**DATASETS COMPILATION/DATASETS USED**

|  |  |  |
| --- | --- | --- |
| **Dataset Name** | **Type** | **Source** |
| Master Plan 2014 Subzone Boundary (Web) | Geospatial | <https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-web> |
| Singapore Childcare Centres | Aspatial | <https://data.gov.sg/dataset/child-care-services> |
| Airbnb Listings 2019 and 2021 | Aspatial | <http://insideairbnb.com/get-the-data.html> |
| Myanmar Township Boundaries | Geospatial | <https://data.humdata.org/dataset/mimu-geonode-myanmar-township-boundaries-mimu> |
| 2014 Myanmar Population and Housing Census Myanmar (ICT penetration) | Aspatial | <https://myanmar.unfpa.org/en/publications/2014-population-and-housing-census-myanmar-data-sheet> |
| London Borough Boundaries | Geospatial | <https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london> |
| Greater London Crimes by month, 2019 | Aspatial | <https://data.london.gov.uk/dataset/recorded_crime_summary>  dataset is in the London\_crime\_2019 folder, which is further split into months. Expectation of input to be one continuous excel file, not over multiple files, so the files should be combined into one year instead.  Also, the Greater London area is covered by the Metropolitan Police Service and the City of London Police data sets. |
| Greater London Crime Summary, 2016 | Aspatial | <https://github.com/dmorison/eda-relationships-between-crime-london>  this dude alr helped to prepare the data (going through folders to join the date + retain relevant columns) so I’ll be using this for analysis |

\* London dataset reference: <https://towardsdatascience.com/plotting-a-map-of-london-crime-data-using-r-8dcefef1c397>, [dataset link 1](https://www.gov.uk/government/statistics/police-recorded-crime-open-data-tables), [dataset link 2](https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datalist?filter=datasets&fromDateDay=&fromDateMonth=&fromDateYear=&query=crime&size=10&sortBy=release_date&toDateDay=1&toDateMonth=1&toDateYear=2021&page=2)

**STANDARDISED DATA FORMAT**

For geospatial data:

* Has to be in shapefile format (for st\_read function to output simple features object)
* Should be the boundaries of the study area: subregions (districts, counties, townships or boroughs) are OK and preferred

For aspatial data:

* (a) has no longitude or latitude, only has subregion information (.csv)
* (b) has longitude and latitude (.csv or .rds or .geojson) 🡪 will be converted to geospatial data ]
  + Have the columns be known as Lat and Lng specifically

Past projects: <https://wiki.smu.edu.sg/1718t2is415g1/Project_Groups>