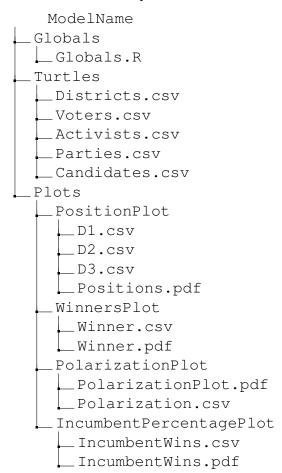# Problem Set 4

R Programming (Due Feb. 20)

Instructions

1. The following questions should each be answered forking the relevant GitHub repository. At each stage of development, you should 'commit' your code. Part of your grade will be how well you begin integrating version control, debugging, etc. into your work routine.

2. You may work in teams, but each student should develop their own R script. To be clear, there should be no copy and paste. Each keystroke in the assignment should be your own.

Reading in data without a clean format

1. he simulations for the 3-dimensional model were run in a program called NetLogo. Your job is to write a function that takes a *generic* NetLogo file as an input and writes out a directory where the various components of the output are written in an organized format.

2. The final directory should be structured as follows:

```
    ModelName
├──Globals
│  └─Globals.R
├──Turtles
│  ├─Districts.csv
│  ├─Voters.csv
│  ├─Activists.csv
│  ├─Parties.csv
│  └─Candidates.csv
└──Plots
   ├──PositionPlot
   │  ├─D1.csv
   │  ├─D2.csv
   │  ├─D3.csv
   │  └─Positions.pdf
   ├──WinnersPlot
   │  ├─Winner.csv
   │  └─Winner.pdf
   ├──PolarizationPlot
   │  ├─PolarizationPlot.pdf
   │  └─Polarization.csv
   └──IncumbentPercentagePlot
      ├─IncumbentWins.csv
      └─IncumbentWins.pdf
```

3. The final model output has three sections of interest to the researcher

   - The first item are the "Globals", which are either parameter settings used to initiate the model, or aggregate quantities of value as recorded in the final period of the simulation.

   - Second, there are traits associated with the various agents (Turtles) in the simulation as they were in the final period of the simulation.

   - Third, there are various "plots" that recorded quantities of interest (e.g., the mean position of "Blue" candidates) periodically throughout the simulation.

4. The Top level directory should be named such that it incorporates the name of the file used (e.g., `4JobTalk3.nlogo`) as well as the date and time the simulation was run.

5. The Turtles directory should contain sepearte csv files containing the `relevant` information for districts, voters, activists, candidates, and parties. For each file:

   - Column names should be meaningful.

   - Vectors should be broken down into separate columns.

   - Information that is constant or missing for all agents of the relevant breed should be dropped.

6. The first "Plot" records the average position of incumbent candidates (Red, Blue), activists, and voters along each of six dimensions. (D1, D2, D3, etc.) Only the first three are relevant here.

   - This directory should contain CSV files for each of the separate dimensions.

   - This directory should also include a PDF file plotting (in some creative, meaningful, and labeled way) how these quantities varied across the simulation.

7. The "Winners" Plot contains information about what percentage of candidates from each party "Won" in each cycle. The numbers may not add to 100. The "y" column shows the percentage, and the "x" column shows the time period. You should again make a CSV and PDF file for this information.

8. The polarization plot shows the Euclidean distance between the mean position of the candidates (TOTAL), voters, and activists associated with the two parties in each time period. The task is the same as before. (Try out some interesting features of plotting please).

9. The "Incumbent" output shows the percentage of incumbent candidates in each party that are "winning" in each time period. The assignment is the same as above.

10. Throughout the file there is extra information (mostly specifying how specific plots are to be displayed, the color they should be plotted etc. in the NetLogo interface).

Extra problems

- JMR Chapter 4, Problems 3 and 4

- JMR Chapter 7, problems 3 and 4