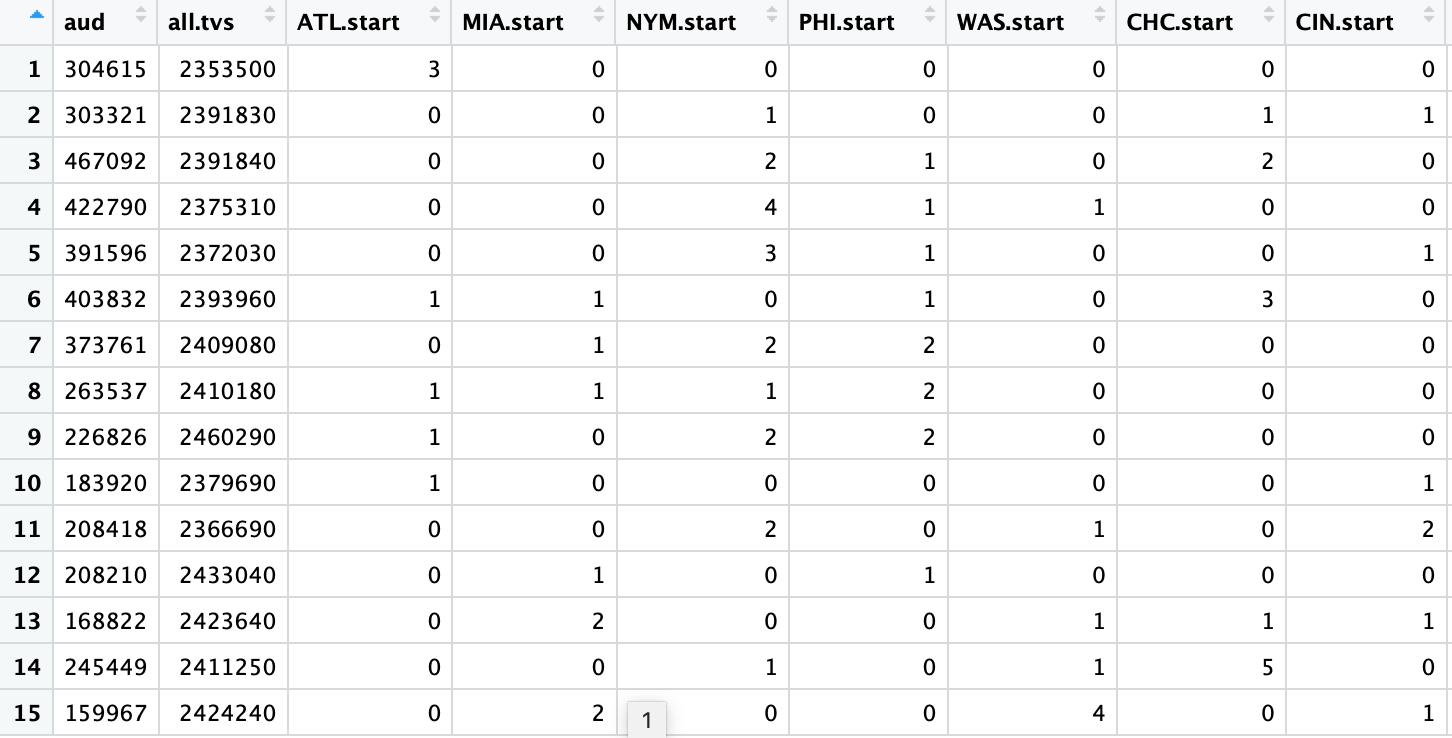
Class Prep

Data

Evidence of correct data

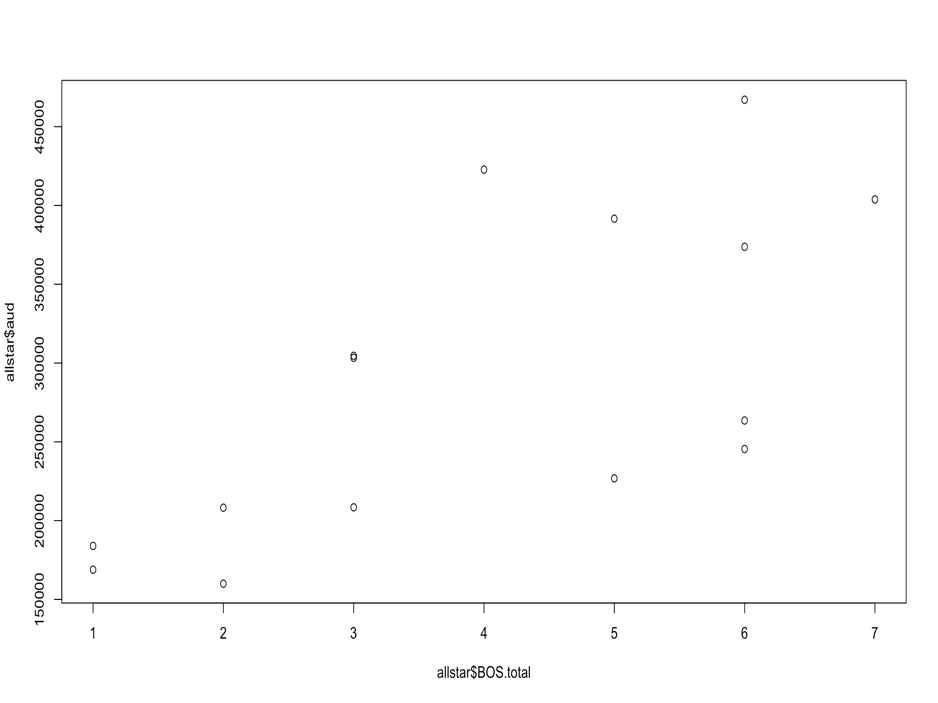


Response variable: The average TV audience (in thousand households)

(Quantitative)

Explanatory variables: 546 explanatory variables describing All-Star Game roster and game characteristics

EDA:



The correlation coefficient is 0.679, It appears a positive linear correlation.

Since the BOSTON TV viewers want to see local players in the All-Star Game, they are more likely to watch it on TV if there are more players from Boston are playing in the All-Star game.

It is not feasible to examine the correlation with all explanatory variables. The appearance of players from outside of BOSTON does not affect the average BOSTON TV audience.

It is impossible to fit all explanatory variables, because some variables provide similar information and some variables are not correlated. Fitting all variables will make the model hard to predict.

I calculated all correlation coefficients between the response variable and explanatory variables. I would also include all variables with coefficients whose absolute values are greater than 0.2. I would include most of the atBreak variables.