Kevin Gong
kg2445
HW02
Stat W4240
Section 2

Homework 2

Problem #1

Part a)

```
> colMeans(set1,na.rm=TRUE)
       x1        x2        x3        x4        x5 
 6.049104 -8.277221  4.665532  7.914270 62.138753 
> rowMeans(set1,na.rm=TRUE)
  [1]  -0.1277116  20.8162864  -8.8984358  25.5999204  -9.7472153  64.0626702  22.0392371  23.3914888  31.7598224 -13.8680290  43.8318898   6.5478369  14.1665143  16.1945993
 [15]  29.6357898  11.0316832  -2.5453007   8.6124471  33.8364419  24.9647839  34.8385372  34.1951748  25.8869897  -0.4545730   9.0418836  21.4051827   3.2291136  35.5748021
 [29]  21.1031545   6.5535668   3.7478608  18.9230712  -9.2447158   6.3811655  16.8358750   7.9628124  16.6264489  16.7027735 -34.4147885   0.4138282  12.6572899  35.4589880
 [43]  17.3456417  17.2383651   0.5124620 -24.7073649  17.1498949  52.3665782   9.6993053   0.3079195  15.6758568 -13.3093667   8.2062088  34.8247664  12.1909900  -3.1939531
 [57]  -5.4779341  10.7689107  36.2253846  19.5034554   8.9492321   4.4008921  14.3901288  14.7207124  27.9510161 -14.3617846  39.3331820  24.0356530  -6.7256757  -4.2948679
 [71]  27.1881673  47.2951022  19.1932996  23.5607379   7.6480638  18.1517706  16.9872267 -46.6660940   7.2223867  28.8378401   6.5043155  26.5206768  -2.4442159  15.3802055
 [85]  16.1739005  26.1705488  20.1409435  63.2646829   9.1977728  29.2026018   1.2105932  21.2145724  -8.4896595  19.0639963  20.9767512   3.5962333  22.3461063   0.7145014
 [99]   6.3080005  64.8829556
```

This matrix has 5 columns and 100 rows, and we can see that the column and row means vary in both magnitude and sign. The second column has a negative mean, while the fifth column's mean is significantly larger than those of the other columns. Among the rows, there are more positive row means than negative ones.

Part b)

```
> set1_emp
           x1         x2         x3         x4         x5
x1  1.0000000 -0.9328319  0.9903675  0.9229496  0.8391148
x2 -0.9328319  1.0000000 -0.9641852 -0.7226496 -0.9787417
x3  0.9903675 -0.9641852  1.0000000  0.8733612  0.8921875
x4  0.9229496 -0.7226496  0.8733612  1.0000000  0.5657457
x5  0.8391148 -0.9787417  0.8921875  0.5657457  1.0000000
```

The diagonal values of the covariance matrix are all 1, while the off-diagonal values are symmetric around the main diagonal. Thus, we know that the covariance matrix is symmetric and standardized.

Part c)

```
> eigen(set1_emp)
$values
[1] 4.486912e+00 5.087101e-01 4.368873e-03 8.952428e-06 5.481364e-07

$vectors
           [,1]        [,2]        [,3]       [,4]        [,5]
[1,]  0.4673470 -0.19431684 -0.42408896  0.7368413  0.14505575
[2,] -0.4596583 -0.31879336  0.25264423  0.2026734  0.76300680
[3,]  0.4711378 -0.04048796  0.85611177  0.1967856 -0.06883262
[4,]  0.4064079 -0.71331462 -0.12223569 -0.5420731  0.13126328
[5,]  0.4279751  0.59173874 -0.09192674 -0.2888262  0.61221849
```

We know that the left eignenvectors are transposes of the right eigenvectors of the transposed covariance matrix. In this case, our left and right eigenvectors are the same since our covariance matrix is symmetric.

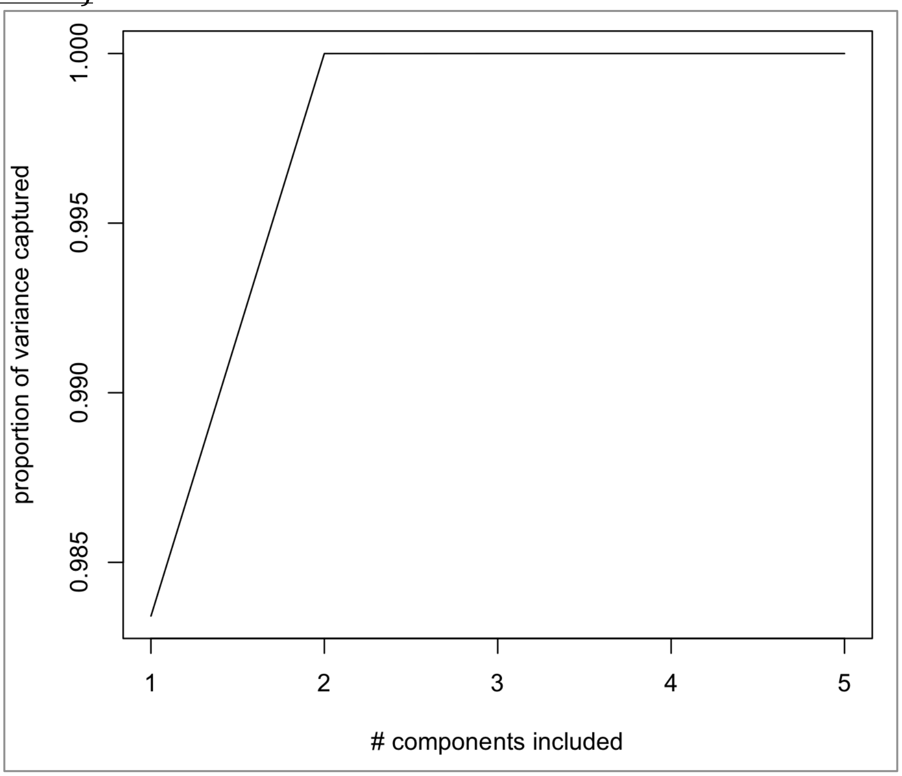
Part d)

```
> set1_pca$scores
       Comp.1       Comp.2        Comp.3        Comp.4        Comp.5
x1  0.9171061 -0.058803201 -2.038744e-03  6.118564e-06  1.221961e-10
x2 -3.2422362  0.007687044  9.159998e-05  4.785668e-07  9.545031e-12
x3  0.9365474  0.018887709  3.482194e-03  1.813234e-06  3.599721e-11
x4  0.6355725 -0.314161290 -5.228455e-04 -4.897621e-06 -9.769873e-11
x5  0.7530102  0.346389738 -1.012205e-03 -3.512744e-06 -7.003975e-11
> set1_pca$loadings

Loadings:
   Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
x1  0.461 -0.209 -0.424  0.735  0.154
x2 -0.469 -0.304  0.253  0.194  0.765
x3  0.469         0.856  0.198
x4  0.384 -0.726 -0.121 -0.544  0.125
x5  0.446  0.578        -0.296  0.609

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
SS loadings       1.0    1.0    1.0    1.0    1.0
Proportion Var    0.2    0.2    0.2    0.2    0.2
Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

Part e)



We should include 2 components, as additional components beyond the first two contribute very insignificantly toward the proportion of variance captured.

Part f)

```
> set2_pca$scores
         Comp.1     Comp.2       Comp.3        Comp.4        Comp.5
[1,] -0.5143948 -1.5502331  0.007730112  0.0007486009 -8.770762e-15
[2,]  1.9261814  0.7596842 -0.028036107  0.0012971909 -1.487699e-14
[3,]  1.8510330 -0.2785482 -0.027645240 -0.0015602335  1.726397e-14
[4,] -2.9793826  0.5349135 -0.045528501 -0.0002362324  1.887379e-15
[5,] -0.2834370  0.5341836  0.093479737 -0.0002493259  3.996803e-15
```

Part g)

Coordinates of the projection in the original space:

```
> Y
          x1          x2         x3         x4          x5
1  1.6021865  0.04605328  1.0635890 -0.5996748  -4.0227381
2 -1.2877658 -4.14192024 -1.1392189  1.7976776 -18.8652256
3 -3.2729844  6.43439645 -2.5496028 -7.6101484 -29.9975327
4  0.9713501  0.74796859  0.4469463 -3.5978859   0.6046244
5  0.8458003 -2.76899602  1.2421640 -0.2266484  21.2489825
```

Euclidean distance:

```
> sqrt(sum((Y-set1_centered))^2)
[1] 46.39402
```

Part h)

```
> as.vector(errors)
[1] 1.838789e-01 6.410107e-01 4.796689e-02 9.753955e-04 9.249128e-09
```

The errors are all positive, though very small, for the 5 new points. The direction of the error for the 5 new points is likely perpendicular to the direction of the projections in the original space since this generates the smallest possible error.

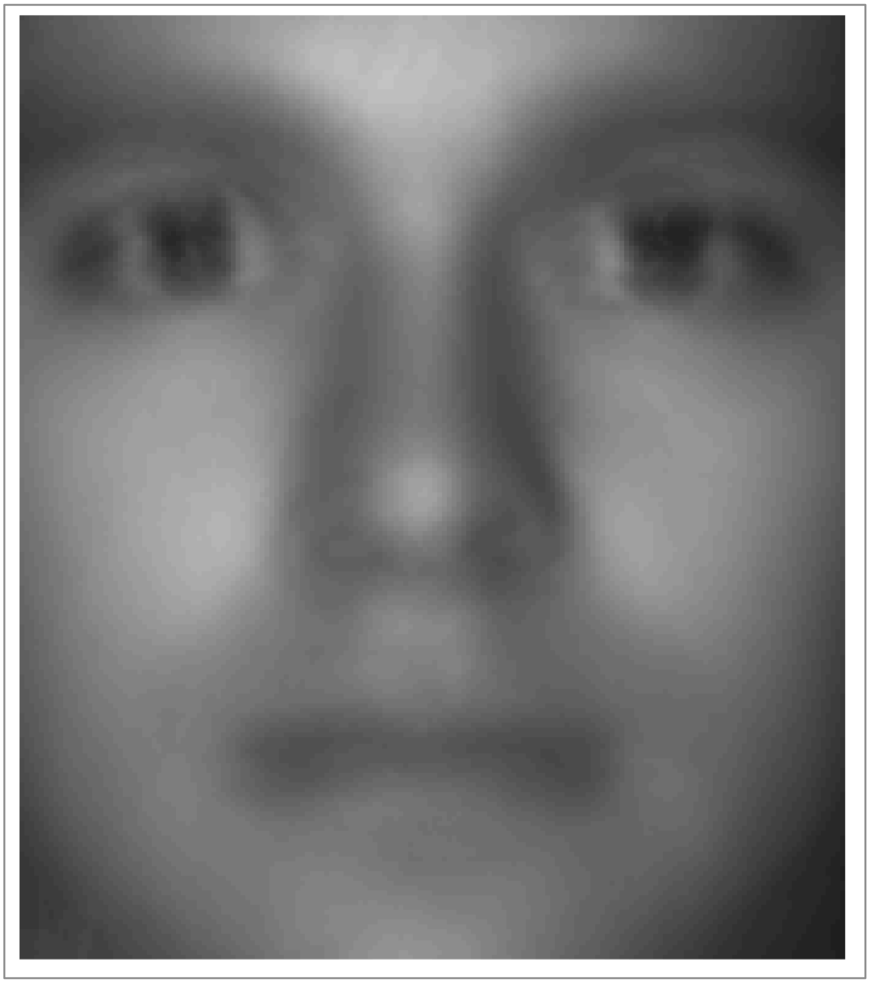
Problem #2

Part a)

```
> dim(faces_matrix)
[1]   152 32256
```

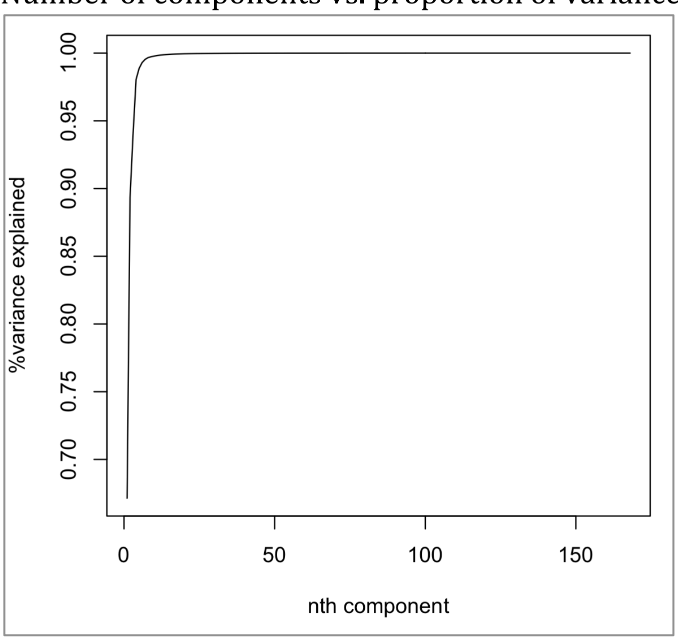
This matrix is 152 rows by 32,256 columns, where each row represents a different picture.

Part b)



Part c)

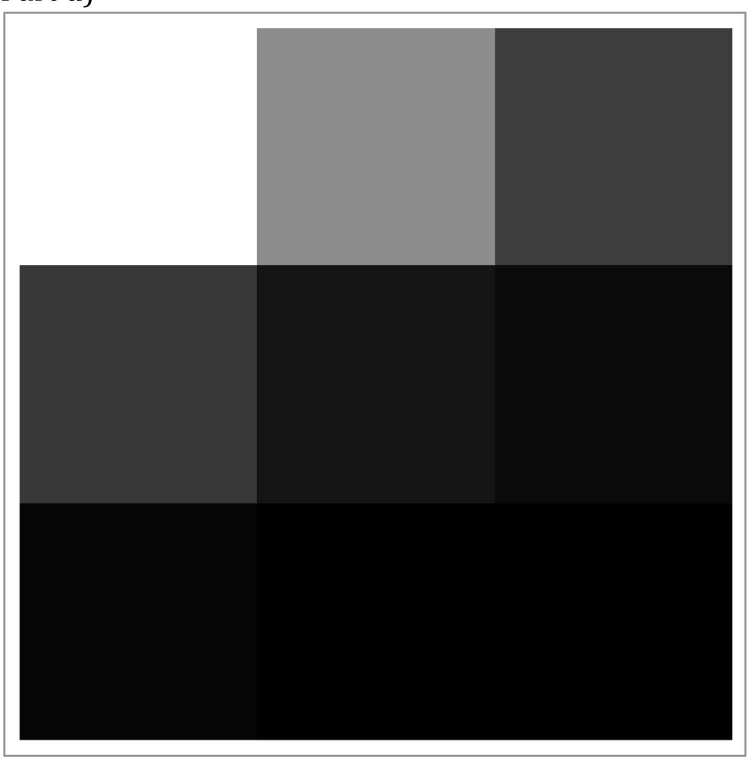
Number of components vs. proportion of variance explained:



Principal components:

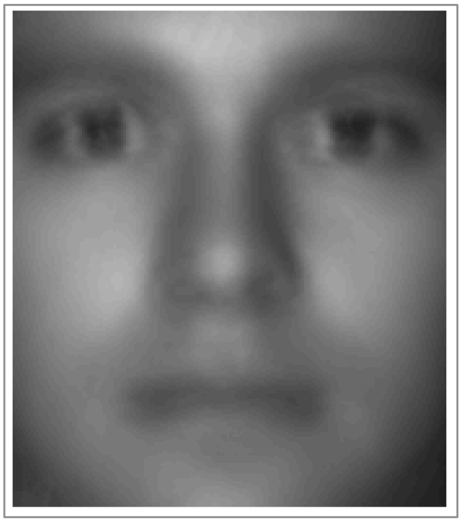
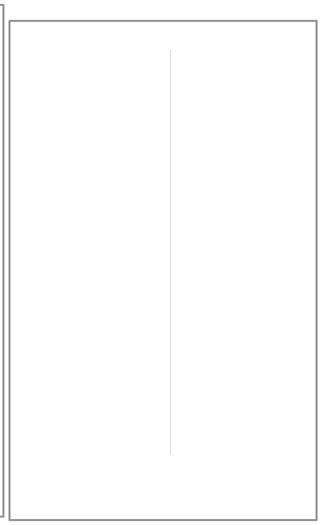
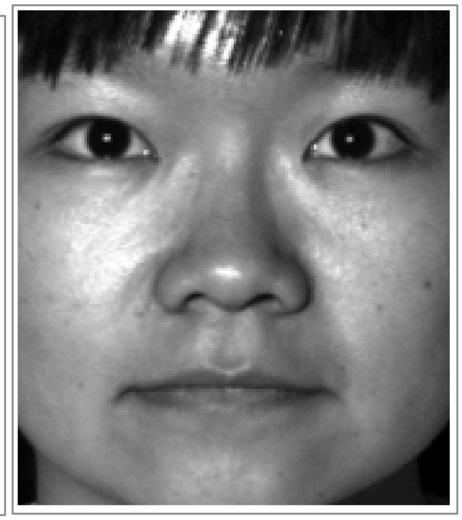
```
> summary(means_pca)
Importance of components:
                          PC1    PC2     PC3     PC4    PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16    PC17    PC18    PC19
Standard deviation     1.1039 0.6352 0.29058 0.27083 0.1213 0.08869 0.06581 0.04903 0.03369 0.02938 0.02659 0.02456 0.02047 0.01833 0.01585 0.01506 0.01329 0.01200 0.01116
Proportion of Variance 0.6714 0.2223 0.04652 0.04041 0.0081 0.00433 0.00239 0.00132 0.00063 0.00048 0.00039 0.00033 0.00023 0.00019 0.00014 0.00012 0.00010 0.00008 0.00007
Cumulative Proportion  0.6714 0.8936 0.94013 0.98054 0.9886 0.99298 0.99536 0.99669 0.99731 0.99779 0.99818 0.99851 0.99874 0.99892 0.99906 0.99919 0.99929 0.99936 0.99943
                          PC20    PC21    PC22     PC23     PC24     PC25     PC26    PC27     PC28     PC29    PC30     PC31     PC32     PC33     PC34     PC35     PC36
Standard deviation     0.01072 0.01023 0.00927 0.008325 0.007995 0.007051 0.006429 0.00622 0.005875 0.005736 0.00525 0.004931 0.004796 0.004665 0.004587 0.004437 0.004139
Proportion of Variance 0.00006 0.00006 0.00005 0.000040 0.000040 0.000030 0.000020 0.00002 0.000020 0.000020 0.00002 0.000010 0.000010 0.000010 0.000010 0.000010 0.000010
Cumulative Proportion  0.99950 0.99955 0.99960 0.999640 0.999680 0.999700 0.999730 0.99975 0.999770 0.999780 0.99980 0.999810 0.999820 0.999840 0.999850 0.999860 0.999870
                           PC37     PC38     PC39     PC40     PC41     PC42    PC43     PC44     PC45     PC46     PC47     PC48     PC49     PC50     PC51     PC52    PC53
Standard deviation     0.004082 0.003806 0.003662 0.003574 0.003204 0.003109 0.00303 0.002938 0.002913 0.002884 0.002715 0.002624 0.002549 0.002383 0.002341 0.002298 0.00218
Proportion of Variance 0.000010 0.000010 0.000010 0.000010 0.000010 0.000010 0.00001 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.00000
Cumulative Proportion  0.999880 0.999890 0.999890 0.999900 0.999910 0.999910 0.99992 0.999920 0.999930 0.999930 0.999930 0.999940 0.999940 0.999940 0.999950 0.999950 0.99995
                           PC54     PC55     PC56     PC57     PC58    PC59     PC60     PC61     PC62    PC63     PC64     PC65     PC66     PC67     PC68     PC69     PC70
Standard deviation     0.002152 0.002063 0.002028 0.001947 0.001922 0.00185 0.001844 0.001766 0.001697 0.00166 0.001607 0.001577 0.001508 0.001494 0.001412 0.001384 0.001342
Proportion of Variance 0.000000 0.000000 0.000000 0.000000 0.000000 0.00000 0.000000 0.000000 0.000000 0.00000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000
Cumulative Proportion  0.999960 0.999960 0.999960 0.999960 0.999960 0.99997 0.999970 0.999970 0.999970 0.99997 0.999970 0.999980 0.999980 0.999980 0.999980 0.999980 0.999980
                           PC71     PC72     PC73     PC74     PC75     PC76     PC77     PC78     PC79      PC80      PC81      PC82      PC83      PC84      PC85      PC86
Standard deviation     0.001312 0.001299 0.001286 0.001258 0.001209 0.001145 0.001108 0.001086 0.001055 0.0009972 0.0009898 0.0009629 0.0009522 0.0009308 0.0009174 0.0008839
Proportion of Variance 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
Cumulative Proportion  0.999980 0.999980 0.999980 0.999990 0.999990 0.999990 0.999990 0.999990 0.999990 0.9999900 0.9999900 0.9999900 0.9999900 0.9999900 0.9999900 0.9999900
                            PC87      PC88      PC89      PC90      PC91      PC92      PC93      PC94     PC95      PC96      PC97      PC98      PC99     PC100     PC101
Standard deviation     0.0008578 0.0008239 0.0008154 0.0007983 0.0007717 0.0007569 0.0007339 0.0006941 0.000669 0.0006639 0.0006531 0.0006334 0.0006244 0.0005997 0.0005917
Proportion of Variance 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
Cumulative Proportion  0.9999900 0.9999900 0.9999900 0.9999900 0.9999900 0.9999900 0.9999900 1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
                           PC102     PC103     PC104     PC105     PC106     PC107     PC108     PC109     PC110     PC111     PC112     PC113     PC114     PC115     PC116
Standard deviation     0.0005641 0.0005506 0.0005441 0.0005311 0.0005133 0.0004959 0.0004883 0.0004798 0.0004686 0.0004619 0.0004459 0.0004353 0.0004239 0.0004101 0.0003975
Proportion of Variance 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
Cumulative Proportion  1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
                           PC117     PC118     PC119     PC120     PC121    PC122     PC123     PC124     PC125     PC126    PC127     PC128     PC129     PC130     PC131
Standard deviation     0.0003962 0.0003827 0.0003791 0.0003677 0.0003623 0.000349 0.0003318 0.0003284 0.0003239 0.0003092 0.000304 0.0002988 0.0002957 0.0002815 0.0002751
Proportion of Variance 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.0000000
Cumulative Proportion  1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.0000000
                           PC132     PC133     PC134     PC135     PC136    PC137     PC138     PC139     PC140     PC141     PC142     PC143     PC144     PC145     PC146
Standard deviation     0.0002649 0.0002562 0.0002436 0.0002379 0.0002334 0.000232 0.0002172 0.0002093 0.0001982 0.0001925 0.0001922 0.0001838 0.0001715 0.0001671 0.0001651
Proportion of Variance 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000 0.0000000
Cumulative Proportion  1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000 1.0000000
                           PC147    PC148     PC149     PC150     PC151    PC152     PC153    PC154     PC155     PC156     PC157     PC158     PC159     PC160     PC161
Standard deviation     0.0001594 0.000153 0.0001516 0.0001494 0.0001369 0.000134 0.0001306 0.000124 0.0001141 0.0001114 0.0001009 9.791e-05 8.512e-05 8.312e-05 8.153e-05
Proportion of Variance 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.000000 0.0000000 0.000000 0.0000000 0.0000000 0.0000000 0.000e+00 0.000e+00 0.000e+00 0.000e+00
Cumulative Proportion  1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.000000 1.0000000 1.000000 1.0000000 1.0000000 1.0000000 1.000e+00 1.000e+00 1.000e+00 1.000e+00
                           PC162     PC163     PC164     PC165     PC166     PC167     PC168
Standard deviation     7.794e-05 7.374e-05 6.306e-05 5.782e-05 5.198e-05 4.563e-05 3.076e-05
Proportion of Variance 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00 0.000e+00
Cumulative Proportion  1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00 1.000e+00
```

Part d)



The eigenfaces take on the standard ranges of pixmap class objects of 0 to 1, where 1 represents white and 0 represents black. These eigenfaces represent eigenvectors of our covariance matrix of faces, or more generally the principal components of a distribution of faces. We generated our principal components in part c, and these form a basis for the images of all of the faces.

Part e)

Unfortunately, we failed to generate the 5x5 plot, as our code only produced an image of a vertical grey line (see middle image above). However, we understand that by adding eigenfaces in the following manner…

mean_face + score[1]*loading[1] + ... + score[m]*loading[m]

…we slowly transforms our mean face into the face of Subject 5. It would likely take us a fairly substantial number of eigenvectors (at least about 80~100) for us to be able to recognize the person.

Part f)

```
> dim(new_matrix)
[1]   148 32256
```

Removing subject 1 leaves us with a new matrix of faces of dimension 148 by 32256. We also successfully run prcomp() to obtain new principal components (see below).

```
> summary(new_pca)
Importance of components:
                           PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8     PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16    PC17    PC18    PC19
Standard deviation     15.5266 7.2094 6.15662 5.99079 5.57408 5.31375 4.94628 4.23696 3.65965 3.44747 3.25434 3.17436 3.02384 2.91959 2.78488 2.64503 2.58397 2.50496 2.35343
Proportion of Variance  0.3678 0.0793 0.05783 0.05476 0.04741 0.04308 0.03733 0.02739 0.02044 0.01813 0.01616 0.01537 0.01395 0.01301 0.01183 0.01067 0.01019 0.00957 0.00845
Cumulative Proportion   0.3678 0.4471 0.50497 0.55973 0.60714 0.65022 0.68755 0.71494 0.73538 0.75351 0.76967 0.78504 0.79900 0.81200 0.82383 0.83451 0.84470 0.85427 0.86272
                          PC20    PC21    PC22    PC23    PC24    PC25    PC26    PC27    PC28    PC29    PC30    PC31   PC32    PC33    PC34    PC35    PC36    PC37    PC38
Standard deviation     2.23564 2.17815 2.11883 2.08600 1.96834 1.93196 1.86006 1.83588 1.81298 1.76646 1.70521 1.64866 1.6394 1.58417 1.49621 1.46367 1.43663 1.38629 1.30114
Proportion of Variance 0.00763 0.00724 0.00685 0.00664 0.00591 0.00569 0.00528 0.00514 0.00502 0.00476 0.00444 0.00415 0.0041 0.00383 0.00342 0.00327 0.00315 0.00293 0.00258
Cumulative Proportion  0.87035 0.87759 0.88444 0.89108 0.89699 0.90268 0.90796 0.91310 0.91812 0.92288 0.92732 0.93146 0.9356 0.93939 0.94281 0.94608 0.94923 0.95216 0.95474
                          PC39    PC40   PC41    PC42    PC43    PC44    PC45    PC46    PC47    PC48    PC49    PC50    PC51    PC52    PC53    PC54    PC55    PC56    PC57
Standard deviation     1.26920 1.19783 1.1732 1.05725 1.03599 0.96921 0.92505 0.90463 0.90216 0.87494 0.86756 0.83367 0.82290 0.79682 0.77706 0.76083 0.73611 0.73041 0.72038
Proportion of Variance 0.00246 0.00219 0.0021 0.00171 0.00164 0.00143 0.00131 0.00125 0.00124 0.00117 0.00115 0.00106 0.00103 0.00097 0.00092 0.00088 0.00083 0.00081 0.00079
Cumulative Proportion  0.95720 0.95939 0.9615 0.96320 0.96483 0.96627 0.96757 0.96882 0.97006 0.97123 0.97238 0.97344 0.97447 0.97544 0.97636 0.97725 0.97807 0.97889 0.97968
                          PC58    PC59    PC60    PC61    PC62    PC63    PC64    PC65    PC66    PC67   PC68    PC69    PC70    PC71    PC72   PC73    PC74    PC75    PC76
Standard deviation     0.71667 0.69545 0.68247 0.66793 0.65500 0.65219 0.63927 0.61782 0.60503 0.58738 0.5732 0.56603 0.53497 0.53026 0.52260 0.5092 0.49680 0.49485 0.49008
Proportion of Variance 0.00078 0.00074 0.00071 0.00068 0.00065 0.00065 0.00062 0.00058 0.00056 0.00053 0.0005 0.00049 0.00044 0.00043 0.00042 0.0004 0.00038 0.00037 0.00037
Cumulative Proportion  0.98046 0.98120 0.98191 0.98259 0.98325 0.98390 0.98452 0.98510 0.98566 0.98619 0.9867 0.98718 0.98761 0.98804 0.98846 0.9889 0.98923 0.98961 0.98997
                          PC77    PC78    PC79    PC80    PC81    PC82   PC83    PC84    PC85    PC86    PC87    PC88    PC89    PC90    PC91    PC92    PC93    PC94   PC95
Standard deviation     0.48523 0.47223 0.46992 0.46132 0.45510 0.44976 0.4454 0.43751 0.42826 0.41397 0.41002 0.40595 0.39723 0.39520 0.38811 0.38240 0.37561 0.37085 0.3649
Proportion of Variance 0.00036 0.00034 0.00034 0.00032 0.00032 0.00031 0.0003 0.00029 0.00028 0.00026 0.00026 0.00025 0.00024 0.00024 0.00023 0.00022 0.00022 0.00021 0.0002
Cumulative Proportion  0.99033 0.99067 0.99101 0.99133 0.99165 0.99196 0.9923 0.99255 0.99283 0.99309 0.99335 0.99360 0.99384 0.99408 0.99431 0.99453 0.99475 0.99496 0.9952
                         PC96    PC97    PC98    PC99   PC100   PC101   PC102   PC103   PC104   PC105   PC106   PC107   PC108   PC109   PC110   PC111   PC112   PC113   PC114
Standard deviation     0.3602 0.35378 0.35102 0.34909 0.34080 0.33564 0.33374 0.32756 0.32285 0.31477 0.31069 0.30445 0.30025 0.29722 0.29136 0.28753 0.28450 0.27622 0.27216
Proportion of Variance 0.0002 0.00019 0.00019 0.00019 0.00018 0.00017 0.00017 0.00016 0.00016 0.00015 0.00015 0.00014 0.00014 0.00013 0.00013 0.00013 0.00012 0.00012 0.00011
Cumulative Proportion  0.9954 0.99555 0.99574 0.99592 0.99610 0.99627 0.99644 0.99661 0.99677 0.99692 0.99706 0.99721 0.99734 0.99748 0.99761 0.99773 0.99786 0.99797 0.99809
                         PC115  PC116  PC117   PC118   PC119   PC120   PC121   PC122   PC123   PC124   PC125   PC126   PC127   PC128   PC129   PC130   PC131   PC132   PC133
Standard deviation     0.26435 0.2588 0.2528 0.23946 0.23648 0.23472 0.23028 0.22461 0.22059 0.21242 0.20712 0.20087 0.19981 0.19643 0.19199 0.18712 0.18534 0.18175 0.17965
Proportion of Variance 0.00011 0.0001 0.0001 0.00009 0.00009 0.00008 0.00008 0.00008 0.00007 0.00007 0.00007 0.00006 0.00006 0.00006 0.00006 0.00005 0.00005 0.00005 0.00005
Cumulative Proportion  0.99819 0.9983 0.9984 0.99848 0.99857 0.99865 0.99873 0.99881 0.99888 0.99895 0.99902 0.99908 0.99914 0.99920 0.99925 0.99931 0.99936 0.99941 0.99946
                         PC134   PC135   PC136   PC137   PC138   PC139   PC140   PC141   PC142   PC143   PC144   PC145   PC146   PC147     PC148
Standard deviation     0.17872 0.17647 0.17424 0.16913 0.16850 0.16412 0.16049 0.15878 0.15595 0.15142 0.14694 0.14053 0.13782 0.13431 1.234e-14
Proportion of Variance 0.00005 0.00005 0.00005 0.00004 0.00004 0.00004 0.00004 0.00004 0.00004 0.00003 0.00003 0.00003 0.00003 0.00003 0.000e+00
Cumulative Proportion  0.99951 0.99956 0.99960 0.99965 0.99969 0.99973 0.99977 0.99981 0.99985 0.99988 0.99991 0.99994 0.99997 1.00000 1.000e+00
```

Unfortunately, we were unable to generate a reconstructed image, though we understand that we are using the eigenfaces of the remaining images to reconstruct subject 1's image. We know that the reconstructed image will look somewhat like the original image since we are eliminating data on the other subjects' faces from the mean face which thus allows us to approximate the image of subject 1.

Code

```r
#############################
# Kevin Gong
# STAT W4240
# Homework 2 , Problem 1
# 2/19/14
#

#############################

#################
# Setup
#################

# make sure R is in the proper working directory
# note that this will be a different path for every machine
setwd("~/Dropbox/SIPA/Data Mining/HW2")

# first include the relevant libraries
# note that a loading error might mean that you have to
# install the package into your R distribution.  From the
# command line, type install.packages("pixmap")
library(pdist)

#################
# Problem 1a
#################

set1 = read.csv(file="hw02_q1_p1.csv")
head(set1)
colMeans(set1,na.rm=TRUE)
rowMeans(set1,na.rm=TRUE)


#################
# Problem 1b
#################

set1_centered = scale(set1, center=TRUE)
set1_emp = cov(set1_centered)
set1_emp

#################
# Problem 1c
#################

eigen(set1_emp)


#################
# Problem 1d
#################

set1_pca = princomp(set1_emp)
set1_pca$scores
set1_pca$loadings


#################
# Problem 1e ??
#################
```

```
set1_pca$sdev^2/sum(set1_pca$sdev^2)

cumsum(set1_pca$sdev^2/sum(set1_pca$sdev^2))

plot(cumsum(set1_pca$sdev^2/sum(set1_pca$sdev^2)),type='l',xlab="# components included",
 ylab="proportion of variance captured")


#################
# Problem 1f
#################

set2 = read.csv(file="hw02_q1_p2.csv")
set2_centered = scale(set2, center=TRUE)
set2_pca = princomp(set2_centered)
set2_pca$scores


#################
# Problem 1g
#################

W = set1_pca$loadings[,1:2]
X = set2_centered[1:5,1:5]
Y = X*t(W)
#euclidean distances
sqrt(sum((Y-set1_centered))^2)
pdist(Y,set1_centered)
dist(Y,set1_centered)

#################
# Problem 1h
#################

errors = set2_pca$sdev-set1_pca$sdev
as.vector(errors)
















############################
# Kevin Gong
# STAT W4240
# Homework 2 , Problem 2
# 2/19/14
#

############################

#################
# Setup
#################
```

```r
# make sure R is in the proper working directory
# note that this will be a different path for every machine
setwd("~/Dropbox/SIPA/Data Mining/HW2")

# first include the relevant libraries
# note that a loading error might mean that you have to
# install the package into your R distribution.  From the
# command line, type install.packages("pixmap")
library(pixmap)

#################
# Problem 2a
#################

# the list of pictures
dir_list_1 = dir(path="CroppedYale/",all.files=FALSE)
view_list = c( 'P00A+000E+00' , 'P00A+005E+10' , 'P00A+005E-10','P00A+010E+00')

# preallocate an empty list
this_face_row = vector()
# initialize an empty matrix of faces data
faces_matrix = vector()
pic_list = c(
 01,02,03,04,05,06,07,08,09,10,11,12,13,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,
 34,35,36,37,38,39)


#generating a matrix of photos of the 4 desired views

for (i in 1:38) {

placeholder_matrix <- NULL

for (j in 1:4) {
 filename= sprintf("CroppedYale/%s/%s_%s.pgm",dir_list_1[i],dir_list_1[i],view_list[j])
print(filename)
face=read.pnm(file=filename)
pic_data <-getChannels(face)
pic_data_vec <- as.vector(pic_data)
placeholder_matrix=rbind(placeholder_matrix,pic_data_vec)
}
faces_matrix=rbind(faces_matrix,placeholder_matrix)
}

dim(faces_matrix)


#################
# Problem 2b
#################

#computing the mean of the faces
faces.matrix_v2=as.matrix(faces_matrix)
means_col=colMeans(faces.matrix_v2)
means_mat=as.matrix(means_col)

#reshape this matrix into 192x168
means_mat_ref=matrix(means_mat,192,168)

#double check dimensions to make sure it worked
dim(means_mat_ref)
```

```
#generate image of mean face
means_face=pixmapGrey(means_mat_ref)
plot(means_face)

#save image of mean face
filename = 'hw02_02b.png'
dev.copy(device=png, file=filename, height=600, width=800)
dev.off()

#################
# Problem 2c
#################

means_pca = prcomp(means_mat_ref)
summary(means_pca)
plot(summary(means_pca)$importance[3,], type="l", ylab="%variance explained", xlab="nth
 component")
means_pca$x

#################
# Problem 2d
#################

eigenfaces_matrix = matrix(summary(means_pca)$sdev[1:9], 3,3, byrow=TRUE)
eigenfaces_matrix
eigenfaces = pixmapGrey(eigenfaces_matrix)
plot(eigenfaces)


#################
# Problem 2e
#################

summary(means_pca)$loadings
means_pca

means_prin = princomp(means_mat_ref)
means_prin$loadings[1,]
means_prin$scores[,1]

dim(means_pca)


yaleB05 P00A+010E+00.pgm
face_05 = read.pnm(file = "CroppedYale/yaleB05/yaleB05_P00A+010E+00.pgm")
plot(face_05)


final_matrix=NULL
p=1

for(i in 1:5){
face_row=NULL
 for(j in 1:5){

 if(i==1 && j==1){
      z=as.vector(means_mat_ref)
      #z=rbind(z,z)
      }
 else {
      means_mat=means_mat_ref+((as.matrix(means_pca$x[,p]))%*%t(as.matrix(means_pca$x[p,])))
      mean_face_t=t(means_mat)
      p=p+1
```

```
      face_vec=as.vector(mean_face_t)
      z=cbind(z,face_vec)
      }
      }
 final_matrix=rbind(final,z)
 }

faces = pixmapGrey(final_matrix)
plot(faces)

as.matrix(means_pca$x[,1])
t(as.matrix(means_pca[1,]))
loadings(means_pca)

head(means_pca)

dim(means_pca$x)




#################
# Problem 2e
#################

new_matrix = faces_matrix[5:152,]
dim(new_matrix)
new_pca = prcomp(new_matrix)
summary(new_pca)


final_matrix2=NULL
p=1

for(i in 1:5){
face_row=NULL
 for(j in 1:5){

 if(i==1 && j==1){
      z=as.vector(means_mat_ref)
      #z=rbind(z,z)
      }
 else {
      means_mat=new_matrx+((as.matrix(new_pca$x[,p]))%*%t(as.matrix(new_pca$x[p,])))
      mean_face_t=t(means_mat)
      p=p+1
      face_vec=as.vector(mean_face_t)
      z=cbind(z,face_vec)
      }
      }
 final_matrix2=rbind(final,z)
 }
```