Quiz 8

1. **(6 Pts.)** Follow the bullet point directions from top to bottom to create a document by term matrix.

- Stem all nouns to singular form
- Remove all stop words
- Remove all terms that occur in only one document
- Create a document by term matrix with terms as columns and documents as rows

**Stoplist:** a about at but for is it me than thing to was you

**Document 1:** to err is human, but to really foul things up you need a computer
**Document 2:** computer science is no more about computers than astronomy is about telescopes
**Document 3:** a computer once beat me at chess, but it was no match for me at kick boxing

2. **(4 pts.)** Consider another unrelated term by document matrix ***A***. ***A*** has ***N*** rows which represent terms and ***p*** columns which represent documents. (We say ***A*** is an ***N*** x ***p*** matrix.) We use singular value decomposition (SVD) to extract ***k*** SVD features from ***A***. Given that SVD follows the well-known equation:

$$A = U\Sigma V^T$$

where ***U*** is an ***N*** x ***k*** matrix and $\boldsymbol{V^T}$ is an ***k*** x ***p*** matrix.

Is ***U*** or ***V*** more ideal to analyze the relationship between topics in the documents and each document?

Is ***U*** or ***V*** more ideal to analyze the relationship between topics in the documents and each term?