Exercise 7.1

Principal Component Analysis on Olympic data - Heptathlon

The data set consists of the results for the 25 athletes that finished the 1988 Olympic women's heptathlon competition in Seoul. The data are online in heptathlon.csv.

The results may also be be found in
http://todor66.com/olim/1988/Athletics/Women_Heptathlon.html

The following variables are available:

1. points - the number of points obtained
2. hurdles - results 100m hurdless
3. highjump - results high jump
4. shot - results shot
5. run200m - results 200m race
6. longjump - results long jump
7. javelin - results javelin
8. run800m - results 800m race

1. Read the data into R, make a scatter plot of the data, and calculate the correlation coefficients. What relationship do you see between the plots and the correlation coefficients?

2. Run a principal component analysis (PCA) on the variables, excluding the number of points obtained, and make the score plots. You can use the `PCA` function or the `princomp` function.

How many principal components should you use?

3. Which variables are explained by the two first principal components? What is the relation to the correlation matrix?