standard deviation bandsUse notes to motivate method for assessing curve cluster quality, if curves were clustered with K-means using DTW as metrics. Explanation for variance is doable in PCA by using the proportions of variance explained by PCs (each PC represents a feature)

* Figure out how this can be done for K-means with DTW

Resources

* [Youtube vid](https://www.youtube.com/watch?v=m6FF69u6rRI) on clustering time series plots with k means and DTW
* [Walkthrough](https://www.youtube.com/watch?v=_K1OsqCicBY) of DTW algo
* [Paper](https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005896) that does time series clustering on transcriptomics seems to just use SD as a metric of quality kind of (See fig 3)

Normalization Techniques (Since we care about trends)

* Z-Scale Normalization
* MinMax Normalization

Notes

* DTW finds the best alignment mapping between 2 sequences
* K Means is a clustering algo
  + Method for deciding on the number of clusters to make is through brute force, measuring the elbow plot (variation per cluster vs number of clusters k)
* Silhouette Score measures how good clusters are
* Might consider adding a prediction interval
  + Not too sure if this makes sense
  + Seems doable since clusters are made of observations
* rCCR: Relative cross coalescence rate can be used to display how populations (idk if this is limited to 2 populations) diverged

**TODO:**

* **Assess cluster quality with something like** [**SD bands**](https://www.fidelity.com/learning-center/trading-investing/technical-analysis/technical-indicator-guide/bollinger-bands#:~:text=Bollinger%20Bands%20are%20envelopes%20plotted,Period%20and%20Standard%20Deviations%2C%20StdDev.) **or inertia**
  + **Higher** [**inertia**](https://tslearn.readthedocs.io/en/stable/gen_modules/clustering/tslearn.clustering.TimeSeriesKMeans.html?highlight=inertia) **is, worse the spread of curves in cluster might be**
* **Short term goal can be figuring out a way to create summaries of clusters**
  + **Peaks**
  + **Troughs**
  + **Slopes/Rates of change**
  + **Observable trends**
* **Maybe try a silhouette analysis or elbow curve to figure best number of clusters to use**
  + **Measures quality of the number of clusters used i think**

Running K-Means Clustering with DTW as distance metric

* Cluster curves (data points over time) as opposed to data points (typical K-Means)
* Centroids for DTW method are lines that minimize DBA

Silhouette analysis

* Used to find optimal number for clusters, K

Applications

* Forecast models
* Ability to classify by which cluster something fits into