Final Exam: Porfolio

December 7, 2023

0.1 The Question

0.1.1 Do the biggest players dominate the sport of Basketball?

The answer might sound obvious, but trust me, it isn't. Most people who are not familiar with the sport of basketball will tell you that height and weight are deciding factors in a player. A taller player might dominate the sport, but that is only true in amateur leagues.

In the NBA, players are performing at near-perfect levels of play. There has been a wide range of all-stars that are relatively short for the NBA to be dominant in the sport (for example, Stephen Curry) because the NBA is more concerned with the philosophy of the game. The game is constantly changing, and it can sometimes be the case that a player who is 6'2'' is preferred over a player who is 6'7''. In this data analysis, I will look at all the players in the three seasons and compare how they play based on their biometrics. The three main factors will be looking at a player's height and weight compared to the net rating. Net rating is the offensive rating minus the defensive rating. Net rating tells us how much better or worse the team is when a specific player is on the court. I will also try to separate players to make the data easy to read.

The three seasons I will look at are 2002-03,2016-17, and 2022-23.

The early 2000s was considered a time when the taller and heavier players dominated the sport. 2016-17 was when the game became more about making three-pointers rather than playing to get close to the basket to shoot. I also wanted to look at 2022-23 to see how much the taller and heavier players dominate the sport now. By looking at these 20 years of data, we can see how the sport has changed, and we can also see if the taller and heavier players are dominating the sport.

0.2 Methodology

By looking at 20 years of data, we will see the different eras of the NBA. The current era is called the modern era and is sometimes called the "new school" by fans, and the era before the modern era is called the transitional era. The transitional era is known as the "old school" era. The modern era is when the game became more about creating space to take three-pointers, while the traditional era was more about getting close to the basket to score.

There is a debate in the NBA community about where the modern era started. Most fans will say the 2014-2015 season is when teams began adopting the new style of the game. It wouldn't be until the 2016 season for all the teams to adopt the new style. We will look at the 2016 season specifically to compare and contrast the shorter and taller players.

We also want to only look at dominant players. For this analysis, a player with a net rating above +10 will be considered a dominating player. Net rating is usually in the range of +15 to -15. A +15 player is usually the MVP. A -15 player should not be allowed to play in the league.

Another thing to consider is who a big player would be and who a small player would be. Most people consider a player that is less than 6'6'' to be short and 7' to be considered tall for the NBA. We will look at the data to determine if this is true.

[2]:
```
#here i am just importing libraries that i will use
import statistics
import numpy as np
from datascience import *

import matplotlib.pyplot as plt
plt.style.use("ggplot")
%matplotlib inline
```

0.2.1 Getting the data

The data was downloaded from kaggle and then uploaded to the server

Dataset: **(https://www.kaggle.com/datasets/justinas/nba-players-data/data)**

[3]:
```
#reading in the data from the dataset
data = Table.read_table("all_seasons.csv")
data
```

[3]:
```
Unnamed: 0 | player_name      | team_abbreviation | age  | player_height |
player_weight | college              | country | draft_year | draft_round |
draft_number | gp   | pts  | reb  | ast  | net_rating | oreb_pct | dreb_pct |
usg_pct | ts_pct | ast_pct | season
0          | Randy Livingston | HOU               | 22   | 193.04        |
94.8007       | Louisiana State      | USA     | 1996       | 2           | 42
| 64   | 3.9  | 1.5  | 2.4  | 0.3        | 0.042    | 0.071    | 0.169   | 0.487
| 0.248   | 1996-97
1          | Gaylon Nickerson | WAS               | 28   | 190.5         |
86.1825       | Northwestern Oklahoma | USA     | 1994       | 2           | 34
| 4    | 3.8  | 1.3  | 0.3  | 8.9        | 0.03     | 0.111    | 0.174   | 0.497
| 0.043   | 1996-97
2          | George Lynch     | VAN               | 26   | 203.2         |
103.419       | North Carolina       | USA     | 1993       | 1           | 12
| 41   | 8.3  | 6.4  | 1.9  | -8.2       | 0.106    | 0.185    | 0.175   | 0.512
| 0.125   | 1996-97
3          | George McCloud   | LAL               | 30   | 203.2         |
102.058       | Florida State        | USA     | 1989       | 1           | 7
| 64   | 10.2 | 2.8  | 1.7  | -2.7       | 0.027    | 0.111    | 0.206   | 0.527
| 0.125   | 1996-97
4          | George Zidek     | DEN               | 23   | 213.36        |
119.748       | UCLA                 | USA     | 1995       | 1           | 22
| 52   | 2.8  | 1.7  | 0.3  | -14.1      | 0.102    | 0.169    | 0.195   | 0.5
| 0.064   | 1996-97
5          | Gerald Wilkins   | ORL               | 33   | 198.12        |
```

```
102.058       | Tennessee-Chattanooga | USA     | 1985       | 2           | 47
| 80   | 10.6 | 2.2  | 2.2  | -5.8       | 0.031    | 0.064    | 0.203   | 0.503
| 0.143   | 1996-97
6          | Gheorghe Muresan | WAS               | 26   | 231.14        |
137.438       | nan                   | USA     | 1993       | 2           | 30
| 73   | 10.6 | 6.6  | 0.4  | 6.9        | 0.098    | 0.217    | 0.185   | 0.618
| 0.024   | 1996-97
7          | Glen Rice        | CHH               | 30   | 203.2         |
99.7902       | Michigan              | USA     | 1989       | 1           | 4
| 79   | 26.8 | 4    | 2    | 3.2        | 0.025    | 0.087    | 0.272   | 0.605
| 0.088   | 1996-97
8          | Glenn Robinson   | MIL               | 24   | 200.66        |
106.594       | Purdue                | USA     | 1994       | 1           | 1
| 80   | 21.1 | 6.3  | 3.1  | -2.9       | 0.051    | 0.144    | 0.278   | 0.528
| 0.146   | 1996-97
9          | Grant Hill       | DET               | 24   | 203.2         |
102.058       | Duke                  | USA     | 1994       | 1           | 3
| 80   | 21.4 | 9    | 7.3  | 6.9        | 0.049    | 0.232    | 0.283   | 0.556
| 0.356   | 1996-97
… (12834 rows omitted)
```

0.2.2 Exploring the data

Now that I have the data, I can explore and manipulate the data so that the tables represent information that will help answer my questions. I want to know who the heaviest and tallest players are and also want to know what their impact they have. I want to know the outliers of the NBA for the seasons.

[3]:
```
#removing the first column cause it doens't tells us anything
data = data.drop(0)
data
```

[3]:
```
player_name      | team_abbreviation | age  | player_height | player_weight |
college              | country | draft_year | draft_round | draft_number | gp
| pts  | reb  | ast  | net_rating | oreb_pct | dreb_pct | usg_pct | ts_pct |
ast_pct | season
Randy Livingston | HOU               | 22   | 193.04        | 94.8007       |
Louisiana State       | USA     | 1996       | 2           | 42          | 64
| 3.9  | 1.5  | 2.4  | 0.3        | 0.042    | 0.071    | 0.169   | 0.487  |
0.248   | 1996-97
Gaylon Nickerson | WAS               | 28   | 190.5         | 86.1825       |
Northwestern Oklahoma | USA     | 1994       | 2           | 34          | 4
| 3.8  | 1.3  | 0.3  | 8.9        | 0.03     | 0.111    | 0.174   | 0.497  |
0.043   | 1996-97
George Lynch     | VAN               | 26   | 203.2         | 103.419       |
North Carolina        | USA     | 1993       | 1           | 12          | 41
| 8.3  | 6.4  | 1.9  | -8.2       | 0.106    | 0.185    | 0.175   | 0.512  |
```

```
0.125   | 1996-97
George McCloud   | LAL               | 30   | 203.2         | 102.058       |
Florida State         | USA     | 1989       | 1           | 7           | 64
| 10.2 | 2.8  | 1.7  | -2.7       | 0.027    | 0.111    | 0.206   | 0.527  |
0.125   | 1996-97
George Zidek     | DEN               | 23   | 213.36        | 119.748       |
UCLA                  | USA     | 1995       | 1           | 22          | 52
| 2.8  | 1.7  | 0.3  | -14.1      | 0.102    | 0.169    | 0.195   | 0.5    |
0.064   | 1996-97
Gerald Wilkins   | ORL               | 33   | 198.12        | 102.058       |
Tennessee-Chattanooga | USA     | 1985       | 2           | 47          | 80
| 10.6 | 2.2  | 2.2  | -5.8       | 0.031    | 0.064    | 0.203   | 0.503  |
0.143   | 1996-97
Gheorghe Muresan | WAS               | 26   | 231.14        | 137.438       |
nan                   | USA     | 1993       | 2           | 30          | 73
| 10.6 | 6.6  | 0.4  | 6.9        | 0.098    | 0.217    | 0.185   | 0.618  |
0.024   | 1996-97
Glen Rice        | CHH               | 30   | 203.2         | 99.7902       |
Michigan              | USA     | 1989       | 1           | 4           | 79
| 26.8 | 4    | 2    | 3.2        | 0.025    | 0.087    | 0.272   | 0.605  |
0.088   | 1996-97
Glenn Robinson   | MIL               | 24   | 200.66        | 106.594       |
Purdue                | USA     | 1994       | 1           | 1           | 80
| 21.1 | 6.3  | 3.1  | -2.9       | 0.051    | 0.144    | 0.278   | 0.528  |
0.146   | 1996-97
Grant Hill       | DET               | 24   | 203.2         | 102.058       |
Duke                  | USA     | 1994       | 1           | 3           | 80
| 21.4 | 9    | 7.3  | 6.9        | 0.049    | 0.232    | 0.283   | 0.556  |
0.356   | 1996-97
… (12834 rows omitted)
```

[4]:
```
#I need to get 3 different tables for each season that i want to compare.
#2022 because it was the most recent and some big players dominated
#2016 because it was when most of the bigger players didn't do much
#2002 because it was when the biggest players were the most wanted
season_2022 = data.where("season", "2021-22").
 ↪drop(1,2,5,6,7,8,9,11,13,12,15,16,17, 18, 19)
season_2016 = data.where("season", "2016-17").
 ↪drop(1,2,5,6,7,8,9,11,13,12,15,16,17, 18, 19)
season_2002 = data.where("season", "2002-03").
 ↪drop(1,2,5,6,7,8,9,11,13,12,15,16,17, 18, 19)
season_2002.show(5)
season_2016.show(5)
season_2022.show(5)
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

Now I need to filiter the data so that I can use useful data. The problem with the data right now is that it has players who usually do not play that many games. If a player only has a few games played they might skew the data. I don't want that so I will try to filiter it so that a player has to have played more than the average player.

[5]:
```
#this is what i used to find the average amount of games
avg_gp_2002 = int(season_2002.column('gp').mean())
avg_gp_2016 = int(season_2016.column('gp').mean())
avg_gp_2022 = int(season_2022.column('gp').mean())



season_2002_filtered = season_2002.where("gp", are.above(avg_gp_2002))
season_2016_filtered = season_2016.where("gp", are.above(avg_gp_2016))
season_2022_filtered = season_2022.where("gp", are.above(avg_gp_2016))

season_2002_filtered.show(5)
season_2016_filtered.show(5)
season_2022_filtered.show(5)
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

```
<IPython.core.display.HTML object>
```

Now that I have the filtered data. I want to find what the min/max of height and weight to better understand who the tallest and biggest players are. This will help in creating the histogram for each table.

[6]:
```
#For 2002
print("2002 min height = " + str(season_2002_filtered.column('player_height').
 ↪min()))
print("2002 max height = " + str(season_2002_filtered.column('player_height').
 ↪max()))
print("2002 min weight = " + str(season_2002_filtered.column('player_weight').
 ↪min()))
print("2002 max weight = " + str(season_2002_filtered.column('player_weight').
 ↪max()))
```

```
2002 min height = 165.1
2002 max height = 228.6
2002 min weight = 60.327736
2002 max weight = 151.95332
```

[7]:
```
#For 2016
print("2016 min height = " + str(season_2016_filtered.column('player_height').
 ↪min()))
```

```
print("2016 max height = " + str(season_2016_filtered.column('player_height').
  ↪max()))
print("2016 min weight = " + str(season_2016_filtered.column('player_weight').
  ↪min()))
print("2016 max weight = " + str(season_2016_filtered.column('player_weight').
  ↪max()))
```

```
2016 min height = 175.26
2016 max height = 220.98
2016 min weight = 68.0388
2016 max weight = 131.088088
```

[8]:
```
#For 2022
print("2022 min height = " + str(season_2022_filtered.column('player_height').
  ↪min()))
print("2022 max height = " + str(season_2022_filtered.column('player_height').
  ↪max()))
print("2022 min weight = " + str(season_2022_filtered.column('player_weight').
  ↪min()))
print("2022 max weight = " + str(season_2022_filtered.column('player_weight').
  ↪max()))
```

```
2022 min height = 177.8
2022 max height = 215.9
2022 min weight = 74.389088
2022 max weight = 131.54168
```

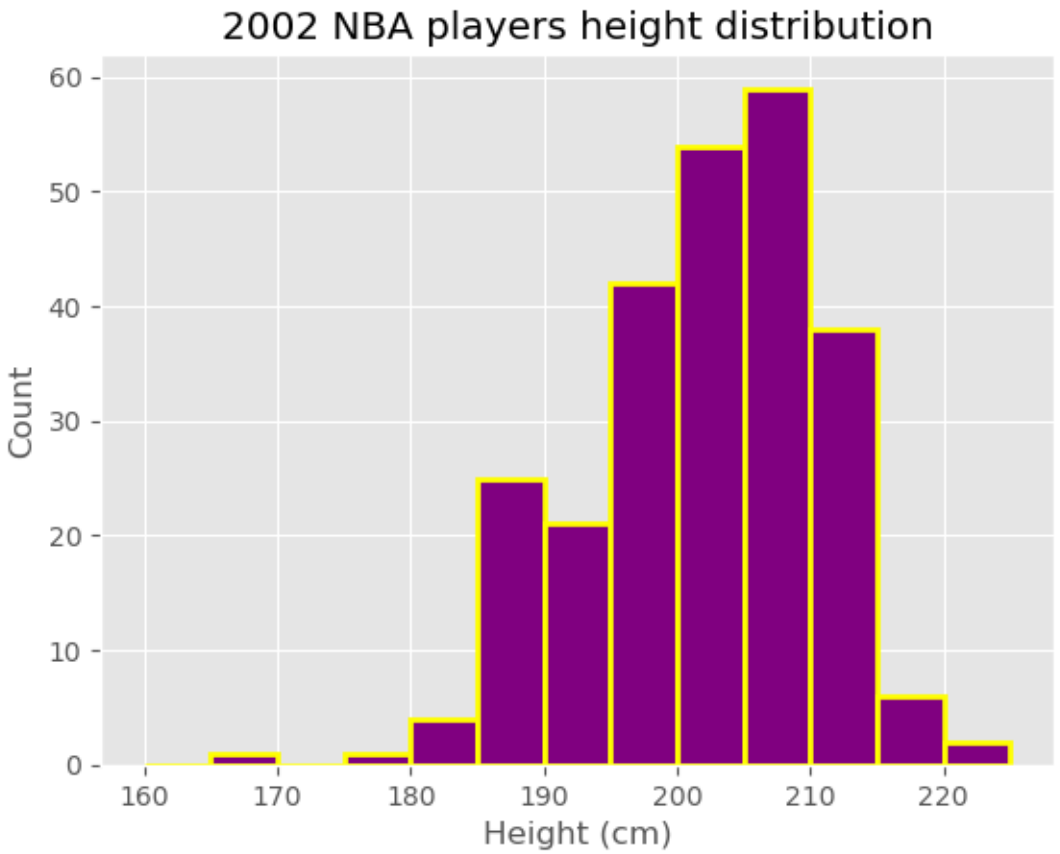
0.3 Modeling the data

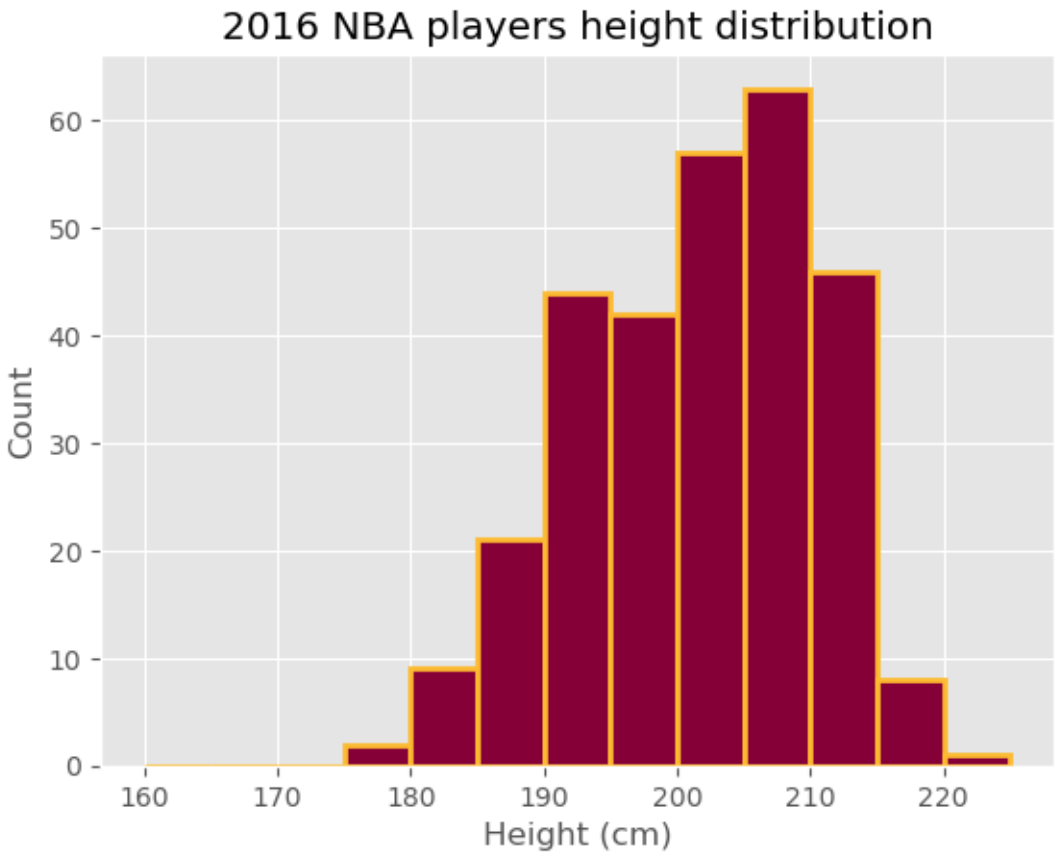
Now that I have all the preliminary data. I can create a histogram for each year and find the tallest and biggest players. The colors are to represent the champions of that season. The idea is that after finding the taller and heaviest players, we can tell if they are dominant or not in the season by looking at their net rating.

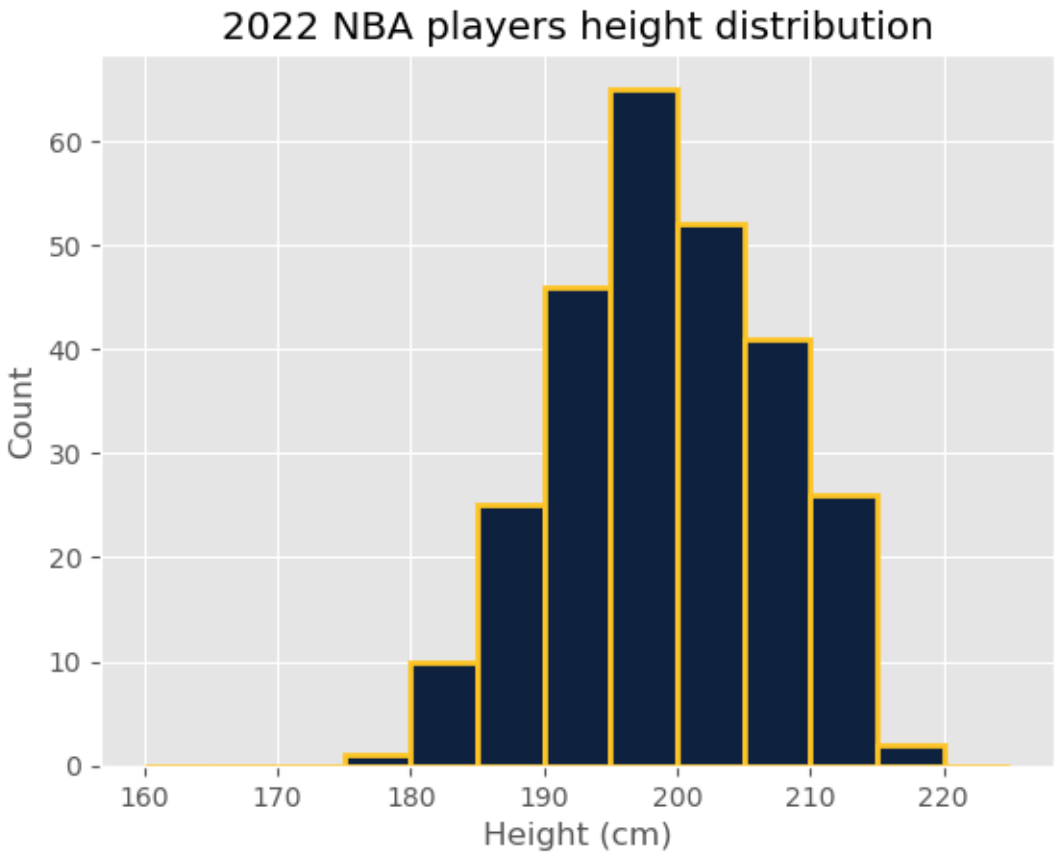
0.3.1 Histograms for height

[9]:
```
#histograms for height
myBin_height = np.arange(160, 230, 5)
plt.title('2002 NBA players height distribution')
plt.xlabel('Height (cm)')
plt.ylabel('Count')
plt.hist(season_2002_filtered.column("player_height"), bins= myBin_height,␣
  ↪color="purple", ec="yellow",
         lw=2, density=False)
plt.show()

plt.title('2016 NBA players height distribution')
```

```
plt.xlabel('Height (cm)')
plt.ylabel('Count')
plt.hist(season_2016_filtered.column("player_height"), bins= myBin_height,␣
 ↪color="#860038", ec="#FDBB30",
         lw=2, density=False)
plt.show()

plt.title('2022 NBA players height distribution')
plt.xlabel('Height (cm)')
plt.ylabel('Count')
plt.hist(season_2022_filtered.column("player_height"), bins= myBin_height,␣
 ↪color="#0E2240", ec="#FEC524",
         lw=2, density=False)
plt.show()
```



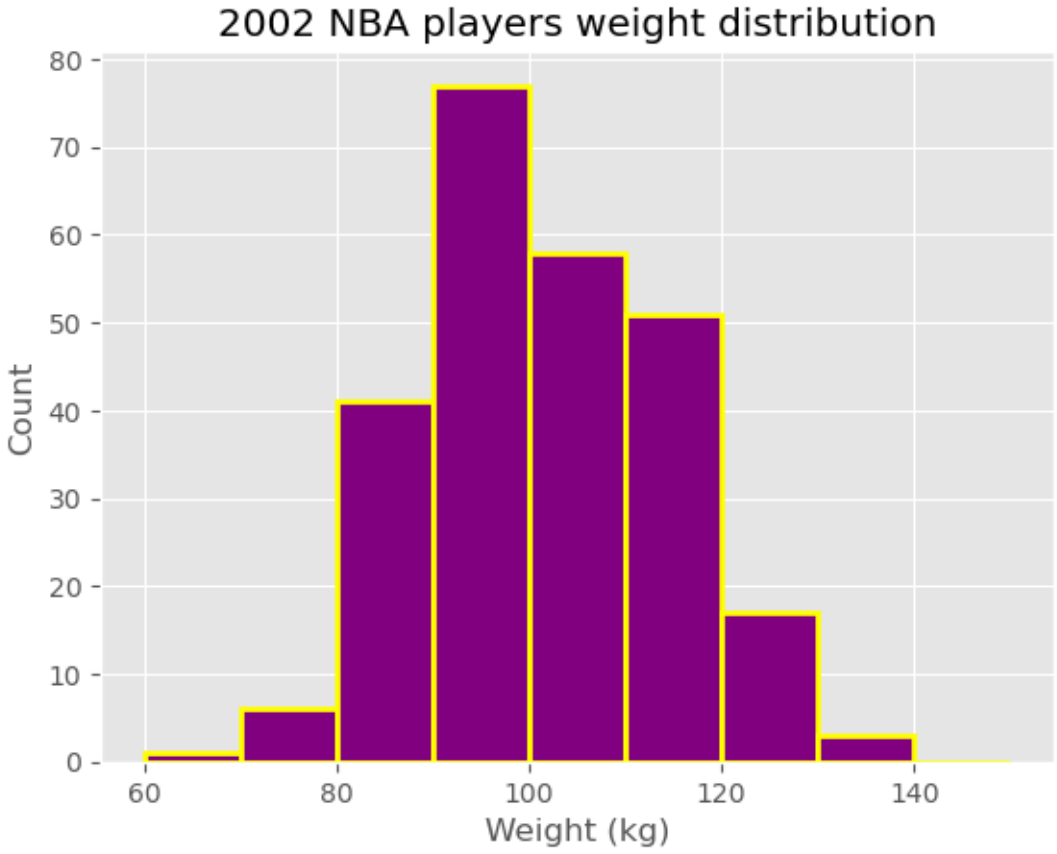

2016 NBA players height distribution
Count
Height (cm)
0
10
20
30
40
50
60
160
170
180
190
200
210
220

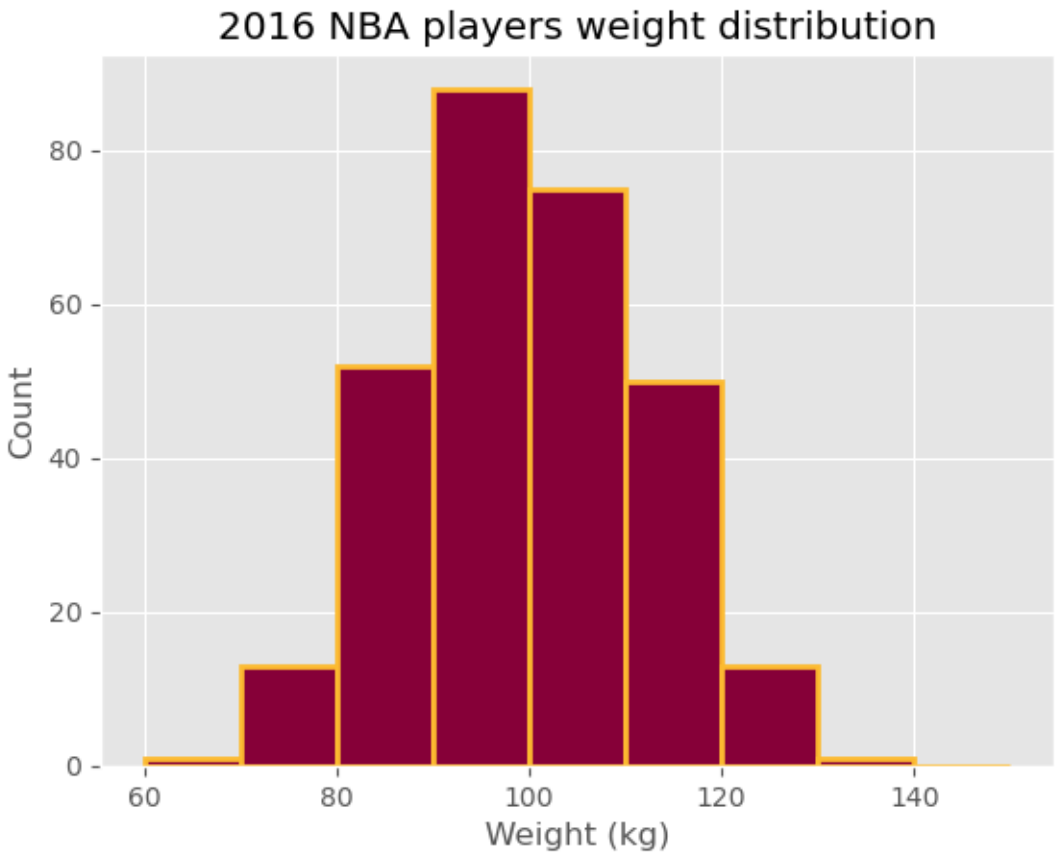


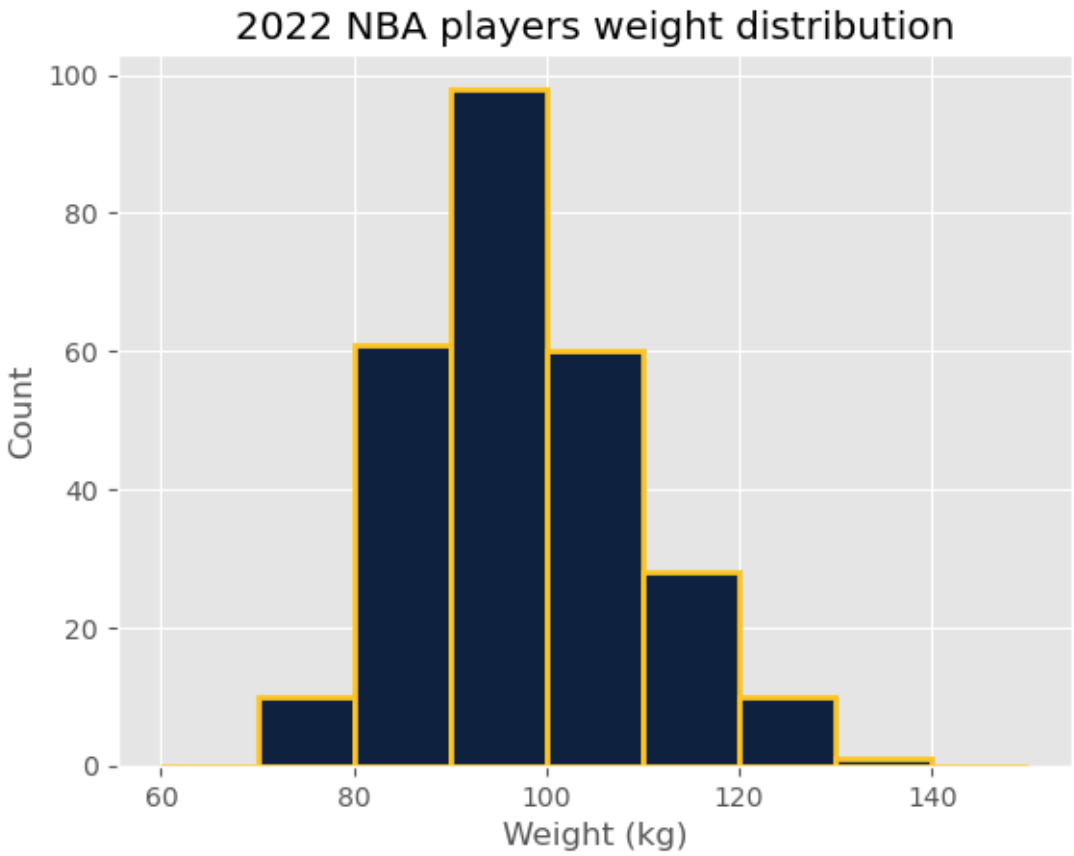
0.3.2 Now doing the same with weight

[11]:
```
myBin_weight = np.arange(60, 160, 10)
plt.title('2002 NBA players weight distribution')
plt.xlabel('Weight (kg)')
plt.ylabel('Count')
plt.hist(season_2002_filtered.column("player_weight"), bins= myBin_weight,␣
 ↪color="purple", ec="yellow",
         lw=2, density=False)
plt.show()

plt.title('2016 NBA players weight distribution')
plt.xlabel('Weight (kg)')
plt.ylabel('Count')
plt.hist(season_2016_filtered.column("player_weight"), bins= myBin_weight,␣
 ↪color="#860038", ec="#FDBB30",
         lw=2, density=False)
plt.show()
```

```
plt.title('2022 NBA players weight distribution')
plt.xlabel('Weight (kg)')
plt.ylabel('Count')
plt.hist(season_2022_filtered.column("player_weight"), bins= myBin_weight,␣
 ↪color="#0E2240", ec="#FEC524",
         lw=2, density=False)
plt.show()
```




2016 NBA players weight distribution
Count
0
20
40
60
80
60
80
100
120
140
Weight (kg)



Looking at these histograms, we can see that the tallest players in the NBA for these three seasons are taller than 210cm (about 6ft 11in), and the heaviest player weighs more than 120kg (264 lbs). By plotting the players into a scatter plot, we can see how many were dominant in their respective seasons.

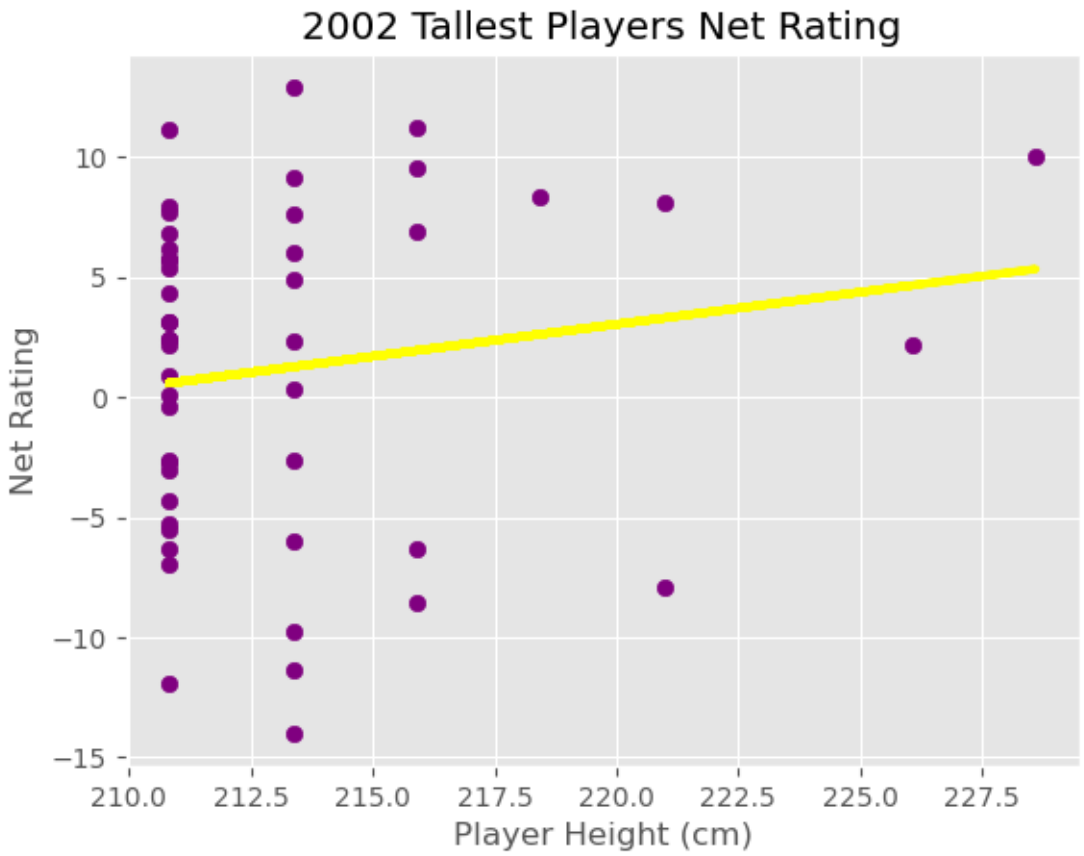
0.3.3 Data Visualization

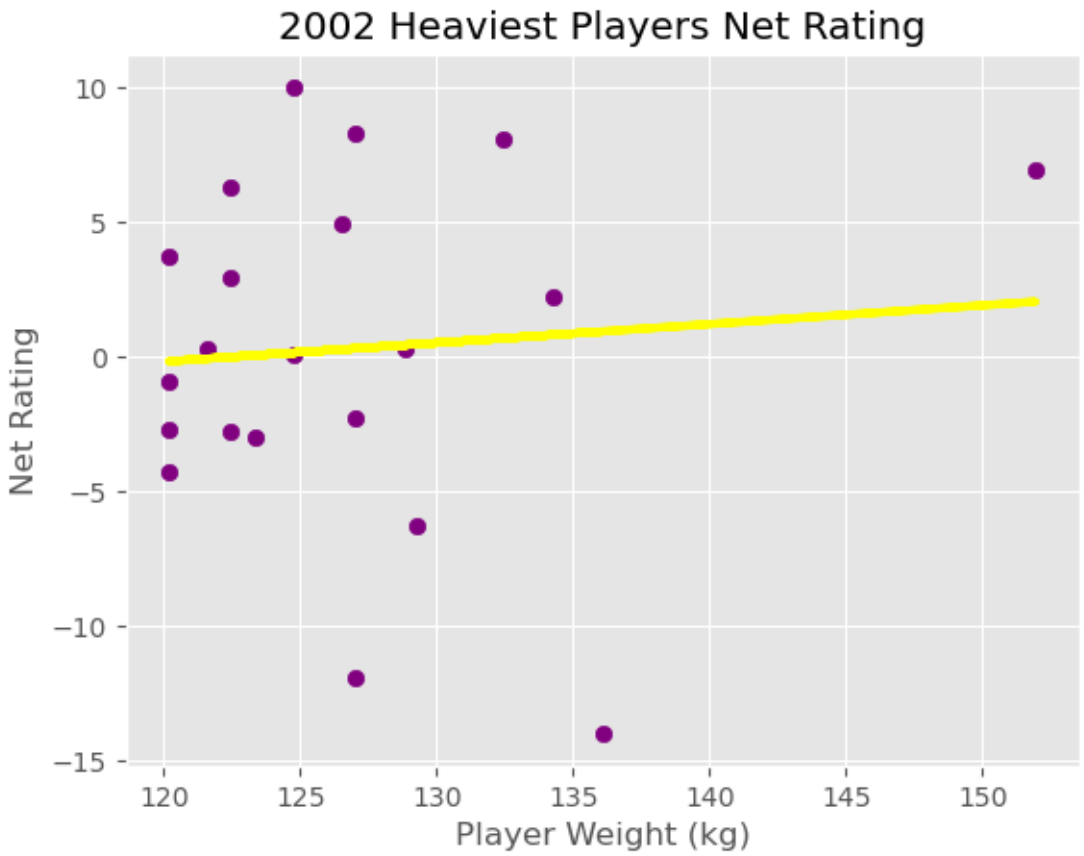
I will use a scatter plot to look at all the data and see how well the tallest and heaviest players play. The scatter plots align with how the histograms were colored.

0.3.4 2002

[12]:
```
## make a table of just the height and the net_rating 2003 -> 2022
season_2002_filtered_tallest = season_2002_filtered.where("player_height", are.
 ↪above(210))
height_rating_2002 = season_2002_filtered_tallest.drop(2,3)
#height
plt.title('2002 Tallest Players Net Rating')
x = height_rating_2002.column('player_height')
y = height_rating_2002.column('net_rating')
```

```
plt.xlabel("Player Height (cm)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = 'purple')
plt.plot(x, a*x+b, color = 'yellow', linewidth=3)
plt.show()

#weight

season_2002_filtered_weight = season_2002_filtered.where("player_weight", are.
 ↪above(120))
weight_rating_2002 = season_2002_filtered_weight.drop(1,3)


plt.title('2002 Heaviest Players Net Rating')
x = weight_rating_2002.column('player_weight')
y = weight_rating_2002.column('net_rating')
plt.xlabel("Player Weight (kg)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = 'purple')
plt.plot(x, a*x+b, color = 'yellow', linewidth=3)
plt.show()
```

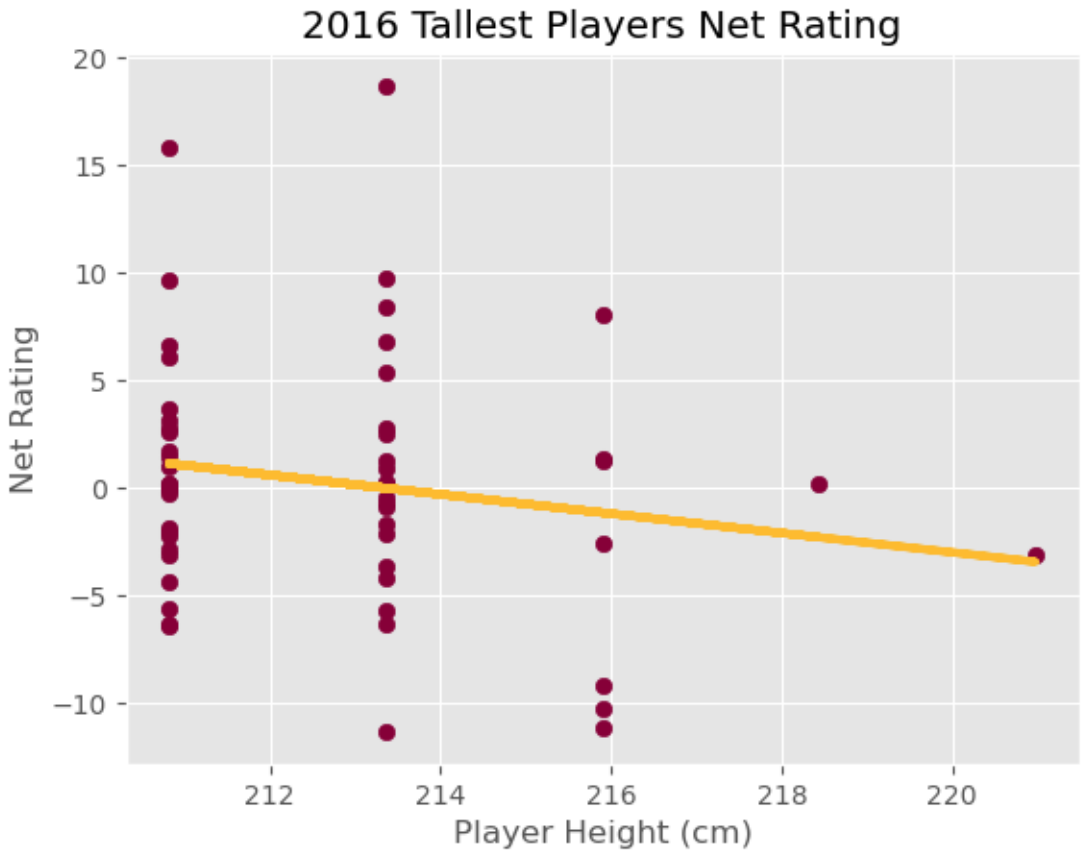

2002 Tallest Players Net Rating
Net Rating
10
5
0
−5
−10
−15
210.0
212.5
215.0
217.5
220.0
222.5
225.0
227.5
Player Height (cm)

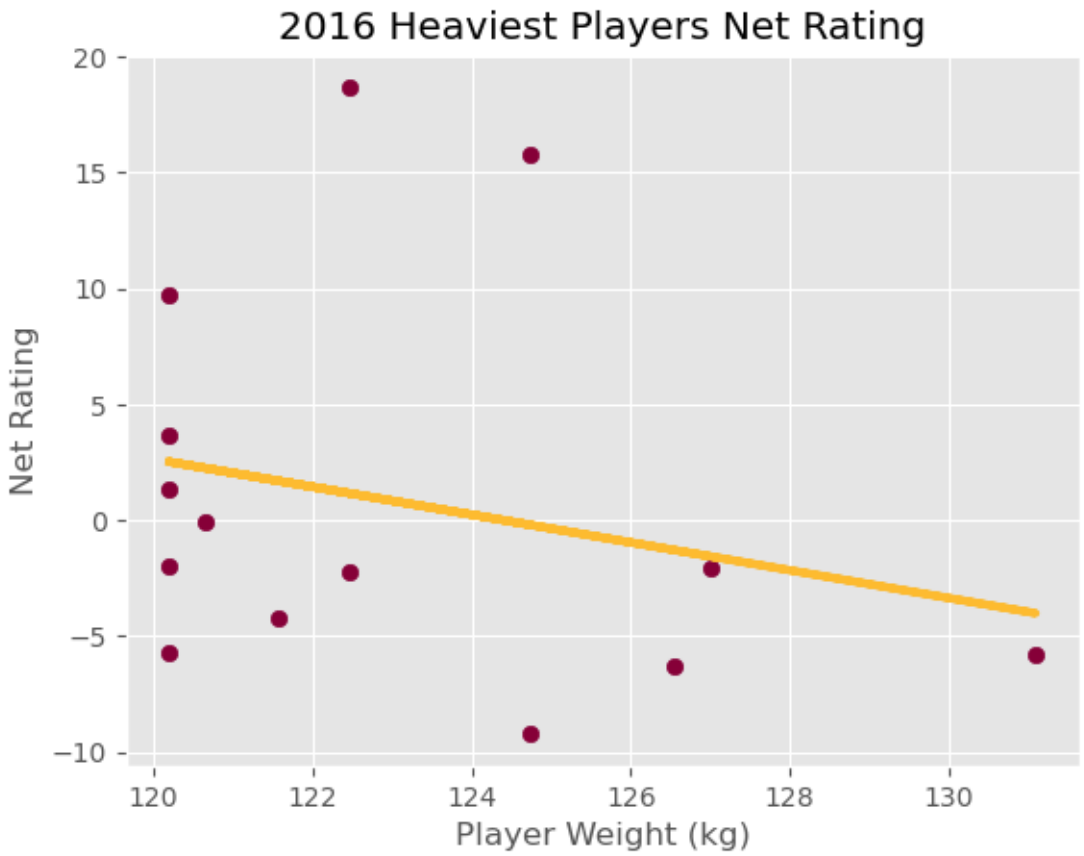


0.3.5 2016

[13]:
```
## 2016
## make a table of just the height and the net_rating 2003 -> 2022
season_2016_filtered_tallest = season_2016_filtered.where("player_height", are.
 ↪above(210))
height_rating_2016 = season_2016_filtered_tallest.drop(2,3)
#height
plt.title('2016 Tallest Players Net Rating')
x = height_rating_2016.column('player_height')
y = height_rating_2016.column('net_rating')
plt.xlabel("Player Height (cm)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = '#860038')
plt.plot(x, a*x+b, color = '#FDBB30', linewidth=3)
plt.show()

#weight
```

```
season_2016_filtered_weight = season_2016_filtered.where("player_weight", are.
 ↪above(120))
weight_rating_2016 = season_2016_filtered_weight.drop(1,3)


plt.title('2016 Heaviest Players Net Rating')
x = weight_rating_2016.column('player_weight')
y = weight_rating_2016.column('net_rating')
plt.xlabel("Player Weight (kg)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = '#860038')
plt.plot(x, a*x+b, color = '#FDBB30', linewidth=3)
plt.show()
```

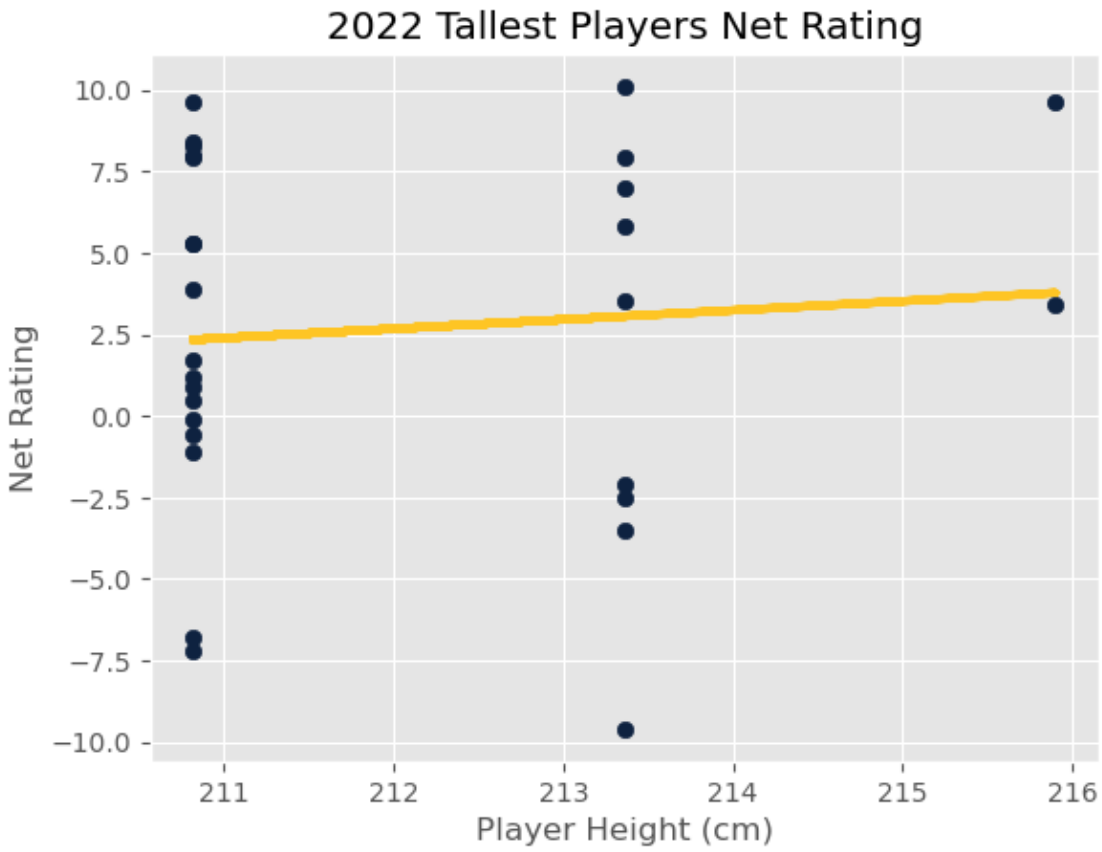


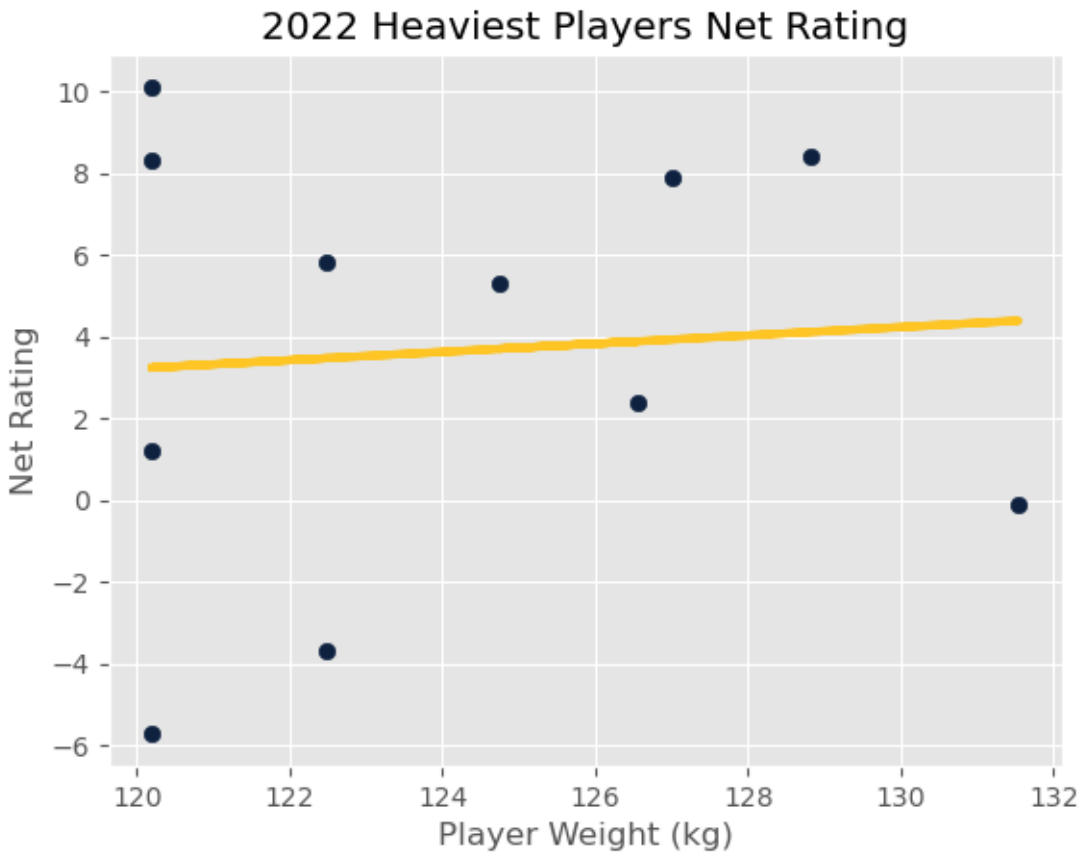


0.3.6 2022

[14]:
```
season_2022_filtered_tallest = season_2022_filtered.where("player_height", are.
  ↪above(210))
height_rating_2022 = season_2022_filtered_tallest.drop(2,3)

plt.title('2022 Tallest Players Net Rating')
x = height_rating_2022.column('player_height')
y = height_rating_2022.column('net_rating')
plt.xlabel("Player Height (cm)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = '#0E2240')
plt.plot(x, a*x+b, color = '#FEC524', linewidth=3)
plt.show()

season_2022_filtered_weight = season_2022_filtered.where("player_weight", are.
  ↪above(120))
weight_rating_2022 = season_2022_filtered_weight.drop(1,3)
```

```
plt.title('2022 Heaviest Players Net Rating')
x = weight_rating_2022.column('player_weight')
y = weight_rating_2022.column('net_rating')
plt.xlabel("Player Weight (kg)")
plt.ylabel("Net Rating")
a, b = np.polyfit(x, y, 1)
plt.scatter(x,y, color = '#0E2240')
plt.plot(x, a*x+b, color = '#FEC524', linewidth=3)
plt.show()
```





What is happening in the scatter plots? The scatter plots compare the net rating to the player's weight and height.

In 2002, we can see that the tallest players dominated the sport since many had above a +5 net rating, and some had a +10 net rating. We observed that the same for the heaviest players in 2002.

In 2016, we can see a stark contrast to 2002. The negative slope in both plots suggests that the tallest and heaviest players struggled in this season. There are two outliers in the plots, and we will look at them in the next part of the analysis, where we look at individual players.

In 2022, we can see that the tallest and heaviest players have started to play decent again. Both plots show that the tallest players and heaviest players are playing well, but they are not dominating the league as they did in 2002.

0.3.7 What happens when we look at individual players to see how game philosophy changes?

We can look at this by specifying Lebron James as the standard dominant player. We can compare and contrast him with other players who are tall (above 6ft 11in), and we can also compare LeBron with dominant short players (below 6 ft 6in). The problem is that no one in the league has had a long career as Lebron. To fix this problem, I can find different players in different eras. I can

separate the ears into three: 1) Late Transitional Era (2002 - 2009) 2) Early Modern Era (2009 - 2013) 3) Modern Era (2013-now)

In the Late Transitional Era, we will have one tall player and one short player. The tall player who dominated in the Era was Yao Ming. Dwyane Wade was a short player who made a name for himself in this Era.

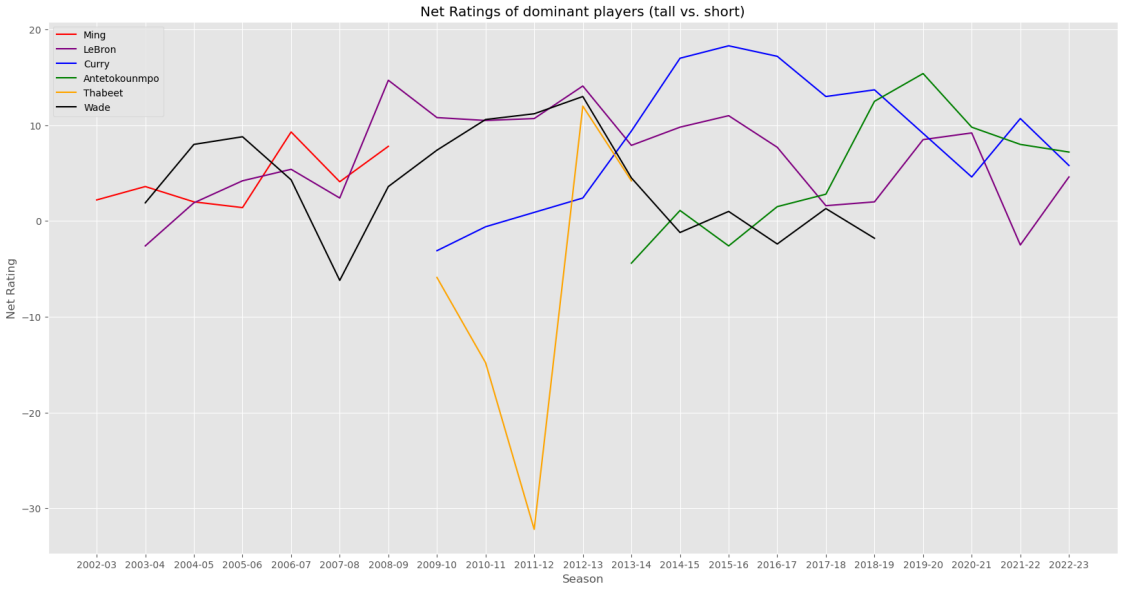
In the Early Modern Era, we will have one tall player and one short player. In this Era, there really wasn't a short or tall player that dominated the sport. Hasheem Thabeet was one of the few tall players in this Era. During this Era, most tall players saw less time on the court. Dwyane Wade also played in this Era, but in this Era, he dominated the sport during this Era.

In the Modern Era, we will also have one tall player and one short player. Stephen Curry and Giannis Antetokounmpo are great for this Era because they both have been dominating the sport. Here, there will be overlap because the players dominating in the Modern Era started playing in the early Modern Era as well. It will be interesting to see because these players represent the "new school" of basketball.

[15]:
```
#taking all the players seasons as a table
LeBron_seasons = data.where("player_name", "LeBron James")

Curry_seasons = data.where("player_name", "Stephen Curry")
Curry_seasons = Curry_seasons.where("gp", are.above(42))

Wade_seasons = data.where("player_name", "Dwyane Wade")

Ming_seasons = data.where("player_name", "Yao Ming")
Ming_seasons = Ming_seasons.where("gp", are.above(42))

Giannis_seasons = data.where("player_name", "Giannis Antetokounmpo")

Thabeet_seasons = data.where("player_name", "Hasheem Thabeet")

#LeBron_seasons.show(20)
#Curry_seasons.show(20)
#Ming_seasons.show(15)
#Kristaps_seasons.show(20)
#Thabeet_seasons.show(10)
#Wade_seasons.show(10)
#Embiid_seasons.show(10)
#Giannis_seasons.show(10)
```

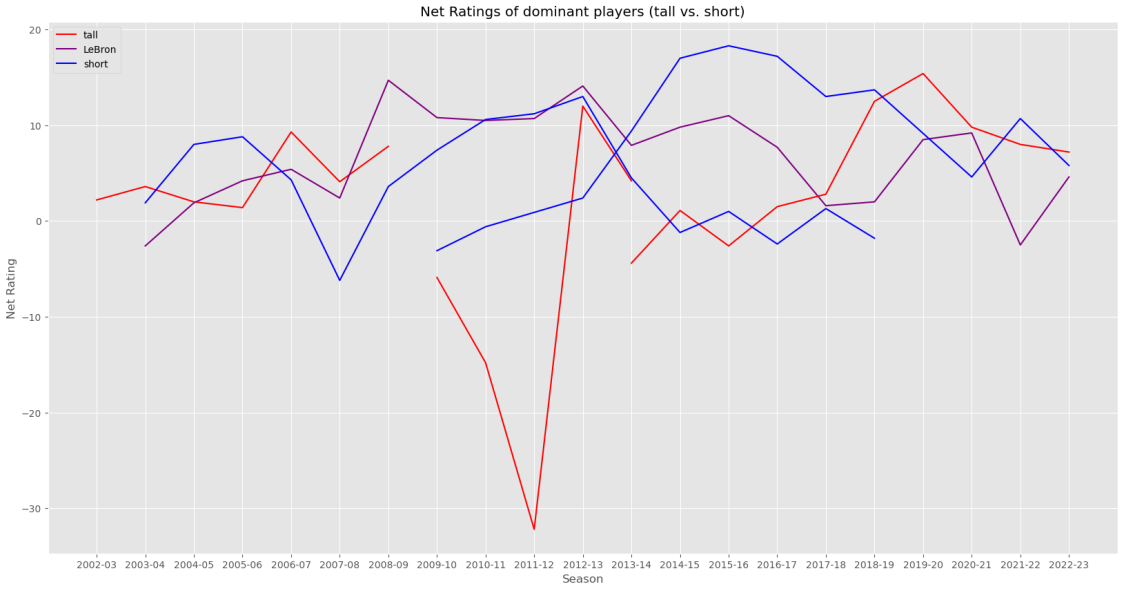
[25]:
```
#now that we have all the players we can try to plot them by seaons and␣
 ↪netRating
LeBron_dates = LeBron_seasons.column("season")
LeBron_rating = LeBron_seasons.column("net_rating")

Curry_dates = Curry_seasons.column("season")
Curry_rating = Curry_seasons.column("net_rating")
```

```
Wade_dates = Wade_seasons.column("season")
Wade_rating = Wade_seasons.column("net_rating")

Ming_dates = Ming_seasons.column("season")
Ming_rating = Ming_seasons.column("net_rating")

Giannis_dates = Giannis_seasons.column("season")
Giannis_rating = Giannis_seasons.column("net_rating")

Thabeet_dates = Thabeet_seasons.column("season")
Thabeet_rating = Thabeet_seasons.column("net_rating")

plt.xlabel('Season')
plt.ylabel('Net Rating')
plt.title('Net Ratings of dominant players (tall vs. short)')
plt.rcParams["figure.figsize"] = (10, 10)

plt.plot(Ming_dates, Ming_rating, color="Red", label="Ming")

plt.plot(LeBron_dates, LeBron_rating, color="Purple", label="LeBron")

plt.plot(Curry_dates, Curry_rating, color="Blue", label="Curry")

plt.plot(Giannis_dates, Giannis_rating, color = "green", label="Antetokounmpo")

plt.plot(Thabeet_dates, Thabeet_rating, color="orange", label="Thabeet")

plt.plot(Wade_dates, Wade_rating, color="black", label="Wade")

plt.legend(loc="upper left")



plt.show()
```



The graph plot above is very confusing to look at. I seperated it into shoter players vs. taller players so that it is easier to interpert.

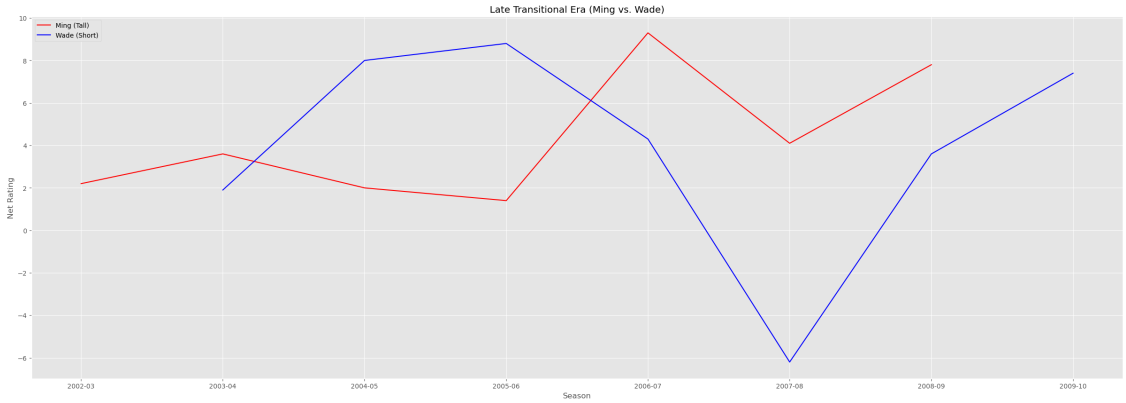
[20]:
```
plt.xlabel('Season')
plt.ylabel('Net Rating')
plt.title('Net Ratings of dominant players (tall vs. short)')
plt.rcParams["figure.figsize"] = (20, 10)



#red = tall, blue = short

plt.plot(Ming_dates, Ming_rating, color="Red", label="tall")
plt.plot(LeBron_dates, LeBron_rating, color="Purple", label="LeBron") #standard␣
 ↪player

plt.plot(Giannis_dates, Giannis_rating, color = "Red")
plt.plot(Thabeet_dates, Thabeet_rating, color="Red")

plt.plot(Curry_dates, Curry_rating, color="Blue", label="short")
plt.plot(Wade_dates, Wade_rating, color="Blue")



plt.legend(loc="upper left")


plt.show()
```



0.4 Let's seperate the huge graph into our three Eras

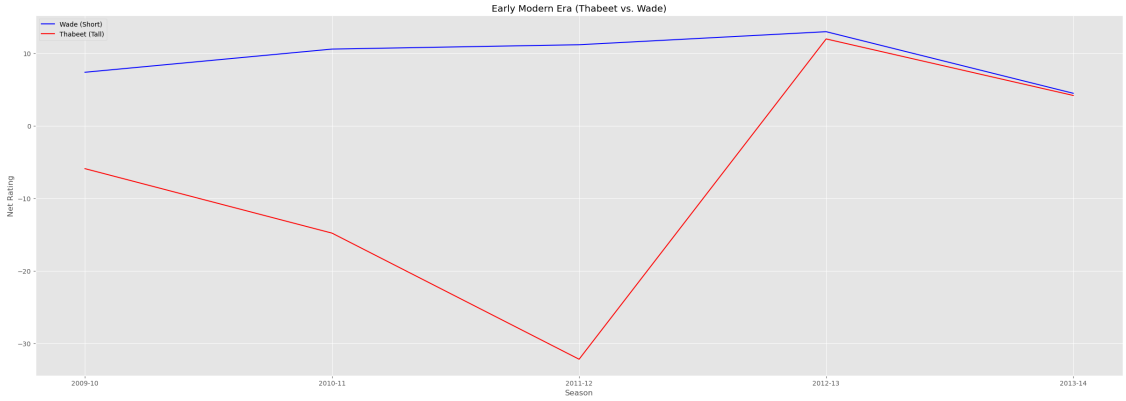
0.4.1 Late Transitional(2002-2009)

[77]:
```
Wade_Late_Transitional = Wade_seasons.exclude(7,8,9,10,11,12,13,14,15)

Wade_Late_Transitional_rating = Wade_Late_Transitional.column("net_rating")
Wade_Late_Transitional_dates = Wade_Late_Transitional.column("season")


plt.xlabel('Season')
plt.ylabel('Net Rating')
plt.title('Late Transitional Era (Ming vs. Wade)')
plt.rcParams["figure.figsize"] = (40, 10)

plt.plot(Ming_dates, Ming_rating, color="Red", label="Ming (Tall)")
plt.plot(Wade_Late_Transitional_dates, Wade_Late_Transitional_rating,␣
 ↪color="Blue", label="Wade (Short)")

plt.legend(loc="upper left")

plt.show()
```



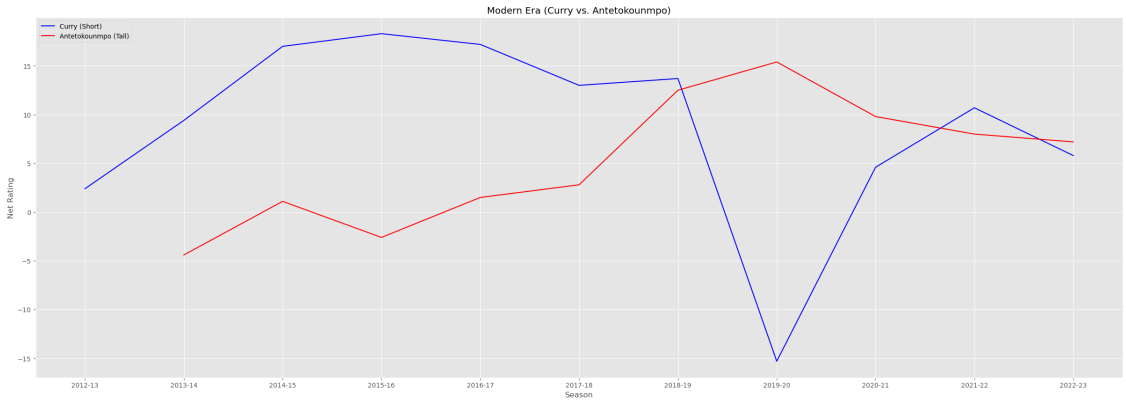
0.4.2 Early Modern Era (2009 - 2013)

[86]:
```
Wade_Early_Modern = Wade_seasons.exclude(0,1,2,3,4,5,11, 12,13,14,15)

Wade_Early_Modern_rating = Wade_Early_Modern.column("net_rating")
Wade_Early_Modern_dates = Wade_Early_Modern.column("season")


plt.xlabel('Season')
plt.ylabel('Net Rating')
plt.title('Early Modern Era (Thabeet vs. Wade)')
plt.rcParams["figure.figsize"] = (30, 10)


plt.plot(Wade_Early_Modern_dates, Wade_Early_Modern_rating, color="Blue",␣
 ↪label="Wade (Short)")
plt.plot(Thabeet_dates, Thabeet_rating, color="Red", label="Thabeet (Tall)")


plt.legend(loc="upper left")

plt.show()
```



0.4.3 Modern Era (2013-now)

[91]:
```
Curry_seasons = data.where("player_name", "Stephen Curry")
Curry_Modern = Curry_seasons.exclude(0, 1,2)
Curry_Modern_rating = Curry_Modern.column("net_rating")
Curry_Modern_seasons = Curry_Modern.column("season")

plt.xlabel('Season')
plt.ylabel('Net Rating')
plt.title('Modern Era (Curry vs. Antetokounmpo)')
plt.rcParams["figure.figsize"] = (30, 10)


plt.plot(Curry_Modern_seasons, Curry_Modern_rating, color="Blue", label="Curry␣
 ↪(Short)")
plt.plot(Giannis_dates, Giannis_rating, color = "Red", label="Antetokounmpo␣
 ↪(Tall)")


plt.legend(loc="upper left")

plt.show()
```



the 2019-2020 season should be ignored for Curry because he only played 5 games that season due to covid.

0.4.4 What are the graphs showing?

The graphs show several things, but the key takeaway is that a 7-foot player did not dominate the sport. In the last twenty years, the shorter players have found more success in the sport, but that could change. As we saw in the scatter plot for 2022, the tallest and heaviest players were playing a lot better compared to 2016. Observing Giannis Antetokounmpo from 2013 to now, we see that the taller players are improving in the Modern Era. Therefore, it is possible that the taller players define a new Era in the NBA.

0.5 Conclusion

In the seasons we looked at, we saw that taller players were doing better in the 2002 season, but as seasons went on, 7-foot-tall players were not doing as well. In 2016, we observed that most of the 7-foot tall players were hurting their teams, and in 2022, we saw that the taller players have started to make a slight comeback.

We also tried to observe dominant tall players vs short players over the last 20 years. We observed that tall players have never dominated the sport and struggled to have long-lasting careers over the other players. We also saw that tall players struggled in the Early Modern Era. However, that seems to be changing as we saw that in 2022, taller players had more of an impact on the game.

0.5.1 Are we in a new Era?

By looking at the analysis, we can come to the conclusion that we could be moving toward a new Era in the NBA. The NBA is changing back to how it was in the Transitional Era as taller players are getting better at shooting from range. It's only a matter of time until we see taller players dominating.

0.5.2 Do the biggest players dominate the sport of Basketball?

Historically, No. Currently, Maybe. Looking at the plot of Tall vs Short in the NBA (and LeBron), we saw that the best tall players were struggling until 2018, but they are still not playing to a point

where they dominate the sport.

0.5.3 Could the analysis be expanded?

Yes, definitely. This report is only a part of the big puzzle that is the sport of basketball. With this report, we concluded that the 7-foot-tall players were not dominant, but we could not answer what height was ideal for the NBA. Further analysis could help in understanding the Early Modern Era of the NBA. We could also take the data we found here to make better future predictions on how the game will change.