Text Analysis Report

Rmarkdown basics: <https://quarto.org/docs/authoring/markdown-basics.html> Tutorial: <https://quarto.org/docs/get-started/authoring/rstudio.html> <https://quarto.org>. The echo: false option disables the printing of code (only output is displayed).

params$report\_rv

## <ReactiveValues>   
## Values: checkbox\_content\_IDs, comp\_freq\_plot, content, content\_freq\_plot, content\_tf\_idf, cor\_test, numFiles, other\_token, stop\_words\_final, tf\_idf\_plot, token, zipfs\_plot   
## Readonly: FALSE

## Summary statistics

A corpus of 8 was submitted for analysis. Text was tokenised by words and were omitted from their contents.

The total count of all tokens in the corpus was \_**, with the mean count per document being** .

The document with the highest token count was, \_\_, whilst the lowest was \_.

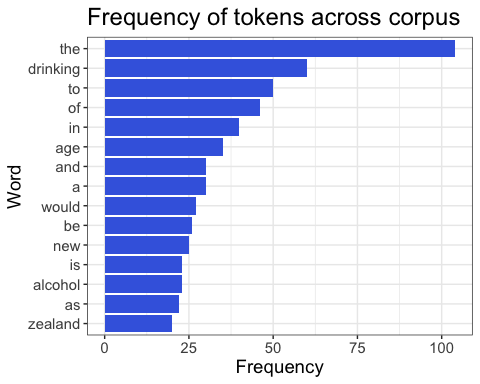
A summary of these statistics is shown in Table 1.

## # A tibble: 896 × 9  
## ID Token Freq. `Token Count` Rank `Term freq.` tf idf tf\_idf  
## <chr> <chr> <int> <int> <int> <dbl> <dbl> <dbl> <dbl>  
## 1 NZ NATIONA… god 3 43 1 0.0698 0.0698 2.08 0.145   
## 2 NZ NATIONA… defe… 2 43 3 0.0465 0.0465 2.08 0.0967  
## 3 1047051.txt with… 7 229 4 0.0306 0.0306 2.08 0.0636  
## 4 1067080.txt adol… 6 167 3 0.0359 0.0359 1.39 0.0498  
## 5 NZ NATIONA… afar 1 43 7 0.0233 0.0233 2.08 0.0484  
## 6 NZ NATIONA… bonds 1 43 10 0.0233 0.0233 2.08 0.0484  
## 7 NZ NATIONA… entr… 1 43 11 0.0233 0.0233 2.08 0.0484  
## 8 NZ NATIONA… feet 1 43 12 0.0233 0.0233 2.08 0.0484  
## 9 NZ NATIONA… free 1 43 13 0.0233 0.0233 2.08 0.0484  
## 10 NZ NATIONA… guard 1 43 15 0.0233 0.0233 2.08 0.0484  
## # … with 886 more rows

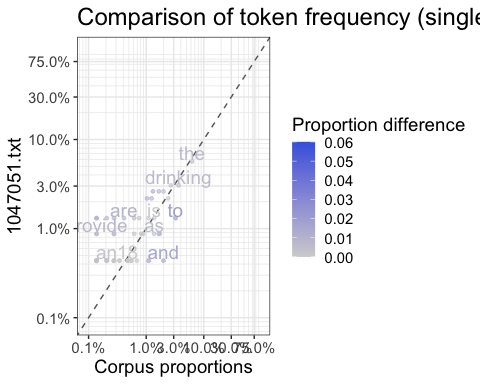
*Table 1. Summary statistics for the submitted current corpus.*

## Token frequency

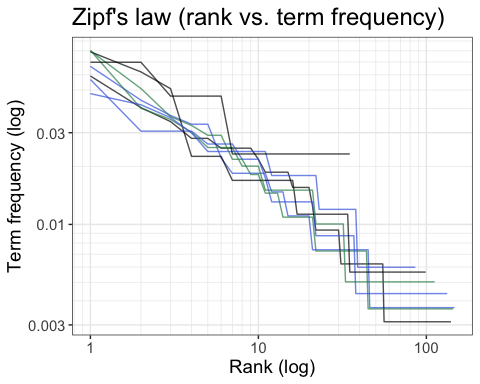
### Most frequent terms

  
*Figure 1. Most frequent tokens across the corpus.*

## Token frequency comparison



## Zipf’s Law



## Term frequency inverse document frequency (tf-idf)

Tf-idf analysis shows some of the most important words to the document 1012888.txt are “being”, “that”, and “addressed”.

