

Q: Why use t-SNE vs. PCA?

A: Some pros and cons can be found here: <https://towardsdatascience.com/t-distributed-stochastic-neighbor-embedding-t-sne-bb60ff109561>. For example, t-SNE is more robust to outliers than PCA.

Q: What are the orange lines in the PCA plot?

A: The orange lines are those orthogonal projections of the points onto the "average line." See: <https://liorpachter.wordpress.com/2014/05/26/what-is-principal-component-analysis/>

Q (Q3): In the t-SNE plot in Section 1.4 above, why did I use `set_aspect('equal')`? Why didn't I choose to use it for the PCA plot?

A: t-SNE creates an embedding in which Euclidean distances between embedded points represent distances between source datapoints. Thus, a horizontal distance x has the same meaning as a vertical distance x . Thus, the plot should have equal aspect ratio.

PCA creates an embedding onto a plane, and there is no inherent meaning to distances on the plane, so there's no implication that horizontal distances count the same as vertical distances.