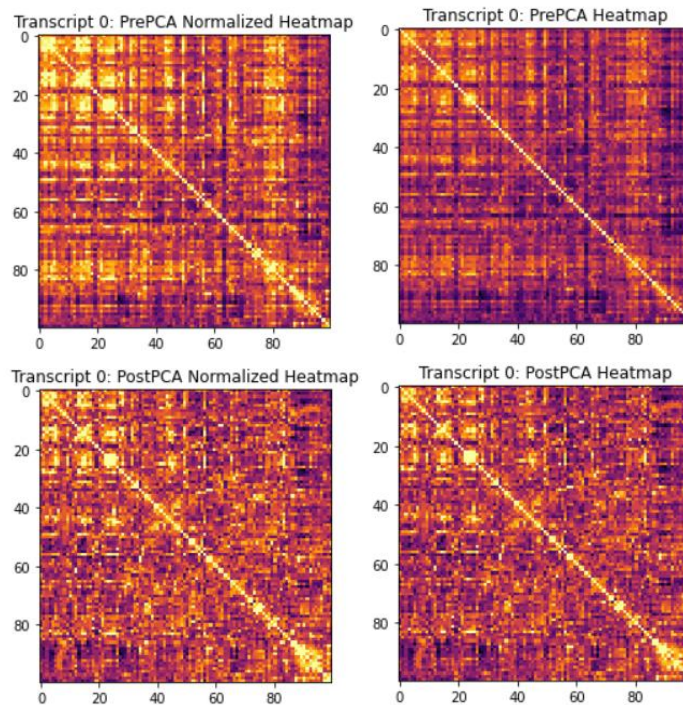# Understanding Results

- There are multiple opportunities for data comparison throughout the program.
- There are outputs of segmentation, cosine similarity heatmaps, embeddings, filtered sentences, excluded sentences, plots of clusters of semantically similar and different ideas.
  - Examples: PrePCA and PostPCA of raw and normalized cosine similarity maps can be seen below. The data is also saved into a .csv matrix so exact number values can be used to further evaluate how the system identifies sentences.

Transcript 0: PrePCA Normalized Heatmap

Transcript 0: PrePCA Heatmap

Transcript 0: PostPCA Normalized Heatmap

Transcript 0: PostPCA Heatmap

Pre PCA						
	0	1	2	3	4	5
0	1	0.336	0.412	0.335	0.412	0.329
1	0.336	1	0.447	0.441	0.613	0.45
2	0.412	0.447	1	0.386	0.579	0.539
3	0.335	0.441	0.386	1	0.495	0.237
4	0.412	0.613	0.579	0.495	1	0.555
5	0.329	0.45	0.539	0.237	0.555	1

Post PCA						
	0	1	2	3	4	5
0	1.001	0.028	0.045	0.085	0.039	-0.052
1	0.028	1	0.075	0.222	0.367	0.121
2	0.045	0.075	1	0.054	0.208	0.184
3	0.085	0.222	0.054	1.001	0.228	-0.143
4	0.039	0.367	0.208	0.228	1	0.208
5	-0.052	0.121	0.184	-0.143	0.208	1

- Comparing segments of related sentences to the heatmap can be a method of evaluating how successful the program was
- Evaluating the excluded sentences by identifying if they are unrelated to the text can help evaluate progress
- Finally, evaluating Clusters of similar/different ideas can be helpful in evaluating the success of the program in semantic embedding of long text documents