* An analysis of fire response time and fire risk in London (Difficulty: 2/5)
  + The London Fire Bridge provides an open data of fire incident records. This dataset can be used to study the fire response time and fire risk in granular spatial units (e.g. LSOA). By linking the fire incident records with other datasets (building density, population density, road density, traffic), we can answer questions like 'why some areas have a high fire risk' and 'why the response time in these neighbourhoods are higher than others'.
* Social Connectedness in Europe: implications for Brexit voting (and other topics) (Difficulty: 3/5)
  + This dataset is an anonymized snapshot of all active Facebook users and their friendship networks to measure the intensity of connectedness between locations. The Social Connectedness Index (SCI) is a measure of the social connectedness between different geographies. Specifically, it measures the relative probability that two individuals across two locations are friends with each other on Facebook.
  + By linking with other data (e.g. Brexit votes, hate crime rate), there are a lot to explore using this dataset. For example, one would assume that if an area is well connected with other EU countries, it is less likely that it would vote for Brexit. Is this true?
  + References
    - <https://data.humdata.org/dataset/social-connectedness-index>
    - <https://osf.io/preprints/socarxiv/3wh67>/
* Spatial analysis of UK local vaccination services: accessibility and coverage (Difficulty: 2/5)
  + This research aims to explore the spatial coverage of the UK vaccination services (https://www.england.nhs.uk/coronavirus/publication/vaccination-sites/) and the accessibility of the public to vaccine services. ONS claims that 96% of the population is within 10 miles of a vaccine service. However, it is still unknown how many people ‘share’ a vaccine site and whether there is spatial inequality in the vaccine service.
  + In this project, you would need to transform the addresses of vaccine sites into coordinates, and use methods of coverage and accessibility analysis. In addition, you could work on the relocation and optimisation of vaccine sites.
* Quantifying the dispersion of 'pockets of deprivation': a case study of Greater London (Difficulty: 2/5)
  + As described in the webpage of London datastore (https://data.london.gov.uk/dataset/indices-of-deprivation), the pockets of deprivation are becoming more dispersed in 2019, compared to 2015. Is this supported by quantitative analysis? What is the definition of pockets of deprivation? Given that the deprivation index is a rank (or an ordinal variable), is it possible to quantify the spatial dispersion of the pockets of deprivation?
  + Methods: clustering methods of ordinal data, spatial statistics
* Interpreting multiple deprivation index using game theory (Difficulty: 3/5)
  + The English Indices of Deprivation 2019 contains 7 domains of deprivation, which are combined using non-linear relationship to create the Index of Multiple Deprivation. This research aims to study how these indices contribute to the individual and overall values of IMD. Potential methods include Shapley values and SHAP (Shapley Additive exPlanations), which stem from game theory and have been used to explain individual predictions in machine learning. Another potential method is the Local Interpretable Model-Agnostic Explanations (LIME).
* Relationship between Geographically weighted regression and Local Interpretable Model-Agnostic Explanations (LIME) (Difficulty: 4/5)
  + This research is a comparison of two methods. Is GWR a specific case of LIME, in the sense that GWR contains the spatial distance metric? If so, is it possible to improve or adjust GWR by borrowing some ideas of LIME?
* Reinforcement Learning and epidemic control (Difficulty: 5/5)
  + Given human movement in a city/region, can we use reinforcement learning to identify the optimal policy that controls the Covid-19 epidemic? This study will leverage the existing code from a KDD paper *Reinforced Epidemic Control: Saving Both Lives and Economy*, and validate this framework using the dataset from UK or the US. You would need to explore the data availability for this project.
  + Github link: <https://github.com/anyleopeace/DURLECA>
  + An introduction on Sina weibo (Chinese only): https://m.weibo.cn/1645372340/4542228212488233
  + An introduction on Bilibili video (Chinese only): https://m.bilibili.com/video/BV1Ja4y177Kr