MID-OHIO REGIONAL
MORPC
PLANNING COMMISSION

111 Liberty St., Suite 100
Columbus, Ohio 43215
www.morpc.org

NOTICE OF A MEETING
REGIONAL INFORMATION & DATA GROUP (RIDG)
MID-OHIO REGIONAL PLANNING COMMISSION
111 Liberty Street, Columbus, Ohio 43215

IN-PERSON Meeting with Hybrid Option Available
February 12, 2025, 2:30 pm – 4:00 pm

AGENDA

1. **Welcome & Introductions –** *Ethan Hug*

2. **Updates**
The RIDG Charter will be altered to remove the requirement for a formal Steering Committee and Chair, aiming to streamline programming and reduce procedural requirements. The desire is to rely on RIDG members to provide ideas and connections for potential RIDG meeting topics. The formal requirements will just be lessened.

3. **Topic Discussion –** *Lanakila Alexander, Matt Bigelow, Carrie Wei, Yiting Wang, OSU*
Community's Built Environment and its Impacts on Tax Productivity
Students from The Ohio State University in the Master of Translational Data Analytics program will discuss their plan for analyzing parcel data from Delaware County's auditor for property tax productivity; "value per acre." Their discussion will include how they address challenges that arose during data exploration and analysis. Their work is informed by prior art conducted the firm Urban3 that has shown built environment developed prior to the proliferation of the automobile tends to produce much more tax revenue per acre than auto-oriented development and also costs less for provision of municipal services. Consequently, pedestrian- and transit-oriented development patterns tend to be more fiscally sustainable for communities. Additionally, newer and wealthier neighborhoods tend to be subsidized by older, poorer, and sometimes "blighted" neighborhoods.

Value Per Acre Analysis
Lanakila Alexander presented the team's analysis comparing old and new business blocks, showing that traditional multi-business blocks are more tax productive than modern single-business blocks. The analysis aims to inform policy decisions and planning, suggesting that improving traditional business blocks may be more beneficial than replacing them with modern developments.

The team referenced Urban 3, a nonprofit that studies development patterns, noting that older, less auto-centric development tends to be more productive and cost-effective.

Data Wrangling Challenges

Matt Bigelow discussed the challenges faced in data wrangling, including obtaining parcel IDs, handling duplicate records, and addressing outliers. Solutions included communication with the county auditor's office, spatial join methodology, and creating an outlier detection package in Python. Matt also discussed the importance of data cleaning, including removing zero-value records and standardizing units and calculations to ensure accurate analysis.

Census Data Integration

Lanakila Alexander explained the integration of census data to analyze social factors such as population distribution, housing vacancies, and poverty levels. This data will help understand the relationship between social factors (such as population density, housing occupancy, and poverty distribution) and tax productivity.

Challenges included changes in census block boundaries over time and the need to join spatial and tabular data using unique Geo IDs.

Future work could include extending the analysis to cover historical population distributions and examining the impact of social factors on tax productivity over time.

Tax Productivity Analysis

Carrie Wei analyzed tax-related attributes, showing how the team decided to use total taxable value instead of annual tax for more reliable insights. The analysis revealed that commercial residential properties have higher taxable values. Properties participating in the Community Reinvestment Area (CRA) program were found to have significantly higher taxable values per acre compared to non-participants. Spatial joins were used to validate the data, ensuring accurate classification and analysis of properties based on their land use categories.

Development Patterns Over Time

Yiting Wang presented the analysis of development patterns over time, showing shifts in tax productivity across different property types. The analysis highlighted the impact of remodeling on tax productivity, with commercial properties seeing the biggest gains. Yiting also discussed the creation of distinct development eras to better reflect the evolution of Delaware County's built environment, revealing clear patterns in tax productivity over time.

The analysis showed that office, commercial, and commercial residential mixed-use properties consistently have high value per acre, while post-World War II development showed a notable dip in productivity.

Future analysis could focus on the impact of location on property values and further exploration of historical development patterns and their social context.

Documentation and Reproducibility
Team members emphasized the importance of documentation and reproducibility in data projects. The team established clear file naming conventions, standardized folder structures, version control practices, and regular validation checkpoints to ensure data integrity and ease of future updates. They emphasized the iterative nature of data wrangling, noting the importance of documenting assumptions and building flexible code to adapt as understanding evolves.

4. **Closing Remarks / Adjourn –** *Ethan Hug*
   - Future Meetings
     - Future Meeting Topics: General discussion
       Staff requested that RIDG Members send ideas for topics to be discussed in the next meeting to the organizer, Ethan Hug, at EHug@morpc.org. Possible topics of discussion at the May 2025 meeting are data governance and/or policies and foundations for AI use
     - General meeting: Wednesday, May 14, 2025, 2:30-4:00pm

Please notify Lynn Kaufman at 614-233-4189 or LKaufman@morpc.org to confirm your attendance for this meeting or if you require special assistance.

The next Meeting of the Regional Information & Data Group will be Wednesday, May 14, 2025, 2:30 - 4:00 pm IN-PERSON with remote option available.

Mid-Ohio Regional Planning Commission
Remote Meeting

Regional Information & Data Group Meeting

February 12, 2025

Attendees Present:

- Mason Alexander, Franklin County Dept. of Jobs & Family Services
- Andrew Bishop, Union County
- Emily Canan, The Columbus Region
- Agata Dryla-Gaca, Chicago Metropolitan Agency for Planning
- Brad Ebersole, Consolidated Cooperative
- Sam Filkins, Knox County Area Development Foundation
- Brad Fisher, Delaware County Regional Planning Commission
- Stephanie Joseph, Source Point
- Jonathan Kabat, Delaware County
- Jeff Kasson, City of Upper Arlington
- Mayor Jeff Kinnell, Village of Galena
- Robert Kramer, South-Western City School District
- Bill LaFayette, Regionomics
- James Mako, City of Westerville
- Drew Merrill, COTA
- Rob Moore, Scioto Analysis
- Tom Noorkah, City of Columbus
- Kristen Pietras, Franklin County
- Renata Ramsini, City of Columbus
- CJ Rhodes,
- Denise Roberts, Franklin County
- Langdon Sanders, City of Dublin
- Michael Shinaba, State of Ohio
- Yiting Wang, OSU
- Fara Waugh, Source Point
- Jason Werner

Staff Present

- Ethan Hug
- Jordan Inskeep
- Lynn Kaufman
- Jonathan Miller
- Adam Porr

A Quiz with No Reward…

Which One Produces More Tax Dollars Per Unit Area?

“Old and blighted”

“Shiny and new”

A Quiz with No Reward…

Which One Produces More Tax Dollars Per Unit Area?

A Quiz with No Reward…

Which One Produces More Tax Dollars Per Unit Area?

"Old and blighted"

"Shiny and new"

"Old and blighted"

"Shiny and new"

"Old and Blighted" Properties	Acres	2011	2014	Change ($)	Change (%)
MATTSON PROPERTIES LLC	0.04	$ 122,800	$ 119,100	$ (3,700)	-3%
MARTIN, KURT W	0.04	$ 105,400	$ 90,100	$ (15,300)	-15%
NORTH CENTRAL STATES REGIONAL	0.08	$ 136,100	$ 149,200	$ 13,100	10%
WILLETTE, ROBERT & DARLENE	0.08	$ 105,000	$ 119,800	$ 14,800	14%
KOERING, PAUL	0.08	$ 100,500	$ 111,500	$ 11,000	11%
FISHER, SCOTT & SUSAN	0.08	$ 98,000	$ 93,400	$ (4,600)	-5%
PJS PROPERTIES, LLC	0.08	$ 88,800	$ 90,800	$ 2,000	2%
MATTHEWS, KEVIN	0.08	$ 68,000	$ 86,600	$ 18,600	27%
S & L PROPERTIES	0.16	$ 238,000	$ 169,000	$ (69,000)	-29%
KOERING, PAUL	0.16	$ 54,700	$ 52,800	$ (1,900)	-3%
S & R PROPERTIES	0.08	$ 19,200	$ 22,200	$ 3,000	16%
TOTAL	**0.96**	**$ 1,136,500**	**$ 1,104,500**	**$ (32,000)**	**-3%**

New Drive Through	Acres	2011	2014	Change ($)	Change (%)
Taco Johns	0.96	$ 803,200	$ 618,500	$ (184,700)	-23%

Source: The cost of auto orientat ion, update, Strong Towns, July 2014

Why This Matters:

Tax productivity informs policy decisions on urban planning, zoning, and municipal services funding

Can we validate this pattern in **Delaware County**?

Unfolding the Story of Tax Productivity: Data, Challenges, and Reproducibility

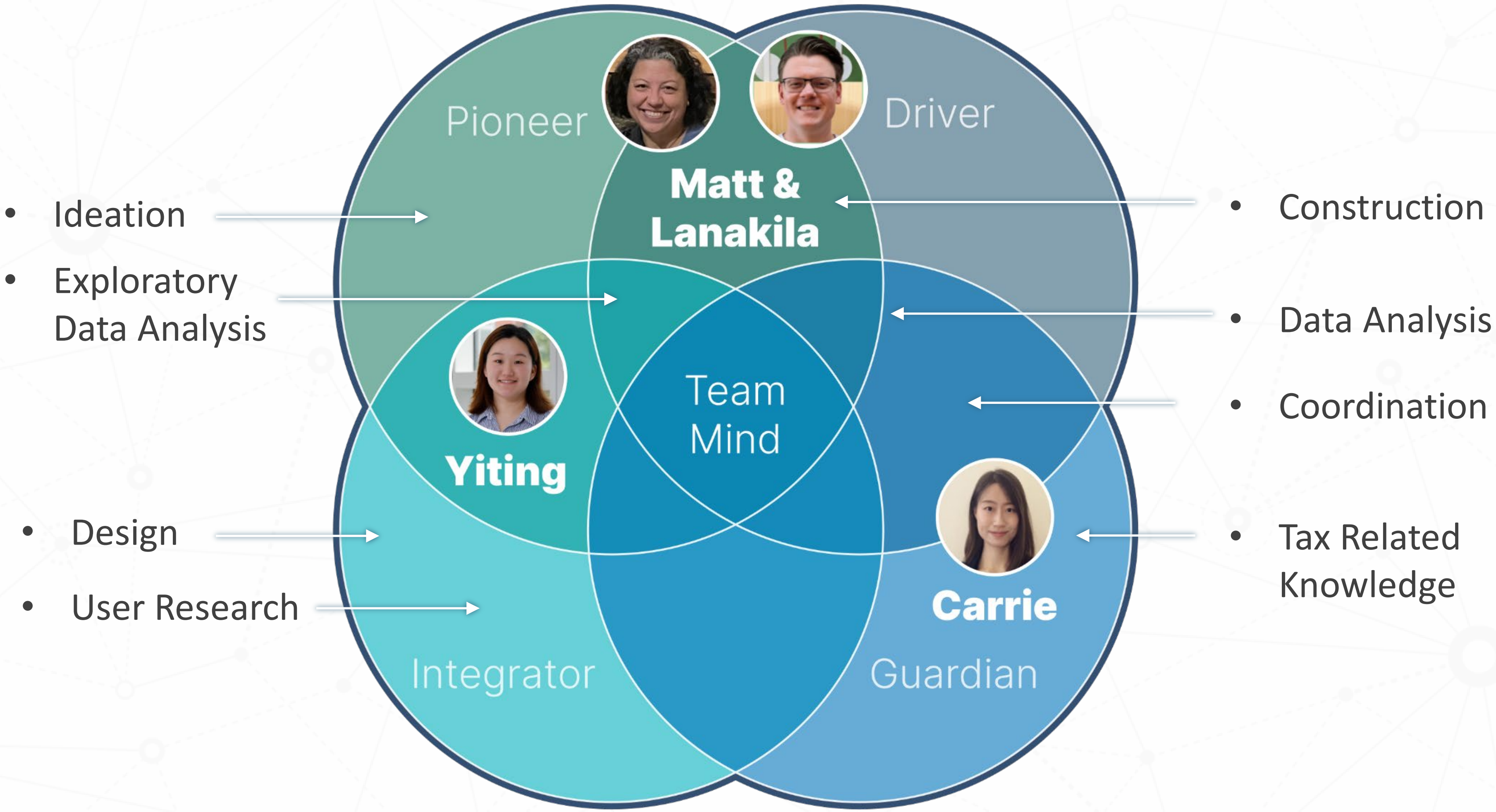
Mid-Ohio Regional Planning Commission (MORPC)

Understanding the Tax Productivity of Development Patterns in Delaware County

Team Members
Lanakila Alexander
Matt Bigelow
Carrie Wei
Yiting Wang

Sponsors
Adam Porr, Research & Data Officer
Jonathan Miller, Principal Planner
Jordan Inskeep, Data Analyst

Our Team and Our Expertise:



What We Will Cover in This Presentation:

1. Why are we doing this project

- Understanding tax productivity's role in urban planning
- Anticipated outcomes

2. How we handled data wrangling

- Data acquisition and cleaning
- Handling anomalies and spatial joins

3. Where we are now

- Key metric: total taxable value per acre
- Current analysis progress

4. What we learned

- Building a reproducible workflow
- Collaborative infrastructure

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

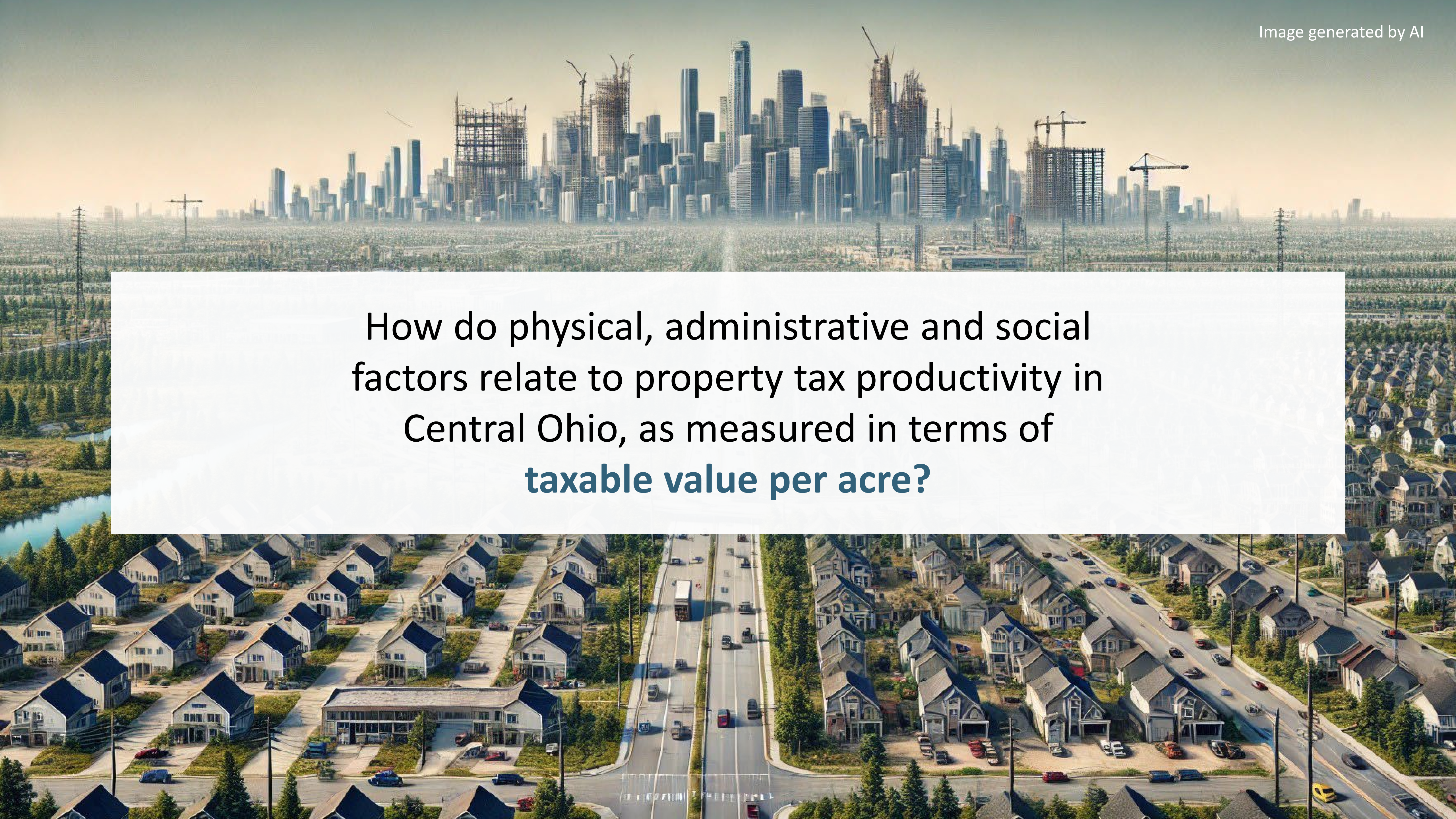
WHY ARE WE DOING THIS PROJECT

Background & Data Scope

1 Competitive market and high poverty rates causing many households to be cost-burdened. **Housing Affordability Crisis**

2 Increasing gap in housing options for low-income families as high-price homes dominate new development. **Limited Affordable Supply**

3 Shifting demographics and lifestyles driving demand for more compact, affordable, and accessible housing options. **Diverse Housing Demand**



How do physical, administrative and social factors relate to property tax productivity in Central Ohio, as measured in terms of **taxable value per acre?**

WHY ARE WE DOING THIS PROJECT

Background & Data Scope

ANTICIPATED BENEFITS

- Evidence for or against mixed-use development, affordable housing, and walkable neighborhoods
- Provide officials with information to make fiscally-responsible decisions

WHY ARE WE DOING THIS PROJECT
Background & Data Scope

ANTICIPATED BENEFITS

- Evidence for or against mixed-use development, affordable housing, and walkable neighborhoods
- Provide officials with information to make fiscally-responsible decisions

PROJECT OBJECTIVES

- Create reproducible processes (using Python and ArcGIS)
- Design compelling visualizations (static or interactive)
- Provide accessible and attractive reports
- Produce thorough technical documentation

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

How to ensure data analysis is **reproducible** for future researches and analysts

HOW WE HANDLED DATA WRANGLING

Data Dictionary

Source of Data

- Delaware County Auditor Parcel GIS Dataset (DCA P)
- Delaware County Auditor Community Reinvestment Areas (DCA CRA)
- Delaware County Condo (DCA Condo)
- MORPC Future Land Use (MORPC L)
- 2020 Census Tabulation Blocks Shapefile with Population and Housing Unit Counts
- 2020 Census Blocks Geodatabase

TITLE	**Delaware County Auditor Parcel GIS Dataset**
LINK	https://drive.google.com/drive/folders/1oLAaZUXfykTwqs9rdYp5ZUx1_PFTg-ry?usp=sharing
DESCRIPTION	A current parcel dataset provided by Delaware County Auditor Employee directly, capturing the latest records of property boundaries, attributes, and details as they stand now. It is designed to provide up-to-date information on parcel characteristics across Delaware County.
Total Records	98,309
Last Updated	10/31/2024
Update Frequency	Unknown

DCA Parcel 2024OCT from Auditor		**DCA Parcel Data from website**					
Attribute Name1	**Data Type1**	**Attribute Name2**	**Data Type2**	**Definition**	**Allow Null Values?**	**Useful? (Y/N)**	**More Info:**
OBJECTID_1	int64	OBJECTID	Int16	Primary Key	N	N	All values are zero.
PARCEL_NO	int64	#N/A	#N/A	Parcel Number	N	Y	
SUB_ID	object	#N/A	#N/A	Subdivision ID ?	Y	TBD	All values are zero.
SPLITDATE	int64	SPLITDATE	Int16	Split Date - ? YYYMMDD Format	N	TBD	
SUB_NAME	object	SUB_NAME	String	Subdivision Name	Y	Y	
CONDO	object	CONDO	String	Condo Name	Y	Y	Used to possibly easily identify Condo complexes
MAPSHEET	int64	MAPSHEET	String	? Map Sheet	N	TBD	
SUBPARCEL	int64	SUBPARCEL	String	Sub Parcel Number	N	Y	
NGHBRHDCD	object	NGHBRHDCD	String	Neighborhood Code	Y	N	
MAILNAME1	object	MAILNAME1	String	Tax Mailing Name of the owner	Y	N	
MAILNAME2	object	MAILNAME2	String	Tax Mailing Name of the owner	Y	N	
MAILADDR1	object	MAILADDR1	String	Tax Mailing Address of the owner	Y	N	
MAILADDR2	object	MAILADDR2	String	Tax Mailing Address 2 of the owner	Y	N	
OWNER1	object	OWNER1	String	Owner Name of Parcel	N	Y	
OWNER2	int64	OWNER2	String	Owner 2 of Parcel	Y	N	All values are zero.
ADDR1	object	ADDR1	String	Property Address	N	Y	
ADDR2	object	ADDR2	String	Property Address Line 2	Y	N	
TAXDIST	float64	TAXDIST	Int16	Tax District	N	Y	
LEGAL	object	LEGAL	String	Legal Description	Y	TBD	
CLASS	object	CLASS	String	Parcel Class, also called Use Code	Y	Y	
ACRES	float64	ACRES	Double	Total Acreage	N	Y	
MARKET_LAN	int64	MARKET_LAN	Int16	Market Land Value	N	Y	
MARKET_IMP	int64	MARKET_IMP	Int16	Market Improvement Value	N	Y	
MARKET_TOT	int64	MARKET_TOT	Int16	Market Total Value	N	Y	
TAXABLE_LA	float64	TAXABLE_LA	Int16	Taxable Land Value	N	Y	
TAXABLE_IM	float64	TAXABLE_IM	Int16	Taxable Improvement Value	N	Y	
TAXABLE_TO	float64	TAXABLE_TO	Int16	Taxable Total Value	N	Y	
CAUV	int64	CAUV	Int16	CAUV Value - Current Agricultural Use Value	N	Y	
ANNUALTAX	float64	ANNUALTAX	Int16	Annual Tax	N	Y	
YRBUILT	int64	YRBUILT	Int16	Year Built	N	Y	
YRREMOD	int64	YRREMOD	Int16	Year Remodeled	N	Y	
ROOMS_TOT	int64	ROOMS_TOT	Int16	Total Rooms Count	N	Y	
BDROOMS	int64	BDROOMS	Int16	Bedrooms	N	Y	
FAMROOMS	int64	FAMROOMS	Int16	Family Rooms Count	N	Y	
DINROOMS	int64	DINROOMS	Int16	Dining Rooms Count	N	Y	
HAFBATHS	int64	HAFBATHS	Int16	Half Bathrooms Count	N	Y	
FULBATHS	int64	FULBATHS	Int16	Full Bathrooms Count	N	Y	

HOW WE HANDLED DATA WRANGLING

Key Challenges in Data Wrangling

INITIAL CHALLENGES

- Obtaining data with Parcel ID
- Handling duplicate records
- Addressing outliers
- Resolving parcel overlaps
- Standardizing units and calculations

HOW WE HANDLED DATA WRANGLING

Key Challenges in Data Wrangling

INITIAL CHALLENGES		KEY SOLUTIONS DEVELOPED
Obtaining data with Parcel ID	→	Communicate with the auditor office
Handling duplicate records	→	Spatial join methodology
Addressing outliers	→	Outlier detection package
Resolving parcel overlaps	→	Data cleaning pipeline
Standardizing units and calculations	→	Validation procedures

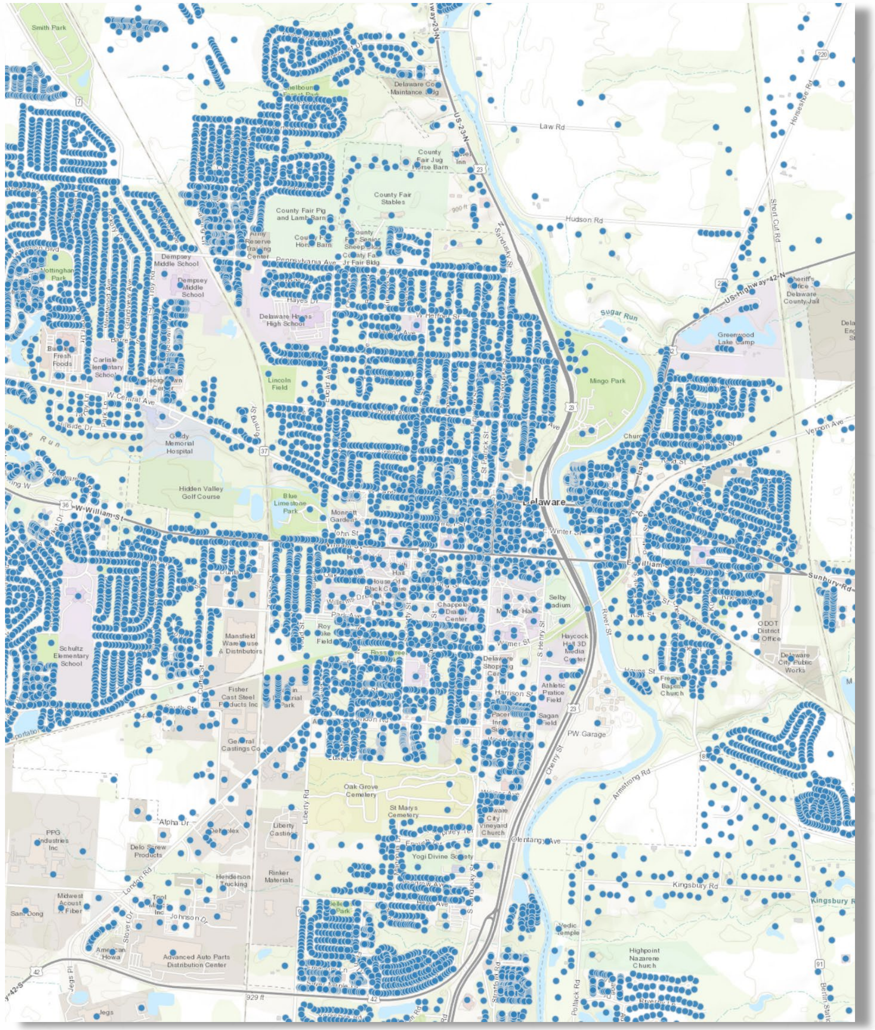
HOW WE HANDLED DATA WRANGLING
Initial EDA and Scoping

Methods

- Analyze the Delaware County Auditor's Parcel dataset, which includes tax and polygon data for all parcels in the county
- Noticed lots of zero values so brought in Condo dataset
- Spatially join all the GIS datasets into single file

Results

- Single dataset to allow for easier manipulation
- The spatial join of the Address Point and Parcel datasets led to a significant number of duplicated records, after reviewing what all the address point dataset brought to the analysis, we decided it wasn't worth the effort. This can be revisited if value is found.

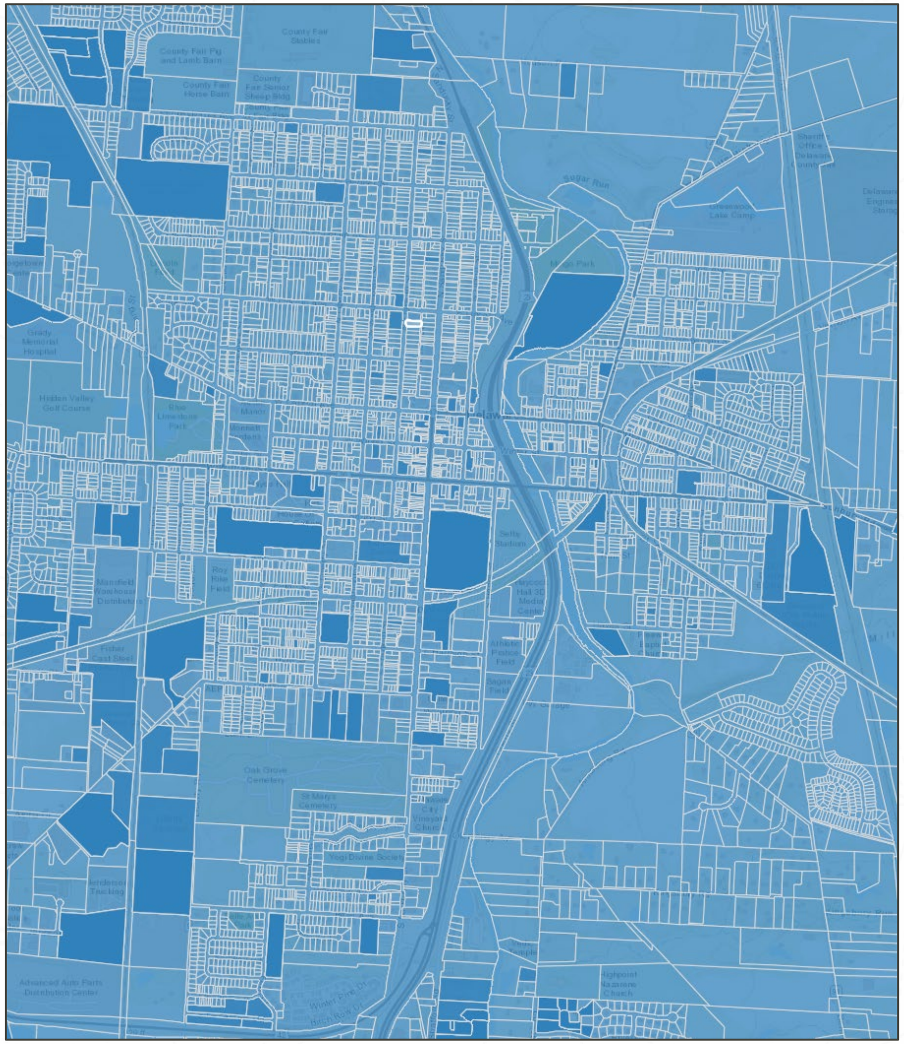


HOW WE HANDLED DATA WRANGLING

Data Collection and Initial Inspection - Parcel Data

Sources

- **Delaware County Auditor's Parcels DB:**
  Public Information Request required for unblinded Parcel IDs
- **Condo Data**:
  Downloaded and spatially joined for use in future analysis
- **Community Reinvestment Areas (CRA)**:
  Added to look at possible annual tax anomalies
- **Street Centerlines:**
  Hasn't been added yet, but likely in future



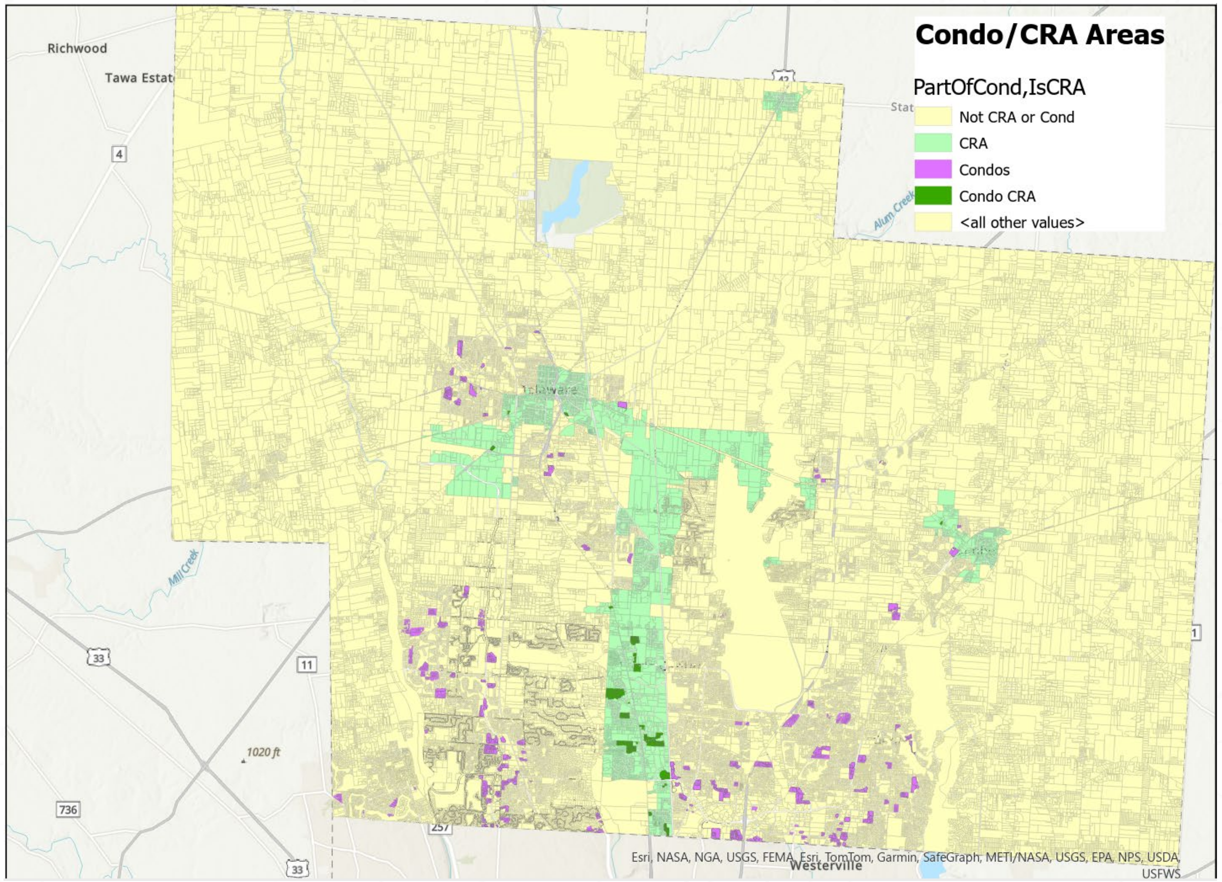
HOW WE HANDLED DATA WRANGLING
Data Integration - Preprocessing and Joining

Spatial Joining:

- **Geo files were opened in ArcGIS Pro**
- Spatial joins were used to combine the data
- This enables the creation of condo and CRA 'flags' for each parcel

Save/Export for Analysis in Python:

- Once joined, the new geodatabase files were saved and exported from ArcGIS Pro



*Larger image available in the appendix

HOW WE HANDLED DATA WRANGLING
Ingestion and Cleaning

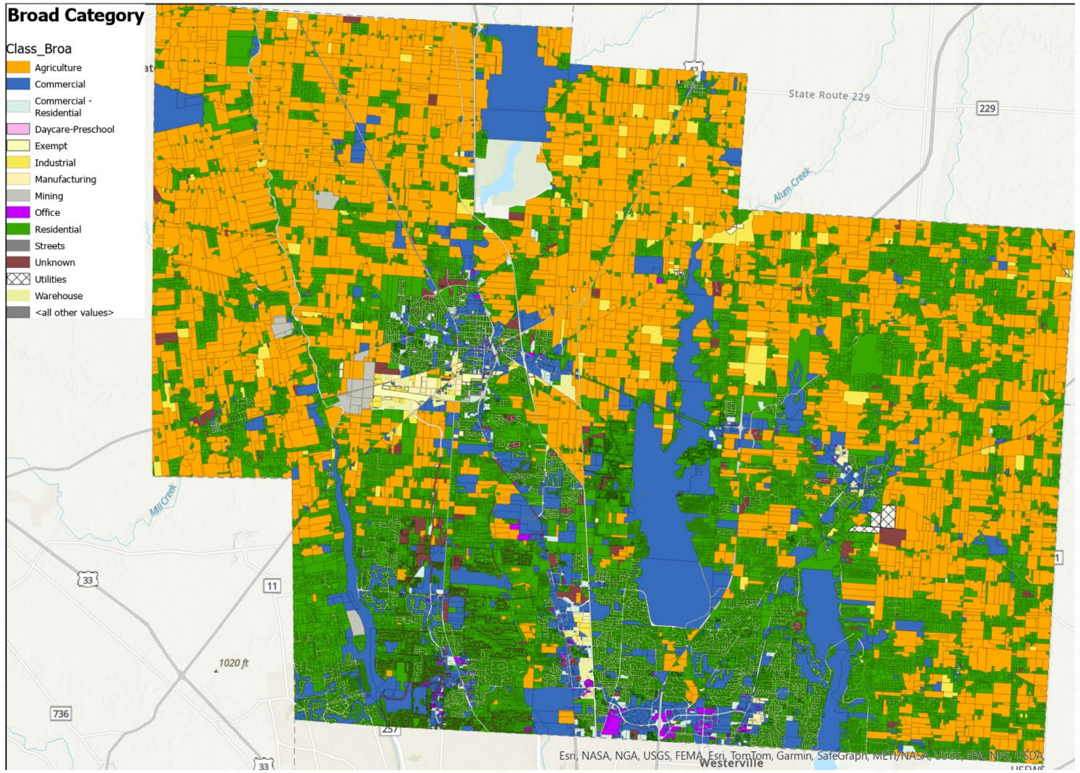
Data Ingestion:
- Geo file loaded using GeoPandas

Column Cleanup/Renaming:
- Useless columns were removed
- Remaining columns renamed to human friendly names

Duplicate removal:
- Duplicate rows were removed

Calculated Columns:
- Useful metrics such as total acreage and Taxable Value per Acre were created for each row

Class Categories:
- MORPC provided class description file was merged with the dataframe
- Broad and Specific categories were added by our team



*Larger image available in the appendix

HOW WE HANDLED DATA WRANGLING
Data Cleanup - Parcel Overlap Analysis

Parcel Overlap Detection:

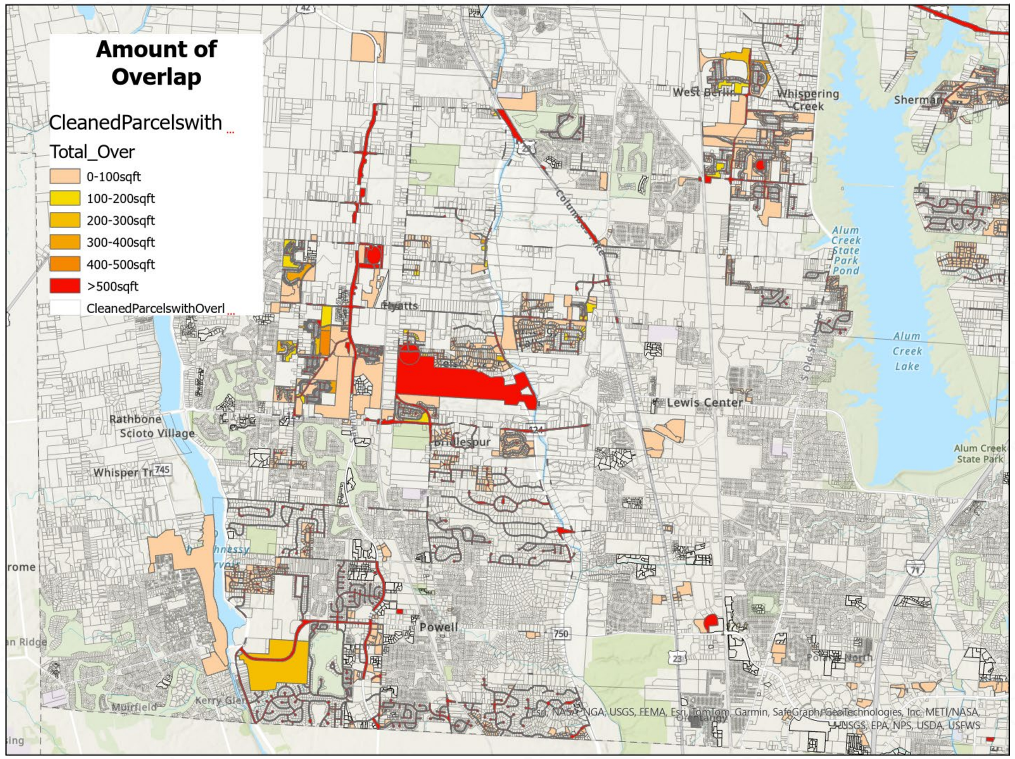
- Used sindex.query() from GeoPandas to look for overlapping geometries
- Utilized parallel processing to do this in ~20 minutes vs many hours

~2500 overlaps were identified:

- Any overlap less than 100ft2 was ignored
- This reduced overlap count to 75

Visual Analysis:

- Flags were added to overlapping parcels during python analysis so that they could be easily visualized in ArcGIS Pro
- Roads, oddly shaped developer parcels and parcels with identical geometries are key culprits
- Further discussion on how to handle these parcels to come



*Larger image available in the appendix

HOW WE HANDLED DATA WRANGLING

Parcel Data Cleanup - Duplicate Geometries

- Finishing data cleanup proved some challenges. We easily removed duplicate parcel IDs, but while checking for parcel overlap, we found many geometries that were duplicated. A strategy was discussed with our sponsor on how to deal with them:
    - If there is a group of duplicate geometries and each parcel ID has its own unique taxable values, the parcel size would be divided by total number of parcels that share that geometry and that would be the new calculated acreage for those parcels.
    - The Taxable Value per Acre would then be recalculated based on the new area

- There were a about 53 overlaps/duplicates that weren't handled and were flagged as anomalies out of 92,427.
    - These mostly consisted of new construction neighborhoods and office/condo sites.

HOW WE HANDLED DATA WRANGLING
Creating an Outlier Detection Package

STATISTICAL:

Standard z-score (+/- 3) and IQR were created to detect outliers statistically.

SPATIAL:

1. **Isolation Forest**
   - Fast processing method (linear time complexity and low memory usage)
   - Works well with higher dimensional data
   - Non-parametric so does not assume normal distributions like z-score and IQR do
2. **Local Outlier Factor (LOF)**
   - Detects outliers relative to their local neighborhood, in our case, geographic areas
   - No need for clustering/grouping ahead of time
   - Provides and outlier ‘score’ giving the ability to measure “outlierness” so we can set certain thresholds
   - By including parcel centroids, this method can highlight parcels that deviate significantly from their “local” peers
3. **Density-Based Spatial Clustering (DBSCAN)**
   - No assumptions about distributions
   - Explicitly finds clusters of similar parcels and labels points in lower density areas as outliers
   - Robust to noise: unlike k-means, doesn’t force all points into a cluster which helps distinguishing true outliers

HOW WE HANDLED DATA WRANGLING
Challenges in Integrating Census Data

Currently using 6 INPUT files:

1. **2010 Census Block contains Population and Housing Unit data**
2. **2020 Census Block**
3. **2020 Demographic and Housing Characteristics**
   - Population
   - Housing Tenure
   - Race
   - Income
   - Poverty Level
4. **Parcel data from Delaware County Auditor**
5. **American Community Survey (ACS)**

It is anticipated that additional data will be required to do time series analyses

HOW WE HANDLED DATA WRANGLING
Integrating Census Data

TIGER/Line Shapefiles for Block and Block Group (BG) used to identify specific GEOIDs for boundaries.

Geodatabase Demographic and Housing Characteristics (DHC)

- GEOIDs will be used to join data files containing Housing Unit counts, vacancy status, tenure of occupants, population, income, and poverty.

Significant data pre-processing using ArcGIS Pro

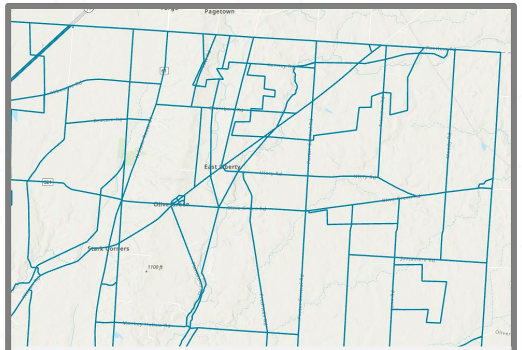
- Parcel data separated into layers specific to land use to facilitate future analysis within Census boundaries.
- Spatial Join of parcel layers and TIGER files to associate GEOIDs to each parcel.

HOW WE HANDLED DATA WRANGLING

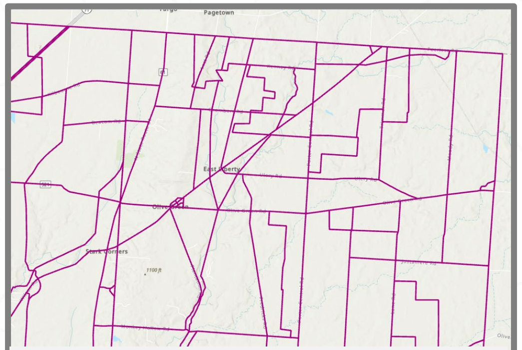
Example of Challenge in Integrating Census Data

In preparation for socioeconomic time series analyses, the Census Block is the smallest geography for which data is published. For our purposes, we'd like to look at parcel data aggregated by Block.

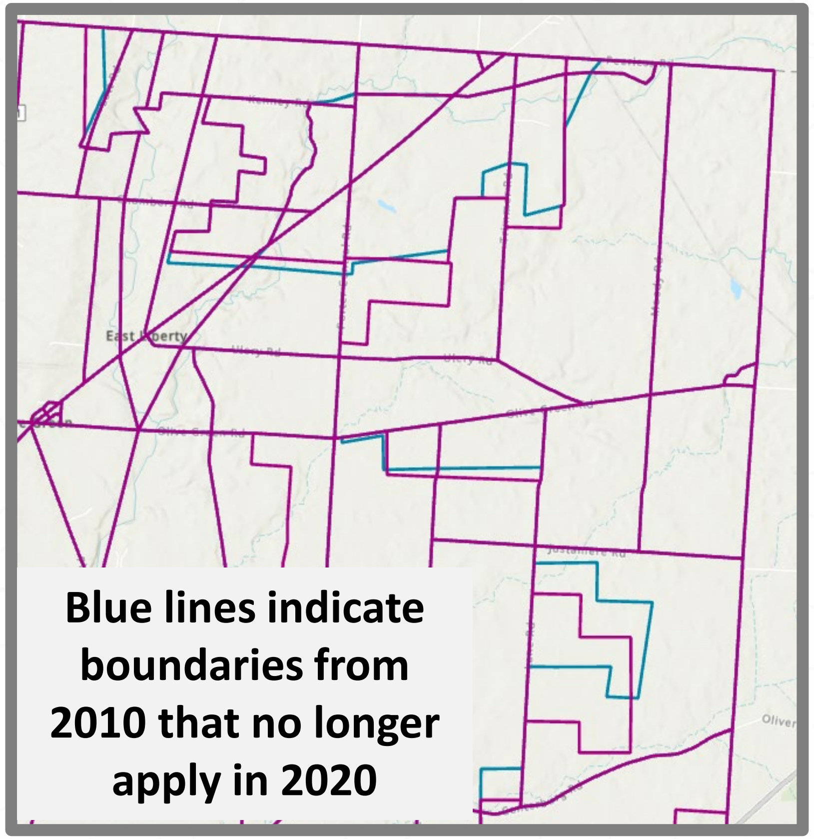
The difficulty we are encountering is the movement of Block geography between decennial censuses



2010 Blocks



2020 Blocks



Blue lines indicate boundaries from 2010 that no longer apply in 2020

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

WHERE WE ARE NOW

Introduction to Tax-Related Attributes

Definition

Annual Tax	VS	Total Taxable Value
Property tax that a property owner is required to pay each year Based on the taxable value of the property and the applicable tax rates		The assessed value of a property that is subject to taxation 35% of the property's true market value

WHERE WE ARE NOW

Why Total Taxable Value is a Better Measure?

Annual Tax	VS	Total Taxable Value
Affected by tax rates, abatements, and tax exemptions		Reflect the full taxable base without those temporary factors
Can change over time		Provide a stable, consistent measure of property value
Misrepresents a property's true contribution		

WHERE WE ARE NOW

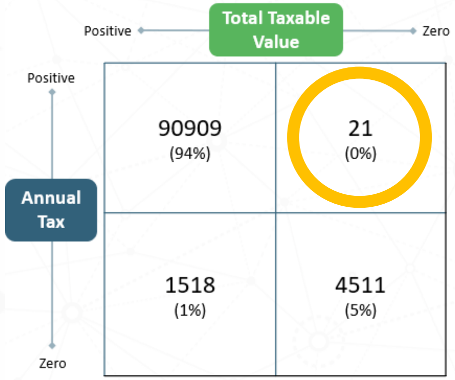
Dividing the Dataset into Four Groups

Annual Tax \ Total Taxable Value	Positive	Zero
Positive	90909 (94%)	21 (0%)
Zero	1518 (1%)	4511 (5%)

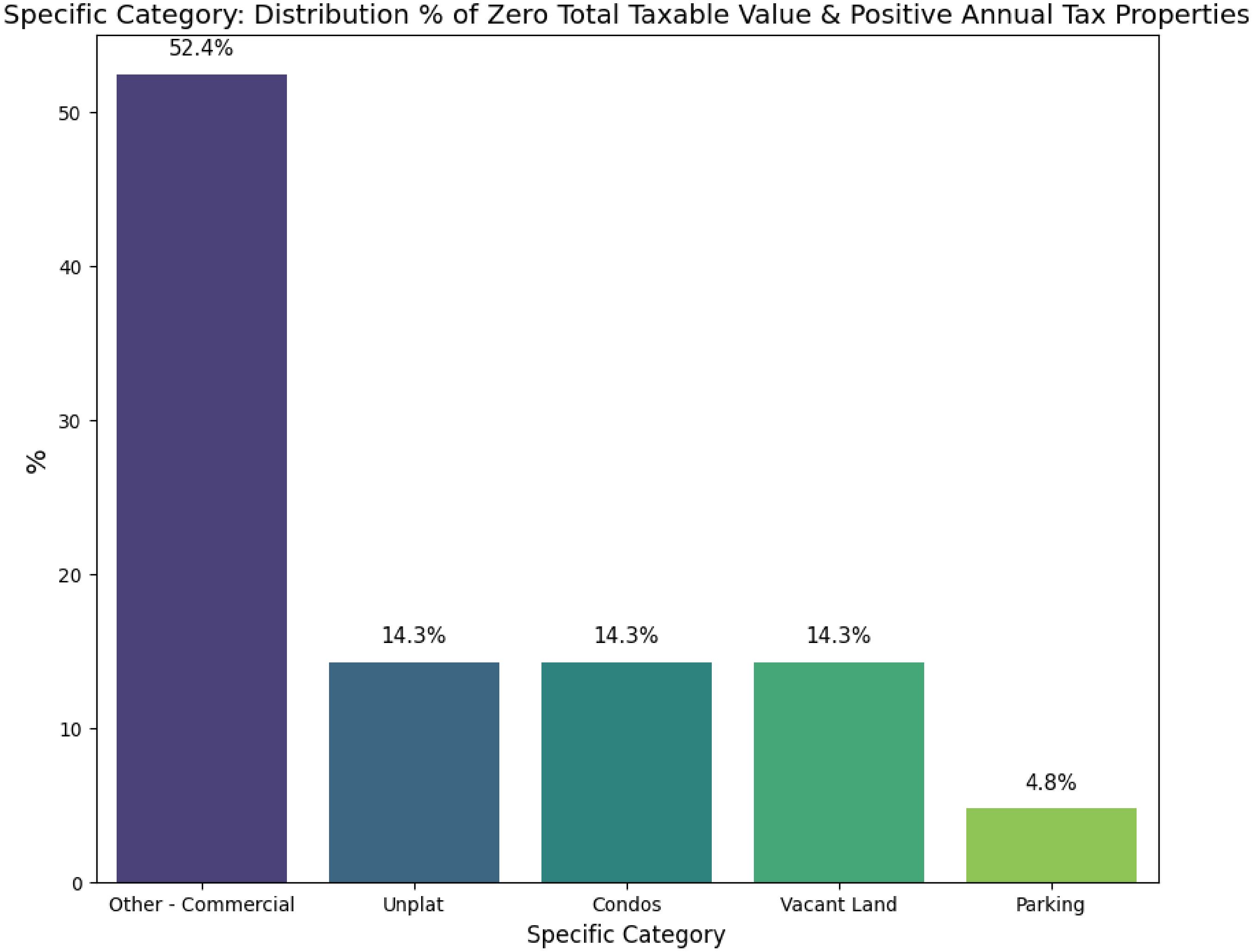
Errors?

WHERE WE ARE NOW

Properties with Zero Taxable Value - Positive Annual Tax (21entries)



- Majority are commercial structures, vacant land, and residential unplatted properties.
- Positive annual tax arises from shared property fees, utility charges, or special assessments.

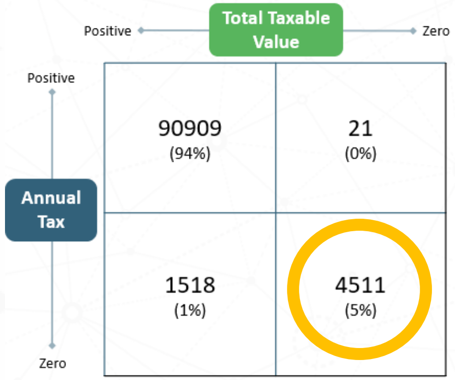


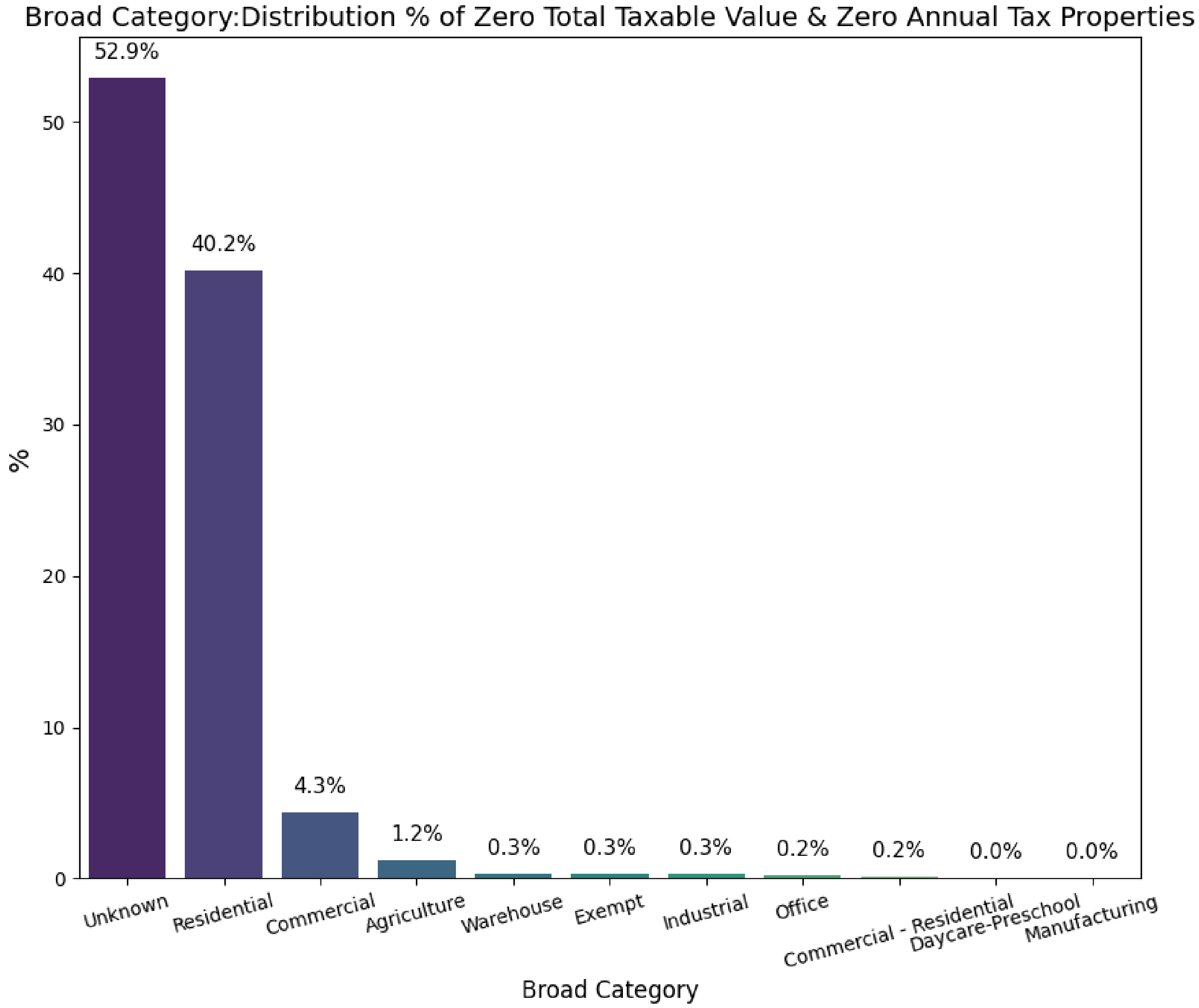
Ohio Department of Taxation. (n.d.). Property tax real property. Retrieved from *https://dam.assets.ohio.gov/image/upload/tax.ohio.gov/communications/publicat*

WHERE WE ARE NOW

Properties with Zero Taxable Value - Zero Annual Tax

(4511 entries)



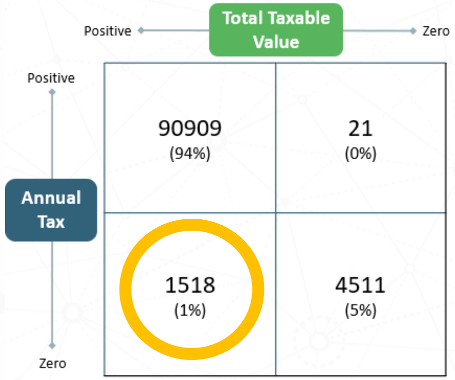


- Most properties of this group (52.9%) are categorized as 'unknown', and 40.2% are residential.

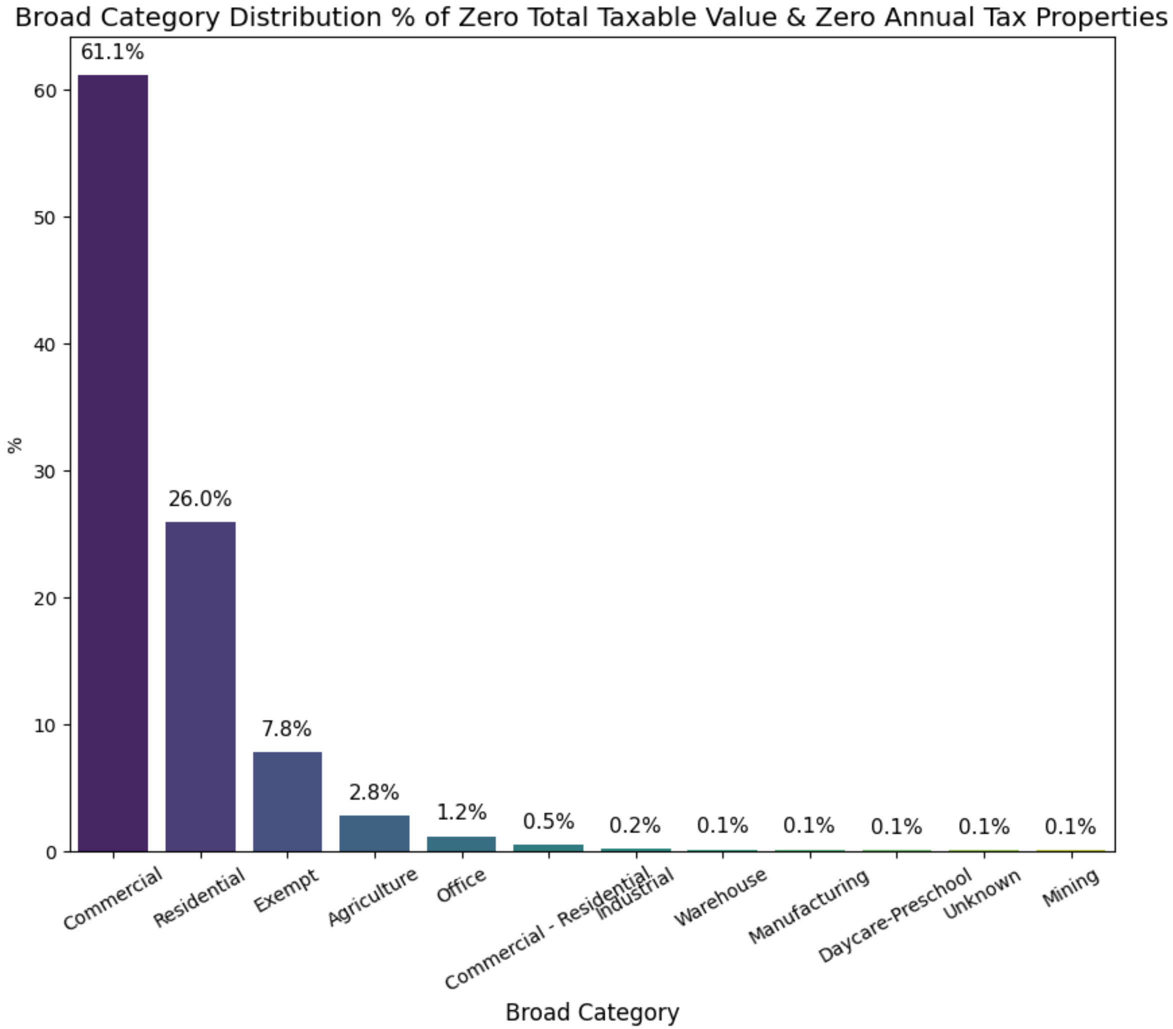
WHERE WE ARE NOW

Properties with Positive Taxable Value - Zero Annual Tax

(1518 entries)



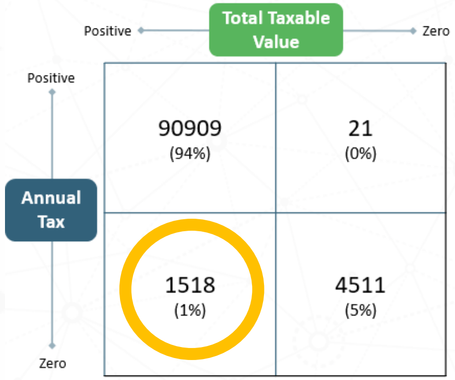
- Mainly commercial (61.1%) and residential (26%) properties.

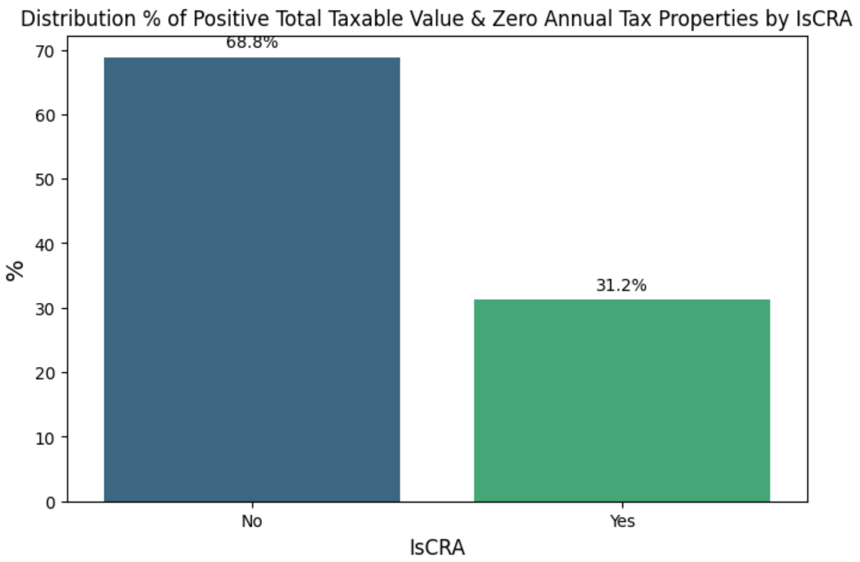


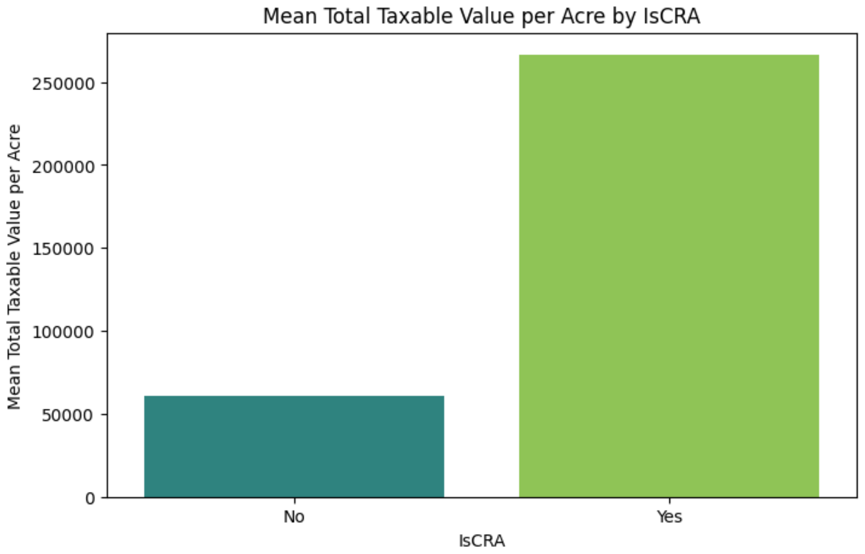
WHERE WE ARE NOW

Properties with Positive Taxable Value - Zero Annual Tax

(1518 entries)







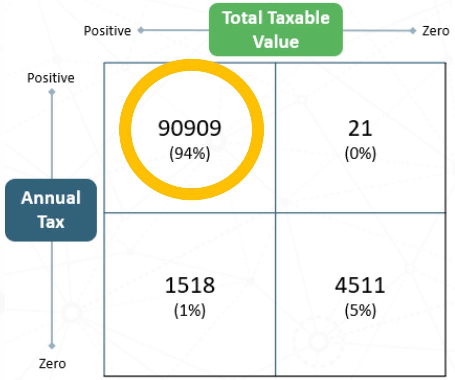
- Mainly commercial (61.1%) and residential (26%) properties.

- Properties in the CRA program (31.2% of total) have a mean taxable value per acre four times higher than non-participants.

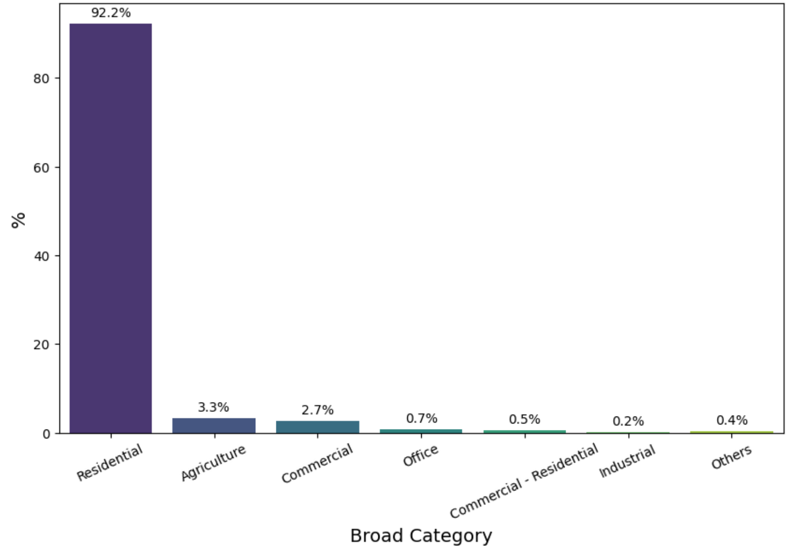
WHERE WE ARE NOW

Largest Group - Properties with Positive Taxable Value -Positive Annual Tax

(90909 entries)

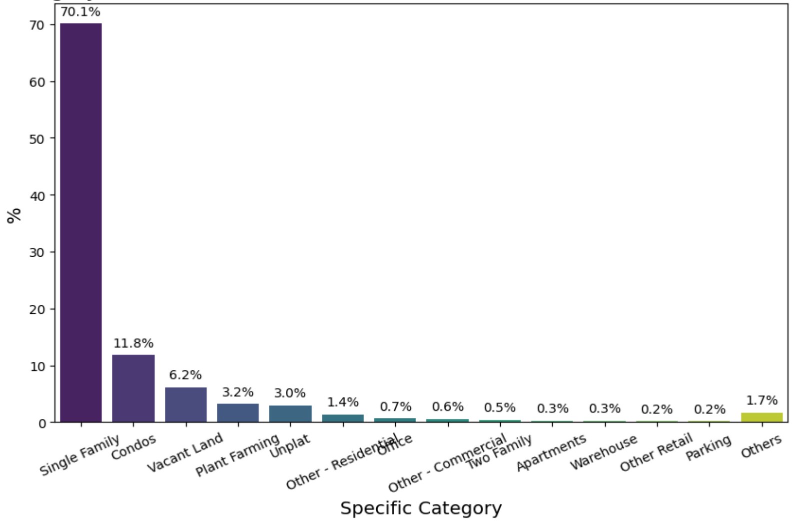


Broad Category: Distribution % of Positive Total Taxable Value & Positive Annual Tax Properties



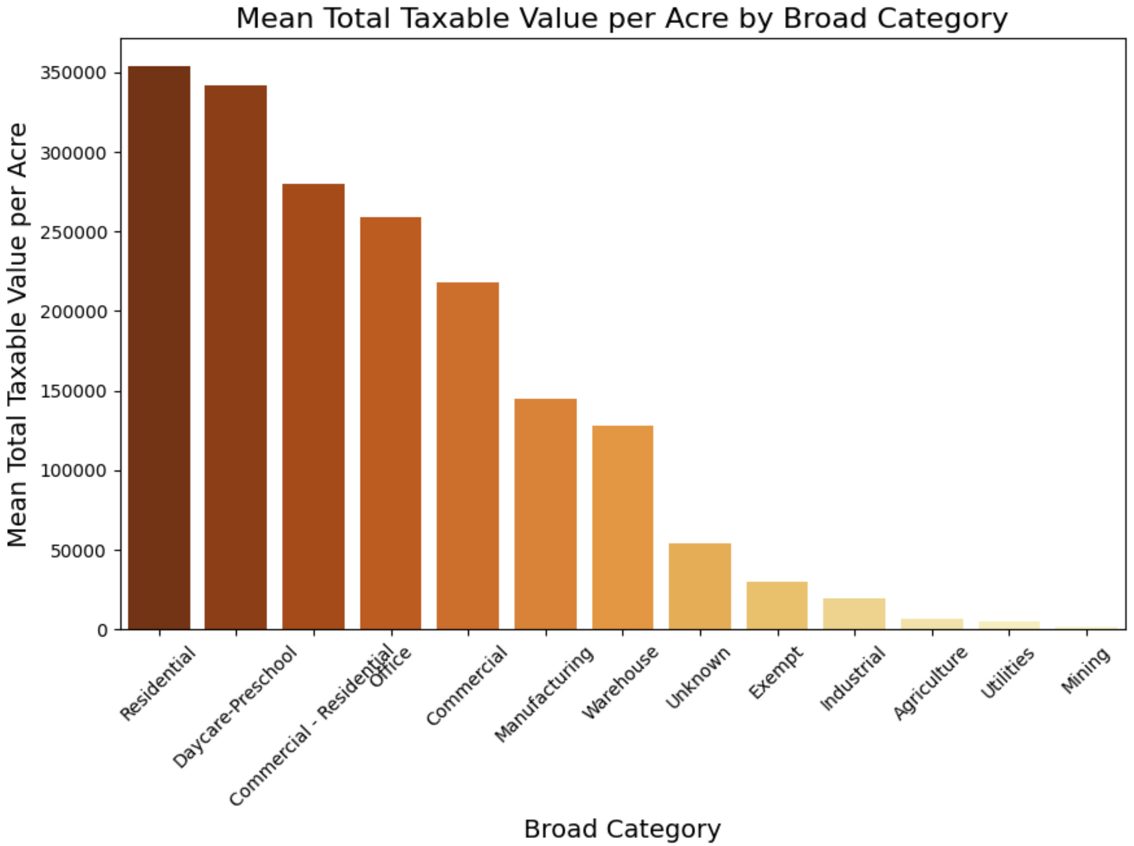
- The data distribution for this group reveals that 92.2% of the entries are Residential properties, with 70.1% categorized as Single-Family homes and 11.8% as Condos

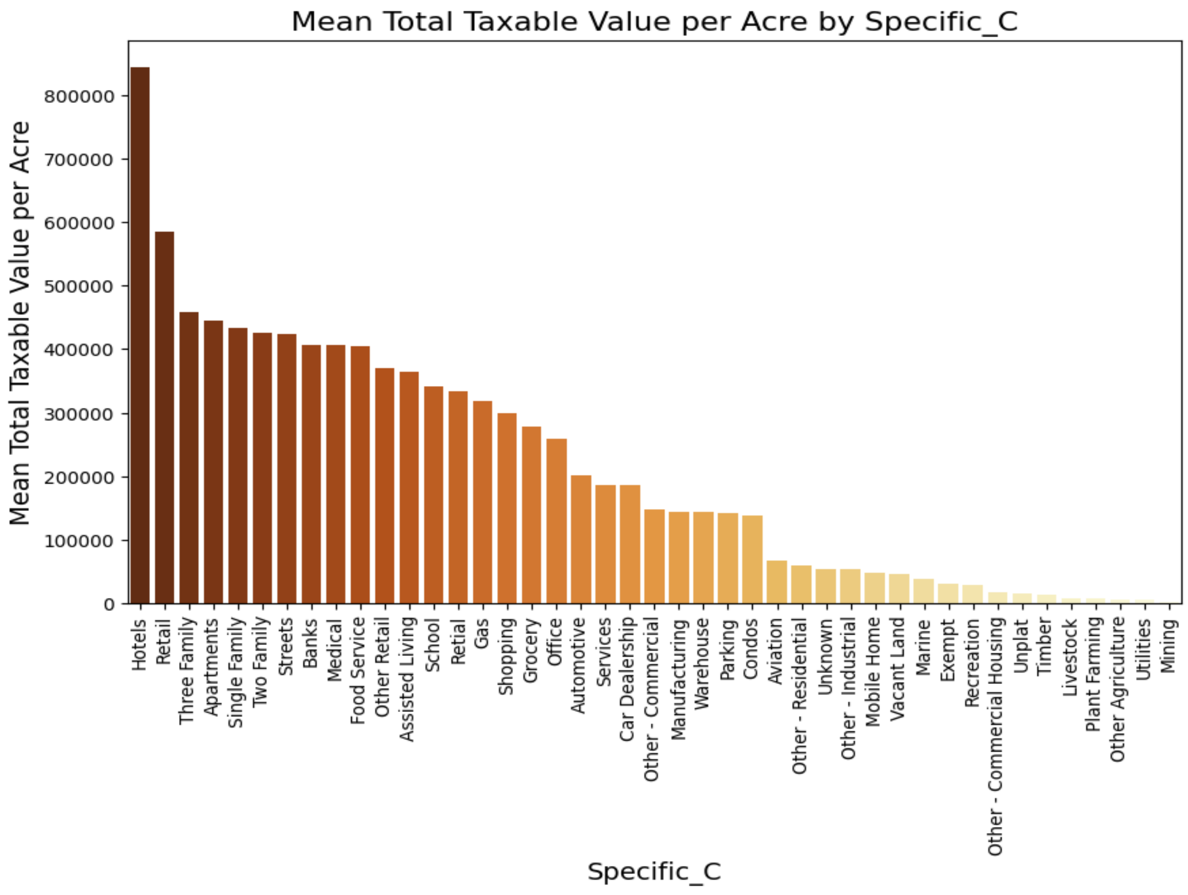
Specific Category: Distribution % of Positive Total Taxable Value & Positive Annual Tax Properties



WHERE WE ARE NOW
Positive Taxable Value -Positive Annual Tax (90909 entries)

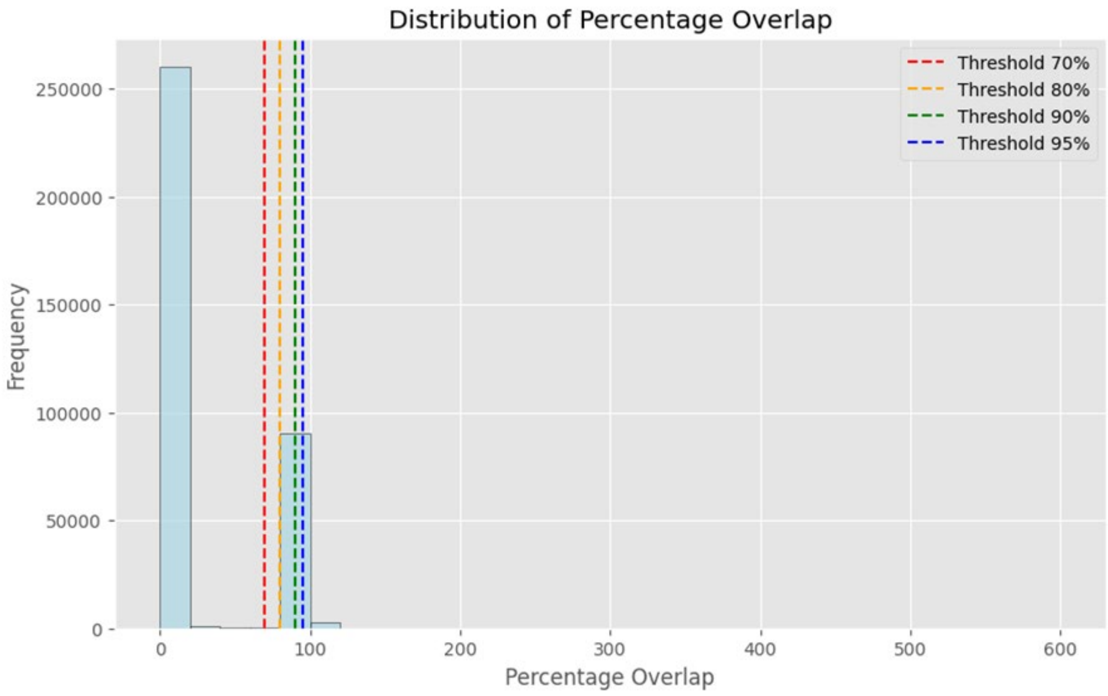
- **High taxable values per acre**: Residential, Daycare, and Commercial-Residential.
- **Top categories**: Hotels and Retail, followed by Residential subcategories.





WHERE WE ARE NOW
Integrating MORPC Land Use Dataset

- **Goal:** Enhance classification by integrating the MORPC Land Use Dataset.
- **Challenges:** No direct attribute linking MORPC to Delaware County parcels; requires spatial joins.
- **Method:** Conduct spatial join, assign land use codes with ≥ 90% overlap for accuracy.
- **Limitation:** Mismatch between 2021 MORPC data and current parcel data causes discrepancies.
- **Next Steps:** validate this spatial join at both the parcel and neighborhood levels.



WHERE WE ARE NOW
Development Patterns Over Time

Tax Productivity Over Time
by Development Eras and Broad Classification

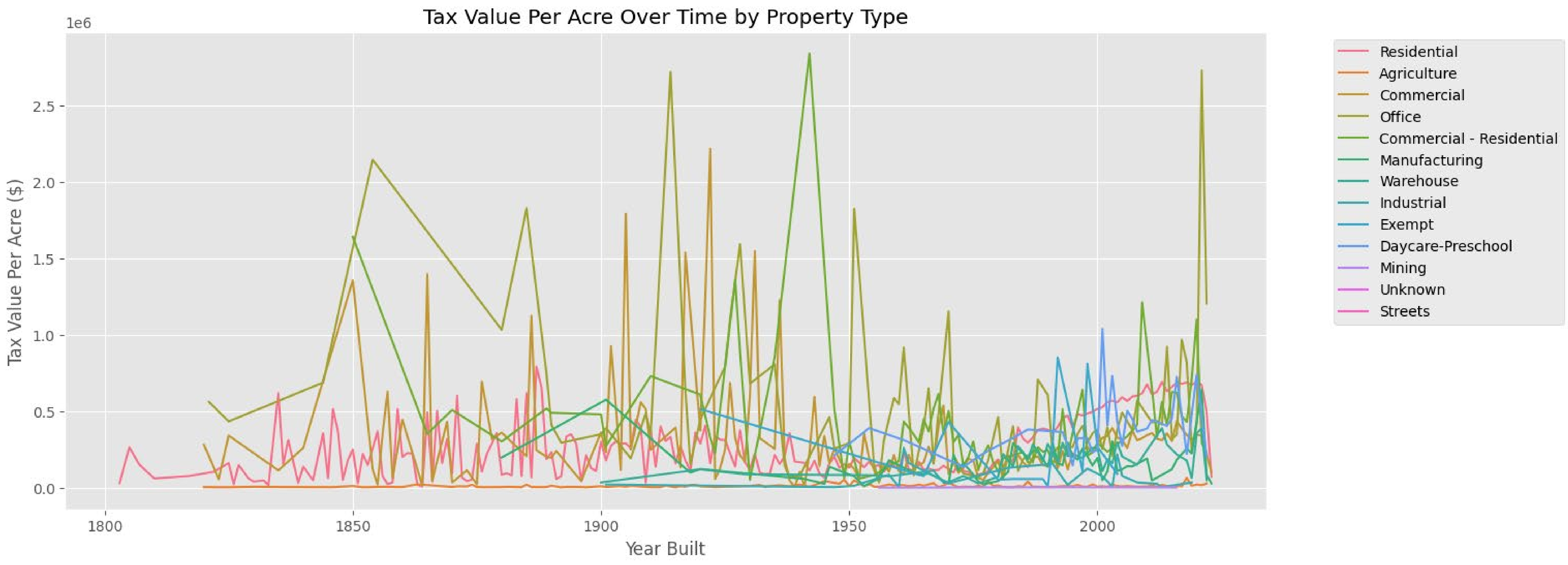
Remodeling Trends
(YRREMOD vs. YRBUILT)

CHALLENGES AND PIVOTS

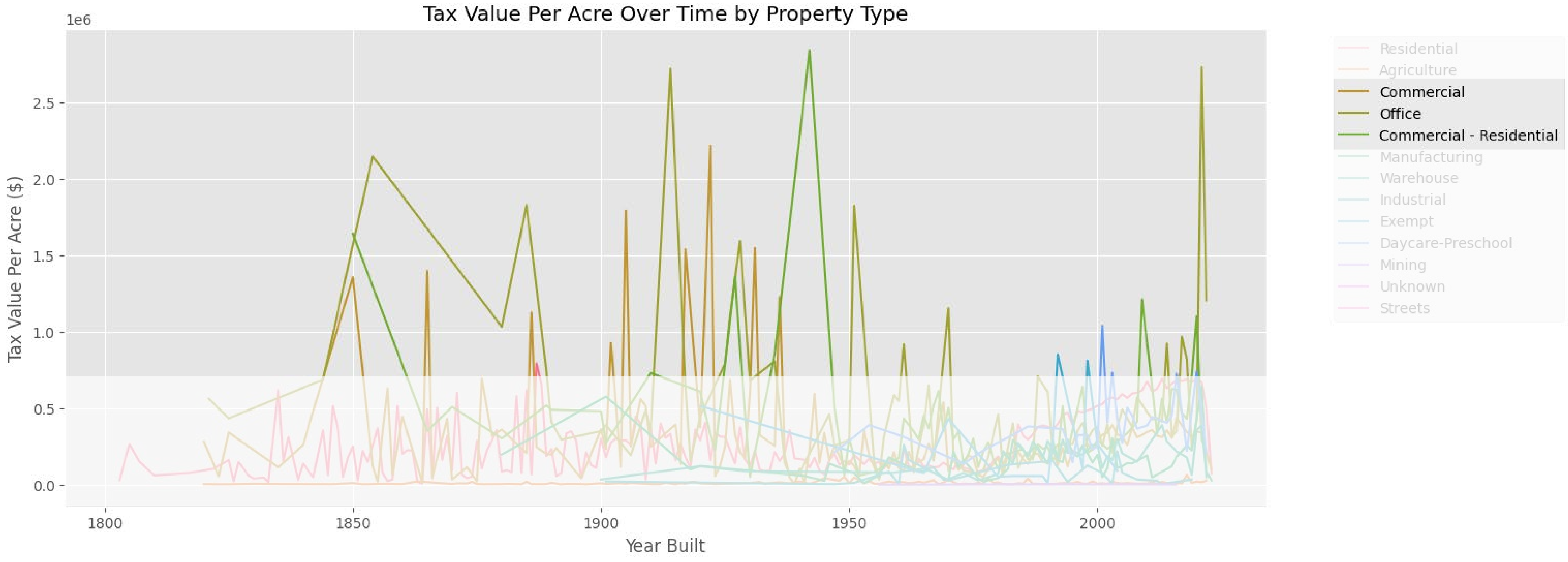
- **Inflation rate:** Initially explored inflation adjustment approaches (Investigated CPI index integration or 3% standard inflation rate), discovered property values already reflect current market conditions through regular reappraisals.

- **Time period grouping:** Started with decade-based analysis, shifted to development eras for better pattern visibility; Created four distinct eras based on historical development patterns.

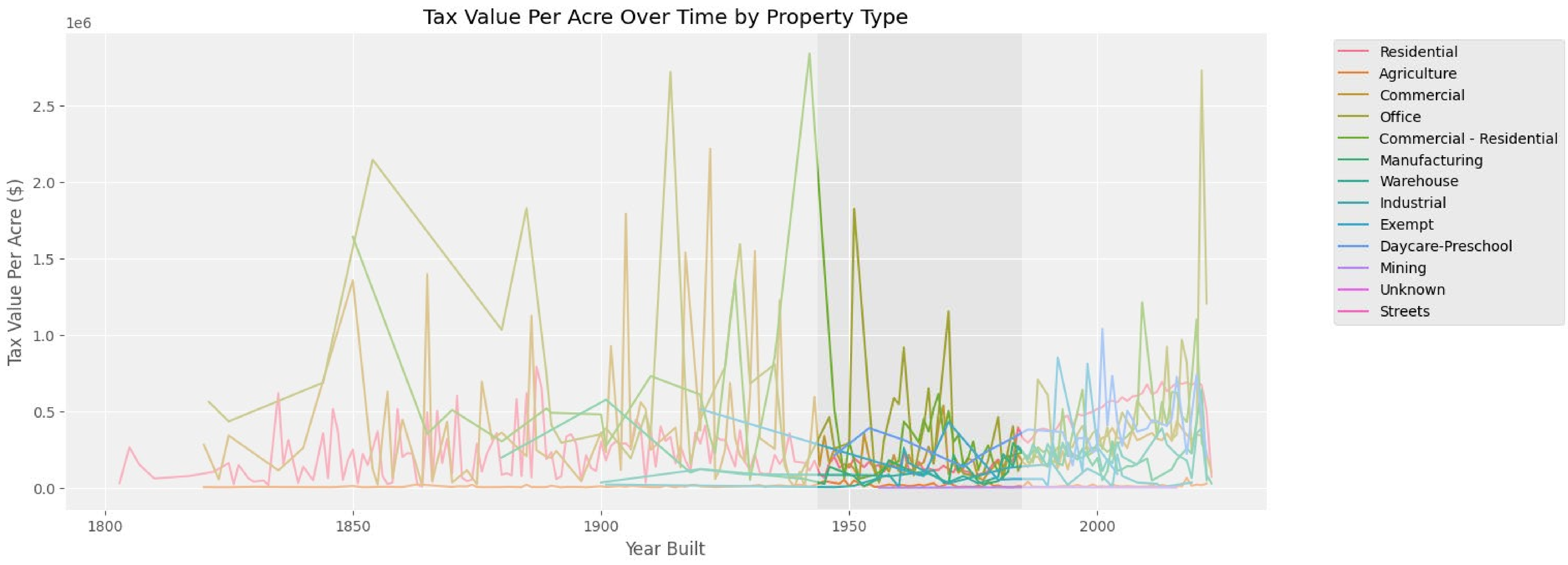
WHERE WE ARE NOW
Development Patterns Over Time



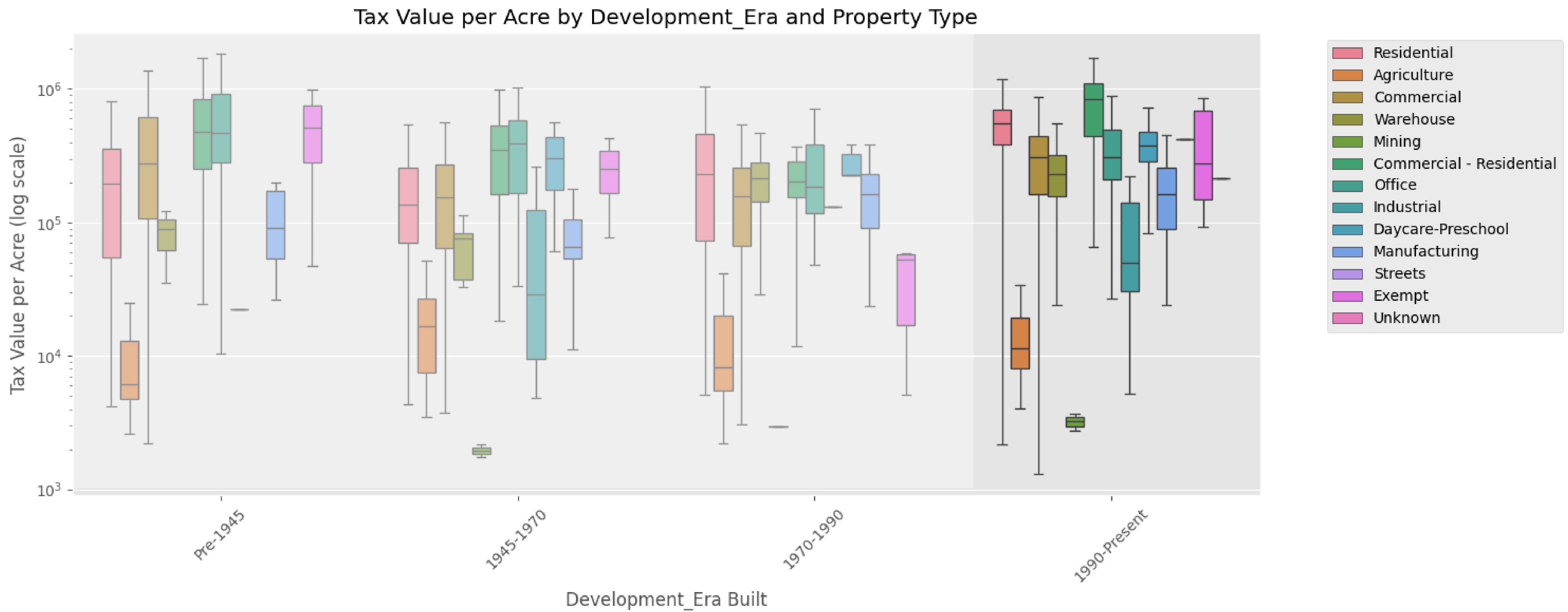
WHERE WE ARE NOW
Development Patterns Over Time



WHERE WE ARE NOW
Development Patterns Over Time



WHERE WE ARE NOW
Development Patterns Over Time



WHERE WE ARE NOW
Development Patterns Over Time

WHERE WE ARE NOW

Development Patterns Over Time – Remodel Status

- Commercial-Residential properties consistently show the highest tax values per acre
- Post-WWII development (1945-1970) generally has lower tax productivity
- Remodeling generally boosts tax productivity, but the impact varies by property type

			TaxValAcre			Calc_Acres
			count	median	mean	median
Development_Era	Class_Broa	Remodel_Status				
1990-Present	Commercial - Residential	Original	244	846722.79	827218.04	0.04
Pre-1945	Commercial - Residential	Remodeled	55	567543.13	757105.12	0.17
1945-1970	Daycare-Preschool	Original	1	566615.52	566615.52	0.24
1990-Present	Residential	Original	52614	553178.43	551511.92	0.27
Pre-1945	Exempt	Remodeled	2	515672.60	515672.60	2.66
	Office	Remodeled	73	498909.79	737852.75	0.17
1990-Present	Residential	Remodeled	2307	495760.04	474078.26	0.34
1945-1970	Commercial - Residential	Remodeled	12	445876.93	370361.94	0.62
	Exempt	Original	1	432061.75	432061.75	0.17
1990-Present	Streets	Original	1	424145.51	424145.51	0.60

Top 10 Most Tax-Productive Development Patterns

WHERE WE ARE NOW
Development Patterns Over Time – Remodel Status

- Commercial-Residential properties consistently show the highest tax values per acre
- Post-WWII development (1945-1970) generally has lower tax productivity
- Remodeling generally boosts tax productivity, but the impact varies by property type

			TaxValAcre			Calc_Acres
			count	median	mean	median
Development_Era	Class_Broa	Remodel_Status				
1990-Present	Commercial - Residential	Original	244	846722.79	827218.04	0.04
Pre-1945	Commercial - Residential	Remodeled	55	567543.13	757105.12	0.17
1945-1970	Daycare-Preschool	Original	1	566615.52	566615.52	0.24
1990-Present	Residential	Original	52614	553178.43	551511.92	0.27
Pre-1945	Exempt	Remodeled	2	515672.60	515672.60	2.66
	Office	Remodeled	73	498909.79	737852.75	0.17
1990-Present	Residential	Remodeled	2307	495760.04	474078.26	0.34
1945-1970	**Commercial - Residential**	**Remodeled**	**12**	**445876.93**	**370361.94**	**0.62**
	Exempt	**Original**	**1**	**432061.75**	**432061.75**	**0.17**
1990-Present	Streets	Original	1	424145.51	424145.51	0.60

Top 10 Most Tax-Productive Development Patterns

WHERE WE ARE NOW

Development Patterns Over Time – Remodel Status

- Commercial-Residential properties consistently show the highest tax values per acre
- Post-WWII development (1945-1970) generally has lower tax productivity
- Remodeling generally boosts tax productivity, but the impact varies by property type

			TaxValAcre			Calc_Acres
			count	median	mean	median
Development_Era	Class_Broa	Remodel_Status				
1990-Present	Commercial - Residential	Original	244	846722.79	827218.04	0.04
Pre-1945	Commercial - Residential	Remodeled	55	567543.13	757105.12	0.17
1945-1970	Daycare-Preschool	Original	1	566615.52	566615.52	0.24
1990-Present	Residential	Original	52614	553178.43	551511.92	0.27
Pre-1945	Exempt	Remodeled	2	515672.60	515672.60	2.66
	Office	Remodeled	73	498909.79	737852.75	0.17
1990-Present	Residential	Remodeled	2307	495760.04	474078.26	0.34
1945-1970	Commercial - Residential	Remodeled	12	445876.93	370361.94	0.62
	Exempt	Original	1	432061.75	432061.75	0.17
1990-Present	Streets	Original	1	424145.51	424145.51	0.60

Top 10 Most Tax-Productive Development Patterns

WHERE WE ARE NOW

Our Next Step

1 Location Impact Analysis

- Define and analyze Central Business Districts (CBDs)
- Study property value relationships with CBD proximity
- Examine neighborhood-level patterns

2 Historical Development Patterns

- Analyze property value changes across development eras
- Assess remodeling impacts over time
- Map urban development evolution

3 Social Context Integration

- Integrate Census Block data (2010-2020)
- Examine relationships between:
  - Population density
  - Housing characteristics
  - Tax productivity
- Create 3D visualizations of key findings

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

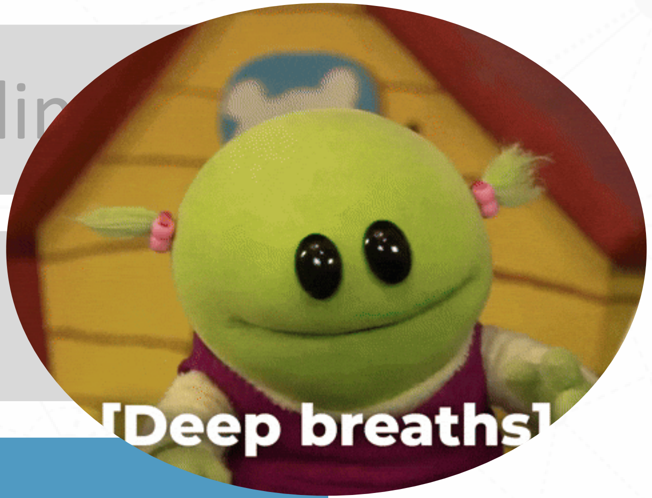
4. What we learned

1. Why are we doing this project

2. How we handled data wrangling

3. Where we are now

4. What we learned

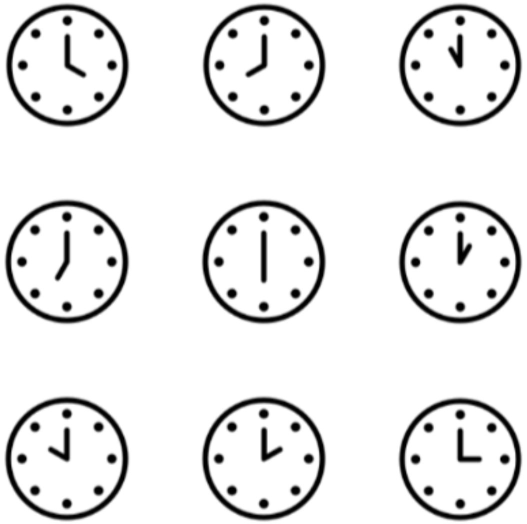


WHAT WE LEARNED

3 Key Takeaways from Our Project Process

1 Invest Time in Data Documentation

- When our team received updated data from the Auditor's office, we discovered data corruption issues
- Having detailed documentation of our original data source and processing steps proved invaluable



WHAT WE LEARNED

3 Key Takeaways from Our Project Process

1. Invest Time in Data Documentation

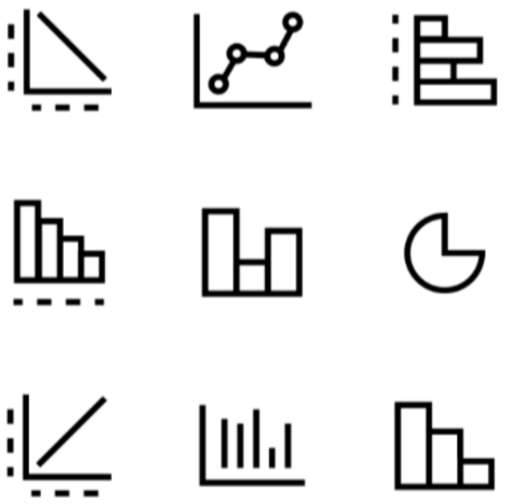
2. **Tackle Reproducibility from the Start**

- Clear file naming conventions - [YYMMDD]_descriptive-name(_initials-version)
- Standardized folder structures
- Version control practices
- Regular validation checkpoints

WHAT WE LEARNED

3 Key Takeaways from Our Project Process

1. Invest Time in Data Documentation
2. Tackle Reproducibility from the Start
3. **Data Wrangling is Iterative**



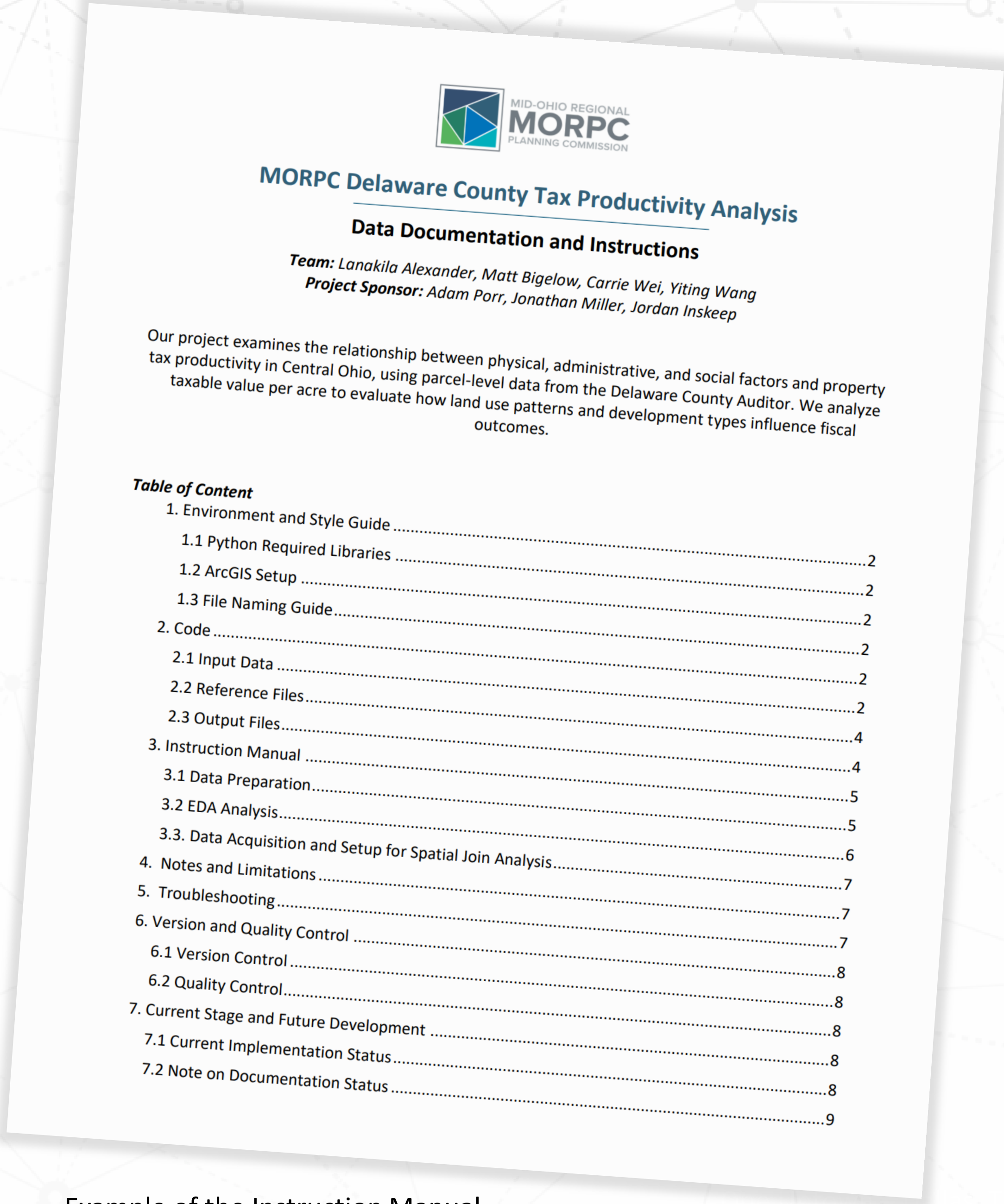
- Perfect data doesn't exist on day one
- Document assumptions and decisions
- Build flexible, modular code
- Plan for future updates

WHAT WE LEARNED
Our Tested Practices

Data Sources & Setup

Critical components:

- Standardized file structure and naming conventions
- Version control practices
- Following a Python Code style guide (PEP 8)
- Comprehensive documentation approach
    - Data Dictionary
    - Instruction Manual
    - Analysis Reports
    - Structured code notebooks



MID-OHIO REGIONAL MORPC PLANNING COMMISSION

MORPC Delaware County Tax Productivity Analysis

Data Documentation and Instructions

Team: *Lanakila Alexander, Matt Bigelow, Carrie Wei, Yiting Wang*
Project Sponsor: *Adam Porr, Jonathan Miller, Jordan Inskeep*

Our project examines the relationship between physical, administrative, and social factors and property tax productivity in Central Ohio, using parcel-level data from the Delaware County Auditor. We analyze taxable value per acre to evaluate how land use patterns and development types influence fiscal outcomes.

Table of Content

1. Environment and Style Guide 2
1.1 Python Required Libraries 2
1.2 ArcGIS Setup 2
1.3 File Naming Guide 2
2. Code 2
2.1 Input Data 2
2.2 Reference Files 4
2.3 Output Files 4
3. Instruction Manual 5
3.1 Data Preparation 5
3.2 EDA Analysis 6
3.3. Data Acquisition and Setup for Spatial Join Analysis 7
4. Notes and Limitations 7
5. Troubleshooting 7
6. Version and Quality Control 8
6.1 Version Control 8
6.2 Quality Control 8
7. Current Stage and Future Development 8
7.1 Current Implementation Status 8
7.2 Note on Documentation Status 9

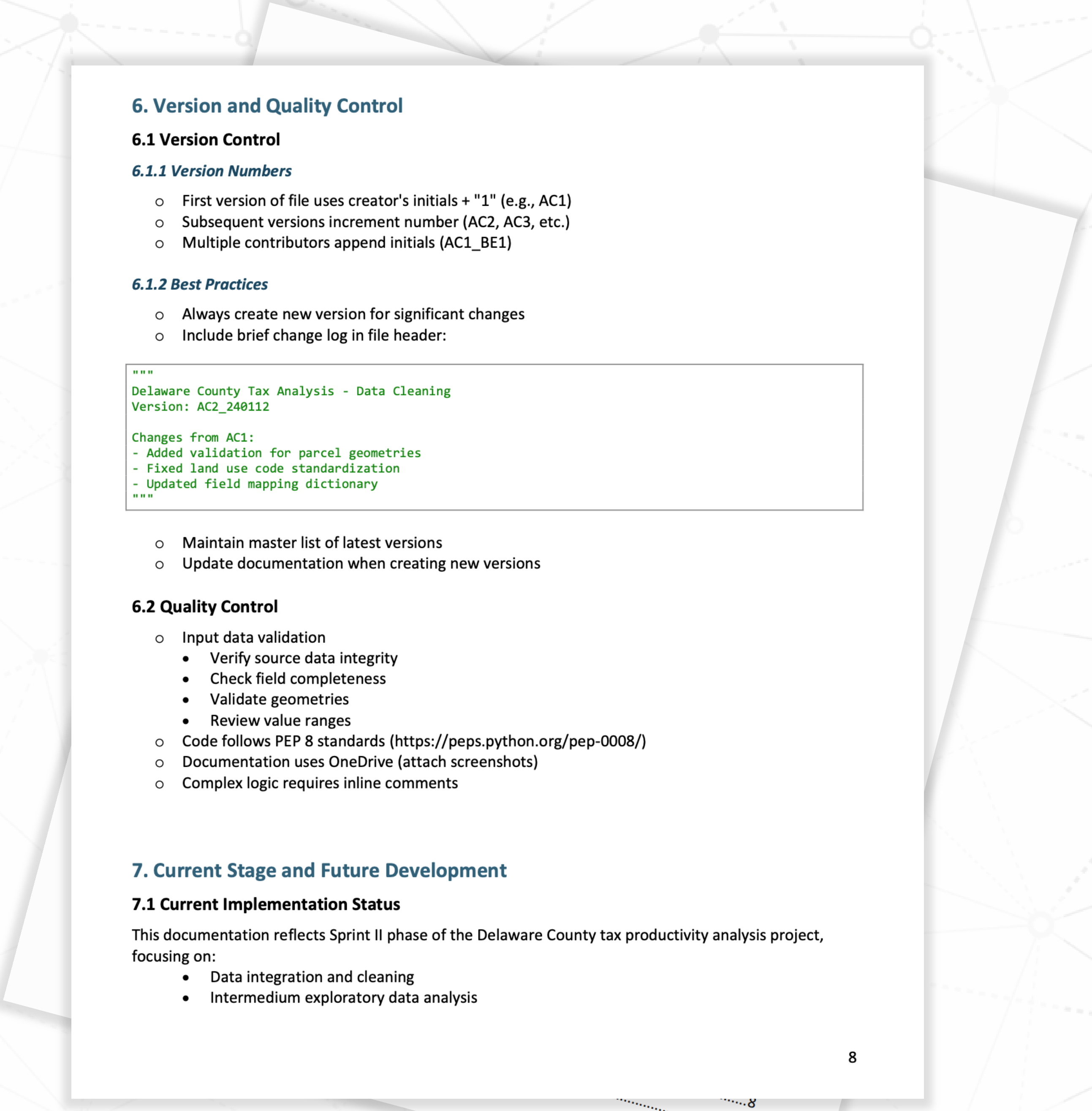
Example of the Instruction Manual

WHAT WE LEARNED
Our Tested Practices

Data Sources & Setup

Critical components:

- Standardized file structure and naming conventions
- Version control practices
- Following a Python Code style guide (PEP 8)
- Comprehensive documentation approach
    - Data Dictionary
    - Instruction Manual
    - Analysis Reports
    - Structured code notebooks



6. Version and Quality Control

6.1 Version Control

6.1.1 Version Numbers

- First version of file uses creator's initials + "1" (e.g., AC1)
- Subsequent versions increment number (AC2, AC3, etc.)
- Multiple contributors append initials (AC1_BE1)

6.1.2 Best Practices

- Always create new version for significant changes
- Include brief change log in file header:

```
"""
Delaware County Tax Analysis - Data Cleaning
Version: AC2_240112

Changes from AC1:
- Added validation for parcel geometries
- Fixed land use code standardization
- Updated field mapping dictionary
"""
```

- Maintain master list of latest versions
- Update documentation when creating new versions

6.2 Quality Control

- Input data validation
    - Verify source data integrity
    - Check field completeness
    - Validate geometries
    - Review value ranges
- Code follows PEP 8 standards (https://peps.python.org/pep-0008/)
- Documentation uses OneDrive (attach screenshots)
- Complex logic requires inline comments

7. Current Stage and Future Development

7.1 Current Implementation Status

This documentation reflects Sprint II phase of the Delaware County tax productivity analysis project, focusing on:

- Data integration and cleaning
- Intermedium exploratory data analysis

8

Example of the Instruction Manual

WHAT WE LEARNED

Our Tested Practices

Collaborative Infrastructure

Team coordination through:

- Defined roles and responsibilities
    - Sprints Planner
- Clear communication channels
- Shared workspace organization
- Regular work review session with team members
- Weekly sponsor check-ins

Sprint Zero	Business Understanding with Initial Exploratory Data Analysis
Goal 1	Ensure data accuracy by validating all fields, aiming for high reliability, with weekly updates on data quality issues and corrections.
Goal 2	Refine the current data dictionary to document clear definitions, data types, and value ranges, maintain consistency across the project, and review on a regular basis.
Goal 3	Generate key descriptive plots for critical variables, update visualizations as needed, and share progress weekly with the team and the sponsor.
Goal 4	Conduct statistical analysis on major metrics, documenting methods and findings in OneNote team work space, and revising weekly as necessary.
Goal 5	Complete 100% of assigned tasks within the sprint, tracking progress weekly to ensure timely completion for all team members.

	Task Description				
Identified Task	Task Size: Amount of Effort Required (Minimum/Moderate /Significant)	Task Complexity (Low/Medium/High)	Risk/ Uncertainty (Low/Moderate/High)	Final Story Point Assignment 2 - Minimum 5 - Moderate 9 - Significant	Notes (Constraints, resources needed, unknowns, etc)
Data Wrangling: Remove incorrect, irrelevant, or duplicate data, handle missing values by filling them with appropriate replacements or dropping rows depending on the situation.	Moderate	Medium	High	9	
Data Wrangling: Join Datasets as appropriate	Minimum	Medium	High	5	Be aware of previous duplication issues
Data Transformation: Change the format or structure of data to make it suitable for analysis, including converting data types, standardizing units, or creating new features.	Moderate	Medium	Low	9	To ensure reproducibility, do not change datatypes without team discussion on how to integrate changes into master code
Code: Develop coding standards to facilitate reproducible and flexible code product (i.e. style guide for code and code-produced graphics)	Moderate	Medium	Low	5	Graphic Outputs: Seaborn, Matplot, Tableau, ArcGIS Function structure
Gross Categorical Analysis: Descriptive statistics	Moderate	Low	Low	2	
Gross Categorical Analysis: Create plots/tables for end-of-term presentation	Moderate	Low	Low	2	
Residential Numerical Analysis: Distribution, Descriptive Statistics, Observations	Moderate	Medium	Low	5	
Residential Categorical Analysis: Distribution, Descriptive Statistics, Observations	Moderate	Medium	Low	5	
Residential Analysis: Detecting and Treating Outliers	Minimum	Medium	Moderate	5	
Residential Analysis: Create complete plots/tables for end-of-term presentation & report	Moderate	Medium	Low	5	

Example of the Sprint Planning Sheet

WHAT WE LEARNED

Key Considerations to Ensure Reproducibility & Effective Project Management

For Data Practitioners - **3 Key Questions To Ask:**

1. Can someone else pick up where I left off?
2. Could they replicate my analysis without asking for help?
3. Is my documentation answering the 'why' as well as the 'how'?

WHAT WE LEARNED

Key Considerations to Ensure Reproducibility & Effective Project Management

For Project Management - **3 Essential Considerations:**

1. Invest in documentation early
2. Build in regular review processes
3. Plan for project handoff from **day one**

Thank you!

What questions do you have?