Instructions for reproducing the analysis

There are several important things I do to the scraped data in order to get to the point of running the models. In the order they should be run:

-Perform sentiment analysis. This is actually done while the data are being scraped, so I’m not sure what you want me to do about that, I imagine it will be hard to reproduce, but if you run scrape\_debates\_FINAL.py for a couple of iterations then you should be able to see what the output is.

-LDA. This is in the file LDA.py

-Categorise MP names as male or female. This is in categorise\_names.py

-Categorise MP names into parties and run the final model. This is in file complete\_analysis.R (because I got sick of bashing my head repeatedly against Python at that point).

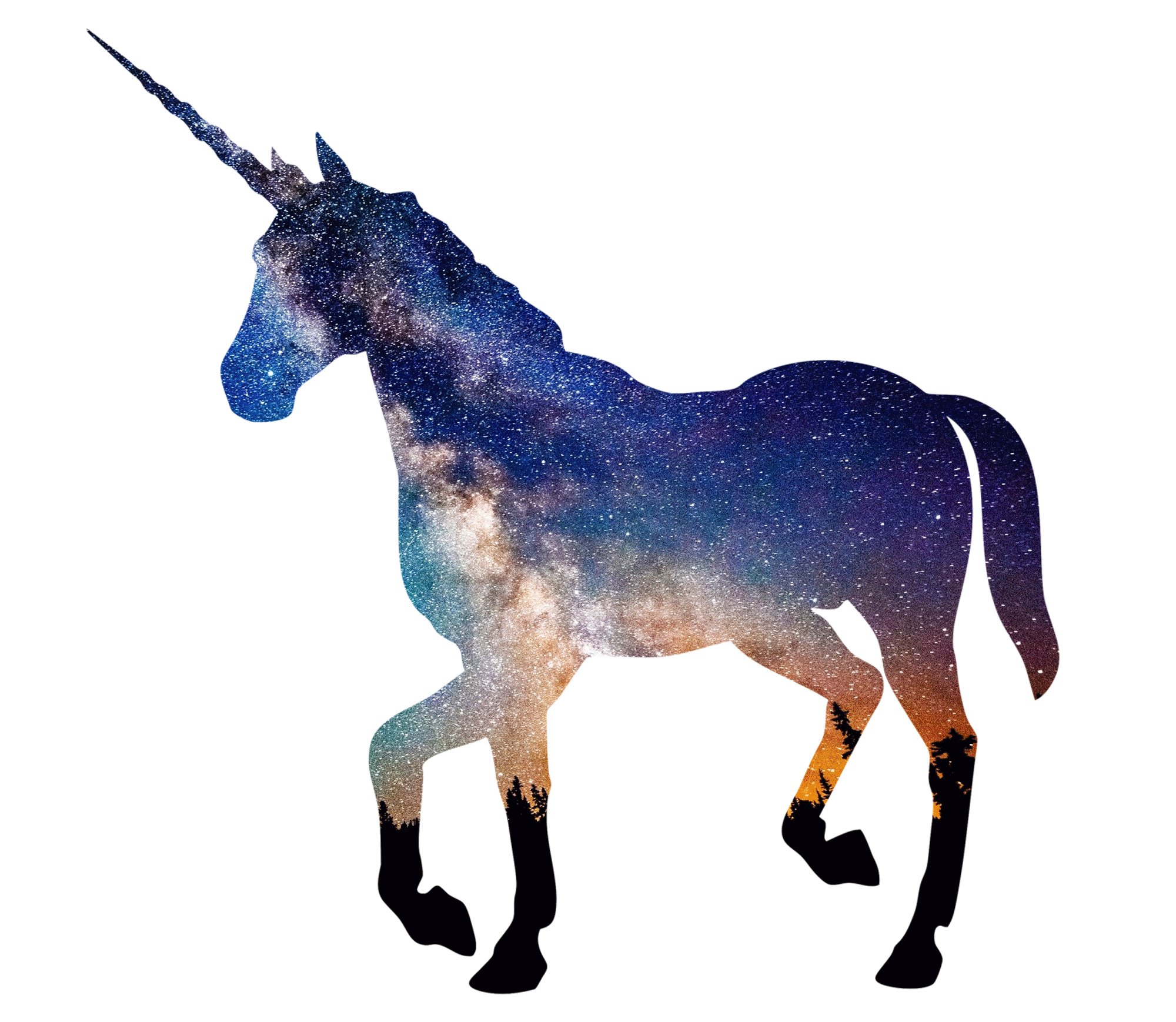
I’ve uploaded the following files to Box so that you can use them if need be:

harvard\_iv.csv, the sentiment analysis dictionary

mps.csv, which was used to categorise MPs by party

baby\_names.csv, which was used to categorise MPs by gender

Good luck with replicating all this.

––