In order to analyze potential correlations between our observed variables we visualized our data as a correlation matrix.

FIFA\_Master<-read.csv("Fifa\_Master.csv")

corrplot(corr=cor(FIFA\_Master[,2:9], use="complete.obs"), method="ellipse")

We could observe strong correltions between our dependent variable Win.Prob and our independent variables TotalGames, Half.Prob, Goals\_per\_game and total.top.scorers. We could observe a negative correlation between Win.Prob and Goalsag\_per\_game. The correlations between Win.Prob and cards\_per\_game describes a positive relationship between awarded cards and probability to win. This would indicate that playing more aggressive would increase a team’s chance to win. In addition, we could observe a negative correlation between Win.Prob and penconv, which indicates that converting penalty shots does not increase a team’s probability to win.



We quickly visualized the relationships between pairs of variables in our data. The scatterplot matrix below shows, that the data for TotalGame, penconv and total.top.scorer is skewed and needs to be transformed for our model.

library(gpairs)

gpairs(FIFA\_Master[, 2:9])

