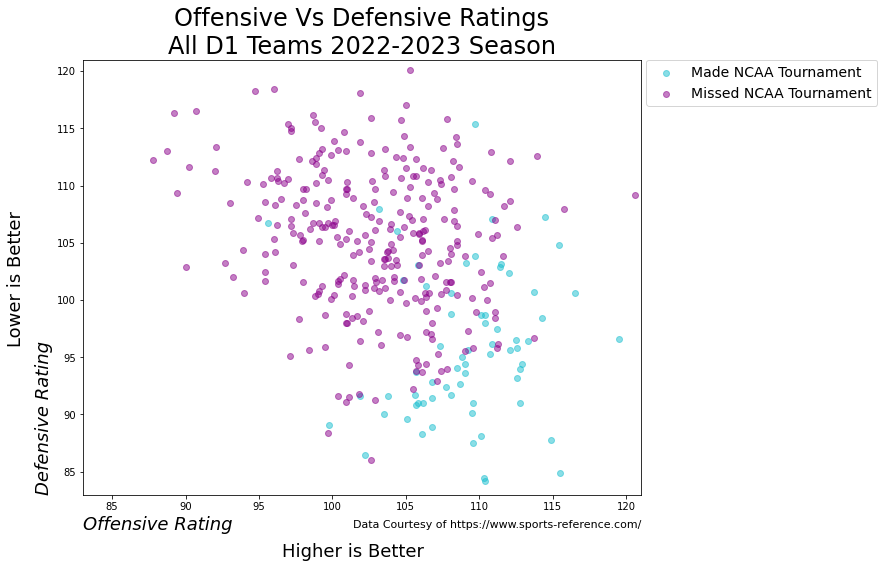
Have you ever wondered how someone made a graph that used pictures or logos instead of just plain boring dots? Well, I did, you see for those who do not know me I love college basketball, to those who do know me it is pretty obvious. Every year when college basketball season rolls around I see some kind of chart that depicts college basketball statistics with the team's logos on the scatter plot. I always found those graphs interesting. This year, curiosity got the best of me and I started researching how these graphs could be made. Lo and behold, I was extremely excited when I came across a [tutorial](https://fcpython.com/visualisation/creating-scatter-plots-with-club-logos-in-pytho) on how to make a scatterplot using English Premier League soccer teams using Matplotlib and Python. As a data analyst, I use Python a lot.

To start we do need some data, those of you who look at sports data a lot may be familiar with [Sports Reference](https://www.sports-reference.com/) a website meant to democratize sports data. The data I used to showcase this method of analysis is taking college basketball Offensive and Defensive Ratings from the 2022-2023 Regular Season. Those stats can be found [here](https://www.sports-reference.com/cbb/seasons/men/2023-ratings.html) on Sports Reference. Sports Reference defines the two metrics as:

* [Offensive Rating](https://www.sports-reference.com/cbb/about/glossary.html#srs) (ORtg): for schools, it is points scored per 100 possessions. The formula for schools is 100 \* (PTS / Poss).
* [Defensive Rating](https://www.sports-reference.com/cbb/about/glossary.html#srs) (DRtg): points allowed per 100 possessions. The formula for schools is 100 \* (Opp PTS / Poss).

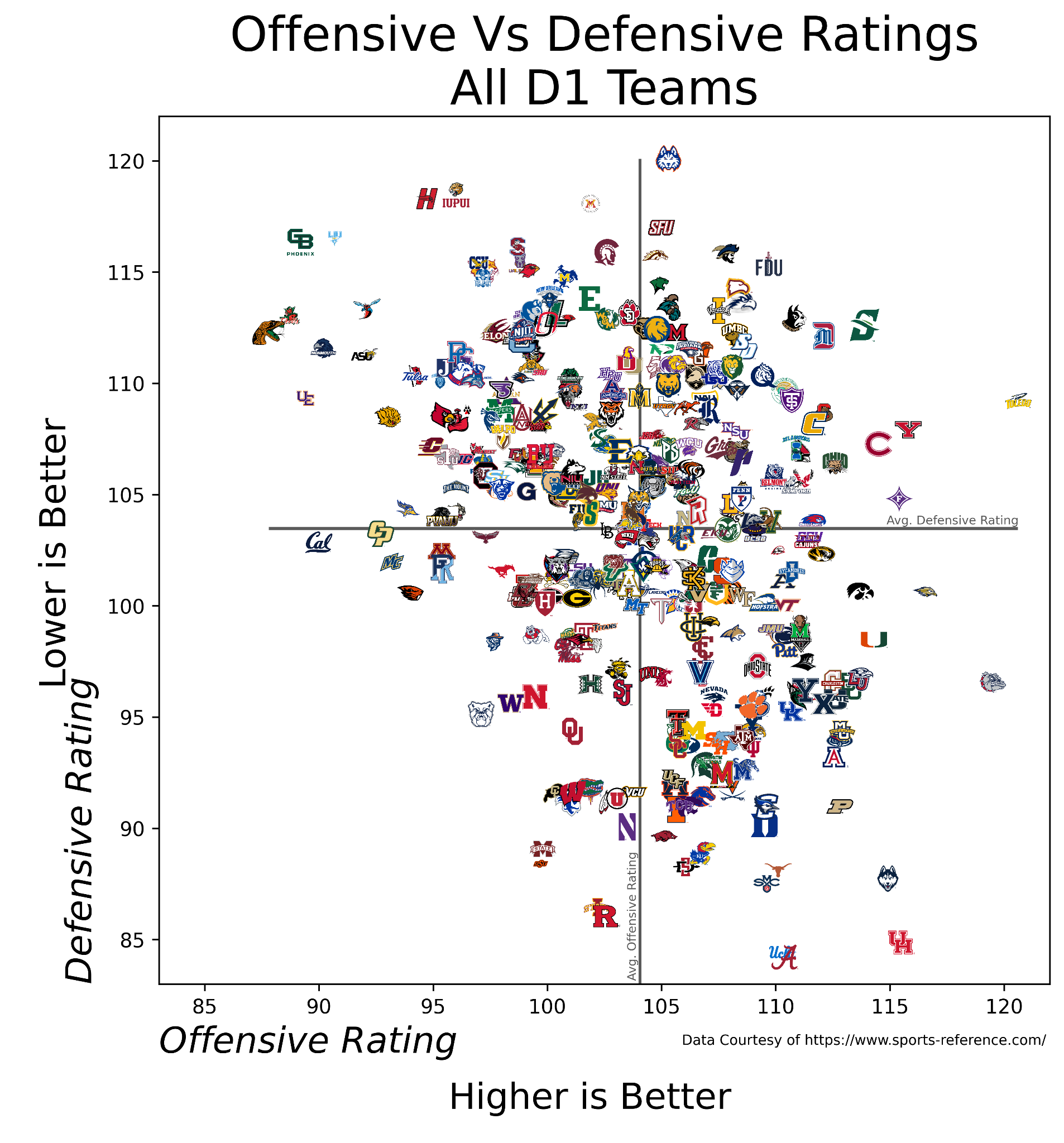
Using some matplotlib code we can generate a simple scatter plot:





This is your simple run-of-the-mill scatterplot. The colors indicate whether or not a basketball team made the coveted NCAA tournament. From the scatter plot, we can see that most of the teams that made the tournament have elite-level Defensive Ratings. We generated this plot using this code:

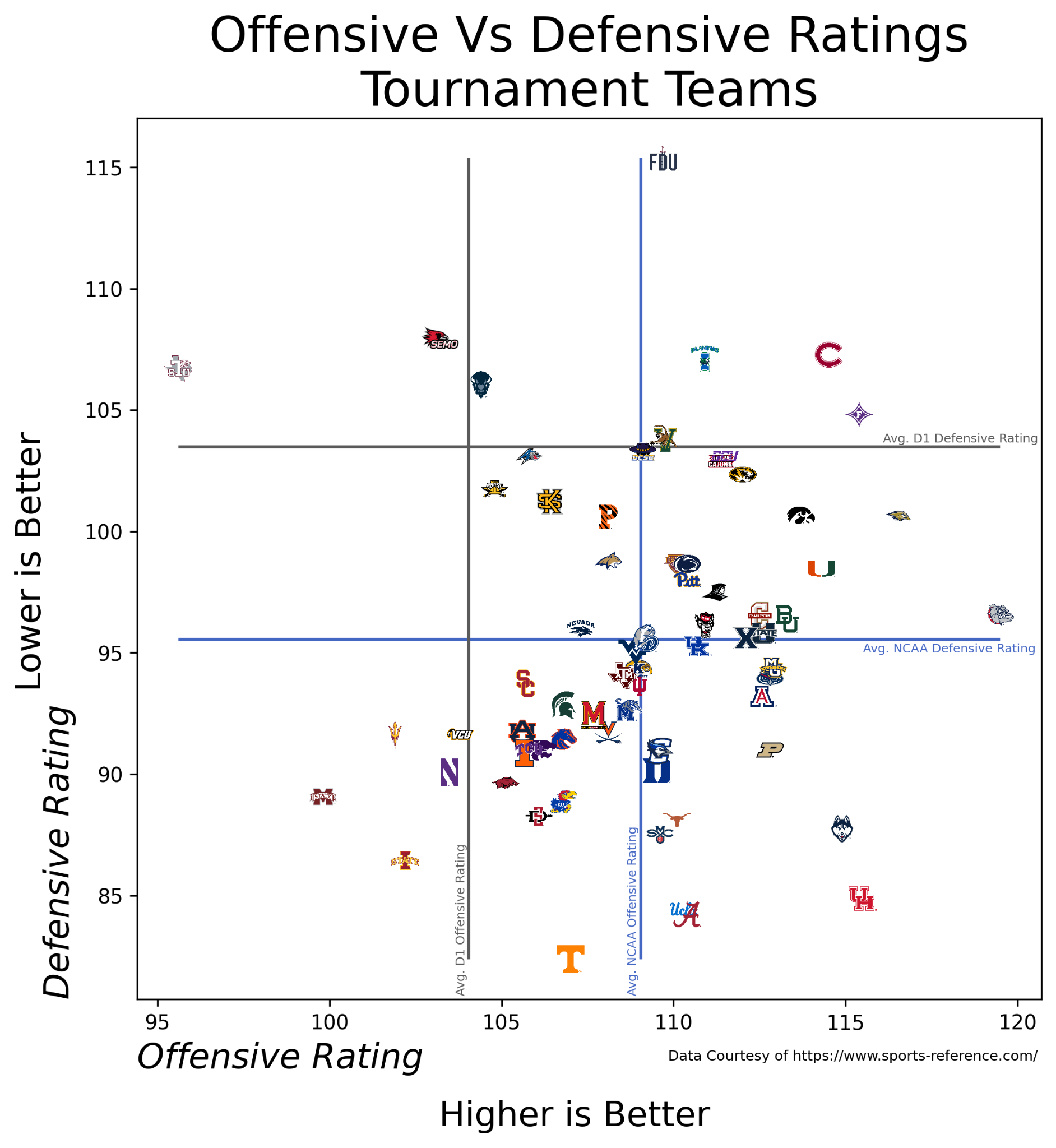


This is not why you clicked on this article though. Here is the scatterplot using college logos.

The logos are courtesy of [SportsLogos.net](https://www.sportslogos.net/) a self-[described](https://www.sportslogos.net/faq/): “virtual museum dedicated to the past, present, and future of graphic design in the world of sports.” This graph is generated by this code:



The two grey lines indicate the averages of defensive and offensive ratings throughout all of D1 Basketball. This means that we can split the teams into 4 quadrants: Bad Offense and Bad Defense (top left quadrant), Good Offense and Bad Defense (top right quadrant), Bad Offense and Good Defense (bottom left quadrant), and Good Offense and Good Defense (bottom right quadrant). This graph is a little messy because there are 363 logos for all the D1 teams. To get some insight from this visual, I filtered to just the 68 teams that made the NCAA Tournament.



In this visual, the grey lines still indicate the average Offensive and Defensive ratings for all of D1. We can see that most of the teams fall into the bottom right quadrant of good defense and good offense. You could also have a lot of different insights from this visual, for example, the teams that may not have deserved a bid based on their offensive ratings: Mississippi State, Iowa State, Northwestern, and Arizona State. I am also very happy to see my favorite team the national champions UConn in almost a league of their own with only Houston near them in Offense and Defense ratings. I am sure you have your own insights and questions from this visual, if so, please leave a response. Or some data you would like to see on the X and Y axis.

The Extra Effort for Logos

If you're not interested in delving into the intricacies behind the creation of this chart, feel free to skip this last part. For those still with me, well let's dive into the process in which I got the logos. Originally, I was going to try and scrape the images from Sports Reference where I got the data from. However, I quickly encountered a hurdle – the logos were not only small and low quality but also featured white backgrounds. After doing that I realized the logos were very small and low quality as well as had white backgrounds. From there I noticed that the logos were courtesy of Sportslogo.net. Realizing that the downloaded logos wouldn't have standardized names, I decided to manually name them, as scraping wouldn't save time in this scenario. From there most of the logos were of high quality. However, some of the logos had white backgrounds which makes it hard to view on a scatterplot because the white would cover up the logos behind it. I employed two methods for dealing with these white backgrounds. The first involved utilizing the remove algorithm in the rembg Python package, while the second utilized the Pillow library to eliminate all white pixels from the images. The dual approach was necessary, as for certain logos, removing all white pixels would inadvertently erase parts of the logo. Although the remove algorithm in rembg proved effective in many cases, there were instances where it didn't perform optimally. The remove background method can be shown here:



This is how I removed white pixels using Pillow:



Fair Use Disclaimer

I do not own the copyright to any of the school logos featured in this article and did not seek authorization from any of the schools to use the logos. However, I am using them under the Fair Use [Section 17 of the Copyright Act](https://www.copyright.gov/title17/92chap1.html#107). I believe this article falls into multiple fair use categories of teaching, research, news reporting, and comment.

<https://fcpython.com/visualisation/creating-scatter-plots-with-club-logos-in-python>

<https://www.sports-reference.com/>

<https://www.sportslogos.net/>