NHL Contracts

Engel, Alec

8/10/2021

team\_colors = c(ANA = '#F47A38', ARI = '#8C2633', BOS = '#FFB81C', BUF = '#002654',  
CAR = '#CC0000', CBJ = '#002654', CGY = '#F1BE48', CHI = '#CF0A2C',  
COL = '#6F263D', DAL = '#006847', DET = '#CE1126', EDM = '#041E42',  
FLA = '#041E42', L.A = '#111111', MIN = '#154734', MTL = '#AF1E2D',  
N.J = '#CE1126', NSH = '#FFB81C', NYI = '#00539B', NYR = '#0038A8',  
OTT = '#C2912C', PHI = '#F74902', PIT = '#FCB514', S.J = '#006D75',  
STL = '#002F87', T.B = '#002868', TOR = '#00205B', VAN = '#00205B',  
VGK = '#B4975A', WPG = '#041E42', WSH = '#041E42')

positionmoney <- players %>% group\_by(Pos) %>%  
 summarise(NetSalary = mean(Salary)) %>%  
 arrange(desc(NetSalary))

## `summarise()` ungrouping output (override with `.groups` argument)

positionmoney$NetSalary <- currency(positionmoney$NetSalary, digits=0L)  
positionmoney <- rename(positionmoney,Position=Pos,AverageSalary=NetSalary)  
positionmoney

## # A tibble: 4 x 2  
## Position AverageSalary  
## <chr> <formttbl>   
## 1 L $2,641,398   
## 2 R $2,569,533   
## 3 D $2,507,515   
## 4 C $2,458,399

agemoney <- players %>% group\_by(Age) %>%  
 summarise(PlayerAge = sum(Salary)) %>%  
 arrange(desc(PlayerAge))

## `summarise()` ungrouping output (override with `.groups` argument)

agemoney$PlayerAge <- currency(agemoney$PlayerAge, digits=0L)  
agemoney <- rename(agemoney,TotalSalary=PlayerAge)  
agemoney <- head(agemoney,10)  
agemoney

## # A tibble: 10 x 2  
## Age TotalSalary   
## <dbl> <formttbl>   
## 1 29 $205,500,000  
## 2 24 $195,425,027  
## 3 28 $178,400,000  
## 4 26 $176,720,800  
## 5 30 $173,820,000  
## 6 31 $163,175,000  
## 7 25 $155,882,675  
## 8 27 $143,827,500  
## 9 32 $119,160,000  
## 10 23 $105,006,625

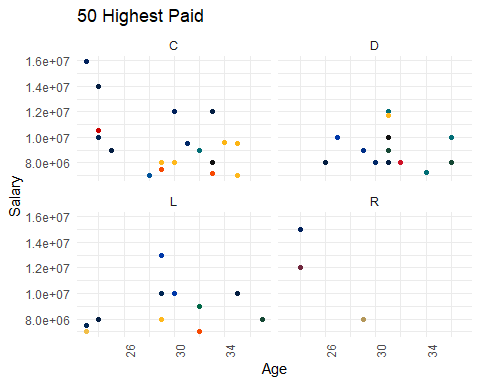
agemoney <- agemoney %>% arrange(TotalSalary)  
agemoney

## # A tibble: 10 x 2  
## Age TotalSalary   
## <dbl> <formttbl>   
## 1 23 $105,006,625  
## 2 32 $119,160,000  
## 3 27 $143,827,500  
## 4 25 $155,882,675  
## 5 31 $163,175,000  
## 6 30 $173,820,000  
## 7 26 $176,720,800  
## 8 28 $178,400,000  
## 9 24 $195,425,027  
## 10 29 $205,500,000

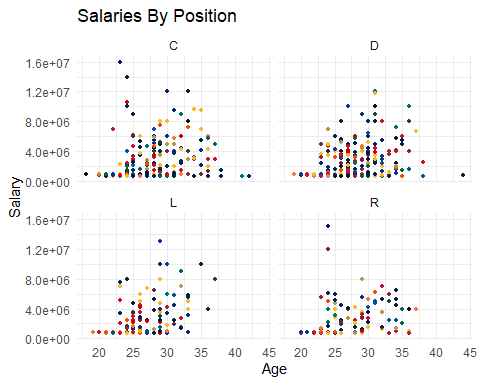
topmoney <- players %>% select(Name,Pos,Age,Team,Salary) %>% arrange(desc(Salary))  
topmoney$Salary <- currency(topmoney$Salary, digits=0L)  
topmoney <- head(topmoney,50)  
topmoney

## # A tibble: 50 x 5  
## Name Pos Age Team Salary   
## <chr> <chr> <dbl> <chr> <formttbl>   
## 1 AUSTON MATTHEWS C 23 TOR $15,900,000  
## 2 MITCHELL MARNER R 24 TOR $15,000,000  
## 3 CONNOR MCDAVID C 24 EDM $14,000,000  
## 4 ARTEMI PANARIN L 29 NYR $13,000,000  
## 5 ERIK KARLSSON D 31 S.J $12,000,000  
## 6 JOHN TAVARES C 30 TOR $12,000,000  
## 7 MIKKO RANTANEN R 24 COL $12,000,000  
## 8 NICKLAS BACKSTROM C 33 WSH $12,000,000  
## 9 ROMAN JOSI D 31 NSH $11,750,000  
## 10 SEBASTIAN AHO C 24 CAR $10,570,000  
## # ... with 40 more rows

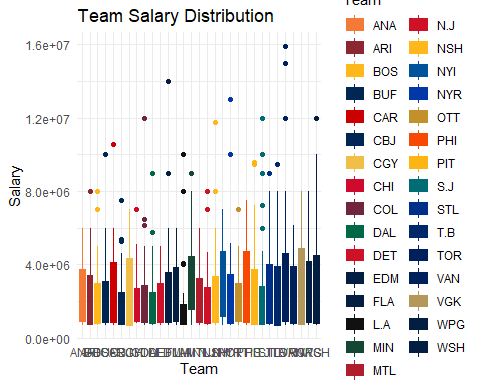
topmoney %>%  
 filter(!(Salary %in% c("$15,900,000", "$15,000,000", "$14,000,000",   
"$13,000,000", "$12,000,000", "$11,750,000", "$10,570,000", "$10,000,000", "$9,600,000",   
"$9,500,000", "$9,000,000", "$8,000,000", "$7,500,000", "$7,250,000", "$7,200,000",   
"$7,000,000")) & !is.na(Salary)) %>%  
 ggplot() +  
 aes(x = Age, y = Salary, Color=Team) +  
 geom\_point(aes(colour = Team)) + scale\_fill\_manual(values=team\_colors) +  
 scale\_color\_manual(values = team\_colors) +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 labs(title = "50 Highest Paid") +  
 facet\_wrap(vars(Pos)) +  
 theme(legend.position = "none")



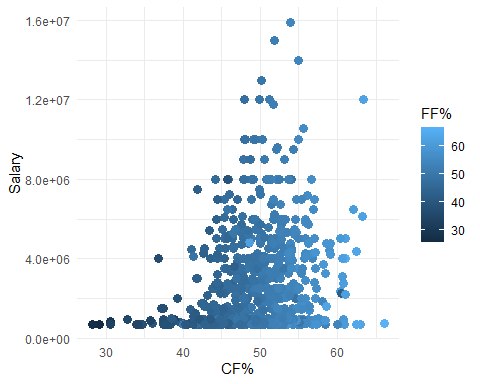
ggplot(players) +  
 aes(x = Age, y = Salary, color = Team) +   
 geom\_point(size = 1L) +   
 scale\_color\_manual(values = team\_colors) +  
 theme\_minimal() +  
 facet\_wrap(vars(Pos)) +  
 labs(title = "Salaries By Position") +  
 ylim(700000L, 16000000L) +  
 theme(legend.position = "none")



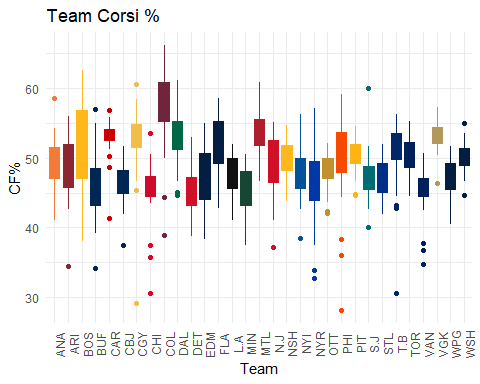
ggplot(players) +  
 aes(x = Team, y = Salary) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team Salary Distribution") +  
theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none") +  
 theme\_minimal()



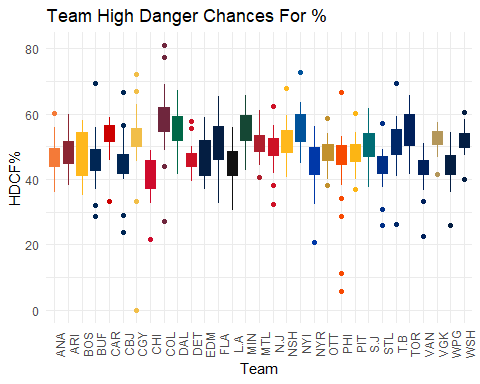
ggplot(players) +  
 aes(x = `CF%`, y = Salary, colour = `FF%`) +  
 geom\_point(size = 3L) +  
 scale\_color\_gradient() +  
 theme\_minimal()



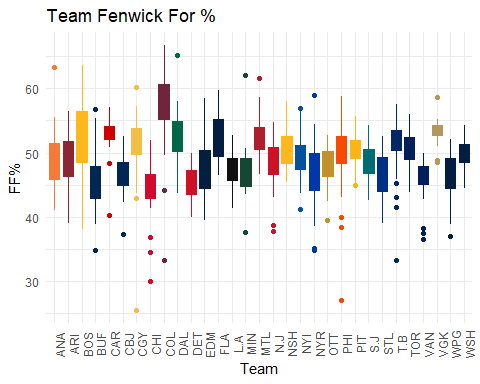
plot1 <- ggplot(players) +  
 aes(x = Team, y = `CF%`) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team Corsi %") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none")  
plot2 <- ggplot(players) +  
 aes(x = Team, y = `HDCF%`) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team High Danger Chances For %") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none")  
plot3 <- ggplot(players) +  
 aes(x = Team, y = `FF%`) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team Fenwick For %") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none")  
plot4 <- ggplot(players) +  
 aes(x = Team, y = `PDO`) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team PDO") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none")  
plot5 <- ggplot(players) +  
 aes(x = Team, y = `SCF%`) +  
 geom\_boxplot(aes(fill = Team,color=Team)) + scale\_fill\_manual(values=team\_colors) + scale\_color\_manual(values = team\_colors) +  
 labs(title = "Team Scoring Chances For %") +  
 theme\_minimal() +  
 theme(axis.text.x = element\_text(angle = 90)) +   
 theme(legend.position = "none")  
plot1



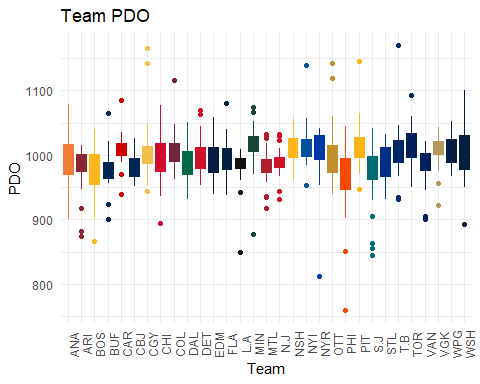
plot2



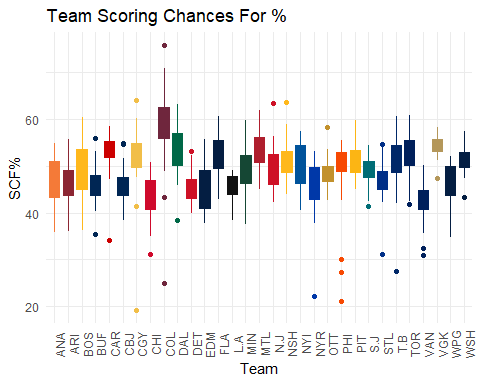
plot3



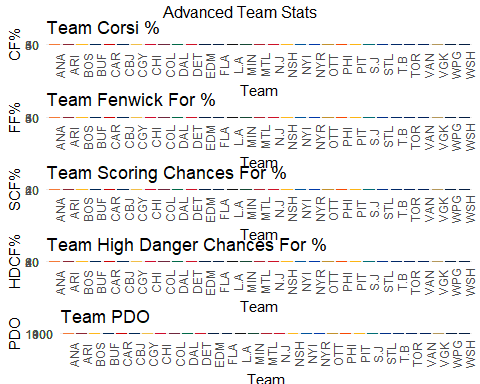
plot4



plot5



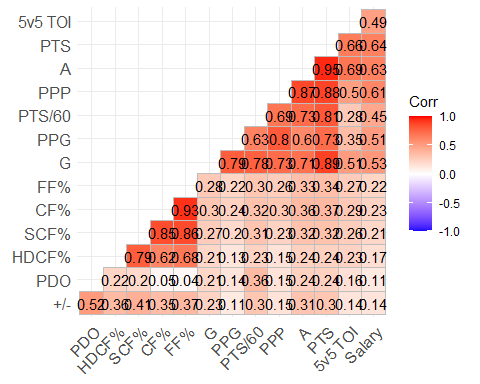
#divide up by top and bottom players for salaries and corsi?  
grid.arrange(plot1,plot3,plot5,plot2,plot4,nrow=5,top="Advanced Team Stats")



playermodel <- players %>%  
 select("G","A","PTS","+/-","PPG","PPP","5v5 TOI","PTS/60","PDO","CF%", "FF%","SCF%","HDCF%","Salary")  
corr <- round(cor(playermodel),2)  
head(corr[,1:8])

## G A PTS +/- PPG PPP 5v5 TOI PTS/60  
## G 1.00 0.71 0.89 0.23 0.79 0.73 0.51 0.78  
## A 0.71 1.00 0.95 0.31 0.60 0.87 0.69 0.73  
## PTS 0.89 0.95 1.00 0.30 0.73 0.88 0.66 0.81  
## +/- 0.23 0.31 0.30 1.00 0.11 0.15 0.14 0.30  
## PPG 0.79 0.60 0.73 0.11 1.00 0.80 0.35 0.63  
## PPP 0.73 0.87 0.88 0.15 0.80 1.00 0.50 0.69

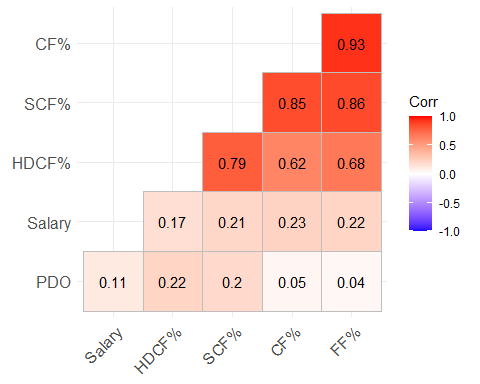
ggcorrplot(corr, hc.order = TRUE, type = "lower", lab = TRUE)



advstatmodel <- players %>%  
 select("PDO","CF%", "FF%","SCF%","HDCF%","Salary")  
corr2 <- round(cor(advstatmodel),2)  
head(corr[,1:6])

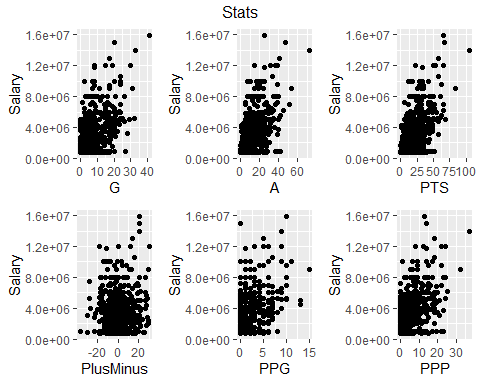
## G A PTS +/- PPG PPP  
## G 1.00 0.71 0.89 0.23 0.79 0.73  
## A 0.71 1.00 0.95 0.31 0.60 0.87  
## PTS 0.89 0.95 1.00 0.30 0.73 0.88  
## +/- 0.23 0.31 0.30 1.00 0.11 0.15  
## PPG 0.79 0.60 0.73 0.11 1.00 0.80  
## PPP 0.73 0.87 0.88 0.15 0.80 1.00

ggcorrplot(corr2, hc.order = TRUE, type = "lower", lab = TRUE)

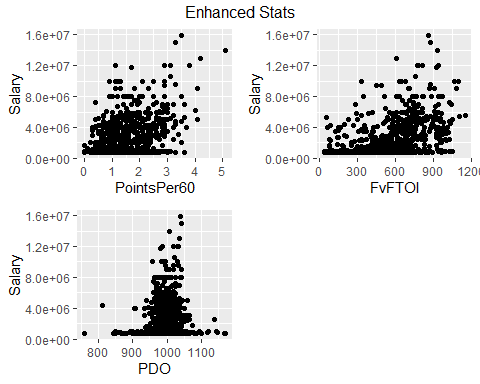


“5v5 TOI”,,“PDO”,“CF%”, “FF%”,“SCF%”,“HDCF%”,“Salary”

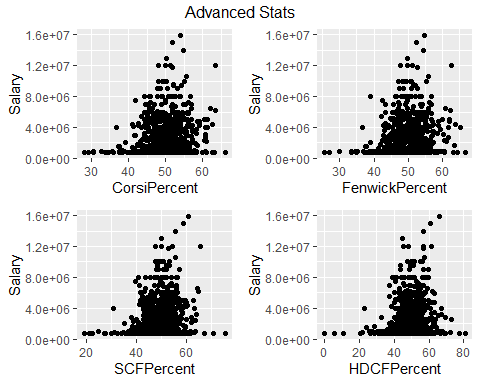
G <- qplot(G,Salary,data = playermodel)  
A <- qplot(A,Salary, data = playermodel)  
PTS <- qplot(PTS,Salary, data = playermodel)  
names(playermodel)[names(playermodel) == "+/-"] <- "PlusMinus"  
PlusMinus <- qplot(PlusMinus,Salary, data = playermodel)  
PPG <- qplot(PPG,Salary, data = playermodel)  
PPP <- qplot(PPP,Salary, data = playermodel)  
  
names(playermodel)[names(playermodel) == "PTS/60"] <- "PointsPer60"  
PointsPer60 <- qplot(PointsPer60,Salary, data = playermodel)  
names(playermodel)[names(playermodel) == "5v5 TOI"] <- "FvFTOI"  
FvFTOI <- qplot(FvFTOI,Salary,data = playermodel)  
PDO <- qplot(PDO,Salary,data = playermodel)  
names(playermodel)[names(playermodel) == "CF%"] <- "CorsiPercent"  
CorsiPercent <- qplot(CorsiPercent,Salary,data = playermodel)  
names(playermodel)[names(playermodel) == "FF%"] <- "FenwickPercent"  
FenwickPercent <- qplot(FenwickPercent,Salary,data = playermodel)  
names(playermodel)[names(playermodel) == "SCF%"] <- "SCFPercent"  
SCFPercent <- qplot(SCFPercent,Salary,data = playermodel)  
names(playermodel)[names(playermodel) == "HDCF%"] <- "HDCFPercent"  
HDCFPercent <- qplot(HDCFPercent,Salary,data = playermodel)  
  
grid.arrange(G,A,PTS,PlusMinus,PPG,PPP,nrow=2,top="Stats")



grid.arrange(PointsPer60,FvFTOI,PDO,nrow=2,top="Enhanced Stats")



grid.arrange(CorsiPercent,FenwickPercent,SCFPercent,HDCFPercent,nrow=2,top="Advanced Stats")



model1 = lm(Salary ~ G+A+PlusMinus+PointsPer60,playermodel)  
model2 = lm(Salary ~ PointsPer60+FvFTOI+PDO,playermodel)  
model3 = lm(Salary ~ CorsiPercent+FenwickPercent+SCFPercent+HDCFPercent,playermodel)  
  
summary(model1)

##   
## Call:  
## lm(formula = Salary ~ G + A + PlusMinus + PointsPer60, data = playermodel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5440329 -948373 -306811 559693 9151161   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1062036 138256 7.682 4.54e-14 \*\*\*  
## G 96886 16811 5.763 1.17e-08 \*\*\*  
## A 143851 10410 13.819 < 2e-16 \*\*\*  
## PlusMinus -14655 7368 -1.989 0.04705 \*   
## PointsPer60 -449724 136052 -3.306 0.00099 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1846000 on 808 degrees of freedom  
## Multiple R-squared: 0.4244, Adjusted R-squared: 0.4216   
## F-statistic: 149 on 4 and 808 DF, p-value: < 2.2e-16

summary(model2)

##   
## Call:  
## lm(formula = Salary ~ PointsPer60 + FvFTOI + PDO, data = playermodel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4419001 -1226469 -266193 904654 10128205   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4611380.4 1878743.9 2.455 0.0143 \*   
## PointsPer60 1075414.0 91652.3 11.734 <2e-16 \*\*\*  
## FvFTOI 3510.4 259.9 13.509 <2e-16 \*\*\*  
## PDO -5438.1 1932.2 -2.814 0.0050 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1959000 on 809 degrees of freedom  
## Multiple R-squared: 0.3511, Adjusted R-squared: 0.3487   
## F-statistic: 145.9 on 3 and 809 DF, p-value: < 2.2e-16

summary(model3)

##   
## Call:  
## lm(formula = Salary ~ CorsiPercent + FenwickPercent + SCFPercent +   
## HDCFPercent, data = playermodel)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -3816872 -1577256 -865748 1080568 12648114   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2924479 841384 -3.476 0.000537 \*\*\*  
## CorsiPercent 94904 47581 1.995 0.046424 \*   
## FenwickPercent -8452 49187 -0.172 0.863615   
## SCFPercent 17343 33558 0.517 0.605428   
## HDCFPercent 7129 17256 0.413 0.679639   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2365000 on 808 degrees of freedom  
## Multiple R-squared: 0.05467, Adjusted R-squared: 0.04999   
## F-statistic: 11.68 on 4 and 808 DF, p-value: 3.151e-09