Objective: understand what effect different factors have on a future match being conceded

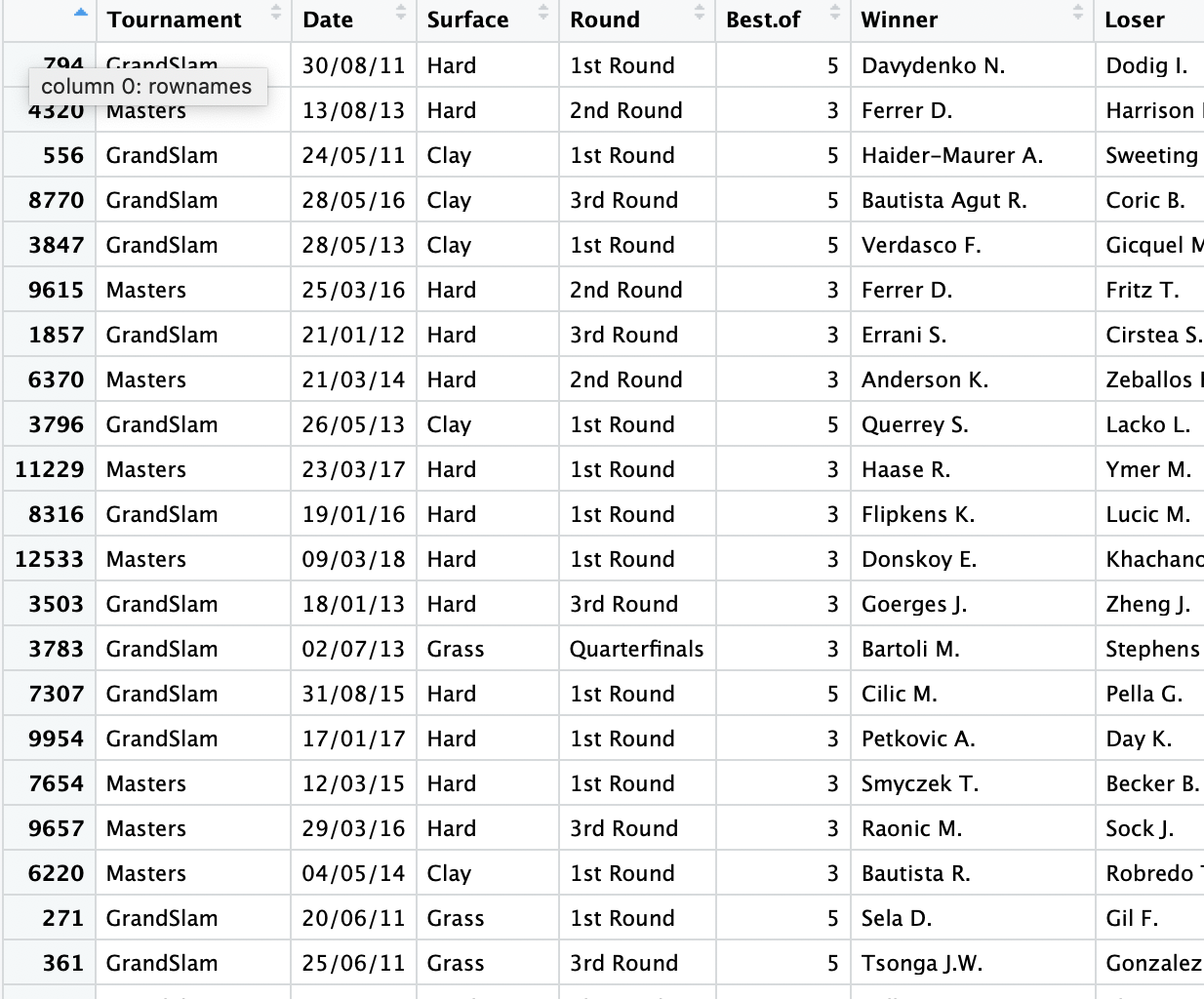
Class Prep

Data

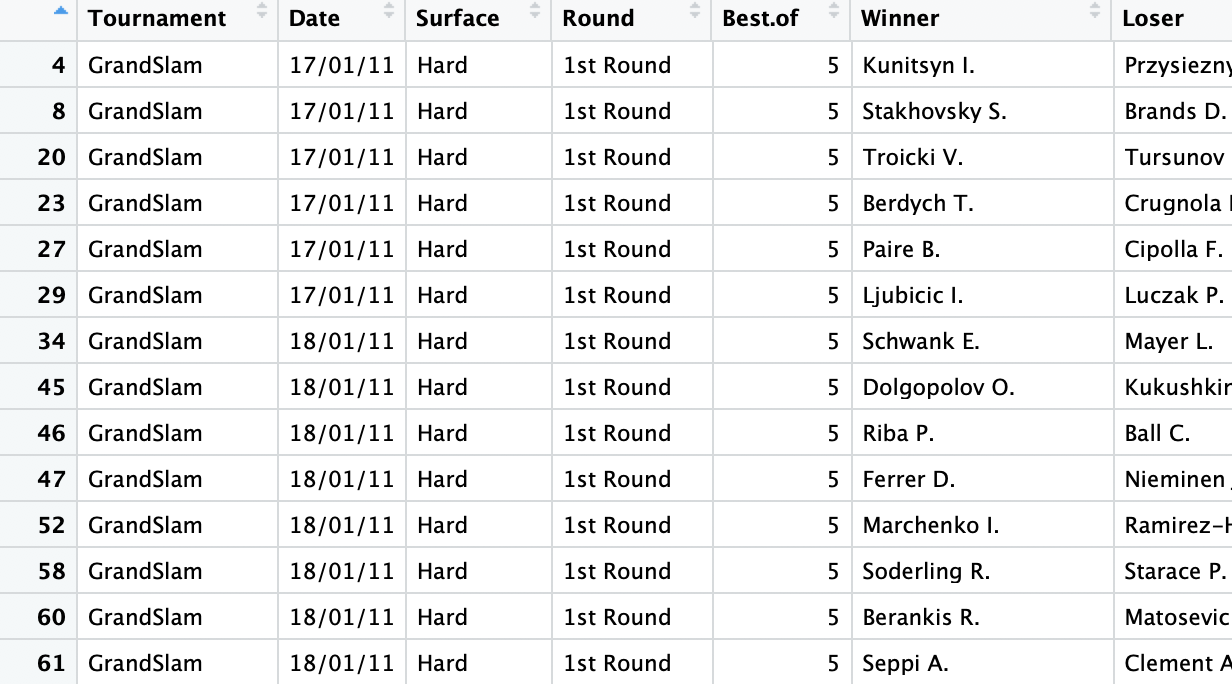
Evidence of correct data



Train set



Test set



Response variable: DNF(whether or not match being conceded)(Categorical)

Explanatory variables: Gender(Categorical), Date(Categorical),Tournament(Categorical), Surface(Categorical), Round(Categorical), Best.of(Categorical), WRank(Quantitative), LRank(Quantitative), WPts(Quantitative), LPts(Quantitative), AvgW(Quantitative), AvgL(Quantitative), Wets(Quantitative), Lsets(Quantitative).

Boxplot of DNF and each quantitative varaiables

