Wesley Addo

QUESTION 1

Man		Woman	
Word	**Similarity**	**Word**	**Similarity**
man	1.0	woman	1.0
woman	0.5876938	child	0.5898086
child	0.33342195	man	0.5876938
doctor	0.28924733	husband	0.44964314
wife	0.2834791	birth	0.42030877
king	0.26449707	wife	0.30068845
husband	0.23411639	nurse	0.25435835
nurse	0.15348099	queen	0.22857241
birth	0.12343916	teacher	0.2040782
scientist	0.11226919	doctor	0.19613354
queen	0.110419504	scientist	0.13731061
professor	0.107622154	king	0.12252855
teacher	0.09874004	professor	0.105198584
president	0.094579265	president	0.084626846
engineer	0.08736357	engineer	0.044264376

QUESTION 2

File -> 3_Encyclopedic_semantics/E01 [country - capital].txt
Target -> [‘capital’]
Other word(s) -> [‘country’]
Protected class (race) words -> ‘black’, ‘white’, ‘asian’

capital	**country**	**similarity**	**black**	**white**	**asian**
abuja	nigeria	N/A	N/A	N/A	N/A
amman	jordan	0.324523	-0.016313	-0.091138	0.138067
ankara	turkey	0.295108	0.024293	-0.008582	0.00251
athens	greece	0.510705	-0.057965	-0.04915	0.152487
baghdad	iraq	0.530551	0.050204	0.021529	0.2039
bangkok	thailand	0.468903	-0.023679	-0.018846	0.250048
beijing	china	0.561427	0.003363	0.026926	0.294069
beirut	lebanon	0.584435	0.006655	-0.066131	0.197885
belgrade	serbia	0.601303	0.021249	-0.017178	0.140576
berlin	germany	0.553622	-0.003668	0.03581	0.131485
bern	switzerland	0.472947	-0.073037	-0.044791	0.105957
brussels	belgium	0.562507	-0.064016	-0.064179	0.146078
bucharest	romania	0.469891	-0.034262	-0.065296	0.151421
budapest	hungary	0.51704	-0.116664	-0.114084	0.11331
cairo	egypt	0.489359	-0.05014	-0.055538	0.134895
canberra	australia	0.343679	-0.099651	-0.063416	0.091348
conakry	guinea	0.525148	0.08642	0.082726	0.163075
copenhagen	denmark	0.446618	-0.063527	-0.11408	0.018329
damascus	syria	0.604804	0.008719	-0.040559	-0.008291
dhaka	bangladesh	0.453235	-0.008924	-0.028551	0.217926
dublin	ireland	0.437284	-0.008579	0.02734	-0.001614
hanoi	vietnam	0.33919	0.062865	-0.021177	0.089597
havana	cuba	0.484984	0.000952	-0.018983	0.034862
helsinki	finland	0.503703	-0.079526	-0.035748	0.159863
islamabad	pakistan	0.574659	-0.02763	-0.100167	0.196573
jakarta	indonesia	0.428815	0.024439	0.01373	0.219421
kabul	afghanistan	0.581063	-0.016735	-0.060399	0.14893
kiev	ukraine	0.480903	-0.063328	-0.12346	0.123904
kingston	jamaica	0.416547	-0.008017	0.007889	-0.005016
lima	peru	0.581407	0.015127	-0.034355	0.096506
lisbon	portugal	0.426096	-0.086989	-0.079879	0.071737
london	england uk britain	0.448425 0.29487267 0.2665021	0.023434	0.05715	0.083953

madrid	spain	0.389742	-0.127163	-0.117954	0.011963
manila	philippines	0.415015	0.033582	0.002149	0.093345
moscow	russia	0.462891	0.021692	0.092954	0.208812
nairobi	kenya	0.589144	0.052983	0.05188	0.292002
oslo	norway	0.416106	-0.063858	-0.085752	0.076945
ottawa	canada	0.507419	-0.046023	-0.023042	0.047322
paris	france	0.484897	-0.065842	-0.042998	-0.001612
rome	italy	0.44624	-0.061149	-0.069985	0.012023
santiago	chile	0.450533	-0.116187	-0.151828	-0.069263
sofia	bulgaria	0.30065	-0.02753	-0.04974	0.120579
stockholm	sweden	0.585047	-0.061972	-0.066089	0.028203
taipei	taiwan	0.65782	-0.012422	-0.003063	0.20021
tbilisi	georgia	0.483393	0.032708	-0.021468	0.242792
tehran	iran	0.440669	-0.116486	-0.120484	0.21861
tokyo	japan	0.478958	-0.025434	-0.019513	0.266332
vienna	austria	0.53641	-0.078941	0.014819	-0.082538
warsaw	poland	0.559183	-0.013961	-0.021814	0.234692
zagreb	croatia	0.498599	-0.109551	-0.07624	0.09361

Of the 50 capitals, only 8 (bolded in table) had all similarity scores positively correlated for the 3 words in the protected class. The remaining 42 had at least one of the 3 protected class words negatively correlated.

QUESTION 3

king is to throne as judge is to bench

a. newmodel.similarity(judge, bench) -> 0.30267334
b. newmodel.most_similar(positive=['judge','throne'], negative=['king'], topn=1) -> [('prosecution', 0.5186458230018616)]

giant is to dwarf as genius is to fool

a. newmodel.similarity(genius, fool) -> 0.2718964
b. newmodel.most_similar(positive=['genius','dwarf'], negative=['giant'], topn=1) -> [('theorist', 0.4280889332294464)]

college is to dean as jail is to warden

a. newmodel.similarity(jail, warden) -> 0.2777742

b. newmodel.most_similar(positive=['jail','dean'], negative=['college'], topn=1) -> [('peress', 0.5444425940513611)]

arc is to circle as line is to <u>square</u>

a. newmodel.similarity(line, square) -> 0.19263238
b. newmodel.most_similar(positive=['line','circle'], negative=['arc'], topn=1) -> [('lines', 0.4287526607513428)]

French is to France as dutch is to <u>netherlands</u>

a. newmodel.similarity(dutch, netherlands) -> 0.41922885
b. newmodel.most_similar(positive=['dutch','france'], negative=['french'], topn=1) -> [('netherlands', 0.6044681668281555)]

man is to woman as king is to <u>queen</u>

a. newmodel.similarity(king, queen) -> 0.5685571
b. newmodel.most_similar(positive=['king','woman'], negative=['man'], topn=1) -> [('queen', 0.5532454252243042)]

water is to ice as liquid is to <u>solid</u>

a. newmodel.similarity(liquid, solid) -> 0.65464735
b. newmodel.most_similar(positive=['liquid','ice'], negative=['water'], topn=1) -> [('solid', 0.4500039219856262)]

bad is to good as sad is to <u>happy</u>

a. newmodel.similarity(sad, happy) -> 0.44885093
b. newmodel.most_similar(positive=['sad','good'], negative=['bad'], topn=1) -> [('glory', 0.4403817057609558)]

nurse is to hospital as teacher is to <u>school</u>

a. newmodel.similarity(teacher, school) -> 0.5326568
b. newmodel.most_similar(positive=['teacher','hospital'], negative=['nurse'], topn=1) -> [('institution', 0.48289817571640015)]

usa is to pizza as japan is to <u>sushi</u>

a. newmodel.similarity(japan, sushi) -> 0.011866339
b. newmodel.most_similar(positive=['japan','pizza'], negative=['usa'], topn=1) -> [('dishes', 0.5763506889343262)]

human is to house as dog is to <u>kennel</u>

a. newmodel.similarity(dog, kennel) -> 0.28415978
b. newmodel.most_similar(positive=['dog','house'], negative=['human'], topn=1) -> [('hound', 0.423166424036026)]

grass is to green as sky is to <u>blue</u>

a. newmodel.similarity(sky, blue) -> 0.4439698
b. newmodel.most_similar(positive=['sky','green'], negative=['grass'], topn=1) -> [('blue', 0.5478643178939819)]

video is to cassette as computer is to memory
a. newmodel.similarity(computer, memory) -> 0.3897226
b. newmodel.most_similar(positive=['computer','cassette'], negative=['video'], topn=1) -> [('peripherals', 0.6654508113861084)]

universe is to planet as house is to hall
a. newmodel.similarity(house, hall) -> 0.31484395
b. newmodel.most_similar(positive=['house','planet'], negative=['universe'], topn=1) -> [('houses', 0.4264702796936035)]

poverty is to wealth as sickness is to health
a. newmodel.similarity(sickness, health) -> 0.19527602
b. newmodel.most_similar(positive=['sickness','wealth'], negative=['poverty'], topn=1) -> [('impious', 0.4960609972476959)]

C. Correlation -> 0.003031016735033952 (Very weak correlation)

QUESTION 4

Age group	0-20	21-40	41-60	61+	Total
Male	1941	901	914	616	4372
Female	2326	1632	751	698	5407
White	1930	1034	1252	1048	5264
Black	180	100	75	70	425
Asian	1017	349	88	99	1553
Indian	607	598	162	85	1452
Others	552	452	88	11	1103
Total	4286	2533	1665	1314	9798

Largest Representations by subgroup:
Age -> 0-20 - Female, White

21-40 - Female, White
41-60 - Male, White
60+ - Female, White

Gender -> Male - 0-20
Female - 0-20

Race -> White - 0-20
Black - 0-20
Asian - 0-20
Indian - 0-20
Others - 0-20

Least Representations by subgroup:
Age -> 0-20 - Male, Black
21-40 - Female, Black
41-60 - Male, Black
60+ - Male, Black

Gender -> Male - 61+
Female - 61+

Race -> White - 21-40
Black - 61+
Asian - 41-60
Indian - 61+
Others - 61+

Which group(s) will be most impacted?
I believe black people and old people, specifically 61+, will be impacted the most. The algorithm will not get training which is properly representative of the population. It does not seem to have enough black people or old (61+) people to train on and so may be biased in favor of other races and ages.