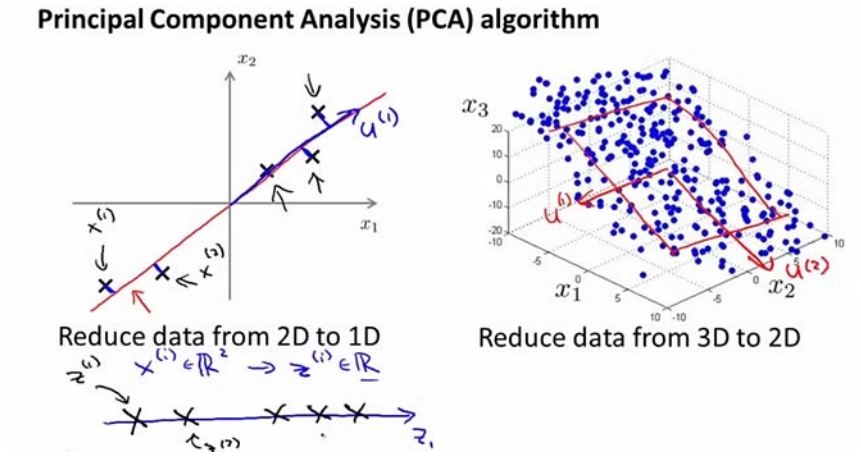
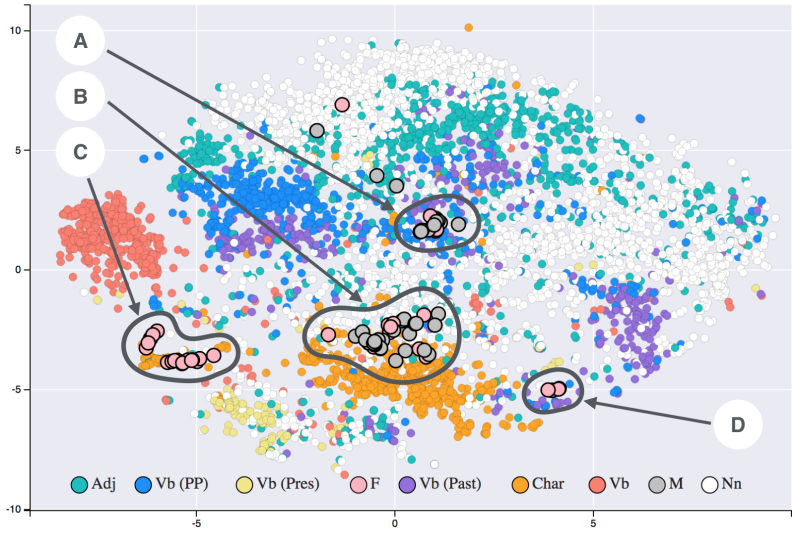
UMAP –adam (working on it now)

PCA (Noah): Principal Component Analysis is used with the unsupervised machine learning side of work, so essentially with the unlabeled/uncategorized data. The purpose of this ML tool is to help reduce the dimensions of our data that we are working with. The overall computational complexity of such algorithms can be simplified through this practice. The below image is an example of PCA used in action, where you can observe the distribution of data points shrinking down dimensionally.



t-SNE (Noah): T-distributed stochastic neighbor embedding is utilized for visualizing non-linear relationships in data that may be difficult to summarize otherwise. There are similar aspects that this technique uses to the PCA approach as well. However, PCA is better suited for applications where the goal is to identify the most important patterns that arise in the data. With t-SNE however, it is better to use when the goal is to visualize and explore the intricate and nuanced relationships between the data points in that low dimensional space. Below is what a t-sne distribution would entail for data point relationship purposes. These 2 methods overall can be used for fish data in such a way that synthesizes the potential patterns observed from the fish that will arise. This simplification can further help us create machine learning models that predict fish behavior in a more computationally less complex manner as well.



SNN - working on

Vst- working on rn

Log Normalize (Noah): This is a good method to utilize when you have data that needs to be categorized. Normalizing the data with the log function can ensure that the data we are working with is standardized as well. A formula that can achieve this for example is $X\text{-}X_{min}/X_{min}$ - X_{max}. This is useful for certain applications with our fish data that could potentially have skewed data or data that has a large range of values.

DeSeq2

GLM