# Quantitative Text Analysis: Assignment 8
# Wordscores

Kenneth Benoit

The objective of this class exercise is to better understand the Wordscores text scaling algorithm, using examples from Laver, Benoit and Garry (2003).

We recommend that you use R for this, along with the quanteda library, which contains the austin package.

To install the libraries you need, follow the instructions at https://github.com/kbenoit/quanteda under the 'How to Install' section.

We will use two sets of files:

- The example from Table 1 of LBG (2003). This data is built in to the `austin` library in R so if you are using R to do this, then you will not need to load this file in at all. If you are using some other software, then you can dowlonad the file LBGexample.csv. This file is in .csv (comma separated value) format and can be loaded directly into Stata or your spreadsheet.
- The Irish 2010 budget speeches data, available as the file http://www.kenbenoit.net/courses/essex2012cta/budget_2010.csv. The R file http://www.kenbenoit.net/courses/essex2012cta/exercise8.R has all of the commands you will need to implement the steps outlined below, including the installation of the austin library.

**Instructions:**

1. Start Rstudio and open the exercise8.R file in Rstudio
2. Install the `quanteda` library as per the instructions in the github website linked above.
3. Estimate the wordscores model for the LBG (2003) example, and inspect the results. Follow the code for precise instructions. Here you will be using the reference scores set at -1.50, -0.75, 0.00, 0.75, and 1.50 for reference texts $r_1$ through $r_5$ respectively. Score the virgin text and compare your results to LBG (2003) Table 1.
4. Run the wordscores scaling procedure on the Irish 2010 Budget speeches. Here we will use the 5th text (Cowen, the FF Prime Minister) as one reference text, and the 6th text (Kenny, the FG opposition leader). We will score all words in the Cowen and Kenny texts, and then score all texts as if they were purely "virgin" documents.
5. Inspect the word scores as shown in the .R file.
6. To run the naive Bayes example using amicus briefs (files here) , follow the instructions under the 'More Documentation' section on the github site. You will need to modify the first command to point to the location of the texts on your machine — on a Windows system, you will need to escape backslashes, e.g.:

```
amicus.texts <- c(getTextDir("C:\\amicuscuriae\\training"),
                  getTextDir("C:\\amicuscuriae\\testing"))
```