Assignment3

My final project is about “Changes in research interests during COVID-19 within the sociology community”. Thus, I gathered the data about COVID-19 and sociology papers. First, I downloaded a COVID-19 case data by date (.csv file) which is posted by WHO. After that, I filtered and grouped the data because I’ll only use the data that refers to Republic of Korea. Additionally, I downloaded academic papers data(.xls file) provided by KCI(Korea Citation Index). Since interest of this research is focused on sociology, I have only downloaded papers in the field of sociology. In addition, I’ll only use KCI registered papers to exclude low-quality papers, which does not reflect the academic circumstances well. Then I manually add column which indicates a month paper published, since KCI data didn’t have it.

My main variables are “Progress of COVID-19”, “Topics of Sociology Papers“, “Keywords of Sociology Papers”. The former represents the independent variable, and the latter two are component variables of the dependent variable “Research Interests of Sociology Papers”. First, in COVID-19 data, there is huge outlier because of the explosive increase in the number of cases due to the Omicron mutation. As you can see, y-scale of 2022 case is fairly large, even in March it peaks to $9,961,089$. In the main study, therefore, the study period will be four years from 2018 to 2021. And about sociology papers, I present changes in the number of academic papers($N = 7,989$) instead of proportion of topics, because the number of published papers shows periodic changes. This data shows that there is 3 cycle increase and it is especially high in June and December. At last, I made a wordcloud of keywords used in papers to see frequency of words. Since there are so many words($N \approx 26,422$, because of preprocessing, it is exact value) and frequencies of words are sparse, I only choosed word that appear more than two times($N = 1058$). The largest frequency word is “life satisfaction”, which is appeared 68 times, and second one is “depression”, 65 times. While this wordcloud shows that there are lots of psychology-related papers in sociology, this should be compared to ones in other periods.

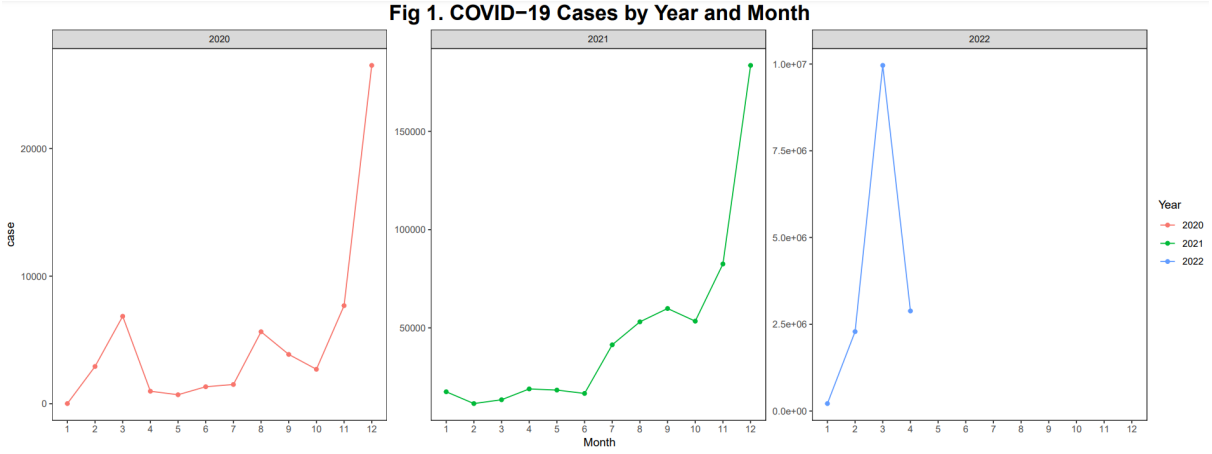

Fig 1. COVID-19 Cases by Year and Month

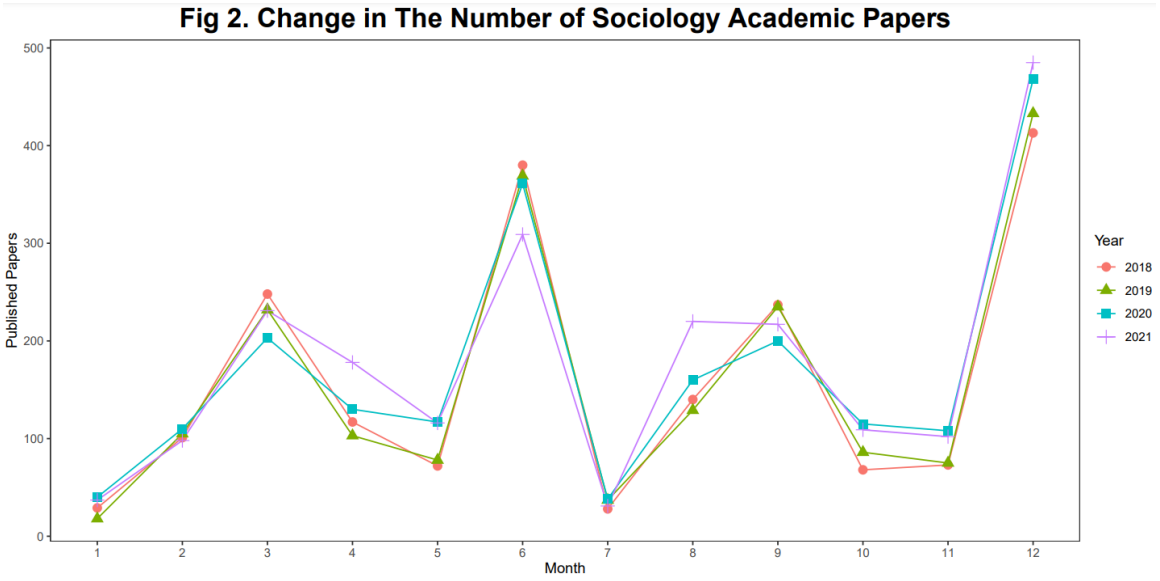

Fig 2. Change in The Number of Sociology Academic Papers

Fig 3. Wordcloud That Presents Keywords of Sociology Papers