1. Diagnostic Table:

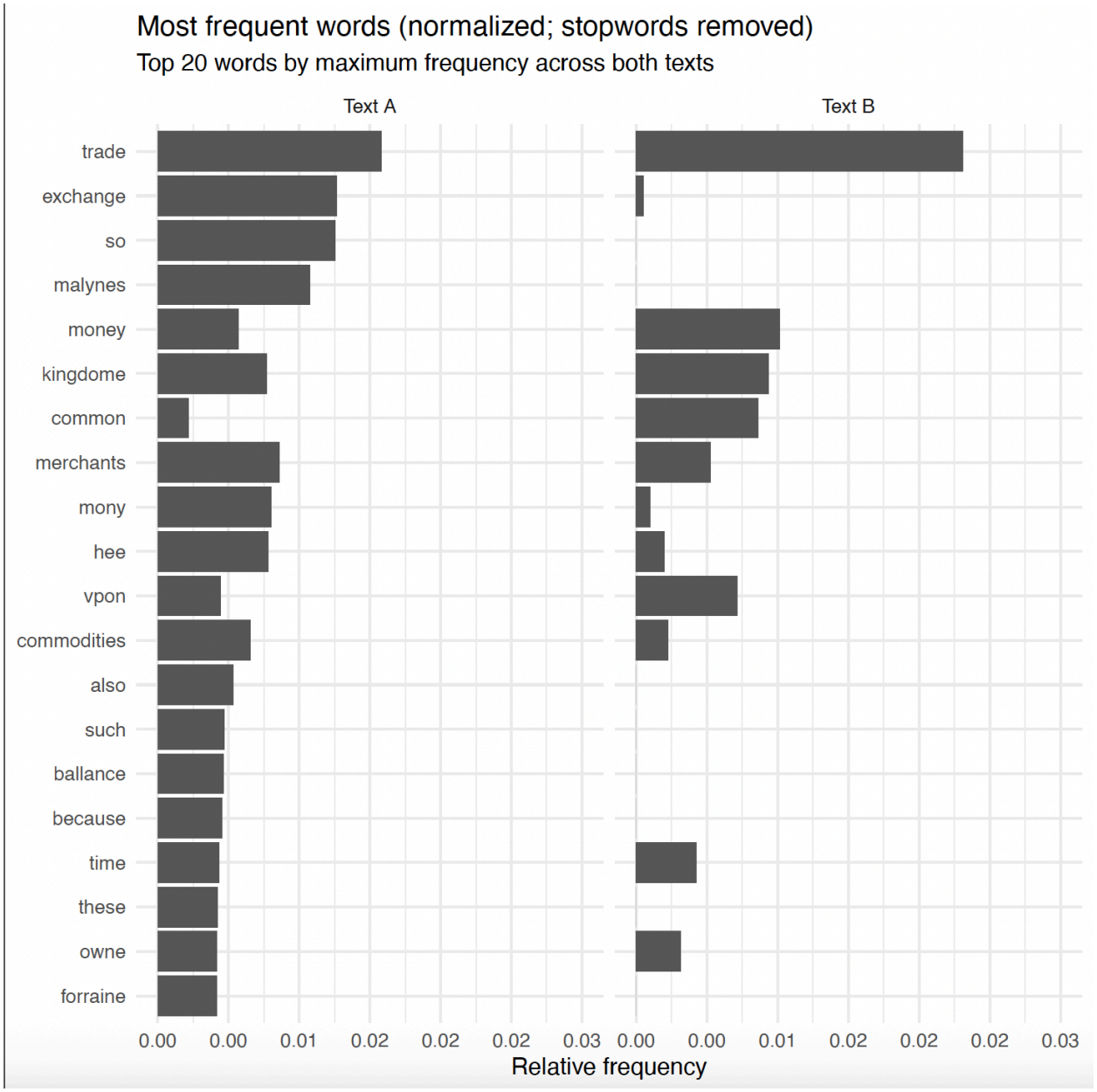
	doc_title	n_chars	n_word_tokens	n_word_types
	<chr>	<int>	<int>	<int>
1	Text A	191605	33889	4773
2	Text B	114211	19922	3332

2. Interpret the diagnostics:

1. Are Text A and Text B comparable in length?: No, Text A is longer than Text B. Text A has 191,605 characters and 33,889 word tokens, while Text B has 114,211 characters and 19922 word tokens. So text A is about 1.7x longer by tokens and about 1.68x longer by characters.
2. If they differ substantially, what does that imply for interpreting raw frequency comparisons?: Because Text A is longer, raw frequency comparisons will make it look like it “uses more” of many words simply because it uses more total language.

3. Compare normalized “trade” across the texts:

1. Does Text A or Text B use "trade" more proportionally? And how does this compare to what the raw counts suggested? : Raw count suggests “trade” more. But once we account for document length text B uses “trade” slightly more proportionally because Text B is slightly shorter.
2. We normalized by dividing each word count by the total words in that document (after stopword removal). How would your results change if you normalized by the original document length (before stopword removal)? Would this be better or worse, and why?: If we normalize by the original document length, each word's proportion would generally get smaller because the denominator would include many additional tokens that we later remove. The comparison could shift depending on whether the two texts have different stopword densities, and that stopword density itself could change if Misselden’s writing style shifts over time, for example if one text contains proportionally more function words than the other. Normalizing after stopward removal is better because it compares words against a pool of other "content carrying” words.


Most frequent words (normalized; stopwords removed)
Top 20 words by maximum frequency across both texts
Text A
Text B
trade
exchange
so
malynes
money
kingdome
common
merchants
mony
hee
vpon
commodities
also
such
ballance
because
time
these
owne
forraine
0.00
0.00
0.01
0.02
0.02
0.02
0.03
0.00
0.00
0.01
0.02
0.02
0.02
0.03
Relative frequency