ALL OVER THE MAP: COMPARING THE ACCURACY OF GEOCODING SOURCES

2019 IEPEC CONFERENCE PRESENTATION

Robert Saul, EMI Consulting

Brian Billing, AEP Ohio

August 21, 2019

EMI CONSULTING

AGENDA

1. **"Why Does it Matter?"**

2. Introduction

3. Methods

4. Results

5. Conclusion

BENEFITS OF SPATIAL ANALYSIS IN EE PROGRAMS

MICRO-TARGETING   **SPATIAL PATTERNS**   **SEGMENTATION**

BENEFITS OF SPATIAL ANALYSIS IN EE PROGRAMS

MICRO-TARGETING   SPATIAL PATTERNS   SEGMENTATION

As energy efficiency programs become more successful, utilities are more frequently using spatial analysis to identify and influence hard-to-reach customers, such as:

- Low-income

- Environmentally disadvantaged

"WHY DOES IT MATTER?"
BENEFITS OF SPATIAL ANALYSIS IN EE PROGRAMS

MICRO-TARGETING

As energy efficiency programs become more successful, utilities are more frequently using spatial analysis to identify and influence hard-to-reach customers, such as:

- Low-income

- Environmentally disadvantaged

SPATIAL PATTERNS

Many utilities are considering new program designs that hinge on specific geographic energy usage patterns, such as:

- Demand-response

- Electric vehicles

- Non-wires solutions

SEGMENTATION

BENEFITS OF SPATIAL ANALYSIS IN EE PROGRAMS

MICRO-TARGETING

As energy efficiency programs become more successful, utilities are more frequently using spatial analysis to identify and influence hard-to-reach customers, such as:

- Low-income
- Environmentally disadvantaged

SPATIAL PATTERNS

Many utilities are considering new program designs that hinge on specific geographic energy usage patterns, such as:

- Demand-response
- Electric vehicles
- Non-wires solutions

SEGMENTATION

To better inform program design, utilities are combining program data with geographic data such as:

- Weather
- Traffic
- Demographics

IEPEC

EMI CONSULTING

LARGE SHIFTS IN GEOCODING SERVICES

1. OLD INFO IN STAND-ALONE SERVICES

2. CHANGE IN API REQUIREMENTS
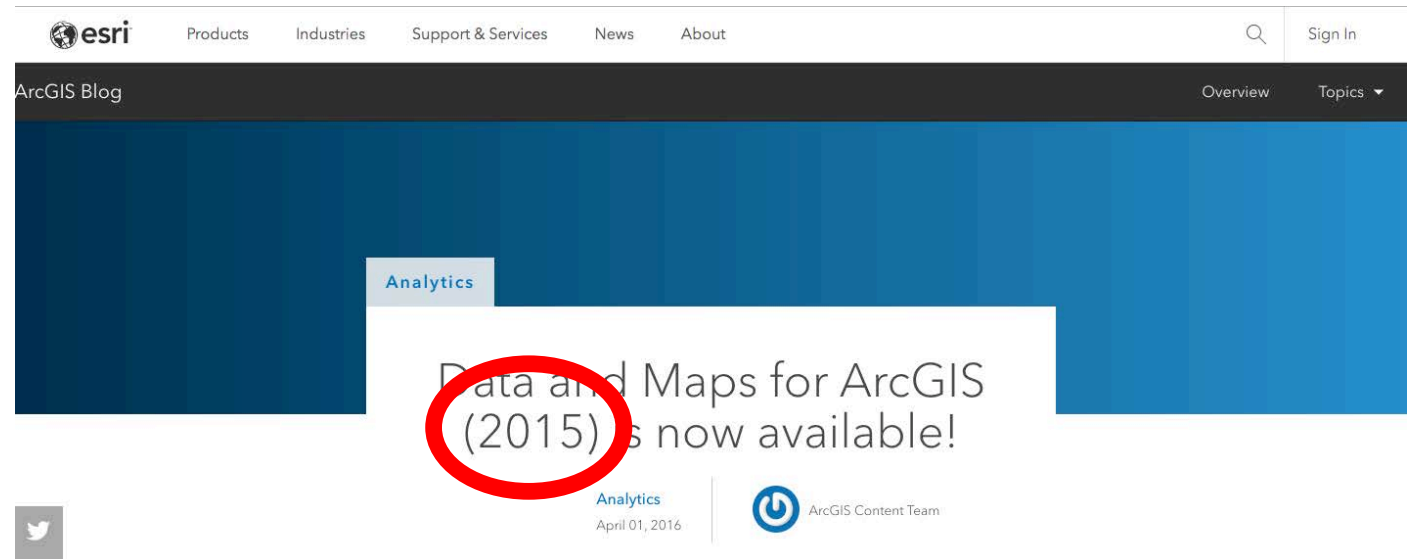
3. CHANGES TO SERVICE CAPABILITIES

IEPEC

EMI CONSULTING

"WHY DOES IT MATTER?"
LARGE SHIFTS IN GEOCODING SERVICES

1. OLD INFO IN STAND-ALONE SERVICES

2. CHANGE IN API REQUIREMENTS

3. CHANGES TO SERVICE CAPABILITIES

https://www.esri.com/arcgis-blog/products/analytics/analytics/data-and-maps-for-arcgis-2015-is-now-available/

LARGE SHIFTS IN GEOCODING SERVICES

1. OLD INFO IN STAND-ALONE SERVICES

2. CHANGE IN API REQUIREMENTS
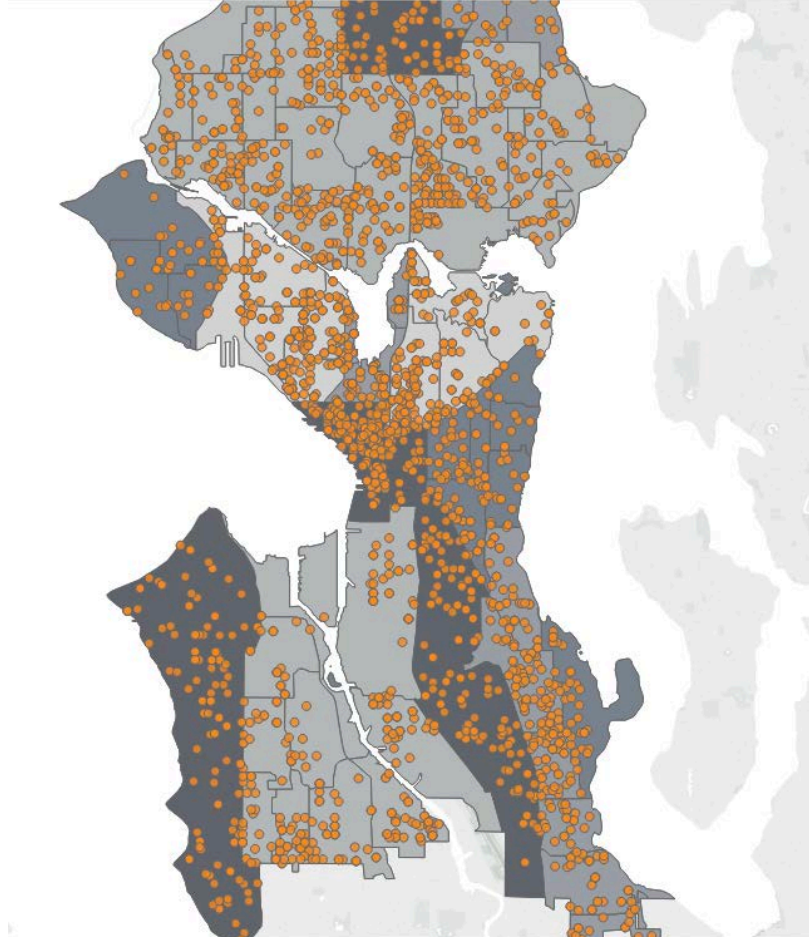
3. CHANGES TO SERVICE CAPABILITIES

"Google has recently changed its API requirements... users are now required to provide an API key and **enable billing**."

https://www.littlemissdata.com/blog/ggmap-updated

1. OLD INFO IN STAND-ALONE SERVICES

2. CHANGE IN API REQUIREMENTS

3. CHANGES TO SERVICE CAPABILITIES

https://multiplottr.com/

https://community.tableau.com/thread/232540

AGENDA

REVIEW OF GEOCODING SERVICES AND SITES

Geocoding Service	Free or Paid Service	Has API?
Alteryx	Free	
ArcGIS	Paid	✓
BatchGeo	Free and Paid	
GeoLytics	Paid	
Google	Free and Paid	✓
Google Sheets	Free and Paid	
HERE	Free and Paid	✓
MapLarge	Paid	✓
OpenStreetMap	Free	✓
Texas A&M Geoservices Geocoding	Free and Paid	✓
TomTom	Free and Paid	✓
United States Census Bureau Geocoder API	Free	✓

CLIENT

AEP Ohio: Electric investor-owned utility serving nearly 1.5 million customers.

AEP OHIO℠
An **AEP** Company

CLIENT

AEP Ohio: Electric investor-owned utility serving nearly 1.5 million customers.

GEOGRAPHY

Majority of AEP Ohio customers located in and around Columbus, OH.

CLIENT

AEP Ohio: Electric investor-owned utility serving nearly 1.5 million customers.

GEOGRAPHY

Majority of AEP Ohio customers located in and around Columbus, OH.

DATA

Examined the geolocational data of 151 AEP Ohio residential survey respondents who completed a Qualtrics web survey in January 2019.

Survey respondents part of residential rebate program evaluation.

AGENDA

RAW DATASET

Raw dataset of survey respondent service addresses as received from AEP Ohio.

DATASETS

RAW DATASET

Raw dataset of survey respondent service addresses as received from AEP Ohio.

CLEAN DATASET

Cleaned dataset of survey respondent service addresses using cleaning best-practices recommended by the Harvard School of Public Health (HSPH 2017).

DATASETS

RAW DATASET

Raw dataset of survey respondent service addresses as received from AEP Ohio.

CLEAN DATASET

Cleaned dataset of survey respondent service addresses using cleaning best-practices recommended by the Harvard School of Public Health (HSPH 2017).

HYPOTHESIS

Differences in the geocoded locations of raw data and the cleaned data would give an indication of how effectively each service handles data quality issues.
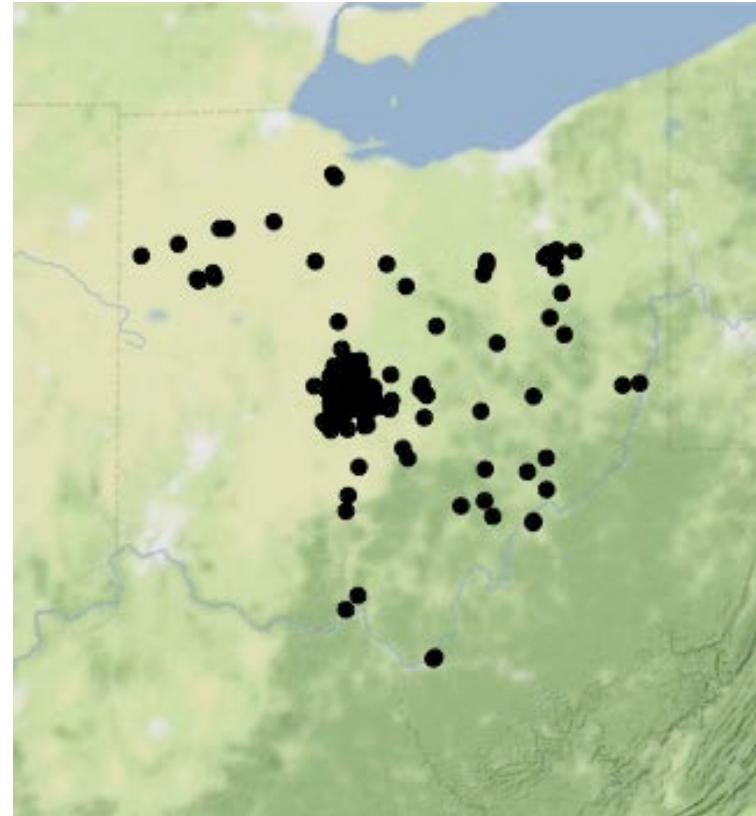
IEPEC

EMI CONSULTING

GEOCODING SOURCES TESTED

1. USED GOOGLE API AS "BASELINE"

2. PERFORMED SPOT-CHECK ON 20 RECORDS

3. COMPARED BASELINE TO 6 OTHER SERVICES

GEOCODING SOURCES TESTED

1. USED GOOGLE API AS "BASELINE"

2. PERFORMED SPOT-CHECK ON 20 RECORDS

3. COMPARED BASELINE TO 6 OTHER SERVICES
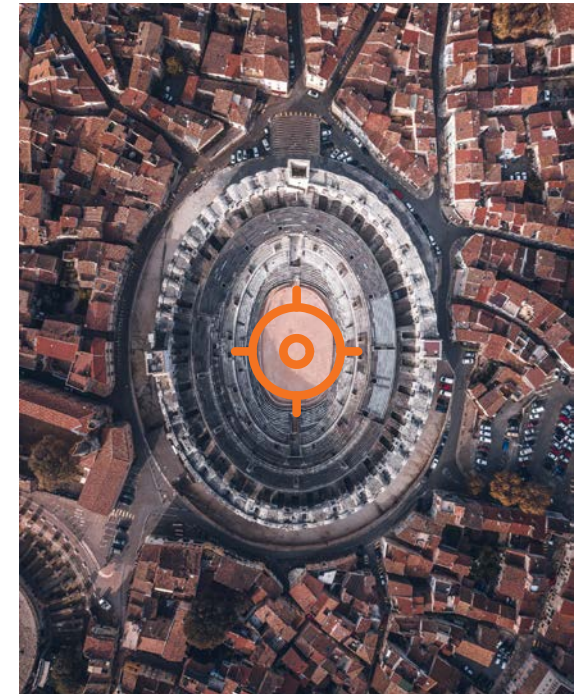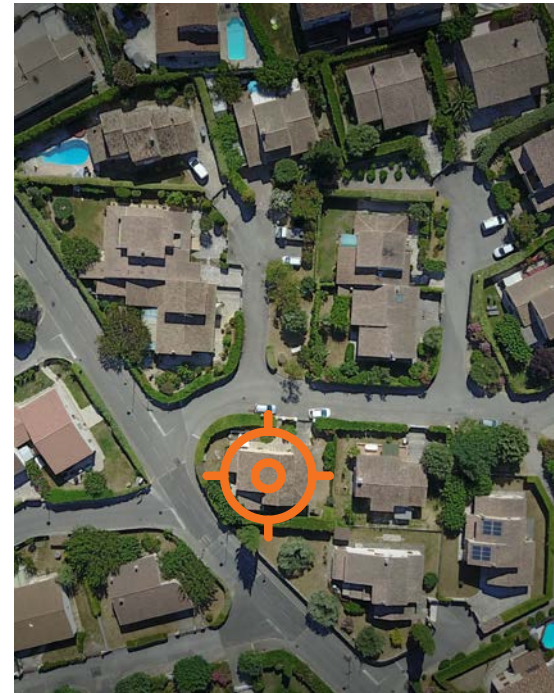
GEOCODING SOURCES TESTED

1. USED GOOGLE API AS "BASELINE"

2. PERFORMED SPOT-CHECK ON 20 RECORDS

3. COMPARED BASELINE TO 6 OTHER SERVICES

✔   X

GEOCODING SOURCES TESTED

1. USED GOOGLE API AS "BASELINE"

2. PERFORMED SPOT-CHECK ON 20 RECORDS

3. COMPARED BASELINE TO 6 OTHER SERVICES

Geocoding Service	Geocoding Type	Geocoded as Baseline	Geocodes Performed on Both Raw and Clean Address Data	Geocodes Based on IP Address
2013 Garmin	Map layer used as address locator tool in ArcMap		✓	
ArcGIS Online	Online application		✓	
Census API	API		✓	
Google API	API	✓		
Google Sheets	Script within online application		✓	
OSM API	API		✓	
Qualtrics	Locational data provided with survey responses			✓

IEPEC

EMI CONSULTING

AGENDA

1. "Why Does it Matter?"

2. Introduction

3. Methods

4. Results

5. Conclusion

SUMMARY STATISTICS FROM COMPARISON TO BASELINE

RAW VS CLEAN

In this particular experiment, found little difference in how the services geocoded the raw addresses and the cleaned addresses.

Likely speaks to good data practices on the part of AEP Ohio rather than the capabilities of the geocoding services.

SUMMARY STATISTICS FROM COMPARISON TO BASELINE

RAW VS CLEAN

In this particular experiment, found little difference in how the services geocoded the raw addresses and the cleaned addresses.

Likely speaks to good data practices on the part of AEP Ohio rather than the capabilities of the geocoding services.

BASELINE VS OTHER SERVICES

- Qualtrics was the outlier.

Geocoding Service	Unmatched Geocoding Records	Average Distance from Baseline Coordinate (in Miles)	Median Distance from Baseline Coordinate (in Miles)
Google Sheets	0	< 0.01	< 0.01
Census API	10	0.05	0.02
2013 Garmin	0	0.22	0.02
OSM API	61	0.32	0.11
Qualtrics	0	95.30	6.15

IEPEC

EMI CONSULTING

SUMMARY STATISTICS FROM COMPARISON TO BASELINE

RAW VS CLEAN

In this particular experiment, found little difference in how the services geocoded the raw addresses and the cleaned addresses.

Likely speaks to good data practices on the part of AEP Ohio rather than the capabilities of the geocoding services.

BASELINE VS OTHER SERVICES

- Qualtrics was the outlier.

- OSM and the Census did not find matches for all addresses.

Geocoding Service	Unmatched Geocoding Records	Average Distance from Baseline Coordinate (in Miles)	Median Distance from Baseline Coordinate (in Miles)
Google Sheets	0	< 0.01	< 0.01
Census API	10	0.05	0.02
2013 Garmin	0	0.22	0.02
OSM API	61	0.32	0.11
Qualtrics	0	95.30	6.15

SUMMARY STATISTICS FROM COMPARISON TO BASELINE

RAW VS CLEAN

In this particular experiment, found little difference in how the services geocoded the raw addresses and the cleaned addresses.

Likely speaks to good data practices on the part of AEP Ohio rather than the capabilities of the geocoding services.

BASELINE VS OTHER SERVICES

- Qualtrics was the outlier.

- OSM and the Census did not find matches for all addresses.

- Excluding Sheets, Census API had results closest to baseline.

Geocoding Service	Unmatched Geocoding Records	Average Distance from Baseline Coordinate (in Miles)	Median Distance from Baseline Coordinate (in Miles)
Google Sheets	0	< 0.01	< 0.01
Census API	10	0.05	0.02
2013 Garmin	0	0.22	0.02
OSM API	61	0.32	0.11
Qualtrics	0	95.30	6.15

FURTHER EXPLORING DIFFERENCES

1. QUALTRICS: SURVEY RESPONDENT LOCATION

2. GEOCODING SERVICES VISUAL COMPARISON

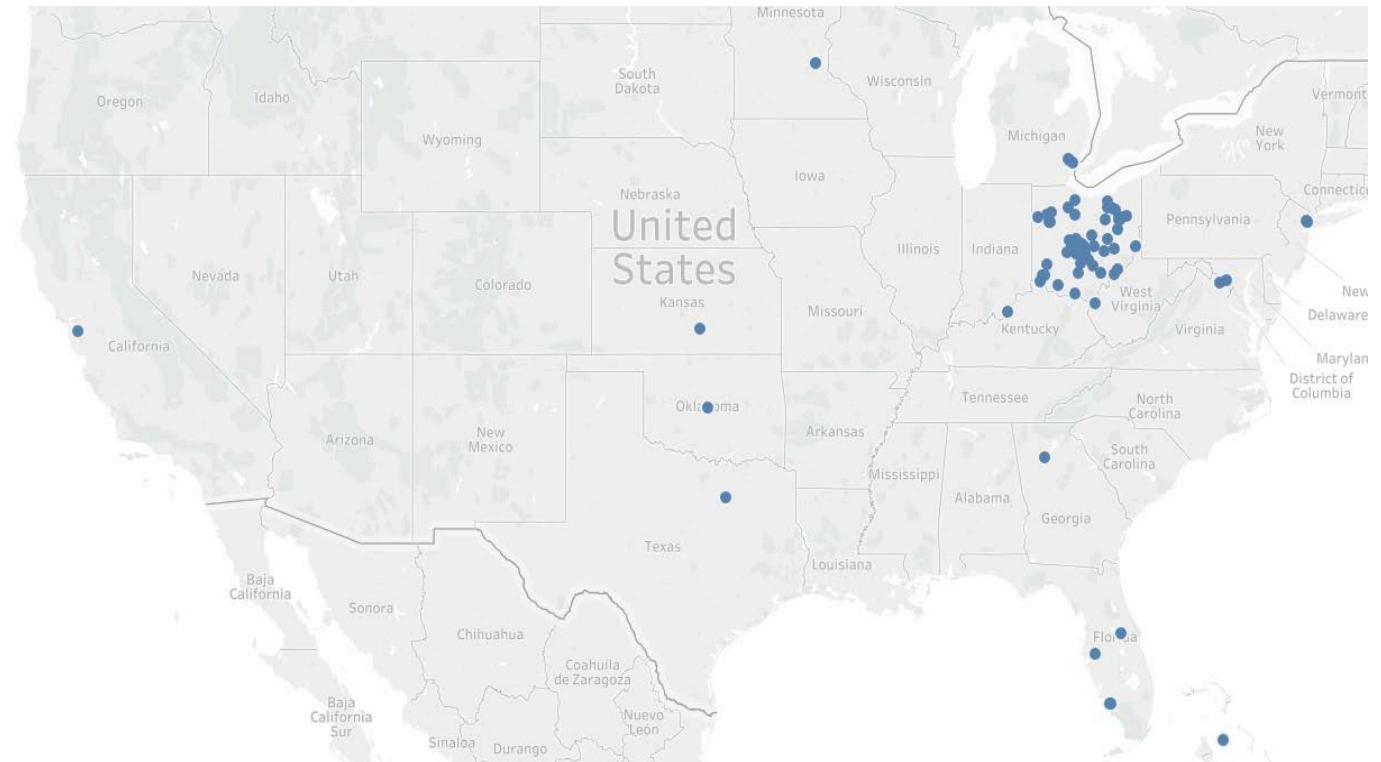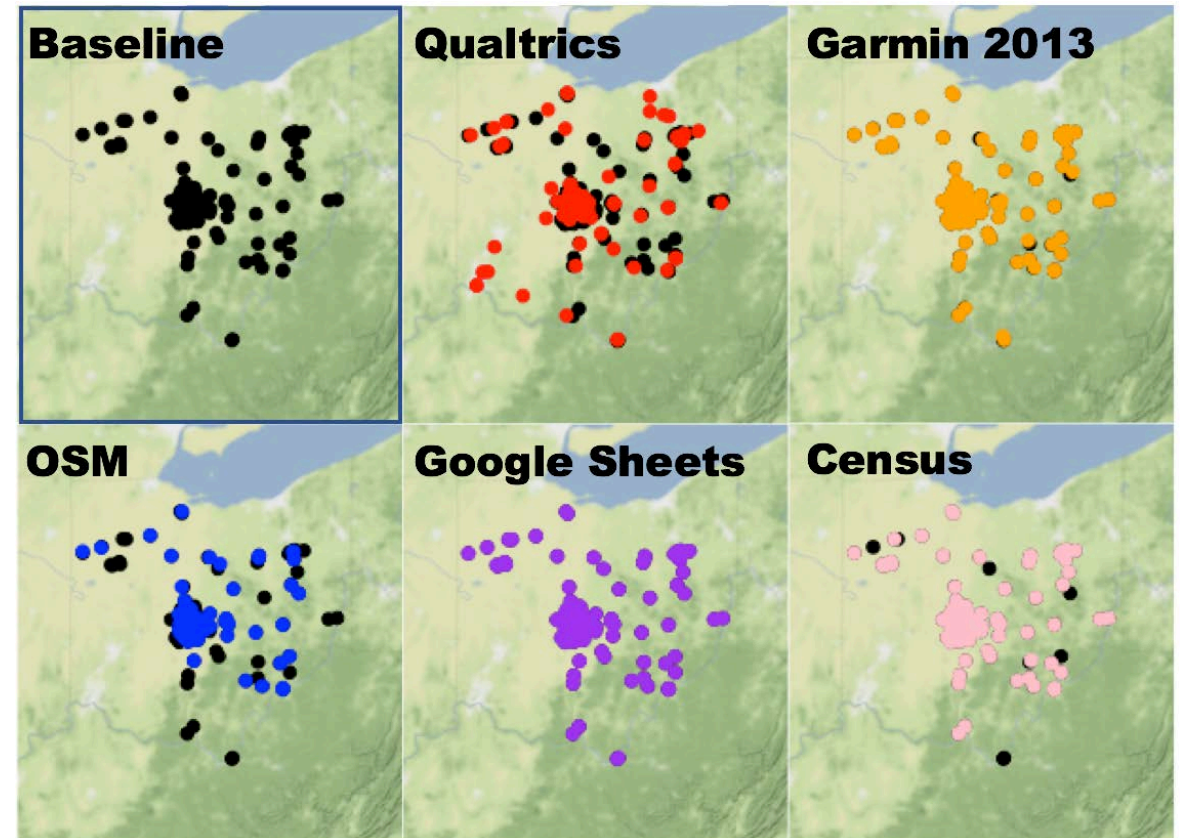**3.** COMPARING BASELINE TO ARCGIS ONLINE

FURTHER EXPLORING DIFFERENCES

1. QUALTRICS: SURVEY RESPONDENT LOCATION

2. GEOCODING SERVICES VISUAL COMPARISON

3. COMPARING BASELINE TO ARCGIS ONLINE

FURTHER EXPLORING DIFFERENCES

1. QUALTRICS: SURVEY RESPONDENT LOCATION

2. GEOCODING SERVICES VISUAL COMPARISON

3. COMPARING BASELINE TO ARCGIS ONLINE

FURTHER EXPLORING DIFFERENCES

1. QUALTRICS: SURVEY
RESPONDENT LOCATION

2. GEOCODING SERVICES
VISUAL COMPARISON

3. COMPARING BASELINE TO
ARCGIS ONLINE

AGENDA

1. "Why Does it Matter?"

2. Introduction

3. Methods

4. Results

5. Conclusion

SMALL BATCH

Geocode a small amount of records (**less than 100**): may be easiest to use to **Google Sheets.**

SMALL BATCH

Geocode a small amount of records (**less than 100**): may be easiest to use to **Google Sheets**.

LARGE BATCH

Geocode a larger amount of records (**less than 40k**): using the **Google API** may be free to use.*
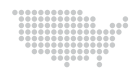
* Statement was true at the time the paper was written.

SMALL BATCH

Geocode a small amount of records (**less than 100**): may be easiest to use to **Google Sheets**.

LARGE BATCH

Geocode a larger amount of records (**less than 40k**): using the **Google API** may be free to use.*

INCREASE ACCURACY

Can perform a second set of geocoding on a subset of records using a different geocoding service.

Can **check the subset** of geocodes that are **flagged as potentially less accurate** based on the geocoders accuracy field.

IEPEC

EMI CONSULTING

SUMMARY TABLE

| GEOCODING SERVICE	RECOMMENDED USE	COST FOR SMALL NUMBER OF GEOCODED LOCATIONS	COST FOR LARGE NUMBER OF GEOCODED LOCATIONS	ACCURACY	INCLUDES ACCURACY METRIC
Google API	Accurate geocoding of a small number of records; have help from a proficient coder	Free (less than 40,000 calls a month)	High (between $4 and $5 per 1,000 calls)	High	Yes
Google Sheets	Accurate geocoding of a small record batch.	Free (unclear when pay barrier starts)	Unknown (unclear when pay barrier starts)	High	No
Census API	Fairly accurate geocoding of large number of records; unmatched records won't pose an issue; have help from a proficient coder; pair with another service for unmatched addresses		Free	Moderate – High	Yes
2013 Garmin North America	Fairly accurate geocoding of large number of records; fast processing time; own Garmin map layers; have help from an ArcGIS user	Cost of acquiring Garmin map layer (cost unknown)		Moderate – High	Yes
OSM API	Less accurate geocoding of large number of records; unmatched records won't pose an issue; have help from a proficient coder; pair with another service for unmatched addresses		Free	Low – Moderate	No
Qualtrics	Quick check that survey respondents are answering in roughly the predicted pattern		Free (no additional cost with Qualtrics service)	Low	No
ArcGIS Online	Accurate geocoding; high price	Approximately $4 for 1,000 geocodes		High	Yes

WHAT ACCURACY IS NEEDED AND AT WHAT COST?

WHAT ACCURACY IS NEEDED AND AT WHAT COST?

https://oehha.ca.gov/calenviroscreen/maps-data
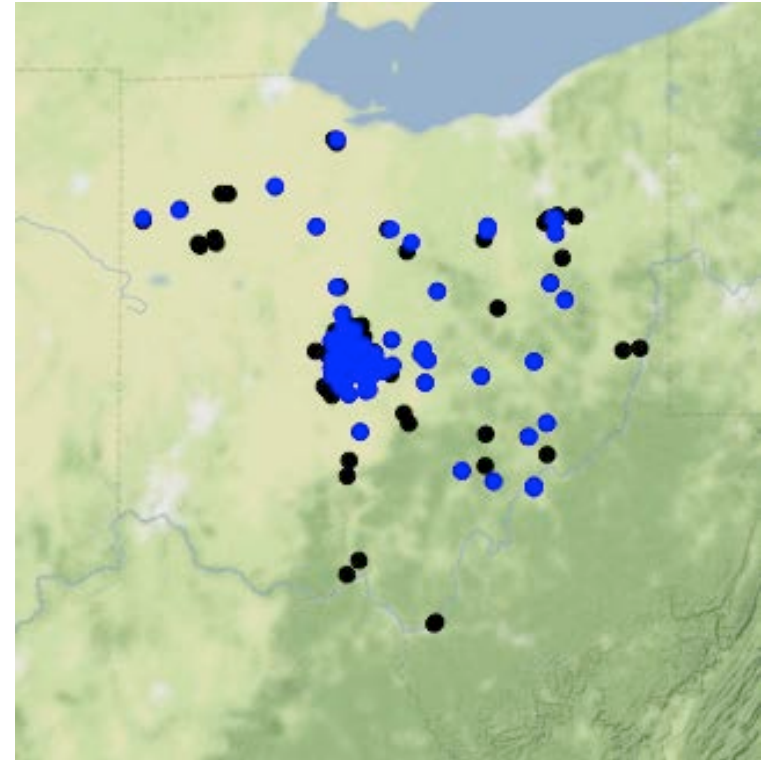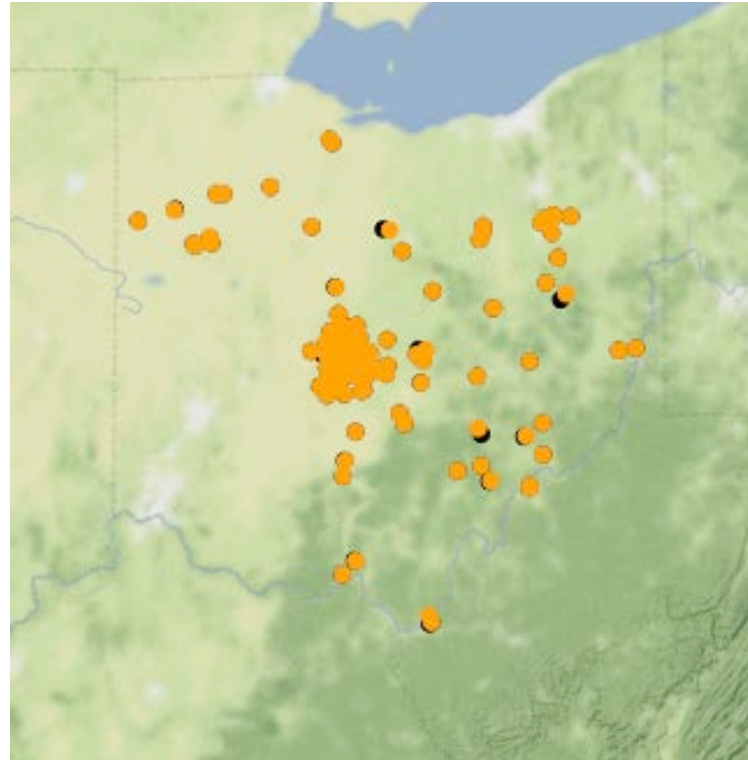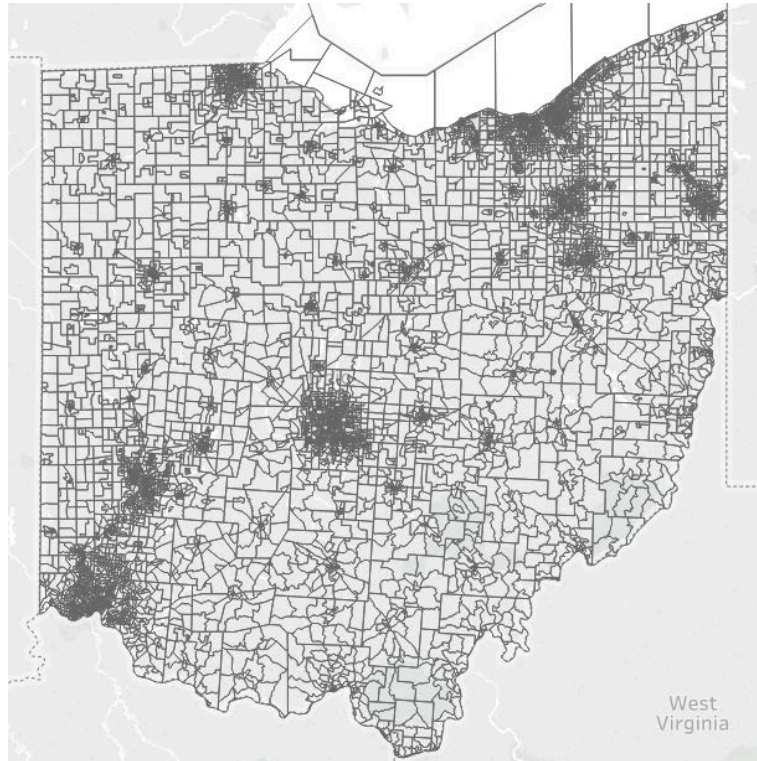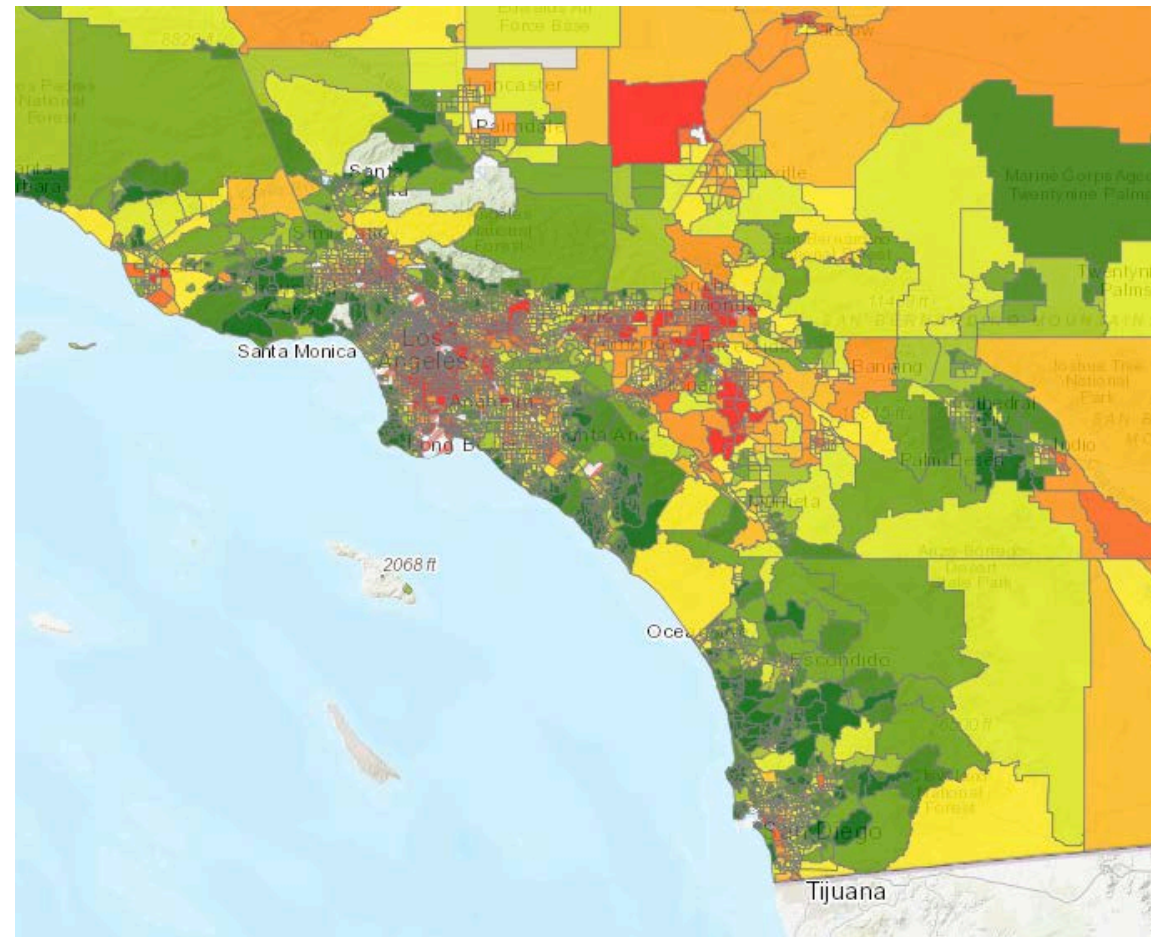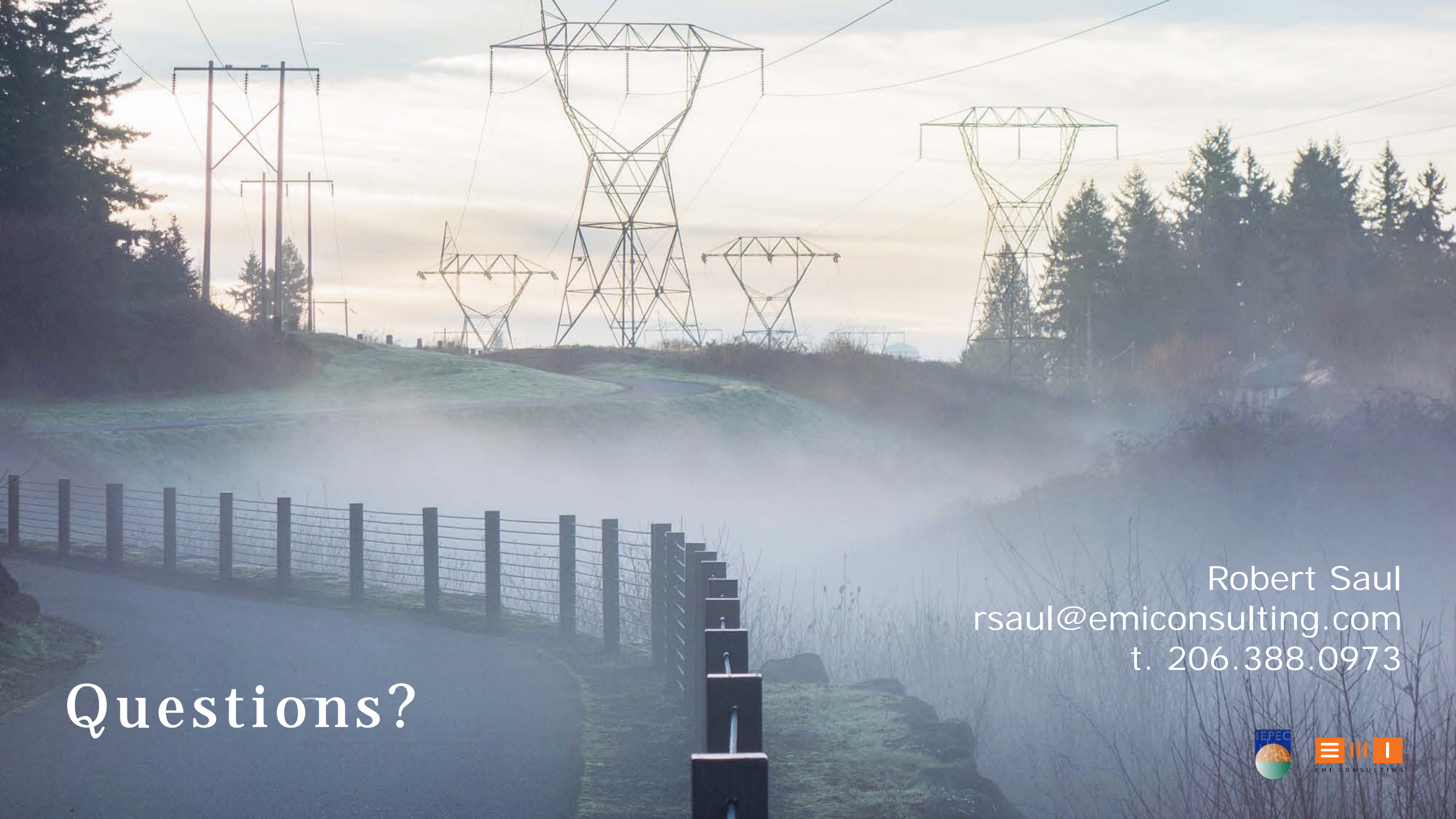
Questions?

Robert Saul
rsaul@emiconsulting.com
t. 206.388.0973