

INTRODUCTION

This document explains what I did to create the word cloud.

The data is about 11,000 comments taken from ratemyprofessor.com on a random subset of professors representing 20 different schools.

NOTATION

Let the quality score be the helpfulness plus clarity score on ratemyprofessor.com.

Let $A_{best} = \{10\}$, $A_{good} = \{8, 9\}$, $A_{okay} = \{7, 6, 5\}$, $A_{bad} = \{4, 3, 2\}$

Also let

$$word_x = \# \text{ of times word appears in comments with quality } \in A_x$$

and

$$word = \# \text{ of times word appears in all comments}$$

FORMULA

I gave all the words a new score denoted $wordscore_x$ that is equal to

$$wordscore_x = \left(\frac{word_x}{word}\right)^k \text{func}(word_x)$$

For the word cloud online I used $k = 1.25$ and $\text{func} = \log$ to get $wordscore_x$.

WORD CLOUD

Then I created a word cloud at worditout.com where the size of the word is proportional to $wordscore_x$.