

# Understanding How Dimension Reduction Tools Work

**Dr. Cynthia Rudin**  
Duke University, USA

## Abstract

Dimension reduction (DR) techniques such as t-SNE, UMAP, and TriMap have demonstrated impressive visualization performance on many real world datasets. They are useful for understanding data and trustworthy decision-making, particularly for biological data. One tension that has always faced these methods is the apparent trade-off between preservation of global structure and preservation of local structure: past methods can either handle one or the other, but not both. In this work, our goal is to understand what aspects of DR methods are important for preserving both local and global structure. We leverage our insights to design a new algorithm for DR, called Pairwise Controlled Manifold Approximation Projection (PaCMAP), which preserves both local and global structure. Our work provides several unexpected insights into what design choices to make when constructing DR algorithms.

I will be discussing work from the following papers, as well as applications to bioinformatics, name-ethnicity classification, finance, and neurology:

Yingfan Wang, Haiyang Huang, Cynthia Rudin, Yaron Shaposhnik  
Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMAP, and PaCMAP for Data Visualization  
Journal of Machine Learning Research (JMLR), 2021  
<https://jmlr.org/papers/v22/20-1061.html>

Haiyang Huang, Yingfan Wang, Cynthia Rudin, and Edward P. Browne  
Towards a Comprehensive Evaluation of Dimension Reduction Methods for Transcriptomic Data Visualization  
Communications Biology (Nature), 2022.  
<https://www.nature.com/articles/s42003-022-03628-x>

Our code for PaCMAP is the winner of the 2023 John M. Chambers Statistical Software Award from the American Statistical Association.