Descriptive Stats, Feature Selection, Estimates and Interpretations

Nicholas Paschall

2023-05-17

Descriptive Statistics

Summaries

```
length( which(combo.df$draft.fac==1) )
```

```
## [1] 210
```

```
length( which(combo.df$draft.fac==0) )
```

```
## [1] 117
```

```
nrow( combo.df )
```

```
## [1] 327
```

Ht

```
summary(combo.df$Ht)[c(1,3,6)]
```

```
##   Min. Median   Max.
##    5.1    6.2    6.6
```

```
sd(combo.df$Ht)
```

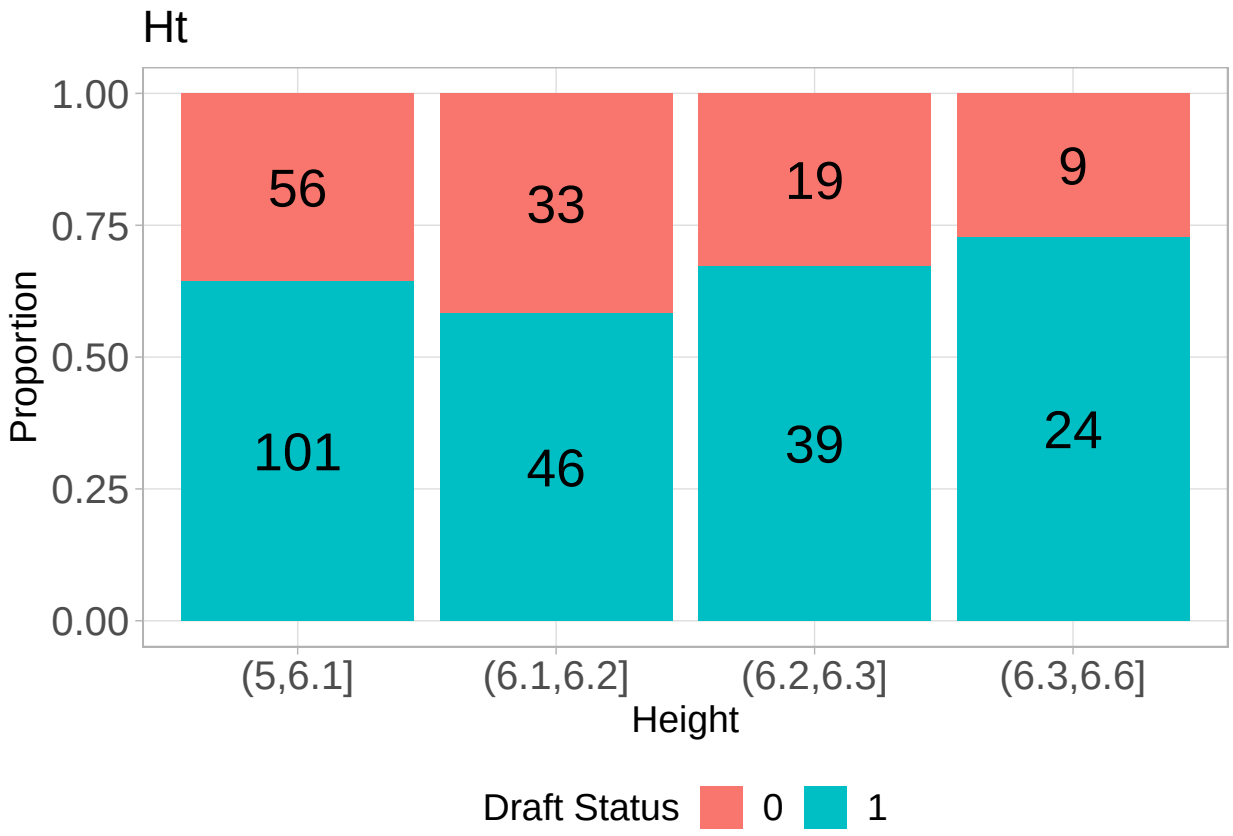
```
## [1] 0.305679
```

```
summary(glm(draft.fac~Ht, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.05580184
```

```
ht.hist <- ggplot(combo.df, aes(x = cut(Ht, breaks=c(5, 6.1, 6.2, 6.3, 6.6)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Height", title="Ht", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
ht.hist
```

```
## Warning: The dot-dot notation (‘..count..‘) was deprecated in ggplot2 3.4.0.

## Warning: Please use ‘after_stat(count)‘ instead.
```



Wt

```
summary(combo.df$Wt)[c(1,3,6)]
```

```
##    Min. Median    Max.
##     202    238     259
```

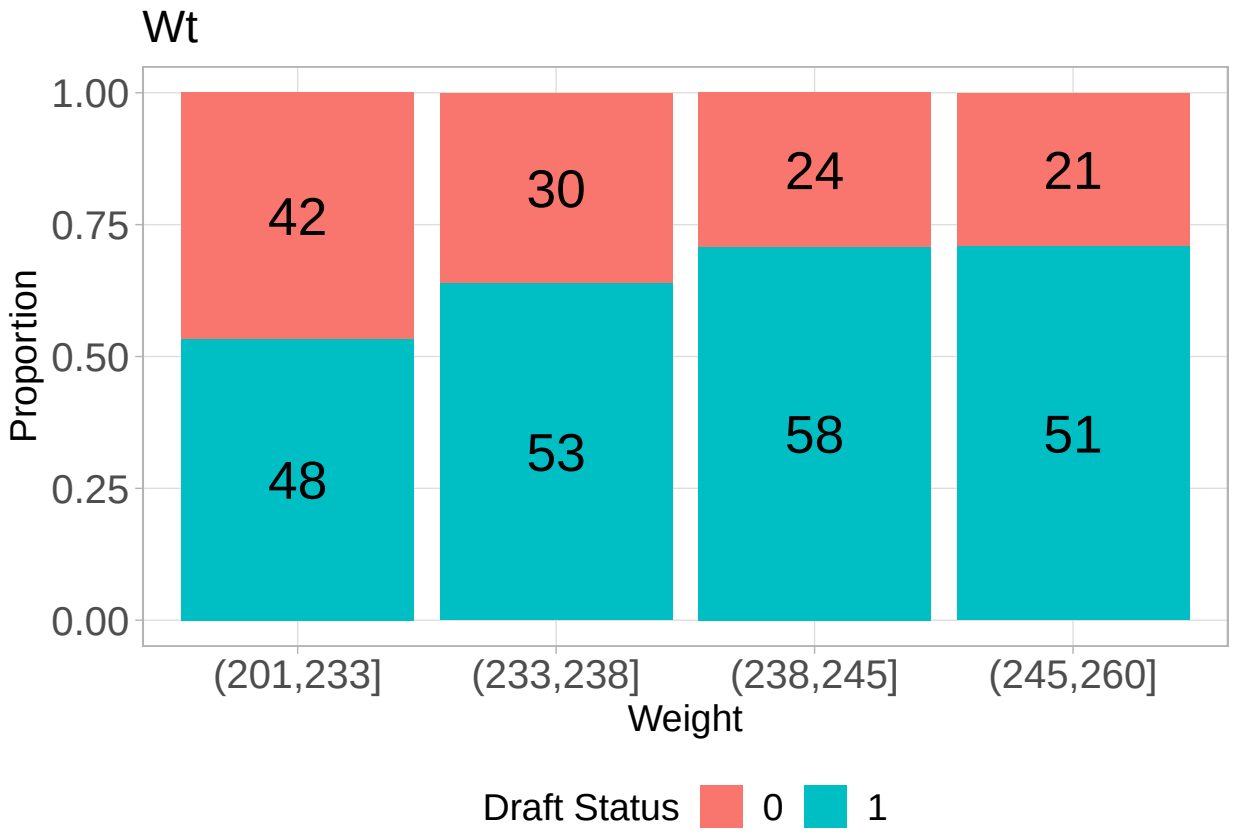
```
sd(combo.df$Wt)
```

```
## [1] 9.032063
```

```
summary(glm(draft.fac~Wt, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.0007178422
```

```
wt.hist <- ggplot(combo.df, aes(x = cut(Wt, breaks=c(201, 233, 238, 245, 260)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Weight", title="Wt", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
wt.hist
```



forty_yd

```r
summary(combo.df$forty_yd)[c(1,3,6)]
```

```
##   Min. Median   Max.
##   4.38   4.69   5.09
```

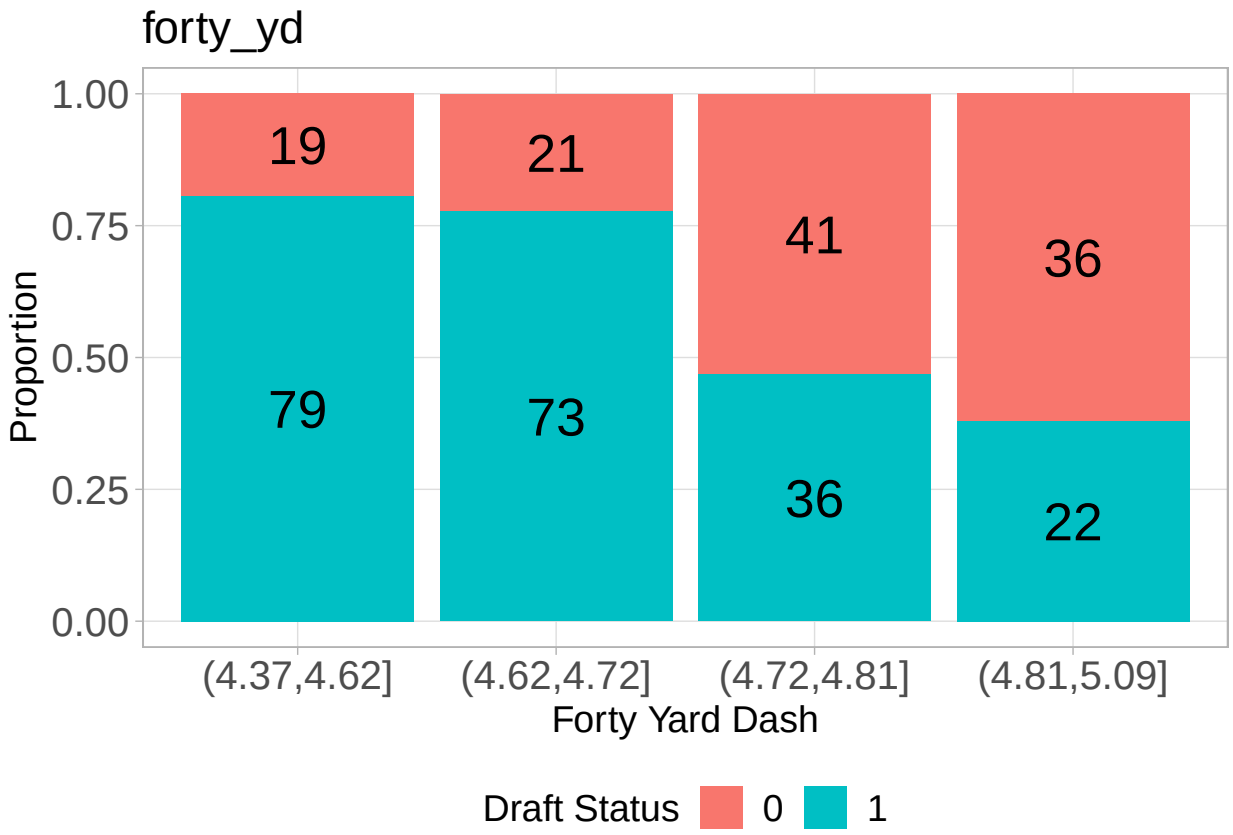
```r
sd(combo.df$forty_yd)
```

```
## [1] 0.1313206
```

```r
summary(glm(draft.fac~forty_yd, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 2.445901e-10
```

```r
forty_yd.hist <- ggplot(combo.df, aes(x = cut(forty_yd, breaks=c(4.37, 4.62, 4.72, 4.81, 5.09)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Forty Yard Dash", title="forty_yd", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
forty_yd.hist
```



vertical

```
summary(combo.df$vertical)[c(1,3,6)]
```

```
##   Min. Median   Max.
##   27.0   33.5   42.5
```

```
sd(combo.df$vertical)
```

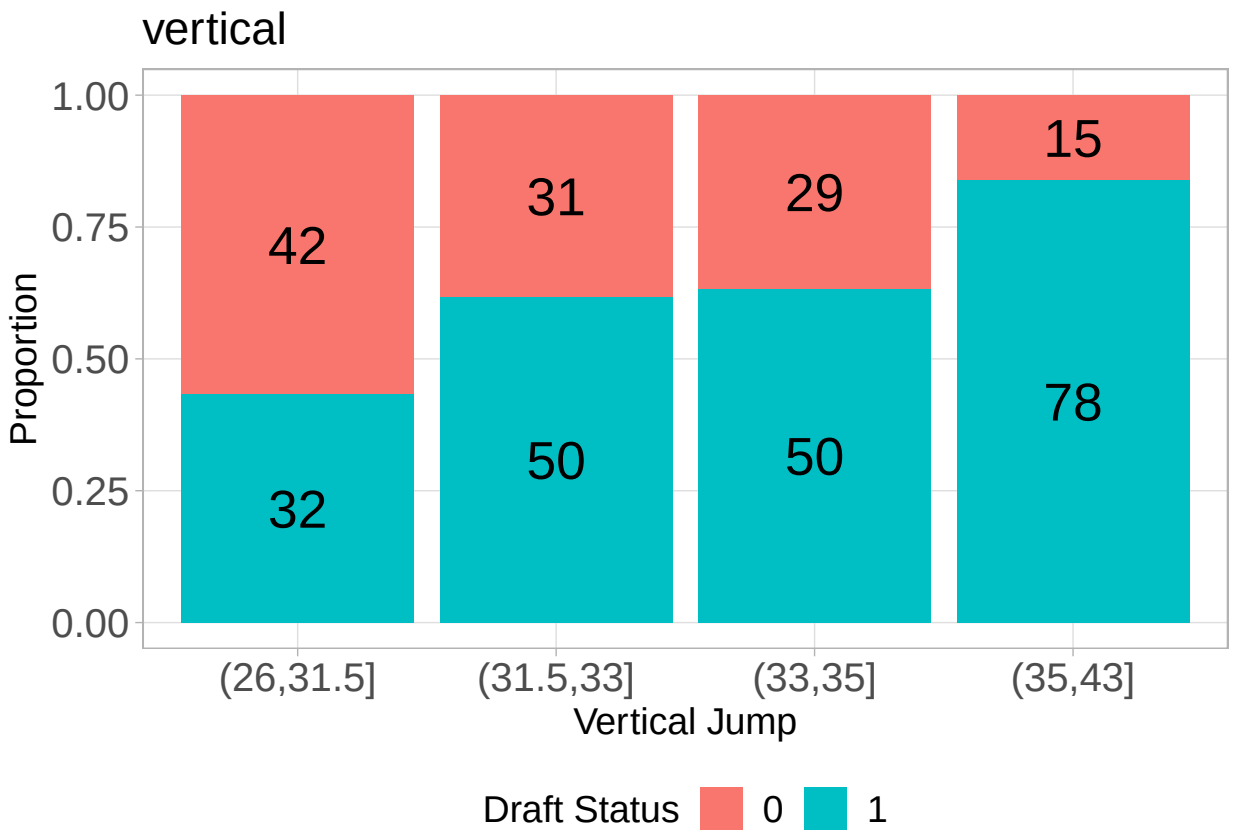
```
## [1] 3.055681
```

```
summary(glm(draft.fac~vertical, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 1.245793e-06
```

```
vertical.hist <- ggplot(combo.df, aes(x = cut(vertical, breaks=c(26, 31.5, 33, 35, 43)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Vertical Jump", title="vertical", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
```

```
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
vertical.hist
```



bench

```
summary(combo.df$bench)[c(1,3,6)]
```

```
##   Min. Median   Max.
##     11     21     35
```

```
sd(combo.df$bench)
```

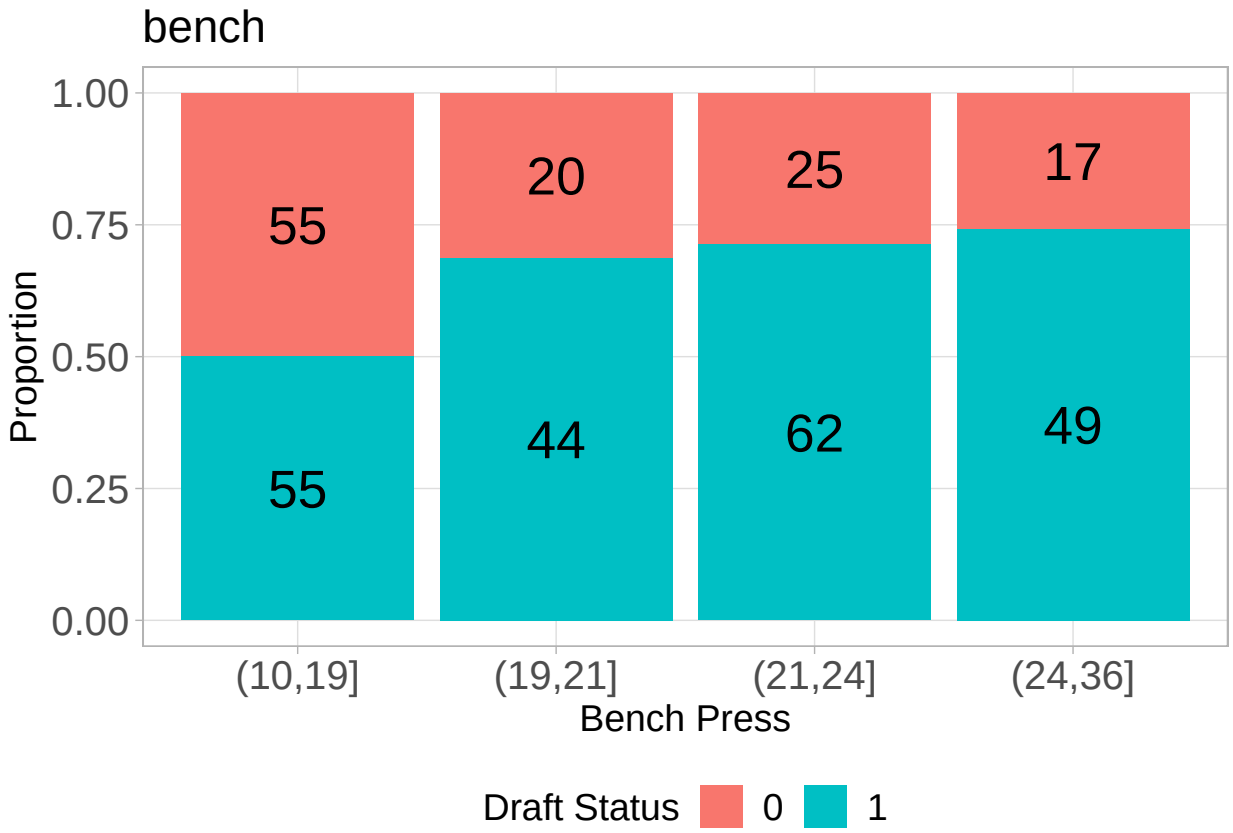
```
## [1] 3.984063
```

```
summary(glm(draft.fac~bench, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.0002307628
```

```
bench.hist <- ggplot(combo.df, aes(x = cut(bench, breaks=c(10, 19, 21, 24, 36)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text", aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
```

```
    labs(y="Proportion", x="Bench Press", title="bench", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
bench.hist
```



broad_jump

```
summary(combo.df$broad_jump)[c(1,3,6)]
```

```
##    Min. Median    Max.
##     104    119     139
```

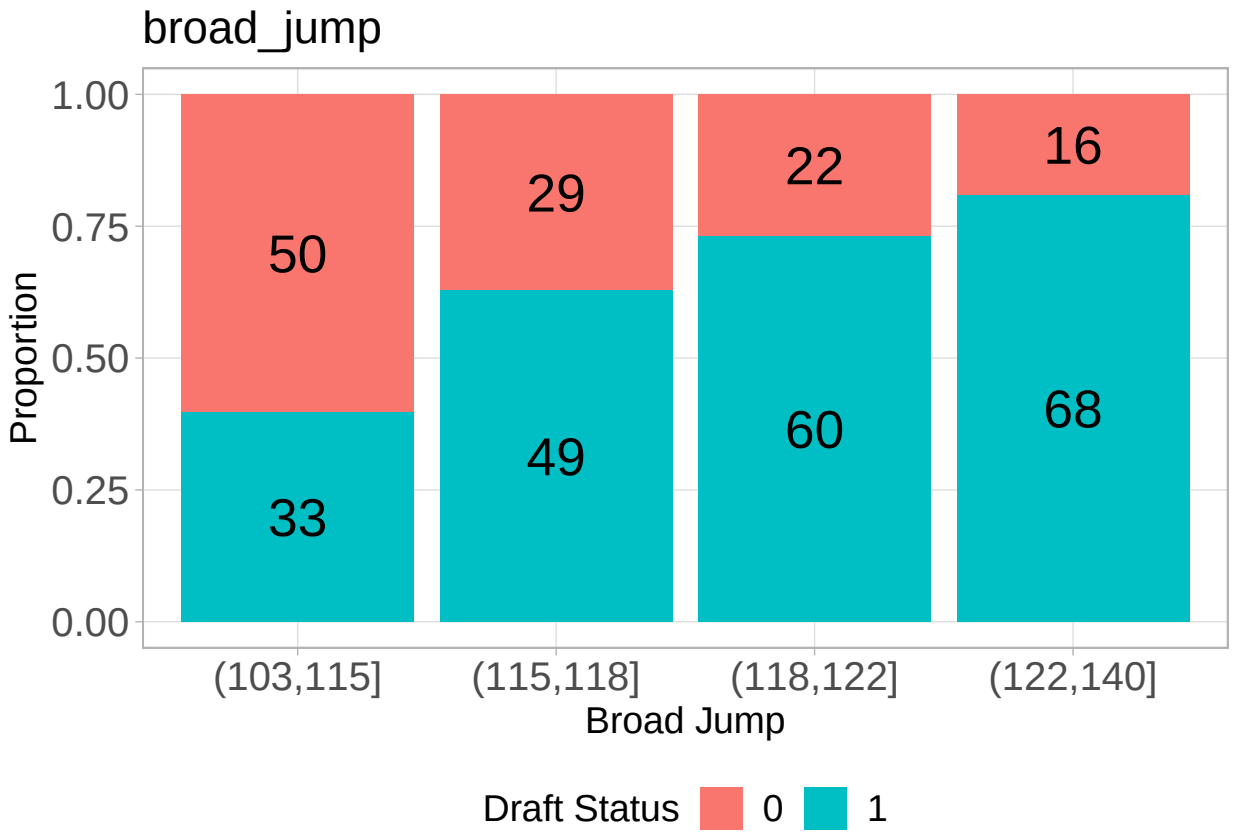
```
sd(combo.df$broad_jump)
```

```
## [1] 5.853789
```

```
summary(glm(draft.fac~broad_jump, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 8.540803e-08
```

```
broad_jump.hist <- ggplot(combo.df, aes(x = cut(broad_jump, breaks=c(103, 115, 118, 122, 140)),
                    fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Broad Jump", title="broad_jump", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
broad_jump.hist
```



cone

```
summary(combo.df$cone)[c(1,3,6)]
```

```
##   Min. Median   Max.
##   6.64   7.12   7.87
```

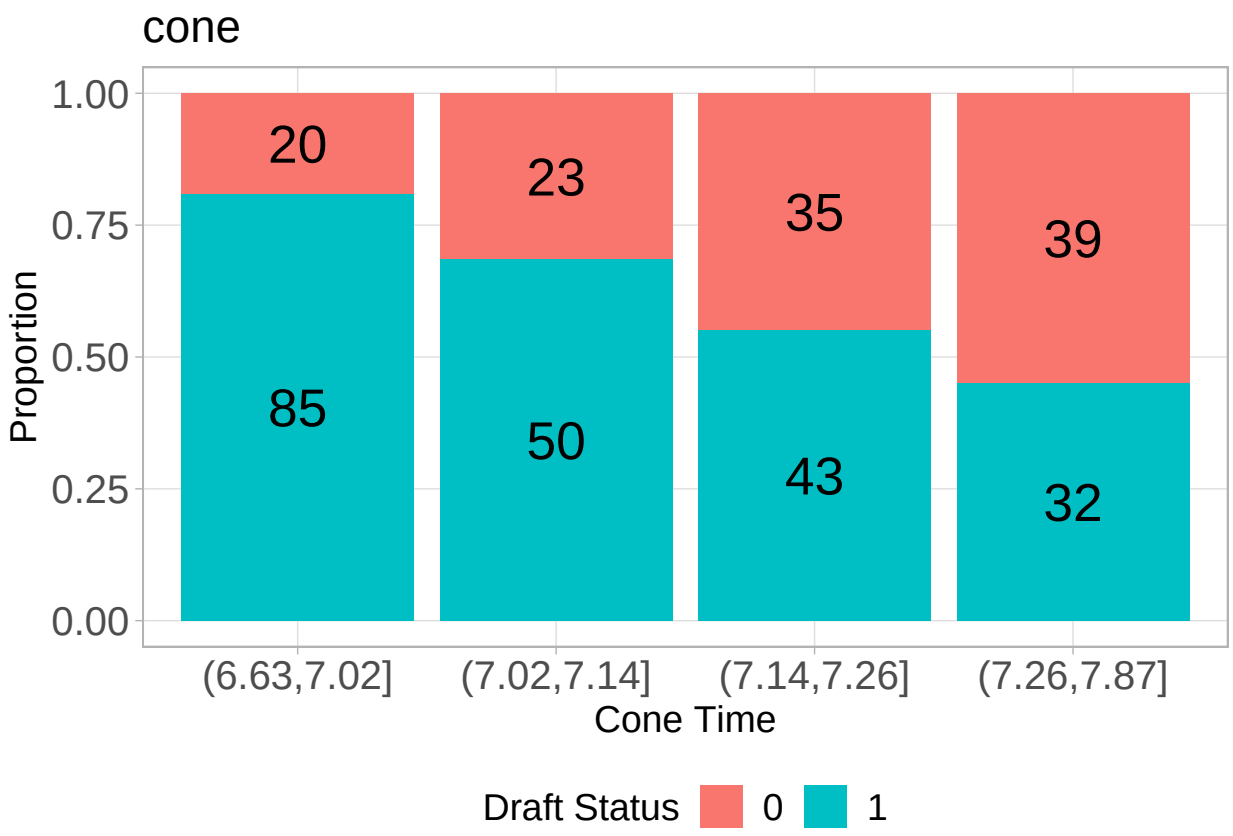
```
sd(combo.df$cone)
```

```
## [1] 0.1954929
```

```
summary(glm(draft.fac~cone, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 1.865381e-06
```

```
cone.hist <- ggplot(combo.df, aes(x = cut(cone, breaks=c(6.63, 7.02, 7.14, 7.26, 7.87)),
                    fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Cone Time", title="cone", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
cone.hist
```



shuttle

```
summary(combo.df$shuttle)[c(1,3,6)]
```

```
##    Min. Median    Max.
##    4.00   4.31   4.96
```

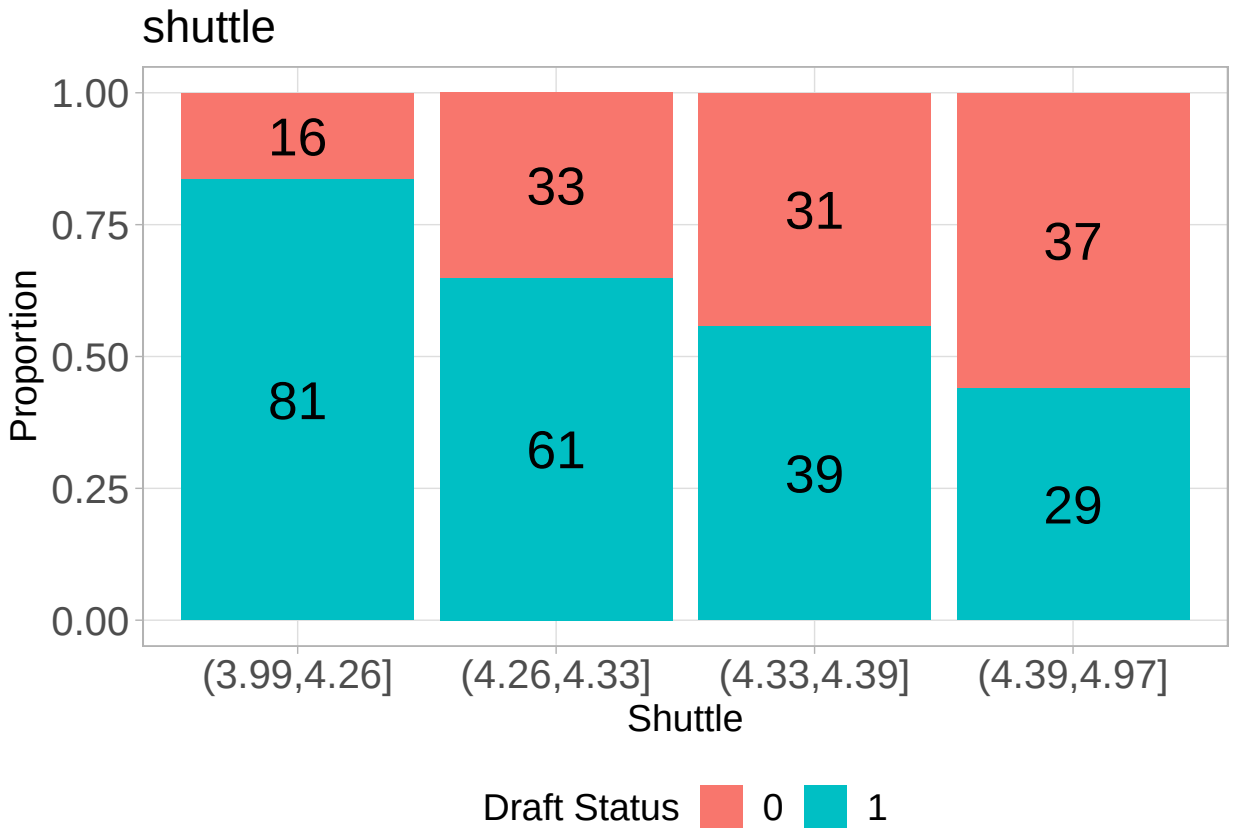
```
sd(combo.df$shuttle)
```

```
## [1] 0.1351995
```

```
summary(glm(draft.fac~shuttle, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 1.96715e-07
```

```
shuttle.hist <- ggplot(combo.df, aes(x = cut(shuttle, breaks=c(3.99, 4.26, 4.33, 4.39, 4.97)),
                       fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Shuttle", title="shuttle", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
shuttle.hist
```



tackles_solo

```
summary(combo.df$tackles_solo)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      6    121    338
```

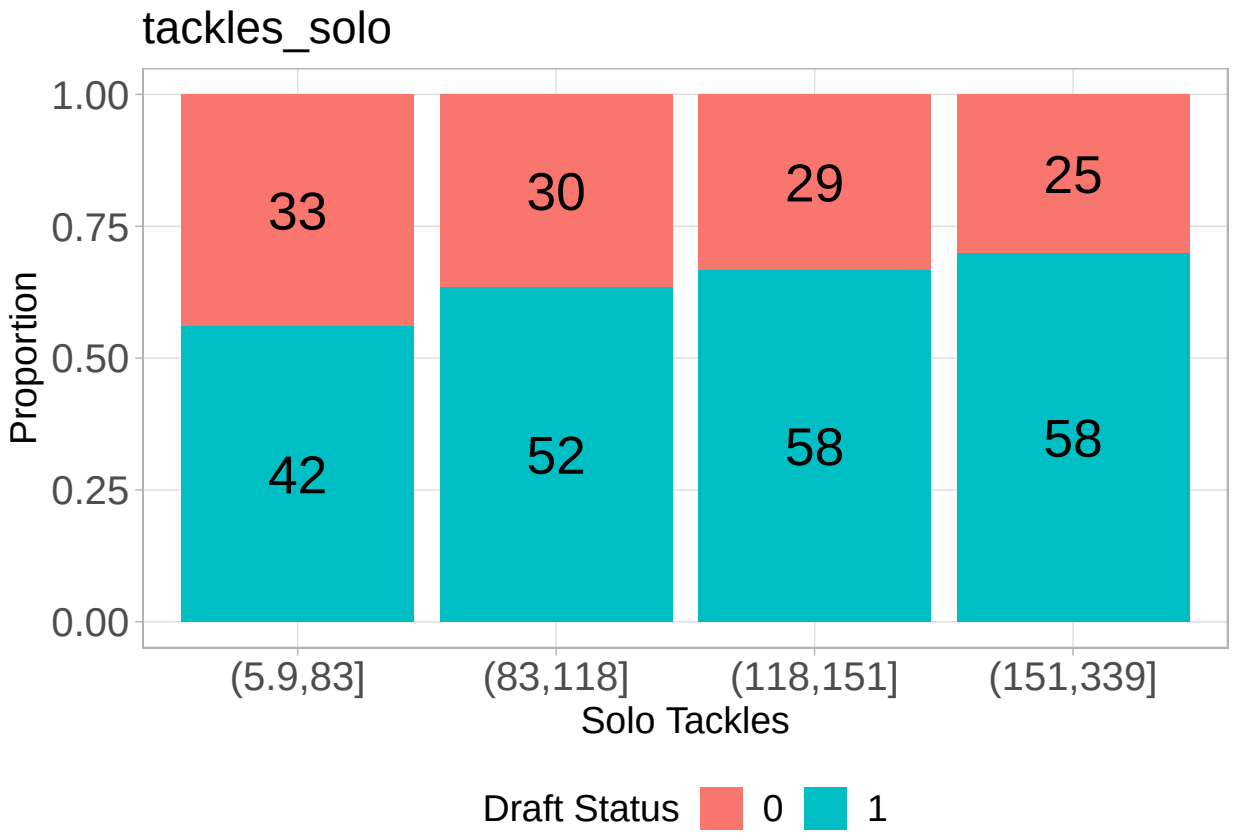
```
sd(combo.df$tackles_solo)
```

```
## [1] 52.16007
```

```
summary(glm(draft.fac~tackles_solo, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.04935869
```

```
tackles_solo.hist <- ggplot(combo.df, aes(x = cut(tackles_solo, breaks=c(5.9, 83, 117.5, 151, 339)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Solo Tackles", title="tackles_solo", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
tackles_solo.hist
```



tackles_ast

```
summary(combo.df$tackles_ast)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      2     92    310
```

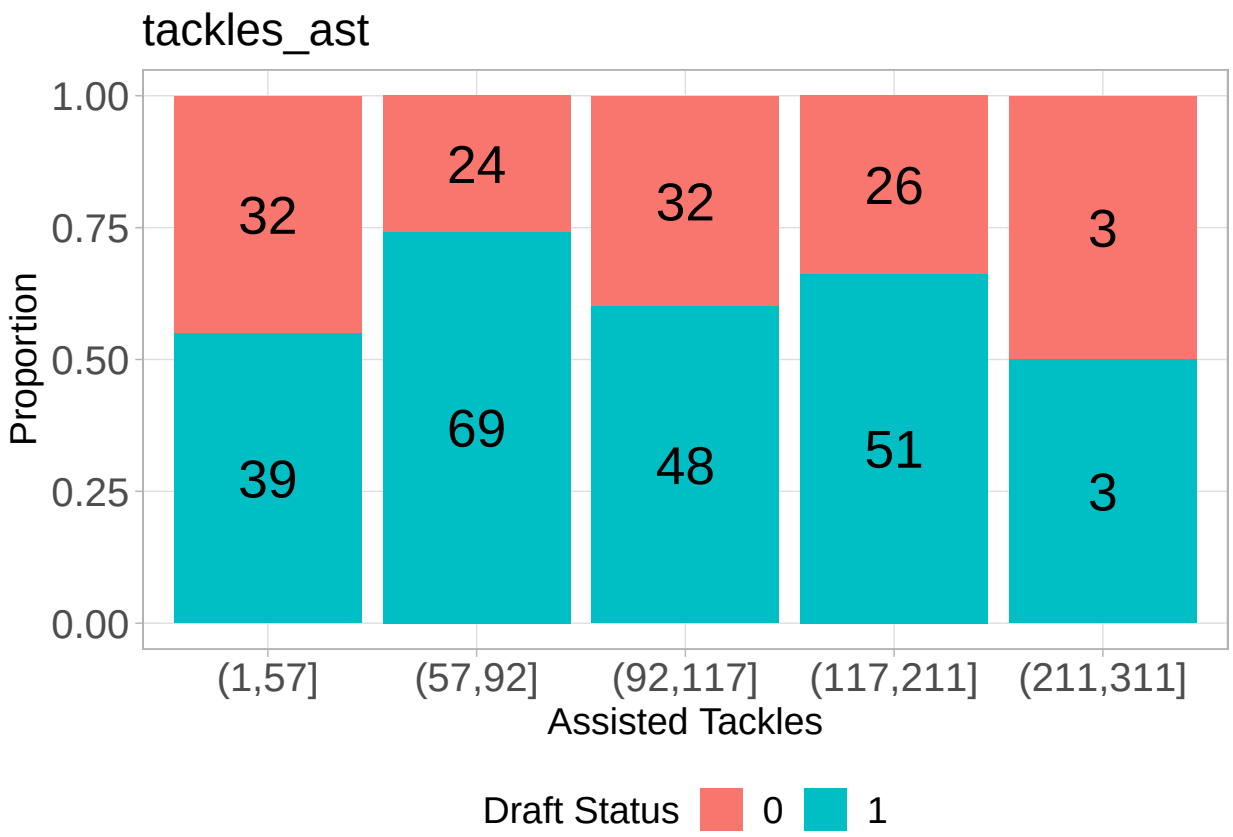
```
sd(combo.df$tackles_ast)
```

```
## [1] 45.77384
```

```
summary(glm(draft.fac~tackles_ast, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.8078726
```

```
tackles_ast.hist <- ggplot(combo.df, aes(x = cut(tackles_ast, breaks=c(1, 57, 92, 117, 211, 311)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Assisted Tackles", title="tackles_ast", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
tackles_ast.hist
```



tackles_tot

```
summary(combo.df$tackles_tot)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      8    218    633
```

```
sd(combo.df$tackles_tot)
```

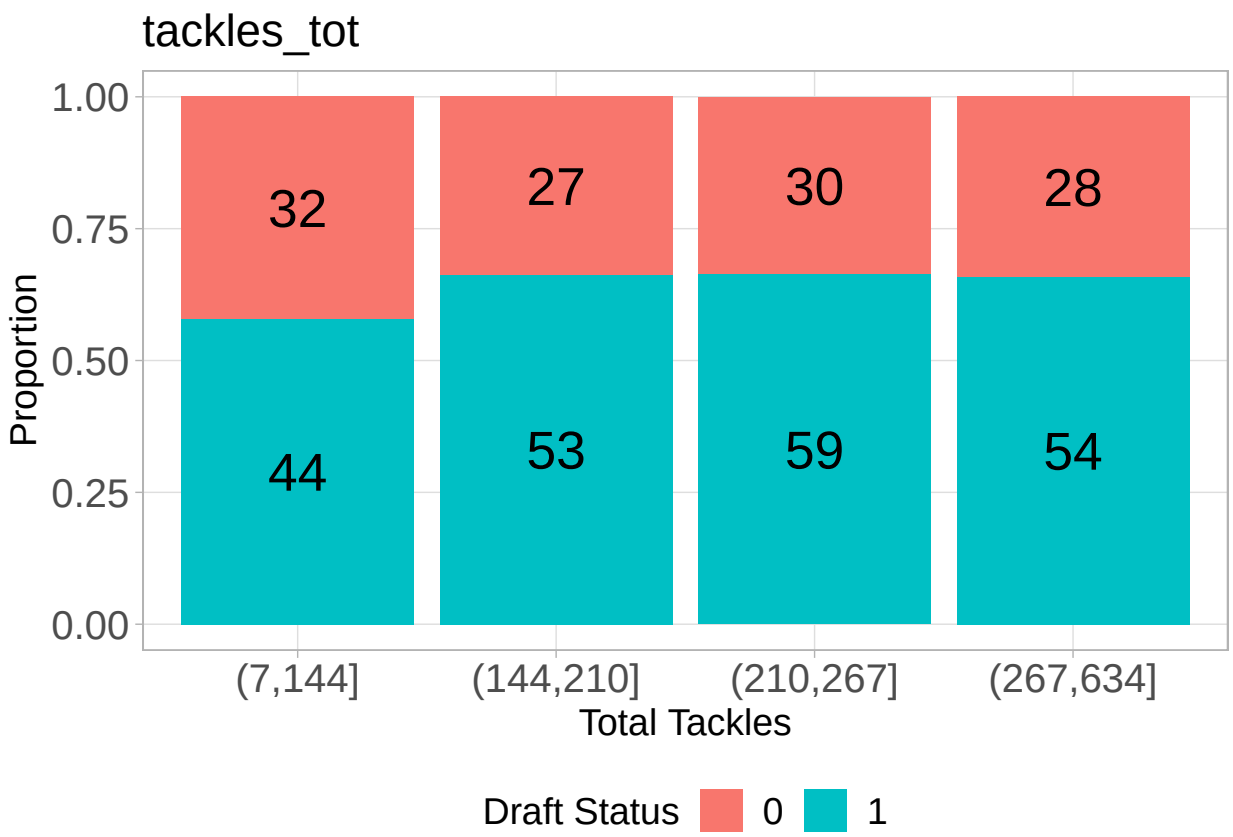
```
## [1] 91.38606
```

```
summary(glm(draft.fac~tackles_tot, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.2289549
```

```
tackles_tot.hist <- ggplot(combo.df, aes(x = cut(tackles_tot, breaks=c(7, 144, 210.5, 267, 633.5)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Total Tackles", title="tackles_tot", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
```

```
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
tackles_tot.hist
```



tackles_loss

```
summary(combo.df$tackles_loss)[c(1,3,6)]
```

```
##   Min. Median   Max.
##    0.0   22.5   74.5
```

```
sd(combo.df$tackles_loss)
```

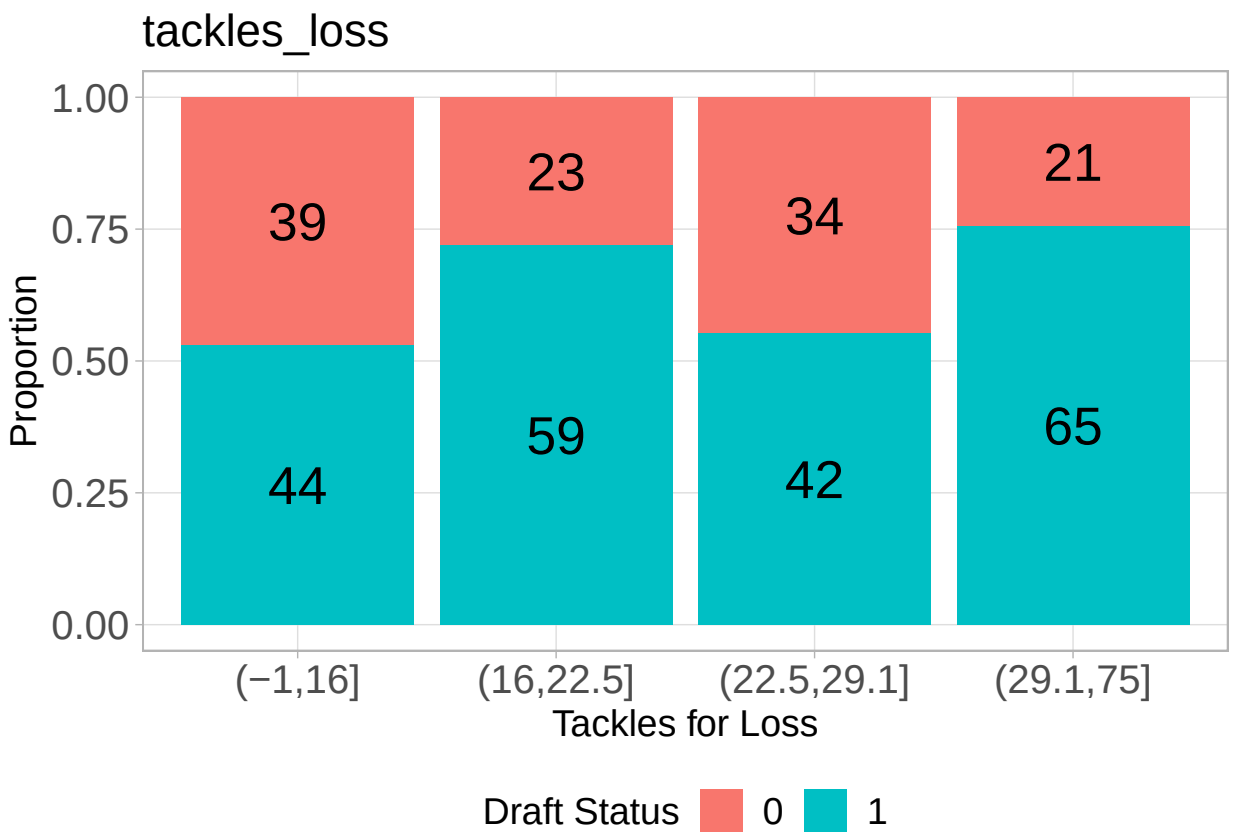
```
## [1] 11.36956
```

```
summary(glm(draft.fac~tackles_loss, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.003122458
```

```
tackles_loss.hist <- ggplot(combo.df, aes(x = cut(tackles_loss, breaks=c(-1, 16, 22.5, 29.125, 75)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
```

```
    theme_light() +
    labs(y="Proportion", x="Tackles for Loss", title="tackles_loss", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
tackles_loss.hist
```



sacks

```
summary(combo.df$sacks)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      0      7     37
```

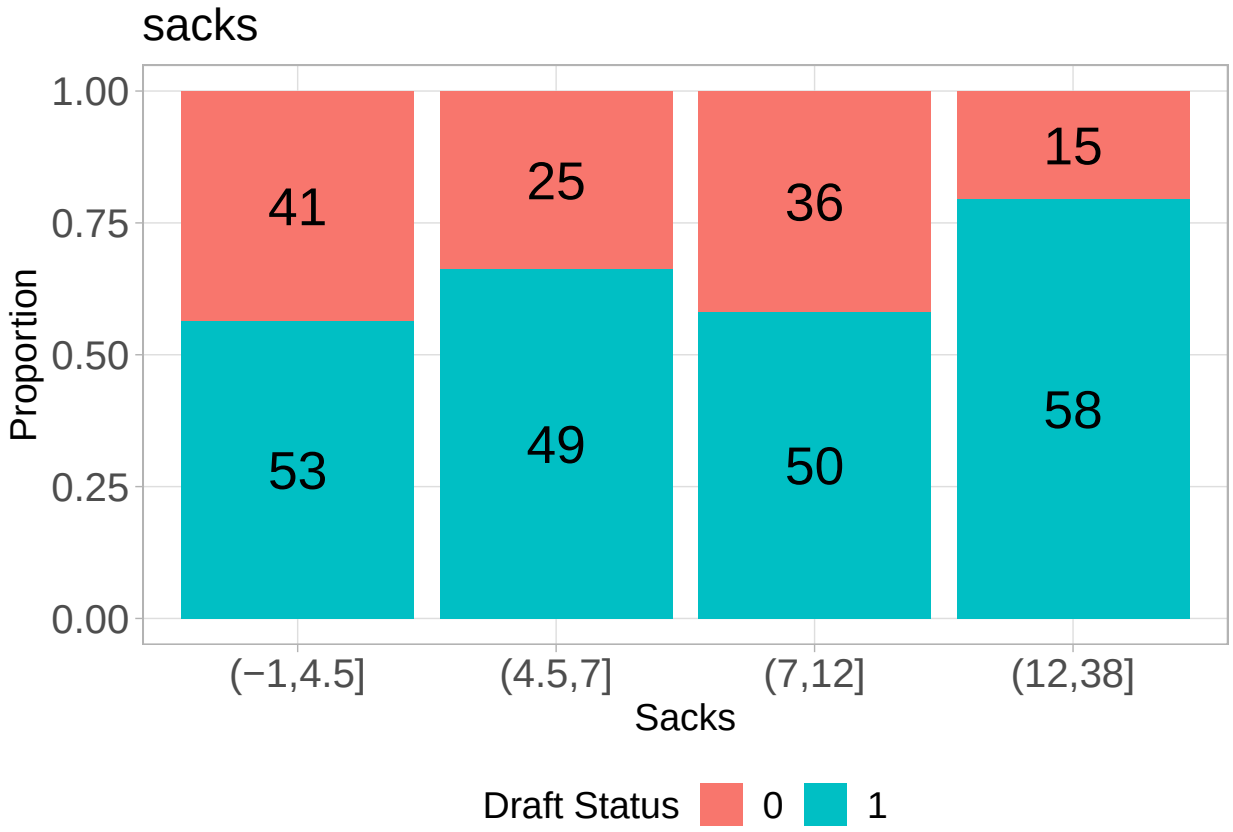
```
sd(combo.df$sacks)
```

```
## [1] 6.612366
```

```
summary(glm(draft.fac~sacks, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.001451477
```

```
sacks.hist <- ggplot(combo.df, aes(x = cut(sacks, breaks=c(-1, 4.5, 7, 12, 38)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Sacks", title="sacks", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
sacks.hist
```



def_int

```
summary(combo.df$def_int)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      0      1     14
```

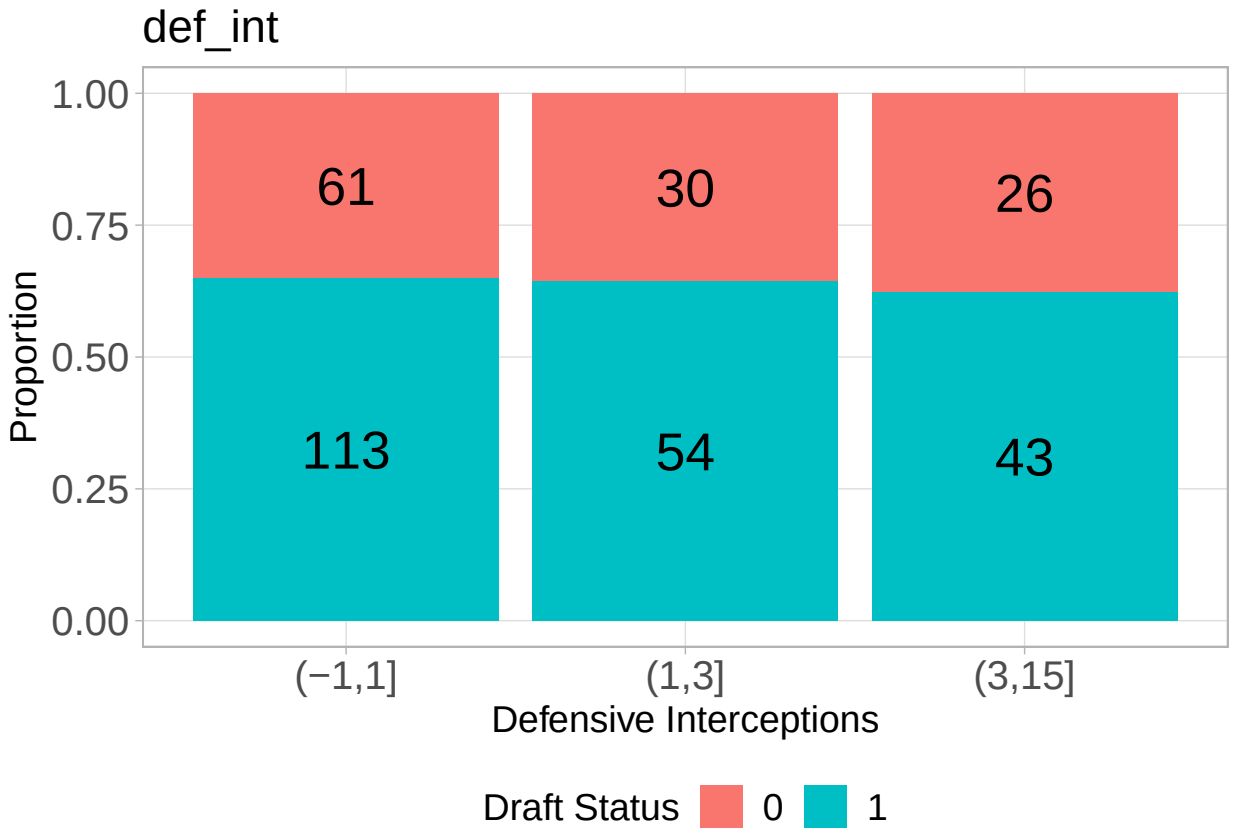
```
sd(combo.df$def_int)
```

```
## [1] 2.117481
```

```
summary(glm(draft.fac~def_int, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.464747
```

```
def_int.hist <- ggplot(combo.df, aes(x = cut(def_int, breaks=c(-1, 1, 3, 15)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Defensive Interceptions", title="def_int", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
def_int.hist
```



pass_defended

```
summary(combo.df$pass_defended)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      0      5     26
```

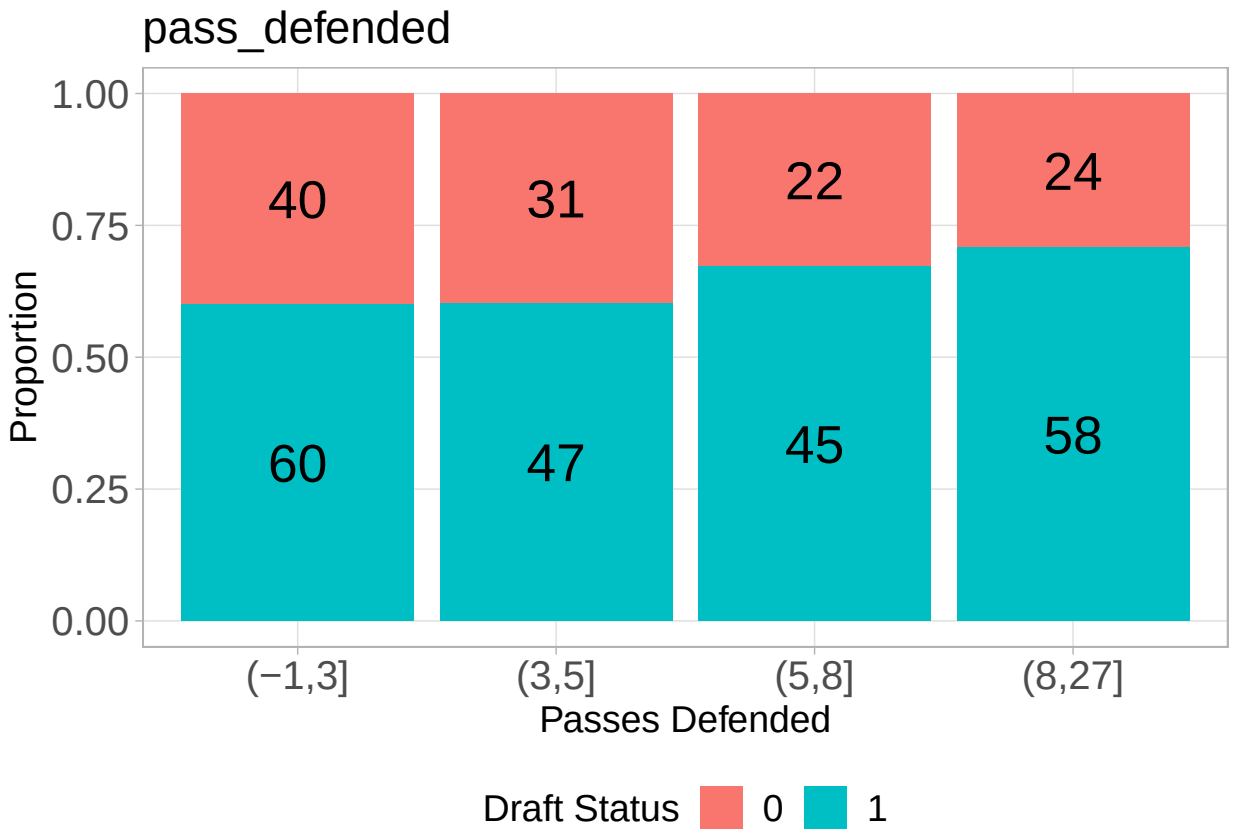
```
sd(combo.df$pass_defended)
```

```
## [1] 4.401568
```

```
summary(glm(draft.fac~pass_defended, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.1255166
```

```
pass_defended.hist <- ggplot(combo.df, aes(x = cut(pass_defended, breaks=c(-1, 3, 5, 8, 27)),
                      fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text", aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Passes Defended", title="pass_defended", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
pass_defended.hist
```



forced_fumble

```r
summary(combo.df$forced_fumble)[c(1,3,6)]
```

```
##   Min. Median   Max.
##      0      2     16
```

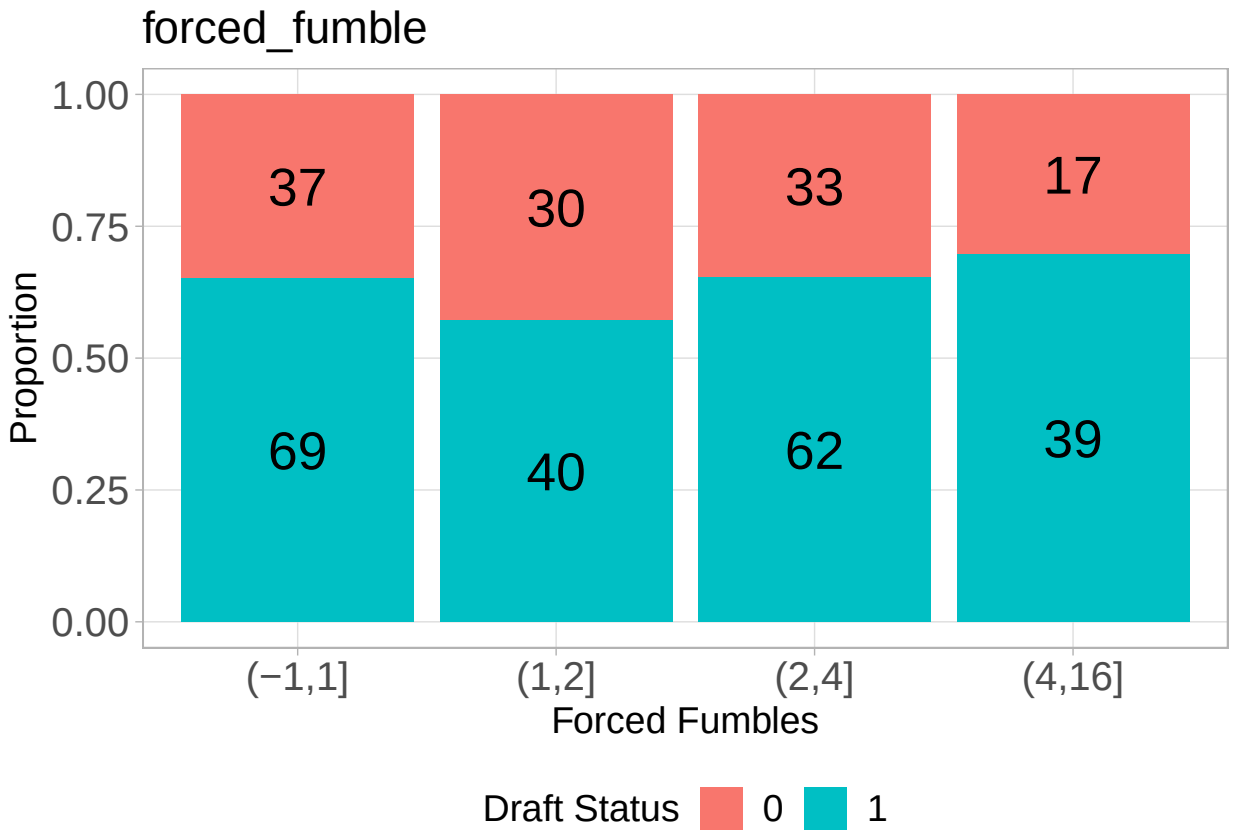
```r
sd(combo.df$forced_fumble)
```

```
## [1] 2.222388
```

```r
summary(glm(draft.fac~forced_fumble, data=combo.df, family=binomial))$coefficients[2,4]
```

```
## [1] 0.3227735
```

```r
forced_fumble.hist <- ggplot(combo.df, aes(x = cut(forced_fumble, breaks=c(-1, 1, 2, 4, 16)),
                     fill = factor(draft.fac)))+
    geom_bar(position = "fill") +
    stat_count(geom = "text",
               aes(label=..count..),
               position = position_fill(vjust = 0.5), size=7) +
    theme_light() +
    labs(y="Proportion", x="Forced Fumbles", title="forced_fumble", fill="Draft Status") +
    theme(panel.grid.minor=element_blank(), legend.position = "bottom",
          axis.text.x=element_text(size=15), plot.title=element_text(size=17),
          plot.subtitle=element_text(size=14), axis.title.x=element_text(size=14),
          legend.text=element_text(size=14), legend.title=element_text(size=14),
          axis.title.y=element_text(size=14), axis.text.y=element_text(size=15) )
forced_fumble.hist
```



Feature Selection

Data Preparation

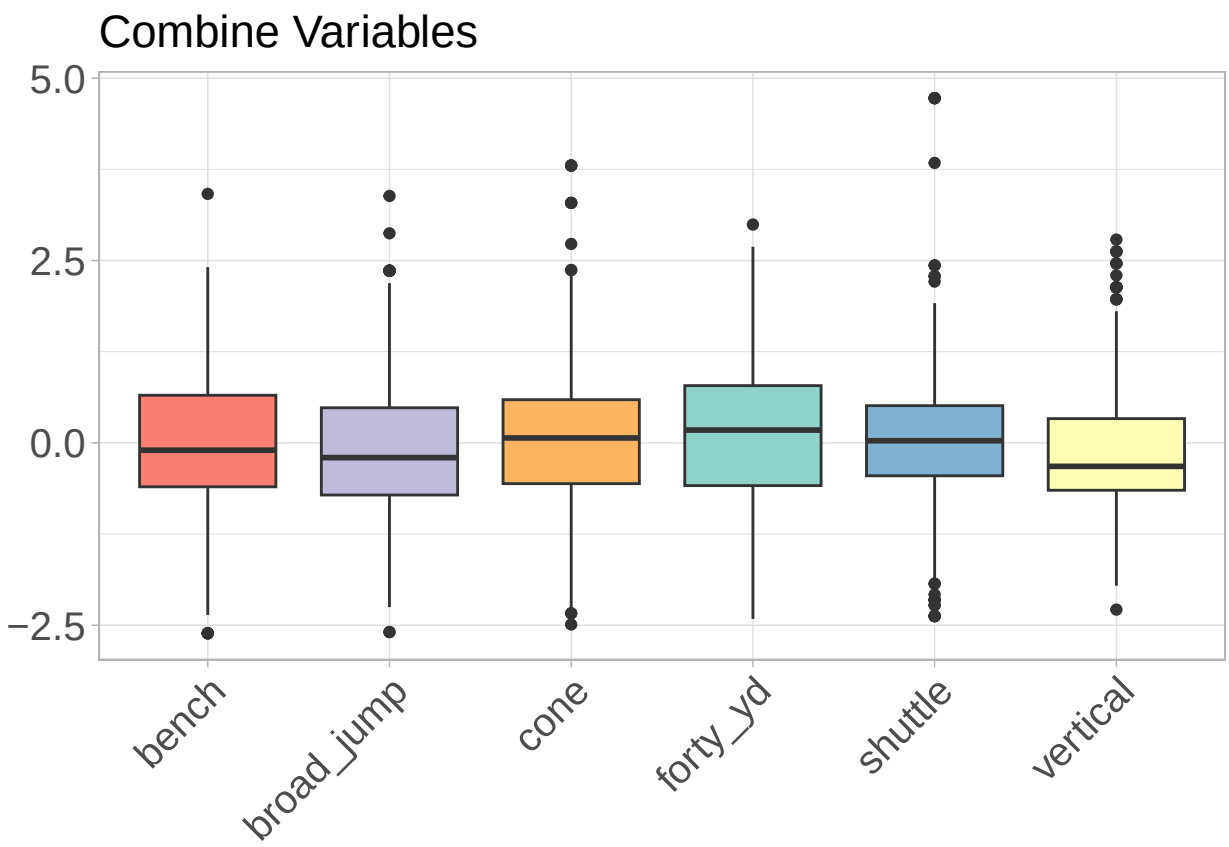
```
set.seed(1234)
# standardization
x <- scale(combo.df[1:16])
y <- combo.df$draft.fac
year <- combo.df$year
combo.df <- as.data.frame( cbind(x, y, year) ) %>%
    rename(draft.fac = y)

# Upsampling for Unrepresented Class (Undrafted Players)
draft.fac.factor <- as.factor( combo.df$draft.fac )
prop.table( table( draft.fac.factor ) ) # very disproportionate classes
```

```
## draft.fac.factor
##         0         1
## 0.3577982 0.6422018
```

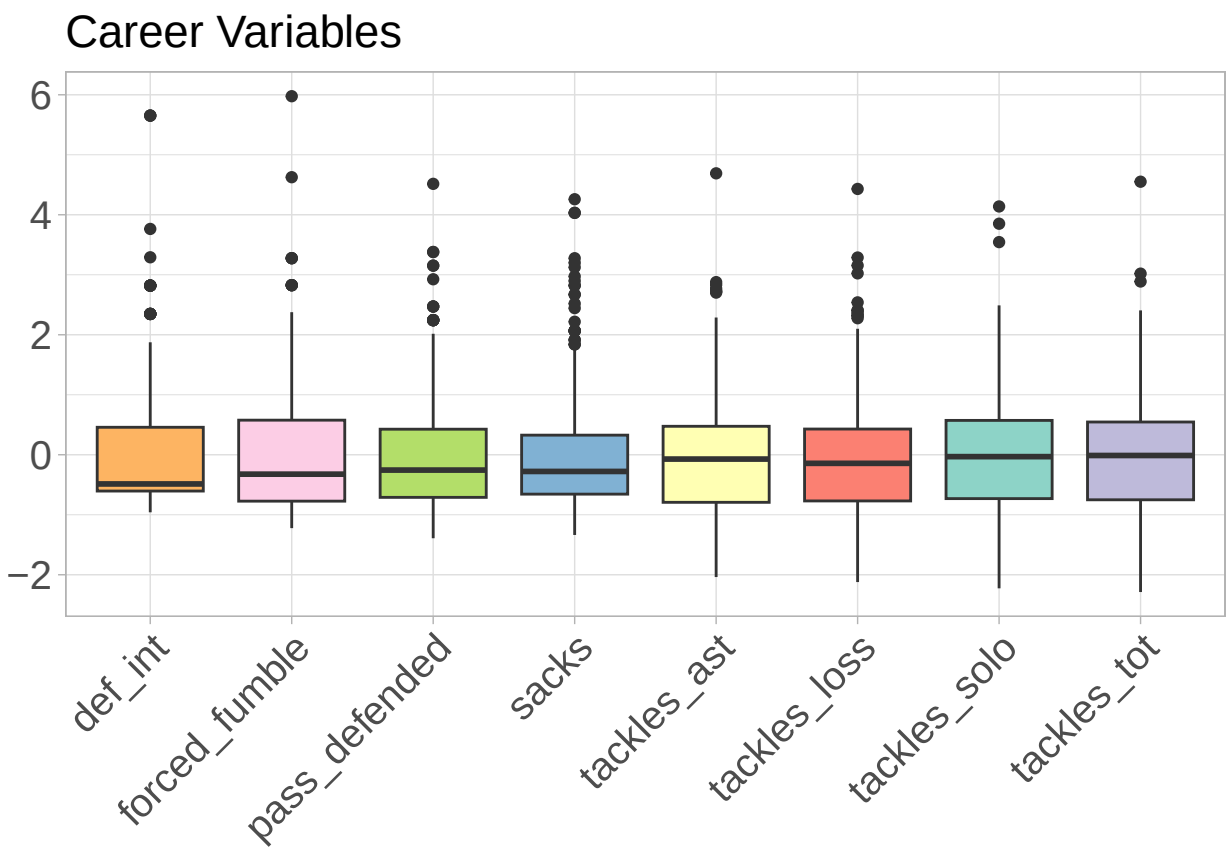
```
predictor.df <- combo.df[ , c(1:16, 18) ]
combo.df <- upSample( x=predictor.df, y=draft.fac.factor ) %>%
    rename( draft.fac=Class )
table( combo.df$draft.fac ) # proportionate classes
```

```
##
##   0   1
## 210 210
```

```
combo.df$draft.fac <- as.numeric( as.character( combo.df$draft.fac ) )

# box plots of combine statistic variables
ggplot( data=combo.df ) +
    geom_boxplot( aes(x="forty_yd", y=forty_yd), fill = brewer.pal(12, "Set3")[1] ) +
    geom_boxplot( aes(x="vertical", y=vertical), fill = brewer.pal(12, "Set3")[2] ) +
    geom_boxplot( aes(x="broad_jump", y=broad_jump), fill = brewer.pal(12, "Set3")[3] ) +
    geom_boxplot( aes(x="bench", y=bench), fill = brewer.pal(12, "Set3")[4] ) +
    geom_boxplot( aes(x="shuttle", y=shuttle), fill = brewer.pal(12, "Set3")[5] ) +
    geom_boxplot( aes(x="cone", y=cone), fill = brewer.pal(12, "Set3")[6] ) +
    theme_light() + labs(title="Combine Variables") +
    theme(axis.title.x = element_blank(), axis.title.y = element_blank(),
          axis.text.x = element_text(angle = 45, hjust = 1, size=15),
          axis.text.y = element_text(size=15), plot.title=element_text( size=17))
```

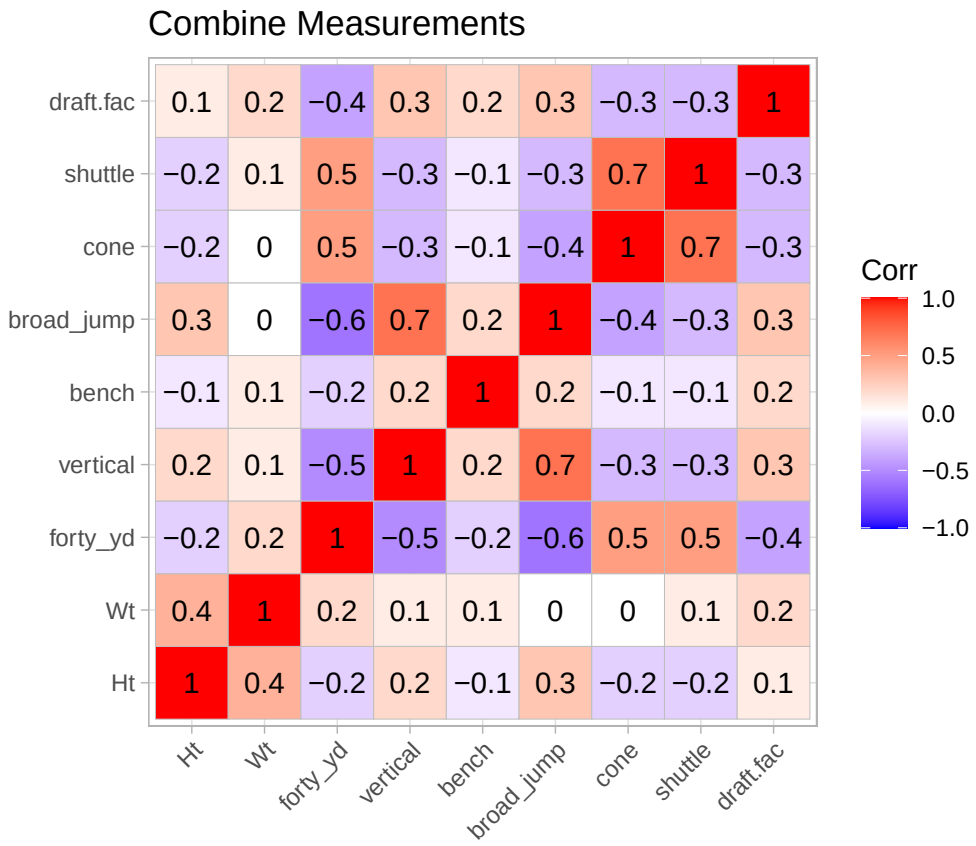



```
# box plots of career statistic variables
ggplot( data=combo.df ) +
    geom_boxplot( aes(x="tackles_solo", y=tackles_solo), fill=brewer.pal(12, "Set3")[1] ) +
    geom_boxplot( aes(x="tackles_ast", y=tackles_ast), fill=brewer.pal(12, "Set3")[2] ) +
    geom_boxplot( aes(x="tackles_tot", y=tackles_tot), fill=brewer.pal(12, "Set3")[3] ) +
    geom_boxplot( aes(x="tackles_loss", y=tackles_loss), fill=brewer.pal(12, "Set3")[4] ) +
    geom_boxplot( aes(x="sacks", y=sacks), fill=brewer.pal(12, "Set3")[5] ) +
    geom_boxplot( aes(x="def_int", y=def_int), fill=brewer.pal(12, "Set3")[6] ) +
    geom_boxplot( aes(x="pass_defended", y=pass_defended), fill=brewer.pal(12, "Set3")[7] ) +
    geom_boxplot( aes(x="forced_fumble", y=forced_fumble), fill=brewer.pal(12, "Set3")[8] ) +
    theme_light() + labs(title="Career Variables") +
    theme(axis.title.x = element_blank(), axis.title.y = element_blank(),
          axis.text.x = element_text(angle = 45, hjust = 1, size=15),
          axis.text.y = element_text(size=15), plot.title=element_text( size=17))
```

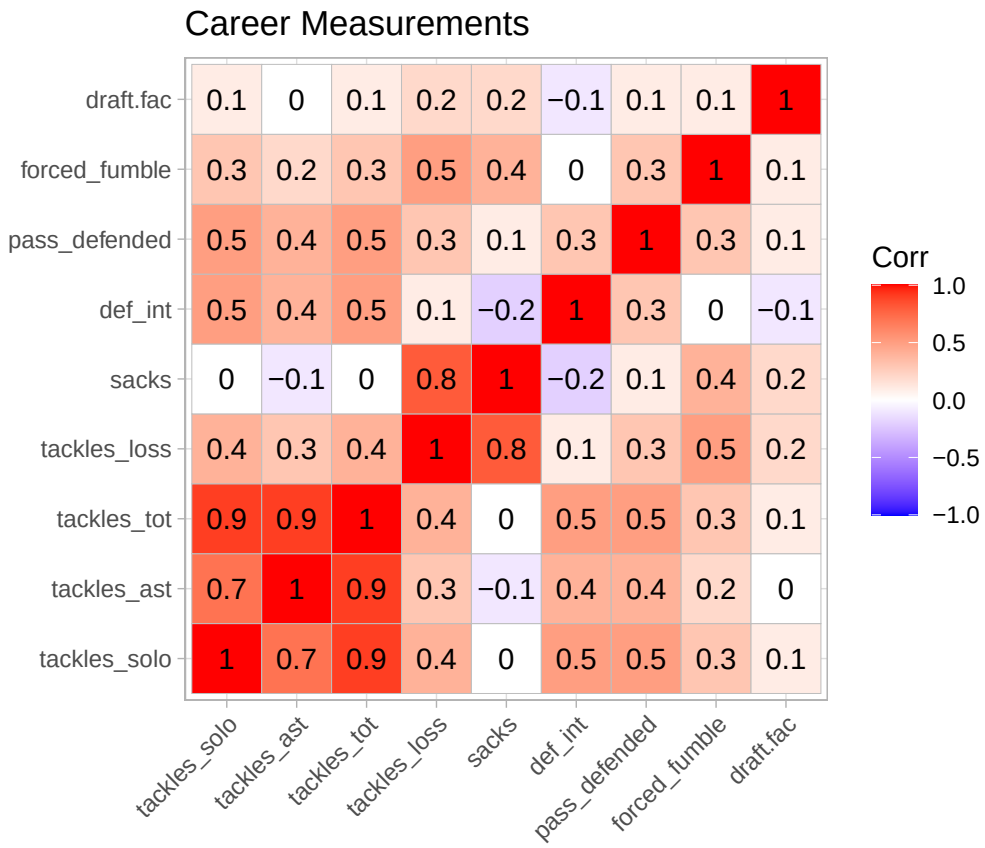


Correlation Matrices

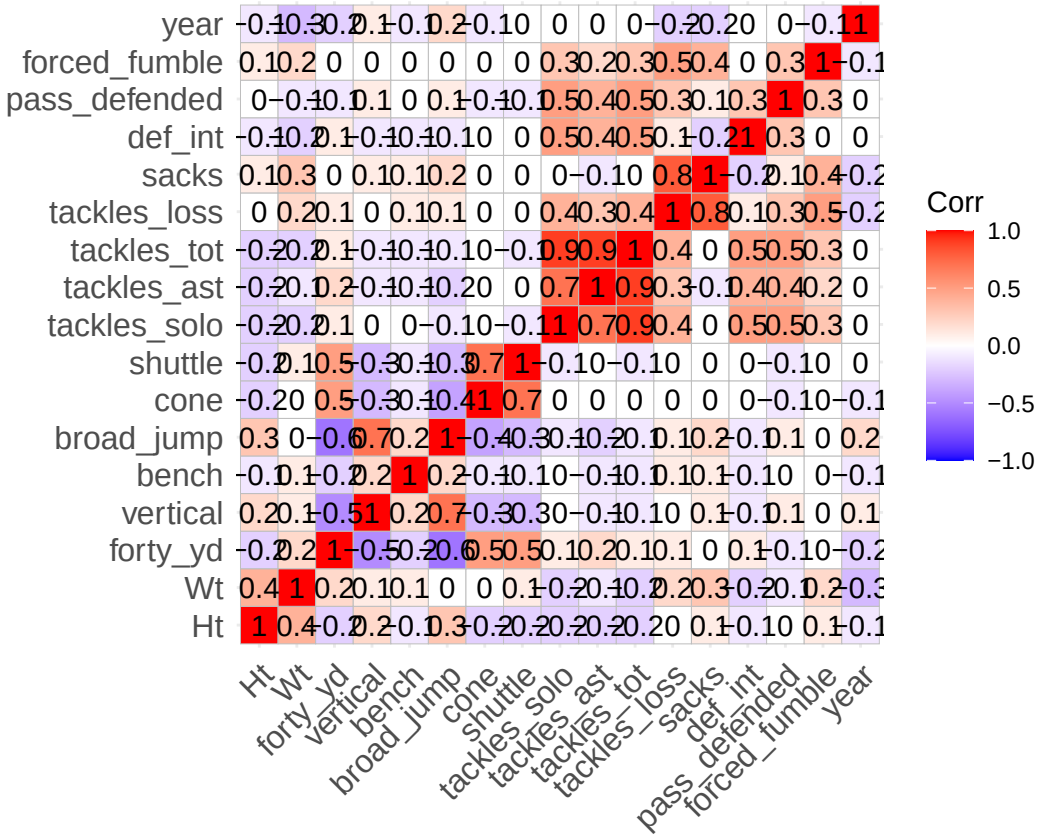
```
# combine correlation matrix
combine.mat <- as.matrix( combo.df[ ,c(1:8, 18) ] )
combine.corr <- round( cor( combine.mat ), 1 )
ggcorrplot( combine.corr, lab=TRUE ) +
    ggtitle( "Combine Measurements" ) +
    theme_light() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1),
          axis.title.x = element_blank(), axis.title.y = element_blank())
```




```
# career stats correlation matrix
career.mat <- as.matrix( combo.df[ ,c(9:16,18) ] )
career.corr <- round( cor( career.mat ), 1 )
ggcorrplot( career.corr, lab=TRUE ) +
    ggtitle( "Career Measurements" ) +
    theme_light() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1),
          axis.title.x = element_blank(), axis.title.y = element_blank())
```




```
full.mat <- as.matrix( combo.df[ ,c( 1:17 ) ] )
full.corr <- round( cor( full.mat ), 1 )
ggcorrplot( full.corr, lab=TRUE )
```



Best Subset Selection

```
set.seed(1234)
draft.glm <- combo.df %>%
    rename( y = draft.fac ) %>% dplyr::select( -year, -tackles_tot ) # get rid of tackles_tot due to mu

# best subset based off AIC
best.sub.AIC <- bestglm( draft.glm,
                         family=binomial,
                         method="exhaustive",
                         IC="AIC" )
```

```
## Morgan-Tatar search since family is non-gaussian.
```

```
# best subset based off BIC
best.sub.BIC <- bestglm( draft.glm,
                         family=binomial,
                         method="exhaustive",
                         IC="BIC" )
```

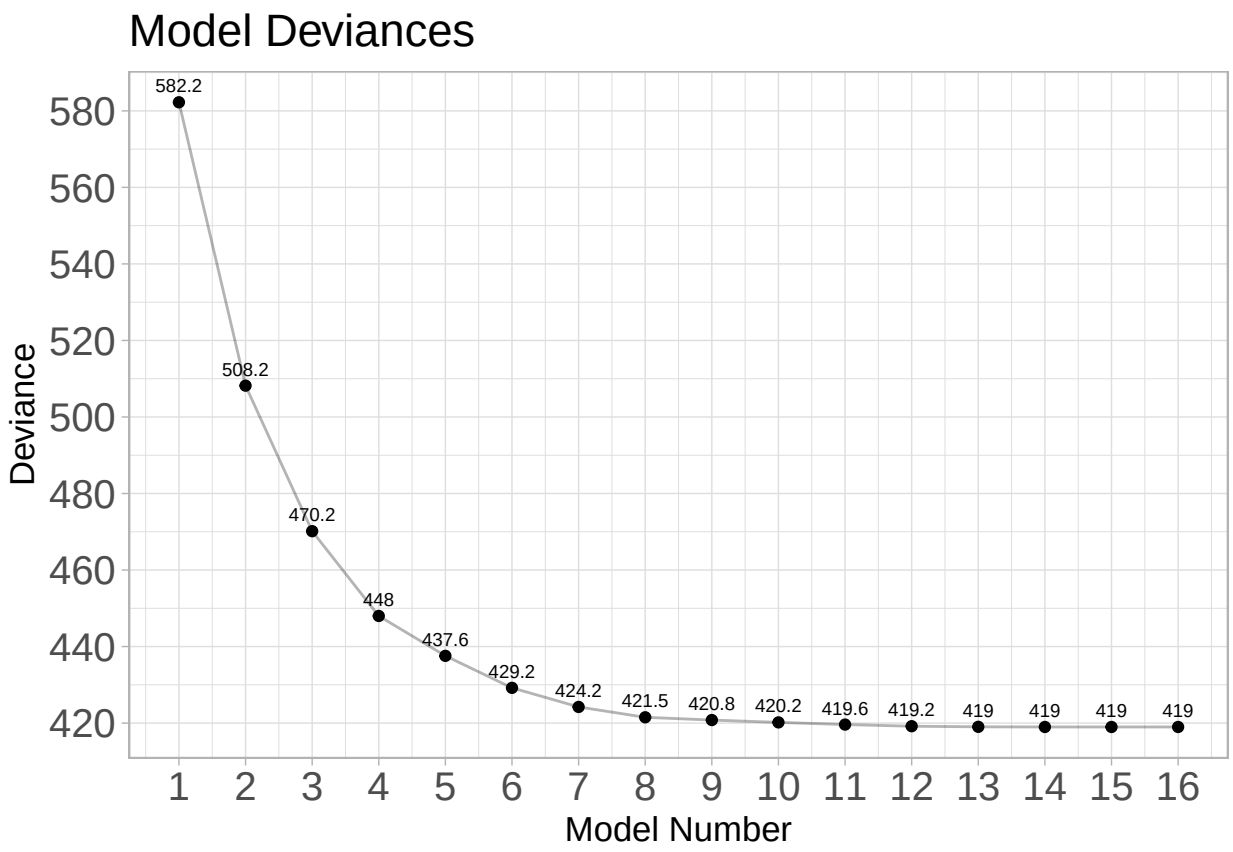
```
## Morgan-Tatar search since family is non-gaussian.
```

```
# creates dataframe for each model and calculates McFadden R^2 and deviance
model.stats <- function( covariates ){
    df <- data.frame( model = rep( NA, times=length( covariates ) ) )

    for( i in 1:length(covariates) ){
        if( i ==1 ){
            tmp.formula <- "draft.fac ~ 1"
            mod <- glm( tmp.formula, data=combo.df, family="binomial" )
        }else{
            tmp.formula <- paste( covariates[ as.matrix( best.sub.AIC$Subsets[i, 1:16] )], collapse=" +
            tmp.formula <- paste( "draft.fac ~", tmp.formula )
            mod <- glm( tmp.formula, data=combo.df, family="binomial" )
        }
        df$mcfad.R[i] <- 1 - logLik(mod) / logLik( glm(draft.fac~1, data=combo.df, family="binomial") )
        df$dev[i] <- mod$deviance
        df$model[i] <- tmp.formula
    }
    return( df )
}

covariates <- colnames( best.sub.AIC$Subsets[, 1:16 ] )
covariates[1] <- "1"
glm.model.stats <- model.stats(covariates)
glm.model.stats$aic <- best.sub.AIC$Subsets$AIC
glm.model.stats$bic <- best.sub.BIC$Subsets$BIC

head( glm.model.stats, n=10 )
```

```
##
## 1                                                                                                draft.fa
## 2                                                                                      draft.fac ~ 1 + fo
## 3                                                                                   draft.fac ~ 1 + Wt + fo
## 4                                                                        draft.fac ~ 1 + Wt + forty_yd + tackle
## 5                                                              draft.fac ~ 1 + Wt + forty_yd + shuttle + tackle
## 6                                                    draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackle
## 7                                         draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo +
## 8                             draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo + sacks + d
## 9                      draft.fac ~ 1 + Wt + forty_yd + bench + cone + shuttle + tackles_solo + sacks + d
## 10 draft.fac ~ 1 + Wt + forty_yd + bench + cone + shuttle + tackles_solo + sacks + def_int + forced_
##      mcfad.R      dev      aic      bic
## 1  0.0000000 582.2436 582.2436 582.2436
## 2  0.1271948 508.1853 510.1853 514.2255
## 3  0.1925143 470.1534 474.1534 482.2339
## 4  0.2305734 447.9938 453.9938 466.1145
## 5  0.2484726 437.5721 445.5721 461.7331
## 6  0.2628348 429.2097 439.2097 459.4110
## 7  0.2714074 424.2184 436.2184 460.4599
## 8  0.2760545 421.5127 435.5127 463.7945
## 9  0.2772904 420.7931 436.7931 469.1151
## 10 0.2783376 420.1833 438.1833 474.5456
```

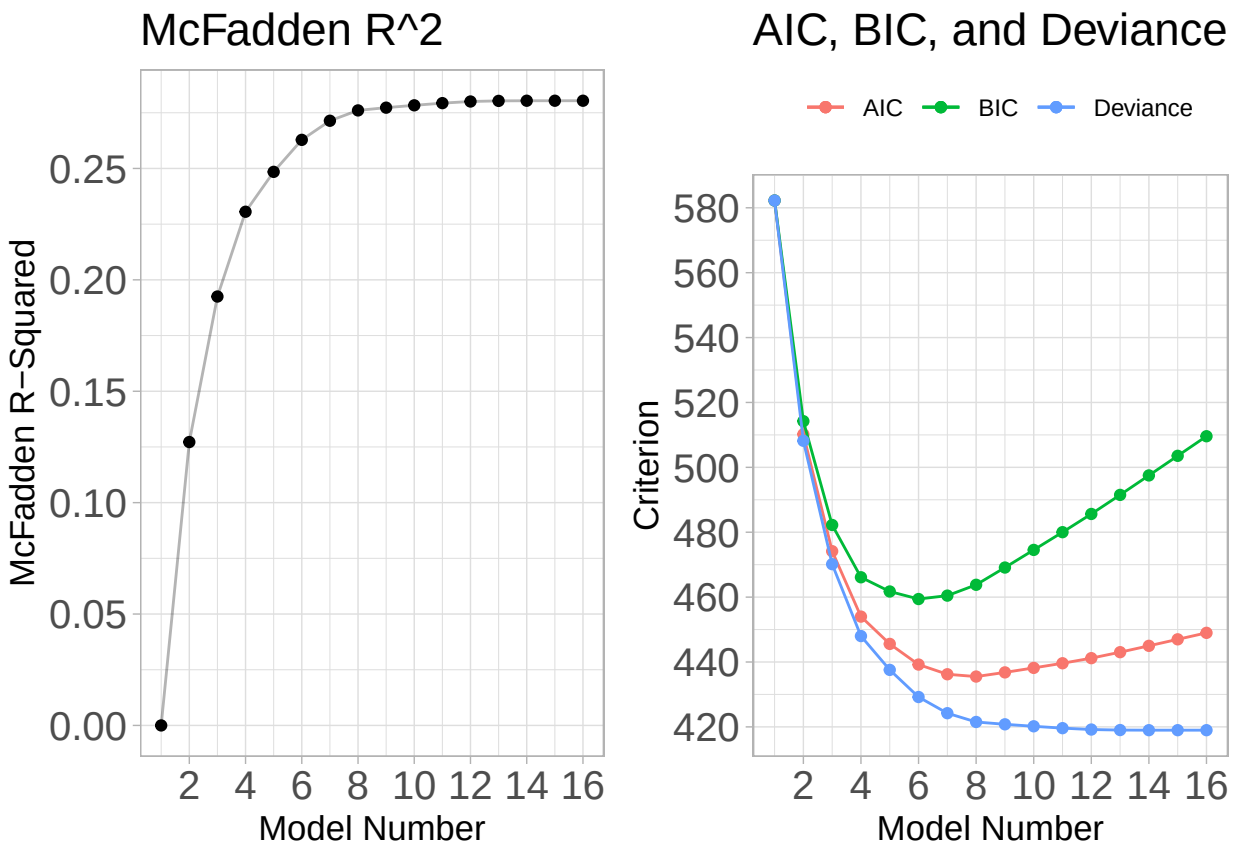
```
# Deviance Plot for Each Model
ggplot( data = glm.model.stats, aes(x=1:16, y=dev, label=round(dev, 1) ) ) +
```

```
    geom_point() +
    geom_path(alpha=0.3) +
    theme_light() + scale_x_continuous( n.breaks=16 ) +
    scale_y_continuous( n.breaks=10 ) +
    labs( y="Deviance", x="Model Number", title="Model Deviances") +
    geom_text( size=2.5, vjust= -0.75 ) +
    theme(axis.text.x=element_text(size=15), axis.text.y=element_text(size=15),
          plot.title=element_text(size=17), axis.title.x=element_text(size=13),
          axis.title.y=element_text(size=13))
```




```
# McFadden R-Squared Values for Each Model
mcfad <- ggplot( data = glm.model.stats, aes(x=1:16, y=mcfad.R, label=round(mcfad.R, 2) ) ) +
    geom_point() +
    geom_path(alpha=0.3) +
    theme_light() + scale_x_continuous( n.breaks=9 ) +
    scale_y_continuous( n.breaks = 10 ) +
    labs( y="McFadden R-Squared", x="Model Number", title="McFadden R^2" ) +
    #geom_text( size=2.5, vjust=-0.75 ) +
    theme(axis.text.x=element_text(size=15), axis.text.y=element_text(size=15),
          plot.title=element_text(size=17), axis.title.x=element_text(size=13),
          axis.title.y=element_text(size=13))

# AIC/BIC Values for Each Model
aicbicdev <- ggplot( data = glm.model.stats ) +
    geom_point( aes(x=1:16, y=aic, color="AIC") ) +
    geom_path( aes(x=1:16, y=aic, color="AIC"), alpha=1 ) +
```

```
    #geom_text( aes(x=1:16, y=aic, label=round(aic, 1) ), size=2.5, vjust=-0.75 ) +
    geom_point( aes(x=1:16, y=bic, color="BIC") ) +
    geom_path( aes(x=1:16, y=bic, color="BIC"), alpha=1 ) +
    #geom_text( aes(x=1:16, y=bic, label=round(bic, 1) ), size=2.5, vjust=-0.75 ) +
    geom_point( aes(x=1:16, y=dev, color="Deviance") ) +
    geom_path( aes(x=1:16, y=dev, color="Deviance"), alpha=1 ) +
    #geom_text( aes(x=1:16, y=dev, label=round(dev, 1) ), size=2.5, vjust=-0.75) +
    theme_light() + scale_y_continuous(n.breaks = 10) +
    scale_x_continuous( n.breaks=9 ) +
    labs( y="Criterion", x="Model Number", title="AIC, BIC, and Deviance", color="" ) +
    theme(axis.text.x=element_text(size=15), axis.text.y=element_text(size=15),
          plot.title=element_text(size=17), axis.title.x=element_text(size=13),
          axis.title.y=element_text(size=13), legend.position = "top")

grid.arrange(mcfad, aicbicdev, nrow=1)
```



Cross Validation

```
set.seed(1234)
cv.errors <- function( df ){
    df <- data.frame( model=df$model )
    for( i in 1:nrow( df ) ){
        glm.fit <- glm( df$model[i], data=combo.df, family="binomial" )
        df$loocv[i] <- cv.glm( combo.df, glm.fit )$delta[1]
```

```
        df$kfcv.5[i] <- cv.glm( combo.df, glm.fit, K=5)$delta[1]
        df$kfcv.10[i] <- cv.glm( combo.df, glm.fit, K=10)$delta[1]
    }
    return( df )
}
error.df <- cv.errors( glm.model.stats )
head(error.df, n=10)
```

```
##
## 1                                                                                          draft.fa
## 2                                                                                draft.fac ~ 1 + for
## 3                                                                           draft.fac ~ 1 + Wt + for
## 4                                                             draft.fac ~ 1 + Wt + forty_yd + tackles
## 5                                                   draft.fac ~ 1 + Wt + forty_yd + shuttle + tackles
## 6                                           draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles
## 7                                   draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo +
## 8                         draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo + sacks + de
## 9                  draft.fac ~ 1 + Wt + forty_yd + bench + cone + shuttle + tackles_solo + sacks + de
## 10 draft.fac ~ 1 + Wt + forty_yd + bench + cone + shuttle + tackles_solo + sacks + def_int + forced_
##        loocv    kfcv.5   kfcv.10
## 1  0.2511947 0.2507334 0.2504787
## 2  0.2085669 0.2092224 0.2096780
## 3  0.1926267 0.1928243 0.1932875
## 4  0.1839780 0.1835728 0.1831627
## 5  0.1792542 0.1792297 0.1786469
## 6  0.1761006 0.1783558 0.1749459
## 7  0.1754850 0.1761809 0.1766245
## 8  0.1750478 0.1777594 0.1777526
## 9  0.1762127 0.1771022 0.1759162
## 10 0.1766679 0.1768522 0.1762456
```

```
error.df[ which.min(error.df$loocv), ]
```

```
##                                                                                 model
## 8 draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo + sacks + def_int
##       loocv    kfcv.5   kfcv.10
## 8 0.1750478 0.1777594 0.1777526
```

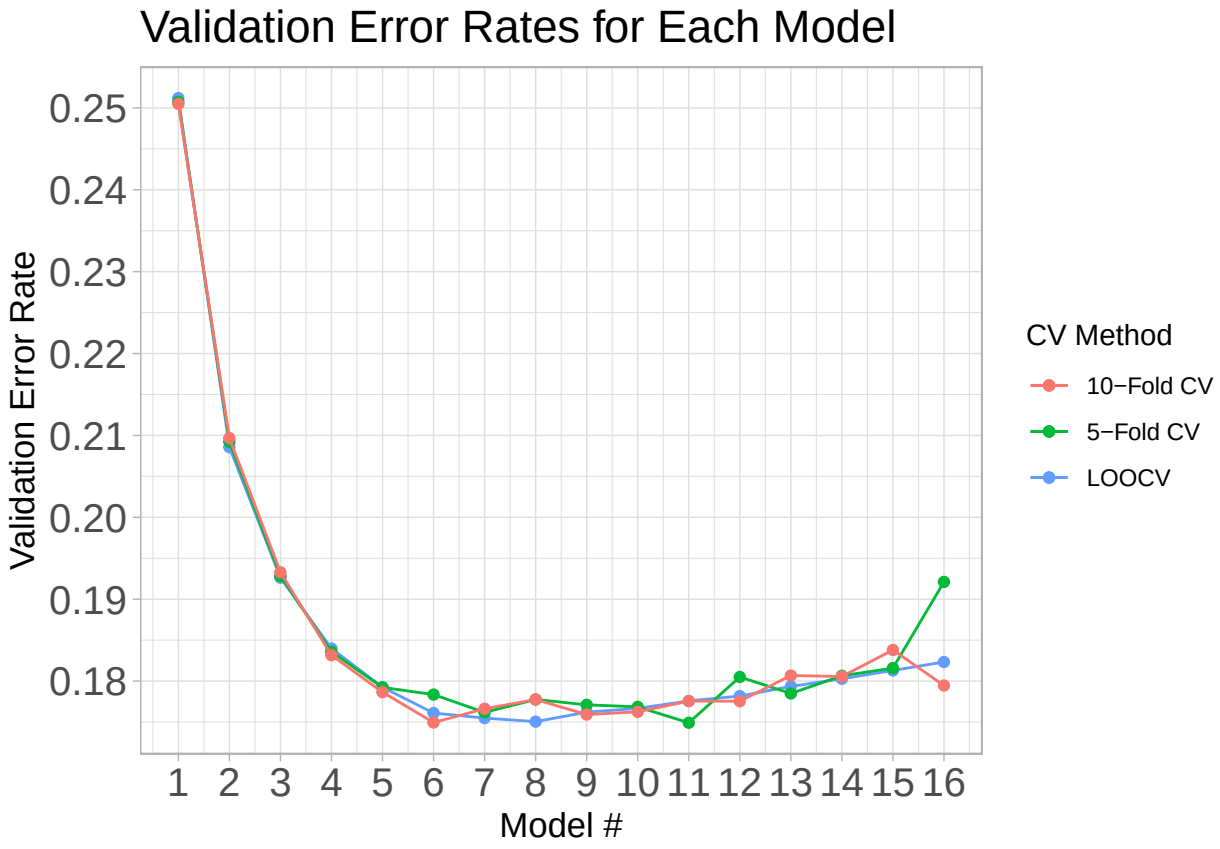
```
error.df[ which.min(error.df$kfcv.5), ]
```

```
##
## 11 draft.fac ~ 1 + Ht + Wt + forty_yd + bench + broad_jump + cone + shuttle + tackles_solo + sacks +
##        loocv   kfcv.5   kfcv.10
## 11 0.1775805 0.174911 0.1775518
```

```
error.df[ which.min(error.df$kfcv.10), ]
```

```
##                                                               model     loocv
## 6 draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo 0.1761006
##      kfcv.5   kfcv.10
## 6 0.1783558 0.1749459
```

```
ggplot( data=error.df ) +
    geom_path( aes(x=1:16, y=loocv, colour="LOOCV") ) +
    geom_point( aes(x=1:16, y=loocv, colour="LOOCV") ) +
    geom_path( aes(x=1:16, y=kfcv.5, colour="5-Fold CV") ) +
    geom_point( aes(x=1:16, y=kfcv.5, colour="5-Fold CV") ) +
    geom_path( aes(x=1:16, y=kfcv.10, colour="10-Fold CV") ) +
    geom_point( aes(x=1:16, y=kfcv.10, colour="10-Fold CV") ) +
    theme_light() + scale_y_continuous(n.breaks=10) + scale_x_continuous(n.breaks=16) +
    labs(y="Validation Error Rate", x="Model #", colour="CV Method", title="Validation Error Rates for E
    theme(axis.text.x=element_text(size=15), axis.text.y=element_text(size=15),
          plot.title=element_text(size=17), axis.title.x=element_text(size=13),
          axis.title.y=element_text(size=13), legend.position = "right")
```



Test Error Rates

```
set.seed(1234)

prediction.errors <- function( formula ){
    training.set <- combo.df %>% filter( year != 2022 )
    testing.set <- combo.df %>% filter( year == 2022 )
    mod <- glm( formula, data=training.set, family=binomial)
    pred <- predict( mod, testing.set, type="response" )
    class <- rep( 0, times=nrow(testing.set) )
    class[ pred > 0.6 ] <- 1
```

```
    tab <- table( class, testing.set$draft.fac)
    return( tab )
}


prediction.errors( "draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo + sacks + def_int" )
```

```
##
## class  0  1
##     0 16  2
##     1  6 14
```

```
( error <- (6+1) / 25 )
```

```
## [1] 0.28
```

ROC Curve

```
set.seed(1234)
roc.curve <- function( formula ){

    training.set <- combo.df %>% filter( year != 2022 )
    testing.set <- combo.df %>% filter( year == 2022 )
    mod <- glm( formula, data=training.set, family=binomial )
    pred <- predict( mod, testing.set, type="response" )
    class <- rep( 0, times=nrow(testing.set) )
    class[ pred > 0.6 ] <- 1

    roc_score <- roc(
        as.factor( class ),
        testing.set$draft.fac )
return( roc_score )
}
roc_data <- roc.curve( "draft.fac ~ 1 + Wt + forty_yd + bench + shuttle + tackles_solo + sacks + def_int
```

```
## Setting levels: control = 0, case = 1
```

```
## Setting direction: controls < cases
```

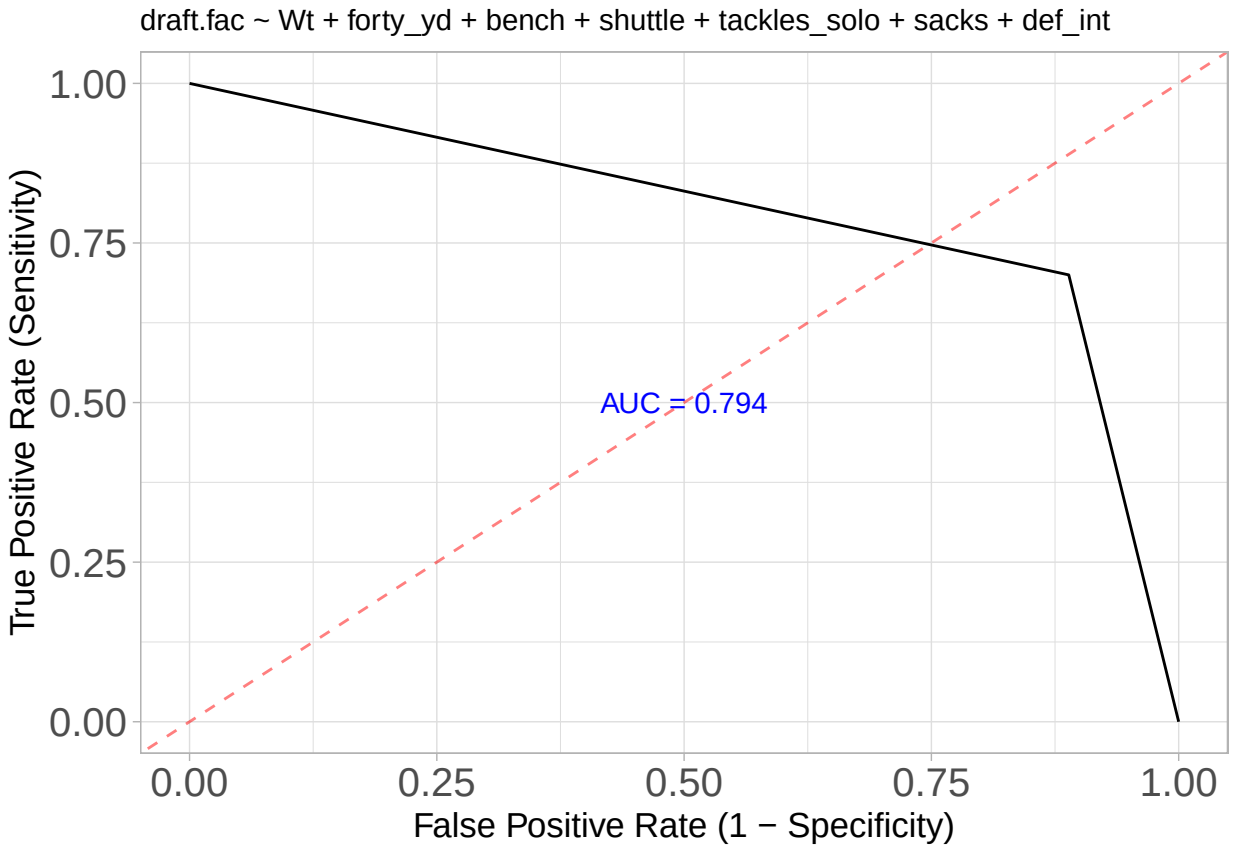
```
roc_df <- data.frame(
  FPR = roc_data$specificities,
  TPR = roc_data$sensitivities
)
roc_df$FPR
```

```
## [1] 0.0000000 0.8888889 1.0000000
```

```
roc_df$TPR
```

```
## [1] 1.0 0.7 0.0
```

```
roc_plot <- ggplot(roc_df, aes(x = FPR, y = TPR)) +
  geom_line() +
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", colour="red", alpha=0.5) +
  labs( x = "False Positive Rate (1 - Specificity)",
        y = "True Positive Rate (Sensitivity)",
        subtitle = "draft.fac ~ Wt + forty_yd + bench + shuttle + tackles_solo + sacks + def_int") +
    theme_light() +
    annotate( "text", x=0.5, y=0.5,
              label=paste( "AUC =", round(roc_data$auc, 3)), color="blue", size=4) +
    theme(axis.text.x=element_text(size=15), axis.text.y=element_text(size=15),
          plot.title=element_text(size=17), axis.title.x=element_text(size=13),
          axis.title.y=element_text(size=13), legend.position = "right")
roc_plot
```



Interpretations

```
setwd("~/Documents/STSCI 4100")
combo.df <- read_csv("combo.df.csv")[,c(6:13, 17:24, 28, 27)]
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Warning: Duplicated column names deduplicated: 'X1' => 'X1_1' [2]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double(),
##   player = col_character(),
##   pos = col_character(),
##   school = col_character(),
##   draft.team = col_character(),
##   pos.fac = col_character(),
##   .html = col_character(),
##   conference = col_character(),
##   forty_yd.fac = col_character(),
##   vertical.fac = col_character(),
##   bench.fac = col_character(),
##   broad_jump.fac = col_character(),
##   cone.fac = col_character(),
##   shuttle.fac = col_character(),
##   tackles_solo.fac = col_character(),
##   tackles_ast.fac = col_character(),
##   tackles_tot.fac = col_character(),
##   tackles_loss.fac = col_character(),
##   sacks.fac = col_character(),
##   def_int.fac = col_character(),
##   pass_defended.fac = col_character()
##   # ... with 1 more columns
## )
## i Use `spec()` for the full column specifications.
```

```
final.mod <- glm(draft.fac ~ Wt+forty_yd+bench+shuttle+tackles_solo+sacks+def_int,
                 data=combo.df, family=binomial)
summary(final.mod)
```

```
##
## Call:
## glm(formula = draft.fac ~ Wt + forty_yd + bench + shuttle + tackles_solo +
##     sacks + def_int, family = binomial, data = combo.df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3264  -0.7704   0.4346   0.7685   2.1352
##
## Coefficients:
```

```
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)  31.282704   7.265601   4.306 1.67e-05 ***
## Wt            0.081648   0.018368   4.445 8.79e-06 ***
## forty_yd     -7.462897   1.336905  -5.582 2.37e-08 ***
## bench         0.068648   0.038232   1.796 0.072568 .
## shuttle      -4.137334   1.315658  -3.145 0.001663 **
## tackles_solo  0.011114   0.003307   3.360 0.000779 ***
## sacks         0.040025   0.024728   1.619 0.105530
## def_int      -0.097254   0.075727  -1.284 0.199045
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 426.50  on 326  degrees of freedom
## Residual deviance: 317.33  on 319  degrees of freedom
## AIC: 333.33
##
## Number of Fisher Scoring iterations: 5
```

$draft.fac_i = 0.2457 + 0.7974(Wt)_i - 1.1096(forty\_yd)_i + 0.3476(bench)_i - 0.533(shuttle)_i + 0.62(tackles\_solo)_i + 0.2853(sacks)_i - 0.2191(def\_int)_i + e_i$

Coefficients and Interpretations

```
( beta1 <- final.mod$coefficients[2] )
```

```
##         Wt
## 0.08164838
```

```
( beta2 <- final.mod$coefficients[3] )
```

```
##  forty_yd
## -7.462897
```

```
( beta3 <- final.mod$coefficients[4] )
```

```
##      bench
## 0.06864789
```

```
( beta4 <- final.mod$coefficients[5] )
```

```
##   shuttle
## -4.137334
```

```
( beta5 <- final.mod$coefficients[6] )
```

```
## tackles_solo
##   0.01111374
```

```r
( beta6 <- final.mod$coefficients[7] )
```

```
##      sacks
## 0.04002482
```

```r
( beta7 <- final.mod$coefficients[8] )
```

```
##     def_int
## -0.09725445
```

```r
( OR.1 <- exp(beta1) )
```

```
##       Wt
## 1.085074
```

```r
( OR.2 <- exp(0.01*beta2) )
```

```
##  forty_yd
## 0.9280878
```

```r
( OR.3 <- exp(beta3) )
```

```
##    bench
## 1.071059
```

```r
( OR.4 <- exp(0.01*beta4) )
```

```
##   shuttle
## 0.9594709
```

```r
( OR.5 <- exp(beta5) )
```

```
## tackles_solo
##     1.011176
```

```r
( OR.6 <- exp(beta6) )
```

```
##    sacks
## 1.040837
```

```r
( OR.7 <- exp(beta7) )
```

```
##   def_int
## 0.9073251
```

```
confint(final.mod)
```

```
## Waiting for profiling to be done...
```

```
##                       2.5 %      97.5 %
## (Intercept)   17.427344281 45.98845442
## Wt             0.046761766  0.11898110
## forty_yd     -10.196452908 -4.93758544
## bench         -0.005262559  0.14513708
## shuttle       -6.771888813 -1.59817899
## tackles_solo   0.004821429  0.01783128
## sacks         -0.007200806  0.09024164
## def_int       -0.248630093  0.04988321
```

```
( OR.CI.1 <- c(exp(0.046761766), exp(0.11898110)) )
```

```
## [1] 1.047872 1.126349
```

```
( OR.CI.2 <- c(exp(0.01*-10.196452908), exp(0.01*-4.93758544)) )
```

```
## [1] 0.9030616 0.9518233
```

```
( OR.CI.3 <- c(exp(-0.005262559), exp(0.14513708)) )
```

```
## [1] 0.9947513 1.1561981
```

```
( OR.CI.4 <- c(exp(0.01*-6.771888813), exp(0.01*-1.59817899)) )
```

```
## [1] 0.9345231 0.9841452
```

```
( OR.CI.5 <- c(exp(0.004821429), exp(0.01783128)) )
```

```
## [1] 1.004833 1.017991
```

```
( OR.CI.6 <- c(exp(-0.007200806), exp(0.09024164)) )
```

```
## [1] 0.9928251 1.0944387
```

```
( OR.CI.7 <- c(exp(-0.248630093), exp(0.04988321)) )
```

```
## [1] 0.7798684 1.0511483
```