

# Clustering NBA Teams by 2019/20 season averages

Tucker Paron

2022

## Finding the Determinant

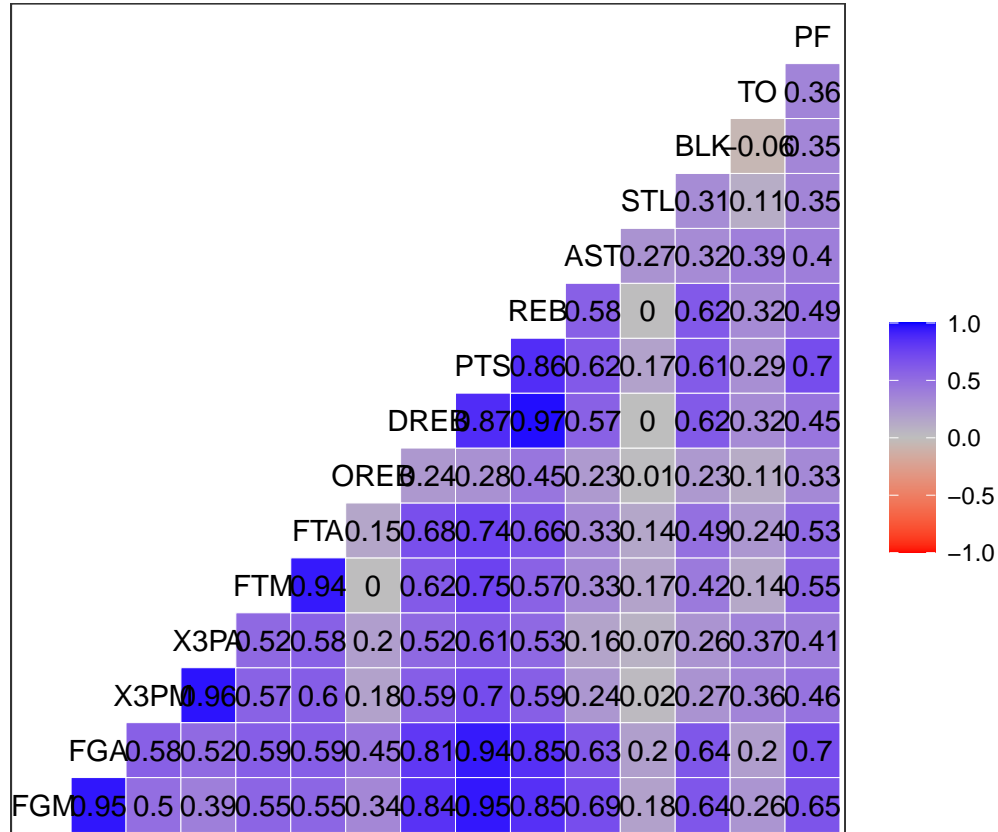
If the value is close to zero, there is evidence of linear dependence - meaning multiple variables are heavily associated and are not both necessary in the analysis. Fouls and yellow cards would be likely be highly correlated for example.

```
## [1] -5.88e-39
```

Since the determinant is extremely close to zero, there is evidence of linear dependence - but it is not for certain.

## Correlation Plot

Here we can see the actually relationship between every variable (metric). Values close to 1 define a strong positive relationship and -1 defines a strong negative relationship. values closer to 0 indicate weak to no relationship.



There are a significant amount of pairings that have strong correlations. Understandably all the attempt variables (FG, 3PT, and FT) are very highly correlated with their corresponding 'makes' variables. Additionally, FGA and FGM have very high correlations with Points. Interestingly, Defensive Rebounds are nearly perfectly correlated with Rebounds, but Offensive Rebounds have only a moderate correlation.

### Principal Component Analysis

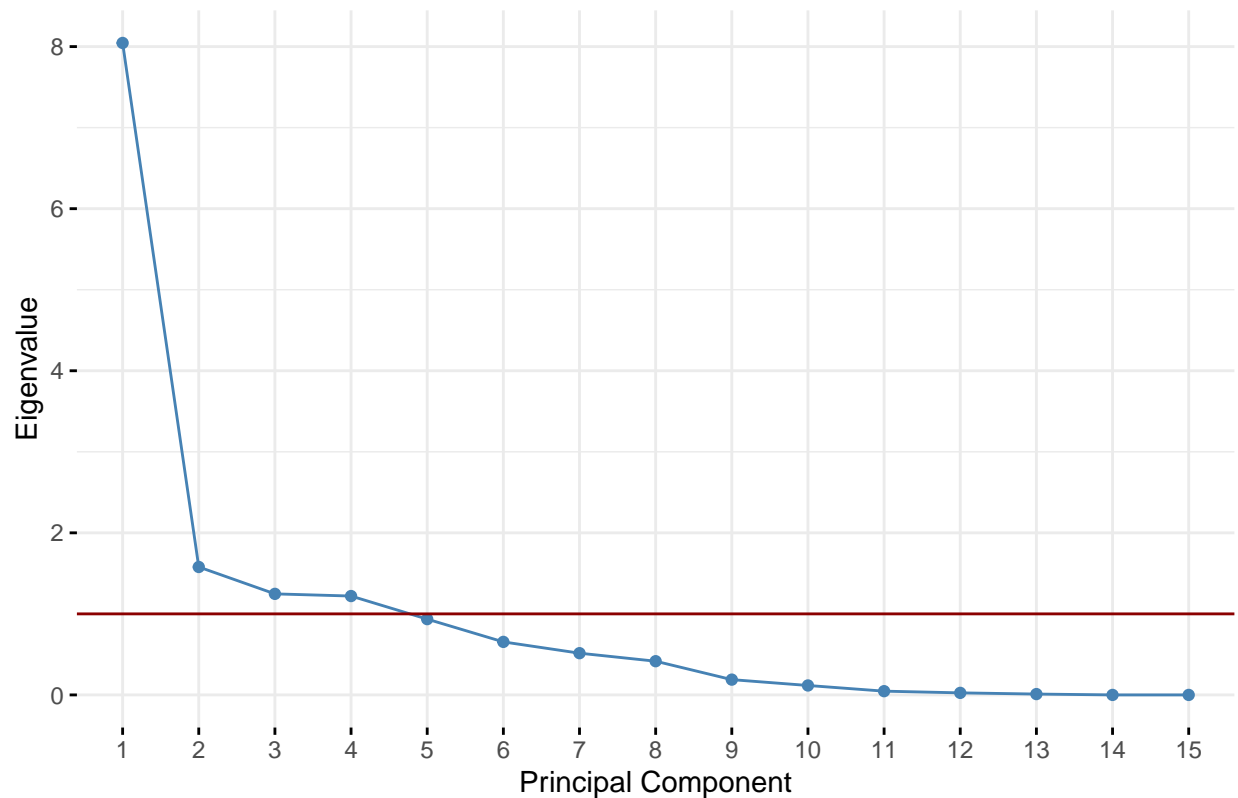
## Importance of components:

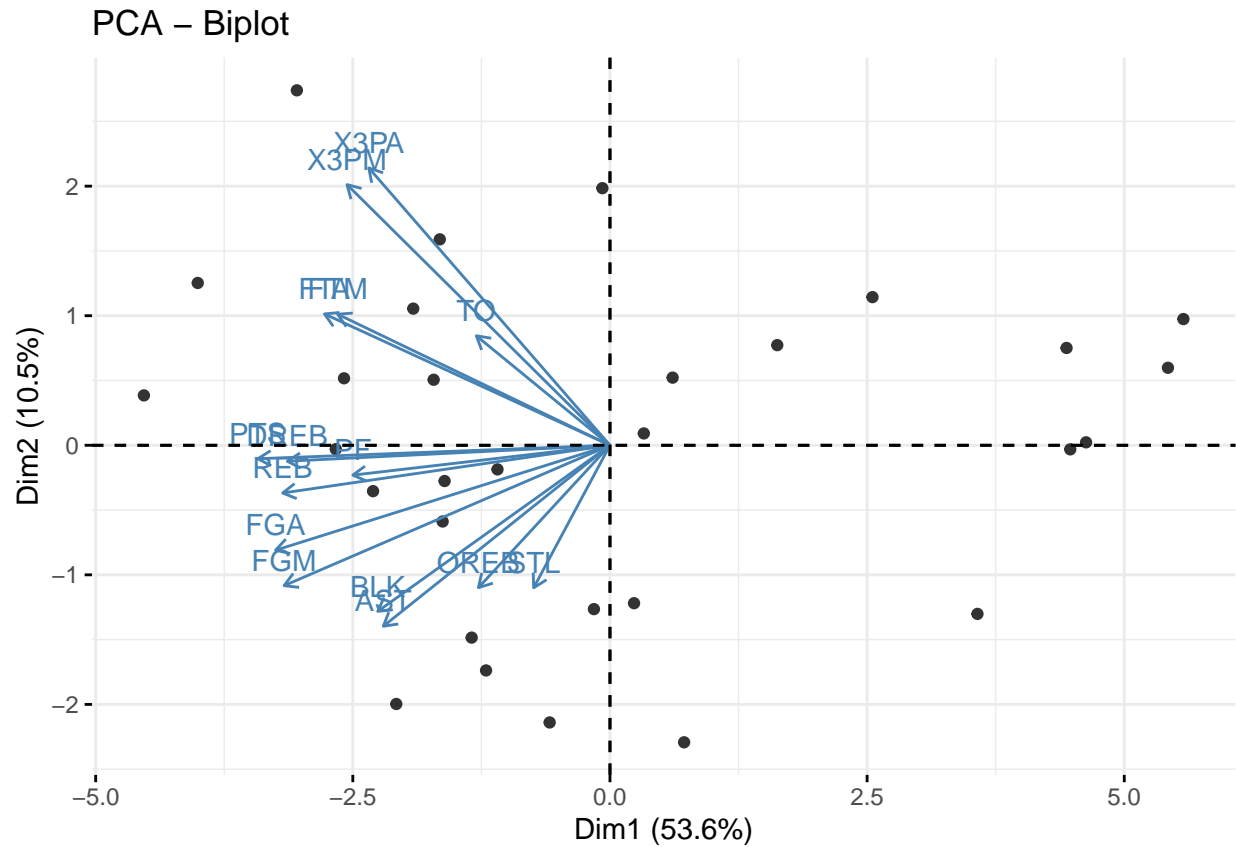
##	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
## Standard deviation	2.836	1.257	1.1167	1.1046	0.9672	0.8090	0.7178	0.6451
## Proportion of Variance	0.536	0.105	0.0831	0.0813	0.0624	0.0436	0.0343	0.0277
## Cumulative Proportion	0.536	0.642	0.7247	0.8061	0.8684	0.9120	0.9464	0.9741

##	PC9	PC10	PC11	PC12	PC13	PC14	PC15
## Standard deviation	0.4345	0.34175	0.21540	0.1596	0.10291	1.04e-15	2.27e-16
## Proportion of Variance	0.0126	0.00779	0.00309	0.0017	0.00071	0.00e+00	0.00e+00
## Cumulative Proportion	0.9867	0.99450	0.99760	0.9993	1.00000	1.00e+00	1.00e+00

Screplot using the Covariance Matrix



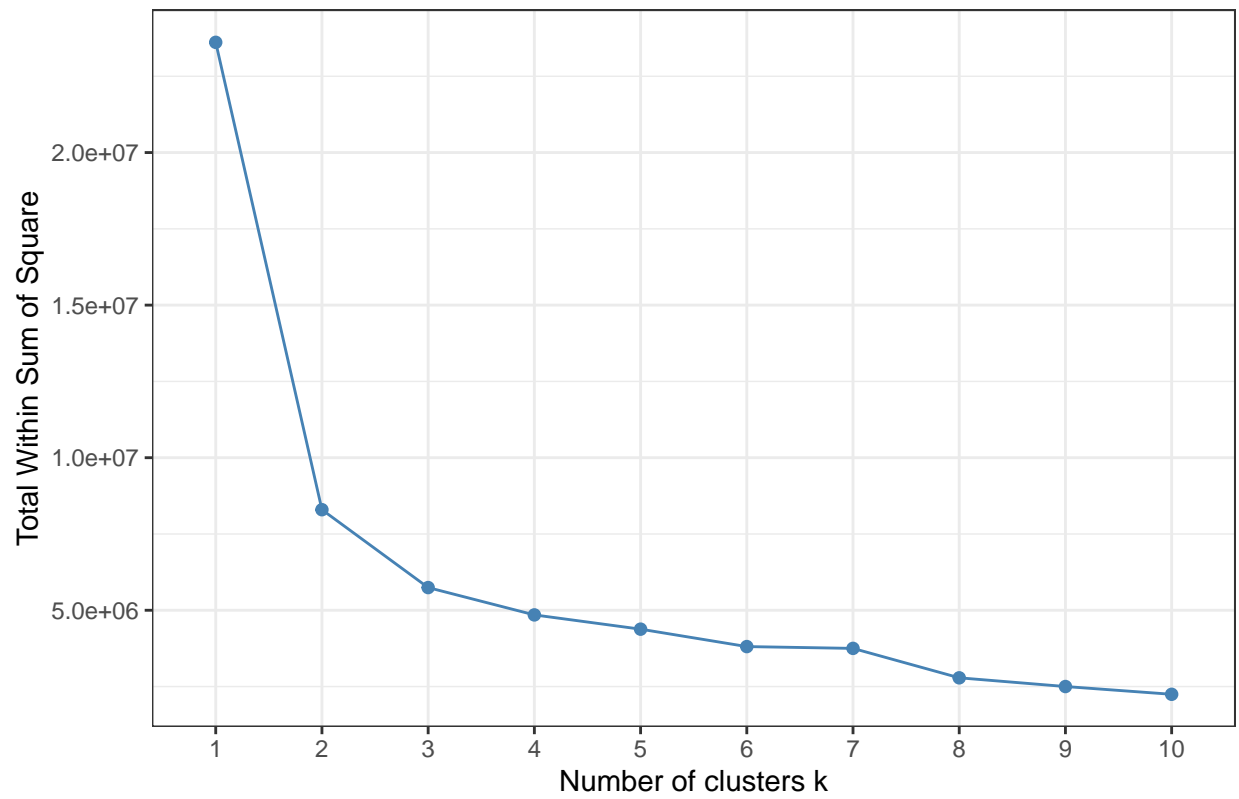


Looking at the biplot, several pairs of variables appear to be linearly dependent. These are primarily the pairs we identified when looking at the correlation plot (FGA-FGM, FTA-FTM, etc.) The scree plot also indicates a need for PCA with no perfect elbow but a suggested k value of 4 (number of clusters).

### Number of Clusters...

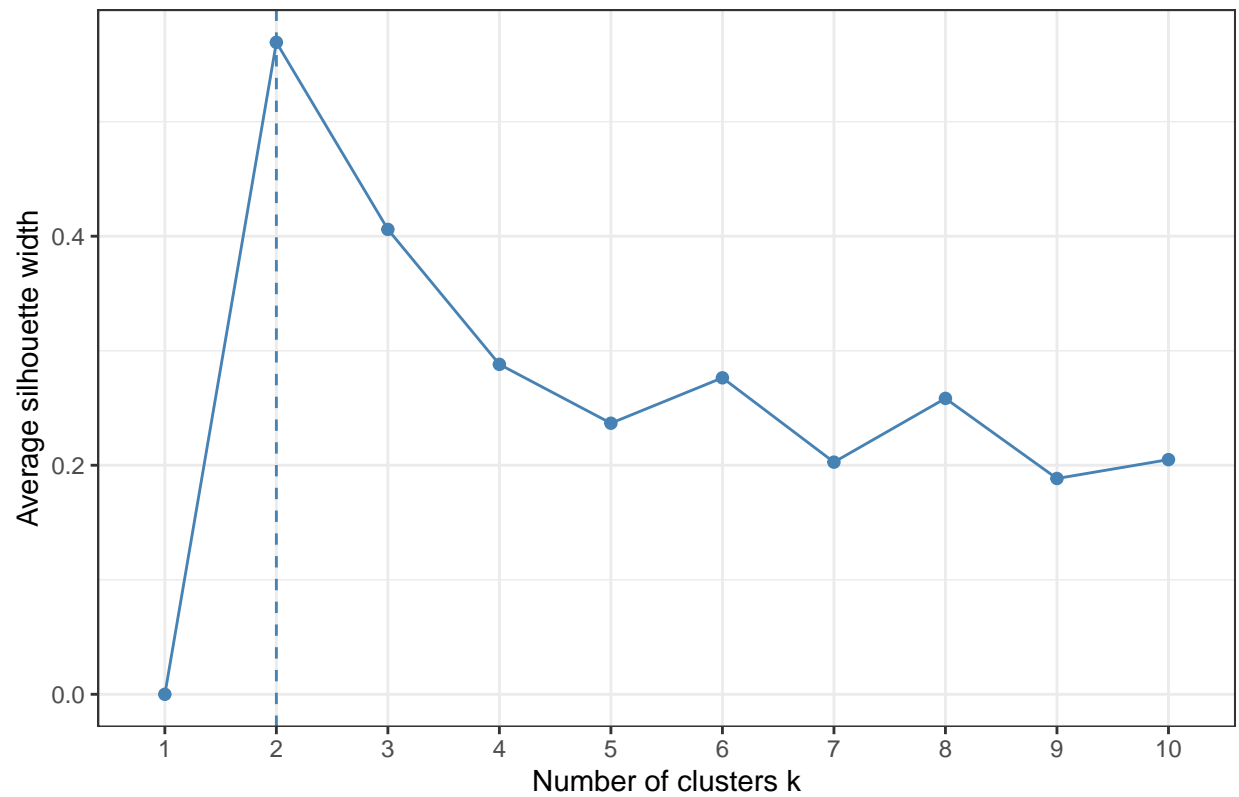
We need to determine how many clusters or categories of teams there are likely to be (ie. high-pressing, possession-based, etc.)

Choosing K for Old Faithful Dataset Using WSS

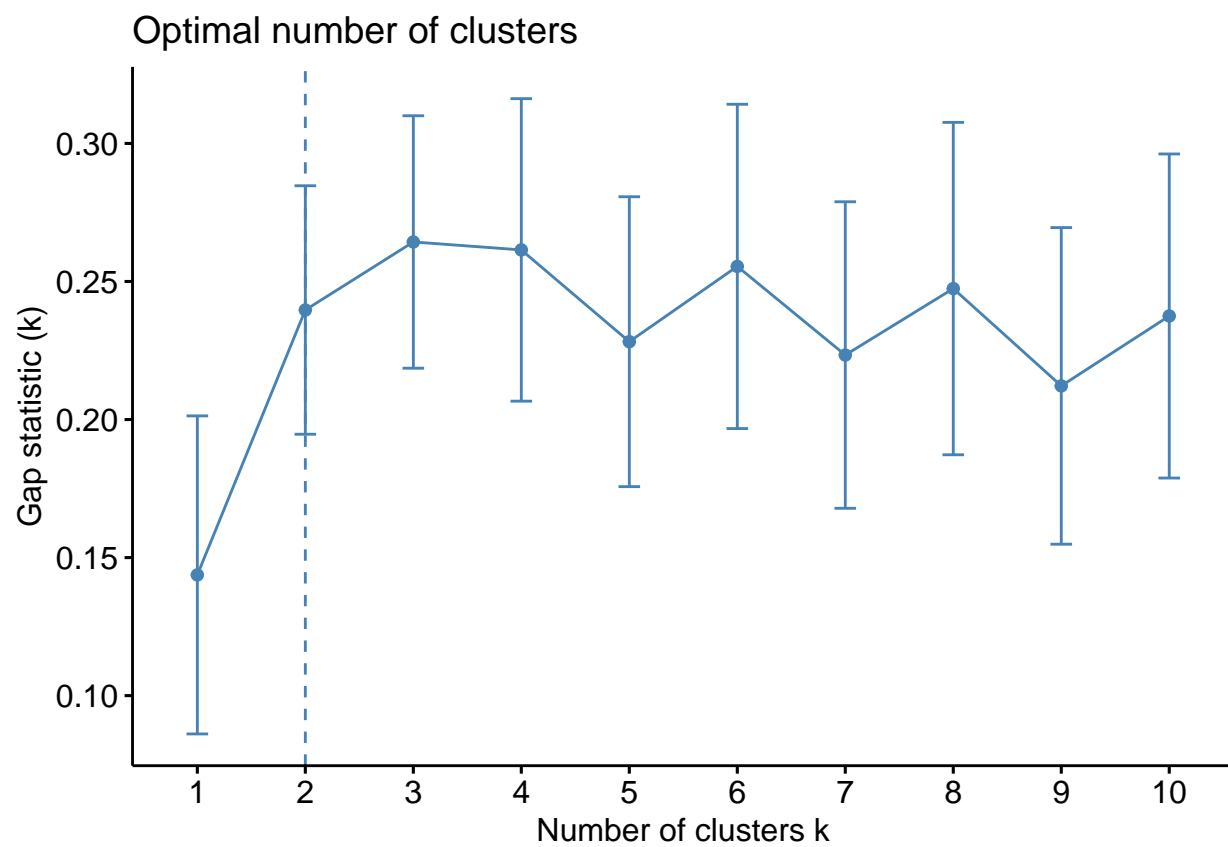


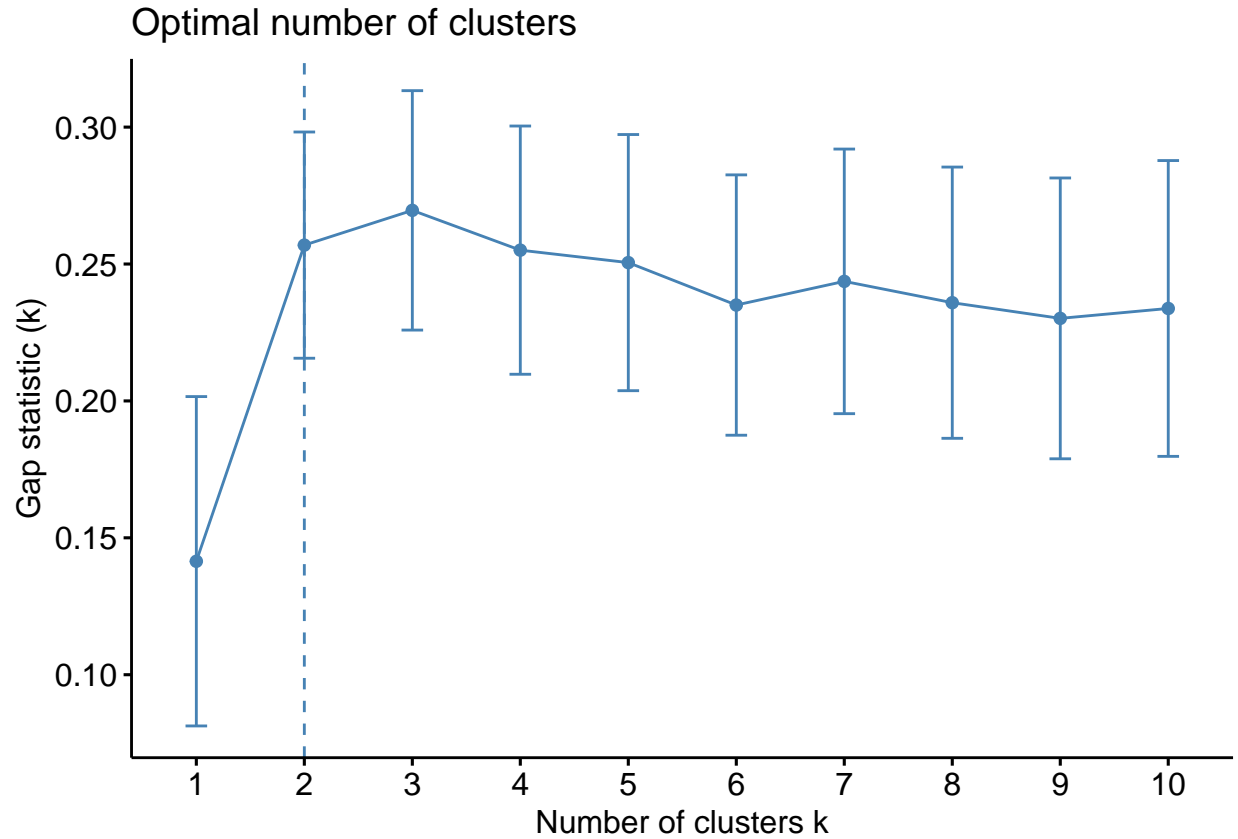
We should use k=3 clusters as this is the value at the elbow of the plot. This is the cutoff where we can capture the most variation within the smallest amount of PC's.

### Choosing K for Old Faithful Dataset Using WSS



We should use  $k=2$  clusters as this is the value with the highest silhouette score. This means the observations in each cluster are closest to one another at this amount of clusters.





The optimal number of clusters here is 2 as is indicated by the gap statistic plot. This is because it is the smallest value of  $k$  where there was the biggest jump in variation within clusters.

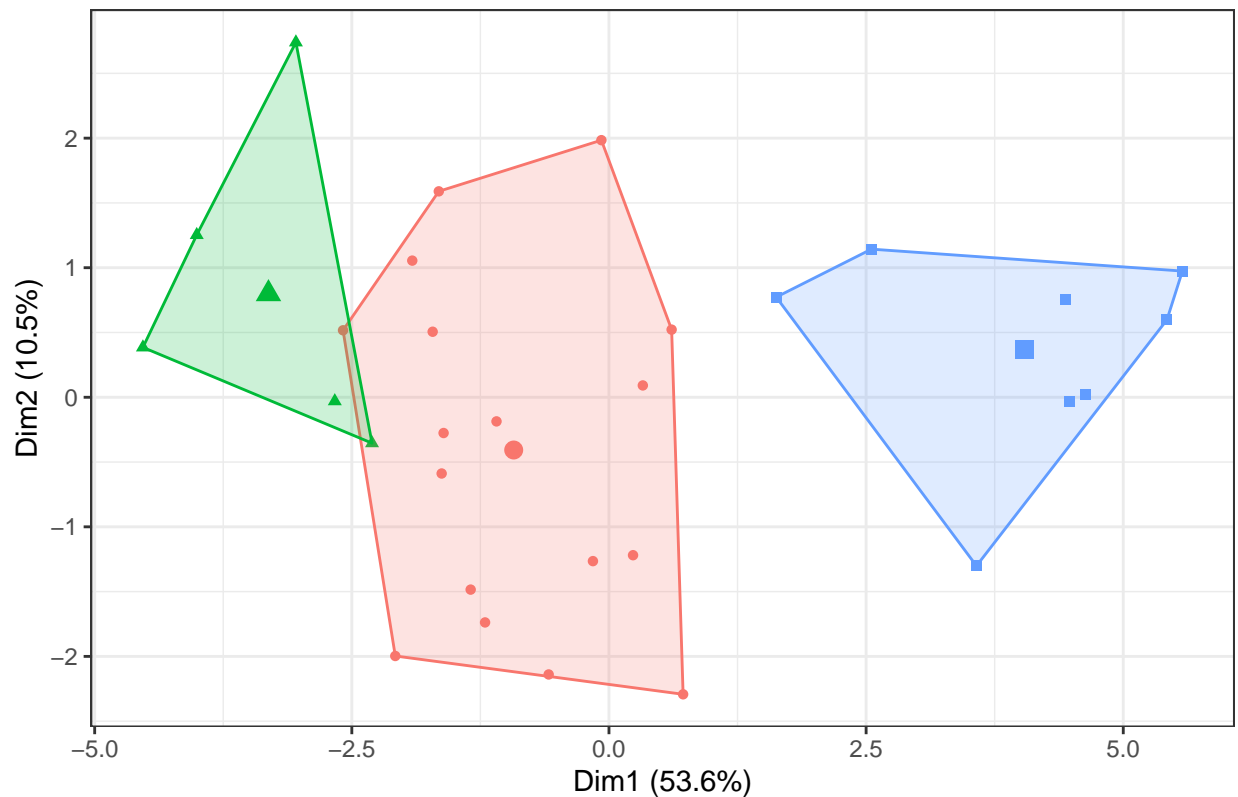
From these three methods I would recommend using  $k=3$  clusters as it could be a reasonable compromise between all of the methods; However, we'll also try using 2 clusters as this was suggested by both the gap statistic and silhouette score plots.

### Clustering the Teams

Now we place the teams in their respective clusters according to how closely they relate to one another based on our variables.

### 3 Clusters

## NBA Data Clustered with k = 3



```
## [1] 17  5  8
```

There is 17 team in cluster 1, 5 teams in cluster 2, and 8 teams in cluster 3. The clusters appear distinct and could indicate different play types or possibly just overall quality. Given the variables, the latter is more likely in my opinion.

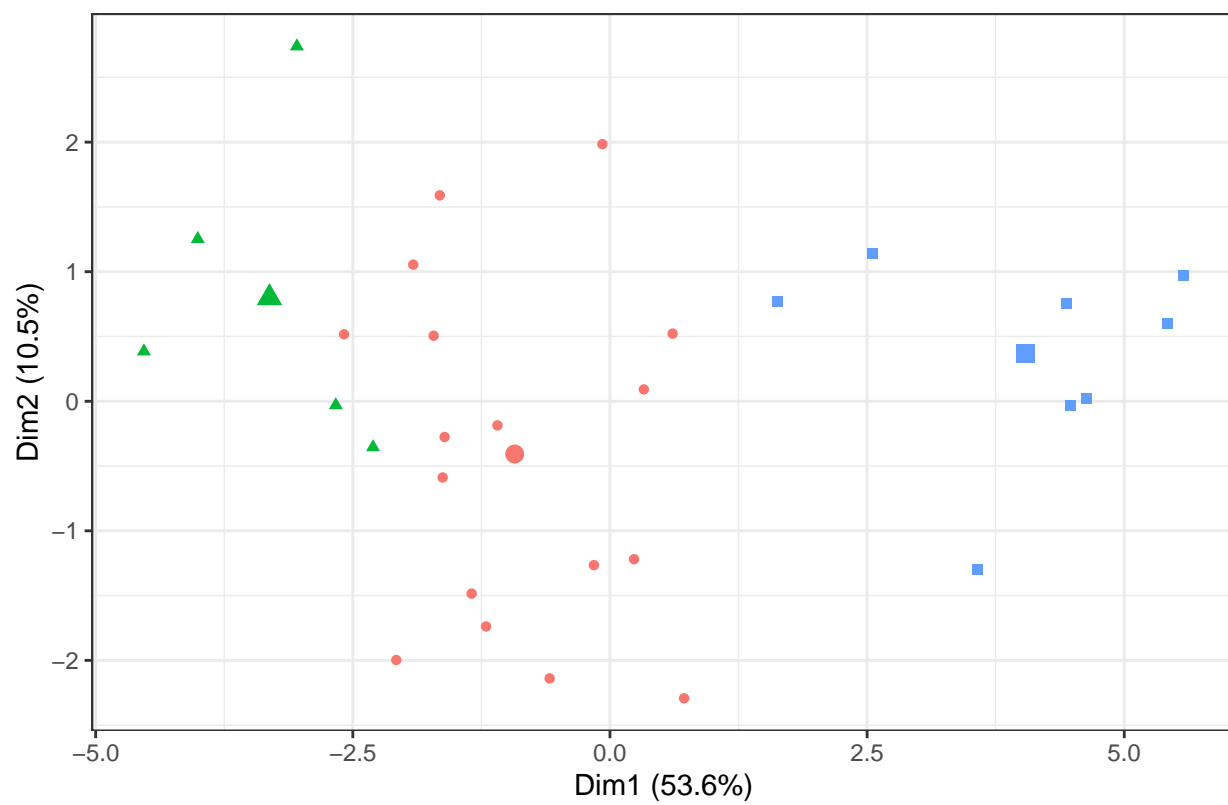
**Find the averages of each metric by cluster...**

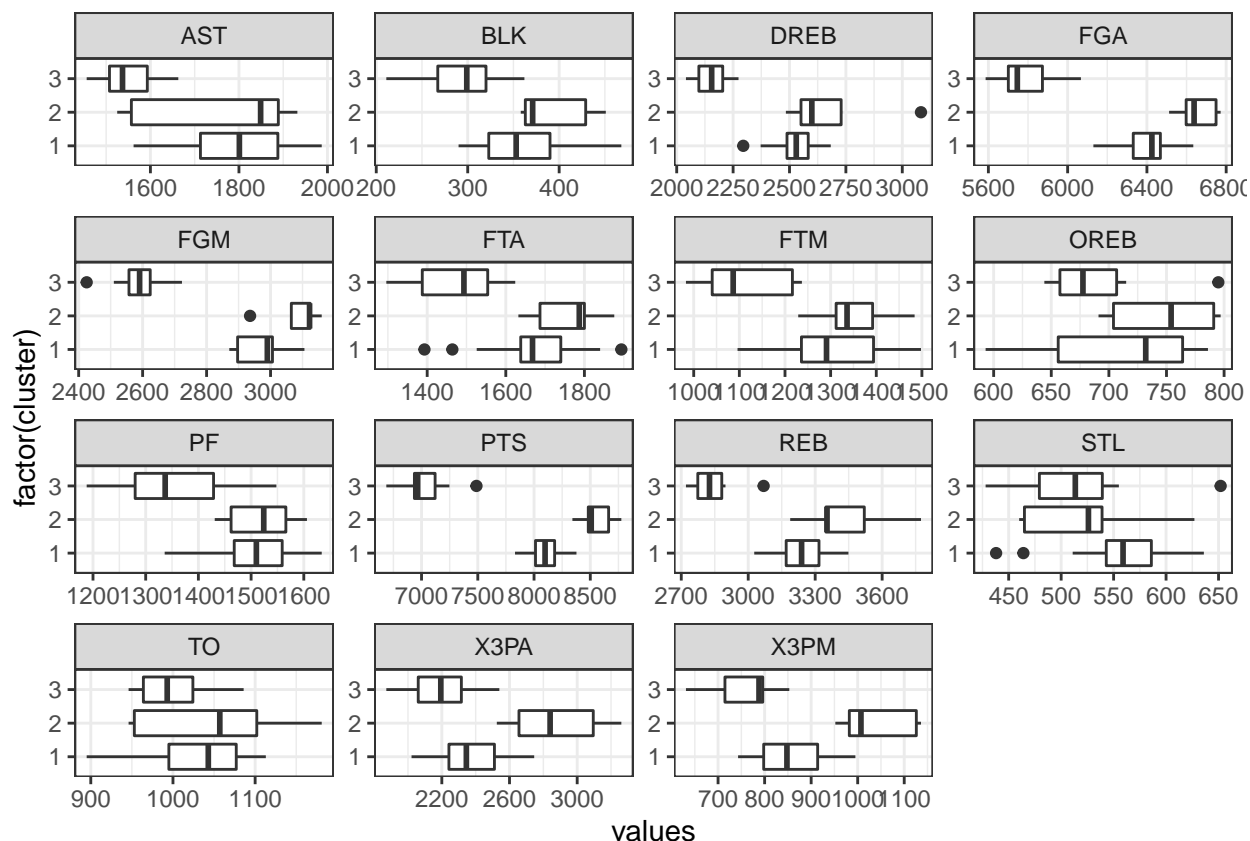
```
##      FGM  FGA  X3PM  X3PA  FTM  FTA  OREB  DREB  PTS  REB  AST  STL  BLK  TO  PF
## 1 2969 6386   859 2369 1305 1671   711 2526 8102 3237 1794 556 356 1032 1506
## 2 3081 6654 1041 2875 1351 1756   747 2689 8554 3437 1750 523 394 1048 1518
## 3 2584 5785   758 2200 1111 1470   690 2154 7036 2844 1550 517 295 1002 1352
```

The above table presents the averages of each variable by cluster. Each row is represented by a cluster and each column by a variable, with the value being the corresponding average.

**Visualize these averages and get a sense for the range of teams within each cluster...**







From the boxplot, it appears it could simply be based on quality with the second cluster having the highest scoring metrics followed by cluster 1 and then by cluster 3.

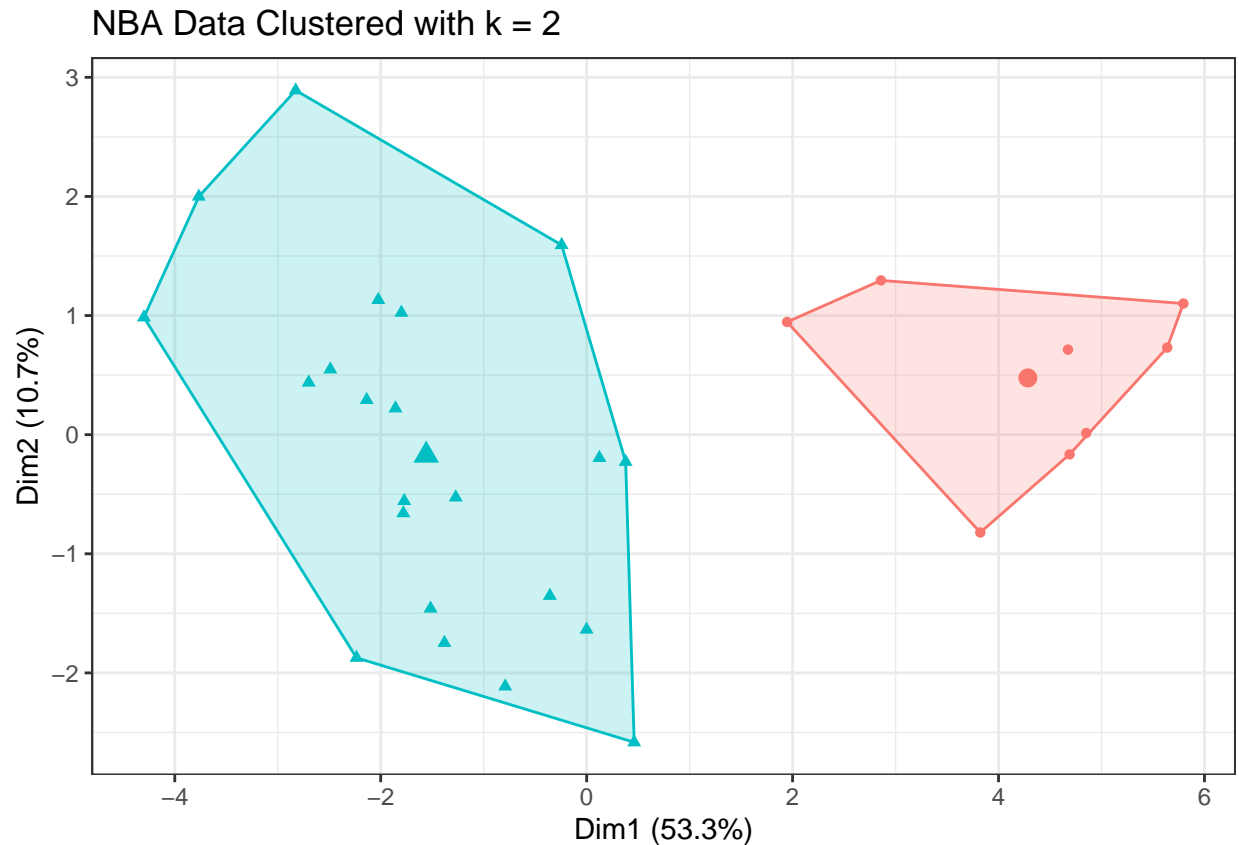
Print cluster observations...

##	Rank	Team	G	FGM	FGA	X3PM	X3PA	FTM	FTA	OREB	DREB	PTS
## 1	1	Milwaukee Bucks	73	3160	6638	1007	2840	1336	1800	691	3083	8663
## 2	2	Houston Rockets	72	2936	6512	1126	3261	1484	1876	704	2484	8482
## 3	3	Dallas Mavericks	75	3124	6772	1136	3095	1392	1787	791	2729	8776
## 4	4	LA Clippers	72	2992	6425	895	2410	1498	1894	767	2664	8377
## 5	5	New Orleans Pelicans	72	3065	6598	982	2656	1229	1687	797	2551	8341
## 6	6	Portland Trail Blazers	74	3122	6749	952	2525	1312	1632	754	2599	8508
## 7	7	Washington Wizards	72	2990	6544	864	2345	1394	1770	732	2295	8238
## 8	8	San Antonio Spurs	71	2995	6350	760	2021	1348	1664	640	2530	8098
## 9	9	Boston Celtics	72	2971	6448	905	2487	1336	1668	768	2549	8183
## 10	10	Phoenix Suns	73	3006	6429	831	2320	1451	1740	712	2465	8294
## 11	11	Los Angeles Lakers	71	3006	6269	782	2242	1260	1728	757	2490	8054
## 12	12	Minnesota Timberwolves	64	2586	5865	853	2540	1223	1624	675	2193	7248
## 13	13	Toronto Raptors	72	2897	6331	995	2663	1329	1670	685	2583	8118
## 14	14	Memphis Grizzlies	73	3106	6634	798	2297	1212	1589	754	2643	8222
## 15	15	Miami Heat	73	2880	6160	979	2584	1440	1840	620	2619	8179
## 16	16	Atlanta Hawks	67	2723	6067	805	2416	1237	1566	661	2237	7488
## 17	17	Brooklyn Nets	72	2908	6498	941	2746	1291	1732	766	2683	8048
## 18	18	Utah Jazz	72	2886	6130	963	2537	1278	1641	656	2577	8013
## 19	19	Denver Nuggets	73	3069	6488	801	2234	1185	1526	786	2436	8124
## 20	20	Philadelphia 76ers	73	3000	6417	848	2307	1236	1638	764	2553	8084
## 21	21	Oklahoma City Thunder	72	2879	6156	770	2171	1422	1787	593	2495	7950

##	22	22		Sacramento Kings	72	2943	6364	914	2511	1127	1464	695	2372	7927
##	23	23		Indiana Pacers	73	3075	6457	743	2046	1096	1393	636	2489	7989
##	24	24		Orlando Magic	73	2871	6468	807	2354	1282	1656	751	2498	7831
##	25	25		Detroit Pistons	66	2595	5658	791	2157	1097	1476	644	2109	7078
##	26	26		Cleveland Cavaliers	65	2619	5715	727	2070	983	1296	704	2172	6948
##	27	27		Chicago Bulls	65	2573	5762	793	2282	1006	1332	680	2041	6945
##	28	28		Golden State Warriors	65	2510	5730	678	2032	1214	1511	647	2138	6912
##	29	29		New York Knicks	66	2638	5896	631	1872	1076	1550	795	2274	6983
##	30	30		Charlotte Hornets	65	2425	5586	785	2231	1052	1406	715	2066	6687
##		REB	AST	STL	BLK	TO	PF	cluster						
##	1	3774	1889	526	429	1102	1431	2						
##	2	3188	1557	627	371	1057	1566	2						
##	3	3520	1849	460	363	953	1462	2						
##	4	3431	1708	511	338	1051	1594	1						
##	5	3348	1932	539	358	1181	1524	2						
##	6	3353	1525	465	451	946	1606	2						
##	7	3027	1801	574	309	1020	1634	1						
##	8	3170	1751	518	390	895	1377	1						
##	9	3317	1655	595	406	995	1553	1						
##	10	3177	1987	559	290	1077	1609	1						
##	11	3247	1803	612	468	1077	1471	1						
##	12	2868	1525	555	362	981	1370	3						
##	13	3268	1816	636	358	1067	1559	1						
##	14	3397	1965	576	398	1113	1546	1						
##	15	3239	1892	547	332	1088	1501	1						
##	16	2898	1605	523	341	1086	1548	3						
##	17	3449	1761	464	323	1099	1510	1						
##	18	3233	1615	438	292	1089	1468	1						
##	19	3222	1952	586	337	1005	1485	1						
##	20	3317	1884	586	389	1037	1529	1						
##	21	3088	1562	550	353	986	1390	1						
##	22	3067	1713	556	293	1043	1595	1						
##	23	3125	1888	543	376	967	1445	1						
##	24	3249	1747	599	396	937	1336	1						
##	25	2753	1589	490	299	1008	1299	3						
##	26	2876	1500	447	211	1073	1188	3						
##	27	2721	1510	652	265	1005	1417	3						
##	28	2785	1663	534	299	969	1304	3						
##	29	3069	1456	504	313	946	1464	3						
##	30	2781	1549	428	268	949	1223	3						

It seems like the previous assumption was correct. The top teams are in cluster 2, the middle-of-the-road teams are generally 1 and those that fall at the bottom are in cluster 3

## 2 Clusters



```
## [1] 8 22
```

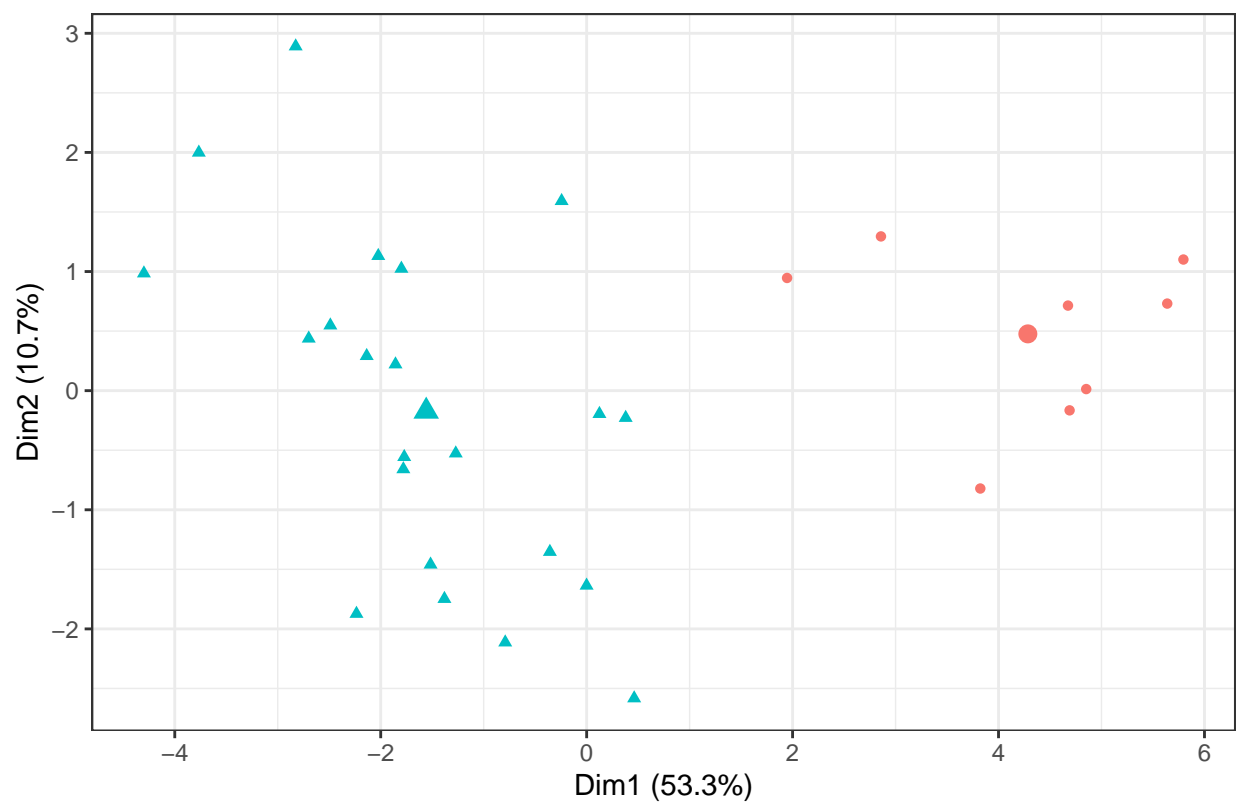
The clusters are of size 22 and 8 respectively. Again, I'd assume this is based generally on quality.

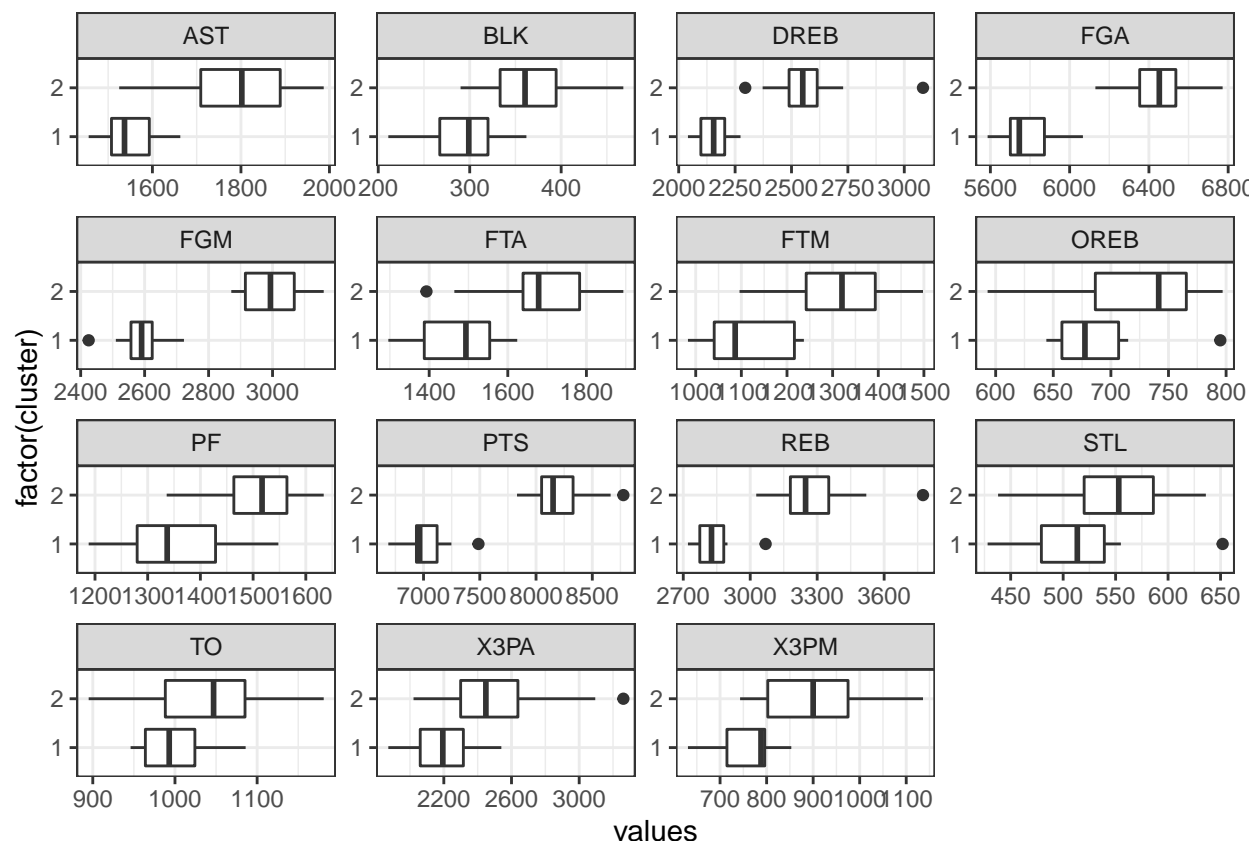
Find the averages of each metric by cluster...

```
##      FGM  FGA  X3PM  X3PA  FTM  FTA  OREB  DREB  PTS  REB  AST  STL  BLK  TO  PF
## 1  2584  5785   758  2200  1111  1470   690  2154  7036  2844  1550  517  295  1002  1352
## 2  2995  6447   900  2484  1315  1690   719  2563  8204  3282  1784  548  365  1036  1509
##   cluster
## 1      3.00
## 2      1.23
```

Again, the above table shows the average of each variable by cluster.

Visualize these averages and get a sense for the range of teams within each cluster...





Cluster 2 appears stronger in every category than cluster 1.

Print cluster observations...

##	Rank	Team	G	FGM	FGA	X3PM	X3PA	FTM	FTA	OREB	DREB	PTS
## 1	1	Milwaukee Bucks	73	3160	6638	1007	2840	1336	1800	691	3083	8663
## 2	2	Houston Rockets	72	2936	6512	1126	3261	1484	1876	704	2484	8482
## 3	3	Dallas Mavericks	75	3124	6772	1136	3095	1392	1787	791	2729	8776
## 4	4	LA Clippers	72	2992	6425	895	2410	1498	1894	767	2664	8377
## 5	5	New Orleans Pelicans	72	3065	6598	982	2656	1229	1687	797	2551	8341
## 6	6	Portland Trail Blazers	74	3122	6749	952	2525	1312	1632	754	2599	8508
## 7	7	Washington Wizards	72	2990	6544	864	2345	1394	1770	732	2295	8238
## 8	8	San Antonio Spurs	71	2995	6350	760	2021	1348	1664	640	2530	8098
## 9	9	Boston Celtics	72	2971	6448	905	2487	1336	1668	768	2549	8183
## 10	10	Phoenix Suns	73	3006	6429	831	2320	1451	1740	712	2465	8294
## 11	11	Los Angeles Lakers	71	3006	6269	782	2242	1260	1728	757	2490	8054
## 12	12	Minnesota Timberwolves	64	2586	5865	853	2540	1223	1624	675	2193	7248
## 13	13	Toronto Raptors	72	2897	6331	995	2663	1329	1670	685	2583	8118
## 14	14	Memphis Grizzlies	73	3106	6634	798	2297	1212	1589	754	2643	8222
## 15	15	Miami Heat	73	2880	6160	979	2584	1440	1840	620	2619	8179
## 16	16	Atlanta Hawks	67	2723	6067	805	2416	1237	1566	661	2237	7488
## 17	17	Brooklyn Nets	72	2908	6498	941	2746	1291	1732	766	2683	8048
## 18	18	Utah Jazz	72	2886	6130	963	2537	1278	1641	656	2577	8013
## 19	19	Denver Nuggets	73	3069	6488	801	2234	1185	1526	786	2436	8124
## 20	20	Philadelphia 76ers	73	3000	6417	848	2307	1236	1638	764	2553	8084
## 21	21	Oklahoma City Thunder	72	2879	6156	770	2171	1422	1787	593	2495	7950
## 22	22	Sacramento Kings	72	2943	6364	914	2511	1127	1464	695	2372	7927

##	23	23		Indiana Pacers	73	3075	6457	743	2046	1096	1393	636	2489	7989
##	24	24		Orlando Magic	73	2871	6468	807	2354	1282	1656	751	2498	7831
##	25	25		Detroit Pistons	66	2595	5658	791	2157	1097	1476	644	2109	7078
##	26	26		Cleveland Cavaliers	65	2619	5715	727	2070	983	1296	704	2172	6948
##	27	27		Chicago Bulls	65	2573	5762	793	2282	1006	1332	680	2041	6945
##	28	28		Golden State Warriors	65	2510	5730	678	2032	1214	1511	647	2138	6912
##	29	29		New York Knicks	66	2638	5896	631	1872	1076	1550	795	2274	6983
##	30	30		Charlotte Hornets	65	2425	5586	785	2231	1052	1406	715	2066	6687
##			REB	AST	STL	BLK	TO	PF	cluster					
##	1	3774	1889	526	429	1102	1431		2					
##	2	3188	1557	627	371	1057	1566		2					
##	3	3520	1849	460	363	953	1462		2					
##	4	3431	1708	511	338	1051	1594		2					
##	5	3348	1932	539	358	1181	1524		2					
##	6	3353	1525	465	451	946	1606		2					
##	7	3027	1801	574	309	1020	1634		2					
##	8	3170	1751	518	390	895	1377		2					
##	9	3317	1655	595	406	995	1553		2					
##	10	3177	1987	559	290	1077	1609		2					
##	11	3247	1803	612	468	1077	1471		2					
##	12	2868	1525	555	362	981	1370		1					
##	13	3268	1816	636	358	1067	1559		2					
##	14	3397	1965	576	398	1113	1546		2					
##	15	3239	1892	547	332	1088	1501		2					
##	16	2898	1605	523	341	1086	1548		1					
##	17	3449	1761	464	323	1099	1510		2					
##	18	3233	1615	438	292	1089	1468		2					
##	19	3222	1952	586	337	1005	1485		2					
##	20	3317	1884	586	389	1037	1529		2					
##	21	3088	1562	550	353	986	1390		2					
##	22	3067	1713	556	293	1043	1595		2					
##	23	3125	1888	543	376	967	1445		2					
##	24	3249	1747	599	396	937	1336		2					
##	25	2753	1589	490	299	1008	1299		1					
##	26	2876	1500	447	211	1073	1188		1					
##	27	2721	1510	652	265	1005	1417		1					
##	28	2785	1663	534	299	969	1304		1					
##	29	3069	1456	504	313	946	1464		1					
##	30	2781	1549	428	268	949	1223		1					

Again our assumption was accurate - cluster two had the higher ranking teams than cluster one. The clustering analysis using  $k=3$  had a bit more detail. We may have even benefitted from trying 4 or 5 clusters to see differences beyond overall quality (such as types of teams: three point heavy teams, post-focused teams, etc.)