Prove: Data Analysis

Author: Lewis Lockhart

Week 13

Part 1

→ 1.1

Calculate the mean and median number of points scored. (In other words, each row is the amount of points a player scored during a particular season. Calculate the median of these values. The result of this is that we have the median number of points players score each season.)

```
p_mean = data.points.mean()
p_median = data.points.median()
print(f"The mean - points scored: {p_mean}")
print(f"The median - points scored: {p_mean}")
```

The mean - points scored: 492.1306892341375
The median - points scored: 329.0

→ 1.2

Determine the highest number of points recorded in a single season. Identify who scored those points and the year they did so.

```
highest_points = data.points.max()
highest_points_year = data[data.points == highest_points][["points", "year",
"firstName", "lastName"]]
print(f"The highest points: {highest_points}")
print(f"The year of highest of points: {highest_points_year}")
```

The highest number of points recorded in a single season was 4029. This was done by Wilton Chamberland in 1961.

→ 1.3

Produce a boxplot that shows the distribution of total points, total assists, and total rebounds (each of these three is a separate box plot, but they can be on the same scale and in the same graphic).

```
data["total_points"] = data.points + data.PostPoints
data["total_assists"] = data.assists + data.PostAssists
data["total_rebounds"] = data.rebounds + data.PostRebounds
data[["total_points", "total_assists", "total_rebounds"]].plot(kind="box")
matplotlib.pyplot.show()
```
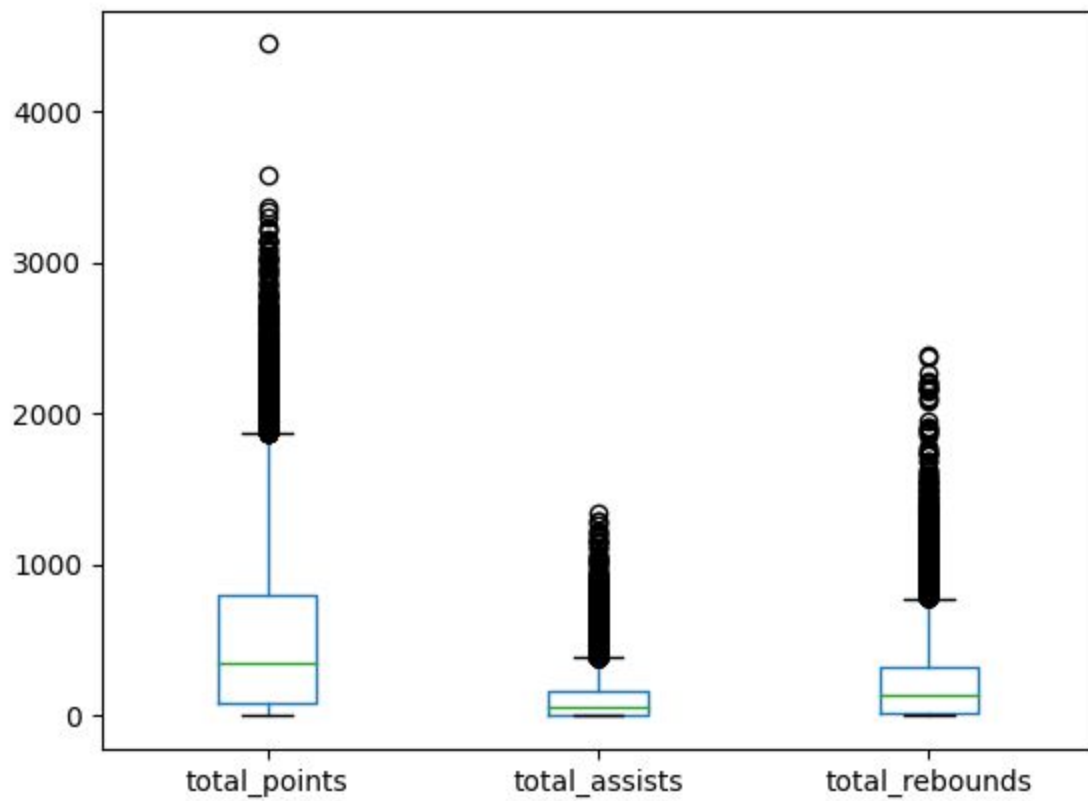
→ 1.4

Produce a plot that shows how the number of points scored has changed over time by showing the median of points scored per year, over time. The x-axis is the year and the y-axis is the median number of points among all players for that year.

```
year_points = data[["year", "points"]].groupby("year").median()
year_points.plot(kind="bar")
matplotlib.pyplot.show()
```
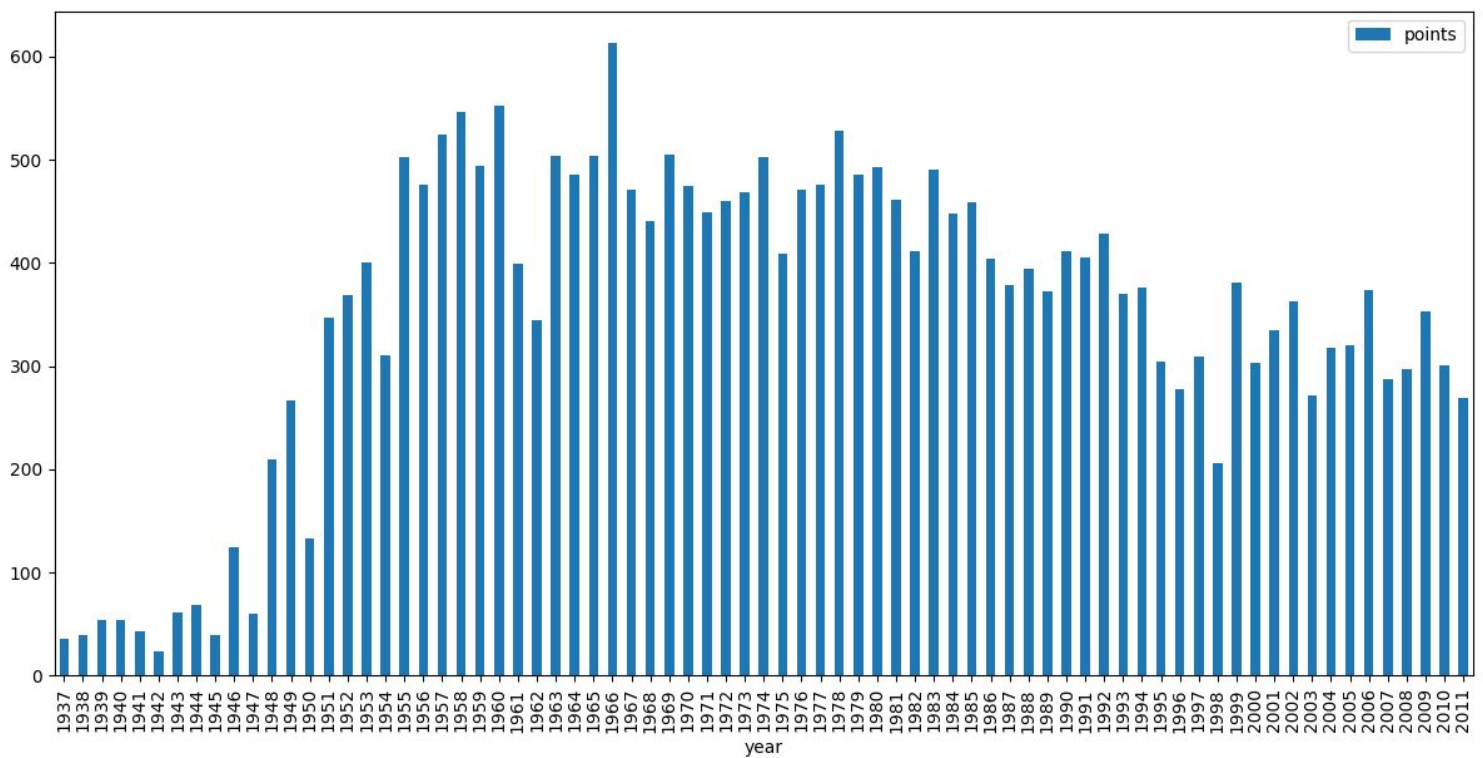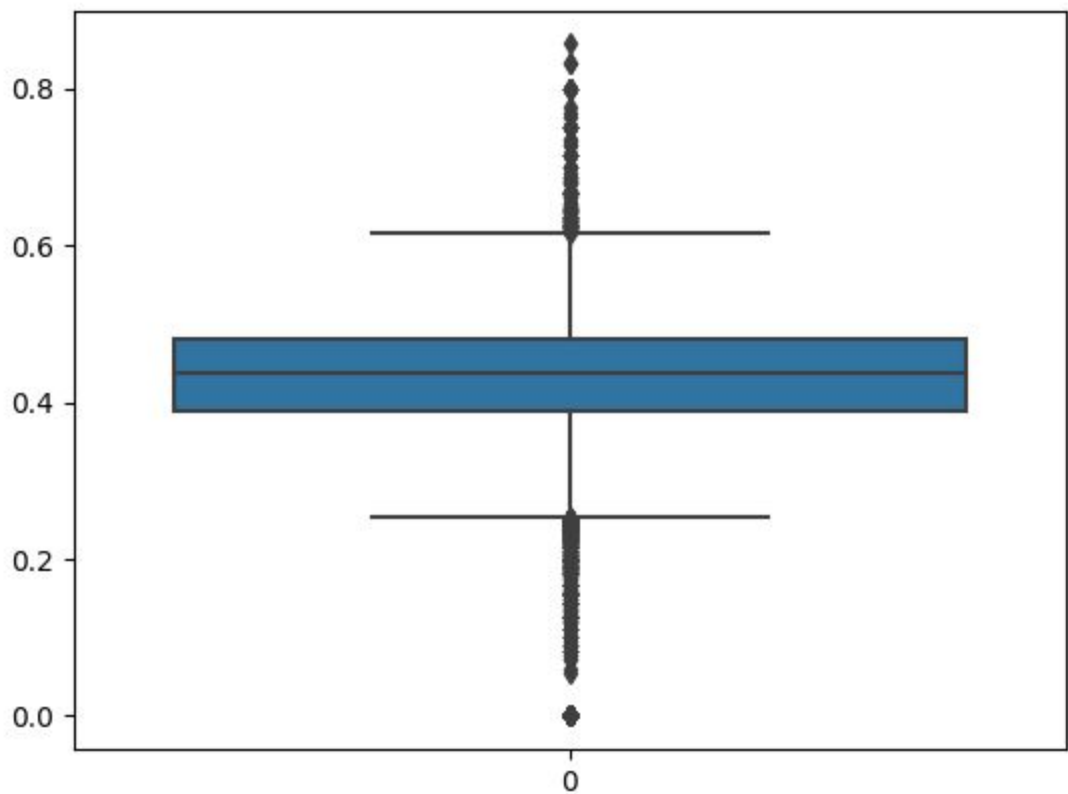
Part 2

→ 2.1

Some players score a lot of points because they attempt a lot of shots. Among players that have scored a lot of points, are there some that are much more efficient (points per attempt) than others?

The players represented at the top of the following graph are those with the highest point to attempt ratio.

→ 2.2

It seems like some players may excel in one statistical category, but produce very little in other areas. Are there any players that are exceptional across many categories?

Below is a list of players that seem to excel in many areas.

firstName	lastName	fgSuccess	ftSuccess	assists	rebounds
Robert	Jones	0.603881	0.695167	303	692
Dave	Twardzik	0.611628	0.841549	247	202
Kareem	Abdul-Jabbar	0.603760	0.764706	371	886
Cedric	Maxwell	0.609333	0.787004	199	704
Darryl	Dawkins	0.606887	0.720395	109	545
Artis	Gilmore	0.670343	0.704887	172	828

→ 2.3

Much has been said about the rise of the three-point shot in recent years. It seems that players are shooting and making more three-point shots than ever. Recognizing that this dataset doesn't contain the very most recent data, do you see a trend of more three-point shots either across the league or among certain groups of players? Is there a point at which popularity increased dramatically?

**Between 1965 and 2011 there is an obvious upward trend of three point shots made.**
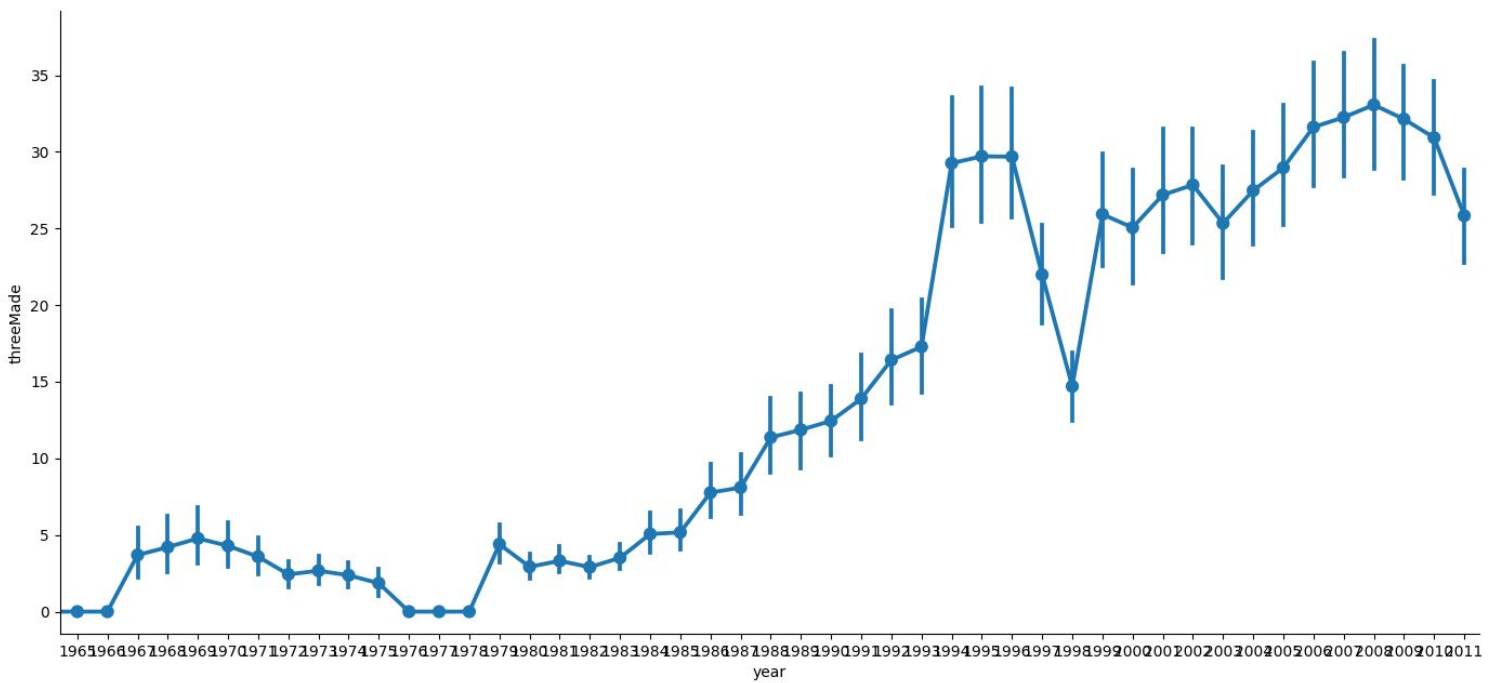
Part 3

→ 3.1

Many sports analysts argue about which player is the GOAT (the Greatest Of All Time). Based on this data, who would you say is the GOAT? Provide evidence to back up your decision.

Based on an aggregation of each players total points, rebounds, assists, steals, blocks, turnovers, I determined the following three players could have a GOAT status.
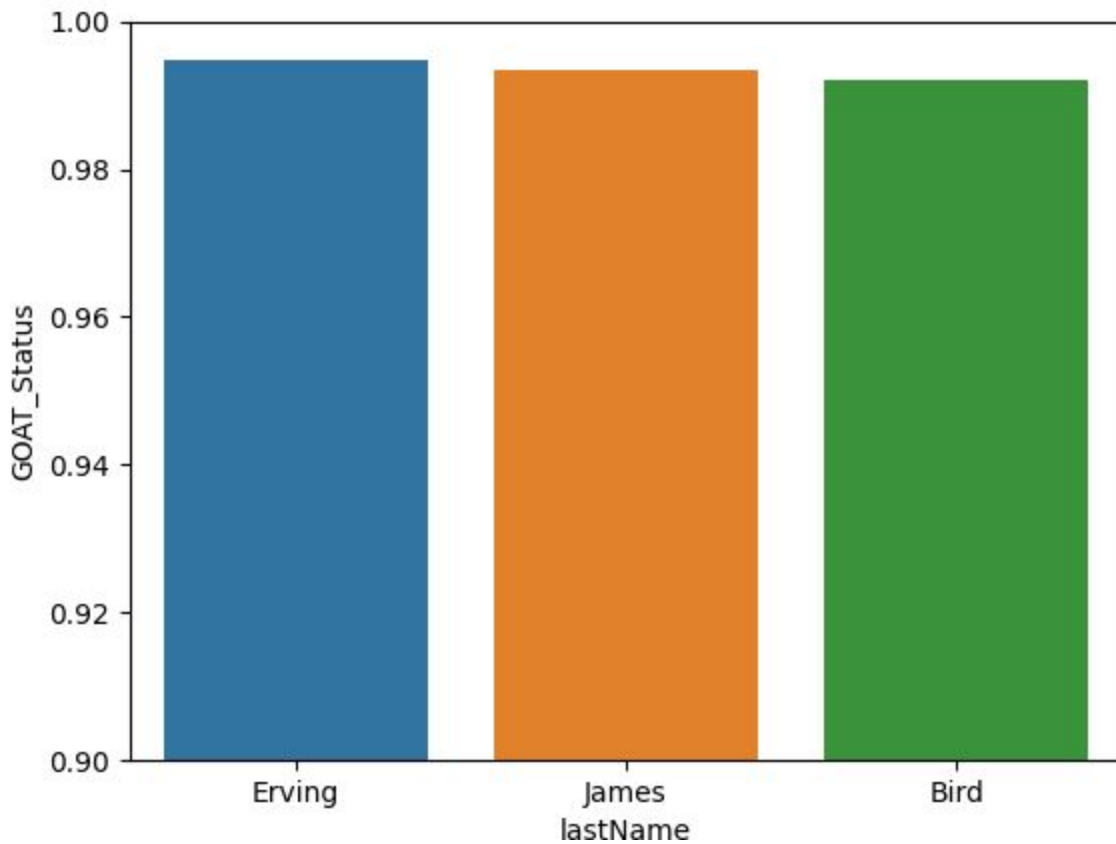
→ 3.2

The biographical data in this dataset contains information about home towns, home states, and home countries for these players. Can you find anything interesting about players who came from a similar location?

There does not seem to be any real significance regarding the particular state they are from, only that more heavily populated states produce the most players. Below are the results from the data and from wikipedia.

California	39,512,223
Texas	28,995,881
Florida	21,477,737
New York	19,453,561
Pennsylvania	12,801,989
Illinois	12,671,821
Ohio	11,689,100
Georgia	10,617,423
North Carolina	10,488,084
Michigan	9,986,857
New Jersey	8,882,190
Virginia	8,535,519
Washington	7,614,893
Arizona	7,278,717
Massachusetts	6,892,503
Tennessee	6,829,174
Indiana	6,732,219

CA	354
NY	348
IL	269
PA	215
OH	195
IN	160
MI	147
TX	144
NJ	117
GA	116
LA	108
KY	108
NC	107
FL	97

→ 3.3

Find something else in this dataset that you consider interesting. Produce a graph to communicate your insight.

I find it interesting that Chicago is the only birth city with more than 100 NBA players - topping out at 139. It is smaller than New York City. It is only ⅓ the size but has produced ⅓ more NBA players.

** Note: I never want to look at another basketball stat for as long as I live.