

Visualizing NBA

Junqi Fu

2023-10-21

Contents

1	NBA Data	5
1.1	Data Type	5
1.2	Preliminary Cleaning	6
1.2.1	Initial Missing Data Imputation	6
1.2.2	Data Cleaning	7
1.3	Preliminary Descriptive Statistics	8
2	Hypothesis	11
2.1	Density Plot for Height and Weight	11
2.1.1	Density Plot for Height Distribution Over Seasons	11
2.1.2	Density Plot for Weight Distribution Over Seasons	12
2.1.3	General Conclusion	12
2.2	Inference & Hypothesis Based On The Density Plot and Initial Summary	13
3	Correlation Analysis	15
3.1	Correlation Between H/W and Others	15
3.2	Correlation Matrix	16
3.3	Correlation Between H/W and R/A	16
3.4	Summary From Scatter Plots	17
3.5	Net Rating By Player's Physical Attributes	18
3.6	Summary From Box Plot	19
3.6.1	Net Rating by Player Height	19
3.6.2	Net Rating by Player Weight	19
3.6.3	General Summary	19
4	In-depth Analysis for H/W	21
4.1	What Can Be Answered For Now?	21
4.2	Time-series Visualization	22
4.3	Summary From 4.2	22
5	Clustering	25
5.1	k-means Clustering Of Players	25

5.2	Summary From the Clustering	29
5.3	Hierachical Clustering	31
6	Transition of NBA Over Last 20 Years	35
6.1	Versatility & Position-less Basketball	35
6.2	Evolution in Playing Style	35
6.3	Potential Future Trends	36

Chapter 1

NBA Data

The data set contains all the players' performance data from season 1996-97 to season 2022-23.

1.1 Data Type

Here's a brief explanation of some important variable in the analysis:

- **player_height**: player's height in the given season.
- **player_weight**: player's weight in the given season.
- **gp**: total games a player has played in the given season.
- **pts**: player's average points per game.
- **reb**: player's average rebound per game.
- **ast**: player's average assist per game.
- **net_rating**: the team's point differential per 100 possessions while a player is on court.
- **oreb_pct**: offensive rebound percentage - an estimate of the percentage of available offensive rebounds a player grabbed.
- **dreb_pct**: defensive rebound percentage - an estimate of the percentage of available defensive rebounds a player grabbed.
- **usg_pct**: usage percentage - an estimate of the percentage of team plays used by a player.
- **ts_pct**: true shooting percentage - a measure of shooting efficiency that takes into account field goals, 3-point field goals, and free throws.
- **ast_pct**: assist percentage - an estimate of the percentage of teammate field goals a player assisted.
- **season**: season - the NBA season for which these stats apply.

1.2 Preliminary Cleaning

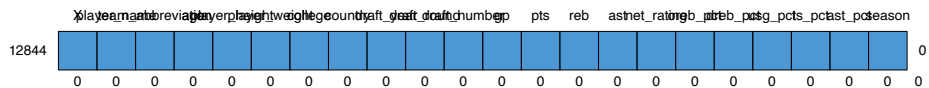
1.2.1 Initial Missing Data Imputation

```
sum(is.na(nba))
```

```
## [1] 0
```

```
md.pattern(nba)
```

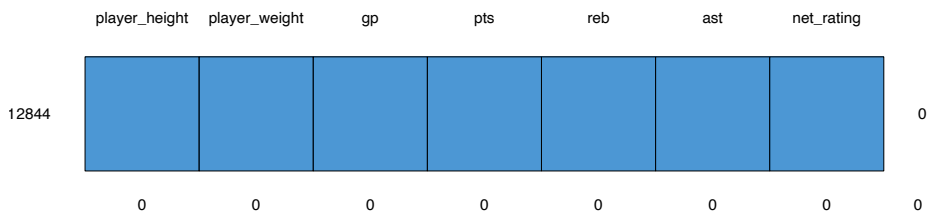
```
## /\      /\
## { `---' }
## { 0    0 }
## ==> V <== No need for mice. This data set is completely observed.
## \  \|/  /
## `-----'
```



```
##      X player_name team_abbreviation age player_height player_weight college
## 12844 1              1                1 1              1              1      1
##      0              0                0 0              0              0      0
##      country draft_year draft_round draft_number gp pts reb ast net_rating
## 12844      1              1                1      1 1 1 1 1 1
##      0              0                0              0 0 0 0 0 0
##      oreb_pct dreb_pct usg_pct ts_pct ast_pct season
## 12844      1              1                1      1 1 1 0
##      0              0                0              0 0 0
```

```
md.pattern(subset(nba, select = c(player_height,
  player_weight, gp, pts, reb, ast, net_rating)))
```

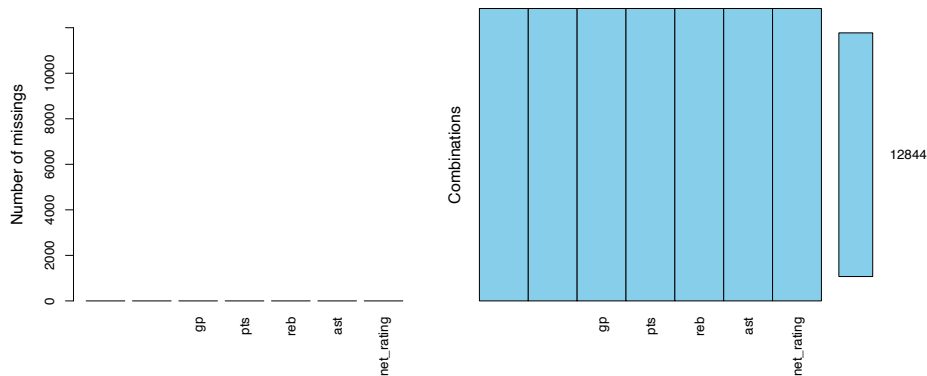
```
## /\      /\
## { `---' }
## { 0    0 }
## ==> V <== No need for mice. This data set is completely observed.
## \  \|/  /
## `-----'
```



```
##      player_height player_weight gp pts reb ast net_rating
```

```
## 12844      1      1 1 1 1 1      1 0
##           0      0 0 0 0 0      0 0
```

```
aggr(subset(nba, select = c(player_height,
  player_weight, gp, pts, reb, ast, net_rating)),
  prop = F, numbers = TRUE)
```



1.2.2 Data Cleaning

identify outlier exists in the net_rating, with the 300 as the max and -250 as the min.

```
summary(nba)
```

```
##      X      player_name      team_abbreviation      age
## Min.   :    0      Length:12844      Length:12844      Min.   :18.00
## 1st Qu.:3211      Class :character      Class :character      1st Qu.:24.00
## Median :6422      Mode  :character      Mode  :character      Median :26.00
## Mean   :6422
## 3rd Qu.:9632
## Max.   :12843
## player_height player_weight      college      country
## Min.   :160.0      Min.   : 60.33      Length:12844      Length:12844
## 1st Qu.:193.0      1st Qu.: 90.72      Class :character      Class :character
## Median :200.7      Median : 99.79      Mode  :character      Mode  :character
## Mean   :200.6      Mean   :100.26
## 3rd Qu.:208.3      3rd Qu.:108.86
## Max.   :231.1      Max.   :163.29
## draft_year      draft_round      draft_number      gp
## Length:12844      Length:12844      Length:12844      Min.   : 1.00
## Class :character      Class :character      Class :character      1st Qu.:31.00
## Mode  :character      Mode  :character      Mode  :character      Median :57.00
##
## Mean   :51.15
## 3rd Qu.:73.00
## Max.   :85.00
```

```
##      pts      reb      ast      net_rating
## Min.   : 0.000   Min.   : 0.000   Min.   : 0.000   Min.   : -250.000
## 1st Qu.: 3.600   1st Qu.: 1.800   1st Qu.: 0.600   1st Qu.:  -6.400
## Median : 6.700   Median : 3.000   Median : 1.200   Median :  -1.300
## Mean   : 8.213   Mean   : 3.558   Mean   : 1.825   Mean   :  -2.226
## 3rd Qu.:11.500   3rd Qu.: 4.700   3rd Qu.: 2.400   3rd Qu.:   3.200
## Max.   :36.100   Max.   :16.300   Max.   :11.700   Max.   : 300.000
##      oreb_pct      dreb_pct      usg_pct      ts_pct
## Min.   :0.00000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.02100   1st Qu.:0.0960   1st Qu.:0.1490   1st Qu.:0.4820
## Median :0.04000   Median :0.1305   Median :0.1810   Median :0.5250
## Mean   :0.05407   Mean   :0.1406   Mean   :0.1846   Mean   :0.5131
## 3rd Qu.:0.08300   3rd Qu.:0.1790   3rd Qu.:0.2170   3rd Qu.:0.5630
## Max.   :1.00000   Max.   :1.0000   Max.   :1.0000   Max.   :1.5000
##      ast_pct      season
## Min.   :0.0000   Length:12844
## 1st Qu.:0.0660   Class :character
## Median :0.1030   Mode  :character
## Mean   :0.1316
## 3rd Qu.:0.1790
## Max.   :1.0000
```

```
nba_clean <- (which(nba$net_rating >= 300 |
  nba$net_rating <= -200))
nba <- nba[-nba_clean] #remove the outliers.
```

1.3 Preliminary Descriptive Statistics

- **age**: the range is between a minimum of 18 years and a maximum of 44 years, with the average age being 26 years old.

```
summary(nba$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.00  24.00   26.00   27.05  30.00   44.00
```

- **height**: the range is between a minimum of 160 cm and a maximum of 231.1 cm, with the average height as 200.7 cm.

```
summary(nba$player_height)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  160.0  193.0   200.7   200.6  208.3   231.1
```

- **weight**: the range is between a minimum of 60.33 kg and a maximum of 163.29 kg, with the average as 100 kg.


```
summary(nba$player_weight)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##  60.33   90.72   99.79  100.26  108.86  163.29
```

- **player:** with a highly competitive threshold, only 2551 players have played in the league since 1996.

```
summary(unique(nba$player_name))
```

```
##      Length      Class      Mode
##      2551 character character
```

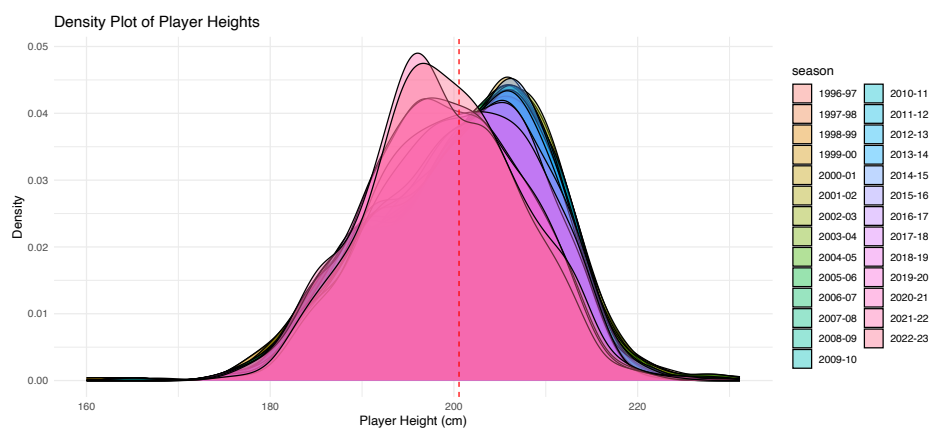

Chapter 2

Hypothesis

2.1 Density Plot for Height and Weight

2.1.1 Density Plot for Height Distribution Over Seasons

```
heightplot <- ggplot(nba, aes(x = player_height)) +  
  geom_density(aes(fill = season), alpha = 0.4) +  
  geom_vline(aes(xintercept = mean(player_height)),  
    linetype = "dashed", color = "red") +  
  ggtitle("Density Plot of Player Heights") +  
  xlab("Player Height (cm)") + ylab("Density") +  
  theme_minimal()  
print(heightplot)
```

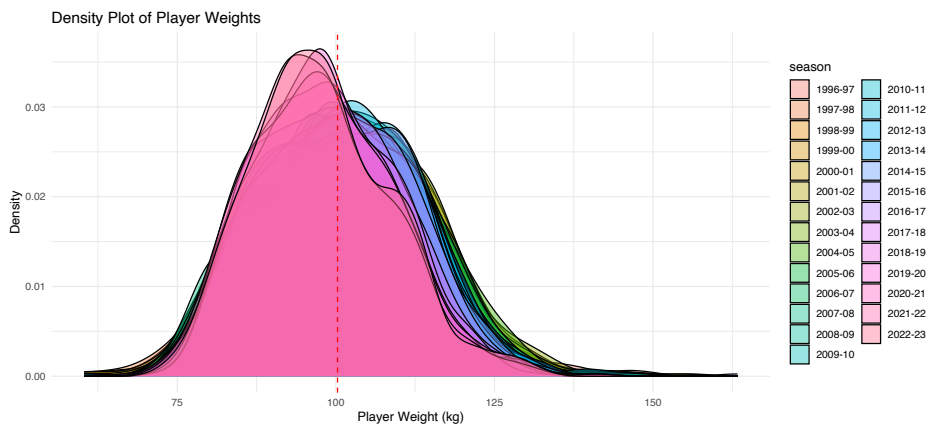


2.1.1.1 Summary

- A more noticeable shift in the mean height of players over seasons. The mean height has decreased steadily.
- A narrower range of heights among players in the recent NBA compared to the past.
- There's a noticeable peak around the 200-210 cm range, suggesting that a significant number of players fall within this height bracket.

2.1.2 Density Plot for Weight Distribution Over Seasons

```
weightplot <- ggplot(nba, aes(x = player_weight)) +
  geom_density(aes(fill = season), alpha = 0.4) +
  geom_vline(aes(xintercept = mean(player_weight)),
    linetype = "dashed", color = "red") +
  ggtitle("Density Plot of Player Weights") +
  xlab("Player Weight (kg)") + ylab("Density") +
  theme_minimal()
print(weightplot)
```



2.1.2.1 Summary

- The mean weight seems to have shifted towards the left, suggesting that players have, on average, become lighter over the years.
- The distributions for the recent seasons appear narrower, indicating less variability in player weights now than in earlier seasons.

2.1.3 General Conclusion

In summary, over the years, NBA players have, on average, become shorter and lighter, with a narrower range of both weights and heights represented in recent

seasons.

2.2 Inference & Hypothesis Based On The Density Plot and Initial Summary

- **H1** Versatility & Position-less Basketball: the narrower range of heights and weights indicates that there might be an increasing trend of “position-less” basketball, specifically players are no longer strictly confined to traditional roles based on their physical attributes.
- **H2** Evolution in Playing Style: the traditional center-focused style of play cannot adapt to the pace of the modern NBA.

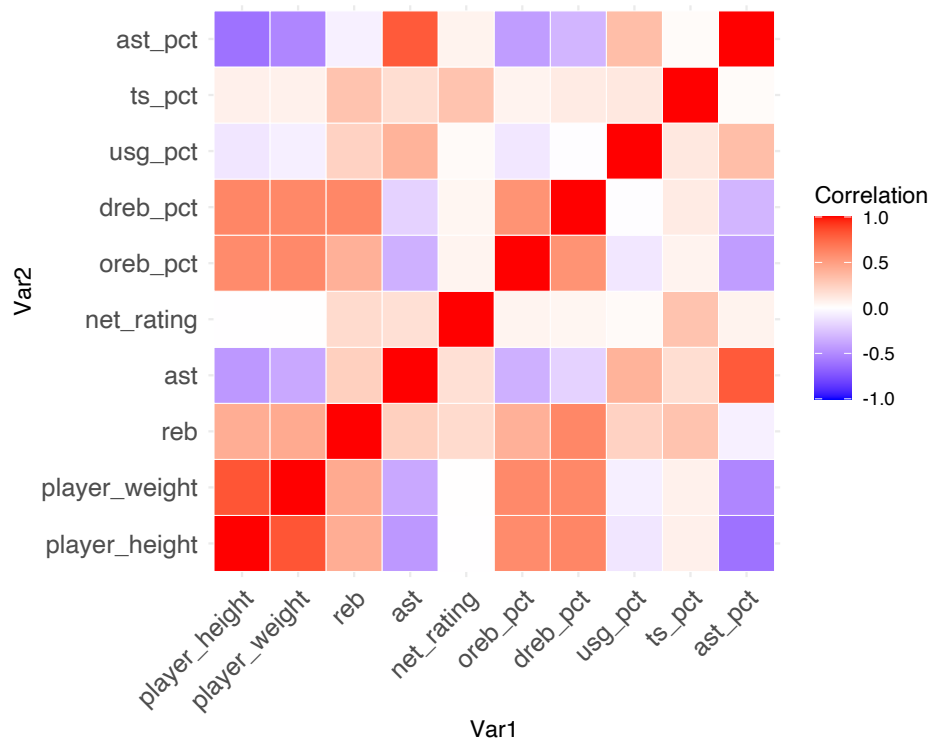
Chapter 3

Correlation Analysis

3.1 Correlation Between H/W and Others

- The correlation matrix visually represents the relationship between height/weight and other basketball metrics.
- Correlations range between -1.0 and 1.0, with darker red indicating a strong positive correlation, blue indicating a strong negative correlation, and lighter shades (towards white) indicating weaker or no correlation.

```
ggplot(data = melted_cormat, aes(x = Var1,  
  y = Var2)) + geom_tile(aes(fill = value),  
  color = "white") + scale_fill_gradient2(low = "blue",  
  high = "red", mid = "white", midpoint = 0,  
  limit = c(-1, 1), space = "Lab", name = "Correlation") +  
  theme_minimal() + theme(axis.text.x = element_text(angle = 45,  
  vjust = 1, size = 12, hjust = 1), axis.text.y = element_text(size = 12)) +  
  coord_fixed()
```



3.2 Correlation Matrix

- A strong positive correlation between height and defensive rebound percentage as well as offensive rebound percentage.
- Weight also shows positive correlations with both rebound percentage, though slightly weaker compared to height.
- Height seems to have a negative correlation with assist percentage, suggesting taller players might assist less.
- Other metrics like true shooting percentage and usage percentage have weaker correlations with height and weight.

3.3 Correlation Between H/W and R/A

Four scatter plots depict the relationships between player height, player weight, rebounds, and assists.

```
height_reb <- ggplot(nba, aes(x = player_height,
                             y = reb))
height_reb <- height_reb + geom_point()

height_ast <- ggplot(nba, aes(x = player_height,
```



```

    y = ast))
height_ast <- height_ast + geom_point()

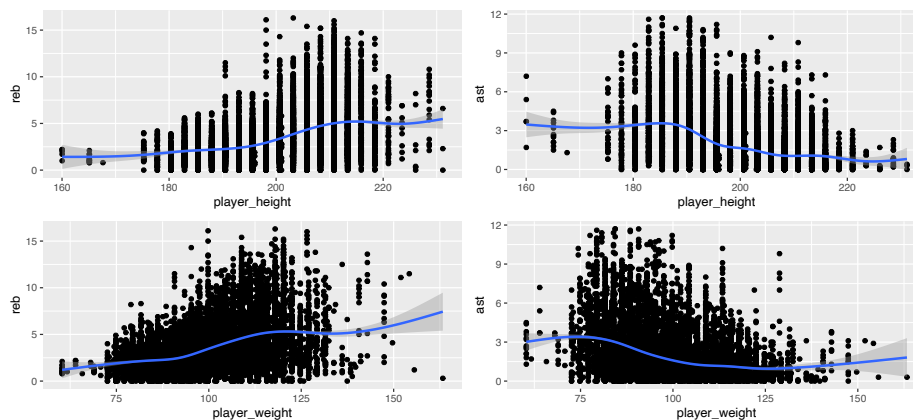
weight_reb <- ggplot(nba, aes(x = player_weight,
    y = reb))
weight_reb <- weight_reb + geom_point()

weight_ast <- ggplot(nba, aes(x = player_weight,
    y = ast))
weight_ast <- weight_ast + geom_point()

grid.arrange(height_reb + geom_smooth(),
    height_ast + geom_smooth(), weight_reb +
    geom_smooth(), weight_ast + geom_smooth(),
    ncol = 2)

## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'

```



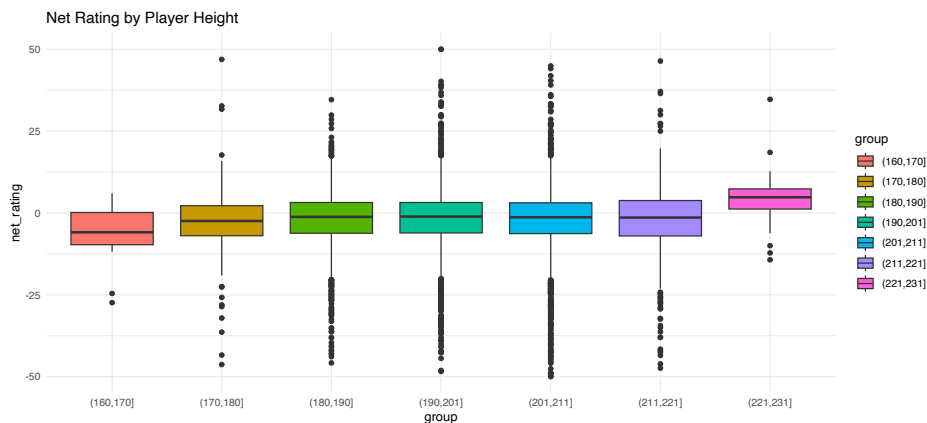
3.4 Summary From Scatter Plots

- Consistent with the correlation matrix, height and rebound displays an upward trend, indicating that taller players tend to have more rebounds.
- Conversely, height and assist depicts a downward trend, signifying that shorter players usually have more assists.

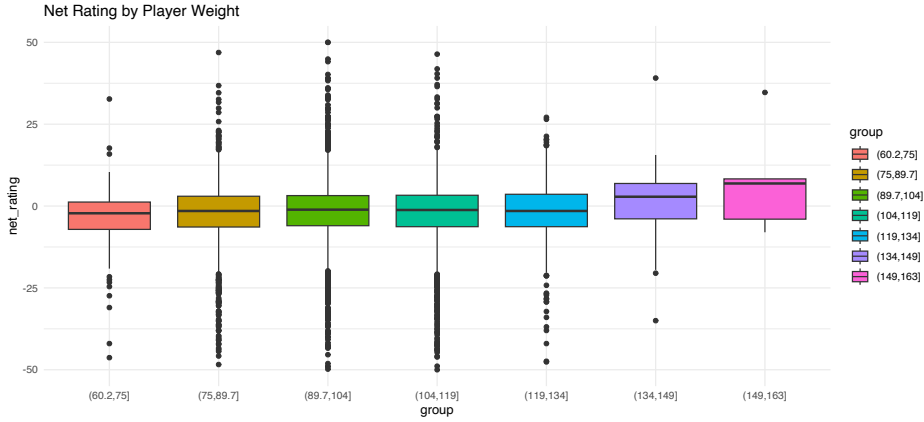
3.5 Net Rating By Player's Physical Attributes

The boxplots depict the distribution of net rating (i.e., the team's point differential per 100 possessions while a player is on court) across various player heights and weights.

```
heightdatapart <- data.frame(player_height = nba$player_height,
  net_rating = nba$net_rating)
heightdatapart <- heightdatapart[(heightdatapart$net_rating >=
  -50 & heightdatapart$net_rating <= 50),
]
heightdatapart$group <- cut(heightdatapart$player_height,
  breaks = 7)
ph <- ggplot(data = heightdatapart, aes(x = group,
  y = net_rating, fill = group)) + geom_boxplot() +
  ggtitle("Net Rating by Player Height") +
  theme_minimal()
print(ph)
```



```
weightdatapart <- data.frame(player_weight = nba$player_weight,
  net_rating = nba$net_rating)
weightdatapart <- weightdatapart[(weightdatapart$net_rating >=
  -50 & weightdatapart$net_rating <= 50),
]
weightdatapart$group <- cut(weightdatapart$player_weight,
  breaks = 7)
pw <- ggplot(data = weightdatapart, aes(x = group,
  y = net_rating, fill = group)) + geom_boxplot() +
  ggtitle("Net Rating by Player Weight") +
  theme_minimal()
print(pw)
```



3.6 Summary From Box Plot

3.6.1 Net Rating by Player Height

- Players in the height range of 190 cm to 203 cm tend to have a wider variation in net rating, with some players in this range even having exceptionally high net ratings.
- Taller players, especially those between 206 cm and 211 cm, have a slightly lower median net rating.
- The shorter players, particularly those below 188 cm, also show a lower median net rating, but with less variability than the taller players.
- Players who are 213 cm and above exhibit a narrower interquartile range (middle 50% of data) but with a few outliers having lower net ratings.

3.6.2 Net Rating by Player Weight

- Players weighing between 70-85 kg and 100-115 kg display a wider range in net rating, with some outliers, particularly on the higher end.
- Players in the weight range of 85-100 kg have a median net rating that's slightly higher, with a more compact interquartile range, indicating less variability in their performance.
- Heavier players, weighing 119 kg and above, exhibit a narrower interquartile range with a few lower outliers. From the boxplots, it can be concluded that:

3.6.3 General Summary

- Players of medium height (190 cm to 203 cm) tend to have varied performances, with some players being exceptionally good based on net rating.
- The shortest and the tallest players tend to have a slightly lower median net rating.

- Players of medium weight (85-100 kg) generally have consistent performances, while players on the lighter and heavier ends show greater variability in net rating.

Chapter 4

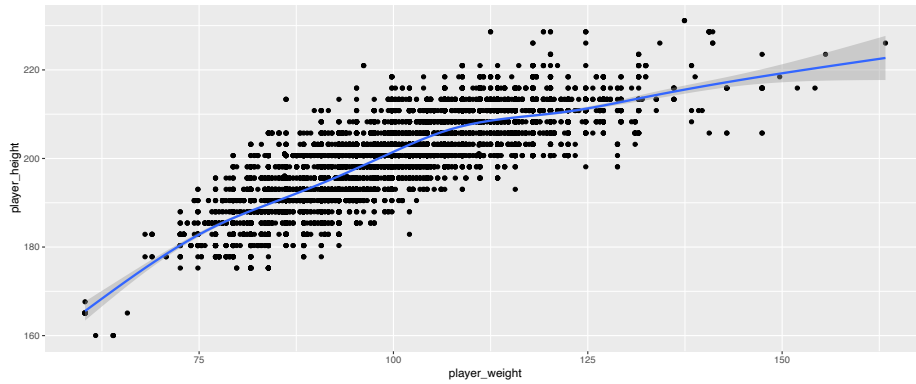
In-depth Analysis for H/W

4.1 What Can Be Answered For Now?

- H1: Versatility & Position-less Basketball Given the narrow range of heights and weights and the varied performance of players irrespective of their physical attributes, it does hint towards the emergence of “position-less” basketball. Players are breaking traditional molds, and their skills are no longer just confined to their height and weight.
- H2: Evolution in Playing Style The NBA’s pace and style have evolved, favoring versatile players over traditional roles. The varied performances of players, especially the medium-height ones, and the shift from a strong center-focused game further supports this hypothesis.

```
p1 <- ggplot(nba, aes(x = player_weight,  
  y = player_height))  
p1 <- p1 + geom_point()  
p1 + geom_smooth()
```

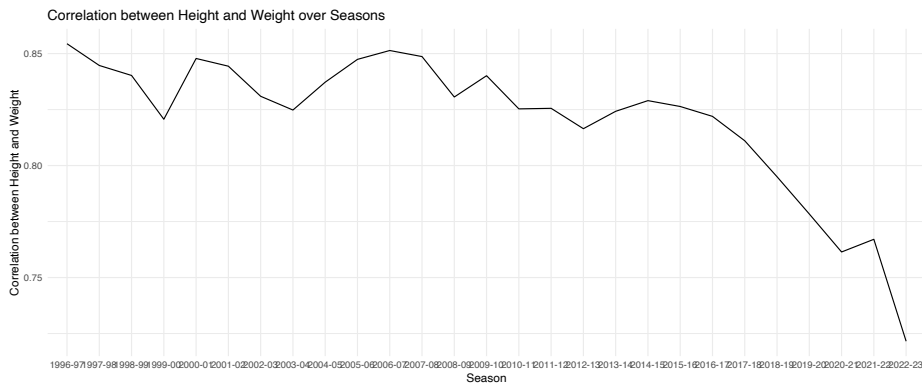
```
## `geom_smooth()` using method = 'gam' and formula = 'y ~ s(x, bs = "cs")'
```



4.2 Time-series Visualization

To further explore H1 and H2, a time-series analysis portraying the fluctuating correlation between height and weight over seasons has been drawn.

```
p <- ggplot(cor_df, aes(x = season,
  y = correlation,
  group = 1)) + geom_line() +
  xlab("Season") +
  ylab("Correlation between Height and Weight") +
  ggtitle("Correlation between Height and Weight over Seasons") +
  theme_minimal()
print(p)
```



4.3 Summary From 4.2

- A significant higher correlation between height and weight can be observed between season 1996-97 and 2009-10, suggesting a stronger relationship between a player's height and weight. This could have been due to the

prevalence of traditional roles and the center-focused game styles, if we recall who were the most dominant players at that time (e.g., Shaquille O’Neal, Hakeem Olajuwon, Tim Duncan, Karl Malone, Dirk Nowitzki, Kevin Garnett)

- Since the season 2009-10, the correlation dips, suggesting a weaker relationship between height and weight. This potentially supports the hypothesis of “position-less” basketball becoming more prevalent, as players’ roles become less defined by their physical attributes.

Chapter 5

Clustering

5.1 k-means Clustering Of Players

To further investigate the transition in playing style and the trend toward position-less basketball, clustering analysis were conducted. This analysis used selected features including player height, weight, rebounds, assists, net rating, offensive rebounds, defensive rebounds, usage percentage, true shooting percentage, and assist percentage.

What Data Was Included

Given that more than 200 players participate in each NBA season and audiences generally remember high-scoring players, I applied specific criteria for more clear and understandable visualization results. For the selected seasons, only players who participated in more than 50 games and had an average score exceeding 10 points were included in the analysis.

Selected Season

- Based on the visualization in Section 4.2, a stronger correlation between height and weight (H/W) is evident between the seasons 1996-97 and 2009-10. This suggests a traditional, center-focused playing style, with a peak occurring in the 2000-01 season. To examine how player performance styles were clustered during periods when traditional styles dominated, we selected the 2000-01 and 2009-10 seasons.
- NBA fans commonly believe that the “small-ball center” and “three-point trend” popularized by Stephen Curry’s Golden State Warriors from 2015 to 2019 were catalysts for the transition in playing styles. To explore how player performance styles have evolved in more recent years, I selected the 2015-16 season and the most current season, 2022-2023.

```

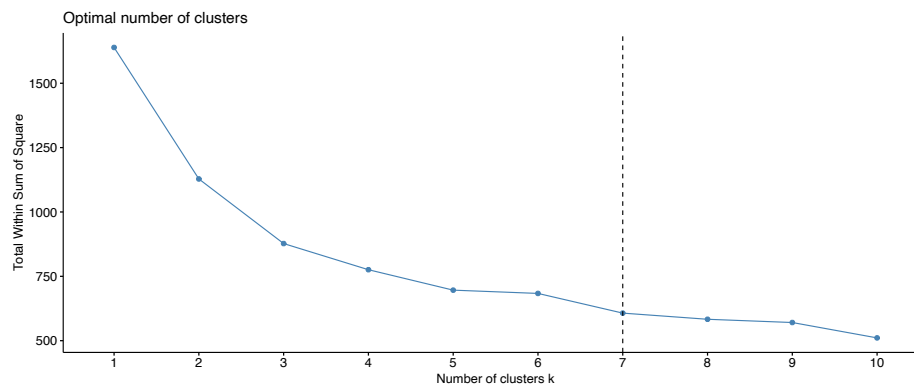
nba_selected_season <- nba[(which(nba$season ==
  "2022-23")), ]
nba_selected_season <- nba_selected_season[nba_selected_season$pts >
  10, ]
nba_selected_season <- nba_selected_season[nba_selected_season$gp >
  50, ]
rownames(nba_selected_season) <- nba_selected_season$player_name

selected_features <- c("player_height",
  "pts", "player_weight",
  "reb", "ast", "net_rating",
  "oreb_pct", "dreb_pct",
  "usg_pct", "ts_pct",
  "ast_pct")
nba_for_clustering <- nba_selected_season %>%
  select(all_of(selected_features))

df <- as.data.frame(scale(nba_for_clustering))

fviz_nbclust(df, kmeans,
  method = "wss") +
  geom_vline(xintercept = 7,
    linetype = 2)

```



```

set.seed(123)
km_result <- kmeans(df,
  centers = 7)

clustering_23 <- fviz_cluster(km_result,
  data = df, ellipse.type = "euclid",
  ellipse.level = 0.5,
  ellipse.ratio = 0.8,

```

```

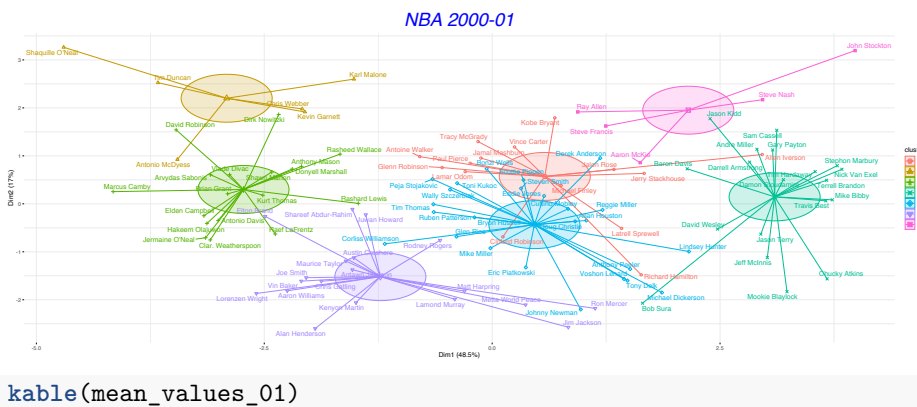
star.plot = TRUE,
repel = TRUE, main = "NBA 2022-2023",
ggtheme = theme_minimal())

clustering_23 <- clustering_23 +
  theme(plot.title = element_text(size = 30,
    face = "italic",
    color = "blue",
    hjust = 0.5,
    vjust = 1, angle = 0,
    lineheight = 1.2))

```

For season 2000-01

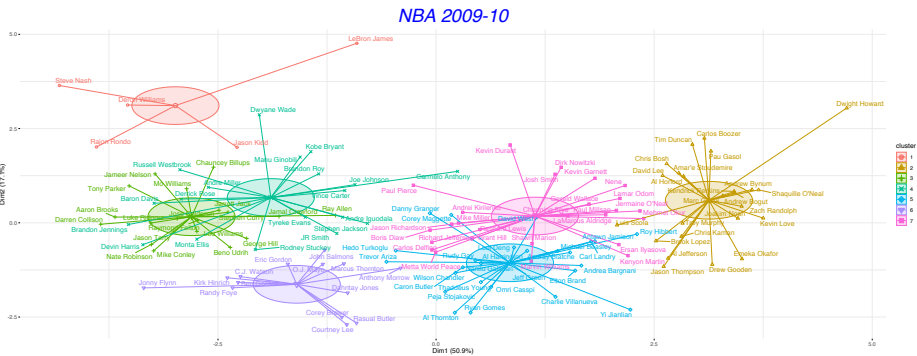
```
print(clustering_01)
```



cluster	player_height	player_weight	pts	reb	ast
1	200.3213	97.58276	23.06667	5.740000	4.360000
2	209.9733	116.49755	24.00000	11.300000	3.750000
3	209.4089	110.37405	14.03889	8.505556	1.650000
4	186.7568	84.94107	15.83158	3.736842	6.742105
5	198.7296	96.77839	13.65200	3.984000	2.472000
6	204.6705	107.38194	14.04211	6.105263	1.921053
7	191.5160	88.63188	16.12000	4.440000	6.420000

For season 2009-10

```
print(clustering_10)
```

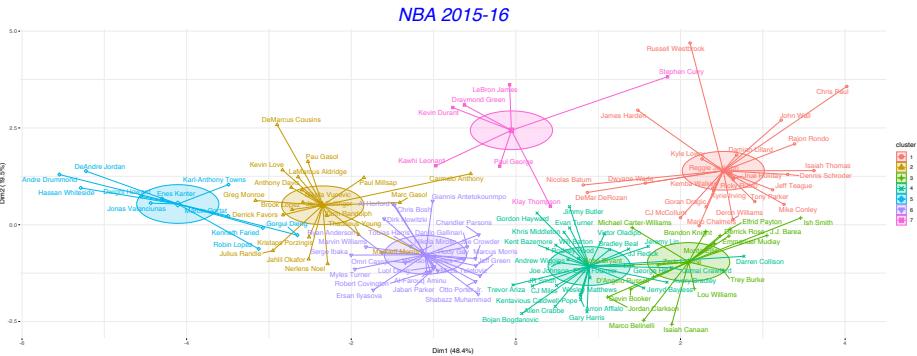


```
kable(mean_values_10)
```

cluster	player_height	player_weight	pts	reb	ast
1	192.5320	92.16989	17.78000	4.920000	9.800000
2	210.0792	117.59373	16.15000	9.754167	1.991667
3	187.4253	84.94107	14.20526	2.705263	4.673684
4	195.1789	93.79805	19.55789	4.378947	5.252632
5	206.3262	106.50689	15.16923	5.715385	1.834615
6	193.9471	91.36639	13.10714	2.964286	2.650000
7	205.5446	107.67576	14.13462	6.519231	2.338462

For season 2015-16

```
print(clustering_16)
```

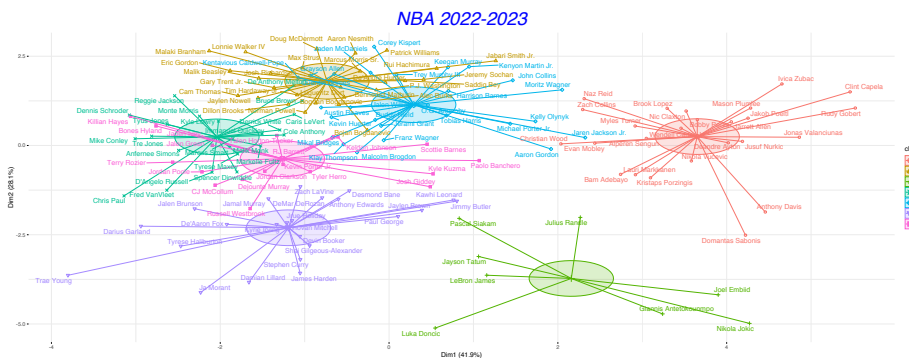


```
kable(mean_values_16)
```

cluster	player_height	player_weight	pts	reb	ast
1	189.7592	88.50714	17.87917	4.016667	6.491667
2	209.5500	114.50930	16.73000	8.585000	2.370000
3	191.5459	87.30312	13.50000	3.147059	3.817647
4	196.7593	94.16894	14.13929	3.557143	2.671429
5	211.0509	114.71754	13.36364	10.263636	1.145455
6	206.3262	104.76231	13.67692	5.723077	1.788461
7	201.0229	102.05820	23.42857	6.871429	4.957143

For season 2022-23

```
print(clustering_23)
```



```
kable(mean_values_23)
```

cluster	player_height	player_weight	pts	reb	ast
1	211.3280	113.25285	15.90000	9.196000	2.352000
2	197.4615	97.10229	13.09259	3.555556	1.855556
3	206.6925	112.09392	28.67500	9.662500	6.125000
4	189.2300	88.18241	13.73182	3.427273	5.009091
5	201.2462	98.91795	14.81154	4.538462	2.276923
6	193.2609	92.27639	24.90000	4.917391	6.239130
7	195.5800	92.41340	18.14737	4.826316	4.700000

5.2 Summary From the Clustering

H1 Versatility & Position-less Basketball

- 2000-01 Season: players are more dispersed, suggesting a distinct division between positions based on physical attributes like height and weight. For example:
 - Traditional Centers: players around 7 feet tall and weighing between 250-300 lbs are prominent. These athletes typically represent traditional centers prevalent during this era (e.g., Tim Duncan, Karl Malone, Kevin Garnett, Dirk Nowitzki, Chris Webber, Shaquille O'Neal).

- Guards: athletes with heights ranging from 6’3” to 6’6” and weights between 180-210 lbs form distinct clusters, representing the roles of shooting and point guards.
- 2009-10 Season: despite the strong correlation between height and weight suggested by Visualization 4.2, players like LeBron James and Dwyane Wade begin to lead the trend toward a more versatile playing style. LeBron James, for instance, while physically dominant enough to be considered a small forward, also demonstrates the skills and agility typical of guards. He can be grouped with well-known point guards like Steve Nash, Rajon Rondo, and Jason Kidd. On the other hand, players like Russell Westbrook, originally considered point guards, start taking on more scoring roles, leading them to be identified with scorers like Kobe Bryant and Carmelo Anthony. The diverse physical attributes within this scoring cluster make the transitional trend more apparent.
- 2015-16 Season: clusters start to merge, indicating the onset of the “position-less” trend. Players of various physical attributes now perform roles previously confined to specific positions. Draymond Green (6’6”, 230 lbs), for example, plays effectively as a point guard, power forward, and even center despite having the height of a traditional small forward. Additionally, dominant forwards who aren’t traditional big men or guards but possess qualities of both are increasingly playing central roles (e.g., LeBron James, Kevin Durant, Kawhi Leonard).
- 2022-23 Season: clusters continue to converge, showcasing increasing versatility among players. Athletes with a narrow height range of 6’5” to 6’10” and weight range of 210-250 lbs are now playing multiple roles on the court—whether it’s shooting, ball-handling, or post plays. Players like Luka Dončić, Giannis Antetokounmpo, LeBron James, and Nikola Jokić can be grouped in the same cluster, despite their varying physical sizes.

H2 Evolution in Playing Style

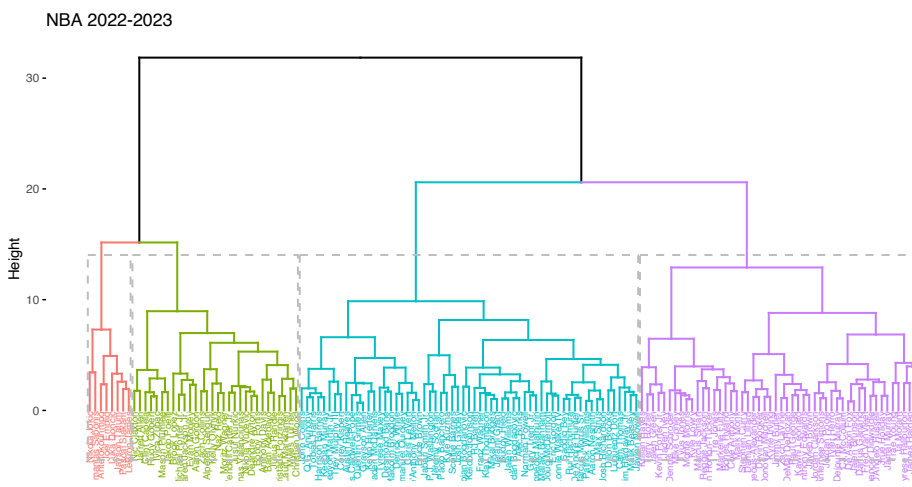
- 2000-01 and 2009-10 Seasons: the dominance of players like Shaquille O’Neal and Tim Duncan indicates a game focused on the post and interior play, supporting the traditional center-focused style. At the same time, players like Kobe Bryant and Vince Carter, typically shooting guards, with a height-weight specification of around 6’6” and 212 lbs, show the traditional divisions in playing roles.
- 2015-16 Season: with players like Stephen Curry and LeBron James leading the clusters from 2010 to 2020, it’s evident that the style has evolved to prioritize agility, shooting range, and versatility. The game sees a rise in sharpshooters and versatile players. Stephen Curry (6’2”, 185 lbs) stands out with his unparalleled three-point shooting prowess.
 - LeBron James (6’9”, 250 lbs) is a classic example of versatility. Though traditionally a forward, players like him and Draymond Green can handle the ball and pass, and effectively play the defense role as either a forward, guard, or center, embodying the essence of position-less basketball.

- 2022-23 Season: The dominance of players like Giannis Antetokounmpo and Nikola Jokić who is traditionally regarded as big player, showcasing the prominent role in ball handling, pass, and play-making like a point guard. Meanwhile, player like Luka Dončić, standing at 6'7" and weighing around 230 lbs, is a prime example of the modern player with a forward style medium-height but can lead the role of play-making and also play inside well.

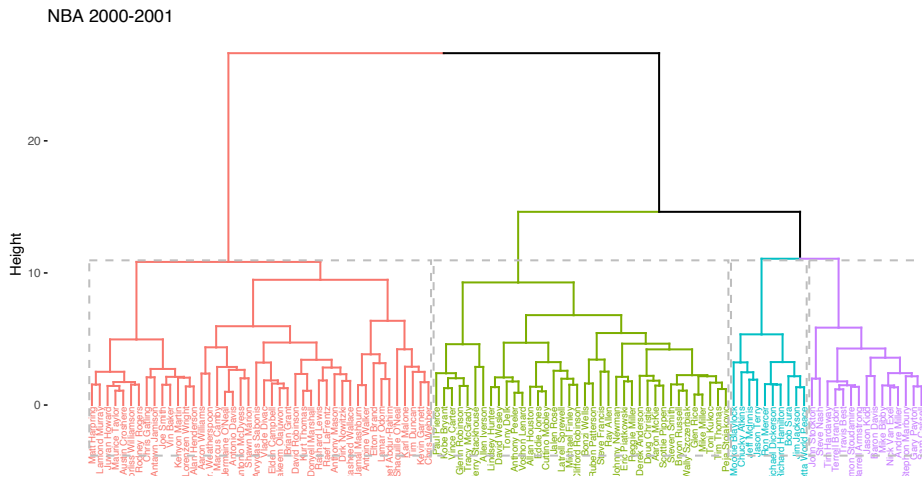
5.3 Hierarchical Clustering

To present a more intuitive depiction of the nested groupings in player clustering results, and to illustrate the gradual changes and transitions between seasons, we conducted hierarchical clustering visualization. The results were consistent with previous findings.

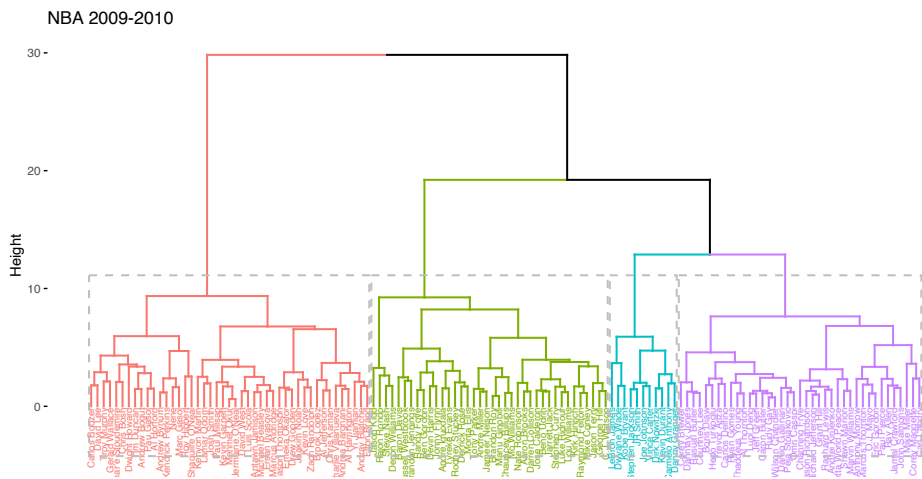
```
result_23 <- dist(df,
  method = "euclidean")
result_hc <- hclust(d = result_23,
  method = "ward.D2")
fviz_dend(result_hc,
  k = 4, cex = 0.5,
  color_labels_by_k = TRUE,
  main = "NBA 2022-2023",
  rect = TRUE)
```



```
fviz_dend(result_hc_01,
  k = 4, cex = 0.5,
  color_labels_by_k = TRUE,
  main = "NBA 2000-2001",
  rect = TRUE)
```

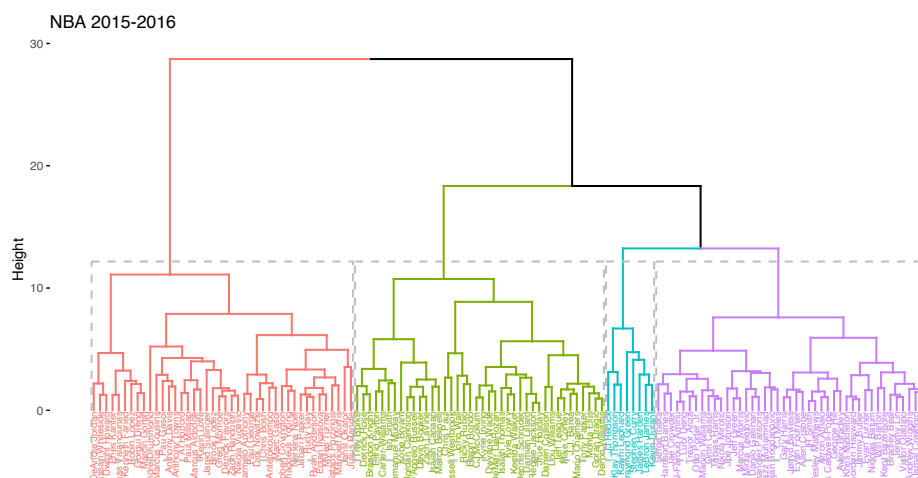


```
fviz_dend(result_hc_10,
  k = 4, cex = 0.5,
  color_labels_by_k = TRUE,
  main = "NBA 2009-2010",
  rect = TRUE)
```



```
fviz_dend(result_hc_16,
  k = 4, cex = 0.5,
  color_labels_by_k = TRUE,
  main = "NBA 2015-2016",
  rect = TRUE)
```


33



Chapter 6

Transition of NBA Over Last 20 Years

Over the past two decades, the NBA has undergone significant transformations, particularly in playing styles and player roles.

6.1 Versatility & Position-less Basketball

The first ten year of this centry, the league showcased traditional player roles based on physical attributes. Centers like Shaquille O’Neal exemplified this trend with their dominance in the post. However, with the progression second ten year, there’s a marked shift towards position-less basketball. Players like Stephen Curry, LeBron James, Giannis Antetokounmpo, Nikola Jokić exemplify this evolution. Initially known for his scoring, LeBron’s game matured to include playmaking, passing, and even post play, adapting to the requirements of his team and the evolving nature of the NBA. This has paved the way for players who, regardless of their size, can perform multiple roles on the court.

6.2 Evolution in Playing Style

The NBA has transitioned from a center-focused play to a more perimeter-oriented game. This shift is evident when comparing the dominance of traditional centers in earlier seasons to the sharpshooting prowess of guards like Stephen Curry in later ones. The league has gravitated towards faster-paced, space-oriented offenses, capitalizing on three-point shots and versatile player skills.

6.3 Potential Future Trends

The increasing value of versatility suggests the NBA will continue to evolve towards an even faster, more skill-oriented game. Traditional roles might further blur, leading to players being valued more for their overall skill set rather than specific positional attributes.

As for physical attributes, while height will always be advantageous, it might not be the dominant factor in player evaluation. Agility, endurance, and skill versatility could become primary attributes scouts look for. For instance, players who, like LeBron, can adapt and reinvent their game will be highly prized. As teams prioritize speed and versatility, we may see an even more diverse range of player heights and weights, with an emphasis on adaptability both in play style and physical conditioning.