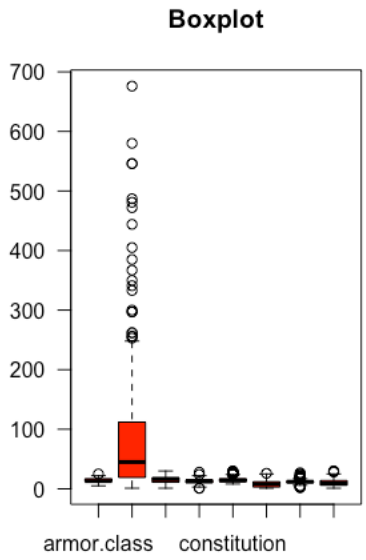
APPLIED STATISTICS EXAM

DATE: 12/07/2022
STUDENT: FILIPPO CIPRIANI
PERSONA CODE: 10956877

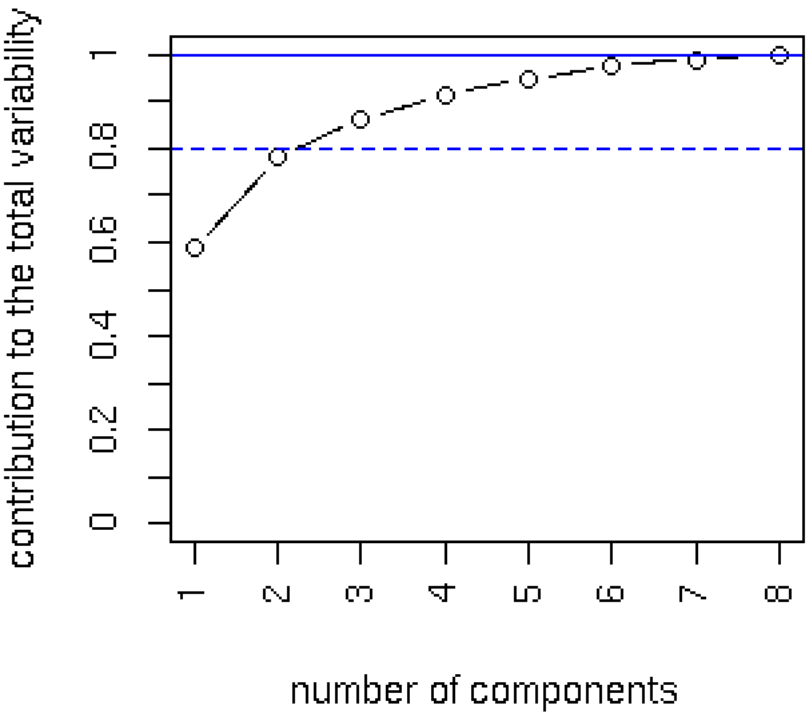
EXERCISE NUMBER 1

POINT A)

In order to perform the PCA, I first do a boxplot to evaluate whether it is appropriate to use the original variables or the scaled ones.



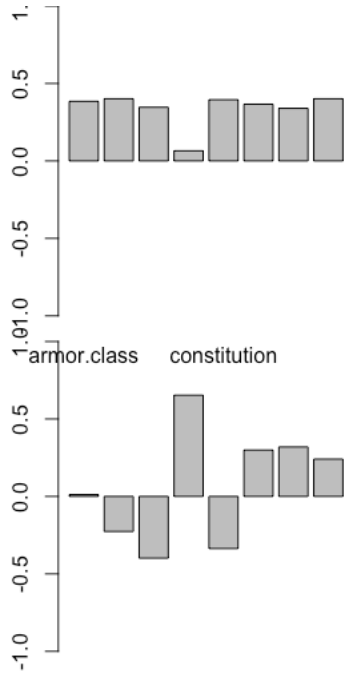
Since one feature has a higher variability scale I decide to standardize the features and perform PCA.



Here is the cumulative Variability explained by the PCs, using an elbow criterion (although not so evident) I can select the first 2 PCs.

POINT B)

Plot of the loadings of the first PCs:

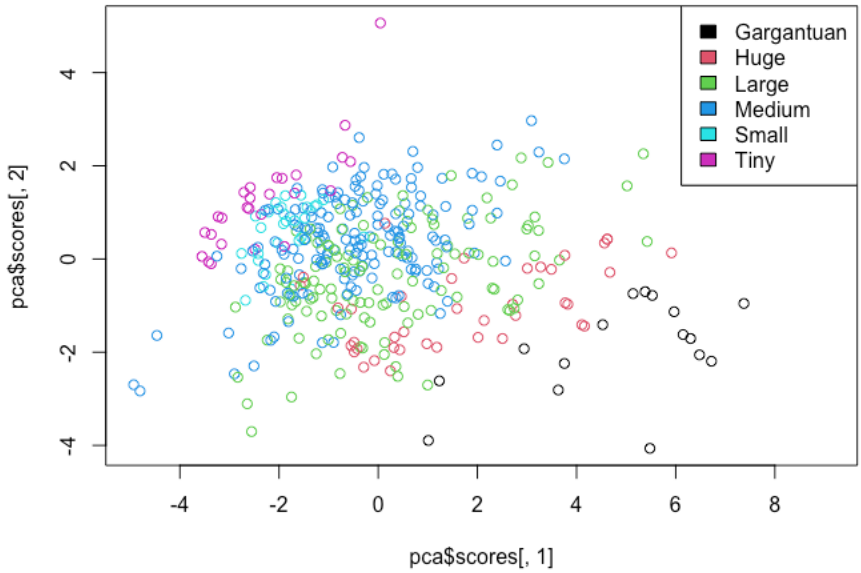


The first one is a weighted mean of everything but the 4th feature (dexterity).

The second principal component is mainly focused on dexterity (with less weight on intelligence, wisdom and charisma), vs hit points, strength and constitution.

POINT C)

I report the scatter plot of the data along the first two PCs:

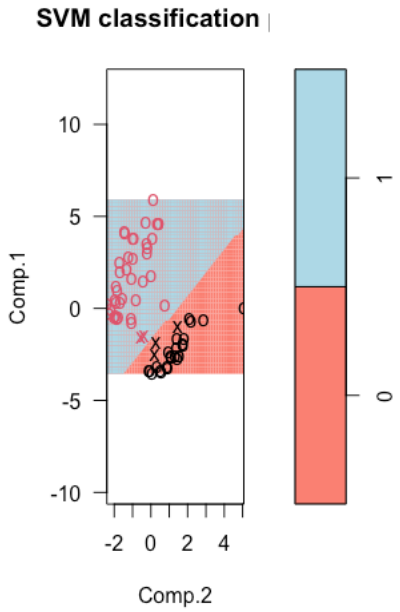


We can see that the bigger the size, the higher the first PC. Meaning that bigger monsters will have in mean greater attributes (except dexterity).

There's also a decreasing trend in the 2nd PC, meaning that bigger monsters will have significantly less dexterity and slightly less intelligence, wisdom as charisma, while generally having more hit points, strength and constitution as expected.

POINT D)

I perform as requested a SVM with linear kernel and cost 1 to classify monsters "Tiny" and "Huge" using the first 2 PCs. I have 5 support vectors. The plot of the classification region is:



And by using the SVM model to predict the new monster, I have a prediction of size Huge.