Name: Aaron Matharoo	Lab: Section [redacted]
Subject/Course: [redacted]	TA: [redacted]
Professor: [redacted]	------

1. **Describing dataset, how you acquired it, why you selected it, the variable of interest, any measurement error, sources of sampling error, bias. Include histogram/table with basic summary statistics. Include dataset with R commands.**

The dataset in this analysis has been collected from an online game log of NBA players' game statistics. This dataset specifically shows the number of 3 point field goals scored by NBA athlete Stephen Curry over the 2020-2021 regular season to date. At this point in time (February 21, 2021), 30 games have transpired in the 2020-2021 regular season which is our sample size. The data has been acquired from an up-to-date online game log of current game and point statistics from athletes participating in the NBA. I decided to select this data on this specific athlete, Stephen Curry, for his well renowned reputation of scoring at and behind the 3 point field goal line on the court. The variable of interest in the data set is the number of 3 point field goals scored by Stephen Curry per game in the 2020-21 regular season to date. Measurement error in collecting this data is minimal to non-existent. This is because points are counted relatively objectively in the NBA as points matter as to where the player is standing on the court when the point is scored. In front of the 3 point line is 2 points and on or behind the line equals 3 points. One potential source of measurement error is if the referee did not see the footing of Stephen Curry at the time the shot was taken. However, this issue is mostly combatted by using replays of footage. Sampling error and bias are not major issues in this data set for the same reason that points are objectively viewed and collected & can be reviewed using replay footage to confirm the points scored.

Figure 1.

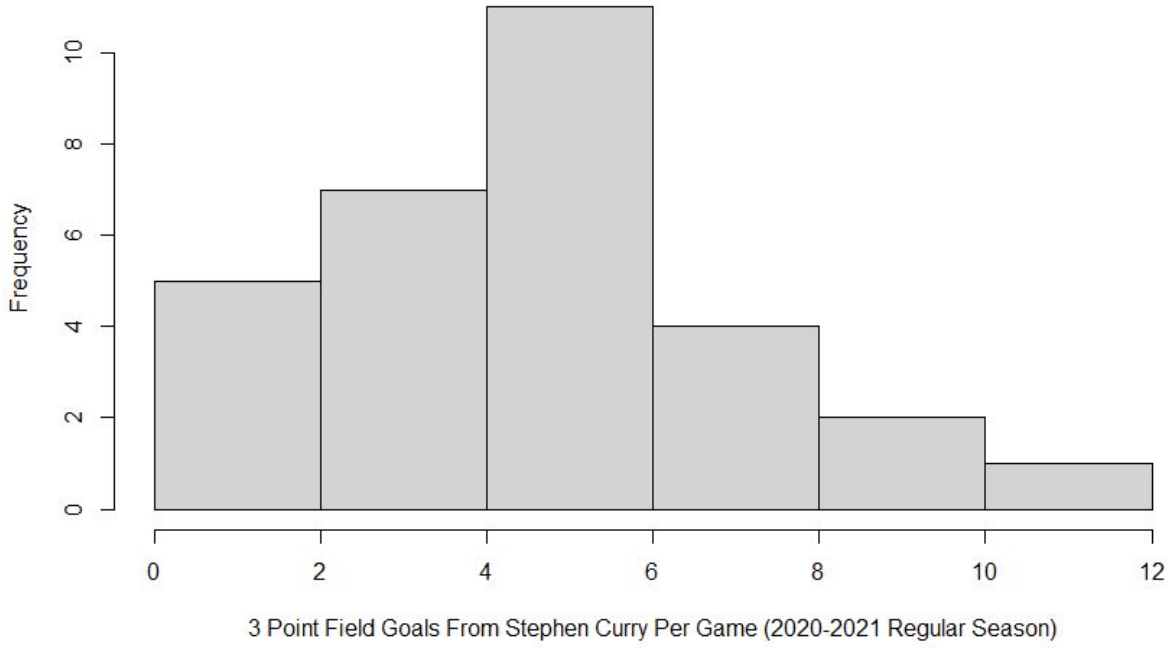


Table 1.
Summary Statistics

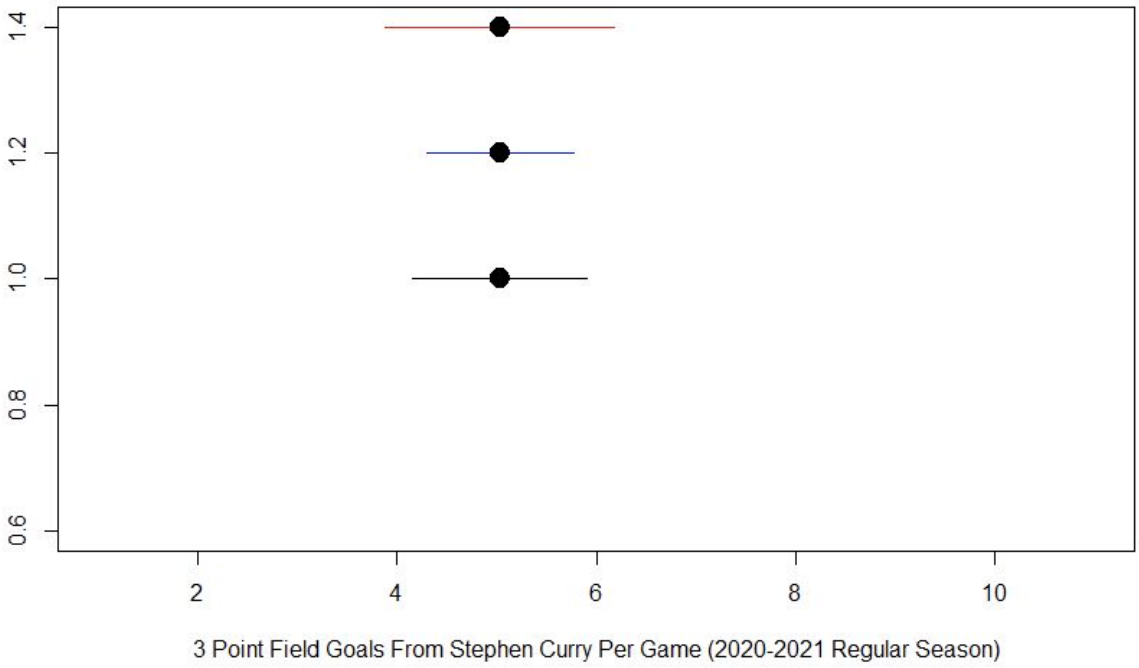
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1.000	4.000	5.000	5.033	6.000	11.000

COMMANDS

#Create Dataset (from https://www.basketball-reference.com/players/c/curryst01/gamelog/2021/)
> x = c(2,2,5,5,4,8,5,1,9,1,3,5,3,4,5,5,7,4,5,6,7,4,11,6,4,10,2,7,5,6)
#Plot Histogram
> hist(x, xlab = "3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)", main = "")
#Create Table with Summary Statistics
> summary(x)

2. **Create 90%, 95% and 99% confidence intervals for your data (using a sample mean). Plot these on a graph. Include a sentence describing what these mean and commands used to generate the answer.** [redacted]

Figure 2.



The intervals in this graph show different confidence levels for the same data. The narrower intervals show confidence that is lower for the same data which means that we are less sure about the interval. Higher probability values have less narrow & longer intervals. We are most confident of the 99% confidence interval of Curry's mean 3 point field goals.

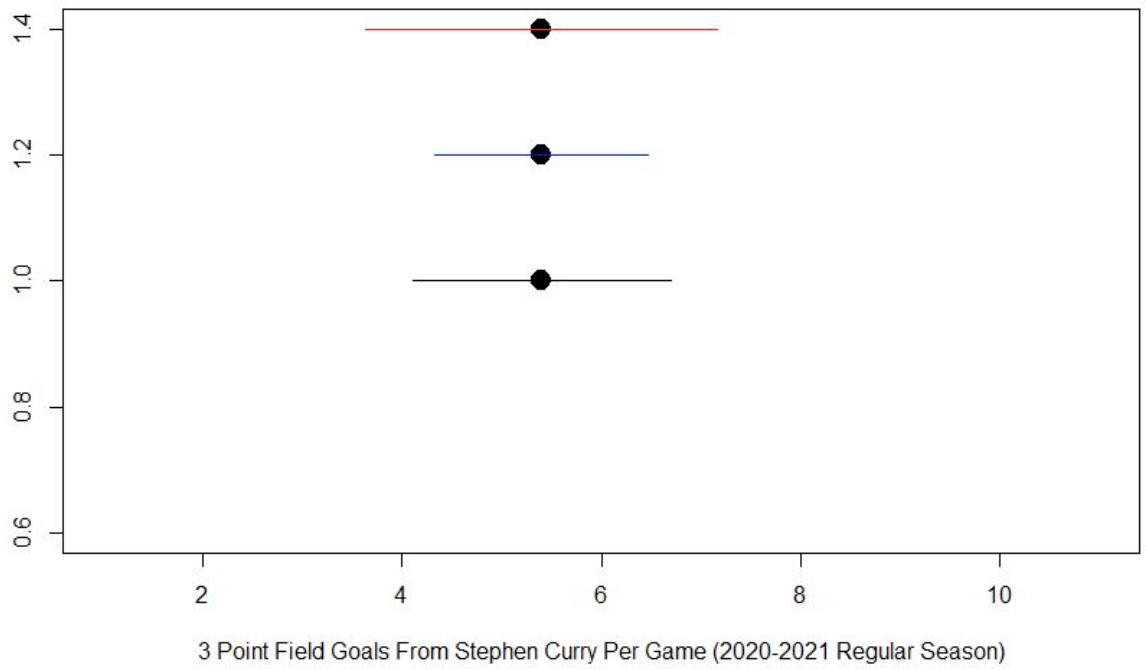
COMMANDS

```
#Define s
s = sd(x)
#Set Probability & Find Associated Z-Score to Use as Critical Value for 95%, 99%, & 90% Confidence Interval
	#95%
z_crit = qnorm(0.05/2, lower.tail = FALSE)
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
plot(x=c(lower, upper), y=c(1, 1), type="l", xlim = c(min(x),max(x)), xlab="3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)", ylab="", col = "Black")
points(x = mean(x), y = 1, pch=20, cex=3)
	#90%
z_crit = qnorm(0.10/2, lower.tail = FALSE)
```

```
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
lines(x = c(lower, upper), y=c(1.2, 1.2), col="blue")
points(x = mean(x), y = 1.2, pch=20, cex=3)
        #99%
z_crit = qnorm(0.01/2, lower.tail = FALSE)
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
lines(x = c(lower, upper), y=c(1.4, 1.4), col="red")
points(x = mean(x), y = 1.4, pch=20, cex=3)
```

3. **Repeat everything for question 2 but this time use a t-critical value (i.e., you will have to use the qt function to find the critical value). You can use the same or a different variable (than that used in question 2) to analyze - use a sample to create a random sample of size n = 20 if you have more than 20 observations in your dataset.**

Figure 3.



The intervals in this graph show confidence intervals for a random sample of 20 from the original data set. This graph uses the t-distribution to find the critical value since the sample size is smaller. For a higher probability value, we are more sure about the interval as the range within

the interval is larger. The narrow intervals are for lower probability values, showing us we are less sure about the interval. Here we are least confident about the 90% confidence interval for Curry's mean 3 point field goals.

COMMANDS

```
#Create Sample Dataset
sample = sample(x, 20)
#Define s_sample
s_sample = sd(sample)
#Set Probability & Find Associated T-Distribution to Use as Critical Value for 95%, 99%, & 90% Confidence Interval
        #95%
t_crit = qt(1-0.05/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
plot(x=c(lower_s, upper_s), y=c(1, 1), type="l", xlim = c(min(sample),max(sample)), xlab="3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)", ylab="", col = "Black")
points(x = mean(sample), y = 1, pch=20, cex=3)
        #90%
t_crit = qt(1-0.10/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
points(x = mean(sample), y = 1.2, pch=20, cex=3)
lines(x = c(lower_s, upper_s), y=c(1.2, 1.2), col="blue")
        #99%
t_crit = qt(1-0.01/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
points(x = mean(sample), y = 1.4, pch=20, cex=3)
lines(x = c(lower_s, upper_s), y=c(1.4,1.4), col="red")
```

4. **Conduct your own analysis of the dataset. This could include calculating z-scores to find unusual values or groups of unusual values and any plots you want to explore. Write a paragraph commenting on the results and interpreting your findings.** [redacted]

Z-Score of Curry Scoring 0 *3 Point Field Goals*: -2.060979
Z-Score of Curry Scoring 5 *3 Point Field Goals*: -0.01364886
Z-Score of Curry Scoring 8 *3 Point Field Goals*: 1.214749

Figure 5.

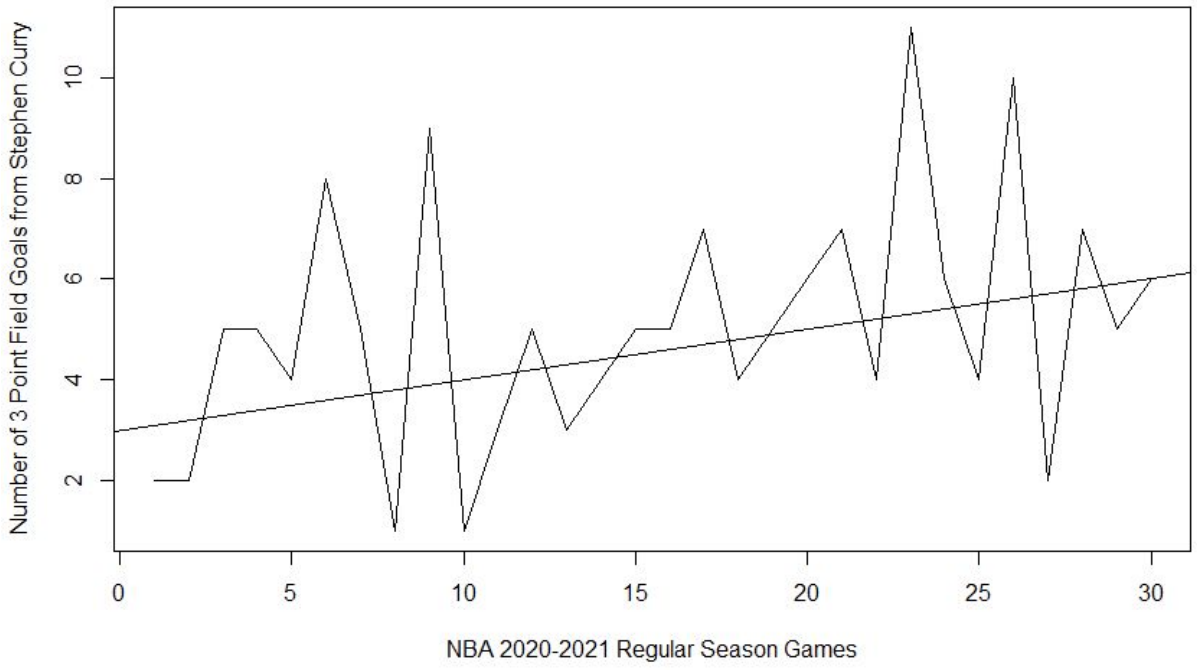
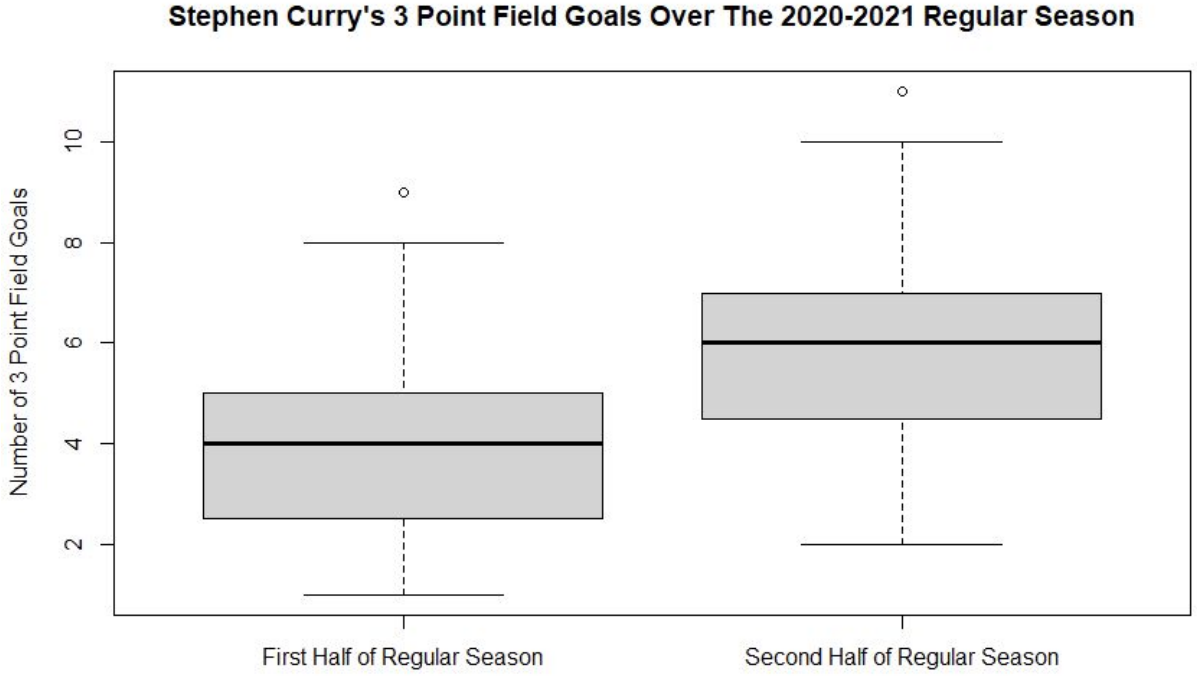


Figure 6.



3 Z-Scores have been calculated for this data set. The Z-Scores tell us how far away the value at hand is away from the mean. For a value of 0 we have a Z-Score of -2.06 which is relatively extremely far away from the mean. This shows that Stephen Curry scoring zero 3 point field goals in the 2020-2021 regular season is unlike his typical game-play since it is extremely far from the mean. For a value of 5 we have a Z-Score of -0.01. This value is extremely close to the mean and shows us that Stephen Curry typically can score five 3 point field goals playing in the 2020-2021 regular season. For a value of 8 we have a Z-Score of 1.21. This value is relatively far from the mean however it is not the farthest out of the 3 Z-Scores computed. We can see from this data that the Z-Score associated with eight 3 point field goals is closer to the mean than zero 3 point field goals. This displays Stephen Curry's immense athletic ability to score from the 3 point line, as he is closer to his mean when scoring eight 3 point field goals compared to when he scores none at all. Figure 5 shows a line graph displaying Stephen Curry's 3 point field goals over each game of the 2020-2021 regular season. By fitting a line of best fit, we can see the trend of the graph is positive. It seems that as the season progresses, Stephen Curry improves his ability to score from the 3 point field goal line, perhaps due to increased practice and repetitive shooting activity from all the games that have transpired thus far. Figure 6 helps us to see this in more depth. The box plots show Stephen Curry and his 3 point field goals over the first half and second half of the 2020-2021 regular season. The conclusions drawn from the box plots are astounding as we can see for the first half of the season Curry had a mean of four 3 point field goals which rose to 6 by the next half of the season. Not only did his mean score increase, but so did Curry's maximum and minimum total of 3 point field goals. Stephen Curry's 3 point field goals overall increased by a factor of about 2 from the first half to the second half of the 2020-2021 regular season, including his anomalies which lead him to score a high of eleven 3 point field goals in the second half of the season. This shows that Curry has been improving regarding his ability to score from the 3 point field goal line.

APPENDIX:

```
# [redacted] Lab [redacted]

#Create Dataset
x = c(2,2,5,5,4,8,5,1,9,1,3,5,3,4,5,5,7,4,5,6,7,4,11,6,4,10,2,7,5,6)

#Plot Histogram
hist(x, xlab = "3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)",
main = "")
#Create Table with Summary Statistics
summary(x)

#Define s
s = sd(x)
```

```
#Set Probability & Find Associated Z-Score to Use as Critical Value for 95%, 99%, & 90% Confidence Interval
  #95%
z_crit = qnorm(0.05/2, lower.tail = FALSE)
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
plot(x=c(lower, upper), y=c(1, 1), type="l", xlim = c(min(x),max(x)), xlab="3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)", ylab="", col = "Black")
points(x = mean(x), y = 1, pch=20, cex=3)
  #90%
z_crit = qnorm(0.10/2, lower.tail = FALSE)
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
lines(x = c(lower, upper), y=c(1.2, 1.2), col="blue")
points(x = mean(x), y = 1.2, pch=20, cex=3)
  #99%
z_crit = qnorm(0.01/2, lower.tail = FALSE)
lower = mean(x) - (z_crit * (s/sqrt(30)))
upper = mean(x) + (z_crit * (s/sqrt(30)))
lines(x = c(lower, upper), y=c(1.4, 1.4), col="red")
points(x = mean(x), y = 1.4, pch=20, cex=3)

#Create Sample Dataset
sample = sample(x, 20)
#Define s_sample
s_sample = sd(sample)
#Set Probability & Find Associated T-Distribution to Use as Critical Value for 95%, 99%, & 90% Confidence Interval
  #95%
t_crit = qt(1-0.05/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
plot(x=c(lower_s, upper_s), y=c(1, 1), type="l", xlim = c(min(sample),max(sample)), xlab="3 Point Field Goals From Stephen Curry Per Game (2020-2021 Regular Season)", ylab="", col = "Black")
points(x = mean(sample), y = 1, pch=20, cex=3)
  #90%
t_crit = qt(1-0.10/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
```

```
points(x = mean(sample), y = 1.2, pch=20, cex=3)
lines(x = c(lower_s, upper_s), y=c(1.2, 1.2), col="blue")
  #99%
t_crit = qt(1-0.01/2, 19, lower.tail = FALSE)
lower_s = mean(sample) - (t_crit * (s_sample/sqrt(20)))
upper_s = mean(sample) + (t_crit * (s_sample/sqrt(20)))
points(x = mean(sample), y = 1.4, pch=20, cex=3)
lines(x = c(lower_s, upper_s), y=c(1.4,1.4), col="red")

#Analysis
#Z-Scores
  #Use Z-Score To Find Probability of Curry Scoring zero 3 Point Field Goals
(0 - mean(x)) / sd(x)
  #Use Z-Score To Find Probability of Curry Scoring five 3 Point Field Goals
(5 - mean(x)) / sd(x)
  #Use Z-Score To Find Probability of Curry Scoring eight 3 Point Field Goals
(8 - mean(x)) / sd(x)
#Line Graph
  #Define games
games = 1:30
  #Plot Line Graph to Note Trends
plot(games, x, xlab = "NBA 2020-2021 Regular Season Games", type = "l", ylab = "Number of 3
Point Field Goals from Stephen Curry")
  #Add ~line of best fit
abline(3, 0.1)

#Box Plots
  #Define First Half Of Regular Season
x_firsthalf = c(2,2,5,5,4,8,5,1,9,1,3,5,3,4,5)
  #Define Second Half Of Regular Season
x_secondhalf = c(5,7,4,5,6,7,4,11,6,4,10,2,7,5,6)
  #Plot Box Plots
season_half <- c("First Half of Regular Season", "Second Half of Regular Season")
boxplot(x_firsthalf, x_secondhalf, names = season_half, main = "Stephen Curry's 3 Point Field
Goals Over The 2020-2021 Regular Season", ylab = "Number of 3 Point Field Goals")
```