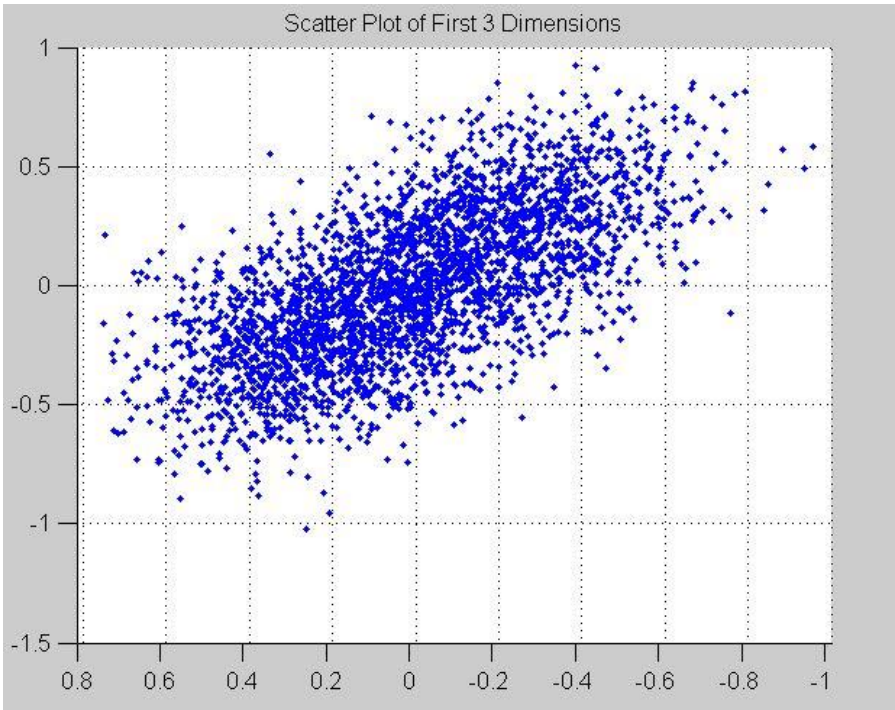
David Young
11/19/2013

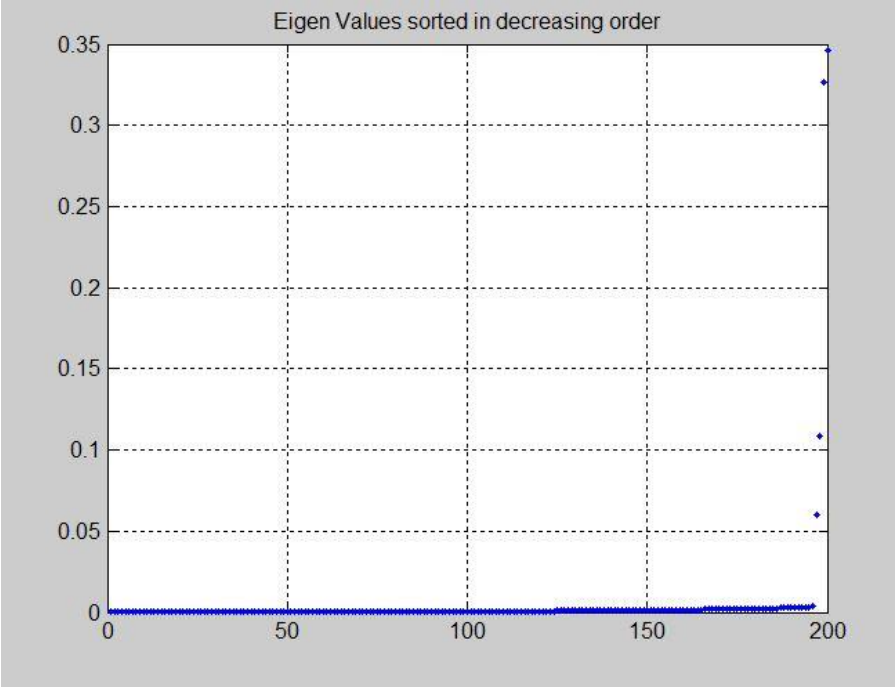
QP HW Nerve

1…

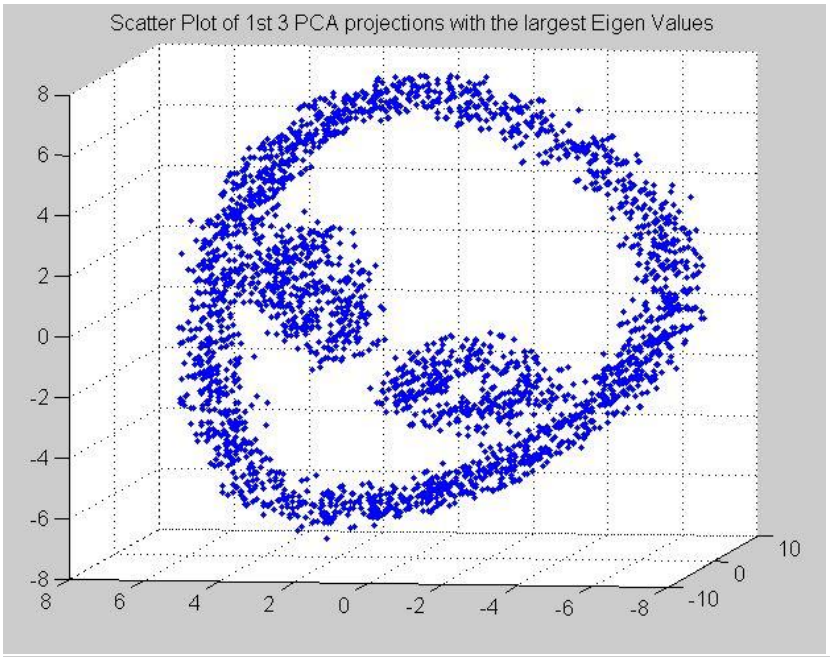
a. No discernable structure is present in the original 3 dimensions.

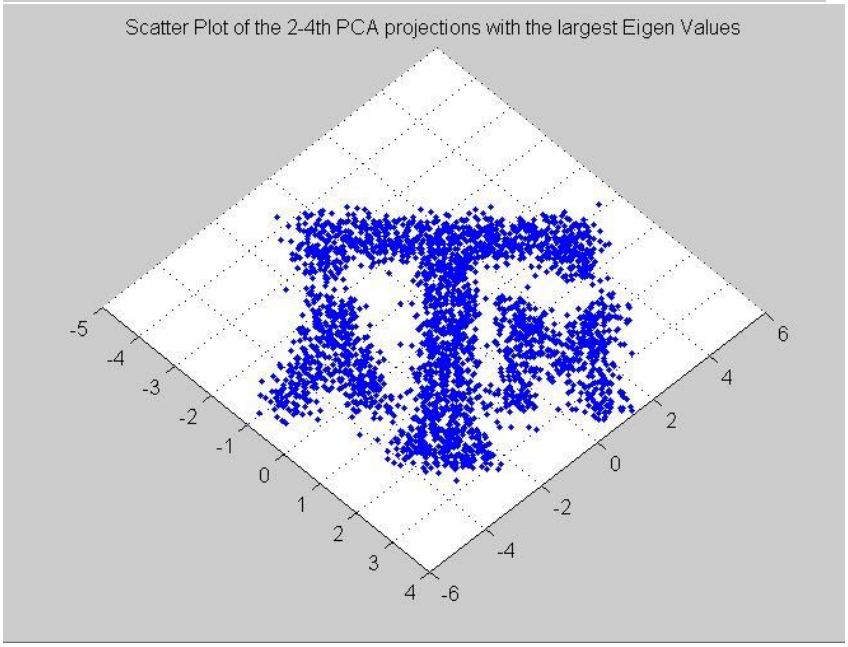


b. After PCA and plotting the eigen values in decreasing order, it is clear that only 3-4 values are responsible for the vast majority of variance in the data.



c. The first 3 PCA projections produce an alien face structure. Additionally a shift of 1 projection (ie: using 2,3, and 4th largest projections) creates the Texas A&M symbol.

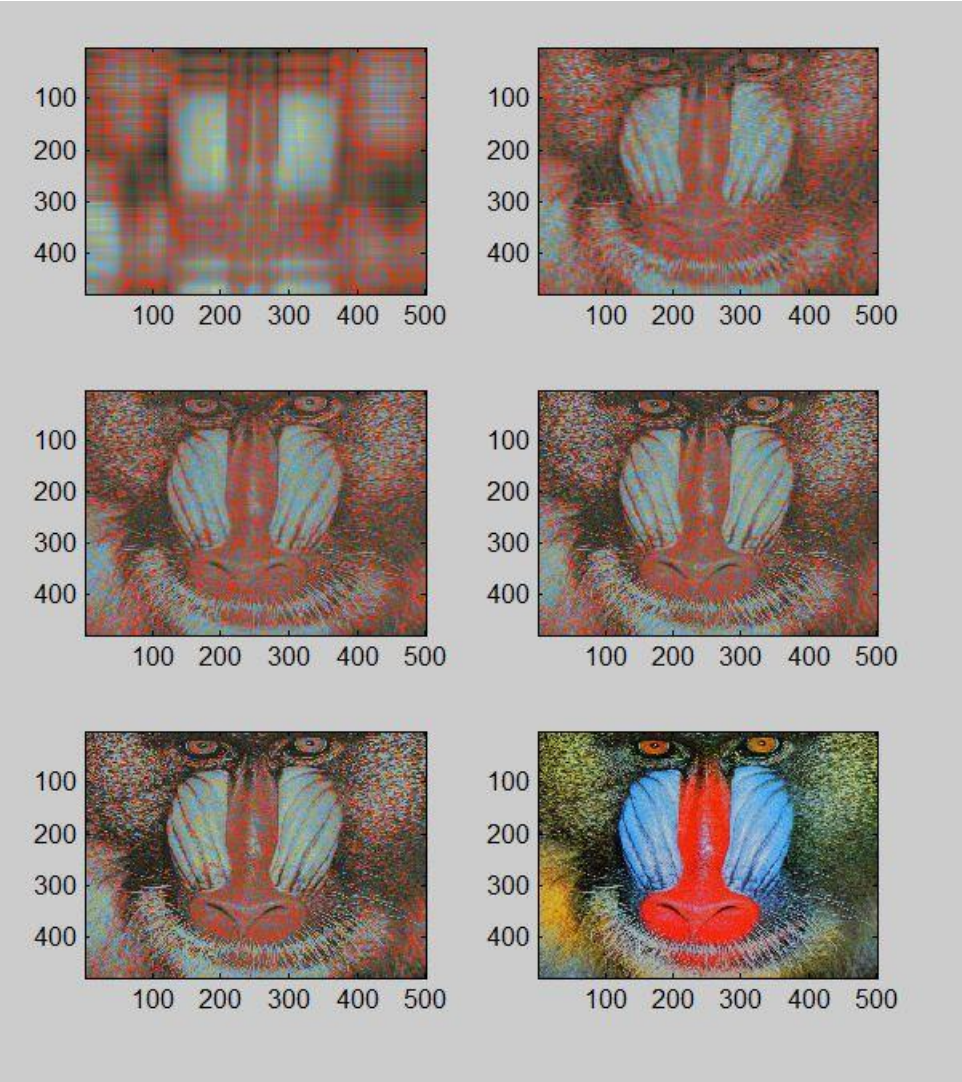




d. PCA analysis and the sorting of the eigen values allowed me to discern the projections with the greatest impact on variance. Normally we do this intuitively in fewer dimensions. Say for example, we wanted to take a picture of an iphone. To determine what angle to look at the iphone that captures the most data, we'd rotate the iphone until we found its long dimension, then hold that axis fixed and rotate the phone around that axis until we found the second longest dimension was shown. In doing this we create the 1st and 2nd principal components (as the two selected axis). Photographing from this angle would convey the most visual information

about the phone. In this problem, with so many dimensions, it would have been impossible to do this by eye… especially given the first graph where no discernable structure was present. PCA is incredibly powerful.

2. Low rank approximations for the given matrix are shown below. They are arranged in increasing order of # of basis vectors (3,30,100,200,300, and the full 480 respectively).



Using <1% of the total vectors for basis vectors meant the monkey was not discernable. Then, as the number of basis vectors increased, the quickly became clear (and only small features because discernable between each jump in basis vector number (from 30-300)). However the final jump from 300 to the full 480 vectors correlated with a drastic color difference that was extremely discernable.

It seems that the explained variance (the ratio of the total amount of variance in the projected data over that in the original) asymptotically approaches 1… after an extremely steep rise from 0. (ie… the first 50 or so vectors seem to get you vary close to an explained variance of 1).

X was originally a 3 dimensional array where each matrix corresponded to a color.

svd(X) computed the matrix singular value decomposition of X where U (square dimensions = rows of X) and W (square dimensions = columns of X) are unitary matrices such that X=U*V*W' , and V is a matrix sharing the same dimensions as X but with its diagonal elements in decreasing order (non-negative).

Discussants: Lauren Bedell, Matt Everett, Ramita Sarakune