The zipcode data are high dimensional, and hence even linear discriminant analysis suffers from high variance. Using the training and test data for the 3s, 5s and 8s, compare the following the four procedures with respect to the training and test misclassification error.

1. LDA on the original 256 dimensional space.

2. LDA on the leading 49 principal components of the features.

3. LDA when you filter the data as follows. Each non-overlapping 2 × 2 pixel block is replaced by its average. Hint: sample code given below.

   ```r
   h1<-kronecker(diag(8),cbind(c(0.5,0.5)))
   h2<-kronecker(h1,h1)
   lda(y ~ I(x%*%h2), data=zip.train)
   ## I(x%*%h2) averages adjacent 2 by 2 pixel blocks
   ```

4. Multiple linear logistic regression using the same filtered data as int previous question. Hint: the `multinom()` function in the R package `nnet` do multiple linear logistic regression.

   This procedure is equivalent to fitting a linear model with the full set of pixels, but the coefficients are constrained to be piecewise constant in 2 × 2 blocks. In other words, it is a linear expansion of Harr (piecewise constant) basis function over the image domain.