1.(a) **By viewing each image as a vector in a high dimensional space, per-  
form PCA on the full dataset. Hand in a plot the sorted eigenvalues of the  
sample covariance matrix.**

**Graphical user interface, text, application

Description automatically generated**

**How many principal components are needed to  
represent 95% of the total variation? How about 99%?**

-41 for 95% and 161 for 99%

**What is the per-centage reduction in dimension in each case?**

-1st case:97%

-2nd case:92%

1. **(b) Hand in a 4×5 array of subplots showing principal eigenvectors (‘eigen-  
   faces’) 0 through 19 as images, treating the sample mean as the zeroth order  
   principal eigenvector.**

Sample Subplots:

**Graphical user interface, application

Description automatically generated**

**2.(a) Please hand in a plot of the data  
points. Use two colors to show different clusters. Include the center of each  
cluster in you plot. Please use black and green colors for the centroids. In  
your report, please include the value of W(ˆc), where ˆcis the clustering map  
that you have found using k-means.**

-The value of WC is: 1991.8757826674994

**Scatter chart

Description automatically generated**

**2.(b) Does k-means work well for this dataset? If not, how should we  
modify the datapoints to cluster them successfully using k-means.**

**Chart

Description automatically generated**