Assignment 1

Advanced Methods in Applied statistics

Figure 1

Name: **Sascha Valentin Brown**

Ku. Id: Jbf383

Date of Submisson:14-02-2024

1

The main part of starting the first exercise is downloading the data. This is done in a Python script called 'data reader,' which creates two separate CSV files that are then loaded into the data analysis Python script file. A data frame is then created using Pandas, from which the conference teams are extracted. To run the script for different conferences, it is only necessary to specify the colors needed for the plot and which conferences.

In the data, there are different columns for AdjD, one of which is adjusted for the strength of the schedule and one which is not. I chose the one not adjusted for the strength of the schedule.

I chose to plot a histogram of the data. However, with five different conferences, the information to be extracted from this is limited, as it is hard to differentiate between different conferences by color. This is not improved with a scatter plot, as it also becomes confusing. I chose a bin width of 1.5 as this reduces the number of gaps between bins within the same conference. A lower bin width provided limited information, as the height of the bins did not exceed much greater than 1, and the different conferences were stacked on top of each other.

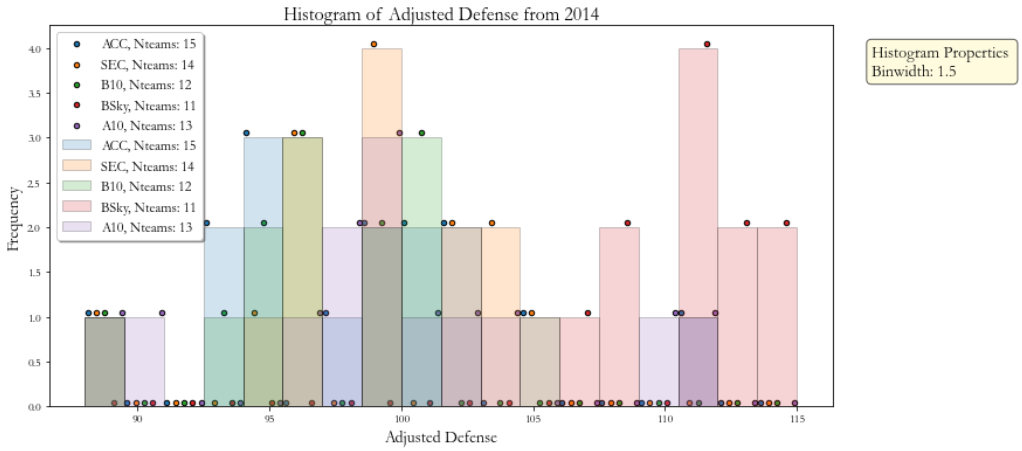

Figure 2: The dots represent the bar height to provide clarity on which conferences are within the bar. If more conferences are in a bar, the color is a mixture of them. The dots have an additional offset in the y-axis and x-axis

2

The next question is somewhat ambiguous. One interpretation involves examining teams that played in the five largest conferences in 2009 and comparing them to 2014, regardless of whether they remained in the same conference.

Another interpretation is that we should focus solely on teams that stayed in the same conference between 2009 and 2014 within the five major conferences.

I opted for the latter interpretation. As evidenced by the histogram in exercise one, particularly in the Bsky conference, there was a significantly higher adjusted defensive rating. This is likely correlated with a higher offensive rating, especially if teams primarily compete within their conference. Additionally, since we are plotting teams within each conference against each other, it's logical to examine developments within each conference. Furthermore, it's unclear which conference a team would belong to if it changed from one of the five conferences to another.

I sorted the data to isolate teams that remained within the same conference and calculated the difference in offensive rating for each of these teams as

$$AdjO_{2014} - AdjO_{2009}$$

Then, I determined the mean and median differences for each conference.

To calculate the mean and median of the difference in adjusted offensive rating outside the 5 conferences, I filtered out all teams that only appeared in the data in either 2009 or 2014. Next, I removed all teams that changed conferences to one of the five conferences during 2009 or 2014. This allows us to analyze the general change in offensive rating for the league, irrespective of conference changes. (Alternatively, one could focus solely on teams that remained in the same conference)

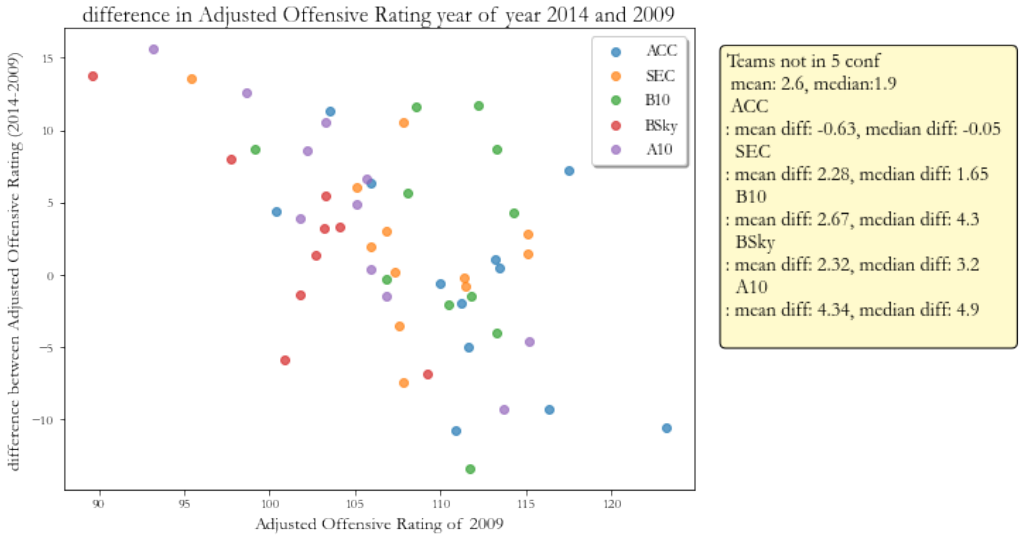

Figure 3: Caption

3

For 3 I followed the procedure as in exercises one and two, but added the BE conference to the 5 other conferences. To get this in the script one just has to remove 'BE' from the conference names.

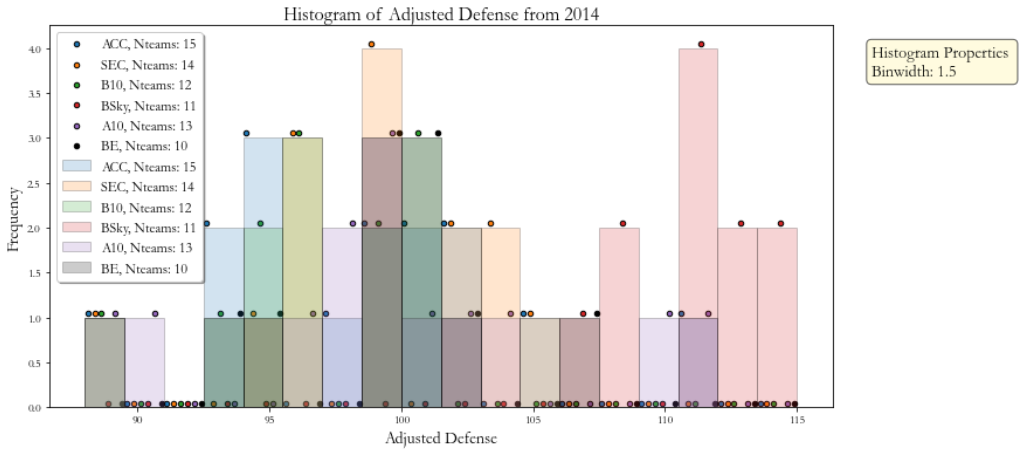

Figure 4: The dots represent the bar height to provide clarity on which conferences are within the bar. If more conferences are in a bar, the color is a mixture of them. The dots have an additional offset in the y-axis and x-axis

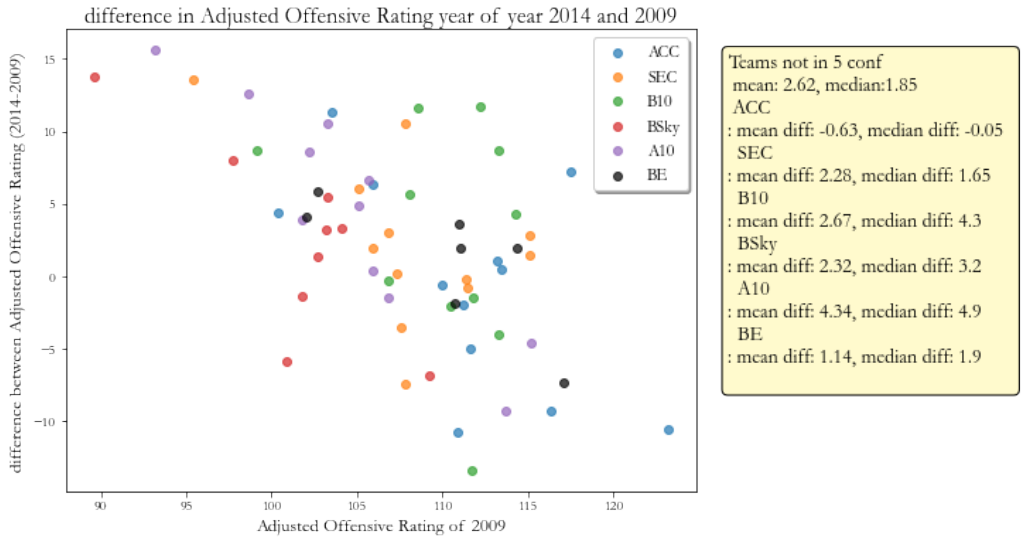

Figure 5: Caption