Clarissa Hoyt and Erin Little

Final Project Paper

Introduction

The goal of our study was to determine whether collegiate athletes youth sports experience has an effect on how much they still enjoy their sport in college. Since we are both student athletes at BU, this topic is very prevalent in our everyday lives and something we found very interesting. There is lots of research available about youth sports, and how much pressure young athletes feel to excel in hopes of competing at the collegiate level. It has been found that specializing in one activity from a younger age does not result in higher performance. Yet, that seems to be the trend we are seeing in youth sports today. As college athletes we have seen first hand how this behavior leads to athletes getting burnt out and causing them to dislike their sport once they reach college, or even before. We wanted to test this theory, and see whether the age college athletes started their sport and how much pressure their parents put on them as a child, has an impact on how much they still enjoy their sport in college. We hypothesized that athletes who started at a younger age and were put under more pressure would now dislike their sport.

Methodology

To conduct our experiment we created a google form with three simple questions: “How old were you when you started playing your sport?”, “On a scale of 1 to 5 how much pressure did you feel from your parents to excel at your sport?”, and “On a scale of 1 to 5 how much do you still like your sport?” For the last two questions, descriptions of each number on the scale were given. For example, for the second question, 1 was no pressure at all and 5 was tons of pressure. We sent this questionnaire out to college athletes of various sports at different schools, and received 93 responses.

After collecting the data we were able to convert the results into a csv file and upload it to GoogleCollab, and then turn the results into a data frame. From there we found the averages

using pivot tables and were able to create various graphs using matplotlib. We constructed bar graphs and histograms in order to display the overall spread of the data in a more visual way. We made 2 bar graphs to show the average enjoyment for each amount of pressure and each age. And created 3 histograms to show the frequency of each response to all three questions.

Using the data we collected, we performed various statistical tests to determine whether or not there is a significant correlation between the age athletes started their sport and how much they still enjoy it; and if there is a significant correlation between how much pressure parents put on athletes and how much they still enjoy their sport. A correlation coefficient quantifies the linear relationship between two variables. After imputing all our collected data into excel we were able to use the correlation function built into excel to determine the correlation coefficient for both relationships. If our hypothesis was correct we would expect a positive correlation between age and continued enjoyment and a negative correlation between pressure and continued enjoyment.

The second statistical analysis test we performed was a chi-square test. A chi-square test determines whether there is a relationship between two variables. The null hypothesis is that the two variables are independent. If the p value is less than 0.05, then there is evidence to reject the null hypothesis, meaning there is a relationship between the two variables, that it is not simply due to chance. We performed our chi-square test using the python libraries numpy and scipy.stats. We were looking to see if there was a relationship between athletes who started their sport at age 12 or younger and whether they now dislike their sport. The data showed 14 athletes who started at age 12 or younger and now dislike their sport, 41 athletes who started at age 12 or younger and now like their sport, 6 athletes who started after 12 and now dislike their sport, and 32 athletes who started after 12 and now like their sport. We also looked at the relationship between athletes who felt less than average pressure and whether they now like their sport. The data showed 57 athletes who experienced less than average pressure and now like their sport, 11 athletes who experienced less than average pressure and now dislike their

sport, 14 athletes who experienced more than average pressure and now like their sport, and 8 athletes who experienced more than average pressure and now dislike their sport.

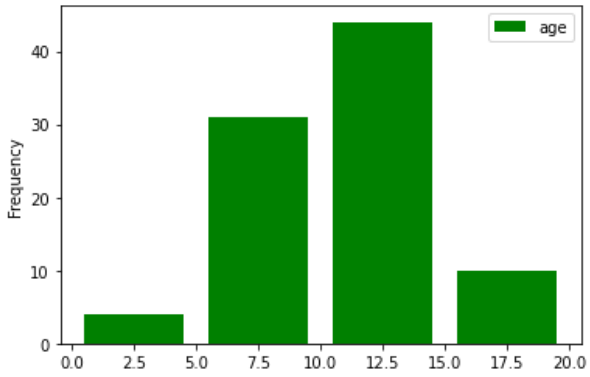
The last statistical tests we performed were the mean, median, and mode of our all three data sets: pressure amount, starting age, and enjoyment. We used the mean, median, and mode function of the data frame in python to find these values.

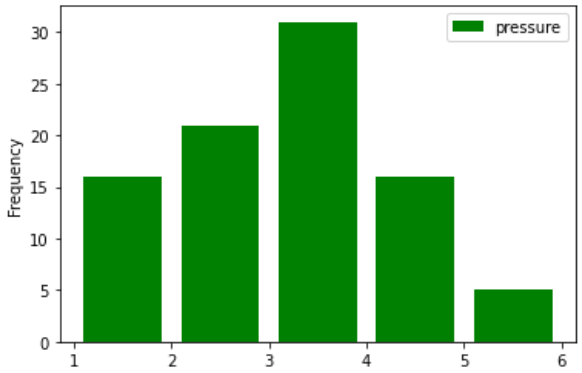
Results

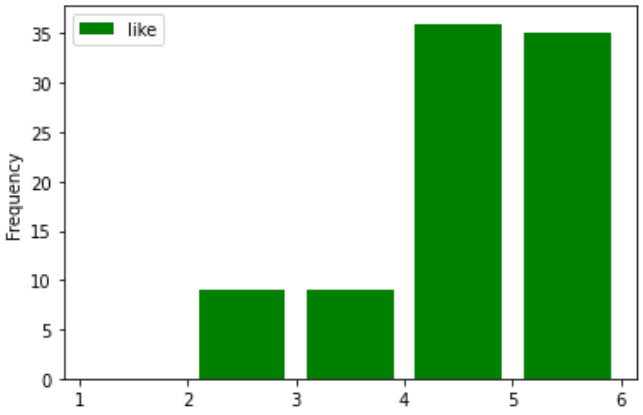
The statistical results include correlation coefficient, chi square test, mean, median, and mode. Our categories measured include amount of pressure, enjoyment level, and age at which an athlete started participating in their sport. Pressure and enjoyment level ranges on a scale from one to five and the age started ranged from three to 18. We found the means for the amount of pressure, enjoyment, and age for the sport. These held values 2.697, 4.09, and 10.629, respectively. The medians for these same categories are 3, 4, and 12. The modes for these categories are 3, 4, and 14. The first statistical analysis test we did was determining the correlation coefficient. The correlation coefficient between age and enjoyment was 0.115, which shows almost no correlation. The correlation coefficient between pressure and enjoyment was -0.178, which shows slightly more correlation but still basically none. We also did two chi square tests. The first calculation tested if there is a significant relationship between starting before age 12 and if you still like your sport. The p-value was 0.3906 which means the two variables are not related, because it is greater than 0.05. The second calculation tested if there is a significant relationship between experiencing below average pressure and still liking your sport. This p-value was 0.1 which also means this relationship is also not significant, because it is greater than 0.05.

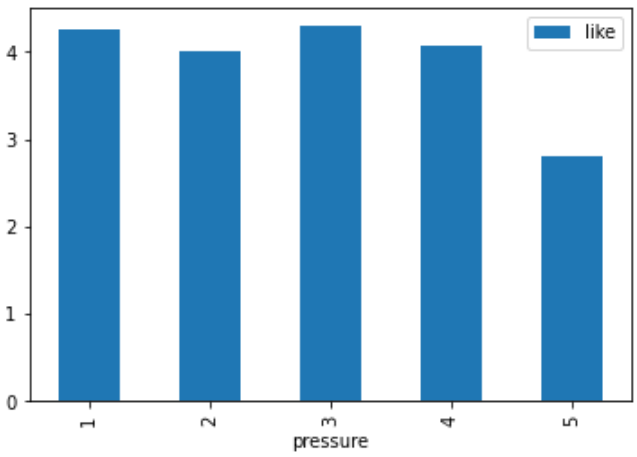
The visualizations include histograms and bar charts. There are three separate histograms to display the frequency of responses for age, pressure, and enjoyment. The bar charts display the average amount of like for each degree of pressure experienced and the

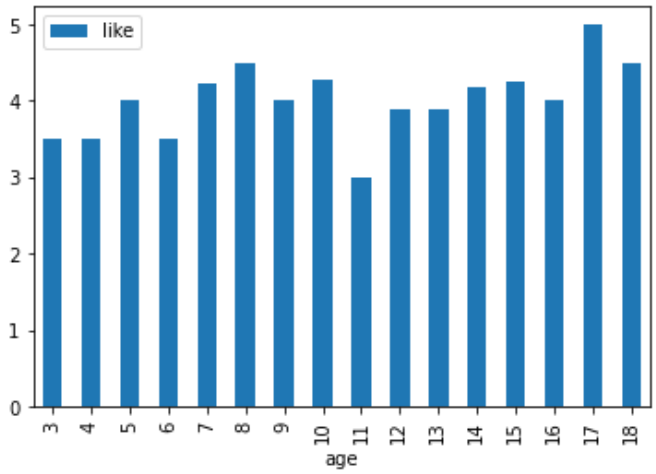
average amount of like for each age an athlete started their sport. Histograms are in green and bar graphs are in blue.











Conclusions

From the results, we conclude that there is no correlation between the age an athlete starts their sport or the amount of pressure experienced early in their career with the current likeness of the sport in collegiate athletes. Our statistical results of the chi square test and correlation coefficient show no correlation between these variables. Additionally, the result of the visualizations show no correlation by displaying an overall equal amount of likeness for the sport despite the age started and amount of pressure experienced.

After seeing the results, my team learned that giving options from one to five to rate an experience is too many options. This range is so broad it seems hard to differentiate between the options. This could be one reason why we found no correlation. Additionally, the data did not show much variation in answers. A much broader sample size can help with this challenge.