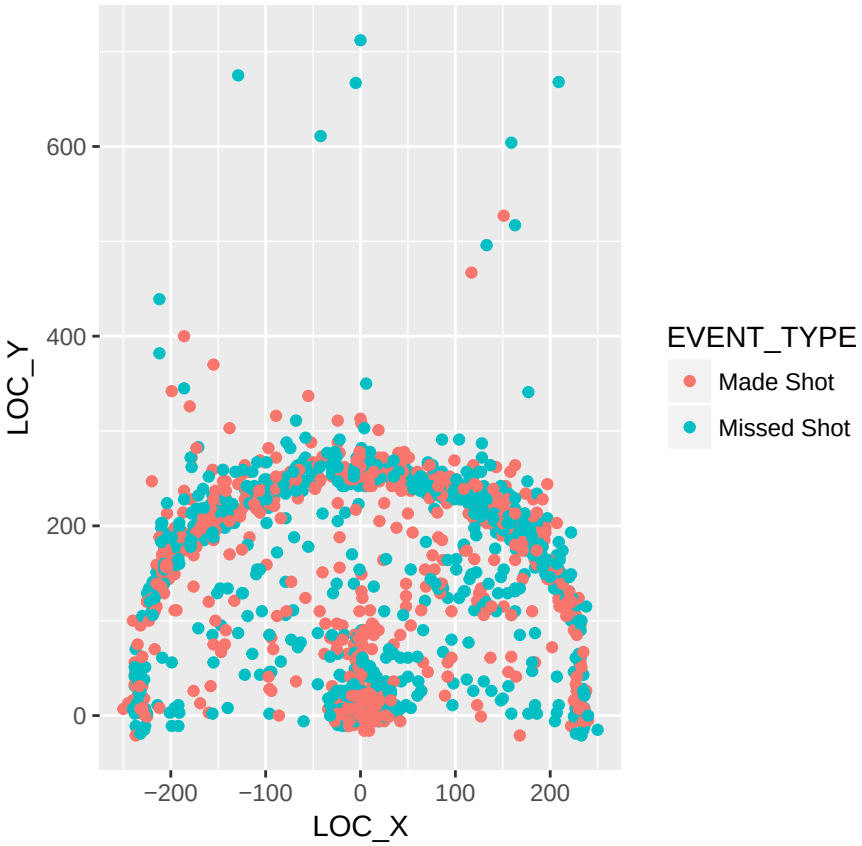
Assignment- 7

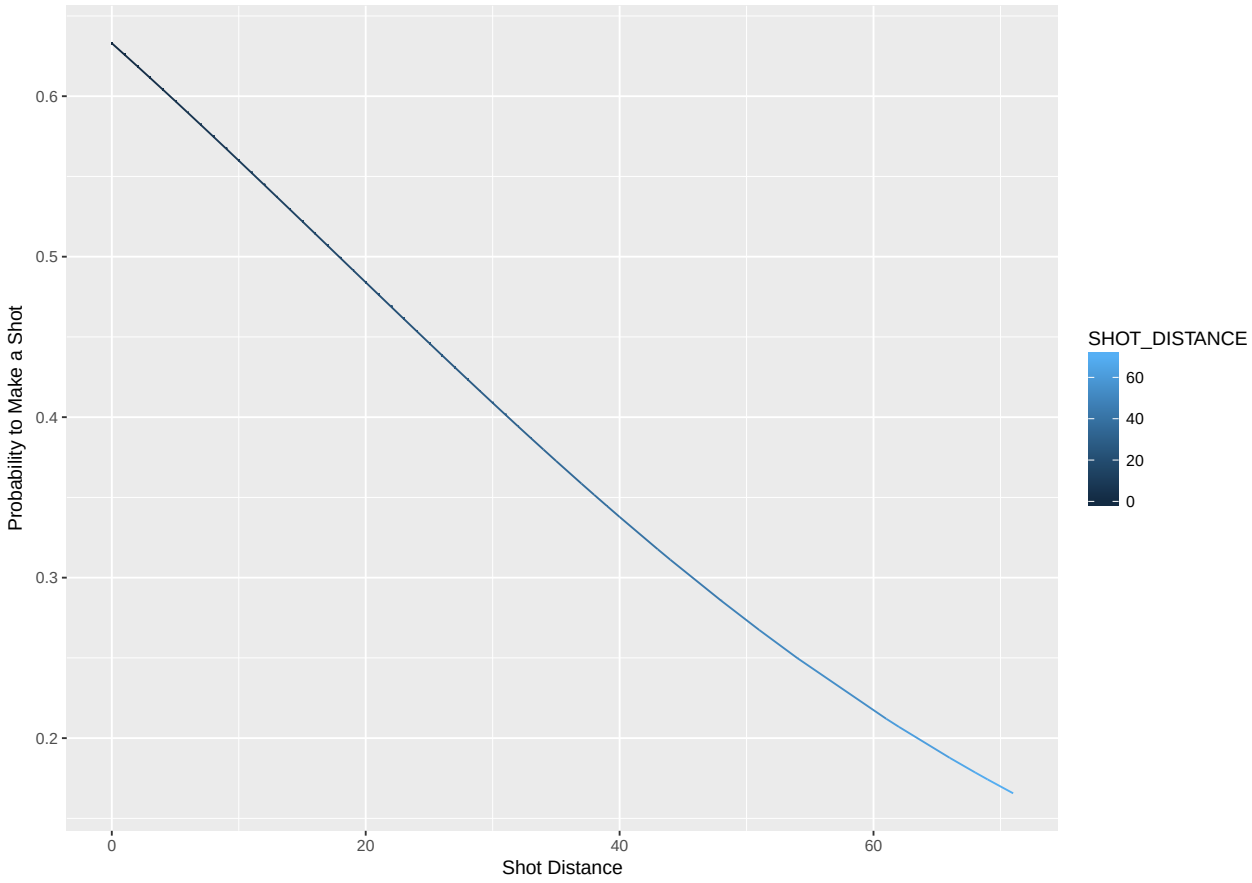
Dwipam

3/29/2017

Question 1:

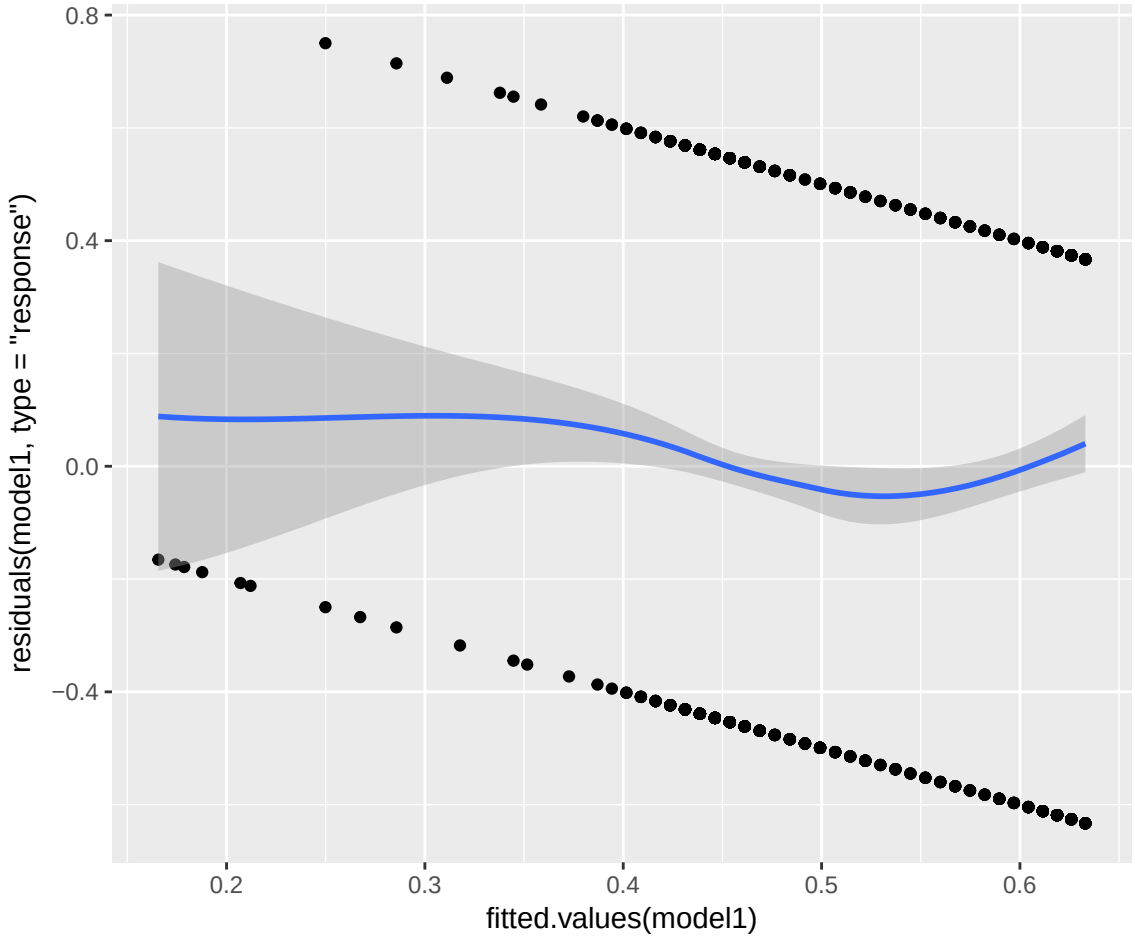


Question 2:



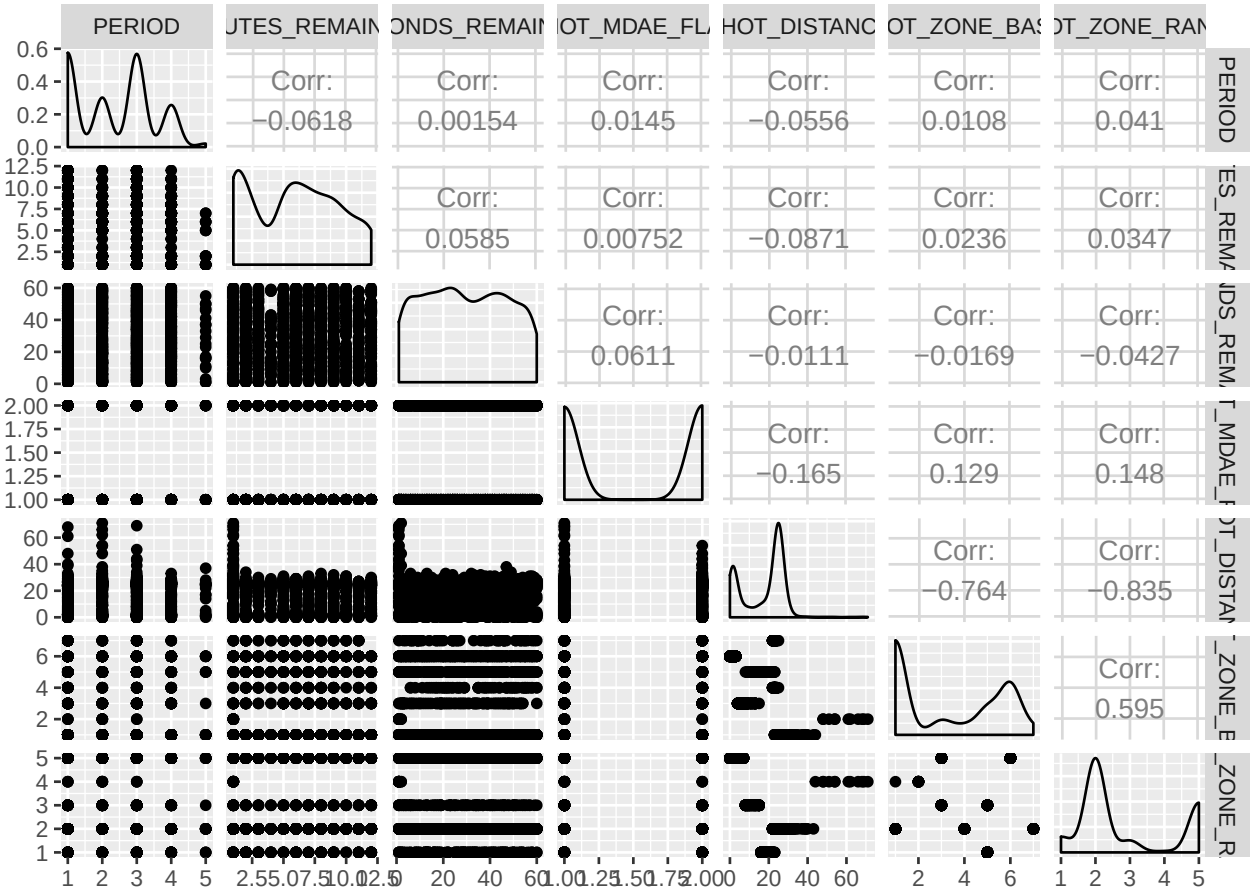
There is a linear decrease in probability to make a shot. As distance increases probability to make a shot decreases. Median distance is 23 and it seems at Median distance there is 50-50 chance of making a shot. Equation for Logistic regression is :
logit[p(MadeShot|ShotDistance)] = 0.54 - 0.03045 x Shot_distance

Question 3:

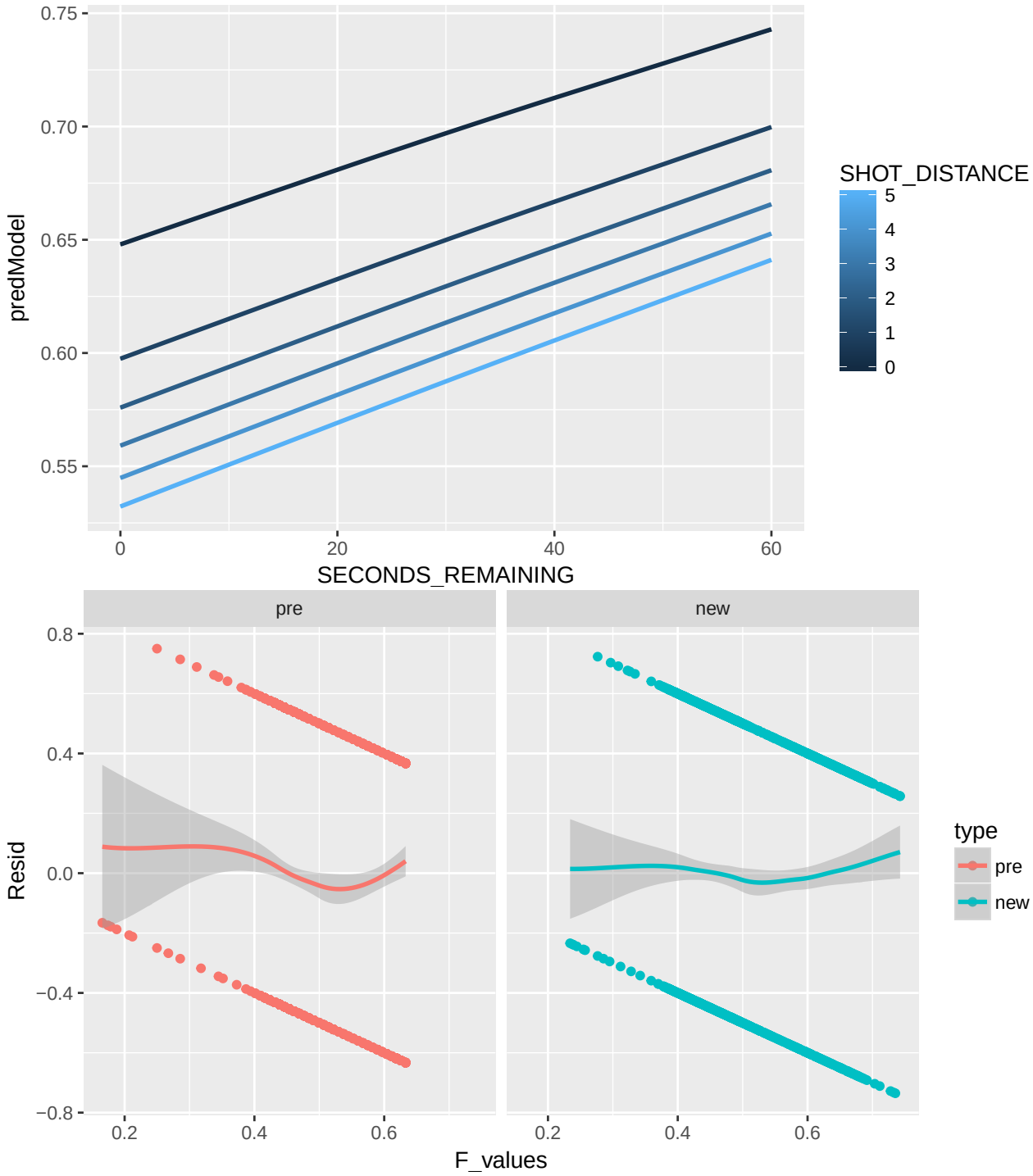


It seems that there is stil much pattern left in residuals. Also the confidence interval band does not have 0. There are several regions where probability is underestimated or over estimated. This might be due to Bi-modal distribution of SHOT_DISTANCE variable. This suggests that we might need transformation of SHOT DISTANCE.

Question 4:



We might want to consider Seconds_remaining as, if there are less seconds remaining a player might try to goal rather than more thinking of accuracy. Also in our previous analysis there was a transformation required, Let's try with square root transformation. We could have also included SHOT ZONE RANGE and SHOT ZONE BASIC, however they are strongly correlated to each other.



Looking at Seconds remaining plot, it seems that as the seconds increases probability to make a shot also increases. Looking at the residual plot it is smoother than old model and also transformation of SHOT_DISTANCE made the residual fit smoother, having confidence band covering 0 line. Hence it seems this is a reasonable model.