Political Data Science
Activity 7

With your group, figure out how to do the following

Write the R code to answer the following questions. Include the documentation.

Recent work in political science has proposed Benford's law as a method for identifying electoral fraud. The idea is that specific integer totals should appear in the *first significant digit* a known number of times if the data is being generated "naturally."

1) Calculating violations

Two ways of testing violations of Benford's law are proposed below. Let X_i represent the observed proportional frequency of the integer i in observed vote totals. So, for example, X_1 would represent the proportion vote totals where the integer 1 appears in the first significant digit.

- Leemis' m statistic

$$m = max_{i=1}^{9}\Big\{(X_i) - log_{10}(1 + 1/i)\Big\}$$

- Cho-Gains' d

$$d = \sqrt{\sum_{i=1}^{9}\Big((X_i) - log_{10}(1 + 1/i)\Big)^2}$$

Write a function to calculate these statistics. The function should take as an input (i) a matrix or vector of election returns and (ii) an option (or options) that controls whether the m statistic should be calculated the d statistic should be calculated or both. The output should be a list containing the results.

2) Critical values

For each statistic, we can reject the null hypothesis of *no fraud* if the statistic reaches the critical values in the table below.

	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
Leemis' m	0.851	0.967	1.212
Cho-Gains' d	1.212	1.330	1.569

Alter your function above so that it appropriately indicates the significance level of the resulting test statistics. You can provide this output in any way you like. Be creative.