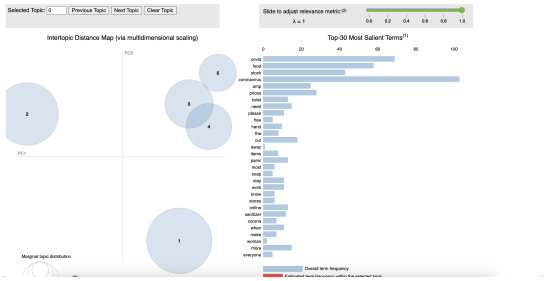
CSC2/447: Homework 1

{Ryo Novoa}

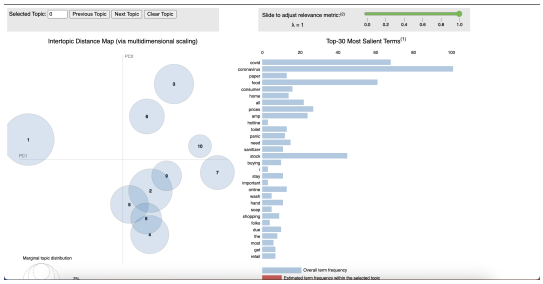
The solutions are my own, and I consulted with no one while working on this homework.

[Q1.1] Visualization of the topic distribution with num_top $= 5, 10, 20$ and num_pass $= 10$. Include images of the topic distribution below.

- num_top $= 5$

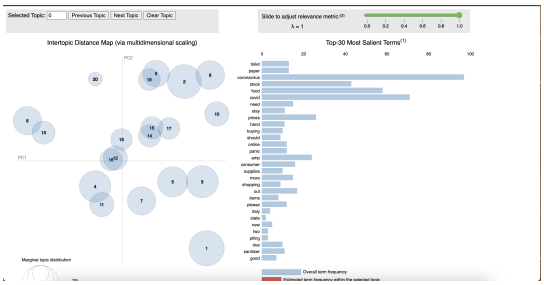


- num_top $= 10$



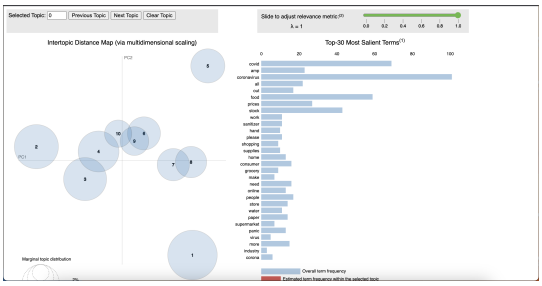
- num_top $= 20$

[Q1.2] Describe how the topic distribution changes when adjusting the number of topics:
There is more overlap to be seen in distributions with more topics. As such, we may also see what appears to be more of a relation in between topics.

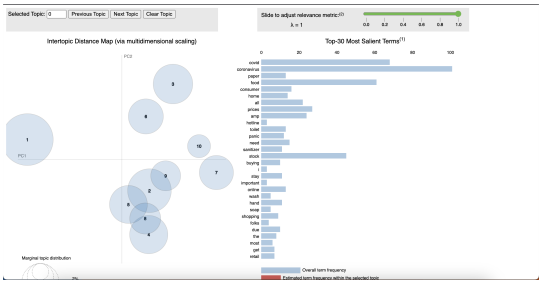


[**Q2.1**] Visualization of the topic distribution with num_top = 10 and num_pass = 5, 10, 20. Include images of the topic distribution below.

- num_pass = 5

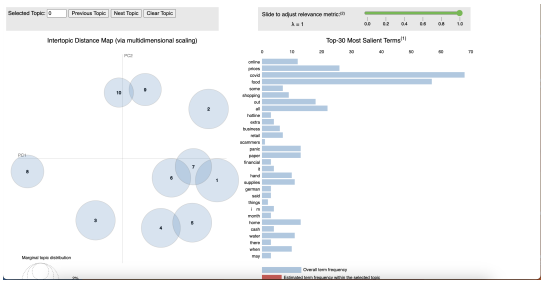


- num_pass = 10



- num_pass = 20

[**Q2.2**] Describe how the topic distribution changes when adjusting the number of passes:
Interestingly enough, here it seems that the increase in number of iterations is separating each individual topic further from their counterparts. We also see an increase in the average size of the topics shown on the graph.


Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(1)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
Marginal topic distribution
Top-30 Most Salient Terms(1)
0
10
20
30
40
50
60
70
online
prices
covid
food
some
shopping
out
all
rushing
extra
business
retail
scammers
panic
paper
financial
k
hand
supplies
german
sold
things
i m
month
home
cash
water
there
when
may
Overall term frequency
Estimated term frequency within the selected topic