Team Members: Anna Krutsinger, Keizo Morgan, Ingrid Tsang

Research Question: How does geography affect the human impact of COVID-19?

Brief Outline of Methods:

We plan to use PCA and LDA to identify relevant geographical factors from research papers by analyzing and extracting information from articles, such as

- Human geography: population density, demographics (e.g. median income, average education level), primary industry, rural/suburban/urban, level of pollution
- Physical geography: climate (e.g. median temperature), latitude

With these topics, supplemented with our own additional factors, we will use regression techniques to compare county-level and state-level data on incident, hospitalization and mortality rates across the US. From this, we hope to employ multiple correspondence analysis (MCA) to aggregate two or more of the above characteristics to grasp a better understanding of environmental impacts on the spread and impact of COVID-19.

Sources:

- Kaggle challenge
    - Help us understand how geography affects virality (https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks?taskId=536)
    - Literature Review Kernel (https://www.kaggle.com/dgunning/transmission-incubation-and-environment-stability)
- US county-level data from John Hopkins (https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data)
    - Incident rate: confirmed cases per 100,000 people
    - People tested
    - People hospitalized
    - Mortality rate
- US state-level data from John Hopkins (https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_daily_reports_us)

- "The Geography of Coronavirus" (https://www.citylab.com/equity/2020/04/coronavirus-spread-map-city-urban-density-suburbs-rural-data/609394/)
- "Do weather conditions influence the transmission of the coronavirus (SARS-CoV-2)?" (https://www.cebm.net/covid-19/do-weather-conditions-influence-the-transmission-of-the-coronavirus-sars-cov-2/)
- "Multiple factor analysis: principal component analysis for multitable and multiblock data sets" (https://personal.utdallas.edu/~herve/abdi-WiresCS-mfa-2013.pdf)