# PLS 300 Assignment 2: Text processing and Word Clouds

In Assignment 2 we will learn how to process textual data and prepare it for further analysis. We will use RStudio to create "Word Clouds" of text. Im sure you've seen these before. In doing so, we'll learn about 'data cleaning of text and the strengths and weaknesses of simple visualizations such as word clouds. If not, Google "word cloud" and youll see some examples and links to websites that will create word clouds for you. This assignment is intended to give you further practice preparing R analysis code.

Accompanying reading in *R for Dummies* for this assignment is Chapter 3. Please read it over and pay close attention to the description of R 'packages'. We have already used a package — the 'lattice' package for graphics — in Assignment 1.We will use R to create word clouds contrasting Presidents Obama, Reagan, and President Trumps Inaugural addresses.

After working through the Assignment 2.R script file examples, and watching the video on YouTube. Do the following: Find the text of President Obama's first inaugural address in 2009. (Hint: The American Presidency Project contains all kinds of info on presidents. Here's the link to it: `http://www.presidency.ucsb.edu/ws/index.php?pid=44`.)

1. Save the text of the speech into a text file, in the same format as the 'reaganspeech.txt file. (Use copy and paste to paste the text from the HTML page into the text file.) Text files are computer files containing only plain ASCII text, identified with a .txt file extension. Any word processor can create a text file. Apps like Note pad (on Windows) and TextEdit (on Mac) are intended for text files.

2. Create, or edit the script file to read in the text file, clean the text, and create a word cloud of Obama's inaugural address. Be sure to use the tm() package functions to process the text, as I show you how to do in the Assignment 2.R script file. To your journal, post

   (a) the text file itself containing the Obama address (upload it as an attachment or copy paste it as a journal entry);

   (b) post the tm() package excerpts from your script file that read in the text file and clean it, and

   (c) post the word cloud function to create the word cloud , and also post the resulting word cloud graphic.

3. Discuss the differences between the clouds; any noticeable differences in themes, illustrated by differences in word frequencies?

4. While interesting visualizations, word clouds have limits. As visualizations of text (also known as "corpora"), what are the strengths and weaknesses of word clouds?

5. Extra Credit question. For extra credit, using R for Dummies as a reference, construct a bar plot of word frequencies, after cleaning the text and preparing it for a word cloud. A hint for doing this question will be posted later.

6. Extra Extra Credit. For extra credit, construct a Comparison Cloud of Clinton and Trumps speeches. Youll need to use your Google-fu, but I will post a hint later.