

2028 Project Proposal

Daniel Paré

A common question for tennis fans, especially ones of my generation is: who is better Rafael Nadal or Roger Federer? It is common for fans to try to cite statistics to prove that the player they prefer is better. The most common way to rank players is by the number of Grand Slams that they won. As a Nadal fan I think that there should be a better way to look at the rivalry. I suspect that simply because of the way Grand Slams are distributed that players of Roger Federer's playing style should be expected to win more Grand Slams than players like Rafael Nadal. In this assignment I would like to prove that this inequality exists by looking at data from players throughout history, and use it to provide context about both of their careers.

I will collect my data through two methods. The first will be with webscraping sites (via Python) that I can get mass data about grand slam results. The other method will involve more brute force and I will just have to go through player profiles and collect career results for players. I will try to display this data using descriptive statistics and analyze the numbers with methods we've discussed in class.

It's not just tennis that fans rank teams or players by the number of championships. For instance it is widely accepted that Joe Montana is the best quarterback mainly because he was able to win four super bowls. This makes some sense because theoretically you have to be among the best players or teams to win these titles, so it provides an easy number for a casual fan to look at. In team sports this assertion is not very strong because it takes more than one person to win, so it is possible the best player could end up on a bad team and never win a championship. In an individual sport that should be a good way rank players, but for tennis I still do not think it is that simple. In tennis there are four tournaments (slams) a year that every top player is striving to win. For the men the slams are a different format than regular tournaments because the match play is three out of five sets instead of the normal two out of three sets. In this drawn out format the matches are long enough that the result is more likely to reflect the better player. However, one problem is that all four slams are not the same. The four slams are played on three surfaces and each surface provides a different challenge. There is a lot of science and physics about why the surfaces are different, but the gist is that grass courts are "fast" with low bounces, clay courts are "slow" with high bounces, and hard courts are "fast" with high bounces. Generally grass court players depend heavily on their serve and aggressive shot making, while clay court players depend on their groundstrokes and defense. Either player can succeed on a hard court, but I suspect that a player with a grass court style of play is more likely to succeed on hard courts. If it is more likely for a grass court player to succeed on a hard court, then they would be expected to win more grand slams than a clay court player because there are simply more opportunities. I'm under the impression that this inequality is significant enough that one can argue that Nadal's accomplishment of 14 Grand Slams is more impressive than Federer's 17 slams.