Iris data

The dataset "Fishers Iris data" is aclassical dataset used in many examples. The dataset consists of 150 observations/objects, 50 Iris Setosa, 50 Iris versicolor and 50 Iris virginica. The flowers on these plant have been measured (mm). The measured variables are sepal length, sepal width, petal length and petalwidth (4 variables). The original hypothesis was that versicolor was a hybrid of the two other species.

1. Read in the data iris.txt.
2. Make descriptive plots.
3. Carry out a PCA on scaled data.
   a. How many principal components are need for an adequate description of the variation in data?
   b. Describe how you chose the number of principal components.
4. How can the principal components be interpreted?
5. Can the principal components be used to distinguish between the sorts?
6. Investigate if the model fits well: look at residuals and influential observations (leverage)