# NBA Clutch Gene Analysis

Andrew Edelman

2022-10-25

## Contents

# Objective

## Background

The "clutch gene" is a hotly debated topic in sports, NBA included. The clutch gene is simple a way of stating that a player can deliver under pressure, usually towards the end of a game, when winning the game is on the line. Arguing the most clutch players is a tired debate - who can deliver, and who chokes under pressure, not to mention the realistic perspective that there is no clutch gene at all. This article from 2011 about LeBron James sums up the clutch gene and the debate better than I ever could.

## Goal

This analysis will look at play-by-play data from the 2015-16 to 2020-21 seasons to identify if players in general perform differently under game winning/losing scenarios, and which players are the most or least clutch.

# Load In Packages Used

```
library(dplyr)
library(ggplot2)
library(tibble)
library(tidyr)
library(readr)
library(knitr)
library(tinytex)
library(formatR)
library(broom)
library(pscl)
```

# Getting the Data

The data for this analysis comes from Kaggle's NBA Play-by-Play Data 2015-2021 data set. This data breaks down every play of the games occurring in 2015-2021.

## Import Data

The data set has quite a few extra variables, so we'll limit to only variables and context around shots taken. We'll then filter down the data set to only plays where shots were taken (as this is all plays during games, not just shots).

```
# the data was split in to six CSV's, so we'll import each
# and combine them
data201516 <- read_csv("NBA_PBP_2015-16.csv", col_types = c("_c_c_ciic_ic_icccd__________cc__________"))
data201617 <- read_csv("NBA_PBP_2016-17.csv", col_types = c("_c_c_ciic_ic_icccd__________cc__________"))
data201718 <- read_csv("NBA_PBP_2017-18.csv", col_types = c("_c_c_ciic_ic_icccd__________cc__________"))
data201819 <- read_csv("NBA_PBP_2018-19.csv", col_types = c("_c_c_ciic_ic_icccd__________cc__________"))
data201920 <- read_csv("NBA_PBP_2019-20.csv", col_types = c("_c_c_ciic_ic_icccd__________cc___________"
```

```
data202021 <- read_csv("NBA_PBP_2020-21.csv", col_types = c("_c_c_ciic_ic_icccd_________cc_________"))

# combine all seasons
all_seasons_raw_data <- data.frame(rbind(data201516, data201617,
    data201718, data201819, data201920, data202021))
str(all_seasons_raw_data)

# limit data to only shots taken plays, all shot plays have
# either a Shooter or a FreeThrowShooter (if the shooter
# was fouled)
all_seasons_shots_raw <- all_seasons_raw_data %>%
    filter(!is.na(Shooter) | !is.na(FreeThrowShooter))
head(all_seasons_shots_raw)
```

# Preparing the Data

## Adding Season Classification

We'll do a few things to clean up/organize the shots data.

First, we'll use the Date column to create a season classification (since each season spans two calendar years we can't just use the year of the date variable).

We'll also replace NA ShotType values with Free Throw, since those are store in a separate variable, so we can have our categorical classifications all in one column.

```
# convert date field to date
all_seasons_shots_raw <- all_seasons_shots_raw %>%
    mutate(Date = as.Date(Date, format = "%B %d %Y"))

# get season from date
all_seasons_shots_raw$Season <- with(all_seasons_shots_raw, ifelse(Date <=
    "2016-06-19", "2015-16", ifelse(Date <= "2017-06-18", "2016-17",
    ifelse(Date <= "2018-06-17", "2017-18", ifelse(Date <= "2019-06-13",
        "2018-19", ifelse(Date <= "2020-10-11", "2019-20", ifelse(Date <=
            "2021-07-22", "2020-21", NA)))))))

# check counts for validation
all_seasons_shots_raw %>%
    group_by(Season) %>%
    count()
```

```
## # A tibble: 6 x 2
## # Groups:   Season [6]
##   Season       n
##   <chr>    <int>
## 1 2015-16 283808
## 2 2016-17 283837
## 3 2017-18 282483
## 4 2018-19 294308
## 5 2019-20 255367
## 6 2020-21  46282
```

### Defining clutch

Second, and more importantly, we need to outline which shots are "clutch" shots. As mentioned above, clutch generally refers to players performing under pressure with the game on the line. This is a bit artibitrary, so we'll use this criteria to note shots as being "clutch" - a decreasing scale of, as the time left in the game is dwindling, the point differential between the two teams is closer. The time remaining and differential specifics were qualitatively decided based on anecdotal knowledge of when it seems realistic that a team can still come back to win, meaning that "it's crunch time and every shot counts".

For example, shots taken and during the last 2 minutes of a game where the score is 125-103 are not clutch, because there is essentially no chance of the outcome being swayed in that time. However, shots taken with 2.5 minutes left when a team is down 5 points are clutch, because there is little time remaining and the pressure is on - every point is needed to keep the comeback alive, or solidify the win.

Therefore, "clutch" shots will be defined as shots taken with:

1. Less than 3-5 minutes remaining in the game, point differential of <=12
2. Less than 1-3 minutes remaining in the game, point differential of <=8
3. Less than <=1 minute remaining in the game, point differential of <=5

```r
# declare prepared dataset, adding column for point
# differential between teams
all_shots <- all_seasons_shots_raw %>%
    mutate(Diff = abs(HomeScore - AwayScore))

# define clutch shots
all_shots$Clutch <- with(all_shots, ifelse(Quarter == 4 & SecLeft <=
    300 & SecLeft > 180 & Diff <= 12, 1, ifelse(Quarter == 4 &
    SecLeft <= 180 & SecLeft > 60 & Diff <= 8, 1, ifelse(Quarter ==
    4 & SecLeft <= 60 & Diff <= 5, 1, 0))))

# recode Clutch variable as factor
all_shots$Clutch <- as.factor(all_shots$Clutch)

# define numeric shot made/free throw made column for
# analysis
all_shots$ShotMade <- with(all_shots, ifelse(ShotOutcome == "make" |
    FreeThrowOutcome == "make", 1, 0))

# replace NA values with 0
all_shots <- all_shots %>%
    replace_na(list(ShotMade = 0L))

# replace NA values in shot type with free throws
all_shots$ShotType <- all_shots$ShotType %>%
    replace_na("Free Throw")
unique(all_shots$ShotType)
```

```
## [1] "2-pt layup"     "2-pt jump shot" "2-pt hook shot" "Free Throw"
## [5] "3-pt jump shot" "2-pt dunk"      "3-pt layup"     "3-pt hook shot"
## [9] "2-pt tip-in"
```

### Weighting Shots

The last thing we'll do is assign a weight to each shot made, relative to the point value for each shot type. 2PT shots will get a weight of 2, 3PT shots will get a weight of 3, and free throws will get a weight of 1 each.

```
# Get shot point value from ShotType column
all_shots <- all_shots %>%
    mutate(ShotWeight = as.integer(substr(ShotType, 1, 1)))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
# Fill NA's as 1 to account for free throws
all_shots <- all_shots %>%
    replace_na(list(ShotWeight = 1L))

# Change game type to a factor for regular and playoff
all_shots <- all_shots %>%
    mutate(GameType = factor(GameType, levels = c("regular",
        "playoff")))
head(all_shots)
```

```
##   GameType       Date WinningTeam Quarter SecLeft AwayTeam AwayScore HomeTeam
## 1  regular 2015-10-27         DET       1     701      DET         0      ATL
## 2  regular 2015-10-27         DET       1     681      DET         2      ATL
## 3  regular 2015-10-27         DET       1     660      DET         2      ATL
## 4  regular 2015-10-27         DET       1     644      DET         4      ATL
## 5  regular 2015-10-27         DET       1     627      DET         4      ATL
## 6  regular 2015-10-27         DET       1     613      DET         4      ATL
##   HomeScore                      Shooter       ShotType ShotOutcome ShotDist
## 1         0      A. Drummond - drumman01     2-pt layup        miss        1
## 2         0        M. Morris - morrima03 2-pt jump shot        make       13
## 3         2       P. Millsap - millspa01 2-pt jump shot        make       12
## 4         2 K. Caldwell-Pope - caldwke01 2-pt jump shot        make        8
## 5         2       A. Horford - horfoal01 2-pt jump shot        miss       20
## 6         2      A. Drummond - drumman01 2-pt hook shot        miss        8
##   FreeThrowShooter FreeThrowOutcome  Season Diff Clutch ShotMade ShotWeight
## 1             <NA>             <NA> 2015-16    0      0        0          2
## 2             <NA>             <NA> 2015-16    2      0        1          2
## 3             <NA>             <NA> 2015-16    0      0        1          2
## 4             <NA>             <NA> 2015-16    2      0        1          2
## 5             <NA>             <NA> 2015-16    2      0        0          2
## 6             <NA>             <NA> 2015-16    2      0        0          2
```

## Identifying Clutch and Non-clutch Players

The analysis above shows that there's no general trend among NBA players indicating players perform better in clutch scenarios - i.e. the clutch gene is not ubiquitous. That does beg the question of whether there are individual players who can be deemed more clutch than others. Going back to the intro, this is something hotly debated, particularly in the NBA.

## Prepping the Data

First, we'll identify clutch players based on difference in make rate for clutch vs. non-clutch shots. To do this we'll need to prep the data to get counts on player shots for clutch/non-clutch scenarios, and calculate respective make rates for each scenario.

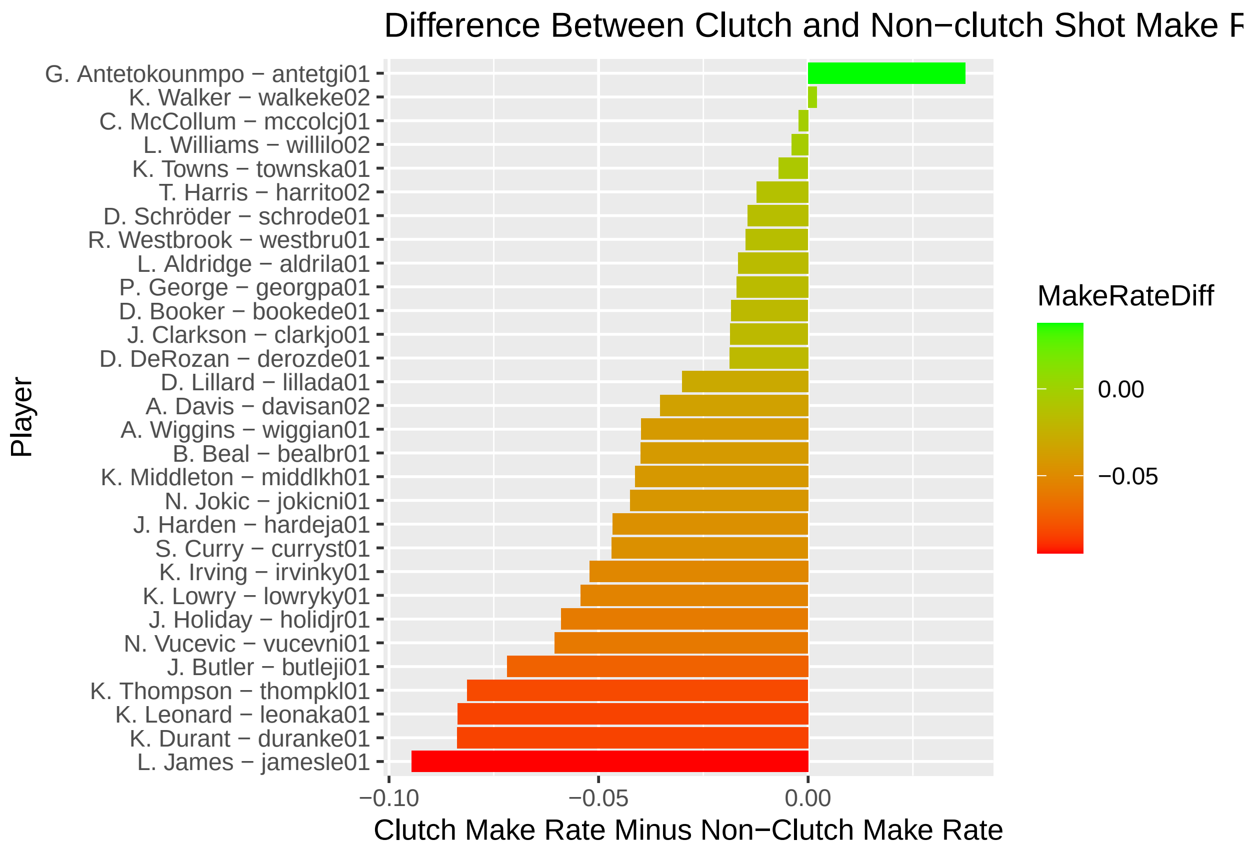
```
# aggregate shots by player for clutch and nonclutch
# scenarios
clutch_shots_player <- all_shots %>%
    filter(Clutch == 1) %>%
    group_by(Shooter) %>%
    summarize(ClutchShotsTaken = n(), ClutchShotsMade = sum(ShotMade))
nonclutch_shots_player <- all_shots %>%
    filter(Clutch == 0) %>%
    group_by(Shooter) %>%
    summarize(NonClutchShotsTaken = n(), NonClutchShotsMade = sum(ShotMade))

# join clutch and nonclutch player shot counts
player_aggregate <- clutch_shots_player %>%
    left_join(nonclutch_shots_player, by = "Shooter")

# create metrics for shots taken and make rates for clutch
# scenarios
player_aggregate <- player_aggregate %>%
    mutate(TotalShotsTaken = ClutchShotsTaken + NonClutchShotsTaken,
        TotalShotsMade = ClutchShotsMade + NonClutchShotsMade)
player_aggregate <- player_aggregate %>%
    mutate(ClutchMakeRate = ClutchShotsMade/ClutchShotsTaken,
        NonClutchMakeRate = NonClutchShotsMade/NonClutchShotsTaken,
        TotalMakeRate = TotalShotsMade/TotalShotsTaken)

# calculate difference in make rate for clutch vs non
# clutch shots
player_aggregate <- player_aggregate %>%
    mutate(MakeRateDiff = ClutchMakeRate - NonClutchMakeRate)

# drop NA value for missing shooter observation
player_aggregate <- player_aggregate %>%
    filter(!is.na(Shooter))
```

## Make Rate for Top Players by Shots Taken

Doing some initial exploration, we'll first look at the difference in Clutch Make Rate - Non Clutch Make Rate for the top 30 players by total number of shots taken. Typically players taking the most shots are the best players, players being given the ball in crunch time during clutch scenarios. So it's reasonable to assume these players are most centered in the conversation of "who's clutch".

```
player_aggregate_top30 <- player_aggregate %>%
    slice_max(order_by = TotalShotsTaken, n = 30) %>%
    arrange(desc(MakeRateDiff))

ggplot(player_aggregate_top30, aes(x = MakeRateDiff, y = reorder(Shooter,
    MakeRateDiff))) + geom_col(aes(fill = MakeRateDiff)) + ggtitle("Difference Between Clutch and Non-cl
```

```
    xlab("Clutch Make Rate Minus Non-Clutch Make Rate") + ylab("Player") +
    scale_fill_gradient(low = "red", high = "green")
```




Based purely on the rate at which they make shots in clutch vs. non-clutch scenarios, only two players have an increase in their total make rate: Giannis Antetokounmpo, and Kemba Walker. Giannis leads the pack by a wide margin, his Non-Clutch to Clutch make rate increases about 0.04 basis points, from 0.536% to 0.573%. On the other hand, LeBron, Kevin Durant, and Kawhi Leonard round out the bottom three least clutch players in terms of difference in rate of shots make for non-clutch and clutch scenarios.

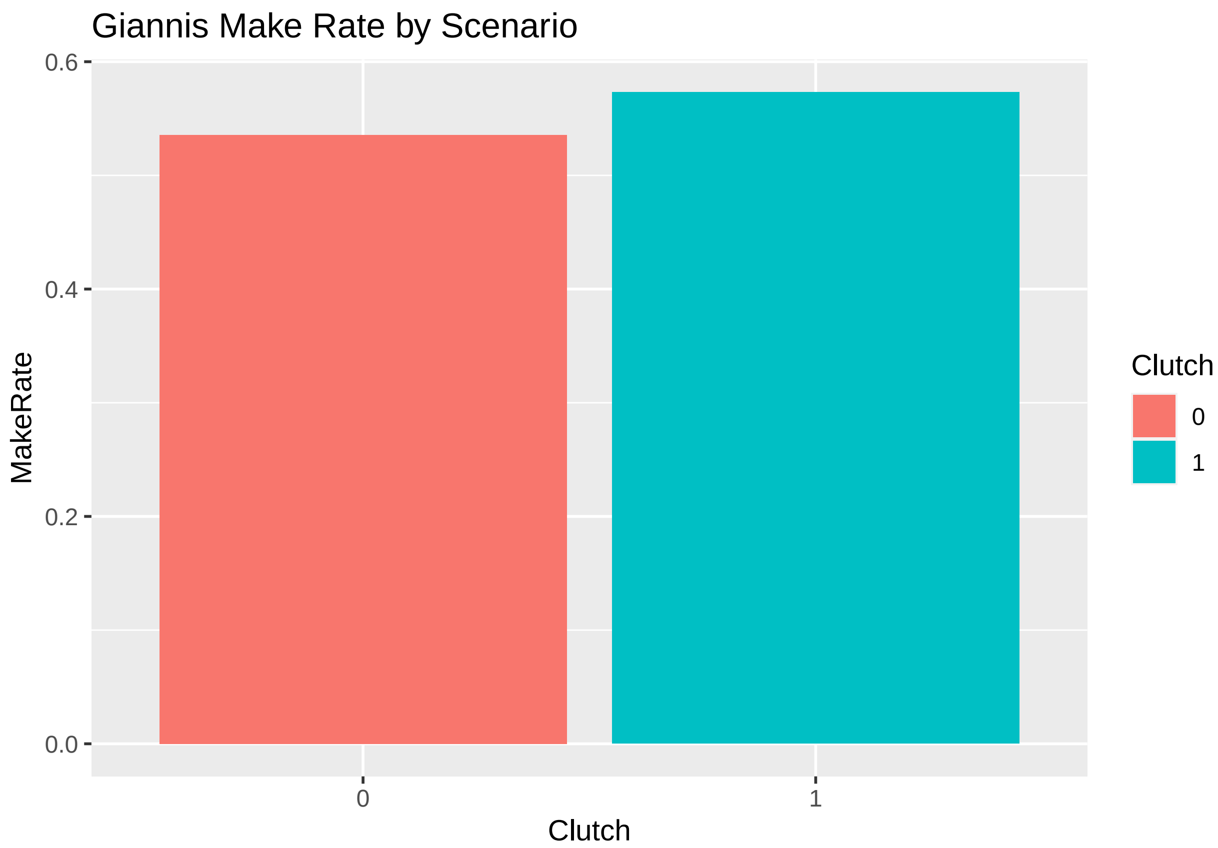
## Testing for Statistical Significance

Essentially what we're seeing is the field-gold % rate going up/down for specific players in clutch scenarios, compared to non-clutch scenarios. For Giannis, an increase of ~8 basis points for field goal % is pretty significant. And on the bottom end of the spectrum, a decrease in ~9 basis points for Lebron is also pretty significant. For the purposes of this analysis, we'll only look at these two players, the top and bottom players by different in make rate. We'll use logistic regression to test if these differences are statistically significant.

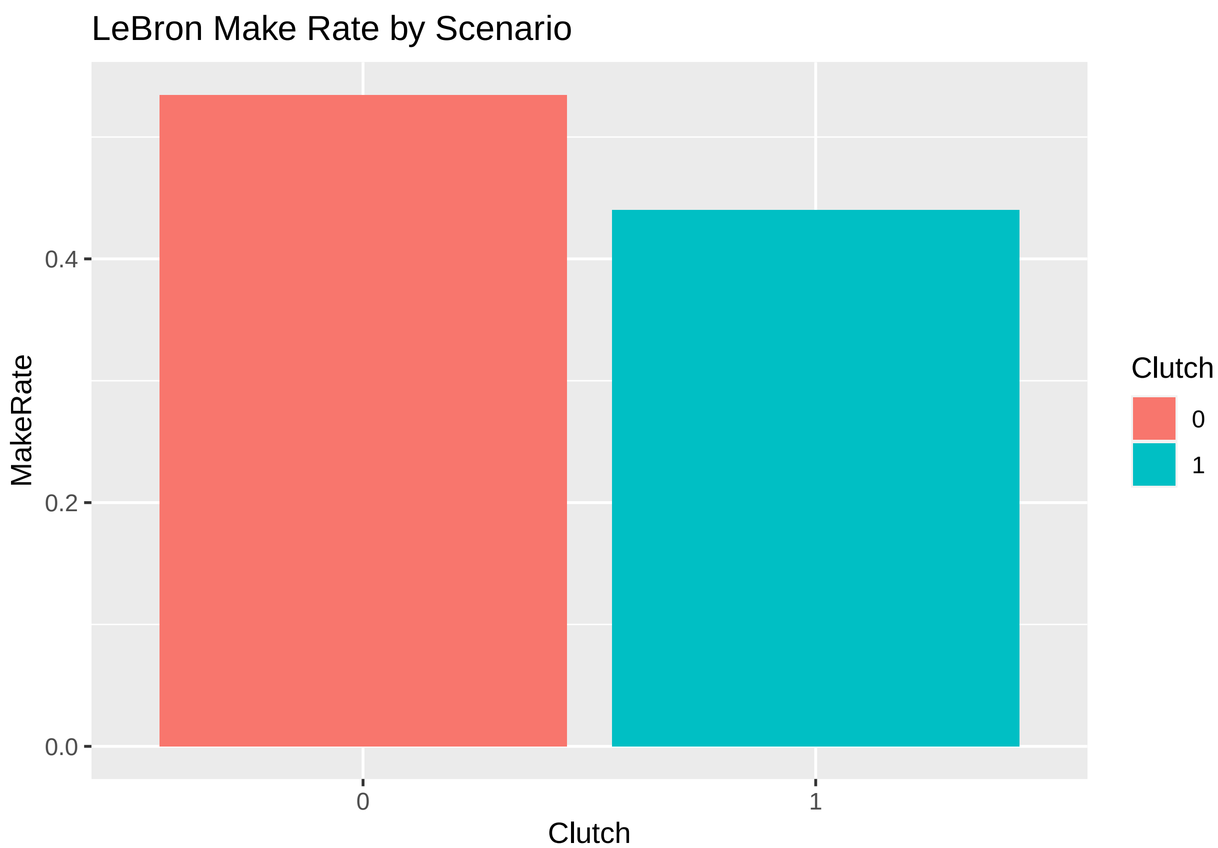
### Filtering Data for Players

```
# filter all shots data for each player
giannis_shots <- all_shots %>%
    filter(Shooter == "G. Antetokounmpo - antetgi01")
lebron_shots <- all_shots %>%
```

```r
    filter(Shooter == "L. James - jamesle01")

# visualize make rate for each player by scenario
giannis_shots %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Giannis Make Rate by Scenario")
```




```r
lebron_shots %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("LeBron Make Rate by Scenario")
```




**Logistic Regression for Significance**

```
# run logistic model for giannis shot data
giannis_model <- glm(ShotMade ~ Clutch, family = "binomial",
    data = giannis_shots)
giannis_model
```

```
##
## Call:  glm(formula = ShotMade ~ Clutch, family = "binomial", data = giannis_shots)
##
## Coefficients:
## (Intercept)      Clutch1
##      0.1426       0.1522
##
## Degrees of Freedom: 7054 Total (i.e. Null);  7053 Residual
## Null Deviance:       9740
## Residual Deviance: 9738  AIC: 9742
```

```
pR2(giannis_model)
```

```
## fitting null model for pseudo-r2
```

```
##           llh       llhNull            G2       McFadden          r2ML
```

```
## -4.868899e+03 -4.870000e+03  2.203216e+00  2.262029e-04  3.122427e-04
##          r2CU
##  4.171202e-04
```

```r
# run logistic model for lebron shot data
lebron_model <- glm(ShotMade ~ Clutch, family = "binomial", data = lebron_shots)
lebron_model
```

```
##
## Call:  glm(formula = ShotMade ~ Clutch, family = "binomial", data = lebron_shots)
##
## Coefficients:
## (Intercept)      Clutch1
##      0.1388      -0.3799
##
## Degrees of Freedom: 8697 Total (i.e. Null);  8696 Residual
## Null Deviance:       12030
## Residual Deviance: 12010     AIC: 12010
```

```r
pR2(lebron_model)
```

```
## fitting null model for pseudo-r2
```

```
##           llh       llhNull            G2      McFadden          r2ML
## -6.004740e+03 -6.016290e+03  2.310015e+01  1.919801e-03  2.652277e-03
##          r2CU
##  3.539822e-03
```

At a significance level of 0.05, the model shows that Giannis's different in performance between scenarios is not significant. However, the results for LeBron's model are statistically significant. Looking at the McFadden pseudo R2 for each model, we see that the McFadden F2 is very small for both models.

### Results

Takeaway 1: there's no statistical significant or predictability in different of Giannis's performance based on clutch scenarios. We can say that the different in Make Rate is statistically lower for LeBron in clutch scenarios, but the pure fact that it is a clutch scenario does not explain this variation, and is not a strong predictor in this difference (i.e. LeBron's drop in field goal make rate in clutch scenarios is not a factor of him "shrinking in the pressure"). As there are many factors that can go in to these scenarios, particularly with star players that are defended heavily towards the end of games, for example, this isn't altogether surprising.

Takeaway 2: Using this basic model, we don't have evidence to say that the "clutch gene" exists, at least for the players on which it was run here.

## Do Playoffs Make a Difference?

Although we didn't find evidence of the clutch gene based on Giannis and Lebron's field goal make rates, one other factor we can take in to account for this is playoff versus regular season games. With a lot more on the line during playoff games, we can re-run the analysis above to determine if players perform differrent under that added pressure.

## Prepping the Data

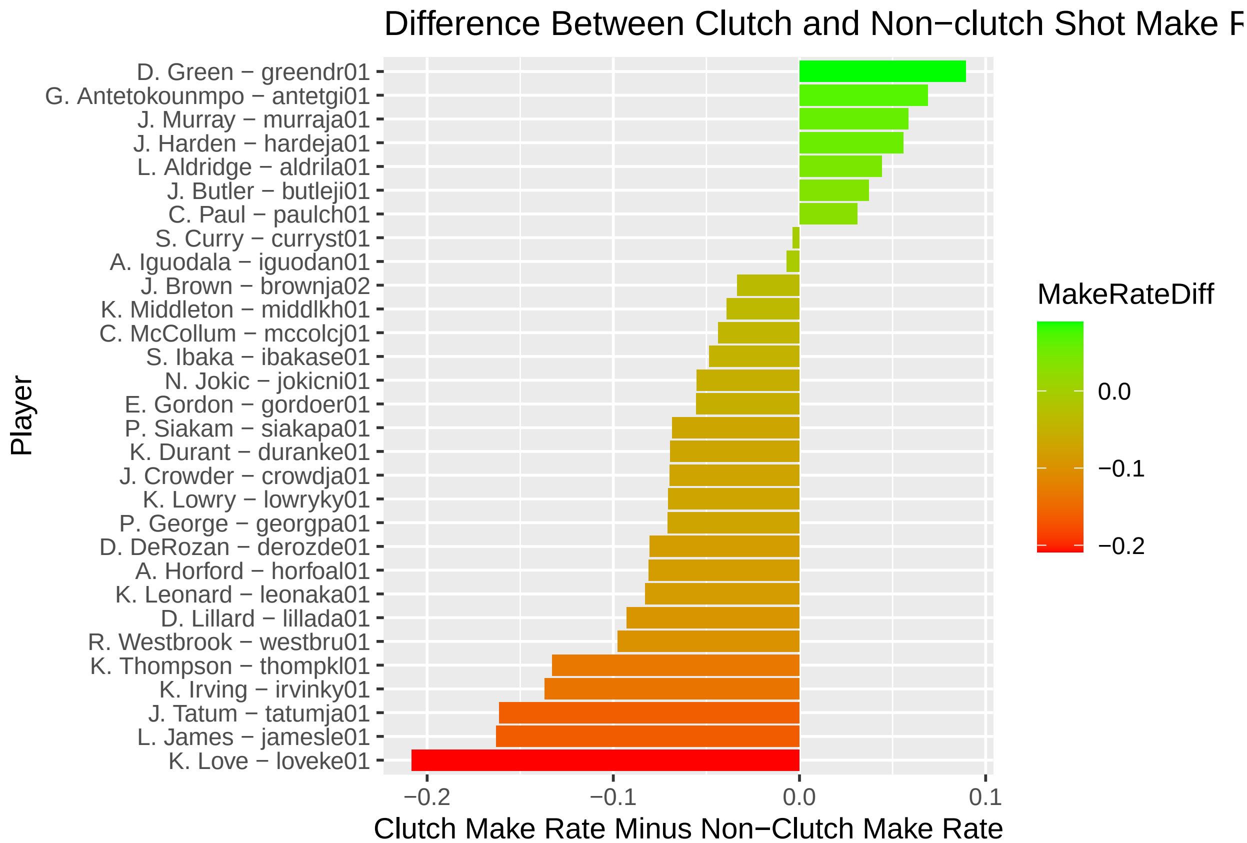
Running the same data prep steps from above, but further filtering the data down to playoff games only.

```r
# aggregate shots by player for clutch and nonclutch
# scenarios
playoff_clutch <- all_shots %>%
    filter(Clutch == 1, GameType == "playoff") %>%
    group_by(Shooter) %>%
    summarize(ClutchShotsTaken = n(), ClutchShotsMade = sum(ShotMade))
playoff_nonclutch <- all_shots %>%
    filter(Clutch == 0, GameType == "playoff") %>%
    group_by(Shooter) %>%
    summarize(NonClutchShotsTaken = n(), NonClutchShotsMade = sum(ShotMade))

# join clutch and nonclutch player shot counts
playoff_aggregate <- playoff_clutch %>%
    left_join(playoff_nonclutch, by = "Shooter")

# create metrics for shots taken and make rates for clutch
# scenarios
playoff_aggregate <- playoff_aggregate %>%
    mutate(TotalShotsTaken = ClutchShotsTaken + NonClutchShotsTaken,
        TotalShotsMade = ClutchShotsMade + NonClutchShotsMade)
playoff_aggregate <- playoff_aggregate %>%
    mutate(ClutchMakeRate = ClutchShotsMade/ClutchShotsTaken,
        NonClutchMakeRate = NonClutchShotsMade/NonClutchShotsTaken,
        TotalMakeRate = TotalShotsMade/TotalShotsTaken)

# calculate difference in make rate for clutch vs non
# clutch shots
playoff_aggregate <- playoff_aggregate %>%
    mutate(MakeRateDiff = ClutchMakeRate - NonClutchMakeRate)

# drop NA value for missing shooter observation
playoff_aggregate <- playoff_aggregate %>%
    filter(!is.na(Shooter))
```

## Make Rate for Top Players by Shots Taken (Playoffs)

Again, identifying top and bottom players by difference in shot make rate based on clutch/non-clutch scenarios, limited to playoff games only.

```r
playoff_aggregate_top30 <- playoff_aggregate %>%
    slice_max(order_by = TotalShotsTaken, n = 30) %>%
    arrange(desc(MakeRateDiff))

ggplot(playoff_aggregate_top30, aes(x = MakeRateDiff, y = reorder(Shooter,
    MakeRateDiff))) + geom_col(aes(fill = MakeRateDiff)) + ggtitle("Difference Between Clutch and Non-cl
    xlab("Clutch Make Rate Minus Non-Clutch Make Rate") + ylab("Player") +
    scale_fill_gradient(low = "red", high = "green")
```




Immediately we can see a change in results here based purely on the difference in make rate between clutch scenarios and non-clutch scenarios, for playoff games. Here we have seven players of the top 30 in shots taken, who have a positive differential (meaning they're making shots at a higher rate in clutch scenarios), whereas with total games we only had 2. Giannis is now second differential, falling behind Draymond Green. On the other end of the spectrum, LeBron is still towards the bottom, but now second worst in terms of differential, and Kevin Love has taken the bottom spot. We're also generally seeing more extreme differences in rate at the top and bottom ends of the spectrum.

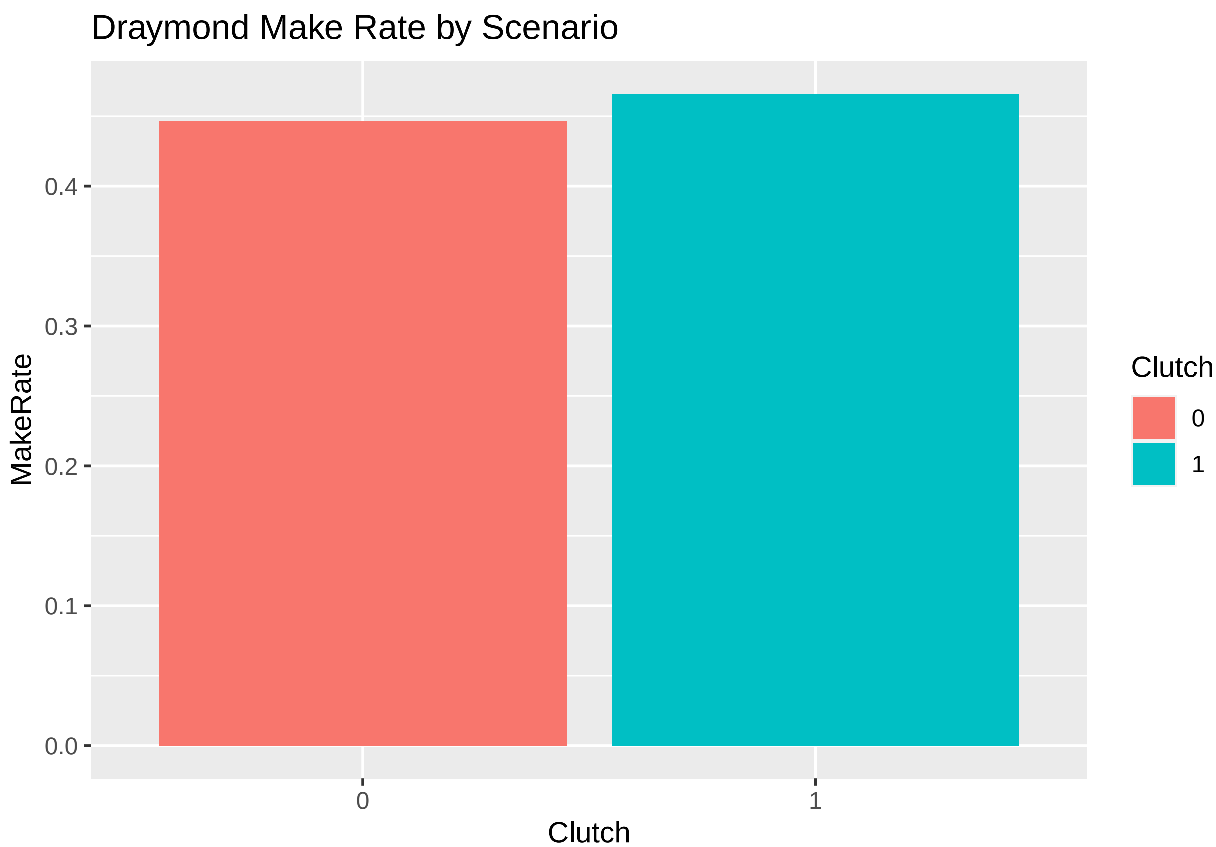
## Statistical Significance for Playoffs

We'll run the same logistic regression above for the top and bottom few players based on the chart above.

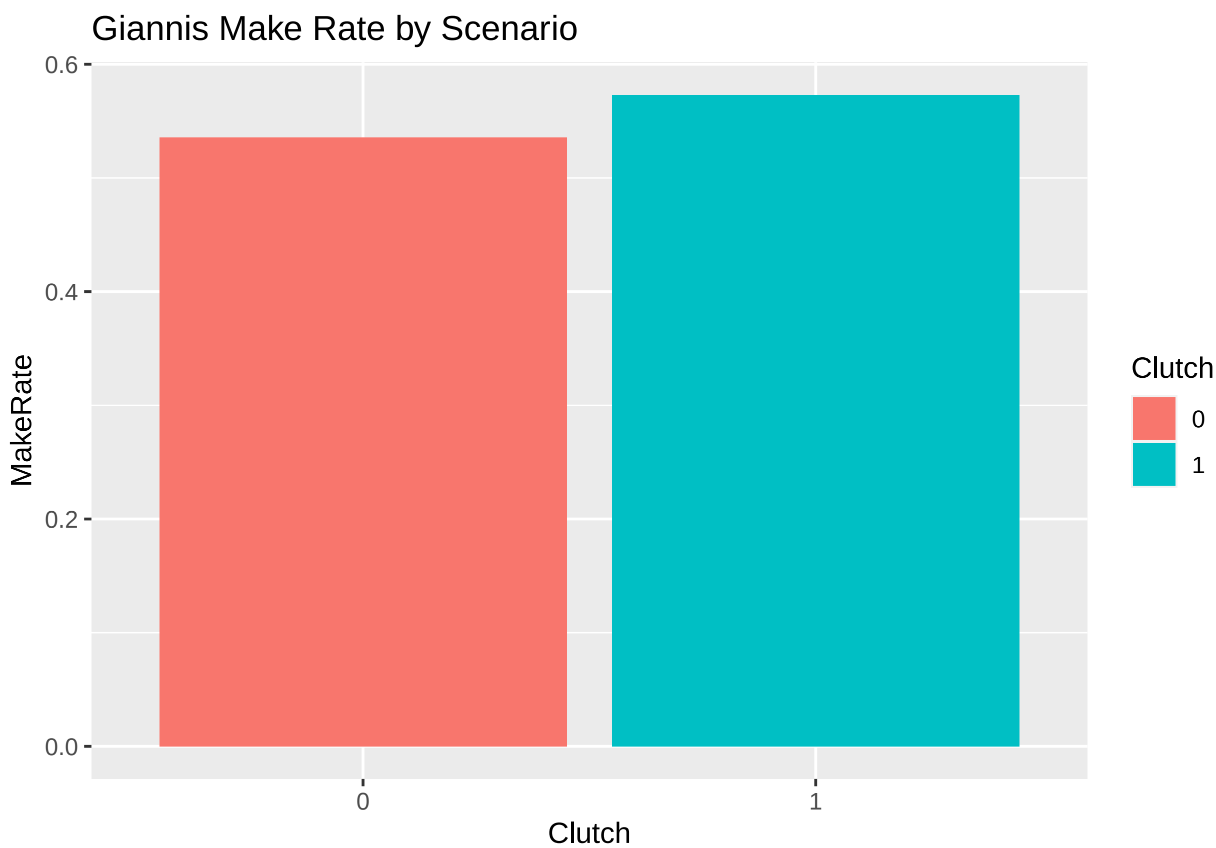
### Filtering Data for Players

```
# filter all shots data for each player
draymond_shots_playoff <- all_shots %>%
    filter(Shooter == "D. Green - greendr01")
giannis_shots_playoff <- all_shots %>%
    filter(Shooter == "G. Antetokounmpo - antetgi01")
jamal_shots_playoff <- all_shots %>%
    filter(Shooter == "J. Murray - murraja01")
lebron_shots_playoff <- all_shots %>%
    filter(Shooter == "L. James - jamesle01")
tatum_shots_playoff <- all_shots %>%
```

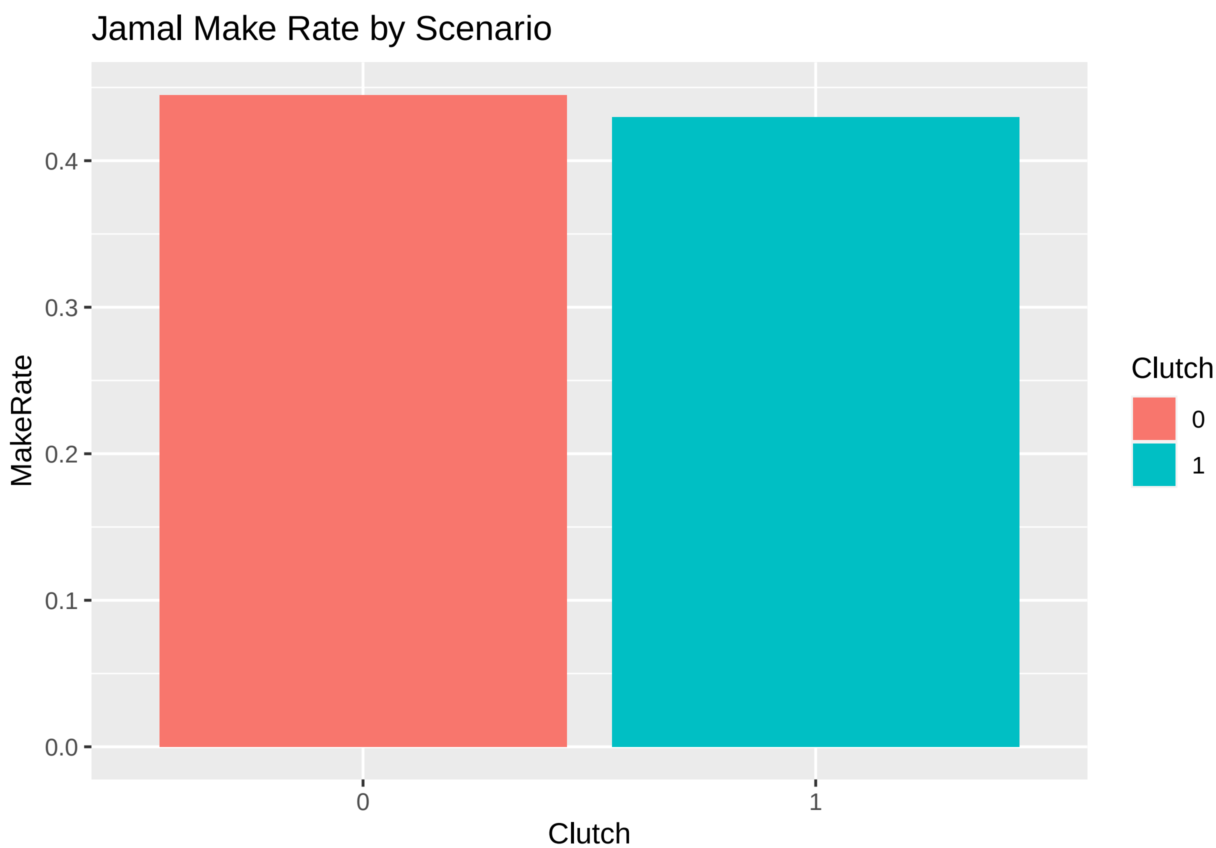
```r
    filter(Shooter == "J. Tatum - tatumja01")
love_shots_playoff <- all_shots %>%
    filter(Shooter == "K. Love - loveke01")

# visualize make rate for each player by scenario
draymond_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Draymond Make Rate by Scenario")
```



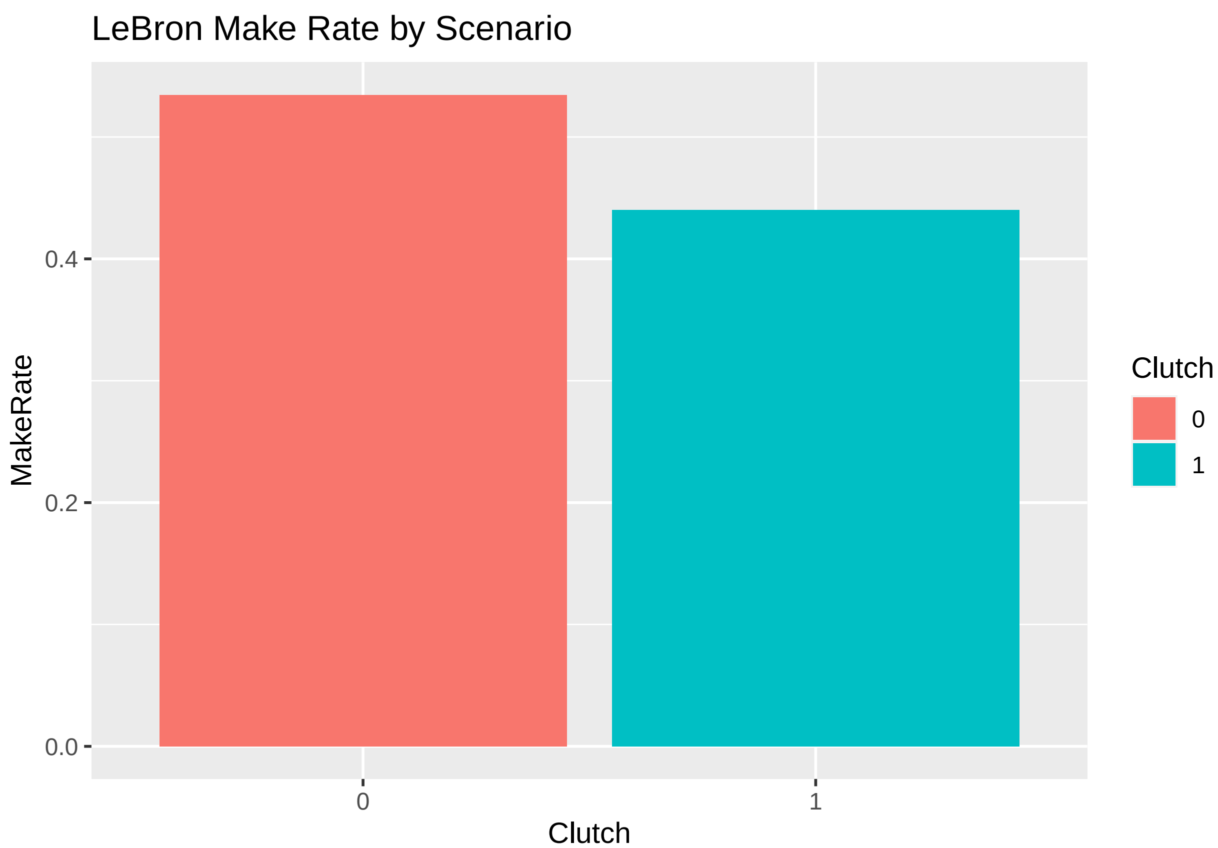

```r
giannis_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Giannis Make Rate by Scenario")
```



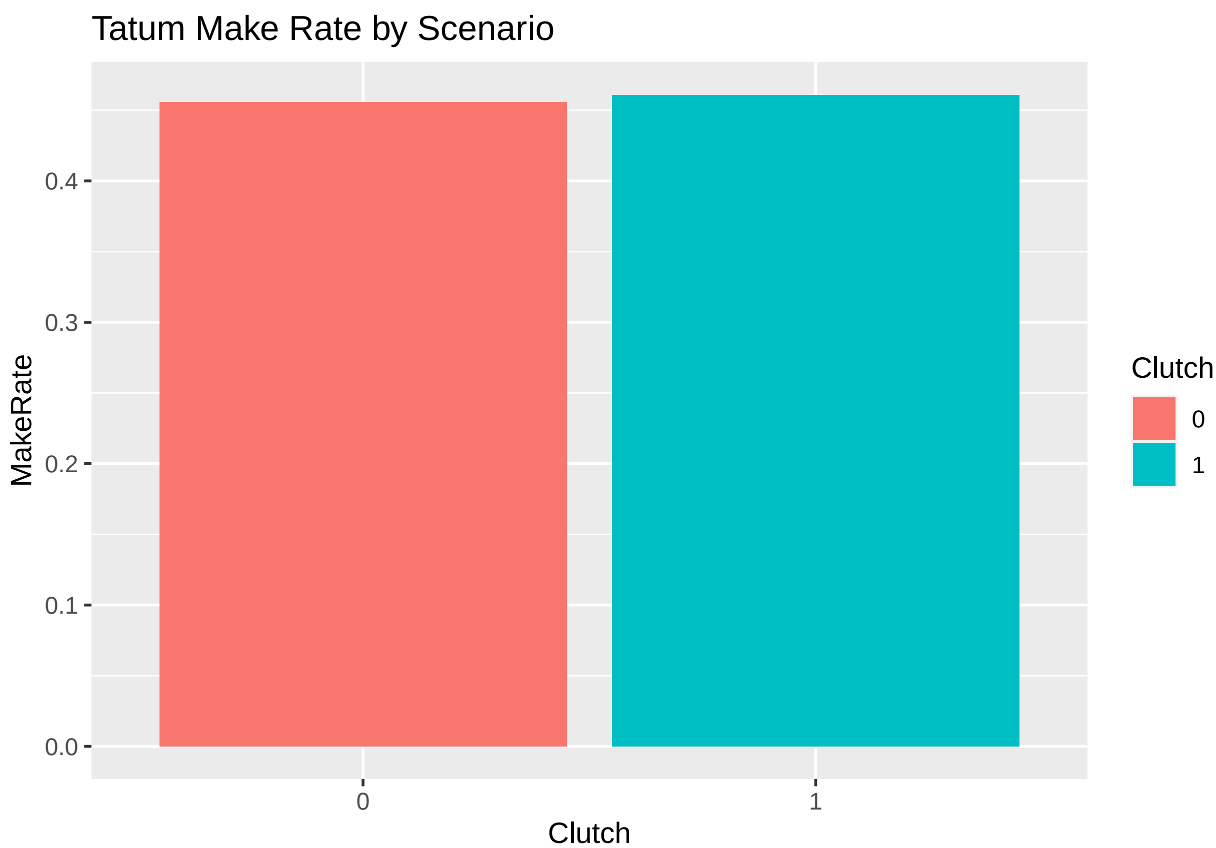

```
jamal_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Jamal Make Rate by Scenario")
```



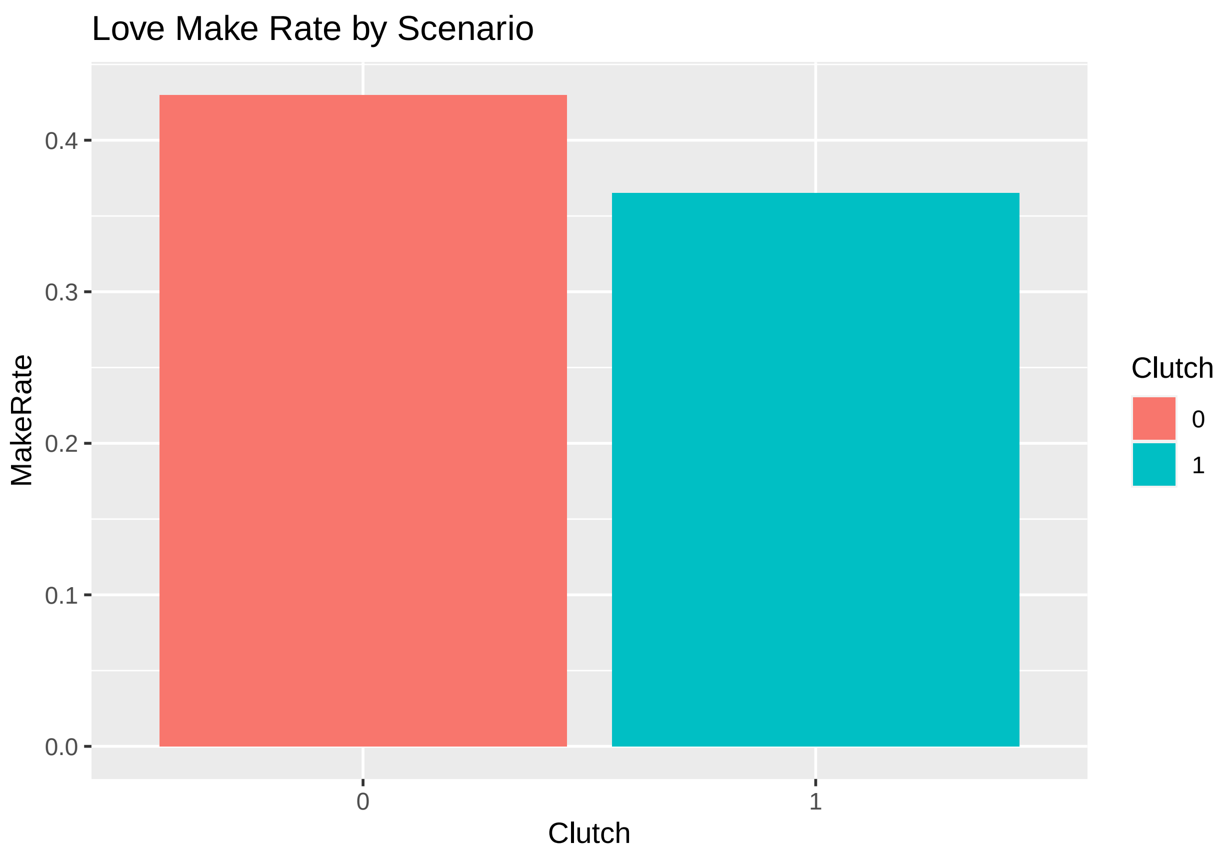

```
lebron_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("LeBron Make Rate by Scenario")
```




```
tatum_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Tatum Make Rate by Scenario")
```




```
love_shots_playoff %>%
    group_by(Clutch) %>%
    summarize(MakeRate = mean(ShotMade)) %>%
    ggplot(aes(x = Clutch, y = MakeRate, fill = Clutch)) + geom_col() +
    ggtitle("Love Make Rate by Scenario")
```




**Logistic Regression for Significance**

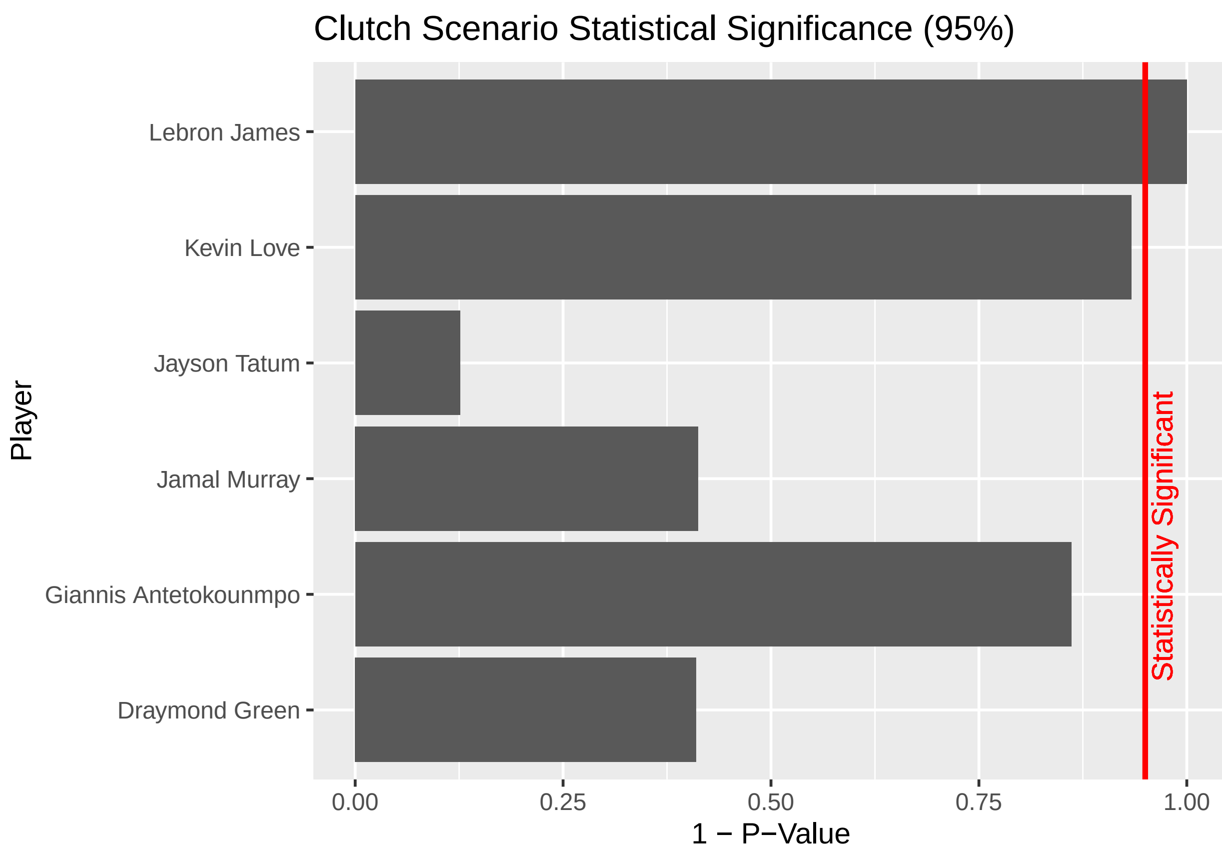
```
# define function to run model and output, since we're
# running this six times
clutch_sig <- function(shot_data) {

    model <- glm(ShotMade ~ Clutch, family = "binomial", data = shot_data)
    fit <- pR2(model)
    pR2 <- fit[4]
    model_tidy <- model %>%
        tidy()
    p_value <- model_tidy[2, 5]
    metrics <- data.frame(p_value, fit[4])
    rownames(metrics) <- NULL
    metrics

}

# run logistic model for each player shot data
draymond_results <- data.frame("Draymond Green", clutch_sig(draymond_shots_playoff)) %>%
    rename(Player = X.Draymond.Green., P_Value = p.value, Pseudo_R2 = fit.4.)
giannis_results <- data.frame("Giannis Antetokounmpo", clutch_sig(giannis_shots_playoff)) %>%
    rename(Player = X.Giannis.Antetokounmpo., P_Value = p.value,
        Pseudo_R2 = fit.4.)
```

```r
jamal_results <- data.frame("Jamal Murray", clutch_sig(jamal_shots_playoff)) %>%
    rename(Player = X.Jamal.Murray., P_Value = p.value, Pseudo_R2 = fit.4.)
lebron_results <- data.frame("Lebron James", clutch_sig(lebron_shots_playoff)) %>%
    rename(Player = X.Lebron.James., P_Value = p.value, Pseudo_R2 = fit.4.)
tatum_results <- data.frame("Jayson Tatum", clutch_sig(tatum_shots_playoff)) %>%
    rename(Player = X.Jayson.Tatum., P_Value = p.value, Pseudo_R2 = fit.4.)
love_results <- data.frame("Kevin Love", clutch_sig(love_shots_playoff)) %>%
    rename(Player = X.Kevin.Love., P_Value = p.value, Pseudo_R2 = fit.4.)
```

```r
# combined dataframes
playoff_results <- rbind(draymond_results, giannis_results, jamal_results,
    lebron_results, tatum_results, love_results)
playoff_results
```

```
##                   Player      P_Value    Pseudo_R2
## 1        Draymond Green 5.897663e-01 5.669573e-05
## 2 Giannis Antetokounmpo 1.388574e-01 2.262029e-04
## 3          Jamal Murray 5.876669e-01 4.579904e-05
## 4          Lebron James 1.697510e-06 1.919801e-03
## 5          Jayson Tatum 8.738938e-01 4.546461e-06
## 6            Kevin Love 6.677920e-02 5.791034e-04
```

```r
playoff_results %>%
    mutate(oneminusp = 1 - P_Value) %>%
    ggplot(aes(x = oneminusp, y = Player)) + geom_col() + geom_vline(xintercept = 0.95,
    color = "red", size = 1) + geom_text(aes(x = 0.97, y = 2.5,
    label = "Statistically Significant"), color = "red", angle = 90) +
    ggtitle("Clutch Scenario Statistical Significance (95%)") +
    xlab("1 - P-Value")
```




### Results

Our playoff game results are not different than the overall results. The only statistically significant difference at a 0.05 confidence level was Lebron James, but even so our Pseudo R2 values are all very close to 0, indicating that the models aren't a great fit. Based on these results and the previous results, we can conclude that the idea of being clutch purely based on end of game win/lose scenarios, is not statistically significant, at least, not in a vacuum. There are likely many other additional factors that affect player performance at end of game - so we can't conclude clutch scenarios are not a factor, but we can say that our definition of clutch scnarios are generally not statistically significant by themselves to predict player performance.

## Conclusion

This analysis assessed player performance in clutch vs. non-clutch scenarios, both for games as a whole, as well as specifically for playoff games. Using only this binary dichotomy of scenarios, we don't have statistical evidence that players perform differently when the pressure is on (at least, for the top players by number of shots taken).

We did see some interesting variation for particular players during clutch scenarios, where players field goal make rates were noticeably different. Ultimately these variations compared to non-clutch scenarios were not statistically significant in most cases, and in all cases, using only that one clutch-scenario variable was not a good fit as a predictor of performance.

To sum up, we do not have evidence of the clutch gene, nor do we have evidence of anti-clutch gene. To further this analysis, we'd need more data around factors going in to clutch time shots. We do have data around type of shot taken and distance of the shot, but one big piece we are missing is the defender of the shot. As defense is typically more locked in at end-of-game scenarios, this can be a huge factor in assessing differences that we do not have data to account for.