13 Prove: Data Analysis

PART I

1. Mean and median of points scored.

mean = players['points'].mean()

median = players['points'].median()

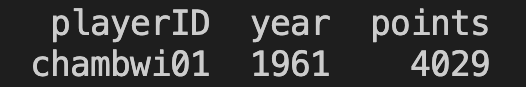
print(f'Mean: {mean:.2f}, Median: {median:.2f}')



1. Highest number of points recorded in a season

print(players[['playerID', 'year', 'points']].sort\_values(

'points', ascending=False).head(1))

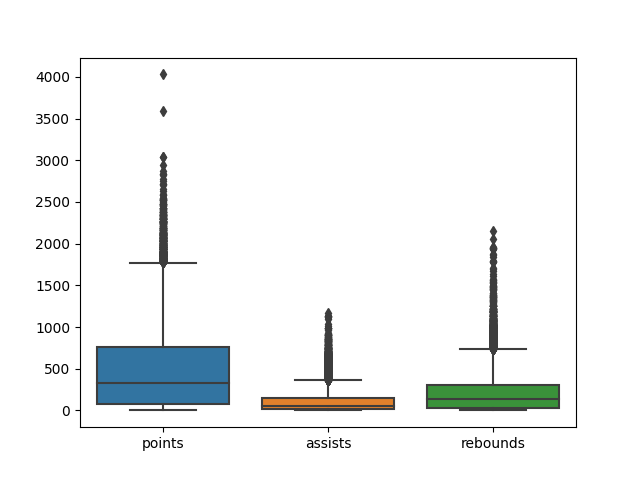


1. Distribution of total points, assists, and rebounds.

sns.boxplot(data=players[['points', 'assists', 'rebounds']])

plt.show()

plt.savefig("boxplot\_rebounds.png")



4. Median of points scored per year, over time.

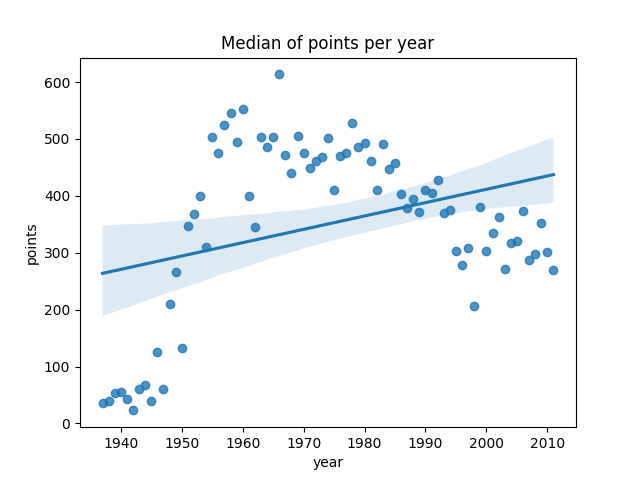
years\_median = players[['points', 'year']].groupby('year').median()

years\_median = years\_median.reset\_index()

sns.regplot(data=years\_median, x='year', y='points').set\_title('Medium of points per year')

plt.show()

plt.savefig("boxplot\_pointYears.png")



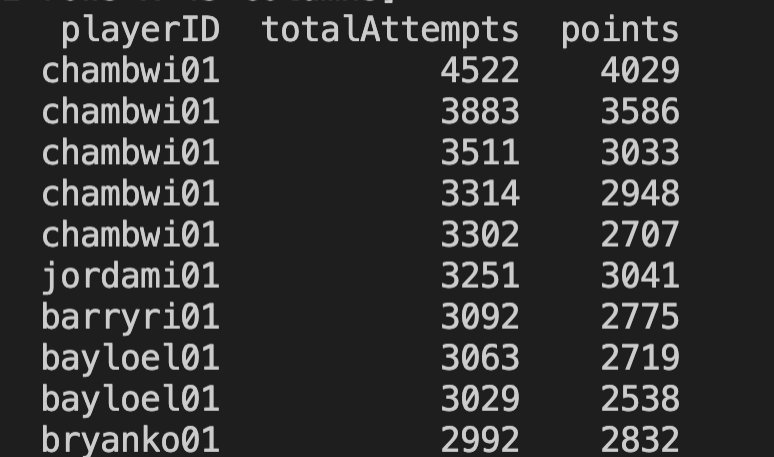
PART II

1. Players that scored more compared to attempts.

sum\_attempts = players['fgAttempted'] + players['ftAttempted']

players['totalAttempts'] = sum\_attempts

print(players[['playerID', 'totalAttempts', 'points']].sort\_values('totalAttempts', ascending=False).head(10))



There is one player who was more efficient at scoring. The player with ID ‘chambwi01’ scored more points per attempt.

1. Exceptional players across many categories.

print(players[['playerID', 'points']].sort\_values(

'points', ascending=False).head(10))

print(players[['playerID', 'rebounds']].sort\_values(

'rebounds', ascending=False).head(10))

print(players[['playerID', 'assists']].sort\_values(

'assists', ascending=False).head(10))

print(players[['playerID', 'steals']].sort\_values(

'steals', ascending=False).head(10))

print(players[['playerID', 'blocks']].sort\_values(

'blocks', ascending=False).head(10))

print(players[['playerID', 'turnovers']].sort\_values(

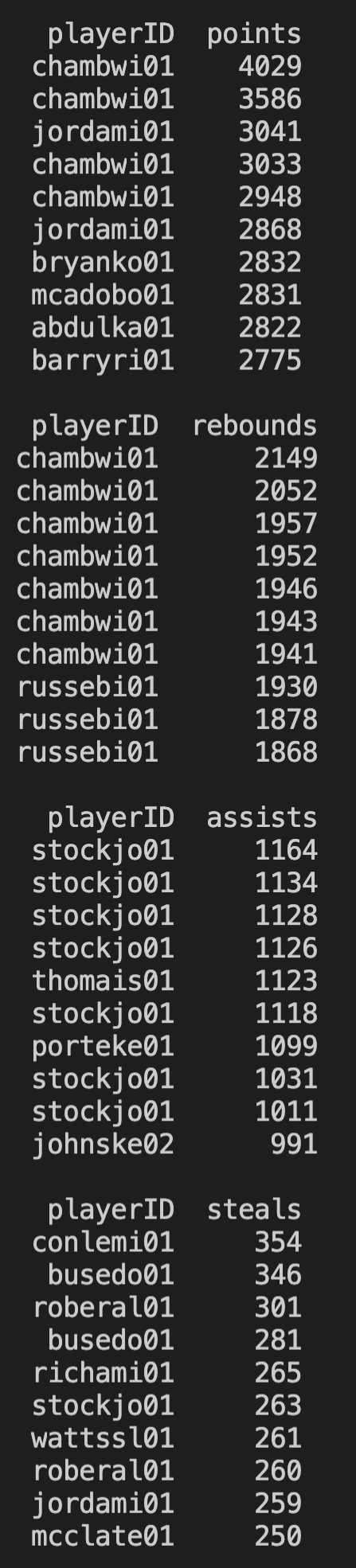
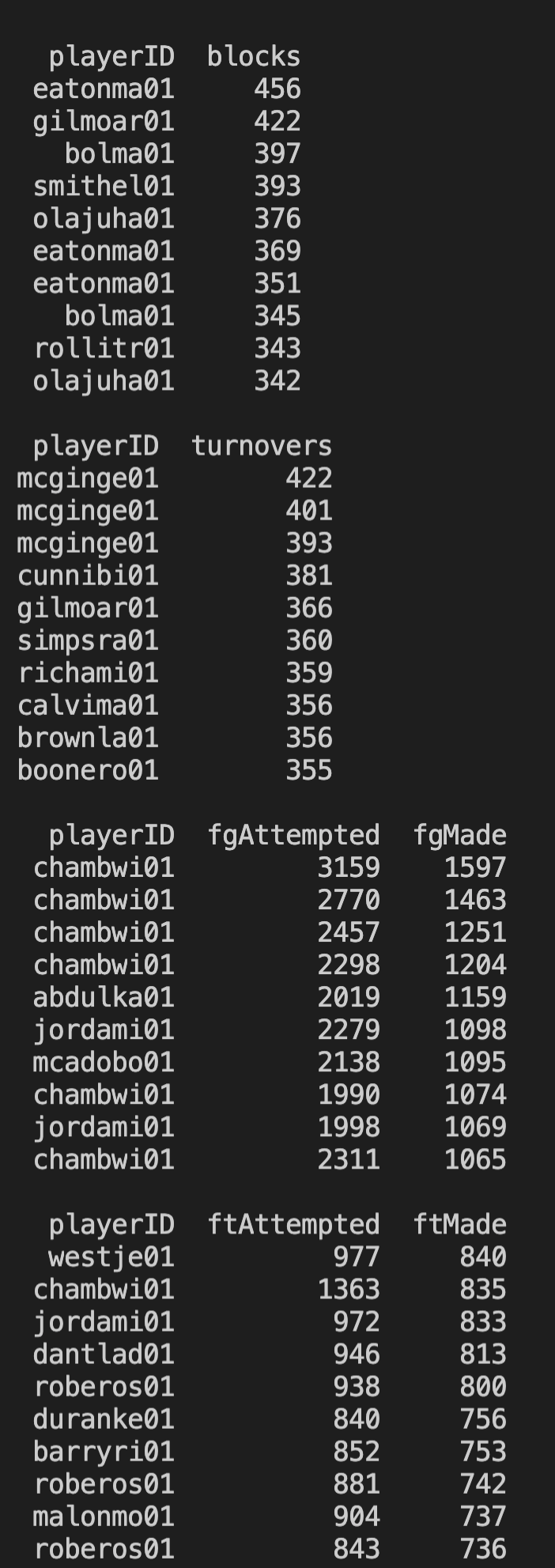
'turnovers', ascending=False).head(10))

print(players[['playerID', 'fgAttempted', 'fgMade']

].sort\_values('fgMade', ascending=False).head(10))

print(players[['playerID', 'ftAttempted', 'ftMade']

].sort\_values('ftMade', ascending=False).head(10))

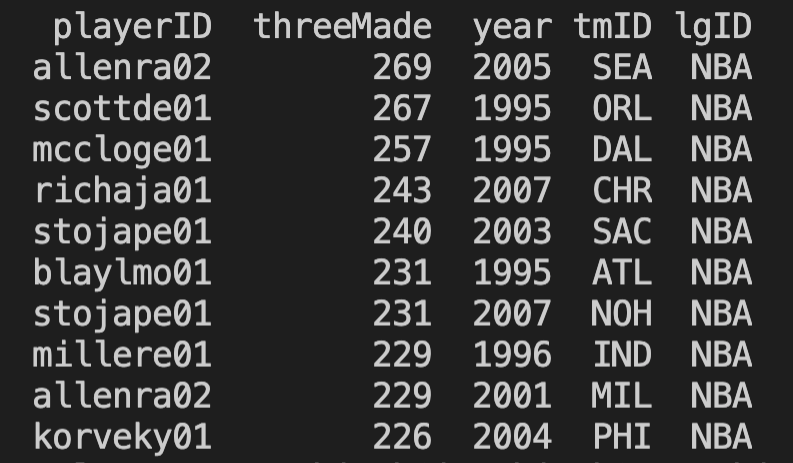
 

The player that repeats in more than one category, as top 10 is chambwi01. He outstands on points, rebounds, field goals, and free-throws categories.

1. Three-point shots.

print(players[['playerID', 'threeMade', 'year', 'tmID', 'lgID']].sort\_values(

'threeMade', ascending=False).head(10))



The top 10 players that scored more three-point shootings played in the NBA league. The popularity increased in 2005.

PART III

1. The GOAT player.

print(players[['playerID', 'points']].sort\_values(

'points', ascending=False).head(10))

print(players[['playerID', 'rebounds']].sort\_values(

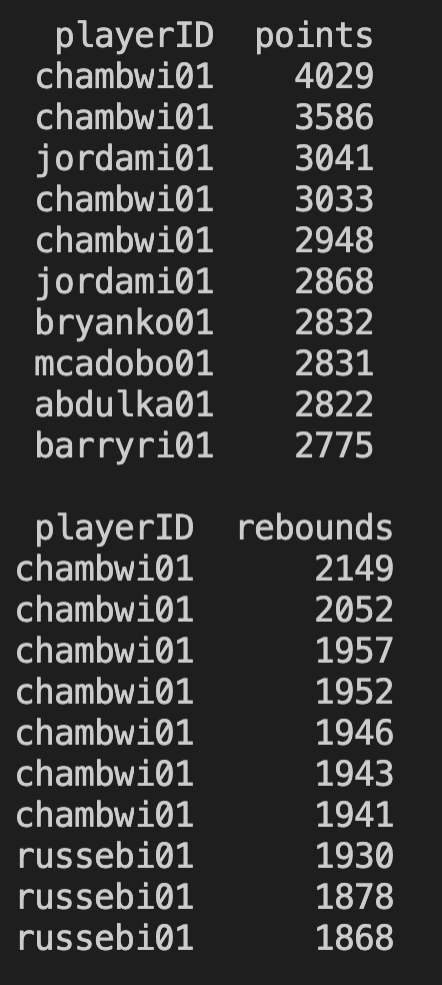
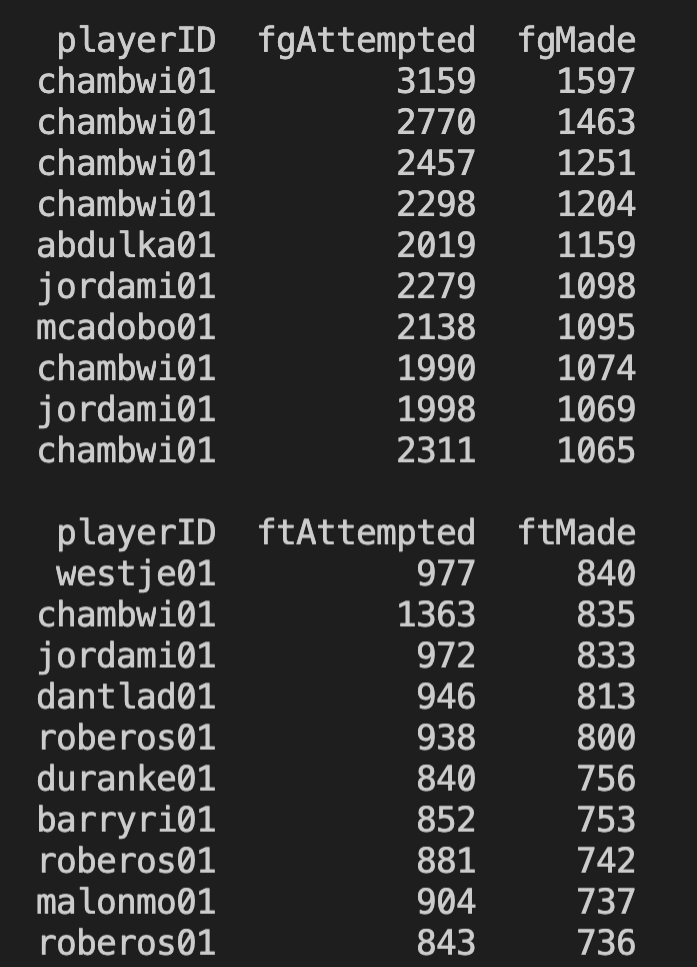
'rebounds', ascending=False).head(10)

print(players[['playerID', 'fgAttempted', 'fgMade']

].sort\_values('fgMade', ascending=False).head(10))

print(players[['playerID', 'ftAttempted', 'ftMade']

].sort\_values('ftMade', ascending=False).head(10))

I think chambwi01 is the greatest of all time because he has the highest score in points, rebounds, and field goals. Only in free-throws he has second place.

1. Something interesting about players from same State.

player\_bio = pd.read\_csv("basketball\_master.csv")

player\_data = pd.merge(players, player\_bio, how="left",

left\_on="playerID", right\_on="bioID")

print(player\_data[['playerID', 'birthCity', 'birthState', 'birthCountry', 'steals']].sort\_values('steals', ascending=False).head(10))

print(player\_data[['playerID', 'birthCity', 'birthState', 'birthCountry', 'turnovers']].sort\_values('turnovers', ascending=False).head(10))



Players who were born in the state of Indiana, like conlemi01, busedo01, and mcginge01, are in the top 10 of more steals and turnovers.

1. Something interesting in the dataset.

print(player\_data[['height', 'points']].sort\_values(

'height', ascending=False).head(10))

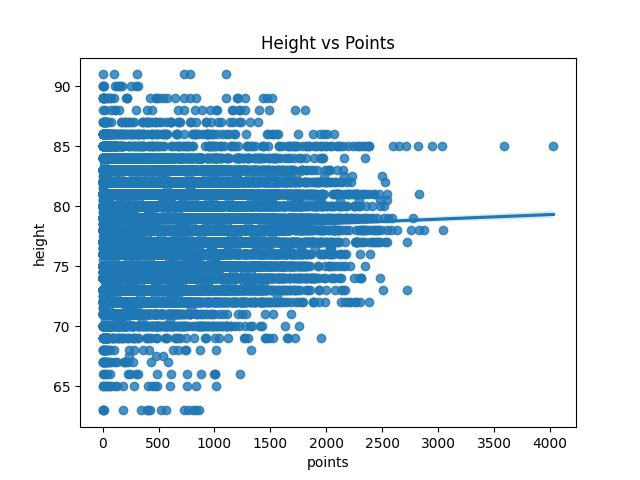
player\_data = player\_data[player\_data["height"] > 0]

sns.regplot(data=player\_data, x='points', y='height').set\_title(

'Height vs Points')

plt.show()

plt.savefig("boxplot\_height.png")



I wanted to know if a player’s height would be directly related to the points scored to see if someone had a physical advantage over other players who scored less. But looking at the graph, I realized that even the tallest players had not scored any points. Nevertheless, the shortest one did not score over 2000 points. The player who had the highest points has around 85 inches of height.