

Playoff1

Abisai Lujan

2025-02-23

Read in Play Off Box Scores 2010-2024 dataset First I will import and examin the starting data that Professor Mario provided.

```
play_off_box_scores_2010_2024 <- read.csv("../Starting Data/play_off_box_scores_2010_2024.csv")
head(play_off_box_scores_2010_2024)
```

```
##   season_year game_date gameId   teamId teamCity teamName teamTricode
## 1    2011-12 2012-06-03 41100304 1610612748    Miami    Heat        MIA
## 2    2011-12 2012-06-09 41100307 1610612748    Miami    Heat        MIA
## 3    2010-11 2011-06-05 41000403 1610612748    Miami    Heat        MIA
## 4    2011-12 2012-04-28 41100111 1610612748    Miami    Heat        MIA
## 5    2011-12 2012-05-15 41100202 1610612748    Miami    Heat        MIA
## 6    2010-11 2011-05-31 41000401 1610612748    Miami    Heat        MIA
##   teamSlug personId  personName position      comment jerseyNum
## 1    heat      436 Juwan Howard           DNP - Coach's Decision      NA
## 2    heat      436 Juwan Howard                               NA
## 3    heat      436 Juwan Howard                               NA
## 4    heat      436 Juwan Howard                               NA
## 5    heat      436 Juwan Howard           DNP - Coach's Decision      NA
## 6    heat      436 Juwan Howard                               NA
##   minutes fieldGoalsMade fieldGoalsAttempted fieldGoalsPercentage
## 1              0              0              0
## 2    0:28              0              0              0
## 3    6:25              0              0              0
## 4    5:12              0              2              0
## 5              0              0              0
## 6    7:37              0              1              0
##   threePointersMade threePointersAttempted threePointersPercentage
## 1              0              0              0
## 2              0              0              0
## 3              0              0              0
## 4              0              0              0
## 5              0              0              0
## 6              0              0              0
##   freeThrowsMade freeThrowsAttempted freeThrowsPercentage reboundsOffensive
## 1              0              0              0.0              0
## 2              0              0              0.0              0
## 3              1              2              0.5              0
## 4              0              0              0.0              0
## 5              0              0              0.0              0
## 6              2              2              1.0              3
```

```
##      reboundsDefensive reboundsTotal assists steals blocks turnovers foulsPersonal
## 1           0           0           0           0           0           0           0
## 2           0           0           0           0           0           0           0
## 3           1           1           1           0           0           0           1
## 4           0           0           0           0           0           0           0
## 5           0           0           0           0           0           0           0
## 6           0           3           0           0           0           0           0
##      points plusMinusPoints
## 1           0           0
## 2           0           0
## 3           1           6
## 4           0           3
## 5           0           0
## 6           2           2
```

Check the dimensions of the dataset

```
dim(play_off_box_scores_2010_2024)
```

```
## [1] 31185    33
```

There are 31,185 rows and 33 variables

Lets explore the data variables and types

```
str(play_off_box_scores_2010_2024)
```

```
## 'data.frame':    31185 obs. of  33 variables:
## $ season_year      : chr  "2011-12" "2011-12" "2010-11" "2011-12" ...
## $ game_date        : chr  "2012-06-03" "2012-06-09" "2011-06-05" "2012-04-28" ...
## $ gameId           : int   41100304 41100307 41000403 41100111 41100202 41000401 41000213 41000401 ...
## $ teamId           : int   1610612748 1610612748 1610612748 1610612748 1610612748 1610612748 1610612748 1610612748 ...
## $ teamCity         : chr  "Miami" "Miami" "Miami" "Miami" ...
## $ teamName         : chr  "Heat" "Heat" "Heat" "Heat" ...
## $ teamTricode      : chr  "MIA" "MIA" "MIA" "MIA" ...
## $ teamSlug         : chr  "heat" "heat" "heat" "heat" ...
## $ personId         : int   436 436 436 436 436 436 436 436 436 436 ...
## $ personName       : chr  "Juwan Howard" "Juwan Howard" "Juwan Howard" "Juwan Howard" ...
## $ position         : chr  "" "" "" "" ...
## $ comment          : chr  "DNP - Coach's Decision" "" "" "" ...
## $ jerseyNum        : logi   NA NA NA NA NA NA ...
## $ minutes          : chr  "" "0:28" "6:25" "5:12" ...
## $ fieldGoalsMade    : int   0 0 0 0 0 0 0 0 0 0 ...
## $ fieldGoalsAttempted : int   0 0 0 2 0 1 0 0 0 0 ...
## $ fieldGoalsPercentage : num  0 0 0 0 0 0 0 0 0 0 ...
## $ threePointersMade : int   0 0 0 0 0 0 0 0 0 0 ...
## $ threePointersAttempted : int   0 0 0 0 0 0 0 0 0 0 ...
## $ threePointersPercentage : num  0 0 0 0 0 0 0 0 0 0 ...
## $ freeThrowsMade    : int   0 0 1 0 0 2 0 0 0 0 ...
## $ freeThrowsAttempted : int   0 0 2 0 0 2 0 0 0 0 ...
## $ freeThrowsPercentage : num  0 0 0.5 0 0 1 0 0 0 0 ...
## $ reboundsOffensive : int   0 0 0 0 0 3 0 0 0 0 ...
## $ reboundsDefensive : int   0 0 1 0 0 0 0 0 0 0 ...
```

```
## $ reboundsTotal      : int  0 0 1 0 0 3 0 0 0 0 ...
## $ assists            : int  0 0 1 0 0 0 0 0 0 0 ...
## $ steals             : int  0 0 0 0 0 0 0 0 0 0 ...
## $ blocks             : int  0 0 0 0 0 0 0 0 0 0 ...
## $ turnovers          : int  0 0 0 0 0 0 0 0 0 0 ...
## $ foulsPersonal      : int  0 0 1 0 0 0 0 0 0 0 ...
## $ points             : int  0 0 1 0 0 2 0 0 0 0 ...
## $ plusMinusPoints    : int  0 0 6 3 0 2 0 0 0 0 ...
```

Are there any missing values? Check for NA values

```
sum(is.na(play_off_box_scores_2010_2024))
```

```
## [1] 31185
```

There are 31,185 missing values

Where are the NA values located

```
colSums((is.na(play_off_box_scores_2010_2024)))
```

```
##          season_year          game_date          gameId
##              0              0              0
##          teamId          teamCity          teamName
##              0              0              0
##          teamTricode          teamSlug          personId
##              0              0              0
##          personName          position          comment
##              0              0              0
##          jerseyNum          minutes          fieldGoalsMade
##          31185              0              0
##          fieldGoalsAttempted          fieldGoalsPercentage          threePointersMade
##              0              0              0
##          threePointersAttempted          threePointersPercentage          freeThrowsMade
##              0              0              0
##          freeThrowsAttempted          freeThrowsPercentage          reboundsOffensive
##              0              0              0
##          reboundsDefensive          reboundsTotal          assists
##              0              0              0
##          steals          blocks          turnovers
##              0              0              0
##          foulsPersonal          points          plusMinusPoints
##              0              0              0
```

We can see that the only column with missing values is jerseyNum. It has just as many rows of NA values as the total amount of rows in the dataset. However, lets make sure whether or not the column *only* has NA values.

```
play_off_box_scores_2010_2024 %>%
  select(jerseyNum) %>%
  unique()
```

```
## jerseyNum
## 1      NA
```

In fact, the jerseyNum column only holds NA values.

Some Data Cleaning I will remove the jerseyNum column as it holds no additional value than the playerId column does

```
v2_play_off_box_scores_2010_2024 <- play_off_box_scores_2010_2024 %>%
  select(-jerseyNum)
head(v2_play_off_box_scores_2010_2024)
```

```
## season_year game_date gameId teamId teamCity teamName teamTricode
## 1 2011-12 2012-06-03 41100304 1610612748 Miami Heat MIA
## 2 2011-12 2012-06-09 41100307 1610612748 Miami Heat MIA
## 3 2010-11 2011-06-05 41000403 1610612748 Miami Heat MIA
## 4 2011-12 2012-04-28 41100111 1610612748 Miami Heat MIA
## 5 2011-12 2012-05-15 41100202 1610612748 Miami Heat MIA
## 6 2010-11 2011-05-31 41000401 1610612748 Miami Heat MIA
## teamSlug personId personName position comment minutes
## 1 heat 436 Juwan Howard DNP - Coach's Decision
## 2 heat 436 Juwan Howard 0:28
## 3 heat 436 Juwan Howard 6:25
## 4 heat 436 Juwan Howard 5:12
## 5 heat 436 Juwan Howard DNP - Coach's Decision
## 6 heat 436 Juwan Howard 7:37
## fieldGoalsMade fieldGoalsAttempted fieldGoalsPercentage threePointersMade
## 1 0 0 0 0
## 2 0 0 0 0
## 3 0 0 0 0
## 4 0 2 0 0
## 5 0 0 0 0
## 6 0 1 0 0
## threePointersAttempted threePointersPercentage freeThrowsMade
## 1 0 0 0
## 2 0 0 0
## 3 0 0 1
## 4 0 0 0
## 5 0 0 0
## 6 0 0 2
## freeThrowsAttempted freeThrowsPercentage reboundsOffensive reboundsDefensive
## 1 0 0.0 0 0
## 2 0 0.0 0 0
## 3 2 0.5 0 1
## 4 0 0.0 0 0
## 5 0 0.0 0 0
## 6 2 1.0 3 0
## reboundsTotal assists steals blocks turnovers foulsPersonal points
## 1 0 0 0 0 0 0 0
## 2 0 0 0 0 0 0 0
## 3 1 1 0 0 0 1 1
## 4 0 0 0 0 0 0 0
## 5 0 0 0 0 0 0 0
## 6 3 0 0 0 0 0 2
```

```
##      plusMinusPoints
## 1              0
## 2              0
## 3              6
## 4              3
## 5              0
## 6              2
```

To confirm, let's observe which variables from the original dataset are not in the new dataset

```
names_v1 <- names(play_off_box_scores_2010_2024)
names_v2 <- names(v2_play_off_box_scores_2010_2024)

setdiff(names_v1, names_v2)
```

```
## [1] "jerseyNum"
```

For filtering purposes, let's change date variables (`season_year` and `game_date`) to numeric types

```
pobs_2010_2024_num <- v2_play_off_box_scores_2010_2024 %>%
  #remove the dashes and convert to numeric
  mutate(season_year = as.numeric(gsub("-", "", season_year)),
         game_date = as.numeric(gsub("-", "", game_date))
  )
pobs_2010_2024_num %>%
  select(season_year,
         game_date) %>%
  str()
```

```
## 'data.frame':   31185 obs. of  2 variables:
## $ season_year: num  201112 201112 201011 201112 201112 ...
## $ game_date : num  20120603 20120609 20110605 20120428 20120515 ...
```

Read in Play off Totals 2010-2024

```
play_off_totals_2010_2024 <- read.csv("../Starting Data/play_off_totals_2010_2024.csv")
head(play_off_totals_2010_2024)
```

##	SEASON_YEAR	TEAM_ID	TEAM_ABBREVIATION	TEAM_NAME	GAME_ID									
## 1	2010-11	1610612737		ATL Atlanta Hawks	41000205									
## 2	2010-11	1610612737		ATL Atlanta Hawks	41000201									
## 3	2010-11	1610612737		ATL Atlanta Hawks	41000202									
## 4	2010-11	1610612737		ATL Atlanta Hawks	41000132									
## 5	2010-11	1610612737		ATL Atlanta Hawks	41000135									
## 6	2010-11	1610612737		ATL Atlanta Hawks	41000131									
##	GAME_DATE	MATCHUP	WL	MIN	FGM	FGA	FG_PCT	FG3M	FG3A	FG3_PCT	FTM	FTA		
## 1	2011-05-10T00:00:00	ATL @ CHI	L	48	34	71	0.479	1	12	0.083	14	17		
## 2	2011-05-02T00:00:00	ATL @ CHI	W	48	40	78	0.513	7	13	0.538	16	20		
## 3	2011-05-04T00:00:00	ATL @ CHI	L	48	26	77	0.338	3	13	0.231	18	24		
## 4	2011-04-19T00:00:00	ATL @ ORL	L	48	32	81	0.395	7	20	0.350	11	17		
## 5	2011-04-26T00:00:00	ATL @ ORL	L	48	25	69	0.362	4	16	0.250	22	32		
## 6	2011-04-16T00:00:00	ATL @ ORL	W	48	38	74	0.514	6	14	0.429	21	29		
##	FT_PCT	OREB	DREB	REB	AST	TOV	STL	BLK	BLKA	PF	PFD	PTS	PLUS_MINUS	GP_RANK
## 1	0.824	8	25	33	20	11	2	5	3	21	16	83	-12	1
## 2	0.800	9	29	38	20	10	8	8	7	17	16	103	8	1
## 3	0.750	10	29	39	14	12	10	10	5	15	17	73	-13	1
## 4	0.647	13	26	39	15	15	7	7	2	27	21	82	-6	1
## 5	0.688	7	36	43	15	13	3	3	4	22	25	76	-25	1
## 6	0.724	5	24	29	19	10	8	1	3	28	24	103	10	1
##	W_RANK	L_RANK	W_PCT_RANK	MIN_RANK	FGM_RANK	FGA_RANK	FG_PCT_RANK	FG3M_RANK						
## 1	82	82		82	13	71	135	37						157
## 2	1	1		1	13	11	69	13						49
## 3	82	82		82	13	159	84	162						130
## 4	82	82		82	13	100	42	128						49
## 5	82	82		82	13	161	143	153						112
## 6	1	1		1	13	27	108	12						68
##	FG3A_RANK	FG3_PCT_RANK	FTM_RANK	FTA_RANK	FT_PCT_RANK	OREB_RANK	DREB_RANK							
## 1	138		161	126	142	39	120	134						
## 2	127		10	112	119	63	98	76						
## 3	127		125	89	86	100	79	76						
## 4	50		81	152	142	148	41	127						
## 5	92		115	49	31	133	129	15						
## 6	113		37	59	47	121	149	142						
##	REB_RANK	AST_RANK	TOV_RANK	STL_RANK	BLK_RANK	BLKA_RANK	PF_RANK	PFD_RANK						
## 1	143	43	32	159	70	18	68	142						
## 2	99	43	16	35	21	109	22	142						
## 3	84	137	51	15	3	66	8	129						
## 4	84	123	107	67	38	6	144	85						
## 5	50	123	63	153	122	42	79	28						
## 6	159	62	16	35	158	18	152	40						
##	PTS_RANK	PLUS_MINUS_RANK	AVAILABLE_FLAG											
## 1	140		139		NA									
## 2	24		39		NA									
## 3	159		145		NA									
## 4	144		106		NA									
## 5	157		159		NA									
## 6	24		28		NA									

Check the dimensions of the dataset

```
dim(play_off_totals_2010_2024)
```

```
## [1] 2362 57
```

There are 2,362 rows and 57 variables.

Lets explore the data variables and types

```
str(play_off_totals_2010_2024)
```

```
## 'data.frame': 2362 obs. of 57 variables:
## $ SEASON_YEAR : chr "2010-11" "2010-11" "2010-11" "2010-11" ...
## $ TEAM_ID : int 1610612737 1610612737 1610612737 1610612737 1610612737 1610612737 1610612737 1610612737 ...
## $ TEAM_ABBREVIATION: chr "ATL" "ATL" "ATL" "ATL" ...
## $ TEAM_NAME : chr "Atlanta Hawks" "Atlanta Hawks" "Atlanta Hawks" "Atlanta Hawks" "Atlanta Hawks" ...
## $ GAME_ID : int 41000205 41000201 41000202 41000132 41000135 41000131 41000215 41000212 41000214 ...
## $ GAME_DATE : chr "2011-05-10T00:00:00" "2011-05-02T00:00:00" "2011-05-04T00:00:00" "2011-05-05T00:00:00" ...
## $ MATCHUP : chr "ATL @ CHI" "ATL @ CHI" "ATL @ CHI" "ATL @ ORL" ...
## $ WL : chr "L" "W" "L" "L" ...
## $ MIN : num 48 48 48 48 48 48 48 48 48 48 ...
## $ FGM : int 34 40 26 32 25 38 33 34 32 40 ...
## $ FGA : int 71 78 77 81 69 74 67 79 75 81 ...
## $ FG_PCT : num 0.479 0.513 0.338 0.395 0.362 0.514 0.493 0.43 0.427 0.494 ...
## $ FG3M : int 1 7 3 7 4 6 7 6 12 3 ...
## $ FG3A : int 12 13 13 20 16 14 15 11 24 12 ...
## $ FG3_PCT : num 0.083 0.538 0.231 0.35 0.25 0.429 0.467 0.545 0.5 0.25 ...
## $ FTM : int 14 16 18 11 22 21 14 17 14 18 ...
## $ FTA : int 17 20 24 17 32 29 20 22 18 24 ...
## $ FT_PCT : num 0.824 0.8 0.75 0.647 0.688 0.724 0.7 0.773 0.778 0.75 ...
## $ OREB : int 8 9 10 13 7 5 4 10 7 9 ...
## $ DREB : int 25 29 29 26 36 24 26 28 32 44 ...
## $ REB : int 33 38 39 39 43 29 30 38 39 53 ...
## $ AST : int 20 20 14 15 15 19 15 18 18 24 ...
## $ TOV : num 11 10 12 15 13 10 17 12 14 18 ...
## $ STL : int 2 8 10 7 3 8 5 7 7 6 ...
## $ BLK : int 5 8 10 7 3 1 5 2 3 6 ...
## $ BLKA : int 3 7 5 2 4 3 3 9 8 6 ...
## $ PF : int 21 17 15 27 22 28 25 24 24 23 ...
## $ PFD : int 16 16 17 21 25 24 19 22 21 25 ...
## $ PTS : int 83 103 73 82 76 103 87 91 90 101 ...
## $ PLUS_MINUS : num -12 8 -13 -6 -25 10 -10 -11 -9 12 ...
## $ GP_RANK : int 1 1 1 1 1 1 1 1 1 1 ...
## $ W_RANK : int 82 1 82 82 82 1 82 82 82 1 ...
## $ L_RANK : int 82 1 82 82 82 1 82 82 82 1 ...
## $ W_PCT_RANK : int 82 1 82 82 82 1 82 82 82 1 ...
## $ MIN_RANK : int 13 13 13 13 13 13 13 13 13 13 ...
## $ FGM_RANK : int 71 11 159 100 161 27 90 71 100 11 ...
## $ FGA_RANK : int 135 69 84 42 143 108 153 55 97 42 ...
## $ FG_PCT_RANK : int 37 13 162 128 153 12 27 92 97 22 ...
## $ FG3M_RANK : int 157 49 130 49 112 68 49 68 5 130 ...
## $ FG3A_RANK : int 138 127 127 50 92 113 106 143 18 138 ...
```

```
## $ FG3_PCT_RANK      : int  161 10 125 81 115 37 21 8 12 115 ...
## $ FTM_RANK          : int  126 112 89 152 49 59 126 98 126 89 ...
## $ FTA_RANK          : int  142 119 86 142 31 47 119 102 134 86 ...
## $ FT_PCT_RANK       : int   39 63 100 148 133 121 129 90 84 100 ...
## $ OREB_RANK         : int  120 98 79 41 129 149 157 79 129 98 ...
## $ DREB_RANK         : int  134 76 76 127 15 142 127 97 41 1 ...
## $ REB_RANK          : int  143 99 84 84 50 159 156 99 84 8 ...
## $ AST_RANK          : int   43 43 137 123 123 62 123 80 80 10 ...
## $ TOV_RANK          : int   32 16 51 107 63 16 134 51 87 141 ...
## $ STL_RANK          : int  159 35 15 67 153 35 125 67 67 101 ...
## $ BLK_RANK          : int   70 21 3 38 122 158 70 146 122 55 ...
## $ BLKA_RANK         : int   18 109 66 6 42 18 18 143 126 94 ...
## $ PF_RANK           : int   68 22 8 144 79 152 124 108 108 94 ...
## $ PFD_RANK          : int  142 142 129 85 28 40 110 70 85 28 ...
## $ PTS_RANK          : int  140 24 159 144 157 24 118 94 98 33 ...
## $ PLUS_MINUS_RANK   : int  139 39 145 106 159 28 131 136 125 19 ...
## $ AVAILABLE_FLAG    : num  NA NA NA NA NA NA NA NA NA NA ...
```

Are there any missing values? Check for NA values

```
sum(is.na(play_off_totals_2010_2024))
```

```
## [1] 336
```

Where are the NA values located

```
colSums((is.na(play_off_totals_2010_2024)))
```

```
##      SEASON_YEAR      TEAM_ID TEAM_ABBREVIATION      TEAM_NAME
##           0           0           0           0
##      GAME_ID      GAME_DATE      MATCHUP      WL
##           0           0           0           0
##           MIN           FGM           FGA      FG_PCT
##           0           0           0           0
##           FG3M          FG3A      FG3_PCT      FTM
##           0           0           0           0
##           FTA           FT_PCT          OREB      DREB
##           0           0           0           0
##           REB           AST           TOV      STL
##           0           0           0           0
##           BLK           BLKA           PF      PFD
##           0           0           0           0
##           PTS      PLUS_MINUS      GP_RANK      W_RANK
##           0           0           0           0
##           L_RANK      W_PCT_RANK      MIN_RANK      FGM_RANK
##           0           0           0           0
##           FGA_RANK      FG_PCT_RANK      FG3M_RANK      FG3A_RANK
##           0           0           0           0
##           FG3_PCT_RANK      FTM_RANK      FTA_RANK      FT_PCT_RANK
##           0           0           0           0
##           OREB_RANK      DREB_RANK      REB_RANK      AST_RANK
##           0           0           0           0
```



```
##          TOV_RANK          STL_RANK          BLK_RANK          BLKA_RANK
##              0              0              0              0
##          PF_RANK          PFD_RANK          PTS_RANK    PLUS_MINUS_RANK
##              0              0              0              0
##    AVAILABLE_FLAG
##              336
```

Only AVAILABLE_FLAG has NA values. Does this column *only* have NA values?

```
play_off_totals_2010_2024 %>%
  select(AVAILABLE_FLAG) %>%
  unique()
```

```
##    AVAILABLE_FLAG
## 1              NA
## 331             1
## 338             0
## 754             2
```

AVAILABLE_FLAG variable represents whether or not the data for this row is available. Im assuming that the NA values are unresolved, 1 represents only one opponent has stats in `play_off_box_scores_2010_2024`, and 2 means that both opponents have stats in `play_off_box_scores_2010_2024`

Lets see how many rows in AVAILABLE_FLAG have NA values

```
play_off_totals_2010_2024 %>%
  filter(is.na(AVAILABLE_FLAG)) %>%
  nrow()
```

```
## [1] 336
```

How many rows in AVAILABLE_FLAG have a value of 0

```
play_off_totals_2010_2024 %>%
  filter(AVAILABLE_FLAG==0) %>%
  nrow()
```

```
## [1] 30
```

How many rows in AVAILABLE_FLAG have a value of 1

```
play_off_totals_2010_2024 %>%
  filter(AVAILABLE_FLAG==1) %>%
  nrow()
```

```
## [1] 1988
```

How many rows in AVAILABLE_FLAG have a value of 2

```
play_off_totals_2010_2024 %>%  
  filter(AVAILABLE_FLAG==2) %>%  
  nrow()
```

```
## [1] 8
```

Some Data Cleaning for Playoff Totals 2010-2024

Read in Regular Season Box Scores 2010-2024 part 1 Regular Season Box regular_season_box_scores_2010_2024_pa

```
regular_season_box_scores_2010_2024_part_1 <- read.csv("../Starting Data/regular_season_box_scores_2010_2024_part_1.csv")
head(regular_season_box_scores_2010_2024_part_1)
```

##	season_year	game_date	gameId	matchup	teamId	teamCity	teamName		
## 1	2010-11	2010-11-10	21000112	NJN @ CLE	1610612751	New Jersey	Nets		
## 2	2010-11	2010-11-17	21000165	NJN @ UTA	1610612751	New Jersey	Nets		
## 3	2010-11	2010-11-27	21000237	NJN @ PHI	1610612751	New Jersey	Nets		
## 4	2010-11	2010-12-12	21000351	NJN vs. LAL	1610612751	New Jersey	Nets		
## 5	2010-11	2010-10-29	21000020	NJN vs. SAC	1610612751	New Jersey	Nets		
## 6	2010-11	2010-12-14	21000361	NJN vs. PHI	1610612751	New Jersey	Nets		
##	teamTricode	teamSlug	personId	personName	position	comment			
## 1	NJN	nets	693	Joe Smith					
## 2	NJN	nets	693	Joe Smith		DNP - Coach's Decision			
## 3	NJN	nets	693	Joe Smith		DNP - Coach's Decision			
## 4	NJN	nets	693	Joe Smith		DNP - Coach's Decision			
## 5	NJN	nets	693	Joe Smith	F				
## 6	NJN	nets	693	Joe Smith		DNP - Coach's Decision			
##	jerseyNum	minutes	fieldGoalsMade	fieldGoalsAttempted	fieldGoalsPercentage				
## 1	NA	4:21	0	0	0				
## 2	NA		0	0	0				
## 3	NA		0	0	0				
## 4	NA		0	0	0				
## 5	NA	9:53	1	1	1				
## 6	NA		0	0	0				
##	threePointersMade	threePointersAttempted	threePointersPercentage						
## 1	0	0	0						
## 2	0	0	0						
## 3	0	0	0						
## 4	0	0	0						
## 5	0	0	0						
## 6	0	0	0						
##	freeThrowsMade	freeThrowsAttempted	freeThrowsPercentage	reboundsOffensive					
## 1	0	0	0	0					
## 2	0	0	0	0					
## 3	0	0	0	0					
## 4	0	0	0	0					
## 5	0	0	0	0					
## 6	0	0	0	0					
##	reboundsDefensive	reboundsTotal	assists	steals	blocks	turnovers	foulsPersonal		
## 1	0	0	0	0	0	0	0		
## 2	0	0	0	0	0	0	0		
## 3	0	0	0	0	0	0	0		
## 4	0	0	0	0	0	0	0		
## 5	1	1	1	0	0	1	1		
## 6	0	0	0	0	0	0	0		
##	points	plusMinusPoints							
## 1	0	-4							
## 2	0	0							
## 3	0	0							
## 4	0	0							
## 5	2	6							
## 6	0	0							

Random Code:

```
play_off_box_scores_2010_2024 %>%  
  filter(gameId==41100111)
```

Lets look at an arbitrary game, the Miami Heat vs the Boston Celtics on 6/3/12 (gameId=41100304)

```
play_off_box_scores_2010_2024 %>%  
  filter(game_date==2012)
```

For the Miami Heat vs Boston Celtics game on 6/3/12, (gameId=41100304), there are 26 rows.

```
play_off_box_scores_2010_2024 %>%  
  filter(gameId == 41100304) %>%  
  select(teamId, teamCity, teamName) %>%  
  unique()
```

There are 2 teams involved which are the Miami Heat(1610612748) and the Boston Celtics(1610612738)

For the Miami Heat vs Boston Celtics game on 6/3/12, Look at same game rows but specifically for Miami Heat. First I will filter the data for team Miami Heat and extract its teamId. I will assign Miami Heat's teamId to its own variable

```
mia_heat_id <- play_off_box_scores_2010_2024 %>%  
  filter(teamCity=="Miami" & teamName=="Heat") %>%  
  select(teamId) %>%  
  unique() %>%  
  #extracts the [1,1]th element  
  .[[1,1]]
```

For the Miami Heat vs Boston Celtics game on 6/3/12, Lets see how many rows Miami Heat has for this specific game

```
play_off_box_scores_2010_2024 %>%  
  filter(gameId == 41100304 & teamId==mia_heat_id)
```

For the Miami Heat vs Boston Celtics game on 6/3/12, There are 13 rows for Miami heat.

For the Miami Heat vs Boston Celtics game on 6/3/12, Lets see if there are 13 rows for each person on the Miami Heat team

```
play_off_box_scores_2010_2024 %>%  
  filter(gameId == 41100304 & teamId==mia_heat_id) %>%  
  select(personId) %>%  
  unique()
```

Looks like all 13 separate Miami Heat rows for this game represent each player's box score statistics

For the Miami Heat vs Boston Celtics game on 6/3/12, Lets look at the rows for Boston Celtics, how many rows there are for the team, and if they represent the 13 players on the team.

```
#First I will create a variable for Boston Celtics' teamId
bos_cel_id <- play_off_box_scores_2010_2024 %>%
  filter(teamCity=="Boston" & teamName=="Celtics") %>%
  select(teamId) %>%
  unique() %>%
  #extract [1,1]th element
  .[[1,1]]

#Next I will filter this game's rows for the Boston Celtics
play_off_box_scores_2010_2024 %>%
  filter(gameId==41100304 & teamId==bos_cel_id)
```