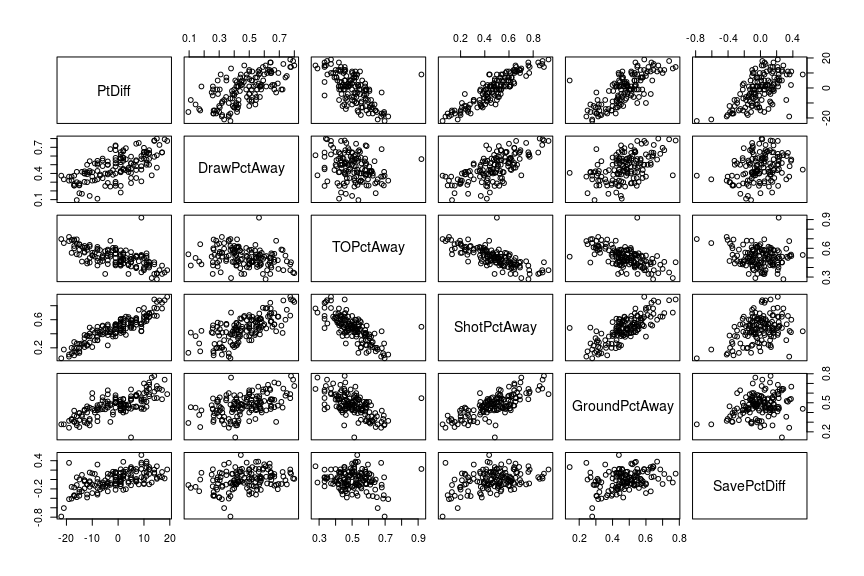
Women's Lacrosse USCOTS 2017

Jeff Witmer

5/22/2017

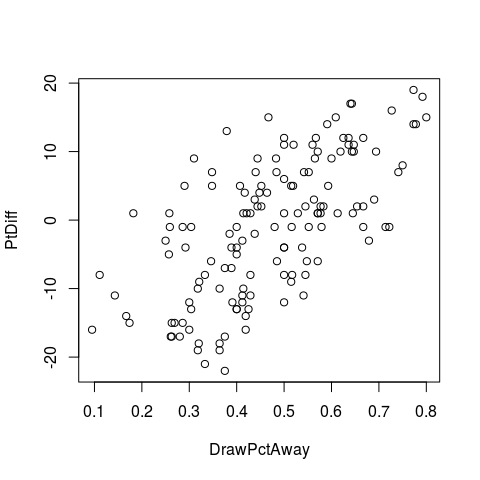
Here are some analyses of the lacrosse data. We start with a scatterplot matrix for some of the variables.

SomeLaX <- LaXdata %>% select(c(PtDiff,DrawPctAway,TOPctAway,ShotPctAway,GroundPctAway,SavePctDiff))  
pairs(SomeLaX)



The coach had suggested that draw percentage is a strong predictor of success, so let's focus on this variable and how it relates to point differential (which determines win/lose).

plot(PtDiff~DrawPctAway,data=LaXdata)

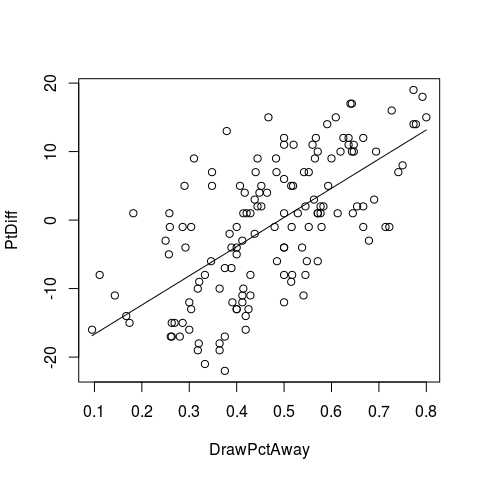


We can add a regression line:

plot(PtDiff~DrawPctAway,data=LaXdata)  
summary(PtModel <- lm(PtDiff~DrawPctAway,data=LaXdata))

##   
## Call:  
## lm(formula = PtDiff ~ DrawPctAway, data = LaXdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -17.059 -5.605 0.055 5.669 17.771   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -20.91 2.07 -10.1 <2e-16 \*\*\*  
## DrawPctAway 42.59 4.18 10.2 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7.63 on 142 degrees of freedom  
## Multiple R-squared: 0.422, Adjusted R-squared: 0.418   
## F-statistic: 104 on 1 and 142 DF, p-value: <2e-16

Predict.Plot(PtModel,pred.var = "DrawPctAway", add=TRUE)

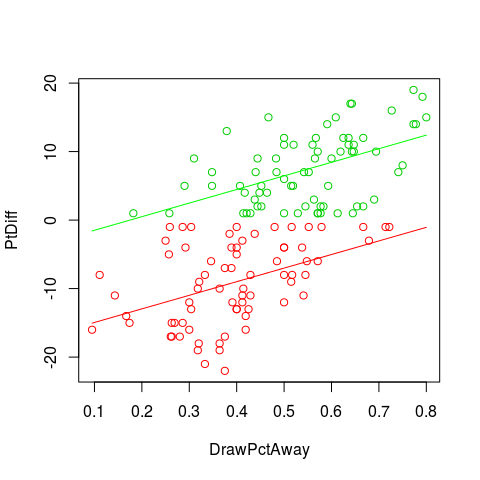


Now color points by win/lose and add regression lines for each colored part of the data.

plot(PtDiff~DrawPctAway,col=Win+2,data=LaXdata)  
summary(PtModelWin <- lm(PtDiff~DrawPctAway+Win,data=LaXdata))

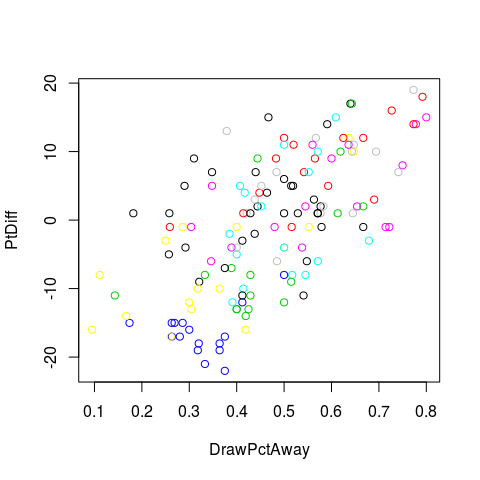
##   
## Call:  
## lm(formula = PtDiff ~ DrawPctAway + Win, data = LaXdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12.516 -3.750 0.221 3.153 10.785   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -16.923 1.375 -12.31 < 2e-16 \*\*\*  
## DrawPctAway 19.838 3.167 6.26 4.3e-09 \*\*\*  
## Win 13.460 0.962 13.99 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.95 on 141 degrees of freedom  
## Multiple R-squared: 0.758, Adjusted R-squared: 0.754   
## F-statistic: 221 on 2 and 141 DF, p-value: <2e-16

Predict.Plot(PtModelWin,pred.var = "DrawPctAway", Win=1, plot.args=list(col="green"),add=TRUE)  
Predict.Plot(PtModelWin,pred.var = "DrawPctAway", Win=0, plot.args=list(col="red"),add=TRUE)



How about coloring points by year?

plot(PtDiff~DrawPctAway,col=AwayTeam,data=LaXdata)

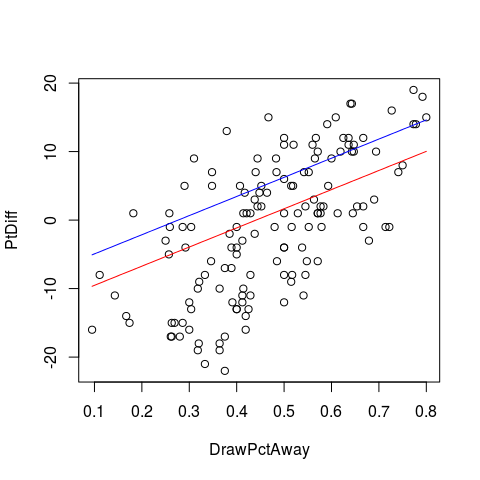


We can add regression lines for some teams.

plot(PtDiff~DrawPctAway,data=LaXdata)  
TeamModel <- lm(PtDiff~DrawPctAway+AwayTeam,data=LaXdata)  
summary(TeamModel)

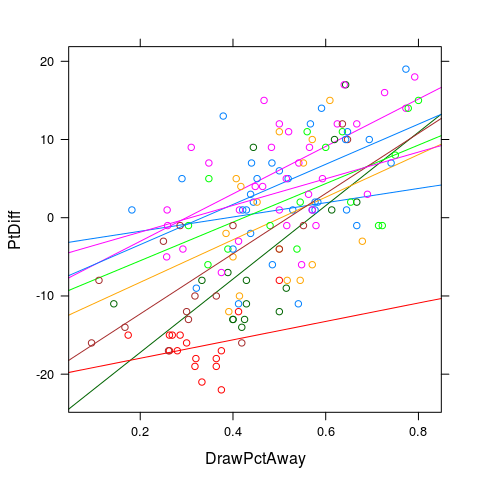
##   
## Call:  
## lm(formula = PtDiff ~ DrawPctAway + AwayTeam, data = LaXdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -13.791 -4.432 -0.271 3.936 16.902   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -12.32099 2.48727 -4.95 2.2e-06 \*\*\*  
## DrawPctAway 27.93395 4.14776 6.73 4.4e-10 \*\*\*  
## AwayTeamDenison 4.59662 2.25765 2.04 0.044 \*   
## AwayTeamDePauw -5.54259 2.21621 -2.50 0.014 \*   
## AwayTeamHiram -13.25056 2.29191 -5.78 5.0e-08 \*\*\*  
## AwayTeamKenyon -1.73886 2.21930 -0.78 0.435   
## AwayTeamOberlin 0.00932 2.26016 0.00 0.997   
## AwayTeamOWU -3.97335 2.26692 -1.75 0.082 .   
## AwayTeamWitt 2.48210 2.24899 1.10 0.272   
## AwayTeamWooster 2.23955 2.21799 1.01 0.314   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.27 on 134 degrees of freedom  
## Multiple R-squared: 0.632, Adjusted R-squared: 0.607   
## F-statistic: 25.5 on 9 and 134 DF, p-value: <2e-16

Predict.Plot(TeamModel,pred.var = "DrawPctAway", AwayTeam="Oberlin", plot.args=list(col='red'), add=TRUE)  
Predict.Plot(TeamModel,pred.var = "DrawPctAway", AwayTeam="Denison", plot.args=list(col='blue'), add=TRUE)



We can add regression lines for all teams.

library(lattice)  
xyplot( PtDiff ~ DrawPctAway, data = LaXdata, groups=AwayTeam, main="", type=c("p","r"))



Now to create the two categories:

LaXnew <- LaXdata %>%  
 mutate(GoodTeam = AllWins >= 18)  
   
with(LaXnew, table(AwayTeam, GoodTeam))

## GoodTeam  
## AwayTeam FALSE TRUE  
## Allegheny 0 16  
## Denison 0 16  
## DePauw 16 0  
## Hiram 16 0  
## Kenyon 16 0  
## Oberlin 0 16  
## OWU 16 0  
## Witt 0 16  
## Wooster 0 16

Now, back to making a scatterplot, with points color-coded by GoodTeam and with regression lines added.

xyplot( PtDiff ~ DrawPctAway, data = LaXnew, groups=GoodTeam, main="", type=c("p","r"), auto.key=list(space="top", columns=3))

## Warning in draw.key(simpleKey(...), draw = FALSE): not enough rows for  
## columns



We see that there is an interaction present: The slope between PtDiff and DrawPctAway is greater for weak team than for strong teams. This suggests that draw percentage is more important for a team that has a losing record. For such a team to be predicted to win, they need to control about 60% of the draws. A strong team will be predicted to win even if they only control 40% of the draws.

summary(lm( PtDiff ~ DrawPctAway\*GoodTeam, data = LaXnew))

##   
## Call:  
## lm(formula = PtDiff ~ DrawPctAway \* GoodTeam, data = LaXnew)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.588 -5.020 -0.858 5.120 14.383   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -25.84 2.61 -9.90 < 2e-16 \*\*\*  
## DrawPctAway 46.08 6.07 7.59 4.0e-12 \*\*\*  
## GoodTeamTRUE 17.26 3.85 4.48 1.5e-05 \*\*\*  
## DrawPctAway:GoodTeamTRUE -21.74 8.01 -2.71 0.0075 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6.7 on 140 degrees of freedom  
## Multiple R-squared: 0.56, Adjusted R-squared: 0.551   
## F-statistic: 59.5 on 3 and 140 DF, p-value: <2e-16

# Here is a model of points differential.

Ptmodel = lm(PtDiff ~ ShotDiff + DrawPctAway + GroundDiff + FreeShotPctDiff + SavePctDiff, data=LaXdata)  
summary(Ptmodel)

##   
## Call:  
## lm(formula = PtDiff ~ ShotDiff + DrawPctAway + GroundDiff + FreeShotPctDiff +   
## SavePctDiff, data = LaXdata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9.001 -1.667 -0.062 1.862 6.014   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1.2191 1.0161 -1.20 0.2323   
## ShotDiff 0.3857 0.0244 15.80 <2e-16 \*\*\*  
## DrawPctAway 3.6810 2.0410 1.80 0.0735 .   
## GroundDiff 0.1296 0.0388 3.34 0.0011 \*\*   
## FreeShotPctDiff 3.3571 0.6988 4.80 4e-06 \*\*\*  
## SavePctDiff 13.8946 1.3238 10.50 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.74 on 138 degrees of freedom  
## Multiple R-squared: 0.927, Adjusted R-squared: 0.925   
## F-statistic: 352 on 5 and 138 DF, p-value: <2e-16

And a prediction of points differential for a particular game: Oberlin at Witt, 2017.

evaluate\_model(Ptmodel,ShotDiff=4,DrawPctAway=0.6,GroundDiff=2,FreeShotPctDiff=0.321,SavePctDiff=0.324,type="response")

## ShotDiff DrawPctAway GroundDiff FreeShotPctDiff SavePctDiff model\_output  
## 1 4 0.6 2 0.321 0.324 8.37

The model predicts that Oberlin (the Away team) will win by 8.37 goals. Oberlin actually won by 9 goals.