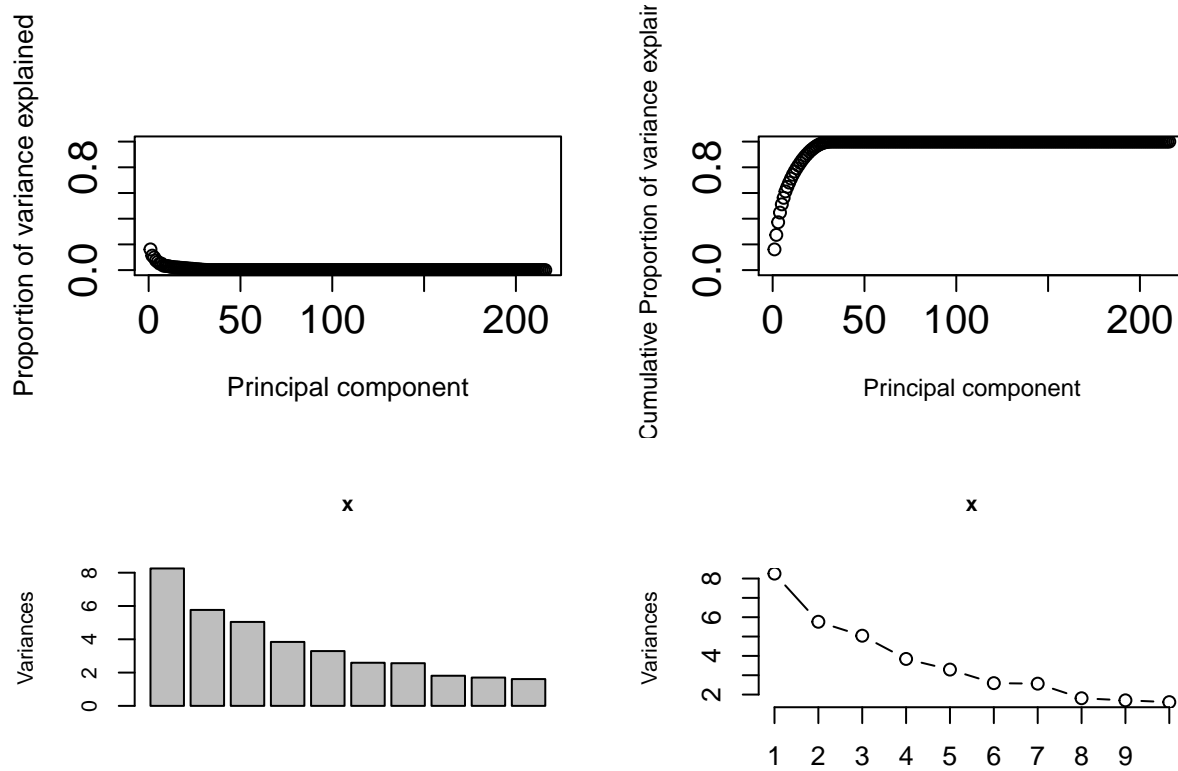


pcr

3. PCR (Iris)



The four graphs above show the variance explained for each component after doing PCA on our model. According to the graphs above, proportion of variance explained and variance are decreasing as number of components increases. Cumulative variance of the model increases as number of principal components increases. As we can see from the upper right graph, more than 90% of variance is explained when there are around 25 principal components.

Then we further explore the details of variance explained for each number of components. We found out that when there are 32 components, the variance explained will reach 100% and the variance will reach 90% when there are 19 components. Since we wanted to obtain as much variance as we could, but also trying to limit the number of predictors, we calculated the cross validation MSE for each number components in a pcr model to get a relatively ideal number of components. As we can see from the graph below, the cross validation MSE is the lowest when there are 28 components in the PCR model. Therefore we applied 28 components to test the result of our testing data set.

SalePrice

