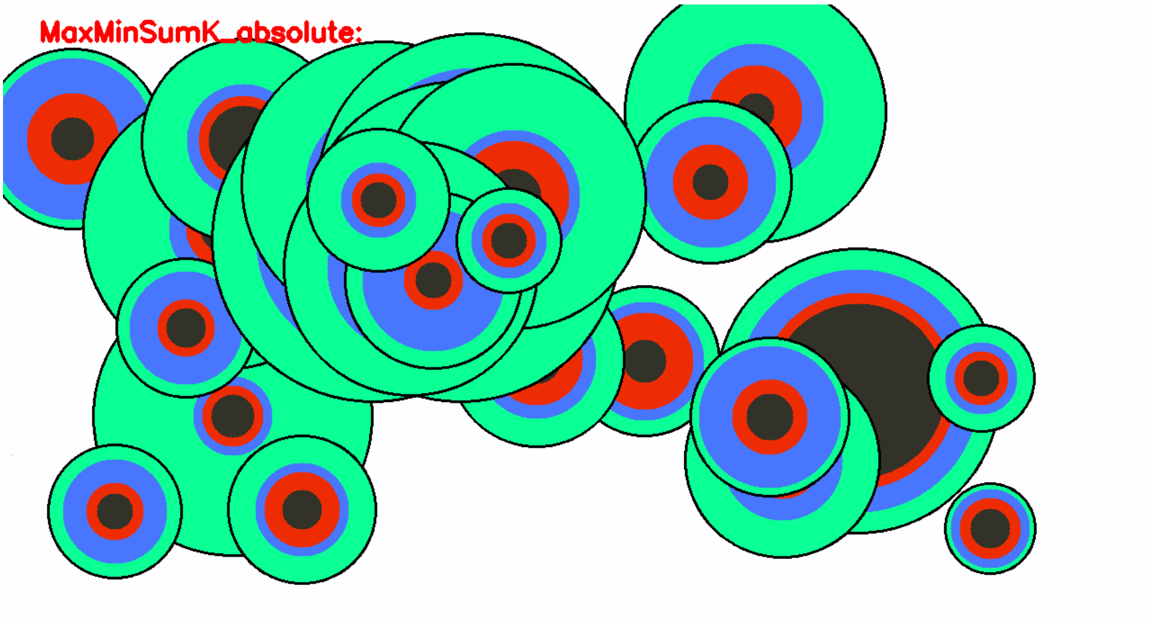
4 Experimental validation

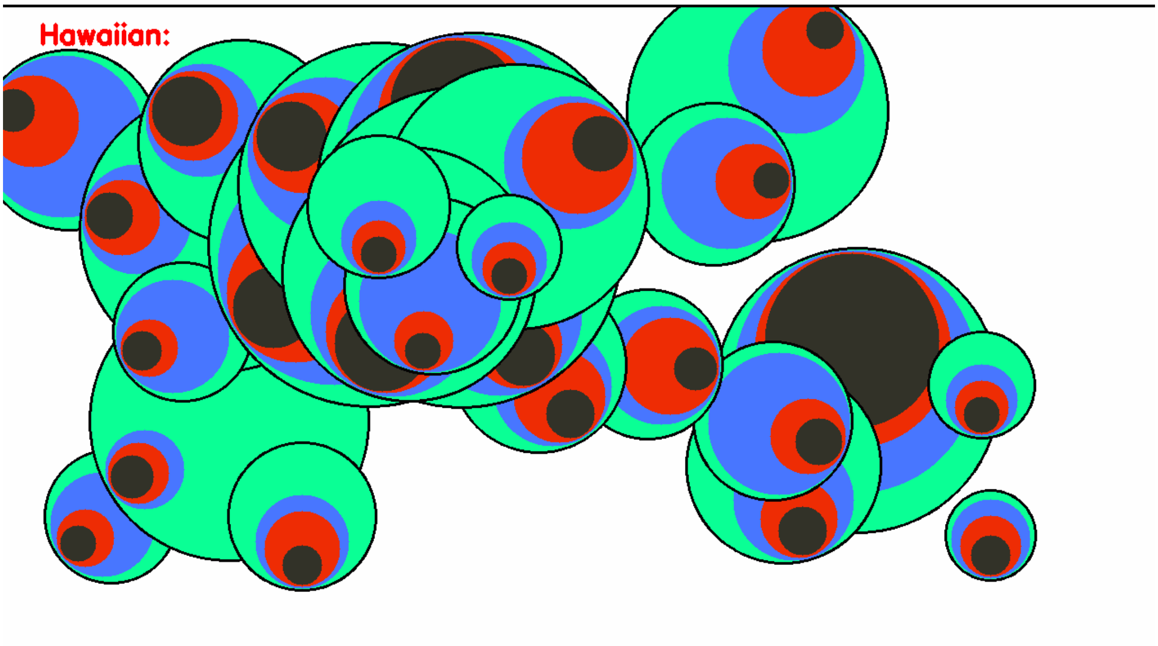


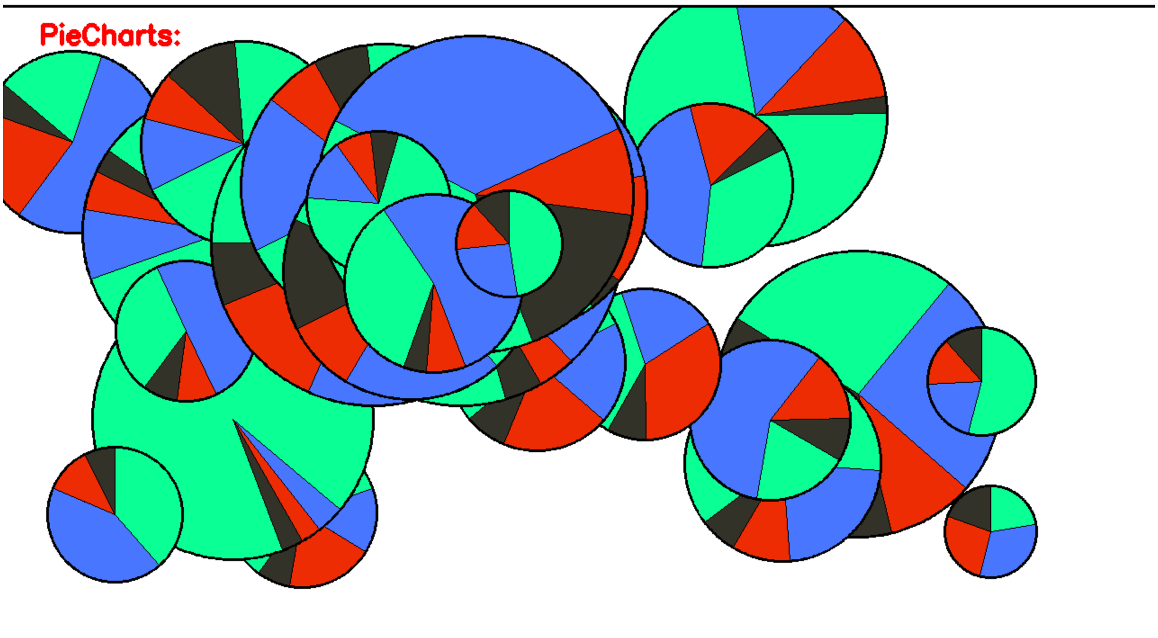

MaxMinSumK_absolute:




PieCharts:


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.