Question-2 After, we upload the data and giving the column names, we analyzed the scatterplot for all futures. We see that according to plots, there are correlation between NOX and DIS, MEDV and RM, LSTAT and MEDV.

A picture containing food

Description automatically generated

Later, we applied pca to our data and had an outcome. As we see in the class, we can determine a threshold value for variance. If we determine a threshold value of 0.85, we can take comp6. Outcome showed us that cumulative proportion of variance.

A close up of a newspaper

Description automatically generated

A picture containing object, antenna

Description automatically generatedLater, we analyzed the barplot of comp6 and coefficients. Comp6 used crim, zn, cas, nox, rm, rad, tax, ptratip, b for the equation and some of them have negative effect. For example, for the 1st observation, positive effect is bigger than the negative effect which means that with correlations crim+zn+nox+rad+tax+b is bigger than others.

Question – 3

library(dplyr)

library(data.table)

After we upload the data, we identified the column names as movie names. We changed 0 values to NA values to fill the data with mean value for each column. We used dplyr package to fill the NA values and as we mentioned before we fill each column with its mean values. We took the transpose of the data to create a square matrix. Then, we applied distance function to data and created a matrix. We applied cmd to the matrix and drew the plot. It is very hard to make comments on plot.

A screenshot of a social media post

Description automatically generated