

In [562]:

output:

I created a single category barplot by calculating the proportion of html tags in spam emails versus that of ham emails. From the barplot below, we can see that the percentage of html tags in spam emails is more than double that of ham emails. This led me to conclude that html tags is a strong indicator for spam emails. However, contrary to my belief before, not all spam emails have html tags, and not all ham emails don't have html tags.



