Analytical Day LAN VU

Step 0:

```
library(tidyverse)
library(dplyr)
library(knitr)

# ---------------------------
# Summary Statistics
# ---------------------------
numeric_summary <- plays %>%
  summarise(
    Yards_Min  = min(yardsGained, na.rm = TRUE),
    Yards_Max  = max(yardsGained, na.rm = TRUE),
    Yards_Mean = mean(yardsGained, na.rm = TRUE),
    Yards_SD   = sd(yardsGained, na.rm = TRUE)
  )

kable(numeric_summary,
      caption = "Summary Statistics for Key Numeric Variables",
      digits = 2,
      booktabs = TRUE)
```

Table 1: Summary Statistics for Key Numeric Variables

Yards_Min	Yards_Max	Yards_Mean	Yards_SD
-68	98	5.46	8.83

```
# ---------------------------
# Dataset Size (N)
# ---------------------------
```

```r
n_plays <- nrow(plays)
cat("Total number of plays (N):", n_plays, "\n")
```

```
Total number of plays (N): 16124
```

```r
# ---------------------------
# Data Dictionary
# ---------------------------
data_dictionary <- tibble(
  Variable = c(
    "yardsGained",
    "offenseFormation",
    "coverage_group",
    "quarter"
  ),
  Type = c(
    "Numeric",
    "Categorical",
    "Categorical",
    "Ordinal"
  ),
  Description = c(
    "Total yards gained on the play",
    "Offensive formation used on the play",
    "Defensive pass coverage category (collapsed)",
    "Quarter of the game (1-5, where 5 = overtime)"
  ),
  Range_or_Levels = c(
    paste(range(plays$yardsGained, na.rm = TRUE), collapse = " to "),
    paste(unique(plays$offenseFormation), collapse = ", "),
    paste(unique(plays$coverage_group), collapse = ", "),
    paste(range(plays$quarter, na.rm = TRUE), collapse = " to ")
  )
)

kable(data_dictionary,
      caption = "Data Dictionary for Key Variables",
      booktabs = TRUE)
```

Table 2: Data Dictionary for Key Variables

Variable	Type	Description	Range_or_Levels
yardsGained	Numeric	Total yards gained on the play	-68 to 98
offenseFormation	Categorical	Offensive formation used on the play	EMPTY, SHOTGUN, PISTOL, SINGLEBACK, NA, JUMBO, I_FORM, WILDCAT
coverage_group	Categorical	Defensive pass coverage category (collapsed)	
quarter	Ordinal	Quarter of the game (1–5, where 5 = overtime)	1 to 5

```
# ---------------------------
# Coverage Recoding Table
# ---------------------------
coverage_recode_table <- tibble(
  Original_Label = c(
    "Cover-3", "Cover-3 Cloud Left", "Cover-3 Cloud Right",
    "Cover-1", "Cover-1 Double",
    "Cover 6-Left", "Cover-6 Right",
    "Bracket", "Prevent", "Goal Line"
  ),
  Recoded_Group = c(
    rep("Cover_3", 3),
    rep("Cover_1", 2),
    rep("Cover_6", 2),
    rep("Other_Rare", 3)
  )
)

kable(coverage_recode_table,
      caption = "Defensive Coverage Recoding Scheme",
      booktabs = TRUE)
```

Table 3: Defensive Coverage Recoding Scheme

Original_Label	Recoded_Group
Cover-3	Cover_3
Cover-3 Cloud Left	Cover_3
Cover-3 Cloud Right	Cover_3

Original_Label	Recoded_Group
Cover-1	Cover_1
Cover-1 Double	Cover_1
Cover 6-Left	Cover_6
Cover-6 Right	Cover_6
Bracket	Other_Rare
Prevent	Other_Rare
Goal Line	Other_Rare

Step 1: Clean data

```
colSums(is.na(plays))
```

```
                          gameId                           playId
                               0                                0
                 playDescription                          quarter
                               0                                0
                            down                        yardsToGo
                               0                                0
                  possessionTeam                     defensiveTeam
                               0                                0
                     yardlineSide                   yardlineNumber
                             224                                0
                       gameClock                  preSnapHomeScore
                               0                                0
              preSnapVisitorScore             playNullifiedByPenalty
                               0                                0
           absoluteYardlineNumber     preSnapHomeTeamWinProbability
                               0                                0
preSnapVisitorTeamWinProbability                   expectedPoints
                               0                                0
                 offenseFormation                receiverAlignment
                             188                              188
                  playClockAtSnap                       passResult
                               1                                0
                       passLength                          targetX
                            7398                             7748
                          targetY                       playAction
                            7748                                0
                     dropbackType                dropbackDistance
                            5803                             5966
```

```
                passLocationType                          timeToThrow
                            6812                                 7419
             timeInTackleBox                              timeToSack
                            7207                                15516
             passTippedAtLine                        unblockedPressure
                            6788                                 6369
                      qbSpike                                 qbKneel
                            6788                                    0
                      qbSneak                          rushLocationType
                            9336                                 9336
                 penaltyYards                    prePenaltyYardsGained
                           15740                                    0
                yardsGained                homeTeamWinProbabilityAdded
                            0                                       0
  visitorTeamWinProbilityAdded                      expectedPointsAdded
                            0                                       0
                  isDropback                     pff_runConceptPrimary
                            0                                    7053
      pff_runConceptSecondary                        pff_runPassOption
                        13303                                       0
             pff_passCoverage                             pff_manZone
                          192                                     192
```

```r
handle_missing_values <- function(df, threshold = 0.7) {
  # Step 1: Calculate percentage of missing values per column
  missing_percentage <- colMeans(is.na(df))

  # Step 2: Drop columns exceeding threshold
  cols_to_drop <- names(missing_percentage[missing_percentage > threshold])
  df <- df[ , !(names(df) %in% cols_to_drop)]
  message("Dropped columns with more than ", threshold * 100, "% missing values: ",
          paste(cols_to_drop, collapse = ", "))

  # Step 3: Separate numeric and categorical columns
  numeric_cols <- sapply(df, is.numeric)
  categorical_cols <- !numeric_cols

  # Step 4: Impute missing values
  # Numeric: replace NA with mean
  for (col in names(df)[numeric_cols]) {
    df[[col]][is.na(df[[col]])] <- mean(df[[col]], na.rm = TRUE)
  }
```

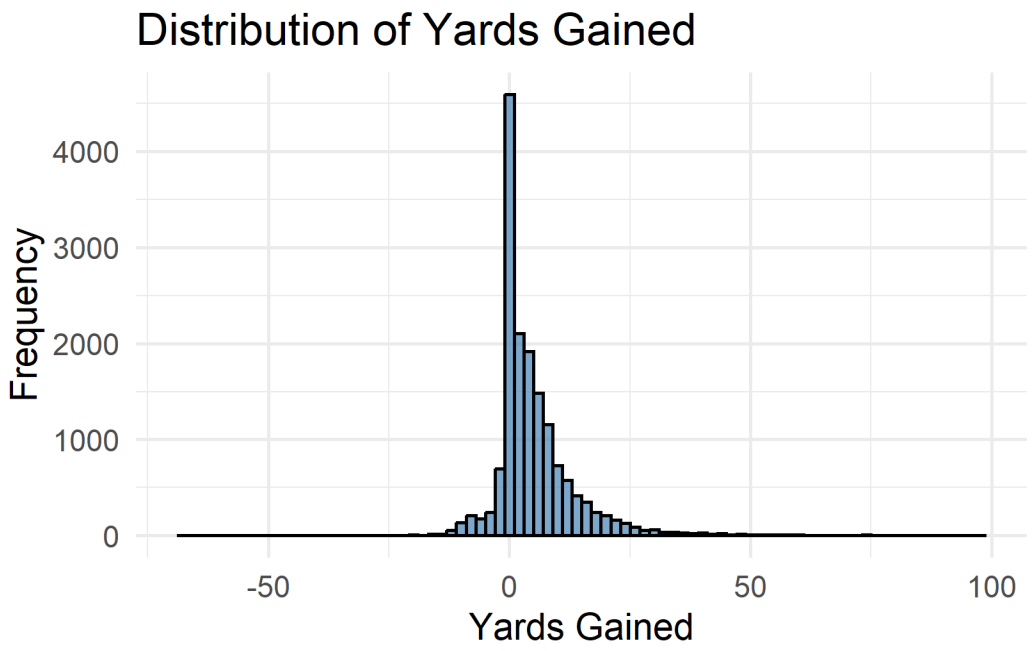
```r
  # Categorical: replace NA with mode (most frequent value)
  mode_value <- function(x) {
    ux <- unique(na.omit(x))
    ux[which.max(tabulate(match(x, ux)))]
  }
  for (col in names(df)[categorical_cols]) {
    df[[col]][is.na(df[[col]])] <- mode_value(df[[col]])
  }

  # Step 5: Verify cleanup
  remaining_missing <- colSums(is.na(df))
  message("\nRemaining missing values after handling:\n")
  print(remaining_missing)

  return(df)
}

clean_plays<-handle_missing_values(plays)
```

```
                          gameId                           playId
                               0                                0
               playDescription                          quarter
                               0                                0
                            down                        yardsToGo
                               0                                0
                  possessionTeam                     defensiveTeam
                               0                                0
                    yardlineSide                   yardlineNumber
                               0                                0
                       gameClock                  preSnapHomeScore
                               0                                0
             preSnapVisitorScore            playNullifiedByPenalty
                               0                                0
          absoluteYardlineNumber    preSnapHomeTeamWinProbability
                               0                                0
preSnapVisitorTeamWinProbability                  expectedPoints
                               0                                0
                 offenseFormation                receiverAlignment
                               0                                0
                  playClockAtSnap                       passResult
                               0                                0
```

```
                    passLength                                      targetX
                             0                                            0
                       targetY                                    playAction
                             0                                            0
                  dropbackType                               dropbackDistance
                             0                                            0
              passLocationType                                    timeToThrow
                             0                                            0
               timeInTackleBox                               passTippedAtLine
                             0                                            0
             unblockedPressure                                       qbSpike
                             0                                            0
                      qbKneel                                       qbSneak
                             0                                            0
               rushLocationType                        prePenaltyYardsGained
                             0                                            0
                   yardsGained                  homeTeamWinProbabilityAdded
                             0                                            0
  visitorTeamWinProbilityAdded                        expectedPointsAdded
                             0                                            0
                    isDropback                        pff_runConceptPrimary
                             0                                            0
              pff_runPassOption                             pff_passCoverage
                             0                                            0
                   pff_manZone
                             0
```

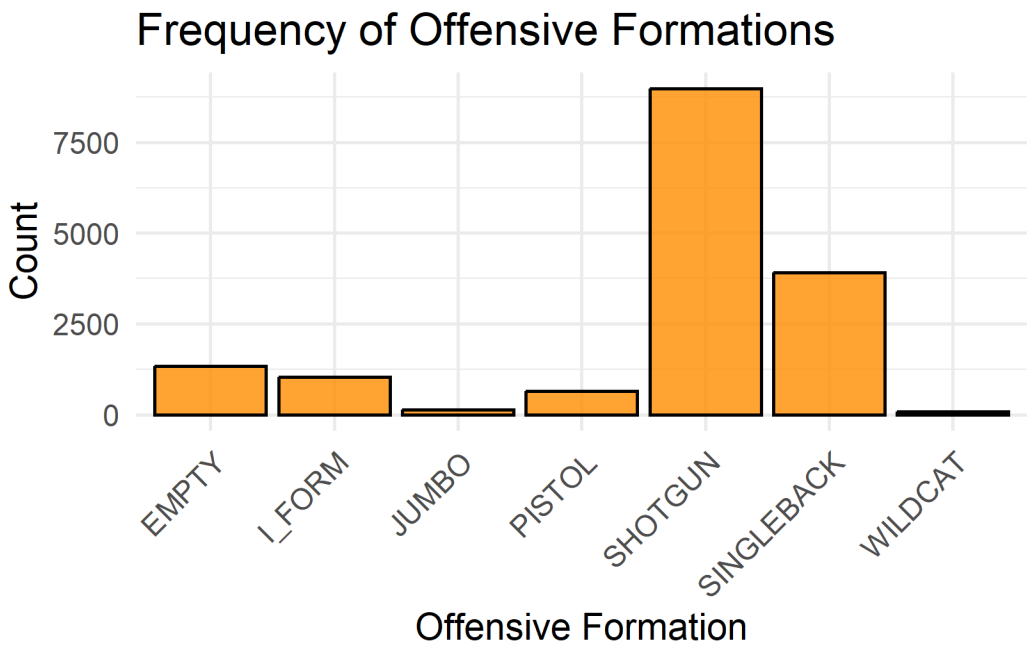
STEP 2: Relationship BETWEEN OFFENSE FORMATION AND YARDGAINED A: Basic distribtuion of YardGained

```
library(ggplot2)

ggplot(clean_plays, aes(x = yardsGained)) +
  geom_histogram(binwidth = 2, fill = "steelblue", color = "black", alpha = 0.7) +
  labs(title = "Distribution of Yards Gained",
       x = "Yards Gained",
       y = "Frequency") +
  theme_minimal(base_size = 14)
```



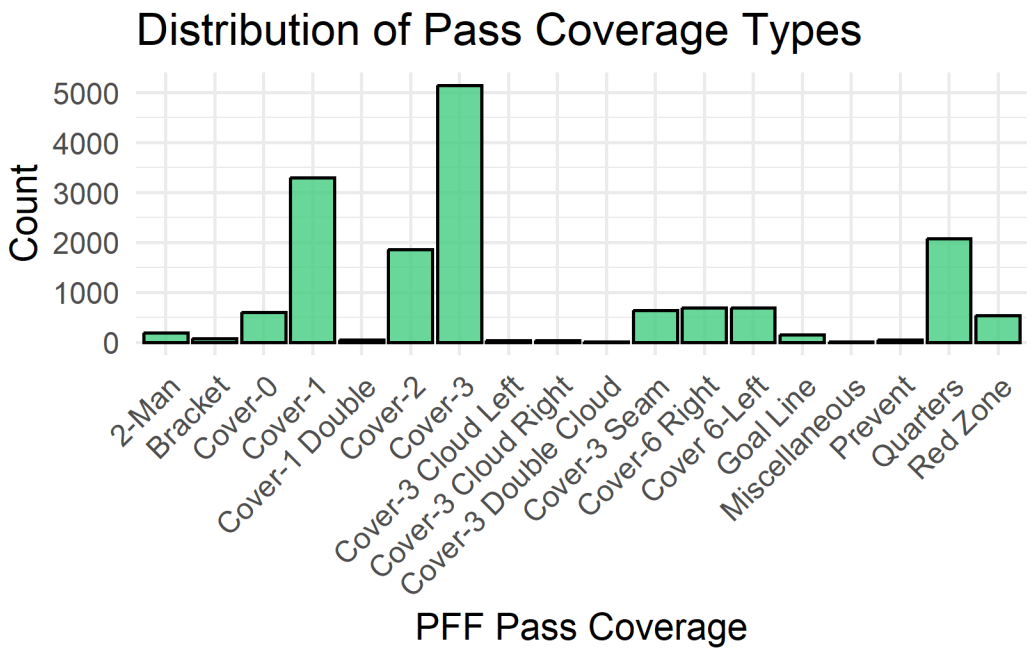
Barplot of OffenseFormation

```
ggplot(clean_plays, aes(x = offenseFormation)) +
  geom_bar(fill = "darkorange", color = "black", alpha = 0.8) +
  labs(title = "Frequency of Offensive Formations",
       x = "Offensive Formation",
       y = "Count") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Bar plot of distribution of passCoverage

```
ggplot(clean_plays, aes(x = pff_passCoverage)) +
  geom_bar(fill = "seagreen3", color = "black", alpha = 0.8) +
  labs(title = "Distribution of Pass Coverage Types",
       x = "PFF Pass Coverage",
       y = "Count") +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
formation_success <- clean_plays %>%
  group_by(offenseFormation) %>%
  summarise(Average_Yards_Gained = mean(yardsGained, na.rm = TRUE)) %>%
  arrange(desc(Average_Yards_Gained))

# Step 2: Print results in styled console output
cat("\nAverage Yards Gained by Formation:\n")
```

```
Average Yards Gained by Formation:
```

```
cat(strrep("=", 50), "\n")
```

```
==================================================
```

```
cat(sprintf("%-25s %-25s\n", "Formation", "Average Yards Gained"))
```

```
Formation                 Average Yards Gained
```

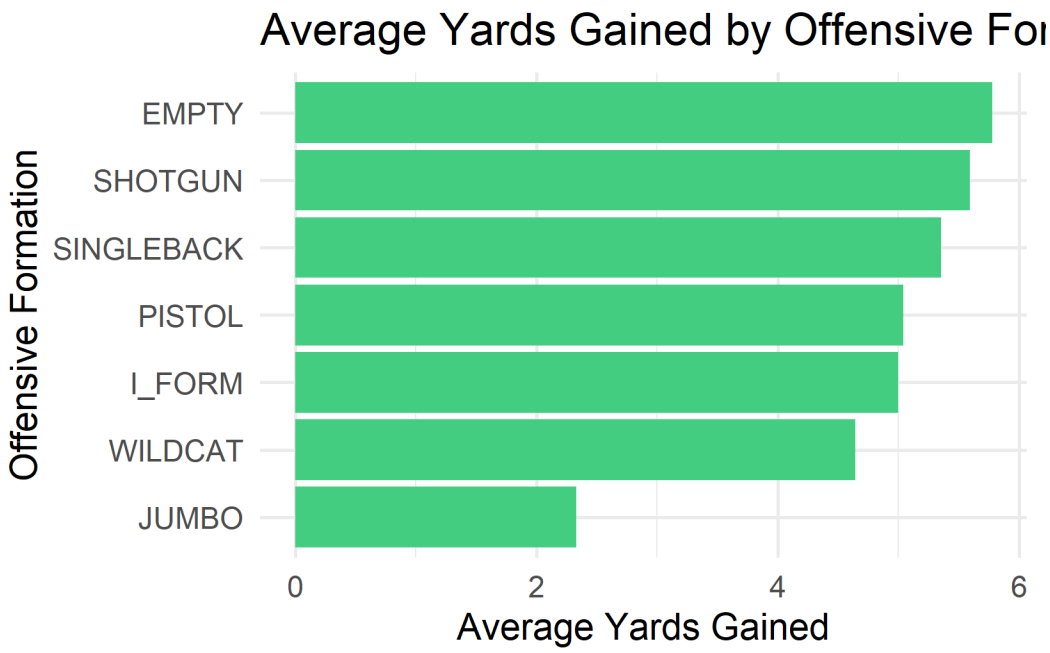
```r
cat(strrep("-", 50), "\n")
```

```
--------------------------------------------------
```

```r
for (i in seq_len(nrow(formation_success))) {
  formation <- formation_success$offenseFormation[i]
  yards <- formation_success$Average_Yards_Gained[i]


  cat(sprintf("%-25s %-25.2f\n", formation, yards))
}
```

```
EMPTY                     5.78
SHOTGUN                   5.59
SINGLEBACK                5.35
PISTOL                    5.04
I_FORM                    5.00
WILDCAT                   4.64
JUMBO                     2.33
```

```r
# Step 3: Visualization (bar chart)
ggplot(formation_success, aes(x = reorder(offenseFormation, Average_Yards_Gained),
                              y = Average_Yards_Gained,
                              fill = Average_Yards_Gained>0))+
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("TRUE" = "seagreen3", "FALSE" = "firebrick2")) +
  coord_flip() +
  labs(
    title = "Average Yards Gained by Offensive Formation",
    x = "Offensive Formation",
    y = "Average Yards Gained"
  ) +
  theme_minimal(base_size = 14) +
  theme(legend.position = "none")
```



Next is OffenseFormation by PassCoverage

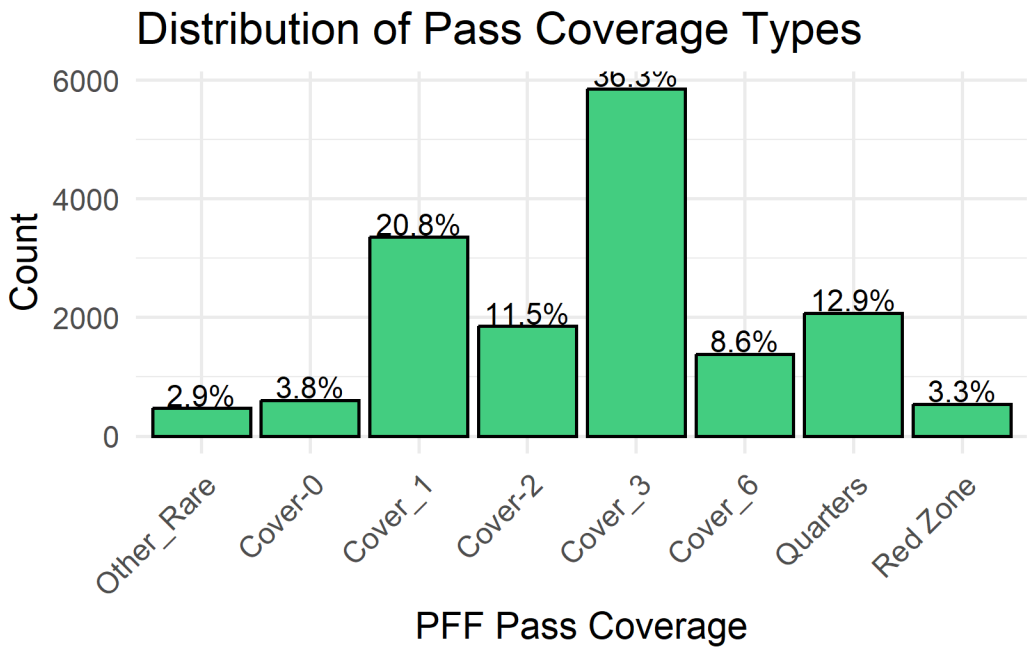
Let's group the pass_coverage to smaller group

```
clean_plays$coverage_group <- clean_plays$pff_passCoverage

clean_plays$coverage_group <- forcats::fct_collapse(
  clean_plays$coverage_group,
  Cover_3 = c("Cover-3", "Cover-3 Cloud Left", "Cover-3 Cloud Right",
              "Cover-3 Seam", "Cover-3 Double Cloud"),
  Cover_6 = c("Cover 6-Left", "Cover-6 Right"),
  Cover_1 = c("Cover-1", "Cover-1 Double"),
  Other_Rare = c("Bracket", "Miscellaneous", "Prevent", "2-Man", "Goal Line")
)
```

See the distribution after grouping

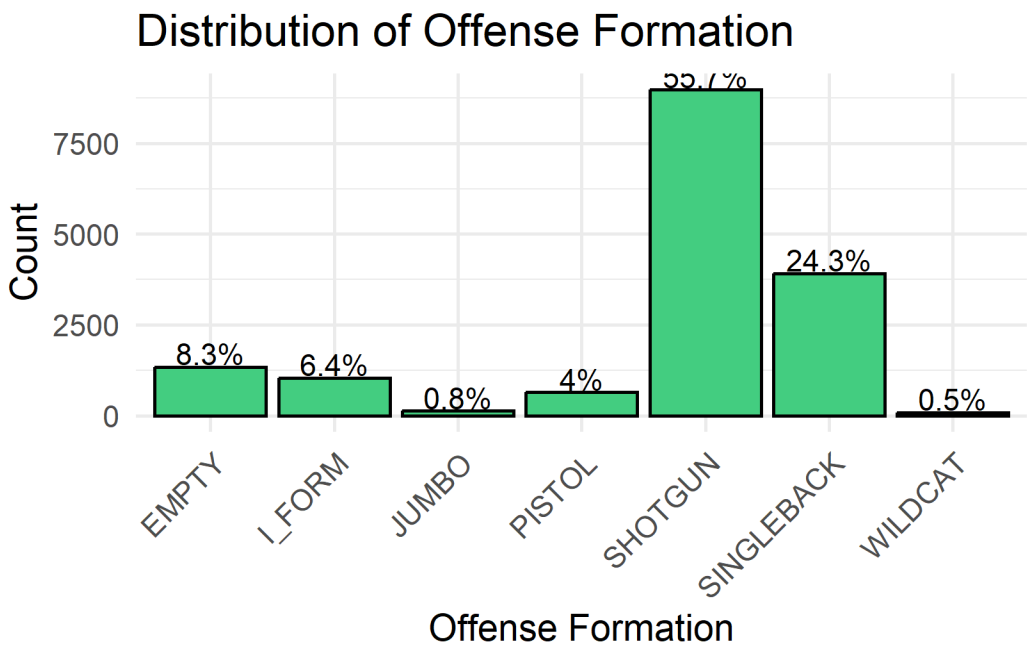
```
ggplot(clean_plays, aes(x = coverage_group)) +
  geom_bar(fill = "seagreen3", color = "black") +
  geom_text(
    stat = "count",
    aes(label = paste0(round((..count..) / sum(..count..) * 100, 1), "%")),
    vjust = -0.1,
```

```
size = 4
  ) +
  labs(
    title = "Distribution of Pass Coverage Types",
    x = "PFF Pass Coverage",
    y = "Count"
  ) +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
ggplot(clean_plays, aes(x = offenseFormation)) +
  geom_bar(fill = "seagreen3", color = "black") +
  geom_text(
    stat = "count",
    aes(label = paste0(round((..count..) / sum(..count..) * 100, 1), "%")),
    vjust = -0.1,
    size = 4
  ) +
  labs(
    title = "Distribution of Offense Formation",
    x = "Offense Formation",
    y = "Count"
```

```
  ) +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



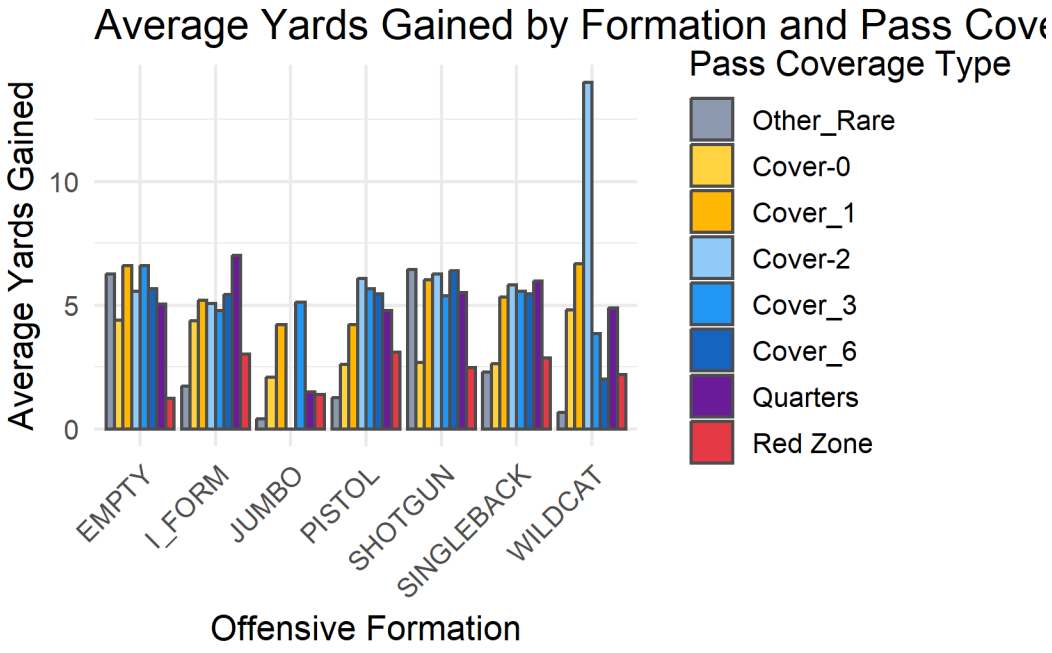
Now run the plot

```
library(dplyr)

yards_by_formation_cov <- clean_plays %>%
  group_by(offenseFormation, coverage_group) %>%
  summarise(
    avg_yards = mean(yardsGained, na.rm = TRUE),
    play_count = n(),
    .groups = "drop"
  )
```

Visualize it

```
library(ggplot2)

ggplot(yards_by_formation_cov,
```

```
      aes(x = offenseFormation, y = avg_yards, fill = coverage_group)) +
geom_bar(stat = "identity", position = position_dodge(width = 0.9), color = "gray30") +
scale_fill_manual(
  values = c(
    "Cover-0"   = "#FFD23F",  # golden yellow
    "Cover_1"   = "#FFB703",  # orange-yellow
    "Cover-2"   = "#90CAF9",  # light blue
    "Cover_3"   = "#2196F3",  # mid blue
    "Cover_6"   = "#1565C0",  # deep blue
    "Quarters"  = "#6A1B9A",  # purple accent
    "Red Zone"  = "#E63946",  # red
    "Other_Rare"= "#8D99AE"   # neutral gray
  )
) +
labs(
  title = "Average Yards Gained by Formation and Pass Coverage",
  x = "Offensive Formation",
  y = "Average Yards Gained",
  fill = "Pass Coverage Type"
)+
theme_minimal(base_size = 13) +
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
# Group 1: First four coverage types
cover_group1 <- c("Other_Rare", "Cover-0", "Cover_1", "Cover-2")
# Group 2: Remaining four coverage types
cover_group2 <- c("Cover_3", "Cover_6", "Quarters", "Red Zone")

# Split data
clean_plays_1 <- subset(clean_plays, coverage_group %in% cover_group1)
clean_plays_2 <- subset(clean_plays, coverage_group %in% cover_group2)
```

OK NOW OFFENSIVE FORMATION IN DIFFERENT QUARTER

Compute the yardGain by formation

```r
library(dplyr)

# Summarize average yards by formation and quarter
yards_by_quarter <- clean_plays %>%
  group_by(offenseFormation, quarter) %>%
  summarise(
    avg_yards = mean(yardsGained, na.rm = TRUE),
    play_count = n(),
    .groups = "drop"
  )

# Print results
cat("\nAverage Yards Gained by Formation and Quarter:\n")
```

```
Average Yards Gained by Formation and Quarter:
```

```r
cat(strrep("=", 60), "\n")
```

```
============================================================
```

```r
cat(sprintf("%-15s %-10s %-20s %-10s\n",
            "Formation", "Quarter", "Average Yards", "Plays"))
```

```
Formation       Quarter    Average Yards        Plays
```

```r
cat(strrep("-", 60), "\n")
```

```
------------------------------------------------------------
```

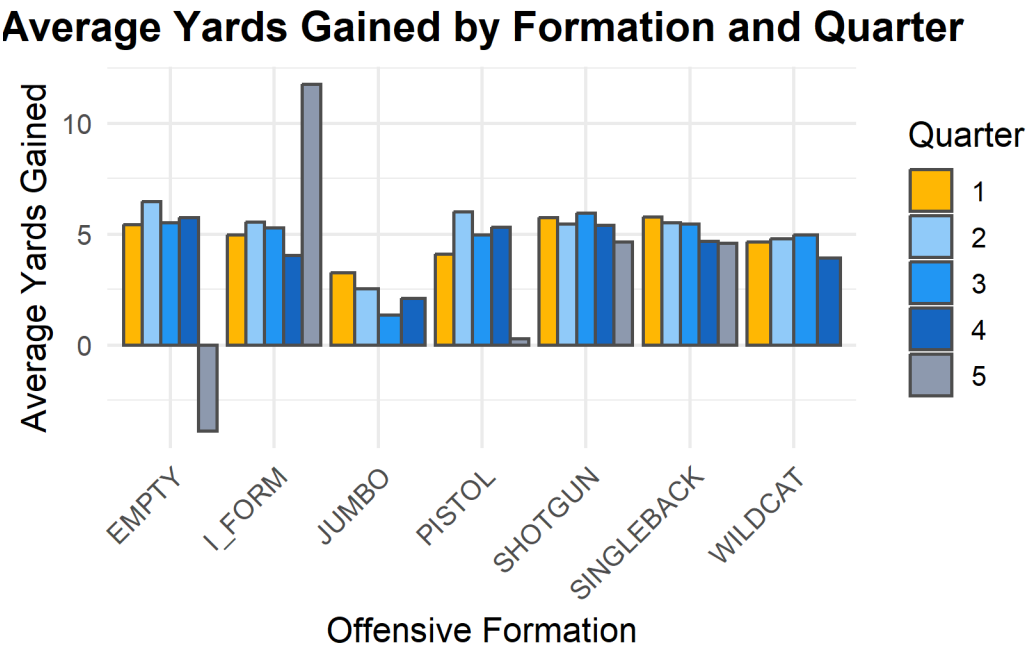
```r
for (i in seq_len(nrow(yards_by_quarter))) {
  formation <- yards_by_quarter$offenseFormation[i]
  quarter <- yards_by_quarter$quarter[i]
  yards <- yards_by_quarter$avg_yards[i]
  count <- yards_by_quarter$play_count[i]

  cat(sprintf("%-15s %-10s %-20.2f %-10d\n", formation, quarter, yards, count))
}
```

```
EMPTY           1          5.42                313
EMPTY           2          6.47                402
EMPTY           3          5.51                278
EMPTY           4          5.74                341
EMPTY           5          -3.88               8
I_FORM          1          4.97                320
I_FORM          2          5.55                251
I_FORM          3          5.29                228
I_FORM          4          4.03                232
I_FORM          5          11.75               4
JUMBO           1          3.26                27
JUMBO           2          2.54                35
JUMBO           3          1.36                25
JUMBO           4          2.11                38
PISTOL          1          4.09                159
PISTOL          2          5.99                164
PISTOL          3          4.97                164
PISTOL          4          5.33                147
PISTOL          5          0.29                7
SHOTGUN         1          5.74                1774
SHOTGUN         2          5.47                2583
SHOTGUN         3          5.96                1844
SHOTGUN         4          5.40                2695
SHOTGUN         5          4.65                83
SINGLEBACK      1          5.77                984
SINGLEBACK      2          5.52                918
SINGLEBACK      3          5.46                1034
SINGLEBACK      4          4.67                947
```

```
SINGLEBACK      5          4.59                32
WILDCAT         1          4.64                25
WILDCAT         2          4.78                23
WILDCAT         3          4.96                24
WILDCAT         4          3.93                15
```

GGPlot

```
library(ggplot2)

ggplot(yards_by_quarter,
       aes(x = offenseFormation, y = avg_yards, fill = factor(quarter))) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.9), color = "gray30") +
  scale_fill_manual(
    values = c(
      "1" = "#FFB703",  # orange-yellow
      "2" = "#90CAF9",  # light blue
      "3" = "#2196F3",  # mid blue
      "4" = "#1565C0",   # deep blue
      "5" = "#8D99AE"   # gray-blue for OT
    )
  ) +
  labs(
    title = "Average Yards Gained by Formation and Quarter",
    x = "Offensive Formation",
    y = "Average Yards Gained",
    fill = "Quarter"
  ) +
  theme_minimal(base_size = 13) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold", hjust = 0.5)
  )
```



CONDUCT AN ANOVA TEST

```
# Fit model with interaction term
model_interaction <- lm(yardsGained ~ offenseFormation *coverage_group , data = clean_plays)

# Show summary
summary(model_interaction)
```

```
Call:
lm(formula = yardsGained ~ offenseFormation * coverage_group,
    data = clean_plays)

Residuals:
    Min      1Q  Median      3Q     Max
-74.024  -5.385  -1.970   2.620  91.560

Coefficients: (1 not defined because of singularities)
                                                    Estimate Std. Error t value
(Intercept)                                           6.2683     1.3730   4.565
offenseFormationI_FORM                               -4.5350     2.1123  -2.147
offenseFormationJUMBO                                -5.8721     1.8285  -3.211
```

offenseFormationPISTOL	-5.0183	4.6053	-1.090
offenseFormationSHOTGUN	0.1714	1.4695	0.117
offenseFormationSINGLEBACK	-3.9720	1.8211	-2.181
offenseFormationWILDCAT	-5.6016	5.2583	-1.065
coverage_groupCover-0	-1.8739	1.7245	-1.087
coverage_groupCover_1	0.3258	1.4732	0.221
coverage_groupCover-2	-0.7191	1.5118	-0.476
coverage_groupCover_3	0.3241	1.4475	0.224
coverage_groupCover_6	-0.6047	1.6087	-0.376
coverage_groupQuarters	-1.2313	1.4843	-0.829
coverage_groupRed Zone	-5.0461	1.8981	-2.658
offenseFormationI_FORM:coverage_groupCover-0	4.5229	2.7971	1.617
offenseFormationJUMBO:coverage_groupCover-0	3.5777	3.4873	1.026
offenseFormationPISTOL:coverage_groupCover-0	3.2239	5.0387	0.640
offenseFormationSHOTGUN:coverage_groupCover-0	-1.8700	1.8703	-1.000
offenseFormationSINGLEBACK:coverage_groupCover-0	2.2205	2.2265	0.997
offenseFormationWILDCAT:coverage_groupCover-0	6.0198	5.7939	1.039
offenseFormationI_FORM:coverage_groupCover_1	3.1387	2.2741	1.380
offenseFormationJUMBO:coverage_groupCover_1	3.4863	2.6171	1.332
offenseFormationPISTOL:coverage_groupCover_1	2.6454	4.7093	0.562
offenseFormationSHOTGUN:coverage_groupCover_1	-0.7411	1.5760	-0.470
offenseFormationSINGLEBACK:coverage_groupCover_1	2.7104	1.9235	1.409
offenseFormationWILDCAT:coverage_groupCover_1	5.6742	5.8631	0.968
offenseFormationI_FORM:coverage_groupCover-2	4.0602	2.3842	1.703
offenseFormationJUMBO:coverage_groupCover-2	0.3228	4.8029	0.067
offenseFormationPISTOL:coverage_groupCover-2	5.5502	4.8681	1.140
offenseFormationSHOTGUN:coverage_groupCover-2	0.5361	1.6201	0.331
offenseFormationSINGLEBACK:coverage_groupCover-2	4.2343	1.9880	2.130
offenseFormationWILDCAT:coverage_groupCover-2	14.0524	7.3358	1.916
offenseFormationI_FORM:coverage_groupCover_3	2.7364	2.1980	1.245
offenseFormationJUMBO:coverage_groupCover_3	4.3908	2.5330	1.733
offenseFormationPISTOL:coverage_groupCover_3	4.0937	4.6568	0.879
offenseFormationSHOTGUN:coverage_groupCover_3	-1.3788	1.5480	-0.891
offenseFormationSINGLEBACK:coverage_groupCover_3	2.9540	1.8893	1.564
offenseFormationWILDCAT:coverage_groupCover_3	2.8492	5.5634	0.512
offenseFormationI_FORM:coverage_groupCover_6	4.2982	2.4712	1.739
offenseFormationJUMBO:coverage_groupCover_6	NA	NA	NA
offenseFormationPISTOL:coverage_groupCover_6	4.8092	4.8650	0.989
offenseFormationSHOTGUN:coverage_groupCover_6	0.5447	1.7172	0.317
offenseFormationSINGLEBACK:coverage_groupCover_6	3.7711	2.0790	1.814
offenseFormationWILDCAT:coverage_groupCover_6	1.9380	10.2784	0.189
offenseFormationI_FORM:coverage_groupQuarters	6.5081	2.3598	2.758
offenseFormationJUMBO:coverage_groupQuarters	2.3350	6.5045	0.359

```
offenseFormationPISTOL:coverage_groupQuarters       4.7776     4.7162   1.013
offenseFormationSHOTGUN:coverage_groupQuarters      0.2967     1.5948   0.186
offenseFormationSINGLEBACK:coverage_groupQuarters   4.9051     1.9525   2.512
offenseFormationWILDCAT:coverage_groupQuarters      5.4469     5.7021   0.955
offenseFormationI_FORM:coverage_groupRed Zone       6.3461     2.9590   2.145
offenseFormationJUMBO:coverage_groupRed Zone        6.0498     4.5299   1.336
offenseFormationPISTOL:coverage_groupRed Zone       6.8913     5.1582   1.336
offenseFormationSHOTGUN:coverage_groupRed Zone      1.0839     2.0309   0.534
offenseFormationSINGLEBACK:coverage_groupRed Zone   5.6270     2.3900   2.354
offenseFormationWILDCAT:coverage_groupRed Zone      6.5794     6.0907   1.080
                                                  Pr(>|t|)
(Intercept)                                       5.02e-06 ***
offenseFormationI_FORM                             0.03181 *
offenseFormationJUMBO                              0.00132 **
offenseFormationPISTOL                             0.27587
offenseFormationSHOTGUN                            0.90713
offenseFormationSINGLEBACK                         0.02919 *
offenseFormationWILDCAT                            0.28676
coverage_groupCover-0                              0.27720
coverage_groupCover_1                              0.82498
coverage_groupCover-2                              0.63435
coverage_groupCover_3                              0.82284
coverage_groupCover_6                              0.70702
coverage_groupQuarters                             0.40684
coverage_groupRed Zone                             0.00786 **
offenseFormationI_FORM:coverage_groupCover-0       0.10589
offenseFormationJUMBO:coverage_groupCover-0        0.30495
offenseFormationPISTOL:coverage_groupCover-0       0.52229
offenseFormationSHOTGUN:coverage_groupCover-0      0.31741
offenseFormationSINGLEBACK:coverage_groupCover-0   0.31863
offenseFormationWILDCAT:coverage_groupCover-0      0.29883
offenseFormationI_FORM:coverage_groupCover_1       0.16755
offenseFormationJUMBO:coverage_groupCover_1        0.18284
offenseFormationPISTOL:coverage_groupCover_1       0.57430
offenseFormationSHOTGUN:coverage_groupCover_1      0.63820
offenseFormationSINGLEBACK:coverage_groupCover_1   0.15883
offenseFormationWILDCAT:coverage_groupCover_1      0.33317
offenseFormationI_FORM:coverage_groupCover-2       0.08859 .
offenseFormationJUMBO:coverage_groupCover-2        0.94641
offenseFormationPISTOL:coverage_groupCover-2       0.25426
offenseFormationSHOTGUN:coverage_groupCover-2      0.74073
offenseFormationSINGLEBACK:coverage_groupCover-2   0.03319 *
offenseFormationWILDCAT:coverage_groupCover-2      0.05544 .
```

```
offenseFormationI_FORM:coverage_groupCover_3       0.21316
offenseFormationJUMBO:coverage_groupCover_3        0.08304 .
offenseFormationPISTOL:coverage_groupCover_3       0.37937
offenseFormationSHOTGUN:coverage_groupCover_3      0.37309
offenseFormationSINGLEBACK:coverage_groupCover_3   0.11794
offenseFormationWILDCAT:coverage_groupCover_3      0.60856
offenseFormationI_FORM:coverage_groupCover_6       0.08200 .
offenseFormationJUMBO:coverage_groupCover_6             NA
offenseFormationPISTOL:coverage_groupCover_6       0.32290
offenseFormationSHOTGUN:coverage_groupCover_6      0.75109
offenseFormationSINGLEBACK:coverage_groupCover_6   0.06971 .
offenseFormationWILDCAT:coverage_groupCover_6      0.85045
offenseFormationI_FORM:coverage_groupQuarters      0.00582 **
offenseFormationJUMBO:coverage_groupQuarters       0.71961
offenseFormationPISTOL:coverage_groupQuarters      0.31107
offenseFormationSHOTGUN:coverage_groupQuarters     0.85243
offenseFormationSINGLEBACK:coverage_groupQuarters  0.01201 *
offenseFormationWILDCAT:coverage_groupQuarters     0.33947
offenseFormationI_FORM:coverage_groupRed Zone      0.03199 *
offenseFormationJUMBO:coverage_groupRed Zone       0.18172
offenseFormationPISTOL:coverage_groupRed Zone      0.18157
offenseFormationSHOTGUN:coverage_groupRed Zone     0.59355
offenseFormationSINGLEBACK:coverage_groupRed Zone  0.01857 *
offenseFormationWILDCAT:coverage_groupRed Zone     0.28005
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.792 on 16069 degrees of freedom
Multiple R-squared:  0.01268,   Adjusted R-squared:  0.009358
F-statistic: 3.821 on 54 and 16069 DF,  p-value: < 2.2e-16
```

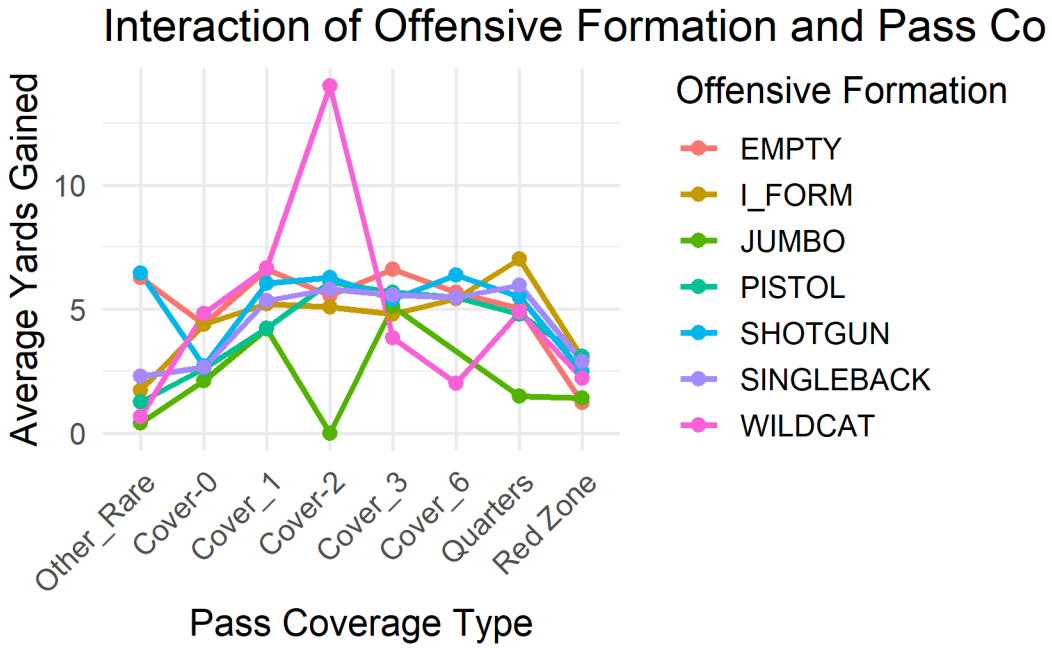
```r
anova(model_interaction)
```

```
Analysis of Variance Table

Response: yardsGained
                                Df  Sum Sq Mean Sq F value    Pr(>F)
offenseFormation                 6    1960  326.61  4.2256 0.0002958 ***
coverage_group                   7    9581 1368.72 17.7081 < 2.2e-16 ***
offenseFormation:coverage_group 41    4406  107.46  1.3903 0.0497594 *
Residuals                       16069 1242028   77.29
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

GGPlot INteraction plot

```
library(ggplot2)
library(dplyr)

interaction_summary <- clean_plays %>%
  group_by(offenseFormation, coverage_group) %>%
  summarise(mean_yards = mean(yardsGained, na.rm = TRUE))

ggplot(interaction_summary,
       aes(x = coverage_group, y = mean_yards, color = offenseFormation, group = offenseForma
  geom_line(linewidth = 1.1) +
  geom_point(size = 2) +
  labs(
    title = "Interaction of Offensive Formation and Pass Coverage on Yards Gained",
    x = "Pass Coverage Type",
    y = "Average Yards Gained",
    color = "Offensive Formation"
  ) +
  theme_minimal(base_size = 14) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



```r
library(emmeans)

# Get estimated marginal means for each combination
emm <- emmeans(model_interaction, ~ offenseFormation | coverage_group)

# Compare formations within each coverage
pairs(emm, adjust = "fdr")
```

```
coverage_group = Other_Rare:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         4.5350 2.110 16069   2.147  0.1113
 EMPTY - JUMBO          5.8721 1.830 16069   3.211  0.0106
 EMPTY - PISTOL         5.0183 4.610 16069   1.090  0.5474
 EMPTY - SHOTGUN       -0.1714 1.470 16069  -0.117  0.9587
 EMPTY - SINGLEBACK     3.9720 1.820 16069   2.181  0.1113
 EMPTY - WILDCAT        5.6016 5.260 16069   1.065  0.5474
 I_FORM - JUMBO         1.3371 2.010 16069   0.666  0.8849
 I_FORM - PISTOL        0.4833 4.680 16069   0.103  0.9587
 I_FORM - SHOTGUN      -4.7064 1.690 16069  -2.788  0.0279
 I_FORM - SINGLEBACK   -0.5630 2.000 16069  -0.281  0.9587
 I_FORM - WILDCAT       1.0667 5.320 16069   0.200  0.9587
 JUMBO - PISTOL        -0.8538 4.560 16069  -0.187  0.9587
 JUMBO - SHOTGUN       -6.0435 1.320 16069  -4.592  0.0001
 JUMBO - SINGLEBACK    -1.9001 1.700 16069  -1.118  0.5474
 JUMBO - WILDCAT       -0.2704 5.220 16069  -0.052  0.9587
 PISTOL - SHOTGUN      -5.1897 4.430 16069  -1.172  0.5474
 PISTOL - SINGLEBACK   -1.0463 4.560 16069  -0.230  0.9587
 PISTOL - WILDCAT       0.5833 6.710 16069   0.087  0.9587
 SHOTGUN - SINGLEBACK   4.1434 1.310 16069   3.173  0.0106
 SHOTGUN - WILDCAT      5.7730 5.100 16069   1.131  0.5474
 SINGLEBACK - WILDCAT   1.6296 5.210 16069   0.312  0.9587

coverage_group = Cover-0:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         0.0120 1.830 16069   0.007  0.9948
 EMPTY - JUMBO          2.2944 2.970 16069   0.773  0.8234
 EMPTY - PISTOL         1.7944 2.040 16069   0.878  0.8234
 EMPTY - SHOTGUN        1.6986 1.160 16069   1.468  0.8234
 EMPTY - SINGLEBACK     1.7515 1.280 16069   1.367  0.8234
 EMPTY - WILDCAT       -0.4181 2.430 16069  -0.172  0.9948
 I_FORM - JUMBO         2.2824 3.160 16069   0.722  0.8234
 I_FORM - PISTOL        1.7824 2.320 16069   0.769  0.8234
```

```
 I_FORM - SHOTGUN       1.6866 1.590 16069   1.062  0.8234
 I_FORM - SINGLEBACK    1.7395 1.680 16069   1.035  0.8234
 I_FORM - WILDCAT      -0.4301 2.670 16069  -0.161  0.9948
 JUMBO - PISTOL        -0.5000 3.290 16069  -0.152  0.9948
 JUMBO - SHOTGUN       -0.5958 2.820 16069  -0.211  0.9948
 JUMBO - SINGLEBACK    -0.5429 2.880 16069  -0.189  0.9948
 JUMBO - WILDCAT       -2.7125 3.540 16069  -0.765  0.8234
 PISTOL - SHOTGUN      -0.0958 1.830 16069  -0.052  0.9948
 PISTOL - SINGLEBACK   -0.0429 1.910 16069  -0.022  0.9948
 PISTOL - WILDCAT      -2.2125 2.810 16069  -0.786  0.8234
 SHOTGUN - SINGLEBACK   0.0529 0.896 16069   0.059  0.9948
 SHOTGUN - WILDCAT     -2.1167 2.250 16069  -0.939  0.8234
 SINGLEBACK - WILDCAT  -2.1696 2.320 16069  -0.935  0.8234

coverage_group = Cover_1:
 contrast              estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         1.3963 0.843 16069   1.657  0.4095
 EMPTY - JUMBO          2.3858 1.870 16069   1.274  0.5905
 EMPTY - PISTOL         2.3729 0.984 16069   2.410  0.2911
 EMPTY - SHOTGUN        0.5697 0.570 16069   1.000  0.6293
 EMPTY - SINGLEBACK     1.2616 0.619 16069   2.038  0.2911
 EMPTY - WILDCAT       -0.0726 2.590 16069  -0.028  0.9948
 I_FORM - JUMBO         0.9895 1.910 16069   0.518  0.7465
 I_FORM - PISTOL        0.9766 1.050 16069   0.927  0.6293
 I_FORM - SHOTGUN      -0.8266 0.681 16069  -1.213  0.5905
 I_FORM - SINGLEBACK   -0.1347 0.723 16069  -0.186  0.9419
 I_FORM - WILDCAT      -1.4689 2.620 16069  -0.561  0.7465
 JUMBO - PISTOL        -0.0129 1.980 16069  -0.007  0.9948
 JUMBO - SHOTGUN       -1.8161 1.810 16069  -1.006  0.6293
 JUMBO - SINGLEBACK    -1.1242 1.820 16069  -0.617  0.7465
 JUMBO - WILDCAT       -2.4583 3.110 16069  -0.791  0.6930
 PISTOL - SHOTGUN      -1.8032 0.851 16069  -2.120  0.2911
 PISTOL - SINGLEBACK   -1.1112 0.884 16069  -1.257  0.5905
 PISTOL - WILDCAT      -2.4454 2.670 16069  -0.916  0.6293
 SHOTGUN - SINGLEBACK   0.6920 0.371 16069   1.866  0.3255
 SHOTGUN - WILDCAT     -0.6422 2.550 16069  -0.252  0.9343
 SINGLEBACK - WILDCAT  -1.3342 2.560 16069  -0.522  0.7465

coverage_group = Cover-2:
 contrast              estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         0.4748 1.110 16069   0.429  0.8202
 EMPTY - JUMBO          5.5492 4.440 16069   1.250  0.4038
 EMPTY - PISTOL        -0.5319 1.580 16069  -0.337  0.8202
```

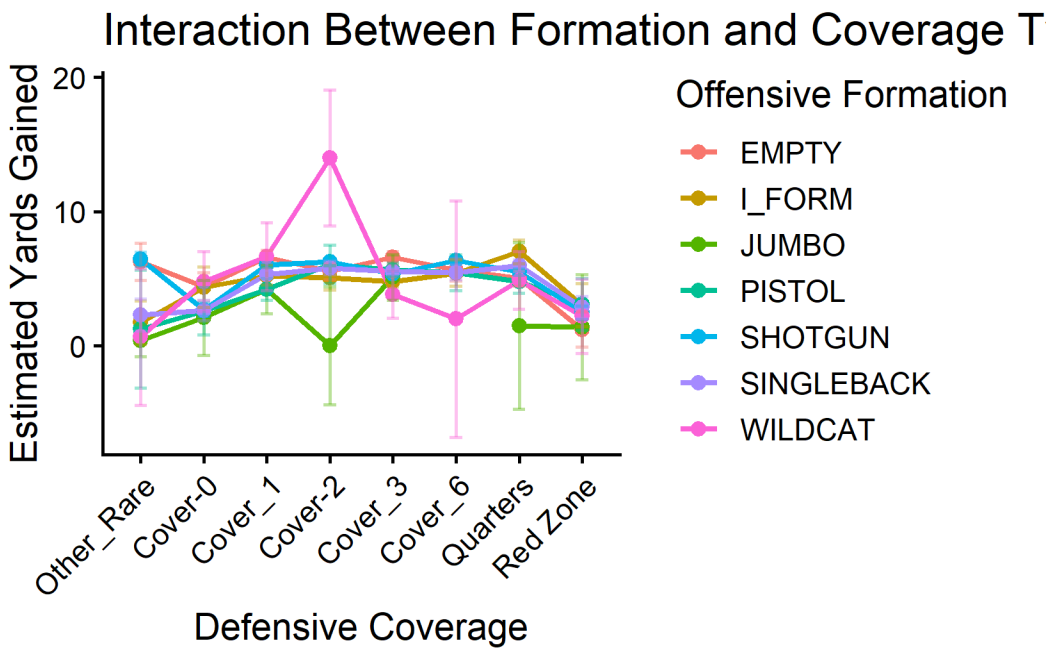
```
 EMPTY - SHOTGUN       -0.7075 0.682 16069  -1.037  0.4841
 EMPTY - SINGLEBACK    -0.2623 0.797 16069  -0.329  0.8202
 EMPTY - WILDCAT       -8.4508 5.120 16069  -1.652  0.4038
 I_FORM - JUMBO         5.0745 4.490 16069   1.131  0.4519
 I_FORM - PISTOL       -1.0066 1.710 16069  -0.590  0.7287
 I_FORM - SHOTGUN      -1.1822 0.942 16069  -1.255  0.4038
 I_FORM - SINGLEBACK   -0.7371 1.030 16069  -0.717  0.6629
 I_FORM - WILDCAT      -8.9255 5.160 16069  -1.731  0.4038
 JUMBO - PISTOL        -6.0811 4.630 16069  -1.314  0.4038
 JUMBO - SHOTGUN       -6.2567 4.400 16069  -1.421  0.4038
 JUMBO - SINGLEBACK    -5.8116 4.420 16069  -1.314  0.4038
 JUMBO - WILDCAT      -14.0000 6.710 16069  -2.085  0.4038
 PISTOL - SHOTGUN      -0.1756 1.470 16069  -0.120  0.9047
 PISTOL - SINGLEBACK    0.2695 1.520 16069   0.177  0.9026
 PISTOL - WILDCAT      -7.9189 5.280 16069  -1.500  0.4038
 SHOTGUN - SINGLEBACK   0.4452 0.548 16069   0.813  0.6243
 SHOTGUN - WILDCAT     -7.7433 5.080 16069  -1.524  0.4038
 SINGLEBACK - WILDCAT  -8.1884 5.100 16069  -1.606  0.4038

coverage_group = Cover_3:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         1.7986 0.608 16069   2.959  0.0649
 EMPTY - JUMBO          1.4813 1.750 16069   0.845  0.6967
 EMPTY - PISTOL         0.9246 0.691 16069   1.338  0.4751
 EMPTY - SHOTGUN        1.2074 0.487 16069   2.480  0.1381
 EMPTY - SINGLEBACK     1.0179 0.503 16069   2.024  0.3010
 EMPTY - WILDCAT        2.7524 1.820 16069   1.515  0.4751
 I_FORM - JUMBO        -0.3173 1.740 16069  -0.183  0.8720
 I_FORM - PISTOL       -0.8740 0.653 16069  -1.338  0.4751
 I_FORM - SHOTGUN      -0.5912 0.432 16069  -1.369  0.4751
 I_FORM - SINGLEBACK   -0.7806 0.450 16069  -1.736  0.4340
 I_FORM - WILDCAT       0.9538 1.800 16069   0.529  0.7907
 JUMBO - PISTOL        -0.5567 1.770 16069  -0.315  0.8720
 JUMBO - SHOTGUN       -0.2739 1.700 16069  -0.161  0.8720
 JUMBO - SINGLEBACK    -0.4633 1.700 16069  -0.272  0.8720
 JUMBO - WILDCAT        1.2711 2.440 16069   0.521  0.7907
 PISTOL - SHOTGUN       0.2829 0.543 16069   0.521  0.7907
 PISTOL - SINGLEBACK    0.0934 0.557 16069   0.168  0.8720
 PISTOL - WILDCAT       1.8278 1.830 16069   0.997  0.6873
 SHOTGUN - SINGLEBACK  -0.1895 0.265 16069  -0.716  0.7654
 SHOTGUN - WILDCAT      1.5450 1.770 16069   0.875  0.6967
 SINGLEBACK - WILDCAT   1.7344 1.770 16069   0.980  0.6873
```

```
coverage_group = Cover_6:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM         0.2368 1.280 16069   0.185  0.9954
 EMPTY - JUMBO          nonEst    NA    NA      NA      NA
 EMPTY - PISTOL         0.2091 1.570 16069   0.133  0.9954
 EMPTY - SHOTGUN       -0.7161 0.889 16069  -0.806  0.9954
 EMPTY - SINGLEBACK     0.2009 1.000 16069   0.200  0.9954
 EMPTY - WILDCAT        3.6636 8.830 16069   0.415  0.9954
 I_FORM - JUMBO         nonEst    NA    NA      NA      NA
 I_FORM - PISTOL       -0.0277 1.640 16069  -0.017  0.9954
 I_FORM - SHOTGUN      -0.9529 1.010 16069  -0.939  0.9954
 I_FORM - SINGLEBACK   -0.0359 1.120 16069  -0.032  0.9954
 I_FORM - WILDCAT       3.4268 8.850 16069   0.387  0.9954
 JUMBO - PISTOL         nonEst    NA    NA      NA      NA
 JUMBO - SHOTGUN        nonEst    NA    NA      NA      NA
 JUMBO - SINGLEBACK     nonEst    NA    NA      NA      NA
 JUMBO - WILDCAT        nonEst    NA    NA      NA      NA
 PISTOL - SHOTGUN      -0.9252 1.360 16069  -0.681  0.9954
 PISTOL - SINGLEBACK   -0.0082 1.440 16069  -0.006  0.9954
 PISTOL - WILDCAT       3.4545 8.890 16069   0.389  0.9954
 SHOTGUN - SINGLEBACK   0.9170 0.624 16069   1.469  0.9954
 SHOTGUN - WILDCAT      4.3798 8.800 16069   0.498  0.9954
 SINGLEBACK - WILDCAT   3.4627 8.810 16069   0.393  0.9954

coverage_group = Quarters:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM        -1.9732 1.050 16069  -1.876  0.7249
 EMPTY - JUMBO          3.5370 6.240 16069   0.567  0.7619
 EMPTY - PISTOL         0.2407 1.020 16069   0.237  0.8984
 EMPTY - SHOTGUN       -0.4681 0.620 16069  -0.755  0.7619
 EMPTY - SINGLEBACK    -0.9331 0.704 16069  -1.325  0.7619
 EMPTY - WILDCAT        0.1547 2.210 16069   0.070  0.9701
 I_FORM - JUMBO         5.5102 6.280 16069   0.877  0.7619
 I_FORM - PISTOL        2.2139 1.230 16069   1.805  0.7249
 I_FORM - SHOTGUN       1.5051 0.925 16069   1.628  0.7249
 I_FORM - SINGLEBACK    1.0401 0.983 16069   1.058  0.7619
 I_FORM - WILDCAT       2.1279 2.310 16069   0.921  0.7619
 JUMBO - PISTOL        -3.2963 6.270 16069  -0.525  0.7619
 JUMBO - SHOTGUN       -4.0051 6.220 16069  -0.644  0.7619
 JUMBO - SINGLEBACK    -4.4701 6.230 16069  -0.717  0.7619
 JUMBO - WILDCAT       -3.3824 6.570 16069  -0.515  0.7619
 PISTOL - SHOTGUN      -0.7088 0.884 16069  -0.802  0.7619
 PISTOL - SINGLEBACK   -1.1738 0.945 16069  -1.242  0.7619
```

```
 PISTOL - WILDCAT      -0.0861 2.290 16069  -0.038  0.9701
 SHOTGUN - SINGLEBACK  -0.4650 0.494 16069  -0.942  0.7619
 SHOTGUN - WILDCAT      0.6228 2.150 16069   0.290  0.8984
 SINGLEBACK - WILDCAT   1.0878 2.170 16069   0.500  0.7619

coverage_group = Red Zone:
 contrast             estimate    SE    df t.ratio p.value
 EMPTY - I_FORM        -1.8111 2.070 16069  -0.874  0.9803
 EMPTY - JUMBO         -0.1778 4.140 16069  -0.043  0.9803
 EMPTY - PISTOL        -1.8730 2.320 16069  -0.806  0.9803
 EMPTY - SHOTGUN       -1.2553 1.400 16069  -0.895  0.9803
 EMPTY - SINGLEBACK    -1.6550 1.550 16069  -1.069  0.9803
 EMPTY - WILDCAT       -0.9778 3.070 16069  -0.318  0.9803
 I_FORM - JUMBO         1.6333 4.250 16069   0.385  0.9803
 I_FORM - PISTOL       -0.0619 2.500 16069  -0.025  0.9803
 I_FORM - SHOTGUN       0.5558 1.680 16069   0.331  0.9803
 I_FORM - SINGLEBACK    0.1561 1.800 16069   0.087  0.9803
 I_FORM - WILDCAT       0.8333 3.210 16069   0.260  0.9803
 JUMBO - PISTOL        -1.6952 4.370 16069  -0.387  0.9803
 JUMBO - SHOTGUN       -1.0776 3.960 16069  -0.272  0.9803
 JUMBO - SINGLEBACK    -1.4772 4.020 16069  -0.368  0.9803
 JUMBO - WILDCAT       -0.8000 4.820 16069  -0.166  0.9803
 PISTOL - SHOTGUN       0.6177 1.980 16069   0.312  0.9803
 PISTOL - SINGLEBACK    0.2180 2.090 16069   0.104  0.9803
 PISTOL - WILDCAT       0.8952 3.380 16069   0.265  0.9803
 SHOTGUN - SINGLEBACK  -0.3996 0.962 16069  -0.415  0.9803
 SHOTGUN - WILDCAT      0.2776 2.820 16069   0.098  0.9803
 SINGLEBACK - WILDCAT   0.6772 2.900 16069   0.234  0.9803

P value adjustment: fdr method for varying numbers of tests
```

```r
library(emmeans)

emm <- emmeans(model_interaction, ~ offenseFormation * coverage_group)

emm_df <- as.data.frame(emm)

ggplot(emm_df, aes(x = coverage_group, y = emmean,
                   color = offenseFormation, group = offenseFormation)) +
  geom_line(size = 1) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = emmean - SE, ymax = emmean + SE),
```

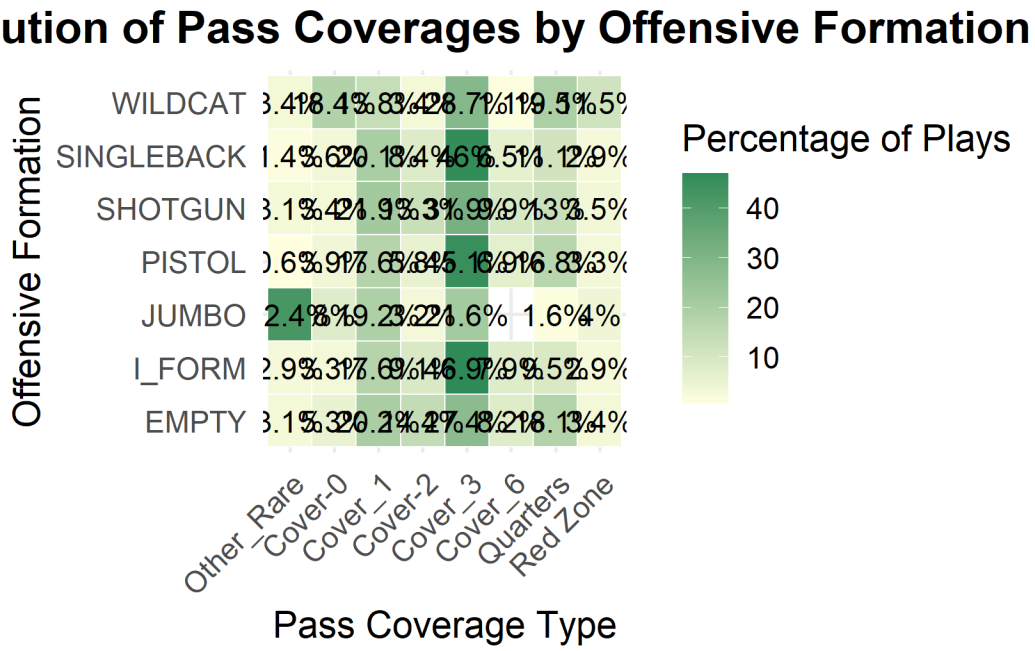
```
                width = 0.2, alpha = 0.4) +
  labs(
    title = "Interaction Between Formation and Coverage Type",
    x = "Defensive Coverage",
    y = "Estimated Yards Gained",
    color = "Offensive Formation"
  ) +
  theme_classic(base_size = 14) +
  theme(
    legend.position = "right",
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```



Match-up frequency

```
formation_coverage_counts <- clean_plays %>%
  filter(!is.na(offenseFormation), !is.na(coverage_group)) %>%
  count(offenseFormation, coverage_group) %>%
  arrange(desc(n))


formation_coverage_pct <- clean_plays %>%
```

```
  filter(!is.na(offenseFormation), !is.na(coverage_group)) %>%
  count(offenseFormation, coverage_group) %>%
  group_by(offenseFormation) %>%
  mutate(pct = n / sum(n) * 100) %>%
  ungroup()


ggplot(formation_coverage_pct, aes(x = coverage_group, y = offenseFormation, fill = pct)) +
  geom_tile(color = "white") +
  geom_text(aes(label = paste0(round(pct, 1), "%")), color = "black", size = 4) +
  scale_fill_gradient(low = "lightyellow", high = "seagreen4") +
  labs(
    title = "Distribution of Pass Coverages by Offensive Formation",
    x = "Pass Coverage Type",
    y = "Offensive Formation",
    fill = "Percentage of Plays"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(face = "bold", hjust = 0.5)
  )
```



Investigate the 4th quarter

```
clean_plays_quarter <- clean_plays %>%
  mutate(game_phase = case_when(
    quarter %in% 1:3 ~ "Early (Q1-Q3)",
    quarter %in% 4:5 ~ "Late / Overtime",
    TRUE ~ "Unknown"
  ))


model_full_4<- lm(yardsGained ~ offenseFormation * coverage_group+game_phase, data = clean_pl
summary(model_full_4)
```

```
Call:
lm(formula = yardsGained ~ offenseFormation * coverage_group +
    game_phase, data = clean_plays_quarter)

Residuals:
    Min      1Q  Median      3Q     Max
-74.134  -5.491  -1.875   2.621  91.409

Coefficients: (1 not defined because of singularities)
                                                Estimate Std. Error t value
(Intercept)                                       6.4666     1.3751   4.703
offenseFormationI_FORM                           -4.5913     2.1120  -2.174
offenseFormationJUMBO                            -5.9462     1.8285  -3.252
offenseFormationPISTOL                           -5.1198     4.6047  -1.112
offenseFormationSHOTGUN                           0.1241     1.4693   0.084
offenseFormationSINGLEBACK                       -4.0054     1.8209  -2.200
offenseFormationWILDCAT                          -5.6709     5.2575  -1.079
coverage_groupCover-0                            -1.9086     1.7243  -1.107
coverage_groupCover_1                             0.2018     1.4738   0.137
coverage_groupCover-2                            -0.8010     1.5120  -0.530
coverage_groupCover_3                             0.2226     1.4478   0.154
coverage_groupCover_6                            -0.6939     1.6088  -0.431
coverage_groupQuarters                           -1.3419     1.4848  -0.904
coverage_groupRed Zone                           -5.1583     1.8983  -2.717
game_phaseLate / Overtime                        -0.3872     0.1551  -2.496
offenseFormationI_FORM:coverage_groupCover-0      4.5637     2.7967   1.632
offenseFormationJUMBO:coverage_groupCover-0       3.6044     3.4868   1.034
offenseFormationPISTOL:coverage_groupCover-0      3.3787     5.0383   0.671
```

```
offenseFormationSHOTGUN:coverage_groupCover-0      -1.8372     1.8701  -0.982
offenseFormationSINGLEBACK:coverage_groupCover-0    2.2507     2.2262   1.011
offenseFormationWILDCAT:coverage_groupCover-0       5.9980     5.7930   1.035
offenseFormationI_FORM:coverage_groupCover_1        3.2313     2.2740   1.421
offenseFormationJUMBO:coverage_groupCover_1         3.6636     2.6177   1.400
offenseFormationPISTOL:coverage_groupCover_1        2.7892     4.7089   0.592
offenseFormationSHOTGUN:coverage_groupCover_1      -0.6583     1.5761  -0.418
offenseFormationSINGLEBACK:coverage_groupCover_1    2.7963     1.9235   1.454
offenseFormationWILDCAT:coverage_groupCover_1       5.7982     5.8624   0.989
offenseFormationI_FORM:coverage_groupCover-2        4.0908     2.3838   1.716
offenseFormationJUMBO:coverage_groupCover-2         0.2806     4.8021   0.058
offenseFormationPISTOL:coverage_groupCover-2        5.5772     4.8673   1.146
offenseFormationSHOTGUN:coverage_groupCover-2       0.5895     1.6199   0.364
offenseFormationSINGLEBACK:coverage_groupCover-2    4.2361     1.9876   2.131
offenseFormationWILDCAT:coverage_groupCover-2      14.0053     7.3347   1.909
offenseFormationI_FORM:coverage_groupCover_3        2.7774     2.1977   1.264
offenseFormationJUMBO:coverage_groupCover_3         4.4542     2.5328   1.759
offenseFormationPISTOL:coverage_groupCover_3        4.1842     4.6562   0.899
offenseFormationSHOTGUN:coverage_groupCover_3      -1.3077     1.5480  -0.845
offenseFormationSINGLEBACK:coverage_groupCover_3    2.9808     1.8890   1.578
offenseFormationWILDCAT:coverage_groupCover_3       2.8991     5.5625   0.521
offenseFormationI_FORM:coverage_groupCover_6        4.3020     2.4708   1.741
offenseFormationJUMBO:coverage_groupCover_6             NA         NA      NA
offenseFormationPISTOL:coverage_groupCover_6        4.8544     4.8642   0.998
offenseFormationSHOTGUN:coverage_groupCover_6       0.6012     1.7171   0.350
offenseFormationSINGLEBACK:coverage_groupCover_6    3.7698     2.0787   1.814
offenseFormationWILDCAT:coverage_groupCover_6       1.8981    10.2768   0.185
offenseFormationI_FORM:coverage_groupQuarters       6.5480     2.3594   2.775
offenseFormationJUMBO:coverage_groupQuarters        2.3215     6.5034   0.357
offenseFormationPISTOL:coverage_groupQuarters       4.8847     4.7156   1.036
offenseFormationSHOTGUN:coverage_groupQuarters      0.3688     1.5948   0.231
offenseFormationSINGLEBACK:coverage_groupQuarters   4.9096     1.9522   2.515
offenseFormationWILDCAT:coverage_groupQuarters      5.4513     5.7012   0.956
offenseFormationI_FORM:coverage_groupRed Zone       6.3938     2.9586   2.161
offenseFormationJUMBO:coverage_groupRed Zone        6.1154     4.5292   1.350
offenseFormationPISTOL:coverage_groupRed Zone       6.9990     5.1575   1.357
offenseFormationSHOTGUN:coverage_groupRed Zone      1.1842     2.0310   0.583
offenseFormationSINGLEBACK:coverage_groupRed Zone   5.6898     2.3898   2.381
offenseFormationWILDCAT:coverage_groupRed Zone      6.6013     6.0897   1.084
                                                  Pr(>|t|)
(Intercept)                                       2.59e-06 ***
offenseFormationI_FORM                             0.02973 *
offenseFormationJUMBO                              0.00115 **
```

```
offenseFormationPISTOL                                0.26621
offenseFormationSHOTGUN                               0.93267
offenseFormationSINGLEBACK                            0.02784 *
offenseFormationWILDCAT                               0.28077
coverage_groupCover-0                                 0.26834
coverage_groupCover_1                                 0.89111
coverage_groupCover-2                                 0.59626
coverage_groupCover_3                                 0.87782
coverage_groupCover_6                                 0.66627
coverage_groupQuarters                                0.36611
coverage_groupRed Zone                                0.00659 **
game_phaseLate / Overtime                             0.01257 *
offenseFormationI_FORM:coverage_groupCover-0          0.10273
offenseFormationJUMBO:coverage_groupCover-0           0.30128
offenseFormationPISTOL:coverage_groupCover-0          0.50249
offenseFormationSHOTGUN:coverage_groupCover-0         0.32590
offenseFormationSINGLEBACK:coverage_groupCover-0      0.31202
offenseFormationWILDCAT:coverage_groupCover-0         0.30050
offenseFormationI_FORM:coverage_groupCover_1          0.15534
offenseFormationJUMBO:coverage_groupCover_1           0.16166
offenseFormationPISTOL:coverage_groupCover_1          0.55365
offenseFormationSHOTGUN:coverage_groupCover_1         0.67620
offenseFormationSINGLEBACK:coverage_groupCover_1      0.14604
offenseFormationWILDCAT:coverage_groupCover_1         0.32265
offenseFormationI_FORM:coverage_groupCover-2          0.08617 .
offenseFormationJUMBO:coverage_groupCover-2           0.95340
offenseFormationPISTOL:coverage_groupCover-2          0.25187
offenseFormationSHOTGUN:coverage_groupCover-2         0.71596
offenseFormationSINGLEBACK:coverage_groupCover-2      0.03309 *
offenseFormationWILDCAT:coverage_groupCover-2         0.05622 .
offenseFormationI_FORM:coverage_groupCover_3          0.20633
offenseFormationJUMBO:coverage_groupCover_3           0.07866 .
offenseFormationPISTOL:coverage_groupCover_3          0.36887
offenseFormationSHOTGUN:coverage_groupCover_3         0.39825
offenseFormationSINGLEBACK:coverage_groupCover_3      0.11460
offenseFormationWILDCAT:coverage_groupCover_3         0.60224
offenseFormationI_FORM:coverage_groupCover_6          0.08168 .
offenseFormationJUMBO:coverage_groupCover_6                NA
offenseFormationPISTOL:coverage_groupCover_6          0.31830
offenseFormationSHOTGUN:coverage_groupCover_6         0.72624
offenseFormationSINGLEBACK:coverage_groupCover_6      0.06977 .
offenseFormationWILDCAT:coverage_groupCover_6         0.85347
offenseFormationI_FORM:coverage_groupQuarters         0.00552 **
```

```
offenseFormationJUMBO:coverage_groupQuarters       0.72112
offenseFormationPISTOL:coverage_groupQuarters      0.30029
offenseFormationSHOTGUN:coverage_groupQuarters     0.81711
offenseFormationSINGLEBACK:coverage_groupQuarters  0.01192 *
offenseFormationWILDCAT:coverage_groupQuarters     0.33900
offenseFormationI_FORM:coverage_groupRed Zone      0.03070 *
offenseFormationJUMBO:coverage_groupRed Zone       0.17697
offenseFormationPISTOL:coverage_groupRed Zone      0.17479
offenseFormationSHOTGUN:coverage_groupRed Zone     0.55987
offenseFormationSINGLEBACK:coverage_groupRed Zone  0.01728 *
offenseFormationWILDCAT:coverage_groupRed Zone     0.27838
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.79 on 16068 degrees of freedom
Multiple R-squared:  0.01306,   Adjusted R-squared:  0.009681
F-statistic: 3.866 on 55 and 16068 DF,  p-value: < 2.2e-16
```

```r
library(dplyr)
formation_phase_counts <- clean_plays_quarter %>%
  count(game_phase, offenseFormation) %>%
  group_by(game_phase) %>%
  mutate(pct = n / sum(n) * 100)




# Exclude JUMBO from both groups because it only appear in specific situation
clean_plays_1 <- subset(
  clean_plays_quarter,
  coverage_group %in% cover_group1 & offenseFormation != "JUMBO"
)

clean_plays_2 <- subset(
  clean_plays_quarter,
  coverage_group %in% cover_group2 & offenseFormation != "JUMBO"
)
```

```r
clean_plays_quarter %>%
  filter(game_phase == "Late / Overtime") %>%  # filter only Cover-2 plays
  group_by(game_phase,possessionTeam
           ) %>%
```

```
  summarise(
    mean_yards = mean(yardsGained, na.rm = TRUE),
    sd_yards   = sd(yardsGained, na.rm = TRUE),
    n          = n()
  ) %>%
  arrange(desc(mean_yards))
```

```
# A tibble: 32 x 5
# Groups:   game_phase [1]
   game_phase      possessionTeam mean_yards sd_yards     n
   <chr>           <chr>               <dbl>    <dbl> <int>
 1 Late / Overtime CAR                  7.22    13.8    134
 2 Late / Overtime NO                   7.12    10.7    163
 3 Late / Overtime MIA                  6.65    11.2    133
 4 Late / Overtime KC                   6.04    11.4    137
 5 Late / Overtime ATL                  5.96     9.73   135
 6 Late / Overtime NYJ                  5.81     9.01   160
 7 Late / Overtime CLE                  5.64     8.57   149
 8 Late / Overtime GB                   5.55     8.27   161
 9 Late / Overtime IND                  5.51     8.38   171
10 Late / Overtime BAL                  5.47     7.55   136
# i 22 more rows
```

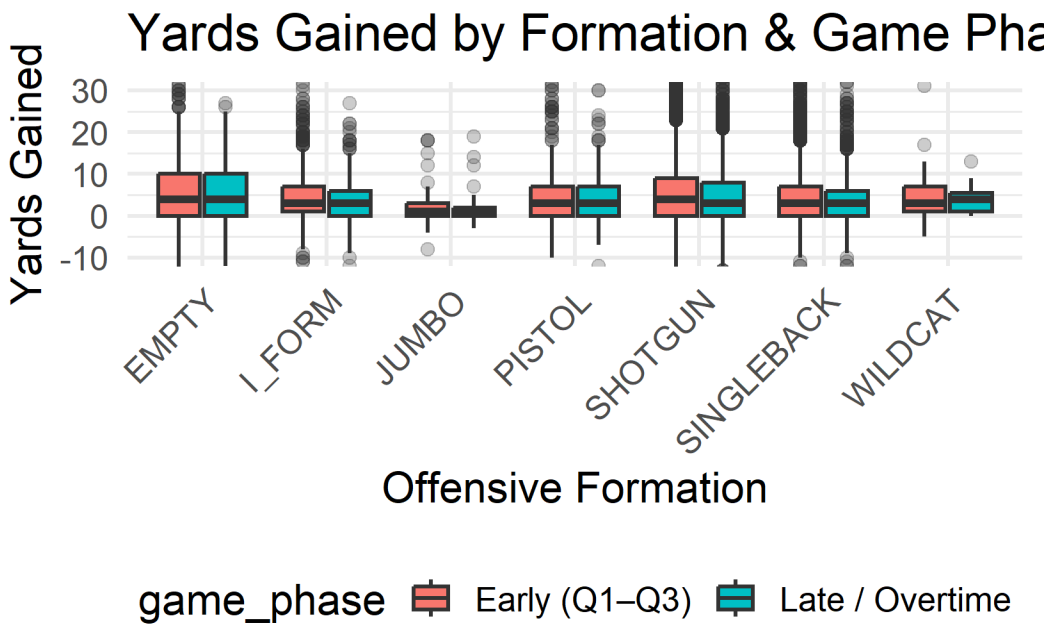
Something else with 3 way interaction

```
library(ggplot2)
library(dplyr)
library(patchwork)

# for multi-panel layout
clean_plays_3way <- clean_plays_quarter %>%
  filter(!is.na(offenseFormation), !is.na(coverage_group), !is.na(game_phase)) %>%
  mutate(
    game_phase = factor(game_phase, levels = c("Early (Q1-Q3)", "Late / Overtime")),
    offenseFormation = factor(offenseFormation),
    coverage_group = factor(coverage_group)
  )

#Plot 1
p1 <- ggplot(clean_plays_quarter, aes(x = offenseFormation, y = yardsGained, fill = game_phas
  geom_boxplot(outlier.alpha = 0.25) +
```

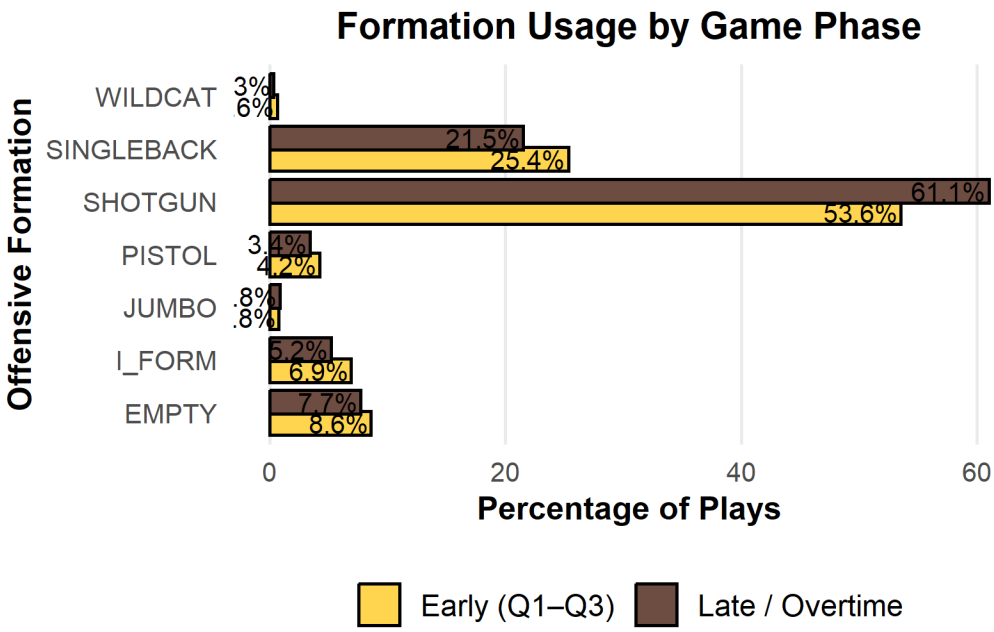
```
  coord_cartesian(ylim = c(-10, 30)) +
  labs(
    title = "Yards Gained by Formation & Game Phase",
    x = "Offensive Formation", y = "Yards Gained"
  ) +
  theme_minimal(base_size = 16) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  )

formation_phase_counts <- clean_plays_quarter %>%
  count(game_phase, offenseFormation) %>%
  group_by(game_phase) %>%
  mutate(pct = n / sum(n) * 100)

defense_phase_counts <- clean_plays_quarter %>%
  count(game_phase, coverage_group) %>%
  group_by(game_phase) %>%
  mutate(pct = n / sum(n) * 100)

p2 <- ggplot(formation_phase_counts,
             aes(y = offenseFormation, x = pct, fill = game_phase)) +
  geom_bar(stat = "identity",
           position = position_dodge(width = 0.8),
           color = "black") +
  geom_text(
    aes(label = paste0(round(pct, 1), "%")),
    position = position_dodge(width = 0.8),
    hjust = 1.05,          # move inside the bar
    color = "black",       # white text contrasts with fill
    size = 3.5
  ) +
  scale_fill_manual(values = c(
    "Early (Q1-Q3)" = "#FFD54F",
    "Late / Overtime" = "#6D4C41"
  )) +
  labs(
    title = "Formation Usage by Game Phase",
    x = "Percentage of Plays",
    y = "Offensive Formation"
  ) +
```

```
  theme_minimal(base_size = 14) +
  theme(
    axis.text.y = element_text(size = 10),
    axis.text.x = element_text(size = 10),
    axis.title = element_text(size = 12, face = "bold"),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank()
  )




#PCT for defense

p_defense <- ggplot(defense_phase_counts,
                    aes(y = coverage_group, x = pct, fill = game_phase)) +
  geom_bar(stat = "identity",
           position = position_dodge(width = 0.8),
           color = "black") +
  geom_text(
    aes(label = paste0(round(pct, 1), "%")),
    position = position_dodge(width = 0.8),
    hjust = -0.1,
    size = 3.5
  ) +
  scale_fill_manual(values = c(
    "Early (Q1-Q3)" = "#FFD54F",   # KSU gold
    "Late / Overtime" = "#6D4C41"  # dark bronze/brown contrast
  )) +
  labs(
    title = "Defensive Coverage Usage by Game Phase",
    x = "Percentage of Plays",
    y = "Coverage Type"
  ) +
  coord_cartesian(xlim = c(0, max(defense_phase_counts$pct) * 1.15)) + # padding for labels
  theme_minimal(base_size = 14) +
  theme(
    axis.text.y = element_text(size = 10),
    axis.text.x = element_text(size = 10),
```

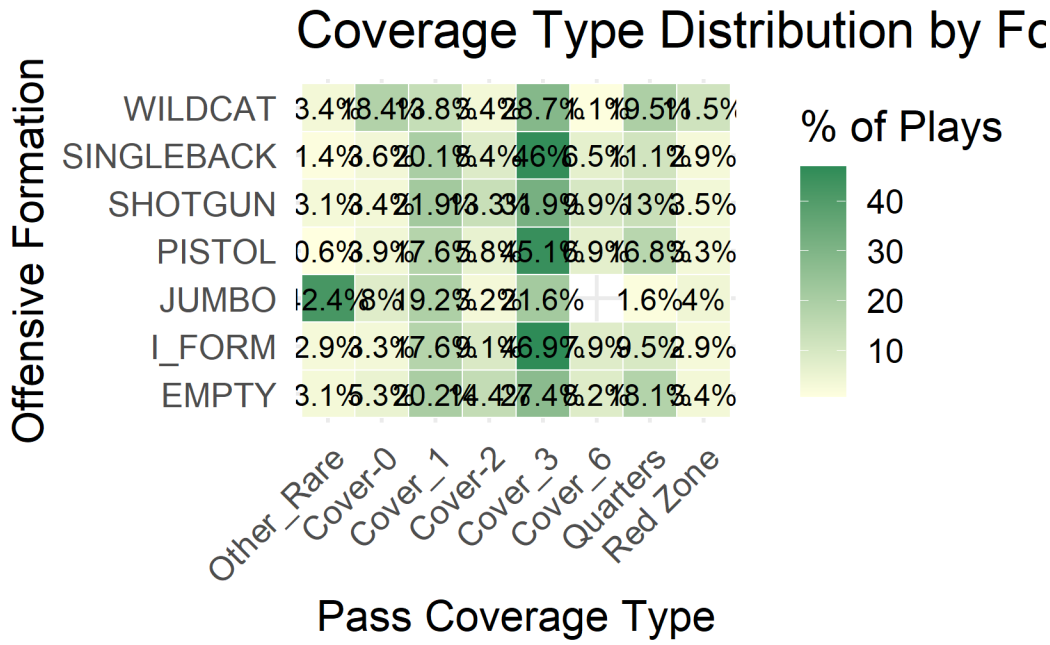
```
    axis.title = element_text(size = 12, face = "bold"),
    plot.title = element_text(size = 14, face = "bold", hjust = 0.5),
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank()
  )

formation_coverage_pct <- clean_plays %>%
  count(offenseFormation, coverage_group) %>%
  group_by(offenseFormation) %>%
  mutate(pct = n / sum(n) * 100)

p3 <- ggplot(formation_coverage_pct, aes(x = coverage_group, y = offenseFormation, fill = pct
  geom_tile(color = "white") +
  geom_text(aes(label = paste0(round(pct, 1), "%")), size = 4) +
  scale_fill_gradient(low = "lightyellow", high = "seagreen4") +
  labs(
    title = "Coverage Type Distribution by Formation",
    x = "Pass Coverage Type", y = "Offensive Formation", fill = "% of Plays"
  ) +
  theme_minimal(base_size = 16) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

p1
```


Yards Gained by Formation & Game Pha
Yards Gained
30
20
10
0
-10
EMPTY
I_FORM
JUMBO
PISTOL
SHOTGUN
SINGLEBACK
WILDCAT
Offensive Formation
game_phase
Early (Q1–Q3)
Late / Overtime

```
p2
```


Formation Usage by Game Phase
Offensive Formation
WILDCAT
SINGLEBACK
SHOTGUN
PISTOL
JUMBO
I_FORM
EMPTY
3%
6%
21.5%
25.4%
61.1%
53.6%
3.4%
4.2%
.8%
.8%
5.2%
6.9%
7.7%
8.6%
0
20
40
60
Percentage of Plays
Early (Q1–Q3)
Late / Overtime

```
p3
```



INVESTIGATE JUMBO AND WILDCAT

```
library(dplyr)

clean_plays_quarter %>%
  filter(offenseFormation %in% c("JUMBO", "WILDCAT")) %>%
  summarise(
    avg_yardsToGo = mean(yardsToGo, na.rm = TRUE),
    median_yardsToGo = median(yardsToGo, na.rm = TRUE),
    avg_yardline = mean(absoluteYardlineNumber, na.rm = TRUE),
    median_yardline = median(absoluteYardlineNumber, na.rm = TRUE),
    mean_yardsGained = mean(yardsGained, na.rm = TRUE),
    n = n()
  )
```

```
  avg_yardsToGo median_yardsToGo avg_yardline median_yardline mean_yardsGained
1      5.051887              3.5     58.02358              54         3.278302
    n
1 212
```

MOVE IT UP A LEVEL

```
library(dplyr)

clean_plays_quarter_formation <- clean_plays_quarter %>%
  mutate(formation_group = case_when(
    offenseFormation %in% c("SHOTGUN", "SINGLEBACK", "PISTOL") ~ "Standard",
    offenseFormation %in% c("EMPTY", "I_FORM") ~ "Contextual",
    offenseFormation %in% c("JUMBO", "WILDCAT") ~ "Situational",
    TRUE ~ "Other"   # catch any undefined or rare formations
  ))
table(clean_plays_quarter_formation$formation_group)
```

```
 Contextual Situational    Standard
       2377         212       13535
```

```
#
clean_plays_quarter_formation %>%
  group_by(formation_group, offenseFormation) %>%
  summarise(
    mean_yards = mean(yardsGained, na.rm = TRUE),
    sd_yards = sd(yardsGained, na.rm = TRUE),
    n = n()
  ) %>%
  arrange(formation_group, desc(mean_yards))
```

```
# A tibble: 7 x 5
# Groups:   formation_group [3]
  formation_group offenseFormation mean_yards sd_yards     n
  <chr>           <chr>                 <dbl>    <dbl> <int>
1 Contextual      EMPTY                  5.78     9.33  1342
2 Contextual      I_FORM                 5.00     7.80  1035
3 Situational     WILDCAT                4.64     7.46    87
4 Situational     JUMBO                  2.33     4.24   125
5 Standard        SHOTGUN                5.59     9.01  8979
6 Standard        SINGLEBACK             5.35     8.78  3915
7 Standard        PISTOL                 5.04     7.80   641
```

```r
model_phase <- aov(yardsGained ~ formation_group * game_phase*coverage_group,
                   data = clean_plays_quarter_formation)
summary(model_phase)
```

```
                                            Df  Sum Sq Mean Sq F value  Pr(>F)
formation_group                              2    1030   515.2   6.661 0.00128
game_phase                                   1     714   713.6   9.226 0.00239
coverage_group                               7    9334  1333.4  17.238 < 2e-
16
formation_group:game_phase                   2      26    13.0   0.168 0.84570
formation_group:coverage_group              14    1429   102.1   1.320 0.18595
game_phase:coverage_group                    7    1220   174.2   2.253 0.02738
formation_group:game_phase:coverage_group   12     604    50.4   0.651 0.79938
Residuals                                16078 1243617    77.3

formation_group                             **
game_phase                                  **
coverage_group                              ***
formation_group:game_phase
formation_group:coverage_group
game_phase:coverage_group                   *
formation_group:game_phase:coverage_group
Residuals
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
lm_model <- lm(
  yardsGained ~ offenseFormation*coverage_group+game_phase,
  data = clean_plays_quarter_formation
)
summary(lm_model)
```

```
Call:
lm(formula = yardsGained ~ offenseFormation * coverage_group +
    game_phase, data = clean_plays_quarter_formation)

Residuals:
    Min      1Q  Median      3Q     Max
-74.134  -5.491  -1.875   2.621  91.409
```

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value
(Intercept)	6.4666	1.3751	4.703
offenseFormationI_FORM	-4.5913	2.1120	-2.174
offenseFormationJUMBO	-5.9462	1.8285	-3.252
offenseFormationPISTOL	-5.1198	4.6047	-1.112
offenseFormationSHOTGUN	0.1241	1.4693	0.084
offenseFormationSINGLEBACK	-4.0054	1.8209	-2.200
offenseFormationWILDCAT	-5.6709	5.2575	-1.079
coverage_groupCover-0	-1.9086	1.7243	-1.107
coverage_groupCover_1	0.2018	1.4738	0.137
coverage_groupCover-2	-0.8010	1.5120	-0.530
coverage_groupCover_3	0.2226	1.4478	0.154
coverage_groupCover_6	-0.6939	1.6088	-0.431
coverage_groupQuarters	-1.3419	1.4848	-0.904
coverage_groupRed Zone	-5.1583	1.8983	-2.717
game_phaseLate / Overtime	-0.3872	0.1551	-2.496
offenseFormationI_FORM:coverage_groupCover-0	4.5637	2.7967	1.632
offenseFormationJUMBO:coverage_groupCover-0	3.6044	3.4868	1.034
offenseFormationPISTOL:coverage_groupCover-0	3.3787	5.0383	0.671
offenseFormationSHOTGUN:coverage_groupCover-0	-1.8372	1.8701	-0.982
offenseFormationSINGLEBACK:coverage_groupCover-0	2.2507	2.2262	1.011
offenseFormationWILDCAT:coverage_groupCover-0	5.9980	5.7930	1.035
offenseFormationI_FORM:coverage_groupCover_1	3.2313	2.2740	1.421
offenseFormationJUMBO:coverage_groupCover_1	3.6636	2.6177	1.400
offenseFormationPISTOL:coverage_groupCover_1	2.7892	4.7089	0.592
offenseFormationSHOTGUN:coverage_groupCover_1	-0.6583	1.5761	-0.418
offenseFormationSINGLEBACK:coverage_groupCover_1	2.7963	1.9235	1.454
offenseFormationWILDCAT:coverage_groupCover_1	5.7982	5.8624	0.989
offenseFormationI_FORM:coverage_groupCover-2	4.0908	2.3838	1.716
offenseFormationJUMBO:coverage_groupCover-2	0.2806	4.8021	0.058
offenseFormationPISTOL:coverage_groupCover-2	5.5772	4.8673	1.146
offenseFormationSHOTGUN:coverage_groupCover-2	0.5895	1.6199	0.364
offenseFormationSINGLEBACK:coverage_groupCover-2	4.2361	1.9876	2.131
offenseFormationWILDCAT:coverage_groupCover-2	14.0053	7.3347	1.909
offenseFormationI_FORM:coverage_groupCover_3	2.7774	2.1977	1.264
offenseFormationJUMBO:coverage_groupCover_3	4.4542	2.5328	1.759
offenseFormationPISTOL:coverage_groupCover_3	4.1842	4.6562	0.899
offenseFormationSHOTGUN:coverage_groupCover_3	-1.3077	1.5480	-0.845
offenseFormationSINGLEBACK:coverage_groupCover_3	2.9808	1.8890	1.578
offenseFormationWILDCAT:coverage_groupCover_3	2.8991	5.5625	0.521
offenseFormationI_FORM:coverage_groupCover_6	4.3020	2.4708	1.741
offenseFormationJUMBO:coverage_groupCover_6	NA	NA	NA

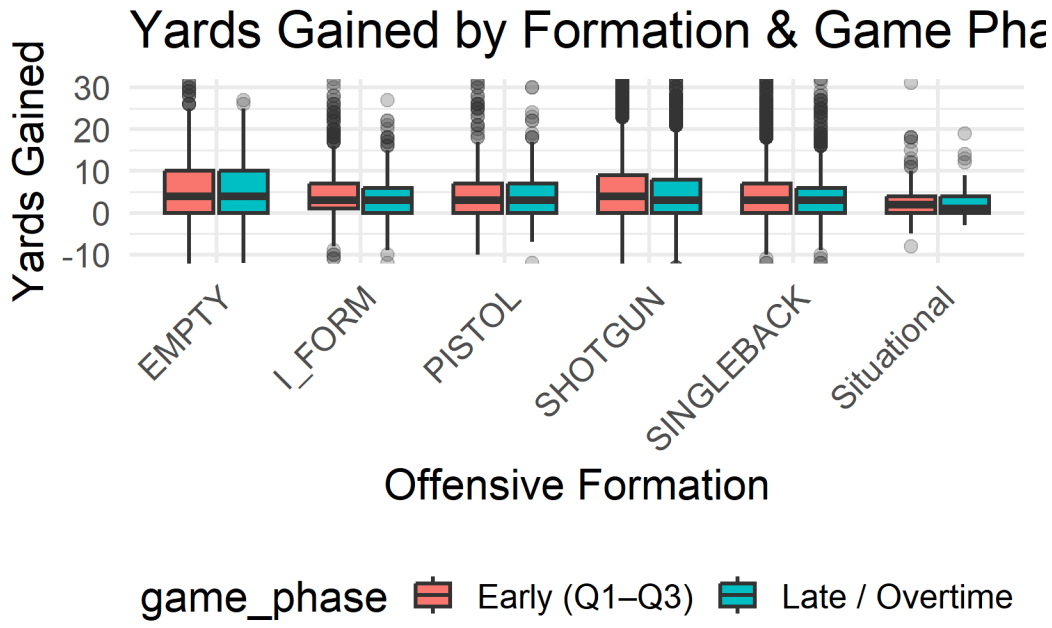
```
offenseFormationPISTOL:coverage_groupCover_6        4.8544     4.8642   0.998
offenseFormationSHOTGUN:coverage_groupCover_6       0.6012     1.7171   0.350
offenseFormationSINGLEBACK:coverage_groupCover_6    3.7698     2.0787   1.814
offenseFormationWILDCAT:coverage_groupCover_6       1.8981    10.2768   0.185
offenseFormationI_FORM:coverage_groupQuarters       6.5480     2.3594   2.775
offenseFormationJUMBO:coverage_groupQuarters        2.3215     6.5034   0.357
offenseFormationPISTOL:coverage_groupQuarters       4.8847     4.7156   1.036
offenseFormationSHOTGUN:coverage_groupQuarters      0.3688     1.5948   0.231
offenseFormationSINGLEBACK:coverage_groupQuarters   4.9096     1.9522   2.515
offenseFormationWILDCAT:coverage_groupQuarters      5.4513     5.7012   0.956
offenseFormationI_FORM:coverage_groupRed Zone       6.3938     2.9586   2.161
offenseFormationJUMBO:coverage_groupRed Zone        6.1154     4.5292   1.350
offenseFormationPISTOL:coverage_groupRed Zone       6.9990     5.1575   1.357
offenseFormationSHOTGUN:coverage_groupRed Zone      1.1842     2.0310   0.583
offenseFormationSINGLEBACK:coverage_groupRed Zone   5.6898     2.3898   2.381
offenseFormationWILDCAT:coverage_groupRed Zone      6.6013     6.0897   1.084
                                                  Pr(>|t|)
(Intercept)                                       2.59e-06 ***
offenseFormationI_FORM                             0.02973 *
offenseFormationJUMBO                              0.00115 **
offenseFormationPISTOL                             0.26621
offenseFormationSHOTGUN                            0.93267
offenseFormationSINGLEBACK                         0.02784 *
offenseFormationWILDCAT                            0.28077
coverage_groupCover-0                              0.26834
coverage_groupCover_1                              0.89111
coverage_groupCover-2                              0.59626
coverage_groupCover_3                              0.87782
coverage_groupCover_6                              0.66627
coverage_groupQuarters                             0.36611
coverage_groupRed Zone                             0.00659 **
game_phaseLate / Overtime                          0.01257 *
offenseFormationI_FORM:coverage_groupCover-0       0.10273
offenseFormationJUMBO:coverage_groupCover-0        0.30128
offenseFormationPISTOL:coverage_groupCover-0       0.50249
offenseFormationSHOTGUN:coverage_groupCover-0      0.32590
offenseFormationSINGLEBACK:coverage_groupCover-0   0.31202
offenseFormationWILDCAT:coverage_groupCover-0      0.30050
offenseFormationI_FORM:coverage_groupCover_1       0.15534
offenseFormationJUMBO:coverage_groupCover_1        0.16166
offenseFormationPISTOL:coverage_groupCover_1       0.55365
offenseFormationSHOTGUN:coverage_groupCover_1      0.67620
offenseFormationSINGLEBACK:coverage_groupCover_1   0.14604
```

```
offenseFormationWILDCAT:coverage_groupCover_1      0.32265
offenseFormationI_FORM:coverage_groupCover-2       0.08617 .
offenseFormationJUMBO:coverage_groupCover-2        0.95340
offenseFormationPISTOL:coverage_groupCover-2       0.25187
offenseFormationSHOTGUN:coverage_groupCover-2      0.71596
offenseFormationSINGLEBACK:coverage_groupCover-2   0.03309 *
offenseFormationWILDCAT:coverage_groupCover-2      0.05622 .
offenseFormationI_FORM:coverage_groupCover_3       0.20633
offenseFormationJUMBO:coverage_groupCover_3        0.07866 .
offenseFormationPISTOL:coverage_groupCover_3       0.36887
offenseFormationSHOTGUN:coverage_groupCover_3      0.39825
offenseFormationSINGLEBACK:coverage_groupCover_3   0.11460
offenseFormationWILDCAT:coverage_groupCover_3      0.60224
offenseFormationI_FORM:coverage_groupCover_6       0.08168 .
offenseFormationJUMBO:coverage_groupCover_6             NA
offenseFormationPISTOL:coverage_groupCover_6       0.31830
offenseFormationSHOTGUN:coverage_groupCover_6      0.72624
offenseFormationSINGLEBACK:coverage_groupCover_6   0.06977 .
offenseFormationWILDCAT:coverage_groupCover_6      0.85347
offenseFormationI_FORM:coverage_groupQuarters      0.00552 **
offenseFormationJUMBO:coverage_groupQuarters       0.72112
offenseFormationPISTOL:coverage_groupQuarters      0.30029
offenseFormationSHOTGUN:coverage_groupQuarters     0.81711
offenseFormationSINGLEBACK:coverage_groupQuarters  0.01192 *
offenseFormationWILDCAT:coverage_groupQuarters     0.33900
offenseFormationI_FORM:coverage_groupRed Zone      0.03070 *
offenseFormationJUMBO:coverage_groupRed Zone       0.17697
offenseFormationPISTOL:coverage_groupRed Zone      0.17479
offenseFormationSHOTGUN:coverage_groupRed Zone     0.55987
offenseFormationSINGLEBACK:coverage_groupRed Zone  0.01728 *
offenseFormationWILDCAT:coverage_groupRed Zone     0.27838
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.79 on 16068 degrees of freedom
Multiple R-squared:  0.01306,   Adjusted R-squared:  0.009681
F-statistic: 3.866 on 55 and 16068 DF,  p-value: < 2.2e-16
```

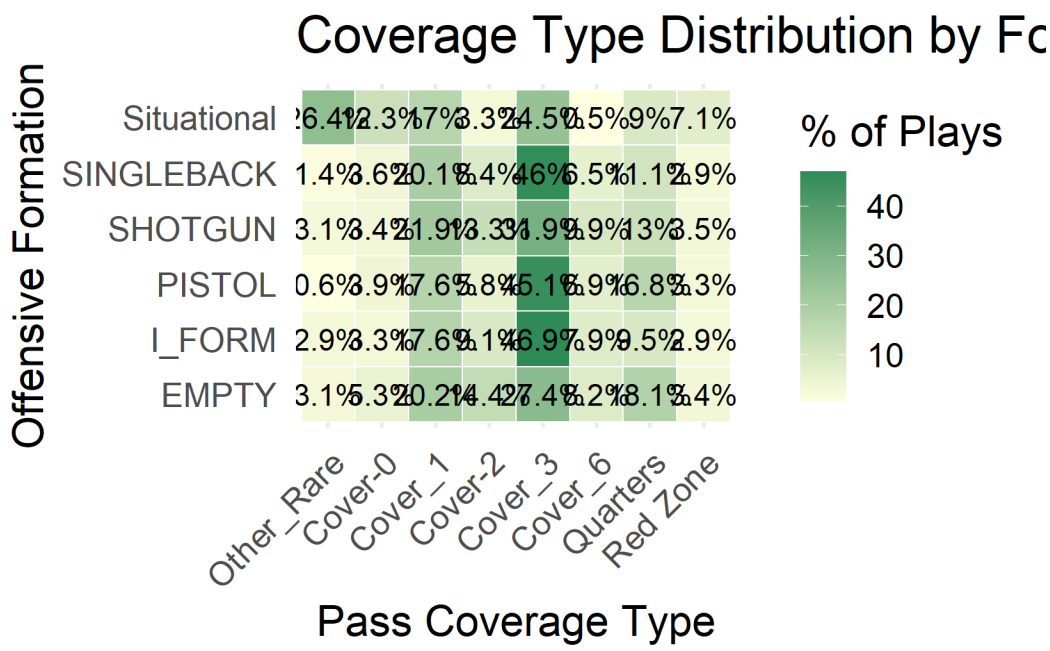
Group WILDCAT AND JUMBO TOGETHER

```r
clean_plays_quarter_formation <- clean_plays_quarter_formation %>%
  mutate(formation_refined = case_when(
```

```
    offenseFormation %in% c("JUMBO", "WILDCAT") ~ "Situational",
    TRUE ~ offenseFormation
  ))

#Now draw some plot


#Plot 1
ggplot(clean_plays_quarter_formation, aes(x = formation_refined, y = yardsGained, fill = game
  geom_boxplot(outlier.alpha = 0.25) +
  coord_cartesian(ylim = c(-10, 30)) +
  labs(
    title = "Yards Gained by Formation & Game Phase",
    x = "Offensive Formation", y = "Yards Gained"
  ) +
  theme_minimal(base_size = 16) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  )
```



```r
formation_phase_counts <- clean_plays_quarter_formation %>%
  count(game_phase, formation_refined) %>%
  group_by(game_phase) %>%
  mutate(pct = n / sum(n) * 100)

#Plot 2
 p1<-ggplot(formation_phase_counts, aes(x = formation_refined, y = pct, fill = game_phase)) +
  geom_bar(stat = "identity", position = "dodge", color = "black") +
   geom_text(
    aes(label = paste0(round(pct, 1), "%")),     # show percentages with one decimal
    position = position_dodge(width = 0.9),      # align text with each bar group
    vjust = -0.4,                                # vertical adjustment above bar
    size = 3.5                                   # text size
  ) +
  labs(
    title = "Formation Usage by Game Phase",
    x = "Offensive Formation", y = "Percentage of Plays"
  ) +
  theme_minimal(base_size = 16) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  )


formation_coverage_pct <- clean_plays_quarter_formation %>%
  count(formation_refined, coverage_group) %>%
  group_by(formation_refined) %>%
  mutate(pct = n / sum(n) * 100)

 ggplot(formation_coverage_pct, aes(x = coverage_group, y = formation_refined, fill = pct)) +
  geom_tile(color = "white") +
  geom_text(aes(label = paste0(round(pct, 1), "%")), size = 4) +
  scale_fill_gradient(low = "lightyellow", high = "seagreen4") +
  labs(
    title = "Coverage Type Distribution by Formation",
    x = "Pass Coverage Type", y = "Offensive Formation", fill = "% of Plays"
  ) +
  theme_minimal(base_size = 16) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```
clean_plays_f1 <- subset(
  clean_plays_quarter_formation,
  coverage_group %in% cover_group1)

clean_plays_2 <- subset(
  clean_plays_quarter,
  coverage_group %in% cover_group2)
```

```
library(dplyr)

cover2_summary <- clean_plays_quarter_formation %>%
  filter(coverage_group == "Cover-2") %>%
  group_by(offenseFormation) %>%
  summarise(
    n = n(),
    mean_yards = mean(yardsGained, na.rm = TRUE),
    median_yards = median(yardsGained, na.rm = TRUE),
    sd_yards = sd(yardsGained, na.rm = TRUE)
  ) %>%
  arrange(desc(mean_yards))

cover2_summary
```

```
# A tibble: 7 x 5
  offenseFormation     n mean_yards median_yards sd_yards
  <chr>            <int>      <dbl>        <dbl>    <dbl>
1 WILDCAT              3      14             7      23.3
2 SHOTGUN           1192       6.26          5       8.99
3 PISTOL              37       6.08          5       6.61
4 SINGLEBACK         329       5.81          4       8.69
5 EMPTY              193       5.55          5       7.20
6 I_FORM              94       5.07          4       6.01
7 JUMBO                4       0             0.5     6.27
```

```r
library(ggsci)
p_box <- ggplot(clean_plays_quarter_formation,
                aes(x = offenseFormation, y = yardsGained, fill = coverage_group)) +
  geom_boxplot(outlier.alpha = 0.2, width = 0.7, color = "gray30") +
  geom_hline(yintercept = 0, linetype = "dashed", color = "gray40") +
  stat_summary(fun = median, geom = "text",
               aes(label = round(..y.., 1)), size = 3, vjust = -0.6, color = "black") +
  facet_wrap(~ game_phase, ncol = 2) +
  scale_fill_npg(name = "Coverage Type") +
  coord_cartesian(ylim = c(-15, 35)) +
  labs(
    title = "Distribution of Yards Gained by Formation and Coverage",
    subtitle = "Faceted by Game Phase (Early vs Late)",
    x = "Offensive Formation",
    y = "Yards Gained"
  ) +
  theme_minimal(base_size = 18) +
  theme(
    panel.grid.major.x = element_blank(),
    strip.text = element_text(face = "bold", size = 16),
    axis.text.x = element_text(angle = 45, hjust = 1, size = 13),
    plot.title = element_text(size = 22, face = "bold", hjust = 0.5),
    legend.position = "bottom"
  )


p_interact <- ggplot(clean_plays_quarter_formation,
                     aes(x = coverage_group, y = yardsGained,
                         color = offenseFormation, group = offenseFormation)) +
  stat_summary(fun = mean, geom = "line", linewidth = 1.2) +
  stat_summary(fun = mean, geom = "point", size = 3) +
```

```
  facet_wrap(~ game_phase, ncol = 2) +
  scale_color_manual(values = c(
  "#E6B800",  # rich gold
  "#FF8F00",  # dark amber
  "#D84315",  # burnt orange
  "#9C27B0",  # royal violet
  "#8D6E63",  # warm taupe
  "#6D4C41",  # dark brownish-gray
  "#455A64"   # cool gray-blue
  )) +
  labs(
    title = "Interaction Between Formation and Coverage",
    subtitle = "Lines represent mean yards gained across game phases",
    x = "Coverage Type", y = "Mean Yards Gained", color = "Formation"
  ) +
  theme_minimal(base_size = 18) +
  theme(
    legend.position = "bottom",
    plot.title = element_text(face = "bold", hjust = 0.5),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )


emm <- emmeans(lm_model, ~ offenseFormation * coverage_group)
emm_df <- as.data.frame(emm)

ggplot(emm_df, aes(coverage_group, emmean, color = offenseFormation, group = offenseFormatio
  geom_line(linewidth = 1.3) +
  geom_point(size = 3) +
  geom_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.1)
```


```
p_pred <- ggplot(emm_df,
                 aes(x = coverage_group, y = emmean,
                     color = offenseFormation, group = offenseFormation)) +
  geom_line(linewidth = 1.1) +
  geom_point(size = 2.2) +
  geom_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.15, alpha = 0.6) +
  labs(title = "Model-Predicted Interaction",
       subtitle = "Adjusted means with 95% CI (from LM)",
       x = "Coverage Type", y = "Estimated Yards Gained",
       color = "Formation") +
  scale_color_manual(values = c(
    "#E6B800", "#FF8F00", "#D84315", "#9C27B0",
    "#8D6E63", "#6D4C41", "#455A64"
  )) +
  theme_minimal(base_size = 16) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, hjust = 1))


p_raw <- ggplot(clean_plays_quarter_formation,
                aes(x = coverage_group, y = yardsGained,
                    color = offenseFormation, group = offenseFormation)) +
  stat_summary(fun = mean, geom = "line", linewidth = 1.1) +
```

```
  stat_summary(fun = mean, geom = "point", size = 2) +
  labs(title = "Observed Interaction",
       subtitle = "Raw mean yards gained by formation and coverage",
       x = "Coverage Type", y = "Mean Yards Gained") +
  scale_color_manual(values = c(
    "#E6B800", "#FF8F00", "#D84315", "#9C27B0",
    "#8D6E63", "#6D4C41", "#455A64"
  )) +
  theme_minimal(base_size = 16) +
  theme(legend.position = "bottom",
        axis.text.x = element_text(angle = 45, hjust = 1))
```

Summarize the model for poster

```
library(broom)
library(dplyr)

# main model
lm_model <- lm(yardsGained ~ offenseFormation * coverage_group + game_phase,
               data = clean_plays_quarter_formation)

# tidy summary table
model_summary <- broom::tidy(lm_model) %>%
  mutate(
    term = gsub("offenseFormation", "Formation: ", term),
    term = gsub("coverage_group", "Coverage: ", term),
    term = gsub("game_phase", "Phase: ", term),
    significance = case_when(
      p.value < 0.001 ~ "***",
      p.value < 0.01  ~ "**",
      p.value < 0.05  ~ "*",
      p.value < 0.1   ~ ".",
      TRUE            ~ ""
    )
  ) %>%
  select(Term = term,
         Estimate = estimate,
         `Std. Error` = std.error,
         `t value` = statistic,
         `p value` = p.value,
         significance)
```

```r
# view top 10 rows
head(model_summary, 10)
```

```
# A tibble: 10 x 6
   Term                  Estimate `Std. Error` `t value`  `p value` significance
   <chr>                    <dbl>        <dbl>     <dbl>      <dbl> <chr>
 1 (Intercept)              6.47          1.38    4.70   0.00000259 "***"
 2 Formation: I_FORM       -4.59          2.11   -2.17   0.0297     "*"
 3 Formation: JUMBO        -5.95          1.83   -3.25   0.00115    "**"
 4 Formation: PISTOL       -5.12          4.60   -1.11   0.266      ""
 5 Formation: SHOTGUN       0.124         1.47    0.0845 0.933      ""
 6 Formation: SINGLEBACK   -4.01          1.82   -2.20   0.0278     "*"
 7 Formation: WILDCAT      -5.67          5.26   -1.08   0.281      ""
 8 Coverage: Cover-0       -1.91          1.72   -1.11   0.268      ""
 9 Coverage: Cover_1        0.202         1.47    0.137  0.891      ""
10 Coverage: Cover-2       -0.801         1.51   -0.530  0.596      ""
```

```r
sig_effects <- model_summary %>%
  filter(`p value` < 0.05)
```

```r
library(dplyr)
library(knitr)
library(lmtest)
library(sandwich)

# Recompute OLS and HC3 safely inside this chunk
ols <- summary(model_interaction)$coefficients
rob <- coeftest(model_interaction, vcov = vcovHC(model_interaction, type = "HC3"))

# Build unified table
combined <- data.frame(
  Term = rownames(ols),
  Estimate = ols[, 1],
  Std.Error.OLS = ols[, 2],
  p.value.OLS = ols[, 4],
  Std.Error.HC3 = rob[, 2],
  p.value.HC3 = rob[, 4]
)

# Keep only predictors significant under HC3
sig_table <- combined %>%
```

```r
  filter(p.value.HC3 < 0.05) %>%
  mutate(
    significance = case_when(
      p.value.HC3 < 0.001 ~ "***",
      p.value.HC3 < 0.01  ~ "**",
      p.value.HC3 < 0.05  ~ "*"
    ),
    Estimate = round(Estimate, 3),
    Std.Error.OLS = round(Std.Error.OLS, 3),
    p.value.OLS = round(p.value.OLS, 4),
    Std.Error.HC3 = round(Std.Error.HC3, 3),
    p.value.HC3 = round(p.value.HC3, 4)
  )

# Final PDF-safe table
kable(
  sig_table,
  caption = "Significant Effects Before and After HC3 Robust Correction",
  booktabs = TRUE,
  align = "c"
)
```

Table 4: Significant Effects Before and After HC3 Robust Correction

	Term	Estimate	Std.Error.OLS	p.value.OLS	Std.Error.HC3	p.value.HC3	significance
(Intercept)	(Intercept)	6.268	1.373	0.0000	1.619	0.0001	***
offenseFormationI_FORM	offenseFormationI_FORM	-4.535	2.112	0.0318	1.820	0.0127	*
offenseFormationJUMBO	offenseFormationJUMBO	-5.872	1.829	0.0013	1.629	0.0003	***
offenseFormationPISTOL	offenseFormationPISTOL	-5.018	4.605	0.2759	1.953	0.0102	*
offenseFormationSINGLEBACK	offenseFormationSINGLEBACK	-3.972	1.821	0.0292	1.775	0.0253	*
offenseFormationWILDCAT	offenseFormationWILDCAT	-5.602	5.258	0.2868	1.669	0.0008	***
coverage_groupRed Zone	coverage_groupRed Zone	-5.046	1.898	0.0079	2.008	0.0120	*

	Term	Estimate	Std.Error.OLS	p.value.OLS	Std.Error.HC3	p.value.HC3	signif-icance
offenseFormationI_FORM:coverage_groupCover-2	offenseFormationI_FORM:coverage_groupCover-2	4.060	2.384	0.0886	1.993	0.0416	*
offenseFormationPISTOL:coverage_groupCover-2	offenseFormationPISTOL:coverage_groupCover-2	5.550	4.868	0.2543	2.302	0.0159	*
offenseFormationSINGLEBACK:coverage_groupCover-2	offenseFormationSINGLEBACK:coverage_groupCover-2	4.234	1.988	0.0332	1.911	0.0267	*
offenseFormationJUMBO:coverage_groupCover_3	offenseFormationJUMBO:coverage_groupCover_3	4.391	2.533	0.0830	2.032	0.0307	*
offenseFormationI_FORM:coverage_groupCover_6	offenseFormationI_FORM:coverage_groupCover_6	4.298	2.471	0.0820	2.089	0.0397	*
offenseFormationPISTOL:coverage_groupCover_6	offenseFormationPISTOL:coverage_groupCover_6	4.809	4.865	0.3229	2.354	0.0411	*
offenseFormationI_FORM:coverage_groupQuarters	offenseFormationI_FORM:coverage_groupQuarters	6.508	2.360	0.0058	2.194	0.0030	**
offenseFormationPISTOL:coverage_groupQuarters	offenseFormationPISTOL:coverage_groupQuarters	4.778	4.716	0.3111	2.103	0.0231	*
offenseFormationSINGLEBACK:coverage_groupQuarters	offenseFormationSINGLEBACK:coverage_groupQuarters	4.905	1.953	0.0120	1.891	0.0095	**
offenseFormationWILDCAT:coverage_groupQuarters	offenseFormationWILDCAT:coverage_groupQuarters	5.447	5.702	0.3395	2.078	0.0088	**
offenseFormationI_FORM:coverage_groupRed Zone	offenseFormationI_FORM:coverage_groupRed Zone	6.346	2.959	0.0320	2.317	0.0062	**
offenseFormationJUMBO:coverage_groupRed Zone	offenseFormationJUMBO:coverage_groupRed Zone	6.050	4.530	0.1817	2.268	0.0076	**
offenseFormationPISTOL:coverage_groupRed Zone	offenseFormationPISTOL:coverage_groupRed Zone	6.891	5.158	0.1816	2.457	0.0050	**

	Term	Estimate	Std.Error.OLS	p.value.OLS	Std.Error.HC3	p.value.HC3	signif-icance
offenseFormationSINGLEBACK:coverage_groupRed Zone	offenseFormationSINGLEBACK:coverage_groupRed Zone	5.627	2.390	0.0186	2.157	0.0091	**
offenseFormationWILDCAT:coverage_groupRed Zone	offenseFormationWILDCAT:coverage_groupRed Zone	6.579	6.091	0.2801	2.232	0.0032	**

Create table to see the raw data

```
clean_plays_quarter %>%
  group_by(offenseFormation, coverage_group) %>%
  summarise(
    mean_yards = mean(yardsGained, na.rm = TRUE),
    sd_yards   = sd(yardsGained, na.rm = TRUE),
    n          = n(),
    .groups = "drop"
  ) %>%
  arrange(desc(mean_yards))
```

```
# A tibble: 55 x 5
   offenseFormation coverage_group mean_yards sd_yards     n
   <chr>            <fct>                <dbl>    <dbl> <int>
 1 WILDCAT          Cover-2              14       23.3      3
 2 I_FORM           Quarters              7.01    11.2     98
 3 WILDCAT          Cover_1               6.67     8.28    12
 4 EMPTY            Cover_1               6.59    11.0    271
 5 EMPTY            Cover_3               6.59     9.95   368
 6 SHOTGUN          Other_Rare            6.44    12.2    282
 7 SHOTGUN          Cover_6               6.38     9.22   890
 8 EMPTY            Other_Rare            6.27    10.2     41
 9 SHOTGUN          Cover-2               6.26     8.99  1192
10 PISTOL           Cover-2               6.08     6.61    37
# i 45 more rows
```

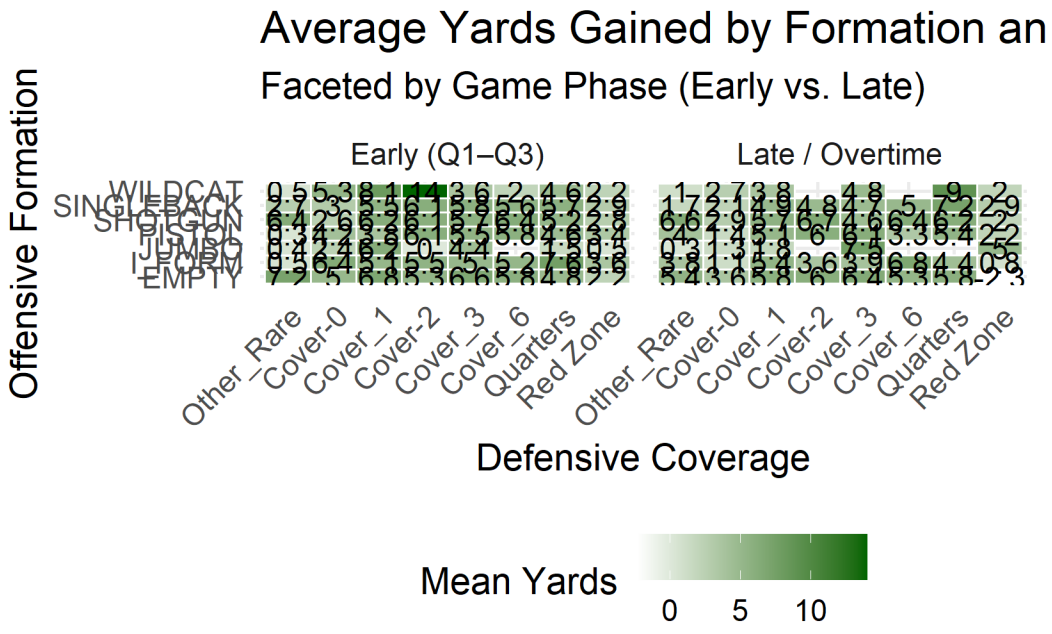
format it and create a heatmap

```r
library(dplyr)
library(knitr)
library(ggplot2)

# Build summary table (same computation as before)
summary_table <- clean_plays_quarter %>%
  group_by(game_phase, offenseFormation, coverage_group) %>%
  summarise(
    mean_yards = mean(yardsGained, na.rm = TRUE),
    sd_yards   = sd(yardsGained, na.rm = TRUE),
    n          = n(),
    .groups = "drop"
  ) %>%
  arrange(desc(mean_yards)) %>%
  mutate(
    mean_yards = round(mean_yards, 2),
    sd_yards = round(sd_yards, 2)
  )

#  PDF-SAFE TABLE (NO HTML)
kable(
  head(summary_table, 10),
  caption = "Average Yards Gained by Formation and Coverage",
  booktabs = TRUE,
  align = "c"
)
```

Table 5: Average Yards Gained by Formation and Coverage

game_phase	offenseFormation	coverage_group	mean_yards	sd_yards	n
Early (Q1–Q3)	WILDCAT	Cover-2	14.00	23.30	3
Late / Overtime	WILDCAT	Quarters	9.00	NA	1
Early (Q1–Q3)	WILDCAT	Cover_1	8.12	9.88	8
Early (Q1–Q3)	I_FORM	Quarters	7.60	12.11	80
Late / Overtime	JUMBO	Cover_3	7.50	8.62	6
Late / Overtime	SINGLEBACK	Quarters	7.24	14.08	66
Early (Q1–Q3)	EMPTY	Other_Rare	7.20	12.42	20
Late / Overtime	I_FORM	Cover_6	6.83	5.59	12
Early (Q1–Q3)	EMPTY	Cover_1	6.79	11.51	219
Late / Overtime	SHOTGUN	Cover-2	6.67	9.48	377

```r
#   PDF-SAFE HEATMAP
ggplot(summary_table,
       aes(x = coverage_group, y = offenseFormation, fill = mean_yards)) +
  geom_tile(color = "white", linewidth = 0.4) +
  geom_text(aes(label = round(mean_yards, 1)), size = 4, color = "black") +
  facet_wrap(~ game_phase) +
  scale_fill_gradient(
    low = "white",
    high = "darkgreen",
    name = "Mean Yards"
  ) +
  labs(
    title = "Average Yards Gained by Formation and Coverage",
    subtitle = "Faceted by Game Phase (Early vs. Late)",
    x = "Defensive Coverage",
    y = "Offensive Formation"
  ) +
  theme_minimal(base_size = 14) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    legend.position = "bottom"
  )
```



Group the 2 freq table for formation and coverage

```
p_form <- ggplot(formation_phase_counts,
                 aes(y = offenseFormation, x = pct, fill = game_phase)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), color = "black") +
  geom_text(aes(label = paste0(round(pct, 1), "%")),
            position = position_dodge(width = 0.8),
            hjust = -0.1, size = 3) +
  scale_fill_manual(values = c(
    "Early (Q1-Q3)" = "#FFD54F",    # KSU gold
    "Late / Overtime" = "#6D4C41"   # dark bronze/brown
  )) +
  labs(
    title = "Formation Usage by Game Phase",
    x = "Percentage of Plays", y = NULL
  ) +
  coord_cartesian(xlim = c(0, max(formation_phase_counts$pct) * 1.2)) +
  theme_minimal(base_size = 12) +
  theme(
    plot.title = element_text(face = "bold", size = 13, hjust = 0.5),
    axis.text.y = element_text(size = 9),
    axis.text.x = element_text(size = 8),
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank()
  )
p_def <- ggplot(defense_phase_counts,
                aes(y = coverage_group, x = pct, fill = game_phase)) +
  geom_bar(stat = "identity", position = position_dodge(width = 0.8), color = "black") +
  geom_text(aes(label = paste0(round(pct, 1), "%")),
            position = position_dodge(width = 0.8),
            hjust = -0.1, size = 3) +
  scale_fill_manual(values = c(
    "Early (Q1-Q3)" = "#FFD54F",
    "Late / Overtime" = "#6D4C41"
  )) +
  labs(
    title = "Defensive Coverage Usage by Game Phase",
    x = "Percentage of Plays", y = NULL
  ) +
  coord_cartesian(xlim = c(0, max(defense_phase_counts$pct) * 1.2)) +
  theme_minimal(base_size = 12) +
```

```
  theme(
    plot.title = element_text(face = "bold", size = 13, hjust = 0.5),
    axis.text.y = element_text(size = 9),
    axis.text.x = element_text(size = 8),
    legend.position = "bottom",
    legend.title = element_blank(),
    panel.grid.major.y = element_blank(),
    panel.grid.minor = element_blank()
  )

library(patchwork)
p_combine1<-p_form + p_def + plot_layout(ncol = 2, widths = c(1, 1))
```

Re-state the model

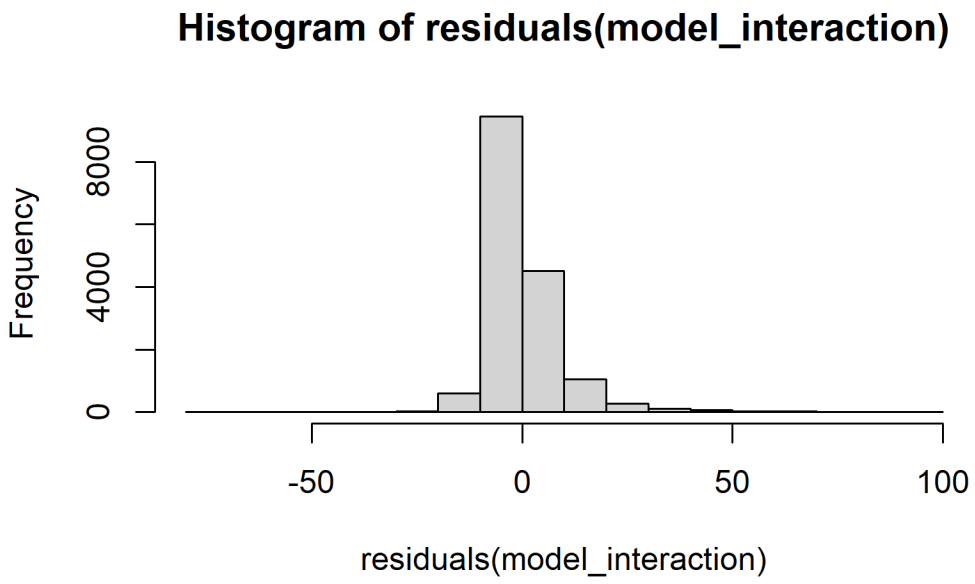
```
# 1. FIT THE MODEL
model_interaction <- lm(yardsGained ~ offenseFormation * coverage_group+game_phase,
                        data = clean_plays_quarter_formation)
```

#THIS PART IS FOR ASSUMPTION TESTING#

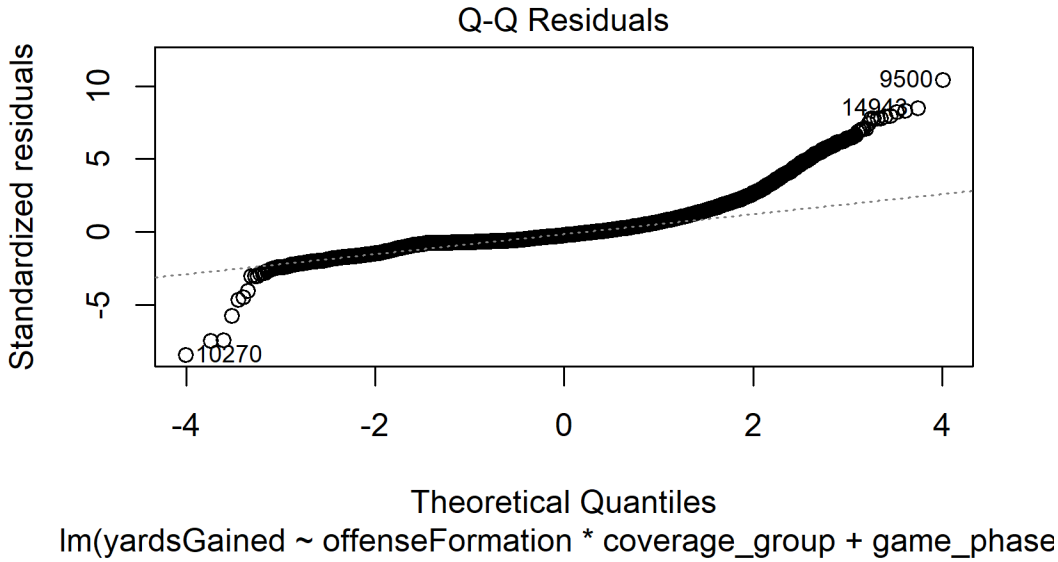
```
#Normality
library(nortest)
ad.test(residuals(model_interaction))
```

```
    Anderson-Darling normality test

data:  residuals(model_interaction)
A = 697.38, p-value < 2.2e-16
```

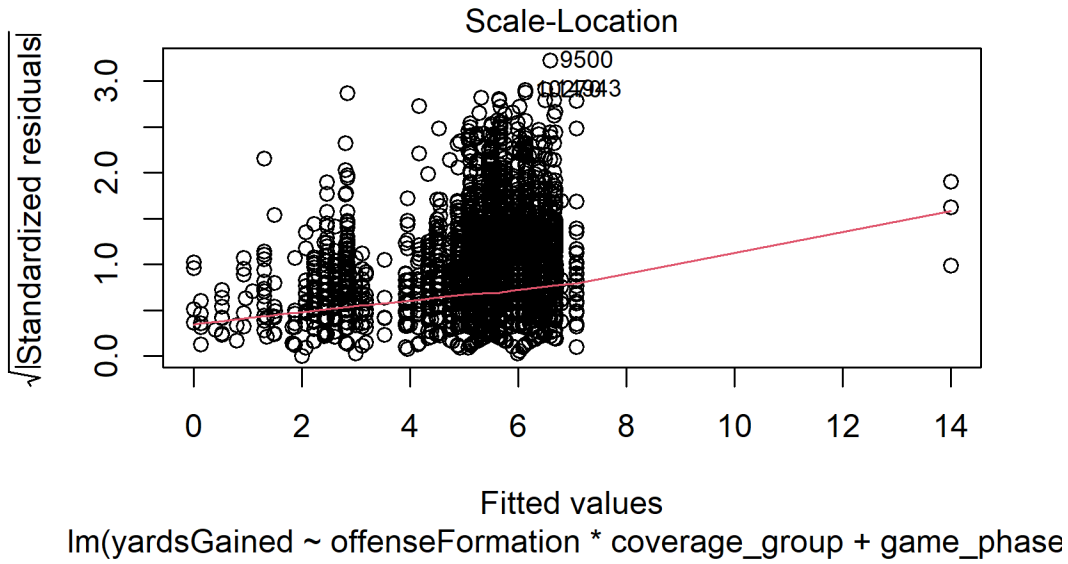
```
hist(residuals(model_interaction))
```




```r
plot(model_interaction, which = 2)
```



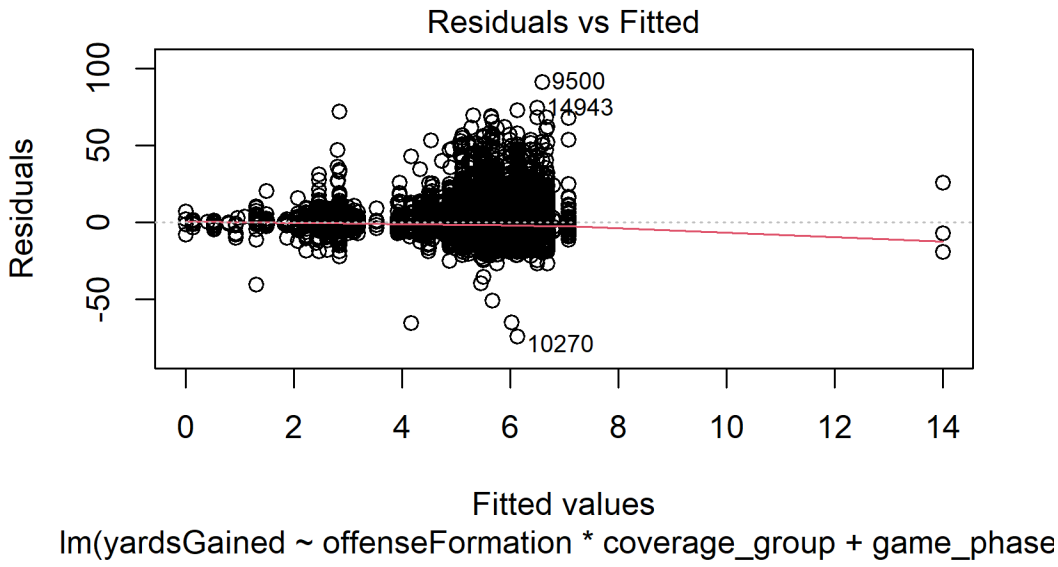
```
library(lmtest)

library(sandwich)
plot(model_interaction, which = 3)
```




```
bptest(model_interaction)
```

```
    studentized Breusch-Pagan test

data:  model_interaction
BP = 125.82, df = 55, p-value = 1.797e-07
```

```
plot(model_interaction, which = 1)
```




```
coeftest(model_interaction,
         vcov = vcovHC(model_interaction, type = "HC3"))
```

```
t test of coefficients:

                                          Estimate Std. Error t value
(Intercept)                                6.46662    1.62051  3.9905
offenseFormationI_FORM                    -4.59131    1.82488 -2.5159
offenseFormationJUMBO                     -5.94619    1.62833 -3.6517
offenseFormationPISTOL                    -5.11982    2.01552 -2.5402
offenseFormationSHOTGUN                    0.12414    1.77225  0.0700
offenseFormationSINGLEBACK                -4.00540    1.77285 -2.2593
offenseFormationWILDCAT                   -5.67088    1.69469 -3.3463
coverage_groupCover-0                     -1.90864    2.14715 -0.8889
coverage_groupCover_1                      0.20178    1.75003  0.1153
coverage_groupCover-2                     -0.80103    1.69905 -0.4715
coverage_groupCover_3                      0.22258    1.69935  0.1310
coverage_groupCover_6                     -0.69386    1.79610 -0.3863
coverage_groupQuarters                    -1.34194    1.68250 -0.7976
coverage_groupRed Zone                    -5.15835    2.00391 -2.5741
game_phaseLate / Overtime                 -0.38720    0.15456 -2.5052
```

```
offenseFormationI_FORM:coverage_groupCover-0       4.56374    2.96285  1.5403
offenseFormationJUMBO:coverage_groupCover-0        3.60438    2.27077  1.5873
offenseFormationPISTOL:coverage_groupCover-0       3.37868    2.81399  1.2007
offenseFormationSHOTGUN:coverage_groupCover-0     -1.83720    2.31023 -0.7952
offenseFormationSINGLEBACK:coverage_groupCover-0   2.25070    2.33525  0.9638
offenseFormationWILDCAT:coverage_groupCover-0      5.99801    3.38976  1.7695
offenseFormationI_FORM:coverage_groupCover_1       3.23135    2.07478  1.5574
offenseFormationJUMBO:coverage_groupCover_1        3.66360    2.05796  1.7802
offenseFormationPISTOL:coverage_groupCover_1       2.78917    2.22060  1.2560
offenseFormationSHOTGUN:coverage_groupCover_1     -0.65828    1.90769 -0.3451
offenseFormationSINGLEBACK:coverage_groupCover_1   2.79627    1.92379  1.4535
offenseFormationWILDCAT:coverage_groupCover_1      5.79822    3.08040  1.8823
offenseFormationI_FORM:coverage_groupCover-2       4.09081    1.99646  2.0490
offenseFormationJUMBO:coverage_groupCover-2        0.28061    4.00360  0.0701
offenseFormationPISTOL:coverage_groupCover-2       5.57717    2.35454  2.3687
offenseFormationSHOTGUN:coverage_groupCover-2      0.58945    1.86529  0.3160
offenseFormationSINGLEBACK:coverage_groupCover-2   4.23610    1.90855  2.2195
offenseFormationWILDCAT:coverage_groupCover-2     14.00530   16.57222  0.8451
offenseFormationI_FORM:coverage_groupCover_3       2.77736    1.92111  1.4457
offenseFormationJUMBO:coverage_groupCover_3        4.45416    2.03464  2.1892
offenseFormationPISTOL:coverage_groupCover_3       4.18419    2.14817  1.9478
offenseFormationSHOTGUN:coverage_groupCover_3     -1.30773    1.85402 -0.7053
offenseFormationSINGLEBACK:coverage_groupCover_3   2.98078    1.85914  1.6033
offenseFormationWILDCAT:coverage_groupCover_3      2.89913    2.04757  1.4159
offenseFormationI_FORM:coverage_groupCover_6       4.30204    2.09327  2.0552
offenseFormationPISTOL:coverage_groupCover_6       4.85440    2.40401  2.0193
offenseFormationSHOTGUN:coverage_groupCover_6      0.60122    1.96131  0.3065
offenseFormationSINGLEBACK:coverage_groupCover_6   3.76979    1.99818  1.8866
offenseFormationWILDCAT:coverage_groupCover_6      1.89812    8.08850  0.2347
offenseFormationI_FORM:coverage_groupQuarters      6.54796    2.19751  2.9797
offenseFormationJUMBO:coverage_groupQuarters       2.32152    2.71279  0.8558
offenseFormationPISTOL:coverage_groupQuarters      4.88465    2.16172  2.2596
offenseFormationSHOTGUN:coverage_groupQuarters     0.36884    1.84528  0.1999
offenseFormationSINGLEBACK:coverage_groupQuarters  4.90959    1.88877  2.5994
offenseFormationWILDCAT:coverage_groupQuarters     5.45134    2.10076  2.5949
offenseFormationI_FORM:coverage_groupRed Zone      6.39382    2.31510  2.7618
offenseFormationJUMBO:coverage_groupRed Zone       6.11537    2.30338  2.6550
offenseFormationPISTOL:coverage_groupRed Zone      6.99898    2.50240  2.7969
offenseFormationSHOTGUN:coverage_groupRed Zone     1.18416    2.14246  0.5527
offenseFormationSINGLEBACK:coverage_groupRed Zone  5.68981    2.15193  2.6441
offenseFormationWILDCAT:coverage_groupRed Zone     6.60133    2.24786  2.9367
                                                  Pr(>|t|)
(Intercept)                                     6.623e-05 ***
```

```
offenseFormationI_FORM                                0.0118809 *
offenseFormationJUMBO                                 0.0002613 ***
offenseFormationPISTOL                                0.0110884 *
offenseFormationSHOTGUN                               0.9441591
offenseFormationSINGLEBACK                            0.0238783 *
offenseFormationWILDCAT                               0.0008210 ***
coverage_groupCover-0                                 0.3740594
coverage_groupCover_1                                 0.9082094
coverage_groupCover-2                                 0.6373188
coverage_groupCover_3                                 0.8957951
coverage_groupCover_6                                 0.6992688
coverage_groupQuarters                                0.4251229
coverage_groupRed Zone                                0.0100577 *
game_phaseLate / Overtime                             0.0122484 *
offenseFormationI_FORM:coverage_groupCover-0          0.1235020
offenseFormationJUMBO:coverage_groupCover-0           0.1124659
offenseFormationPISTOL:coverage_groupCover-0          0.2298966
offenseFormationSHOTGUN:coverage_groupCover-0         0.4264822
offenseFormationSINGLEBACK:coverage_groupCover-0      0.3351634
offenseFormationWILDCAT:coverage_groupCover-0         0.0768375 .
offenseFormationI_FORM:coverage_groupCover_1          0.1193850
offenseFormationJUMBO:coverage_groupCover_1           0.0750609 .
offenseFormationPISTOL:coverage_groupCover_1          0.2091189
offenseFormationSHOTGUN:coverage_groupCover_1         0.7300504
offenseFormationSINGLEBACK:coverage_groupCover_1      0.1460997
offenseFormationWILDCAT:coverage_groupCover_1         0.0598143 .
offenseFormationI_FORM:coverage_groupCover-2          0.0404748 *
offenseFormationJUMBO:coverage_groupCover-2           0.9441237
offenseFormationPISTOL:coverage_groupCover-2          0.0178629 *
offenseFormationSHOTGUN:coverage_groupCover-2         0.7519983
offenseFormationSINGLEBACK:coverage_groupCover-2      0.0264636 *
offenseFormationWILDCAT:coverage_groupCover-2         0.3980637
offenseFormationI_FORM:coverage_groupCover_3          0.1482789
offenseFormationJUMBO:coverage_groupCover_3           0.0285993 *
offenseFormationPISTOL:coverage_groupCover_3          0.0514574 .
offenseFormationSHOTGUN:coverage_groupCover_3         0.4806032
offenseFormationSINGLEBACK:coverage_groupCover_3      0.1088850
offenseFormationWILDCAT:coverage_groupCover_3         0.1568280
offenseFormationI_FORM:coverage_groupCover_6          0.0398781 *
offenseFormationPISTOL:coverage_groupCover_6          0.0434734 *
offenseFormationSHOTGUN:coverage_groupCover_6         0.7591987
offenseFormationSINGLEBACK:coverage_groupCover_6      0.0592301 .
offenseFormationWILDCAT:coverage_groupCover_6         0.8144683
```

```
offenseFormationI_FORM:coverage_groupQuarters     0.0028895 **
offenseFormationJUMBO:coverage_groupQuarters      0.3921390
offenseFormationPISTOL:coverage_groupQuarters     0.0238584 *
offenseFormationSHOTGUN:coverage_groupQuarters    0.8415759
offenseFormationSINGLEBACK:coverage_groupQuarters 0.0093484 **
offenseFormationWILDCAT:coverage_groupQuarters    0.0094693 **
offenseFormationI_FORM:coverage_groupRed Zone     0.0057551 **
offenseFormationJUMBO:coverage_groupRed Zone      0.0079397 **
offenseFormationPISTOL:coverage_groupRed Zone     0.0051655 **
offenseFormationSHOTGUN:coverage_groupRed Zone    0.5804709
offenseFormationSINGLEBACK:coverage_groupRed Zone 0.0081999 **
offenseFormationWILDCAT:coverage_groupRed Zone    0.0033217 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
library(car)
library(knitr)
library(olsrr)
model_main <- lm(yardsGained ~ offenseFormation + coverage_group+game_phase,
                 data = clean_plays_quarter_formation)

vif_table <- vif(model_main)

# Convert to data frame for kable
vif_df <- as.data.frame(vif_table)

# Add row names as a column
vif_df$Predictor <- rownames(vif_df)
rownames(vif_df) <- NULL

# Reorder columns
vif_df <- vif_df[, c("Predictor", "GVIF", "Df", "GVIF^(1/(2*Df))")]

# Print clean table
kable(
  vif_df,
  caption = "Multicollinearity Diagnostics (GVIF Results)",
  digits = 6
)
```

Table 6: Multicollinearity Diagnostics (GVIF Results)

Predictor	GVIF	Df	GVIF^(1/(2*Df))
offenseFormation	1.090163	6	1.007220
coverage_group	1.090878	7	1.006232
game_phase	1.010826	1	1.005398

```r
library(sandwich)
library(lmtest)

# OLS coefficients
ols <- summary(model_interaction)$coefficients

# HC3 robust coefficients
rob <- coeftest(model_interaction, vcov = vcovHC(model_interaction, type = "HC3"))
```

```r
library(dplyr)
library(gt)

# Get OLS + HC3 results
ols <- summary(model_interaction)$coefficients
rob <- coeftest(model_interaction, vcov = vcovHC(model_interaction, type = "HC3"))

# Build unified table
combined <- data.frame(
  Term = rownames(ols),
  Estimate = ols[, 1],
  `Std.Error.OLS` = ols[, 2],
  `p.value.OLS` = ols[, 4],
  `Std.Error.HC3` = rob[, 2],
  `p.value.HC3` = rob[, 4]
)

# Filter only predictors significant under HC3
sig_table <- combined %>%
  filter(`p.value.HC3` < 0.05) %>%
  mutate(
    significance = case_when(
      `p.value.HC3` < 0.001 ~ "***",
      `p.value.HC3` < 0.01 ~ "**",
      `p.value.HC3` < 0.05 ~ "*"
```

```
    )
  )


sig_table <- sig_table %>%
  rename(
    `Std. Error (OLS)` = Std.Error.OLS,
    `p value (OLS)` = p.value.OLS,
    `Std. Error (HC3)` = Std.Error.HC3,
    `p value (HC3)` = p.value.HC3
  )


sig_table %>%
  gt() %>%
  tab_header(title = "Significant Effects Before and After HC3 Robust Correction") %>%

  # Yellow header row
  tab_style(
    style = list(
      cell_fill(color = "#FFD54F"),
      cell_text(weight = "bold")
    ),
    locations = cells_column_labels(everything())
  ) %>%

  # Round numeric columns
  fmt_number(
    columns = c(Estimate, `Std. Error (OLS)`, `p value (OLS)`,
                `Std. Error (HC3)`, `p value (HC3)`),
    decimals = 3
  ) %>%


  tab_style(
    style = cell_text(color = "red"),
    locations = cells_body(columns = c(`p value (OLS)`, `p value (HC3)`))
  ) %>%


  tab_style(
    style = cell_text(weight = "bold"),
```

```
    locations = cells_body(columns = Term)
  )
```

[

Significant Effects Before and After HC3 Robust Correction] Significant Effects Before and After HC3 Robust Correction

Term	Estimate	Std. Error (OLS)
(Intercept)	6.467	1.375
offenseFormationI_FORM	-4.591	2.112
offenseFormationJUMBO	-5.946	1.828
offenseFormationPISTOL	-5.120	4.605
offenseFormationSINGLEBACK	-4.005	1.821
offenseFormationWILDCAT	-5.671	5.258
coverage_groupRed Zone	-5.158	1.898
game_phaseLate / Overtime	-0.387	0.155
offenseFormationI_FORM:coverage_groupCover-2	4.091	2.384
offenseFormationPISTOL:coverage_groupCover-2	5.577	4.867
offenseFormationSINGLEBACK:coverage_groupCover-2	4.236	1.988
offenseFormationJUMBO:coverage_groupCover_3	4.454	2.533
offenseFormationI_FORM:coverage_groupCover_6	4.302	2.471
offenseFormationPISTOL:coverage_groupCover_6	4.854	4.864
offenseFormationI_FORM:coverage_groupQuarters	6.548	2.359
offenseFormationPISTOL:coverage_groupQuarters	4.885	4.716
offenseFormationSINGLEBACK:coverage_groupQuarters	4.910	1.952
offenseFormationWILDCAT:coverage_groupQuarters	5.451	5.701
offenseFormationI_FORM:coverage_groupRed Zone	6.394	2.959
offenseFormationJUMBO:coverage_groupRed Zone	6.115	4.529
offenseFormationPISTOL:coverage_groupRed Zone	6.999	5.158
offenseFormationSINGLEBACK:coverage_groupRed Zone	5.690	2.390
offenseFormationWILDCAT:coverage_groupRed Zone	6.601	6.090