Midterm Project

Christopher Flores and Zack Chipperfield

library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

data\_url <- 'https://raw.githubusercontent.com/florescss14/STAT1341\_Midterm\_Project/main/Bundesliga.csv'  
bund <- readr::read\_csv(data\_url, col\_names = TRUE)

## Rows: 72 Columns: 29

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Season, Club  
## dbl (27): MP, W, D, L, GF, GA, GD, Pts, Total\_Comp, Total\_Att, Total\_Cmp\_Pct...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_url <- 'https://raw.githubusercontent.com/florescss14/STAT1341\_Midterm\_Project/main/La\_Liga.csv'  
liga <- readr::read\_csv(data\_url, col\_names = TRUE)

## Rows: 80 Columns: 29

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Season, Club  
## dbl (27): MP, W, D, L, GF, GA, GD, Pts, Total\_Comp, Total\_Att, Total\_Cmp\_Pct...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_url <- 'https://raw.githubusercontent.com/florescss14/STAT1341\_Midterm\_Project/main/Ligue\_1.csv'  
ligue <- readr::read\_csv(data\_url, col\_names = TRUE)

## Rows: 80 Columns: 29

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Season, Club  
## dbl (27): MP, W, D, L, GF, GA, GD, Pts, Total\_Comp, Total\_Att, Total\_Cmp\_Pct...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_url <- 'https://raw.githubusercontent.com/florescss14/STAT1341\_Midterm\_Project/main/Premier\_League.csv'  
epl <- readr::read\_csv(data\_url, col\_names = TRUE)

## Rows: 80 Columns: 29

## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Season, Club  
## dbl (27): MP, W, D, L, GF, GA, GD, Pts, Total\_Comp, Total\_Att, Total\_Cmp\_Pct...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_url <- 'https://raw.githubusercontent.com/florescss14/STAT1341\_Midterm\_Project/main/Serie\_A.csv'  
serie <- readr::read\_csv(data\_url, col\_names = TRUE)

## Rows: 80 Columns: 29

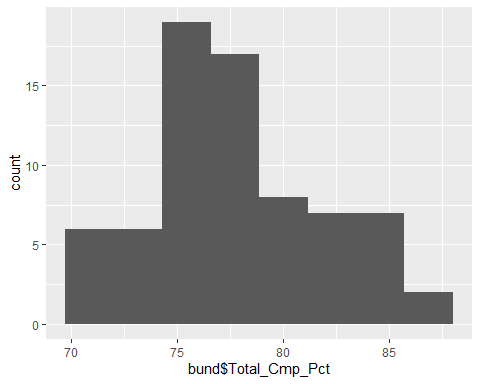
## -- Column specification --------------------------------------------------------  
## Delimiter: ","  
## chr (2): Season, Club  
## dbl (27): MP, W, D, L, GF, GA, GD, Pts, Total\_Comp, Total\_Att, Total\_Cmp\_Pct...

##   
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

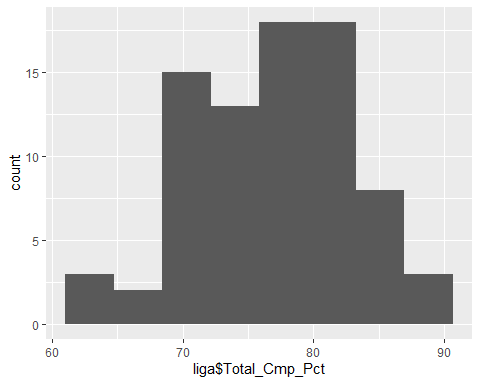
As some leagues have a different number of matches played we want to normalize it to number of points scored per match in each league.

bund$Pts.MP <- bund$Pts / bund$MP  
liga$Pts.MP <- liga$Pts / liga$MP  
ligue$Pts.MP <- ligue$Pts / ligue$MP  
epl$Pts.MP <- epl$Pts / epl$MP  
serie$Pts.MP <- serie$Pts / serie$MP  
  
  
all <- rbind(bund, liga, ligue, epl, serie)

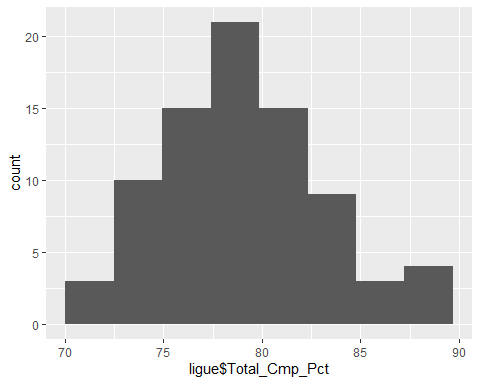
ggplot() +  
geom\_histogram(mapping = aes(x = bund$Total\_Cmp\_Pct), bins = 8)



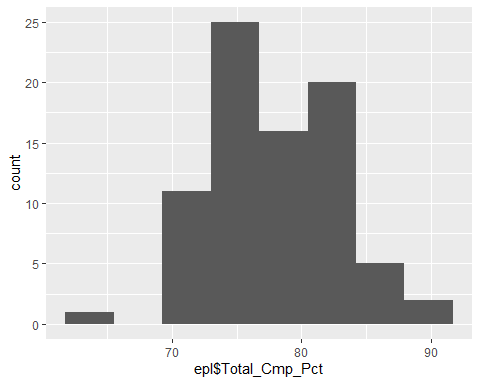
ggplot() +  
geom\_histogram(mapping = aes(x = liga$Total\_Cmp\_Pct), bins = 8)



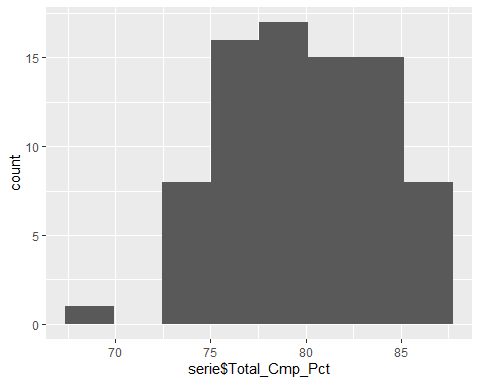
ggplot() +  
 geom\_histogram(mapping = aes(x = ligue$Total\_Cmp\_Pct), bins = 8)



ggplot() +  
 geom\_histogram(mapping = aes(x = epl$Total\_Cmp\_Pct), bins = 8)



ggplot() +  
 geom\_histogram(mapping = aes(x = serie$Total\_Cmp\_Pct), bins = 8)



summary(bund$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 70.60 75.28 77.25 77.82 80.80 86.60

summary(liga$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 63.20 72.45 77.80 77.01 80.50 89.10

summary(ligue$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 71.50 76.47 79.00 79.09 81.22 88.70

summary(epl$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 62.10 74.05 77.25 77.96 82.12 88.30

summary(serie$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 69.50 76.78 79.75 79.77 83.03 87.30

sd(bund$Total\_Cmp\_Pct)

## [1] 4.05822

sd(liga$Total\_Cmp\_Pct)

## [1] 5.876715

sd(ligue$Total\_Cmp\_Pct)

## [1] 3.996561

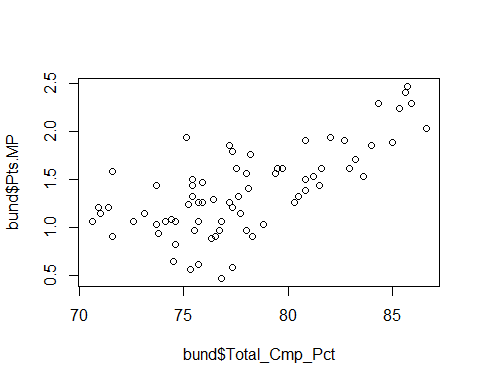
sd(epl$Total\_Cmp\_Pct)

## [1] 5.148746

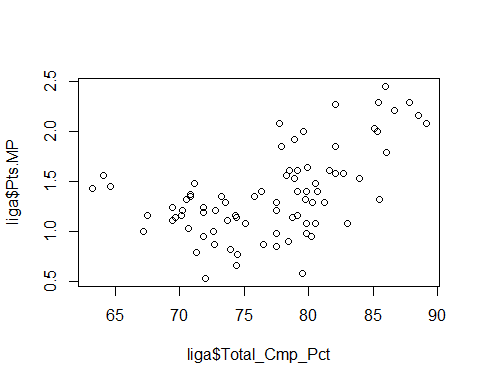
sd(serie$Total\_Cmp\_Pct)

## [1] 3.940542

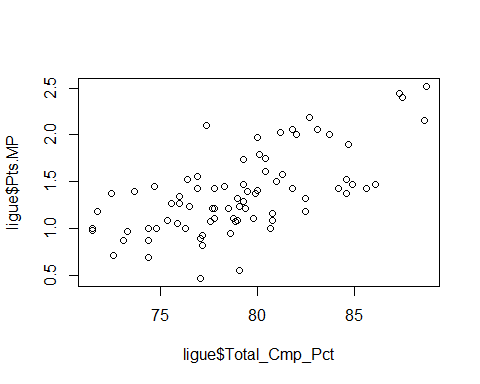
plot(bund$Total\_Cmp\_Pct, bund$Pts.MP)



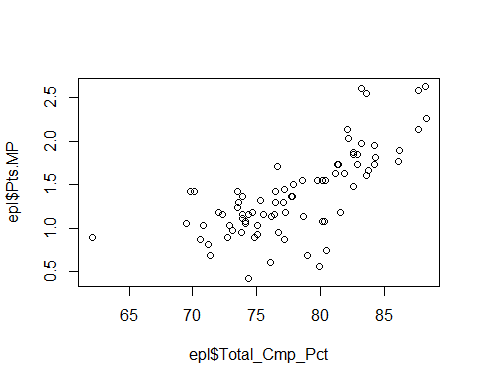
plot(liga$Total\_Cmp\_Pct, liga$Pts.MP)



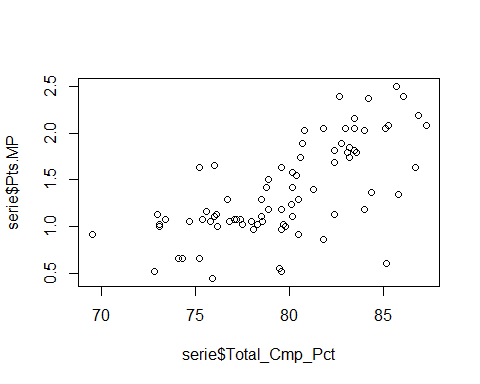
plot(ligue$Total\_Cmp\_Pct, ligue$Pts.MP)



plot(epl$Total\_Cmp\_Pct, epl$Pts.MP)



plot(serie$Total\_Cmp\_Pct, serie$Pts.MP)

 Let’s look at the correlation between Passing completion and points for each league.

cor(bund$Total\_Cmp\_Pct, bund$Pts.MP)

## [1] 0.6918816

cor(liga$Total\_Cmp\_Pct, liga$Pts.MP)

## [1] 0.5465845

cor(ligue$Total\_Cmp\_Pct, ligue$Pts.MP)

## [1] 0.6535265

cor(epl$Total\_Cmp\_Pct, epl$Pts.MP)

## [1] 0.7229535

cor(serie$Total\_Cmp\_Pct, serie$Pts.MP)

## [1] 0.6701023

summary(bund$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 70.60 75.28 77.25 77.82 80.80 86.60

summary(liga$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 63.20 72.45 77.80 77.01 80.50 89.10

summary(ligue$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 71.50 76.47 79.00 79.09 81.22 88.70

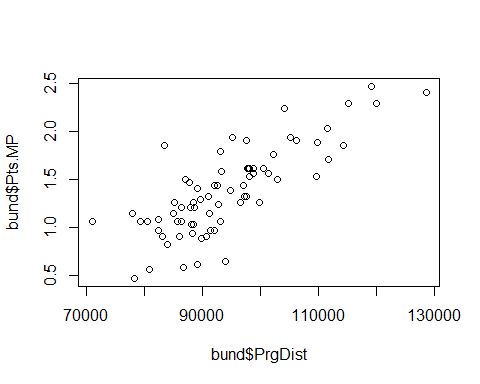
summary(epl$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 62.10 74.05 77.25 77.96 82.12 88.30

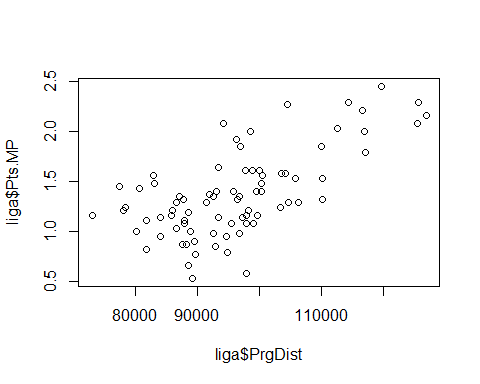
summary(serie$Total\_Cmp\_Pct)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 69.50 76.78 79.75 79.77 83.03 87.30

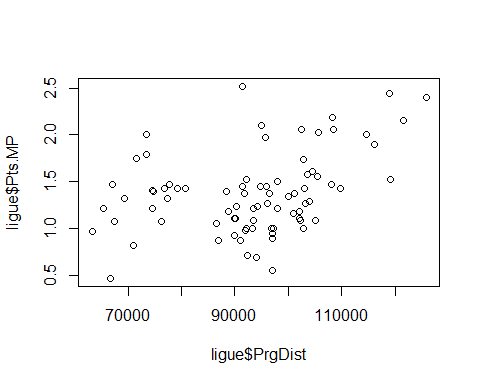
plot(bund$PrgDist, bund$Pts.MP)



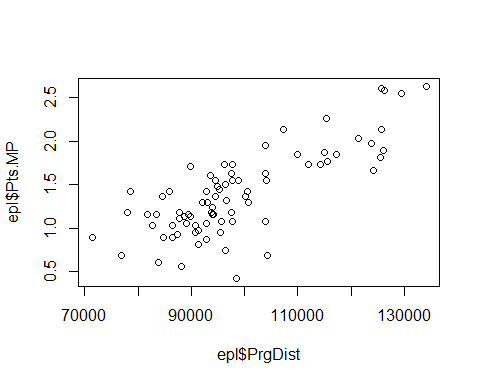
plot(liga$PrgDist, liga$Pts.MP)



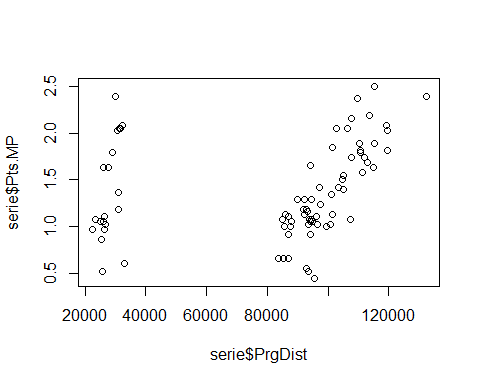
plot(ligue$PrgDist, ligue$Pts.MP)



plot(epl$PrgDist, epl$Pts.MP)



plot(serie$PrgDist, serie$Pts.MP)



cor(bund$PrgDist, bund$Pts.MP)

## [1] 0.7926936

cor(liga$PrgDist, liga$Pts.MP)

## [1] 0.6602864

cor(ligue$PrgDist, ligue$Pts.MP)

## [1] 0.3793589

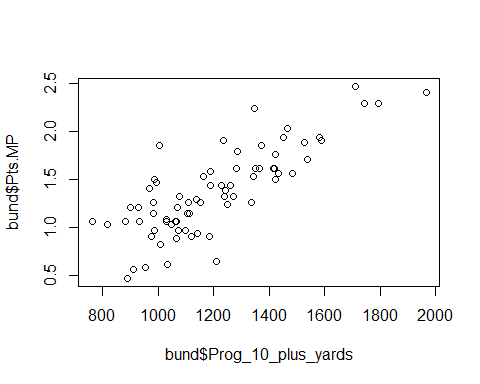
cor(epl$PrgDist, epl$Pts.MP)

## [1] 0.7803466

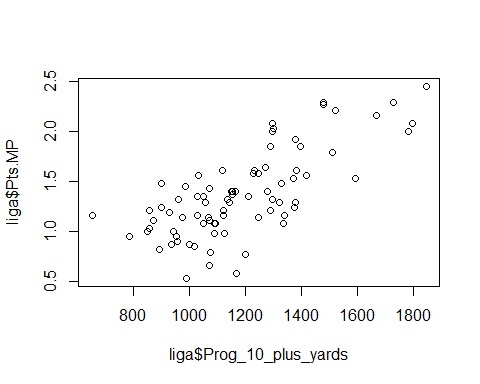
cor(serie$PrgDist, serie$Pts.MP)

## [1] 0.2121837

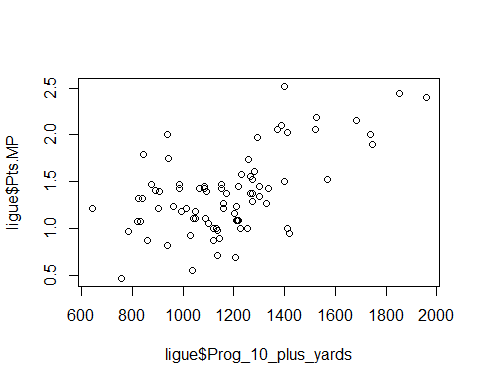
plot(bund$Prog\_10\_plus\_yards, bund$Pts.MP)



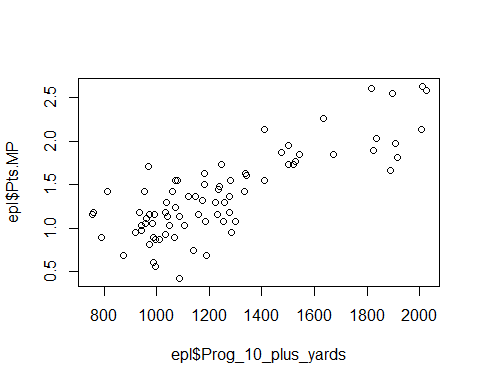
plot(liga$Prog\_10\_plus\_yards, liga$Pts.MP)



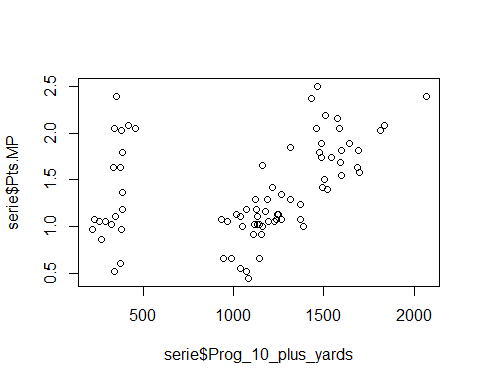
plot(ligue$Prog\_10\_plus\_yards, ligue$Pts.MP)



plot(epl$Prog\_10\_plus\_yards, epl$Pts.MP)



plot(serie$Prog\_10\_plus\_yards, serie$Pts.MP)



cor(bund$Prog\_10\_plus\_yards, bund$Pts.MP)

## [1] 0.8021635

cor(liga$Prog\_10\_plus\_yards, liga$Pts.MP)

## [1] 0.7329325

cor(ligue$Prog\_10\_plus\_yards, ligue$Pts.MP)

## [1] 0.5835585

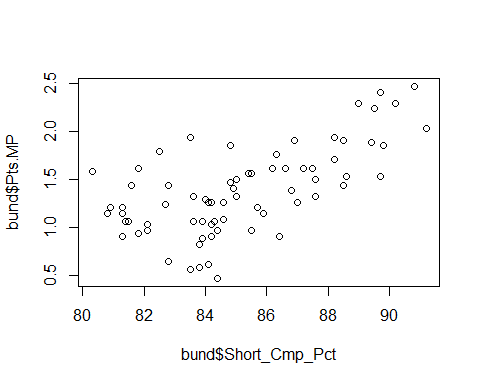
cor(epl$Prog\_10\_plus\_yards, epl$Pts.MP)

## [1] 0.8045364

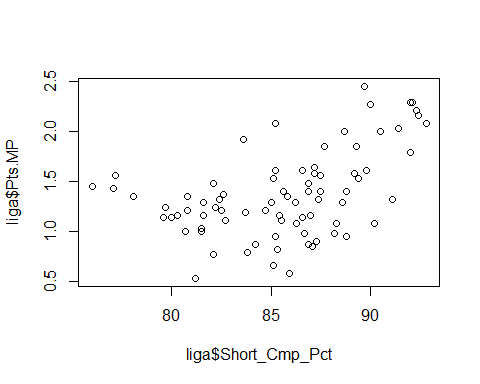
cor(serie$Prog\_10\_plus\_yards, serie$Pts.MP)

## [1] 0.3258032

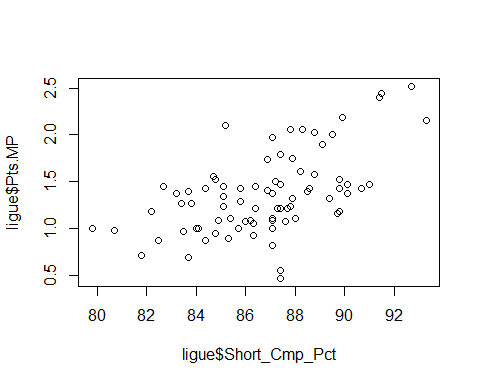
plot(bund$Short\_Cmp\_Pct, bund$Pts.MP)



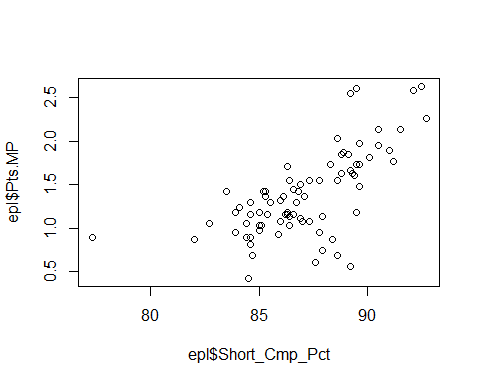
plot(liga$Short\_Cmp\_Pct, liga$Pts.MP)



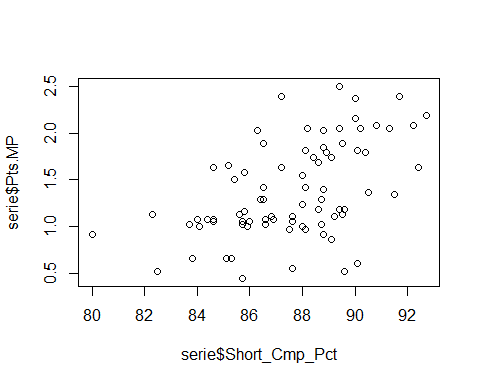
plot(ligue$Short\_Cmp\_Pct, ligue$Pts.MP)



plot(epl$Short\_Cmp\_Pct, epl$Pts.MP)



plot(serie$Short\_Cmp\_Pct, serie$Pts.MP)



cor(bund$Short\_Cmp\_Pct, bund$Pts.MP)

## [1] 0.6448393

cor(liga$Short\_Cmp\_Pct, liga$Pts.MP)

## [1] 0.4686661

cor(ligue$Short\_Cmp\_Pct, ligue$Pts.MP)

## [1] 0.5645258

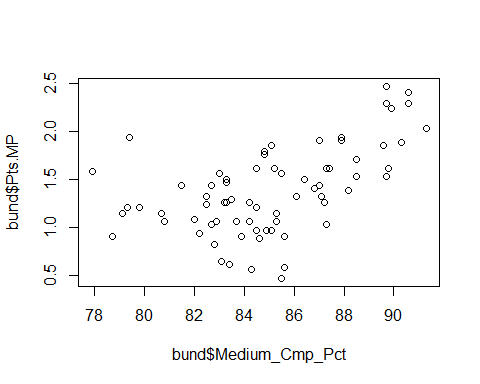
cor(epl$Short\_Cmp\_Pct, epl$Pts.MP)

## [1] 0.6687712

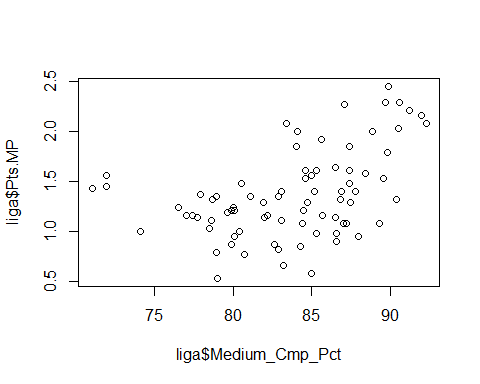
cor(serie$Short\_Cmp\_Pct, serie$Pts.MP)

## [1] 0.5148248

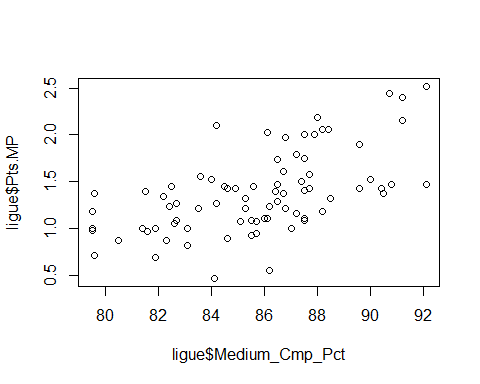
plot(bund$Medium\_Cmp\_Pct, bund$Pts.MP)



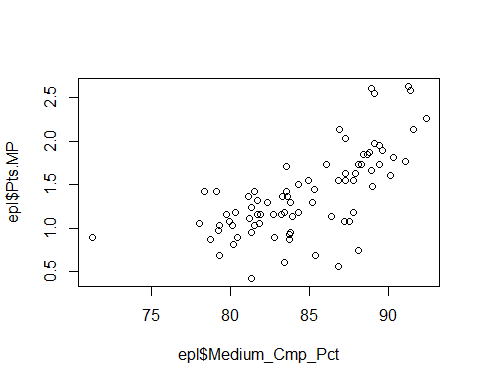
plot(liga$Medium\_Cmp\_Pct, liga$Pts.MP)



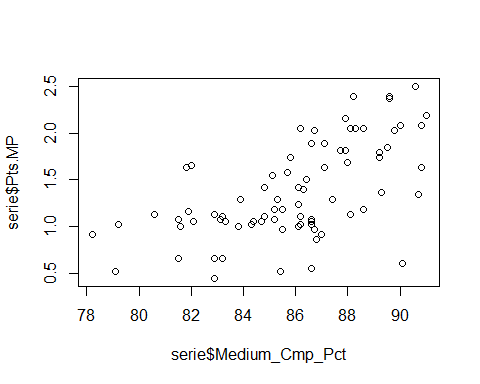
plot(ligue$Medium\_Cmp\_Pct, ligue$Pts.MP)



plot(epl$Medium\_Cmp\_Pct, epl$Pts.MP)



plot(serie$Medium\_Cmp\_Pct, serie$Pts.MP)



cor(bund$Medium\_Cmp\_Pct, bund$Pts.MP)

## [1] 0.5414421

cor(liga$Medium\_Cmp\_Pct, liga$Pts.MP)

## [1] 0.4526962

cor(ligue$Medium\_Cmp\_Pct, ligue$Pts.MP)

## [1] 0.5863414

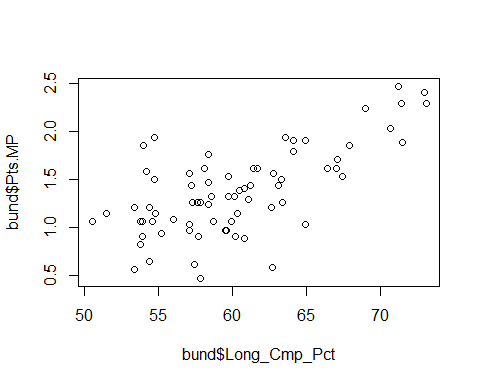
cor(epl$Medium\_Cmp\_Pct, epl$Pts.MP)

## [1] 0.6799676

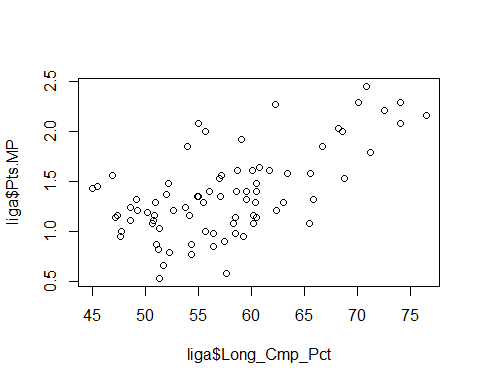
cor(serie$Medium\_Cmp\_Pct, serie$Pts.MP)

## [1] 0.605784

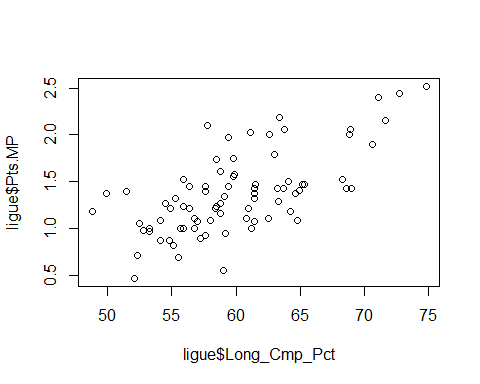
plot(bund$Long\_Cmp\_Pct, bund$Pts.MP)



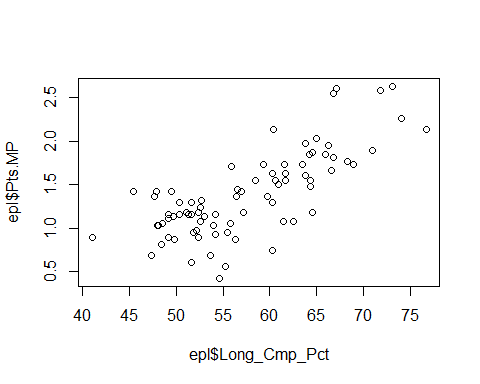
plot(liga$Long\_Cmp\_Pct, liga$Pts.MP)



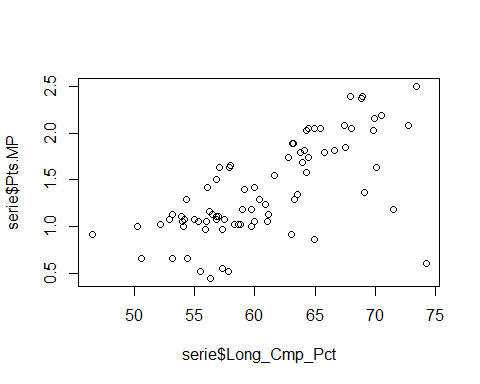
plot(ligue$Long\_Cmp\_Pct, ligue$Pts.MP)



plot(epl$Long\_Cmp\_Pct, epl$Pts.MP)



plot(serie$Long\_Cmp\_Pct, serie$Pts.MP)



cor(bund$Long\_Cmp\_Pct, bund$Pts.MP)

## [1] 0.6724197

cor(liga$Long\_Cmp\_Pct, liga$Pts.MP)

## [1] 0.6191942

cor(ligue$Long\_Cmp\_Pct, ligue$Pts.MP)

## [1] 0.6646574

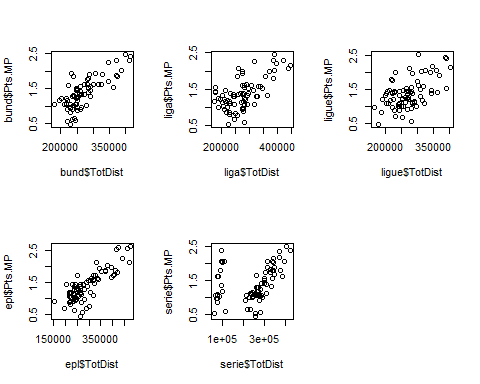
cor(epl$Long\_Cmp\_Pct, epl$Pts.MP)

## [1] 0.75147

cor(serie$Long\_Cmp\_Pct, serie$Pts.MP)

## [1] 0.6802777

par(mfrow=c(2,3))  
plot(bund$TotDist, bund$Pts.MP)  
plot(liga$TotDist, liga$Pts.MP)  
plot(ligue$TotDist, ligue$Pts.MP)  
plot(epl$TotDist, epl$Pts.MP)  
plot(serie$TotDist, serie$Pts.MP)



cor(bund$TotDist, bund$Pts.MP)

## [1] 0.7700547

cor(liga$TotDist, liga$Pts.MP)

## [1] 0.6388306

cor(ligue$TotDist, ligue$Pts.MP)

## [1] 0.5554981

cor(epl$TotDist, epl$Pts.MP)

## [1] 0.812418

cor(serie$TotDist, serie$Pts.MP)

## [1] 0.3349027

multi\_model <- lm( Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = all)  
summary(multi\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA +   
## Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct +   
## Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = all)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.06204 -0.16652 0.01007 0.16736 0.91965   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.014e+00 7.982e-01 1.270 0.204907   
## Total\_Cmp\_Pct 3.798e-02 2.764e-02 1.374 0.170165   
## Ent\_Final\_Third 7.020e-04 1.838e-04 3.819 0.000156 \*\*\*  
## Ent\_Opposing\_PA 1.095e-03 4.936e-04 2.218 0.027137 \*   
## Cross\_into\_PA -1.157e-03 9.662e-04 -1.197 0.231866   
## PrgDist -1.270e-05 2.446e-06 -5.192 3.39e-07 \*\*\*  
## Short\_Cmp\_Pct -1.278e-03 1.721e-02 -0.074 0.940848   
## Medium\_Cmp\_Pct -4.933e-02 1.966e-02 -2.509 0.012515 \*   
## Long\_Cmp\_Pct 2.619e-02 7.541e-03 3.473 0.000575 \*\*\*  
## Prog\_10\_plus\_yards 1.576e-04 2.469e-04 0.638 0.523638   
## KP 6.044e-04 4.743e-04 1.274 0.203317   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2901 on 381 degrees of freedom  
## Multiple R-squared: 0.6099, Adjusted R-squared: 0.5997   
## F-statistic: 59.57 on 10 and 381 DF, p-value: < 2.2e-16

intercept\_only <- lm(Pts.MP ~ 1, data=all)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = all)  
step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(step\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ Long\_Cmp\_Pct + Ent\_Opposing\_PA + PrgDist +   
## Ent\_Final\_Third + Prog\_10\_plus\_yards, data = all)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.03153 -0.17631 0.00429 0.17558 0.87569   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.430e-01 1.765e-01 -1.377 0.1695   
## Long\_Cmp\_Pct 2.495e-02 2.790e-03 8.943 < 2e-16 \*\*\*  
## Ent\_Opposing\_PA 1.058e-03 4.348e-04 2.434 0.0154 \*   
## PrgDist -1.480e-05 2.132e-06 -6.944 1.63e-11 \*\*\*  
## Ent\_Final\_Third 8.432e-04 1.760e-04 4.792 2.36e-06 \*\*\*  
## Prog\_10\_plus\_yards 2.965e-04 2.073e-04 1.430 0.1535   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2917 on 386 degrees of freedom  
## Multiple R-squared: 0.6005, Adjusted R-squared: 0.5953   
## F-statistic: 116 on 5 and 386 DF, p-value: < 2.2e-16

all$pred\_Pts.MP <- predict(step\_model, all)  
all$error <- all$Pts.MP - all$pred\_Pts.MP

intercept\_only <- lm(Pts.MP ~ 1, data=bund)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = bund)  
bund\_step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(bund\_step\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ PrgDist + Ent\_Opposing\_PA, data = bund)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.65394 -0.14919 -0.00399 0.15507 0.80295   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.207e-01 3.631e-01 -2.536 0.01348 \*   
## PrgDist 1.669e-05 5.730e-06 2.913 0.00481 \*\*  
## Ent\_Opposing\_PA 2.640e-03 8.198e-04 3.221 0.00195 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2591 on 69 degrees of freedom  
## Multiple R-squared: 0.6769, Adjusted R-squared: 0.6676   
## F-statistic: 72.29 on 2 and 69 DF, p-value: < 2.2e-16

bund$pred\_Pts.MP <- predict(bund\_step\_model, bund)  
bund$error <- bund$Pts.MP - bund$pred\_Pts.MP

intercept\_only <- lm(Pts.MP ~ 1, data=liga)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = liga)  
liga\_step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(liga\_step\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ Cross\_into\_PA + Ent\_Opposing\_PA + Ent\_Final\_Third,   
## data = liga)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.75315 -0.19345 0.01924 0.18810 0.57510   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.4707648 0.1782245 2.641 0.01002 \*   
## Cross\_into\_PA -0.0040574 0.0014623 -2.775 0.00695 \*\*  
## Ent\_Opposing\_PA 0.0025415 0.0009208 2.760 0.00724 \*\*  
## Ent\_Final\_Third 0.0004500 0.0002899 1.552 0.12474   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2817 on 76 degrees of freedom  
## Multiple R-squared: 0.5851, Adjusted R-squared: 0.5688   
## F-statistic: 35.73 on 3 and 76 DF, p-value: 1.639e-14

liga$pred\_Pts.MP <- predict(liga\_step\_model, liga)  
liga$error <- liga$Pts.MP - liga$pred\_Pts.MP

intercept\_only <- lm(Pts.MP ~ 1, data=ligue)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = ligue)  
ligue\_step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(ligue\_step\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ Ent\_Opposing\_PA + Total\_Cmp\_Pct + PrgDist +   
## KP + Ent\_Final\_Third, data = ligue)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.69615 -0.12899 -0.01421 0.13729 0.54989   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.559e+00 6.731e-01 -3.802 0.000293 \*\*\*  
## Ent\_Opposing\_PA 1.409e-03 8.104e-04 1.738 0.086283 .   
## Total\_Cmp\_Pct 4.929e-02 8.489e-03 5.807 1.49e-07 \*\*\*  
## PrgDist -2.016e-05 4.245e-06 -4.748 9.80e-06 \*\*\*  
## KP 2.690e-03 9.553e-04 2.816 0.006232 \*\*   
## Ent\_Final\_Third 7.029e-04 2.994e-04 2.348 0.021571 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2514 on 74 degrees of freedom  
## Multiple R-squared: 0.6818, Adjusted R-squared: 0.6603   
## F-statistic: 31.72 on 5 and 74 DF, p-value: < 2.2e-16

ligue$pred\_Pts.MP <- predict(ligue\_step\_model, ligue)  
ligue$error <- ligue$Pts.MP - ligue$pred\_Pts.MP

intercept\_only <- lm(Pts.MP ~ 1, data=epl)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = epl)  
epl\_step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(epl\_step\_model)

##   
## Call:  
## lm(formula = Pts.MP ~ Ent\_Final\_Third, data = epl)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.82842 -0.16659 0.02317 0.15018 0.61119   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -5.655e-02 1.141e-01 -0.495 0.622   
## Ent\_Final\_Third 1.278e-03 9.741e-05 13.119 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2725 on 78 degrees of freedom  
## Multiple R-squared: 0.6881, Adjusted R-squared: 0.6841   
## F-statistic: 172.1 on 1 and 78 DF, p-value: < 2.2e-16

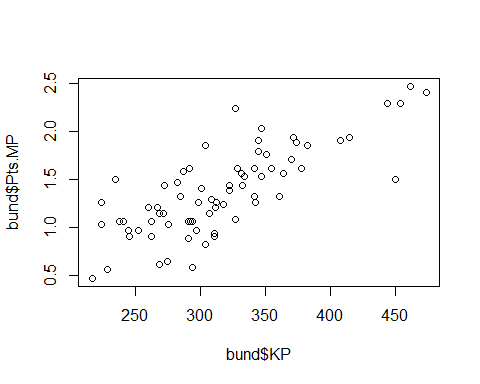
epl$pred\_Pts.MP <- predict(epl\_step\_model, epl)  
epl$error <- epl$Pts.MP - epl$pred\_Pts.MP

intercept\_only <- lm(Pts.MP ~ 1, data=serie)  
all\_predictors <- lm(Pts.MP ~ Total\_Cmp\_Pct + Ent\_Final\_Third + Ent\_Opposing\_PA + Cross\_into\_PA + PrgDist + Short\_Cmp\_Pct + Medium\_Cmp\_Pct + Long\_Cmp\_Pct + Prog\_10\_plus\_yards + KP, data = serie)  
serie\_step\_model <- step(intercept\_only, direction = "both", scope=formula(all\_predictors), trace=0)  
summary(serie\_step\_model)

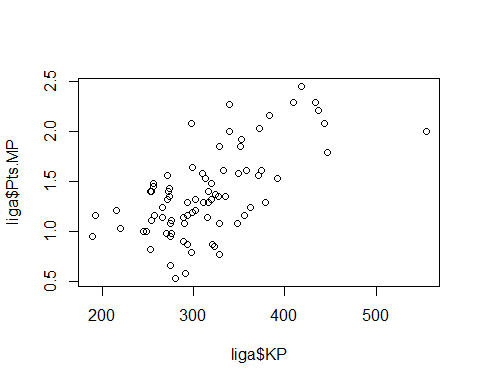
##   
## Call:  
## lm(formula = Pts.MP ~ Long\_Cmp\_Pct + Ent\_Final\_Third + PrgDist +   
## Ent\_Opposing\_PA, data = serie)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.16831 -0.14064 0.02565 0.16763 0.79838   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -9.577e-01 4.864e-01 -1.969 0.05266 .   
## Long\_Cmp\_Pct 3.668e-02 7.484e-03 4.901 5.37e-06 \*\*\*  
## Ent\_Final\_Third 1.105e-03 3.635e-04 3.041 0.00324 \*\*   
## PrgDist -1.669e-05 4.530e-06 -3.685 0.00043 \*\*\*  
## Ent\_Opposing\_PA 1.693e-03 9.977e-04 1.697 0.09384 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.3221 on 75 degrees of freedom  
## Multiple R-squared: 0.612, Adjusted R-squared: 0.5913   
## F-statistic: 29.58 on 4 and 75 DF, p-value: 9.108e-15

serie$pred\_Pts.MP <- predict(serie\_step\_model, serie)  
serie$error <- serie$Pts.MP - serie$pred\_Pts.MP

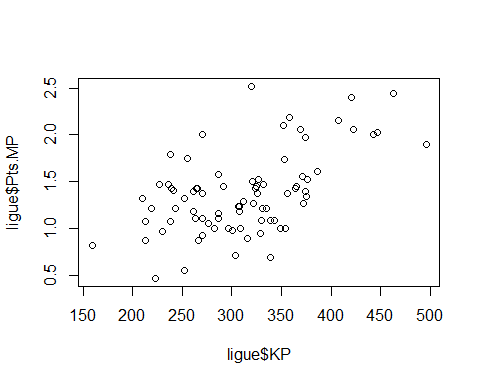
plot(bund$KP, bund$Pts.MP)



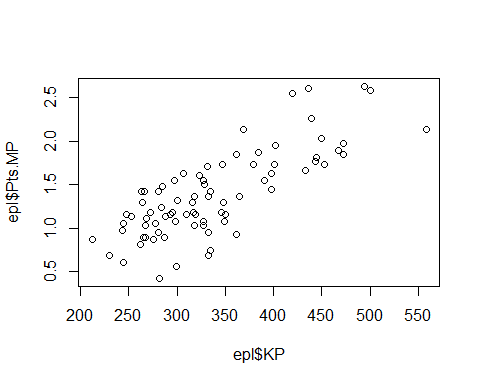
plot(liga$KP, liga$Pts.MP)



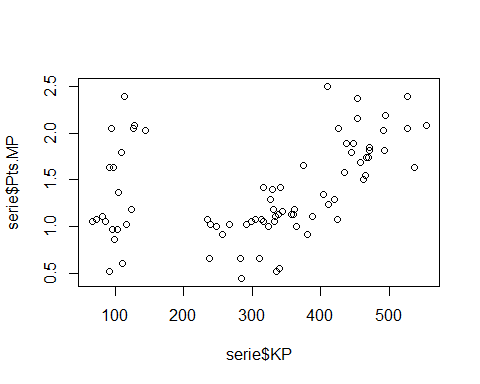
plot(ligue$KP, ligue$Pts.MP)



plot(epl$KP, epl$Pts.MP)



plot(serie$KP, serie$Pts.MP)



cor(bund$KP, bund$Pts.MP)

## [1] 0.7752488

cor(liga$KP, liga$Pts.MP)

## [1] 0.6386289

cor(ligue$KP, ligue$Pts.MP)

## [1] 0.5505307

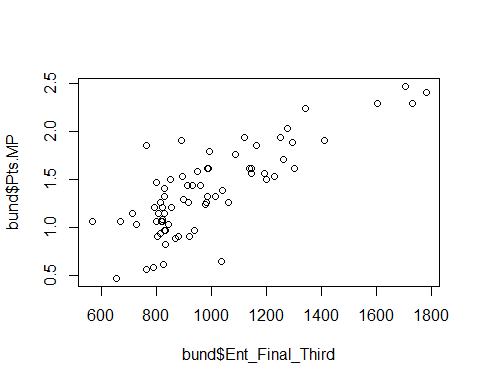
cor(epl$KP, epl$Pts.MP)

## [1] 0.7794895

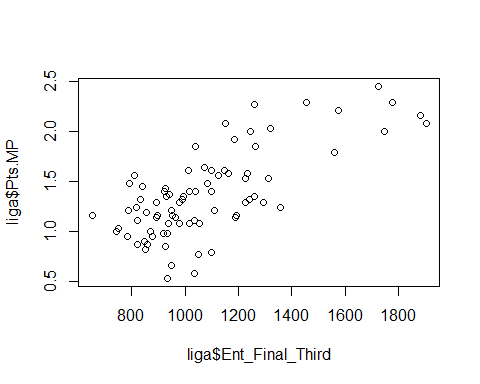
cor(serie$KP, serie$Pts.MP)

## [1] 0.3638758

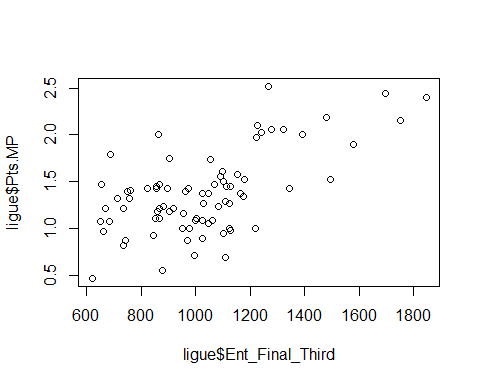
plot(bund$Ent\_Final\_Third, bund$Pts.MP)



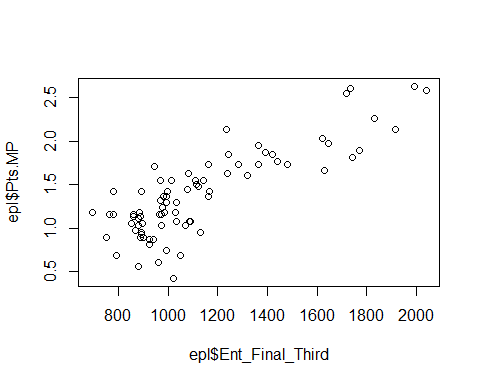
plot(liga$Ent\_Final\_Third, liga$Pts.MP)



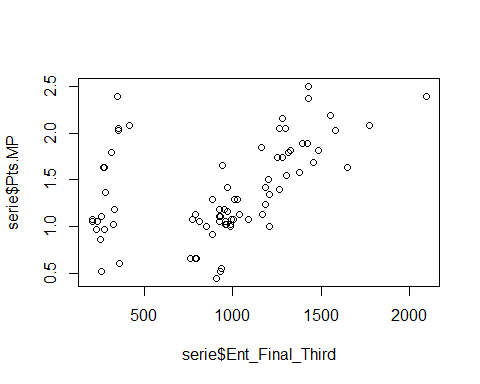
plot(ligue$Ent\_Final\_Third, ligue$Pts.MP)



plot(epl$Ent\_Final\_Third, epl$Pts.MP)



plot(serie$Ent\_Final\_Third, serie$Pts.MP)



cor(bund$Ent\_Final\_Third, bund$Pts.MP)

## [1] 0.7992186

cor(liga$Ent\_Final\_Third, liga$Pts.MP)

## [1] 0.7193965

cor(ligue$Ent\_Final\_Third, ligue$Pts.MP)

## [1] 0.614141

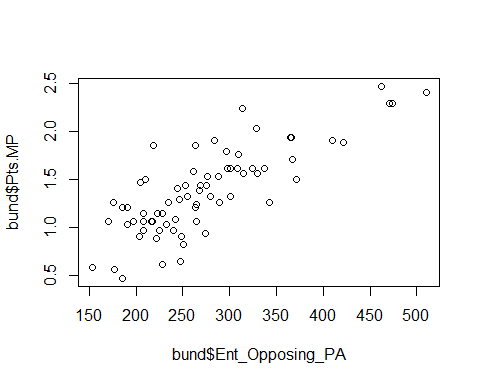
cor(epl$Ent\_Final\_Third, epl$Pts.MP)

## [1] 0.8295344

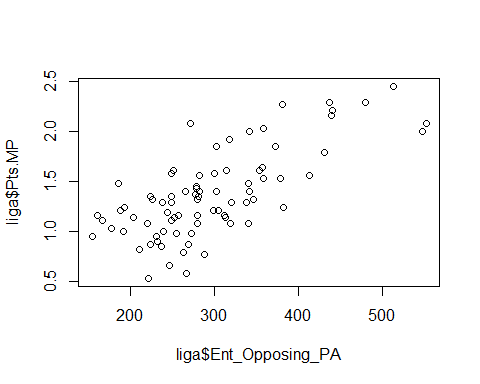
cor(serie$Ent\_Final\_Third, serie$Pts.MP)

## [1] 0.4015767

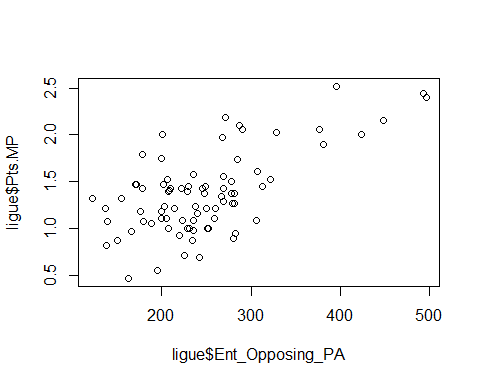
plot(bund$Ent\_Opposing\_PA, bund$Pts.MP)



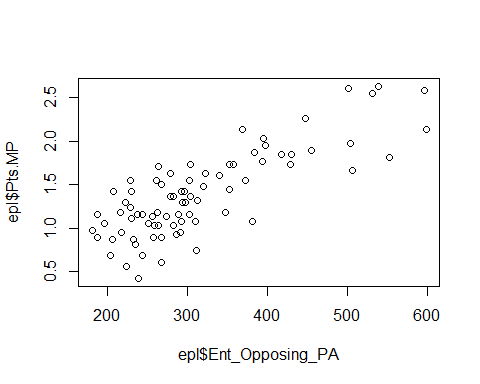
plot(liga$Ent\_Opposing\_PA, liga$Pts.MP)



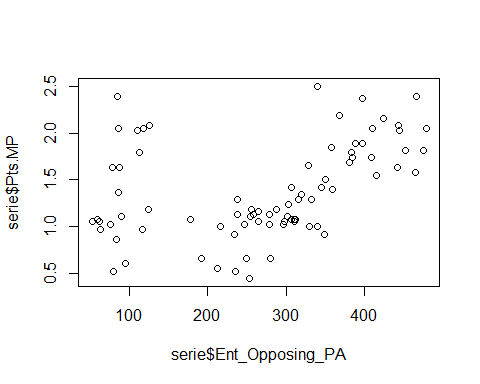
plot(ligue$Ent\_Opposing\_PA, ligue$Pts.MP)



plot(epl$Ent\_Opposing\_PA, epl$Pts.MP)



plot(serie$Ent\_Opposing\_PA, serie$Pts.MP)



cor(bund$Ent\_Opposing\_PA, bund$Pts.MP)

## [1] 0.7982419

cor(liga$Ent\_Opposing\_PA, liga$Pts.MP)

## [1] 0.7201683

cor(ligue$Ent\_Opposing\_PA, ligue$Pts.MP)

## [1] 0.6435303

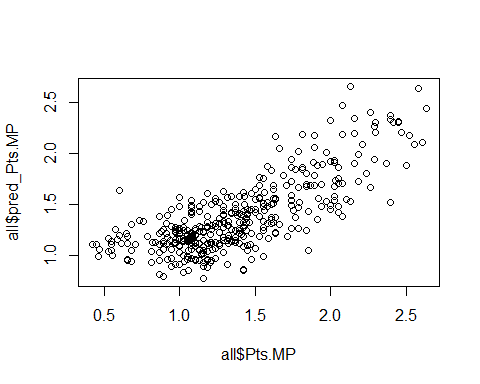
cor(epl$Ent\_Opposing\_PA, epl$Pts.MP)

## [1] 0.8012852

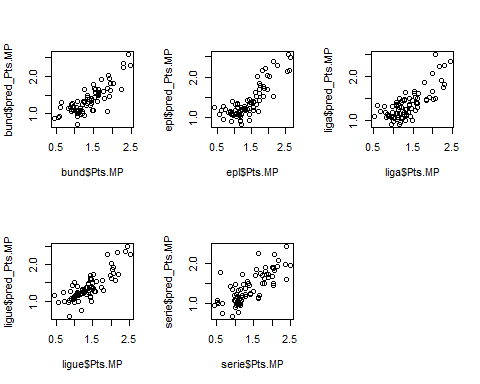
cor(serie$Ent\_Opposing\_PA, serie$Pts.MP)

## [1] 0.3831452

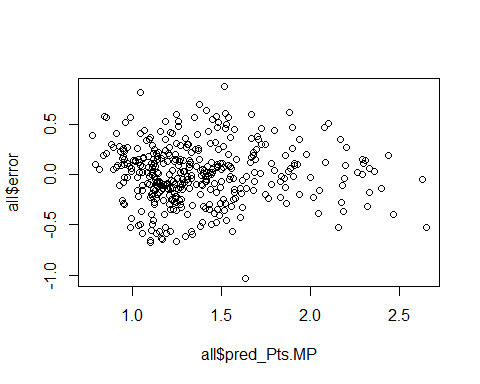
plot(all$Pts.MP, all$pred\_Pts.MP)



par(mfrow=c(2,3))  
plot(bund$Pts.MP, bund$pred\_Pts.MP)  
plot(epl$Pts.MP, epl$pred\_Pts.MP)  
plot(liga$Pts.MP, liga$pred\_Pts.MP)  
plot(ligue$Pts.MP, ligue$pred\_Pts.MP)  
plot(serie$Pts.MP, serie$pred\_Pts.MP)



plot(all$pred\_Pts.MP, all$error)



par(mfrow=c(2,3))  
plot(bund$pred\_Pts.MP, bund$error)  
plot(epl$pred\_Pts.MP, epl$error)  
plot(liga$pred\_Pts.MP, liga$error)  
plot(ligue$pred\_Pts.MP, ligue$error)  
plot(serie$pred\_Pts.MP, serie$error)

