Study Questions for “The art of using t-SNE for single-cell transcriptomics” by Kobak & Berens

1) What statistic do the authors use to quantify how well embeddings capture local structure in high-dimensional single-cell gene expression profiles? What about global structure? Think about what local and global mean in this context.

2) What does the perplexity parameter do in t-SNE? Why did the authors choose to use two different values in their final pipeline?

3) Why do the initial values for the t-SNE algorithm change the result? Do you think it makes sense to initialize with PCA?

4) Would you use the authors’ approach to placing new points on a t-SNE plot?

5) The authors argue that tuning how you use an embedding algorithm is potentially more important than which algorithm you use (e.g., t-SNE vs. UMAP). Does this “art” aspect comfort you or make you nervous?