Introduction to Machine Learning and Text Mining for Trade Policy
Problem: Lecture 2.

Ben Shepherd, Principal.

1. Amend the demonstration script to use XML files numbers 3 and 7. Both are trade agreements signed by Chile, with Japan and China respectively. We are interested in knowing to what extent a small country signing agreements with much larger ones is subject to common language, or different language.
2. Using the demonstration script as a guide, conduct an exploratory text analysis of both PTAs. At a minimum, you should look at word counts and frequencies.
3. Compare frequency of word use between the two agreements using a scatter plot. Which terms stand out as being more frequent in one agreement than the other?
4. Is there anything you can conclude from this analysis in a policy sense?