MATH232FinalProject

2025-04-26

#https://jaseziv.github.io/worldfootballR/articles/extract-fbref-data.html#get-player-scouting-report  
  
library(worldfootballR)  
  
pull\_team\_data <- function(team\_url){  
 team\_standard <- na.omit(fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "standard"))  
 fullgames\_player <- (team\_standard["Min\_Playing\_Time"] / 90)  
 fullgames\_player <- as.vector(fullgames\_player)  
 #head(team\_standard, 10)  
   
 #Player,  
 #Pos,  
 #Age,  
 #Min\_Playing\_Time,  
 #CrdY / games (Not red cards)  
 #npxG\_Expected / games  
 #xAG\_Expected / games  
 #PrgC\_Progression / games  
 #PrgR\_Progression / games  
   
 standard\_cols\_unadj <- c("Player", "Pos", "Age", "Min\_Playing\_Time")  
 standard\_cols\_unadj <- team\_standard[, standard\_cols\_unadj]  
 standard\_cols\_adj <- c("CrdY", "npxG\_Expected", "xAG\_Expected", "PrgC\_Progression", "PrgR\_Progression")  
   
 standard\_cols\_adj <- team\_standard[, standard\_cols\_adj] / fullgames\_player  
 team\_standard\_clean <- cbind(standard\_cols\_unadj, standard\_cols\_adj)  
   
 #head(team\_standard\_clean)  
   
 team\_shooting <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "shooting")  
 #head(team\_shooting, 10)  
   
 #SoT\_percent\_Standard  
 #Sh\_per\_90\_Standard  
 #SoT\_per\_90\_Standard  
 #Dist\_Standard  
 #npxG\_per\_Sh\_Expected NO  
 #np:G\_minus\_xG\_Expected / games  
   
 shooting\_cols\_unadj <- c("SoT\_percent\_Standard", "Sh\_per\_90\_Standard", "SoT\_per\_90\_Standard", "Dist\_Standard", "npxG\_per\_Sh\_Expected")  
 shooting\_cols\_unadj <- team\_shooting[, shooting\_cols\_unadj]  
 shooting\_cols\_unadj[is.na(shooting\_cols\_unadj)] <- 0  
 shooting\_cols\_adj <- c("np:G\_minus\_xG\_Expected")  
   
   
 #shooting\_cols\_adj <- lapply(team\_shooting[, shooting\_cols\_adj], as.numeric) / fullgames\_player  
 #team\_shooting\_clean <- cbind(shooting\_cols\_unadj, shooting\_cols\_adj)  
   
 team\_shooting\_clean <- na.omit(shooting\_cols\_unadj)  
   
 team\_passing <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "passing")  
 #head(team\_passing, 10)  
   
 #TotDist\_Total / games  
 #PrgDist\_Total / games  
 #Cmp\_Short / games  
 #Att\_Short / games  
 #Cmp\_percent\_Short  
 #Cmp\_Medium / games  
 #Att\_Medium / games  
 #Cmp\_percent\_Medium  
 #Cmp\_Long / games  
 #Att\_Long / games  
 #Cmp\_percent\_Long  
 #A\_minus\_xAG\_Expected / games NO  
 #KP / games  
 #Final\_Third / games  
 #PPA / games  
 #CrsPA / games  
 #PrgP / games  
   
   
 passing\_cols\_unadj <- c("Cmp\_percent\_Short", "Cmp\_percent\_Medium", "Cmp\_percent\_Long")  
 passing\_cols\_unadj <- team\_passing[, passing\_cols\_unadj]  
 passing\_cols\_unadj[is.na(passing\_cols\_unadj)] <- 0  
   
   
 passing\_cols\_adj <- c("TotDist\_Total", "PrgDist\_Total", "Cmp\_Short", "Att\_Short", "Cmp\_Medium", "Att\_Medium", "Cmp\_Long", "Att\_Long", "KP", "Final\_Third", "PPA", "CrsPA", "PrgP")  
 passing\_cols\_adj <- na.omit(team\_passing[, passing\_cols\_adj]) / fullgames\_player  
   
 team\_passing\_clean <- na.omit(cbind(passing\_cols\_unadj, passing\_cols\_adj))  
   
   
 team\_passing\_types <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "passing\_types")  
 #head(team\_passing\_types, 10)  
   
 #TB\_Pass\_Types / games  
 #SW\_Pass\_Types / games  
 #Crs\_Pass\_Types / games  
 #TI\_Pass\_types / games  
   
 #head(team\_passing\_types, 10)  
   
 passing\_types\_cols\_adj <- c("TB\_Pass\_Types", "Sw\_Pass\_Types", "Crs\_Pass\_Types", "TI\_Pass\_Types")  
 passing\_types\_cols\_adj <- team\_passing\_types[, passing\_types\_cols\_adj] / fullgames\_player  
 team\_passing\_types\_clean <- passing\_types\_cols\_adj  
   
   
 team\_gca <- na.omit(fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "gca"))  
 #head(team\_gca, 10)  
   
 #SCA90\_SCA  
 #PassLive\_SCA\_Types / games  
 #TO\_SCA\_Types / games  
 #Sh\_SCA\_Types / games  
 #Fld\_SCA\_Types / games  
 #Def\_SCA\_Types / games  
   
 gca\_cols\_unadj <- c("SCA90\_SCA")  
 gca\_cols\_unadj <- team\_gca[, gca\_cols\_unadj]  
 gca\_cols\_adj <- c("PassLive\_SCA\_Types", "TO\_SCA\_Types", "Sh\_SCA\_Types", "Fld\_SCA\_Types", "Def\_SCA\_Types")  
   
 gca\_cols\_adj <- team\_gca[, gca\_cols\_adj] / fullgames\_player  
 team\_gca\_clean <- cbind(gca\_cols\_unadj, gca\_cols\_adj)  
   
 #No GCA because really sparse, high colinearity  
   
 team\_defense <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "defense")  
 #head(team\_defense, 10)  
   
 #TklW\_Tackles / games  
 #"Def 3rd\_Tackles" / games  
 #"Mid 3rd\_Tackles" / games  
 #"Att 3rd\_Tackles" / games  
 #Tkl\_Challenges / games  
 #Att\_Challenges / games  
 #Tkl\_percent\_Challenges  
 #Sh\_Blocks / games  
 #Pass\_Blocks / games  
 #Int / games  
 #Clr / games  
 #Err / games  
   
 defense\_cols\_unadj <- c("Tkl\_percent\_Challenges")  
 defense\_cols\_unadj <- team\_defense[, defense\_cols\_unadj]  
 defense\_cols\_unadj[is.na(defense\_cols\_unadj)] <- 0  
   
 defense\_cols\_adj <- c("TklW\_Tackles", "Def 3rd\_Tackles", "Mid 3rd\_Tackles", "Att 3rd\_Tackles", "Tkl\_Challenges", "Att\_Challenges", "Sh\_Blocks", "Pass\_Blocks", "Int", "Clr", "Err")  
   
 defense\_cols\_adj <- na.omit(team\_defense[, defense\_cols\_adj]) / fullgames\_player  
 team\_defense\_clean <- na.omit(cbind(defense\_cols\_unadj, defense\_cols\_adj))  
   
  
 team\_possession <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "possession")  
 #head(team\_possession, 10)  
   
 #Def Pen\_Touches / games  
 #Def 3rd\_Touches / games  
 #Mid 3rd\_Touches / games  
 #Att 3rd\_Touches / games  
 #Att Pen\_Touches / games  
 #Att\_Take\_Ons / games  
 #Succ\_Take\_Ons / games  
 #Succ\_percent\_Take\_Ons  
 #Carries\_Carries / games  
 #TotDist\_Carries / games  
 #PrgDist\_Carries / games  
 #PrgC\_Carries / games  
 #Final\_Third\_Carries / games  
 #CPA\_Carries / games  
 #Mis\_Carries / games  
 #Rec\_Receiving / games  
 #PrgR\_Receiving / games  
   
 possession\_cols\_unadj <- c("Succ\_percent\_Take\_Ons")  
 possession\_cols\_unadj <- team\_possession[, possession\_cols\_unadj]  
 possession\_cols\_unadj[is.na(possession\_cols\_unadj)] <- 0  
   
 possession\_cols\_adj <- c("Def Pen\_Touches", "Def 3rd\_Touches", "Mid 3rd\_Touches", "Att 3rd\_Touches", "Att Pen\_Touches", "Att\_Take\_Ons", "Succ\_Take\_Ons", "Carries\_Carries", "TotDist\_Carries", "PrgDist\_Carries", "PrgC\_Carries", "Final\_Third\_Carries", "CPA\_Carries", "Mis\_Carries", "Rec\_Receiving", "PrgR\_Receiving")  
   
 possession\_cols\_adj <- na.omit(team\_possession[, possession\_cols\_adj]) / fullgames\_player  
 team\_possession\_clean <- na.omit(cbind(possession\_cols\_unadj, possession\_cols\_adj))  
   
   
   
   
 team\_misc <- fb\_team\_player\_stats(team\_urls = team\_url, stat\_type = "misc")  
 #head(team\_misc, 10)  
   
 #Fls / games  
 #Fld / games  
 #Off / games  
 #Recov / games  
 #Won\_Aerial\_Duels / games  
 #Lost\_Aerial\_Duels / games  
 #Won\_percent\_Aerial\_Duels  
   
 misc\_cols\_unadj <- c("Won\_percent\_Aerial\_Duels")  
 misc\_cols\_unadj <- team\_misc[, misc\_cols\_unadj]  
 misc\_cols\_unadj[is.na(misc\_cols\_unadj)] <- 0  
   
 misc\_cols\_adj <- c("Fls", "Fld", "Off", "Recov", "Won\_Aerial\_Duels", "Lost\_Aerial\_Duels")  
   
 misc\_cols\_adj <- na.omit(team\_misc[, misc\_cols\_adj] / fullgames\_player)  
 team\_misc\_clean <- na.omit(cbind(misc\_cols\_unadj, misc\_cols\_adj))  
   
 final\_result <- cbind(team\_standard\_clean, team\_shooting\_clean, team\_passing\_clean, team\_passing\_types\_clean, team\_gca\_clean, team\_defense\_clean, team\_possession\_clean, team\_misc\_clean)  
 #final\_result <- team\_passing\_clean  
   
 final\_result <- subset(final\_result, final\_result$Pos != "GK")  
 final\_result <- subset(final\_result, final\_result$Min\_Playing\_Time >= 500)  
 final\_result$Pos <- substr(final\_result$Pos, 1, 2)  
   
   
   
   
 return(final\_result)  
   
   
}

#set.seed(123) #reproducibility  
  
liverpool\_2425\_url <- "https://fbref.com/en/squads/822bd0ba/Liverpool-Stats"  
liverpool\_2425\_logs <- pull\_team\_data(liverpool\_2425\_url)  
  
liverpool\_2425\_logs  
  
# Perform PCA  
pca\_result <- prcomp(liverpool\_2425\_logs[ , 5:76], scale. = TRUE)  
pca\_data <- as.data.frame(pca\_result$x[, 1:2]) # Use the first 2 principal components  
pca\_data$cluster <- as.factor(liverpool\_2425\_logs$Pos)  
pca\_data$label <- as.factor(liverpool\_2425\_logs$Player)  
  
  
loadings\_pc1\_pc2 <- pca\_result$rotation[, 1:2]  
# Sort for PC1 by absolute value  
sorted\_pc1 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,1]), decreasing = TRUE), 1]  
# Sort for PC2 by absolute value  
sorted\_pc2 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,2]), decreasing = TRUE), 2]  
cat("Top variables for PC1:\n")  
sorted\_pc1  
cat("\nTop variables for PC2:\n")  
sorted\_pc2  
  
  
# Plot PCA with clusters  
ggplot(pca\_data, aes(x = PC1, y = PC2, color = cluster)) +  
 geom\_point(size = 3) +  
 scale\_x\_continuous(expand = expansion(mult = 0.1)) +  
 scale\_y\_continuous(expand = expansion(mult = 0.1)) +  
 theme\_minimal() +  
 labs(title = "PCA of Liverpool Outfield Players, by Position", x = "PC1", y = "PC2") +   
 geom\_text(aes(label = label), vjust = -1, hjust = 0.5, color = "black", size = 2.5) # Add labels

chelsea\_2425\_url <- "https://fbref.com/en/squads/cff3d9bb/Chelsea-Stats"  
chelsea\_2425\_logs <- pull\_team\_data(chelsea\_2425\_url)  
  
chelsea\_2425\_logs  
  
# Perform PCA  
pca\_result <- prcomp(chelsea\_2425\_logs[ , 5:76], scale. = TRUE)  
pca\_data <- as.data.frame(pca\_result$x[, 1:2]) # Use the first 2 principal components  
pca\_data$cluster <- as.factor(chelsea\_2425\_logs$Pos)  
pca\_data$label <- as.factor(chelsea\_2425\_logs$Player)  
  
  
loadings\_pc1\_pc2 <- pca\_result$rotation[, 1:2]  
# Sort for PC1 by absolute value  
sorted\_pc1 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,1]), decreasing = TRUE), 1]  
# Sort for PC2 by absolute value  
sorted\_pc2 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,2]), decreasing = TRUE), 2]  
cat("Top variables for PC1:\n")  
sorted\_pc1  
cat("\nTop variables for PC2:\n")  
sorted\_pc2  
  
  
# Plot PCA with clusters  
ggplot(pca\_data, aes(x = PC1, y = PC2, color = cluster)) +  
 geom\_point(size = 3) +  
 scale\_x\_continuous(expand = expansion(mult = 0.1)) +  
 scale\_y\_continuous(expand = expansion(mult = 0.1)) +  
 theme\_minimal() +  
 labs(title = "PCA of Chelsea Outfield Players, by Position", x = "PC1", y = "PC2") +   
 geom\_text(aes(label = label), vjust = -1, hjust = 0.5, color = "black", size = 2.5) # Add labels

bmouth\_2425\_url <- "https://fbref.com/en/squads/4ba7cbea/Bournemouth-Stats"  
bmouth\_2425\_logs <- pull\_team\_data(bmouth\_2425\_url)  
  
bmouth\_2425\_logs

## Player Pos Age Min\_Playing\_Time CrdY npxG\_Expected  
## 1 Milos Kerkez DF 21-172 2987 0.12052226 0.01506528  
## 2 Antoine Semenyo FW 25-111 2907 0.27863777 0.28792570  
## 3 Illia Zabarnyi DF 22-239 2752 0.09811047 0.03597384  
## 4 Lewis Cook MF 28-084 2726 0.26412326 0.02971387  
## 6 Ryan Christie MF 30-065 2117 0.38261691 0.07652338  
## 7 Justin Kluivert MF 25-358 2097 0.34334764 0.22317597  
## 8 Evanilson FW 25-204 1972 0.00000000 0.46551724  
## 9 Dean Huijsen DF 20-014 2144 0.33582090 0.07136194  
## 10 Dango Ouattara FW 23-076 1932 0.09316770 0.39130435  
## 11 Tyler Adams MF 26-073 1620 0.38888889 0.08888889  
## 12 Adam Smith DF 33-364 1409 0.38325053 0.04471256  
## 13 Marcus Tavernier FW 26-037 1580 0.28481013 0.29620253  
## 14 Marcos Senesi DF 27-353 1014 0.44378698 0.07100592  
## 15 David Brooks FW 27-294 815 0.33128834 0.24294479  
## 16 Alex Scott MF 21-250 683 0.26354319 0.09224012  
## xAG\_Expected PrgC\_Progression PrgR\_Progression SoT\_percent\_Standard  
## 1 0.10244392 2.8021426 4.9112822 29.4  
## 2 0.17647059 4.0866873 8.0185759 31.4  
## 3 0.01962209 0.7848837 0.1308140 0.0  
## 4 0.13536317 1.0895084 1.5847395 28.6  
## 6 0.14879547 1.4454417 1.8705716 36.4  
## 7 0.20600858 3.0901288 6.2660944 43.1  
## 8 0.07302231 1.1866126 3.9705882 45.2  
## 9 0.08395522 1.5111940 0.4197761 11.8  
## 10 0.24223602 3.5869565 8.7111801 32.3  
## 11 0.05000000 0.6666667 0.5000000 25.0  
## 12 0.01916253 0.7665011 1.9801278 0.0  
## 13 0.15949367 2.2215190 5.8670886 31.3  
## 14 0.01775148 1.0650888 0.4437870 0.0  
## 15 0.18773006 2.7607362 8.2822086 45.5  
## 16 0.10541728 1.8448023 4.2166911 38.5  
## Sh\_per\_90\_Standard SoT\_per\_90\_Standard Dist\_Standard npxG\_per\_Sh\_Expected  
## 1 0.51 0.15 20.4 0.03  
## 2 3.65 1.15 17.2 0.08  
## 3 0.33 0.00 9.8 0.11  
## 4 0.46 0.13 23.4 0.07  
## 6 1.40 0.51 19.6 0.06  
## 7 2.19 0.94 20.3 0.10  
## 8 2.83 1.28 12.9 0.17  
## 9 0.71 0.08 12.3 0.10  
## 10 2.89 0.93 13.5 0.14  
## 11 0.44 0.11 16.1 0.19  
## 12 0.19 0.00 16.9 0.22  
## 13 2.73 0.85 20.5 0.11  
## 14 0.18 0.00 6.0 0.40  
## 15 2.43 1.10 16.8 0.10  
## 16 1.71 0.66 20.0 0.05  
## Cmp\_percent\_Short Cmp\_percent\_Medium Cmp\_percent\_Long TotDist\_Total  
## 1 89.1 77.9 36.5 552.0221  
## 2 84.9 76.6 37.8 322.1672  
## 3 90.1 90.6 52.8 840.3488  
## 4 84.6 84.8 45.4 650.5026  
## 6 88.2 87.9 48.8 632.6783  
## 7 83.4 76.1 54.4 341.2446  
## 8 74.6 70.7 64.3 109.9442  
## 9 89.1 91.4 60.5 973.2090  
## 10 72.5 62.5 37.9 258.4472  
## 11 88.2 87.1 55.8 608.2778  
## 12 88.0 78.1 37.5 452.7466  
## 13 85.9 78.2 43.6 510.4367  
## 14 88.0 85.9 49.7 906.3018  
## 15 84.7 76.2 50.0 413.8896  
## 16 88.7 91.5 59.4 532.4890  
## PrgDist\_Total Cmp\_Short Att\_Short Cmp\_Medium Att\_Medium Cmp\_Long Att\_Long  
## 1 186.80951 19.404084 21.784399 12.926013 16.601942 2.1694007 5.9357215  
## 2 83.74613 13.529412 15.944272 7.182663 9.380805 0.9597523 2.5386997  
## 3 312.41642 10.726744 11.904070 23.284884 25.704942 5.4614826 10.3343023  
## 4 224.30668 16.342627 19.314013 16.771827 19.776229 3.4336023 7.5605282  
## 6 181.48795 15.644780 17.727917 16.069910 18.280586 3.4860652 7.1421823  
## 7 94.42060 12.103004 14.506438 7.253219 9.527897 1.8454936 3.3905579  
## 8 22.68256 5.887424 7.895538 1.871197 2.647059 0.4107505 0.6389452  
## 9 385.31250 15.363806 17.252799 25.900187 28.334888 6.4225746 10.6203358  
## 10 74.72050 9.456522 13.043478 5.357143 8.571429 1.1645963 3.0745342  
## 11 146.61111 19.166667 21.722222 16.555556 19.000000 1.6111111 2.8888889  
## 12 151.51171 13.605394 15.457771 11.625266 14.882896 1.7246274 4.5990064  
## 13 153.51266 14.582278 16.974684 12.474684 15.949367 2.5063291 5.7531646  
## 14 472.10059 15.621302 17.751479 20.591716 23.964497 7.1005917 14.2899408  
## 15 124.67485 15.239264 18.000000 8.834356 11.595092 1.9877301 3.9754601  
## 16 166.29575 16.471449 18.579795 12.781845 13.967789 2.5036603 4.2166911  
## KP Final\_Third PPA CrsPA PrgP TB\_Pass\_Types  
## 1 0.9340475 2.3803147 1.05456980 0.66287245 4.308671 0.06026113  
## 2 1.3622291 1.4860681 1.36222910 0.24767802 3.281734 0.27863777  
## 3 0.3597384 4.2841570 0.19622093 0.00000000 4.087936 0.13081395  
## 4 1.5517241 4.9192957 0.82538518 0.33015407 5.018342 0.09904622  
## 6 1.7430326 5.0165328 1.40292867 0.21256495 5.824280 0.34010392  
## 7 1.9313305 1.7167382 1.37339056 0.25751073 4.077253 0.17167382  
## 8 1.0496957 0.6389452 0.09127789 0.00000000 0.867140 0.18255578  
## 9 0.5876866 5.9608209 0.29384328 0.08395522 4.701493 0.04197761  
## 10 1.4906832 0.9316770 1.11801242 0.46583851 1.909938 0.23291925  
## 11 0.8333333 2.9444444 0.11111111 0.05555556 3.388889 0.11111111  
## 12 0.4471256 2.6188786 0.51100071 0.31937544 2.555004 0.06387509  
## 13 2.0506329 3.4746835 1.65189873 0.51265823 5.639241 0.17088608  
## 14 0.2662722 6.4792899 1.06508876 0.26627219 4.526627 0.00000000  
## 15 1.7668712 1.6564417 1.65644172 0.55214724 2.981595 0.22085890  
## 16 1.3177160 4.8755490 0.52708638 0.00000000 6.325037 0.26354319  
## Sw\_Pass\_Types Crs\_Pass\_Types TI\_Pass\_Types gca\_cols\_unadj PassLive\_SCA\_Types  
## 1 0.06026113 3.6156679 7.74355541 2.26 1.8078339  
## 2 0.15479876 1.6099071 1.26934985 3.78 2.4458204  
## 3 0.16351744 0.1635174 0.03270349 0.92 0.8175872  
## 4 0.06603081 4.1599413 2.24504769 2.84 2.1790169  
## 6 0.38261691 2.0406235 0.12753897 3.62 3.0609353  
## 7 0.25751073 3.3476395 0.38626609 4.03 2.9613734  
## 8 0.13691684 0.0000000 0.00000000 2.56 1.5973631  
## 9 0.62966418 0.5037313 0.20988806 1.60 1.4692164  
## 10 0.23291925 3.9596273 1.21118012 3.17 2.5155280  
## 11 0.00000000 0.1666667 0.22222222 2.00 1.7222222  
## 12 0.12775018 1.9162527 6.57913414 1.28 0.9581263  
## 13 0.39873418 5.0696203 1.31012658 4.67 3.2468354  
## 14 0.08875740 0.8875740 0.44378698 1.42 1.4201183  
## 15 0.00000000 4.8588957 0.99386503 3.87 3.2024540  
## 16 0.52708638 0.6588580 0.13177160 3.55 2.3718887  
## TO\_SCA\_Types Sh\_SCA\_Types Fld\_SCA\_Types Def\_SCA\_Types defense\_cols\_unadj  
## 1 0.03013057 0.06026113 0.15065283 0.09039170 61.1  
## 2 0.58823529 0.43343653 0.15479876 0.06191950 38.1  
## 3 0.00000000 0.00000000 0.00000000 0.06540698 69.8  
## 4 0.00000000 0.06603081 0.00000000 0.00000000 53.8  
## 6 0.00000000 0.12753897 0.08502598 0.04251299 42.2  
## 7 0.25751073 0.21459227 0.21459227 0.04291845 20.0  
## 8 0.13691684 0.41075051 0.41075051 0.00000000 60.0  
## 9 0.00000000 0.04197761 0.04197761 0.00000000 60.0  
## 10 0.09316770 0.46583851 0.04658385 0.04658385 42.5  
## 11 0.00000000 0.05555556 0.00000000 0.22222222 55.6  
## 12 0.06387509 0.00000000 0.12775018 0.06387509 48.0  
## 13 0.17088608 0.34177215 0.22784810 0.00000000 36.0  
## 14 0.00000000 0.00000000 0.00000000 0.00000000 46.7  
## 15 0.22085890 0.11042945 0.00000000 0.00000000 7.7  
## 16 0.00000000 0.26354319 0.39531479 0.39531479 44.7  
## TklW\_Tackles Def 3rd\_Tackles Mid 3rd\_Tackles Att 3rd\_Tackles Tkl\_Challenges  
## 1 0.7833947 0.7231336 0.5724807 0.18078339 0.6628724  
## 2 1.0526316 0.4953560 0.7430341 0.40247678 0.4953560  
## 3 1.0465116 0.8502907 0.6213663 0.03270349 0.9811047  
## 4 1.3536317 1.3206163 1.1225238 0.09904622 1.1555393  
## 6 1.4029287 0.8077468 1.1478507 0.34010392 1.1478507  
## 7 0.3862661 0.1716738 0.4291845 0.30042918 0.1716738  
## 8 0.4107505 0.2738337 0.2281947 0.13691684 0.2738337  
## 9 0.9654851 0.6716418 0.6716418 0.04197761 0.6296642  
## 10 1.1645963 0.6055901 0.5124224 0.46583851 0.7919255  
## 11 2.5555556 1.3333333 2.4444444 0.38888889 1.9444444  
## 12 0.6387509 0.6387509 0.3193754 0.38325053 0.7665011  
## 13 1.3101266 0.7405063 0.9683544 0.51265823 0.5126582  
## 14 1.8639053 1.9526627 0.9763314 0.17751479 1.2426036  
## 15 0.4417178 0.2208589 0.6625767 0.33128834 0.1104294  
## 16 1.5812592 1.3177160 1.0541728 0.92240117 2.2401171  
## Att\_Challenges Sh\_Blocks Pass\_Blocks Int Clr Err  
## 1 1.0847004 0.33143622 0.7532641 1.2052226 3.1034483 0.06026113  
## 2 1.3003096 0.24767802 1.5789474 0.3095975 1.0835913 0.03095975  
## 3 1.4062500 0.65406977 0.4251453 1.1446221 5.1998547 0.06540698  
## 4 2.1460015 0.33015407 0.5942773 1.3206163 2.1129861 0.00000000  
## 6 2.7208314 0.46764289 1.4029287 1.2753897 2.1681625 0.12753897  
## 7 0.8583691 0.00000000 1.0300429 0.5579399 0.9871245 0.04291845  
## 8 0.4563895 0.09127789 0.3651116 0.0000000 0.9127789 0.00000000  
## 9 1.0494403 1.13339552 0.3358209 2.0569030 7.2201493 0.16791045  
## 10 1.8633540 0.41925466 1.1645963 0.5124224 2.3291925 0.00000000  
## 11 3.5000000 0.55555556 1.2777778 1.7222222 2.1111111 0.00000000  
## 12 1.5968772 0.31937544 0.4471256 0.8942512 3.0021292 0.06387509  
## 13 1.4240506 0.17088608 0.6835443 0.2278481 0.6265823 0.00000000  
## 14 2.6627219 0.62130178 0.7100592 0.9763314 4.7928994 0.00000000  
## 15 1.4355828 0.22085890 1.5460123 0.5521472 0.7730061 0.11042945  
## 16 5.0073206 0.39531479 1.5812592 1.3177160 1.7130307 0.00000000  
## possession\_cols\_unadj Def Pen\_Touches Def 3rd\_Touches Mid 3rd\_Touches  
## 1 45.5 3.1335788 14.884499 24.255105  
## 2 52.0 1.3622291 5.603715 16.099071  
## 3 62.5 6.9985465 27.536337 30.087209  
## 4 58.8 2.7072634 13.734409 30.704329  
## 6 47.9 3.1459613 10.883325 32.862541  
## 7 40.2 0.9871245 3.991416 14.034335  
## 8 31.6 1.0040568 2.236308 8.351927  
## 9 40.0 9.6128731 32.994403 34.589552  
## 10 46.8 2.7950311 7.453416 13.881988  
## 11 18.8 2.8888889 16.166667 34.611111  
## 12 40.0 2.4911285 13.860894 21.206529  
## 13 42.3 1.0253165 6.151899 20.392405  
## 14 55.6 5.9467456 30.976331 33.372781  
## 15 26.7 1.3251534 5.852761 18.000000  
## 16 20.0 2.7672035 10.278184 31.229868  
## Att 3rd\_Touches Att Pen\_Touches Att\_Take\_Ons Succ\_Take\_Ons Carries\_Carries  
## 1 19.253432 1.7174422 1.3257449 0.60261132 28.68430  
## 2 25.294118 5.4489164 3.8080495 1.98142415 27.80186  
## 3 2.191134 0.4251453 0.2616279 0.16351744 36.75872  
## 4 14.823918 0.6272927 0.5612619 0.33015407 25.71900  
## 6 15.814832 1.9555975 2.0406235 0.97779877 30.86443  
## 7 24.420601 4.9356223 3.5193133 1.41630901 24.72103  
## 8 11.409736 5.1572008 0.8671400 0.27383367 12.73327  
## 9 5.121269 1.2173507 0.2098881 0.08395522 39.33302  
## 10 22.826087 5.9161491 3.5869565 1.67701863 22.31366  
## 11 10.500000 0.9444444 0.8888889 0.16666667 28.50000  
## 12 10.347764 0.3832505 0.9581263 0.38325053 19.48190  
## 13 29.107595 2.7911392 2.9620253 1.25316456 30.75949  
## 14 5.946746 1.0650888 0.7988166 0.44378698 42.86982  
## 15 26.282209 3.5337423 3.3128834 0.88343558 30.36810  
## 16 16.734993 2.2401171 3.2942899 0.65885798 30.57101  
## TotDist\_Carries PrgDist\_Carries PrgC\_Carries Final\_Third\_Carries CPA\_Carries  
## 1 168.55039 98.46669 2.8021426 1.9584868 0.45195849  
## 2 209.78328 103.93189 4.0866873 1.9814241 1.98142415  
## 3 237.68895 141.50799 0.7848837 0.6540698 0.00000000  
## 4 128.62803 69.39839 1.0895084 0.9244314 0.13206163  
## 6 161.63439 75.37553 1.4454417 1.7005196 0.34010392  
## 7 158.45494 89.78541 3.0901288 2.1459227 1.45922747  
## 8 69.14300 33.49899 1.1866126 0.6845842 0.63894523  
## 9 251.78172 156.36660 1.5111940 1.0074627 0.00000000  
## 10 161.64596 79.47205 3.5869565 1.5372671 1.77018634  
## 11 136.27778 60.00000 0.6666667 0.9444444 0.05555556  
## 12 82.71824 37.36693 0.7665011 0.9581263 0.00000000  
## 13 189.45570 89.37342 2.2215190 2.1075949 0.39873418  
## 14 221.71598 130.02959 1.0650888 1.2426036 0.00000000  
## 15 202.96933 88.56442 2.7607362 2.3190184 0.88343558  
## 16 169.32650 77.21816 1.8448023 1.0541728 0.26354319  
## Mis\_Carries Rec\_Receiving PrgR\_Receiving misc\_cols\_unadj Fls Fld  
## 1 1.2956143 31.09474 4.9112822 35.2 0.8135253 1.3257449  
## 2 2.4148607 31.95046 8.0185759 45.5 2.0123839 1.1145511  
## 3 0.4578488 36.88953 0.1308140 63.6 0.5559593 0.1962209  
## 4 0.9574468 29.31768 1.5847395 54.1 1.5187087 0.3301541  
## 6 1.8705716 33.03259 1.8705716 41.7 1.8280586 1.3179027  
## 7 2.1030043 27.08155 6.2660944 21.9 1.7167382 1.2446352  
## 8 1.4148073 15.38032 3.9705882 40.4 1.4148073 1.9624746  
## 9 0.6716418 44.07649 0.4197761 60.6 1.1753731 0.3777985  
## 10 2.7950311 26.04037 8.7111801 42.2 0.6987578 0.8850932  
## 11 1.5000000 32.38889 0.5000000 63.0 2.2777778 1.0555556  
## 12 0.6387509 21.27040 1.9801278 48.0 0.9581263 0.7026260  
## 13 2.0506329 34.86076 5.8670886 43.6 1.4240506 1.1392405  
## 14 0.5325444 43.75740 0.4437870 46.0 1.6863905 0.6213018  
## 15 2.9815951 33.46012 8.2822086 43.8 0.8834356 1.6564417  
## 16 2.8989751 34.52416 4.2166911 46.2 1.3177160 3.4260615  
## Off Recov Won\_Aerial\_Duels Lost\_Aerial\_Duels  
## 1 0.03013057 4.609977 0.7532641 1.3860060  
## 2 0.43343653 4.551084 1.8575851 2.2291022  
## 3 0.00000000 4.578488 2.2892442 1.3081395  
## 4 0.09904622 5.876742 1.3206163 1.1225238  
## 6 0.04251299 7.737364 0.8502598 1.1903637  
## 7 0.21459227 3.047210 0.3004292 1.0729614  
## 8 0.77586207 2.555781 1.7342799 2.5557809  
## 9 0.08395522 3.903918 2.6445896 1.7210821  
## 10 0.65217391 4.099379 2.0031056 2.7484472  
## 11 0.00000000 5.333333 1.6111111 0.9444444  
## 12 0.00000000 2.746629 0.7665011 0.8303762  
## 13 0.22784810 5.468354 1.3670886 1.7658228  
## 14 0.00000000 5.946746 2.0414201 2.3964497  
## 15 0.00000000 4.858896 0.7730061 0.9938650  
## 16 0.00000000 6.456808 0.7906296 0.9224012

# Perform PCA  
pca\_result <- prcomp(bmouth\_2425\_logs[ , 5:76], scale. = TRUE)  
pca\_data <- as.data.frame(pca\_result$x[, 1:2]) # Use the first 2 principal components  
pca\_data$cluster <- as.factor(bmouth\_2425\_logs$Pos)  
pca\_data$label <- as.factor(bmouth\_2425\_logs$Player)  
  
  
loadings\_pc1\_pc2 <- pca\_result$rotation[, 1:2]  
# Sort for PC1 by absolute value  
sorted\_pc1 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,1]), decreasing = TRUE), 1]  
# Sort for PC2 by absolute value  
sorted\_pc2 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,2]), decreasing = TRUE), 2]  
cat("Top variables for PC1:\n")

## Top variables for PC1:

sorted\_pc1

## Cmp\_Medium Def 3rd\_Touches Att\_Medium   
## -0.168106727 -0.167804020 -0.165840064   
## TotDist\_Total PrgR\_Progression PrgR\_Receiving   
## -0.165586153 0.164282323 0.164282323   
## SoT\_per\_90\_Standard Att Pen\_Touches Sh\_per\_90\_Standard   
## 0.163617078 0.159124242 0.158704203   
## PrgDist\_Total Mid 3rd\_Touches Att 3rd\_Touches   
## -0.156785861 -0.156242310 0.152974744   
## Def Pen\_Touches Clr CPA\_Carries   
## -0.150149829 -0.149930770 0.149259324   
## Cmp\_Long Final\_Third Sh\_SCA\_Types   
## -0.148920639 -0.148119393 0.146242122   
## Att\_Take\_Ons Att\_Long Sh\_Blocks   
## 0.145399241 -0.144912557 -0.144855319   
## Cmp\_percent\_Medium npxG\_Expected Int   
## -0.143881085 0.142853730 -0.142711126   
## SoT\_percent\_Standard Mis\_Carries xAG\_Expected   
## 0.141126677 0.140664186 0.139887483   
## gca\_cols\_unadj Succ\_Take\_Ons PrgC\_Carries   
## 0.139302854 0.138437627 0.134282421   
## PrgC\_Progression KP misc\_cols\_unadj   
## 0.134282421 0.132818051 -0.125085442   
## TO\_SCA\_Types Cmp\_percent\_Short Def 3rd\_Tackles   
## 0.123645426 -0.121960278 -0.120827003   
## PassLive\_SCA\_Types Carries\_Carries Off   
## 0.120505474 -0.118944239 0.117834730   
## TB\_Pass\_Types defense\_cols\_unadj Rec\_Receiving   
## 0.114146261 -0.110527789 -0.109990411   
## Final\_Third\_Carries PPA Crs\_Pass\_Types   
## 0.108588956 0.095263143 0.091986886   
## Tkl\_Challenges PrgDist\_Carries Fld\_SCA\_Types   
## -0.089116548 -0.086302569 0.085888471   
## TklW\_Tackles Pass\_Blocks Fld   
## -0.085562177 0.081160085 0.078956083   
## Won\_Aerial\_Duels Dist\_Standard Cmp\_Short   
## -0.076517763 0.075629222 -0.073828752   
## Att 3rd\_Tackles PrgP CrsPA   
## 0.072278766 -0.069743844 0.068351899   
## Mid 3rd\_Tackles TotDist\_Carries npxG\_per\_Sh\_Expected   
## -0.062558505 -0.061244448 -0.060953645   
## CrdY Att\_Short Att\_Challenges   
## -0.057584535 -0.057456004 -0.050916210   
## Recov Err Lost\_Aerial\_Duels   
## -0.050136497 -0.047313814 0.044468849   
## Cmp\_percent\_Long possession\_cols\_unadj Sw\_Pass\_Types   
## -0.044097672 -0.043167813 -0.019085566   
## Def\_SCA\_Types Fls TI\_Pass\_Types   
## -0.014262721 0.005677751 0.001732765

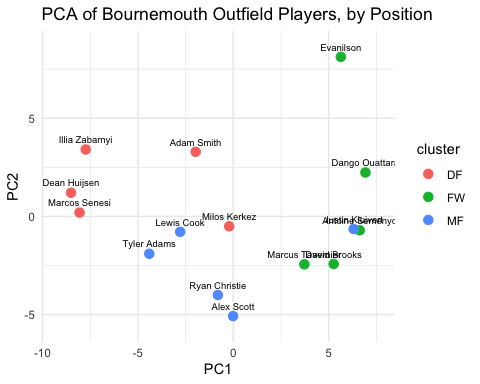
cat("\nTop variables for PC2:\n")

##   
## Top variables for PC2:

sorted\_pc2

## PrgP Recov Att\_Short   
## -0.2319799379 -0.2307632451 -0.2297512503   
## Cmp\_Short Pass\_Blocks Att\_Challenges   
## -0.2244998634 -0.2095028472 -0.1876638641   
## PassLive\_SCA\_Types CrdY Off   
## -0.1866897480 -0.1732490866 0.1694963752   
## Att 3rd\_Tackles Mid 3rd\_Tackles gca\_cols\_unadj   
## -0.1680056213 -0.1633849244 -0.1586849541   
## Dist\_Standard PPA Final\_Third\_Carries   
## -0.1582729654 -0.1559267438 -0.1502098259   
## Rec\_Receiving Cmp\_percent\_Short Lost\_Aerial\_Duels   
## -0.1498420431 -0.1492374780 0.1485722496   
## KP Mis\_Carries Mid 3rd\_Touches   
## -0.1440990947 -0.1421861338 -0.1368770492   
## defense\_cols\_unadj Att\_Take\_Ons TklW\_Tackles   
## 0.1361855313 -0.1348748065 -0.1326349652   
## Def\_SCA\_Types Won\_Aerial\_Duels Carries\_Carries   
## -0.1318567585 0.1315769433 -0.1310734164   
## Tkl\_Challenges Att 3rd\_Touches Cmp\_percent\_Medium   
## -0.1285259368 -0.1274897021 -0.1226529911   
## Final\_Third npxG\_Expected Fls   
## -0.1081715547 0.1078698157 -0.1055725820   
## TB\_Pass\_Types npxG\_per\_Sh\_Expected Fld   
## -0.1042849013 0.1020243499 -0.1015484230   
## xAG\_Expected TotDist\_Carries Crs\_Pass\_Types   
## -0.1007365175 -0.0998163178 -0.0994959542   
## Int Def 3rd\_Tackles SoT\_percent\_Standard   
## -0.0959047649 -0.0913579051 -0.0837207278   
## Succ\_Take\_Ons Sw\_Pass\_Types TotDist\_Total   
## -0.0836769022 -0.0731312595 -0.0724765159   
## Clr possession\_cols\_unadj CrsPA   
## 0.0705831152 0.0676679282 -0.0676678222   
## PrgC\_Progression PrgC\_Carries Att\_Medium   
## -0.0633443020 -0.0633443020 -0.0627412899   
## Cmp\_Medium Att Pen\_Touches Def Pen\_Touches   
## -0.0595086273 0.0490895489 0.0423185185   
## PrgDist\_Carries PrgR\_Receiving PrgR\_Progression   
## -0.0385443370 -0.0385063516 -0.0385063516   
## Cmp\_percent\_Long Sh\_SCA\_Types Fld\_SCA\_Types   
## 0.0344862452 0.0327325461 0.0323816580   
## TI\_Pass\_Types Def 3rd\_Touches Err   
## 0.0294653525 0.0264960060 -0.0263650681   
## Att\_Long misc\_cols\_unadj PrgDist\_Total   
## -0.0247432506 0.0239600323 -0.0236820673   
## Cmp\_Long SoT\_per\_90\_Standard CPA\_Carries   
## -0.0223858043 0.0088226025 0.0055800408   
## Sh\_Blocks Sh\_per\_90\_Standard TO\_SCA\_Types   
## -0.0008078287 -0.0007846066 -0.0006103608

# Plot PCA with clusters  
ggplot(pca\_data, aes(x = PC1, y = PC2, color = cluster)) +  
 geom\_point(size = 3) +  
 scale\_x\_continuous(expand = expansion(mult = 0.1)) +  
 scale\_y\_continuous(expand = expansion(mult = 0.1)) +  
 theme\_minimal() +  
 labs(title = "PCA of Bournemouth Outfield Players, by Position", x = "PC1", y = "PC2") +   
 geom\_text(aes(label = label), vjust = -1, hjust = 0.5, color = "black", size = 2.5) # Add labels



manu\_2425\_url <- "https://fbref.com/en/squads/19538871/Manchester-United-Stats"  
manu\_2425\_logs <- pull\_team\_data(manu\_2425\_url)  
  
manu\_2425\_logs

## Player Pos Age Min\_Playing\_Time CrdY npxG\_Expected  
## 2 Bruno Fernandes MF 30-232 2758 0.13052937 0.23168963  
## 3 Diogo Dalot DF 26-041 2788 0.16140603 0.06456241  
## 4 Noussair Mazraoui DF 27-165 2639 0.06820765 0.02387268  
## 5 Matthijs de Ligt DF 25-259 2089 0.12924844 0.10770704  
## 6 Alejandro Garnacho MF 20-301 2056 0.08754864 0.30642023  
## 7 Lisandro Martínez DF 27-100 1751 0.35979440 0.06681896  
## 8 Rasmus Højlund FW 22-083 1744 0.10321101 0.19610092  
## 9 Manuel Ugarte Ribeiro MF 24-017 1616 0.50123762 0.05569307  
## 10 Amad Diallo FW 22-291 1594 0.16938519 0.23149310  
## 11 Harry Maguire DF 32-054 1483 0.42481457 0.08496291  
## 12 Kobbie Mainoo MF 20-009 1416 0.31779661 0.03813559  
## 13 Casemiro MF 33-064 1335 0.20224719 0.11460674  
## 14 Joshua Zirkzee FW 23-341 1402 0.12838802 0.30813124  
## 15 Marcus Rashford FW 27-179 978 0.18404908 0.15644172  
## 16 Leny Yoro DF 19-166 1069 0.42095416 0.05893358  
## 17 Christian Eriksen MF 33-073 994 0.36217304 0.02716298  
## 18 Patrick Dorgu DF 20-184 585 0.46153846 0.06153846  
## xAG\_Expected PrgC\_Progression PrgR\_Progression SoT\_percent\_Standard  
## 2 0.24800580 2.1211022 3.8506164 30.1  
## 3 0.06456241 2.0014347 5.2941176 21.7  
## 4 0.04433498 1.2618416 1.7392952 23.1  
## 5 0.01292484 0.3015797 0.2584969 50.0  
## 6 0.15758755 5.8657588 11.9941634 34.6  
## 7 0.05653912 1.1821816 0.3597944 35.3  
## 8 0.06708716 1.3417431 4.2832569 40.0  
## 9 0.06126238 0.6683168 1.0581683 16.7  
## 10 0.22020075 5.1944793 8.9774153 32.5  
## 11 0.01820634 0.3641268 0.3641268 41.7  
## 12 0.10169492 1.3347458 2.4152542 15.4  
## 13 0.15505618 0.2696629 1.7528090 28.6  
## 14 0.09629101 0.8987161 4.4293866 39.3  
## 15 0.19325153 1.8404908 7.2699387 56.3  
## 16 0.00000000 1.5996258 0.3367633 20.0  
## 17 0.23541247 0.9054326 2.6257545 11.1  
## 18 0.06153846 3.8461538 8.1538462 22.2  
## Sh\_per\_90\_Standard SoT\_per\_90\_Standard Dist\_Standard npxG\_per\_Sh\_Expected  
## 2 2.71 0.82 21.9 0.09  
## 3 0.74 0.16 15.6 0.09  
## 4 0.44 0.10 20.6 0.05  
## 5 0.78 0.39 7.9 0.14  
## 6 3.55 1.23 17.1 0.09  
## 7 0.87 0.31 15.9 0.08  
## 8 1.29 0.52 12.8 0.15  
## 9 1.00 0.17 22.0 0.05  
## 10 2.26 0.73 17.7 0.10  
## 11 0.73 0.30 7.9 0.12  
## 12 0.83 0.13 20.8 0.04  
## 13 1.89 0.54 18.8 0.06  
## 14 1.80 0.71 13.1 0.17  
## 15 1.47 0.83 16.4 0.11  
## 16 0.42 0.08 8.6 0.15  
## 17 0.81 0.09 24.8 0.03  
## 18 1.38 0.31 16.2 0.04  
## Cmp\_percent\_Short Cmp\_percent\_Medium Cmp\_percent\_Long TotDist\_Total  
## 2 86.7 80.4 52.6 872.6867  
## 3 89.9 81.9 55.4 693.9491  
## 4 91.3 89.1 54.3 717.9879  
## 5 94.1 94.5 50.5 928.3916  
## 6 89.0 82.9 57.4 366.6537  
## 7 94.2 93.0 55.2 1042.9926  
## 8 82.9 71.7 80.0 125.1950  
## 9 89.2 93.3 74.0 652.3329  
## 10 90.4 85.3 52.4 511.0351  
## 11 91.8 91.0 58.3 920.6339  
## 12 91.9 87.7 71.8 527.7966  
## 13 86.4 83.7 59.8 747.5730  
## 14 78.1 74.4 50.0 231.0984  
## 15 83.6 65.9 50.0 322.8221  
## 16 95.5 93.9 65.0 867.2498  
## 17 91.9 80.3 49.1 828.4708  
## 18 88.6 73.3 37.5 542.0000  
## PrgDist\_Total Cmp\_Short Att\_Short Cmp\_Medium Att\_Medium Cmp\_Long Att\_Long  
## 2 306.97244 23.886875 27.541697 17.849891 22.189993 6.2001450 11.7802756  
## 3 213.34648 22.306313 24.824247 16.753945 20.466284 3.0021521 5.4232425  
## 4 256.39257 26.635089 29.158772 17.836302 20.018947 2.1485411 3.9560440  
## 5 254.92101 19.344184 20.550503 30.890378 32.699856 2.2833892 4.5236955  
## 6 47.40759 16.590467 18.647860 7.222763 8.711089 1.1819066 2.0573930  
## 7 362.62136 26.830383 28.475157 30.890919 33.203883 3.2895488 5.9623073  
## 8 23.01606 8.256881 9.959862 1.961009 2.735092 0.2064220 0.2580275  
## 9 138.56436 20.216584 22.667079 15.426980 16.540842 3.1745050 4.2883663  
## 10 106.14806 22.979925 25.407779 10.445420 12.252196 1.8632371 3.5570891  
## 11 311.63183 18.327714 19.966285 26.277815 28.887390 4.0660823 6.9790964  
## 12 132.13983 25.296610 27.521186 9.978814 11.377119 1.7796610 2.4788136  
## 13 237.43820 21.438202 24.808989 18.000000 21.505618 3.9101124 6.5393258  
## 14 57.83880 12.582026 16.112696 3.915835 5.263909 0.5777461 1.1554922  
## 15 79.87730 15.460123 18.496933 5.153374 7.822086 1.2883436 2.5766871  
## 16 285.65949 24.752105 25.930776 25.762395 27.446211 2.1889616 3.3676333  
## 17 255.15091 26.800805 29.154930 15.845070 19.738431 5.0704225 10.3219316  
## 18 130.15385 26.307692 29.692308 11.384615 15.538462 0.9230769 2.4615385  
## KP Final\_Third PPA CrsPA PrgP TB\_Pass\_Types  
## 2 2.74111675 5.7432922 2.38216099 0.35895577 9.1696882 0.88107324  
## 3 0.83931133 1.9691535 0.80703013 0.25824964 3.2926829 0.12912482  
## 4 0.61386889 4.0242516 0.61386889 0.17051914 3.4785904 0.23872679  
## 5 0.25849689 2.7142173 0.08616563 0.04308281 2.3695548 0.00000000  
## 6 1.31322957 0.8754864 0.96303502 0.13132296 2.3638132 0.08754864  
## 7 0.46259280 4.6773272 0.92518561 0.10279840 6.0137065 0.25699600  
## 8 0.72247706 0.6192661 0.25802752 0.05160550 0.8256881 0.10321101  
## 9 0.50123762 2.8960396 0.50123762 0.16707921 3.0631188 0.11138614  
## 10 2.20200753 1.4115433 1.63739021 0.28230866 2.9924718 0.11292346  
## 11 0.36412677 2.5488874 0.18206339 0.00000000 2.6095752 0.00000000  
## 12 0.88983051 2.8601695 0.57203390 0.00000000 3.5593220 0.12711864  
## 13 1.34831461 4.5842697 1.34831461 0.40449438 4.7191011 0.60674157  
## 14 0.77032810 1.6690442 0.38516405 0.00000000 2.8245364 0.38516405  
## 15 1.38036810 1.1963190 1.47239264 0.73619632 1.9325153 0.55214724  
## 16 0.08419083 3.6202058 0.33676333 0.08419083 4.0411600 0.16838167  
## 17 2.08249497 4.7987928 0.72434608 0.18108652 4.4366197 0.45271630  
## 18 1.07692308 2.3076923 0.92307692 0.30769231 3.2307692 0.00000000  
## Sw\_Pass\_Types Crs\_Pass\_Types TI\_Pass\_Types gca\_cols\_unadj PassLive\_SCA\_Types  
## 2 0.62001450 5.61276287 0.58738216 5.51 3.9158811  
## 3 0.25824964 2.09827834 7.36011478 2.39 1.8400287  
## 4 0.17051914 0.68207654 5.32019704 1.71 1.3641531  
## 5 0.38774533 0.25849689 0.04308281 0.60 0.5169938  
## 6 0.39396887 2.18871595 0.56906615 3.24 2.4075875  
## 7 0.15419760 0.46259280 0.82238721 1.34 1.1821816  
## 8 0.00000000 0.30963303 0.00000000 1.55 1.0321101  
## 9 0.27846535 0.38985149 0.11138614 1.67 1.2809406  
## 10 0.05646173 2.14554580 2.03262233 4.52 3.0489335  
## 11 0.48550236 0.12137559 0.06068780 1.03 0.7282535  
## 12 0.19067797 0.25423729 0.31779661 1.84 1.5254237  
## 13 0.53932584 1.28089888 0.13483146 3.31 3.0337079  
## 14 0.12838802 0.25677603 0.00000000 1.99 1.7974322  
## 15 0.46012270 2.48466258 0.36809816 2.67 2.2085890  
## 16 0.08419083 0.08419083 0.58933583 0.59 0.3367633  
## 17 0.72434608 6.70020121 0.45271630 4.07 2.6257545  
## 18 0.15384615 4.30769231 7.23076923 2.00 1.8461538  
## TO\_SCA\_Types Sh\_SCA\_Types Fld\_SCA\_Types Def\_SCA\_Types defense\_cols\_unadj  
## 2 0.16316171 0.19579405 0.19579405 0.06526468 38.7  
## 3 0.22596844 0.12912482 0.00000000 0.03228121 65.4  
## 4 0.03410383 0.06820765 0.10231148 0.06820765 60.0  
## 5 0.00000000 0.08616563 0.00000000 0.00000000 71.9  
## 6 0.13132296 0.39396887 0.26264591 0.00000000 32.4  
## 7 0.00000000 0.10279840 0.00000000 0.05139920 58.3  
## 8 0.15481651 0.20642202 0.15481651 0.00000000 0.0  
## 9 0.00000000 0.11138614 0.11138614 0.16707921 41.7  
## 10 0.39523212 0.11292346 0.22584693 0.22584693 57.7  
## 11 0.00000000 0.24275118 0.00000000 0.00000000 76.5  
## 12 0.25423729 0.00000000 0.06355932 0.00000000 61.8  
## 13 0.00000000 0.20224719 0.00000000 0.00000000 62.1  
## 14 0.00000000 0.12838802 0.06419401 0.00000000 50.0  
## 15 0.18404908 0.18404908 0.00000000 0.00000000 33.3  
## 16 0.00000000 0.16838167 0.00000000 0.08419083 64.7  
## 17 0.00000000 0.18108652 0.00000000 0.00000000 45.8  
## 18 0.00000000 0.00000000 0.15384615 0.00000000 70.6  
## TklW\_Tackles Def 3rd\_Tackles Mid 3rd\_Tackles Att 3rd\_Tackles Tkl\_Challenges  
## 2 1.5337201 0.7831762 1.2073967 0.48948513 0.9463379  
## 3 1.2589670 1.2589670 0.7101865 0.19368723 1.0975610  
## 4 2.0462296 1.8075028 1.2277378 0.44334975 1.7392952  
## 5 0.9478219 0.7324079 0.5169938 0.04308281 0.9909047  
## 6 0.7879377 0.4377432 0.2626459 0.52529183 0.4815175  
## 7 1.2335808 1.2849800 0.6167904 0.15419760 1.0793832  
## 8 0.0516055 0.0000000 0.1032110 0.05160550 0.0000000  
## 9 2.3948020 1.7264851 1.7264851 0.44554455 1.3923267  
## 10 1.4680050 0.3387704 1.0727729 0.45169385 0.8469260  
## 11 0.7889413 1.2137559 0.3034390 0.06068780 0.7889413  
## 12 1.5254237 1.0169492 0.6991525 0.76271186 1.3347458  
## 13 3.1011236 3.0337079 1.5505618 0.60674157 2.7640449  
## 14 0.5135521 0.1925820 0.5135521 0.19258203 0.3209700  
## 15 0.3680982 0.1840491 0.1840491 0.27607362 0.2760736  
## 16 0.8419083 1.3470533 0.6735267 0.08419083 0.9260992  
## 17 0.8148893 0.3621730 1.0865191 0.27162978 0.9959759  
## 18 2.6153846 1.8461538 1.3846154 0.30769231 1.8461538  
## Att\_Challenges Sh\_Blocks Pass\_Blocks Int Clr Err  
## 2 2.4474257 0.1305294 0.9137056 0.8158086 1.1094996 0.06526468  
## 3 1.6786227 0.2905308 0.4196557 1.1944046 2.3242468 0.00000000  
## 4 2.8988253 0.2728306 0.8525957 1.1595301 3.0011368 0.03410383  
## 5 1.3786501 0.9047391 0.3446625 1.4217329 3.9636190 0.00000000  
## 6 1.4883268 0.0000000 0.8317121 0.5690661 0.4377432 0.08754864  
## 7 1.8503712 0.4111936 0.7195888 1.6447744 2.9297544 0.05139920  
## 8 0.2064220 0.0000000 0.2580275 0.0516055 0.3096330 0.00000000  
## 9 3.3415842 0.1670792 1.3366337 1.3923267 0.8910891 0.05569307  
## 10 1.4680050 0.0000000 0.5646173 0.6210790 0.9033877 0.00000000  
## 11 1.0316925 0.7282535 0.1213756 2.2454484 4.3088334 0.06068780  
## 12 2.1610169 0.1271186 0.8898305 1.2076271 0.4449153 0.06355932  
## 13 4.4494382 0.6741573 1.5505618 0.8089888 3.2359551 0.20224719  
## 14 0.6419401 0.2567760 0.7061341 0.3851641 0.5135521 0.06419401  
## 15 0.8282209 0.0000000 0.7361963 0.3680982 0.6441718 0.00000000  
## 16 1.4312442 0.2525725 0.1683817 1.4312442 3.3676333 0.00000000  
## 17 2.1730382 0.3621730 0.8148893 0.6338028 0.6338028 0.00000000  
## 18 2.6153846 0.0000000 1.2307692 0.6153846 2.6153846 0.30769231  
## possession\_cols\_unadj Def Pen\_Touches Def 3rd\_Touches Mid 3rd\_Touches  
## 2 41.5 1.9253082 12.987672 36.87455  
## 3 50.9 3.6477762 21.241033 24.75968  
## 4 55.3 5.3884047 26.566881 30.79576  
## 5 50.0 7.7979895 35.026328 31.10579  
## 6 30.6 0.5690661 5.252918 14.88327  
## 7 85.7 8.0182753 36.647630 35.82524  
## 8 31.0 0.3096330 1.599771 11.50803  
## 9 54.8 2.2277228 16.707921 32.02351  
## 10 50.0 0.7340025 8.469260 21.11669  
## 11 40.0 8.2535401 33.985165 30.89009  
## 12 50.0 1.2076271 11.122881 33.24153  
## 13 50.0 4.6516854 22.044944 37.41573  
## 14 31.4 1.0271041 4.557775 19.45078  
## 15 23.1 0.2760736 3.220859 17.39264  
## 16 33.3 6.7352666 31.824135 31.73994  
## 17 0.0 2.1730382 14.486922 34.04427  
## 18 33.3 2.0000000 18.000000 20.76923  
## Att 3rd\_Touches Att Pen\_Touches Att\_Take\_Ons Succ\_Take\_Ons Carries\_Carries  
## 2 28.977520 2.3168963 2.1211022 0.88107324 45.00000  
## 3 17.980631 2.1305595 1.7109039 0.87159254 35.86442  
## 4 10.845017 0.9208033 1.6028799 0.88669951 37.03676  
## 5 2.412638 1.2494016 0.1723313 0.08616563 45.79703  
## 6 27.840467 7.1789883 3.7208171 1.13813230 34.27529  
## 7 6.424900 0.9251856 0.3597944 0.30839520 52.27299  
## 8 10.372706 2.9931193 1.4965596 0.46444954 14.19151  
## 9 10.581683 0.8353960 1.7264851 0.94678218 33.36015  
## 10 28.456713 6.1543287 4.0652447 2.03262233 37.03890  
## 11 3.762643 2.2454484 0.3034390 0.12137559 39.93257  
## 12 14.936441 2.0338983 2.9237288 1.46186441 36.10169  
## 13 14.629213 2.5617978 0.4044944 0.20224719 38.76404  
## 14 14.058488 3.5948645 2.2467903 0.70613409 20.99144  
## 15 20.429448 3.1288344 3.5889571 0.82822086 26.96319  
## 16 5.135641 1.0944808 0.2525725 0.08419083 46.55753  
## 17 22.273642 1.1770624 0.2716298 0.00000000 38.39034  
## 18 24.615385 3.5384615 1.3846154 0.46153846 35.53846  
## TotDist\_Carries PrgDist\_Carries PrgC\_Carries Final\_Third\_Carries CPA\_Carries  
## 2 211.71864 112.12473 2.1211022 1.9579405 0.35895577  
## 3 158.56528 86.64275 2.0014347 1.1944046 0.51649928  
## 4 172.59947 91.39826 1.2618416 0.9890110 0.10231148  
## 5 202.31690 95.60077 0.3015797 0.1292484 0.00000000  
## 6 260.50097 132.85506 5.8657588 3.0642023 2.84533074  
## 7 264.24329 160.82810 1.1821816 1.1821816 0.10279840  
## 8 71.00917 32.25344 1.3417431 0.6708716 0.82568807  
## 9 143.13119 64.32550 0.6683168 0.6683168 0.05569307  
## 10 253.06148 129.52321 5.1944793 2.5972396 2.93601004  
## 11 246.08901 99.77073 0.3641268 0.3641268 0.18206339  
## 12 163.22034 69.21610 1.3347458 0.8898305 0.63559322  
## 13 119.59551 55.48315 0.2696629 0.6741573 0.00000000  
## 14 95.19971 33.57347 0.8987161 1.3480742 0.25677603  
## 15 155.88957 66.99387 1.8404908 1.7484663 1.10429448  
## 16 235.98690 131.08513 1.5996258 0.9260992 0.08419083  
## 17 138.25956 61.29779 0.9054326 0.9959759 0.00000000  
## 18 191.84615 122.00000 3.8461538 2.0000000 0.61538462  
## Mis\_Carries Rec\_Receiving PrgR\_Receiving misc\_cols\_unadj Fls Fld  
## 2 1.2400290 54.33285 3.8506164 42.9 0.8484409 1.2073967  
## 3 1.1944046 38.28551 5.2941176 64.4 1.2912482 0.2905308  
## 4 0.7843880 40.58355 1.7392952 62.7 1.2618416 0.7502842  
## 5 0.2154141 49.20057 0.2584969 66.3 0.6893250 0.5169938  
## 6 2.6702335 36.28891 11.9941634 11.5 0.4815175 1.7509728  
## 7 0.0513992 54.94575 0.3597944 55.2 1.1307824 0.3597944  
## 8 2.9415138 18.16514 4.2832569 21.9 0.7740826 1.3417431  
## 9 1.6150990 36.20050 1.0581683 48.6 1.9492574 1.7264851  
## 10 1.9761606 40.31368 8.9774153 37.5 1.0727729 1.2986198  
## 11 0.3641268 43.81659 0.3641268 70.8 1.0316925 0.4855024  
## 12 1.9067797 41.69492 2.4152542 40.0 1.6525424 1.2076271  
## 13 0.6741573 46.98876 1.7528090 64.6 1.1460674 1.2134831  
## 14 2.6319544 26.83310 4.4293866 28.0 0.5777461 0.8345221  
## 15 1.2883436 31.38037 7.2699387 41.7 0.7361963 0.6441718  
## 16 0.2525725 49.50421 0.3367633 51.6 0.7577175 0.1683817  
## 17 0.9959759 46.53924 2.6257545 28.6 0.7243461 0.2716298  
## 18 1.5384615 40.00000 8.1538462 68.8 2.0000000 2.9230769  
## Off Recov Won\_Aerial\_Duels Lost\_Aerial\_Duels  
## 2 0.06526468 6.265410 0.6852792 0.9137056  
## 3 0.12912482 5.294118 1.2266858 0.6779053  
## 4 0.10231148 4.808640 1.6028799 0.9549072  
## 5 0.08616563 2.843466 2.4557204 1.2494016  
## 6 0.43774319 4.508755 0.1313230 1.0068093  
## 7 0.00000000 4.163335 0.8223872 0.6681896  
## 8 0.56766055 1.548165 1.1869266 4.2316514  
## 9 0.00000000 6.181931 0.9467822 1.0024752  
## 10 0.50815558 5.250941 0.5081556 0.8469260  
## 11 0.06068780 3.216453 3.0950775 1.2744437  
## 12 0.12711864 4.894068 1.1440678 1.7161017  
## 13 0.20224719 6.000000 2.0898876 1.1460674  
## 14 0.19258203 3.659058 1.3480742 3.4664765  
## 15 0.64417178 2.208589 0.4601227 0.6441718  
## 16 0.08419083 3.451824 1.3470533 1.2628625  
## 17 0.00000000 5.342052 0.1810865 0.4527163  
## 18 0.46153846 4.000000 1.6923077 0.7692308

# Perform PCA  
pca\_result <- prcomp(manu\_2425\_logs[ , 5:76], scale. = TRUE)  
pca\_data <- as.data.frame(pca\_result$x[, 1:2]) # Use the first 2 principal components  
pca\_data$cluster <- as.factor(manu\_2425\_logs$Pos)  
pca\_data$label <- as.factor(manu\_2425\_logs$Player)  
  
  
loadings\_pc1\_pc2 <- pca\_result$rotation[, 1:2]  
# Sort for PC1 by absolute value  
sorted\_pc1 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,1]), decreasing = TRUE), 1]  
# Sort for PC2 by absolute value  
sorted\_pc2 <- loadings\_pc1\_pc2[order(abs(loadings\_pc1\_pc2[,2]), decreasing = TRUE), 2]  
cat("Top variables for PC1:\n")

## Top variables for PC1:

sorted\_pc1

## Def 3rd\_Touches Cmp\_Medium Att\_Medium   
## 0.186107368 0.180347272 0.180091239   
## Mis\_Carries Def Pen\_Touches TotDist\_Total   
## -0.177754461 0.177544579 0.176699542   
## PrgDist\_Total PrgR\_Progression PrgR\_Receiving   
## 0.174387373 -0.169374997 -0.169374997   
## Int Mid 3rd\_Touches Att\_Take\_Ons   
## 0.165453464 0.164835655 -0.163556145   
## Off Clr Att Pen\_Touches   
## -0.160257489 0.160067152 -0.159800228   
## misc\_cols\_unadj npxG\_Expected CPA\_Carries   
## 0.152231624 -0.149999771 -0.149855116   
## Sh\_Blocks SoT\_per\_90\_Standard Cmp\_percent\_Medium   
## 0.149113455 -0.145760548 0.145385084   
## Carries\_Carries Rec\_Receiving Final\_Third\_Carries   
## 0.144825238 0.141364242 -0.137886160   
## Sh\_per\_90\_Standard defense\_cols\_unadj Cmp\_percent\_Short   
## -0.137865302 0.135974657 0.135081938   
## Final\_Third Att 3rd\_Touches Fld\_SCA\_Types   
## 0.134410778 -0.133176450 -0.132685522   
## PrgC\_Carries PrgC\_Progression Def 3rd\_Tackles   
## -0.127753060 -0.127753060 0.121557702   
## TO\_SCA\_Types Succ\_Take\_Ons Won\_Aerial\_Duels   
## -0.115350582 -0.115258730 0.114894299   
## Cmp\_Short Tkl\_Challenges Cmp\_Long   
## 0.114594054 0.105911256 0.104578759   
## xAG\_Expected Att\_Short CrdY   
## -0.104409351 0.102734314 0.096520248   
## Att\_Long KP gca\_cols\_unadj   
## 0.093001862 -0.091622343 -0.089696179   
## PassLive\_SCA\_Types Fld possession\_cols\_unadj   
## -0.088449268 -0.086783469 0.085495473   
## Att\_Challenges Lost\_Aerial\_Duels PrgP   
## 0.081333603 -0.074857235 0.070230317   
## TklW\_Tackles Sh\_SCA\_Types PPA   
## 0.068593341 -0.067359826 -0.063541881   
## Mid 3rd\_Tackles CrsPA Fls   
## 0.061566653 -0.055220644 0.054751905   
## SoT\_percent\_Standard PrgDist\_Carries TotDist\_Carries   
## -0.051798233 0.048229505 0.047123617   
## Att 3rd\_Tackles Crs\_Pass\_Types npxG\_per\_Sh\_Expected   
## -0.045568215 -0.041765484 -0.038502586   
## Recov Dist\_Standard Sw\_Pass\_Types   
## 0.036715412 -0.036148597 0.034961175   
## TB\_Pass\_Types Pass\_Blocks TI\_Pass\_Types   
## -0.017977522 -0.010165773 0.009713711   
## Err Cmp\_percent\_Long Def\_SCA\_Types   
## 0.005676614 -0.003655817 -0.002472114

cat("\nTop variables for PC2:\n")

##   
## Top variables for PC2:

sorted\_pc2

## Recov PassLive\_SCA\_Types PPA   
## 0.213607743 0.204183164 0.201968984   
## gca\_cols\_unadj Dist\_Standard KP   
## 0.198254994 0.196524083 0.190225787   
## Att\_Short npxG\_per\_Sh\_Expected Mid 3rd\_Tackles   
## 0.187239887 -0.186058494 0.180666195   
## Att 3rd\_Touches Att\_Challenges Att 3rd\_Tackles   
## 0.179780780 0.177857212 0.177627289   
## PrgP Crs\_Pass\_Types Lost\_Aerial\_Duels   
## 0.176742488 0.176578248 -0.171579963   
## Pass\_Blocks xAG\_Expected Cmp\_Short   
## 0.171480267 0.169941893 0.168873154   
## TklW\_Tackles Att\_Long Cmp\_Long   
## 0.162161775 0.151286092 0.145405727   
## Final\_Third Tkl\_Challenges Rec\_Receiving   
## 0.140760897 0.138188943 0.132734227   
## TB\_Pass\_Types Final\_Third\_Carries SoT\_percent\_Standard   
## 0.128659949 0.125761025 -0.123161444   
## CrsPA Sw\_Pass\_Types Carries\_Carries   
## 0.122143836 0.112670251 0.108069039   
## Mid 3rd\_Touches Won\_Aerial\_Duels Sh\_per\_90\_Standard   
## 0.107621834 -0.106228210 0.102869012   
## Err Cmp\_percent\_Long PrgC\_Progression   
## 0.095548484 -0.094304906 0.091086141   
## PrgC\_Carries Def 3rd\_Tackles Succ\_Take\_Ons   
## 0.091086141 0.087978219 0.084298990   
## Fld\_SCA\_Types PrgDist\_Carries Def\_SCA\_Types   
## 0.082851418 0.082781942 0.080407610   
## Fls Fld TotDist\_Total   
## 0.080284303 0.080236622 0.075155339   
## PrgR\_Receiving PrgR\_Progression TotDist\_Carries   
## 0.065197974 0.065197974 0.065005243   
## TI\_Pass\_Types PrgDist\_Total Def Pen\_Touches   
## 0.063640295 0.060919774 -0.057760816   
## Sh\_Blocks Att\_Take\_Ons TO\_SCA\_Types   
## -0.056443596 0.055999037 0.053464704   
## CPA\_Carries Clr defense\_cols\_unadj   
## 0.042663072 -0.039162979 0.037436087   
## Cmp\_percent\_Short Off Att Pen\_Touches   
## 0.037160812 -0.033246951 0.033220234   
## SoT\_per\_90\_Standard Mis\_Carries Att\_Medium   
## 0.030321913 -0.026244287 0.025194717   
## Int Def 3rd\_Touches CrdY   
## -0.024902584 -0.021522639 0.019798062   
## npxG\_Expected Sh\_SCA\_Types possession\_cols\_unadj   
## -0.016265970 -0.016163147 0.014101946   
## misc\_cols\_unadj Cmp\_Medium Cmp\_percent\_Medium   
## 0.012667910 0.005238427 0.000378773

# Plot PCA with clusters  
ggplot(pca\_data, aes(x = PC1, y = PC2, color = cluster)) +  
 geom\_point(size = 3) +  
 scale\_x\_continuous(expand = expansion(mult = 0.1)) +  
 scale\_y\_continuous(expand = expansion(mult = 0.1)) +  
 theme\_minimal() +  
 labs(title = "PCA of Man. United Outfield Players, by Position", x = "PC1", y = "PC2") +   
 geom\_text(aes(label = label), vjust = -1, hjust = 0.5, color = "black", size = 2.5) # Add labels

