Problem description:

1. Get 40 more CNN news articles online (try to find them in different categories, e.g., sports and finance). Record the category of each article.
   1. Find software- no copy pastarino!
   2. Keep URL
   3. Make sure DM has a header
2. Convert all 50 CNN news articles to data matrix (each row is an article and each column is a unique term).
3. Run K-means clustering with Euclidean, Cosine and Jarcard similarity. (Specify K as the number of categories of your 50 CNN news articles)
   1. Specify k <- # of categories
4. Evaluate K-means clustering results with SSE

Deadline: pls submit your results by April 12 . You may compress your 50 news articles, data matrix, K-means results and your own discussion into a zip file to submit.