

## Ex3 - Lior Ziv - 305742611

1. (a) The importance of this rule is to avoid stemming words which 'ss' is important at their end, such as 'sterss' which is not plural. This will prevent the s -> " to turn 'stress' into 'sters'.

(b) PONIES will turn into PONI(IES\rightarrow I), CIRCUS into CIRCU (S\rightarrow) , TIES into TI (IES \rightarrow I)

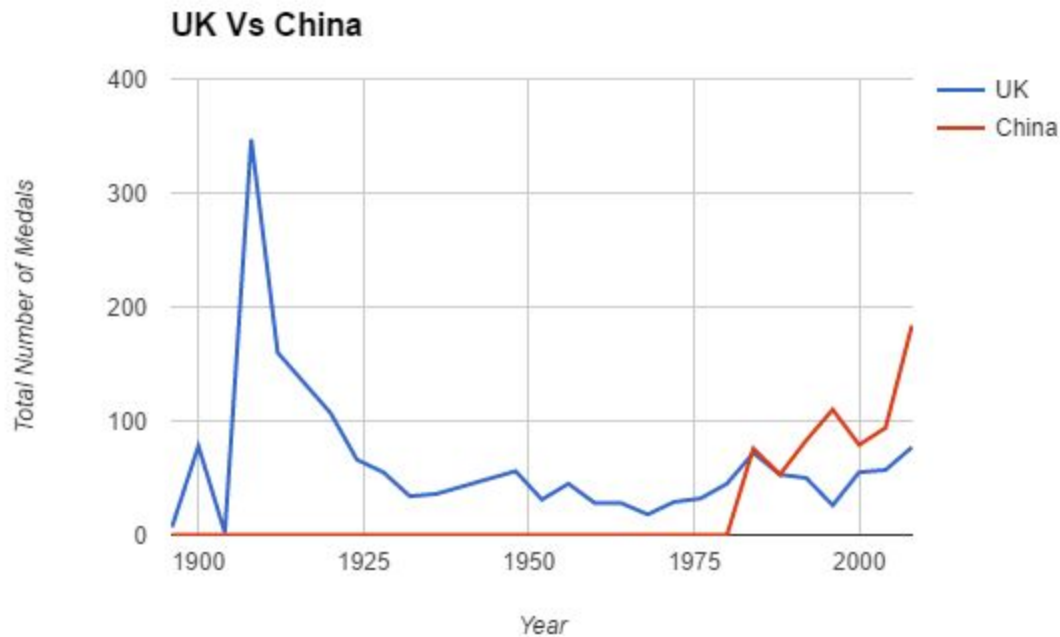
(c) CARESSES if we would stay with the original rules we will get CARESSE (S\rightarrow), in order to get the correct stemming we should add (ES\rightarrow) and then we get CARESS.

(d) True or false (and why): Stemming increases the size of the vocabulary , False. stemming will decrease the vocabulary size , since it's rules change words in a way that or a few words change to one word PONIES,PONI -> PONI , or that if a rule is suitable only for one word it changes it, which means it doesn't change the size of vocabulary(adds one and negates one).

(e) True of false (and why): In a search engine, if you stem the documents and the query, it never lowers precision. False, in the stemming process we might get to words that are not really connected to the connotation of the topic for example CARESSES (S\rightarrow) we will get CARESSE

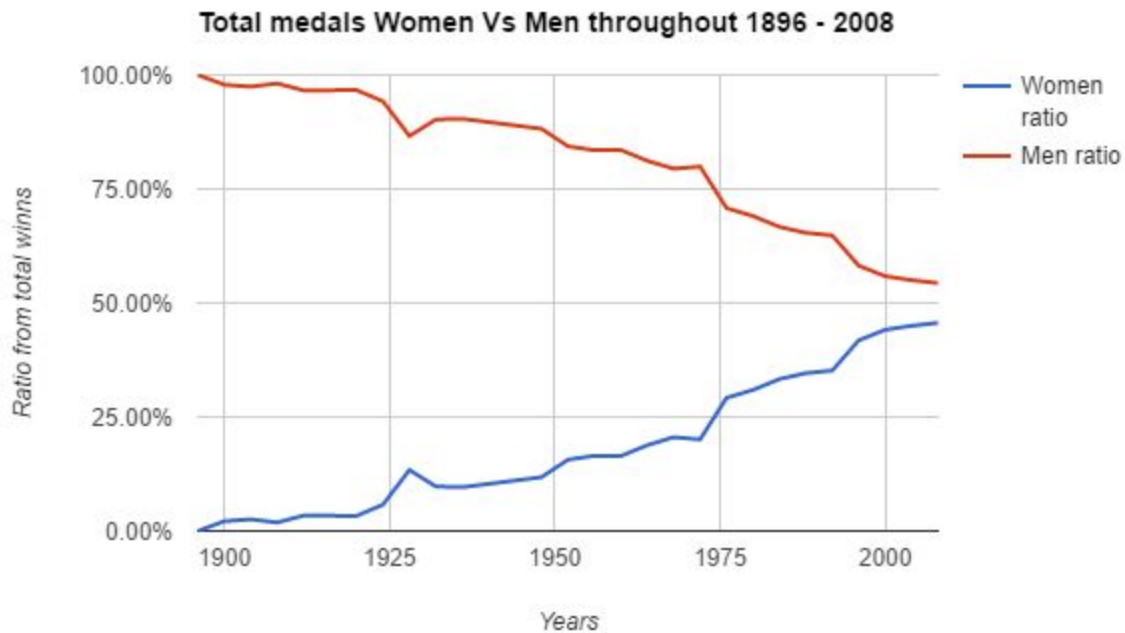
(f) True of false (and why): In a search engine, if you stem the documents and the query, it never lowers recall. True, since stemming only adds to the vocabulary we get that, if the wanted result was already found without stemming, than it will also be contained in all the results retrieved after the stemming.

## Explore the Intersection



Due to the dispute over the political status of China, the PRC did not participate in the Olympics again until the **1980**, as we can see in the graph. In that part we do see zero medals win, but since it started we can see it has a really steep improvement in oppose to Britain which generally doesn't improve in those year. We can see a big part of intersection in the years of 1984,1988.

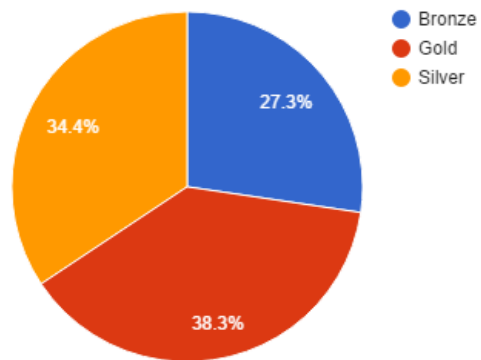
## Narrate Change over Time



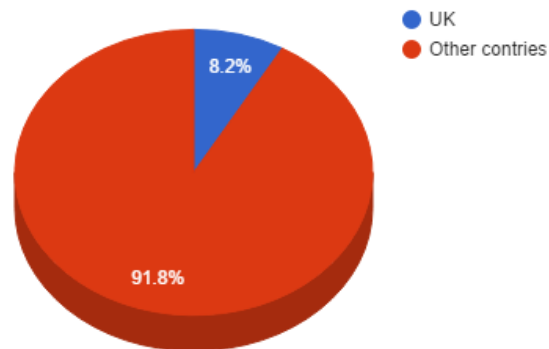
Women have competed in the Olympics since 1900, in the graph we can see very clearly that since 1896 there's been a constant rise in women winning medals percentage. This can be related directly to women's right, which got better and better throughout the years. It is interesting to see that it looks like we got to an equality point by now :)

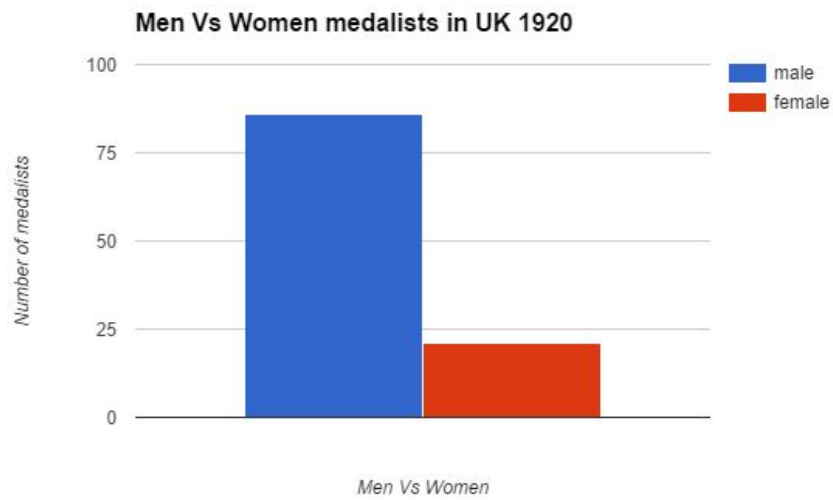
# Start Big and Drill Down

Total Gold, Silver and Bronze medals



Total Gold, Silver and Bronze medals in UK Vs all other contries





I started with the total number of medals from all kind won in 1920, then narrowed down to only medals of the UK in oppose to other countries, which then i narrowed to victories of men vs women.

3. (a) DRACULA by Bram Stoker

The code, pictures of the plot, and output files are attached.