NBA All Star Case Study

Nikhil Thomas A

7/28/2021

The age old question

In this phase the main question which came to mind was the age old argument that NBA players nowadays are teaming up and creating super teams which in turn is taking the easy way out according to some purists. So I used the number of All Stars in an NBA team as a metric and tried to identify the trends from the past 30 years and test this hypothesis.

Also in this phase I decided that I would be using Basketball Refrence's All Star Game database for this case study.

basketball-reference.com/allstar/NBA_1991.html

All-Star Games | Box Scores ▾ | Voting ▾ | Leaders ▾ | Contests | All-Star Career Stats | Rising Stars ▾

West Share & Export ▾ Glossary

		Totals																Shooting		
Starters	**Tm**	**MP**	**FG**	**FGA**	**3P**	**3PA**	**FT**	**FTA**	**ORB**	**DRB**	**TRB**	**AST**	**STL**	**BLK**	**TOV**	**PF**	**PTS**	**FG%**	**3P%**	**FT%**
Karl Malone	UTA	31:00	6	11	0	0	4	6	4	7	11	4	1	1	3	1	16	.545		.667
Magic Johnson	LAL	28:00	7	16	2	5	0	0	1	3	4	3	0	0	3	1	16	.438	.400	
Chris Mullin	GSW	24:00	4	8	1	1	4	4	0	2	2	2	2	0	2	2	13	.500	1.000	1.000
Kevin Johnson	PHO	23:00	2	5	0	0	1	2	1	1	2	7	3	1	3	2	5	.400		.500
David Robinson	SAS	18:00	6	13	0	0	4	5	3	3	6	0	2	3	2	5	16	.462		.800
		Totals																**Shooting**		
Reserves	**Tm**	**MP**	**FG**	**FGA**	**3P**	**3PA**	**FT**	**FTA**	**ORB**	**DRB**	**TRB**	**AST**	**STL**	**BLK**	**TOV**	**PF**	**PTS**	**FG%**	**3P%**	**FT%**
James Worthy	LAL	21:00	3	11	0	0	3	4	0	2	2	0	2	1	0	2	9	.273		.750
Clyde Drexler	POR	19:00	4	9	0	0	4	4	2	2	4	2	1	1	0	3	12	.444		1.000
Kevin Duckworth	POR	19:00	2	3	0	0	2	2	2	2	4	0	1	0	2	3	6	.667		1.000
Tom Chambers	PHO	18:00	4	11	0	1	0	0	2	2	4	1	1	0	4	3	8	.364	.000	
Terry Porter	POR	15:00	2	6	0	2	0	0	1	2	3	4	2	1	3	2	4	.333	.000	
Tim Hardaway	GSW	12:00	2	7	1	2	0	0	2	1	3	4	2	0	0	1	5	.286	.500	
John Stockton	UTA	12:00	1	6	0	1	2	4	0	1	1	2	0	0	0	2	4	.167	.000	.500
Team Totals		**240**	**43**	**106**	**4**	**12**	**24**	**31**	**18**	**28**	**46**	**29**	**17**	**8**	**22**	**27**	**114**	**.406**	**.333**	**.774**

Larry Bird (East): Did not play due to injury; replaced by Hersey Hawkins.
Hersey Hawkins (East): Replacement for Larry Bird.
Isiah Thomas (East): Did not play due to injury; no replacement named.

Preparing and Cleaning

After extracting the data from the website the data needed to be cleaned and organised for analysis. An important point to be noted is that the injury replacements for the All Star game are also considered in this dataset.

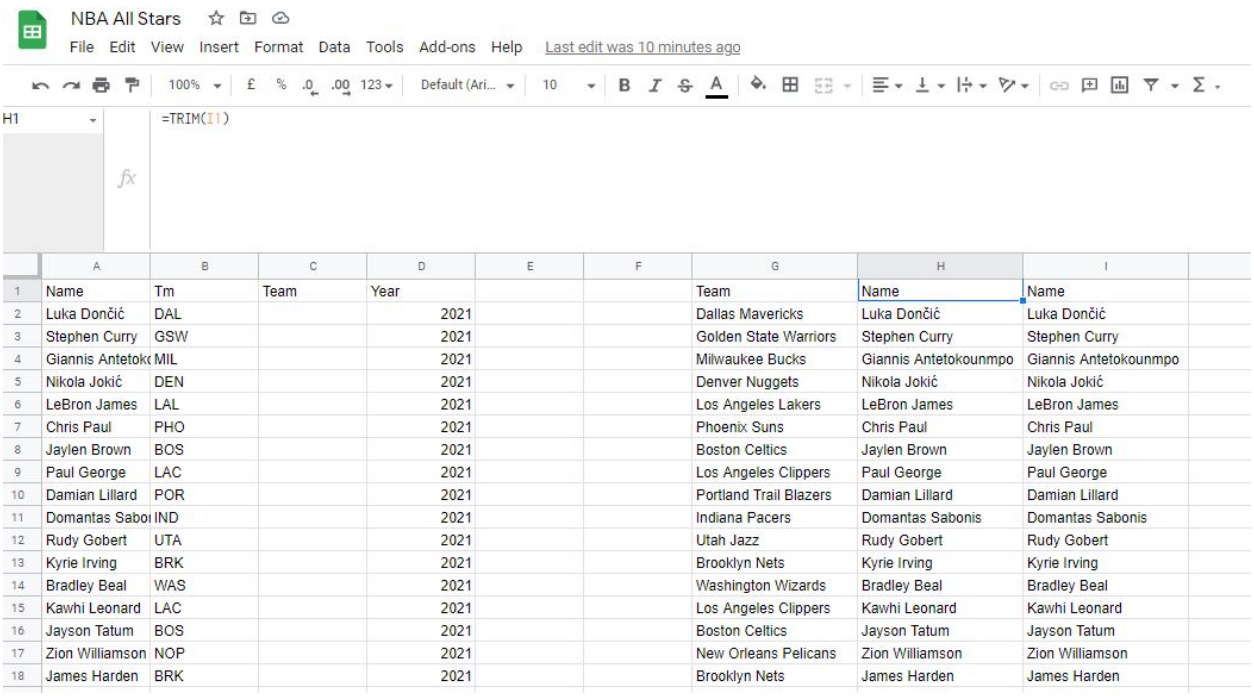
NBA All Stars

File Edit View Insert Format Data Tools Add-ons Help

	Name	Tm	Team	Year
5	Bernard King	WSB		1991
6	Joe Dumars	DET		1991
7	Dominique Wil	ATL		1991
8	Ricky Pierce	MIL		1991
9	Hersey Hawkir	PHI		1991
10	Kevin McHale	BOS		1991
11	Brad Daughert	CLE		1991
12	Alvin Robertso	MIL		1991
13	Robert Parish	BOS		1991
14	Karl Malone	UTA		1991
15	Magic Johnson	LAL		1991
16	Chris Mullin	GSW		1991
17	Kevin Johnson	PHO		1991
18	David Robinso	SAS		1991
19	James Worthy	LAL		1991
20	Clyde Drexler	POR		1991
21	Kevin Duckwo	POR		1991
22	Tom Chambers	PHO		1991
23	Terry Porter	POR		1991
24	Tim Hardaway	GSW		1991
25	John Stockton	UTA		1991
26	Larry Bird (Eas	BOS		1991
27	Isiah Thomas (I	DET		1991

Processing data tables

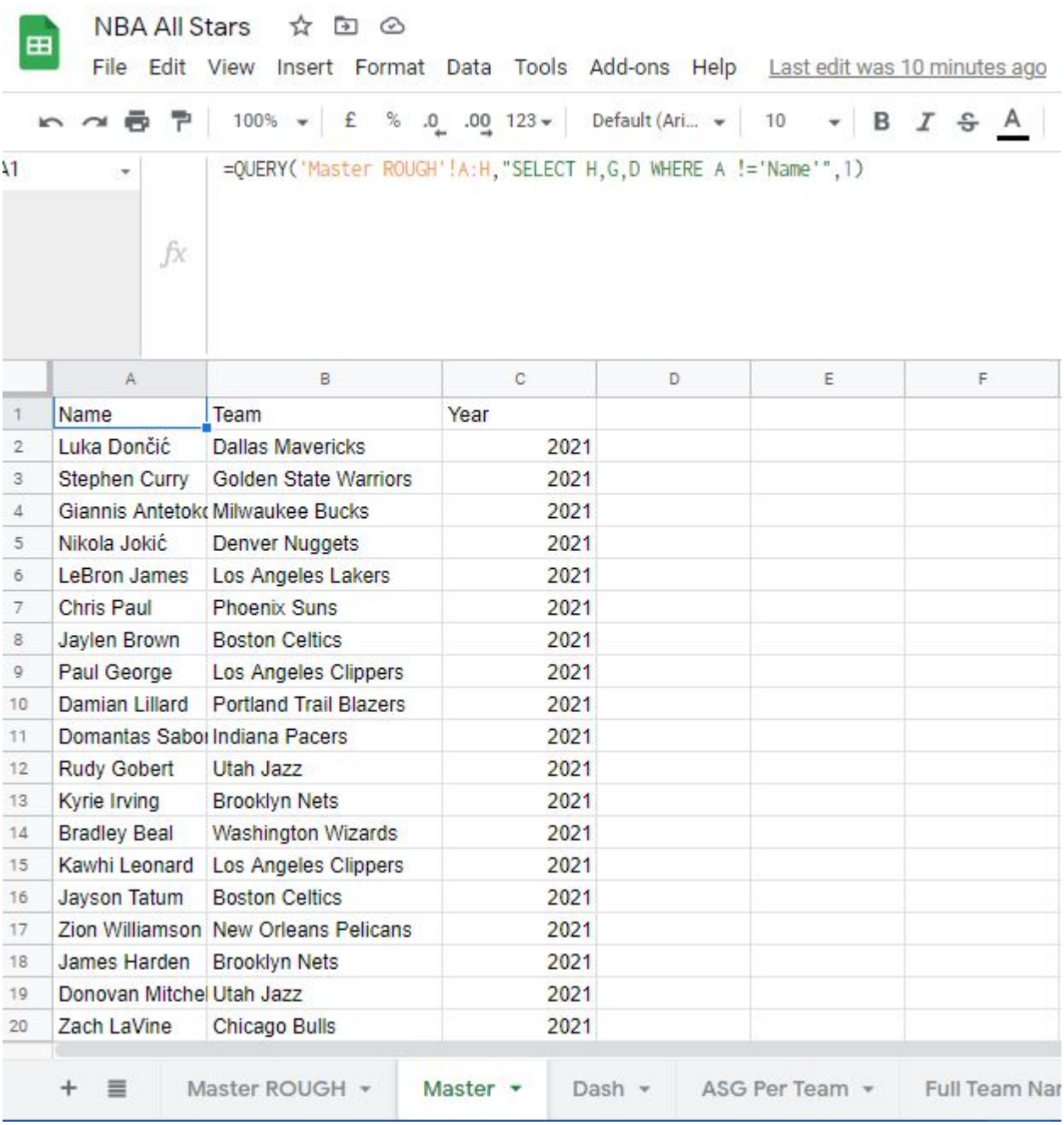
After pulling the data into a spreadsheet the data was formatted accordingly and the player names were cleaned using trim() and split() functions. Later all the year tables was combined using the Google sheets query function.



	A	B	C	D	E	F	G	H	I
1	Name	Tm	Team	Year			Team	Name	Name
2	Luka Dončić	DAL		2021			Dallas Mavericks	Luka Dončić	Luka Dončić
3	Stephen Curry	GSW		2021			Golden State Warriors	Stephen Curry	Stephen Curry
4	Giannis Antetok	MIL		2021			Milwaukee Bucks	Giannis Antetokounmpo	Giannis Antetokounmpo
5	Nikola Jokić	DEN		2021			Denver Nuggets	Nikola Jokić	Nikola Jokić
6	LeBron James	LAL		2021			Los Angeles Lakers	LeBron James	LeBron James
7	Chris Paul	PHO		2021			Phoenix Suns	Chris Paul	Chris Paul
8	Jaylen Brown	BOS		2021			Boston Celtics	Jaylen Brown	Jaylen Brown
9	Paul George	LAC		2021			Los Angeles Clippers	Paul George	Paul George
10	Damian Lillard	POR		2021			Portland Trail Blazers	Damian Lillard	Damian Lillard
11	Domantas Sabo	IND		2021			Indiana Pacers	Domantas Sabonis	Domantas Sabonis
12	Rudy Gobert	UTA		2021			Utah Jazz	Rudy Gobert	Rudy Gobert
13	Kyrie Irving	BRK		2021			Brooklyn Nets	Kyrie Irving	Kyrie Irving
14	Bradley Beal	WAS		2021			Washington Wizards	Bradley Beal	Bradley Beal
15	Kawhi Leonard	LAC		2021			Los Angeles Clippers	Kawhi Leonard	Kawhi Leonard
16	Jayson Tatum	BOS		2021			Boston Celtics	Jayson Tatum	Jayson Tatum
17	Zion Williamson	NOP		2021			New Orleans Pelicans	Zion Williamson	Zion Williamson
18	James Harden	BRK		2021			Brooklyn Nets	James Harden	James Harden

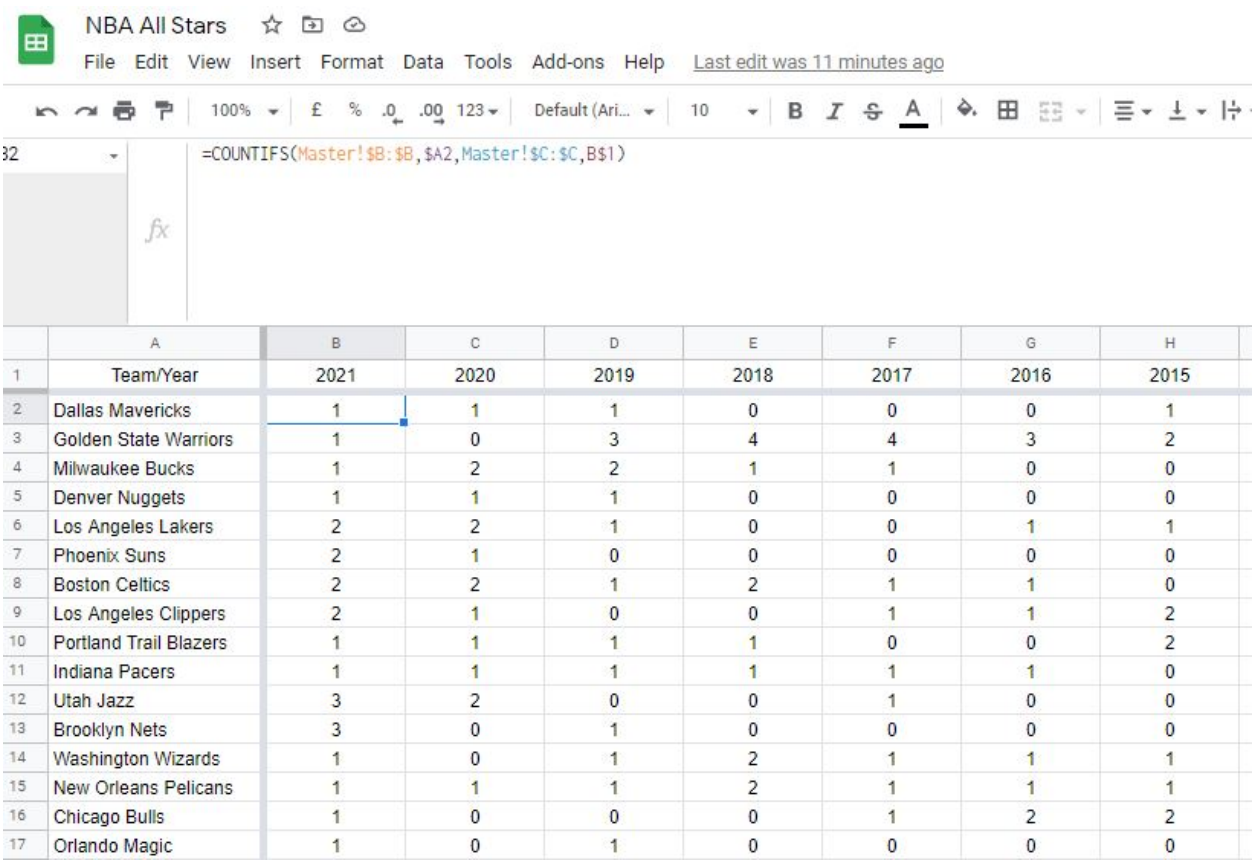
Analyzing the data

After combining the tables of All Stars from the past 30 years, the full team name was bought into the dataset with a basic vlookup(). Then later the final table for analysis was prepared having the name, team and year of all All Stars from the past 30 years.



	A	B	C	D	E	F
1	Name	Team	Year			
2	Luka Dončić	Dallas Mavericks	2021			
3	Stephen Curry	Golden State Warriors	2021			
4	Giannis Antetoko	Milwaukee Bucks	2021			
5	Nikola Jokić	Denver Nuggets	2021			
6	LeBron James	Los Angeles Lakers	2021			
7	Chris Paul	Phoenix Suns	2021			
8	Jaylen Brown	Boston Celtics	2021			
9	Paul George	Los Angeles Clippers	2021			
10	Damian Lillard	Portland Trail Blazers	2021			
11	Domantas Sabo	Indiana Pacers	2021			
12	Rudy Gobert	Utah Jazz	2021			
13	Kyrie Irving	Brooklyn Nets	2021			
14	Bradley Beal	Washington Wizards	2021			
15	Kawhi Leonard	Los Angeles Clippers	2021			
16	Jayson Tatum	Boston Celtics	2021			
17	Zion Williamson	New Orleans Pelicans	2021			
18	James Harden	Brooklyn Nets	2021			
19	Donovan Mitchel	Utah Jazz	2021			
20	Zach LaVine	Chicago Bulls	2021			

A simplified table was created as well using countifs() statement which will help us in further analysis of the data when we move forward in the case study.



	A	B	C	D	E	F	G	H
1	Team/Year	2021	2020	2019	2018	2017	2016	2015
2	Dallas Mavericks	1	1	1	0	0	0	1
3	Golden State Warriors	1	0	3	4	4	3	2
4	Milwaukee Bucks	1	2	2	1	1	0	0
5	Denver Nuggets	1	1	1	0	0	0	0
6	Los Angeles Lakers	2	2	1	0	0	1	1
7	Phoenix Suns	2	1	0	0	0	0	0
8	Boston Celtics	2	2	1	2	1	1	0
9	Los Angeles Clippers	2	1	0	0	1	1	2
10	Portland Trail Blazers	1	1	1	1	0	0	2
11	Indiana Pacers	1	1	1	1	1	1	0
12	Utah Jazz	3	2	0	0	1	0	0
13	Brooklyn Nets	3	0	1	0	0	0	0
14	Washington Wizards	1	0	1	2	1	1	1
15	New Orleans Pelicans	1	1	1	2	1	1	1
16	Chicago Bulls	1	0	0	0	1	2	2
17	Orlando Magic	1	0	1	0	0	0	0

Visualising and Further Analysis

Now we import the data into R for visualization and in depth analysis which will lead to insight derivation. First we import the data.

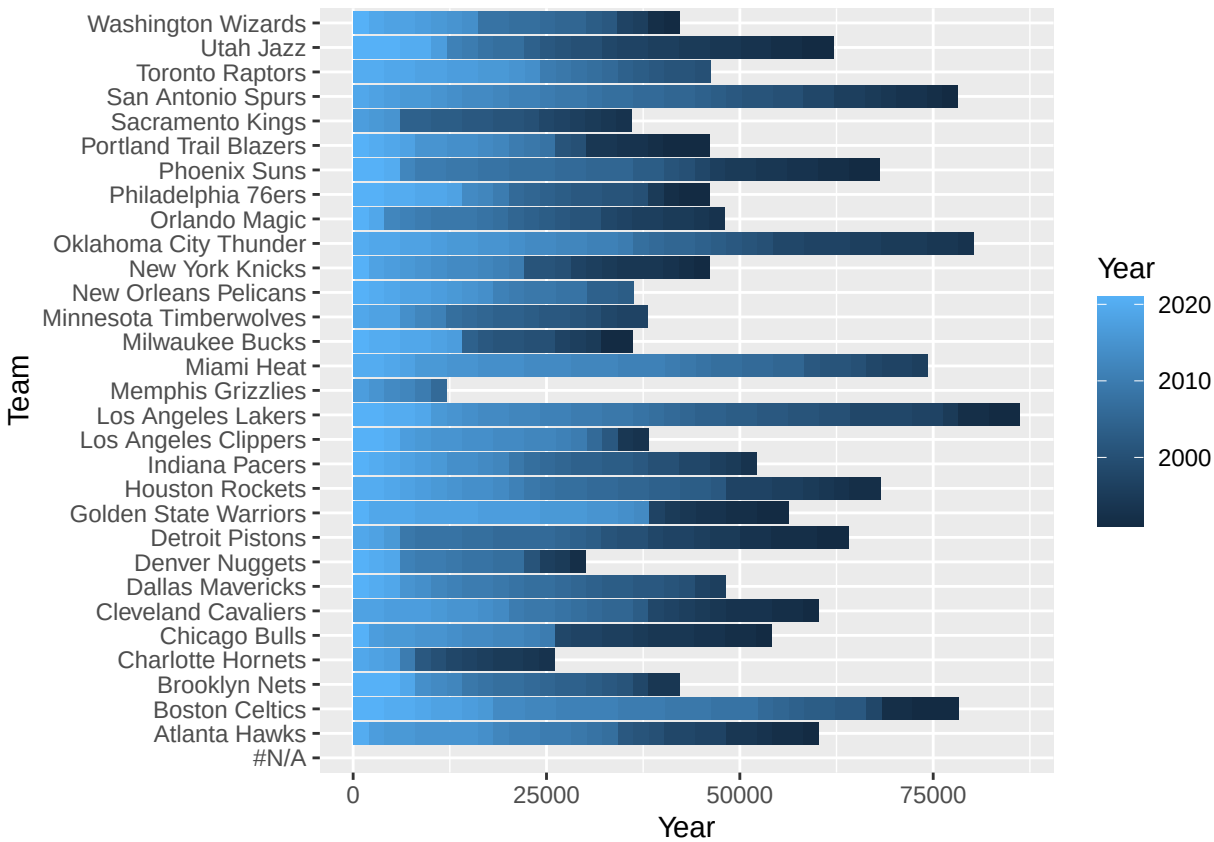
```
maindata <- read.csv("NBA All Stars - Master.csv")
head(maindata)
```

```
##                    Name                  Team Year
## 1           Luka Dončić      Dallas Mavericks 2021
## 2         Stephen Curry Golden State Warriors 2021
## 3 Giannis Antetokounmpo       Milwaukee Bucks 2021
## 4          Nikola Jokić        Denver Nuggets 2021
## 5          LeBron James    Los Angeles Lakers 2021
## 6            Chris Paul          Phoenix Suns 2021
```

Then we plot the data using ggplot

```
library(ggplot2)
ggplot((maindata),aes(Year,Team,fill=Year))+geom_col()
```

```
## Warning: Removed 112 rows containing missing values (position_stack).
```



This visualization is good but is not conclusive hence we must dive deeper. Now we import the second table where we will continue our analysis

```
descdata <- read.csv("NBA Project CSV - Sheet1.csv")
head(descdata)
```

```
##                   Count X2021 X2020 X2019 X2018 X2017 X2016 X2015 X2014 X2013
## 1      Dallas Mavericks     1     1     1     0     0     0     1     1     0
## 2 Golden State Warriors     1     0     3     4     4     3     2     1     1
## 3       Milwaukee Bucks     1     2     2     1     1     0     0     0     0
## 4        Denver Nuggets     1     1     1     0     0     0     0     0     0
## 5    Los Angeles Lakers     2     2     1     0     0     1     1     1     2
## 6          Phoenix Suns     2     1     0     0     0     0     0     0     0
##   X2012 X2011 X2010 X2009 X2008 X2007 X2006 X2005 X2004 X2003 X2002 X2001 X2000
## 1     1     1     2     1     1     2     1     1     1     2     2     1     1
## 2     0     0     0     0     0     0     0     0     0     0     0     0     0
## 3     0     0     0     0     0     0     0     0     1     0     1     2     2
## 4     0     1     2     1     2     2     0     0     0     0     0     1     0
## 5     2     2     2     2     1     1     1     1     2     2     2     2     2
## 6     1     0     2     2     2     3     2     3     0     2     0     1     1
##   X1998 X1997 X1996 X1995 X1994 X1993 X1992 X1991
## 1     0     1     1     0     0     0     0     0
## 2     0     1     0     1     1     2     2     2
## 3     0     1     1     1     0     0     0     2
## 4     0     0     1     1     0     0     1     0
## 5     4     2     0     1     0     0     2     2
## 6     1     0     1     2     2     2     2     2
```

Now we find out the average All Star from the past 30 years for each team and calculate which team has averaged the highest number of All Stars

```
descdata$average <- rowMeans(descdata[,c(2:30)],na.rm = TRUE)
head(descdata)
```

```
##                   Count X2021 X2020 X2019 X2018 X2017 X2016 X2015 X2014 X2013
## 1      Dallas Mavericks     1     1     1     0     0     0     1     1     0
## 2 Golden State Warriors     1     0     3     4     4     3     2     1     1
## 3       Milwaukee Bucks     1     2     2     1     1     0     0     0     0
## 4        Denver Nuggets     1     1     1     0     0     0     0     0     0
## 5    Los Angeles Lakers     2     2     1     0     0     1     1     1     2
## 6          Phoenix Suns     2     1     0     0     0     0     0     0     0
##   X2012 X2011 X2010 X2009 X2008 X2007 X2006 X2005 X2004 X2003 X2002 X2001 X2000
## 1     1     1     2     1     1     2     1     1     1     2     2     1     1
## 2     0     0     0     0     0     0     0     0     0     0     0     0     0
## 3     0     0     0     0     0     0     0     0     1     0     1     2     2
## 4     0     1     2     1     2     2     0     0     0     0     0     1     0
## 5     2     2     2     2     1     1     1     1     2     2     2     2     2
## 6     1     0     2     2     2     3     2     3     0     2     0     1     1
##   X1998 X1997 X1996 X1995 X1994 X1993 X1992 X1991   average
## 1     0     1     1     0     0     0     0     0 0.8275862
## 2     0     1     0     1     1     2     2     2 0.8965517
## 3     0     1     1     1     0     0     0     2 0.5517241
## 4     0     0     1     1     0     0     1     0 0.5172414
## 5     4     2     0     1     0     0     2     2 1.4137931
## 6     1     0     1     2     2     2     2     2 1.1034483
```

```
AvgTeam<- select(descdata,Count,average)
TopAvgTeam<- AvgTeam[order(-AvgTeam$average),]
head(TopAvgTeam)
```

```
##                     Count  average
## 5     Los Angeles Lakers 1.413793
## 20 Oklahoma City Thunder 1.379310
## 25     San Antonio Spurs 1.310345
## 23            Miami Heat 1.275862
## 7         Boston Celtics 1.241379
## 19       Houston Rockets 1.172414
```

Here we can see in the top 5 teams with the highest average All Stars 3 if not 4 of the teams are considered big market teams. Big market teams continue to lure in big names in the free agency period and this trend is justified here.

Now we further analyse by figuring out teams with 2 All Stars as well as 3 All Stars in each year which will help us gauge the trends better

```
overTwo<-colSums(descdata[,c(1:32)]>=2,na.rm = TRUE)
overTwo
```

```
##   Count   X2021   X2020   X2019   X2018   X2017   X2016   X2015   X2014   X2013
##      30       7       8       5       8       3       7       8       5       8
##   X2012   X2011   X2010   X2009   X2008   X2007   X2006   X2005   X2004   X2003
##       6       6       6       7       6      10       5       7       7       7
##   X2002   X2001   X2000   X1998   X1997   X1996   X1995   X1994   X1993   X1992
##       6       9       6       5       9       6       6       8       8      10
##   X1991 average
```

```
##       9       0
```

```
overThree<-colSums(descdata[,c(1:32)]>=3,na.rm = TRUE)
overThree
```

```
##   Count   X2021   X2020   X2019   X2018   X2017   X2016   X2015   X2014   X2013
##      30       2       0       1       1       2       1       1       1       1
##   X2012   X2011   X2010   X2009   X2008   X2007   X2006   X2005   X2004   X2003
##       1       2       1       2       2       1       1       1       0       0
##   X2002   X2001   X2000   X1998   X1997   X1996   X1995   X1994   X1993   X1992
##       0       0       0       1       2       0       1       2       1       1
##   X1991 average
##       2       0
```

Now we see the mean of two all stars and three all stars per team over the years and add the initial vector to the dataset

```
mean(overTwo[c(2:31)],na.rm = TRUE)
```

```
## [1] 6.933333
```

```
mean(overThree[c(2:31)],na.rm = TRUE)
```

```
## [1] 1.033333
```

```
descdata[nrow(descdata)+1,]=overTwo
descdata[nrow(descdata)+1,]=overThree
head(descdata)
```

```
##                   Count X2021 X2020 X2019 X2018 X2017 X2016 X2015 X2014 X2013
## 1      Dallas Mavericks     1     1     1     0     0     0     1     1     0
## 2 Golden State Warriors     1     0     3     4     4     3     2     1     1
## 3       Milwaukee Bucks     1     2     2     1     1     0     0     0     0
## 4        Denver Nuggets     1     1     1     0     0     0     0     0     0
## 5    Los Angeles Lakers     2     2     1     0     0     1     1     1     2
## 6          Phoenix Suns     2     1     0     0     0     0     0     0     0
##   X2012 X2011 X2010 X2009 X2008 X2007 X2006 X2005 X2004 X2003 X2002 X2001 X2000
## 1     1     1     2     1     1     2     1     1     1     2     2     1     1
## 2     0     0     0     0     0     0     0     0     0     0     0     0     0
## 3     0     0     0     0     0     0     0     0     1     0     1     2     2
## 4     0     1     2     1     2     2     0     0     0     0     0     1     0
## 5     2     2     2     2     1     1     1     1     2     2     2     2     2
## 6     1     0     2     2     2     3     2     3     0     2     0     1     1
##   X1998 X1997 X1996 X1995 X1994 X1993 X1992 X1991   average
## 1     0     1     1     0     0     0     0     0 0.8275862
## 2     0     1     0     1     1     2     2     2 0.8965517
## 3     0     1     1     1     0     0     0     2 0.5517241
## 4     0     0     1     1     0     0     1     0 0.5172414
## 5     4     2     0     1     0     0     2     2 1.4137931
## 6     1     0     1     2     2     2     2     2 1.1034483
```

As we can see there is an average of one team with 3 all stars in the past 30 years and nearly 7 teams average close to 7 all stars in the same time period. Then we calculate the standard deviation of both and plot the same.

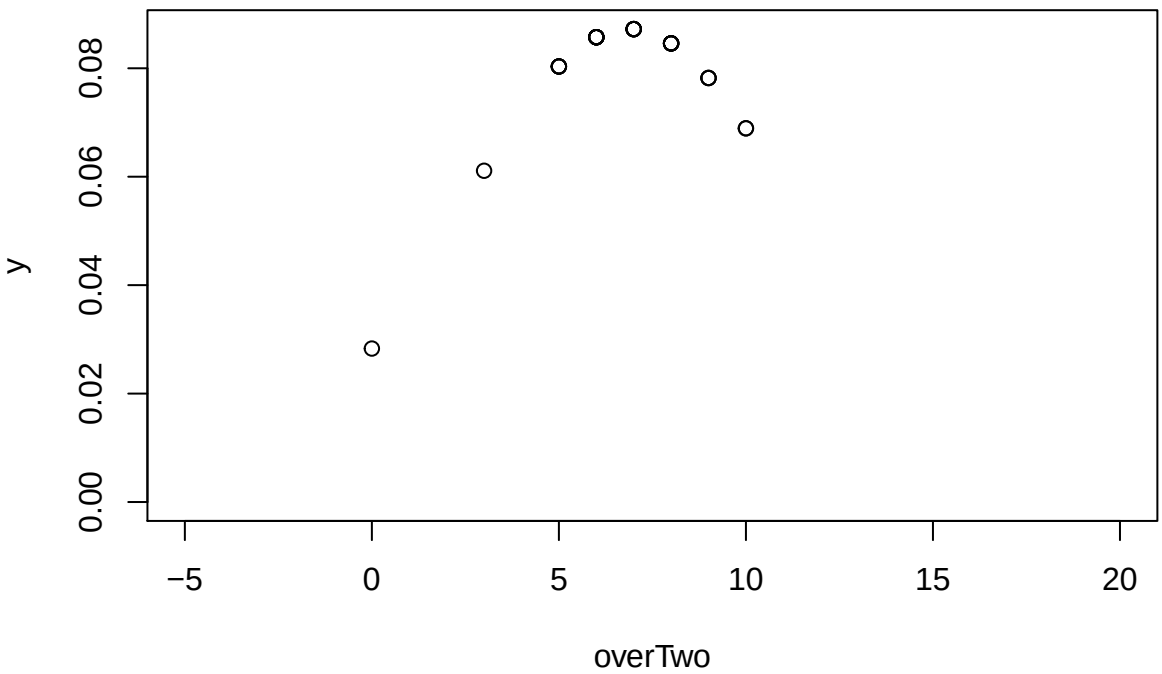
```
sd(overTwo)
```

```
## [1] 4.571564
```

```r
sd(overThree)
```

```
## [1] 5.176618
```

```r
y<-dnorm(overTwo,mean = 6.86, sd = 4.571564)
plot(overTwo,y,
     xlim=c(-5,20))
```



Conclusion

From the overTwo plot we can see that the over two All Stars per team per year follows a normal distribution curve. This in turn shows that all star caliber players teaming up is now recent anomaly.