

1. Run “model1_for_uab.R” in code folder to make sure it works. (it uses demo file which I put in data folder)
2. Rename the pair file that you created by MINDFIRL software to “pairs.csv” and copy it to data folder. It should be like this:

```
ID,voter_reg_num,first_name,last_name,dob,sex,race,info1,info2,info3,info4,info5,type,file_id,cntfn,cntln,same
1,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
1,1000013655,CLARETTA,ADAMS,10/31/1944,F,B,o@gmail.com,77840,TX,9796542356,ADAMS,1,A15,6,6,0
2,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
2,1000013651,CLARETTA,ADAMS,10/31/1946,F,B,p@gmail.com,77840,TX,9796542356,ADAMS,1,A16,6,6,0
3,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
3,1000013651,CLARETTA,ADAMS,10/31/1945,F,B,q@gmail.com,77840,TX,6636500439,ADAMS,1,A17,6,6,0
4,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
4,1000013651,CLARETTA,ADAMS,10/31/1944,M,B,r@gmail.com,77064,TX,6636500439,ADAMS,1,A18,6,6,0
5,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
5,1000013651,CLARETTA,ADAMS,10/31/1944,F,W,s@gmail.com,77064,TX,6636500439,ADAMS,1,A19,6,6,0
6,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
6,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,B14,6,6,0
7,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
7,1000013655,CLARETTA,ADAMS,10/31/1944,F,B,o@gmail.com,77840,TX,9796542356,ADAMS,1,B15,6,6,0
8,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
8,1000013651,CLARETTA,ADAMS,10/31/1946,F,B,p@gmail.com,77840,TX,9796542356,ADAMS,1,B16,6,6,0
9,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
9,1000013651,CLARETTA,ADAMS,10/31/1945,F,B,q@gmail.com,77840,TX,6636500439,ADAMS,1,B17,6,6,0
10,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
10,1000013651,CLARETTA,ADAMS,10/31/1944,M,B,r@gmail.com,77064,TX,6636500439,ADAMS,1,B18,6,6,0
11,1000013651,CLARETTA,ADAMS,10/31/1944,F,B,n@gmail.com,77840,TX,9796542356,ADAMS,1,A14,6,6,0
11,1000013651,CLARETTA,ADAMS,10/31/1944,F,W,s@gmail.com,77064,TX,6636500439,ADAMS,1,B19,6,6,0
12,1000001971,JEAN EWART,ABERNATHY,08/27/1930,F,W,i@gmail.com,77064,CA,9796542356,Gorge Bush,1,A9,5,5,0
12,1000001972,JEAN EWART,ABERNATHY,08/28/1930,F,W,j@gmail.com,77064,CA,9796542356,Gorge Bush,1,A10,5,5,0
13,1000001971,JEAN EWART,ABERNATHY,08/27/1930,F,W,i@gmail.com,77064,CA,9796542356,Gorge Bush,1,A9,5,5,0
13,1000001972,JEAN EWART,ABERNATHY,08/28/1930,F,W,k@gmail.com,77064,CA,9796542356,Gorge Bush,1,A11,5,5,0
14,1000001971,JEAN EWART,ABERNATHY,08/27/1930,F,W,i@gmail.com,77064,CA,9796542356,Gorge Bush,1,A9,5,5,0
14,1000001971,JEAN EWART,ABERNATHY,08/27/1930,M,W,l@gmail.com,77064,CA,9796542356,Gorge Bush,1,A12,5,5,0
15,1000001971,JEAN EWART,ABERNATHY,08/27/1930,F,W,i@gmail.com,77064,CA,9796542356,Gorge Bush,1,A9,5,5,0
15,1000001976,JEAN EWART,ABERNATHY,08/27/1930,F,B,m@gmail.com,77064,CA,9796542356,Gorge Bush,1,A13,5,5,0
```

3. Rename your frequency files that you created for first name frequency and last name frequency to “ffreq.csv” and “lfreq.csv” and copy them to the data folder. They should look like:

a) ffreq.csv:

```
","fname","ffreq_raw"
1,lynn,10
2,jean ewart,2
3,claretta,5
```

b) lfreq.csv:

```
","lname","lfreq_raw"
1,WILDING,6
2,ABERNATHY,1
3,ADAMS,4
```

4. Now you can run “model1_for_uab.R” with your data.
5. After running the code, you can find the result in the result folder. Please report the number of pairs in these files (Do not send the files, we just need the number of

pairs):

- a) "df_test_final.csv"
- b) "review pairs.csv"

Note: in the first file each row is a pair: $\#pairs = \#row$
but in the second file each 2 rows are one pair: $\#pairs = \#row / 2$