

PS6

Bradley Rann

March 2023

1 Data Cleaning and Figures

I'm using data from my previous problem set regarding the champions from the Wimbledon tennis tournament. My first step was converting the data from HTML code into a data frame, first I had to parse the HTML then take the put the selected table into a data frame. Most of the data was in good shape, the dates were only by year and none of the data needed to be standardized to be more understandable. There was no competition in 2020 though due to COVID, so I removed that from my data frame. I also extended the max number of columns that could be seen after printing the pandas DF so that all of the data could be seen.

I then used the data to show three different graphs, a pie and bar chart of the champions by country and a bar chart showing the most successful champions at Wimbledon. The graphs clearly show easy information about who are the best grass tennis players and what countries have traditionally dominated tennis' most prestigious tournament. It makes it easy to compare by era and by country and summarizes the important aspects of the data set. Here are those images.

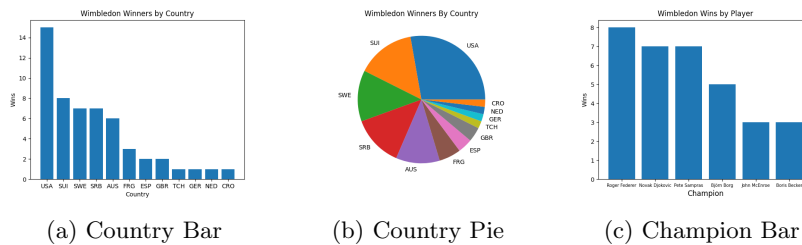


Figure 1: Wimbledon Champions