Jacob Jashinsky

Final Project Write Up

PART I - SPECIFIC ASSIGNMENTS

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.)

   Mean of points per season: 492.1

      Median of points per season: 329

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: 4029

      By: Wilt Chamberlain in year 1961

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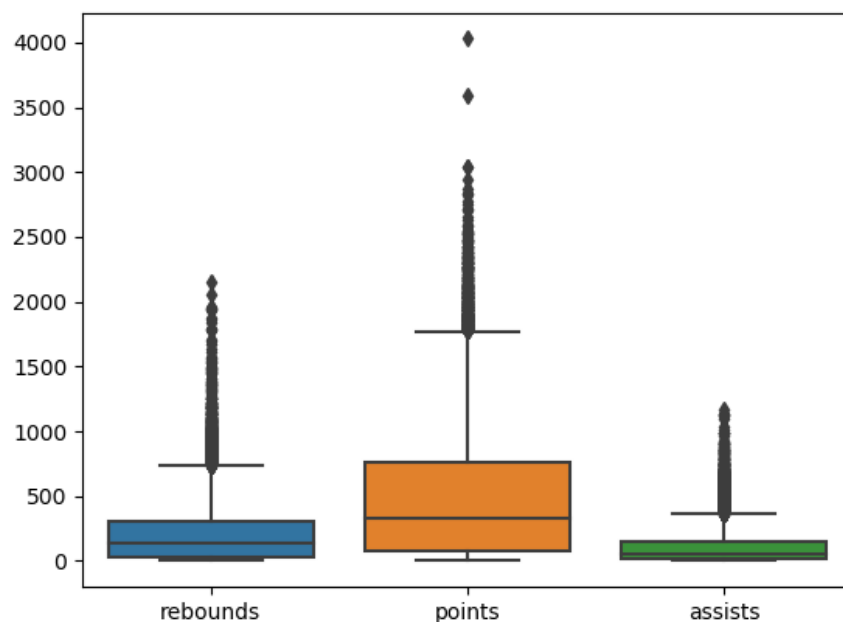
4. Produce a plot that shows how the number of points scored has changed over time by showing the median of points scored per year, over time. The x-axis is the year and the y-axis is the median number of points among all players for that year.

> To produce this plot I grouped by year and took the median of each year. We can see that the median is much lower on the beginning and but increased over time.
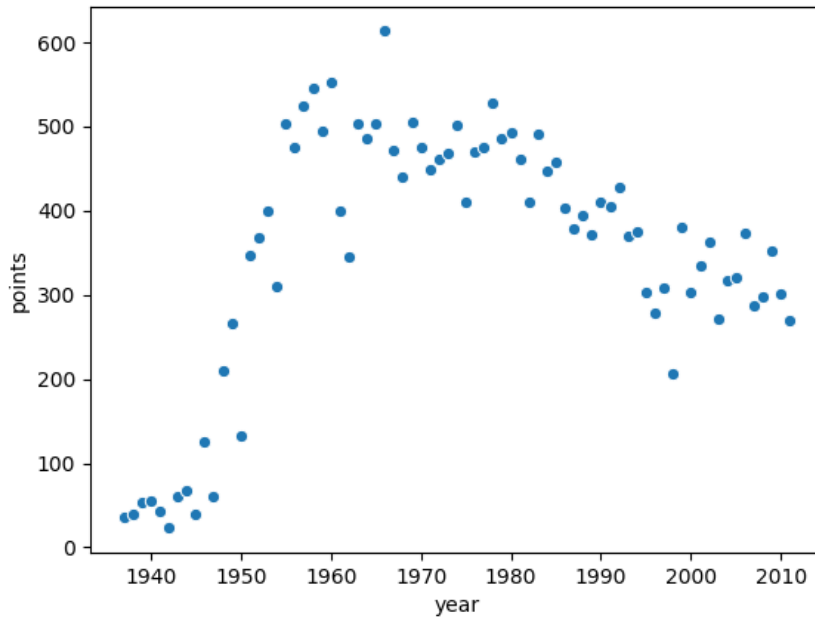
PART II - COME UP WITH SUPPORTING EVIDENCE

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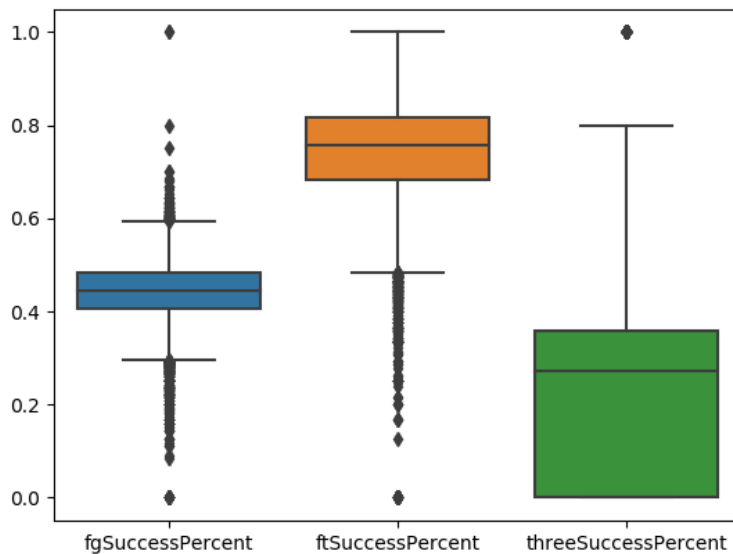
2. It seems like some players may excel in one statistical category, but produce very little in other areas. Are there any players that are exceptional across many categories?

> View the script file to see more of this and the data set. I filtered down to people who seemed to be the best at most categories. The top players after the filter are shown in the output.

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?

> To show the difference among the leagues I did a scatter plot and colored by the league, but it is difficult to see the whole distribution because of the density of the points.
>
> To counter this I showed the mean and the median.
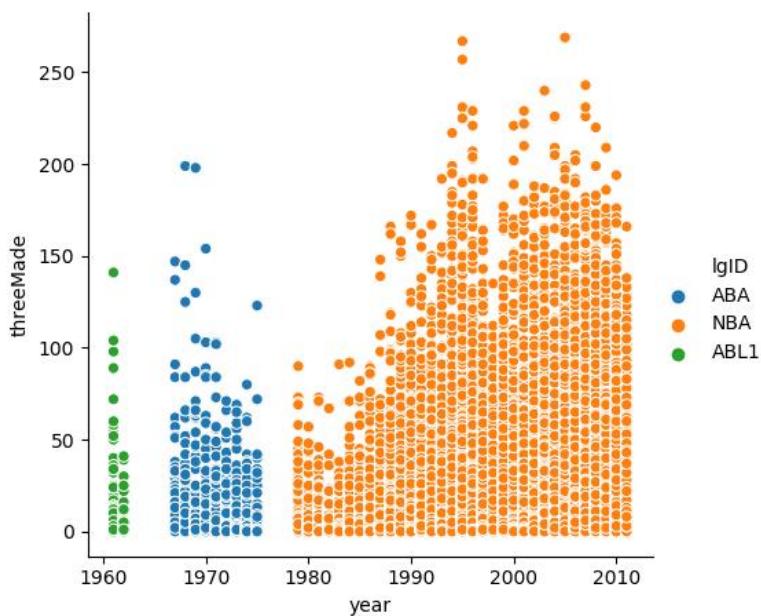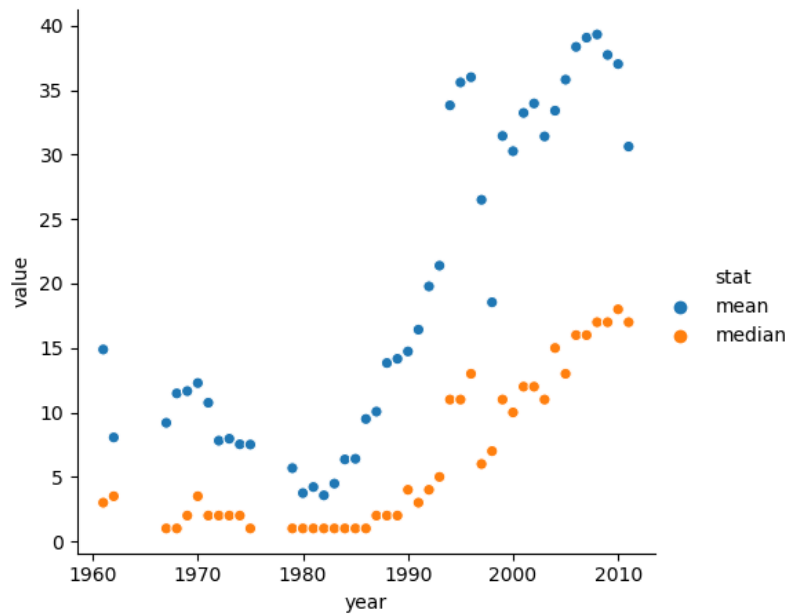
PART III - SHOW CREATIVITY

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.

My results revealed Michael Jordan to be the GOAT

The GOAT needs to have scored a lot of points. He needs great accuracy. He needs to be a team player, so he needs assists, rebounds blocks, and steals. I looked for the best well rounded

player. This player also needs to do really well, consistently. I did not sum up the players stats or find an average. The GOAT needs to be consistent, but a bad season or injury should not totally cut him out.

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?

I looked for a while but I couldn't find much interesting relations.

I looked at a few states and a city within CA.

I also looked at how many players come from each state and the mean number of points from players for each college.

You can view my script to see my more in depth search.

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

This first one correlated minutes players with games played and points. All three showed a strong positive correlation. This wasn't all the surprising to me.
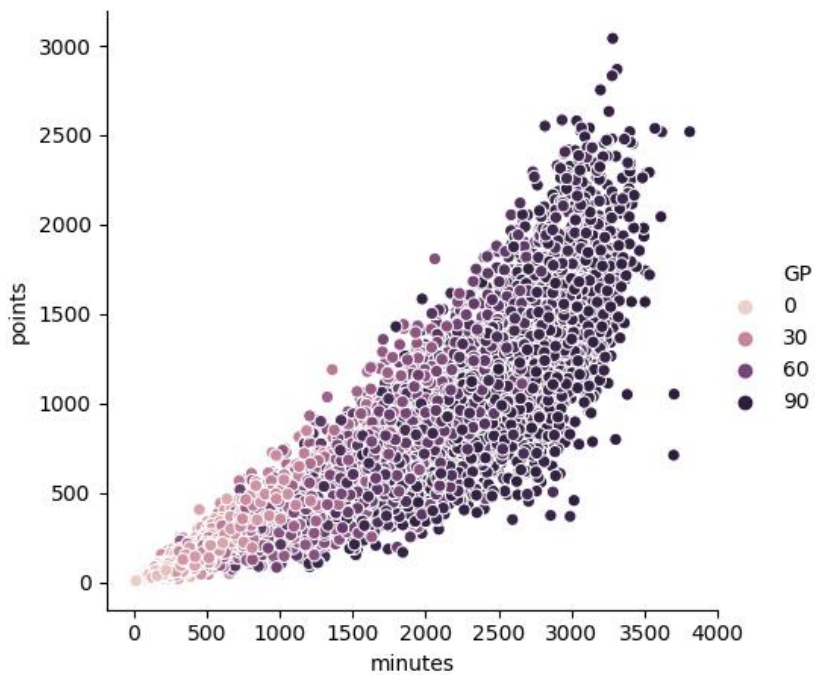
Next I filtered down to just the North Carolina college. I wanted to see if all students who come from the college have similar stats. It appears the students coming from NC range from decent to exceptional players.
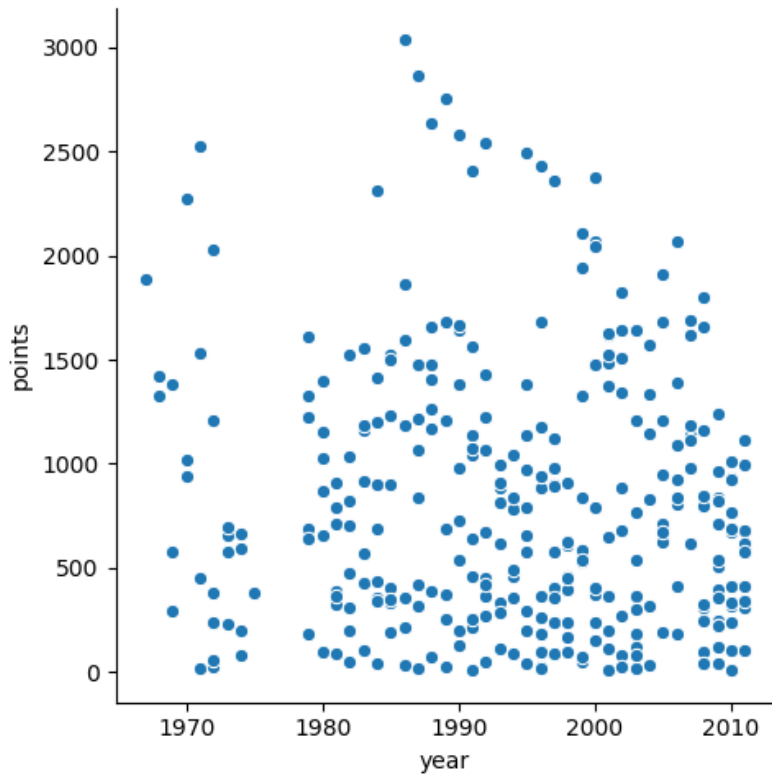
I noticed that Michael Jordan and some other players scored progressively worse as they aged. I wanted to see if this same pattern held for other players. I found a few top point scoring players throughout the years and created the line plot below.