

NFL Combine Project

David Deniziuk

11/4/2021

Imports

```
#preemptively loading various libraries
library(readr)
library(tidyverse)
library(dplyr)
library(tidyr)
library(leaps)
library(devtools)
library(data.table)
```

```
data <- read_csv("/Users/david/Code/STAT 1341/midterm project/combine_data_since_2000_PROCESSED_2018-04-
```

Quarterback Section

Creation of Dataframe

predictor variables

```
#modification of dataset of predictor variables
data <- data %>% filter(Pos == "QB", Forty != "NA", Shuttle != "NA", BroadJump != "NA", Cone != "NA", V
data <- data %>%
  select(Player = Player, CombineYear = Year, Forty = Forty, Vertical = Vertical, BroadJump = BroadJump
head(data, 10)
```

```
## # A tibble: 10 x 8
##   Player      CombineYear Forty Vertical BroadJump Cone Shuttle Pick
##   <chr>          <dbl> <dbl>   <dbl>    <dbl> <dbl>  <dbl> <dbl>
## 1 Tom Brady      2000  5.28   24.5      99    7.2    4.38  199
## 2 Todd Husak     2000  5.2    27.5      99    7.17   4.41  202
## 3 Chad Pennington 2000  4.81   33.5     111    7.12   4.16   18
## 4 Tim Rattay     2000  4.91   25.5     100    7.34   4.21  212
## 5 Chris Redman   2000  5.37   26.5      98    7.8    4.78   75
## 6 Spergon Wynn   2000  4.91    34     108    7.71   4.59  183
## 7 Josh Booty     2001  4.9    29.5     105    7.54   4.48  172
## 8 Drew Brees     2001  4.83    32     105    7.09   4.21   32
## 9 Quincy Carter  2001  4.58   35.5     113    7.12   4.12   53
## 10 Josh Heupel   2001  4.85   37.5     114    7.18   4.15  177
```

response variables

```
#creation and modification of dataset of response variables
data_ <- read_csv("/Users/david/Code/STAT 1341/midterm project/archive/QBStats_all.csv") #reading in c
```

```
data_ <- data_ %>% filter(year >= 2000) %>% group_by(qb, year) %>% summarise(rate = mean(rate), games =
  filter(games >= 10) # filtered for seasons where QB played at least 10 games in order to select for
data_ %>% arrange(desc(year))
```

```
## # A tibble: 505 x 4
## # Groups:   qb [133]
##   qb                year rate games
##   <chr>             <dbl> <dbl> <int>
## 1 Aaron RodgersA. Rodgers 2016 106.    16
## 2 Alex SmithA. Smith    2016 92.9    15
## 3 Andrew LuckA. Luck     2016 97.1    15
## 4 Andy DaltonA. Dalton   2016 93.7    16
## 5 Ben RoethlisbergerB. Roethlisberger 2016 95.2    14
## 6 Blake BortlesB. Bortles 2016 78.3    16
## 7 Brock OsweilerB. Osweiler 2016 71.1    15
## 8 Cam NewtonC. Newton    2016 76.1    14
## 9 Carson PalmerC. Palmer  2016 88.9    15
## 10 Carson WentzC. Wentz   2016 80.4    16
## # ... with 495 more rows
```

merging both dataframes

```
#this code is designed to merge dataframes together. It was tricky due to the player names being listed
data <- data.table(data, key = "Player") #set up key to join
data_ <- data.table(data_, key = "qb") #set up key to join
```

```
temp <- data[,data[,agrep(Player, qb)], by = .(Player, CombineYear, Forty, Vertical, BroadJump, Cone, Shuttle)
temp$qb <- NULL #removed duplicate column
```

final dataframe

```
#this code creates a temporary dataframe in order to get QBs from a given QB's first year.
```

```
firstyear_rate <- temp %>%
  group_by(Player) %>% filter(year == min(year)) %>% #selects for minimum year
  arrange(Player) #orders alphabetically
colnames(firstyear_rate)[10] <- "firstyearrate" #rename variable
colnames(firstyear_rate)[9] <- "firstyear" #rename variable
```

```
final <- temp %>% group_by(Player, CombineYear, Forty, Vertical, BroadJump, Cone, Shuttle) %>% summarise(
  arrange(Player)
final$firstyearrate <- firstyear_rate$firstyearrate #add first year variable to dataframe
final$CombineYear <- NULL #remove unnecessary variable
```

```
#final dataframe for modeling
head(final, 10)
```

```
## # A tibble: 10 x 8
## # Groups:   Player, Forty, Vertical, BroadJump, Cone [10]
##   Player          Forty Vertical BroadJump Cone Shuttle rate firstyearrate
##   <chr>          <dbl>   <dbl>   <dbl> <dbl>   <dbl> <dbl>   <dbl>
## 1 Andrew Luck    4.59     36     124  6.8     4.28  89.9    77.7
## 2 Andy Dalton    4.83    29.5    106  6.93    4.27  91.4    81.1
## 3 Blaine Gabbert 4.61    33.5    120  6.84    4.26  75.9    68.8
## 4 Blake Bortles  4.93    32.5    115  7.08    4.21  78.9    69.1
```

```
## 5 Bruce Gradkowski 4.59 34 107 7.12 4.16 68.5 68.5
## 6 Cam Newton 4.56 35 126 6.92 4.18 87.9 88.6
## 7 Carson Wentz 4.77 30.5 118 6.86 4.15 80.4 80.4
## 8 Chad Henne 4.92 25.5 106 7.17 4.4 76.4 76.4
## 9 Chad Pennington 4.81 33.5 111 7.12 4.16 92.0 98.4
## 10 Charlie Frye 4.79 33 113 6.94 4.08 77.2 77.2
```

Quarterback Descriptive Statistics

descriptive statistics: predictor variables

```
attach(final) #attach dataset for ease of use
mean(Forty) #various means
```

```
## [1] 4.777317
```

```
mean(Vertical)
```

```
## [1] 32.37805
```

```
mean(BroadJump)
```

```
## [1] 112.4146
```

```
mean(Cone)
```

```
## [1] 7.072439
```

```
mean(Shuttle)
```

```
## [1] 4.245122
```

```
sd(Forty) #various standard deviations
```

```
## [1] 0.1919118
```

```
sd(Vertical)
```

```
## [1] 3.236087
```

```
sd(BroadJump)
```

```
## [1] 6.851918
```

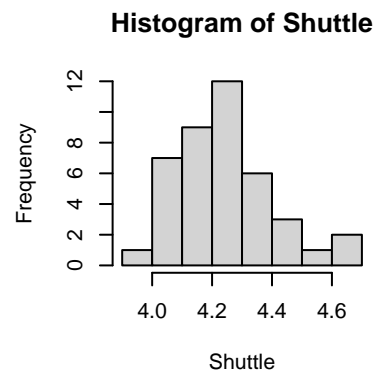
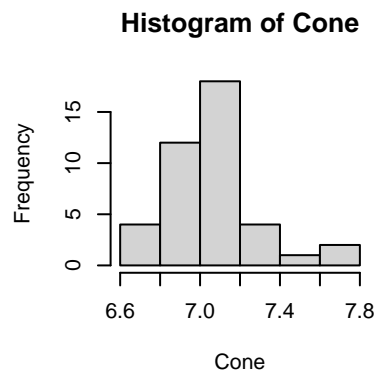
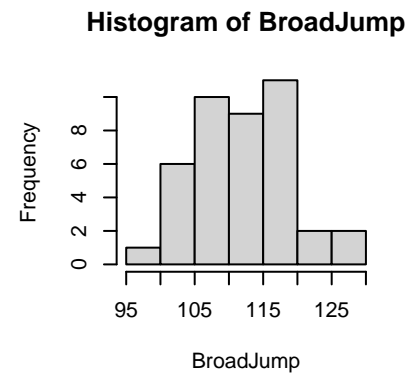
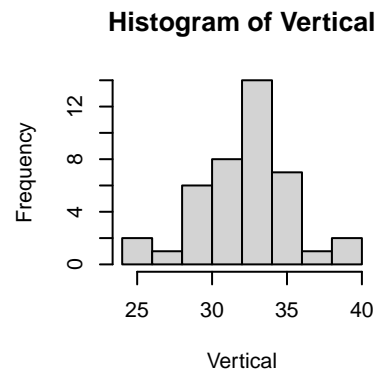
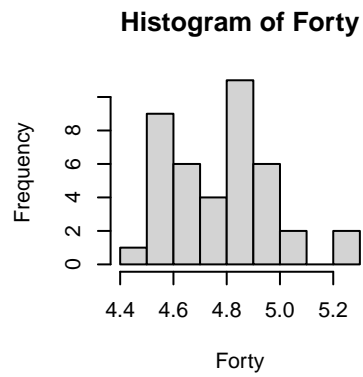
```
sd(Cone)
```

```
## [1] 0.2304537
```

```
sd(Shuttle)
```

```
## [1] 0.1604076
```

```
par(mfrow = c(2, 3))
hist(Forty) #various histograms
hist(Vertical)
hist(BroadJump)
hist(Cone)
hist(Shuttle)
```



descriptive statistics: response variables

```
mean(rate)      #means
```

```
## [1] 82.84862
```

```
mean(firstyearrate)
```

```
## [1] 79.17493
```

```
sd(rate)        #standard deviations
```

```
## [1] 10.09856
```

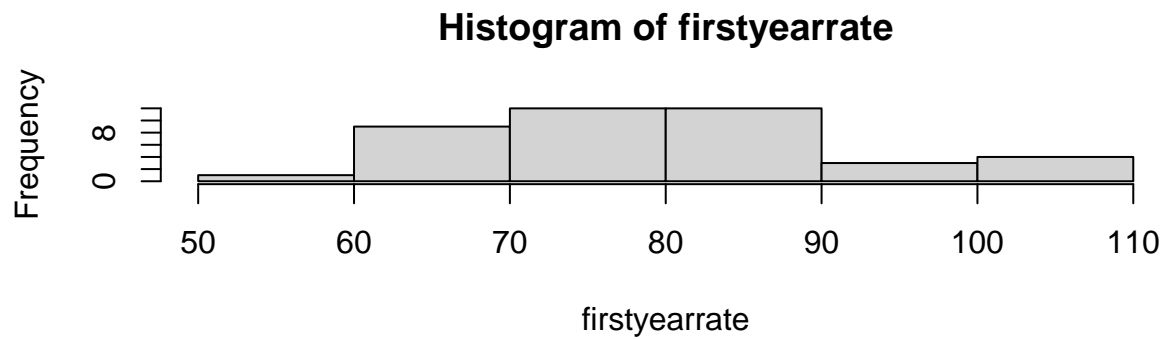
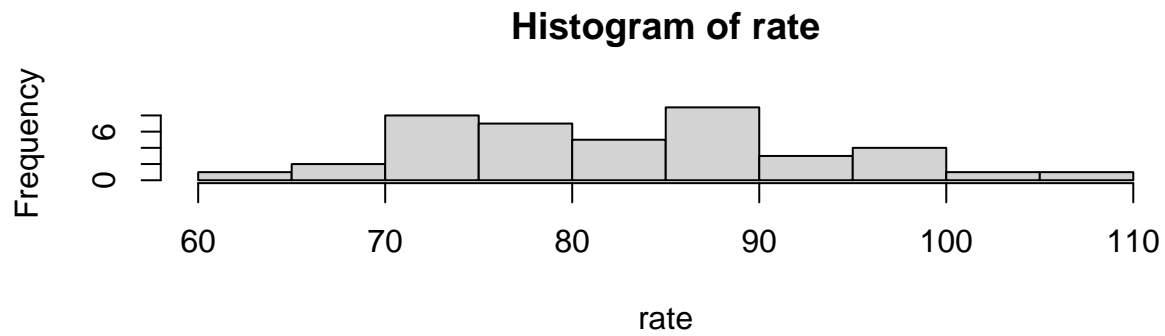
```
sd(firstyearrate)
```

```
## [1] 12.07074
```

```
par(mfrow = c(2, 1))
```

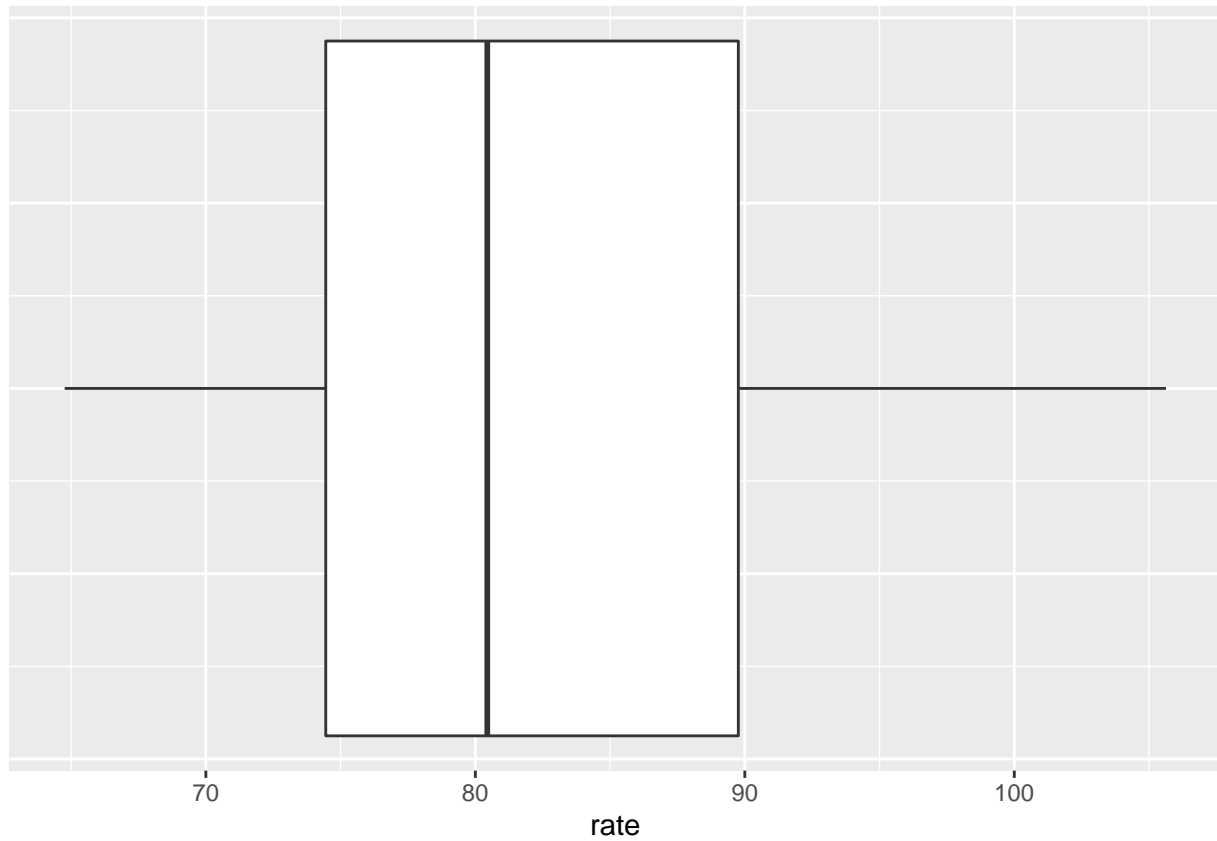
```
hist(rate)      # histograms
```

```
hist(firstyearrate)
```



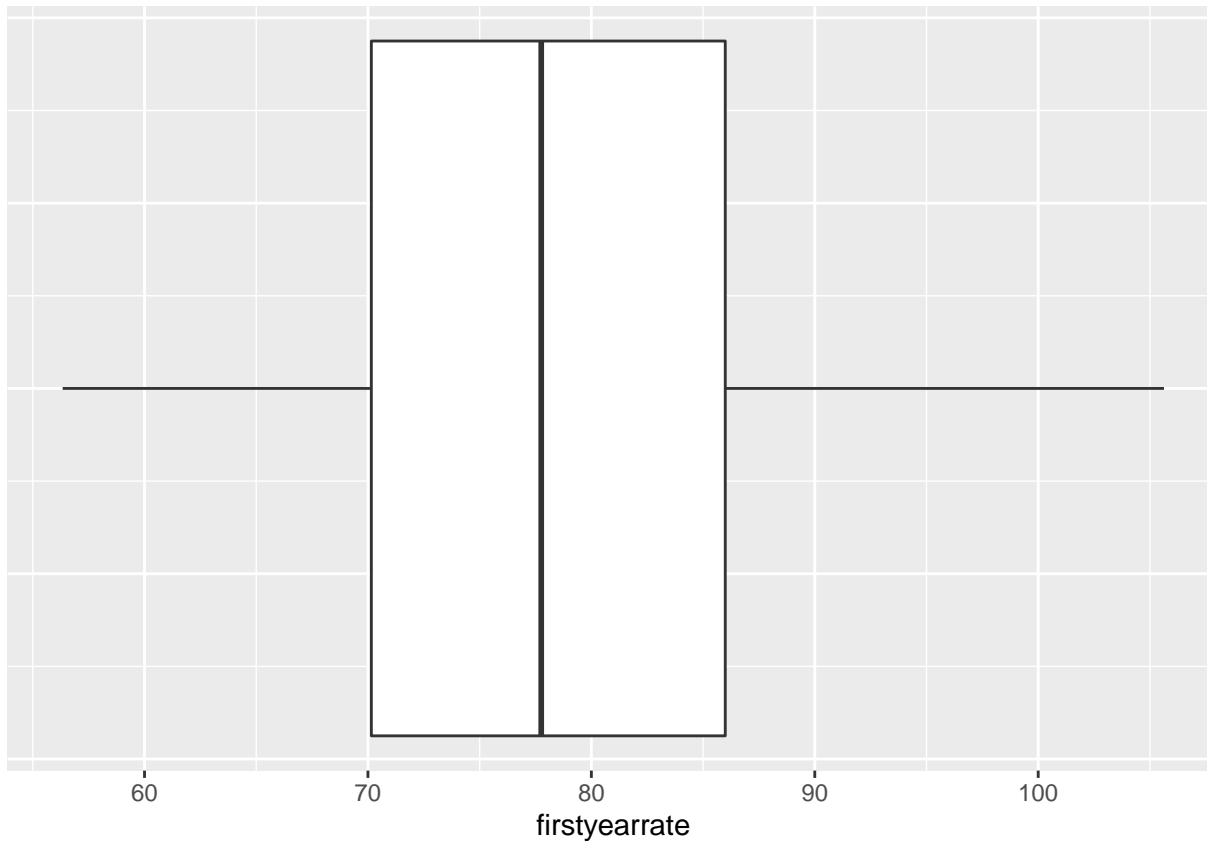
boxplot for career average QBR

```
ggplot(final, aes(y=rate)) + #creating boxplot
  geom_boxplot() +
  coord_flip() + #flipping coordinates
  theme(axis.ticks.y = element_blank(), axis.text.y = element_blank()) #removing axis label
```



boxplot for first year only QBR

```
ggplot(final, aes(y=firstyearrate)) + #creating boxplot
  geom_boxplot() +
  coord_flip()+ #flipping coordinates
  theme(axis.ticks.y = element_blank(), axis.text.y = element_blank()) #removing axis label
```



```
#detach(final)
```

Quarterback Modeling

fitting model 1: career average

```
#fitting model on average QBR across years. "rate"
fit.full <- lm(rate ~ Forty+Vertical+BroadJump+Cone+Shuttle, data=final) #fitting full model
fit.empty <- lm(rate~1, data=final) #fitting empty model
fit.final <- step(fit.empty, direction = 'both', scope = formula(fit.full), trace = 0) #iterating thro
summary(fit.final) #printing summary of stepwise model
```

```
##
## Call:
## lm(formula = rate ~ 1, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -18.089  -8.399  -2.405   6.911  22.783
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   82.849      1.577   52.53  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.1 on 40 degrees of freedom
```

fitting model 2: first year only

```
#fitting model on first year QBR. "firstyearrate"
fit.full_ <- lm(firstyearrate ~ Forty+Vertical+BroadJump+Cone+Shuttle, data=final) #fitting full model
fit.empty_ <- lm(firstyearrate~1, data=final) #fitting empty model
fit.final_ <- step(fit.empty_, direction = 'both', scope = formula(fit.full_), trace = 0) #iterating t
summary(fit.final_) #printing summary of stepwise model

##
## Call:
## lm(formula = firstyearrate ~ 1, data = final)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -22.839  -9.022  -1.419   6.818  26.456
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   79.175      1.885     42  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.07 on 40 degrees of freedom
```

Runningback Section

Creation of Dataframe

predictor variables

```
#creation and modification of dataset of predictor variables
dataRB <- read_csv("/Users/david/Code/STAT 1341/midterm project/combine_data_since_2000_PROCESSED_2018-2019.csv")
dataRB <- dataRB %>% filter(Pos == "RB", Forty != "NA", Shuttle != "NA", BroadJump != "NA", Cone != "NA")
dataRB <- dataRB %>%
  select(Player = Player, CombineYear = Year, Forty = Forty, Vertical = Vertical, BroadJump = BroadJump, Cone = Cone, Shuttle = Shuttle, Pick = Pick)
dataRB %>% arrange(desc(CombineYear)) #displaying dataframe ordered by combine year
```

```
## # A tibble: 167 x 8
##   Player          CombineYear Forty Vertical BroadJump Cone Shuttle Pick
##   <chr>          <dbl> <dbl>   <dbl>   <dbl> <dbl> <dbl> <dbl>
## 1 Dalvin Cook      2017  4.49    30.5     116  7.27  4.53   41
## 2 Wayne Gallman    2017  4.6     29.5     120  7.17  4.28  140
## 3 Brian Hill       2017  4.54    34      125  7.03  4.32  156
## 4 Aaron Jones      2017  4.56    37.5     127  6.82  4.2   182
## 5 Christian McCaffrey 2017  4.48    37.5     121  6.57  4.22   8
## 6 Elijah McGuire    2017  4.53    36      120  7.26  4.56  188
## 7 Jeremy McNichols  2017  4.49    35.5     121  6.93  4.28  162
## 8 Samaje Perine     2017  4.65    33      116  7.26  4.37  114
## 9 Joe Williams     2017  4.41    35      125  7.19  4.19  121
## 10 Jamaal Williams  2017  4.59    30      123  7.25  4.53  134
## # ... with 157 more rows
```

response variables

```
#creation and modification of dataset of response variables
dataRB_ <- read_csv("/Users/david/Code/STAT 1341/midterm project/Game_Logs_Runningback.csv") #reading
```



```
dataRB_
```

```
## # A tibble: 67,661 x 25
##   `Player Id`   Name      Position  Year Season   Week `Game Date` `Home or Away`
##   <chr>         <chr>    <chr>    <dbl> <chr>    <dbl> <chr>         <chr>
## 1 bobbyfowler/~ Fowler~ NA      1985 Regula~    7 10/20       Away
## 2 bobbyfowler/~ Fowler~ NA      1985 Regula~    8 10/27       Home
## 3 bobbyfowler/~ Fowler~ NA      1985 Regula~    9 11/03       Away
## 4 bobbyfowler/~ Fowler~ NA      1985 Regula~   10 11/10       Home
## 5 bobbyfowler/~ Fowler~ NA      1985 Regula~   11 11/17       Away
## 6 bobbyfowler/~ Fowler~ NA      1985 Regula~   12 11/24       Away
## 7 bobbyfowler/~ Fowler~ NA      1985 Regula~   13 12/01       Home
## 8 bobbyfowler/~ Fowler~ NA      1985 Regula~   14 12/08       Away
## 9 bobbyfowler/~ Fowler~ NA      1985 Regula~   15 12/15       Home
## 10 bobbyfowler/~ Fowler~ NA      1985 Regula~   16 12/22       Home
## # ... with 67,651 more rows, and 17 more variables: Opponent <chr>,
## #   Outcome <chr>, Score <chr>, Games Played <dbl>, Games Started <chr>,
## #   Rushing Attempts <chr>, Rushing Yards <chr>, Yards Per Carry <chr>,
## #   Longest Rushing Run <chr>, Rushing TDs <chr>, Receptions <chr>,
## #   Receiving Yards <chr>, Yards Per Reception <chr>, Longest Reception <chr>,
## #   Receiving TDs <chr>, Fumbles <chr>, Fumbles Lost <chr>
```

```
colnames(dataRB_)[16] <- "YardsPerCarry"
```

```
dataRB_$YardsPerCarry <- as.numeric(dataRB_$YardsPerCarry)
```

```
dataRB_ <- dataRB_ %>% filter(Year >= 2000) %>% filter(Season == "Regular Season") %>% drop_na(YardsPerCarry)
  filter(games >= 10) # filtered for seasons where RB played at least 10 games in order to select
```

```
dataRB_ %>% arrange(desc(Year)) #displaying dataframe ordered by year
```

```
## # A tibble: 734 x 4
## # Groups:   Player Id [224]
##   `Player Id`   Year YardsPerCarry games
##   <chr>         <dbl>         <dbl> <int>
## 1 aaronripkowski/2552477 2016         3.59    10
## 2 alfredblue/2543600    2016         4.51    14
## 3 alfredmorris/2533457  2016         4.3     13
## 4 bilalpowell/2495328   2016         5.6     16
## 5 carloshyde/2543743    2016         4.48    13
## 6 charcandrickwest/2550268 2016         2.84    14
## 7 chrisivory/2507999    2016         3.9     11
## 8 christhompson/2540011  2016         5.62    16
## 9 christinemichael/2539322 2016         3.89    15
## 10 damienwilliams/2550512 2016         3.98    13
## # ... with 724 more rows
```

merging both data frames

```
#this code is designed to merge dataframes together. It was tricky due to the player names being listed
```

```
dataRB <- data.table(dataRB, key = "Player") #set up key so join works
```

```
dataRB_ <- data.table(dataRB_, key = "Player Id") #set up key so join works
```

```
tempRB <- dataRB[,dataRB_[agrep(Player, `Player Id`, max.distance = 0.35)], by = .(Player, CombineYear,
```

```
tempRB <- tempRB %>% filter(Player != "Adrian Peterson-02") #removed duplicate Adrian Peterson lol
```

```
tempRB
```

```
##           Player CombineYear Forty Vertical BroadJump Cone Shuttle
## 1: Adrian Peterson-01      2002 4.68    34.0      119 7.59    4.60
## 2: Adrian Peterson-01      2002 4.68    34.0      119 7.59    4.60
## 3: Adrian Peterson-01      2002 4.68    34.0      119 7.59    4.60
## 4: Adrian Peterson-01      2002 4.68    34.0      119 7.59    4.60
## 5: Adrian Peterson-01      2002 4.68    34.0      119 7.59    4.60
## ---
## 305:      Toby Gerhart      2010 4.50    38.0      118 6.94    4.25
## 306:      Toby Gerhart      2010 4.50    38.0      118 6.94    4.25
## 307:      Toby Gerhart      2010 4.50    38.0      118 6.94    4.25
## 308:      Vick Ballard      2012 4.59    33.0      115 7.03    4.19
## 309: Wendell Smallwood      2016 4.47    33.5      120 6.83    4.28
##           Player Id Year YardsPerCarry games
## 1:  adrianpeterson/2505173 2005      5.318182    11
## 2:  adrianpeterson/2505173 2007      3.781250    16
## 3:  adrianpeterson/2507164 2007      5.300000    14
## 4:  adrianpeterson/2507164 2008      4.762500    16
## 5:  adrianpeterson/2507164 2009      4.431250    16
## ---
## 305:      tobygerhart/497176 2011      5.353333    15
## 306:      tobygerhart/497176 2012      3.742857    14
## 307:      tobygerhart/497176 2014      3.257143    14
## 308:      vickballard/2533012 2012      3.562500    16
## 309: wendellsmallwood/2555461 2016      3.541667    12
```

final dataframe

```
#this code creates a temporary dataframe in order to get yards per carry from a given RB's first year.
firstyear_YPC <- tempRB %>%
  group_by(Player) %>% filter(Year == min(Year)) %>% #selects for minimum year
  arrange(Player)      #orders alphabetically
colnames(firstyear_YPC)[10] <- "firstyearYardsPerCarry" #rename variable
colnames(firstyear_YPC)[9] <- "firstyear" #rename variable
firstyear_YPC <- firstyear_YPC %>% filter(Player != "Dan Herron") #remove mismatched observation

finalRB <- tempRB %>% group_by(Player, CombineYear, Forty, Vertical, BroadJump, Cone, Shuttle) %>% sum
finalRB$firstyearYardsPerCarry <- firstyear_YPC$firstyearYardsPerCarry #add first year variable to data
finalRB$CombineYear <- NULL #remove unnecessary variable

#final dataframe for modeling
#finalRB
```

Runningback Descriptive Statistics

descriptive statistics: predictor variables

```
attach(finalRB) #attach dataset for ease of use
mean(Forty)      #various means

## [1] 4.514868
mean(Vertical)

## [1] 35.33553
```

```

mean(BroadJump)

## [1] 119.7895
mean(Cone)

## [1] 7.010395
mean(Shuttle)

## [1] 4.258289
sd(Forty)      #various standard deviations

## [1] 0.09795909
sd(Vertical)

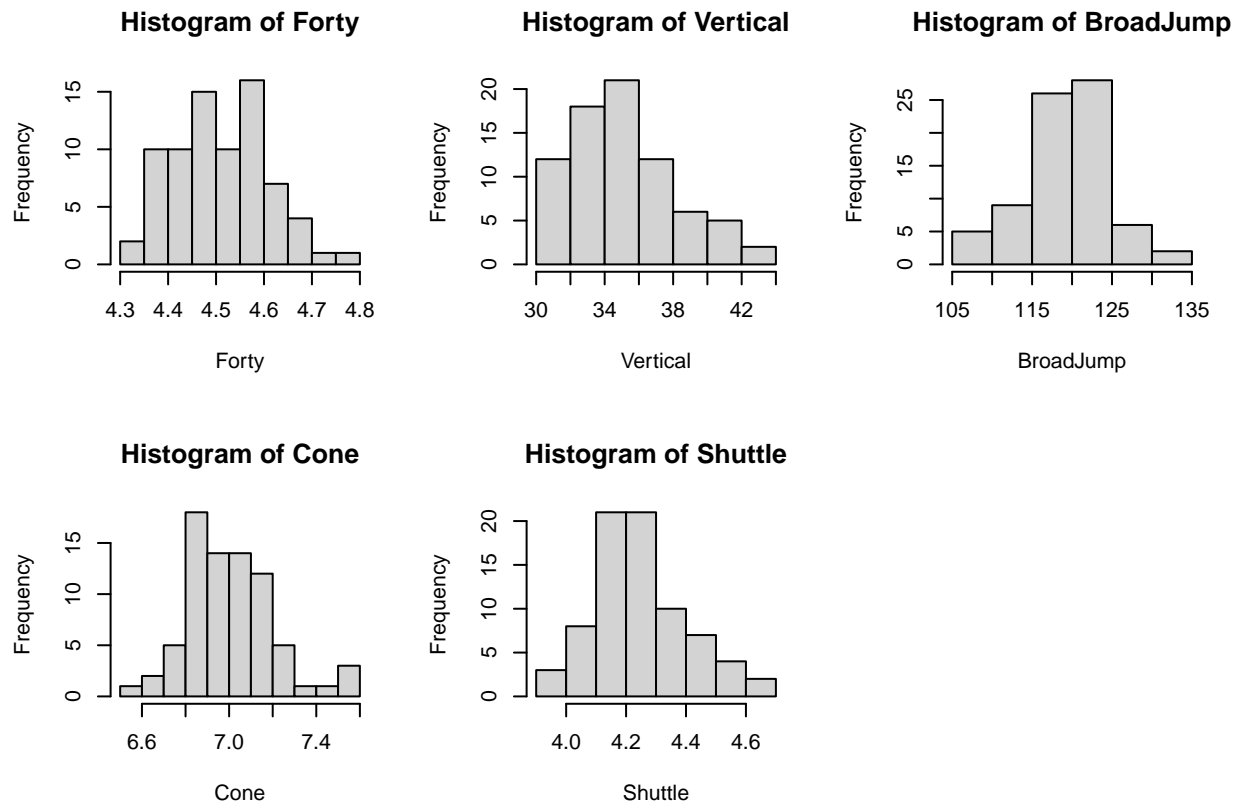
## [1] 3.066364
sd(BroadJump)

## [1] 5.446261
sd(Cone)

## [1] 0.2041205
sd(Shuttle)

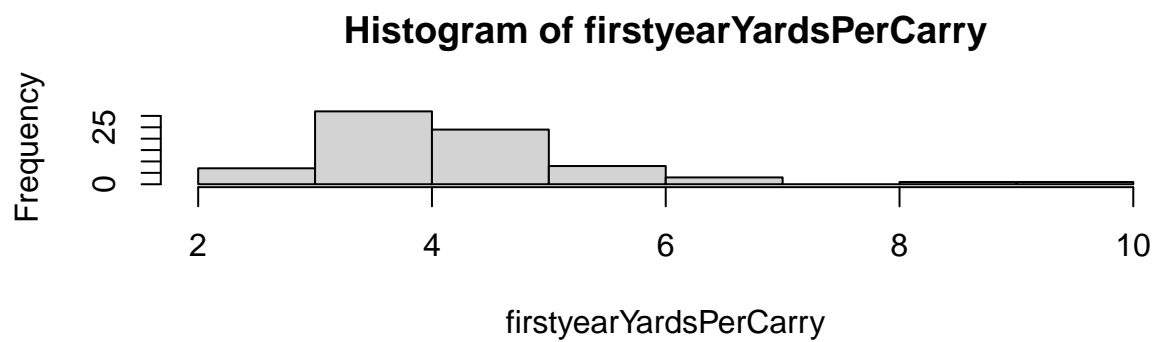
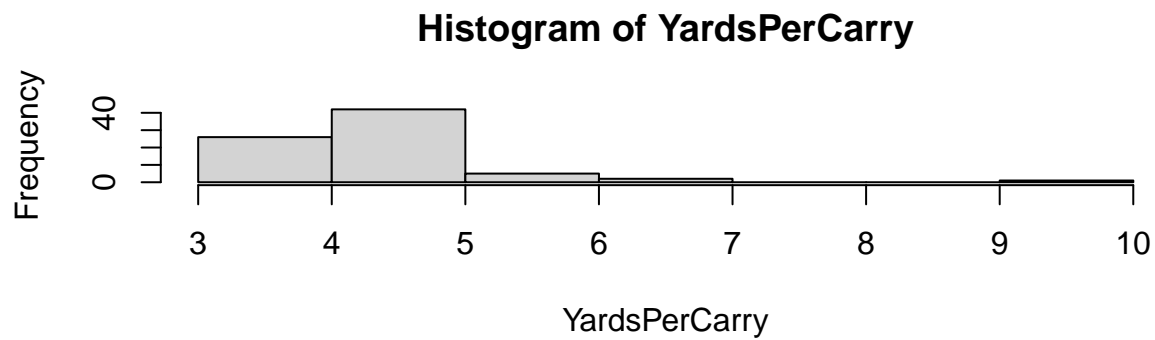
## [1] 0.1590839
par(mfrow = c(2, 3))
hist(Forty)      #various histograms
hist(Vertical)
hist(BroadJump)
hist(Cone)
hist(Shuttle)

```



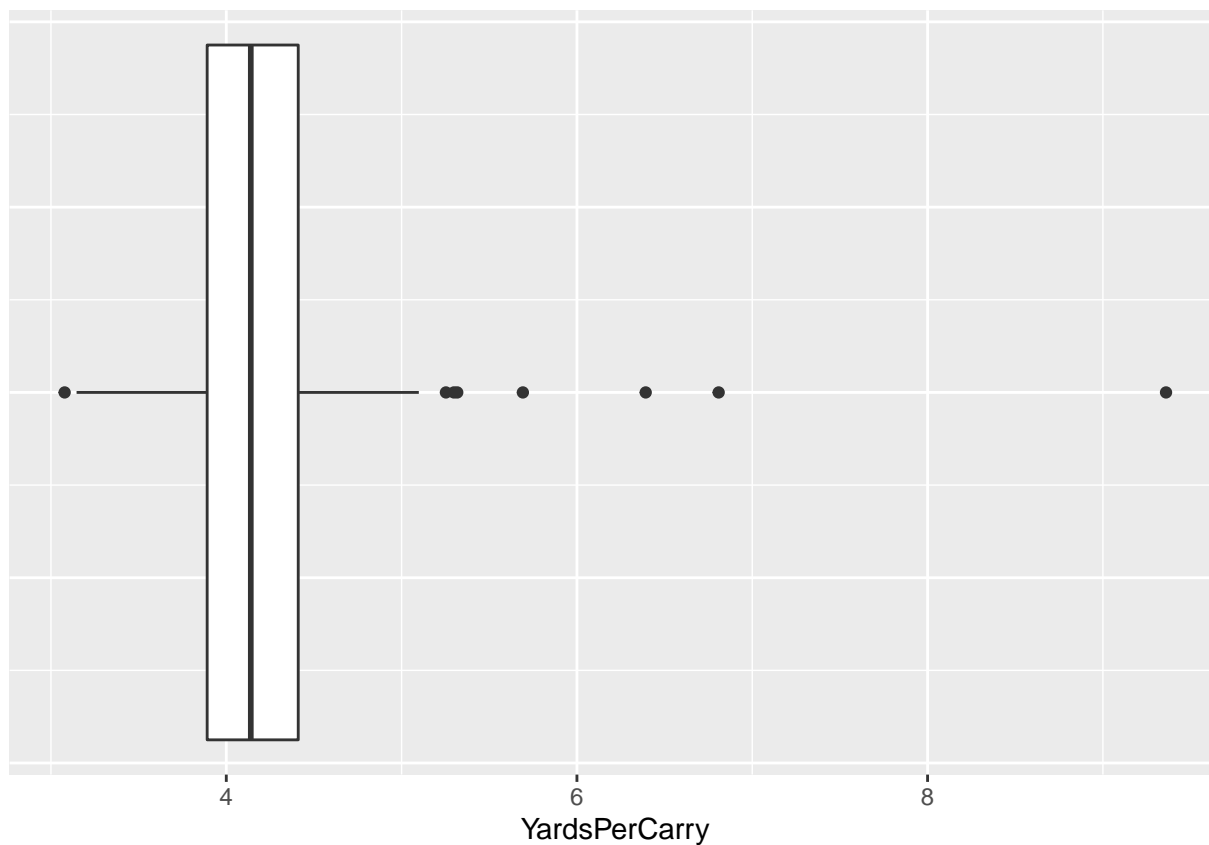
descriptive statistics: response variables

```
mean(YardsPerCarry)      #means
## [1] 4.26972
mean(firstyearYardsPerCarry)
## [1] 4.215346
sd(YardsPerCarry)        #standard deviations
## [1] 0.8741762
sd(firstyearYardsPerCarry)
## [1] 1.210113
par(mfrow = c(2, 1))
hist(YardsPerCarry)      # histograms
hist(firstyearYardsPerCarry)
```



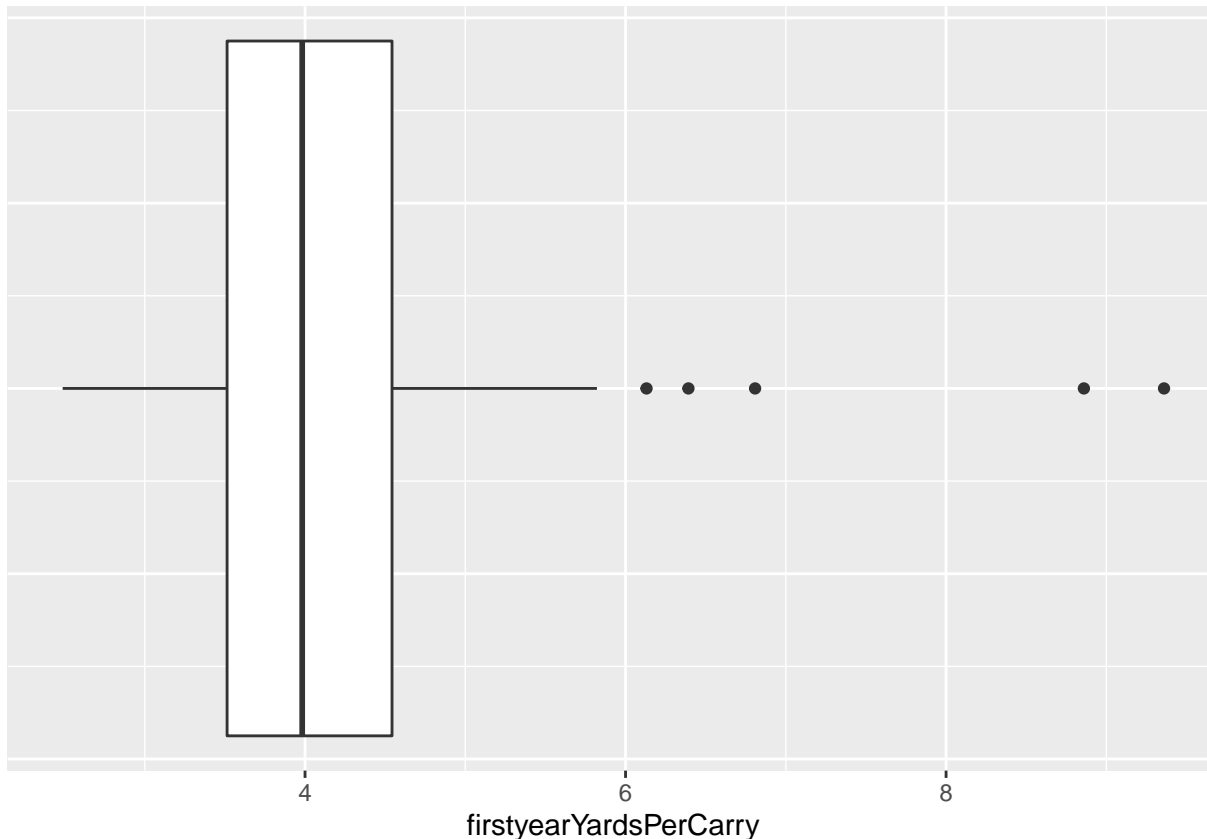
boxplot for career average YPC

```
ggplot(finalRB, aes(y=YardsPerCarry)) + #creating boxplot
  geom_boxplot() +
  coord_flip() + #flipping coordinates
  theme(axis.ticks.y = element_blank(), axis.text.y = element_blank()) #removing axis label
```



boxplot for first year only YPC

```
ggplot(finalRB, aes(y=firstyearYardsPerCarry)) +      #creating boxplot
  geom_boxplot() +
  coord_flip()+      #flipping coordinates
  theme(axis.ticks.y = element_blank(), axis.text.y = element_blank()) #removing axis label
```



```
detach(finalRB)
```

Runningback Modeling

fitting model 1: career average

```
#fitting model on average Yards Per Carry across years. "YardsPerCarry"
fit.fullRB <- lm(YardsPerCarry ~ Forty+Vertical+BroadJump+Cone+Shuttle, data=finalRB) #fitting full model
fit.emptyRB <- lm(YardsPerCarry~1, data=finalRB) #fitting empty model
fit.finalRB <- step(fit.emptyRB, direction = 'both', scope = formula(fit.fullRB), trace = 0) #iterative stepAIC
summary(fit.finalRB)
```

```
##
## Call:
## lm(formula = YardsPerCarry ~ Forty + BroadJump, data = finalRB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.2034 -0.4139 -0.1120  0.2648  4.5312
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  20.55166    6.09678   3.371  0.0012 **
## Forty        -2.58074    1.07332  -2.404  0.0187 *
## BroadJump    -0.03865    0.01931  -2.002  0.0490 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.8453 on 73 degrees of freedom
## Multiple R-squared:  0.08986,    Adjusted R-squared:  0.06492
## F-statistic: 3.604 on 2 and 73 DF,  p-value: 0.03217
```

fitting model 2: first year only

```
#fitting model on first year YPC. "firstyearYardsPerCarry"
fit.fullRB_ <- lm(firstyearYardsPerCarry ~ Forty+Vertical+BroadJump+Cone+Shuttle, data=finalRB) #fitti
fit.emptyRB_ <- lm(firstyearYardsPerCarry~1, data=finalRB) #fitting empty model
fit.finalRB_ <- step(fit.emptyRB_, direction = 'both', scope = formula(fit.fullRB_), trace = 0) #itera
summary(fit.finalRB_)
```

```
##
## Call:
## lm(formula = firstyearYardsPerCarry ~ 1, data = finalRB)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7278 -0.7017 -0.2331  0.3273  5.1447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.2153     0.1388   30.37  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.21 on 75 degrees of freedom
```