Beyond the Trade-off Curve: Multivariate and Advanced Risk-Utility Maps for Evaluating Anonymized and Synthetic Data

Oscar Thees* Roman Müller† Matthias Templ‡

November 11, 2025

Abstract

Anonymizing microdata requires balancing the reduction of disclosure risk with the preservation of data utility. Traditional evaluations often rely on single measures or two-dimensional risk-utility (R-U) maps, but real-world assessments involve multiple, often correlated, indicators of both risk and utility. Pairwise comparisons of these measures can be inefficient and incomplete. We therefore systematically compare six visualization approaches for simultaneous evaluation of multiple risk and utility measures: heatmaps, dot plots, composite scatterplots, parallel coordinate plots, radial profile charts, and PCA-based biplots. We introduce blockwise PCA for composite scatterplots and joint PCA for biplots that simultaneously reveal method performance and measure interrelationships. Through systematic identification of Pareto-optimal methods in all approaches, we demonstrate how multivariate visualization supports a more informed selection of anonymization methods.

Keywords: statistical disclosure control, risk-utility, Pareto optimality, multivariate statistics, visualization

1 Introduction

In microdata anonymization, modifications to the original dataset, such as suppression, generalization, perturbation, or synthetic data generation, are essential to reduce disclosure risk (Hundepool et al. 2012; Templ 2017). However, these modifications inevitably result in a loss of information, potentially limiting the analytical value of the data. Effective anonymization therefore requires balancing the confidentiality gained from lowering disclosure risk with the preservation of data utility. A possibility for evaluating this trade-off visually is the risk–utility (R-U) confidentiality map (Duncan, Keller-McNulty, and Stokes 2001). R–U maps depict anonymized datasets according to their measured disclosure risk and data utility, facilitating the systematic comparison of alternative anonymization approaches. By anonymization approaches, we refer to the different strategies by which datasets can be anonymized. These may involve method-specific variations, such as adjusting the level of noise in noise addition (Brand 2002) or comparisons between methods – e.g., the effect of microaggregation (Defays and Anwar 1998) and post randomization (Gouweleeuw et al. 1998) on disclosure risk. They can also include differences in synthetic data, either through employing distinct data generators and/or by producing multiple datasets with the same generator.

In practice, it is common to report one or more disclosure risk indicators and utility measures separately, often presented as summary statistics in tables or text, or visualized through simplified

*University of Applied Sciences and Arts Northwestern Switzerland. ORCID: 0009-0001-9378-4988. Corresponding author.

†University of Applied Sciences and Arts Northwestern Switzerland. ORCID: 0009-0007-2142-3896

‡University of Applied Sciences and Arts Northwestern Switzerland. ORCID: 0000-0002-8638-5276

two-dimensional R–U maps (Muralidhar and Sarathy 2006; Templ and Meindl 2008; Hornby and Hu 2021; Little, Elliot, and Allmendinger 2022; Little, Allmendinger, and Elliot 2025). Although such representations, especially R-U maps, aid in interpretation, displaying several risk and utility quickly becomes cumbersome. One remedy is to use faceting and arrange multiple R–U maps as *small multiples*: by enforcing comparisons of change, of differences between objects, small multiples are often the best solution for a wide range of presentation problems (Tufte 1990, p. 67-68). However, as the number of measures grows and thereby the number of panels needed for their display, this approach becomes more fragmented and inefficient (Hosseinpour et al. 2025). For example, with five distinct risk measures and five utility measures, 25 different R–U plots would be required to examine all pairwise relationships, complicating overall interpretation.

This challenge stems from the inherently multivariate nature of the evaluation problem: different risk and utility measures reflect distinct, often uncorrelated, aspects of data protection and analytical validity. Presenting them separately obscures potential interactions and trade-offs across dimensions. A comprehensive assessment therefore requires methods that jointly account for this multivariate structure, moving beyond simple pairwise comparisons to enable simultaneous consideration of multiple criteria. By framing risk and utility as a multi-objective optimization problem, we provide tools for more holistic evaluation of data releases – whether synthetically or traditionally anonymized (Dankar, Ibrahim, and Ismail 2022; Dankar and Ibrahim 2022).

In this article, we extend the classical risk-utility map to a multivariate setting through systematic comparison of six visualization approaches: heatmaps, dot plots, composite scatterplots, parallel coordinate plots, radial profile charts, and PCA-based biplots. We introduce three methodological innovations using principal component analysis: (1) blockwise PCA that extracts principal components separately for risk and utility measure blocks, (2) alignment analysis that validates dimensionality reduction by correlating composite measures with principal components, and (3) PCA-based biplots that simultaneously visualize method performances and measure relationships. To our knowledge, this represents the first systematic application of PCA to risk-utility visualization in statistical disclosure control. We further demonstrate systematic Pareto-optimal approach identification across all visualization approaches and evaluate each approach across twelve analytical capabilities. The design of these tools follows Tufte's data-ink principle (Tufte 1983) and key guidelines for visual data communication Franconeri et al. 2021.

The remainder of the paper is structured as follows. Section 2 introduces the conceptual and methodological foundation, framing evaluation as multi-objective optimization and presenting our visualization toolkit. Section 3 applies the framework to synthetic microdata from the EU-SILC dataset, illustrating how different approaches reveal performance trade-offs among synthetic data generators (SDGs). Section 4 discusses practical implications, method-specific considerations, and future directions. Section 5 provides key recommendations for practitioners and synthesizes our main contributions.

2 Methodology

This section develops a visual approach to the multi-objective optimization problem underlying the disclosure risk-utility trade-off. We introduce methods well suited to evaluating and visualizing anonymized and synthesized data across multiple risk and utility dimensions.

2.1 Pareto-Optimal/Efficient Trade-offs

The multivariate evaluation of disclosure risk and data utility naturally constitutes a multi-objective optimization problem, since improvements in one dimension often come at the expense of another. In such settings, the objectives are said to be at least partly conflicting (Miettinen 1998, p. 5), and a solution (or observation) is called non-dominated, or Pareto-optimal/Pareto-

efficient, if none of the objectives can be improved without degrading at least one of the others. Intuitively, this means that a Pareto-optimal solution represents an efficient trade-off: it cannot be improved in one aspect (e.g., utility) without performing worse in another (e.g., risk). Any method that is not Pareto-optimal is strictly inferior to at least one alternative (Mas-Colell, Whinston, and Green 1995, p. 547), a concept originating with Vilfredo Pareto's *Cours d'économie politique* (1896–1897).

Formally, let $\mathbf{u}_i \in \mathbb{R}^p$ denote the vector of p utility measures and $\mathbf{r}_i \in \mathbb{R}^q$ the vector of q risk measures for an anonymization approach i. We say that anonymization approach i dominates anonymization approach j if

$$u_{ik} \geq u_{jk} \quad \text{for all } k = 1, \ldots, p, \quad \text{and} \quad r_{i\ell} \leq r_{j\ell} \quad \text{for all } \ell = 1, \ldots, q,$$

with at least one strict inequality (Miettinen 1998, p. 11) (Emmerich and Deutz 2018, Def. 5, p. 588). An anonymization approach i is Pareto-optimal if there exists no other anonymization approach j that dominates it. This corresponds to strong Pareto optimality, since at least one inequality must be strict; the weaker notion, which also allows ties, is not considered here (Miettinen 1998, p. 19).

The collection of all non-dominated solutions forms the Pareto frontier, which makes the trade-off between risk and utility explicit by identifying efficient configurations and distinguishing them from dominated alternatives. Without additional preference information, there may exist an infinite set of Pareto-optimal solutions, all formally equally valid.

Additional decision criteria can also be applied. One option is to impose a risk-tolerance threshold and select the Pareto-optimal solution with the highest utility subject to this bound. Another intuitive "bang-for-buck" heuristic, applicable when the frontier is smooth and concave toward the origin, is to choose the knee (or elbow) point – where marginal increases in risk begin to yield only diminishing gains in utility, corresponding to the location of greatest curvature along the curve (Thorndike 1953, p. 275).

In practice, Pareto-optimal observations can be highlighted in visualizations using different colors or shapes, thereby conveying multidimensional information within a two-dimensional plot (Nagar, Ramu, and Deb 2023). However, the frontier itself can only be directly visualized in two dimensions.

2.2 Visualization and Evaluation Tools

As argued in Section 1, the risk-utility problem is fundamentally multivariate – joint feasibility matters more than marginal checks. We therefore introduce concise visualization strategies: different PCA biplots, R-U maps, origami charts and related multivariate views with Pareto-optimal solutions highlighted. We illustrate these on real data in Section 3.

2.2.1 Heatmaps

Heatmaps represent the values of a data matrix by encoding them as colour intensities (Wilkinson and Friendly 2009). In the context of statistical disclosure control, they provide a compact overview of how multiple anonymization strategies perform across a range of risk and utility measures.

Two main variants can be distinguished. In the first form, used in Section 3 (cf. Figure 1), the heatmap displays anonymization approaches as rows and evaluation measures as columns. Each tile then shows the performance of one method on one risk or utility metric, with darker or lighter colours indicating higher or lower values, depending on the scale. This layout facilitates visual comparison both across metrics (horizontally) and across anonymization strategies (vertically). When multiple datasets are evaluated, faceting and small multiples can be used to assign one panel per dataset, supporting cross-dataset comparisons. Hierarchical clustering

may be applied to reorder rows or columns based on similarity, thereby highlighting patterns or grouping structures.

In an alternative representation, each heatmap corresponds to a single anonymization method, with utility measures along one axis and risk measures along the other. Here, the tile colour reflects the performance on a particular utility–risk pair. This variant highlights joint behaviour across metrics and can be useful for examining interaction structures or identifying imbalanced profiles (e.g., methods that perform well in utility but poorly in specific risk dimensions). However, comparisons between methods are then made across different plots, rather than within a single unified display.

In both formats, enhancements such as numeric value labels, normalization per measure, or visual markers for Pareto-optimal methods (e.g., asterisks or outlines) can improve interpretability. Despite their strengths, heatmaps represent only univariate values per tile, so inter-metric correlations and higher-order patterns remain hidden. Colour encoding may distort perception, especially when a few large values compress the visual range of the smaller ones.

2.2.2 Dot Plots

Dot plots offer a position-based alternative to heatmaps by placing values as points on a common scale along horizontal or vertical axes. Each dot represents a single risk or utility measure for a given anonymization approach or dataset (cf. Figure 2). Faceting can again separate risk-like and utility-like metrics, and within each row, dot plots may display dispersion, overlay summary statistics such as medians, or include local trade-off annotations (e.g., ΔRisk and ΔUtility).

Dot plots provide greater accuracy for magnitude comparison, as position encodings are perceptually more effective than colour (Cleveland and McGill 1984). They make it easier to spot outliers, skewed profiles, and the relative influence of individual metrics. Pareto-optimal methods can be highlighted using colour or shape coding, and overlays such as risk caps or baseline comparisons can be added for richer interpretation.

2.2.3 Composite Risk/Utility scatterplots

Composite scatterplots summarizes the performance of different anonymization approaches by reducing multiple risk and utility measures into two composite scores (see, e.g., Little, Elliot, and Allmendinger 2022). The horizontal axis shows composite utility e.g. an average of all utility measures, while the vertical axis shows composite risk, similarly an average of all risk measures. Each point in the scatterplot corresponds to an single anonymization approach. The goal of the plot is to highlight trade-offs between disclosure risk and data utility, making it possible to identify anonymization strategies that strike a good balance. The quadrant structure is the same as in a classical R-U map and provides an intuitive interpretation (see Figure 3):

- Bottom-right: desirable region – high utility with low risk (best trade-off).
- Top-right: high utility but also high risk, posing disclosure concerns.
- Bottom-left: low utility but also low risk, often of little practical value.
- Top-left: worst case – low utility combined with high risk.

The strengths of this visualization are its simplicity and interpretability: it reduces a n-dimensional evaluation to two axes, enabling a quick diagnostic view and straightforward comparisons across anonymization strategies.

At the same time, several limitations need to be acknowledged. Collapsing all measures into averages inevitably leads to information loss, hides variability across metrics, and implicitly assumes equal weighting of measures. The approach also ignores metric correlations and it does

not display uncertainty such as standard errors or confidence intervals. Finally, results may depend on whether measures have been standardized prior to aggregation.

To enhance interpretability, the composite R–U map can be augmented with several features. First, highlighting the composite Pareto-optimal set separates genuinely efficient anonymization solutions from dominated ones, transforming the plot from a summary into a decision aid. Second, local trade-off annotations reveal the marginal "price" of utility: for each anonymization approach, showing the incremental move to the next better one (e.g., labels with Δ Utility and Δ Risk) indicates how much additional disclosure risk is incurred per unit of utility gained. Third, error bars showing the standard deviation of risk and utility along their respective axes indicate dispersion across the underlying measures.

While composite scores are commonly used to summarize performance across multiple risk or utility metrics, their interpretability hinges on the degree to which the constituent measures reflect a coherent underlying construct. To assess this, internal consistency metrics such as Cronbach's α or McDonald's ω can be used as heuristic diagnostics (Cronbach 1951; McDonald 1999; Zinbarg et al. 2005; Hayes and Coutts 2020). A value of $\alpha \geq 0.70$ is often cited as indicative of acceptable consistency among items within a block (Taber 2018; Nunnally and Bernstein 1994, p. 230). A high value then means that the composite score is a reliable summary of the underlying set of risk or utility measures, reflecting shared variation rather than averaging over unrelated or inconsistent metrics. Note that the mentioned threshold is a rule-of-thumb and context-dependent, and α assumes tau-equivalence – i.e., equal contributions of all indicators to the latent construct – which may not hold in practice. McDonald's ω is more general, as it allows heterogeneous loadings across items. In the context of multivariate risk–utility evaluation, internal consistency estimates serve only as rough checks for whether averaging across metrics (to form composite scores) is justifiable.

2.2.4 Parallel Coordinate Plot (PCP)

Parallel coordinate plots (Inselberg 1985) are a classical technique for visualizing multivariate data. Shown to be effective for many-objective optimization (Nagar, Ramu, and Deb 2023), this visualization likewise applies to the utility–risk formulation. In a PCP, each observation is represented as a polyline intersecting a series of parallel vertical axes, each corresponding to a (min-max normalized) variable (cf. Figure 5). This enables the joint visual inspection of multiple attributes, such as disclosure risk scores and utility indicators across different anonymization methods. PCPs can be useful for identifying outliers, indicating trade-offs between risk and utility, or revealing clusters of methods with similar risk or utility profiles.

2.2.5 Radial Profile Plots (Radar / Origami)

Radial profile plots place each metric on a separate radial axis and connect the values for each anonymization approach or dataset to form a polygonal profile similar to PCP but on a circular basis (cf. Figure 6). Although originally proposed for variables with a circular course (Mayr 1877, p. 78), the visualization is now widely used for general multi-criteria comparison.

One constructs the plot by placing the measure names on the radial axes, plotting the measures's min–max normalized values as points, and connecting them to form a polygon. This polygonal profile provides a comprehensive view of the performance of an anonymization approach. Area-based summaries of the polygon and its radial profile expose trade-offs: a bulge on one utility axis may coincide with a dip elsewhere, highlighting opposing tendencies. Radar charts provide a holistic view of each anonymization approach's multivariate measure profile, revealing both how measures vary within an approach and how approaches compare across all measures. However, they can become visually cluttered when many approaches or measures are displayed, inherently assume equal weighting across measures, and produce shapes that depend on the ordering of measures around the circle, which may bias interpretation.

The Origami plot (Duan et al. 2023) addresses some of these limitations by making the polygonal profile order-invariant through auxiliary axes with points equidistant from the center. In addition, the method allows explicit weighting of variables, so that selected measures can be emphasized more strongly in the visual profile. For a clearer view, one may limit comparisons to Pareto-optimal approaches or a ranking-based subset and group the risk and utility axes and distinguish their labels by color to facilitate interpretation (cf. Figure 6).

2.2.6 Multivariate PCA-based R-U Maps

Multivariate PCA-based R-U maps rely on principal component analysis (PCA) (Pearson 1901; Hotelling 1933), which reduces the dimensionality of multivariate data by projecting it into a low-dimensional space while retaining as much variance as possible. We present two approaches for applying PCA to risk-utility evaluation. The joint PCA approach combines all risk and utility measures in a single analysis and visualizes them in a biplot. The blockwise PCA approach applies PCA separately to risk and utility measure blocks, then plots the resulting principal components in a composite scatterplot.

1. Biplot

   Biplots, introduced by Gabriel (1971), provide a graphical framework to represent both observations and variables in the same plot. By combining the PCA projection with variable loadings, biplots make it possible to explore patterns, relationships, and group structures in complex datasets. In the context of data anonymization, biplots provide an effective visualization for exploring the trade-off between disclosure risk and data utility while also revealing correlations among multiple evaluation metrics within a single, interpretable representation. Although dimensional reduction methods have previously been used to evaluate the utility of anonymized datasets (e.g., Pau et al. 2025), to the best of our knowledge, biplots have not yet been applied to assess multiple risk and utility measures simultaneously.

   To construct the biplot, PCA is applied to the min-max standardized risk and utility measures, where the first two principal components define the axes on which observations and variable loadings are jointly visualized (cf. Figure 7). Each plotted point represents an anonymization approach; colors and shapes can further distinguish different characteristics of each anonymization approach (e.g., whether the anonymization approach is Pareto-optimal or not) and can be further augmented with confidence ellipses (e.g. in the context of the comparison of anonymization approaches across multiple datasets). In the biplot, approaches that lie close together have similar risk–utility profiles, while isolated points may indicate outliers (e.g., very low utility and/or very low risk). Loading arrows indicate how the measures correlate, highlighting which risk and utility measures align or contrast. Note, however, that the PCA axes are linear combinations of the original measures and may mix risk and utility; interpretation depends on how the measures load on the PCs. Results also depend on standardization and linearity assumptions; nonlinear relationships are not directly represented.

   Principal components are sign-indeterminate: flipping a component and its loadings leaves the general structure of the biplot unchanged. Although one could fix signs in the joint PCA for readability – e.g., enforce $\text{corr}(PC1, U) \geq 0$ and $\text{corr}(PC2, -R) \geq 0$ and apply the same flips to scores and loadings – we keep the native orientations returned by the algorithms throughout. To aid interpretation, figures annotate the directions of higher utility (U) and lower risk (R) and display loading arrows.

   To assess whether an anonymization approach satisfies predefined disclosure-risk and utility requirements, one can specify per-metric acceptance thresholds. In the original metric space, these thresholds define an axis-aligned hyperrectangle, with each axis ranging from

its lower bound (typically zero) to its cutoff. When projected into PCA space via the fitted loadings, this acceptance region appears as a polygon delineating the feasible area.

When outliers are a concern, a robust PCA variant (e.g., ROBPCA; Hubert, Rousseeuw, and Vanden Branden 2005) is preferable. We further propose to use the score–distance / orthogonal-distance (SD-OD) diagnostic plot (Hubert, Rousseeuw, and Vanden Branden 2005) to screen anonymization approaches for outliers. Following (p. 66), the robust score distance and orthogonal distance are

$$\mathrm{SD}_i = \sqrt{\sum_{j=1}^{k} \frac{t_{ij}^2}{\ell_j}}, \qquad \mathrm{OD}_i = \left\| x_i - \mu - P_{p,k}\, t_i \right\|_2,$$

where t_{ij} are (robust) scores on PC j, ℓ_j the corresponding eigenvalues, μ the robust center, and $P_{p,k}$ the loading matrix which all need to be estimated. Intuitively, SD measures leverage within the k-dimensional PC subspace, i.e. the space spanned by the first k components retained in the model. It is proportional to the Mahalanobis distance of the score vector in this reduced space. Large SD means the observation has unusually large scores on one or more principal components – i.e., it lies far from the center within the PC subspace. OD measures residual distance orthogonal to that subspace: it is the Euclidean distance between the observation and its projection onto the k-dimensional PCA space. So a large OD means the point is poorly represented by the first k PCs. For a quick look,

Table 1: SD–OD regions and their interpretation.

Region (SD, OD)	**Meaning**	**Typical follow-up**
SD low, OD low	Regular / typical; near the center of the data cloud and well represented by the first k PCs.	No concern; representative SDG.
SD high, OD low	Good Leverage outlier; far within the PC subspace (large scores on one or more PCs) yet small reconstruction error.	Check influence; may be a valid extreme SDG.
SD low, OD high	Orthogonal outlier; not far in the PC plane but poorly reconstructed by the first k PCs – structure outside the captured subspace.	Inspect variables not explained by PCs; consider increasing k or data issues.
SD high, OD high	Bad leverage; both far in the PC plane and poorly represented – extreme and structurally unusual.	Strong candidate for anomaly; scrutinize or exclude.

the same diagnostic plot can be made with classical PCA; for outlier detection and stable cutoffs we prefer robust PCA, comparing SD to $\sqrt{\chi^2_{k,.975}}$ and the OD to $\Phi^{-1}(.975)$ (p. 66). Points beyond either cutoff are flagged as outliers cf. Table 1.

PC1 alignment with risk and utility composites: To support interpretation, we investigate the loadings in detail and relate the PC1 scores to the composite measures of risk and utility (cf. Table 4).

While the loadings show how risk and utility measures contribute to the principal components, our aim is to assess whether these constructs are distinguishable, i.e., whether they occupy separate directions in feature space. The idea is thus to align PC's to composite indicators on risk and utility. In what follows we use classical joint PCA with $k = 2$ retained components, corresponding to the two conceptual dimensions of risk and utility. Let $x_i \in \mathbb{R}^p$ denote observation i (the row vector of min–max standardized risk–utility measures), and let μ be the column mean vector. Define the loading matrix of the k retained

principal components

$$P_{p,k} = [\,\mathbf{p}_1, \ldots, \mathbf{p}_k\,] \in \mathbb{R}^{p\times k},$$

The score vector for observation i is

$$t_i \;=\; P_{p,k}^{\top}(x_i - \mu) \in \mathbb{R}^k,$$

so the PC1 score for observation i is $t_{i1} = \mathbf{p}_1^{\top}(x_i - \mu)$. Stacking $t_i^{\top}$ as rows yields the score matrix $S \in \mathbb{R}^{n\times k}$.

Composites for utility and risk are constructed directly from the min–max standardized measures by taking the mean as we did for the composite R-U map in Section 2.2.3. Because correlation is invariant to centering and scaling, we use the original composite scores U and R for correlation analysis.

We then individually align the composites with t_1 (the vector of first-PC scores across observations)

$$\rho_{t_1,U} = \mathrm{corr}(t_1, U), \qquad \rho_{t_1,R} = \mathrm{corr}(t_1, R).$$

These values capture how strongly the dominant principal component t_1 aligns with each composite considered separately. Because the utility and risk composites may themselves be correlated, separate correlations can overstate the overall alignment. We therefore also consider a joint model in which t_1 is regressed on both composites simultaneously

$$s = \beta_0 + \beta_U U + \beta_R R + \varepsilon.$$

The coefficient of determination R^2 from this regression expresses the total fraction of variance in the component scores that is explained jointly by the two composites. Thus, while $\rho_{t_1,U}$ and $\rho_{t_1,R}$ describe individual associations, R^2 provides a single summary of joint alignment.

For interpretation we report three sets of results:

(a) the proportion of variance in the measures explained by the first principal component (from the PCA output);

(b) the individual correlations $\rho_{t_1,U}$ and $\rho_{t_1,R}$, together with their squared values $\rho^2_{t_1,U}$ and $\rho^2_{t_1,R}$, which indicate the proportion of variance in each composite accounted for by the component scores;

(c) the coefficient of determination R^2 from the joint regression, which provides a single summary measure of how well the component scores are aligned with the two composites taken together. A large R^2 indicates that PC1 is well captured by the utility and risk composites, whereas a small R^2 signals that other patterns (e.g. method-specific effects or dataset baselines) are driving the component. This helps clarify what PC1 represents and whether it can be treated as a single summary axis.

Caveats are that this approach relies on linear correlation and assumes that the risk and utility composites meaningfully capture the underlying constructs. When a PC shows a clear association with risk or utility – regardless of sign – it can be interpreted as capturing a general risk–utility contrast. If such alignment is weak or absent, additional components can be examined to identify secondary or dataset-specific patterns. Otherwise, the PCA is interpreted descriptively, with loadings indicating which measures contribute most to the observed structure.

Compare anonymization approaches across multiple datasets: When multiple datasets are anonymized (e.g., to compare the performance of different SDGs across datasets), each point in the biplot represents a dataset–anonymization method pair. The biplot then

separates datasets and positions anonymization approaches. With a small amount of anonymization approaches, we can show per-dataset centroids with ellipses/convex hulls to display uncertainty. This complements joint PCA, alignment, and blockwise PCA (see following section) for multi-dataset comparisons.

2. Blockwise PCA

   In the blockwise PCA approach, we conduct two separate principal component analyses: one on the set of disclosure risk measures and one on the set of utility measures (for utility measures only, see Dankar and Ibrahim 2022). We then plot the first principal component of the utility block on the x-axis and the PC1 of the risk block on the y-axis. Each point in the resulting two-dimensional plot represents an anonymization approach, while the two axes serve as composite indices summarizing the respective blocks. Unlike the simple means used for the composites in the composite R-U map in Section 2.2.3, this PCA approach takes into account the correlation structure within each block and assigns weights to measures based on their contribution to overall variance. This allows for a more informed aggregation, where strongly varying and interrelated measures are emphasized, rather than treating all measures as equally important. However, this summarization comes at a cost: the axes are unitless and data-driven, and their interpretation depends on the specific loadings of the first principal components. To aid interpretation, the squared normalized loadings of PC1 can be visualized as stacked bar charts aligned to each axis (Utility PC1 and Risk PC1), showing which measures contribute most to the composite scores (cf. Figure 9). Optionally, Pareto-optimal anonymization approaches can be highlighted, and, when comparing multiple datasets for each anonymization approach, group structures can be visualized using ellipses to indicate clustering or method families.

3 Illustrative Applications

This section applies the proposed visualization strategies to real-world data.

3.1 Data and measures

The European Union Statistics on Income and Living Conditions (EU-SILC) is a flagship data source in official statistics, widely used to monitor poverty, social inclusion, and related policy targets in Europe. We use the Austrian EU-SILC public-use file from 2013. A concise description of key variables and further details is available in the manual of the R package simPop (Templ et al. 2017); a comprehensive variable catalogue is provided in (Gesis 2024) and online at https://www.gesis.org/en/missy/materials/EU-SILC/documents/codebooks (accessed 2025-06-25).

We synthesize the data with a set of commonly used, methodologically diverse synthetic data generators (SDGs). A description of the SDGs used and their software packages is provided in Table 6 in the Appendix A. For each synthetic dataset we compute a standard set of disclosure risk and utility measures; these measures and their definitions are given in Table 2. We analyze the resulting matrix of risk and utility measures. We apply min-max scaling to $[0, 1]$ and harmonize the directions of the measures for interpretability, such that larger values indicate higher utility for utility metrics and higher disclosure risk for risk metrics. All risk and utility measure calculations, data manipulation, and visualization were performed in `R` (R Core Team 2025) using the `tidyverse` ecosystem (Wickham et al. 2019), specifically `ggplot2` (Wickham 2016) for visualization and `dplyr` (Wickham et al. 2022) for data manipulation. Additional packages are cited where used.

Table 2: Definition of the risk and utility measures

Type	Measure & Abbreviation	Author	Description
Risk	Replicated Uniques (`RepU`)	(Raab, Nowok, and Dibben 2025)	Percentage of replicated sample uniques in synthetic dataset.
Risk	Disclosure by Successful Correct Outcome (`DiSCO`)	(Raab, Nowok, and Dibben 2025)	Percentage of disclosive in synthetic and correct in the original.
Risk	Disclosure by Correct Attribution Probability (Direct) (`DCAP`)	(Taub et al. 2018)	Probability that a synthetic record can be correctly attributed to an original record (direct case).
Risk	Targeted Correct Attribution Probability (`TCAP`)	(Taub et al. 2019)	Probability of correct attribution when targeting specific records.
Risk	Risk of Attribute Prediction-Induced Disclosure (`RAPID`)	(Currently under development by the authors; a preprint will be available within the next month.)	Expected share of re-identifiable records.
Utility	Confidence Interval Proximity (`Proximity`)		Deviation of confidence intervals for sensitive variables.
Utility	Propensity Mean Squared Error (`pMSE`)	(Snoke et al. 2017)	Predictive score from distinguishing real vs. synthetic data.
Utility	Wasserstein Distance (`Wasserstein`)	(Vasershtein 1969)	Wasserstein distance between numeric distributions.
Utility	Hellinger Distance (`Hellinger`)	(Hellinger 1909)	Hellinger distance between categorical distributions.
Utility	Energy Distance (`Energy`)	(Rizzo and Székely 2016)	Multivariate energy distance on numeric variables.

3.2 Visualization Approaches

Clustered Heatmap As an initial overview, Figure 1 presents a clustered heatmap of min-max normalized measures to reveal cross-method patterns at a glance. To indicate Pareto-optimal SDGs we applied a standard two-objective dominance rule (minimizing Risk, maximizing Utility) explained in Section 2.1, i.e., for SDGs i and j, we say that j dominates i if $\text{Risk}_j \leq \text{Risk}_i$ and $\text{Utility}_j \geq \text{Utility}_i$, with at least one strict inequality. SDGs that are not dominated by any other form the Pareto set and are indicated with an asterisk in Figure 1. For the row-wise ordering of the SDGs, hierarchical clustering was used. The heatmap shows that `MostlyAI`, `synthpop`, and `simPop` are the only SDGs that are composite Pareto-optimal (cf. Paragraph 3.2), whereas the dataset generated with `synthpop` has the highest utility but also the highest disclosure risk.

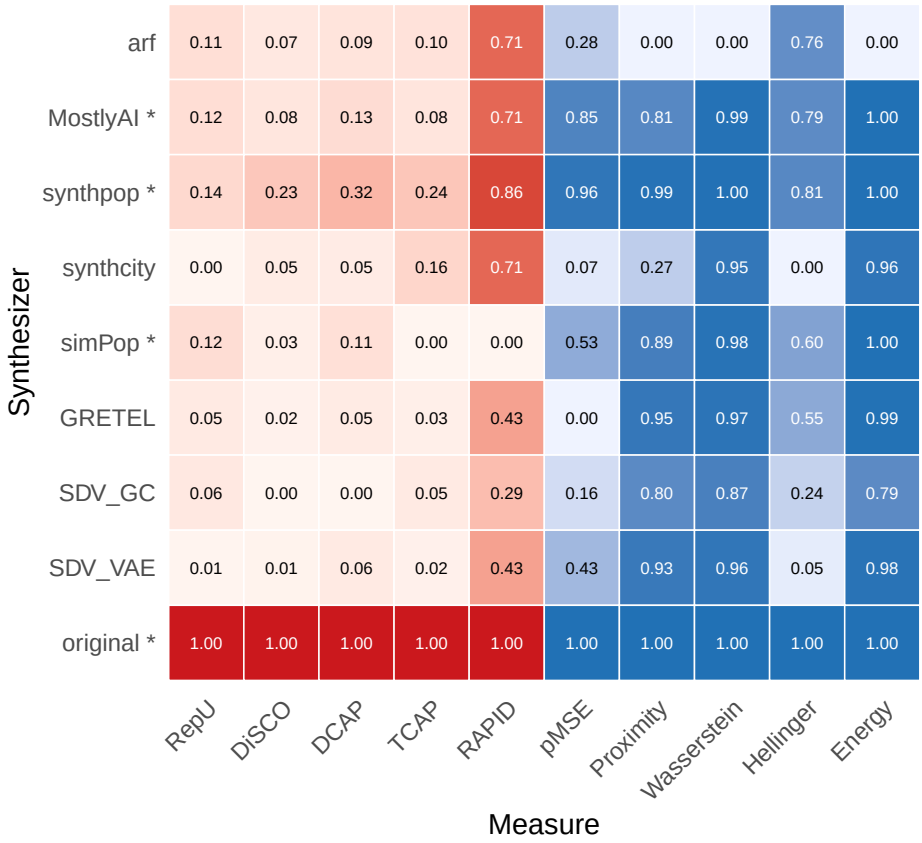

Figure 1: Performance of synthetic data generators across utility (bluish) and risk measures (reddish), with values min-max normalized to [0,1] (rows = SDGs, columns = measures). Asterisks indicate composite Pareto-optimal SDGs.

Dot Plots Because a heatmap does not convey within-SDG dispersion, we complement it with the dot plot in Figure 2. This figure shows the distribution of individual measures per SDG relative to their median. For example, `SDV_VAE` performs competitively on several metrics but shows substantially lower performance in *Hellinger-distance*, which lowers its overall standing.

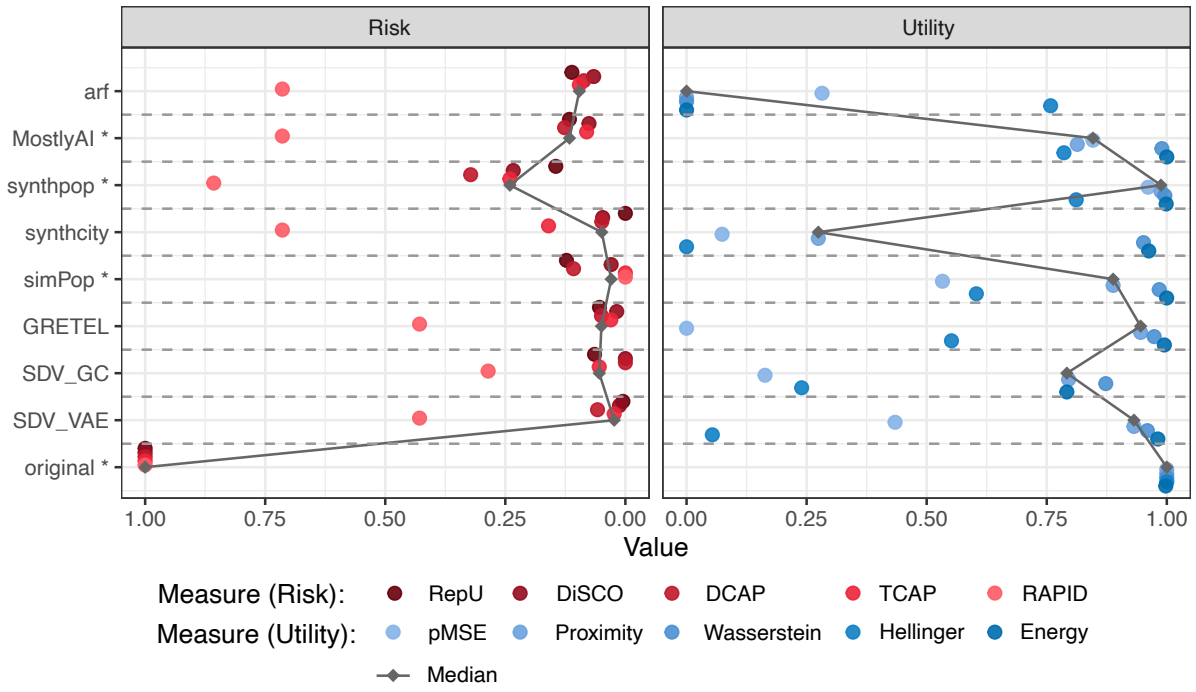

Figure 2: Dot plot of risk-like (left, red scale) and utility-like (right, blue scale) measures across SDGs. Points show individual measure values; the gray line with diamond marker indicates the per-SDG median.

Composite risk–utility (R-U) map: The second visualization is a composite R-U map. For each SDG we compute composite scores by averaging the normalized constituent measures separately for risk (lower is better) and utility (higher is better). The resulting scatterplot in Figure 3 places each SDG in the R-U plane. We report Cronbach's α and McDonald's ω (cf. Section 2.2.3) for the sets of metrics entering each composite; both indicate that the composites capture a coherent underlying construct. Pareto-optimal SDGs are shown in blue and connected to form the empirical Pareto front. We highlight `synthpop` as the knee point (cf. Section 2.1), defined as the point on the front with the largest perpendicular distance from the straight line joining the two extreme Pareto points, that is, the location of maximum curvature. Further we display the slope values $\Delta R/\Delta U$, which represent the trade-off from one SDG to another on the Pareto front.

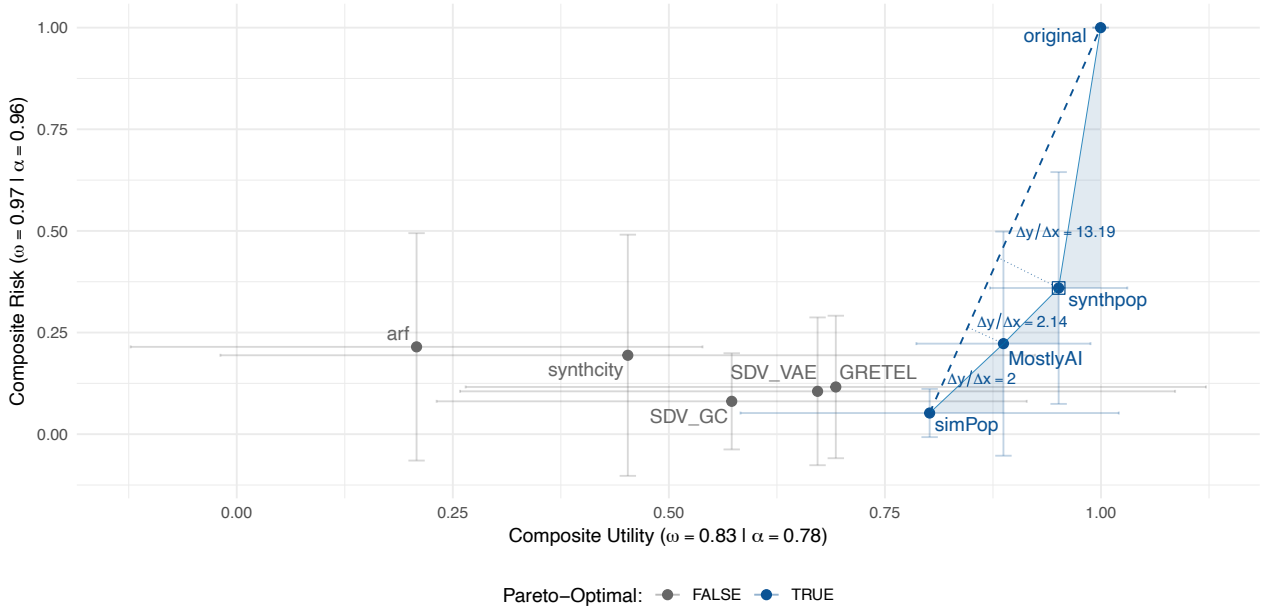

Figure 3: Composite risk vs. utility for SDGs. Pareto-optimal methods are in blue; reliability of composites is indicated by α and ω. Error bars denote the standard deviation of risk and utility along the axes.

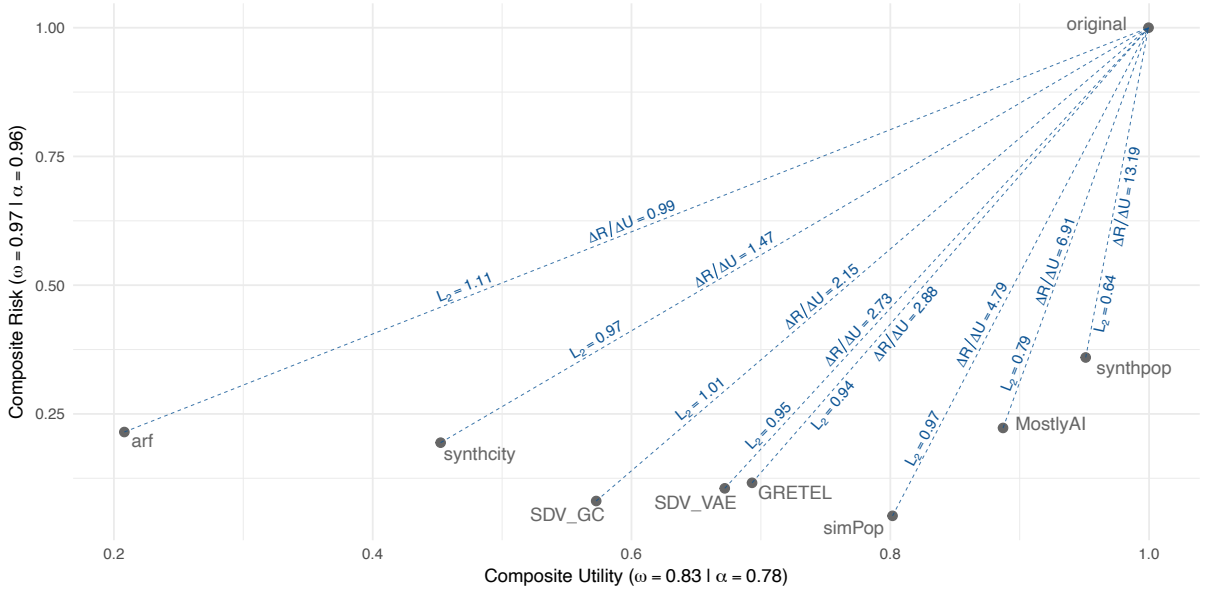

Figure 4: Rays from each SDG to the original (U_0, R_0) with labels $\Delta R/\Delta U$ and Euclidean distance L_2.

In Figure 4 we draw, for each SDG point (U_i, R_i), the ray to the original (U_0, R_0) and report the Euclidean L_2 distance and the slope

$$\text{slope}_i^{(\text{orig})} = \frac{\Delta R}{\Delta U} = \frac{R_i - R_0}{U_i - U_0},$$

interpreted (for $\Delta U > 0$) as the incremental risk paid per unit of utility gained relative to the original. A lower $\text{slope}_i^{(\text{orig})}$ means a cheaper marginal trade-off from the original, but it is not an overall ranking: global desirability depends on the joint (U_i, R_i) and the full metric set. A

similar picture is revealed when we look at simple Euclidean distance (L_2) in the R-U plane. Without explicit scaling/weighting, these can be misleading; for example, `synthcity` and `simPop` can be equally distant from the original even though the latter Pareto-dominates the former. We therefore use Pareto dominance to identify "best", and use slopes only to summarize marginal cost.

Parallel Coordinates Plot (PCP): A PCP, introduced in Section 2.2.4, is depicted in Figure 5. To account for the inverse scaling of the two dimensions, we separated risk and utility into distinct facets. The composite Pareto-optimal SDGs are highlighted by coloring, while the remaining ones are shown in neutral tones for context. This representation makes the multivariate nature of the evaluation particularly clear: instead of reducing performance to aggregated scores, the PCP allows us to directly and easily compare SDGs across all individual metrics. Patterns of trade-offs and strengths become more apparent, helping to identify which methods perform consistently well across dimensions and which excel only in specific areas.

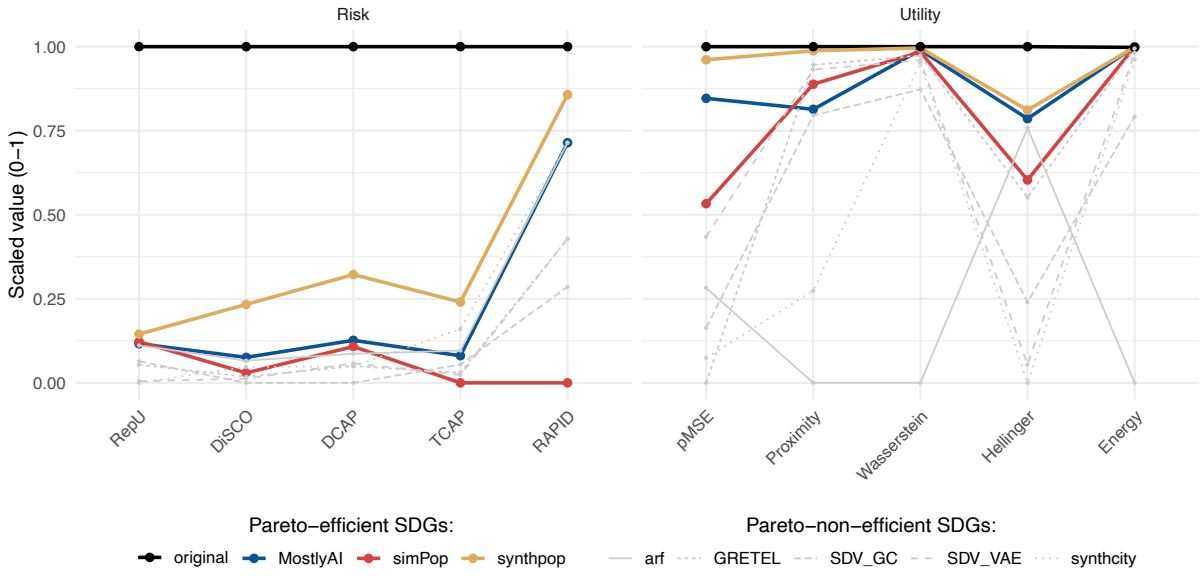

Figure 5: Parallel coordinates of scaled risk (left) and utility (right) measures for SDGs. Pareto-optimal SDGs (colored) are contrasted with non-optimal ones (gray), with the original dataset shown as a benchmark.

Radial Profile Plots: Figure 6 shows an origami plot (cf. Section 2.2.5). Although origami plots provide a compact representation of multivariate performance, they can become cluttered when more than two anonymization approaches are displayed simultaneously. In such cases, the interpretability suffers, leading to the same "overview problem" encountered in the classical R-U map when we need many small multiples to show all SDGs at once. To mitigate this, we restrict the visualization to the Pareto-optimal SDGs and present them only in a bivariate way in three separate plots. This focused approach preserves readability while still allowing direct comparison of the leading methods.

Furthermore, polygonal shapes permit area-based summaries, offering an additional quantitative lens on overall performance (cf. Table 3). However, such areas can be misleading: unless risk axes are inverted, a larger area does not imply a better performing SDG. For these reasons, we report areas only as a secondary descriptor and do not use them for ranking.

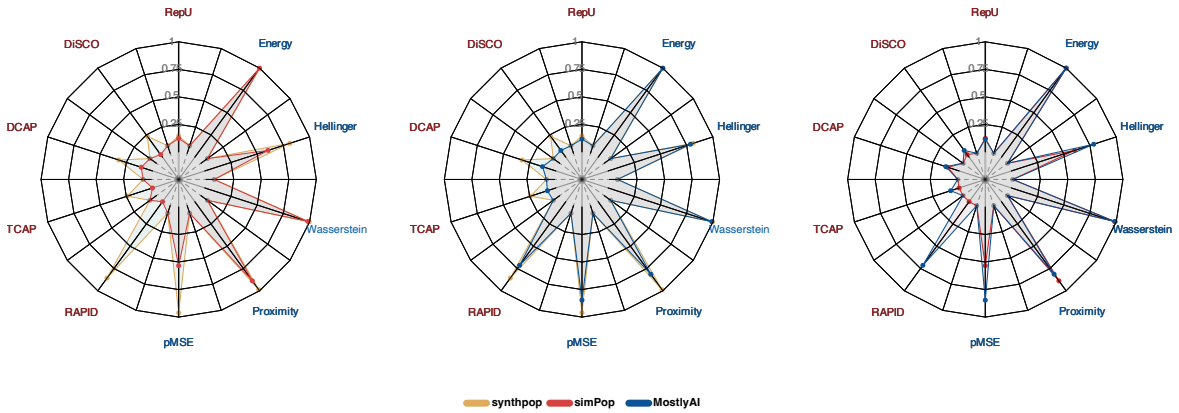

Figure 6: Bivariate Origami plots of disclosure-risk (↓ better) and utility (↑ better) measures for all SDGs. Each axis corresponds to one normalized metric, scaled to $[0, 1]$. The polygon shape of each SDG reflects its performance profile across measures, with smaller areas on the risk dimensions indicating lower disclosure risk and larger areas on the utility dimensions indicating higher utility.

Table 3: Polygon areas per SDG computed from the origami plot (figure not shown).

SDG	**Area**
original	1.000
synthpop	0.660
MostlyAI	0.550
simPop	0.430
GRETEL	0.400
SDV_VAE	0.390
SDV_GC	0.330
synthcity	0.320
arf	0.210

Biplots and other PCA-Related Plots: A biplot based on a joint PCA of risk and utility measures, as outlined in Section 2.2.6 is presented in Figure 7. The biplot summarizes risk and utility measures through two principal components, capturing approximately 81% of the total variation. Risk measures (red arrows) are strongly aligned with the PC1 axis, where most risk and utility measures are highly positively correlated, reflecting the risk-utility trade-off. For example, generators such as `MostlyAI` and `synthpop` are associated with high data utility but also with high risk relative to the other generators. The plot additionally reveals groups of SDGs with similar risk-utility profiles as well as outliers like `arf` with its particularly low utility regarding the measures *Wasserstein*, *Energy* and *Proximity*.

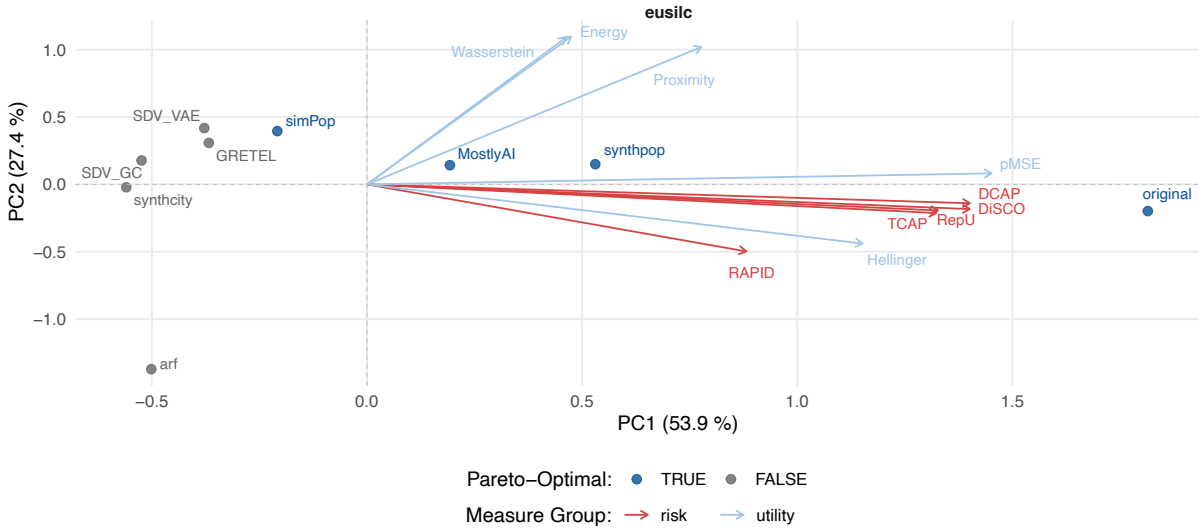

Figure 7: Joint PCA biplot. Points are SDGs (blue = Pareto-optimal); arrows are variable loadings with color indicating measure group (utility vs risk). Arrow length shows a measure's influence on the PCs; a point's projection approx. in direction of an arrow indicates association with that measure.

We also apply the proposed measure of alignment with PC1 (cf. Item 1 in Subsection 2.2.6). Table 4 shows that the first principal component (PC1) correlates strongly with both the utility composite ($\rho = 0.74$) and the risk composite ($\rho = 0.94$). The linear regression yielded $R^2 = 0.99$, indicating that PC1 is almost entirely explained by the two composites. Thus, PC1 can be interpreted as a dominant axis capturing both higher utility and higher risk simultaneously.

df	ρ_{s,U_c}	ρ_{s,R_c}	ρ^2_{s,U_c}	ρ^2_{s,R_c}	R^2_{joint}
eusilc	0.74	0.94	0.54	0.88	0.99

Table 4: Alignment of PC1 with utility and risk composites

Figure 8 shows the diagnostic plot of a robust version of the PCA (Hubert, Rousseeuw, and Vanden Branden 2005) of risk and utility measures, as described in Subsection 2.2.6 in Item 1. This diagnostic plot can provide valuable insights into which anonymization approaches exhibit distinctive performance.

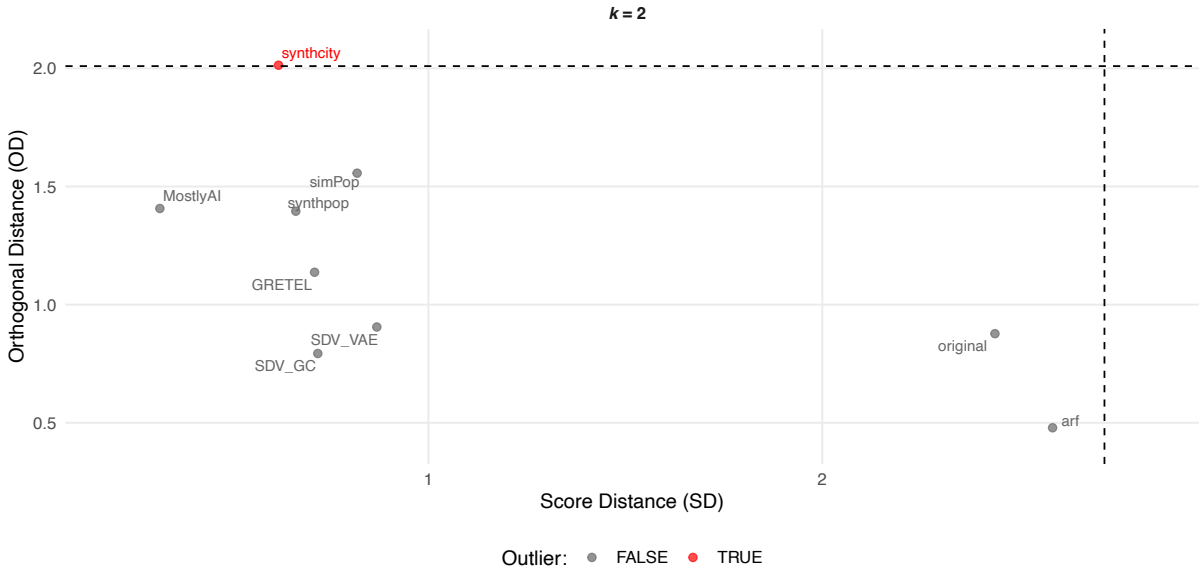

Figure 8: Robust PCA diagnostics ("outlier map"). The x-axis shows the Score Distance (SD) in the retained PC space (spanned by $k = 2$ components), and the y-axis shows the Orthogonal Distance (OD) to that space. Each point is an SDG; color indicates the robust outlier flag. Large SD indicates leverage within the PC space; large OD indicates poor reconstruction (the observation lies far outside the subspace). Cutoffs follow Hubert, Rousseeuw, and Vanden Branden (2005, p. 66): SD is compared to $\sqrt{\chi^2_{k,0.975}}$ and OD to $\Phi^{-1}(0.975)$.

We employ a blockwise PCA approach: one PCA on the utility indicators and a separate PCA on the risk indicators. We then plot PC1 from the utility block against PC1 from the risk block (PC1(utility) vs. PC1(risk)). With monotone, min-max standardized measures, these first components capture the dominant "high-utility" and "high-risk" directions, so their joint scatter approximates the composite R-U map while remaining model-based. As illustrated in Figure 9, the blockwise PCA (cf. Section 2.2.6, item 2) produces a representation that closely resembles the composite risk-utility (R–U) map (cf. Figure 3). Since PC1 from each block is a standardized linear transformation of the original measures, it effectively functions as a weighted mean. The stacked bar charts in Figure 9 display each measure's contribution to its corresponding PC1, calculated from squared loadings (representing variance contributions) and normalized to 100%. The relatively uniform distribution of these contributions indicates that no single measure dominates either principal component, thereby explaining the similarity between the blockwise PCA representation and the mean-based composite approach. The implicit slope (cf. Figure 3) of the line connecting each SDG to the next on the Pareto front reflects the trade-off between risk and utility. While the blockwise PCA approach adds computational complexity and reduces interpretability compared to simple averaging, it provides empirical validation that our measures are indeed homogeneous (uniform loadings) and justifies dimensionality reduction. In scenarios with heterogeneous or redundant measures, this approach would reveal structure that simple averaging obscures.

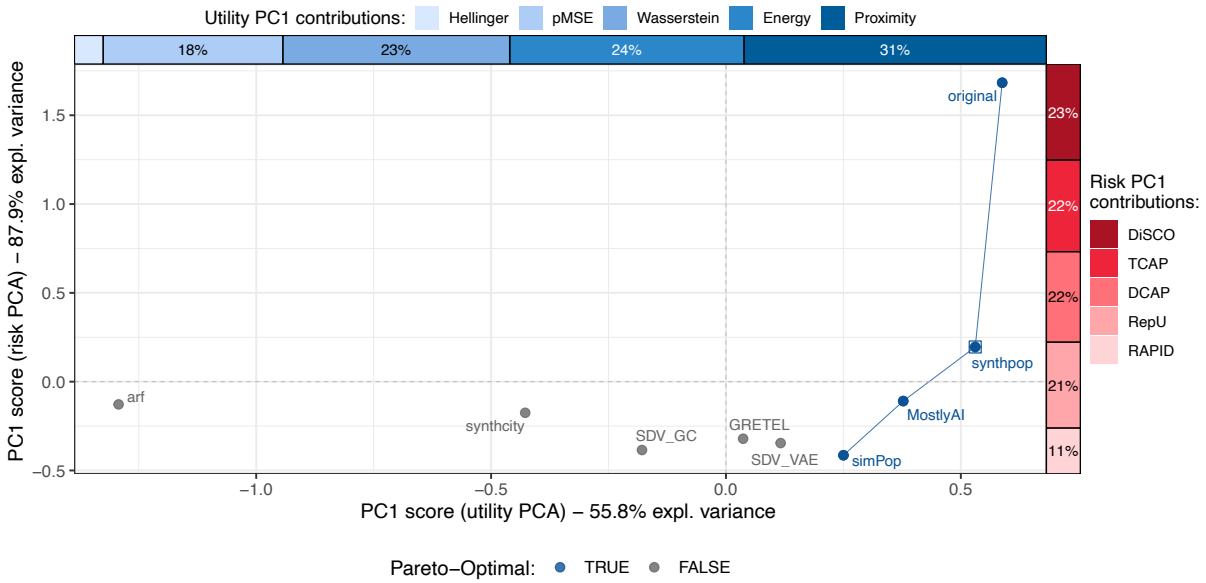

Figure 9: Blockwise PCA summary. X-axis: PC1 from utility measures; Y-axis: PC1 from risk measures (points = SDGs; blue = Pareto-optimal). Stacked bars show each measure's contribution to PC1, computed from squared loadings and normalized to 100% (labels shown for contributions $\geq 5\%$). Edge colors indicate loading sign (black = positive, red = negative).

4 Discussion

4.1 Comparative Overview of Visualization Methods

We established criteria to compare the strengths and weaknesses of the different visualization methods. The evaluation of these capabilities is shown in Table 5. No single method excels across all criteria, suggesting that combining multiple visualization approaches often provides the most complete understanding of risk-utility trade-offs.

Capability	**Heat-maps**	**Dot Plots**	**Composite Scatterplots**	**PCP**	**Radial Profile Plots**	**PCA Biplots**
Analyitcal Capabilities						
Within-block correlations between measures in same block	∼	✓	×	✓	∼	✓
Detecting systematic differences across methods	✓	✓	✓	✓	✓	✓
Distribution risk and utility	∼	∼	∼	∼	×	✓
Uncertainty depiction	×	×	✓	×	×	✓
Outlier detection	∼	∼	✓	∼	∼	✓
Trade-off Visualization						
Between-block correlations: Risk-utility tradeoff visualization	✓	✓	∼	∼	∼	✓
Displaying of Pareto-optimal methods	✓	✓	✓	✓	✓	✓
Displaying Pareto-Front	×	×	✓	×	×	✓
Support of acceptable thresholds	✓	✓	✓	✓	∼	✓
Usability & Scalability						
Scalability with number of methods	✓	∼	✓	✓	×	✓
Compare methods across multiple datasets	∼	×	✓	×	×	✓
Intuitive interpretation for non-technical stakeholders	✓	✓	✓	✓	✓	∼

Table 5: Capability comparison across visualization methods, organized by analytical capabilities, trade-off visualization, and usability (✓ = good, × = poor/unsupported, ∼ = partial/mixed).

4.2 Method-Specific Considerations

When using composite scores for Pareto identification, reliability depends on whether measures within each block show reasonably consistent values for each method. A method with highly disparate measure values may appear Pareto-optimal based on its mean but has critical vulnerabilities the composite obscures (because of outliers). Heatmaps reveal within-method disparities through mixed colors within rows, while dot plots clearly show the spread of individual measure values. Alternatively, the range or standard deviation within methods can be calculated, or blockwise PCA used to check if PC1 explains most variance with uniform loadings. When substantial disparities exist, measure-specific examination provides necessary context beyond composite analysis.

For initial screening and detailed inspection of individual measures, heatmaps and dot plots provide complementary strengths. Heatmaps are especially well suited for initial diagnostics and for communicating metric-level performance in a concise and structured way. The choice of layout depends on the comparison task: methods-as-rows (with measures as columns) favors holistic evaluation across methods, while methods-as-columns (with measures as rows) benefits detailed per-method inspection. Dot plots can enhance heatmap analysis when visualizing variation across risk and utility dimensions is needed, which heatmaps cannot deliver as effectively. Dot plots further excel in showing individual measure values with minimal chart junk. However, dot plots may suffer from overplotting when too many methods or measures are displayed; in such cases, jittering or small multiples can still preserve readability.

Composite scatterplots – whether based on simple averages or blockwise PCA – provide an intuitive risk-utility visualization that is accessible to both technical and non-technical audiences. The main advantage of using blockwise PCA over simple averaging is empirical validation: if PC1 in each block explains a high proportion of variance (e.g., >70%) and loadings are relatively uniform, this justifies the dimensionality reduction. When loadings are heterogeneous, blockwise

PCA reveals which measures dominate each composite, information that simple averaging obscures. However, the number of observations – i.e., the number of anonymization approaches – is typically small in practice, which limits the statistical reliability of internal consistency measures such as Cronbach's α and McDonald's ω for composite scores. For this reason, we report these coefficients when composite scores are used but interpret them with caution. In low-n settings, a more robust strategy may be to rely on conceptual grouping of measures and direct inspection of correlation patterns rather than on formal internal consistency statistics alone.

For multivariate profile visualization, radial profile charts can be used to explore and communicate individual performance patterns. To preserve readability, the number of anonymization approaches and measures displayed should be limited, as overlapping polygons become difficult to distinguish. Their distinctive "fingerprint" appearance makes them memorable for stakeholder communication, as the multivariate structure is conveyed in a single polygon for each method – a gestalt representation that PCPs cannot provide as intuitively. However, PCPs often provide a more scalable alternative to radial profile charts, particularly when the number of measures is high, as they preserve readability and pattern detection capabilities at higher dimensions.

PCA-based biplots not only visualize method performance but also reveal the relationships among measures themselves. While the biplot visualization facilitates the simultaneous comparison of multiple risk and utility measures, its interpretive value depends on the extent to which the first two principal components capture the overall variance in the data. If these two dimensions explain only a small proportion of the variance, the plot may provide a distorted view of the risk and utility properties of the anonymization strategies. In such cases, examining additional components or using alternative dimensionality reduction techniques may be necessary.

PCA-based biplots are only applicable when all anonymization approaches under comparison use the same set of risk and utility measures. This becomes problematic when the disclosure risk of a non-perturbative anonymization approach is evaluated using metrics such as *k-anonymity* (Sweeney 2002), which are not applicable to perturbative anonymization approaches that, e.g., swap values of quasi-identifiers. Because the set of relevant metrics may not overlap, direct comparison in a single biplot is not appropriate in such scenarios.

In the right use cases, such as for comparing the performance of various SDGs, biplots can provide clear insights into the trade-off between disclosure risk and data utility. In this context, the biplot not only highlights overall performance but also reveals how risk and utility measures relate to each other, supporting a more nuanced understanding of anonymization quality. To enhance biplot interpretability, a sign convention can be established to ensure that the first principal component always correlates positively with utility measures (as briefly discussed in Section 2.2.6, Item 1). This convention makes the visualization more intuitive: anonymization approaches positioned toward the upper-right would consistently represent high utility and high risk, while those in the lower-left would indicate low utility and low risk. Without such a convention, the arbitrary sign of principal components can lead to confusion when comparing across datasets or analyses.

4.3 Extensions and Alternative Approaches

For specialized use cases, bivariate color scales (von Mayr 1874; Wainer and Francolini 1980) can encode two variables simultaneously – for instance, displaying both risk and utility dimensions through color intensity and hue in geographic or network visualizations.

We considered but did not pursue several alternative approaches. Nonlinear dimensionality reduction methods like t-SNE (van der Maaten and Hinton 2008) or UMAP (McInnes, Healy, and Melville 2020) could capture complex relationships but lack PCA's interpretable loadings. The interpretable self-organizing map (iSOM; Nagar, Ramu, and Deb, 2023) shows promise for dense Pareto-optimal sets with many candidates. However, in anonymization contexts where a limited number of approaches are typically compared, SOM grids become under-determined and highly sensitive to hyperparameters. The visualization approaches presented here are better

suited to small-sample scenarios common in practice.

Beyond the methods explored here, several directions warrant further investigation. Alternative composite construction methods such as blockwise-PCA-based, median-based, or stakeholder-weighted composites could improve Pareto identification robustness. Developing comparison frameworks that accommodate incomplete measure coverage would enhance practical applicability.

5 Conclusion

5.1 Summary of Contributions

This paper addresses anonymization approach selection as a genuine multivariate optimization problem. Traditional risk-utility visualizations typically compare a single risk measure against a single utility measure, though multiple measures are often calculated for each dimension. We present and systematically compare six visualization approaches for simultaneous evaluation of multiple risk and utility measures: heatmaps, dot plots, composite scatterplots, parallel coordinate plots, radial profile charts, and PCA-based biplots. Through systematic Pareto-optimal approach identification applied across all approaches, we demonstrate that simultaneously visualizing multiple measures provides richer evaluation than selecting single representatives. Our comparative analysis reveals that visualization choice should align with analytical objectives: PCA approaches excel at revealing measure relationships and multivariate structure, while simpler approaches facilitate initial screening or intuitive assessment.

5.2 Key Recommendations

Effective anonymization approach selection requires integrating technical risk-utility assessment with broader organizational considerations including legal frameworks, data sensitivity, and institutional risk tolerance (Templ 2017; Hundepool et al. 2012). For the technical assessment component, we recommend employing multiple complementary visualizations tailored to analytical objectives. For initial screening, heatmaps efficiently reveal overall performance patterns while dot plots clearly display individual measure values across risk and utility dimensions. For deeper analysis, composite scatterplots provide intuitive risk-utility trade-off visualization, PCA-based biplots reveal measure relationships and multivariate structure, and parallel coordinate plots effectively display high-dimensional profiles. Pareto-optimal approaches can be identified and highlighted consistently across these visualization types. However, composite-based identification of Pareto-optimal approaches requires checking whether aggregation adequately represents each approach's performance across measures.

Acknowledgment & Disclosure

Acknowledgment

This work was funded by the Swiss National Science Foundation (SNSF) with grant "Harnessing event and longitudinal data in industry and health sector through privacy preserving technologies" (Grant Number 211751).

Disclosure of Interests

The authors have no competing interests to declare that are relevant to the content of this article.

A Synthetic Data Generators

Table 6: **Overview of the used synthetic data generators**

Name	Method(s)	Software	Authors	Version
synthpop	CART	R package	Nowok and Raab 2016	1.9.1
Synthetic Data Vault	GaussianCopula; VAE	Python package	Patki, Wedge, and Veeramachaneni 2016	1.18.0
simPop	Multinomial log-linear models; random draws; random forest (alt.)	R package	Templ et al. 2017	2.1.3
Mostly AI	Transformers; GANs; VAEs; autoregressive networks	Mostly AI (AT)	Mostly AI 2025	4.2.3
Gretel	Synthetic ACTGAN	Gretel Labs (US)	Gretel 2025	0.22.16
arf	Advanced random forests	R package	Watson et al. 2023	0.2.0
synthcity	Tabular GAN	Python package	Qian, Cebere, and van der Schaar 2023	0.2.11

References

Brand, Ruth. 2002. "Microdata protection through noise addition." In *Inference Control in Statistical Databases: From Theory to Practice,* 1st ed., edited by Josep Domingo-Ferrer, 2316:97–116. Lecture Notes in Computer Science. Berlin, Heidelberg, Germany: Springer, April 23, 2002. https://doi.org/10.1007/3-540-47804-3_8.

Cleveland, William S., and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *Journal of the American Statistical Association* 79 (387): 531–554. JSTOR: 2288400.

Cronbach, Lf. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *Psychometrika* 16 (3): 297–334. https://doi.org/10.1007/BF02310555.

Dankar, F. K., and M. K. Ibrahim. 2022. *A new PCA-based utility measure for synthetic data evaluation.* https://doi.org/10.48550/arXiv.2212.05595. arXiv: 2212.05595 [cs.DB].

Dankar, Fida K., Mahmoud K. Ibrahim, and Leila Ismail. 2022. "A Multi-Dimensional Evaluation of Synthetic Data Generators." *IEEE Access* 10:11147–11158. https://doi.org/10.1109/ACCESS.2022.3144765.

Defays, D., and M. N. Anwar. 1998. "Masking microdata using micro-aggregation." *Journal of Official Statistics* 14 (4): 449–461.

Duan, Rui, Jiayi Tong, Alex J. Sutton, David A. Asch, Haitao Chu, Christopher H. Schmid, and Yong Chen. 2023. "Origami Plot: A Novel Multivariate Data Visualization Tool That Improves Radar Chart." *Journal of Clinical Epidemiology* 156 (April): 85–94. https://doi.org/10.1016/j.jclinepi.2023.02.020.

Duncan, George, Sallie Keller-McNulty, and Lynne Stokes. 2001. *Disclosure risk vs data utility: The R-U confidentiality map.* Technical Report LA-UR-01-6428. Los Alamos, New Mexico: National Institute of Statistical Sciences.

Emmerich, Michael T. M., and André H. Deutz. 2018. "A Tutorial on Multiobjective Optimization: Fundamentals and Evolutionary Methods." *Natural Computing* 17, no. 3 (September): 585–609. https://doi.org/10.1007/s11047-018-9685-y.

Franconeri, Steven L., Lace M. Padilla, Priti Shah, Jeffrey M. Zacks, and Jessica Hullman. 2021. "The Science of Visual Data Communication: What Works." *Psychological Science in the Public Interest* 22, no. 3 (December): 110–161. https://doi.org/10.1177/15291006211051956.

Gabriel, K. R. 1971. "The Biplot Graphic Display of Matrices with Application to Principal Component Analysis." *Biometrika* 58 (3): 453–467. https://doi.org/10.2307/2334381.

Gesis. 2024. *Series: European Union Statistics on Income and Living Conditions (EU-SILC).* https://www.gesis.org/en/missy/metadata/EU-SILC/. Accessed: 2024-05-13.

Gouweleeuw, J. M., P. Kooiman, Leon Willenborg, and Peter-Paul de Wolf. 1998. "Post randomisation for statistical disclosure control: Theory and Implementation." *Journal of Official Statistics* 14 (4): 463–478.

Gretel. 2025. *Gretel Synthetic Data Platform.*

Hayes, Andrew F., and Jacob J. Coutts. 2020. "Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…" *Communication Methods and Measures* 14, no. 1 (January): 1–24. https://doi.org/10.1080/19312458.2020.1718629.

Hellinger, E. 1909. "Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen." *Journal für die reine und angewandte Mathematik* 136:210–271.

Hornby, Ryan, and Jingchen Hu. 2021. *Identification Risks Evaluation of Partially Synthetic Data with the IdentificationRiskCalculation R Package,* arXiv:2006.01298, April. https://doi.org/10.48550/arXiv.2006.01298. arXiv: 2006.01298 [stat].

Hosseinpour, Helia, Laura E. Matzen, Kristin M. Divis, Spencer C. Castro, and Lace Padilla. 2025. "Examining Limits of Small Multiples: Frame Quantity Impacts Judgments With Line Graphs." *IEEE Transactions on Visualization and Computer Graphics* 31, no. 3 (March): 1875–1887. https://doi.org/10.1109/TVCG.2024.3372620.

Hotelling, Harold. 1933. "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology* 24 (6): 417–441. https://doi.org/10.1037/h0071325.

Hubert, Mia, Peter J Rousseeuw, and Karlien Vanden Branden. 2005. "ROBPCA: A New Approach to Robust Principal Component Analysis." *Technometrics* 47, no. 1 (February): 64–79. https://doi.org/10.1198/004017004000000563.

Hundepool, Anco, Josep Domingo-Ferrer, Luisa Franconi, Sarah Giessing, Eric Schulte Nordholt, Keith Spicer, and Peter-Paul De Wolf. 2012. *Statistical Disclosure Control.* 1st ed. Wiley, August 17, 2012. https://doi.org/10.1002/9781118348239.

Inselberg, Alfred. 1985. "The Plane with Parallel Coordinates." *The Visual Computer* 1, no. 2 (August): 69–91. https://doi.org/10.1007/BF01898350.

Little, Claire, Richard Allmendinger, and Mark Elliot. 2025. "Synthetic Census Microdata Generation: A Comparative Study of Synthesis Methods Examining the Trade-Off Between Disclosure Risk and Utility." *Journal of Official Statistics* 41, no. 1 (March): 255–308. https://doi.org/10.1177/0282423X241266523.

Little, Claire, Mark Elliot, and Richard Allmendinger. 2022. "Comparing the Utility and Disclosure Risk of Synthetic Data with Samples of Microdata." In *Privacy in Statistical Databases,* edited by Josep Domingo-Ferrer and Maryline Laurent, 13463:234–249. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-13945-1_17.

Mas-Colell, Andreu, Michael D Whinston, and Jerry R. Green. 1995. "Chapter 16: Equilibrium and Its Basic Welfare Properties." In *Microeconomic Theory.* Oxford University Press.

Mayr, Georg von. 1877. *Die Gesetzmäßigkeit im Gesellschaftsleben: Statistische Studien.* München: Oldenbourg.

McDonald, Roderick P. 1999. *Test Theory A Unified Treatment.* 1. New York: Psychology Press. https://doi.org/110.4324/9781410601087.

McInnes, Leland, John Healy, and James Melville. 2020. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction,* arXiv:1802.03426, September. https://doi.org/10.48550/arXiv.1802.03426. arXiv: 1802.03426 [stat].

Miettinen, Kaisa. 1998. *Nonlinear Multiobjective Optimization.* Edited by Frederick S. Hillier. Vol. 12. International Series in Operations Research & Management Science. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-5563-6.

Mostly AI. 2025. *Mostly AI Synthetic Data Platform.*

Muralidhar, Krishnamurty, and Rathindra Sarathy. 2006. "Data Shuffling—A New Masking Approach for Numerical Data." *Management Science* 52, no. 5 (May): 658–670. https://doi.org/10.1287/mnsc.1050.0503.

Nagar, Deepak, Palaniappan Ramu, and Kalyanmoy Deb. 2023. "Visualization and Analysis of Pareto-optimal Fronts Using Interpretable Self-Organizing Map (iSOM)." *Swarm and Evolutionary Computation* 76 (February): 101202. https://doi.org/10.1016/j.swevo.2022.101202.

Nowok, Beata, and Gillian M. Raab. 2016. "synthpop: Bespoke Creation of Synthetic Data in R." *Journal of Statistical Software* 74 (11): 1–26. https://doi.org/10.18637/jss.v074.i11.

Nunnally, J. C., and I. H. Bernstein. 1994. *Psychometric Theory.* 3rd ed. New York: McGraw-Hill.

Patki, Neha, Roy Wedge, and Kalyan Veeramachaneni. 2016. "The Synthetic Data Vault." In *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA),* 399–410. Montreal, QC, Canada: IEEE, October. https://doi.org/10.1109/DSAA.2016.49.

Pau, David, Camille Bachot, Charles Monteil, Laetitia Vinet, Mathieu Boucher, Nadir Sella, and Romain Jegou. 2025. "Comparison of anonymization techniques regarding statistical reproducibility." *PLOS Digital Health* 4, no. 2 (February): 1–19. https://doi.org/10.1371/journal.pdig.0000735.

Pearson, Karl. 1901. "LIII. On lines and planes of closest fit to systems of points in space." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2 (11): 559–572. https://doi.org/10.1080/14786440109462720.

Qian, Zhaozhi, CeberBogdan-Constantine Cebere, and Mihaela van der Schaar. 2023. *Synthcity: Facilitating Innovative Use Cases of Synthetic Data in Different Data Modalities,* https://arxiv.org/abs/2301 https://doi.org/10.48550/arxiv.2301.07573.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing. Vienna.

Raab, Gillian M., Beata Nowok, and Chris Dibben. 2025. *Practical privacy metrics for synthetic data.* Pages: 1-23, arXiv:2406.16826. https://doi.org/10.48550/arXiv.2406.16826.

Rizzo, Maria L., and Gábor J. Székely. 2016. "Energy Distance." *WIREs Computational Statistics* 8, no. 1 (January): 27–38. https://doi.org/10.1002/wics.1375.

Snoke, Joshua, Gillian Raab, Beata Nowok, Chris Dibben, and Aleksandra Slavkovic. 2017. *General and specific utility measures for synthetic data* [in en]. ArXiv:1604.06651 [stat], June.

Sweeney, Latanya. 2002. "k-anonymity: A model for protecting privacy." *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10, no. 05 (October): 557–570. https://doi.org/10.1142/S0218488502001648.

Taber, Keith S. 2018. "The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education." *Research in Science Education* 48, no. 6 (December): 1273–1296. https://doi.org/10.1007/s11165-016-9602-2.

Taub, Jennifer, M. J. Elliot, Gillian Raab, Anne-Sophie Charest, Cong Chen, Christine M. O'Keefe, Michelle Pistner Nixon, Joshua Snoke, and Aleksandra Slavković. 2019. "Creating the best risk-utility profile: The synthetic data challenge." In *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality,* 1–21. Conference of European Statisticians. The Hague, Netherlands: UNECE.

Taub, Jennifer, Mark Elliot, Maria Pampaka, and Duncan Smith. 2018. "Differential Correct Attribution Probability for Synthetic Data: An Exploration." In *Privacy in Statistical Databases,* edited by Josep Domingo-Ferrer and Francisco Montes, 122–137. Cham: Springer International Publishing.

Templ, Matthias. 2017. *Statistical disclosure control for microdata: Methods and applications in R.* 1st ed. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-50272-4.

Templ, Matthias, and Bernhard Meindl. 2008. "Robust statistics meets SDC: New disclosure risk measures for continuous microdata masking." In *Privacy in Statistical Databases,* edited by Josep Domingo-Ferrer and Yücel Saygın, 5262:177–189. Lecture Notes in Computer Science. ISSN: 0302-9743, 1611-3349, PSD 2008, Istanbul, Turkey. Berlin, Heidelberg, Germany: Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-87471-3_15.

Templ, Matthias, Bernhard Meindl, Alexander Kowarik, and Olivier Dupriez. 2017. "Simulation of Synthetic Complex Data: The *R* Package **simPop**." *Journal of Statistical Software* 79 (10). https://doi.org/10.18637/jss.v079.i10.

Thorndike, Robert L. 1953. "Who Belongs in the Family?" *Psychometrika* 18, no. 4 (December): 267–276. https://doi.org/10.1007/BF02289263.

Tufte, Edward. 1983. *The Visual Display of Quantitative Information.* Cheshire: Graphics Press.

———. 1990. *Envisioning Information.* Cheshire: Graphics Press.

van der Maaten, Laurens, and Geoffrey Hinton. 2008. "Visualizing Data Using T-SNE." *Journal of Machine Learning Research* 9 (86): 2579–2605.

Vasershtein, L. N. 1969. "Markovskie processy na schetnom proizvedenii prostranstv, opisyvayushchie bol'shie sistemy avtomatov." In Russian, *Problemy peredachi informatsii* 5 (3): 64–72.

von Mayr, Georg. 1874. *Gutachten Über Die Anwendung Der Graphischen Und Geographischen Methode in Der Statistik.* München: J. Gotteswinter & Mössl.

Wainer, Howard, and Carl M. Francolini. 1980. "An Empirical Inquiry Concerning Human Understanding of Two-Variable Color Maps." *The American Statistician* 34 (2): 81–93. https://doi.org/10.2307/2684111. JSTOR: 2684111.

Watson, David S, Kristin Blesch, Jan Kapar, and Marvin N Wright. 2023. "Adversarial Random Forests for Density Estimation and Generative Modeling." In *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS),* vol. 206. Valencia: PMLR.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4, no. 43 (November): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.*

Wilkinson, Leland, and Michael Friendly. 2009. "The History of the Cluster Heat Map." *The American Statistician* 63, no. 2 (May): 179–184. https://doi.org/10.1198/tas.2009.0033.

Zinbarg, Richard E., William Revelle, Iftah Yovel, and Wen Li. 2005. "Cronbach's α , Revelle's β , and Mcdonald's ω_H : Their Relations with Each Other and Two Alternative Conceptualizations of Reliability." *Psychometrika* 70, no. 1 (March): 123–133. https://doi.org/10.1007/s11336-003-0974-7.