Lab Exercise #10: Creating a composite Index

CRP 5080: Introduction to GIS for planners
Fall 2024

Due: Monday,

Part 1: Create a composite Index

An index is a number that measures a subject of interest, often something that is difficult to directly measure or define, such as social vulnerability or business innovation. In this lab, we will create an environmental justice screening tool to determine which areas of the state of Ohio are experiencing the most environmental and social burdens compared to other areas of the state. The analysis question you want to answer is: What are the cumulative environmental and social burdens are communities experiencing across the state?

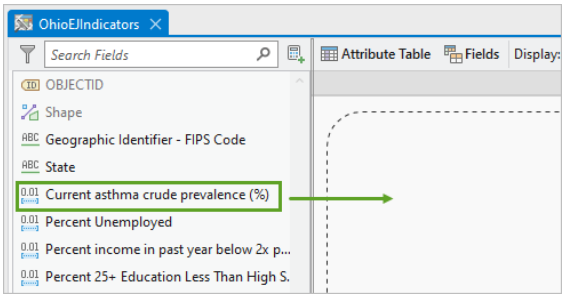
1. Explore Indicators

First, you will explore the indicators. Unzip the Ohio EJ Indicators geodatabase (OhioEJIndicators.gdb). Open the file and add the Ohio EJ Indicators feature class. Open the **Attribute Table**:

- Current asthma crude prevalence (%)
- Percent Unemployed
- Percent income in past year below 2x poverty level
- Percent 25+ Education Less Than High School
- PerSevHousingBurden
- ChildLeadRisk - Mean
- Distance to nearest park (miles)
- Avg PM2.5 2014--2016
- Sum Traffic
- Toxic Release Chemicals (lb/km2) within 1 mile

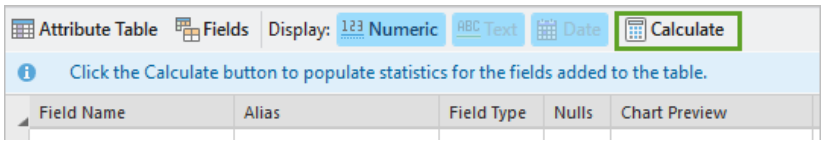
This list of indicators is not meant to be prescriptive or comprehensive for an environmental justice index. These 10 indicators will be used in this lab to serve as a learning resource on how to design and create a composite index. We will use the Data Engineering view to generate and investigate statistics on each of the indicator fields. Close the table.

- In the **Contents** pane, right-click the **OhioEJIndicators** layer and choose **Data Engineering**.
- Drag the **Current asthma crude prevalence (%)** field into the statistics panel.



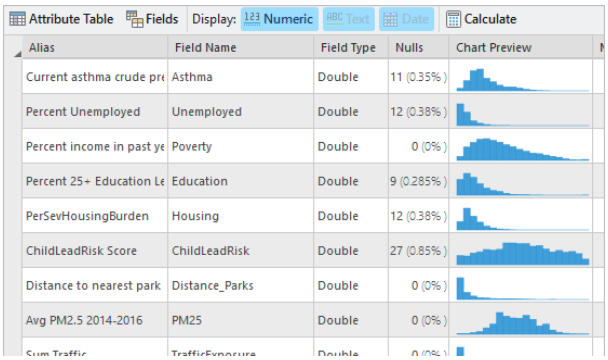
Next, you will add the remaining nine indicators into the statistics panel.

- At the top of the Data Engineering view, click **Calculate**.



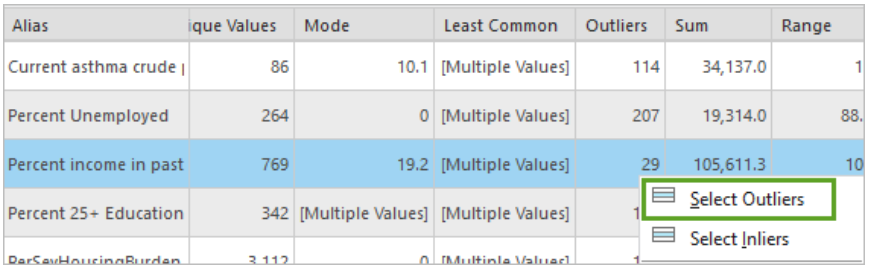
Statistics appear for each of the fields. Before you explore the data, you will freeze the **Alias** column so it continues to be visible as you scroll through the statistics panel.

- Right-click **Alias** and click **Freeze/Unfreeze**. The **Alias** column is now locked to the first column in the statistics panel.

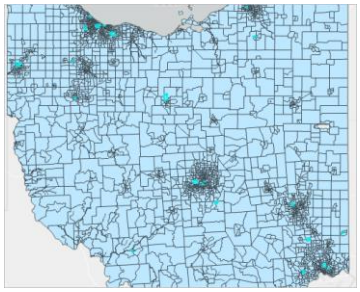


The **Chart Preview** column shows a histogram of each indicator's values. This is a helpful way to notice the data distribution and to consider what methods would be best to use for your composite index.

- Scroll to the **Outliers** column. In the **Outliers** column, you can view the number of data values that are outliers. You can also use the Data Engineering view to highlight them on the map so you can visualize where they are located.
- For the **Percent income in past year below 2x poverty level** row, right-click the **Outlier** value and choose **Select Outliers**.



The census tracts with outlier values for the **Percent income in past year below 2x poverty level** indicator highlight on the map.



Now clear the selection.

- You can also use the Data Engineering view to select values by quartiles for each indicator. In the Data Engineering view, in the statistics panel, scroll to the **Third Quartile** column.
- For the **Percent Unemployed** row, right-click the **Third Quartile** value, point to **Select**, and click **Above Quartile**.

Values above the third quartile represent values that are in the top 25 percent of all the values for a field. The map selects the census tracts with the highest 25 percent of unemployment rates.

The Data Engineering view also identifies any null values. scroll to the **Nulls** column. For the **Education** row, right-click the **Nulls** value, and click **Select Nulls**.

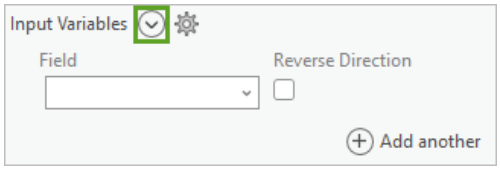
The tracts with null values for the **Percent 25+ Education Less Than High School** indicator highlight on the map. These tracts represent areas that do not have any population, so they have null values. The Calculate Composite Index tool will automatically ignore these records and exclude them from the index calculation.

Continue exploring the indicators. When you are finished, clear any selection and close the Data Engineering view.

2. **Use the Calculate Composite Index tool**

In the **Geoprocessing** pane, on the search bar, type *calculate composite index*. In the **Calculate Composite Index** tool pane, for **Input Table**, choose **OhioEJIndicators**. Under Output Features or Table, navigate to the OhioEJIndicatos.gdb and save your output as Ohio_EJIndex. By saving into an existing geodatabase, we will be able to avoid losing our existing column names.

Next to **Input Variables**, click the **Add Many** button.



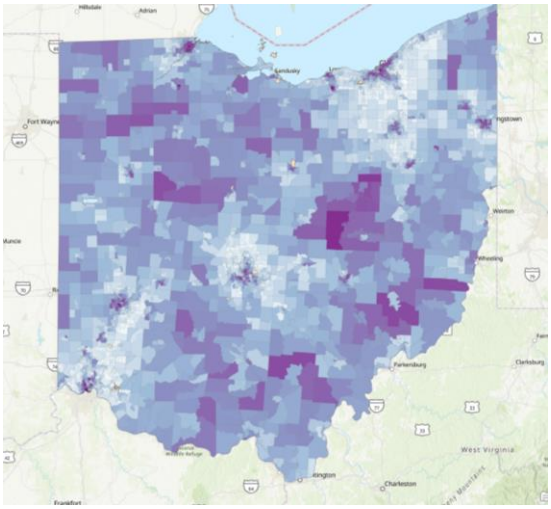
On the **Add Many** menu, check the 10 indicator fields and click **Add**.

- Next, you will choose the parameters for preprocessing and combining the indicators. You can use the **Preset Method to Scale and Combine Variables** parameter to choose common methods of scaling and combining the indicators.
- For now, you will keep the default selection - **Combine values (mean of scaled values)** - which is to combine values by mean of scaled values using the minimum to maximum method of scaling. Ensure **Preset Method to Scale and Combine Variables** is set to **Combine values (Mean of scaled values)**.
- Next, you will review the **Variable Weights** settings. Expand the **Variable Weights** section.

At this point, you want to treat each indicator with the same weight, because you are concerned about how they will combine and reflect the cumulative burden communities face. You will not change the weights.

Finally, you will configure the **Output Settings** options and postprocessing parameters. Expand the **Output Settings** section.

- For **Output Index Name**, type Ohio_EJIndex. For **Minimum**, type 0, and for **Maximum**, type 100.
- Click **Run**. The resulting composite index map appears.

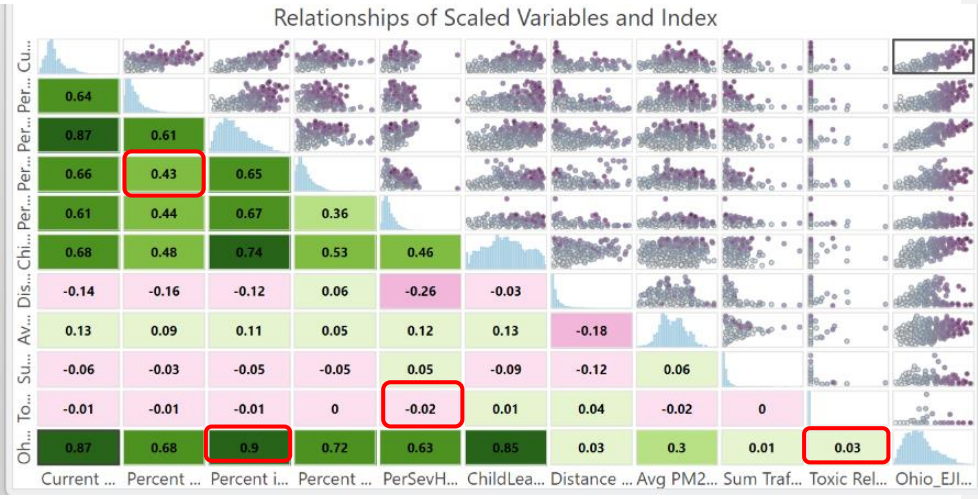


In the **Contents** pane, the **Ohio_EJIndex Layers** group layer contains all of the **Calculate Composite Index** tool outputs. Two maps were created. **Ohio_EJIndex - Mean** is styled by the raw index values scaled between 0 to 100. It also contains a number of charts for investigating the index output in more detail. **Ohio_EJIndex - Percentile** styles the index mean values by percentile values.

Before you continue, save the project.

3. Examine the relationship of scaled variables

The Calculate Composite Index tool produces resulting maps of the composite index as well as charts to help you evaluate and assess your index results. You will focus on the matrix relationship chart. In the Contents pane, double-click **Relationship of Scaled Variables and Index Variables** and open it.
The matrix chart appears. The most critical row to pay attention to is the bottom row, which shows any correlations between the index score and each indicator.



The green cells denote high correlations. In other words, as one variable goes up, so does the other. For example, the **Percent Unemployed and Percent below 2x the poverty level** indicators are 43 percent correlated.

The pink cells donate negative correlations. In other words, as one variable increases the other decreases. For example, the **Distance to toxic releases** indicator is negatively correlated with the **Percent severe housing burden** indicator (r = -0.02). This could be because housing burden is more correlated with urban areas and toxic releases may be more correlated with rural areas.

Another finding that stands out among the correlations between the index score and each indicator is that there are some very high values—0.90 for income, which means that the final index is 90 percent correlated with this indicator. There are also some very low values—0.03 for distance to toxic releases. Although you weighted each input equally, the correlations are different from each other. This could be due to the ranges of the variables or the correlation among the inputs. Note that the more environmental indicators tend to have a lower correlation with the index compared with the social indicators.

One way to gain more control of the indicators and the correlations among them is to group them into subindices. We will create two subindices (environmental and social) and combine them again in the next iteration of the index.

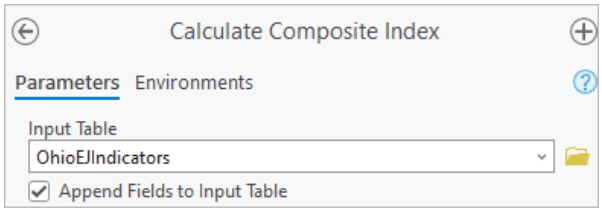
Note that the distribution of the **Toxic Release Chemicals** indicator values is especially interesting because most of the values are very low values and only a few are very large. This means that the minimum-maximum method of preprocessing may not be the best choice given the intense skew for this indicator. It may be better to use a percentile preprocessing method in your next iteration. Close the chart.

Part 2: Create subindices for the final index

We have created an environmental justice index by combining 10 indicators and preprocessed it using the minimum-maximum preprocessing method. Upon examining the resulting composite index, we determined that it would be better to use the percentile preprocessing method and to create subindices for each of the dimensions of the index.

You will prepare two subindices—one for the social vulnerability indicators and one for the environmental indicators. You will use the Calculate Composite Index tool to create each of the subindices.

- In the **Calculate Composite Index** tool pane, enter the following:
  - For **Input Table**, choose **OhioEJIndicators**.
  - Check the box for **Append Fields to Input Table**.



Since you will eventually combine the subindex scores together, it is recommended that you append the tool output to the input table instead of starting a new feature layer and table.

Next, you will add the indicators related to social vulnerability and outcomes.

For **Input Variables**, include the following fields:

- Current asthma crude prevalence (%)
- Percent Unemployed
- Percent 25+ Education Less Than High School
- PerSevHousingBurden
- Percent income in past year below 2x poverty level

For **Preset Method to Scale and Combine Variables**, choose **Combine ranks (Mean of percentiles)**.

Expand **Output Settings** and enter the following:

- For **Output Index Name**, type SV subindex.
- Under **Output Index Minimum and Maximum Values**, for **Minimum**, type 0.
- For **Maximum**, type 100.

Click **Run**. The **SV subindex** fields have been created and added to the attribute table in the **OhioEJIndicators** layer.

In the **Contents** pane, right-click the **OhioEJIndicators** layer and click **Attribute Table**. In the table, scroll until you see the **SV subindex** fields.

The fields that were added include the preprocessed values of the five social vulnerability indicators and the subindex mean values.

Next, you will use the Calculate Composite Index tool to create the environmental subindex. In the **Calculate Composite Index** tool pane, under **Input Variables**, add the fields related to environmental indicators:

- ChildLeadRisk Score
- Distance to nearest park
- Avg PM2.5 2014-2016
- Sum Traffic
- Toxic Release Chemicals

Confirm that **Preset Method to Scale and Combine Variables** is set to **Combine ranks (Mean of percentiles)**.

Under **Output Settings**, for **Output Index Name**, type ENV subindex.

Click **Run**. The fields for the ENV subindex are added to the **OhioEJIndicators** attribute table

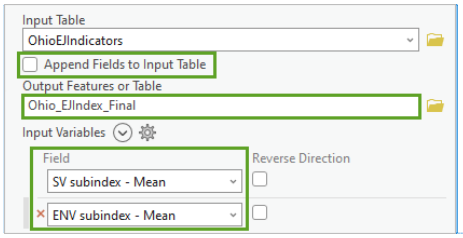
Part 3: Combine the subindices and examine results

In this section, you will use the Calculate Composite Index tool to combine the two subindices, adding a weight to the environmental subindex.

You no longer need to append the output to the **OhioEJIndicators** attribute table, so you will uncheck the **Append Fields to Input Table** option.

In the **Calculate Composite Index** tool pane, update the following parameters:

- Under **Input Table**, uncheck the box for **Append Fields to Input Table**.
- Under Output Features or Table, navigate to the OhioEJIndicatos.gdb and save your output as Ohio_EJIndex_Final.
- Under **Input Variables**, remove the existing fields and add **SV subindex - Mean** and **ENV subindex - Mean**.

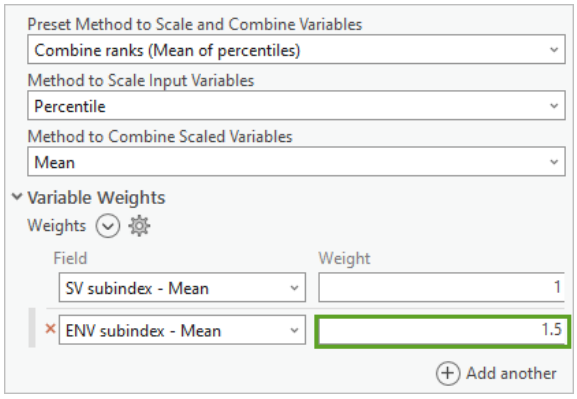


Commented [SJS1]: Calculate composite index only shows Index and Index_1 instead of SV subindex and ENV subindex....

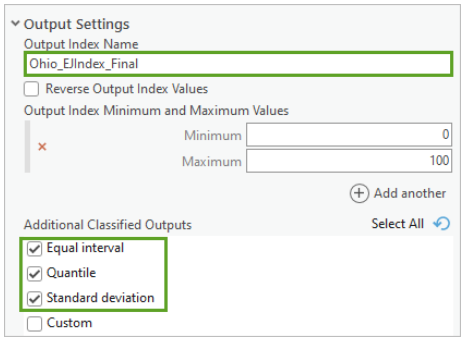
Earlier you observed that there was a high level of correlation between the social vulnerability indicators. Their strong correlation impacted the overall index score and caused the environmental related indicators to not contribute very much to the resulting index score.

By creating the subindices, you corrected for this. In addition to the use of subindices, your expert panel advising the development of this environmental justice screening tool recommended an additional 50 percent weight to the environmental factors.

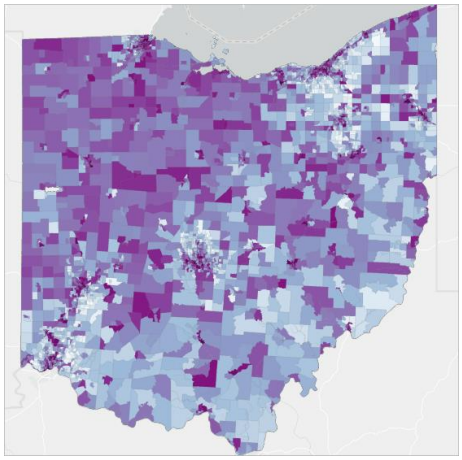
Expand the **Variable Weights** section. For the **ENV subindex - Mean** field, type 1.5.



Under **Output Settings**, for **Output Index Name**, type Ohio_EJIndex_Final. Under **Additional Classified Outputs**, check the box for **Equal interval**, **Quantile**, and **Standard deviation**.

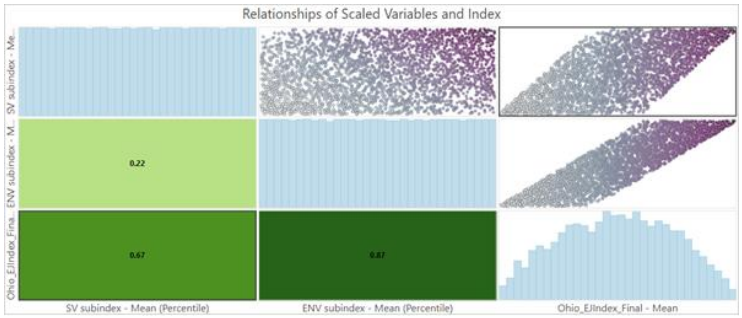


Click **Run**. The **Ohio_EJIndex_Final** layer appears on the map.



Next, you will review the matrix chart to better understand the correlations between the two subindices in the final index.

In the **Contents** pane, under the **Ohio_EJIndex_Final** layer, double-click the **Relationships of Scaled Variables and Index** chart. The chart appears.



Earlier, the matrix chart for the first index you created showed each variables' correlations to one another and to the index score. In this chart, you see each subindex's correlations.

The environmental subindex is more correlated with the final index than the social vulnerability subindex. This is expected because you added the weight to the environmental subindex.

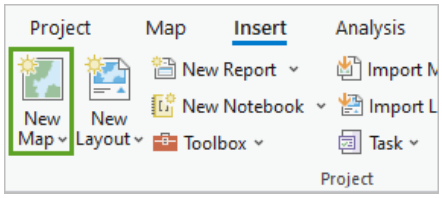
The blue bars are histograms of each subindex and the final index score. For the subindices, the distribution forms a flat line. This is the effect of using percentiles as opposed to minimum-maximum preprocessing. Using the percentile method does not preserve the original distribution.

Close the chart.

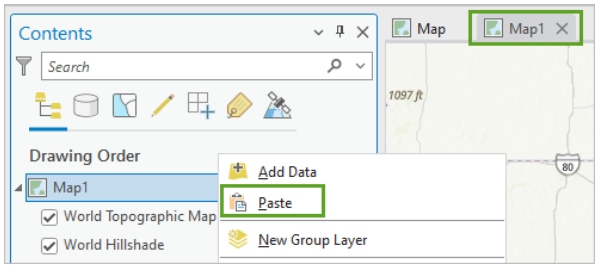
Next, you will compare this resulting index with the initial index you created that did not use subindices and preprocessed the indicators by minimum-maximum.

In the **Contents** pane, Right-click **Ohio_EJIndex** and click **Copy**.

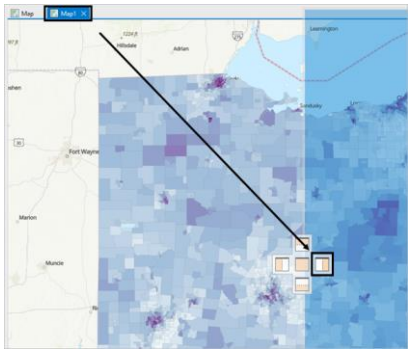
You will also turn off all the layers in this map except for the final index layer (like the OhioEJIndicators layer) On the ribbon, click the **Insert** tab. click **New Map**.



If necessary, click the Map1 view tab. In the **Contents** pane, right-click **Map1** and click **Paste**.

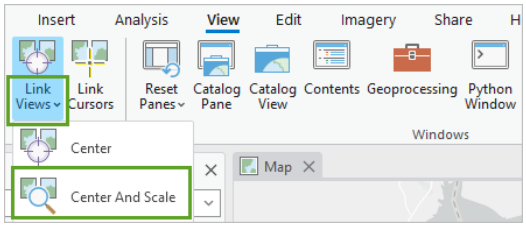


Next, you will dock **Map1** to the side of the first map so you can view them at the same time side by side. Drag **Map1** tab and drop it in the right dock of the **Map** tab.



You can now view the two maps side by side. Next, you will also link the view of the two maps so that their extent matches when you navigate either map.

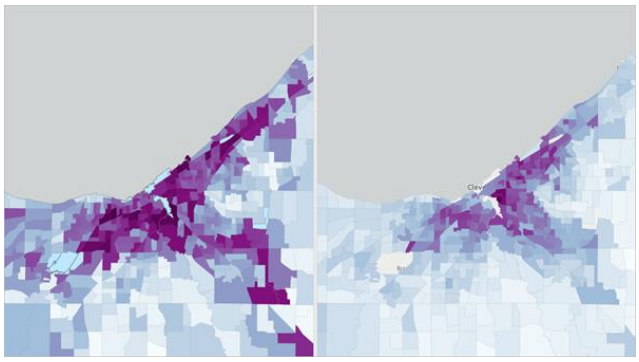
On the ribbon, click the **View** tab. click the bottom half of the **Link Views** button and choose **Center And Scale**.



Now the extent of the maps will match one another when you zoom or pan one of the maps.

Next, you will explore specific areas in Ohio to compare the two index results. On either map, zoom in to the city of Cleveland, located in the northeast edge of Ohio.

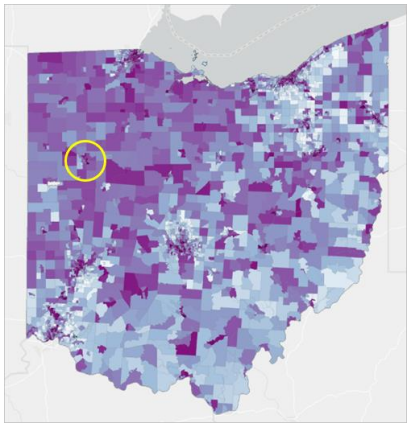
Seeing the two maps side by side, you can see how the results of the first index differ from the second iteration where you combined subindices.



Cleveland is one of the largest urban centers in the state. The initial index results, shown on the right, emphasize the social vulnerabilities. They correlated strongly with populated areas but do not appear to account as much for the potential environmental hazards and risks.

The second index, which provided more control over the environmental and social dimensions, appears to have captured more of the impacts of environmental exposure in the first map. There are many toxic release facilities along the border of the lake. It is also important to note that the revised index did not simply emphasize environmental hazards; it still maintained priority for areas that had high social vulnerability indicator values.

Zoom to the less populated city of Lima, Ohio.



The city of Lima contains the toxic release facility responsible for the most amount of chemicals released in the entire state. In the initial index results, shown in the second image, the index primarily emphasized areas of social vulnerability. The use of the subindex in the resulting index in the first image not only emphasized the substantial environmental impact that a single facility might contribute to the area, but it also still maintains the areas of social vulnerability in the dark purple areas.

Save the project

Homework

1. The index creation process is often a cycle where you create the index and learn something new, resulting in the need to add or remove a variable, adjust weights, or change the combining methodology. Experiment and create another version of the composite index. Weight the initial variables differently.
2. Use additional tools in ArcGIS Pro for further analysis of index results. Consider using Hot Spot Analysis or Cluster and Outlier Analysis to help identify significant areas of risk. Interpret your findings.