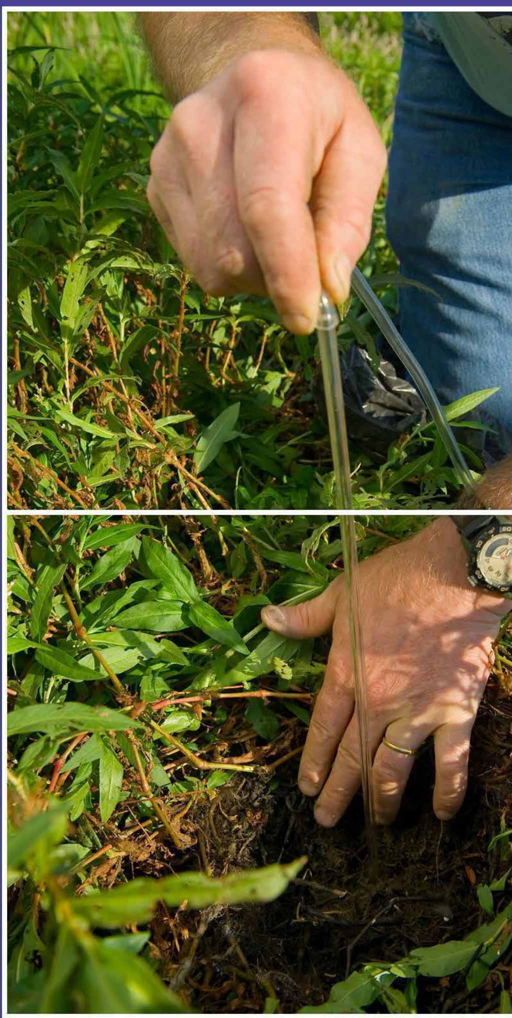
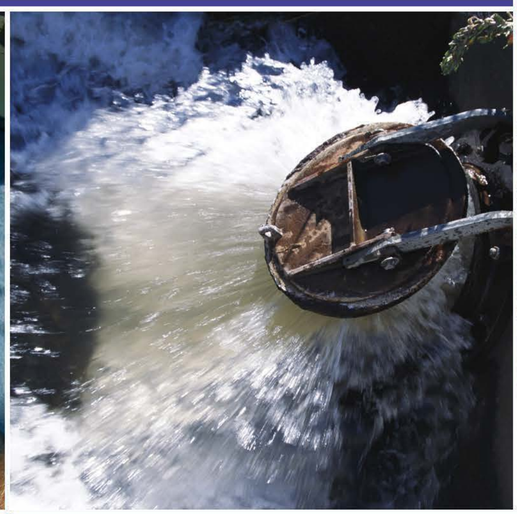
United States Environmental Protection Agency

EPA/600/R-20/367 December 2020 www.epa.gov/ord





Environmental Quality Index 2006-2010

Technical Report

Office of Research and Development
Center for Public Health and Environmental Assessment
Public Health and Environmental Systems Division

EPA/600/R-20/367

United States
Environmental Protection
Agency

ENVIRONMENTAL QUALITY INDEX 2006-2010, Technical Report

Public Health and Environmental Systems Division

Epidemiology Branch
Chapel Hill, NC

Office of Research and Development
Public Health and Environmental Systems Division

Acknowledgments

Project Personnel

Danelle T. Lobdell, United States Environmental Protection Agency (EPA), Office of Research and Development (ORD), Center for Public Health and Environmental Assessment (CPHEA)

Kristen M. Rappazzo, EPA, ORD, CPHEA

Stephanie DeFlorio-Barker, EPA, ORD, CPHEA

Alison K. Krajewski, Oak Ridge Institute for Science and Education (ORISE) Postdoctoral Grantee

Lynne C. Messer, Oregon Health and Science University-Portland State University School of Public Health, Support Contractor

Jyotsna S. Jagai, University of Illinois at Chicago, Support Contractor

Christine L. Gray, Duke University, ORISE Grantee

Monica P. Jimenez, Oak Ridge Associated Universities (ORAU) Student Services Contractor

Achal Patel, ORAU Student Services Contractor

Barbara Rosenbaum, General Dynamics Information Technology/Woolpert, Inc. (GDIT/Woolpert), Geographic information systems (GIS) Contractor Support

Steven Jett, GDIT/Woolpert, GIS Contractor Support

External Peer Reviewers

Sheryl Magzamen, Department of Environmental and Radiological Health Sciences at Colorado State University's College of Veterinary Medicine and Biomedical Sciences.

Anne M. Roubal, County Health Rankings & Roadmap at University of Wisconsin Population Health Institute

Ying Zhou, Infant Outcomes Monitoring, Research, and Prevention Branch, Division of Birth Defects and Infant Disorders, National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention

Internal Peer Reviewers

Tom Brody, EPA Region 5

Linda Harwell, EPA ORD, Center for Environmental Measurement and Modeling

Table of Contents

List of Table

List of Figures

List of Acronyms

ACRES	Assessment, Cleanup, and Redevelopment Exchange
AQS	Air Quality System
C-CAP	Coastal Change Analysis Program
CO	Carbon monoxide
CWA	Clean Water Act
EPA	United States Environmental Protection Agency
EQI	Environmental Quality Index
FARS	Fatality Annual Reporting System
FBI UCR	Federal Bureau of Investigation Uniform Crime Report
FIPS	Federal Information Processing Standard
GIS	Geographic information systems
GTFS	General Transit Feed Specification
HAP	Hazardous air pollutant
HUD	Housing and Urban Development
LEHD	Longitudinal Employer-Household Dynamics
LQG	Large Quantity Generators
MRLC	Multi-Resolution Land Characteristics
MSHA	Mine Safety Health Administration
NADP	National Atmospheric Deposition Program
NATA	National-Scale Air Toxics Assessment
NCOD	National Contaminant Occurrence Database
NGS	National Geochemical Survey
NLCD	National Land Cover Database
NO2	Nitrogen dioxide
NPDES	National Pollutant Discharge Elimination System
NPL	National Priorities List
NPUD	National Pesticide Use Database
NWI	National Walkability Index
PCA	Principal component analysis
PM	Particulate matter
PM_{10}	Particulate matter below 10 micrometers (µm) in aerodynamic diameter
$PM_{2.5}$	Particulate matter below 2.5 micrometers (µm) in aerodynamic diameter
PWS	Public water systems
RAD	REACH Address Database
RCRA	Resource Conservation and Recovery Act
ROE	Report on the Environment
RUCC	Rural-urban continuum code
SD	Standard deviation
SDWIS	Safe Drinking Water Information System
SO2	Sulfur dioxide
SSTS	Section Seven Tracking System
TIGER	Topologically Integrated Geographic Encoding and Referencing
TOD	Transit Oriented Development
TRI	Toxic Release Inventory
TSD	Treatment, Storage, and Disposal
U.S.	United States
USDA ERS	United States Department of Agriculture Economic Research Service
WATERS	Watershed Assessment, Tracking, and Environmental Results
WQS	Water quality standards

1.0
Overview of Report

An overall Environmental Quality Index (EQI), which represents multiple domains of the ambient environment, including air, water, land, built, and sociodemographic, for all counties in the United States, was created for the period 2000-2005[1]. It was developed to provide a better estimate of overall environmental quality and to improve the understanding of the relationship between environmental conditions and human health. This report describes the efforts to update the EQI for all counties in the United States for the 2006-2010 period. The EQI was created for two main purposes: (1) as an indicator of ambient conditions/exposure in environmental health modeling and (2) as a covariate to adjust for ambient conditions in environmental models. However, with the public release of the EQI and variables that constructed the EQI, other uses may emerge. The methods applied provide a reproducible approach that capitalizes almost exclusively on publicly available data sources.

This report is written for audiences interested in the construction of the EQI and is technical in nature. The created variables, EQI, domain-specific indices, and EQI stratified by rural-urban continuum codes (RUCCs) are available publicly at the United States Environmental Protection Agency's (EPA's) Environmental Dataset Gateway. Also, an interactive map of the EQI is available at EPA's GeoPlatform.

2.0
Background

Conceptually, the EQI accounts for the multiple domains of the environment with which humans interact (see Figure 1). These domains include chemical, natural, built, and sociodemographic environments that have both positive and negative influences on health. People move in and out of these positive and negative influences. Also, the positive and negative influences often are co-located.

Brief Overview of EQI 2000-2005

The EQI 2000-2005 was developed in four steps: (1) The five domains were identified, (2) data for each of the five domains were located and reviewed, (3) environmental variables were developed from the data sources, and (4) data were combined in each of the environmental domains; then these domain indices were used to create the overall EQI. The EQI relied on data sources that were mostly available to the public. Below is a summary of the creation of the county level EQI 2000-2005. For more detailed technical information, see the technical report for EQI 2000-2005 [1] located at the Environmental Dataset Gateway.

EQI 2000-2005, Summary of Creation

Domain Identification. Based on three sources, (1) the Report on the Environment (ROE) [2], (2) literature review, and (3) experts, five environmental domains were identified and developed for the EQI: (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic.

Data Source Identification and Review. Predetermined constructs were identified to represent each domain. Based on those constructs, data sources were explored to provide variables representing those constructs.

Air Domain: Three data types were considered: (1) monitoring data, (2) emissions data, and (3) modeled estimates representing two constructs: concentrations of either criteria air pollutants or hazardous air pollutants (toxics). Twelve data sources were identified, and seven were considered for the EQI. Two were used for the air domain of the EQI because they were the most complete.

Water Domain: Five broad data types within the water domain were identified: (1) modeled, (2) monitoring, (3) reported, (4) survey/study, and (5) miscellaneous data. Eighty data sources were identified. Five were used for the water domain of the EQI representing seven constructs: water quality, general water contamination, recreational water quality, domestic use, deposition, drought, and chemical contamination.

Land Domain: Land domain data sources were grouped into five constructs: (1) agriculture, (2) pesticides, (3) contaminants, (4) facilities, and (5) radon. Eighty sources were identified. Eleven were retained.

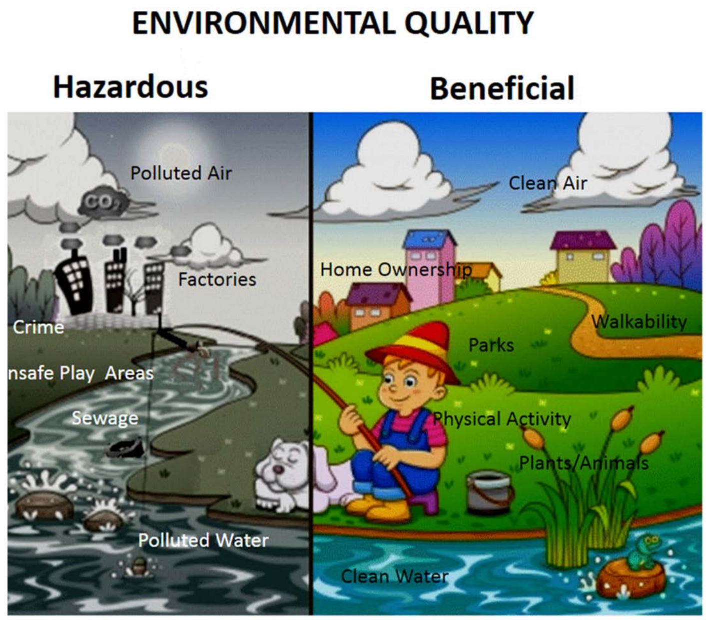


Figure 1. Conceptual environmental quality - Hazardous and beneficial aspects.

Built-Environment Domain: Built environment considered five data types: (1) traffic-related, (2) transit access, (3) pedestrian safety, (4) access to various business environments (such as the food, recreation, health care, and educational environments), and (5) the presence of subsidized housing. Twelve data sources were identified, and four were retained for the built-environment domain of the EQI for five constructs: (1) roads, (2) highway road safety, (3) public transit behavior, (4) business environments (physical activity, food, health care, and educational), and (5) subsidized housing.

Sociodemographic Domain: The sociodemographic domain is represented by crime and socioeconomic constructs. Only two data sources were identified for the sociodemographic domain of the EQI, one for each of the constructs.

Variable Construction. After researching and choosing data sources, variables were created to represent each of the five domains. New variables were created because raw data sources were not always appropriate for statistical analysis.

The process for selecting and creating variables included

- making variables for each domain for each available year of data (2000-2005),
- looking for highly correlated variables that are giving the same information statistically and deciding which of the variables best represents the environmental domain (and remove the extra variables),

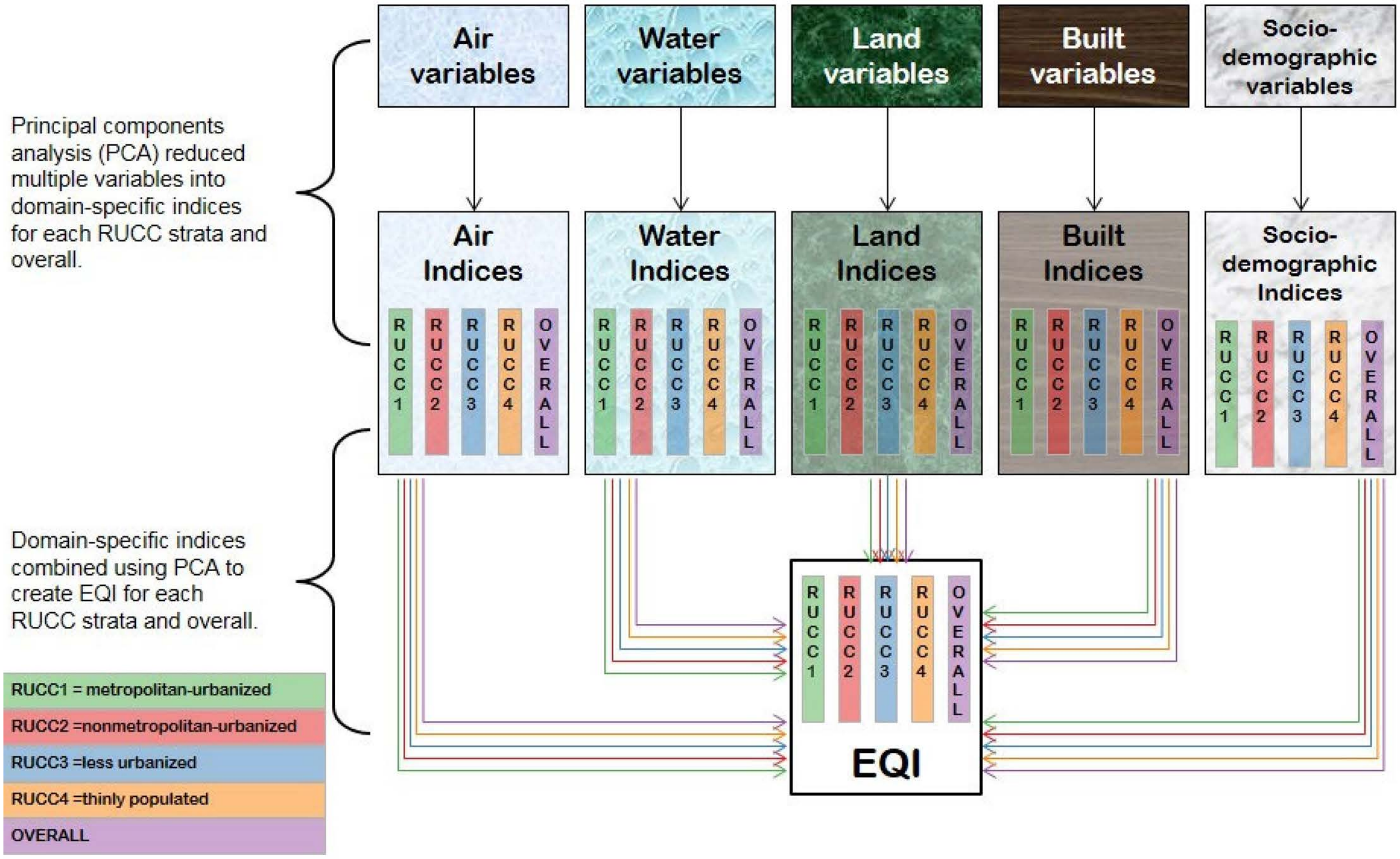


Figure 2. Principal component analysis for the Environmental Quality Index (EQI). All counties included with four rural-urban continuum codes (RUCCs).

- looking for missing data,
- looking at the distribution and statistical properties of each variable and deciding how it should be scaled for analysis, and
- averaging variables from 2000-2005 for each county.

Data Reduction and Index Construction. After variables were created, they were combined into a single index (the EQI) using statistical methods. Each domain has its own index (air domain index, water domain index, etc.). Next, each of the domain-specific indices was used to create the overall EQI. The statistical process used to add these variables together is called principal component analysis (PCA). Figure 2 shows the steps that include

Since the creation of the EQI 2000-2005, multiple studies were conducted examining the relationship between overall environmental quality and health outcomes, including preterm birth [3], mortality [4], cancer incidence [5], asthma prevalence [6], physical inactivity and obesity [7], infant mortality [8], and pediatric multiple sclerosis [9]. A complete list of references related to EQI and health outcomes is shown in Appendix I.

3.0
Development of the EQI 2006-2010

Overview

The development of the EQI 2006-2010 followed mostly the same protocol as the EQI 2000-2005. The majority of constructs identified for each of the five domains in the EQI 2000-2005 were maintained as the basis for variable identification, with the exception of one deletion each in the water domain and land domain and constructs added to the water domain, land domain, sociodemographic domain, and the built-environment domain. Most data sources remained unchanged. Principal components analysis was used to develop the indices. However, using lessons learned from the creation of the EQI 2000-2005, some modifications were adopted to improve the EQI 2006-2010; these modifications included exploring new data sources that were not available during EQI 2000-2005 development, assessment of all variables for continued inclusion in the EQI, and assessment of variables' valence within a domain and valence correction. This section outlines the development of the EQI 2006-2010 through (1) data source identification and review, (2) variable construction, and (3) data reduction and index construction.

Data Source Identification and Review

Approach

Data Selection

An index that comprehensively captures the total environment relating to human health requires numerous variables representing the full range of health-influencing exposures. From within each domain identified in the conceptual model (air, water, land, sociodemographic, and built environments), specific constructs or major areas were identified (Table 1). In general, the identified constructs from EQI 2000-2005 were maintained for the EQI 2006-2010. However, in the water domain, we removed the ”recreational water quality” construct as it only provided data for 231 counties in the United States with beach recreational waters. Because of this low representation, the variables in this domain had extremely low loading values in the Principal Components Analysis; therefore, they were removed in the 2006-2010 EQI. In addition, a dataset representing drinking water quality was identified and, therefore, we were able to include ”Drinking water quality” construct. In the land domain, the ”Contaminants” construct was eliminated. We eliminated these data because they were not the same quality as the rest of the data for the EQI. There was a lack of updated contaminants data,

Table 1. Constructs for each environmental domain.

Domain	Constructs
Air	Criteria air pollutants Hazardous air pollutants
Water	Overall water quality General water contamination Domestic use Atmospheric deposition Drought Chemical contamination Drinking water quality (new 2006-2010)
Land	Agriculture Pesticides Facilities Radon Mining activity (new 2006-2010)
Sociodemographic	Socioeconomic Crime Political character (new 2006-2010) Creative class representation (new 2006-2010)
Built Environment	Roads Highway/road safety Commuting behavior Business environment Housing environment Walkability (new 2006-2010) Green space (new 2006-2010)

and, because of the high correlation between this construct and constructs in other domains, contaminants of this type were better represented by water contaminant data. Also, in the land domain, a ”Mining activity” construct was added. The sociodemographic domain added two new constructs: (1) political character and (2) creative class representation. There was a change in how educational attainment was represented in the 2006-2010 EQI. That change in the education variable from percent of adults with greater than high school education in the 2000-2005 EQI to percent of adults with a college education in the 2006-2010 EQI resulted from inclusion of an education variable with more variability, as almost all citizens have a high school education at this time. The built-environment domain added two new constructs: (1) walkability and (2) green space. Data sources were explored to identify variables that represent the identified constructs for construction. All data sources used for EQI 2000-2005 were reviewed for data updates, and a subsequent search was conducted to identify potential new data sources.

We had solid representation of data for most domains, and we sought to ensure continuity and comparability for the 2006-2010 EQI. Still, our update required identification of new data sources to ensure representation of identified constructs. Because the team came to appreciate the limitations and knowledge gaps in data from the original EQI, the data source identification process was different for the 2006-2010 period than that undertaken for the original (2000-2005) EQI. For example, because of limitations in the National Geochemical Survey representing the geology construct in the land domain, we looked for alternative sources and are now using mines data in the land domain. In recognition of gaps, such as the absence of walkability in the built domain and absence of political climate in the sociodemographic domain, we sought additional data sources to represent the new constructs that we believed would represent more fully the environmental quality of a county.

The details of the new data sources that were identified and included in the EQI 2006-2010 are included in the data source descriptions below.

Data Source Search

Once the desired constructs were identified, the research team conducted an extensive search for potential sources for data to represent those constructs. In general, a broad approach to searching for data sources was undertaken to

- identify EPA and non-EPA domain-specific environmental data sources for all counties in the 50 states of the United States;
- summarize environmental data source availability, quality, spatial and temporal coverage, storage requirements, and acquisition steps; and
- obtain the identified data.

Possible data sources were identified using Web-based search engines (e.g., Google), site-specific search engines (e.g., federal and state data sites), literature-reported data sources (e.g., PubMed, ScienceDirect, TOXNET), and personal communications from data owners. Data that were available at or had the potential to be aggregated to the U.S. county level were sought. Data were restricted to represent the years 2006-2010.

Data Quality and Coverage Assessment

Once potential data sources were identified, several criteria were used to assess sources for inclusion in the EQI. First, constructs representing each domain were identified. Data sources were evaluated as to whether variables could be developed to represent the construct. If a data source could provide variables for a construct in the domain, then data quality and data coverage were used to evaluate data sources for use in the EQI. Data sources of the highest quality were sought. Quality was assessed by one or more of the following ways: Through documentation and discussion with the data source managers, in data reports and internal documentation, project investigators, and the larger field of environmental research through use and critique of the various data sources. Data coverage, which included spatial and temporal components, was more challenging to achieve. Coverage for the entire United States, including Alaska and Hawaii, was one important spatial criterion. Often, it was relatively straightforward to identify high-quality data on a few individual locations or a small geographic area, but the EQI was developed to represent all counties (N=3143) in all 50 States. A second spatial criterion was county-level representation, so data had to be constructible at the county-level for inclusion (e.g., average of point measures or census tract values). Temporally, ideal sources would have had annual data for the 2006-2010 period. At minimum, however, at least some data must have fallen within the 2006-2010 period or close to this time. In theory, a “perfect” data source would have variable measurements at high temporal and spatial resolutions. In practice, data often met one but not both criteria, and evaluation of trade-off values was required, along with consideration of data quality. Unfortunately, some of the data sources used in EQI 2000-2005 did not have any updates for the 2006-2010 period. Redundant data sources that were determined to meet the criteria for inclusion but were not selected for inclusion were retained for use in sensitivity analyses.

Summary of Activities

Table 2 identifies the data sources that were acquired and used for the construction of the EQI and includes a description of the data source and variables constructed from data source.

Table 2 Sources of data for air, water, land, built-environment, and sociodemographic domains for use in the county Environmental Quality Index 20006-2010

Source of Data	Description	Variables*	EQI version
Air Domain			
Air Quality System (AQS 2006-2010) [10]	Repository of ambient air quality data, including both criteria and hazardous air pollutants (HAPs)	PM_{10} - Particulate matter under 10 µg in aerodynamic diameter (µg/m^3 5-year average); PM_{25} - Particulate matter under 2.5 µg in aerodynamic diameter (µg/m^3 5-year average); NO_2 - Nitrogen dioxide (parts per billion [ppb] 5-year average); SO_2 - Sulfur dioxide (ppb 5-year average); O_3 - Ozone (parts per million (ppm) 5-year average); CO - Carbon -monoxide (ppm 5-year average)	2000-2005 and updated 2006-2010
National-Scale Air Toxics Assessment (NATA 2005)[11]	Estimates of HAP concentrations using emissions information from the National Emissions Inventory and meteorological data input into the Assessment System for Population Exposure Nationwide model	A_TeCA - 1,1,2,2-tetrachloroethane (tons emitted per year); A_112TCA - 1,1,2-trichloroethane (tons emitted per year); A_DBCP - 1,2-dibromo-3-chloropropane (tons emitted per year); A_Acrylic_acid - Acrylic acid (tons emitted per year); A_Benzidine - Benzidine (tons emitted per year); A_Benzyl_Cl - Benzyl chloride (tons emitted per year); A_Be - Beryllium compounds (tons emitted per year); A_DEHP - *bis*-2-ethylhexyl phthalate (tons emitted per year); A_CCl4 - Carbon tetrachloride (tons emitted per year); A_CS - Carbon sulfide (tons emitted per year); A_Cl - Chlorine; A_C6H5Cl - Chlorobenzene (tons emitted per year); A_chloroform - Chloroform (tons emitted per year); A_Chloroprene - Chloroprene (tons emitted per year); A_Cr - Chromium compounds (tons emitted per year); A_Co - Cobalt compounds (tons emitted per year); A_CN - Cyanide compounds (tons emitted per year); A_DBP - Dibutylphthalate (tons emitted per year); A_EtCl - Ethyl chloride (tons emitted per year); A_EDB - Ethylene dibromide (tons emitted per year); A_EDC - Ethylene dichloride (tons emitted per year); A_Formaldehyde - Formaldehyde (tons emitted per year); A_Glycol_ethers - Glycol ethers (tons emitted per year); A_N2H2 - Hydrazine (tons emitted per year); A_HCl - Hydrochloric acid (tons emitted per year); A_Isophorone - Isophorone (tons emitted per year); A_Mn - Manganese compounds (tons emitted per year); A_MeBr - Methyl bromide (tons emitted per year); A_MeCl - Methyl chloride (tons emitted per year); A_PH3 - Phosphine (tons emitted per year); A_PCBs - Polychlorinated biphenyls (tons emitted per year); A_ProCl2 - Propylene dichloride (tons emitted per year); A_Quinolin - Quinoline (tons emitted per year); A_C2HCl3 - Trichloroethylene (tons emitted per year); A_VyCl - Vinyl chloride (tons emitted per year)	2000-2005 and 2006-2010 (used 2005 NATA only)
Water Domain			
Source of Data	**Description**	**Variables†**	**EQI version**
Watershed Assessment, Tracking and Environmental Results Program Database (WATERS)[12]	Collection of EPA water assessments programs, including impairment, water quality standards, pollutant discharge permits, and beach violations	ALLNPDESperKM_ln - All NPDES permits per 1000 km of stream in county (permits per 1000 km stream length)	2000-2005 and updated 2006-2010

Table 2. continued

National Atmospheric Deposition Program (NADP 2006-2010) [13]	Samples both regulated and unregulated contaminants in public water supplies; maintained by EPA to satisfy statutory requirements for Safe Drinking Water Act	CaAve_ln - Calcium (Ca) precipitation weighted mean (mg/L); KAve_ln - Potassium (K) precipitation weighted mean (mg/L); NO3Ave - Nitrate (NO_3) precipitation weighted mean (mg/L); ClAve_ln - Chloride (Cl) precipitation weighted mean (mg/L); SO4_mean_ave - Sulfate (SO_4) precipitation weighted mean (mg/L); HgAve - Total mercury deposition (ng/M_2)	2000-2005 and updated 2006-2010
Estimated Use of Water in the United States (2010)[14]	County-level estimates of water withdrawals for domestic, agricultural, and industrial use calculated by the United States Geological Survey	Per_TotPopSS - Percent of population on self supply (percent); Per_PSWithSW - Percent of public supply population that is on surface water (percent)	2000-2005 and updated 2006-2010
Drought Monitor Data (2006-2010) [15]	Geographic information systems raster files reporting weekly modeled drought conditions; a collaboration that includes the National Atmospheric and Oceanic Administration, the U.S. Department of Agriculture, and academic partners.	AvgOfD3_ave - Percent of county drought – extreme (D3-D4) (percent)	2000-2005 and updated 2006-2010
	Measures deposition of various pollutants, such as calcium, sodium, potassium, and sulfate, from rainfall	W_As_ln - Arsenic (mg/L); W_Ba_ln - Barium (mg/L); W_Cd_ln - Cadmium (mg/L); W_Cr_ln - Chromium (total) (mg/L); W_CN_ln - Cyanide (mg/L); W_FL_ln - Fluoride (mg/L); W_HG_ln - Mercury (inorganic) (mg/L); W_NO3_ln - Nitrate (as N) (mg/L); W_NO2_ln - Nitrite (as N) (mg/L); W_SE_ln - Selenium (mg/L); W_Sb_ln - Antimony (mg/L); W_Endrin_ln - Endrin (µg /L); W_methoxychlor_ln - Methoxychlor (ug/L); W_Dalapon_ln - Dalapon (µg /L); W_DEHA_ln - Di(2-ethylhexyl)adipate (DEHA) (µg /L); W_Simazine_ln - Simazine (µg /L); W_DEHP_ln - Di(2-ethylhexyl) phthalate (DEHP)(µg /L); W_Picloram_ln - Picloram (µg /L); W_Dinoseb_ln - Dinoseb (µg /L); W_atrazine_ln - Atrazine (µg /L); W_24D_ln - 2,4-D (2,4-Dichlorophenoxyacetic acid) (µg /L); W_BenzoAP_ln - Benzo[a]pyrene (µg /L); W_PCP_ln - Pentachlorophenol (µg /L); W_PCB_ln - Polychlorinated biphenyls (PCBs) (µg /L); W_DBCP_ln - 1,2-Dibromo-3-chloropropane (DBCP) (µg /L); W_EDB_ln - Ethylene dibromide (EDB) (µg /L); W_xylenes_ln - Xylenes (Total)(µg /L); W_Chlordane_ln - Chlordane (µg /L); W_DCM_ln - Dichloromethane (methylene chloride) (µg /L); W_PDCB_ln - 1,4-Dichlorobenzene (p-dichlorobenzene) (µg /L); W_111trichlorane_ln - 1,1,1-Trichloroethane (µg /L); W_Trichlorene_ln - Trichloroethylene (µg /L); W_C2Cl4_ln - Tetrachloroethylene (µg /L); W_benzene_ln - Monochlorobenzene (chlorobenzene) (µg /L); W_Toluene_ln - Toluene (µg /L); W_ethylbenz_ln - Ethylbenzene (µg /L); W_styrene_ln - Styrene (µg /L); W_Alpha - Alpha Particles (Gross Alpha, excluding radon and uranium) (pCi/L); W_DCE_ln - cis-1,2-Dichloroethylene (µg /L)	2000-2005 and 2006-2010 (not updated, used same variables from 2000-2005)
Safe Drinking Water Information System (SDWIS 2006-2010) [17]{United States Environmental Protection Agency (EPA), #966}	Monitoring of public water systems for health-based violations	Coliform_proportion_ln - Total coliform proportion (average number of violations*(population served/ county population)	2006-2010
Land Domain			
Source of Data	**Description**	**Variables†**	**EQI version**
National Pesticide Use Database: 2009[18]	Delineates state-level pesticide usage rates for cropland applications; contains estimates for active ingredients, of which 68 are insecticides, and 22 are other pesticides	insecticide_ln - Insecticide applied (lb); herbicide_ln-Herbicides applied (lb); fungicide_ln - Fungicides applied (lb)	2000-2005 and updated 2006-2010

Table 2. continued

2007 Census of Agriculture Full Report[19]	Summary of agricultural activity, including number of farms by size and type, inventory and values for crops and livestock, and operator characteristics	pct_manure_acres_ln - Manure, acres applied per county acres (percent); pct_nematode_acres_ln - Chemicals used to control nematodes, acres applied per county acres (percent); pct_disease_acres_ln - Chemicals used to control diseases in crops and orchards, acres applied per county acres (percent); pct_defoliate_acres_ln - Chemicals used to control growth, thin fruit, or defoliate, acres applied per county acres (percent); Pct_AU_ln - Animal units, animal units per county acres (percent); farms_per_acre_ln - Number of farms (number); pct_irrigated_acres_ln - Irrigated acres, acres irrigated per county acres (percent); pct_harvested_acres_ln - Harvested acres, acres harvested per county acres (percent)	2000-2005 and updated 2006-2010
EPA Geospatial Data Download Service (2006-2010)[20]	Maintained by EPA and provides locations of and information on facilities throughout the United States; different datasets within this database are updated at different intervals, but most are updated monthly; no set spatial scale across datasets. Some provide addresses, some geocoded addresses, etc.	facilities_rate_ln – Log transformed rate of all facilities per county (proportion)	2000-2005 and updated 2006-2010
Map of Radon Zones[21]	Identifies areas of the United States with the potential for elevated indoor radon levels; maintained by EPA	Radon - Radon zone (ordinal value)	2000-2005 and 2006-2010 (not updated, used same variable from 2000-2005)
Mine Safety and Health Administration (MSHA) Mines Data Set(2006-2010)[22]	Includes status of coal/metal/nonmetal mines under MSHA jurisdiction since 1970	std_coal_prim_pop_ln - Primarily coal mines, mines per county population (proportion); std_metal_prim_pop_ln - Primarily metal mines, mines per county population (proportion); std_nonmetal_prim_pop_ln - Primarily nonmetal mines, mines per county population (proportion); std_sandandgravel_prim_pop_ln - Primarily sand and gravel mines, mines per county (proportion); std_stone_prim_pop_ln - Primarily stone mines, mines per county population (proportion)	2006-2010
National Geochemical Survey[23]	Geochemical data (arsenic, selenium, mercury, lead, zinc, magnesium, manganese, iron, etc.) for the United States based on stream sediment samples		2000-2005; not used in 2006-2010. These data are represented in the water domain with the National Contaminant Occurrence Database (2006-2010) and the National Atmospheric Deposition Program (2006-2010)
Sociodemographic Domain			
Source of Data	**Description**	**Variables‡**	**EQI version**
United States Census (2010)[24]	County-level population and housing characteristics, including density, race, spatial distribution, education, socioeconomics, home and neighborhood features, and land use	Pct_RenterOcc - Percent renter-occupied units (percent); Pct_Vacant_Housing - Percent vacant units (percent); Med_HH_Value - Median household value (dollars); ln_HH_Inc - Natural log transformed median household income (dollars); pct_fam_pov - Percent of families living below federal poverty level (percent); pct_BS - Percent of persons with bachelor's degree or higher, age 25+ (percent); pct_unemp_total - Percent of persons who are unemployed (percent); ln_Occs_Room - Natural log transformed number of occupants per room (count); GINI_est - Measure of income inequality (proportion)	2000-2005 and updated 2006-2010
Uniform Crime Reports (2006-2010) [25]	County-level reports of violent crime	ln_ViolAv - Natural log transformed violent crime rate (log of count of violent crimes / county population)	2000-2005 and updated 2006-2010
Dave Leip's Atlas of U.S. Presidential Elections (2008)[26]	2008 Election results	DEMO2008 - Percent county voting Democratic in 2008 (percent)	2006-2010

Table 2. continued

Source of Data	Description	Variables	EQI version
United States Department of Agriculture Economic Research Service Creative Class County Codes (2010)[27]	An index of a county's share of population employed in occupations that require thinking creatively"	num_CreatClass - Percent county employed in a creative class (percent)	2006-2010
Built-Environment Domain			
Source of Data	**Description**	**Variables†**	**EQI version**
Dun and Bradstreet North American Industry Classification System codes (2008)[28]	Description of physical activity environment (recreation facilities, parks, physical-fitness-related businesses), food environment (fast food restaurants, groceries, convenience stores), and education environment (schools, daycares, universities) per county	al_pwn_gm_env_rate_ln - Natural log transformed rate of vice-related businesses per county (log of count of businesses / county population); ed_env_rate_ln - Natural log transformed rate of education-related businesses per county (log of count of businesses / county population); neg_food_rate_ln - Natural log transformed rate of negative food resources per county (log of count of businesses / county population); pos_food_rate_ln - Natural log transformed rate of positive food resources per county (log of count of businesses / county population); hc_env_rate_ln - Natural log transformed rate of health-care-related businesses per county (log of count of businesses / county population); rec_env_rate_ln - Natural log transformed rate of recreation-related businesses per county (log of count of businesses / county population); ss_env_rate_ln - Natural log transformed rate of social service agencies per county (log of count of businesses / county population); civic_env_rate_ln - Natural log transformed rate of civic-related businesses per county (log of count of businesses / county population	2000-2005 and updated 2006-2010
Topologically Integrated Geographic Encoding and Referencing (2009) [29] and NAVTEQ map data[30]	Road type and length per county; road types by county created by joining NAVTEQ map data to Topologically Integrated Geographic Encoding and Referencing (TIGER) county definitions	SecondaryRoadProportion - Proportion of all roads that are secondary roads (proportion)	2000-2005 and updated 2006-2010
Fatality Annual Reporting System (2006-2010)[31]	Annual pedestrian-related fatality per 100,000 population; maintained by National Highway Safety Commission	Ln_fatalities - Natural log transformed rate (count / county population) of fatal car crashes per county (log transformed count / county population)	2000-2005 and updated 2006-2010
Housing and Urban Development Data (2010)[32]	Housing authority profiles provide general housing details (low-rent and subsidized/Section 8 housing); information updated by individual public housing agencies.	total_units_ln - Natural log transformed rate of the sum of the following two variables (low_rent_units - Count of low-rent units per county [count] and section_eight_units - Count of section eight units per county [count]) (log of summation of units / county population)	2000-2005 and updated 2006-2010
United States Census (2010)[24]	County-level population characteristics, including density, race, spatial distribution, education, socioeconomics, home and neighborhood features, and land use	CommuteTime - Time it takes to travel from home to work (min); ln_PubTrans -Natural log of percent of county residents who report using public transportation (percent)	2006-2010
EnviroAtlas Green space dataset (2011, 2005-2011)[33]	Description of 20 different land covers for National Land Cover Database (NLCD)[34] and 24 for Coastal Change Analysis Program (C-CAP)[35]; given as percent of county	NINDEX_open - Percent of county land area classified as natural land cover and open space developed land cover (percent)	2006-2010
EPA's National Walkability Index (NWI) (2010)[36]	Characterizes every census block group walkability on a score from 0 to 20 based on four variables: (1) mix of employment types and occupied housing, (2) mix of employment types in a block group, (3) street intersection density, and (4) predicted commute mode split – proportion of workers in the block group who carpool	sum_NWIBG - Walkability score (ordinal)	2006-2010

*Air domain: All variables are natural log transformed with the exceptions of A_edb, A_formaldehyde, O_3, PM_{10}, and PM_{25}.

†Water, Land, and Built domains: Variables with _ln indicated natural log transformation.

‡Sociodemographic domain: ln_ indicates natural log transformation.

Data sources highlighted in blue are new data sources added to 2006-2010 EQI. Data sources highlighted in gree are data sources used in 2000-2005 EQI but are not included in 2006-2010 EQI.

Air Domain

Two constructs represent the air domain: (1) criteria air pollutants and (2) hazardous air pollutants (HAPs). The Air Quality System (AQS)[10] was used to construct variables for the criteria air pollutants and the National-Scale Air Toxics Assessment (NATA) database[11] was used to construct variables for the HAPs.

The AQS is a repository for criteria ambient air pollution data collected by federal, state, local, and tribal agencies from thousands of monitors for EPA's ambient air monitoring program across the United States. Monitored pollutants include all criteria air pollutants, PM species, and approximately 60 ozone precursors. Major strengths of the AQS are that data are measured, rather than modeled, and these measurements are synchronized across the country. Monitors in the network and the reported data are audited regularly for accuracy and precision. However, most of the ambient air monitors are located in or near urban areas, leaving many U.S. counties without reported data. In addition, the AQS provides sparse and limited data collection for HAPs.

The NATA database uses data from the National Emissions Inventory[37] to construct air dispersion models for estimating ambient concentrations of HAPs at the county and census-tract levels. Beginning in 1996, the National Emissions Inventory data are constructed every 3 years, providing annual estimates. The NATA databases contain estimated ambient concentrations for 177 to 180 of the 187 HAPs and use validated models that take meteorology and chemical dispersion into account. The methodology for estimating concentrations may change between assessments, but these modifications are well documented and justified. Although the ambient concentrations may be comparable over time, some differences between estimates are attributable to these minor methodological modifications. The temporal resolution of the assessments is adequate for the intended EQI, but, because of the 3-year release schedule, there are gaps in temporal coverage. NATA 2008 was not developed and thus, for EQI 2006-2010, NATA 2005 was used.

Water Domain

The water domain included six data sources: (1) the WATERS program database[12], (2) Estimated Use of Water in the United States.[14], (3) the National Atmospheric Deposition Program (NADP)[13], (4) the Drought Monitor Network[15], (5) the National Contaminant Occurrence Database (NCOD)[16], and (6) the Safe Drinking Water Information System (SDWIS)[17]. Using these six data sources, variables were created to represent seven constructs that describe the overall water environment. The seven constructs were (1) overall water quality, (2) general water contamination, (3) drinking water quality, (4) domestic use, (5) atmospheric deposition, (6) drought, and (7) chemical contamination.

The Watershed Assessment, Tracking, and Environmental Results (WATERS) Program[12] database represents the surface water assessment programs under the Clean Water Act (CWA). A limitation of this data source is that data are maintained at the state level and reported to the federal system. Although all states report county-level data, there is little consistency in the temporal reporting and type of data reported across states. These data were first geocoded to a specific stream length in the National Hydrography Dataset[38] via the REACH Address Database (RAD)[39]. The geocoded WATERS program data were used to calculate human-exposure-related variables, such as percentage of stream length impaired for recreational use. This dataset is the only database maintaining information on EPA CWA regulations, which is a strength.

The National Contaminant Occurrence Database (NCOD)[16] is a surveillance database maintained to satisfy the requirements of the Safe Drinking Water Act. This database includes information on contaminants in public water supplies that are not measured elsewhere. The survey is conducted every 6 years, and data are provided by public water suppliers. The data are limited, as they are provided by public water suppliers, and, therefore, spatial aggregation was needed to get county-level estimates. *Estimated Use of Water in the United States*[14], which is modeled by the United States Geological Survey, provided county-level estimates of water withdrawals (an indication of water stress in a county) for domestic, irrigation, livestock, and industrial use. This dataset already is provided at the county level, which is a strength; however, it is limited, as the estimates are based on several different data sources.

Two data sources provided information on meteorological impacts on water quality. The Drought Monitor Data[15] are modeled weekly drought conditions. Weekly coverage for the entire country is a strength of this dataset. The National Atmospheric Deposition Program (NADP)[13] provided weekly measures and national coverage of the deposition of various pollutants from rainfall using monitors around the country. Again, this database provided weekly information for the entire country; however, it was reported by monitors and required spatial aggregation to achieve county-level estimates.

Drinking water quality data was gathered from the Safe Drinking Water Information System[17] (SDWIS), which is a repository maintained for compliance with federal regulations. This is a new data source to the water domain. SDWIS provides publicly available data based on requirements from the Safe Drinking Water Act. States are required to report basic information about the public water systems (PWS), violations, and enforcement information. The health-based violations provided in SDWIS are not measured elsewhere. Of the SDWIS measures, only total coliform health-based violations were considered for inclusion in the 2006-2010 EQI, as the other contaminant categories have a high frequency of missing data (arsenic: 87.18%; ground water: 97.8%; inorganic chemicals: 97.04%; lead and copper: 90.87%; long-term enhanced surface water treatment rule 1 and 2: 87.69%; nitrates: 91.92%; radionuclides: 89.76%; disinfection and disinfectant by-products: 66.43%; surface water treatment: 90.84%; synthetic organics: 98.79%; and volatile organic chemicals: 98.5%) for health-based violations. Average total coliform health-based violations were used to estimate the proportion of the county population affected by coliform violations between 2006 and 2010.

Land Domain

The land domain included five data sources representing five constructs: (1) Agriculture, (2) Pesticides, (3) Facilities, (4) Radon, and (5) Mining Activity. The data sources identified for this domain include: 2007 Census of Agriculture[19], 2009 National Pesticide Use Database[18], EPA Geospatial Data

Download Service[20], Map of Radon Zones[21], and Mine Safety and Health Administration (MSHA) mines data[22]. The MSHA mines database is a data source new to EQI 2006-2010. Also, the National Geochemical Survey database used in EQI 2000-2005 was not used in EQI 2006-2010.

The 2007 Census of Agriculture Full Report[19] was used to represent agricultural factors. Information on nonpesticide chemicals used in farming, animal units, harvested acreage, irrigated acreage, manure acreage, and proportion of farms was taken from the 2007 Census of Agriculture. The Census of Agriculture[19] data provided mostly farm-related summary characteristics and did not offer direct pesticide measures or probable exposure information. As a strictly environmental indicator, the Census of Agriculture was useful, but its ability to link to human health was somewhat limited. Eight variables from the census of agriculture were included in the EQI.

The 2009 National Pesticide Use Database (NPUD)[18] provides county-level rates of pesticide use. A limitation of the NPUD was its availability only for contiguous states. Pesticides were classified into three pesticide classes and then summed to estimate county-level pesticide use (in kilograms) for herbicides, fungicides, and insecticides. These three pesticide categories were included in the EQI.

The industrial facilities data source, the EPA Geospatial Data Download Service[21], was used to find the following types of sites: Brownfield sites; Superfund sites; Toxic Release Inventory sites; pesticide-producing-location sites; large-quantity generator sites; and treatment, storage, and disposal sites. All facilities-related data were retained for inclusion in the EQI with extensive information on each facility for the years 2006-2010.

The EPA Radon Zone[21] map assigned a radon potential level to each county in the United States. As the data source provided radon potential, not actual measurement, these data were limited. The three-level radon categorization masked important radon-level heterogeneity across the United States. Despite these limitations, the data sources provided land-related data not available elsewhere.

The Mine Safety and Health Administration (MSHA) Mines Data Set[22] was used to create the mining activity construct. The MSHA's dataset includes current and historical coal, metal, and nonmetal mines. The list included the status of each mine (Abandoned, Abandoned and Sealed, Active, Intermittent, New Mine, Nonproducing, Temporarily Idled) and in which county the mine was located. The dataset does not include the size of each mine, so it is possible a mine may span two counties, but only the county indicated by its official address is reported.

The National Geochemical Survey (NGS)[23], used in the 2000-2005 version of the EQI to determine the contaminant construct, was not included in the updated version. The NGS data provided the mean and standard deviations for multiple soil chemicals. However, these values were calculated from multiple surveys of soil samples collected over several years based on local agencies' interests and resources and, therefore, were combining many varying sources of data. Because of high correlation between the NGS and the National Contaminant Occurrence Database and the National Atmospheric Deposition Program, the decision to drop the NGS was made.

Sociodemographic Domain

The original sociodemographic domain included only two constructs: (1) socioeconomics and (2) crime. In an effort to better reflect each county's sociodemographic character, the updated Sociodemographic Domain for EQI 2006-2010 has four constructs: (1) Socioeconomic, (2) Crime, (3) County creative typology (new for EQI 2006-2010), and (4) County political valence (new for EQI 2006-2010). Because counties can be characterized as "working class" or "tech savvy," we added the creative typology to help capture these characteristics. Similarly, counties may be known for their political valence (e.g., a "red" county in a "blue" state); the percent voting Democratic in the 2008 election was added to capture this county characteristic. Only four data sources were identified and retained for the sociodemographic domain: (1) the United States Census Bureau[24], (2) the Federal Bureau of Investigation Uniform Crime Reports (FBI UCRs)[25], (3) the United States Department of Agriculture Economic Research Service (USDA ERS)[27], and (4) Dave Leip's Atlas of U.S. Presidential Elections (2008)[26].

The United States Census[24] reports county-level population and housing characteristics, including population density, race, spatial distribution, socioeconomic characteristics, home and neighborhood features, and land use. One strength of this data source is its national coverage and consistency of data collection with standard methods. One weakness of this data source is its decennial collection.

The FBI UCR[25] provides annual violent and property crime counts and rates for reporting areas. These data are a valuable source of crime exposure, but reporting is not mandatory and may vary by jurisdiction.

The USDA ERS[27] creates a "creative class" index, derived from census data, to identify what proportion of the population may be employed in creative pursuits. This variable helps to characterize counties as being attractive to people in creative work (e.g., physicians, professors, architects). Because this variable is based on census data, it has the same strengths and weaknesses of the United States Census.

Dave Leip's Atlas of U.S. Presidential Elections[26] tracks the political valence of the counties. Political valence tracks with a number of county-level attributes, such as provision of social supports, levels of school funding, etc. Capturing this variability may be useful for differentiating counties from each other. One strength of Dave Leip's Atlas of U.S. Presidential Elections data source is its data quality, and one weakness of this data source is its infrequency of publication.

Each of these data sources represents critical aspects of the human sociodemographic environment and is updated regularly and available at the county-level for the entire country.

Built-Environment Domain

Built-environment data sources were identified for the following constructs: Business environment, Highway safety, Housing, Roads, Commuting practices, Walkability, and Green Space. For EQI 2006-2010, we added two new data constructs with new data sources: one representing green space and another estimating county walkability.

For the road construct, NAVTEQ road map data[30] were joined to Topologically Integrated Geographic Encoding and Referencing (TIGER)[29] county definitions to result in road types by county. The road data from NAVTEQ, whose underlying map database was based on first-hand observation of geographic features, rather than relying on official government maps, is the majority supplier for car navigation systems (around 85% of car makers). The TIGER files provide relatively uniform and nationwide coverage. From these files, county-specific proportions were characterized for various road types. Unfortunately, considerable heterogeneity may be lost; for instance, a tertiary road in Maryland may not be qualitatively equivalent to one located in Wyoming.

The Fatality Analysis Reporting System[31] of the National Highway Safety Commission was retained as part of traffic safety because of its national coverage. The data are regularly updated and available from the Web site. A limitation of these data is that traffic fatalities result from diverse types of events (e.g., from road conditions or substance-involved fatalities), but this diversity is not captured well.

North American Industry Classification System codes through Dun and Bradstreet[28] were used as the data source to estimate five different business environment topics: (1) physical activity, (2) food, (3) educational, (4) social, and (5) health care environments. These data are available as geocoded business addresses. Although these data have sometimes been criticized for inadequate spatial resolution (e.g., inaccurate geocoding to small units of aggregation, such as census tracts), they should be sufficient as a construct for county-level business environments of food, physical activity, and education.

The Housing and Urban Development database[32] includes data on Section 8 and low-income housing. These housing units are a feature of built environments associated with known and suspected health risks and disamenities.

The EPA's National Walkability[36] data is the source of the walkability index. It combines data from 2010 Census TIGER/Line shapefiles, 2010 Census Summary File 1, Census Longitudinal Employer-Household Dynamics (LEHD) 2010, InfoUSA 2011, NAVTEQ NAVSTREETS 2011, General Transit Feed Specification (GTFS) data for 228 transit agencies, and the Center for Transit Oriented Development (TOD) Database 2012 to produce a block group score, which was aggregated to the county level.

The Landcover data derive from the EPA's National Land Cover Database (NLCD)[34]. It represents land cover across the contiguous 48 states, circa 2011. Each 30-m2 pixel has been classified using a standard land cover classification scheme, and some of these categories have been aggregated further according to procedures outlined in EPA's Report on the Environment[40]. Data originally were processed and compiled by the Multi-Resolution Land Characteristics Consortium (MRLC)[41], a United States federal interagency group, based on Landsat satellite imagery. These data are combined with NOAA's C-CAP Land[35] cover county data to represent land cover for all 3143 counties.

Summary of Changes to 2006-2010 data sources from original 2000-2005 EQI

- Air Domain – No changes to data sources
- Water Domain - One data source was added for 2006-2010 (SDWIS), and some variables developed from the WATERS database for 2000-2005 were not used in 2006-2010.
- Land Domain – One data source was eliminated for 2006-2010 (National Geochemical
- Survey). One data source was added for 2006-2010.
- Mine Safety and Health Administration (MSHA) Mines Data Set (2006-2010)
- Sociodemographic Domain – No data sources were eliminated for 2006-2010. Two data
- sources were added to the 2006-2010 EQI.
- USDA ERS Creative class data
- 2008 Presidential Election results data
- Built Domain – No data sources were eliminated for 2006-2010. Two data sources were added to the 2006-2010 EQI.
- EPA National Walkability data
- EPA NLCD + C-CAP data

Variable Construction

Approach

We followed the same approach in developing variables for EQI 2006-2010 that we used for EQI 2000-2005. Most variables throughout the different domains were identified previously and developed as part of the EQI 2000-2005, then were updated for the 2006-2010 period. For the newly added data sources, we developed new variables. We assessed all variables as to whether the new variables needed to be standardized, as a proportion of geographical space (e.g., road proportions) or as a rate per population (e.g., violent crimes per capita) for use in the EQI. Additionally, some data were not available for all counties but required spatial kriging to provide national coverage. Kriging is a geospatial technique that uses known data points to interpolate data at locations with unknown measurements[42].

The overall process for variable development for 2006-2010 was as follows.

- Update or identify and develop relevant variables within each domain for each available year (2006-2010)
- Assess collinearity among the variables within each domain and eliminate redundant variables
- Assess missing data and variability of each variable
- Assess normality of variables and transform as necessary

Appendix II lists all the variables included in the EQI for each of the five domains for 2006-2010 and includes notes about whether the variables were used in the previous version of the EQI or if they are newly created variables. Appendix III provides the variables that were used in EQI 2000-2005 but were not used in the EQI 2006-2010 update. The created variables are available publicly at EPA's Environmental Dataset Gateway.

Identification and Construction of Variables from Data Sources

For each domain, all variables from EQI 2000-2005 were reviewed and assessed for continued inclusion in the EQI 2006-2010. Variables were created from selected data sources to represent the constructs. Variables were developed in a variety of manners, including kriging and standardization by area or population. Each domain section below provides the details of variable construction.

Assessing Variables

The data reduction method Principal Component Analysis (PCA) is based on the variability between variables[43]; therefore, collinearity of variables was assessed. This assessment was done by developing correlation matrices for each domain. Variables with any correlation coefficient >0.70 were examined; representative variables were chosen for each pair or group of highly correlated variables (Appendix IV).

Ideally, developed variables would have measured or estimated values for each county of the United States. When this criterion was not met, or when a majority (>50%) of values were zero, the proportion of missing data and zero values were evaluated for variable inclusion. If a particular variable had information missing for many counties, the nature of the missing data was evaluated. When it was determined that the missing data could be interpreted as meaningful zeros (i.e., no measures were taken because that condition did not occur in that county), the missing values were set to zero. For instance, the counties with no reported public housing were set to zero because public housing is truly absent from some counties. When counties were missing data because reporting areas were centralized, but the data could not be assumed to be truly missing, the data were spatially kriged when possible. For instance, crime was reported only for specific counties, even though it likely occurred in counties other than those in which it was reported as well. Therefore, crime rates were averaged spatially over adjacent counties to create an estimate for a county with no official reported crime. If the missing data could not be determined to be legitimate zeros, the data could not be reasonably kriged or averaged over geography, and the number of counties with missing data was too high (more than 50% of counties), the variable was not used in the EQI.

In some instances, there may have been more than one data source that could represent a particular domain construct. In that case, the data source deemed to have better data quality and coverage was used.

Finally, normality of variables was evaluated. Using PCA, the chosen data reduction technique, a key assumption is that variables are distributed normally[43]. If data were nonnormal, transformations were applied (typically log-transformation) to increase normality. For those variables with zero values, half of the nonzero minimum value was added to all observations before log-transformation.

When data were updated on an annual or regular basis, variable consistency (mean and standard deviation) was compared across each year of the 5-year period (2006-2010).

Summary of Activities

Domain-Specific Variable Descriptions

Air Domain

The air domain consists of two data sources, (1) the AQS[10] and (2) the NATA[11], representing criteria air pollutants and HAPs.

Criteria Air Pollutants

Daily concentration data from the EPA's AQS monitors (point scale) were downloaded for ozone, carbon monoxide (CO), sulfur dioxide (SO2), nitrogen dioxide (NO2), particulate matter under 10 µm in aerodynamic diameter (PM10), and particulate matter under 2.5 µm in aerodynamic diameter (PM2.5). Annual averages were calculated for each of the six pollutants at each monitor with data. These averages then were used in a kriging procedure to estimate annual concentration at each county's center point for each year from 2006 to 2010.

For the EQI spanning 2006 to 2010., a single average concentration was calculated from the annual average concentrations for each county from the kriged estimates. When indicated (i.e., lognormal distribution) half of the minimum nonzero value was added, and variables were log transformed.

Hazardous Air Pollutants (HAPs)

County-level concentrations estimates from NATA were used for all HAPs included in the EQI. HAPs were selected for inclusion from the full NATA pollutant list. Using data from 2005, variables were evaluated for collinearity and variability. Variables with any correlation coefficient >0.70 were examined, and representative variables were chosen for each pair or group of highly correlated variables (see Appendix IV). Correlations were determined after assessing for missingness/zeros and assessing normality. The variable that is correlated with the most other variables was chosen. For example, if variable A was highly correlated with variables B, C, D, and E, but each of those were correlated with a lower number of variables, A was chosen as the representative variable. The nonchosen variables (B, C, D, and E) then were removed from consideration within other groupings. If the correlation group was isolated (i.e., no variables in it were associated with any other variables outside the isolated group), then a representative variable was chosen without particular criteria. By the end, all variables remaining had correlation less than 0.7 with each other. All variables excluded were highly correlated with (represented by) at least one variable that was retained. Of the remaining variables, all missing values were set to zero, with the assumption that lack of estimate for an area indicated low concern for contamination with a particular HAP, and the number of zero values was evaluated for each variable. Pollutants with more than 50% zero values were dropped. This process left 37 HAPs included in the EQI. When indicated (i.e., log-normal distribution), half of the minimum nonzero value was added, and variables were log transformed.

Table 3. 2005 NATA variables included in EQI 2006-2010

1,1,2,2-tetrachloroethane
1,1,2-trichloroethane
1,2-dibromo-3-chloropropane
1-3-dichloropropene
Acrylic acid
Benzidine
Benzyl chloride
Beryllium compounds
bis-2-ethylhexyl phthalate
Carbon tetrachloride
Carbonyl sulfide
Chlorine
Chlorobenzene
Chloroform
Chloroprene
Chromium compounds
Cobalt compounds
Cyanide compounds
Dibutylphthalate
Ethyl benzene
Ethyl chloride
Ethylene dibromide
Ethylene dichloride
Formaldehyde
Glycol ethers
Hydrazine
Hydrochloric acid
Isophorone
Manganese compounds
Methyl bromide
Methylene chloride
Phosphine
Polychlorinated biphenyls
Propylene dichloride
Quinoline
Trichloroethylene
Vinyl chloride

The air domain includes 43 variables representing criteria and HAPs.

Water Domain

The water domain included six data sources: (1) the WATERS program database[12], (2) Estimated Use of Water in the United States[14], (3) the National Atmospheric Deposition Program (NADP)[13], (4) the Drought Monitor Network[15], (5) the National Contaminant Occurrence Database (NCOD)[16], (6) the Safe Drinking Water Information System (SDWIS)[17] Using these six data sources, variables were created to represent seven constructs that describe the overall water environment. The seven constructs were (1) overall water quality, (2) general water contamination, (3) drinking water quality, (4) domestic use, (5) atmospheric deposition, (6) drought, and (7) chemical contamination.

Overall Water Quality

Impairment and water quality standards (WQS) data were obtained for the most recent state reported data that were collected under Sections 303(d) and 305(b) of the Clean Water Act (CWA)[44]. The CWA is administered at the state level, and data are reported voluntarily from the states to the federal level. The dates of the reported data ranged from 2004 to 2010, as the federal reporting system maintains only the most recent data reported by each state. Under Section 305(b) of the CWA, states establish WQS for each hydrological feature based on the expected use (or uses) of these waters. Under Section 303(d) of the CWA, states assess whether waters are impaired (do not meet the standards) for the uses established in the WQS. This assessment is conducted biennially, and the states voluntarily report these data to the federal level.

County-level impaired stream length was estimated for the contiguous United States using impairment and WQS data (from the WATERS database). With the designated uses listed for each state, the WQS was classified into five broad categories of water use: (1) agriculture, (2) drinking water, (3) recreation, (4) wildlife, and (5) industry. Using geographic information systems (GIS), county-level percentages of impairment were calculated. WQS and impairment datasets were joined to the map layer of hydrologic features in EPA's RAD[39]. RAD is a replicate of the National Hydrography Dataset Plus[38] augmented for reporting water quality data. The defined broad water use categories were joined to the WQS data, and a table summarizing hydrologic features with multiple uses was created. WQS and impairment tables were assigned to features in the RAD using GIS Network and Event tools. These tools link tabular database information with linear or polygon features. Stream lengths were clipped by county boundaries to calculate percent impairment by county. Only linear water features were included in each category. Polygon features, such as lakes, were excluded because of the lack of well-defined county and state boundaries across water bodies. Next, county and state designations were linked with linear features in RAD. Once all data were associated to linear hydrologic features, lengths were calculated for water features impaired for any use, drinking water use, or recreational use and for all stream lengths within a county. The final variable was cumulative measure of percent of water impaired for any use.

General Water Contamination

Water contamination can be caused by several sources. Unfortunately, EPA only has consistent data on the point sources of contamination in the form of the number of National Pollutant Discharge Elimination System (NPDES)[45] permits. Therefore, the number of permits in a county was used as a proxy for general water contamination. Using permit information in the WATERS database, 13 variables were calculated for the number of discharge permits in a county. Permits that were current during the period 2006-2010 were selected. The 10 variables that were calculated based on individual permit types had too many missing data; therefore, three composite variables were created for inclusion in the EQI. A composite variable was developed for the number of sewage permits per 1000 km of stream length in a county. The number of animal feeding operations and concentrated animal feeding operations NPDES permits, combined sewer overflow NPDES permits, and NPDES permits for sludge in each county were summed and divided by the total stream length in the county. Similarly, composite variables were calculated for industrial permits (combining the total of pretreatment NPDES permits, general facilities NPDES permits, and individual facilities NPDES permits) and stormwater permits (combining the total of general stormwater NPDES permits, industrial stormwater NPDES permits) by county per 1000 km of stream length. Preliminary analyses demonstrated low loadings for the grouped variables; therefore, only one variable was maintained: the total number of discharge permits per 1000km of stream length in the county.

Drinking Water Quality

In the United States, drinking water quality is measured and maintained by the public water system (PWS) treating and distributing drinking water. Based on the Safe Drinking Water Act, states are required to report basic information about PWS, violation information for each PWS, and enforcement information to the federal system. The SDWIS data is publicly available data through the Fed Data Warehouse[17]. The basic information for the PWSs were merged with the violations reports, so that the county and city served by the violations were together in one report. In instances where there were multiple counties served by a PWS, the counties were separated to account for these violations in both counties served by the PWS. Variables were created for each rule within the Safe Drinking Water Act, such as the Lead and Copper Rule. A time period average for each rule name violation by PWS was calculated as the frequency divided by the number of years in the time period of interest, in this case five (2006-2010). This time period average was then multiplied by the population served for each PWS, and these values were summed for the county to estimate the proportion of the population in the county affected by the violation. Most counties did not report violations for the majority of rules; therefore, only one variable constructed provided sufficient variability to be included, which was that calculated from violations to the Total Coliform Rule.

Domestic Use

Data from the Estimated Use of Water in the United States database[14] were used as a proxy for domestic water quality. If water is being withdrawn for competing uses (agriculture, industry, etc.), it will put stress on water supplies, which, in turn, will affect water quality. This database includes county-level estimates of water withdrawals for domestic, agricultural, and industrial use. Initially, 15 variables of water withdrawals for domestic, agricultural, and industrial use were developed. These data are estimated every 5 years and were included in the EQI as averaged data for 2006 and 2010. Two variables were included in the EQI after evaluation for collinearity (four variables removed) and missing data (nine variables removed). The two variables were (1) the percent of population on self-supplied water supplies and (2) the percent of those on public water supplies that are on surface waters. For these variables, higher values are not necessarily a marker for poor water quality. The data were provided at the county level and normally distributed; therefore, no additional transformation was required.

Atmospheric Deposition

The atmospheric deposition of chemicals can affect water quality. The NADP dataset[13] provides measures for the concentration of nine chemicals in precipitation: (1) calcium, (2) magnesium, (3) potassium, (4) sodium, (5) ammonium, (6) nitrate, (7) chloride, (8) sulfate, and (9) mercury. Annual summary data from each monitoring site for each year 2006-2010 were kriged spatially to achieve national coverage and county-level estimates. The annual estimates for each pollutant then were averaged over the 5-year study period. The data for all pollutants, except sulfate, were skewed and, therefore, were natural log transformed to achieve normal distributions. Magnesium, sodium, and ammonium were removed as they were highly correlated with potassium, chloride, and nitrate, respectively.

Drought

Drought affects the concentration of pathogens and chemicals in water bodies and, therefore, can affect water quality. The Drought Monitor dataset[15] provides raster data on six possible drought status conditions for the entire United States on a weekly basis. The data were aggregated spatially to the county level to estimate the percentage of the county in each drought status condition. The weekly data were averaged to achieve annual estimates for 2006-2010 and, then, averaged to create a composite for the entire period. From this data, the percentage of the county in extreme or exceptional drought (intensity levels D3 and D4, respectively) was used in the EQI. The remaining five drought status conditions were removed because all of the drought statuses were highly correlated.

Chemical Contamination

Chemical contamination of water supplies can directly affect human health. The NCOD dataset[16] provides data on 69 contaminants provided by public water supplies throughout the country for the period from 1998-2005. Data for all samples in a county for each contaminant were averaged over the entire period of the dataset, 1998-2005. More recent data were not available. The data also were natural log transformed to achieve normal distributions. Missing values were set to zero, with the assumption that lack of measurement for an area indicated low concern for contamination with that particular contaminant. Nine contaminants, (1) asbestos, (2) beryllium, (3) diquat, (4) endothall, (5) glyphosate, (6) dioxin, (7) radium, (8) beta particles, and (9) uranium, did not include data for enough counties (missing data) to be included in the EQI construction. Twenty-one variables were deleted because of high correlation with other contaminants: (1) lindane, (2) thallium, (3) toxaphene, (4) oxamyl, (5) alachlor, (6) 2,4,5-TP (Silvex), (7) hexachlorocyclopentadiene, (8) carbofuran, (9) heptachlor, (10) heptachlor epoxide, (11) hexachlorobenzene, (12) 1,2,4-trichlorobenzene, (13) 1,2-dichlorobenzene, (14) vinyl chloride, (15) 1,1-dichloroethylene, (16) trans-1,2-dichloroethylene, (17) 1,2-dichloroethane, (18) carbon tetrachloride, (19) 1,2-dichloropropane, (20) 1,1,2-trichloroethane, (21) benzene.

Land Domain

The land domain consisted of five data sources, representing five constructs: (1) agriculture, (2) pesticide use, (3) facilities, (4) radon zone, and (5) mining activity.

Agriculture

Information on nonpesticide chemicals used in farming, animal units, harvested acreage, irrigated acreage, manure acreage, and proportion of farms was taken from the 2007 Census of Agriculture[19]. Final acreage for each item then was divided by total acreage for each county to return a percentage (e.g., percentage of irrigated acres out of total acres in a county). In some cases, county-level acreage for items was suppressed. In these, case estimates were imputed based on unaccounted for and total state-level acreage. Known acreage was subtracted from total state acreage, leaving an "unassigned" total acreage for each state. This total number was divided by the total number of farms in counties with suppressed acreage to return an average acreage for each farm. This average acreage then was multiplied by the number of farms in each county with suppressed acreage to estimate acreage. Animal units were estimated by multiplying the number of livestock (cows, hogs, and poultry) by the animals per animal unit statistic[46] and, then, adding together all livestock categories for each county. Eight variables representing agriculture were included in the EQI.

Pesticide Use

Pesticide use for each county was estimated using county-pesticide-use data from the 2009 National Pesticide Use Dataset[18]. Each pesticide was categorized into one of three categories: (1) herbicide, (2) fungicide, or (3) insecticide. The average weight (in kilograms) of each pesticide was calculated for the years available (2006-2009) for each county, then summed by pesticide type. If a county did not have information for one of the pesticide categories, the national average was used. Despite the choice of high spatial coverage, there are recognized uncertainties in estimating the geographic distribution of compounds applied to specific crops as described by Baker et al. (2015) in prior literature [47]. These three pesticide categories were included in the EQI. Pesticide variables were evaluated for normality and log transformed.

Facilities

Large facilities have the capacity to affect land quality. The facilities included in the land domain are those represented on the EPA Geospatial Data Download Service[20]. Because many counties had at least one, but no counties had all six of the facility types present, a composite facilities data variable was constructed by summing the count of any one of the six facilities types (Brownfield sites (n=1273)[48]; Superfund sites (n=719)[49]; Toxic Release Inventory sites (n=2671)[20]; pesticide-producing-location sites (n=2099)[50]; large-quantity generator sites (n=1963)[51]; and treatment, storage, and disposal sites (n=874)[52]) across the counties. Facilities were included in the count if they were identified during the 2006-2010 period. The count of facilities was divided by the county population, which produced a facilities rate. The facilities rate variable was assessed for normality and log transformed.

Radon Zone

The potential for elevated indoor radon levels was represented using the county score from the EPA Radon Zone map[21], which was available for 3142 counties (one county, Broomfield, Colorado, was missing). The EPA Radon Zone map identified areas of the United States with the potential for elevated indoor radon levels. Each United States county was assigned to one of three zones based on radon-level elevation potential.

Mines

Mines, like large facilities, have the capacity to affect land quality. The mines included in the land domain are those found in the MSHA dataset[22], which includes those mines under MSHA jurisdiction since 1970. Mines were included if they were active at any point before 2010 and were not abandoned and sealed after 2006. Those excluded most likely do not continue to pose any environmental impact. Any mines already represented in Superfund data were excluded. Mines were separated by the five primary commodity types: (1) coal, (2) metal, (3) nonmetal, (4) sand and gravel, and (5) stone, and a county could have more than one type of mine. The counts of the mines were divided by the county population, producing a mine rate. Of the 3143 counties, 2904 had at least one mine. For those counties that had zero values for the different mine types, zeros were replaced with the minimum value of the mine type/2 was added to the standardized population variables. The mine variables were assessed for normality and log transformed.

Sociodemographic Domain

This domain was constructed to explore the sociodemographic features of counties in the United States. These features were used to approximate the social stress associated with residing in more deprived (low education, high unemployment, high violent crime, high poverty, etc.) or more affluent (high employment rates, low property crime, high proportion of college graduates, etc.) counties. This domain includes variables from the 2010 United States Census)[24], the FBI Uniform Crime Reports (UCR)[25], the 2008 Presidential election results[26], and the United States Department of Agriculture Economic Research Service Creative Class data[27]. Because the sociodemographic domain is related to population density, by virtue of the data's collection and reporting, variables were developed as population rates (denominator: count of persons per county), rather than area-based rates (denominator: square miles per county).

Nine variables were obtained from the 2010 United States Census[24]. The nine variables were (1) percent earning a bachelor's degree or higher among persons aged 25 years or older; (2) percent persons unemployed; (3) percent of families living below the federal poverty line; (4) percent vacant housing units; (5) median household value; (6) median household income; (7) percent renter-occupied units; (8) count of occupants per room; and (9) the Gini coefficient, a marker of income inequality. Owing to the skewed nature of the household income and count of occupants per room data, these variables were log transformed for inclusion in the EQI. The sociodemographic domain contains a mix of positive and negative features; therefore, when the sociodemographic domain was constructed, positive variables were reverse-coded to ensure that a higher amount of the sociodemographic domain will represent adverse environmental conditions.

The area-level crime environment was represented using the FBI UCRs[25]. The first step in constructing crime data was to assign each jurisdiction or place to a county using county Federal Information Processing Standards code[53]. In cases when a jurisdiction covered more than one county, the reported crime was assigned to both counties. Although this double assignment results in a slight inflation of crime reports for a state, there was no way to determine which county should receive the crime report. Further, if police or municipal jurisdictions crossed county lines, it is likely residents of both counties were "exposed" to the crime environment. Crime data attributed to more than one county occurred in approximately 15 counties. Second, because crime was reported for less than half the United States counties, crime data were kriged spatially and temporally to estimate values for counties with no reported crime. The decision was made to krig these data because data reporting was voluntary, and it seemed unlikely that no crime occurred in the nonreported areas. Because zeros could not be reasonably assigned to the missing counties, the data were interpolated spatially and temporally instead. Based on experience with the 2000-2005 county-level EQI, and in acknowledgement that the correlation between the property and violent crime rates was very high (0.96), only log violent crime was included in the EQI.

The political climate of a county was represented by Leip's election map[26]. On this Web site, county-specific percents voting Republican or Democratic are reported. These data were downloaded for each county. The report voting Democratic in the 2008 presidential election are included in the EQI. One county in Hawaii that had been an independent county unit, FIPS 15005, was subsumed by Maui for the presidential election data, so the same Democratic percentage was applied to county 15005 as to Maui.

One creative class variable was included in the 2006-2010 EQI. The creative class thesis—that towns need to attract engineers, architects, artists, and people in other creative occupations to compete in today's economy—may be particularly relevant to rural communities, which tend to lose much of their talent when young adults leave. The ERS creative class codes[27] indicate a county's share of population employed in occupations that require "thinking creatively." The percent employed in creative class occupations index was included in the EQI.

Built Domain

Seven data sources were included in the built domain, representing (1) the subsidized housing environment, (2) traffic safety, (3) public transportation usage and commuting times, (4) road properties (road type and density), (5) the business and service environments (e.g., food, recreation), (6) county walkability, and (7) green space.

Housing Environment

The subsidized housing environment was represented by the Housing and Urban Development data[32]. These data provide a count of the low-rent and Section 8 housing in each housing authority data area. The housing authority areas correspond to cities, which were assigned county codes. Data were collected in 2010, but, because low-rent and Section 8 housing does not change substantially over time, these data were considered representative of the 2006-2010 period. The variables were summed to result in the count of any low-rent or Section 8 housing in each county. The rate of subsidized housing was constructed by dividing the count of subsidized housing units per county by the county population. The data were log transformed prior to inclusion in the EQI.

Traffic Safety

Traffic fatalities, an important feature and consequence of the built environment, were estimated using the Fatality Analysis Reporting System (FARS) data[31]. The FARS is a national census providing the National Highway Traffic Safety administration yearly reports of fatal injuries suffered in motor vehicle crashes. Rates for the 2006-2010 counts of fatal crashes per county were constructed by dividing the count of county-level fatal crashes by the county-level population. Many counties had no fatal crashes. To accommodate the large number of meaningful zeros in the data, the log of this rate variable was used in the built domain of the EQI.

Public Transportation Usage and Commuting Time

The percent of county residents who use public transportation was estimated using the 2010 United States Census[24] variable in the EQI. For many counties, the percent of the population that reports using public transportation is near zero. Therefore, this variable was log transformed prior to its use in the built domain of the EQI. Also obtained from the United States Census was the average number of minutes employed persons spent on the commute home from work.

Road Properties

For the built-environment domain, characterizing the relative proportions of each county that was served by highways, secondary roads, and primary roads were of interest, as these types of roads confer different risks (related to speed and safety) and benefits (related to neighborhood walking or ease of transit). Road type for the year 2008 was approximated using the NAVTEQ road data[30] associated to TIGER county boundary [29] data. Three proportion variables were constructed by dividing the mileage of each road type (e.g., secondary roads) by the total road mileage in each county. The proportions of all roadways that were secondary roads were included.

Business and Service Environments

Businesses represent an important component of the built environment and can contribute to the risk and amenity landscape. Variables representing various built-environmental features were constructed using the proprietary 2008 Dun and Bradstreet data[28], which include commercial information on businesses and data on more than 195 million records. Eight rate variables were constructed by dividing the county-level count of a business type by the county-level population count. The eight variables included the (1) positive food environment, (2) negative food environment, (3) vice environment (alcohol, pawn, and gaming), (4) health care business environment, (5) recreation environment, (6) education environment, (7) social-service environment, and (8) civic-related environment. Note: Positive food environments included those that sold healthier foods, like grocery stores, sit-down restaurants, and organic shops, whereas the negative food environment included businesses like fast-food restaurants, convenience stores, and pretzel trucks. Although related, these two food environments comprise different businesses and are not 100% inversely correlated. Nonnormally distributed variables were log transformed, and all eight were included in the EQI.

Walkability

Walkability is an important feature of the built environment, and variability across walkability may help explain poor or good health. The National Walkability Index (NWI)[36] was used to determine walkability as a mode of travel for each county. The scores, ranging from zero to 20 are calculated using a weighted rank of four variables: (1) mix of employment types (such as office, retail, and service) and occupied housing, (2) mix of employment types in a block group (such as office, retail, and service), (3) street intersection density (pedestrian-oriented intersections), and (4) predicted commute mode split – proportion of workers in the block group who carpool. A higher rank indicates an increased likelihood of walking being used as the mode of travel. The block group scores were added, and, then, a mean of the block group scores based on county population proportions was created. The county walkability scores ranged from 1.00 to 16.23.

Green Space

Exposure to green space also has been associated with improved health. The green space variable was created by EPA's EnviroAtlas[33] using National Land Cover Database (NLCD) [34] and Coastal Change Analysis Program[35] data. Three possible constructions were considered: The NINDEX variable was created by EnviroAtlas as a natural land cover variable and includes barren land, forest, shrub/scrub, grassland, sedge, lichens, moss, and wetlands. NINDEX_open is the NINDEX variable with developed open space, such as parks and golf courses, included. The Richardson index[54] is based on a green space paper and includes the NINDEX and also developed open space, low intensity, and medium intensity. For the sake of dissemination outside academic communities and ease of data availability/construction, the 2006-2010 EQI used the NINDEX_open variable. The variables represented percentages of up to 24 possible land cover types. To create a green space variable, five total land cover groups were combined, those classified as (1) natural land cover (barren land, rock/sand/clay/tundra/perennial ice), (2) forest, (3) shrubland/scrub land, (4) herbaceous, and (5) wetlands) and those classified as developed open space, where impervious surfaces make up less than 20% of total cover and includes recreational areas, such as grassy lawns, parks, and golf courses. This combined variable of natural land cover and developed open space gave a percentage of the county that had green space and ranged from 3.88% to 99.99%. The variable then was assessed for normality.

Changes to 2006-2010 variable construction from original 2000-2005 EQI

Air Domain

Variables eliminated from the 2006-2010 EQI

- The following air variables were eliminated because of high collinearity to one or more variables.

• **Variable**	**Represented by**
• 2-4-toluene diisocyanate	Ethylbenzene
• 2-chloroacetophenone	Benzyl chloride
• 2-nitropropane	Chloroprene
• 4-nitrophenol	Ethylbenzene
• Acetophenone	Ethylbenzene
• Acrolein	Ethylbenzene
• Acrylonitrile	Chloroprene
• Biphenyl	Ethylbenzene
• Bromoform	Benzyl chloride
• Cadmium compounds	Chromium compounds
• Carbon disulfide	Ethylbenzene
• Cresol cresylic acid	Ethylbenzene
• Cumene	Ethylbenzene
• Diesel engine emissions	Ethylbenzene
• Dimethyl formamide	Ethyl chloride
• Dimethyl phthalate	Ethylbenzene
• Dimethyl sulfate	Benzyl chloride
• Epichlorohydrin	Chloroprene
• Ethyl acrylate	Chloroprene
• Ethylene glycol	Ethylbenzene
• Ethylene oxide	Ethylene dichloride
• Ethylidene dichloride	Vinyl chloride
• Hexachlorobenzene	Polychlorinated biphenyls
• Hexachlorobutadiene	Chloroprene
• Hexachlorocyclopentadiene	Chloroprene
• Hexane	Ethylbenzene
• Lead compounds	Chromium compounds
• Mercury compounds	Ethylbenzene
• Methanol	Ethylbenzene
• Methyl chloride	Carbon tetrachloride
• Methyl isobutyl ketone	Ethylbenzene
• Methyl methacrylate	Ethylbenzene
• Methylhydrazine	Benzyl chloride
• MTBE	Ethylbenzene
• Nitrobenzene	Chloroprene
• n-n-dimethylaniline	Chloroprene
• o-toluidine	Chloroprene
• PAH/POM	Ethylbenzene
• Propylene oxide	Chloroprene
• Selenium compounds	Ethylbenzene
• Styrene	Ethylbenzene
• Tetrachloroethylene	Ethylbenzene
• Toluene	Ethylbenzene
• Triethylamine	Ethylbenzene
• Vinyl acetate	Ethylbenzene
• Vinylidene chloride	Ethylbenzene

Water Domain

New variables added to the 2006-2010 EQI

- Total coliform health-based violations added

Variables removed in the recreational water construct

- Number of days closed per event in county 2000-2005
  numDays_Close_Activity_tot
- Number of days per contamination advisory event in county 2000-2005
  numDays_Cont_Activity_tot
- Number of days per rain advisory event in county 2000-2005
  numDays_Rain_Activity_tot

Variables removed in the chemical contamination construct from the 2006-2010 EQI because of correlation with other variables

- Beryllium - W_Be_ln (mg/L)
- Lindane - W_Lindane_ln (mg/L)
- Thallium - W_Tl_ln (mg/L) 1996
- Toxaphene - W_Toxaphene_ln (µg/L)
- Oxamyl (Vydate) - W_Oxamyl_ln (µg/L)
- Alachlor - W_Alachlor_ln (µg/L)
- 2,4,5-TP (Silvex) - W_silvex_ln (µg/L)
- Hexachlorocyclopentadiene - W_HCCPD_ln (µg/L)
- Carbofuran - W_Carbofuran_ln (µg/L)
- Heptachlor - W_Heptachlor_ln (µg/L)
- Heptachlor Epoxide - W_Heptachlor_epox_ln (µg/L)
- Hexachlorobenzene - W_HCB_ln (µg/L)
- 1,2,4-Trichlorobenzene - W_124TCIB_ln (µg/L)
- 1,2-Dichlorobenzene (o-Dichlorobenzene) - W_ODCB_ln (µg/L)
- Vinyl Chloride - W_VCM_ln (µg/L)
- 1,1-Dichloroethylene - W_11DCE_ln (µg/L)
- trans-1,2-Dichloroethylene - W_t12DCE_ln (µg/L)
- 1,2-Dichloroethane (Ethylene Dichloride) - W_EDC_ln (µg/L)
- Carbon Tetrachloride - W_CCl4_ln (µg/L)
- 1,2-Dichloropropane - W_PDC_ln (µg/L)
- 1,1,2-Trichloroethane - W_112TCA_ln (µg/L)
- Benzene - W_Cl1benz_ln (µg/L)

Land Domain

Variables eliminated from the 2006-2010 EQI

- The following variables were eliminated because content was represented in the NCOD and NADP.
- Mean level of arsenic
- Mean level of selenium
- Mean level of mercury
- Mean level of lead
- Mean level of zinc
- Mean level of copper
- Mean level of aluminum
- Mean level of sodium
- Mean level of magnesium
- Mean level of phosphourous
- Mean level of titanium
- Mean level of calcium
- Mean level of iron

New variables added to the 2006-2010 EQI

- Primarily coal mines per county population
- Primarily metal mines per county population
- Primarily nonmetal mines per county population
- Primarily sand and gravel mines per county population
- Primarily stone mines per county population

Sociodemographic Domain

Variables eliminated from the 2006-2010 EQI

- Percent management occupation – eliminated because content better covered in creative class index data
- Housing built before 1939 – eliminated because of unclear association with health
- Percent with no English – eliminated because of unclear association with health and increasing subjectivity

Variables substitutions for the 2006-2010 EQI

- Percent bachelor's degree (>25 years old) substituted for percent greater than high school
- Percent family poverty substituted for percent persons in poverty
- Count of occupants per room replaced median number of rooms

New variables added to the 2006-2010 EQI

- Percent of persons working in creative occupations
- Percent of county that voted Democratic in the 2008 presidential election
- Built domain

Variables eliminated from the 2006-2010 EQI

- Entertainment environment – eliminated because of unclear association with health
- Transportation environment – because the data contained in this variable is better covered using other data sources

Variables substitutions for the 2006-2010 EQI

- Percent secondary roads replaced percent primary roads

New variables added to the 2006-2010 EQI

- Walkability score added
- Proportion of county in green space added

Data Reduction and Index Construction

Overall Approach

- After variable development, all the variables were combined into an index representing the overall environmental quality. The specific tasks required for index construction were as follows.
- Included all the variables from one domain in a PCA to empirically summarize that domain-specific environmental context (retaining the first component as the domain index) for each of the five domains
- Assessed the positive/negative direction (valence) of the variable loadings for each domain; if loadings were not in the correct direction to ensure a higher value on the index corresponded to worse environmental quality, corrected valence when necessary
- Combined each of the five domain-specific indices in another PCA to empirically summarize the overall environmental context into one index of environmental quality and retained the initial component as the overall EQI
- Repeated the three previous steps for each of the four RUCC strata (e.g., RUCC stratum 1 air domain; RUCC stratum 2 air domain, etc.), such that each RUCC had its own set of domain-specific indices, as well as its own overall index

The EQI, domain-specific indices, and EQI stratified by rural-urban data are available publicly at EPA's Environmental Dataset Gateway. Also, an interactive map of the EQI is available at EPA's GeoPlatform.

Principal Components Analysis (PCA)

PCA is a data reduction technique frequently used to create sociodemographic scales or indices for inclusion in statistical models[43, 55]. PCA analyzes total variance, and the loading represents the correlation between the variable and the component. PCA assumes no underlying latent variable structure but, rather, seeks to empirically summarize multiple possible domains. Three major goals of PCA are to

1. summarize the patterns of correlations among observed or measured variables,
2. provide an operational definition—in this case, a regression equation—for underlying processes by using observed or measured variables, and
3. reduce a large number of observed variables into a smaller number of factors or a single component.

PCA was chosen for data reduction for several reasons. Production of an empirical summary of the various constituent components of the EQI was desired. Various data sources measured on multiple scales needed to be combined. PCA standardizes these measures prior to combining. Therefore, the differing scales are less problematic. To assess variables influences on the index, variables cannot simply be added together. To do so would mean knowledge for most of the variables would not be available to indicate if any one variable would prove to be more "influential" for environmental quality than another. PCA enables variable loadings to vary by their relative importance to the total component. This feature enabled exploration of variable loading differences for interpretation purposes.

The PCA steps included

- selecting the set of variables to be used,
- preparing the correlation matrices,
- extracting the set of components from the correlation matrix,
- determining the number of components observed, and
- interpreting the findings.

The sole modification to the PCA methodology in the county 2006-2010 EQI compared to that of the 2000-2005 EQI is "valence correction." We also have created a 2000-2005 valence-corrected version of the EQI.

"Valence correction" refers to reorientation of PCA output for uniformity of interpretation of domain indices and uniformity in orientation of domain indices input into the second PCA for EQI construction. In this instance, we are defining valence as the departure from neutrality along a continuum; generally, we are interested how attributes depart from neutrality in opposite directions. The PCA loadings are a function of the program's starting point, or seed, which is not easily manipulable. Therefore, the loading valence needed to be corrected prior to the construction of the indices to ensure that higher values on a given index, and on the overall EQI, signify worse environmental quality [56, 57].

Domain and EQI indices are designed such that lower (more negative) values represent "better" quality and higher (more positive) values represent "worse" quality. Under this setup, health beneficial variables should load negative in the PCA output ("+" or "–" loading sign for a variable in the component variable loadings vector represents positive or negative correlation between that variable and the component, respectively). Given that the first principal component was taken to represent domain or environmental quality and that the orientation of these indices was designated as going from better to worse quality (negative to positive index value), it was necessary to reverse the component variable loadings vector from a PCA output if a high proportion of variables was deemed beneficial loaded "+", and a high proportion of variables was deemed detrimental loaded "–"[55]. Determination of variables as beneficial or detrimental to human health across domains was done a priori based on literature evidence and content matter judgment. Reorientation of PCA-derived indices through multiplication of the component variables loading vector by –1 preserves (1) the direction of the relationship among the variables for a given PCA (i.e., variables that loaded with same signs will retain same signs, and variables that loaded opposite to each other will retain opposite signs after reversal, and, therefore, the pattern of correlations among the variables will remain intact); and (2) the magnitude of correlation among variables (reversal of loading signs does not impact the magnitude of the loading)[58]. The sum of squares of variable loadings in a PCA output equals 1, and, therefore, each square of a variable loading can be viewed as a measure of the contribution of that variable toward the principal component (domain indices and EQI in this case), enabling estimation of the "correctness" of the orientation of the index. We used the square of variable loadings in a given PCA output in combination with aforementioned a priori designations of benefit or harm to guide choice of index reorientations.

PCA analyzes the total variance. Therefore, in the PCA correlation matrix, "1" is in the positive diagonal[55]. To construct the EQI, variables from each domain were entered into domain-specific PCAs. PCA produced variable loadings, which were roughly equivalent to the "weight" or contribution that each variable made toward explaining the total variance. The weights, however, need not sum to 1.0 because the loadings were for the total variance, rather than just the shared variance. The loading associated with each variable then was multiplied by its mean value for the given geography (county, for the EQI), and these weighted mean values were summed.

Rural-Urban Continuum

Both the domain-specific indices and the overall EQI were created for each county in the United States. Recognizing that environments differ dramatically across the rural-urban continuum[59], the decision was made that the EQI would be most useful if it accommodated rural-urban environmental differences. The EQI was stratified by RUCCs. The RUCC is a nine-item categorization code of proximity to or influence of major metropolitan areas[60]. The nine-item categories were condensed into four, where RUCC1 represents metropolitan-urbanized = codes 1+2+3, RUCC2 nonmetropolitan-urbanized = 4+5, RUCC3 less urbanized = 6+7, and RUCC4 thinly populated (rural) = 8+9 (see Figure 3)[61-64]. For the 2006-2010 EQI, the 2013 RUCC was used. RUCC-stratified EQIs, and an overall EQI was constructed. Loadings on the stratified and nonstratified sets of indices were assessed to determine loading heterogeneity across counties. Because these loadings differed meaningfully by RUCC level, RUCC-stratified EQIs were constructed for each county.

Although it was possible to form as many independent linear combinations as there were variables in PCA, only the first principal component was retained. The first principal component was the unique linear combination that accounted for the largest possible proportion of the total variability in the component measures. Therefore, the first component from each of these domain-specific indices was retained (e.g., air index, water index). Domain-specific indices were then entered into another PCA, where the first component was retained as the EQI (Figure 2). This process was undertaken separately for each of the four RUCC strata.

Within each RUCC strata, domain-specific variable loadings were evaluated based on the value of variable loading and the variable's hypothesized relevance to health. For instance, although arsenic may occur in low frequency in a lot of counties and, therefore, may have a relatively small component loading, it is an important health hazard when present. Based on variable loading magnitude alone, dropping arsenic from an EQI may be a reasonable conclusion. However, it was retained for the EQI based on its relevance to human health.

The first principal component, the domain-specific EQI (e.g., air domain EQI), then was standardized to have a mean of 0 and standard deviation (SD) of 1 by dividing the index by the square of its eigenvalue. Each domain-specific index was then included in a second PCA procedure (Figure 2) to result in the overall EQI for each stratum of RUCC.

For orientation to the results, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) indicate lower environmental quality.

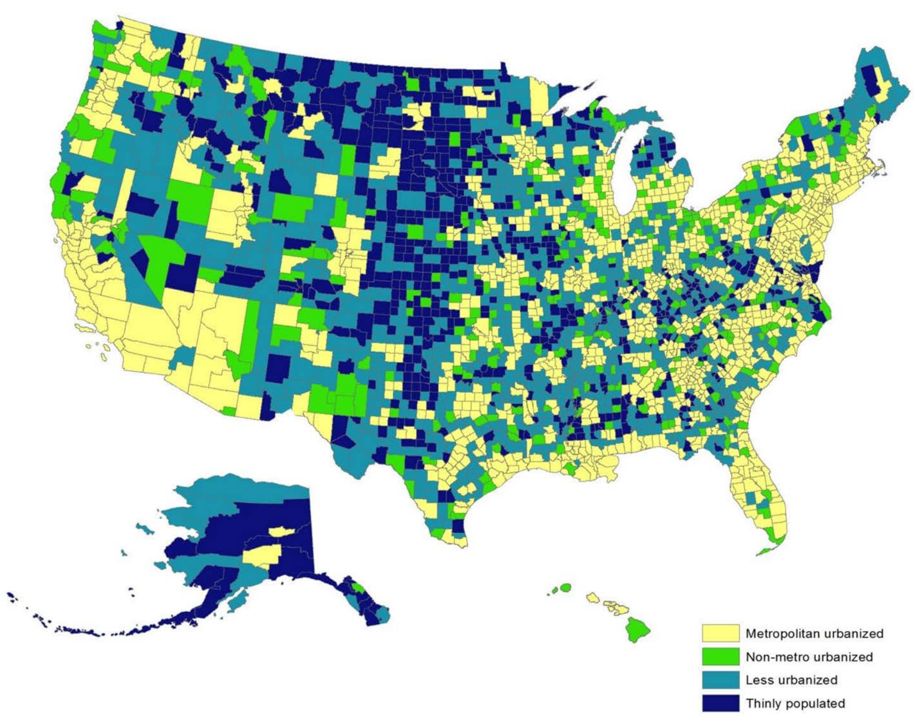


Figure 3. Rural-urban continuum code (RUCC) stratification for all counties in the United States.

Results

Description of Variables Comprising Environmental Quality Index Domains

Air Domain

Criteria air pollutants were distributed relatively evenly across the rural-urban gradient (Table 4). Some hazardous air pollutants varied in emissions across rural-urban strata; however, there was no discernable pattern for most. For example, 1,1,2-trichloroethane's highest levels were observed in the less urbanized stratum, whereas levels were similar across other strata, and emissions for manganese compounds were highest in the most metropolitan areas then steadily decreased across more rural strata.

Table 4. Air domain variable means, standard deviations (SDs), and ranges - Overall and rural-urban continuum codes (RUCCs) stratified

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	] Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range	Total (3143) Mean (SD) [Range]
Construct: Criteria Air Pollutants						
PM10	µg/m3	2.0E+01 (4.7E+00) [4.1E-01, 5.4E+01]	1.95E+01 (5.07E+00) [6.00E+00, 6.60e+01]	1.95E01 (4.37E+00) [5.39E+00, 5.25E+01]	1.89E+01 (4.88E+00) [4.01E-01, 3.42E+01]	2.0E+01 (4.7E+00) [4.0E-01, 6.6E+01]
PM2.5	µg/m3	1.1E+01 (2.1E+00) [4.1E+00, 2.4E+01]	1.02E+01 (2.19E+00) [4.28E+00, 1.48E+01]	9.99E+00 (2.20E+00) [3.35E+00, 1.80E+01]	9.05E+00 (2.39E+00) [4.28E+00, 1.79E+01]	1.0E+01 (2.3E+00) [3.3E+00, 2.4E+01]
Ozone	ppm	4.5E-02 (4.4E-03) [2.2E-02, 5.9E-02]	4.46E-02 (4.99E-03) [2.22E-02, 5.76E-02]	4.47E+02 (3.99E+03) [2.99E-02, 5.72E-02]	4.46E-02 (4.47E-03) [2.90E-02, 5.65E-02]	4.5E-02 (4.4E-03) [2.2E-02, 5.9E-02]
Nitrogen oxide	ppb	9.2E+00 (4.6E+00) [5.9E-01, 3.1E+01]	7.93E+00 (3.93E+00) [5.92E-01, 2.81E+01]	3.85E-01 (8.36E-02) [2.41E-01, 8.89E-01]	6.65E+00 (4.37E+00) [5.91E-01, 2.84E+01]	8.0E+00 (4.4E+00) [2.6E-01, 3.1E+01]
Sulfur dioxide	ppb	2.2E+00 (1.5E+00) [7.3E-03, 9.7E+00]	1.97E+00 (2.22E+00) [1.10E-02, 3.09E+01]	7.53E+00 (4.00E+00) [2.65E-01, 2.84E-01]	1.47E+00 (1.39E+00) [2.21E-02, 9.23E+00]	1.9E+00 (1.5E+00) [7.3E-03, 3.1E+01]
Carbon monoxide	ppm	3.9E-01 (8.2E-02) [2.5E-01, 8.7E-01]	3.87E-01 (7.49E-02) [2.49E-01, 7.38E-01]	4.32E-03 (4.91E-04) [3.90E-03, 8.19E-03]	3.93E-01 (9.57E-02) [2.61E-01, 8.90E-01]	3.9E-01 (8.5E-02) [2.4E-01, 8.9E-01]
Ethylene dibromide	Tons emitted	5.5E-04 (3.1E-04) [5.5E-05, 2.0E03]	5.47E-04 (3.47E-04) [1.65E-04, 1.64E-03]	5.50E-04 (3.14E-04) [1.65E-04, 1.79E-03]	4.77E-04 (2.75E-04) [5.50E-05, 1.68E-03]	5.4E-04 (3.1E-04) [5.5E-05, 2.0E-03]
Formaldehyde	Tons emitted	1.9E+00 (6.0E-01) [2.1E-01, 5.6E+00]	1.75E+00 (5.57E-01) [6.83E-01, 3.20E+00]	1.79E+00 (5.80E-01) [6.25E-01, 3.86E+00]	1.61E+00 (6.05E-01) [2.08E-01, 3.36E+00]	1.8E+00 (6.0E-01) [2.1E-01, 5.6E+00]
1,1,2,2-Tetrachloroethane	Tons emitted	4.4E-03 (7.5E-04) [1.3E-03, 1.4E-02]	4.46E-03 (9.07E-04) [3.90E-03, 1.33E-02]	1.39E-04 (2.79E-03) [1.76E-13, 8.10E-02]	4.20E-03 (6.61E-04) [1.30E-03, 1.60E-02]	4.4E-03 (6.7E-04) [1.3E-03, 1.6E-02]
1,1,2-Trichloroethane	Tons emitted	4.0E-04 (6.6E-03) [1.8E-13, 2.1E-01]	2.00E-05 (1.24E-04) [1.76E-13, 1.73E-03]	5.25E-06 (9.53E-06) [1.95E-06, 1.87E-04]	9.61E-05 (1.58E-03) [1.76E-03, 3.59E-02]	2.1E-04 (4.4E-03) [1.8E-13, 2.1E-01]
1,2-Dibromo-3-chloropropane	Tons emitted	5.2E-06 (7.3E-06) [6.5E-07, 9.1E-05]	5.98E-06 (2.29E-05) [1.95E-06, 3.52E-04]	8.41E-03 (2.26E-02) [5.00E-16, 3.75E-01]	4.34E-06 (6.27E-06) [6.50E-07, 6.60E-05]	5.1E-06 (1.0E-05) [6.5E-07, 3.5E-04]
1,2-Dichloropropane	Tons emitted	1.1E-02 (3.4E-02) [5.0E-16, 4.9E-01]	1.06E-02 (2.13E-02) [5.00E-16, 1.40E-1]	6.41E-05 (5.31E-04) [3.00E-015, 1.01E-02]	5.00E-03 (1.38E-02) [5.00E-016, 1.18E-01]	9.1E-03 (2.6E-02) [5.0E-16, 4.9E-01]
Acrylic acid	Tons emitted	1.4E-04 (2.4E-03) [3.0E-15, 7.2E-02]	2.06E-04 (2.45E-03) [3.00E-15, 4.23E-02]	3.43E-07 (7.89E-07) [1.46E-08, 7.29E-06]	9.76E-05 (1.39E-03) [3.00E-15, 3.36E-02]	1.1E-04 (1.8E-03) [3.0E-15, 7.2E-02]
Benzidine	Tons emitted	3.3E-07 (1.2E-06) [4.9E-09, 3.6E-05]	3.22E-07 (1.98E-06) [1.48E-08, 3.39E-05]	1.26E-05 (2.92E-05) [4.69E-12, 3.90E-04]	3.14E-07 (1.60E-06) [4.88E-09, 3.72E-05]	3.3E-07 (1.3E-06) [4.9E-09, 3.7E-05]
Benzyl chloride	Tons emitted	1.4E-05 (3.9E-05) [4.7E-12, 8.5E-04]	1.40E-05 (4.08E-05) [4.69E-12, 4.20E-04]	1.26E-05 (2.92E-05) [4.69E-12, 3.90E-04]	1.10E-05 (4.97E-05) [4.69E-12, 1.16E-03]	1.3E-05 (3.9E-05) [4.7E-12, 1.2E-03]
Beryllium compounds	Tons emitted	4.4E-05 (4.4E-05) [7.5E-06, 7.7E-04]	4.55E-05 (6.00E-05) [2.25E-05, 6.93E-04]	4.66E-05 (8.23E-05) [2.25E-05, 1.56E-03]	3.57E-05 (2.93E-05) [7.50E-06, 6.26E-04]	4.3E-05 (5.9E-05) [7.5E-06, 1.6E-03]
bis-2-Ethylhexyl phthalate	Tons emitted	8.4E-03 (1.9E-03) [2.6E-03, 6.3E-02]	8.22E-03 (5.39E-04) [7.80E-03, 1.30E-02]	8.31E-03 (1.77E-03) [7.80E-03, 4.36E-02]	8.08E-03 (6.40E-04) [2.60E-03, 1.22E-02]	8.3E-03 (1.6E-03) [2.6E-03, 6.3E-02]
Carbon tetrachloride	Tons emitted	9.1E-01 (1.8E-02) [3.0E-01, 9.2E-01]	9.11E-01 (3.75E-04) [9.11E-01, 9.15E-01]	9.11E-01 (9.67E-04) [9.03E-01, 9.28E-01]	9.06E-01 (5.36E-02) [3.01E-01, 9.27E-01]	9.1E-01 (2.7E-02) [3.0E-01, 9.3E-01]
Carbonyl sulfide	Tons emitted	1.8E-03 (1.1E-02) [5.0E-16, 1.6E-01]	5.14E-03 (7.25E-02) [5.00E-16, 1.27E+00]	9.25E-04 (4.94E-03) [5.00E-16, 7.78E-02]	2.13E-03 (2.26E-02) [5.00E-16, 4.39E-01]	1.9E-03 (2.6E-02) [5.0E-16, 1.35E+00]
Chlorine	Tons emitted	2.4E-03 (1.9E-02) [3.4E-13, 5.6E-01]	3.25E-03 (2.48E-02) [3.41E-13, 3.58E-01]	1.57E-03 (9.72E-03) [3.41E-13, 1.76E-01]	1.34E-03 (8.28E-03) [3.41E-13, 1.13E-01]	2.0E-03 (1.6E-02) [3.4E-13, 5.6E-01]
Chlorobenzene	Tons emitted	4.2E-03 (1.5E-02) [3.4E-11, 2.3E-01]	3.40E-03 (1.17E-02) [2.77E-07, 1.63E-01]	2.73E-03 (9.33E-03) [1.01E-10, 1.74E-01]	1.60E-03 (5.08E-03) [3.36E-11, 5.42E-02]	3.1E-03 (1.1E-02) [3.4E-11, 2.3E-01]

Table 4. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	] Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range	Total (3143) Mean (SD) [Range]
Chloroform	Tons emitted	1.0E-01 (2.6E-02) [3.0E-02, 6.6E-01]	9.77E-02 (1.61E-02) [8.85E-02, 2.02E-01]	9.58E-02 (1.41E-02) [8.85E-02, 2.26E-01]	9.36E-02 (1.31E-02) [2.95E-02, 2.11E-01]	9.7E-02 (2.0E-02) [3.0E-02, 6.6E-01]
Chloroprene	Tons emitted	1.9E-04 (3.1E-03) [1.6E-013, 8.8E-02]	1.06E-03 (1.81E-02) [1.57E-13, 3.17E-01]	2.05E-04 (5.31E-03) [1.57E-13, 1.69E-01]	2.68E-05 (3.84E-04) [1.57E-13, 7.24E-03]	2.4E-04 (6.7E-03) [1.6E-13, 3.2E-01]
Chromium compounds	Tons emitted	4.1E-04 (7.0E-04) [2.1E-05, 6.6E-03]	3.44E-04 (6.25E-04) [6.15E-05, 5.63E-03]	3.28E-04 (7.70E-04) [6.15E-05, 1.04E-02]	2.18E-04 (4.00E-04) [2.05E-05, 6.24E-03]	3.4E-04 (6.5E-04) [2.1E-05, 1.0E-02]
Cobalt compounds	Tons emitted	3.9E-05 (3.5E-04) [2.2E-14, 8.5E-03]	2.66E-05 (1.12E-04) [2.20E-14, 1.66E-03]	2.91E-05 (2.56E-04) [2.20E-014, 6.95E-03]	3.80E-05 (2.92E-04) [2.20E-14, 4.67E-03]	3.5E-05 (2.9E-04) [2.2E-14, 8.5E-03]
Cyanide compounds	Tons emitted	2.5E-02 (6.1E-02) [8.1E-14, 1.4E+00]	2.50E-02 (5.74E-02) [8.10E-14, 8.76E-01]	1.76E-02 (2.15E-02) [8.10E-014, 2.54E-01]	1.49E-2 (3.50E-02) [8.10E-14, 8.00E-01]	2.1E-02 (4.6E-02) [8.1E-14, 1.4E+00]
Dibutylphthalate	Tons emitted	3.5E-03 (5.3E-02) [1.3E-09, 1.7E+00]	5.63E-03 (2.92E-02) [3.81E-08, 4.02E-01]	2.21E-03 (1.38E-02) [7.18E-09, 2.19E-01]	1.76E-03 (2.94E-02) [1.30E-09, 7.40E-01]	2.9E-03 (3.7E-02) [1.3E-09, 1.7E+00]
Ethyl chloride	Tons emitted	1.8E-03 (1.5E-02) [7.6E-09, 5.1E-01]	1.18E-03 (1.67E-03) [4.97E-08, 1.31E-02]	1.42E-03 (9.95E-03) [7.59E-09, 2.34E-01]	8.36E-04 (1.88E-03) [7.59E-09, 2.93E-02]	1.4E-03 (1.1E-02) [7.6E-09, 5.5E-01]
Ethyl benzene	Tons emitted	7.7E-02 (1.2E-01) [3.5E-05, 1.9E+00]	6.56E-02 (8.41E-02) [1.78E-04, 5.41E-01]	5.88E-02 (8.87E-02) [2.49E-04, 8.86E-01]	4.86E-02 (8.28E-02) [3.46E-05, 8.46E-01]	6.4E-02 (1.0E-01) [3.5E-05, 1.9E+00]
Ethyl dichloride	Tons emitted	4.2E-03 (2.5E-03) [9.0E-04, 3.9E-02]	4.17E-03 (3.10E-03) [2.70E-03, 3.04E-02]	4.30E-03 (4.07E-03) [2.70E-03, 7.73E-02]	3.89E-03 (4.38E-03) [9.00E-04, 9.84E-02]	4.2E-03 (3.6E-03) [9.0E-04, 9.8E-02]
Glycol ethers	Tons emitted	3.4E-03 (1.4E-02) [1.8E-11, 2.5E-01]	2.68E-03 (8.45E-03) [1.83E-11, 7.92E-02]	3.59E-03 (1.55E-02) [1.83E-11, 2.66E-01]	2.63E-03 (1.35E-02) [1.83E-11, 2.43E-01]	3.2E-03 (1.4E-02) [1.8E-11, 2.7E-01]
Hydrazine	Tons emitted	4.2E-06 (1.4E-05) [6.5E-08, 1.4E-04]	4.60E-06 (1.46E-05) [1.95E-07, 1.21E-04]	3.27E-06 (1.25E-05) [1.95E-07, 1.83E-04]	3.34E-06 (1.67E-05) [6.50E-08, 2.80E-04]	3.8E-06 (1.4E-05) [6.5E-08, 2.8E-04]
Hydrochloric acid	Tons emitted	4.7E-01 (1.9E+00) [3.7E-06, 2.5E+01]	2.08E-01 (1.04E+00) [7.72E-05, 1.16E+01]	2.80E-01 (1.30E+00) [1.11E-05, 2.52E+01]	1.96E-01 (1.09E+00) [3.69E-06, 2.15E+01]	3.3E-01 (1.5E+00) [3.7E-06, 2.5E+01]
Isophorone	Tons emitted	1.1E-04 (9.4E-04) [5.4E-14, 3.1E-02]	1.31E-04 (8.65E-04) [5.40E-14, 1.46E-02]	9.79E-05 (6.31E-04) [5.40E-14, 1.71E-02]	4.55E-05 (1.63E-04) [5.40E-14, 2.45E-03]	9.4E-05 (7.3E-04) [5.4E-14, 3.1E-02]
Manganese compounds	Tons emitted	2.4E-03 (1.8E-02) [2.9E-04, 5.6E-01]	2.21E-03 (1.19E-02) [8.70E-04, 2.03E-01]	1.58E-03 (3.79E-03) [8.70E-04, 9.02E-02]	1.49E-03 (3.39E-03) [2.90E-04, 6.50E-02]	1.9E-03 (1.2E-02) [2.9E-04, 5.6E-01]
Methyl bromide	Tons emitted	6.8E-02 (5.2E-02) [1.8E-02, 7.5E-01]	6.38E-02 (3.00E-02) [5.25E-02, 2.90E-01]	6.19E-02 (3.21E-02) [5.25E-02, 5.94E-01]	5.77E-02 (1.66E-02) [1.75E-02, 2.22E-01]	6.3E-02 (3.8E-02) [1.8E-02, 7.5E-01]
Methyl chloride	Tons emitted	2.4E-01 (1.9E-01) [5.5E-02, 4.7E+00]	2.31E-01 (1.29E-01) [1.65E-01, 1.64E+00]	2.13E-01 (8.85E-02) [1.65E-01, 1.04E+00]	1.96E-01 (6.98E-02) [5.50E-02, 1.01E+00]	2.2E-01 (1.4E-01) [5.5E-02, 4.7E+00]
Phosphine	Tons emitted	3.8E-05 (7.5E-05) [2.6E-13, 8.3E-04]	3.72E-05 (6.85E-05) [2.64E-13, 4.70E-04]	4.20E-05 (8.84E-05) [2.64E-13, 1.64E-03]	4.33E-05 (1.23E-04) [2.64E-13, 2.59E-03]	4.0E-05 (9.1E-05) [2.6E-13, 2.6E-03]
Polychlorinated biphenyls	Tons emitted	3.8E-05 (1.1E-04) [2.1E-13, 3.7E-03]	3.66E-05 (3.78E-05) [2/06E-013, 2.99E-04]	3.14E-05 (3.47E-05) [2.06E-013, 4.21E-04]	2.87E-05 (3.70E-05) [2.06E-13, 4.88E-04]	3.4E-05 (7.4E-05) [2.1E-13, 3.7E-03]
Propylene dichloride	Tons emitted	1.6E-03 (2.2E-03) [2.3E-04, 4.5E-02]	1.21E-03 (1.06E-03) [6.90E-04, 7.98E-03]	1.03E-03 (8.81E-04) [6.90E-04, 8.60E-03]	9.74E-04 (8.25E-04) [2.30E-04, 7.00E-03]	1.3E-03 (1.6E-03) [2.3E-04, 4.5E-02]
Quinoline	Tons emitted	1.4E-04 (2.7E-04) [4.4E-07, 1.7E-03]	1.51E-03 (3.27E-04) [1.32E-06, 2.06E-03]	1.05E-04 (2.59E-04) [1.32E-06, 1.89E-03]	5.10E-05 (1.49E-04) [4.40E-07, 1.25E-03]	1.1E-04 (2.5E-04) [4.4E-07, 2.1E-03]
Trichloroethylene	Tons emitted	5.2E-02 (4.9E-02) [2.5E-03, 7.6E-01]	4.69E-02 (4.06E-02) [7.50E-03, 2.21E-01]	4.45E-02 (4.13E-02) [7.50E-03, 2.84E-01]	3.48E-02 (4.08E-02) [2.50E-03, 4.36E-01]	4.5E-02 (3.1E-03) [2.8E-10, 7.0E-2]
Vinyl chloride	Tons emitted	7.8E-04 (3.8E-03) [2.8E-10, 7.0E-02]	5.35E-04 (1.87E-03) [2.84E-10, 2.35E-02]	6.01E-04 (2.89E-03) [2.84E-10, 5.59E-02]	4.55E-04 (2.64E-03) [2.84E-10, 4.77E-02]	6.3E-04 (1.5E+00) [7.3E-03, 3.1E+01]

Water Domain

Variables included in the water domain (Table 5) suggest that urban counties were more likely to have impaired stream length (20%) compared with rural counties (9%). Additionally, urban counties had higher mercury deposition, chloride precipitation, sulfate precipitation, and the percentage of the county in drought status. Chemical contamination varied by urban-rural status depending on the chemical.

Table 5. Water domain variable means, standard deviations (SDs), and ranges - Overall and rural-urban continuum codes (RUCCs) stratified

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Construct: Domestic Use						
Percent pop. on self-supply	%	4.47E+01 (4.29E+01) [0.00E+00, 1.00E+02]	4.35E+01 (4.24E+01) [0.00E+00, 1.00E+02]	3.26E+01 (4.13E+01) [0.00E+00, 1.00E+02]	2.30E+01 (3.83E+01) [0.00E+00, 1.00E+02]	3.62E+01 (4.23E+01) [0.00E+00, 1.00E+02]
Percent pop. on self supply that is surface water	%	2.33E+01 (2.10E+01) [-2.62E-04, 1.00E+02]	2.40E+01 (1.72E+01) [0.00E+00, 8.20E+01]	2.99E+01 (2.10E+01) [-4.17E-02, 9.21E+01]	3.38E+01 (2.46E+01) [-6.78E-02, 1.00E+02]	2.77E+01 (2.18E+01) [-6.78E-02, 1.00E+02]
Construct: Overall Water Quality						
Percent of stream length impaired	%	1.97E+01 (2.35E+01) [1.00E-03, 1.56E+02]	1.72E+01 (2.30E+01) [1.00E-03, 1.00E+02]	1.28E+01 (1.84E+01) [1.00E-03, 1.08E+02]	9.02E+00 (1.33E+01) [1.00E-03, 1.00E+02]	1.50E+01 (2.05E+01) [1.00E-03, 1.56E+02]
Construct: General Water Contamination						
NPDES permits per 1000 km of stream	proportion	9.08E+01 (1.91E+02) [1.00E-03, 2.39E+03]	3.44E+01 (3.90E+01) [1.00E-03, 2.97E+02]	2.42E+01 (4.34E+01) [1.00E-03, 7.05E+02]	1.20E+01 (2.40E+01) [1.00E-03, 3.55E+02]	4.74E+01 (1.25E+02) [1.00E-03, 2.39E+03]
Construct: Atmospheric Deposition						
Calcium precipitation weighted mean	mg/L	1.63E-01 (9.69E-02) [1.22E-02, 5.94E-01]	1.83E-01 (1.10E-01) [1.22E-02, 7.48E-01]	2.03E-01 (1.20E-01) [3.80E-02, 1.06E+00]	2.23E-01 (1.09E-01) [3.66E-02, 8.63E-01]	1.90E-01 (1.11E-01) [1.22E-02, 1.06E+00]
Potassium precipitation weighted mean	mg/L	2.57E-01 (3.63E-02) [1.22E-01, 4.91E-01]	2.57E-01 (3.98E-02) [1.22E-01, 4.44E-01]	2.67E-01 (5.60E-02) [1.68E-01, 1.01E+00]	2.83E-01 (7.16E-02) [1.58E-01, 1.11E+00]	2.66E-01 (5.31E-02) [1.22E-01, 1.11E+00]
Nitrate precipitation	mg/L	7.34E-01 (2.11E-01) [0.00E+00, 1.13E+00]	7.38E-01 (2.40E-01) [0.00E+00, 1.14E+00]	7.44E-01 (2.03E-01) [1.93E-02, 1.14E+00]	7.55E-01 (2.07E-01) [5.47E-03, 1.14E+00]	7.42E-01 (2.10E-01) [0.00E+00, 1.14E+00]
Chloride precipitation weighted mean	mg/L	2.98E-01 (2.44E-01) [3.47E-02, 1.91E+00]	2.37E-01 (2.19E-01) [3.47E-02, 1.56E+00]	2.22E-01 (1.79E-01) [6.94E-02, 2.15E+00]	1.88E-01 (1.77E-01) [7.19E-02, 1.58E+00]	2.44E-01 (2.13E-01) [3.47E-02, 2.15E+00]
Sulfate precipitation weighted mean	mg/L	1.10E+00 (3.39E-01) [1.00E-01, 1.89E+00]	1.05E+00 (3.78E-01) [1.00E-01, 1.96E+00]	1.02E+00 (3.10E-01) [2.00E-01, 2.09E+00]	9.26E-01 (2.76E-01) [2.03E-01, 1.92E+00]	1.03E+00 (3.28E-01) [1.00E-01, 2.09E+00]
Total mercury deposition	ng/m2	9.44E+00 (2.59E+00) [2.81E-02, 1.84E+01]	9.02E+00 (2.67E+00) [2.62E-02, 1.76E+01]	9.29E+00 (2.66E+00) [3.62E-01, 1.55E+01]	8.43E+00 (2.88E+00) [1.60E-01, 1.46E+01]	9.15E+00 (2.71E+00) [2.62E-02, 1.84E+01]
Construct: Drough						
Percent of county drought extreme	%	4.16E+00 (7.38E+00) [0.00E+00, 4.52E+01]	3.70E+00 (6.67E+00) [0.00E+00, 3.87E+01]	3.76E+00 (6.51E+00) [0.00E+00, 4.82E+01]	3.43E+00 (5.92E+00) [0.00E+00, 4.43E+01]	3.84E+00 (6.75E+00) [0.00E+00, 4.82E+01]
Construct: Chemical Contamination						
Arsenic	mg/L	3.59E-03 (5.10E-03) [1.00E-03, 1.34E-01]	3.61E-03 (3.53E-03) [1.00E-03, 3.90E-02]	3.75E-03 (5.13E-03) [1.00E-03, 7.20E-02]	2.67E-03 (3.24E-03) [1.00E-03, 3.10E-02]	3.46E-03 (4.66E-03) [1.00E-03, 1.34E-01]

Table 5. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Barium	mg/L	8.08E-02 (3.93E-01) [1.00E-02, 1.31E+01]	8.34E-02 (2.37E-01) [1.00E-02, 3.98E+00]	6.81E-02 (9.96E-02) [1.00E-02, 1.03E+00]	4.84E-02 (7.72E-02) [1.00E-02, 6.70E-01]	7.03E-02 (2.59E-01) [1.00E-02, 1.31E+01]
Cadmium	mg/L	1.71E-03 (8.60E-04) [1.00E-03, 6.00E-03]	1.66E-03 (7.77E-04) [1.00E-03, 7.00E-03]	1.66E-03 (7.67E-04) [1.00E-03, 8.00E-03]	1.45E-03 (6.96E-04) [1.00E-03, 7.00E-03]	1.64E-03 (7.96E-04) [1.00E-03, 8.00E-03]
Chromium	mg/L	6.21E-03 (7.16E-03) [1.00E-03, 1.46E-01]	6.09E-03 (5.69E-03) [1.00E-03, 3.60E-02]	6.27E-03 (7.48E-03) [1.00E-03, 5.60E-02]	4.21E-03 (6.36E-03) [1.00E-03, 1.01E-01]	5.81E-03 (7.02E-03) [1.00E-03, 1.46E-01]
Cyanide	mg/L	1.51E-02 (2.85E-02) [1.00E-03, 2.67E-01]	1.68E-02 (2.92E-02) [1.00E-03, 2.11E-01]	1.57E-02 (3.18E-02) [1.00E-03, 3.39E-01]	1.39E-02 (4.12E-02) [1.00E-03, 8.16E-01]	1.52E-02 (3.26E-02) [1.00E-03, 8.16E-01]
Fluoride	mg/L	1.16E+00 (7.81E+00) [2.00E-02, 1.50E+02]	4.31E-01 (4.20E-01) [2.00E-02, 2.65E+00]	4.83E-01 (6.44E-01) [2.00E-02, 8.71E+00]	3.50E-01 (6.63E-01) [2.00E-02, 1.14E+01]	7.02E-01 (4.80E+00) [2.00E-02, 1.50E+02]
Mercury (inorganic)	mg/L	1.15E-03 (1.13E-03) [1.00E-03, 3.60E-02]	1.08E-03 (2.74E-04) [1.00E-03, 2.00E-03]	1.09E-03 (3.08E-04) [1.00E-03, 5.00E-03]	1.08E-03 (3.44E-04) [1.00E-03, 7.00E-03]	1.11E-03 (7.33E-04) [1.00E-03, 3.60E-02]
Nitrate	mg/L	8.07E-01 (1.64E+00) [1.00E-02, 2.00E+01]	6.59E-01 (1.19E+00) [1.00E-02, 1.46E+01]	7.37E-01 (2.80E+00) [1.00E-02, 8.10E+01]	6.22E-01 (2.01E+00) [1.00E-02, 3.28E+01]	7.32E-01 (2.13E+00) [1.00E-02, 8.10E+01]
Nitrite	mg/L	6.78E-02 (1.76E-01) [1.00E-02, 3.60E+00]	6.70E-02 (1.39E-01) [1.00E-02, 1.90E+00]	5.84E-02 (1.17E-01) [1.00E-02, 1.54E+00]	5.18E-02 (1.71E-01) [1.00E-02, 3.41E+00]	6.13E-02 (1.55E-01) [1.00E-02, 3.60E+00]
Selenium	mg/L	4.19E-03 (5.46E-03) [1.00E-03, 9.50E-02]	3.82E-03 (3.48E-03) [1.00E-03, 3.10E-02]	3.96E-03 (4.21E-03) [1.00E-03, 3.10E-02]	3.21E-03 (4.50E-03) [1.00E-03, 4.80E-02]	3.88E-03 (4.72E-03) [1.00E-03, 9.50E-02]
Antimony	mg/L	2.51E-03 (1.76E-03) [1.00E-03, 2.00E-02]	2.50E-03 (1.59E-03) [1.00E-03, 7.00E-03]	2.49E-03 (1.63E-03) [1.00E-03, 7.00E-03]	2.00E-03 (1.44E-03) [1.00E-03, 7.00E-03]	2.40E-03 (1.65E-03) [1.00E-03, 2.00E-02]
Endrin	mg/L	8.05E-02 (2.01E-01) [1.00E-02, 1.01E+00]	7.26E-02 (1.84E-01) [1.00E-02, 1.01E+00]	7.86E-02 (2.03E-01) [1.00E-02, 1.01E+00]	5.71E-02 (1.75E-01) [1.00E-02, 1.01E+00]	7.43E-02 (1.95E-01) [1.00E-02, 1.01E+00]
Methoxychlor	μg/L	6.90E-01 (1.97E+00) [1.00E-02, 1.00E+01]	5.66E-01 (1.65E+00) [1.00E-02, 1.00E+01]	4.06E-01 (1.23E+00) [1.00E-02, 9.65E+00]	1.59E-01 (6.02E-01) [1.00E-02, 8.01E+00]	4.76E-01 (1.52E+00) [1.00E-02, 1.00E+01]
Dalapon	μg/L	7.28E+00 (2.27E+01) [8.00E-02, 1.00E+02]	8.47E+00 (2.44E+01) [8.00E-02, 1.00E+02]	8.47E+00 (2.50E+01) [8.00E-02, 1.00E+02]	7.78E+00 (2.49E+01) [8.00E-02, 1.00E+02]	7.89E+00 (2.41E+01) [8.00E-02, 1.00E+02]
Di(2-ethylhexyl) adipate	μg/L	1.12E+01 (2.94E+02) [6.00E-02, 1.00E+04]	3.15E+00 (9.77E+00) [6.00E-02, 5.01E+01]	3.03E+00 (1.77E+01) [6.00E-02, 5.01E+02]	1.30E+00 (5.69E+00) [6.00E-02, 5.01E+01]	5.74E+00 (1.79E+02) [6.00E-02, 1.00E+04]
Simazine	μg/L	2.25E-01 (3.12E-01) [5.00E-02, 4.89E+00]	2.38E-01 (3.77E-01) [5.00E-02, 5.05E+00]	2.19E-01 (2.77E-01) [5.00E-02, 1.85E+00]	1.56E-01 (2.31E-01) [5.00E-02, 1.05E+00]	2.10E-01 (2.94E-01) [5.00E-02, 5.05E+00]
Di(2-ethylhexyl) pthalate	μg/L	8.55E-01 (1.26E+00) [8.00E-02, 9.41E+00]	8.72E-01 (1.20E+00) [8.00E-02, 6.08E+00]	7.87E-01 (1.29E+00) [8.00E-02, 1.59E+01]	4.79E-01 (8.87E-01) [8.00E-02, 9.15E+00]	7.57E-01 (1.21E+00) [8.00E-02, 1.59E+01]
Picloram	μg/L	2.44E+00 (1.00E+01) [4.00E-02, 5.00E+01]	3.64E+00 (1.25E+01) [4.00E-02, 1.00E+02]	2.54E+00 (1.00E+01) [4.00E-02, 5.00E+01]	1.22E+00 (6.36E+00) [4.00E-02, 5.00E+01]	2.34E+00 (9.71E+00) [4.00E-02, 1.00E+02]

Table 5. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Dinoseb	µg/L	2.94E-01 (4.19E-01) [8.00E-02, 3.08E+00]	3.32E-01 (4.45E-01) [8.00E-02, 2.08E+00]	2.92E-01 (4.64E-01) [8.00E-02, 9.08E+00]	2.48E-01 (3.87E-01) [8.00E-02, 2.08E+00]	2.88E-01 (4.31E-01) [8.00E-02, 9.08E+00]
Atrazine	µg/L	2.05E-01 (3.12E-01) [3.00E-02, 2.53E+00]	2.24E-01 (3.42E-01) [3.00E-02, 3.78E+00]	2.73E-01 (2.37E+00) [3.00E-02, 7.53E+01]	1.34E-01 (2.35E-01) [3.00E-02, 2.28E+00]	2.15E-01 (1.37E+00) [3.00E-02, 7.53E+01]
2,4-Dichlorophenoxyacetic acid	µg/L	1.40E-01 (1.08E-01) [9.00E-02, 2.51E+00]	1.42E-01 (5.41E-02) [9.00E-02, 4.00E-01]	1.42E-01 (2.27E-01) [9.00E-02, 7.19E+00]	1.20E-01 (5.30E-02) [9.00E-02, 8.10E-01]	1.37E-01 (1.49E-01) [9.00E-02, 7.19E+00]
Benzo[a]pyrene	µg/L	4.78E-02 (5.40E-02) [1.00E-02, 3.47E-01]	5.03E-02 (5.82E-02) [1.00E-02, 3.34E-01]	5.33E-02 (5.93E-02) [1.00E-02, 3.10E-01]	3.84E-02 (4.93E-02) [1.00E-02, 2.10E-01]	4.79E-02 (5.56E-02) [1.00E-02, 3.47E-01]
Pentachlorophenol	µg/L	7.84E-02 (1.63E-01) [1.00E-02, 1.71E+00]	8.91E-02 (1.81E-01) [1.00E-02, 1.01E+00]	8.82E-02 (1.76E-01) [1.00E-02, 1.01E+00]	6.16E-02 (1.36E-01) [1.00E-02, 1.01E+00]	7.92E-02 (1.65E-01) [1.00E-02, 1.71E+00]
Polychlorinated biphenyls	µg/L	1.65E-01 (1.19E+00) [6.00E-02, 4.04E+01]	1.13E-01 (1.24E-01) [6.00E-02, 1.06E+00]	1.13E-01 (1.88E-01) [6.00E-02, 4.31E+00]	8.13E-02 (6.53E-02) [6.00E-02, 1.06E+00]	1.26E-01 (7.35E-01) [6.00E-02, 4.04E+01]
1,2-Dibromo-3-chloropropane	µg/L	2.19E-02 (1.93E-02) [1.00E-02, 5.45E-01]	2.01E-02 (9.92E-03) [1.00E-02, 3.00E-02]	2.05E-02 (9.96E-03) [1.00E-02, 4.50E-02]	1.86E-02 (9.86E-03) [1.00E-02, 3.00E-02]	2.06E-02 (1.42E-02) [1.00E-02, 5.45E-01]
Ethylene dibromide	µg/L	8.28E-02 (1.60E-01) [1.00E-02, 1.17E+00]	7.14E-02 (1.39E-01) [1.00E-02, 5.10E-01]	6.94E-02 (1.41E-01) [1.00E-02, 8.70E-01]	8.19E-02 (1.59E-01) [1.00E-02, 5.10E-01]	7.72E-02 (1.52E-01) [1.00E-02, 1.17E+00]
Xylenes	µg/L	8.44E-01 (6.05E+00) [1.00E-01, 2.00E+02]	8.60E-01 (3.26E+00) [1.00E-01, 5.08E+01]	2.00E+00 (4.37E+01) [1.00E-01, 1.40E+03]	2.01E+00 (3.94E+01) [1.00E-01, 1.00E+03]	1.46E+00 (3.09E+01) [1.00E-01, 1.40E+03]
Chlordane	µg/L	1.08E-01 (9.94E-02) [2.00E-02, 9.70E-01]	1.17E-01 (9.62E-02) [2.00E-02, 2.76E-01]	1.12E-01 (9.77E-02) [2.00E-02, 2.87E-01]	8.43E-02 (9.23E-02) [2.00E-02, 2.20E-01]	1.06E-01 (9.77E-02) [2.00E-02, 9.70E-01]
Dichloromethane	µg/L	4.99E-01 (4.91E-01) [1.00E-01, 1.03E+01]	4.90E-01 (2.67E-01) [1.00E-01, 1.98E+00]	4.95E-01 (3.09E-01) [1.00E-01, 4.05E+00]	4.29E-01 (5.13E-01) [1.00E-01, 1.18E+01]	4.83E-01 (4.27E-01) [1.00E-01, 1.18E+01]
p-Dichlorobenzene	µg/L	5.09E-01 (5.13E+00) [2.00E-02, 1.75E+02]	3.72E-01 (2.41E-01) [2.00E-02, 1.54E+00]	3.62E-01 (2.57E-01) [2.00E-02, 2.77E+00]	3.11E-01 (3.55E-01) [2.00E-02, 6.02E+00]	4.07E-01 (3.13E+00) [2.00E-02, 1.75E+02]
1,1,1-Trichloroethane	µg/L	6.77E-01 (1.03E+01) [1.00E-02, 3.51E+02]	7.94E-01 (7.15E+00) [1.00E-02, 1.25E+02]	3.99E-01 (9.67E-01) [1.00E-02, 3.03E+01]	3.03E-01 (2.51E-01) [1.00E-02, 2.16E+00]	5.21E-01 (6.67E+00) [1.00E-02, 3.51E+02]
Trichloroethylene	µg/L	4.39E-01 (4.89E-01) [2.00E-02, 6.50E+00]	4.06E-01 (2.67E-01) [2.00E-02, 2.03E+00]	4.00E-01 (2.70E-01) [2.00E-02, 3.75E+00]	3.27E-01 (2.54E-01) [2.00E-02, 1.93E+00]	4.00E-01 (3.67E-01) [2.00E-02, 6.50E+00]
Carbon tetrachloride	µg/L	4.62E-01 (5.79E-01) [1.00E-02, 8.01E+00]	4.13E-01 (3.76E-01) [1.00E-02, 5.12E+00]	4.22E-01 (7.75E-01) [1.00E-02, 2.38E+01]	3.26E-01 (2.96E-01) [1.00E-02, 4.34E+00]	4.16E-01 (5.95E-01) [1.00E-02, 2.38E+01]
Benzene	µg/L	4.92E-01 (3.48E-01) [1.10E-01, 4.24E+00]	4.87E-01 (2.43E-01) [1.10E-01, 1.74E+00]	4.94E-01 (2.47E-01) [1.10E-01, 3.24E+00]	4.22E-01 (2.49E-01) [1.10E-01, 1.55E+00]	4.78E-01 (2.90E-01) [1.10E-01, 4.24E+00]
Toluene	µg/L	7.60E-01 (6.22E+00) [7.00E-02, 2.01E+02]	2.59E+00 (2.27E+01) [7.00E-02, 3.34E+02]	1.07E+00 (1.26E+01) [7.00E-02, 3.50E+02]	4.43E-01 (1.34E+00) [7.00E-02, 3.37E+01]	9.74E-01 (1.08E+01) [7.00E-02, 3.50E+02]
Ethylbenzene	µg/L	5.00E-02 (0.00E+00) [5.00E-02, 5.00E-02]	5.00E-02 (0.00E+00) [5.00E-02, 5.00E-02]	5.00E-02 (0.00E+00) [5.00E-02, 5.00E-02]	5.00E-02 (0.00E+00) [5.00E-02, 5.00E-02]	5.00E-02 (0.00E+00) [5.00E-02, 5.00E-02]

Table 5. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1,026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Styrene	µg/L	5.67E-01 (2.37E+00) [1.00E-01, 7.86E+01]	4.91E-01 (3.40E-01) [1.00E-01, 3.58E+00]	4.93E-01 (3.12E-01) [1.00E-01, 5.00E+00]	4.14E-01 (2.73E-01) [1.00E-01, 2.80E+00]	5.04E-01 (1.47E+00) [1.00E-01, 7.86E+01]
Alpha particles	pCi/L	1.05E+00 (2.32E+00) [0.00E+00, 3.58E+01]	1.24E+00 (3.42E+00) [0.00E+00, 5.15E+01]	1.34E+00 (3.19E+00) [0.00E+00, 3.47E+01]	7.33E-01 (2.01E+00) [0.00E+00, 1.81E+01]	1.10E+00 (2.71E+00) [0.00E+00, 5.15E+01]
cis-1,2-Dichloroethylene	µg/L	3.87E-01 (4.21E-01) [2.00E-02, 1.19E+01]	4.02E-01 (3.53E-01) [2.00E-02, 5.19E+00]	3.92E-01 (2.22E-01) [2.00E-02, 1.22E+00]	3.28E-01 (2.53E-01) [2.00E-02, 2.09E+00]	3.78E-01 (3.28E-01) [2.00E-02, 1.19E+01]
Construct: Drinking Water Quality						
Total coliform proportion	Proportion	1.20E-01 (3.55E-01) [1.00E-03, 4.93E+00]	2.86E-01 (1.26E+00) [1.00E-03, 1.34E+01]	2.03E-01 (8.82E-01) [1.00E-03, 1.84E+01]	2.22E-01 (8.41E-01) [1.00E-03, 9.71E+00]	1.84E-01 (7.76E-01) [1.00E-03, 1.84E+01]

Land Domain

In the land domain, the metropolitan-urbanized counties had lower agricultural-related variables (percent harvested and percent irrigated) than did nonmetropolitan-urbanized, less urban, and thinly populated counties (Table 6). Pesticides and animal units showed no clear pattern in variation across the strata. For example, average pounds of herbicides applied were 58,700, 78,400, 75,100, and 61,500 for most urban to most rural strata, respectively. There was little variation in the distribution of radon zones across the urban/rural strata.

Table 6. Land domain variable means, standard deviations (SDs), and ranges - Overall and rural-urban continuum codes (RUCCs) stratified

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Construct: Agriculture						
Farms per acre	Number	1.53E-03 (1.10E-03) [2.34E-06, 7.87E-03]	1.49E-03 (1.06E-03) [2.34E-06, 6.48E-03]	1.34E-03 (1.03E-03) [2.34E-06, 5.95E-03]	9.15E-04 (8.72E-04) [2.34E-06, 5.18E-03]	1.34E-03 (1.05E-03) [2.34E-06, 7.87E-03]
Irrigated acreage	%	2.20E+00 (6.72E+00) [3.62E-04, 7.42E+01]	3.46E+00 (9.15E+00) [3.62E-04, 5.65E+01]	3.45E+00 (8.73E+00) [3.62E-04, 7.14E+01]	2.81E+00 (7.39E+00) [3.62E-04, 6.07E+01]	2.86E+00 (7.83E+00) [3.62E-04, 7.42E+01]
Chemicals used to control nematodes, acres applied per county acres	%	1.01E-02 (1.28E-02) [1.32E-06, 1.07E-01]	1.14E-02 (1.54E-02) [1.32E-06, 1.30E-01]	1.27E-02 (1.60E-02) [1.32E-06, 1.50E-01]	8.75E-03 (1.08E-02) [1.32E-06, 9.63E-02]	1.08E-02 (1.39E-02) [1.32E-06, 1.50E-01]
Manure, acres applied per county acres	%	1.69E-02 (2.56E-02) [1.56E-06, 2.63E-01]	2.10E-02 (2.71E-02) [1.56E-06, 1.68E-01]	1.96E-02 (2.83E-02) [1.56E-06, 2.52E-01]	1.12E-02 (1.78E-02) [1.56E-06, 1.54E-01]	1.70E-02 (2.55E-02) [1.56E-06, 2.63E-01]
Chemicals used to control diseases in crops and orchards, acres applied per county acres	%	1.48E-02 (2.62E-02) [8.78E-07, 2.25E-01]	1.68E-02 (2.63E-02) [8.78E-07, 1.59E-01]	1.86E-02 (3.06E-02) [8.78E-07, 2.60E-01]	1.95E-02 (3.32E-02) [8.78E-07, 3.05E-01]	1.72E-02 (2.93E-02) [8.78E-07, 3.05E-01]
Chemicals used to defoliate/ control growth/thin fruit, acres applied per county acres	%	1.46E-02 (2.91E-02) [8.49E-07, 3.84E-01]	1.67E-02 (3.28E-02) [8.49E-07, 3.63E-01]	1.91E-02 (3.37E-02) [8.49E-07, 4.15E-01]	1.32E-02 (1.92E-02) [8.49E-07, 2.12E-01]	1.60E-02 (2.95E-02) [8.49E-07, 4.15E-01]
Harvested acreage, acres harvested per county acres	%	1.90E-01 (2.12E-01) [2.59E-05, 9.94E-01]	2.47E-01 (2.50E-01) [2.59E-05, 9.16E-01]	2.51E-01 (2.60E-01) [2.59E-05, 9.43E-01]	2.18E-01 (2.25E-01) [2.59E-05, 9.21E-01]	2.21E-01 (2.37E-01) [2.59E-05, 9.94E-01]
Animal Units, animal units per county acres	%	2.62E-04 (1.01E-03) [1.31E-08, 1.75E-02]	1.11E-04 (2.08E-04) [1.31E-08, 2.36E-03]	1.29E-04 (4.09E-04) [1.31E-08, 6.14E-03]	1.32E-04 (5.43E-04) [1.31E-08, 6.75E-03]	1.77E-04 (7.11E-04) [1.31E-08, 1.75E-02]

Table 6. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1,167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	Total (3143) Mean (SD) [Range]
Construct: Pesticides						
Fungicides, applied	Pounds	2.66E+04 (2.00E+05) [3.75E-01, 5.17E+06]	8.56E+03 (2.44E+04) [3.00E-01, 2.24E+05]	6.37E+03 (1.74E+04) [2.00E-01, 2.37E+05]	3.96E+03 (9.61E+03) [4.33E-01, 1.59E+05]	1.36E+04 (1.23E+05) [2.00E-01, 5.17E+06]
Herbicides, applied	Pounds	5.87E+04 (8.30E+04) [2.23E+00, 8.68E+05]	7.84E+04 (9.32E+04) [7.00E-01, 6.17E+05]	7.51E+04 (8.39E+04) [1.42E+01, 4.75E+05]	6.15E+04 (7.00E+04) [2.00E-01, 4.28E+05]	6.65E+04 (8.22E+04) [2.00E-01, 8.68E+05]
Insecticides, applied	Pounds	9.61E+03 (3.23E+04) [2.00E-01, 5.72E+05]	8.96E+03 (2.11E+04) [2.01E+01, 2.30E+05]	8.11E+03 (1.42E+04) [1.85E+00, 2.57E+05]	5.18E+03 (7.47E+03) [1.00E-01, 9.77E+04]	8.15E+03 (2.26E+04) [1.00E-01, 5.72E+05]
Construct: Mines						
Primarily coal mines, mines per county pop.	Proportion	1.11E-04 (7.38E-04) [6.25E-07, 1.25E-02]	1.35E-04 (5.64E-04) [6.25E-07, 4.67E-03]	4.05E-04 (2.18E-03) [6.25E-07, 2.82E-02]	5.67E-04 (3.75E-03) [6.25E-07, 5.78E-02]	3.03E-04 (2.17E-03) [6.25E-07, 5.78E-02]
Primarily metal mines, mines per county pop.	Proportion	3.29E-05 (3.24E-04) [2.44E-07, 6.43E-03]	4.14E-05 (2.19E-04) [2.44E-07, 2.54E-03]	1.19E-04 (7.78E-04) [2.44E-07, 1.43E-02]	5.18E-04 (3.84E-03) [2.44E-07, 7.41E-02]	1.61E-04 (1.81E-03) [2.44E-07, 7.41E-02]
Primarily nonmetal mines, mines per county pop.	Proportion	3.16E-05 (2.57E-04) [2.86E-07, 7.67E-03]	3.08E-05 (7.09E-05) [2.86E-07, 6.35E-04]	7.76E-05 (3.34E-04) [2.86E-07, 6.41E-03]	1.43E-04 (8.15E-04) [2.86E-07, 1.66E-02]	6.94E-05 (4.46E-04) [2.86E-07, 1.66E-02]
Primarily sand and gravel mines, mines per county pop.	Proportion	1.40E-04 (3.49E-04) [2.00E-07, 6.87E-03]	2.07E-04 (2.38E-04) [2.00E-07, 1.25E-03]	3.47E-04 (4.78E-04) [2.00E-07, 4.43E-03]	8.32E-04 (1.34E-03) [2.00E-07, 1.24E-02]	3.56E-04 (7.49E-04) [2.00E-07, 1.24E-02]
Primarily stone mines, mines per county pop.	Proportion	9.42E-05 (3.10E-04) [3.06E-07, 5.66E-03]	1.12E-04 (1.78E-04) [3.06E-07, 1.95E-03]	2.04E-04 (5.12E-04) [3.06E-07, 9.32E-03]	3.40E-04 (1.32E-03) [3.06E-07, 2.42E-02]	1.82E-04 (7.00E-04) [3.06E-07, 2.42E-02]
Construct: Radon						
Radon	Ordinal	2.02E+00 (8.14E-01) [0.00E+00, 3.00E+00]	1.97E+00 (8.23E-01) [1.00E+00, 3.00E+00]	2.03E+00 (8.24E-01) [1.00E+00, 3.00E+00]	1.88E+00 (8.09E-01) [1.00E+00, 3.00E+00]	1.99E+00 (8.19E-01) [0.00E+00, 3.00E+00]
Construct: Facilities						
Facilities per county	Proportion	3.69E-04 (2.82E-04) [5.60E-06, 3.22E-03]	4.99E-04 (3.25E-04) [3.69E-05, 2.24E-03]	5.60E-04 (4.63E-04) [5.60E-06, 6.65E-03]	8.25E-04 (2.08E-03) [5.60E-06, 4.58E-02]	5.38E-04 (1.01E-03) [5.60E-06, 4.58E-02]

Sociodemographic Domain

Socioeconomic variables included in the sociodemographic domain indicated that rural counties generally were more deprived than were more urban counties (Table 7), with both the lowest household income ($30,300) and lowest household value ($94,900). From the crime perspective, however, rural areas were at an advantage compared with more urban areas; the mean violent crime rate per county population for rural counties was 385.5 compared with 619.8 for the most urban counties.

Table 7. Sociodemographic domain variable means, standard deviations (SDs), and ranges - Overall and rural-continuum codes (RUCCs) stratified

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	OVERALL (n=3143) Mean (SD) [Range]
Sociodemographic Domain						
Construct: Socioeconomic						
Percent bachelor's degree	%	15.1 (5.8) [2.6, 37.2]	12.7 (4.6) [5.4, 34.7]	10.5 (4.0) [3.0, 42.2]	11.4 (4.6) [1.9, 36.1]	12.6 (5.3) [1.9, 42.2]
Percent unemployed	%	7.6 (2.5) [0, 27.5]	8.1 (2.6) [2.2, 20.2]	7.9 (3.4) [0.3, 26.3]	6.7 (4.6) [0.0, 30.9]	7.5 (3.6) [0, 30.9]
Percent families less than poverty level	%	9.8 (4.5) [0, 39.6]	11.9 (4.8) [3.1, 35.1]	12.7 (5.8) [1.4, 44.9]	11.9 (6.4) [0.0, 44.4]	11.4 (5.5) [0, 44.9]

Table 7. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	OVERALL (n=3143) Mean (SD) [Range]
Percent vacant housing	%	12.1 (6.5) [1.7, 60.1]	14.8 (7.7) [5.5, 63.9]	18.5 (9.3) [4.9, 68.0]	25.8 (12.3) [7.2, 83.3]	17.3 (10.3) [1.7, 83.3]
Median household value (X1000)	Dollar value	175.4 (103.9) [0, 868k]	135.4 (78.7) [57.0, 583.2k]	106.6 (64.9) [18.6, 100.0k]	94.9 (55.5) [29.7, 4965.6k]	133.5 (88.4) [0, 1000k]
Household income (X1000)	Dollars	82.6 (17.0) [67.0, 3217.9k]	23.1 (9.7) [5.9, 76.7k]	8.7 (4.9) [1.1, 30.7k]	3.0 (2.4) [0.2, 15.4k]	36.3k (109.9) [22, 321.8k]
Count of occupants per room	Count	0.6 (0.6) [0.1, 6.1]	0.6 (0.6) [0.1, 5.4]	0.8 (1.2) [0.1, 20.2]	0.9 (1.4) [0.1, 31.5]	0.7 (1.0) [0.1, 31.5]
Percent renter-occupied housing	%	28.0 (9.3) [8.7, 100]	30.0 (6.3) [16.8, 51.0]	26.2 (5.9) [11.3, 53.7]	23.6 (7.0) [8.7, 71.4]	26.7 (7.8) [8.7, 100]
Gini coefficient	Proportion	0.43 (0.04) [0.3, 0.6]	0.44 (0.03) [0.35, 0.54]	0.4 (0.0) [0.3, 0.6]	0.4 (0.0) [0.2, 0.6]	0.43 (0.04) [0.21, 0.65]
Construct: Crime						
Mean number of violent crimes per capita	Rate per county population	619.8 (441.4) [22.6, 6628.6]	472.3 (308.2) [19.52, 1735.0]	446.7 (249.8) [7.3, 1710.7]	385.5 (195.1) [69.9, 1420.1]	500.9 (344.5) [7.3, 6628.6]
Construct: County typology						
Creative class	%	0.2 (0.1) [0, 0.51]	0.2 (0.0) [0.1, 0.4]	0.2 (0.0) [0.0, 0.5]	0.15 (0.0) [0, 0.4]	0.18 (0.06) [0, 0.51]
Construct: County political valence						
Percent Democratic voters	%	44.8 (13.7) [5.5, 92.5]	43.9 (12.4) [12.5, 84.5]	40.2 (12.9) [7.8, 88.7]	36.4 (14.3) [4.9, 86.8]	41.5 (13.8) [4.9, 92.5]

NOTE: Means calculated using nontransformed variables
k = 1000

Built Domain

The most urban counties had a higher rate of traffic fatalities and residents reporting spending more time commuting compared with more rural areas (Table 8). Urban counties also had a higher walkability score but contained less green space and undeveloped areas than rural counties.

Table 8. Built-environment domain variable means, standard deviations (SDs), and ranges - Overall and rural-urban continuum codes (RUCCs) stratified

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	OVERALL (n=3143) Mean (SD) [Range]
Built Domain						
Construct: Business environment						
Vice-related environment	Count / county population	4.9e-4 (3.1e-4) [1.5e-5, 3.4e-3]	5.8e-4 (2.9e-4) [6.3e-5, 1.8e-3]	6.4e-4 (4.3e-4) [1.5e-5, 2.8e-3]	8.9e-4 (8.9e-3) [1.5e-5, 7.2e-3]	6.3e-4 (5.3e-4) [1.5e-5, 7.2e-3]
Civic-related environment	Count / county population	2.9e-3 (9.4e-4) [2.5e-4, 8.4e-4]	3.3e-3 (8.6e-4) [9.5e-4, 7.2e-3]	3.8e-3 (1.1e-3) [5.9e-4, 6.5e-3]	4.3e-3 (1.7e-3) [2.5e-4, 1.6e-2]	3.5e-3 (1.3e-3) [2.5e-4, 1.6e-2]
Education-related environment	Count / county population	1.2e-3 (4.2e-4) [1.8e-4, 4.5e-3]	1.3e-3 (3.6e-4) [6.3e-4, 3.2e-3]	1.5e-3 (6.0e-4) [5.9e-4, 6.5e-3]	2.5e-3 (1.8e-3) [1.8e-4, 1.8e-2]	1.6e-3 (1.0e-3) [1.8e-4, 1.8e-2]
Health care-related environment	Count / county population	3.4e-3 (1.6e-3) [3.4e-3, 1.6e-3]	3.7e-3 (1.1e-3) [1.0e-3, 1.1e-2]	3.2e-3 (1.3e-3) [6.0e-4, 2.0e-2]	2.8e-3 (1.4e-3) [1.0e-4, 9.1e-3]	3.2e-3 (1.4e-3) [1.0e-4, 2.0e-2]
Negative food environment	Count / county population	1.2e-3 (3.4e-4) [7.0e-5, 3.4e-3]	1.4e-3 (3.8e-4) [6.4e-4, 4.3e-3]	1.4e-3 (4.2e-4) [1.7e-4, 4.7e-3]	1.3e-3 (8.5e-4) [7.0e-5, 1.3e-2]	1.3e-3 (5.2e-4) [7.0e-5, 1.3e-2]
Positive food environment	Count / county population	2.2e-3 (7.7e-7) [1.3e-4, 8.1e-3]	2.3e-3 (8.5e-4) [1.0e-3, 7.8e-3]	2.4e-3 (8.9e-4) [4.4e-4, 9.0e-3]	2.9e-3 (1.7e-3) [1.3e-4, 2.0e-2]	2.4e-3 (1.1e-3) [1.3e-4, 2.0e-2]

Table 8. continued

Variable	Units	Metropolitan-Urbanized (RUCC1 = 1167) Mean (SD) [Range]	Nonmetropolitan-Urbanized (RUCC2 = 306) Mean (SD) [Range]	Less Urbanized (RUCC3 = 1026) Mean (SD) [Range]	Thinly Populated (RUCC4 = 644) Mean (SD) [Range]	OVERALL (n=3143) Mean (SD) [Range]
Recreation environment	Count / county population	1.3e-3 (6.1e-4) [4.7e-5, 1.1e-2]	1.6e-3 (8.5e-4) [3.0e-4, 8.8e-3]	1.7e-3 (1.0e-3) [1.2e-4, 1.0e-2]	2.2e-3 (1.9e-3) [4.7e-5, 1.8e-2]	1.6e-3 (1.2e-3) [4.7e-5, 1.8e-2]
Social service-related environment	Count / county population	1.5e-3 (5.9e-4) [9.2e-5, 5.1e-3]	1.8e-3 (6.5e-4) [6.2e-4, 4.8e-3]	1.8e-3 (7.8e-4) [3.0e-4, 5.2e-3]	1.9e-3 (1.1e-3) [9.2e-5, 8.4e-3]	1.7e-3 (8.0e-4) [9.2e-5, 8.4e-3]
Construct: Highway safety						
Traffic fatality rate	Fatality count / county population	23.2 (39.0) [1.0, 685.8]	11.2 (6.5) [1.3, 59.6]	5.7 (3.5) [1.0, 39.4]	2.8 (1.8) [1.0, 14.0]	12.1 (25.5) [1, 685.8]
Construct: Housing						
Rate of low-rent + Section 8 housing	Unit count / county population	0.2 (0.4) [0.0, 1.0]	0.2 (0.4) [0.0, 1.0]	0.4 (0.5) [0.0, 1.0]	0.6 (0.5) [0.0, 1.0]	0.4 (0.5) [0, 1]
Construct: Roads						
Proportion of roads that are secondary	Secondary road mile / total road miles	0.2 (0.1) [0.0, 0.5]	0.1 (0.1) [0.0, 0.44]	0.14 (0.1) [0.0, 0.4]	0.1 (0.1) [0.1, 24.1]	(0.1) [0.2, 0.5]
Construct: Commuting practices						
Residents who report using public transport	%	1.8 (4.6) [0.1, 60.5]	0.7 (1.2) [0.1, 12.8]	0.7 (1.0) [0.1, 13.0]	0.9 (1.2) [0.1, 24.1]	1.2 (3.0) [0.1, 60.5]
Commute time	Minutes	25.0 (5.1) [6.2, 60.5]	20.7 (3.6) [12.3, 31.8]	21.6 (5.0) [5.4, 38.5]	21.4 (6.4) [4.3, 44.2]	22.7 (5.5) [4.3, 44.2]
Construct: Walkability						
Walkability score	Ordinal	7.1 (2.3) [1.7, 16.2]	6.6 (1.1) [4.1, 13.8]	5.9 (1.1) [2.0, 10.5]	5.3 (1.2) [1.0, 9.5]	6.3 (1.8) [1.0, 16.2]
Construct: Green space						
County land area classified as natural cover and open space	%	61.5 (24.4) [3.9, 99.7]	62.3 (28.0) [5.3, 99.8]	63.2 (28.6) [6.9, 100.0]	68.5 (27.7) [6.2, 100.0]	63.5 (27.0) [3.9, 100.0]

NOTE: Means calculated using nontransformed variable

Variable Loadings on Environmental Quality Index Domains

Air Domain

The loadings for the variables comprising the air domain are displayed in Table 9. Each variable has been annotated with a "+" or "–"that is the predicted direction for the loading. Because we want to ensure that higher values of the EQI are associated with worse environmental quality, those variables that we anticipate being associated with poor environmental quality are assigned a "+" indicating more of this attribute would be a negative for health. All variables except for SO2 and benzidine (in certain strata) loaded as intended; loadings for SO2 and benzidine were relatively low. Most variables loaded consistently across rural-urban strata.

Table 9. Variable loadings, valence determination of variables - Air domain

Air Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan-Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	OVERALL (n=3143)
Construct: Criteria Air Pollutants					
PM10 (+)	0.0007	-0.0086	0.0016	0.0687	0.0272
PM2.5 (+)	0.1054	0.1191	0.1220	0.1204	0.1278
Ozone (+)	0.0224	0.0402	0.0273	0.0728	0.0398
Sulfur dioxide (+)	-0.0036	-0.0141	-0.0200	-0.0535	-0.0221
Nitrogen oxide (+)	0.1306	0.1665	0.1652	0.1626	0.1514
Carbon monoxide (+)	0.1345	0.1215	0.1458	0.1745	0.1513
Construct: Hazardous Air Pollutants					
Ethylene dibromide (+)	0.1120	0.1181	0.1131	0.1042	0.1179
Formaldehyde (+)	0.0443	0.0718	0.0794	0.0738	0.0798
1,1,2,2-Tetrachloroethane (+)	0.1410	0.1208	0.1478	0.1551	0.1475
1,1,2-Trichloroethane (+)	0.1654	0.1508	0.1583	0.1648	0.1616

Table 9. continued.

Air Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan-Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	OVERALL (n=3143)
1,2-Dibromo-3-chloropropane (+)	0.0722	0.0657	0.0416	0.0879	0.0688
1,2-Dichloropropane (+)	0.1069	0.1090	0.1095	0.1143	0.1129
Acrylic acid (+)	0.1714	0.1785	0.1727	0.1422	0.1661
Benzidine (+)	-0.0031	0.0023	-0.0058	0.0592	0.0135
Benzyl chloride (+)	0.1976	0.1926	0.1968	0.1850	0.1917
Beryllium compounds (+)	0.1761	0.1460	0.1343	0.1688	0.1557
bis-2-Ethylhexyl phthalate (+)	0.1046	0.1343	0.0872	0.1654	0.1192
Carbon tetrachloride (+)	0.0649	0.1127	0.0761	0.1272	0.0823
Carbonyl sulfide (+)	0.1524	0.1322	0.1439	0.1664	0.1580
Chlorine (+)	0.1791	0.1972	0.1877	0.1775	0.1866
Chlorobenzene (+)	0.2065	0.1810	0.1998	0.1995	0.2014
Chloroform (+)	0.1880	0.1674	0.1705	0.1713	0.1740
Chloroprene (+)	0.1724	0.1560	0.1479	0.1443	0.1537
Chromium compounds (+)	0.2012	0.2010	0.2010	0.1676	0.1904
Cobalt compounds (+)	0.2120	0.2223	0.2093	0.1908	0.2081
Cyanide compounds (+)	0.1722	0.1532	0.2033	0.1910	0.1825
Dibutylphthalate (+)	0.1923	0.2087	0.2029	0.1988	0.2000
Ethyl chloride (+)	0.1890	0.2047	0.1830	0.1946	0.1875
Ethyl benzene (+)	0.2407	0.2313	0.2343	0.2138	0.2306
Ethyl dichloride (+)	0.1275	0.1183	0.1299	0.1500	0.1344
Glycol ethers (+)	0.1882	0.1987	0.1965	0.1673	0.1884
Hydrazine (+)	0.1219	0.1434	0.1261	0.1186	0.1246
Hydrochloric acid (+)	0.1910	0.1987	0.2066	0.1974	0.1994
Isophorone (+)	0.1597	0.1775	0.1630	0.1667	0.1647
Manganese compounds (+)	0.1229	0.1369	0.1358	0.1187	0.1250
Methyl bromide (+)	0.1404	0.0889	0.1183	0.1355	0.1247
Methyl chloride (+)	0.1931	0.1905	0.1887	0.1756	0.1825
Phosphine (+)	0.0041	0.0014	0.0054	0.0439	0.0089
Polychlorinated biphenyls (+)	0.0971	0.1004	0.0933	0.1288	0.1040
Propylene dichloride (+)	0.1585	0.1529	0.1349	0.1254	0.1428
Quinoline (+)	0.1805	0.1881	0.1915	0.1560	0.1799
Trichloroethylene (+)	0.2283	0.2288	0.2296	0.1995	0.2210
Vinyl chloride (+)	0.1781	0.1577	0.1696	0.1767	0.1770

Water Domain

The loadings for the variables that comprise the water domain are displayed in Table 10. Each variable has been annotated with a “+” or “–“ that is the predicted direction for the loading. Because we want to ensure that higher values of the EQI are associated with worse environmental quality, those variables that we anticipate being associated with poor environmental quality are assigned a “+” indicating more of this attribute would be a negative for health. The variables in the drought, chemical contamination, and drinking water quality constructs loaded in the direction intended; however, some of the variables in the remaining constructs loaded in the opposite direction intended.

Table 10. Variable loadings, valence determination of variables - Water domain

Water Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	Total (All = 3143)
Construct: Domestic Use					
Percent of population on self-supply (+)	0.0028	0.0155	0.0203	0.0279	0.0096
Percent of public supply population on surface water (+)	0.0197	0.0155	-0.0004	0.0251	0.0191

Table 10. continued

Water Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	Total (All = 3143)
Construct: Overall Water Quality					
Percent of stream length impaired in county (+)	0.0142	-0.0174	-0.0053	0.0160	0.0111
Construct: General Water Contamination					
ALL NPDES permits per 1000 km of stream (+)	-0.0161	-0.0415	-0.0225	0.0164	-0.0009
Construct: Atmospheric Deposition					
Calcium precipitation weighted mean (+)	0.0378	0.0199	0.0347	-0.0039	0.0206
Potassium precipitation weighted mean (+)	-0.0108	-0.0236	-0.0075	-0.0291	-0.0204
Nitrate precipitation weighted mean (+)	0.0239	0.0014	0.0182	0.0009	0.0140
Chloride precipitation weighted mean (+)	-0.0408	-0.0329	-0.0457	-0.0077	-0.0278
Sulfate precipitation weighted mean (+)	-0.0162	-0.0217	-0.0086	0.0209	-0.0035
Total mercury deposition (+)	-0.0730	-0.0632	-0.0596	0.0015	-0.0462
Construct: Drought					
Percent of county drought - extreme (+)	0.0066	0.0179	0.0008	0.0142	0.0084
Construct: Chemical Contamination					
Arsenic (+)	0.1669	0.1674	0.1605	0.1584	0.1641
Barium (+)	0.1673	0.1684	0.1609	0.1628	0.1655
Cadmium (+)	0.1460	0.1475	0.1533	0.1615	0.1523
Chromium (+)	0.1661	0.1658	0.1592	0.1596	0.1636
Cyanide (+)	0.1369	0.1383	0.1181	0.1230	0.1291
Fluoride (+)	0.1736	0.1770	0.1804	0.1729	0.1765
Mercury (inorganic) (+)	0.0634	0.0494	0.0478	0.0614	0.0575
Nitrate (+)	0.1666	0.1600	0.1485	0.1417	0.1565
Nitrite (+)	0.1356	0.1322	0.1212	0.1231	0.1298
Selenium (+)	0.1661	0.1740	0.1644	0.1626	0.1663
Antimony (+)	0.1639	0.1541	0.1538	0.1586	0.1597
Endrin (+)	0.1392	0.1369	0.1387	0.1480	0.1412
Methoxychlor (+)	0.1670	0.1650	0.1676	0.1752	0.1690
Dalapon (+)	0.1462	0.1444	0.1409	0.1473	0.1449
Di(2-ethylhexyl) adipate (+)	0.1614	0.1576	0.1568	0.1624	0.1605
Simazine (+)	0.1674	0.1635	0.1651	0.1666	0.1671
Di(2-ethylhexyl) phthalate (+)	0.1682	0.1607	0.1594	0.1580	0.1638
Picloram (+)	0.1344	0.1301	0.1308	0.1445	0.1350
Dinoseb (+)	0.1599	0.1570	0.1550	0.1591	0.1584
Atrazine (+)	0.1758	0.1747	0.1738	0.1763	0.1759
2,4-Dichlorophenoxyacetic acid (+)	0.1612	0.1695	0.1565	0.1671	0.1623
Benzo[a]pyrene (+)	0.1578	0.1510	0.1538	0.1589	0.1561
Pentrachlorophenol (+)	0.1652	0.1622	0.1689	0.1715	0.1674
Polychlorinated biphenyls (+)	0.1244	0.1169	0.1081	0.1189	0.1185
1,2,-Dibromo-3-chloropropane (+)	0.1606	0.1552	0.1622	0.1631	0.1613
Ethylene dibromide (+)	0.0947	0.1043	0.1051	0.1035	0.1000
Xylenes (+)	0.1685	0.1654	0.1790	0.1816	0.1744
Chlordane (+)	0.1734	0.1755	0.1755	0.1763	0.1751
Dichloromethane (+)	0.1877	0.1950	0.1986	0.1900	0.1921
p-Dichlorobenzene (+)	0.1814	0.1886	0.1807	0.1814	0.1820
1,1,1-Trichloroethane (+)	0.1885	0.1917	0.1977	0.1906	0.1920
Trichloroethylene (+)	0.1893	0.1954	0.1992	0.1914	0.1932
Carbon tetrachloride (+)	0.1919	0.1968	0.2008	0.1926	0.1951
Benzene (+)	0.1880	0.1957	0.2008	0.1901	0.1929

Table 9. continued.

Water Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	Total (All = 3143)
Toluene (+)	0.1839	0.1736	0.1908	0.1876	0.1859
Styrene (+)	0.1822	0.1927	0.1980	0.1905	0.1896
Alpha particles (+)	0.0670	0.0537	0.0609	0.0771	0.0639
cis1,2-Dichloroethylene (+)	0.1892	0.1958	0.1998	0.1904	0.1930
Total coliform proportion (+)	0.0084	-0.0088	0.0008	0.0105	0.0067

Land Domain

The loadings for the variables that comprise the mines construct of the land domain varied by RUCC (Table 11), but loadings for the variables that comprise the other constructs (agriculture, pesticides, radon, and facilities) were consistent across RUCCs. Each variable again has been annotated with a "+" or "–" that is the predicted direction for the loading to ensure that higher values of the EQI represent worse environmental quality.

Table 11. Variable loadings, valence determination of variables - Land domain

Land Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	Total (All = 3143)
Construct: Agriculture					
Farms per acre (+)	0.3742	0.3148	0.3275	0.3501	0.3487
Irrigated acreage (+)	0.2750	0.1364	0.1789	0.1720	0.2109
Chemicals used to control nematodes (+)	0.3127	0.2753	0.2883	0.3297	0.3070
Manure (+)	0.3701	0.3049	0.3174	0.3561	0.3483
Chemicals used to control diseases in crops and orchards (+)	0.3589	0.3384	0.3302	0.3420	0.3479
Chemicals used to defoliate/control growth/thin fruit (+)	0.2796	0.2486	0.2630	0.3209	0.2793
Harvested acreage (+)	0.4173	0.3943	0.4039	0.4074	0.4156
Animal units (+)	0.1876	0.1135	0.1118	0.1603	0.1479
Construct: Pesticides					
Fungicides (+)	0.1055	0.2088	0.2125	0.0972	0.1582
Herbicides (+)	0.2007	0.3285	0.3177	0.2388	0.2742
Insecticides (+)	0.1759	0.2893	0.2604	0.1676	0.2272
Construct: Mines					
Primarily coal mines, mines per county population (+)	-0.0220	-0.0497	-0.0966	-0.0583	-0.0611
Primarily metal mines, mines per county population (+)	-0.0836	-0.2283	-0.1961	-0.2172	-0.1754
Primarily nonmetal mines, mines per county population (+)	0.0076	-0.0798	-0.0904	-0.0676	-0.0521
Primarily sand and gravel mines, mines per county population (+)	0.1181	-0.0229	-0.0341	0.0058	0.0270
Primarily stone mines, mines per county population (+)	0.0740	-0.0971	-0.1101	-0.1088	-0.0515
Construct: Radon					
Radon zone (+)	-0.0680	-0.0838	-0.0517	-0.1475	-0.0827
Construct: Facilities					
Facilities (+)	0.1389	0.2361	0.1930	0.1322	0.1598

Sociodemographic Domain

The loadings for the variables that comprise the sociodemographic domain varied by RUCC (Table 12), indicating some variables were more influential on the domain score in urban counties, whereas others exerted more of an effect in rural counties. For instance, percent unemployed loaded on the RUCC 1 sociodemographic domain at 0.16 compared with its loading on RUCC 4 sociodemographic domain of 0.44. Each variable has been annotated with a "+" or "–" that is the predicted direction for the loading. Because we want to ensure that higher values of the EQI are associated with worse environmental quality, those variables that we anticipate being associated with poor environmental quality are assigned a "+" indicating more of this attribute would be a negative for health. Most of the variables initially loaded in nearly the opposite direction intended. The loadings are a function of the program's starting point, or seed, which is not easily manipulable. Therefore, the loading valence needed to be corrected prior to the construction of the indices to ensure that higher values on a given index, and on the overall EQI, signify worse environmental quality. One important item to note is that the patterns of association within the socioeconomic construct across RUCC levels were not consistent. For instance, percent Democratic voting in the 2008 election loaded negatively in the most urban counties (RUCC 1 and 2) but positively in the less urban counties (RUCC 3 and 4). Percent of individuals earning a bachelor's degree, percent unemployed, percent of families in poverty, median household value, and creative class are variables that loaded in a consistent direction across rural-urban strata. Appendix V provides the original and modified valence corrected variable loadings.

Table 12. Valence corrected variable loadings, valence determination of variables - Sociodemographic domain

Sociodemographic Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan-Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	OVERALL (n=3143)
Socioeconomic Construct					
Percent bachelor's degree (-)	-0.4689	-0.4621	-0.4174	-0.4416	-0.4585
Percent unemployed (+)	0.1625	0.3274	0.3546	0.4418	0.1269
Percent families less than poverty level (+)	0.2591	0.4293	0.4737	0.4904	0.298
Percent vacant housing (+)	0.2306	-0.1331	-0.0555	-0.1381	0.1979
Median household value (-)	-0.4034	-0.4002	-0.3476	-0.2216	-0.4331
Household income (-)	-0.3700	-0.0874	-0.0640	0.2578	-0.3824
Count of occupants per room (+)	0.0055	0.1371	0.1116	-0.0141	0.1085
Percent renter-occupied housing (+)	-0.1827	0.0141	0.1523	0.0603	-0.1458
Gini coefficient (+)	-0.1162	0.1604	0.2725	0.2766	0.0118
Crime Construct					
Log violent crime (+)	-0.0094	0.2386	0.2997	0.2012	-0.0234
Creative class construct					
Creative class (-)	-0.4668	-0.4463	-0.3829	-0.2458	-0.4833
2008 Political valence construc					
Percent Democratic (-)	-0.2625	-0.0929	0.0374	0.2313	-0.211

Built Domain

Similar to the sociodemographic domain, the loadings for the variables that comprise the built domain varied by RUCC (Table 13), indicating some variables were more influential on the domain score in urban counties, whereas others exerted more of an effect in rural counties. Each variable again has been annotated with a "+" or "–" that is the predicted direction for the loading to ensure that higher values of the EQI represent worse environmental quality. Also, similar to the sociodemographic domain, many of the initial variable loadings are opposite to that intended. These loading valences needed to be valence corrected prior to the construction of the indices to ensure that higher values on a given index, and on the overall EQI, signify worse environmental quality. The business-related environments loaded consistently across RUCC levels, as did the public transportation, commute time and walkability score (Table 13). Appendix V provides the original and modified valence corrected variable loadings.

Table 13. Valence corrected variable loadings, valence determination of variables - Built domain

Built Domain	Metropolitan-Urbanized (RUCC1 = 1167)	Nonmetropolitan-Urbanized (RUCC2 = 306)	Less Urbanized (RUCC3 = 1026)	Thinly Populated (RUCC4 = 644)	OVERALL (n=3143)
Socioeconomic Construct					
Vice-related environment (+)	-0.2676	-0.0331	-0.2724	-0.2595	-0.2930
Civic-related environment (-)	-0.1238	-0.2057	-0.1890	-0.3102	-0.3071
Education-related environment (-)	-0.2409	-0.2626	-0.3278	-0.3285	-0.3495
Health care-related environment (-)	-0.4189	-0.3856	-0.3179	-0.2742	-0.2798
Negative food environment (+)	-0.3239	-0.2707	-0.2306	-0.1527	-0.2280
Positive food environment (-)	-0.3405	-0.2752	-0.2660	-0.2524	-0.3179
Recreation environment (-)	-0.2354	-0.3484	-0.3212	-0.3222	-0.3590
Social service-related environment (-)	-0.3446	-0.3503	-0.3644	-0.2793	-0.3629
Highway safety construct					
Traffic fatality rate (+)	-0.1978	0.2340	0.2197	0.2312	0.1751
Housing construct					
Rate of low-rent + Section 8 housing (+)	0.1230	-0.0459	-0.0697	0.0178	-0.0581
Road construct					
Proportion of secondary roads (+)	-0.0950	0.1319	0.1761	0.2054	0.1777
Commuting behavior construct					
Commute time (+)	0.1886	0.2808	0.3230	0.3546	0.3329
Public transportation (-)	-0.2253	-0.1111	-0.0777	-0.0256	-0.0463
Walkability construct					
Walkability score (-)	-0.3516	-0.3310	-0.3542	-0.3787	-0.1585
Green space construct					
Proportion green space (-)	0.1065	-0.0253	0.0418	0.1370	0.0451

Changes to 2006-2010 index construction from original 2000-2005 EQI

Valence Assignment

The sole modification to the PCA methodology in the county 2006-2010 EQI compared to that of the 2000-2005 EQI is "valence correction." We also have created a 2000-2005 valence corrected version of the EQI.

The loading pattern for the air domain, which is comprised of established pollutants, served as the reference for our index orientation. The vast majority of variables for the air domain loaded "+" for both the overall United States and across the rural-urban continuum. Thus, orientation for valence correction, if needed, was toward variables with known poor environmental attributes toward "+" loadings. Valence correction was applied only to the sociodemographic and built-environment domains. This is because only the sociodemographic and built domains had variables that were assigned as poor environmental attributes that loaded initially as "–". For instance, we were reasonably certain that a high percentage of unemployed per county (variable in sociodemographic domain) is anticipated to have deleterious effects (and, therefore, could be assigned a "+" loading sign based on our determined index orientation). Appendix V provides the modified loadings, when applicable, along with the rationale for valence correction.

Comparison of 2000-2005 EQI to the 2000-2005 valence corrected EQI

To assess the impact of valence correction, we computed Pearson and Spearman correlation coefficients between the nonvalence-corrected and valence-corrected 2000-2005 EQI. For the overall EQI, both the Pearson and Spearman correlation coefficients were roughly 1. For RUCC1, they were 0.99 across both. For RUCC2, the Pearson correlation coefficient was 0.99, whereas the Spearman correlation coefficient was 0.98. For RUCC3, the Pearson and Spearman correlation coefficients were -0.97 and -0.96, respectively. And, finally, for RUCC4, they were -0.97 and -0.97, respectively.

Comparison of 2000-2005 valence corrected EQI to the 2006-2010 EQI

We additionally computed Pearson and Spearman correlation coefficients between the valence corrected 2000-2005 EQI and the 2006-2010 EQI. The domain-specific loadings for the overall EQI differed over the two time periods in terms of magnitude, rank, and direction. These differential loadings contributed to the relatively low correlation between the 2000-2005 and 2006-2010 periods. For the overall EQI, the Pearson and Spearman correlation coefficients were both 0.34. For RUCC1, they were -0.71 and -0.72, respectively. For RUCC2, the Pearson correlation coefficient was -0.35, whereas the Spearman correlation coefficient was -0.37. For RUCC3, the Pearson and Spearman correlation coefficients were 0.64 and 0.69, respectively. And, finally, for RUCC4, they were 0.57 and 0.59, respectively. The loadings may have differed over the two time periods because of inputs that were included in the domains, valence correction procedures, and potential changes in environmental quality. It is for these reasons that we recommend the two indices not be compared over time.

Domain-Specific Index Description and Loadings on Overall EQI

The means, standard deviations, and ranges for each domain-specific index are presented in Table 14. As expected, the index loadings on the overall EQI index were mean (0) and standard deviation (1). In examining the ranges of each RUCC-stratified index, the larger the negative number (the smaller the minimum), the better the environmental quality, whereas the larger the maximum value, the worse the environmental quality. In general, higher values of each domain's index was found in the more metropolitan areas, and the maximum values went down as counties became more thinly populated.

Table 14. Description of the domain indices contributing to the overall and rural-urban continuum codes (RUCCs) stratified Environmental Quality Index for 3143 U.S. counties (2006-2010)

	Mean	Standard Deviation	Minimum	Maximum
All Counties (n=3143)				
Air Environment Index	-4.39E-10	1	-6.72	3.71
Water Environment Index	-3.48E-12	1	-1.46	2.05
Land Environment Index	-9.70E-10	1	-4.54	1.84
Built-Environment Index	1.20E-09	1	-4.71	5.66
Sociodemographic Environment Index	-2.11E-11	1	-5.13	2.76
Metropolitan-Urbanized (n=1167)				
Air Environment Index	-2.20E-10	1	-7.29	3.68
Water Environment Index	-1.38E-09	1	-1.48	1.93
Land Environment Index	1.28E-09	1	-4.30	1.80
Built-Environment Index	-1.93E-09	1	-3.62	7.29
Sociodemographic Environment Index	-7.23E-10	1	-4.28	2.78
Non-Metropolitan-Urbanized (n=306)				
Air Environment Index	-2.96E-09	1	-2.92	2.37
Water Environment Index	-1.59E-09	1	-1.61	1.56
Land Environment Index	-2.11E-09	1	-3.86	1.62
Built-Environment Index	-2.34E-09	1	-3.50	3.28
Sociodemographic Environment Index	-1.45E-10	1	-4.14	2.84
Less Urbanized (n=1026)				
Air Environment Index	8.32E-10	1	-2.67	3.31
Water Environment Index	2.94E-10	1	-3.95	2.37
Land Environment Index	7.79E-10	1	-3.88	1.61
Built-Environment Index	6.18E-10	1	-3.22	3.77
Sociodemographic Environment Index	7.34E-10	1	-4.79	3.64
Thinly Populated (n=644)				
Air Environment Index	1.40E-09	1	-5.69	2.17
Water Environment Index	1.30E-10	1	-1.21	1.96
Land Environment Index	5.36E-10	1	-4.32	1.51
Built-Environment Index	-4.06E-10	1	-2.64	4.20
Sociodemographic Environment Index	-1.17E-09	1	-3.51	3.81

Description of Overall EQI

The pattern of association for the domain-specific loadings differed by rural-urban status (Table 15). In the most urban areas, RUCC1, the sociodemographic and built-environment domains were both influential, as indicated by their loading values (0.68 and 0.67, respectively), followed by the land domain (0.23). For the nonmetropolitan-urbanized areas (RUCC2), the built and sociodemographic domains loaded similarly on the overall EQI (0.58 and 0.53, respectively), followed more closely by the air domain. In all but the overall EQI, the water domain was least influential, based on its low PCA coefficients. In the most thinly populated counties, RUCC4, the water and land domains were characterized by the lowest loadings (0.13 and 0.14, respectively), whereas the built, sociodemographic, and air domains were the most influential (loadings of 0.60, 0.56, and 0.54, respectively).

The built and the air domains loaded approximately equally on the overall EQI, and, unlike the loadings observed on the RUCC-stratified EQIs, the sociodemographic domain was relatively unimportant to the overall quality. Similar to the loadings for each domain, the loadings for each RUCC-stratified EQI was valence corrected to ensure that a higher EQI score corresponds to worse environmental quality. Appendix VI contains county mapping of the overall EQI 2006-2010 and RUCC-stratified domain-specific indices.

Table 15. Loadings of the domain indices contributing to the overall and rural-urban continuum codes (RUCCs) stratified Environmental Quality Index for 3143 U.S. counties (2006-2010)

Overall (n=3143)	Coefficient/Loading	95% Confidence Interval
Air Domain	0.6678	0.6238, 0.7118
Water Domain	0.2209	0.0940, 0.3479
Land Domain	0.3038	0.2054, 0.4021
Built-Environment Domain	0.6240	0.5582, 0.6898
Sociodemographic Domain	-0.1536	-0.2966, -0.0107
Metropolitan-Urbanized RUCC1 (n=1167)		
Air Domain	-0.1280	-0.2414, -0.0146
Water Domain	-0.0906	-0.2522, 0.7010
Land Domain	0.2340	0.0856, 0.3824
Built-Environment Domain	0.6730	0.6377, 0.7083
Sociodemographic Domain	0.6839	0.6476, 0.7201
Nonmetropolitan Urbanized Areas RUCC 2 (n=306)		
Air Domain	0.4128	0.2771, 0.5484
Water Domain	-0.2407	-0.4204, -0.0611
Land Domain	0.3926	0.2514, 0.5337
Built-Environment Domain	0.5274	0.4136, 0.6414
Sociodemographic Domain	0.5825	0.4939, 0.6712
Less Urbanized Areas RUCC 3 (n=1026)		
Air Domain	0.4785	0.4049, 0.5520
Water Domain	-0.1569	-0.2693, -0.0445
Land Domain	0.1769	0.0672, 0.2866
Built-Environment Domain	0.6370	0.5939, 0.6802
Sociodemographic Domain	0.5562	0.4939, 0.6184
Thinly Populated RUCC 4 (n=644)		
Air Domain	0.5402	0.4809, 0.5994
Water Domain	0.1323	0.0177, 0.2469
Land Domain	0.1430	0.0233, 0.2627
Built-Environment Domain	0.5960	0.5469, 0.6450
Sociodemographic Domain	0.5612	0.5064, 0.6160

4.0 Discussion

This report describes the efforts to update the Environmental Quality Index (EQI) for all counties in the United States for the 2006-2010 period. The EQI was created for two main purposes: (1) as an indicator of ambient conditions/exposure in environmental health modeling and (2) as a covariate to adjust for ambient conditions in environmental models. However, with the public release of the EQI and variables that constructed the EQI, other uses may emerge. The methods applied provide a reproducible approach that capitalizes almost exclusively on publicly available data sources.

The EQI holds promise for improving the environmental estimation in public health. The EQI describes the ambient county-level conditions to which residents are exposed, whether they are at home, at school, or at work, provided these multiple human activity spaces occur in the same county. Since the creation of the EQI 2000-2005, multiple studies have been conducted examining the relationship between overall environmental quality and health outcomes, including preterm birth[3], mortality[4], cancer incidence[5], asthma prevalence[6], physical inactivity and obesity[7], infant mortality[8], and pediatric multiple sclerosis[9]. A complete list of references related to EQI and health outcomes is listed in Appendix I.

With the updated EQI 2006-2010, the hope is that the EQI can continue to be used to help public health researchers investigate cumulative impact of various diverse constructs that typically are viewed in isolation. Each of the domain-specific pieces of information that contributes to the EQI is also informative. Because most environmental health practice occurs on a domain-specific basis, this domain-specific information may be important to policymakers and environmental health practitioners. The domain-specific loadings to the EQI indicate which of the environmental domains accounts for the largest portion of the variability in the EQI; in essence, these loadings answer the question about which domain is making the biggest contribution to the total environment. In addition, the variable loadings on each of the domains are also informative for the same reason.

The development of the EQI 2006-2010 followed mostly the same protocol as the EQI 2000-2005. Most of the constructs and the data sources identified for each of the five domains in the EQI 2000-2005 were maintained. Principal components analysis was used to develop the indices. However, using lessons learned from the creation of the EQI 2000-2005, some modifications were adopted to improve the EQI 2006-2010.

Summary of changes made to 2006-2010 version compared with 2000-2005

Modifications to the EQI 2006-2010 included exploring new data sources that were not available during EQI 2000-2005 development, assessment of all variables for continued inclusion in the EQI, and assessment of variables' valence within a domain and valence correction. Although most constructs were carried over from the EQI 2000-2005 to the updated EQI 2006-2010, the exceptions to this were the following: One deletion each in the water domain and land domain and constructs added to the water domain, land domain, sociodemographic domain, and the built-environment domain. For data sources, we added seven new data sources and discontinued use of one data source. Lastly, we assessed the valence of each domain to ensure that the orientation of the PCA output would have uniformity for interpretation of the domain indices and uniformity for orientation as input into the second PCA.

Strengths and Limitations

Because modifications were made to the updated EQI 2006-2010, direct comparisons between EQI 2000-2005 and EQI 2006-2010 should not be made. The two indices should not be examined as being continuous over time (e.g., if a study period covers 2004-2007, only one of the indices should be chosen or study population should be stratified by time period matched to the appropriate EQI).

The EQI offers a comprehensive measure of environmental quality for all counties in the United States and is comprised of many of the best environmental measures currently available. The EQI can be used as an ambient exposure metric to help identify environmental issues related to community health. It provides information on overall environmental exposures faced in a community. In addition, because data sources were used for all U.S. counties, the EQI is comparable across counties to help identify areas of better and worse overall environmental quality. The development of domain-specific indices enables counties to assess the drivers of poor environmental quality in their county. Additionally, because it is comparable across counties, areas that are burdened most by poor environmental quality can be identified. Finally, the EQI can be used in a variety of environmental health research activities as a control variable to adjust for overall environmental exposure, while trying to isolate a specific effect. Such a control variable will provide better estimates of effects by reducing confounding by co-occurring environmental factors.

The EQI is a national-level index that potentially can provide a better understanding into how multiple environmental conditions affect U.S. counties. At its current county-level scale, the EQI may not reveal environmental injustices seen at the local community level. However, it does highlight those counties experiencing an increased burden of environmental impacts. Further, the EQI can contribute to environmental justice endeavors by describing the process by which EQI data were obtained and how the EQI was constructed and by indicating the Web sites containing available data that can be used to construct indices at different levels of aggregation.

The EQI can be a tool for interested investigators to consider constructing local EQIs and adding relevant local-level data for more focused comparisons.

Use of the EQI as a measure of exposure assumes exposure to "environment" is consistent for all individuals, but the extent of individual environmental exposure was not assessable. The EQI was focused solely on the outside environment, which may not be the most relevant exposure in relation to human health and disease. Finally, population-level analyses offer little predictive utility for individual-level risk. Therefore, although the index may be useful at identifying less healthy county environments, it will not be useful for predicting individual-level adverse outcomes.

The EQI was developed for research purposes and is not meant to be a diagnostic tool. The EQI would be useful to identify potential areas of concern for counties to target future research, but it should not be used to target regulatory purposes.

Data

Data sources evaluated represented each of the five environmental domains. Each data source was reasonably well documented. Despite finding a considerable number of data sources applicable to each environmental domain, significant data gaps exist.

The data used to create the index balanced quality measurement with geographic breadth of coverage. Therefore, the index does a solid job estimating the ambient environment but may be less useful for estimating specific environments (e.g., in a particular noncounty location in the United States at a specific time). Not all relevant environmental exposures necessarily were included in the index. Data inclusion was dependent on data collection and coverage; if relevant data were not being collected, the information was not captured in the EQI. Relatedly, in areas where little data collection occurs, the data may be overrepresenting the environmental profile of those areas. For example, a county that contains a National Park without data collected and a town with data collection will be represented solely by the town data, although that may be inaccurate for the entire county. Conversely, environments with a wealth of environmental measurements, like urban areas, will be better estimated by the EQI.

Environmental data sources often are plagued by inadequate spatial and temporal coverage. Most of the data sources obtained for the EQI required spatial interpolation to achieve county-level estimates. For example, even with extensive air monitoring networks, the measured spatial coverage of the United States was incomplete, particularly in rural areas. Some types of measures were located disproportionately in urban areas (e.g., PM air pollution), whereas other sorts are found in rural areas (e.g., industrial livestock operations). The nonrandom distribution of environmental risk meant that virtually all interpolated data were inaccurate, impairing the assessment of how pollutants differentially impacted urban and rural areas.

From a human health perspective, probably the biggest limitation to existing environmental data sources is that data are collected with little thought given to potential health impacts. For instance, monitoring sites may collect relevant air pollutant data, but their location (e.g., air monitors located on top of buildings) is inappropriate for assessing the street-level values to which humans are exposed. Pesticide data, from the land domain, usually reports pesticide sales in relation to crops and livestock, not application, handling, or disbursement. Even the United States Census, which is widely used in health research, primarily is collected for tax and political districting purposes. Some of the data sources identified have not been used in human health research and, as such, are a limitation. Regularly collected, high-quality data that considers probable human health impacts would make the task of assessing differential exposures considerably easier.

Environmental data also were collected rarely with adequate temporal frequency. Although data on some parameters were collected on a consistent and frequent basis, the majority were not. Water data, for instance, were collected only sporadically in response to a particular query or based on regulatory statute. Within the sociodemographic domain, the complete United States Census was collected decennially, which limits investigators' capacity to explore temporal changes. Some characteristics of places can change rapidly, but, under current data collection schedules, these changes cannot be assessed. Initially, the EQI sought to estimate yearly measures. However, ultimately, only the 5-year (2006-2010) and 6-year (2000-2005) measures were created because of the lack of yearly data for some of the variables.

Many environmental parameters were compiled at a smaller unit of aggregation (e.g., for a municipality or city), and most were not maintained in a single source, such as a data repository. Although national repositories for some domains exist (e.g., water, air), often in response to federal regulations, no built-environment repository exists (for transit, walkability/physical activity, street connectivity, presence of sidewalks, or pedestrian lighting measures). Localities with limited funds may not be motivated or able to collect these data.

PCA Methodology

The use of PCA was not without limitations. Normality is an important assumption for PCA, and not all the data were distributed normally in their raw form. Many of the nonnormal variables were those with a substantial number of meaningful zeros (e.g., there were no public housing units contained within these counties). This "absence" of attribute is important information to convey, and, yet, it was problematic from a score-construction perspective. Although transforming the data improved their distribution, it reduced each variable's interpretability. A PCA-derived score also can be challenging to interpret. Outliers in the data also can be a limitation. However, with 3143 counties and normality checks, this is less problematic in the EQI.

Although limited, the use of PCA was also an important strength of this project. PCA provided a means to overcome one of the significant limitations in the field of environmental health and combine multiple environmental domains into one index of ambient environmental quality; the whole endeavor would not have been possible without this data reduction strategy. The resulting scale is standardized, which will facilitate its comparison to other scales constructed in different countries or at different units of aggregation. Further, it is the approach that has been used in other scale or score construction activities[65, 66].

Conclusion

The updated EQI 2006-2010 was constructed for all counties (n=3143) in the United States, incorporating data for five environmental domains, (1) air, (2) water, (3) land, (4) built, and (5) sociodemographic, and stratified by RUCCs. Mostly, the same reproducible approach used to create EQI 2000-2005 also was used to create EQI 2006-2010, with some noted changes that incorporate lessons learned from the first version. The EQI will be used as a measure in environmental health research. This broad-based effort acknowledges the many factors that together impact environmental quality and, more generally, recognizes that these factors work together to impact public health. Updates to the EQI for future years are planned, and the research team is actively creating a census tract version as a first step to explore other, finer spatial aggregations.

5.0
References

1. United States Environmental Protection Agency (EPA), *Creating an Overall Environmental Quality Index - Technical Report.* 2014. National Health and Environmental Effects Research Laboratory: Chapel Hill, NC.
2. United States Environmental Protection Agency (EPA), *EPA's 2008 Report on the Environment.* 2008: Washington, DC.
3. Rappazzo, K.M., et al., *The associations between environmental quality and preterm birth in the United States, 2000-2005: A cross-sectional analysis.* Environ Health, 2015. **14**: p. 50.
4. Jian, Y., et al., *Associations between Environmental Quality and Mortality in the Contiguous United States, 2000-2005.* Environ Health Perspect, 2017. **125**(3): p. 355-362.
5. Jagai, J.S., et al., *County-level cumulative environmental quality associated with cancer incidence.* Cancer, 2017. **123**(15): p. 2901-2908.
6. Gray, C.L., et al., *Associations between environmental quality and adult asthma prevalence in medical claims data.* Environ Res, 2018. **166**: p. 529-536.
7. Gray, C.L., et al., *The association between physical inactivity and obesity is modified by five domains of environmental quality in U.S. adults: A cross-sectional study.* PLoS One, 2018. **13**(8): p. e0203301.
8. Patel, A.P., et al., *Associations between environmental quality and infant mortality in the United States, 2000-2005.* Arch Public Health, 2018. **76**: p. 60.
9. Lavery, A.M., et al., *Examining the contributions of environmental quality to pediatric multiple sclerosis.* Mult Scler Relat Disord, 2017. **18**: p. 164-169.
10. United States Environmental Protection Agency (EPA), *Air Quality System Data Mart. The Ambient Air Monitoring Program.* 2010.
11. United States Environmental Protection Agency (EPA), *National Air Toxics Assessments.* 2005.
12. United States Environmental Protection Agency (EPA), *Watershed Assessment, Tracking, and Environmental Results (WATERS).* 2010.
13. Program, N.A.D., *National Atmospheric Deposition Program.* 2010.
14. United States Geological Survey (USGS), *Estimated Use of Water in the United States.* 2010.
15. United States Drought Monitor (USDM), *Drought Monitor Data Downloads.* 2010.
16. United States Environmental Protection Agency (EPA), *National Contaminant Occurrence Database (NCOD).* 2005.
17. United States Environmental Protection Agency (EPA), *Safe Drinking Water Information System.* 2010.
18. Stone, W.W., *Estimated annual agricultural pesticide use for counties of the conterminous United States,* 1992-2009. 2013, U.S. Geological Survey.
19. United States Department of Agriculture (USDA), *2007 Census of Agriculture full report.* 2009.
20. United States Environmental Protection Agency (EPA), *EPA Geospatial Data Download Service.* 2017.
21. United States Environmental Protection Agency (EPA), *Map of radon zones.* 2017.
22. United States Department of Labor Mines Safety Health Administration (MSHA), *Mines Data Set.* 2017.
23. United States Geological Survey (USGS), *National Geochemical Survey.* 2006.
24. Bureau, U.S.C., *American FactFinder.* 2017.
25. Federal Bureau of Investigation (FBI), *Uniform Crime Reports.* 2014.
26. Leip, D., *Dave Leip's Atlas of U.S. Presidential Elections.* 2016.
27. United States Department of Agriculture (USDA), *Economic Research Service (ERS) Creative Class County Codes.* 2017.
28. Bradstreet, D.a., *Dun and Bradstreet Products.* 2017.
29. Bureau, U.S.C., *Topologically Integrated Geographic Encoding and Referencing.* 2017.
30. HERE. *NAVTEQ traffic mapping.* 2019 [cited 2019 April 2]; Available from: https://www.here.com/navteq.
31. National Highway Traffic Safety Administration (NHTSA), N.C.f.S.a.A.N., *Fatality Analysis Reporting System (FARS).* 2017.
32. Development, U.S.D.o.H.a.U., *Multifamily Assistance and Section 8 Contracts Database.* 2017.
33. United States Environmental Protection Agency (EPA), *EnviroAtlas Green space dataset.* 2017.
34. Homer, C., et al., *Completion of the 2011 National Land Cover Database for the Conterminous United States–representing a decade of land cover change information.* 2015. **81**(5): p. 345-354.
35. National Oceanic and Atmospheric Administration, O.f.C.M., *Coastal Change Analysis Program (C-CAP) Regional Land Cover.* 2017.
36. United States Environmental Protection Agency (EPA), *National Walkability Index (NWI).* 2017.

37. United States Environmental Protection Agency (EPA). *National Emissions Inventory.* 2019 [cited 2019 April 2]; Available from: https://www.epa.gov/air-emissions-inventories/national-emissions-inventory-nei.
38. United States Geologic Services (USGS). *National Hydrography Dataset.* 2019 [cited 2019 April 2]; Available from: https://www.usgs.gov/core-science-systems/ngp/national-hydrography.
39. United States Environmental Protection Agency (EPA). *Reach Address Database.* 2010 [cited 2013 May 31]; Available from: http://www.epa.gov/waters/doc/rad/index.html.
40. United States Environmental Protection Agency (EPA). *EPA Report on the Environment.* 2019 [cited 2019 April 2]; Available from: https://www.epa.gov/report-environment.
41. Mult-Resolution Land Cover Characteristics (MRLC) Consortium. 2019 [cited 2019 April 2]; Available from: https://www.mrlc.gov/.
42. Cressie, N., *The origins of kriging.* Mathematical Geology, 1990. 22(3): p. 239-252.
43. Tabachnick, B.G., Fidell, L.S., *Using Multivariate Statistics*. 5th ed. 2007, Boston: Pearson Allyn and Bacon.
44. *Clean Water Act of 1972.*
45. United States Environmental Protection Agency (EPA). *National Pollutant Discharge Elimination System (NPDES).* December 12, 2018 [cited 2019 April 3]; Available from: https://www.epa.gov/npdes.
46. Kellog, R.L., Lander, C.H., Moffitt, D.C., Gollehon. N., *Manure Nutrients Relative to the Capacity of Cropland and Pastureland to Assimilate Nutrients: Spatial and Temporal Trends for the United States.* 2000, United States Department of Agriculture.
47. Baker, N.T., Stone, W.W., *Estimated annual agricultural pesticide use for counties of the conterminous United States,* 2008-12. 2014: U.S. Department of the Interior, U.S. Geological Survey.
48. United States Environmental Protection Agency (EPA). *Assessment, Cleanup, and Redevelopment Exchange (ACRES) Brownfield Sites.* 2010 [cited 2010 August 26]; Available from: http://www.epa.gov/brownfields/.
49. United States Environmental Protection Agency (EPA). *Superfund National Priorities List (NPL) Sites.* 2010; Available from: http://www.epa.gov/superfund/sites/npl/index.htm.
50. United States Environmental Protection Agency (EPA). *Section Seven Tracking System (SSTS) Pesticide Producing Site Locations.* 2019 [cited 2019 April 3]; Available from: https://www.epa.gov/frs/geospatial-data-download-service
51. United States Environmental Protection Agency (EPA). *Resource Conservation and Recovery Act (RCRA) Large Quantity Generators (LQG).* 2010 [cited 2010 August 26]; Available from: http://www.epa.gov/osw/hazard/generation/lqg.htm.
52. United States Environmental Protection Agency (EPA). *Resource Conservation and Recovery Act (RCRA) Treatment, Storage, and Disposal Facilities (TSD) and (RCRA) Corrective Action Facilities*. 2010 [cited 2010 August 26]; Available from: http://www.epa.gov/osw/hazard/tsd/index.htm.
53. National Technical Information Service. *Federal Information Processing Standards Publications (FIPS PUBS).* [cited 2013 August 1]; Available from: http://www.nist.gov/itl/fips.cfm.
54. Richardson, E.A., et al., *Green cities and health: a question of scale?* J Epidemiol Community Health, 2012. **66**(2): p. 160-5.
55. Access, G.B.D.H., et al., *Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990-2015: A novel analysis from the Global Burden of Disease Study 2015.* Lancet, 2017. **390**(10091): p. 231-266.
56. Friesen, C.E., Seliske, P., Papadopoulos, A., *Using principal component analysis to identify priority neighbourhoods for health services delivery by ranking socioeconomic status.* 2016. **8**(2).
57. Vyas, S., Kumaranayake, L., *Constructing socio-economic status indices: how to use principal components analysis.* 2006. **21**(6): p. 459-468.
58. Jolliffe, I.T., Cadima, J., *Principal component analysis: a review and recent developments.* Philos Trans A Math Phys Eng Sci, 2016. **374**(2065): p. 20150202.
59. Hall, S.A., Kaufman, J.S., Ricketts, T.C., *Defining urban and rural areas in U.S. epidemiologic studies.* J Urban Health, 2006. **83**(2): p. 162-75.
60. United States Department of Agriculture (USDA). *Measuring rurality: Rural-urban continuum codes.* [cited 2019 April 3]; Available from: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes//.
61. Langlois, P.H., et al., *Occurrence of conotruncal heart birth defects in Texas: a comparison of urban/rural classifications.* J Rural Health, 2010. **26**(2): p. 164-74.
62. Langlois, P.H., et al., *Urban versus rural residence and occurrence of septal heart defects in Texas.* Birth Defects Res A Clin Mol Teratol, 2009. **85**(9): p. 764-72.
63. Luben, T.J., et al., *Urban-rural residence and the occurrence of neural tube defects in Texas, 1999-2003.* Health Place, 2009. **15**(3): p. 848-54.
64. Messer, L.C., et al., *Urban-rural residence and the occurrence of cleft lip and cleft palate in Texas, 1999-2003.* Ann Epidemiol, 2010. **20**(1): p. 32-9.
65. Emerson, J., et al., *2012 Environmental Performance Index and Pilot Trend Environmental Performance Index - Full Report.* 2012, Yale Center for Environmental Law and Policy: New Haven, CT.
66. Messer, L.C., et al., *The development of a standardized neighborhood deprivation index.* J Urban Health, 2006. **83**(6): p. 1041-62.

Appendix I: List of References Related to 2000-2005 Environmental Quality Index

1. Lobdell DT, Jagai JS, Rappazzo K, Messer LC. (2011) Data sources for environmental assessment: determining availability, quality and utility, American Journal of Public Health Suppl 1:S277-85.
2. Jagai JS, Rosenbaum BJ, Pierson SM, Messer LC, Rappazzo K, Naumova EN, Lobdell DT. (2013) Putting Regulatory Data to Work at the Service of Public Health: Utilizing Data Collected Under the Clean Water Act. Water Quality, Exposure, and Health 5:117-125; DOI: 10.1007/s12403-013-0095-1.
3. Messer LC, Jagai JS, Rappazzo KM, Lobdell DT. (2014) Construction of an environmental quality index for public health research. Environmental Health 13:39; DOI: 10.1186/1476-069X-13-39.
4. Rappazzo KM, Messer LC, Jagai JS, Gray CL, Grabich SC, Lobdell DT. (2015) The association between environmental quality and preterm birth in the United States, 2000-2005: a cross-sectional analysis. Environmental Health 14:50;DOI: 10.1186/s12940-015-0038-3.
5. Grabich SC, Horney J, Konrad C, Lobdell DT. (2015). Measuring the Storm: Methods of Quantifying Hurricane Exposure with Pregnancy Outcomes. Natural Hazards Review; DOI: 10.1061/(ASCE)NH.1527-6996.0000204.
6. Grabich SC, Rappazzo KM, Gray CL, Jagai JS, Jian Y, Messer LM, Lobdell DT. (2016) Additive interaction between heterogeneous environmental quality domains (air, water, land, sociodemographic, and built environment) on preterm birth. Frontiers in Public Health, http://dx.doi.org/10.3389/fpubh.2016.00232.
7. Jian Y, Messer LC, Jagai JS, Rappazzo KM, Gray CL, Grabich SC, Lobdell DT. (2017) The associations between environmental quality and mortality in the contiguous United States 2000-2005. Environmental Health Perspectives 125:355-362, http://dx.doi.org/10.1289/EHP119.
8. Jagai JS, Messer LC, Rappazzo KM, Gray CL, Grabich SC, Lobdell DT. (2017) County-level cumulative environmental quality associated with cancer incidence. Cancer, http://dx.doi.org/10.1002/cncr.30709.
9. Lavery AM, Waldman AT, Charles Casper T, Roalstad S, Candee M, Rose J, Belman A, Weinstock-Guttman B, Aaen G, Tillema JM, Rodriguez M, Ness J, Harris Y, Graves J, Krupp L, Benson L, Gorman M, Moodley M, Rensel M, Goyal M, Mar S, Chitnis T, Schreiner T, Lotze T, Greenberg B, Kahn I, Rubin J, Waubant E; U.S. Network of Pediatric MS Centers. (2017) Examining the contributions of environmental quality to pediatric multiple sclerosis. Multiple Sclerosis and Related Disorders 18:164-169, https://doi.org/10.1016/j.msard.2017.09.004.
10. Jian Y, Wu CYH, Gohike JM. (2017) Effect modification by environmental quality on the association between heatwaves and mortality in Alabama, United States. International Journal of Environmental Research and Public Health 14:1143, https://doi.org/10.3390/ijerph14101143.
11. Gray CL, Lobdell DT, Rappazzo KM, Jian Y, Jagai JS, Messer LC, Patel AP, DeFlorio-Barker SA, Lyttle C, Solway J, Rzhetsky A. (2018) Associations between environmental quality and adult asthma prevalence in medical claims data. Environmental Research 166:529-536, https://doi.org/10.1016/j.envres.2018.06.020.
12. Gray CL, Messer LC, Rappazzo KM, Jagai JS, Grabich SC, Lobdell DT. (2018) The association between physical inactivity and obesity is modified by five domains of environmental quality in U.S. adults: A cross-sectional study. PLoS One, https://doi.org/10.1371/journal.pone.0203301.
13. Patel AP, Jagai JS, Messer LC, Gray CL, Rappazzo KM, Deflorio-Barker SA, Lobdell DT. (2018) Associations between environmental quality and infant mortality in the United States, 2000-2005. Archives of Public Health 76:60, https://doi.org/10.1186/s13690-018-0306-0.
14. Kosnik MB, Reif DM, Lobdell DT, Astell-Burt T, Feng X, Hader JD, Hoppin JA. (2019) Associations between access to healthcare, environmental quality, and end-stage renal disease survival time: proportional-hazards models of over 1,000,000 people over 14 years. PLoS One, https://doi.org/10.1371/journal.pone.0214094.
15. Jagai JS, Krajewski AK, Shaikh S, Lobdell DT, Sargis RM. (2020) Association between environmental quality and diabetes in the USA. Journal of Diabetes Investigation 11(2):315-324, https://doi.org/10.1111/jdi.13152.
16. Huanga M, Xiaob J, Nasca PC, Liu C, Lu Y, Lawrence WR, Wang L, Chen Q, Lin S. (2019) Do multiple environmental factors impact four cancers in women in the contiguous United States? Environmental Research 179:108782, https://doi.org/10.1016/j.envres.2019.108782.
17. Wang M, Wasserman E, Geyer N, Carroll RM, Zhao S, Zhang L, Hohl R, Lengerich EJ, McDonald AC. (2020) Spatial patterns in prostate cancer-specific mortality in Pennsylvania using Pennsylvania Cancer registry data, 2004-2014.

18. Gearhart-Serna LM, Hoffman K, Devi GR. (2020) Environmental Quality and Invasive Breast Cancer. Cancer Epidemiology, Biomarkers & Prevention; DOI: 10.1158/1055-9965.EPI-19-1497.

19. Li X, Xiao J, Huang M, Liu T, Guo L, Zeng W, Chen Q, Zhang J, Ma W. (2020) Associations of county-level cumulative environmental quality with mortality of chronic obstructive pulmonary disease and mortality of tracheal, bronchus, and lung cancers. Science of the Total Environment 703:135523, https://doi.org/10.1016/j.scitotenv.2019.135523.

Appendix II: Identified Variables by Source for Each Domain

Variables by Data Source - Air Domain

AIR QUALITY SYSTEM (AQS)

Variable	Variable Name	Counties/Monitors	Variable Notes	EQI Version
Particulate Matter <10 micrometers in aerodynamic diameter (PM10)	PM10	3143 / 303	µg/m3	2000-2005; 2006-2010
Particulate Matter <2.5 micrometers in aerodynamic diameter (PM2.5)	PM25	3143 / 1146	µg/m3	2000-2005; 2006-2010
Nitrogen Dioxide (NO2)	ln_NOx	3143 / 442	ppm, log transformed	2000-2005; 2006-2010
Sulfur Dioxide (SO2)	ln_SO2	3143 / 575	ppb, log transformed	2000-2005; 2006-2010
Ozone (O3)	O3	3143 / 1187	ppb	2000-2005; 2006-2010
Carbon Monoxide (CO)	ln_CO	3143 / 499	ppm, log transformed	2000-2005; 2006-2010

NATIONAL AIR TOXICS ASSESSMENT (NATA)

NOTES: WHEN DATA IS MISSING/NOT RECORDED, ZERO VALUES WERE DEEMED APPROPRIATE. MOST VARIABLES KEPT FOR EQI HAVE BEEN LOG TRANSFORMED. EQI 2006-2010 = NATA 2005. ALL VARIABLES REPORTED IN TONS EMITTED PER YEAR. UNLESS OTHERWISE NOTED, ALL VARIABLES ARE LOG TRANSFORMED. VARIABLES WERE DROPPED DUE TO INSUFFICIENT DATA (HIGH NUMBERS OF MISSING OR ZERO OBSERVATIONS) OR DUE TO HIGH CORRELATION WITH OTHER VARIABLES.

Variable	Variable Name	Counties	Variable Notes	EQI Version
1,1,2,2-tetrachloroethane	A_TeCA_ln	3137		2000-2005; 2006-2010
1,1,2-trichloroethane	A_112TCA_ln	3137		2000-2005; 2006-2010
1,2-dibromo-3-chloropropane	A_DBCP_ln	3137		2000-2005; 2006-2010
1,3-dichloropropene	A_DCl_propene_ln	3061		2006-2010
Acrylic acid	A_Acrylic_acid_ln	3107		2000-2005; 2006-2010
Benzidine	A_Benzidine_ln	3137		2000-2005; 2006-2010
Benzyl chloride	A_Benzyl_Cl_ln	3137		2000-2005; 2006-2010
Beryllium compounds	A_Be_ln	3137		2000-2005; 2006-2010
bis-2-ethylhexyl phthalate	A_DEHP_ln	3137		2000-2005; 2006-2010
Carbon tetrachloride	A_CCl4	3137		2000-2005; 2006-2010
Carbonyl sulfide	A_CylS_ln	3137		2006-2010
Chlorine	A_Cl_ln	3137		2000-2005; 2006-2010
Chlorobenzene	A_C6H5Cl_ln	3137		2000-2005; 2006-2010
Chloroform	A_chloroform_ln	3137		2000-2005; 2006-2010
Chloroprene	A_Chloroprene_ln	3137		2000-2005; 2006-2010
Chromium compounds	A_Cr_ln	3137		2000-2005; 2006-2010
Cobalt compounds	A_Co_ln	3132		2006-2010
Cyanide compounds	A_CN_ln	3137		2000-2005; 2006-2010
Dibutylphthalate	A_DBP_ln	3137		2000-2005; 2006-2010
Ethyl chloride	A_EtCl_ln	3136		2000-2005; 2006-2010
Ethylbenzene	A_Ebenzine	3137		2006-2010
Ethylene dibromide	A_EDB	3137		2000-2005; 2006-2010
Ethylene dichloride	A_EDC_ln	3137		2000-2005; 2006-2010
Formaldehyde	A_Formaldehyde	3137		2006-2010
Glycol ethers	A_Glycol_ethers_ln	3057		2000-2005; 2006-2010
Hydrazine	A_N2H2_ln	3137		2000-2005; 2006-2010

Variable	Variable Name	Counties	Variable Notes	EQI Version
Hydrochloric acid	A_HCl_ln	3137		2000-2005; 2006-2010
Isophorone	A_Isophorone_ln	3131		2000-2005; 2006-2010
Manganese compounds	A_Mn_ln	3137		2000-2005; 2006-2010
Methyl bromide	A_Me_Br_ln	3137		2006-2010
Methylene chloride	A_MeCl2_ln	3137		2000-2005; 2006-2010
Phosphine	A_PH3_ln	3062		2000-2005; 2006-2010
Polychlorinated biphenyls	A_PCBs_ln	3137		2000-2005; 2006-2010
Propylene dichloride	A_ProCl2_ln	3137		2000-2005; 2006-2010
Quinoline	A_Quinolin_ln	3137		2000-2005; 2006-2010
Trichloroethylene	A_C2HCl3_ln	3137		2000-2005; 2006-2010
Vinyl chloride	A_VyCl_ln	3137		2000-2005; 2006-2010

Variables by Data Source - Water Domain

WATERS PROGRAM DATABASE/REACH ADDRESS DATABASE

NOTES: THESE MEASURES WERE COMPUTED; LOTS OF MISSING DATA, SO SEVERAL VARIABLES CANNOT BE USED. VARIABLES CALCULATED USING REACH STREAM LENGTH DATABASE. DATA FOR 2006, 2008, AND 2010 WERE AVERAGED. DATA WAS UPDATED BASED ON 2010 FIPS CODES.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Percent of stream length impaired in county	D303_Percent	2513	Calculated with REACH database information	2000-2005; 2006-2010	
All NPDES Permits grouped per 1000km of stream length in county	ALLNPDESperKM	3141	All types of NPDES Permits	2006-2010	Grouped variable of Sewage Permits per 1000 km of Stream in County; Industrial Permits per 1000 km of Stream in County; Stormwater Permits per 1000 km of Stream in County

ESTIMATE USE OF WATER IN THE UNITED STATES

NOTES: THESE MEASURES WERE COMPUTED FOR 2005 AND 2010 DATA AND AVERAGED. USGS PROVIDES ESTIMATES AT COUNTY LEVEL, SO NO ADDITIONAL MANIPULATION REQUIRED.

Variable	Variable Name	Counties	Variable Notes
Percent of Population on Self Supply, 2005, 2010	Per_TotPopSS	3141	Estimate provided at county level
Percent of Public Supply Population that is on Surface Water, 2005, 2010	Per_PSWithSW	3067	Estimate provided at county level

NATIONAL ATMOSPHERIC DEPOSITION PROGRAM

NOTES: MEASURES PROVIDED AT VARIOUS MONITORING STATIONS. VALUES FOR 2006-2010 WERE KRIGED TO NATIONAL LEVEL COVERAGE. DATA FOR ALL YEARS WAS AVERAGED TOGETHER.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Calcium (Ca) precipitation weighted mean (mg/L)	CaAve_ln	3141	Kriged & log transformed	2000-2005; 2006-2010
Potassium (K) precipitation weighted mean (mg/L)	KAve_ln	3141	Kriged & log transformed	2000-2005; 2006-2010
Nitrate (NO3) precipitation weighted mean (mg/L)	NO3Ave	3141	Kriged – transformation not needed	2000-2005; 2006-2010
Chloride (Cl) deposition	ClAve_ln	3141	Kriged & log transformed	2000-2005; 2006-2010
Sulfate (SO4) deposition	SO4Ave_ln	3141	Kriged & log transformed	2000-2005; 2006-2010
Total Mercury deposition (ng/M2) Use only values with A or B quality rating	HgAve	3141	Kriged – transformation not needed	2000-2005; 2006-2010

DROUGHT MONITOR DATA

NOTES: RASTER DATA AGGREGATED TO THE COUNTY LEVEL. DATA FOR ALL YEARS 2006-2010 WAS AVERAGED TOGETHER.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Percent of county drought – extreme (D3-D4)	AvgOfD3_ave	3141		2000-2005; 2006-2010

NATIONAL CONTAMINANT OCCURRENCE DATABASE (NCOD)

NOTES: WILL USE 6 YEAR REVIEW 2 (DATA COLLECTED BETWEEN 1998-2005).

CALCULATE THE FOLLOWING VARIABLES FOR EACH CHEMICAL FOR EACH COUNTY (AGGREGATING ALL PWS IN COUNTY) FOR ALL YEARS COMBINED; MISSING FOR THOSE COUNTIES WITHOUT ANY DATA; DID NOT KEEP DETECTS.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Arsenic - average	W_As_ln (mg/L)	2017	Average for all samples in county, log transformed	2000-2005; 2006-2010
Barium - average	W_Ba_ln (mg/L)	1990	Average for all samples in county, log transformed	2000-2005; 2006-2010
Cadmium - average	W_Cd_ln (mg/L)	1991	Average for all samples in county, log transformed	2000-2005; 2006-2010
Chromium (total) - average	W_Cr_ln (mg/L)	1989	Average for all samples in county, log transformed	2000-2005; 2006-2010
Cyanide - average	W_CN_ln (mg/L)	1385	Average for all samples in county, log transformed	2000-2005; 2006-2010
Fluoride - average	W_FL_ln (mg/L)	2138	Average for all samples in county, log transformed	2000-2005; 2006-2010
Mercury (inorganic) - average	W_HG_ln (mg/L)	2056	Average for all samples in county, log transformed	2000-2005; 2006-2010
Nitrate (as N) - average	W_NO3_ln (mg/L)	1988	Average for all samples in county, log transformed	2000-2005; 2006-2010
Nitrite (as N) - average	W_NO2_ln (mg/L)	1583	Average for all samples in county, log transformed	2000-2005; 2006-2010
Selenium - average	W_SE_ln (mg/L)	1986	Average for all samples in county, log transformed	2000-2005; 2006-2010
Antimony - average	W_Sb_ln (mg/L)	1994	Average for all samples in county, log transformed	2000-2005; 2006-2010
Endrin - average	W_Endrin_ln (ug/L)	1509	Average for all samples in county, log transformed	2000-2005; 2006-2010
Methoxychlor - average	W_methoxychlor_ln (ug/L)	1512	Average for all samples in county, log transformed	2000-2005; 2006-2010
Dalapon - average	W_Dalapon_ln (ug/L)	1292	Average for all samples in county, log transformed	2000-2005; 2006-2010
Di(2-ethylhexyl)adipate (DEHA) - average	W_DEHA_ln (ug/L)	1456	Average for all samples in county, log transformed	2000-2005; 2006-2010
Simazine - average	W_Simazine_ln (ug/L)	1669	Average for all samples in county, log transformed	2000-2005; 2006-2010
Di(2-ethylhexyl) phthalate (DEHP) - average	W_DEHP_ln (ug/L)	1449	Average for all samples in county, log transformed	2000-2005; 2006-2010
Picloram - average	W_Picloram_ln (ug/L)	1352	Average for all samples in county, log transformed	2000-2005; 2006-2010
Dinoseb - average	W_Dinoseb_ln (ug/L)	1347	Average for all samples in county, log transformed	2000-2005; 2006-2010
Atrazine - average	W_atrazine_ln (ug/L)	1726	Average for all samples in county, log transformed	2000-2005; 2006-2010
2,4-D (2,4-Dichlorophenoxyacetic acid) - average	W_24D_ln (ug/L)	1360	Average for all samples in county, log transformed	2000-2005; 2006-2010
Benzo[a]pyrene - average	W_BenzoAP_ln (ug/L)	1430	Average for all samples in county, log transformed	2000-2005; 2006-2010
Pentachlorophenol - average	W_PCP_ln (ug/L)	1547	Average for all samples in county, log transformed	2000-2005; 2006-2010
Polychlorinated biphenyls (PCBs) - average	W_PCB_ln (ug/L)	848	Average for all samples in county, log transformed	2000-2005; 2006-2010
1,2-Dibromo-3-chloropropane (DBCP) - average	W_DBCP_ln (ug/L)	1652	Average for all samples in county, log transformed	2000-2005; 2006-2010
Ethylene dibromide (EDB) - average	W_EDB_ln (ug/L)	1630	Average for all samples in county, log transformed	2000-2005; 2006-2010
Xylenes (Total) - average	W_xylenes_ln (ug/L)	2203	Average for all samples in county, log transformed	2000-2005; 2006-2010
Chlordane - average	W_Chlordane_ln (ug/L)	1498	Average for all samples in county, log transformed	2000-2005; 2006-2010
Dichloromethane (Methylene chloride) - average	W_DCM_ln (ug/L)	2245	Average for all samples in county, log transformed	2000-2005; 2006-2010
1,4-Dichlorobenzene (p-Dichlorobenzene) - average	W_PDCB_ln (ug/L)	2165	Average for all samples in county, log transformed	2000-2005; 2006-2010
1,1,1-Trichloroethane - average	W_111trichlorane_ln (ug/L)	2238	Average for all samples in county, log transformed	2000-2005; 2006-2010
Trichloroethylene - average	W_Trichlorene_ln (ug/L)	2250	Average for all samples in county, log transformed	2000-2005; 2006-2010
Tetrachloroethylene - average	W_C2Cl4_ln (ug/L)	224	Average for all samples in county, log transformed	2000-2005; 2006-2010
Monochlorobenzene (Chlorobenzene) - average	W_benzene_ln (ug/L)	2239	Average for all samples in county, log transformed	2000-2005; 2006-2010
Toluene - average	W_Toluene_ln (ug/L)	2245	Average for all samples in county, log transformed	2000-2005; 2006-2010
Ethylbenzene - average	W_ethylbenz_ln (ug/L)	2241	Average for all samples in county, log transformed	2000-2005; 2006-2010
Styrene - average	W_styrene_ln (ug/L)	2235	Average for all samples in county, log transformed	2000-2005; 2006-2010
cis-1,2-Dichloroethylene - average	W_DCE_ln (ug/L)	2238	Average for all samples in county, log transformed	2000-2005; 2006-2010
Alpha Particles (Gross Alpha, excl.Radon&U) - average	W_alpha (PCi/L)	1243	Average for all samples in county	

SAFE DRINKING WATER INFORMATION SYSTEM (SDWIS)

NOTES: CUMULATIVE COUNT OF VIOLATIONS FOR ALL PWS IN COUNTY FOR THE YEAR. DATA IS AVAILABLE ANNUALLY. DATA WERE COMPILED FOR 2006-2010.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Total Coliform, Proportion	Coliform_Sum	2034		2006-2010

Variables by Source - Land Domain

2007 CENSUS OF AGRICULTURE

NOTES: ACRES OF CROP OR TREATMENT WERE DIVIDED BY TOTAL COUNTY ACRES TO GET PERCENTAGE OF ITEM PER COUNTY. SOME COUNTIES HAD SUPPRESSED ACREAGE DUE TO IDENTIFIABILITY ISSUES. FOR THESE, THE UNACCOUNTED-FOR ACREAGE FOR EACH STATE WAS CALCULATED (TOTAL STATE ACREAGE – LISTED COUNTY ACREAGE). THE ACREAGE WAS DIVIDED EQUALLY AMONG THE FARMS IN COUNTIES WITH SUPPRESSED INFORMATION. DATA FOR HAWAII AND ALASKA ARE NOT AVAILABLE. THESE DATA ARE REFRESHED EVERY 5 YEARS. THE NEXT AVAILABLE DATA IS FOR 2012.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Commercial fertilizer, lime, and soil conditioners	pct_lime_acres	3065		2000-2005; 2006-2010
Manure	pct_manure_acres_ln	2975		2000-2005; 2006-2010
Chemicals used to control insects	pct_insecticide_acres	3141		2000-2005; 2006-2010
Chemicals used to control weeds, grass, or brush	pct_weed_acres	3061		2000-2005; 2006-2010
Chemicals used to control nematodes	pct_nematode_acres_ln	1933		2000-2005; 2006-2010
Chemicals used to control diseases in crops and orchards	pct_disease_acres_ln	2530		2000-2005; 2006-2010
Chemicals used to control growth, thin fruit, or defoliate	pct_defoliate_acres_ln	1980		2000-2005; 2006-2010
Corn for grain (bushels)	pct_corn_acres	2588		2000-2005; 2006-2010
Potatoes (cwt)	Pct_potato_acres	1565		2000-2005; 2006-2010
Soybeans for beans (bushels)	pct_soybean_acres	2082		2000-2005; 2006-2010
Wheat for grain, all (bushels)	pct_wheat_acres	2520		2000-2005; 2006-2010
Animal units	pct_au_ln	3078	1 AU is equal to 0.94 cattle and calves, 5.88 hogs and pigs, 250 egg laying chickens, and 455 broiler chickens.	2000-2005; 2006-2010
Number of farms	farms_per_acre_ln	3039		2000-2005; 2006-2010
Irrigated acres	pct_irrigated_acres_ln	2815		2000-2005; 2006-2010
Harvested acres	pct_harvest_acres	2755		2000-2005; 2006-2010

2009 NATIONAL PESTICIDE USE DATASET (NPUD)

NOTES: PESTICIDE CONCENTRATIONS WERE GROUPED BY CLASS AND ADDED TOGETHER TO GET CLASS-LEVEL ESTIMATES OF PESTICIDE APPLICATION. THESE DATA ARE REFRESHED EVERY 5 YEARS. THE NEXT AVAILABLE DATA IS FOR 2012.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Insecticides	insecticides_ln	2761		2000-2005; 2006-2010
Herbicides	herbicides_ln	2907		2000-2005; 2006-2010
Fungicides	fungicides_ln	2256		2000-2005; 2006-2010

MAP OF RADON ZONE (EPA)

NOTES: THE EPA RADON ZONE MAP IDENTIFIES AREAS OF THE UNITED STATES WITH THE POTENTIAL FOR ELEVATED INDOOR RADON LEVELS. EACH UNITED STATES COUNTY (3142) IS ASSIGNED TO ONE OF THREE ZONES BASED ON RADON POTENTIAL. DATA YEARS UNAVAILABLE. PRESUMABLY, RADON IS A STABLE FEATURE, AND THE MAP IS NOT VARIABLE, BUT REFRESH DATES ARE NOT AVAILABLE. NO OTHER INFORMATION AVAILABLE IN DATA DOCUMENTATION.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Radon zones	Radon_zone	3142	3-level variable	2000-2005; 2006-2010

SUPERFUND NATIONAL PRIORITIES LIST (NPL) SITES

NOTES: NPL SITE LOCATIONS AVAILABLE THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED MONTHLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of Superfund National Priority List sites per county	sf_county_count	719		2000-2005; 2006-2010	Included as part of composite count variable

RESOURCE CONSERVATION AND RECOVERY ACT (RCRA) TREATMENT, STORAGE, AND DISPOSAL FACILITIES (TSD) AND RCRA CORRECTIVE ACTION FACILITIES

NOTES: RCA TSD AND CORRECTION ACTION FACILITIES SITE LOCATIONS AVAILABLE THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED MONTHLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of RCRA TSD and corrective action facilities per county	rcra_tsd_count_by_fips	874		2000-2005; 2006-2010	Included as part of composite count variable

RESOURCE CONSERVATION AND RECOVERY ACT (RCRA) LARGE QUANTITY GENERATORS (LQG)

NOTES: RCA LQG SITE LOCATIONS THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED MONTHLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of RCRA LQG facilities per county	rcralqg_count	1963		2000-2005; 2006-2010	Included as part of composite count variable

TOXIC RELEASE INVENTORY (TRI) SITES

NOTES: TRI SITES AVAILABLE THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED MONTHLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of TRI sites per county	tri_county_count	2671		2000-2005; 2006-2010	Included as part of composite count variable

ASSESSMENT, CLEANUP, AND REDEVELOPMENT EXCHANGE (ACRES) BROWNFIELD SITES

NOTES: BROWNFIELD SITE LOCATIONS AVAILABLE THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED MONTHLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of ACRES sites per county	acres_county_count	1273		2000-2005; 2006-2010	Included as part of composite count variable

SECTION SEVEN TRACKING SYSTEM (SSTS) PESTICIDE PRODUCING SITE LOCATIONS

NOTES: SSTS PESTICIDE-PRODUCING SITE LOCATIONS AVAILABLE THROUGH THE EPA GEOSPATIAL DATA ACCESS PROJECT. SITES WERE INCLUDED IN THE COUNTS IF THEY WERE IDENTIFIED BETWEEN 2006-2010. PUBLISHED AUGUST 2016. START AND END DATES NOT AVAILABLE. DATA REFRESHED BUT NOT ANNUALLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Count of SSTS sites per county	ssts_county_count	2099		2000-2005; 2006-2010	Included as part of composite count variable

MINE SAFETY AND HEALTH ADMINISTRATION (MSHA)

NOTES: THE MINE DATASET LISTS ALL COAL AND METAL/NON-METAL MINES UNDER MSHA'S JURISDICTION SINCE 1/1/1970. IT INCLUDES SUCH INFORMATION AS THE CURRENT STATUS OF EACH MINE (ACTIVE, ABANDONED, NONPRODUCING, ETC.), THE CURRENT OWNER AND OPERATING COMPANY, COMMODITY CODES AND PHYSICAL ATTRIBUTES OF THE MINE. MINE ID IS THE UNIQUE KEY FOR THIS DATA (HTTPS://ARLWEB.MSHA.GOV/OPENGOVERNMENTDATA/OGIMSHA.ASP). DATA REFRESHED WEEKLY. COUNTIES WITH ZERO MINES WERE GIVEN A VALUE OF MINIMUM VALUE/2. THESE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Primarily coal mines, mines per county population	Std_coal_prim_pop_ln	464	See notes above	2006-2010	
Primarily metal mines, mines per county population	Std_coal_prim_pop_ln	386	See notes above	2006-2010	
Primarily nonmetal mines, mines per county population	Std_coal_prim_pop_ln	1135	See notes above	2006-2010	
Primarily sand and gravel mines, mines per county population	Std_coal_prim_pop_ln	2342	See notes above	2006-2010	
Primarily stone mines, mines per county population	Std_coal_prim_pop_ln	1965	See notes above	2006-2010	

Variables by Source - Sociodemographic Domain

UNITED STATES CENSUS SUMMARY FILES

NOTES: MANY, MANY MORE VARIABLES ARE AVAILABLE FROM THE UNITED STATES CENSUS THAN WILL BE DESCRIBED HERE. THE VARIABLES IDENTIFIED HERE ARE THOSE THAT WILL BE USED IN THE EQI AND NOT THE PLETHORA OF VARIABLES THAT COULD BE CONSTRUCTED. DATA ARE AVAILABLE FOR MULTIPLE UNITS OF GEOGRAPHIC AGGREGATION, INCLUDING THE COUNTY-LEVEL. FULL POPULATION DATA ARE COLLECTED DECENNIALLY; SAMPLE DATA ARE COLLECTED MORE FREQUENTLY. DATA ARE AVAILABLE FOR DOWNLOAD FROM THE UNITED STATES CENSUS BUREAU WEB SITE.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Percent renter-occupied units	Pct_RenterOcc	3143		2000-2005; 2006-2010	
Percent vacant units	Pct_Vacant_Housing	3143		2000-2005; 2006-2010	
Median household value	med_hh_value	3143		2000-2005; 2006-2010	
Median household income	ln_HH_Inc	3143		2000-2005; 2006-2010	
Bachelor's degree or higher, percent of persons age 25 years+	Pct_BS	3143		2006-2010	This variable replaced percent < HS
Percent of persons who are unemployed	Pct_Unemp_total	3143		2000-2005; 2006-2010	
Percent of families in poverty	Pct_Fam_Pov	3143		2006-2010	This variable replaced percent families in poverty
Occupants per Room	ln_Occs_Room	3143		2006-2010	This variable replaced number rooms / house
Measure of income inequality (proportion)	GINI_est	3143		2006-2010	

FBI UNIFORM CRIME REPORTS

NOTES: FBI UCR DATA WERE DOWNLOADED FOR EACH COUNTY IN EACH STATE FROM THE WEBSITE (HTTPS://WWW.UCRDATATOOL.GOV/). DATA ARE AVAILABLE BY YEAR AND BY CRIME TYPE (VIOLENT = MURDER AND NONNEGLIGENT MANSLAUGHTER, FORCIBLE RAPE, ROBBERY, AND AGGRAVATED ASSAULT; PROPERTY = BURGLARY, LARCENY-THEFT, AND MOTOR VEHICLE THEFT). DATA FROM 2006-2010 WERE TEMPORALLY AND SPATIALLY KRIGED FOR USE IN THE EQI. DATA REPORTING IS VOLUNTARY. DATA ARE AVAILABLE AT THE CITY AND COUNTY LEVELS, BUT MANY COUNTIES DO NOT REPORT THESE DATA. DATA FOR LAW ENFORCEMENT AGENCIES SERVING CITY JURISDICTIONS WITH POPULATIONS OF 10,000 OR MORE AND COUNTY AGENCIES OF 25,000 OR MORE. THEREFORE, DATA MAY NOT BE AVAILABLE FOR EACH JURISDICTION EACH YEAR. DATA ARE AVAILABLE FROM 1960 TO CURRENT YEAR. RATES WERE OBTAINED FROM THE FBI. THE VIOLENT CRIME RATE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Violent crime rate	ln_ViolAv	3143	Variable kriged to estimate values for counties with no reported violent crime data	2000-2005; 2006-2010	
Murder-manslaughter crime rate	murder_manslaughter_rate	1062	Variable kriged to estimate values for counties with no reported violent crime data	No	Constituent of violent crime rate
Rape crime rate	rape_rate	1055	Variable kriged to estimate values for counties with no reported violent crime data	No	Constituent of violent crime rate
Robbery crime rate	rob_rate	1062	Variable kriged to estimate values for counties with no reported violent crime data	No	Constituent of violent crime rate
Aggravated assault crime rate	agg_assault_rate	1062	Variable kriged to estimate values for counties with no reported violent crime data	No	Constituent of violent crime rate

UNITED STATES DEPARTMENT OF AGRICULTURE ECONOMIC RESEARCH SERVICE CREATIVE CLASS INDEX

NOTES: THE ECONOMIC RESEARCH SERVICE (ERS) CLASS CODES INDICATE A COUNTY'S SHARE OF POPULATION EMPLOYED IN OCCUPATIONS THAT REQUIRE "THINKING CREATIVELY." THIS SKILL ELEMENT IS DEFINED AS "DEVELOPING, DESIGNING, OR CREATING NEW APPLICATIONS, IDEAS, RELATIONSHIPS, SYSTEMS, OR PRODUCTS, INCLUDING ARTISTIC CONTRIBUTIONS. " DATA ARE AVAILABLE FOR DOWNLOAD FROM THE USDA ERS WEBSITE.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Percent county employed in creative class	Num_CreatClass	3143		2006-2010

UNITED STATES ELECTION ATLAS

NOTES: THE POLITICAL CLIMATE OF A COUNTY WAS REPRESENTED BY THE DAVID LEIP ELECTION MAP. COUNTY-SPECIFIC PERCENTS VOTING REPUBLICAN OR DEMOCRATIC WERE REPORTED. THE REPORT VOTING DEMOCRATIC IN THE 2008 PRESIDENTIAL ELECTION WERE INCLUDED IN THE EQI.

Variable	Variable Name	Counties	Variable Notes	EQI Version
Percent county voting Democratic in 2008	DEMO2008	3143		2006-2010

Variables by Source - Built Domain

HOUSING AND URBAN DEVELOPMENT (HUD) DATA

NOTES: THESE DATA PROVIDE A COUNT OF THE LOW-RENT AND SECTION 8 HOUSING IN EACH HOUSING AUTHORITY AREA. THESE HOUSING AUTHORITY AREAS CORRESPOND TO CITIES, WHICH ARE THEN ASSIGNED FIPS CODES. COUNTIES WITHOUT HOUSING AUTHORITY CITIES ARE GIVEN A COUNT OF ZERO FOR LOW-RENT AND/OR SECTION-EIGHT HOUSING. THESE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA. DATA ARE REFRESHED FREQUENTLY), BUT UPDATE FREQUENCY NOT PROVIDED. HISTORIC DATA DOES NOT APPEAR TO BE AVAILABLE FROM WEB SITE. DATA WERE COLLECTED IN 2010, BUT, SINCE LOW-RENT AND SECTION 8 HOUSING DOES NOT CHANGE SUBSTANTIALLY OVER TIME, THESE DATA ARE CONSIDERED REPRESENTATIVE OF THE 2006-2010 TIME PERIOD. RATES FOR EACH VARIABLE CONSTRUCTED BY DIVIDING COUNT BY COUNTY POPULATION.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Rate of low-rent + section 8 units in county	total_units_ln	3143	Variable transformed (log) to allow it to approximate normal distribution	2000-2005; 2006-2010	Zeros considered meaningful zeros (lack of public housing)
Count of low-rent units per county	low_rent_units	2080	Variable transformed (log) to allow it to approximate normal distribution	No	Constituent of total unit rate
Count of section 8 units per county	section_eight_units	2080	Variable transformed (log) to allow it to approximate normal distribution	No	Constituent of total unit rate

FATALITY ANALYSIS REPORTING SYSTEM (FARS) DATA

NOTES: THE FATALITY ANALYSIS REPORTING SYSTEM (FARS) IS A NATIONWIDE CENSUS PROVIDING THE NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION YEARLY DATA REGARDING FATAL INJURIES SUFFERED IN MOTOR VEHICLE TRAFFIC CRASHES. FARS DATA ARE AVAILABLE FROM 1975 (HTTP://WWW.NHTSA.GOV/FARS/). RATES FOR THE COUNT OF FATAL CRASHES PER COUNTY FOR 2006-2010 WERE CONSTRUCTED BY DIVIDING COUNT BY COUNTY POPULATION. THESE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA. THESE DATA CAN BE UPDATED ANNUALLY.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Rate of fatal car crashes per county	ln_fatalities	3143	Variable transformed (log) to allow it to approximate normal distribution	2000-2005; 2006-2010	

2010 UNITED STATES CENSUS SUMMARY FILES

NOTES: MANY, MANY MORE VARIABLES ARE AVAILABLE FROM THE UNITED STATES CENSUS THAN WILL BE DESCRIBED HERE. THE VARIABLES IDENTIFIED HERE ARE THOSE THAT WILL BE USED IN THE EQI AND NOT THE PLETHORA OF VARIABLES THAT COULD BE CONSTRUCTED. DATA ARE AVAILABLE FOR MULTIPLE UNITS OF GEOGRAPHIC AGGREGATION, INCLUDING THE COUNTY-LEVEL. FULL POPULATION DATA ARE COLLECTED DECENNIALLY; SAMPLE DATA ARE COLLECTED MORE FREQUENTLY. THESE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA. DATA ARE AVAILABLE FOR DOWNLOAD FROM THE UNITED STATES CENSUS BUREAU WEB SITE.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Percent of county residents who report using public transportation	ln_PubTrans	3143	Variable transformed (log) to allow it to approximate normal distribution	2000-2005; 2006-2010	
Time it takes from home to go to work	CommuteTime	3143	Recorded in minutes	2006-2010	

TIGER FILES

NOTES: TOPOLOGICALLY INTEGRATED GEOGRAPHIC ENCODING AND REFERENCING PRODUCTS PROVIDE MAPS AND ROAD LAYERS WORLDWIDE. INCLUDING THE UNITED STATES. THESE DATA ARE UPDATED REGULARLY BUT DO NOT CHANGE SUBSTANTIALLY OVER TIME. THE DATA USED IN THE EQI ARE FROM 2009. DATA ARE AVAILABLE AT CENSUS GEOGRAPHY. FOR THE STREET TYPES, THE HIGHWAY AND SECONDARY AND LOCAL ROADS (TERTIARY ROADS) PER COUNTY PER STATE WERE DOWNLOADED. PROPORTION OF EACH ROAD TYPE WAS CONSTRUCTED BY DIVIDING THE DISTANCE OF EACH ROAD TYPE BY THE TOTAL AMOUNT OF EACH ROAD.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Proportion of all roads that are secondary roads	SecondaryRoadProportion	3143		2006-2010	This single variable replaced proportion primary road and highways

DUN AND BRADSTREET

NOTES: DUN AND BRADSTREET COLLECT COMMERCIAL INFORMATION ON BUSINESS. ITS DATABASE CONTAINS MORE THAN 195 MILLION RECORDS AND IS PROPRIETARY. THE DATA ARE PUT THROUGH AN EXTENSIVE QUALITY ASSURANCE PROCESS, WHICH INCLUDES OVER 2000 SEPARATE AUTOMATED AND SEVERAL MANUAL CHECKS. DATA ARE UPDATED DAILY. RATES OF EACH TYPE OF BUSINESS IN 2008 WERE CALCULATED BY DIVIDING THE COUNTS OF EACH VARIABLE BY THE COUNTY POPULATION. THESE DATA WERE TRANSFORMED (LOG) TO ACCOUNT FOR THE LARGE NUMBER OF ZEROS AND TO RESULT IN NEARLY NORMALLY DISTRIBUTED DATA.

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Rate of positive food environment businesses per county	pos_food_rate_ln	3140		2000-2005; 2006-2010	
Rate of negative food environment businesses per county	neg_food_rate_ln	3117		2000-2005; 2006-2010	
Rate of alcohol, pawn, gaming businesses per county	al_pwn_gm_env_rate_ln	3039		2000-2005; 2006-2010	
Rate of health care-related businesses per county	hc_env_rate_ln	3119		2000-2005; 2006-2010	
Rate of recreation-related businesses per county	rec_env_rate_ln	3133		2000-2005; 2006-2010	
Rate of education-related businesses per county	ed_env_rate_ln	3141		2000-2005; 2006-2010	
Rate of social-service-related businesses per county	ss_env_rate_ln	3125		2000-2005; 2006-2010	
Rate of civic-related businesses per county	civic_env_rate_ln	3138		2006-2010	

ENVIROATLAS LAND COVER CONTERMINOUS UNITED STATES (EPA)

NOTES: THIS ENVIROATLAS DATASET REPRESENTS THE PERCENTAGE OF LAND AREA THAT IS CLASSIFIED AS NATURAL, BARREN, FOREST, TUNDRA, SHRUBLAND, HERBACEOUS, WETLAND, WOODY WETLAND, EMERGENT WETLAND, ALL HUMAN LAND USE, DEVELOPED, OPEN SPACE DEVELOPED, LOW-INTENSITY DEVELOPED, MEDIUM-INTENSITY DEVELOPED, HIGH-INTENSITY DEVELOPED, AGRICULTURAL, PASTURE/HAY, AND CULTIVATED CROP USING THE 2011 NATIONAL LAND COVER DATASET (NLCD) FOR EACH COUNTY IN THE CONTERMINOUS UNITED STATES. THIS DATASET WAS PRODUCED BY THE UNITED STATES EPA TO SUPPORT RESEARCH AND ONLINE MAPPING ACTIVITIES RELATED TO ENVIROATLAS. ENVIROATLAS (HTTPS://WWW.EPA.GOV/ENVIROATLAS) ENABLES THE USER TO INTERACT WITH A WEB-BASED, EASY-TO-USE, MAPPING APPLICATION TO VIEW AND ANALYZE MULTIPLE ECOSYSTEM SERVICES FOR THE CONTIGUOUS UNITED STATES. THE DATASET IS AVAILABLE AS DOWNLOADABLE DATA (HTTPS://EDG.EPA.GOV/DATA/PUBLIC/ORD/ENVIROATLAS) OR AS AN ENVIROATLAS MAP SERVICE. ADDITIONAL DESCRIPTIVE INFORMATION ABOUT EACH ATTRIBUTE IN THIS DATASET CAN BE FOUND IN ITS ASSOCIATED ENVIROATLAS FACT SHEET (HTTPS://WWW.EPA.GOV/ENVIROATLAS/ENVIROATLAS-FACT-SHEETS).

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Combined natural land cover and open space developed	NINDEX_open	3109	Green space composite variable	2006-2010	
Percentage of county land area that is classified as natural land cover	NINDEX	3109	Composite variable of barren, forest, tundra, shrubland, herbaceous, and wetland land cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as barren land cover	pbar	3109	Vegetation accounts for <15% total cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as forest land cover	pfor	3109	Composite variable of deciduous, evergreen, and mixed forests. Areas dominated by trees generally greater than 5-meters tall, and greater than 20% total vegetation cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as tundra land cover	ptun	3109	Alaska only areas	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as shrubland land cover	pshb	3109	Areas dominated by shrubs; less than 5-meters tall; shrub canopy greater than 20% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as herbaceous land cover	phrb	3109	Areas dominated by graminoid and herbaceous vegetation, usually greater than 80% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as wetland land cover	pwtl	3109	Composite variable of woody and emergent wetlands.	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as woody wetland land cover	pwtlw	3109	Soil or substrate is periodically saturated with or covered with water, and forest or shrubland vegetation account for >20% vegetative cover	2006-2010	Included as part of green space composite variable

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
Percentage of county land area that is classified as emergent wetland land cover	pwtle	3109	Soil or substrate is periodically saturated with or covered with water, and perennial herbaceous vegetation accounts for >80% vegetative cover	No	Included as part of green space composite variable
Percentage of county land area that is classified as all human land use land cover	UINDEX	3109	Composite variable of developed and agricultural land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as developed land cover	pdev	3109	All developed land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as open space developed land cover	pdevo	3109	Mixture of some constructed materials but mostly vegetation; < 20% impervious surface	No	Included as part of green space composite variable
Percentage of county land area that is classified as low-intensity developed land cover	pdevl	3109	Mixture of constructed materials and vegetation; 20% to 49% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as medium-intensity developed land cover	pdevm	3109	Mixture of constructed materials and vegetation; 50% to 79% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as high-intensity developed land cover	pdevh	3109	Highly developed areas; 80% to 100% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as agricultural land cover	pagr	3109	Composite variable of pasture/hay and cultivated crop land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as pasture/hay land cover	pagrp	3109	Grasses, legumes, or grass-legume mixtures for livestock grazing; production of seed or hay crops; pasture/hay vegetation accounts for >20% total vegetation	No	Does not meet definition of green space
Percentage of county land area that is classified as cultivated crop land cover	pagrc	3109	Production of annual crops; crop vegetation accounts for >20% total vegetation; includes land being actively tilled	No	Does not meet definition of green space

ENVIROATLAS LAND COVER ALASKA (EPA)

NOTES: THIS ENVIROATLAS DATASET REPRESENTS THE PERCENTAGE OF LAND AREA THAT IS CLASSIFIED AS NATURAL, BARREN, FOREST, TUNDRA, SHRUBLAND, HERBACEOUS, WETLAND, WOODY WETLAND, EMERGENT WETLAND, ALL HUMAN LAND USE, DEVELOPED, OPEN SPACE DEVELOPED, LOW-INTENSITY DEVELOPED, MEDIUM-INTENSITY DEVELOPED, HIGH-INTENSITY DEVELOPED, AGRICULTURAL, PASTURE/HAY, CULTIVATED CROP, AND PERENNIAL SNOW/ICE USING THE 2011 NATIONAL LAND COVER DATASET (NLCD) FOR EACH COUNTY IN ALASKA. THIS DATASET WAS PRODUCED BY THE UNITED STATES EPA TO SUPPORT RESEARCH AND ONLINE MAPPING ACTIVITIES RELATED TO ENVIROATLAS. ENVIROATLAS (HTTPS://WWW.EPA.GOV/ENVIROATLAS) ENABLES THE USER TO INTERACT WITH A WEB-BASED, EASY-TO-USE, MAPPING APPLICATION TO VIEW AND ANALYZE MULTIPLE ECOSYSTEM SERVICES FOR THE CONTIGUOUS UNITED STATES. THE DATASET IS AVAILABLE AS DOWNLOADABLE DATA (HTTPS://EDG.EPA.GOV/DATA/PUBLIC/ORD/ENVIROATLAS) OR AS AN ENVIROATLAS MAP SERVICE. ADDITIONAL DESCRIPTIVE INFORMATION ABOUT EACH ATTRIBUTE IN THIS DATASET CAN BE FOUND IN ITS ASSOCIATED ENVIROATLAS FACT SHEET (HTTPS://WWW.EPA.GOV/ENVIROATLAS/ENVIROATLAS-FACT-SHEETS).

Variable	Variable Name	Counties	Variable Notes	EQIVersion	Notes
Combined natural land cover and open space developed	NINDEX_open	29	Green space composite variable	2006-2010	
Percentage of county land area that is classified as natural land cover	NINDEX	29	Composite variable of barren, forest, tundra, shrubland, herbaceous, and wetland land cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as barren land cover	pbar	29	Vegetation accounts for <15% total cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as forest land cover	pfor	29	Composite variable of deciduous, evergreen, and mixed forests. Areas dominated by trees generally greater than 5-meters tall, and greater than 20% total vegetation cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as tundra land cover	ptun	29	Alaska only areas; includes dwarf scrub, sedge/herbaceous, lichens, and moss land cover	2006-2010	Included as part of green space composite variable

Variable	Variable Name	Counties	Variable Notes	EQIVersion	Notes
Percentage of county land area that is classified as shrubland land cover	pshb	29	Areas dominated by shrubs; less than 5-meters tall; shrub canopy greater than 20% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as herbaceous land cover	phrb	29	Areas dominated by graminoid and herbaceous vegetation, usually greater than 80% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as wetland land cover	pwtl	29	Composite variable of woody and emergent wetlands	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as woody wetland land cover	pwtlw	29	Soil or substrate is periodically saturated with or covered with water and forest or shrubland vegetation account for >20% vegetative cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as emergent wetland land cover	pwtle	29	Soil or substrate is periodically saturated with or covered with water, and perennial herbaceous vegetation accounts for >80% vegetative cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as all human land use land cover	UINDEX	29	Composite variable of developed and agricultural land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as developed land cover	pdev	29	All developed land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as open space developed land cover	pdevo	29	Mixture of some constructed materials but mostly vegetation; <20% impervious surface	No	Included as part of green space composite variable
Percentage of county land area that is classified as low-intensity developed land cover	pdevl	29	Mixture of constructed materials and vegetation; 20% to 49% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as medium-intensity developed land cover	pdevm	29	Mixture of constructed materials and vegetation; 50% to 79% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as high-intensity developed land cover	pdevh	29	Highly developed areas; 80% to 100% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as agricultural land cover	pagr	29	Composite variable of pasture/hay and cultivated crop land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as pasture/hay land cover	pagrp	29	Grasses, legumes, or grass-legume mixtures for livestock grazing; production of seed or hay crops; pasture/hay vegetation accounts for >20% total vegetation	No	Does not meet definition of green space
Percentage of county land area that is classified as cultivated crop land cover	pagrc	29	Production of annual crops; crop vegetation accounts for >20% total vegetation; includes land being actively tilled	No	Does not meet definition of green space
Percentage of county land area that is classified as forest and woody wetland cover	Pfor90	29	Composite variable of forest and woody wetland	No	Included as part of green space composite variable
Percentage of county land area that is classified as forest and emergent wetland cover	Pwetl95	29	Composite of forest and emergent wetland	No	Included as part of green space composite variable
Percentage of county land area that is classified as perennial snow/ice	pice	29	Characterized by perennial cover of ice and/or snow, generally >25% total cover	No	Does not meet definition of green space

ENVIROATLAS LAND COVER HAWAII (EPA)

NOTES: THIS ENVIROATLAS DATASET REPRESENTS THE PERCENTAGE OF LAND AREA THAT IS CLASSIFIED AS NATURAL, BARREN, FOREST, TUNDRA, SHRUBLAND, HERBACEOUS, WETLAND, WOODY WETLAND, EMERGENT WETLAND, ALL HUMAN LAND USE, DEVELOPED, OPEN SPACE DEVELOPED, LOW-INTENSITY DEVELOPED, MEDIUM-INTENSITY DEVELOPED, HIGH -INTENSITY DEVELOPED, AGRICULTURAL, PASTURE/HAY, AND CULTIVATED CROP LAND COVER USING THE ENVIROATLAS COMPOSITE OF THE 2005-2011 COASTAL CHANGE ANALYSIS PROGRAM (C-CAP) LAND COVER DATASET FOR EACH 12-DIGIT HYDROLOGIC UNIT CODE (HUC) IN HAWAII. THIS DATASET WAS PRODUCED BY THE UNITED STATES EPA TO SUPPORT RESEARCH AND ONLINE MAPPING ACTIVITIES RELATED TO ENVIROATLAS. ENVIROATLAS (HTTPS://WWW.EPA.GOV/ENVIROATLAS) ENABLES THE USER TO INTERACT WITH A WEB-BASED, EASY-TO-USE, MAPPING APPLICATION TO VIEW AND ANALYZE MULTIPLE ECOSYSTEM SERVICES FOR THE CONTIGUOUS UNITED STATES. THE DATASET IS AVAILABLE AS DOWNLOADABLE DATA (HTTPS://EDG.EPA.GOV/DATA/PUBLIC/ORD/ENVIROATLAS) OR AS AN ENVIROATLAS MAP SERVICE. ADDITIONAL DESCRIPTIVE INFORMATION ABOUT EACH ATTRIBUTE IN THIS DATASET CAN BE FOUND IN ITS ASSOCIATED ENVIROATLAS FACT SHEET (HTTPS://WWW.EPA.GOV/ENVIROATLAS/ENVIROATLAS-FACT-SHEETS).

Variable	Variable Name	Counties	Variable Notes	EQIVersion	Notes
Combined natural land cover and open space developed	NINDEX_open	5	Green space composite variable	2006-2010	
Percentage of county land area that is classified as natural land cover	NINDEX	5	Composite variable of barren, forest, tundra, shrubland, herbaceous, and wetland land cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as barren land cover	pbar	5	Vegetation accounts for <15% total cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as forest land cover	pfor	5	Composite variable of deciduous, evergreen, and mixed forests. Areas dominated by trees generally greater than 5-meters tall, and greater than 20% total vegetation cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as tundra land cover	ptun	5	Alaska only areas	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as shrubland land cover	pshb	5	Areas dominated by shrubs; less than 5-meters tall; shrub canopy greater than 20% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as herbaceous land cover	phrb	5	Areas dominated by graminoid and herbaceous vegetation, usually greater than 80% of total vegetation	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as wetland land cover	pwtl	5	Composite variable of woody and emergent wetlands	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as woody wetland land cover	pwtlw	5	Soil or substrate is periodically saturated with or covered with water and forest or shrubland vegetation account for >20% vegetative cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as emergent wetland land cover	pwtle	5	Soil or substrate is periodically saturated with or covered with water and perennial herbaceous vegetation accounts for >80% vegetative cover	2006-2010	Included as part of green space composite variable
Percentage of county land area that is classified as all human land use land cover	UINDEX	5	Composite variable of developed and agricultural land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as developed land cover	pdev	5	All developed land cover	No	Does not meet definition of green space
Percentage of county land area that is classified as open space developed land cover	pdevo	5	Mixture of some constructed materials but mostly vegetation; < 20% impervious surface	No	Included as part of green space composite variable
Percentage of county land area that is classified as low-intensity developed land cover	pdevl	5	Mixture of constructed materials and vegetation; 20% to 49% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as medium-intensity developed land cover	pdevm	5	Mixture of constructed materials and vegetation; 50% to 79% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as high-intensity developed land cover	pdevh	5	Highly developed areas; 80% to 100% impervious surface	No	Does not meet definition of green space
Percentage of county land area that is classified as agricultural land cover	pagr	5	Composite variable of pasture/hay and cultivated crop land cover	No	Does not meet definition of green space

Variable	Variable Name	Counties	Variable Notes	EQIVersion	Notes
Percentage of county land area that is classified as pasture/hay land cover	pagrp	5	Grasses, legumes, or grass-legume mixtures for livestock grazing; production of seed or hay crops; pasture/hay vegetation accounts for >20% total vegetation	No	Does not meet definition of green space
Percentage of county land area that is classified as cultivated crop land cover	pagrc	5	Production of annual crops; crop vegetation accounts for >20% total vegetation; includes land being actively tilled	No	Does not meet definition of green space

NATIONAL WALKABILITY INDEX (EPA)

NOTES: THE NATIONAL WALKABILITY INDEX IS A NATIONWIDE GEOGRAPHIC DATA RESOURCE THAT RANKS BLOCK GROUPS ACCORDING TO THEIR RELATIVE WALKABILITY. THE NATIONAL DATASET INCLUDES WALKABILITY SCORES FOR ALL BLOCK GROUPS, AS WELL AS THE UNDERLYING ATTRIBUTES THAT ARE USED TO RANK THE BLOCK GROUPS. DATA ARE AVAILABLE FOR DOWNLOAD FROM THE EPA SMARTGROWTH WEB SITE (HTTPS://WWW.EPA.GOV/SMARTGROWTH/SMART-LOCATION-MAPPING#WALKABILITY).

Variable	Variable Name	Counties	Variable Notes	EQI Version	Notes
National walkability index score	Sum_NWIBG	3143	Scores were available at block group; county score created by adding block group scores, then taking mean of the block group scores based on county population proportions	2006-2010	

Appendix III: Changes in Variables from EQI 2000-2005 to EQI 2006-2010

Table A: Variables Added

Domain	Data Source	Variable	Variable Name	Notes
Water	Safe Drinking Water Information System (SDWIS)	Total coliform, proportion	Coliform_Sum	Added to drinking water quality construct
Land	Mine Safety and Health Administration (MSHA) Mines Data Set	Primarily coal mines, mines per county population	Std_coal_prim_pop_ln	Part of new mining activity construct
		Primarily metal mines, mines per county population	Std_coal_prim_pop_ln	Part of new mining activity construct
		Primarily nonmetal mines, mines per county population	Std_coal_prim_pop_ln	Part of new mining activity construct
		Primarily sand and gravel mines, mines per county population	Std_coal_prim_pop_ln	Part of new mining activity construct
		Primarily stone mines, mines per county population	Std_coal_prim_pop_ln	Part of new mining activity construct
Sociodemographic	United States Census	Measure of income inequality (proportion)	GINI_est	Added to socioeconomic construct
	United States Department of Agriculture Economic Research Service Creative Class	Percent county employed in creative class	Num_CreatClass	County creative typology construct
	United States Election Atlas	Percent county voting Democratic in 2008	DEMO2008	County political valence construct
Built	TIGER Files	Proportion of all roads that are secondary roads	SecondaryRoadProportion	Replaced proportion primary road and highways
	EnviroAtlas Land Cover	Combined natural land cover and open space developed	NINDEX_open	Green Space construct
	National Walkability Index (EPA)	National walkability index score	Sum_NWIBG	Walkability construct

Table B: Variables Changed

Domain	Data Source	Variable	Variable Name	Variable Replaced	Variable Replaced Name
Sociodemographic	United States Census	Bachelor's degree or higher, percent of persons age 25 years+	Pct_BS	Percent of persons with more than a high school education	Pct_hs_more
		Percent of families in poverty	Pct_Fam_Pov	Percent of persons less than poverty level	Pct_pers_lt_pov
		Occupants per room	ln_Occs_Room	Median number of rooms in residence	Med_rooms

Table C: Variables Deleted

Domain	Data Source	Variable	Variable Name	Reason Not Used
Land	National Geochemical Survey	Mean level of arsenic from sampled county sources	Mean_as_ln	Data quality
		Mean level of selenium from sampled county sources	Mean_se_ln	Data quality
		Mean level of mercury from sampled county sources	Mean_hg_ln	Data quality
		Mean level of lead from sampled county sources	Mean_pb_ln	Data quality
		Mean level of zinc from sampled county sources	Mean_zn_ln	Data quality
		Mean level of copper from sampled county sources	Mean_cu_ln	Data quality
		Mean level of aluminum from sampled county sources	Mean_al_pct	Data quality
		Mean level of sodium from sampled county sources	Mean_na_pct	Data quality
		Mean level of magnesium from sampled county sources	Mean_mg_pct_ln	Data quality

Table C: continued

Domain	Data Source	Variable	Variable Name	Reason Not Used
		Mean level of titanium from sampled county sources	Mean_ti_pct_ln	Data quality
		Mean level of calcium from sampled county sources	Mean_ca_pct_ln	Data quality
		Mean level of manganese from sampled county sources	Mean_mn	Data quality
		Mean level of iron from sampled county sources	Mean_fe_pct_ln	Data quality
		Mean level of phosphorus from sampled county sources	mean_al_pct	Data quality
Built	Dun & Bradstreet	Rate of transportation-related businesses per county	rate_trans_env_log	Captured by public transportation, commuting times and roads
		Rate of entertainment businesses per county	rate_ent_env_log	Dropped because there was no clear association with health
Built	TIGER files	Proportion of all roads that are highways Proportion of all roads that are primary roads	hwyprop primaryprop	Both variables replaced with secondary roads
Sociodemographic	United States Census	Percent of persons less than poverty level	pct_pers_lt_pov	Replaced with percent of families below poverty level
		Percent of persons who do not speak English	pct_no_eng	
		Percent of persons with more than high school education	pct_hs_more	Replaced with percent of persons with a bachelor's degree
		Percent of persons who work outside their county of residence	work_out_co	
		Median number of rooms in residence	med_rooms	Replaced with occupants per room
		Percent of residences with more than 10 units	pct_mt_10units_log	
Water	Watershed Assessment, Tracking and Environmental Results Program Database/ REACH Address Database	Sewage Permits per 1000 km of Stream in County	SEWAGENPDESperKM	Used group variable
		Industrial Permits per 1000 km of stream in county	INDNPDESperKM	Used group variable
		Stormwater Permits per 1000 km of stream in county	STORMNPDESperKM	Used group variable
		Number of days closed per event in county 2002	numDays_Close_Activity_2002	Not enough counties
		Number of days per contamination advisory event in county 2002	numDays_Cont_Activity_2002	Not enough counties
		Number of days per rain advisory event in county 2002	numDays_Rain_Activity_2002	Not enough counties
Water	National Atmospheric Deposition Program	Magnesium (Mg) precipitation weighted mean (mg/L)	Mg_ln	Correlated
		Sodium (Na) precipitation weighted mean (mg/L)	Na_ln	Correlated
		Ammonium (NH4) precipitation weighted mean (mg/L)	NH4_mean	Correlated
Water	National Contaminant Occurrence Database	Beryllium - average	W_Be_ln (mg/L)	Zeros
		Thallium - average	W_Tl_ln (mg/L)	Correlated
		Lindane - average	W_Lindane_ln (mg/L)	Correlated
		Toxaphene - average	W_Toxaphene_ln (ug/L)	Correlated
		Oxamyl (Vydate) - average	W_Oxamyl_ln (ug/L)	Correlated
		Hexachlorocyclopentadiene - average	W_HCCPD_ln (ug/L)	Correlated
		Carbofuran - average	W_Carbofuran_ln (ug/L)	Correlated
		Alachlor - average	W_Alachlor_ln (ug/L)	Correlated
		Heptachlor - average	W_Heptachlor_ln (ug/L)	Correlated
		Heptachlor epoxide - average	W_Heptachlor_epox_ln (ug/L)	Correlated
		2,4,5-TP (Silvex) - average	W_silvex_ln (ug/L)	Correlated
		Hexachlorobenzene - average	W_HCB_ln (ug/L)	Correlated
		1,2,4-Trichlorobenzene - average	W_124TCIB_ln (ug/L)	Correlated
		1,2-Dichlorobenzene (o-Dichlorobenzene) - average	W_ODCB_ln (ug/L)	Correlated
		Vinyl chloride - average	W_VCM_ln (ug/L)	Correlated

Table C: continued

Domain	Data Source	Variable	Variable Name	Reason Not Used
		Carbon Tetrachloride - average	W_CCl4_ln (ug/L)	Correlated
		1,1,2-Trichloroethane - average	W_112TCA_ln (ug/L)	Correlated
		1,1-Dichloroethylene - average	W_11DCE_ln (ug/L)	Correlated
		trans-1,2-Dichloroethylene - average	W_t12DCE_ln (ug/L)	Correlated
		1,2-Dichloroethane (Ethylene Dichloride) - average	W_EDC_ln (ug/L)	Correlated
		1,2-Dichloropropane - average	W_PDC_ln (ug/L)	Correlated
		Benzene - average	W_Cl1benz_ln (ug/L)	Correlated
Air	National-Scale Air Toxics Assessment	2,4-toluene diisocyanate	A_TDI_ln	Correlated
		2-chloroacetophenone	A_2Clacephen_ln	Correlated
		2-nitropropane	A_2NP_ln	Correlated
		4-nitrophenol	A_PNP_ln	Correlated
		Acetonitrile	A_CH3CN_ln	Correlated
		Acetophenone	A_Acetophenone_ln	Correlated
		Acrolein	A_Aroclein_ln	Correlated
		Acrylonitrile	A_C3H3N_ln	Correlated
		Antimony compounds	A_Sb_ln	Correlated
		Biphenyl	A_biphenyl_ln	Correlated
		Bromoform	A_Bromoform_ln	Correlated
		Cadmium compounds	A_Cd_ln	Correlated
		Carbon disulfide	A_CS2_ln	Correlated
		Carbon sulfide	A_CS_ln	Correlated
		Cresol/cresylic acid	A_Cresol_ln	Correlated
		Cumene	A_Cumene_ln	Correlated
		Diesel engine emissions	A_Diesel_ln	Correlated
		Dimethyl formamide	A_DMF_ln	Correlated
		Dimethyl phthalates	A_Me2_phatalte_ln	Correlated
		Dimethyl sulfate	A_Me2SO4_ln	Correlated
		Epichlorohydrin	A_ECH_ln	Correlated
		Ethyl acrylate	A_Etacrylate_ln	Correlated
		Ethylene glycol	A_EGLY_ln	Correlated
		Ethylene oxide	A_EOx_ln	Correlated
		Ethylidene dichloride	A_EdCl2_ln	Correlated
		Hexachlorobenzene	A_HCB_ln	Correlated
		Hexachlorobutadiene	A_HCBD_ln	Correlated
		Hexachlorocyclopentadiene	A_HCCPD_ln	Correlated
		Hexane	A_Hexane_ln	Correlated
		Lead compounds	A_Pb_ln	Correlated
		Mercury compounds	A_Hg_ln	Correlated
		Methanol	A_MeOH_ln	Correlated
		Methyl isobutyl ketone	A_MIBK_ln	Correlated
		Methyl methacrylate	A_MMA_ln	Correlated
		Methyl chloride	A_MeCl_ln	Correlated
		Methylhydrazine	A_Mehydrazine_ln	Correlated
		MTBE	A_MTBE_ln	Correlated
		Nitrobenzene	A_nitrobenzene_ln	Correlated
		N,N-dimethylaniline	A_DMA_ln	Correlated
		o-toluidine	A_otoluidine_ln	Correlated
		PAH/POM	A_PAHPOM_ln	Correlated
		Pentachlorophenol	A_PCP_ln	Correlated

Table C: continued

Domain	Data Source	Variable	Variable Name	Reason Not Used
		Phosphorus	A_P_ln	Correlated
		Propylene oxide	A_ProO_ln	Correlated
		Selenium compounds	A_Se_ln	Correlated
		Styrene	A_Styrene_ln	Correlated
		Tetrachloroethylene	A_Cl4C2_ln	Correlated
		Toluene	A_Toluene_ln	Correlated
		Triethylamine	A_Et3N_ln	Correlated
		Vinyl acetate	A_VyAc_ln	Correlated
		Vinylidene chloride	A_11DCE_ln	Correlated

Appendix IV: Table of Highly Correlated Variables for Each Domain

Air Domain

Variable	Correlated Variable	Correlation Coefficient	Variable Used to Represent Group
1-1-1-trichloroethane	Methylene chloride	0.73	Methylene chloride
	1-4-dichlorobenzene	0.70	
Vinylidene chloride	Ethylbenzene	0.73	Ethylbenzene
	2-2-4-trimethylpentane	0.72	
	Carbon disulfide	0.80	
	Cumene	0.72	
	Diesel engine emissions	0.71	
	Ethylene glycol	0.75	
	Hexane	0.74	
	Methanol	0.75	
	Methyl isobutyl ketone	0.71	
	MTBE	0.71	
	Naphthalene	0.71	
	Toluene	0.71	
	Xylenes	0.74	
2-2-4-trimethylpentane	Ethylbenzene	0.95	Ethylbenzene
	Vinylidene chloride	0.72	
	4-4-methylenediphenyl diisocyanate	0.82	
	Acetophenone	0.75	
	Acrolein	0.74	
	Benzene	0.82	
	Biphenyl	0.76	
	1-3-butadiene	0.83	
	Tetrachloroethylene	0.72	
	Cresol cresylic acid	0.71	
	Cumene	0.85	
	Diesel engine emissions	0.88	
	Ethylene glycol	0.86	
	Triethylamine	0.76	
	Hexane	0.92	
	Mercury compounds	0.82	
	Dimethyl phthalate	0.72	
	Methanol	0.85	
	Methyl isobutyl ketone	0.83	
	Methyl methacrylate	0.75	
	MTBE	0.79	
	Naphthalene	0.88	
	Pahpom	0.77	
	4-nitrophenol	0.82	
	Propionaldehyde	0.73	
	Selenium compounds	0.76	
	Styrene	0.82	
	2-4-toluene diisocyanate	0.72	
	Toluene	0.88	
	Vinyl acetate	0.78	
	Xylenes	0.95	
2-chloroacetophenone	Benzyl chloride	0.71	Benzyl chloride
	Bromoform	0.95	
	Methylhydrazine	0.96	
2-nitropropane	Chloroprene	0.70	Chloroprene
	Allyl chloride	0.76	
	n-n-dimethylaniline	0.77	
	2-4-dinitrotoluene	0.74	
	Nitrobenzene	0.76	
	o-toluidine	0.72	

Air Domain Variable	Correlated Variable	Correlation Coefficient	Variable Used to Represent Group
4-4-methylenediphenyl diisocyanate	Ethylbenzene	0.83	Ethylbenzene
	2-2-4-trimethylpentane	0.82	
	Acetophenone	0.74	
	Acrolein	0.72	
	Benzene	0.73	
	Biphenyl	0.70	
	1-3-butadiene	0.76	
	Cumene	0.84	
	Diesel engine emissions	0.75	
	Ethylene glycol	0.86	
	Triethylamine	0.79	
	Hexane	0.82	
	Mercury compounds	0.76	
	Dimethyl phthalate	0.76	
	Methanol	0.83	
	Methyl isobutyl ketone	0.82	
	Methyl methacrylate	0.77	
	MTBE	0.74	
	Naphthalene	0.79	
	Pahpom	0.72	
	Phenol	0.71	
	4-nitrophenol	0.78	
	Selenium compounds	0.71	
	Styrene	0.79	
	2-4-toluene diisocyanate	0.75	
	Toluene	0.77	
	Vinyl acetate	0.80	
	Xylenes	0.84	
Acetophenone	Ethylbenzene	0.76	Ethylbenzene
	2-2-4-trimethylpentane	0.75	
	4-4-methylenediphenyl diisocyanate	0.74	
	Biphenyl	0.78	
	1-3-butadiene	0.72	
	Cresol cresylic acid	0.76	
	Cumene	0.78	
	Ethylene glycol	0.78	
	Triethylamine	0.71	
	Hexane	0.74	
	Mercury compounds	0.75	
	Methanol	0.78	
	Methyl isobutyl ketone	0.76	
	MTBE	0.74	
	Naphthalene	0.76	
	Pahpom	0.76	
	Phenol	0.73	
	4-nitrophenol	0.81	
	Selenium compounds	0.72	
	Toluene	0.70	
	Vinyl acetate	0.70	
	Xylenes	0.77	
Acrolein	Ethylbenzene	0.77	Ethylbenzene
	2-2-4-trimethylpentane	0.74	
	4-4-methylenediphenyl diisocyanate	0.72	
	1-3-butadiene	0.74	
	Cresol cresylic acid	0.81	
	Cumene	0.74	
	Ethylene glycol	0.76	
	Hexane	0.73	
	Methanol	0.76	
	Methyl isobutyl ketone	0.75	
	MTBE	0.73	
	Naphthalene	0.75	
	Pahpom	0.71	
	Propionaldehyde	0.75	
	Xylenes	0.77	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Allyl chloride	Chloroprene	0.90	Chloroprene
	2-nitropropane	0.76	
	Acetonitrile	0.81	
	n-n-dimethylaniline	0.96	
	Epichlorohydrin	0.85	
	Ethyl acrylate	0.78	
	Hexachlorobutadiene	0.73	
	Hexachlorocyclopentadiene	0.70	
	Nitrobenzene	0.96	
	o-toluidine	0.85	
	Propylene oxide	0.77	
	1-2-4-trichlorobenzene	0.78	
Arsenic compounds	Chromium compounds	0.80	Chromium compounds
	Cadmium compounds	0.80	
	Lead compounds	0.74	
Benzene	Ethylbenzene	0.85	Ethylbenzene
	2-2-4-trimethylpentane	0.82	
	4-4-methylenediphenyl diisocyanate	0.73	
	1-3-butadiene	0.90	
	Tetrachloroethylene	0.85	
	Cumene	0.77	
	Diesel engine emissions	0.76	
	Ethylene glycol	0.80	
	Hexane	0.81	
	Mercury compounds	0.71	
	Methanol	0.79	
	Methyl isobutyl ketone	0.74	
	MTBE	0.70	
	Naphthalene	0.80	
	4-nitrophenol	0.74	
	Styrene	0.70	
	Toluene	0.96	
	Xylenes	0.85	
Biphenyl	Ethylbenzene	0.75	Ethylbenzene
	2-2-4-trimethylpentane	0.76	
	4-4-methylenediphenyl diisocyanate	0.70	
	Acetophenone	0.78	
	1-3-butadiene	0.70	
	Cresol cresylic acid	0.74	
	Cumene	0.77	
	Ethylene glycol	0.77	
	Hexane	0.73	
	Mercury compounds	0.76	
	Methanol	0.77	
	Methyl isobutyl ketone	0.74	
	MTBE	0.71	
	Naphthalene	0.77	
	Pahpom	0.80	
	Phenol	0.74	
	4-nitrophenol	0.74	
	Selenium compounds	0.72	
	Toluene	0.71	
	Xylenes	0.76	
Bromoform	Benzyl chloride	0.70	Benzyl chloride
	Methylhydrazine	0.94	

Air Domain

Variable	Correlated Variable	Correlation Coefficient	Variable Used to Represent Group
1-3-butadiene	Ethylbenzene	0.84	Ethylbenzene
	2-2-4-trimethylpentane	0.83	
	4-4-methylenediphenyl diisocyanate	0.76	
	Acetophenone	0.72	
	Acrolein	0.74	
	Benzene	0.90	
	Biphenyl	0.70	
	Tetrachloroethylene	0.74	
	Cresol cresylic acid	0.71	
	Cumene	0.80	
	Diesel engine emissions	0.72	
	Ethylene glycol	0.83	
	Triethylamine	0.74	
	Hexane	0.81	
	Mercury compounds	0.76	
	Methanol	0.81	
	Methyl isobutyl ketone	0.79	
	Methyl methacrylate	0.71	
	MTBE	0.73	
	Naphthalene	0.81	
	Pahpom	0.72	
	4-nitrophenol	0.77	
	Selenium compounds	0.70	
	Styrene	0.73	
	2-4-toluene diisocyanate	0.70	
	Toluene	0.94	
	Vinyl acetate	0.73	
	Xylenes	0.84	
Acrylonitrile	Trichloroethylene	0.74	Trichloroethylene
Cadmium compounds	Chromium compounds	0.71	Chromium compounds
	Arsenic compounds	0.80	
Acetonitrile	Chloroprene	0.80	Chloroprene
	Allyl chloride	0.81	
	n-n-dimethylaniline	0.80	
	2-4-dinitrotoluene	0.75	
	Epichlorohydrin	0.76	
	Nitrobenzene	0.79	
	o-toluidine	0.75	
	Propylene oxide	0.77	
Tetrachloroethylene	Ethylbenzene	0.72	Ethylbenzene
	2-2-4-trimethylpentane	0.72	
	Benzene	0.85	
	1-3-butadiene	0.74	
	Naphthalene	0.73	
	Toluene	0.82	
	Xylenes	0.72	
Cresol cresylic acid	Ethylbenzene	0.77	Ethylbenzene
	2-2-4-trimethylpentane	0.71	
	Acetophenone	0.76	
	Acrolein	0.81	
	Biphenyl	0.74	
	1-3-butadiene	0.71	
	Cumene	0.73	
	Ethylene glycol	0.75	
	Triethylamine	0.71	
	Mercury compounds	0.73	
	Methanol	0.74	
	Methyl isobutyl ketone	0.75	
	Naphthalene	0.78	
	Pahpom	0.76	
	Phenol	0.75	
	Propionaldehyde	0.71	
	Xylenes	0.78	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Carbon disulfide	Ethylbenzene	0.72	Ethylbenzene
	Vinylidene chloride	0.80	
	Cumene	0.70	
	Ethylene glycol	0.74	
	Methanol	0.74	
	Methyl isobutyl ketone	0.73	
	Xylenes	0.72	
Cumene	Ethylbenzene	0.87	Ethylbenzene
	Vinylidene chloride	0.72	
	2-2-4-trimethylpentane	0.85	
	4-4-methylenediphenyl diisocyanate	0.84	
	Acetophenone	0.78	
	Acrolein	0.74	
	Benzene	0.77	
	Biphenyl	0.77	
	1-3-butadiene	0.80	
	Cresol cresylic acid	0.73	
	Carbon disulfide	0.70	
	Diesel engine emissions	0.77	
	Ethylene glycol	0.89	
	Triethylamine	0.82	
	Hexane	0.88	
	Mercury compounds	0.81	
	Dimethyl phthalate	0.74	
	Methanol	0.88	
	Methyl isobutyl ketone	0.86	
	Methyl methacrylate	0.76	
	MTBE	0.83	
	Naphthalene	0.84	
	Pahpom	0.79	
	Phenol	0.78	
	4-nitrophenol	0.81	
	Selenium compounds	0.76	
	Styrene	0.81	
	2-4-toluene diisocyanate	0.77	
	Toluene	0.81	
	Vinyl acetate	0.80	
	Xylenes	0.88	
1-4-dichlorobenzene	Methylene chloride	0.80	Methylene chloride
	1-1-1-trichloroethane	0.70	
Diesel engine emissions	Ethylbenzene	0.86	Ethylbenzene
	Vinylidene chloride	0.71	
	2-2-4-trimethylpentane	0.88	
	4-4-methylenediphenyl diisocyanate	0.75	
	Benzene	0.76	
	1-3-butadiene	0.72	
	Cumene	0.77	
	Ethylene glycol	0.78	
	Triethylamine	0.70	
	Hexane	0.85	
	Mercury compounds	0.75	
	Methanol	0.78	
	Methyl isobutyl ketone	0.74	
	MTBE	0.73	
	Naphthalene	0.78	
	4-nitrophenol	0.74	
	Selenium compounds	0.71	
	Styrene	0.74	
	2-4-toluene diisocyanate	0.71	
	Toluene	0.78	
	Vinyl acetate	0.72	
	Xylenes	0.85	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
n-n-dimethylaniline	Chloroprene	0.92	Chloroprene
	2-nitropropane	0.77	
	Allyl chloride	0.96	
	Acetonitrile	0.80	
	2-4-dinitrotoluene	0.92	
	Epichlorohydrin	0.86	
	Ethyl acrylate	0.77	
	Hexachlorobutadiene	0.72	
	Hexachlorocyclopentadiene	0.72	
	Nitrobenzene	0.95	
	o-toluidine	0.86	
	Propylene oxide	0.78	
	1-2-4-trichlorobenzene	0.78	
Dimethyl formamide	Ethyl chloride	0.71	Ethyl chloride
2-4-dinitrotoluene	Chloroprene	0.88	Chloroprene
	2-nitropropane	0.74	
	Allyl chloride	0.89	
	A_CH3CN	0.75	
	n-n-dimethylaniline	0.92	
	Epichlorohydrin	0.84	
	Ethyl acrylate	0.76	
	Hexachlorocyclopentadiene	0.70	
	Nitrobenzene	0.88	
	o-toluidine	0.86	
	Propylene oxide	0.70	
	1-2-4-trichlorobenzene	0.76	
Epichlorohydrin	Chloroprene	0.84	Chloroprene
	Allyl chloride	0.85	
	Acetonitrile	0.76	
	n-n-dimethylaniline	0.86	
	2-4-dinitrotoluene	0.84	
	Ethyl acrylate	0.77	
	Nitrobenzene	0.81	
	o-toluidine	0.80	
	Propylene oxide	0.75	
	1-2-4-trichlorobenzene	0.74	
Ethylidene dichloride	Vinyl chloride	0.82	Vinyl chloride

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Ethylene glycol	Ethylbenzene	0.88	Ethylbenzene
	Vinylidene chloride	0.75	
	2-2-4-trimethylpentane	0.86	
	4-4-methylenediphenyl diisocyanate	0.86	
	Acetophenone	0.78	
	Acrolein	0.76	
	Benzene	0.80	
	Biphenyl	0.77	
	1-3-butadiene	0.83	
	Cresol cresylic acid	0.75	
	Carbon disulfide	0.74	
	Cumene	0.89	
	Diesel engine emissions	0.78	
	Triethylamine	0.83	
	Hexane	0.87	
	Mercury compounds	0.84	
	Dimethyl phthalate	0.76	
	Methanol	0.93	
	Methyl isobutyl ketone	0.91	
	Methyl methacrylate	0.79	
	MTBE	0.81	
	Naphthalene	0.86	
	Pahpom	0.78	
	Phenol	0.75	
	4-nitrophenol	0.83	
	Propionaldehyde	0.73	
	Selenium compounds	0.78	
	Styrene	0.81	
	2-4-toluene diisocyanate	0.78	
	Toluene	0.84	
	Vinyl acetate	0.82	
	Xylenes	0.90	
Ethylene oxide	Ethylene dichloride	0.72	Ethylene dichloride
Triethylamine	Ethylbenzene	0.79	Ethylbenzene
	2-2-4-trimethylpentane	0.76	
	4-4-methylenediphenyl diisocyanate	0.79	
	Acetophenone	0.71	
	1-3-butadiene	0.74	
	Cresol cresylic acid	0.71	
	Cumene	0.82	
	Diesel engine emissions	0.70	
	Ethylene glycol	0.83	
	Hexane	0.79	
	Mercury compounds	0.75	
	Methanol	0.80	
	Methyl isobutyl ketone	0.81	
	Methyl methacrylate	0.72	
	MTBE	0.70	
	Naphthalene	0.80	
	Pahpom	0.70	
	4-nitrophenol	0.71	
	Styrene	0.73	
	2-4-toluene diisocyanate	0.74	
	Toluene	0.74	
	Vinyl acetate	0.77	
	Xylenes	0.81	
Ethyl acrylate	Chloroprene	0.80	Chloroprene
	Allyl chloride	0.78	
	n-n-dimethylaniline	0.77	
	2-4-dinitrotoluene	0.76	
	Epichlorohydrin	0.77	
	Nitrobenzene	0.75	
	o-toluidine	0.76	
Hexachlorobenzene	Polychlorinated biphenyls	0.83	Polychlorinated biphenyls

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Hexachlorobutadiene	Chloroprene	0.70	Chloroprene
	Allyl chloride	0.73	
	n-n-dimethylaniline	0.72	
	Hexachlorocyclopentadiene	0.93	
	Nitrobenzene	0.73	
Hexachlorocyclopentadiene	Chloroprene	0.71	Chloroprene
	Allyl chloride	0.70	
	n-n-dimethylaniline	0.72	
	2-4-dinitrotoluene	0.70	
	Hexachlorobutadiene	0.93	
Hexane	Ethylbenzene	0.92	Ethylbenzene
	Vinylidene chloride	0.74	
	2-2-4-trimethylpentane	0.92	
	4-4-methylenediphenyl diisocyanate	0.82	
	Acetophenone	0.74	
	Acrolein	0.73	
	Benzene	0.81	
	Biphenyl	0.73	
	1-3-butadiene	0.81	
	Cumene	0.88	
	Diesel engine emissions	0.85	
	Ethylene glycol	0.87	
	Triethylamine	0.79	
	Mercury compounds	0.80	
	Dimethyl phthalate	0.72	
	Methanol	0.87	
	Methyl isobutyl ketone	0.83	
	Methyl methacrylate	0.76	
	MTBE	0.81	
	Naphthalene	0.86	
	Pahpom	0.73	
	4-nitrophenol	0.79	
	Selenium compounds	0.72	
	Styrene	0.80	
	2-4-toluene diisocyanate	0.77	
	Toluene	0.85	
	Vinyl acetate	0.79	
	Xylenes	0.92	
Hydrogen fluoride	Hydrochloric acid	0.91	Hydrochloric acid
Mercury compounds	Ethylbenzene	0.82	Ethylbenzene
	2-2-4-trimethylpentane	0.82	
	4-4-methylenediphenyl diisocyanate	0.76	
	Acetophenone	0.75	
	Benzene	0.71	
	Biphenyl	0.76	
	1-3-butadiene	0.76	
	Cresol cresylic acid	0.73	
	Cumene	0.81	
	Diesel engine emissions	0.75	
	Ethylene glycol	0.84	
	Triethylamine	0.75	
	Hexane	0.80	
	Methanol	0.82	
	Methyl isobutyl ketone	0.81	
	Methyl methacrylate	0.72	
	MTBE	0.74	
	Naphthalene	0.84	
	Pahpom	0.75	
	Phenol	0.72	
	4-nitrophenol	0.80	
	Propionaldehyde	0.73	
	Selenium compounds	0.91	
	Styrene	0.74	
	Toluene	0.76	
	Vinyl acetate	0.76	
	Xylenes	0.82	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Dimethyl phthalate	Ethylbenzene	0.73	Ethylbenzene
	2-2-4-trimethylpentane	0.72	
	4-4-methylenediphenyl diisocyanate	0.76	
	Cumene	0.74	
	Ethylene glycol	0.76	
	Hexane	0.72	
	Methanol	0.74	
	Methyl isobutyl ketone	0.73	
	Methyl methacrylate	0.76	
	Naphthalene	0.71	
	Styrene	0.75	
	Xylenes	0.74	
Dimethyl sulfate	Benzyl chloride	0.90	Benzyl chloride
Methyl chloride	Carbon tetrachloride	0.94	Carbon tetrachloride
Methylhydrazine	Benzyl chloride	0.71	Benzyl chloride
	2-chloroacetophenone	0.96	
	Bromoform	0.94	
Methanol	Ethylbenzene	0.88	Ethylbenzene
	Vinylidene chloride	0.75	
	2-2-4-trimethylpentane	0.85	
	4-4-methylenediphenyl diisocyanate	0.83	
	Acetophenone	0.78	
	Acrolein	0.76	
	Benzene	0.79	
	Biphenyl	0.77	
	1-3-butadiene	0.81	
	Cresol cresylic acid	0.74	
	Carbon disulfide	0.74	
	Cumene	0.88	
	Diesel engine emissions	0.78	
	Ethylene glycol	0.93	
	Triethylamine	0.80	
	Hexane	0.87	
	Mercury compounds	0.82	
	Dimethyl phthalate	0.74	
	Methyl isobutyl ketone	0.89	
	Methyl methacrylate	0.78	
	MTBE	0.82	
	Naphthalene	0.84	
	Pahpom	0.78	
	Phenol	0.76	
	4-nitrophenol	0.82	
	Propionaldehyde	0.72	
	Selenium compounds	0.77	
	Styrene	0.81	
	2-4-toluene diisocyanate	0.76	
	Toluene	0.82	
	Vinyl acetate	0.79	
	Xylenes	0.89	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Methyl isobutyl ketone	Ethylbenzene	0.86	Ethylbenzene
	Vinylidene chloride	0.71	
	2-2-4-trimethylpentane	0.83	
	4-4-methylenediphenyl diisocyanate	0.82	
	Acetophenone	0.76	
	Acrolein	0.75	
	Benzene	0.74	
	Biphenyl	0.74	
	1-3-butadiene	0.79	
	Cresol cresylic acid	0.75	
	Carbon disulfide	0.73	
	Cumene	0.86	
	Diesel engine emissions	0.74	
	Ethylene glycol	0.91	
	Triethylamine	0.81	
	Hexane	0.83	
	Mercury compounds	0.81	
	Dimethyl phthalate	0.73	
	Methanol	0.89	
	Methyl methacrylate	0.77	
	MTBE	0.81	
	Naphthalene	0.82	
	Pahpom	0.77	
	Phenol	0.78	
	4-nitrophenol	0.79	
	Selenium compounds	0.76	
	Styrene	0.81	
	2-4-toluene diisocyanate	0.77	
	Toluene	0.79	
	Vinyl acetate	0.76	
	Xylenes	0.89	
Methyl methacrylate	Ethylbenzene	0.77	Ethylbenzene
	2-2-4-trimethylpentane	0.75	
	4-4-methylenediphenyl diisocyanate	0.77	
	1-3-butadiene	0.71	
	Cumene	0.76	
	Ethylene glycol	0.79	
	Triethylamine	0.72	
	Hexane	0.76	
	Mercury compounds	0.72	
	Dimethyl phthalate	0.76	
	Methanol	0.78	
	Methyl isobutyl ketone	0.77	
	Naphthalene	0.74	
	4-nitrophenol	0.72	
	Styrene	0.83	
	Toluene	0.71	
	Vinyl acetate	0.72	
	Xylenes	0.78	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
MTBE	Ethylbenzene	0.79	Ethylbenzene
	Vinylidene chloride	0.71	
	2-2-4-trimethylpentane	0.79	
	4-4-methylenediphenyl diisocyanate	0.74	
	Acetophenone	0.74	
	Acrolein	0.73	
	Benzene	0.70	
	Biphenyl	0.71	
	1-3-butadiene	0.73	
	Cumene	0.83	
	Diesel engine emissions	0.73	
	Ethylene glycol	0.81	
	Triethylamine	0.70	
	Hexane	0.81	
	Mercury compounds	0.74	
	Methanol	0.82	
	Methyl isobutyl ketone	0.81	
	Naphthalene	0.78	
	Pahpom	0.71	
	Phenol	0.71	
	4-nitrophenol	0.74	
	Selenium compounds	0.70	
	Styrene	0.72	
	2-4-toluene diisocyanate	0.73	
	Toluene	0.73	
	Xylenes	0.79	
Naphthalene	Ethylbenzene	0.87	Ethylbenzene
	Vinylidene chloride	0.71	
	2-2-4-trimethylpentane	0.88	
	4-4-methylenediphenyl diisocyanate	0.79	
	Acetophenone	0.76	
	Acrolein	0.75	
	Benzene	0.80	
	Biphenyl	0.77	
	1-3-butadiene	0.81	
	Tetrachloroethylene	0.73	
	Cresol cresylic acid	0.78	
	Cumene	0.84	
	Diesel engine emissions	0.78	
	Ethylene glycol	0.86	
	Triethylamine	0.80	
	Hexane	0.86	
	Mercury compounds	0.84	
	Dimethyl phthalate	0.71	
	Methanol	0.84	
	Methyl isobutyl ketone	0.82	
	Methyl methacrylate	0.74	
	MTBE	0.78	
	Pahpom	0.84	
	Phenol	0.73	
	4-nitrophenol	0.79	
	Propionaldehyde	0.74	
	Selenium compounds	0.77	
	Styrene	0.76	
	2-4-toluene diisocyanate	0.70	
	Toluene	0.83	
	Vinyl acetate	0.78	
	Xylenes	0.88	
Nickel compounds	Chromium compounds	0.79	Chromium compounds

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Nitrobenzene	Chloroprene	0.88	Chloroprene
	2-nitropropane	0.76	
	Allyl chloride	0.96	
	Acetonitrile	0.79	
	n-n-dimethylaniline	0.95	
	2-4-dinitrotoluene	0.88	
	Epichlorohydrin	0.81	
	Ethyl acrylate	0.75	
	Hexachlorobutadiene	0.70	
	o-toluidine	0.82	
	Propylene oxide	0.77	
	1-2-4-trichlorobenzene	0.76	
o-toluidine	Chloroprene	0.84	Chloroprene
	2-nitropropane	0.72	
	Allyl chloride	0.85	
	Acetonitrile	0.75	
	n-n-dimethylaniline	0.86	
	2-4-dinitrotoluene	0.86	
	Epichlorohydrin	0.80	
	Ethyl acrylate	0.76	
	Nitrobenzene	0.82	
	Propylene oxide	0.77	
	1-2-4-trichlorobenzene	0.76	
Pahpom	Ethylbenzene	0.76	Ethylbenzene
	2-2-4-trimethylpentane	0.77	
	4-4-methylenediphenyl diisocyanate	0.72	
	Acetophenone	0.76	
	Acrolein	0.71	
	Biphenyl	0.80	
	1-3-butadiene	0.72	
	Cresol cresylic acid	0.76	
	Cumene	0.79	
	Ethylene glycol	0.78	
	Triethylamine	0.70	
	Hexane	0.73	
	Mercury compounds	0.75	
	Methanol	0.78	
	Methyl isobutyl ketone	0.77	
	MTBE	0.71	
	Naphthalene	0.84	
	Phenol	0.79	
	4-nitrophenol	0.76	
	Selenium compounds	0.72	
	Styrene	0.73	
	Xylenes	0.78	
Lead compounds	Chromium compounds	0.74	Chromium compounds
	Arsenic compounds	0.74	
Phenol	Ethylbenzene	0.71	Ethylbenzene
	4-4-methylenediphenyl diisocyanate	0.71	
	Acetophenone	0.73	
	Biphenyl	0.74	
	Cresol cresylic acid	0.75	
	Cumene	0.78	
	Ethylene glycol	0.75	
	Mercury compounds	0.72	
	Methanol	0.76	
	Methyl isobutyl ketone	0.78	
	MTBE	0.71	
	Naphthalene	0.73	
	Pahpom	0.79	
	Styrene	0.74	
	Xylenes	0.72	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
4-nitrophenol	Ethylbenzene	0.81	Ethylbenzene
	2-2-4-trimethylpentane	0.82	
	4-4-methylenediphenyl diisocyanate	0.78	
	Acetophenone	0.81	
	Benzene	0.74	
	Biphenyl	0.74	
	1-3-butadiene	0.77	
	Cumene	0.81	
	Diesel engine emissions	0.74	
	Ethylene glycol	0.83	
	Triethylamine	0.71	
	Hexane	0.79	
	Mercury compounds	0.80	
	Methanol	0.82	
	Methyl isobutyl ketone	0.79	
	Methyl methacrylate	0.72	
	MTBE	0.74	
	Naphthalene	0.79	
	Pahpom	0.76	
	Propionaldehyde	0.71	
	Selenium compounds	0.75	
	Styrene	0.75	
	2-4-toluene diisocyanate	0.70	
	Toluene	0.77	
	Vinyl acetate	0.73	
	Xylenes	0.81	
Propylene oxide	Chloroprene	0.75	Chloroprene
	Allyl chloride	0.77	
	Acetonitrile	0.77	
	n-n-dimethylaniline	0.78	
	2-4-dinitrotoluene	0.70	
	Epichlorohydrin	0.75	
	Nitrobenzene	0.77	
	o-toluidine	0.73	
Propionaldehyde	Ethylbenzene	0.74	Ethylbenzene
	2-2-4-trimethylpentane	0.73	
	Acrolein	0.75	
	Cresol cresylic acid	0.71	
	Ethylene glycol	0.73	
	Mercury compounds	0.73	
	Methanol	0.72	
	Naphthalene	0.74	
	4-nitrophenol	0.71	
	Selenium compounds	0.70	
	Xylenes	0.73	
Selenium compounds	Ethylbenzene	0.76	Ethylbenzene
	2-2-4-trimethylpentane	0.76	
	4-4-methylenediphenyl diisocyanate	0.71	
	Acetophenone	0.72	
	Biphenyl	0.72	
	1-3-butadiene	0.70	
	Cumene	0.76	
	Diesel engine emissions	0.71	
	Ethylene glycol	0.78	
	Hexane	0.72	
	Mercury compounds	0.91	
	Methanol	0.77	
	Methyl isobutyl ketone	0.76	
	MTBE	0.70	
	Naphthalene	0.77	
	Pahpom	0.72	
	4-nitrophenol	0.75	
	Propionaldehyde	0.70	
	Xylenes	0.77	

Air Domain Variable	Correlated Variable	Correlation Coefficient	Variable Used to Represent Group
Styrene	Ethylbenzene	0.82	Ethylbenzene
	2-2-4-trimethylpentane	0.82	
	4-4-methylenediphenyl diisocyanate	0.79	
	Benzene	0.70	
	1-3-butadiene	0.73	
	Cumene	0.81	
	Diesel engine emissions	0.74	
	Ethylene glycol	0.81	
	Triethylamine	0.73	
	Hexane	0.80	
	Mercury compounds	0.74	
	Dimethyl phthalate	0.75	
	Methanol	0.81	
	Methyl isobutyl ketone	0.81	
	Methyl methacrylate	0.83	
	MTBE	0.72	
	Naphthalene	0.76	
	Pahpom	0.73	
	Phenol	0.74	
	4-nitrophenol	0.75	
	Toluene	0.74	
	Vinyl acetate	0.73	
	Xylenes	0.83	
1-2-4-trichlorobenzene	Chloroprene	0.70	Chloroprene
	Allyl chloride	0.78	
	n-n-dimethylaniline	0.78	
	2-4-dinitrotoluene	0.76	
	Epichlorohydrin	0.74	
	Nitrobenzene	0.76	
	o-toluidine	0.74	
2-4-toluene diisocyanate	Ethylbenzene	0.77	Ethylbenzene
	2-2-4-trimethylpentane	0.72	
	4-4-methylenediphenyl diisocyanate	0.75	
	1-3-butadiene	0.70	
	Cumene	0.77	
	Diesel engine emissions	0.71	
	Ethylene glycol	0.78	
	Triethylamine	0.74	
	Hexane	0.77	
	Methanol	0.76	
	Methyl isobutyl ketone	0.77	
	MTBE	0.73	
	Naphthalene	0.70	
	4-nitrophenol	0.70	
	Toluene	0.71	
	Vinyl acetate	0.70	
	Xylenes	0.77	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Toluene	Ethylbenzene	0.88	Ethylbenzene
	Vinylidene chloride	0.71	
	2-2-4-trimethylpentane	0.88	
	4-4-methylenediphenyl diisocyanate	0.77	
	Acetophenone	0.70	
	Benzene	0.96	
	Biphenyl	0.71	
	1-3-butadiene	0.94	
	Tetrachloroethylene	0.82	
	Cumene	0.81	
	Diesel engine emissions	0.78	
	Ethylene glycol	0.84	
	Triethylamine	0.74	
	Hexane	0.85	
	Mercury compounds	0.76	
	Methanol	0.82	
	Methyl isobutyl ketone	0.79	
	Methyl methacrylate	0.71	
	MTBE	0.73	
	Naphthalene	0.83	
	4-nitrophenol	0.77	
	Styrene	0.74	
	2-4-toluene diisocyanate	0.71	
	Vinyl acetate	0.73	
	Xylenes	0.88	
Vinyl acetate	Ethylbenzene	0.79	Ethylbenzene
	2-2-4-trimethylpentane	0.78	
	4-4-methylenediphenyl diisocyanate	0.80	
	Acetophenone	0.70	
	1-3-butadiene	0.73	
	Cumene	0.80	
	Diesel engine emissions	0.72	
	Ethylene glycol	0.82	
	Triethylamine	0.77	
	Hexane	0.79	
	Mercury compounds	0.76	
	Methanol	0.79	
	Methyl isobutyl ketone	0.76	
	Methyl methacrylate	0.72	
	Naphthalene	0.78	
	4-nitrophenol	0.73	
	Styrene	0.73	
	2-4-toluene diisocyanate	0.70	
	Toluene	0.73	
	Xylenes	0.88	

Air Domain			
Variable	**Correlated Variable**	**Correlation Coefficient**	**Variable Used to Represent Group**
Xylenes	Ethylbenzene	0.99	Ethylbenzene
	Vinylidene chloride	0.74	
	2-2-4-trimethylpentane	0.95	
	4-4-methylenediphenyl diisocyanate	0.84	
	Acetophenone	0.77	
	Acrolein	0.77	
	Benzene	0.85	
	Biphenyl	0.76	
	1-3-butadiene	0.84	
	Tetrachloroethylene	0.72	
	Cresol cresylic acid	0.78	
	Carbon disulfide	0.72	
	Cumene	0.88	
	Diesel engine emissions	0.85	
	Ethylene glycol	0.90	
	Triethylamine	0.81	
	Hexane	0.92	
	Mercury compounds	0.82	
	Dimethyl phthalate	0.74	
	Methanol	0.89	
	Methyl isobutyl ketone	0.89	
	Methyl methacrylate	0.78	
	MTBE	0.79	
	Naphthalene	0.88	
	Pahpom	0.78	
	Phenol	0.72	
	4-nitrophenol	0.81	
	Propionaldehyde	0.73	
	Selenium compounds	0.77	
	Styrene	0.83	
	2-4-toluene diisocyanate	0.77	
	Toluene	0.88	
	Vinyl acetate	0.80	

Water Domain			
Variable	**Correlated Variable(s)**	**Correlation Coefficient**	**Variable Used To Represent Group**
Percent of county abnormally dry	Percent of county without drought, Percent of county drought - moderate, Percent of county drought - severe, Percent of county drought - extreme	0.94 0.94 0.86 0.71	Percent of county drought - extreme
Percent of county drought - moderate	Percent of county without drought, Percent of county abnormally dry, Percent of county drought - severe, Percent of county drought - extreme	0.94 0.94 0.86 0.71	Percent of county drought - extreme
Percent of county drought - severe	Percent of county without drought, Percent of county abnormally dry, Percent of county drought - moderate, Percent of county drought - extreme	0.86 0.86 0.94 0.71	Percent of county drought - extreme
Percent of county drought - exceptional	Percent of county drought - moderate, Percent of county drought - severe, Percent of county drought - extreme	0.94 0.86 0.80	Percent of county drought - extreme
Lindane - average	Barium - average	0.75	Barium - average
Thallium - average	Cadmium - average	0.76	Cadmium - average
Toxaphene - average	Endrin - average	0.80	Endrin - average
Oxamyl (Vydate) – average	Dalapon - average	0.70	Dalapon - average
Alachlor - average	Simazine - average	0.72	Simazine - average
2,4,5-TP (Silvex) - average	Picloram - average	0.73	Picloram - average
Hexachlorocyclopentadiene - average	Ethylene dibromide (EDB) - average	0.80	Ethylene dibromide (EDB) - average
Carbofuran - average	Chlordane - average	0.79	Chlordane - average

Water Domain

Variable	Correlated Variable(s)	Correlation Coefficient	Variable Used To Represent Group
Heptachlor - average	Di(2-ethylhexyl) phthalate (DEHP) - average Hexachlorobenzene - average Heptachlor - average	0.77 0.70 0.81	Di(2-ethylhexyl) phthalate (DEHP) - average
Heptachlor Epoxide - average	Di(2-ethylhexyl) phthalate (DEHP) - average Hexachlorobenzene - average Heptachlor - average	0.73 0.74 0.81	Di(2-ethylhexyl) phthalate (DEHP) - average
Hexachlorobenzene - average	Di(2-ethylhexyl) phthalate (DEHP) - average Heptachlor - average Heptachlor Epoxide - average	0.77 0.70 0.74	Di(2-ethylhexyl) phthalate (DEHP) - average
1,2,4-Trichlorobenzene - average	Ethylbenzene - average Vinyl chloride - average Benzene - average	0.77 0.71 0.82	Ethylbenzene - average
1,2-Dichlorobenzene (o-Dichlorobenzene) - average	1,2,4-Trichlorobenzene - detect Ethylbenzene - average Benzene - average	0.80 0.77 0.88	Ethylbenzene - average
Vinyl chloride - average	1,2-Dichlorobenzene (o-Dichlorobenzene) - average 1,2,4-Trichlorobenzene - detect Ethylbenzene - average Benzene - average	0.73 0.80 0.77 0.82	Ethylbenzene - average
Benzene - average	1,2-Dichlorobenzene (o-Dichlorobenzene) - average 1,2,4-Trichlorobenzene - detect Ethylbenzene - average Vinyl chloride - average	0.88 0.82 0.72 0.82	Ethylbenzene - average
1,1-Dichloroethylene - average	cis1,2-Dichloroethylene - average Dichloroethylene - average cis-1,2-Dichloroethylene - average	0.70 0.70 0.81	cis-1,2-Dichloroethylene - average
W_t12DCE_ln	cis-1,2-Dichloroethylene - average 1,1-Dichloroethylene - average cis-1,2-Dichloroethylene - average	0.82 0.70 0.75	cis-1,2-Dichloroethylene - average
cis-1,2-Dichloroethylene - average	cis-1,2-Dichloroethylene - average 1,1-Dichloroethylene - average Dichloroethylene - average	0.82 0.81 0.75	cis-1,2-Dichloroethylene - average
Carbon Tetrachloride - average	1,1,1-Trichloroethane - average	0.71	1,1,1-Trichloroethane - average
1,2-Dichloropropane - average	1,4-Dichlorobenzene (p-Dichlorobenzene) - average	0.72	1,4-Dichlorobenzene (p-Dichlorobenzene) - average
1,1,2-Trichloroethane - average	Tetrachloroethylene - average	0.80	Tetrachloroethylene - average

Land Domain

Variable	Correlated Variable(s)	Correlation Coefficient	Variable Used To Represent Group
Mean manganese	Mean iron percent	0.90	Mean iron percent
Percent weed acres	Percent harvested acres, percent lime acres	0.96 0.95	Percent harvested acres
Percent lime acres	Percent harvested acres, percent weed acres	0.97 0.95	Percent harvested acres

Sociodemographic Domain

Variable	Correlated Variable(s)	Correlation Coefficient	Variable Used To Represent Group
Property crime rate	Violent crime rate	0.91	Violent crime rate

Built Domain

Variable	Correlated Variable(s)	Correlation Coefficient	Variable Used To Represent Group
Secondary road proportion	Street proportion	0.94	Street proportion

Appendix V: Sociodemographic and Built-Domain Valence Correction

Sociodemographic Overall

	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Percent bachelor's degree	Beneficial	" - "	0.4585	No	Yes	0.2102	-0.4585
Percent unemployed	Harmful	" + "	-0.1269	No	Yes	0.0161	0.1269
Percent families less than poverty level	Harmful	" + "	-0.298	No	Yes	0.0888	0.298
Percent vacant housing	Harmful	" + "	-0.1979	No	Yes	0.0392	0.1979
Median household value	Beneficial	" - "	0.4331	No	Yes	0.1876	-0.4331
Household income	Beneficial	" - "	0.3824	No	Yes	0.1462	-0.3824
Count of occupants per room	Harmful	" + "	-0.1085	No	Yes	0.0118	0.1085
Percent renter-occupied housing	Harmful	" + "	0.1458	Yes	Yes	0.0213	-0.1458
Violent crime	Harmful	" + "	0.0234	Yes	Yes	0.0005	-0.0234
Percent creative class	Beneficial	" - "	0.4833	No	Yes	0.2336	-0.4833
Percent Democratic	Beneficial	" - "	0.211	No	Yes	0.0445	-0.211
GINI coefficient	Harmful	" + "	-0.0118	No	Yes	0.0001	0.0118

Sociodemographic RUCC 1

	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Percent bachelor's degree	Beneficial	" - "	0.4689	No	Yes	0.2199	-0.4689
Percent unemployed	Harmful	" + "	-0.1625	No	Yes	0.0264	0.1625
Percent families less than poverty level	Harmful	" + "	-0.2591	No	Yes	0.0671	0.2591
Percent vacant housing	Harmful	" + "	-0.2306	No	Yes	0.0532	0.2306
Median household value	Beneficial	" - "	0.4034	No	Yes	0.1627	-0.4034
Household income	Beneficial	" - "	0.3700	No	Yes	0.1369	-0.3700
Count of occupants per room	Harmful	" + "	-0.0055	No	Yes	0.0000	0.0055
Percent renter-occupied housing	Harmful	" + "	0.1827	Yes	Yes	0.0334	-0.1827
Violent crime	Harmful	" + "	0.0094	Yes	Yes	0.0001	-0.0094
Percent creative class	Beneficial	" - "	0.4668	No	Yes	0.2179	-0.4668
Percent Democratic	Beneficial	" - "	0.2625	No	Yes	0.0689	-0.2625
GINI coefficient	Harmful	" + "	0.1162	Yes	Yes	0.0135	-0.1162

Sociodemographic RUCC 2	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Percent bachelor's degree	Beneficial	" - "	0.4621	No	Yes	0.2136	-0.4621
Percent unemployed	Harmful	" + "	-0.3274	No	Yes	0.1072	0.3274
Percent families less than poverty level	Harmful	" + "	-0.4293	No	Yes	0.1843	0.4293
Percent vacant housing	Harmful	" + "	0.1331	Yes	Yes	0.0177	-0.1331
Median household value	Beneficial	" - "	0.4002	No	Yes	0.1602	-0.4002
Household income	Beneficial	" - "	0.0874	No	Yes	0.0076	-0.0874
Count of occupants per room	Harmful	" + "	-0.1371	No	Yes	0.0188	0.1371
Percent renter-occupied housing	Harmful	" + "	-0.0141	No	Yes	0.0002	0.0141
Violent crime	Harmful	" + "	-0.2386	No	Yes	0.0569	0.2386
Percent creative class	Beneficial	" - "	0.4463	No	Yes	0.1992	-0.4463
Percent Democratic	Beneficial	" - "	0.0929	No	Yes	0.0086	-0.0929
GINI coefficient	Harmful	" + "	-0.1604	No	Yes	0.0257	0.1604

Built (Overall)	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Vice-related environment	Harmful	" + "	0.2930	Yes	Yes	0.0858	-0.2930
Civic-related environment	Beneficial	" - "	0.3071	No	Yes	0.0943	-0.3071
Education-related environment	Beneficial	" - "	0.3495	No	Yes	0.1222	-0.3495
Health care-related environment	Beneficial	" - "	0.2798	No	Yes	0.0783	-0.2798
Negative food environment	Harmful	" + "	0.2280	Yes	Yes	0.0520	-0.2280
Positive food environment	Beneficial	" - "	0.3179	No	Yes	0.1011	-0.3179
Recreation environment	Beneficial	" - "	0.3590	No	Yes	0.1289	-0.3590
Social service-related environment	Beneficial	" - "	0.3629	No	Yes	0.1317	-0.3629
Traffic fatality rate	Harmful	" + "	-0.1751	No	Yes	0.0307	0.1751
Rate of low-rent + Section 8 housing	Harmful	" + "	0.0581	Yes	Yes	0.0034	-0.0581
Proportion of secondary roads	Harmful	" + "	-0.1777	No	Yes	0.0316	0.1777
Commute time	Harmful	" + "	-0.3329	No	Yes	0.1108	0.3329
Public transportation	Beneficial	" - "	0.0463	No	Yes	0.0021	-0.0463
Walkability score	Beneficial	" - "	0.1585	No	Yes	0.0251	-0.1585
Proportion green space	Beneficial	" - "	-0.0451	Yes	Yes	0.0020	0.0451

Built RUCC 1	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Vice-related environment	Harmful	" + "	0.2676	Yes	Yes	0.0716	-0.2676
Civic-related environment	Beneficial	" - "	0.1238	No	Yes	0.0153	-0.1238
Education-related environment	Beneficial	" - "	0.2409	No	Yes	0.0580	-0.2409
Health care-related environment	Beneficial	" - "	0.4189	No	Yes	0.1755	-0.4189

Built RUCC 1	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Negative food environment	Harmful	" + "	0.3239	Yes	Yes	0.1049	-0.3239
Positive food environment	Beneficial	" - "	0.3405	No	Yes	0.1159	-0.3405
Recreation environment	Beneficial	" - "	0.2354	No	Yes	0.0554	-0.2354
Social service-related environment	Beneficial	" - "	0.3446	No	Yes	0.1187	-0.3446
Traffic fatality rate	Harmful	" + "	0.1978	Yes	Yes	0.0391	-0.1978
Rate of low-rent + Section 8 housing	Harmful	" + "	-0.1230	No	Yes	0.0151	0.1230
Proportion of secondary roads	Harmful	" + "	0.0950	Yes	Yes	0.0090	-0.0950
Commute time	Harmful	" + "	-0.1886	No	Yes	0.0356	0.1886
Public transportation	Beneficial	" - "	0.2253	No	Yes	0.0508	-0.2253
Walkability score	Beneficial	" - "	0.3516	No	Yes	0.1236	-0.3516
Proportion green space	Beneficial	" - "	-0.1065	Yes		0.0113	0.1065

Built RUCC 2	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Vice-related environment	Harmful	" + "	0.0331	Yes	Yes	0.0331	-0.0331
Civic-related environment	Beneficial	" - "	0.2057	No	Yes	0.2057	-0.2057
Education-related environment	Beneficial	" - "	0.2626	No	Yes	0.2626	-0.2626
Health care-related environment	Beneficial	" - "	0.3856	No	Yes	0.3856	-0.3856
Negative food environment	Harmful	" + "	0.2707	Yes	Yes	0.2707	-0.2707
Positive food environment	Beneficial	" - "	0.2752	No	Yes	0.2752	-0.2752
Recreation environment	Beneficial	" - "	0.3484	No	Yes	0.3484	-0.3484
Social service-related environment	Beneficial	" - "	0.3503	No	Yes	0.3503	-0.3503
Traffic fatality rate	Harmful	" + "	-0.2340	No	Yes	-0.2340	0.2340
Rate of low-rent + Section 8 housing	Harmful	" + "	0.0459	Yes	Yes	0.0459	-0.0459
Proportion of secondary roads	Harmful	" + "	-0.1319	No	Yes	-0.1319	0.1319
Commute time	Harmful	" + "	-0.2808	No	Yes	-0.2808	0.2808
Public transportation	Beneficial	" - "	0.1111	No	Yes	0.1111	-0.1111
Walkability score	Beneficial	" - "	0.3310	No	Yes	0.3310	-0.3310
Proportion green space	Beneficial	" - "	0.0253	No	Yes	0.0253	-0.0253

Built RUCC 3

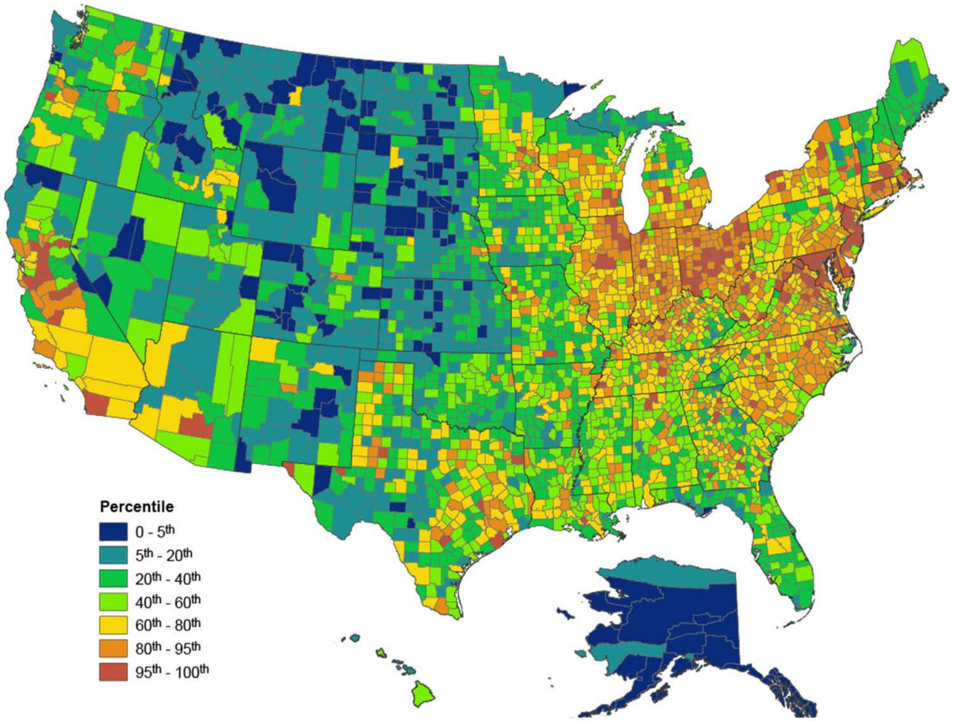
	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Vice-related environment	Harmful	" + "	0.2724	Yes	Yes	0.0742	-0.2724
Civic-related environment	Beneficial	" - "	0.1890	No	Yes	0.0357	-0.1890
Education-related environment	Beneficial	" - "	0.3278	No	Yes	0.1074	-0.3278
Health care-related environment	Beneficial	" - "	0.3179	No	Yes	0.1011	-0.3179
Negative food environment	Harmful	" + "	0.2306	Yes	Yes	0.0532	-0.2306
Positive food environment	Beneficial	" - "	0.2660	No	Yes	0.0707	-0.2660
Recreation environment	Beneficial	" - "	0.3212	No	Yes	0.1032	-0.3212
Social service-related environment	Beneficial	" - "	0.3644	No	Yes	0.1328	-0.3644
Traffic fatality rate	Harmful	" + "	-0.2197	No	Yes	0.0483	0.2197
Rate of low-rent + Section 8 housing	Harmful	" + "	0.0697	Yes		0.0049	-0.0697
Proportion of secondary roads	Harmful	" + "	-0.1761	No		0.0310	0.1761
Commute time	Harmful	" + "	-0.3230	No		0.1043	0.3230
Public transportation	Beneficial	" - "	0.0777	No		0.0060	-0.0777
Walkability score	Beneficial	" - "	0.3542	No		0.1255	-0.3542
Proportion green space	Beneficial	" - "	-0.0418	Yes		0.0017	0.0418

Built RUCC 4

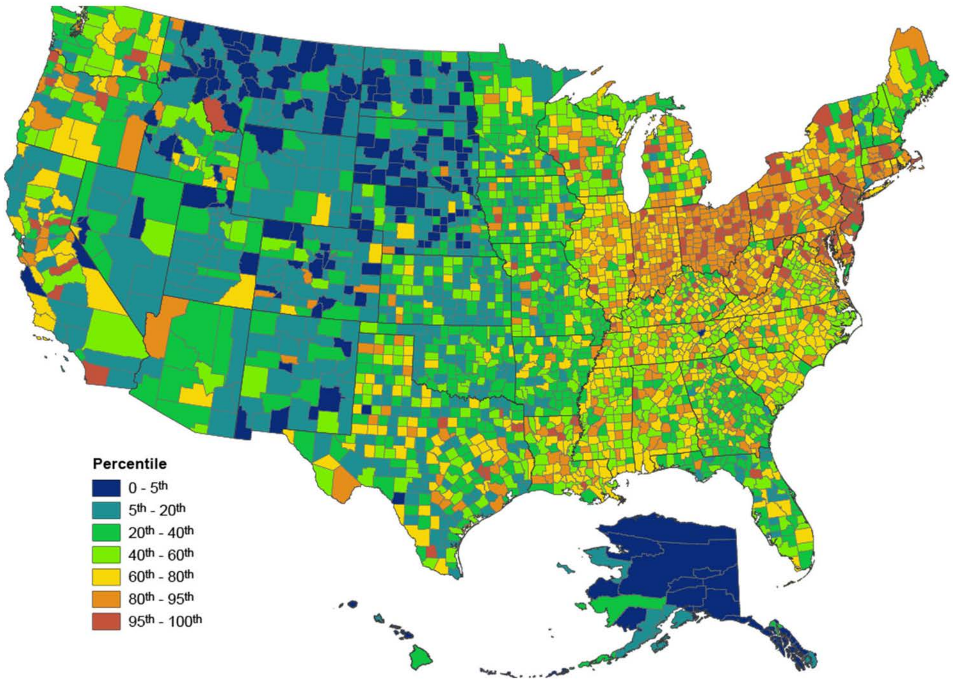
	A priori Variable Characteristic	Loading (Expected Sign)	Loading (Actual)	Match (Expected versus Observed)	Necessary To Multiply Vector of Loadings by -1?	(Loading)^2	Modified Loadings
Vice-related environment	Harmful	" + "	0.2595	Yes	Yes	0.0673	-0.2595
Civic-related environment	Beneficial	" - "	0.3102	No	Yes	0.0962	-0.3102
Education-related environment	Beneficial	" - "	0.3285	No	Yes	0.1079	-0.3285
Health care-related environment	Beneficial	" - "	0.2742	No	Yes	0.0752	-0.2742
Negative food environment	Harmful	" + "	0.1527	Yes	Yes	0.0233	-0.1527
Positive food environment	Beneficial	" - "	0.2524	No	Yes	0.0637	-0.2524
Recreation environment	Beneficial	" - "	0.3222	No	Yes	0.1038	-0.3222
Social service-related environment	Beneficial	" - "	0.2793	No	Yes	0.0780	-0.2793
Traffic fatality rate	Harmful	" + "	-0.2312	No	Yes	0.0535	0.2312
Rate of low-rent + Section 8 housing	Harmful	" + "	-0.0178	No	Yes	0.0003	0.0178
Proportion of secondary roads	Harmful	" + "	-0.2054	No	Yes	0.0422	0.2054
Commute time	Harmful	" + "	-0.3546	No	Yes	0.1257	0.3546
Public transportation	Beneficial	" - "	0.0256	No	Yes	0.0007	-0.0256
Walkability score	Beneficial	" - "	0.3787	No	Yes	0.1434	-0.3787
Proportion green space	Beneficial	" - "	-0.1370	Yes	Yes	0.0188	0.1370

Appendix VI: County Maps of Environmental Quality Index 2006- 2010

Overall Environmental Quality Index by County 2006-2010

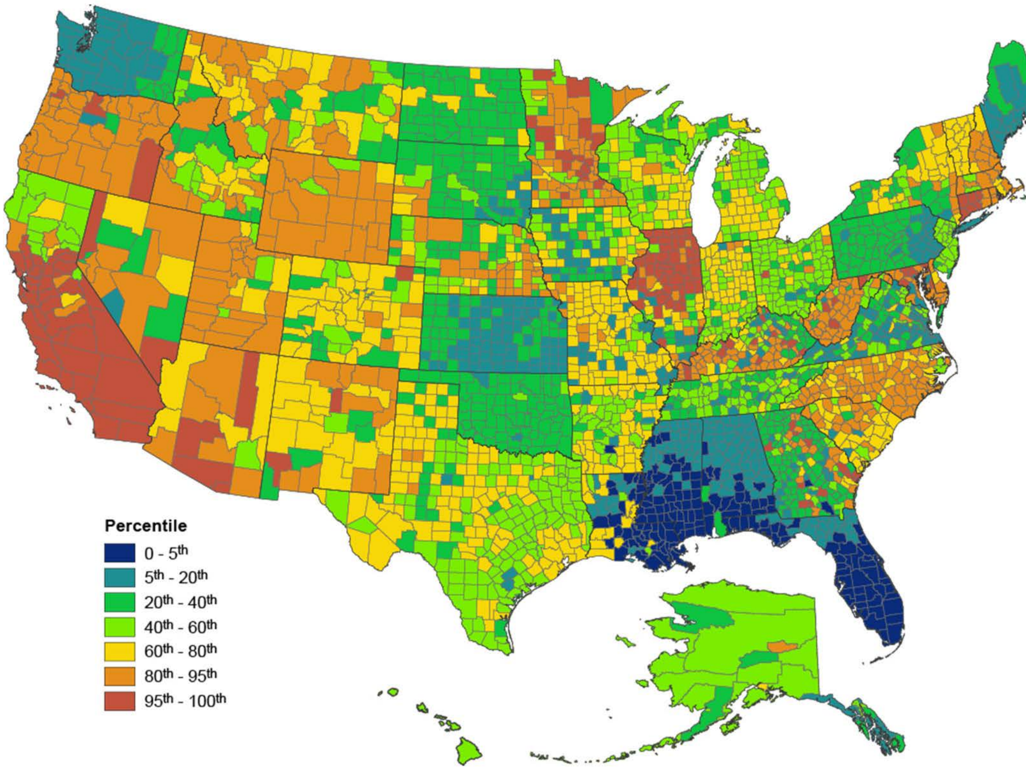


Air Domain Index by County 2006-2010

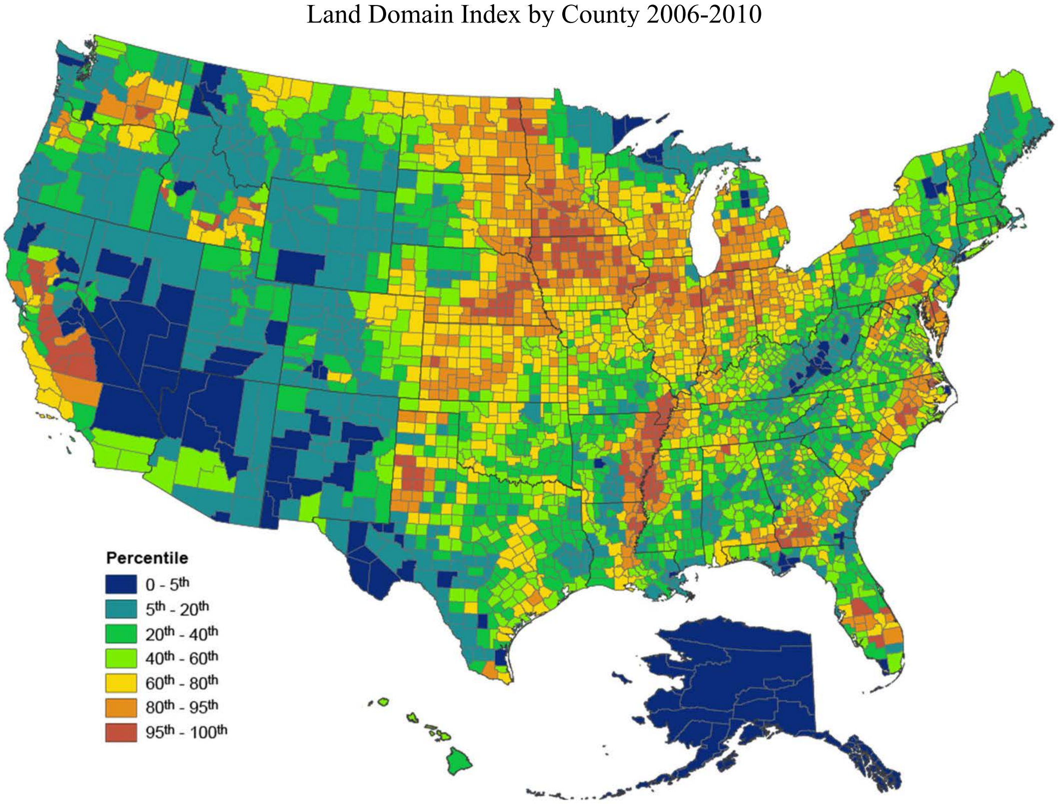


* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Water Domain Index by County 2006-2010

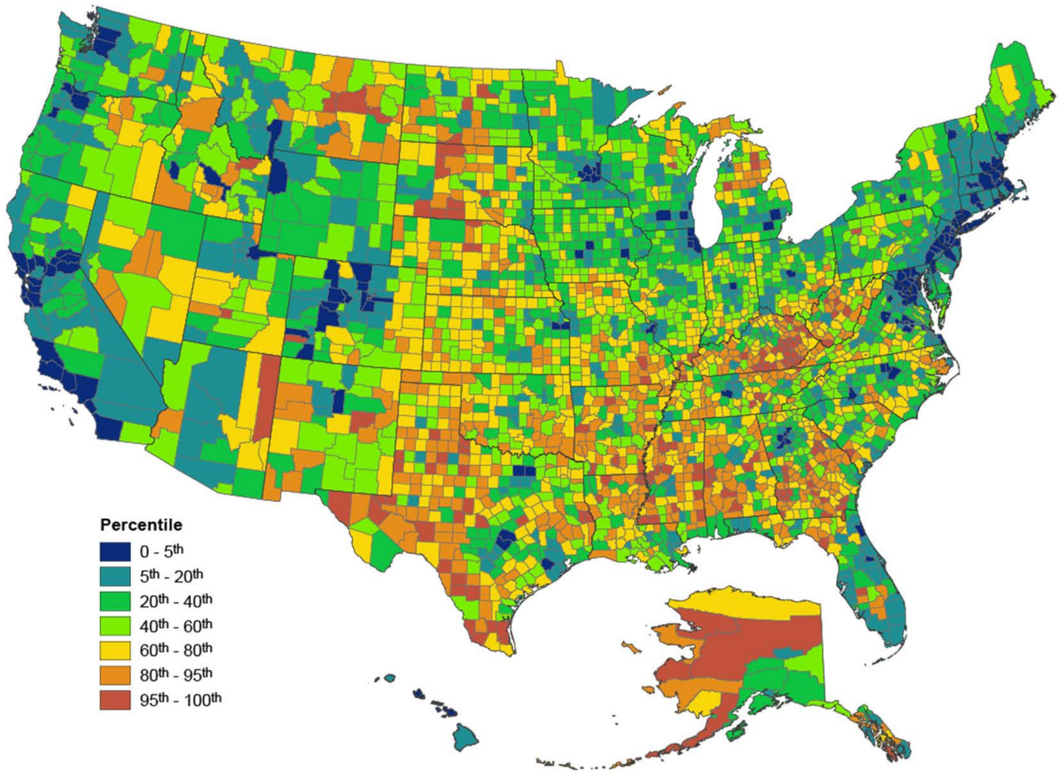


Land Domain Index by County 2006-2010

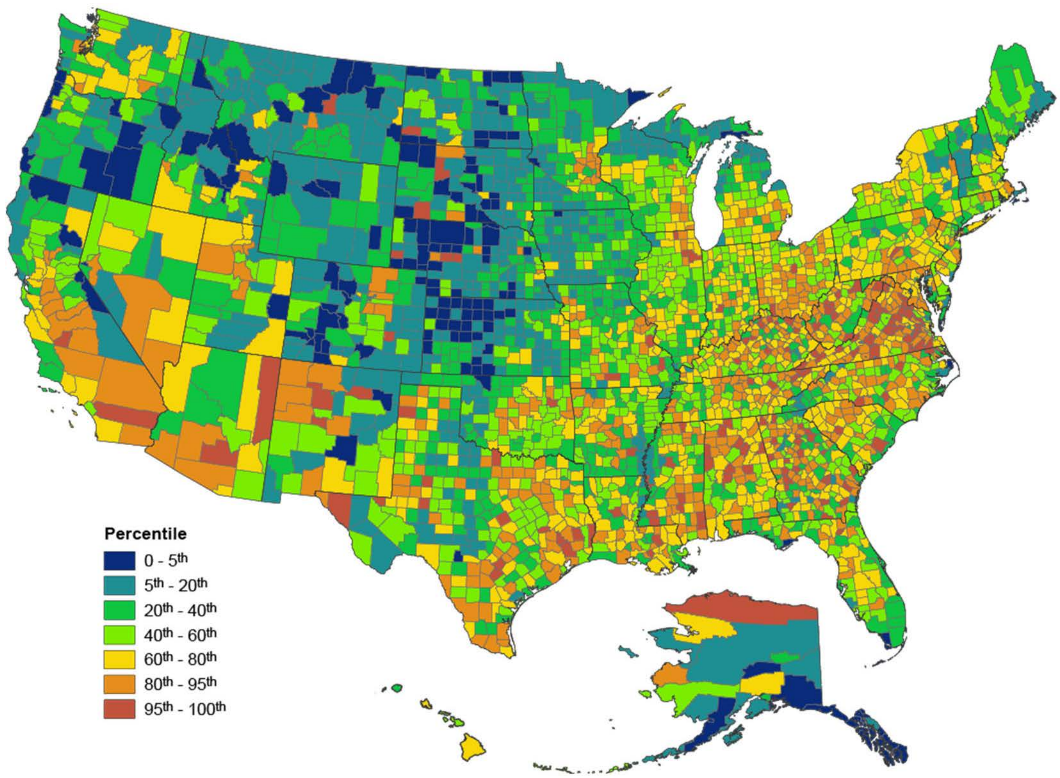


* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Sociodemographic Domain Index by County 2006-2010

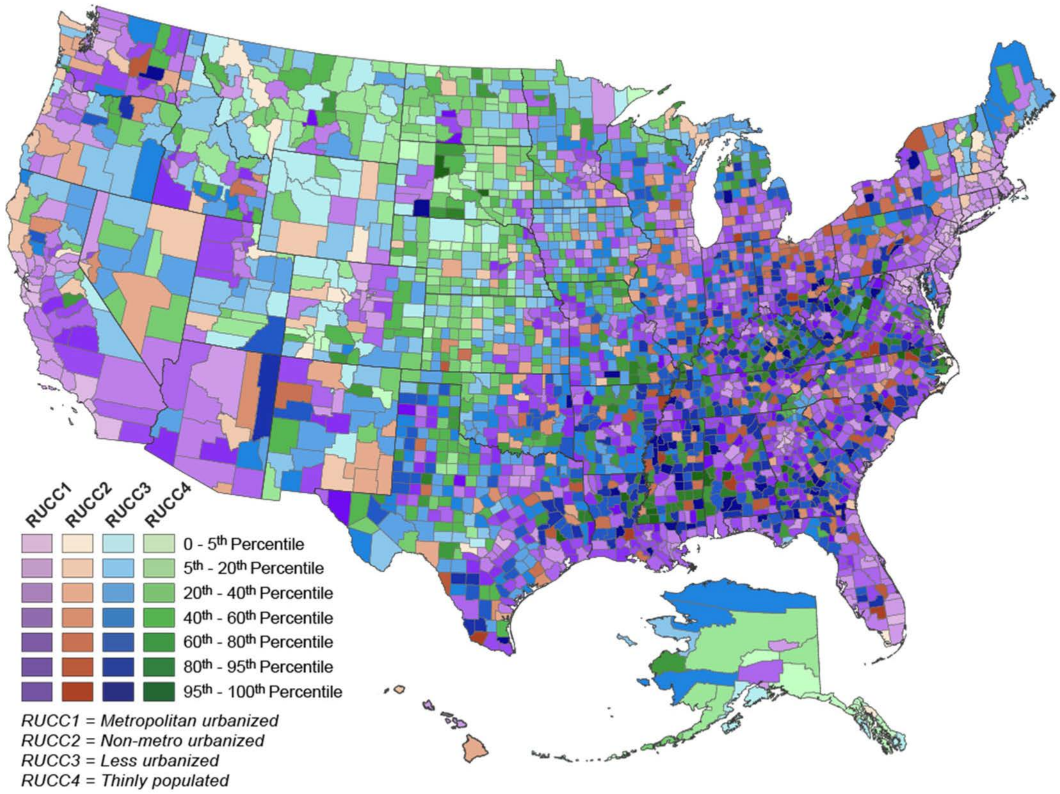


Built Domain Index by County 2006-2010

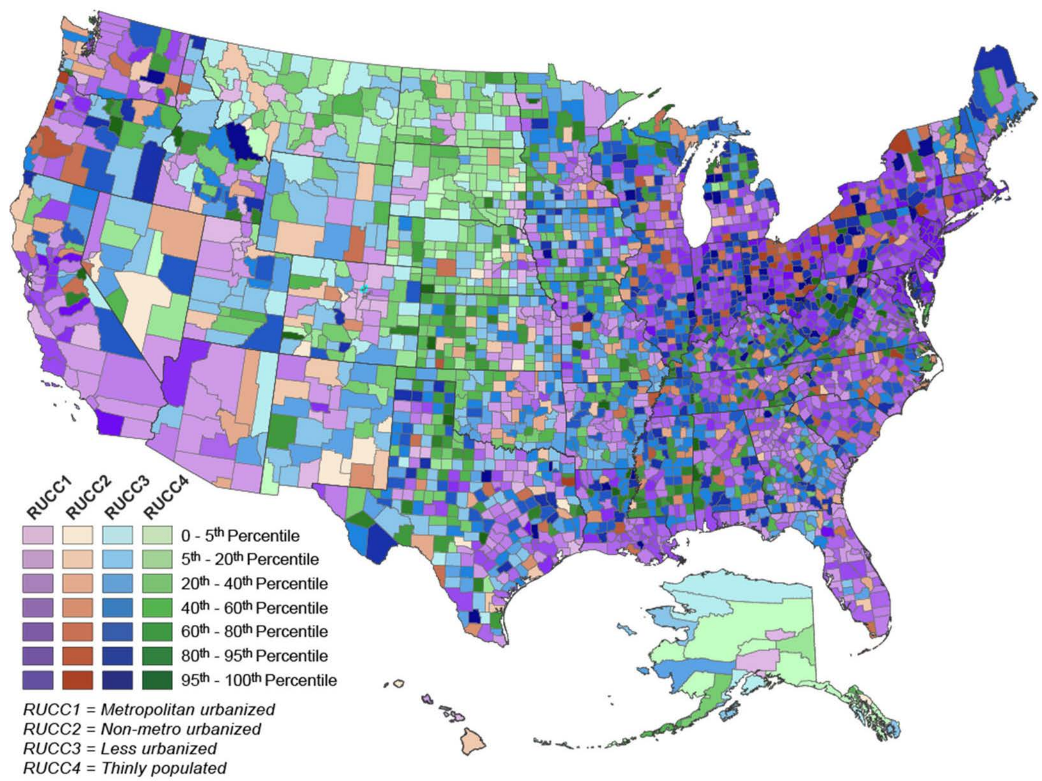


* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Overall Environmental Quality Index Stratified by Rural-Urban Continuum Codes by County 2006-2010

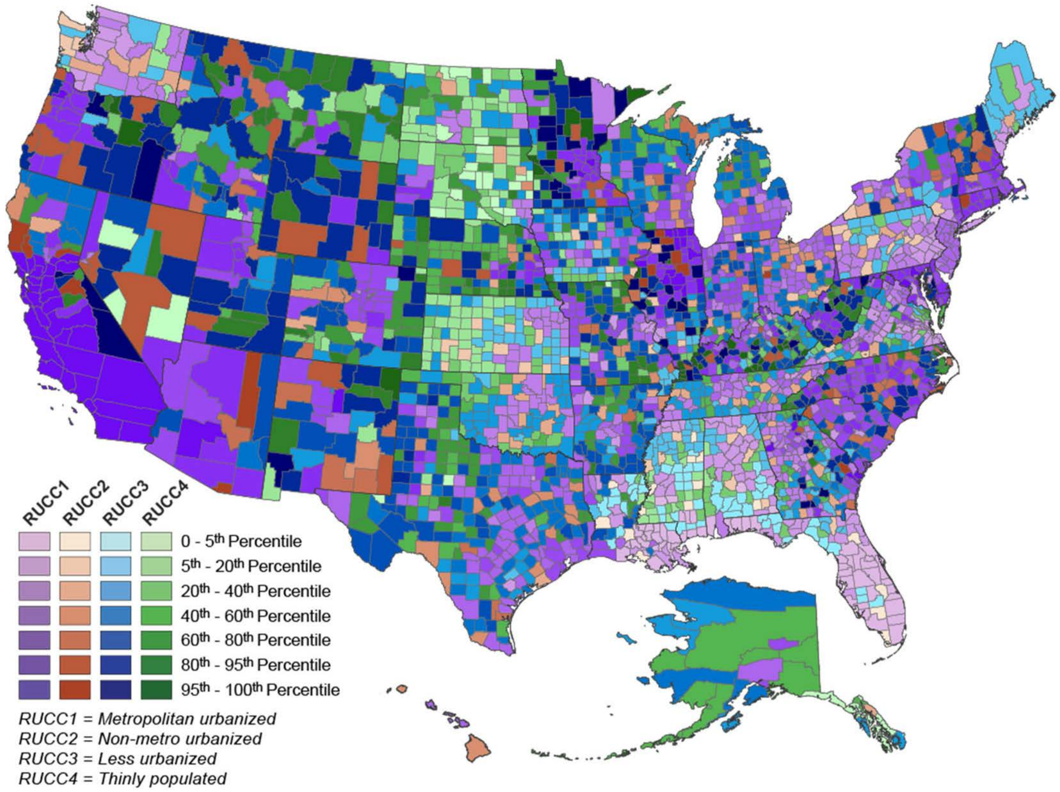


Built Domain Index Stratified by Rural-Urban Continuum Codes by County 2006-2010

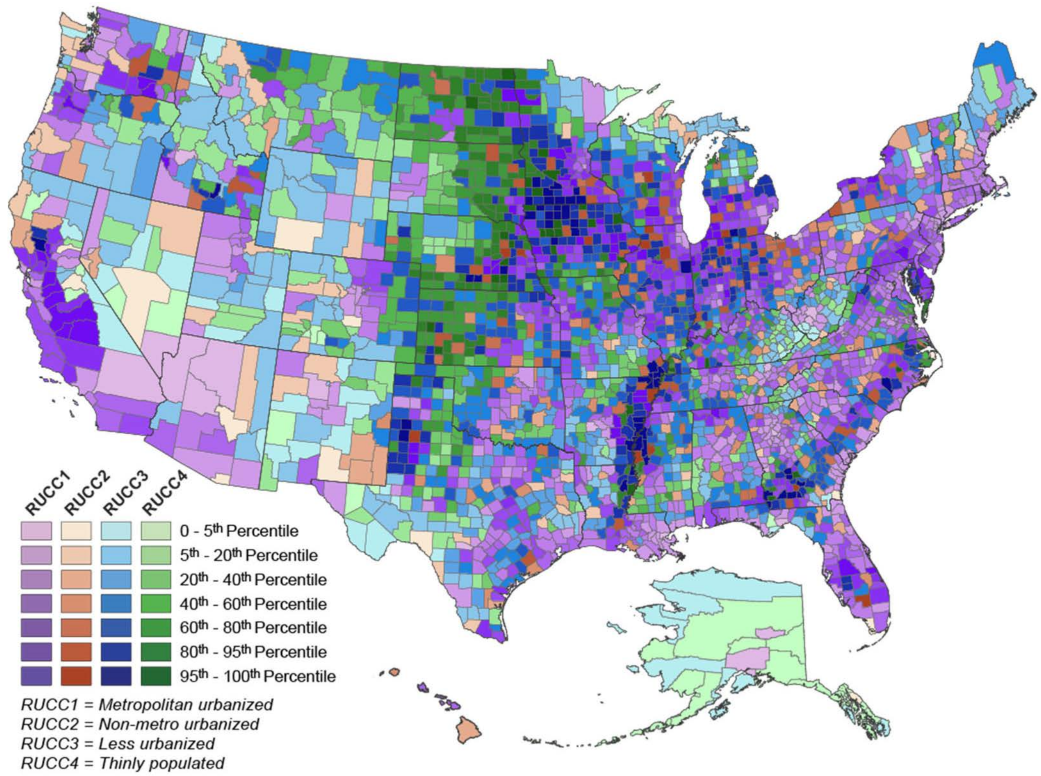


* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Water Domain Index Stratified by Rural-Urban Continuum Codes by County 2006-2010

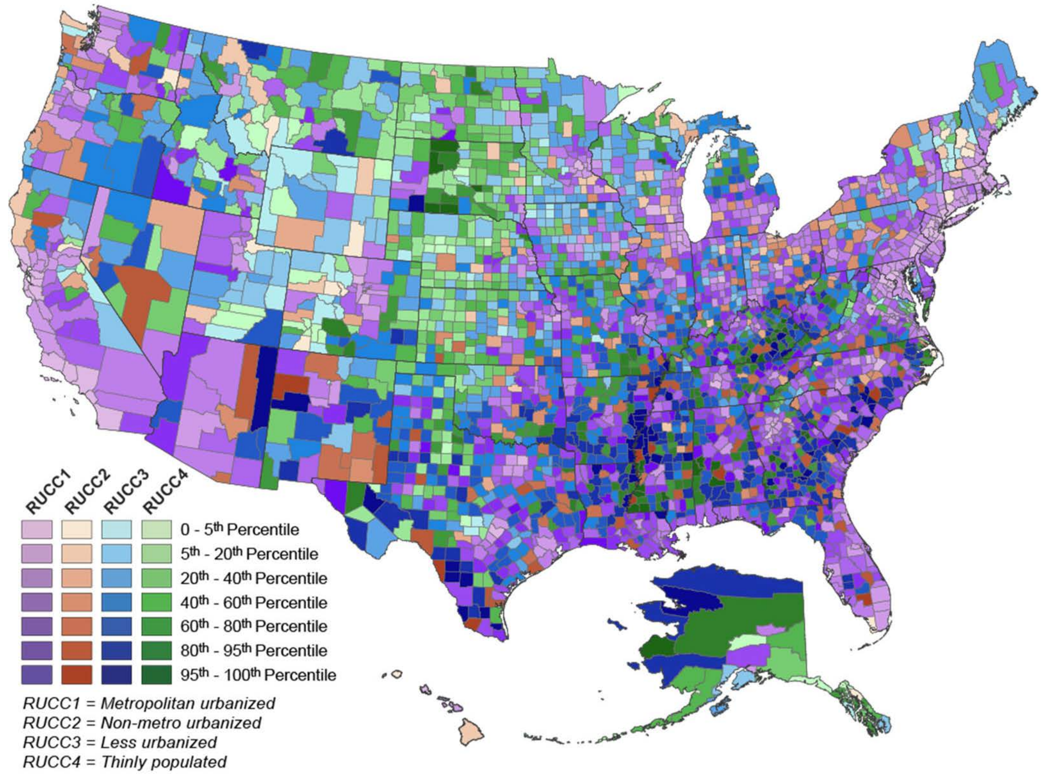


Land Domain Index Stratified by Rural-Urban Continuum Codes by County 2006-2010

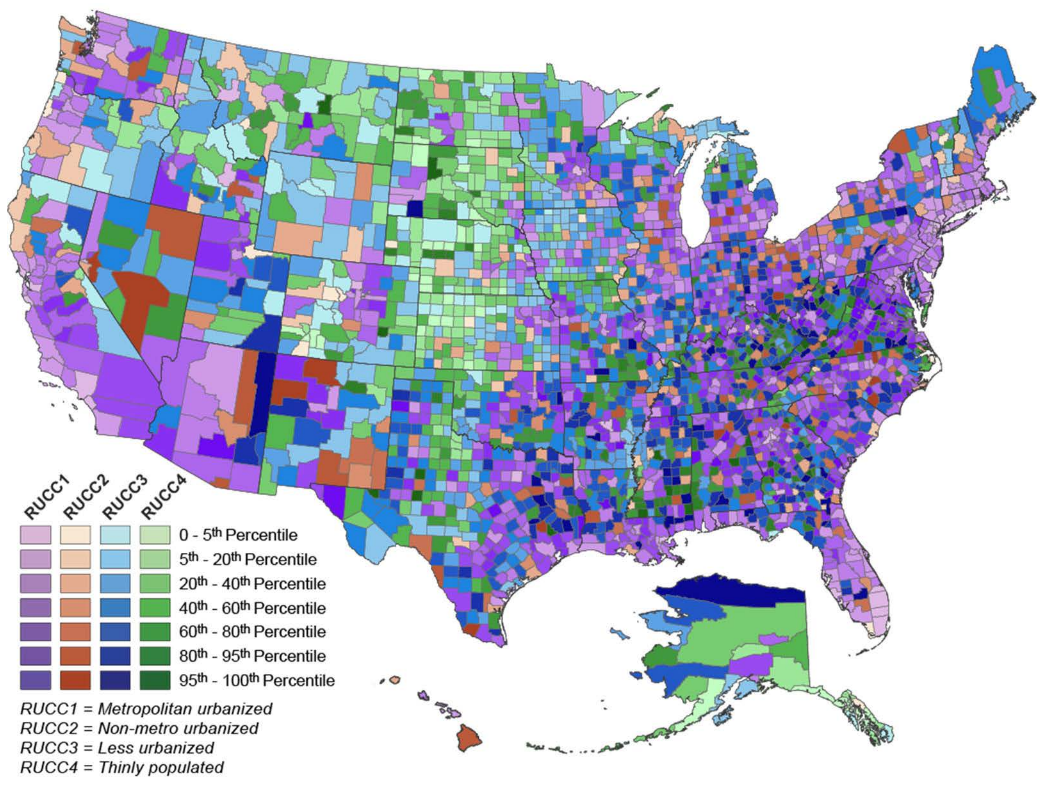


* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Sociodemographic Domain Index Stratified by Rural-Urban Continuum Codes by County 2006-2010



Built Domain Index Stratified by Rural-Urban Continuum Codes by County 2006-2010



* For orientation to the maps, low index scores (EQI and domain-specific) indicate higher environmental quality, and higher index scores (EQI and domain-specific) mean lower environmental quality

Appendix VII: Quality Assurance

The approved Center for Public Health and Environmental Assessment, Public Health and Environmental Systems Division, Quality Assurance Project Plan for this project is "Creating an Overall Environmental Quality Index," with Document Control Number IRP-NHEERL/HSD/EBB/DL/2008-01-QP-1-7. An internal EPA review of this report was conducted in April 2019. An external peer review was conducted in March 2020.

The data sources used to create the EQI and the criteria used to select the data sources are mentioned in this report in the Development of the EQI 2006-2010 section.

Information about uses of the EQI, as well as strengths and limitations of the EQI, is located within the Discussion section of the report.

SCIENCE

EPA
United States
Environmental Protection
Agency

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35

Office of Research and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300

Recycled/Recyclable Printed on paper that contains a minimum of
50% postconsumer fiber content processed chlorine free