APPENDIX: R PROGRAM FILE FOR NFLDRAFT.CSV

### PART A ###

#1a. Reading in the data file "NFLdraft.csv"

file <- read.csv("NFLdraft.csv", head=T, strip.white=T, stringsAsFactors=F)

View(file)

str(file)

#1b. Switching Long Snapper position to the Center position

gsub('LS', 'C', file$Pos)

#1c. Making a new factor to group players according to their positions

positions <- file$Pos

positionsFac <- factor(positions, levels = c('C', 'CB', 'DE', 'DT', 'FB', 'FS', 'ILB', 'LS', 'OG', 'OLB', 'OT','P','QB','RB', 'SS', 'TE', 'WR'), labels = c('Linemen', 'Small Backs', 'Linemen', 'Linemen',

'Big Back', 'Small Backs', 'Big Back', 'Big Back', 'Linemen', 'Big Back', 'Linemen', 'Big Back', 'Big Back', 'Big Back', 'Big Back', 'Linemen', 'Small Backs'))

positionsFac <- factor(positionsFac)

file$Pos2 <- positionsFac

Pos2 <- file$Pos2

#1d. Converting ft into inches

height <- file$Ht

heights <- gsub('-', '.', height)

heights <- as.numeric(heights)

file$Ht <- heights\*12

#1e. Splitting the last column into four - Team, Round, Pick, DraftYear

draft <- file$Drafted

final <- unlist(strsplit(draft, " / "))

final <- matrix(final, ncol = 4, byrow = TRUE)

final <- as.data.frame(final)

colnames(final) = c("Team", "Round", "Pick", "DraftYear")

file$Team <- as.character(final$Team)

file$Round <- as.character(final$Round)

file$Pick <- as.character(final$Pick)

file$DraftYear <- as.integer(as.character(final$DraftYear))

file$Drafted <- NULL

#1f. Storing pick as a number

file$Pick <- gsub("[a-zA-Z]\*", "", file$Pick) # Remove all letters

file$Pick <- as.integer(file$Pick)

#2a. Summary statistics for the Team column

team <- factor(file$Team)

summary(team)

#2b. Summary statistics for the CollegeTeam column

college = factor(file$CollegeTeam)

summary(college)

#2c. Player position displayed using a barplot in pareto order

pos <- factor(file$Pos)

barplot(table(factor(pos, levels = levels(pos)[order(-table(pos))])), main = "Number of Players picked by Player Position")

#2d. 5-number summary for player heights

summary(file$Ht)

#2e.Histogram for player heights

hist(file$Ht, main="Histogram for Player Heights", xlab="Heights (in inches)")

#2f. Shortest Players

ind <- which(grepl(min(file$Ht), file$Ht))

file[ind,]$Name

#2g. Plot of 40-yd dash time against Weight

Weights <- file$Wt

FortyYardDashTime <- file$Yd40

plot(Weights, FortyYardDashTime)

#2h. Plot of 3-Cone Drill Time vs shuttle drill time

plot(file$Shuttle,file$Cone3)

#2i. Plot of 3-Cone Drill Time vs shuttle drill time, grouped by position

ShuttleDrillTime <- file$Shuttle

ThreeConeDrillTime <- file$Cone3

plot(ShuttleDrillTime, ThreeConeDrillTime, pch=c(1,19,2), col=c("black", "red", "green")[file$Pos2])

legend(x="topright", legend = levels(file$Pos2), col=c("black", "red", "green"), pch=c(1,19,2))

#2j. Broad jump score Vs. Bench press for linemen

positionSub <- subset(file$Pos2, file$Pos2 == "Linemen")

BenchPressScore <- file$Bench

BroadJumpScore <- file$Broad

plot(BenchPressScore, BroadJumpScore, col=positionSub)

#2k. Finding player with shortest broad jump

minimum <- min(file$Broad, na.rm=TRUE)

ind <- which(grepl(minimum, file$Broad))

#players with the smallest scores

names <-file[ind,]$Name

#their weights

weights <- file[ind,]$Wt

#index of the lighest player

index <- which(grepl(min(weights), weights))

names[index]

file$Round[ind[2]]

#2l. Finding player with longest broad jump

maximum <- max(file$Broad, na.rm=TRUE)

ind <- which(grepl(maximum, file$Broad))

#player with the longest scores

names <-file[ind,]$Name

names

file$Bench[ind]

### PART B ###

#1a. Fitting a linear model

fit <- lm(Cone3 ~ Shuttle, data=file)

summary(fit)

#1c. 92% Confidence Interval

confint(fit, level= 0.92)

#1d. 95% interval for shuttle time = 4.5s

new <- data.frame(Shuttle = 4.5)

predict(fit, new, interval = "confidence") #default is 95% interval

#1e. 95% interval for new player with shuttle time = 4.7s

newData <- data.frame(Shuttle = 4.7)

predict(fit, newData, interval = "prediction")

#2. Testing correlation of draft picks with different test methods

cor(file$Yd40, file$Pick, use="complete")

cor(file$Vertical,file$Pick , use="complete")

cor(file$Pick, file$Bench, use="complete")

cor(file$Pick, file$Broad, use="complete")

cor(file$Pick, file$Cone3, use="complete")

cor(file$Pick, file$Shuttle, use="complete")