



NICHOLAS SCHOOL OF THE
ENVIRONMENT AND EARTH SCIENCES
DUKE UNIVERSITY



ENVIRON 859

Geospatial Data Analytics

Data Engineering

John Fay – Instructor (NSOE)



Data

State Code	County Code	Site Number	Latitude	Longitude	Datum	Elevation	Land Use	Location Setting	Site Established Date	Site	
"01"	"001"	"0001"	32.437458	-86.472891	"WGS84"	64	"RESIDENTIAL"	"SUBURBAN"	"1974-05-01"	"1976-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"001"	"0002"	32.42847	-86.443585	"WGS84"	0	"AGRICULTURAL"	"RURAL"	"1980-01-01"	"1982-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"001"	"0003"	32.332659	-86.791521	"WGS84"	41	"FOREST"	"RURAL"	"1989-08-31"	"1990-11-30"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"003"	"0001"	0,0	"NAD27"	0	"UNKNOWN"	"RURAL"	"1959-01-01"	"1959-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "US EPA National Exposure Research Lab", ""	
"01"	"003"	"0002"	30.552367	-87.706911	"WGS84"	0	"COMMERCIAL"	"RURAL"	"1974-01-01"	"1977-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"003"	"0003"	30.555472	-87.713599	"WGS84"	49	"COMMERCIAL"	"URBAN AND CENTER CITY"	"1977-08-12"	"1987-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"003"	"0010"	30.497478	-87.880258	"NAD83"	37.19	"COMMERCIAL"	"SUBURBAN"	"2000-01-01"	"", "", "", "", "", "ON-SITE MET EQUIP", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""	
"01"	"005"	"0001"	31.891272	-85.13521	"WGS84"	85	"COMMERCIAL"	"SUBURBAN"	"1974-01-01"	"1994-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"005"	"0002"	31.664302	-85.606175	"WGS84"	143.26	"FOREST"	"SUBURBAN"	"2003-10-06"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", "CLIO", "437"	
"01"	"007"	"0001"	32.904048	-87.233057	"WGS84"	445	"AGRICULTURAL"	"RURAL"	"1991-07-15"	"", "", "", "", "", "", "", "", "", "", "-6", "Environmental Science & Engineering", ""	
"01"	"009"	"0001"	33.95702	-86.464443	"WGS84"	250	"MOBILE"	"URBAN AND CENTER CITY"	"1957-01-01"	"1978-05-18"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"009"	"0002"	33.949542	-86.473042	"WGS84"	6	"COMMERCIAL"	"URBAN AND CENTER CITY"	"1979-01-01"	"1997-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"015"	"0001"	33.656771	-85.824409	"WGS84"	237	"COMMERCIAL"	"URBAN AND CENTER CITY"	"1972-01-01"	"1998-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"015"	"0002"	33.685659	-85.837744	"WGS84"	247	"INDUSTRIAL"	"SUBURBAN"	"1974-05-03"	"1985-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"021"	"0001"	32.83956	-86.637809	"WGS84"	213	"RESIDENTIAL"	"SUBURBAN"	"1972-01-01"	"1976-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"021"	"0002"	32.83957	-86.628871	"WGS84"	171	"COMMERCIAL"	"URBAN AND CENTER CITY"	"1976-01-01"	"1993-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"023"	"0001"	32.209311	-88.051961	"WGS84"	39	"INDUSTRIAL"	"RURAL"	"1974-01-01"	"1980-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"023"	"0002"	32.2057	-88.053628	"WGS84"	37	"INDUSTRIAL"	"RURAL"	"1981-01-01"	"1987-01-01"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", "P."
"01"	"025"	"0001"	31.966669	-87.782035	"WGS84"	160	"AGRICULTURAL"	"RURAL"	"1972-01-01"	"1977-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"025"	"0002"	31.694265	-87.773463	"WGS84"	152	"COMMERCIAL"	"SUBURBAN"	"1974-01-01"	"1977-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"027"	"0001"	33.284928	-85.803608	"NAD83"	344.42	"RESIDENTIAL"	"RURAL"	"1991-03-25"	"", "", "", "", "", "ON-SITE MET EQUIP", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""	
"01"	"031"	"0002"	31.056847	-87.259142	"WGS84"	0	"UNKNOWN"	"UNKNOWN"	"1975-01-01"	"1997-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", "UN"
"01"	"033"	"0001"	34.737034	-87.886701	"WGS84"	143	"AGRICULTURAL"	"RURAL"	"1973-01-01"	"1997-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0002"	34.781787	-87.646127	"WGS84"	152	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1976-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0003"	34.770676	-87.646127	"WGS84"	152	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1974-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0004"	34.771231	-87.628348	"WGS84"	168	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1976-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0005"	34.762064	-87.64946	"WGS84"	158	"AGRICULTURAL"	"RURAL"	"1973-01-01"	"1974-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0006"	34.764009	-87.661683	"WGS84"	168	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1974-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0007"	34.760089	-87.665029	"WGS84"	162	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1997-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0008"	34.760922	-87.643085	"WGS84"	155	"AGRICULTURAL"	"RURAL"	"1973-01-01"	"1997-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0009"	34.783453	-87.651127	"WGS84"	152	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1977-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""
"01"	"033"	"0010"	34.780676	-87.636126	"WGS84"	168	"INDUSTRIAL"	"RURAL"	"1973-01-01"	"1974-12-31"	"", "", "", "", "", "", "", "", "", "", "-6", "Al Dept Of Env Mgt", ""

Wisdom

Knowledge

Information

Data



Incorrect coordinate system

Highly skewed

Duplicate records

Incompatible geometries

Missing geometry

Missing values

Spatial outliers

Value outliers

Data is messy!

Duplicate fields

Inconsistent range

Incomprehensible field names

Bad topology

Incorrect data types

Missing locations

Missing records

Stored in multiple databases

Incompatible data formats

Multicollinearity

Data Engineering

messy data  usable information

Data Engineering is for...

Operationalizing data



Analyzing data



Data Engineering for operationalizing data



Accessibility



Integration



Transformation



Automation

Data Engineering for Spatial Analysis and Data Science



Data
Engineering
and Integration



Visualization
& Exploration



Spatial
Analysis



GeoAI



Big Data
Analytics



Modeling
& Scripting



Sharing
& Collaboration

Data Engineering for Spatial Analysis and Data Science



Data
Engineering



Visualization
& Exploration



Spatial
Analysis



GeoAI



Big Data
Analytics



Modeling
& Scripting



Sharing
& Collaboratio

Data Engineering in ArcGIS Pro



Explore
Visualize
Prepare



Document
Replicate
Automate



Air Data: Air Quality Data Collected at Outdoor Monitors Across the US

Outdoor Air Quality Data

Use the tools on this site to access recent and historical data.

For current air quality, visit [AirNow.gov](#).

During fire events, use the [Fire and Smoke map](#).

1

2

3

Download Data

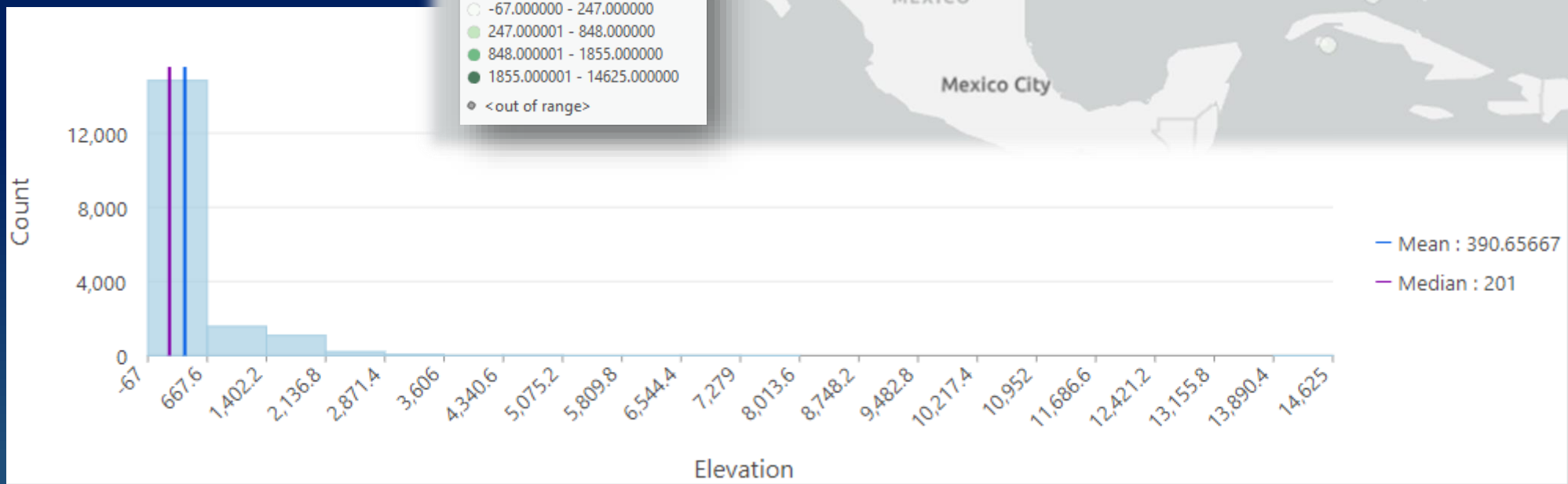
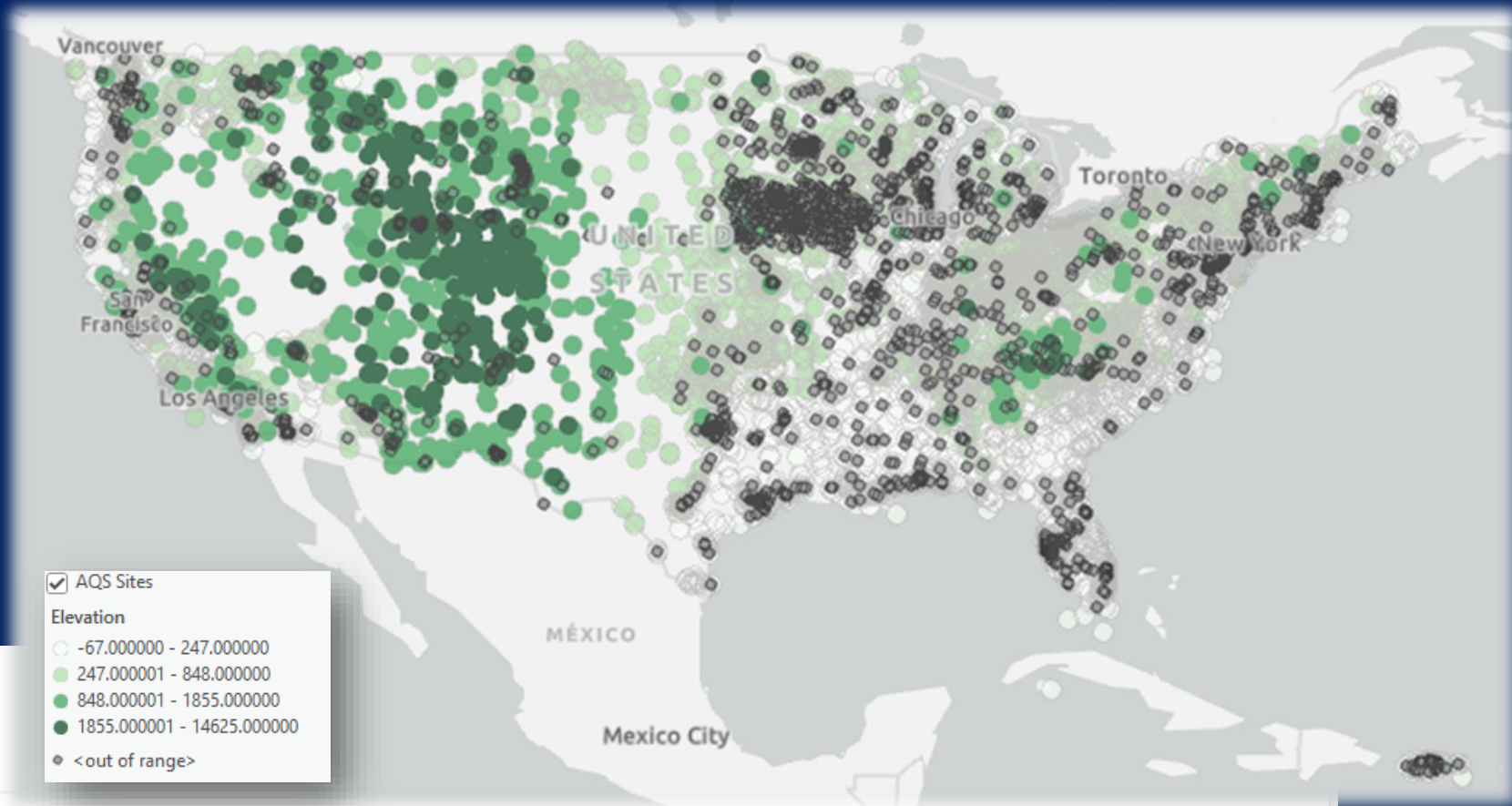
- [Pre-generated Data Files](#)
- [Download Daily Data](#)
- [Download Raw Data \(API\)](#)

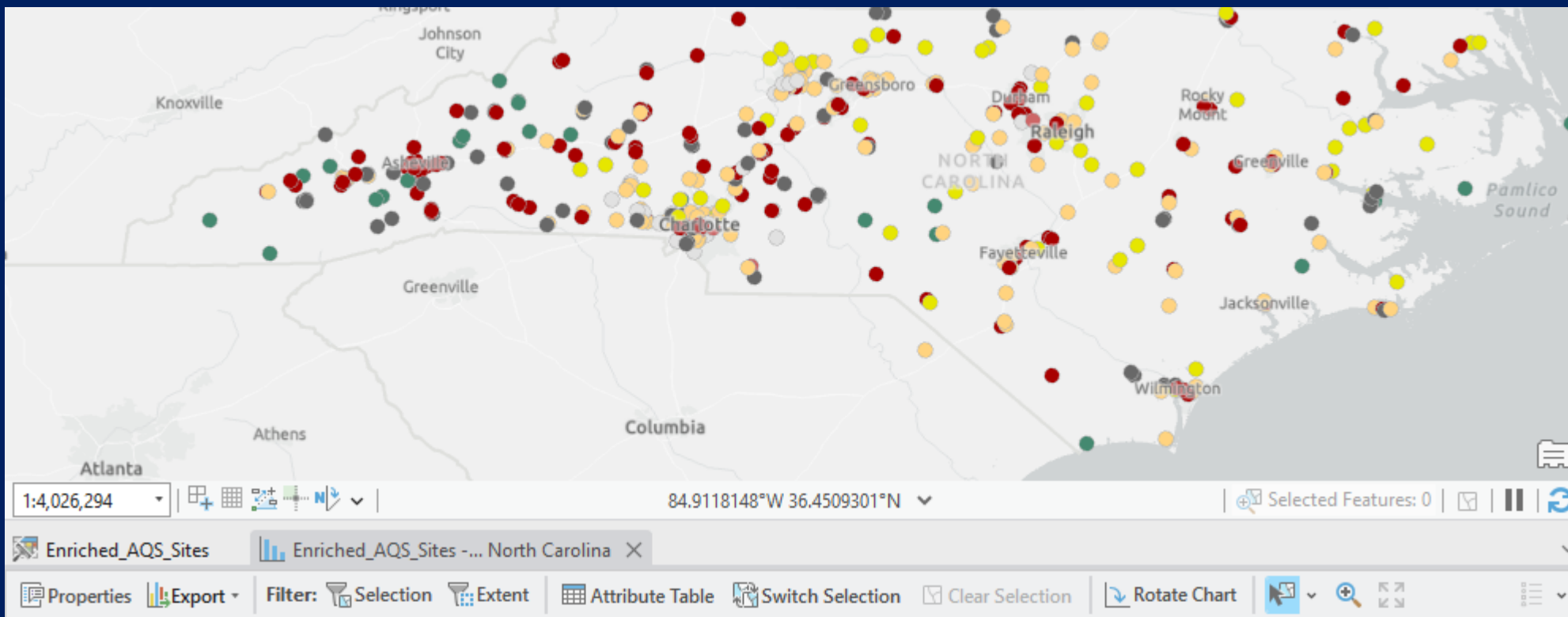
Monitor Locations

- [Interactive Map of Air Quality Monitors](#)

About Air Data

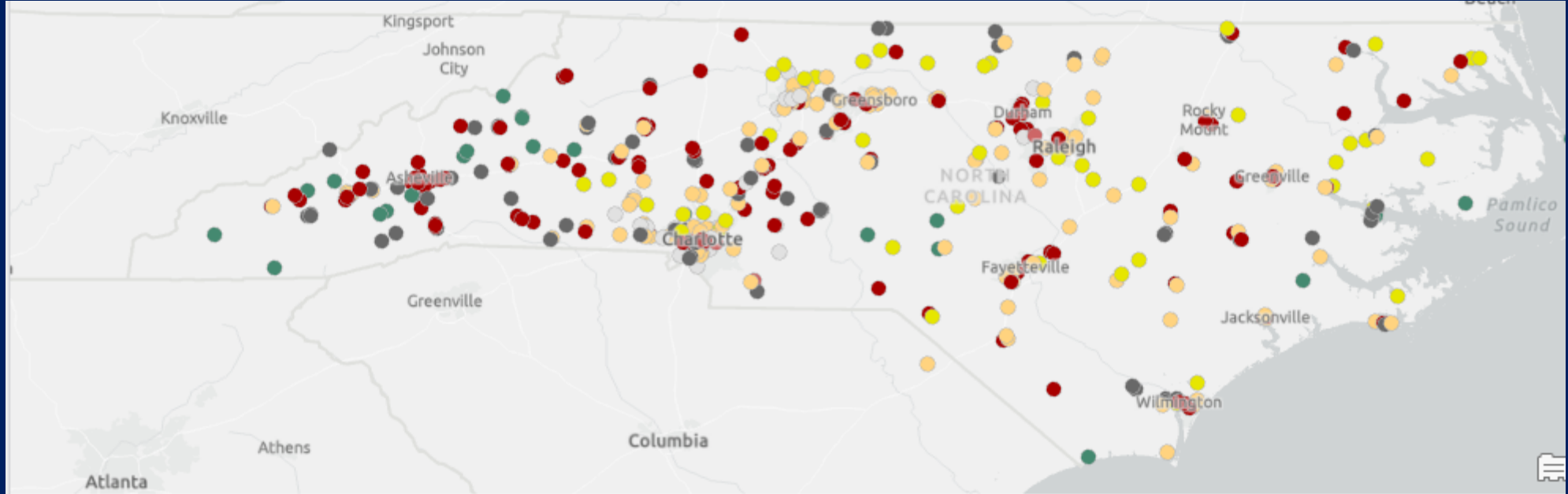
- [Basic Information](#)
- [Frequent Questions](#)
- [Subscribe to RSS feed](#)





Mean 2021 PM2.5 concentration - North Carolina



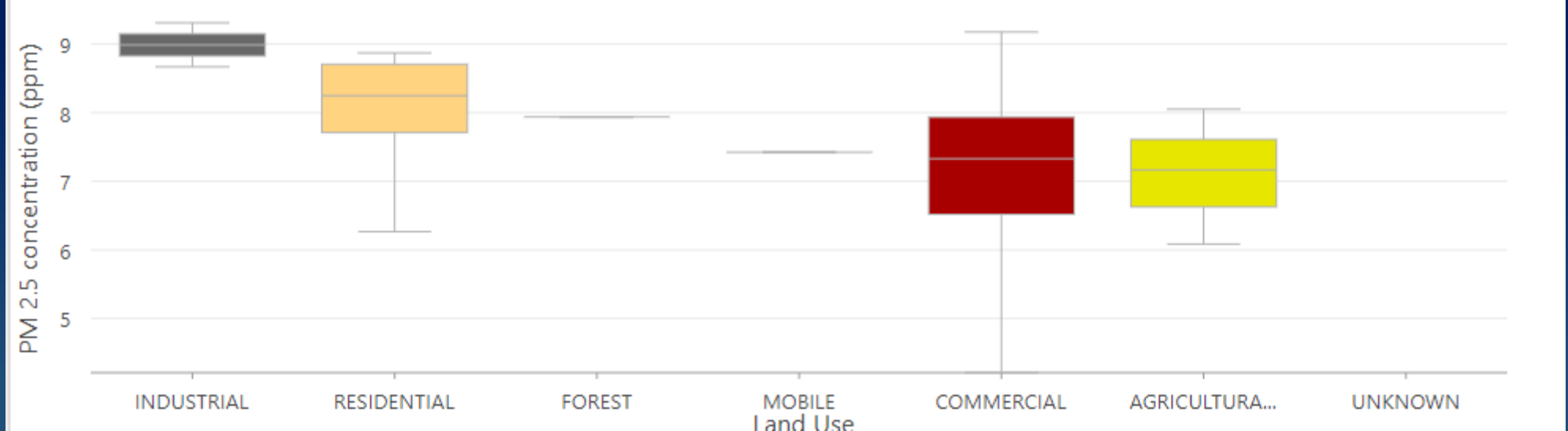


1:4,026,294 | 83.9931251°W 36.3743726°N | Selected Features: 0

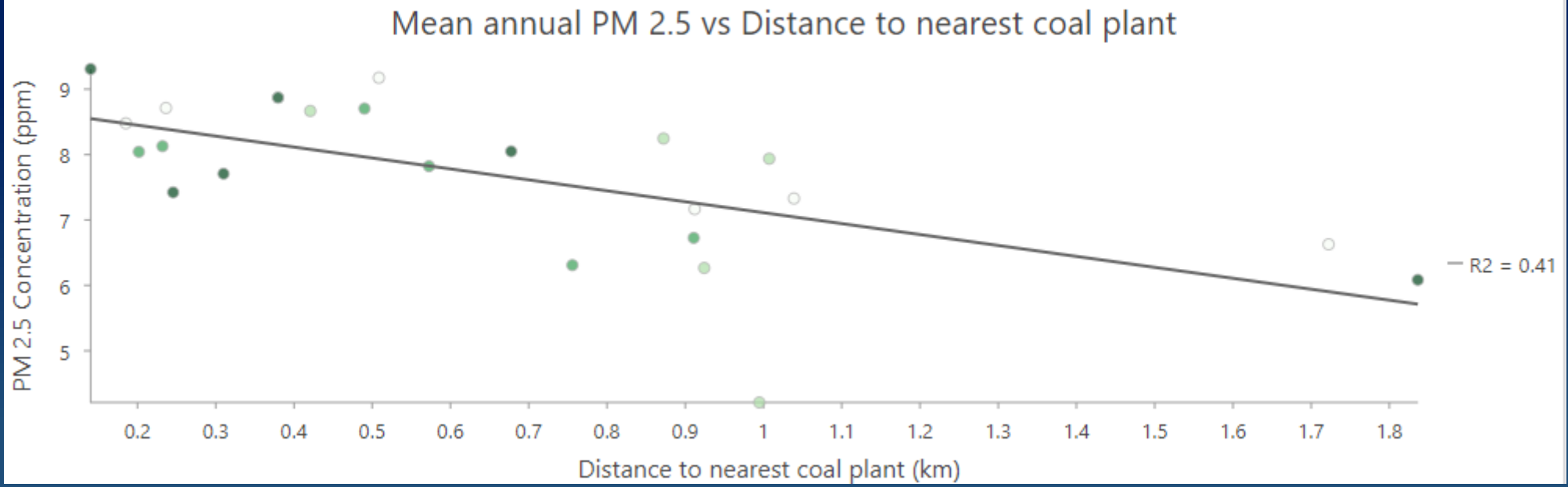
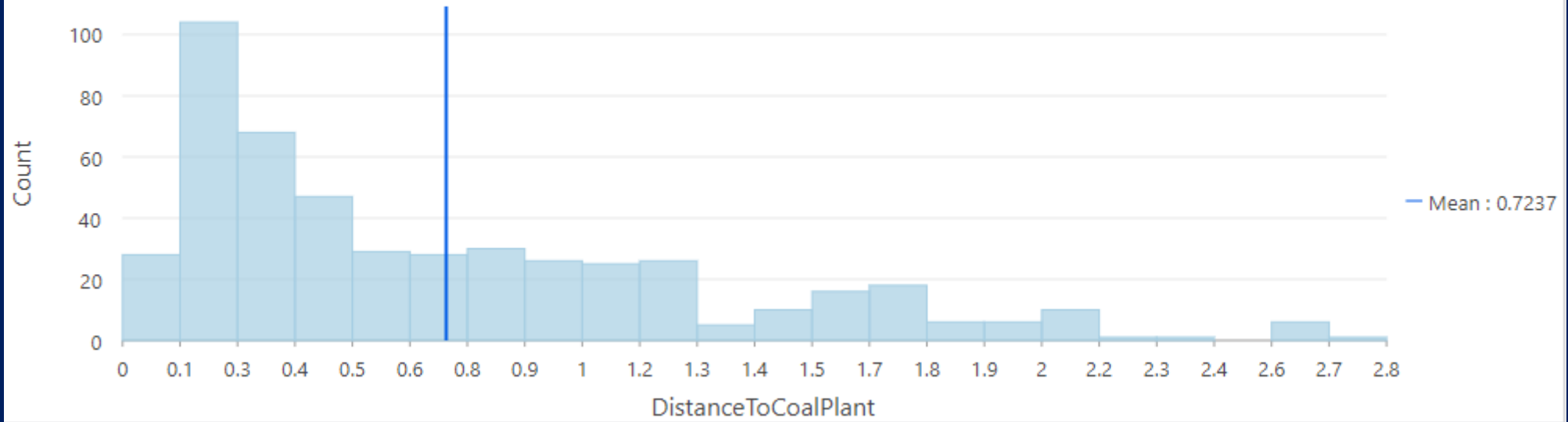
Enriched_AQS_Sites | Enriched_AQS_Sites -... North Carolina | Enriched_AQS_Sites -... North Carolina X

Properties | Export | Filter: Selection | Extent | Attribute Table | Switch Selection | Clear Selection | Rotate Chart

Mean 2021 PM2.5 concentration - North Carolina



Distribution of DistanceToCoalPlant



Workflow

1. Prepare for analysis
2. Fetch the data
3. Import and explore the data
4. Convert data to features and explore more
5. Enrich and enhance the data
6. Transform the data for additional analyses
7. More explorations...