Project #1 Answers

STAT 494/873

Fall 2025

The National Football League (NFL) holds a scouting combine every year for college football players who would like to play football in the NFL. These players go through a number of evaluations during the combine so that NFL teams can assess their ability. For more information, please see <http://en.wikipedia.org/wiki/NFL_Scouting_Combine> and <http://www.nfl.com/combine>.

The NFLcombine.csv data file is available from my course website, and it contains information on some of the players who participated in a combine. The columns in the data file represent the following information:

* Player: Name of player being evaluated
* College: College that the player attended
* Position: The position of the player where DB = defensive back, LB = linebacker, OL = offensive linemen, RB = running back, S = safety, TE = tight end, WO = wide receiver; players who played other positions were excluded from the data file
* OverallGrade: The overall grade of the player based on the evaluations
* Height: Height in inches
* ArmLength: Arm length in inches
* HandLength: Hand length in inches
* Weight: Weight in pounds
* Dash40: 40-yard dash time in seconds
* BenchPress: Number of bench press repetitions of 225 pounds
* VerticalJump: Vertical jump in inches
* BroadJump: Broad jump in inches
* Cone3Drill: 3-cone drill time in seconds
* Shuttle20: 20-yard shuttle run in seconds

Complete the following problems below using the data. While you are welcome to use your football knowledge to help with interpretations, this is not needed to perform well on this project. Include your R program output with code inside of it for each part and any additional information needed to explain your answer. Your R code and output should be formatted in the same manner as in the lecture notes.

1. This part focuses on plotting the data and interpreting the corresponding plots.
   1. (4 points) Construct side-by-side scatter plots for the numerical variables. Compute the estimated correlation matrix for these same variables. Interpret the plots and the matrix in the context of the data.

It is difficult to get the variable names and axis labels to be shown large enough without part of the labels being excluded.

> fb <- read.csv(file = "NFLcombine.csv", stringsAsFactors = TRUE))

> head(fb)

Player College Position OverallGrade

1 Abbrederis, Jared Wisconsin WO 5.20

2 Adams, Davante Fresno St. WO 5.96

3 Amaro, Jace Texas Tech TE 5.40

4 Andrews, Antonio Western Kentucky RB 5.16

5 Archer, Dri Kent St. RB 5.45

6 Atkinson, George Notre Dame RB 5.10

Height ArmLength Weight HandLength Dash40 BenchPress

1 73 31.375 195 9.625 4.50 4

2 73 32.625 212 9.000 4.56 14

3 77 34.000 265 9.000 4.74 28

4 70 31.250 225 9.500 4.82 20

5 68 31.000 173 8.875 4.26 20

6 73 33.250 218 9.375 4.48 19

VerticalJump BroadJump Cone3Drill Shuttle20

1 30.5 117 6.80 4.08

2 39.5 123 6.82 4.30

3 33.0 118 7.42 4.30

4 29.5 106 7.24 4.49

5 38.0 122 6.86 4.06

6 38.0 121 7.07 4.46

> table(fb$Position)

DB LB OL RB S TE WO

18 18 30 16 11 8 28

> pairs(formula = ~ OverallGrade + Height + ArmLength + Weight + HandLength + Dash40 + BenchPress + VerticalJump + BroadJump + Cone3Drill + Shuttle20, data = fb)

A grid of black dots

AI-generated content may be incorrect.

> round(cor(fb[,-c(1:3)]),2)

OverallGrade Height ArmLength Weight HandLength Dash40 BenchPress

OverallGrade 1.00 0.21 0.31 0.19 0.16 0.05 0.08

Height 0.21 1.00 0.75 0.77 0.40 0.67 0.40

ArmLength 0.31 0.75 1.00 0.68 0.49 0.57 0.33

Weight 0.19 0.77 0.68 1.00 0.46 0.91 0.71

HandLength 0.16 0.40 0.49 0.46 1.00 0.42 0.28

Dash40 0.05 0.67 0.57 0.91 0.42 1.00 0.57

BenchPress 0.08 0.40 0.33 0.71 0.28 0.57 1.00

VerticalJump 0.02 -0.57 -0.50 -0.77 -0.28 -0.80 -0.46

BroadJump 0.02 -0.56 -0.49 -0.78 -0.29 -0.86 -0.45

Cone3Drill 0.07 0.51 0.53 0.81 0.34 0.79 0.58

Shuttle20 0.06 0.53 0.57 0.76 0.35 0.79 0.44

VerticalJump BroadJump Cone3Drill Shuttle20

OverallGrade 0.02 0.02 0.07 0.06

Height -0.57 -0.56 0.51 0.53

ArmLength -0.50 -0.49 0.53 0.57

Weight -0.77 -0.78 0.81 0.76

HandLength -0.28 -0.29 0.34 0.35

Dash40 -0.80 -0.86 0.79 0.79

BenchPress -0.46 -0.45 0.58 0.44

VerticalJump 1.00 0.87 -0.72 -0.71

BroadJump 0.87 1.00 -0.73 -0.70

Cone3Drill -0.72 -0.73 1.00 0.88

Shuttle20 -0.71 -0.70 0.88 1.00

Comments:

* The OverallGrade variable is not closely correlated with any one variable, where the largest value is 0.31 for ArmLength.
* There are some strong correlations between some variable pairs. These are shown below as TRUE values:

> save.cor <- cor(fb[,-c(1:3)])

> large.cor <- abs(save.cor) > 0.8

> large.cor

OverallGrade Height ArmLength Weight HandLength Dash40 BenchPress

OverallGrade TRUE FALSE FALSE FALSE FALSE FALSE FALSE

Height FALSE TRUE FALSE FALSE FALSE FALSE FALSE

ArmLength FALSE FALSE TRUE FALSE FALSE FALSE FALSE

Weight FALSE FALSE FALSE TRUE FALSE TRUE FALSE

HandLength FALSE FALSE FALSE FALSE TRUE FALSE FALSE

Dash40 FALSE FALSE FALSE TRUE FALSE TRUE FALSE

BenchPress FALSE FALSE FALSE FALSE FALSE FALSE TRUE

VerticalJump FALSE FALSE FALSE FALSE FALSE TRUE FALSE

BroadJump FALSE FALSE FALSE FALSE FALSE TRUE FALSE

Cone3Drill FALSE FALSE FALSE TRUE FALSE FALSE FALSE

Shuttle20 FALSE FALSE FALSE FALSE FALSE FALSE FALSE

VerticalJump BroadJump Cone3Drill Shuttle20

OverallGrade FALSE FALSE FALSE FALSE

Height FALSE FALSE FALSE FALSE

ArmLength FALSE FALSE FALSE FALSE

Weight FALSE FALSE TRUE FALSE

HandLength FALSE FALSE FALSE FALSE

Dash40 TRUE TRUE FALSE FALSE

BenchPress FALSE FALSE FALSE FALSE

VerticalJump TRUE TRUE FALSE FALSE

BroadJump TRUE TRUE FALSE FALSE

Cone3Drill FALSE FALSE TRUE TRUE

Shuttle20 FALSE FALSE TRUE TRUE

Of course, the diagonal elements will be TRUE. The off-diagonal elements are where the strong correlations can exist between variables. For example, the correlation between Weight and Dash40 is 0.91. Thus, there is a strong, positive linear relationship between a player’s weight and their time in the 40-yard dash (the larger the weight, the slower the player).

* 1. (6 points) Create a parallel coordinates plot for the numerical variables. The color of the lines for each player should correspond to their position in the following way:

*#Suppose the fb data frame contains the data*

library(plyr)

color.position <- palette()[1:length(levels(fb$Position))]

Position.color <- revalue(x = fb$Position, replace = c(DB = color.position[1],

LB = color.position[2], OL = color.position[3], RB = color.position[4],

S = color.position[5], TE = color.position[6], WO = color.position[7]))

Are there any trends in the plot corresponding to the position? If there are trends, explain what characteristics of the plot lead you to this conclusion. Are there any outliers? If there are outliers, identify them by observation number and player name along with discussing the characteristics of the plot that lead you to this conclusion.

> library(plyr)

> color.position <- palette()[1:length(levels(fb$Position))]

> Position.color <- revalue(x = fb$Position, replace = c(DB=color.position[1],

LB=color.position[2], OL=color.position[3], RB=color.position[4],

S=color.position[5], TE=color.position[6], WO=color.position[7]))

> fb2 <- data.frame(player = 1:nrow(fb), fb[,4:14])

> head(fb2)

player OverallGrade Height ArmLength Weight HandLength Dash40 BenchPress VerticalJump

1 1 5.20 73 31.375 195 9.625 4.50 4 30.5

2 2 5.96 73 32.625 212 9.000 4.56 14 39.5

3 3 5.40 77 34.000 265 9.000 4.74 28 33.0

4 4 5.16 70 31.250 225 9.500 4.82 20 29.5

5 5 5.45 68 31.000 173 8.875 4.26 20 38.0

6 6 5.10 73 33.250 218 9.375 4.48 19 38.0

BroadJump Cone3Drill Shuttle20

1 117 6.80 4.08

2 123 6.82 4.30

3 118 7.42 4.30

4 106 7.24 4.49

5 122 6.86 4.06

6 121 7.07 4.46

> library(package = MASS)

> dev.new(width = 10, height = 6, pointsize = 8)

> parcoord(x = fb2, col = Position.color, main = "Parallel coordinate plot")

> legend(locator(1), legend = levels(fb$Position)[1:4], lty = 1, col =

color.position[1:4], cex = 1, bty = "n")

> legend(locator(1), legend = levels(fb$Position)[5:7], lty = 1, col =

color.position[5:7], cex = 1, bty = "n")

A diagram of lines and colors

AI-generated content may be incorrect.

There are many different comments that one can make about this plot. Below are a few:

* OL players tend to have much larger weights and timed variables values (Dash40m Cone3Drill, Shuttle20) than players at other positions. The OL players also tend to have small vertical and broad jumps. By far, OL players distinguish themselves from other types of football players.
* DB players tend to have long jumps (large VerticalJump and BroadJump) while also being smaller in height and weight, weaker in strength (BenchPress), and fastest in timed variables (Dash40, Cones3Drill, Shuttle20).
* With respect to OverallGrade, Sammy Watkins by far stands out as having the largest overall grade value. Note that he was the 4th pick of the first round of the draft. While not as distinguished from the other players, Phillip Gaines has the lowest overall grade value. He was the 23rd pick of the third round. This is still a good draft position, but perhaps he was drafted a little lower due to his overall grade (I am speculating here!).

> fb[fb$OverallGrade == max(fb$OverallGrade),] #Watkins, Sammy: Round 1, pick 4

Player College Position OverallGrade Height ArmLength Weight HandLength

122 Watkins, Sammy Clemson WO 7 71 32 211 9.625

Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

122 4.43 16 34 126 6.95 4.34

> fb[fb$OverallGrade == min(fb$OverallGrade),] #Gaines, Phillip: Round 3, pick 23

Player College Position OverallGrade Height ArmLength Weight HandLength

50 Gaines, Phillip Rice DB 4.5 72 31.875 193 9.625

Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

50 4.38 11 36.5 122 6.62 4.04

I would categorize Watkins as an outlier.

* Additional players that stood out include Trent Murphy. He has by far the largest hand length.

> fb[fb$HandLength == max(fb$HandLength),] #Murphy, Trent: Round 2, pick 15

Player College Position OverallGrade Height ArmLength Weight HandLength

90 Murphy, Trent Stanford LB 5.65 77 33.875 250 11.125

Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

90 4.86 19 35.5 118 6.78 4.2

I would categorize him as an outlier.

* The discreteness of some measures (like Height) make it difficult to follow some players across the axes of the plot.
  1. (6 points) Create a stars plot for the numerical variables. Sort the data by position in the following manner before plotting:

fb[order(fb$Position),]

and use 10 columns of stars per row. Are there any trends in the plot corresponding to player position? If there are trends, explain what characteristics of the plot lead you to this conclusion. Are there any outliers? If there are outliers, identify them by observation number and player name along with discussing the characteristics of the plot that lead you to this conclusion.

> # helps determine the player positions on the plot

> table(fb$Position)

DB LB OL RB S TE WO

18 18 30 16 11 8 28

> dev.new(width = 10, pointsize = 10)

> stars(x = fb[order(fb$Position),-c(1:2)], draw.segments = TRUE, ncol = 10,

cex = 0.75, main = "Combine data ordered by position", key.loc = c(-4, 5))

A chart of a pie chart

AI-generated content may be incorrect.

Below is another version of the plot that I used to help denote where the different player positions were located:

> stars(x = fb[order(fb$Position),-c(1:2)], draw.segments = TRUE, ncol = 10,

cex=0.75, main = "Combine data ordered by position", key.loc = c(-4, 5), axes =

TRUE)

> text(x = 25, y = 29.5, label = "DB", col = "red")

> segments(x0 = 19.5, x1 = 19.5, y0 = 26.5, y1 = 28.2, lwd = 5, col = "red")

> text(x = 25, y = 28, label = "LB", col = "red")

> segments(x0 = 15, x1 = 15, y0 = 22, y1 = 24, lwd = 5, col = "red")

> text(x = 25, y = 23, label = "OL", col = "red")

> segments(x0 = 15, x1 = 15, y0 = 15, y1 = 17, lwd = 5, col = "red")

> text(x = 25, y = 16, label = "RB", col = "red")

> segments(x0 = 6, x1 = 6, y0 = 10, y1 = 12, lwd = 5, col = "red")

> text(x = 25, y = 11.5, label = "S", col = "red")

> segments(x0 = 8, x1 = 8, y0 = 8, y1 = 10, lwd = 5, col = "red")

> text(x = 25, y = 9.5, label = "TE", col = "red")

> segments(x0 = 3.5, x1 = 3.5, y0 = 3, y1 = 5, lwd = 5, col = "red")

> text(x = 25, y = 4.5, label = "WO", col = "red")

A screenshot of a computer game

AI-generated content may be incorrect.

I left the axes = FALSE argument in the stars() function call so that you could see my choices for the segments() and text() argument values.

This plot allows us to see some of the same items as in the parallel coordinate plot (although perhaps it more difficult with the stars plot). In particular, the OL players stand out in comparison to the rest (notice the large weight values denoted by blue rays in the stars). Also, Sammy Watkins (#122) has the largest ray corresponding to OverallGrade.

* 1. (3 points) Would it be difficult to represent ALL of the numerical variables on one Trellis plot? If your answer is yes, fully explain why. If your answer is no, construct the plot and interpret it.

It would be difficult because the numerical values tend to have many different values (some are essentially continuous variables). While shingles could be formed, there are many different shingles needed and the cut-off points to form the shingles are somewhat arbitrary.

1. This part focuses on using PCA with the correlation matrix. **Exclude the OverallGrade variable in the analysis.**
   1. (3 points) Why is the correlation matrix more appropriate to use here than the covariance matrix with this data?

The variables are measured on different numerical scales.

* 1. (5 points) Determine the number of PCs needed. Fully justify your answer.

> options(width = 70)

> pca.cor<-princomp(formula = ~ Height + ArmLength + Weight + HandLength + Dash40 + BenchPress + VerticalJump + BroadJump + Cone3Drill + Shuttle20, data = fb, cor = TRUE, scores = TRUE)

> summary(pca.cor, loadings = TRUE, cutoff = 0.0)

Importance of components:

Comp.1 Comp.2 Comp.3 Comp.4

Standard deviation 2.5545696 1.0226811 0.85582492 0.77866560

Proportion of Variance 0.6525826 0.1045877 0.07324363 0.06063201

Cumulative Proportion 0.6525826 0.7571702 0.83041385 0.89104586

Comp.5 Comp.6 Comp.7 Comp.8

Standard deviation 0.67472122 0.48047428 0.39720261 0.34054042

Proportion of Variance 0.04552487 0.02308555 0.01577699 0.01159678

Cumulative Proportion 0.93657073 0.95965628 0.97543328 0.98703005

Comp.9 Comp.10

Standard deviation 0.296387839 0.204581818

Proportion of Variance 0.008784575 0.004185372

Cumulative Proportion 0.995814628 1.000000000

Loadings:

Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8

Height 0.300 0.354 0.237 0.507 0.097 0.470 0.371 0.181

ArmLength 0.285 0.483 0.264 0.233 -0.307 -0.606 -0.295 -0.045

Weight 0.375 0.017 -0.143 0.172 0.046 0.191 -0.126 -0.101

HandLength 0.195 0.635 -0.285 -0.626 0.259 0.037 0.112 0.042

Dash40 0.365 -0.115 0.023 -0.054 0.174 0.345 -0.499 -0.315

BenchPress 0.252 -0.113 -0.811 0.342 -0.006 -0.219 0.046 -0.066

VerticalJump -0.333 0.264 -0.213 0.064 -0.370 0.396 -0.579 0.272

BroadJump -0.337 0.271 -0.212 0.103 -0.410 0.154 0.349 -0.407

Cone3Drill 0.342 -0.215 -0.079 -0.213 -0.446 0.037 0.156 0.645

Shuttle20 0.334 -0.151 0.131 -0.294 -0.540 0.158 0.120 -0.440

Comp.9 Comp.10

Height 0.197 0.179

ArmLength -0.081 0.064

Weight -0.280 -0.816

HandLength 0.048 0.006

Dash40 -0.352 0.478

BenchPress 0.236 0.196

VerticalJump 0.251 -0.055

BroadJump -0.519 0.097

Cone3Drill -0.359 0.124

Shuttle20 0.483 -0.058

At least two principal components are needed to achieve 75% of the total variation being accounted for. Also, only two principal components have estimated eigenvalues greater than 1. Therefore, at least two are needed, but I am o.k. with using three because 3D plots are relatively easy to construct and will help one see if they are missing something with only two PCs.

We see that PC #1 is VERY important because it accounts for 65% of the total variation among the 10 variables!

* 1. (6 points) Interpret the PCs chosen from b).

PC #1: This is a contrast between jumping variables and all of the remaining variables.

PC #2: This is mainly a body length measurement PC (Height, ArmLength, HandLength). Perhaps one could say it is a contrast between Height, ArmLength, HandLength, VerticalJump, BroadJump and Dash40, BenchPress, Cone3Drill, Shuttle20; however, the variables with negative values have small coefficients.

PC #3: This is a measurement of strength as given by BenchPress. Again, one could include the other variables in the interpretation like for PC #2, but their values are small relative to the BenchPress coefficient.

* 1. (4 points) Show how the first PC score for the first observation is found using matrix algebra in R and through using by-hand calculations. Use predict() or the scores component from princomp() to check your answer.

> #Scores with the adjustment

> pca.cor$scale <- apply(X = fb[,-c(1:4)], MARGIN = 2, FUN = sd)

> score.cor <- predict(pca.cor, newdata = fb)

> head(score.cor[,1:3])

Comp.1 Comp.2 Comp.3

[1,] -1.9482744 0.2871572 1.714398081

[2,] -1.9821342 0.4156005 0.565226520

[3,] 1.0851349 0.2293685 -0.313895601

[4,] 0.3062786 -1.5321963 -0.001247874

[5,] -3.4559516 -0.9225880 -0.737615715

[6,] -0.9588560 0.6536310 0.026777909

> pca.cor$loadings[,1] #a^\*\_1

Height ArmLength Weight HandLength Dash40

0.2997879 0.2847605 0.3752775 0.1948522 0.3654330

BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

0.2522646 -0.3334001 -0.3366015 0.3423159 0.3338007

> Z <- scale(fb[,-c(1:4)])

> z1 <- Z[1,]

> z1

Height ArmLength Weight HandLength Dash40

-0.1408421 -0.6548626 -0.9151200 0.1153049 -0.8151051

BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

-2.2239914 -0.5958581 0.1339942 -0.9906777 -1.0609538

> pca.cor$loadings[,1]%\*%z1

[,1]

[1,] -1.948274



* 1. (4 points) Construct a bubble plot of the first three PC scores. Include the observation number (black color) in the middle of each bubble. Colorize each bubble by position using

fg = fb$Position

in the symbols() function (the colors will be the same as given in 1)b)).

> symbols(x = pca.cor$scores[,1], y = pca.cor$scores[,2], circles =

pca.cor$scores[,3]-min(pca.cor$scores[,3]), inches = 0.25, xlab = "Principal

component 1", ylab = "Principal component 2", fg = fb$Position,

main = "Bubble plot for first three principal components \n NFL data",

xlim = c(min(pca.cor$scores[,1:2]), max(pca.cor$scores[,1:2])),

ylim = c(min(pca.cor$scores[,1:2]), max(pca.cor$scores[,1:2])))

> abline(h = 0, lty = 1, lwd = 2)

> abline(v = 0, lty = 1, lwd = 2)

> text(x = pca.cor$scores[,1], y = pca.cor$scores[,2], col = "black", cex = 0.75)

> legend(locator(1), legend = levels(fb$Position), pch = c(1,1,1,1,1,1,1),

col = color.position, cex = 1, bty = "n")

A diagram of a bubble plot

AI-generated content may be incorrect.

* 1. (4 points) Identify outliers on the plot from e) with their observation number and player name. Explain why they are outliers relative to the characteristics of the PCs.

With respect to PC #1, there do not appear to be any outliers (no plotting points are significantly separated on the plot). With respect to PC #2, observations 90 and 124 are somewhat separated from the other points suggesting they could be outliers. These observations are

> fb[90,]

Player College Position OverallGrade Height ArmLength Weight HandLength

90 Murphy, Trent Stanford LB 5.65 77 33.875 250 11.125

Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

90 4.86 19 35.5 118 6.78 4.2

> score.cor[90,1:3]

Comp.1 Comp.2 Comp.3

-0.6687963 -3.3811481 -0.4159878

> fb[124,]

Player College Position OverallGrade Height ArmLength Weight HandLength

124 White, James Wisconsin RB 5 69 29.25 204 8.25

Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

124 4.57 23 32 114 7.05 4.2

> score.cor[124,1:3]

Comp.1 Comp.2 Comp.3

2.0868474 3.1425432 -0.5476279

Observation 90 was mentioned as standing out in the parallel coordinate plot too. This player has the largest overall hand length. Because HandLength has a positive coefficient in PC #2, this is a large contributing factor to observation 90 having the largest PC #2 value.

With respect to observation #124, I highlighted the observation in the parallel coordinate plot:

> dev.new(width = 10, height = 6, pointsize = 8)

> new.lwd <- c(rep(x = 1, times = 123), 5, rep(x = 1, times = 129-124))

> parcoord(x = fb2, col = Position.color, lwd = new.lwd, main = "Parallel

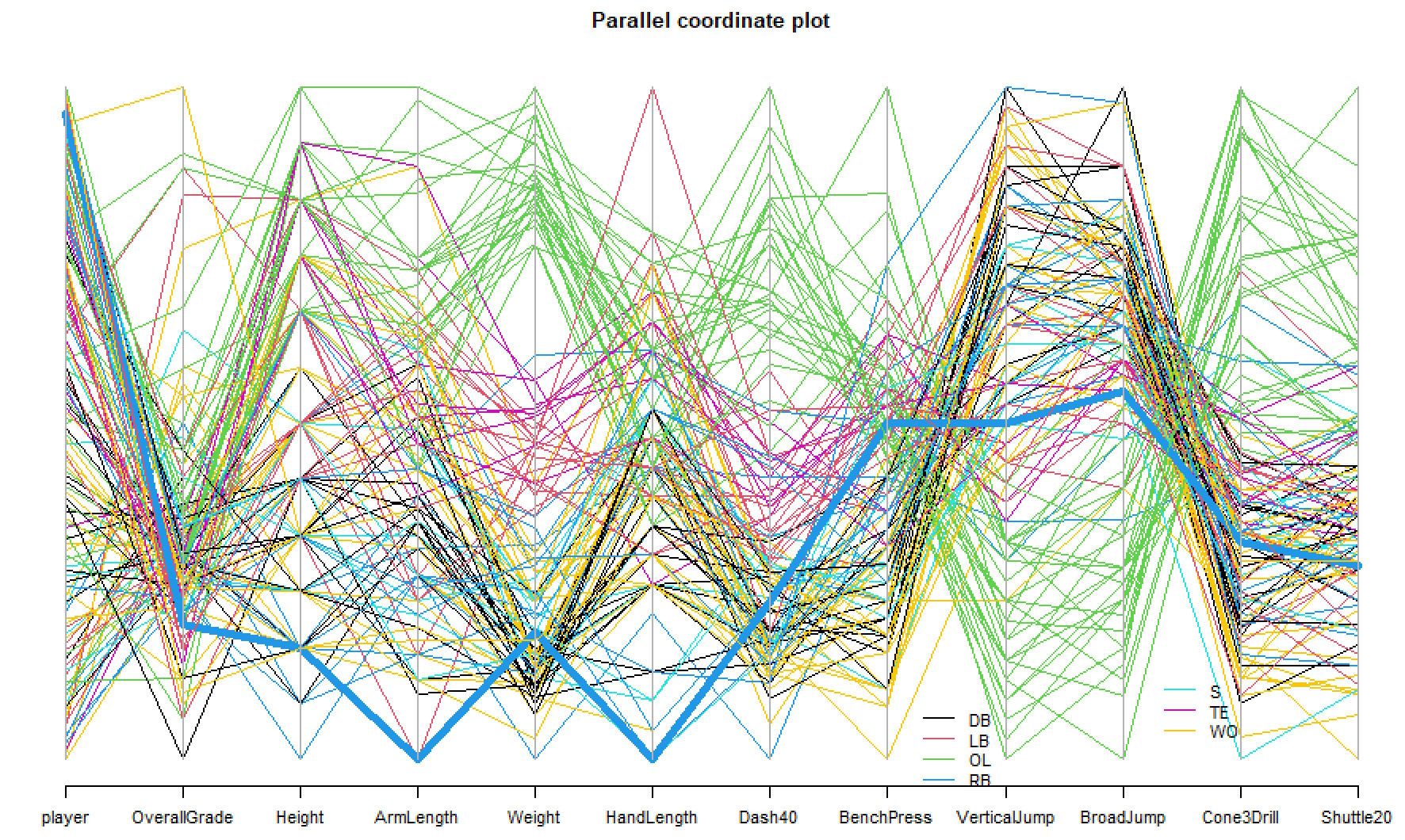
coordinate plot")

> legend(locator(1), legend = levels(fb$Position)[1:4], lty = 1, col =

color.position[1:4], cex = 1, bty = "n")

> legend(locator(1), legend = levels(fb$Position)[5:7], lty = 1, col =

color.position[5:7], cex = 1, bty = "n")



This particular player has a small arm and hand length. Because these variables have a positive coefficient for PC #2, this is why observation #124 has the smallest PC #2 value.

The examination of the parallel coordinate plot suggested observations like 122 stood out from the rest. Why didn’t this occur here? These observations stood out with respect to the OverallGrade variable, which was not included for the PCA. Unfortunately, we do not know exactly how this variable is constructed.

* 1. (4 points) Again using the plot from e), discuss the differences between the PC scores with respect to the position of the player. Include characteristics of the PCs in your discussion.

The OL players all have large PC #1 values relative to the other players. As shown on the parallel coordinate plot, these players had low jumping values while having larger weights and timed variables values. Because the coefficients for jumping variables are negative and for weight and timed variables are positive, this leads to the small PC #1 values.

The DB players all have relatively small PC #1 values. These players tend to have long jumps (negative coefficients for PC #1) while also being smaller in height and weight, smaller bench press values, and faster timed variables (positive coefficients for PC #1).

For those with a football background, having the TE players as the next closest to OL players should be expected because tight ends tend to line up next to the offensive lineman before a play starts.

* 1. (4 points) Construct a 3D plot of the PC scores using plot3d() and common x-, y-, and z-axis limits. Color the plotting points using the position of players (same color scheme in 1)b)). Are there any new items that were not apparent with the bubble plot? Explain.

> library(rgl)

> common.limits <- c(min(score.cor[,1:3]), max(score.cor[,1:3]))

> plot3d(x = score.cor[,1], y = score.cor[,2], z = score.cor[,3], xlab = "PC #1",

ylab = "PC #2", zlab = "PC #3", type = "h", xlim = common.limits, ylim =

common.limits, zlim = common.limits)

> plot3d(x = score.cor[,1], y = score.cor[,2], z = score.cor[,3], add = TRUE, col =

Position.color, size = 6)

> persp3d(x = common.limits, y = common.limits, z = matrix(data = c(0,0,0,0), nrow

= 2, ncol = 2), add = TRUE, col = "green") #Put a plane on the plot

> grid3d(side = c("x", "y", "z"), col = "lightgray")

> text3d(x = score.cor[,1], y = score.cor[,2], z = score.cor[,3] + 0.2, text =

1:nrow(fb))

A diagram of a cube with numbers and dots

AI-generated content may be incorrect.

Yes, there are some outliers more easily detected with this plot. In particular, observation #15 has a distinctly different PC #3 value in comparison to the other observations. This player is

> fb[15,]

Player College Position OverallGrade Height ArmLength Weight

15 Bodine, Russell North Carolina OL 5.3 75 32.5 310

HandLength Dash40 BenchPress VerticalJump BroadJump Cone3Drill Shuttle20

15 10 5.18 42 29 109 8.26 4.66

> score.cor[15,1:3]

Comp.1 Comp.2 Comp.3

-4.1835794 0.9203304 -2.6214240

Below is his line highlighted on the parallel coordinate plot:

> new.lwd <- c(rep(x = 1, times = 14), 5, rep(x = 1, times = 129-15))

> parcoord(x = fb2, col = Position.color, lwd = new.lwd, main = "Parallel

coordinate plot")

> legend(locator(1), legend = levels(fb$Position)[1:4], lty = 1, col =

color.position[1:4], cex = 1, bty = "n")

> legend(locator(1), legend = levels(fb$Position)[5:7], lty = 1, col =

color.position[5:7], cex = 1, bty = "n")

A diagram of lines and colors

AI-generated content may be incorrect.

This player has the largest bench press, but also the largest time in the Cone3Drill. The bench press value is what causes him to have a low PC #3 value.