# Replication Instructions for Estimation Uncertainty in Voter-Candidate Distances

The R-code for data analysis and plots can be found in two files: "Replication Ternary and Simulation code.R" and "ImputingCode.R"

Variable names:

- *drdistp* = physical distance between voter and candidate
- *aq16_* = party feeling thermometer, relabeled in some files as *partyfeeling*
- *votechoice* = binary indicating whether the individual voted for the candidate represented in that row
- *absdepdist* = absolute value of measure of socio-economic separation between the candidate and the voter
- *party vote*= indicator variable for which party the candidate is in: 1 = Labour, 2 = Conservative, 3 = Liberal Democrat
- *conservatives*= binary indicator of whether the candidate is a Conservative
- *liberal_democrats*=binary indicator of whether the candidate is a Liberal Democrat
- *outsideCon*= binary indicating whether the candidate lives outside the constituency that he/she lives in
- *inccand*=binary indicating whether the candidate is incumbent

The "Replication Ternary and Simulation Code.R" file contains code which will allow the user to replicate everything that can be found in the paper before the data imputation. It requires the data file "PGreplication.csv". This file is in 'long' form, so each individual voter occupies three rows in the dataset: 1 for each candidate they can choose. The R code enables the user to:

- Format the data
- Replicate two results from Arzheimer and Evans' original paper
- Simulate to account for estimation uncertainty
- Create ternary plots shown in Figure 1
- Create confidence interval plots shown in Figure 2
- Test the original results for model dependence[1]

[1] To test every scenario will require changing the input values on the counterfactual

To replicate the new graphs found in the paper section titled "Additional Improvements", use the R-code called "ImputingCode.R". This requires a different data file called "imputable.csv". This is the same data as used in the original analysis except that the inside/outside constituency variable has been added, and it has been changed to 'wide' rather than 'long' form in order to do the multiple imputation. In wide form, each individual occupies only a single row in the dataset because we need to impute the data based on the actual unit of analysis, the individual. For ease of access, the imputed data sets are provided here separately as Idata1.csv, Idata2.csv, Idata3.csv, Idata4.csv, and Idata5.csv, which are all converted back to long form. Note that some of the code has been recycled from the previous file such that the original and new results can be shown on the same graph, so this code will also require the file "PGreplication.csv". The code in this file will allow the user to:

- Use multiple imputation to eliminate missingness in the data
- Simulate conditional logic results using the new imputed data
- Combine the imputed data
- Make the new ternary plot in Figure 3
- Make the new confidence interval plots in Figure 4