1. The file *cars-fixed04.csv* on the Datasets web page contains the data on 388 cars from the 2004 model year, with 18 features (from http://www.amstat.org/publications/jse/datasets/04cars.txt, with incomplete records removed). The data is comma-separated; you may want to import it to Excel before working with it in Matlab. Eight features are binary indicators (yes/no); the other 11 features are numerical. PCA only works with numerical features. Use the Principal Components Analysis to analyze this dataset.

    a) What factors generate most variance in the data? What may the first two principal components stand for? How many principal components are essential for describing the variability in the data?

    b) What factors affect the price most (either Dealer price or Retail price)?

    c) PCA may help you see the correlations within the data. Which attributes are highly correlated? Are those correlations reflected in the r^2 values?

    d) Explore the dataset further. What other information can be inferred?