# Math E-23C Term Project

An analysis of the Lakers 2019 Championship season

Maria Cristina An, Daniel Lebedinsky, Julio Solis Arce

06 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 A glimpse at the dataset** **2**

**2 Graphical analysis** **3**

2.1 Barplot of wins vs losses for our team . . . 3

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

2.3 Barplot of Opponent blocks . . . 4

2.4 Histogram of 3 point shots . . . 5

2.5 Histogram of Opponent blocks . . . 6

2.6 Histogram of our team's number of Free Throws . . . 7

**3 Probability analysis** **9**

3.1 Probability density graph of our team's scores in a game . . . 9

3.2 Probability density graph of the opponents' scores in a game . . . 10

3.3 Contingency tables . . . 12

**4 Inferential and statistical analysis** **14**

4.1 Permutation test . . . 14

4.2 Data for Eastern and Western Conference teams . . . 14

4.3 Total number of games played and differences of mean scores . . . 14

# 1 A glimpse at the dataset

```r
#Import dataset
Lakers <- read.csv("Lakers\ 2019-20\ Game\ log.csv"); head(Lakers, n = c(6, 8)) %>% knitr::kable()
```

| Rk | G | Date | W_E | Opp | W_L | Tm | Opp_pts |
|---:|---:|---|---|---|---|---:|---:|
| 1 | 1 | 2019-10-22 | W | LAC | L | 102 | 112 |
| 2 | 2 | 2019-10-25 | W | UTA | W | 95 | 86 |
| 3 | 3 | 2019-10-27 | E | CHO | W | 120 | 101 |
| 4 | 4 | 2019-10-29 | W | MEM | W | 120 | 91 |
| 5 | 5 | 2019-11-01 | W | DAL | W | 119 | 110 |
| 6 | 6 | 2019-11-03 | W | SAS | W | 103 | 96 |

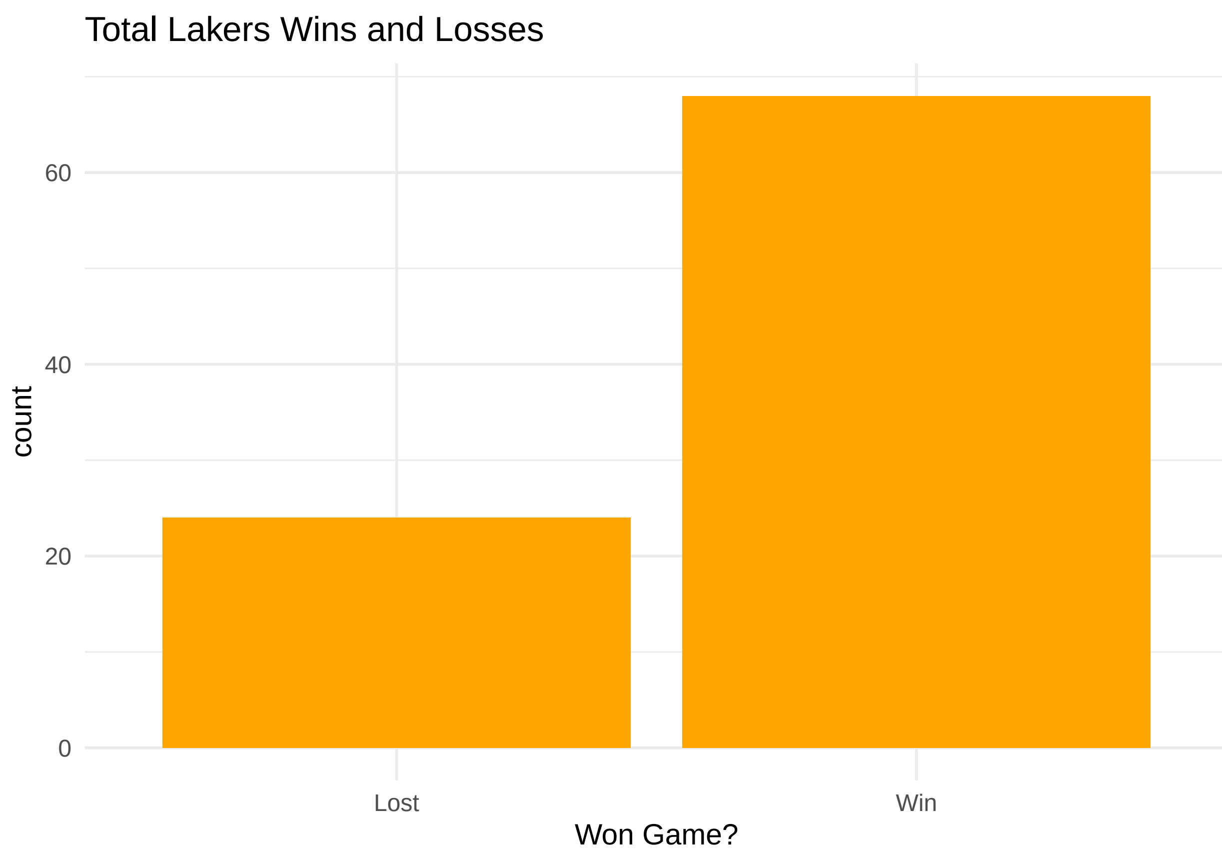
```r
# Convert to proper date format
Lakers$Dates <- as.Date(Lakers$Date)
attach(Lakers)

# Creating categorical columns
WonLost <- ifelse(W_L == "W", "Win", "Lost")
```

# 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?". The Lakers season was a majorly successful one: they won 68 games in total and lost 24 of them. 52 of those wins were in regular season, with the rest done in post season.

```
Lakers %>%
  ggplot(aes(WonLost)) +
  geom_bar(fill="orange") +
  xlab("Won Game?") +
  ggtitle("Total Lakers Wins and Losses") +
  theme_minimal()
```

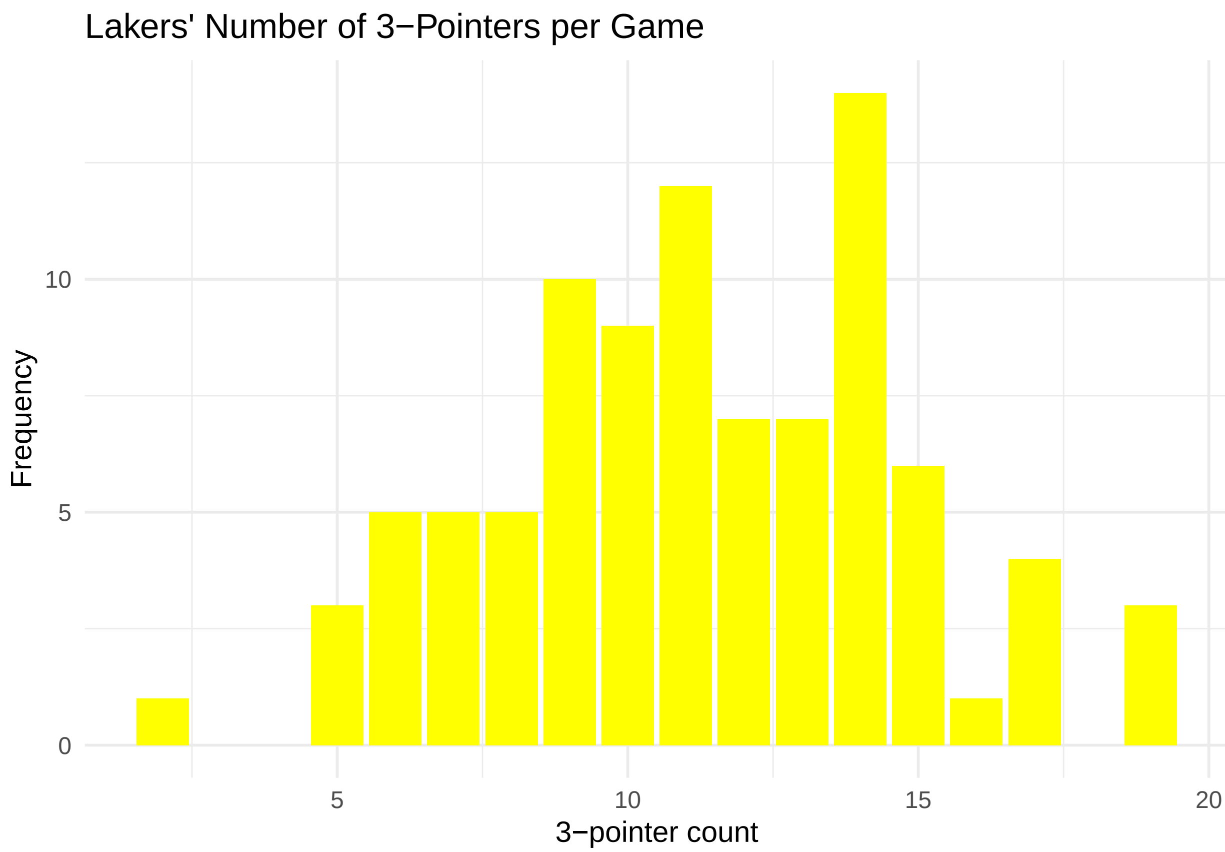



## 2.2 Barplot of 3 point shots from our team

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. The mean three points per game where 11.26, with a mode of 11 and a median of 11.

```
Lakers %>% ggplot(aes(X3P_LA))+
  geom_bar(fill="yellow") +
```

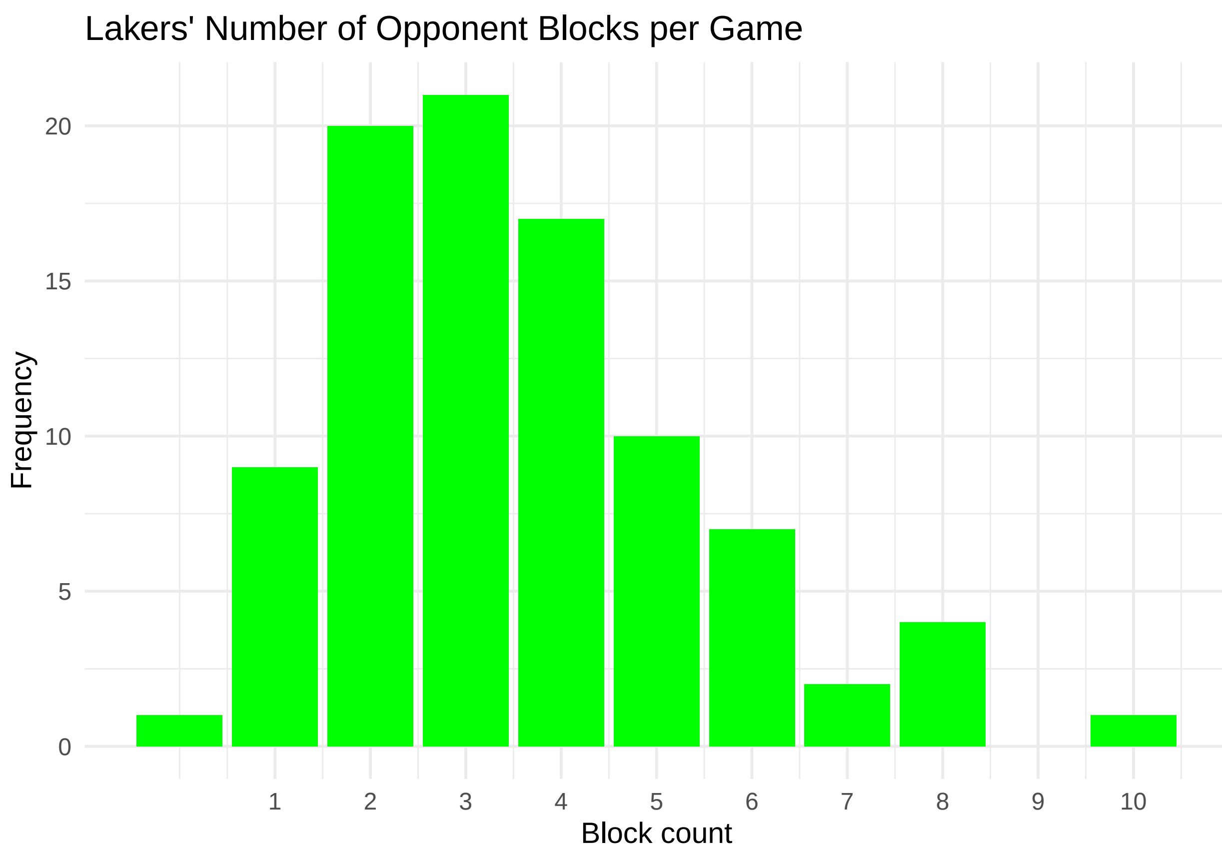
```
  xlab("3-pointer count") +
  ylab("Frequency") +
  ggtitle("Lakers' Number of 3-Pointers per Game") +
  theme_minimal()
```




## 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.

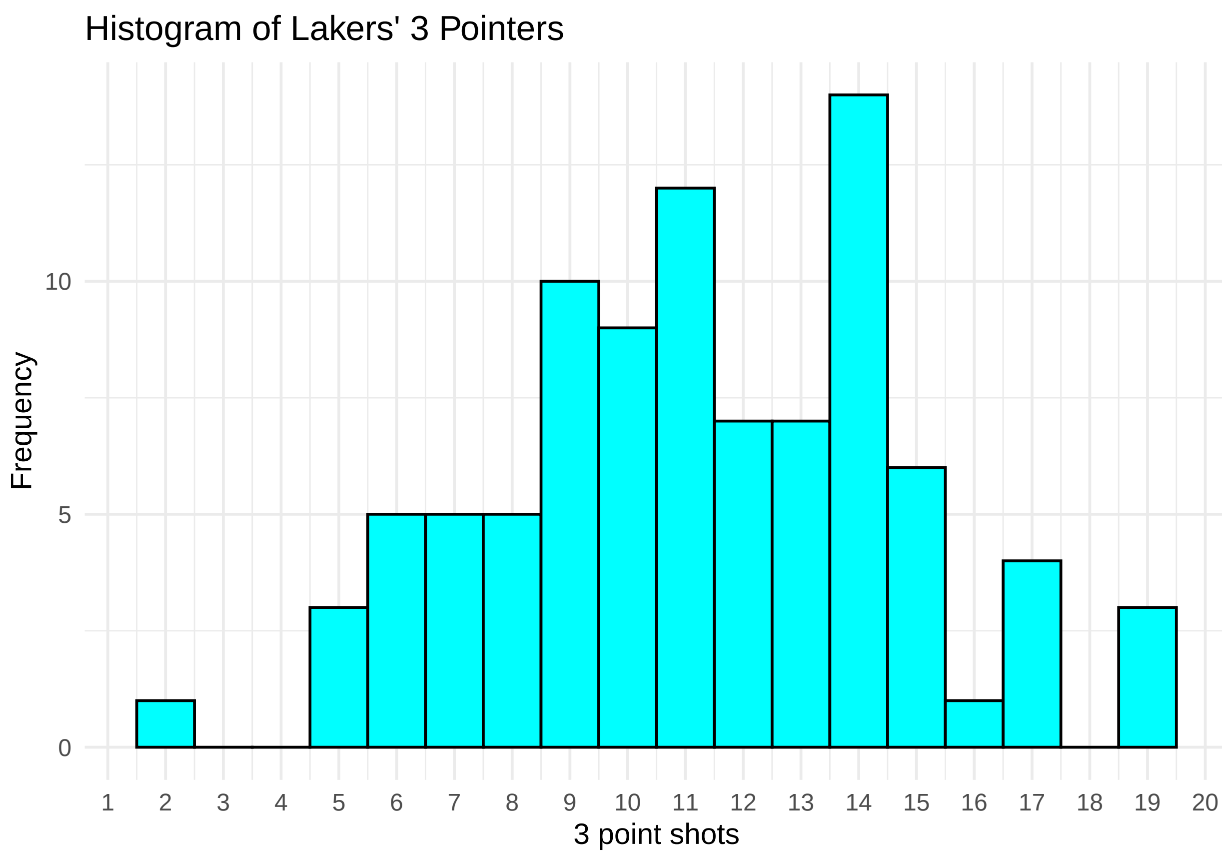
```
Lakers %>% ggplot(aes(BLK))+geom_bar(fill="green")+
  scale_x_continuous(breaks=c(1:10))+
  xlab("Block count") +
  ylab("Frequency")+
  ggtitle("Lakers' Number of Opponent Blocks per Game") +
  theme_minimal()
```




## 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.

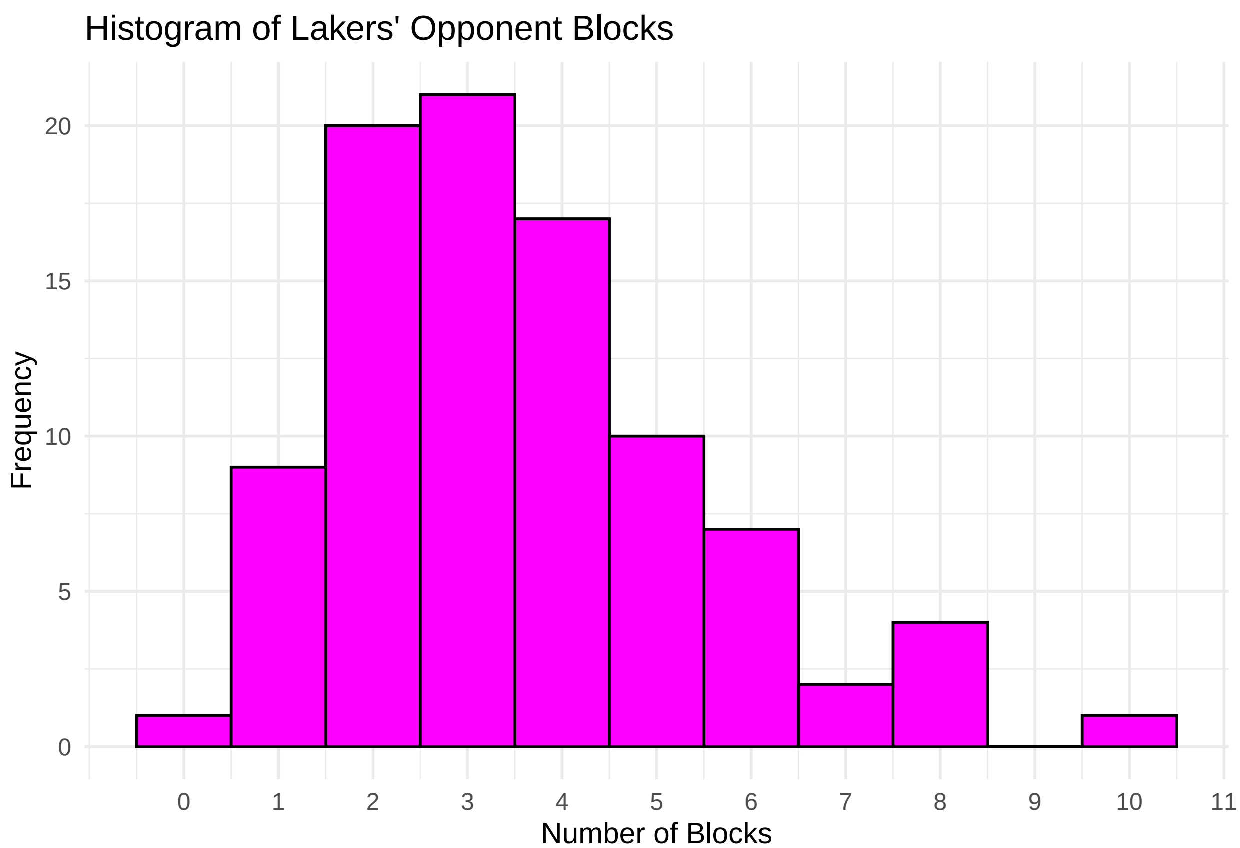
```
Lakers %>%
  ggplot(aes(X3P_LA)) +
  geom_histogram(binwidth = 1, fill="cyan", col = "black")+
  scale_x_continuous(breaks=c(0:20))+
  ylab("Frequency")+
  xlab("3 point shots")+
  ggtitle("Histogram of Lakers' 3 Pointers") +
  theme_minimal()
```




## 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.

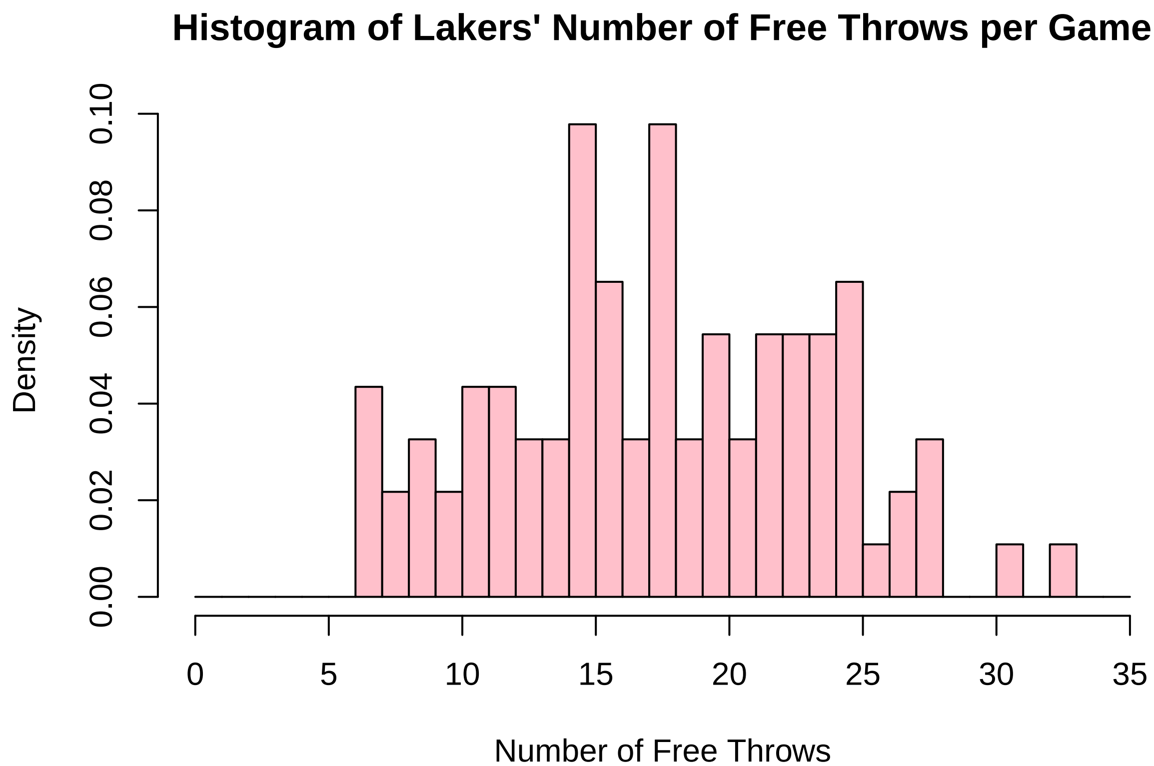
```
Lakers %>%
  ggplot(aes(BLK)) +
  geom_histogram(binwidth = 1, fill="magenta", col = "black")+
  scale_x_continuous(breaks=c(0:12))+
  ylab("Frequency")+
  xlab("Number of Blocks")+
  ggtitle("Histogram of Lakers' Opponent Blocks")  +
  theme_minimal()
```




## 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 of Free Throws",
     col = "pink",
     breaks=(0:35))
```



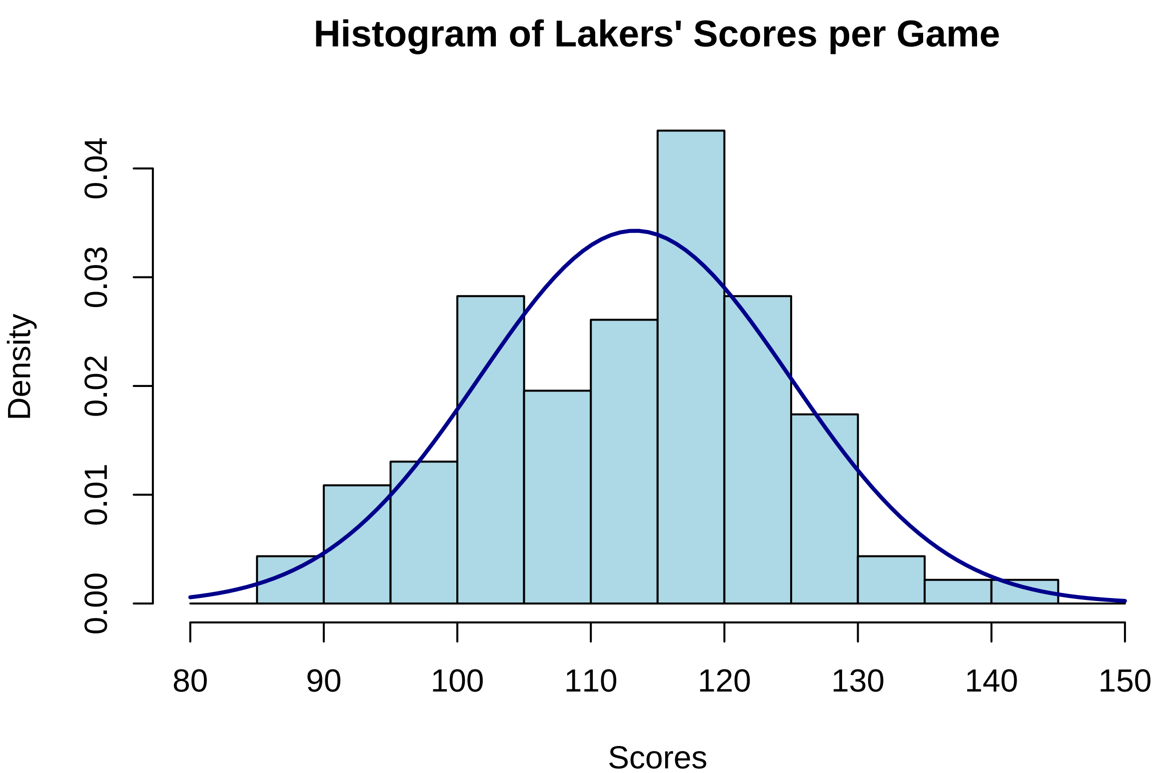

```
# free throws from the team per game
```

# 3 Probability analysis

## 3.1 Probability density graph of our team's scores in a game

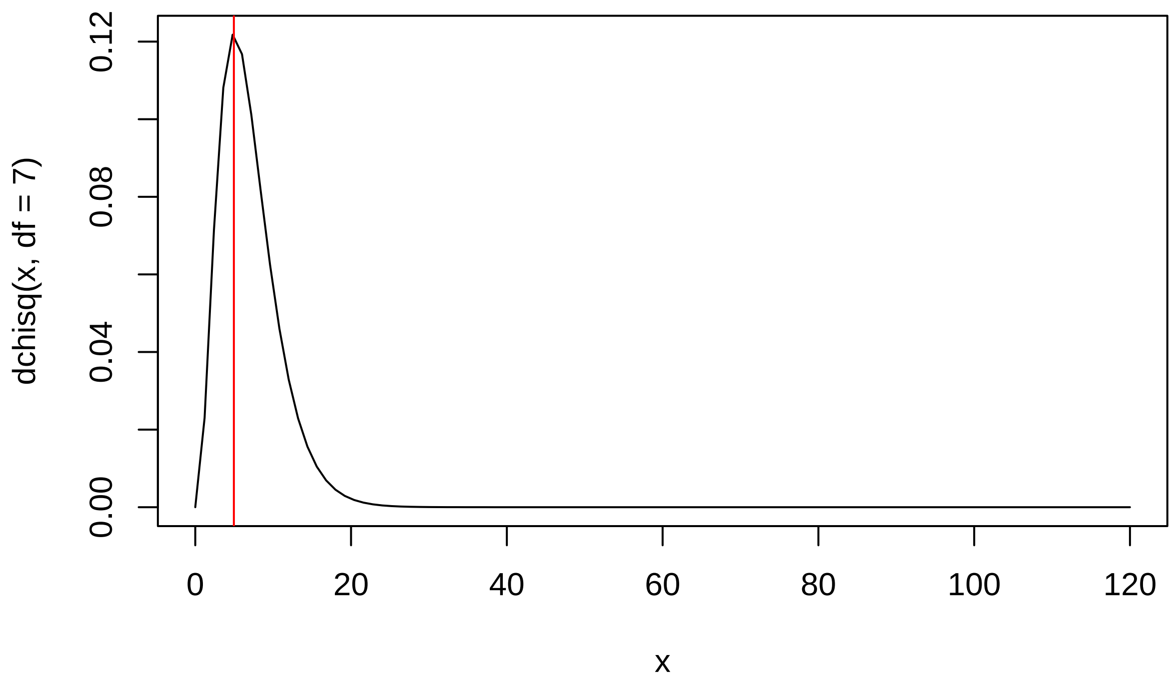
We go back to our histogram of our team's score and its probability in a game. By looking at it, it is plausible that we can model it by a normal distribution. Here we overlay a normal distribution to the data. Then, we provide evidence that this is right way to model it by using a chi-square goodness 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 Game",
     xlab="Scores",
     col = "light blue") # team's score frequencies

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)   #10 bins
Exp <- rep(length(Tm)/10,10)     #expected scores per bin
#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]) )
#Test for uniformity using chi square.
Chi2 <- sum((binscores - Exp)^2/Exp)
#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")
```

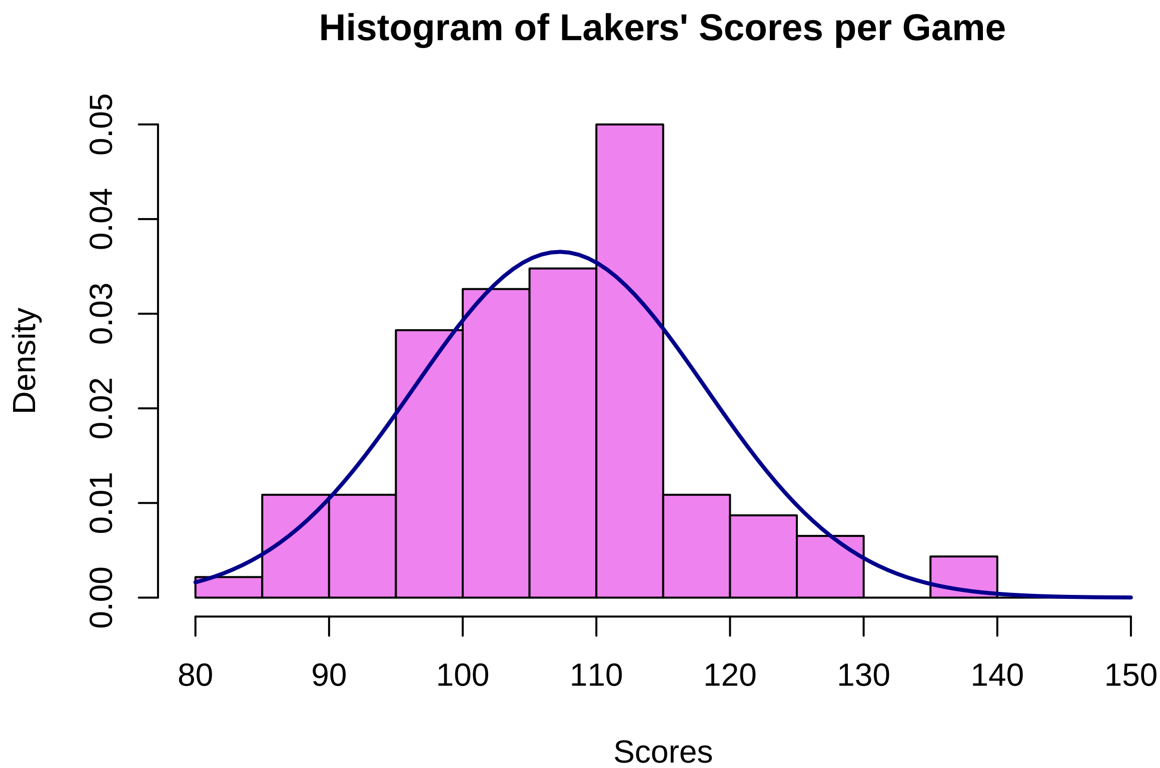



```
#The probability of this chi-square value is relatively large
#The normal distribution was a good model
```

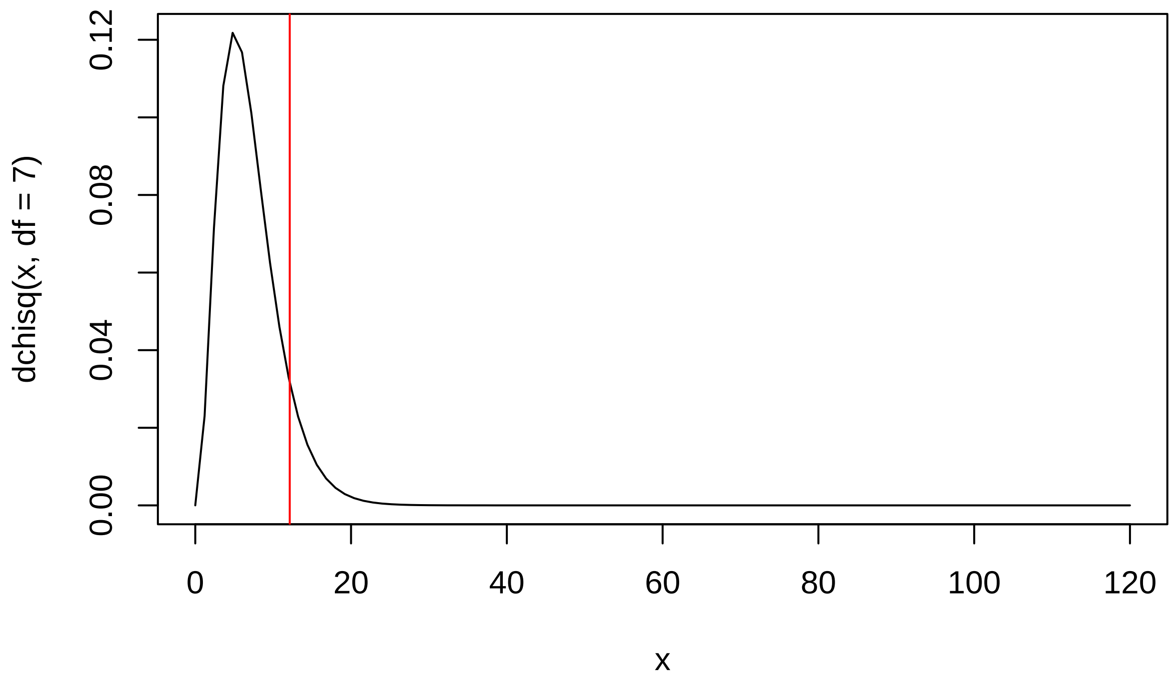
### 3.2 Probability density graph of the opponents' scores in a game

The following analysis shows that the Lakers had a slightly higher mean score than their average opponents did.

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




```r
#Binning and Chisq test to compare this histogram to a Normal Distribution
dec <- qnorm(seq(0.0, 1, by = 0.1), mu, stdv)
Exp <- rep(length(Opp_pts)/10,10)
#Now we can count how many scores are in each bin
binscores <- numeric(10)
for (i in 1:10)
  binscores[i] <- sum((Opp_pts >= dec[i]) & (Opp_pts <= dec[i+1]) )
#Test for uniformity using chi square.
Chi2 <- sum((binscores - Exp)^2/Exp)
#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")
```




```
#The probability of this chi-square value is relatively large
#The normal distribution was a decent, but not ideal 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". In 68 games the opposition scored more than 100 points. We looked at how many of these games we won or lost. In 45 games, the opposition scored more than 100 points and we won, and in 23 times the opposition scored more than 100 points and we lost.

```
Opp100 <- Opp_pts > 100
table(Opp100, W_L) %>%
  knitr::kable()
```

| | L | W |
|---|---|---|
| FALSE | 1 | 23 |
| TRUE | 23 | 45 |

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
table(Less100) %>%
  knitr::kable()
```

| Less100 | Freq |
|---|---:|
| FALSE | 81 |
| TRUE | 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

## 4.2 Data for Eastern and Western Conference teams

How many opponents were from each, and what were their mean scores?

```
Eastern <- W_E == "E"
East <- sum(Eastern) # count of Eastern opponents = 31
ScorevsEast <- mean(Tm*Eastern); ScorevsEast # mean team score vs East = 37.03261
```

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

```
Western <- W_E == "W"
West <- sum(Western) # Western opponents = 61
ScorevsWest <- mean(Tm*Western); ScorevsWest # mean team score vs West = 76.26087
```

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

Lakers played against 31 opponents from the East and 61. The mean Lakers team Points against Eastern teams was of 37.0326087 and against Western teams was 76.2608696.

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

## 4.3 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
```

```r
# 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)

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

head(Score_diffs)
```

```
## [1]  1.6493919  0.5790587 -6.3294553 -0.3453199 -0.7831835  3.4981491
```

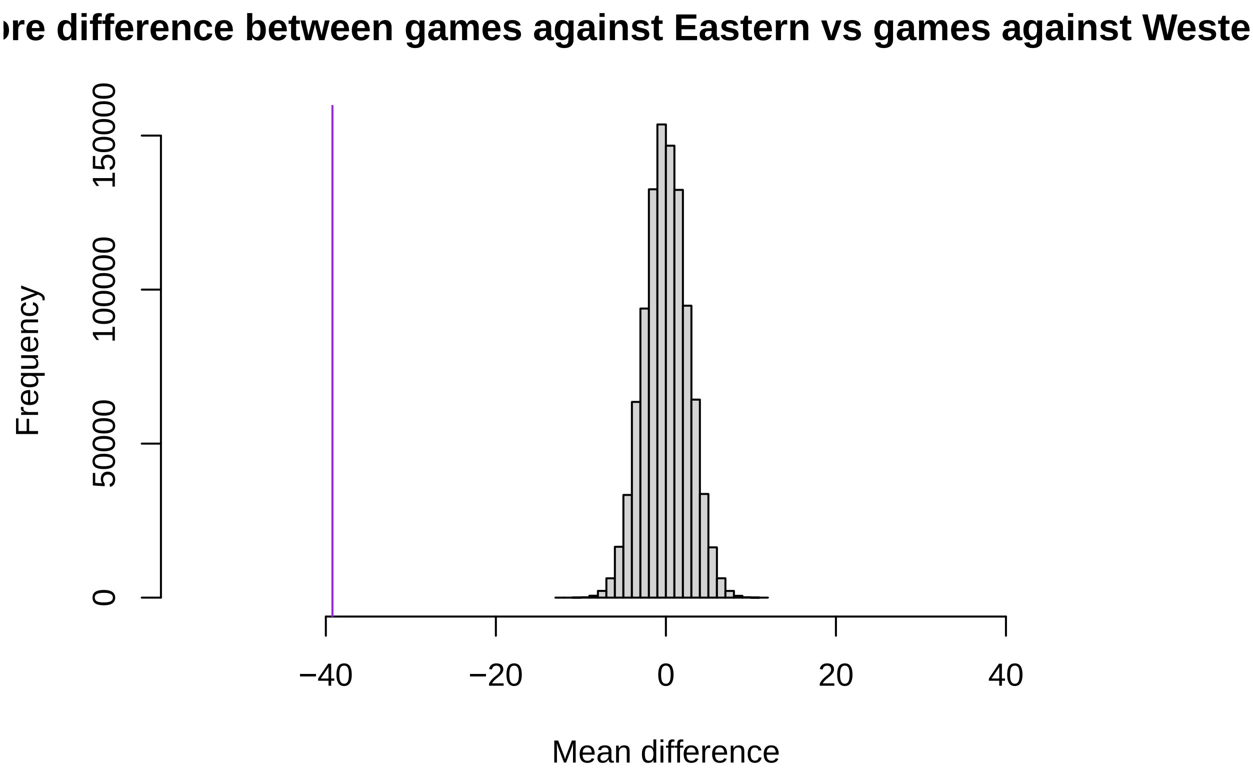
```r
summary(Score_diffs)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -12.167636  -1.756214  -0.004759   0.000459   1.746695  11.914860
```

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

```
## [1] 0.0004589254
```

```r
hist(Score_diffs,
     main="Mean Score difference between games against Eastern vs games against Western Opponents",
     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.