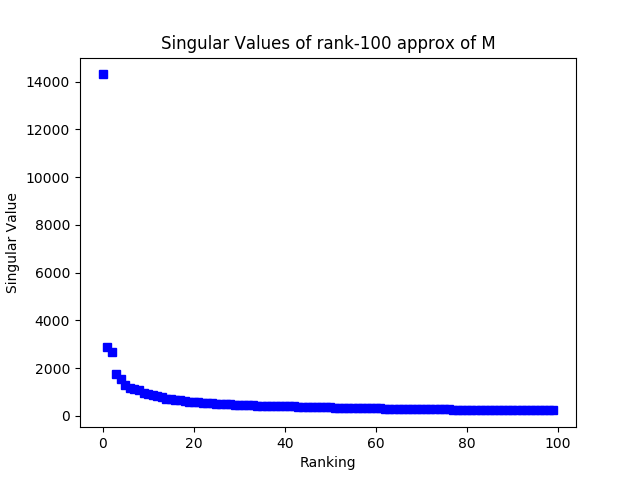
CS168 Project 5

1.

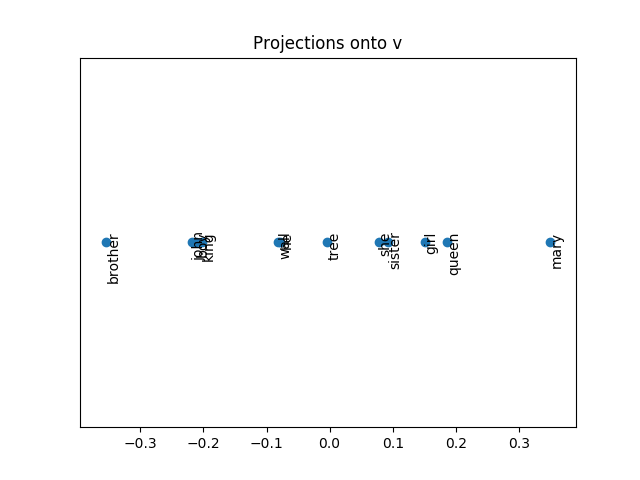
B.



Yes, M seems to be close to a low rank matrix because only the first few singular values are notably higher than the rest.

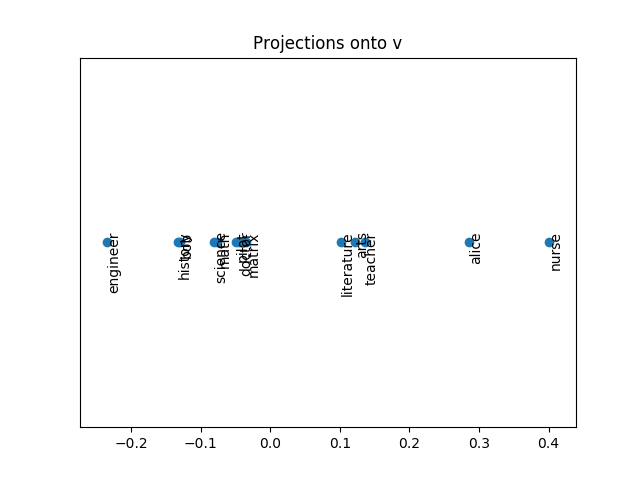
C. :s UNFINISHED

D.

i. 

The highest values are female-associated words, such as “Mary” (a typical name for a woman, “queen”, “girl”, “sister”, and “she”. The lowest values are male-associated words, such as “brother”, “John”, “boy”, and “king”. “Tree”, being a gender-neutral word is near a value of 0, and “wall” and “he” are tied slightly closer to the male side.

ii.



Professions typically associated with women, such as “nurse”, “teacher”, “arts”, and “literature”, have positive values while professions typically associated with men, such as “engineer”, “history”, “science”, “math”, and “doctor” have negative values; since these professions have historically been dominated by one gender or another, Wikipedia articles reflect this by containing more articles with men as doctors and women as nurses, for example. This could be harmful; for example, if LinkedIn used these embeddings to improve their “search for qualified job candidates” option, men would automatically be seen as more qualified for engineering and math positions while women would be seen as more qualified for nursing and teaching positions before any consideration for the rest of the parts of their resumes.

E.

i. The words most similar to “stanford” are other highly ranked schools; “harvard”, “cornell”, “ucla”, “yale”, “princeton”, “penn”, “auburn”, “mit”, and “berkeley”.

ii. The accuracy was 18.9%. UNFINISHED