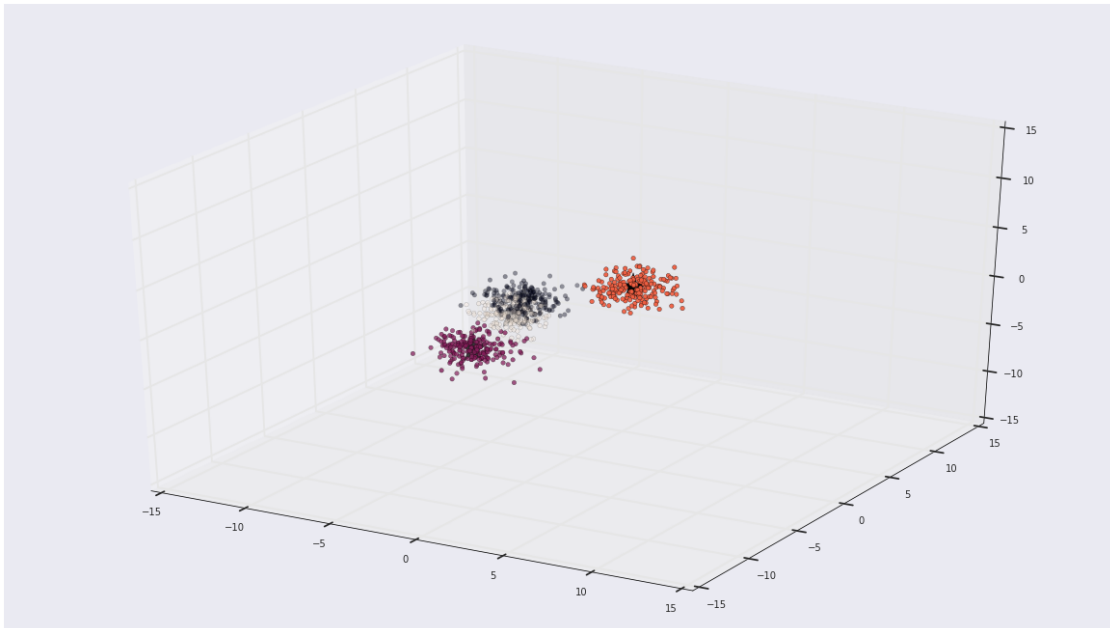
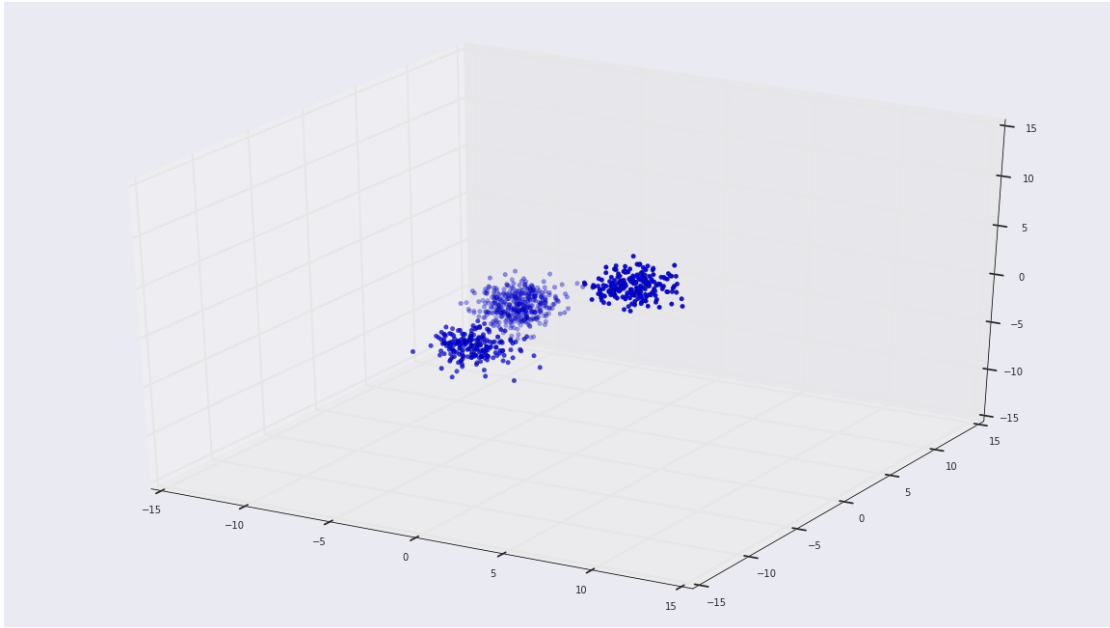


**Scott Viteri**  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
sviteri@mit.edu

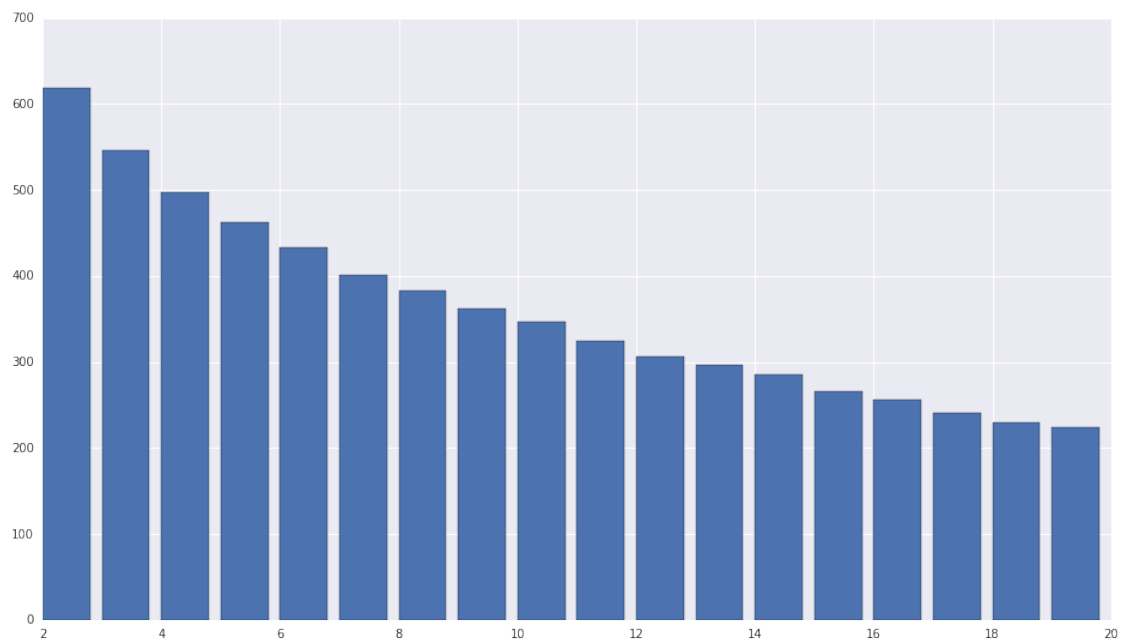
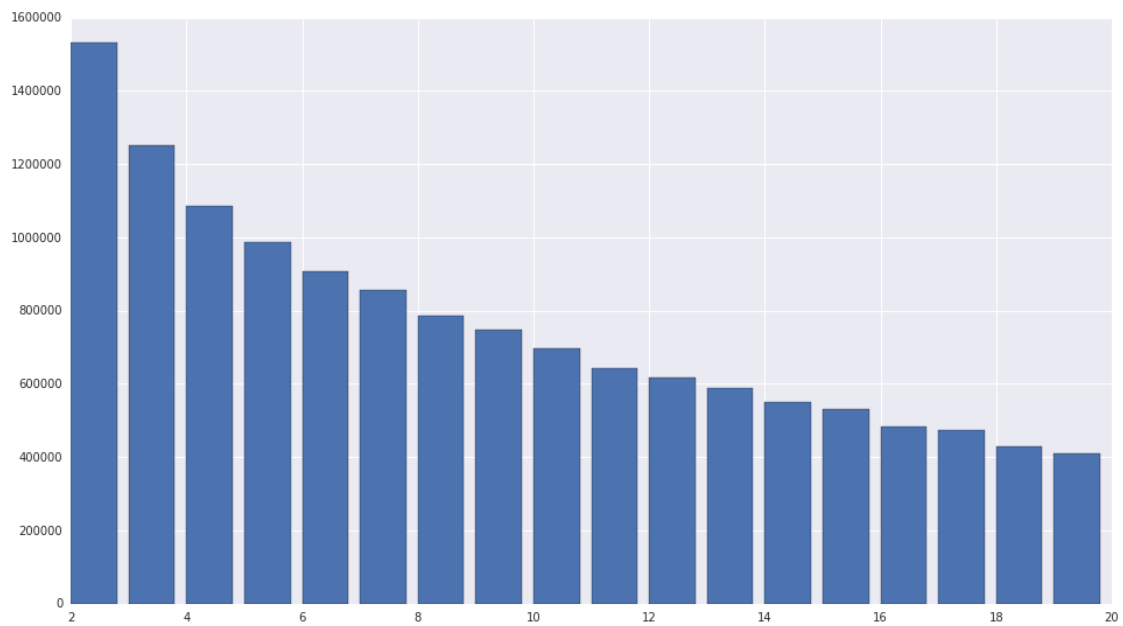
We apply the unsupervised learning methods of principle component analysis, multi-dimensional scaling, and gaussian mixture models to the animal feature dataset created by Osherson et al 1991. In the process, we compare and contrast structures found and generalization performance.

First, we load the relevant animal data.



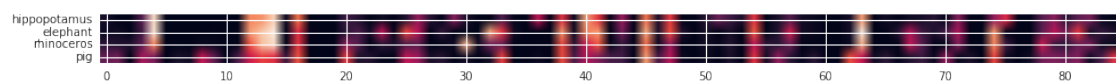


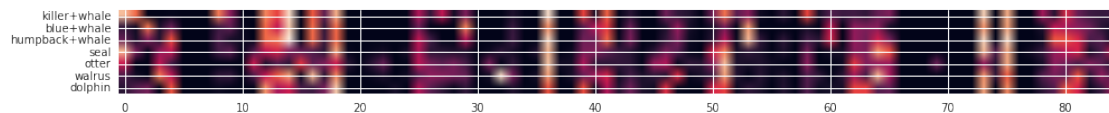
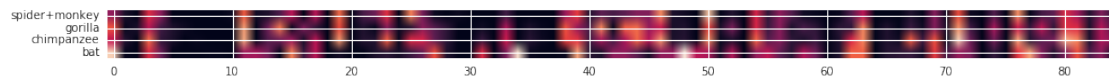
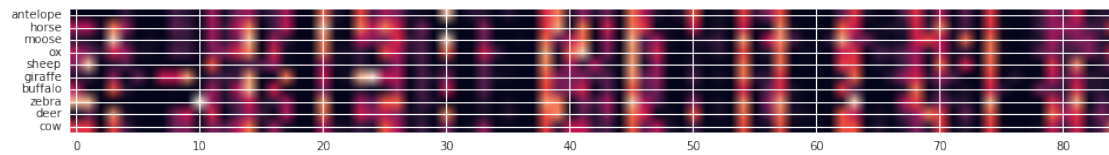
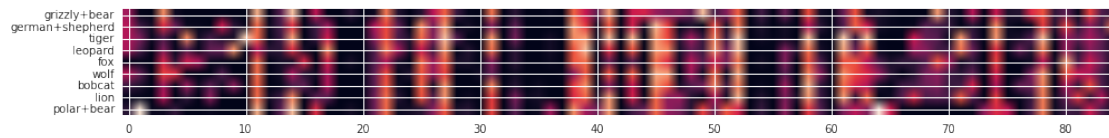
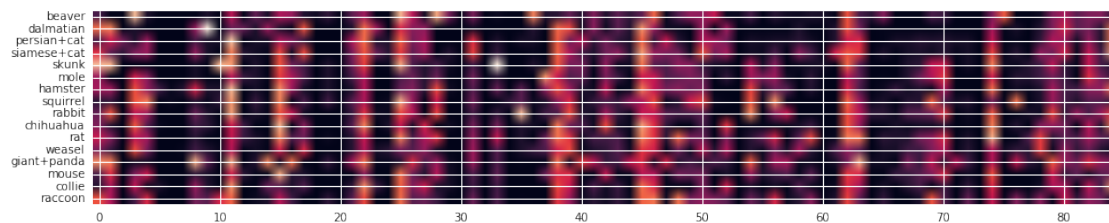
Then we plot the accuracy of clustering on the animal data as a function of the number of clusters  $k$ .



We learn that the k-means error rate for this data set drops off slowly with the number of variables. We set the future number of clusters to 6 as a tradeoff between small size and low error rate.

Plotted below are the six clusters.

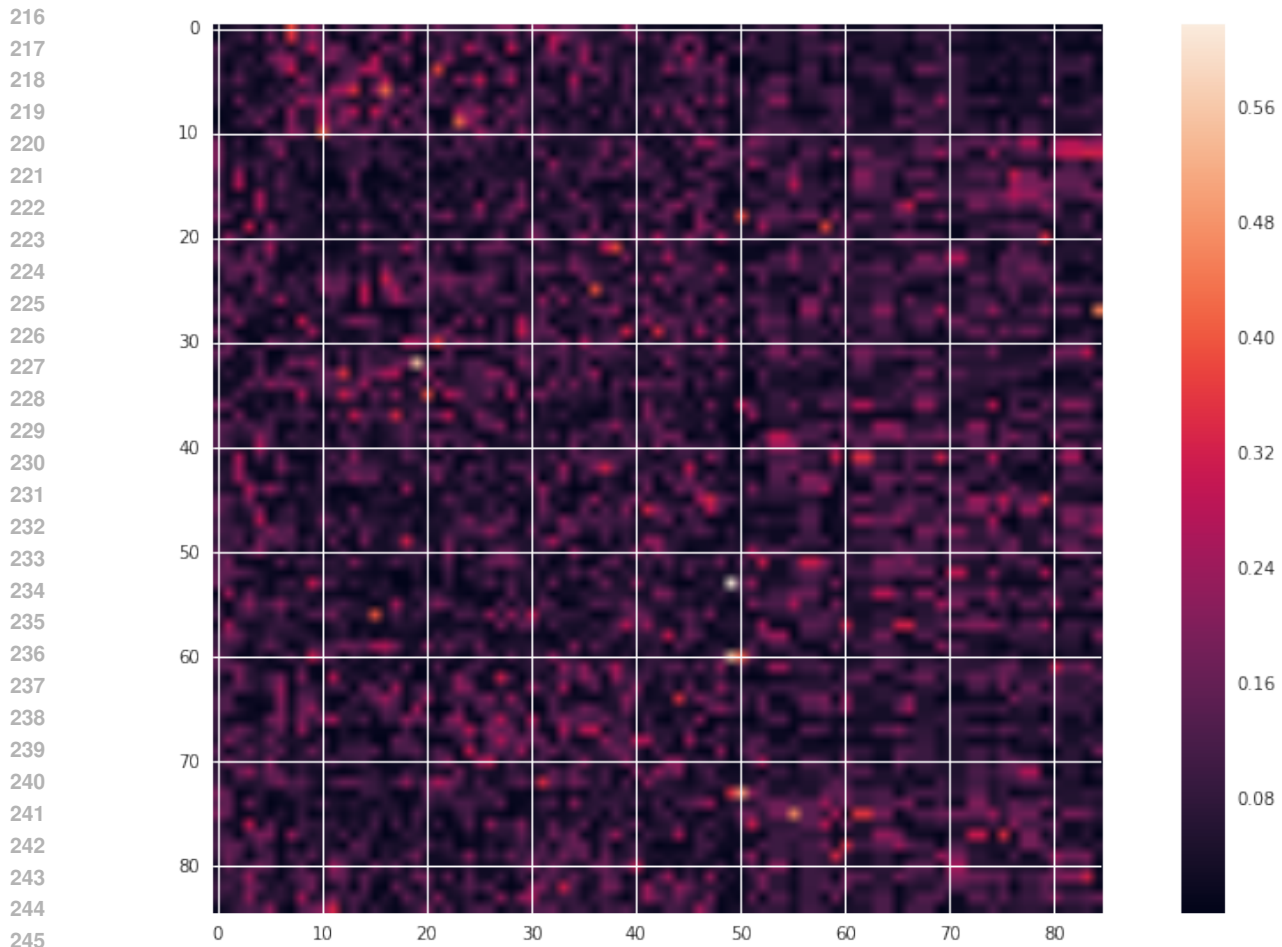




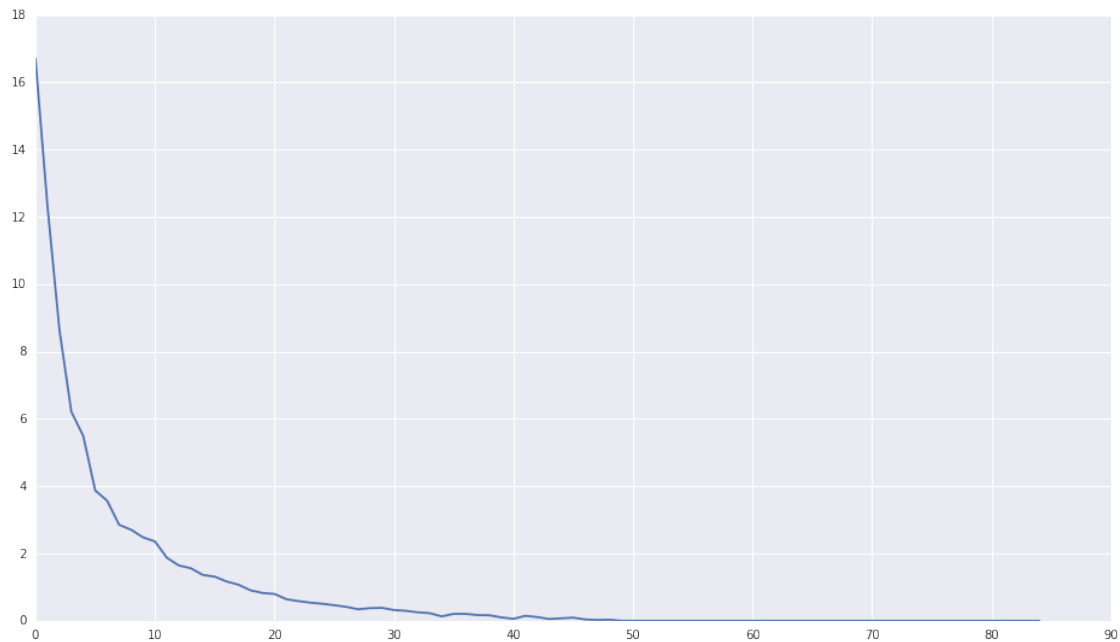
The clustering is reasonably intuitive with k-means. Perhaps we can get clearer separation by first running principle component analysis and projecting into the data into a lower dimensional subspace.

The steps are as follows: 1. Standardize data 2. Obtain the eigenvecs and eigenvals from the covariance or correlation matrix or perform SVD 3. Sort eigenvals in descending order and choose the k eigenvecs that correspond to the k largest eigenvals (choose some  $k \leq d$ ) 4. Construct the projection matrix  $W$  from the selected k eigenvecs 5. Transform the original dataset  $X$  via  $W$  to obtain a k-dim feature subspace  $Y$

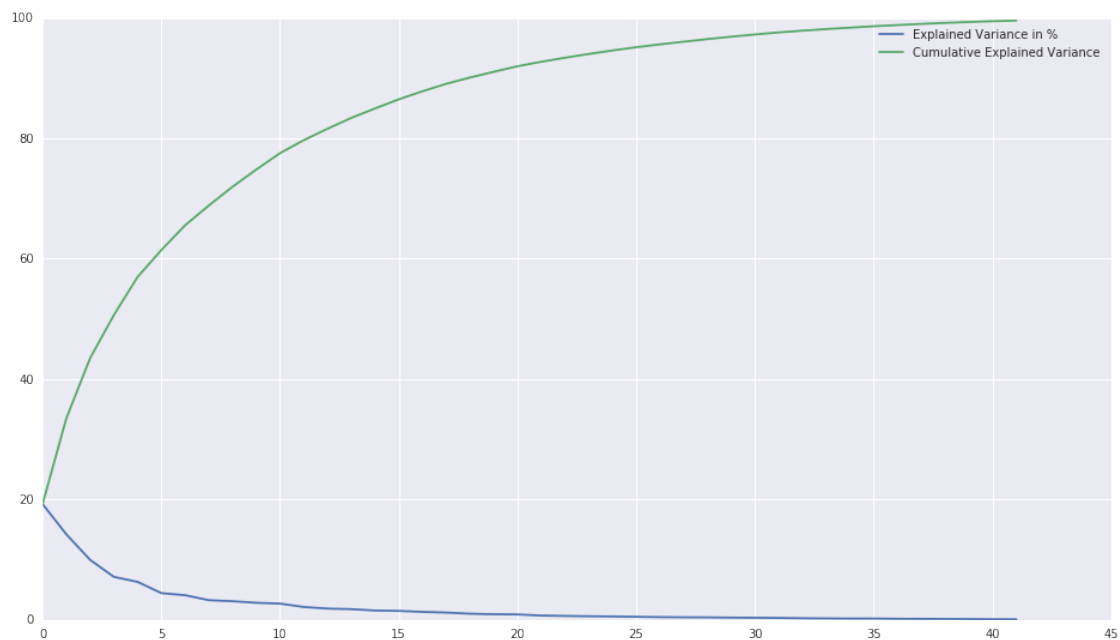
Performed steps 1 and 2 on data, and plotted eigenvals below.



3. Sort by largest eigenvalue, and plot below to verify sorted correctly.



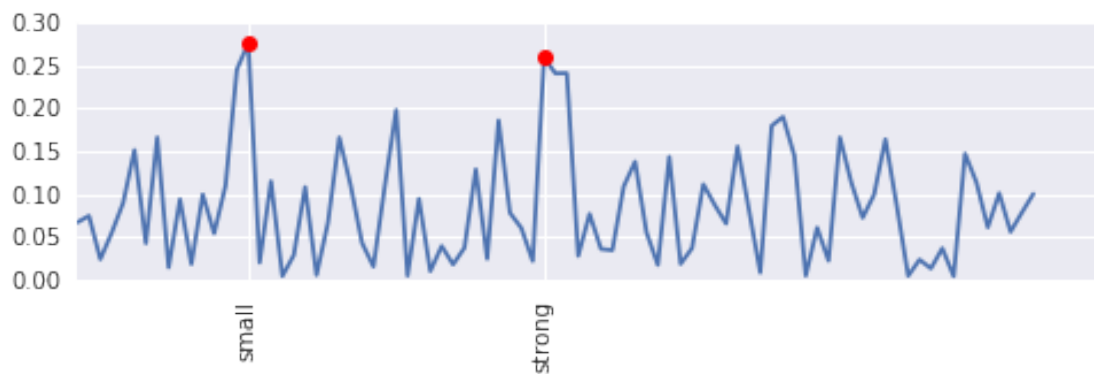
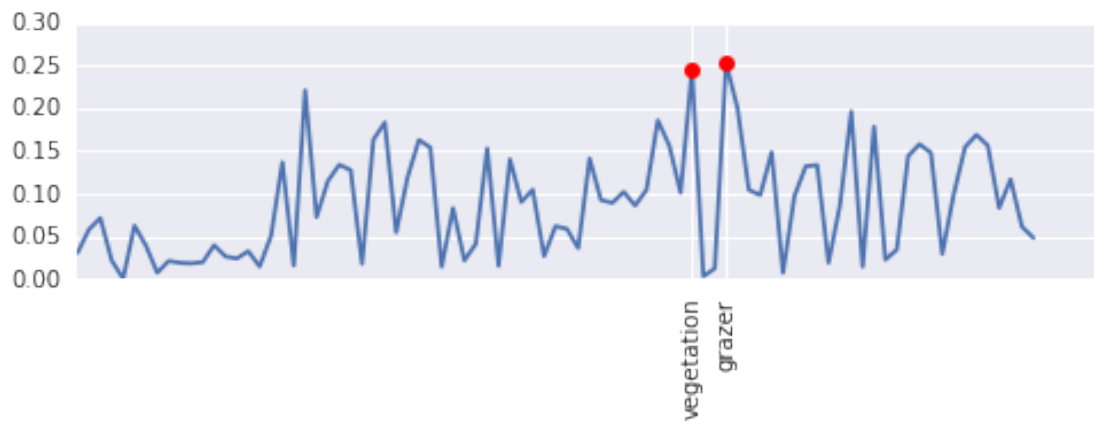
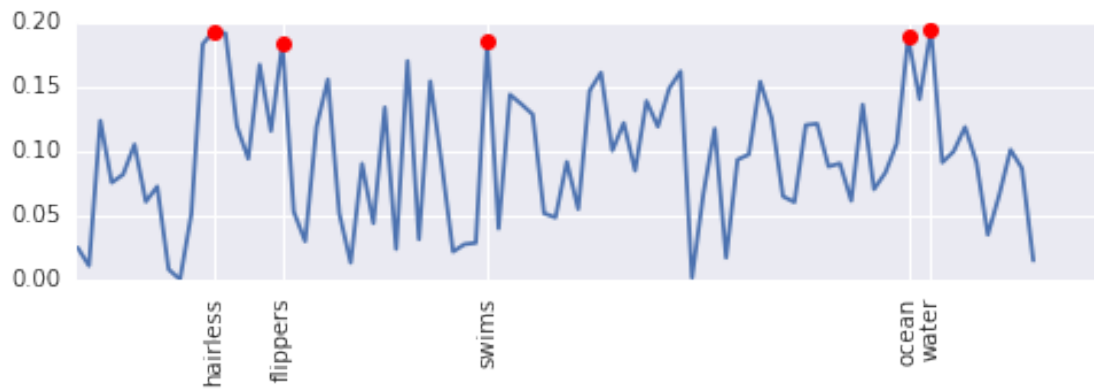
4. Constructed projection matrix, and plotted the amount of percentage of variance each 2D invariant eigenspace accounts for. The accumulated percentage is plotted above it.



6 eigenvectors strikes a good tradeoff between information and simplicity. Plot the animal data in the projected feature space.

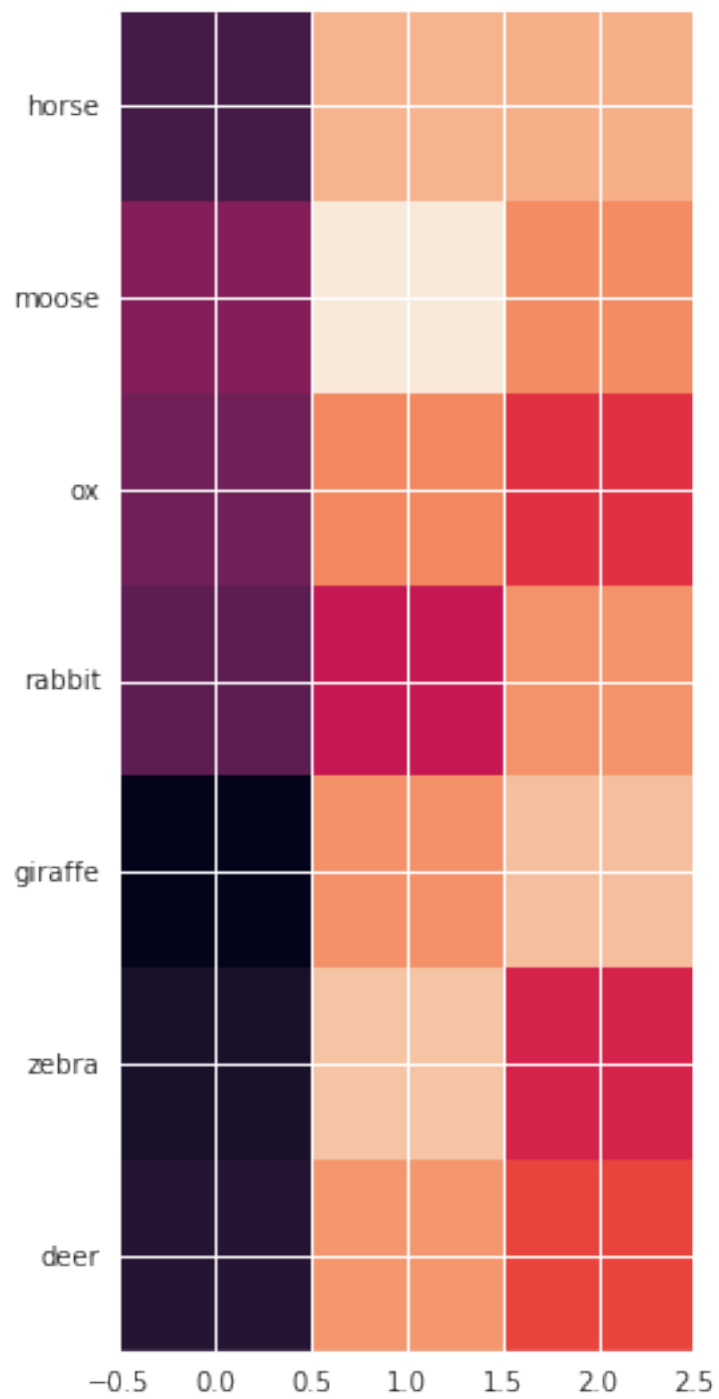


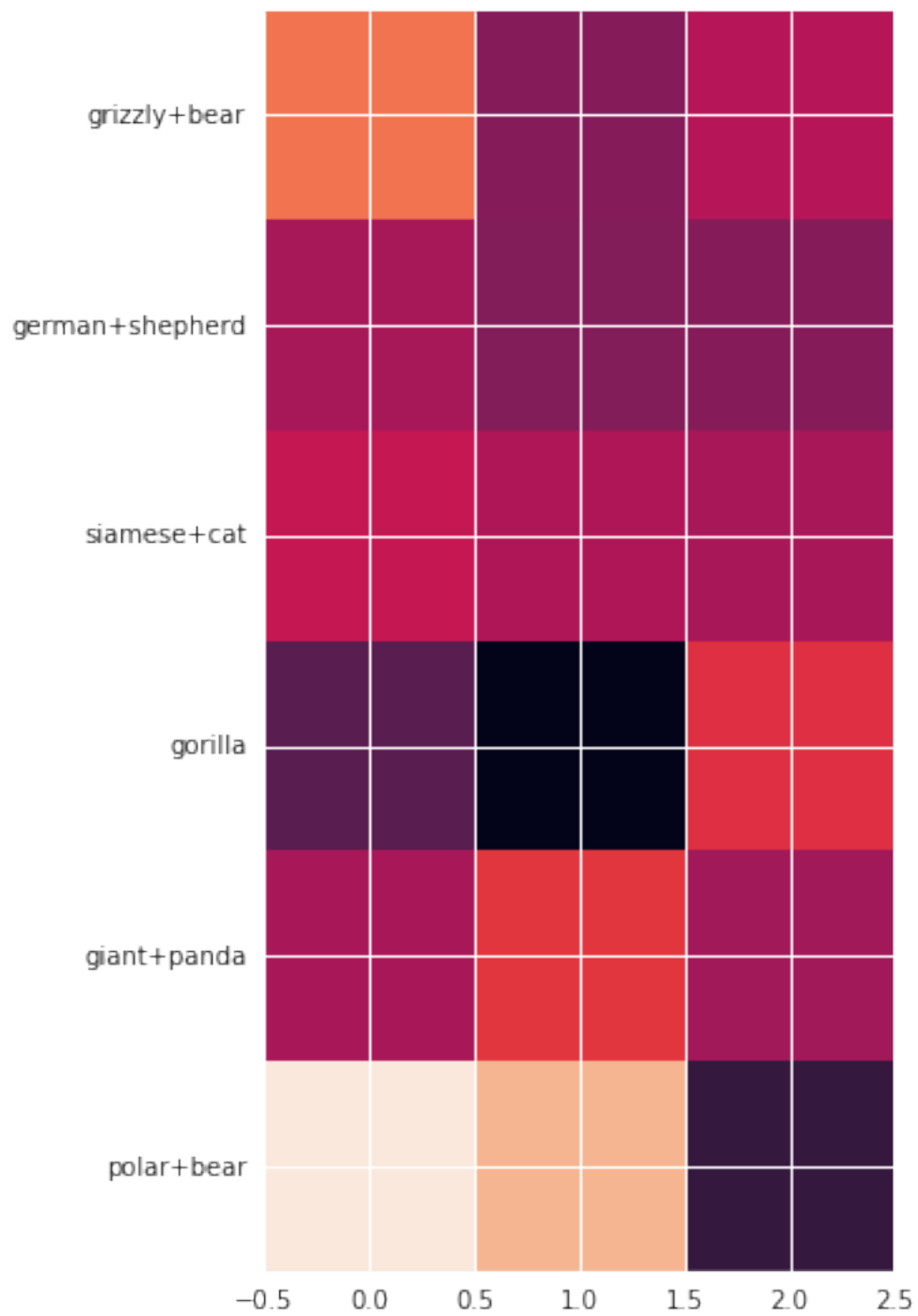
The x axis no longer corresponds directly to labelled features. Rather they correspond to linear combinations of labelled features. Plotted below are the first few eigenvectors, and the corresponding most relevant features in those eigenvectors.

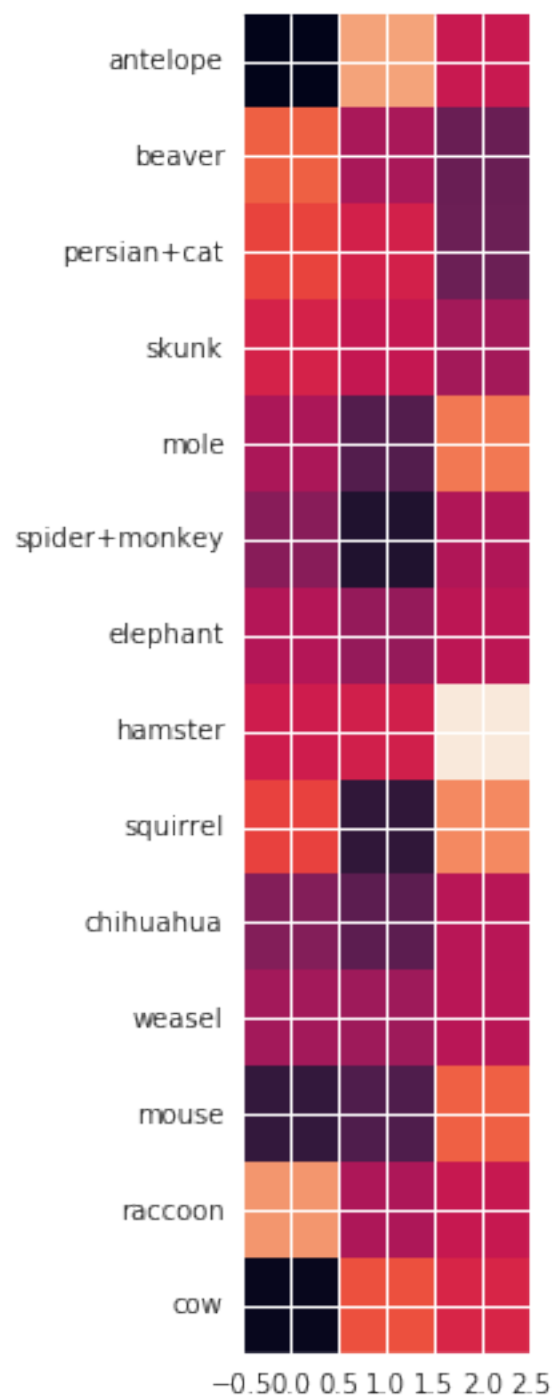


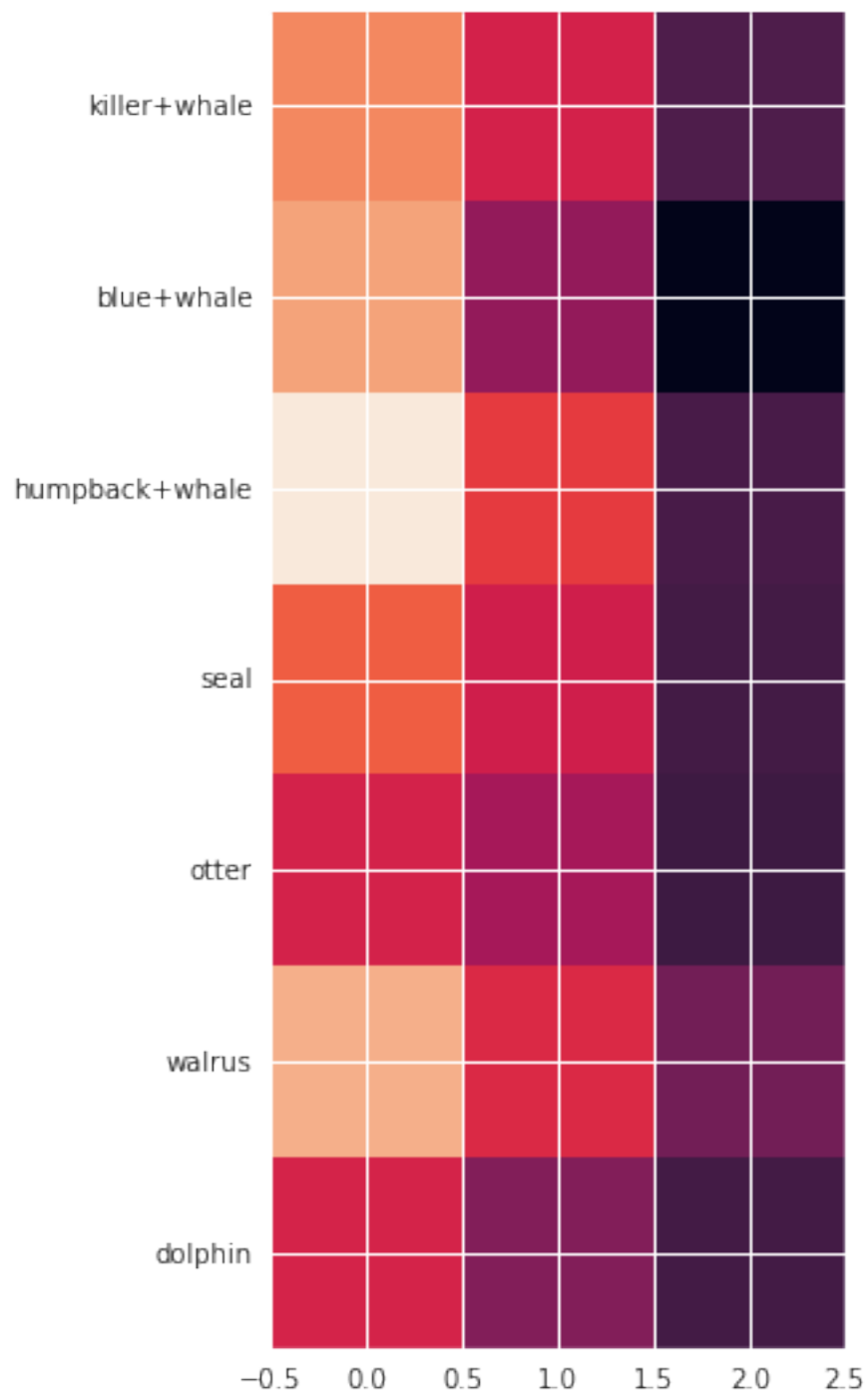
Plotted next are k-means clusters of 6, used on the reduced-dimensional data. The data has been reduced to 3 dimensions, but the output is less intuitive than the k-means case. Maybe this was not the appropriate dimensionality reduction.

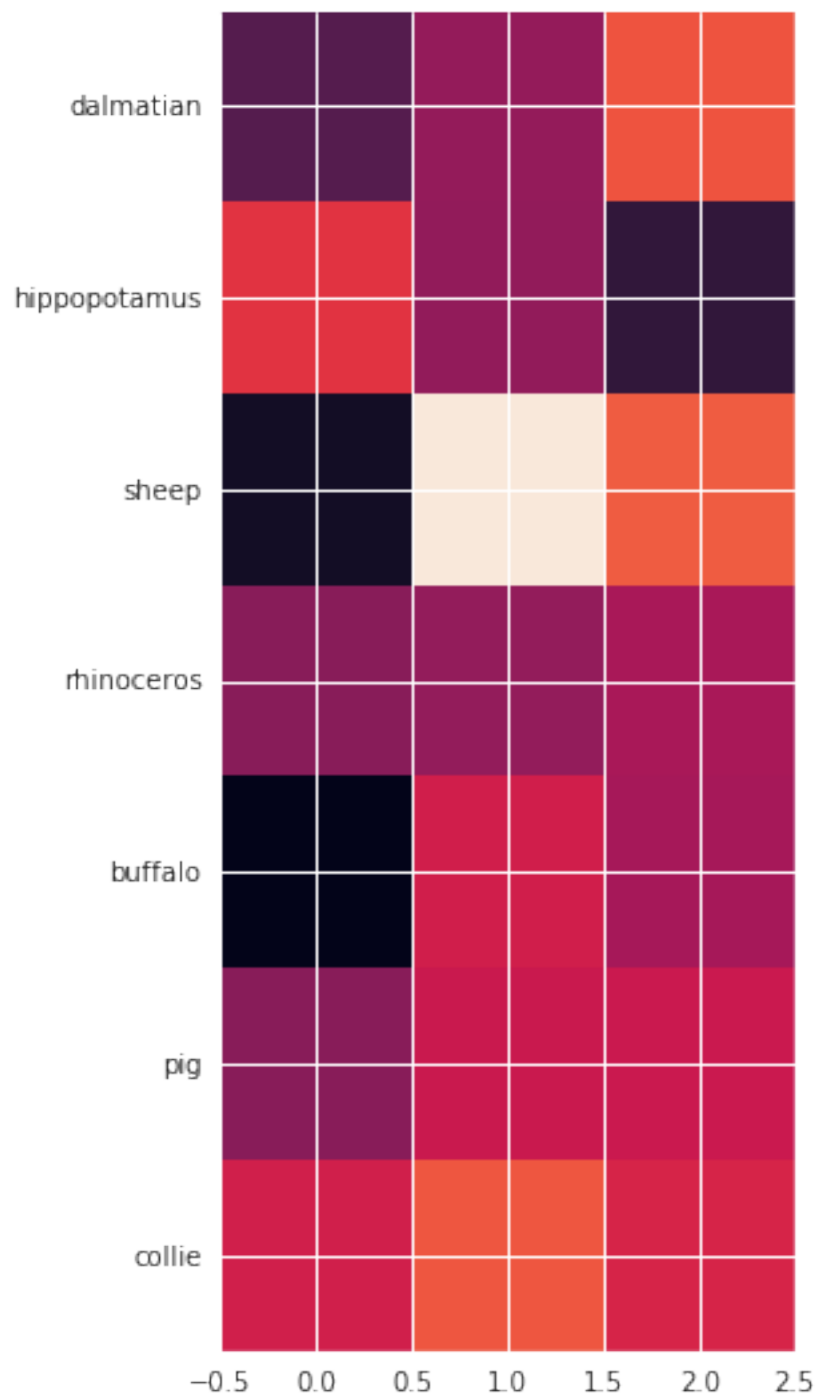


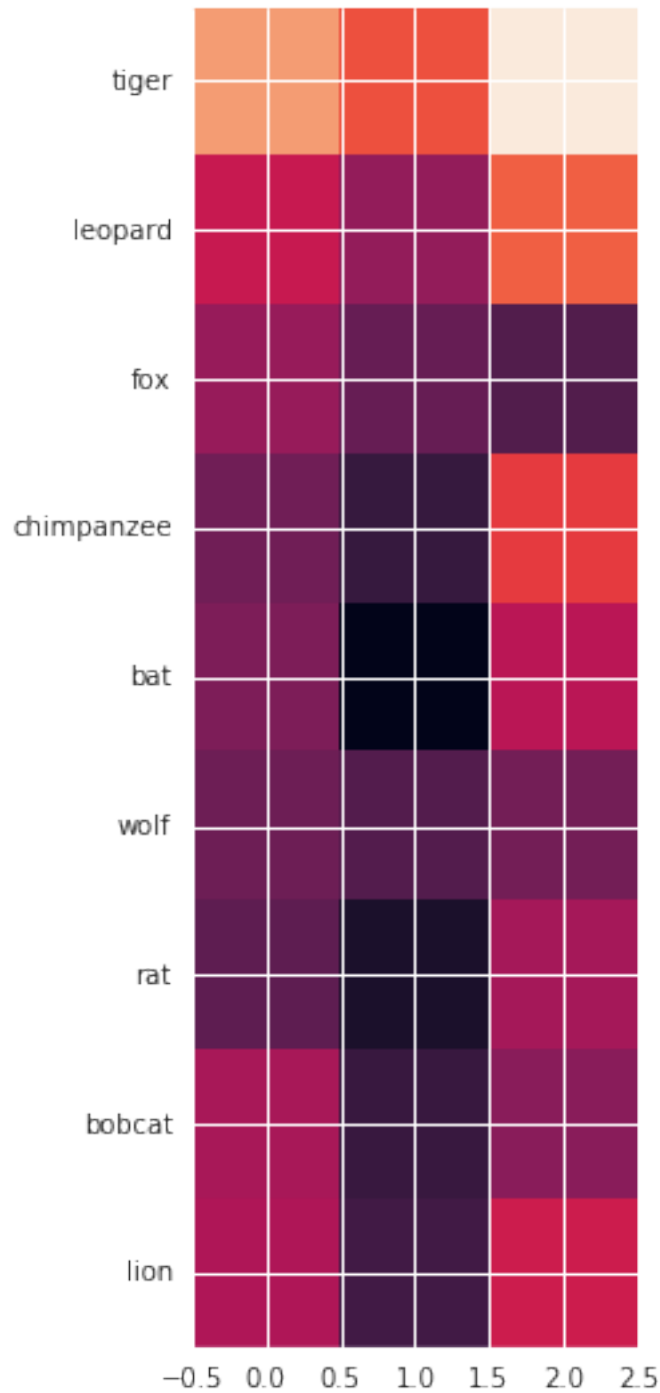










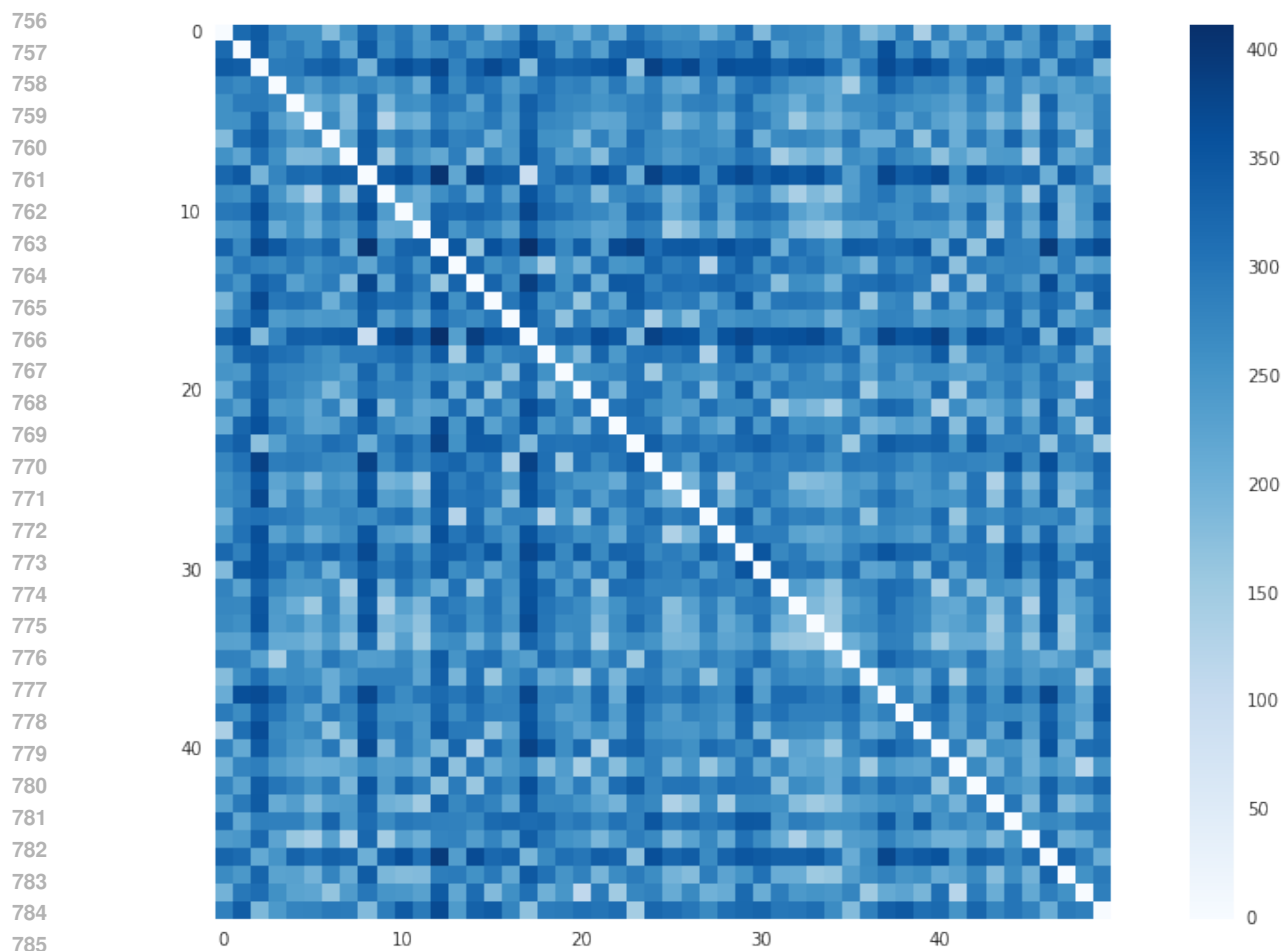


PCA is betting that there is a lower dimensional subspace that contains a decent amount of the data. But it seems that is not the case for our animals data set.

Try MDS.

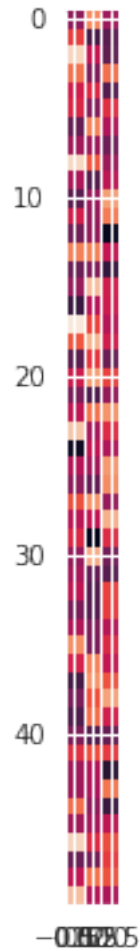
First get pairwise distances.

Try a dimensionality reduction with 3 components.



Do dimensionality reduction.

810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863

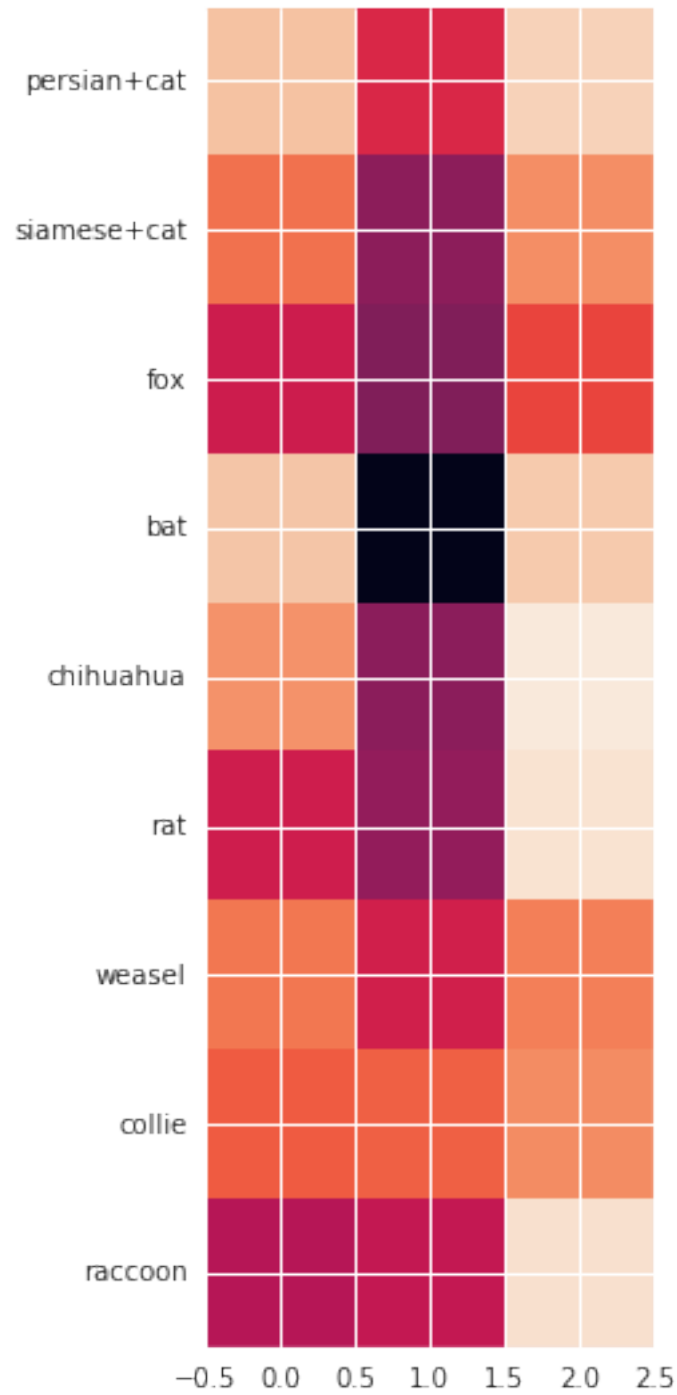


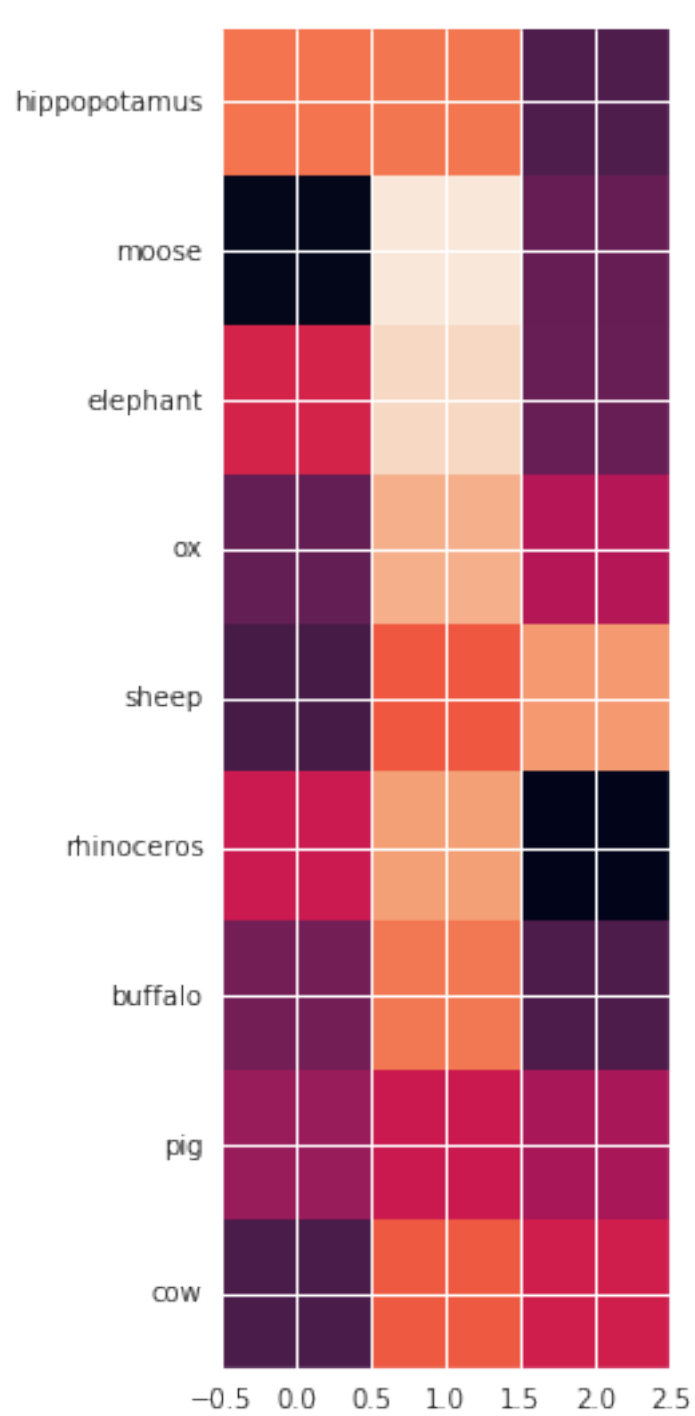
Use k-means on the MDS-reduced data. This is doing much better than PCA with 3 components. This seems to hint that the animal data is more affected by pairwise distances between points than by improper rotation.

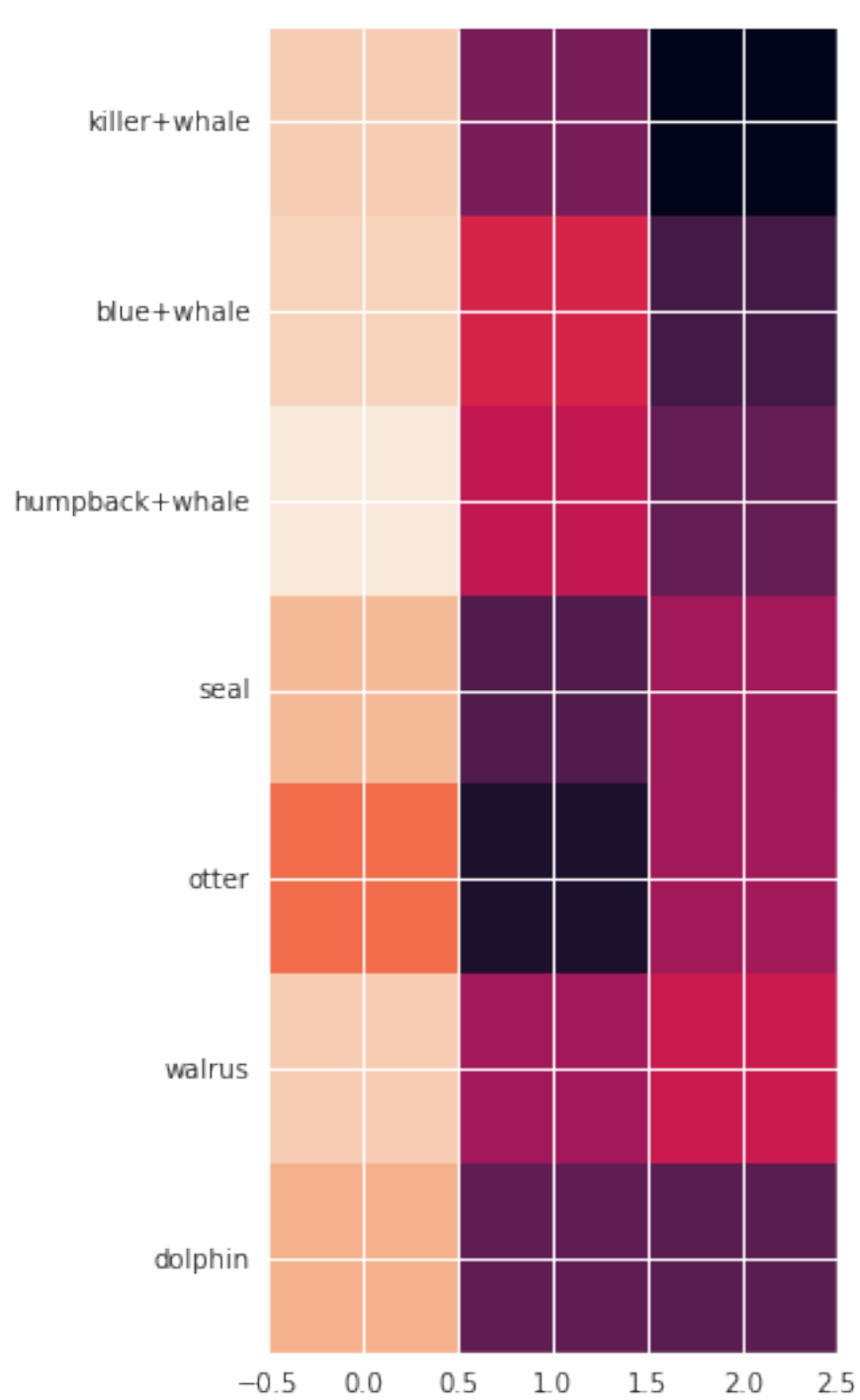
Still the MDS clustering is not perfect, and it seems strange that it groups spider monkey with horse and zebra.

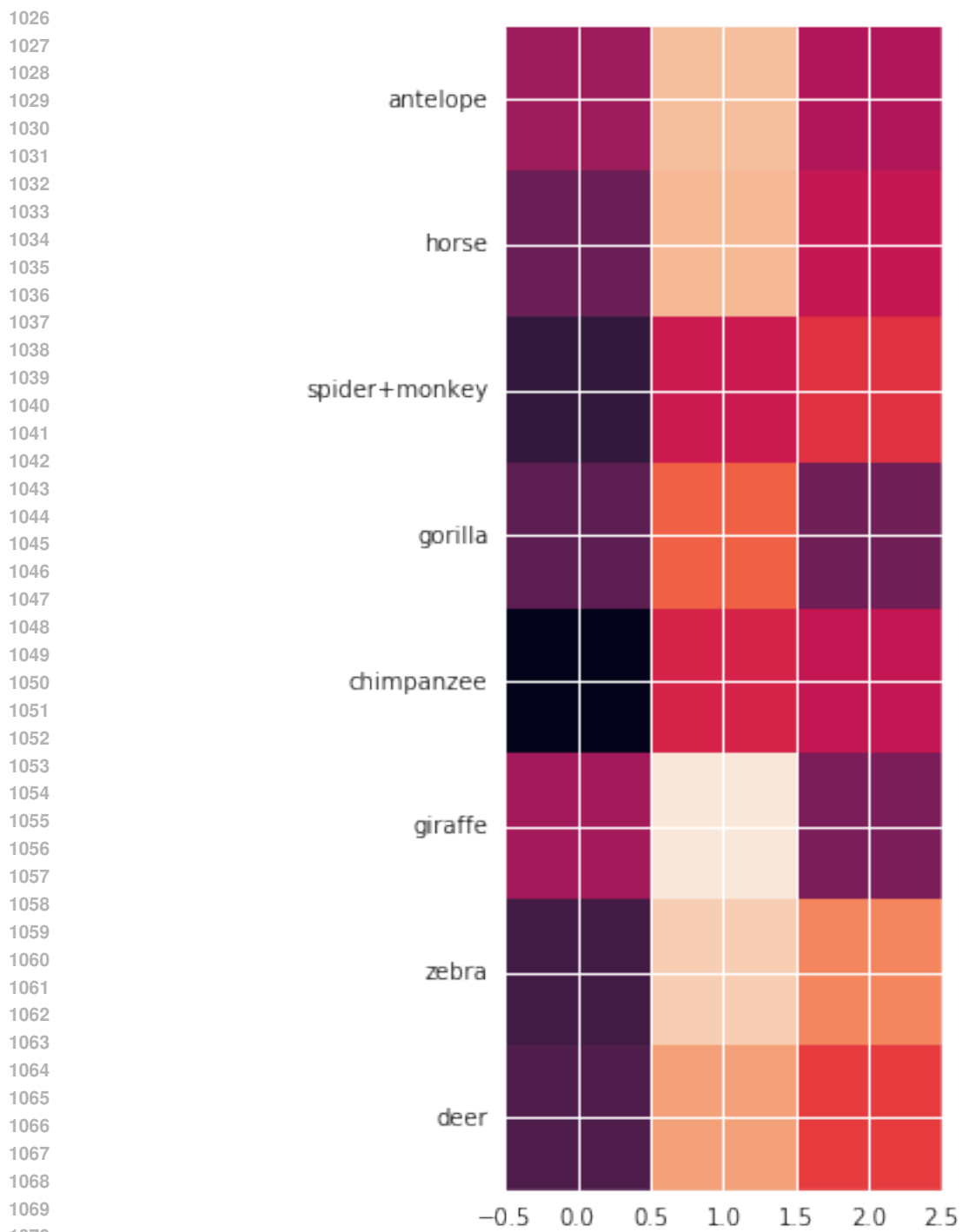


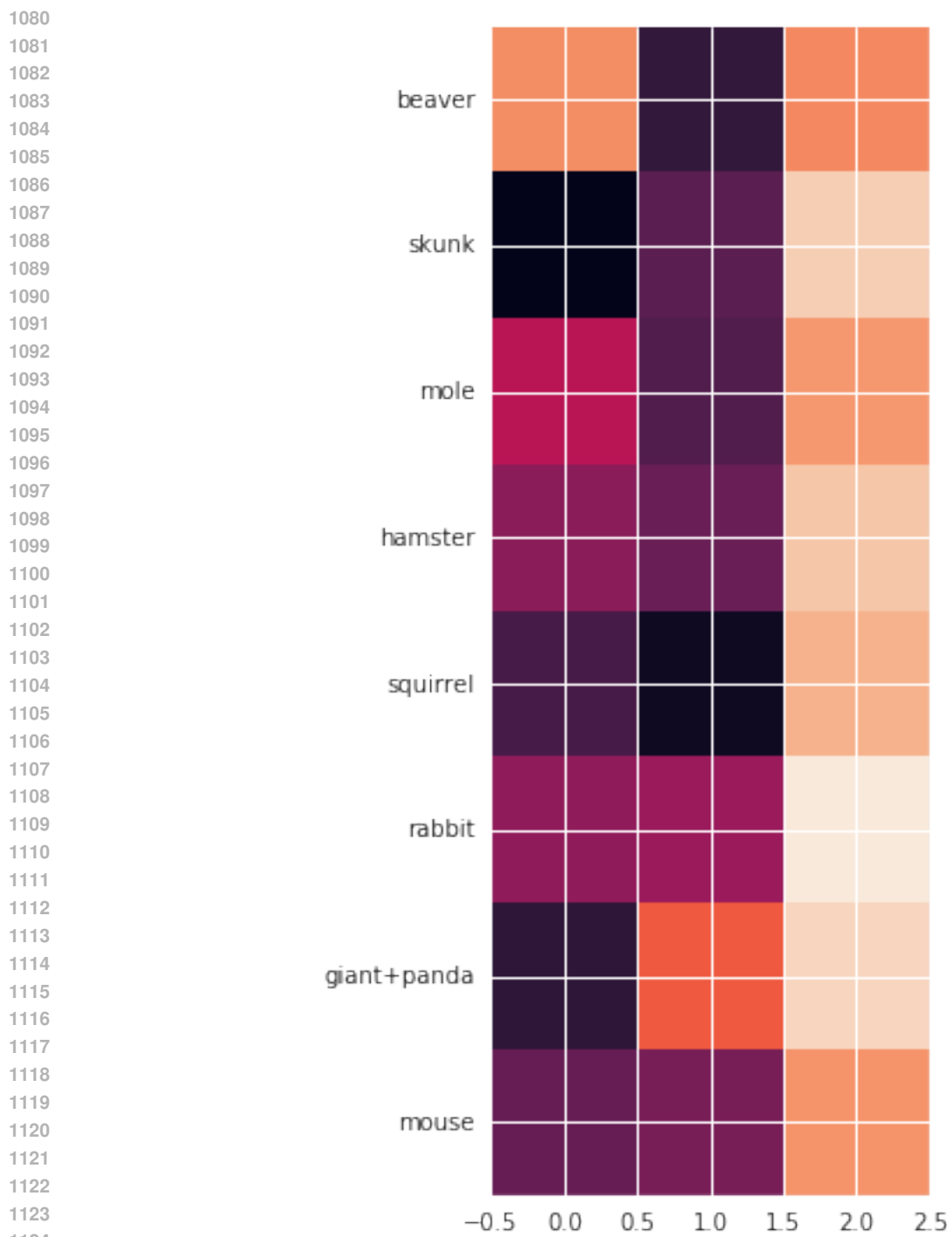
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917

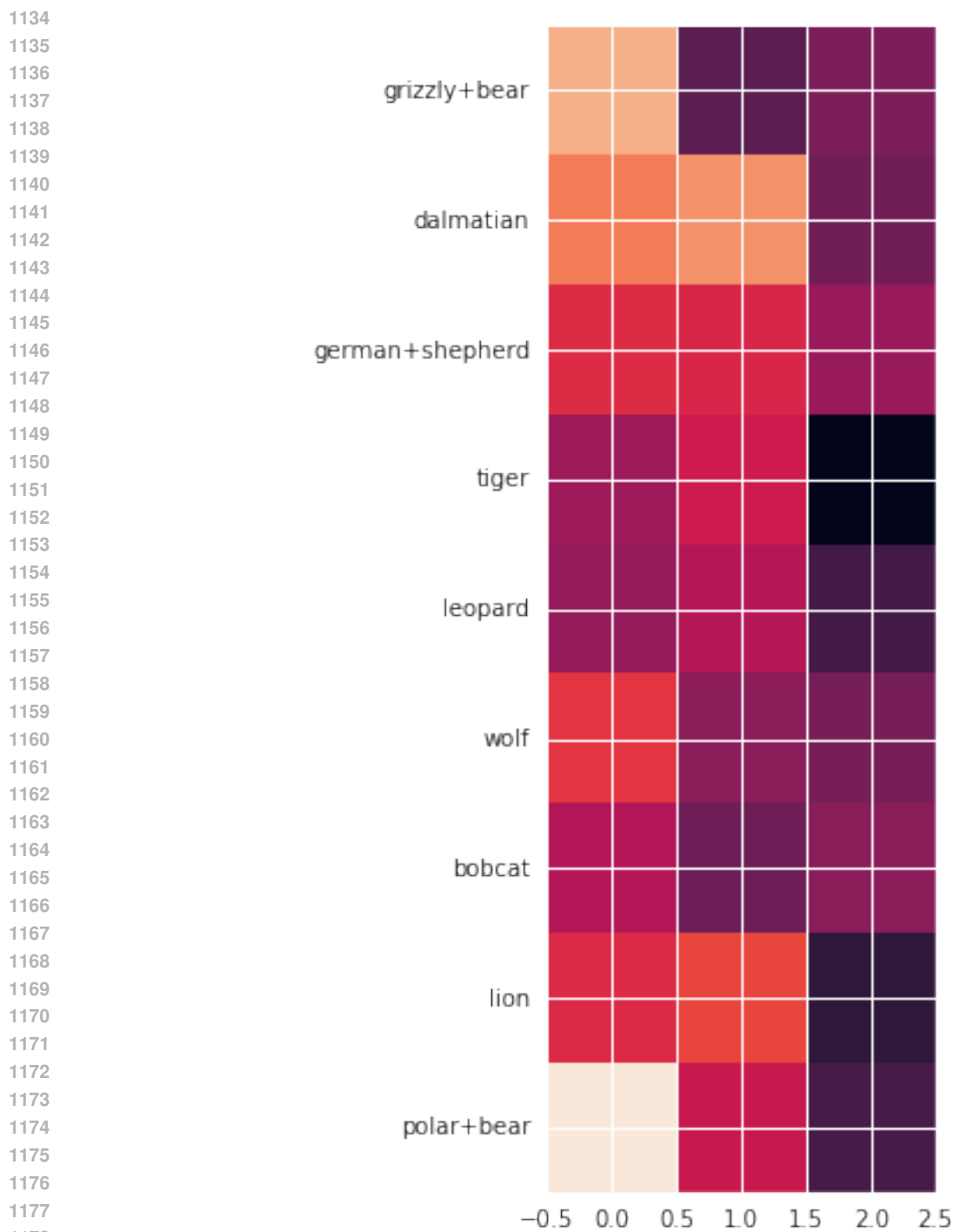










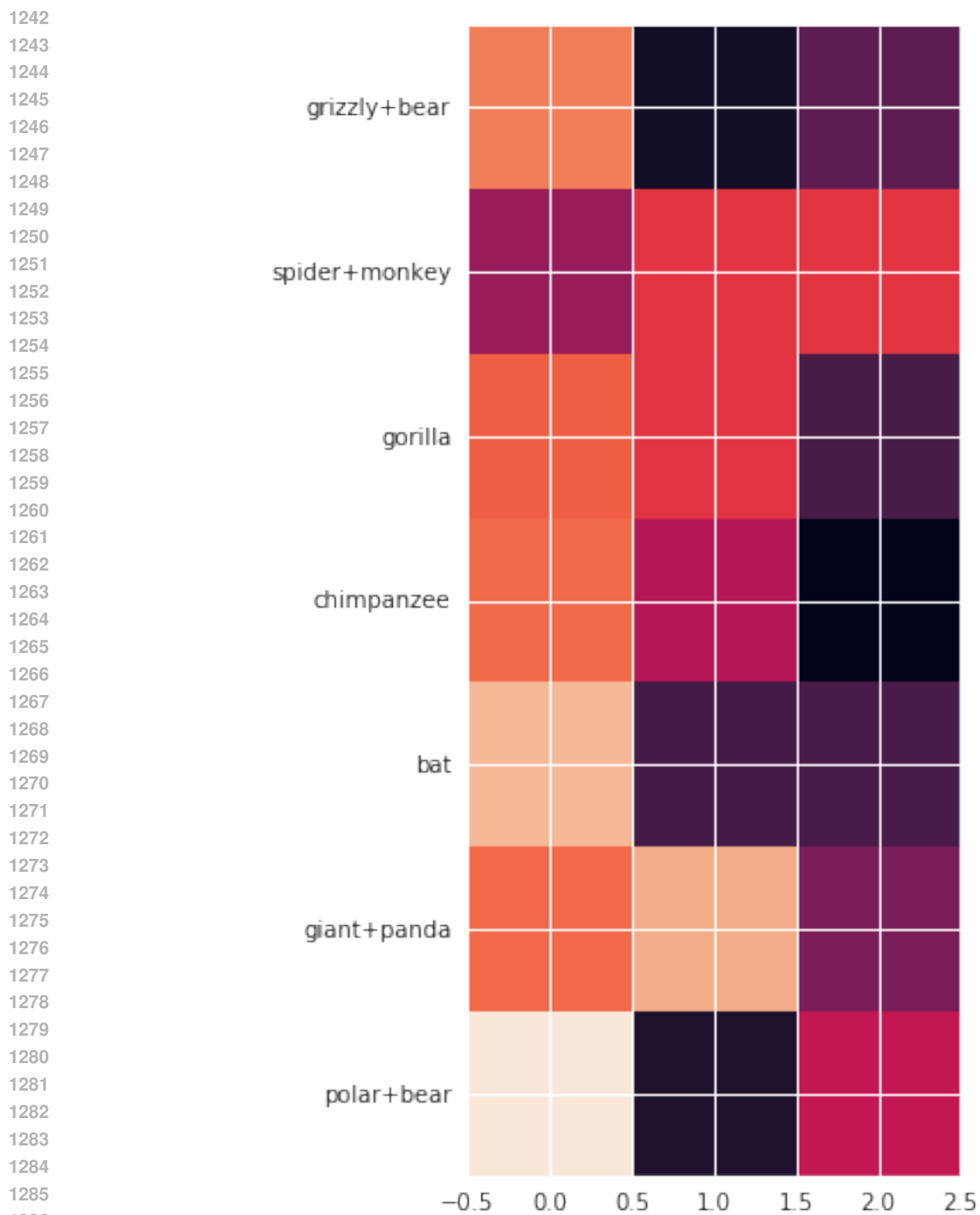


Since the previous attempts has been unsuccessful, it is reasonable to think that a linear embedding will not work given the shape of the animal data. Try a 3 components of locally linear embedding of the data.

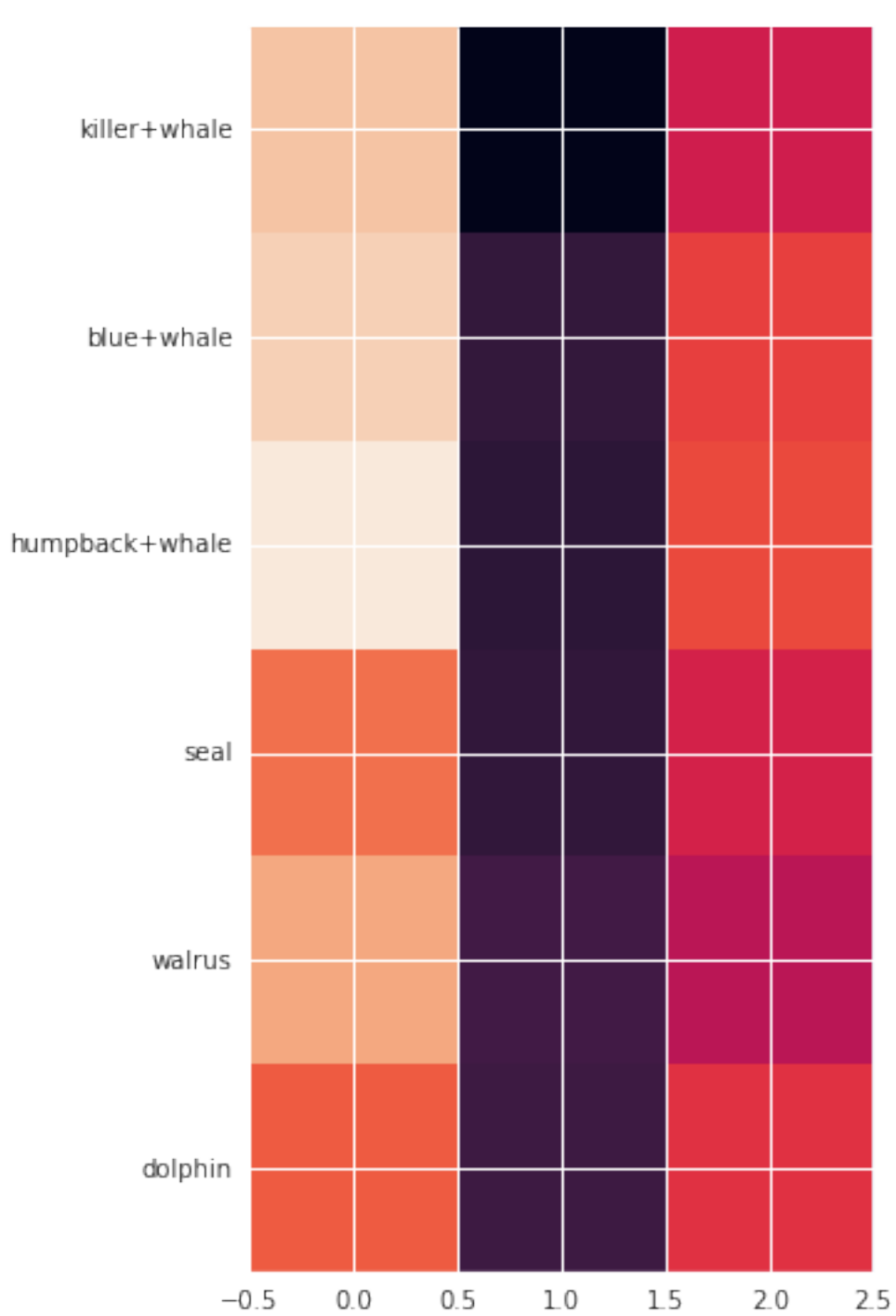
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241

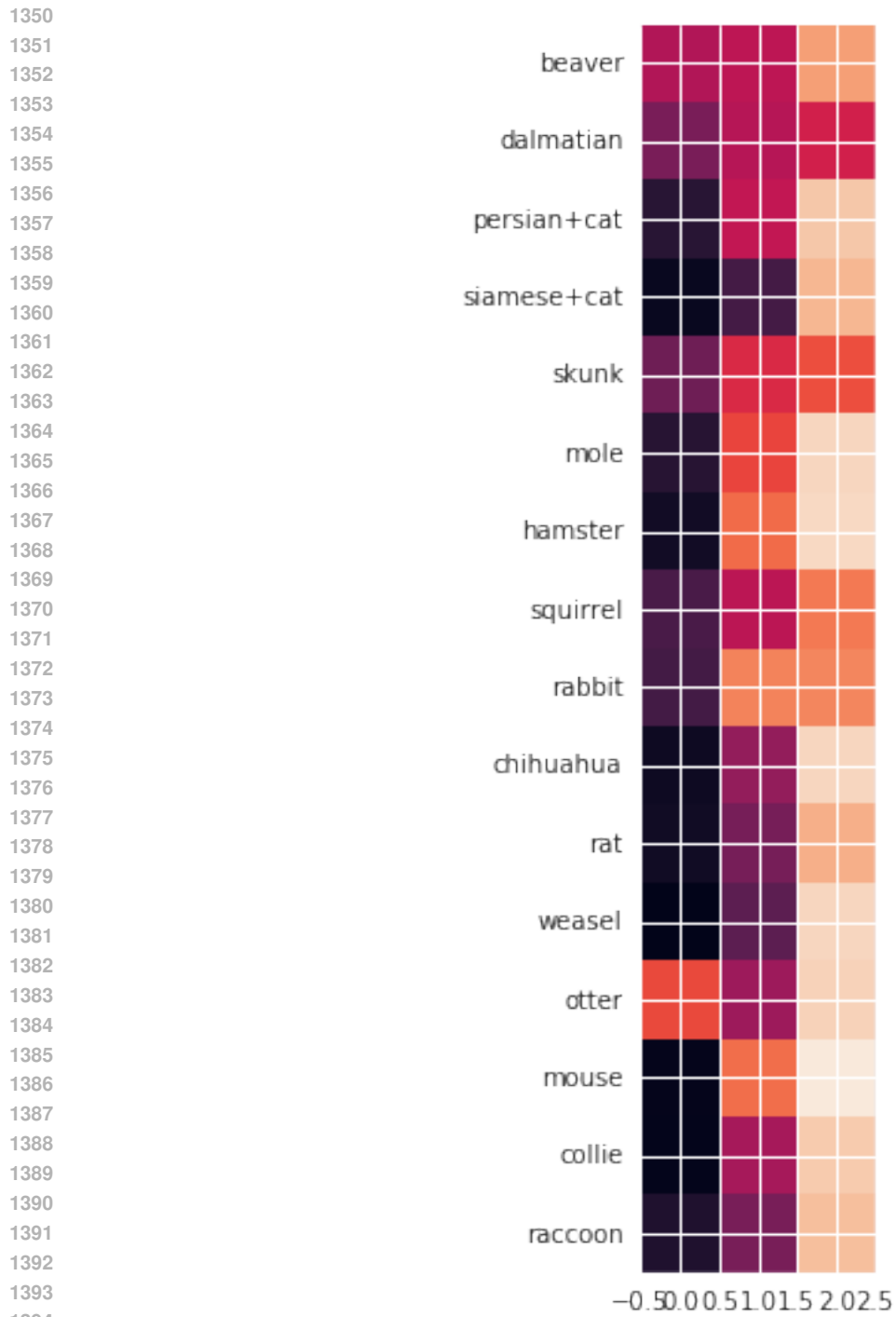


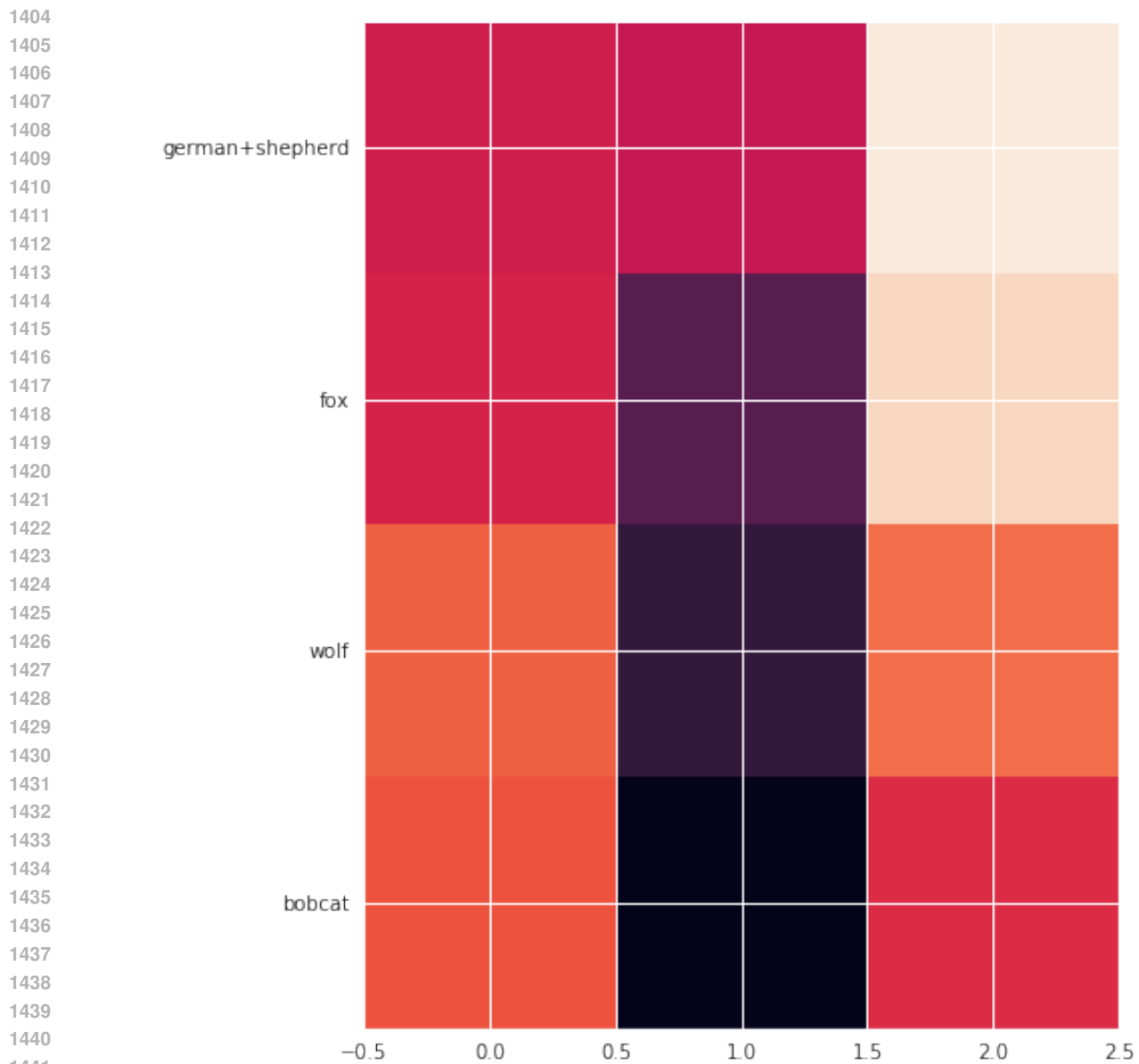
Try to cluster reduced representation.

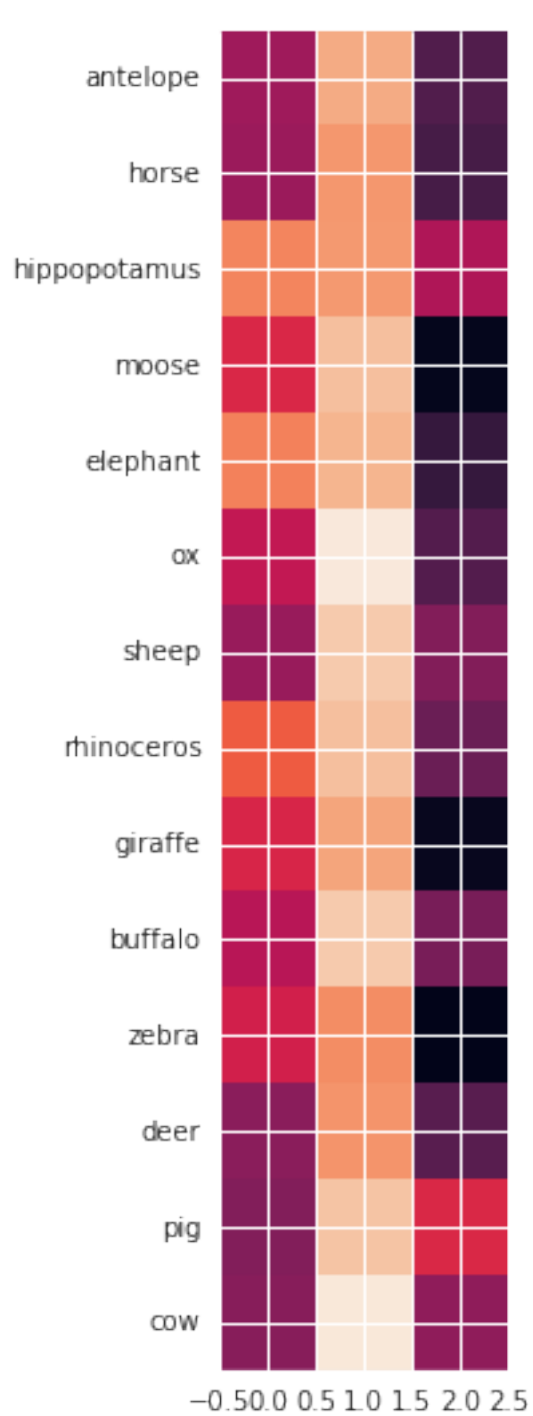


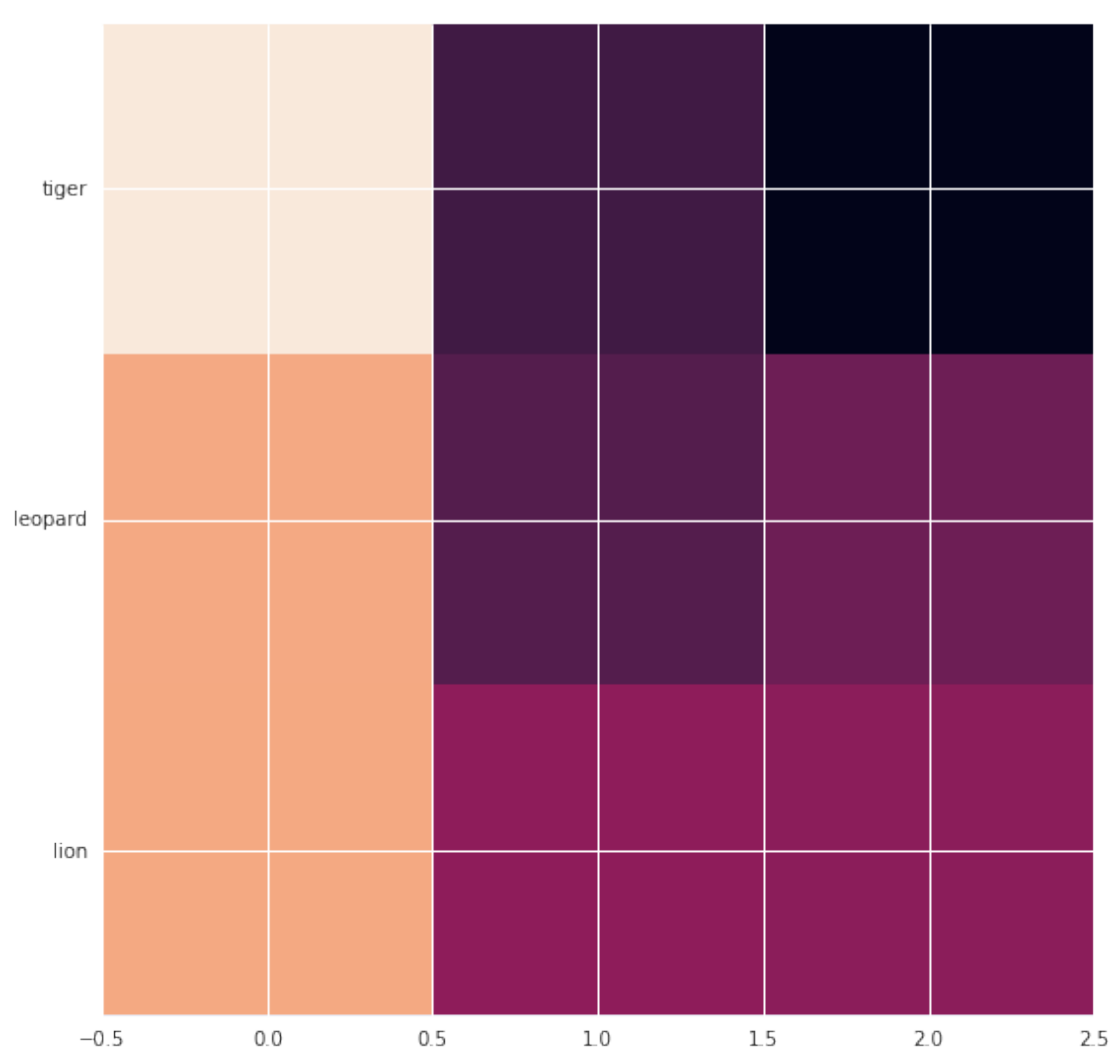








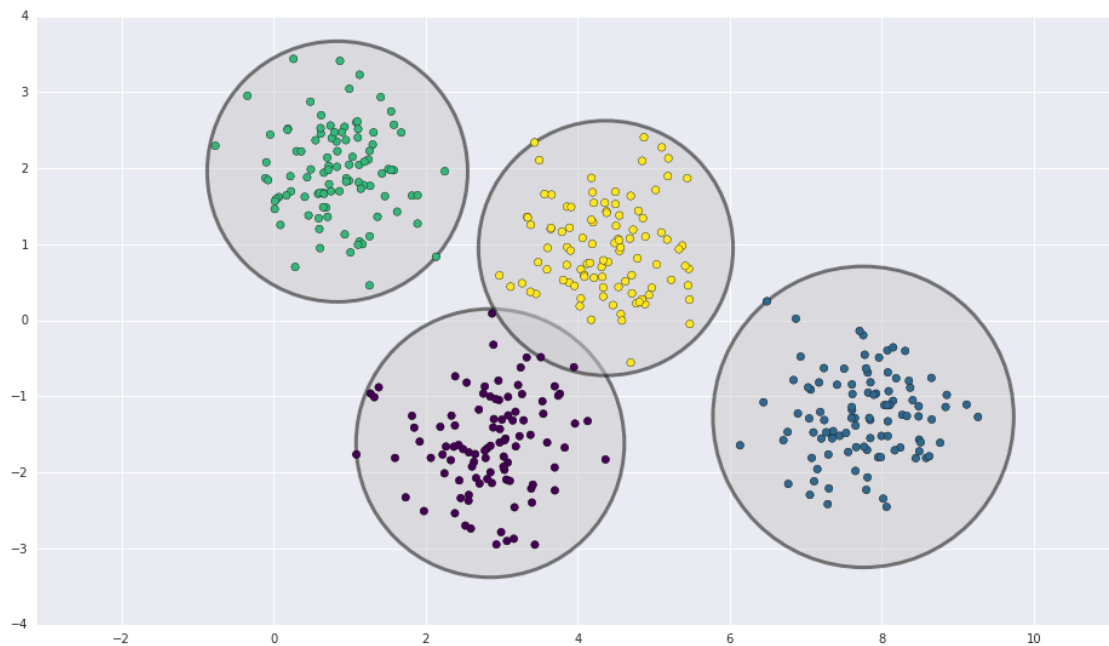




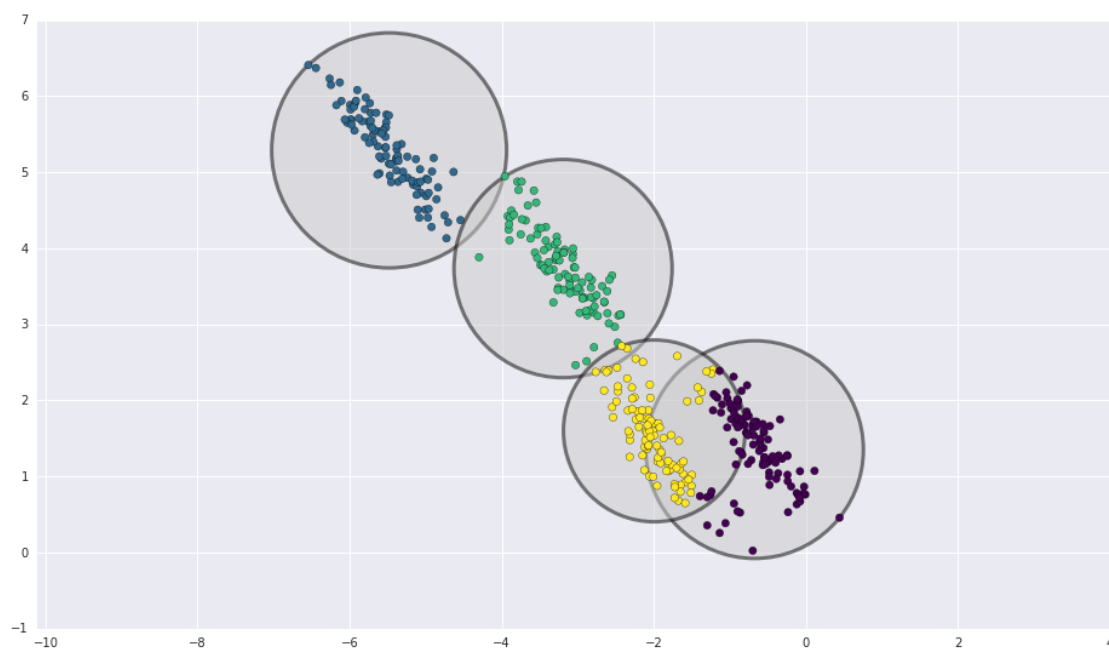
Using 3 components, this method did a much better job of clustering animals in a way that matches my intuition. But for a large number of components, like 8, the predictions one again become unintuitive.

Lastly, try GMM to compare with k-means clustering. k-means is nice and simple, but its non-probabilistic nature and its use of a single distance from a point to a cluster limits its possible use cases. It can lead to poor performance in many real-world situations. GMM can be viewed as an extension of k-means.

Let's view k-means again.



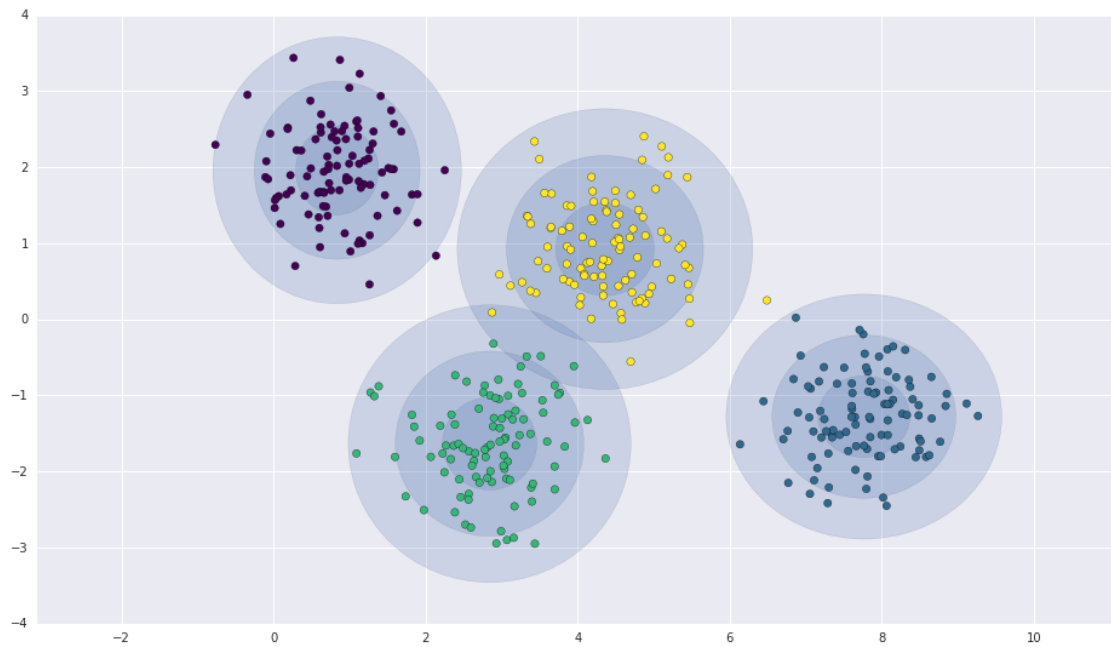
Plotted below is a case where k-means will struggle. It cannot change the shape of its clusters from circles. K-means does pretty well on the animal dataset anyway, because of the shape and high-dimensionality of the data.



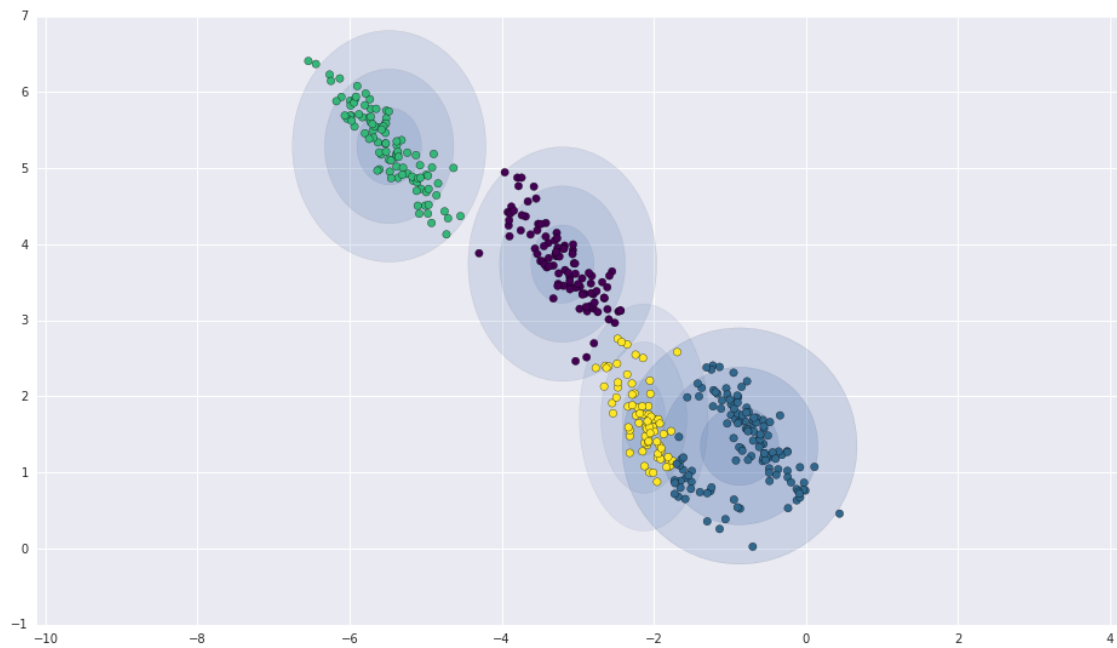
Test the Gaussian Mixture model for a simple case.



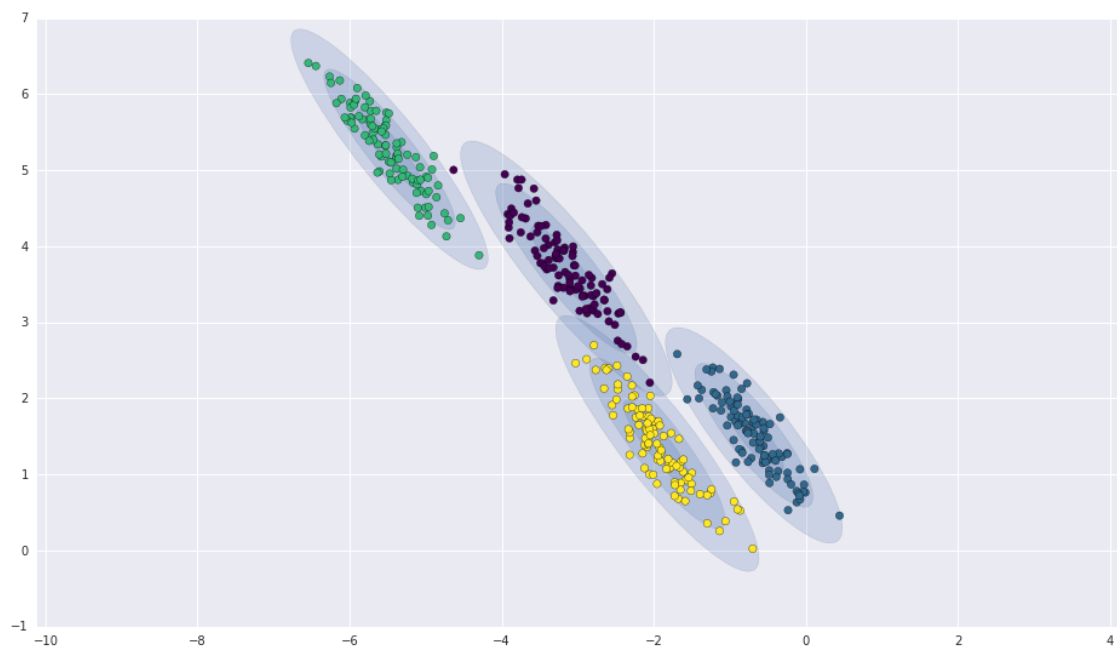
Plot the probability distributions around the clusters.



Do for problem with non-circular clusters.

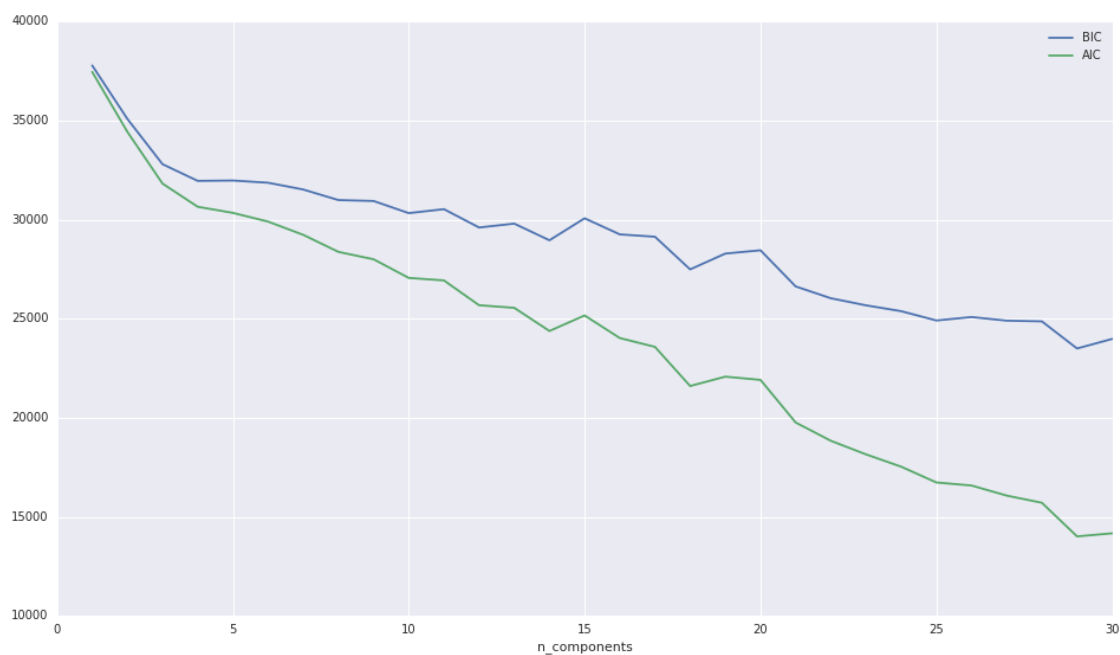
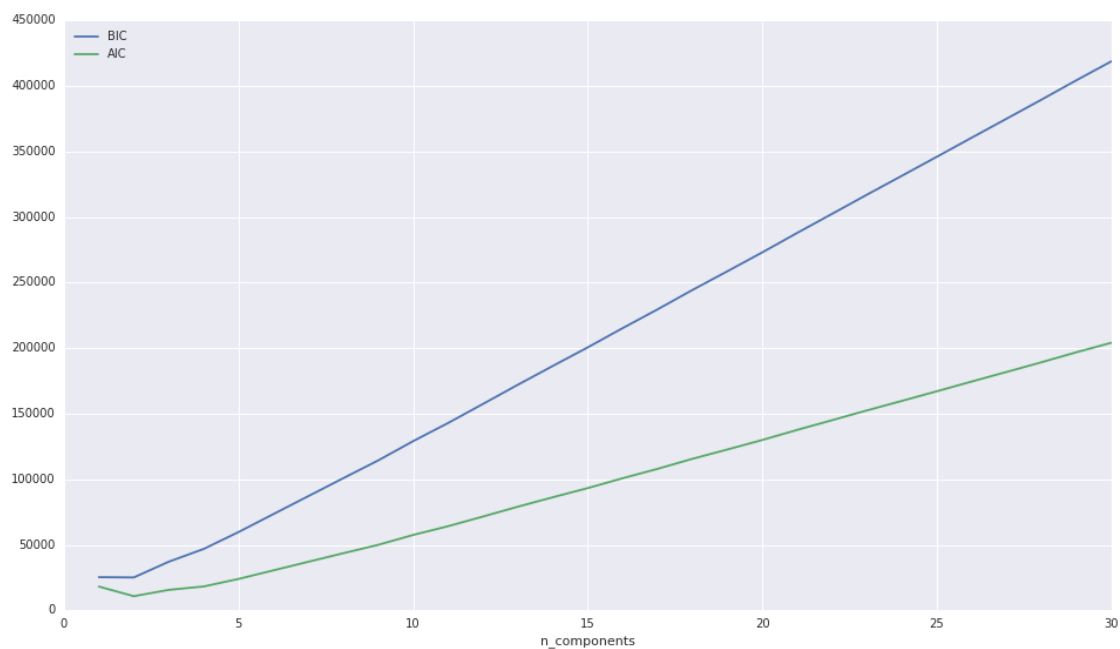


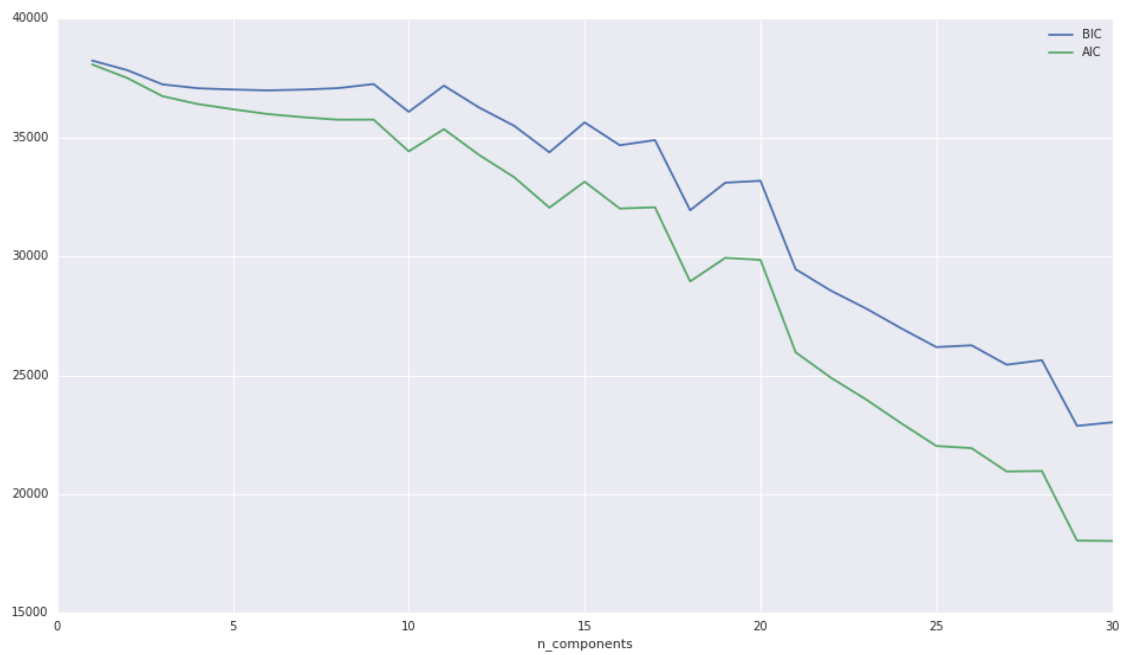
Do for problem with very non-circular clusters



AIC-Akaike information criterion and BIC-Bayesian information criterion are plotted below. These tests are inconclusive for the different types of GMM (full, diag, and spherical)

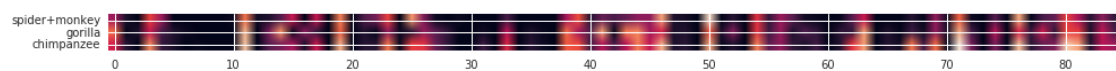
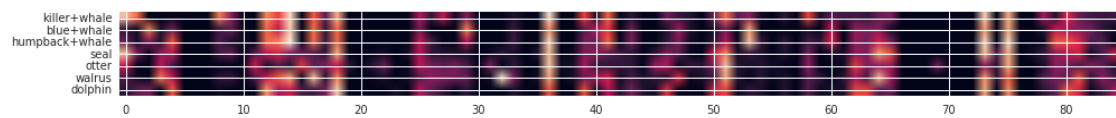
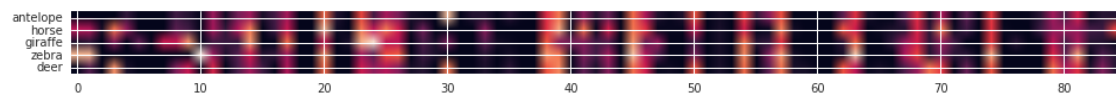
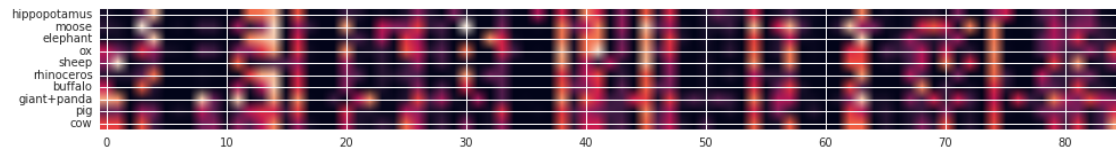


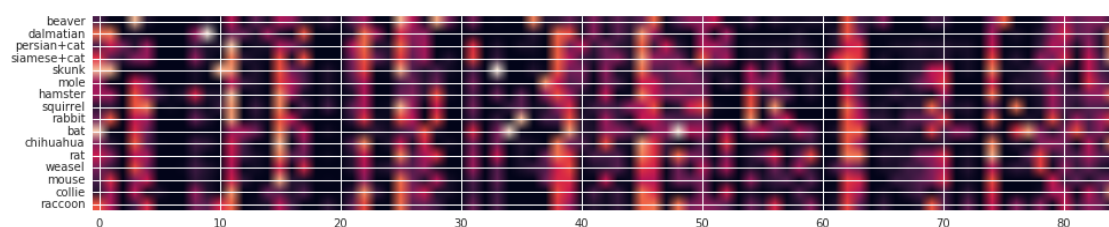




Try grouping animals based on 6 components of GMM. The results are pretty reasonable!

It would most likely be redundant to do something like PCA on top of this.





## References

- [1] D. N. Osherson, J. Stern, O. Wilkie, M. Stob, and E. E. Smith. "Default probability". Cognitive Science, 15(2), 1991.
- [2] C. Kemp, J. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda. "Learning systems of concepts with an infinite relational model". In AAAI, 2006.