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 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/

```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)
```

1.3 Summary of variables

```r
summary(Lakers) %>%
  t() %>%
  as.data.frame() %>%
  select(-Var2) %>%
  rename(Variable=Var1, Description=Freq) %>%
  filter(!is.na(Description)) %>%
  arrange(Variable) %>%
  knitr::kable()
```

Variable	Description
Rk	Min. : 1.00
Rk	1st Qu.:12.00
Rk	Median :25.50
Rk	Mean :30.29
Rk	3rd Qu.:48.25
Rk	Max. :71.00
G	Min. : 1.00
G	1st Qu.:12.00
G	Median :25.50
G	Mean :30.29
G	3rd Qu.:48.25
G	Max. :71.00
Date	Length:92
Date	Class :character
Date	Mode :character
W_E	Length:92
W_E	Class :character
W_E	Mode :character
Opp	Length:92
Opp	Class :character
Opp	Mode :character
W_L	Length:92
W_L	Class :character
W_L	Mode :character
Tm	Min. : 86.0
Tm	1st Qu.:104.0
Tm	Median :114.0
Tm	Mean :113.3
Tm	3rd Qu.:122.0
Tm	Max. :142.0
Opp_pts	Min. : 80.0
Opp_pts	1st Qu.:100.0
Opp_pts	Median :108.0
Opp_pts	Mean :107.3
Opp_pts	3rd Qu.:114.0
Opp_pts	Max. :139.0
FG_LA	Min. :29.00
FG_LA	1st Qu.:38.00
FG_LA	Median :43.00
FG_LA	Mean :42.04
FG_LA	3rd Qu.:46.00
FG_LA	Max. :55.00

Variable	Description
FGA_LA	Min. : 74.00
FGA_LA	1st Qu.: 83.00
FGA_LA	Median : 87.00
FGA_LA	Mean : 87.32
FGA_LA	3rd Qu.: 91.00
FGA_LA	Max. :102.00
FGpc_LA	Min. :0.3510
FGpc_LA	1st Qu.:0.4512
FGpc_LA	Median :0.4855
FGpc_LA	Mean :0.4815
FGpc_LA	3rd Qu.:0.5160
FGpc_LA	Max. :0.5880
X3P_LA	Min. : 2.00
X3P_LA	1st Qu.: 9.00
X3P_LA	Median :11.00
X3P_LA	Mean :11.26
X3P_LA	3rd Qu.:14.00
X3P_LA	Max. :19.00
X3PA_LA	Min. :19.00
X3PA_LA	1st Qu.:29.00
X3PA_LA	Median :32.00
X3PA_LA	Mean :32.17
X3PA_LA	3rd Qu.:36.00
X3PA_LA	Max. :47.00
X3Ppc_LA	Min. :0.1050
X3Ppc_LA	1st Qu.:0.2965
X3Ppc_LA	Median :0.3435
X3Ppc_LA	Mean :0.3487
X3Ppc_LA	3rd Qu.:0.4042
X3Ppc_LA	Max. :0.5480
FT_LA	Min. : 7.00
FT_LA	1st Qu.:14.00
FT_LA	Median :18.00
FT_LA	Mean :17.95
FT_LA	3rd Qu.:23.00
FT_LA	Max. :33.00
FTA_LA	Min. : 8.00
FTA_LA	1st Qu.:19.00
FTA_LA	Median :24.00
FTA_LA	Mean :24.39
FTA_LA	3rd Qu.:29.00
FTA_LA	Max. :43.00
FTpc_LA	Min. :0.4710
FTpc_LA	1st Qu.:0.6670
FTpc_LA	Median :0.7340
FTpc_LA	Mean :0.7371
FTpc_LA	3rd Qu.:0.8135
FTpc_LA	Max. :0.9470
ORB_LA	Min. : 3.00
ORB_LA	1st Qu.: 9.00
ORB_LA	Median :11.00
ORB_LA	Mean :10.64

Variable	Description
ORB_LA	3rd Qu.:12.00
ORB_LA	Max. :19.00
TRB_LA	Min. :25.00
TRB_LA	1st Qu.:41.00
TRB_LA	Median :45.00
TRB_LA	Mean :45.36
TRB_LA	3rd Qu.:49.00
TRB_LA	Max. :62.00
AST_LA	Min. :17.00
AST_LA	1st Qu.:22.75
AST_LA	Median :25.00
AST_LA	Mean :25.45
AST_LA	3rd Qu.:28.25
AST_LA	Max. :39.00
STL_LA	Min. : 2.000
STL_LA	1st Qu.: 6.750
STL_LA	Median : 9.000
STL_LA	Mean : 8.489
STL_LA	3rd Qu.:10.000
STL_LA	Max. :14.000
BLK_LA	Min. : 1.000
BLK_LA	1st Qu.: 4.000
BLK_LA	Median : 6.000
BLK_LA	Mean : 6.304
BLK_LA	3rd Qu.: 8.000
BLK_LA	Max. :20.000
TOV_LA	Min. : 7.00
TOV_LA	1st Qu.:12.75
TOV_LA	Median :15.00
TOV_LA	Mean :14.61
TOV_LA	3rd Qu.:16.25
TOV_LA	Max. :24.00
PF_LA	Min. :13.00
PF_LA	1st Qu.:18.00
PF_LA	Median :21.50
PF_LA	Mean :21.16
PF_LA	3rd Qu.:24.00
PF_LA	Max. :30.00
FG	Min. :28.00
FG	1st Qu.:35.00
FG	Median :38.50
FG	Mean :38.55
FG	3rd Qu.:42.00
FG	Max. :54.00
FGA	Min. : 65.00
FGA	1st Qu.: 81.00
FGA	Median : 86.00
FGA	Mean : 85.83
FGA	3rd Qu.: 91.00
FGA	Max. :102.00
FGpc	Min. :0.3260
FGpc	1st Qu.:0.4155

Variable	Description
FGpc	Median :0.4470
FGpc	Mean :0.4497
FGpc	3rd Qu.:0.4813
FGpc	Max. :0.5840
X3P	Min. : 4.00
X3P	1st Qu.: 9.00
X3P	Median :11.00
X3P	Mean :11.68
X3P	3rd Qu.:14.00
X3P	Max. :22.00
X3PA	Min. :15.00
X3PA	1st Qu.:28.00
X3PA	Median :33.00
X3PA	Mean :33.22
X3PA	3rd Qu.:37.00
X3PA	Max. :57.00
X3Ppc	Min. :0.1710
X3Ppc	1st Qu.:0.2930
X3Ppc	Median :0.3520
X3Ppc	Mean :0.3506
X3Ppc	3rd Qu.:0.4100
X3Ppc	Max. :0.5650
FT	Min. : 6.00
FT	1st Qu.:14.00
FT	Median :18.00
FT	Mean :18.47
FT	3rd Qu.:23.00
FT	Max. :32.00
FTA	Min. : 7.00
FTA	1st Qu.:19.00
FTA	Median :23.00
FTA	Mean :23.46
FTA	3rd Qu.:29.00
FTA	Max. :39.00
FTpc	Min. :0.5160
FTpc	1st Qu.:0.7390
FTpc	Median :0.7860
FTpc	Mean :0.7886
FTpc	3rd Qu.:0.8800
FTpc	Max. :1.0000
ORB	Min. : 1.000
ORB	1st Qu.: 6.000
ORB	Median : 8.000
ORB	Mean : 8.957
ORB	3rd Qu.:12.000
ORB	Max. :18.000
TRB	Min. :26.00
TRB	1st Qu.:37.00
TRB	Median :41.00
TRB	Mean :41.13
TRB	3rd Qu.:45.00
TRB	Max. :61.00

Variable	Description
AST	Min. :12.00
AST	1st Qu.:20.00
AST	Median :23.00
AST	Mean :23.03
AST	3rd Qu.:25.25
AST	Max. :37.00
STL	Min. : 2.000
STL	1st Qu.: 6.000
STL	Median : 8.000
STL	Mean : 7.946
STL	3rd Qu.:10.000
STL	Max. :15.000
BLK	Min. : 0.000
BLK	1st Qu.: 2.000
BLK	Median : 3.000
BLK	Mean : 3.565
BLK	3rd Qu.: 5.000
BLK	Max. :10.000
TOV	Min. : 5.00
TOV	1st Qu.:12.00
TOV	Median :15.00
TOV	Mean :14.83
TOV	3rd Qu.:17.25
TOV	Max. :26.00
PF	Min. :11.0
PF	1st Qu.:19.0
PF	Median :22.0
PF	Mean :21.9
PF	3rd Qu.:25.0
PF	Max. :32.0
Dates	Min. :2019-10-22
Dates	1st Qu.:2019-12-07
Dates	Median :2020-01-28
Dates	Mean :2020-03-16
Dates	3rd Qu.:2020-08-08
Dates	Max. :2020-10-11

1.4 Number of columns and rows in dataset

```
length(Lakers)
```

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

```
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
```

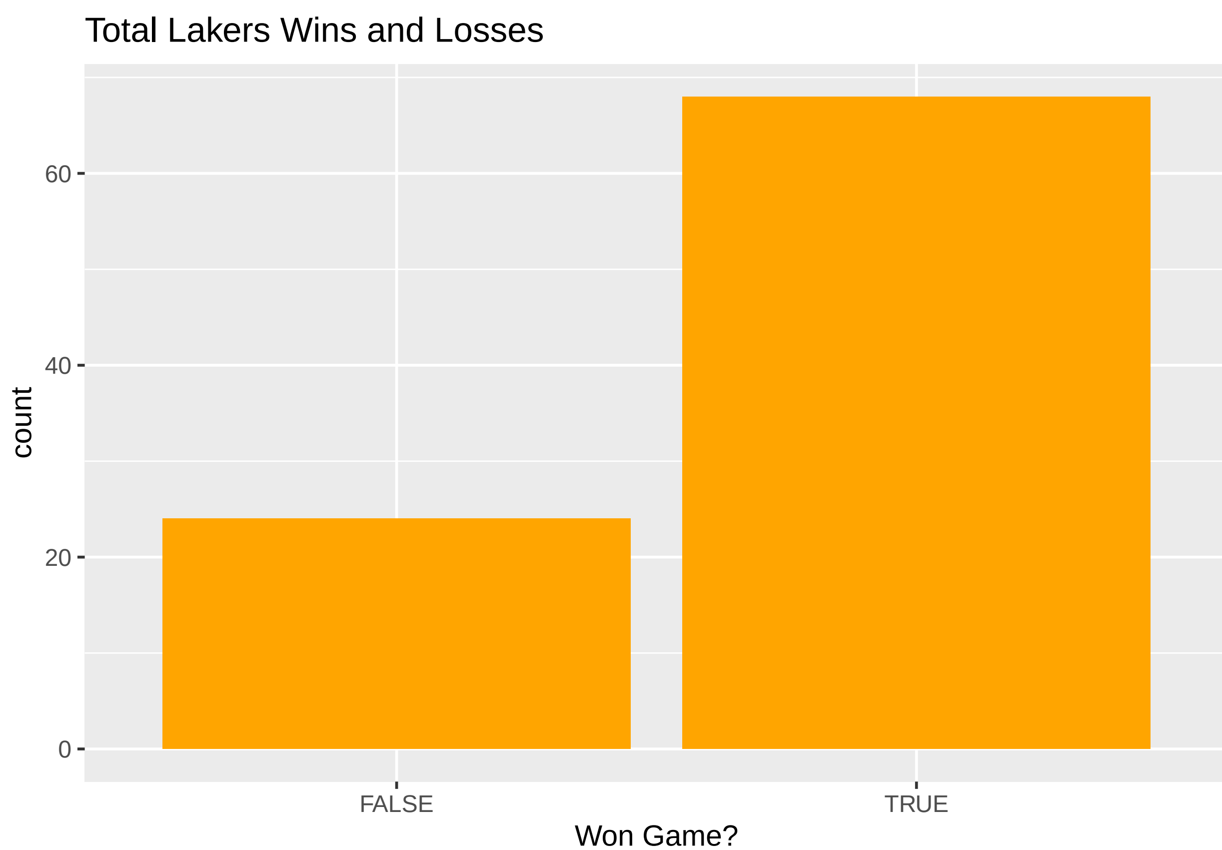
```
## [1] 68
```

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.

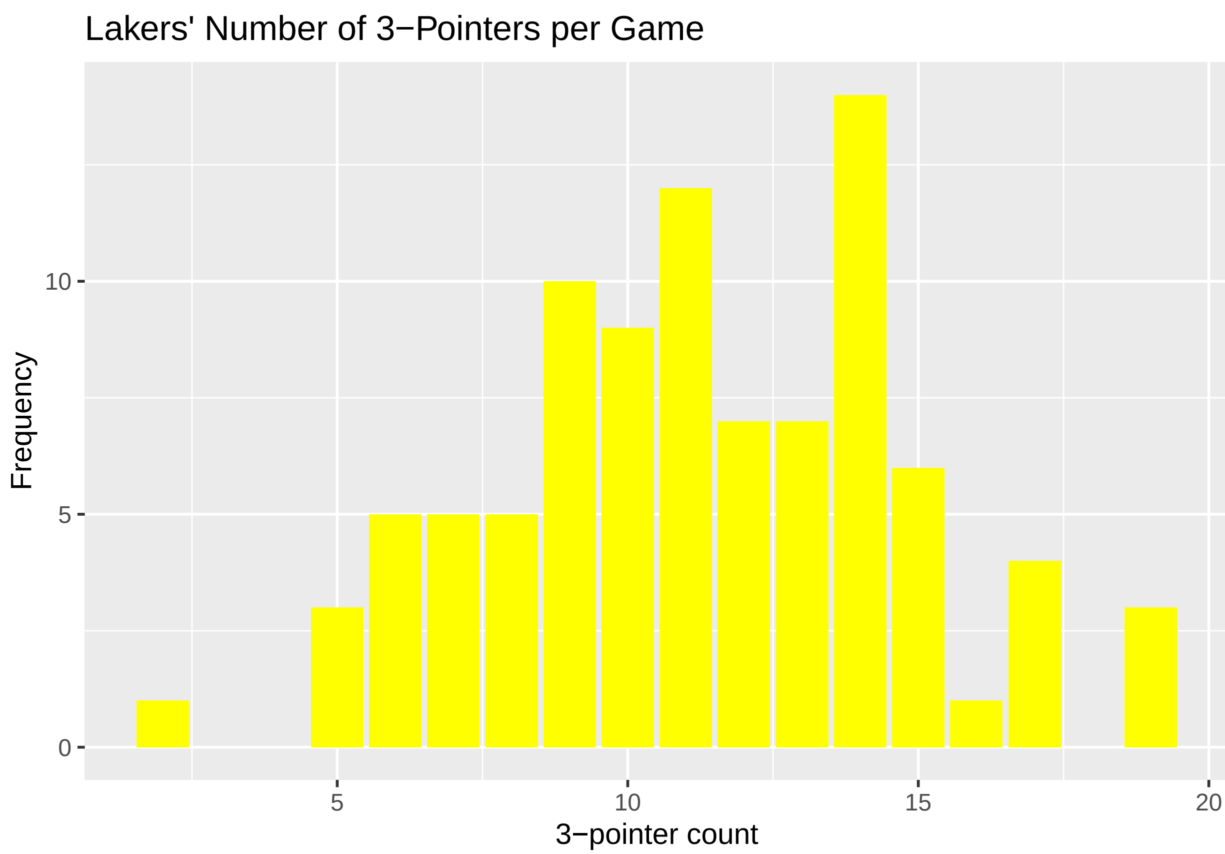
```
library(ggplot2)
Lakers %>% ggplot(aes(WonLost))+geom_bar(fill="orange")+
  xlab("Won Game?") +
  ggtitle("Total Lakers Wins and Losses")
```



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.

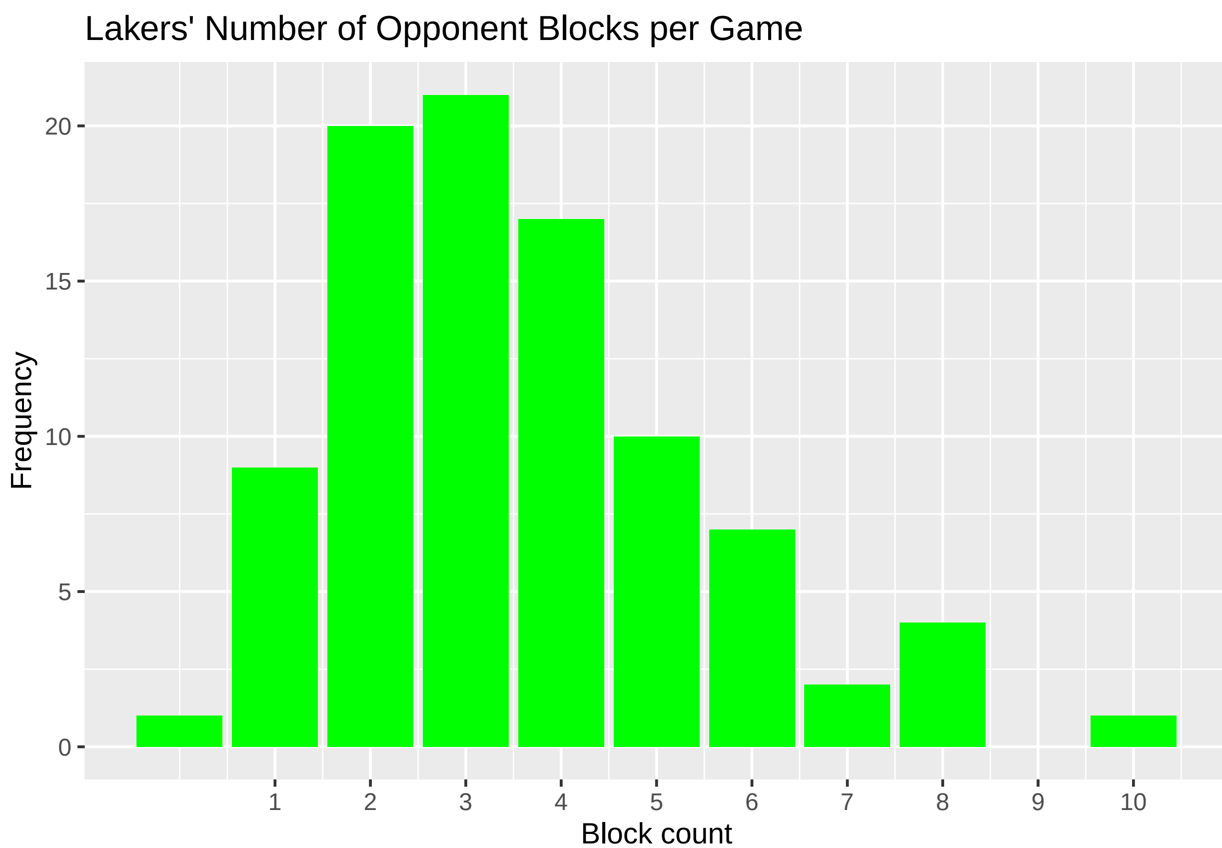
```
Lakers %>% ggplot(aes(X3P_LA))+geom_bar(fill="yellow")+
  xlab("3-pointer count") +
  ylab("Frequency")+
  ggtitle("Lakers' Number of 3-Pointers per Game")
```



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.

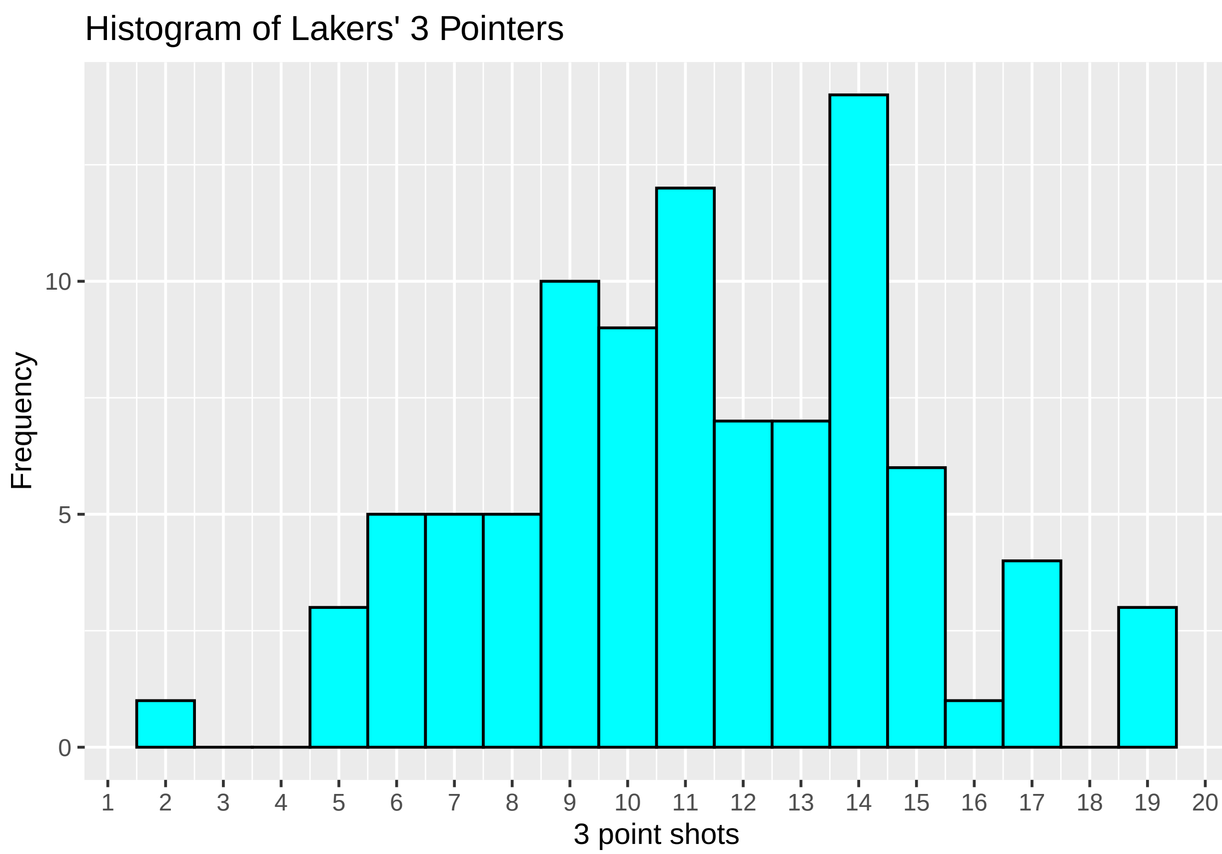
```
Lakers %>% ggplot(aes(BLK))+geom_bar(fill="green")+
  scale_x_continuous(breaks=c(1:10))+
  xlab("Block count") +
  ylab("Frequency")+
  ggtitle("Lakers' Number of Opponent Blocks per Game")
```



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.

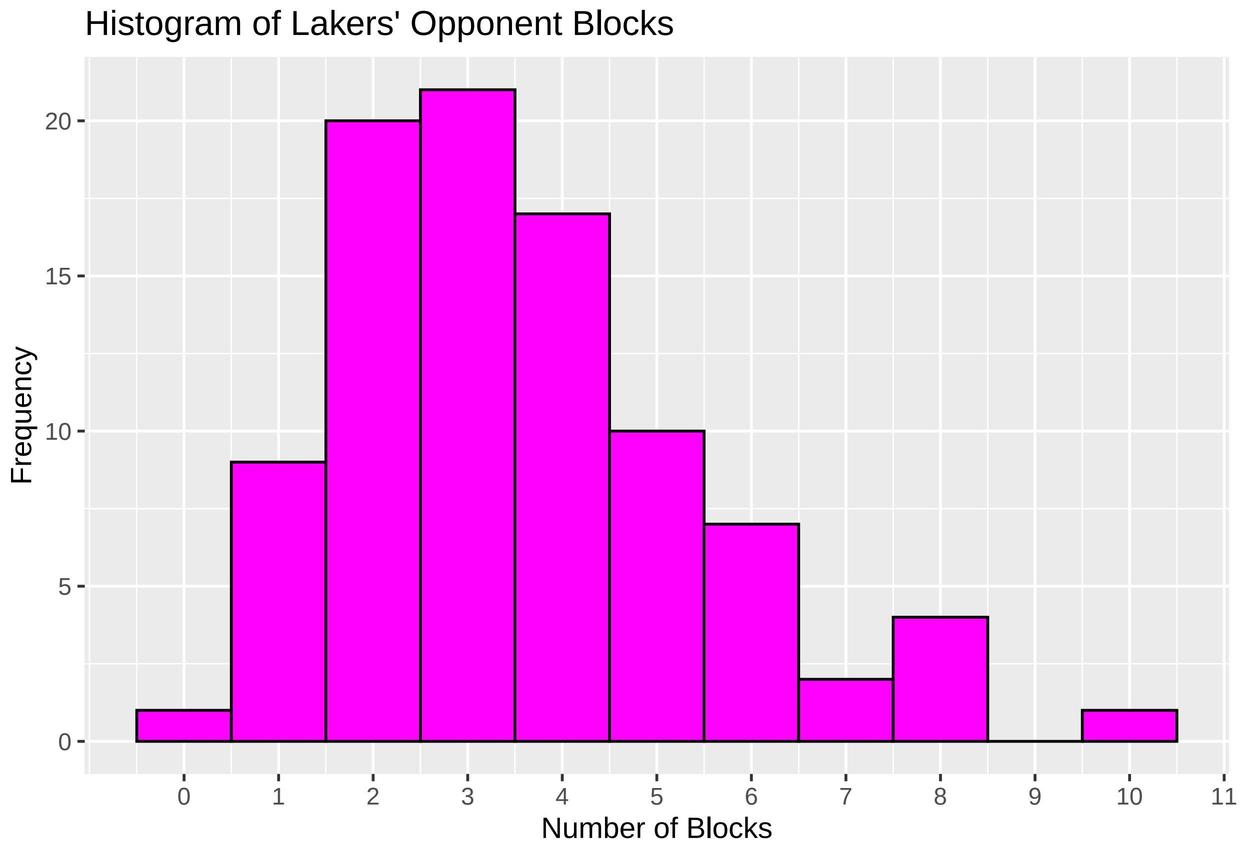
```
Lakers %>%
  ggplot(aes(X3P_LA)) +
  geom_histogram(binwidth = 1, fill="cyan", col = "black")+
  scale_x_continuous(breaks=c(0:20))+
  ylab("Frequency")+
  xlab("3 point shots")+
  ggtitle("Histogram of Lakers' 3 Pointers")
```



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.

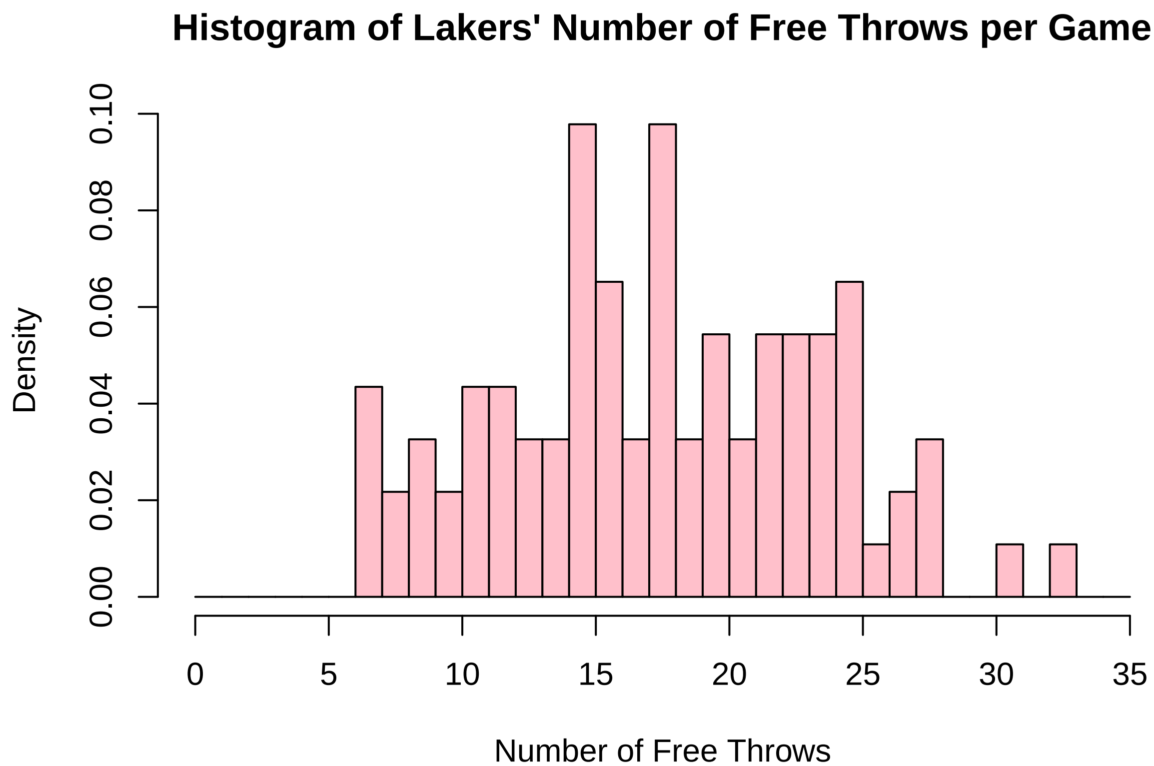
```
Lakers %>%
  ggplot(aes(BLK)) +
  geom_histogram(binwidth = 1, fill="magenta", col = "black")+
  scale_x_continuous(breaks=c(0:12))+
  ylab("Frequency")+
  xlab("Number of Blocks")+
  ggtitle("Histogram of Lakers' Opponent Blocks")
```



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
```



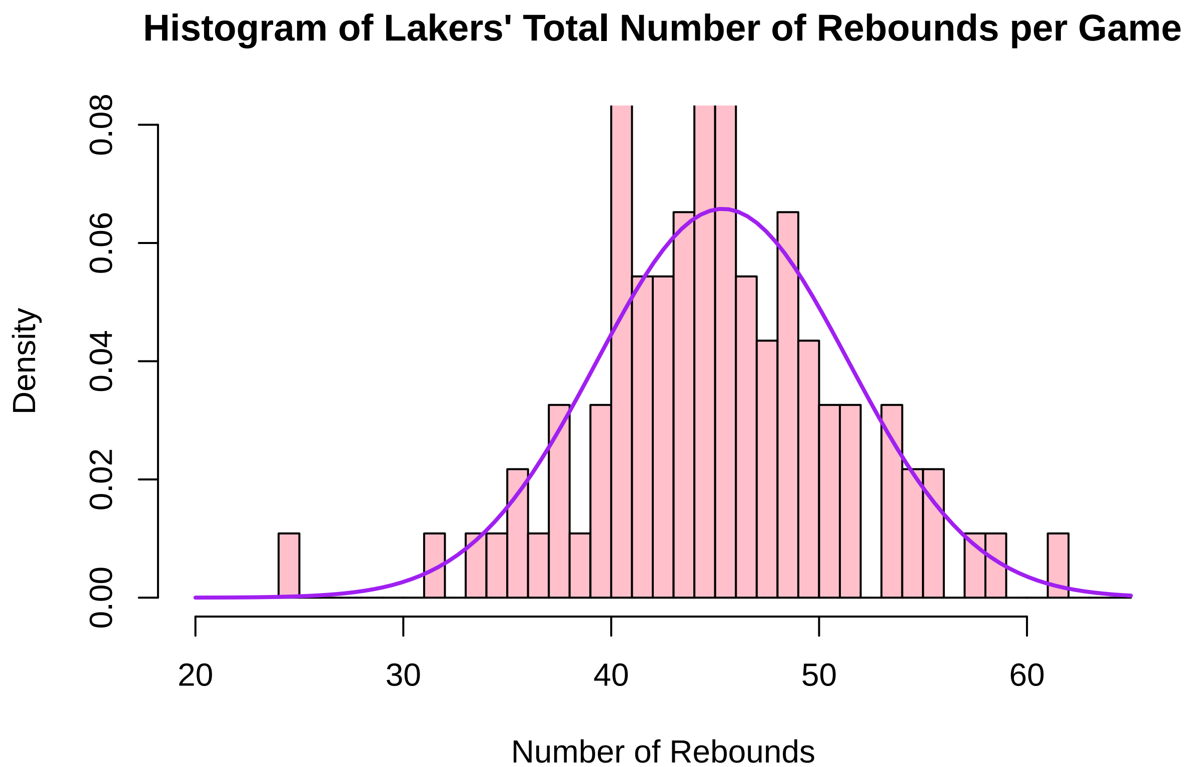

```
# free throws from the team per game
```

3 Probability analysis

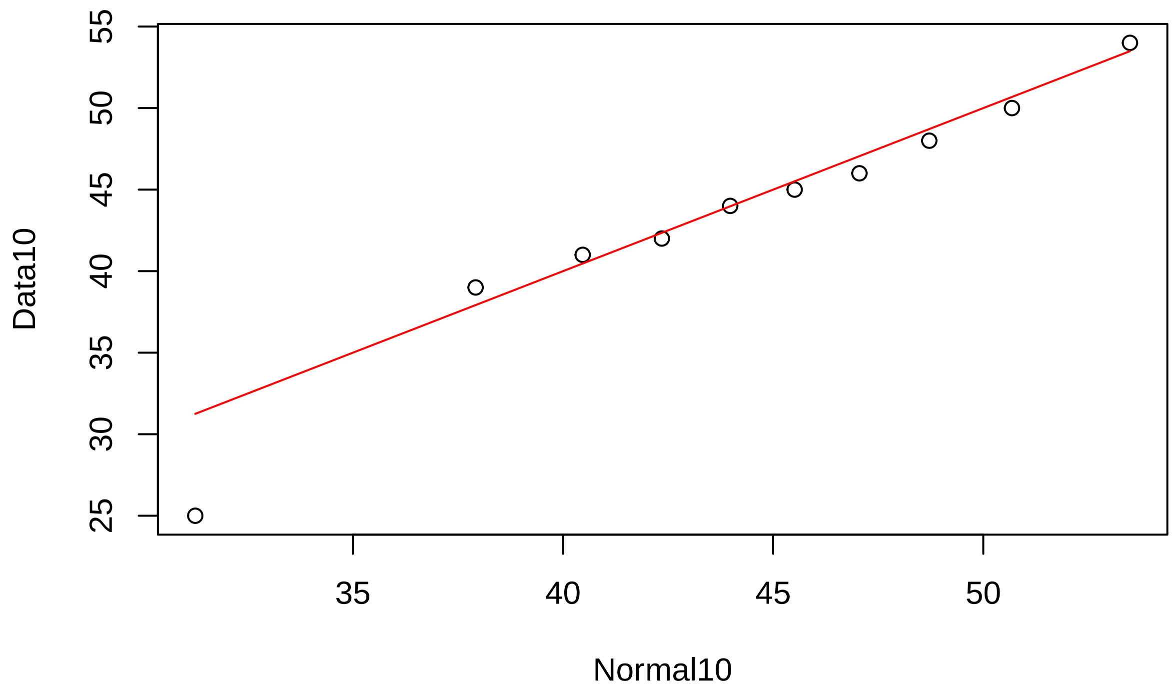
3.1 Probability density graph overlay on the our team's total number of rebounds in a game

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 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. 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.

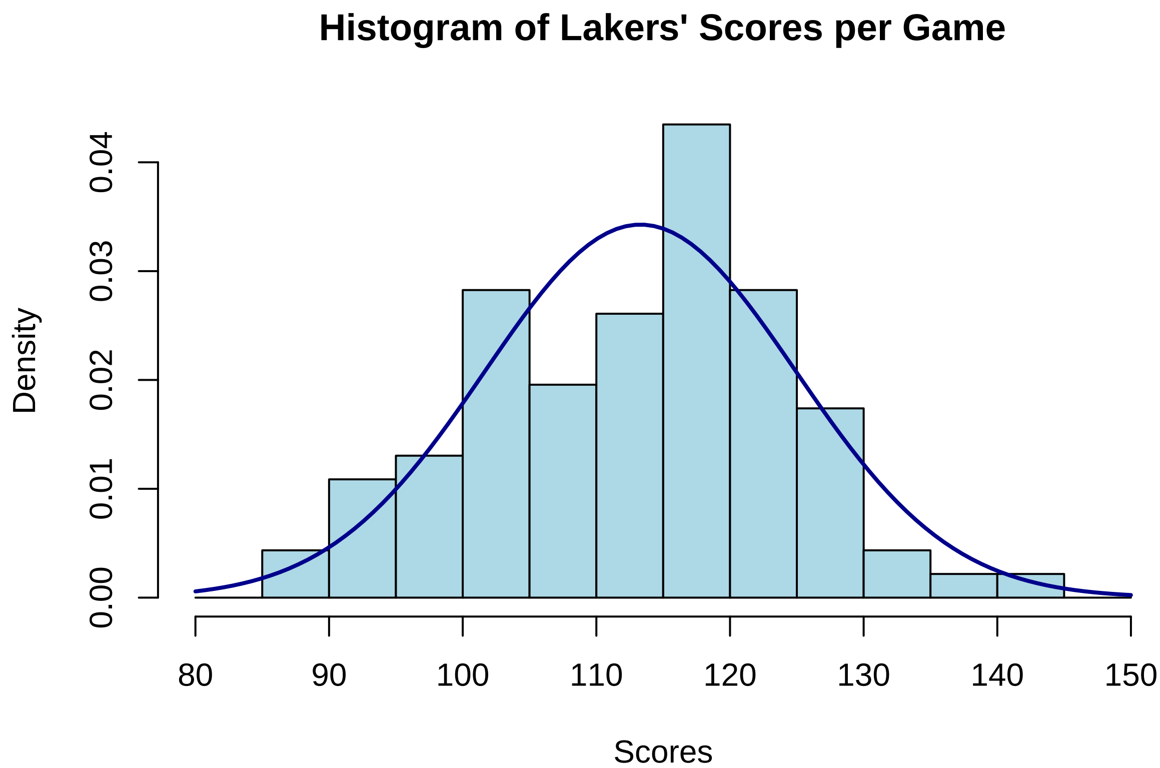
```
mu<-mean(Tm); mu
```

```
## [1] 113.2935
```

```
stdv<-sd(Tm); stdv
```

```
## [1] 11.63977
```

```
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)
```



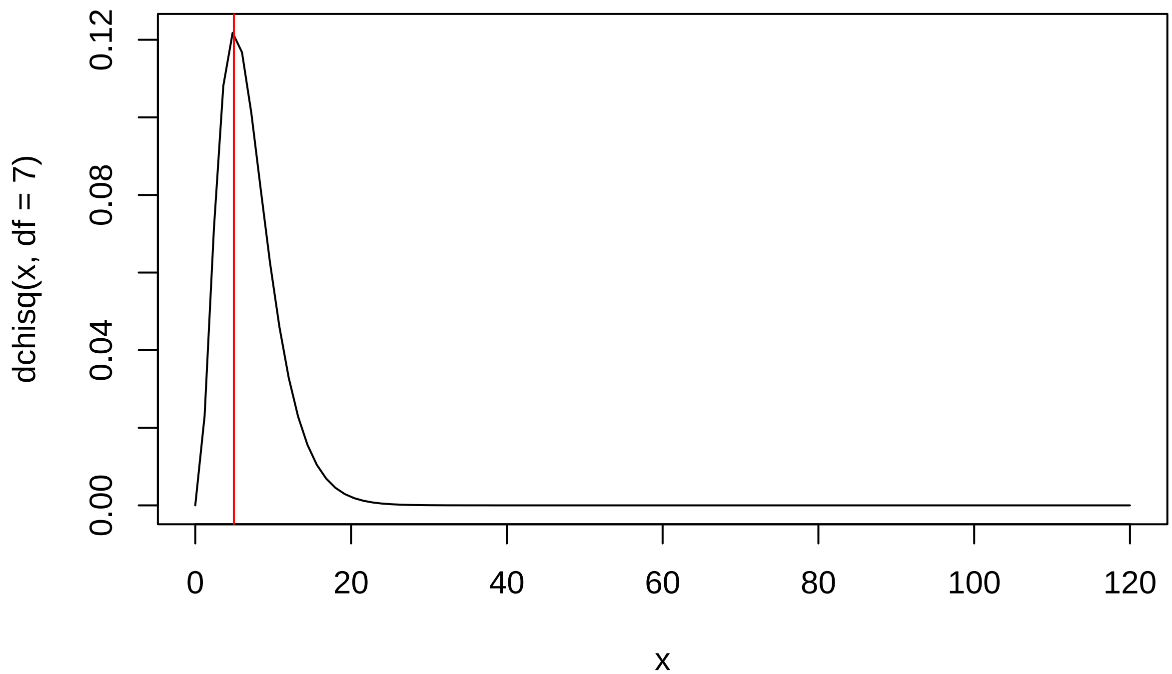

```
#Binning and Chisq test to compare this histogram to a Normal Distribution
dec <- qnorm(seq(0.0, 1, by = 0.1), mu, stdv)    #10 bins
Exp <- rep(length(Tm)/10,10)     #expected scores per bin
#Now we can count how many scores are in each bin
binscores <- numeric(10)
for (i in 1:10)
  binscores[i] <- sum((Tm >= dec[i]) & (Tm <= dec[i+1]) ); 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")
```




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

3.3 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.

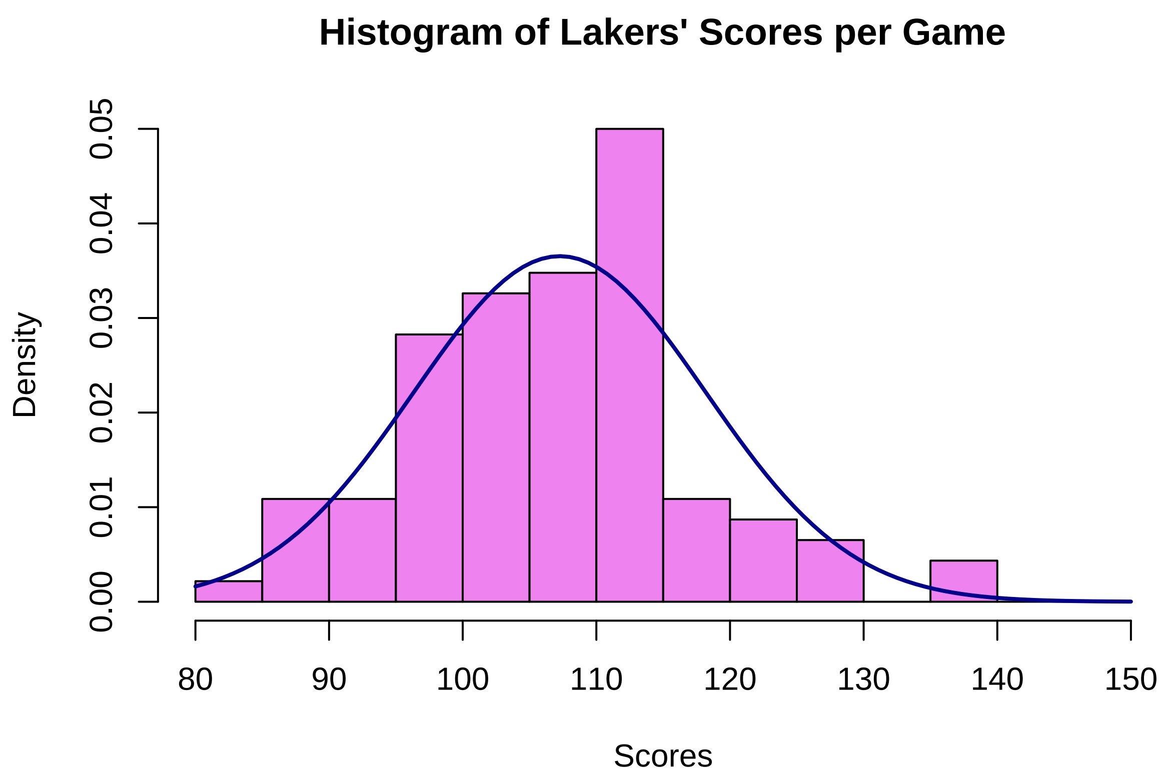
```
mu<-mean(Opp_pts); mu
```

```
## [1] 107.2609
```

```
stdv<-sd(Opp_pts); stdv
```

```
## [1] 10.91815
```

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



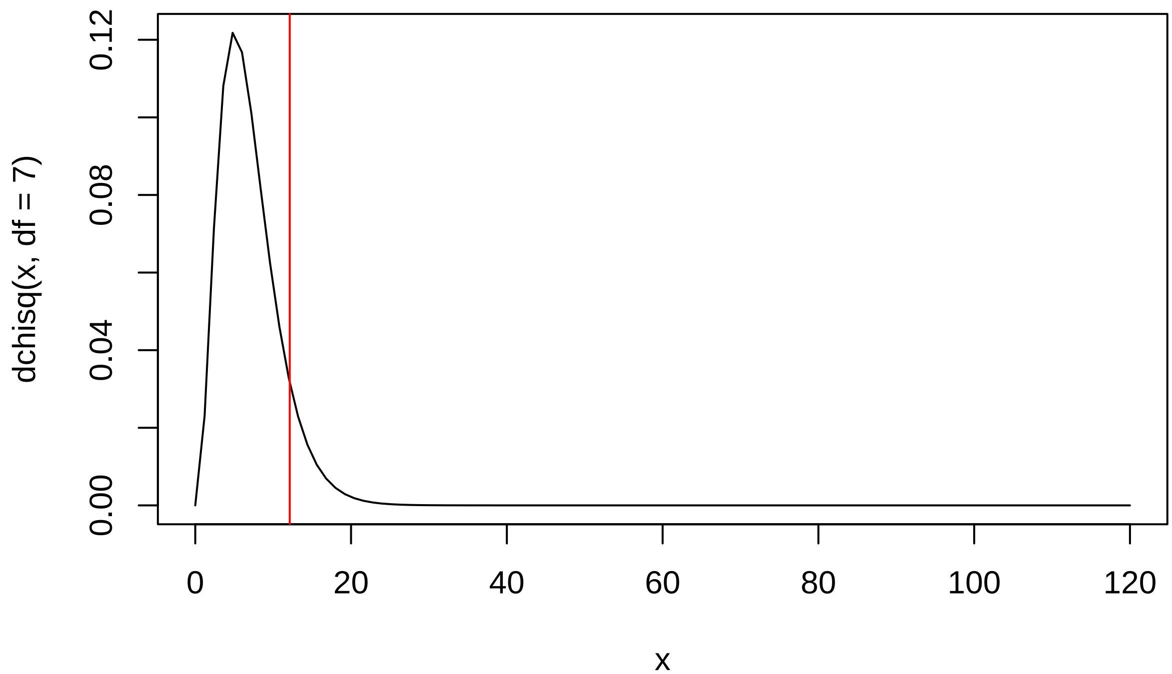

```
#Binning and Chisq test to compare this histogram to a Normal Distribution
dec <- qnorm(seq(0.0, 1, by = 0.1), mu, stdv)
Exp <- rep(length(Opp_pts)/10,10)
#Now we can count how many scores are in each bin
binscores <- numeric(10)
for (i in 1:10)
  binscores[i] <- sum((Opp_pts >= dec[i]) & (Opp_pts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

```
##  [1]  9 10  8  8  9 11  7 18  5  7
```

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

```
## [1] 12.13043
```

```
#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
```

3.4 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.

```
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
```

```
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.

```
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
```

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

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

```r
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]  1.7466949  1.4061343 -1.1723956  2.0386039  3.7414067 -0.5885775
```

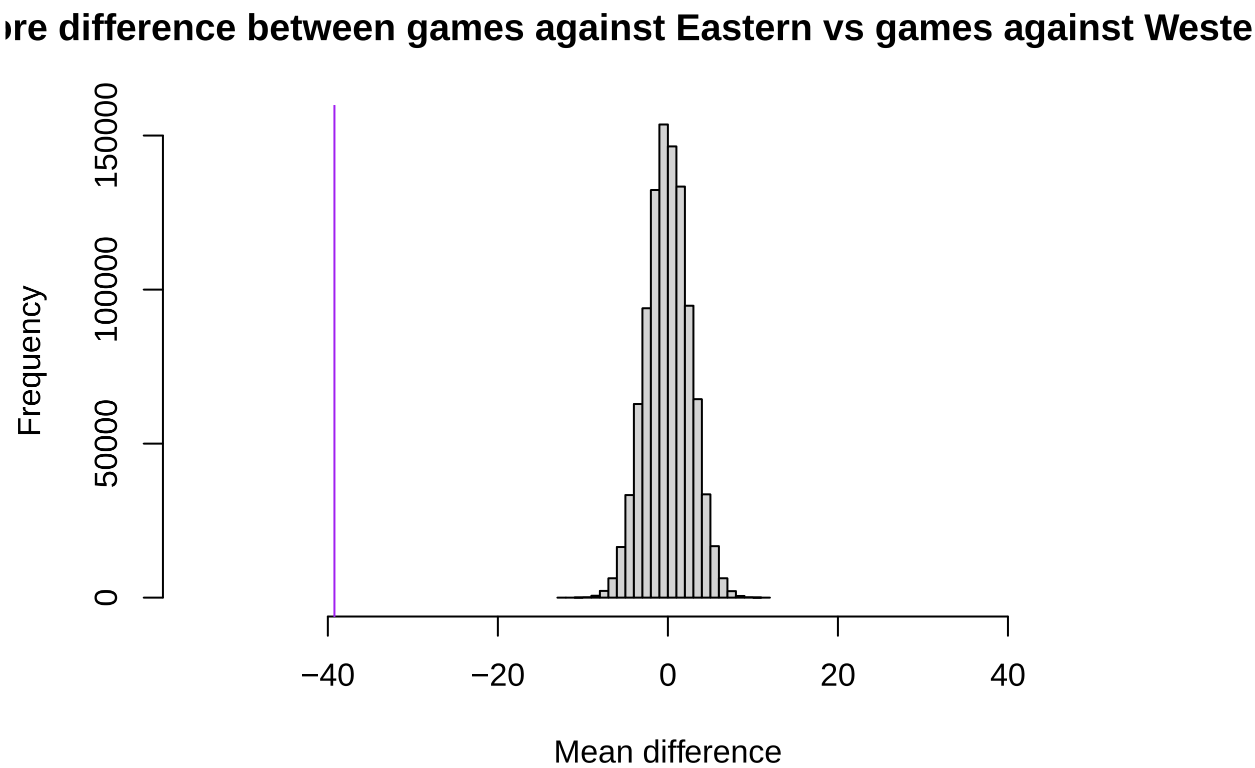
```r
summary(Score_diffs)
```

```
##       Min.    1st Qu.     Median       Mean    3rd Qu.       Max.
## -12.118985  -1.756214  -0.004759   0.005185   1.746695  11.817557
```

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

```
## [1] 0.005184884
```

```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.