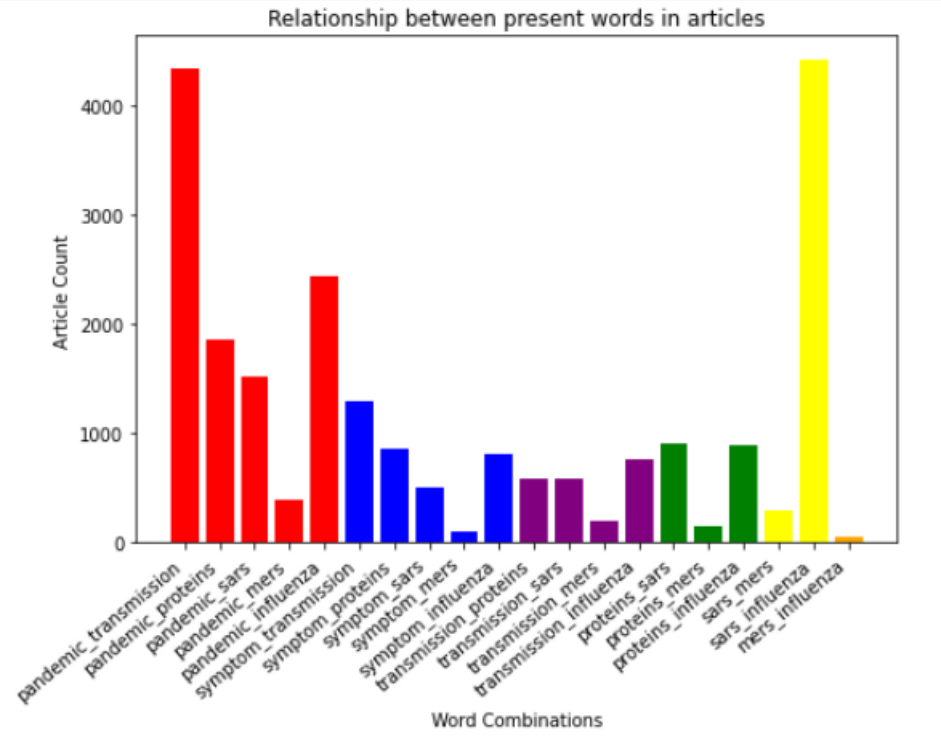
Final Exam Project

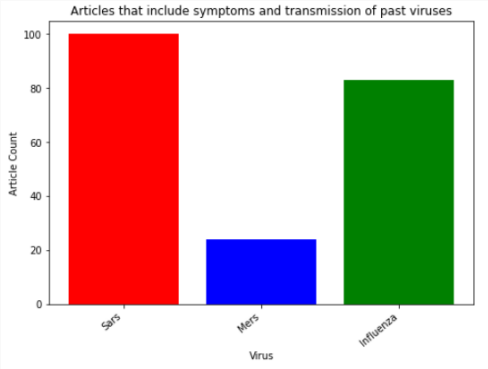
Supporting research into COVID-19

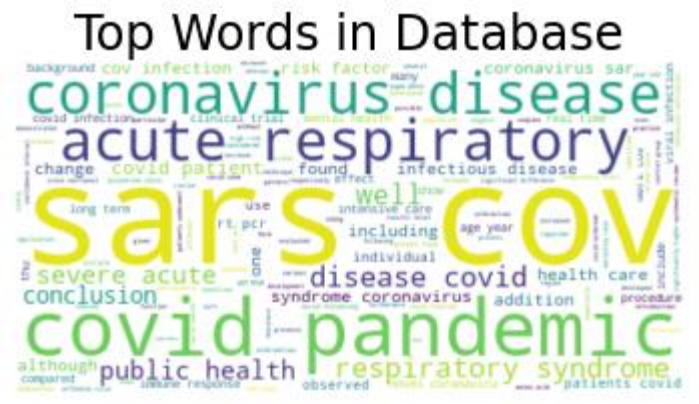
Melina Fartaj
Student Number:
1002348596

Exploratory Analysis

- The relationship between present words in articles was created by searching for keywords associated to Pandemics, symptoms, past viruses, cell breakdowns and transmission.
- The Articles that include symptoms and transmission of past viruses graphs shows the amount of articles related to each virus in the database
- The word cloud was created to view the top words that were present in the database

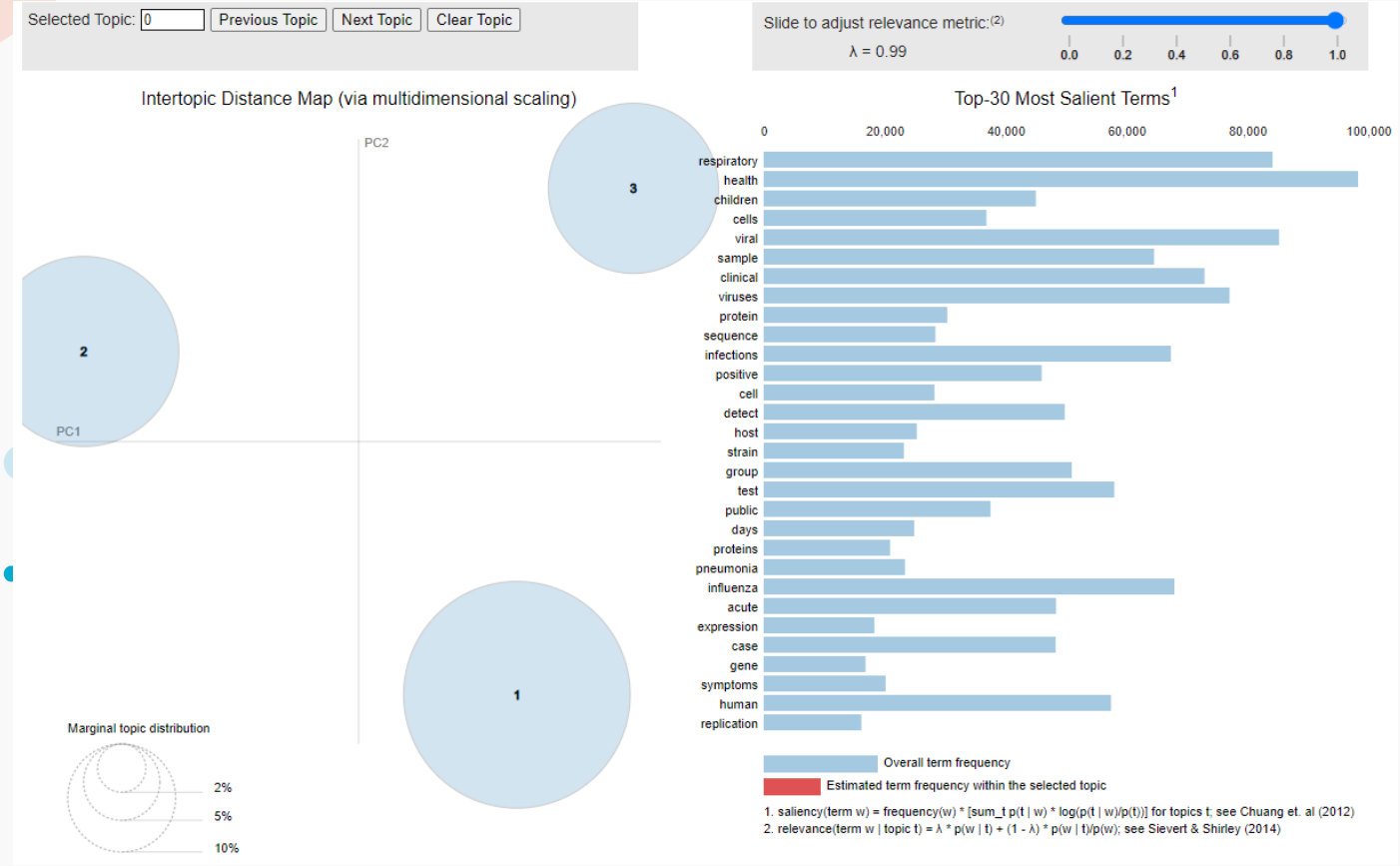


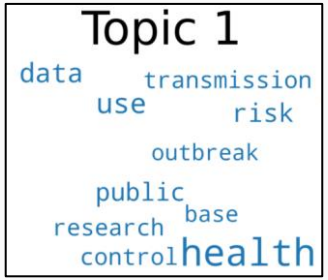


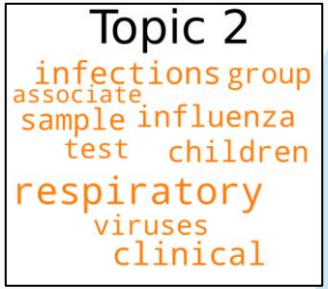


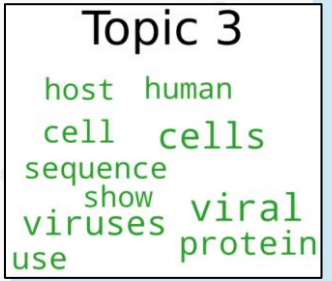
LDA Model Features

- Key words for each topic were explored using gensims interactive visual which allows you to see the keywords of the topics in relation to each other. A word cloud of the 10 key words for each topic is also created for easier visualization.



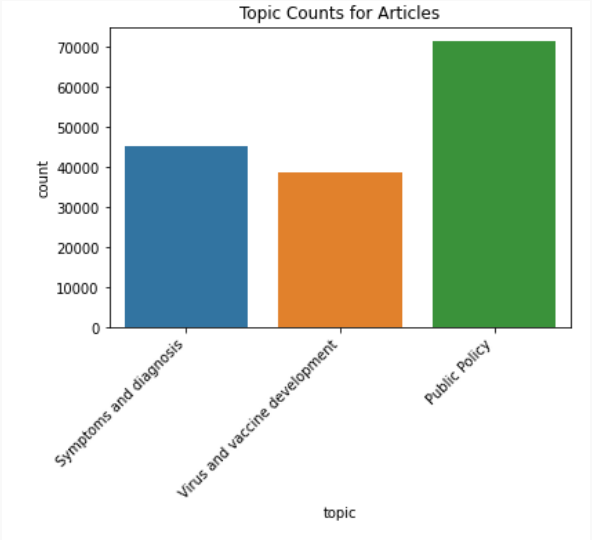


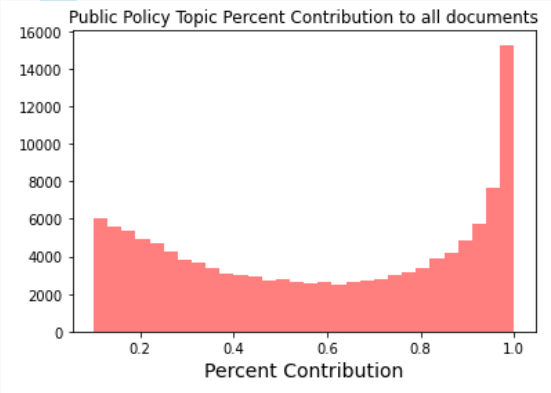


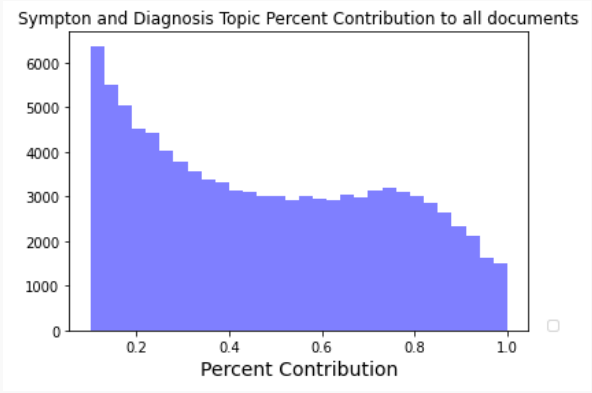


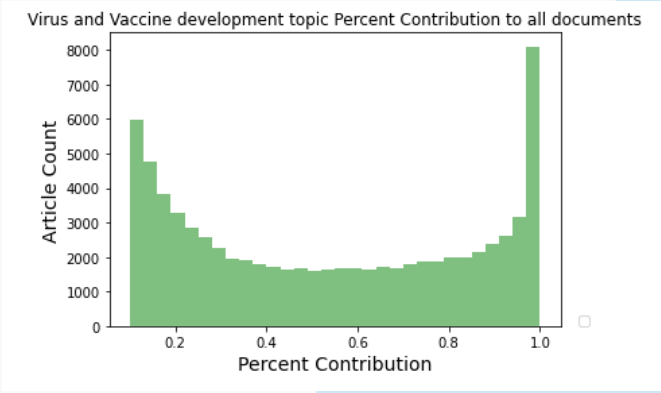
Exploring Topics found by LDA model

- The top counts for the articles by topic can be seen on the right. The highest topic count was discovered to be Public Policy.
- The distribution of topics for each document was also calculated. The distributions of each topic can be seen below. It was found that public policy and Virus and Vaccine Development not only had the most amount of articles but it also had the articles with the highest topic percentage (meaning some articles almost 100% were about public policy)



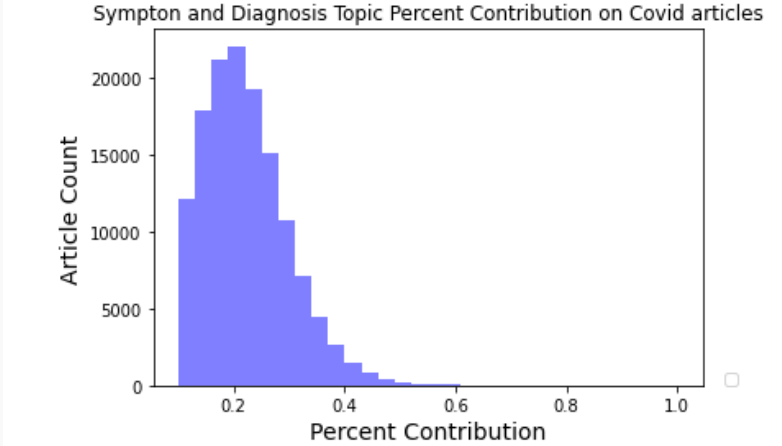


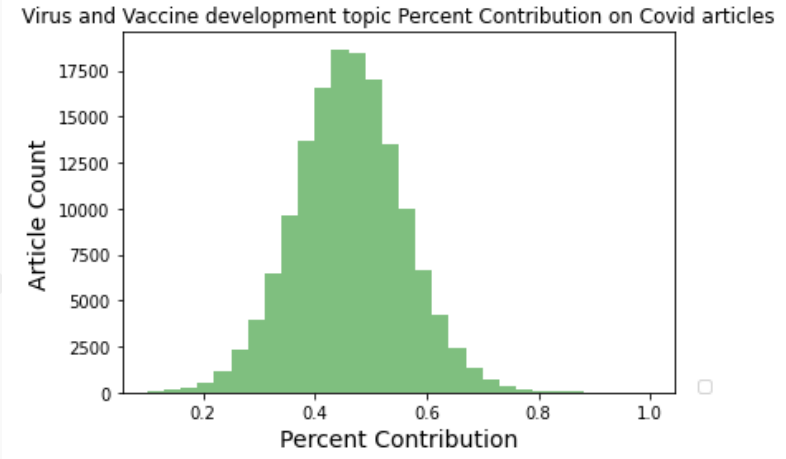


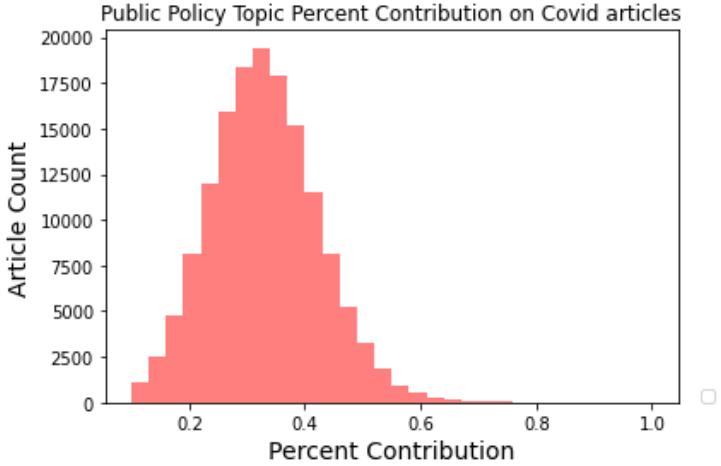


LDA Modelling with Unseen Covid Papers

- By looking at the topic distributions for the unseen covid data it is noticed that not many documents had a high topic percentage for any of the topics. This can be a sign that our topic model may be overfit or that the unseen covid papers discuss relatively different topics than the ones found.







Insights and Guidance Example

- Example: Looking to gain insight about Virus and Vaccine development from pre-coivid-19 articles based on the topics that were allocated from the model
- First we filter for only the virus and vaccine development papers and then the paper with the highest precent contribution in those papers
- The found article is below:

"influenza viruses are enveloped, negative stranded, segmented rna viruses belonging to orthomyxoviridae family. each virion consists of three major subviral components, namely (i) a viral envelope decorated with three transmembrane proteins hemagglutinin (ha), neuraminidase (na) and m2, (ii) an intermediate layer of matrix protein (m1), and (iii) an innermost helical viral ribonucleocapsid [vrnp] core formed by nucleoprotein (np) and negative strand viral rna (vrna). since complete virus particles are not found inside the cell, the processes of assembly, morphogenesis, budding and release of progeny virus particles at the plasma membrane of the infected cells are critically important for the production of infectious virions and pathogenesis of influenza viruses as well. morphogenesis and budding require that all virus components must be brought to the budding site which is the apical plasma membrane in polarized epithelial cells whether in vitro cultured cells or in vivo infected animals. ha and na forming the outer spikes on the viral envelope possess apical sorting signals and use exocytic pathways and lipid rafts for cell surface transport and apical sorting. np also has apical determinant(s) and is probably transported to the apical budding site similarly via lipid rafts and/or through cortical actin microfilaments. m1 binds the np and the exposed rnas of vrnps, as well as to the cytoplasmic tails (ct) and transmembrane (tm) domains of ha, na and m2, and is likely brought to the budding site on the piggy-back of vrnp and transmembrane proteins. budding processes involve bud initiation, bud growth and bud release. presence of lipid rafts and assembly of viral components at the budding site can cause asymmetry of lipid bilayers and outward membrane bending leading to bud initiation and bud growth. bud release requires fusion of the apposing viral and cellular membranes and scission of the virus buds from the infected cellular membrane. the processes involved in bud initiation, bud growth and bud scission/release require involvement both viral and host components and can affect bud closing and virus release in both positive and negative ways. among the viral components, m1, m2 and na play important roles in bud release and m1, m2 and na mutations all affect the morphology of buds and released viruses. disassembly of host cortical actin microfilaments at the pinching-off site appears to facilitate bud fission and release. bud scission is energy dependent and only a small fraction of virus buds present on the cell surface is released. discontinuity of m1 layer underneath the lipid bilayer, absence of outer membrane spikes, absence of lipid rafts in the lipid bilayer, as well as possible presence of m2 and disassembly of cortical actin microfilaments at the pinching off site appear to facilitate bud fission and bud release. we provide our current understanding of these important processes leading to the production of infectious influenza virus particles."

- This provides insight into bud initiation, bud growth and bud scission/release and its affects which is related to released viruses. Information like this can lead researchers to investigate bud initiation, growth and release to determine the development of the COVID-19 virus which is important because once there is an understanding of how the virus is developed then the path to find a treatment is smoother.
- Since this is only the abstract, not all the information can be pulled from here but if a researcher reads something that they are interested in learning more about they can easily access the whole article.