1) How many documents are in this corpus?

Answer : 636 Documents

# 2) What are the top 10 tokens and their frequency?

Answer

Feature frequency

1 mr 517849

2 s 294143

3 act 245887

4 section 226094

5 states 203579

6 shall 199942

7 united 189683

8 house 183585

9 b 180166

10 bill 157303

3) From your review of the frequency table, what are the additional stopwords that you would like to drop from the 2018 to 2020 file?

Function

n\_toks = 100

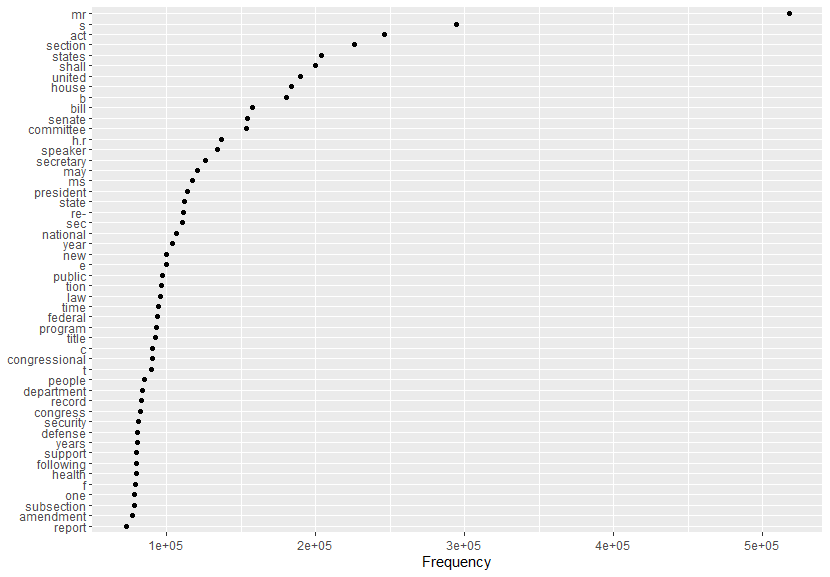
tokens <- textstat\_frequency(cr\_dfm\_clean)[1:n\_toks, 1]

Answer:

'mr, s, b, h.r, ms, re-, e, c, f, con-, in-, ii, sep, d, d-, com-, pro-, fm,

use, us,

4. Generate Frequency plot top 50 tokens



5. Generate Word Cloud

