# Math E-23C Term Project

An analysis of the Lakers 2019 Championship season

Maria Cristina An, Daniel Lebedinsky, Julio Solis Arce

04 May, 2022

**Abstract**

The current documents offer a comprehensive analysis of the Lakers 2019 season, in which they successfully defended the NBA championship

## Contents

# 1 Dataset description

## 1.1 Codebook

### 1.1.1 Game level variables

- Rk – Rank
- G – Season Game
- W_E – opponent is from East or West
- Opp – Opponent
- W_L – Lakers team won or lost
- Tm – Lakers team Points
- Opp_pts – Opponent Points

### 1.1.2 Lakers relevant variables

- FG_LA – Field Goals
- FGA_LA – Field Goal Attempts
- FGpc_LA – Field Goal Percentage
- X3P_LA – 3-Point Field Goals
- X3PA_LA – 3-Point Field Goal Attempts
- X3Ppc_LA – 3-Point Field Goal Percentage
- FT_LA – Free Throws
- FTA_LA – Free Throw Attempts
- FTpc_LA – Free Throw Percentage
- ORB_LA – Offensive Rebounds
- TRB_LA – Total Rebounds
- AST_LA – Assists
- STL_LA – Steals
- BLK_LA – Blocks
- TOV_LA – Turnovers
- PF_LA – Personal Fouls

### 1.1.3 Opponent relevant variables

- FG – Opponent Field Goals
- FGA – Opponent Field Goal Attempts
- FGpc – Opponent Field Goal Percentage
- X3P – Opponent 3-Point Field Goals
- X3PA – Opponent 3-Point Field Goal Attempts
- X3Ppc – Opponent 3-Point Field Goal Percentage
- FT – Opponent Free Throws

- FTA – Opponent Free Throw Attempts
- FTpc – Opponent Free Throw Percentage
- ORB – Opponent Offensive Rebounds
- TRB – Opponent Total Rebounds
- AST – Opponent Assists
- STL – Opponent Steals
- BLK – Opponent Blocks
- TOV – Opponent Turnovers
- PF – Opponent Personal Fouls

## 1.2 Importing and cleaning data

Reference: https://www.basketball-reference.com/teams/LAL/2020/gamelog/

```
#Import dataset
Lakers <- read.csv("Lakers\ 2019-20\ Game\ log.csv"); head(Lakers)
```

```
##   Rk G       Date W_E Opp W_L  Tm Opp_pts FG_LA FGA_LA FGpc_LA X3P_LA X3PA_LA
## 1  1 1 2019-10-22   W LAC   L 102     112    37     85   0.435     13      33
## 2  2 2 2019-10-25   W UTA   W  95      86    34     86   0.395      8      26
## 3  3 3 2019-10-27   E CHO   W 120     101    49     93   0.527     11      31
## 4  4 4 2019-10-29   W MEM   W 120      91    40     86   0.465      7      26
## 5  5 5 2019-11-01   W DAL   W 119     110    47     96   0.490      9      32
## 6  6 6 2019-11-03   W SAS   W 103      96    42     87   0.483      6      24
##   X3Ppc_LA FT_LA FTA_LA FTpc_LA ORB_LA TRB_LA AST_LA STL_LA BLK_LA TOV_LA PF_LA
## 1    0.394    15     21   0.714      9     41     20      4      7     14    24
## 2    0.308    19     24   0.792     11     42     19     14      9     14    21
## 3    0.355    11     14   0.786     10     47     30      7      8     13    22
## 4    0.269    33     39   0.846     10     54     25      9      8     18    19
## 5    0.281    16     21   0.762      6     41     24     11      7     11    21
## 6    0.250    13     19   0.684      7     46     21      8     10     18    22
##   FG FGA  FGpc X3P X3PA X3Ppc FT FTA  FTpc ORB TRB AST STL BLK TOV PF
## 1 42  81 0.519  11   31 0.355 17  24 0.708  11  45  24   8   5  14 25
## 2 29  70 0.414   8   25 0.320 20  21 0.952   3  40  19   8   2  22 25
## 3 35  91 0.385  10   35 0.286 21  23 0.913  13  45  19   7   4  14 20
## 4 31  95 0.326   7   34 0.206 22  23 0.957  13  46  20  10   3  21 27
## 5 40 102 0.392  14   36 0.389 16  23 0.696  18  61  23   5   1  22 21
## 6 33  83 0.398   6   25 0.240 24  29 0.828   8  47  23  11   4  17 18
```

```
# Convert to proper date format
Lakers$Dates <- as.Date(Lakers$Date)
attach(Lakers)
```

## 1.3 Summary of variables

```
summary(Lakers)
```

```
##        Rk              G             Date                W_E
##  Min.   : 1.00   Min.   : 1.00   Length:92          Length:92
##  1st Qu.:12.00   1st Qu.:12.00   Class :character   Class :character
##  Median :25.50   Median :25.50   Mode  :character   Mode  :character
```

```
##  Mean   :30.29   Mean   :30.29  
##  3rd Qu.:48.25   3rd Qu.:48.25  
##  Max.   :71.00   Max.   :71.00  
##      Opp                W_L                  Tm           Opp_pts     
##  Length:92          Length:92          Min.   : 86.0   Min.   : 80.0  
##  Class :character   Class :character   1st Qu.:104.0   1st Qu.:100.0  
##  Mode  :character   Mode  :character   Median :114.0   Median :108.0  
##                                        Mean   :113.3   Mean   :107.3  
##                                        3rd Qu.:122.0   3rd Qu.:114.0  
##                                        Max.   :142.0   Max.   :139.0  
##      FG_LA           FGA_LA          FGpc_LA           X3P_LA     
##  Min.   :29.00   Min.   : 74.00   Min.   :0.3510   Min.   : 2.00  
##  1st Qu.:38.00   1st Qu.: 83.00   1st Qu.:0.4512   1st Qu.: 9.00  
##  Median :43.00   Median : 87.00   Median :0.4855   Median :11.00  
##  Mean   :42.04   Mean   : 87.32   Mean   :0.4815   Mean   :11.26  
##  3rd Qu.:46.00   3rd Qu.: 91.00   3rd Qu.:0.5160   3rd Qu.:14.00  
##  Max.   :55.00   Max.   :102.00   Max.   :0.5880   Max.   :19.00  
##     X3PA_LA         X3Ppc_LA          FT_LA           FTA_LA     
##  Min.   :19.00   Min.   :0.1050   Min.   : 7.00   Min.   : 8.00  
##  1st Qu.:29.00   1st Qu.:0.2965   1st Qu.:14.00   1st Qu.:19.00  
##  Median :32.00   Median :0.3435   Median :18.00   Median :24.00  
##  Mean   :32.17   Mean   :0.3487   Mean   :17.95   Mean   :24.39  
##  3rd Qu.:36.00   3rd Qu.:0.4042   3rd Qu.:23.00   3rd Qu.:29.00  
##  Max.   :47.00   Max.   :0.5480   Max.   :33.00   Max.   :43.00  
##     FTpc_LA           ORB_LA          TRB_LA          AST_LA     
##  Min.   :0.4710   Min.   : 3.00   Min.   :25.00   Min.   :17.00  
##  1st Qu.:0.6670   1st Qu.: 9.00   1st Qu.:41.00   1st Qu.:22.75  
##  Median :0.7340   Median :11.00   Median :45.00   Median :25.00  
##  Mean   :0.7371   Mean   :10.64   Mean   :45.36   Mean   :25.45  
##  3rd Qu.:0.8135   3rd Qu.:12.00   3rd Qu.:49.00   3rd Qu.:28.25  
##  Max.   :0.9470   Max.   :19.00   Max.   :62.00   Max.   :39.00  
##      STL_LA           BLK_LA           TOV_LA          PF_LA      
##  Min.   : 2.000   Min.   : 1.000   Min.   : 7.00   Min.   :13.00  
##  1st Qu.: 6.750   1st Qu.: 4.000   1st Qu.:12.75   1st Qu.:18.00  
##  Median : 9.000   Median : 6.000   Median :15.00   Median :21.50  
##  Mean   : 8.489   Mean   : 6.304   Mean   :14.61   Mean   :21.16  
##  3rd Qu.:10.000   3rd Qu.: 8.000   3rd Qu.:16.25   3rd Qu.:24.00  
##  Max.   :14.000   Max.   :20.000   Max.   :24.00   Max.   :30.00  
##        FG             FGA              FGpc             X3P       
##  Min.   :28.00   Min.   : 65.00   Min.   :0.3260   Min.   : 4.00  
##  1st Qu.:35.00   1st Qu.: 81.00   1st Qu.:0.4155   1st Qu.: 9.00  
##  Median :38.50   Median : 86.00   Median :0.4470   Median :11.00  
##  Mean   :38.55   Mean   : 85.83   Mean   :0.4497   Mean   :11.68  
##  3rd Qu.:42.00   3rd Qu.: 91.00   3rd Qu.:0.4813   3rd Qu.:14.00  
##  Max.   :54.00   Max.   :102.00   Max.   :0.5840   Max.   :22.00  
##       X3PA           X3Ppc              FT             FTA       
##  Min.   :15.00   Min.   :0.1710   Min.   : 6.00   Min.   : 7.00  
##  1st Qu.:28.00   1st Qu.:0.2930   1st Qu.:14.00   1st Qu.:19.00  
##  Median :33.00   Median :0.3520   Median :18.00   Median :23.00  
##  Mean   :33.22   Mean   :0.3506   Mean   :18.47   Mean   :23.46  
##  3rd Qu.:37.00   3rd Qu.:0.4100   3rd Qu.:23.00   3rd Qu.:29.00  
##  Max.   :57.00   Max.   :0.5650   Max.   :32.00   Max.   :39.00  
##       FTpc             ORB              TRB             AST       
##  Min.   :0.5160   Min.   : 1.000   Min.   :26.00   Min.   :12.00  
```

```
##  1st Qu.:0.7390   1st Qu.: 6.000   1st Qu.:37.00   1st Qu.:20.00
##  Median :0.7860   Median : 8.000   Median :41.00   Median :23.00
##  Mean   :0.7886   Mean   : 8.957   Mean   :41.13   Mean   :23.03
##  3rd Qu.:0.8800   3rd Qu.:12.000   3rd Qu.:45.00   3rd Qu.:25.25
##  Max.   :1.0000   Max.   :18.000   Max.   :61.00   Max.   :37.00
##       STL              BLK              TOV              PF
##  Min.   : 2.000   Min.   : 0.000   Min.   : 5.00   Min.   :11.0
##  1st Qu.: 6.000   1st Qu.: 2.000   1st Qu.:12.00   1st Qu.:19.0
##  Median : 8.000   Median : 3.000   Median :15.00   Median :22.0
##  Mean   : 7.946   Mean   : 3.565   Mean   :14.83   Mean   :21.9
##  3rd Qu.:10.000   3rd Qu.: 5.000   3rd Qu.:17.25   3rd Qu.:25.0
##  Max.   :15.000   Max.   :10.000   Max.   :26.00   Max.   :32.0
##      Dates
##  Min.   :2019-10-22
##  1st Qu.:2019-12-07
##  Median :2020-01-28
##  Mean   :2020-03-16
##  3rd Qu.:2020-08-08
##  Max.   :2020-10-11
```

## 1.4 Number of columns and rows in dataset

```r
length(Lakers)
```

```
## [1] 41
```

```r
nrow(Lakers)
```

```
## [1] 92
```

```r
# Creating categorical columns
WonLost <- W_L == "W"; head(WonLost)
```

```
## [1] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE
```

```r
sum(WonLost) # total wins
```

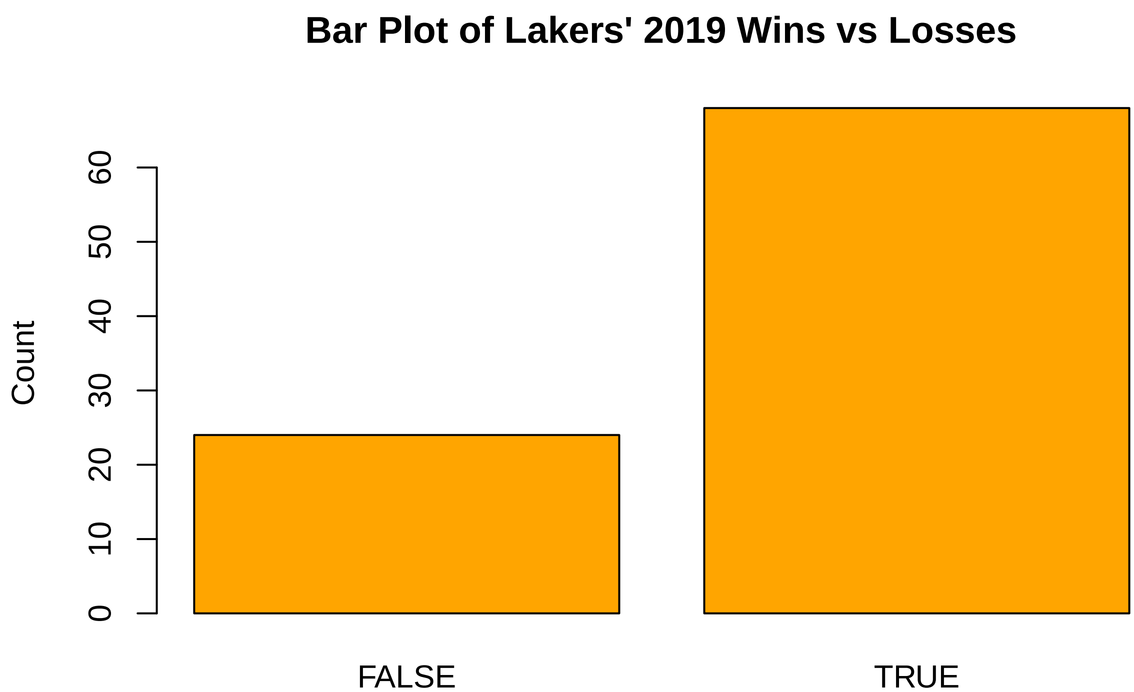
```
## [1] 68
```

## 2 Graphical analysis

### 2.1 Barplot of wins vs losses for our team

This plot shows the number of losses compared to the number wins by our team. The y-axis gives the count of wins and losses. The x-label answers the question “did the team win?”, FALSE means losing and TRUE means winning.

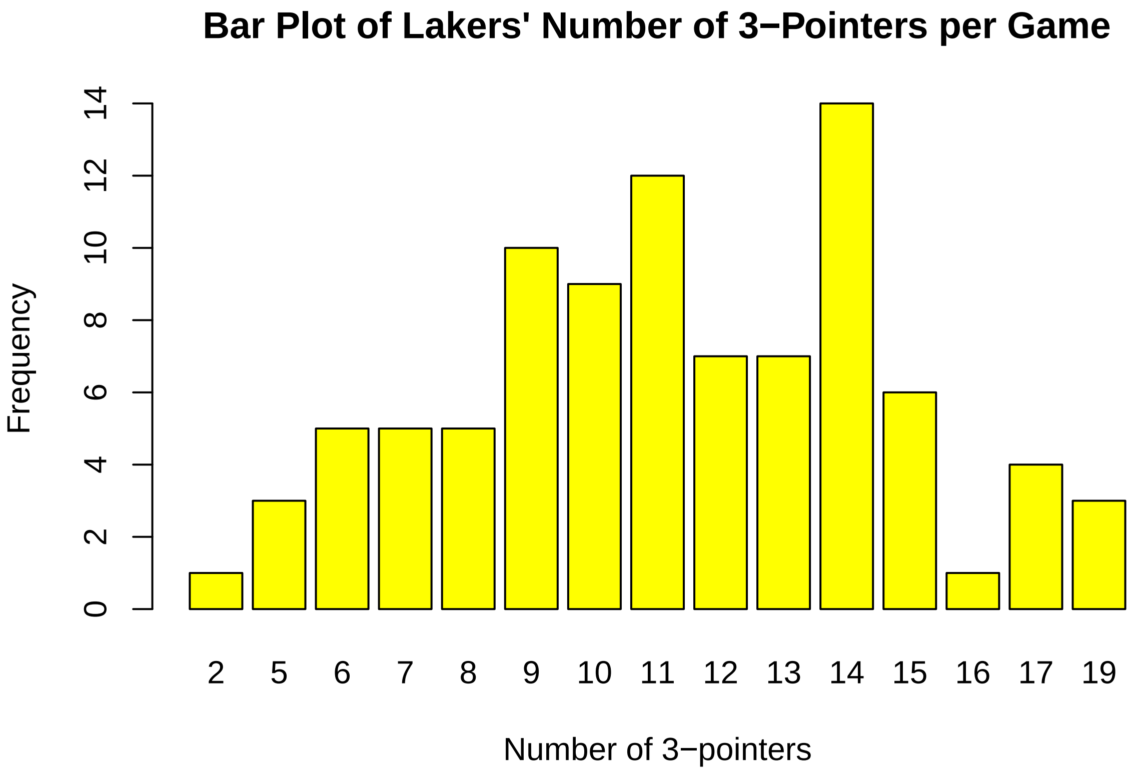
```
barplot(table(WonLost), main = "Bar Plot of Lakers' 2019 Wins vs Losses", col = "orange", ylab="Count")
```




### 2.2 Barplot of 3 point shots from our team

This plot shows the number of times the team scores a 3-point shot in a game. How often a 3-pointer occurs ranges from the minimum of 2 to a maximum of 19 in a single game.

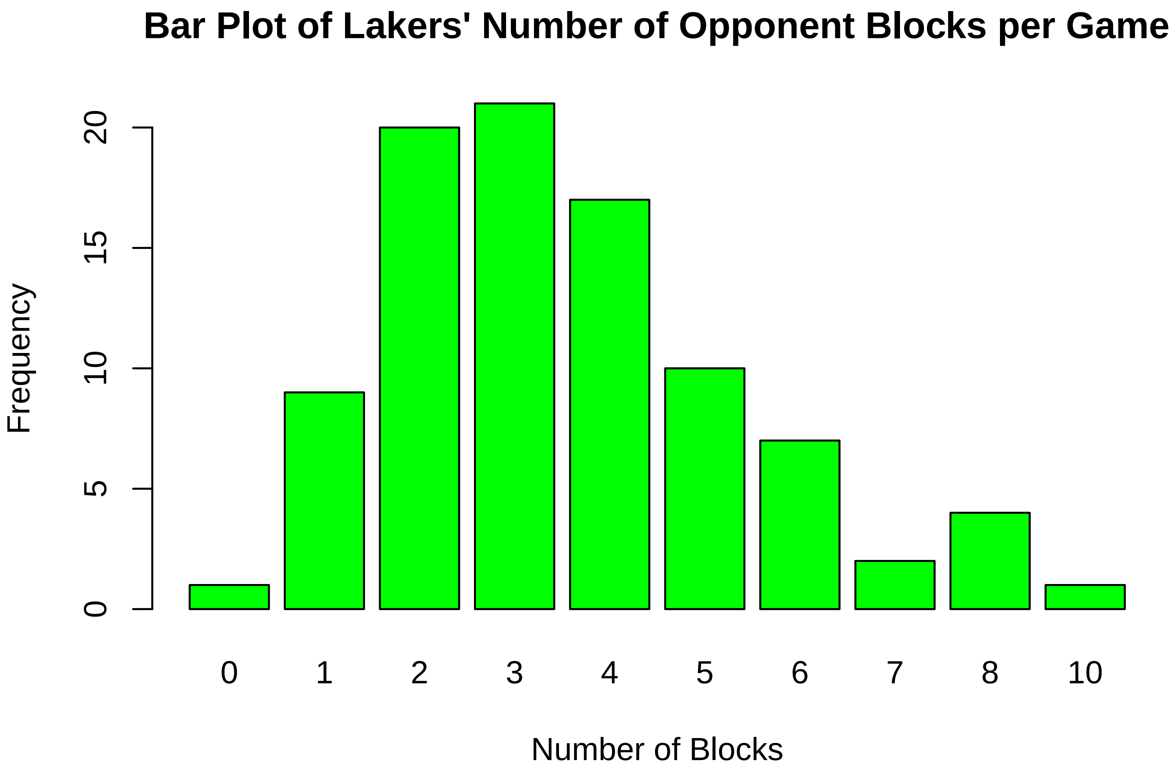
```
barplot(table(X3P_LA), main = "Bar Plot of Lakers' Number of 3-Pointers per Game", xlab="Number of 3-poi
```




## 2.3 Barplot of Opponent blocks

This plot shows how often an opponent successfully blocks our team in a game. Most often (the median), an opponent is able to block 3 times and, of course, depending on the opponent's abilities, blocks can range from 0 to 9 in a game.

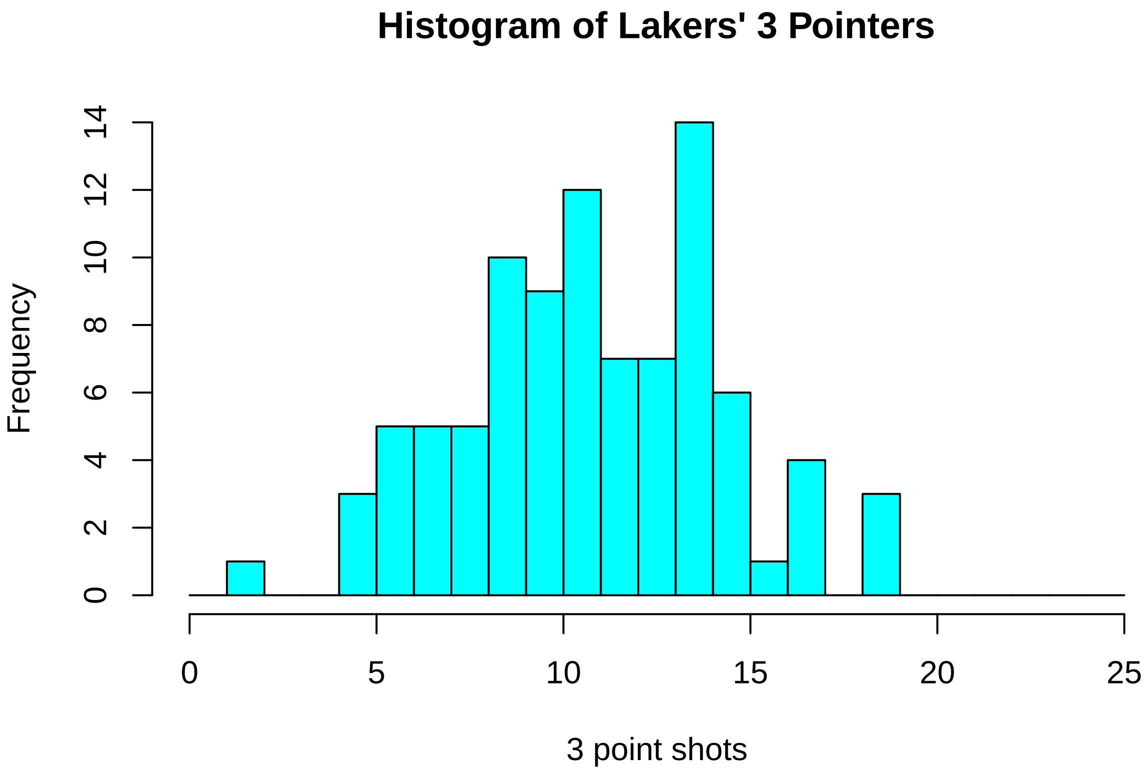
```
barplot(table(BLK), main = "Bar Plot of Lakers' Number of Opponent Blocks per Game", xlab="Number of Bl
```




### 2.4 Histogram of 3 point shots

This is a histogram version of the barplot which shows the number of times the team scores a 3-point shot in a game. The data is the same, and so is the result, a 3-pointer event ranges from the minimum of 2 to a maximum of 19.

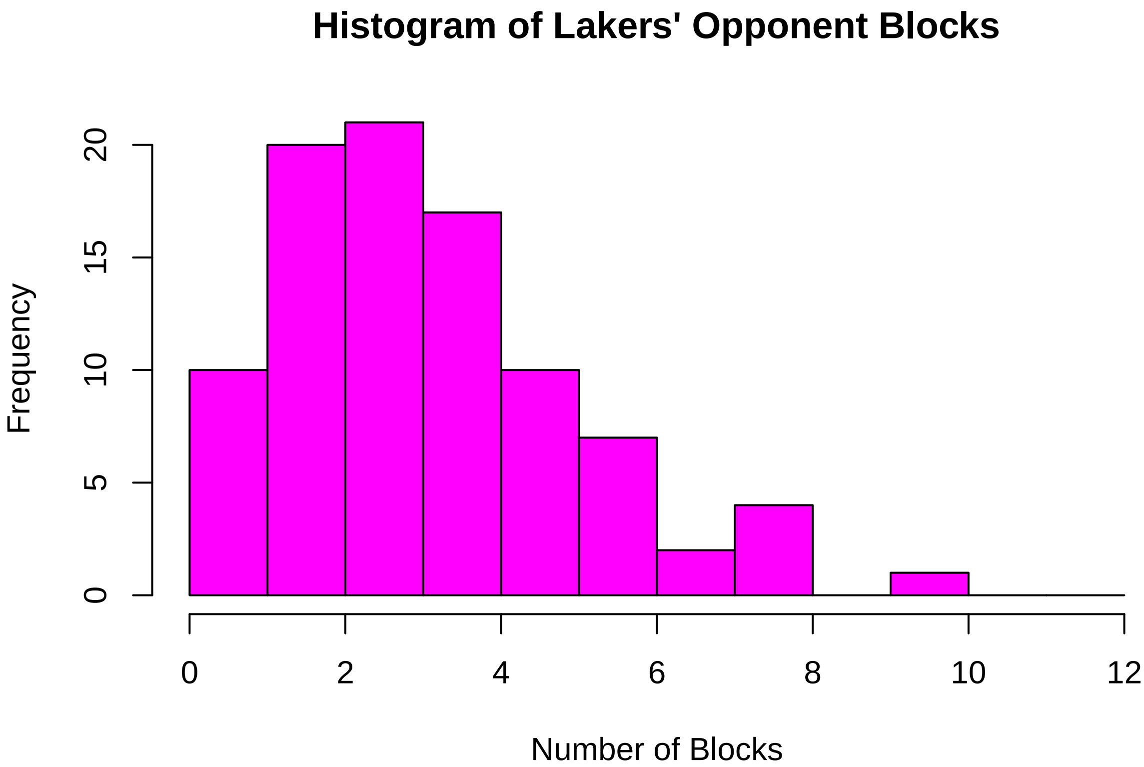
```
hist(X3P_LA, xlab = "3 point shots", main = "Histogram of Lakers' 3 Pointers", col = "cyan", breaks=(0:
```




### 2.5 Histogram of Opponent blocks

This is a histogram version that shows how often an opponent successfully blocks our team in a game. Same results, an opponent is able to block 3 times and ranges from 0 to 9 in a game.

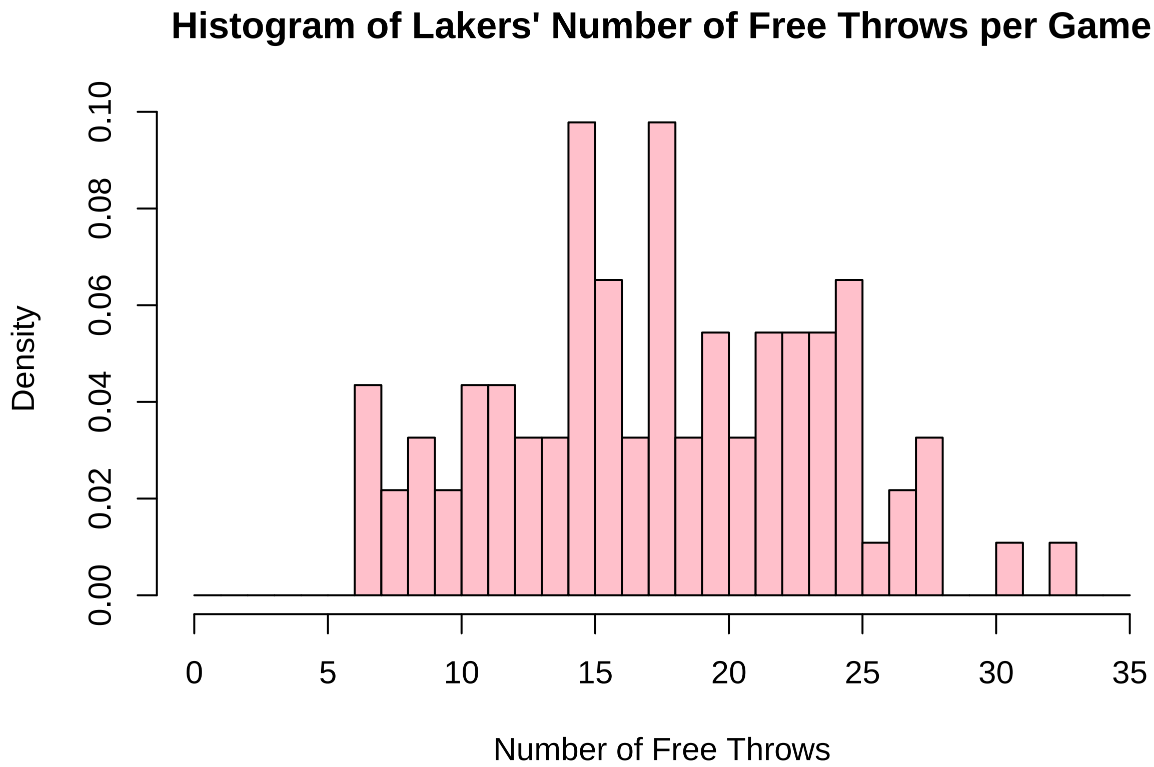
```
hist(BLK, main = "Histogram of Lakers' Opponent Blocks", xlab="Number of Blocks", col = "magenta", breal
```




### 2.6 Histogram of our team's number of Free Throws

This histogram shows the probability of the number of free throws done by our team in a game. The range is from a minimum of 7 to a maximum of 33 with a maximum density at 14 and 17.

```
hist(FT_LA, probability=TRUE, main="Histogram of Lakers' Number of Free Throws per Game", xlab="Number o
```

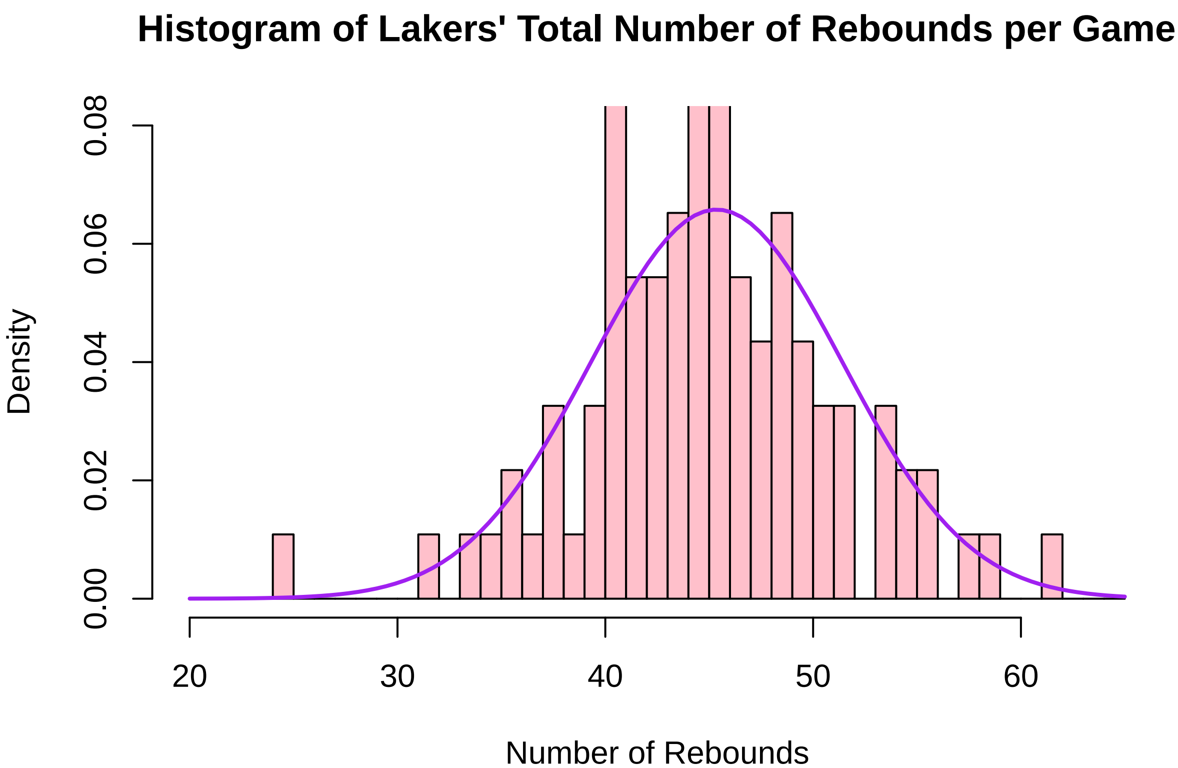

Histogram of Lakers' Number of Free Throws per Game
Density
0.00
0.02
0.04
0.06
0.08
0.10
0
5
10
15
20
25
30
35
Number of Free Throws

# 3 Probability analysis

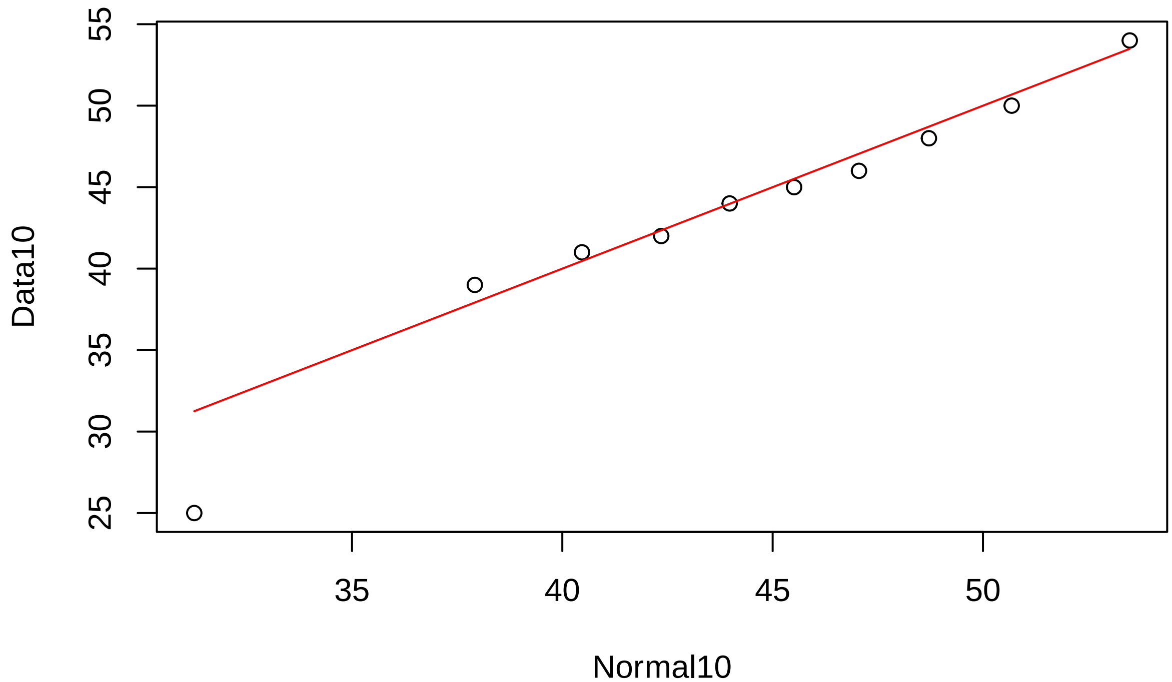
## 3.1 Probability density graph overlay on the our team's total number of rebounds in a game

Here we have our team's probability of the total number of rebounds in a game. Ranging from a minimum of 25 to a maximum of 62, the graph seems like it came from a normal distribution. Overlaying a normal distribution shows like it does.

```
mu<-mean(TRB_LA); stdv<- sd(TRB_LA)
hist(TRB_LA, probability=TRUE, main = "Histogram of Lakers' Total Number of Rebounds per Game", ylim=c(
# total rebounds per game
curve(dnorm(x, mu, stdv), col="purple", add=TRUE, lwd=2)
```




```
#Appears to be a normal distribution. We will compare it with Deciles and a Q-Q plot
Normal10<- qnorm(seq(0.01, 0.99, by=0.1), mean=mu, sd=stdv);
Data10<- quantile(TRB_LA, seq(0.01, 0.99, by=0.1), type=2);
plot(Normal10, Data10)
f<-function(x) x
curve(f, col="red", add=TRUE)
```

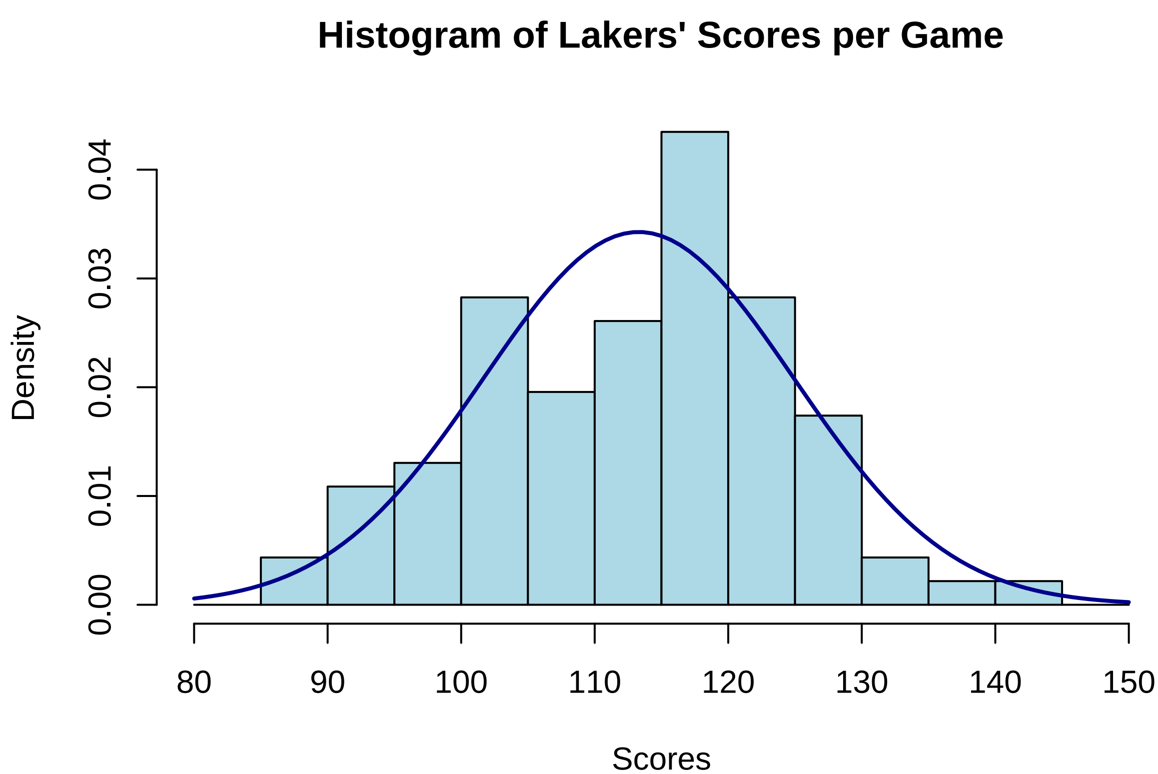



```
#Our data does not differ much from the Normal Distribution, except in the first decile.
```

### 3.2 Probability density graph on the our team's scores in a game

We go back to our histogram of our team's score and its probability in a game. It also seems as if it fits a normal distribution like it shows with a normal distribution overlay. We can provide more proof later with a chi-square goodness of fit test.

```
mu<-mean(Tm); stdv<-sd(Tm)
hist(Tm,breaks=seq(from=80, to = 150, by =5), probability=TRUE, main = "Histogram of Lakers' Scores per
curve(dnorm(x, mu, stdv), col="dark blue", add=TRUE, lwd=2)
```




```
#Binning and Chisq test to compare this histogram to a Normal Distribution
dec <- qnorm(seq(0.0, 1, by = 0.1), mu, stdv); dec   #10 bins
```

```
## [1]      -Inf  98.37651 103.49720 107.18957 110.34458 113.29348 116.24238
## [8] 119.39738 123.08976 128.21045       Inf
```

```
Exp <- rep(length(Tm)/10,10); Exp      #expected scores per bin
```

```
## [1] 9.2 9.2 9.2 9.2 9.2 9.2 9.2 9.2 9.2 9.2
```

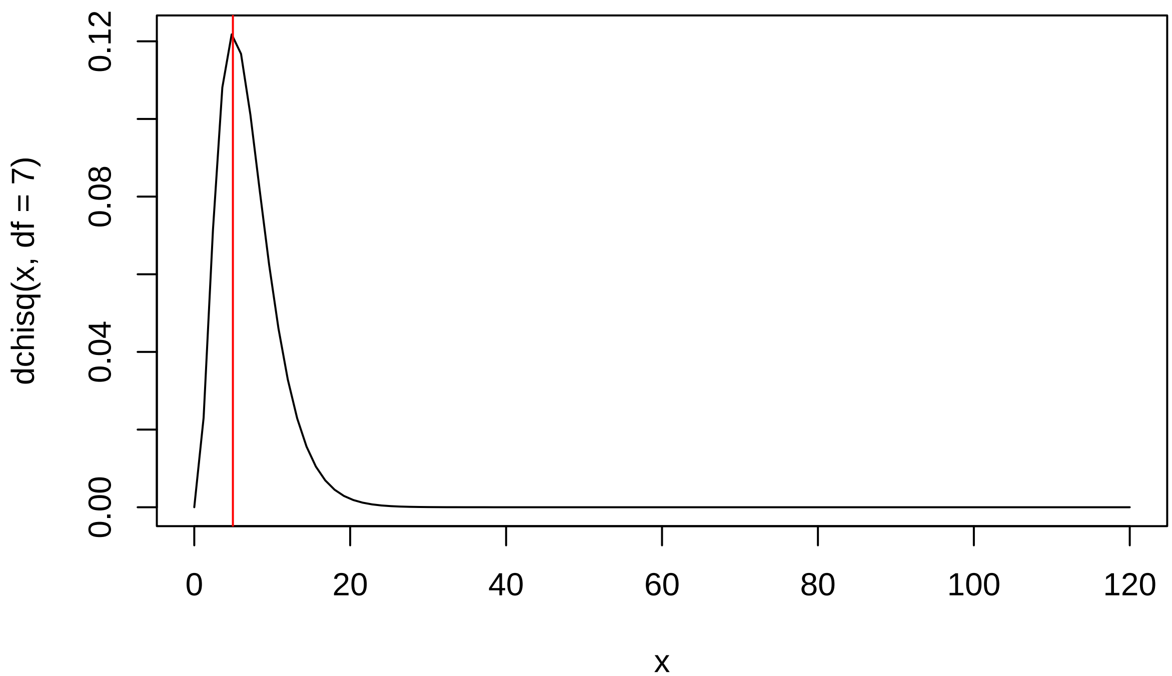
```
#Now we can count how many scores are in each bin
binscores <- numeric(10)
for (i in 1:10)
  binscores[i] <- sum((Tm >= dec[i]) & (Tm <= dec[i+1]) ); binscores
```

```
## [1] 10 10 11  4  8  8 11 11 11  8
```

```
#Test for uniformity using chi square.
Chi2 <- sum((binscores - Exp)^2/Exp); Chi2
```

```
## [1] 4.956522
```

```
#There were 10 bins. We estimated 2 parameters (mu and stdv), which costs two degrees of freedom
#Also we "made the totals match", costing another 1. So there are 10-2-1=7 df.
curve(dchisq(x, df = 7), from = 0, to = 120 )
abline(v=Chi2, col = "red")
```




```r
#The probability of this chi-square value is relatively large
#The normal distribution was a good model
```

### 3.3 Contingency tables

This table answers the question "how many times the opponents scored greater than 100 points in a game with our team". This does not say that the opponent won or not. But in the times that they did not, it only shows that our team scored more and that such a high scoring game would have been a truly awesome event to watch.

```r
Opp100 <- Opp_pts > 100; Opp100
```

```
##  [1]  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE
## [13]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE
## [25] FALSE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE FALSE
## [37] FALSE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE
## [49]  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
## [61]  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE
## [73] FALSE  TRUE  TRUE  TRUE  TRUE  TRUE  TRUE FALSE FALSE  TRUE  TRUE  TRUE
## [85]  TRUE  TRUE FALSE  TRUE  TRUE FALSE  TRUE FALSE
```

```r
table(Opp100)
```

```
## Opp100
## FALSE  TRUE
##    24    68
```

This table shows the opposite of the table above, that despite being the all-around champion during that year, our team scored less than 100 in some games and this table shows that.

```r
Less100 <- Tm < 100; Less100
```

```
##  [1] FALSE  TRUE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE  TRUE
## [13] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [25]  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
## [37] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## [49] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE  TRUE FALSE FALSE
## [61] FALSE FALSE FALSE FALSE  TRUE FALSE  TRUE  TRUE FALSE FALSE FALSE  TRUE
## [73] FALSE FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## [85] FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
```

```
table(Less100)
```

```
## Less100
## FALSE  TRUE
##    81    11
```

# 4 Inferential and statistical analysis

## 4.1 Permutation test

Our permutation test will be comparing our team's scores in games played vs the opponents from the East and the opponents from the West

H_o: The mean score of the opponents coming from the Eastern Conference is equal to the mean score of the opponents coming from the Western Conference H_a: The mean score of the opponents coming from the Eastern Conference is not equal to the mean score of the opponents coming from the Western Conference

#Data for Eastern and Western Conference teams How many opponents were from each, and what were their mean scores?

```
Eastern <- W_E == "E"; head(Eastern)
```

```
## [1] FALSE FALSE  TRUE FALSE FALSE FALSE
```

```
East <- sum(Eastern); East # count of Eastern opponents = 31
```

```
## [1] 31
```

```
ScorevsEast <- mean(Tm*Eastern); ScorevsEast # mean team score vs East = 37.03261
```

```
## [1] 37.03261
```

```
Western <- W_E == "W"; head(Western)
```

```
## [1]  TRUE  TRUE FALSE  TRUE  TRUE  TRUE
```

```
West <- sum(Western); West # Western opponents = 61
```

```
## [1] 61
```

```
ScorevsWest <- mean(Tm*Western); ScorevsWest # mean team score vs West = 76.26087
```

```
## [1] 76.26087
```

It shows that there are more opponents coming from the West and that their mean score obviously would have a higher range.

## 4.2 Total number of games played and differences of mean scores

This will be our Observed value.

```
EW <- sum(Eastern) + sum(Western); EW # total number of games = 92
```

```
## [1] 92
```

```
Score_diff <- ScorevsEast - ScorevsWest; Score_diff # -39.22826
```

```
## [1] -39.22826
```

```
Observed <- Score_diff; Observed
```

```
## [1] -39.22826
```

```
# Let's see if this score difference is significant
# We repeat 10^6 times
N <- 10^6
Score_diffs <- numeric(N)
for (i in 1:N){
  # Permute West indices
  E <- sample(EW, East, replace = FALSE)
```

```r
  # Get the difference of the 2 opponent groups
  Score_diffs[i] <- mean(Tm[E]) - mean(Tm[-E])
}

head(Score_diffs)
```

```
## [1]  2.2818614 -0.6858805 -2.1454257 -1.9994712 -1.1723956  2.3305130
```

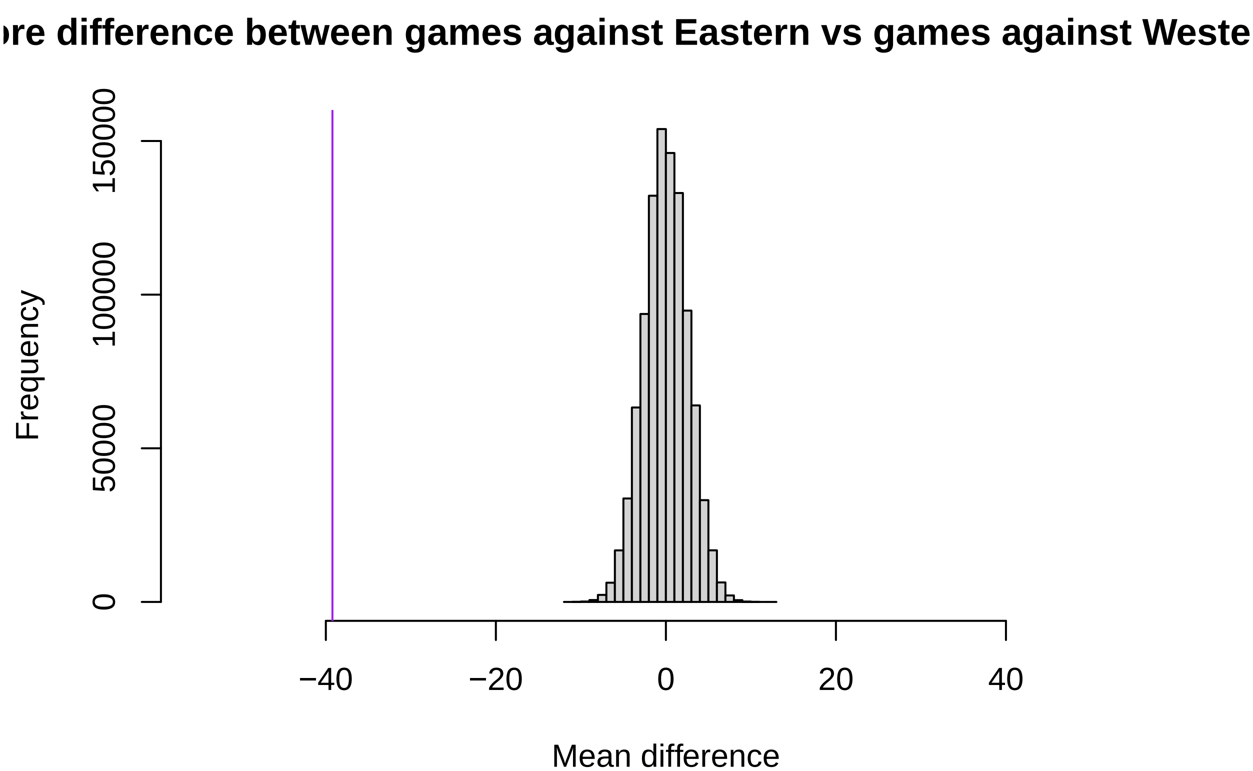
```r
summary(Score_diffs)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -11.632470  -1.756214  -0.004759  -0.002543   1.746695  12.255420
```

```r
mean(Score_diffs) # 0.001794507 close to zero
```

```
## [1] -0.002543359
```

```r
hist(Score_diffs, main="Mean Score difference between games against Eastern vs games against Western Op
     col="light gray", xlab="Mean difference", xlim=c(-55, 55))
#Now display the observed value on the histogram
abline(v = Observed, col = "purple")
```




```r
#What is the probability (the P value) that a difference this large
#could have arisen with a random subset?
pvalue <- (sum(Score_diffs >= Observed)+1)/(N+1); pvalue # 1
```

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

This goes to show that the data observed has a significant likelihood to have come about by chance. Therefore, there is insufficient evidence to to reject the null hypothesis.