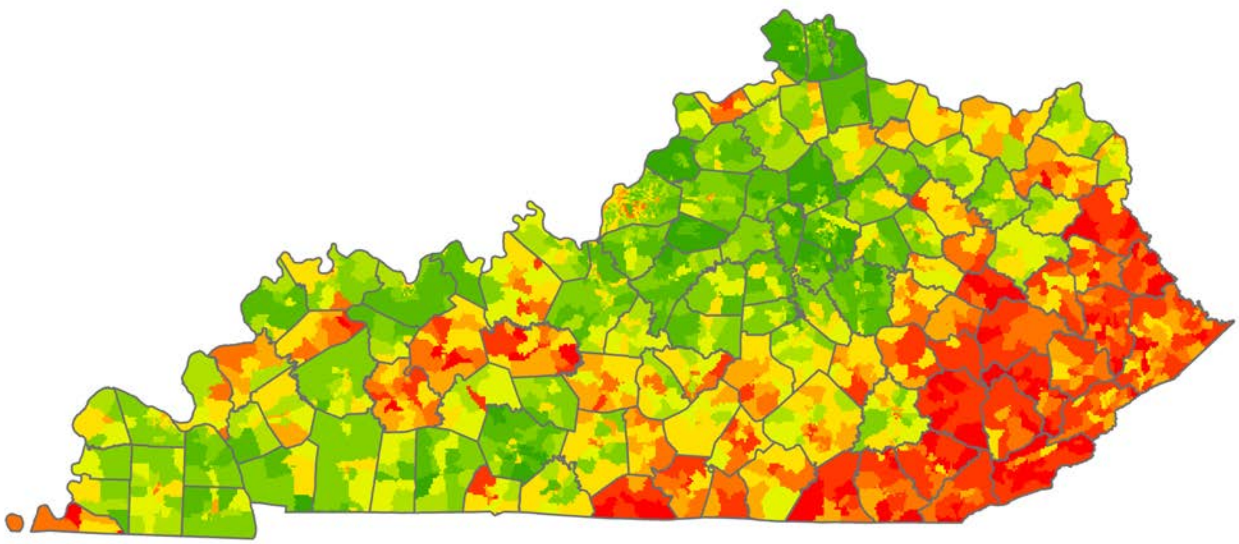
Empower Kentucky Environmental Justice Analysis Documentation



Version 1
November 2016

Kentuckians for the Commonwealth (KFTC),
New Energy and Transition Committee (NET),
Environmental Justice Work Team

Table of Contents

1) Introduction

1.1 Context and Overview of Empower KY's EJ Analysis

The Clean Power Plan (CPP), finalized in October 2015 by the U.S. Environmental Protection Agency (EPA), encourages, but does not require, states to conduct an environmental justice (EJ) analysis to identify vulnerable communities both within the current energy system, as well as during any clean energy transition. The EPA suggests states should use this analysis to prioritize public outreach and engagement with those communities and to ensure that the energy implementation plan, requiring states to reduce CO2 emissions from the electric power sector by 2030, does not result in increased pollution and disinvestment in these communities.

The EPA defines environmental justice as "the fair treatment and meaningful involvement of all people regardless of race, color, national origin, or income with respect to the development, implementation, and enforcement of environmental laws, regulations, and policies" (U.S. EPA 2016 *Tech Doc*). When Kentuckians for the Commonwealth (KFTC) took on the challenge of developing a clean power plan for all Kentuckians, a priority was to also challenge itself, as an organization and a clean energy movement, to prioritize justice, recognizing that, across the US, low-income and people of color communities face the greatest exposure to air and water pollution that is harmful to health. In Kentucky, a shift to cleaner energy sources can generate significant improvements in the health of Kentuckians, including lower rates of premature death, asthma, lung cancer, and COPD (chronic obstructive pulmonary disease). However, it's not inevitable that those benefits will occur in the most affected communities and not impossible that well-intentioned solutions could inadvertently increase pollution burden in the already disproportionately-affected areas. For these reasons, the Empower Kentucky Environmental Justice Analysis was born.

This documentation presents the preliminary product of the Empower KY EJ Analysis, consisting of, as the EPA classifies it in their EJ technical guidance (U.S. EPA 2016, *Tech Guidance)*, a visual display, featuring three main map products focusing on various types of vulnerability: Cumulative Pollution Exposure, Exposure-Related Health Problems, and Demographic Vulnerability, as well as individual maps for all the datasets included as pollution, health, and demographic indicators (see Section 5 "Visual Map Products"). As this is preliminary work, opportunity for further products to augment the analysis are discussed in Section 7 "Next Steps." These maps show visual concentrations of various pollutions and health problems stretching across the energy landscape of Kentucky, not just the coal landscape. In doing this, the maps examine the energy landscape as a *system.* Through this spatial analysis, KFTC and people across Kentucky can begin to identify the communities that are most vulnerable and stand to benefit the most from a clean energy transition.

The KFTC EJ work team, a sub-group of the New Energy Transition (NET) Committee, brought together a variety of experiences, skillsets, and perspectives to work on initial project shaping, data collection, data processing, public presentation, and product creation. From May to November 2016, this team met and worked to engage many voices, not just their own, to inform the project. This work also draws heavily on EPA documentation of EJSCREEN, the EPA EJ screening tool, and documentation for EJ Analysis in general (for reference citations, see Section 7 "Works Cited"), allowing the project to draw upon existing and robust resources. This analysis, although a preliminary look into a multi-dimensioned, complex topic, provides meaningful first-steps and a foundation with which to begin informing conversation, directing resources and actions, and recognizing the spatial inequity that has persisted systemically in our world for too long.

It is hoped that this document is used as a work of transparency, collaboration, and encouraged continued generation of environmental justice analysis and action in Kentucky. Here is an outline of what's in this document:

SECTION 1.2- *goals of the project*
SECTION 1.3 – *general process outline*
SECTION 2 – *Analytic Methods – choice of EPA-recommended best practices for conducting EJ analysis; information on Weighted Sum Overlay analysis conducted in ArcGIS; decisions made, such as scope, etc.; strengths/limitations of analysis*
SECTION 3 – *Data Details in Empower KY EJ Analysis – data details for indicators used overlay maps, including source, description, and processing workflow for each dataset*
SECTION 4 – *Data Discussion – commentary on quality/availability of data; data not included in the analysis*
SECTION 5 – *Empower KY EJ Visual Map Products – presentation of the EJ overlay maps and individual indicator maps, divided by category: Demographics, Pollution, and Health*
SECTION 6 – *Uncertainty and Bias Discussion – choices that were made and how the analysis is impacted*
SECTION 7 – *Next Step Research Priorities - opportunities for next steps and additional research opportunities*
SECTION 8 – *Works Cited – list of sources cited directly within this documentation, not comprehensive list of sources references throughout this project (in development and coming soon)*
APPENDIX 1 – *Processing Data for Weighted Sum Overlay*
APPENDIX 2 – *Generating an Overlay Using Weighted Sum Analysis*
APPENDIX 3 – *Calculating Proximity*

1.2 Goals of Empower KY's EJ Analysis

- Analyze sets of indicators to identify communities that should be prioritized as part of planning for a just transition to a clean energy economy in Kentucky
- Develop analysis which examines pollution and health indicators individually and overlaid and also seeks to identify communities where key demographic indicators intersect with pollution exposures and health indicators
- Make the data, maps, and methodologies transparent and publicly accessible and, ultimately, make the analysis/maps as interactive as possible
- Identify key research questions and indicators that should be considered by KFTC, other stakeholder groups, the state, and the EPA as part of a broader EJ analysis. Given the complexity

of this task and limited time and resources, we recognize that our initial analysis will be both meaningful and necessarily limited in scope

- Use analysis to shape policies that minimize harm and direct benefits to identified priority communities. Ideally, this analysis can inform decisions which will direct benefits such as health, jobs, job training, direct support for displaced workers, bill assistance, and Renewable Energy/Energy Efficiency investments
- Be intentional, inclusive, and direct. The words we use to frame and describe our analysis matter. While words "vulnerable communities" are commonplace in EJ analysis and discussions, we want to emphasize that – in many case- more so than vulnerable, these are communities that *stand to benefit* from a clean energy transition
- Check ourselves throughout the process: **Are we engaging the communities highlighted?** If no, that is not acceptable and KFTC needs to support the voices in this communities. **Are the solutions proposed making things worse?** If yes, that is not acceptable and plans need to change. **Are solutions directing resources to the areas which need it most?** If no, plans will change
- No matter the choice Kentucky makes of either implementing the Clean Power Plan or its own plan (KFTC's or another), KFTC can present this analysis to display the importance of recognizing environmental injustice and conducting an EJ analysis, as well as offer experience on methodologies to generate the analysis

1.3 General Project Process Outline

This documentation provides detailed information on data processing (see Section 3) and step-by-step methodologies on the processes used in the Empower KY EJ Analysis (see Appendices) to meet the project's goal of transparency and accessibility, as well as offer methodological options to the spatial analytic communities who wish to conduct EJ analysis. However, below is a brief summary of the process to answer, in part, the question, "how did we get to where we are today?"

- First, the Empower KY EJ work team developed a list of indicators to be included in the cumulative pollution and health overlay maps, drawing upon research and studies at both the national level (EPA-suggested and -utilized indicators) and state level
- The work team looked for data that was valid and well-supported, working within limitations of time and resources
- The data was analyzed in two groups: all of the indicators which speak towards cumulative pollution exposure are in one group, they will be analyzed together, and indicators of health problems are in their own group
- Vulnerability overlay maps were created for each Demographics, Pollution, and Health by layering the indicators for that group and running an analysis which combines them. What this does is creates a new layer - which is a cumulative value of all the indicators that went into it
- Then, for each map, the Demographic Overlay, which is comprised of demographic indicators used by the EPA in their environmental justice analysis tools, was overlaid with each of the Pollution and Health overlay maps

2) Analytic Methods

2.1 EPA-Recommended Analytic Method Chosen

In the EPA's documentation "Technical Guidance for Assessing Environmental Justice in Regulatory Analysis" (2016) several best-practice methods for conducting such analysis are presented and recommended. The "Technical Guidance" document's primary audience, according to the introduction, is EPA analysts. However, this documentation has been referred to frequently throughout the Empower KY EJ Analysis project to gain understanding and information about best-practices for EJ analysis. This documentation stresses the importance of choosing a method based on data availability, time, resources, audience, and purpose and needs to be feasible and appropriate (U.S. EPA 2016, *Tech Guidance*).

The various methods include Statistical Significance, Visual Displays, Proximity-Based Analysis, Use of Exposure Data, and Qualitative Approaches. The "Technical Guidance" provides a description and a list of both advantages and disadvantages of each method and is very beneficial to read (the citation of this documentation is in the works cited section of this document). The analysis chosen for the Empower KY EJ Analysis, an overlay analysis which combines data sets to produce a map of cumulative values, falls under the category of Visual Displays. This analytic method "can communicate baseline levels of air pollutants or clusters of hazardous waste sites and then overlay the demographic profile and baseline health status of various population groups of concern. In this way, analysts can identify potential hot spots where high levels of pollution are found in communities with minority populations, low-income populations, or indigenous peoples" (U.S. EPA 2016, *Tech Guidance*). Visual displays consist of maps, charts, graphs, and other visualizations.

This project's visual display consists of overlay maps generated from the ArcGIS Weighted Sum Overlay tool (see Section 2.2 and Appendices 1 and 2 for more information about this process). The overall goal of the project was to provide as set of preliminary visualizations of environmental justice issues in Kentucky to a broad audience (i.e. not just researchers and policy-makers but all Kentuckians), working within the constraints of time and resources.

2.2 Analytic Method Decisions

Some big decisions were made that are worth discussing as they informed the analytic methods in this project. These decisions were made keeping in mind the goal of this project, including audience, purpose, and resource constraints. They are Scope, Weighted Sum Overlay Analysis, and Data Collection.

Scope

First, the scope of the analysis is an important decision for any EJ analysis. There is an undeniable and significant value to site-specific analysis where data collection can be focused, narratives can be gathered and shared locally, and the ultimate produce is entirely unique to its place. However, this project's scope is a statewide scale. The rationale for choosing this scope is that the project needed to, importantly, match the scope of the Empower Kentucky plan (which is plan for all of Kentucky) and that the goal of the project is to provide a broad overview which provides a starting foundation with which to identify areas for further research.

Weighted Sum Overlay Analysis

A second important decision is the use of overlay analysis, specifically ArcGIS's Weighted Sum Overlay tool, to produce the EJ maps. This type of analysis has been utilized before in community vulnerability as well as site suitability analysis and maps. In this project, an overlay analysis expands upon what EJSCREEN, the EPA EJ Screening Tool, provides, as it is truly overlays the data, meaning it compresses the datasets together, producing a map of cumulative values. Users can then look at the map and see a visualization of concentration of all the indicators that went into the analysis, for example, a cumulative map of pollution using indicators such as Ozone, PM2.5, etc. The tool's capability to weight the data sets adds flexibility and creates a dynamic product, i.e. the analysis can be re-run with various weighting schemes.

More information about how Weighted Sum Overlay works can be found within ArcGIS online documentation: http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-weighted-sum-works.htm. As described in this documentation, "the Weighted Sum tool provides the ability to weight and combine multiple inputs to create an integrated analysis…multiple raster inputs, representing multiple factors, can be easily combined incorporating weights or relative importance" (ESRI 2016). It works by "multiplying the designated field values for each input raster by the specified weight. It then sums (adds) all input rasters together to create an output raster" (ESRI 2016). A diagram below visualizes this analysis:

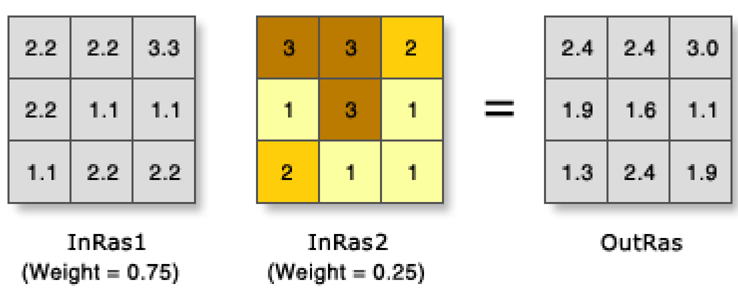


Figure 1. ESRI Weighted Sum Diagram *Image Source: http://pro.arcgis.com/en/pro-app/tool-reference/spatial-analyst/GUID-D7ABDBB3-B782-4BF9-89AC-4FACDE4FC6BD-web.gif*

Below is a diagram created for this project's documentation further explaining how this analysis is used to generate an environmental justice analysis, i.e. an analysis examining the most vulnerable areas in our state's current energy landscape.

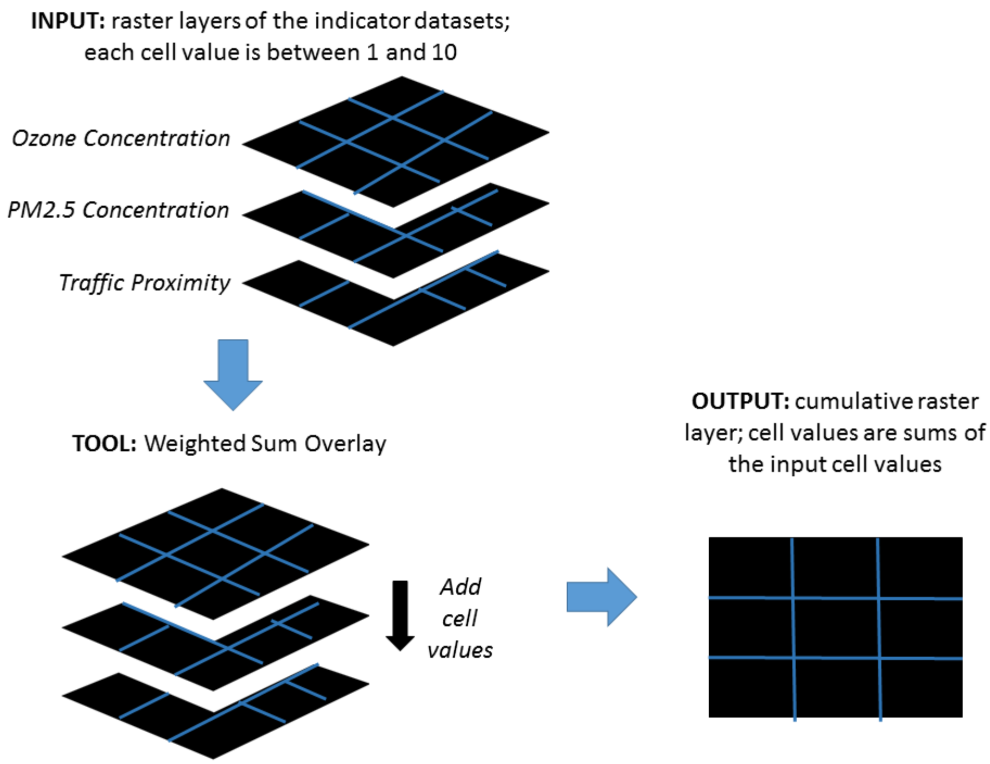


Figure 2. Project-Specific Weighted Sum Diagram *Image Source: Laura Greenfield (KFTC EJ work team)*

The benefit of this analysis is that it produces a cumulative visual of concentrations of values, i.e. the indicators that went into the analysis. Each dataset is classified individually so that every geography (block group, census tract, county polygon, etc.) has an index value of 1-10. Thus, when the dataset is rasterized before going into the Weighted Sum tool, the datasets can be compared using these index values. A limitation is that the analysis opens up uncertainties and bias depending on the extent to which the analyst chooses to weight the indicators. Additionally, while the analysis presents visuals of concentrations, without spatial statistical supplementary analysis, it does not speak to the extent to which the indicators are significantly correlated (see Section 7 for more information on statistical opportunity within this project).

Data Collection

A third important decision was the approach with which the work team for this project collected data. If a reader of this document refers to Section 3 “Data Details,” they will find that the data is from all different ranges of years. The data doesn’t cover all indicators of environmental justice in Kentucky and, therefore, does not tell the full story. However, the decision was made to approach data collection and the project processes as a whole as a preliminary analysis, remaining open always to feedback and improvements.

2.3 Strengths of Analysis

Below is a list of the strengths of the analytic methods outlined in this section and used to generate the preliminary work of the Empower KY EJ Analysis.

- Kentucky-wide scope allows the analysis to match the Empower Kentucky plan for a clean-energy transition
- Analysis provides opportunity to begin building a Kentucky-specific environmental justice database which KFTC, cities, organizations, and researchers can use to generate and refine EJ analysis
- The analysis, in ways, stands on the shoulders of giants by drawing upon EPA resources (data, methods) and Kentucky researchers and activists (for information on pollution and health issues specific to the state, data availability, etc.)
- The visual display method is accessible to a broad audience which creates space for open conversation, refinement and improvements, and a strong foundation for informing discussions on policy and further discussions
- The method's goal of visualizing cumulative exposure and occurrence is an important one. Pollution exposures may occur from multiple sources and, regarding the Health overlay, the energy landscape in Kentucky affects many health issues. By aiming to measure cumulative impacts, the analysis works to tell as much of the story as possible

2.4 Limitations of Analysis

Below is a list of the limitations of the analytic methods outlined in this section and used to generate the preliminary work of the Empower KY EJ Analysis. These limitations are fuel for exploring other opportunities for research and engagement in this work. See Section 7 "Next Steps" for specific ideas which work to fill the gaps in this preliminary project.

- Analysis does not include a statistical analysis component to look at individual dataset significance as well as correlation between datasets
- Qualitative data, such as narratives of communities, is not worked into the final visual displays
- Analysis has not yet explored community capacity or resiliency literature/data which could be weighted negatively in the overlay analysis as it represents aspects of an area which make it "less vulnerable" and/or more resilient to shifts in the energy landscape
- Many datasets, generated by this project's work team as well as downloaded from EJSCREEN (the EPA's environmental justice screening tool), use proximity calculations as a proxy (stand-in) for pollution exposure to a source
- An analysis such as this cannot contain all the relevant issues that should be considered (environmental issues, pollution sources, health problems). Rather, it is a product of time and resource constraints, as well as maintaining the state-wide scope. For example, there may be an issue and data that is specific to a certain community in Kentucky but, if that data is not available state-wide, it was not included
- Although none of the overlays in this preliminary analysis contain weighted indicators, this doesn't mean that there is truly "equal-weighting." As the EJSCREEN Technical Documentation

discusses, “there is no objective version of ‘equal weighting’… [it is] just an artifact of the units (scaling) and aggregation method” (U.S. EPA 2016, *Tech Doc*). This underlines the importance of examining datasets individually, engaging in case studies, and developing statistical reports to augment the visual display

3) Data Details in Empower Kentucky EJ Analysis

The goal of this section is provide detail on the datasets used to generate the Empower KY EJ Analysis overlay maps. There are three sub-sections (hyperlinked below) each containing a summary table of the indicators used for that overlay map and specifics on each dataset, including source, year, description, and GIS processing**.

Click on one of the headings below to jump to that section

3.1 Pollution Exposure Data Details

3.2 Health Effects Data Details

3.3 Demographic Data Details

3.1 Pollution Exposure Data Details

Table 1. Summary Table of Pollution Indicators and Data Sources

Indicator	Data Source	Data Year
Proximity to Coal Mines (surface and underground)	Kentucky Division of Mine Permits	2016
Proximity to Coal Ash Impoundments	Southeast Coal Ash	2013
Coal Haul Highway System Proximity	Kentucky Transportation Center	2015
Concentration of Particulate Matter (PM2.5)	EPA Office of Research and Development via EJSCREEN	2012
Concentration of Ozone	EPA EJSCRREEN	2012
Lead Exposure (percent of pre-1960 housing)	Census/ACS via EJSCREEN	2010-2014 estimates
Proximity to Risk Management Plan (RMP) sites	EPA RMP database via EJSCREEN	2012
Proximity to Major Direct Water Dischargers (NPDES)	EPA PCS/ICIS database via EJSCREEN	2012
Proximity to National Priorities List sites (NPL)	PEA CERCUS database via EJSCREEN	2013
Proximity to Treatment, Storage, or Disposal Facilities (TSDFs)	EPA RECRA database via EJSCREEN	2012
Traffic Proximity	U.S. Department of Transportation via EJSCREEN	2014

Proximity to Coal Mines (surface and underground)

Data Source and Year

Kentucky Division of Mine Permits (http://kygisserver.ky.gov/geoportal/catalog/main/home.page , search for "permitted mine boundaries"), 2016, geographic level: polygons representing individual mine boundaries.

Description

This original dataset contains approved permitted mine boundaries for surface and underground mines in Kentucky, including western and eastern coal fields and all active, inactive, and released permits. The dataset, for this project, was filtered to only contain Active* coal mines, then divided into two separate datasets for processing: surface coal mines and underground coal mines.

The rationale for dividing the mine data into two separate datasets is to recognize and highlight that each type of mine is a source for a different type of pollution. For surface mines, the pollution is mainly the environmental impact upon surface water, scarring of the landscape, and various degrees of soil erosion and deforestation, depending on the extent of the surface mine. Underground mines can have surface effects but, chiefly, affect the groundwater (in most, not all underground mines) and fissures in the land's surface in the case of a mine collapse. Additionally, the 2010 report "Assessing the True Cost of Coal" adds important polluting effects of mines, including destruction of local habitat and biodiversity, acid mine drainage, risk of incomplete reclamation following mine use, and methane emissions from coal leading to climate change (Epstein et. al 2010).

Proximity was calculated using a methodology which replicates the EPA's calculation of proximity for EJSCREEN datasets (U.S. EPA 2016, *Tech Doc*). Block groups received a proximity value (inverse of sum of all distances from block group centroid to coal mine centroids within 5km; for block group centroids with no mine within 5km, the score is the inverse distance from block group centroid to nearest coal mine centroid). See Appendix 3 for detailed write-up of this project's proximity calculations.

**Kentucky Division of Mine Permits applies a Mine Status Code to each mine. The following Mine Status Codes were included as "Active" mines for this project: A1, A2, D3, D6, FF, IA, O1, O2, P1, P2, SP. This project's definition for an "Active" mine is any mine with a permit that is not completely released and reclamation of the mine has not fully been completed.*

*Data Processing***

1) **Input Dataset:** Permitted Mine Boundaries (active surface coal mines), Permitted Mine Boundaries (active underground coal mines)
2) Proximity values generated at block group level using Calculating Proximity Layer methodology (Appendix 3)
3) Group data into similar values using Geometrical Interval classification method, to 10 classes
4) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
5) Rasterize polygon layer based on index value
6) **Output Layers:** Pollution_CoalMineProx_SF (surface coal mine) and Pollution_CoalMineProx_UG (underground coal mine) raster layers

Proximity to Coal Ash Impoundments

Data Source and Year

Southeast Coal Ash, via www.southeastcoalash.org, 2013, exported impoundment location data for Kentucky.

Description

This original dataset contains each coal ash impoundment in Kentucky, including location, utility company managing it, impoundment type, and EPA hazardous rating. For this project, location data for each impoundment is the only component incorporated into the overlay analyses.

According to Southeast Coal Ash, "coal ash is the toxic-laden waste left behind after coal is burned for energy" and "contains many heavy metals and other toxic elements, which are concentrated in the ash" (www.southeastcoalash.org). Including in these contaminants are arsenic, cadmium, chromium, lead, mercury, and selenium. Pollution effects from coal ash impoundments include leakage into ground water, risk dam collapse, and damage to air quality, drinking water, and animal life (www.southeastcoalash.org). Proximity to impoundments increase risk of these polluting effects. Southeast Coal Ash provides further information about impoundments and effects on nearby infrastructure and communities, specifically for Kentucky (see http://www.southeastcoalash.org/wp-content/uploads/2013/09/KYCoalAshFactSheet_V2.pdf). A report from 2010, "Assessing the True Costs of Coal," discusses fly ash ponds, which is a certain type of coal ash impoundment, stating that "up to 1 in 50 residents in Kentucky, including 1 in 100 children, living near one of the fly ash ponds is at risk of developing cancer as a result of waterborne and airborne exposure to the waste" and that a breach of the impoundment "would likely cause significant property damage, illness, and deaths," citing the Tennessee Valley Authority Kingston coal-fired plant fly ash spill in 2008 (Epstein et. al 2010).

Proximity was calculated using a methodology which replicates the EPA's calculation of proximity for EJSCREEN datasets (U.S. EPA 2016, *Tech Doc*). Block groups received a proximity value (inverse of sum of all distances from block group centroid to coal ash impoundment centroid within 5km; for block group centroids with no mine within 5km, the score is the inverse distance from block group centroid to nearest coal ash impoundment centroid). See Appendix 3 for detailed write-up of this project's proximity calculations.

*Data Processing***

1) **Input Dataset:** Coal Ash Impoundments in Kentucky
2) Proximity values generated at block group level using Calculating Proximity Layer methodology (Appendix 3)
3) Group data into similar values using Geometrical Interval classification method, to 10 classes
4) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
5) Rasterize polygon layer based on index value
6) **Output Layers:** Pollution_AshImpoundProx raster layer

Coal Haul Highway System Proximity

Data Source and Year

Kentucky Transportation Cabinet (KTC) and the Kentucky Division of Planning, http://datamart.business.transportation.ky.gov/, 2015

Description

This dataset contains the public highways on which coal was reported to be transported by truck in 2015. The KTC is required to report these roads, as well as include the coal haul ton-miles in each county using the reported information. The Department for Local Government uses this ton-miles info as one factor to determine the distribution of coal severance tax revenue to coal producing and coal impact counties. The dataset contains both the regular coal haul highway system as well as the Extended Weight system which consists of road segments over which coal or coal by-products in excess of 50,000 tons were transported by motor vehicles. This description was generated from information on KTC's coal haul webpage: http://transportation.ky.gov/planning/pages/coal-haul.aspx. The value of this dataset is that it highlights both coal producing and coal impact counties as well as highlights pollution emitting from the transportation stage of coal production in Kentucky.

The regular coal haul highway system and Extended Weight system were merged into a single dataset representing the coal haul highway system in Kentucky as a single feature. Proximity was calculated using a methodology which replicates the EPA's calculation of proximity for EJSCREEN datasets (U.S. EPA 2016, *Tech Doc*). Block groups received a proximity value (inverse of sum of all distances from block group centroid to coal haul road segment within 5km; for block group centroids with no road segment within 5km, the score is the inverse distance from block group centroid to nearest road segment. See Appendix 3 for detailed write-up of this project's proximity calculations. The regular highway system and Extended Weight system were merged into a single dataset representing the coal haul highway system in Kentucky as a single feature.

*Data Processing***

1) **Input Dataset:** Permitted Mine Boundaries (active surface coal mines), Permitted Mine Boundaries (active underground coal mines)
2) Proximity values generated at block group level using Calculating Proximity Layer methodology (Appendix 3)
3) Group data into similar values using Geometrical Interval classification method, to 10 classes
4) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
5) Rasterize polygon layer based on index value
6) **Output Layers:** Pollution_CoalMineProx_SF (surface coal mine) and Pollution_CoalMineProx_UG (underground coal mine) raster layers

Concentration of Particulate Matter (PM2.5)

Data Source and Year

Downloaded from EPA EJSCREEN which sources from a combination of monitoring data and air quality monitoring from EPA's Office of Research and Development, 2012, exported data for Kentucky.

Description

PM2.5 is particulate matter that is 2.5 microns or less in diameter, commonly released by power plants and industrial facilities (U.S. EPA 2016, *Tech Doc*). This dataset contains the annual average PM2.5 concentrations in micrograms per cubic meter. The EPA rationalizes inclusion of this dataset in EJSCREEN due to the documentation of the "health effects associated with exposure to PM2.5, including elevated risk of premature mortality from cardiovascular diseases or lung cancer, and increased health problems such as asthma attacks" (U.S. EPA 2016, *Tech Doc*).

More information about particulate matter can be found on the EPA's PM2.5 web page: http://www.epa.gov/pm.

*Data Processing***

1) **Input Dataset:** Particulate Matter (PM2.5) – Kentucky
2) Proximity values generated at block group level using Calculating Proximity Layer methodology (Appendix 3)
3) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
4) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
5) Rasterize polygon layer based on index value
6) **Output Layers:** Pollution_PM25 raster layer

Concentration of Ozone

Data Source and Year

Downloaded from EPA EJSCREEN which sources from a combination of monitoring data and (Congestion Mitigation and Air Quality Improvement) CMAQ air quality monitoring, 2012, exported data for Kentucky.

Description

This dataset contains the May-September (summer/ozone season) average of daily-maximum 8-hour-average ozone concentrations, in parts per billion (ppb) (U.S. EPA 2016, *Tech Doc*). The EPA rationalizes inclusion of this dataset in EJSCREEN by citing "toxicological and epidemiological studies [which] have established an association between exposure to ambient ozone and a variety of health outcomes, including reduction in lung function, increased inflammation, and increased hospital admissions and mortality" (U.S. EPA 2016, *Tech Doc*).

More information about ozone and its effects can be found on the EPA's ozone webpage: http://www.epa.gov/air/ozonepollution/.

*Data Processing***

1) **Input Dataset:** Ozone – Kentucky
2) Proximity values generated at block group level using Calculating Proximity Layer methodology (Appendix 3)
3) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
4) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
5) Rasterize polygon layer based on index value
6) **Output Layers:** Pollution_Ozone raster layer

Lead Exposure

Data Source and Year

Downloaded from EPA EJSCREEN which sourced from the American Community Survey (ACS) 2010-2014 estimates.

Description

This dataset contains the percentage of housing per block group which was built before 1960. This data is being used as a proxy for lead exposure because lead-based paint was banned in 1978 but still remains in housing built before that time. Therefore, "the percentage of occupied housing units before 1960 is selected as an indicator of the likelihood of having significant lead-based paint hazards in the home" (U.S. EPA 2016, *Tech Doc*).

The EPA rationalizes inclusion of this dataset in EJSCREEN due to the fact that "elevated blood lead levels are a well-documented public health concern of particular interest to [Environmental Justice] stakeholders, and represent an important environmental health issue" (U.S. EPA 2016, *TECH DOC*). Additionally, there are strong racial and socioeconomic aspects to lead exposure susceptibility as well as "recent research [which] demonstrated that children can experience neurological damage even at low levels of exposure to lead" (U.S. EPA 2016, *Tech Doc*).

More information about lead and its effects can be found on the EPA's ozone webpage: http://www.epa.gov/lead.

*Data Processing***

1) **Input Dataset:** Percentage of Housing Units Built Pre-1960 – Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_LeadExposure raster layer

Proximity to Risk Management Plan (RMP) sites

Data Source and Year

Downloaded from EPA EJSCREEN which sourced and calculated data from the EPA RMP database, 2012, exported data for Kentucky.

Description

The Clean Air Act (CAA) requires certain facilities (RMP facilities) to file risk management plans due to their risk of chemical accidents due to handling/processing/storage/etc. of substances with high toxicity and flammable/explosive potential (U.S. EPA 2016, *Tech Doc*).

This dataset contains the count of RMP facilities within 5km, divided by distance, presented as population-weighted averages of blocks in each block group. Adjustments are made if there are none within 5 km (closest is used). The EPA rationalizes inclusion of this dataset in EJSCREEN due to the multiple dimensions of both routine pollution exposure as well as heightened risk regarding proximity to RMP sites. With some industrial facilities, "there may be routine releases to the air and water…Thus, people may be exposed to some substances directly through inhalation or indirectly through water routes or via ingestion of food" (U.S. EPA 2016, *Tech Doc*). However, the largest concern is surrounding the accidental explosions, fires, and releases. As stated in the EPA EJSCREEN Technical Documentation, "the sudden release of relatively large quantities of acutely toxic substances can cause serious health effects including death…[and] these effects may be prompt or may occur or persist sometime after the exposure" (2016).

More information about RMPs can be found on the EPA's RMP webpage: http://www.epa.gov/rmp. The RMP database is stored in Envirofacts: http://www.epa.gov/enviro/facts/rcrainfo/search.html.

*Data Processing***

1) **Input Dataset:** Proximity to Risk Management Plan Facilities - Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_RMPProx raster layer

Proximity to Major Direct Water Dischargers (NPDES)

Data Source and Year

Downloaded from EPA EJSCREEN which sourced and calculated data from the EPA PCS/ICIS database, 2012, exported Kentucky data.

Description

"Major Direct Water Dischargers" are facilities, part of the National Pollutant Discharge Elimination System (NPDES) and regulated by the Clean Water Act, which discharge pollutants from point sources to waters specifically including "industrial direct dischargers (facilities that discharge pollutants directly into water bodies) and Publicly Owned Treatment Works (POTWs) (which receive and treat domestic and municipal waste and industrial wastewater and discharge treated water into water bodies)" (U.S. EPA 2016, *Tech Doc*).

This dataset contains the count of major direct discharger facilities within 5km, divided by distance, presented as population-weighted averages of blocks in each block group. Adjustments are made if there are none within 5km. The EPA rationalizes inclusion of this dataset because water pollutants can cause many adverse ecological and human health effects by both direct (people swimming downstream) and indirect (pollutant traces in drinking water even after it's been processed by the drinking water utility) (U.S. EPA 2016, *Tech Doc*).

More information about RMPs can be found on the EPA's NPDES webpage: http://www.epa.gov/npdes. Additionally, the EPA's water website is also a good resource: http://www.epa.gov/water.

*Data Processing***

1) **Input Dataset:** Proximity to NPDES Major Direct Water Dischargers - Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_NPDESProx raster layer

Proximity to National Priority List (NPL) sites

Data Source and Year

Downloaded from EPA EJSCREEN which sourced and calculated data from the EPA CERCLIS database, 2013, exported Kentucky data.

Description

National Priorities List (NPL) sites are a subset of Superfund sites, i.e. "uncontrolled abandoned hazardous waste sites" which the EPA began monitoring in 1980 with the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) (U.S. EPA 2016, *Tech Doc*). Superfund sites get placed on the NPL through one of the following ways: states/territories designate the site as a top-priority, the site has a score at or above 28.5 for the EPA's Hazard Ranking System (HRS), or a public health threat has been determined, either by the EPA, U.S. Public Health Service, or the Agency for Toxic Substance and Disease Registry (ATSDR) (U.S. EPA 2016, *Tech Doc*). Additionally, if "the EPA anticipates it will be more cost-effective to use its remedial authority (available only at NPL sites) than to use its emergency removal authority to respond to the site," the site becomes an NPL site (U.S. EPA 2016, *Tech Doc*).

This dataset contains the count of proposed and listed NPL sites, each represented by a point on the map (not a polygon of the site itself), within 5 km of the average resident in a block group, divided by distance, calculated as the population-weighted average of blocks in each block group. Adjustments are made if there are no NPL sites within 5km. The EPA rationalizes inclusion of this dataset for multiple reasons. First, research has been working to examine “the locations, listing decisions, and pace of cleanup at NPL sites in low-income and minority communities” (U.S. EPA 2016, *Tech Doc*). Also, NPL site contaminants have multiple pathways by which to reach humans, including surface contaminants that become airborne in dry seasons and climates and can be inhaled, contaminant other surfaces, and migrate onto agricultural land or into groundwater (U.S. EPA 2016, *Tech Doc*).

More information about NPL and Superfund sites can be found on the EPA’s Superfund webpage: http://www.epa.gov/superfund.

*Data Processing***

1) **Input Dataset:** Proximity to NPL/Superfund Sites - Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, “Processing Data for Weighted Sum Overlay,” for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_NPLProx raster layer

Proximity to Treatment, Storage, or Disposal Facilities (TSDFs)

Data Source and Year

Downloaded from EPA EJSCREEN which sourced and calculated data from the EPA RCRA (Resource Conservation and Recovery Act) database, 2012, exported Kentucky data

Description

Treatment, Storage, or Disposal Facilities (TSDFs) monitoring was established “to manage hazardous wastes from ‘cradle to grave,’ or from generation to disposal, to ensure that hazardous waste is managed in a manner that protects human wealth and the environment” (U.S. EPA 2016, *Tech Doc*).

This dataset contains the count of all active, commercial TSDF facilities within 5km, divided by distance, presented as population-weighted averages of blocks in each block group. Adjustments are made if there are none within 5km. The EPA rationalizes inclusion of this dataset by citing the multiple pathways substances at TSDFs can reach humans and the environment, including both through the atmosphere and migration to groundwater and agricultural land (U.S. EPA 2016, *Tech Doc*). Additionally, the EPA EJSCREEN Technical Documentation cites numerous studies which have examined disparities in proximity to TSDFs (2016).

More information about hazardous waste monitoring can be found on the EPA’s hazardous waste webpage: http://www.epa.gov/epawaste/hazard and the TSD webpage: http://www.epa.gov/epawaste/hazard/tsd. The RCRA database can be found on its webpage: http://www.epa.gov/epawaste/inforesources/data.

*Data Processing***

1) **Input Dataset:** Proximity to TSDFs - Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes

3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_TSDFProx raster layer

Traffic Proximity

Data Source and Year

Downloaded from EPA EJSCREEN which sourced and calculated data from U.S. Department of Transportation (DOT) traffic data, 2014, exported Kentucky data

Description

Proximity to motor vehicle traffic "is associated with increased exposures to ambient noise, toxic gases, and particulate matter (PM2.5) including diesel particulates" (U.S. EPA 2016, *Tech Doc*). While there are benefits to living close to major roads (access to jobs, healthcare, etc.), this dataset focuses on the adverse health effects of proximity to traffic, "particularly asthma exacerbation…as well as mortality rates…[and it has] been associated with subclinical atherosclerosis (a key pathology underlying cardiovascular disease (CVD)), prevalence of CBD and coronary heart disease (CHD), incidence of myocardial infarction, and CVD mortality" (U.S. EPA 2016, *Tech Doc*).

This dataset contains the count of vehicles per day within 500m of a block centroid, divided by distance in meters, presented as population-weighted average of blocks in each block group.

More information about conditions and effects regarding traffic proximity can be found on the EPA's near-roadway website: http://epa.gov/otaq/nearroadway.htm.

*Data Processing**

1) **Input Dataset:** Traffic Proximity - Kentucky
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Pollution_TrafficProximity raster layer

3.2 Health Effects Data Details

Table 2. Summary Table of Health Indicators and Data Sources

Indicator	Data Source	Data Year
Prevalence of Asthma in Adults (age 18 and older)	Behavioral Risk Factor Surveillance System (BRFSS) via Kentucky Health Facts	2012-2014
Hospitalization Rates of Asthma in Children (age 17 and under)	Kentucky Cabinet for Health and Family Sources and Kentucky State Data Center, via Kentucky's Kids Count database	2009-2011
Hypertension in Adults	Behavioral Risk Factor Surveillance System (BRFSS) via Kentucky Health Facts	2011-2013
Heart Disease Deaths	Kentucky State Data Center and Kentucky Department for Public Health Office of Vital Statistics via Kentucky Health Facts	2010-2014
Lung and Bronchus Cancer	Kentucky Cancer Registry	2009-2013
Premature Deaths	Kentucky State Data Center and Kentucky Department for Public Health Office of Vital Statistics via Kentucky Health Facts	2010-2013

Prevalence of Asthma in Adults (age 18 and older)

Data Source and Year

Downloaded from Kentucky Health Facts (http://www.kentuckyhealthfacts.org/data/topic/show.aspx?ind=48), sourcing from the Behavioral Risk Factor Surveillance System (BRFSS), 2012-2014

Description

This dataset contains the percent of adults (age 18 and older) with asthma at the county level. This dataset is important to include, especially when considering the coal landscape in Kentucky. Pollution from coal-fired plants contains particulates which are a cause of many respiratory ailments, including asthma (Epstein et. al 2010).

*Data Processing***

1) **Input Dataset:** Prevalence of Asthma (percent adults)
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Health_AsthmaPrevAdult raster layer

Hospitalization Rates of Asthma in Children (age 17 and under)

Data Source and Year

Downloaded from Kentucky's Kids Count database (http://datacenter.kidscount.org/data/tables/7216-rate-of-asthma-hospitalizations?loc=19&loct=5#detailed/5/2924-3043/false/995,932,757,470,116/any/14259), sourcing from the Kentucky Cabinet for Health and Family Sources and the Kentucky State Data Center, 2009-2011

Description

This dataset contains the rate of inpatient hospitalizations of children (ages 0-17) due to an asthma attack per 10,000 children. Rates are calculated using a 3-year average. Rates are not calculated for counties with fewer than 6 incidents. This data is at the county level. This dataset is important to include when regarding Kentucky's coal landscape because particulate emissions and ozone created from NOx and VOCs are a cause of respiratory ailments (including emphysema, asthma, and bronchitis) (Epstein et. al 2010). Additionally, research shows that "infants living in areas with high levels of particulate emissions face a 40% increase risk of death from respiratory complications" (Epstein et. al 2010). The effects of asthma in children is significant to note in Kentucky as it is the third-leading cause of hospitalization for children in Kentucky (Kentucky Department for Public Health 2013). The 2013 Asthma Surveillance Document for Kentucky states that asthma is "also one of the leading causes of school absenteeism in Kentucky [as] children with asthma miss an average of four school days each year, and some children in Kentucky miss many more days, resulting in an estimated annual loss of $10 million to school districts" (Kentucky Department for Public Health 2013).

*Data Processing***

1) **Input Dataset:** Rate of Asthma Hospitalizations (children)
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Health_AsthmaHospChild raster layer

Hypertension in Adults

Data Source and Year

Downloaded from Kentucky Health Facts website (http://www.kentuckyhealthfacts.org/data/topic/map.aspx?ind=68), sourcing from the Behavioral Risk Factor Surveillance System (BRFSS), 2011-2013

Description

This dataset contains the percent of adults who report they have been told they have high blood pressure or hypertension by a health care professional. Data for counties with fewer than 50 respondents have been suppressed. Data is at the county level.

This dataset is important to include, especially when regarding the coal landscape in Kentucky. As the 2012 report "Assessing the True Cost of Coal" and the corresponding True Cost of Coal Catalogue (version from 8/5/2011) state, "the odds for hypertension hospitalization increases 1% for each 1873 tons of coal burned" (Epstein et. al 2010). For perspective, the Energy Information Administration's State Profile and Energy Estimates for Kentucky published preliminary 2015 figures for coal production at 61.4 million tons. Using the 2011 Catalogue information, this would mean, in 2015 in Kentucky, the odds for hypertension hospitalization increased 32781%.

*Data Processing***

1) **Input Dataset:** Percent of Hypertension
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Health_Hypertension raster layer

Heart Disease Deaths

Data Source and Year

Downloaded from Kentucky Health Facts website (http://www.kentuckyhealthfacts.org/data/topic/show.aspx?ind=56 sourcing from Kentucky State Data Center analysis of records from the Kentucky Department for Public Health Office of Vital Statistics, 2010-2014

Description

This dataset contains the age-adjusted rate of deaths due to heart disease per year per 100,000 persons. The data is at the county level. This dataset is important to include, especially when considering the coal landscape in Kentucky. Pollution from coal-fired plants contains particulates which are a cause of heart disease (Epstein et. al 2010).

*Data Processing***

1) **Input Dataset:** Heart Disease Deaths (per 100,000 population)
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Health_HeartDisease raster layer

Lung and Bronchus Cancer

Data Source and Year

Kentucky Cancer Registry (http://www.cancer-rates.info/ky/), 2009-2013

Description

This dataset contains the age-adjusted rate of invasive cancer (lung and bronchus) incidence per 100,000 persons. Rates are age-adjusted to the 2000 U.S. Standard Population. For more information about age-adjustments within and across datasets, read here: http://health.mo.gov/data/mica/CDP_MICA/AARate.html . This dataset is important to include, especially when considering the coal landscape in Kentucky. Pollution from coal-fired plants contains particulates which are a cause of lung and bronchus cancer (Epstein et. al 2010).

*Data Processing***

1) **Input Dataset:** Age-Adjusted Invasive Cancer Incidence Rates by County in Kentucky
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Health_Cancer raster layer

Premature Deaths

Data Source and Year

Downloaded from Kentucky Health Facts website, sourcing from Kentucky State Data Center analysis of records from the Kentucky Department for Public Health Office of Vital Statistics

Description

Premature Death is the years of potential life lost before age 75 (YPLL-75). Every death occurring before the age of 75 contributes to the total number of years of potential life lost. For example, a person dying at age 25 contributes 50 years of life lost, whereas a person who dies at age 65 contributes 10 years of life to a county's YPLL. This dataset contains the YPLL measure presented as a rate per 100,000 populations. The data is at the county level.

This dataset is important to include, especially when considering the coal landscape in Kentucky. According to the 2010 report "Assessing the True Cost of Coal" and the corresponding True Cost of Coal Catalogue (version from 8/5/2011), "risk of death for those living within 30 miles of a coal-fired power plant is 3-4 times greater than for those living further away," and "for every million tons of coal mined, the death rate for those in proximity to the plant increases by 21 out of 100,000 people" (Epstein et. al 2010).

*Data Processing***

6) **Input Dataset:** Premature Death (years lost per 100,000 population)
7) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
8) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
9) Rasterize polygon layer based on index value
10) **Output Layers:** Health_PrematureDeath raster layer

3.3 Demographic Data Details

Table 3. Summary Table of Demographic Indicators and Data Sources

Indicator	Data Source	Data Year
Percent Under 5 Years of Age	ACS 2010-2014 estimates via EJSCREEN	2010-2014
Percent Over 64 Years of Age	ACS 2010-2014 estimates via EJSCREEN	2010-2014
Percent Minority	ACS 2010-2014 estimates via EJSCREEN	2010-2014
Percent Low Income	ACS 2010-2014 estimates via EJSCREEN	2010-2014
Percent Less Than High School Education	ACS 2010-2014 estimates via EJSCREEN	2010-2014
Percent Linguistic Isolation	ACS 2010-2014 estimates via EJSCREEN	2010-2014

Percent Under 5 Years of Age

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of population per block group that are 5 years of age and younger. See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen "to represent the 'social vulnerability' characteristics of a disadvantaged population" (2016).

*Data Processing***

1) **Input Dataset:** Percent Under Age 5
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_Under5 raster layer

Percent Over 64 Years of Age

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of population per block group that are 64 years of age and older. See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen "to represent the 'social vulnerability' characteristics of a disadvantaged population" (2016).

*Data Processing***

1) **Input Dataset:** Percent Over Age 64
2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_Over64 raster layer

Percent Minority

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of individuals in a block group who list their racial status as a race other than white along and/or list their ethnicity as Hispanic or Latino (i.e. all people other than non-Hispanic white-alone individuals; a non-Hispanic individual who is half-white and half American Indian would be county as a minority by this definition) (U.S. EPA 2016, *Tech Doc*). See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen "to represent the 'social vulnerability' characteristics of a disadvantaged population" (2016).

*Data Processing***

1) **Input Dataset:** Percent Minority
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, "Processing Data for Weighted Sum Overlay," for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_Minority raster layer

Percent Low Income

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of a block group's population where the household income is less than or equal to twice the federal "poverty level." See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen "to represent the 'social vulnerability' characteristics of a disadvantaged population" (2016).

*Data Processing***

1) **Input Dataset:** Percent Low Income

2) Group data into similar values using Natural Breaks (Jenks) classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, “Processing Data for Weighted Sum Overlay,” for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_LowIncome raster layer

Percent Less Than High School Education

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of people age 25 or older in a block group whose educational attainment is less than high school. See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen “to represent the ‘social vulnerability’ characteristics of a disadvantaged population” (2016).

*Data Processing***

1) **Input Dataset:** Percent Less Than High School Education
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, “Processing Data for Weighted Sum Overlay,” for more information about this script)
4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_LessHS raster layer

Percent Linguistic Isolation

Data Source and Year

Downloaded from EJSCREEN, sourcing from American Community Survey (ACS) data, estimates 2010-2014

Description

This dataset contains the percent of people in a block group living in linguistically isolated household (in which all members age 14 years and over speak a non-English language and also speak English less than “very well”) (U.S. EPA 2016, *Tech Doc*). See EPA EJSCREEN Technical Documentation for more information on the Demographic indicators which were chosen “to represent the ‘social vulnerability’ characteristics of a disadvantaged population” (2016).

*Data Processing***

1) **Input Dataset:** Percent Linguistic Isolation
2) Group data into similar values using Geometrical Interval classification method, to 10 classes
3) Assign index value (1-10) to each block group using Index Value Script (see Appendix 1, “Processing Data for Weighted Sum Overlay,” for more information about this script)

4) Rasterize polygon layer based on index value
5) **Output Layers:** Demographics_LingIso raster layer

4) Data Discussion

This section outlines observations of the availability and quality of data used to shape the Empower KY EJ Analysis Project. Availability and quality largely inform the data sets and indicators discussed in Section 3. Data that was unable to be obtained for this project and isn't included is also listed.

Data Availability and Quality Observations

- EJSCREEN, the EPA environmental justice screening tool, is great data resource. All indicators included in the tool are available for download, providing easy-to-access and quality processed data. All Demographic indicators used in this project and 8 of the 11 Pollution Exposure indicators were downloaded from EJSCREEN. The screening tool as well as accompanying documentation, which describes clearly the methodology and studies supporting the data, can be found on the EJSCREEN webpage: https://www.epa.gov/ejscreen. Data can be downloaded from the EJSCREEN interface for any state in the U.S.
- Data is not always available from a consistent time frame. The data sets used in the Health overlay is a good example of this. As seen in the screenshot below, taken from Section 3 of this documentation, each data set measures a different time frame:

Indicator	Data Source	Data Year
Prevalence of Asthma in Adults (age 18 and older)	Behavioral Risk Factor Surveillance System (BRFSS) via Kentucky Health Facts	2012-2014
Hospitalization Rates of Asthma in Children (age 17 and under)	Kentucky Cabinet for Health and Family Sources and Kentucky State Data Center, via Kentucky's Kids Count database	2009-2011
Hypertension in Adults	Behavioral Risk Factor Surveillance System (BRFSS) via Kentucky Health Facts	2011-2013
Heart Disease Deaths	Kentucky State Data Center and Kentucky Department for Public Health Office of Vital Statistics via Kentucky Health Facts	2010-2014
Lung and Bronchus Cancer	Kentucky Cancer Registry	2009-2013
Premature Deaths	Kentucky State Data Center and Kentucky Department for Public Health Office of Vital Statistics via Kentucky Health Facts	2010-2013

  These data sets were already packaged by the respective source offices/departments/organizations and available to the public for free download and use. A workaround for this in a future project to increase quality is to contact the sources and request a common time frame, critically thinking about the balance between current and accurate data.
- Datasets available do not always match measurement method. It is intuitive that data will not always be measured the same way. For example, PM2.5 and Ozone are commonly measured in micrograms per cubic meter or parts per billion, respectively. Likewise, demographic data is often presented as percentages of a population per geography. It is important to recognize these differences, especially when they occur within one overlay, such as the Health overlay in this project. Within these data sets, there is a mixture of percentages, rates and data sets that

are and aren't age-adjusted. While these differences are managed in part by the overlay processing (see Appendix 1), ideally the data sets can reflect the same type of measurement.

- EPA Technical Documentation for EJSCREEN shares the various processes the EPA analysts used to generate the EJSCREEN proximity datasets, aggregate data from the block level to the block group level, and more. The processes are transparent and research-backed which increases their quality. For example, the proximity thresholds (5km for a site/facility and 500m for roads) and the inverse distance-weighting methodology derives from air dispersion models, which the EPA sites in their documentation. However, these processes, while used for the EPA-sourced data, were not replicated fully for the other datasets. A replication of the EPA proximity processing was developed and can be found in Appendix 3.
- Minimizing overlap between the various indicators chosen to be included in the overlays was a goal from the beginning of the project. There can be multiple ways to represent a pollution source, for example, so it is important to include the "best" data to use as the indicator. The "best" data is dependent on availability, accuracy, and measurement. An example of "best" measurement is that, when wanting an indicator for pollution exposure, exposure data is a better measurement than proximity to the source, which would be a stand-in for exposure data.
- For Kentucky data, such as mine locations and oil and gas wells (not included in this analysis because it wasn't able to be obtained), there are multiple interfaces and webpages which map the data, provide it for download, and link to national database webpages. These datasets do not always match. When searching for data which existed on these multiple platforms, the work team reached out to personnel familiar with these departments and datasets to inform the decision of which to download and work with.

Data Not Included

Below is a list of the factors which are not included in the current EJ analysis due to limited data availability, time constraints, processing issues, and, occasionally, uncertainty as to whether or not the data exists and in what form it is in:

- Statewide water quality/pollution data, including both drinking water and surface water quality
- More comprehensive asthma (what we have, what we don't have)
- Community resiliency
- Abandoned mines
- Oil and gas wells
- Updated household energy costs as percent of income data/map. See Section 7 "Next Steps" for a look at the current map which exists for the state
- Employment within the energy industries, including mining, transportation, utilities, and calculating the percentage loss/gain over the years
- Renewable potential as replacement
- NATA Diesel PM (provided by EJSCREEN) but missing data
- Air pollutants other than PM2.5 and Ozone
- Proximity to other point sources which are sources of pollution

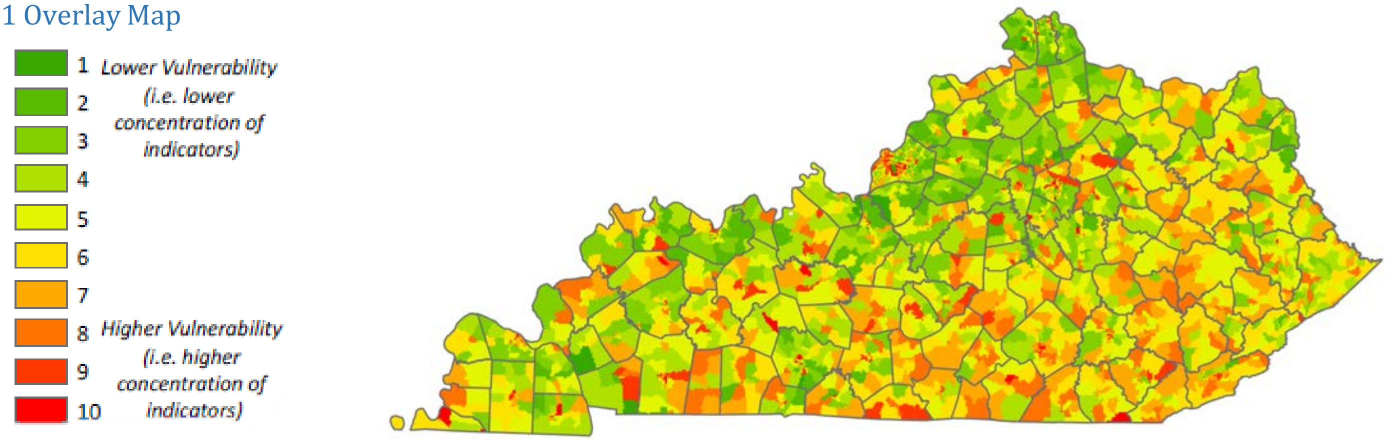
- Pollution exposure and increased air pollution release from power plant/facility startup, shutdown, malfunctions, etc.

5) Empower Kentucky Environmental Justice Visual Map Products

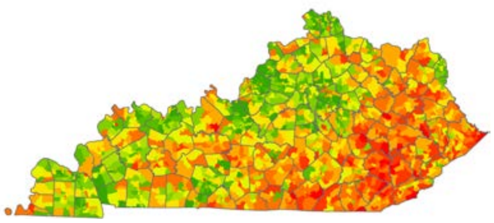
This section presents the visual map products from the Empower KY EJ Analysis. The overlay maps as well as individual indicator maps are shared for each category: Demographics, Cumulative Pollution Exposure, and Exposure-Related Health Problems. These maps can assist KFTC and all Kentuckians in examining environmental justice issues within our state, working to answer the question, “Where the most vulnerable areas in our current energy landscape?” or, , rather, “Who stands to benefit the most from a clean energy transition?” Section 5.4 presents Kentucky's Energy Infrastructure map series which visualize the overlay maps and the locations of Kentucky's coal-fired power plants.

5.1 Demographics Overlay

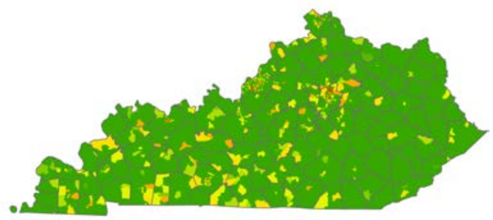
5.1.1 Overlay Map



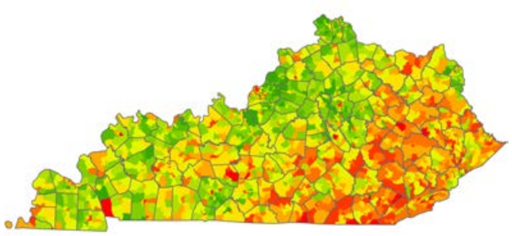
5.1.2 Individual Maps



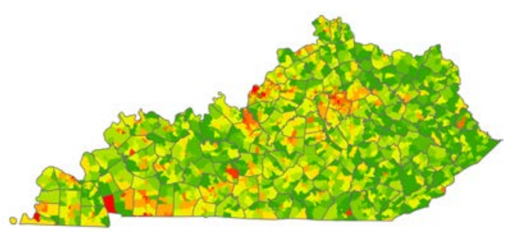
Percent Less Than High School Degree



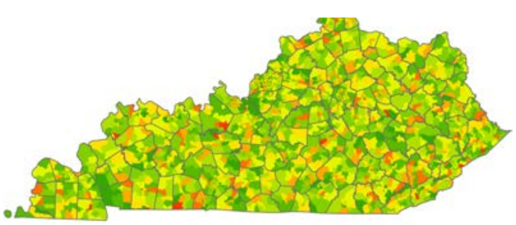
Percent Linguistic Isolation



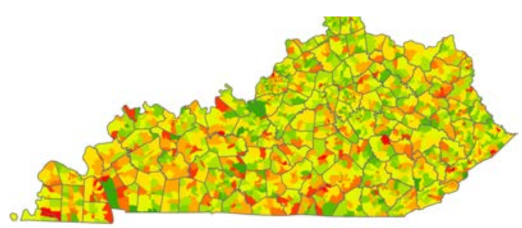
Percent Low Income



Percent Minority



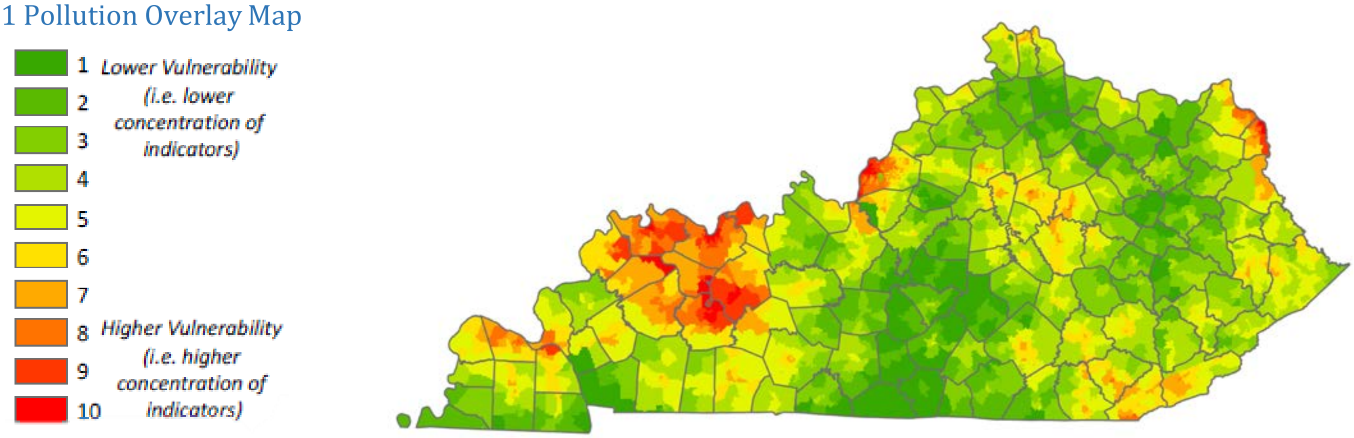
Percent Under Age 5



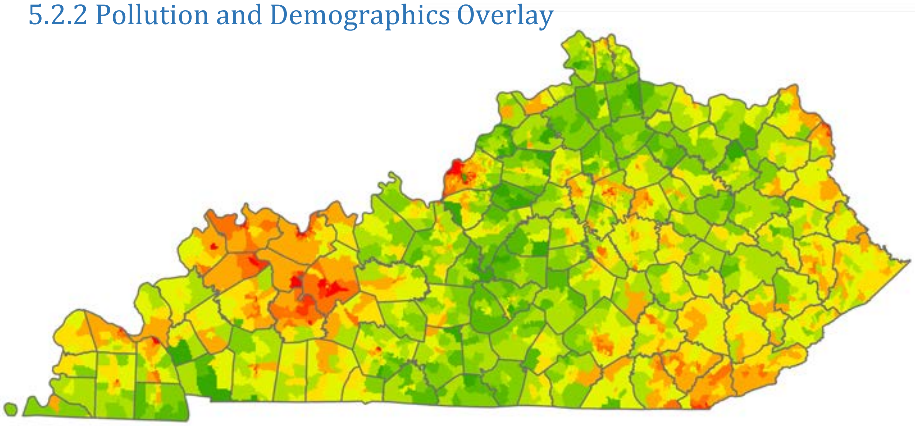
Percent Over Age 64

5.2 Cumulative Pollution Exposure Overlay

5.2.1 Pollution Overlay Map

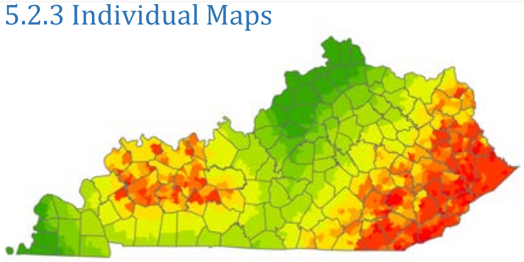


5.2.2 Pollution and Demographics Overlay

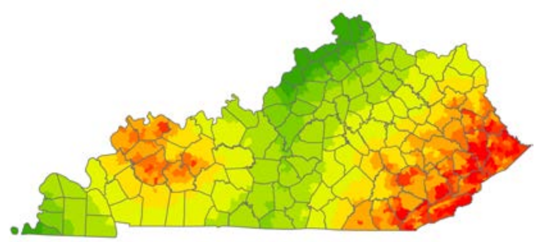


Note: The overlay maps are re-classified to 10 classes by Natural Breaks (Jenks)

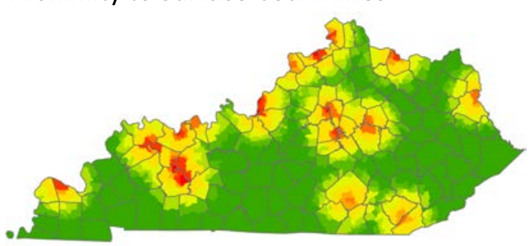
5.2.3 Individual Maps



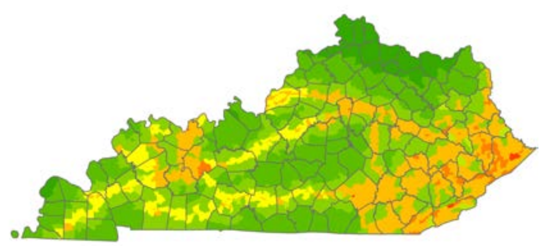
Proximity to Surface Coal Mines



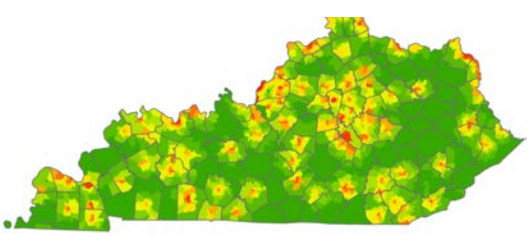
Proximity to Underground Coal Mine



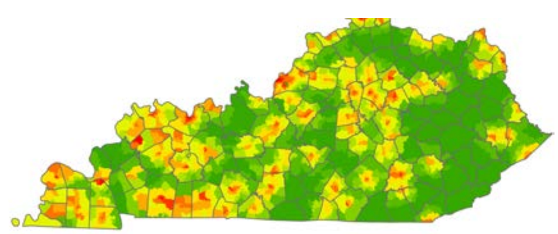
Proximity to Coal Ash Impoundments



Proximity to Coal Haul Highway System

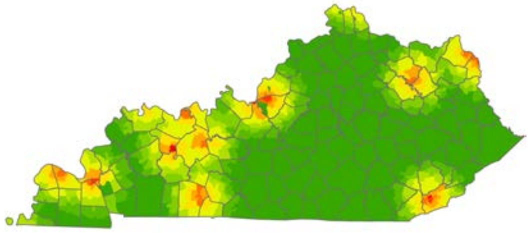


Proximity to Major Direct Water Dischargers (NPDES)

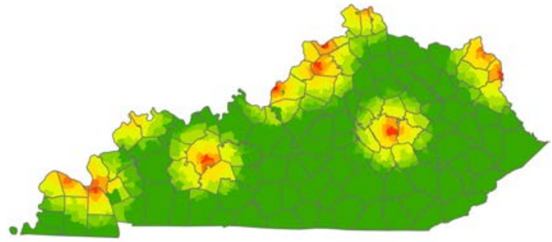


Proximity to Risk Management Plan (RMP) Facilities

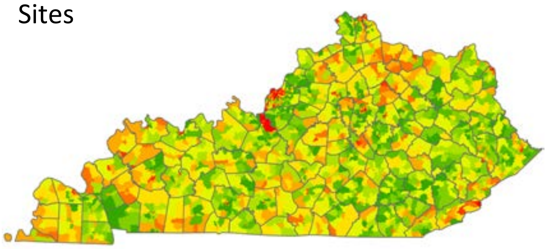
5.2.3 Individual Maps Continued



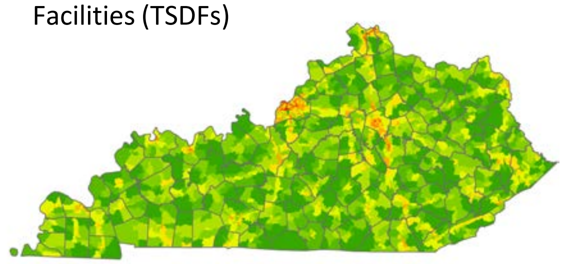
Proximity to National Priorities List (NPL) Sites



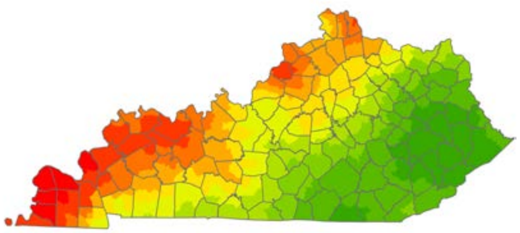
Proximity to Transfer, Storage, Disposal Facilities (TSDFs)



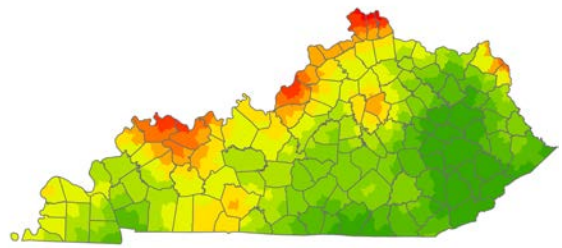
Lead Exposure



Traffic Proximity



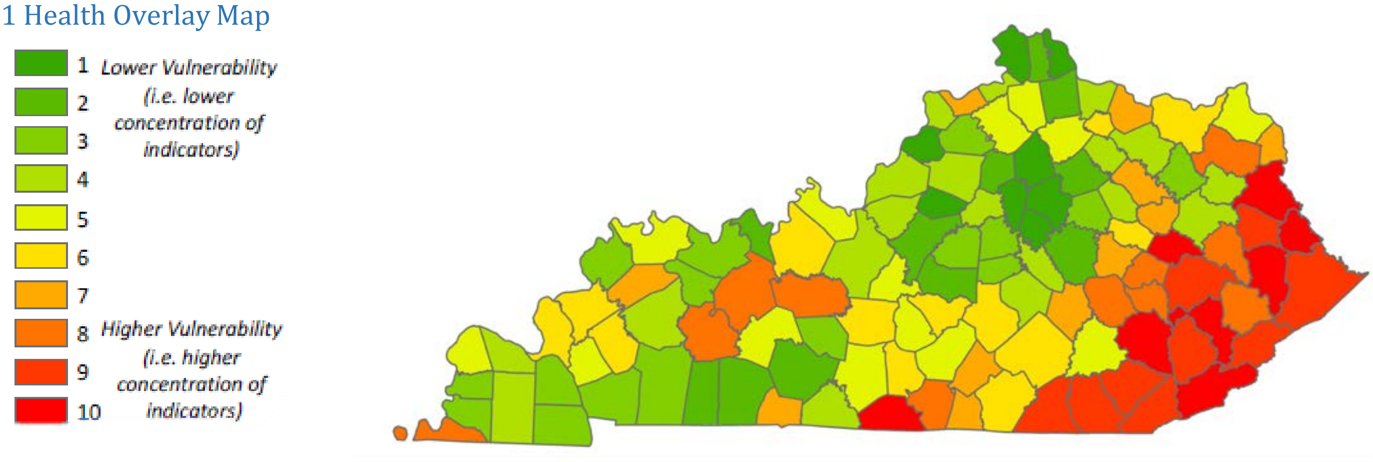
Concentration of Ozone



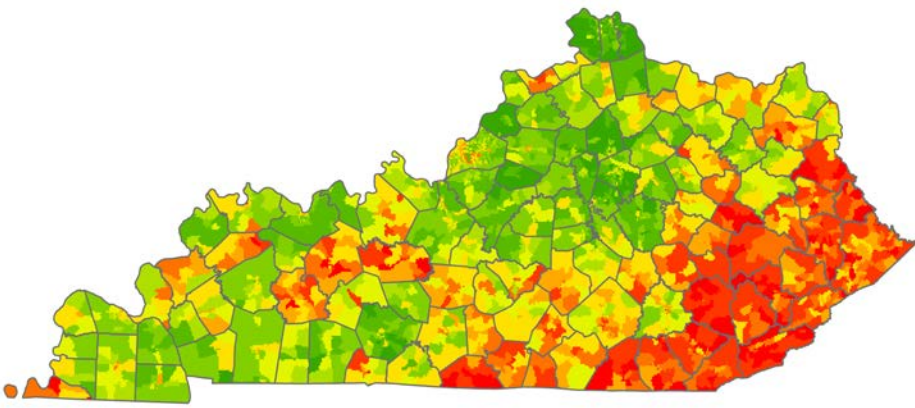
Concentration of PM2.5

5.3 Exposure-Related Health Problems Overlay

5.3.1 Health Overlay Map

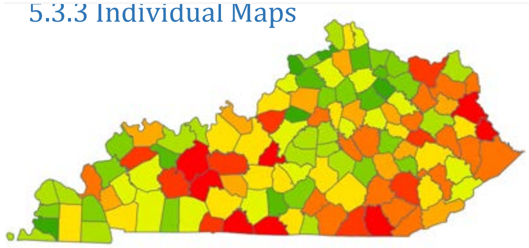


5.3.2 Health and Demographics Overlay Map

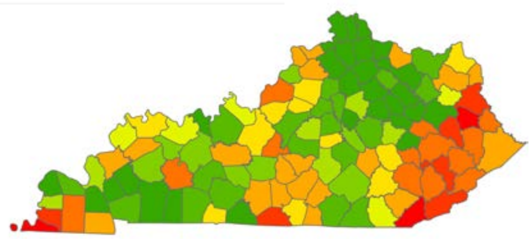


Note: The overlay maps are re-classified to 10 classes by Natural Breaks (Jenks)

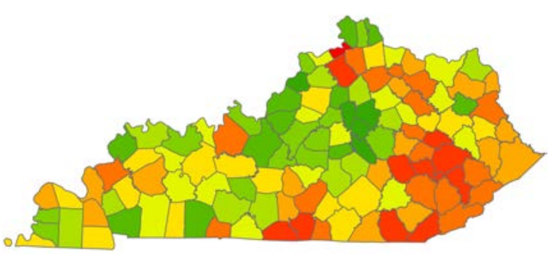
5.3.3 Individual Maps



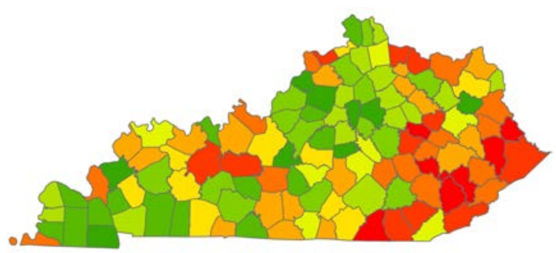
Prevalence of Asthma in Adults (age 18 and older)



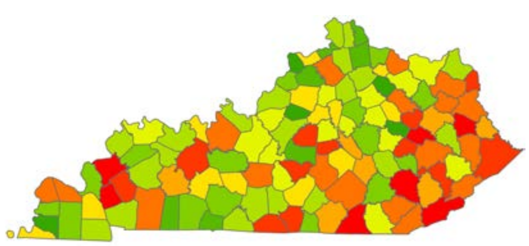
Hospitalization for Asthma of Children (age 17 and under)



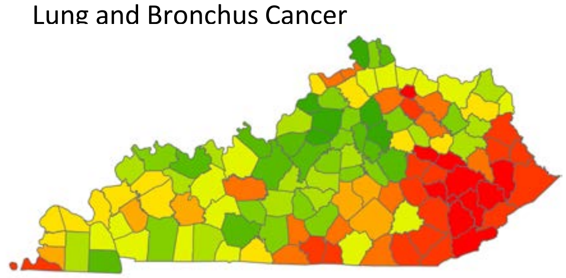
Heart Disease Death Rate



Lung and Bronchus Cancer



Adult Hypertension



Premature Death Rate

5.4 Kentucky Energy Infrastructure Maps

These maps are generated for use by KFTC's New Energy and Transition (NET) Committee. The maps display the Demographic Overlay and the Pollution/Demographic Overlay maps with the locations of Kentucky's fleet of coal-fired power plants and some data about these plants. Thus, these maps show the places in Kentucky where a set of demographic factors overlap with a set of pollution factors and a user can visualize relationships between these locations and the power plant locations.

5.4.1 Demographic Overlay Map with Coal Power Plants

(see page 35)

5.4.1 Pollution Demographic Overlay Map with Coal Power Plants

(see page 36)

Legend

▲ Coal Power Plants

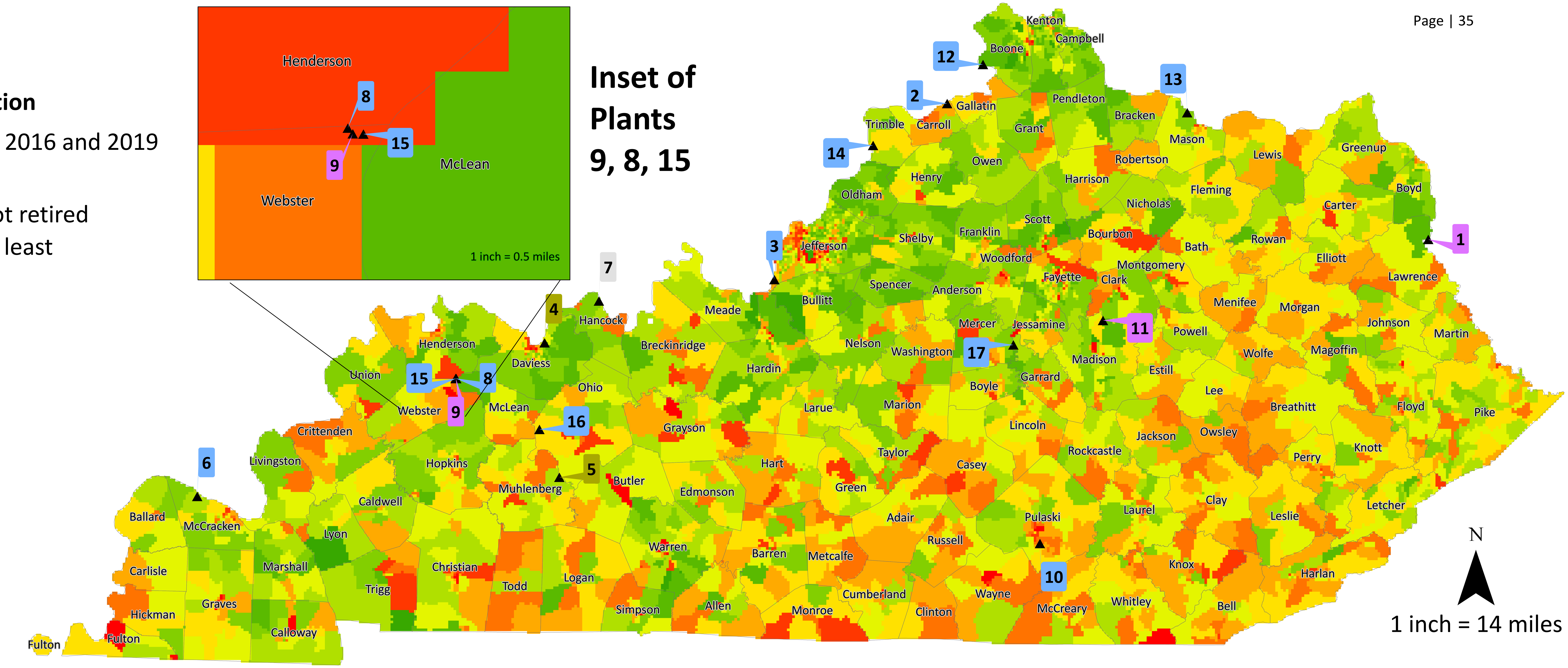
Plant Status Symbolization

- Retiring between 2016 and 2019
- Operating
- Not operating, not retired
- Multiple units, at least 1 unit will retire in near future

Lower Vulnerability

1
2
3
4
5
6
7
8
9
10

Higher Vulnerability

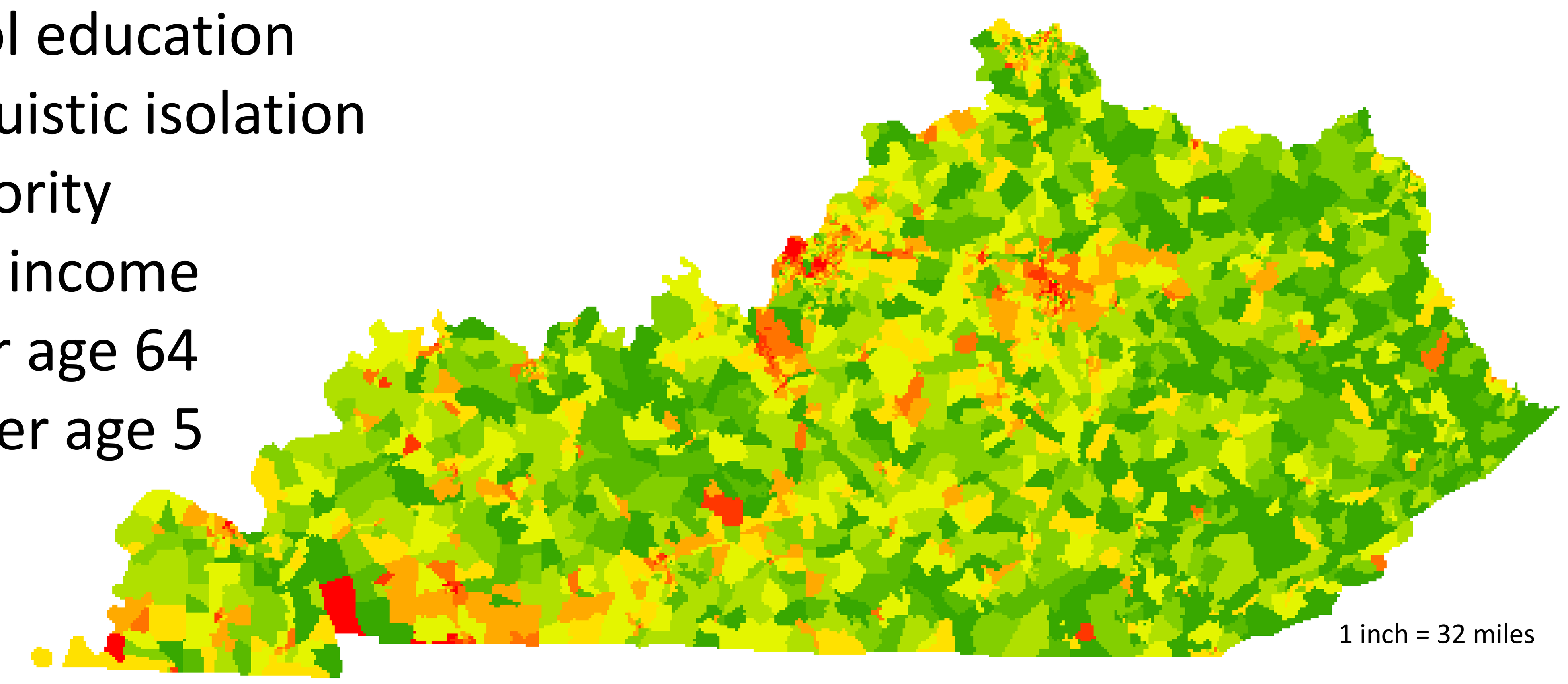


	Plant Name	Utility Name	Total MW	Source(s)
1	Big Sandy	Kentucky Power Co	260	Coal (260 MW)
2	Ghent	Kentucky Utilities Co	1919	Coal (1919 MW)
3	Mill Creek (KY)	Louisville Gas & Electric Co	1472	Coal (1472 MW)
4	Elmer Smith	City of Owensboro - (KY)	398.7	Coal (398.7 MW)
5	Paradise	Tennessee Valley Authority	2201	Coal (2201 MW)
6	Shawnee	Tennessee Valley Authority	1206	Coal (1206 MW)
7	Kenneth C Coleman	Big Rivers Electric Corp	443	Coal (443 MW)
8	HMP&L Station Two Henderson	Big Rivers Electric Corp	312	Coal (312 MW)
9	Robert A Reid	Big Rivers Electric Corp	123	Coal (65 MW), Natural Gas (58 MW)
10	Cooper	East Kentucky Power Coop, Inc	341	Coal (341 MW)
11	Dale	East Kentucky Power Coop, Inc	149	Coal (149 MW)
12	East Bend	Duke Energy Kentucky Inc	600	Coal (600 MW)
13	H L Spurlock	East Kentucky Power Coop, Inc	1346	Coal (1346 MW)
14	Trimble County	Louisville Gas & Electric Co	2185	Coal (1243 MW), Natural Gas (942 MW)
15	R D Green	Big Rivers Electric Corp	454	Coal (454 MW)
16	D B Wilson	Big Rivers Electric Corp	417	Coal (417 MW)
17	E W Brown	Kentucky Utilities Co	1588	Coal (682 MW), Natural Gas (906 MW)

Demographic Overlay Information

Indicators Included:
- Less than high school education
- Linguistic isolation
- Minority
- Low income
- Over age 64
- Under age 5

Map: Percent Minority



NOTE: all plants are conventional steam coal and natural gas fired combustine turbine (when natural gas is a MW source)

Data Sources: U.S. Energy Information Administration (EIA) dated March 2016 (power plant names, MW, etc.); KFTC (operational statuses)

Demographic Overlay Map with Coal Power Plants

Map by: Laura Greenfield
REV: 20161129

Legend

▲ Coal Power Plants

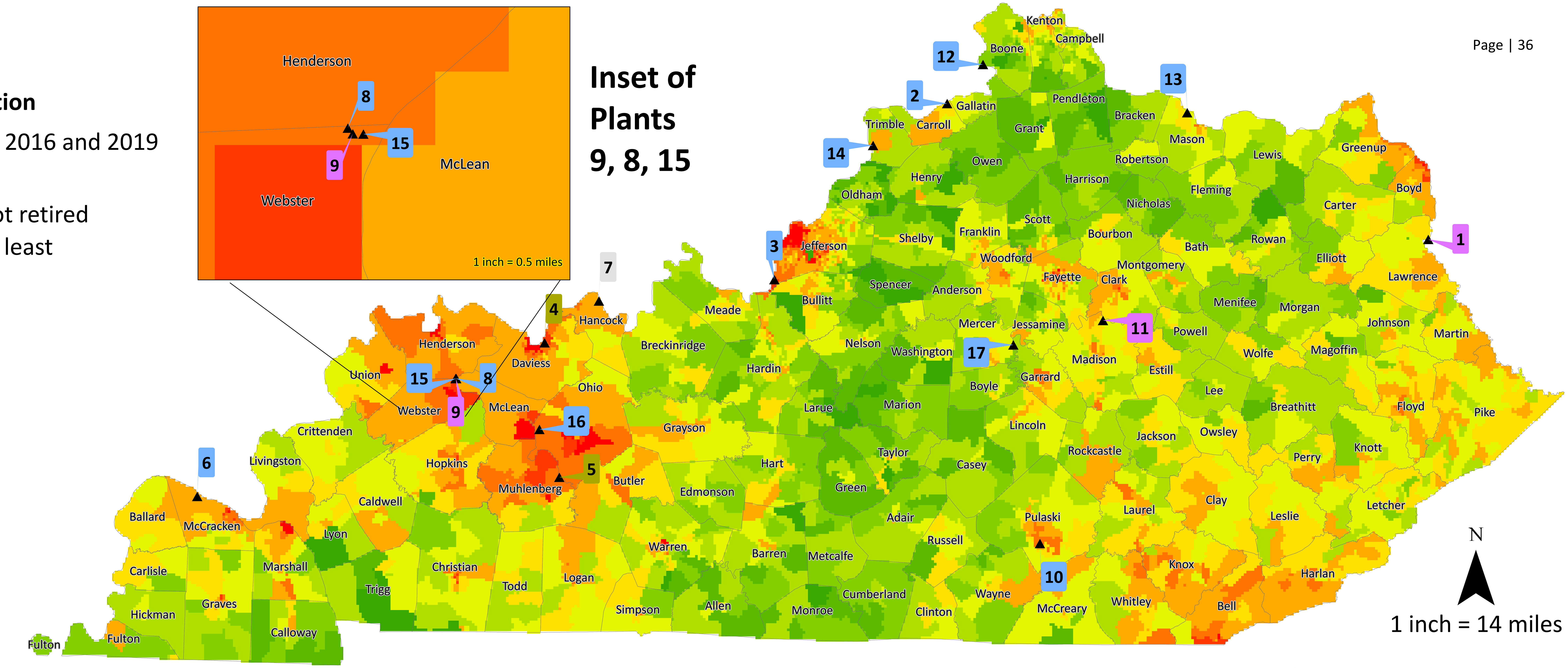
Plant Status Symbolization

- Retiring between 2016 and 2019
- Operating
- Not operating, not retired
- Multiple units, at least 1 unit will retire in near future

Lower Vulnerability

1
2
3
4
5
6
7
8
9
10

Higher Vulnerability

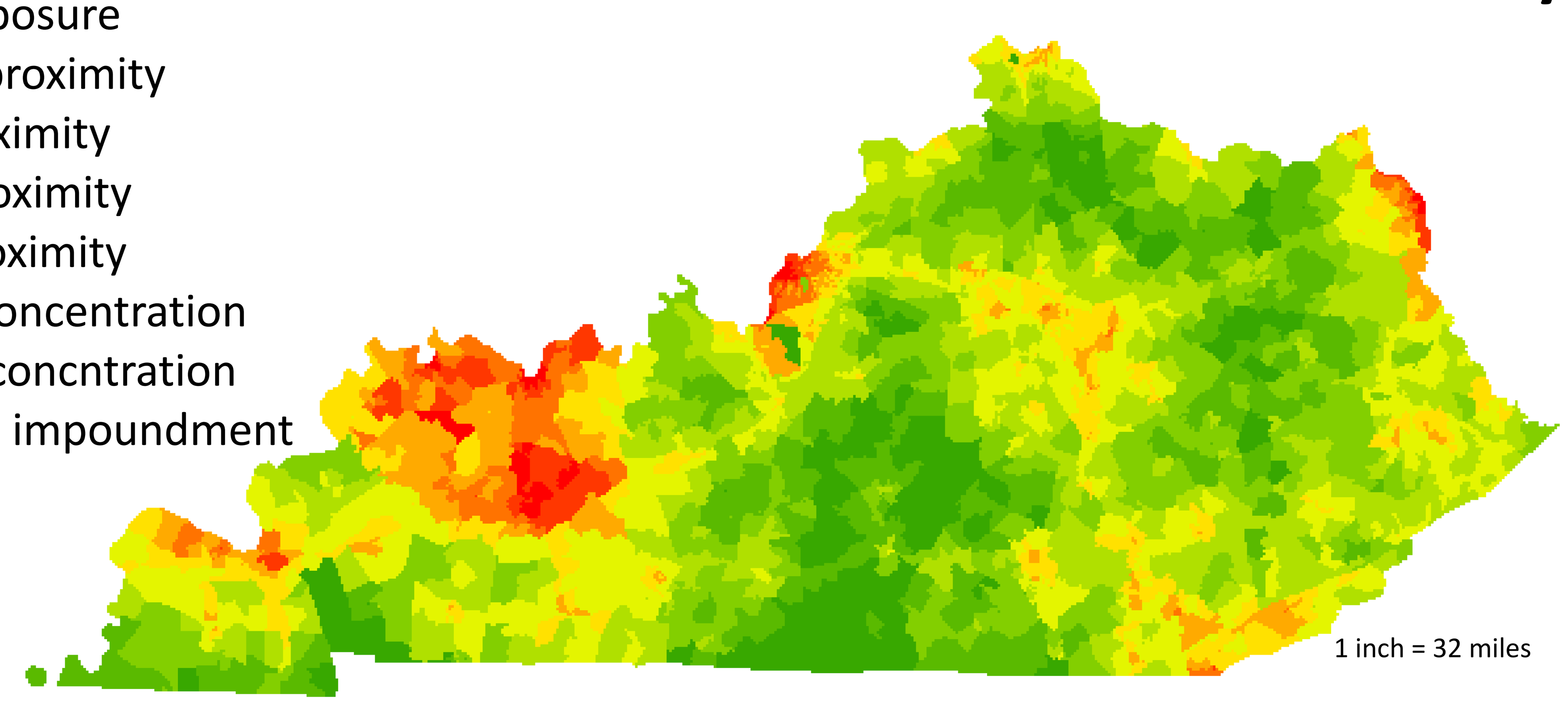


	Plant Name	Utility Name	Total MW	Source(s)
1	Big Sandy	Kentucky Power Co	260	Coal (260 MW)
2	Ghent	Kentucky Utilities Co	1919	Coal (1919 MW)
3	Mill Creek (KY)	Louisville Gas & Electric Co	1472	Coal (1472 MW)
4	Elmer Smith	City of Owensboro - (KY)	398.7	Coal (398.7 MW)
5	Paradise	Tennessee Valley Authority	2201	Coal (2201 MW)
6	Shawnee	Tennessee Valley Authority	1206	Coal (1206 MW)
7	Kenneth C Coleman	Big Rivers Electric Corp	443	Coal (443 MW)
8	HMP&L Station Two Henderson	Big Rivers Electric Corp	312	Coal (312 MW)
9	Robert A Reid	Big Rivers Electric Corp	123	Coal (65 MW), Natural Gas (58 MW)
10	Cooper	East Kentucky Power Coop, Inc	341	Coal (341 MW)
11	Dale	East Kentucky Power Coop, Inc	149	Coal (149 MW)
12	East Bend	Duke Energy Kentucky Inc	600	Coal (600 MW)
13	H L Spurlock	East Kentucky Power Coop, Inc	1346	Coal (1346 MW)
14	Trimble County	Louisville Gas & Electric Co	2185	Coal (1243 MW), Natural Gas (942 MW)
15	R D Green	Big Rivers Electric Corp	454	Coal (454 MW)
16	D B Wilson	Big Rivers Electric Corp	417	Coal (417 MW)
17	E W Brown	Kentucky Utilities Co	1588	Coal (682 MW), Natural Gas (906 MW)

Pollution Overlay Information

Indicators Included:
- Coal Haul Highway proximity
- Active coal mine proximity
- Lead exposure
- NPDES proximity
- NPL proximity
- TSDF proximity
- RMP proximity
- Ozone concentration
- PM 2.5 concntration
- Coal ash impoundment proximity

Map: Cumulative Pollution Overlay



NOTE: All plants are conventional steam coal and natural gas fired combustine turbine (when natural gas is a MW source)

Data Sources: U.S. Energy Information Administration (EIA) dated March 2016 (power plant names, MW, etc.); KFTC (operational statuses)

Pollution and Demographic Overlay Map with Coal Power Plants

Map by: Laura Greenfield
REV: 20161129

6) Uncertainty and Bias Discussion

This section discusses some key sources of uncertainty and potential bias in the data and analysis used to create the Empower KY EJ Analysis and how results may be impacted. Some of these sources have been shared in other places in this document, especially in Section 2, but are all summarized together here.

First, as mentioned in the data discussion in Section 4, this analysis relies on datasets which are proximity measurements. The proximity data is being used as a stand-in for exposure data, when no exposure data is available or when the indicator cannot be as readily quantified. As explained in the EPA Technical Documentation for EJSCREEN, "there are other aspects of an individual's or a community's environmental concerns that are less readily quantified in terms of emissions, concentrations, or risk estimates. People may be concerned about living near facilities that handle hazardous waste substances, and other potential sources of pollution, such as highways or abandoned waste sites" (2016). The methods of calculating proximity are included in the EJSCREEN documentation as well as in Appendix 3 of this document. Results are impacted because proximity measurements do not accurately reflect the dispersion characteristics of pollution.

Second, it is recognized that this project is a hybrid of methodologies and, thus, leaves exciting room for refinement and improvement. The EPA's method for calculating proximity, explained in detail in the EJSCREEN Technical Documentation, was unable to be replicated exactly for this project (see Appendix 3) due to limitations in resources and processing power. As part of a preliminary project, however, this approach was appropriate and follows the EPA's rationale and 5km/500m buffer standard for calculating proximity. This hybrid affects results because it limits the consistency across methodologies.

Third, while there remains opportunity to weight each of the indicators as they are combined to produce a cumulative overlay map, no weights were incorporated in this preliminary generation of the analysis. There are many challenges to choosing weights and the work team decided that, without conducting expert surveys and statistical analysis, there could be no justification for any weights at this time. The EPA EJSCREEN Technical Documentation touches on this topic, stating that "it is important to acknowledge that there is no objective version of 'equal weighting.'" Meaning, that, although no weights are applied to produce the overlay maps (presented in Section 5), we cannot say that all indicators are "weighted equally" because the values in each dataset are determined via the particular measurements, scaling, and classification for that dataset. Results are impacted because there is the appearance of "equal weight" when, in fact, the overlay maps are just a product of the datasets and their generation.

Fourth, expanding further on thinking about classification of the datasets, it is important to note that the datasets were not originally classified in the same way. Classification (grouping data into similar values) occurs in the first stage of processing the data to be put into the Weighted Sum Overlay (see Appendix 1). One approach is to classify the datasets all the same way to maintain consistency. However, the distribution of data varies from dataset to dataset. Therefore, different classification schemes were applied to best-fit each dataset. Results are impacted because each dataset is examined individually and is classified to best show its “story,” however, the classification schemes were chosen solely based on a data distribution chart and no further analysis, which indicates that a different classification scheme can greatly alter the final product. See Section 4 for information on how each dataset was classified in the overlay prep processing, again outlined in Appendix 1.

Finally, a challenge in this project is incorporating *all* indicators of environmental injustice in Kentucky. This analysis was preliminary but worked to consider many different points of view and experiences in order to create a map which can tell a story of Kentuckians who are impacted. However, the entire story is more complex than this single project. When the first draft of maps was presented at the Empower Kentucky Summit in Fall 2016, this was recognized and all were invited to send feedback and respond to the maps and indicators chosen to represent environmental injustice in Kentucky. The feedback and responses received were incorporated to produce the final products presented in this documentation but it is the hope that any uncertainty and bias currently deciding the indicators which shape these maps can decrease gradually with more opportunities for community engagement.

7) Next Step Research Priorities

The Empower KY EJ Analysis provides a meaningful yet preliminary look at the disproportionate landscape of environmental justice issues in Kentucky relating to pollution exposure and concentration of health problems. The EPA's Environmental Justice Analysis Technical Guidance encourages developers of an EJ project to drive their work by resources, time and data availability, and audience and purpose. This section presents the opportunities for next steps and additional research opportunities which were not able to be completed in this preliminary work but would augment this project for Kentuckians to utilize. These ideas can work to fill the research, methodological, and data gaps in the project.

Statistical Analysis

While this project provides a visual display of concentrations of indicators chosen as environmental justice issues for Kentucky, the project lacks statistical analysis. Spatial statistical analysis can provide increased understanding and a stronger foundation for the datasets and analysis represented in this preliminary work, working to answer questions such as "Is the data clustered (indicating spatial significance) or is it randomly or uniformly distributed?" and "How are the datasets related? How strong is their correlation?" Particular tools to utilize are spatial autocorrelation tools and regression analysis processes.

Continued Public Presentation

Empower Kentucky has continuously worked to engage Kentuckians in conversation about developing a people's energy plan, one that works for all Kentuckians. This Environmental Justice analysis is an important effort within Empower Kentucky. The first draft maps of this project were presented at the Empower Kentucky Summit in Fall 2016 and, as this stage of the project is completed, the work team has brainstormed other avenues to enhance the public presentation. These ideas included a webinar which provides a comprehensive and accessible overview of the project, its processes, data, implications, and next steps and development of an interactive web map which can be used to explore the EJ overlay maps and individual datasets, creating an EJ screening tools for Kentucky.

Data Accessibility

An original goal of this project was to make the data used in the overlay maps accessible to the public. This goal still remains. Transparency in both process and data are integrally important to any sort of analysis like this. In building an accessible database, in addition to publication of this documentation, citizens, researchers, activists, students, teachers, and anyone else can take part in Kentucky-specific environmental justice data exploration and analysis. For example, if a group in Lexington would like to start creating an environmental justice analysis for their city, they can begin by drawing upon the robust database created during this project, augmenting their analysis with other sources and indicators as they see fit.

Increased Data, Map, Analysis Specificity

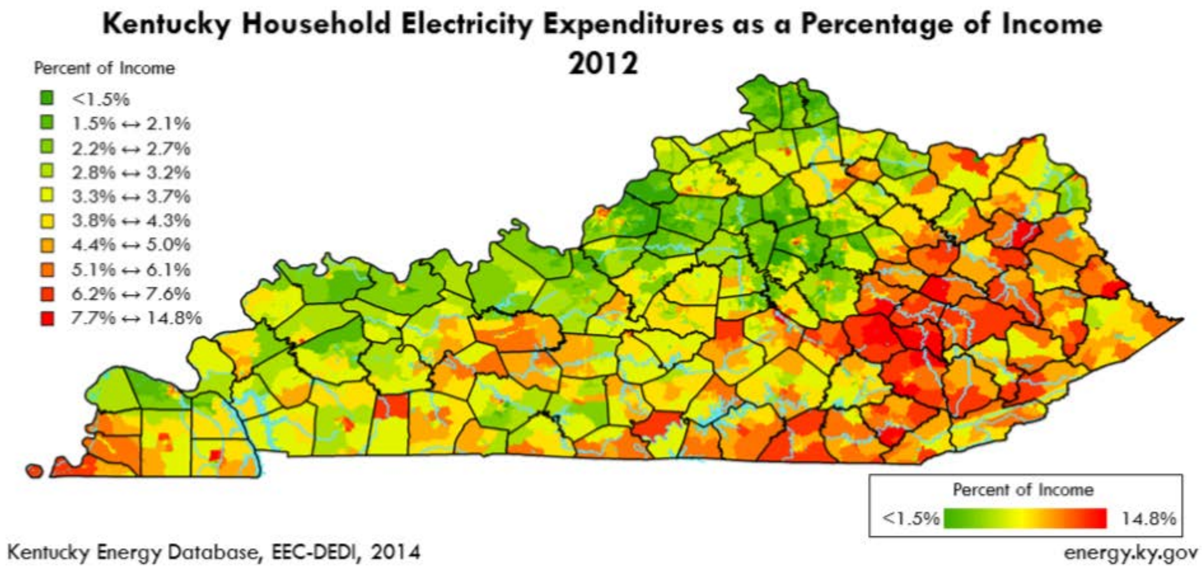
This project examines broad indicators for three categories: Cumulative Pollution Exposure and Concentration of Health Problems, and Demographic Vulnerability. The indicators chosen for each overlay analysis were research-backed and chosen intentionally to generate an EJ exploration of the energy landscape in Kentucky. The analyses would benefit from more in-depth analyses which separates, and even adds indicators, to produce specialized looks at Kentucky's energy landscape. One example is conducting different analysis for rural and urban areas in Kentucky where there are different demographic vulnerabilities and pollution sources. Another example is examining overlay analysis for specific pollution sources, i.e. cumulative pollution exposure due to the coal landscape in Kentucky, etc. The analysis currently focuses broadly on multiple pollution sources in the state, not just coal.

Case Studies

In addition to increased data, map, and analysis specificity, the analysis would benefit from case studies looking at the areas which are highlighted as "most vulnerable" in the current EJ overlay maps and in any overlay that is conducted at a future time. EJSCREEN, the EPA EJ screening tool, is able to generate reports of a specific area chosen by the user. These reports delineate the demographic data for the given area as well as links to a report from the CDC, all specific to the area the user chose. In addition to the EJSCREEN reports, qualitative data, such as narratives and experiences from the communities in the highlighted area, can augment the case studies. In these studies, questions can be explored such as, "What are the particular pollution sources for this area?", "What policies and resources can be directed to this area to support environmental justice work?", and, importantly, "Has KFTC engaged this community? What does the community understand to be the pollution sources and appropriate response?"

Community Distress Maps

We also want to include in this project an analysis of communities that are economically vulnerable in ways that extend beyond pollution and health. Initial conversations and research into this category of maps shaped the idea of generating two maps. The first is a recreation and update of the following map created by the Kentucky Energy and Environment Cabinet and the Department for Energy Development and Independence, appearing in their 2015 Kentucky Energy Profile (found here: http://energy.ky.gov/Kentucky_Energy_Profile/Kentucky%20Energy%20Profile%202015.pdf). This is a map of energy expenditures as a percent of household income.



The second map will be visualizing employment data for various energy sectors, such as mining, transportation, and utilities. Preliminary work has begun to examine the various industry categories within the Quarterly Census of Employment and Wages (QCEW). See the QCEW webpage here: http://www.bls.gov/cew/. The interface being used to look at industry codes can be found here: http://data.bls.gov/cew/apps/data_views/data_views.htm#tab=Tables. This map can be visualized as percent employment loss as well as current employment data to highlight economic vulnerability within a clean energy transition.

These maps will augment the current analysis by providing a look at vulnerability as it has existed and currently exists besides solely focusing on environmental and health factors. Thus, the analysis can become truly an economic and environmental justice analysis.

8) Works Cited

Epstein, M.D., M.P.H., et al. True Cost of Coal, 2010. Harvard Medical School.

ESRI. "How Weighted Sum Works" 2016. *ArcGIS for Desktop*. desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/how-weighted-sum-works.htm. U.S.

Kentucky Department for Public Health. Kentucky Department for Public Health. *Kentucky Asthma Surveillance Report 2008-2012*, 2013. www.chfs.ky.gov/asthma.

U.S. Environmental Protection Agency (EPA). *EJSCREEN Technical Documentation*, 2016. https://www.epa.gov/ejscreen.

U.S. Environmental Protection Agency (EPA). *Technical Guidance for Assessing Environmental Justice in Regulatory Analysis*, 2016. https://www.epa.gov/environmentaljustice.

Appendix 1: Processing Data for Weighted Sum Overlay

Important Note:

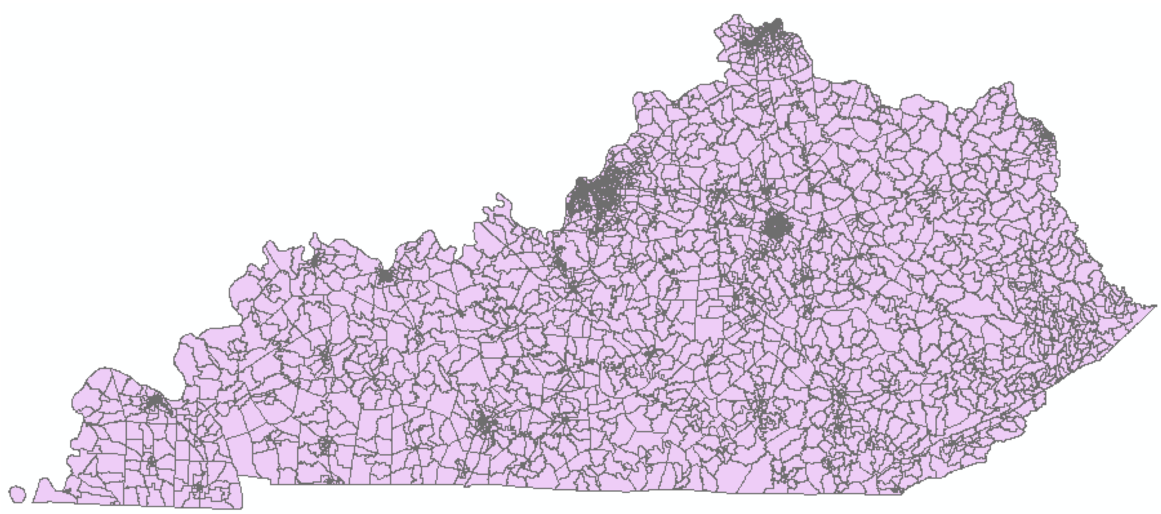
This section outlines the preparation processing required for datasets so that they may be utilized in a Weighted Sum Overlay Analysis (see Appendix 2), i.e. contribute to an overall overlay of the various topics examined in this project: Demographic Vulnerabilities, Cumulative Pollution Exposure, and Exposure-Related Health Problems. This process is completed for each dataset which originates as a vector file (made up of points, lines, or polygons). The Weighted Sum Overlay Analysis necessitates raster layers which have the same cell size and spatial extent (in this case, the state of Kentucky). The goal of this process is to create the layers necessary for the Overlay Analysis. For this project, this process is completed using ArcGIS Desktop 10.3.1.

Process:

I. *Classify Vector File*
II. *Generate Index Value*
III. *Produce Raster Layer*

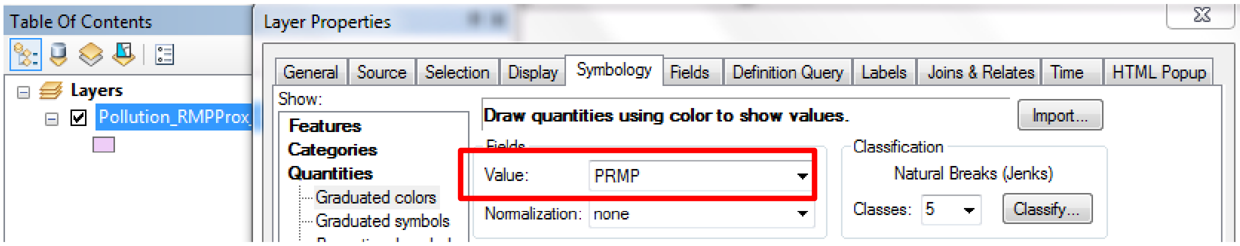
I. *Classify Vector File*

a. Load vector file (a dataset containing points, lines, or polygons, versus a raster which contains pixels/cells) into ArcGIS. Below, the vector file for the indicator of Ozone is an example. This dataset was exported from EPA's EJSCREEN and holds values for Proximity to Risk Management Plan (RMP) facilities at the block group level

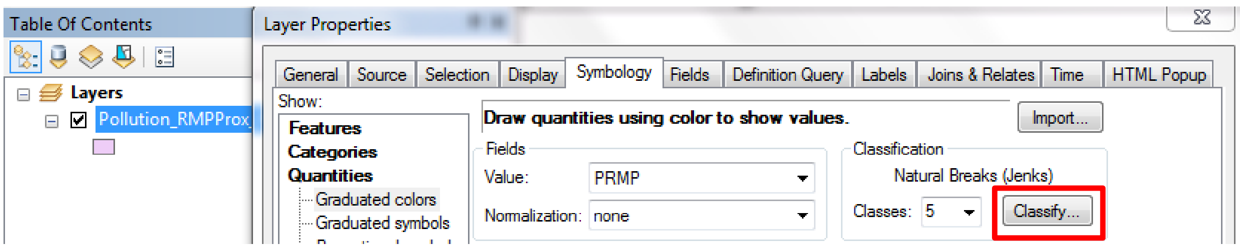


b. Examine the data's distribution
    i. Open Classification Window
        1. Right-click on the layer, select "Properties"
        2. Go to tab "Symbology" and select "Quantities"

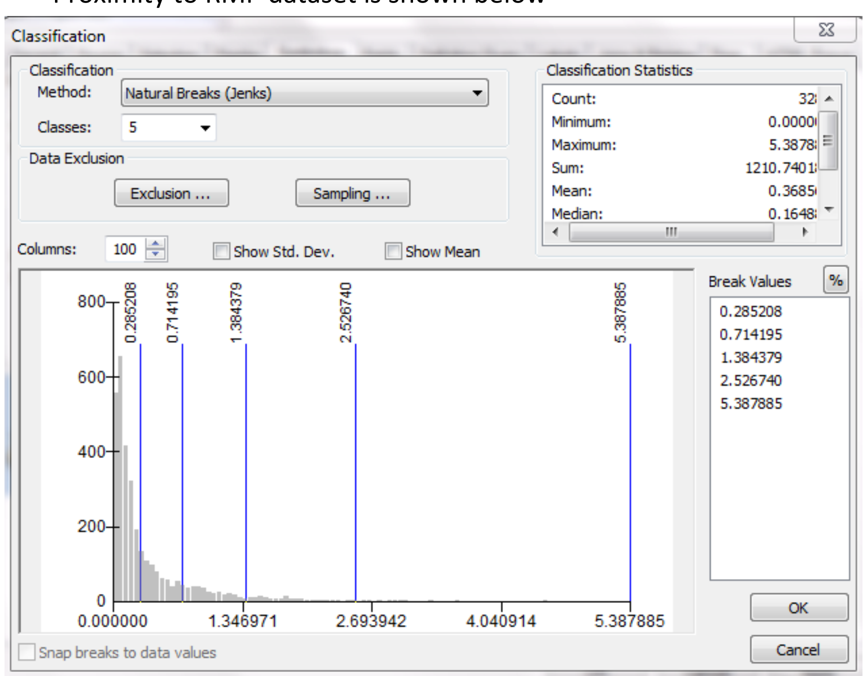
3. For Fields, click the drop-down menu to select the data attribute which contains values for what you are wanting to measure. In this example, we select “PRMP” which is the attribute which holds the values of calculated proximity to RMP plans across Kentucky



4. Next, click the “Classify” button to open the Classification window, shown below



5. In the Classification window, the data’s distribution is shown in a frequency table with the range of data values on the x-axis and frequency (number of times of occurrence of the values) on the y-axis. The Classification window and data distribution frequency table for the Proximity to RMP dataset is shown below



ii. Classify the data

1. Determine Best-Fit Classification

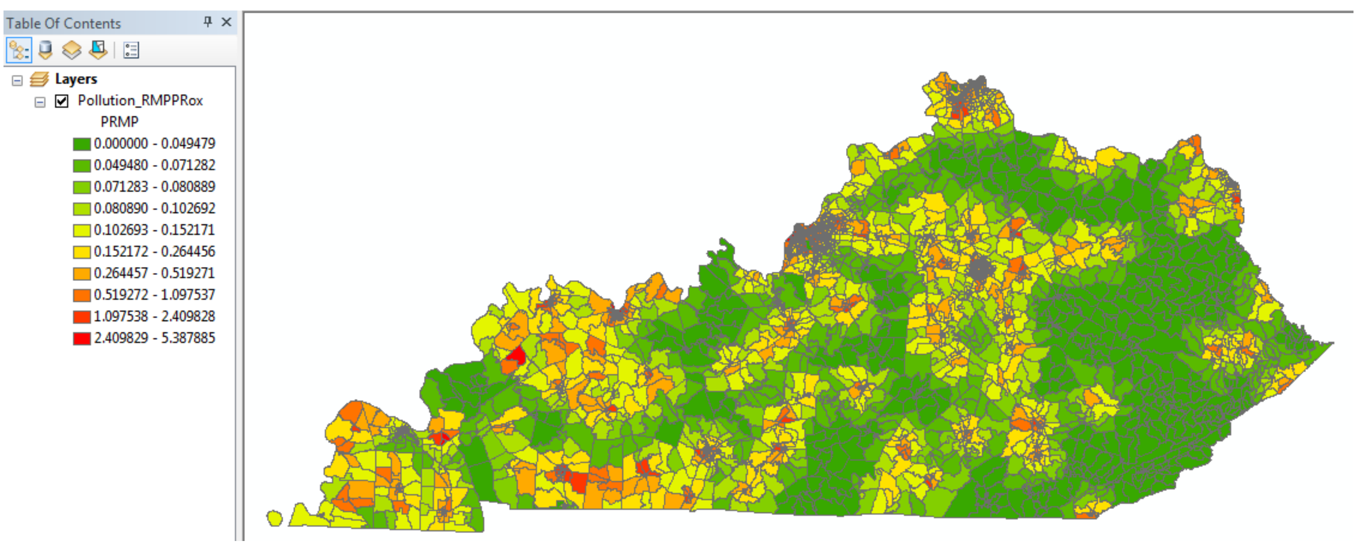
   a. Choose a classification method that best-fits the data's distribution. For more information on choosing classification methods, refer to ESRI documentation (LINK). For the Proximity to RMP dataset, Geometrical Interval is chosen as the best-fit method, as it is encouraged to be utilized for datasets that are not equally distributed and have an extreme concentration of values that "drops-off" as the value range on the frequency table increases

2. Classify

   a. In the Classification Window, select the chosen method and set the number of classes to 10

   b. Select "OK" in the Classification Window and "Apply" in the Layer Properties window. The dataset will now visually be classified to the chosen method to 10 classes

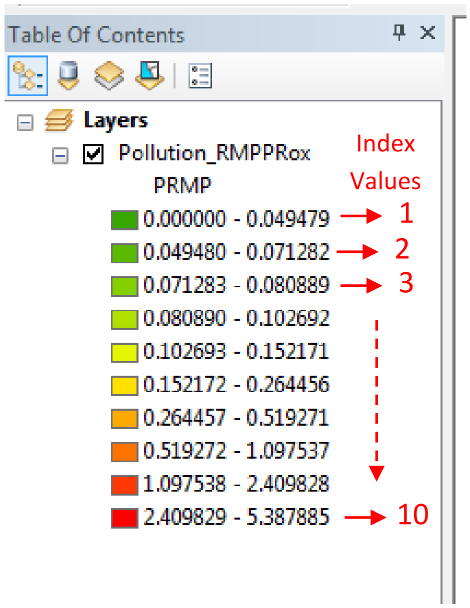
      i. Below, Proximity to RMP is shown classified to Geometrical Interval to 10 classes



II. *Generate Index Value*

a. *NOTE:* Index values, ranging from 1-10, will be created for each of the datasets that will go through a Weighted Sum Overlay. These values will correspond to the 10 classes that were created in Step I. Using the Proximity to RMP example, all block groups with values in the first class (data values ranging from 0-0.049479) will receive an index value of 1, All block groups in the second class (data values ranging from 0.049480-0.071282) will receive an index value of 2, etc. That way, datasets can be compared with each other, despite having original data values which could not be compared

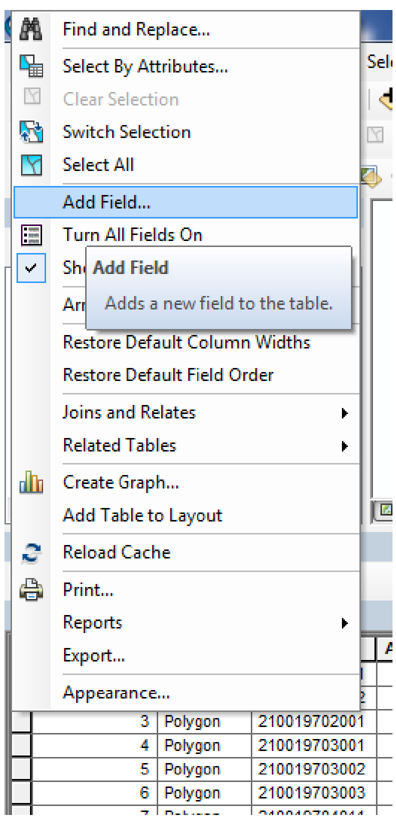
i. There is more discussion about the choices inherent in this process, including regarding classification in Section 5 of this documentation, "Uncertainty and Bias Discussion"



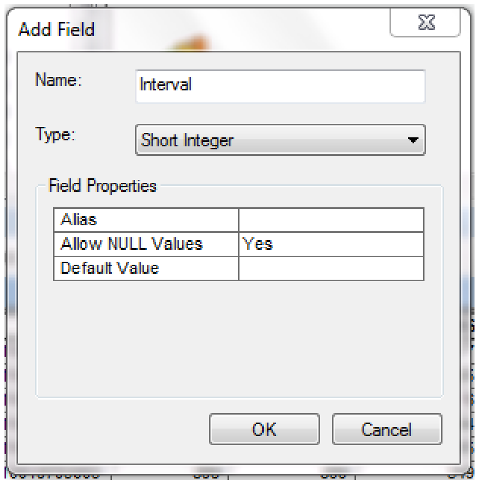
b. Add new value field to the data

i. Right-click on the dataset in the Table of Contents and select "Open Attribute Table"

ii. Open the drop-down menu in the attribute table and select "Add Field..."



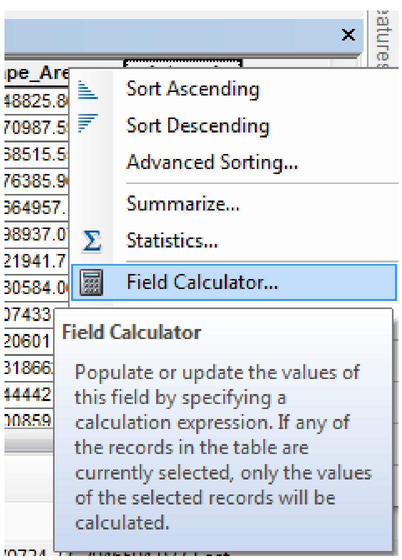
iii. Name the new field. It helps to use a common name for this field across all datasets. This field will hold all index values (per block group, in the case of this Proximity to RMP dataset) so the "Type" of the data in this field will be Short Integer



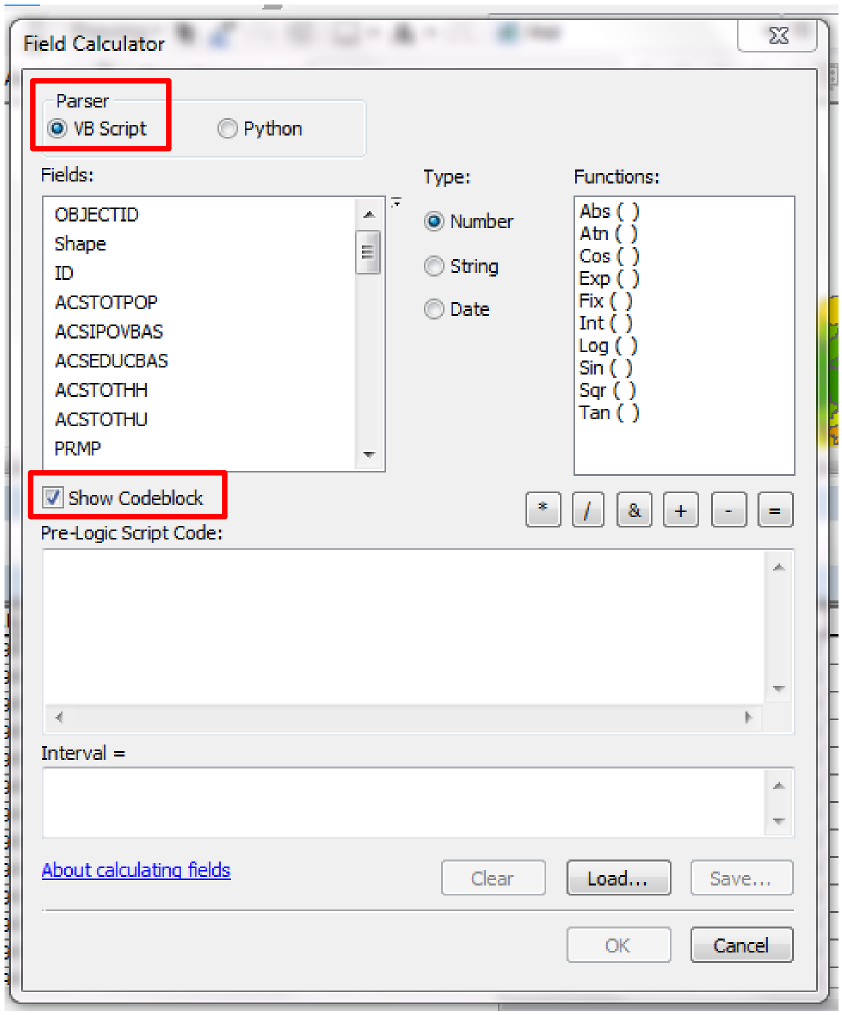
iv. Click "OK" and the new field is added to the dataset's attribute table

c. Run the Index Value Script

i. To populate each of the rows in the attribute table (all block groups in Kentucky, in this case), right-click on the newly created field's name and select "Field Calculator"



ii. In the Field Calculator window, shown below, select VB Script as the Parser and select "Show Codeblock"



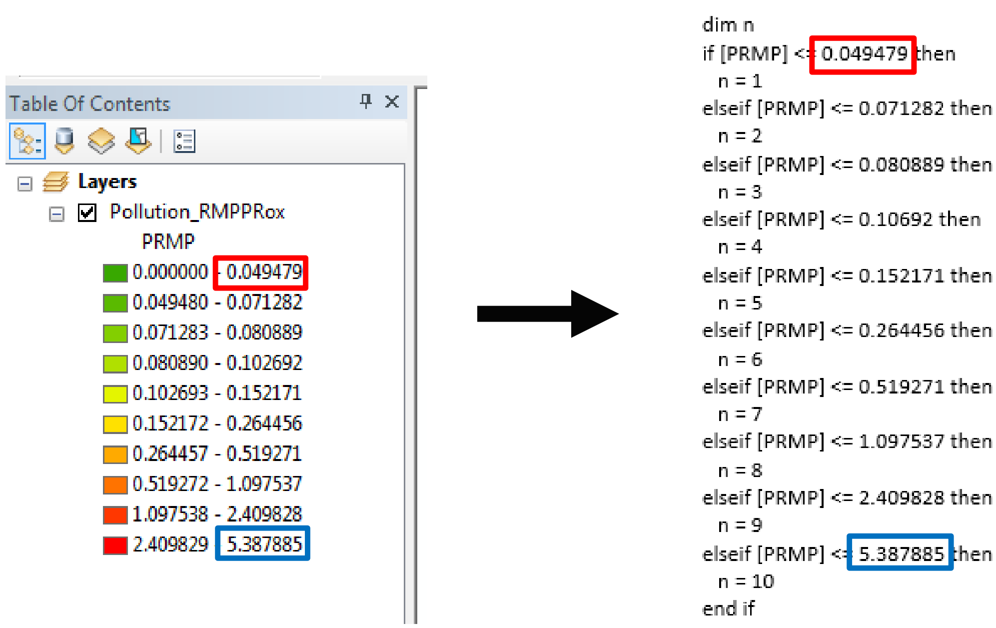
iii. In the box underneath “Interval =” type “n”. The function “n” will be defined in the “Pre-Logic Script Code” box

iv. Shown below is the Index Value Script. This script is written in the “Pre-Logic Script Code” box. Field Calculator, as it processes the dataset row by row, will see that the column, Interval, should have a value of “n” which is defined by this VB Script. The script below uses if/elseif statements to determine the index value of the row, based on the original data value. The portions of the script highlighted need to be replaced in each if/elseif statement to match the current dataset being processed ([PRMP] is the field that contains the original data values; the number after the <= sign comes from the dataset’s classification, see Step 2.c.v.)

```
dim n
if [PRMP] <= 0.049479 then
  n = 1
elseif [PRMP] <= 0.071282 then
  n = 2
elseif [PRMP] <= 0.080889 then
```

```
  n = 3
elseif [PRMP] <= 0.10692 then
  n = 4
elseif [PRMP] <= 0.152171 then
  n = 5
elseif [PRMP] <= 0.264456 then
  n = 6
elseif [PRMP] <= 0.519271 then
  n = 7
elseif [PRMP] <= 1.097537 then
  n = 8
elseif [PRMP] <= 2.409828 then
  n = 9
elseif [PRMP] <= 5.387885 then
  n = 10
end if
```

v. NOTE: Each of the if/elseif statements above correspond to the 10 classes created when the dataset was classified:



vi. When the Index Value Script has been edited in the "Pre-Logic Script Code" box to match the current dataset being processed, select "OK" in the Field Calculator window and the index values for this dataset will be generated in the newly created field

III. *Produce Raster Layer*

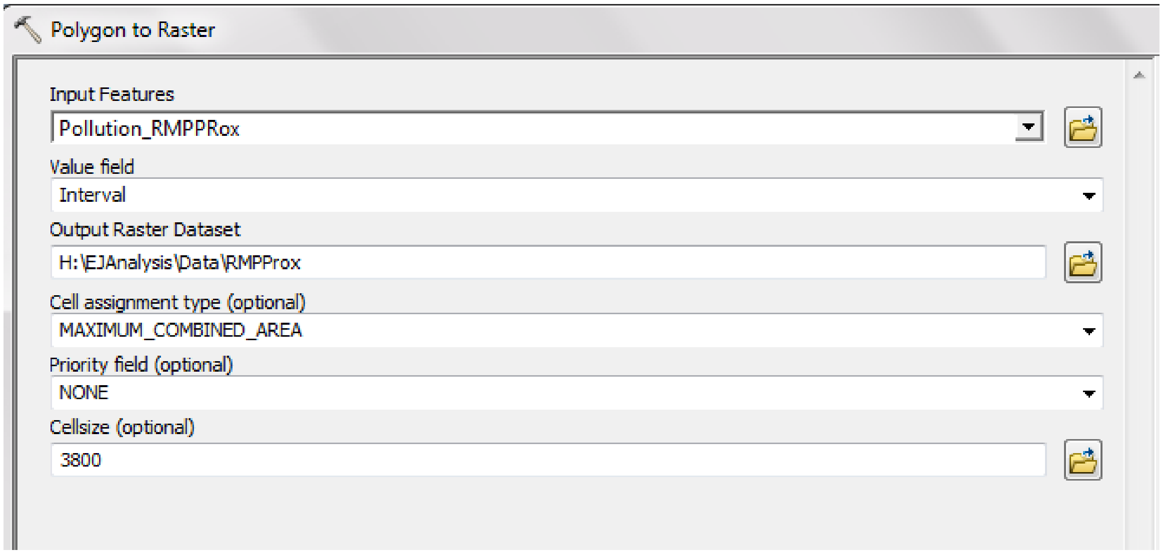
a. In ArcToolbox, navigate to Conversion Tools > To Raster > Polygon to Raster (shown below)

   i. Input Features: the vector layer currently being processed

   ii. Value Field: Interval (the field containing newly generated index values, ranging 1-10)

iii. Priority Field: None

iv. Cellsize: This will depend on the spatial scope of the dataset. This tool will generate a cellsize but a user is able to change it. Cellsize needs to be the same across datasets to be used in the Weighted Sum Overlay

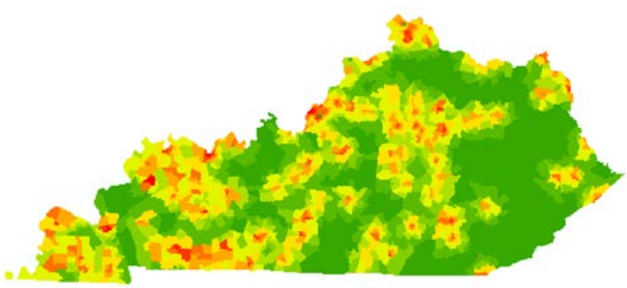


v. Click “OK” to run the tool and convert the vector data to a raster layer

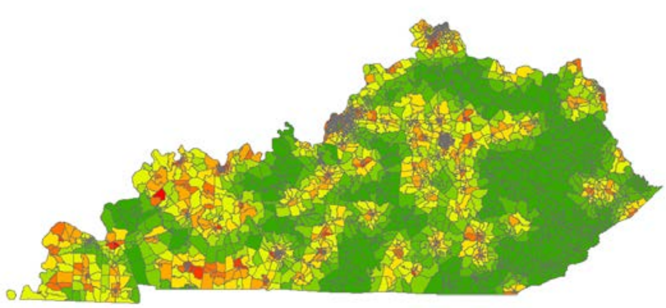
b. Examine the resulting raster layer

i. As an example, the raster layer produced from the Proximity to RMP dataset is shown below, compared to the original vector dataset, classified by index value

Proximity to RMP Raster Layer



Proximity to RMP Vector Dataset



Appendix 2: Generating an Overlay Using Weighted Sum Analysis

Important Note:

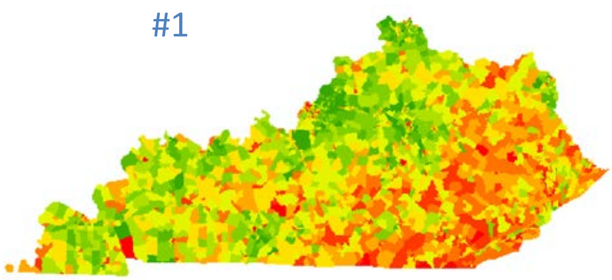
This section outlines how ArcGIS's Weighted Sum Overlay Analysis is utilized in this project to produce overlays to show concentrations of multiple indicators across the state of Kentucky on the topics of Demographic Vulnerabilities, Cumulative Pollution Exposure, and Exposure-Related Health Problems. For this project, this analysis is completed using ArcGIS Desktop 10.3.1. The Weighted Sum Overlay Analysis is completed for each group of indicators (i.e. all Demographic indicators are in an overlay analysis together, all pollution indicators are in another analysis, etc.). The product of this analysis is a raster layer which can then be overlaid again, depending on the user's intent and purpose.

Process:

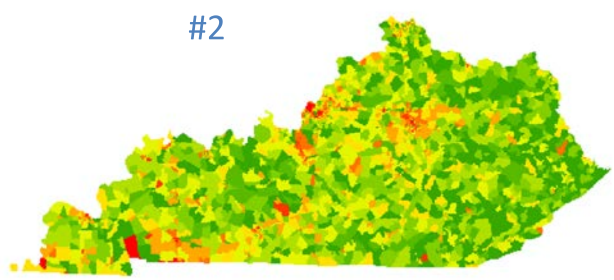
I. *Load Rasters into ArcGIS*

a. In ArcGIS, add all the raster layers which are to be overlaid. These raster layers have the same cell size, the same spatial extent (state of Kentucky), and all have cell values ranging from 1-10, as produced in the process outlined in Appendix 1
b. As example, below are shown 3 of the raster layers used as indicators in this project's Cumulative Pollution Exposure Overlay: Percent Low Income (#1), Percent Minority (#2), and Percent with Less Than High School Education (#3). The layers are symbolized so lower concentration of the indicator (starting with index value of 1) is green, scaling to yellow, orange, and red, which indicates higher concentration of the indicator, i.e. higher vulnerability

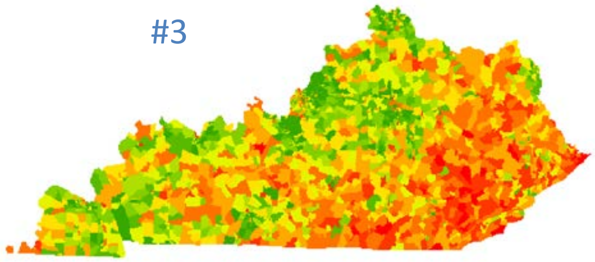
#1



#2



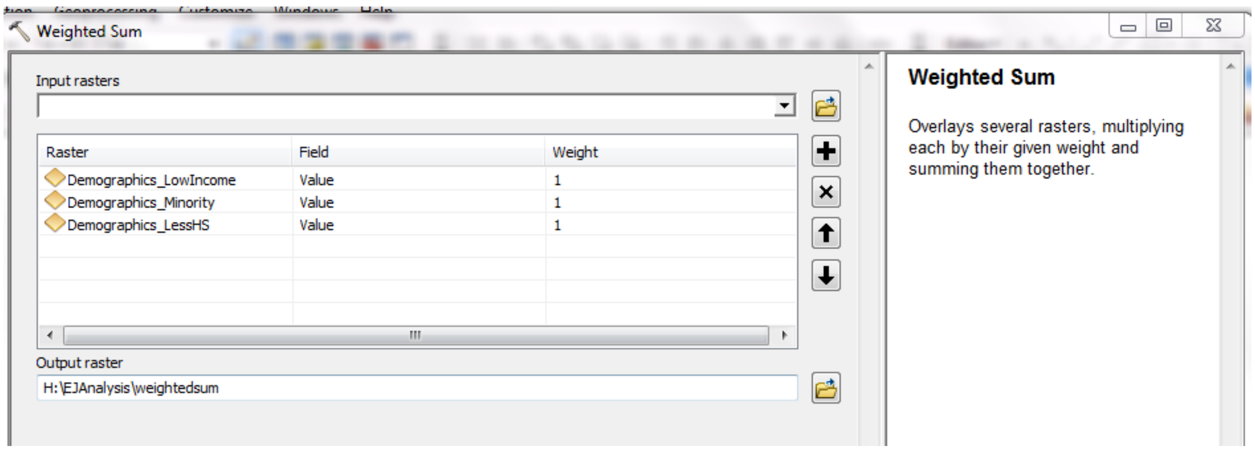
#3



II. *Run Weighted Sum Overlay Analysis*

a. Navigate to ArcToolbox > Spatial Analyst Tools > Overlay > Weighted Sum

b. Add all raster layers that are to be overlaid, as shown below:



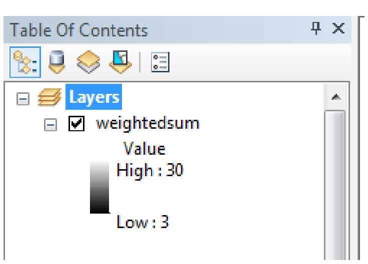
c. Weight the datasets, if intentional and applicable

   i. In this window, a user can change the weight of the datasets (what each cell value will be multiplied by before being added to the other layers' cell values). The weights can be any positive or negative decimal value

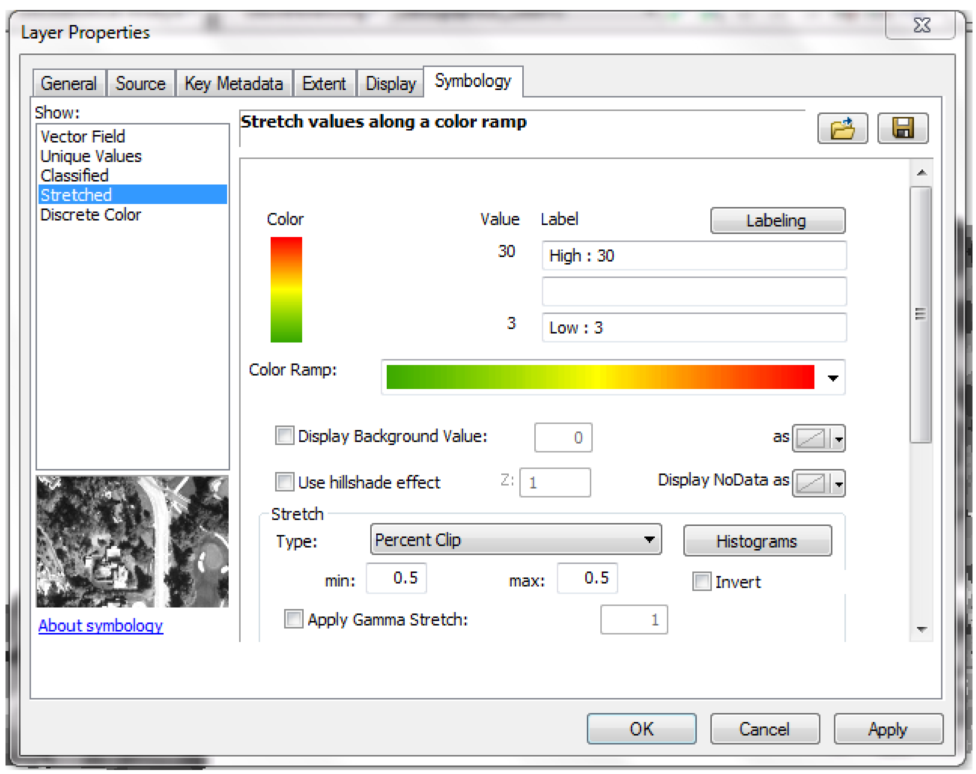
      1. In this example, no weights will be applied to the analysis

d. Symbolize the resulting raster overlay

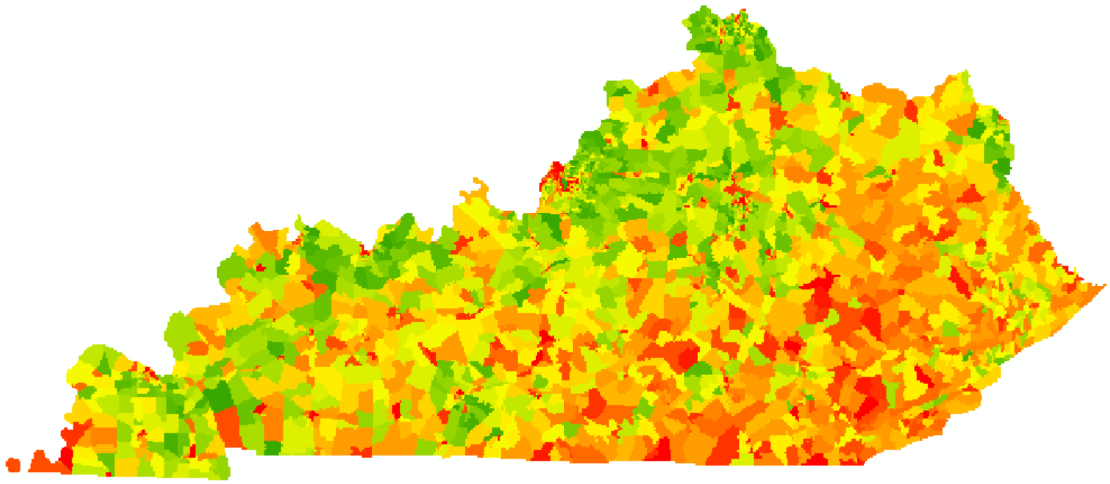
   i. In this example, the Weighted Sum Overlay Analysis produced the following overlay, which contains the sums of index values for the datasets Percent Low Income, Percent Minority, and Percent Less Than High School Education. As shown in the Table of Contents (also below), the tool automatically produces a layer which is symbolized so that a lighter color represents higher concentration and a darker color represents lower concentration



ii. To re-symbolize for a more intuitive visual display, right-click on the raster layer in the Table of Contents, click on Properties, and navigate to the Symbology tab
iii. Choose a color-scheme which indicates higher concentration of indicators where there was a higher sum of index values in the Overlay Analysis. An example is shown below:



iv. The resulting layer is a visual display of cumulative concentration of 3 demographic indicators (with no weights applied to individual indicators):



v. *NOTE:* This raster can be re-classified so the overlay's cell values are again within a range of 1-10 (i.e. can be compared more appropriately with another overlay analysis or used in another run of the Weighted Sum tool). The cell values ranged from 3-30 when produced by the Weighted Sum tool. Re-classification was completed for each of the resulting overlays within this project using the Reclassify tool, found in ArcToolbox > Spatial Analyst Tools > Reclass

Appendix 3: Calculating Proximity

Important Note:

This section outlines the method used in the Empower Kentucky Environmental Justice Analysis to generate proximity indicators, ex. Proximity to surface mines, proximity to coal mines, proximity to coal ash impoundments, etc. Many datasets sourced from EJSCREEN, the EPA's environmental justice screening tool, are proximity indicators (proximity to Risk Management Plan facilities, major direct water dischargers, etc. See Section 3 for complete listing). The EPA's method for calculating proximity is described in the EJSCREEN Technical Documentation and will not be wholly recounted here. Essentially, proximity is a function of the inverse distance (d) from a block's centroid to the given location of a facility, shown below:

$$f(d) = {}^{1}/_{d}$$

Each block receives a proximity score which is the sum of the inverse distances of all the facilities within 5km (or 500m when measuring proximity to roads/traffic). If there is no facility (or road segment) within 5km, the block's proximity score is the inverse distance to the nearest facility. In using inverse distance, facilities that are within a kilometer of a block's centroid contributes twice the proximity score as a facility 2km from the same block centroid. In addition to this, "block-level measures are then aggregated among all the blocks within a block group, weighted by the number of people in the different blocks" (U.S. EPA 2016, *Tech Doc*). Below is an excerpt from the EPA's description of this method, describing the validation for using inverse-distance. For more details, refer to the EPA documentation, cited in the Citations section of this document.

> *We note that we have made a choice in using inverse distance for this function. Air dispersion modeling for pollutants following Gaussian plume assumptions would show a generally greater drop-off in concentration, roughly with the second power to 2.5 power of one over distance. But actual concentrations around individual [facilities] follow often-complex patterns that depend on particular [factors]…* (U.S. EPA 2016, *Tech Doc*).

To attempt to maintain consistency in generating and including proximity indicators other than the ones downloaded from EJSCREEN, this project developed its own process for calculating proximity, using the inverse distance function from the EPA. This method, however, is not able to follow the EPA process exactly. There are significant differences, including:

- Due to limitations in processing power, block group centroids were used, instead of block centroids. There are 161672 blocks in Kentucky versus 3285 block groups
- This method does not incorporate population weighting, meaning block groups with little to no population are given a proximity the same as a block group in the middle of a city. While ideally, population would be considered and, additionally, land-use, measuring proximity in this process can also speak to adverse effects (real or risk of) to our environment as a whole

Therefore, the goal of this process is to replicate the EPA's method for generating proximity indicators to the best ability of current resources and other limitations. For this project, the process is completed using ArcGIS Desktop 10.3.1.

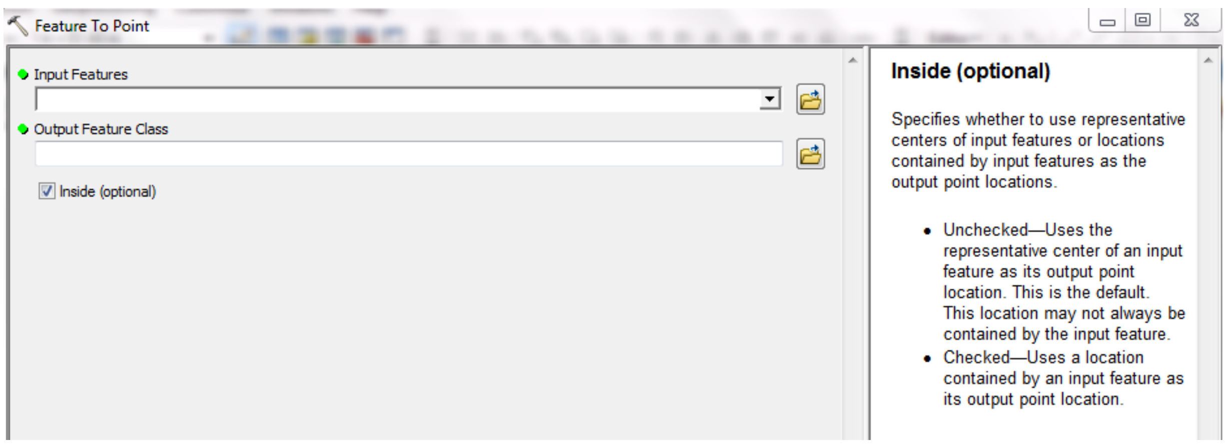
Process:

I. Generate Near Table

a. Convert Block Group Vector to Centroids

   i. Navigate to the Data Management Tools > Features > Feature to Point

   ii. Check the box "Inside" to make sure the point feature class created has centroids that fall within the polygons



   iii. The output from this tool is a feature class containing the centroids of the block groups. Distances to the facilities/road segments will be calculated from these centroids to the facilities/road segments

b. Generate Near Table

   Note: This tool calculates distances and other proximity information between features in one or more layer. The tool produces a stand-alone table and supports finding more than one feature.

   i. Navigate to the tool: Analysis Tools > Proximity > Generate Near Table

      1. Input Features: the newly created block group centroid feature class

      2. Near Features: the features you want to calculate proximity to (i.e. facilities, road segments, etc.)

         a. Note: Near Features can be points, lines, or polygons. For calculating proximity to mines (both surface and underground) and coal ash impoundments, the centroids of these respective features were used as the Near Features

3. Search Radius, Location, Angle, and Find only closest feature are left blank/unchecked
4. Method: PLANAR

ii. A sample of the output Near Table is shown below. For more information on how to read and use the table, refer to ESRI's documentation on the Generate Near Table tool. A new OBJECTID is created for this table. The IN_FID corresponds to the OBJECTID from the original centroid feature class. The NEAR_FID corresponds to the OBJECTID from the near features (facility/mine centroids, etc.). The NEAR_RANK column ranks the near features in order to proximity to the input features per block group centroid.

Table_CoalHaul_BlockGroup_Prox

OBJECTID *	IN_FID	NEAR_FID	NEAR_DIST	NEAR_RANK
1	1	692	0.042481	1
2	1	849	0.056553	2
3	1	920	0.069527	3
4	1	46	0.219249	4
5	1	148	0.219249	4
6	1	737	0.219249	4
7	1	805	0.219249	4
8	1	503	0.222935	5
9	1	961	0.222935	5
10	1	955	0.224608	6
11	1	461	0.240692	7
12	1	597	0.242296	8
13	1	541	0.258576	9
14	1	635	0.258576	9

c. Create Converted Distance Column

i. The NEAR_DIST column will have the same units as the input feature class' coordinate system. If these units are not kilometers or meters (used by the EPA proximity calculations), create new column in the table
1. Add Field (define Type as Floating Point)
2. Use Field Calculator to convert NEAR_DIST values to the new units

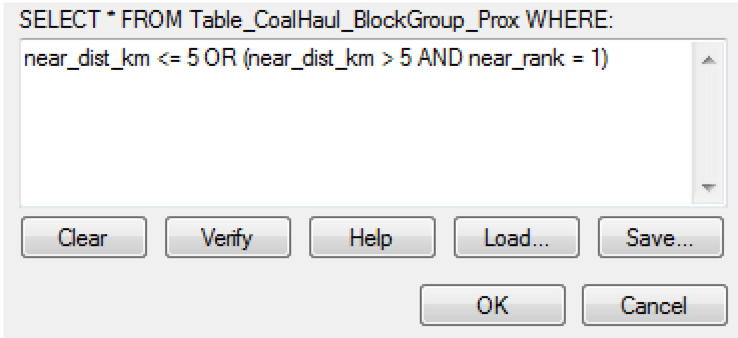
II. *Generate Inverse Near Distances*

a. Query table to capture all the near distances for facilities/sites per block group centroid within 5km (or 500m for road segments) and the closest facility/site to a block group centroid when none are within 5km (or 500m for road segments)

i. Navigate to the Properties of the Near Table (right-click on the Near Table in the Table of Contents and select "Properties")

ii. Navigate to tab "Definition Query" and select "Query Builder"

iii. Copy and paste the following query, replacing field names when appropriate:
near_dist_km <= 5 OR (near_dist_km > 5 AND near_rank = 1)



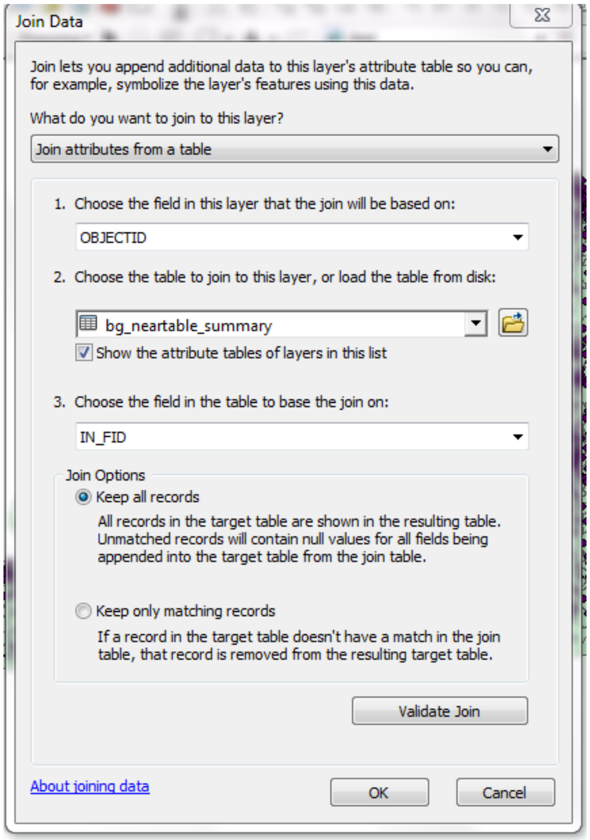
b. Calculate Inverse Distances
    i. Add Field, Type: Float
    ii. Use Field Calculator to calculate inverse of near distance values (1/near_dist_km)

c. Use Summary Statistics to Calculate Sum
    i. In ArcToolbox, navigate to Analysis Tools > Statistics > Summary Statistics
    ii. Choose the field which holds the inverse distances (in km or m, depending on which proximity indicator you are calculating)
    iii. Case Field: IN_FID
    iv. Stat: Sum
    v. The output is a table with, for each IN_FID (i.e. original block group centroid ID), a sum of all inverse distances that fell within the specified distance (5km or 500m) or the closest inverse distance if there was no feature within the specified distance
    vi. Note: make sure that the number of features in this table equal the number of features in the original geography feature class (i.e. 3285 block groups)

III. *Rejoin Data*

a. The summary table needs to be joined with the original centroid feature class, which will then be joined to the original geography (block group polygon) feature class for use in the overlay analysis

b. Join Centroid and Summary Table
    i. Right-click on the centroid feature class in the Table of Contents
    ii. Select “Joins and Relates” > “Join…”
    iii. The join will be completed using the OBJECTID field in the centroid feature class and the IN_FID field in the summary table, as shown below:



iv. Complete this join. All attributes from the summary table are now joined to the centroid feature class

c. Join Centroid and Geography Feature Classes
   i. In ArcToolbox, navigate to Analysis Tools > Overlay > Spatial Join
   ii. Target Features: original geography feature class
   iii. Join Features: centroid feature class (which now includes the summary table information)
   iv. Join Operation: Join One to One
   v. Match Option: Have Their Center In (because the centroid feature class holds points which are in the center of the original geographies)
   vi. Complete this join. All attributes from the centroid feature class are now joined to the original geography feature class, including, importantly, the sum column holding the data which will be analyzed and mapped as the proximity indicator