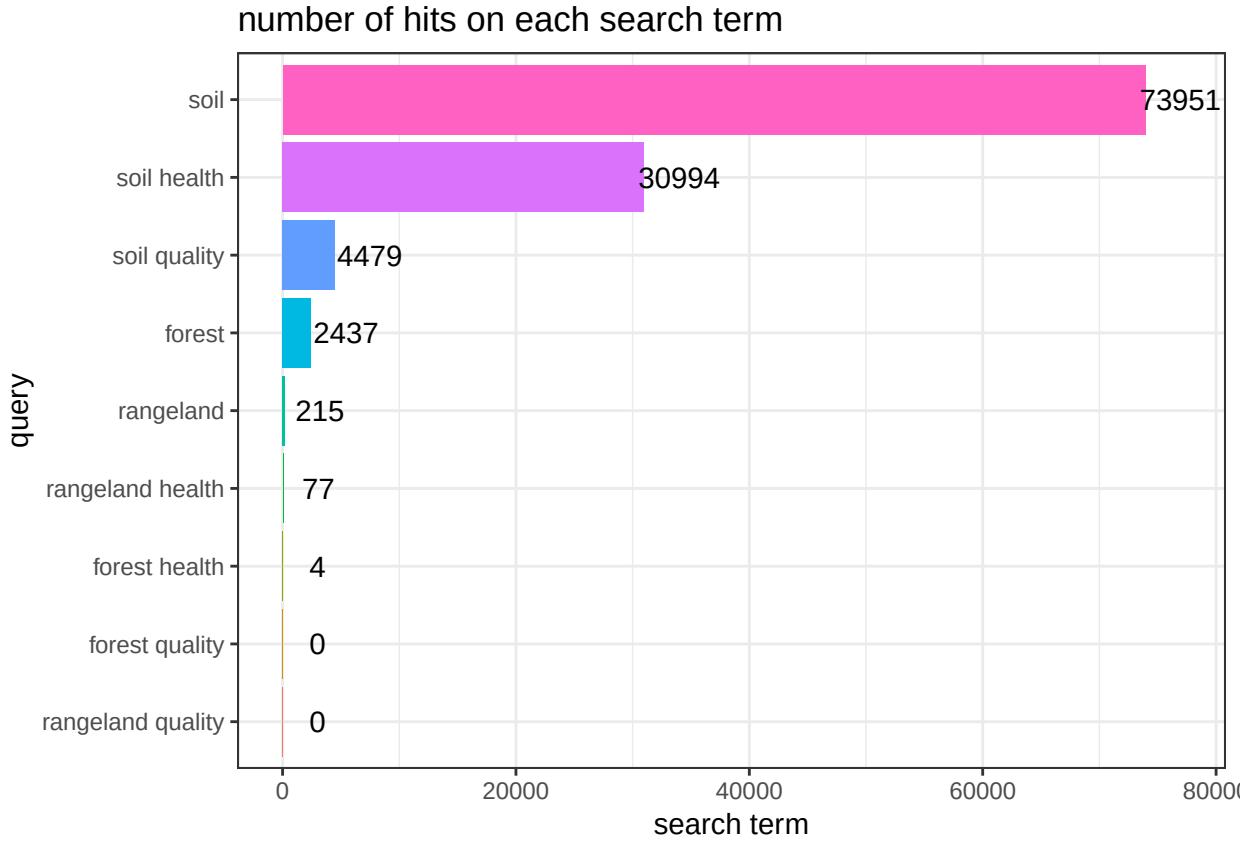
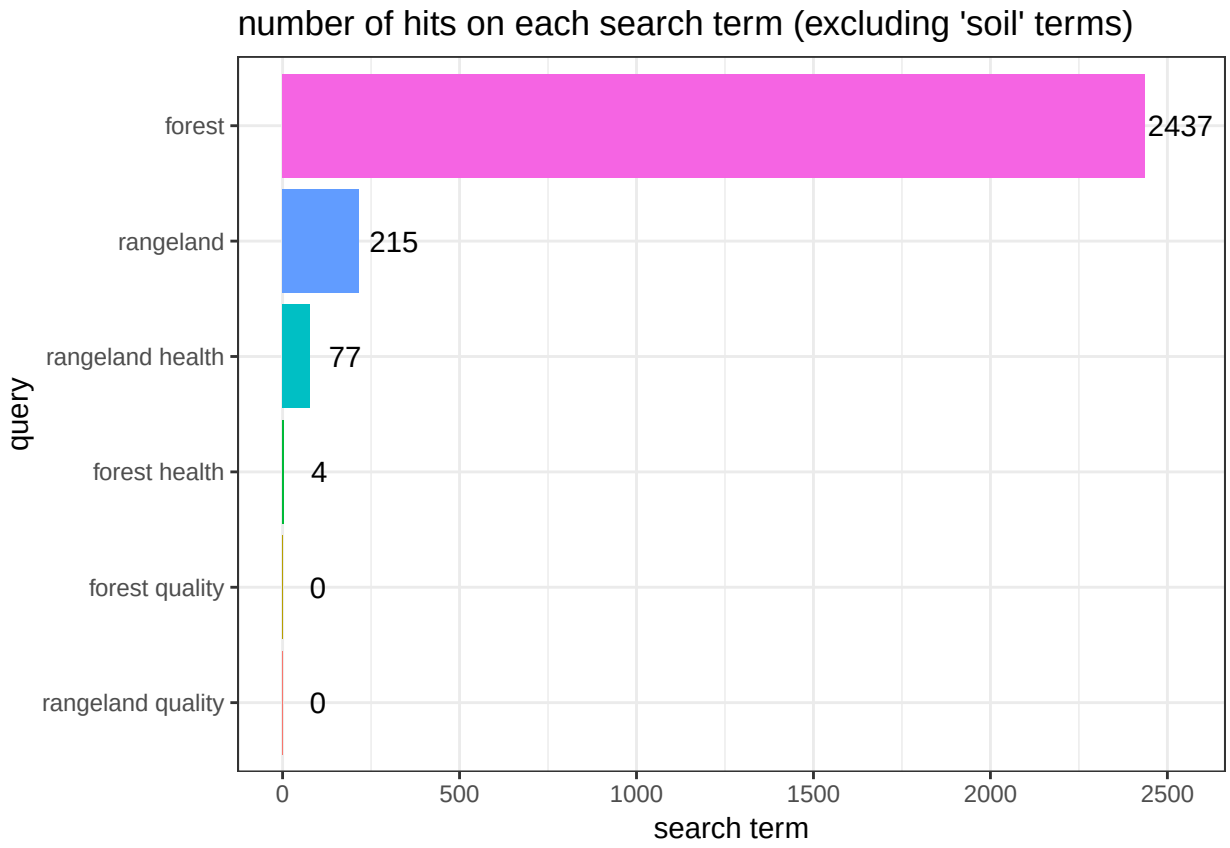
word_assoc

Length of dataframes produced by each "category"

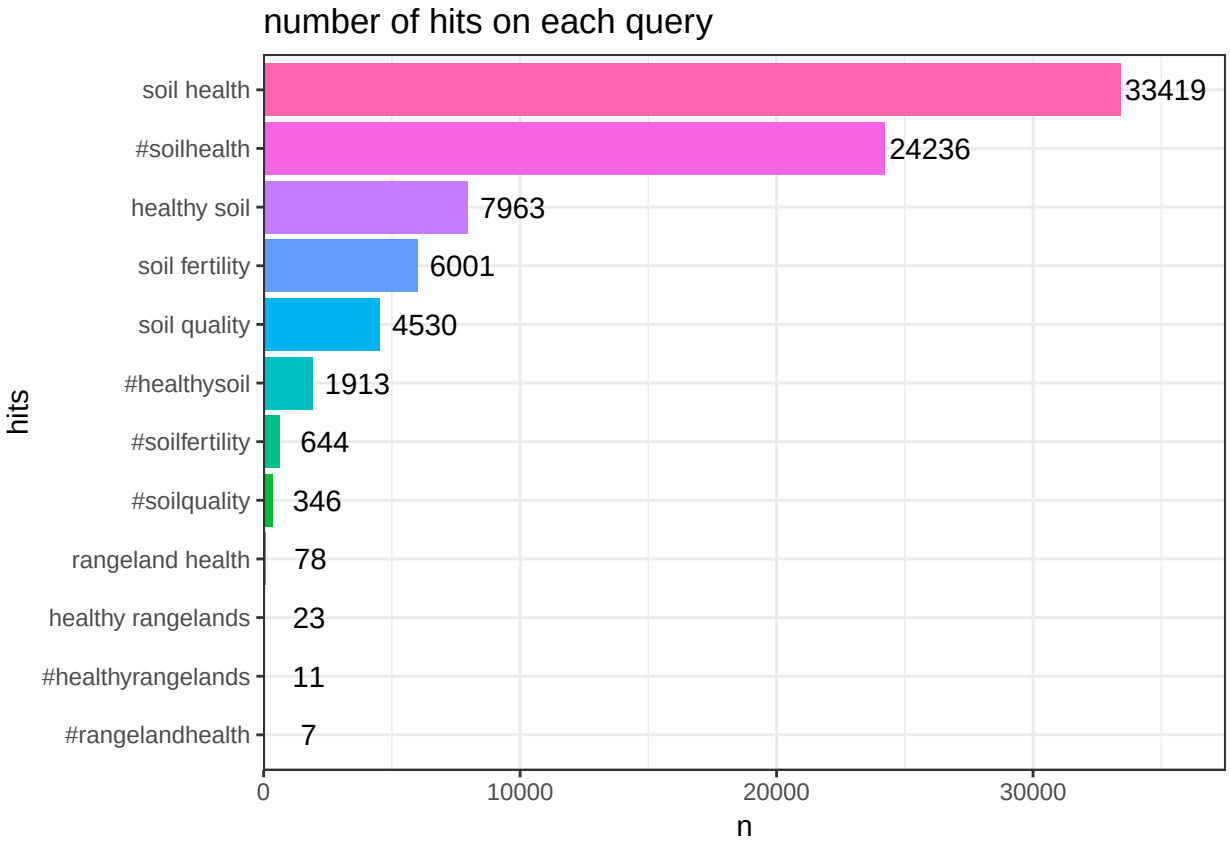
it is import to keep in mind for all results (including here, in wordclouds, and bigram plots) that the number of hits for each query term varies dramatically, with "soil" terms being used far more frequently than "rangeland" or "forest" terms. For this reason, it may not be very valuable to perform this sort of analysis while looking at the most influential tweets. Nevertheless, this it seems valuable to get a feel for the different language being used within each category



same as above without 'soil' terms



looking at the relative hits on query terms (i.e those originally used to querry the API)

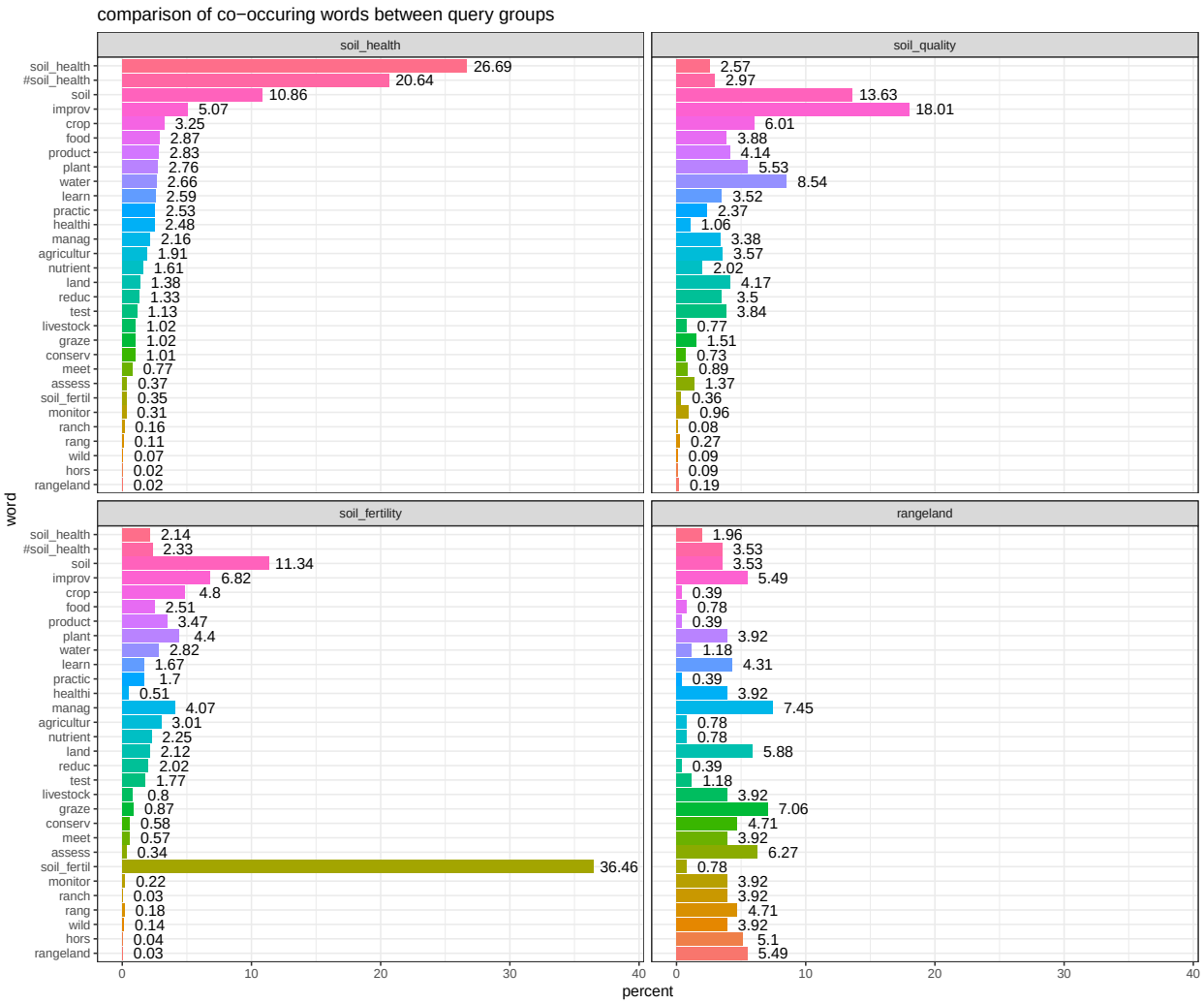


compare if/how content changes when a tweet contains each group of query terms

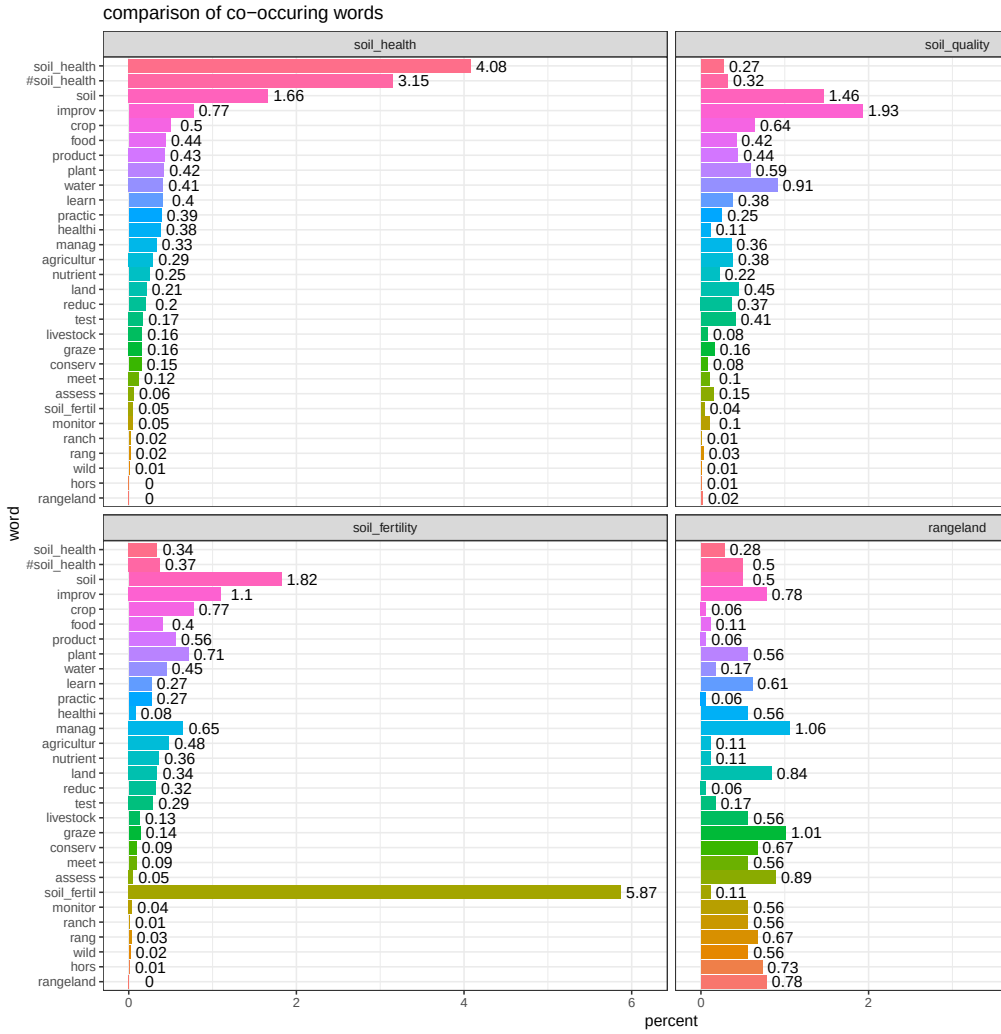
groupings are:

- soil health = 'soil health','#soilhealth','healthy soil','#healthysoil'
- soil quality = 'soil quality','#soilquality'
- soil fertiliy = 'soil fertility','#soilfertility'
- rangeland = 'rangeland health','#rangelandhealth','healthy rangelands','#healthyrangelands'

caclculate percentages based relative to word counts of top 20

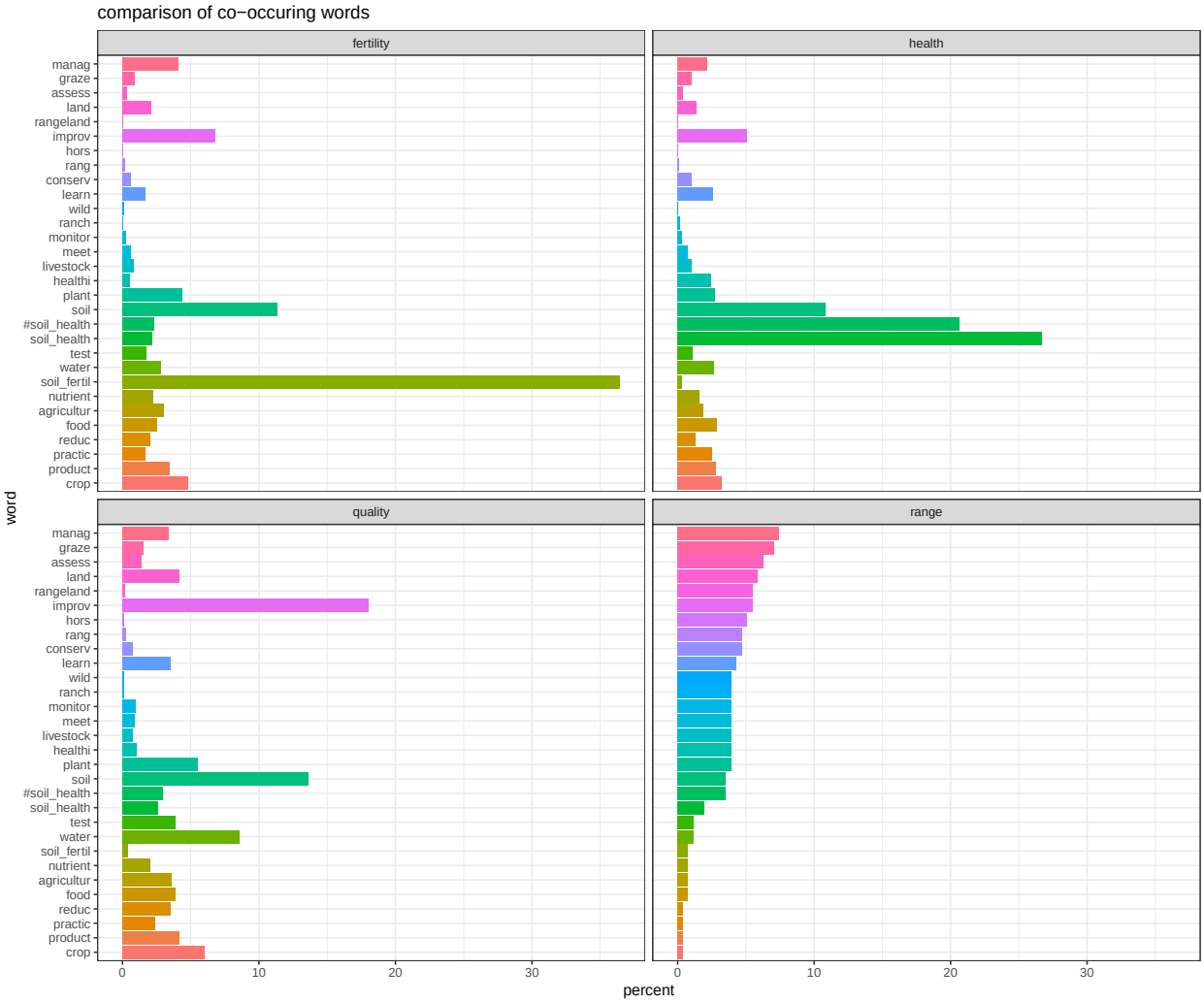


caclculate percentages based relative to total word counts of each category



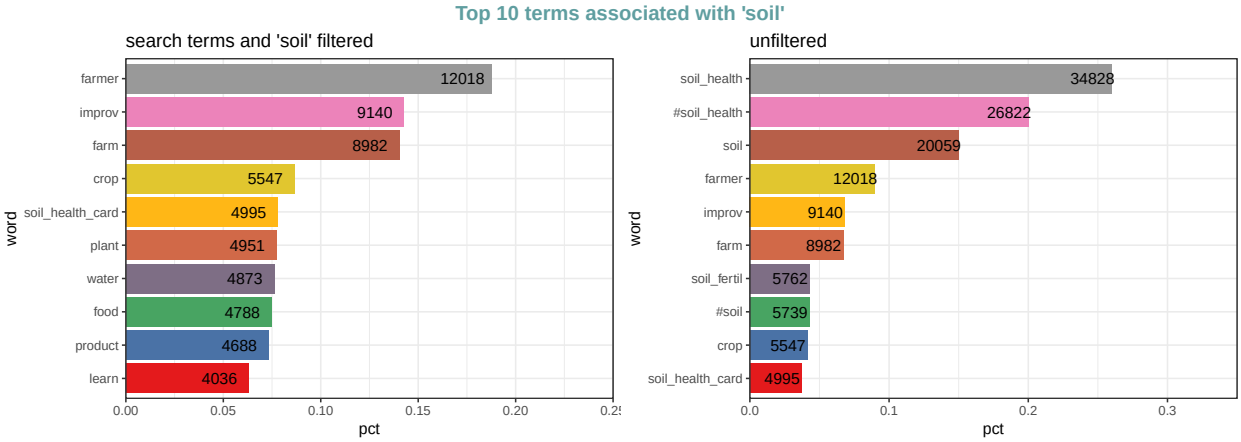
no change in trend, only in percentages

same comparison, but visually comparing based on word counts of 'rangeland'


comparison of co−occuring words
fertility
health
quality
range
word
percent
manag
graze
assess
land
rangeland
improv
hors
rang
conserv
learn
wild
ranch
monitor
meet
livestock
healthi
plant
soil
#soil_health
soil_health
test
water
soil_fertil
nutrient
agricultur
food
reduc
practic
product
crop
0
10
20
30

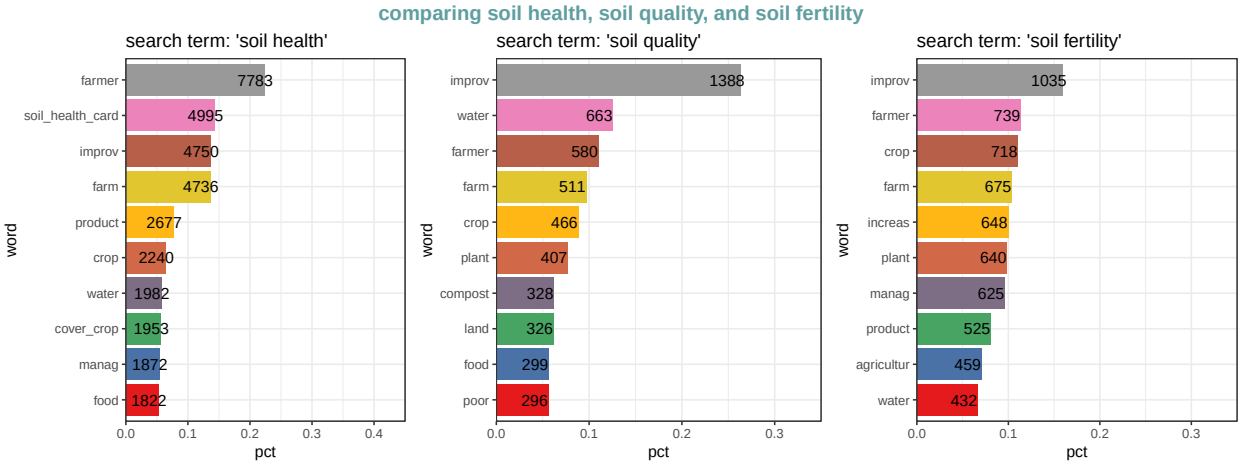
visualizing propotions

looking at 'soil' terms alone



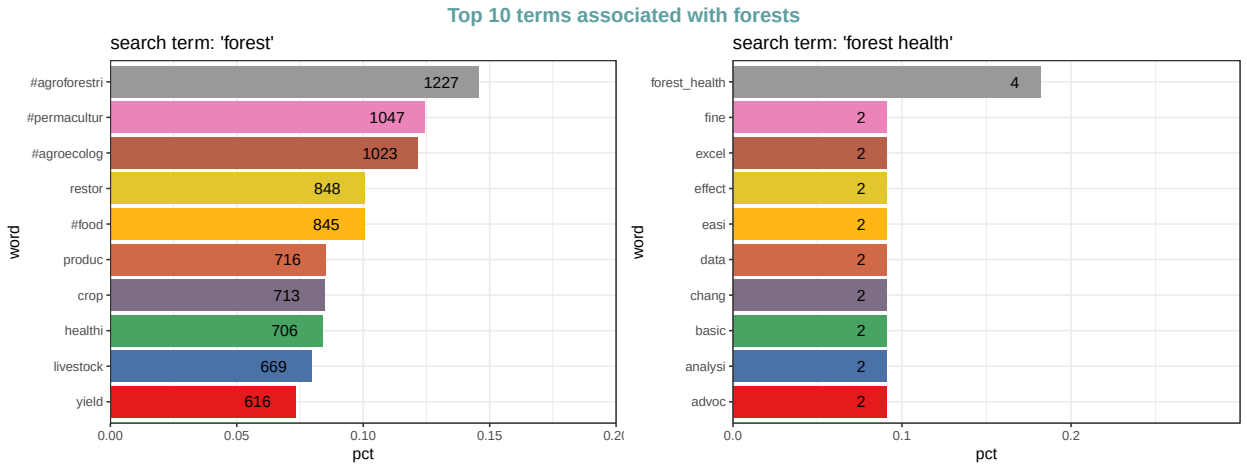
percents are relative to the top 10 terms for each respective graph. Values in each column are the raw count of each word

comparing: soil health, soil qualitly, soil fertility



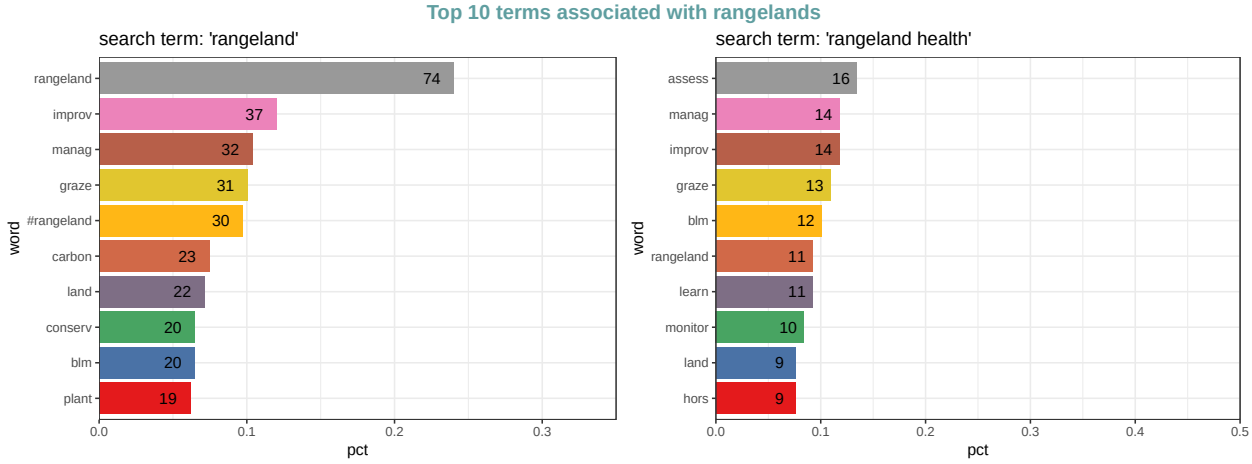
percents are relative to the top terms for each respective graph. Values in each column are the raw count of each word. the terms 'soil quality' and 'soil fertility' were filtered out of ther respective graphs as they were major outliers

forest



percents are relative to the top terms for each respective graph. Values in each column are the raw count of each word. The 'forest health' search was filtered for words counts > 1

rangeland



percents are relative to the top terms for each respective graph. Values in each column are the raw count of each word.