High School Marks of All-Americans in Division III Track & Field

(Revised March 2022)

Leo Pittsley, IEEE

**Abstract**—The nature of this project was to build a database of the High School marks of the All-Americans in NCAA Division III Track & Field dating back to 2016 and researched the data to learn what High School Track & Field marks have a good chance of becoming All-Americans in Division III in the future. Worked with the Head Coach of Track & Field at the University of Wisconsin- River Falls (UWRF) to create the guidelines of this project.

**Index Terms**—Data, Databases, Statistical Databases, Database Design, Database Management.

————————————————

1. Leo Pittsley is with the University of Wisconsin-River Falls, River Falls, WI 54022. E-mail: leo.pittsley@my.uwrf.edu.

—————————— u ——————————

# 1 Introduction

T

HIS project is an important recruiting tool for the UWRF Track & Field team to find High School Track & Field athletes that can compete at a high level in college at the Division III level. In NCAA Division III there are no athletic scholarships to draw in athletes, as Division I and Division II are allowed athletic scholarships. For the Division III level merely recruiting the top High School track athletes in your state/region/nation is not feasible. There are many more reasons Division III colleges will have a disadvantage to recruit the top athletes, such as better facilities, better coaches, more funding for the team, better competition, bigger spotlight, better NIL deals (Name, Image, Likeness), and more. This project was to find the best high school track and field athletes that would possibly compete at the Division III level and for the UWRF Track and Field team. Achieved this goal by finding the high school times and marks of past All-Americans gave the insight of what high school times and marks were at the level of past All-Americans.

The internet is full of databases of all sorts, but a database with high school times and marks of past All-Americans in Division III Track and Field was not one of them. For this project’s database would need to be created and finding who was an All-American since 2016 in Division III and finding their high schooltimes and marks. After the database is created and finished the process of analyzing the data began, using R as the program to analyze it.

# 2 Procedure for Collecting Data

## 2.1 Guidelines of Data

Datasets contain information about a certain thing, giving the necessary information. In this project the head coach of UWRF’s Track and Field team, Andrew Eggerth, helped to give what the necessary information would be needed for this project. It was decided that the name of the All-American, the year they were an All-American, the event they were All-American in, their college personal best (PR) in the specific event, the state the All-American is from, the high school they went to, their high school PR in the specific event, and the high school second best in the specific event, there are also columns for the high school best and second best in feet for those that do not use the metric system. Not all events in the Division III Outdoor Track and Field championship are also in all the states in the United States. Javelin is an event in college and in few US state high school meets, not all and was decided to take them out of the project. There are some events that can’t be compared at an individual level, such as relays and that would exclude the 4x100 meter relay and 4x400 meter relay from this project. There are events that are different from college to high school but were decided to use other high school events to compare for such college events. For Hammer throw the Discus, for 400-meter Hurdles the 300-meter Hurdles, and for 5,000-meter, 10,000-meter, and 3,000-meter Steeplechase the 3,200 meter.

## 2.2 Finding the All-Americans

To find the Outdoor Track and Field championship in Division III dating back to 2016 with the website TFRRS, except the year 2020 as the outdoor season was cancelled due to COVID-19. TFRRS gives the results of the Outdoor Championships by event and listed the athletes in order of finish, with the top 8 finishers for each event becoming All-Americans, except for certain circumstances where there are ties in the field events which can have more than eight athletes finishing in the top 8. All these athletes are All-Americans and were added to the project. By clicking on each name of the All-Americans it sends you to their bio on the website that lists the PR of all the events they did in college and that is how the college PR was inputted into the database for all the All-Americans. In the bios of the All-Americans on TFRRS there was also the name of the college/university the All-American attends/attended, this helped contribute to finding the All-American's high school information.

## 2.3 Finding High School Information

The websites found that have the high school information required for this project are Mile Split, Athletic.net, and Maxpreps. Athetic.net was choose as the website to use for the project since it did not require a subscription fee to obtain the information needed for the project. Athletic.net is simple to use but may be hard to find the correct person you are looking for. The names of the All-Americans may be common and finding the correct profile on Athletic.net, without knowing the high school they attended leads to problems. The method used to find what high school and state the All-American is from, was to look up their bio on their college rosters. TFRRS had the name of the college the All-American attended in the top right corner, making it feasible to look on the internet for the roster and bio of the All-Americans. The college bios would detail the high school and hometown of the athlete, now being able to use that information to find the correct profile on Athletic.net to find the high school marks of the given All-American. However, this did not work with every athlete since their full name or nick name might be used on Athletic.net, leading to their profile not showing up when searched. Since the information on what their high school's name was found through their college bio, now we can search for the high school in Athletic.net. Scrolling down the page of the high school of the All-American on Athletic.net, there is a list of past athletes that competed for the high school, by clicking on the name of the of the athlete needed it will go directly to the athlete’s profile on Athletic.net. Not all high school profiles were found on Athletic.net, as some All-Americans were from other countries, where it is rare for them to be on Athletic.net. There were two All-Americans where no information on where they were from or what high school they attended were found, their names are Sam Anderson and Josh Thorson.

## Finding High School Information for Multis

The information for Decathalon and Heptathlon All-Americans are different from each other and from the rest of the All-Americans and is why they have their own datasets. Decathalon men compete in ten events, not all these ten events are able to be compared to high school marks, such as Javelin, but 100-meter, 400-meter, 110-meter Hurdles, 1,500-meter, High Jump, Long Jump, Pole Vault, Shot Put, and Discus are compared in this project. The Heptathlon has seven events with Javelin being the only event able to be compared to high school marks, but 200-meter, 800-meter, 100-meter Hurdles, High Jump, Long Jump, and Shot Put are compared in this project. The process of finding the information is the same as other All-Americans by using Athetic.net. The database for the Decathalon will contain the high school PR marks of the events listed above minus Javelin, will also contain college PR marks for those same events, the state they are from, the high school they are from, and their PR Decathalon score. The database for the Heptathlon will contain the same as the Decathlon, with the events being slightly different, as they are listed above.

# 3 Structuring the Data

The Database for the All-Americans cover a vast amount of information and the database needs to be structured in a way that can be used properly in R. While collecting the data each event was given its own excel sheet, 100-meter men, 100-meter women and so on. If left this way there would be many small datasets, making it hard for a user to compare more than one event at a time. The process that was used to combine all the excel sheets into one single excel sheet, was make a column that listed the event, example is “M 100m” standing for 100-meter men, making it clear what event each All-American was a part of. Adding columns for the marks to be in feet for the field events, columns high school PR in feet and high school second-best in feet were created, since high school marks are generally listed in feet and inches. The field event marks were converted into meters and added the columns of high school best and high school second best accordingly, to be able to compare to the college PR and to also keep the “HS\_Best” and “HS\_2nd\_Best” columns as decimal format. The next step was to convert all the high school times to be the same, such as converting 1,500-meter times or 1 Mile times to 1,600-meter times and 3,000-meter times or 2 Mile times to 3,200-meter times. To convert these times the Running Conversion Calculator on the website Mile Split was used to convert the high school times and to convert the college PR times of 1,500-meter. The high school times of the 400-meter Hurdles was converted by multiplying the high school time by 0.75 to convert it to 300-meter Hurdle times, as there was no website on how to calculate the times another way. The “HS\_Best”, “HS\_2nd\_Best”, and “College\_PR” columns would only work in R if all the information was in decimal form, meaning that times that were over one minute needed to be converted into seconds. Then the final step was to combine the data into one excel sheet, to make one unified database. The same steps were also taken for the Decathlon and Heptathlon datasets to keep the same structure throughout all three datasets.

# 4 Analyzing the Datasets

## 4.1 Datasets

Decathlon and Heptathlon datasets will have similar things analyzed, analyzing the multi datasets to figure out what type of high schoolers would have a potential to be good multis in college. The All-American dataset will analyze a vast more variety of events and were analyzed differently at times. All datasets were looked at to see the states that had the most All-Americans, to learn where the hot spots are for recruiting. Other areas that were investigated were the average improvement from high school to college, the median of high school times per event, and more.

## 4.2 The States All-Americans are From

After analyzing all the datasets there was a state that was on top for all three datasets, this state was Wisconsin. The top three states in the Decathlon were Wisconsin with 11, New York with 4, and California, Minnesota, Ohio, Oregon, and Virginia all tied with 3. The top three states in the Heptathlon were Wisconsin with 9, New York with 8, and California with 5. Both Heptathlon and Decathlon had similar results with Wisconsin, New York, and California in the top three states for both. The All-American dataset has a larger sample size with All-Americans that are from outside of the United States, such as England, Israel, Germany, Sweden, Bahamas, Belgium, Canada, Jamacia, Ghana, Japan, Nigeria, and Saudi Arabia. The top three states in the All-American dataset were Wisconsin with 187, Illinois 180, and New York with 109. Wisconsin and Illinois are the top states to recruit from but there are other states that just missed the top three as you can see below in figure 1.

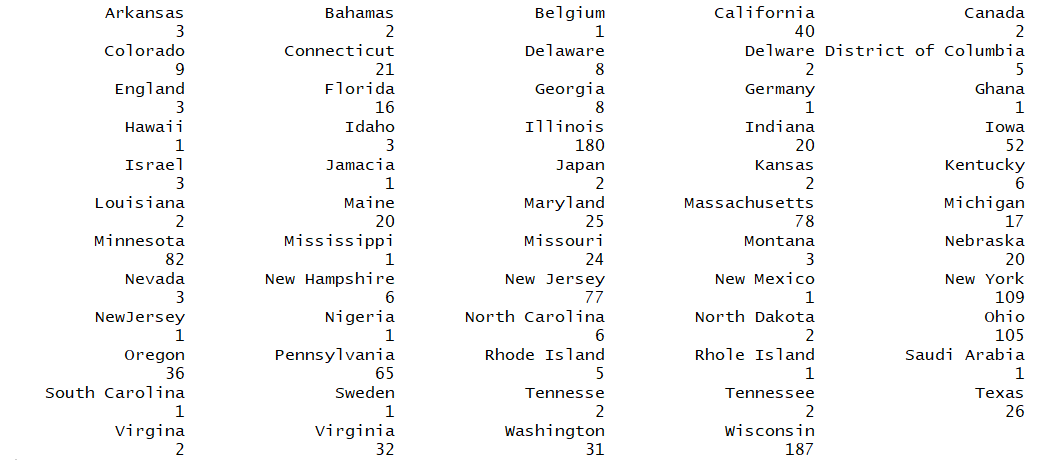


Fig. 1. Number of All-Americans from each state or country, note that two All-Americans are missing from this figure due to no information found on them.

## 4.3 All-Americans High School Times by Event

This area will break down the high school times by event showing the boxplot and the numbers of the max, min median, and those sorts of things. I am in the process of doing the coding for that and have not completed it yet. This is the part of the paper that will explain what I have found for each event, dividing it by men’s and women’s as well. I might break it down by paragraph for each event to break it up make it easier to read.

## 4.4 College PR vs High School PR

In this section I will compare the college PR with the HS PR to see the average improvement by each event, this also tells the reader on average how much you can expect an athlete to improve in college. Men’s and Women’s will get separate comparisons as they preform differently in track and field. I will not be able to compare the fallowing events due to the high school times and marks not being the same event, 5,000-meter, 10,000-meter, 400-meter Hurdles, 3,000-meter Steeplechase, and Hammer throw.

## 4.5 Decathlon

This is where I will write about the Decathlon dataset. I will compare the high school PRs by all the events the multis might have, since high schoolers might not have competed in every event that the Decathlon has. This will also show what the average improvement is for each event in the dataset and show which events have more improvement. I will also look at what events the decathlon athletes did in high school to see what high school events make it better to become a multi in college.

## 4.6 Heptathlon

This is where I will write about the Heptathlon dataset and do all the same analytical work that I did for the Decathlon athletes. I will have more figures through these sections to better show the data for the reader. Unfortunately, I do not have more data analysis for you as I ran into a few problems setting of the datasets, but now it should be clear sailing.

# 5 Acknowledgments

I would like to thank Andrew Eggerth the head coach at UWRF for bringing this project to my attention, and to sit down with me and work out the details of the project. From the information I should collect about the All-Americans to where I could look to find the information needed to complete this project. Mr. Eggerth also went through what events should be in the project, which events shouldn’t, and which events could be compared to other high school events. I’m thankful to give back to the team as I am an Alum of the team, and I am thankful for the experienced gained by completing this project.

# 6 Conclusion

Here would be the conclusion of the paper but as I have not yet completed my full analysis of the datasets it is not fair to make a conclusion of the project. I can say is that if give more time and resources on this project I could gave the information better, since I could call the athletic programs of colleges and high schools to find the PRs of the All-Americans' times and marks.

**References**

1. Athletic.net
2. TFRRS.org
3. <https://ny.milesplit.com/calc>