Dataset – NBA Player Information and Statistics for the 2019-2020 NBA Season

For this project, I used two datasets containing statistical and other information about every NBA player who competed in the 2019-2020 NBA season. Both NBA.com and Basketball-Reference.com keep great databases, each of which had something different that I wanted to use.

These datasets both include player name, some basic stats (points scored, games played, etc.), some general information (age, team, and when they were drafted in the case of NBA.com's data), and some advanced metrics which I will describe later on. The figures below show the data from the sources, and then saved into excel.

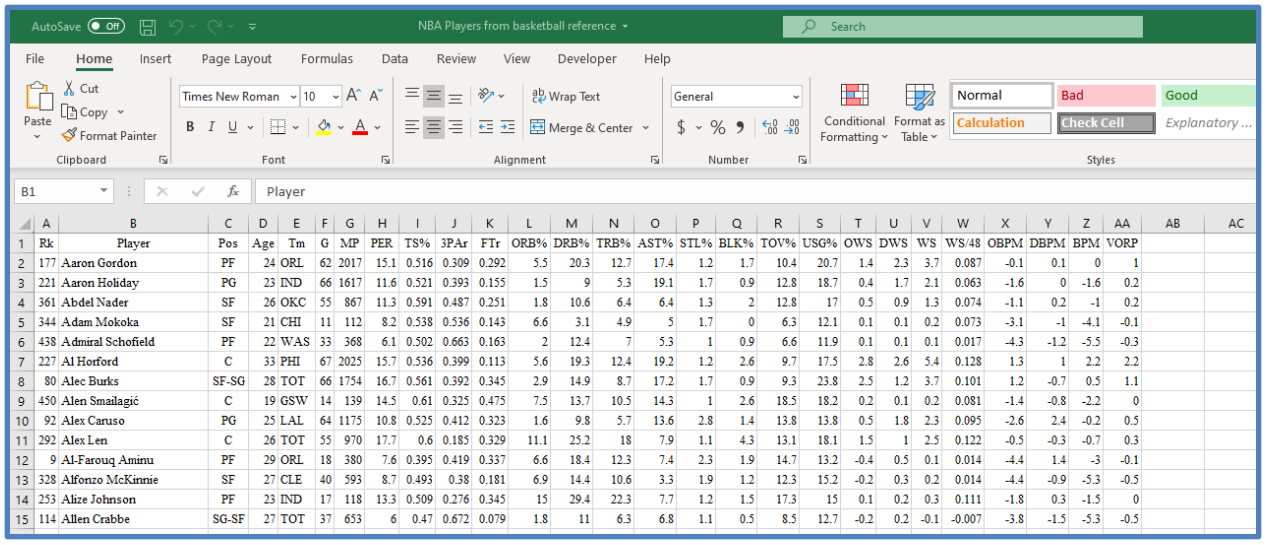
Advanced Share & more ▾ ☑ When table is sorted, hide non-qualifiers for rate stats Glossary Hide Partial Rows

Rk	Player	Pos	Age	Tm	G	MP	PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP
1	Steven Adams	C	26	OKC	63	1680	20.5	.604	.006	.421	14.0	24.0	19.2	13.2	1.5	3.4	14.2	17.3	3.8	2.7	6.5	.185	1.9	1.1	2.9	2.1
2	Bam Adebayo	PF	22	MIA	72	2417	20.3	.598	.018	.484	8.5	24.9	17.0	24.2	1.7	3.8	17.6	21.2	4.6	3.9	8.5	.168	1.4	2.0	3.4	3.3
3	LaMarcus Aldridge	C	34	SAS	53	1754	19.7	.571	.198	.241	6.3	17.8	12.0	11.4	1.0	4.4	7.8	23.4	3.0	1.4	4.5	.122	1.8	-0.5	1.4	1.5
4	Kyle Alexander	C	23	MIA	2	13	4.7	.500	.000	.000	17.9	8.3	12.9	0.0	0.0	0.0	33.3	10.2	0.0	0.0	0.0	-0.003	-6.1	-3.5	-9.6	0.0
5	Nickeil Alexander-Walker	SG	21	NOP	47	591	8.9	.473	.500	.139	1.6	13.5	7.5	21.1	1.3	1.1	16.1	23.3	-0.7	0.4	-0.2	-0.020	-3.2	-1.4	-4.6	-0.4
6	Grayson Allen	SG	24	MEM	38	718	12.0	.609	.562	.179	1.2	11.1	6.2	10.0	0.7	0.2	10.9	17.6	0.8	0.4	1.2	.082	-0.1	-1.2	-1.3	0.1
7	Jarrett Allen	C	21	BRK	70	1852	20.7	.664	.013	.581	12.3	24.9	18.7	8.9	1.0	4.2	11.7	14.9	5.2	2.9	8.2	.212	1.4	1.0	2.3	2.0

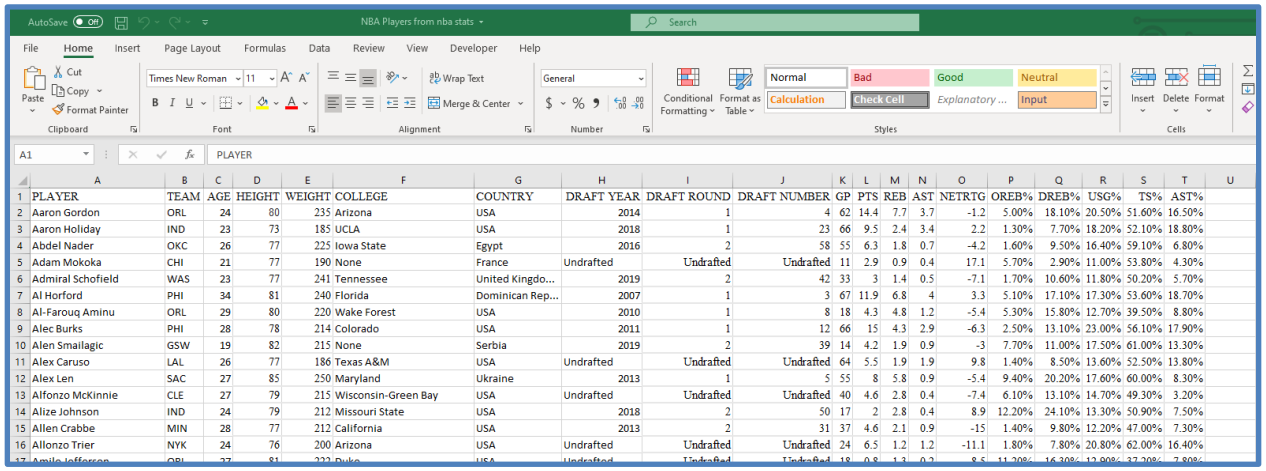
1 - Basketabll-Reference.com

PLAYER	TEAM	AGE	HEIGHT	WEIGHT	COLLEGE	COUNTRY	DRAFT YEAR	DRAFT ROUND	DRAFT NUMBER	GP	PTS	REB	AST	NETRTG	O
Aaron Gordon	ORL	24	6-8	235	Arizona	USA	2014	1	4	62	14.4	7.7	3.7	-1.2	
Aaron Holiday	IND	23	6-0	185	UCLA	USA	2018	1	23	66	9.5	2.4	3.4	2.2	
Abdel Nader	OKC	26	6-5	225	Iowa State	Egypt	2016	2	58	55	6.3	1.8	0.7	-4.2	
Adam Mokoka	CHI	21	6-5	190	None	France	Undrafted	Undrafted	Undrafted	11	2.9	0.9	0.4	17.1	

2 - NBA.com

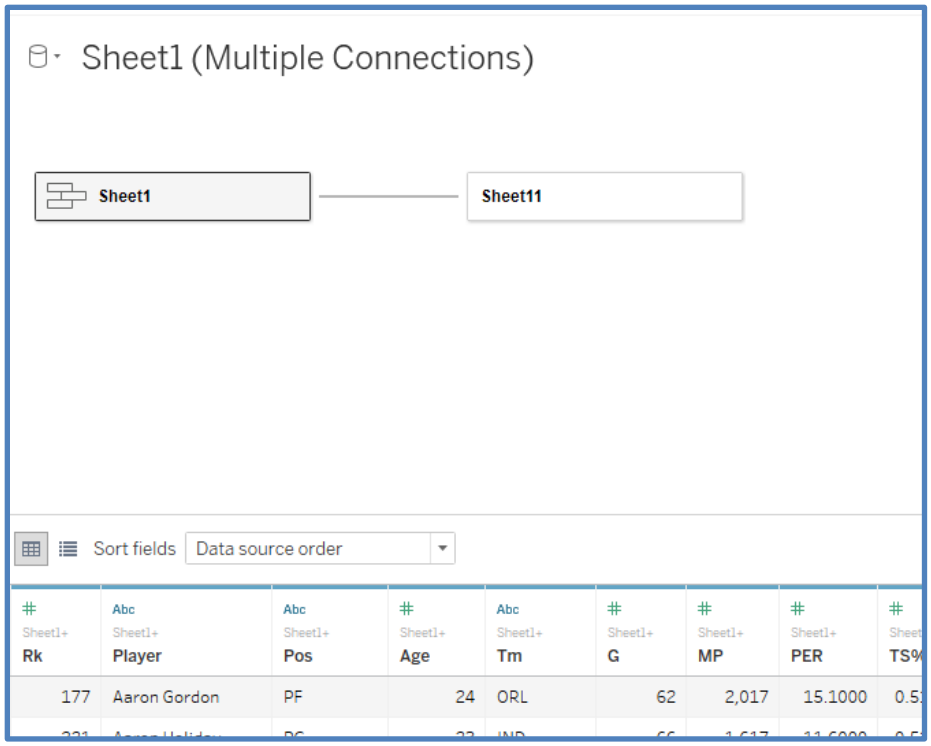


Rk	Player	Pos	Age	Tm	G	MP	PER	TS%	3PAr	FTr	ORB%	DRB%	TRB%	AST%	STL%	BLK%	TOV%	USG%	OWS	DWS	WS	WS/48	OBPM	DBPM	BPM	VORP
177	Aaron Gordon	PF	24	ORL	62	2017	15.1	0.516	0.309	0.292	5.5	20.3	12.7	17.4	1.2	1.7	10.4	20.7	1.4	2.3	3.7	0.087	-0.1	0.1	0	1
221	Aaron Holiday	PG	23	IND	66	1617	11.6	0.521	0.393	0.155	1.5	9	5.3	19.1	1.7	0.9	12.8	18.7	0.4	1.7	2.1	0.063	-1.6	0	-1.6	0.2
361	Abdel Nader	SF	26	OKC	55	867	11.3	0.591	0.487	0.251	1.8	10.6	6.4	6.4	1.3	2	12.8	17	0.5	0.9	1.3	0.074	-1.1	0.2	-1	0.2
344	Adam Mokoka	SF	21	CHI	11	112	8.2	0.538	0.536	0.143	6.6	3.1	4.9	5	1.7	0	6.3	12.1	0.1	0.1	0.2	0.073	-3.1	-1	-4.1	-0.1
438	Admiral Schofield	PF	22	WAS	33	368	6.1	0.502	0.663	0.163	2	12.4	7	5.3	1	0.9	6.6	11.9	0.1	0.1	0.1	0.017	-4.3	-1.2	-5.5	-0.3
227	Al Horford	C	33	PHI	67	2025	15.7	0.536	0.399	0.113	5.6	19.3	12.4	19.2	1.2	2.6	9.7	17.5	2.8	2.6	5.4	0.128	1.3	1	2.2	2.2
80	Alec Burks	SF-SG	28	TOT	66	1754	16.7	0.561	0.392	0.345	2.9	14.9	8.7	17.2	1.7	0.9	9.3	23.8	2.5	1.2	3.7	0.101	1.2	-0.7	0.5	1.1
450	Alen Smailagić	C	19	GSW	14	139	14.5	0.61	0.325	0.475	7.5	13.7	10.5	14.3	1	2.6	18.5	18.2	0.2	0.1	0.2	0.081	-1.4	-0.8	-2.2	0
92	Alex Caruso	PG	25	LAL	64	1175	10.8	0.525	0.412	0.323	1.6	9.8	5.7	13.6	2.8	1.4	13.8	13.8	0.5	1.8	2.3	0.095	-2.6	2.4	-0.2	0.5
292	Alex Len	C	26	TOT	55	970	17.7	0.6	0.185	0.329	11.1	25.2	18	7.9	1.1	4.3	13.1	18.1	1.5	1	2.5	0.122	-0.5	-0.3	-0.7	0.3
9	Al-Farouq Aminu	PF	29	ORL	18	380	7.6	0.395	0.419	0.337	6.6	18.4	12.3	7.4	2.3	1.9	14.7	13.2	-0.4	0.5	0.1	0.014	-4.4	1.4	-3	-0.1
328	Alfonzo McKinnie	SF	27	CLE	40	593	8.7	0.493	0.38	0.181	6.9	14.4	10.6	3.3	1.9	1.2	12.3	15.2	-0.2	0.3	0.2	0.014	-4.4	-0.9	-5.3	-0.5
253	Alize Johnson	PF	23	IND	17	118	13.3	0.509	0.276	0.345	15	29.4	22.3	7.7	1.2	1.5	17.3	15	0.1	0.2	0.3	0.111	-1.8	0.3	-1.5	0
114	Allen Crabbe	SG-SF	27	TOT	37	653	6	0.47	0.672	0.079	1.8	11	6.3	6.8	1.1	0.5	8.5	12.7	-0.2	0.2	-0.1	-0.007	-3.8	-1.5	-5.3	-0.5



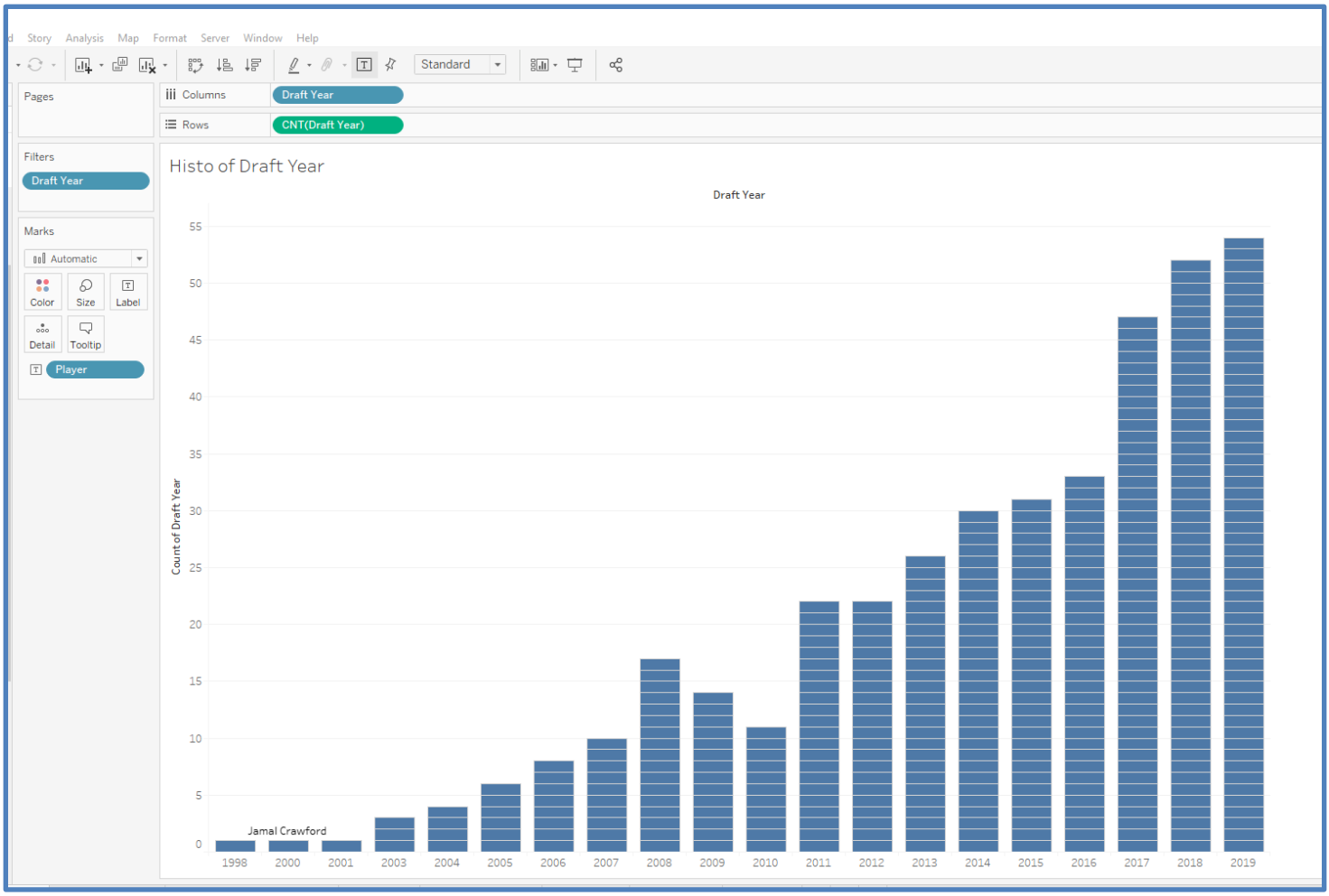
PLAYER	TEAM	AGE	HEIGHT	WEIGHT	COLLEGE	COUNTRY	DRAFT YEAR	DRAFT ROUND	DRAFT NUMBER	GP	PTS	REB	AST	NETRTG	OREB%	DREB%	USG%	TS%	AST%
Aaron Gordon	ORL	24	80	235	Arizona	USA	2014	1	4	62	14.4	7.7	3.7	-1.2	5.00%	18.10%	20.50%	51.60%	16.50%
Aaron Holiday	IND	23	73	185	UCLA	USA	2018	1	23	66	9.5	2.4	3.4	2.2	1.30%	7.70%	18.20%	52.10%	18.80%
Abdel Nader	OKC	26	77	225	Iowa State	Egypt	2016	2	58	55	6.3	1.8	0.7	-4.2	1.60%	9.50%	16.40%	59.10%	6.80%
Adam Mokoka	CHI	21	77	190	None	France	Undrafted	Undrafted	Undrafted	11	2.9	0.9	0.4	17.1	5.70%	2.90%	11.00%	53.80%	4.30%
Admiral Schofield	WAS	23	77	241	Tennessee	United Kingdo...	2019	2	42	33	3	1.4	0.5	-7.1	1.70%	10.60%	11.80%	50.20%	5.70%
Al Horford	PHI	34	81	240	Florida	Dominican Rep...	2007	1	3	67	11.9	6.8	4	3.3	5.10%	17.10%	17.30%	53.60%	18.70%
Al-Farouq Aminu	ORL	29	80	220	Wake Forest	USA	2010	1	8	18	4.3	4.8	1.2	-5.4	5.30%	15.80%	12.70%	39.50%	8.80%
Alec Burks	PHI	28	78	214	Colorado	USA	2011	1	12	66	15	4.3	2.9	-6.3	2.50%	13.10%	23.00%	56.10%	17.90%
Alen Smailagic	GSW	19	82	215	None	Serbia	2019	2	39	14	4.2	1.9	0.9	-3	7.70%	11.00%	17.50%	61.00%	13.30%
Alex Caruso	LAL	26	77	186	Texas A&M	USA	Undrafted	Undrafted	Undrafted	64	5.5	1.9	1.9	9.8	1.40%	8.50%	13.60%	52.50%	13.80%
Alex Len	SAC	27	85	250	Maryland	Ukraine	2013	1	5	55	8	5.8	0.9	-5.4	9.40%	20.20%	17.60%	60.00%	8.30%
Alfonzo McKinnie	CLE	27	79	215	Wisconsin-Green Bay	USA	Undrafted	Undrafted	Undrafted	40	4.6	2.8	0.4	-7.4	6.10%	13.10%	14.70%	49.30%	3.20%
Alize Johnson	IND	24	79	212	Missouri State	USA	2018	2	50	17	2	2.8	0.4	8.9	12.20%	24.10%	13.30%	50.90%	7.50%
Allen Crabbe	MIN	28	77	212	California	USA	2013	2	31	37	4.6	2.1	0.9	-15	1.40%	9.80%	12.20%	47.00%	7.30%
Allonzo Trier	NYK	24	76	200	Arizona	USA	Undrafted	Undrafted	Undrafted	24	6.5	1.2	1.2	-11.1	1.80%	7.80%	20.80%	62.00%	16.40%

Then I imported the data to Tableau. Since I was using two different datasets, I had to define a relationship in Tableau to connect them. I did this by setting the 'Player' columns – which contained the names of each player – in each dataset to be recognized as the same.

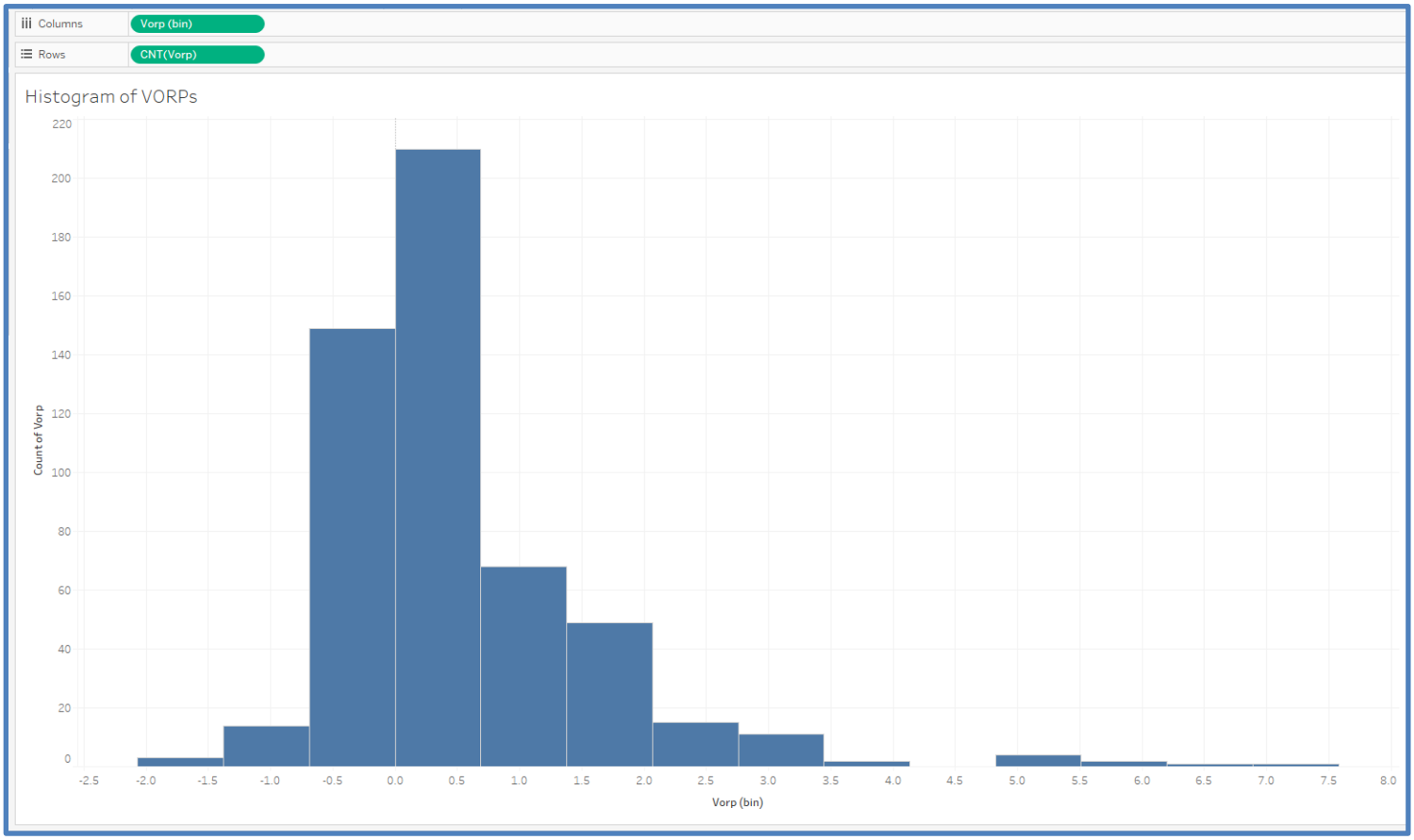


Then, I started the data visualization process. Below, I analyzed player data by a histogram of draft year. As expected, there are more players in the league drafted recently compared to longer ago. NBA players usually retire in their mid-thirties. It was surprising to see that there was still a player from the 1998 draft (Vince Carter), and that the 2008 draft has more players left than either of 2009 or 2010. This indicates that the 2008 draft had more quality and longevity.

To create this graph, I used the draft year in the columns and the count of draft year in the rows and labeled by player, which allowed me to hover over a section in the histogram and see which player it contained.

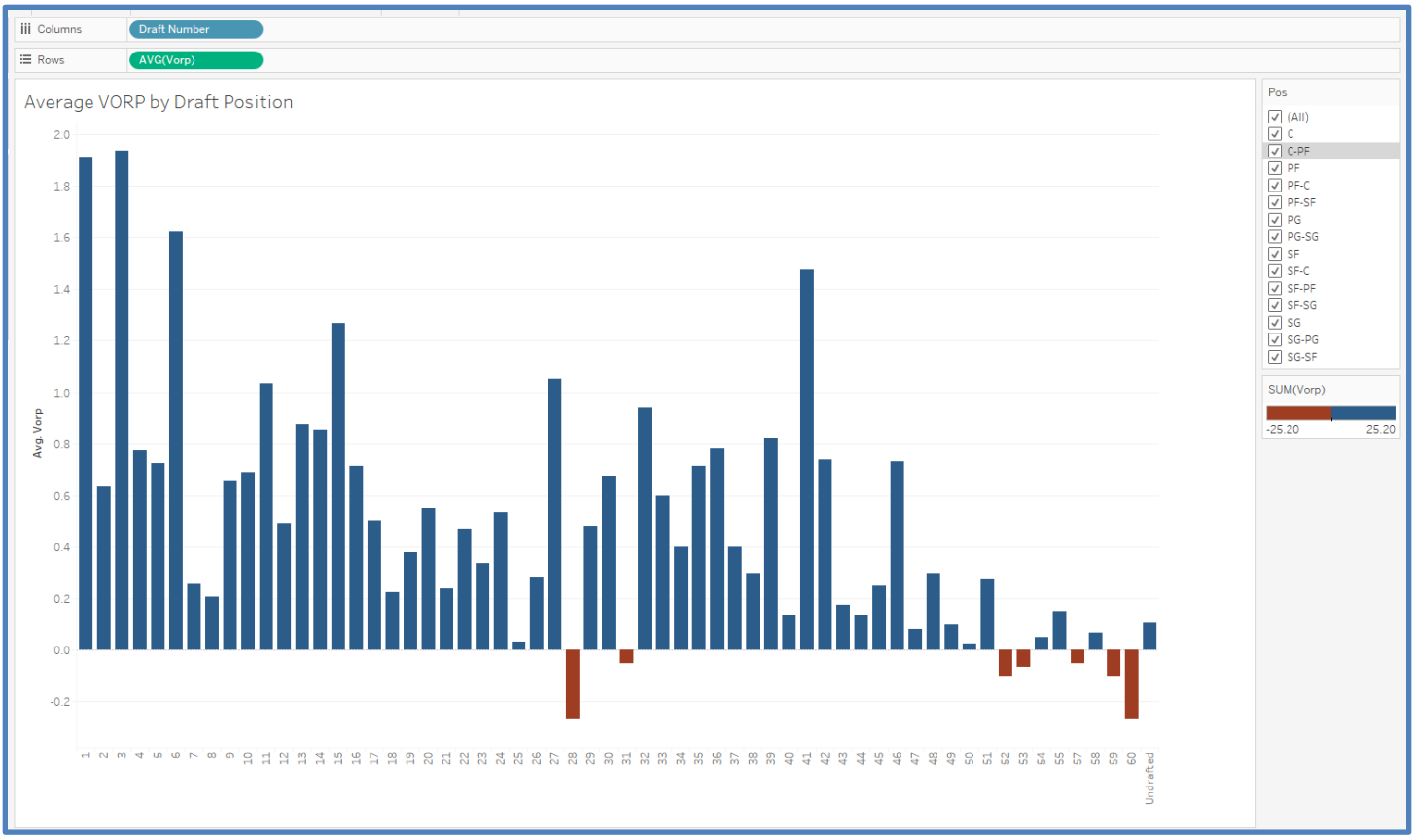


Next, I created a histogram of VORP statistics. VORP stands for value over replacement player and is a popular advanced metric around the NBA for comparing the values of players to other players. I wanted to see the distribution of VORPs so, I created the histogram below.



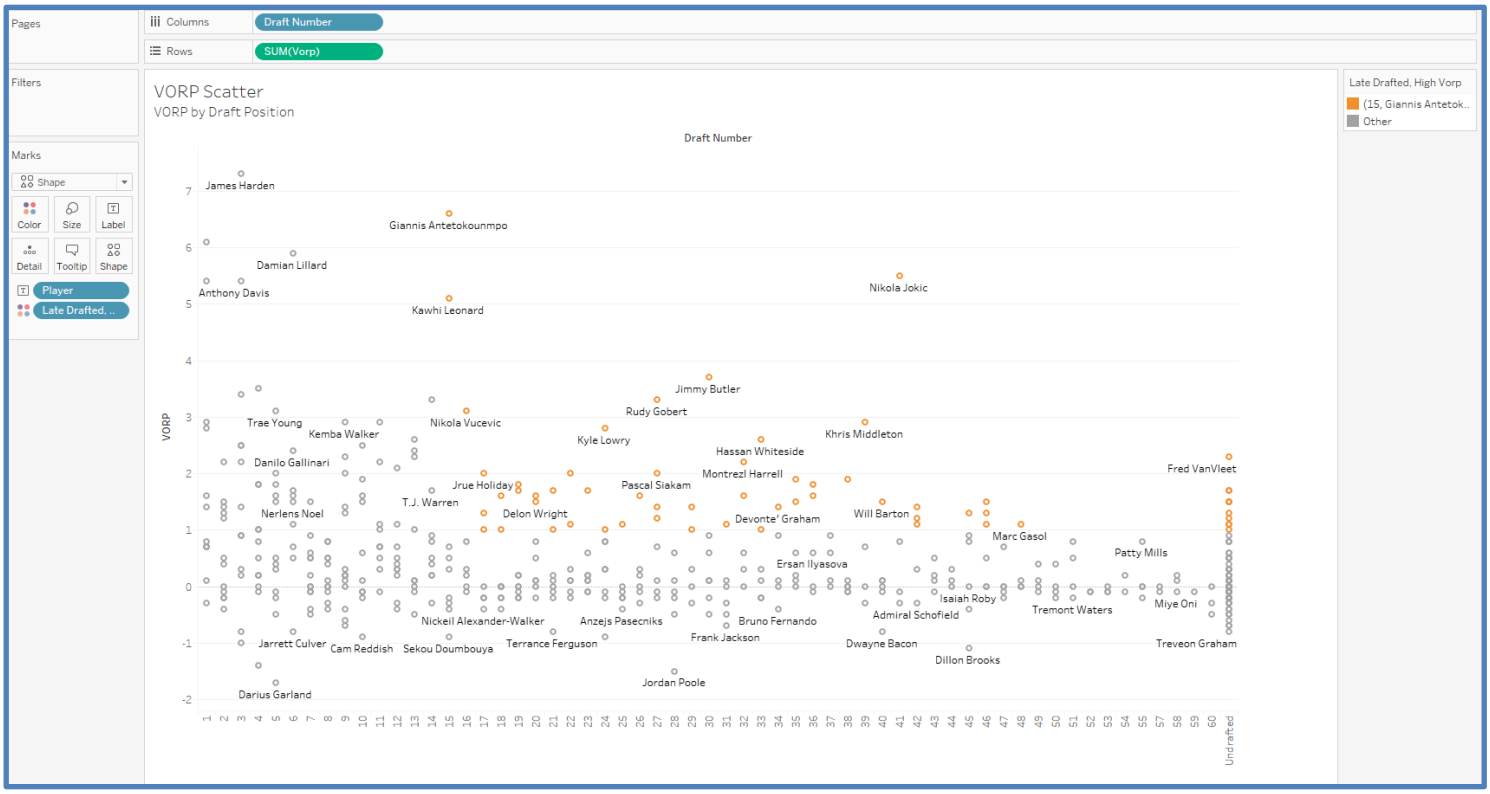
Most of the players fall in the -0.5 to 0.5 range, which basically means that the player is average. This makes sense that most players would fall in here. VORP is a relative measure. It also makes sense that there would not be many players on the low side as they would be out of the league fairly quickly because of their poor play. Also, it is rare to find superstar players so the low number of high VORP players aligns with this idea.

Then, with the NBA draft upcoming this week, I wanted to see how good teams are at drafting. That is, do the players chosen first, have higher VORPs than those chosen later. I created the following chart to find out.



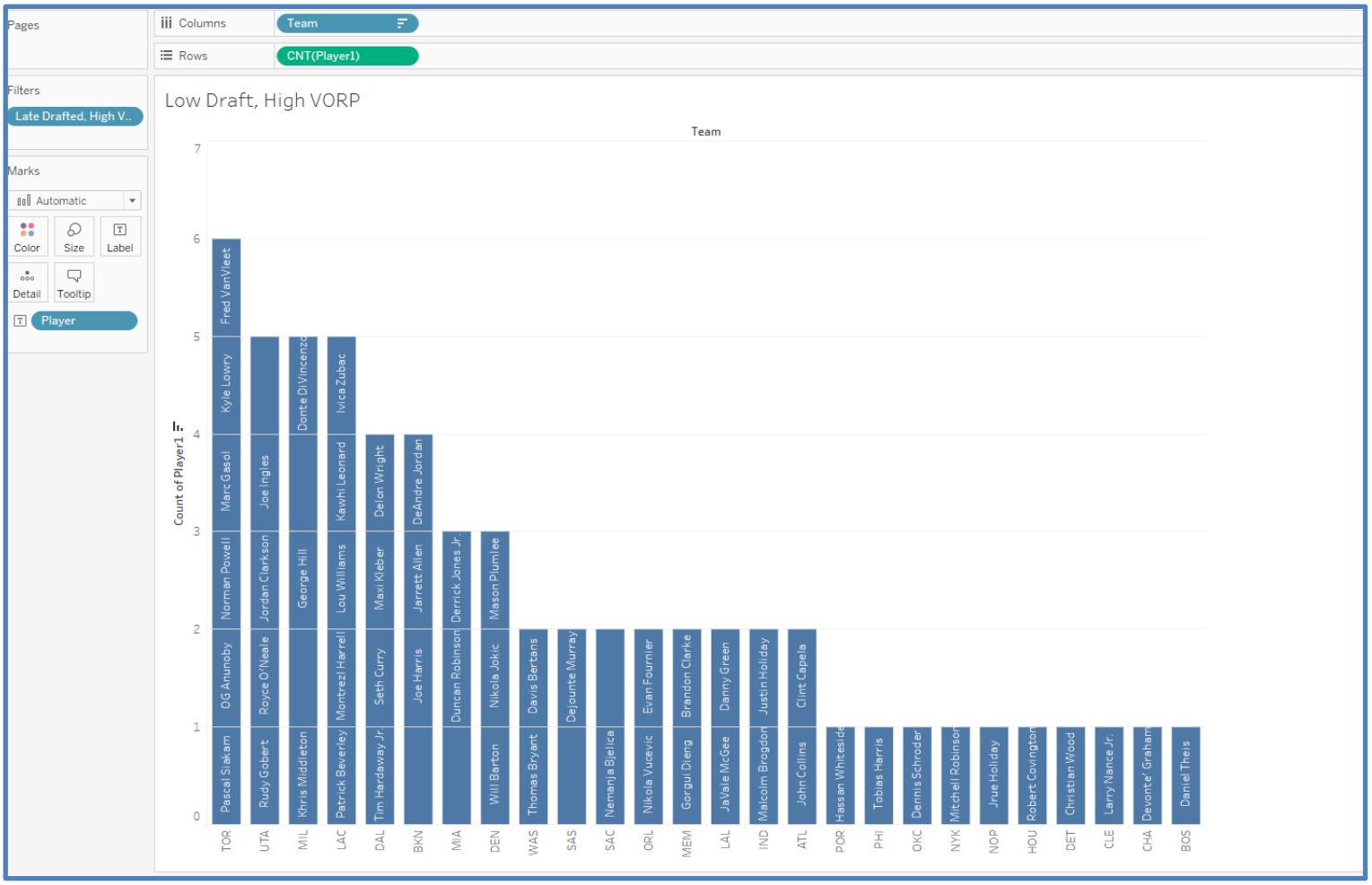
As the chart shows, there tends to be higher average VORP for higher draft picks. However, drafting players is not an exact science, as many later picks end up having higher average VORP than earlier ones. I also added a filter to this chart to differentiate by position. Filtering for each position looked similar in patterns to the overall chart. This showed that it is just as hard to scout a particular position as it is to scout any player in general.

Next, I wanted to see if there were any specific players who made the average VORP for the draft position especially high. I was looking for outliers. To find out, I created the following scatterplot.



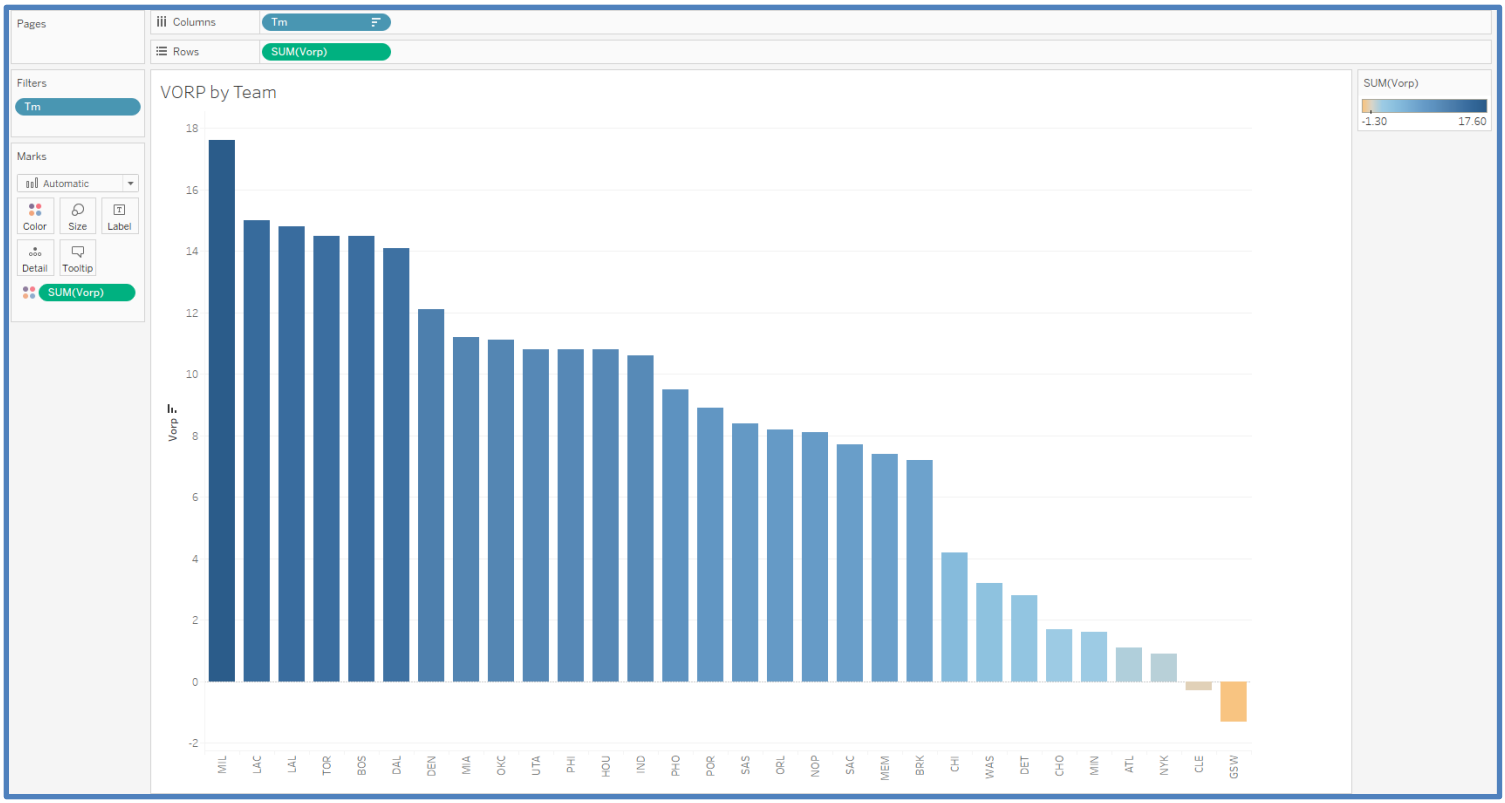
This plotted every player according to their draft position (x-axis) and their individual VORP (y-axis). The players in the top half could be considered outliers for their draft slot - they have high VORP, which drives up the average VORP for their draft position. For example, Nikola Jokic was drafted 41st overall and has a VORP of 5.5. He is an extreme outlier. The next highest VORP player in the 41st slot, Pat Connaughton, has a VORP of 0.8. I also noticed a high number of players with relatively high VORPs at the far right of the graph. These players were undrafted. It makes sense that some would have higher VORPs, because this dataset includes only current players, and to stay in the league after not being drafted, the player must have proved himself with high on-court performance.

Plotting VORP relative to draft position made me curious whether some teams are better at drafting than others. Put more clearly, have some teams drafted high VORP players late more frequently than other teams. To answer this, I grouped the players in the previous chart who had a low draft position (later than 14th overall) and a high VORP (>2) and created the following histogram with that group.



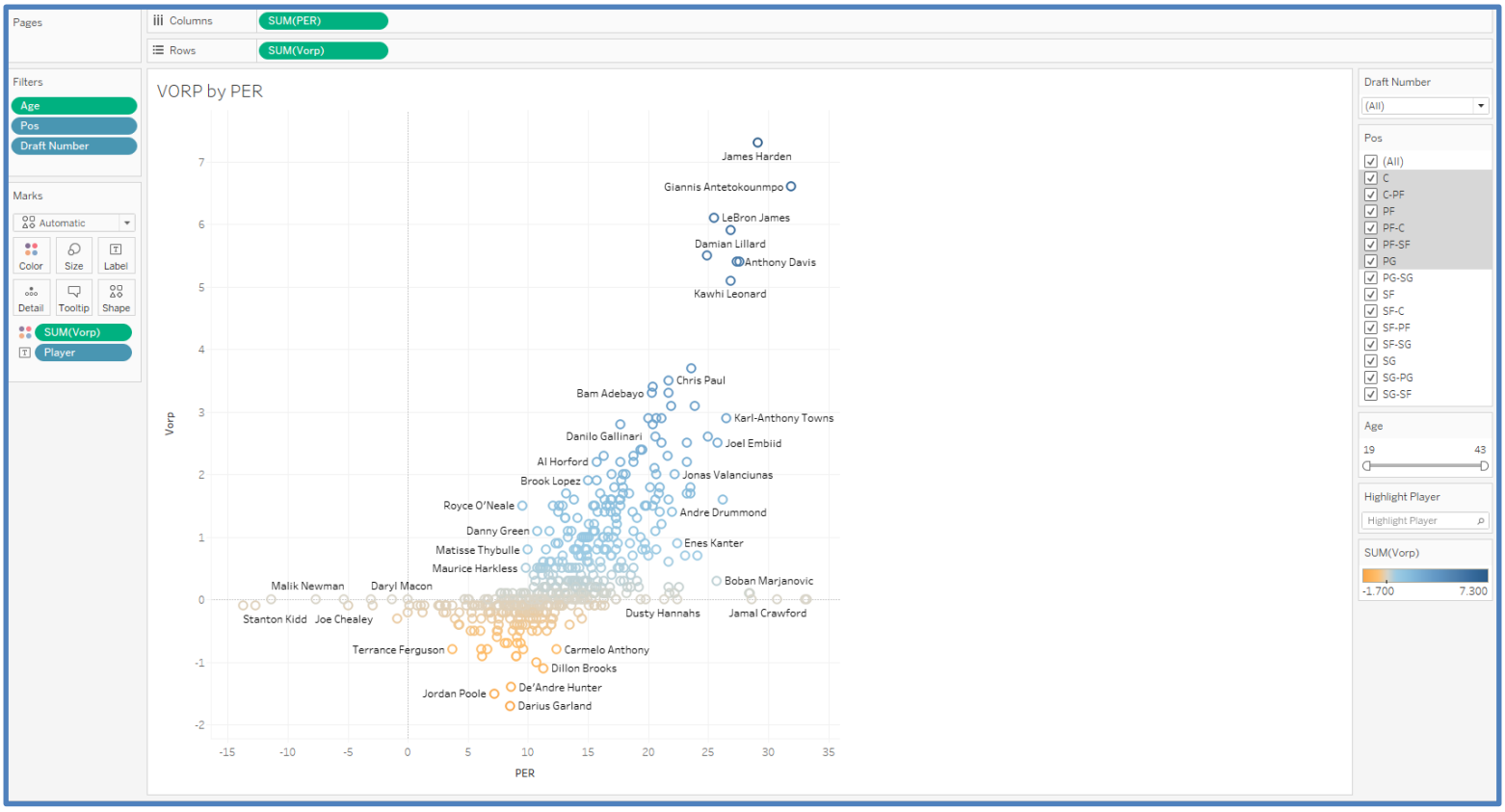
This chart showed me that the Toronto Raptors, Utah Jazz, Milwaukee Bucks, and L.A. Clippers seem to be successful at finding quality players late in the draft. Qualitatively, Toronto has a reputation in the league of excellent scouting and player development; this graph confirms that. To create this graph, I used 'Team' in the columns and 'Player' count in the rows. I sorted by ascending count and added player name to the label marks.

Next, I wanted to see which team had the best cumulative VORP. To find this out I had to graph which team's players had the highest sum of VORPs. The following chart shows the results.



The teams with the highest VORPs included Milwaukee, L.A. Clippers, L.A. Lakers, and Toronto Raptors. I saw in the previous step how Toronto accumulates VORP by drafting well, but the other teams in this list may draft star players early in the draft or sign them via free agency. This list makes sense given that the teams at the top had deep playoff runs and the worst (Cleveland and Golden State) lost most of their games this season.

Next, I wanted to compare VORP with another advanced metric used widely in the NBA – Player Efficiency Rating (PER). Like VORP, PER tries to combine all of a player's effect on the court into a single number. There is sometimes disputes in the media about which metric is better, so I wanted to see how they compare. It turns out that, for this past season at least, they provide consistent results, as shown in the graph below.



Overall, this graph shows that players with the highest VORP also have the highest PER, and vice versa. One interesting thing to me though, is that a group of the league's superstars are clustered far above the rest of the players. Players like Giannis Antetokounmpo and LeBron James are part of a group that had much higher VORP and PER than the rest of the league. The cluster above shows that there is a definitive tier system in the league, at least between the 'superstars' and the rest of the league. I did not expect such a big gap.

I added a position filter and an age parameter to this graph. It was interesting to see which player was the best in his position. No position stood above the others though. And it was also interesting to sort by age, and see who the next generation of superstars are, the players like Luka Doncic and Nikola Jokic who are young and already are performing incredibly well on the court. These are the players that team general managers should build their teams around.