**Analyse COVID-19 risk using ArcGIS Pro**

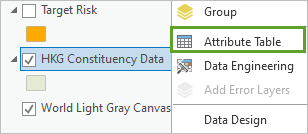
Data source ESRI China

Coronavirus disease 2019, or COVID-19, poses a threat to global health. While researchers work to distribute a vaccine, the rest of the world does their best to prepare for, contain, and help those suffering with the virus. To assist these efforts, you'll create risk maps for transmission, susceptibility, and resource scarcity. You'll also create a map of risk profiles to help pinpoint targeted intervention areas.

**Examine the data**

First, you'll download the ArcGIS Pro project package and explore the data.

1. Double-click the COVID19RiskData.ppkx file to open the project in ArcGIS Pro.
2. In the Contents pane, right-click the HKG Constituency Data layer and choose Attribute Table.

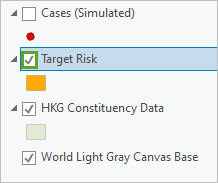


The table has data for 431 Hong Kong subdivisions (also called constituencies). The data includes demographic and spatial information that can be used to assess the risk of disease transmission, COVID- 19 susceptibility, and healthcare resource scarcity.

The project includes several layers:

* + The Total Population 2018, Pop Density per SqKM 2018, Seniors (60+) per 1000 people, 2018 Tobacco: Index, and 2018 Purchasing Power: Index fields were added using the [Enrich](https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/enrich.htm) tool.
  + For the 2018 Tobacco: Index and 2018 Purchasing Power: Index fields, values less than 100 are below average for Hong Kong, while values more than 100 are above average. The tobacco index will be useful for identifying risk because COVID-19 is a respiratory disease, while the purchasing power index will help determine income and poverty levels.
  + The Healthcare Resource Index field reflects the number of hospital beds and reflects a constituency's ability to respond to and treat COVID-19 cases.
  + The Spatial Interaction Index field was computed based on road network connectivity.
  + The Relative Case Distance field indicates the distance (in meters) from each constituency's centre point to the closest 10 percent of the simulated COVID-19 cases.

1. Close the table.
2. In the Contents pane, check the Target Risk layer to turn it on.



1. Right-click the Target Risk layer and choose Attribute Table.

The attributes for the Target Risk layer are similar to those for the HKG Constituency Data layer. The values in the Target layer, however, are the worst-case values found throughout Hong Kong. They include the densest population value, the largest number of seniors per 1,000 people, the highest index for tobacco, the smallest purchasing power value, and so on. These worst-case values will be used as the target against which all other constituencies will be ranked in order to determine risk.

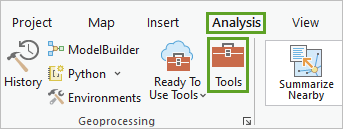
1. Close the table. Uncheck the Target Risk layer.
2. In the Contents pane, turn on the Cases (Simulated) layer.

There were 71 known COVID-19 cases in Hong Kong between January 22, 2020, and February 23, 2020. The locations and dates associated with each case have been fictionalized for this exercise.

**Map transmission risk**

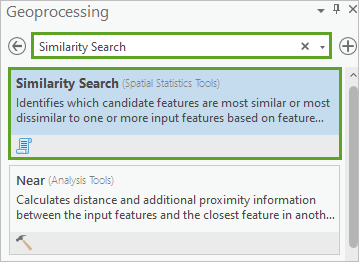
The highest potential for disease transmission is in locations with dense populations and lots of spatial interaction. You'll use the [Similarity Search](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/similarity-search.htm) tool to create a transmission risk map using the population density variable and the spatial interaction index.

1. On the ribbon, click the Analysis tab. In the Geoprocessing group, click the Tools button.



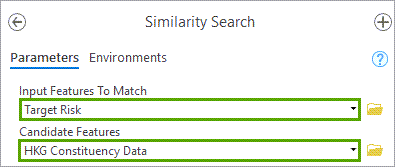
The Geoprocessing pane appears.

1. In the Geoprocessing pane, in the search bar, type Similarity Search and press Enter. In the result list, click the Similarity Search tool to open it.

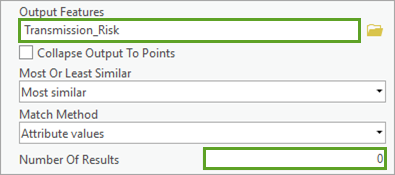


This tool determines which candidate features are most and least similar to an input feature based on attribute values. You'll use it to determine which constituencies are most similar to the worst-case Target Risk feature.

1. In the Similarity Search tool pane, for Input Features to Match, choose the Target Risk layer. For Candidate Features, choose the HGK Constituency Data layer.



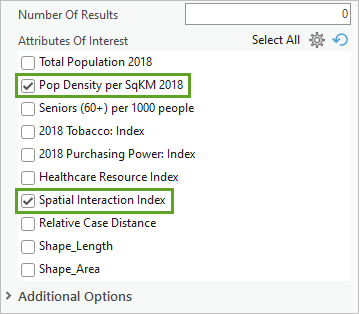
1. For Output Features, type Transmission\_Risk. For Number Of Results, type 0.



Note: The Collapse Output to Points parameter specifies whether the geometry for the Output Features parameter will be collapsed to points or will match the original geometry (lines or polygons) of the input features if the Input Features To Match and Candidate Features parameter values are both either lines or polygons. The Collapse Output to Points parameter is only available in ArcGIS Pro Advanced. Learn more about the [Similarity Search Parameters](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/similarity-search.htm#GUID-934752CB-3675-43B9-8E7F-C34F2083B569) tool.

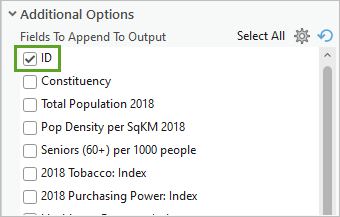
By setting the number of results to 0, you're telling the tool to rank all of the candidate features. Next, you'll choose the attributes of interest. For transmission risk, you'll choose the population density and spatial interaction index fields.

1. For Attributes Of Interest, check Pop Density per SqKM 2018 and Spatial Interaction Index.



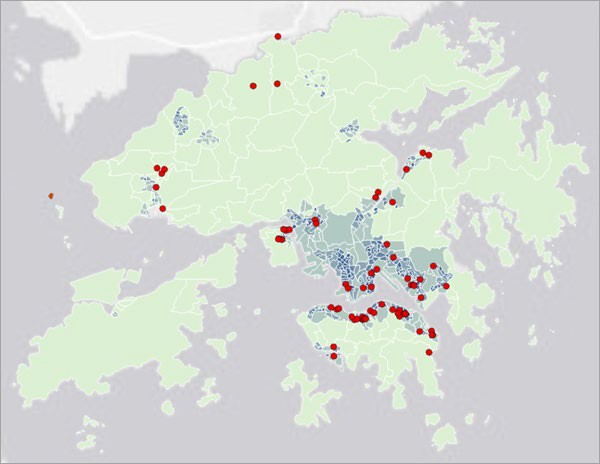
You'll also append an ID field to the output.

1. Expand Additional Options and check ID.



1. Click Run.

The tool runs and a new layer is added to the map. Each constituency is ranked from 1 (most like the worst-case feature) to 431 (least like the worst-case feature). Features with darker colors have values for population density and spatial interaction most like those in the Target Risk layer. You can compare the highest-risk areas to the simulated location of COVID-19 cases.



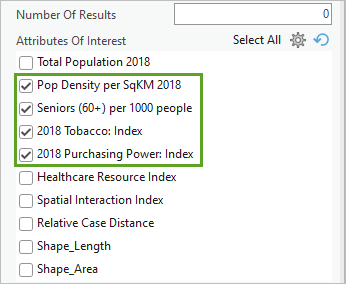
Efforts to minimize transmission (such as encouraging the use of facial masks, increasing hand sanitizing stations, cancelling large group events, and dispersing information about best practices for reducing transmission) would be most important in the places associated with the darkest regions of the map.

*What additional practices might be effective in the highest transmission risk areas? Are there practices, policies, or measures that are unique to your own community?*

**Map susceptibility risk**

Many people who contract COVID-19 will have mild symptoms, and most people will recover completely. Most children and adults recover well, but mortality rates are high for [older adults](https://www.cdc.gov/aging/covid19/) and those suffering from existing chronic illnesses or other health factors like smoking. Risk increases where large numbers of susceptible people live in densely populated areas. Using the Similarity Search tool, you'll create a risk map showing where death rates could be highest. You already used the tool in the previous step, so you only need to adjust some of the parameters.

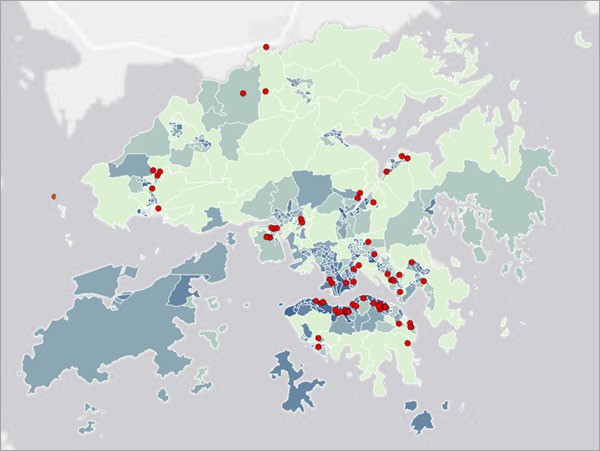
1. In the Similarity Search tool, for Output Features, type Susceptibility\_Risk.
2. For Attributes Of Interest, check Seniors (60+) per 1000 people, 2018 Tobacco: Index, and 2018 Purchasing Power: Index. Keep Pop Density per SqKM 2018 checked, and uncheck Spatial Interaction Index.



The remaining fields, including the input features, candidate features, and number of results, should be set correctly from when you previously ran the tool.

1. Click Run.

The tool runs, and a new layer is added to the map. Similar to the transmission risk layer, the susceptibility risk layer ranks each constituency by its similarity to the worst-case values for each of the attributes of interest. Darker-coloured constituencies on the map likely have higher numbers of seniors per 1,000 people, larger amounts of tobacco spending, and lower purchasing power.



Efforts to minimize impacts to the most vulnerable populations (restricting access to nursing homes, establishing quarantine centres, limiting travel, and working from home) would be most important in the places associated with the darkest regions of the map.

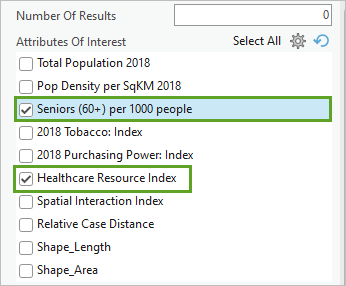
*What additional efforts, policies, or interventions might help reduce risk to the most vulnerable populations?*

**Map healthcare resource scarcity risk**

Another concern regarding the COVID-19 outbreak is that healthcare resources will become strained, possibly increasing the outbreak's negative effects. You'll run the Similarity Search tool again to determine the areas at highest risk for healthcare resource scarcity. This time, you'll use the Healthcare Resource Index field as an attribute of interest. As a proxy for people who would be most severely impacted if they contracted COVID-19 (and thus place higher strain on healthcare resources), you'll use the Seniors (60+) per 1000 people field.

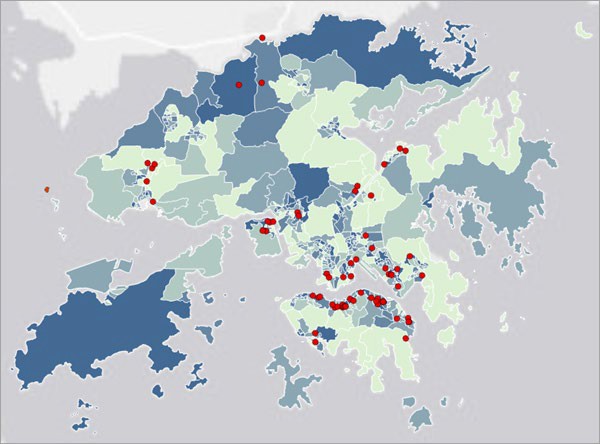
1. In the Similarity Search tool, for Output Features, type Insufficient\_Resource\_Risk.
2. For Attributes Of Interest, check Healthcare Resource Index. Uncheck Pop Density per SqKM 2018, 2018 Tobacco: Index, and 2018 Purchasing Power: Index.

Only the Seniors (60+) per 1000 people and Healthcare Resources Index attributes are checked.



1. Click Run.

The tool runs and a new layer is added to the map. The darkest areas are the places least prepared to deal with large numbers of COVID-19 cases among the most vulnerable populations.



If COVID-19 begins to spread rapidly, it's possible healthcare resources will be quickly overwhelmed. Having a plan in place to ramp up healthcare resources (such as proposed quarantine sites, test kits, ventilators, and protective clothing and masks) will be essential.

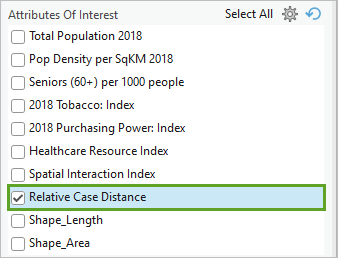
*What additional efforts, policies, or strategies would be important in places with insufficient resources in the case of rampant COVID-19 outbreak? Specifically, what can be done ahead of time to prepare for potential mass outbreak?*

**Map exposure risk**

Next, you'll create a map showing areas with the highest risk of exposure to COVID-19. The Relative Case Distance field is the summed distance from each constituency centroid to the closest 10 percent of all COVID-19 cases. Constituencies closer to large numbers of known cases have a higher risk of exposure than those farther away.

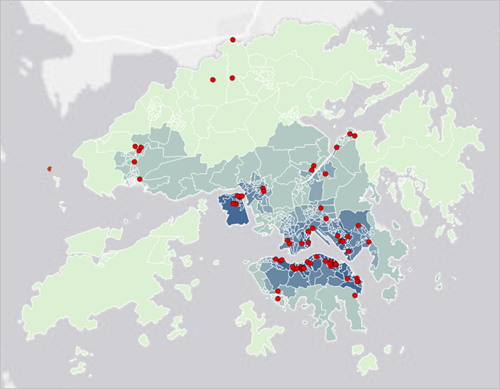
1. In the Similarity Search tool, for Output Features, type Exposure\_Risk.
2. For Attributes Of Interest, check Relative Case Distance. Uncheck Seniors (60+) per 1000 people and Healthcare Resources Index.

Only Relative Case Distance is checked.



1. Click Run.

The tools runs and a layer is added to the map.

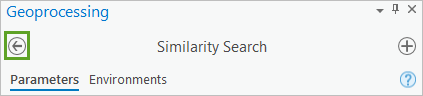


You now have four layers that rank risk factors for COVID-19.

**Map risk profiles**

Next, you'll create a map of risk profiles that combine all of the risk factors you've analyzed. Your final map will show areas that face similar challenges regarding COVID-19 and can be used to develop targeted interventions. First, you'll add the rankings from all four layers to the HKG Constituency Data layer. To do so, you'll add fields for each of the four rankings to the attribute table.

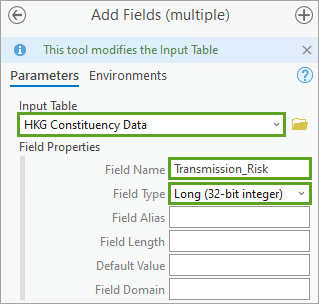
1. In the Geoprocessing pane, click the Back button.



1. Search for and open the Add Fields (multiple) tool.

Note: Learn more about the [Add Fields (multiple)](https://pro.arcgis.com/en/pro-app/latest/tool-reference/data-management/add-fields.htm) tool.

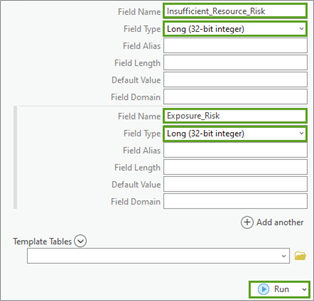
1. In the Add Fields (multiple) tool pane, for Input Table, choose HKG Constituency Data. For Field Name, type Transmission\_Risk, and for Field Type, choose Long (32-bit integer).



1. Click the Add another button.

Six new parameters appear.

1. For Field Name, type Susceptibility\_Risk. For Field Type, choose Long (32-bit integer).
2. Add two more fields, one with a Field Name of Insufficient\_Resource\_Risk and a Field Type of Long (32-bit integer), and one with a Field Name of Exposure\_Risk and a Field Type of Long (32-bit integer).



The Add Fields tool now includes four new fields that will be added to the table.

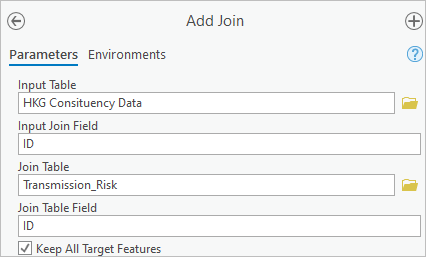
1. Click Run.

The tool runs and four fields are added to the attribute table of the HKG Constituency Data layer.

1. Open the attribute table for the HKG Constituency Data layer to confirm that the fields were added to the end of the table.

By default, the fields are empty and have no values. You'll join the rankings from the layers you created to the HKG Constituency Data layer. When you ran the Similarity Search tool, you made sure to append an ID field to the results. You'll use that ID field to perform the joins.

1. In the Geoprocessing pane, click the Back button. Search for and open the Add Join tool.
2. In the Add Join tool pane, enter the following parameters:
   * For Input Table, choose HKG Constituency Data.
   * For Input Join Field, choose ID.
   * For Join Table, choose Transmission\_Risk.
   * For Output Join Field, choose ID.



A warning appears for the Input Join Field parameter. This warning states that the chosen field isn't indexed, which may reduce tool running speed. The tool runs quickly either way, so you'll ignore the warning.

1. Click Run.

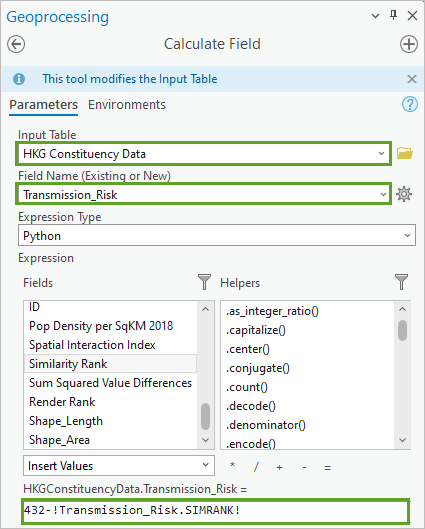
The tool runs and the tables are joined. If you still have the HKG Constituency Data table open, you'll see several new fields added, not just the field that contains the ranking.

Next, you'll calculate the Transmission\_Risk field that you added with the appropriate ranking values and then remove the join.

1. In the Geoprocessing pane, click the Back button. Search for and open the Calculate Field (Data Management Tools) tool.

All of the rankings are on a scale from 1 to 431, with 1 being the highest risk. You'll reverse the rankings so that higher numbers correspond to higher risk.

1. In the Calculate Field tool, set the following parameters:
   * For Input Table, choose HKG Constituency Data.
   * For Field Name (Existing or New), choose Transmission\_Risk.
   * For Expression, create the expression 432 - !Transmission\_Risk.SIMRANK!.

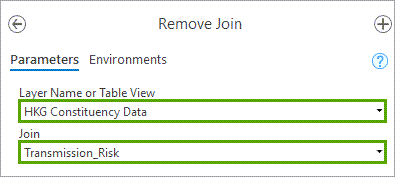


The expression will subtract the rankings by 432, which will lead to the highest risk ranking being 431 and the lowest being 1.

1. Click Run.

The field is calculated. Next, you'll remove the join.

1. In the Geoprocessing pane, click the Back button. Search for and open the Remove Join tool.
2. In the Remove Join tool pane, for Layer Name or Table View, choose HKG Constituency Data. For Join, choose Transmission\_Risk.

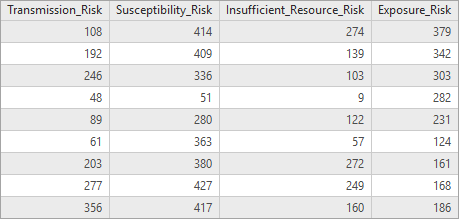


1. Click Run.

The join is removed. Next, you'll repeat the process for the other three risk layers.

1. Run the Add Join tool, the Calculate Field tool, and the Remove Join tool using the same parameters as before, but replace the following:
   * In the Add Join tool, for Join Table, choose Susceptibility\_Risk.
   * In the Calculate Field tool, for Field Name, choose Susceptibility\_Risk and for the expression, type 432 - !Susceptibility\_Risk.SIMRANK! as the expression.
   * In the Remove Join tool, for Join, choose Susceptibility\_Risk.
2. Repeat the process for the Insufficient\_Resource\_Risk and Exposure\_Risk layers.

All four of the fields you added have been calculated.



1. Close the table.

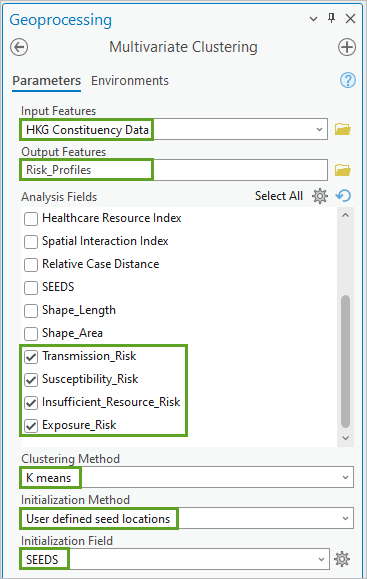
Next, you'll cluster constituencies with similar characteristics based on the four fields using the Multivariate Clustering (Spatial Statistics) tool.

**Use the Multivariate Clustering tool**

The Multivariate Clustering tool creates groupings such that the values within each group are as similar as possible, and the groups themselves are as different as possible.

Note: Learn more about the [Multivariate Clustering (Spatial Statistics)](https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/multivariate-clustering.htm) tool.

1. On the Analysis tab, in the Geoprocessing group, click Tools.
2. In the Geoprocessing pane, search for and open the Multivariate Clustering tool.
3. In the Multivariate Clustering tool pane, set the following parameters:
   * For Input Features, choose HKG Constituency Data.
   * For Output Features, type Risk\_Profiles.
   * For Analysis Fields, check Transmission\_Risk, Susceptibility\_Risk, Insufficient\_Resource\_Risk, and Exposure\_Risk.
   * For Clustering Method, choose K means.
   * For Initialization Method, choose User defined seed locations.
   * For Initialization Field, choose SEEDS.

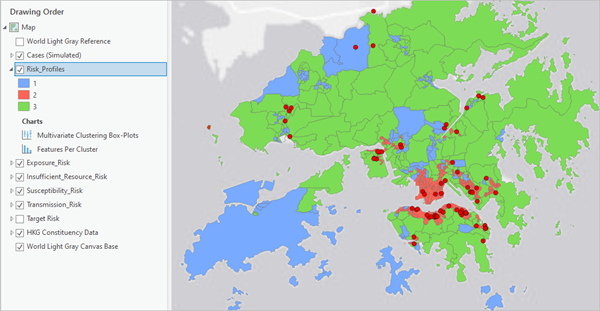


The SEED field marks (with a value of 1) the constituency that has the highest transmission risk, susceptibility risk, and exposure risk. The Multivariate Clustering tool is a heuristic; it hones in on an optimal result but doesn't guarantee the very best result possible. By setting seed values, the tool starts its search for an optimal result with the risk extremes. This has the advantage of efficiency but also ensures you will get the exact same result every time you run the tool. When you don't use seed values, you are likely to get the same result, but the colours associated with each group will likely be different (swapped).

There are three seeds, so the tool will find three clusters. This is appropriate for identifying high, medium and low-risk groupings. With three clusters, you'll create one cluster of high-risk groupings, one cluster of medium-risk groupings, and one cluster of low-risk groupings.

1. Click Run.

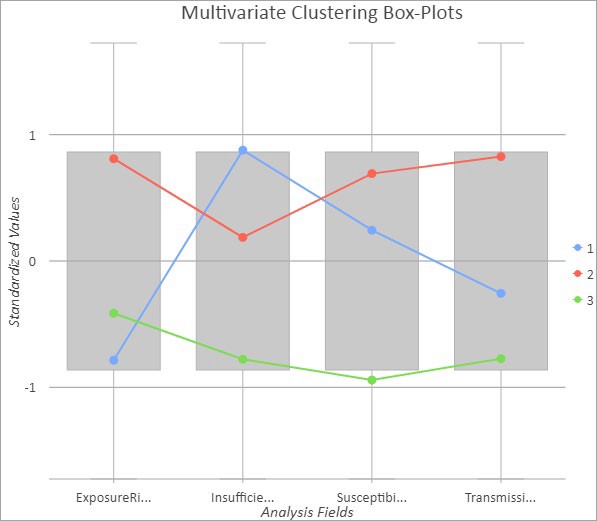
The tool runs and a new layer is added to the map.



The tool also created a box-plot chart that contains the characteristics of each profile.

1. In the Contents pane, double-click the Multivariate Cluster Box-Plots chart to open it.

The chart contains three lines that correspond to the clusters on the map. The nodes on the lines indicate whether the risk is relatively high or low for each category.



Because the rankings were reversed (432 – ranking), the largest values (at the top of the chart) represent the highest risk. In this tutorial example, the constituencies shown in red have the second highest risk for insufficient resources and are the highest risk for exposure, susceptibility risk, and transmission risk. These locations should be the highest priority for interventions that minimize person-to-person interaction: keeping children home from school, limiting visits to senior communities and hospitals, cancelling events, and encouraging work from home.

The blue group is also of concern because those constituencies have the highest risk of insufficient resources and the second highest risk for susceptibility and transmission risk. In addition to practising social distancing, these constituencies will benefit from putting plans in place for quarantine centres and healthcare training.

1. Close the chart. Save the project.

In this tutorial, you analysed the risk of transmission, susceptibility, healthcare resource scarcity, and exposure in the face of the COVID-19 pandemic. You also created profiles of similar risk groups, which will allow officials to plan targeted intervention programs.