

### **Introduction:**

On March 11th, 2020, the NBA announced that a player on the Utah Jazz has tested positive for COVID-19 which subsequently led to the suspension of play (Aschburner). On July 30th, 2020, the NBA season restarted but inside a bubble instead of team arenas. The idea of the NBA bubble was to seclude all players, coaches, and team/league personnel into one area to prevent the spread of COVID-19. The NBA invited twenty-two teams to participate in the NBA bubble to finish the rest of the season: sixteen teams that were in the playoffs based on NBA standings when the season was suspended in March and the six teams that were six games or fewer behind the 8th seed in their respective conferences (Haislop). The 2019-2020 season successfully concluded within the bubble and a champion was crowned, however, COVID-19 was still prevalent and would seep into the next season. The 2020-2021 season was not played inside the bubble but would face several environmental and travel changes due to the health and safety protocols: a shortened 72 game season, limits on traveling parties, suspensions for players who break COVID-19 protocols, suspensions for players who contract COVID-19, and no fan presence inside NBA arenas (ESPN).

Studies on the uniqueness of the 2019-2020 season and 2020-2021 season have already been done. For example, a research project under *Scientific Reports* utilized the NBA's COVID-19 restart "bubble" to uncover the impact of travel and circadian disruption on athletic performance and a team's ability to win regular season games (McHill and Chinoy). This explored which aspects of regular season team performance changed when teams were traveling within and across time zones, and when no teams traveled in the bubble (eFG%, ORB%, defensive rating, etc.). In short, the combination of travel and circadian disruption during the regular season might hinder the away team's ability to rebound and shoot accurately.

## COVID-19 and the NBA

Utilizing the context of the COVID-19 pandemic, we aimed to examine how environmental and travel restrictions impacted the players and teams when comparing the 2019-2020 season with the 2020-2021 season. Sports teams and statisticians could benefit from this research because it can further reveal the underlying effects of home court advantage or the dominance of overall talent. For example, a player or team may be more affected by a certain atmosphere whether that be moral support from the home team or hostility towards the away team. Other examples aside from the team may be how the games are shifted due to the referees being subconsciously swayed by a specific environment. Contrastingly, teams may even show minimal effects when playing at home or away.

Using this information, sports teams and statisticians can be better suited in making certain decisions about the players and teams that will be prevalent as the NBA lifts its restrictions regarding COVID-19. By analyzing the changes from the 2019-2020 season to the 2020-2021 season, sports teams can gather an understanding of how being out of the bubble affected their players' performance. The trends found in the analysis between the two seasons can carry through to future seasons.

**Data Collection:**

The final datasets used in this analysis were found on *NBAStuffer* (2019-2020 NBA Regular Season Player Stats, 2019-2020 NBA Playoffs Player Stats, 2020-2021 NBA Regular Season Player Stats, and 2020-2021 NBA Playoffs Player Stats) and *Basketball-Reference* (2018-2019 NBA Team Stats, 2019-2020 NBA Team Stats, and 2020-2021 NBA Team Stats). Before these files were read into R, several columns were renamed because of its invalid characters and also for readability.

For the datasets that consisted of player regular season statistics, columns such as rank, position, age, minutes per game, minutes percentage, free throw attempts, two point attempts, two point percentage, three point attempts, three point percentage, total rebound percentage, and assist percentage were dropped. These columns were dropped either because they were not going to be used in the analysis or if there was potential for collinearity. For example, keeping both assists per game and assists percentage could cause collinearity. Also, instances where there were null values (NA) were removed because they are considered incomplete observations. After the columns were dropped, duplicates were removed. If the player was on two teams, the team that had more games played was kept. Then, players that did not play in both seasons were removed.

For the datasets that consisted of team statistics, columns such as rank, age, pythagorean wins, pythagorean losses, margin of victory, strength of schedule, simple rating system, adjusted net rating, free throw rate, three point attempt rate, column 18, column 23, column 24, column 25, defensive rebound percentage, column 27, column 28, arena, and attendance. These columns were dropped either because they were not going to be used in the analysis or if there was potential for collinearity. For example, the offensive four factors were kept and the defensive four factors were removed. The last row of the team statistics which showed league averages was

removed because it had null values and is considered an extraneous variable. Lastly, there were some teams with “NA” under attendance per game which were set to 0.

### **Descriptive Statistics:**

The player data between the 2019-2020 season and the 2020-2021 season are graphically and numerically displayed to discover any obvious trends from the environmental and travel changes implemented due to COVID-19.

## **Summary Statistics**

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
GP	417	47.029	20.066	1	33	63	74
USG_percentage	417	18.932	5.496	7	15.1	22.4	37.5
TO_percentage	417	12.686	4.651	0	10.1	14.7	50
FT_percentage	417	0.73	0.163	0	0.667	0.828	1
eFG_percentage	417	0.523	0.08	0	0.484	0.559	0.786
TS_percentage	417	0.554	0.073	0	0.516	0.593	0.751
PPG	417	10.099	6.429	0	5.4	13.7	34.3
RPG	417	4.101	2.501	0.1	2.3	5.2	15.8
APG	417	2.213	1.841	0	0.9	3	10.2
SPG	417	0.7	0.389	0	0.4	0.93	2.11
BPG	417	0.445	0.422	0	0.17	0.56	2.93
TOPG	417	1.266	0.838	0	0.7	1.63	4.8
VIVersatility	417	7.361	2.069	0	5.9	8.4	17.2
ORTG	417	109.458	11.391	27.4	103.3	115.9	141.7
DRTG	417	104.353	5.34	77.2	101.3	108.1	118.5

**Fig. 1: 2019-2020 Season Player Statistics**

Fig.1. displays the summary statistics for players in the 2019-2020 regular season. The mean, standard deviation, and quartiles are given for each metric that was kept in the dataset.

## Summary Statistics

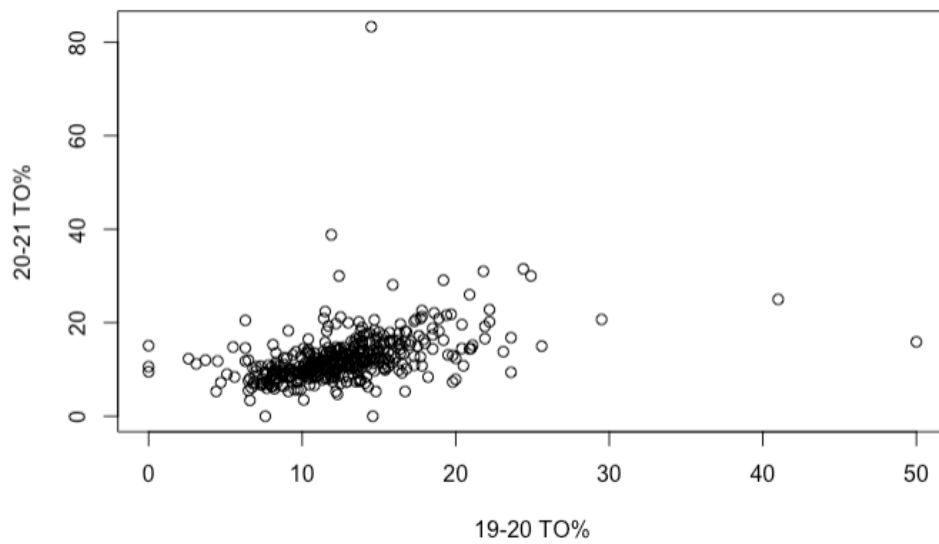
Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
GP	417	43.341	20.542	1	27	61	72
USG_percentage	417	18.797	5.772	3	14.8	22.1	39.4
TO_percentage	417	12.435	5.824	0	9.3	14.4	83.3
FT_percentage	417	0.735	0.183	0	0.669	0.845	1
eFG_percentage	417	0.527	0.093	0	0.491	0.573	1
TS_percentage	417	0.556	0.092	0	0.524	0.604	1.078
PPG	417	9.988	6.662	0	5	13.6	32
RPG	417	3.988	2.434	0.4	2.2	5.2	14.3
APG	417	2.217	1.933	0	0.9	3	11.7
SPG	417	0.682	0.384	0	0.4	0.92	2.08
BPG	417	0.444	0.425	0	0.16	0.6	3.38
TOPG	417	1.197	0.844	0	0.62	1.58	5
VIVersatility	417	7.217	2.148	0	5.8	8.4	15.8
ORTG	417	110.2	16.265	0	103.8	118.6	231.8
DRTG	417	106.346	5.362	83.2	103.5	110.1	117.4

**Fig. 2: 2020-2021 Season Player Statistics**

Fig.2. displays the summary statistics for players in the 2020-2021 regular season. The mean, standard deviation, and quartiles are given for each metric that was kept in the dataset. When analyzing the summary statistics, the mean values for half of the metrics lowered from the 2019-2020 season heading into the 2020-2021 season: GP, USG\_percentage, TO\_percentage, PPG, RPG, SPG, BPG, TOPG, VIVersatility. The standard deviations for the 2020-2021 season were greater than the standard deviations from the 2019-2020 season in all metrics except for: RPG and SPG. The minimum values for ORTG were significantly different from one another in each season: 0 and 27.4. There was an increase in minimum value in the 2020-2021 season compared to the 2019-2020 in these metrics: USG\_percentage, RPG, ORTG. There was a decrease in minimum value in the 2020-2021 season compared to the 2019-2020 season in this metric: DRTG. There was an increase in maximum value in the 2020-2021 season compared to

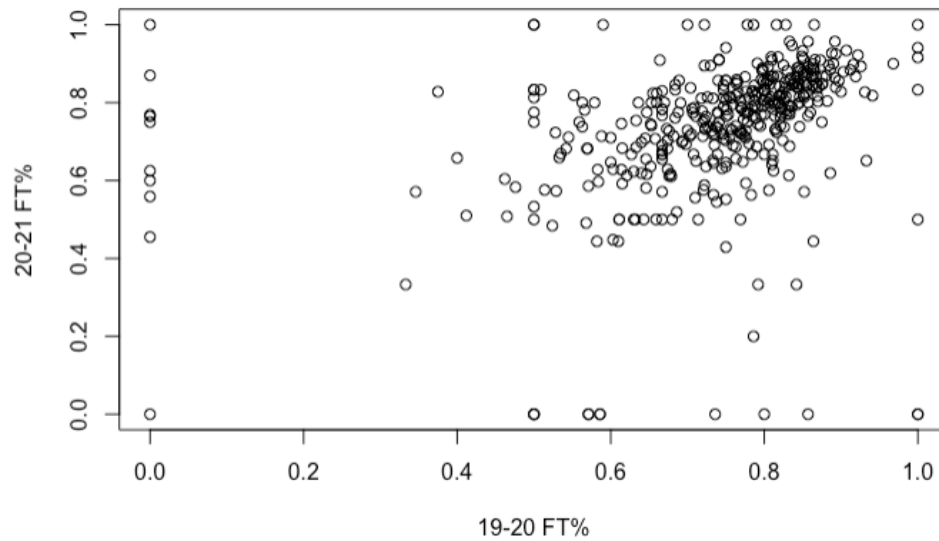
the 2019-2020 season in these metrics: USG\_percentage, TO\_percentage, eFG\_percentage, TS\_percentage, APG, BPG, TOPG, and ORTG. There was a decrease in maximum value in the 2020-2021 season compared to the 2019-2020 season in these metrics: GP, PPG, RPG, SPG, VIVersatility, and DRTG. The maximum value that stayed the same between the two seasons occurred in this metric: FT\_percentage.

Below are the scatterplots that displayed low correlations and displayed moderately-weak to weak relationship between the metric in both seasons. This means that these metrics were not repeatable unlike the other metrics.



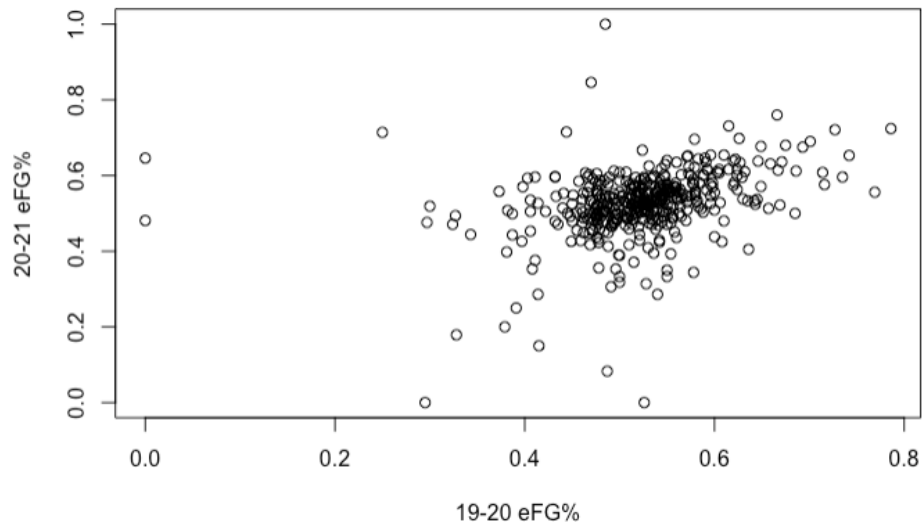
**Fig. 3: 2019-2020 TO% vs 2020-2021 TO%**

Fig. 3. displays the scatterplot for turnover percentage between the 2019-2020 and 2020-2021 regular season. The correlation for turnover percentage is 0.4126832. This scatterplot indicates a linear, moderately weak, positive relationship for turnover percentage between the two seasons. The scatterplot is most dense at around 13.



**Fig. 4: 2019-2020 FT% vs 2020-2021 FT%**

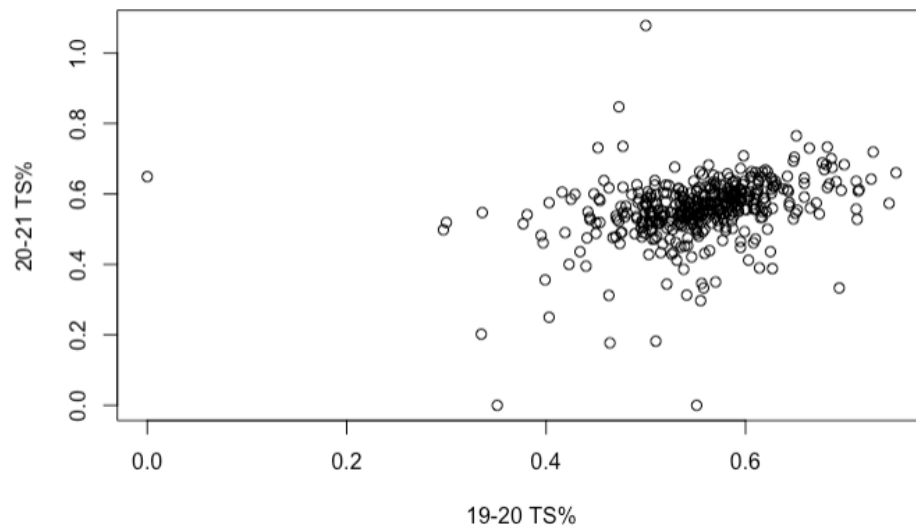
Fig. 4 displays the scatterplot for free throw percentage between the 2019-2020 and 2020-2021 regular season. The correlation for free throw percentage is 0.3159867. This scatterplot indicates a linear, moderately weak, positive relationship for free throw percentage between the two seasons. The scatterplot is most dense at around 0.85.



**Fig. 5: 2019-2020 eFG% vs 2020-2021 eFG%**

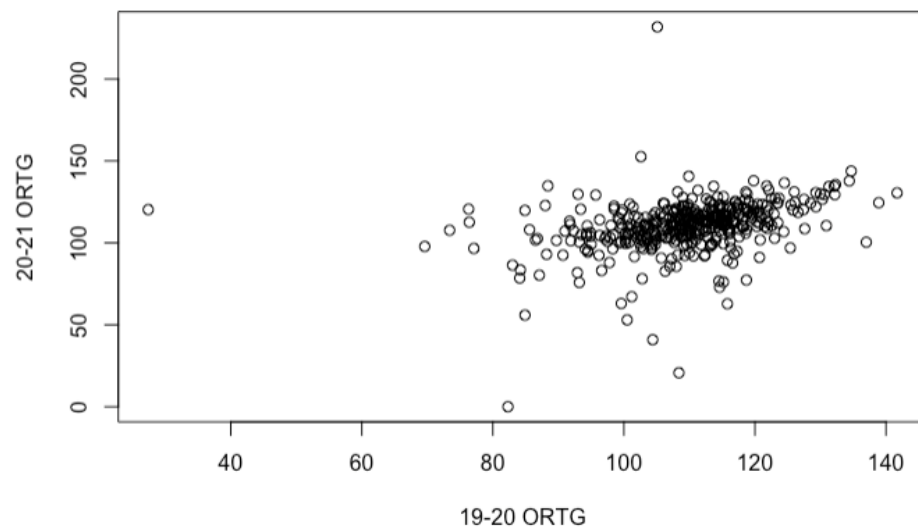
Fig. 5 displays the scatterplot for effective field goal percentage between the 2019-2020 and 2020-2021 regular season. The correlation for effective field goal percentage is 0.3354806. This scatterplot indicates a linear, moderately weak, positive relationship for effective field goal percentage between the two seasons. The scatterplot is most dense at around 0.5.





**Fig. 6: 2019-2020 TS% vs 2020-2021 TS%**

Fig. 6 displays the scatterplot for true shooting percentage between the 2019-2020 and 2020-2021 regular season. The correlation for true shooting percentage is 0.3175335. This scatterplot indicates a linear, moderately weak, positive relationship for true shooting percentage between the two seasons. The scatterplot is most dense at around 0.55.



**Fig. 7: 2019-2020 ORTG vs 2020-2021 ORTG**

Fig. 7 displays the scatterplot for offensive rating between the 2019-2020 and 2020-2021 regular season. The correlation for offensive rating is 0.3306984. This scatterplot indicates a linear, moderately weak, positive relationship for offensive rating between the two seasons. The scatterplot is most dense at around 110.

The team data between the 2019-2020 season and the 2020-2021 season are graphically and numerically displayed to discover any obvious trends from the environmental and travel changes implemented due to COVID-19.

## Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
W	30	35.3	11.914	15	23.5	44	56
L	30	35.3	9.603	17	28	42.75	50
ORtg	30	110.5	2.708	105.2	108.6	112.4	116.7
DRtg	30	110.68	3.084	102.9	109.125	112.925	115.5
Pace	30	100.31	2.217	95.8	98.725	101.475	105.1
TS_percentage	30	0.564	0.014	0.531	0.554	0.574	0.587
eFG_percentage	30	0.528	0.014	0.497	0.52	0.536	0.552
TOV_percentage	30	12.827	0.85	11.2	12.325	13.3	14.6
ORB_percentage	30	22.48	1.587	19.3	21.525	23.8	25.8
FT_FGA	30	0.201	0.018	0.17	0.186	0.212	0.234
Attend._G	30	17786.667	1582.969	15066	16511.5	19050.25	20629

**Fig. 8: 2019-2020 Team Statistics**

Fig.8. displays the summary statistics for teams in the 2019-2020 regular season. The mean, standard deviation, and quartiles are given for each metric that was kept in the dataset.

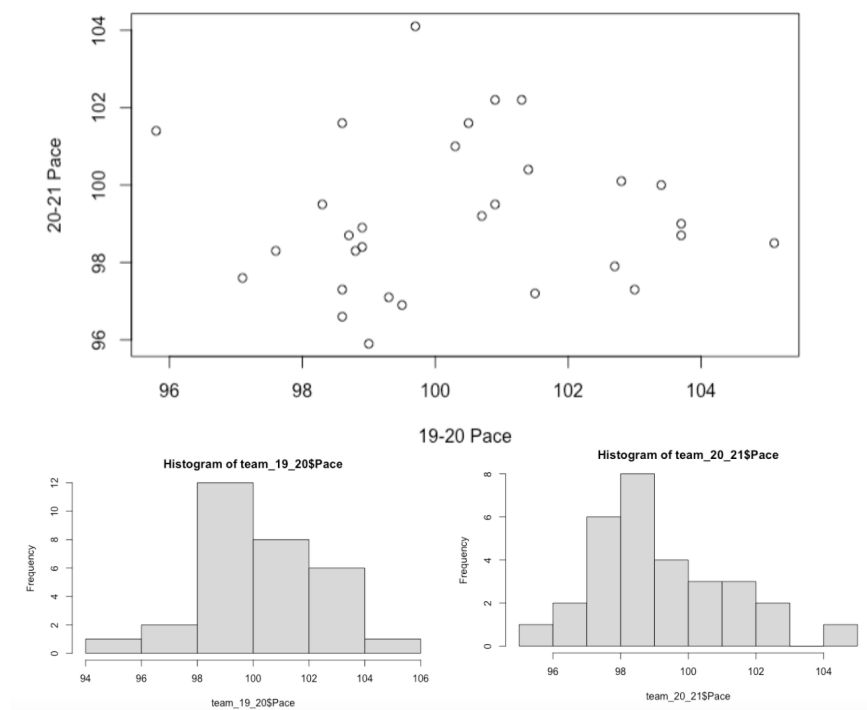
## Summary Statistics

Variable	N	Mean	Std. Dev.	Min	Pctl. 25	Pctl. 75	Max
W	30	36	10.014	17	31	42	52
L	30	36	10.014	20	30	41	55
ORtg	30	112.353	4	103.5	110.675	115.625	118.3
DRtg	30	112.33	2.377	107.1	111.225	113.8	117.2
Pace	30	99.18	1.955	95.9	97.675	100.325	104.1
TS_percentage	30	0.572	0.019	0.527	0.559	0.582	0.61
eFG_percentage	30	0.538	0.02	0.49	0.525	0.55	0.575
TOV_percentage	30	12.36	0.998	9.9	11.9	13.075	14.2
ORB_percentage	30	22.193	1.844	17.9	21.125	23.45	26.3
FT_FGA	30	0.192	0.016	0.156	0.181	0.198	0.226
Attend._G	30	1374.367	1161.882	0	406.5	1902.75	4203

**Fig. 9: 2020-2021 Team Statistics**

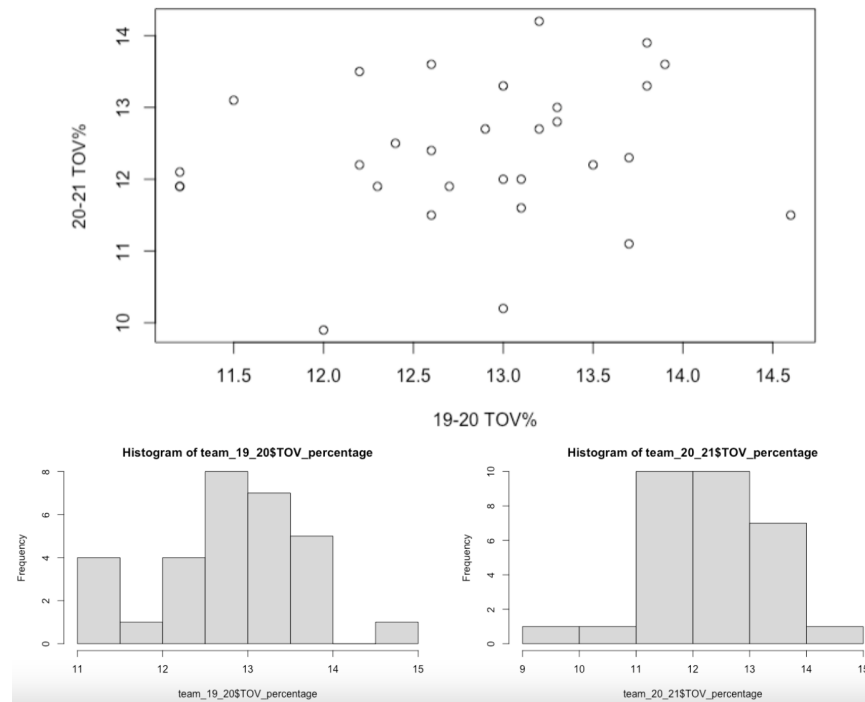
Fig. 9 displays the summary statistics for teams in the 2019-2020 regular season. The mean, standard deviation, and quartiles are given for each metric that was kept in the dataset. When analyzing the summary statistics, the mean values for over half of the metrics increased from the 2019-2020 season heading into the 2020-2021 season: W, L, ORtg, DRtg, TS\_percentage, and eFG\_percentage. The mean values that decreased from the 2019-2020 season heading into the 2020-2021 seasons are: Pace, TOV\_percentage, ORB\_percentage, FT\_FGA, and Attend\_G. The standard deviations for the 2020-2021 season were greater than the standard deviations from the 2019-2020 season in all metrics except for: W, DRTG, Pace, FT\_FGA, and Attend\_G. There was a decrease in minimum value in the 2020-2021 season compared to the 2019-2020 season in all metrics except for: W, L, DRTG, and Pace. There was an increase in maximum value in the 2020-2021 season compared to the 2019-2020 season in all metrics except for: W, Pace, TOV\_percentage, FT\_FGA, and Attend\_G.

Below are the scatterplots that displayed low correlations and displayed a somewhat weak relationship between the metric in both seasons. This means that these metrics were not repeatable unlike the other metrics. The histograms display how skewed the data is for each metric in their given season. This allows for analysis on how the trend has changed from one season to the next.



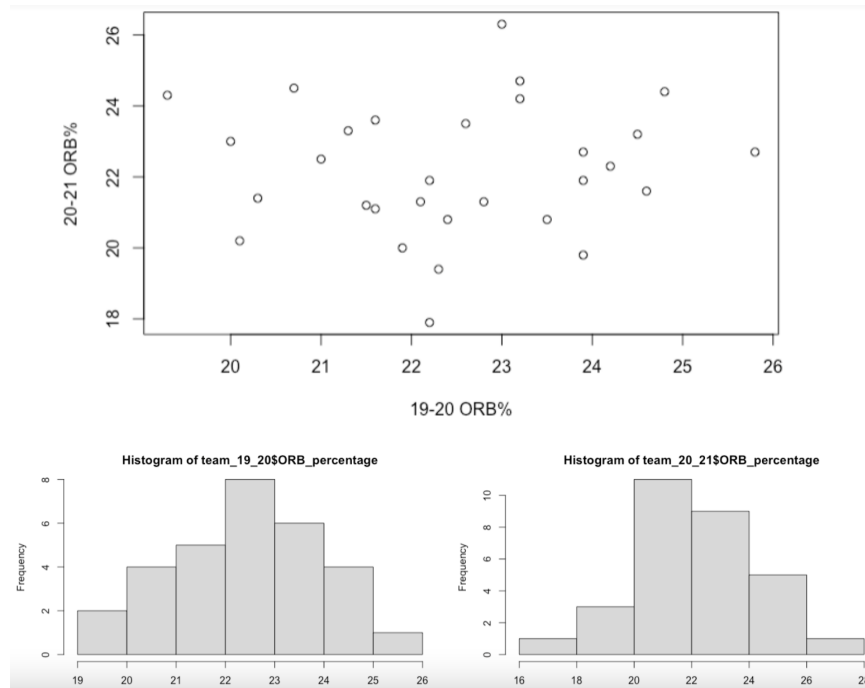
**Fig. 10: 2019-2020 Pace vs. 2020-2021 Pace**

Fig. 10 displays the scatterplot for pace between the 2019-2020 and 2020-2021 season. The correlation for pace is 0.03759867. This scatterplot shows no relationship at all between the two seasons. Also, the histograms of both seasons show that pace in the 2020-2021 season decreased from the pace in the 2019-2020 season.



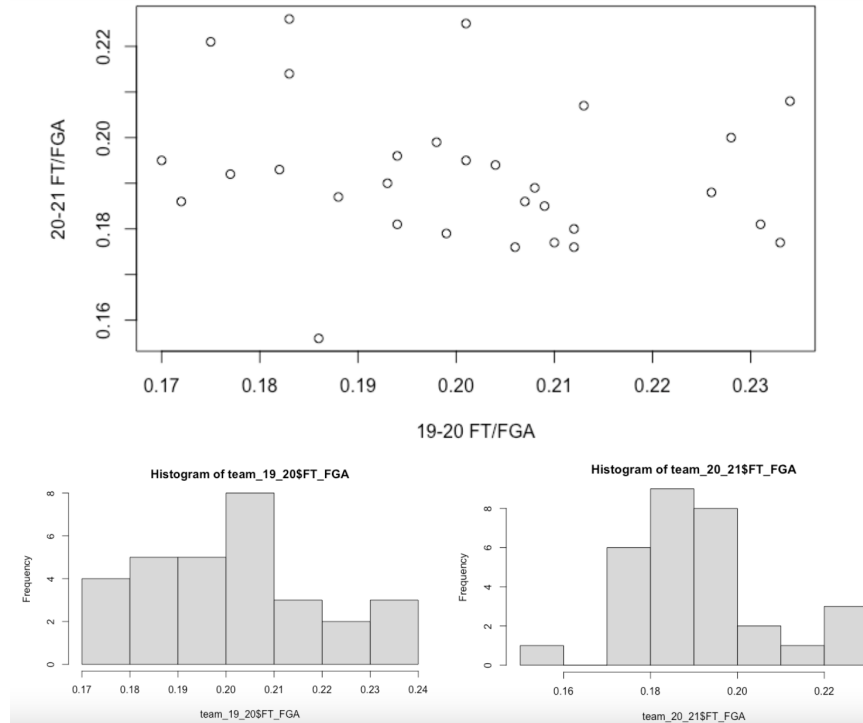
**Fig. 11: 2019-2020 TOV% vs 2020-2021 TOV%**

Fig. 11 displays the scatterplot for TOV% between the 2019-2020 and 2020-2021 season. The correlation for turnover percentage is 0.172061. This scatterplot shows no relationship at all between the two seasons. Also, the histograms of both seasons show that TOV% in the 2020-2021 season is less spread from the TOV% in the 2019-2020 season.



**Fig. 12: 2019-2020 ORB% vs 2020-2021 ORB%**

Fig. 12 displays the scatterplot for ORB% between the 2019-2020 and 2020-2021 season. The correlation for offensive rebound percentage is 0.05439447. This scatterplot shows no relationship at all between the two seasons. Also, the histograms of both seasons show that ORB% in the 2020-2021 season was more sporadic from the ORB% in the 2019-2020 season.



**Fig. 13: 2019-2020 FT/FGA vs. 2020-2021 FT/FGA**

Fig. 13 displays the scatterplot for FT/FGA between the 2019-2020 and 2020-2021 season. The correlation for FT/FGA is -0.1970078. This scatterplot shows no relationship at all between the two seasons. Also, the histograms of both seasons show that FT/FGA in the 2020-2021 season was more sporadic from the FT/FGA in the 2019-2020 season.



**Inferential Statistics:**

As we saw in the summary statistics for the players, GP, USG\_percentage, TO\_percentage, PPG, RPG, SPG, BPG, TOPG, VIVersatility all decreased from the 2019-2020 season heading into the 2020-2021 season. The mean value of GP lowered because the 2020-2021 season was shortened so that the NBA could return to its normal schedule that typically lasts from October to April. The regular season for the 2019-2020 season lasted 74 games at most for the teams that went into the bubble due to COVID-19. Also, the 2020-2021 regular season had 72 games. The USG\_percentage decreased which meant that players were not as involved while they were on the floor. This could be due to fatigue because players had a shortened off-season prior to the 2020-2021 season. Because players were not as rested and were not as prepared to play, their impact on the court lessened. TO\_percentage and TOPG also decreased in the 2020-2021 season. This could be linked to the decrease in USG\_percentage. Players typically need the ball to be involved in the team's plays. Since there was a decrease in USG\_percentage, players are less likely to handle the ball which means less opportunity for turnovers. PPG decreased in the 2020-2021 season which is also linked with USG\_percentage. Players need to be involved with the team's plays in order to score points. By being less involved on the court, players have less opportunities to score points. RPG, SPG, BPG all decreased in the 2020-2021 season. This could be due to a lack of hustle on the court. Since players are fatigued, they are less likely to hustle for the ball while playing and tend to settle on defense. VIVersatility decreased which is a metric that measures a player's ability to produce in points, assists, and rebounds. This could be due to fatigue, lessened minutes, lessened games due to contracting COVID-19, or a lack of motivation to perform due to external circumstances such as no fans or travel restrictions.

## COVID-19 and the NBA

The metrics for players that displayed a low correlation and a moderately-weak to weak relationship between the metric in both seasons were TO%, FT%, eFG%, TS%, and ORTG. TO% displayed a low correlation between the 2019-2020 season and the 2020-2021 season. As mentioned in the summary statistics, turnovers went down in the 2020-2021 season. This could be due to the decreased usage rate which decreases the likelihood of players creating turnovers. In addition, players could have been more focused when playing due to the lack of fans in the stadium. The improved concentration helped when handling the ball. FT%, eFG%, TS%, and ORTG all displayed a low correlation between the 2019-2020 season and the 2020-2021 season. All of these metrics displayed an increase in the summary statistics. A lack of fan presence in the stadium could have increased the focus of the players which enhanced their performance and allowed them to perform better on offense, allowing them to be more efficient. Since the NBA tried to limit travel as much as they can, players found themselves playing teams in the same area multiple times during the span of their home stay or away trip. By playing the same team multiple times, strategies can be created which allows players to perform better against those teams.

As we saw in the summary statistics for the teams, Pace, TOV\_percentage, ORB\_percentage, FT\_FGA, and Attend\_G all decreased from the 2019-2020 season heading into the 2020-2021 season. In both the summary statistics and the histogram, Pace decreased which means that players were playing at a slower rate, resulting in fewer possessions per game. This slower rate could be due to fatigue from the shortened off-season, or it could be due to the amount of injuries occurring to players throughout the season. Since there were many star players who suffered injuries, the role players or bench players had to step up to fill in. Since those players are not as good, they are less likely to be able to control the tempo of the game

## COVID-19 and the NBA

which can result in a slower game overall. In both the summary statistics and the histogram, TOV\_percentage decreased which means that players were committing less turnovers during the course of the 2020-2021 regular season. The histogram displays less spread which means that teams were consistent in the number of turnovers they were committing in the 2020-2021 season. For ORB\_percentage, the summary statistics show that the mean ORB\_percentage decreased, but for the histogram, it shows that ORB% is left-skewed which means at least more than half of the teams grabbed offensive rebounds higher than the mean. The mean ORB\_percentage decreased because there was less variability within the histogram. The histogram for ORB% in the 2019-2020 season was unimodal and symmetric which suggests that teams recorded the amount of offensive rebounds that they should have. However, the histogram for ORB% in the 2019-2020 season increased which suggests that there were more opportunities for players to get offensive rebounds since offensive rebounds came out of teammates' misses. For FT/FGA, the summary statistics show a decrease in its mean value, and the histogram shows that the values are more sporadic. The inconsistency of the histogram shows that players either got worse at free throws, or there were players taking free throws that were not as good. Since the 2020-2021 NBA season was riddled with injuries, less-skilled players took the floor more often which typically means that less-skilled free throw shooters stepped up to the line, decreasing the mean value. Attendance per game saw a decrease in its mean value and that is due to the lack of fans throughout the entire season. During the beginning of the season, most teams did not allow fans to attend games. As the season progressed and fans began to get the vaccine, the NBA removed some restrictions and allowed fans to attend the game. However, the fan attendance dropped significantly when compared to the 2019-2020 season even at the end of the 2020-2021 season.

## COVID-19 and the NBA

```
Call:
lm(formula = W ~ ORtg + DRtg + Pace + TS_percentage + eFG_percentage +
    TOV_percentage + ORB_percentage + FT_FGA, data = team_18_19)

Residuals:
    Min       1Q   Median       3Q      Max
-5.5754 -1.0158  0.3069  1.4070  3.6869

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   19.13145   46.59789   0.411   0.686
ORtg           1.04356    1.70013   0.614   0.546
DRtg          -2.33324    0.23347  -9.994 1.96e-09 ***
Pace          -0.05404    0.28480  -0.190   0.851
TS_percentage  637.11795   411.14269   1.550   0.136
eFG_percentage -354.79939   323.53745  -1.097   0.285
TOV_percentage -1.47121    2.50671  -0.587   0.564
ORB_percentage  1.16130    0.99213   1.171   0.255
FT_FGA        -46.43744    65.57498  -0.708   0.487
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.463 on 21 degrees of freedom
Multiple R-squared:  0.9696,    Adjusted R-squared:  0.9581
F-statistic: 83.83 on 8 and 21 DF,  p-value: 3.448e-14
```

**Fig. 14: 2018-2019 Season Multiple Linear Regression Model**

In this model, the response variable was wins and the predictor variables were all the rest of the columns in the 2018-2019 season team dataset except losses and attendance per game.

## COVID-19 and the NBA

```
Call:
lm(formula = W ~ ORtg + DRTg + Pace + TS_percentage + eFG_percentage +
    TOV_percentage + ORB_percentage + FT_FGA, data = team_19_20)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5052 -1.1461  0.1311  2.2186  4.5784

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   112.7521    64.6826   1.743  0.0959 .
ORtg          -1.4638     2.2215  -0.659  0.5171
DRTg          -2.3353     0.2627  -8.889 1.46e-08 ***
Pace          -0.4151     0.3043  -1.364  0.1870
TS_percentage  892.8376   546.7621   1.633  0.1174
eFG_percentage -178.4664   372.9298  -0.479  0.6372
TOV_percentage -5.0213     3.0685  -1.636  0.1166
ORB_percentage  2.0987     1.3691   1.533  0.1402
FT_FGA        -39.1972    92.6153  -0.423  0.6764
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.378 on 21 degrees of freedom
Multiple R-squared:  0.9418,    Adjusted R-squared:  0.9196
F-statistic: 42.46 on 8 and 21 DF,  p-value: 2.969e-11
```

**Fig. 15: 2019-2020 Season Multiple Linear Regression Model**

In this model, the response variable was wins and the predictor variables were all the rest of the columns in the 2019-2020 season team dataset except losses and attendance per game.

```
Call:
lm(formula = W ~ ORtg + DRtg + Pace + TS_percentage + eFG_percentage +
    TOV_percentage + ORB_percentage + FT_FGA, data = team_20_21)

Residuals:
    Min       1Q   Median       3Q      Max
-6.5039 -1.3910  0.0697  2.1753  6.5121

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   157.9488    76.8620   2.055  0.0525 .
ORtg          -2.2836     2.3227  -0.983  0.3367
DRtg          -1.7361     0.3083  -5.630 1.38e-05 ***
Pace          -0.4810     0.3938  -1.222  0.2354
TS_percentage  440.5402   635.6779   0.693  0.4959
eFG_percentage 294.3531   414.4499   0.710  0.4854
TOV_percentage -6.3518     3.3286  -1.908  0.0701 .
ORB_percentage  1.7503     1.2819   1.365  0.1866
FT_FGA         35.2544    83.5655   0.422  0.6774
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.398 on 21 degrees of freedom
Multiple R-squared:  0.9166,    Adjusted R-squared:  0.8849
F-statistic: 28.86 on 8 and 21 DF,  p-value: 1.195e-09
```

**Fig. 16: 2020-2021 Season Multiple Linear Regression Model**

In this model, the response variable was wins and the predictor variables were all the rest of the columns in the 2020-2021 season team dataset except losses and attendance per game.

After analyzing the descriptive statistics, multiple linear regression models of three different regular seasons in the NBA were utilized to further answer the research question. Wins were used as the response variable because it indicates how well a team played in that season alone. Seen above, the linear models of all three seasons have defensive rating as the most significant predictor for the number of team wins. The “\*\*\*” indicates that the p-value is smaller than the significant value which is between 0 and 0.001. The 2018-2019 season model and 2019-2020 season model outputted very similar results. However, the 2020-2021 season model had slight changes. Out of all the predictor coefficients, turnover percentage came out to be more significant than the previous models. This can be seen by the “.” in the significant codes which means that the coefficient resulted in a p-value smaller than the significant value which is

## COVID-19 and the NBA

between 0.05 and 1. With the significant decrease in fan attendance, turnovers could have decreased which made it more significant in predicting a team's performance.

Also there were many indicators that the 2020-2021 model was the worst model out of the three. When looking at the  $R^2$  values, the 2018-2019 model had a  $R^2$  value of 0.9696, the 2019-2020 model had a  $R^2$  value of 0.9418, and the 2020-2021 model had a  $R^2$  value of 0.9166. The  $R^2$  value was the lowest in the third model. This is another sign that there may exist more underlying factors outside of the coefficients that could have affected the teams' wins or performance in general. For example, there was a shortened offseason and travel restrictions that could have made the players in worse condition. The model was not able to see these factors. When looking at the residual standard errors in the models, the 2018-2019 model resulted in 2.463, the 2019-2020 model resulted in 3.378, and the 2020-2021 model resulted in 3.398. Although there isn't a significant difference, the third model had the worst model because it had the highest residual standard error. Again, this means that the model did not fit the dataset well compared to others, meaning there could be more hidden components in the 2020-2021 NBA season such as the decrease in fan attendance which could have deterred the team's performance. Lastly, the F-statistic for the 2018-2019 model was 83.83, the 2019-2020 model was 42.46, and the 2020-2021 model was 28.86. Knowing that a high F-statistic is a good indicator of whether there is a relationship between our predictor and response variables, all the models were significant. However, the third model had the smallest F-statistic which means that it was the worst model. Ultimately, the model summary statistics all point to the 2020-2021 season team model to be the worst fitted model out of the three. Using this as evidence, the potential underlying determinant is the result of various changes in the NBA such as travel restrictions, fan attendance, and shortened offseason caused by the COVID-19 pandemic.

**Discussion/Conclusion:**

Given that the 2020-2021 NBA season results were somewhat unique compared to the previous seasons due to COVID-19 restrictions, one practical implication that the coaches and front offices can assume is that a shortened offseason, shortened season, travel restrictions, and fan attendance can negatively affect the players' and team's performance. For example, the shortened offseason may have been a huge factor in the increase of injuries. Assuming this is true, coaches and front offices may find more favor in durable players.

These results are generalizable to other sports and leagues because COVID-19 resulted in restrictions on all sports in the United States. For example, the MLB had a shortened preseason which led to the increase in injuries. The insufficient preparation prior to the start of the season could have placed athletes at a greater risk of injury which ultimately can affect the team's performance. Another example is the NFL. Games had mostly empty stands which essentially got rid of home team advantage. Not only that, but the way the game was played was different with coaches coaching differently and officials officiating differently. According to ESPN, road teams had their highest win percentages in decades and saw a drop in turnovers (Seifert). Generally, all sports faced different environmental factors that altered their experience playing professional sports.

There were certain limitations that couldn't be accomplished due to time and resources. Overall, the challenge was separating COVID-19 caused differences from natural variability. However, for the datasets that were used in this analysis, the data was cleaned so that the players included in the datasets played both seasons. First, the player statistics datasets had to have the same players to make the scatterplots and correlations consistent. As a result, our descriptive statistics did not account for players injured in one season and returning in the next season. Also,



## COVID-19 and the NBA

it did not account for traded players who played on different teams in the same season because the data was cleaned so that the players would only be considered for the team they played more games for. Lastly, the analysis did not control for players progressing or regressing due to age.

Research that could be done in the future is to compare home and away statistics/records in each season. This would check if there was a significant league wide difference between how players and teams performed at home versus on the road. Given that every team plays an equal number of home and away games with relatively equal competition when averaged out over the season, if the only difference between the 2019-2020 and 2020-2021 regular seasons is COVID-19 (fan attendance, travel restrictions, shortened offseason, shortened season, etc.), then any difference in performance can be attributed to COVID-19. This can be possible because everything else has been controlled for other than progression/regression due to player aging.

Works Cited

Aschburner, Steve. "Coronavirus Pandemic Causes NBA to Suspend Season after Player Tests Positive." *NBA.com*, NBA.com, 10 Oct. 2020,  
<https://www.nba.com/news/coronavirus-pandemic-causes-nba-suspend-season>.

Haislop, Tadd. "NBA Bubble, Explained: A Complete Guide to the Rules, Teams, Schedule & More for Orlando Games." *Sporting News*, (Getty Images/SN Illustration), 26 Aug. 2020,  
<https://www.sportingnews.com/us/nba/news/nba-bubble-rules-teams-schedule-orlando/zha>  
[p66a9hcwq1khmcex3ggabo](https://www.sportingnews.com/us/nba/news/nba-bubble-rules-teams-schedule-orlando/zha).

McHill, Andrew W., and Evan D. Chinoy. "Utilizing the National Basketball Association's Covid-19 Restart 'Bubble' to Uncover the Impact of Travel and Circadian Disruption on Athletic Performance." *Nature News*, Nature Publishing Group, 11 Dec. 2020,  
<https://www.nature.com/articles/s41598-020-78901-2>.

"NBA 2020-21 Season Explainer: Schedule, COVID-19 Protocols, Playoff Format and Big Questions." *ESPN*, ESPN Internet Ventures, 6 Dec. 2020,  
[https://www.espn.com/nba/story/\\_/id/30430912/nba-2020-21-season-explainer-schedule-co](https://www.espn.com/nba/story/_/id/30430912/nba-2020-21-season-explainer-schedule-covid-19-protocols-playoff-format-big-questions)  
[vid-19-protocols-playoff-format-big-questions](https://www.espn.com/nba/story/_/id/30430912/nba-2020-21-season-explainer-schedule-covid-19-protocols-playoff-format-big-questions).

Platt, Brooks N., et al. "Injury Rates in Major League Baseball during the 2020 Covid-19 Season - Brooks N. Platt, Timothy L. Uhl, Aaron D. Sciascia, Anthony J. Zacharias, Nicole G. Lemaster, Austin v. Stone, 2021." *SAGE Journals*,  
<https://journals.sagepub.com/doi/full/10.1177/2325967121999646>.

## COVID-19 and the NBA

Seifert, Kevin. "10 Ways Covid-19 Protocols Impact NFL Game Play: No Home-Field

Advantage, Fewer Turnovers and More." *ESPN*, ESPN Internet Ventures, 15 Dec. 2020,

[https://www.espn.com/nfl/insider/story/\\_/id/30512756/10-ways-covid-19-protocols-imp](https://www.espn.com/nfl/insider/story/_/id/30512756/10-ways-covid-19-protocols-impact-nfl-game-play-no-home-field-advantage-fewer-turnovers-more)

[ct-nfl-game-play-no-home-field-advantage-fewer-turnovers-more](https://www.espn.com/nfl/insider/story/_/id/30512756/10-ways-covid-19-protocols-impact-nfl-game-play-no-home-field-advantage-fewer-turnovers-more).