

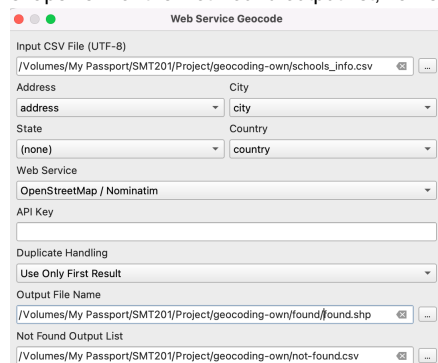
## Data preparation for distance matrix


### 1. Data extraction for subzone boundary

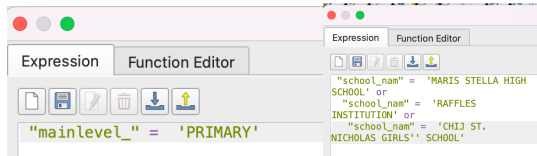
- a. Subzone boundary data can be extracted from data.gov.sg → master plan 2014 subzone boundary (<https://data.gov.sg/dataset/master-plan-2014-subzone-boundary-no-sea>)
- b. Import the kml MP14\_SUBZONE\_NO\_SEA\_PL file and save it into geopackage, converting the CRS to EPSG3414/ SVY21. Rename it as 'MP14 subzone'.

### 2. Data extraction for schools from online sources

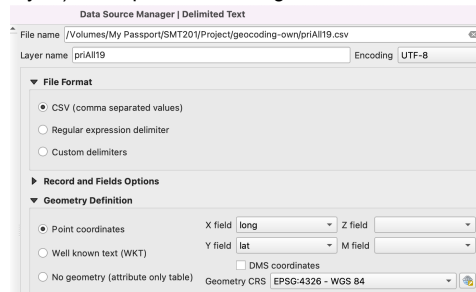
- a. 2019 data
  - i. Location of schools extracted from data.gov.sg (<https://data.gov.sg/dataset/school-directory-and-information>) → general-information-of-schools.csv
  - ii. Since general-information-of-schools.csv does not contain latitude or longitude data/ coordinate data, geocoding using the plugin MMQGIS needs to be done. Before geocoding, open the csv in excel and add 2 more columns 'City' and 'Singapore' and fill both columns with 'Singapore'. Save changes and rename the csv to 'schools\_info.csv'.
  - iii. Download the MMQGIS plugin on QGIS and navigate to Geocode > Geocode CSV with web service. Select schools\_info.csv for input CSV file, and OpenStreetMap/ Nominatim for geocoding web service. For output file name create a new folder 'found' to save geocoded shapefile. For the Not Found output list, name it not-found.csv.



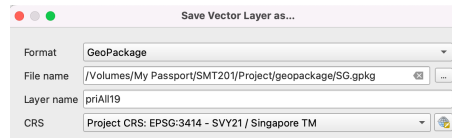
- iv. The resulting 'found.shp' should contain only 324 records once loaded onto the layers panel. Open the attribute table and select the primary schools using these queries using the Select by expression icon (  )



- v. Save the selected features as csv (Export > Save selected features as). Name it 'priAll19-173'
  - vi. Open the not-found.csv file, copy over the details of the following primary schools into 'priAll19-173.csv'. As for the column 'latlong', for each of the following primary schools, retrieve the latlong coordinates from google maps.
- |                                    |
|------------------------------------|
| DAZHONG PRIMARY SCHOOL             |
| FERN GREEN PRIMARY SCHOOL          |
| FRONTIER PRIMARY SCHOOL            |
| GEYLANG METHODIST SCHOOL (PRIMARY) |
| LIANHUA PRIMARY SCHOOL             |
| NAVAL BASE PRIMARY SCHOOL          |
| PEIYING PRIMARY SCHOOL             |
| PIONEER PRIMARY SCHOOL             |
| PRINCESS ELIZABETH PRIMARY SCHOOL  |
| RAFFLES GIRLS' PRIMARY SCHOOL      |
| SOUTH VIEW PRIMARY SCHOOL          |
| ST. ANTHONY'S PRIMARY SCHOOL       |
| TAO NAN SCHOOL                     |
- vii. Separate the latlong data into 2 columns 'lat' and 'long'. Select the 'latlong' column, Data > text to columns > choose the delimiter as comma. 2 new columns 'lat' and 'long' should be created. Save this csv named 'priAll19.csv'.
  - viii. Import 'priAll19.csv' into QGIS (layer > add layer > add delimited text layer ) and input the following details:



- ix. Save the layer priAll19 into geopackage format with the CRS EPSG3414



b. 2017 data

- i. Download 2017 primary school locations from (<https://github.com/hxchua/datadoubleconfirm/blob/master/datasets/priaryschools.csv> )

- ii. It should contain the 7 schools that were merged into other schools in 2019

BALESTIER HILL PRIMARY SCHOOL
EAST VIEW PRIMARY SCHOOL
DA QIAO PRIMARY SCHOOL
EAST COAST PRIMARY SCHOOL
CORAL PRIMARY SCHOOL
LOYANG PRIMARY SCHOOL
MACPHERSON PRIMARY SCHOOL

- iii. Save it into geopackage with EPSG3414 projection, and named 'priAll17'.

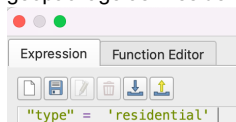
### 3. Data extraction for residential buildings from online sources

- a. Download the shapefile from Geofabrik

(<https://download.geofabrik.de/asia/malaysia-singapore-brunei.html> ) →

malaysia-singapore-brunei-latest-free.shp.zip

- b. Import the file 'gis\_osm\_buildings\_a\_free\_1.shp' into QGIS
- c. Clip the buildings that fall within singapore (vector > geoprocessing tools > clip). For input layer, select 'gis\_osm\_buildings\_a\_free\_1' layer. For overlay layer, select 'MP14 subzone'.
- d. A new temporary layer 'Clipped' should be added to layers panel. Open the attribute table, and select the residential type buildings. Save this layer into geopackage as 'Residential-buildings'.



- e. To get residential building centroids, use the plugin RealCentroid. For polygon layer, select 'Residential-buildings'. For output point on surface layer, create a folder and name the output file 'residential-centroids'. Save the newly created layer to geopackage named 'Resid-centroid'.

Commented [1]: KIV  
navigate to vector > geometry tools > centroids. Select 'Residential-buildings' for input layer and run. Save the generated layer into geopackage as 'Resid-centroid'.

### 4. Distance matrix for all merged schools

- a. From the priAll17 layer, select 14 schools affected by the 2019 merger ( <https://www.moe.gov.sg/-/media/files/news/press/2017/annex-a---named-of-pri-and-sec-schools-merging-in-2019.pdf> ). Save these selected features to geopackage named 'mergedPri'.
  - i. Save unmerged17: Select the 2017 schools from mergedPri and save them into a separate layer named unmerged17

BALESTIER HILL PRIMARY SCHOOL
EAST VIEW PRIMARY SCHOOL
DA QIAO PRIMARY SCHOOL
EAST COAST PRIMARY SCHOOL
CORAL PRIMARY SCHOOL
LOYANG PRIMARY SCHOOL
MACPHERSON PRIMARY SCHOOL

- ii. **Save merged19:** Select the 2019 schools from mergedPri and save them into a separate layer named merged19

BENDEMEER PRIMARY SCHOOL  
CEDAR PRIMARY SCHOOL  
DAMAI PRIMARY SCHOOL  
CASUARINA PRIMARY SCHOOL  
WHITE SANDS PRIMARY SCHOOL  
JUNYUAN PRIMARY SCHOOL  
JING SHAN PRIMARY SCHOOL

- b. Create 2km dissolved buffers around the 14 affected schools (vector > geoprocessing tools > buffer) using the mergedPri layer. Enter the fields below, ensuring the 'Dissolve result' checkbox is checked.

**Buffer**

Parameters Log

Input layer  
SG - mergedPri [EPSG:3414]

☐ Selected features only

Distance  
2000.000000 meters

Segments  
100

End cap style  
Round

Join style  
Round

Miter limit  
2.000000

☒ Dissolve result

Buffered  
[Create temporary layer]

☒ Open output file after running algorithm

- c. Select the centroids that fall within the 2km buffers (vector > research tools > select by location) according to the screenshot below and save selected features as 'ALL-study area'.

**Select by Location**

Parameters Log

Select features from  
SG - Resid-centroid [EPSG:3414]

Where the features (geometric predicate)  
☒ intersect ☐ touch  
☐ contain ☐ overlap  
☐ disjoint ☐ are within  
☐ equal ☐ cross




By comparing to the features from  
SG - all-2km buffers [EPSG:3414]

☒ Open output file after running algorithm

- d. To obtain distance matrix for unmerged17 layer, (vector > analysis tools > distance matrix). For input point layer, select the ALL-study area layer. For target point layer, select unmerged17 layer. For output matrix type, select summary distance matrix. A temporary layer 'Distance Matrix' will be created, rename this to 'Distance Matrix17'.




**Distance Matrix**

Parameters Log

Input point layer  
 ALL-study area [EPSG:3414]   

☐ Selected features only

Input unique ID field  
 123 fid


Target point layer  
 SG — unmerged17 [EPSG:3414]   

☐ Selected features only

Target unique ID field  
 123 fid

Output matrix type  
 Summary distance matrix (mean,std.dev.,min,max)

Use only the nearest (k) target points  
 0




Distance matrix  
 [Create temporary layer] 

☒ Open output file after running algorithm

- e. Repeat step 4d) with the merged19 layer. A temporary layer 'Distance Matrix' will be created, rename this to 'Distance Matrix19'




**Distance Matrix**

Parameters Log

Input point layer  
 ALL-study area [EPSG:3414]   

☐ Selected features only

Input unique ID field  
 123 fid

Target point layer  
 SG — merged19 [EPSG:3414]   

☐ Selected features only

Target unique ID field  
 123 fid

Output matrix type  
 Summary distance matrix (mean,std.dev.,min,max)

- f. Perform a relational join on 'Distance Matrix19', joining 'Distance Matrix17' into 'Distance Matrix19' on the fid as Join field and Target field.
- g. Next, on 'Distance Matrix19' create a new field named MIN DIFF with the following query to get the difference in minimum distance. Save this layer into geopackage named 'ALL-matrix diff'.

☒ Create a new field






☐ Create virtual field

Output field name  
 MIN DIFF

Output field type  
 1,2 Decimal number (real)

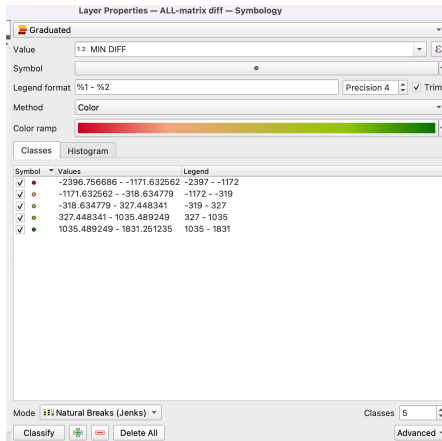
Output field length  
 0 Precision 3

Expression Function Editor

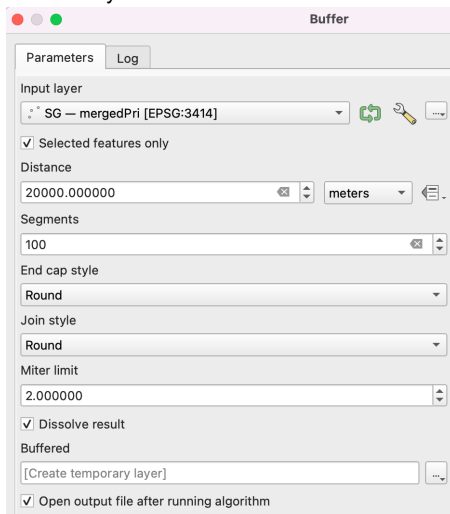
    

"Distance matrix17\_MIN" - "MIN"

- h. To symbolise 'ALL-matrix diff', (properties > symbology) select the graduated symbology and MIN DIFF as value field. Select Natural Breaks for symbology mode.



5. Distance matrix for bendemeer and balestier hill primary school
  - a. Select bendemeer and balestier hill primary school from 'mergedPri' layer and create 2km dissolved buffers around the 2 schools (vector > geoprocessing tools > buffer). Enter the fields below, ensuring the 'Selected feature only' and 'Dissolve result' checkboxes are checked.



- b. Select the centroids that fall within the 2km buffers using step 4c) and save selected features as 'BB-study area'.
  - c. Repeat steps 4d) to 4g) to obtain the MIN DIFF and name the final layer 'BB-matrix diff'.
  - d. Symbolise 'BB-matrix diff' using step 4h)

## Data preparation for choropleth map

### 1. Data extraction from online sources

- Population data can be extracted from singstat (<https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data>) → Singapore Residents by Planning Area / Subzone, Age Group, Sex and Type of Dwelling, June 2011-2020

### 2. Import the data

- Bring in the csv 'respopagesex2011to2020' : layer > add layer > add delimited text layer and select the csv to import.
- Ensure that 'No geometry' option is selected

**▼ Geometry Definition**

☐ Point coordinates

☐ Well known text (WKT)

☒ No geometry (attribute only table)

### 3. Extraction of relevant data

- Create a new virtual layer to extract the relevant data for 2017
- Layer > create layer > new virtual layer
- Name the layer 'pop2017-pri' and import the population csv 'respopagesex2011to2020' from the layers panel. Then enter the following SQL statement to extract the population of primary school children and total population in 2017

**Create a Virtual Layer**

Layer name:

Build a layer by referring to the current project's vector layer names and/or any configured embedded layer names as tables in the custom SQL query.

**▼ Embedded layers**

Local Name	Source	Provider
respopagesex2011to2020	pes=yes&geomType=none&subsetIndex=no&watchFile=no	Delimited text data provider

**Query**

```
1 SELECT temp1.PA AS PA, temp1.SZ AS SZ, PopYoung, PopSZ, temp1.Yr AS Yr FROM
2 (SELECT "PA","SZ", SUM("POP") AS PopYoung, "Time" AS Yr FROM respopagesex2011to2020 WHERE "Age" IN ('7',
3 '8','9','10','11','12') AND "Time" = '2017' GROUP BY "PA","SZ","Time") AS temp1
4 INNER JOIN
5 (SELECT "PA","SZ", SUM("POP") AS PopSZ, "Time" AS Yr FROM respopagesex2011to2020 WHERE "Time" = '2017'
6 GROUP BY "PA","SZ","Time") AS temp2
7 ON temp1.SZ = temp2.SZ AND temp1.PA = temp2.PA;
```

☐ Unique identifier column

**► Geometry**

- Click Add and a virtual table should be created on the layers panel named 'pop2017-pri'. Save this layer into geopackage format.
- Repeat steps 3a) to 3d) for 2019.


- i. Using this SQL query instead:

```
1 SELECT temp1.PA AS PA, temp1.SZ AS SZ, PopYoung, PopSZ, temp1.Yr AS Yr FROM
2 (SELECT "PA","SZ", SUM("POP") AS PopYoung, "Time" AS Yr FROM respagesex2011to2020 WHERE "Age" IN ('7',
3 '8','9','10','11','12') AND "Time" = '2019' GROUP BY "PA","SZ","Time") AS temp1
4 INNER JOIN
5 (SELECT "PA","SZ", SUM("POP") AS PopSZ, "Time" AS Yr FROM respagesex2011to2020 WHERE "Time" = '2019'
6 GROUP BY "PA","SZ","Time") AS temp2
7 ON temp1.SZ = temp2.SZ AND temp1.PA = temp2.PA;
```

- ii.

- f. Current layers should be 'pop2017-pri' and 'pop2019-pri' saved into geopackage.

#### 4. Data manipulation - calculating percentage of primary school children

- a. Open the data table for pop2017-pri layer.
- b. Open the field calculator () and create a new field named 'SUBZONE\_N' that capitalises the subzone names. Ensure the output field type is Text(string)

☒ Create a new field


☐ Create virtual field

Output field name: SUBZONE\_N

Output field type: abc Text (string)

Output field length: 0 Precision: 3

Expression Function Editor



upper("SZ")

- i.

- c. Next create another field that calculates the percentage of primary school children and ensure the output field type is Decimal Number (real).

☐ Only update 0 selected feature(s)

☒ Create a new field


☐ Create virtual field

Output field name: PER

Output field type: 1.2 Decimal number (real)

Output field length: 0 Precision: 3

Expression Function Editor




round("PopYoung" / "PopSZ" \* 100, 2)

- i.

- d. Next, for the newly created PER column, convert the cells with NULL values to 0. By checking the update existing field checkbox.

Expression Function Editor



if("PER" is null, 0, "PER")

- i.

- e. Save and close the attribute table which should have the following columns




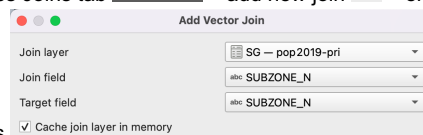
	fId	PA	SZ	PopYoung	PopSZ	Yr	SUBZONE_N	PER
1	234	Seletar	Seletar	0	50	2017	SELETAR	0
2	172	Orchard	Tanglin	10	220	2017	TANGLIN	4.55
3	28	Bukit Batok	Brickworks	0	0	2017	BRICKWORKS	0
4	276	Tampines	Tampines No...	0	0	2017	TAMPINES N...	0
5	273	Sungei Kadut	Turf Club	10	530	2017	TURF CLUB	1.89
6	117	Jurong East	Jurong Gate...	0	90	2017	JURONG GAT...	0
7	210	Queenstown	One North	50	570	2017	ONE NORTH	8.77
8	91	Downtown C...	Bugis	0	790	2017	BUGIS	0
9	320	Yishun	Yishun Central	70	1520	2017	YISHUN CEN...	4.61

- i.
- f. Repeat steps 4a) to 4e) for the layer pop2019-pri

## 5. Data manipulation - calculating percentage difference

- a. Duplicate the layer pop2017-pri from the layers panel and rename this layer as 'perDiff- 17-19'.
- b. Open the attribute table for 'perDiff- 17-19' and perform a relational join on the SUBZONE\_N field by navigating as follows:

- i. From properties Joins tab  > add new join  > enter the



Join layer: SG — pop2019-pri

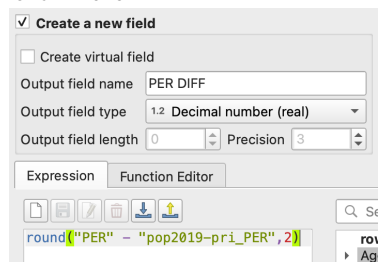
Join field: abc SUBZONE\_N

Target field: abc SUBZONE\_N

☒ Cache join layer in memory

following fields

- ii. Columns from 'pop2019-pri' should be added to pop2017-pri attribute table
- c. Create a new field named 'PER DIFF' and enter the following query that calculates the difference between the percentage of primary school children in 2017 and in 2019.



☒ Create a new field

☐ Create virtual field

Output field name: PER DIFF

Output field type: 1.2 Decimal number (real)

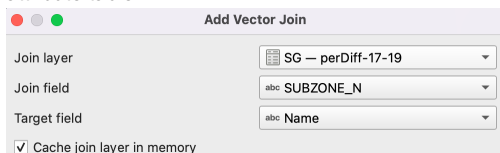
Output field length: 0 Precision: 3

Expression: round(("PER" - "pop2019-pri\_PER", 2)

- i.

## 6. Creating choropleth maps

- a. Perform relational join on MP14 subzone boundary layer 'MP14 subzone' and 'perDiff- 17-19' using steps 5b), joining on the subzone name as shown below. Columns from 'perDiff- 17-19' should be added to the 'MP14 subzone' attribute table.



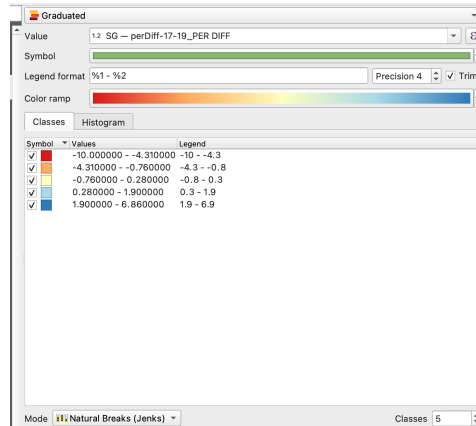
Join layer: SG — perDiff-17-19

Join field: abc SUBZONE\_N

Target field: abc Name

☒ Cache join layer in memory

- b. Symbolise the 'MP14 subzone' layer using the Graduated symbology, selecting the PER DIFF field and Natural Breaks (Jenks) classification mode.



#### 7. Creating 1000 x 1000 Grid

- Select Vector > Research Tools > Create Grid
- For Grid Type, select Hexagon
- For Grid Extent, select the MP14 subzone
- For Horizontal and Vertical spacing, input 1000
- Run the program
- Next, select Vector > Geometry Tools > Centroids, and select the input layer as the Grid Layer that was just created.
- Clip both the Grid Layer and the Centroid Layer to the boundary of MP14 Subzone.
- Save the Clipped Grid Layer as 'Hex Singapore' and the Clipped Centroid Layer as 'Centroids Singapore'

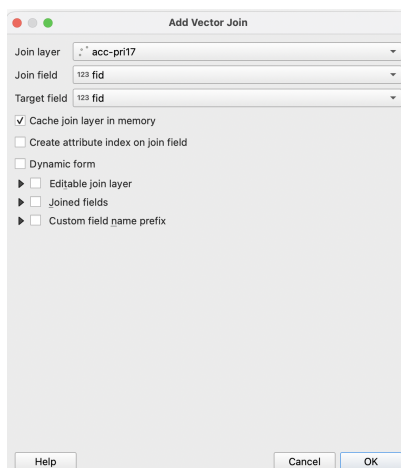
#### 8. Calculating the Distance Matrix across the entire Singapore

- Using the QNEAT3 plugin, select 'OD Matrix from Layers as Table'
- For the Network Layer, import the 'roads\_free\_1.shp' from OSM, and clip it to the boundary of MP14.
- Use this as the Network Layer.
- For the From Point layer, select the centroids obtained from the above.
- For To-Point Layer, select from 'priAll17' the drop-down list. For Unique Point ID Field, select fid from the drop-down list. For Optimization Criterion, select Shortest Path (distance optimization) from the drop-down list. For Entry Cost calculation method, select Ellipsoidal from the drop-down list. For Direction field, select oneway from the drop-down list. For Value for forward direction, type F. For Value for backward direction, type T. For Value for both direction, type B.
- Run the algorithm.
- Next, we will use the SQL tool of QGIS to select destination points with the shortest distance.
- At the Search pane of Processing Toolbox, type SQL.

- i. SQL function appears on the list.
- j. Double-click on Execute SQL of Vector general.
- k. Execute SQL dialog window appears.
- l. For Additional input datasources, select on the button at the right end.
- m. Click on the checkbox Output OD Matrix.
- n. Click on OK button.
- o. At SQL query panel, type the following SQL
- p. For Geometry type, select No Geometry from the drop-down list.
- q. Notice that a temporary table called SQL Output is added onto Layers panel. It consists of four fields. The values in shortest\_distance field are shortest distance between demand points and its nearest primary schools.
- r. Save this layer as 'acc-pri17' in the geopackage
- s. Repeat the above steps for priAll19 and save the layer as 'acc-pri19' in the geopackage.

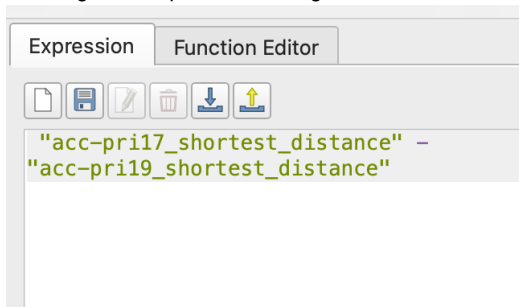
#### 9. Obtaining the Hexagons in Singapore

- a. Using the hexagon layer obtained above, perform a join on the fid of acc-pri17.










- b. Likewise, perform a join on the fid of acc-pri19.

- c. Under symbology, select Graduated. For value, click on the summation sign at the right and input the following formula.



Expression    Function Editor

`"acc-pri17_shortest_distance" -  
"acc-pri19_shortest_distance"`

- d. For color ramp, select Cividis and classify via equal count.

