

PS5*Garcia*

jorgevladimir.garcia

February 2018

The first data set I collected corresponds to a table from Wikipedia with historical information about medal count by country from the Winter and Summer Olympics. Given that the event is currently undergoing in Pyeongchang, South Korea; I thought that it would be interesting to know what countries are leading the list. Although the data itself might not be useful, perhaps for news reports purposes, it was useful to learn how to gather data from Wikipedia sources and convert it into csv files only with the help of a short code. I did not see online tutorials to parse the data because the data set was in a table already.

On the other hand, I used data from twitter and used the wordcloud and tm packages to analyze what are the most frequent words related to the gun control debate, which I did by gathering the data from a twitter search. I followed different tutorials and videos but the results were rather disappointing since it was difficult to clean the data. Nonetheless, it was interesting to see what people are saying about this topic on the internet.